FM Crystal Radios?

copyright © 2006 by Larry J Solomon

I have heard, even from a physicist, that it is impossible to build FM crystal radios. On the other hand some experimenters claim that they have built them. This argument intrigued me to try and build an FM crystal radio, which I have done successfully. To my surprise, the result is an astounding performer, pulling in four local stations in Tucson. When connected as a receiver to a good sound system the sound fidelity is as good or better than more expensive AM radios. In fact, it sounds "high-fidelity".

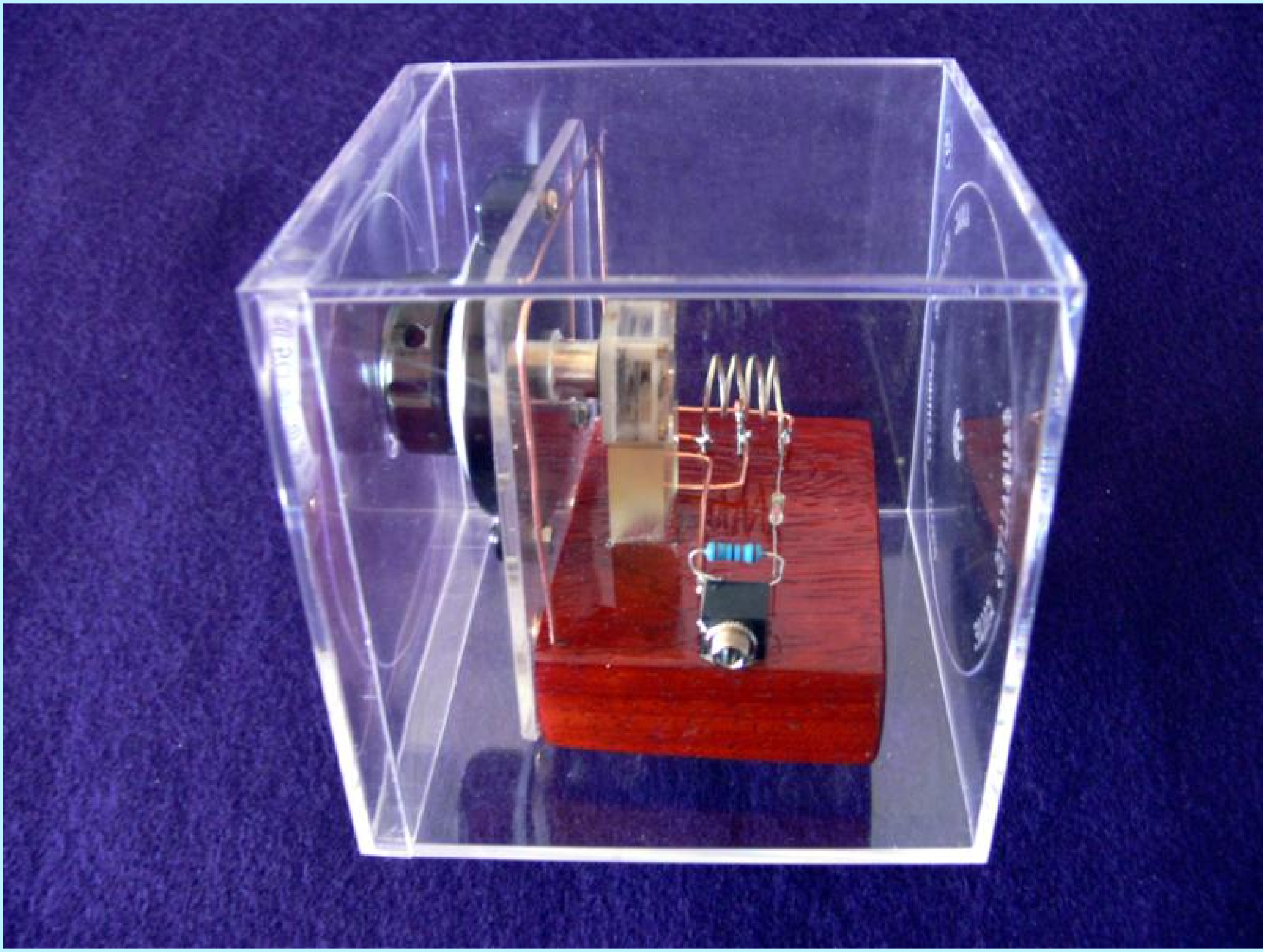

This picture shows the Solomon FM Crystal Set in an acrylic display case. I made the set specifically to fit inside this case (the case came first).

My definition of a crystal radio is one that is not powered, except by the radio transmission itself and employs a crystal detector. So, it should work without any batteries or AC power. An FM crystal receiver must be able to detect and receive FM signals well enough to be heard in earphones without any such extra power.

This FM receiver is an amazing performer. It has crystal clear reception (pun intended), good sensitivity, but only fair selectivity. This set was a discovery for me. I started out by designing and building the normal AM sets. Then one day while testing the "Mystery" set (see my other web links), to my surprise, in addition to the expected panoply of AM stations, I heard a very faint signal that I could not tune out. At first, it seemed too weak to identify. When I tuned out all the AM stations, I was astonished to hear the announcement "KiiM FM, 99.5"! This is a

country music FM station here in Tucson. It was all over the dial, untunable, but the much louder AM signals masked it when they were tuned in.

I set myself the task of trying to improve the FM reception. I tried some simple circuit modifications that did not seem to improve anything. Then I connected a dipole antenna instead of the AM antenna I normally use. Suddenly, the FM signal was much clearer, although still weak. By using the audio output and sound system amplifier, I was even more amazed that four different FM stations came in loud (or rather medium) and clear. I found that changing the telescoping antenna length and position I could tune the stations in and out. They were KRQ, KLPX, KiiM, and KHYT all local FM stations with transmitters nearby. Their reception was also affected by the length and position of the audio output cable.

After doing some research, I discovered that there was a physical theory that claimed that FM reception was possible and even probable using the same circuit as an AM receiver. The theory is called "slope detection". So, I set out to find circuit improvements. A web search yielded little, mostly theory. But there was enough information that I thought I could make some modifications to the AM circuits to make them more tunable to FM signals and less tunable to AM. Since FM operates at higher frequencies, all I had to do, I thought, was make the coil and caps smaller. After much "tinkering" I arrived at the current circuit.

The circuit looks identical to a classic AM crystal circuit but is even simpler to build. The components were reduced in dimension to resonate at higher frequencies. This was done by experimenting with smaller and smaller coils and capacitors. The antenna is also much reduced in size (from that of AM) to resonate at higher frequencies (the antenna is crucial). The air variable capacitor I used has two trimmers in it which should be adjusted for best reception.I have found that a commonly available vernier dial and knob will fit the capacitor nicely. See end of article for a picture of the variable. C3 is a ceramic capacitor of 18 pf, but may be anywhere from 10 to 50pf. A detected FM signal is converted to AM due to an effect called slope detection that modulates amplitude.

This FM Crystal Set works best near the transmitter (I have not tested it beyond about 10 miles). Secondly, the sound level is quiet, especially without an amplifier. A quiet room is needed for listening with earphones. One must be willing to move the set around to find a location for the best reception of signals. However, in addition to listening with high impedance earphones (crystal or otherwise), the set can be connected directly to an audio amplifier's low level magnetic input which can then play amplified through a sound system at any volume -- sounds GREAT. In fact, I recommend starting tests with the FM crystal set by connecting it to the low-level phono inputs of a receiver or preamplifier. (Nowadays, many receivers don't even have a phono input!) That way you can crank up the volume, which makes it more likely to find the FM stations. If no signals are detected, I also recommend connecting an external "rabbit ear" antenna or hanging a short wire (12 inches or so) in various positions next to the internal antenna. The variable length of rabbitt ears can help to tune in stations.

No additional wiring or antenna is necessary (the antenna is optimized in length for FM.)

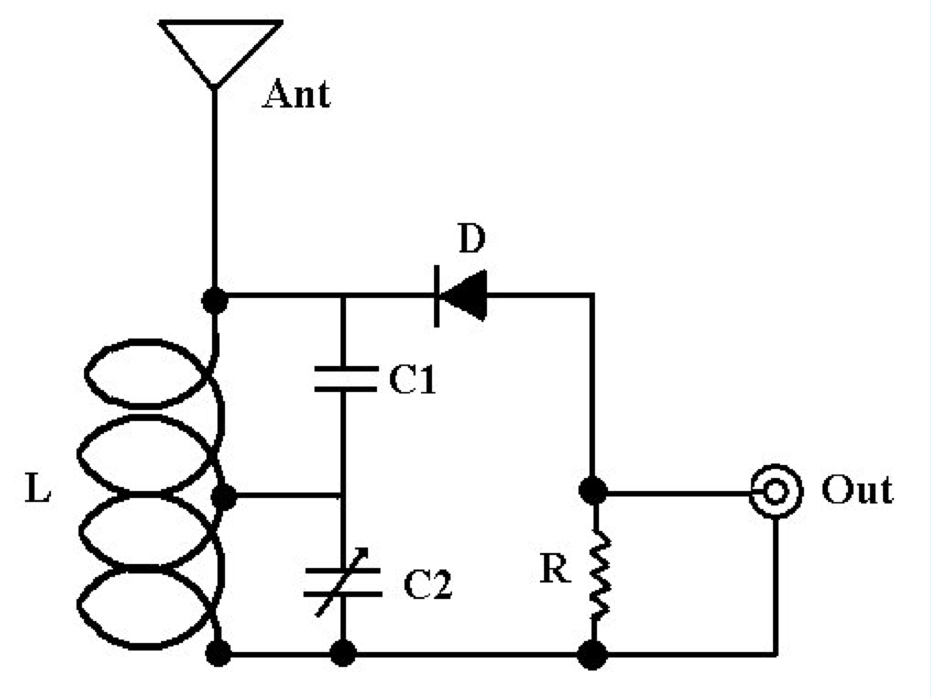

L - 4 turns #18 copper or silver wire, 12mm inside diameter, tapped at 2.5 turns
Ant - 7 inches of #18 bare copper wire
C1 - 18 pf ceramic capacitor
C2 - 50 pf air variable capacitor
D - 1N34 diode or rock crystal
R - 150K resistor

The diode is tapped directly to the antenna. The vernier dial fits directly on the tuning capacitor. The antenna parallels the perimeter of the acrylic face plate. "Military style" #18 AWG wiring is used without any insulation. It is important to keep the components physically close together. The component specifications are the same as in circuit #2. The coil is silver rather than copper, but copper does just as well. I think that the contrast of the silver and copper is beautiful. The coil was wrapped around a Sharpie Permanent Marker, then slipped off and expanded slightly. The wooden base is made from lacquered, polyurethane padouk.

I consider this set a work of art as well as science and think it is the most elegant crystal receiver I have created. I love the contrast of the silver coil, the copper antenna, the clear acrylic faceplate, the black vernier dial, the white and transparent variable capacitor, and the subtle colorings on the resistor, the diode, and the lucite base. Yet the circuit is so ridiculously simple that some will not believe it is possible without building it themselves. No shielding is necessary, and there is no problem with hand capacitance. However, the output cable position may affect reception sensitivity.

Photos of wired circuit

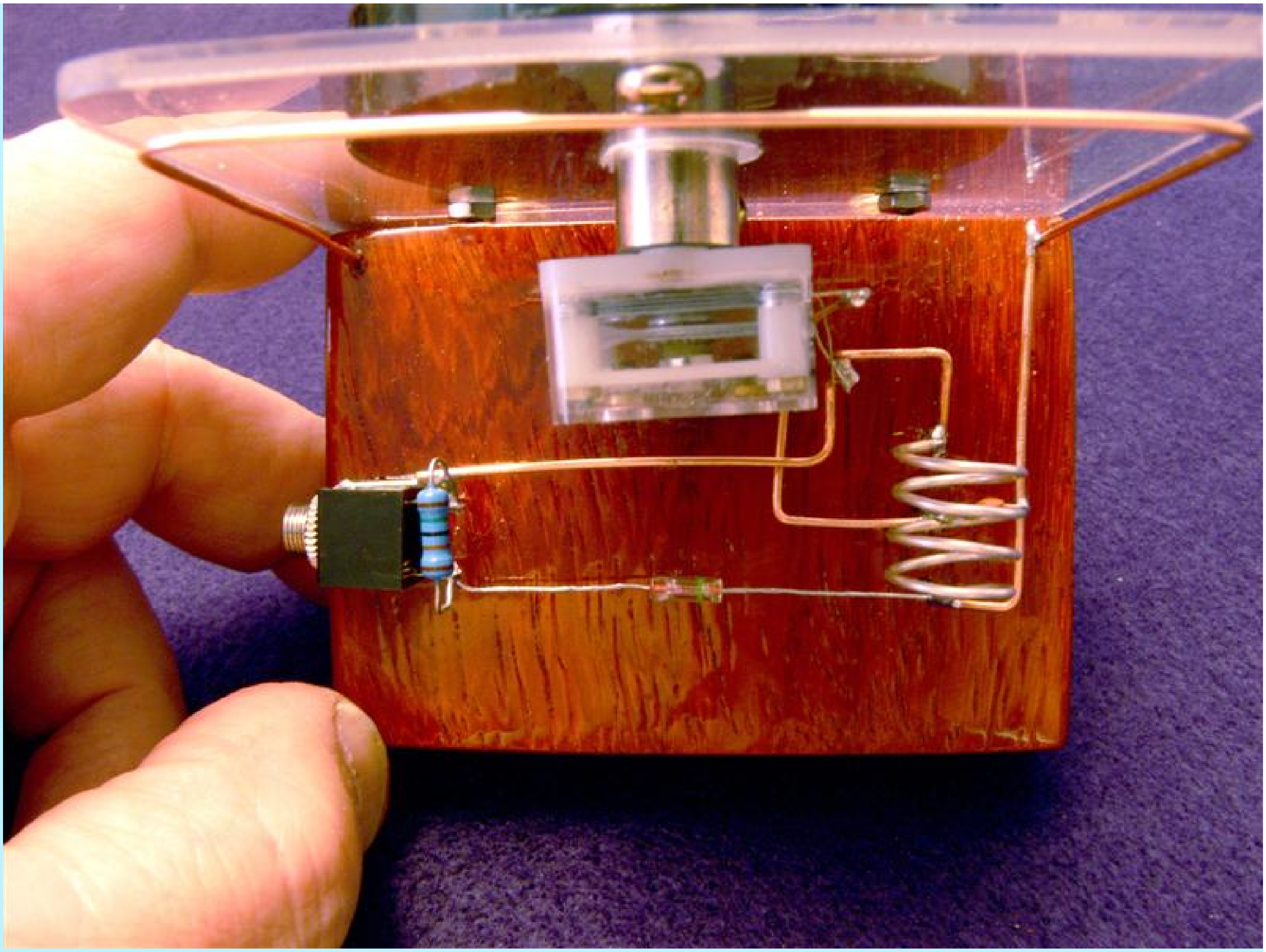

A hand is included in this photograph to show scale. Note the military style wiring, diode, and antenna. I wanted the wiring to create a modern design similar to a Mondrian painting. Not only is this set beautiful, it works! No power and no long antenna! It looks like a work of fiction.

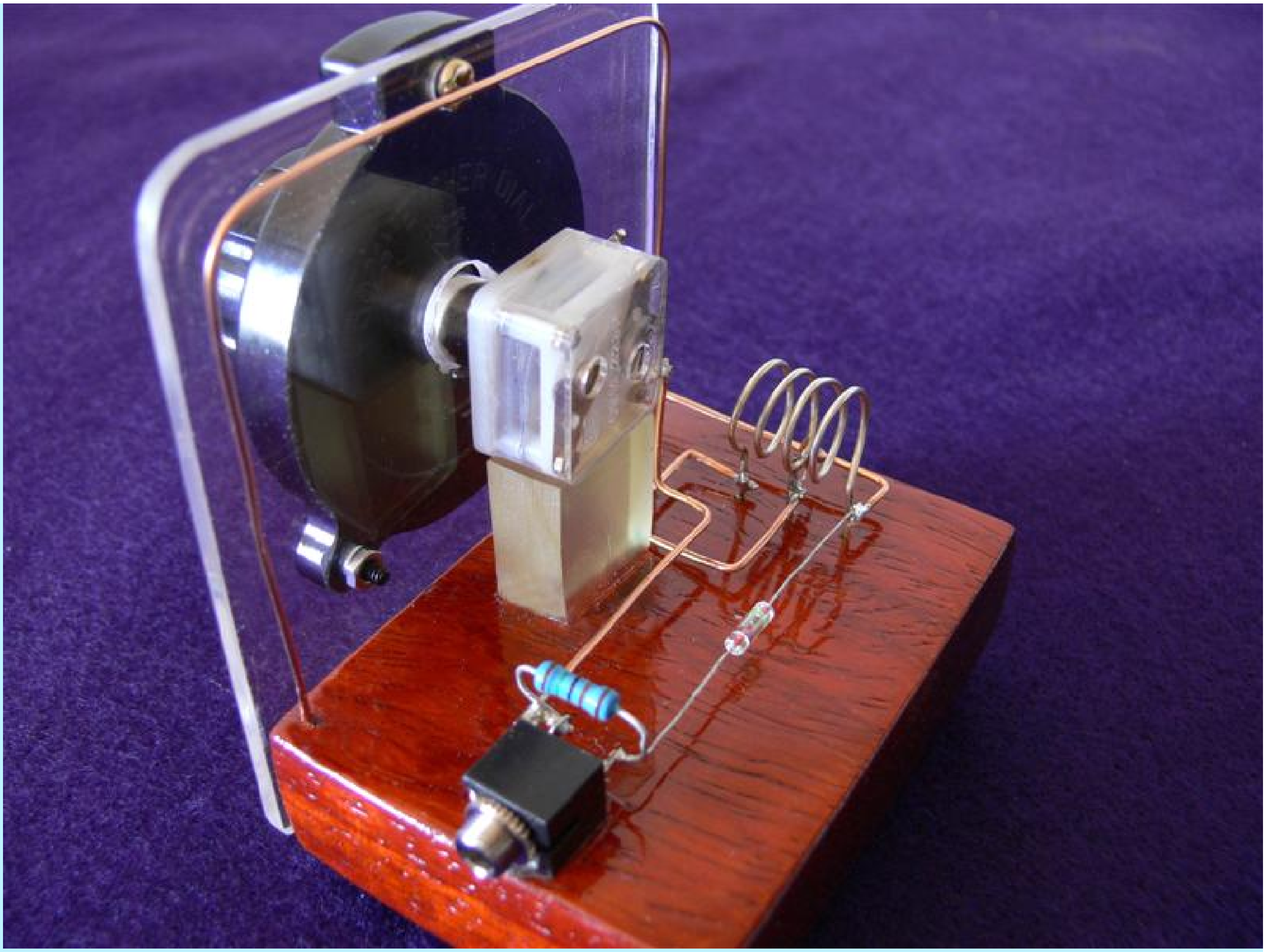

Is this thing imaginary -- science fiction? Well, imagination did play a part, but it is definitely not science fiction. This shot shows the elegance of the FM set best, I think. There is only one resistor and one fixed capacitor.

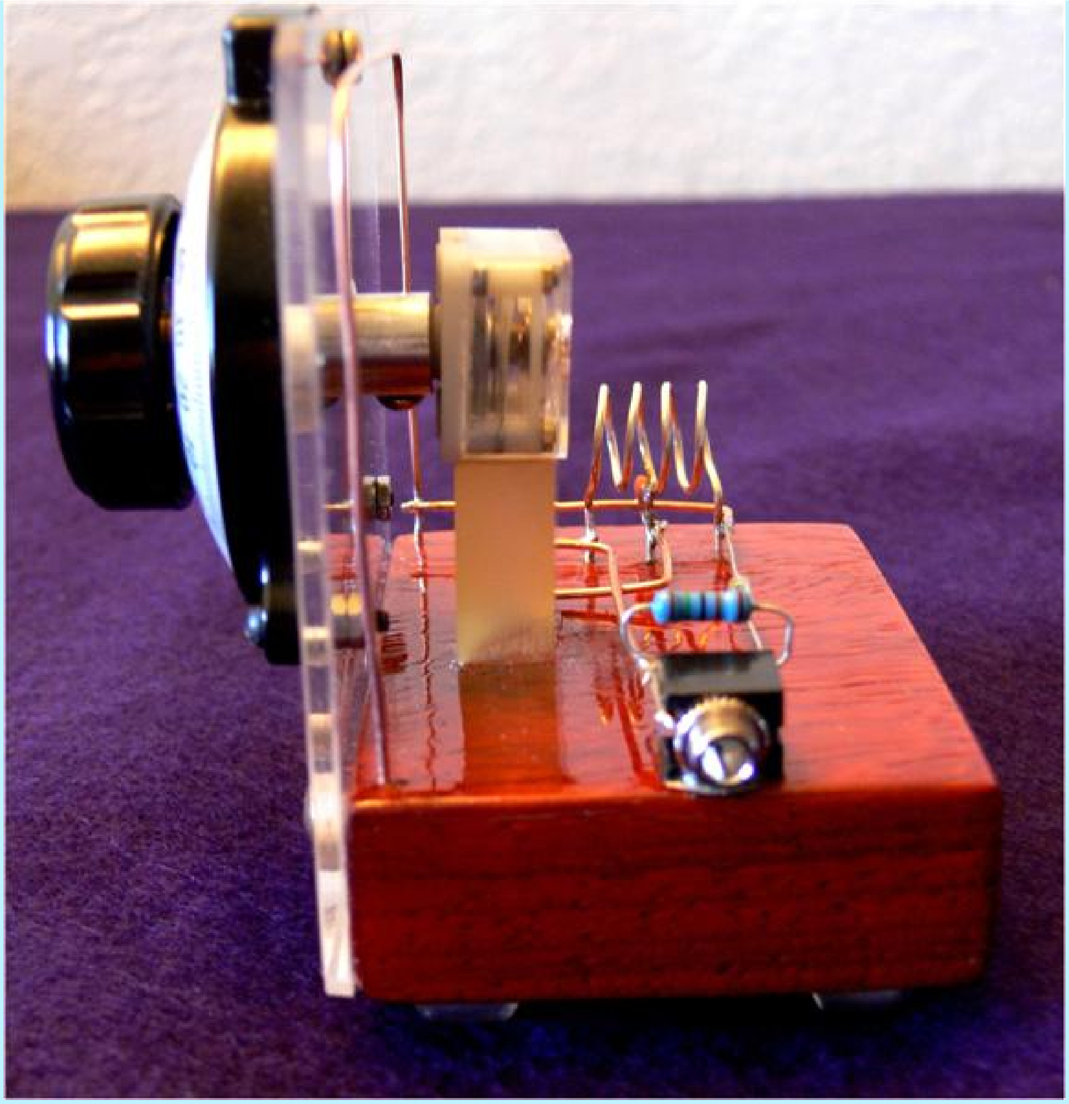

The inside of the tuning capacitor and the phono jack/output can be seen here. Can you spot the fixed ceramic capacitor? Note the polished edge of the face plate and the reflection in the wooden base.

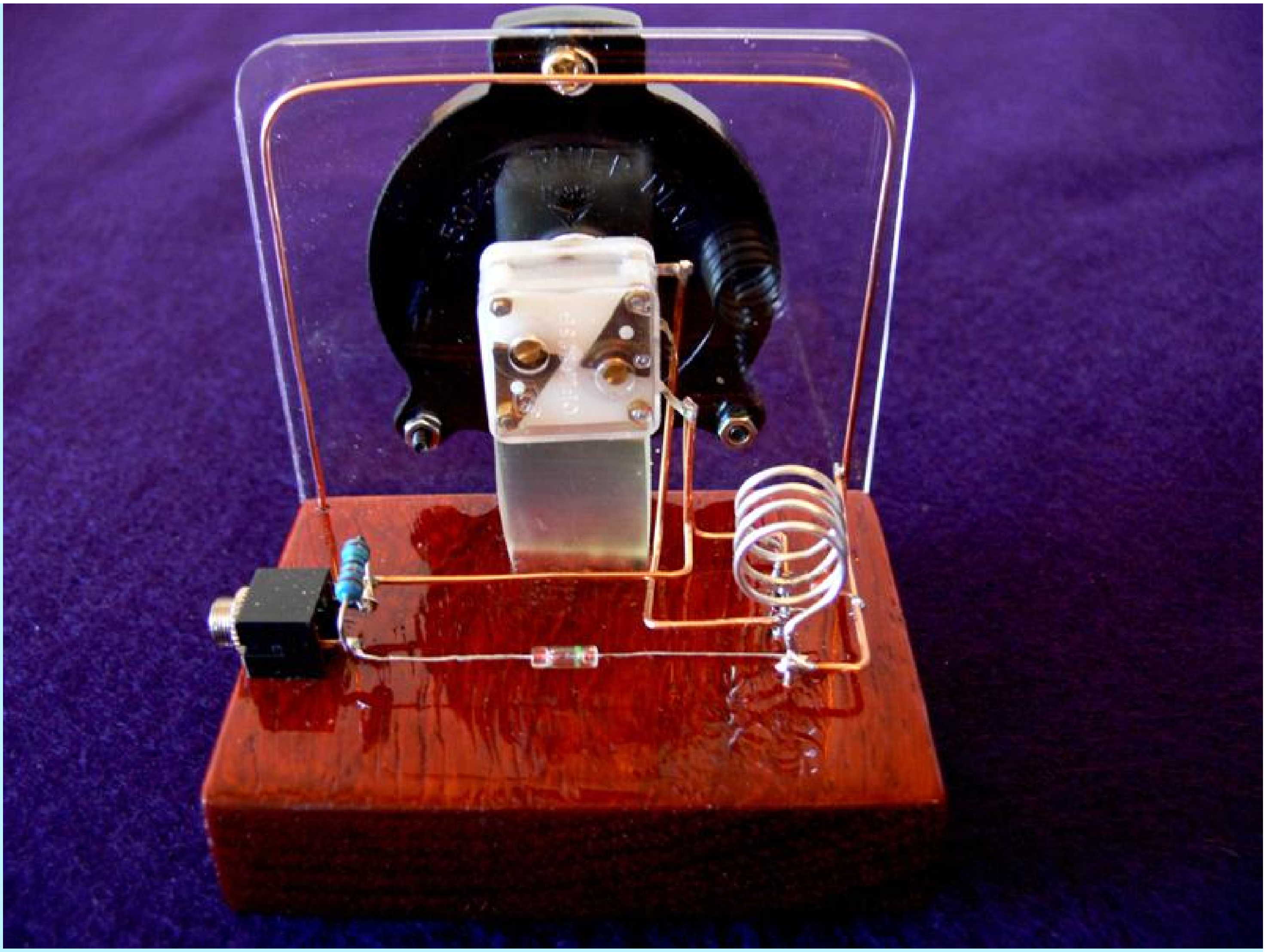

A quarter-inch piece of lucite was fitted under the tuning capacitor to anchor it. Note the two tiny trimmers on the back of the tuning capacitor. Brass screws were used to enhance appearance.

100
90
80
70
60
50
40
30
20
10
0

The vernier dial is large to accomodate ease of tuning, and the vernier makes it easy to separate stations. Two golden (brass) wood screws fix the face plate to the base. Holes for the face plate were made with special plastic drills, but ordinary drills may be used if drilled very SLOWLY. The knob is removable.

FM Crystal Circuit #2

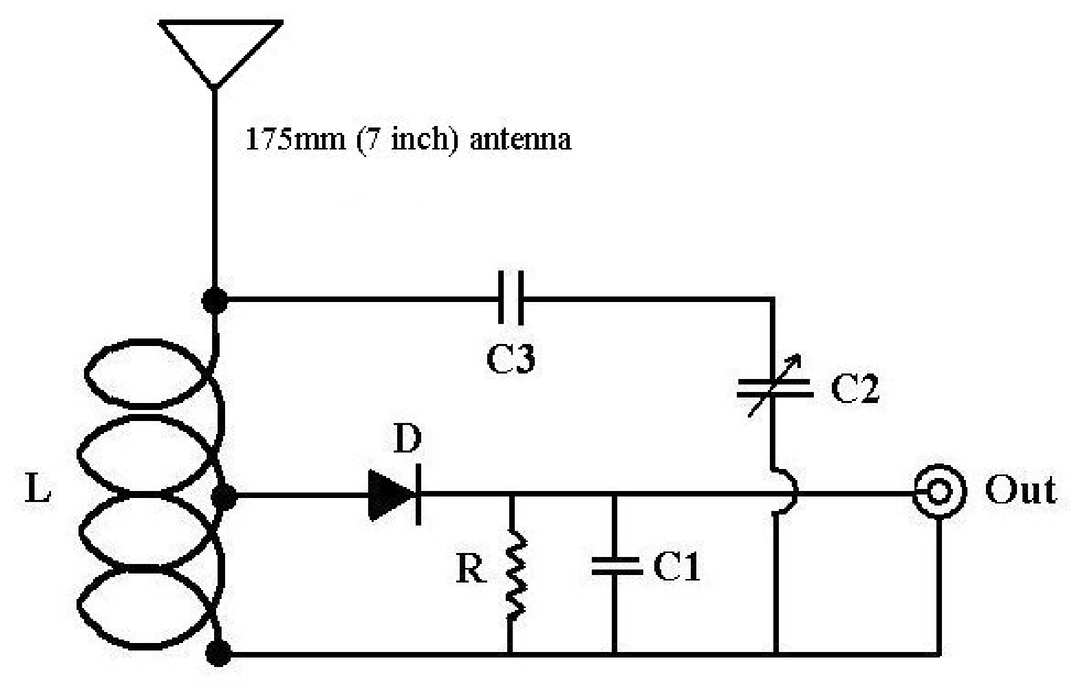

L - 5 turns AWG#18 bare copper or silver wire, 12mm inside diameter, tapped at 2.5 turns
D - 1N34 or rock crystal diode
C1 - 82 pf capacitor
C2 - 80 pf air variable capacitor
C3 - 18 pf capacitor
R - 150K resistor

The following photographs show the circuit wired with the handmade Saturn Dial. and knob. It is perhaps not as visually striking as set No. 1, but it works just as well. In fact, this set was the original version. Notice that all the wiring and coil are copper.

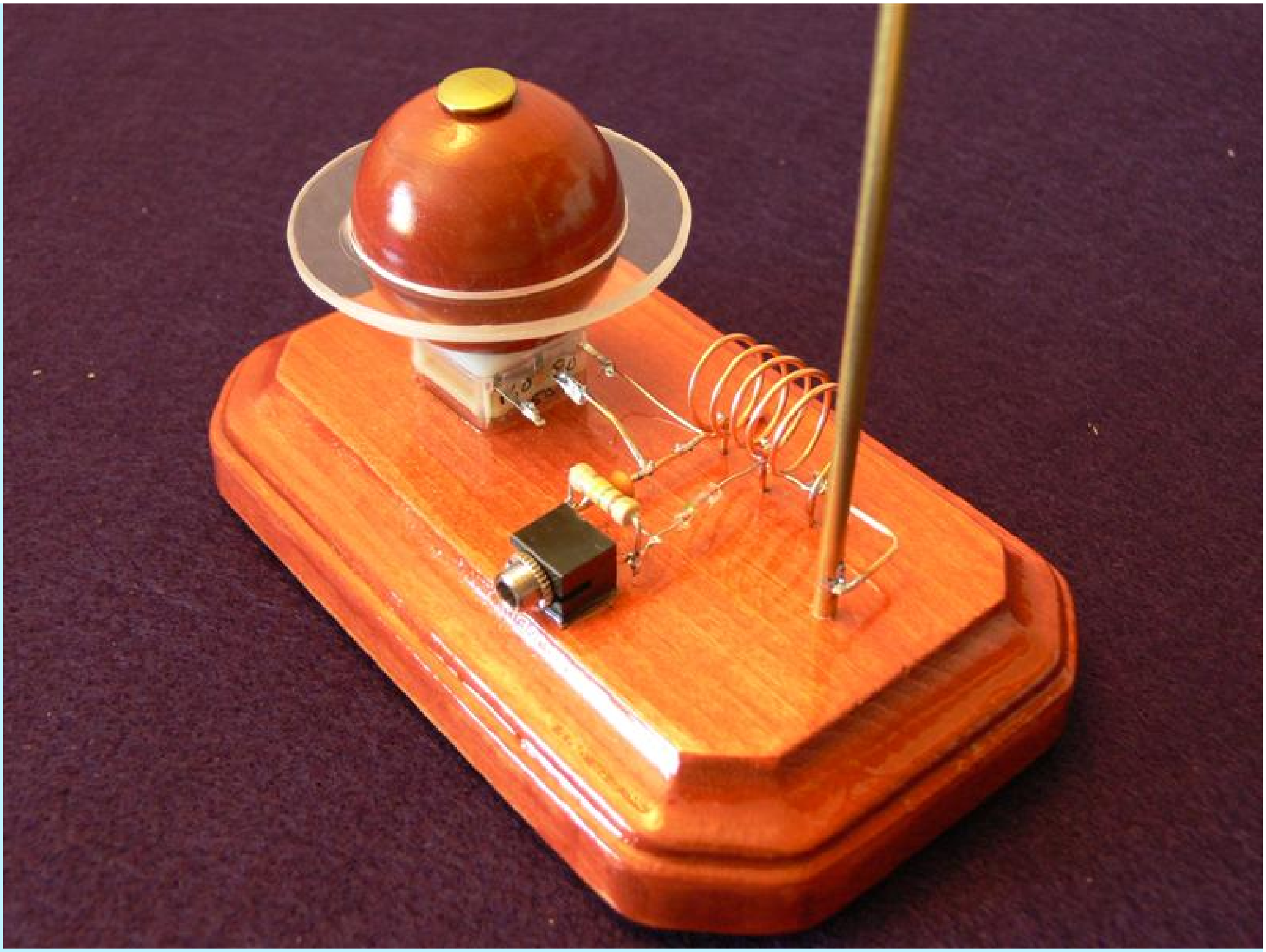

The Saturn dial and knob were fashioned from a "doll's head" from Michael's Arts and Crafts, a piece of lucite cut with two circle cutters, and a brass paper fastener. The knob is fixed to the tuning capacitor with a small machine screw that fits in the hole below the brass fastener. The most difficult part of this was fashioning "Saturn's rings". This must be done very carefully and slowly. The inside edge should be cut slightly undersized and then sanded with a drum sander to fit snugly. The outside edges can be sanded with fine sandpaper and polished with a plastic polisher.

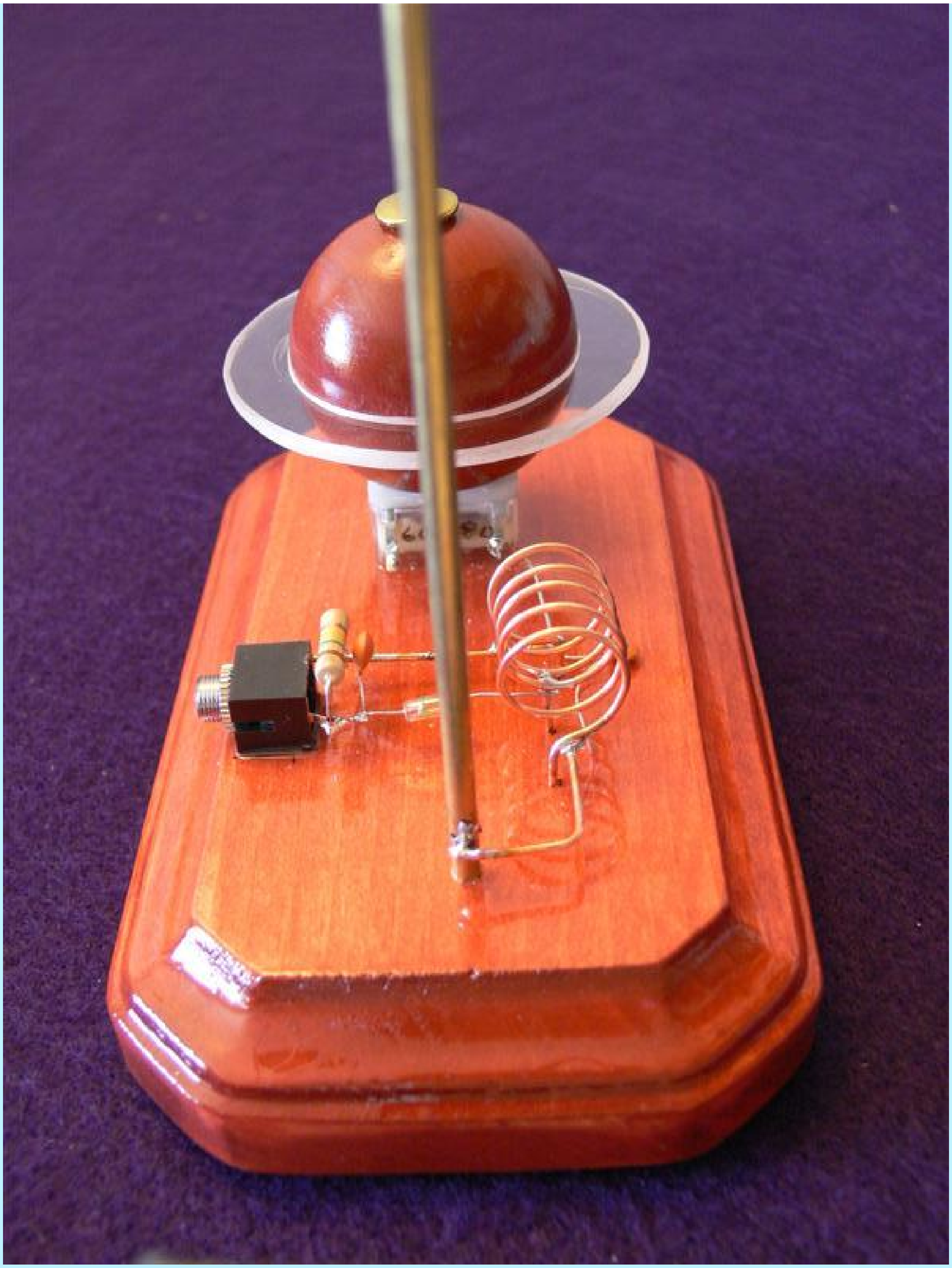

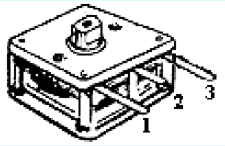

The air variable capacitor may be obtained from Electronix Express at http://www.elexp.com/. Part number 14VCRF10-280P. The 80 pf side is recommended for the second circuit, contacts 2-3. Contacts 1 and 3 were used for the first circuit (50pf).

- OSC: 5-59 pf
- ANT: 5-142 pf
- OSC and ANT Trimmer 10pf range

Home -- If you want to contact me go to the home page and go through the Music Clinic link.

aesthetika

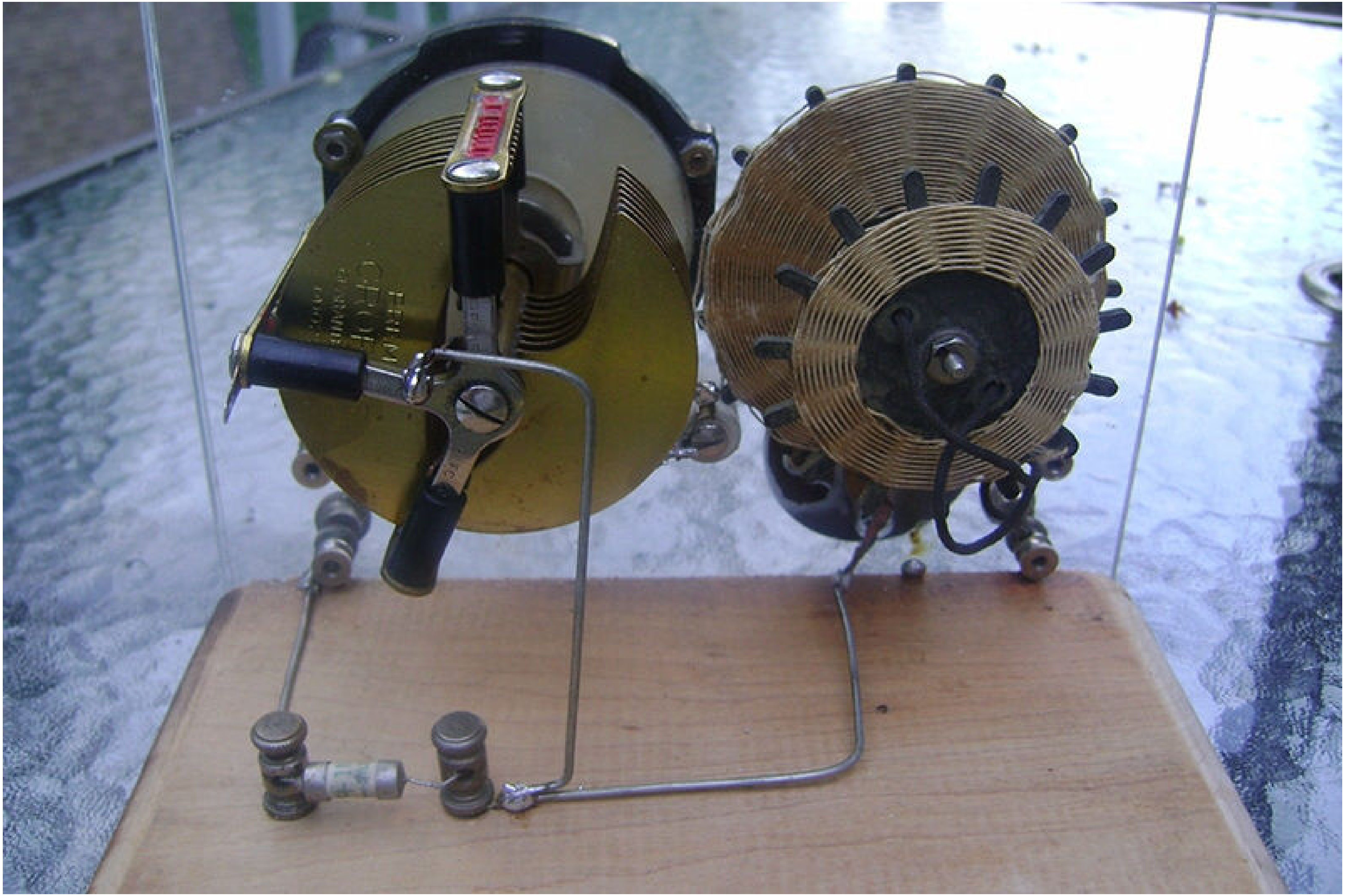

BLACK CUBE- AUSTRALIA

Manhattan, NYC

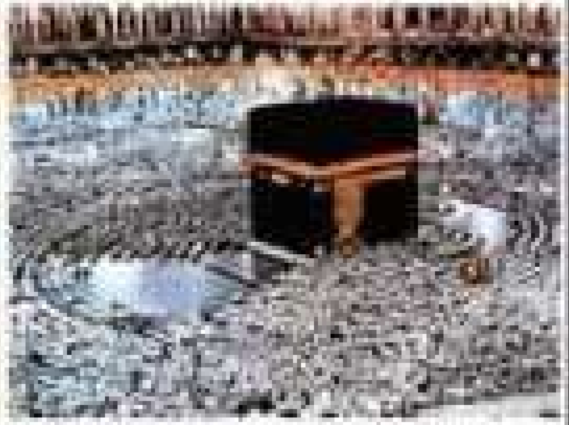

Kaba in Mecca

BLACK CUBE- DENMARK

Aple Store cube- NYC

"Rings going around SATURN"

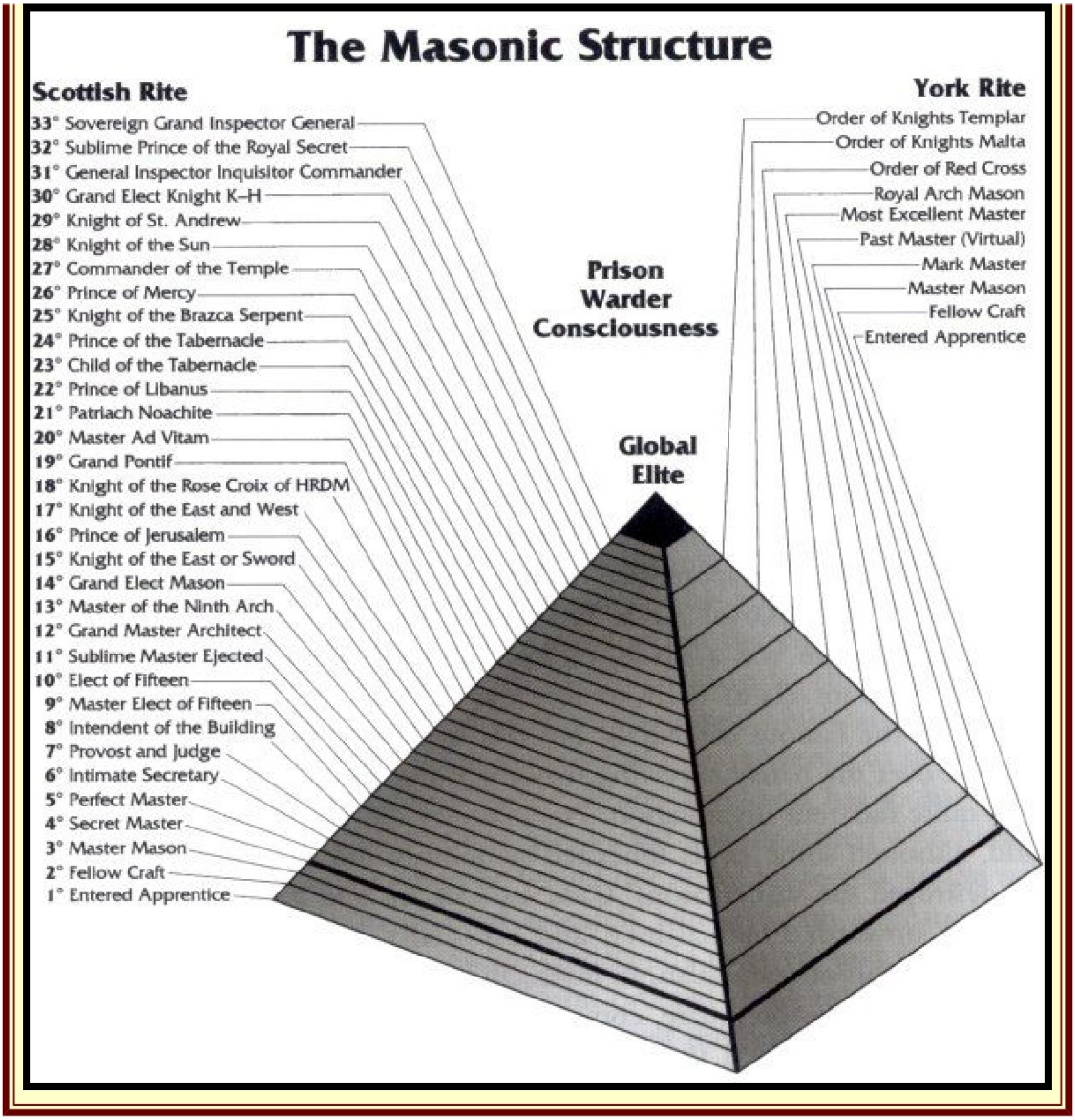
The Masonic Structure
Scottish Rite
33° Sovereign Grand Inspector General
32° Sublime Prince of the Royal Secret
31° General Inspector Inquisitor Commander
30° Grand Elect Knight K–H
29° Knight of St. Andrew
28° Knight of the Sun
27° Commander of the Temple
26° Prince of Mercy
25° Knight of the Brazca Serpent
24° Prince of the Tabernacle
23° Child of the Tabernacle
22° Prince of Libanus
21° Patriach Noachite
20° Master Ad Vitam
19° Grand Pontif
18° Knight of the Rose Croix of HRDM
17° Knight of the East and West
16° Prince of Jerusalem
15° Knight of the East or Sword
14° Grand Elect Mason
13° Master of the Ninth Arch
12° Grand Master Architect
11° Sublime Master Ejected
10° Elect of Fifteen
9° Master Elect of Fifteen
8° Intendent of the Building
7° Provost and Judge
6° Intimate Secretary
5° Perfect Master
4° Secret Master
3° Master Mason
2° Fellow Craft
1° Entered Apprentice
Prison Warder Consciousness
Global Elite
York Rite
Order of Knights Templar
Order of Knights Malta
Order of Red Cross
Royal Arch Mason
Most Excellent Master
Past Master (Virtual)
Mark Master
Master Mason
Fellow Craft
Entered Apprentice

FIG. 9 TERMINAL STRIP

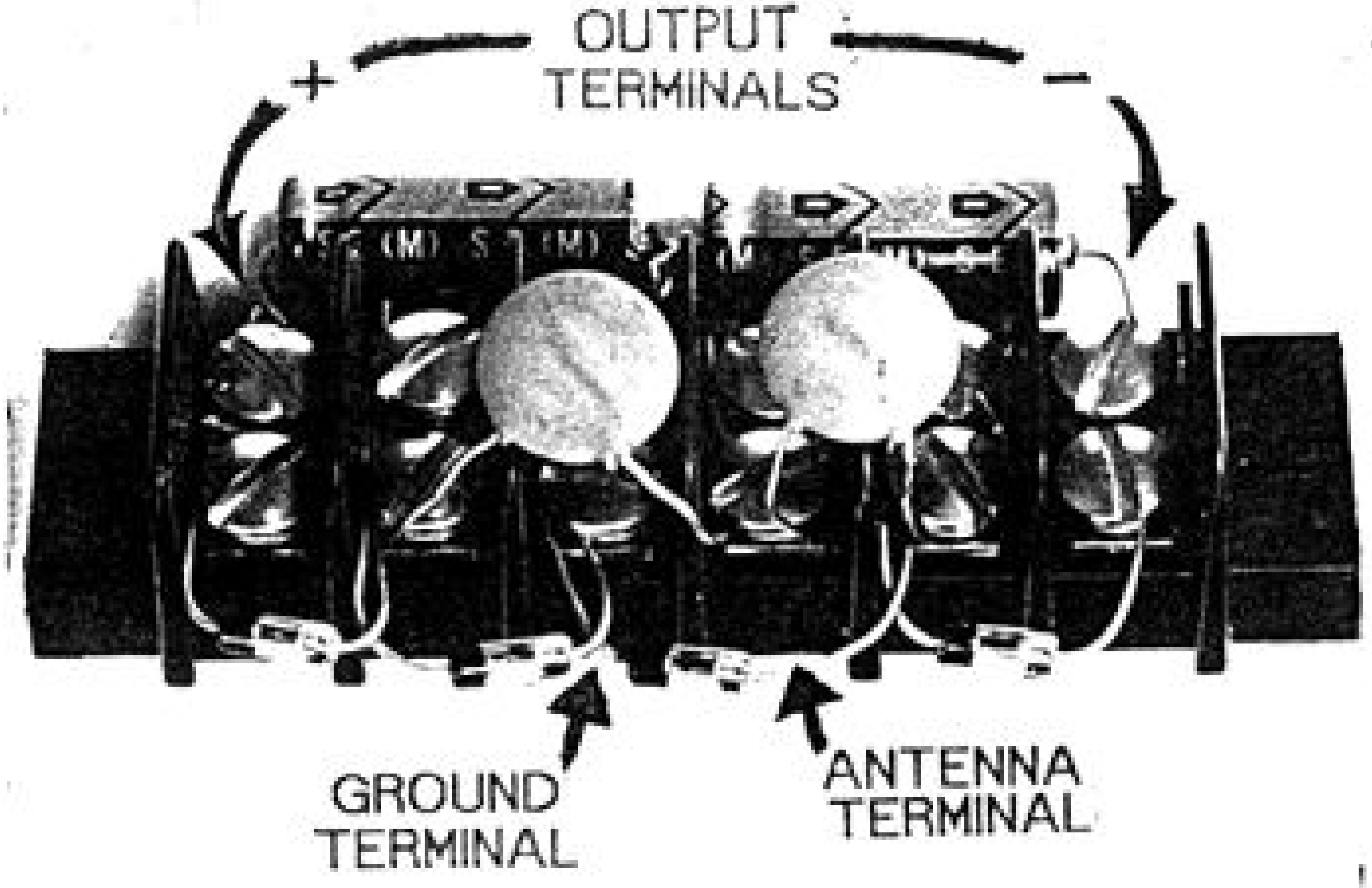
+
OUTPUT TERMINALS
−
GROUND TERMINAL
ANTENNA TERMINAL

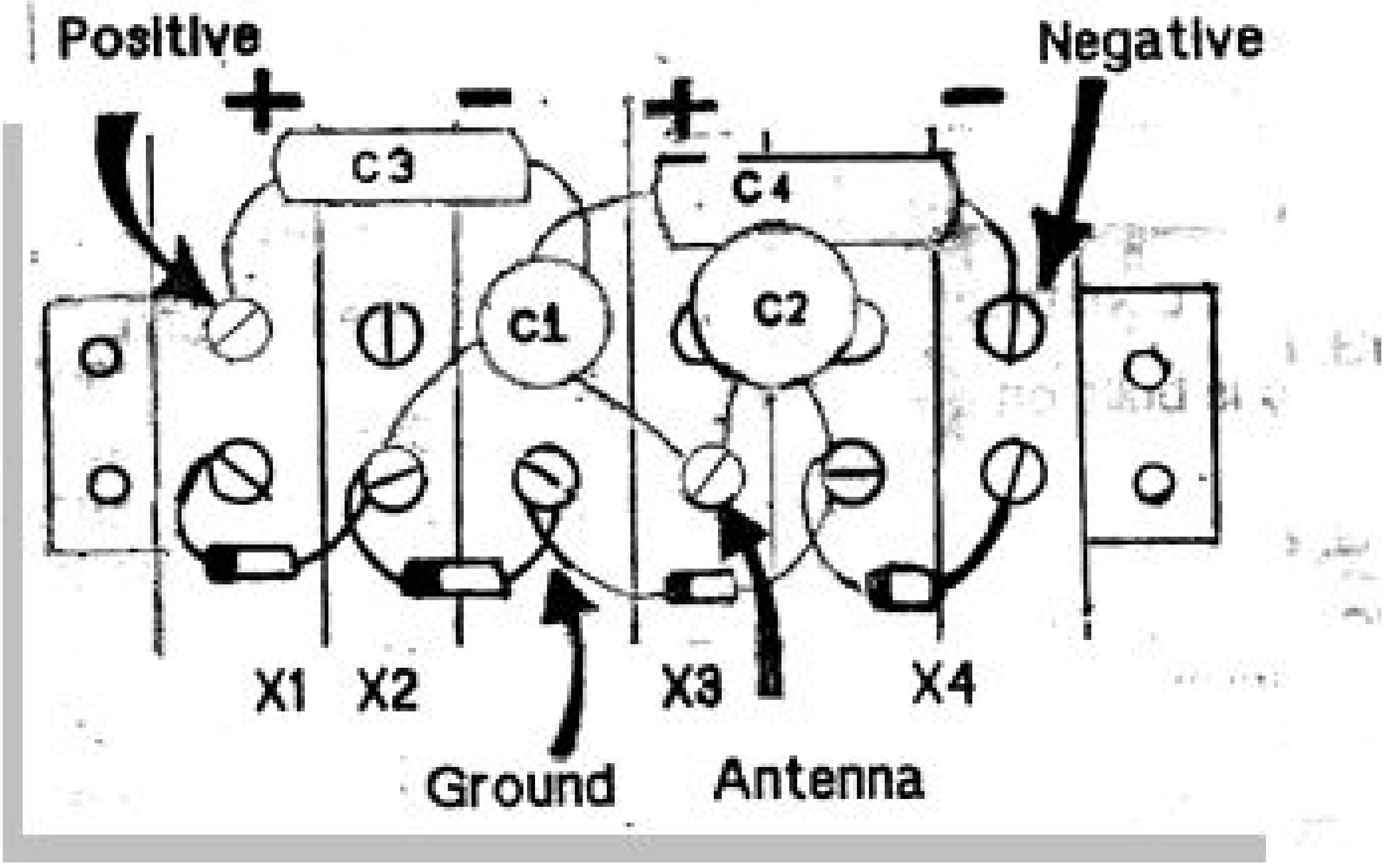
Positive
+
−
+
−
Negative
C3
C4
C1
C2
X1
X2
X3
X4
Ground
Antenna

FIG. 9 TERMINAL STRIP

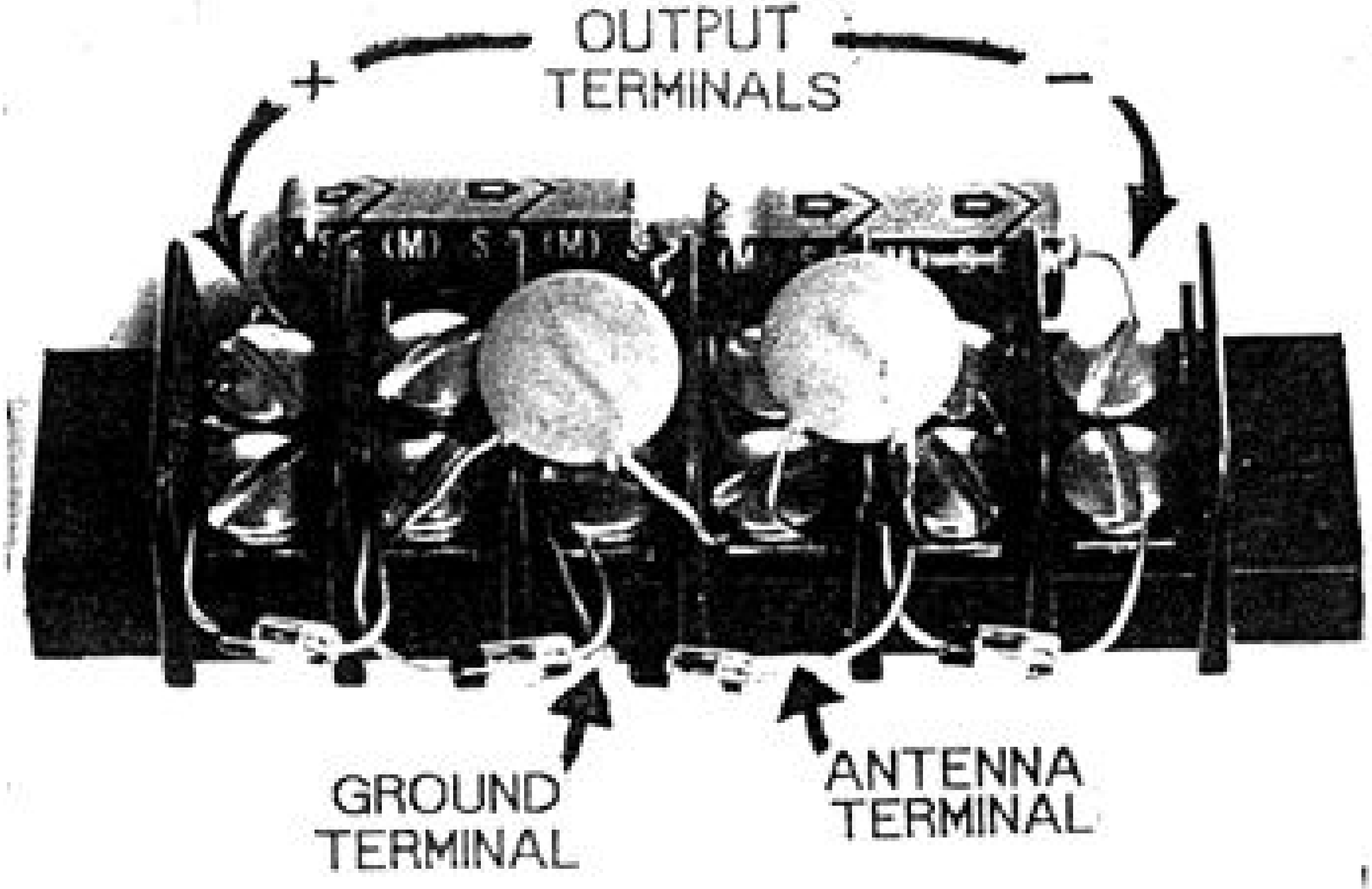
+
OUTPUT TERMINALS
−
GROUND TERMINAL
ANTENNA TERMINAL

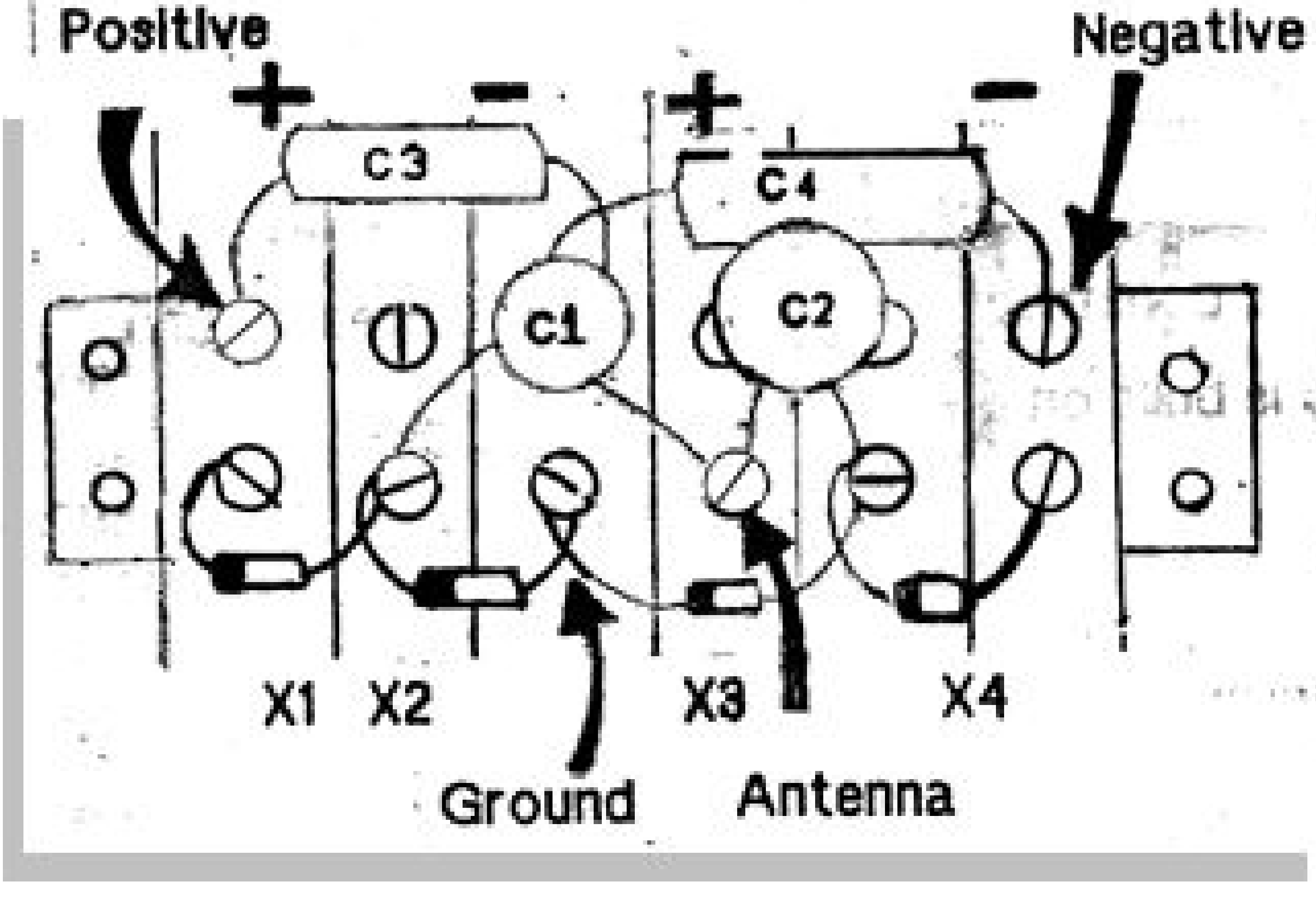
Positive
+
−
+
−
Negative
C3
C4
C1
C2
X1
X2
X3
X4
Ground
Antenna

FIG. 9 TERMINAL STRIP

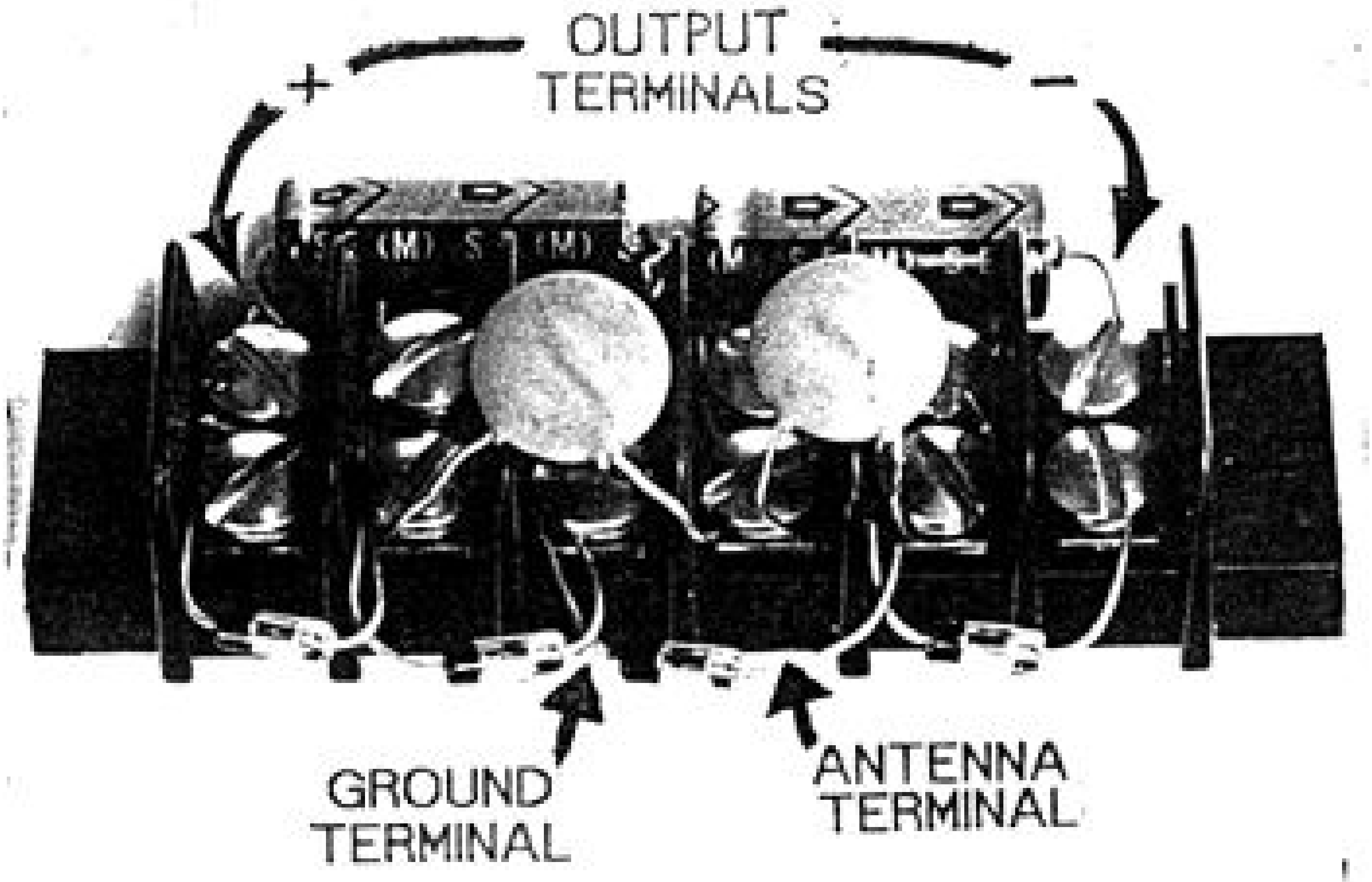
+
OUTPUT TERMINALS
−
GROUND TERMINAL
ANTENNA TERMINAL

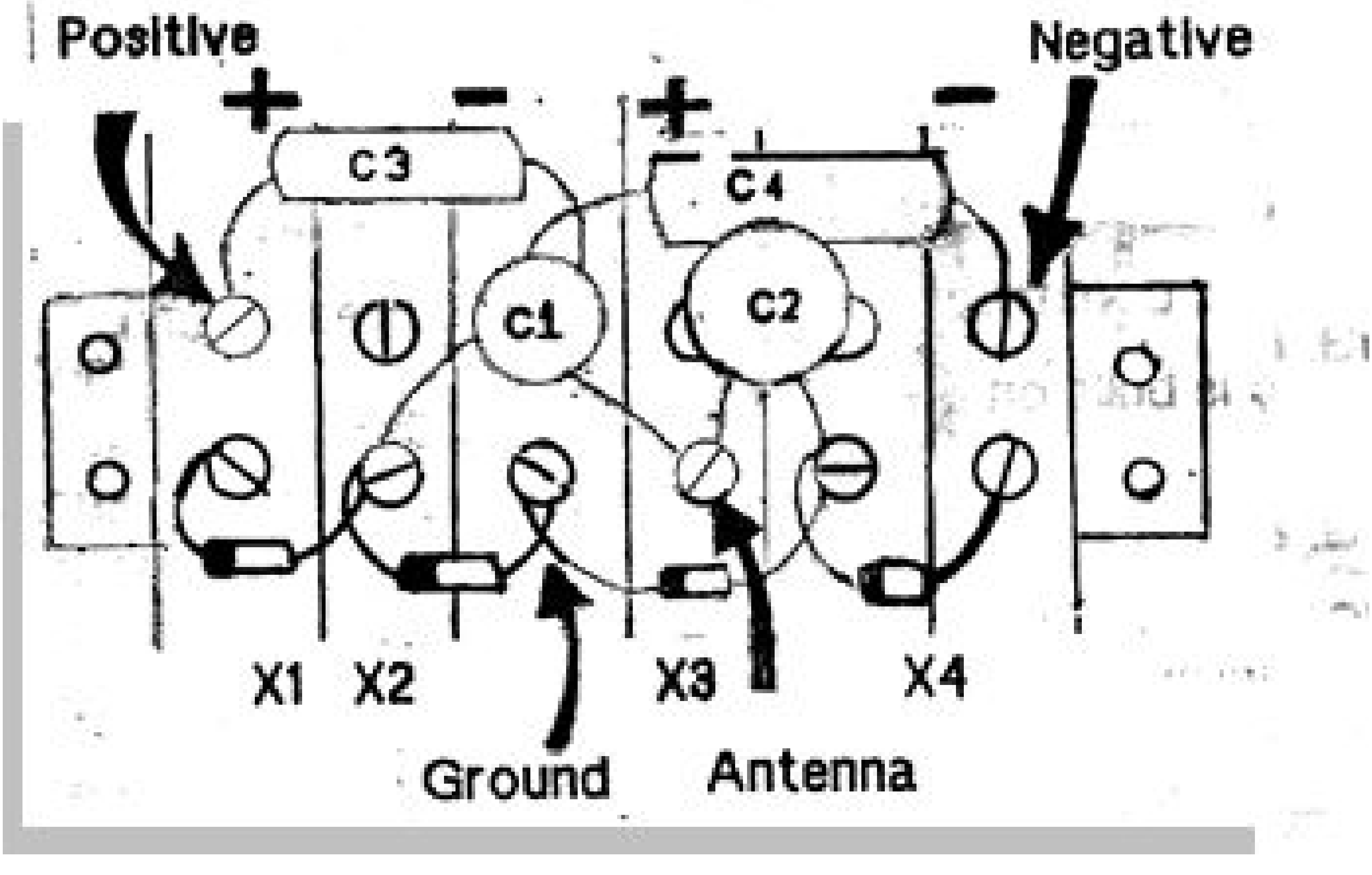
Positive
+
−
+
−
Negative
C3
C4
C1
C2
X1
X2
X3
X4
Ground
Antenna

FIG. 9 TERMINAL STRIP

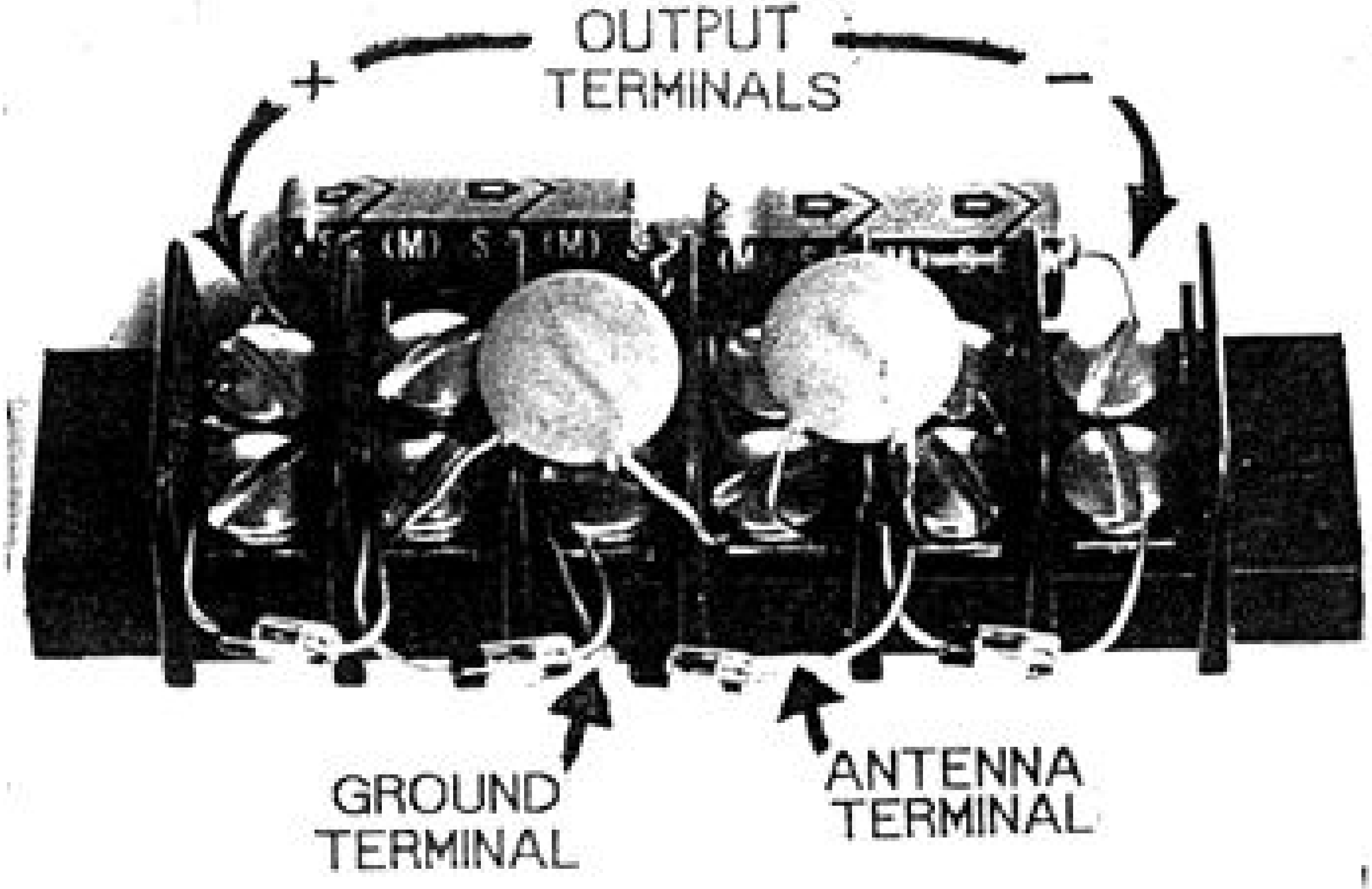
+
OUTPUT TERMINALS
−
GROUND TERMINAL
ANTENNA TERMINAL

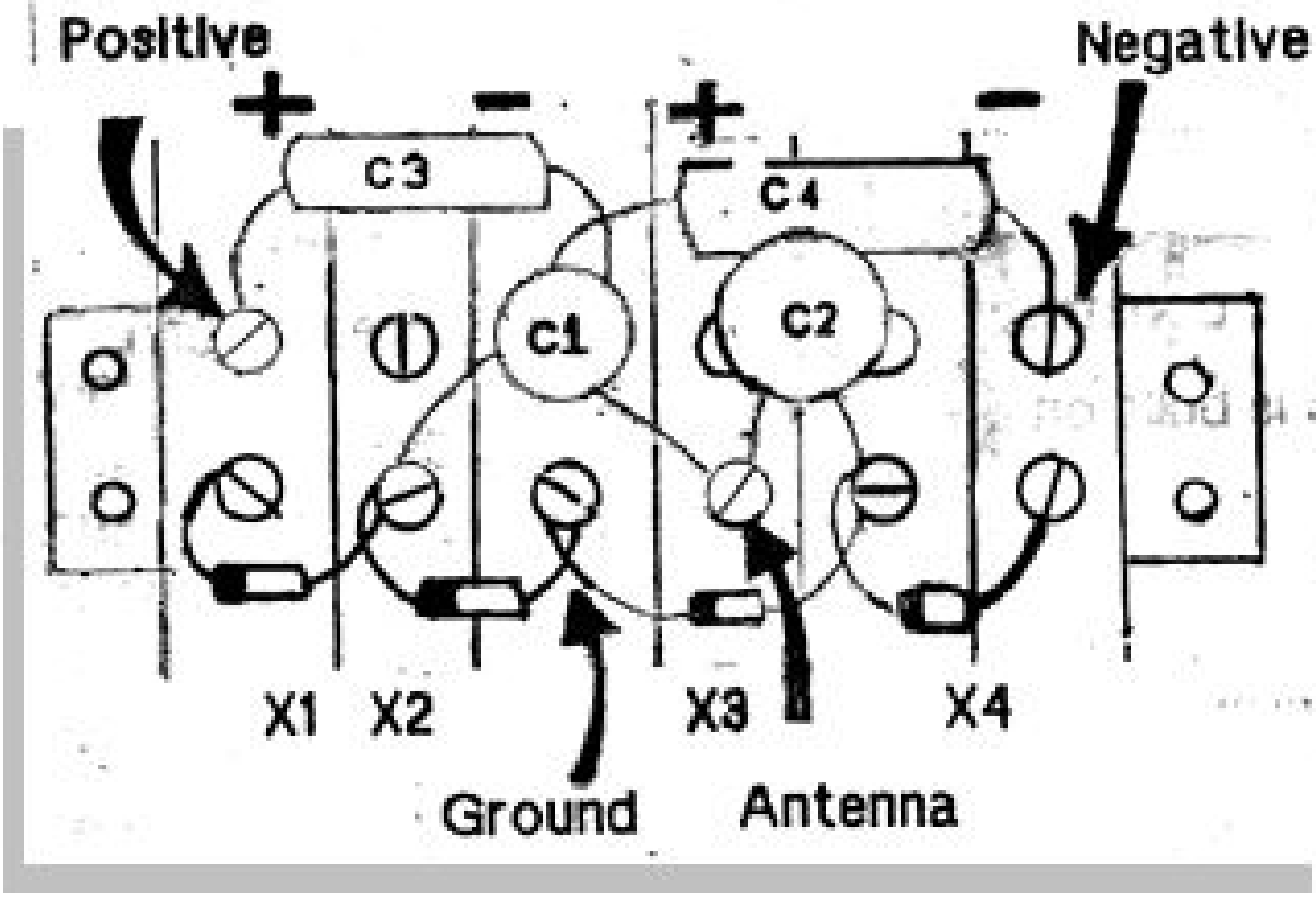
Positive
+
−
+
−
Negative
C3
C4
C1
C2
X1
X2
X3
X4
Ground
Antenna

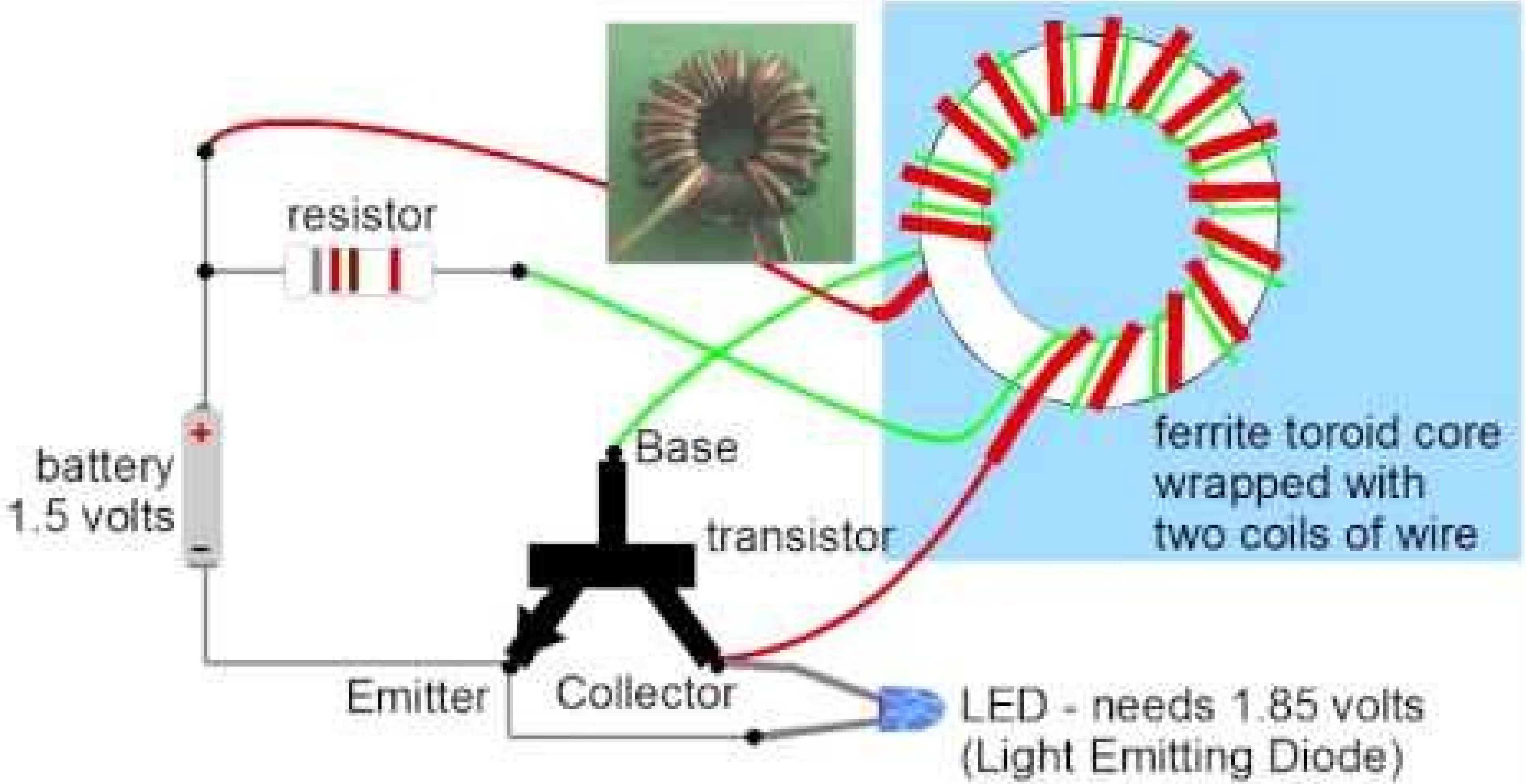
resistor
battery
1.5 volts
+
-
Base
transistor
Emitter
Collector
ferrite toroid core
wrapped with
two coils of wire
LED - needs 1.85 volts
(Light Emitting Diode)

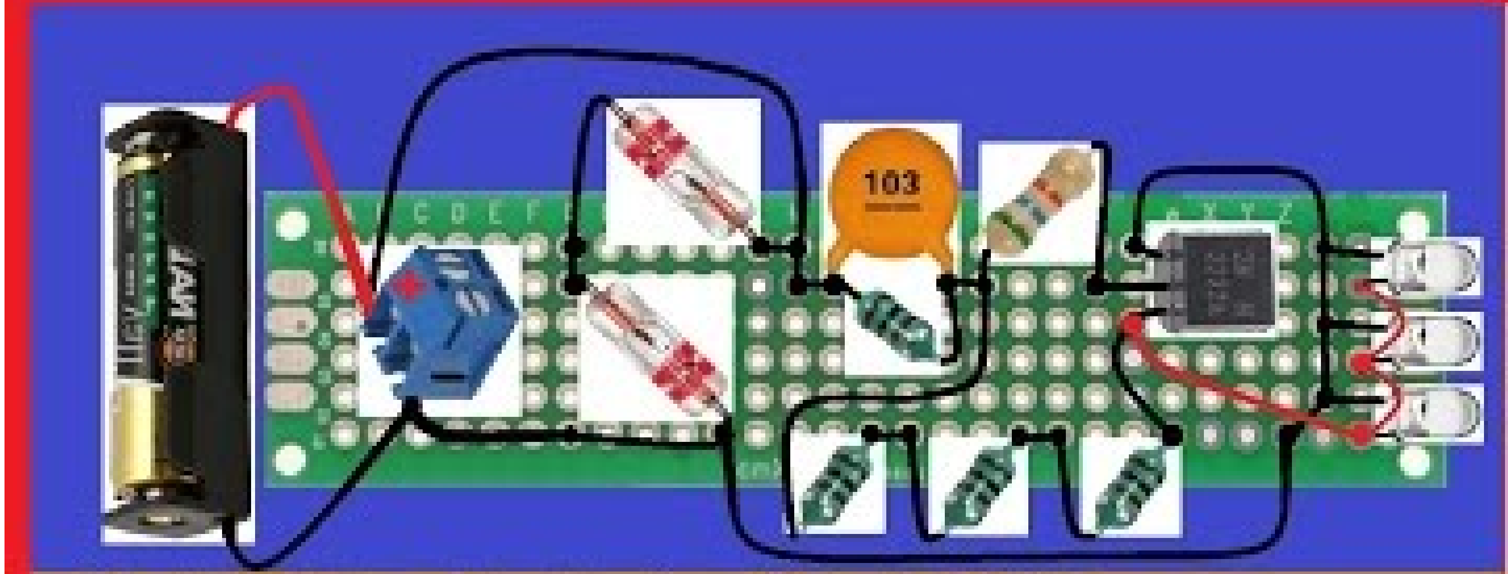

3* Axial lead inductor 82uH 1/4W

1* 10nf ceramic capacitor 103

2* AO90 Germanium signal diode

1* 2N2222 NPN Transistor bipolar 0.6A 40V

1* RESISTOR 5.6K Ohm 1/4W

3* 5mm white Emitting led Diode 20 mA 3.2 V

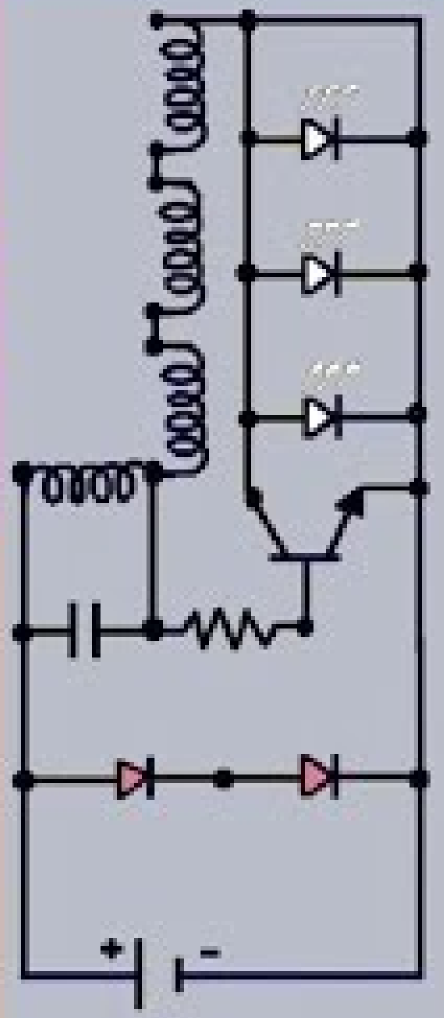

Spy Crystal Radio

The Little Whippersnapper's Little Spy Radio

© www.petervis.com

"MINDSNAPPER" ATTACK ON SPECIAL SHIPS & ENERGETICS WEAPONS SITES

"We have a new, fantastic weapon ... so powerful that, if unrestrainedly used, it could wipe out all life on earth."

Khrushchev,
speaking to Presidium,
Jan. 1960

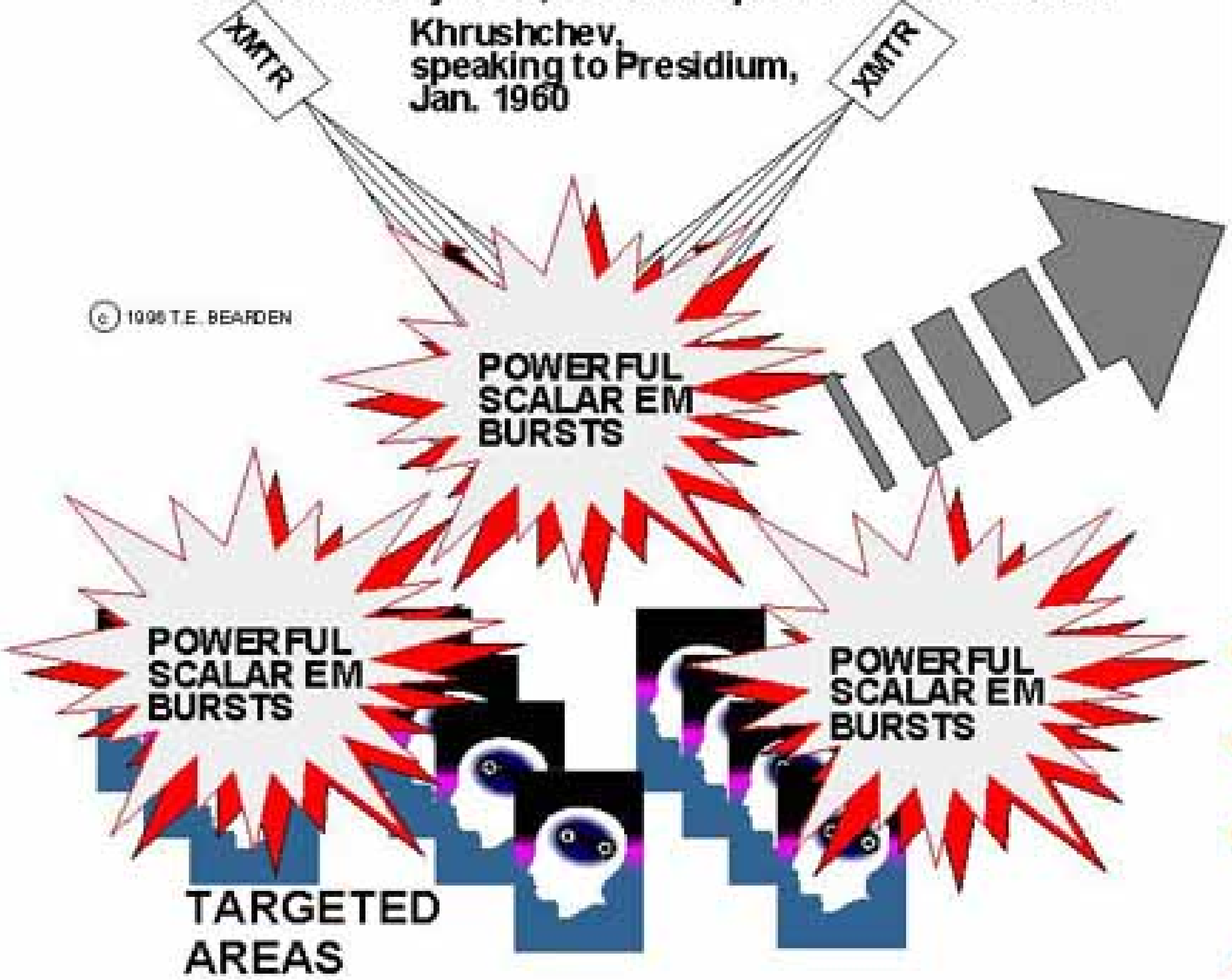
XMTR
XMTR
POWERFUL SCALAR EM BURSTS
POWERFUL SCALAR EM BURSTS
POWERFUL SCALAR EM BURSTS
TARGETED AREAS

(c) 1996 T.E. BEARDEN

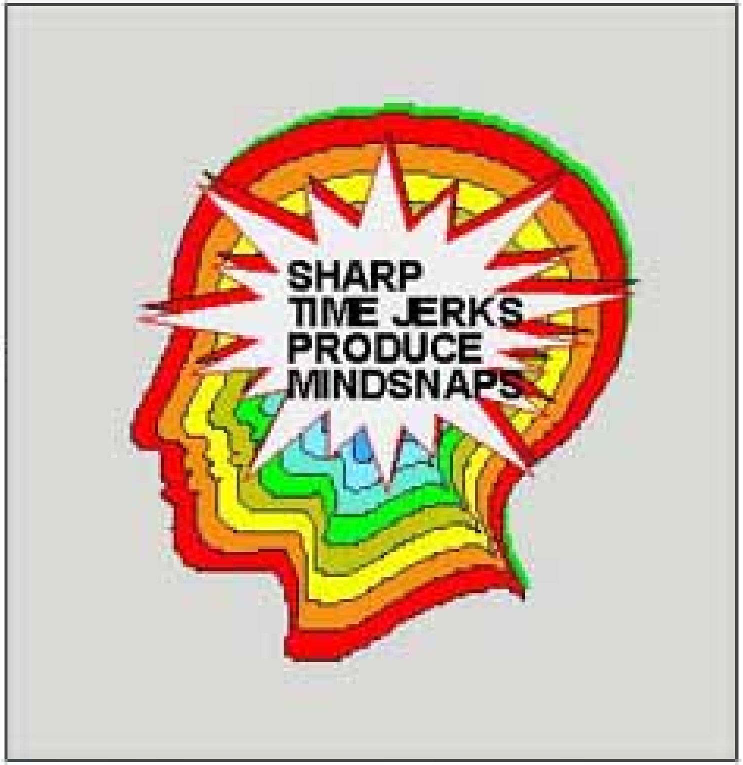
SHARP TIME JERKS PRODUCE MINDSNAPS

- Minds snapped loose from bodies (every living thing)
- Struck areas completely sterilized of all life
- LWs emitted; slow decay of time-charge makes contamination persistent
- Tested twice in Afghanistan

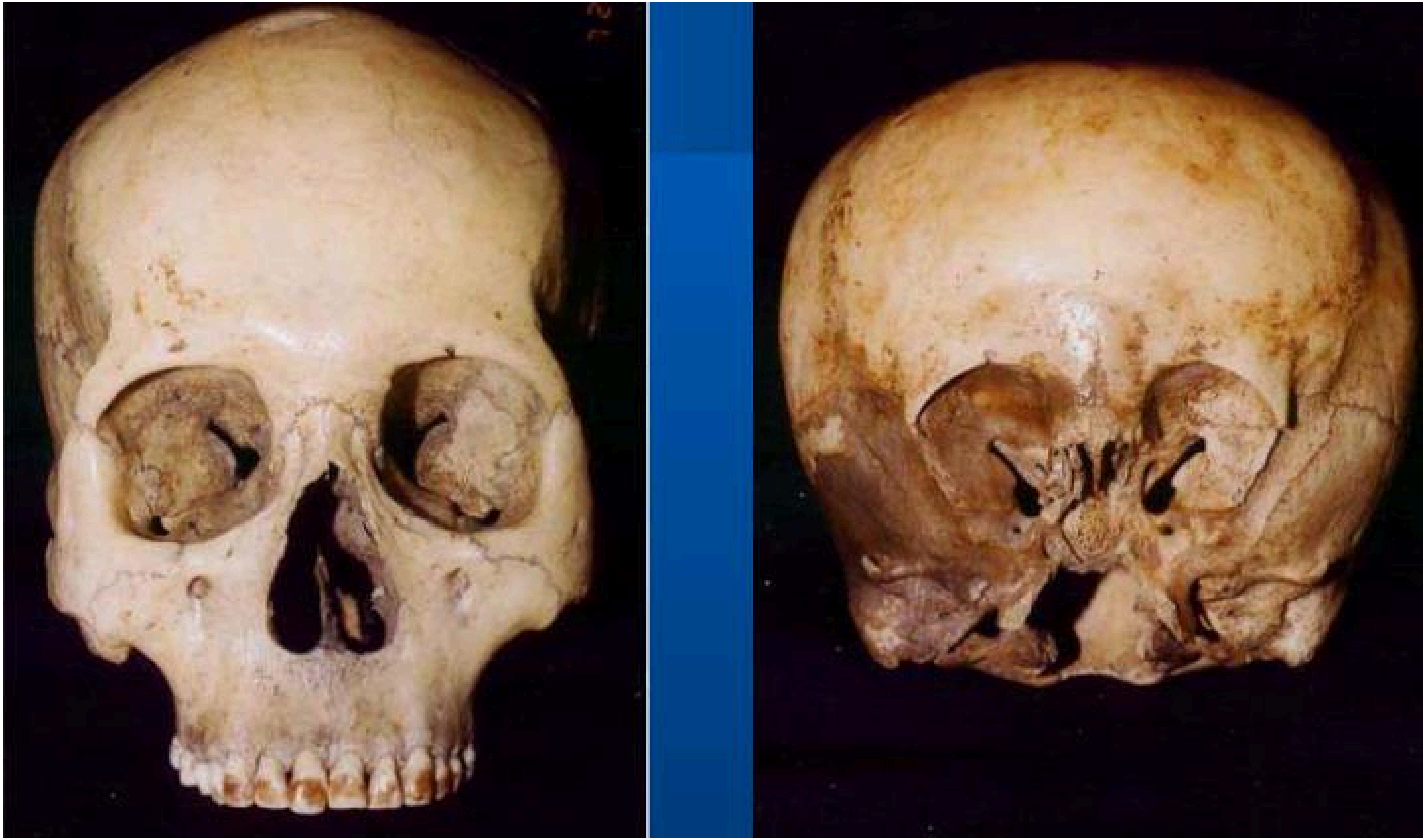

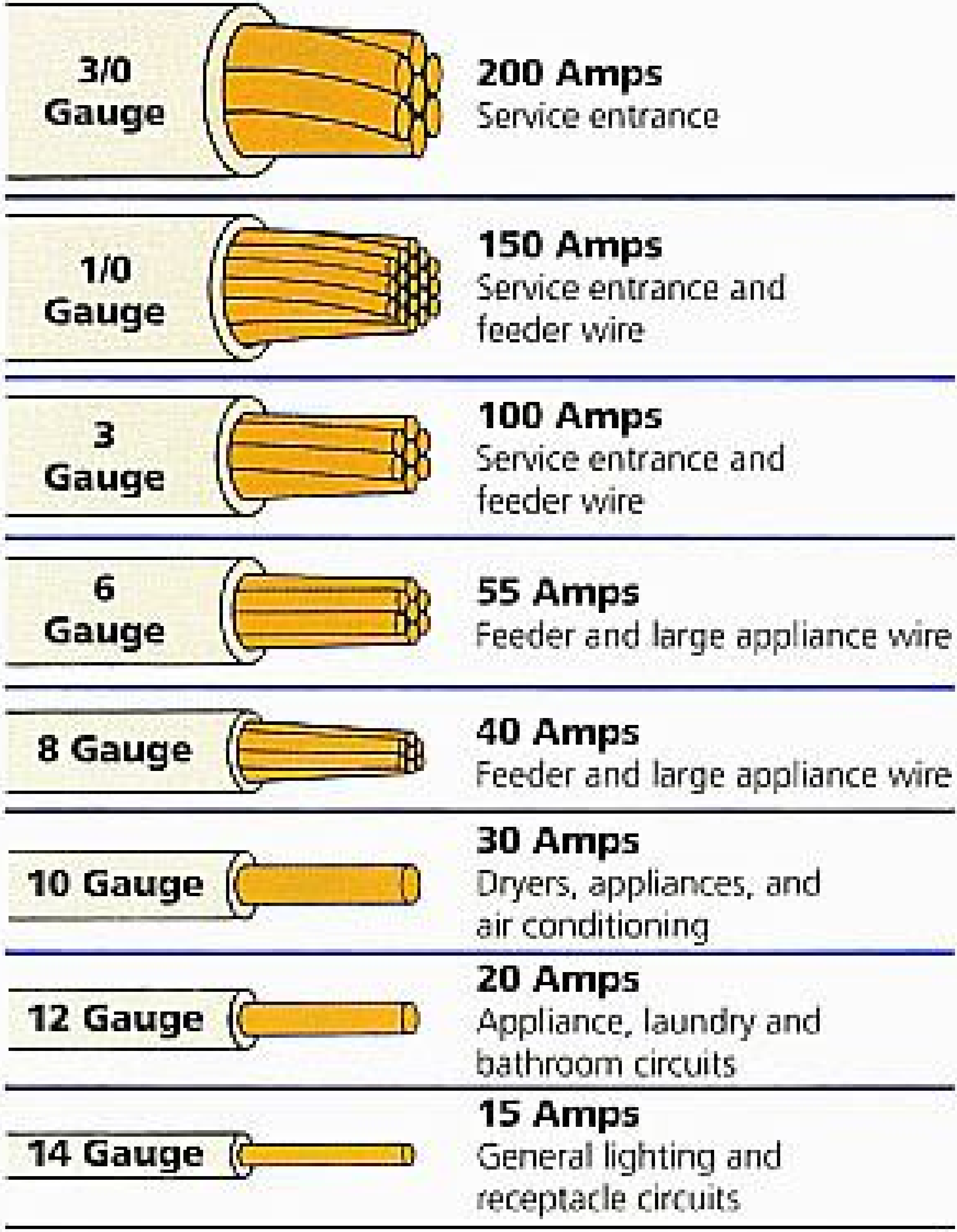

3/0 Gauge	**200 Amps**	Service entrance
1/0 Gauge	**150 Amps**	Service entrance and feeder wire
3 Gauge	**100 Amps**	Service entrance and feeder wire
6 Gauge	**55 Amps**	Feeder and large appliance wire
8 Gauge	**40 Amps**	Feeder and large appliance wire
10 Gauge	**30 Amps**	Dryers, appliances, and air conditioning
12 Gauge	**20 Amps**	Appliance, laundry and bathroom circuits
14 Gauge	**15 Amps**	General lighting and receptacle circuits

Mosfet Audio Amplifier
Use Only 1 Mosfet !!
SAMWHA
160 V
47 µF
105°C

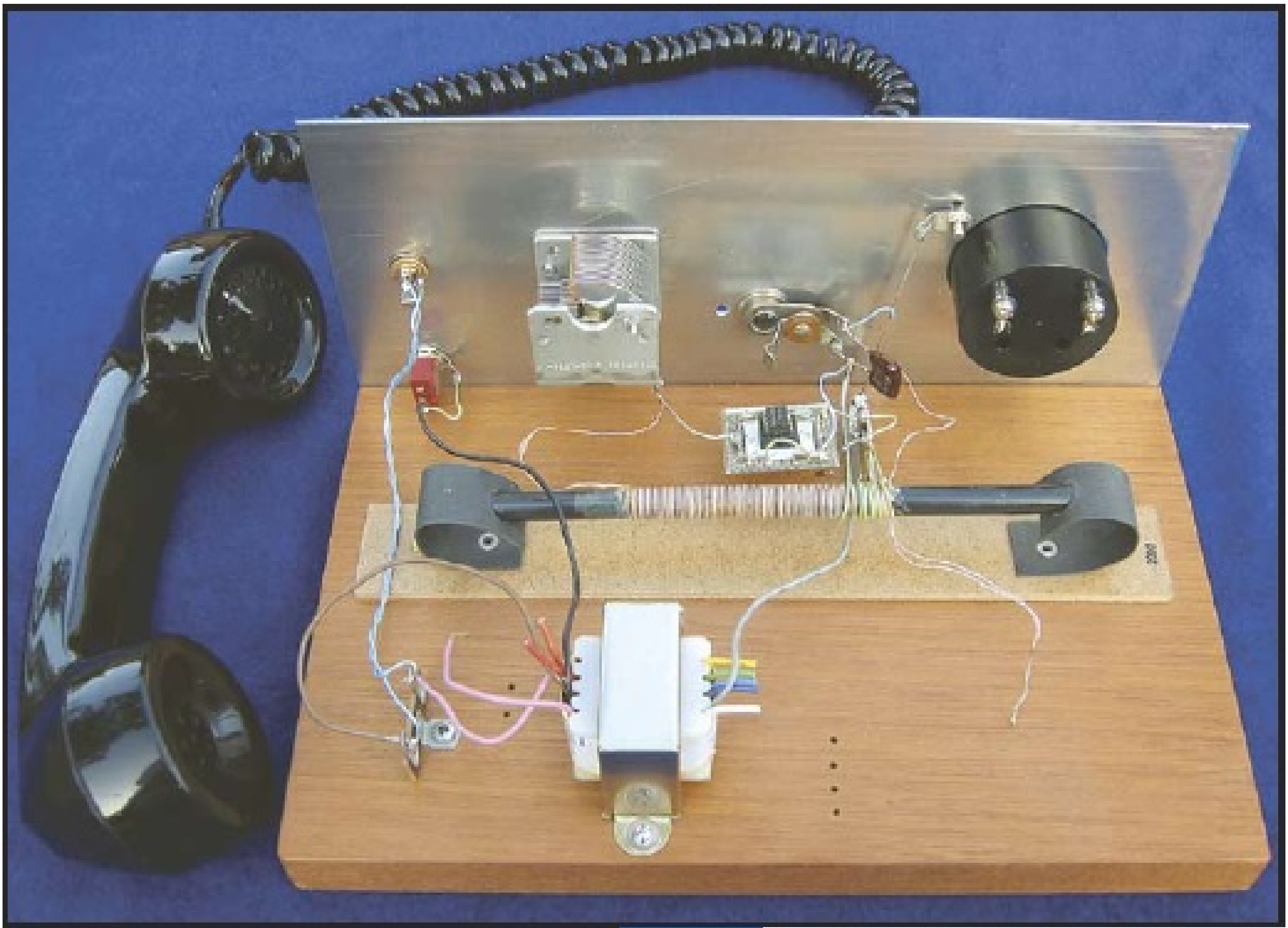

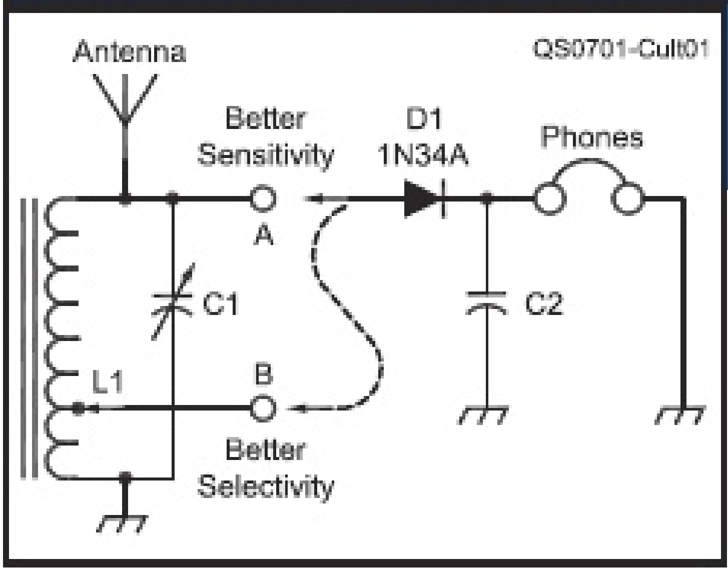
QS0701-Cult01
Antenna
Better
Sensitivity
A
D1
1N34A
Phones
C1
C2
L1
B
Better
Selectivity

Top of right foot

On the top of the right foot are reflex areas relating to the right side of the body, such as the right arm and leg. A point halfway down each foot is known as the "waistline". The upper back and its organs are mapped above this point, and the lower back and the internal organs it encases are below this guideline. The lymph glands and the groin reflex areas wrap around the ankle.

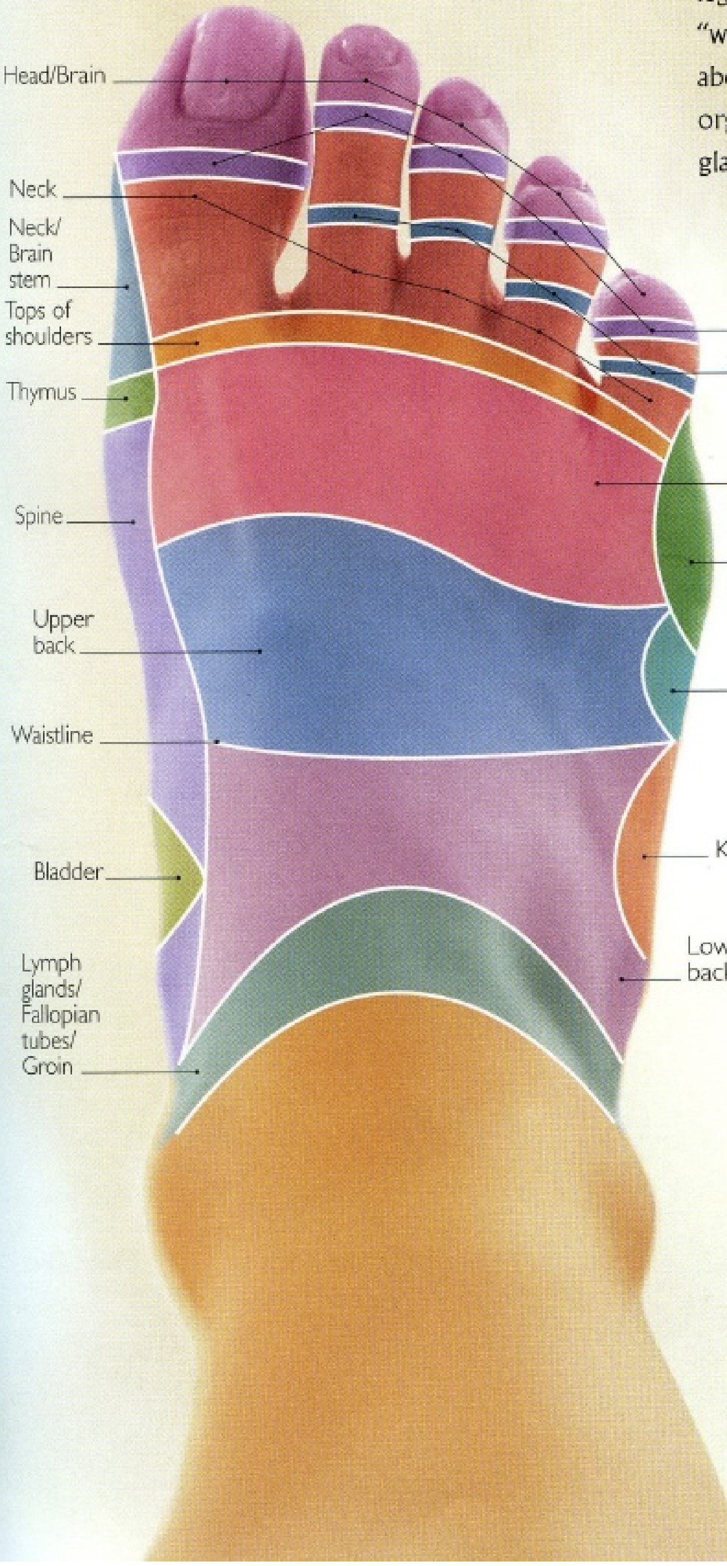
Head/Brain
Neck
Neck/Brain stem
Tops of shoulders
Thymus
Spine
Upper back
Waistline
Bladder
Lymph glands/Fallopian tubes/Groin
Face/Sinus
Teeth/Gums/Jaw
Lung/Chest/Breast/Upper back
Arm
Elbow
Knee/Leg
Lower back

Outside foot

The reflex area for the top of the shoulders runs across the toes, with the areas corresponding to the arm and elbow at the side of the foot. The reflex areas for the reproductive organs and for the sciatic nerve and hip, which curve round the ankle bone, can be clearly seen in this view.

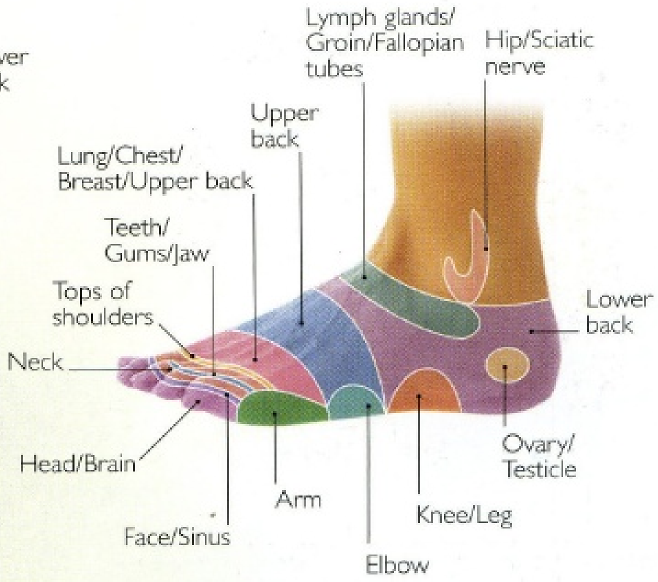
Lymph glands/Groin/Fallopian tubes
Hip/Sciatic nerve
Upper back
Lung/Chest/Breast/Upper back
Teeth/Gums/Jaw
Tops of shoulders
Neck
Lower back
Ovary/Testicle
Head/Brain
Arm
Knee/Leg
Face/Sinus
Elbow

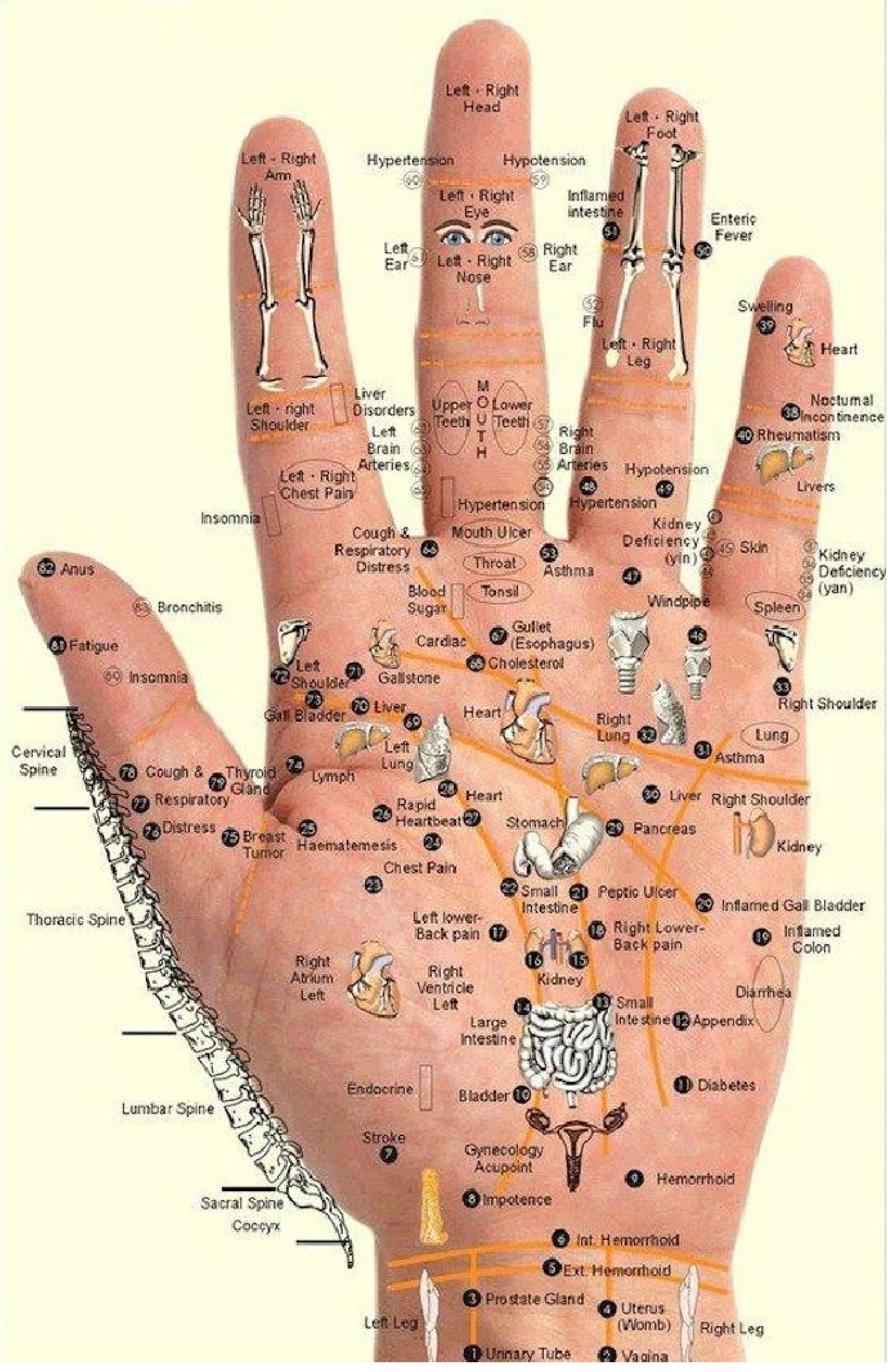
Left · Right
Head
Left · Right
Foot
Left - Right
Arm
Hypertension
60
Hypotension
59
Left · Right
Eye
Inflamed
intestine
51
Enteric
Fever
50
Left
Ear
Left · Right
Nose
58 Right
Ear
52
Flu
Left · Right
Leg
Swelling
37
Heart
Left · right
Shoulder
Liver
Disorders
Upper
Teeth
MOUTH
Lower
Teeth
57
Left
Brain
Arteries
Right
Brain
Arteries
56
55
54
Nocturnal
38 Incontinence
40 Rheumatism
Hypotension
48
49
Hypertension
Livers
Left · Right
Chest Pain
Hypertension
Insomnia
Cough &
Respiratory
Distress
66
Mouth Ulcer
53
Asthma
Kidney
Deficiency
(yin)
45 Skin
Kidney
Deficiency
(yan)
Throat
47
Tonsil
Blood
Sugar
Windpipe
Spleen
82 Anus
83 Bronchitis
81 Fatigue
80 Insomnia
Cardiac
67 Gullet
(Esophagus)
46
68 Cholesterol
72
Left
Shoulder
71
Gallstone
33
Right Shoulder
73
Gall Bladder
70 Liver
69
Heart
Left
Lung
Right
Lung 32
Lung
31 Asthma
Cervical
Spine
76 Cough &
77 Respiratory
78 Distress
79 Thyroid
Gland
74
Lymph
28 Heart
Rapid
26 Heartbeat 27
Stomach
30 Liver
29 Pancreas
Right Shoulder
Kidney
75 Breast
Tumor
25
Haematemesis
24
Chest Pain
23
22 Small
Intestine
21 Peptic Ulcer
20 Inflamed Gall Bladder
Thoracic Spine
Left lower-
Back pain 17
18 Right Lower-
Back pain
19 Inflamed
Colon
Right
Atrium
Left
Right
Ventricle
Left
16
15
Kidney
Diarrhea
14
Large
Intestine
13 Small
Intestine
12 Appendix
Endocrine
Bladder 10
11 Diabetes
Lumbar Spine
Stroke
9
Gynecology
Acupoint
Hemorrhoid
8 Impotence
Sacral Spine
Coccyx
6 Int. Hemorrhoid
5 Ext. Hemorrhoid
3 Prostate Gland
4 Uterus
(Womb)
Left Leg
Right Leg
1 Urinary Tube
2 Vagina

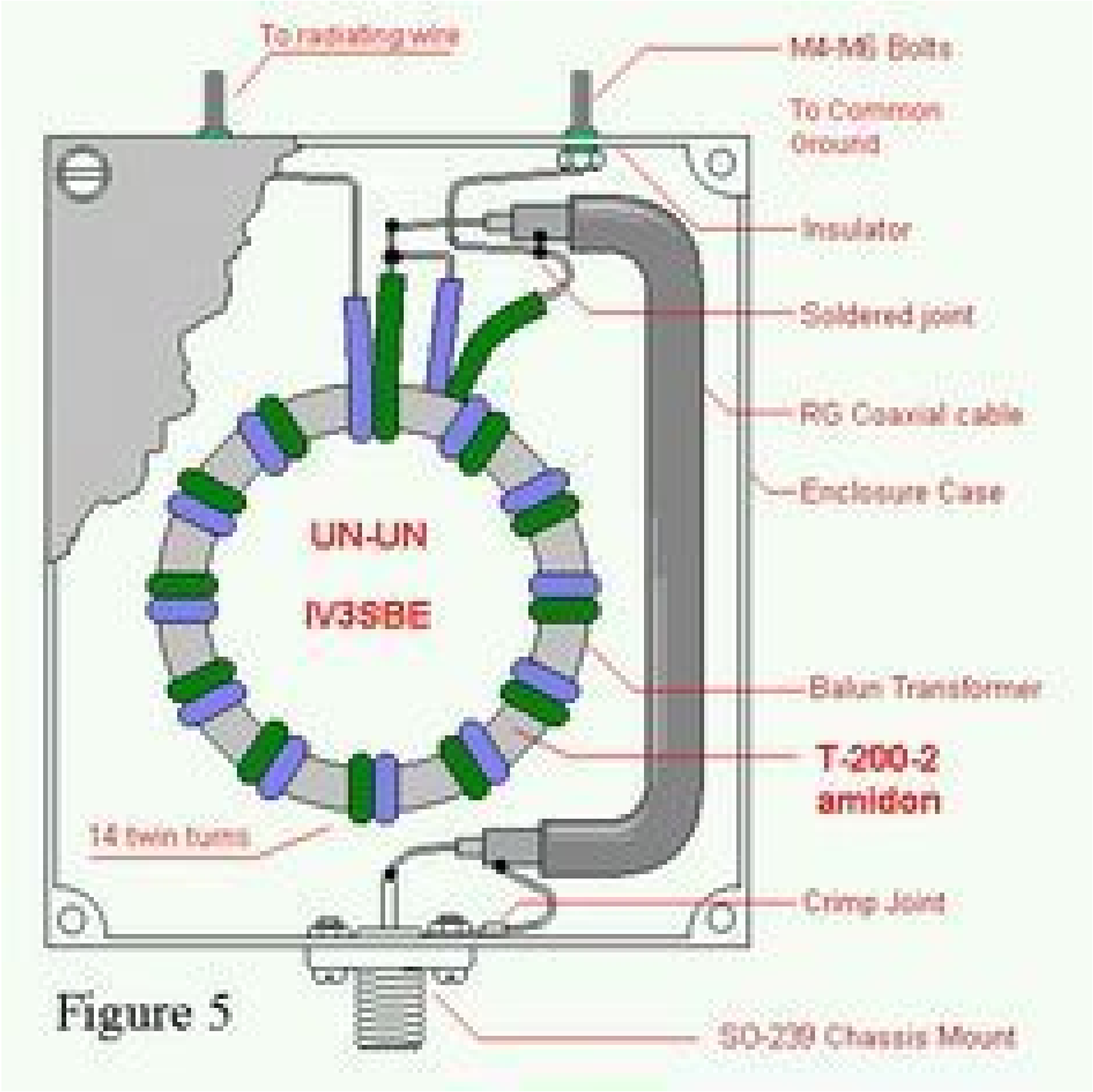
To radiating wire
M4-M6 Bolts
To Common Ground
Insulator
Soldered joint
RG Coaxial cable
Enclosure Case
UN-UN
IV3SBE
Balun Transformer
T-200-2 amidon
14 twin turns
Crimp Joint
SO-239 Chassis Mount

Figure 5

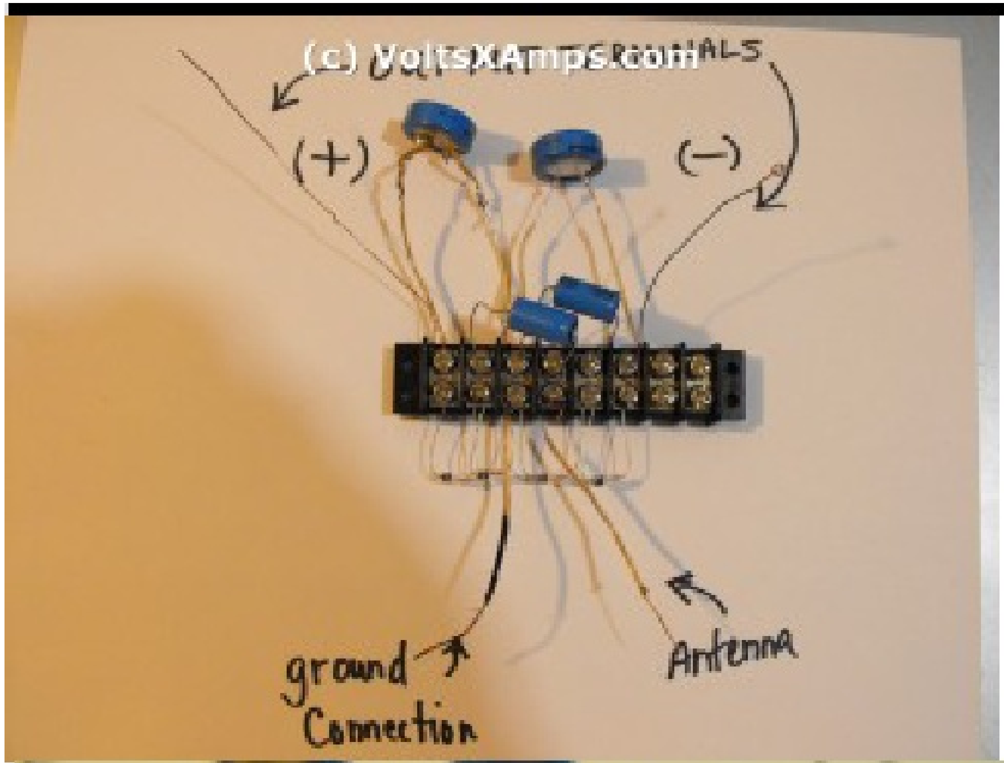
(c) VoltsXAmps.com
OUTPUT TERMINALS
(+)
(−)
ground Connection
Antenna

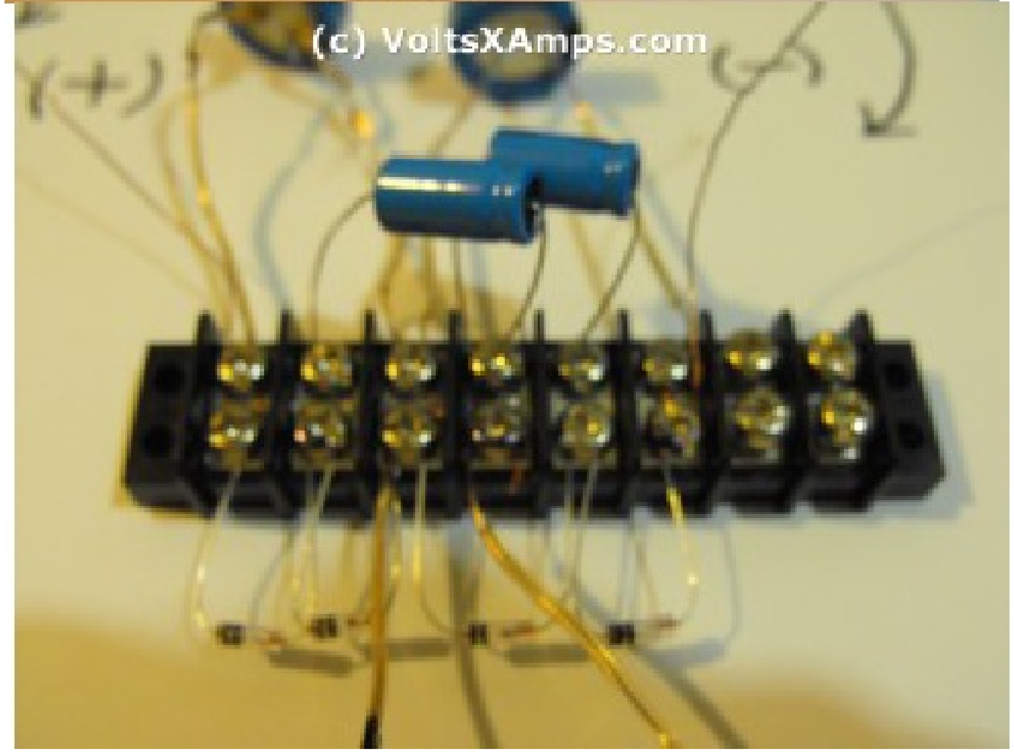
(c) VoltsXAmps.com

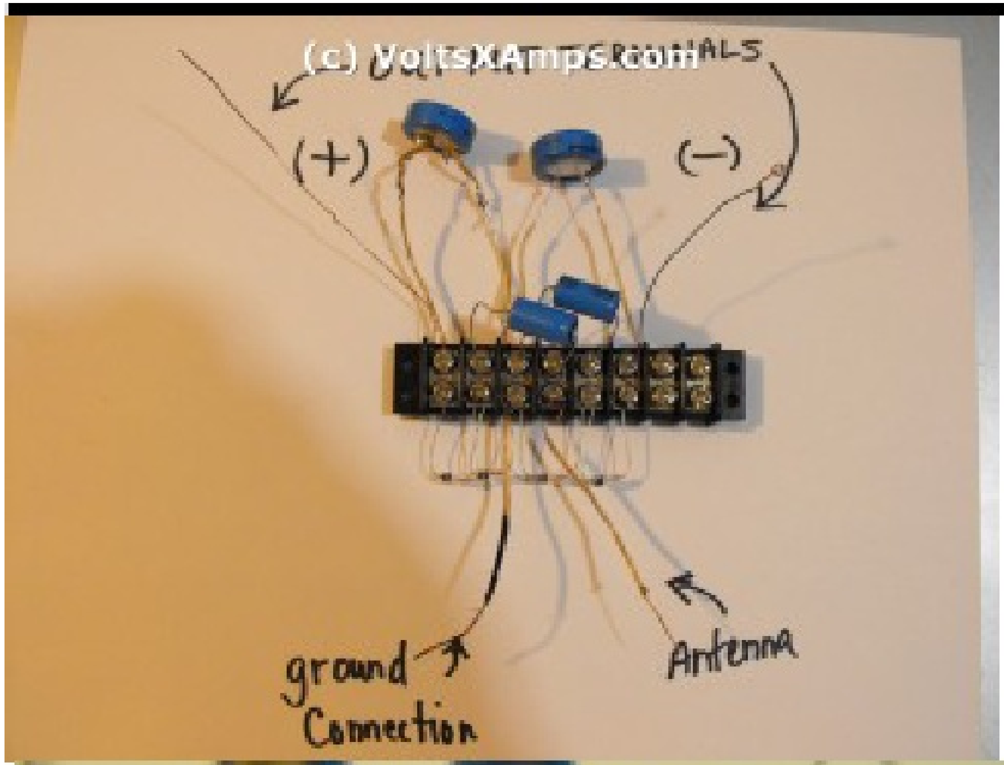
(c) VoltsXAmps.com
ALS
(+)
(–)
ground Connection
Antenna

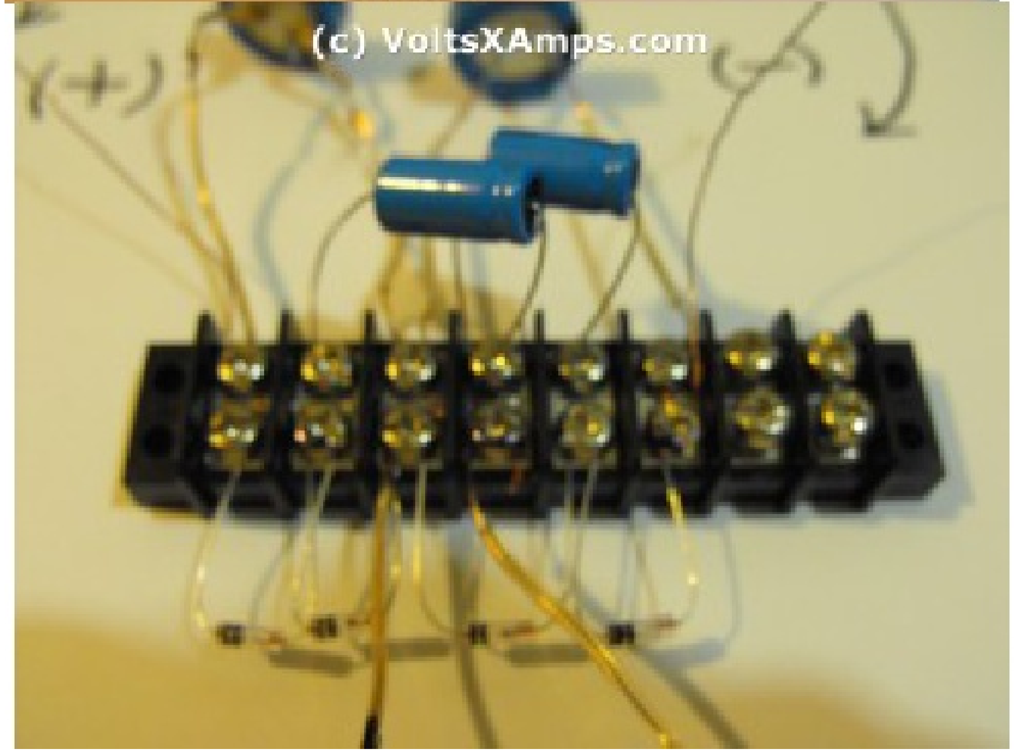
(c) VoltsXAmps.com
(+)
(–)

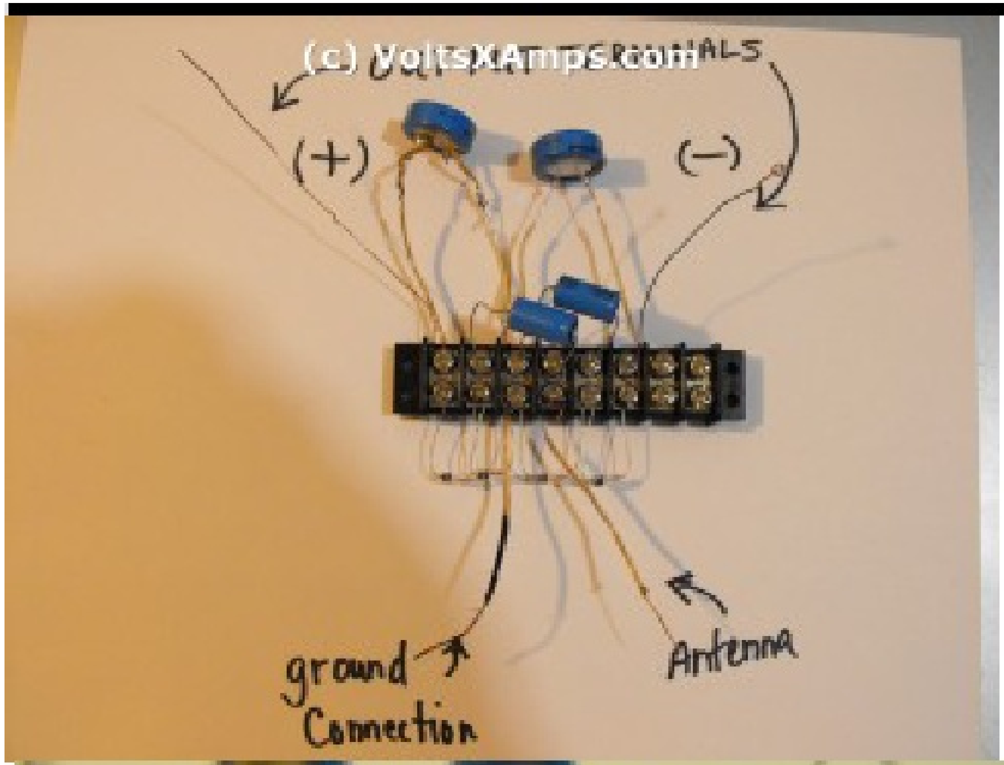
(c) VoltsXAmps.com
(+)
(−)
ground Connection
Antenna

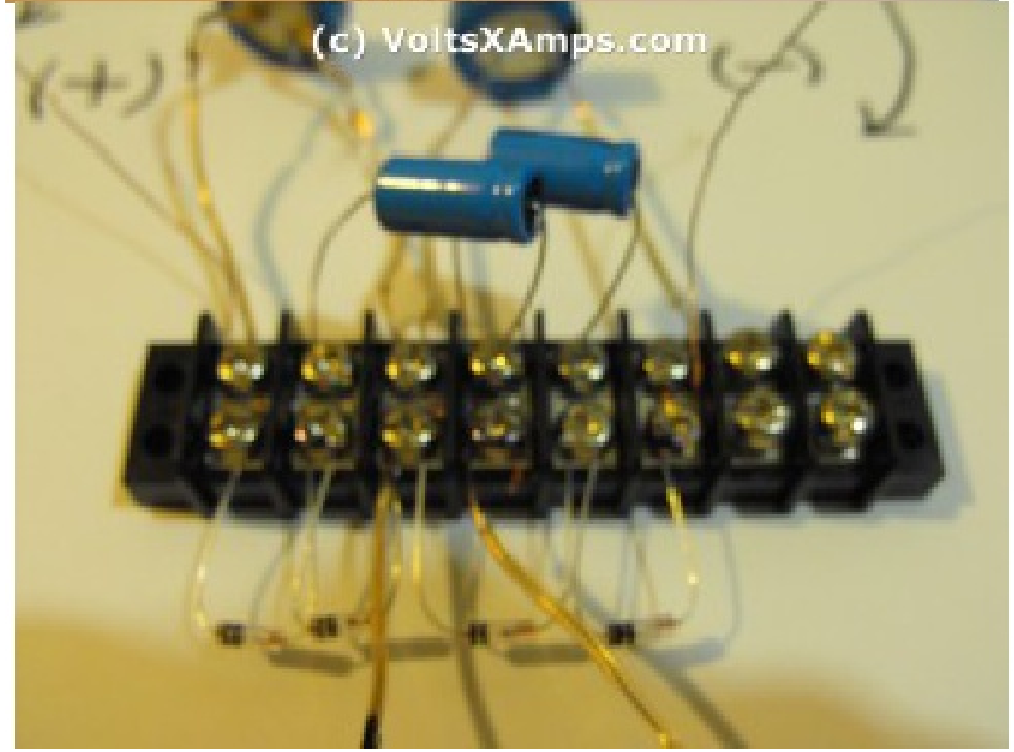
(c) VoltsXAmps.com
(+)
(−)

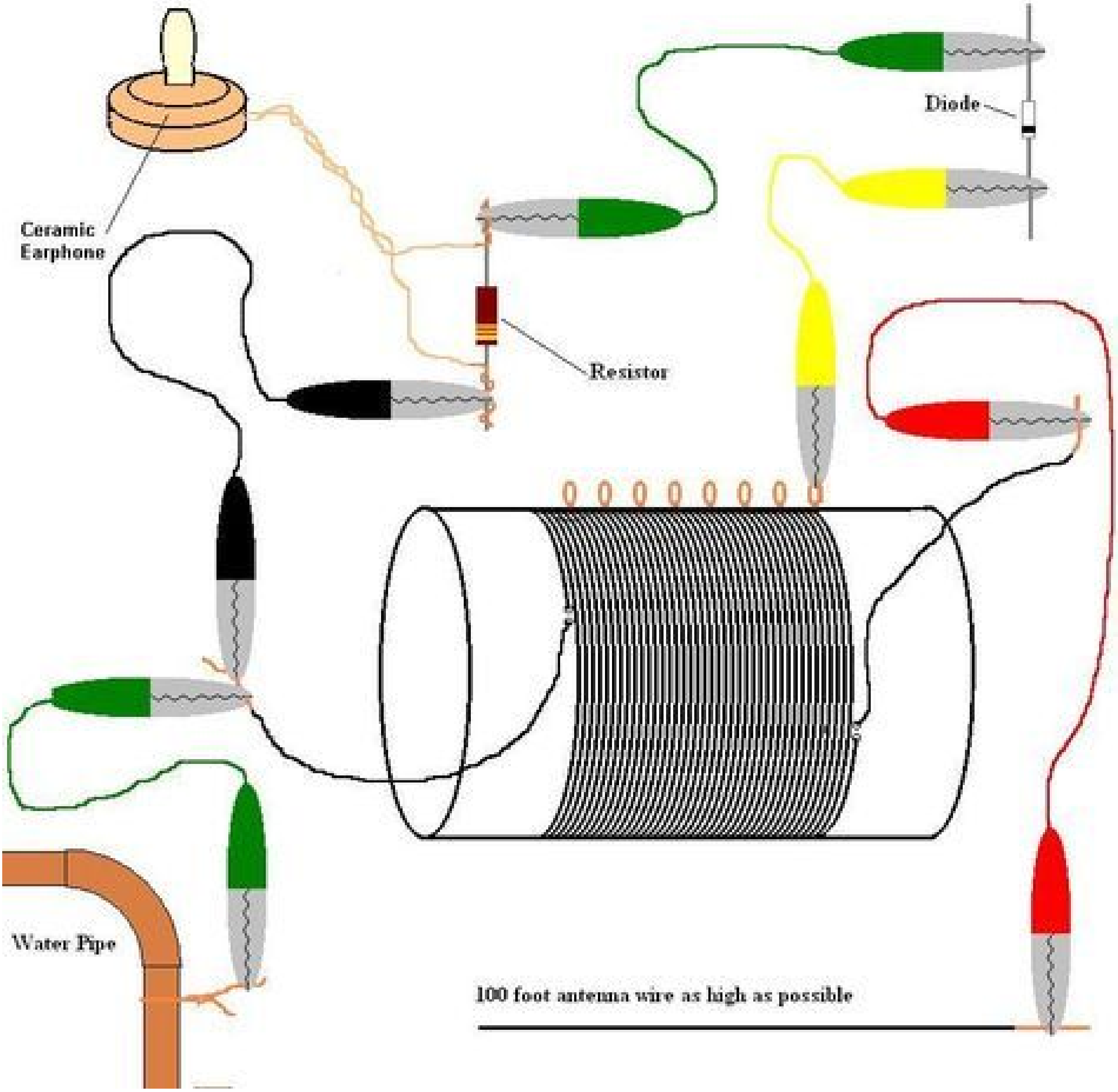
Ceramic Earphone
Diode
Resistor
Water Pipe
100 foot antenna wire as high as possible

1. Slide
2. Barrel
3. Recoil spring guide rod
4. Recoil spring assembly
5. Firing pin
6. Spacer sleeve
7. Firing pin spring
8. Spring cups
9. Firing pin safety
10. Firing pin safety spring
11. Extractor
12. Extractor depressor plunger
13. Extractor depressor plunger spring
14. Spring-loaded bearing
15. Slide cover plate
16. Sights
17. Frame
18. Magazine catch spring
19. Magazine catch
20. Slide lock spring
21. Slide lock
22. Locking block
23. Trigger mechanism housing
24. Connector
25. Trigger spring
26. Trigger with trigger bar
27. Slide stop lever
28. Trigger pin
29. Trigger housing pin
30. Follower
31. Magazine spring
32. Magazine floorplate
32a. Magazine insert
33. Magazine tube
34. Locking block pin

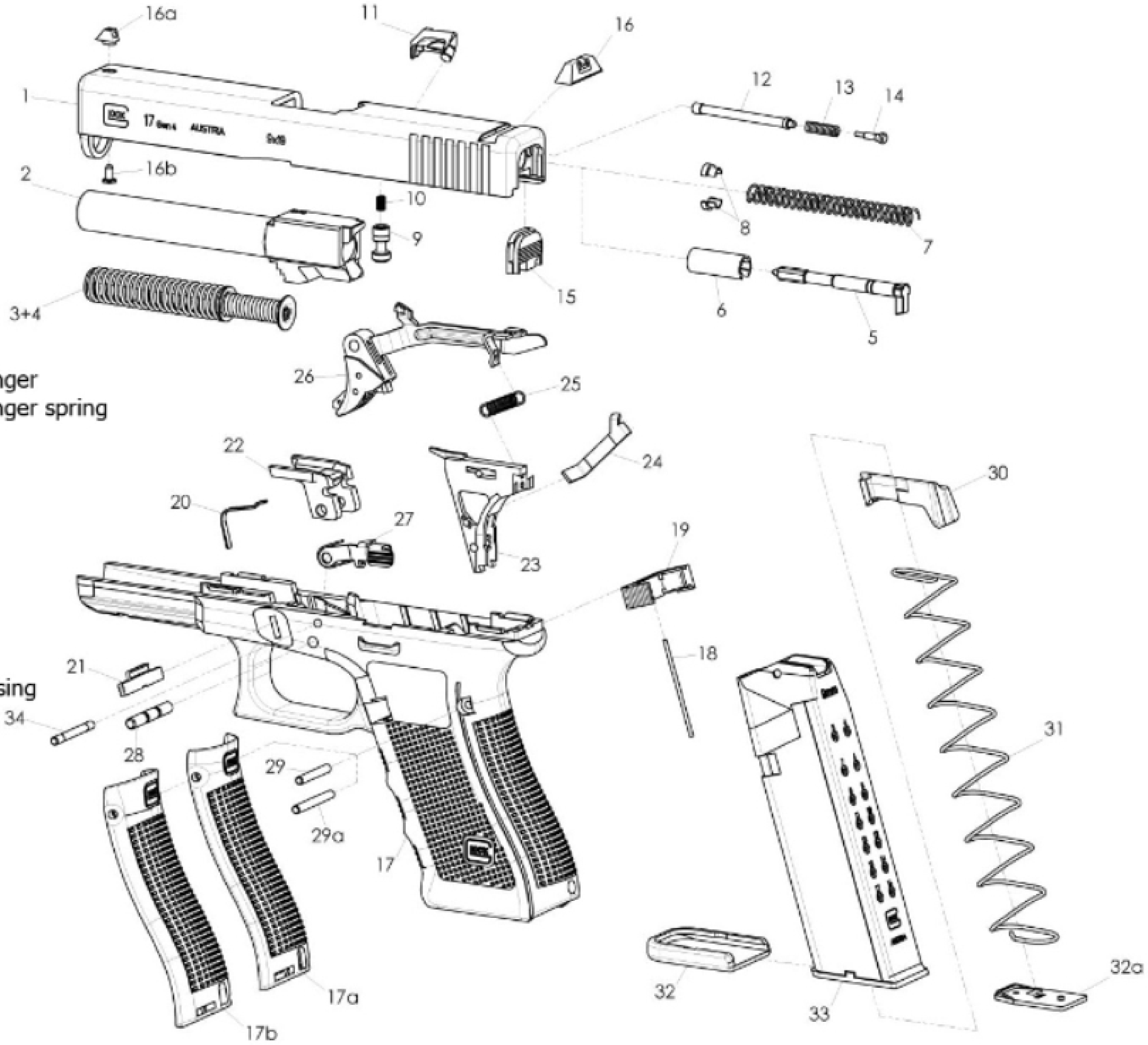
16a
11
16
1
12
13
14
2
16b
10
9
8
7
3+4
15
6
5
26
25
22
24
20
27
30
19
23
21
18
34
28
29
31
29a
17
32a
32
33
17a
17b

REFLEXOLOGIA DE LAS MANOS

EL PODER CURATIVO DE LOS REFLEJOS DE LA MANO

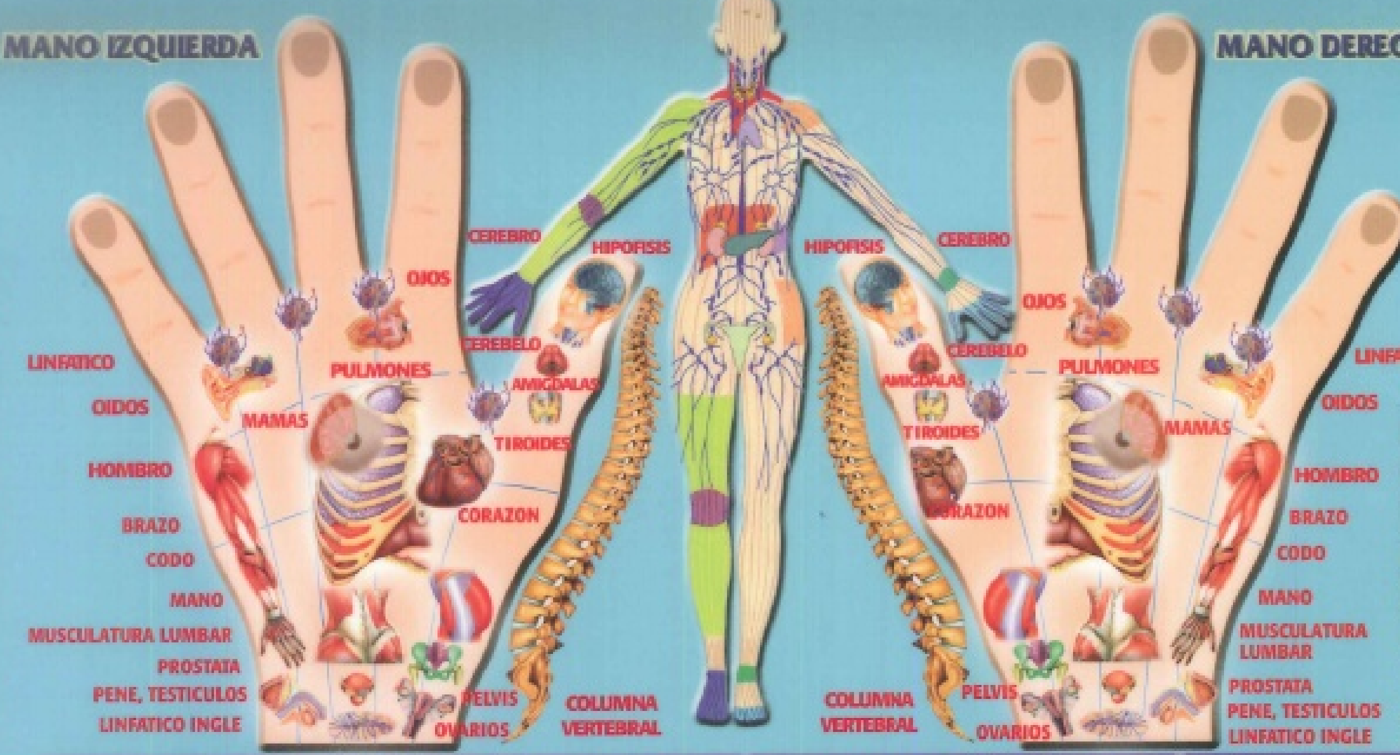
MANO IZQUIERDA
LINFATICO
OIDOS
HOMBRO
BRAZO
CODO
MANO
MUSCULATURA LUMBAR
PROSTATA
PENE, TESTICULOS
LINFATICO INGLE
MAMAS
PULMONES
OJOS
CEREBRO
CEREBELO
AMIGDALAS
TIROIDES
CORAZON
PELVIS
OVARIOS
HIPOFISIS
COLUMNA VERTEBRAL
HIPOFISIS
CEREBRO
OJOS
CEREBELO
AMIGDALAS
TIROIDES
CORAZON
PELVIS
OVARIOS
COLUMNA VERTEBRAL
PULMONES
MAMAS
MANO DERECHA
LINFATICO
OIDOS
HOMBRO
BRAZO
CODO
MANO
MUSCULATURA LUMBAR
PROSTATA
PENE, TESTICULOS
LINFATICO INGLE

LOS REFLEJOS CRUZADOS

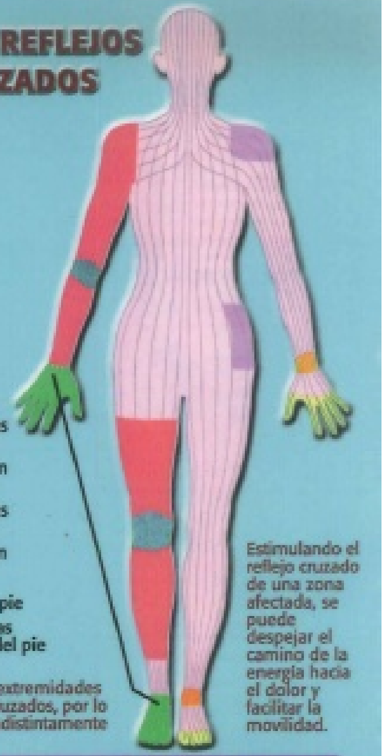

- 🟥 Los brazos en las piernas
- 🟪 Los hombros en las caderas
- 🟦 Los codos en las rodillas
- 🟧 Las muñecas en los tobillos
- 🟩 La mano en el pie
- 🟨 Los dedos de las manos en los del pie

Los reflejos entre las extremidades del mismo lado son cruzados, por lo que puede actuarse indistintamente sobre uno u otro.

Estimulando el reflejo cruzado de una zona afectada, se puede despejar el camino de la energía hacia el dolor y facilitar la movilidad.

EJERCICIOS PREVIOS DE RELAJACIÓN (fondo lila) Y SECUENCIA BÁSICA DE TRATAMIENTO (fondo verde)

Rotación de lado

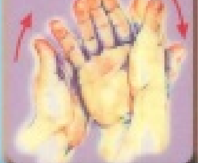

Calentar palma y dorso

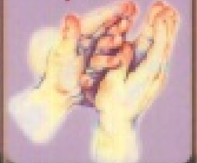

Rotación con el puño en la palma

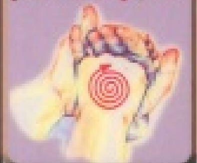

Presión con el puño en la palma

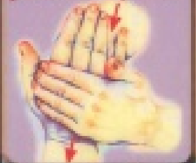

Masaje del diafragma

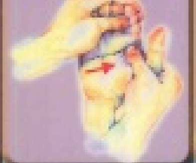

Masaje de la caja torácica

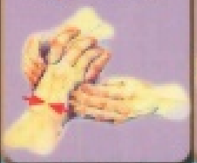

Masaje de Muñeca

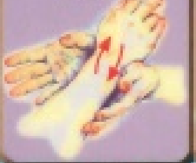

Giro de la muñeca arriba y abajo

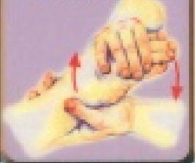

Giro de la muñeca a los lados

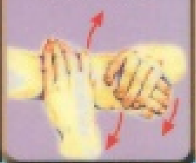

Agitar la mano

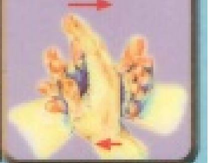

CABEZA

Migrañas • Dolores de cabeza • Ansiedad • Depresión

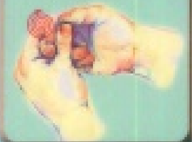

GARGANTA. TIROIDES. CUELLO Y CERVICALES

• Dolor de cabeza • Migrañas • Obesidad • Amigdalitis • Faringitis • Laringitis • Afonía • Dolor de cuello • Dolor cervical • Rigidez de nuca

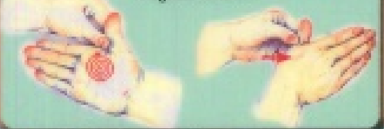

SENOS OSEOS

• Catarro de nariz • Sinusitis • Rinitis • Alergias

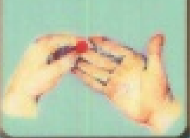

OJOS Y OIDOS

• Otitis • Zumbidos • Glaucoma • Conjuntivitis • Fatiga ocular

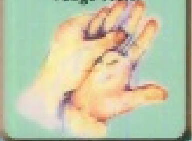

COLUMNA

• Ciáticas • Artrosis de cadera • Lumbagos • Hernia discal • Escoliosis y cifosis dolorosas

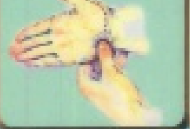

CADERA

• Falta de fuerza en la pierna • Artrosis de cadera • Osteoporosis

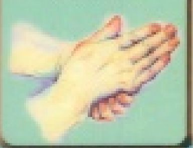

COXIS

• Pubalgias • Falta de fuerza en la pierna • Estimulación general

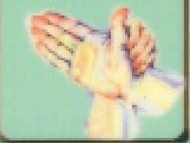

HOMBRO

• Artritis y artrosis • Tendinitis del Supraespinoso

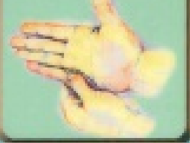

PULMON y CORAZON

• Taquicardia • Asma • Hipertensión • Infecciones Respiratorias

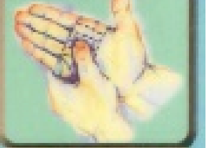

PULMON

• Bronquitis • Asma • Neumonía • Infecciones respiratorias

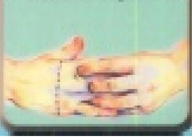

ESTOMAGO. BAZO. PANCREAS

• Úlceras • Gastritis • Gases • Mala digestión • Acidez • Diabetes • Hipoglucemia • Anemias • Infecciones

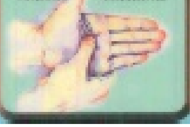

INTESTINOS

• Estreñimiento • Diarreas • Apendicitis • Colon irritable • Gases

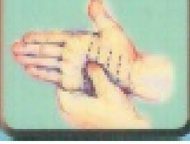

VEJIGA. URETER

• Cistitis • Infecciones • Inflamación

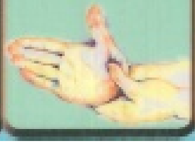

RIÑONES

• Infecciones del riñón • Piedras riñón • Retención de líquidos

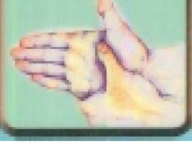

HIGADO VESICULA BILIAR

• Hepatitis • Cálculos de vesícula • Digestión lenta

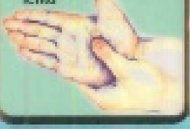

UTERO. PROSTATA

• Hemorragias uterinas • Miomas • Prostatitis

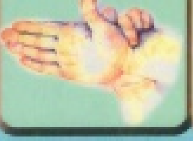

OVARIOS TESTICULOS

• Menstruaciones dolorosas • Infecciones genitales • Inflamaciones

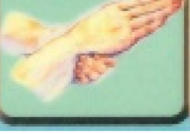

TROMPAS FALOPIO COND. DEFERENTES

• Menstruaciones dolorosas • Infecciones • Inflamaciones

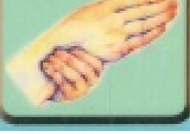

PLEXO SOLAR DIAFRAGMA

• Distonías neurovegetativas • Ansiedad • Miedo • Obsesión • Presión pecho • Dolor estómago

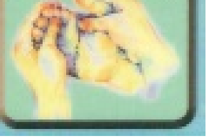

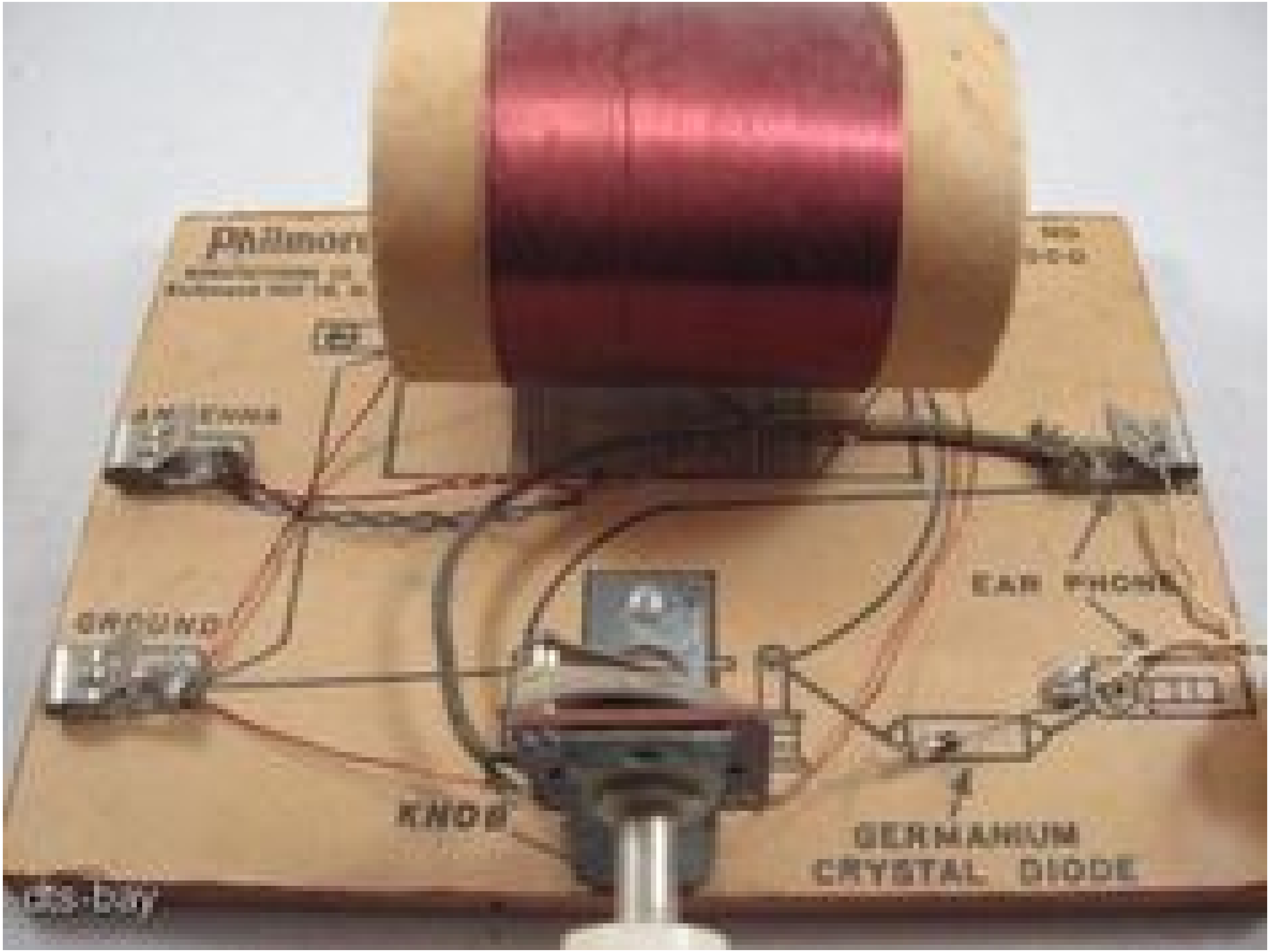
Philmore
ANTENNA
GROUND
EAR PHONE
KNOB
GERMANIUM
CRYSTAL DIODE
ebay

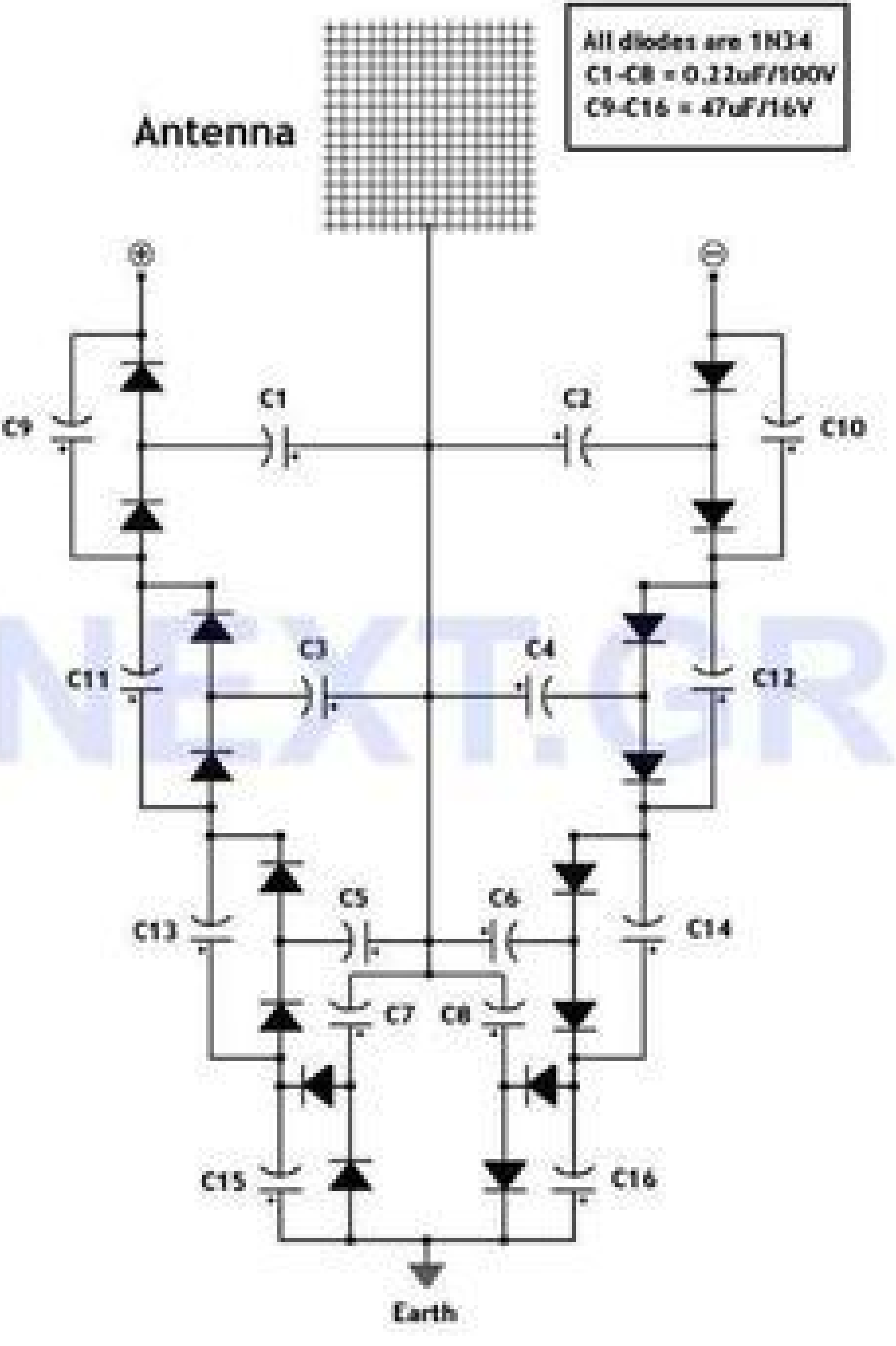
Antenna
⊕
⊖
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
C13
C14
C15
C16
Earth

All diodes are 1N34
C1-C8 = 0.22uF/100V
C9-C16 = 47uF/16V

NEXT.GR

According to **Prof. K. Meyl** out of the book **"Scalarwaves"**

Explanation of the longitudinal Wave

and the functional principle of this exhibit

The special properties of the longitudinal wave (scalar wave) have been discovered over 100 years ago by Nicola Tesla. Through the different behaviour, compared to the Hertzian wave (radio wave), there resulted in a scientific dispute. (Similar to the war of currents a couple of years before).
Tesla's experiment, the Wardenclyff-Tower, and the patent drawings that describe the wireless power transmission.

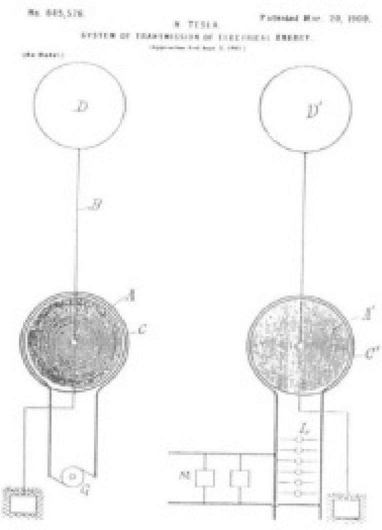
No. 645,576.
N. TESLA.
Patented Mar. 20, 1900.
SYSTEM OF TRANSMISSION OF ELECTRICAL ENERGY.
D
D'
B
A
C
A'
C'
L
M
G

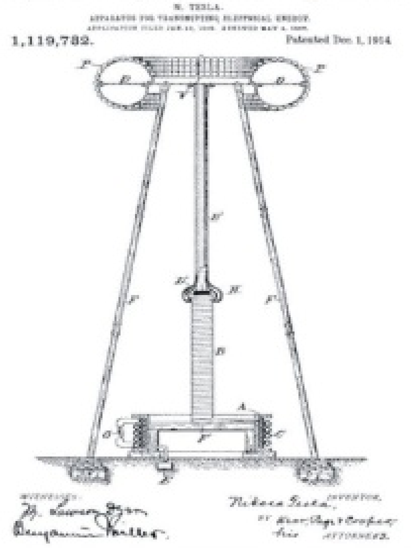
N. TESLA.
APPARATUS FOR TRANSMITTING ELECTRICAL ENERGY.
1,119,782.
Patented Dec. 1, 1914.
P
D
B'
E
H
F
B
A
C
WITNESSES:
INVENTOR
ATTORNEYS.

The principle of scalarwavetransmission between two pancake coils

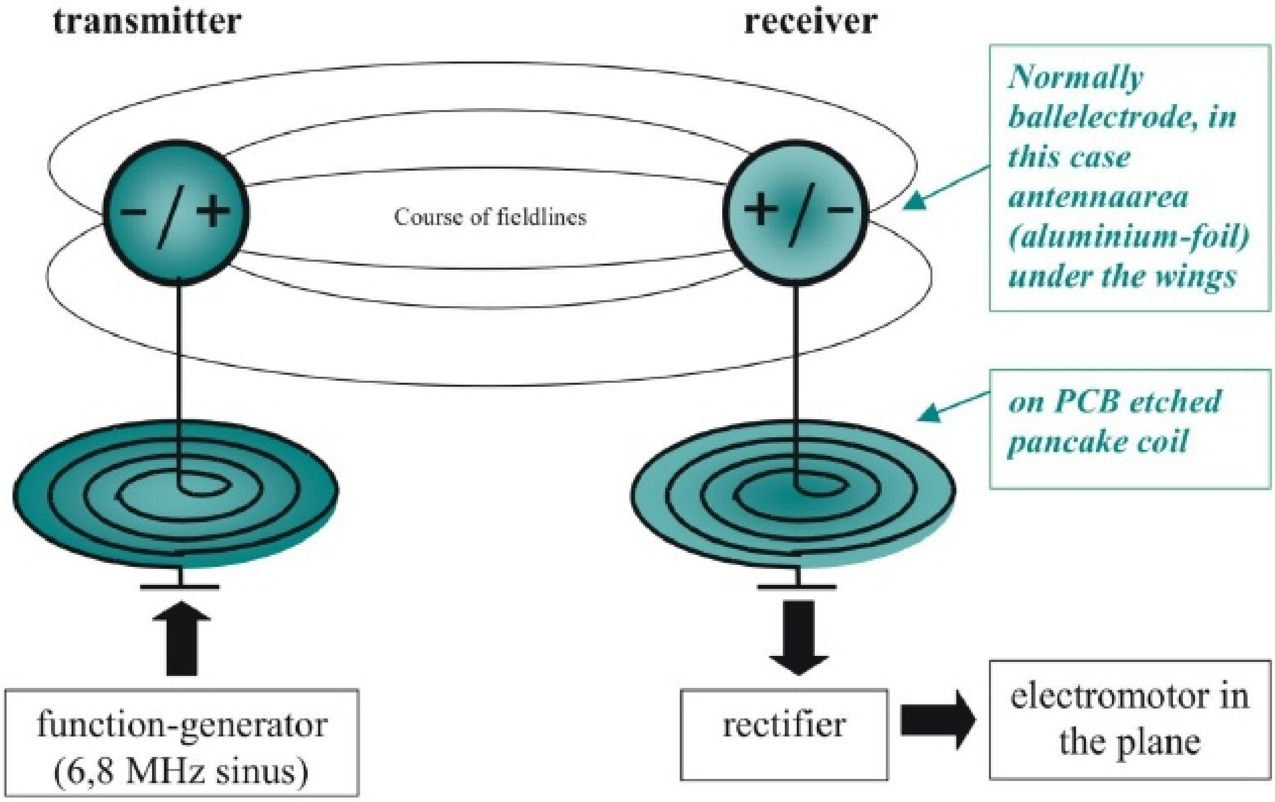

transmitter

receiver

Normally ballelectrode, in this case antennaarea (aluminium-foil) under the wings

on PCB etched pancake coil

–/+
+/–
Course of fieldlines

function-generator
(6,8 MHz sinus)

rectifier

electromotor in the plane

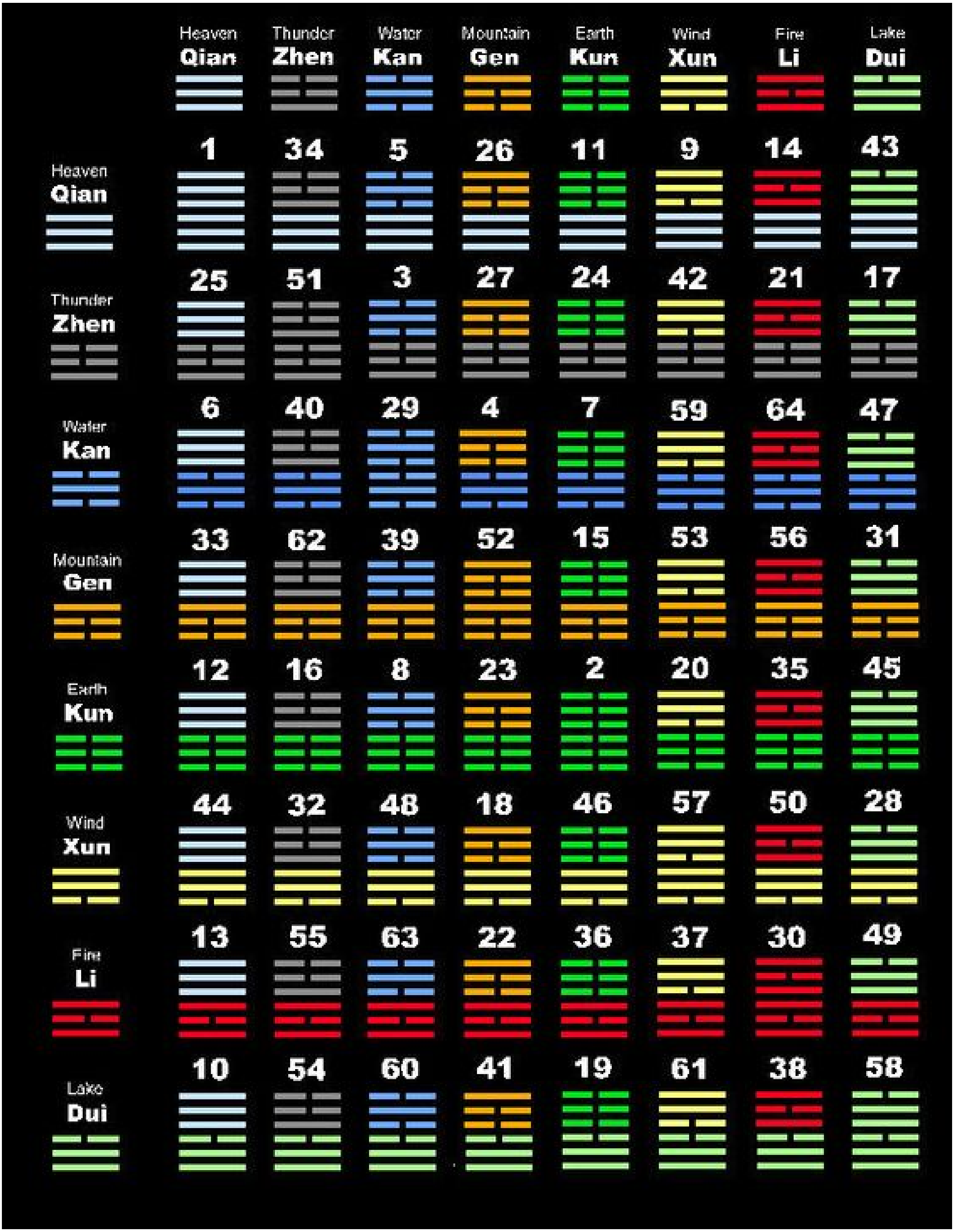

	Heaven Qian ☰	Thunder Zhen ☳	Water Kan ☵	Mountain Gen ☶	Earth Kun ☷	Wind Xun ☴	Fire Li ☲	Lake Dui ☱
Heaven Qian ☰	1	34	5	26	11	9	14	43
Thunder Zhen ☳	25	51	3	27	24	42	21	17
Water Kan ☵	6	40	29	4	7	59	64	47
Mountain Gen ☶	33	62	39	52	15	53	56	31
Earth Kun ☷	12	16	8	23	2	20	35	45
Wind Xun ☴	44	32	48	18	46	57	50	28
Fire Li ☲	13	55	63	22	36	37	30	49
Lake Dui ☱	10	54	60	41	19	61	38	58

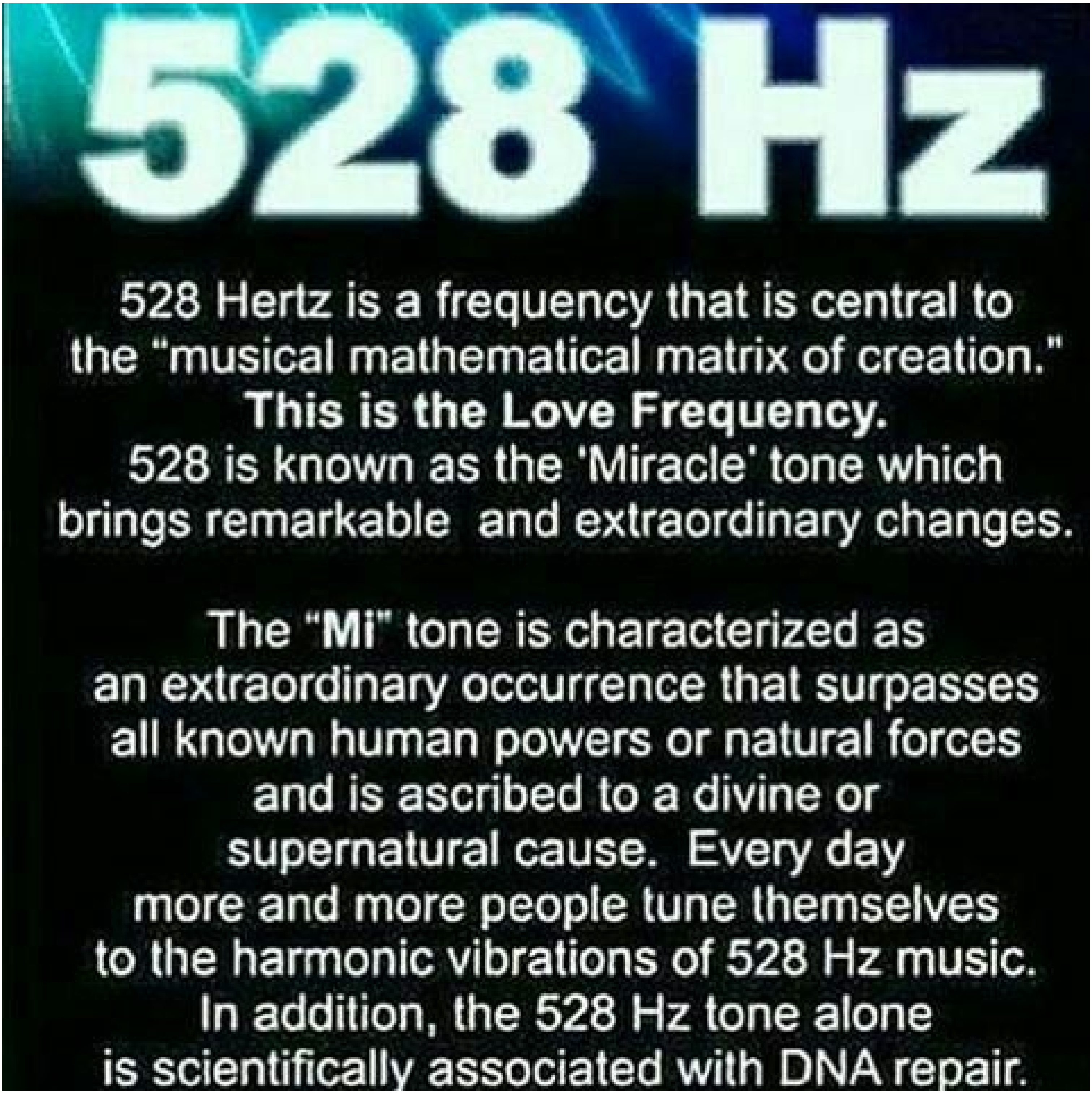

528 Hz

528 Hertz is a frequency that is central to the "musical mathematical matrix of creation."
This is the Love Frequency.
528 is known as the 'Miracle' tone which brings remarkable and extraordinary changes.

The "**Mi**" tone is characterized as an extraordinary occurrence that surpasses all known human powers or natural forces and is ascribed to a divine or supernatural cause. Every day more and more people tune themselves to the harmonic vibrations of 528 Hz music. In addition, the 528 Hz tone alone is scientifically associated with DNA repair.

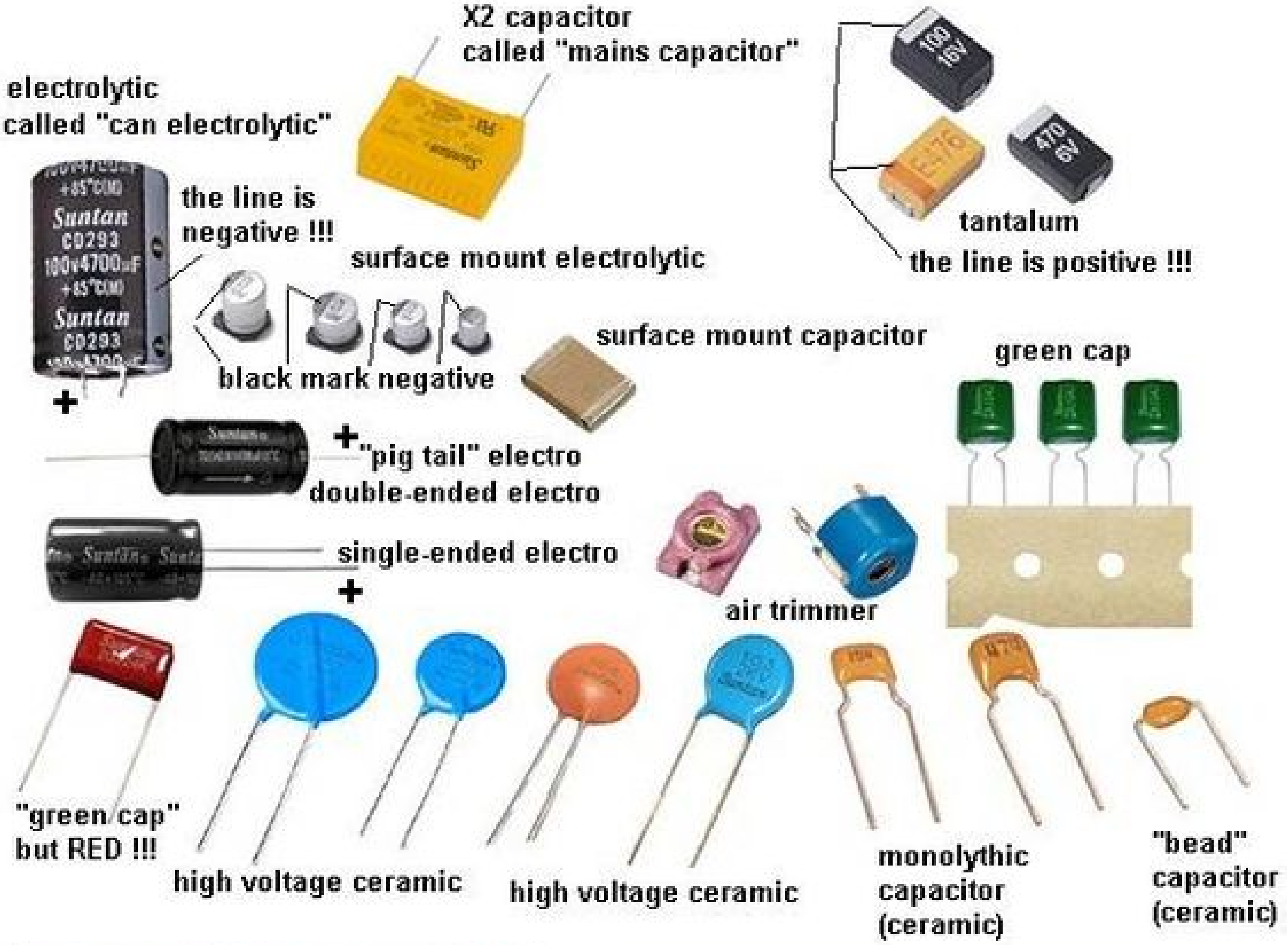
X2 capacitor
called "mains capacitor"
electrolytic
called "can electrolytic"
the line is
negative !!!
surface mount electrolytic
tantalum
the line is positive !!!
black mark negative
surface mount capacitor
green cap
+
+"pig tail" electro
double-ended electro
single-ended electro
+
air trimmer
"green cap"
but RED !!!
high voltage ceramic
high voltage ceramic
monolythic
capacitor
(ceramic)
"bead"
capacitor
(ceramic)

tantalum capacitors can
"go up in smoke"

"bead" tantalum

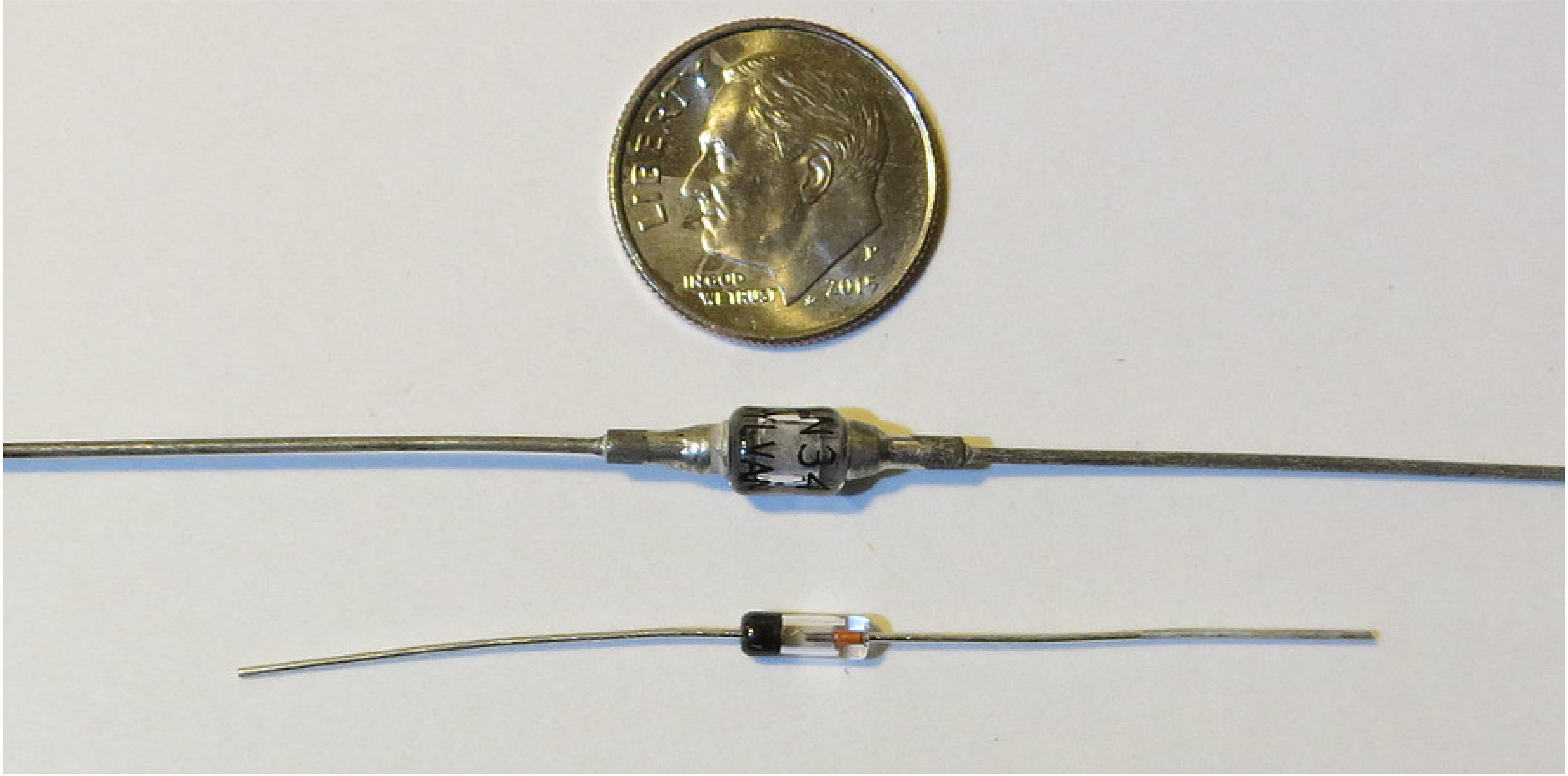
LIBERTY
IN GOD WE TRUST
2015
P

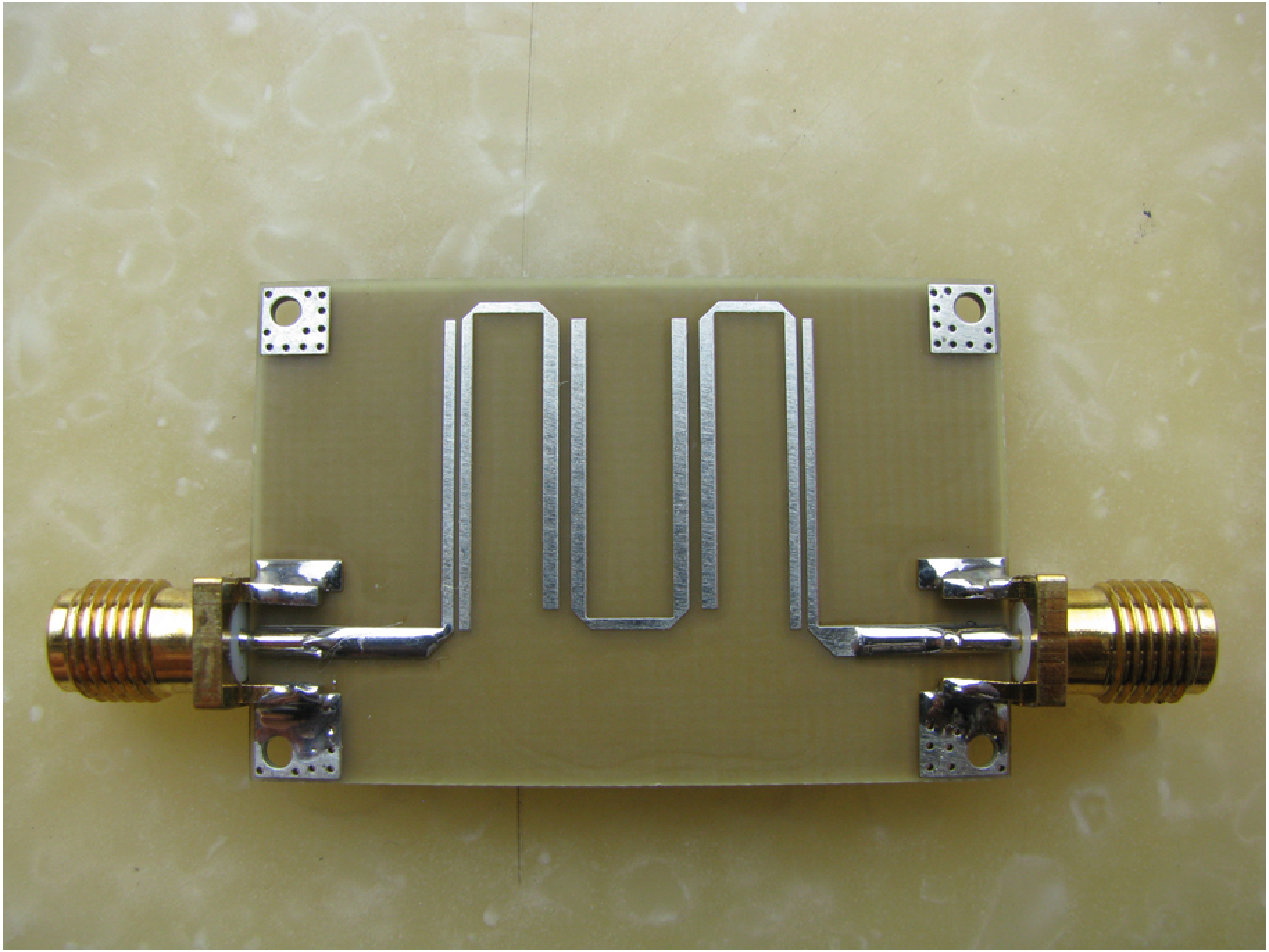

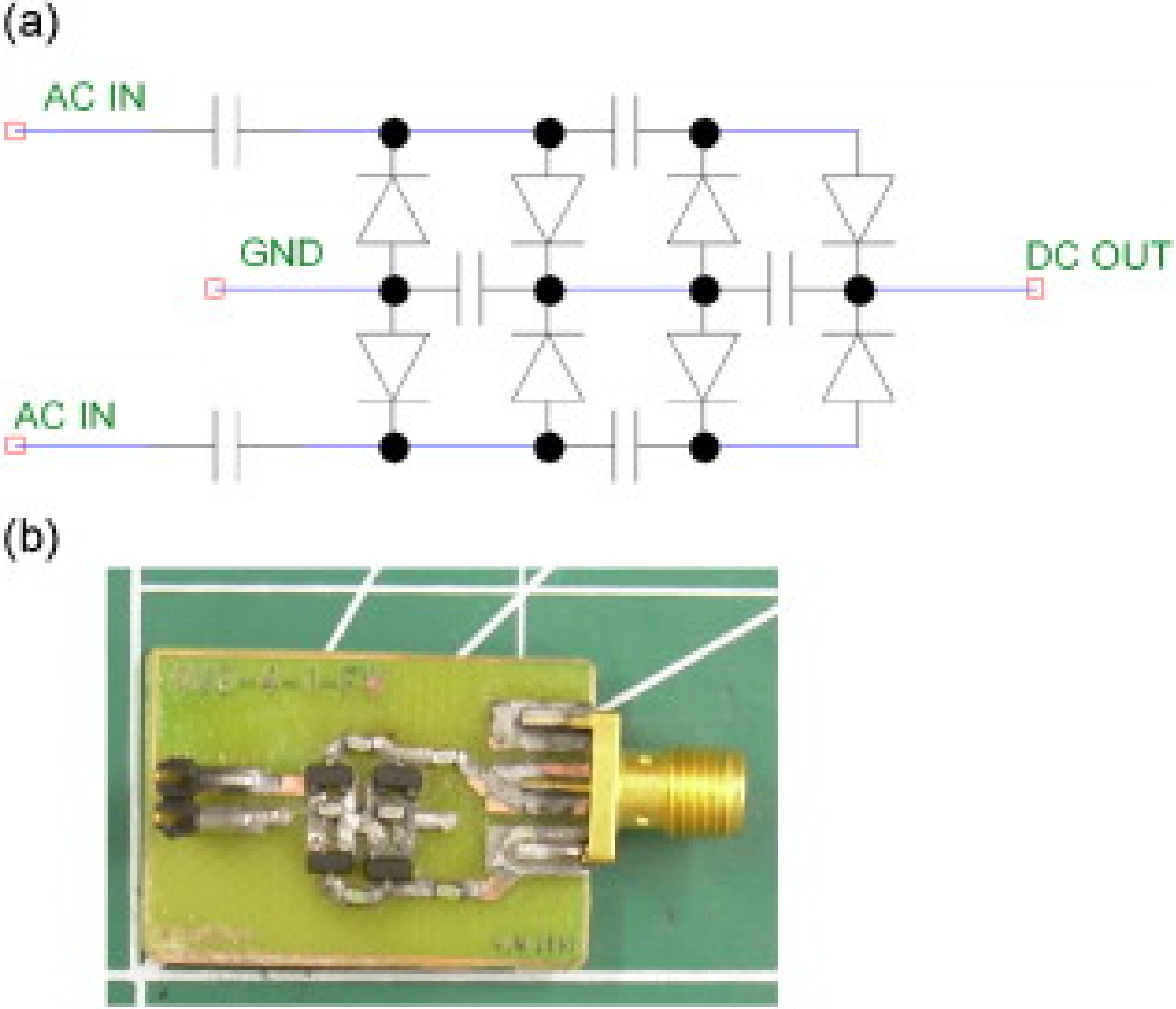
(a)
AC IN
GND
DC OUT
AC IN
(b)

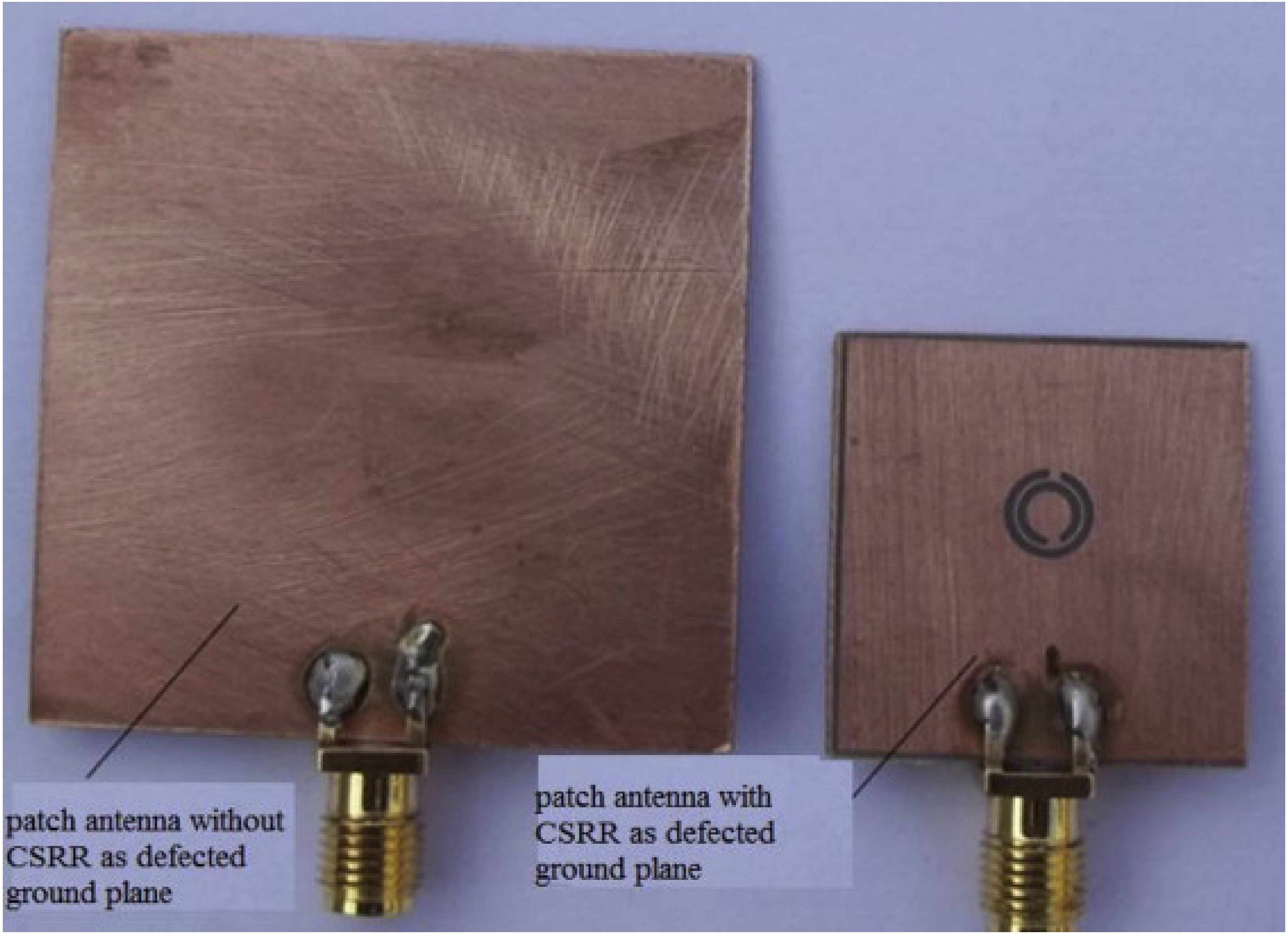
patch antenna without CSRR as defected ground plane
patch antenna with CSRR as defected ground plane

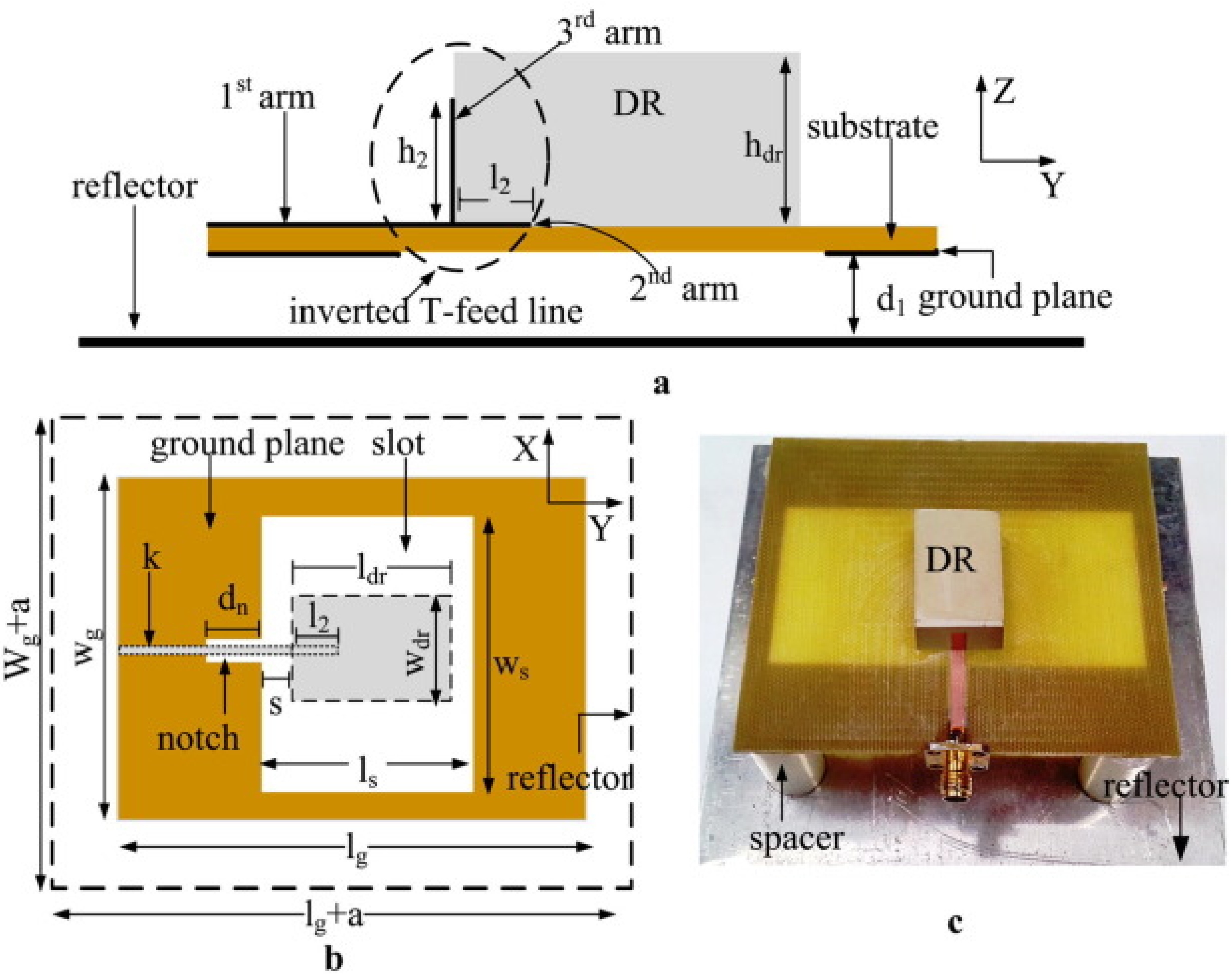
3rd arm
1st arm
DR
substrate
Z
Y
h2
hdr
l2
reflector
2nd arm
inverted T-feed line
d1 ground plane
a
ground plane
slot
X
Y
k
ldr
dn
l2
wdr
ws
Wg+a
wg
s
notch
ls
reflector
lg
lg+a
b
DR
spacer
reflector
c

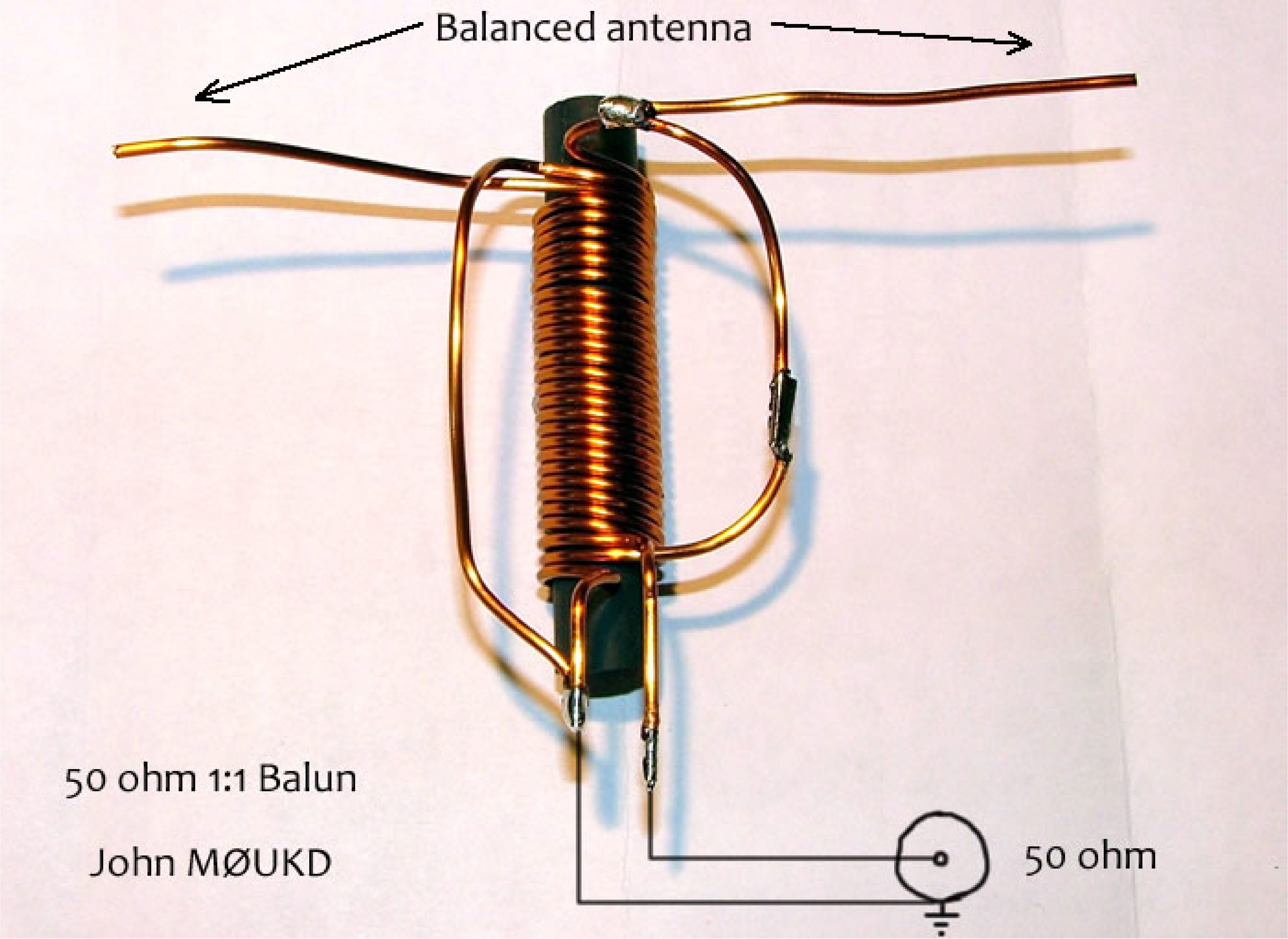
Balanced antenna
50 ohm 1:1 Balun
John MØUKD
50 ohm

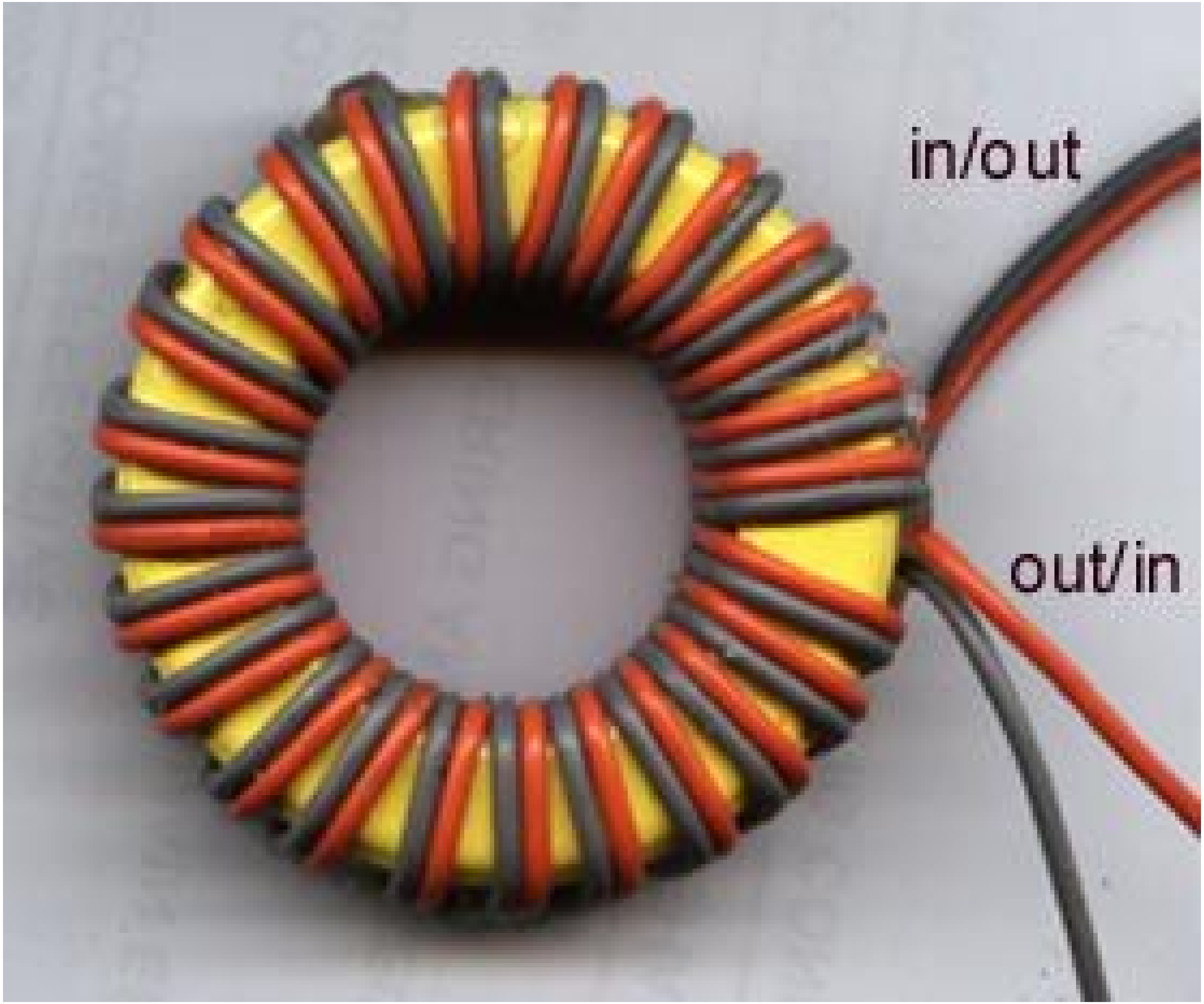
in/out
out/in

1:1 Voltage balun ("Ruthroff")

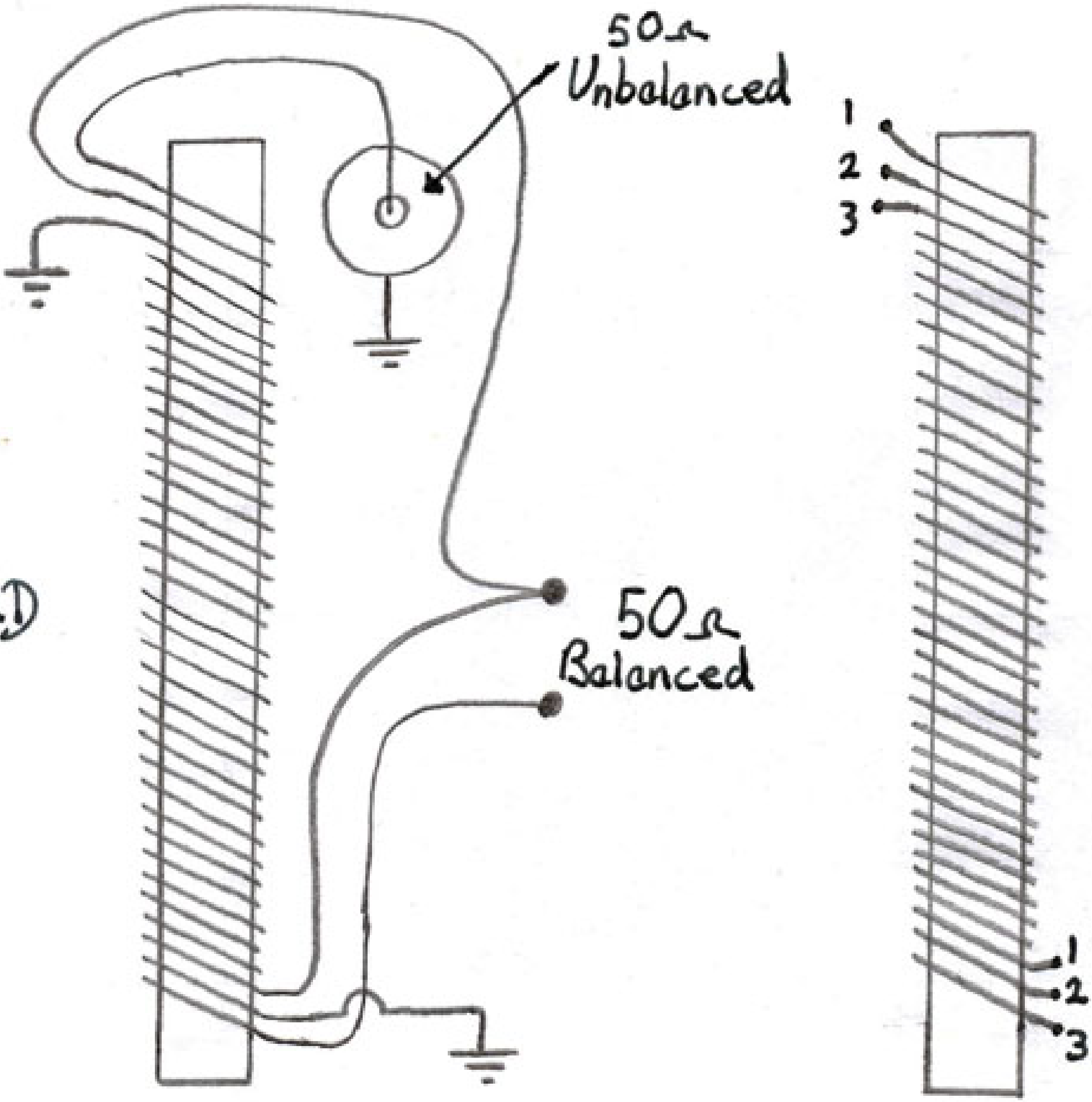
50Ω
Unbalanced
50Ω
Balanced
1
2
3
1
2
3

M0UKD

10 - 15 trifilar turns on ferrite rod.

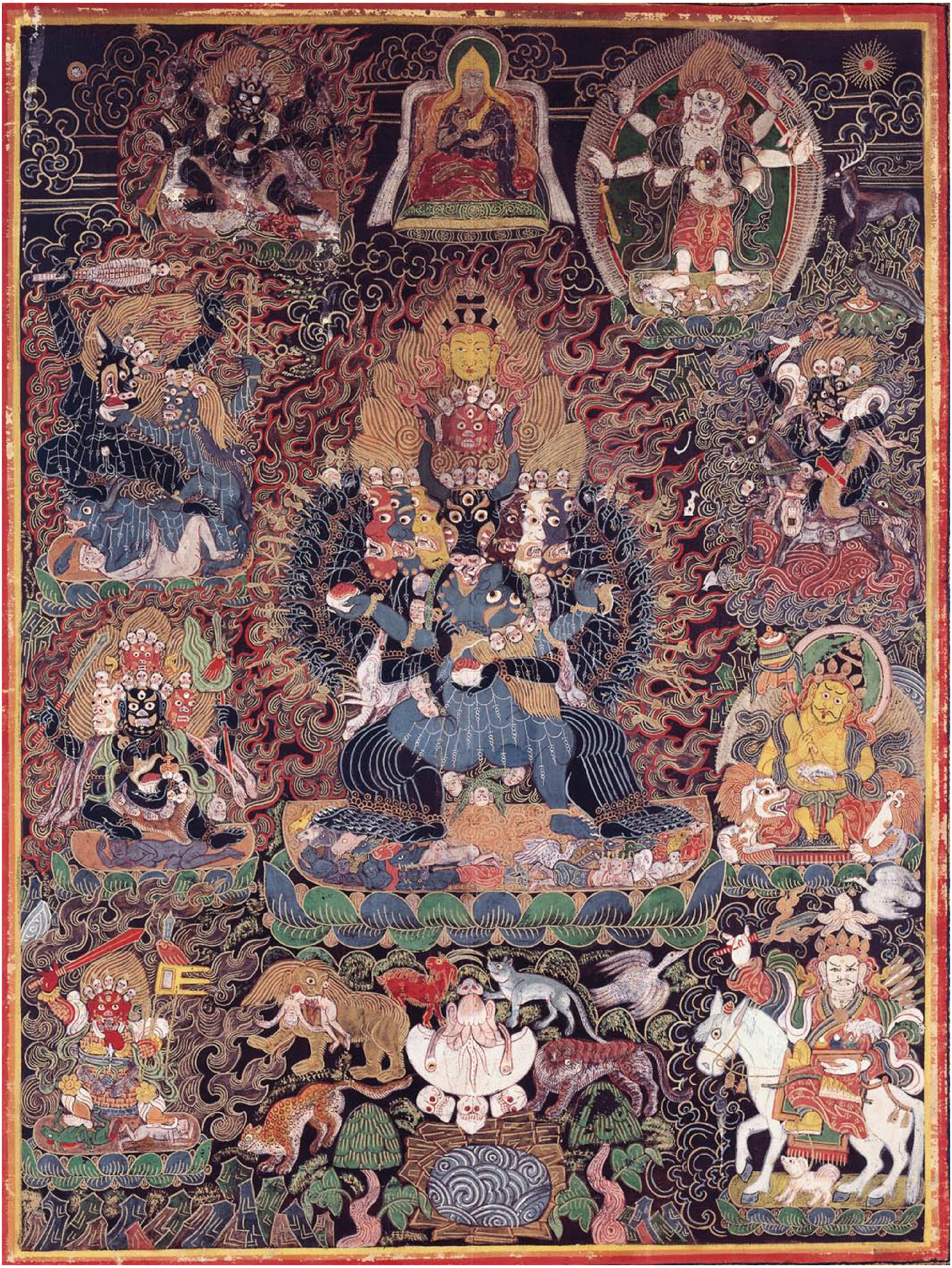

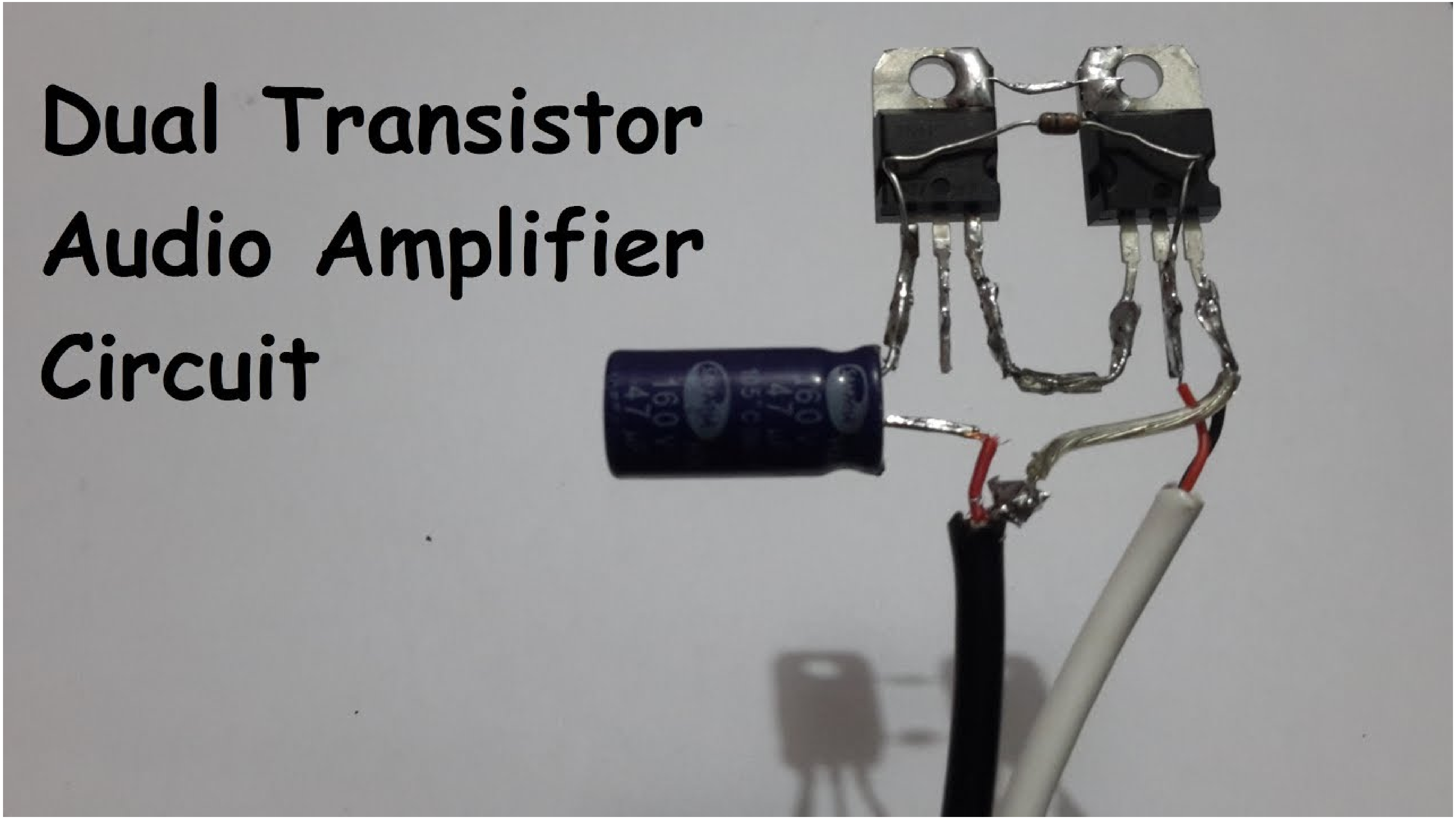
Dual Transistor
Audio Amplifier
Circuit

RECEIVERS FOR BEGINNERS . . .

local broadcasting stations in dual-type headsets with an impedance of 2000 ohms, or higher. The antenna should be a high, well-insulated outdoor type at least 100 feet long, although antennas as short as 40 feet will work. The ground should be made to a cold-water pipe.

The practical receiving range of any crystal receiver is limited to about 40 miles for powerful broadcasting stations. Set No. 2 is double-tuned for improved selectivity. Two coils like the one used in set No. 1 are employed and each coil is tuned by one section of a dual variable condenser. Coupling between the two tuned circuits is by means of a "gimmick" made by twisting two pieces of insulated wire together for a length of about 4 in. This is condenser C3. The longer this "gimmick" the louder the signals, but the poorer the selectivity.

Set No. 3 is an experimental arrangement using two 1N34 diode crystals in push-pull to provide louder signals than conventional crystal circuits. The coils are hand-wound to provide a split secondary winding with a primary winding between. Both primary and secondary are tuned. The dual-section variable condenser tunes the secondary of the coil to the frequency of the station. Condenser C2 in primary L3 tunes the station but it has its greatest effect on very long antennas. If the desired station operates on 850 kc. or higher, "open" the toggle switch on the bracket. Detailed student material list is R-396.

C

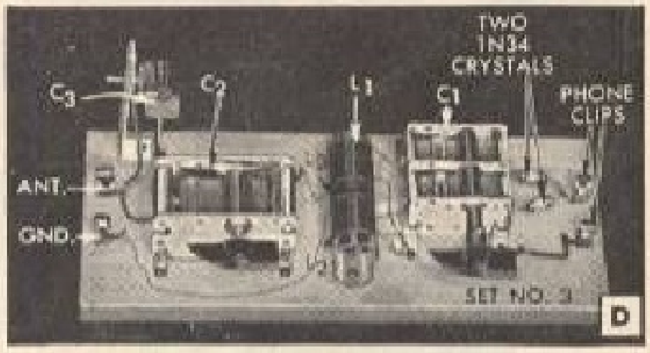
TWO
1N34
CRYSTALS
C3
C2
L1
C1
PHONE
CLIPS
ANT.
GND.
L2
SET NO. 3
D

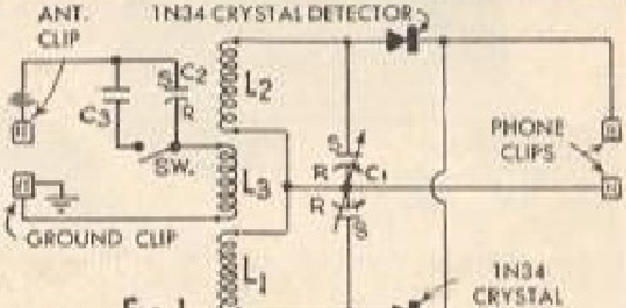
ANT.
CLIP
1N34 CRYSTAL DETECTOR
C2
S
R
C3
L2
SW.
L3
R
C1
S
R
S
PHONE
CLIPS
GROUND CLIP
L1
1N34
CRYSTAL
DETECTOR
SET NO. 3 FIG. 1

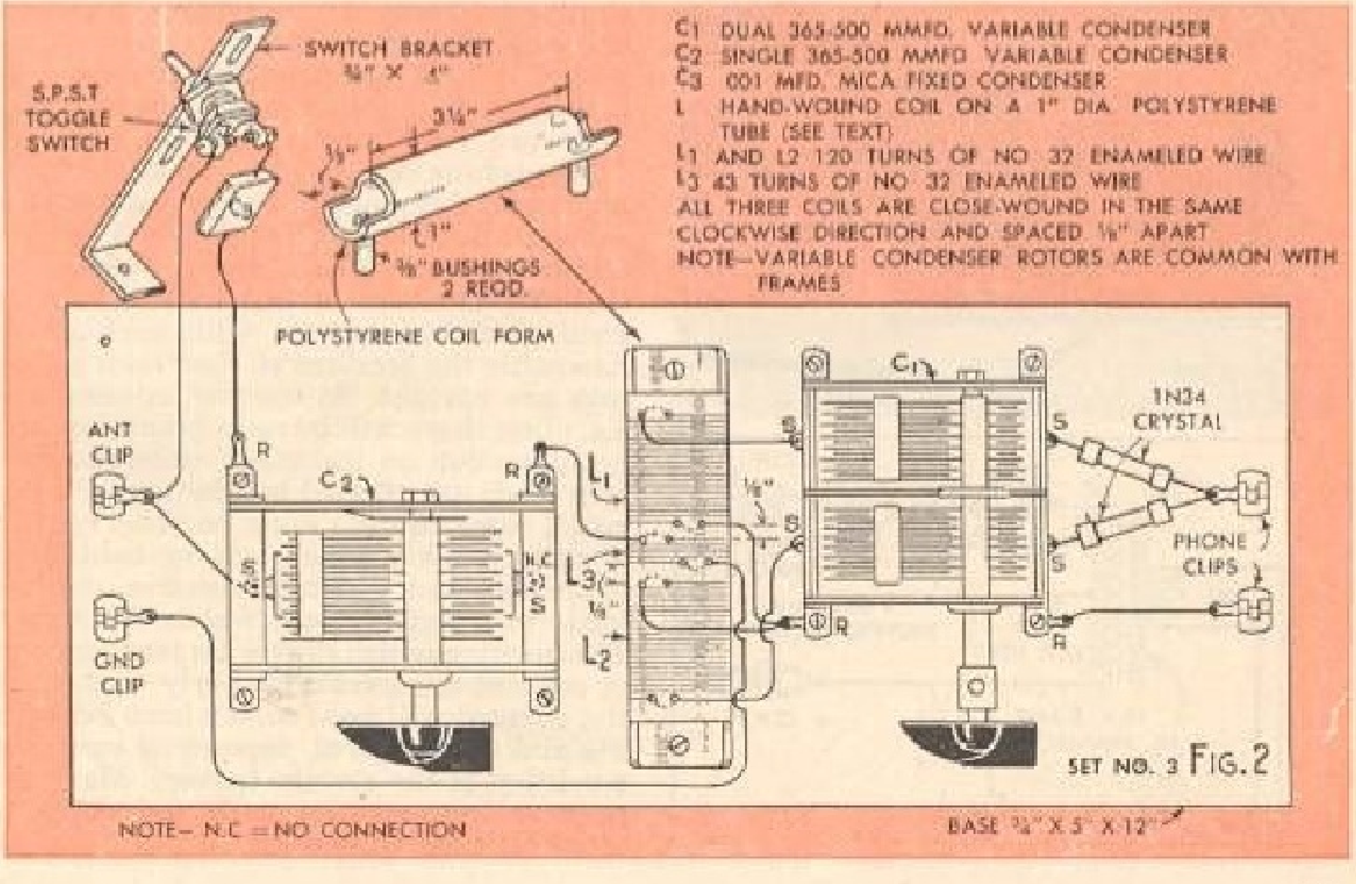
SWITCH BRACKET
5/8" X 4"
S.P.S.T
TOGGLE
SWITCH
3 1/4"
1/2"
1"
1/8" BUSHINGS
2 REQD.
POLYSTYRENE COIL FORM
ANT
CLIP
R
C2
R
L1
C1
S
1N34
CRYSTAL
S
N.C
S
L3
1/8"
1/8"
S
S
PHONE
CLIPS
GND
CLIP
L2
R
R
NOTE— N.C = NO CONNECTION
BASE 1/2" X 5" X 12"

C1 DUAL 365-500 MMFD. VARIABLE CONDENSER
C2 SINGLE 365-500 MMFD VARIABLE CONDENSER
C3 .001 MFD. MICA FIXED CONDENSER
L HAND-WOUND COIL ON A 1" DIA. POLYSTYRENE TUBE (SEE TEXT)
L1 AND L2 120 TURNS OF NO. 32 ENAMELED WIRE
L3 43 TURNS OF NO. 32 ENAMELED WIRE
ALL THREE COILS ARE CLOSE-WOUND IN THE SAME CLOCKWISE DIRECTION AND SPACED 1/8" APART
NOTE—VARIABLE CONDENSER ROTORS ARE COMMON WITH FRAMES

SET NO. 3 FIG. 2

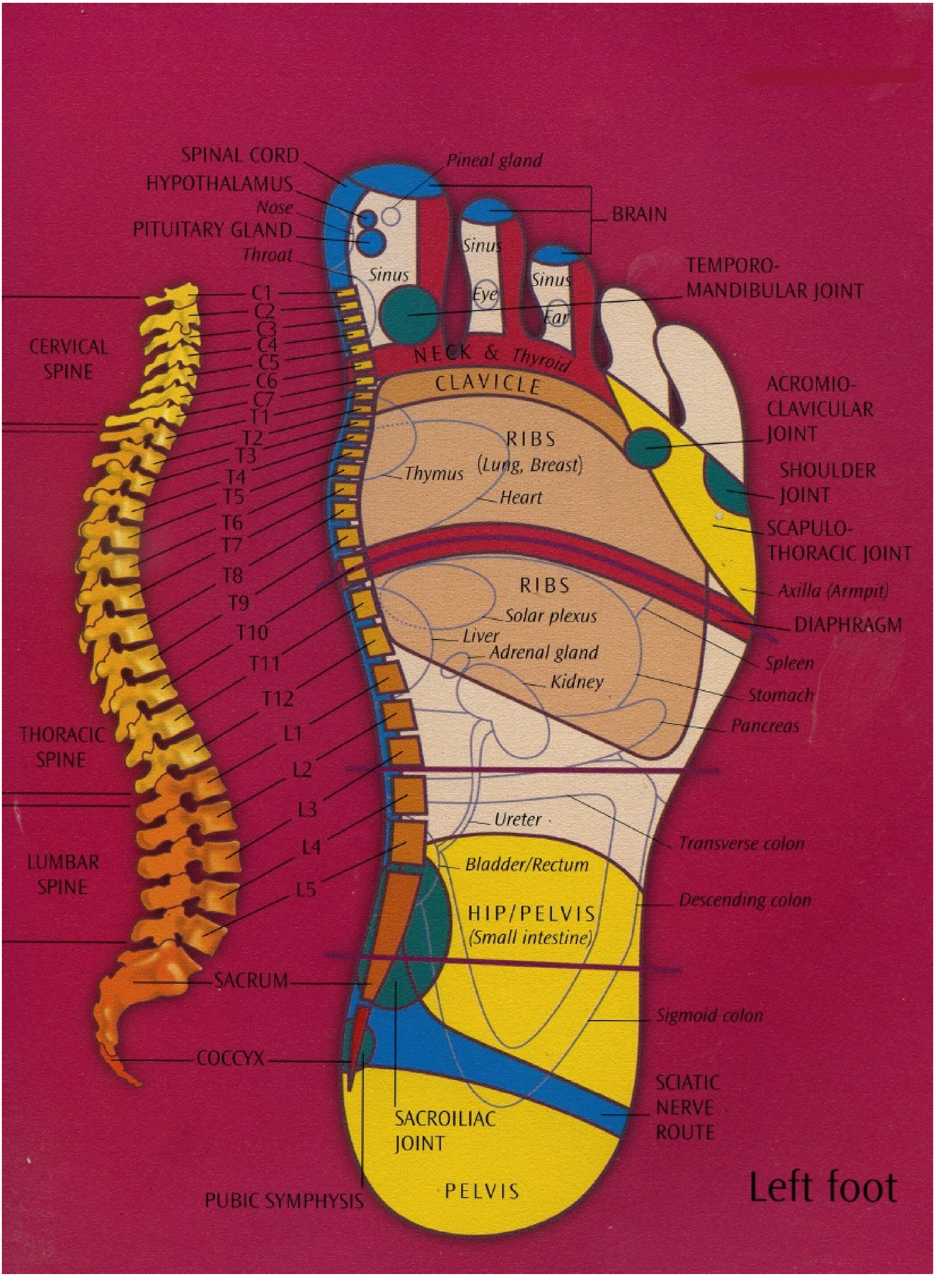
SPINAL CORD
HYPOTHALAMUS
Nose
PITUITARY GLAND
Throat
Pineal gland
BRAIN
Sinus
Sinus
Sinus
Eye
Ear
TEMPORO-MANDIBULAR JOINT
CERVICAL SPINE
C1
C2
C3
C4
C5
C6
C7
T1
T2
T3
T4
T5
T6
T7
T8
T9
T10
T11
T12
L1
L2
L3
L4
L5
NECK & Thyroid
CLAVICLE
ACROMIO-CLAVICULAR JOINT
RIBS
(Lung, Breast)
Thymus
Heart
SHOULDER JOINT
SCAPULO-THORACIC JOINT
RIBS
Axilla (Armpit)
Solar plexus
DIAPHRAGM
Liver
Adrenal gland
Spleen
Kidney
Stomach
THORACIC SPINE
Pancreas
Ureter
Transverse colon
LUMBAR SPINE
Bladder/Rectum
Descending colon
HIP/PELVIS
(Small intestine)
SACRUM
Sigmoid colon
COCCYX
SCIATIC NERVE ROUTE
SACROILIAC JOINT
PUBIC SYMPHYSIS
PELVIS

Left foot

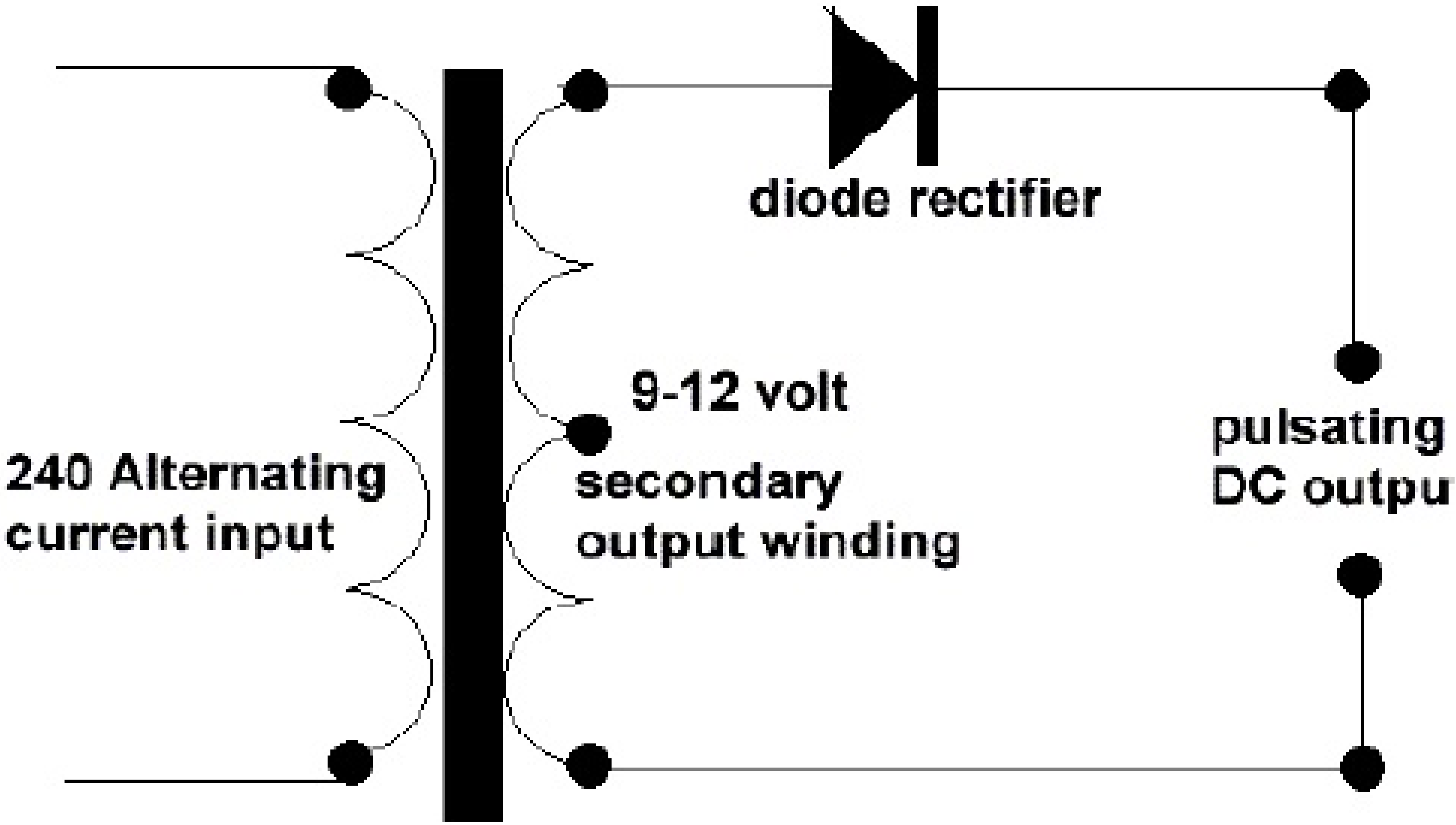
diode rectifier
9-12 volt
240 Alternating current input
secondary output winding
pulsating DC outpu

typical schematic for 1/2 wave rectifier

ELF, GWEN Towers, and HAARP Connection
These things are pretty sinister and all over the U.S.
Click "see more" to read article

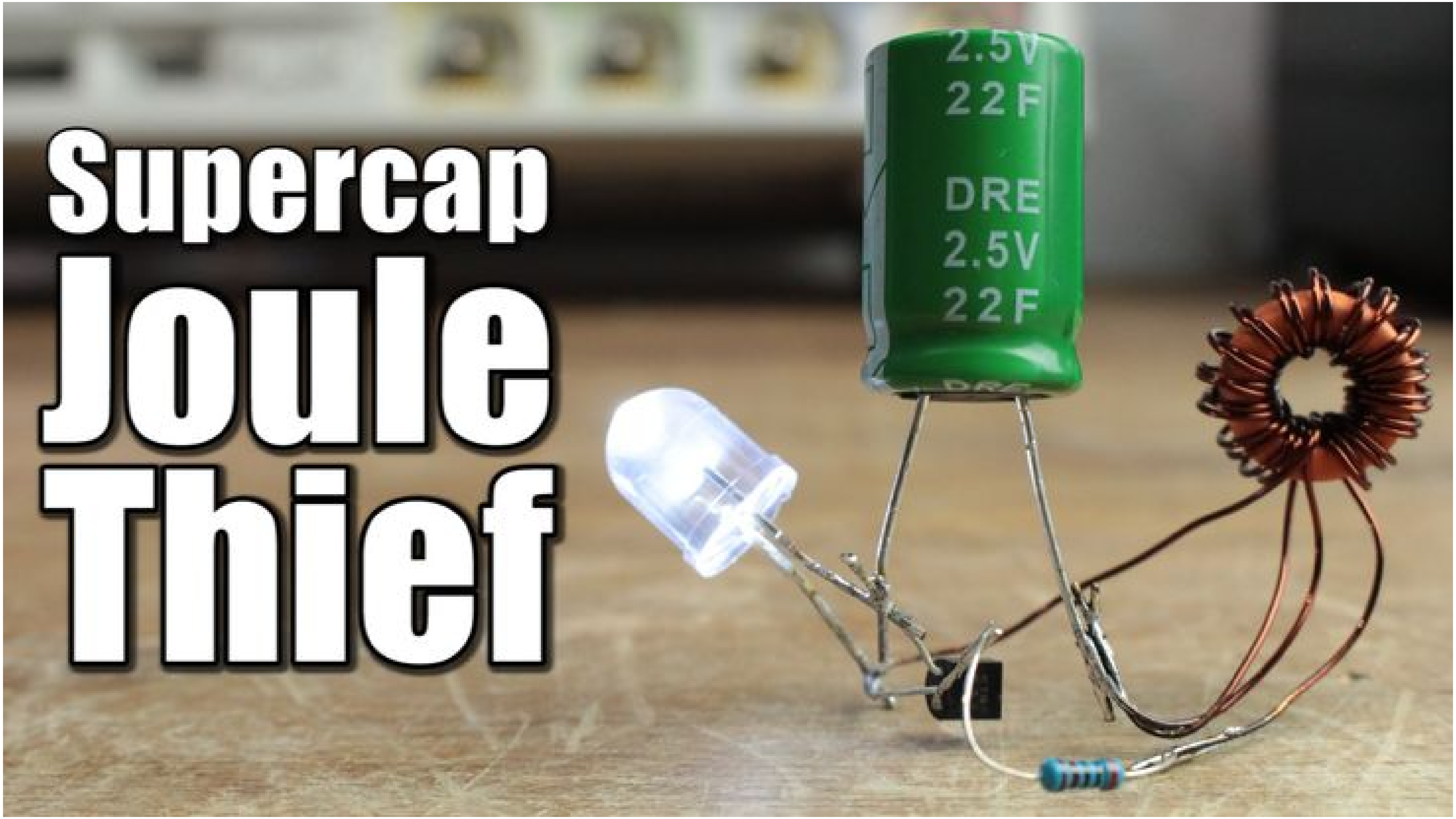
Supercap
Joule
Thief
2.5V
22F
DRE
2.5V
22F

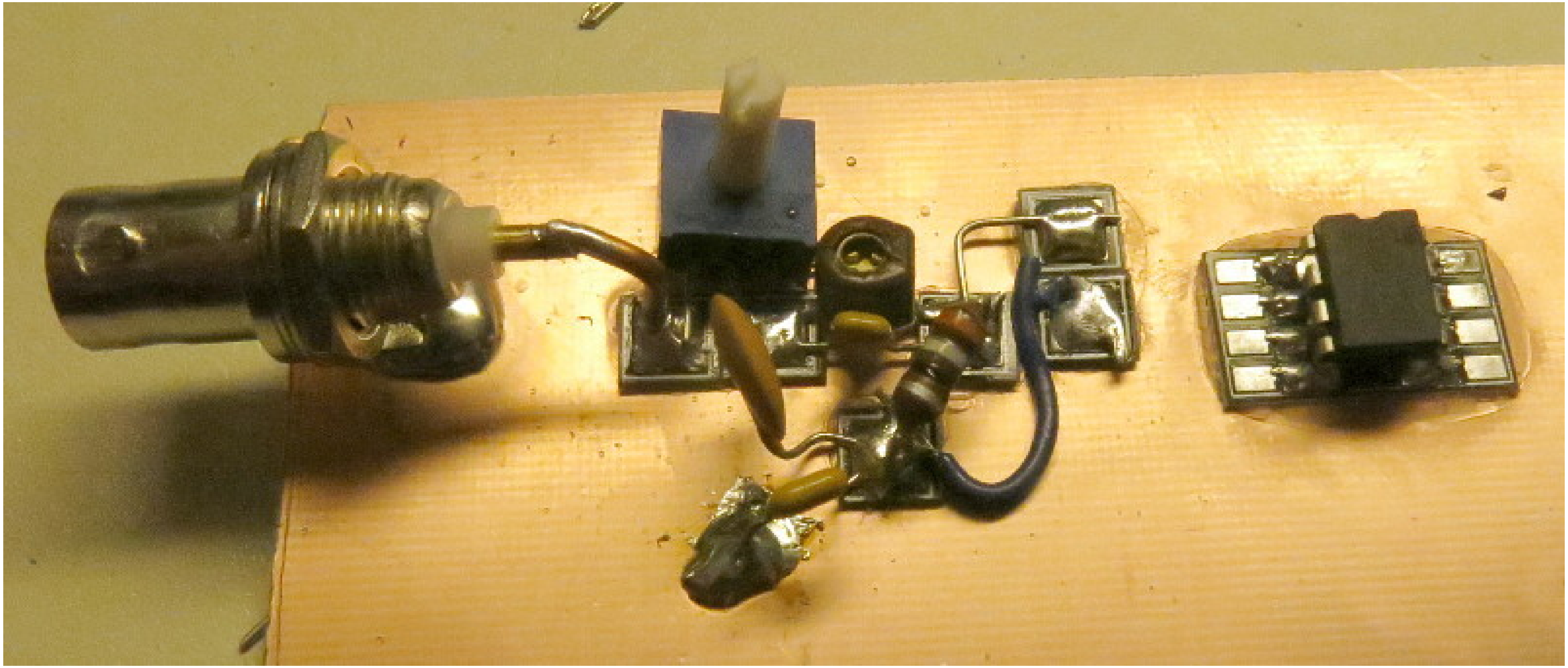

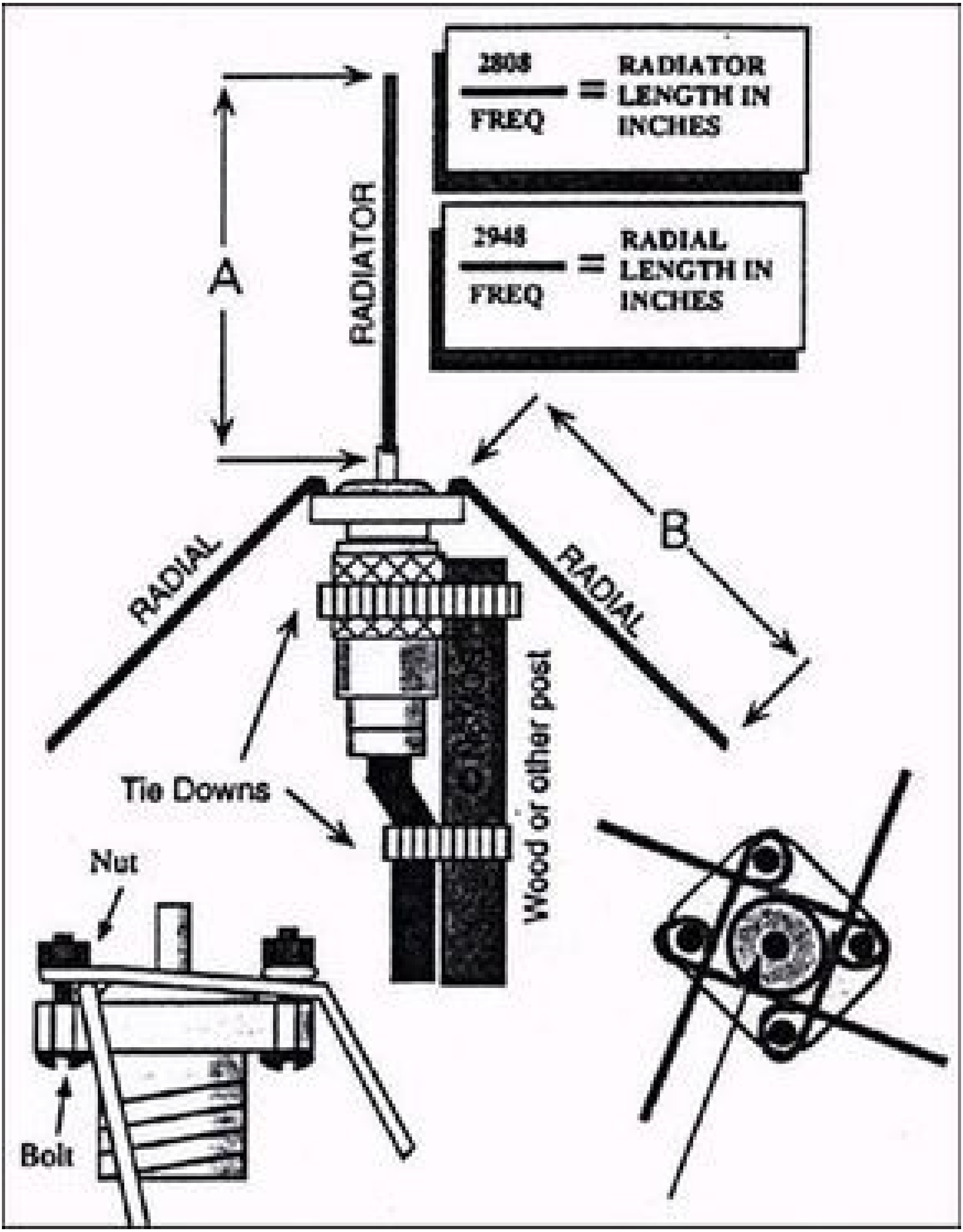
2808
FREQ
= RADIATOR LENGTH IN INCHES
2948
FREQ
= RADIAL LENGTH IN INCHES
A
RADIATOR
B
RADIAL
RADIAL
Tie Downs
Wood or other post
Nut
Bolt

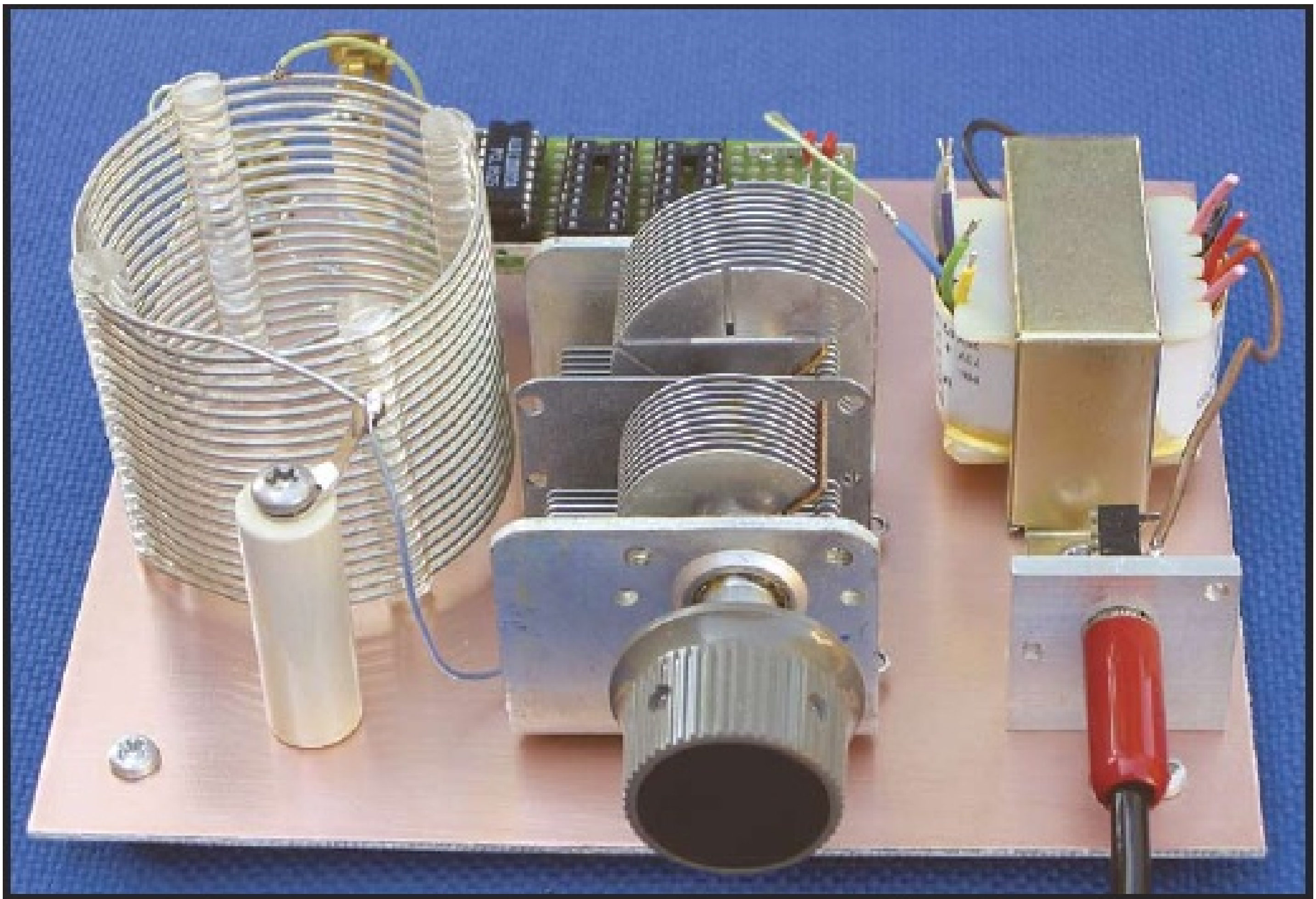

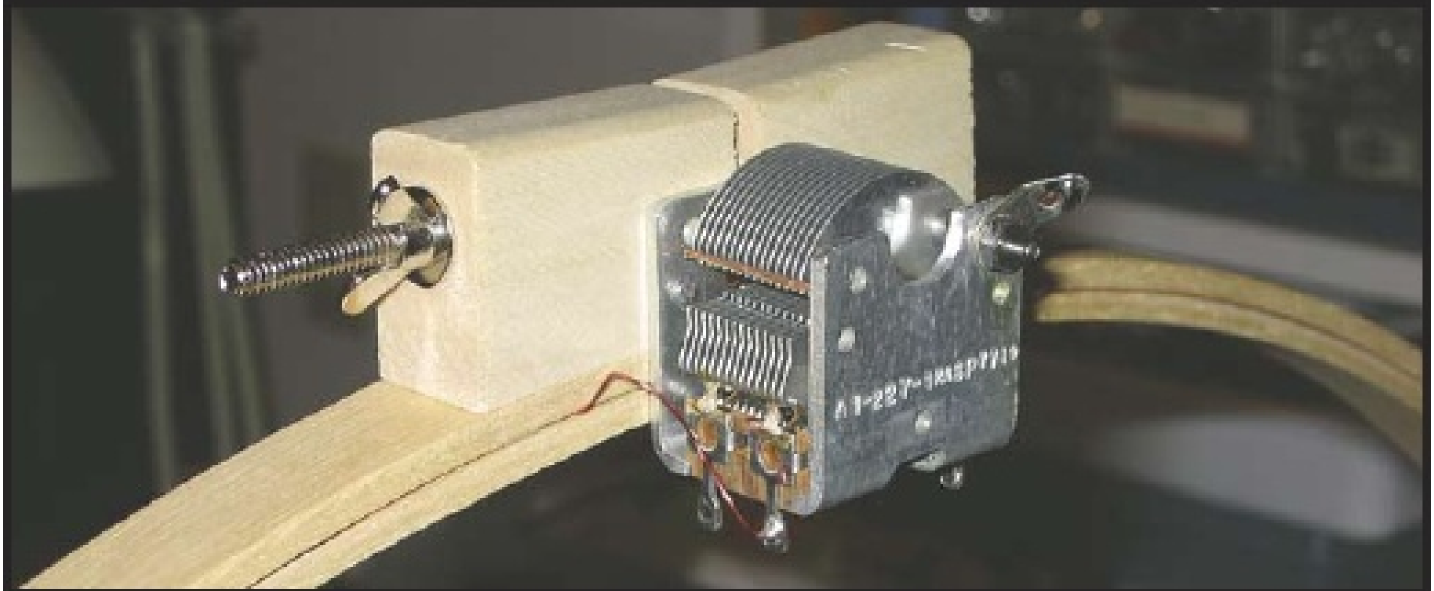

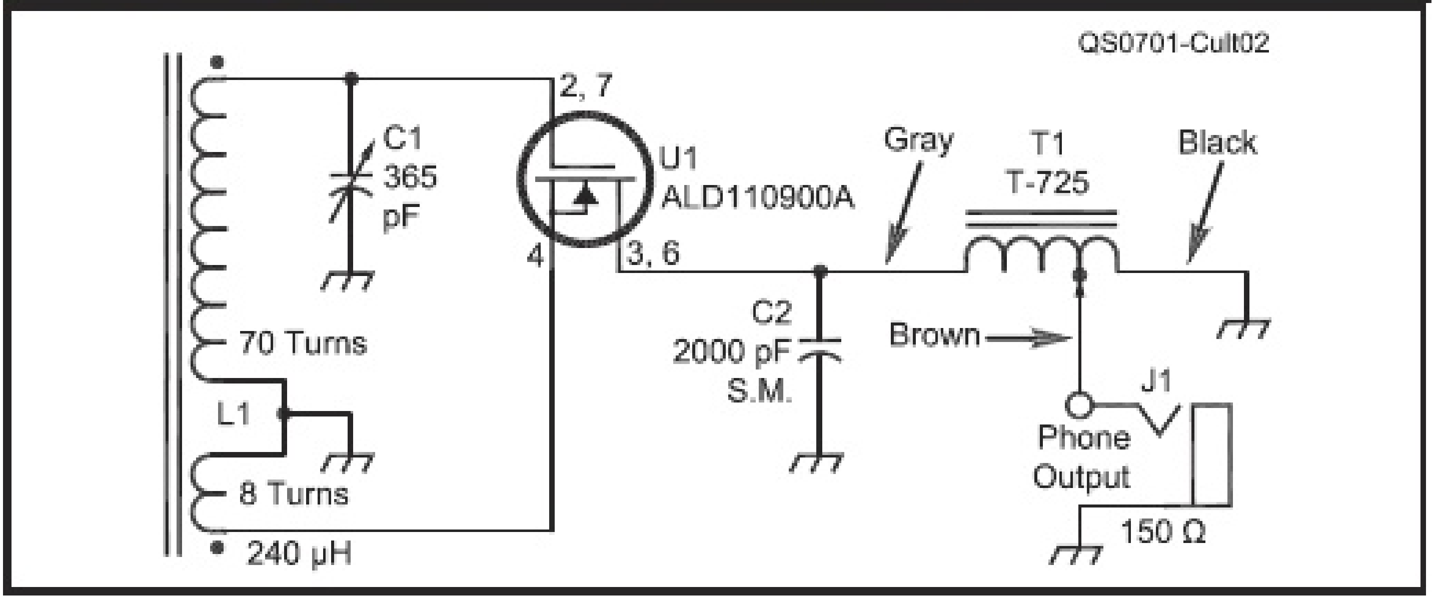
QS0701-Cult02
2, 7
C1
365
pF
U1
ALD110900A
4
3, 6
Gray
T1
T-725
Black
Brown
C2
2000 pF
S.M.
70 Turns
L1
8 Turns
240 µH
J1
Phone
Output
150 Ω

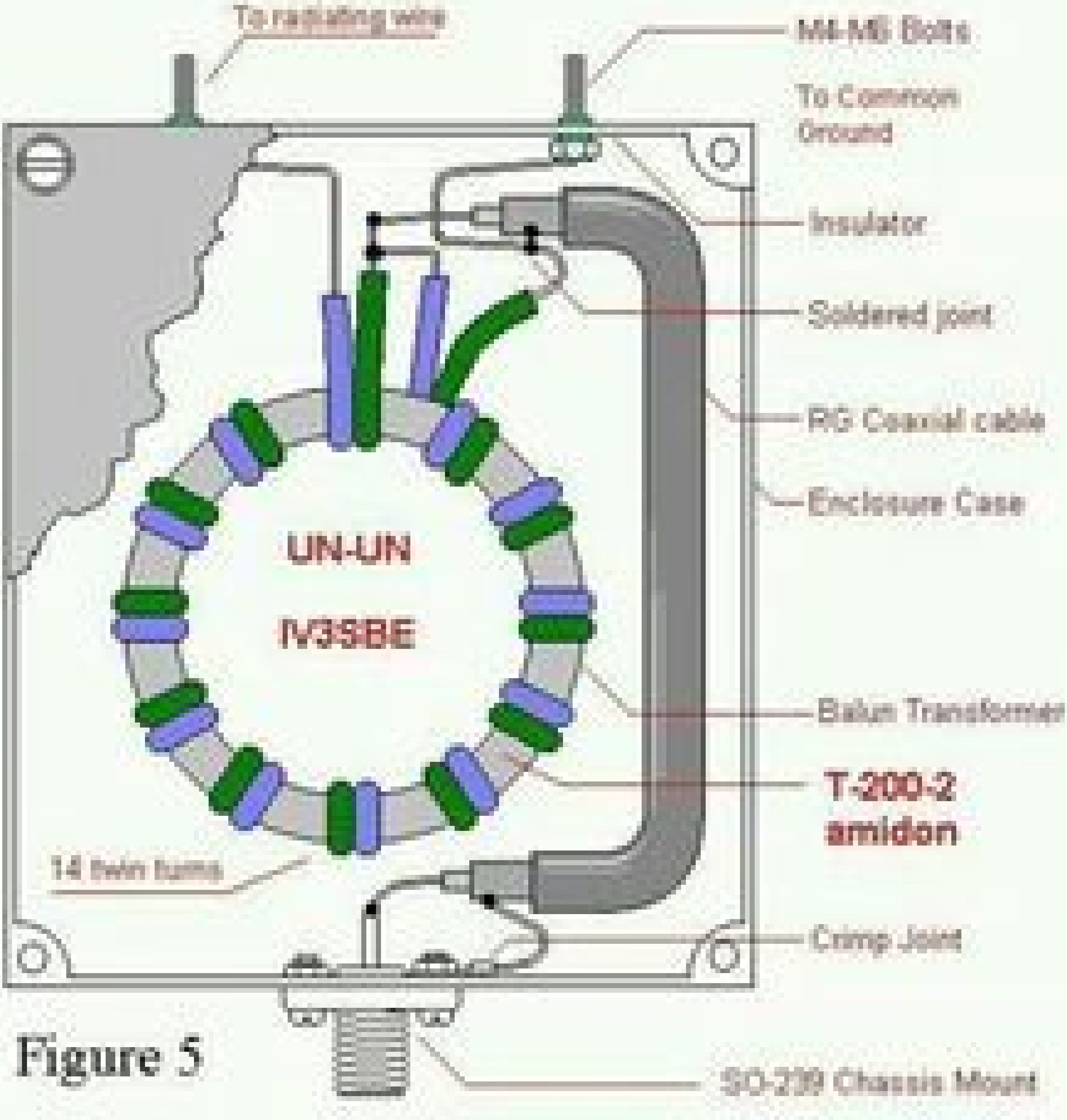
To radiating wire
M4-M6 Bolts
To Common Ground
Insulator
Soldered joint
RG Coaxial cable
Enclosure Case
UN-UN
IV3SBE
Balun Transformer
T-200-2
amidon
14 twin turns
Crimp Joint
SO-239 Chassis Mount

Figure 5

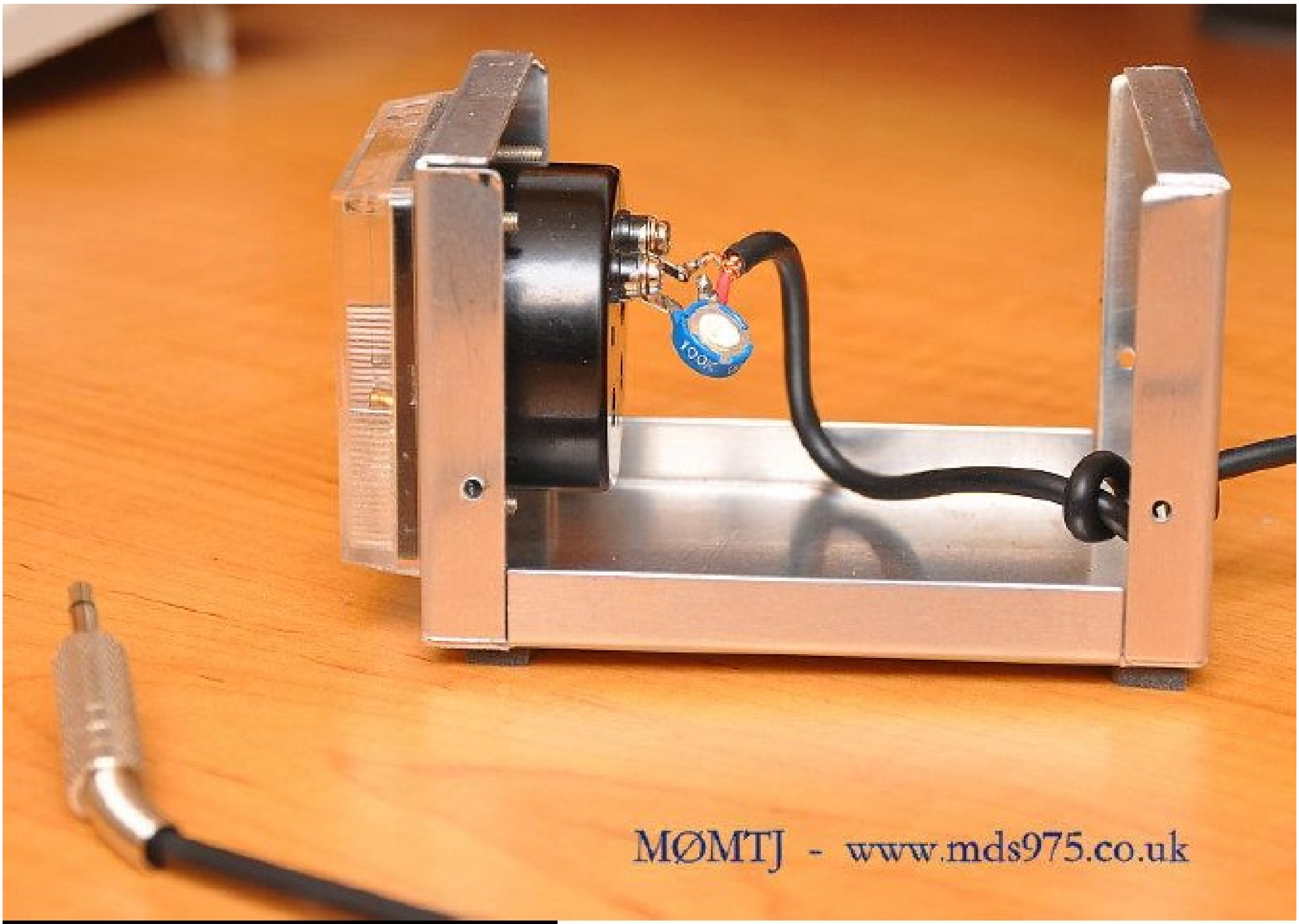
100K
M0MTJ - www.mds975.co.uk

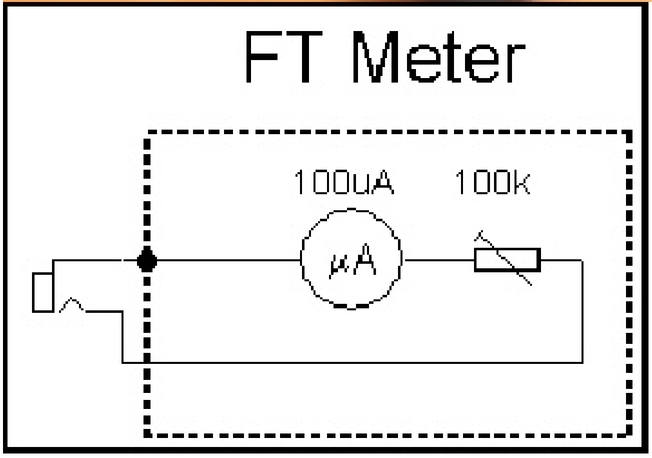
FT Meter
100uA
100k
μA

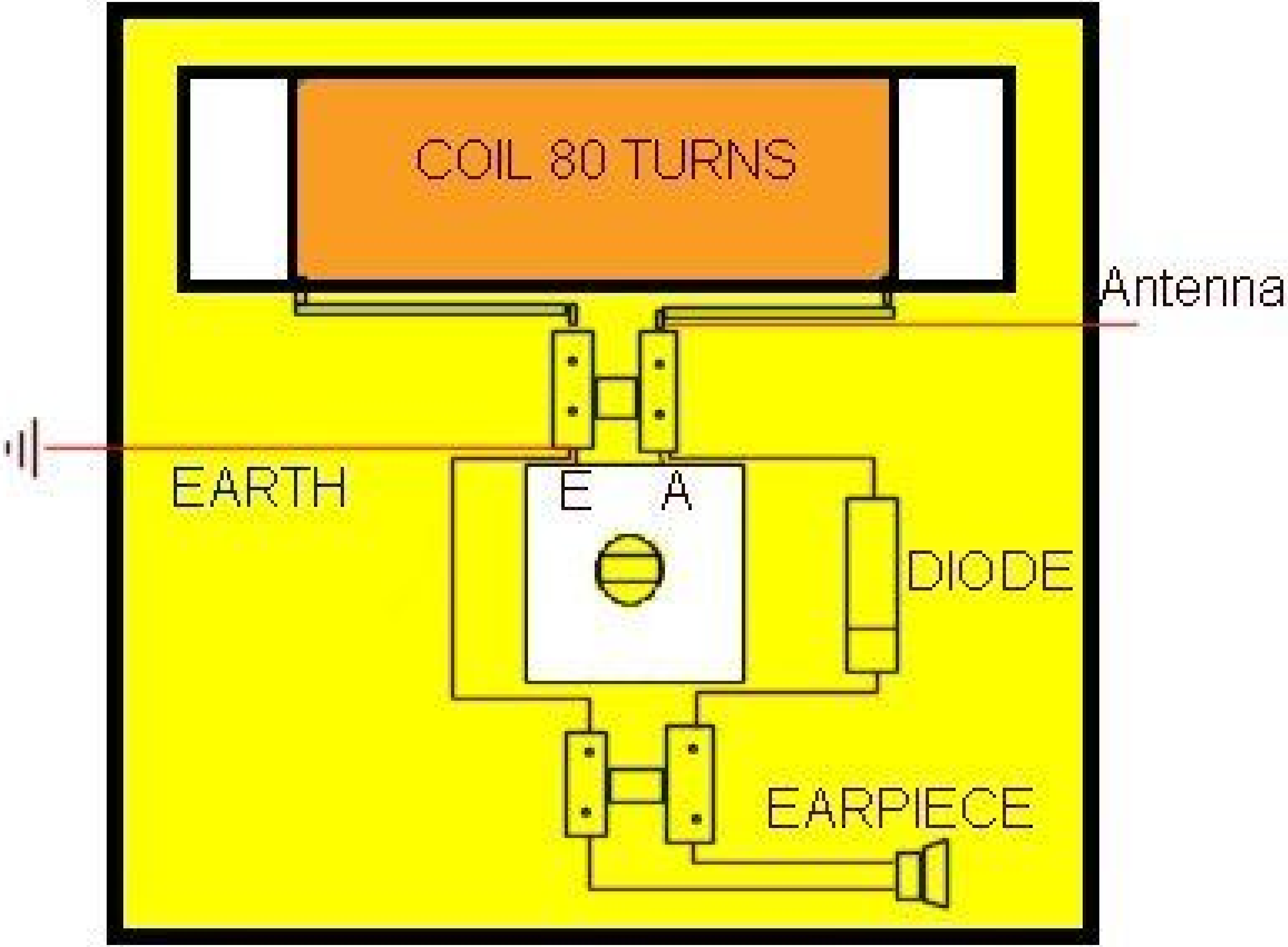
COIL 80 TURNS
Antenna
EARTH
E
A
DIODE
EARPIECE

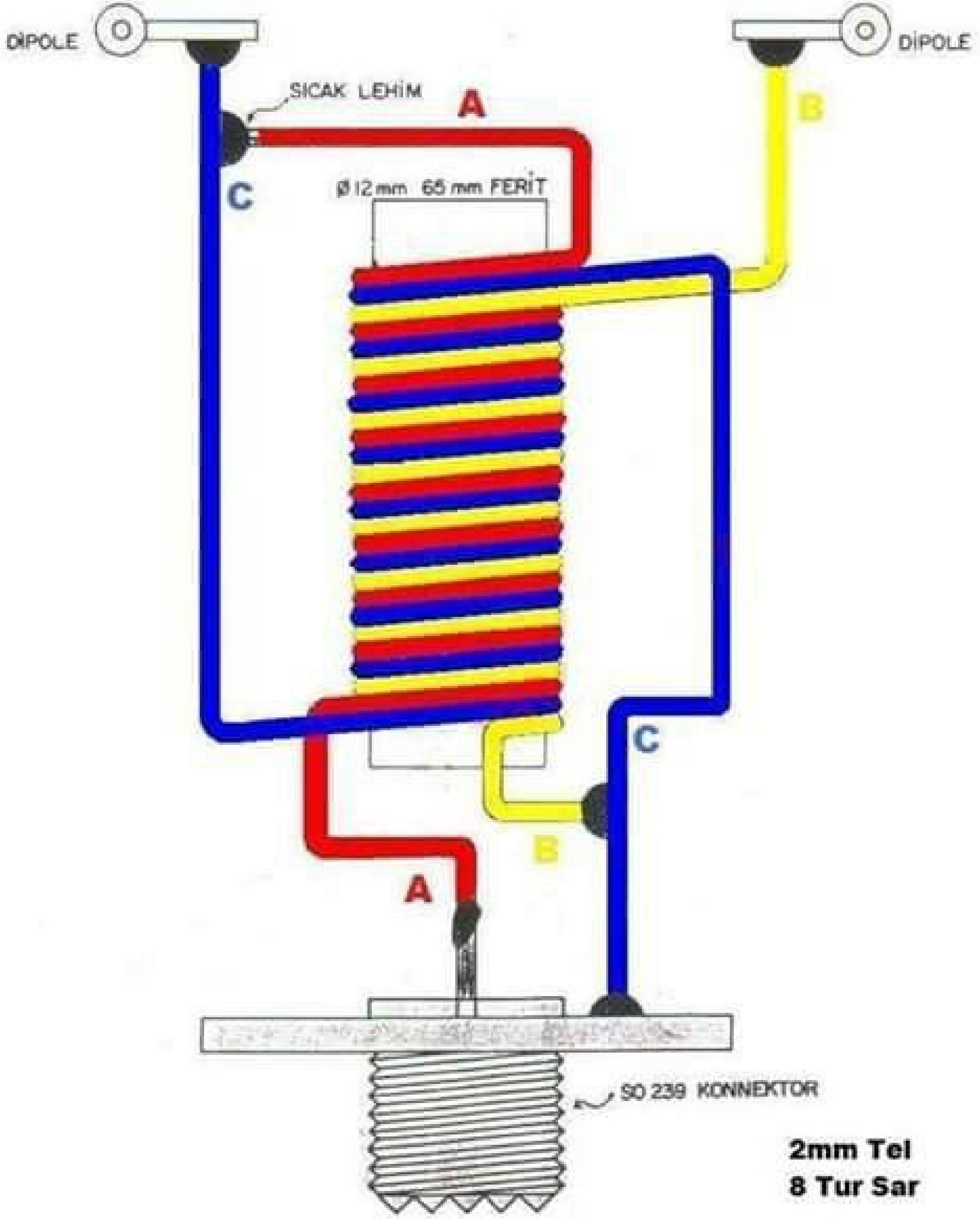
DİPOLE
DİPOLE
SICAK LEHİM
A
B
C
Ø 12 mm 65 mm FERİT
C
B
A
SO 239 KONNEKTOR

2mm Tel
8 Tur Sar

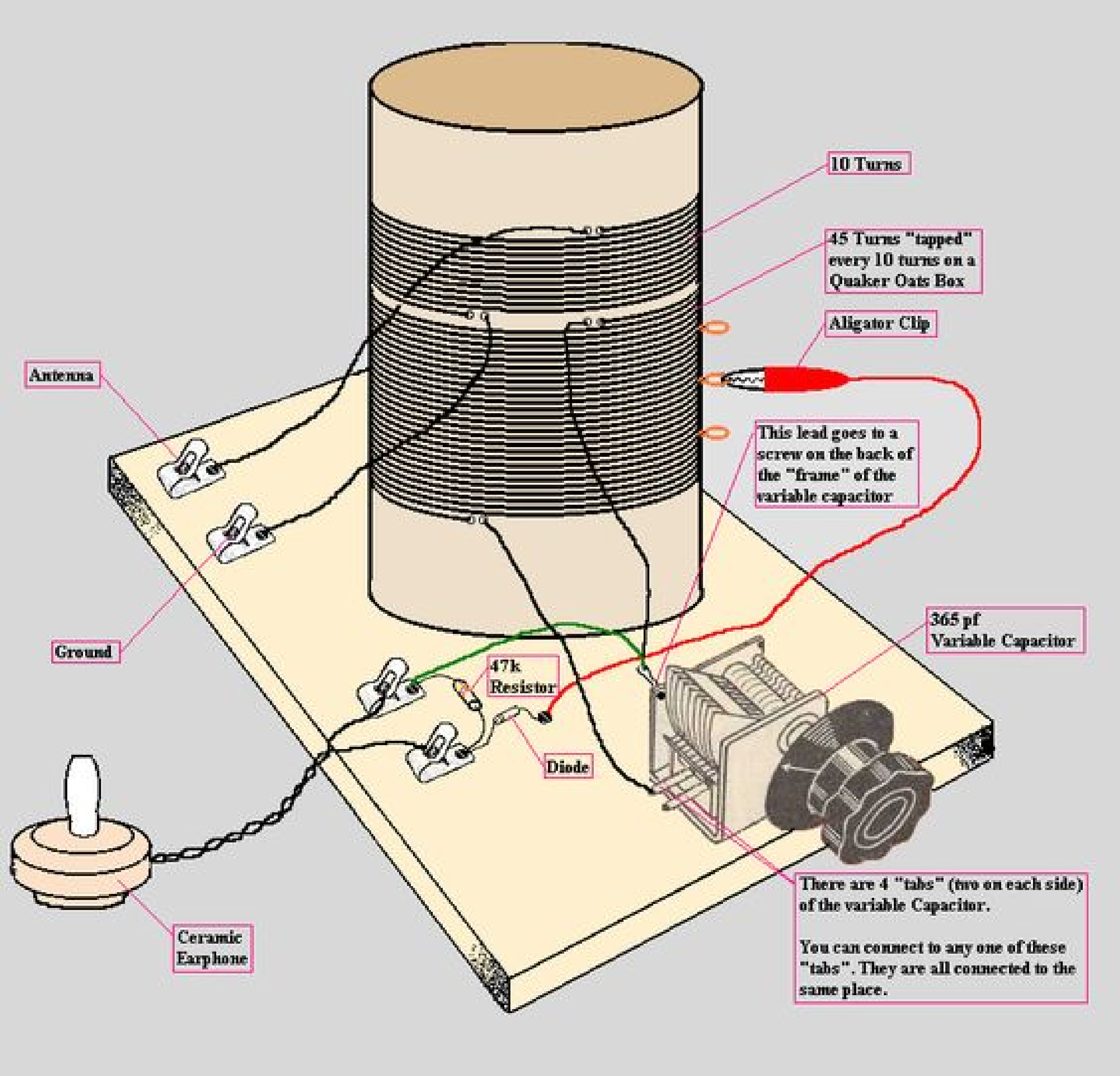
10 Turns
45 Turns "tapped" every 10 turns on a Quaker Oats Box
Aligator Clip
Antenna
This lead goes to a screw on the back of the "frame" of the variable capacitor
Ground
47k Resistor
365 pf Variable Capacitor
Diode
Ceramic Earphone
There are 4 "tabs" (two on each side) of the variable Capacitor.
You can connect to any one of these "tabs". They are all connected to the same place.

0
1
2
3
4
5
6
7
8
9

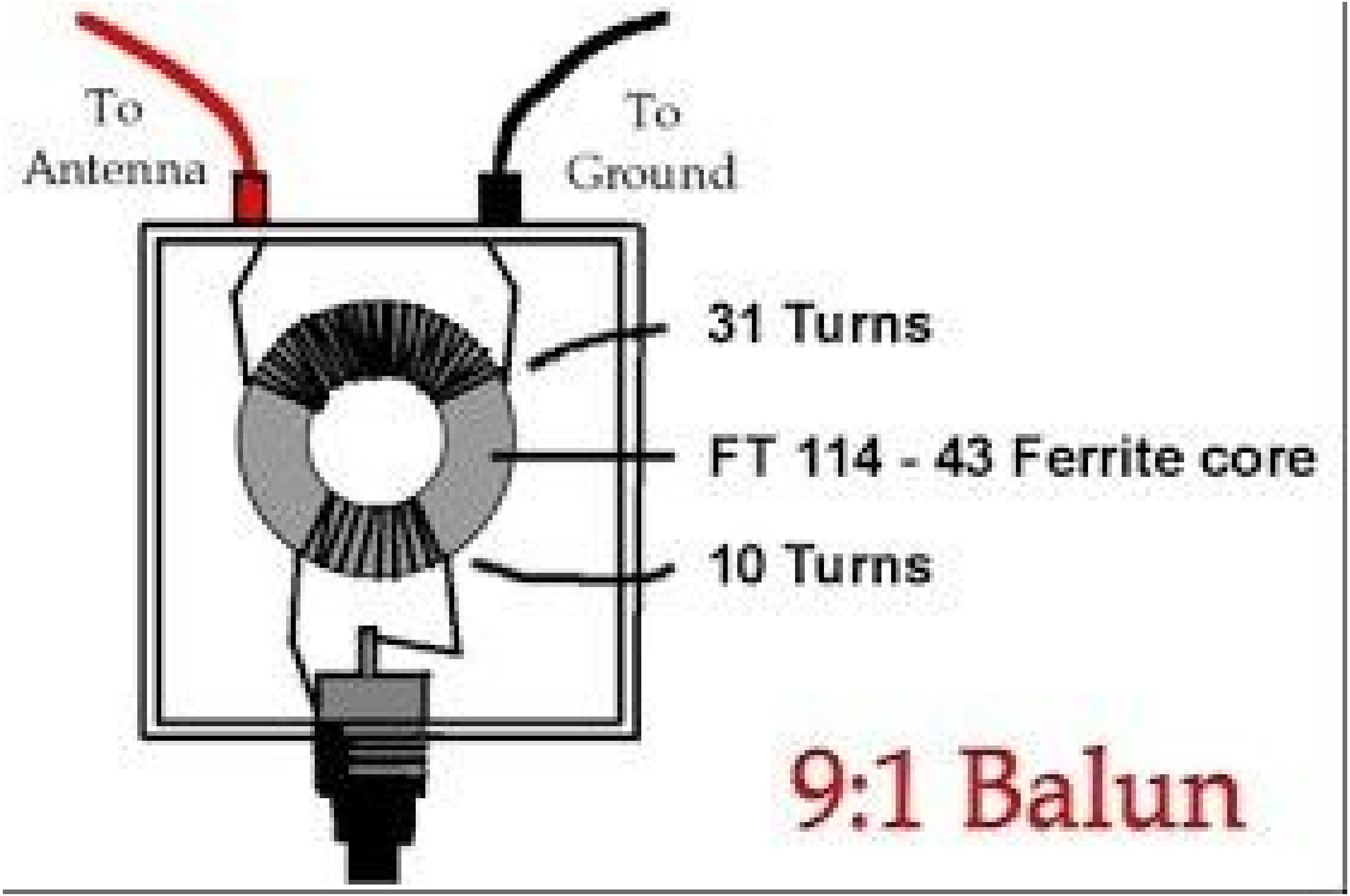
To Antenna
To Ground
31 Turns
FT 114 - 43 Ferrite core
10 Turns
9:1 Balun

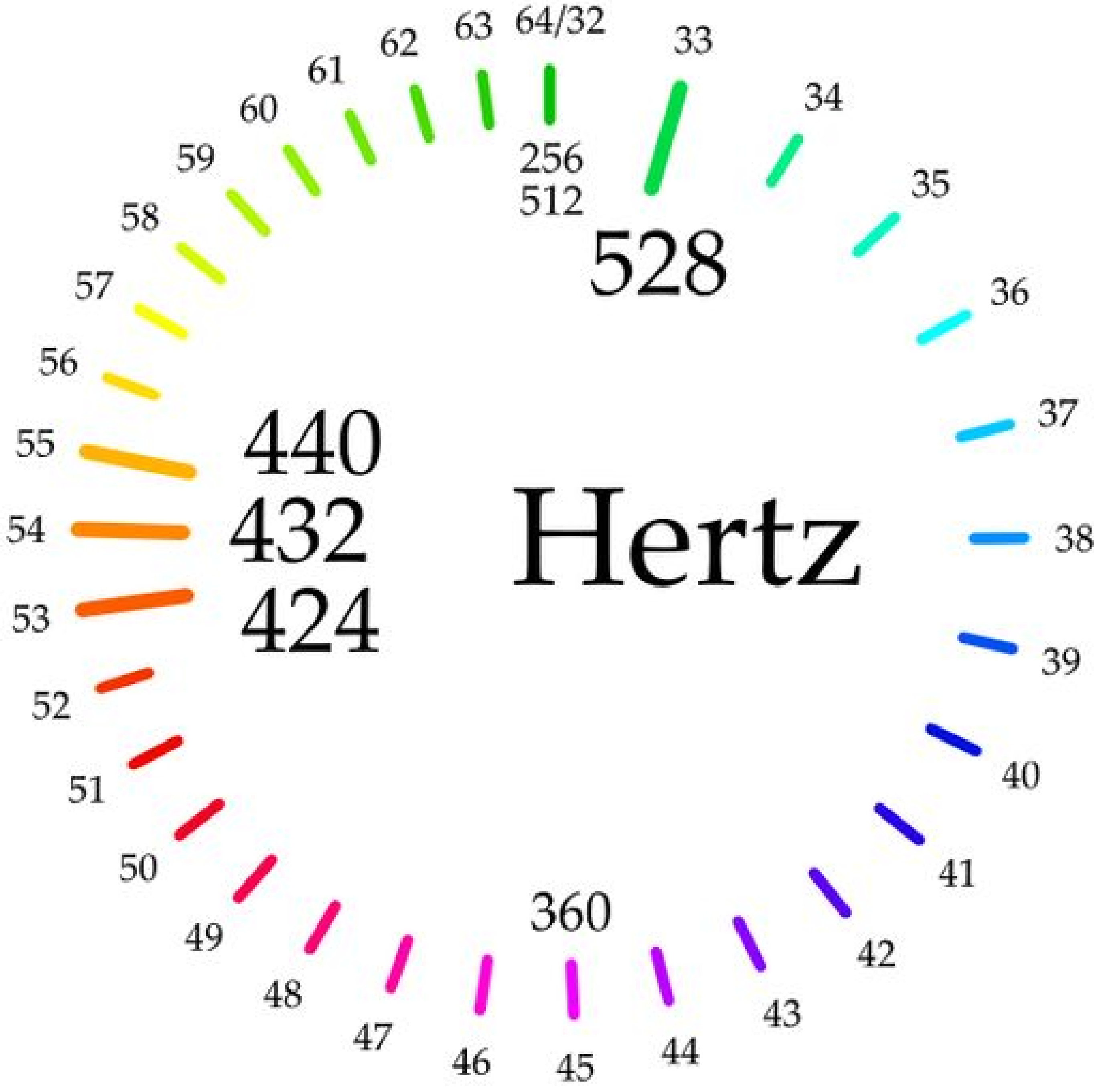
64/32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
256
512
528
440
432
424
Hertz
360

whatmusicreallyis.com

Seventh/Crown

Crown Chakra - located at the top of the head. Its function is understanding. Its inner state is: bliss. Its color is Violet and its planet is Uranus. Its stone is Amethyst. Its meditation is: I understand. Balancing this chakra is said to give vitality to the cerebrum and affects the development of psychic abilities. Energies: Air, Meditative, Intuition, Promotes thought.

Sixth/Third Eye

Third Eye Chakra - located in the center of the forehead above the eyebrows. Its function is seeing, intuiting. Its inner state is: I know. Its color is indigo and its planet is Jupiter. Its stone is: Iolite. Its meditation is: I see. Balancing this chakra helps psychic perception and balances the pineal gland. Energies: Air, Meditative, Intuition, Promotes thought.

Fifth/Throat

Throat Chakra - located in the throat Its function is communication, creativity. Its inner state is synthesis of ideas into symbols. Its color is bright blue and its planets are Mercury and Neptune. Its stones are: Sodalite, Blue Lace Agate, Lapis Lazuli. Meditation on: I speak. Balancing this chakra is important for the speech and communication areas of the brain. Energies: Water, Calming, Sooths, Relaxes.

Fourth/Heart

Heart Chakra - located in the center of the chest. Its function is: love, its inner state is compassion, love, its color is green, and its planet is Venus. Its stones are: Green/Pink Stones - Peridot, Rose Quartz, Malachite. Meditation on: I love. Balancing this chakra is important for the circulatory system, heart and thymus. It also affects spiritual love, compassion and universal oneness. Energies: Water, Calming, Soothes, Relaxes.

Third/Solar Plexus

Solar Plexus Chakra - located in the area above the navel area. Its function is Will, power, its inner state is laughter, joy, anger, its color is yellow, and its planets are Mars and the Sun. Its stones are: Amber, Topaz and Citrine. Meditation on: I do. Balancing this chakra is associated with calming emotions and frustration, easing tension and helping to better utilize intuition. Energies: Fire, Energizing, Charging, Lends Energy.

Second/Sacral

Navel/Sacral Chakra - located in the lower abdomen, genitals, womb, its function is desire, sexuality, pleasure, procreation. Its inner state is tears, its color is orange and its celestial body is the Moon. Its stones are: Coral and Carnelian. Meditation on: I feel. Balancing this chakra is associated with sexual vitality, physical power and fertility. Energies: Fire, Energizing, Charging.

First/Root Chakra

Root/support chakra - located at the base of the spine. Its function is survival and grounding, its inner state is stillness and stability, its color is red and its planets are Earth and Saturn. Its stones are: Garnet, Ruby, Onyx, Obsydian. Meditation on: I am. Balancing this chakra gives energy to the physical body, controls fear, increases overall health and helps in grounding. Energies: Earth, Grounding, Focusing, Centering.

The word chakra is Sanskrit for wheel or disk and signifies one of seven basic energy centers in the body. Each of these centers correlates to major nerve ganglia branching forth from the spinal column.

Created by: Izolda Trakhtenberg
Source: Wheels of Life by Anodea Judith

DIODE

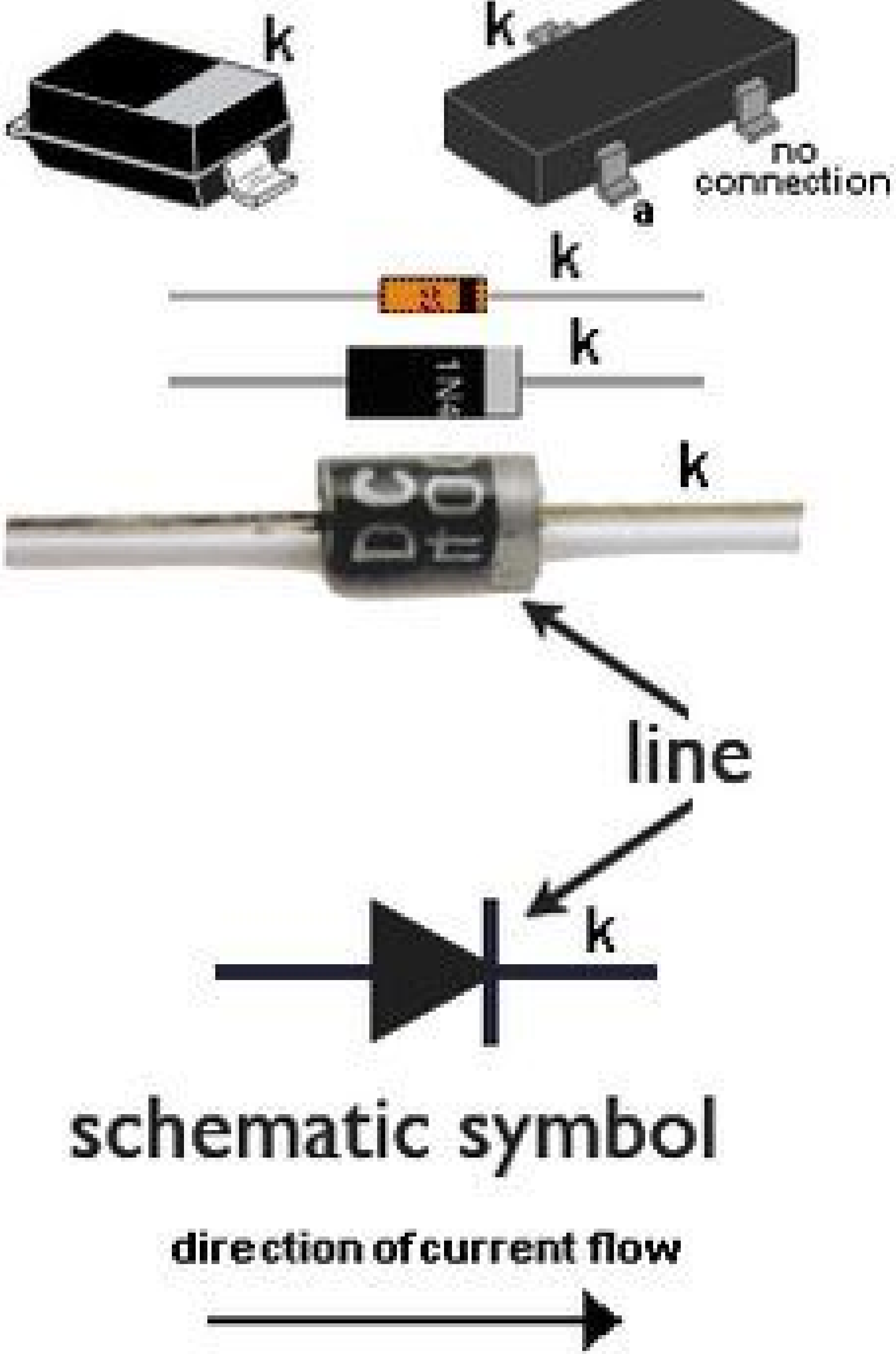
k
k
no connection
a
k
k
k
line
k
schematic symbol
direction of current flow

Remember:
104 = 100n
105 = 1u
106 = 10u
107 - 100u

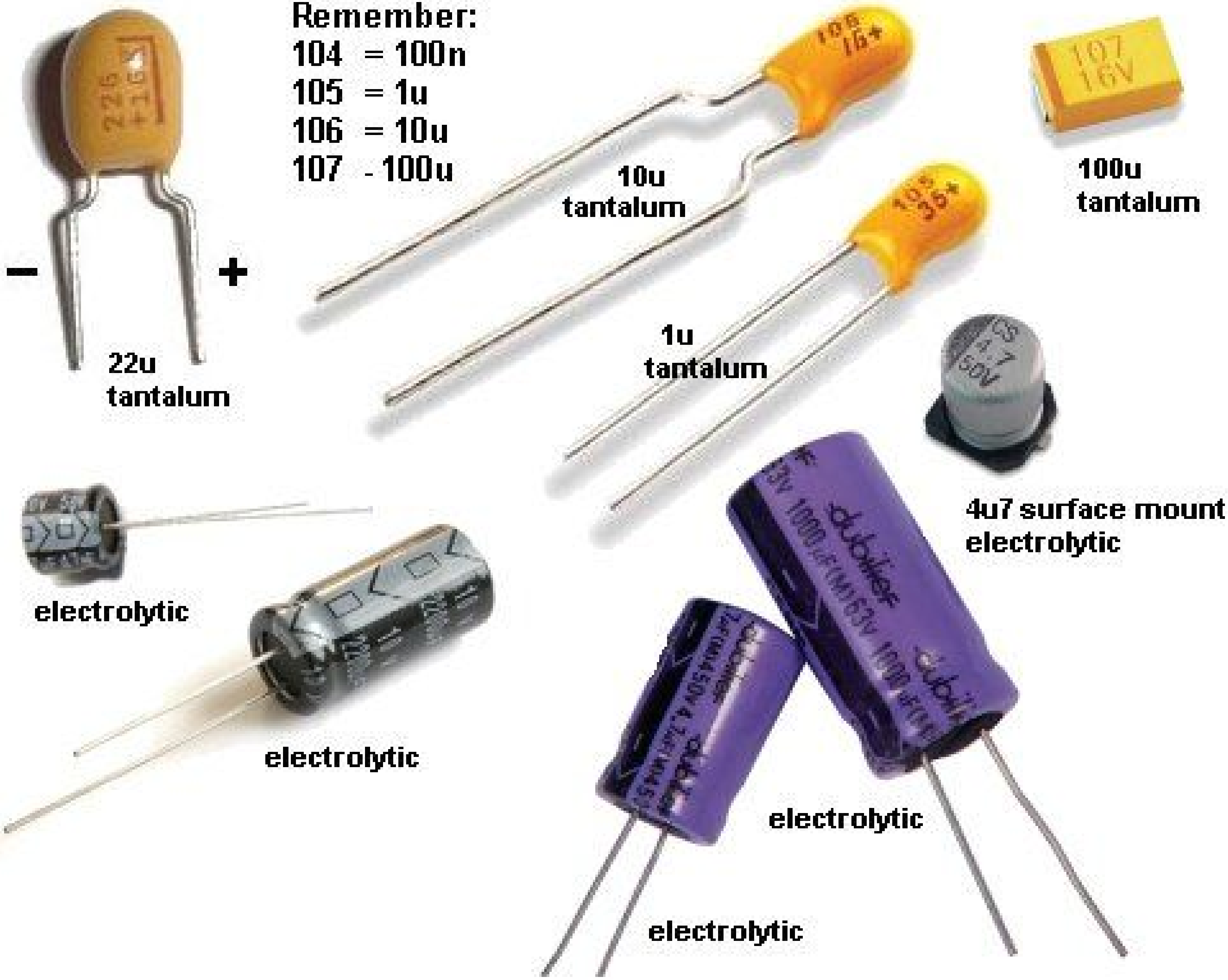
226
16
107
16V
106
16+
105
35+
CS
4.7
50V
dubilier
1000

−
+

22u
tantalum

10u
tantalum

1u
tantalum

100u
tantalum

4u7 surface mount
electrolytic

electrolytic

electrolytic

electrolytic

electrolytic

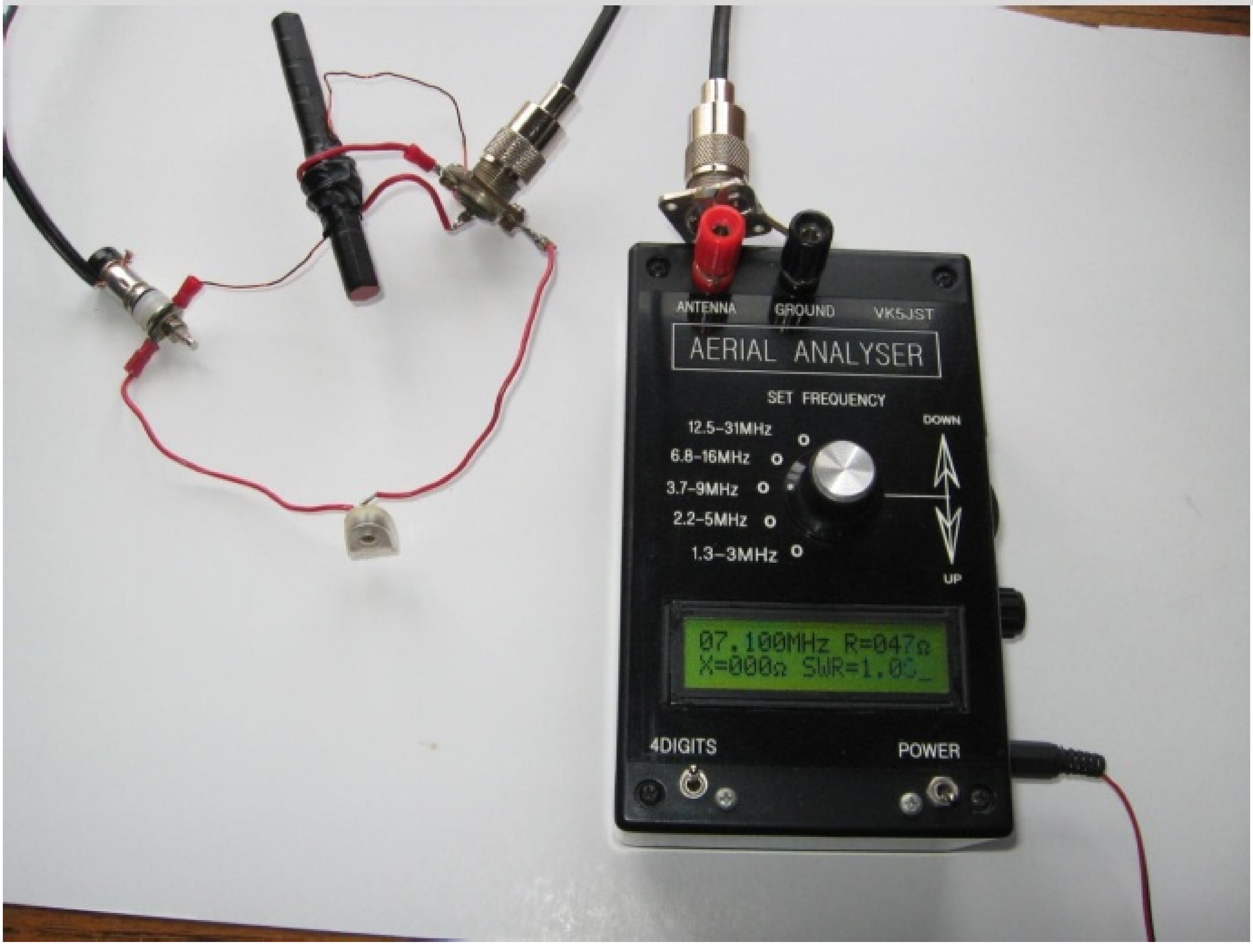
ANTENNA
GROUND
VK5JST
AERIAL ANALYSER
SET FREQUENCY
12.5-31MHz
6.8-16MHz
3.7-9MHz
2.2-5MHz
1.3-3MHz
DOWN
UP
07.100MHz R=047Ω
X=000Ω SWR=1.05_
4DIGITS
POWER

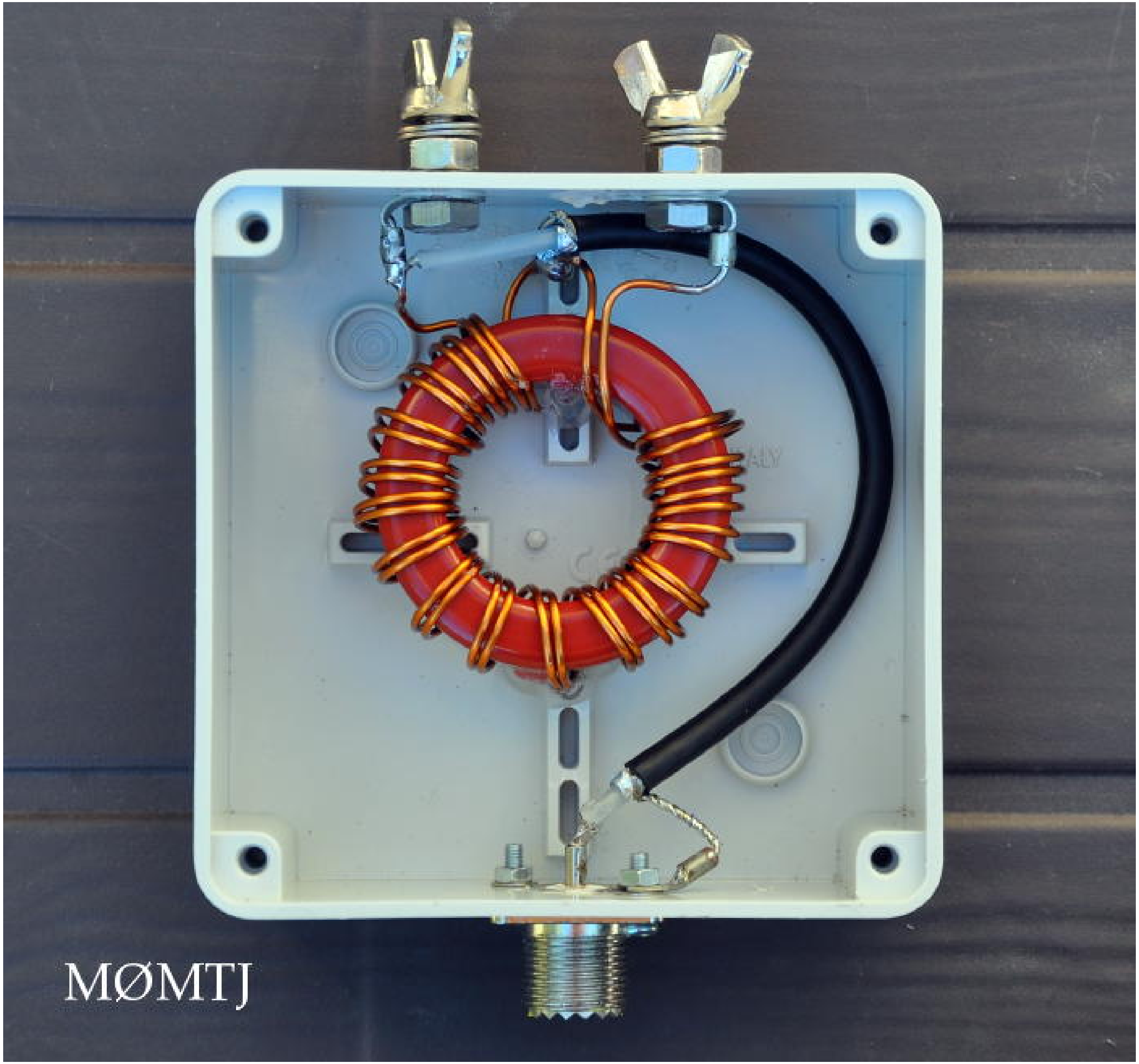

M0MTJ

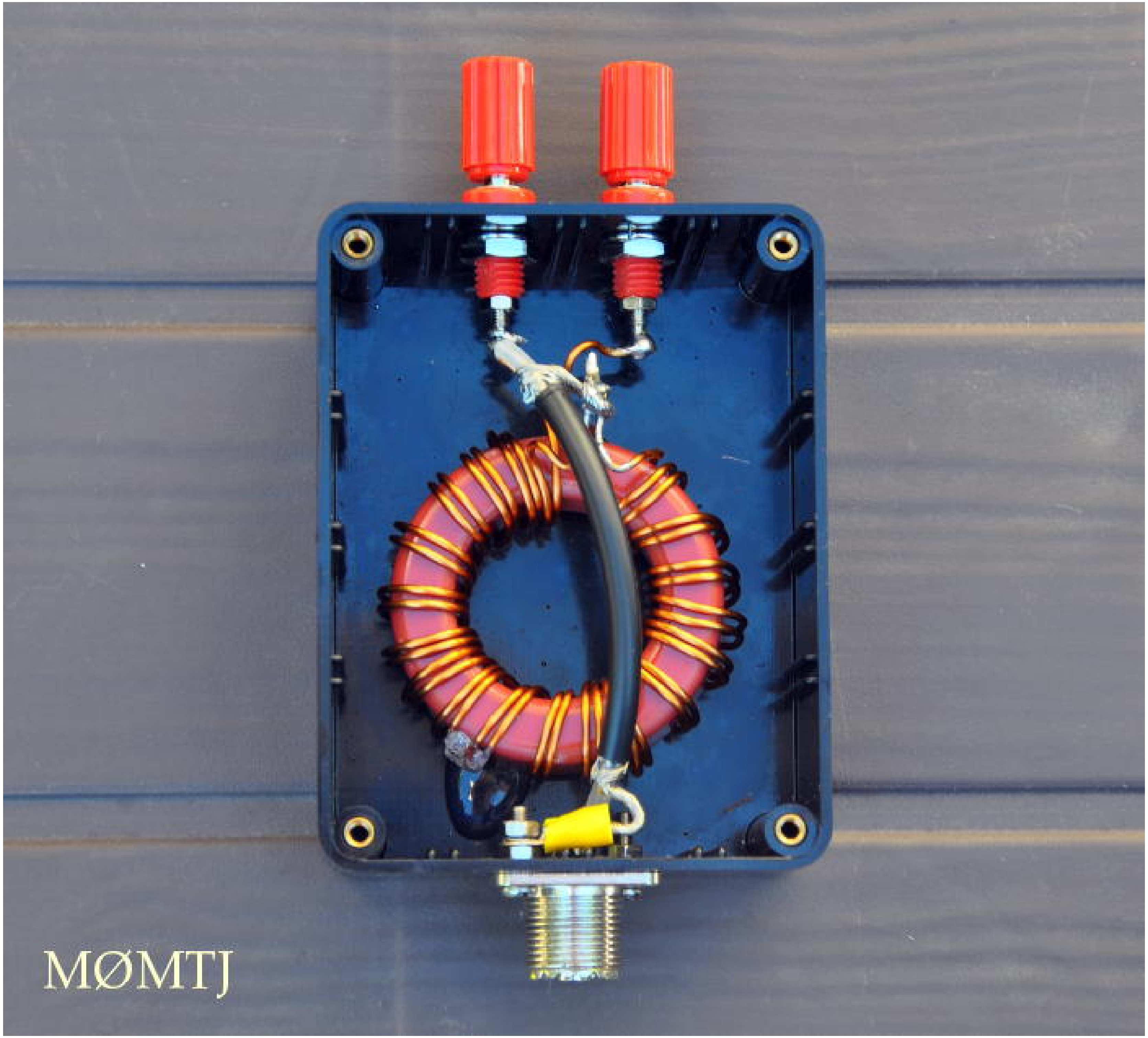
M0MTJ

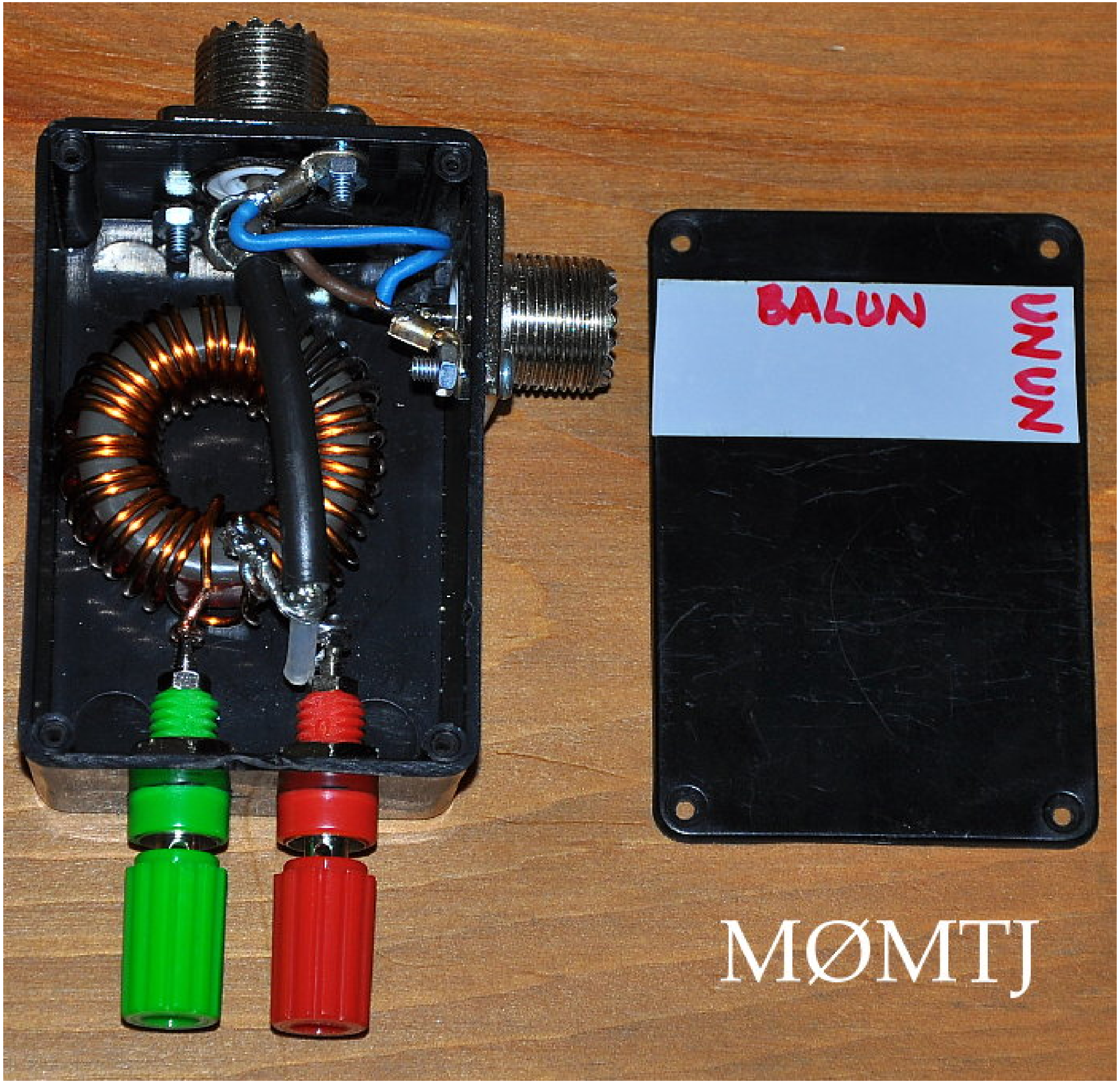
BALUN
UNUN
M0MTJ

FIG. 3 AMP Power Envelope

Volts vs. Resistance

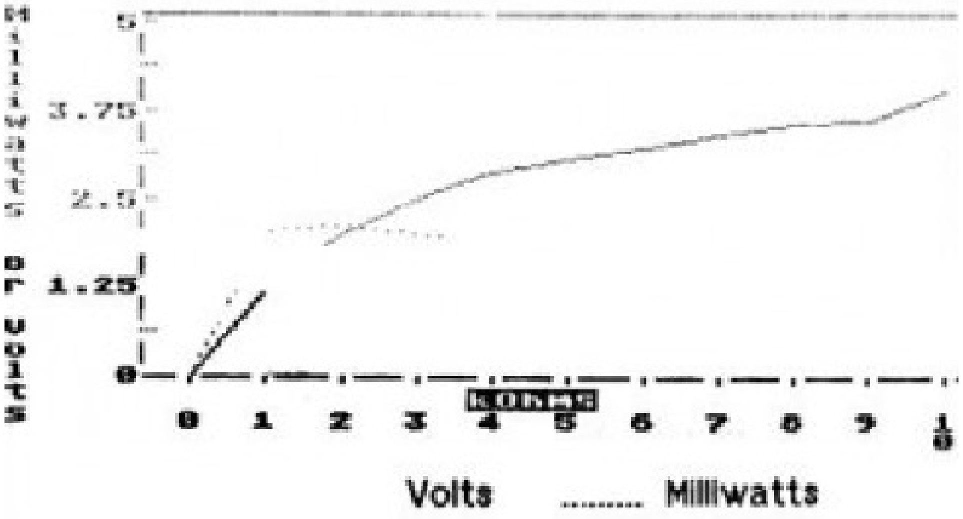

FIG. 4 Clock powered by APM

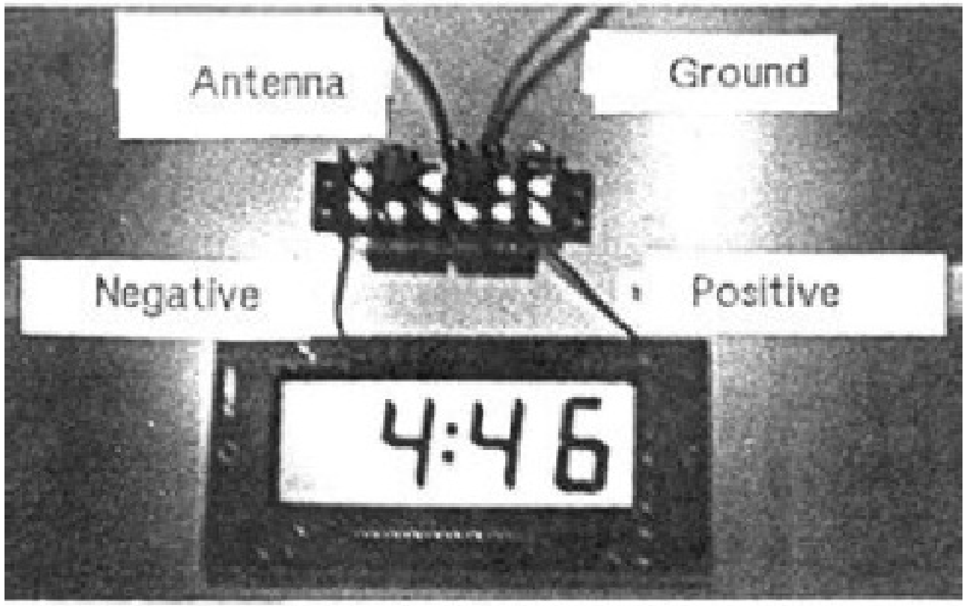
Antenna
Ground
Negative
Positive
4:46

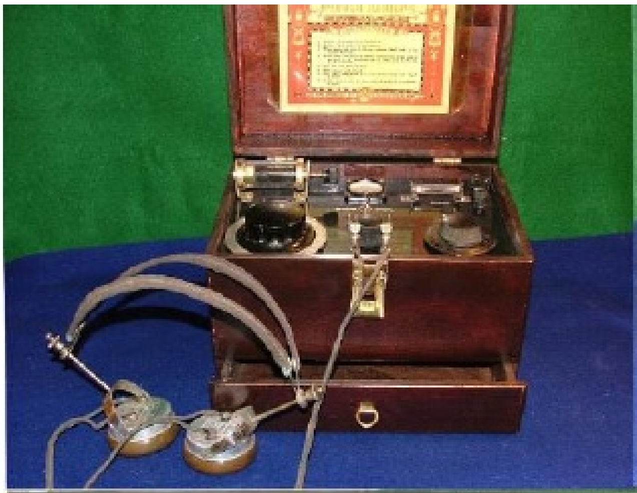

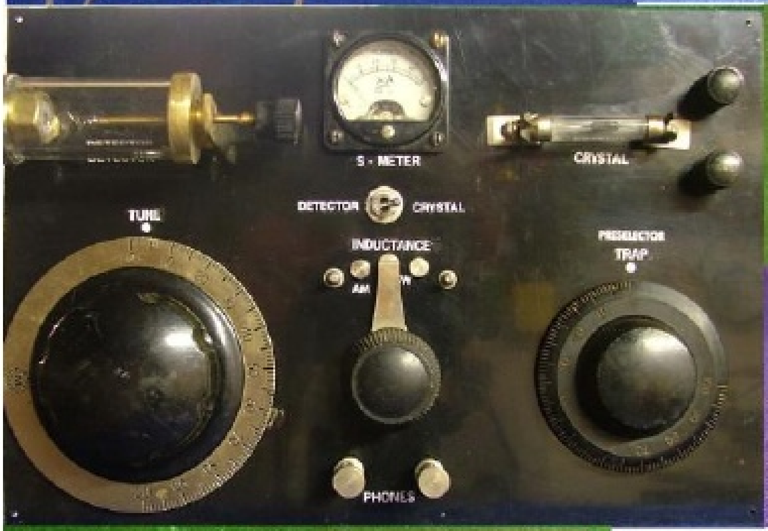
DETECTOR
S - METER
CRYSTAL
DETECTOR
CRYSTAL
TUNE
INDUCTANCE
AM
PRESELECTOR
TRAP
PHONES

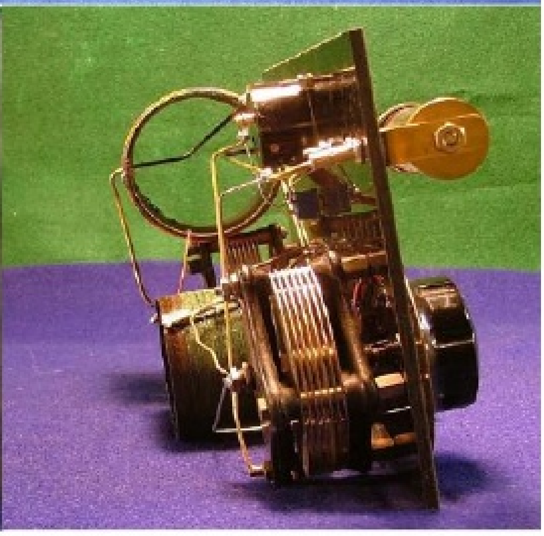

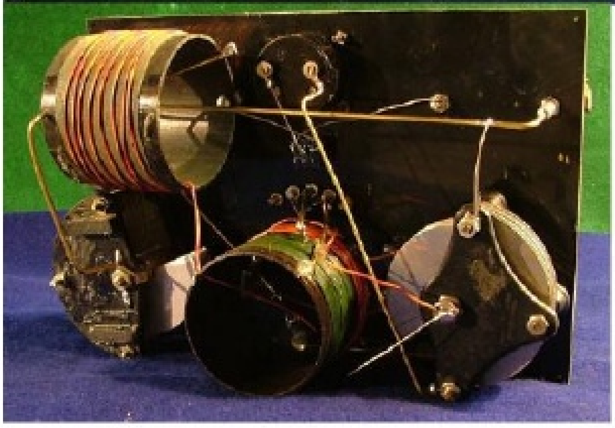

"The Ultimate Crystal Radio"

Very nice, 2 stage radio.
Tuning, Wavetrap and
Inductance controls.
Fixed and Cat Whisker
Detectors and S-Meter.
Vintage 80+ year old
finger joint box.

Headphones Included

$459.00 plus Shipping

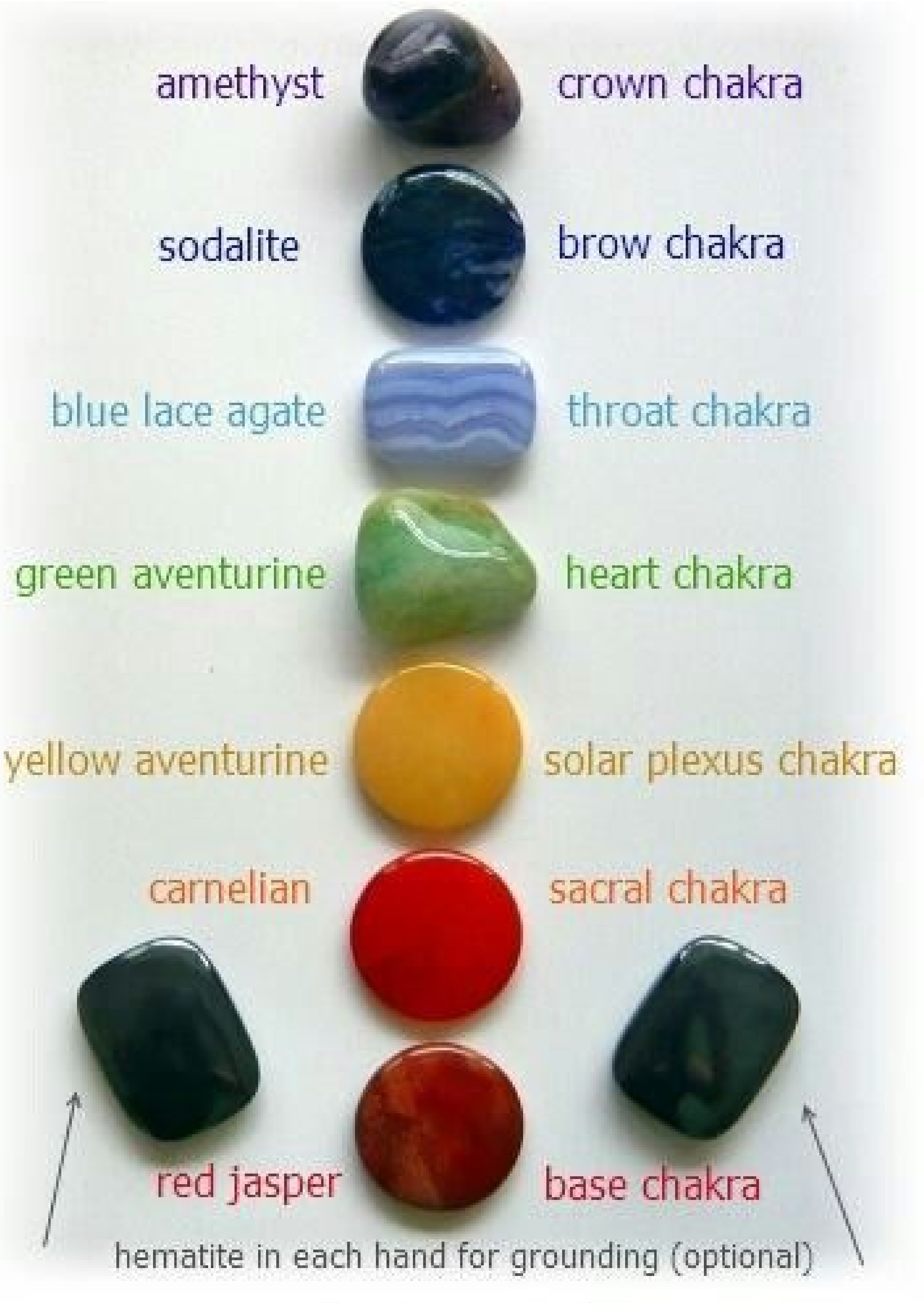

amethyst
crown chakra
sodalite
brow chakra
blue lace agate
throat chakra
green aventurine
heart chakra
yellow aventurine
solar plexus chakra
carnelian
sacral chakra
red jasper
base chakra
hematite in each hand for grounding (optional)

a basic crystal
chakra balancing layout

© henny designs

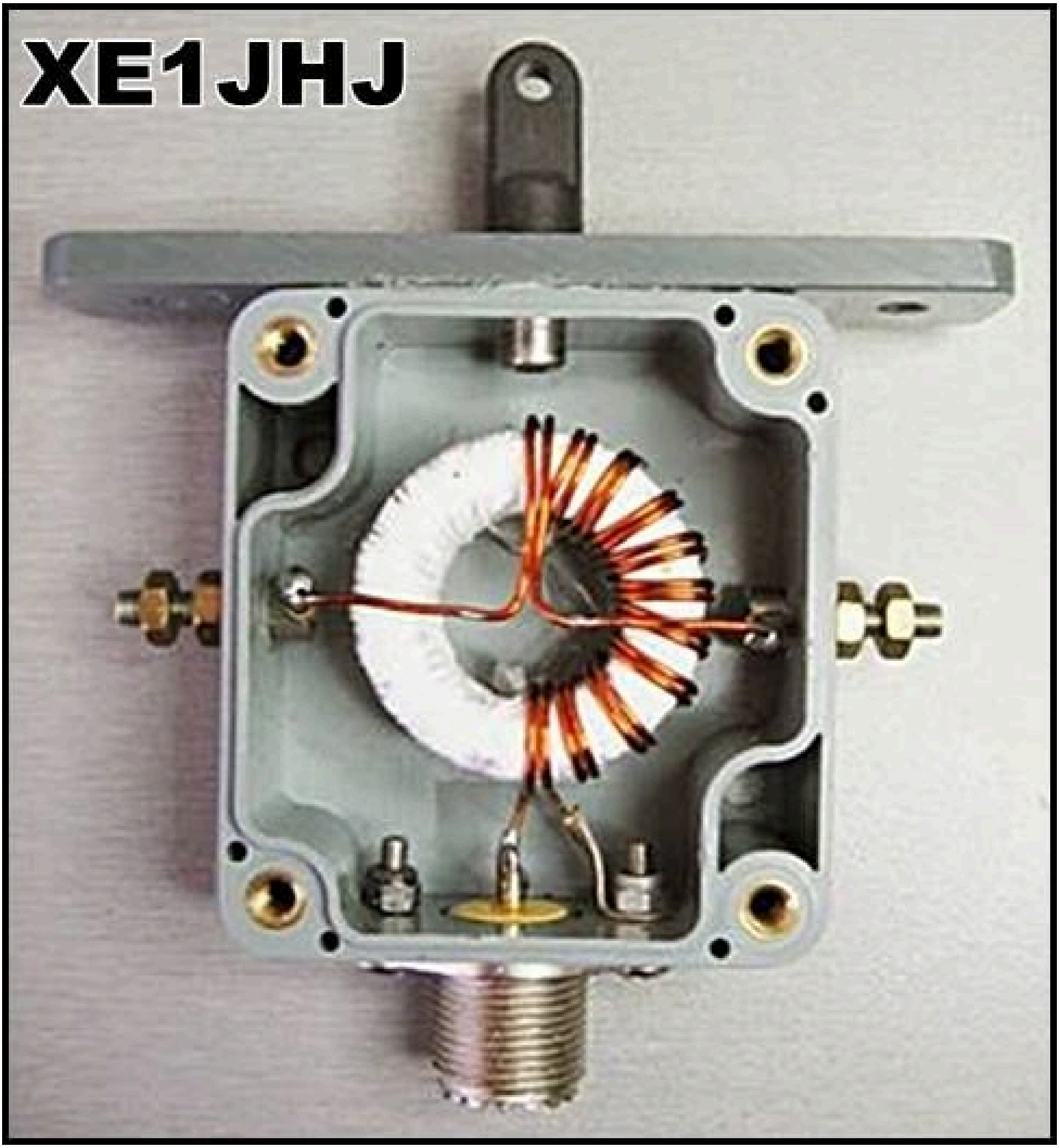
XE1JHJ

How to hang a BALUN without damaging BALUN Antenna terminals.

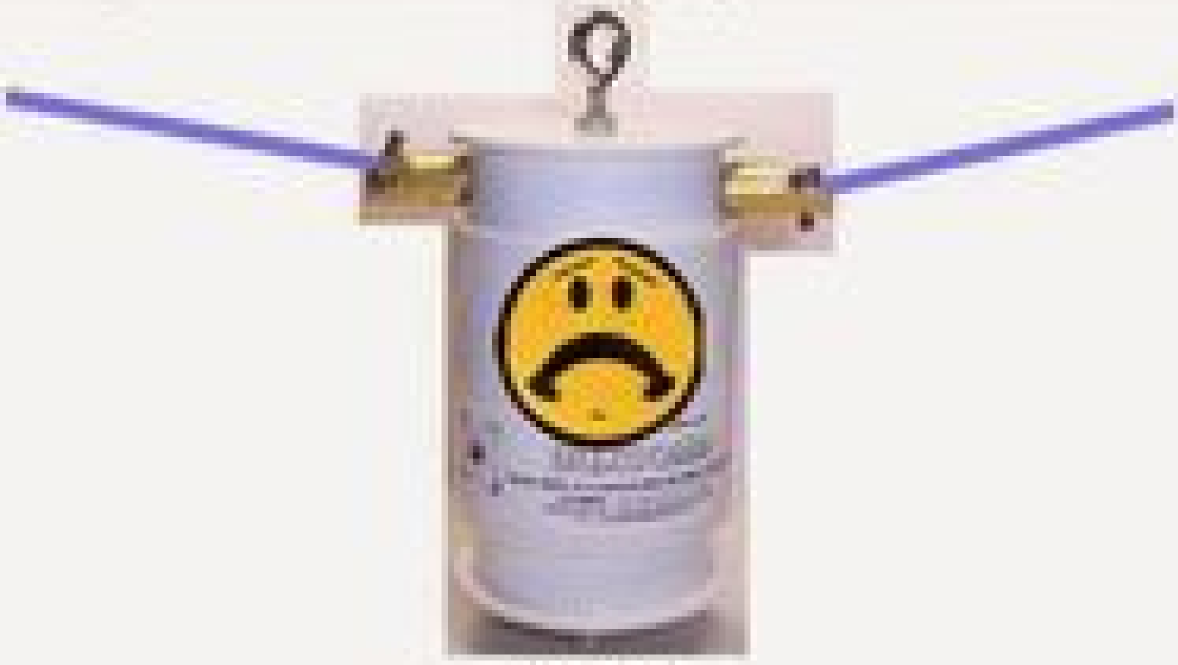

"WRONG"

Do not support BALUN by Antenna Wire Elements,

Use Eyebolt to support BALUN from Center Insulator.

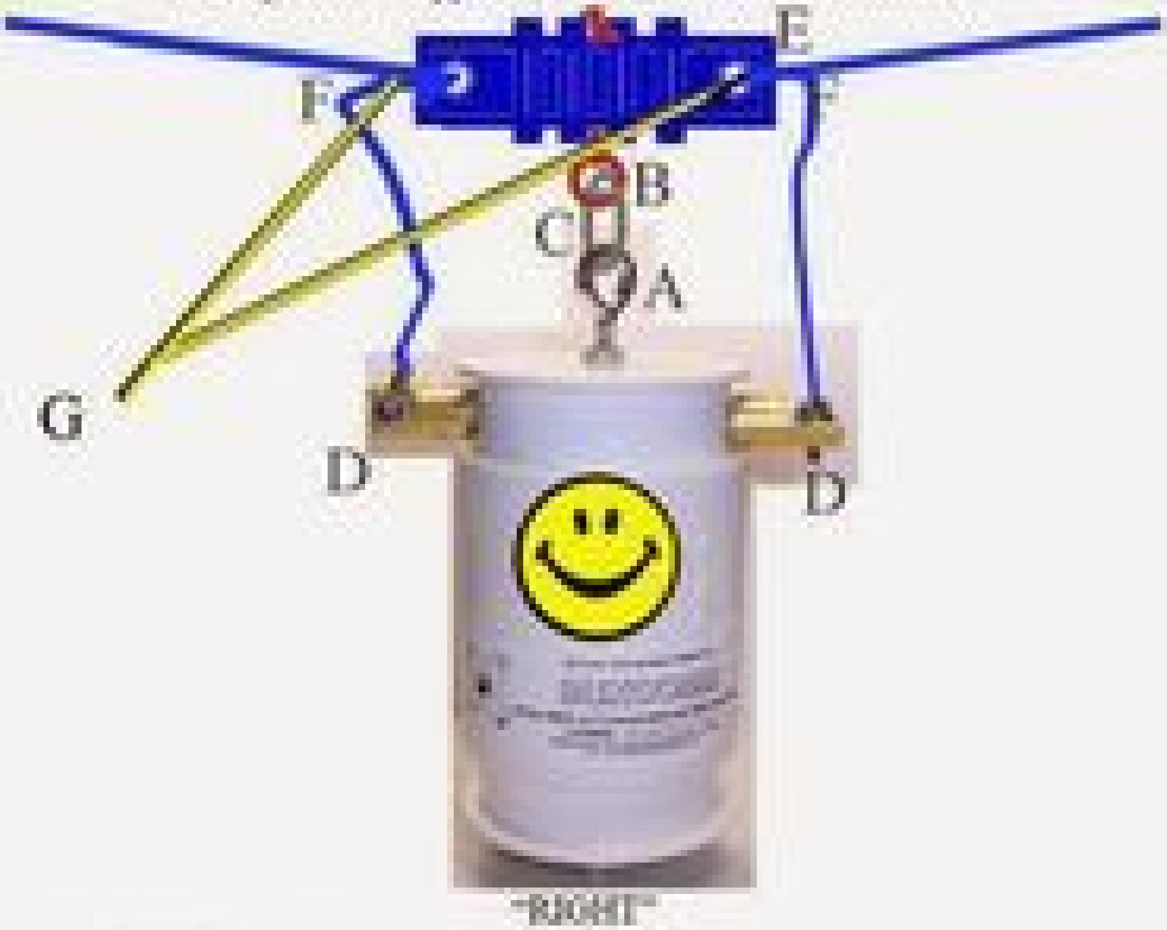

"RIGHT"

A= BALUN Support Eyebolt,
B= Center Insulator Eyebolt Support
C= Support Link between Center Insulator and BALUN Eyebolt
D= Antenna Terminals BALUN Connectors
E= Center Insulator
F/F=Loop antenna wire element(s) through insulator holes .
G= Fold back wire element onto itself, secure with Ty-Wrap and cover with tape.

Visit: www.HamRadioExpress.com

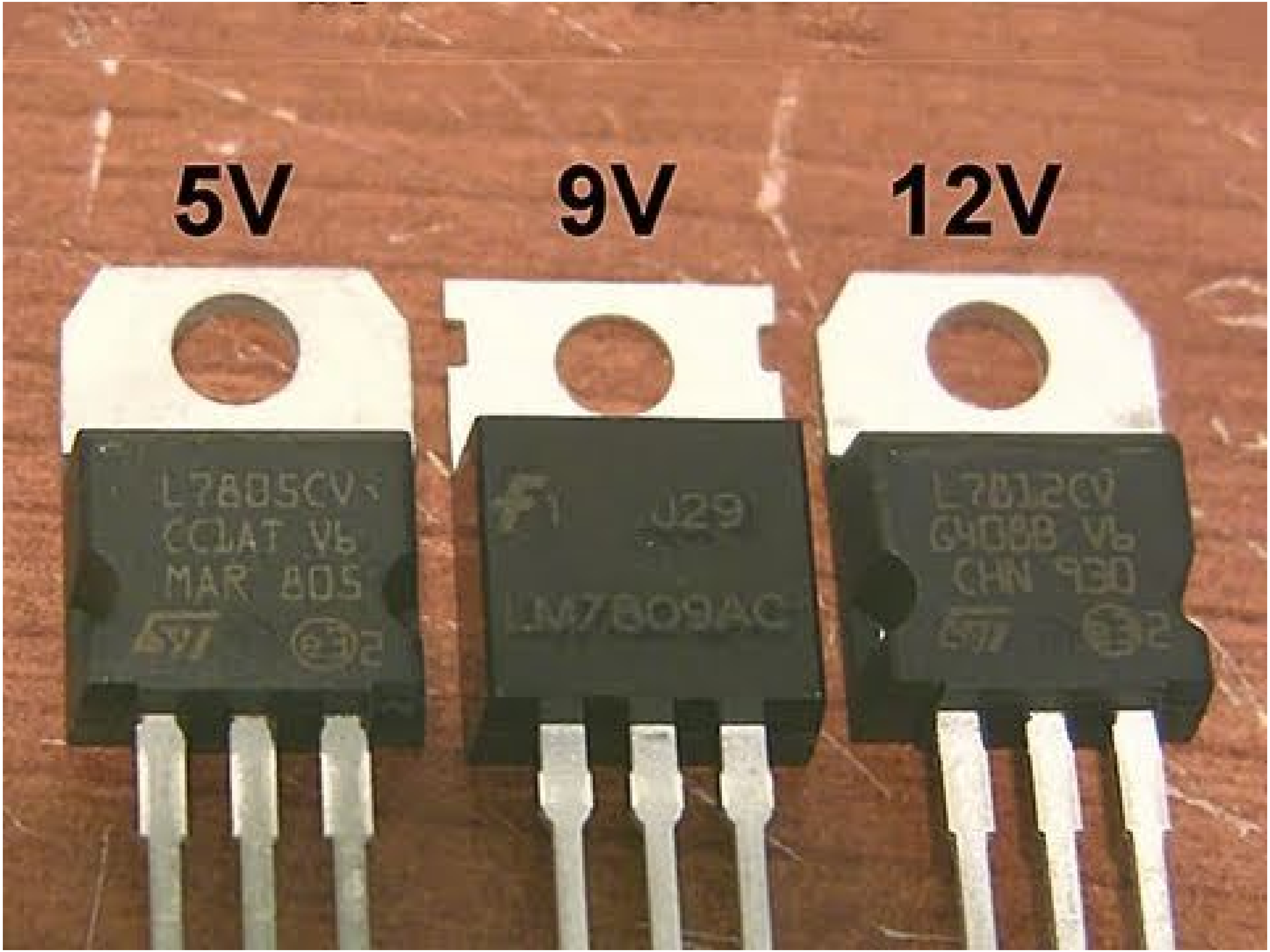
5V
9V
12V
L7805CV
CC1AT V6
MAR 805
J29
LM7809AC
L7812CV
G408B V6
CHN 930

Common US Spiders

Deadly & Dangerous

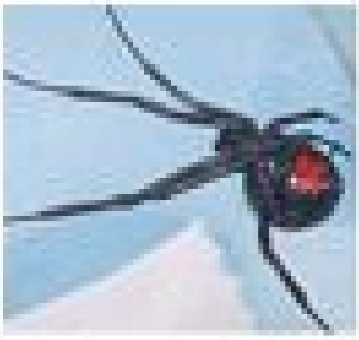

Black Widow Spider

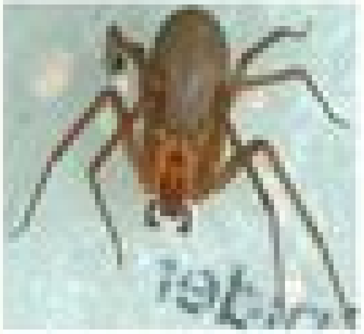

Brown Recluse Spider

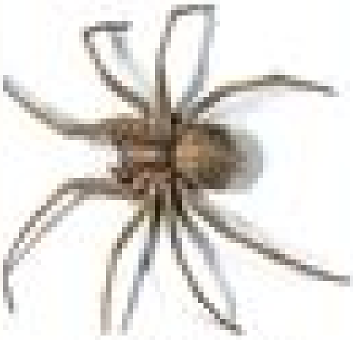

Hobo Spider

Toxic with Painful Bite

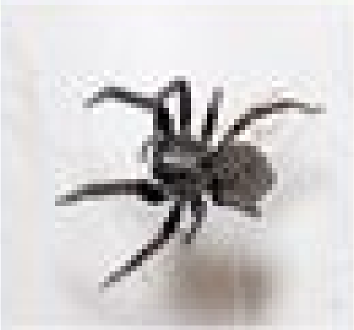

Black House Spider

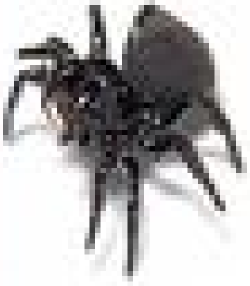

Female Mouse Spider

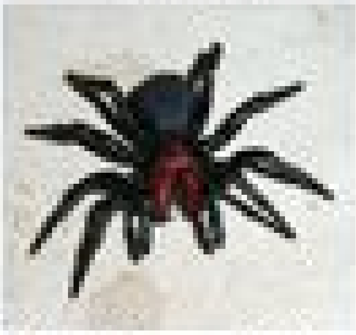

Male Mouse Spider

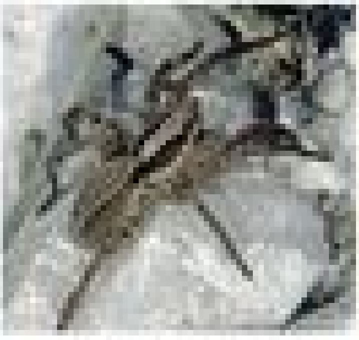

Wolf Spider

Low Risk

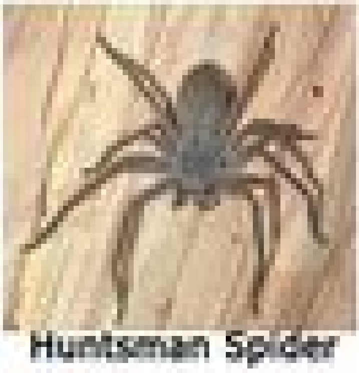

Huntsman Spider

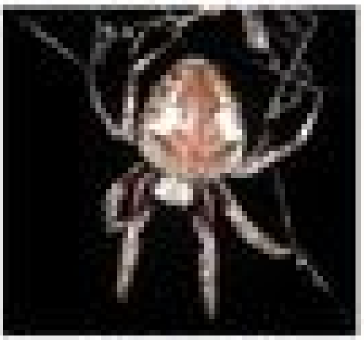

Garden Orb-Weaving Spider

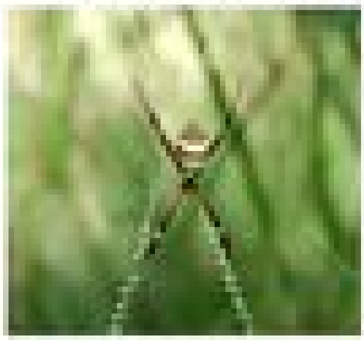

St Andrew's Cross Spider

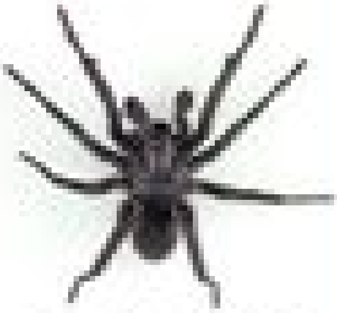

Trapdoor Spider

Call 641-427-1677 Today!

Copyright 2011, BestColumbusPestControl.com

Grog's Index of Boating Knots

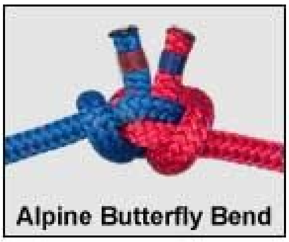
Alpine Butterfly Bend

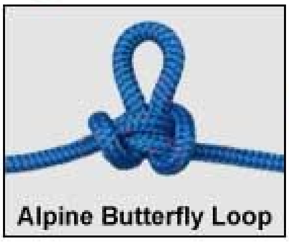
Alpine Butterfly Loop

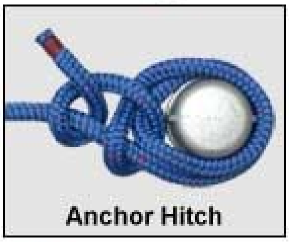
Anchor Hitch

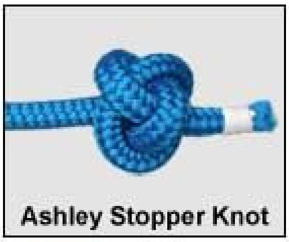
Ashley Stopper Knot

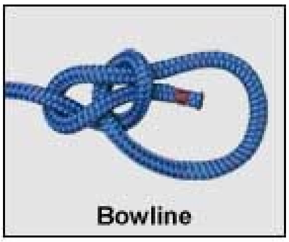
Bowline

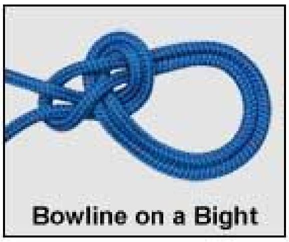
Bowline on a Bight

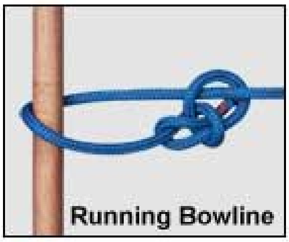
Running Bowline

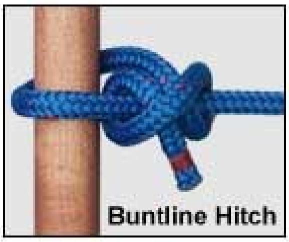
Buntline Hitch

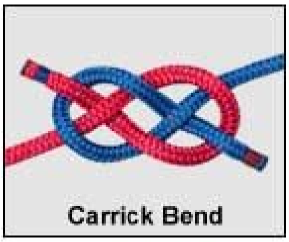
Carrick Bend

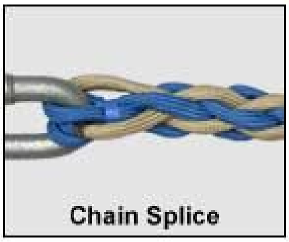
Chain Splice

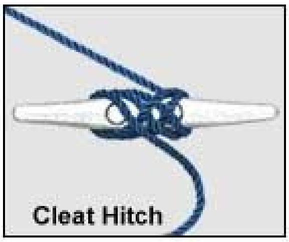
Cleat Hitch

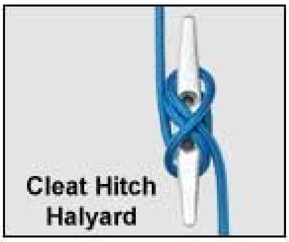
Cleat Hitch
Halyard

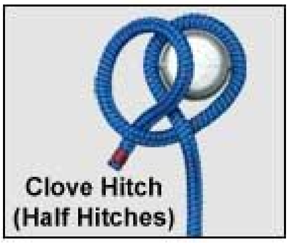
Clove Hitch
(Half Hitches)

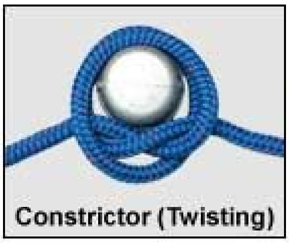
Constrictor (Twisting)

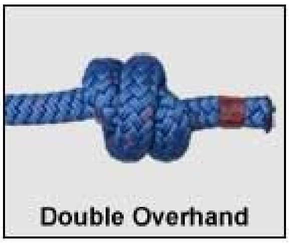
Double Overhand

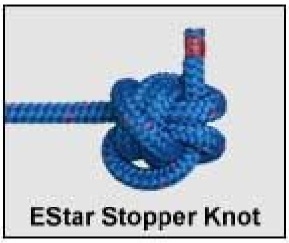
EStar Stopper Knot

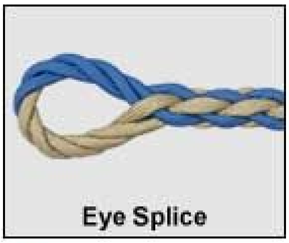
Eye Splice

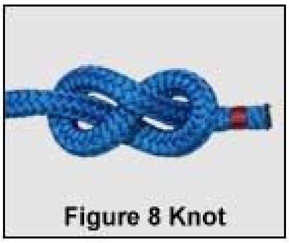
Figure 8 Knot

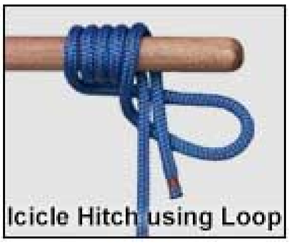
Icicle Hitch using Loop

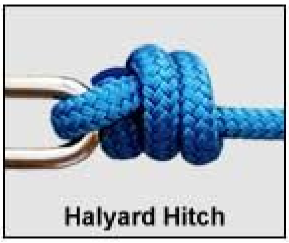
Halyard Hitch

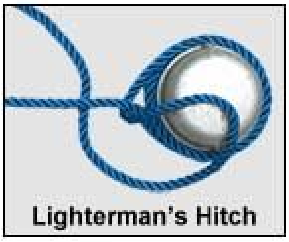
Lighterman's Hitch

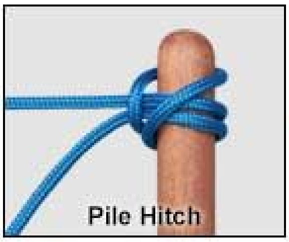
Pile Hitch

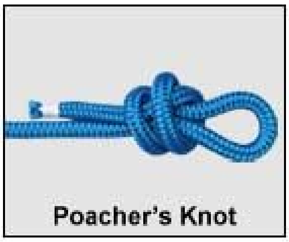
Poacher's Knot

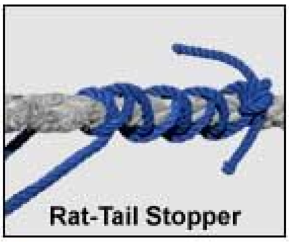
Rat-Tail Stopper

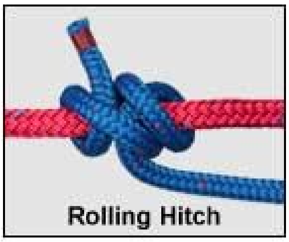
Rolling Hitch

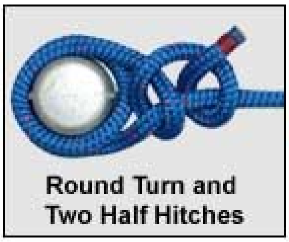
Round Turn and
Two Half Hitches

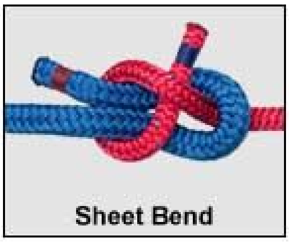
Sheet Bend

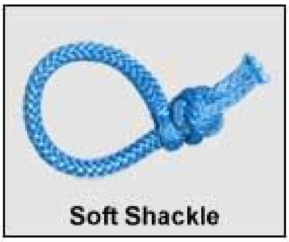
Soft Shackle

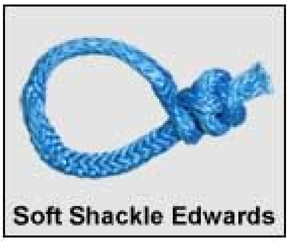
Soft Shackle Edwards

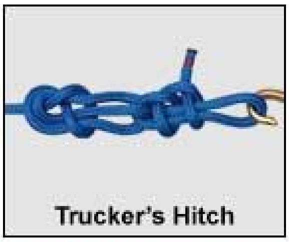
Trucker's Hitch

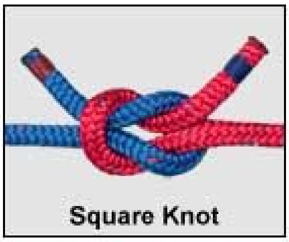
Square Knot

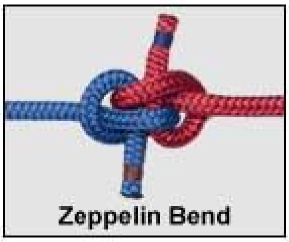
Zeppelin Bend

www.animatedknots.com

Grog's Index of Scouting Knots

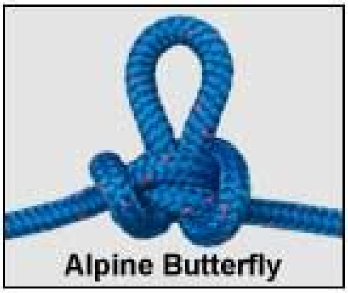
Alpine Butterfly

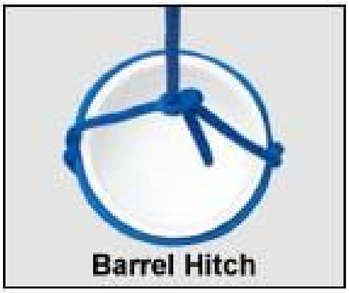
Barrel Hitch

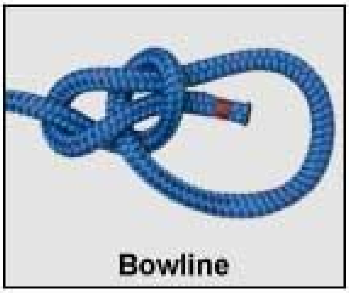
Bowline

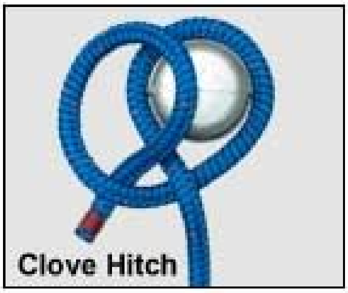
Clove Hitch

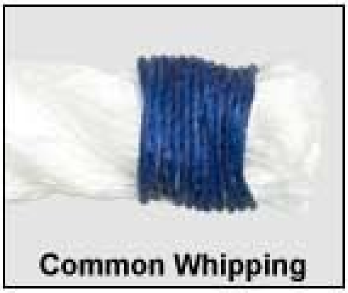
Common Whipping

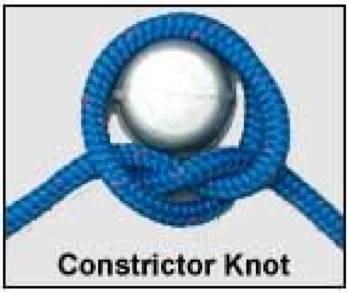
Constrictor Knot

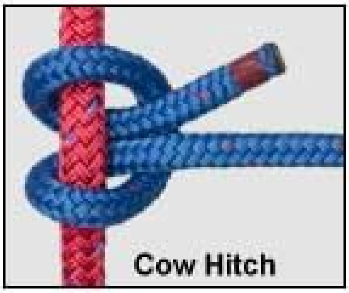
Cow Hitch

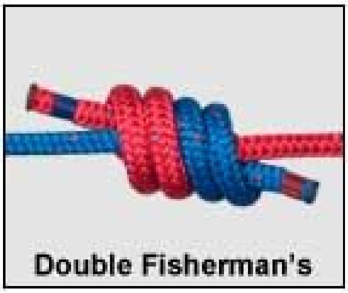
Double Fisherman's

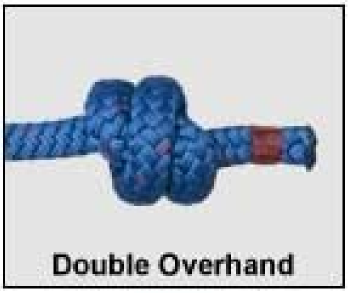
Double Overhand

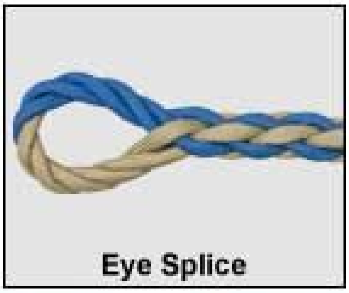
Eye Splice

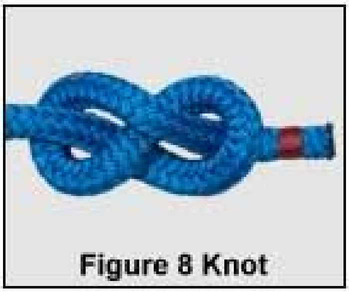
Figure 8 Knot

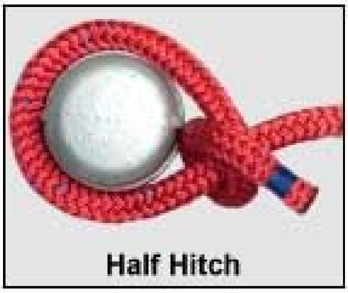
Half Hitch

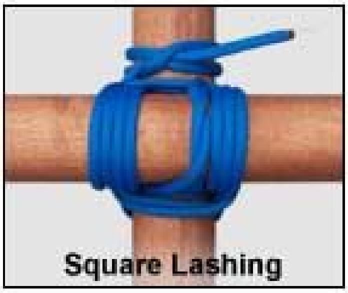
Square Lashing

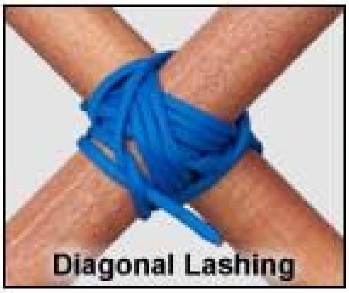
Diagonal Lashing

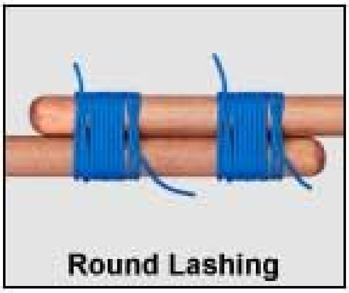
Round Lashing

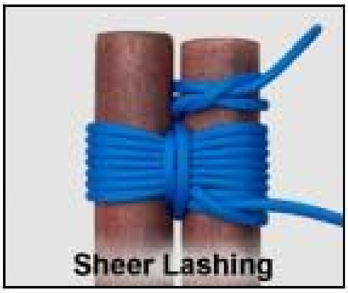
Sheer Lashing

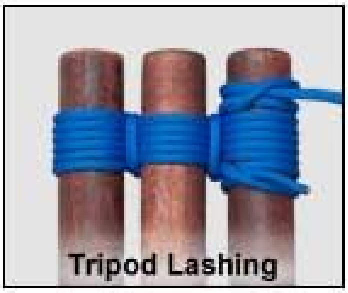
Tripod Lashing

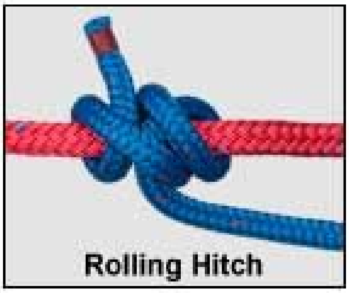
Rolling Hitch

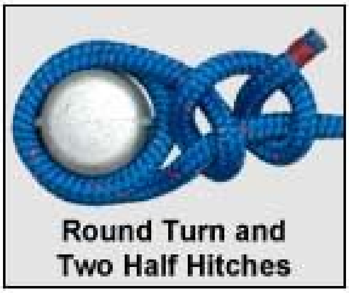
Round Turn and
Two Half Hitches

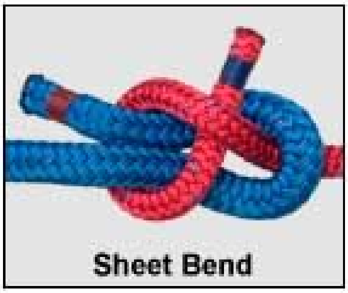
Sheet Bend

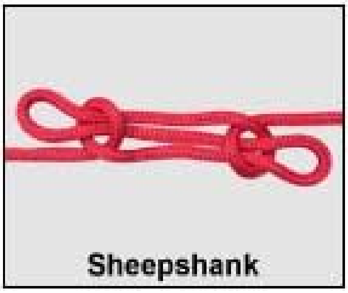
Sheepshank

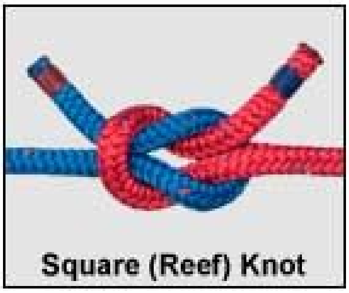
Square (Reef) Knot

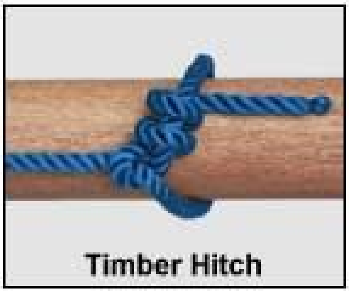
Timber Hitch

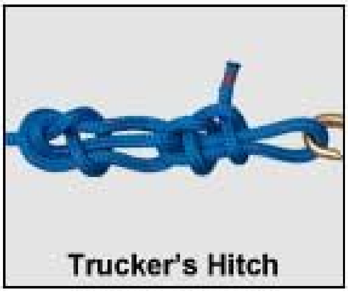
Trucker's Hitch

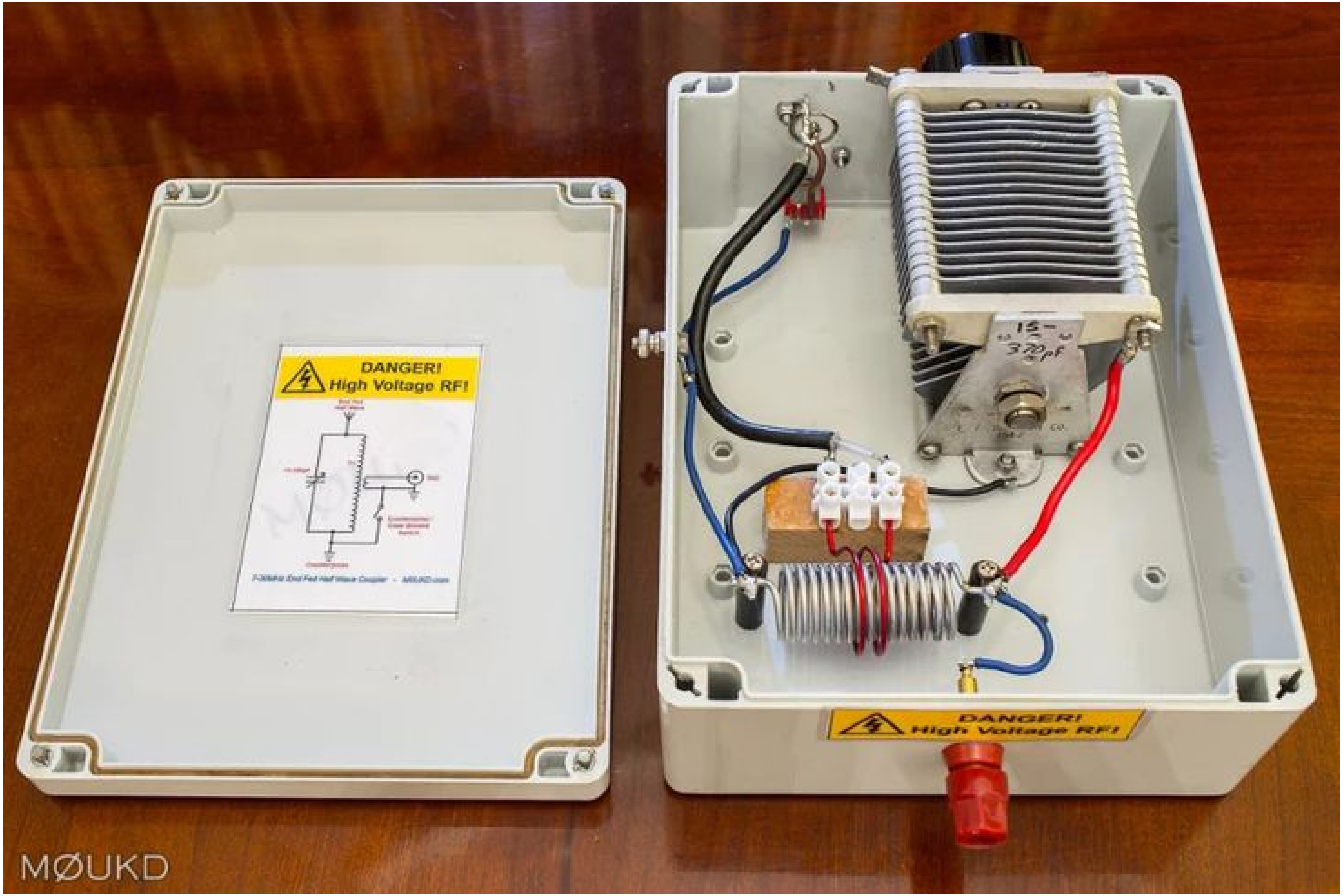
DANGER!
High Voltage RF!
7-30MHz End Fed Half Wave Coupler - M0UKD.com
15-370pF
DANGER!
High Voltage RF!
MØUKD

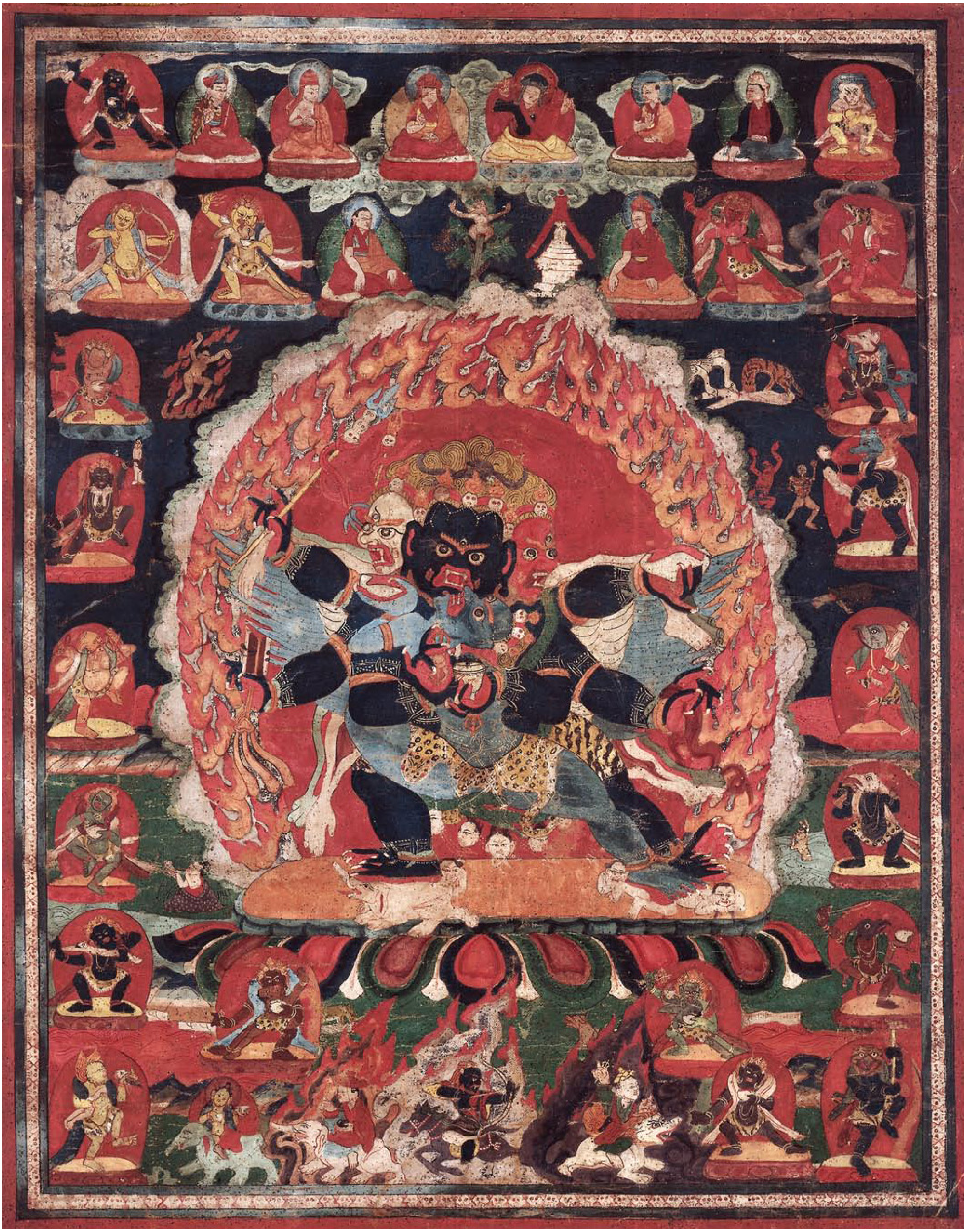

Voltage Balun / "Magnetic Balun"

From GW0VMR's plan

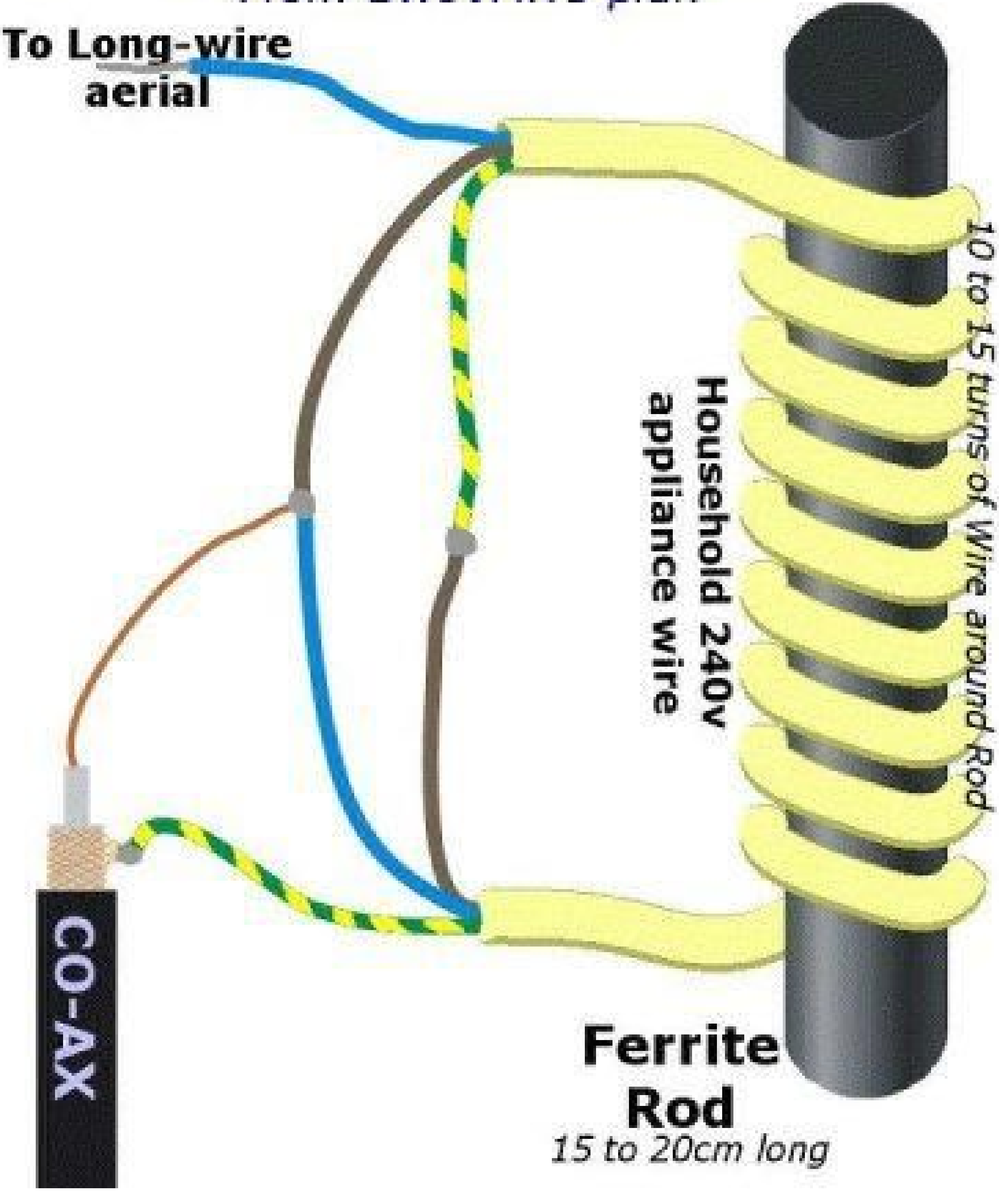
To Long-wire aerial
10 to 15 turns of Wire around Rod
Household 240v appliance wire
CO-AX
Ferrite Rod
15 to 20cm long

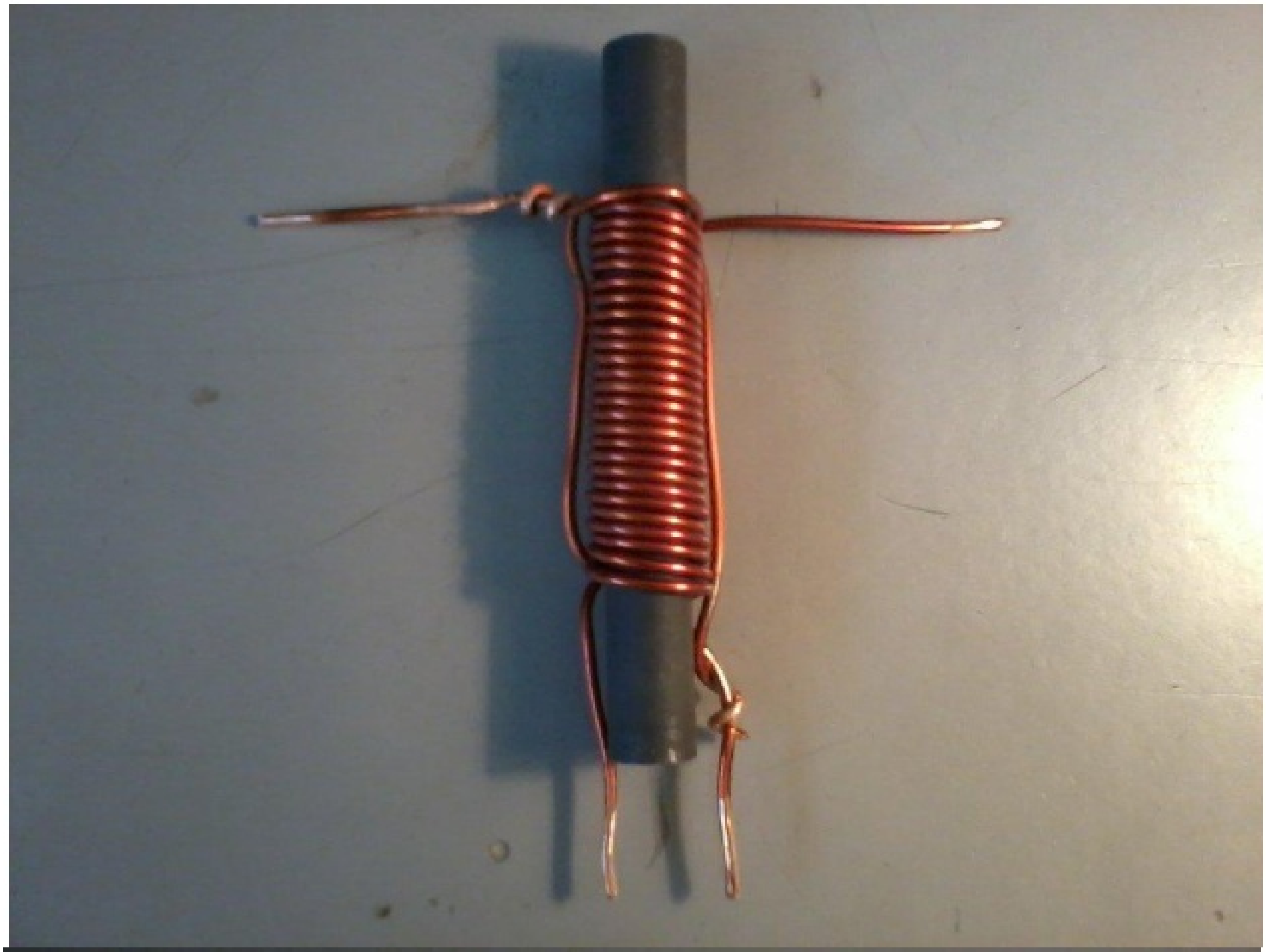

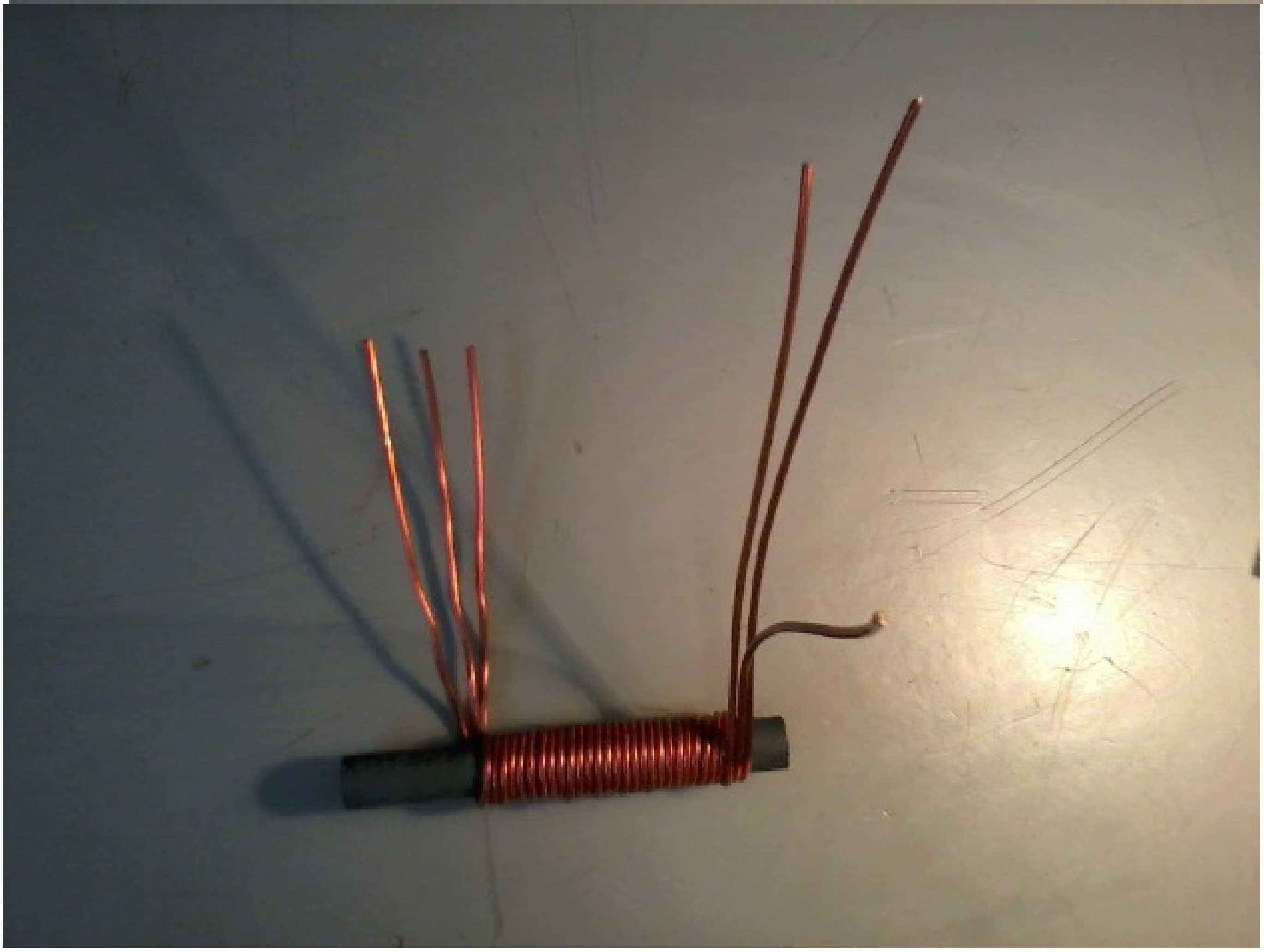

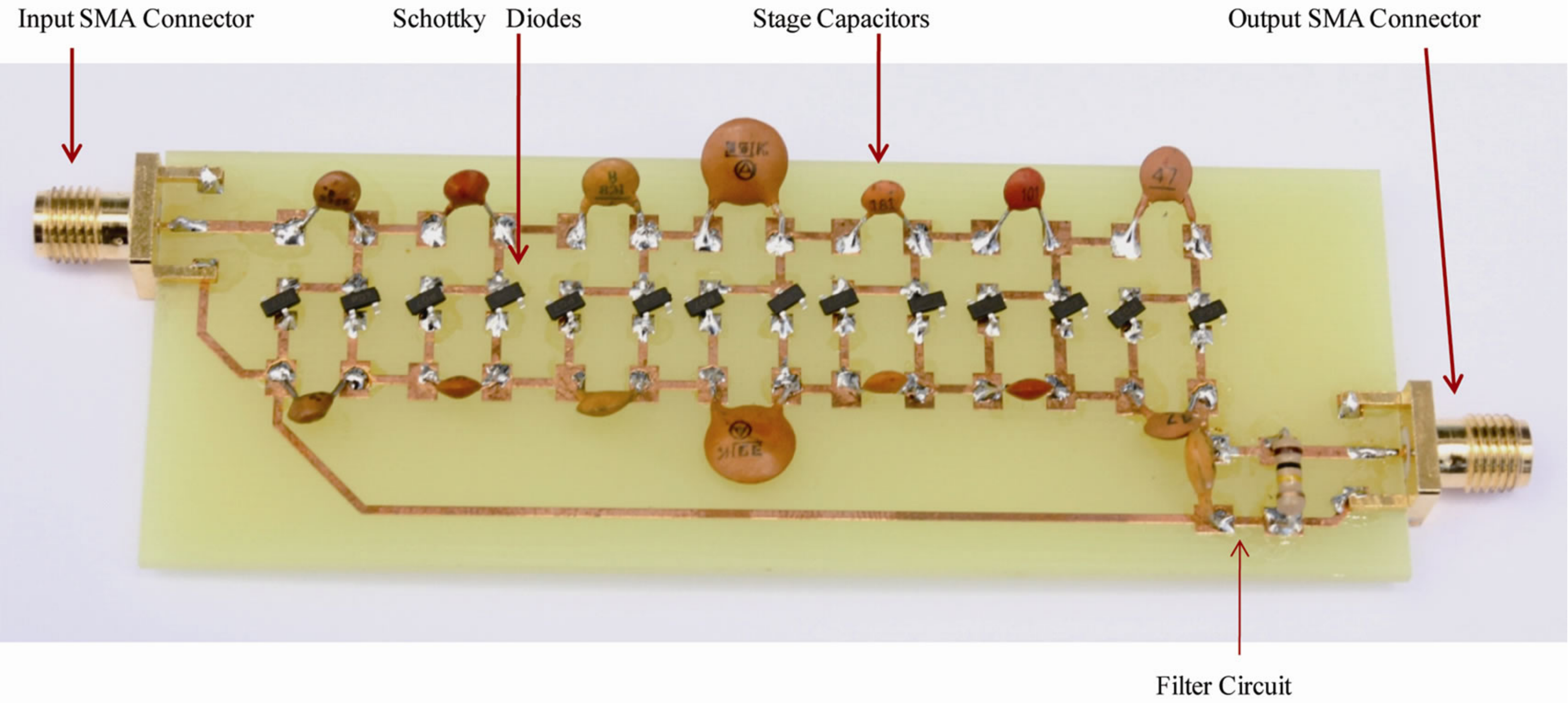
Input SMA Connector
Schottky Diodes
Stage Capacitors
Output SMA Connector
Filter Circuit

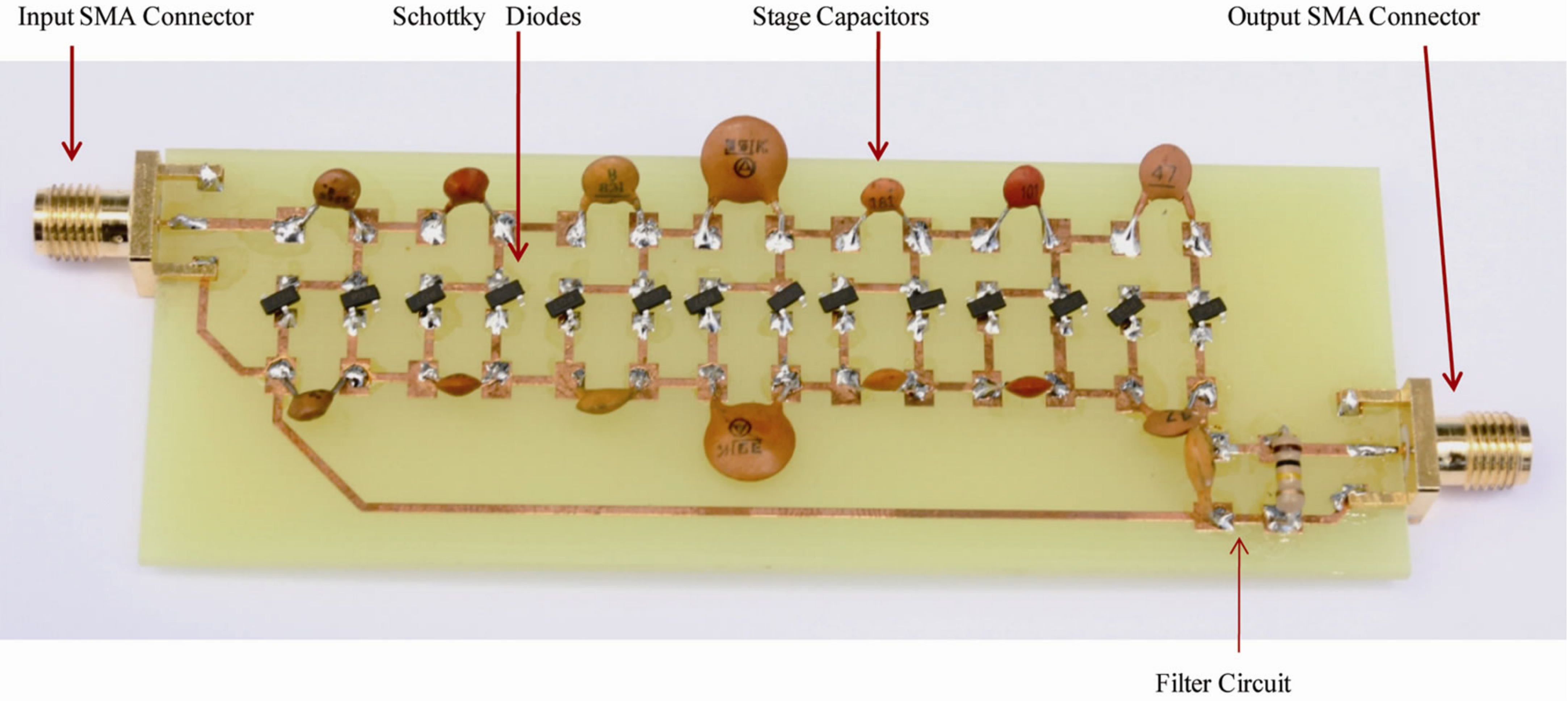
Input SMA Connector
Schottky Diodes
Stage Capacitors
Output SMA Connector
Filter Circuit

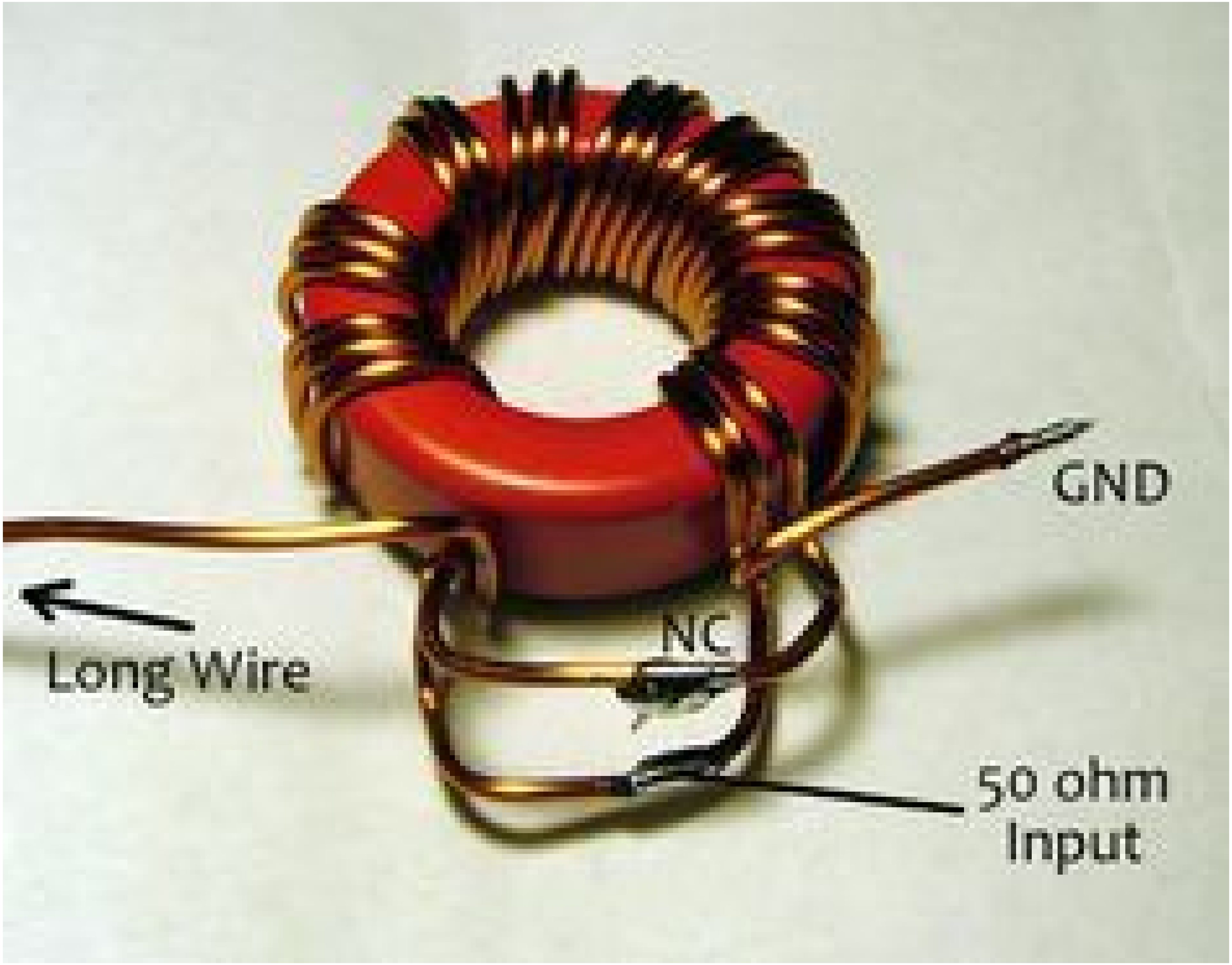
GND
Long Wire
NC
50 ohm
Input

How to hang a BALUN without damaging BALUN Antenna terminals.

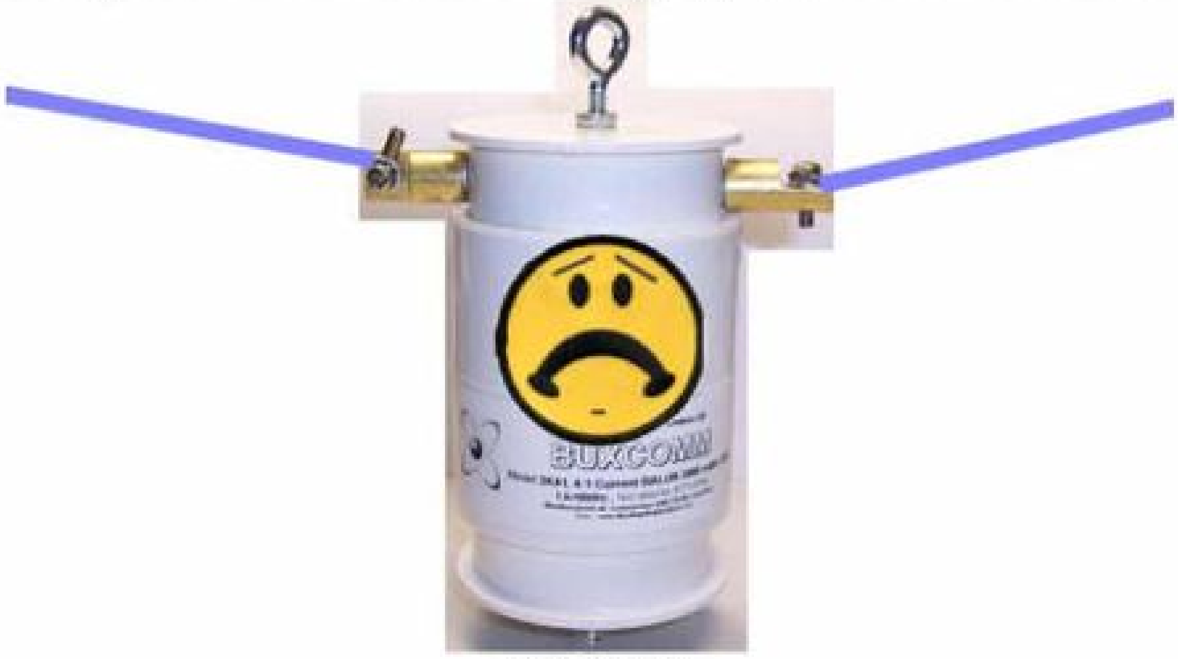

"WRONG"
Do not support BALUN by Antenna Wire/Elements,

Use Eyebolt to support BALUN from Center Insulator.

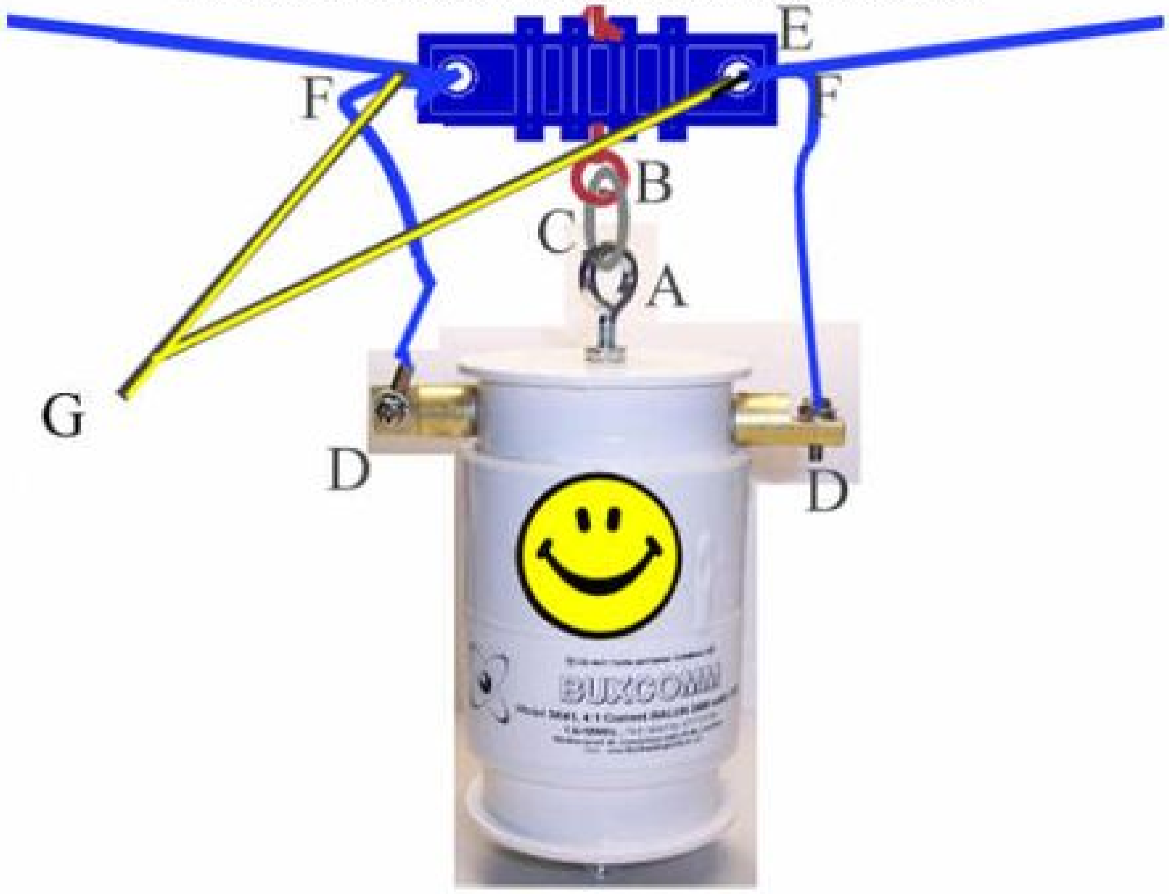

"RIGHT"

A= BALUN Support Eyebolt,
B= Center Insulator Eyebolt Support
C= Support Link between Center Insulator and BALUN Eyebolt
D= Antenna Terminals/BALUN Connectors
E= Center Insulator
F/F=Loop antenna wire element(s) through insulator holes ,
G= Fold back wire element onto itself, secure with Ty-Wrap and cover with tape.
Visit: www.HamRadioExpress.com

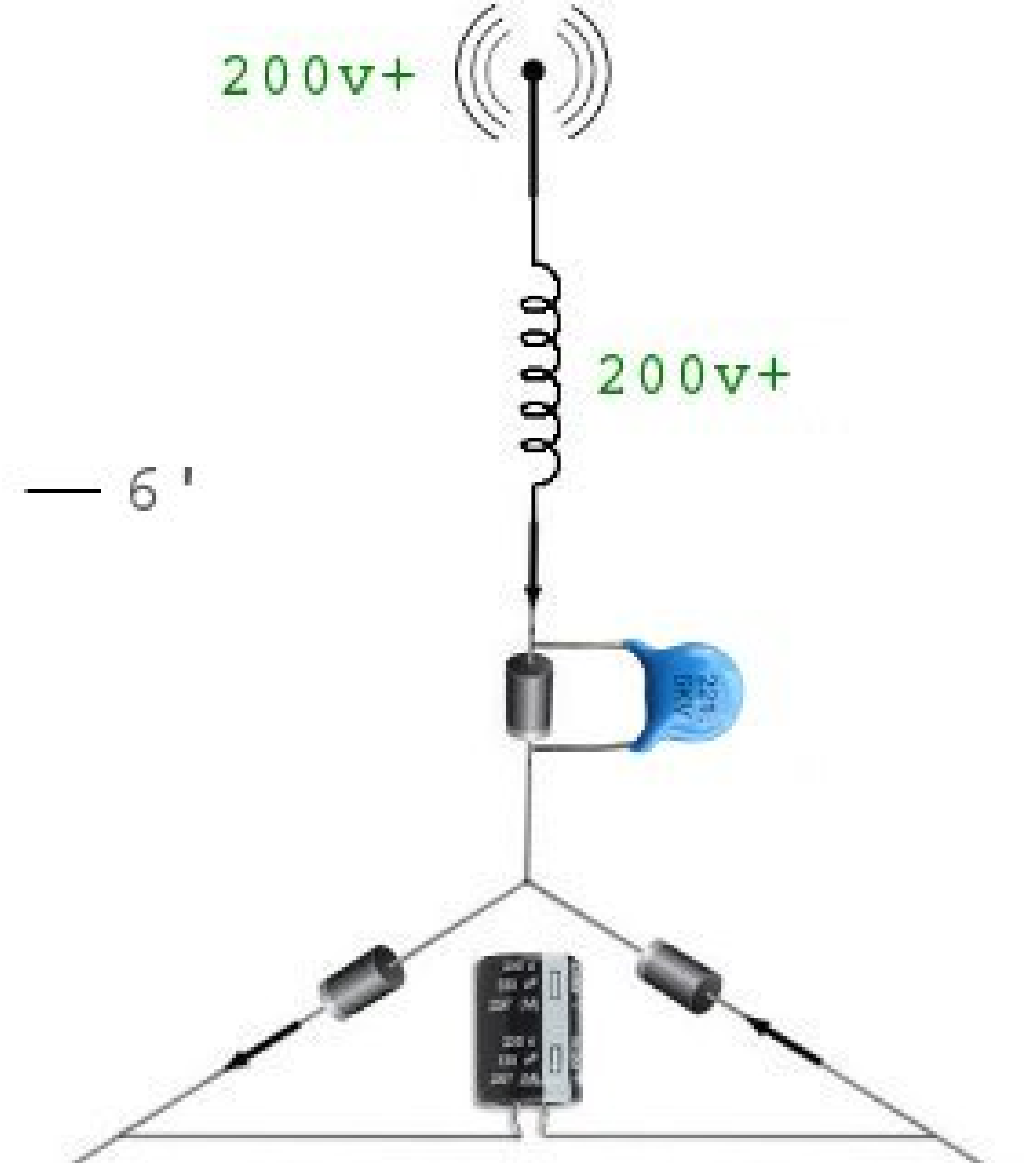
200v+
200v+
— 6 '

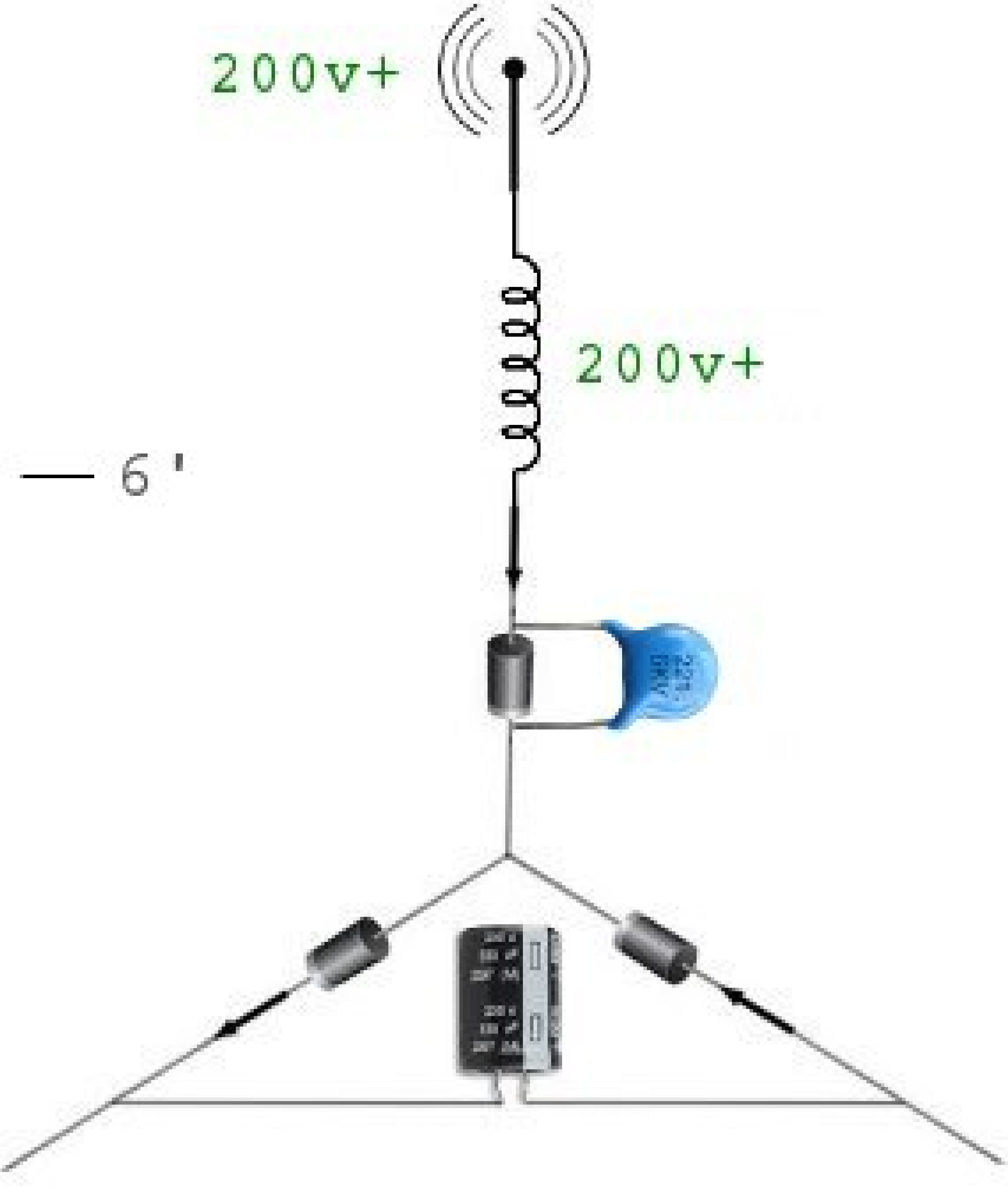
200v+
200v+
— 6'

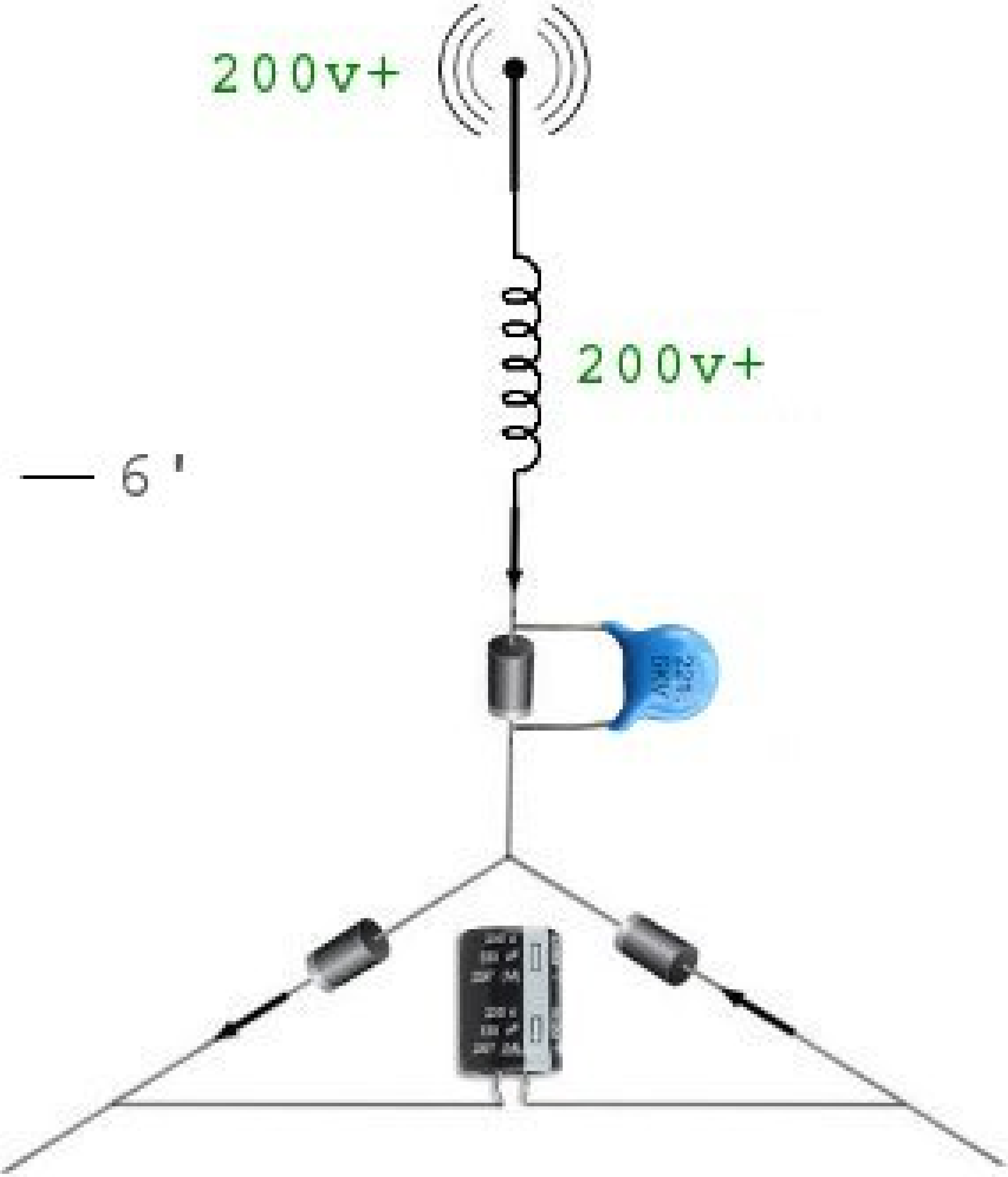
200v+
200v+
— 6 '

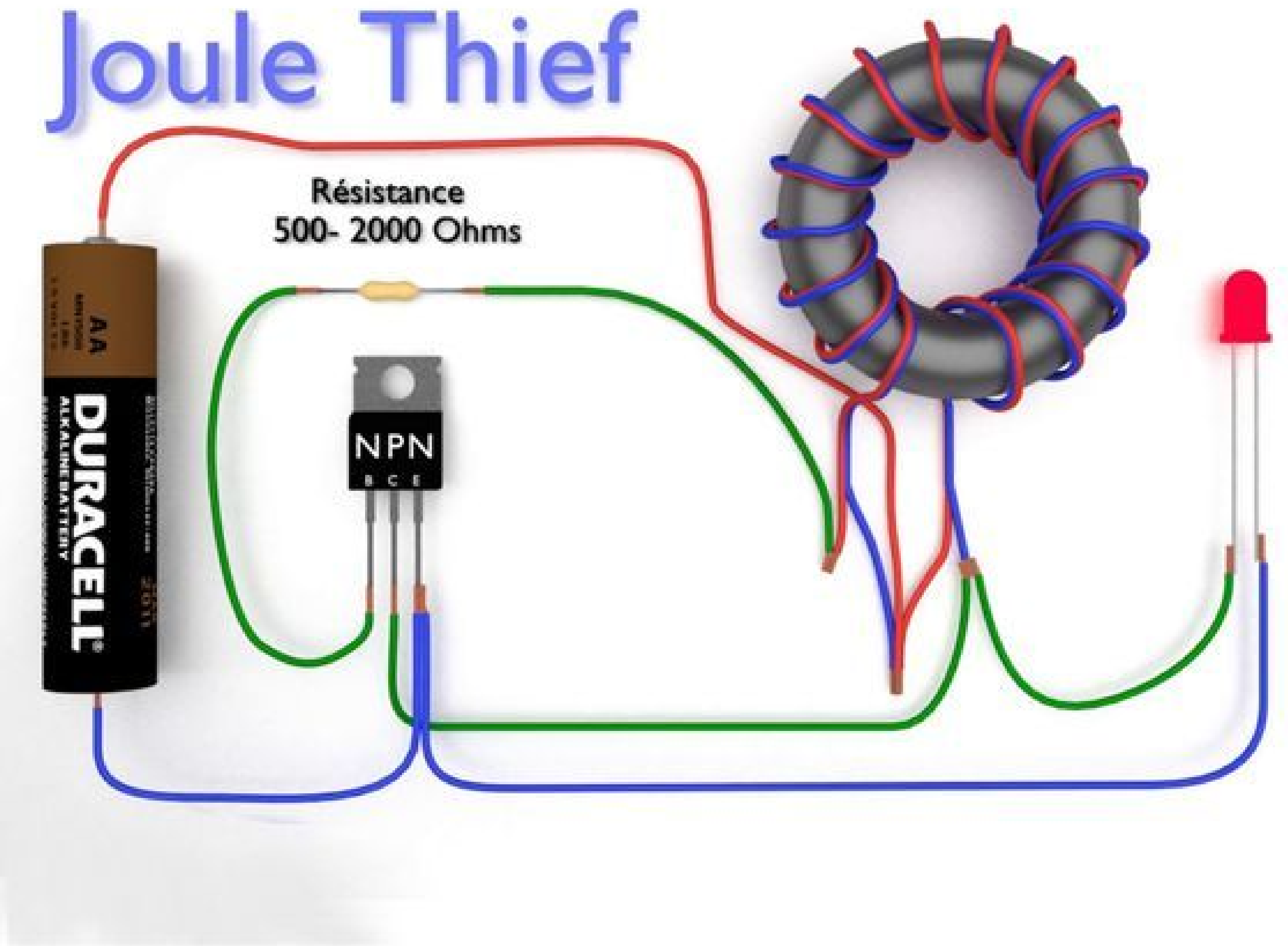
Joule Thief
Résistance
500- 2000 Ohms
NPN
B C E
AA
DURACELL
ALKALINE BATTERY

FIG. 5 APM charging a battery

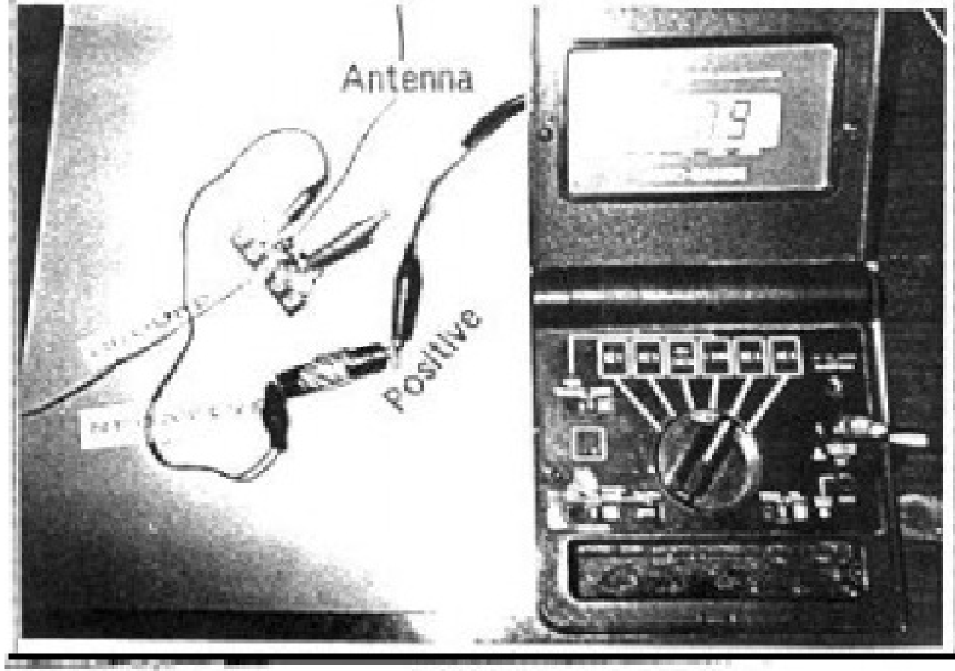
Antenna
Positive

FIG. 6

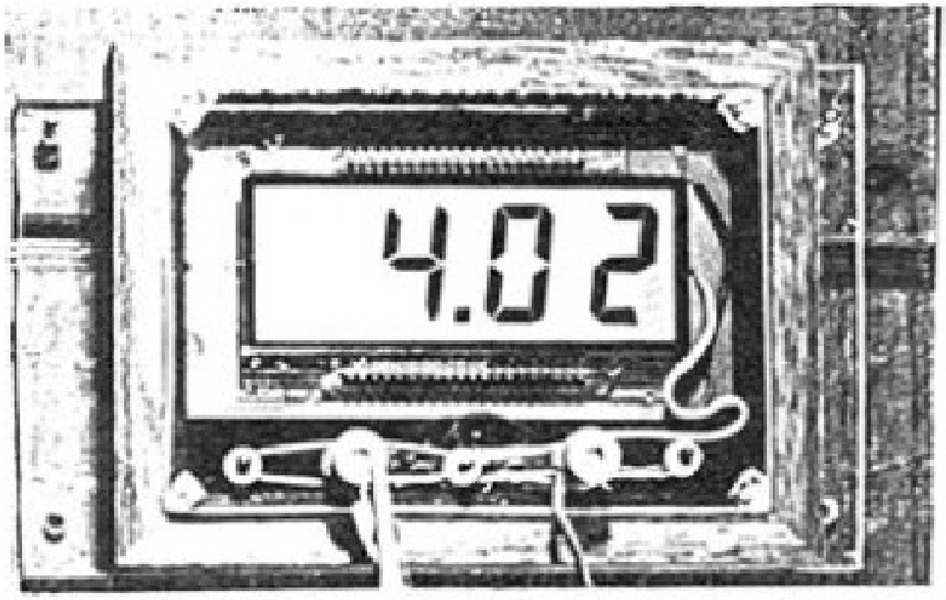

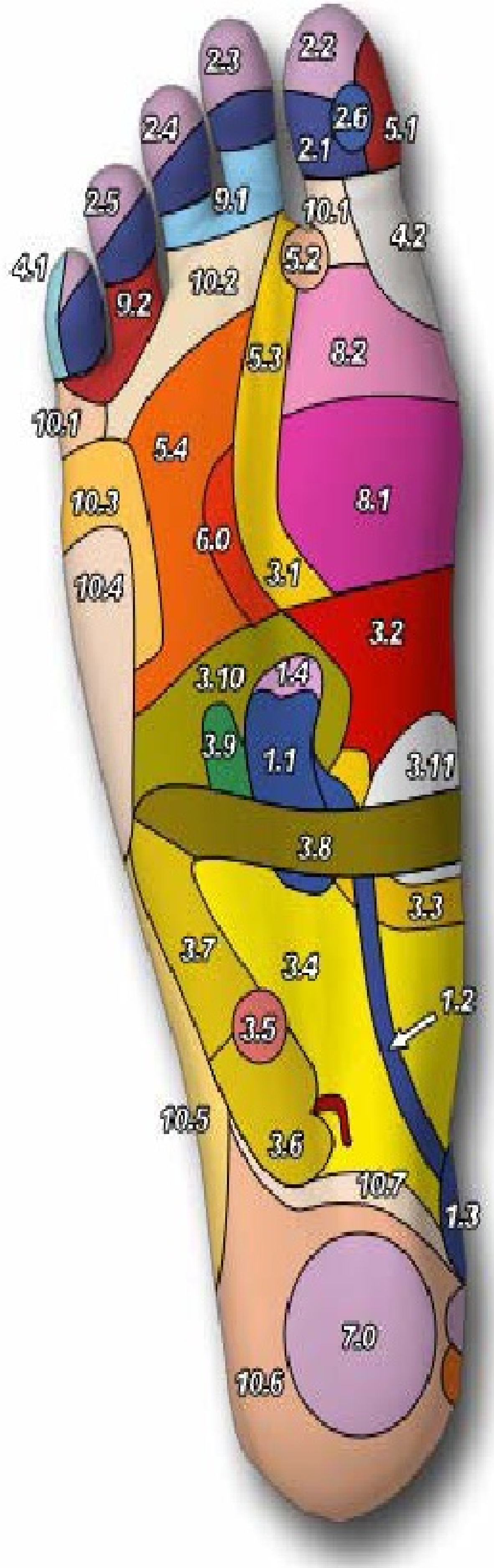
2.2
2.3
2.6
5.1
2.4
2.1
2.5
9.1
10.1
4.2
4.1
9.2
10.2
5.2
5.3
8.2
10.1
5.4
10.3
8.1
6.0
3.1
10.4
3.2
3.10
1.4
3.9
1.1
3.11
3.8
3.3
3.7
3.4
1.2
3.5
10.5
3.6
10.7
1.3
7.0
10.6

Reflexology Map of the Male Right Foot (sole)

1.1 Kidney
1.2 Ureter
1.3 Bladder
1.4 Adrenal gland

2.1 Cerebellum
2.2 Cerebrum
2.3 Frontal lobe
2.4 Temporal lobe
2.5 Occipital lobe
2.6 Pituitary gland

3.1 Oesophagus
3.2 Stomach and liver
3.3 Duodenum
3.4 Small intestine
3.5 Ileocaecal valve
3.6 Appendix (caecum)
3.7 Ascending colon
3.8 Transverse colon
3.9 Gall bladder
3.10 Liver
3.11 Pancreas

4.1 Occiput
4.2 Cervical vertebrae

5.1 Nose
5.2 Larynx
5.3 Trachea, bronchi
5.4 Lung, bronchi

6.0 Heart

7.0 Gential organs

8.1 Thymus gland
8.2 Thyroid glands

9.1 Eye, inside
9.2 Ear, inside

10.1 Nape
10.2 Trapezoid musculature
10.3 Shoulder blade and joint
10.4 Upper arm
10.5 Knee, elbow, leg, forearm
10.6 Pelvis, seat
10.7 Groin

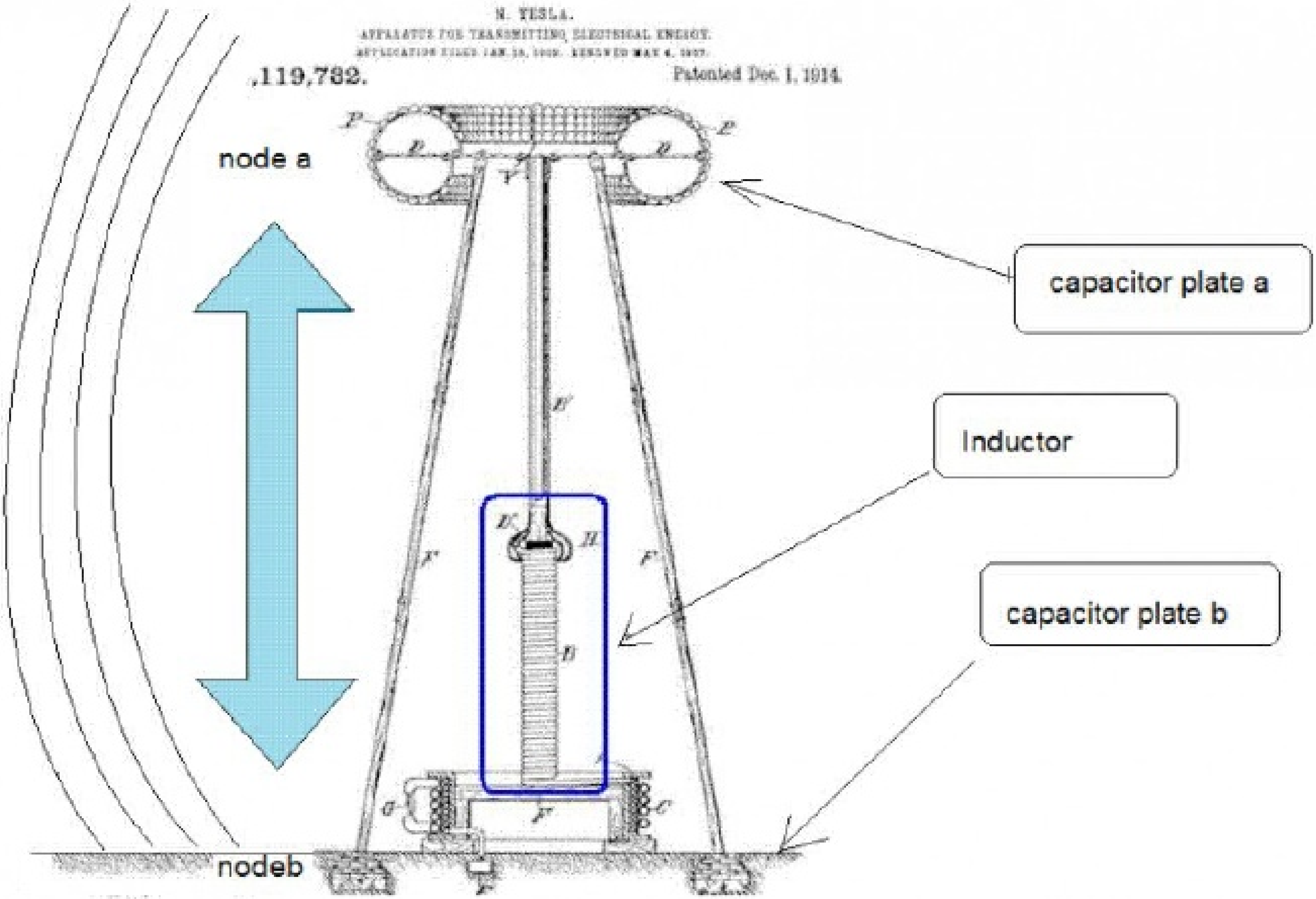
N. TESLA.
APPARATUS FOR TRANSMITTING ELECTRICAL ENERGY.
APPLICATION FILED JAN 18, 1902. RENEWED MAY 4, 1907.
,119,732.
Patented Dec. 1, 1914.
P
P
D
D
V
E
E'
H
B
F
F
F
G
C
F

node a

capacitor plate a

Inductor

capacitor plate b

nodeb

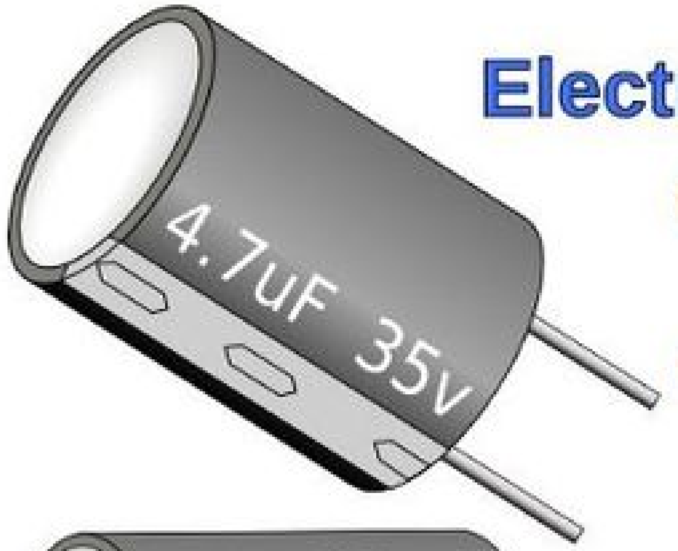

Electrolytic Capacitors

- Value shown in micro-farads [uF]
- Negative terminal shown by stripe or arrow

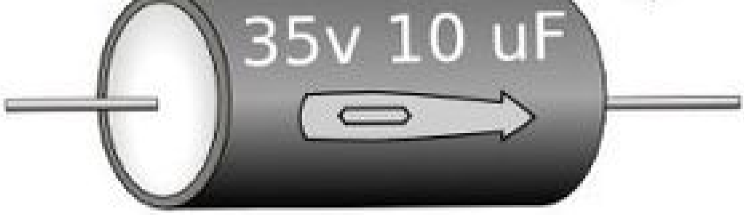

ROBOT MILL

Disc / Ceramic Capacitors

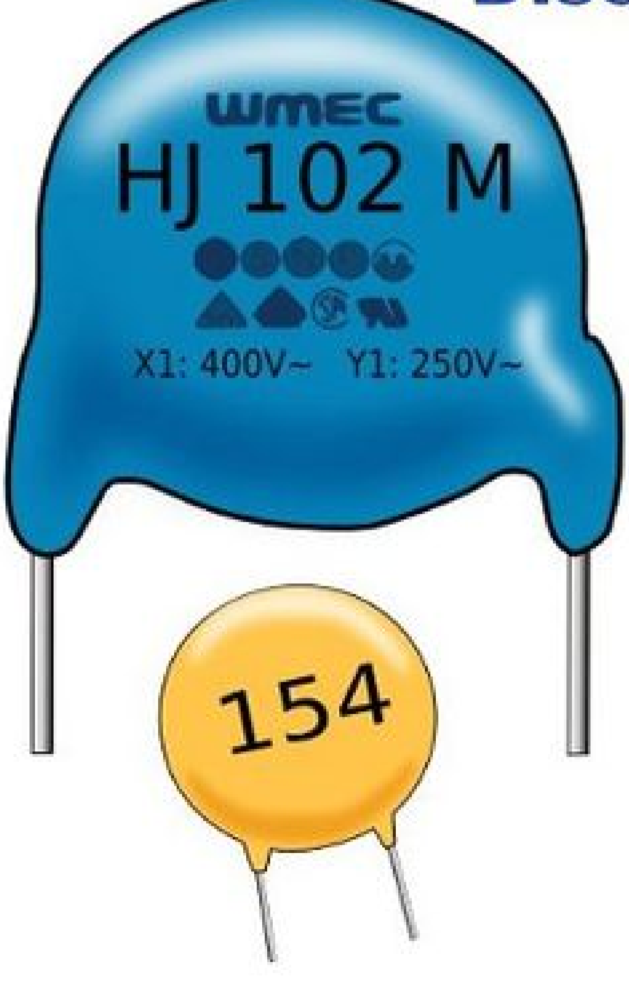
WMEC
HJ 102 M
X1: 400V~ Y1: 250V~
154

- Value shown in pico-farads [pF]

  (1 pF = 0.000001 uF)
- Non-polar (no + / - terminals)
- 3rd digit is multiplier

  ex: 154 = 150000 pF

0 = x1
1 = x10
2 = x100
3 = x1000
4 = x10000
5 = x100000
8 = x .01
9 = x .1

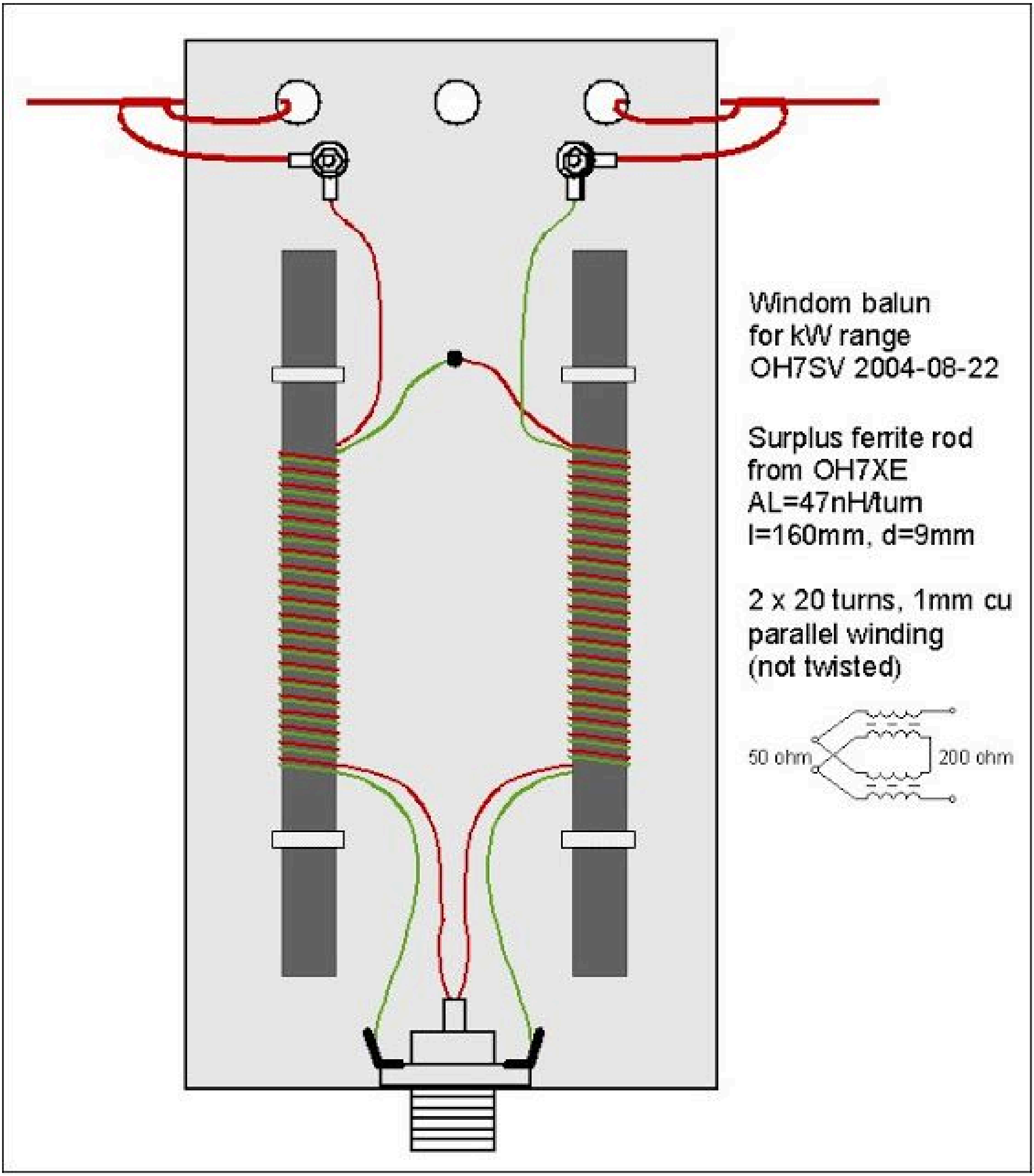

Windom balun
for kW range
OH7SV 2004-08-22

Surplus ferrite rod
from OH7XE
AL=47nH/turn
l=160mm, d=9mm

2 x 20 turns, 1mm cu
parallel winding
(not twisted)

Vest Pocket Crystal Receiver

Here's a little crystal receiver which is no larger than a package of cigarettes.

by L. B. Robbins

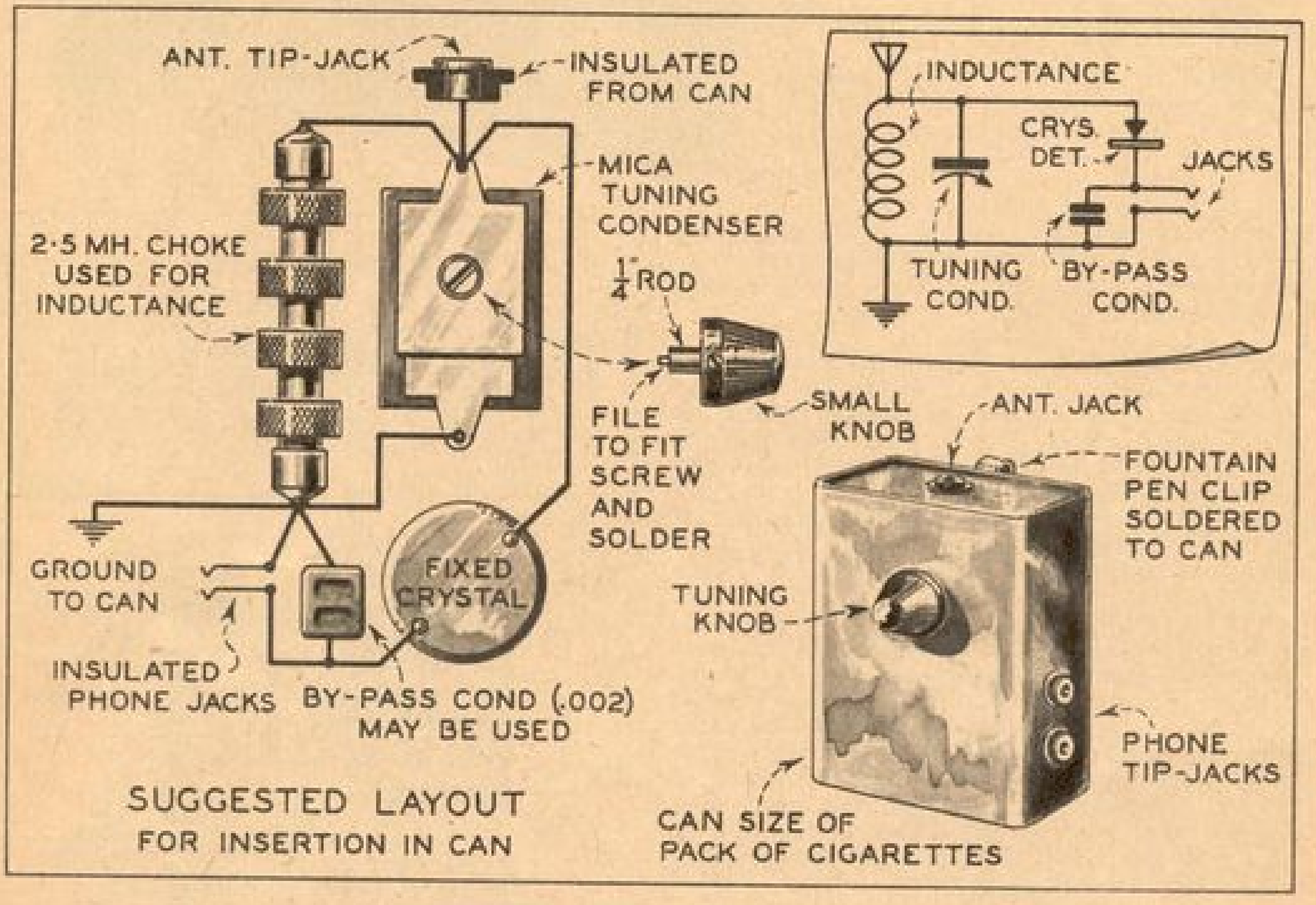
ANT. TIP-JACK
INSULATED FROM CAN
MICA TUNING CONDENSER
2·5 MH. CHOKE USED FOR INDUCTANCE
1/4" ROD
FILE TO FIT SCREW AND SOLDER
SMALL KNOB
GROUND TO CAN
INSULATED PHONE JACKS
FIXED CRYSTAL
BY-PASS COND (.002) MAY BE USED
SUGGESTED LAYOUT
FOR INSERTION IN CAN
INDUCTANCE
CRYS. DET.
JACKS
TUNING COND.
BY-PASS COND.
ANT. JACK
FOUNTAIN PEN CLIP SOLDERED TO CAN
TUNING KNOB
PHONE TIP-JACKS
CAN SIZE OF PACK OF CIGARETTES

ADJUSTABLE in tuning and carried in the upper breast pocket, this tiny receiver is held securely in place by an ordinary fountain-pen clip. With but two or three feet of antenna wire and a pair of head-phones one can clearly receive broadcasts within a radius of two or three miles, or within city limits. The can used to house the original set was one constructed to hold just one pack of cigarettes. Some dime stores sell them or they may be procured at cigar stores.

The circuit used is the simplest known and is shown in the diagram. It consists of a tuning coil in series with a fixed crystal and midget phone tip-jacks; the circuit being shunted by a suitable tuning condenser of the compression type commonly known as a padding condenser. Compactness is obtained mainly by utilizing a 2.5 millihenry choke, used in short wave work, as the tuning coil. The four pies, tuned by a condenser of about 250 mmfd. maximum capacity, will cover most of the broadcast band quite well.

Lay out the parts on the bench in a compact form as shown in the drawing, arranging them according to their design so they will fit into the can with the condenser tuning shaft in a position to enter a suitable hole in the upper center of the can. Any parts that may come in contact with the can should be covered with bits of electricians' tape. Wire and connect the various elements together with buss wire, if possible, to gain solidity. Fit the two phone tip-jacks in the lower edge as shown, and arrange the insulated tip-jack for the antenna in the center of the bottom (used as the top). The open top (bottom) can be closed later if desired. Grounded portions can be caught to the can with a bit of solder when the assembly finally is made. When completely assembled in the can, a short piece of brass rod can be filed and fitted into [*Continued on page 144*]

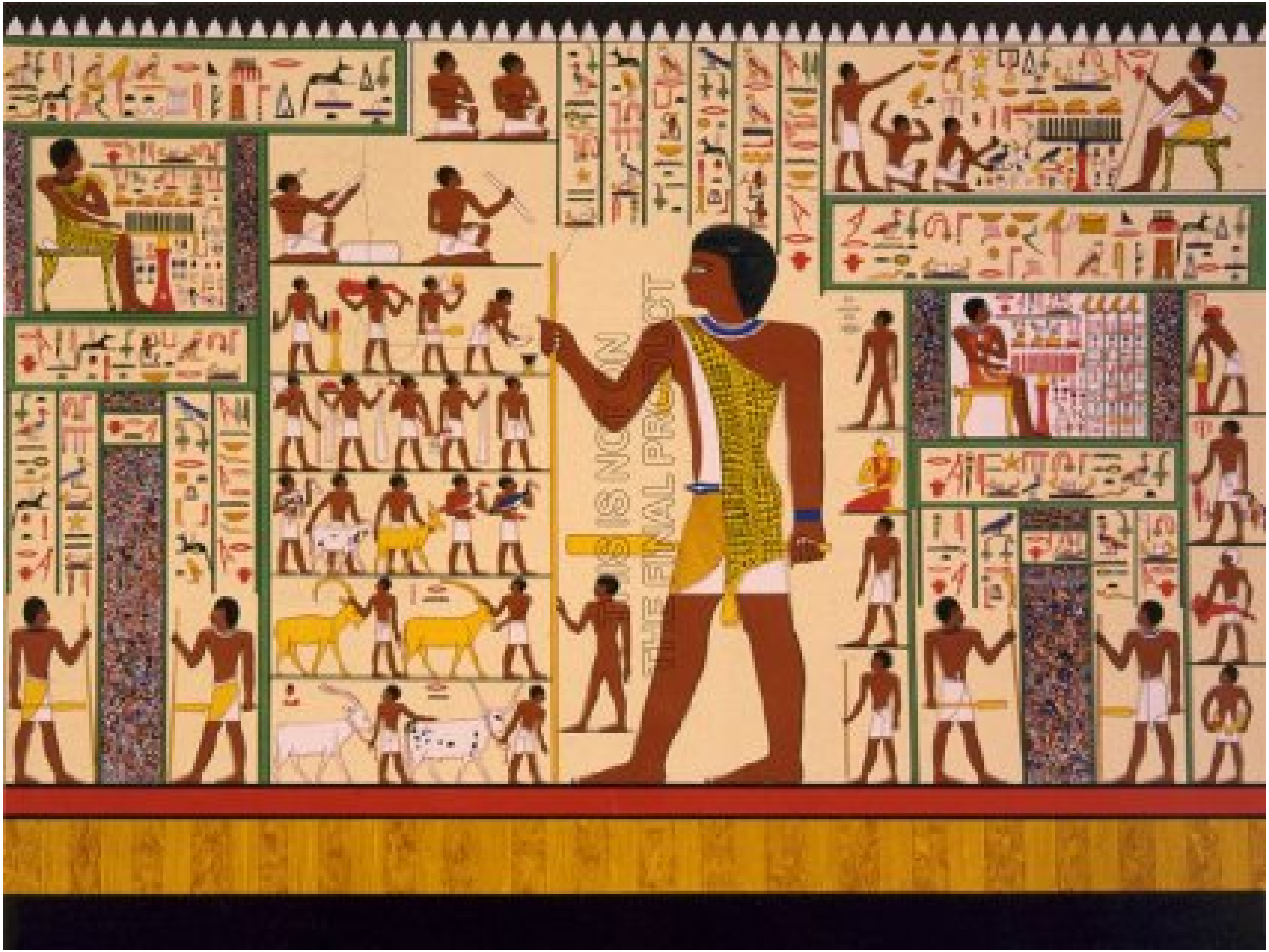

THIS IS NOT THE FINAL PRODUCT

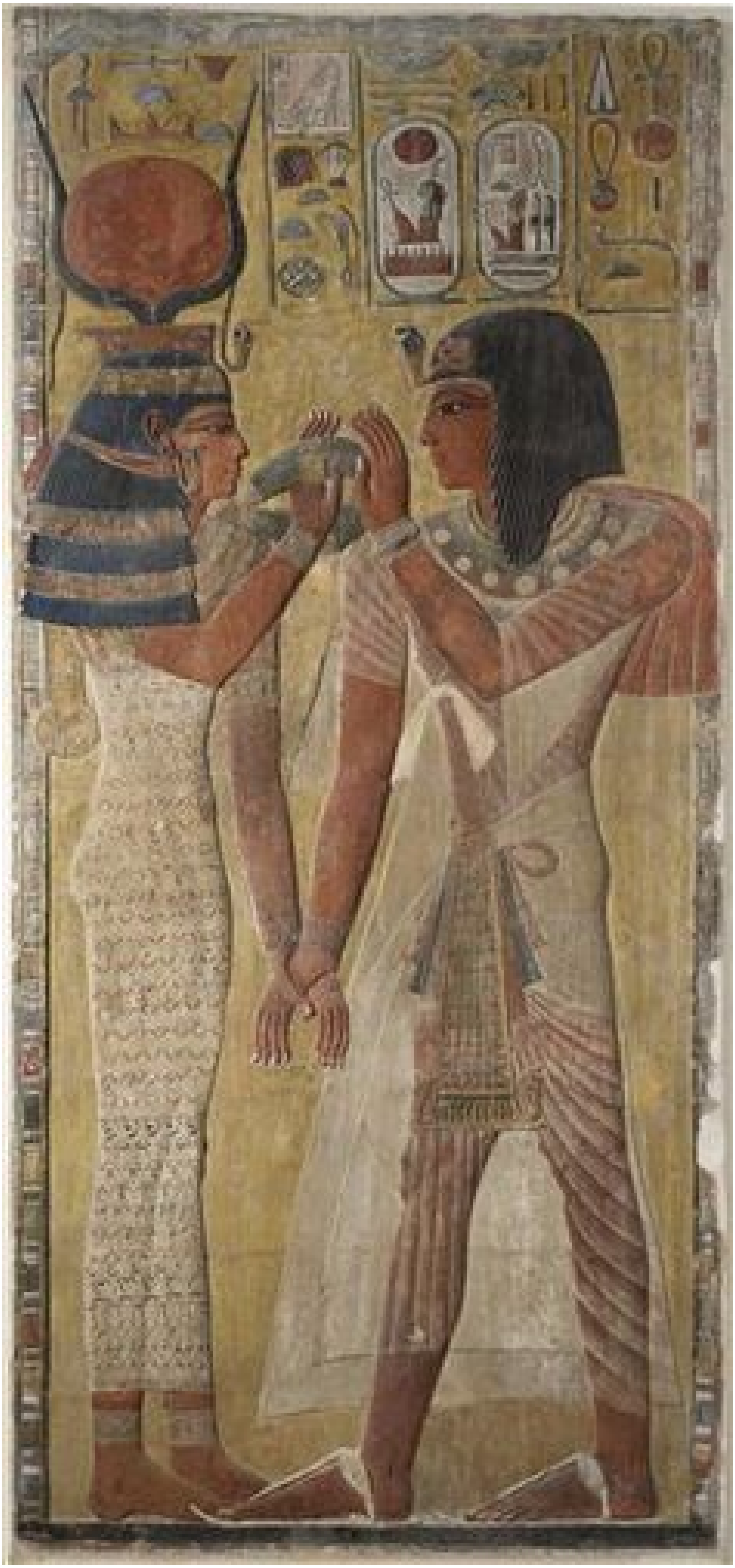

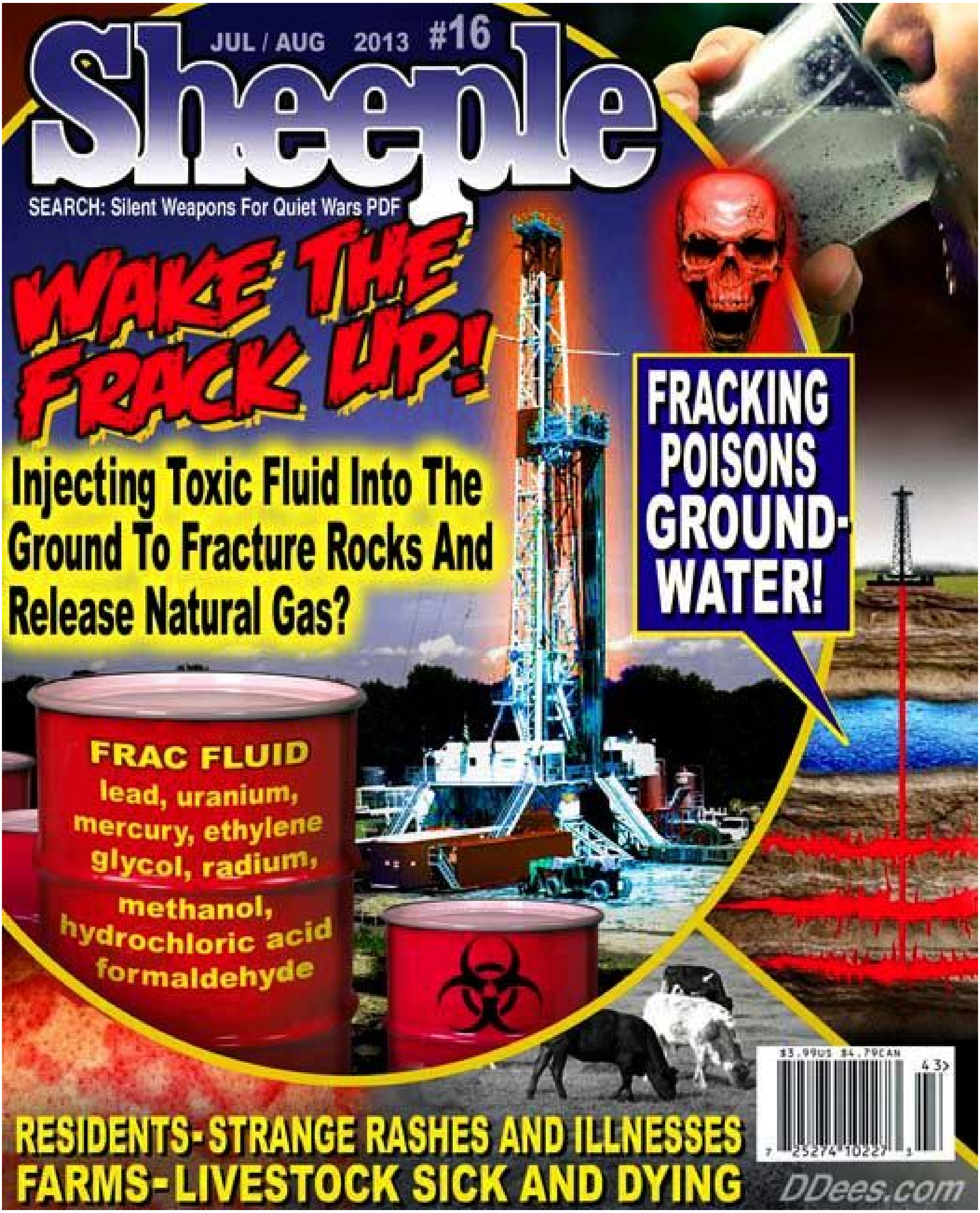
JUL / AUG 2013 #16
Sheeple
SEARCH: Silent Weapons For Quiet Wars PDF
WAKE THE FRACK UP!
Injecting Toxic Fluid Into The Ground To Fracture Rocks And Release Natural Gas?
FRACKING POISONS GROUND-WATER!
FRAC FLUID
lead, uranium, mercury, ethylene glycol, radium, methanol, hydrochloric acid formaldehyde
$3.99US $4.79CAN
RESIDENTS- STRANGE RASHES AND ILLNESSES
FARMS- LIVESTOCK SICK AND DYING
DDees.com

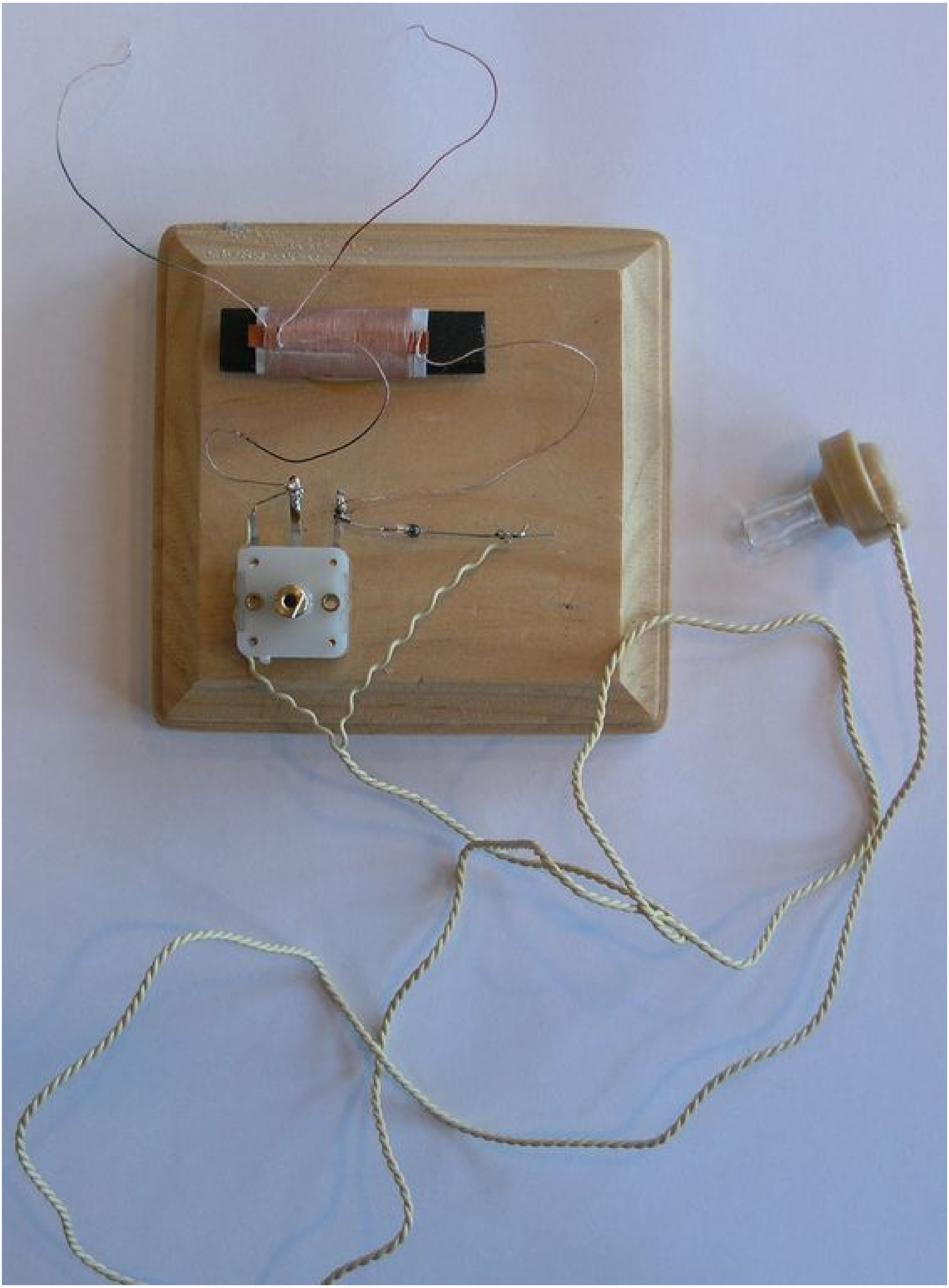

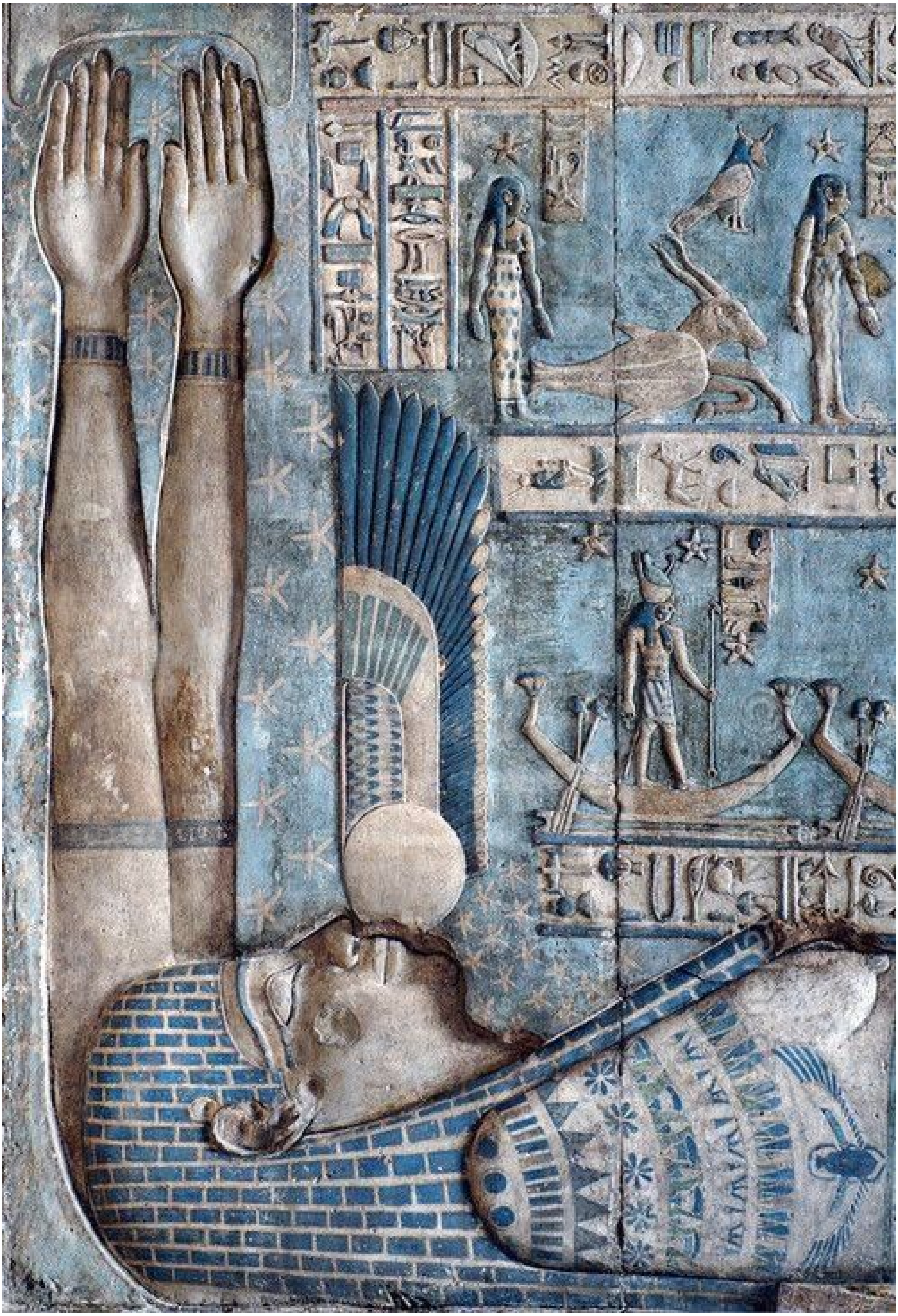

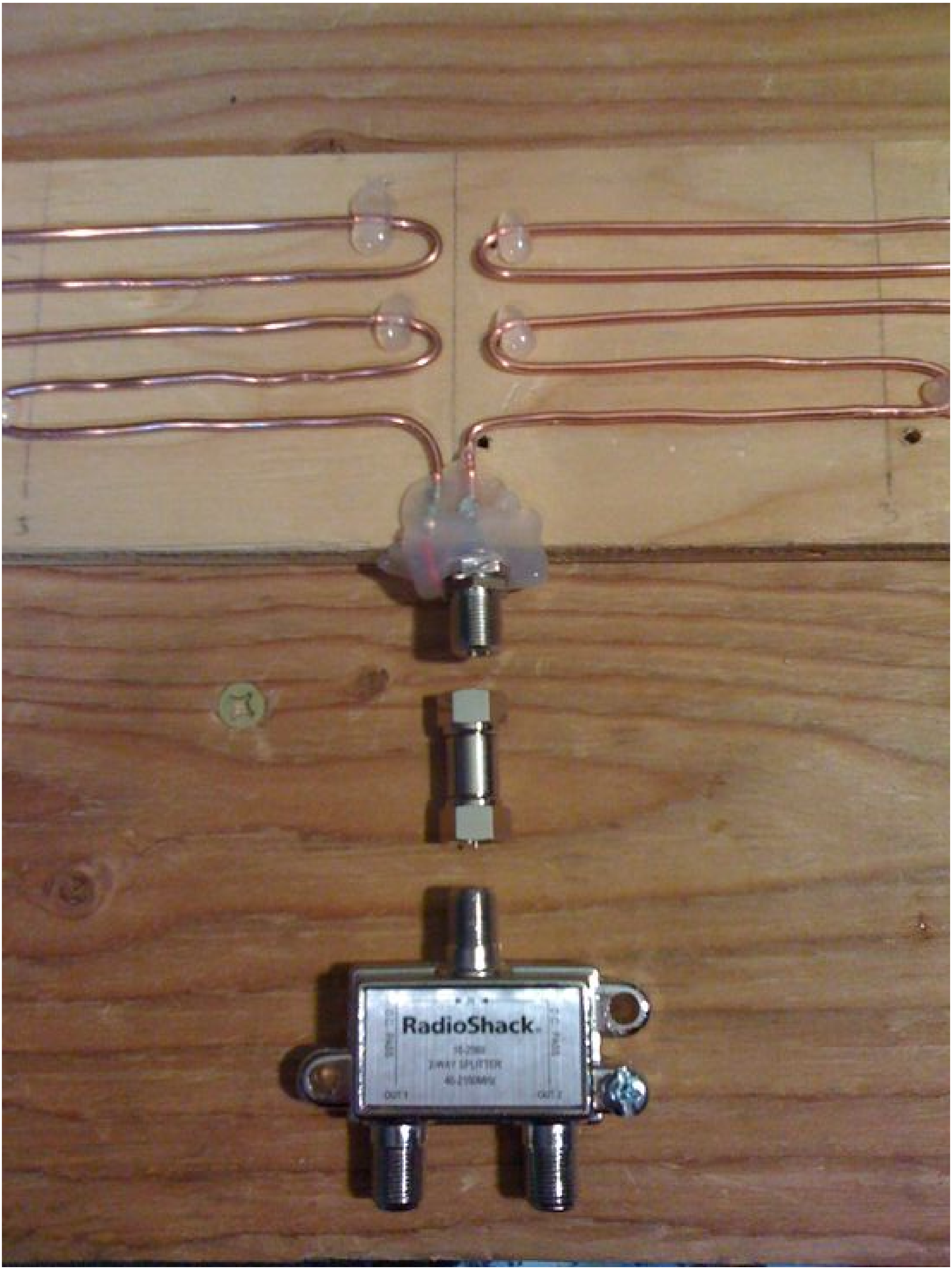
RadioShack
2-WAY SPLITTER
OUT 1
OUT 2

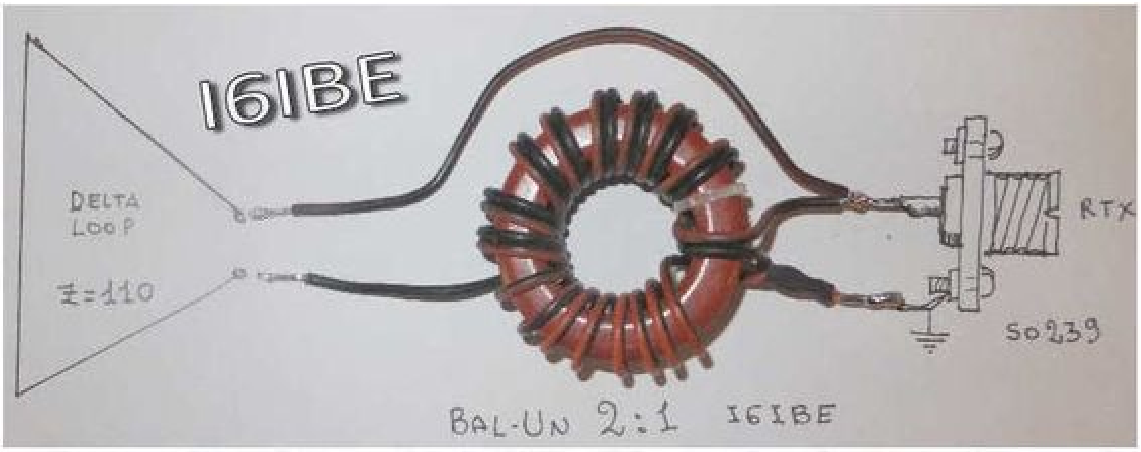
I6IBE
DELTA LOOP
Z=110
RTX
SO239
BAL-UN 2:1 I6IBE

17 spire su T-200/2 Rosso, 8 Spire bifilari (filo rosso/nero) + 9 spire (filo rosso)

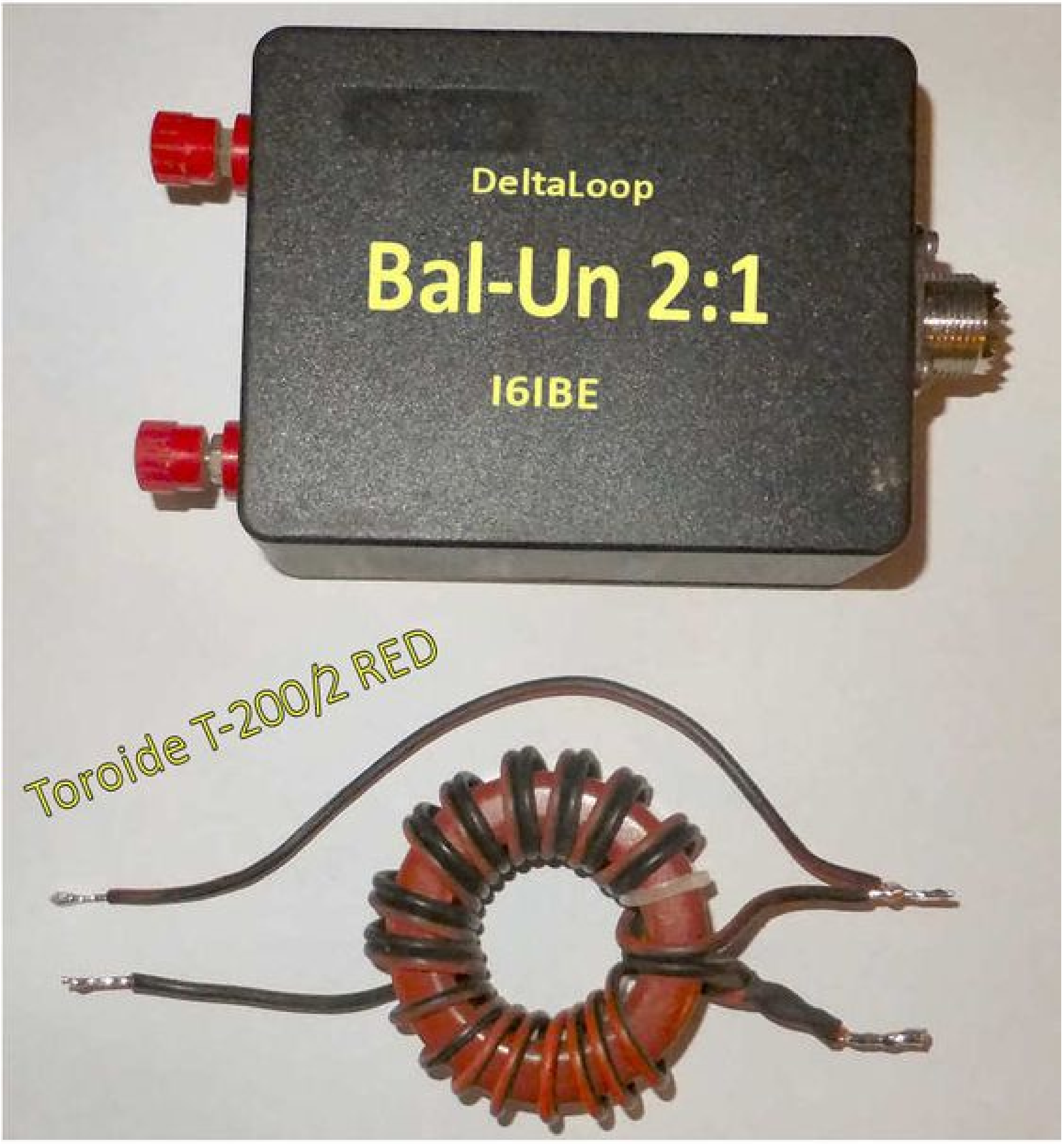
DeltaLoop
Bal-Un 2:1
I6IBE
Toroide T-200/2 RED

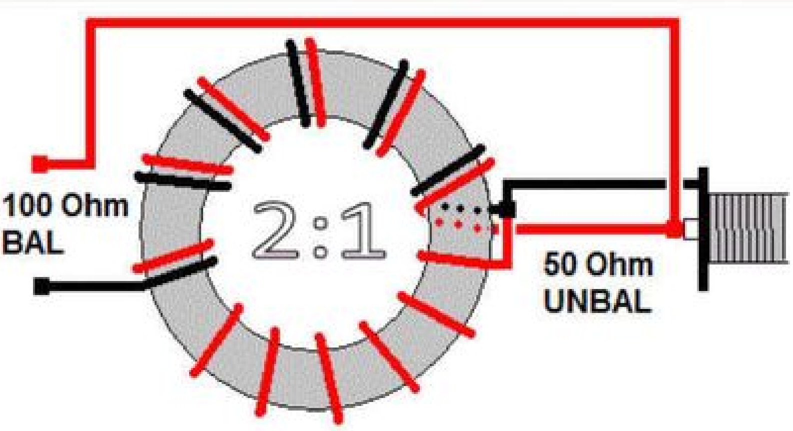
100 Ohm
BAL
2:1
50 Ohm
UNBAL

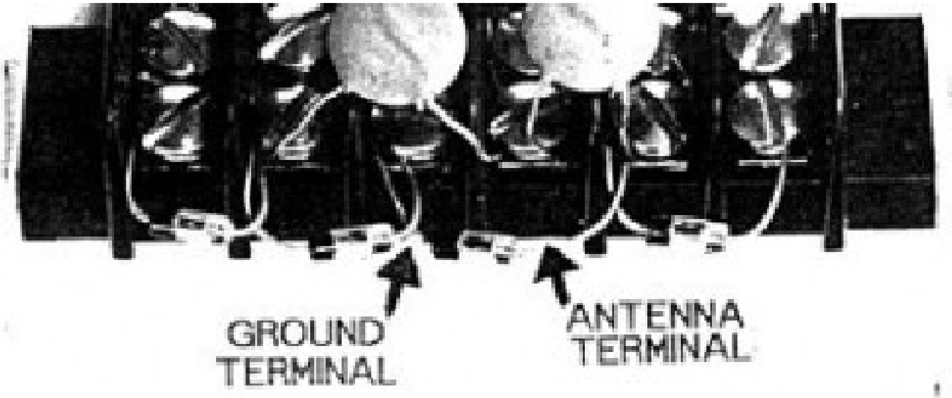
GROUND TERMINAL
ANTENNA TERMINAL

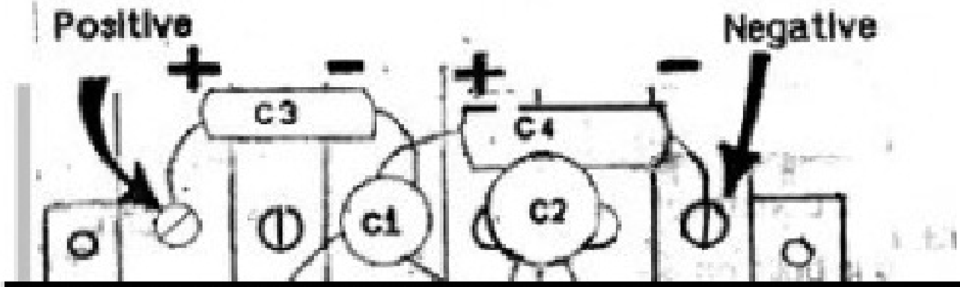
Positive
+
-
C3
+
-
C4
Negative
C1
C2

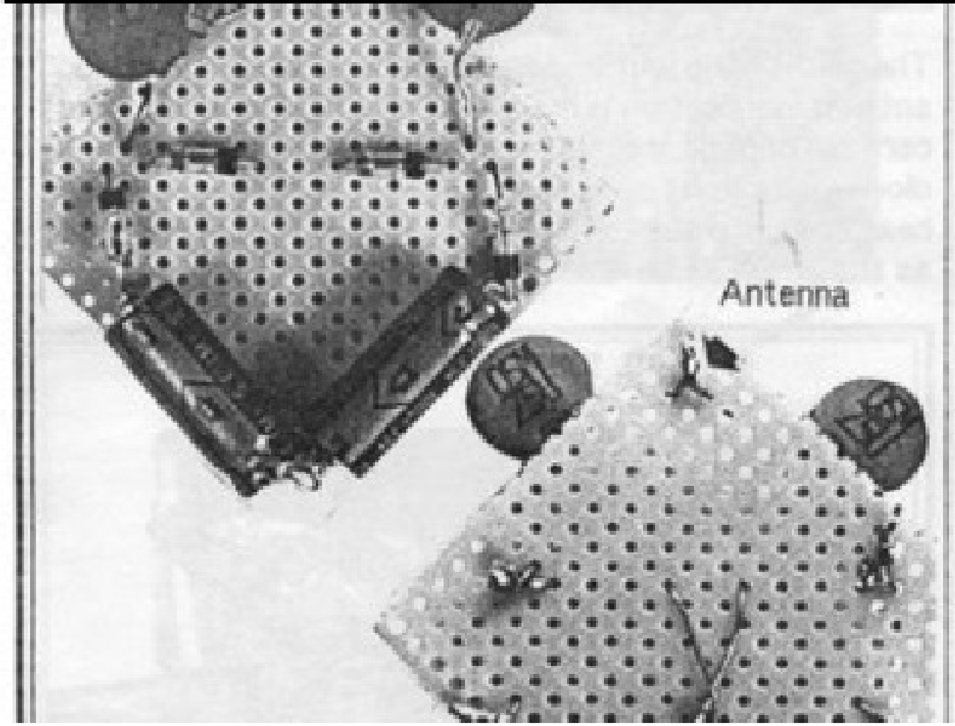
Antenna

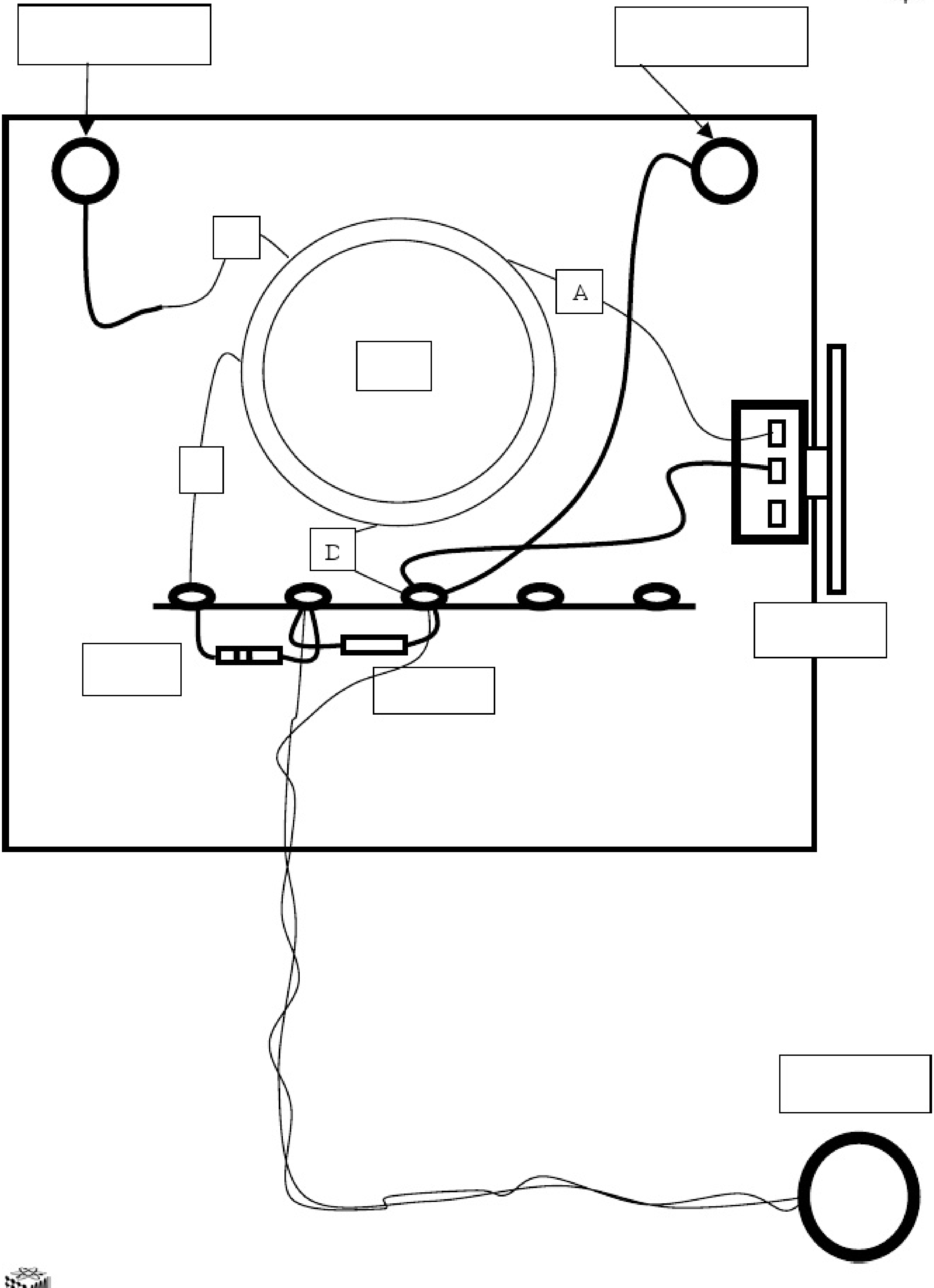
A
D

HAM AND SHORTWAVE RADIO
MICHIGAN MIGHTY MITE

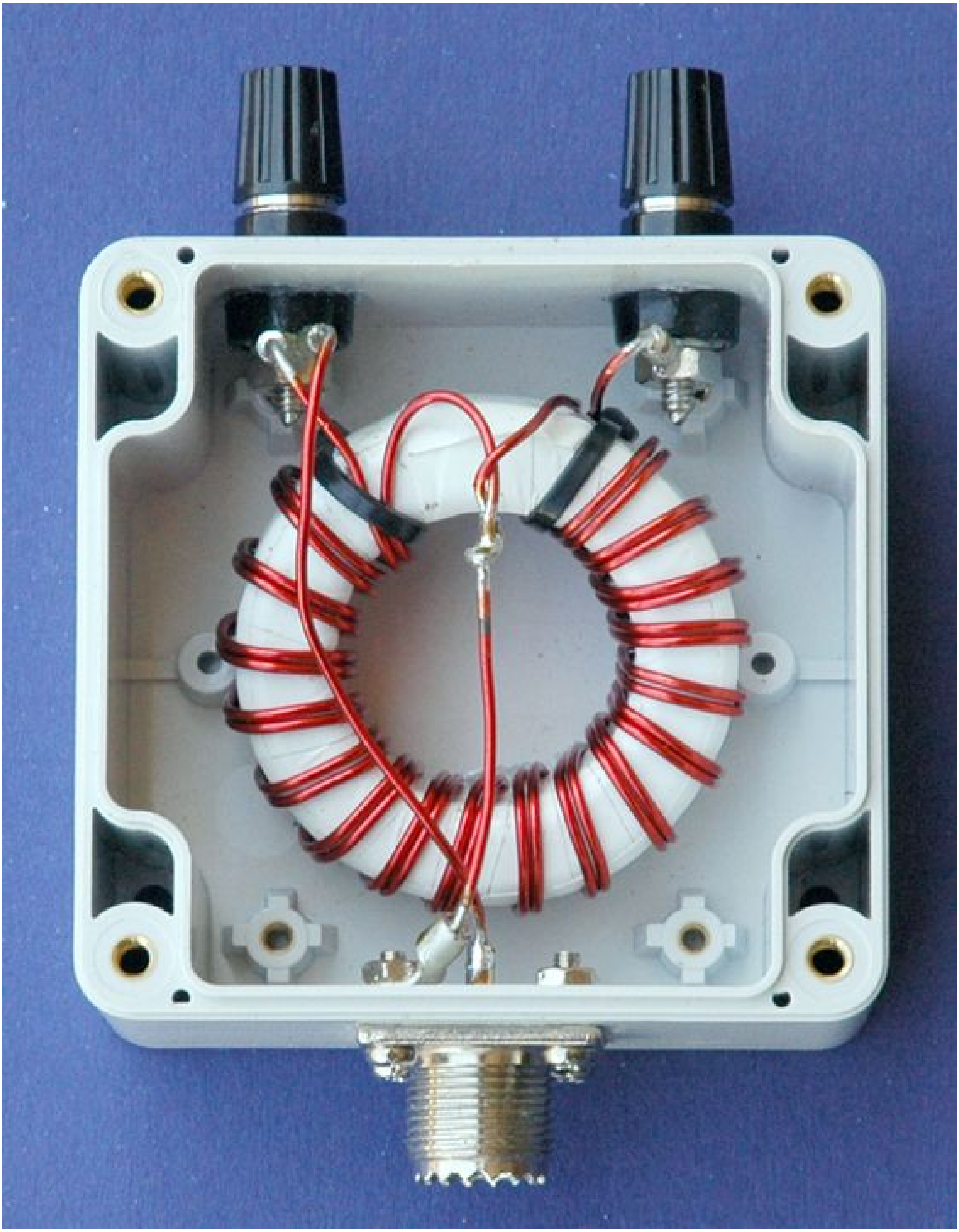

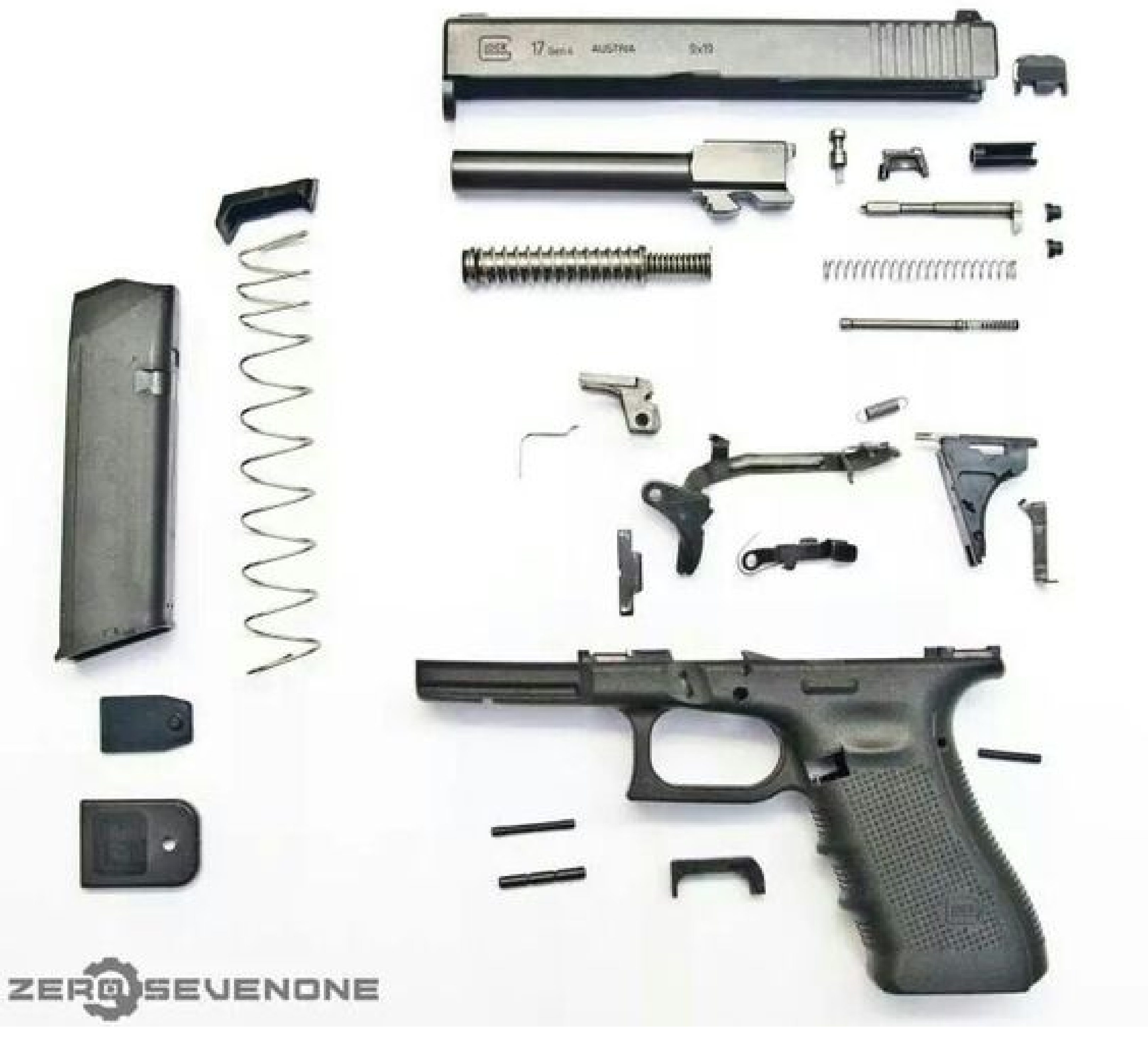
GLOCK 17 Gen4 AUSTRIA 9x19

ZERO SEVENONE

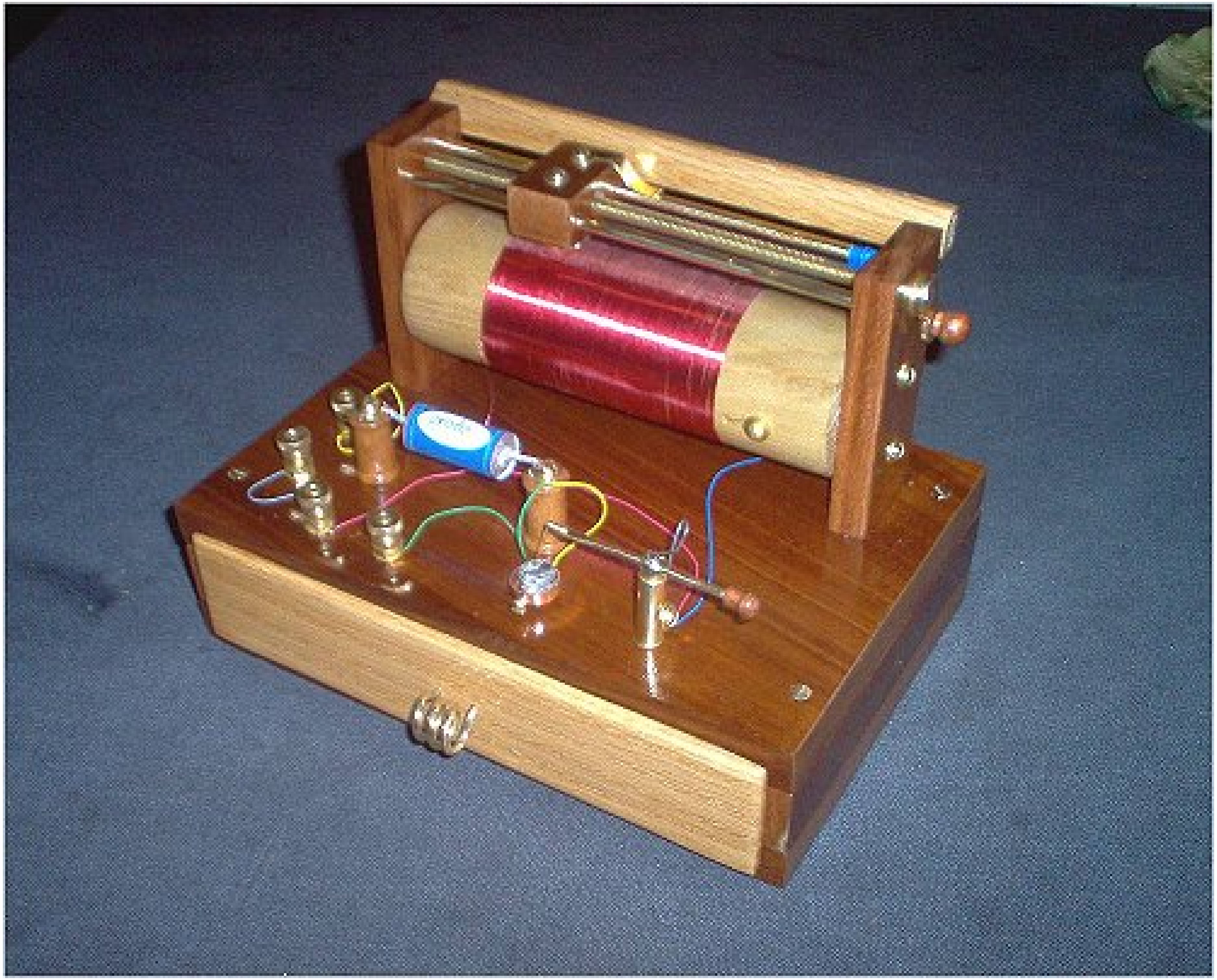

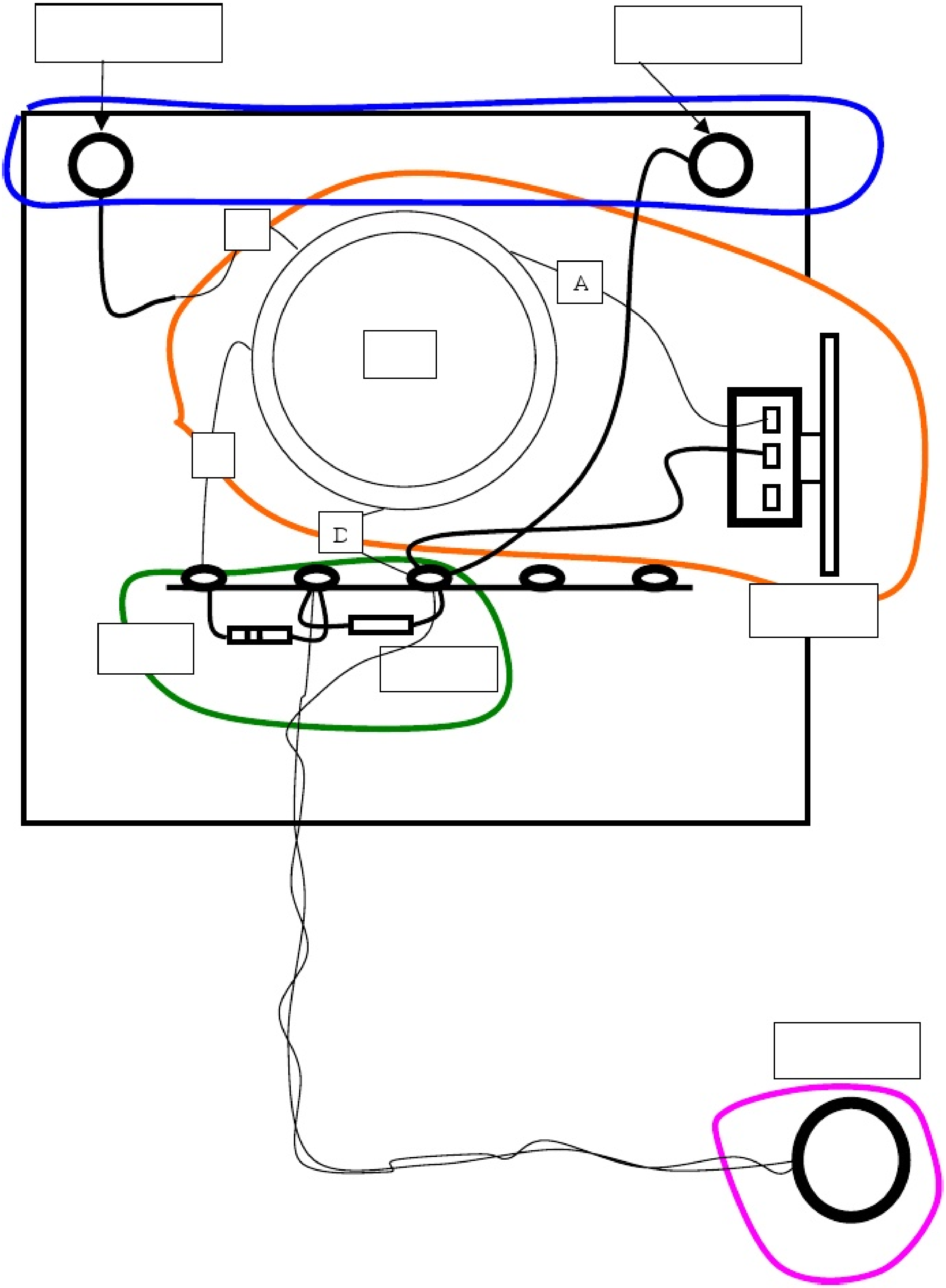
A
D

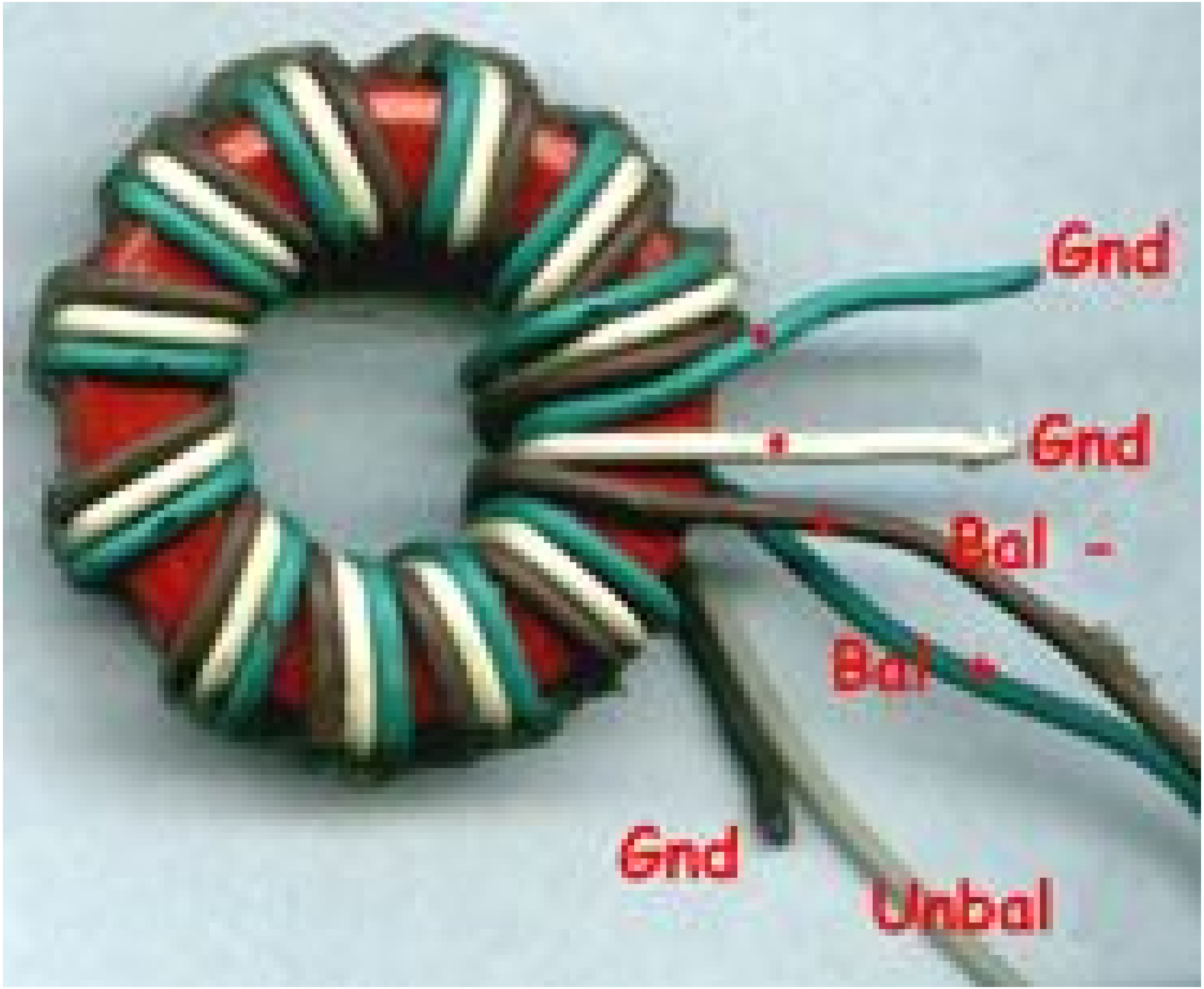
Gnd
Gnd
Bal -
Bal +
Gnd
Unbal

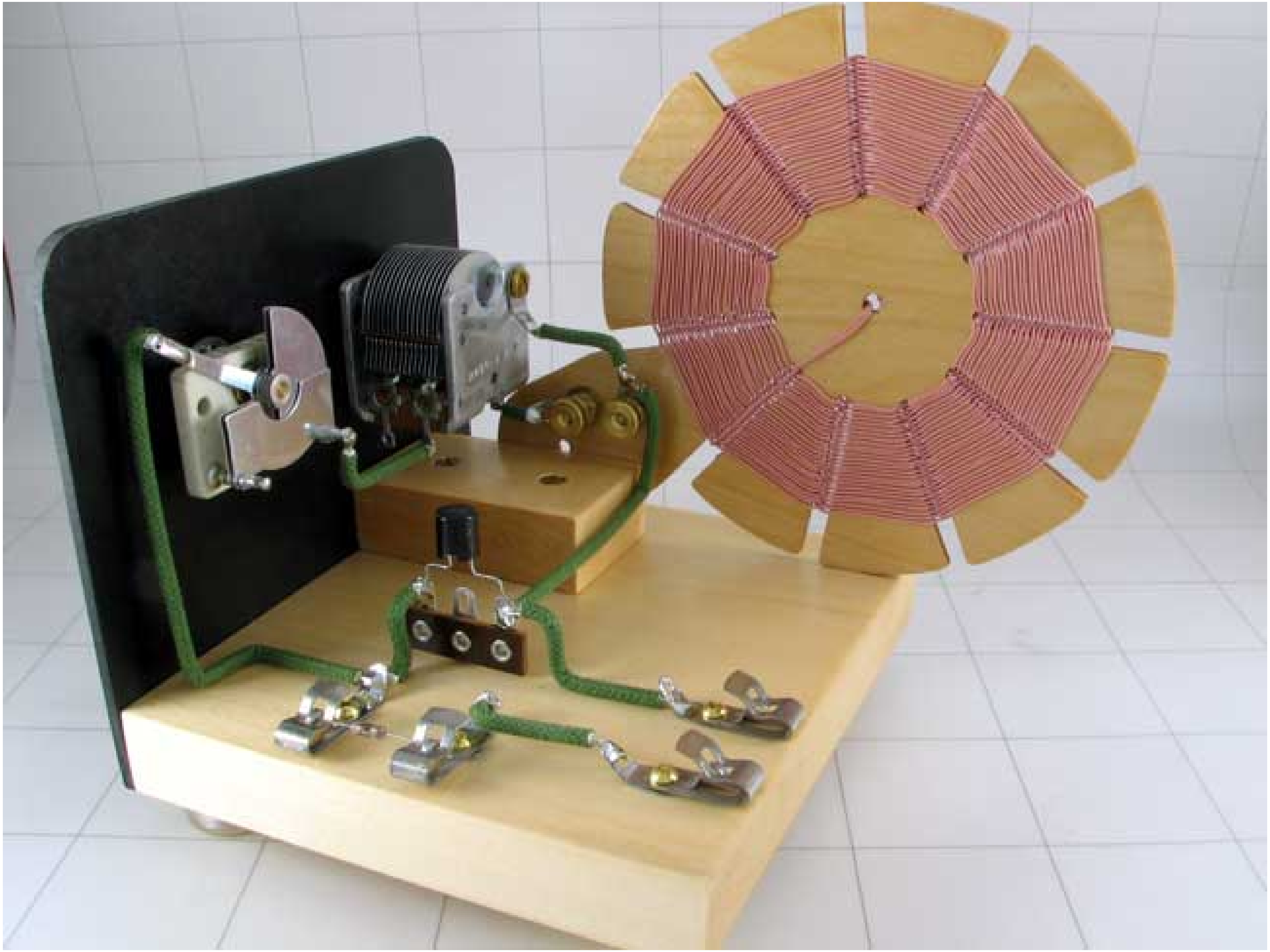

FT
243
FREQ.-KC
10133

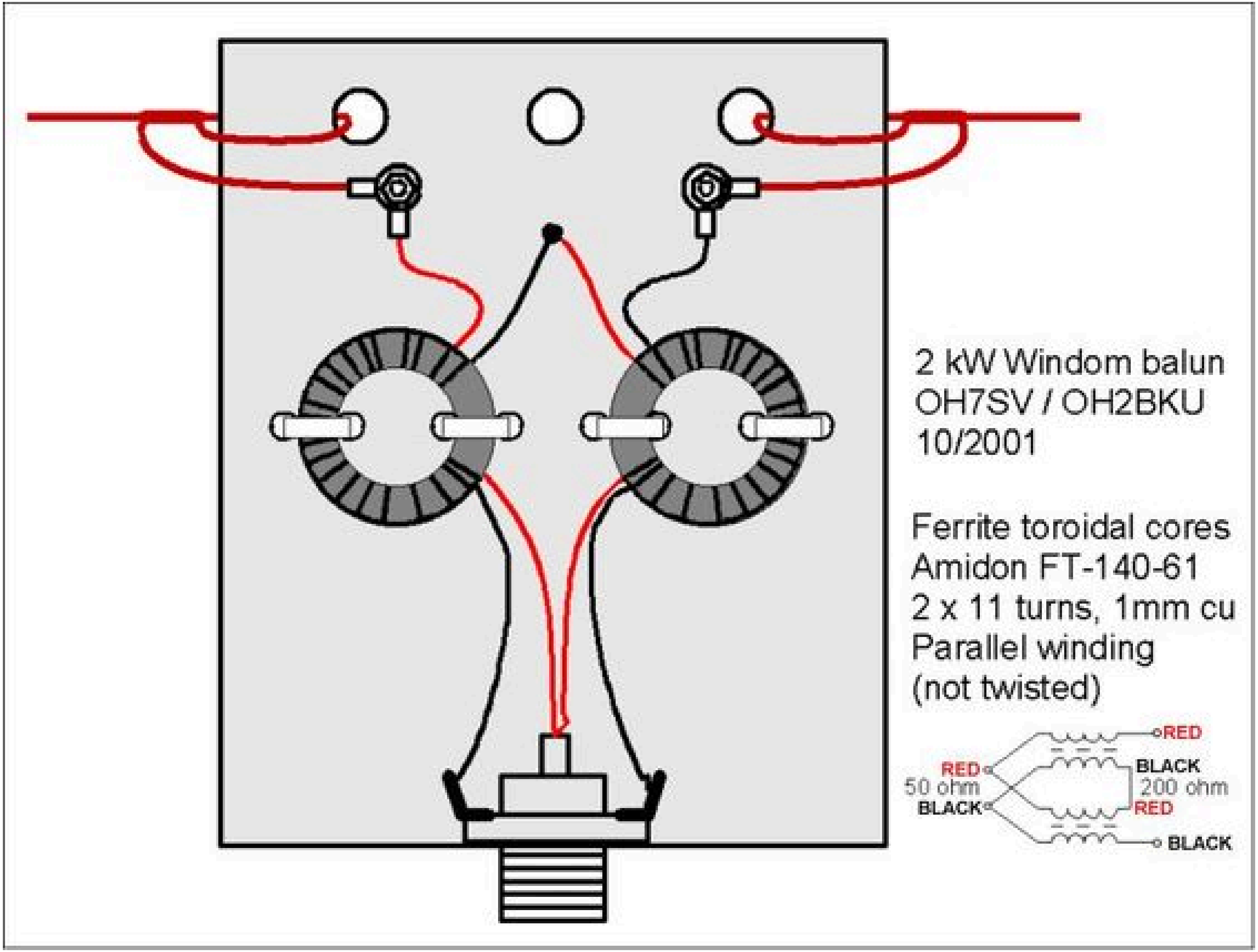

2 kW Windom balun
OH7SV / OH2BKU
10/2001

Ferrite toroidal cores
Amidon FT-140-61
2 x 11 turns, 1mm cu
Parallel winding
(not twisted)

RED
RED
50 ohm
BLACK
BLACK
200 ohm
RED
BLACK

LIMITLESS POWER AND PERPETUAL MOTION

WARNING: THIS WORKS QUITE WELL BUT IS VERY DANGEROUS.

INVENTOR N. TESLA

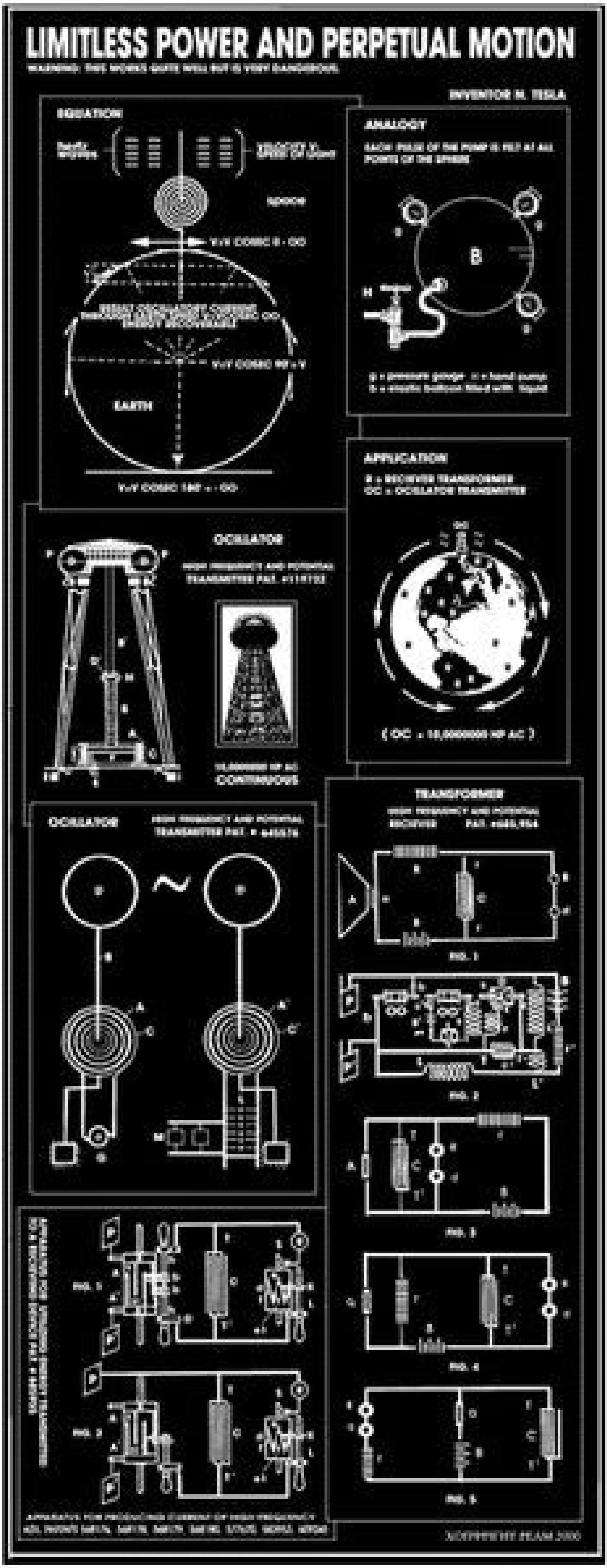
EQUATION
EARTH WAVES
VELOCITY V = SPEED OF LIGHT
space
V=V COSEC 0 = OO
V=V COSEC 90 = V
EARTH
V=V COSEC 180 = OO
ANALOGY
B
H
OCILLATOR
P
APPLICATION
OC
OCILLATOR
~
A
C
M
TRANSFORMER
FIG. 1
FIG. 2
FIG. 3
FIG. 4
FIG. 5
FIG. 1
FIG. 2
C

ANALOGY: EACH PULSE OF THE PUMP IS FELT AT ALL POINTS OF THE SPHERE

g = pressure gauge H = hand pump
B = elastic balloon filled with liquid

APPLICATION

R = RECIEVER TRANSFORMER
OC = OCILLATOR TRANSMITTER

(OC = 10,0000000 HP AC)

OCILLATOR

HIGH FREQUENCY AND POTENTIAL
TRANSMITTER PAT. #1119732

10,000000 HP AC
CONTINUOUS

OCILLATOR

HIGH FREQUENCY AND POTENTIAL
TRANSMITTER PAT. # 645576

TRANSFORMER

HIGH FREQUENCY AND POTENTIAL
RECEIVER PAT. #685,954

APPARATUS FOR PRODUCING CURRENT OF HIGH FREQUENCY
ADL. PATENTS 568176, 568178, 568179, 568180, 577670, 583953, 609245

COPYRIGHT PLAM 2000

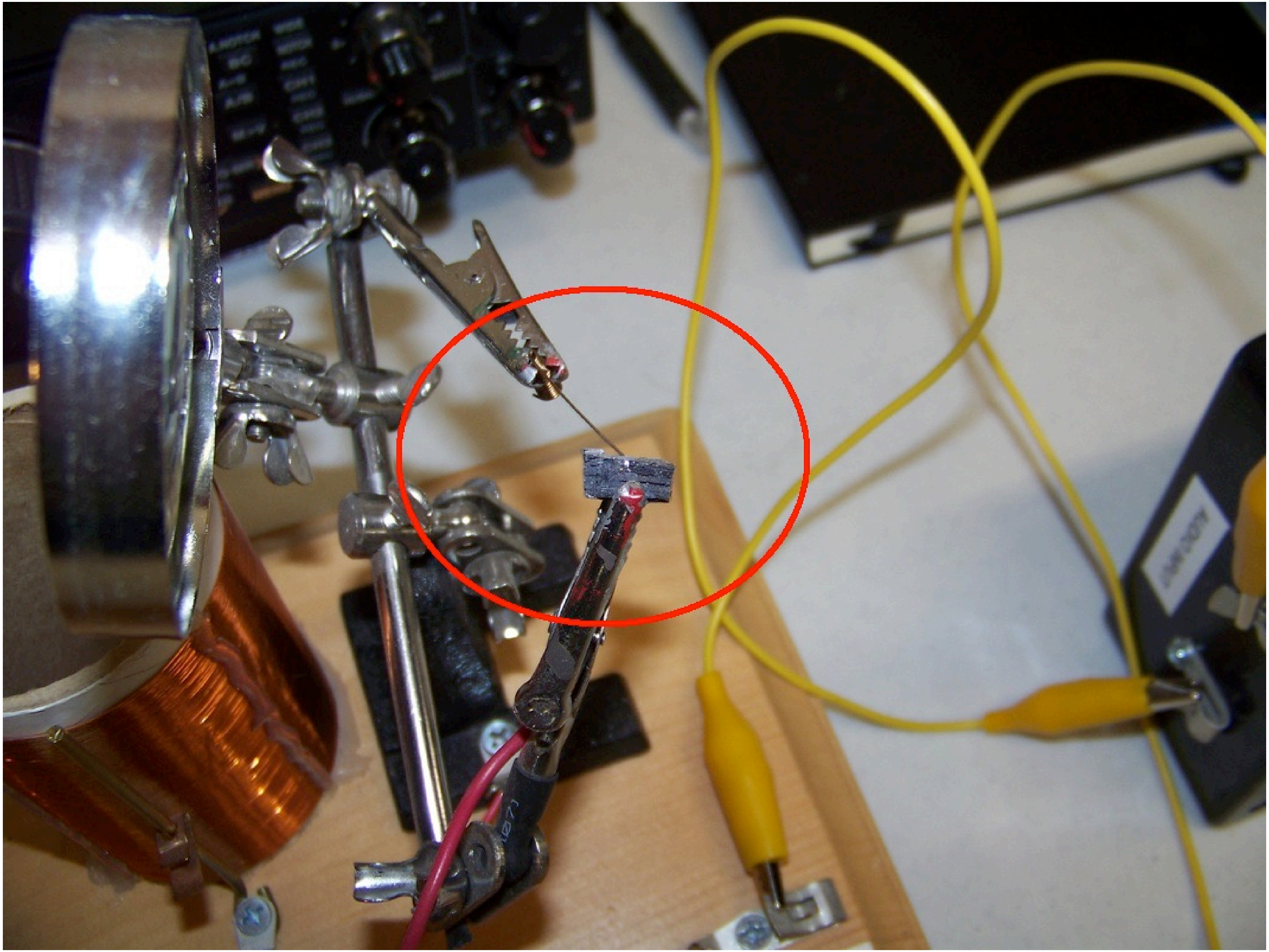

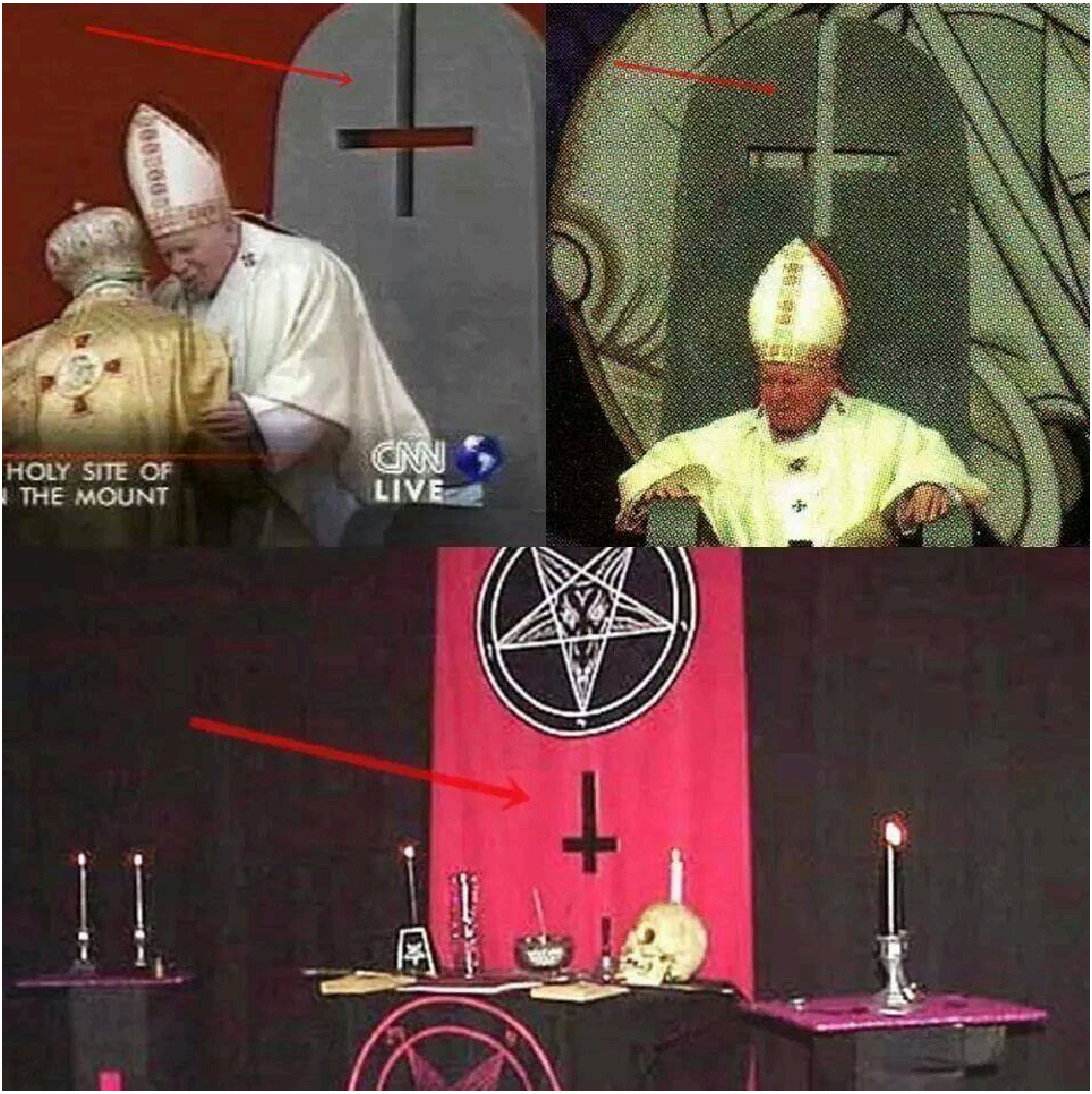
HOLY SITE OF
THE MOUNT
CNN
LIVE

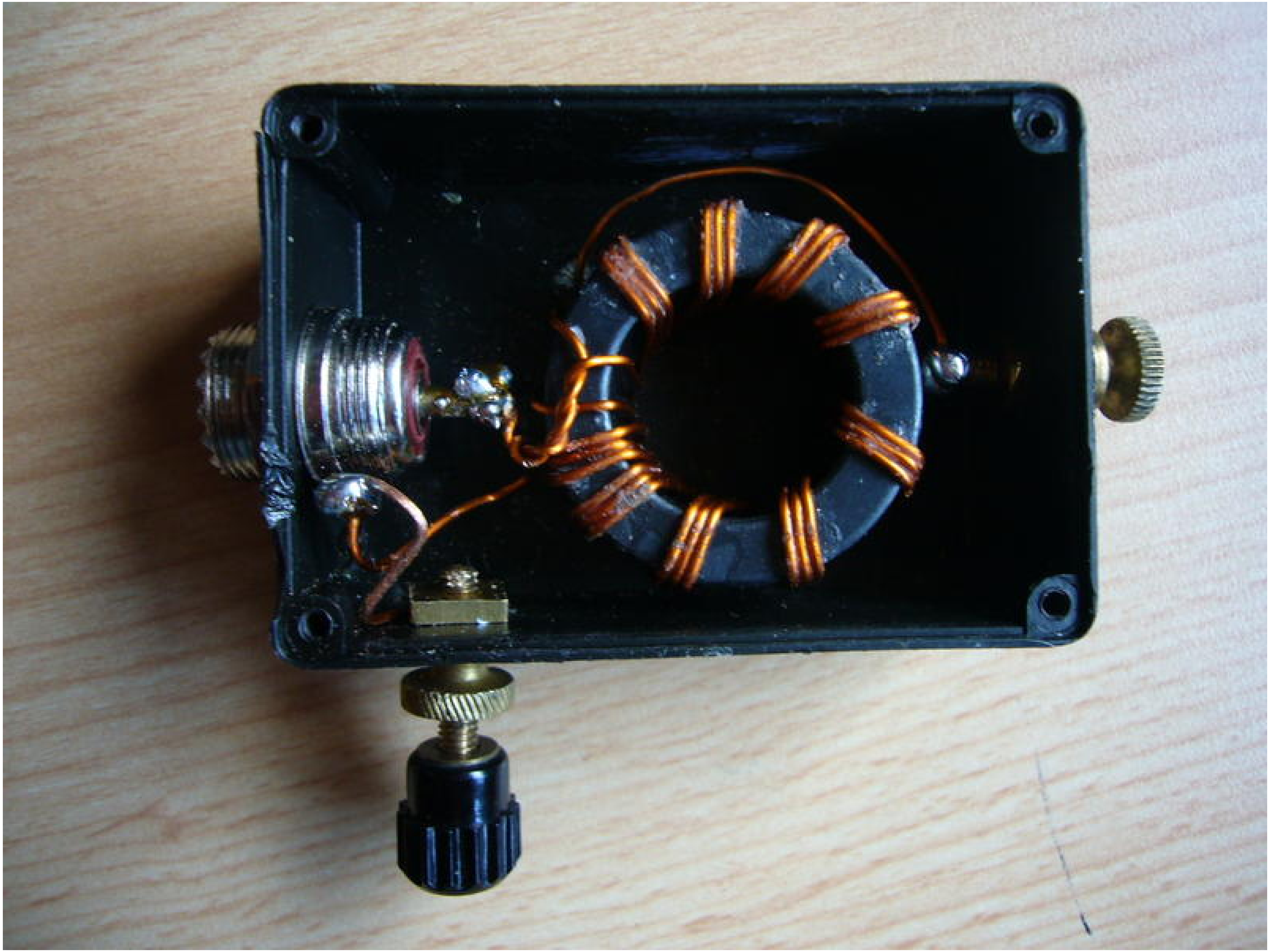

THE COAST OF ANTARCTICA, AND THE ANTARCTIC SHELF.
THIS IS THE ICE WALL, THE CONTAINER OF THE EARTH,
A GIANT LAND MASS SURROUNDING THE EARTH.
WATER ALWAYS LEVELS OUT AND ALWAYS TAKES
THE SHAPE OF ITS CONTAINER. THERE IS NO EDGE.

IF SOMEONE ASKS FOR A
PICTURE OF THE ICE WALL
#FLATEARTHMATTERS
SAR-27
#MOONHOAX
#SPACEHOAX
#GRAVITYHOAX
#RELATIVITYHOAX
#TRUSTFLATWATER
#FLATEARTH
SHOW THEM THIS PICTURE

FISH TANKS, TUBS, POOLS & LAKES OF WATER ARE FLAT AND LEVEL.

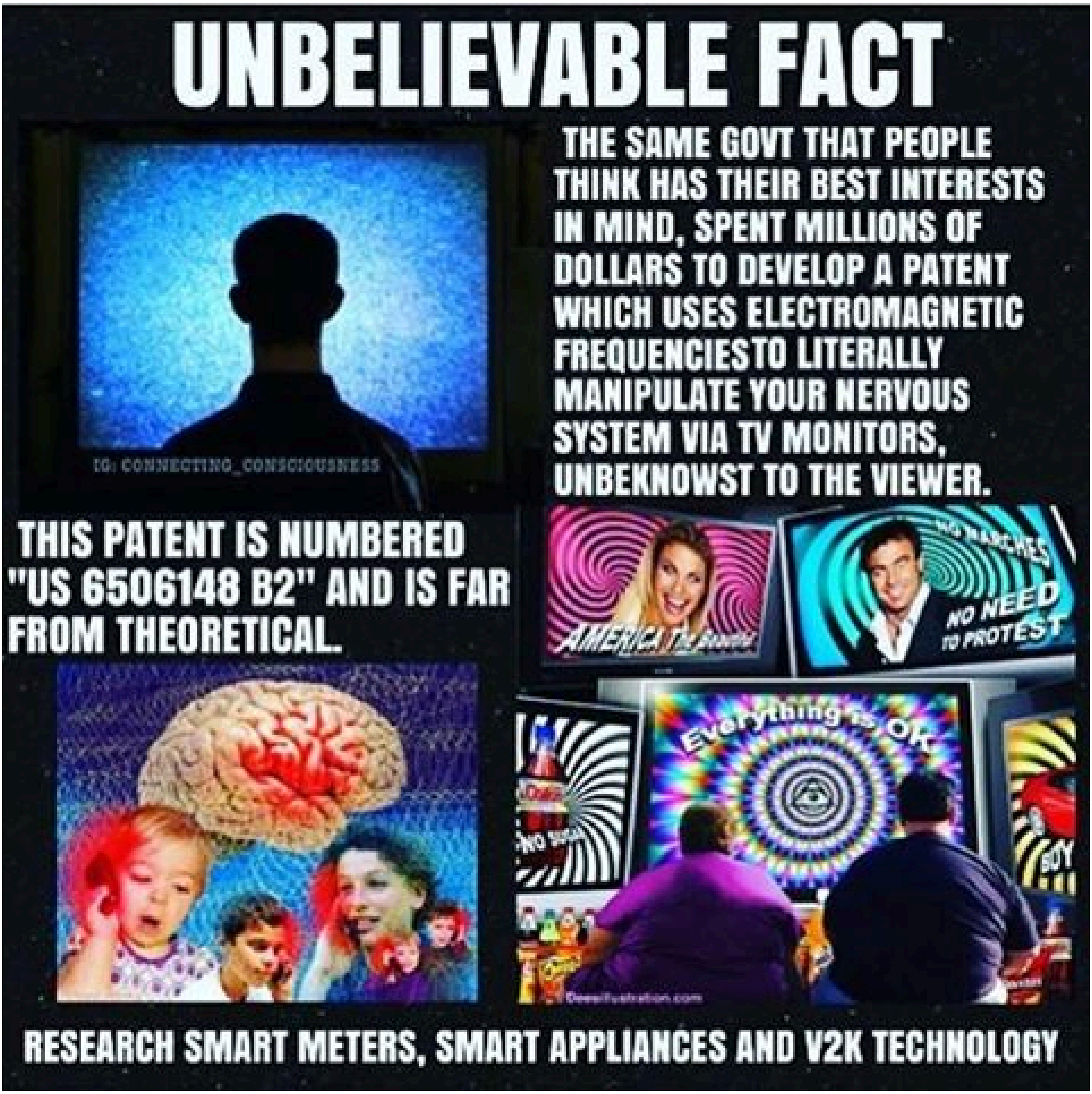

UNBELIEVABLE FACT

THE SAME GOVT THAT PEOPLE THINK HAS THEIR BEST INTERESTS IN MIND, SPENT MILLIONS OF DOLLARS TO DEVELOP A PATENT WHICH USES ELECTROMAGNETIC FREQUENCIESTO LITERALLY MANIPULATE YOUR NERVOUS SYSTEM VIA TV MONITORS, UNBEKNOWST TO THE VIEWER.

THIS PATENT IS NUMBERED "US 6506148 B2" AND IS FAR FROM THEORETICAL.

RESEARCH SMART METERS, SMART APPLIANCES AND V2K TECHNOLOGY

IG: CONNECTING_CONSCIOUSNESS
AMERICA THE BRAVE
NO MARCHES
NO NEED TO PROTEST
Everything is OK
NO
BUY
Doosillustration.com

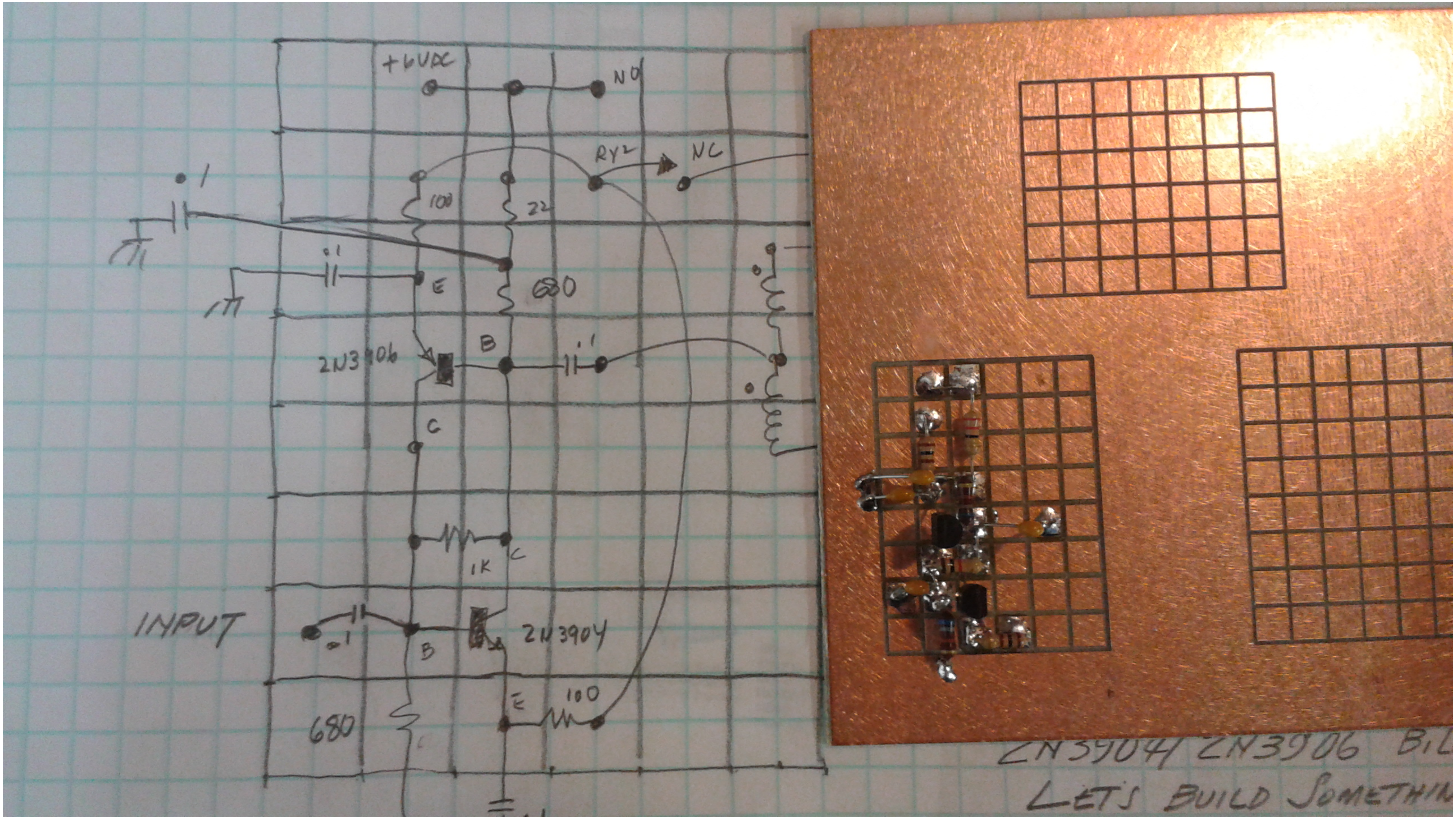
+6VDC
NO
RY2
NC
.1
100
22
.1
E
680
2N3906
B
.1
C
1K
C
INPUT
.01
B
2N3904
E
100
680
2N3904/2N3906 Bi
LET'S BUILD SOMETHIN

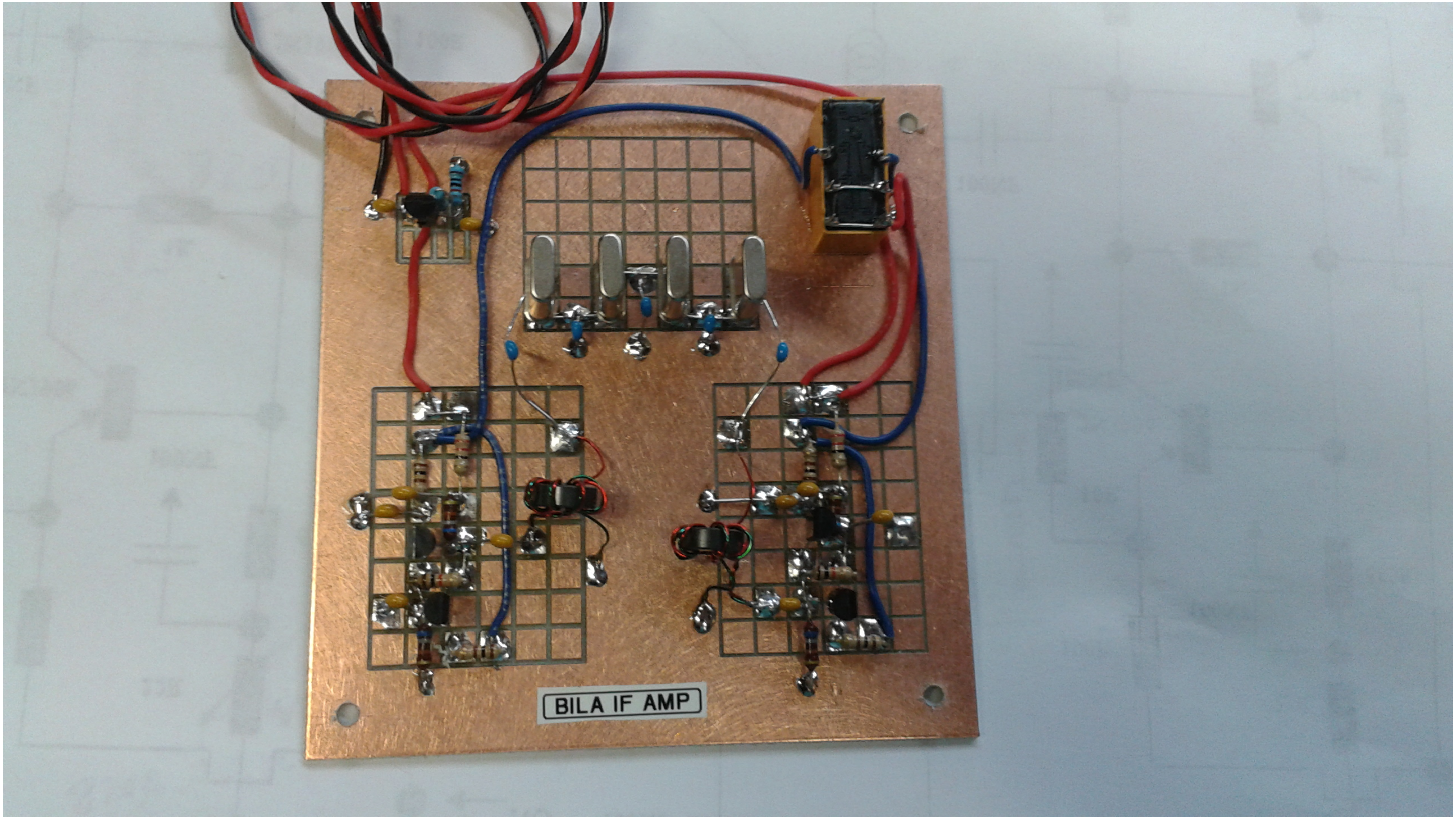
BILA IF AMP

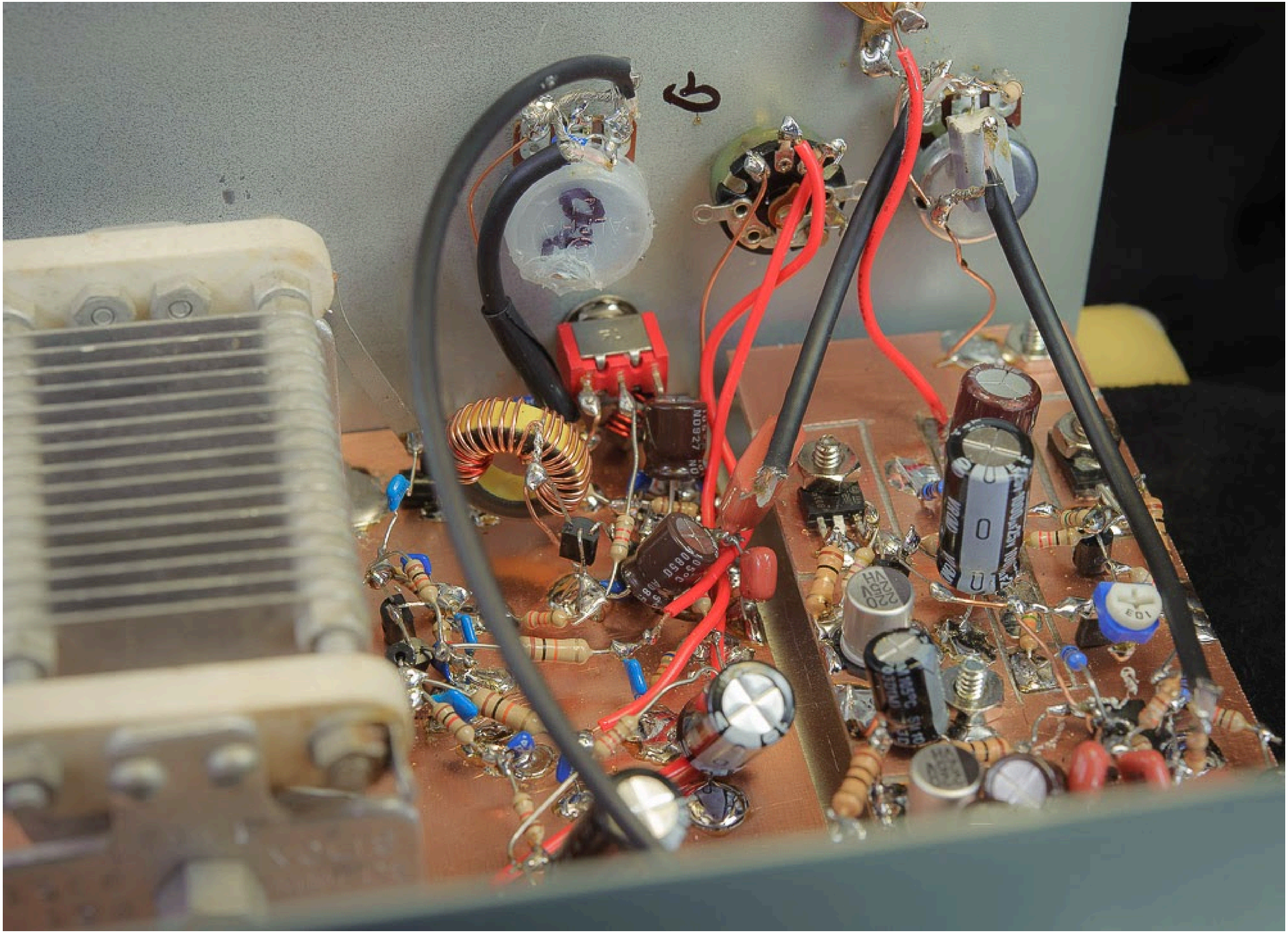

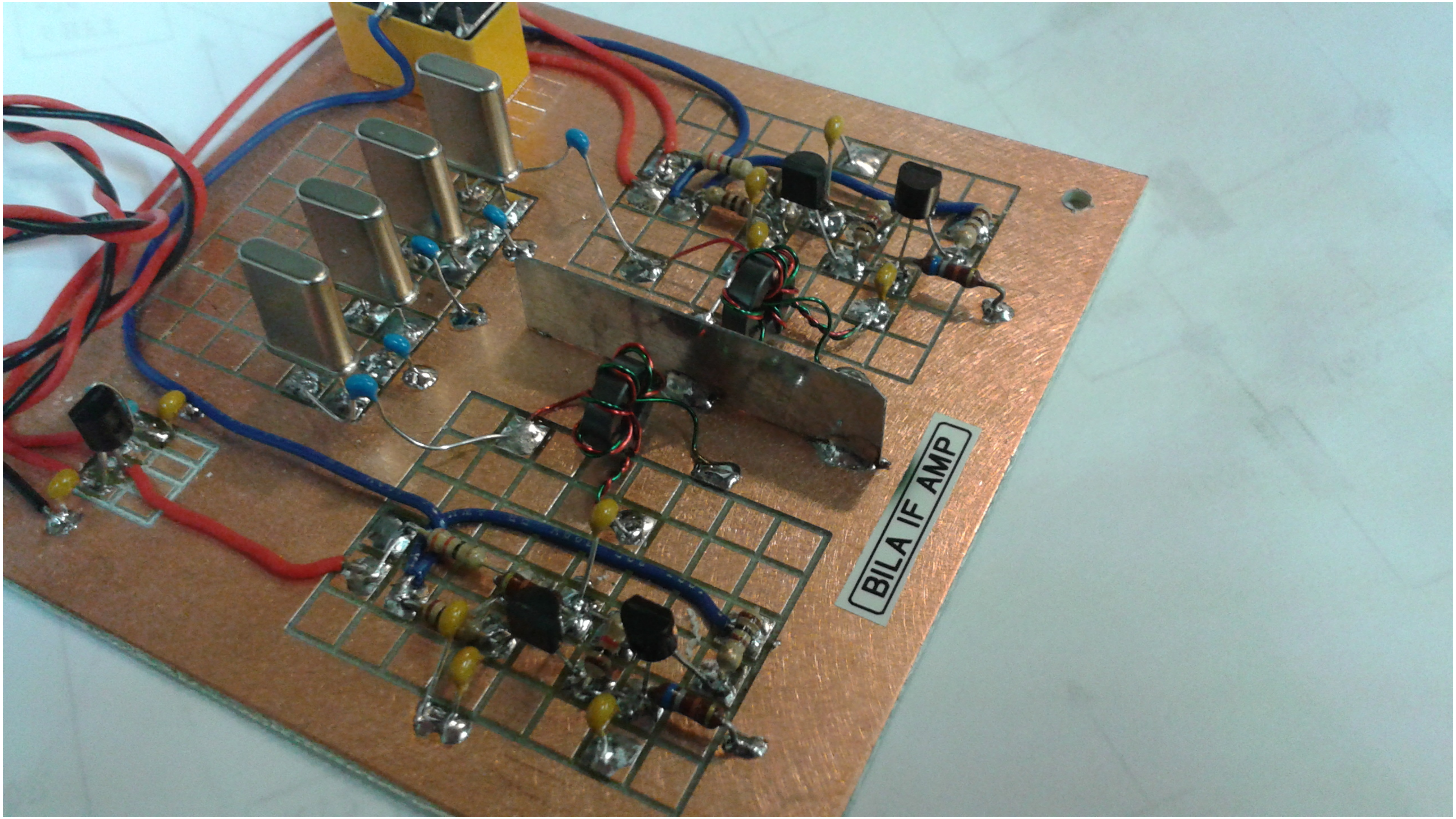
BILA IF AMP

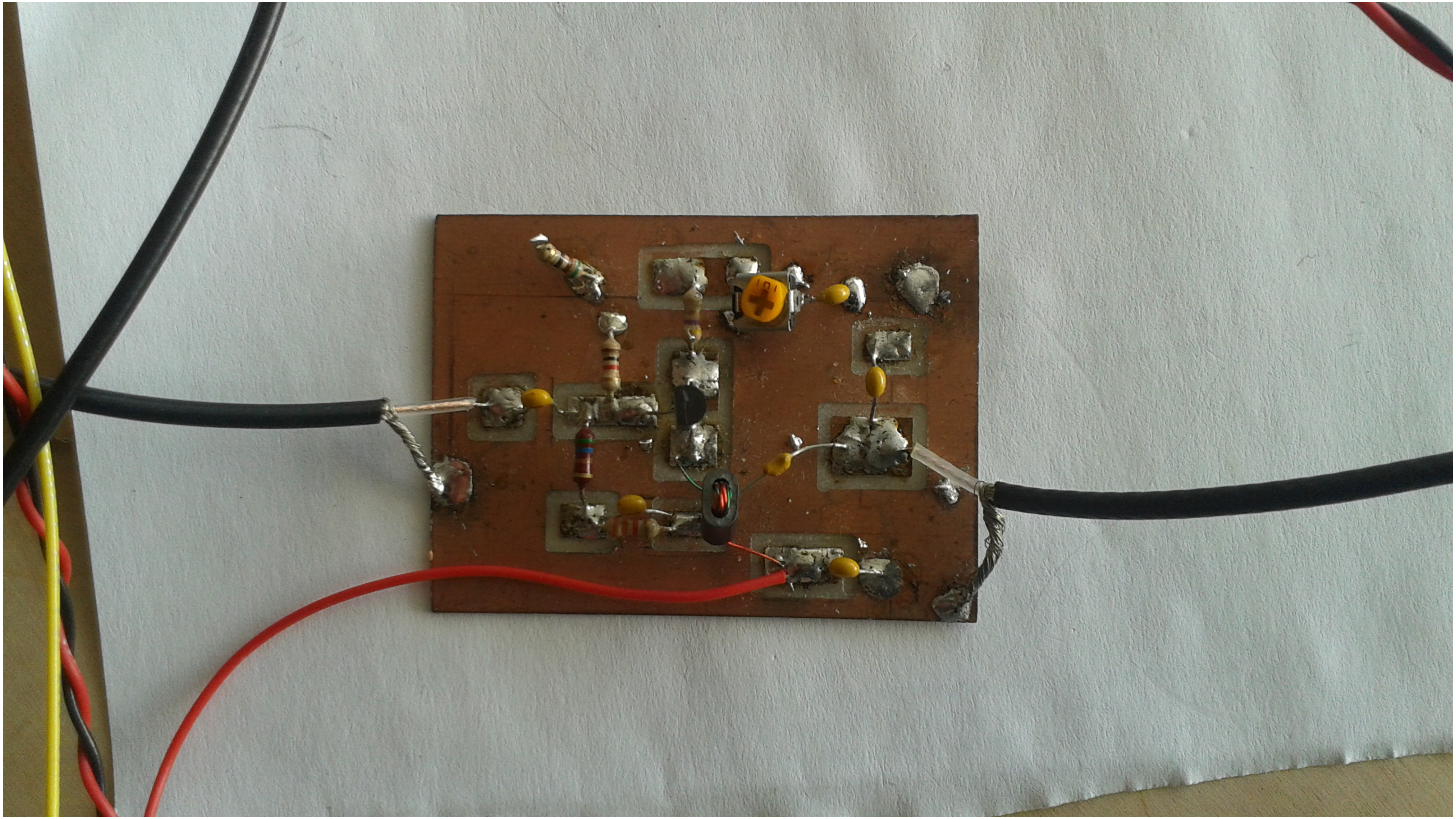
101

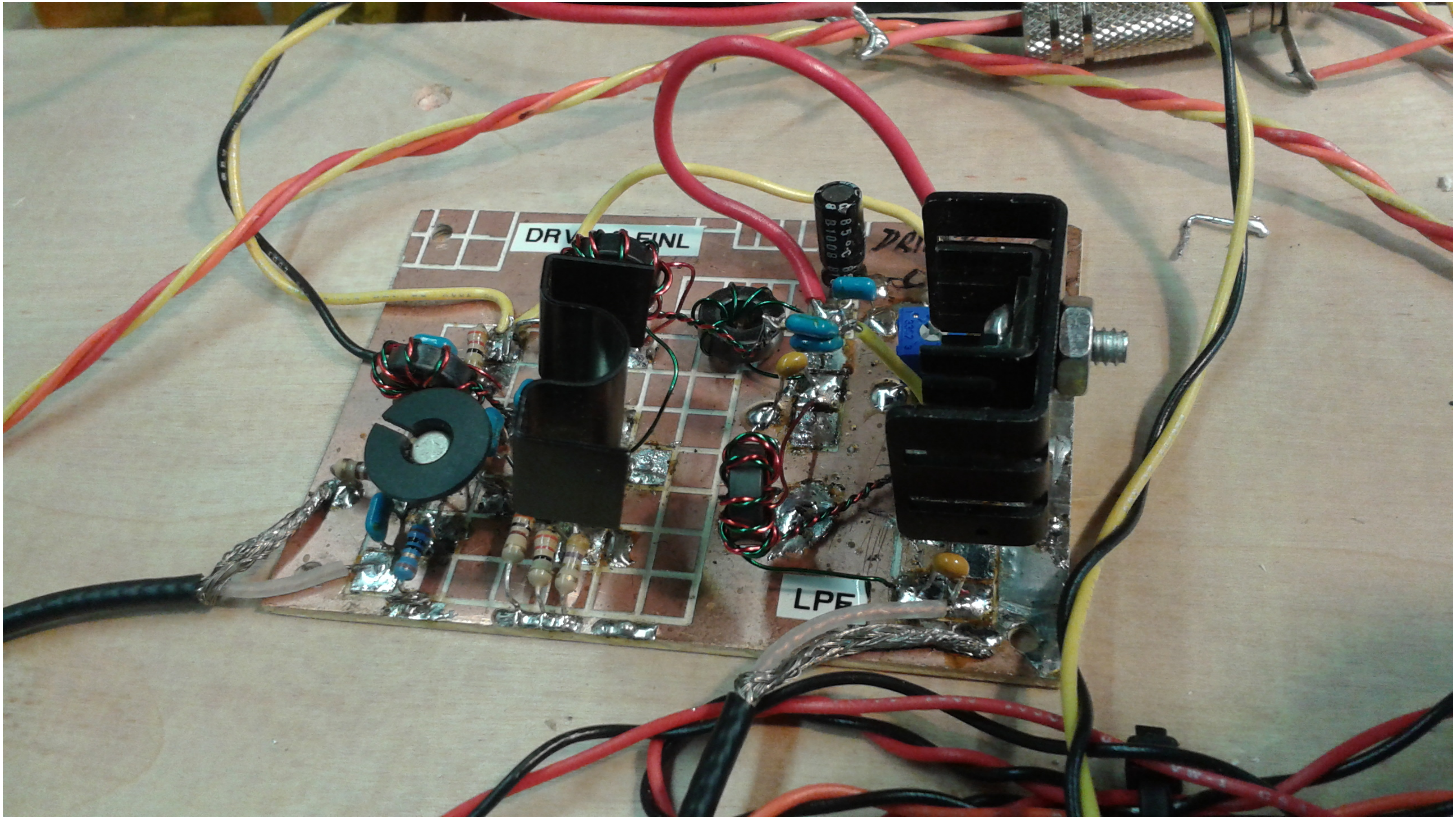
LPF

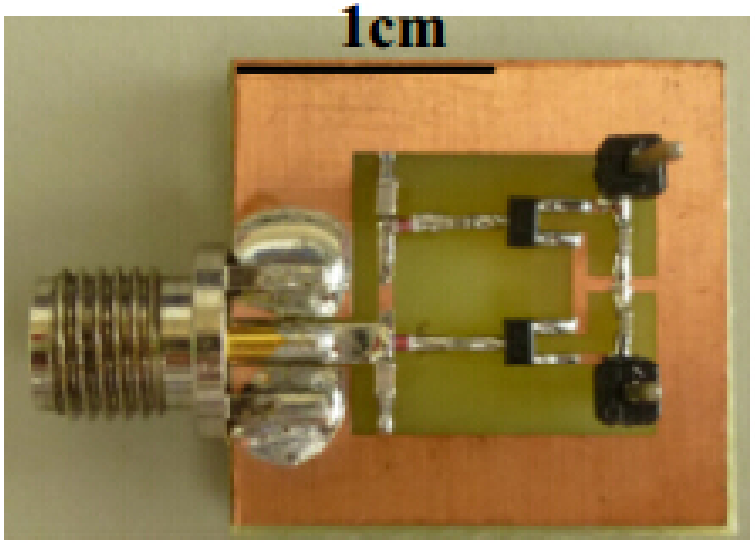
1cm

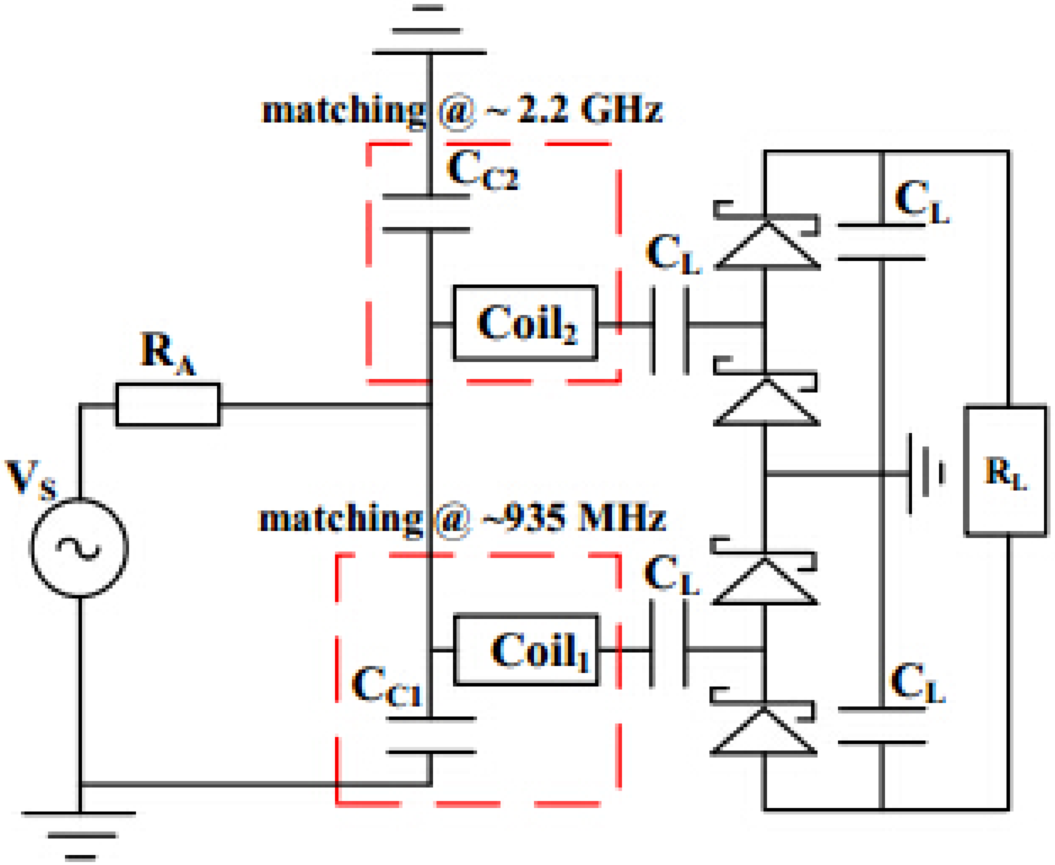
matching @ ~ 2.2 GHz
C_{C2}
Coil$_2$
C_L
C_L
R_A
V_S
R_L
matching @ ~935 MHz
C_L
Coil$_1$
C_{C1}
C_L

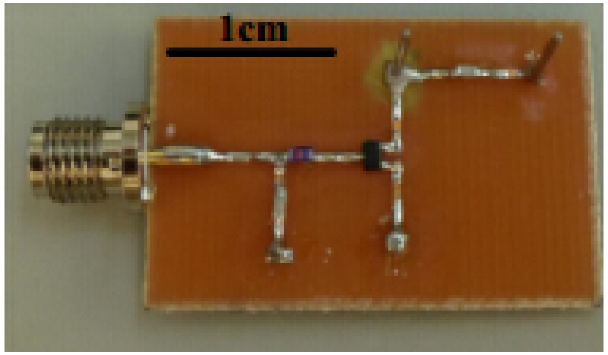
1cm

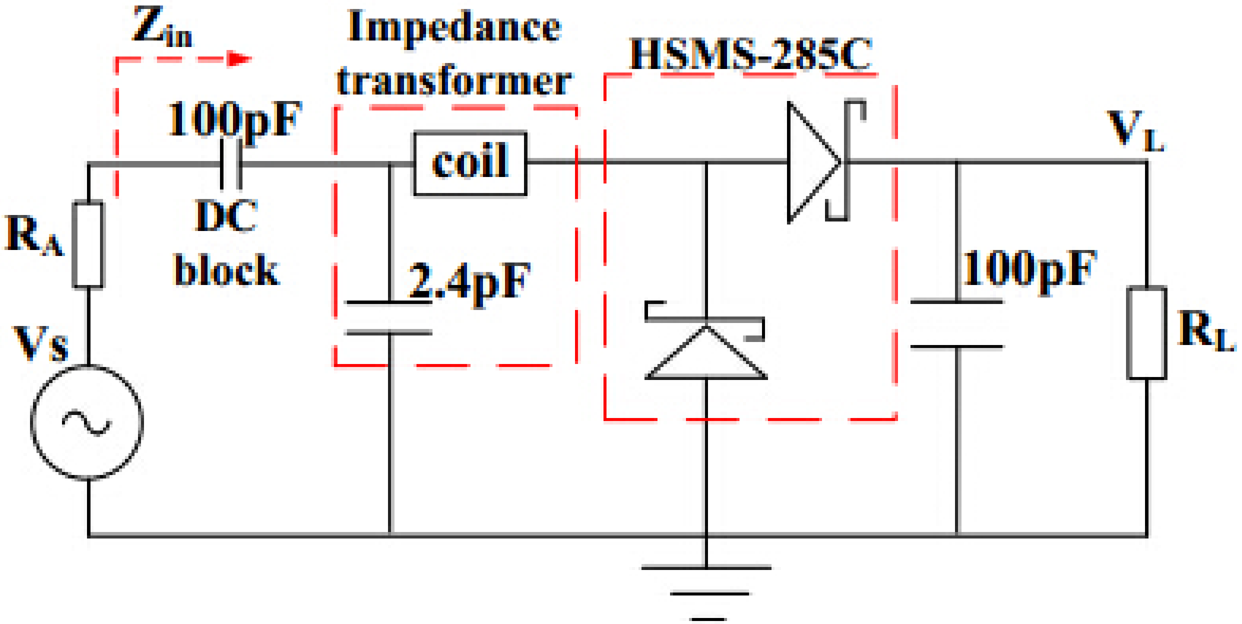
Z$_{in}$
Impedance transformer
HSMS-285C
100pF
DC block
coil
V$_L$
R$_A$
2.4pF
100pF
Vs
R$_L$

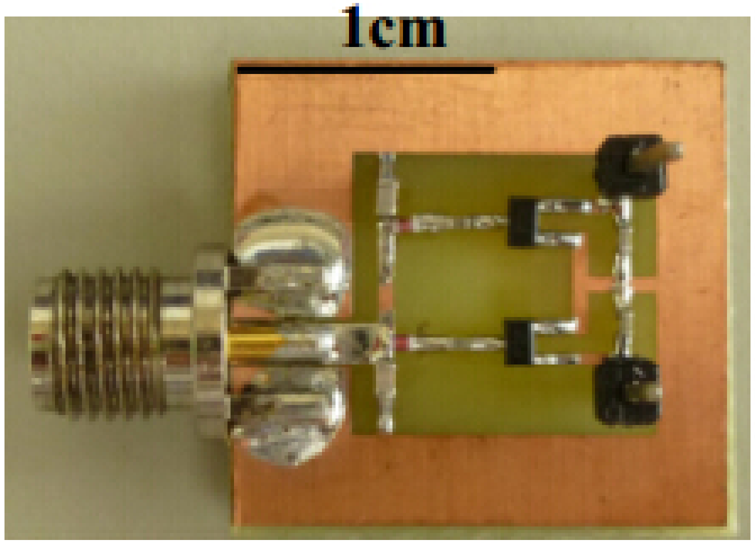
1cm

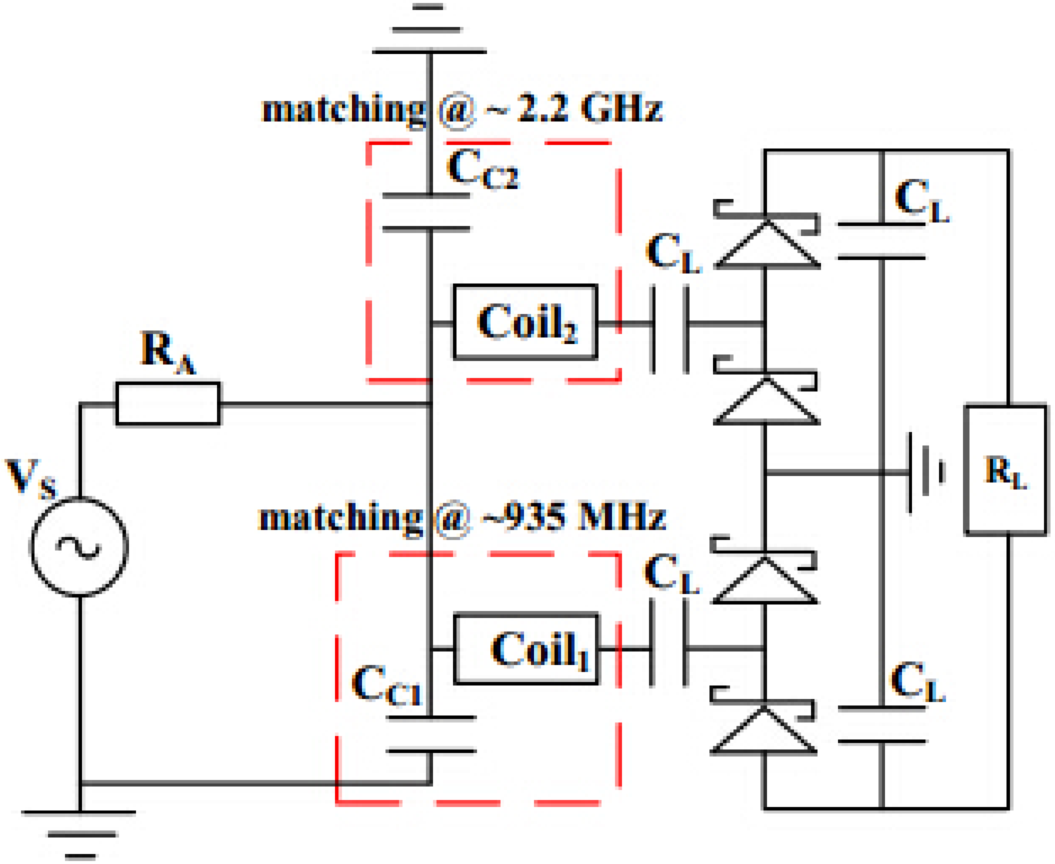
matching @ ~ 2.2 GHz
C_{C2}
C_L
C_L
Coil$_2$
R_A
V_S
R_L
matching @ ~935 MHz
C_L
Coil$_1$
C_{C1}
C_L

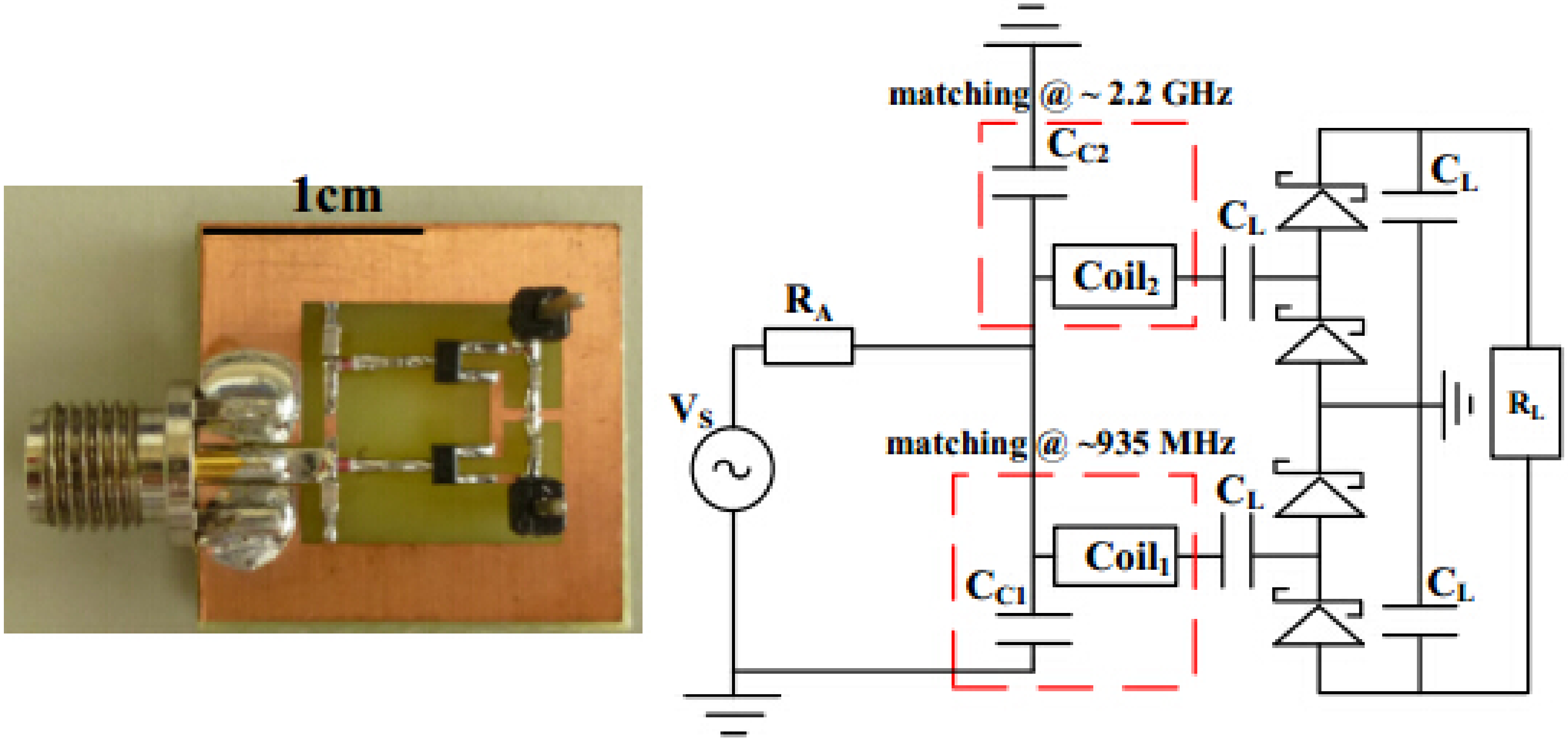
1cm
matching @ ~ 2.2 GHz
C_C2
Coil_2
C_L
C_L
R_A
V_S
R_L
matching @ ~935 MHz
C_L
Coil_1
C_C1
C_L

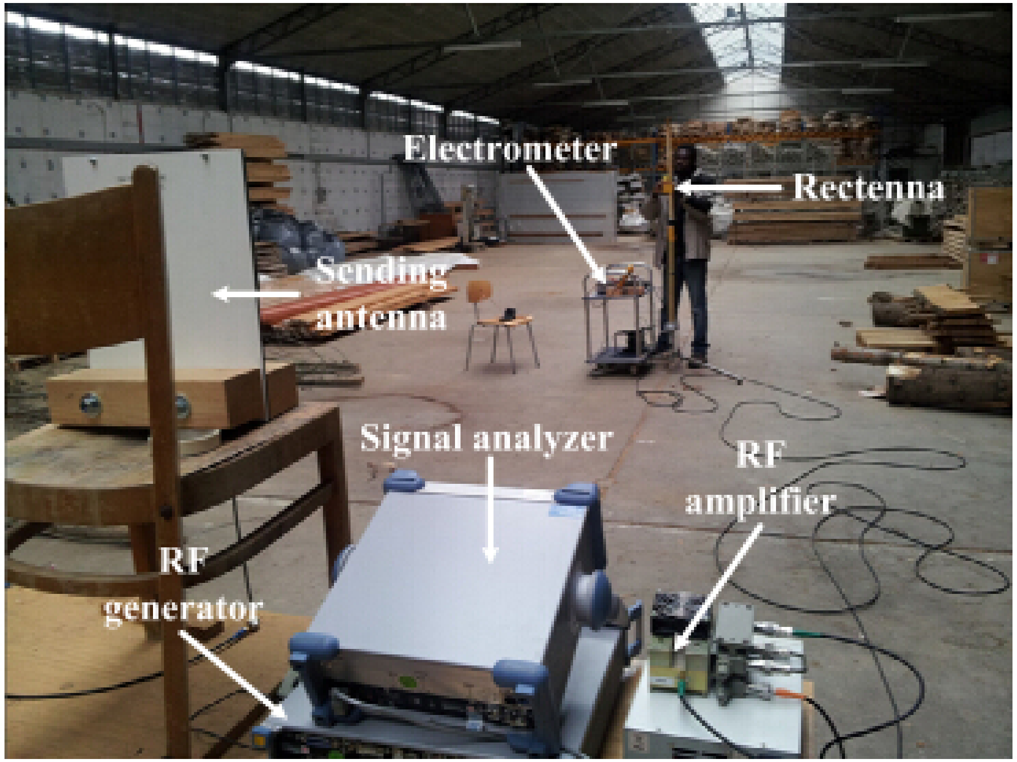
Electrometer
Rectenna
Sending antenna
Signal analyzer
RF amplifier
RF generator

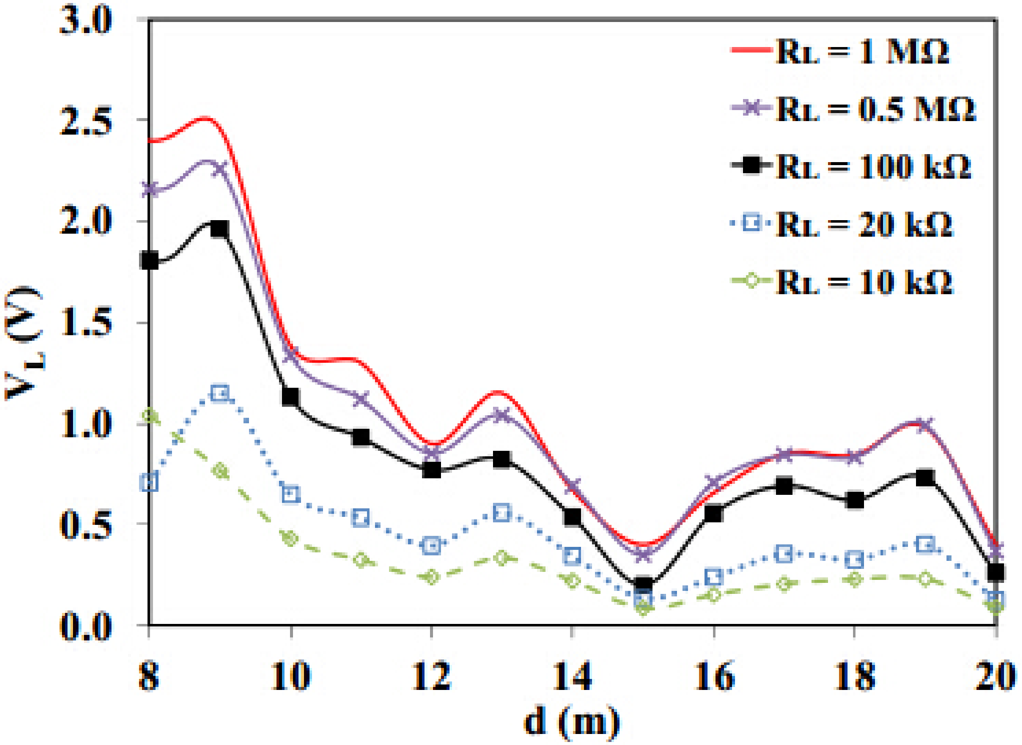
3.0
2.5
2.0
1.5
1.0
0.5
0.0
V_L (V)
8
10
12
14
16
18
20
d (m)
RL = 1 MΩ
RL = 0.5 MΩ
RL = 100 kΩ
RL = 20 kΩ
RL = 10 kΩ

28.9 °C
26%
°C / °F
MAX / MIN
MEM CLEAR

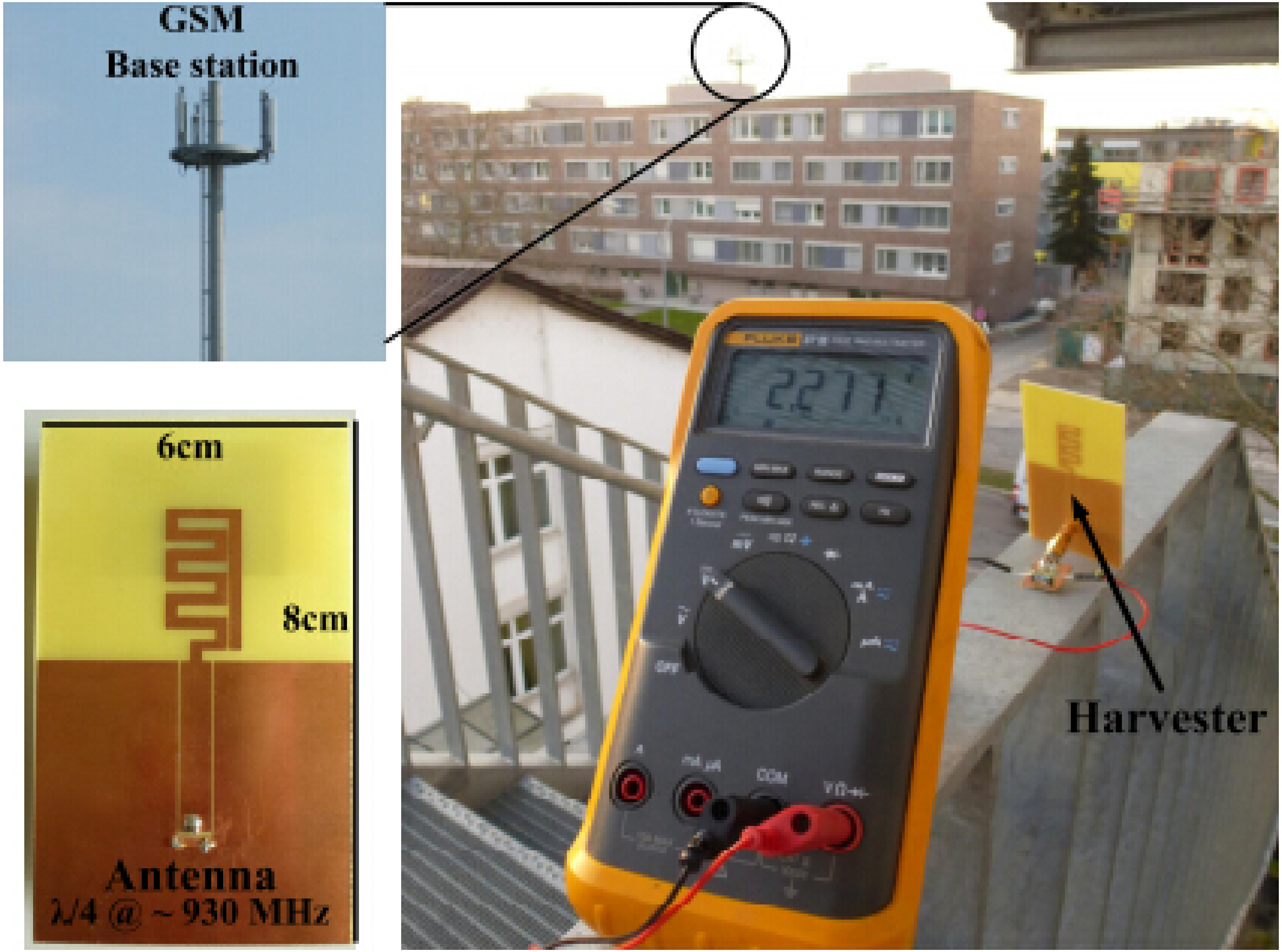
GSM
Base station
6cm
8cm
Antenna
λ/4 @ ~ 930 MHz
Harvester

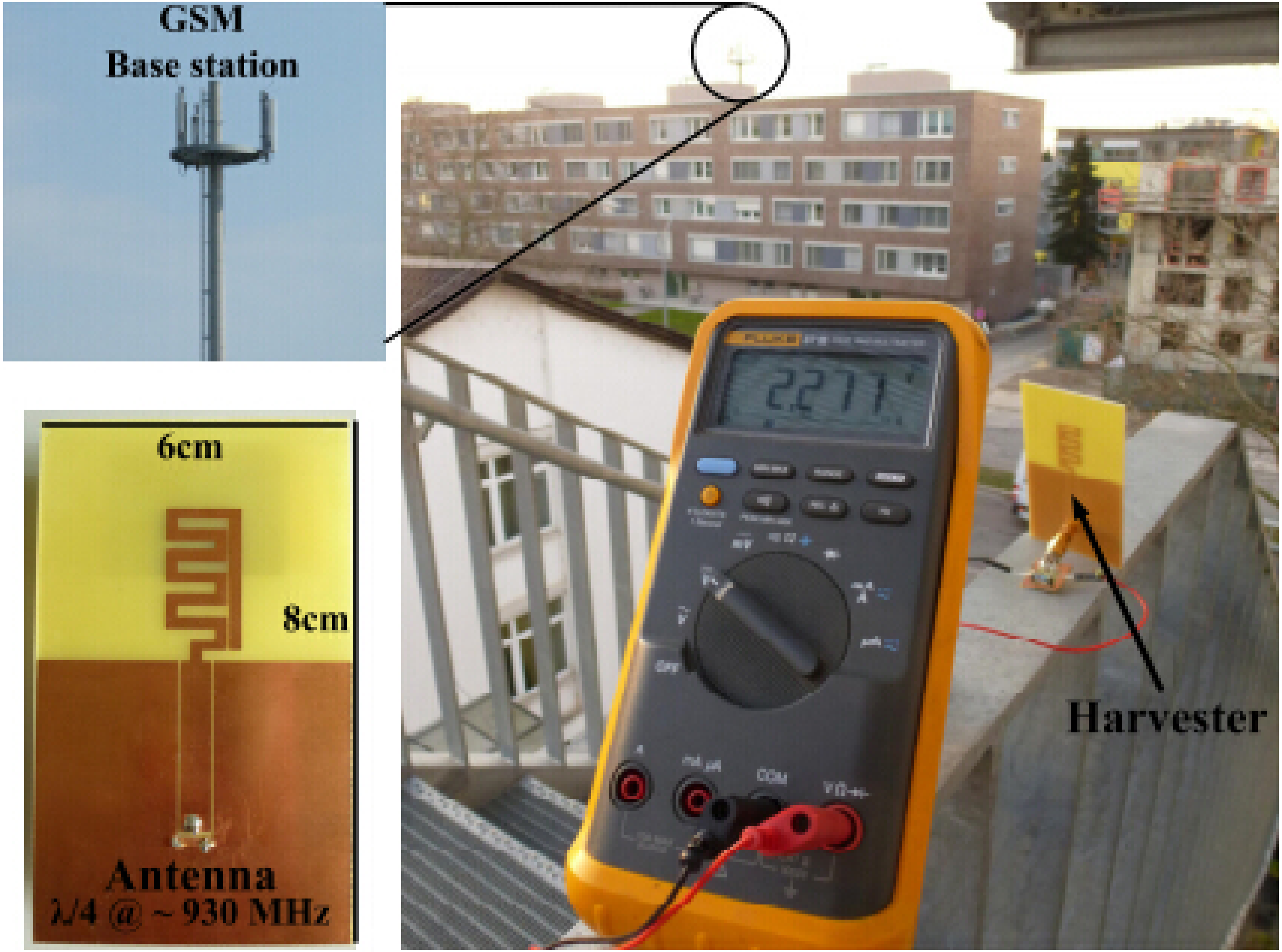
GSM
Base station
6cm
8cm
Antenna
λ/4 @ ~ 930 MHz
2277
Harvester

Seed Sounds in the Chakra Petals–Chakra Petal Bijas

Creation occurred due to the first sound Aum. The universe and everything in it is condensed energy, that energy is vibrating and all has sound. The body itself is sound, called sabda in Sanskrit. The body can be recreated and regenerated with sound as many modern technologies are now tapping into DNA Repair using sound waves. In the ancient Puranas (sacred story texts), there are elaborate stories of the creation of the world, this is called Shristi Karana. Meditating on the bijas (seed sounds) of the chakra petals is also called Shristi Karana, for it is the recreation of the body through sabda (sound).

Dan Mirahorian

Copyright © 2008
All Rights Reserved

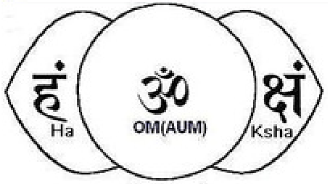
हं
Ha
ॐ
OM(AUM)
क्षं
Ksha

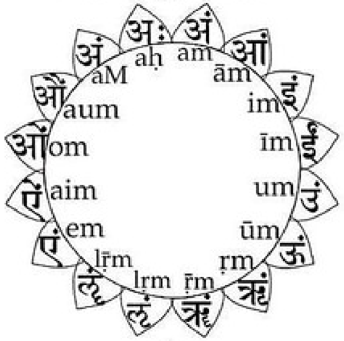
अः
aḥ
अं
aṁ
आं
ām
इं
im
ईं
īm
उं
um
ऊं
ūm
ऋं
ṛm
ॠं
ṝm
लृं
lṛm
ॡं
lṝm
एं
em
ऐं
aim
ओं
om
औं
aum
अं
aM

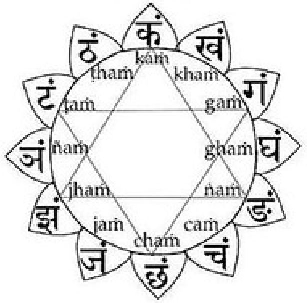
कं
kaṁ
खं
khaṁ
गं
gaṁ
घं
ghaṁ
ङं
ṅaṁ
चं
caṁ
छं
chaṁ
जं
jaṁ
झं
jhaṁ
ञं
ñaṁ
टं
ṭaṁ
ठं
ṭhaṁ

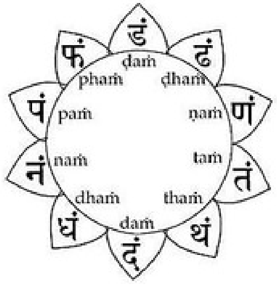
डं
ḍaṁ
ढं
ḍhaṁ
णं
ṇaṁ
तं
taṁ
थं
thaṁ
दं
daṁ
धं
dhaṁ
नं
naṁ
पं
paṁ
फं
phaṁ

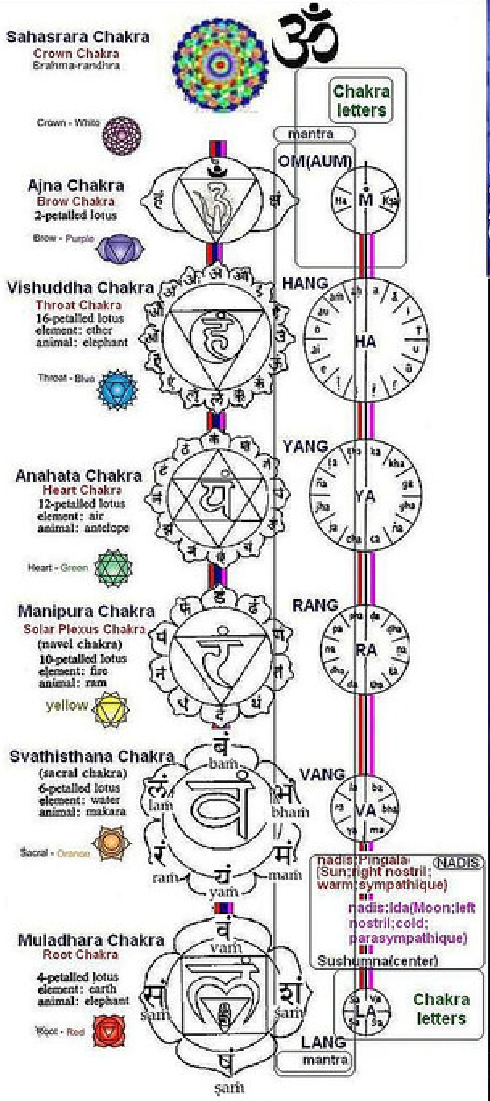
Sahasrara Chakra
Crown Chakra
Brahma-randhra
Crown - White
Chakra letters
mantra
OM(AUM)
Ha
M
Ksa
Ajna Chakra
Brow Chakra
2-petalled lotus
Brow - Purple
Vishuddha Chakra
Throat Chakra
16-petalled lotus
element: ether
animal: elephant
Throat - Blue
HANG
HA
Anahata Chakra
Heart Chakra
12-petalled lotus
element: air
animal: antelope
Heart - Green
YANG
YA
Manipura Chakra
Solar Plexus Chakra
(navel chakra)
10-petalled lotus
element: fire
animal: ram
yellow
RANG
RA
Svathisthana Chakra
(sacral chakra)
6-petalled lotus
element: water
animal: makara
Sacral - Orange
baṁ
bhaṁ
maṁ
yaṁ
raṁ
laṁ
VANG
VA
nadis:Pingala (Sun;right nostril; warm;sympathique)
NADIS
nadis:Ida(Moon;left nostril;cold; parasympathique)
Sushumna(center)
Muladhara Chakra
Root Chakra
4-petalled lotus
element: earth
animal: elephant
Root - Red
vaṁ
śaṁ
ṣaṁ
saṁ
LANG
LA
mantra
Chakra letters

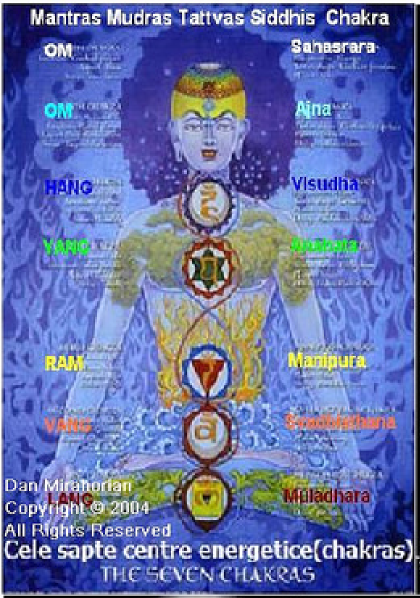
Mantras Mudras Tattvas Siddhis Chakra
OM
Sahasrara
OM
Ajna
HANG
Visudha
YANG
Anahata
RAM
Manipura
VANG
Svadhisthana
LANG
Muladhara
Dan Mirahorian
Copyright © 2004
All Rights Reserved
Cele sapte centre energetice(chakras)
THE SEVEN CHAKRAS

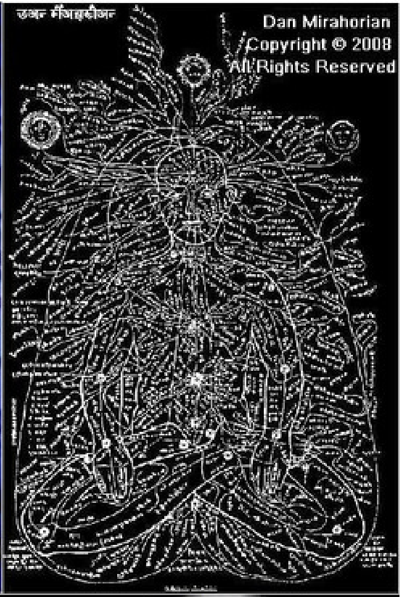
Dan Mirahorian
Copyright © 2008
All Rights Reserved

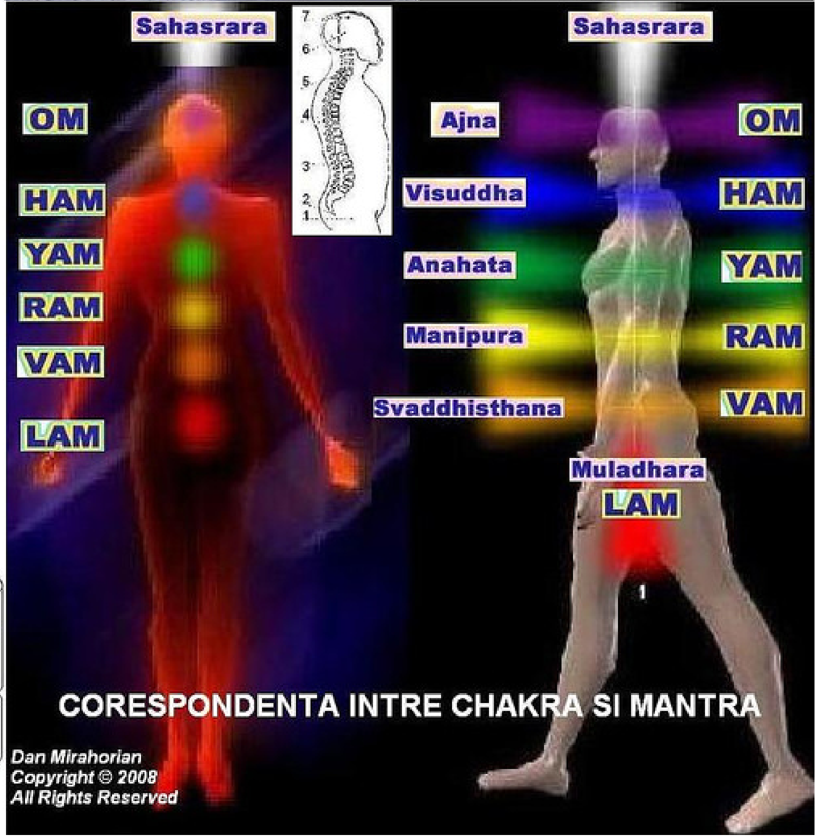
Sahasrara
Sahasrara
OM
HAM
YAM
RAM
VAM
LAM
7
6
5
4
3
2
1
Ajna
Visuddha
Anahata
Manipura
Svaddhisthana
Muladhara
LAM
OM
HAM
YAM
RAM
VAM
CORESPONDENTA INTRE CHAKRA SI MANTRA
Dan Mirahorian
Copyright © 2008
All Rights Reserved

3.74
OFF
ON
CEN-TECH
DIGITAL
MULTIMETER

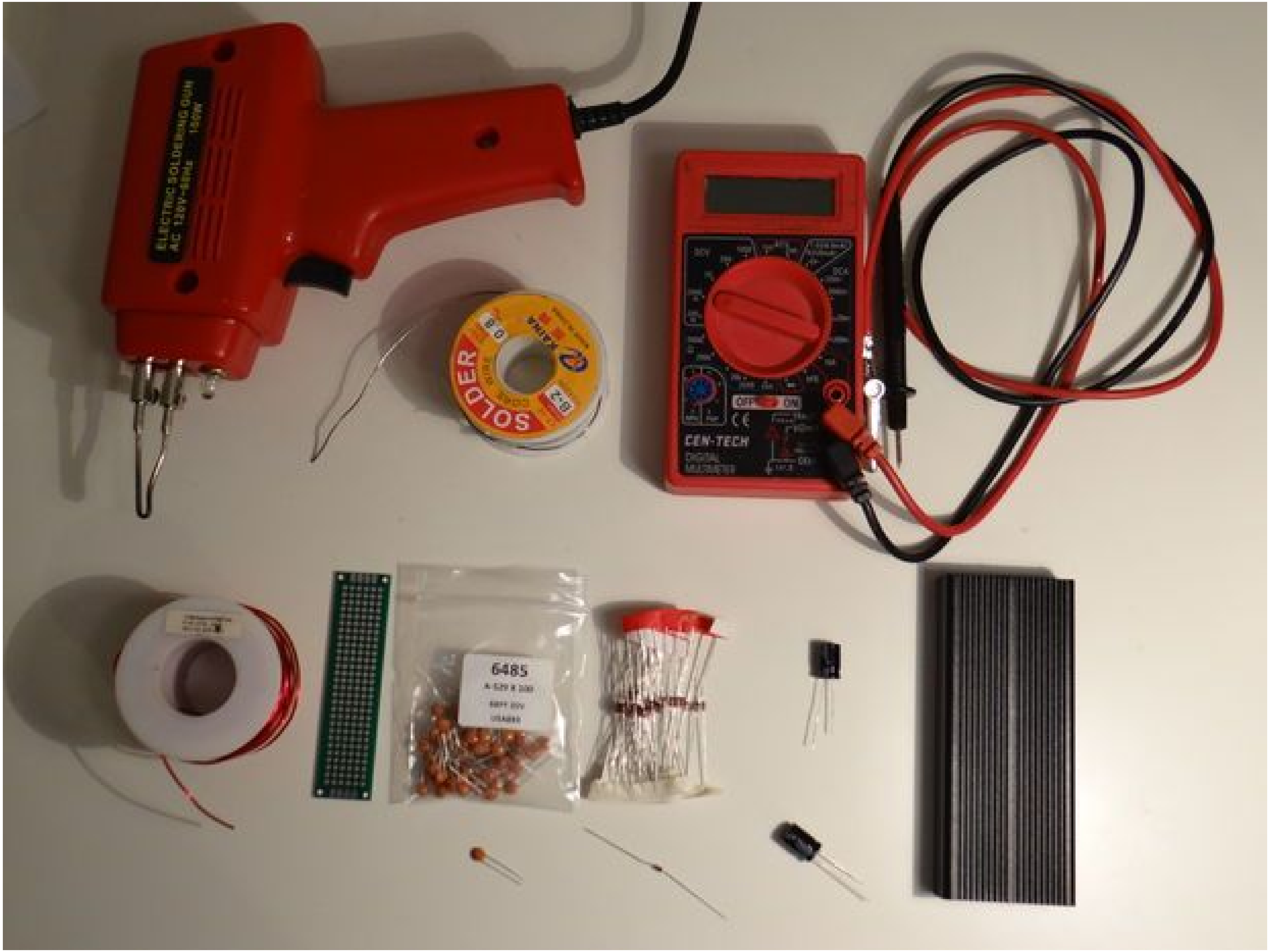
ELECTRIC SOLDERING GUN
AC 120V-60Hz 150W
SOLDER
0.8
CEN-TECH
DIGITAL MULTIMETER
OFF
ON
6485

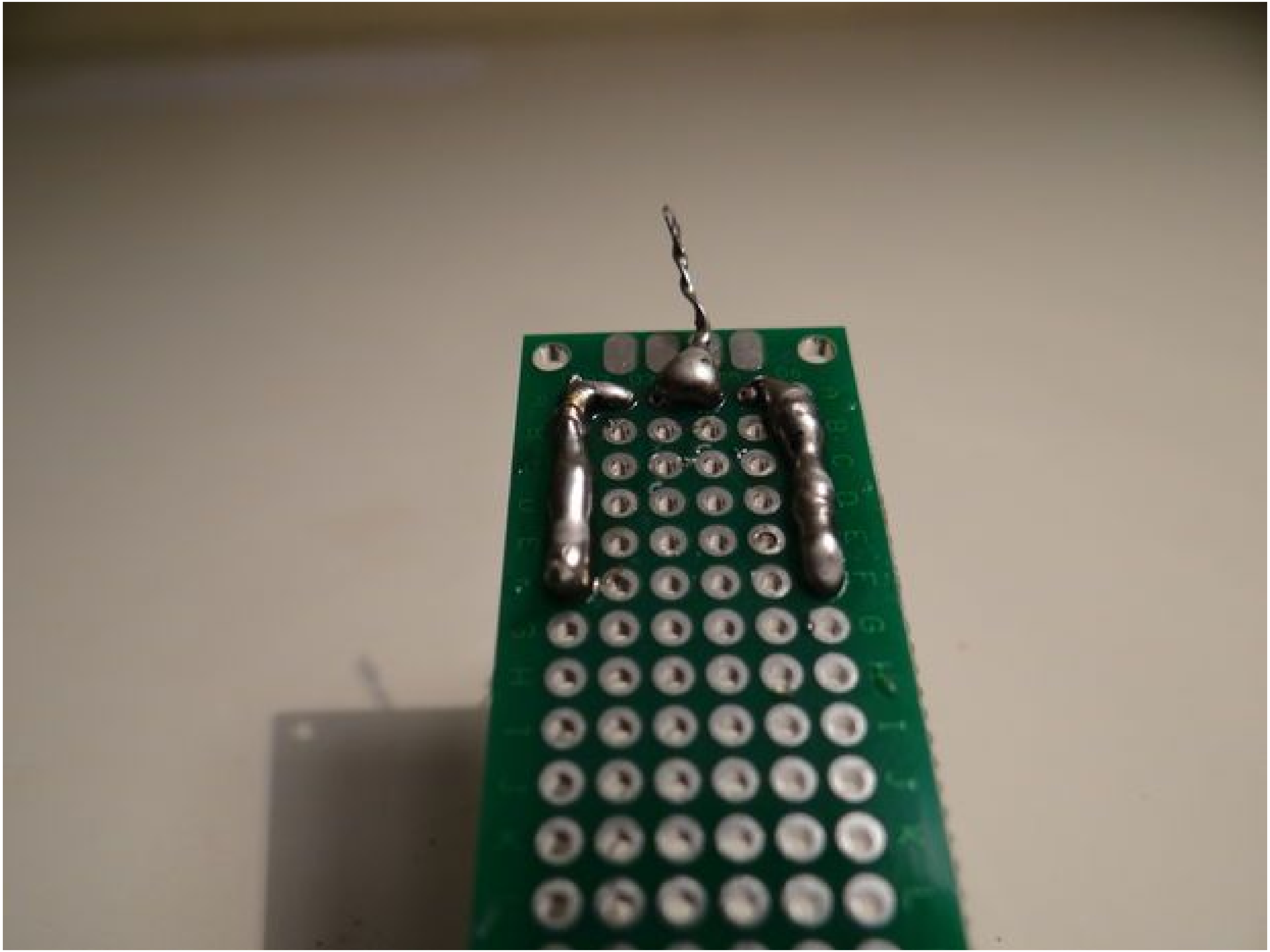

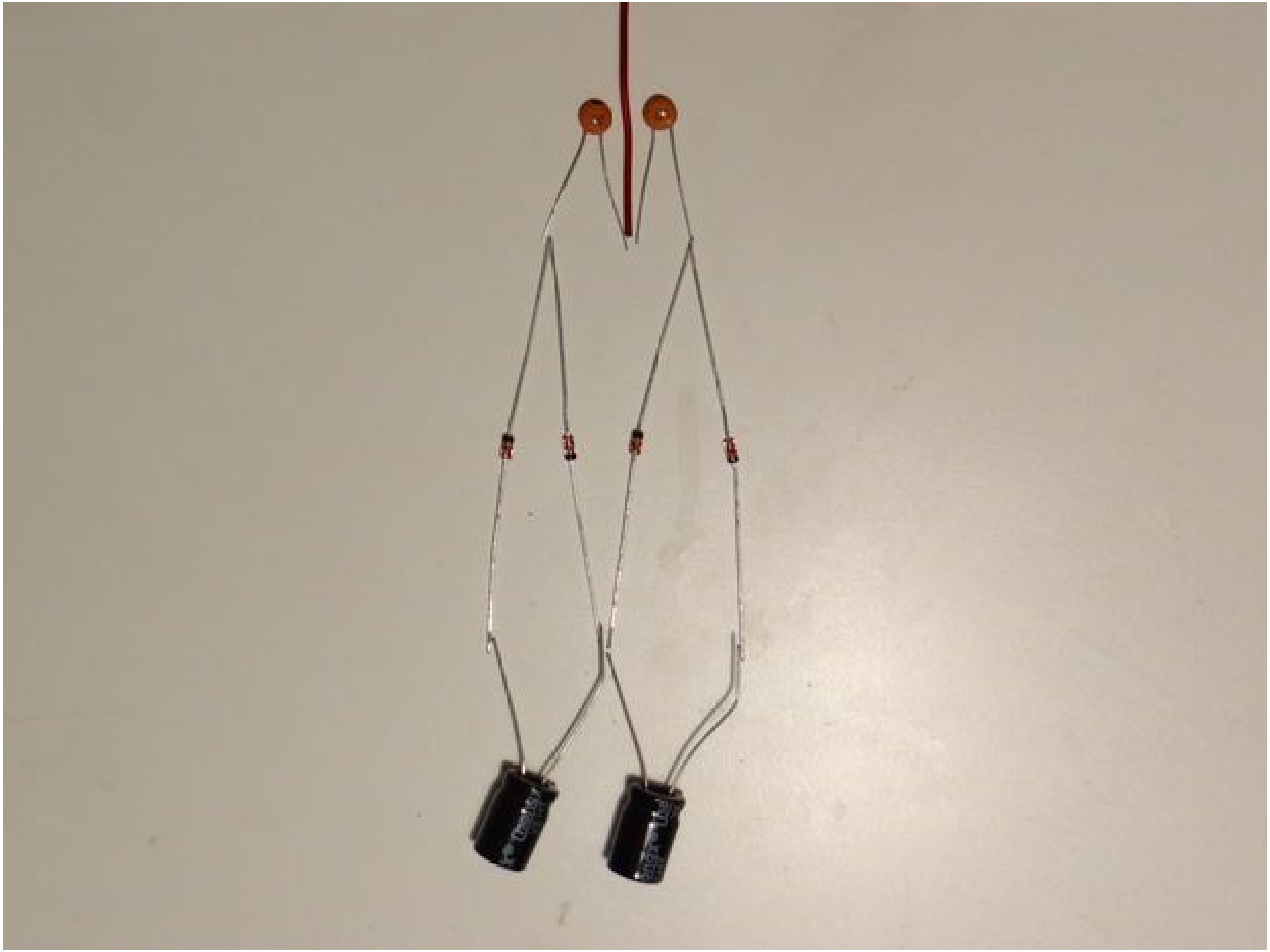

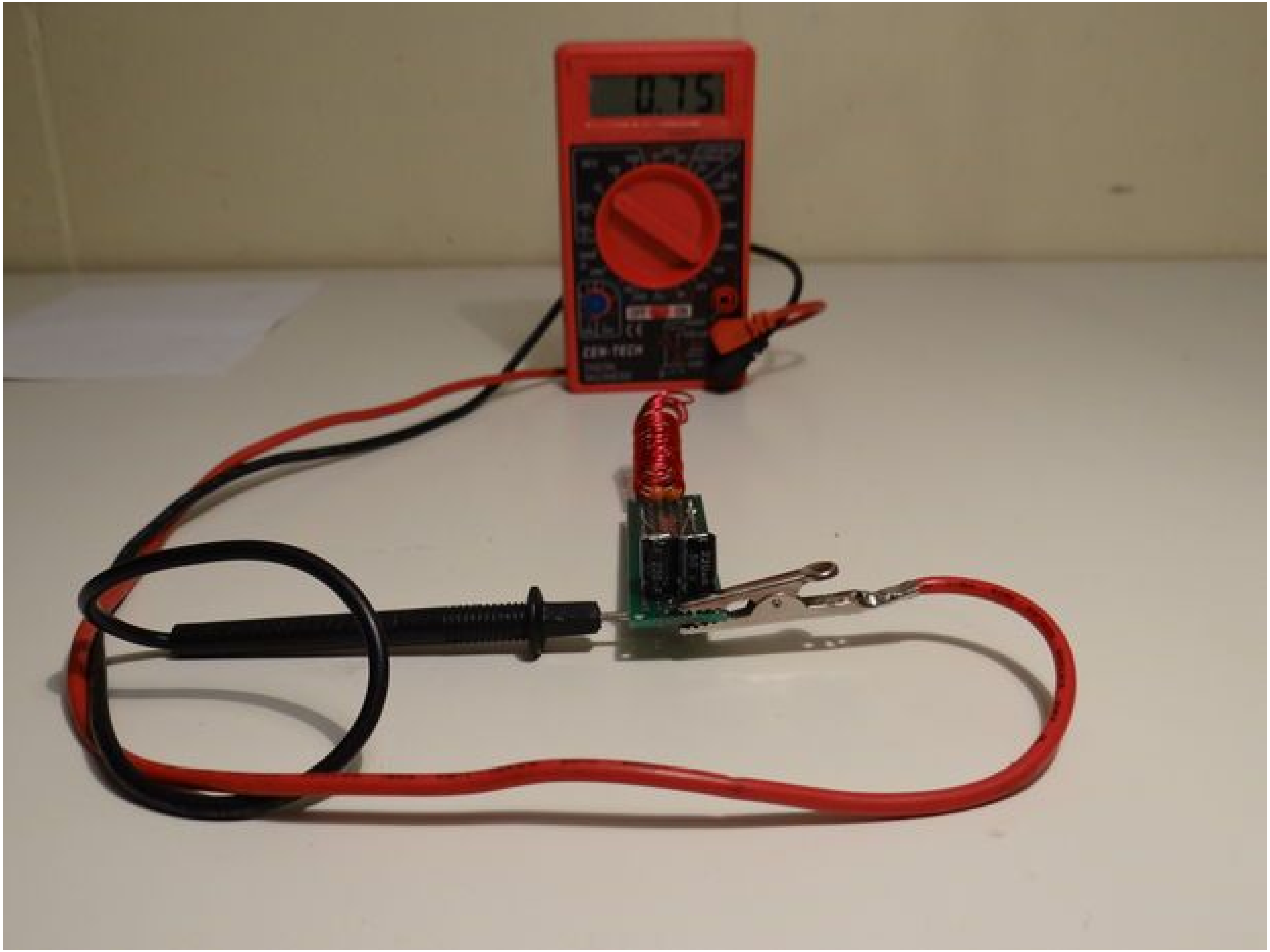
0.75
CEN-TECH

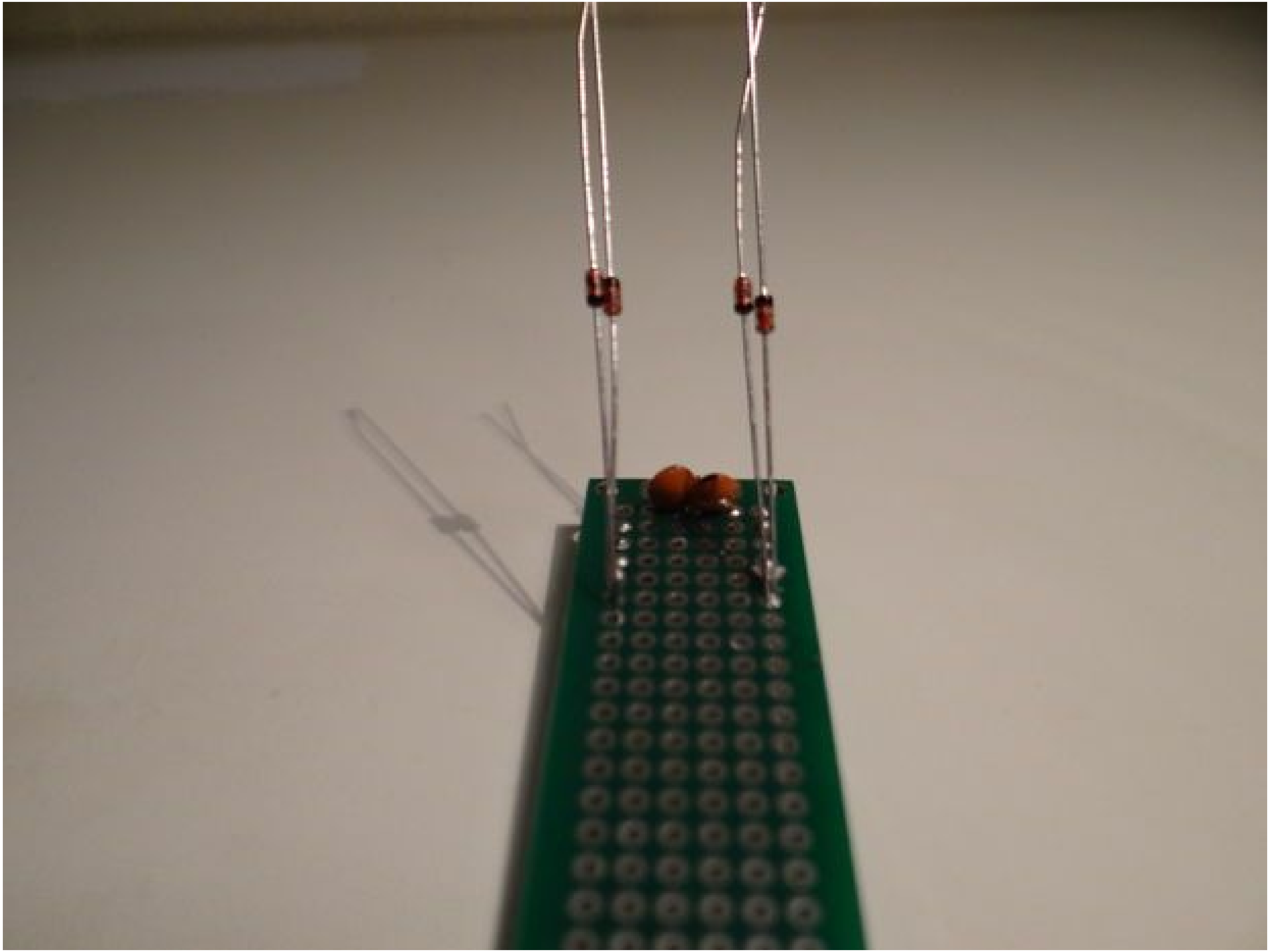

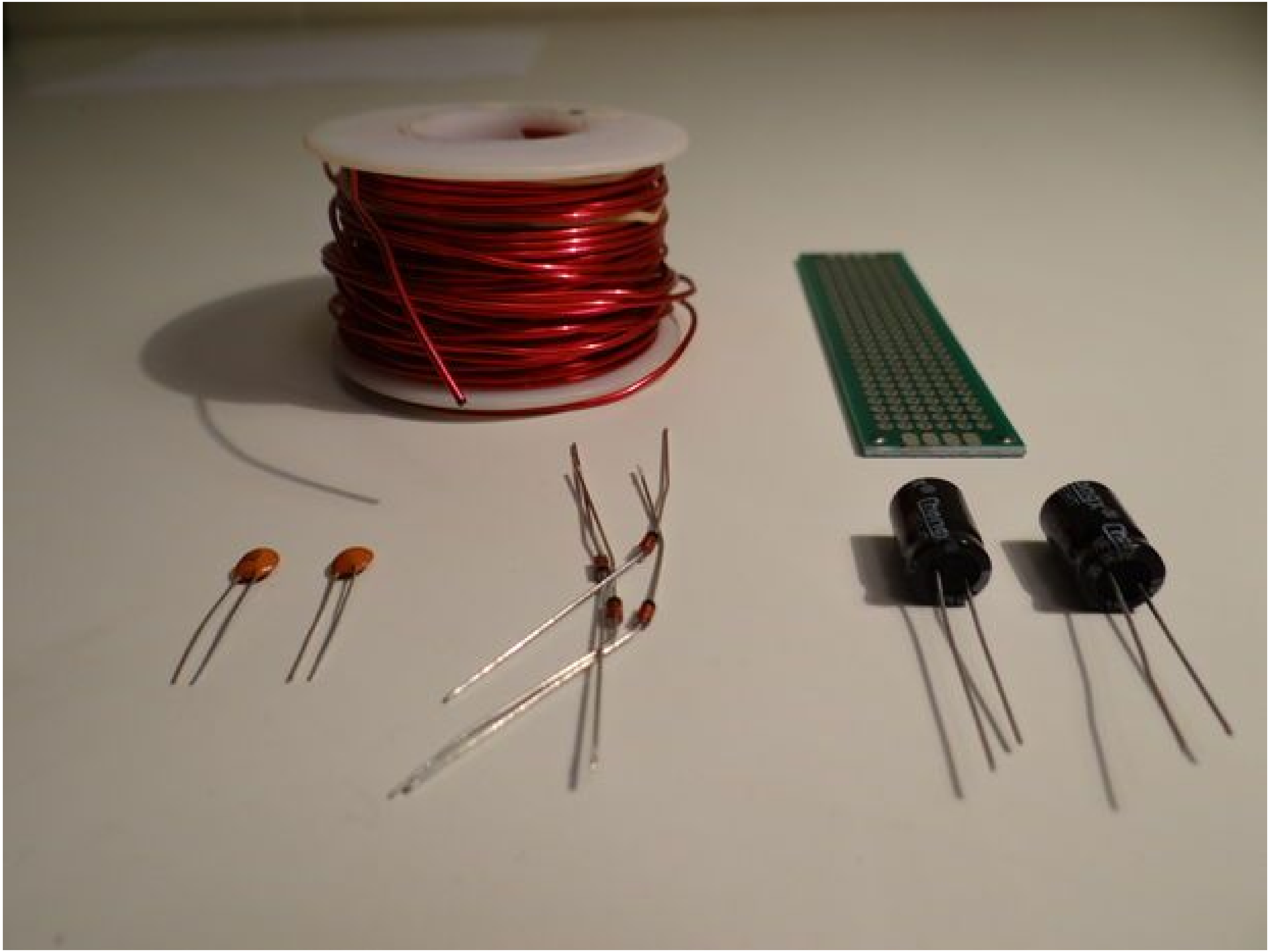

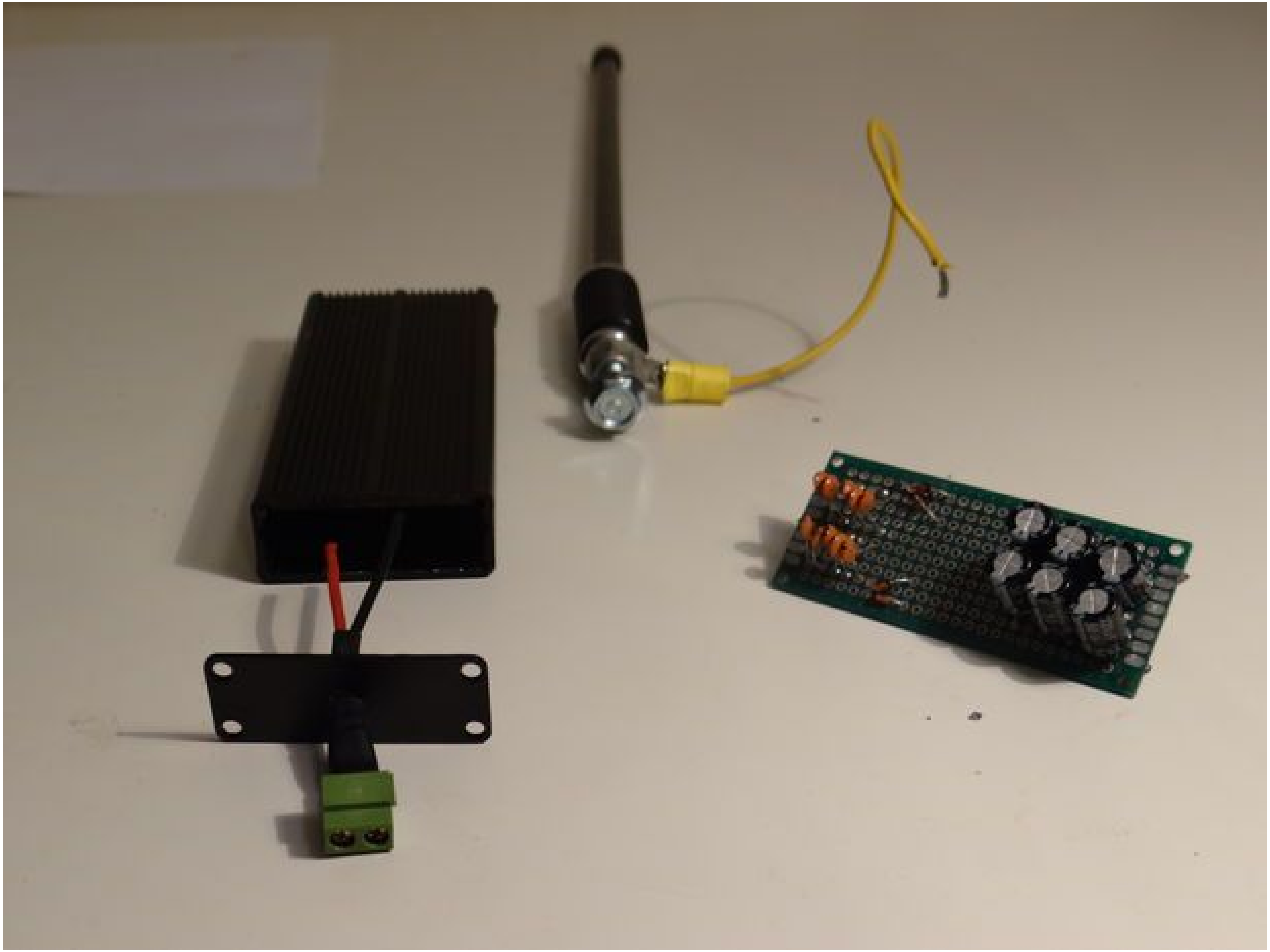

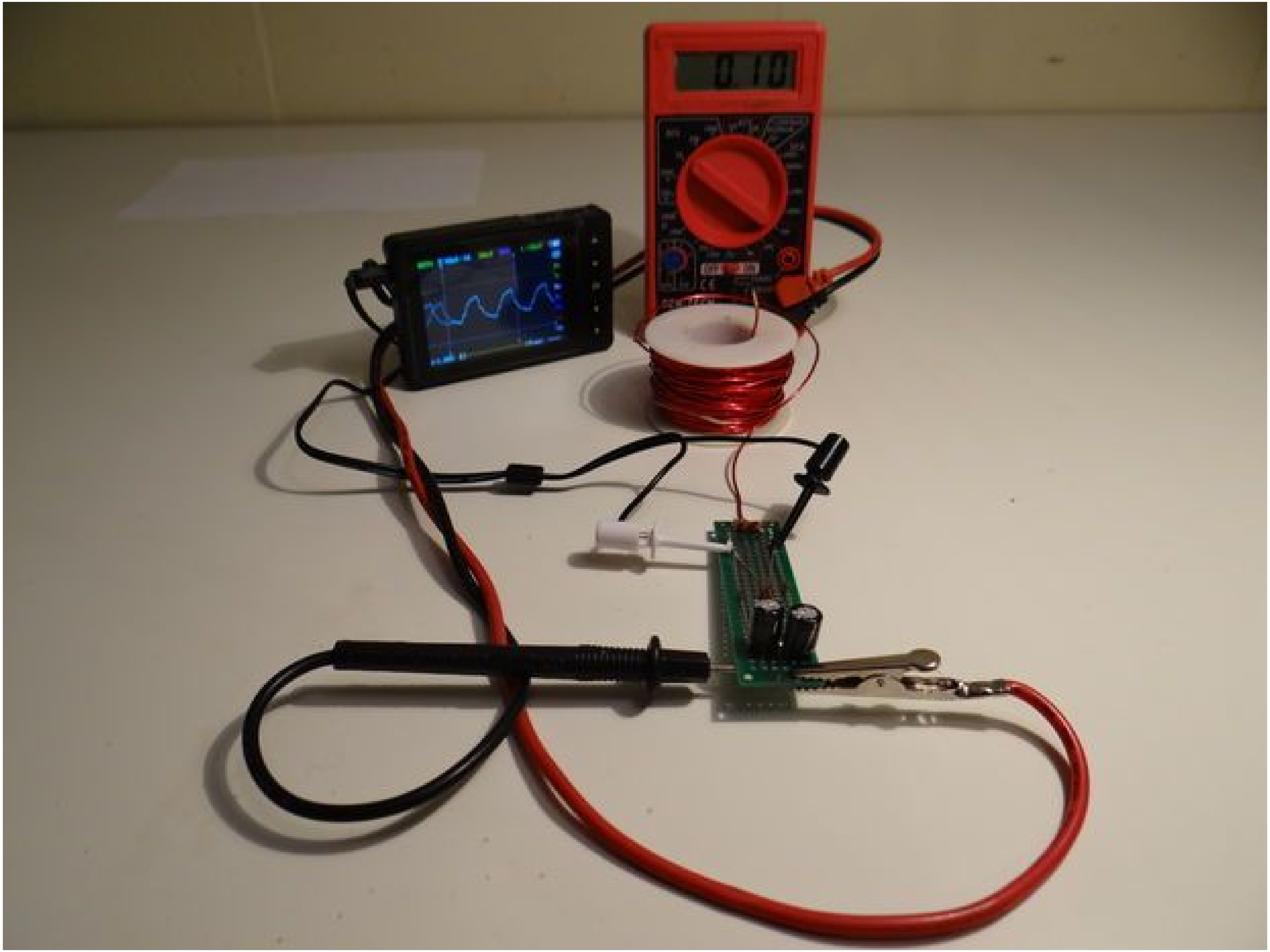
0.10

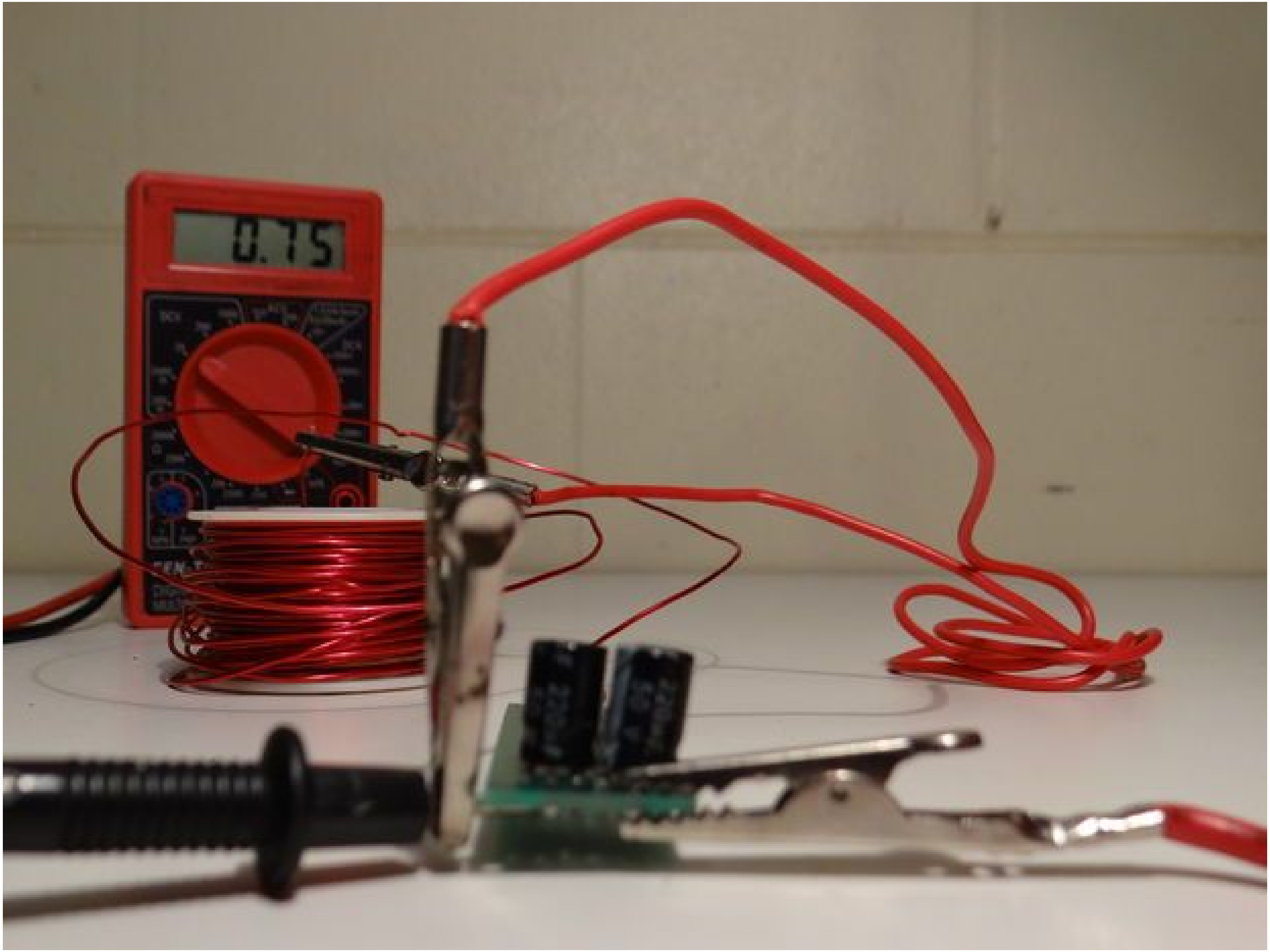
0.75

6485
A-329 X 100
68PF 50V

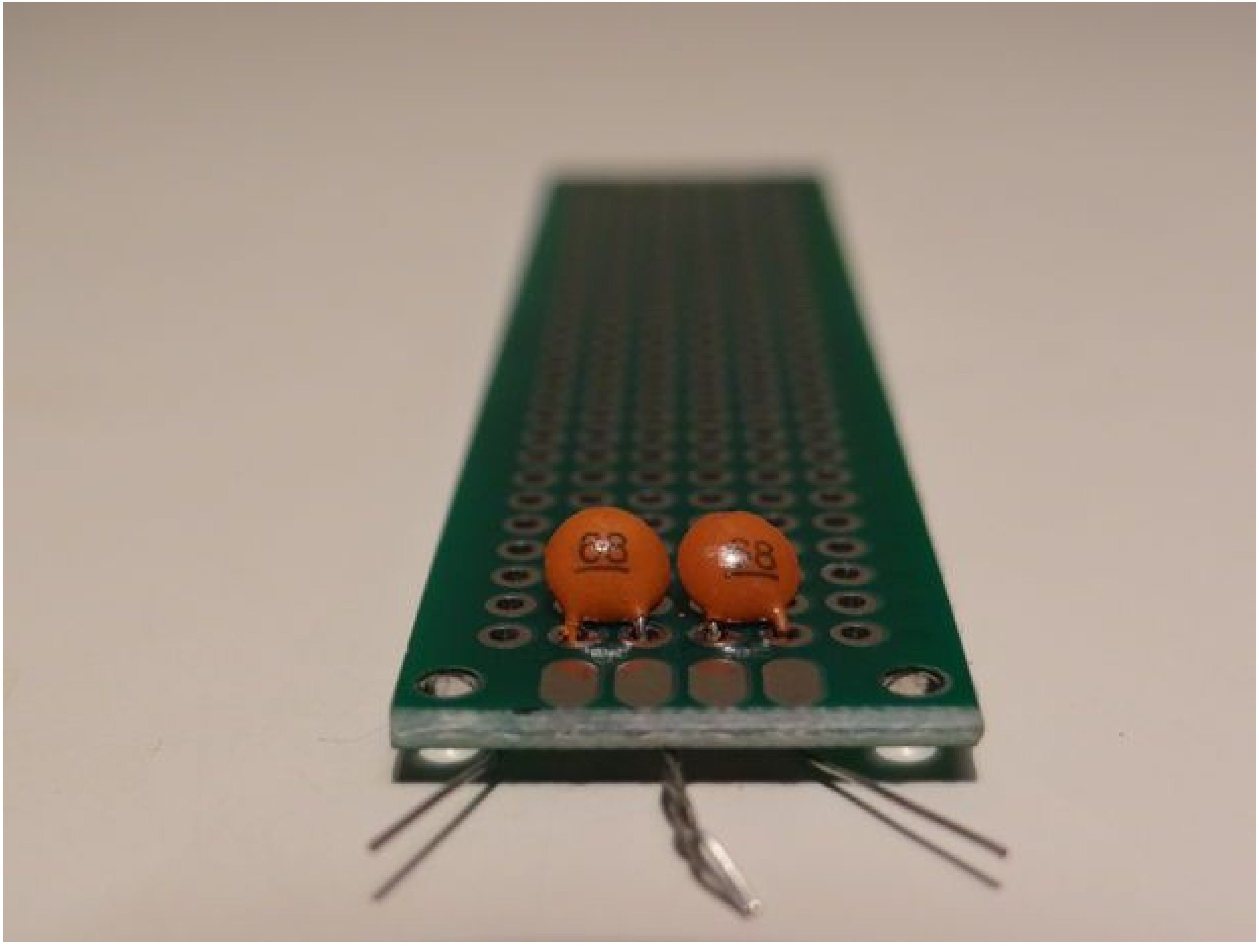
68
68

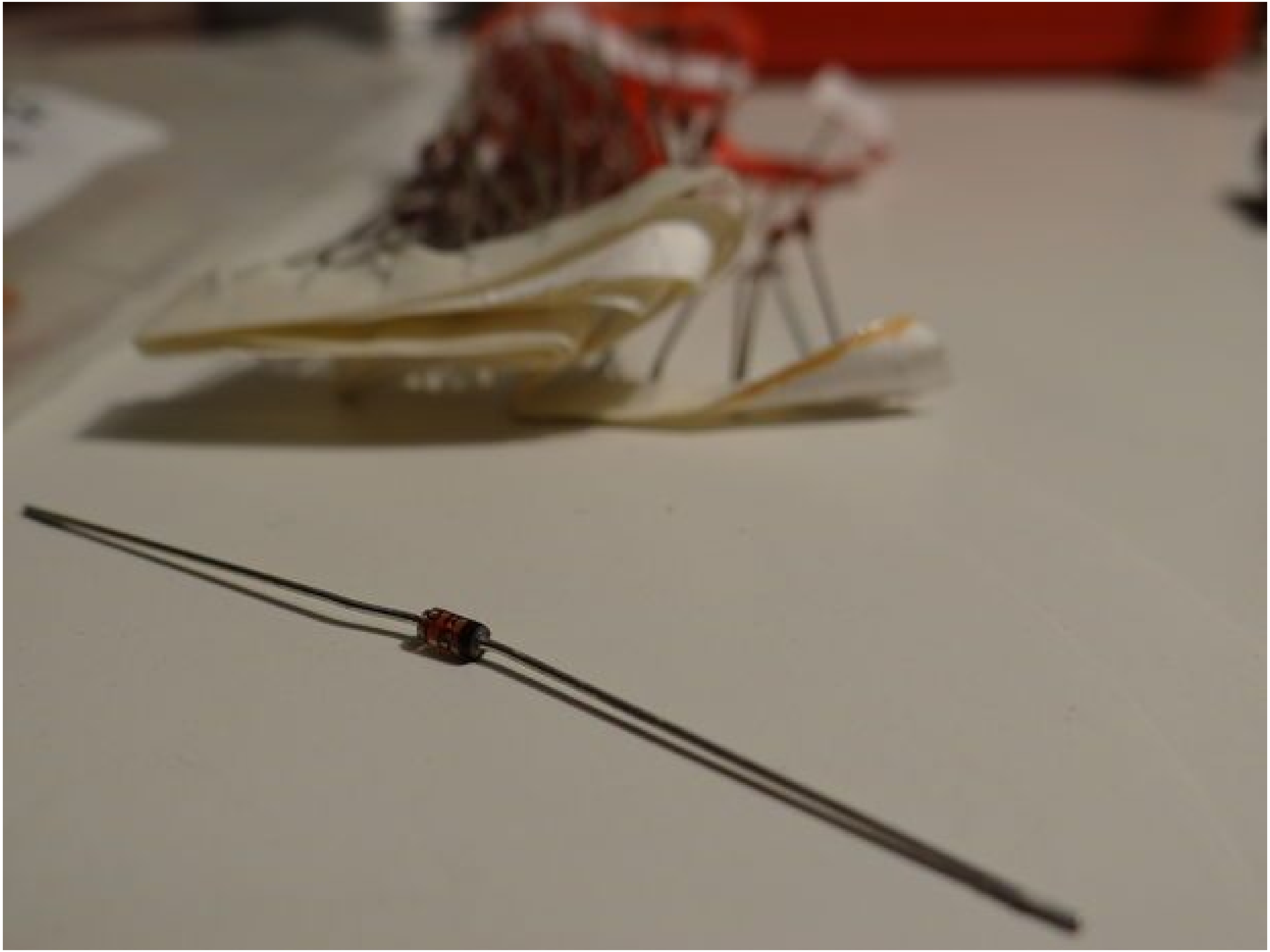

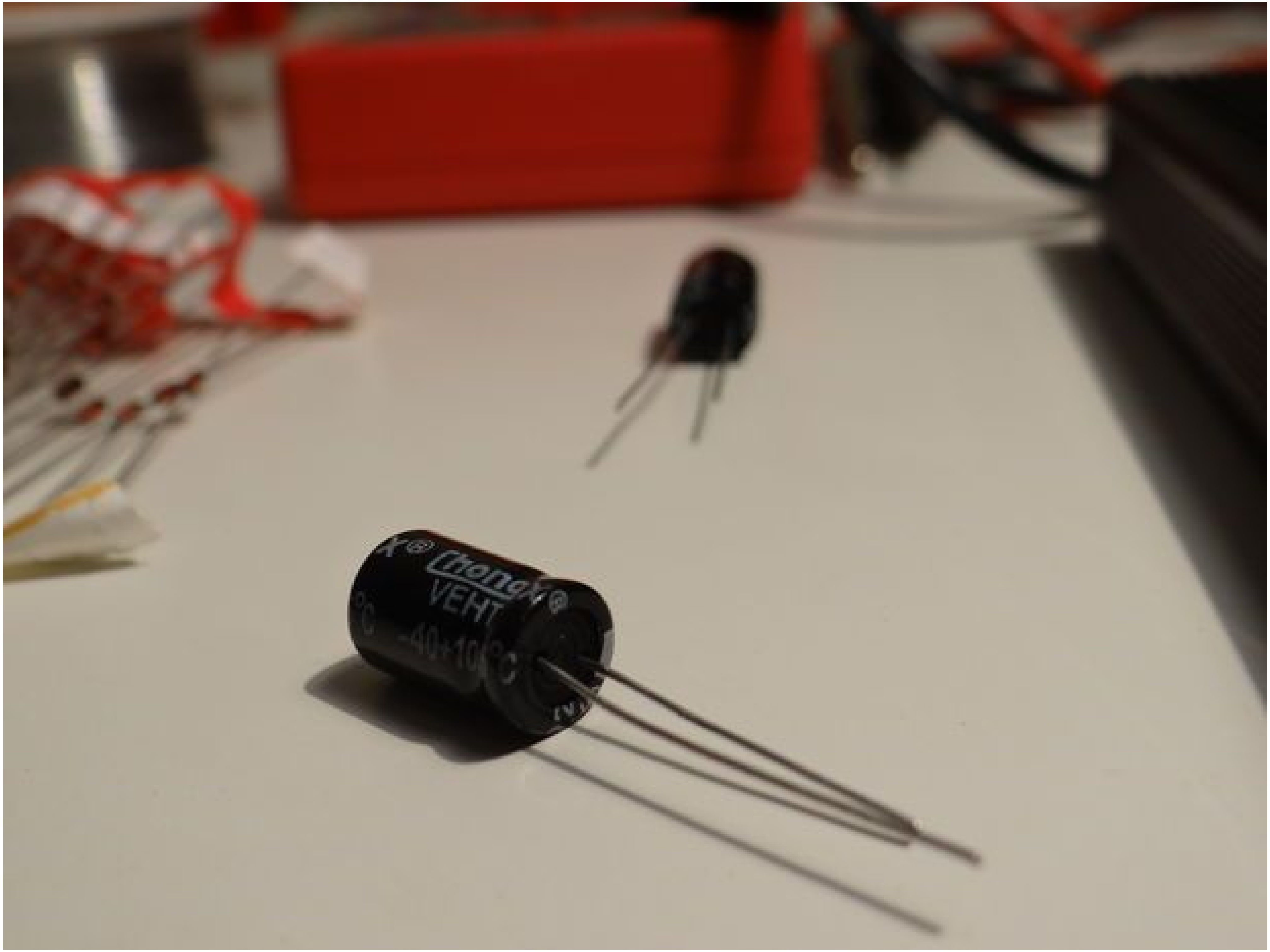

VEHT

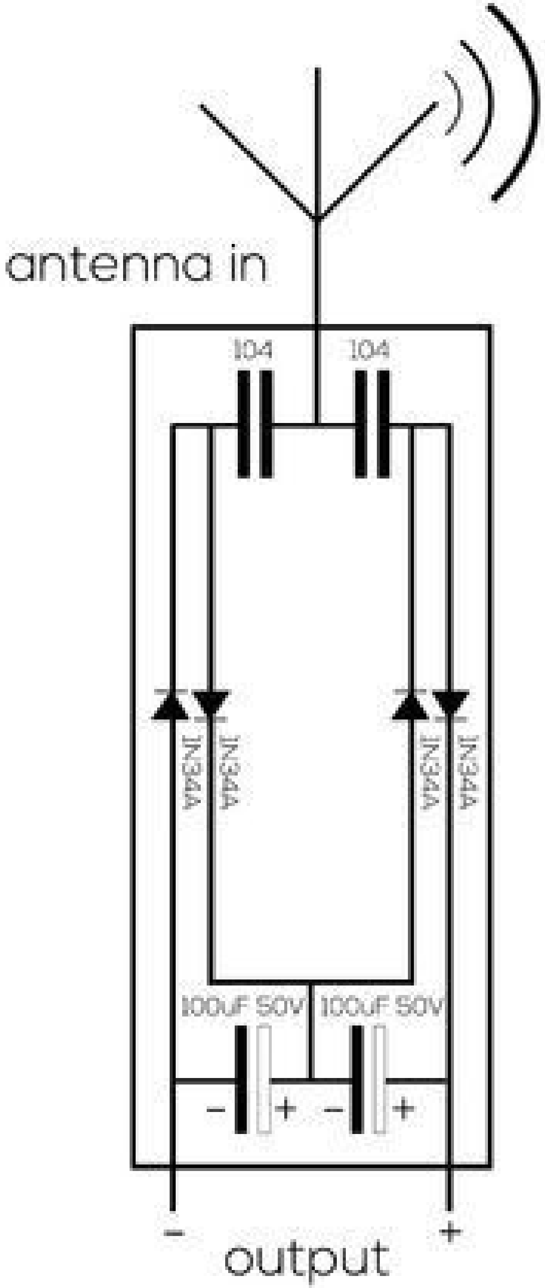
antenna in
104
104
1N34A
1N34A
1N34A
1N34A
100uF 50V
100uF 50V
–
+
–
+
–
+
output

RF to DC

Circuit Diagam (simplified)

Component list:
(2) Ceramic Capacitors (104)
(4) Germanium Diodes (1N34A)
(2) Electrolytic Capacitors (100uF 50V)

Copyright 2016. Drew Paul Designs. All rights reserved.

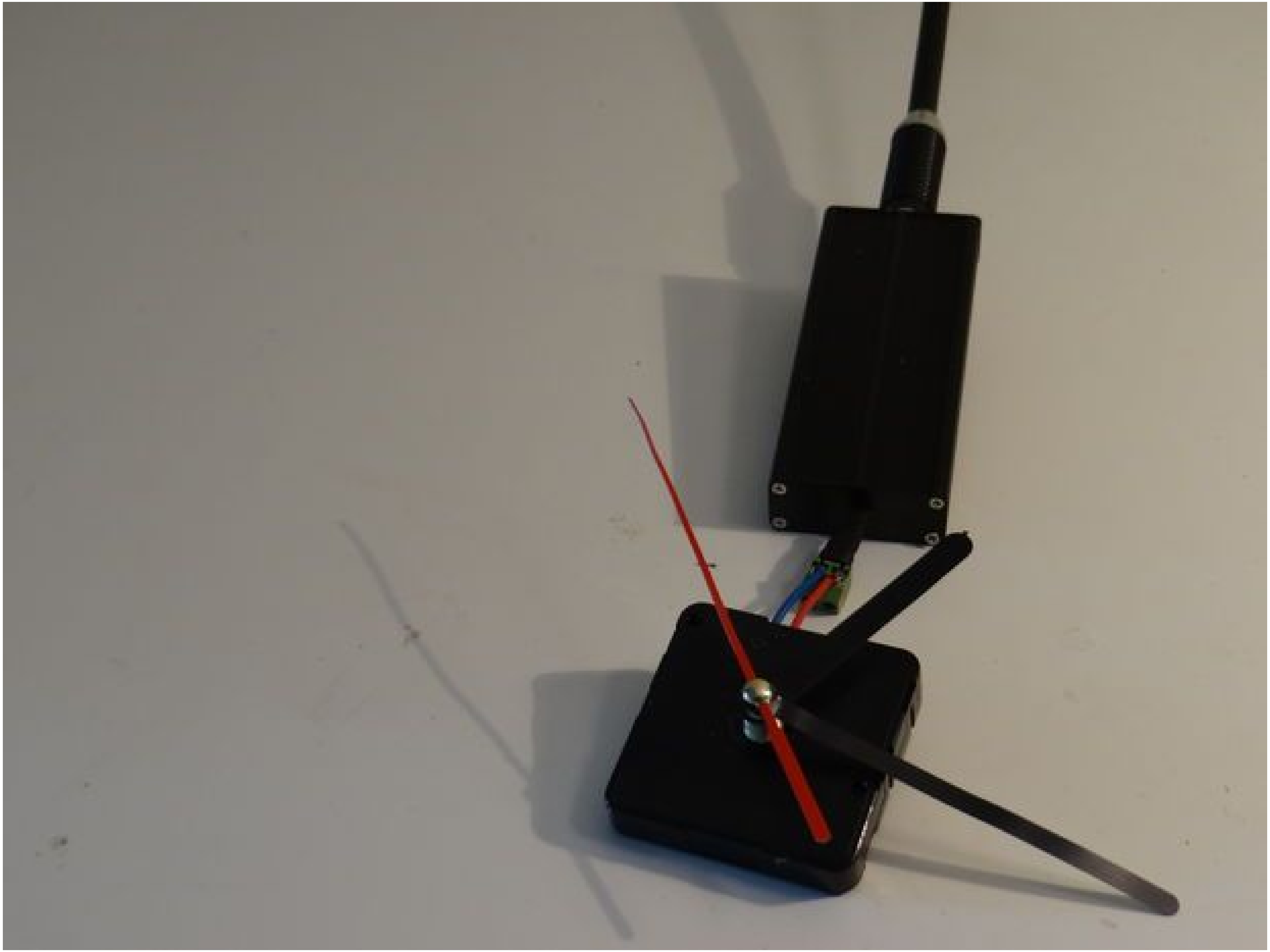

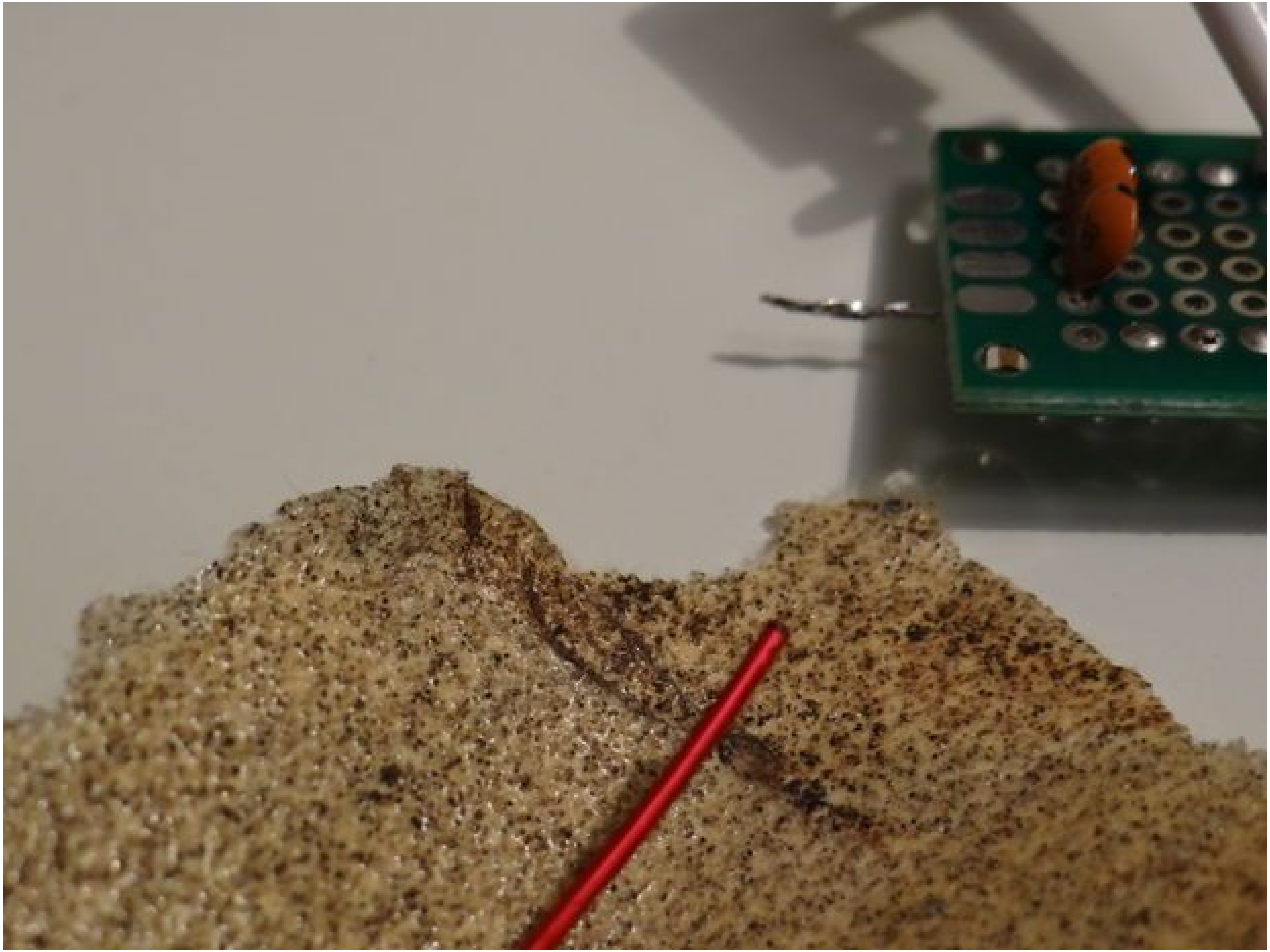

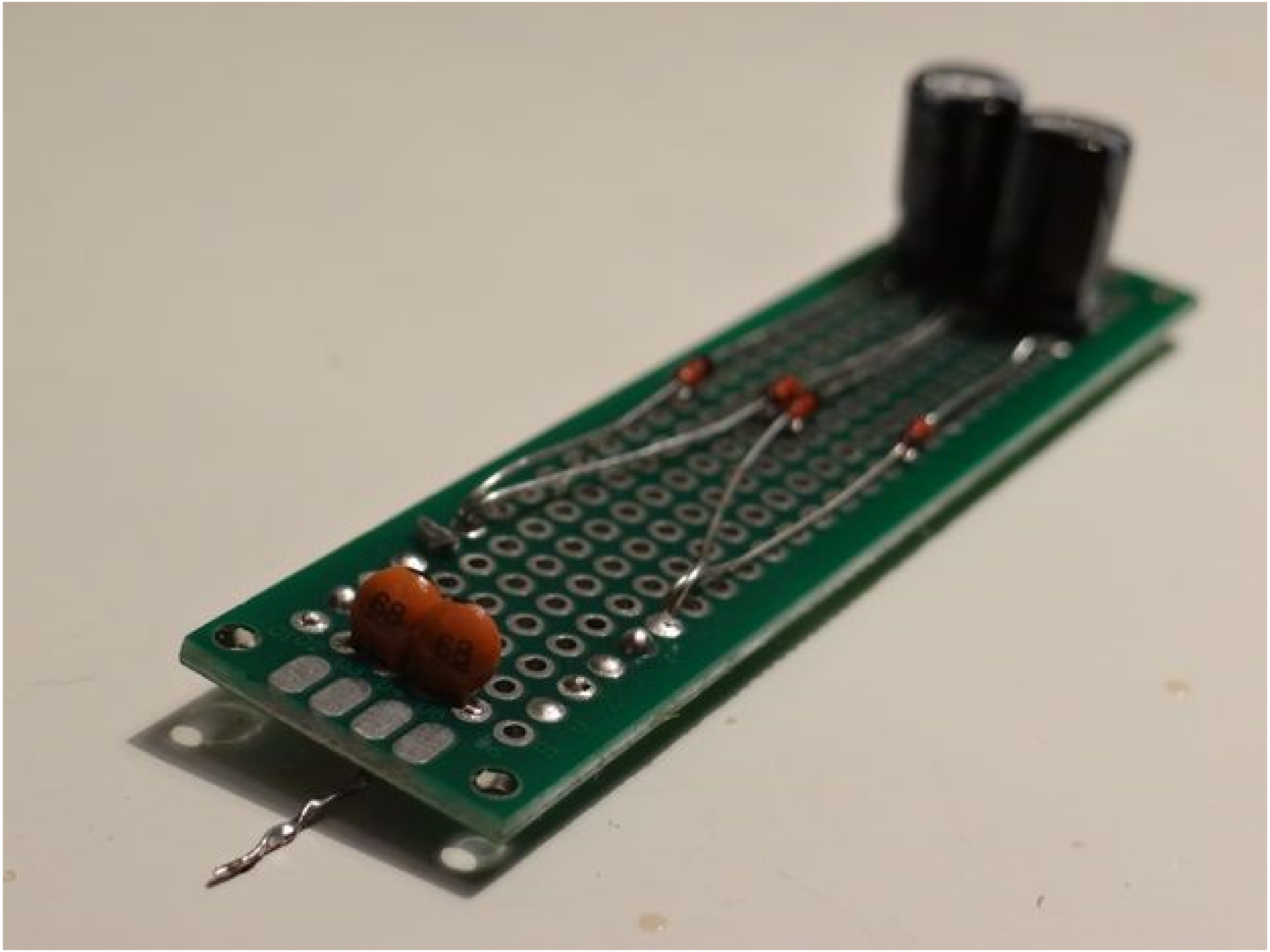
68
68

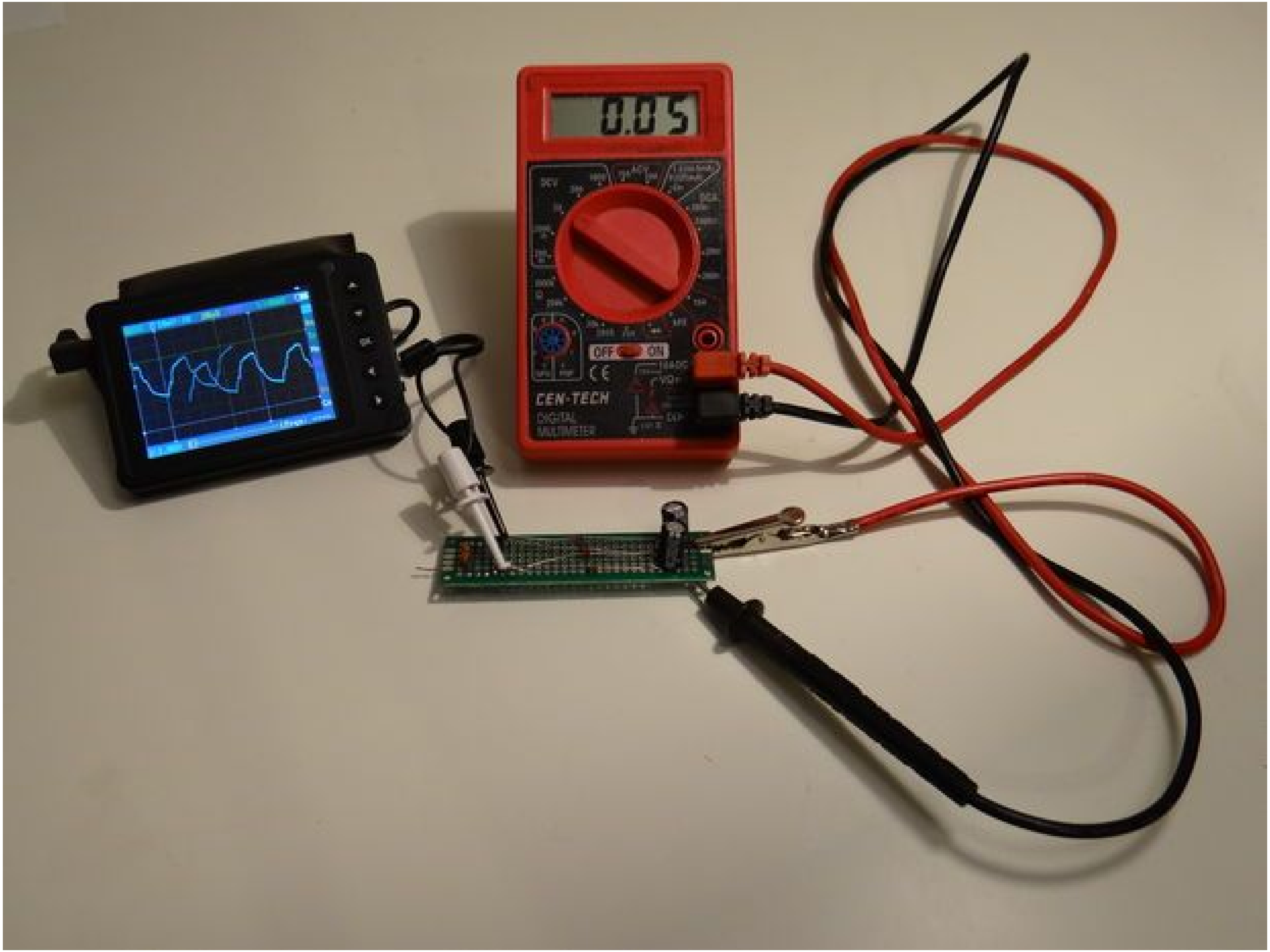
0.05
OFF ON
CEN-TECH
DIGITAL
MULTIMETER

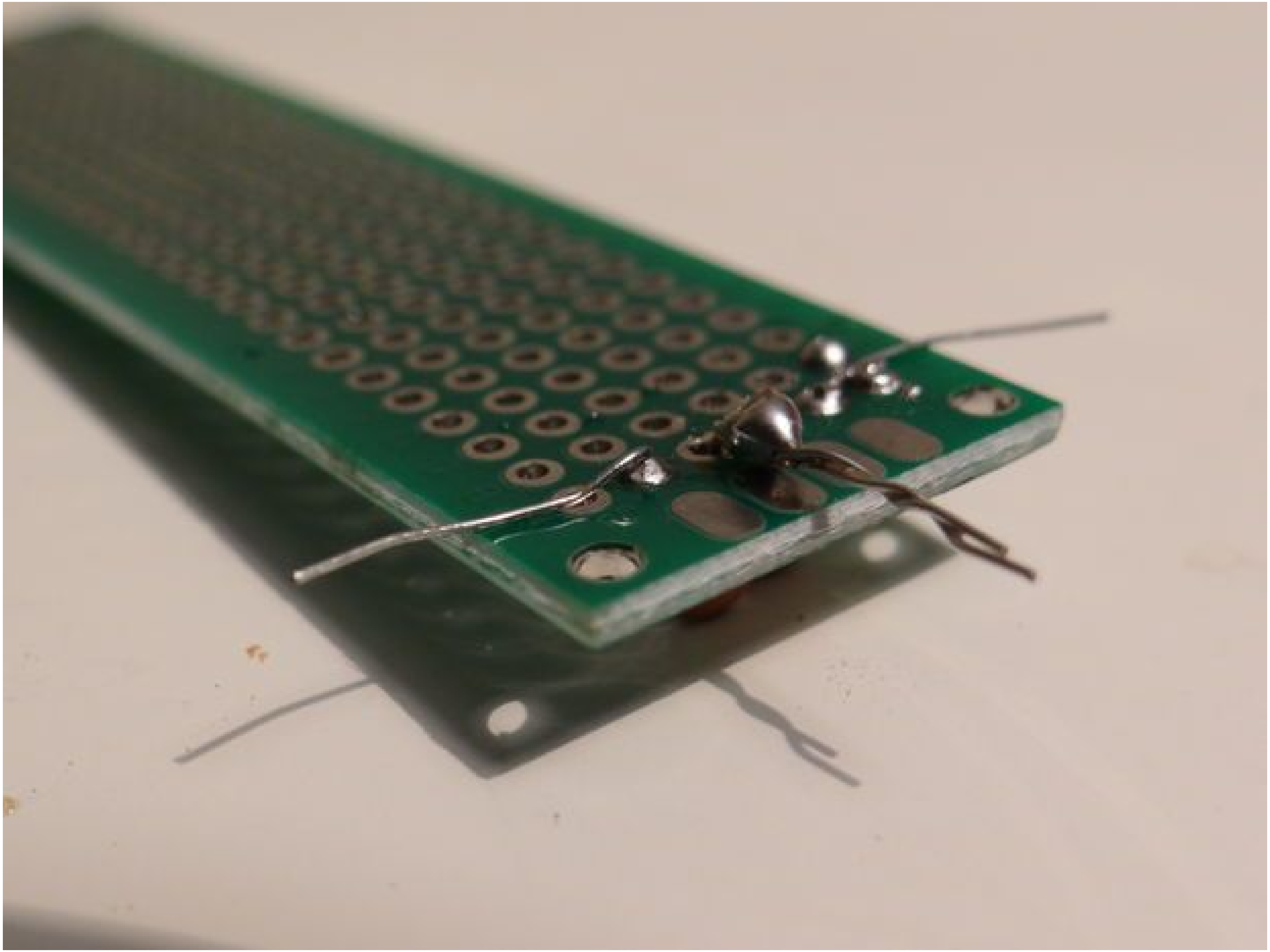

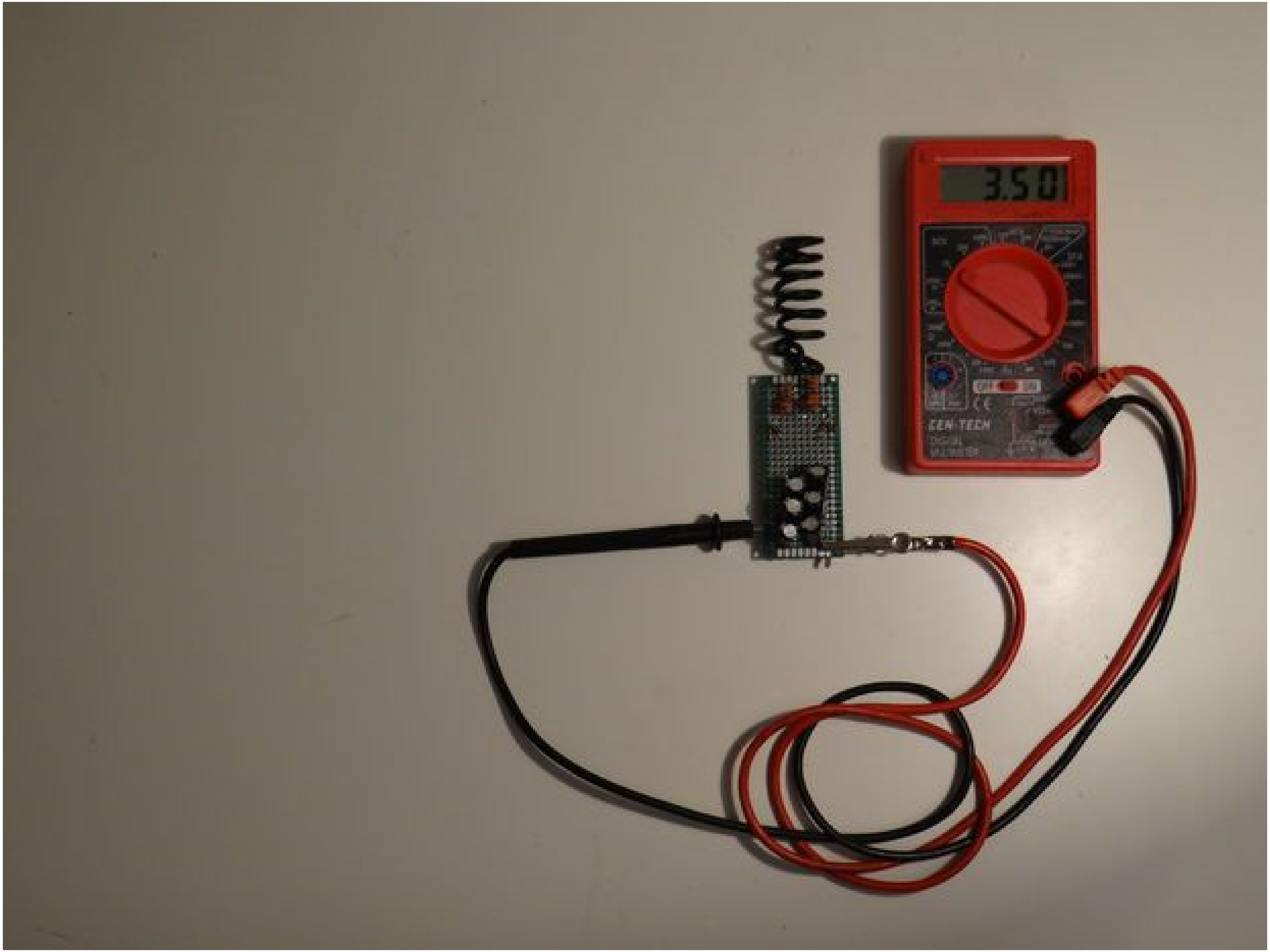
3.50
CEN-TECH
DIGITAL
MULTIMETER

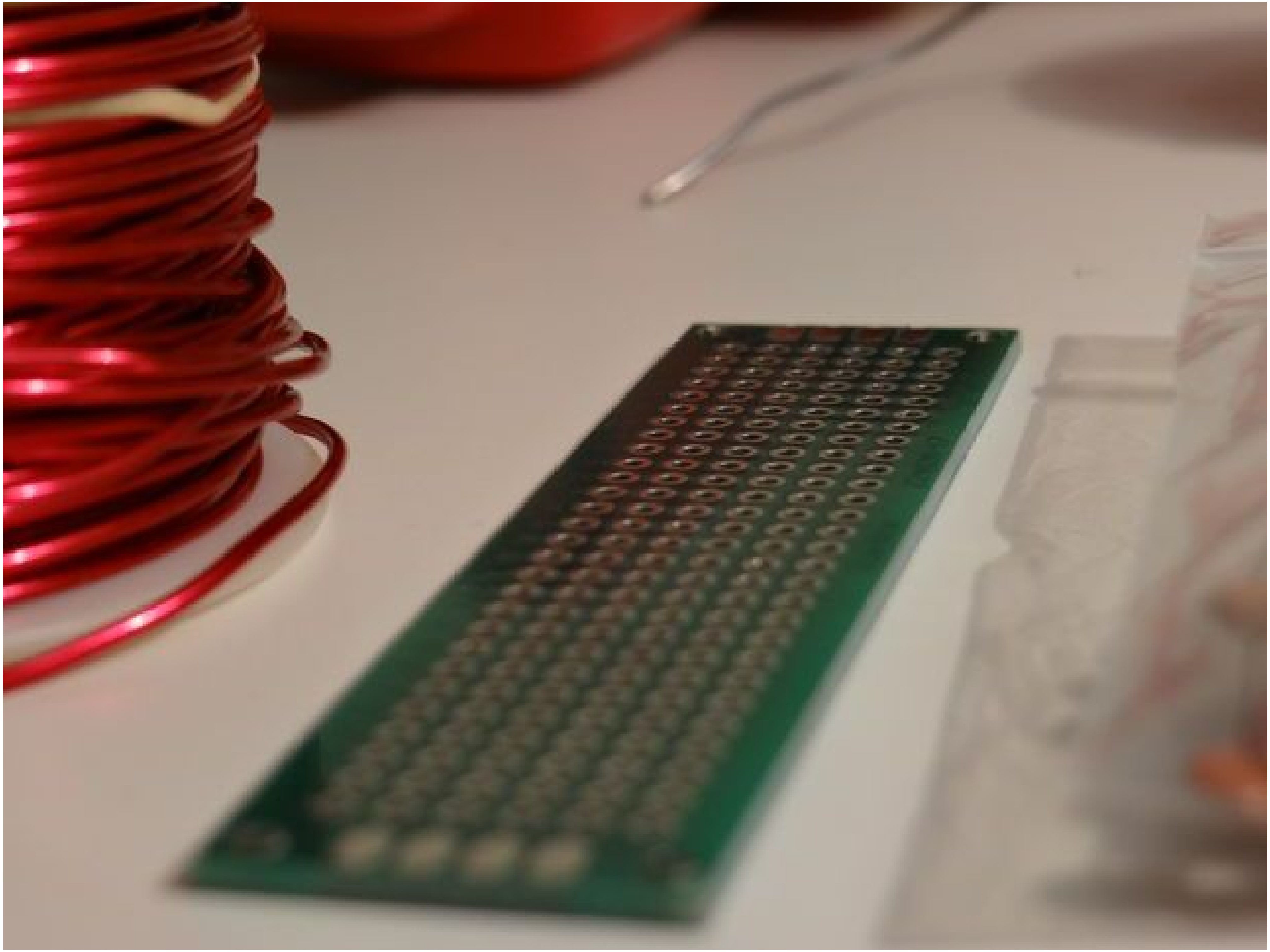

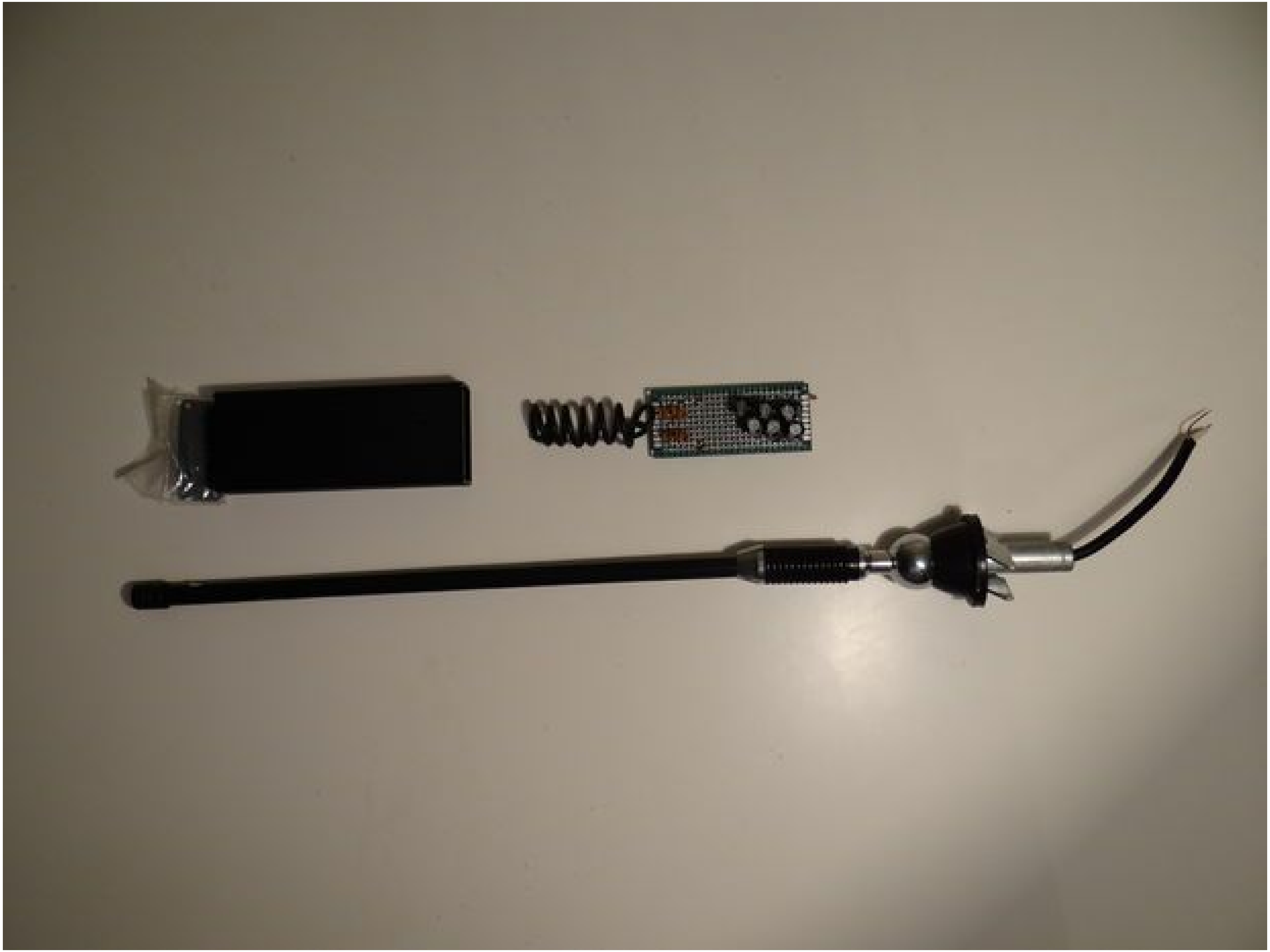

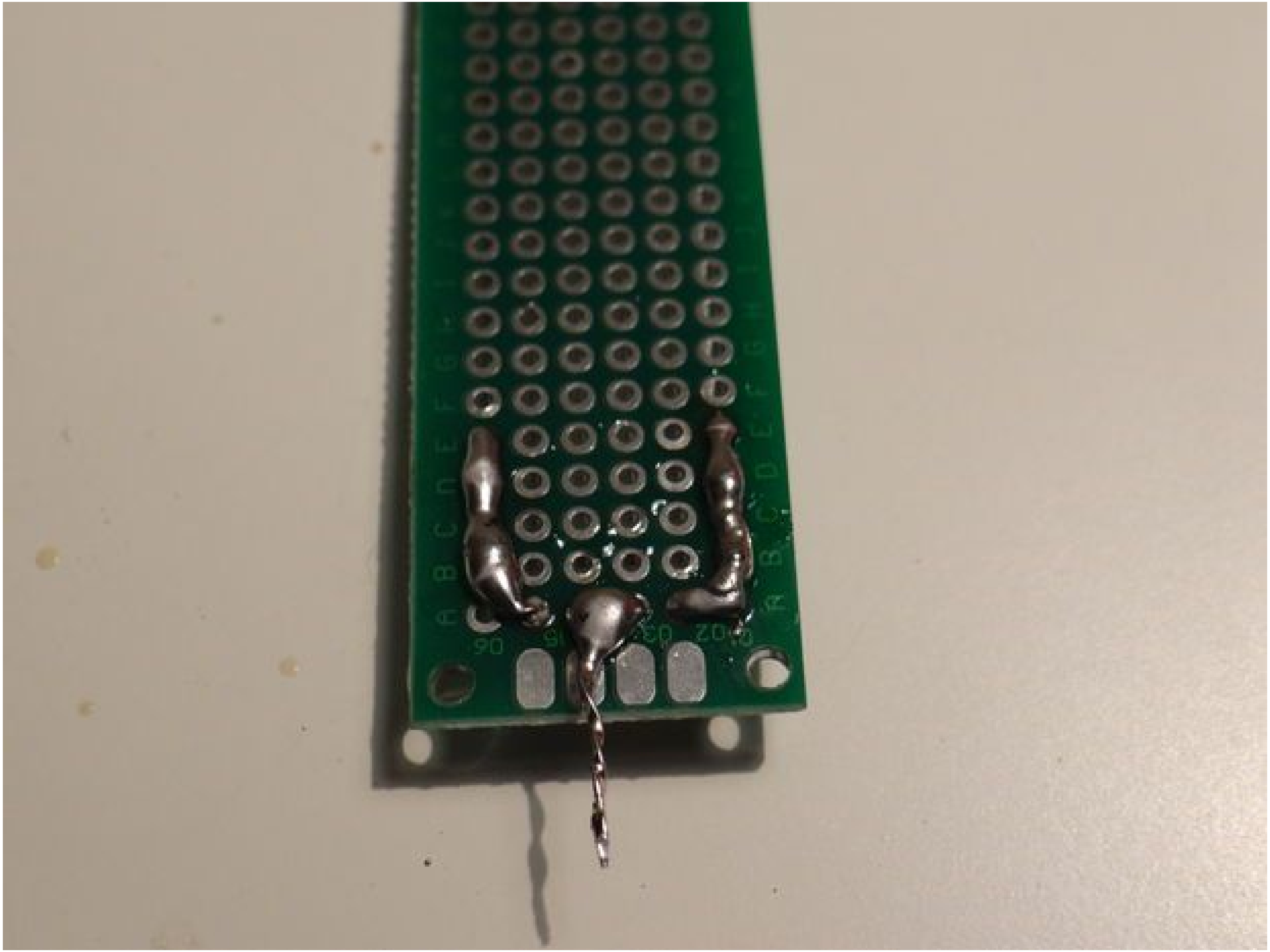
A
B
C
D
E
F
G
H
06
03
02

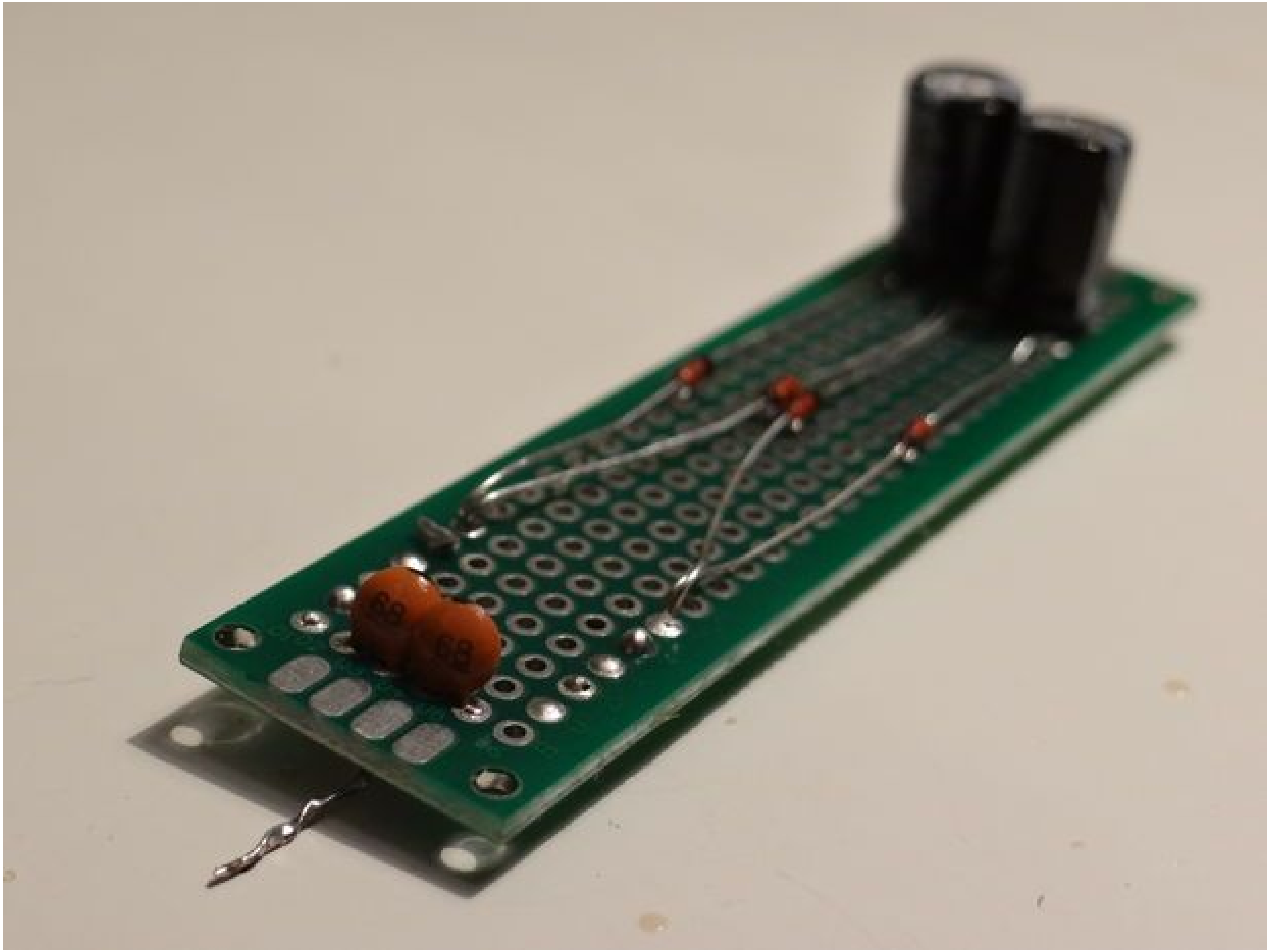
68
68

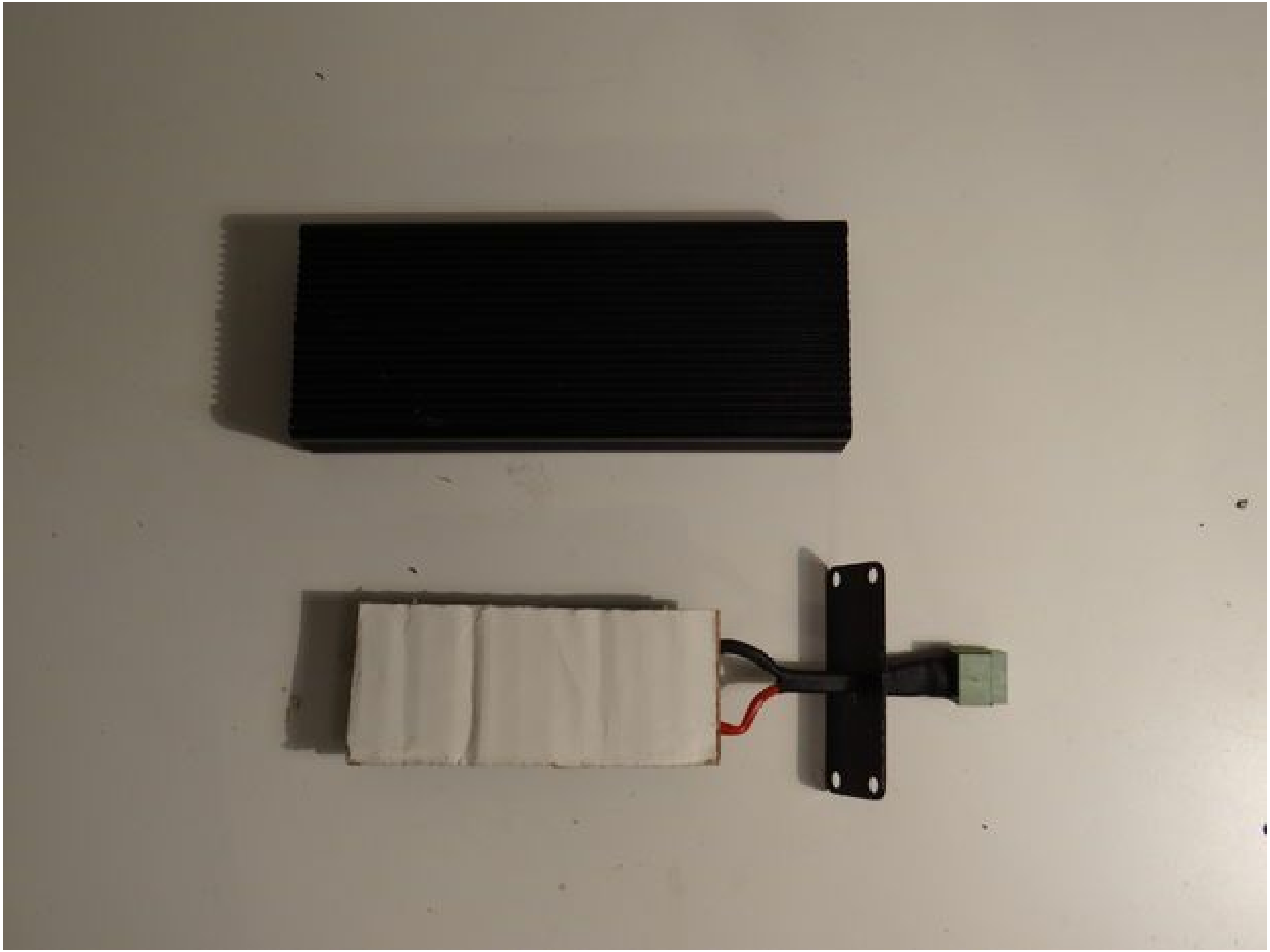

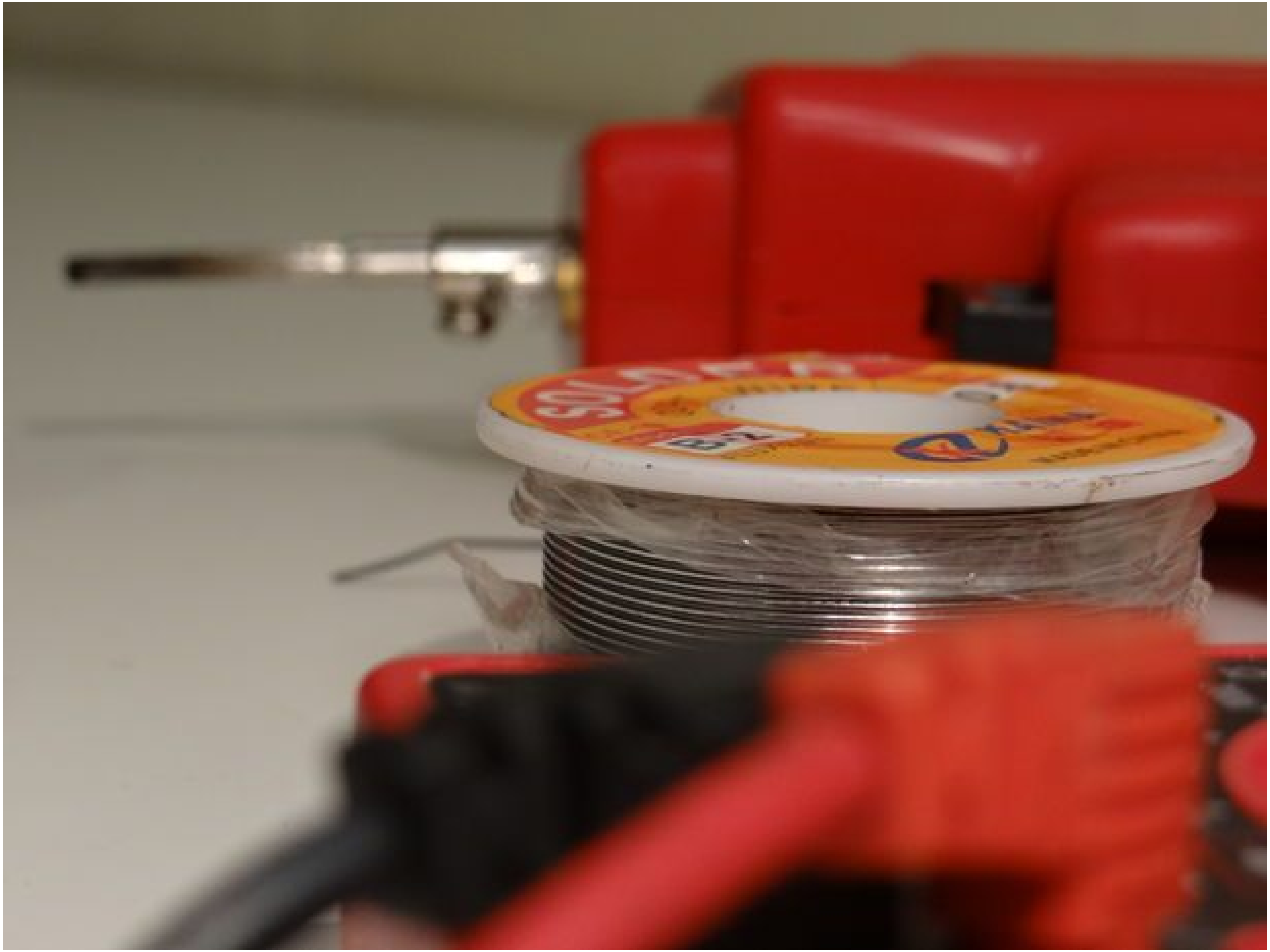

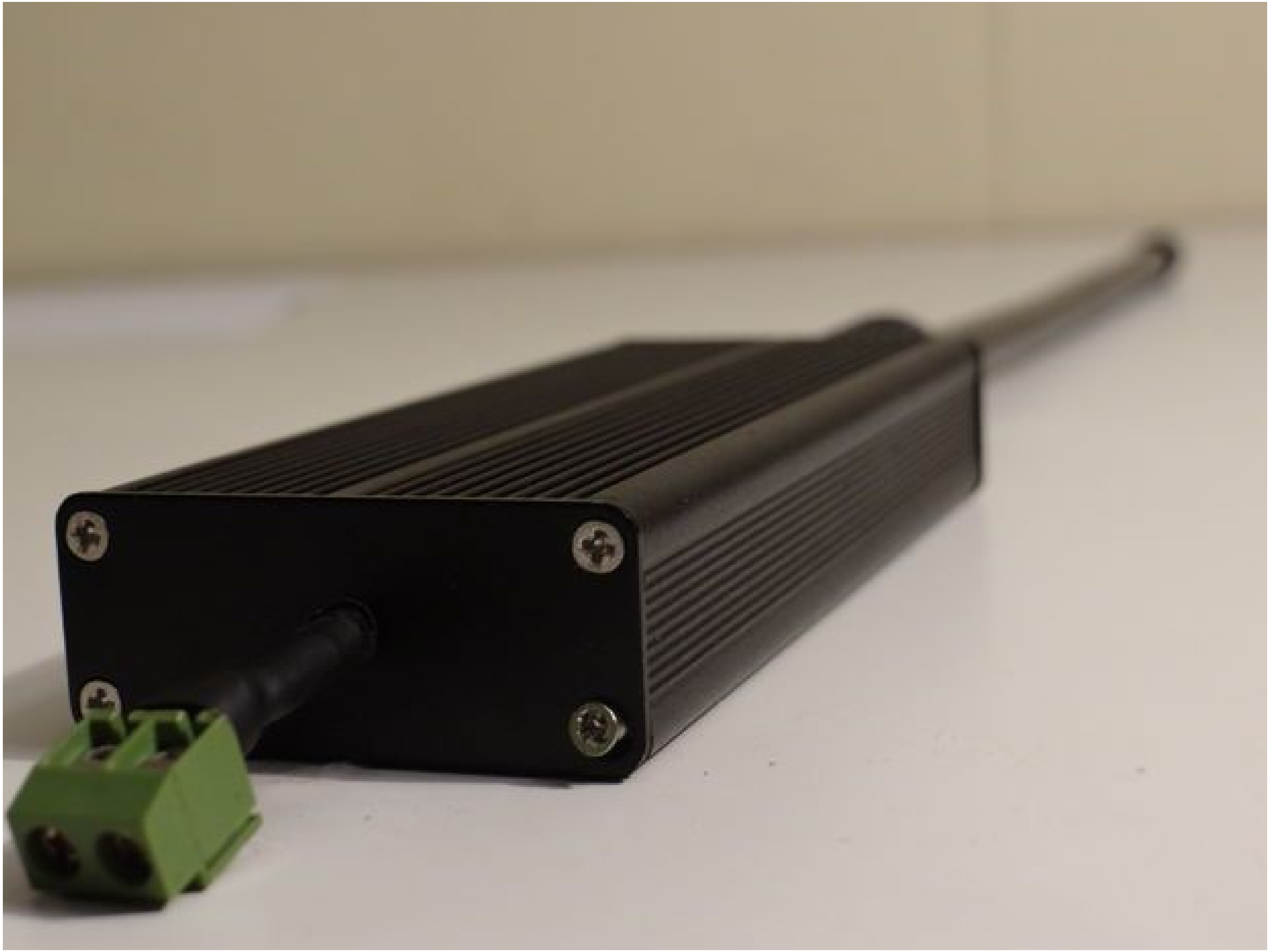

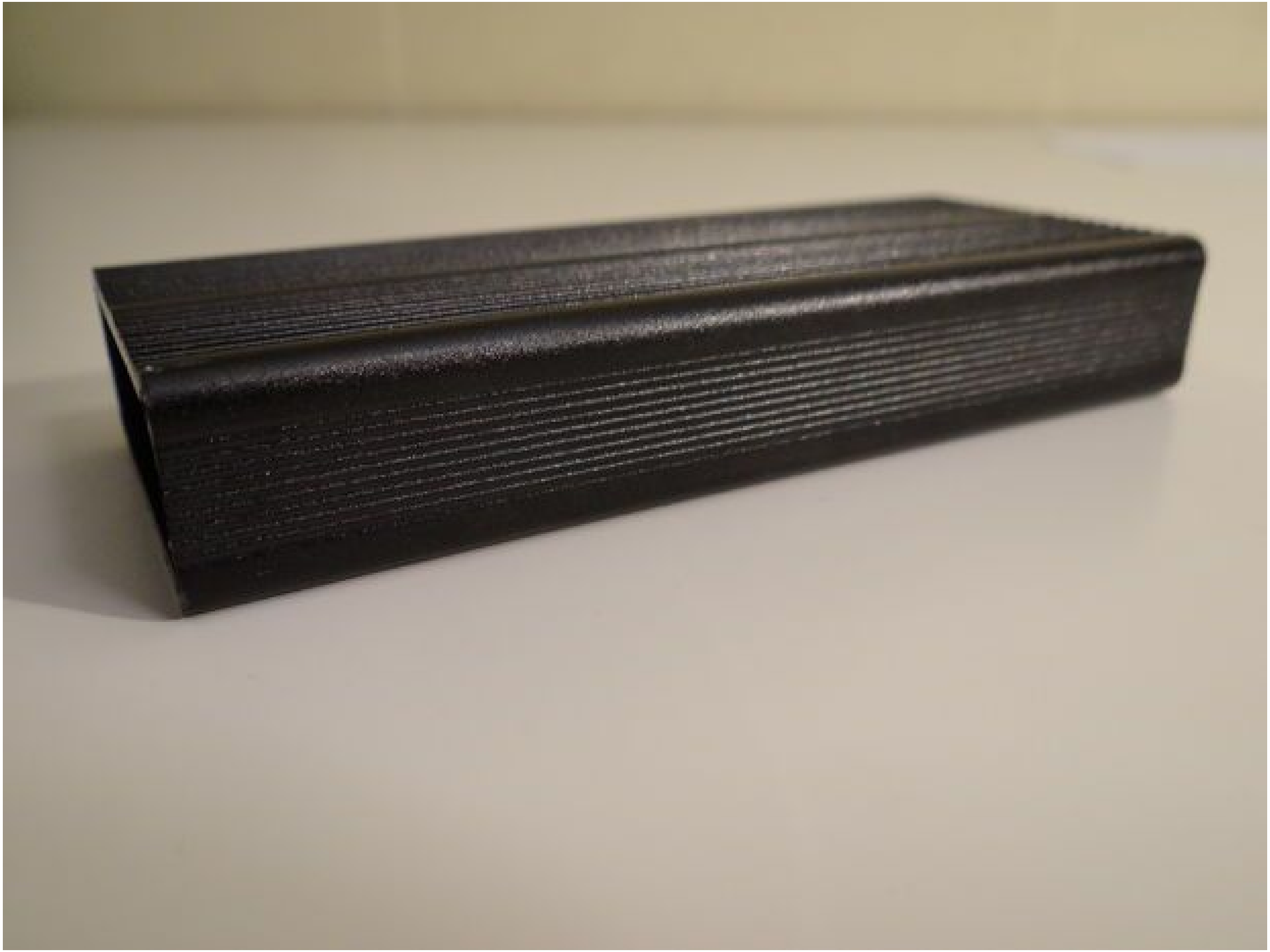

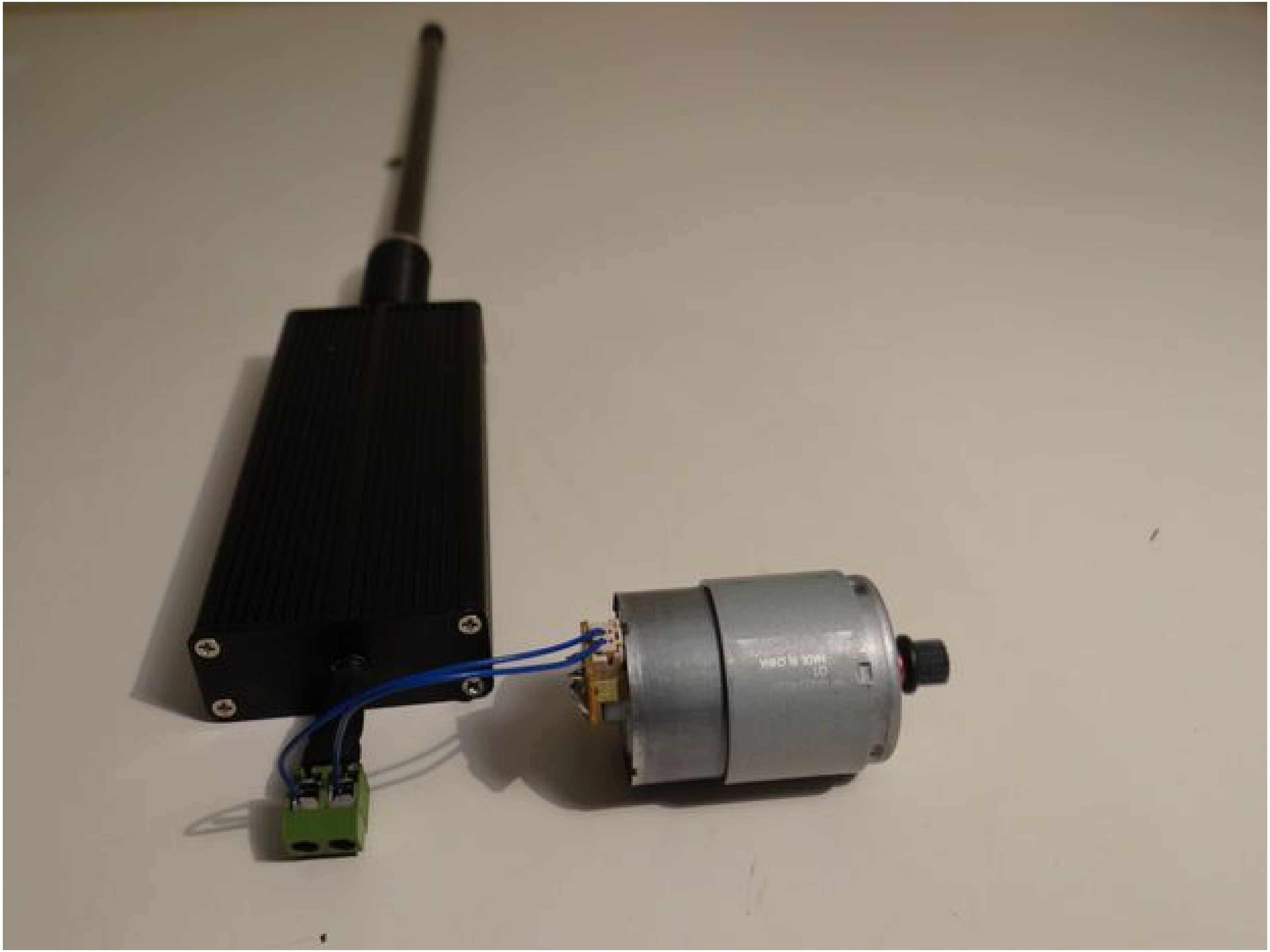

13:00

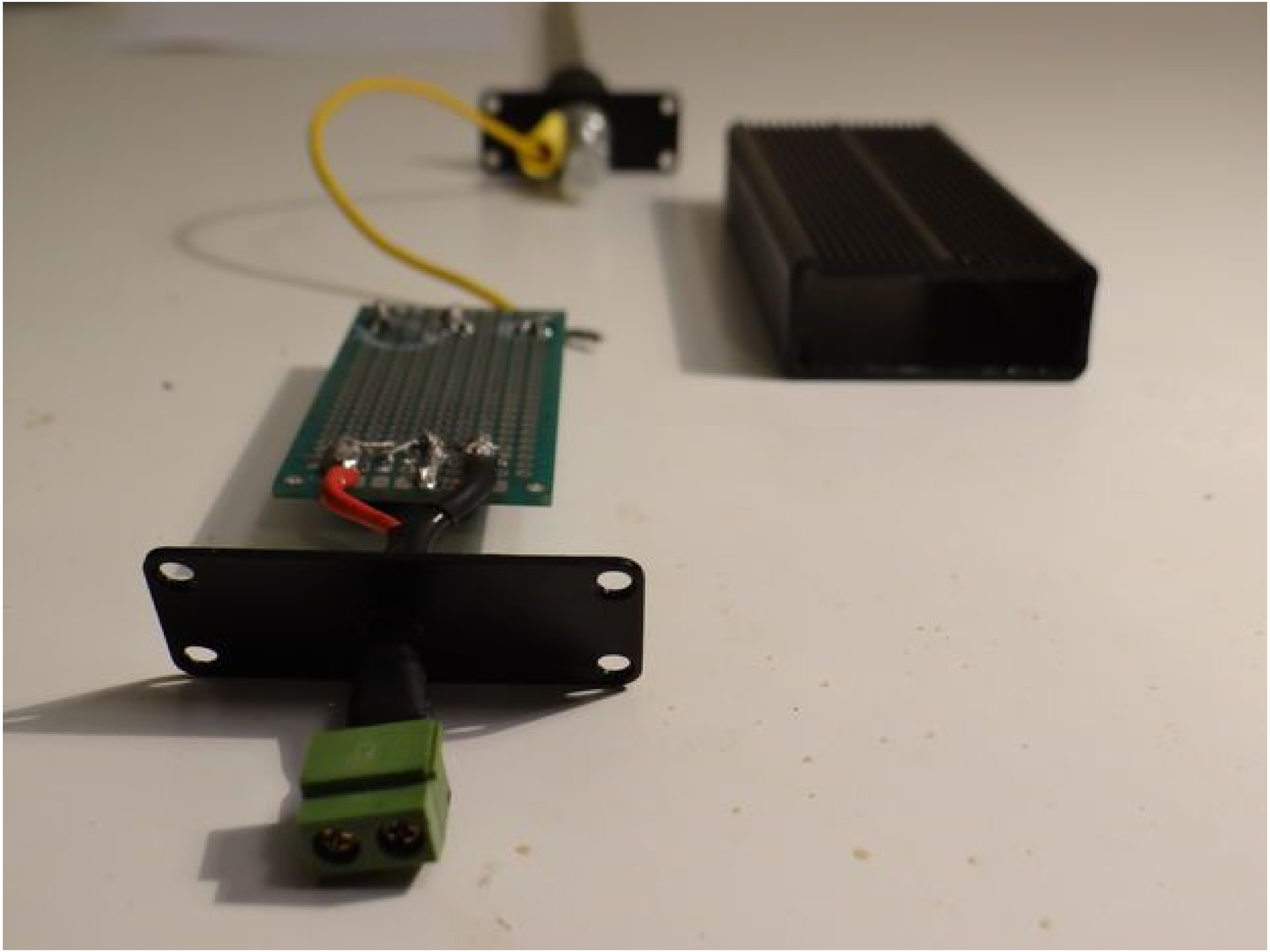

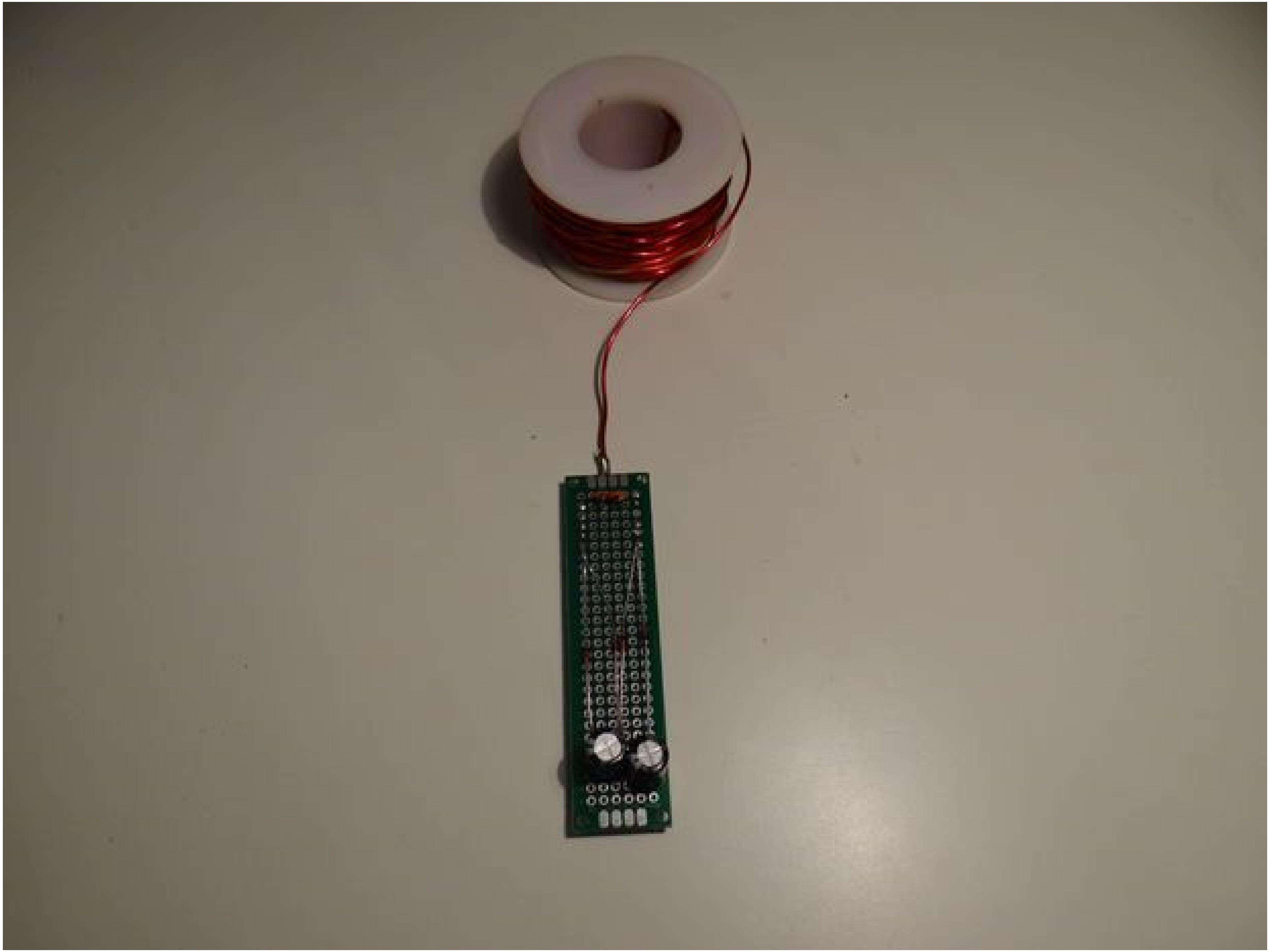

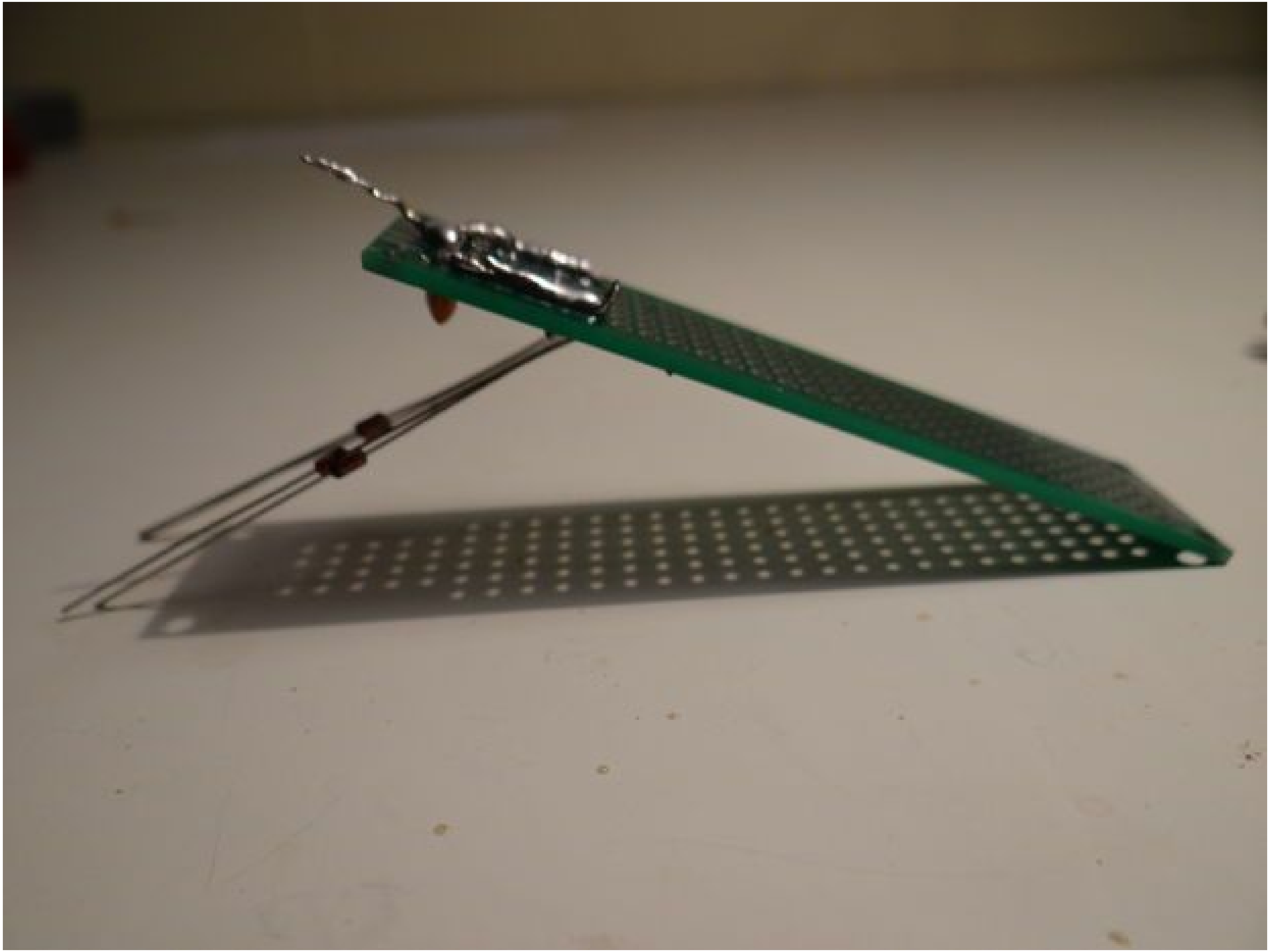

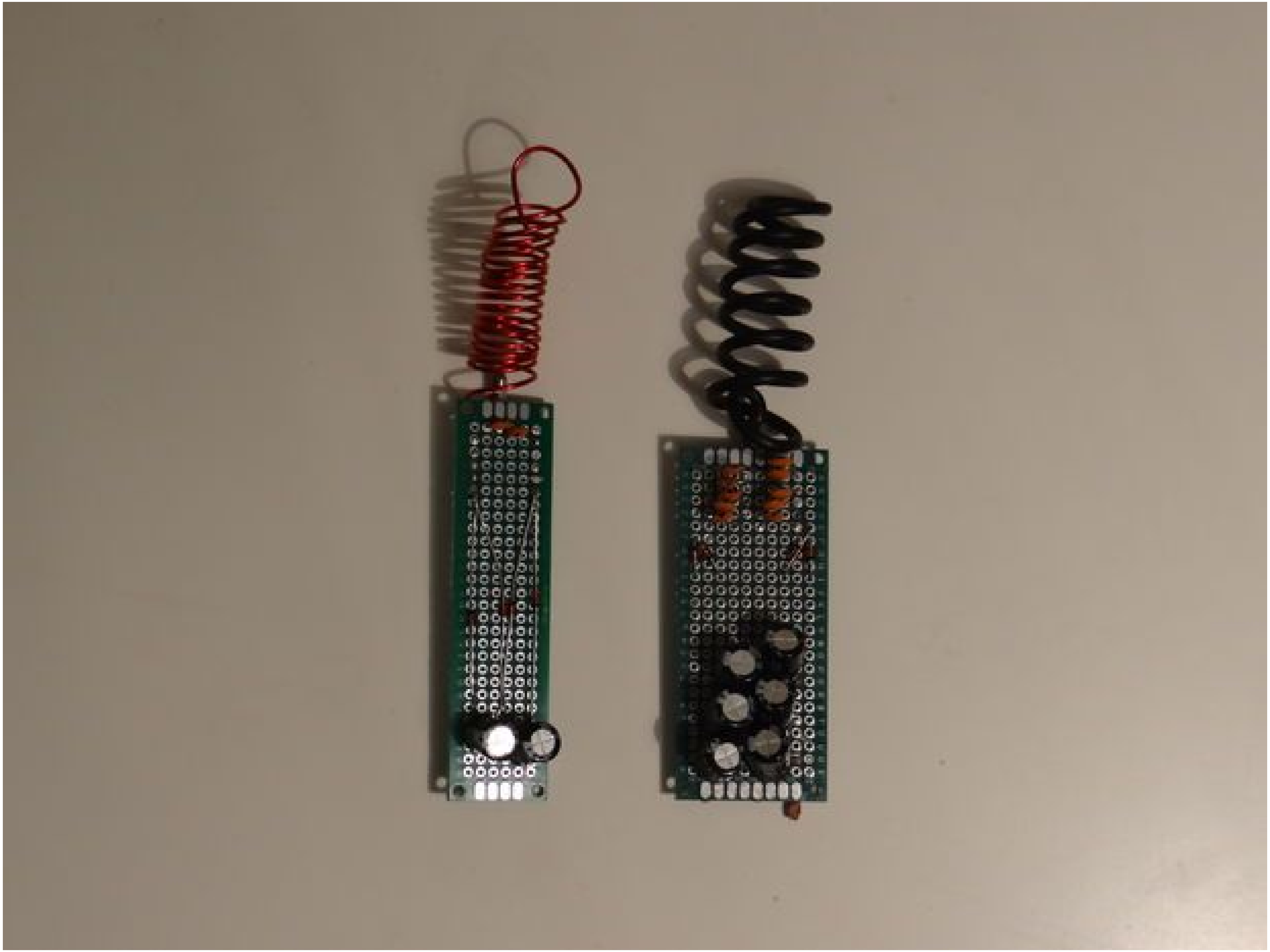

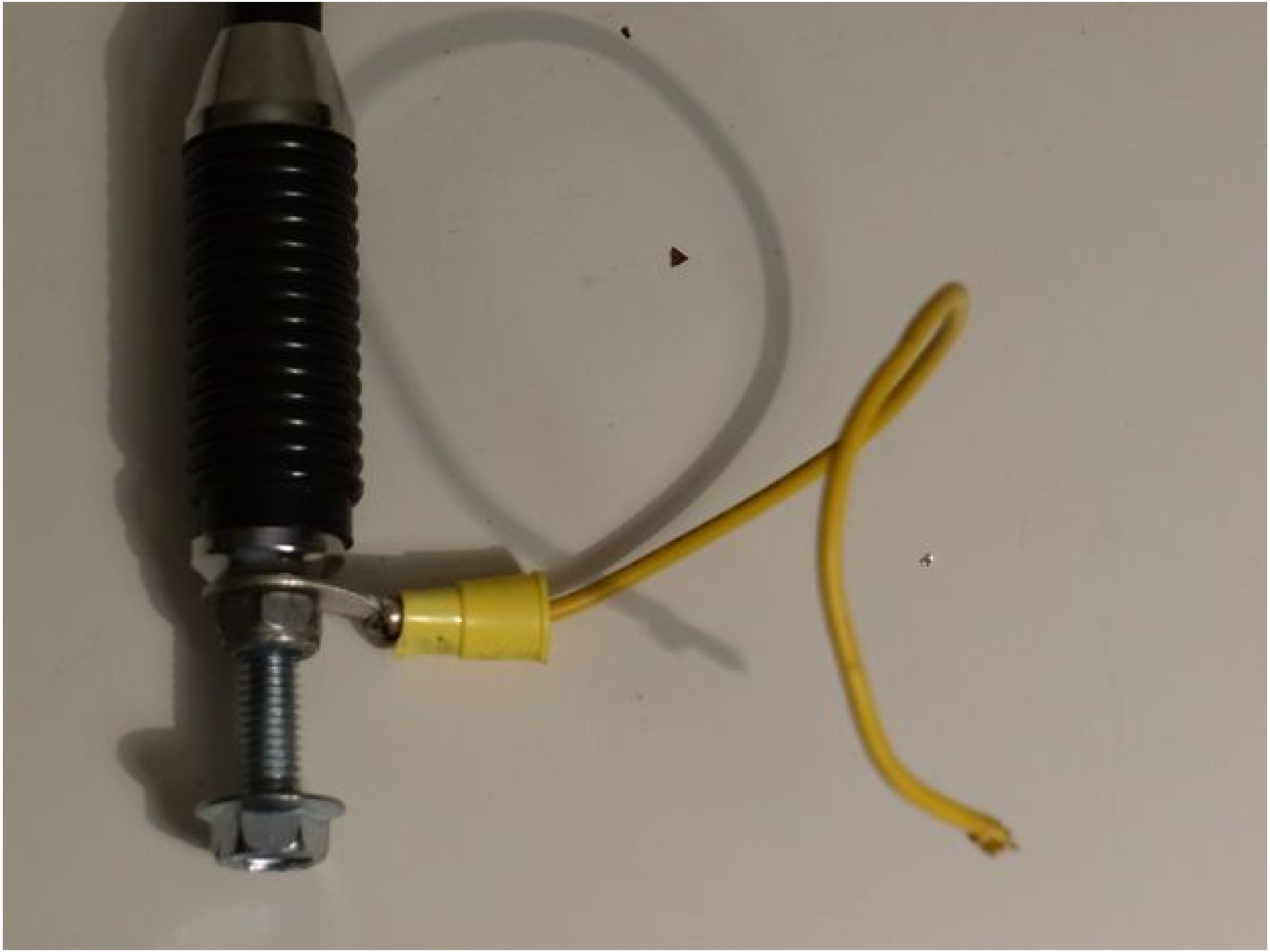

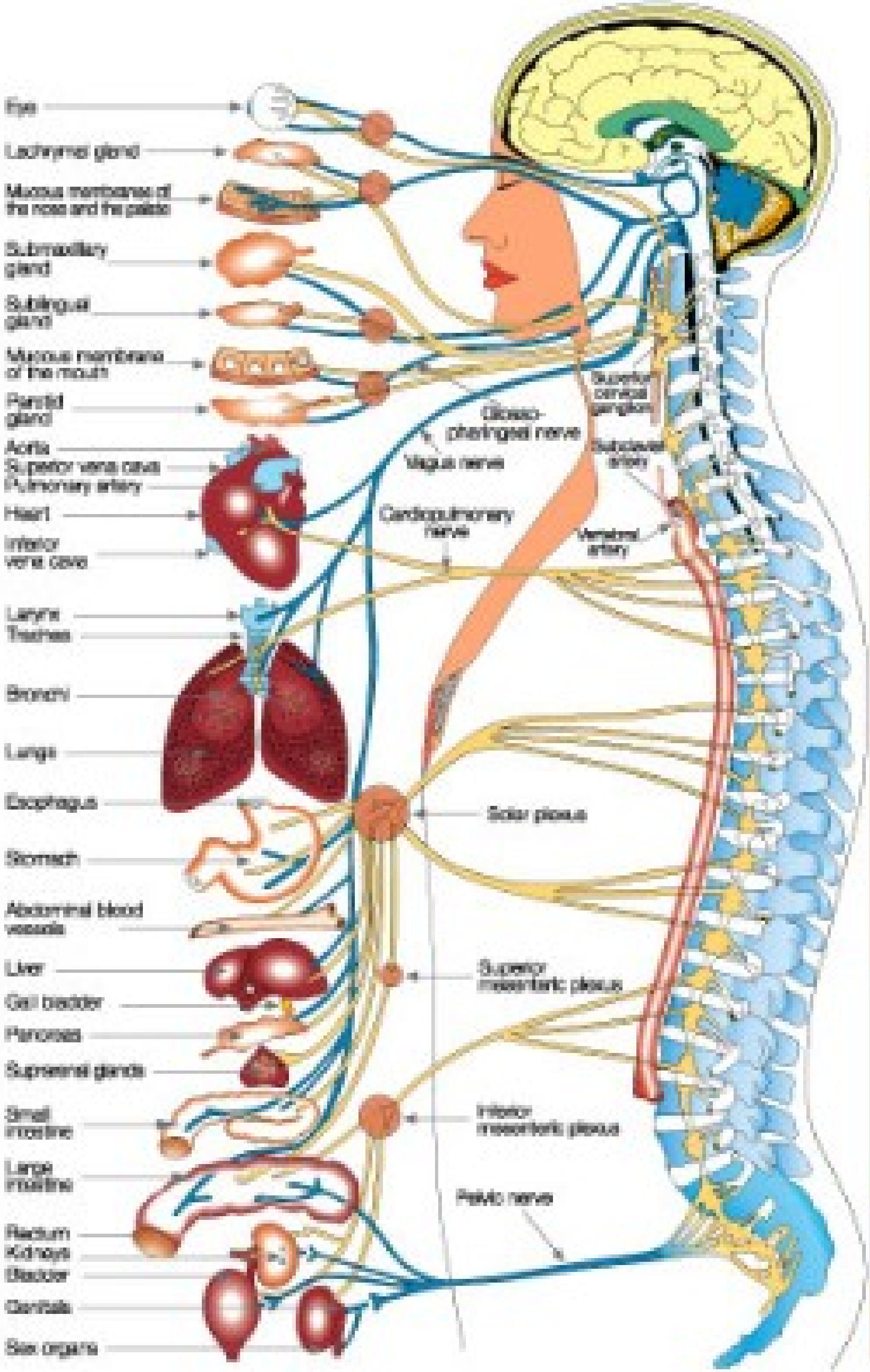
Eye
Lachrymal gland
Mucous membranes of the nose and the palate
Submaxillary gland
Sublingual gland
Mucous membrane of the mouth
Parotid gland
Aorta
Superior vena cava
Pulmonary artery
Heart
Inferior vena cava
Larynx
Trachea
Bronchi
Lungs
Esophagus
Stomach
Abdominal blood vessels
Liver
Gall bladder
Pancreas
Suprarenal glands
Small intestine
Large intestine
Rectum
Kidneys
Bladder
Genitals
Sex organs
Glosso-pharyngeal nerve
Vagus nerve
Superior cervical ganglion
Subclavian artery
Cardiopulmonary nerve
Vertebral artery
Solar plexus
Superior mesenteric plexus
Inferior mesenteric plexus
Pelvic nerve

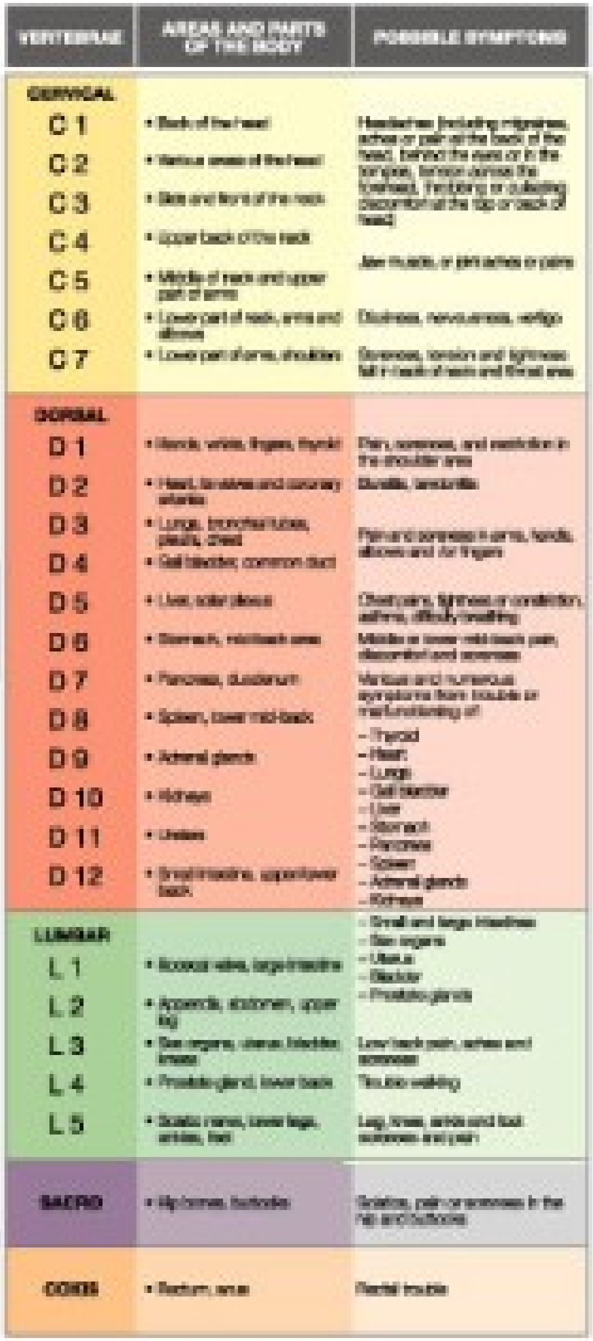

VERTEBRAE	AREAS AND PARTS OF THE BODY	POSSIBLE SYMPTOMS
CERVICAL		
C 1	• Back of the head	Headaches (including migraines, aches or pain at the back of the head, behind the eyes or in the temples, tension across the forehead, throbbing or pulsating discomfort at the top or back of head)
C 2	• Various areas of the head	
C 3	• Side and front of the neck	
C 4	• Upper back of the neck	
C 5	• Middle of neck and upper part of arms	Jaw muscle, or joint aches or pains
C 6	• Lower part of neck, arms and elbows	Dizziness, nervousness, vertigo
C 7	• Lower part of arms, shoulders	Soreness, tension and tightness felt in back of neck and throat area
DORSAL		
D 1	• Hands, wrists, fingers, thyroid	Pain, soreness, and restriction in the shoulder area
D 2	• Heart, its valves and coronary arteries	Bursitis, tendonitis
D 3	• Lungs, bronchial tubes, pleura, chest	Pain and soreness in arms, hands, elbows and /or fingers
D 4	• Gall bladder, common duct	
D 5	• Liver, solar plexus	Chest pains, tightness or constriction, asthma, difficulty breathing
D 6	• Stomach, mid-back area	Middle or lower mid-back pain, discomfort and soreness
D 7	• Pancreas, duodenum	Various and numerous symptoms from trouble or malfunctioning of:
D 8	• Spleen, lower mid-back	
D 9	• Adrenal glands	– Thyroid – Heart – Lungs – Gall bladder – Liver – Stomach – Pancreas – Spleen – Adrenal glands – Kidneys
D 10	• Kidneys	
D 11	• Ureters	
D 12	• Small intestine, upper/lower back	
LUMBAR		– Small and large intestines – Sex organs – Uterus – Bladder – Prostate glands
L 1	• Ileocecal valve, large intestine	
L 2	• Appendix, abdomen, upper leg	
L 3	• Sex organs, uterus, bladder, knees	Low back pain, aches and soreness
L 4	• Prostate gland, lower back	Trouble walking
L 5	• Sciatic nerve, lower legs, ankles, feet	Leg, knee, ankle and foot weakness and pain
SACRO	• Hip bones, buttocks	Sciatica, pain or soreness in the hip and buttocks
COXIS	• Rectum, anus	Rectal trouble

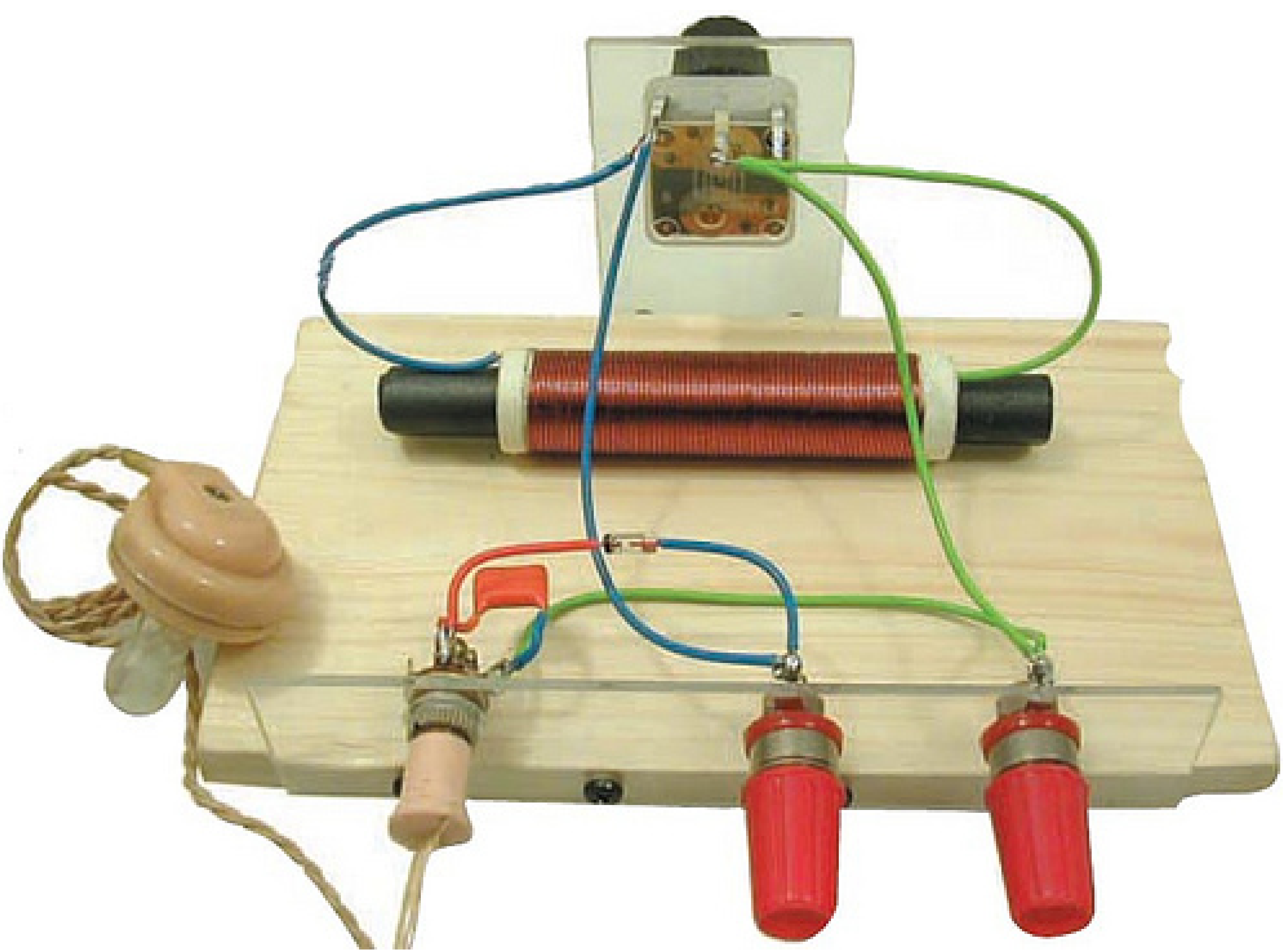

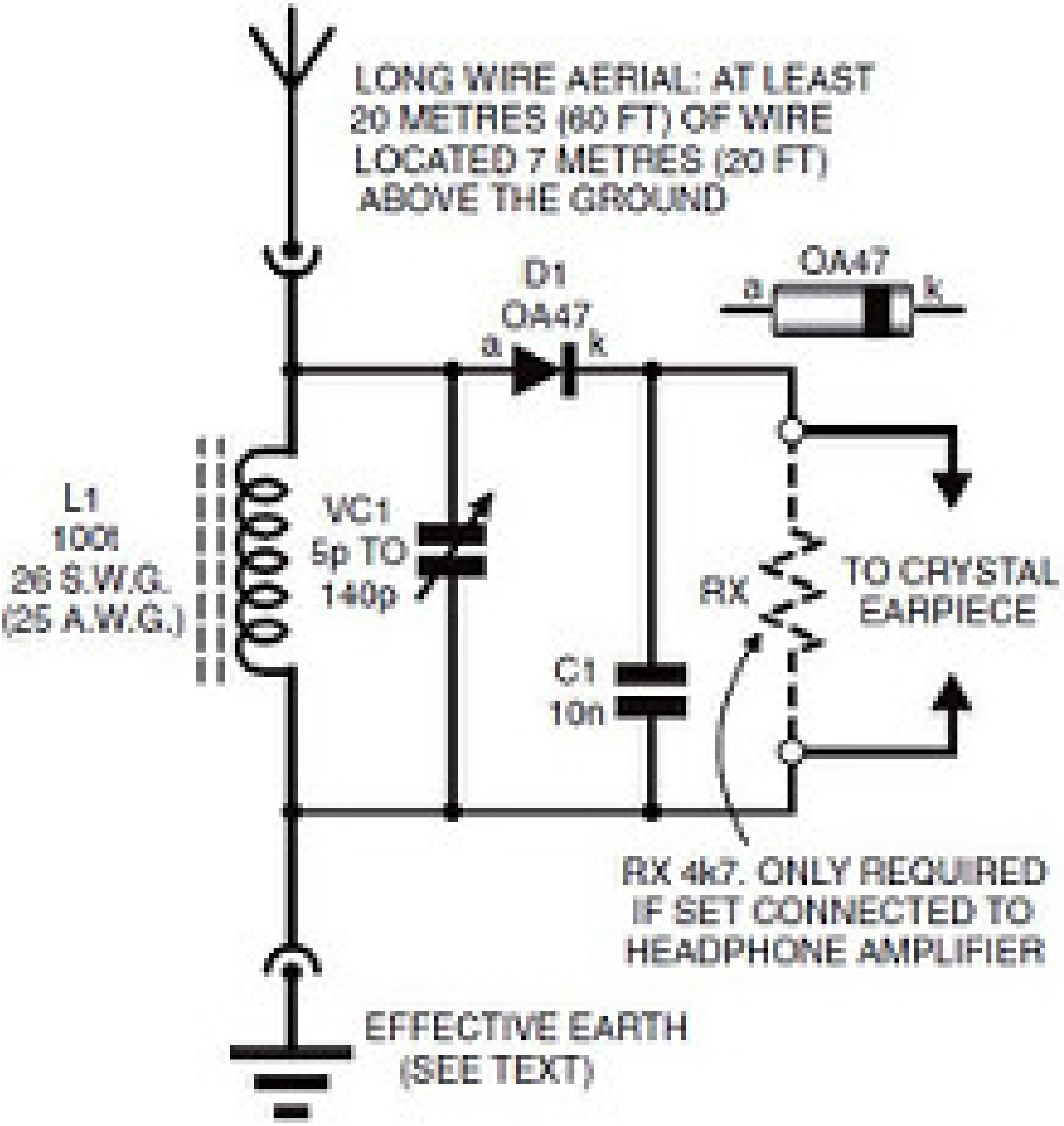
LONG WIRE AERIAL: AT LEAST
20 METRES (60 FT) OF WIRE
LOCATED 7 METRES (20 FT)
ABOVE THE GROUND
D1
OA47
a
k
OA47
a
k
L1
100t
26 S.W.G.
(25 A.W.G.)
VC1
5p TO
140p
C1
10n
RX
TO CRYSTAL
EARPIECE
RX 4k7. ONLY REQUIRED
IF SET CONNECTED TO
HEADPHONE AMPLIFIER
EFFECTIVE EARTH
(SEE TEXT)

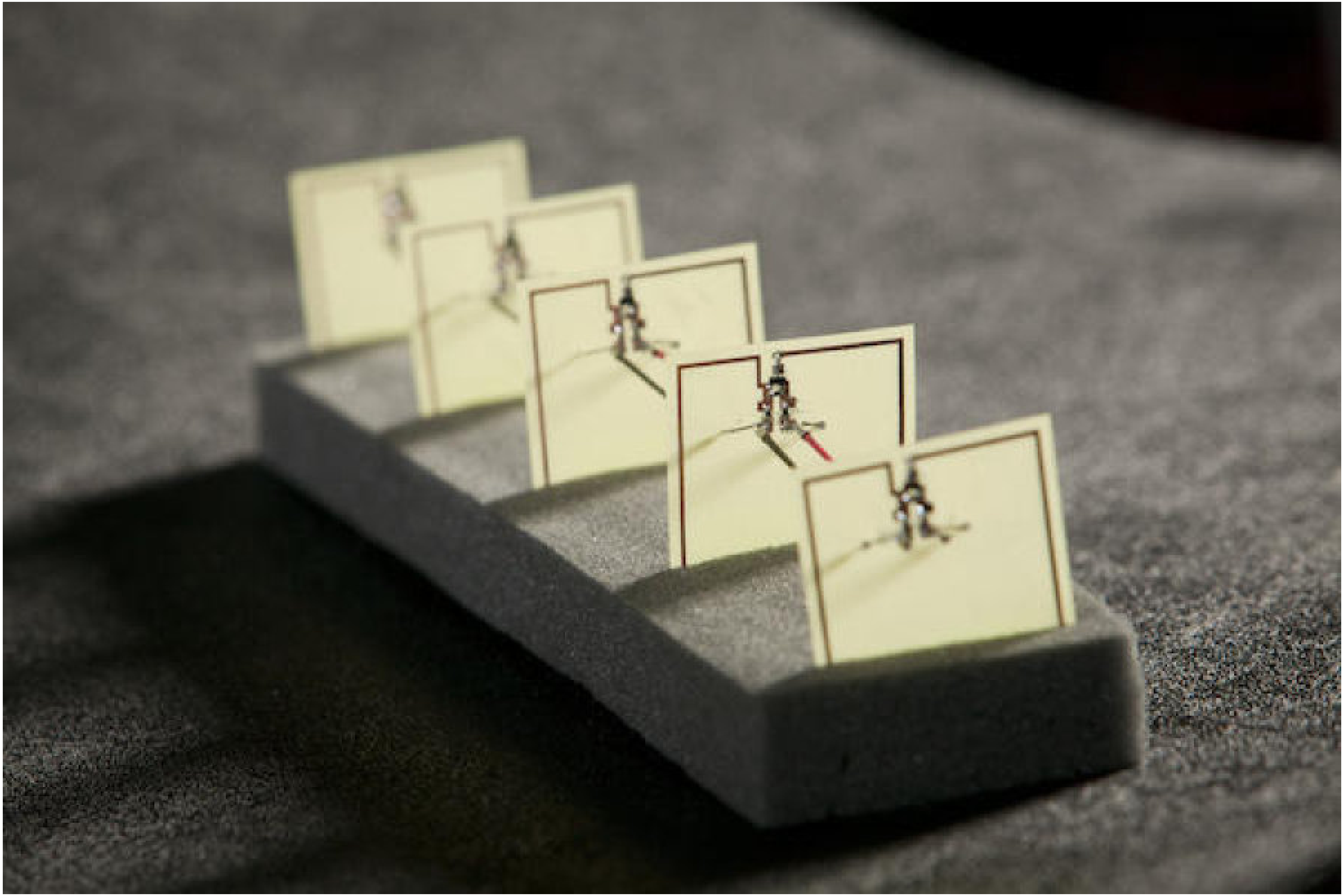

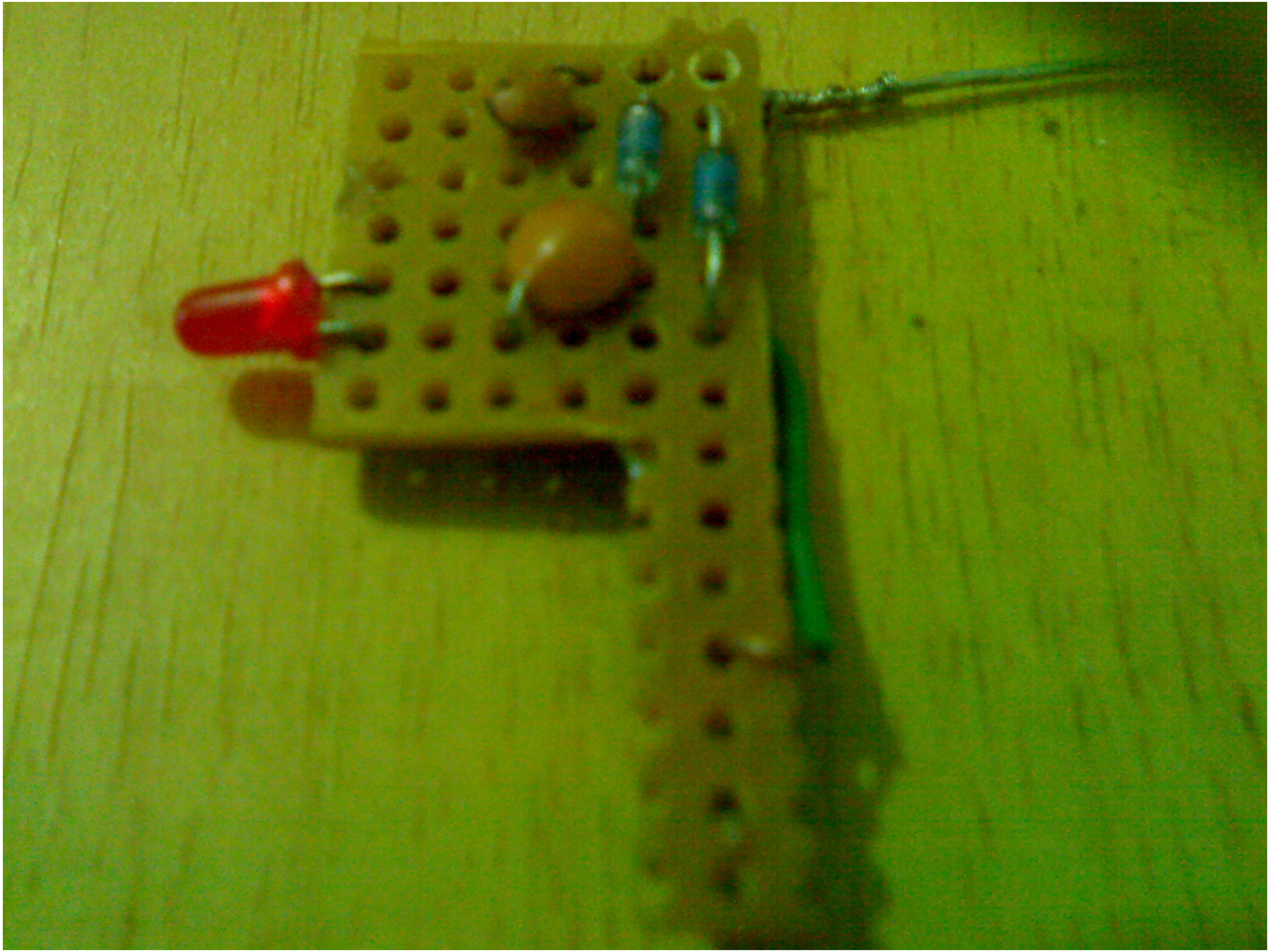

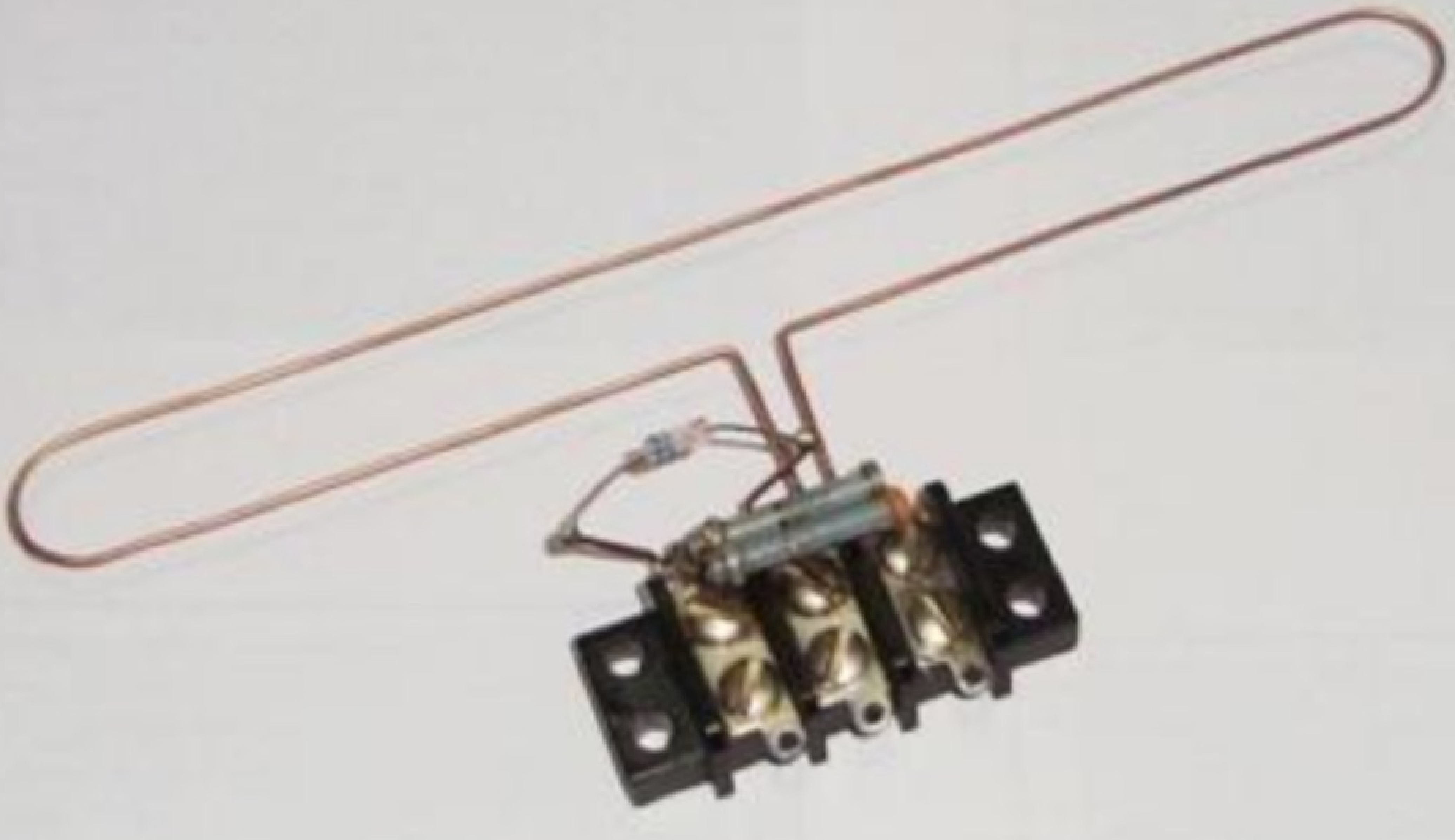

28.9 °C
26%
°C / °F
MAX / MIN
MEM CLEAR

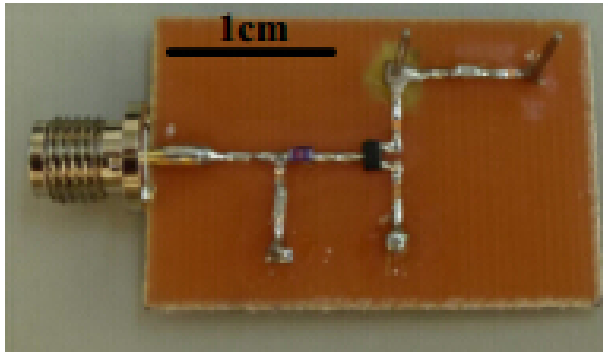
1cm

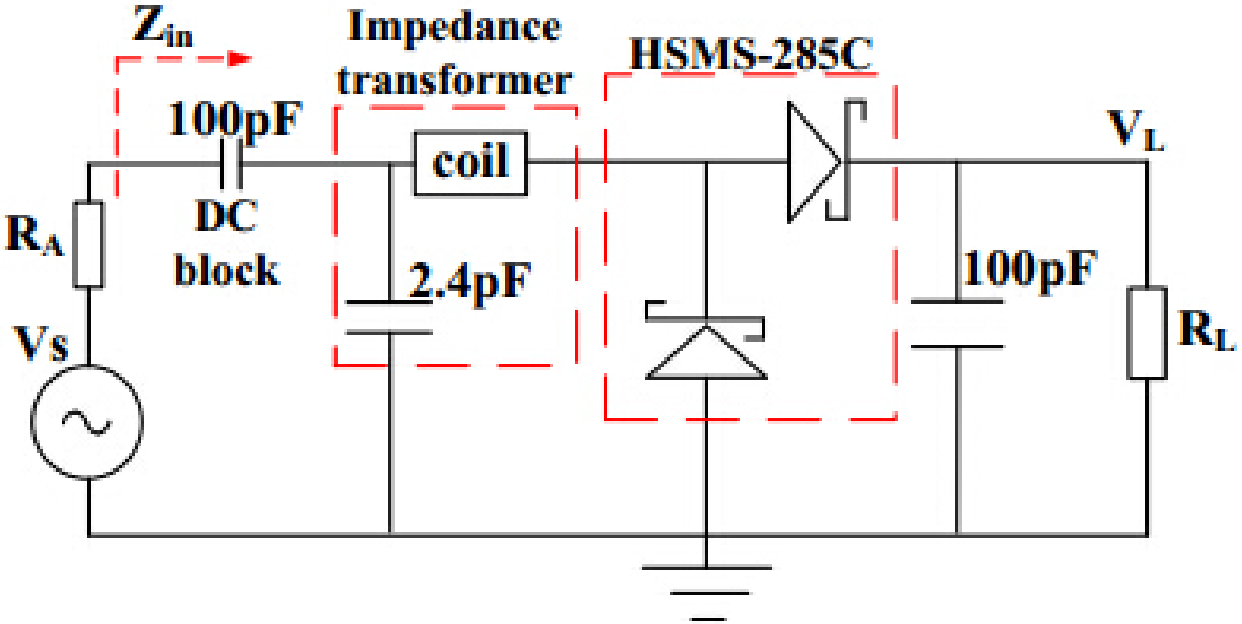
Z_{in}
Impedance
transformer
HSMS-285C
100pF
DC
block
coil
V_L
R_A
2.4pF
100pF
R_L
Vs

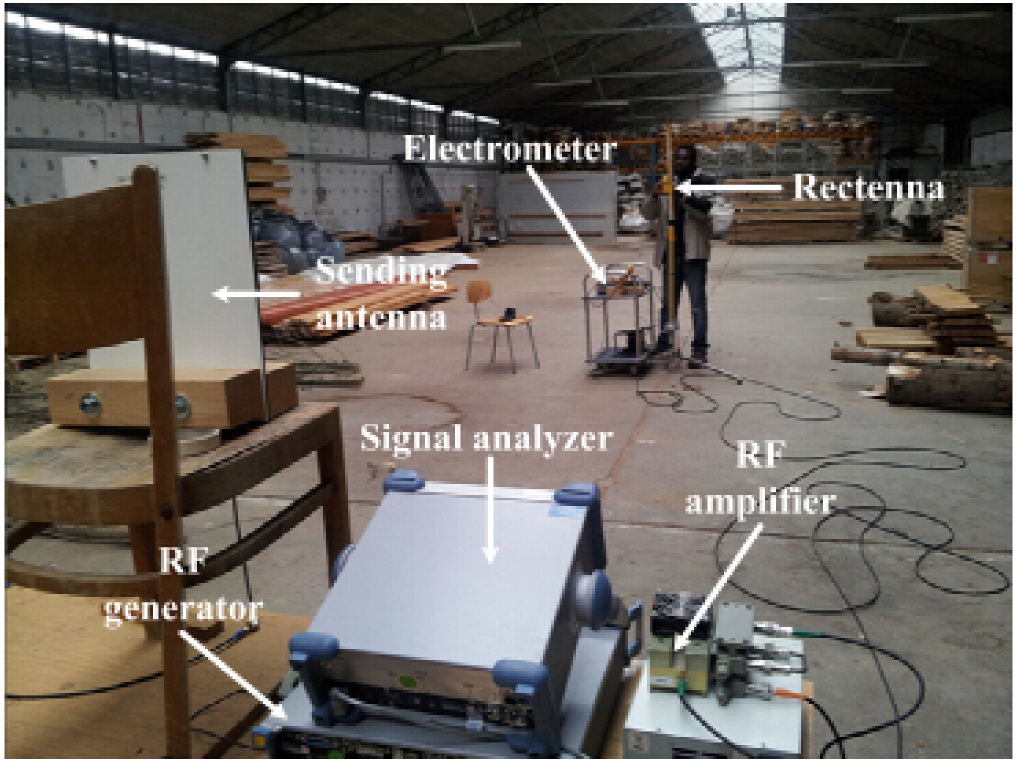
Electrometer
Rectenna
Sending antenna
Signal analyzer
RF amplifier
RF generator

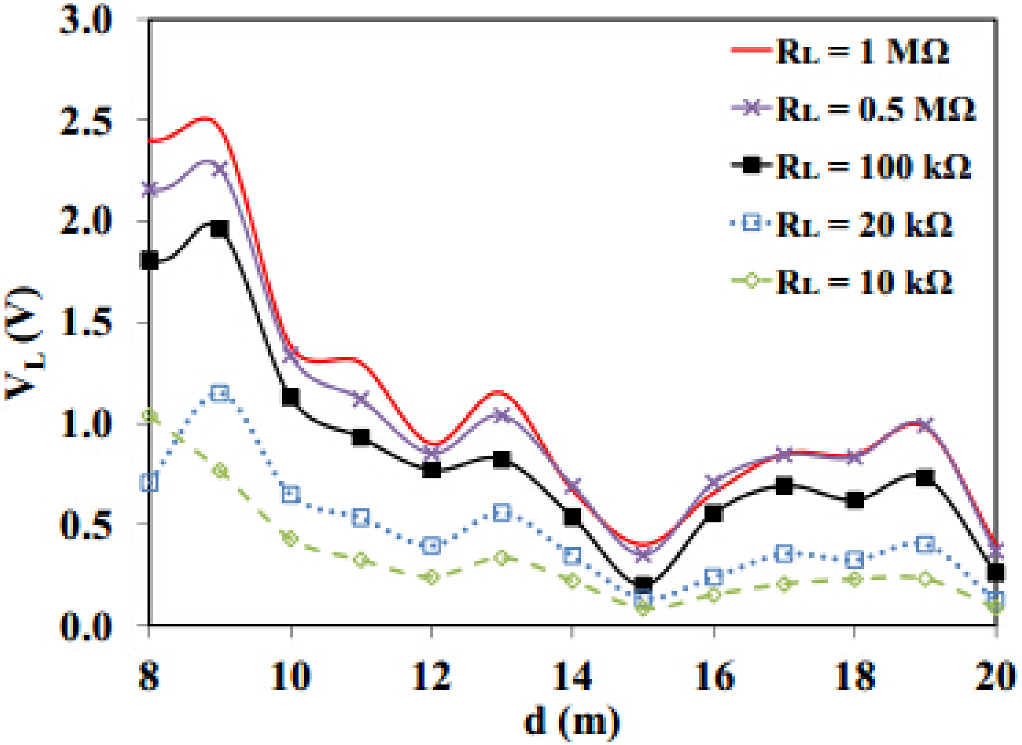
3.0
2.5
2.0
1.5
1.0
0.5
0.0
VL (V)
8
10
12
14
16
18
20
d (m)
RL = 1 MΩ
RL = 0.5 MΩ
RL = 100 kΩ
RL = 20 kΩ
RL = 10 kΩ

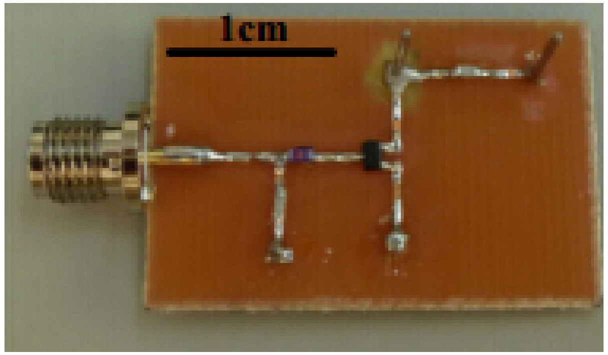
1cm

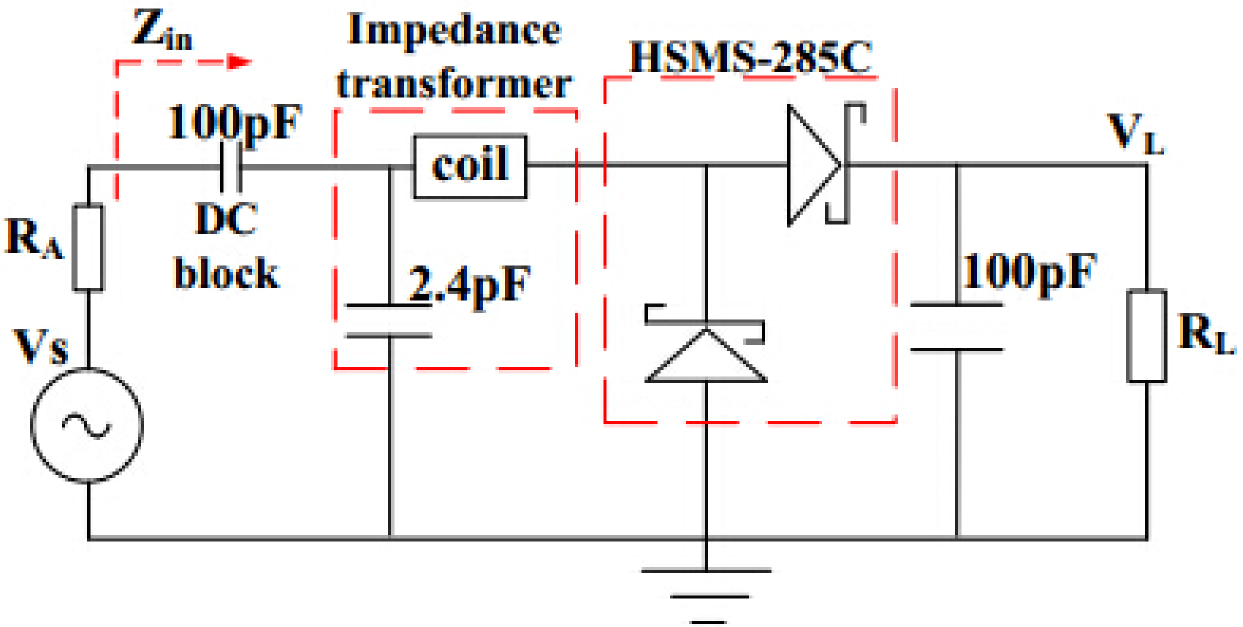
Z_in
Impedance transformer
HSMS-285C
100pF
coil
V_L
R_A
DC block
2.4pF
100pF
Vs
R_L

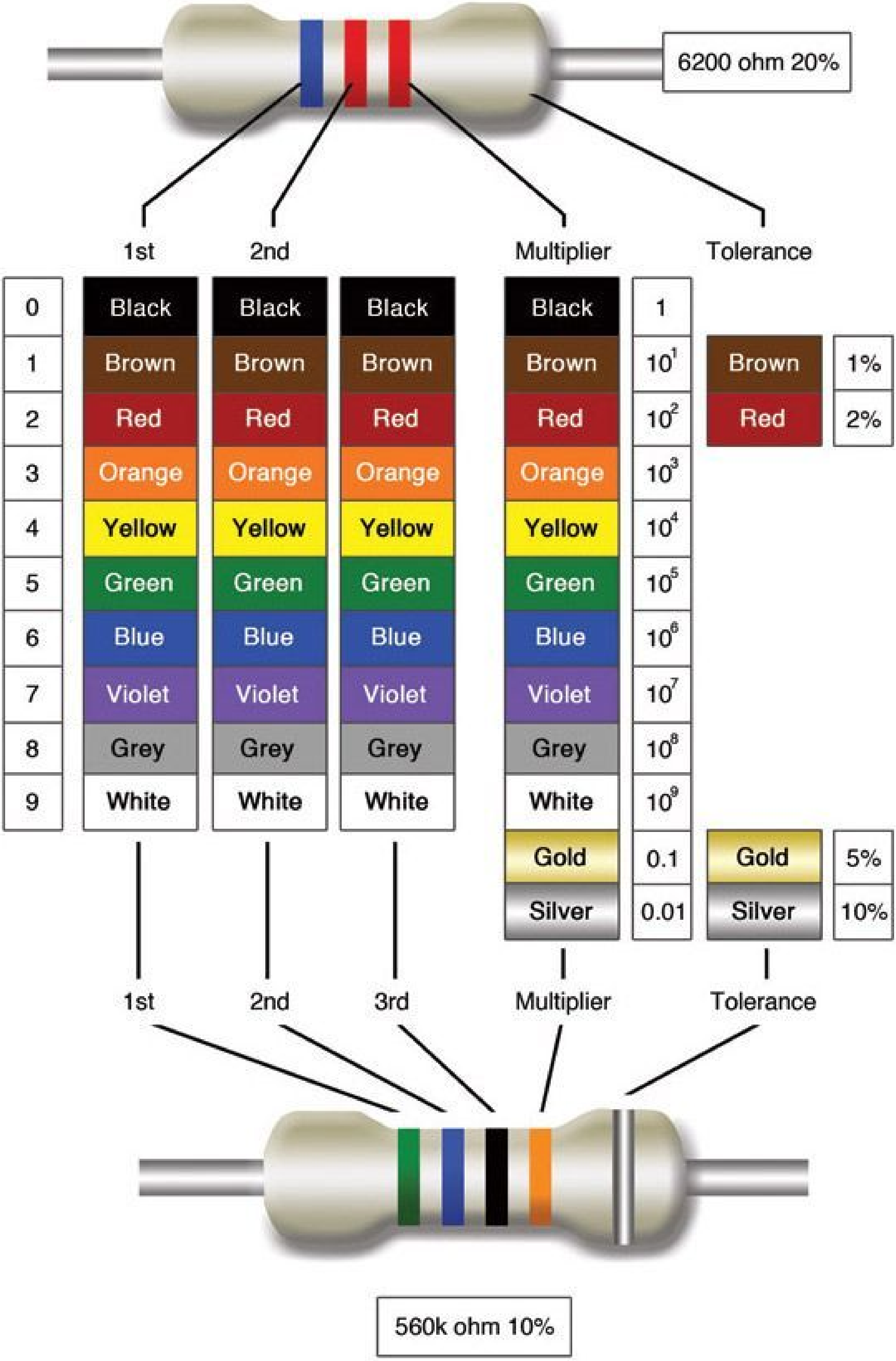
6200 ohm 20%
1st
2nd
Multiplier
Tolerance
560k ohm 10%

	1st	2nd	3rd	Multiplier		Tolerance	
0	Black	Black	Black	Black	1		
1	Brown	Brown	Brown	Brown	10^1	Brown	1%
2	Red	Red	Red	Red	10^2	Red	2%
3	Orange	Orange	Orange	Orange	10^3		
4	Yellow	Yellow	Yellow	Yellow	10^4		
5	Green	Green	Green	Green	10^5		
6	Blue	Blue	Blue	Blue	10^6		
7	Violet	Violet	Violet	Violet	10^7		
8	Grey	Grey	Grey	Grey	10^8		
9	White	White	White	White	10^9		
				Gold	0.1	Gold	5%
				Silver	0.01	Silver	10%

FIG. 9 TERMINAL STRIP

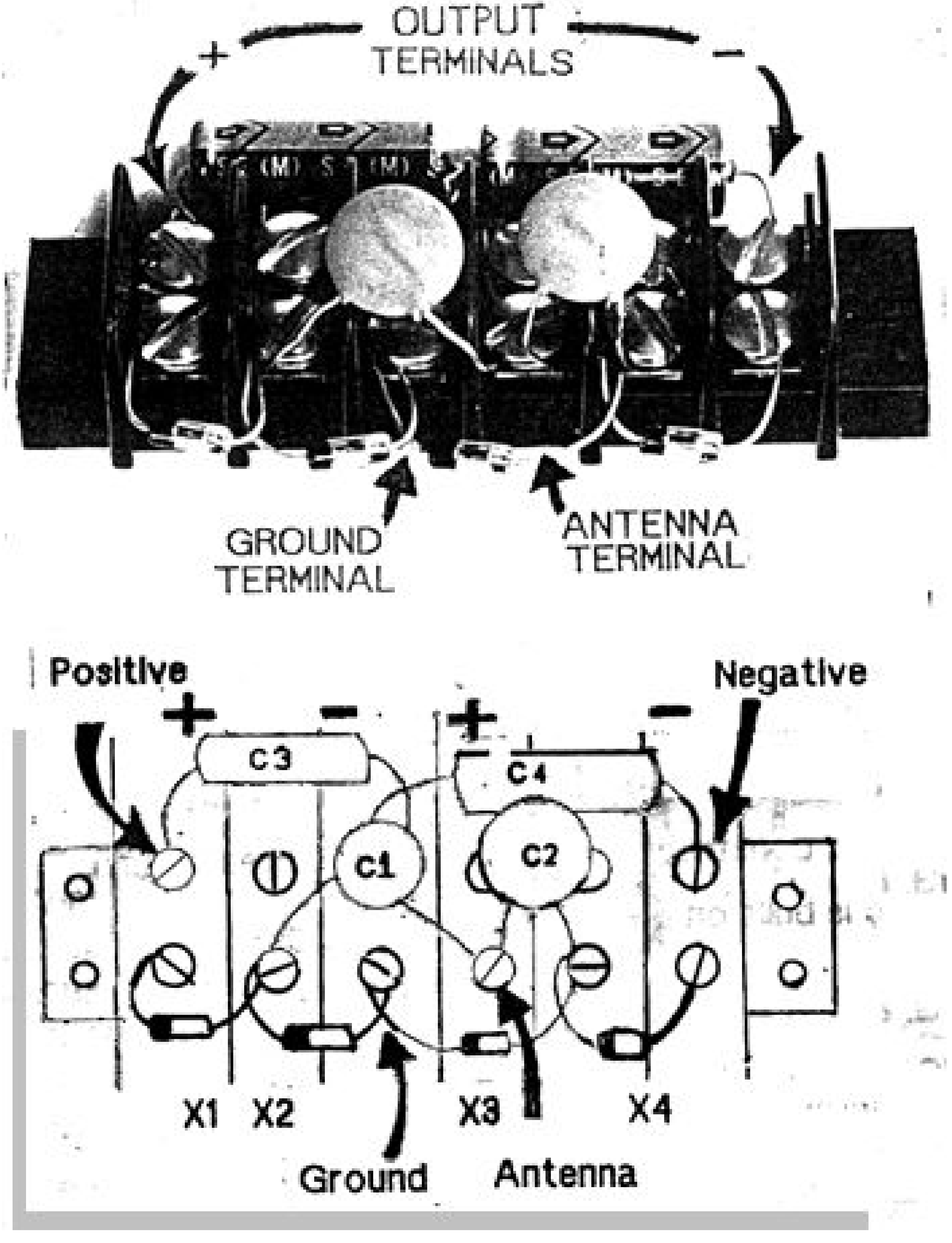

+
OUTPUT TERMINALS
−
GROUND TERMINAL
ANTENNA TERMINAL
Positive
Negative
+
−
+
−
C3
C4
C1
C2
X1
X2
X3
X4
Ground
Antenna

FIG. 9 TERMINAL STRIP

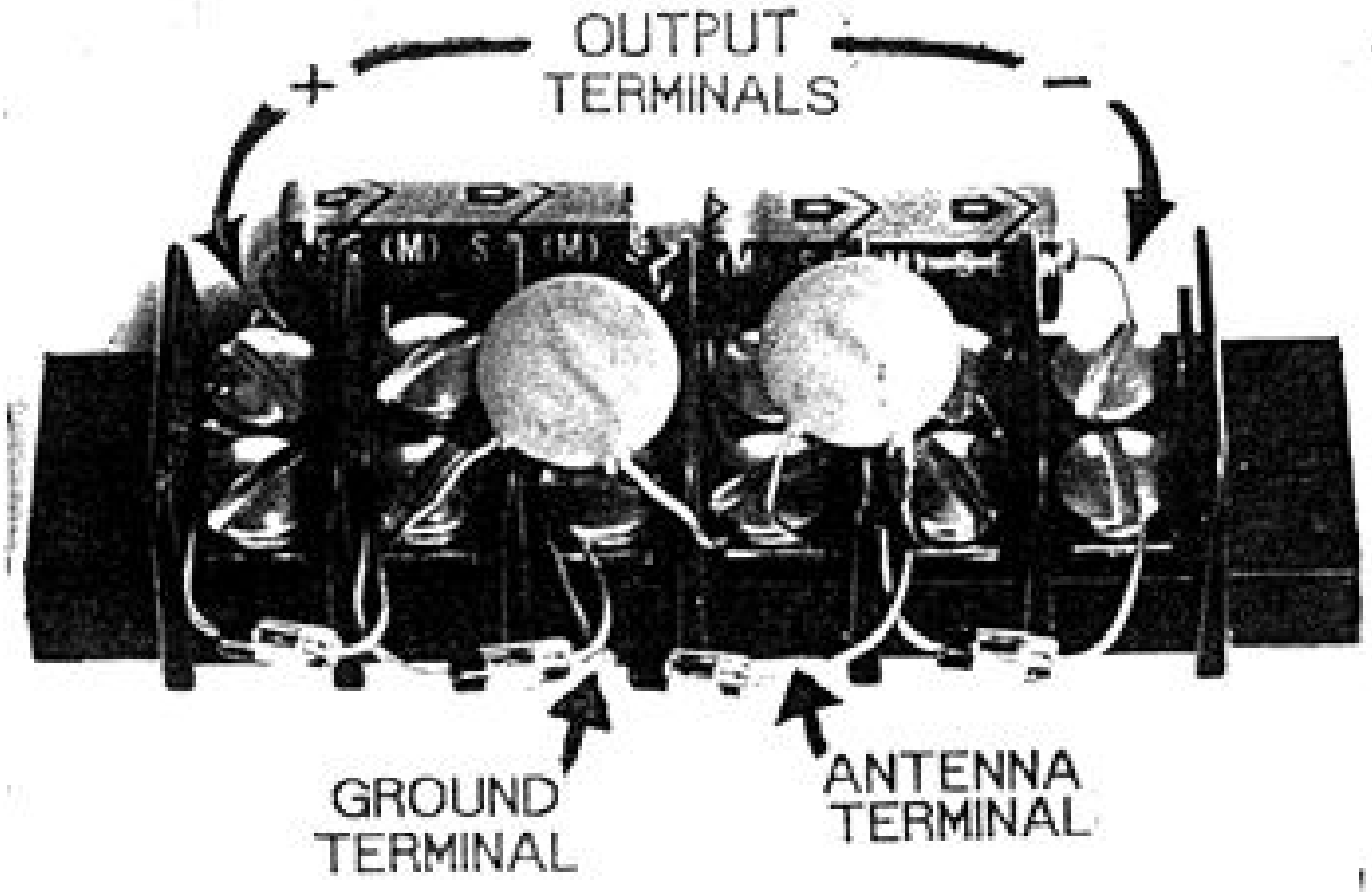
+
OUTPUT TERMINALS
−
GROUND TERMINAL
ANTENNA TERMINAL

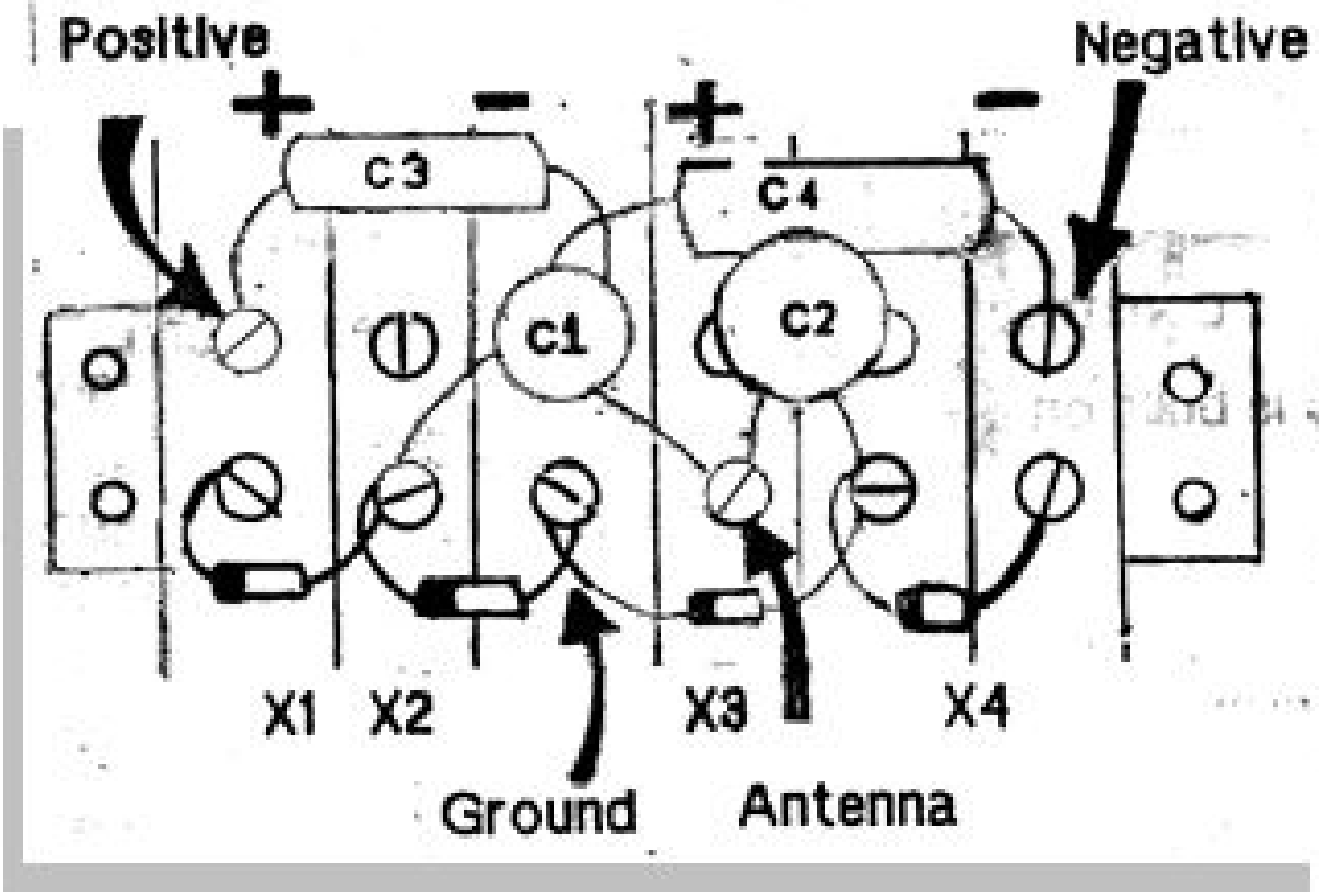
Positive
+
−
+
−
Negative
C3
C4
C1
C2
X1
X2
X3
X4
Ground
Antenna

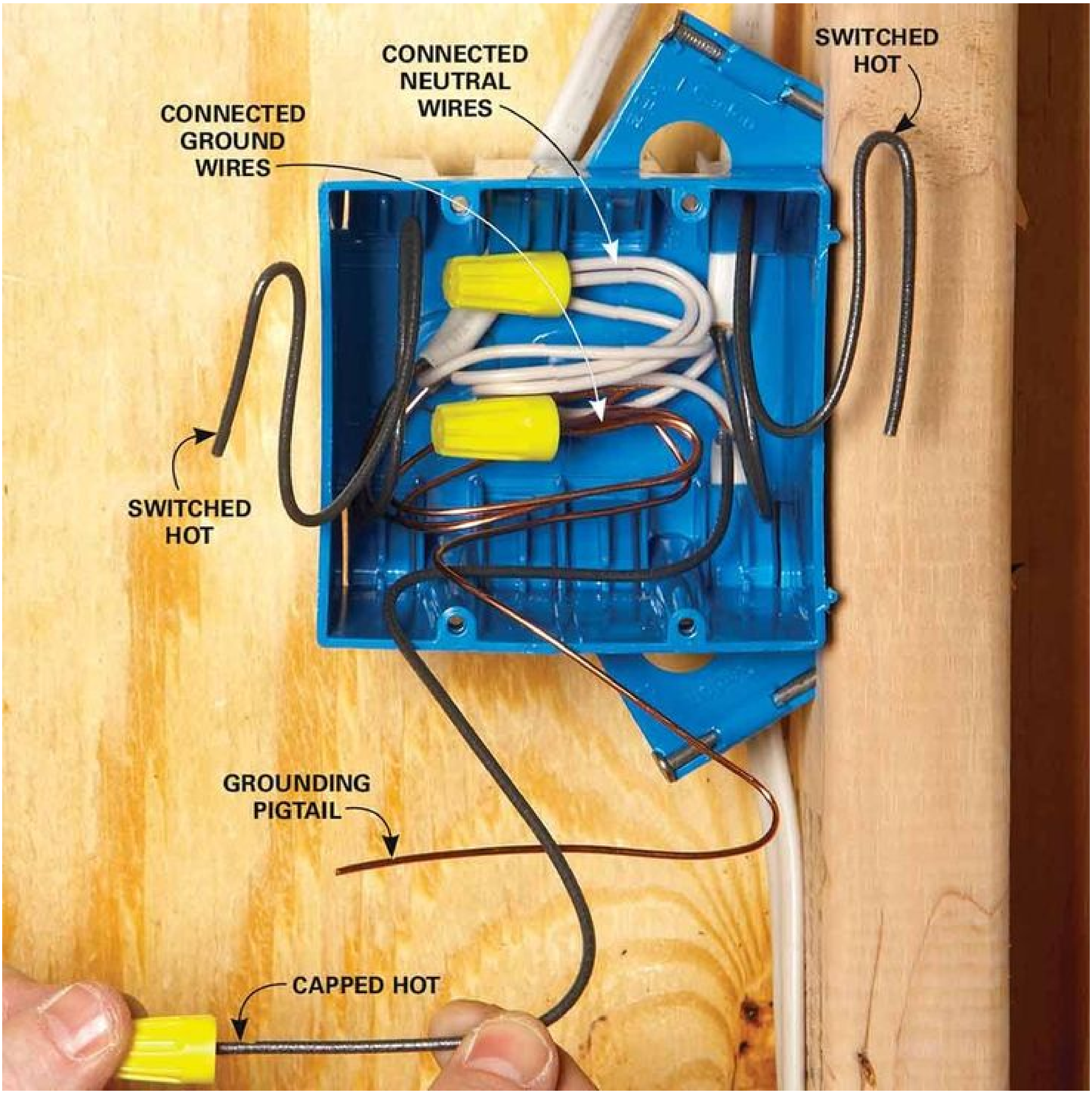
CONNECTED NEUTRAL WIRES
CONNECTED GROUND WIRES
SWITCHED HOT
SWITCHED HOT
GROUNDING PIGTAIL
CAPPED HOT

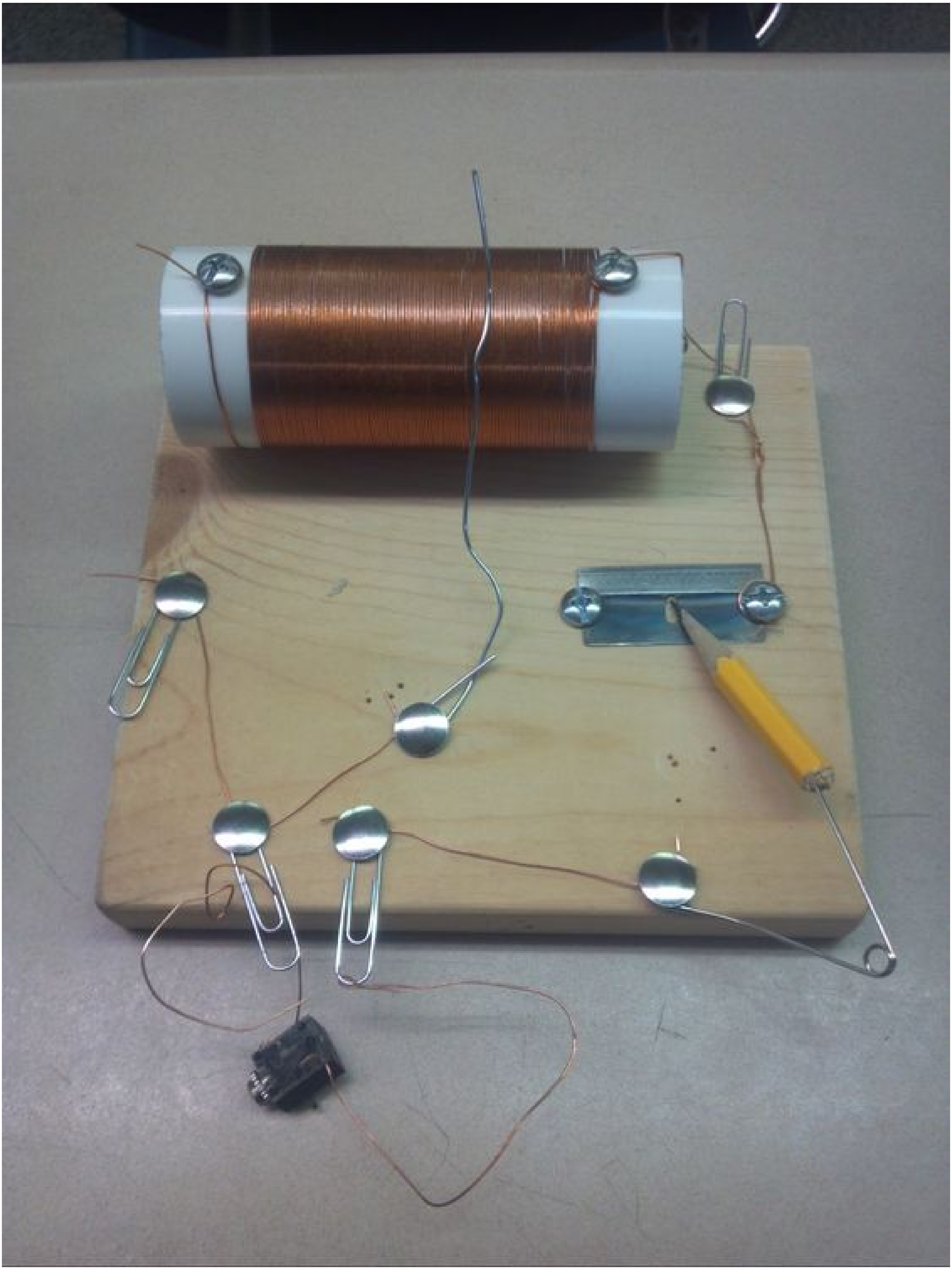

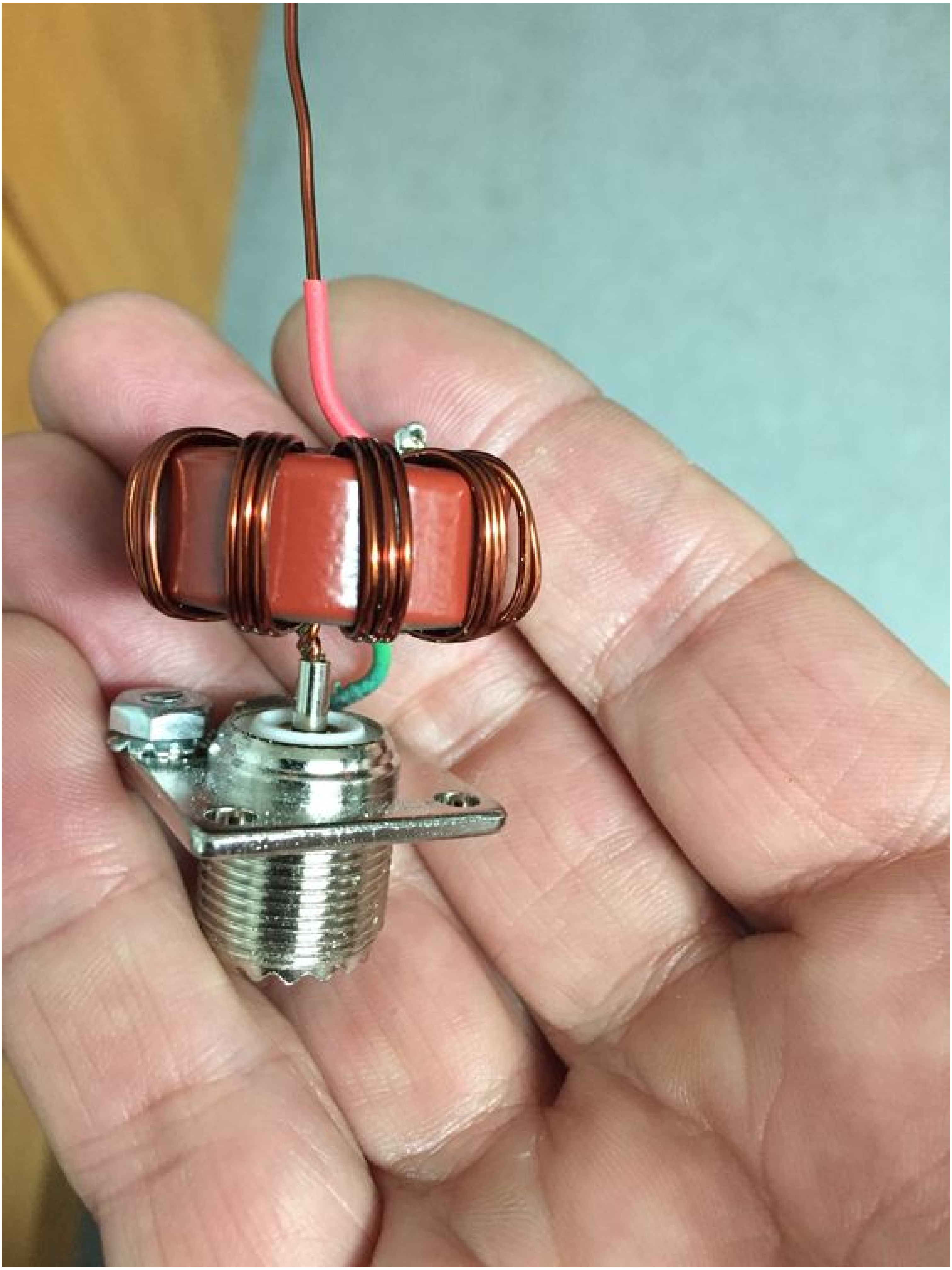

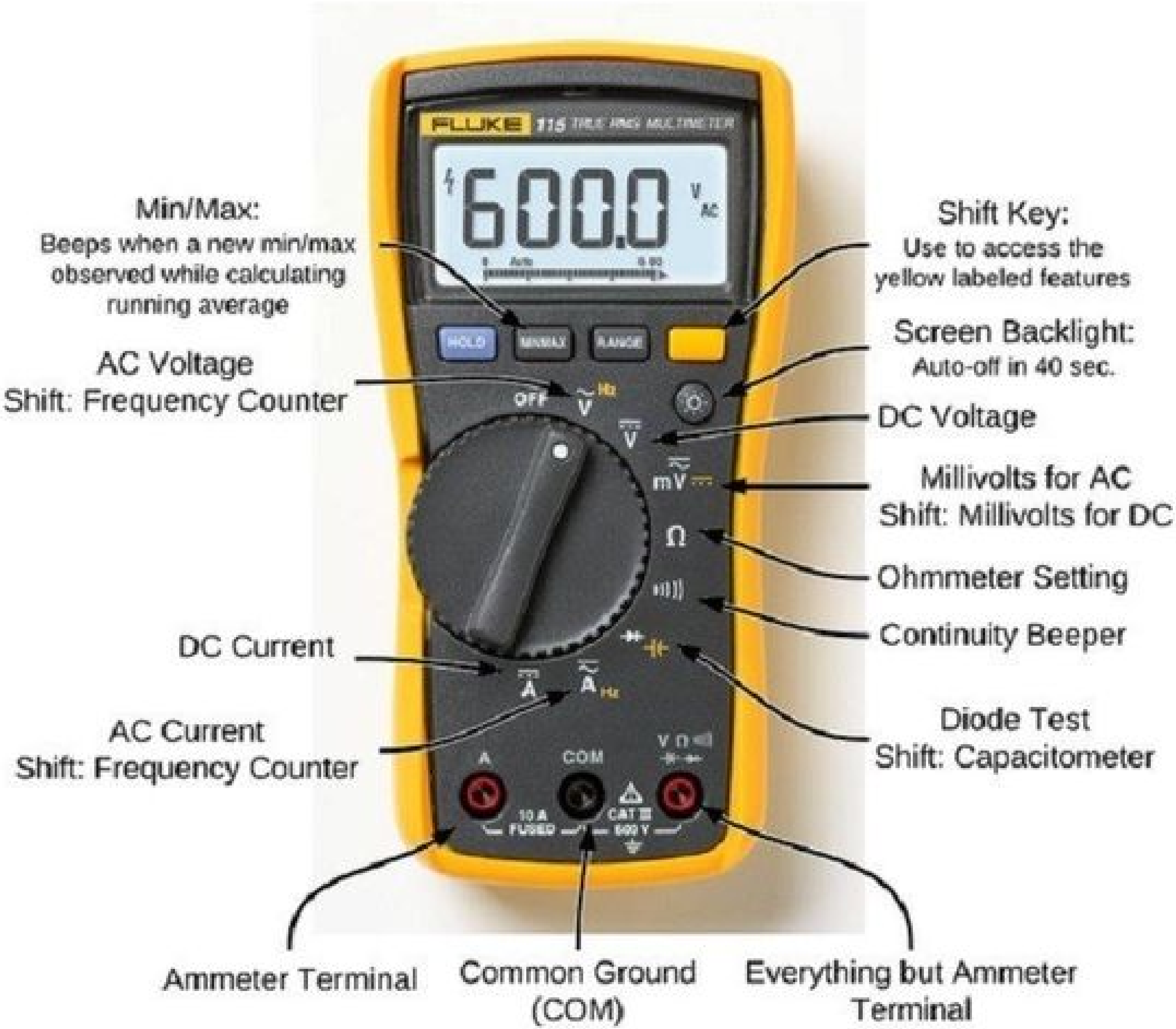
FLUKE 115 TRUE RMS MULTIMETER
600.0 V AC
HOLD
MIN MAX
RANGE
OFF
A
COM
10 A FUSED
CAT III 600 V

Min/Max:
Beeps when a new min/max observed while calculating running average

Shift Key:
Use to access the yellow labeled features

Screen Backlight:
Auto-off in 40 sec.

AC Voltage
Shift: Frequency Counter

DC Voltage

Millivolts for AC
Shift: Millivolts for DC

Ohmmeter Setting

Continuity Beeper

DC Current

AC Current
Shift: Frequency Counter

Diode Test
Shift: Capacitometer

Ammeter Terminal

Common Ground (COM)

Everything but Ammeter Terminal

MP1RP Specifications

Bands:
40m-30m-20m-17m
15m-12m-11m-10m*

Frequency Range:
HF 7MHz~30MHz
continuous

* Optional SuperCoil MC80 for 80 meter 75 meter and MC60 for 60m bands are available separately

Mil-Spec Titanium
ship survives
real world field
deployments
Length 44: (1.1m)

Good SWR: 1.5 : 1 or better, no analyzer or tuner needed

Power: 500W SSB
300W CW / DIGITAL

Antenna weight:
14 ounces (370grams)

Fittings: Standard 3/8"-24

Color: Natural
Metallic and Black

Assembly:
2 minutes
no tools needed

Collapsed Size:
Entire antenna
packs down to
12" (30cm) circle
for portability

Max Size:
Extends to
7ft (2.1m) long
for operation

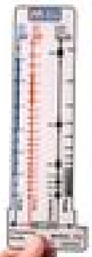

Includes FG1 Frequency Guide SWR Ruler for Band Change

Mounts to many types of backpacks and surfaces

Contents:
1 MP1B SuperStick Antenna
1 SW1 SuperWhip Titanium Whip
1 UM1 SuperMount Mount + Clamp
1 FG1 Frequency Guide SWR Ruler
1 MR1C SuperWire Counterpoise

© 2015 Super Antenna Systems Corp. All rights reserved. Super Antenna® is a registered trademark of Super Antenna Systems Corp. SuperWhip™ SuperWire™ SuperMount™ SuperSlider™ are trademarks of, and Get The Power of a Super Antenna℠ is a Service Mark of, Super Antenna Systems Corporation, of USA.

SUPER ANTENNA®

newsuperantenna.com

MP1RP SuperWhip

Ruggedized Antenna for HF / Pedestrian Mobile / Backpack Portable

Get the Power of a SuperWhip.

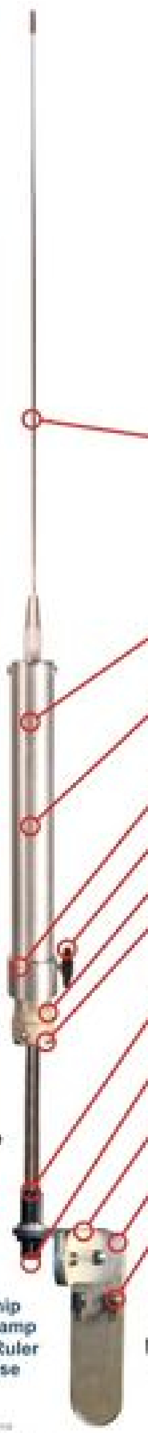

- SW1 SuperWhip
  Super Flexible Ruggedized Titanium Alloy
  Bends in a 12 inch Circle to Fit in Backpack
- All HF Bands - All Frequencies - Good SWR
- 6061-T6 Aircraft Aluminum SuperSlider
- Nickel Beryllium Weatherproof Contacts
- Tuning Lock for Secure Frequency in Motion
- High Q SuperCoil for Maximum RF Efficiency
- Durable Polymer SuperCoil Form
- All Parts Modular Standard 3/8"-24 Fittings
- Standard SO-239 Coaxial Connector
  PL-259 Compatible 50 ohms
- 360 Degree Swivel for All Positions
- UM1 SuperMount
- Dual Counterpoise Connectors
  Quick - Release Push-On
- MR1C SuperWire Stealth Teflon
  Drag Wire Counterpoise with
  Quick - Release Push-On Connectors

ER1 Extension Rod

TW1 Telescopic Whip
for VHF/UHF

CC1 Clamp For Attachments

(backpack not included)

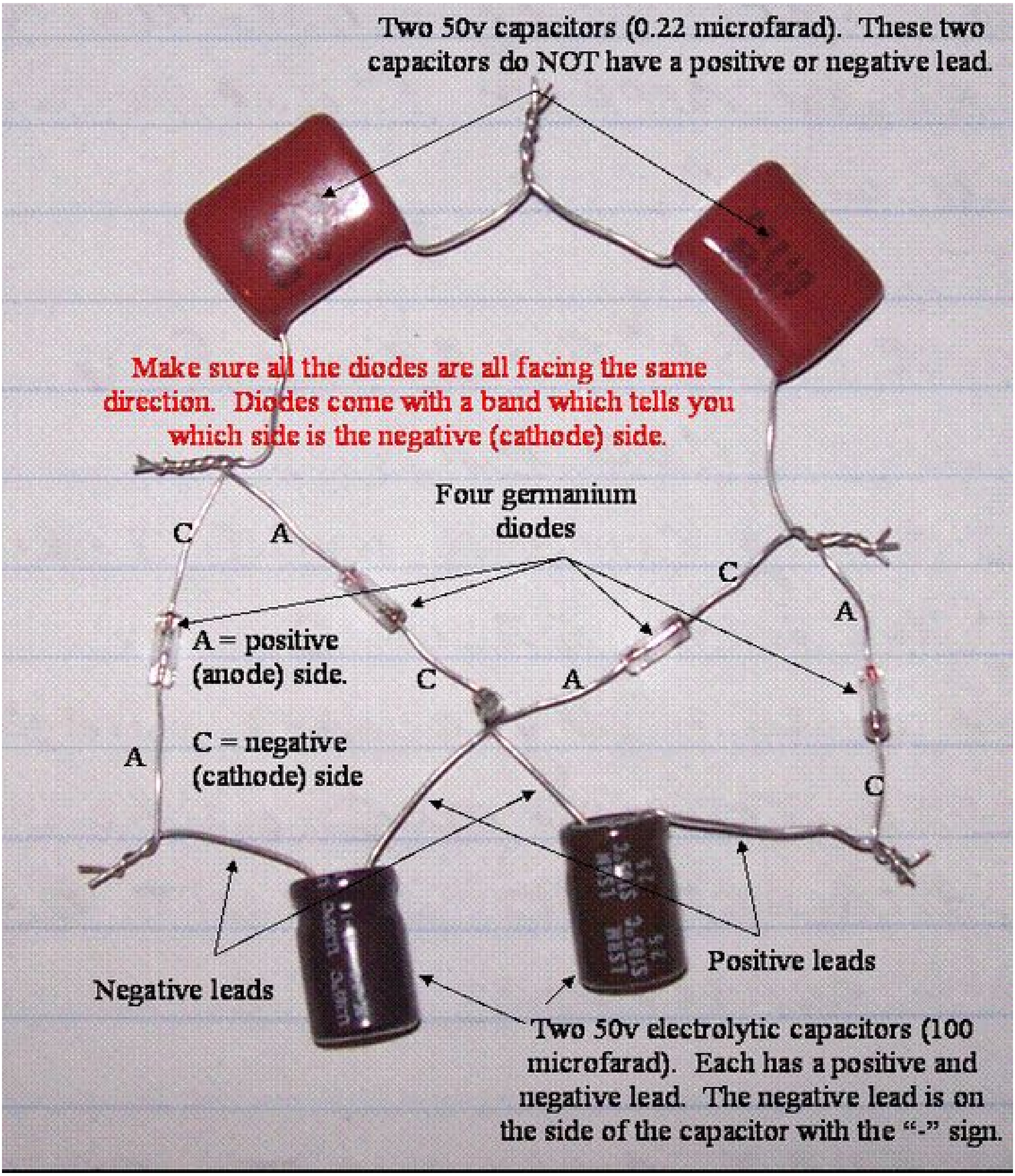

Two 50v capacitors (0.22 microfarad). These two capacitors do NOT have a positive or negative lead.

Make sure all the diodes are all facing the same direction. Diodes come with a band which tells you which side is the negative (cathode) side.

Four germanium diodes

A = positive (anode) side.

C = negative (cathode) side

Negative leads

Positive leads

Two 50v electrolytic capacitors (100 microfarad). Each has a positive and negative lead. The negative lead is on the side of the capacitor with the "-" sign.

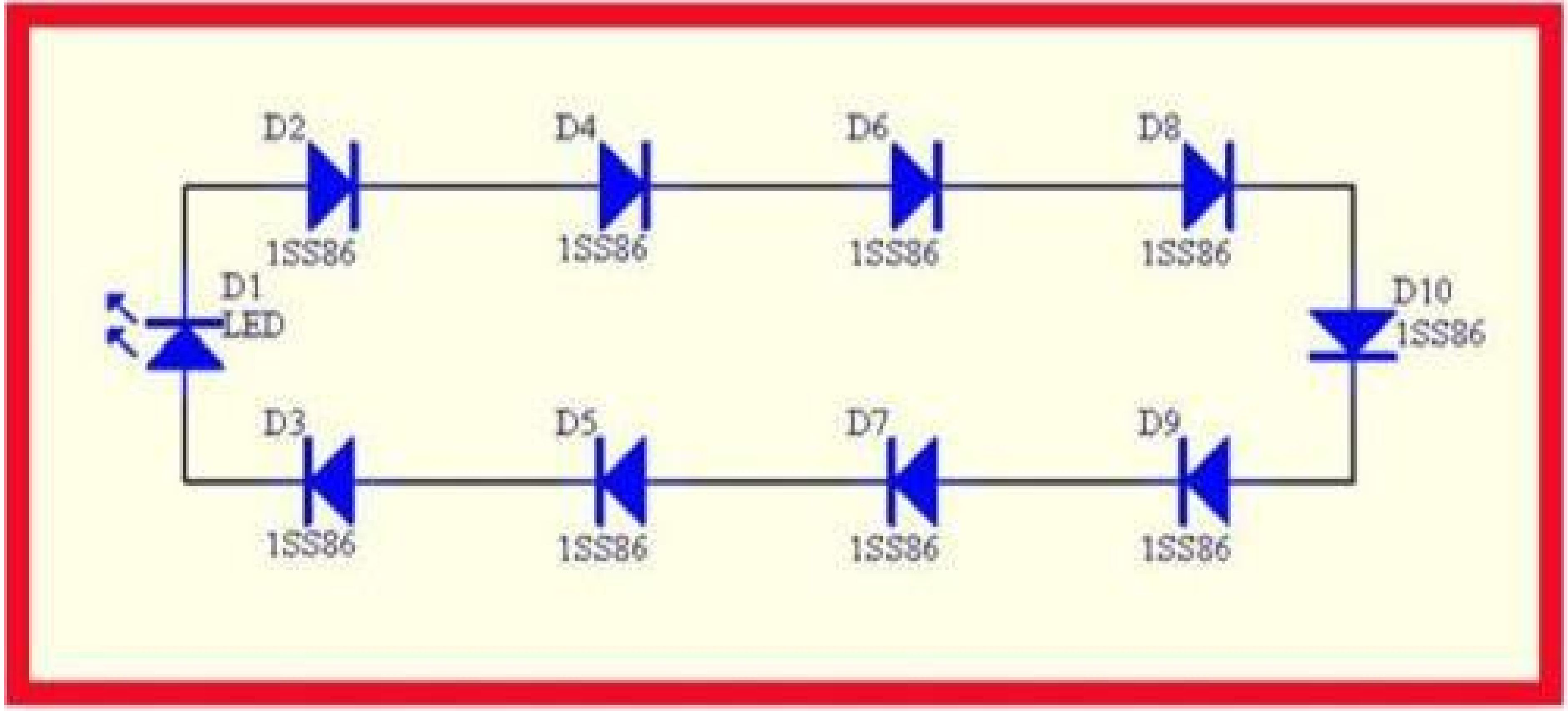
D2
1SS86
D4
1SS86
D6
1SS86
D8
1SS86
D1
LED
D10
1SS86
D3
1SS86
D5
1SS86
D7
1SS86
D9
1SS86

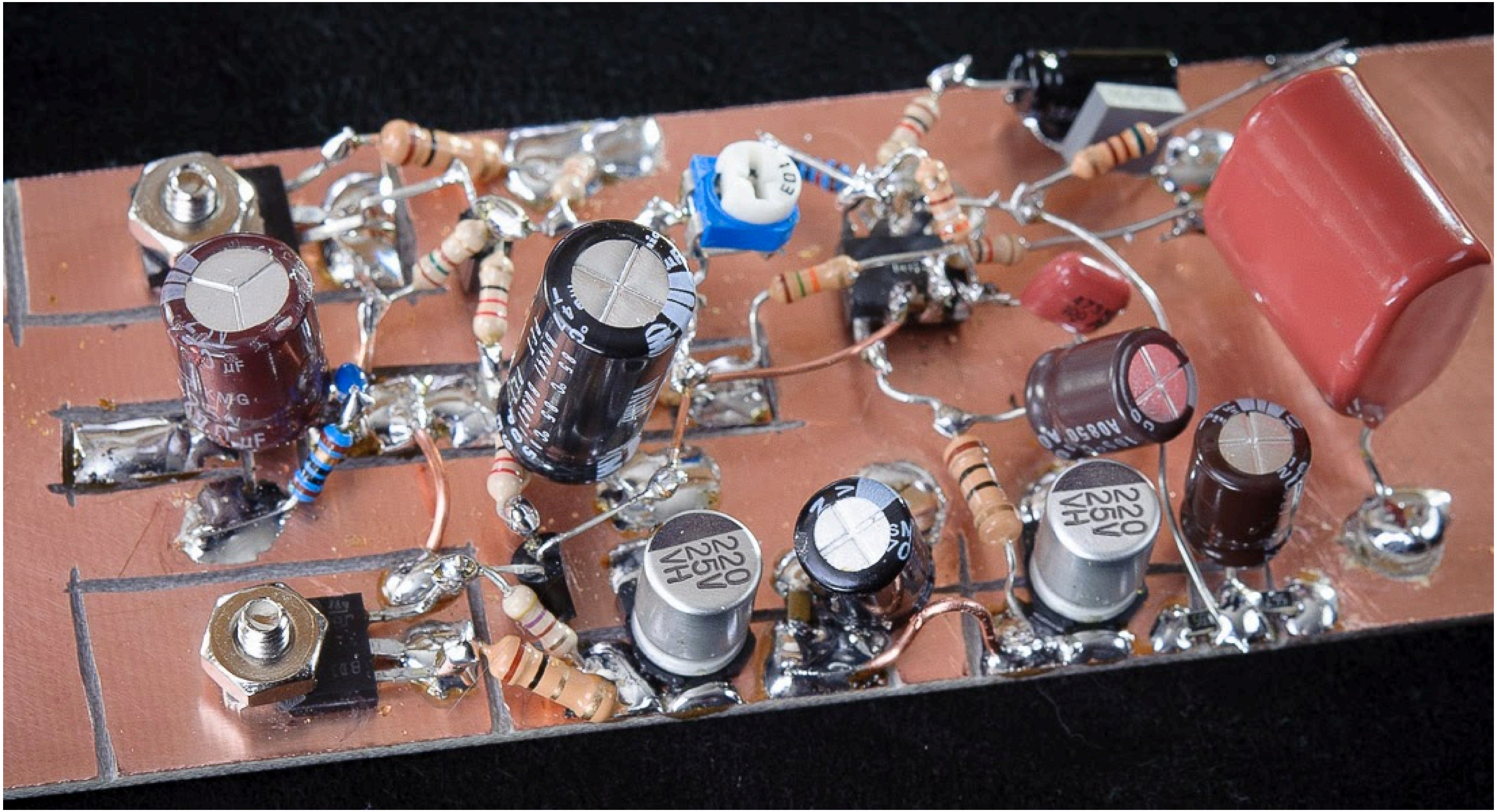
220
25V
VH
103

Shopping List:

- 4 - Germanium Diodes (1N34)
- 2 - 100 µF 50V electrolytic capacitors
- 2 - 0.2 µF 50V ceramic capacitors

Total cost = $2

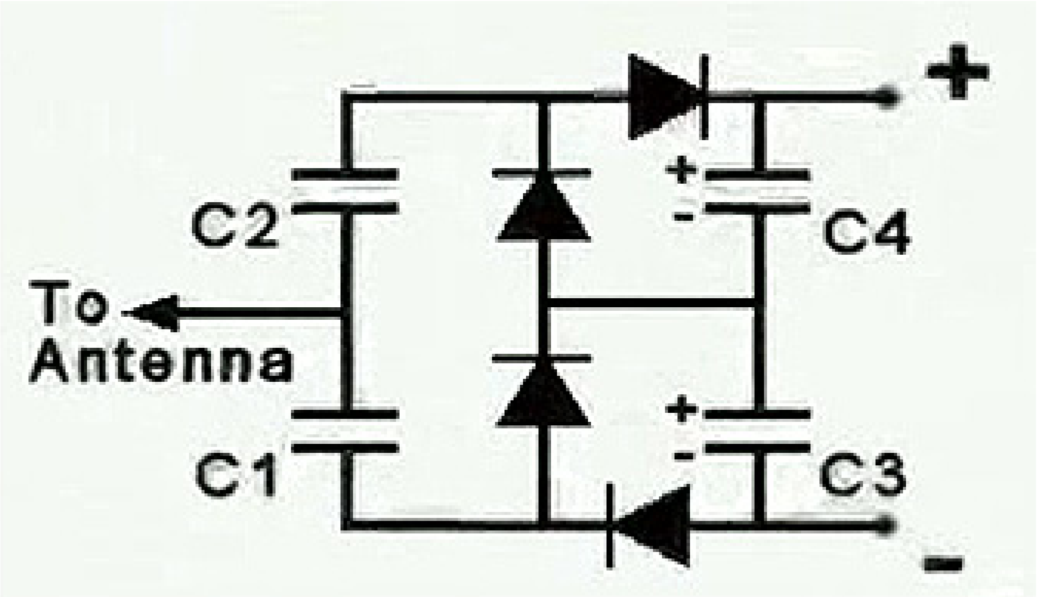
+
C2
C4
To Antenna
C1
C3
+
-
+
-
-

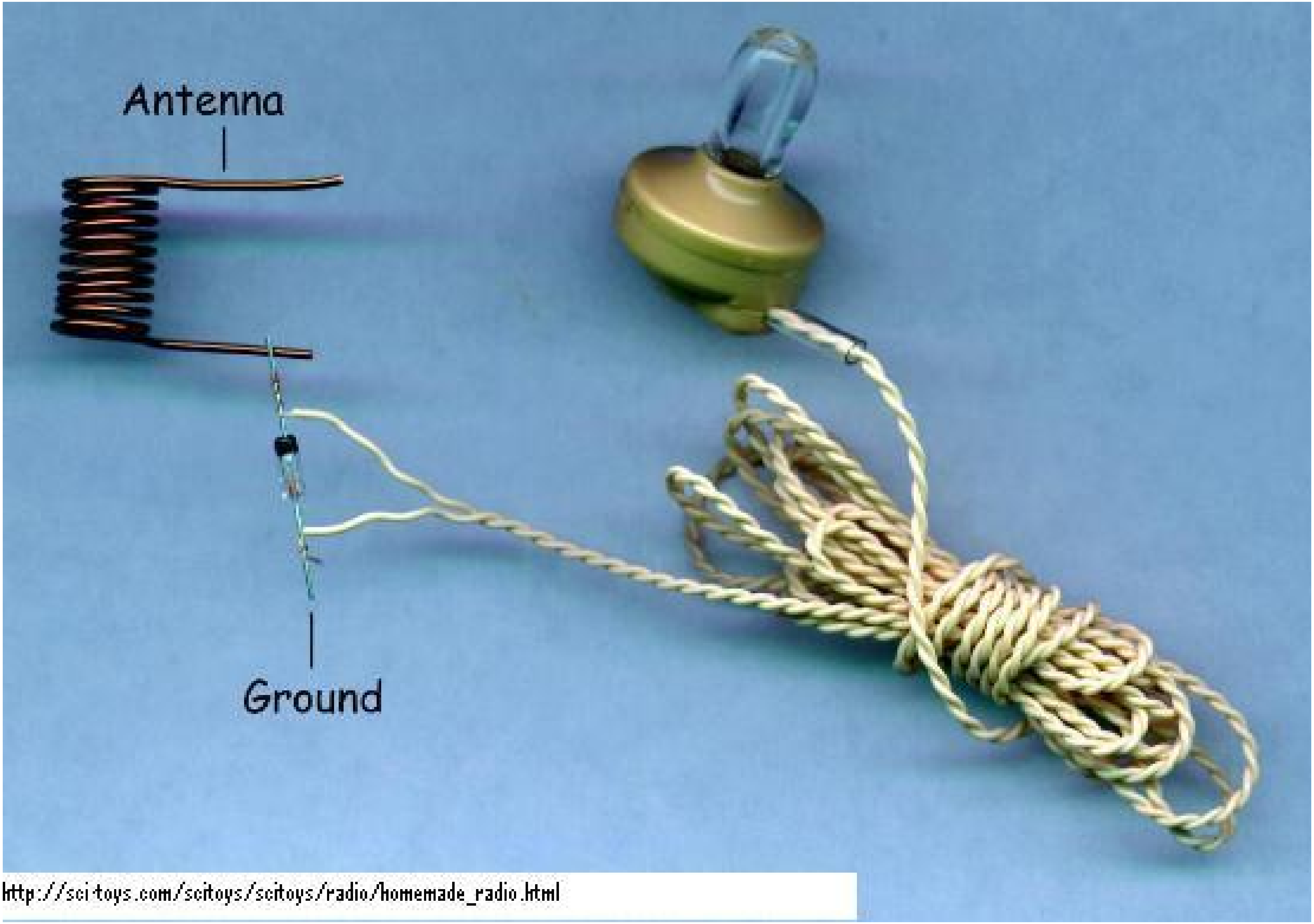
Antenna
Ground

http://sci-toys.com/scitoys/scitoys/radio/homemade_radio.html

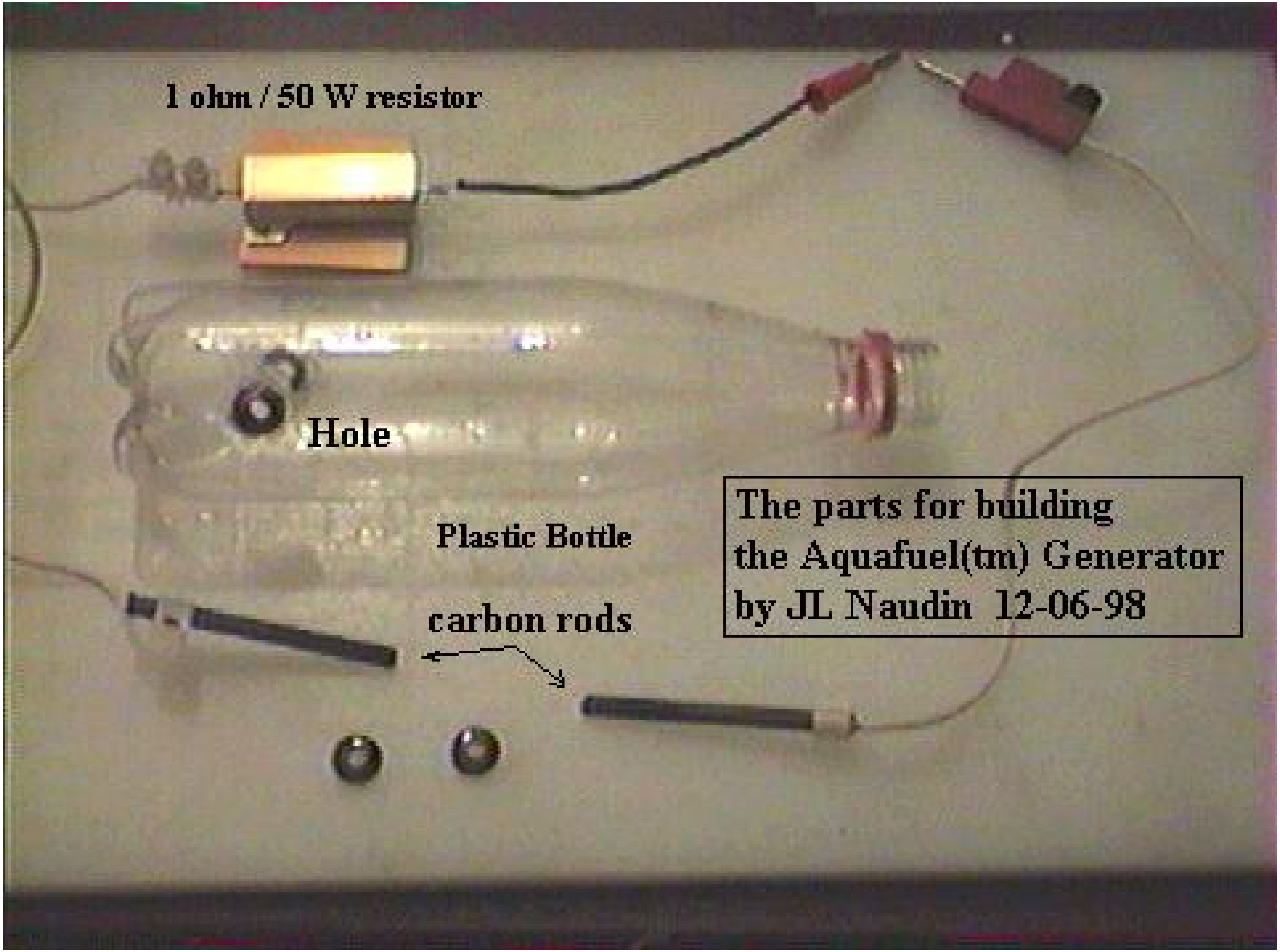
1 ohm / 50 W resistor
Hole
Plastic Bottle
carbon rods

The parts for building
the Aquafuel(tm) Generator
by JL Naudin 12-06-98

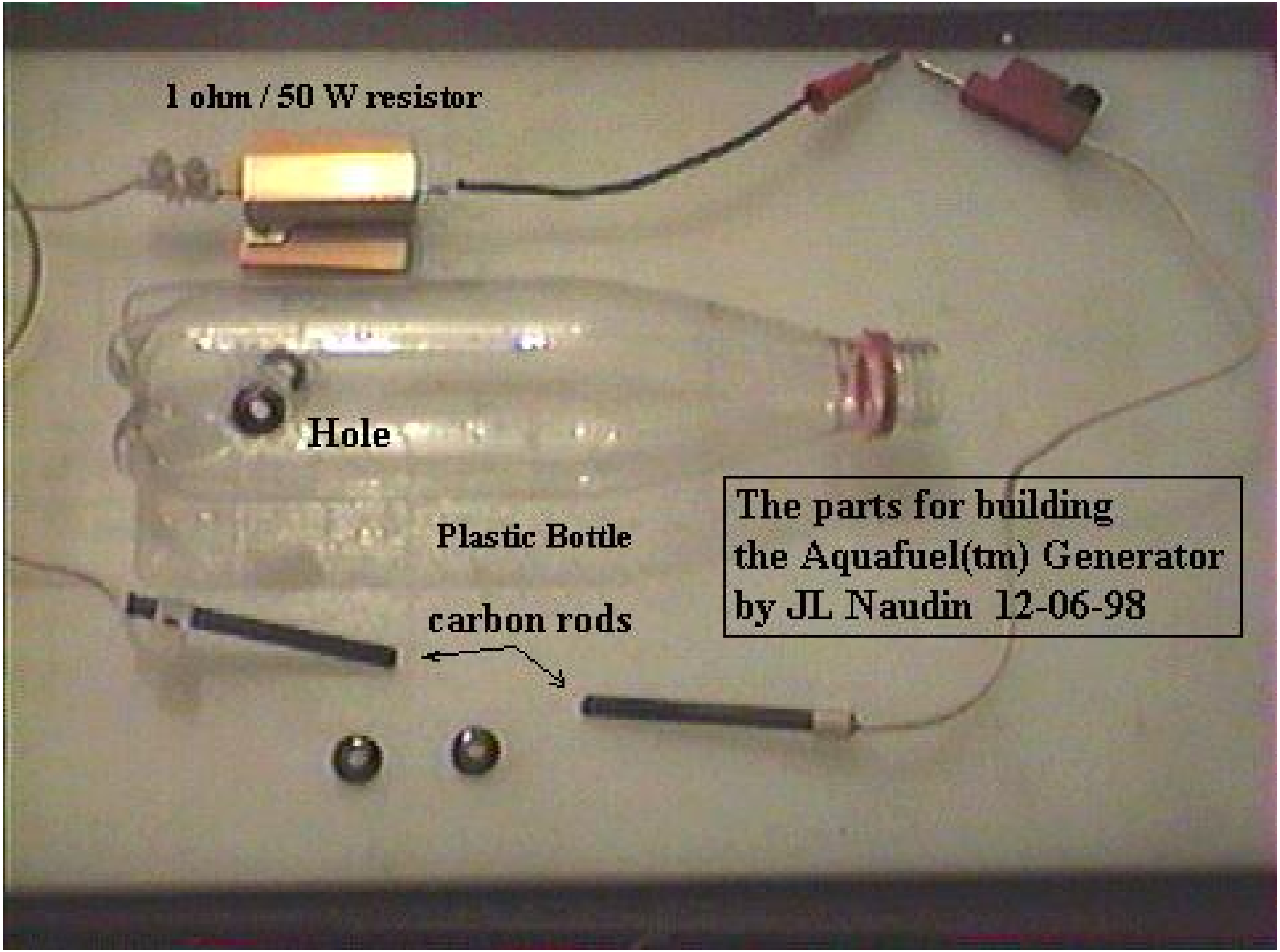
1 ohm / 50 W resistor
Hole
Plastic Bottle
carbon rods

The parts for building
the Aquafuel(tm) Generator
by JL Naudin 12-06-98

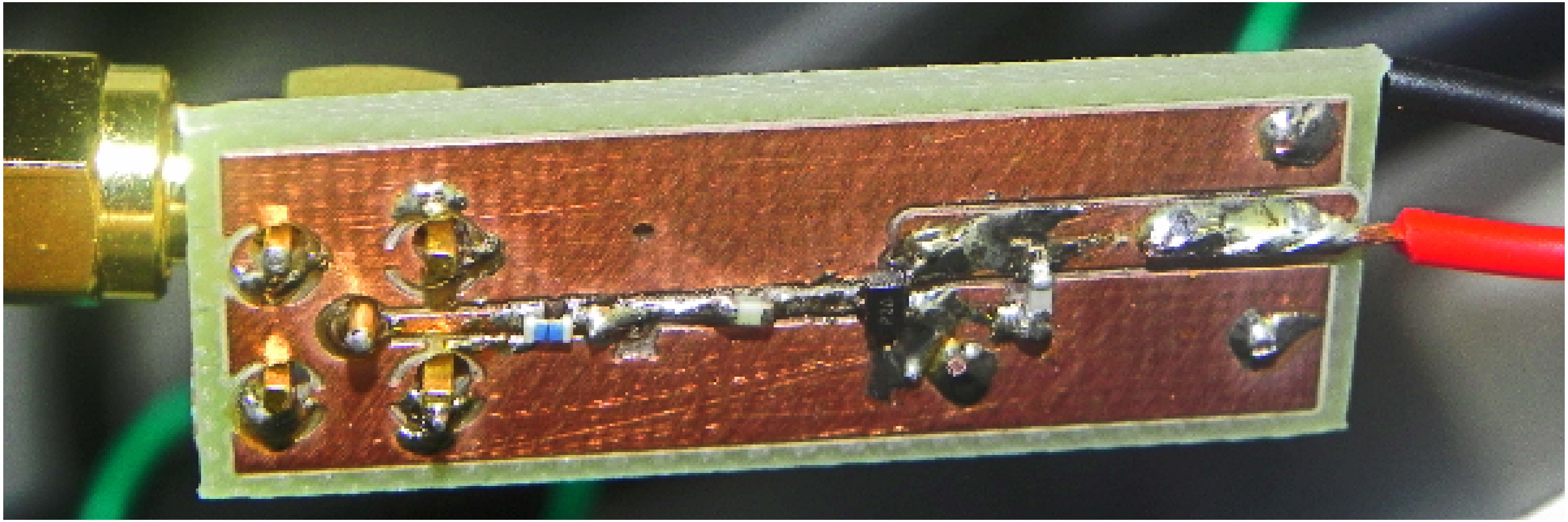

The Joule Thief

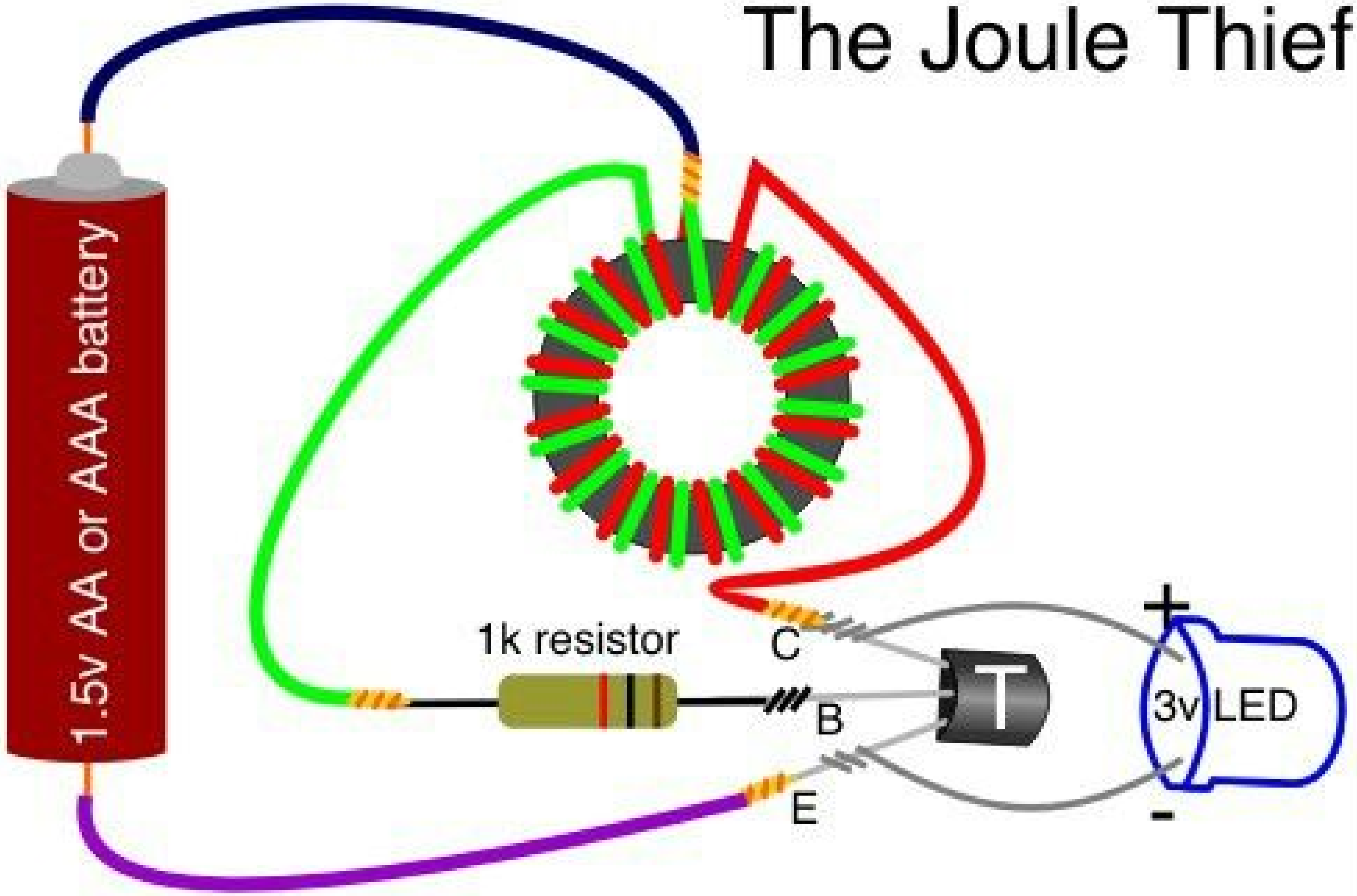
1.5v AA or AAA battery
1k resistor
C
B
E
T
+
3v LED
-

Transistor shown (T) is 2N3904 (can also use 2N2222).

Note LED pins - negative is shorter and has flat on bulb edge.

2 wires round ferrite ring are wound together, then start of one joined to end of other.

Try adding more and more LEDs...

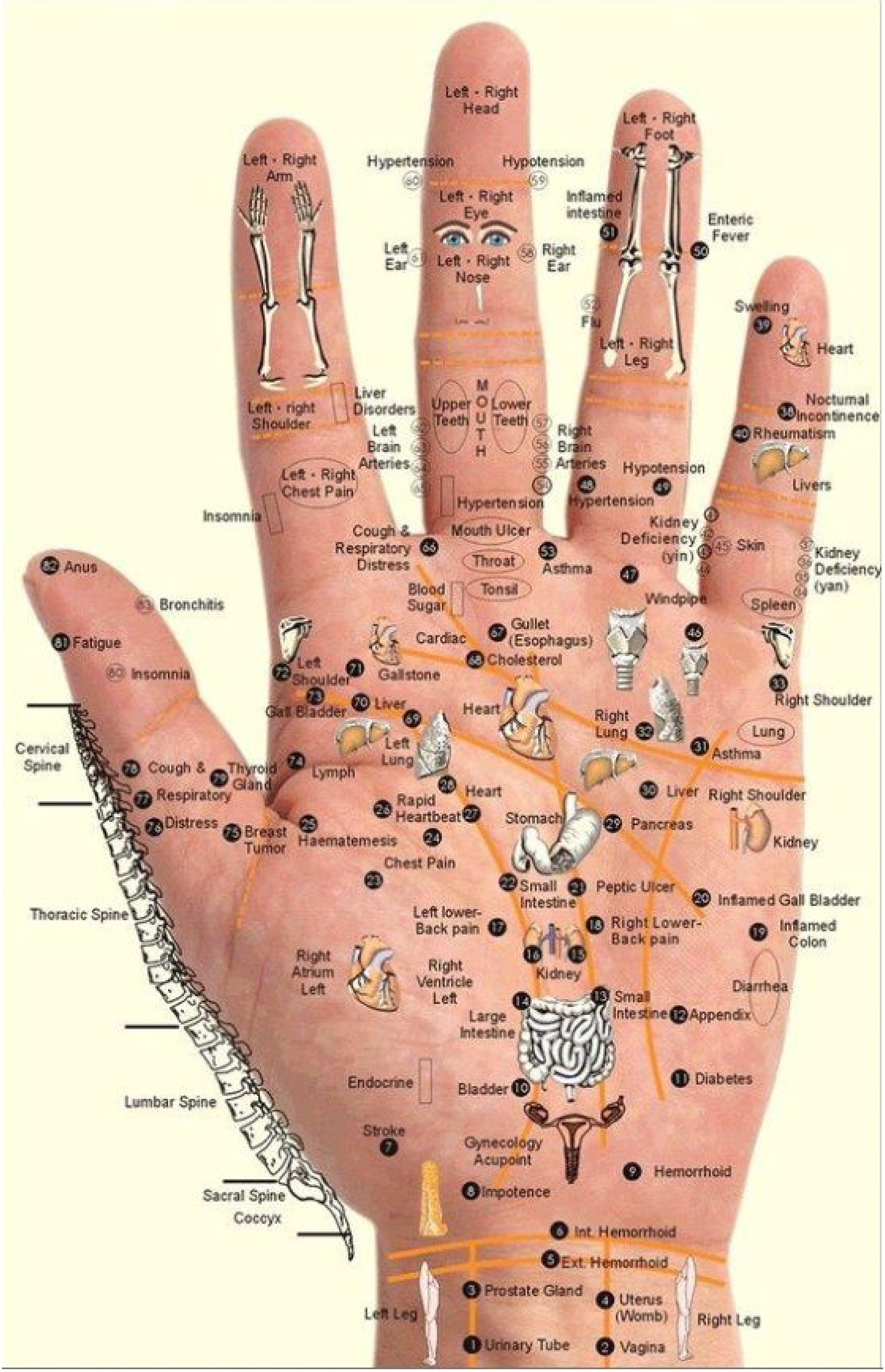

Left - Right
Head
Left - Right
Arm
Hypertension
60
Hypotension
59
Left - Right
Eye
Left
Ear
Left - Right
Nose
58
Right
Ear
Left - Right
Foot
Inflamed
intestine
51
Enteric
Fever
50
Flu
Left - Right
Leg
Swelling
39
Heart
Left - right
Shoulder
Liver
Disorders
Upper
Teeth
M
O
U
T
H
Lower
Teeth
Left
Brain
Arteries
57
Right
56
Brain
55
Arteries
54
Noctumal
38
Incontinence
40
Rheumatism
Livers
Left - Right
Chest Pain
Insomnia
Hypertension
48
Hypertension
Hypotension
49
Kidney
Deficiency
(yin)
45
Skin
Kidney
Deficiency
(yan)
Cough &
Respiratory
Distress
65
Mouth Ulcer
Throat
Tonsil
53
Asthma
47
Blood
Sugar
Windpipe
Spleen
82
Anus
83
Bronchitis
81
Fatigue
80
Insomnia
67
Gullet
(Esophagus)
Cardiac
68
Cholesterol
46
72
Left
Shoulder
71
Gallstone
33
Right Shoulder
73
Gall Bladder
70
Liver
69
Heart
Right
Lung
32
Lung
Left
Lung
31
Asthma
Cervical
Spine
78
Cough &
79
Thyroid
Gland
77
Respiratory
76
Distress
75
Breast
Tumor
74
Lymph
28
Heart
26
Rapid
Heartbeat
27
25
Haematemesis
24
Stomach
29
Pancreas
30
Liver
Right Shoulder
Kidney
Chest Pain
23
22
Small
Intestine
21
Peptic Ulcer
20
Inflamed Gall Bladder
Thoracic Spine
Left lower-
Back pain
17
18
Right Lower-
Back pain
19
Inflamed
Colon
Right
Atrium
Left
Right
Ventricle
Left
16
15
Kidney
Diarrhea
14
Large
Intestine
13
Small
Intestine
12
Appendix
Endocrine
Bladder
10
11
Diabetes
Lumbar Spine
Stroke
7
Gynecology
Acupoint
9
Hemorrhoid
8
Impotence
Sacral Spine
Coccyx
6
Int. Hemorrhoid
5
Ext. Hemorrhoid
3
Prostate Gland
4
Uterus
(Womb)
Left Leg
Right Leg
1
Urinary Tube
2
Vagina

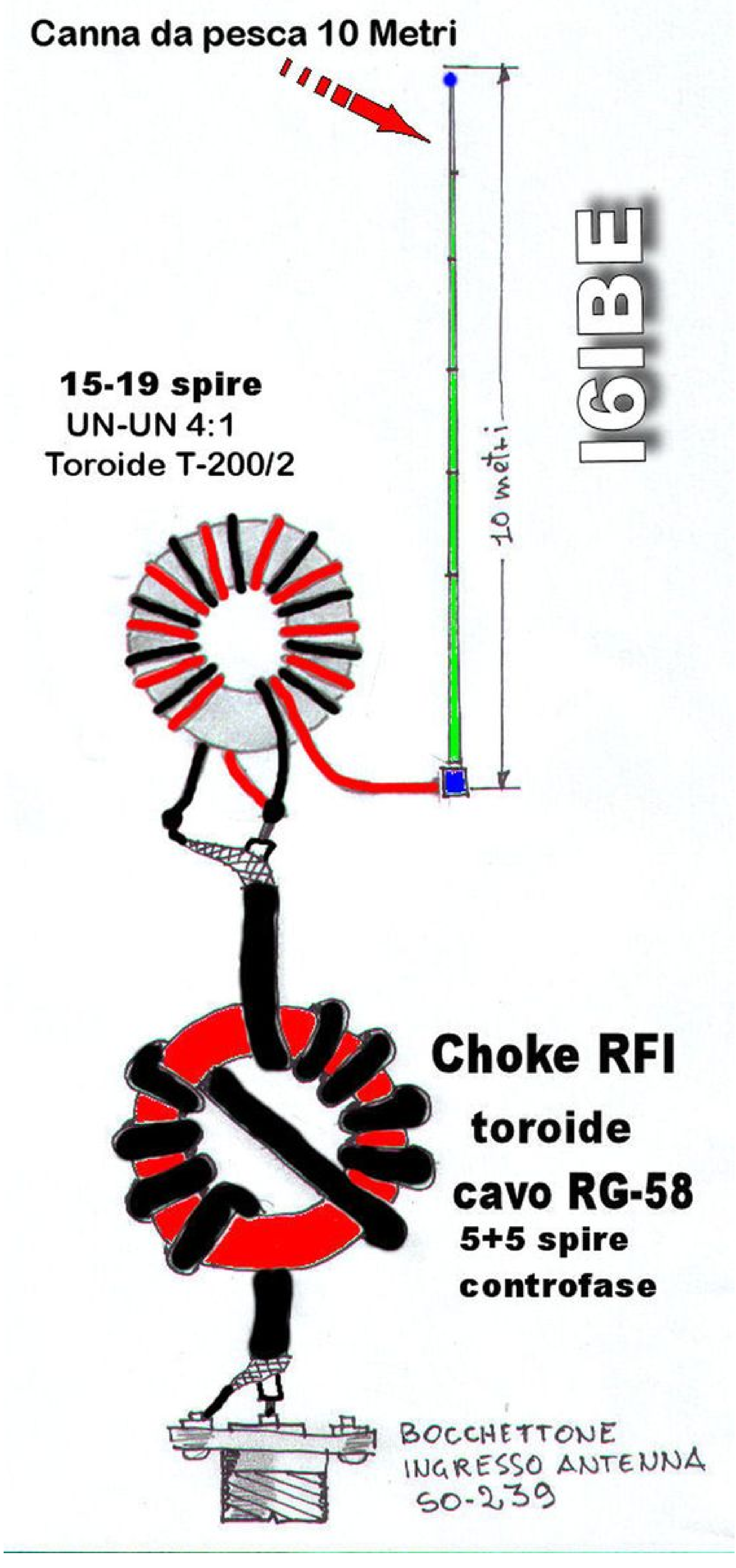
Canna da pesca 10 Metri
I6IBE
10 metri
15-19 spire
UN-UN 4:1
Toroide T-200/2
Choke RFI
toroide
cavo RG-58
5+5 spire
controfase
BOCCHETTONE
INGRESSO ANTENNA
SO-239

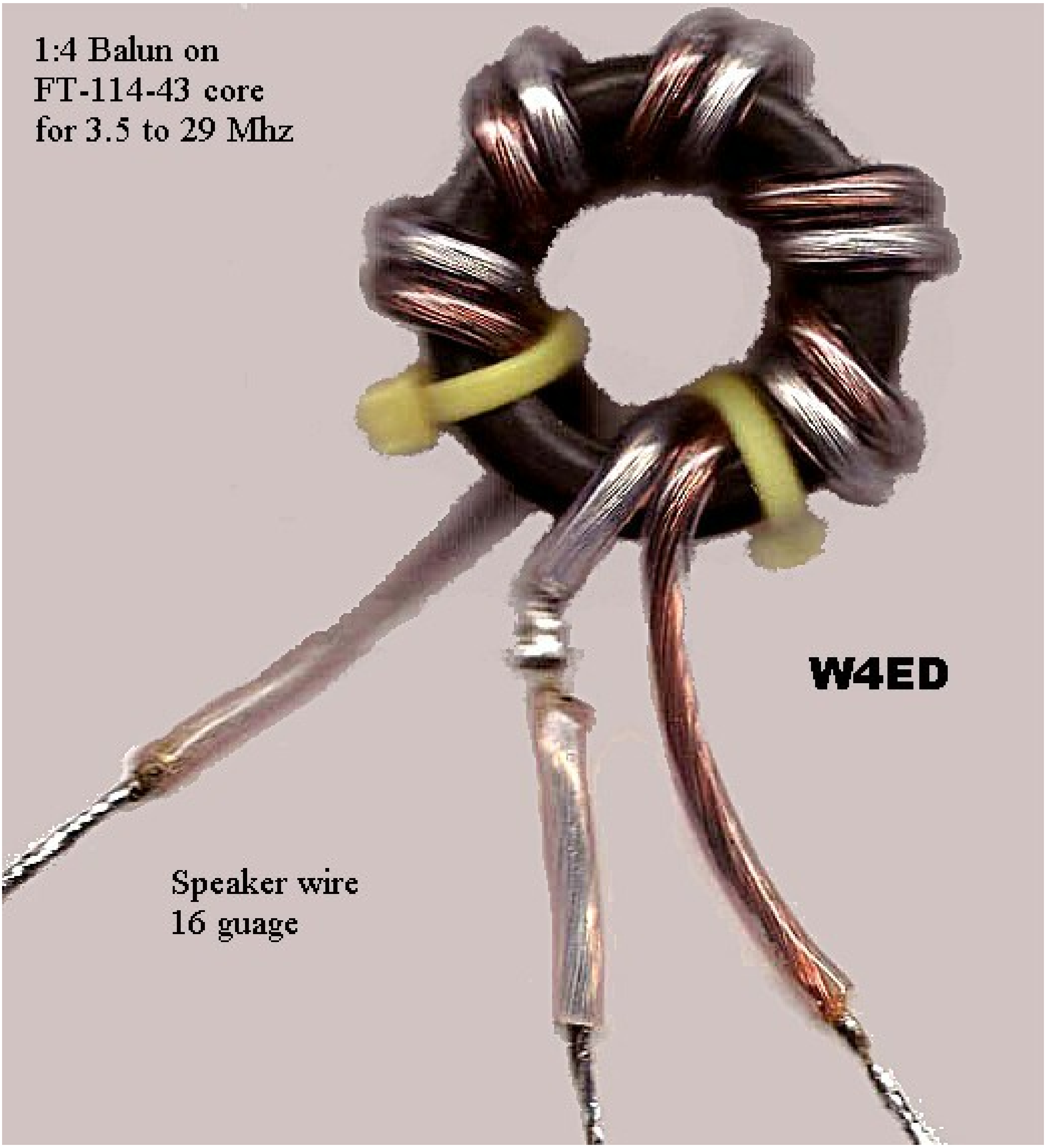

1:4 Balun on
FT-114-43 core
for 3.5 to 29 Mhz

W4ED

Speaker wire
16 guage

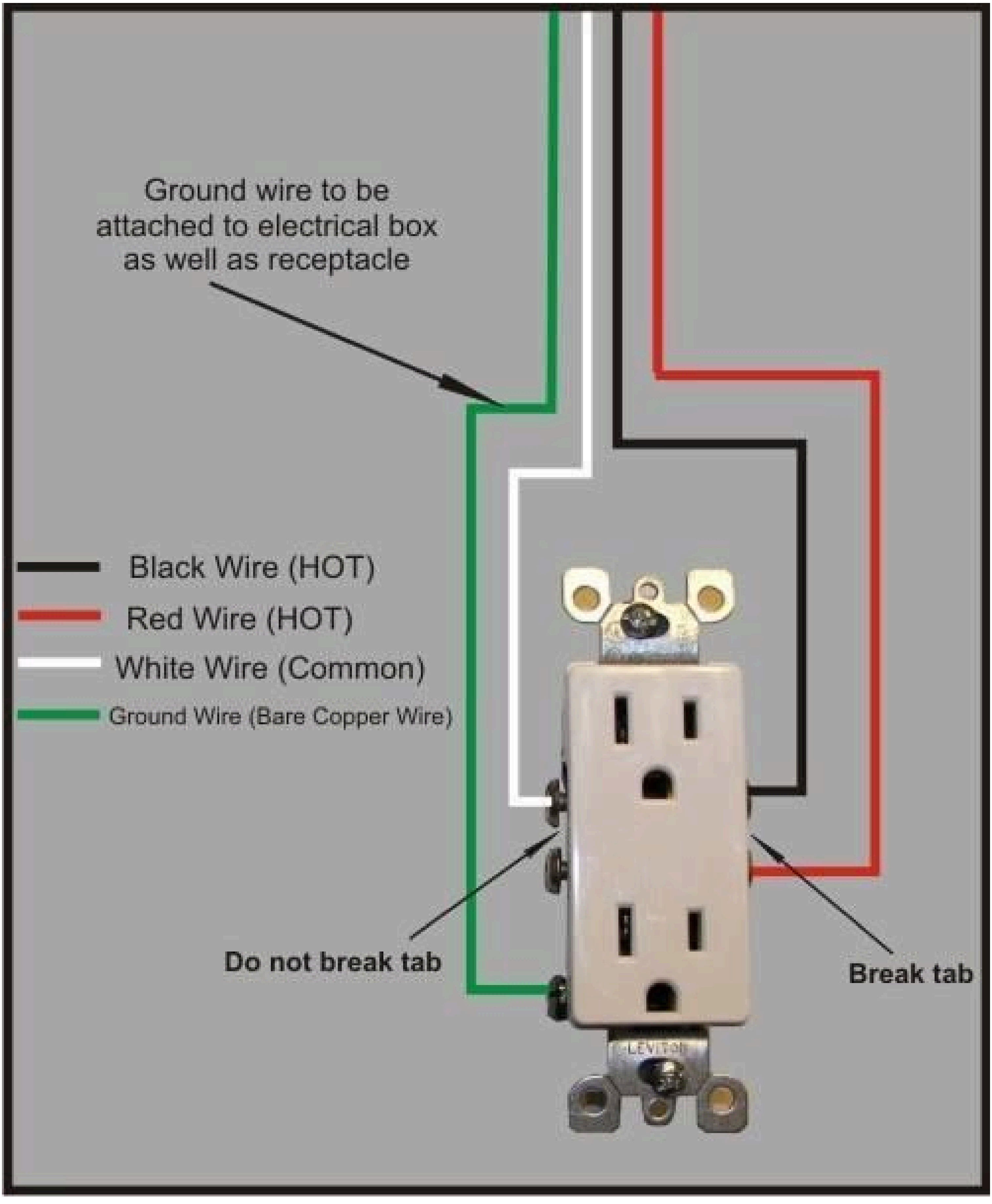
Ground wire to be attached to electrical box as well as receptacle
Black Wire (HOT)
Red Wire (HOT)
White Wire (Common)
Ground Wire (Bare Copper Wire)
Do not break tab
Break tab
LEVITON

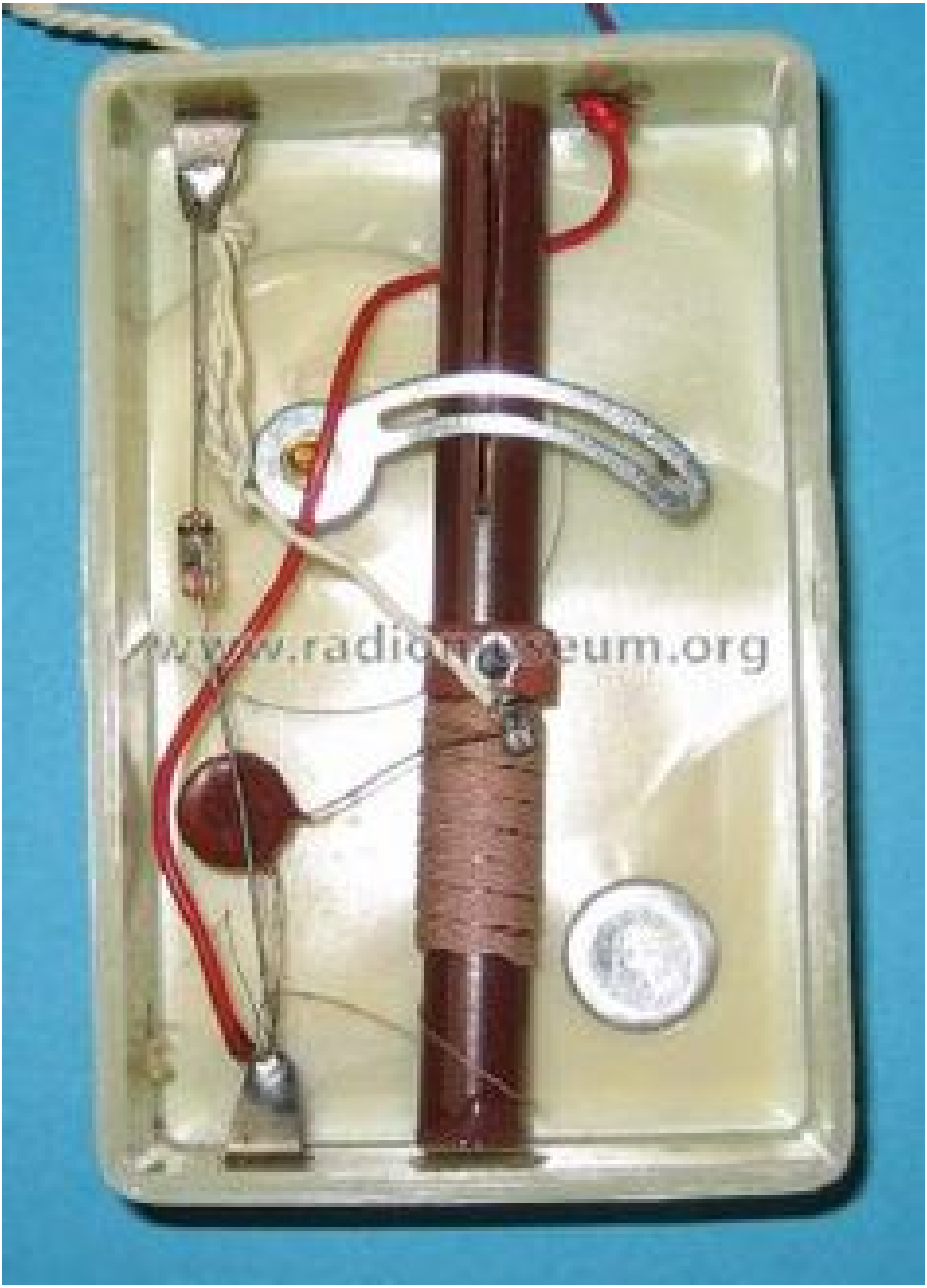

www.radiomuseum.org

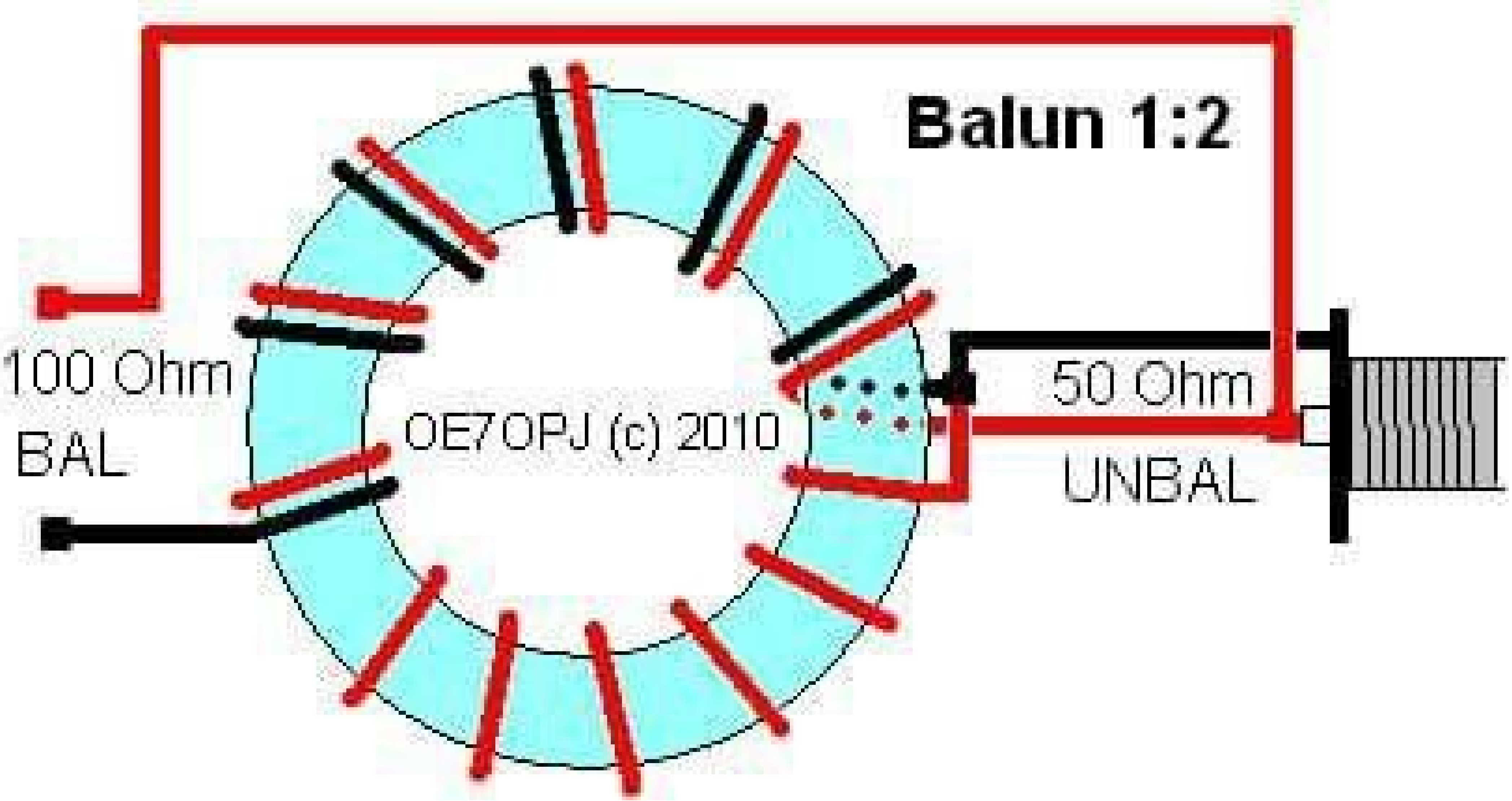
Balun 1:2
100 Ohm
BAL
OE7OPJ (c) 2010
50 Ohm
UNBAL

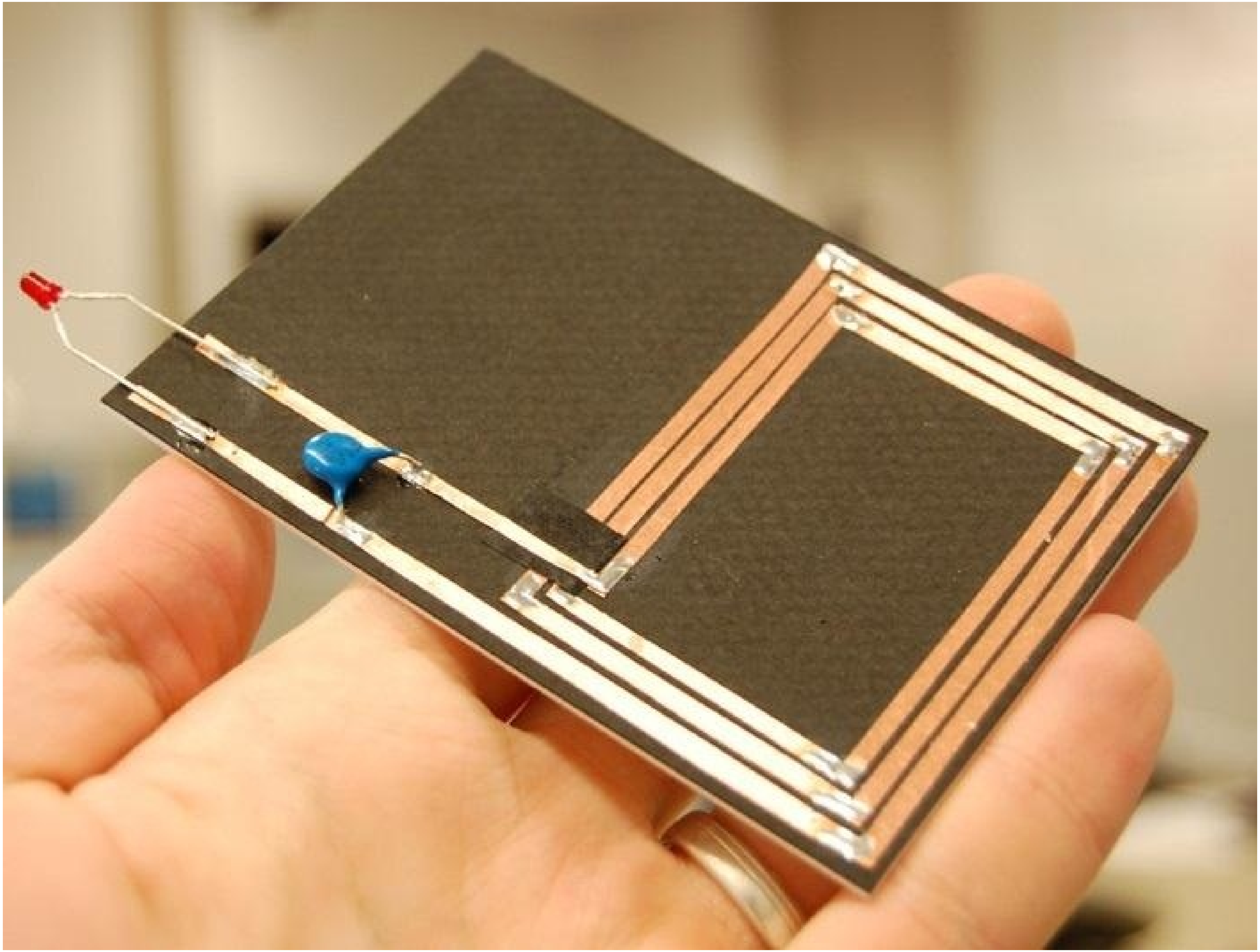

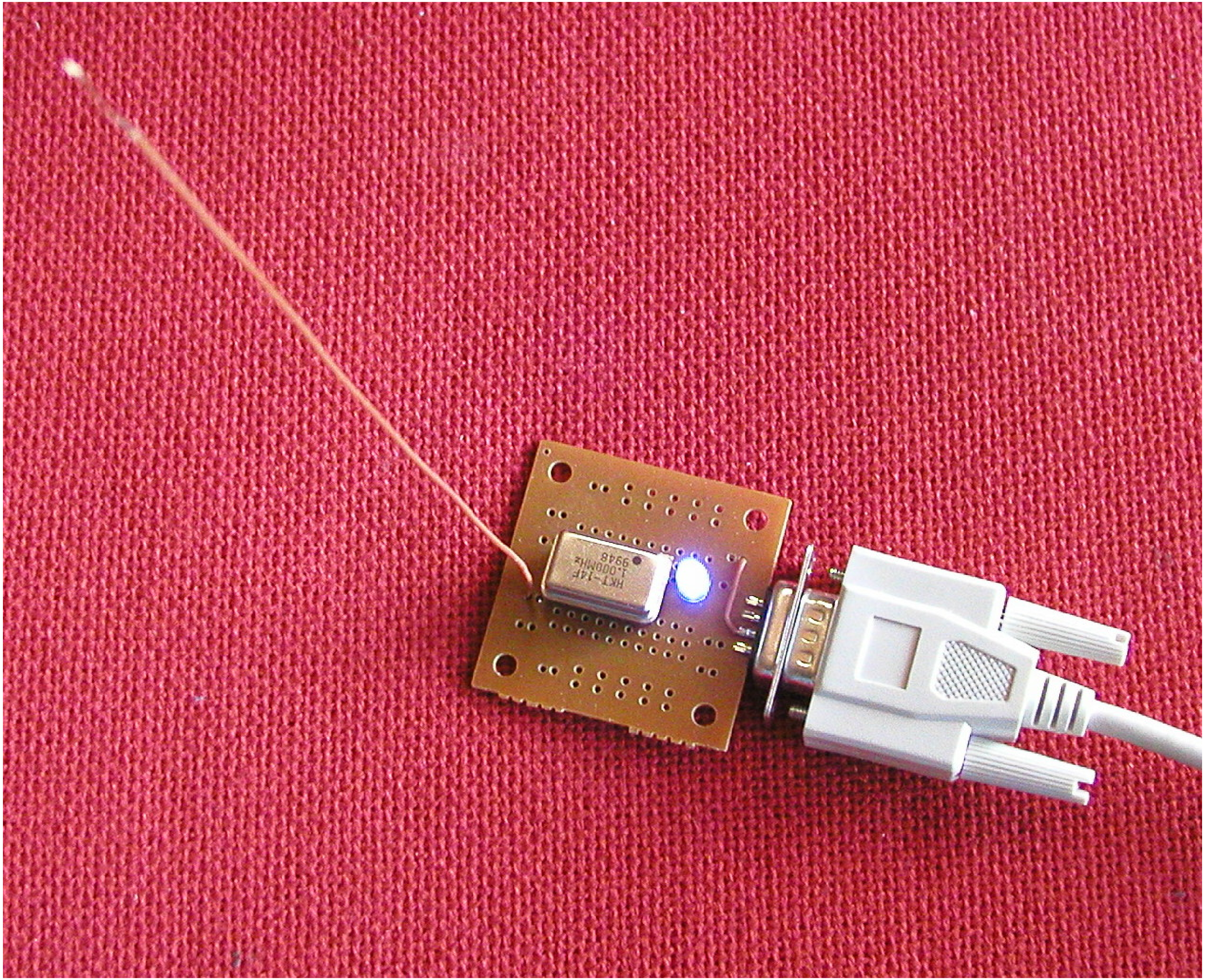
HKT-14F
1.000MHz
9948

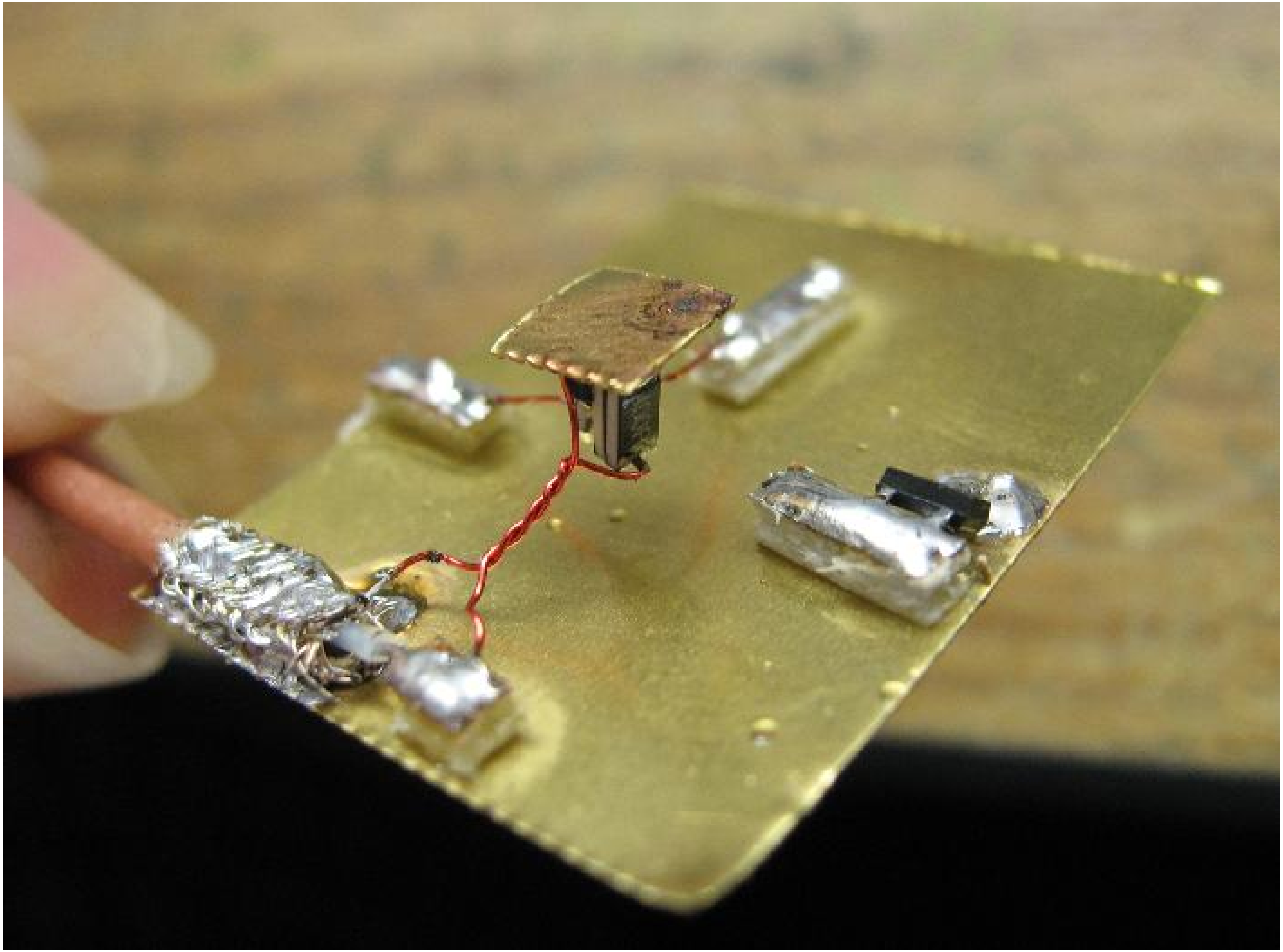

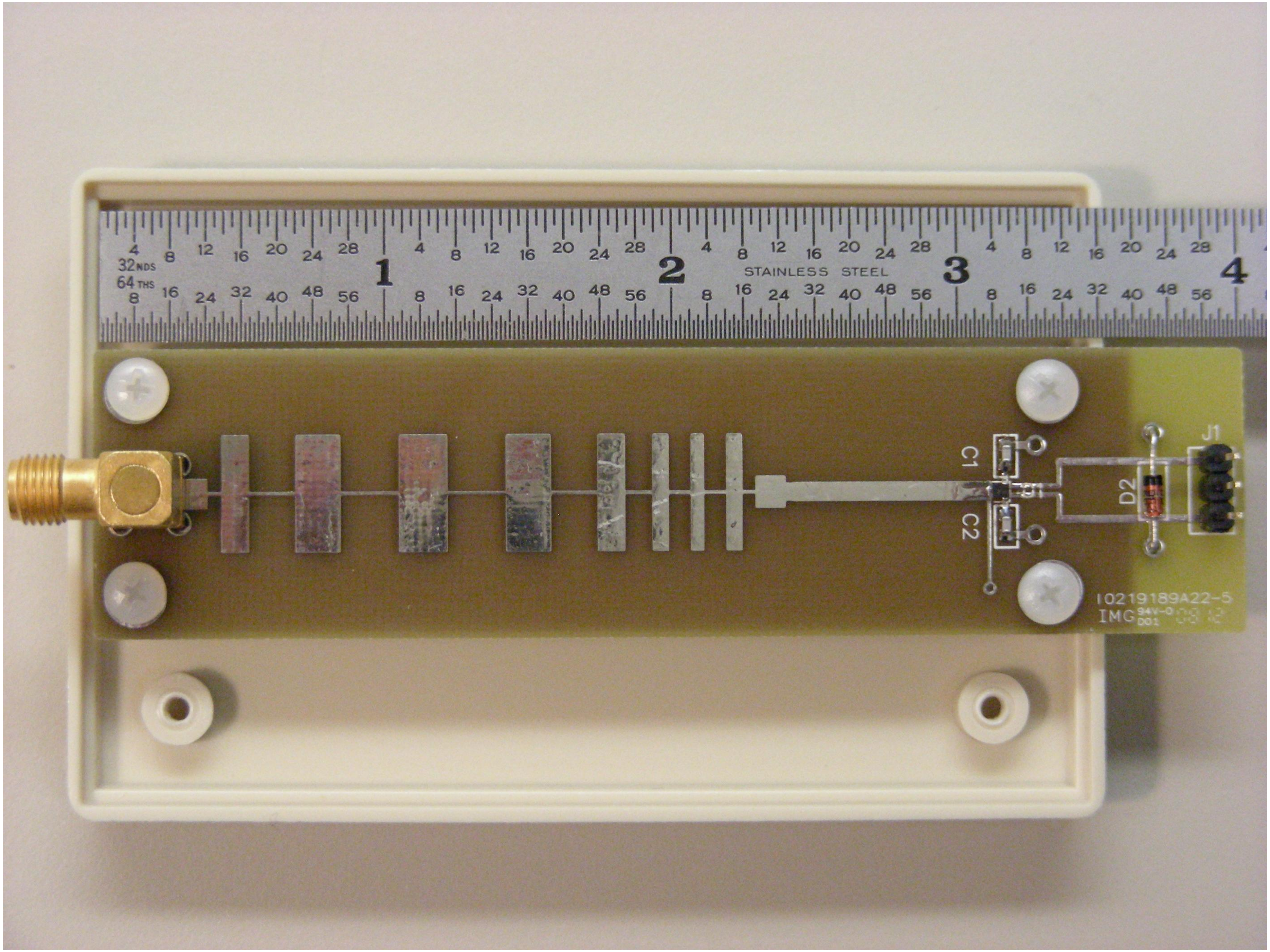
32NDS
64THS
1
2
STAINLESS STEEL
3
4
C1
C2
D2
J1
10219189A22-5
IMG 94V-0 D01 08 12

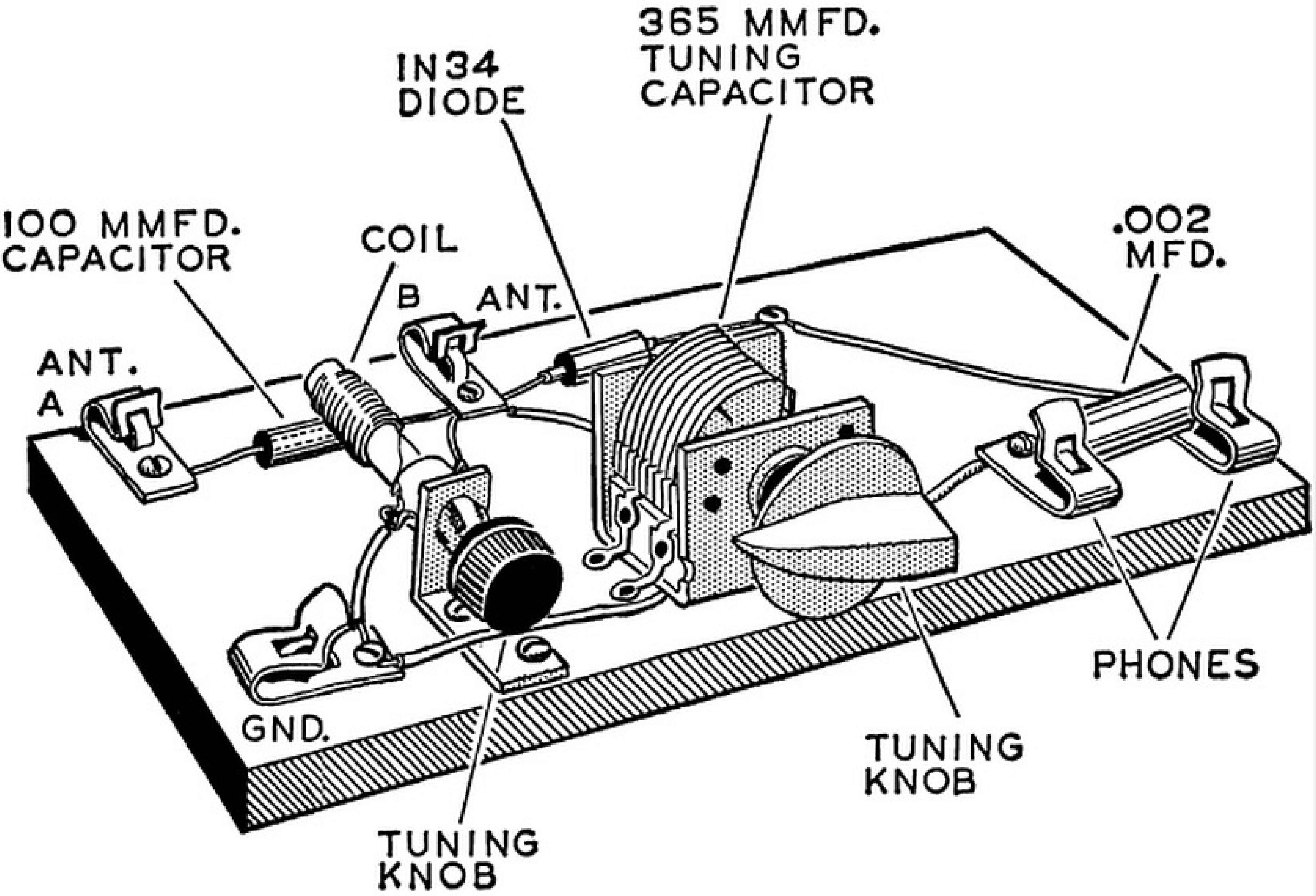

365 MMFD. TUNING CAPACITOR
IN34 DIODE
100 MMFD. CAPACITOR
COIL
B
ANT.
.002 MFD.
ANT. A
GND.
TUNING KNOB
TUNING KNOB
PHONES

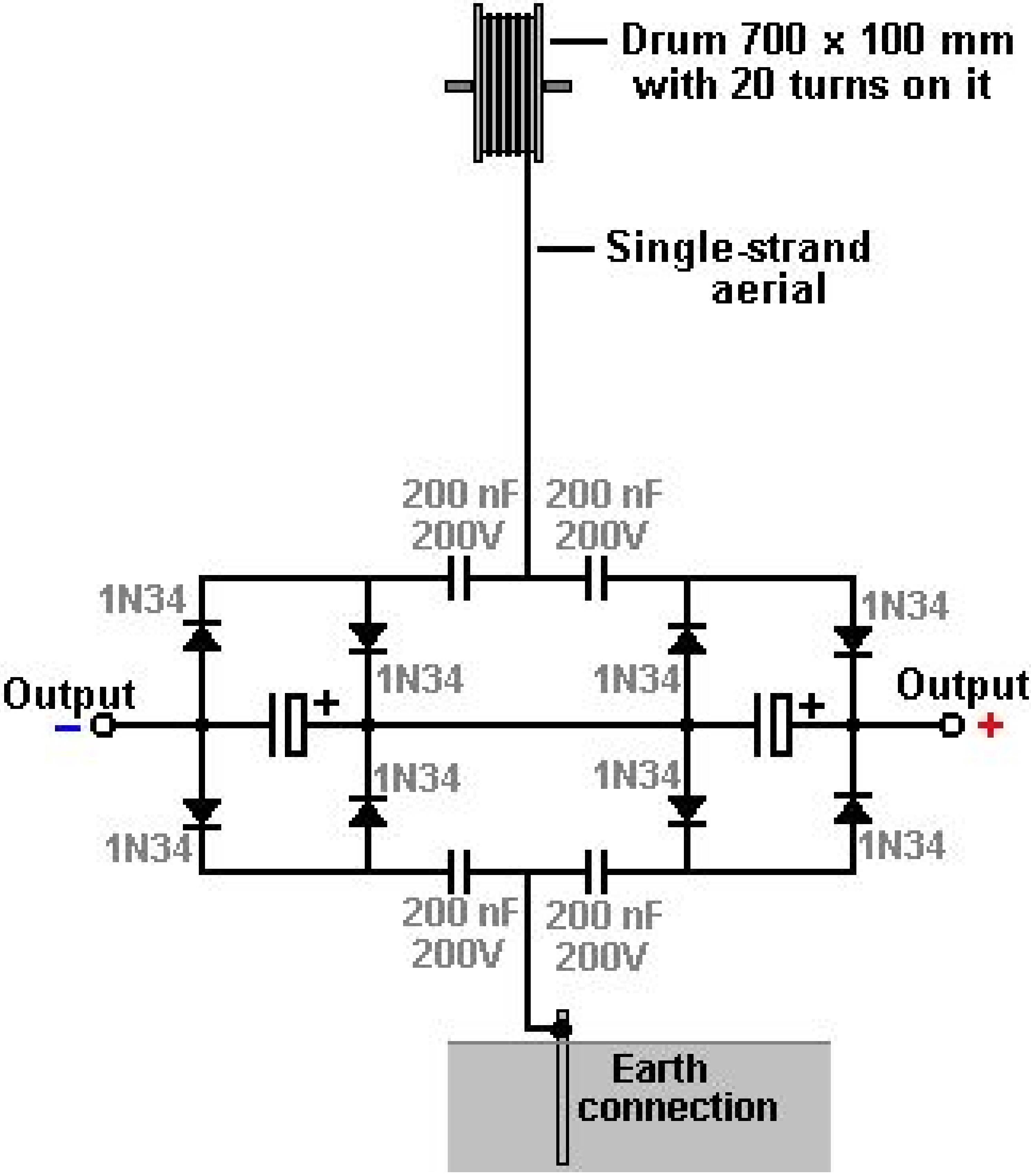
Drum 700 x 100 mm
with 20 turns on it
Single-strand
aerial
200 nF
200V
200 nF
200V
1N34
1N34
1N34
1N34
1N34
1N34
1N34
1N34
Output
–
+
+
Output
+
200 nF
200V
200 nF
200V
Earth
connection

4:1 Current Balun

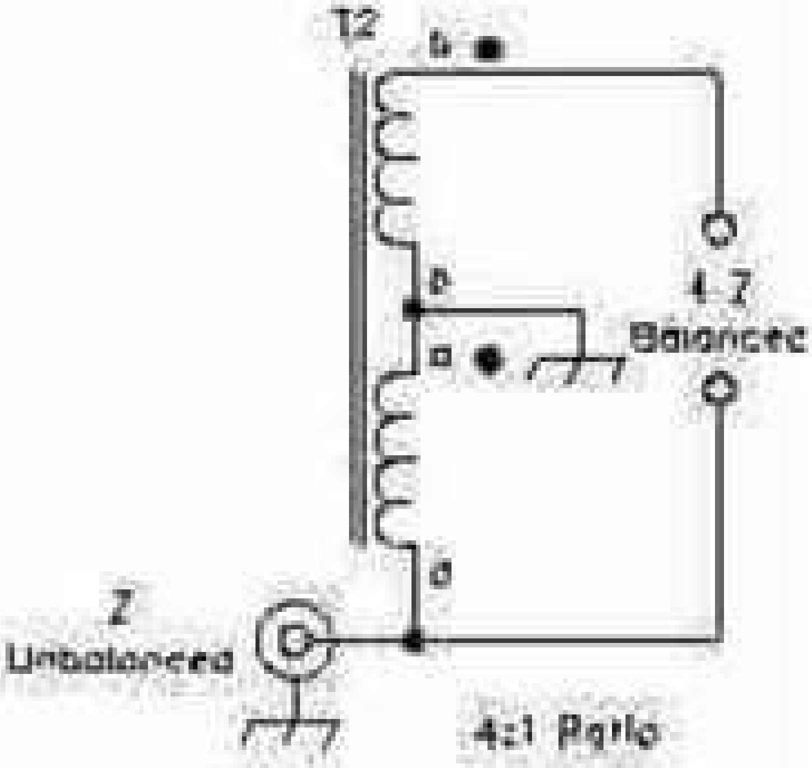
T2
b
b
a
a
4 Z
Balanced
Z
Unbalanced
4:1 Ratio

Enamle 1.5mm
Each 9-10 turns
On a Radio Broadcast (MW) Ferrite

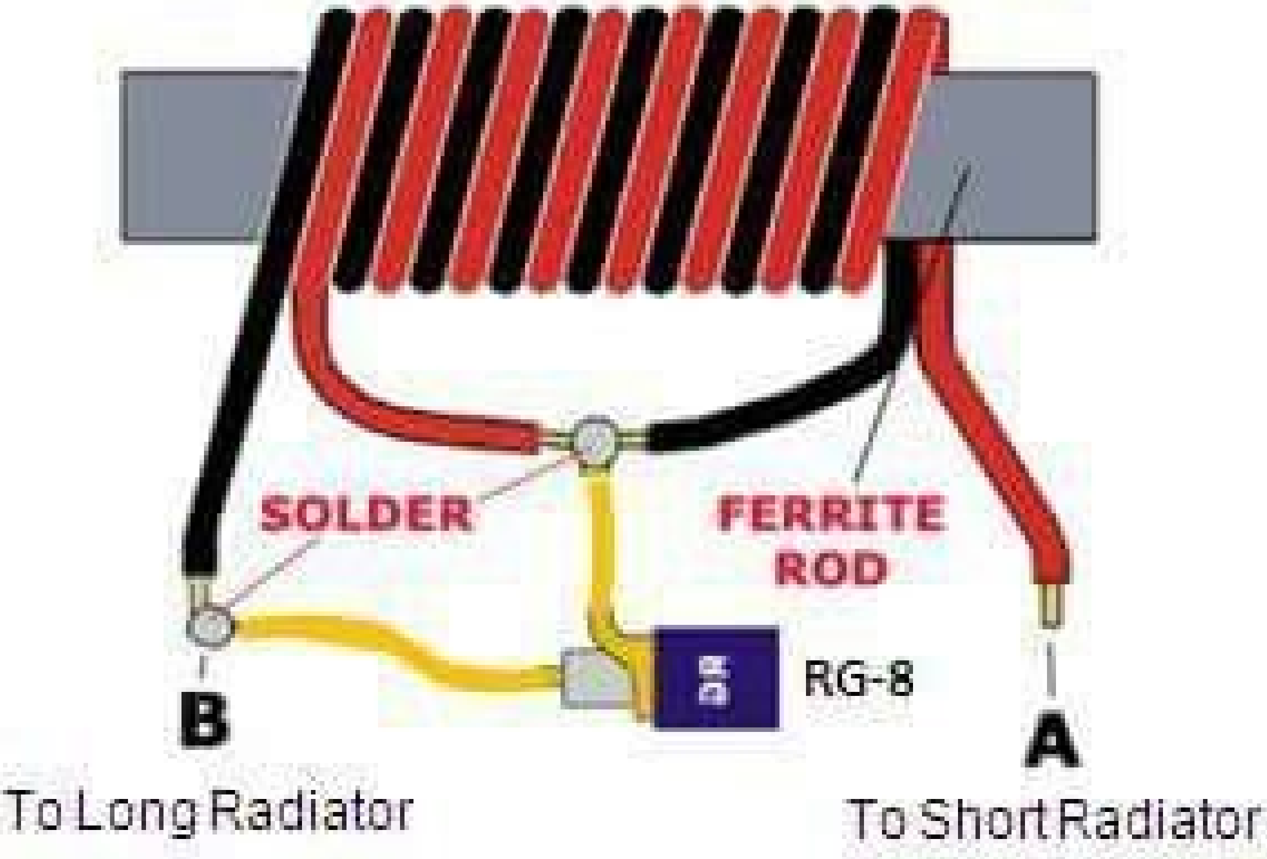
SOLDER
FERRITE
ROD
RG-8
B
To Long Radiator
A
To Short Radiator

4:1 Current Balun

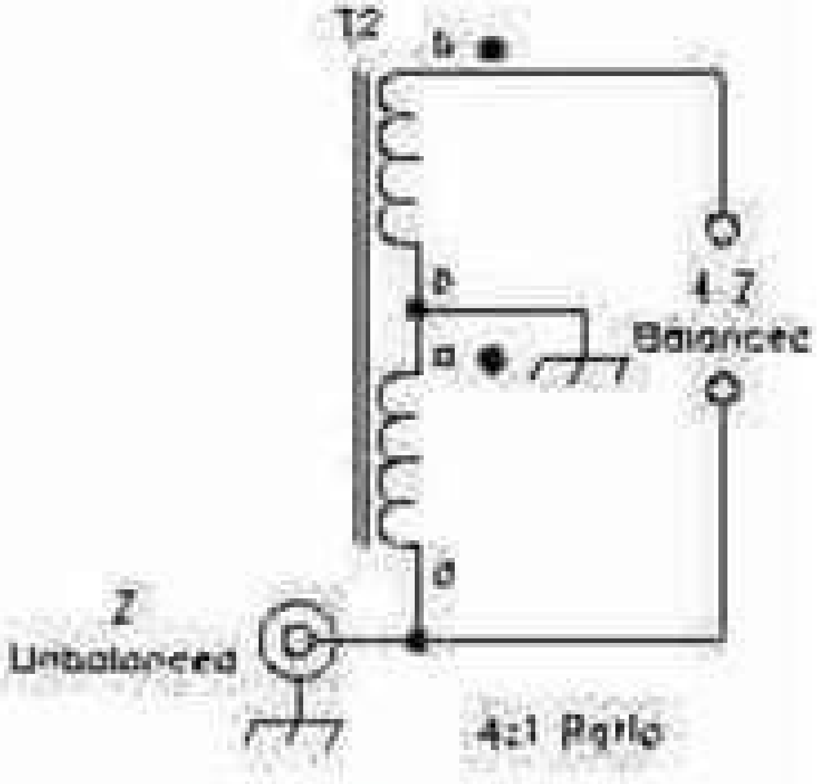
T2
b
b
a
a
4 Z
Balanced
Z
Unbalanced
4:1 Ratio

Enamle 1.5mm
Each 9-10 turns
On a Radio Broadcast (MW) Ferrite

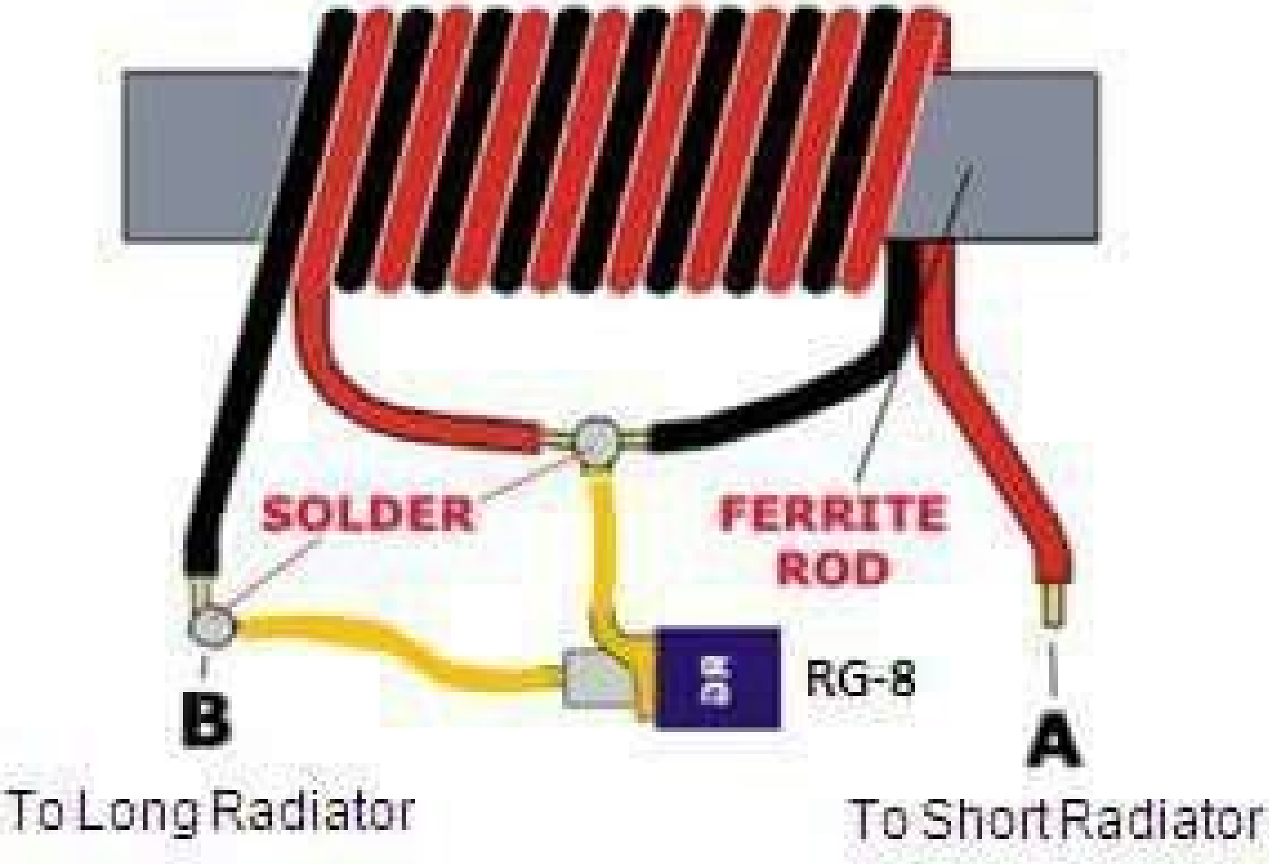
SOLDER
FERRITE
ROD
RG-8
B
To Long Radiator
A
To Short Radiator

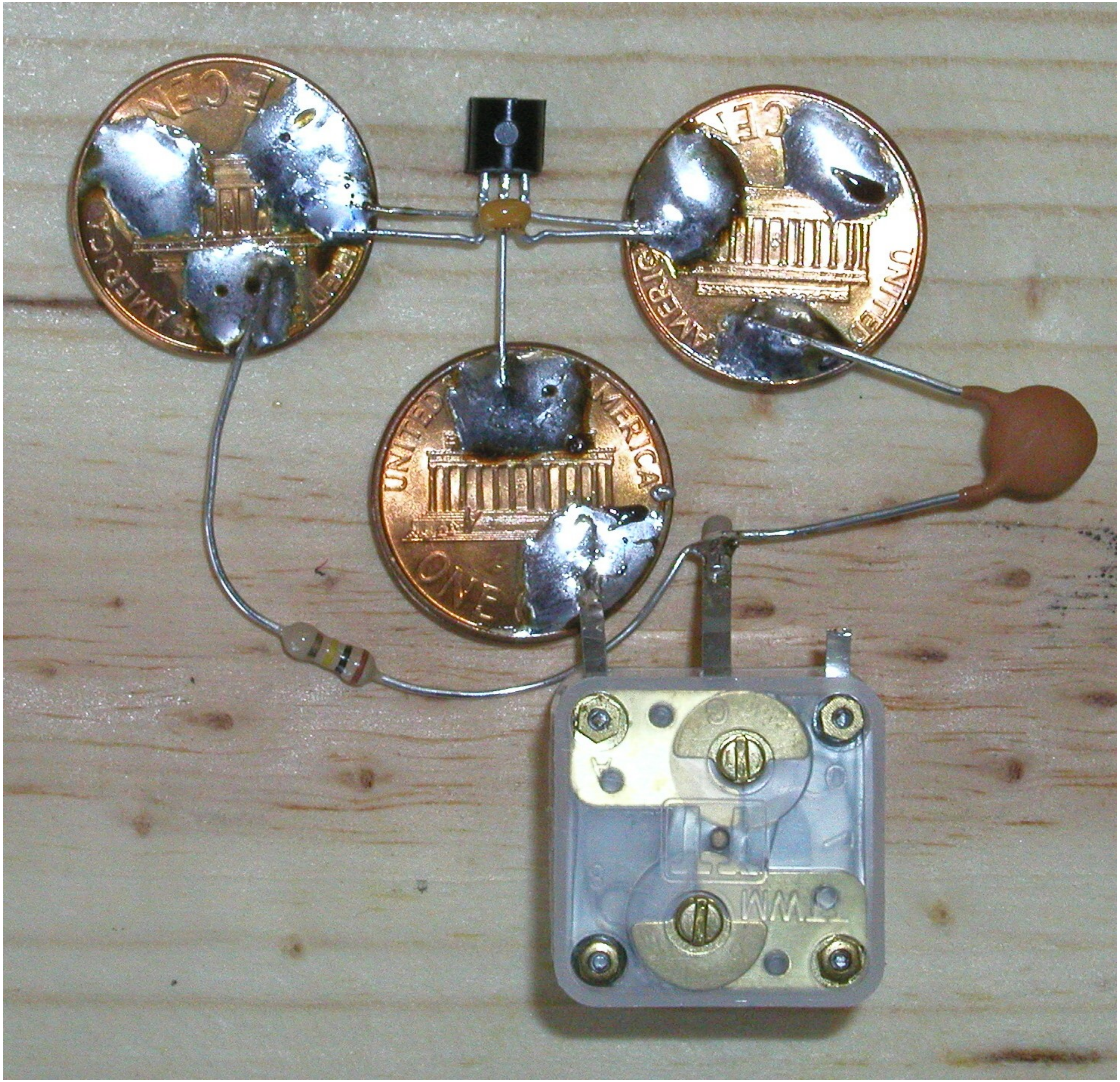

Multi-Band HF Vertical Dipole

On a Spiderbeam 12m HD Telescoping Fiberglass Pole

for: 80/40/30/20/17/15/12/10m

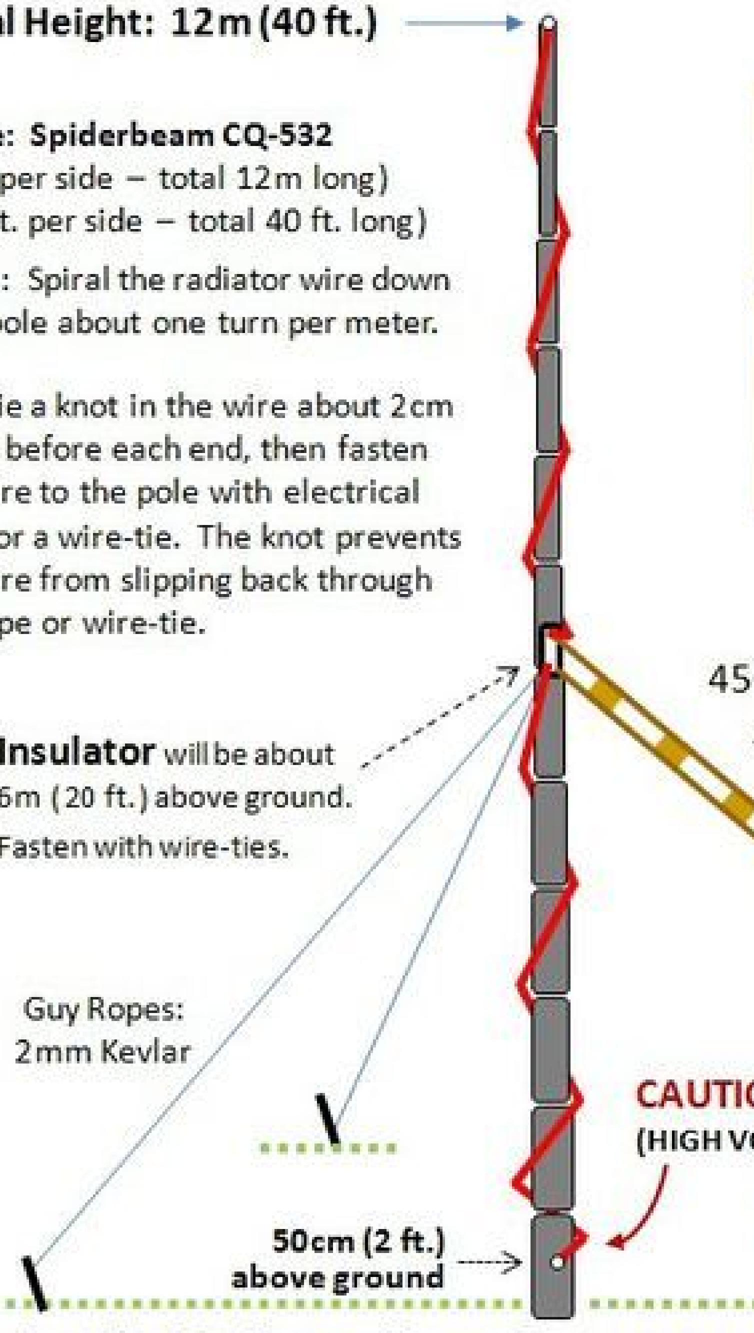

Total Height: 12m (40 ft.)

Wire: Spiderbeam CQ-532
(6m per side – total 12m long)
(20 ft. per side – total 40 ft. long)

Note: Spiral the radiator wire down the pole about one turn per meter.

Tip: tie a knot in the wire about 2cm (1 in.) before each end, then fasten the wire to the pole with electrical tape or a wire-tie. The knot prevents the wire from slipping back through the tape or wire-tie.

Insulator will be about 6m (20 ft.) above ground. Fasten with wire-ties.

Guy Ropes: 2mm Kevlar

50cm (2 ft.) above ground

Guy stakes 5 to 7m away from pole. (16 to 23 ft.) spaced equal distance around the pole. Or, use tree or fence to fasten ends.

CAUTION HV!
(HIGH VOLTAGE)

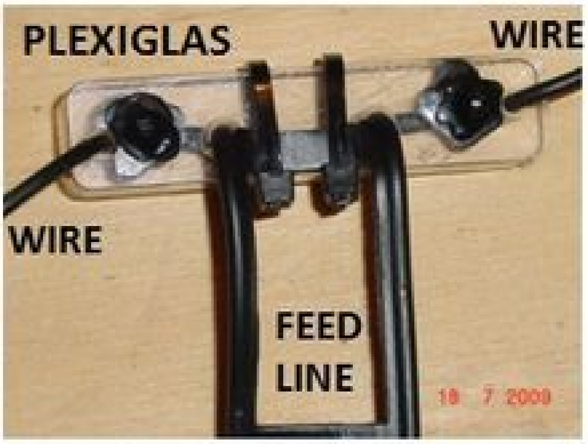

Simple lightweight insulator

450Ω Openwire Feedline

Although the length is not critical, there are some lengths which can cause trouble. See text.

Good Antenna Matchbox

MFJ-974B*

Guy Rope

*Alternatives described in text.

Spiraling the wire down the pole distributes its weight evenly around the pole and prevents it from flopping in the wind. Since the lower half of the pole is thicker (wider) than the upper half, the wire will end about 50cm above the ground.

IT IS IMPORTANT TO KEEP THE WIRE THIS DISTANCE FROM THE GROUND!

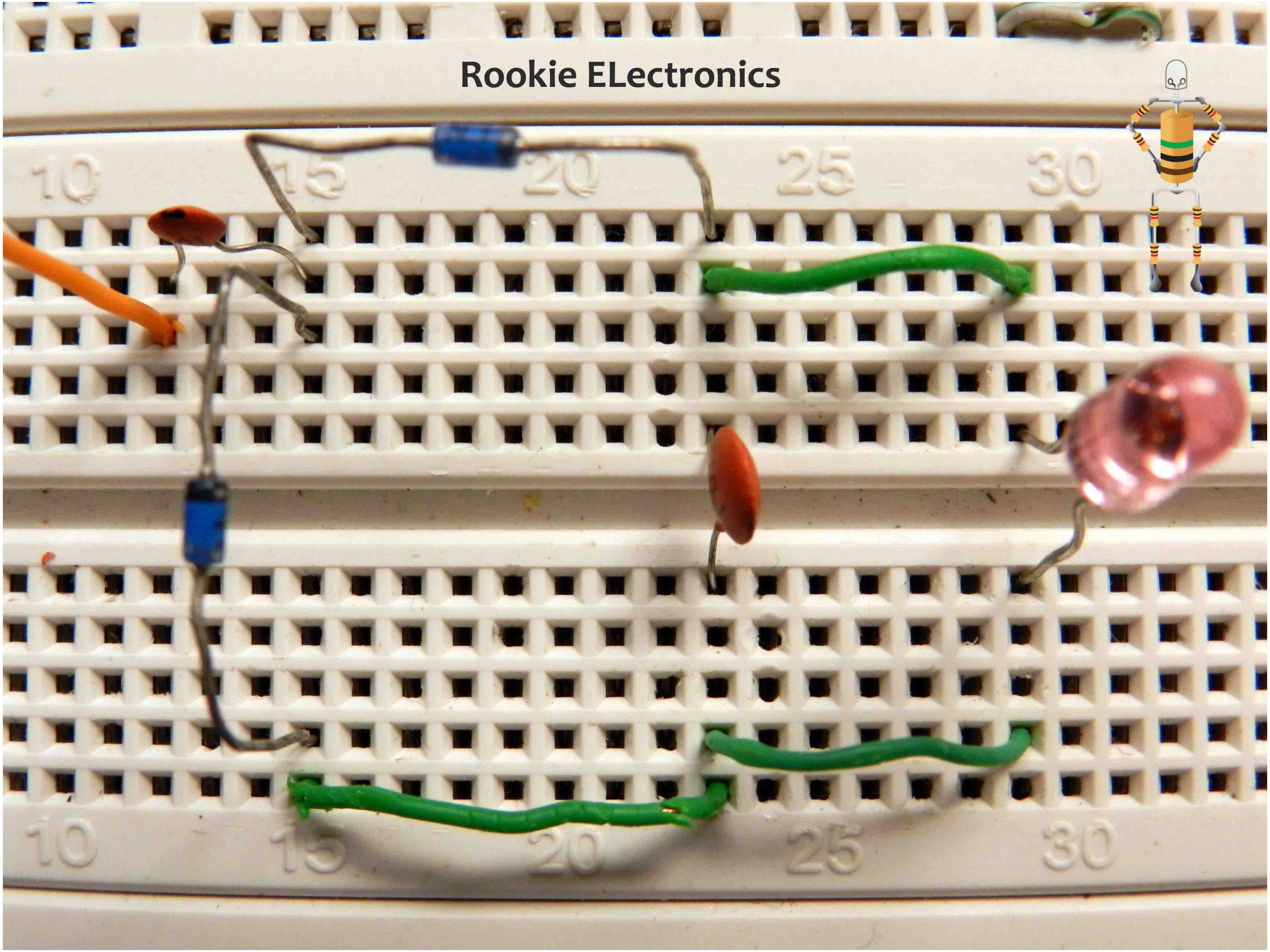
10
15
20
25
30
10
15
20
25
30

Rookie ELectronics

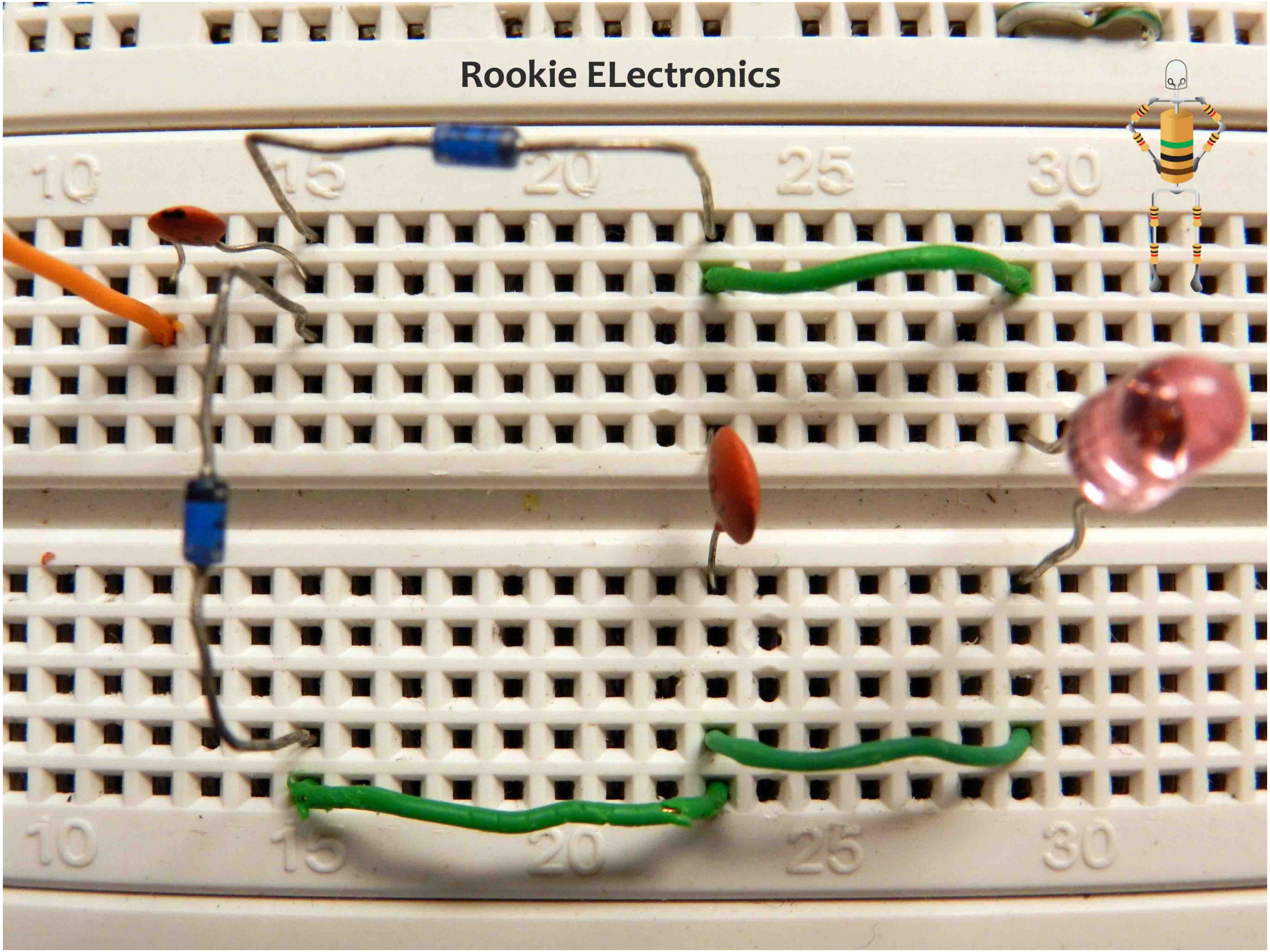
10
15
20
25
30
10
15
20
25
30

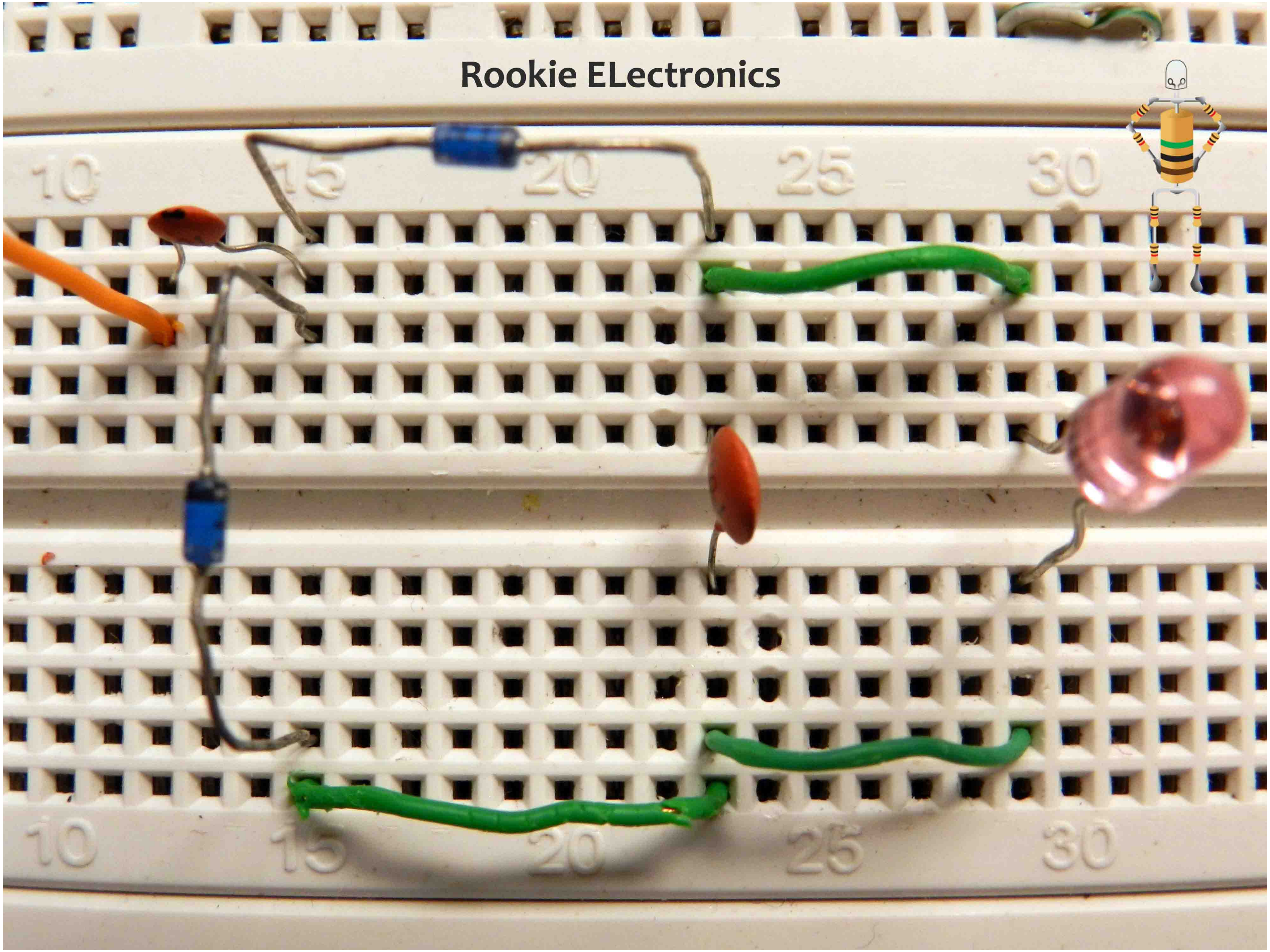
10
15
20
25
30
10
15
20
25
30

Rookie ELectronics

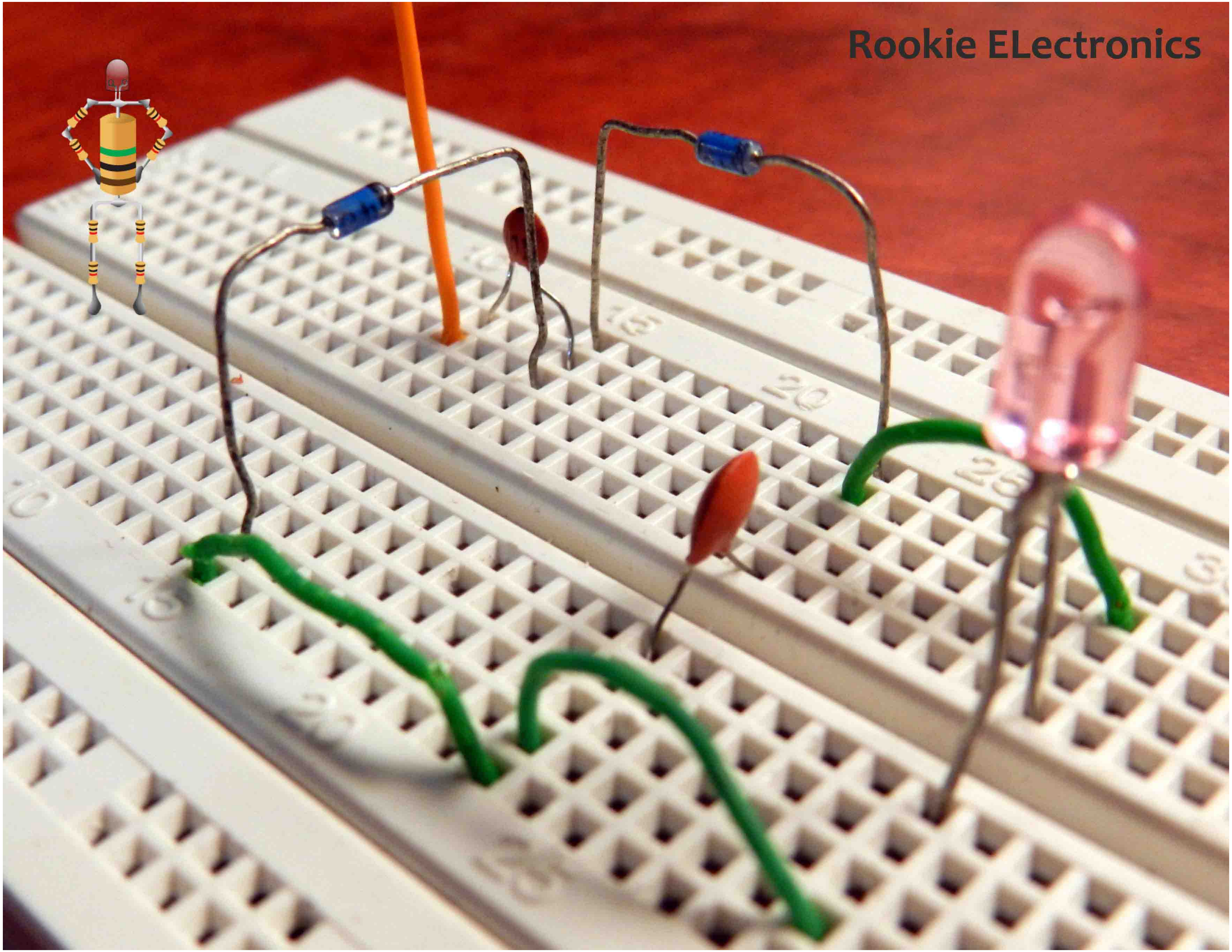
15
20
25
10
15

Rookie ELectronics

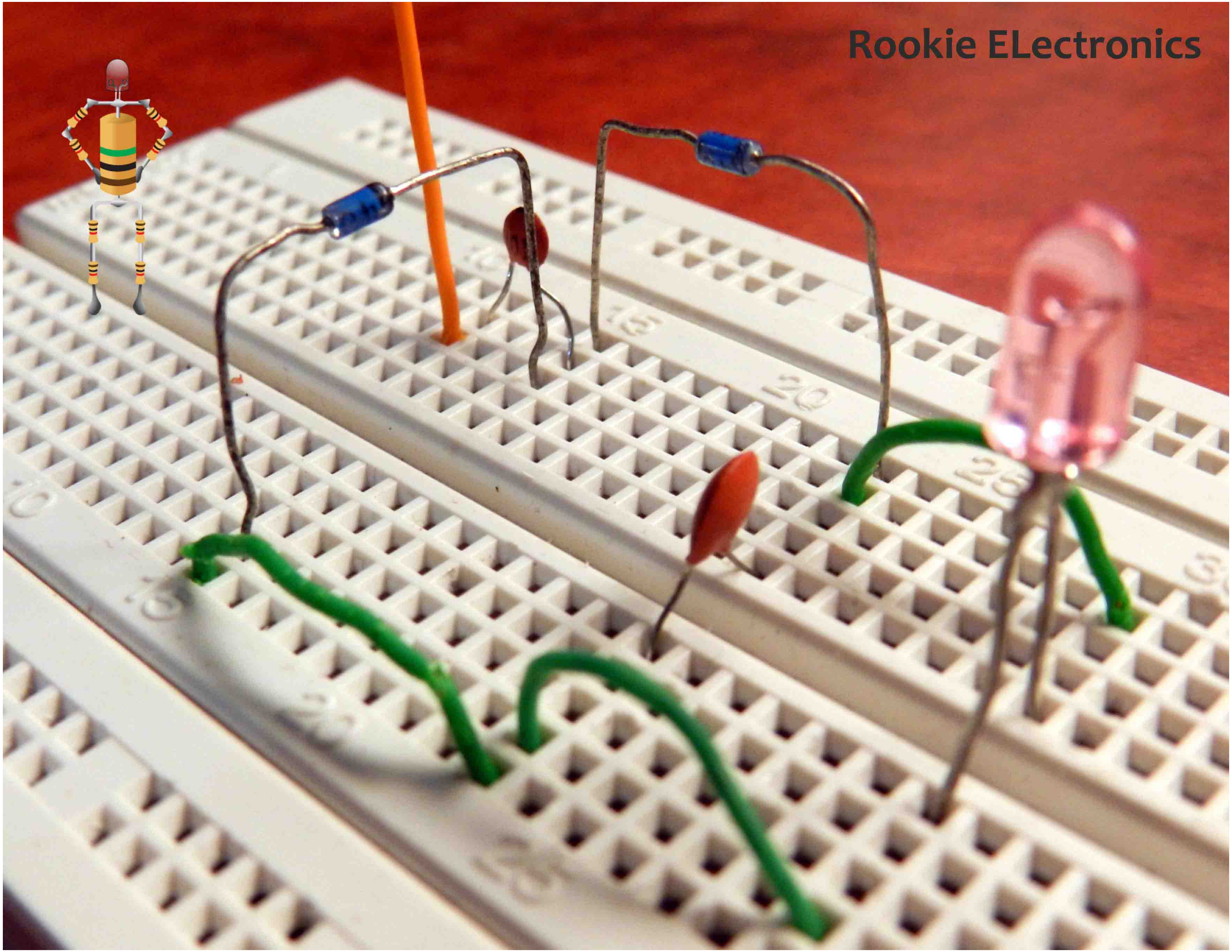

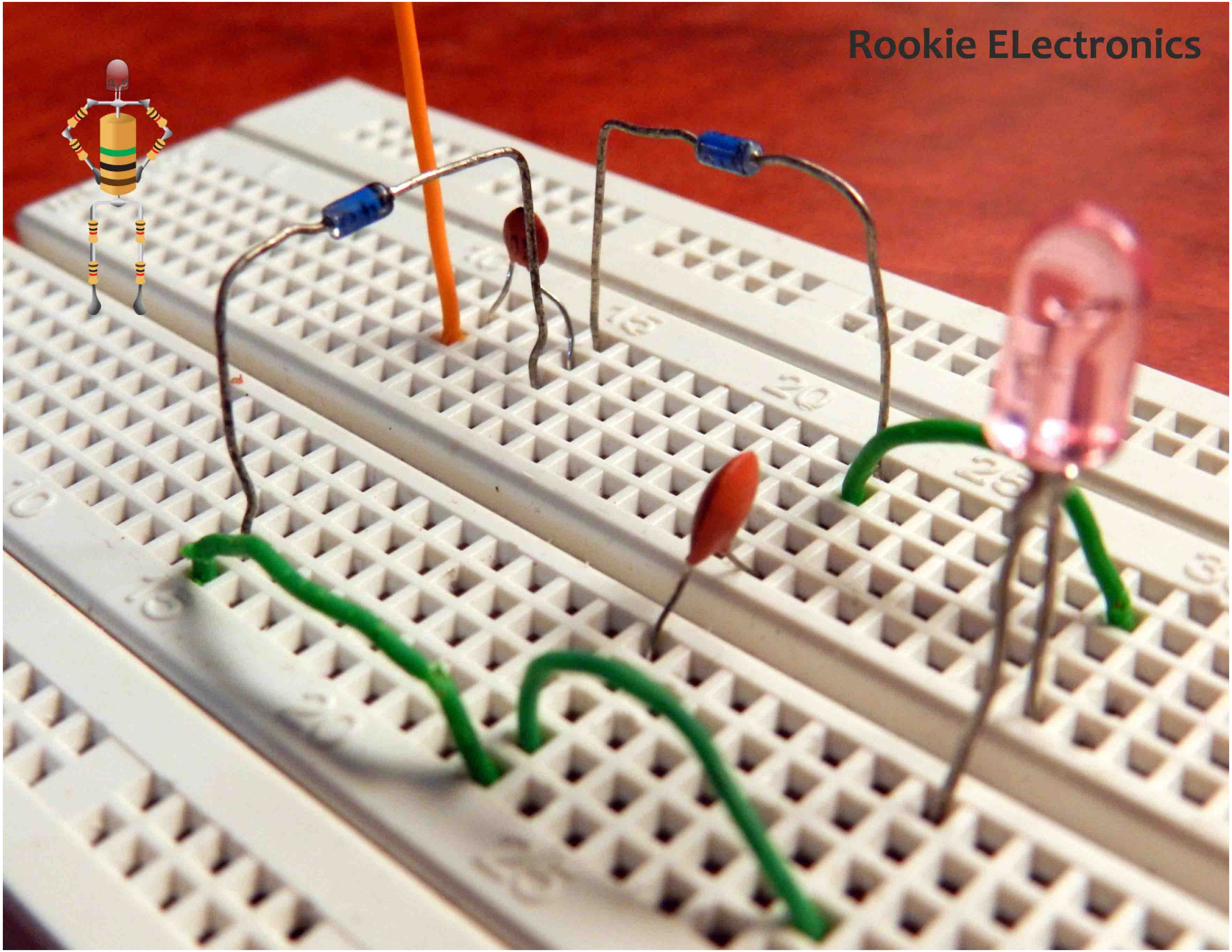

Rookie ELectronics

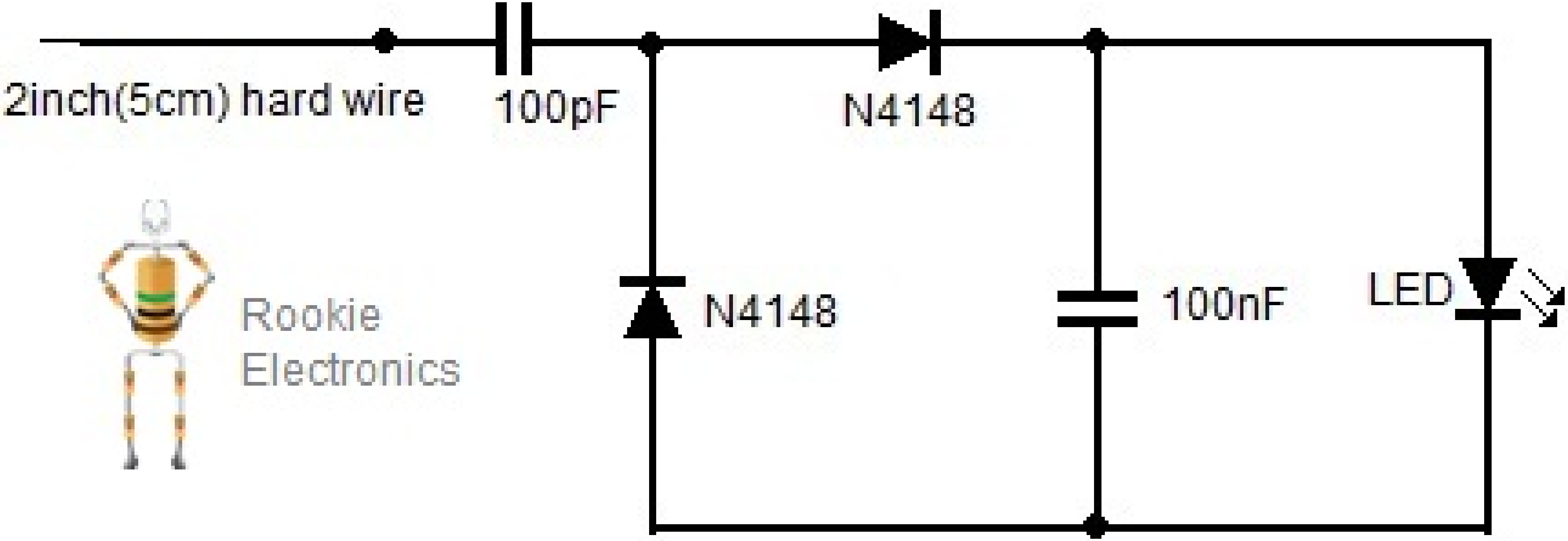
2inch(5cm) hard wire
100pF
N4148
N4148
100nF
LED

Rookie
Electronics

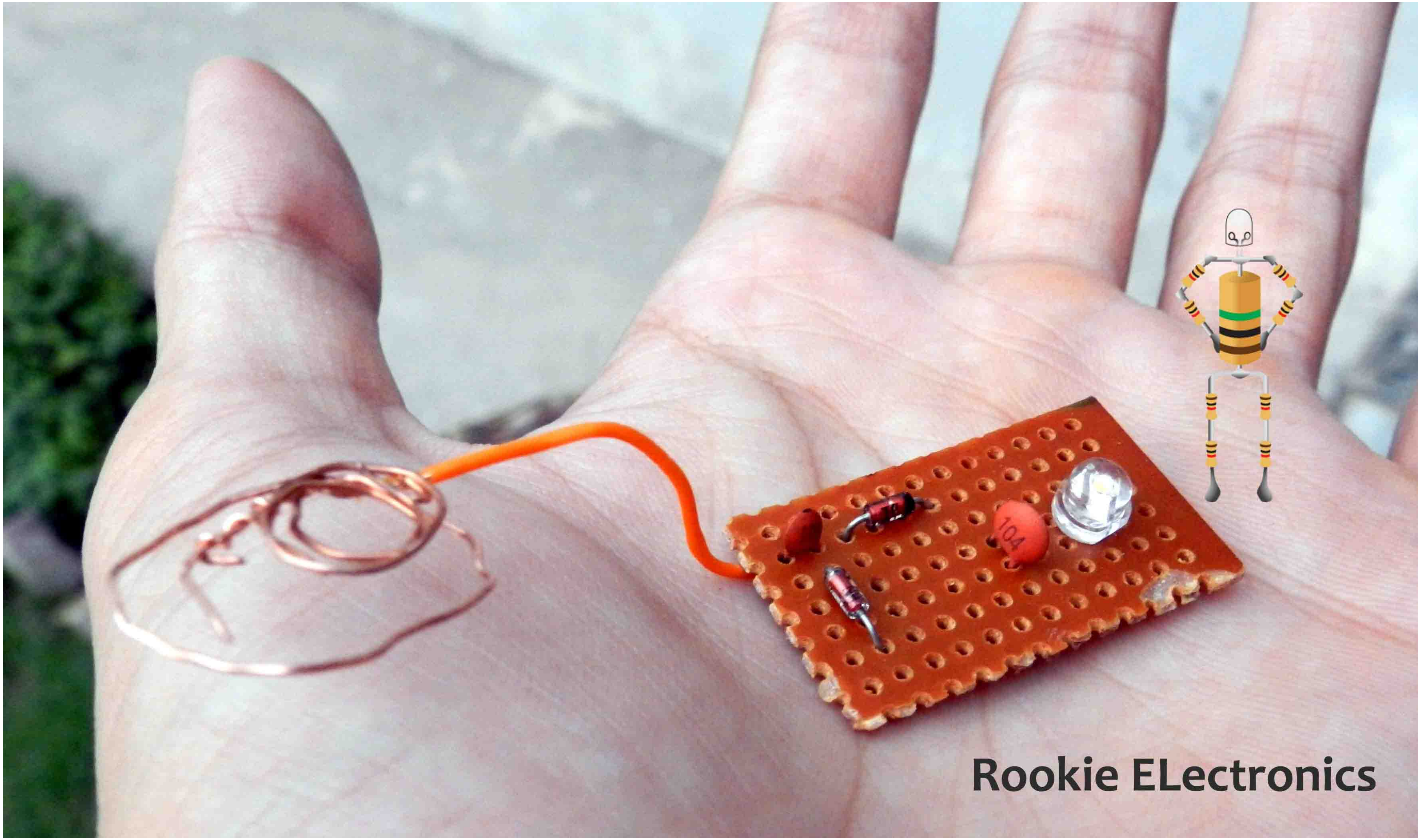

Rookie ELectronics

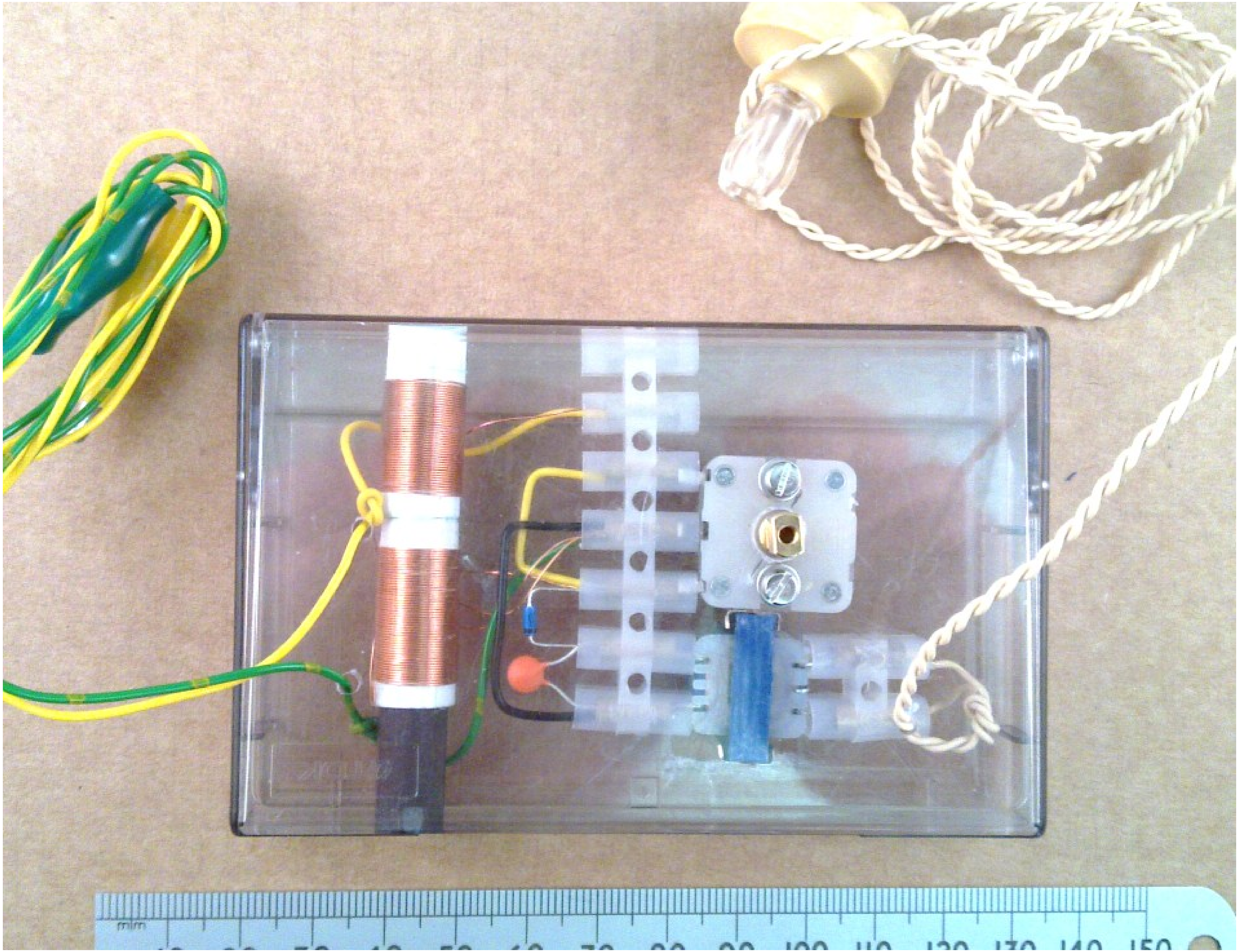

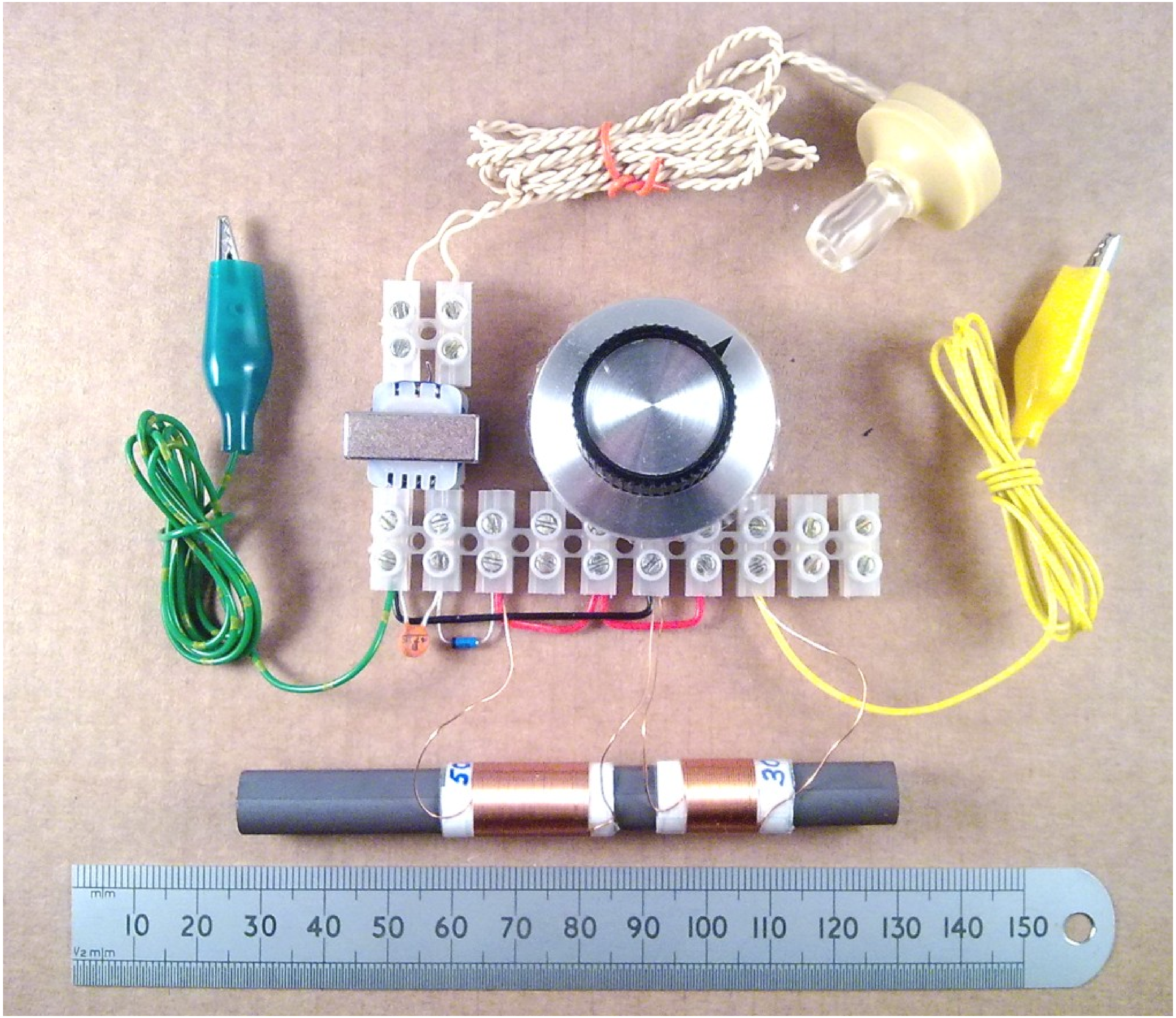
50
30
m|m
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150
½ m|m

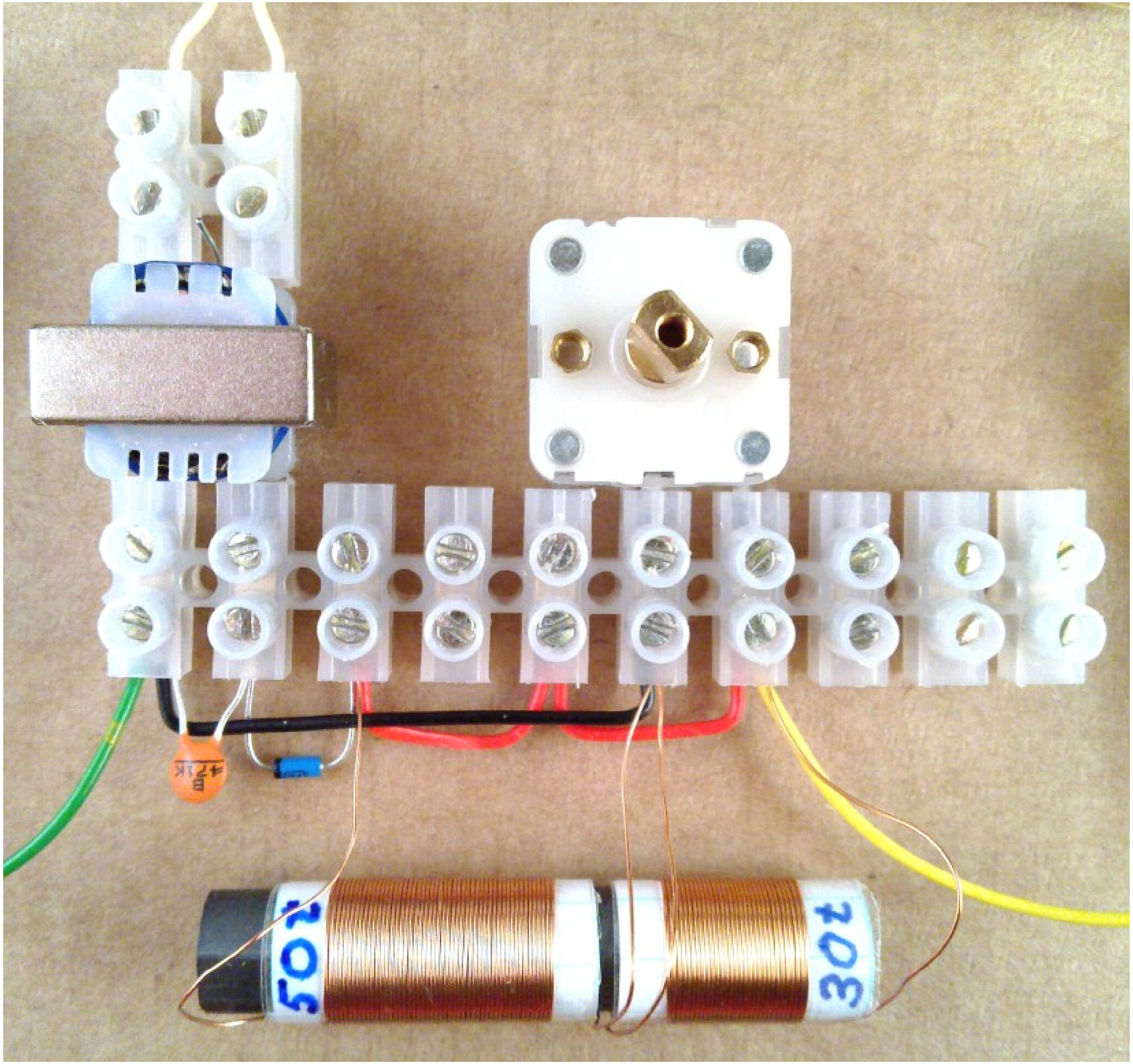
50t
30t

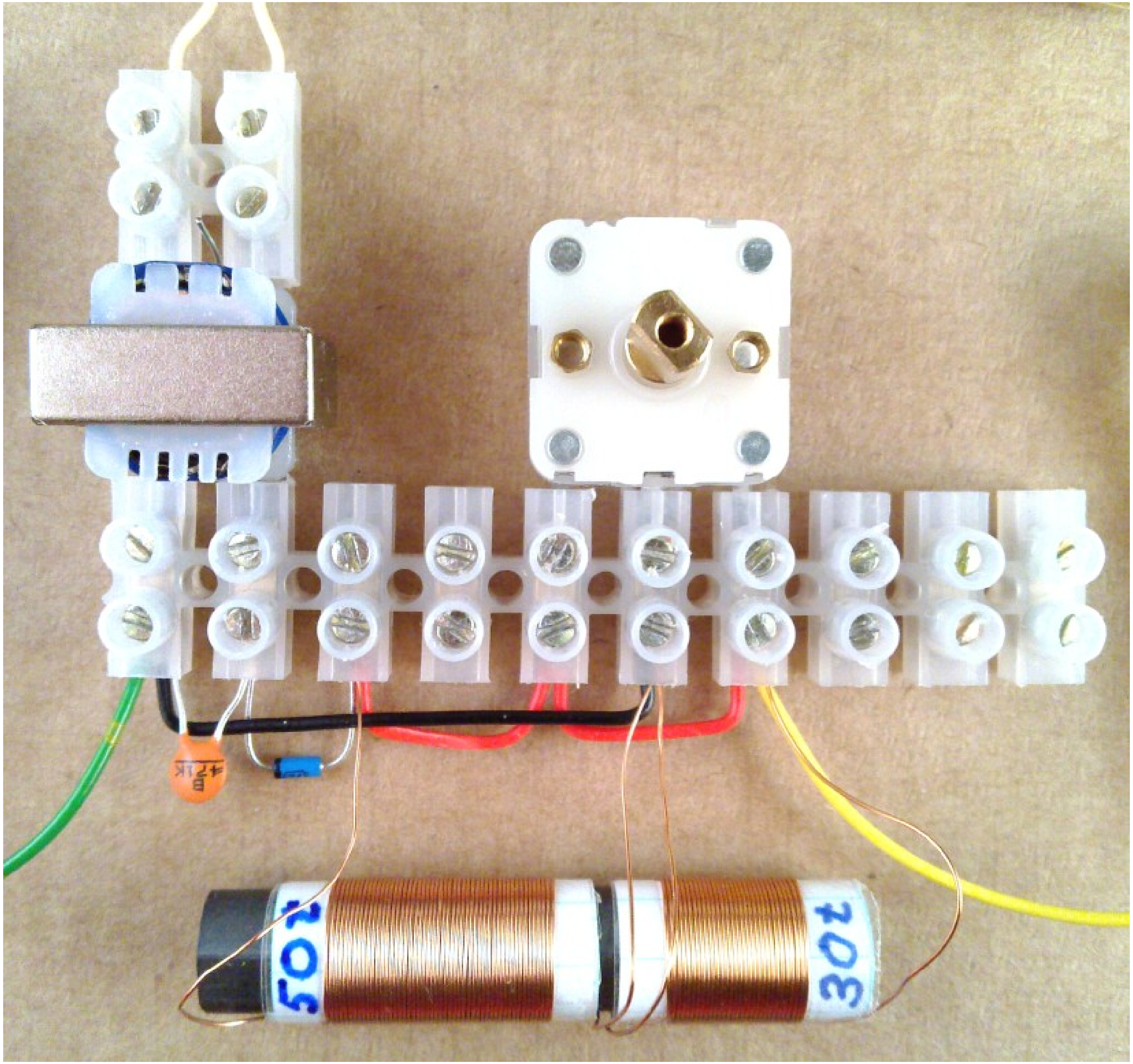
50t
30t

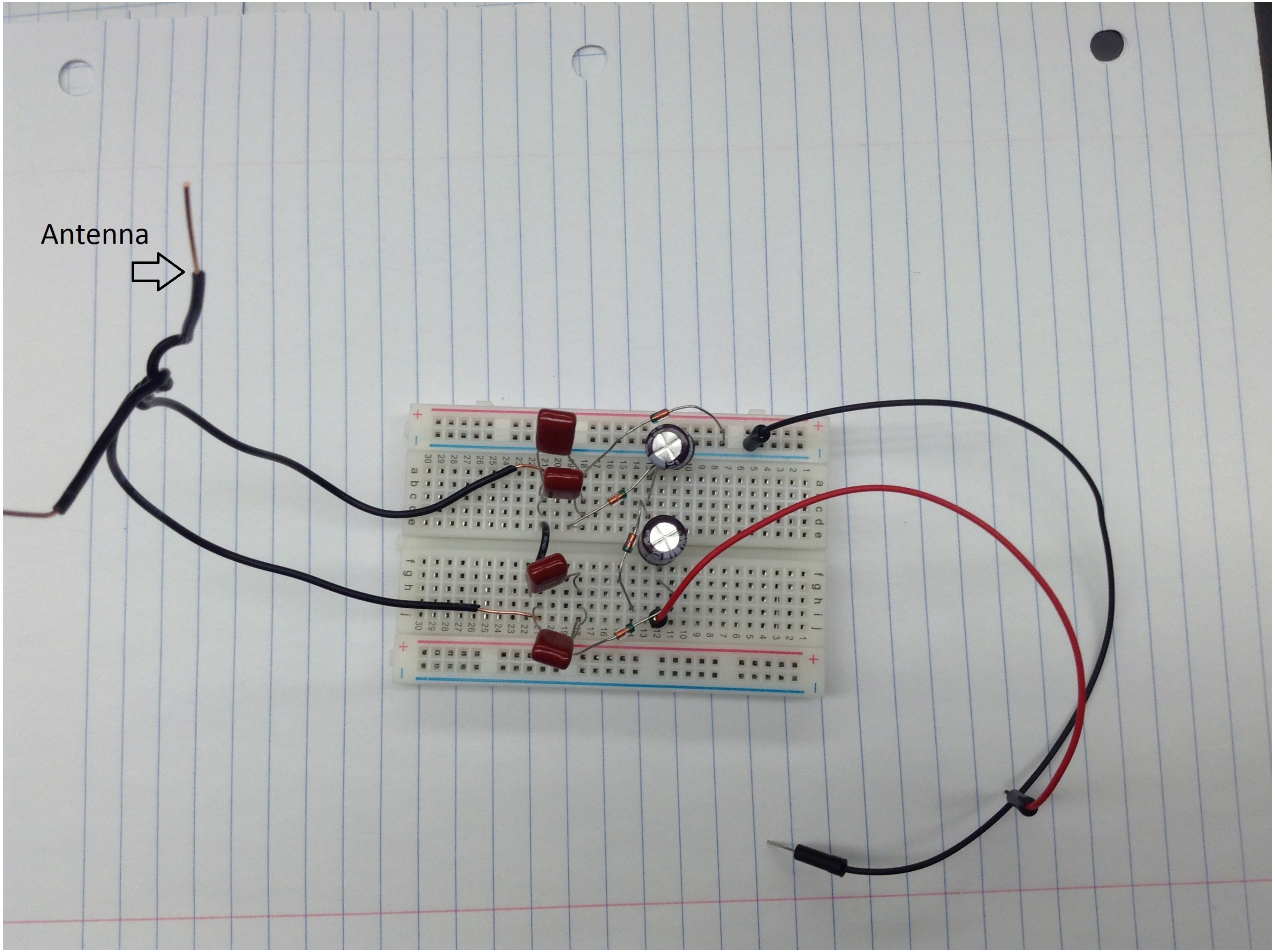
Antenna

US006960975B1

(12) United States Patent
Volfson

(10) **Patent No.:** **US 6,960,975 B1**
(45) **Date of Patent:** **Nov. 1, 2005**

(54) SPACE VEHICLE PROPELLED BY THE PRESSURE OF INFLATIONARY VACUUM STATE

(76) Inventor: **Boris Volfson**, 5707 W. Maple Grove Rd., Apt. 3046, Huntington, IN (US) 46750

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 8 days.

(21) Appl. No.: **11/079,670**

(22) Filed: **Mar. 14, 2005**

Related U.S. Application Data

(63) Continuation of application No. 10/633,778, filed on Aug. 4, 2003, now abandoned.

(51) **Int. Cl.**^{7} **H01F 6/00**; F03H 5/00
(52) **U.S. Cl.** **335/216**; 60/200.1
(58) **Field of Search** 335/216; 60/200.1

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,626,605	A	12/1971	Wallace
3,626,606	A	12/1971	Wallace
3,823,570	A	7/1974	Wallace
5,197,279	A	3/1993	Taylor
6,353,311	B1	3/2002	Brainard et al.

OTHER PUBLICATIONS

M.T. French, "To the Stars by Electromagnetic Propulsion", http://www.mtjf.demon.co.uk/antigravp2.htm#cforce.

Evgeny Podkletnov, "Weak Gravitational Shielding Properties of Composite Bulk $YBa_2Cu_3 3O_{(7-x)}$ Superconductor Below 70K Under E.M. Field", LANL database number cond-mat/9701074, v. 3, 10 pages, Sep. 16, 1997.

N. LI & D.G. Torr, "Effects of a Gravitomagnetic Field on Pure Superconductors", Physical Review, vol. 43, p. 457, 3 pages, Jan. 15, 1991.

Evgeny Podkletnov, Giovanni Modanese "Impulse Gravity Generator Based on Charged $YBa_2Cu_3 3O_{7-y}$ Superconductor with Composite Crystal Structure", arXiv.org/physics database, #0108005 vol. 2, 32 pages, 8 figures, Aug. 30, 2001.

S. Kopeikin & E. Fomalont, "General Relativistic Model for Experimental Measurement of the Speed of Propagation of Gravity by VLBI", Proceedings of the 6th European VLBI Network Symposium Jun. 25-28, 2002, Bonn, Germany, 4 pages.

Sean M. Carroll, "The Cosmological Constant", http://pancake.uchicago.edu/~carroll/encyc/, 6 pages.

Chris Y. Taylor and Giovanni Modanese, "Evaluation of an Impulse Gravity Generator Based Beamed Propulsion Concept", American Institute of Aeronautics and Astronautics, Inc., 2002.

Peter L. Skeggs, "Engineering Analysis of the Podkletnov Gravity Shielding Experiment", Quantum Forum, Nov. 7, 1997, http://www.inetarena'.com/~noetic/pls/podlev.html).

Primary Examiner—Ramon M. Barrera

(57) **ABSTRACT**

A space vehicle propelled by the pressure of inflationary vacuum state is provided comprising a hollow superconductive shield, an inner shield, a power source, a support structure, upper and lower means for generating an electromagnetic field, and a flux modulation controller.

A cooled hollow superconductive shield is energized by an electromagnetic field resulting in the quantized vortices of lattice ions projecting a gravitomagnetic field that forms a spacetime curvature anomaly outside the space vehicle. The spacetime curvature imbalance, the spacetime curvature being the same as gravity, provides for the space vehicle's propulsion. The space vehicle, surrounded by the spacetime anomaly, may move at a speed approaching the light-speed characteristic for the modified locale.

13 Claims, 6 Drawing Sheets

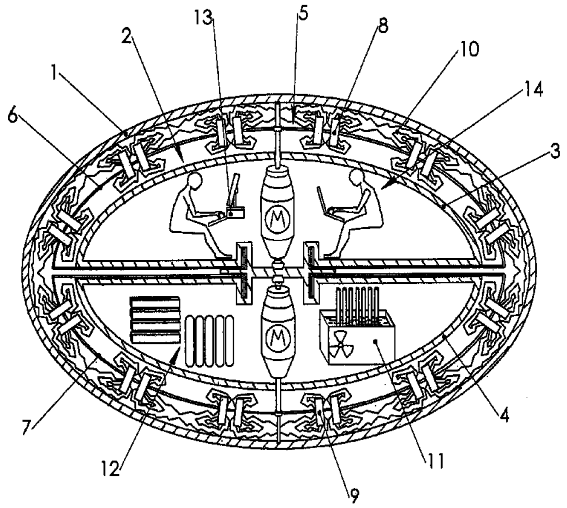

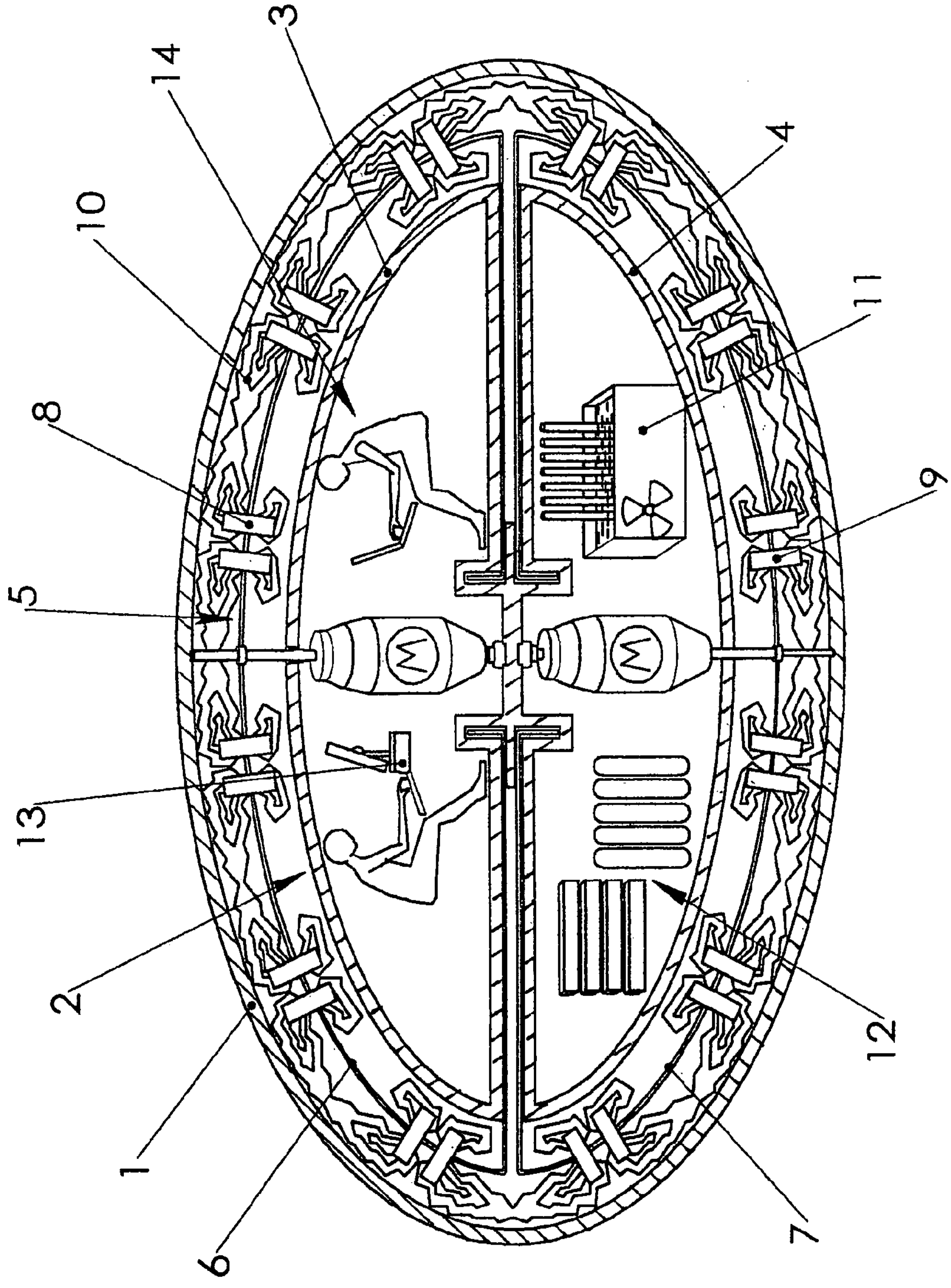
3
14
10
8
5
13
2
1
6
4
11
9
12
7

FIG. 1

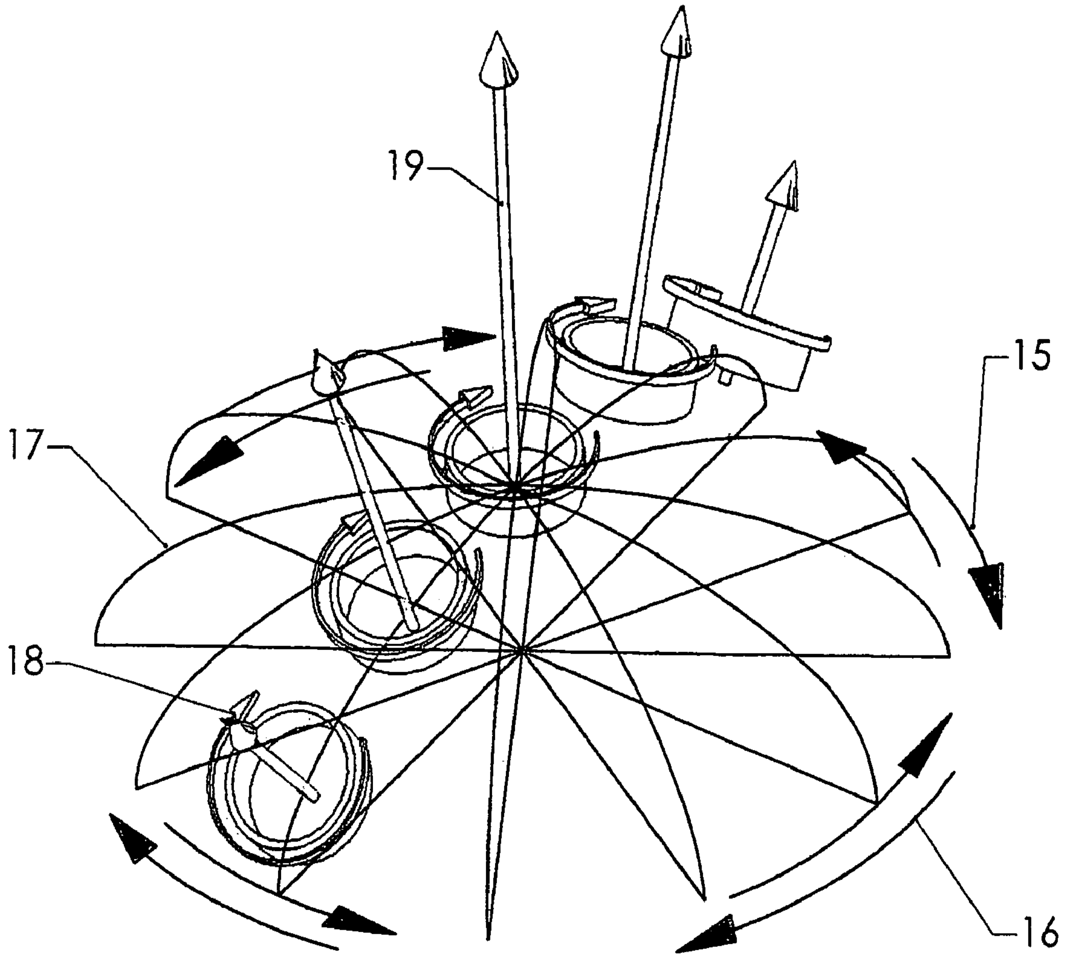
19
15
17
18
16

FIG. 2A

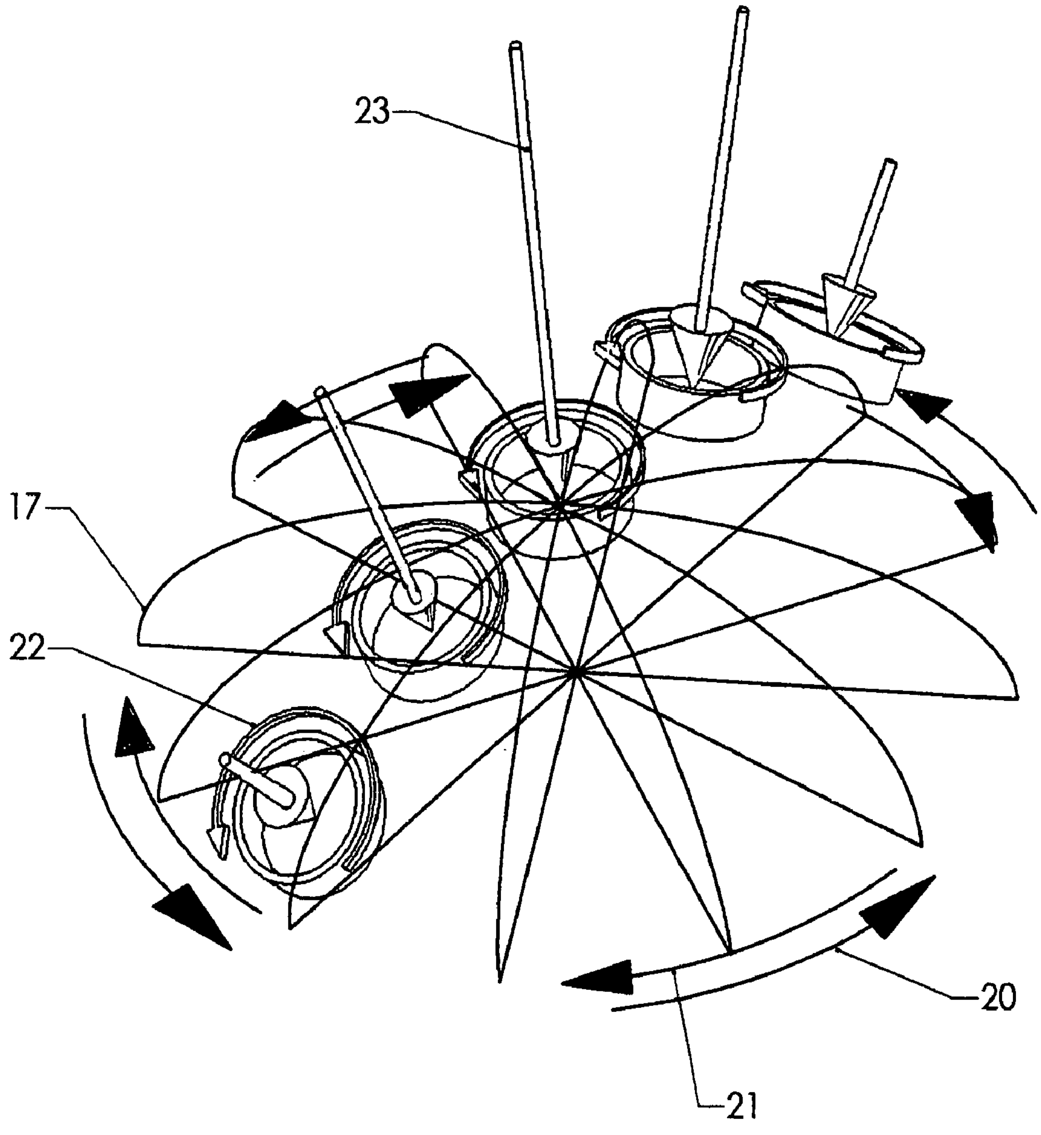
23
17
22
20
21

FIG. 2B

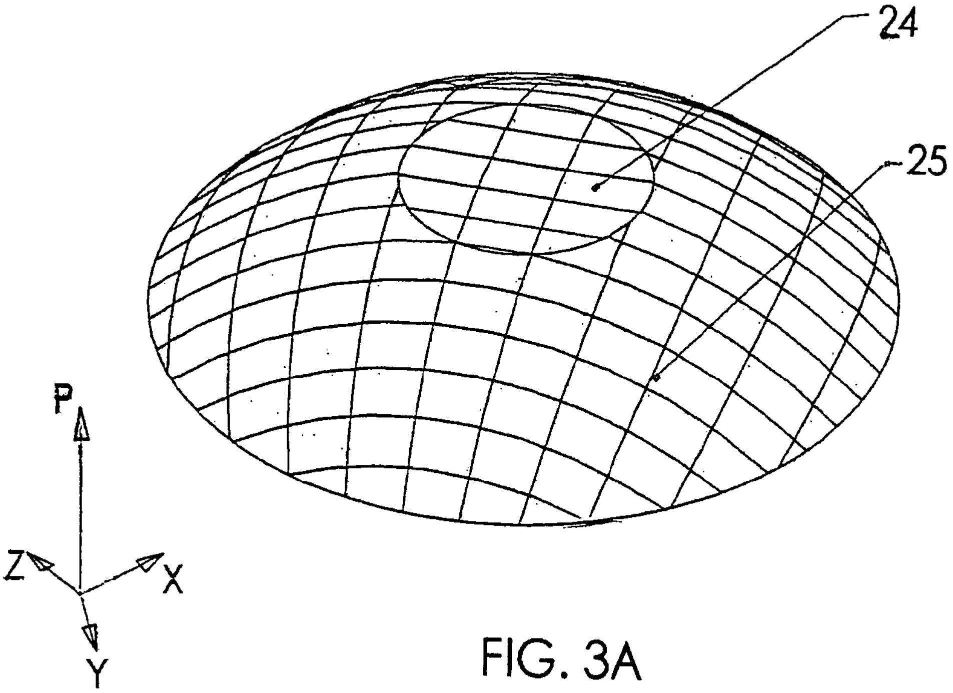

FIG. 3A

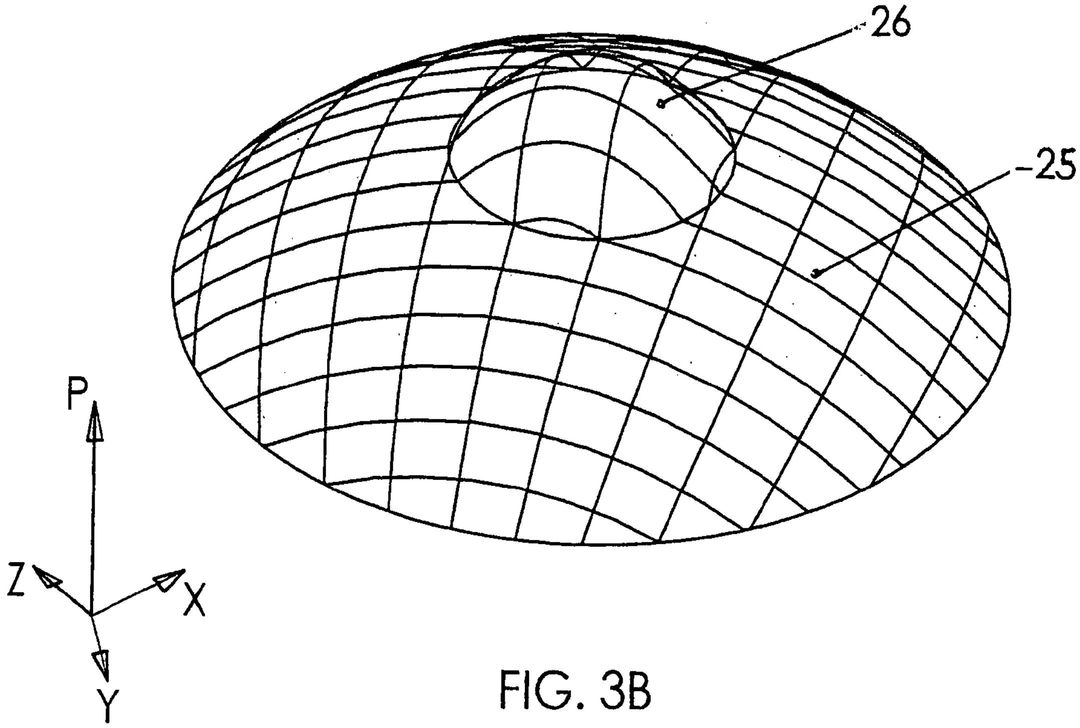

FIG. 3B

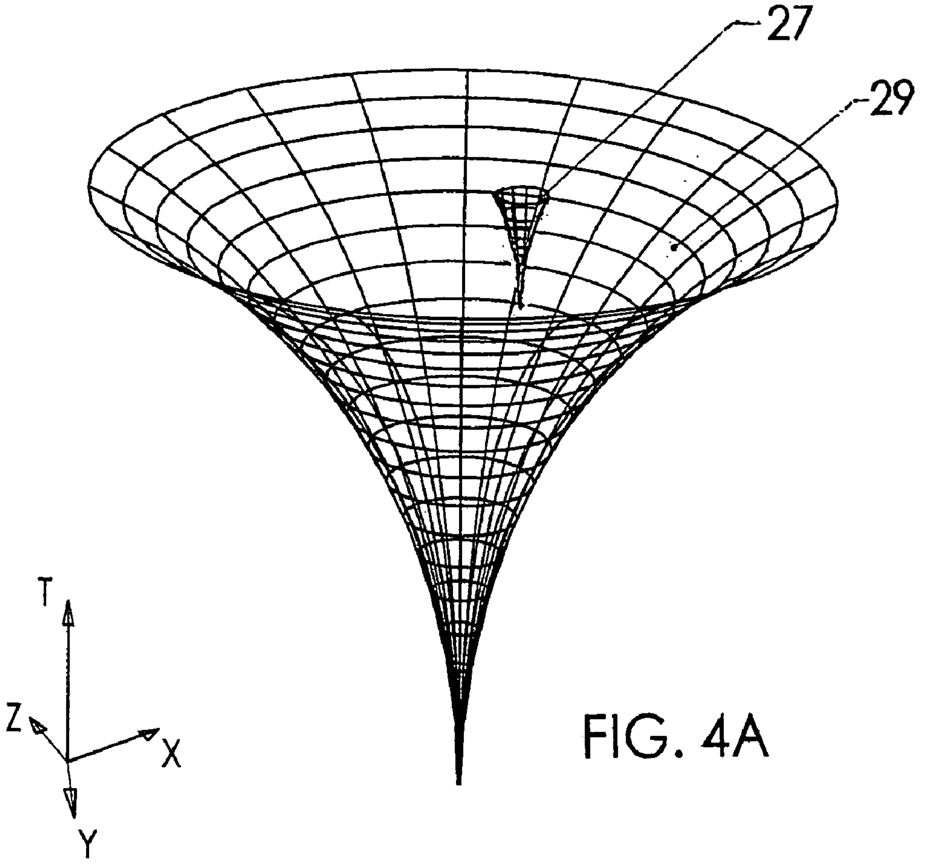

FIG. 4A

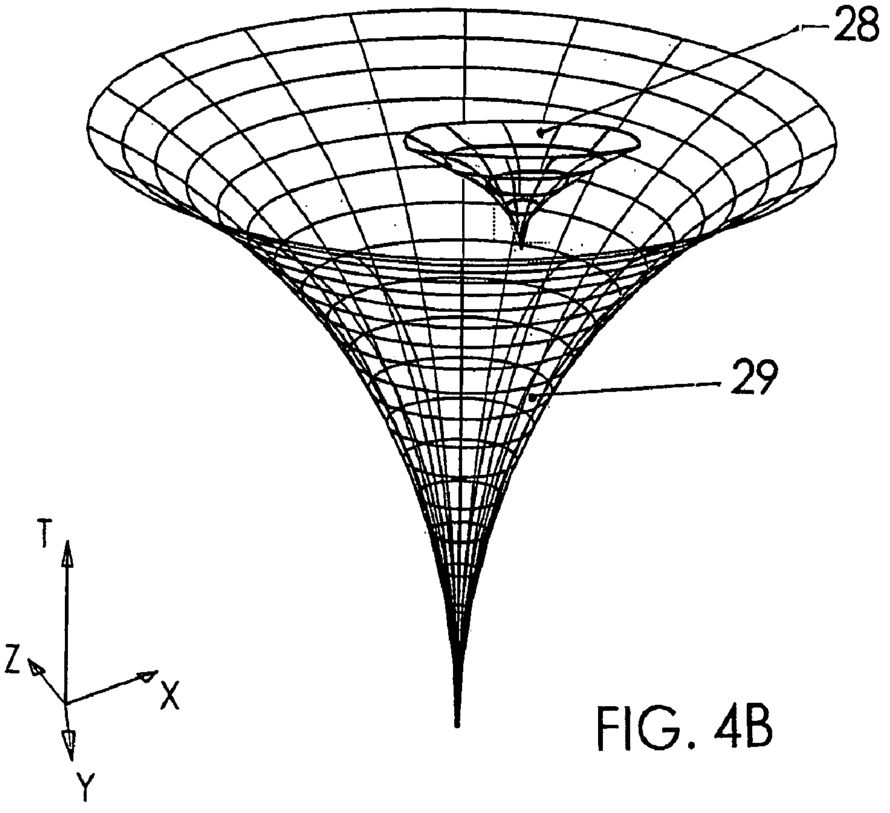

FIG. 4B

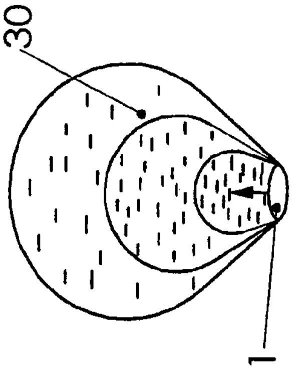

FIG.5A

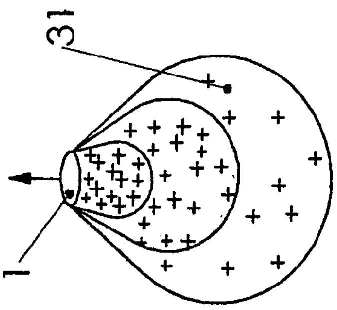

FIG.5B

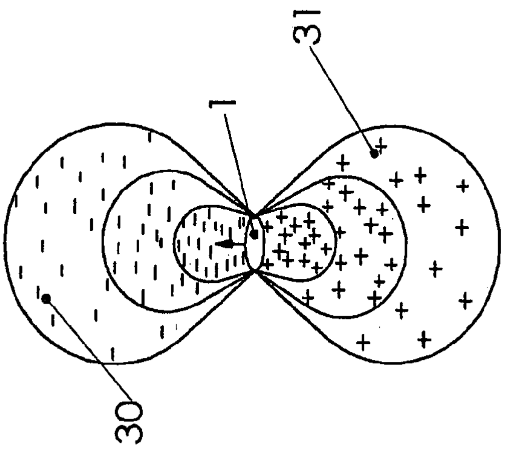

FIG.6

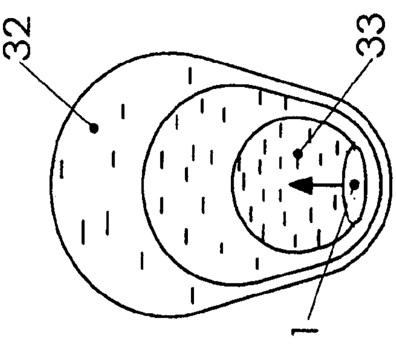

FIG.7A

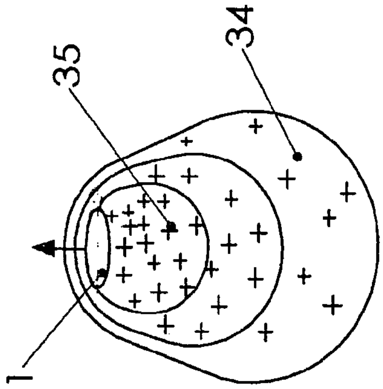

FIG.7B

SPACE VEHICLE PROPELLED BY THE PRESSURE OF INFLATIONARY VACUUM STATE

CROSS REFERENSE TO RELATED APPLICATIONS

This is a continuation of application Ser. No. 10/633,778 filed on Aug. 4, 2003, now abandoned.

TECHNICAL FIELD

The technical field of this disclosure represents space vehicles propelled by the electro-magnetically changed properties of vacuum, particularly the pressure of inflationary vacuum state.

BACKGROUND OF THE INVENTION

The existence of a magnetic-like gravitational field has been well established by physicists for general relativity, gravitational theories, and cosmology. The consequences of the effect of electromagnetically-affected gravity could be substantial and have many practical applications, particularly in aviation and space exploration.

There are methods known for converting electromagnetism into a propulsive force that potentially generates a large propulsive thrust. According to these methods, the machine thrust is produced by rotating, reciprocating masses in the following ways: centrifugal thrust, momentum thrust, and impulse thrust. ("To the Stars by Electromagnetic Propulsion", M. T. French, http://www.mtjf.demon.co.uk/antigravp2.htm#cforce).

However, the electromagnetic propulsion in an ambient space, or space that is not artificially modified, is not practical for interstellar travel because of the great distances involved. No interstellar travel is feasible without some form of distortion of space. In turn, no alteration of space is possible without the corresponding deformation of time. Gravitomagnetic alteration of space, resulting in the space-time curvature anomaly that could propel the space vehicle, could be a feasible approach to future space travel.

In the late 1940s, H. B. G. Casimir proved that the vacuum is neither particle nor field-free. It is a source of zero-point-fluctuation (ZPF) of fields such as the vacuum gravitomagnetic field. ZPF fields lead to real, measurable physical consequences such as the Casimir force. The quantized hand-made electromagnetic processes, such as those occurring in superconductors, affect the similarly quantized ZPFs. The most likely reason is the electron-positron creation and annihilation, in part corresponding to the "polarization effect" sited by Evgeny Podkletnov in explaining the gravitomagnetic effect reportedly observed by him in 1992. ("Weak Gravitational Shielding Properties of Composite Bulk $YBa_2Cu_33O_{(7-x)}$ Superconductor Below 70 K Under E.M. Field", Evgeny Podkletnov, LANL database number cond-mat/9701074, v. 3, 10 pages, 16 Sep. 1997).

The investigation of gravitomagnetism, however, started well before Podkletnov. In the U.S. Pat. No. 3,626,605, Henry Wm. Wallace describes an experimental apparatus for generating and detecting a secondary gravitational field. He also shows how a time-varying gravitomagnetic field can be used to shield the primary background of a gravitoelectric field.

In the U.S. Pat. No. 3,626,606, Henry Wm. Wallace provides a variation of his earlier experiment. A type III-V semiconductor material, of which both components have unpaired nuclear spin, is used as an electronic detector for the gravitomagnetic field. The experiment demonstrates that the material in his gravitomagnetic field circuit has hysterisis and remanence effects analogous to magnetic materials.

In the U.S. Pat. No. 3,823,570, Henry Wm. Wallace provides an additional variation of his experiment. Wallace demonstrates that, by aligning the nuclear spin of materials having an odd number of nucleons, a change in specific heat occurs.

In the U.S. Pat. No. 5,197,279, James R. Taylor discloses Electromagnetic Propulsion Engine where solenoidal windings generate an electromagnetic field that, without the conversion into a gravitomagnetic field, generates the thrust necessary for the propulsion.

In the U.S. Pat. No. 6,353,311 B1, John P. Brainard et al. offer a controversial theory of Universal Particle Flux Field, and in order to prove it empirically, provide a shaded motor-type device. This device is also intended for extracting energy from this hypothetical Field.

In the early 1980s, Sidney Coleman and F. de Luca noted that the Einsteinean postulate of a homogeneous Universe, while correct in general, ignores quantized local fluctuation of the pressure of inflationary vacuum state, this fluctuation causing local cosmic calamities. While the massless particles propagate through large portions of Universe at light speed, these anomaly bubbles, depending on their low or high relative vacuum density, cause a local increase or decrease of the propagation values for these particles. Scientists disagree about the possibility, and possible ways, to artificially create models of such anomalies.

In the early 1990s, Ning Li and D. G Torr described a method and means for converting an electromagnetic field into a gravitomagnetic field. Li and Torr suggested that, under the proper conditions, the minuscule force fields of superconducting atoms can "couple", compounding in strength to the point where they can produce a repulsion force ("Effects of a Gravitomagnetic Field on Pure Superconductors", N. Li and D. G. Torr, Physical Review, Volume 43, Page 457, 3 pages, 15 Jan. 1991).

A series of experiments, performed in the early 1990s by Podkletnov and R. Nieminen, reportedly resulted in a reduction of the weights of objects placed above a levitating, rotating superconductive disk subjected to high frequency magnetic fields. These results substantially support the expansion of Einstainean physics offered by Li & Torr. Podkletnov and Giovanni Modanese have provided a number of interesting theories as to why the weight reduction effect could have occurred, citing quantum gravitational effects, specifically, a local change in the cosmological constant. The cosmological constant, under ordinary circumstances, is the same everywhere. But, according to Podkletnov and Modanese, above a levitating, rotating superconductive disk exposed to high frequency magnetic fields, it is modified. ("Impulse Gravity Generator Based on Charged $YBa_2Cu_33O_{7-y}$ Superconductor with Composite Crystal Structure", Evgeny Podkletnov, Giovanni Modanese, arXiv.org/physics database, #0108005 volume 2, 32 pages, 8 figures, Aug. 30, 2001).

In the July 2004 paper, Ning Wu hypothesized that exponential decay of the gravitation gauge field, characteristic for the unstable vacuum such as that created by Podkletnov and Nieminen, is at the root of the gravitational shielding effects (Gravitational Shielding Effects in Gauge Theory of Gravity, Ning Wu, arXiv:hep-th/ 0307225 v 1 23 Jul. 2003, 38 pages incl. 3 figures, July 2004).

In 2002, Edward Fomalont and Sergei Kopeikin measured the speed of propagation of gravity. They confirmed that the speed of propagation of gravity matches the speed of light. ("General Relativistic Model for Experimental Measurement of the Speed of Propagation of Gravity by VLBI", S. Kopeikin and E. Fomalont, Proceedings of the 6th European VLBI Network Symposium Jun. 25–28 2002, Bonn, Germany, 4 pages).

String theory unifies gravity with all other known forces. According to String theory, all interactions are carried by fundamental particles, and all particles are just tiny loops of space itself forming the spacetime curvature. Gravity and bent space are the same thing, propagating with the speed of light characteristic of the particular curvature. In light of the Fomalont and Kopeikin discovery, one can conclude that if there is a change in the speed of propagation of gravity within the spacetime curvature, then the speed of light within the locality would also be affected.

In general relativity, any form of energy affects the gravitational field, so the vacuum energy density becomes a potentially crucial ingredient. Traditionally, the vacuum is assumed to be the same everywhere in the Universe, so the vacuum energy density is a universal number. The cosmological constant Λ is proportional to the vacuum pressure ρ_Λ: $\Lambda=(8\pi G/3c^2)\rho_\Lambda$, where G is Newton's constant of gravitation and c is the speed of light ("The Cosmological Constant", Sean M. Carroll, http://pancake.uchicago.edu/~carroll/encyc/, 6 pages). Newer theories, however, permit local vacuum fluctuations where even the "universal" constants are affected: $\Lambda_1=(8\pi G_1/3c_1^2)\rho_{\Lambda_1}$.

Analyzing physics laws defining the cosmological constant, a conclusion can be drawn that, if a levitating, rotating superconductive disk subjected to high frequency magnetic fields affects the cosmological constant within a locality, it would also affect the vacuum energy density. According to the general relativity theory, the gravitational attraction is explained as the result of the curvature of spacetime being proportional to the cosmological constant. Thus, the change in the gravitational attraction of the vacuum's subatomic particles would cause a local anomaly in the curvature of the Einsteinean spacetime.

Time is the fourth dimension. Lorentz and Einstein showed that space and time are intrinsically related. Later in his life, Einstein hypothesized that time fluctuates both locally and universally. Ruggero Santilli, recognized for expanding relativity theory, has developed the isocosmology theory, which allows for variable rates of time. Time is also a force field only detected at speeds above light speed. The energy of this force field grows as its propagation speed declines when approaching light-speed. Not just any light-speed: the light-speed of a locale. If the conditions of the locale were modified, this change would affect the local time rate relative to the rate outside the affected locale, or ambient rate. The electromagnetically-generated gravitomagnetic field could be one such locale modifier.

Analyzing the expansion of Einstainean physics offered by Li & Torr, one could conclude that gravity, time, and light speed could be altered by the application of electromagnetic force to a superconductor.

By creating a spacetime curvature anomaly associated with lowered pressure of inflationary vacuum state around a space vehicle, with the lowest vacuum pressure density located directly in front of the vehicle, a condition could be created where gravity associated with lowered vacuum pressure density pulls the vehicle forward in modified spacetime.

By creating a spacetime curvature anomaly associated with elevated pressure of inflationary vacuum state around the space vehicle, with the point of highest vacuum pressure density located directly behind the vehicle, a condition could be created where a repulsion force associated with elevated vacuum pressure density pushes the space vehicle forward in modified spacetime. From the above-mentioned cosmological constant equation, re-written as:

$$\rho_\Lambda = \frac{3c^2}{8\pi G}\Lambda$$

it is clear that the increase in the vacuum pressure density could lead to a substantial increase in the light-speed. If the space vehicle is moving in the anomaly where the local light-speed is higher than the light-speed of the ambient vacuum, and if this vehicle approaches this local light-speed, the space vehicle would then possibly exceed the light-speed characteristic for the ambient area.

The levitating and rotating superconductor disk, which Podkletnov used to protect the object of experiment from the attraction produced by the energy of the vacuum, was externally energized by the externally-powered solenoid coils. Thus, Podkletnov's system is stationary by definition and not suitable for travel in air or space. Even if the superconductive disk is made part of the craft, and if it is energized by the energy available on the craft, the resulting anomaly is one-sided, not enveloping, and not providing the variable speed of light (VSL) environment for the craft.

In a recent (2002) article, Chris Y. Tailor and Modanese propose to employ an impulse gravity generator directing, from an outside location, an anomalous beam toward a spacecraft, this beam acting as a repulsion force field producing propulsion for the spacecraft. ("Evaluation of an Impulse Gravity Generator Based Beamed Propulsion Concept", Chris Y. Taylor and Giovanni Modanese, American Institute of Aeronautics and Astronautics, Inc., 2002, 21 pages, 10 figures). The authors of the article, however, didn't take into account the powerful quantized processes of field dispersion, which would greatly limit the distance of propagation of the repulsive force. At best, the implementation of this concept could assist in acceleration and deceleration at short distances from the impulse gravity generator, and only along a straight line of travel. If the travel goal is a space exploration mission rather than the shuttle-like commute, the proposed system is of little use.

Only a self-sufficient craft, equipped with the internal gravity generator and the internal energy source powering this generator, would have the flexibility needed to explore new frontiers of space. The modification of the spacetime curvature all around the spacecraft would allow the spacecraft to approach the light-speed characteristic for the modified locale, this light-speed, when observed from a location in the ambient space, being potentially many times higher than the ambient light-speed. Then, under sufficient local energies, that is, energies available on the spacecraft, very large intergalactic distances could be reduced to conventional planetary distances.

In "The First Men in the Moon" (1903), H. G. Wells anticipates gravitational propulsion methods when he describes gravity repelling "cavorite." Discovered by Professor Cavor, the material acts as a "gravity shield" allowing Cavor's vehicle to reach the Moon. Prof. Cavor built a large spherical gondola surrounded on all sides by cavorite shutters that could be closed or opened. When Prof. Cavor closed all the shutters facing the ground and opened the shutters facing the moon, the gondola took off for the Moon.

Until today, no cavorite has been discovered. However, recent research in the area of superconductivity, nanomaterials and quantum state of vacuum, including that of Li, Torr, Podkletnov, and Modanese, has resulted in important new information about the interaction between a gravitational field and special states of matter at a quantum level. This new research opens the possibility of using new electromagnetically-energized superconductive materials allowing stable states of energy, the materials useful not only in controlling the local gravitational fields, but also in creating new gravitomagnetic fields.

BACKGROUND OF INVENTION: OBJECTS AND ADVANTAGES

There are four objects of this invention:

The first object is to provide a method for generating a pressure anomaly of inflationary vacuum state that leads to electromagnetic propulsion.

The second object is to provide a space vehicle capable of electromagnetically-generated propulsion. The implementation of these two objects leads to the development of the space vehicle propelled by gravitational imbalance with gravity pulling, and/or antigravity pushing, the space vehicle forward.

The third object is to provide a method for generating a pressure anomaly of inflationary vacuum state, specifically, the local increase in the level of vacuum pressure density associated with the greater curvature of spacetime. The speed of light in such an anomaly would be higher than the speed of light in the ambient space.

The fourth object is to provide the space vehicle capable of generating an unequally-distributed external anomaly all around this vehicle, specifically the anomaly with the elevated level of vacuum pressure density. The anomaly is formed in such a way that gravity pulls the space vehicle forward in the modified spacetime at a speed possibly approaching the light-speed specific for this modified locale. If the vacuum pressure density of the locale is modified to be substantially higher than that of the ambient vacuum, the speed of the vehicle could conceivably be higher than the ambient light-speed.

SUMMARY OF THE INVENTION

This invention concerns devices self-propelled by the artificially changed properties of the pressure of inflationary vacuum state to speeds possibly approaching the light-speed specific for this modified locale. Furthermore, this invention concerns devices capable of generating the spacetime anomaly characterized by the elevated vacuum pressure density. The devices combining these capabilities may be able to move at speeds substantially higher than the light-speed in the ambient space.

The device of this invention is a space vehicle. The outside shell of the space vehicle is formed by a hollow disk, sphere, or the like hollowed 3-dimensional shape made of a superconductor material, hereinafter a hollow superconductive shield. An inner shield is disposed inside the hollow superconductive shield. The inner shield is provided to protect crew and life-support equipment inside.

A support structure, upper means for generating an electromagnetic field and lower means for generating an electromagnetic field are disposed between the hollow superconductive shield and the inner shield. A flux modulation controller is disposed inside the inner shield to be accessible to the crew.

Electrical energy is generated in a power source disposed inside the hollow superconductive shield. The electrical energy is converted into an electromagnetic field in the upper means for generating an electromagnetic field and the lower means for generating an electromagnetic field.

Electrical motors, also disposed inside the hollow superconductive shield, convert the electrical energy into mechanical energy.

The mechanical energy and the electromagnetic field rotate the hollow superconductive shield, and the upper and the lower means for generating an electromagnetic field, against each other.

The electromagnetic field is converted into a gravitomagnetic field in the hollow superconductive shield.

The gravitomagnetic field, propagated outward, orthogonally to the walls of the hollow superconductive shield, forms a pressure anomaly of inflationary vacuum state in the area of propagation. The pressure anomaly of inflationary vacuum state is comprised of an area of relatively lower vacuum pressure density in front of the space vehicle and an area of relatively higher vacuum pressure density behind the vehicle.

The difference in the vacuum pressure density propels the space vehicle of this invention forward.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a cross-sectional view through the front plane taken along the central axis of a space vehicle provided by the method and device of this invention.

FIGS. **2**A & **2**B are diagrams, presented as perspective views, showing some of the physical processes resulting from a dynamic application of an electromagnetic field to a hollow superconductive shield. Only one line of quantized vortices, shown out of scale, is presented for illustration purposes.

FIGS. **3**A & **3**B are diagrams, presented as perspective views, showing a vacuum pressure density anomaly associated with lowered pressure of inflationary vacuum state and a vacuum pressure density anomaly associated with elevated pressure of inflationary vacuum state, respectively. Both anomalies are shown on the background of Universal curvature of inflationary vacuum state.

FIGS. **4**A & **4**B are diagrams, presented as perspective views, showing a spacetime anomaly associated with lowered pressure of inflationary vacuum state and a spacetime anomaly associated with elevated pressure of inflationary vacuum state, respectively. Both anomalies are shown on the background of Universal spacetime.

FIGS. **5**A, **5**B, **6**, **7**A, & **7**B are diagrams of spacetime curvature anomalies generated by the space vehicle of the current invention, these anomalies providing for the propulsion of the space vehicle.

DRAWINGS—REFERENCE NUMERALS

1 hollow superconductive shield
2 inner shield
3 upper shell
4 lower shell
5 support structure
6 upper rotating element
7 lower rotating element
8 upper means for generating an electromagnetic field
9 lower means for generating an electromagnetic field
10 flux lines
11 power source

12 life-support equipment
13 flux modulation controller
14 crew
15 clockwise shield motion vector
16 counterclockwise EMF motion vector
17 wire grid
18 clockwise quantized vortices of lattice ions
19 outward gravitomagnetic field vector
20 counterclockwise shield motion vector
21 clockwise EMF motion vector
22 counterclockwise quantized vortices of lattice ions
23 inward gravitomagnetic field vector
24 vacuum pressure density anomaly associated with lowered pressure of inflationary vacuum state
25 Universal curvature of inflationary vacuum state
26 vacuum pressure density anomaly associated with elevated pressure of inflationary vacuum state
27 spacetime anomaly associated with lowered pressure of inflationary vacuum state
28 spacetime anomaly associated with elevated pressure of inflationary vacuum state
29 Universal spacetime
30 substantially droplet-shaped spacetime curvature anomaly associated with lowered pressure of inflationary vacuum state
31 substantially droplet-shaped spacetime anomaly associated with elevated pressure of inflationary vacuum state
32 substantially egg-shaped spacetime anomaly associated with lowered pressure of inflationary vacuum state
33 area of the lowest vacuum pressure density
34 substantially egg-shaped spacetime anomaly associated with elevated pressure of inflationary vacuum state
35 area of the highest vacuum pressure density

DESCRIPTION OF THE PRESENTLY PREFERRED EMBODIMENT

FIG. **1** is a cross-sectional view through the front plane taken along the central axis of a space vehicle provided by the method and device of this invention. A hollow superconductive shield **1** forms a protective outer shell of the space vehicle. The hollow superconductive shield **1** may be shaped as a hollow disk, sphere, or the like 3-dimensional geometrical figure formed by the 2-dimensional rotation of a curve around the central axis.

In the preferred embodiment, the hollow superconductive shield **1** is made of a superconductor such as YBa2Cu3O7-y, or a like high-temperature superconductor with a composite crystal structure cooled to the temperature of about 40 degrees K. The skilled in the art may envision the use of many other low- and high-temperature superconductors, all within the scope of this invention.

An inner shield **2** is disposed inside the hollow superconductive shield **1**. The inner shield **2** is comprised of an upper shell **3** and a lower shell **4**, the shells **3** and **4** adjoined with each other. Executed from insulation materials such as foamed ceramics, the inner shield **2** protects the environment within the shield from the electromagnetic field and severe temperatures.

A support structure **5** is disposed between the hollow superconductive shield **1** and the inner shield **2**, concentric to the hollow superconductive shield. The support structure **5** is comprised of an upper rotating element **6** and a lower rotating element **7**.

The upper rotating element **6** is pivotably disposed inside the hollow superconductive shield **1** and may envelope the upper shell **3**. The lower rotating element **7** is pivotably disposed inside the hollow superconductive shield **1** and may envelope the lower shell **4**. Even though the preferred embodiment has two rotating elements, a skilled in the art may envision only one rotating element, or three or more rotation elements, all within the scope of this invention.

Upper means for generating an electromagnetic field **8** are disposed between the hollow superconductive shield **1** and the upper shell **3**. The upper means for generating an electromagnetic field **8** are affixed to the upper rotating element **6** at an electromagnetic field-penetrable distance to the hollow superconductive shield **1**.

Lower means for generating an electromagnetic field **9** are disposed between the hollow superconductive shield **1** and the lower shell **4**. The lower means for generating an electromagnetic field **9** are affixed to the lower rotating element **7** at an electromagnetic field-penetrable distance to the hollow superconductive shield **1**.

The upper means for generating an electromagnetic field **8** and the lower means for generating an electromagnetic field **9** could be solenoid coils or electromagnets. In the process of operation of the space vehicle, the electromagnetic field identified by flux lines **10**, is controllably and variably applied to the hollow superconductive shield **1**.

Electric motors are disposed inside the hollow superconductive shield along its central axis.

A power source **11** is disposed inside the hollow superconductive shield **1** and may be disposed inside the lower shell **4**. The power source **11** is electrically connected with the upper means for generating an electromagnetic field **8**, the lower means for generating an electromagnetic field **9**, and the electric motors. In-turn, the upper means for generating an electromagnetic field **8**, the lower means for generating an electromagnetic field **9**, and the electric motors provide for the rotation of the upper rotating element **6** and the lower rotating element **7**. The power source **11** may be executed as a nuclear power generator.

Life-support equipment **12** is disposed inside the inner shield **2**, and may be disposed inside the lower shell **4**. The life-support equipment **12** may include oxygen, water, and food.

A flux modulation controller **13** is disposed inside the inner shield **2**, and may be disposed inside the upper shell **3**. The flux modulation controller **13** is in communication with the upper means for generating an electromagnetic field **8**, the lower means for generating an electromagnetic field **9**, the power source **11**, and the electric motors.

The flux modulation controller **8** may be executed as a computer or a microprocessor. The flux modulation controller **8** is provided with a capability of modulating the performance parameters of the upper means for generating an electromagnetic field **8**, the lower means for generating an electromagnetic field **9**, the power source **11**, and the electric motors.

A crew **14** may be located inside the upper shell **3** of the inner shield **2** and may consist of one or more astronauts. The crew has a free access to the life-support equipment **12** and the flux modulation controller **8**. A skilled in the art may envision a fully-automated, pilotless craft, which is also within the scope of this invention.

A skilled in the art may also envision the embodiment (not shown), also within the scope of this invention, where the hollow superconductive shield is pivotable, and the support structure with the means for generating an electromagnetic field is affixed on the outside of the inner shield.

FIGS. **2**A & **2**B are diagrams showing the results of the quantized electromagnetic turbulence within the superconductive shell of the hollow superconductive shield provided by the relative rotational motion of the hollow superconductive shield against the upper means for generating an electromagnetic field.

FIG. **2A** shows the clockwise relative rotational motion of the hollow superconductive shield, this motion identified by a clockwise shield motion vector **15**, and the counterclockwise relative rotational motion of upper means for generating an electromagnetic field, this motion identified by a counterclockwise EMF motion vector **16**.

The electromagnetic field, controllably and variably applied by the upper means for generating an electromagnetic field, whose various positions are identified by a wire grid **17**, to the hollow superconductive shield (not shown), causes quantized electromagnetic turbulence within the hollow superconductive shield. This turbulence is represented by a plurality of clockwise quantized vortices of lattice ions **18**. Only one line of the clockwise quantized vortices of lattice ions **18**, out of scale, is shown for illustration purposes only. Each of the clockwise quantized vortices of lattice ions **18** generates a gravitomagnetic field identified by an outward gravitomagnetic field vector **19** directed orthogonally away from the hollow superconductive shield.

FIG. **2B** shows the counterclockwise relative rotational motion of the hollow superconductive shield, this motion identified by a counterclockwise shield motion vector **20**, and the clockwise relative rotational motion of upper means for generating an electromagnetic field, this motion identified by a clockwise EMF motion vector **21**.

The electromagnetic field, controllably and variably applied by the upper means for generating an electromagnetic field identified by the wire grid **17**, to the hollow superconductive shield (not shown), causes quantized electromagnetic turbulence within the hollow superconductive shield, this turbulence represented by a plurality of counterclockwise quantized vortices of lattice ions **22**. Only one line of the counterclockwise quantized vortices of lattice ions **22**, out of scale, is shown for illustration purposes only. Each of the counterclockwise quantized vortices of lattice ions **22** generates a gravitomagnetic field identified by an inward gravitomagnetic field vector **23** directed orthogonally toward the hollow superconductive shield.

The electrical requirements for providing the Li-Torr effect are as follows:

Podkletnov has reported using the high frequency current of 10^5 Hz. He also used 6 solenoid coils @ 850 Gauss each. The reported system's efficiency reached 100% and the total field in the Podkletnov's disk was about 0.5 Tesla. The maximum weight loss reported by Podkletnov was 2.1%.

The preferred embodiment of the device of current invention is capable of housing 2–3 astronauts and therefore is envisioned to be about 5 meters in diameter at the widest point. The preferred space vehicle's acceleration is set at 9.8 m/s/s providing that gravity on board is similar to that on the surface of Earth.

The means for generating an electromagnetic field may be comprised of 124 solenoid coils. At the same 100% efficiency reported by Podkletnov, the total field required providing the acceleration of 9.8 m/s/s is 5,000 Tesla, or about 40 Tesla per coil. Skeggs suggests that on the Podkletnov device, out of 850 Gauss developed on the coil surface, the field affecting the superconductor and causing the gravitomagnetism is only 400 Gauss ("Engineering Analysis of the Podkletnov Gravity Shielding Experiment, Peter L. Skeggs, Quantum Forum, Nov. 7, 1997, http://www.inetarena.com/~noetic/pls/podlev.html, 7 pages). This translates into 47% device efficiency.

In this 47%-efficient space vehicle, the total field required achieving the 9.8 m/s/s acceleration is about 10,600 Tesla, or 85.5 Tesla per each of 124 solenoid coils. It must be noted that at this acceleration rate, it would take nearly a year for the space vehicle to reach the speed of light.

It also must be noted that Skeggs has detected a discrepancy between the Li-Torr estimates and Podkletnov's practical results. If Podkletnov's experimental results are erroneous while the Li-Torr estimates are indeed applicable to the space vehicle of this invention, then the energy requirements for achieving the sought speed would be substantially higher than the above estimate of 10,600 Tesla.

Podkletnov has concluded that, in order for the vacuum pressure density anomaly to take place, the Earth-bound device must be in the condition of Meissner levitation. As are all space bodies, the space vehicle is a subject to the pressure inflationary vacuum state and the gravitational force, which, within the migrating locality of the expanding Universe, in any single linear direction, are substantially in equilibrium. Thus, for the space vehicle, the requirement of Meissner levitation is waved.

The propagation of the gravitomagnetic field identified by the outward gravitomagnetic field vector **19** and the inward gravitomagnetic field vector **23** would cause exotic quantized processes in the vacuum's subatomic particles that include particle polarization, ZPF field defects, and the matter-energy transformation per $E=mc^2$. The combination of these processes would result in the gravitational anomaly. According to the general relativity theory, gravitational attraction is explained as the result of the curvature of spacetime being proportional to the gravitational constant. Thus, the change in the gravitational attraction of the vacuum's subatomic particles would cause a local anomaly in the curvature of the Einsteinean spacetime.

Gravity is the same thing as bent space, propagating with the speed of light characteristic for the particular spacetime curvature. When bent space is affected, there is a change in the speed of propagation of gravity within the spacetime curvature anomaly. The local speed of light, according to Fomalont and Kopeikin always equal to the local speed of propagation of gravity, is also affected within the locality of spacetime curvature anomaly.

Creation of spacetime curvature anomalies adjacent to, or around, the space vehicle, these anomalies characterized by the local gravity and light-speed change, has been the main object of this invention.

FIG. **3A** shows a diagram of a vacuum pressure density anomaly associated with lowered pressure of inflationary vacuum state **24** on the background of Universal curvature of inflationary vacuum state **25**. The vacuum pressure density anomaly associated with lowered pressure of inflationary vacuum state **24** is formed by a multitude of the inward gravitomagnetic field vectors. According to the cosmological constant equation, $\Lambda=(8\pi G/3c^2)\rho_\Lambda$, where the cosmological constant Λ is proportional to the vacuum energy pressure ρ_Λ, G is Newton's constant of gravitation, and c is the speed of light, the curvature of spacetime is proportional to the gravitational constant. According to the general relativity theory, the change in the vacuum pressure density is proportional to the change in the spacetime curvature anomaly. By replacing ρ_Λ with the vacuum pressure density, P times the vacuum energy coefficient κ, and replacing c with Δdistance/Δtime, we derive to $\Lambda=[8\pi G/3(\Delta\text{distance}/\Delta\text{time})^2]P_\kappa$, and could now construct a vacuum pressure density curvature diagram.

The vacuum pressure density curvature anomaly associated with lowered pressure of inflationary vacuum state **24** is shown here as a flattened surface representing the lowered pressure of the inflationary vacuum state. This anomaly is the result of the exotic quantized processes in the subatomic particles caused by the quantized turbulence occurring in the hollow superconductive shield. The XYZ axes represent three dimensions of space and the P axis represents the vacuum pressure density.

FIG. **3**B shows a diagram of a vacuum pressure density anomaly associated with elevated pressure of inflationary vacuum state **26** on the background of the Universal curvature of inflationary vacuum state **25**. The vacuum pressure density anomaly associated with elevated pressure of inflationary vacuum state **26** is formed by a multitude of the outward gravitomagnetic field vectors. The anomaly is shown here as a convex surface representing the elevated pressure of inflationary vacuum state. The diagrams of FIGS. **3**A & **3**B are not to scale: the anomaly sizes are exaggerated for clarity.

FIGS. **4**A & **4**B show diagrams of a spacetime anomaly associated with lowered pressure of inflationary vacuum state **27**, and a spacetime anomaly associated with elevated pressure of inflationary vacuum state **28**, respectively, each on the background a diagram of Universal spacetime **29**.

The quaterized Julia set $Q_{n+1}=Q_n^2+C_0$ is assumed to be an accurate mathematical representation of the Universal spacetime. The generic quaternion Q_0 belongs to the Julia set associated with the quaternion C, and n tends to infinity. If we assume that the quaternion value C_0 is associated with the Universal spacetime **29**, C_1 is the value of quaternion C for the spacetime anomaly associated with lowered pressure of inflationary vacuum state **27**, and C_2 is the value of quaternion C for the spacetime anomaly associated with elevated pressure of inflationary vacuum state **28**, then we can construct two diagrams.

The diagram of FIG. **4**A shows the spacetime anomaly associated with lowered pressure of inflationary vacuum state **27** as a quaterized Julia set contained in a 4-dimensional space: $Q_{n+1}=Q_n^2+C_1$ on the background of the Universal spacetime **29** represented by $Q_{n+1}=Q_n^2+C_0$.

The diagram of FIG. **4**B shows the spacetime anomaly associated with elevated pressure of inflationary vacuum state **28** as a quaterized Julia set $Q_{n+1}=Q_n^2+C_2$, also on the background of the Universal spacetime **29** represented by $Q_{n+1}=Q_n^2+C_0$. On both diagrams, the XYZ axes represent three dimensions of space, and the T axis represents time. The diagrams are not to scale: the anomaly sizes are exaggerated for clarity, and the halves of quaterized Julia sets, conventionally associated with the hypothetical Anti-Universe, are omitted.

FIGS. **5**A, **5**B, **6**, **7**A, & **7**B show simplified diagrams of spacetime curvature anomalies generated by the space vehicle of the current invention, these anomalies providing for the propulsion of the space vehicle. In each case, the pressure anomaly of inflationary vacuum state is comprised of an area of relatively lower vacuum pressure density in front of the space vehicle and an area of relatively higher vacuum pressure density behind the space vehicle. Because the lower pressure of inflationary vacuum state is associated with greater gravity and the higher pressure is associated with the higher repulsive force, the space vehicle is urged to move from the area of relatively higher vacuum pressure density toward the area of relatively lower vacuum pressure density.

FIG. **5**A illustrates the first example of spacetime curvature modification. This example shows a substantially droplet-shaped spacetime curvature anomaly associated with lowered pressure of inflationary vacuum state **30** adjacent to the hollow superconductive shield **1** of the space vehicle. The anomaly **30** is provided by the propagation of a gravitomagnetic field radiating orthogonally away from the front of the hollow superconductive shield **1**. This gravitomagnetic field may be provided by the relative clockwise motion of the upper means for generating an electromagnetic field, and relative counterclockwise motion of the hollow superconductive field, as observed from above the space vehicle.

In this example, the difference between the spacetime curvature within the substantially droplet-shaped spacetime anomaly associated with lowered pressure of inflationary vacuum state, and the ambient spacetime curvature, the spacetime curvature being the same as gravity, results in the gravitational imbalance, with gravity pulling the space vehicle forward.

FIG. **5**B illustrates the second example of spacetime curvature modification. This example shows a substantially droplet-shaped spacetime anomaly associated with elevated pressure of inflationary vacuum state **31** adjacent to the hollow superconductive shield **1** of the space vehicle. The anomaly **31** is provided by the propagation of a gravitomagnetic field radiating orthogonally away from the back of the hollow superconductive shield. This gravitomagnetic field may be provided by the relative counterclockwise motion of the lower means for generating an electromagnetic field, and relative clockwise motion of the hollow superconductive field, as observed from below the space vehicle.

In this example, the difference between the spacetime curvature within the substantially droplet-shaped spacetime anomaly associated with elevated pressure of inflationary vacuum state, and the ambient spacetime curvature, the spacetime curvature being the same as gravity, results in the gravitational imbalance, with the repulsion force pushing the space vehicle forward.

FIG. **6** illustrates the third example of spacetime curvature modification. This example shows the formation of the substantially droplet-shaped spacetime anomaly associated with lowered pressure of inflationary vacuum state **30** combined with the substantially droplet-shaped spacetime anomaly associated with elevated pressure of inflationary vacuum state **31**. This combination of anomalies may be provided by the relative clockwise motion of the upper means for generating an electromagnetic field and relative clockwise motion of the hollow superconductive field, combined with the relative clockwise motion of the lower means for generating an electromagnetic field, as observed from above the space vehicle.

In this example, the difference between the spacetime curvature within the substantially droplet-shaped spacetime anomaly associated with lowered pressure of inflationary vacuum state, and the spacetime curvature of the substantially droplet-shaped spacetime anomaly associated with elevated pressure of inflationary vacuum state, the spacetime curvature being the same as gravity, results in the gravitational imbalance, with gravity pulling, and the repulsion force pushing, the space vehicle forward.

FIG. **7**A illustrates the fourth example of spacetime curvature modification. This example shows the formation of a substantially egg-shaped spacetime anomaly associated with lowered pressure of inflationary vacuum state **32** around the hollow superconductive shield **1** of the space vehicle. The anomaly **32** is provided by the propagation of gravitomagnetic field of unequally-distributed density, this gravitomagnetic field radiating in all directions orthogonally away from the hollow superconductive shield. The propagation of the unequally-distributed gravitomagnetic field leads to the similarly unequally-distributed spacetime curvature anomaly. This unequally-distributed gravitomagnetic field may be provided by the relatively faster clockwise motion of the upper means for generating an electromagnetic field relative to the hollow superconductive field, combined with the relatively slower counterclockwise motion of the lower means for generating an electromagnetic field, as observed from above the space vehicle.

An area of the lowest vacuum pressure density **33** of the substantially egg-shaped spacetime anomaly associated with lowered pressure of inflationary vacuum state **32** is located directly in front of the space vehicle.

In this example, the variation in the spacetime curvature within the substantially egg-shaped spacetime anomaly associated with lowered pressure of inflationary vacuum state, the spacetime curvature being the same as gravity, results in a gravitational imbalance, with gravity pulling the space vehicle forward in modified spacetime.

FIG. **7**B illustrates the fifth example of spacetime curvature modification, also with the purpose of providing for a propulsion in modified spacetime. This example shows the formation of a substantially egg-shaped spacetime anomaly associated with elevated pressure of inflationary vacuum state **34** around the hollow superconductive shield **1** of the space vehicle. The anomaly **34** is provided by the propagation of gravitomagnetic field of unequally-distributed density, this gravitomagnetic field radiating in all directions orthogonally away from the hollow superconductive shield. The propagation of the unequally-distributed gravitomagnetic field leads to the similarly unequally-distributed spacetime curvature anomaly. This unequally-distributed gravitomagnetic field may be provided by the relatively slower counterclockwise motion of the upper means for generating an electromagnetic field relative to the hollow superconductive field, combined with the relatively faster clockwise motion of the lower means for generating an electromagnetic field, as observed from above the space vehicle.

An area of the highest vacuum pressure density **35** of the substantially egg-shaped spacetime anomaly associated with elevated pressure of inflationary vacuum state **34** is located directly behind the space vehicle.

In this example, the variation in the spacetime curvature within the substantially egg-shaped spacetime anomaly associated with elevated pressure of inflationary vacuum state, the spacetime curvature being same as gravity, results in a gravitational imbalance, with the repulsion force pushing the space vehicle forward in modified spacetime at speeds approaching the light-speed characteristic for this modified area. This light-speed might be much higher than the light-speed in the ambient space.

By creating alternative anomalies and modulating their parameters, the space vehicle's crew would dilate and contract time and space on demand. The space vehicle, emitting a vacuum pressure modifying, controllably-modulated gravitomagnetic field in all directions, would rapidly move in the uneven spacetime anomaly it created, pulled forward by gravity or pushed by the repulsion force. The time rate zone of the anomaly is expected to have multiple quantized boundaries rather than a single sudden boundary affecting space and time in the immediate proximity of the vehicle. Speed, rate of time, and direction in space could be shifted on demand and in a rapid manner. The modulated light-speed could make the space vehicle suitable for interstellar travel. Because of the time rate control in the newly created isospace, the accelerations would be gradual and the angles of deviation would be relatively smooth. The gravity shielding would further protect pilots from the ill-effects of gravity during rapid accelerations, directional changes, and sudden stops.

What is claimed is:

1. A space vehicle propelled by the pressure of inflationary vacuum state is claimed comprising:

a hollow superconductive shield,

an inner shield, the inner shield disposed inside said hollow superconductive shield, said inner shield comprising an upper shell and a lower shell,

a support structure, the support structure disposed between said hollow superconductive shield and said inner shield concentrically to said hollow superconductive shield, said support structure comprised of an upper rotating element and a lower rotating element,

upper means for generating an electromagnetic field, the upper means for generating an electromagnetic field disposed between said hollow superconductive shield and said upper shell, affixed to said upper rotating element at an electromagnetic field-penetrable distance to said hollow superconductive shield,

lower means for generating an electromagnetic field, the lower means for generating an electromagnetic field disposed between said hollow superconductive shield and said lower shell, affixed to said lower rotating element at an electromagnetic field-penetrable distance to said hollow superconductive shield,

electric motors, the electric motors disposed inside said hollow superconductive shield along the central axis of said hollow superconductive shield,

a power source, the power source disposed inside said hollow superconductive shield, said power source electrically connected with said upper means for generating an electromagnetic field, said lower means for generating an electromagnetic field, and said electric motors,

life-support equipment, the life-support equipment disposed inside said inner shield,

a flux modulation controller, the flux modulation controller disposed inside said inner shield, said flux modulation controller in communication with said upper means for generating an electromagnetic field, said lower means for generating an electromagnetic field, said power source, and said electric motors, and

a crew, the crew disposed inside said inner shield accessibly to said life-support and said flux modulation controller.

2. The space vehicle of claim **1** wherein said hollow superconductive shield is composed of a high-temperature ceramic superconductor with a composite crystal structure, and is cooled to a temperature of about 40 degrees K.

3. The space vehicle of claim **1** wherein said inner shield is composed of insulating materials such as foamed ceramics.

4. The space vehicle of claim **1** wherein said power source is a nuclear generator.

5. The space vehicle of claim **1** wherein said flux modulation controller is a computer.

6. The space vehicle of claim **1** wherein said upper rotating element and said lower rotating element are pivoted, urged to rotate by said upper means for generating an electromagnetic field, said lower means for generating an electromagnetic field, and said electric motors.

7. The space vehicle of claim **1** wherein said hollow superconductive shield is pivoted and urged to rotate by said upper means for generating an electromagnetic field, said lower means for generating an electromagnetic field, and said electric motors.

8. A method for generating pressure anomaly of inflationary vacuum state is claimed comprising the following steps, some occurring simultaneously:

providing a space vehicle comprising a hollow superconductive shield, an inner shield, a support structure, upper means for generating an electromagnetic field, lower means for generating an electromagnetic field, electric motors, a power source, life-support equipment, a flux modulation controller, and a crew,

generating electrical energy in said power source,

converting said electrical energy into an electromagnetic field in said upper means for generating an electromagnetic field and said lower means for generating an electromagnetic field,

converting said electrical energy into mechanical energy in said electrical motors,

rotating said upper means for generating an electromagnetic field and said lower means for generating an electromagnetic field relative to said hollow superconductive shield,

converting said electromagnetic field into a gravitomagnetic field in said hollow superconductive shield,

propagating said gravitomagnetic field outward, orthogonal to the walls of said hollow superconductive shield, and

forming a pressure anomaly of inflationary vacuum state in the area of propagation of said gravitomagnetic field, the pressure anomaly of inflationary vacuum state comprised of an area of relatively lower vacuum pressure density in front of said space vehicle and an area of relatively higher vacuum pressure density behind said vehicle,

whereby said space vehicle is urged to move from said area of relatively higher vacuum pressure density toward said area of relatively lower vacuum pressure density.

9. The method of claim **8** wherein said pressure anomaly of inflationary vacuum state is a substantially droplet-shaped anomaly associated with lowered pressure of inflationary vacuum state provided by the propagation of a gravitomagnetic field radiating orthogonally away from the front of said hollow superconductive shield,

whereby providing for the gravitational imbalance where the lowered pressure of inflationary vacuum state is pulling said space vehicle forward.

10. The method of claim **8** wherein said pressure anomaly of inflationary vacuum state is a substantially droplet-shaped anomaly associated with elevated pressure of inflationary vacuum state provided by the propagation of a gravitomagnetic field radiating orthogonally away from the back of said hollow superconductive shield,

whereby providing for the gravitational imbalance where the elevated pressure of inflationary vacuum state is pushing said space vehicle forward.

11. The method of claim **8** wherein said pressure anomaly of inflationary vacuum state is comprised of a substantially droplet-shaped anomaly associated with lowered pressure of inflationary vacuum state and a substantially droplet-shaped anomaly associated with elevated pressure of inflationary vacuum state,

whereby providing for the gravitational imbalance where the lowered pressure of inflationary vacuum state is pulling, and the elevated pressure of inflationary vacuum state is pushing, said space vehicle forward.

12. The method of claim **8** wherein said pressure anomaly of inflationary vacuum state is a substantially egg-shaped anomaly associated with lowered pressure of inflationary vacuum state provided by the unequally-distributed gravitomagnetic field propagated in all directions orthogonally away from said hollow superconductive shield with an area of lowest vacuum pressure density disposed directly in front said space vehicle,

whereby providing for the gravitational imbalance such that the lowered pressure of inflationary vacuum state is pulling said space vehicle forward in modified space-time.

13. The method of claim **8** wherein said pressure anomaly of inflationary vacuum state is a substantially egg-shaped anomaly associated with elevated pressure of inflationary vacuum state provided by the unequally-distributed gravitomagnetic field propagated in all directions orthogonally away from said hollow superconductive shield with an area of highest vacuum pressure density disposed directly behind said space vehicle,

whereby providing for the gravitational imbalance such that the highest pressure of inflationary vacuum state is pushing said space vehicle forward in modified space-time at a speed possibly approaching a local light-speed, the local light-speed which may be substantially higher than the light-speed in the ambient space.

* * * * *

MIND CONTROL
TECHNOLOGY
PATENTS

The following pages contain facsimile reproductions of patents that have been granted by the Patent Office of the United States of America for the purposes of mental monitoring and mind alteration:

- *Hearing Device*—United States Patent 4,858,612.
- *Apparatus and Method for remotely Monitoring and Altering Brain Waves*
  —United States Patent 3,951,134.
- *Biomagnetic Analytical System Using Fiber-Optic Magnetic Sensors*
  —United States Patent 4,951,674.
- *Brain Electrical Activity Mapping*—United States Patent 4,408,616.
- *Method of Inducing Mental, Emotional and Physical States of Consciousness, Including Specific Mental Activity, in Human Beings*
  —United States Patent 5,213,562.
- *Method of and Apparatus for Inducing Desired States of Consciousness*
  —United States Patent 5,356,368.

United States Patent No. 4,858,612—*Hearing Device*:

> ABSTRACT
>
> A method and apparatus for simulation of hearing in mammals by introduction of a plurality of microwaves into the region of the auditory cortex... A microphone is used to transform sound signals into electrical signals which are in turn analyzed and processed to provide controls for generating a plurality of microwave signals at different frequencies. The multifrequency microwaves are then applied to the brain in the region of the auditory cortex. By this method sounds are perceived by the mammal which are representative of the original sound perceived by the microphone.

You will not find a better description of the means by which voices can be electronically transmitted into a human mind. Although the text of the patent presents the invention as a therapeutic aid for individuals who have hearing impairments, the potential mind control applications of the device are obvious.

But why stop at the mere electronic transmission of voices into the human brain? Why not electronically alter the brain waves themselves by means of radio transmissions? Consider, for example, United States Patent No. 3,951,134 – *Apparatus and Method for Remotely Monitoring and Altering Brain Waves*. The abstract says:

> ABSTRACT
> Apparatus for and method of sensing brain waves at a position remote from a subject whereby electromagnetic signals of different frequencies are simultaneously transmitted to the brain of the subject in which the signals interfere with one another to yield a waveform which is modulated by the subject's brain waves. The interference waveform which is representative of the brain wave activity is re-transmitted by the brain to a receiver where it is demodulated and amplified. The demodulated waveform is then displayed for visual viewing and routed to a computer for further processing and analysis. The demodulated waveform also can be used to produce a compensating signal which is transmitted back to the brain to effect a desired change in electrical activity therein.

Quoting again from the text of the patent:

> SUMMARY OF THE INVENTION
> The present invention relates to apparatus and a method for monitoring brain waves wherein all components of the apparatus employed are remote from the test subject. More specifically, high frequency transmitters are operated to radiate electromagnetic energy of different frequencies through antennas which are capable of scanning the entire brain of the test subject or any desired region thereof. The signals of different frequencies penetrate the skull of the subject and impinge upon the brain where they mix to yield an interference wave modulated by radiations from the brain's natural electrical activity. The modulated interference wave is re-transmitted by the brain and received by an antenna at a remote station where it is demodulated, and processed to provide a profile of the subject's brain waves. In addition to passively monitoring his brain waves, the subject's neurological processes may be affected by transmitting to his brain, through a transmitter, compensating signals. The latter signals can be derived from the received and processed brain waves.

The import of the invention could not be clearer. This patent spells out, in black and white, an electronic method and apparatus for remotely monitoring and altering the brain waves of human beings, by use of radio transmitters and computers. The subject's brain waves are electronically monitored, received, routed to a computer, modified, and then transmitted back to his/her brain, with the stated intention of altering the "electrical activity" of the subject's brain. As the patent states:

> ... the subject's neurological processes may be affected by transmitting to his brain, through a transmitter, compensating signals. The latter signals can be derived from the received and processed brain waves.

The potential for electronic mind control is transparent.

Thanks to modern electronic technology the electrical activity of the human brain can be precisely monitored and displayed. The subtle electromagnetic fields and accompanying electrical activity in the brain can be electronically monitored and recorded in real time. United States Patent No. 4,951,674—*Biomagnetic Analytical System Using Fiber-Optic Magnetic Sensors* describes one such method and apparatus:

> ABSTRACT
> A biomagnetic analytical system for sensing and indicating minute magnetic fields emanating from the brain or from any other tissue region of interest in a subject under study. The system includes a magnetic pickup device constituted by an array of fiber-optic magnetic sensors mounted at positions distributed throughout the inner confines of a magnetic shield configured to conform generally to the head of the subject or what-

> ever other body region is of interest. Each sensor yields a light beam whose phase or other parameter is modulated in accordance with the magnetic field emanating from the related site in the region. The modulated beam from each sensor is compared in an interferometer with a reference light beam to yield an output signal that is a function of the magnetic field being emitted at the related site. The output signals from the interferometer are processed to provide a display or recording exhibiting the pattern or map of magnetic fields resulting from emanations at the multitude of sites encompassed by the region.

This device makes possible the mapping of the brain's biomagnetic activity. The data can be displayed visually, on a CRT monitor (as on a computer screen), permitting a specialist to electronically peer into a person's head and to:

> ..."see" a functional image of the brain on a CRT, ... the image will be a profile of the electromganetically "active" portions of the brain, as shown by the magnetic pattern derived from data reduction...

The data can also be digitized for subsequent storage and manipulation. In the words of the patent:

> The data can also be directed to a storage medium for the purpose of recording the digitized biomagnetic data for archiving and later retrieval and processing.

Patent No. 4,951,674 goes on to say that during the data processing stage, data from:

> ... modalities such as EEG, EKG, MRI and X-ray ... can be combined with the biomagnetic data ...

In other words, the practitioner can reconcile the biomagnetic data with an individual's EEG pattern. This means a specific pattern of biomagnetic activity can be correlated with a particular pattern of electrical activity in the brain. (See the appendix at the end of the book for the complete patent. Look at illustrations 1 to 3. Note the special helmet with the magnetic sensors that fits around the head of the person whose brain's biomagnetic activity is being monitored and digitally recorded.)

United States Patent No. 4,408,616, *Brain Electrical Activity Mapping*, provides further information on how the human brain can be electronically monitored.

The patent abstract says, in part:

> Topographic displays of brain electrical activity are produced from matrices of data derived from evoked potential (EP) and steady-state responses of skull transducers ... the rate of data sampling is sufficient to capture rapid transient events ...

In other words, the transition from one mental state to another can be observed and electronically mapped. The text of the patent explains how this is done:

> Twenty electrodes (e.g., Grass gold cup) are attached to subject's skull ... Twenty leads from electrodes are connected through switch to conventional 24-channel polygraph ...

The patent succinctly explains how electrodes are attached to a person's skull and connected to a polygraph machine to record the electrical activity of their brain in real time, as it responds to various stimuli. It neatly sums up:

> ... (T)he brain electrical activity mapping system creates color topographic displays reflecting brain electrical activity using, as input, continuous electrical waveforms recorded from a number of points on the skull.

The data are then converted to digital form. Data are stored in individual files or combined with others into "group" files, so that a group profile of brain electrical activity in response to a variety of stimuli can be constructed. The stimulus may be a flashing light or something more complex, like being asked to distinguish between two similar, but slightly different, spoken words.

Over time, a library of digitized data (data that can be stored and analyzed by computer) about the electrical activity of the brains of individual people and groups of people is built up. This is precisely the type of data that can be used to modulate a radio wave that is transmitted into a target subject's brain to alter his/her brainwaves. In principle, by using averaged data derived from group files of brain electrical activity, it should be possible to modulate radio broadcasts to affect the brain electrical activity of large numbers of a target population.

In many cases people's minds and brains have been tampered with and altered without their informed consent. Psychoactive drugs, microwaves, hypnosis, brain washing and electronic implants are some of the ways in which mind control victims have been attacked by these shadowy agencies.

Electronic and psychotronic mind control technologies are absolutely antithetical to the fundamental exercise of human freedom. Unfortunately, the patents and technologies described here are real. It is time for us all to wake up and see things as they are. A particularly insidious slavery is mental enslavement, accomplished by electronically or psychotronically altering people's thoughts, such that their very thoughts are not their own, but are ones that their self-appointed masters would have them to think.

United States Patent [19]
Stocklin

[11] Patent Number: 4,858,612
[45] Date of Patent: Aug. 22, 1989

[54] HEARING DEVICE

[76] Inventor: Philip L. Stocklin, P.O. Box 2111, Satellite Beach, Fla. 32937

[21] Appl. No.: 562,742

[22] Filed: Dec. 19, 1983

[51] **Int. Cl.**⁴ **A61N 1/36**
[52] **U.S. Cl.** .... **128/422; 178/419 S**
[58] **Field of Search** .... 128/419 R, 419 S, 422, 128/653, 771, 732, 741, 746, 791, 804; 340/407

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,490,458	1/1970	Allison	128/421
3,751,605	8/1973	Michelson	128/1 R
3,951,134	4/1976	Malech	128/131
4,428,377	1/1984	Zollner et al.	128/419 R

FOREIGN PATENT DOCUMENTS

893311	2/1972	Canada	128/422
2811120	9/1978	Fed. Rep. of Germany	128/419 R
591196	1/1978	U.S.S.R.	128/419 R

OTHER PUBLICATIONS

Gerkin, G., "Electroencephalography & Clinical Neurophysiology", vol. 135, No. 6, Dec. 1973, pp. 652-653.
Frye et al., "Science", vol. 181, Jul. 27, 1973, pp. 356-358.
Bise, William, "Low Power Radio-Frequency and Microwave Effects on Human Electroencephalogram and Behavior", Physiol. Chem. & Physics 10 (1978).

Primary Examiner—William E. Kamm
Attorney, Agent, or Firm—Wegner & Bretschneider

[57] **ABSTRACT**

A method and apparatus for simulation of hearing in mammals by introduction of a plurality of microwaves into the region of the auditory cortex is shown and described. A microphone is used to transform sound signals into electrical signals which are in turn analyzed and processed to provide controls for generating a plurality of microwave signals at different frequencies. The multifrequency microwaves are then applied to the brain in the region of the auditory cortex. By this method sounds are perceived by the mammal which are representative of the original sound received by the microphone.

29 Claims, 7 Drawing Sheets

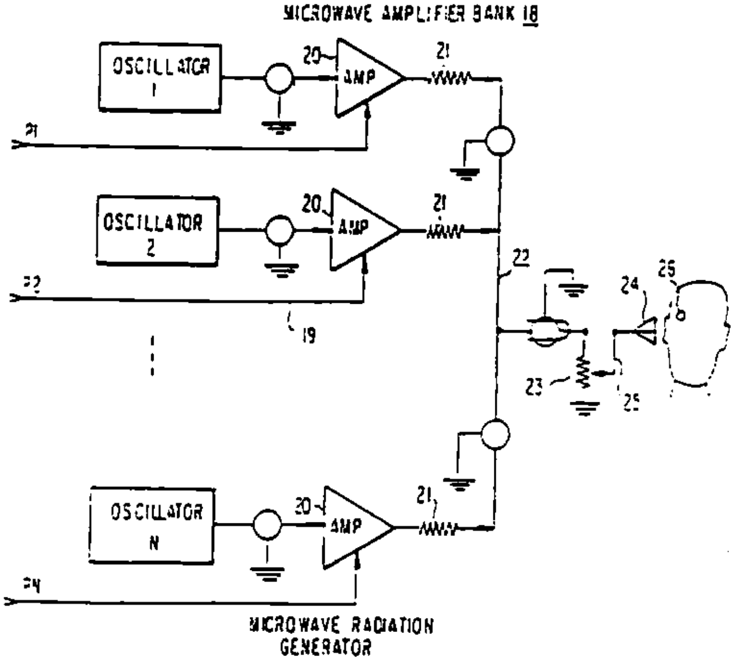

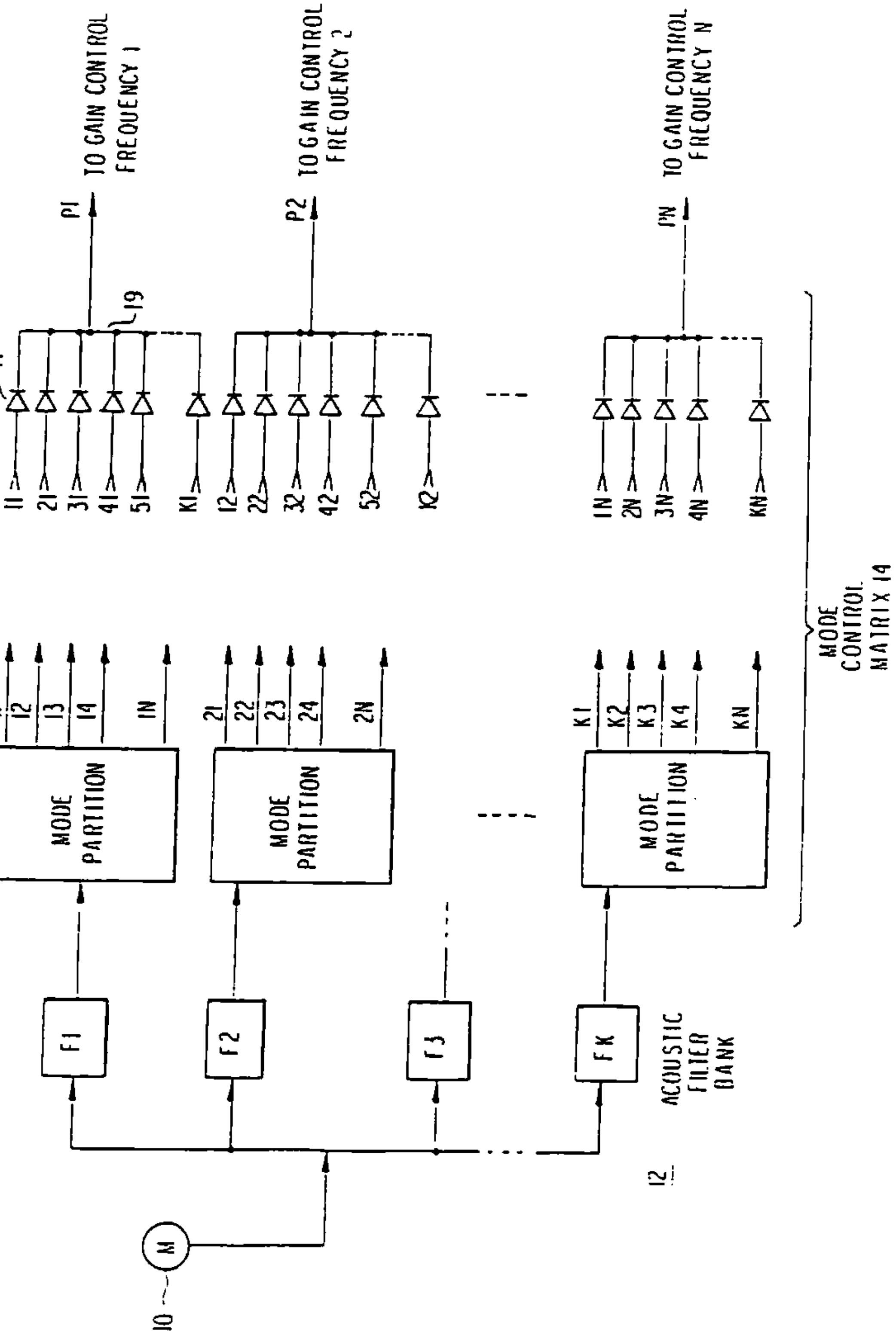
10
M
ACOUSTIC FILTER BANK
12
F1
F2
F3
FK
MODE PARTITION
MODE PARTITION
MODE PARTITION
11
12
13
14
1N
21
22
23
24
2N
K1
K2
K3
K4
KN
MODE CONTROL MATRIX 14
MODE SUMMATION
16
17
19
11
21
31
41
51
K1
12
22
32
42
52
K2
1N
2N
3N
4N
KN
P1
P2
PN
TO GAIN CONTROL FREQUENCY 1
TO GAIN CONTROL FREQUENCY 2
TO GAIN CONTROL FREQUENCY N

FIG 1

FIG. 2

MICROWAVE AMPLIFIER BANK 18

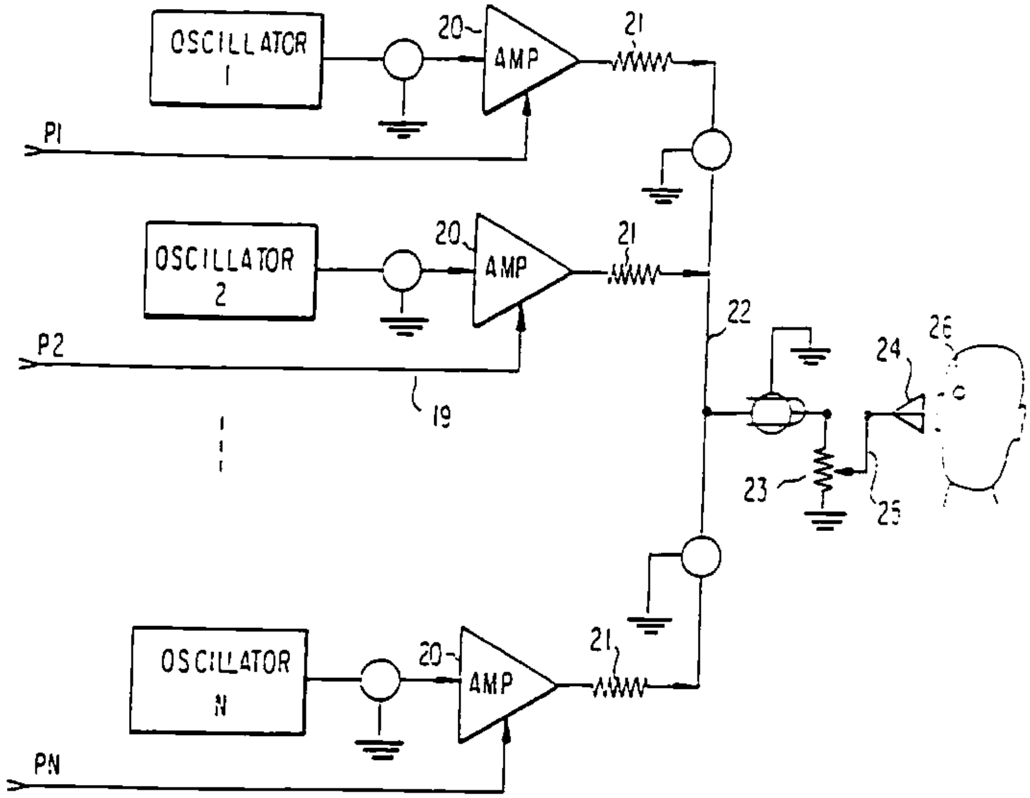
OSCILLATOR 1
20
AMP
21
P1
OSCILLATOR 2
20
AMP
21
22
P2
19
26
24
23
25
OSCILLATOR N
20
AMP
21
PN

MICROWAVE RADIATION GENERATOR

FIG 2a

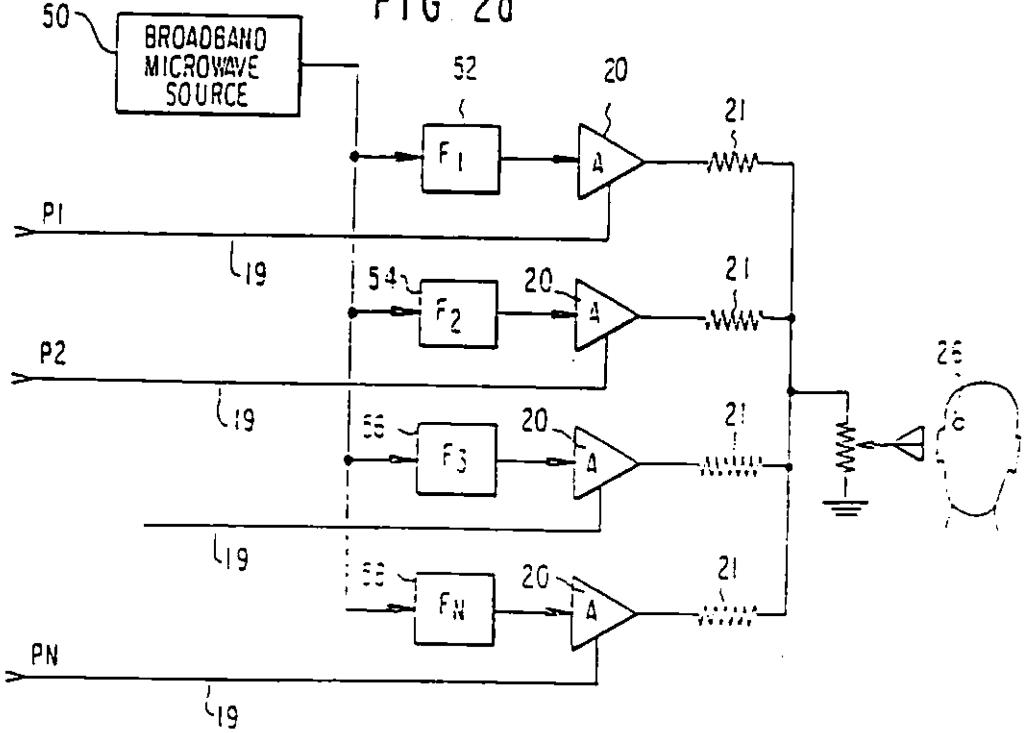
50
BROADBAND MICROWAVE SOURCE
52
20
F1
A
21
P1
19
54
F2
20
A
21
P2
19
56
F3
20
A
21
26
19
58
FN
20
A
21
PN
19

FIG. 5

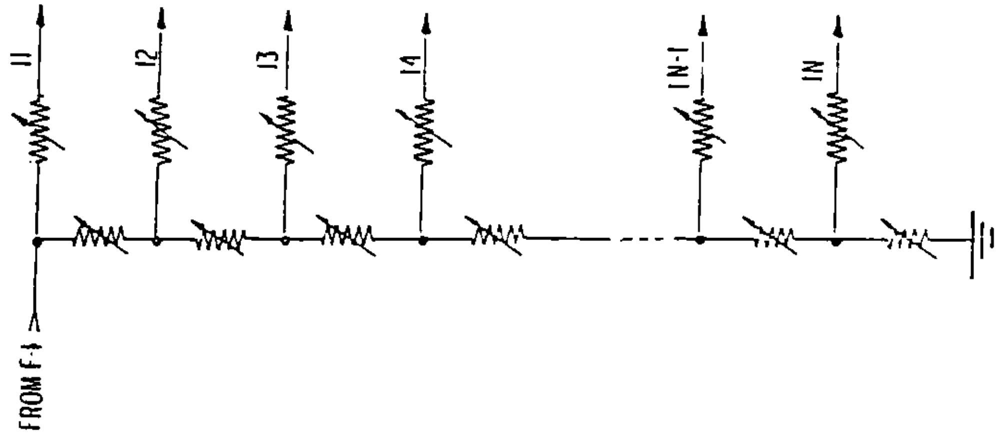
FROM F-1
I1
I2
I3
I4
IN-1
IN

FIG. 4

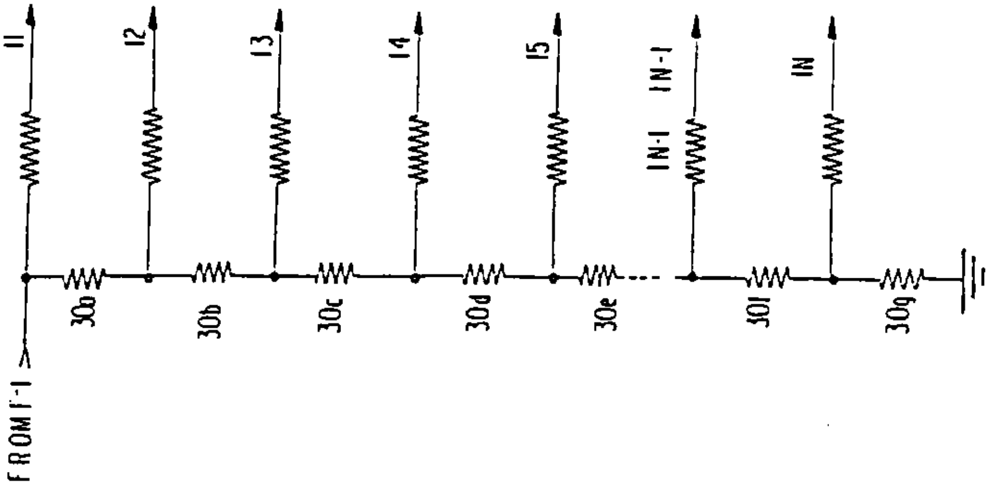
FROM F-1
I1
I2
I3
I4
I5
IN-1 IN-1
IN-1
IN
30a
30b
30c
30d
30e
30f
30g

FIG. 3

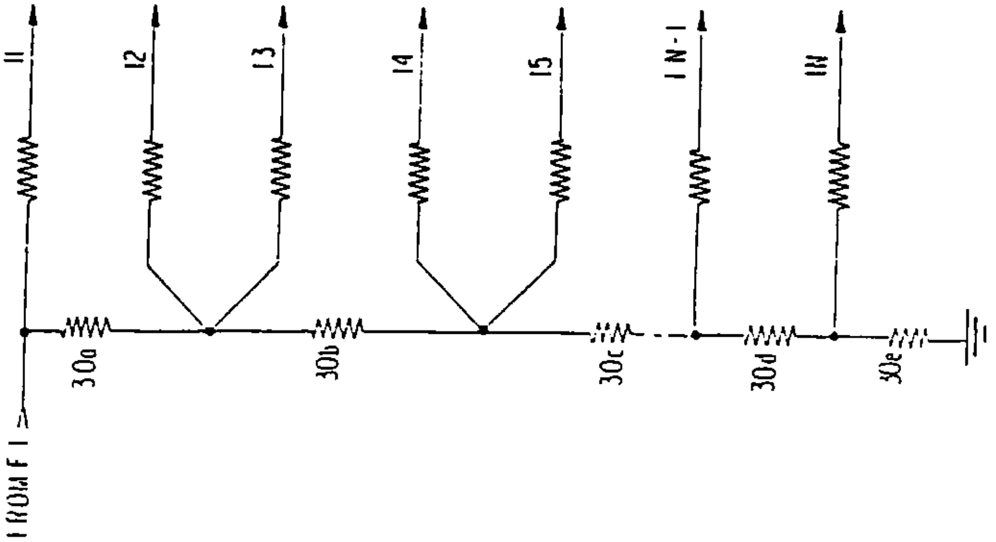
FROM F-1
I1
I2
I3
I4
I5
IN-1
IN
30a
30b
30c
30d
30e

8

FIG 6

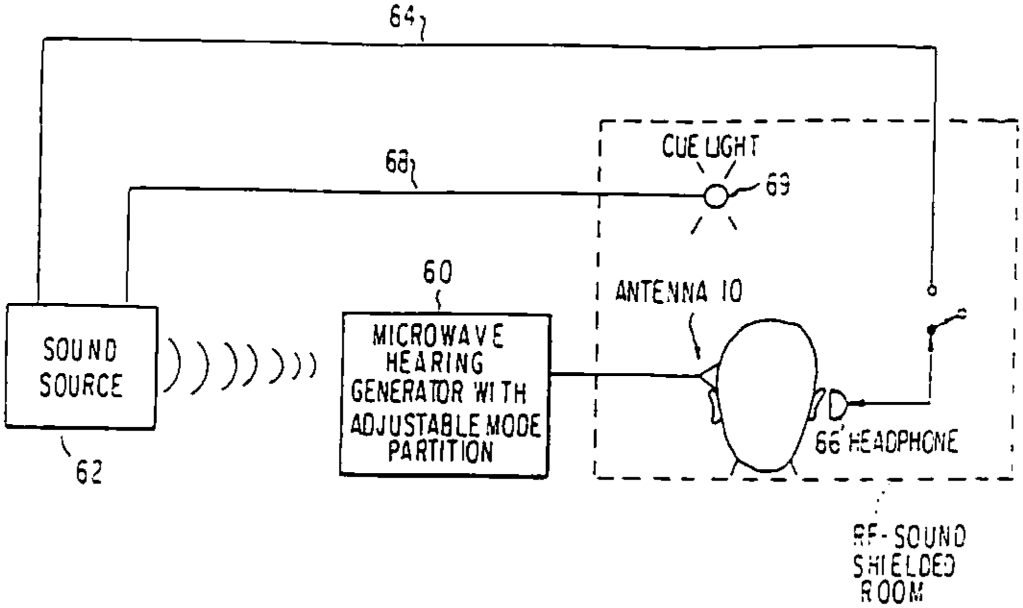
64
68
CUE LIGHT
69
60
ANTENNA 10
SOUND SOURCE
MICROWAVE HEARING GENERATOR WITH ADJUSTABLE MODE PARTITION
62
66 HEADPHONE
RF-SOUND SHIELDED ROOM

FIG 7

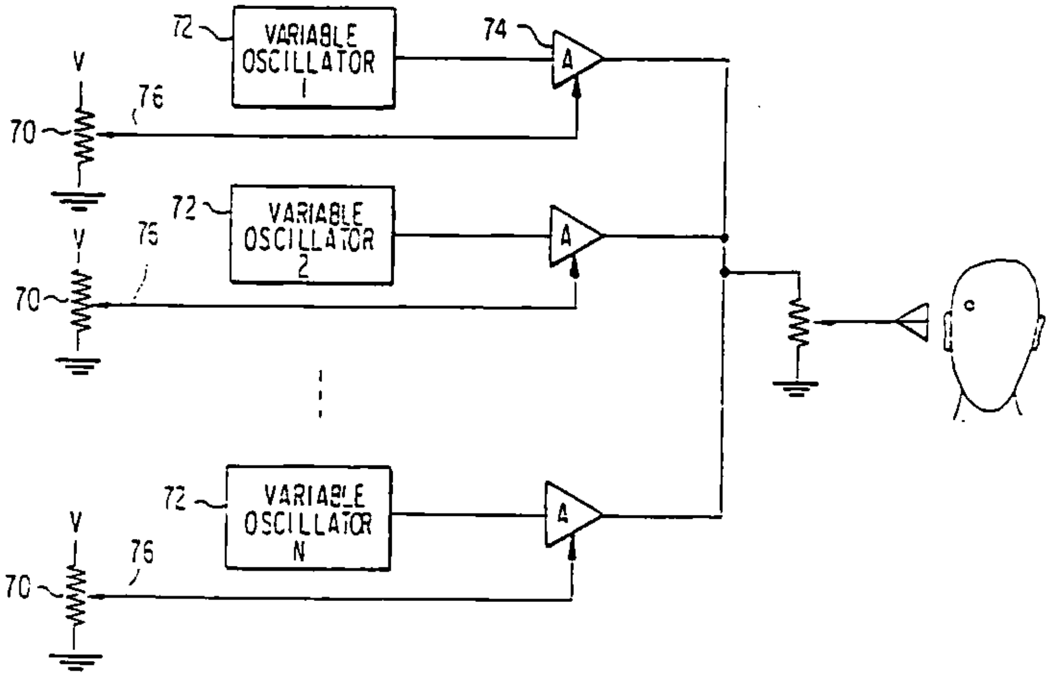
72
VARIABLE OSCILLATOR 1
74
A
V
76
70
72
VARIABLE OSCILLATOR 2
A
V
76
70
72
VARIABLE OSCILLATOR N
A
V
76
70

9

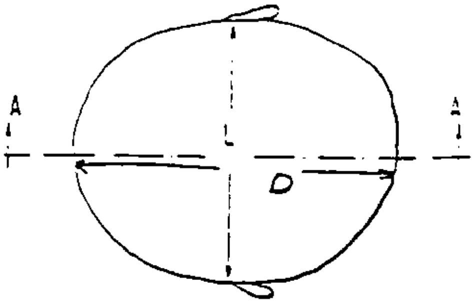

FIG 8

FIG 9

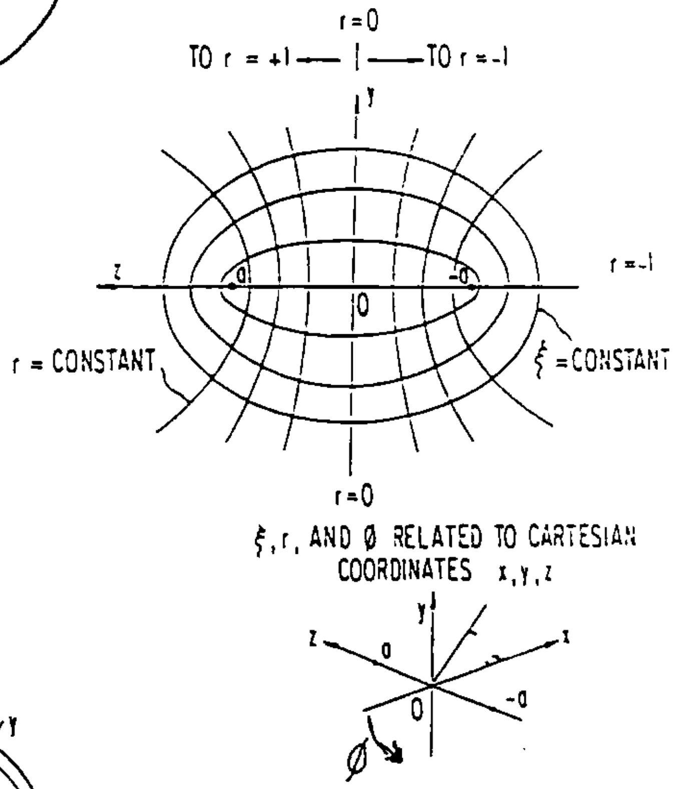
r=0
TO r = +1
TO r = -1
y
z
a
-a
0
r = -1
r = CONSTANT
ξ = CONSTANT
r=0
ξ, r, AND Ø RELATED TO CARTESIAN COORDINATES x, y, z
y
z
a
x
0
-a
Ø

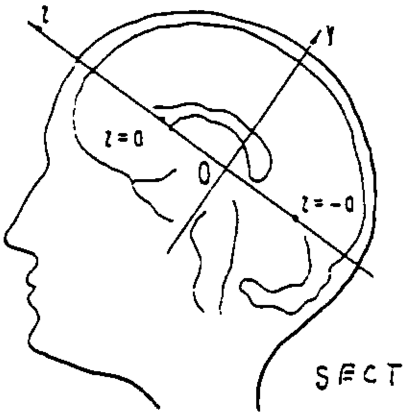

FIG 10

SECTION A-A

TRANSFORMATION EQUATIONS

$x = a(\xi^2 - 1)^{1/2} (1 - r^2)^{1/2} \cos Ø$ $\quad 1 \leq \xi \leq$

$y = -a(\xi^2 - 1)^{1/2} (1 - r^2)^{1/2} \sin Ø$ $\quad -1 \leq Ø \leq +1$

$z = a \xi r$ $\quad 0 \leq Ø \leq 2\pi$

10

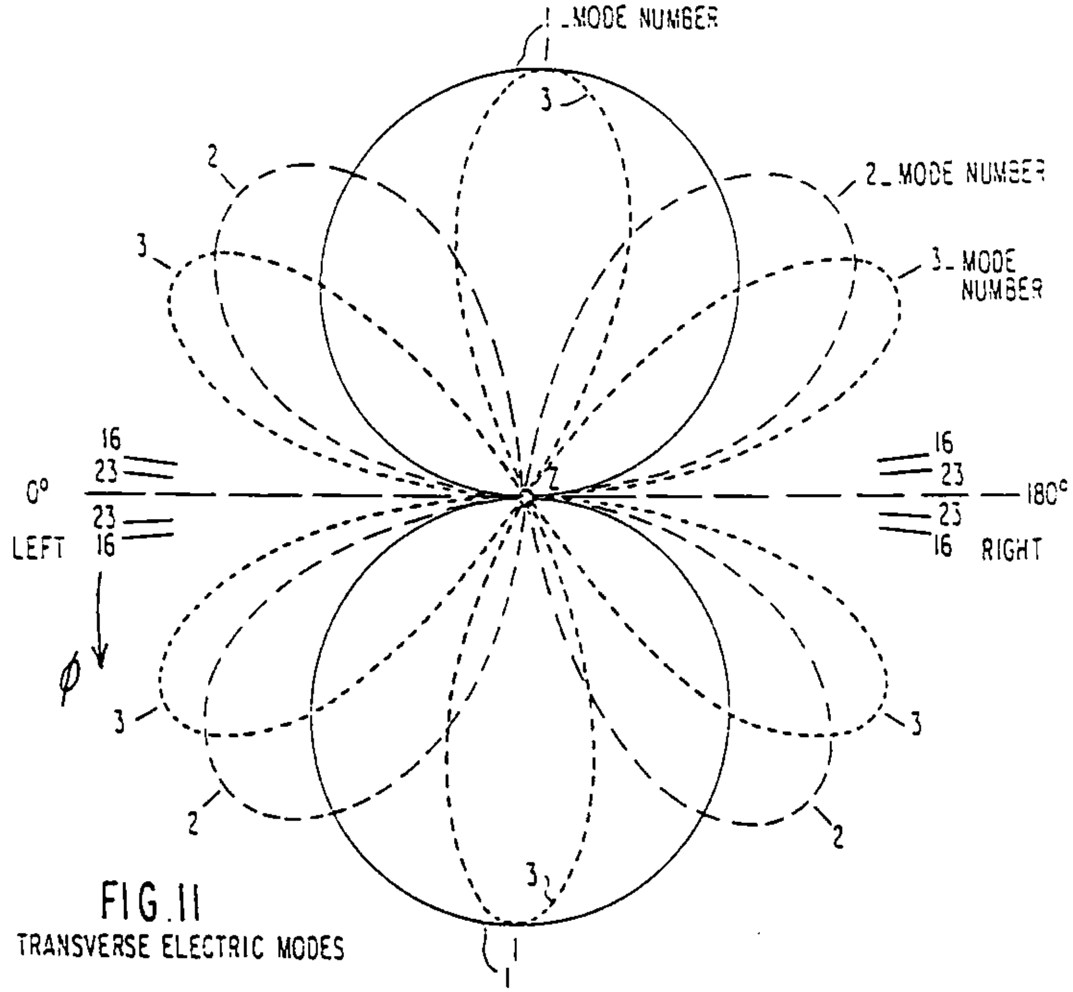
1_MODE NUMBER
3
2
2_MODE NUMBER
3
3_MODE NUMBER
16
23
0°
23
LEFT 16
Z
16
23
180°
23
16 RIGHT
ϕ
3
3
2
2
3
1
1

FIG. 11
TRANSVERSE ELECTRIC MODES

FIG. 12
PRIMARY AUDITORY CORTEX

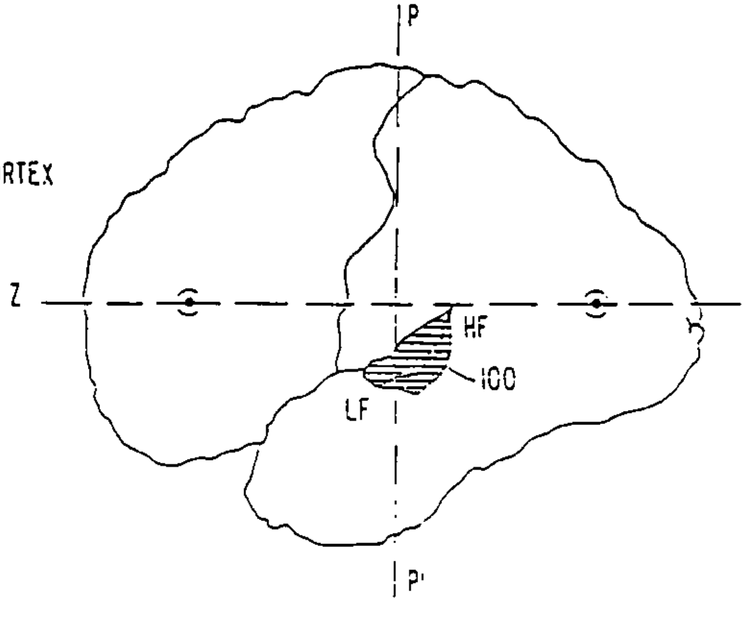
P
Z
HF
100
LF
P'

1(

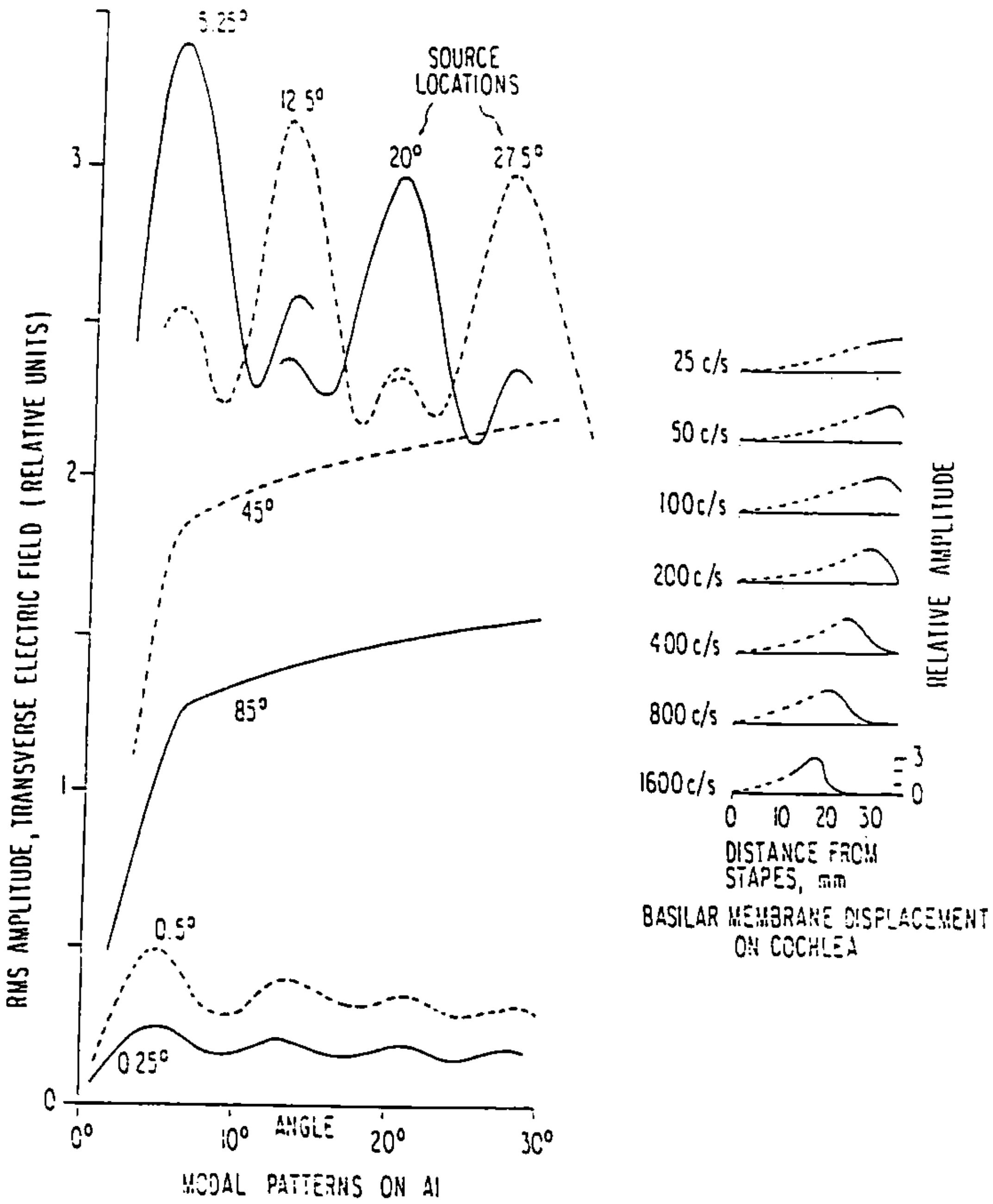
5.25°
SOURCE LOCATIONS
12.5°
20°
27.5°
45°
85°
0.5°
0.25°
3
2
1
0
RMS AMPLITUDE, TRANSVERSE ELECTRIC FIELD (RELATIVE UNITS)
0°
10°
ANGLE
20°
30°
MODAL PATTERNS ON AI
25 c/s
50 c/s
100c/s
200c/s
400 c/s
800 c/s
1600c/s
RELATIVE AMPLITUDE
3
0
0
10
20
30
DISTANCE FROM STAPES, mm
BASILAR MEMBRANE DISPLACEMENT ON COCHLEA

FIG 13

12

HEARING DEVICE

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates to devices for aiding of hearing in mammals. The invention is based upon the perception of sounds which is experienced in the brain when the brain is subjected to certain microwave radiation signals.

2. Description of the Prior Art

In prior art hearing devices for human beings, it is well known to amplify sounds to be heard and to apply the amplified sound signal to the ear of the person wearing the hearing aid. Hearing devices of this type are however limited to hearing disfunctions where there is no damage to the auditory nerve or to the auditory cortex. In the prior art, if there is damage to the auditory cortex or the auditory nerve, it cannot be corrected by the use of a hearing aid.

During World War II, individuals in the radiation path of certain radar installations observed clicks and buzzing sounds in response to the microwave radiation. It was through this early observation that it became known to the art that microwaves could cause a direct perception of sound within a human brain. These buzzing or clicking sounds however were not meaningful, and were not perception of sounds which could otherwise be heard by the receiver. This type of microwave radiation was not representative of any intelligible sound to be perceived. In such radar installations, there was never a sound which was generated which resulted in subsequent generation of microwave signals representative of that sound.

Since the early perception of buzzing and clicking, further research has been conducted into the microwave reaction of the brain. In an article entitled "Possible Microwave Mechanisms of the Mammalian Nervous System" by Philip L. Stocklin and Brain F. Stocklin, published in the TIT Journal of Life Sciences, Tower International Technomedical Institute, Inc. P.O. Box 4594, Philadelphia, Pa. (1979) there is disclosed a hypothesis that the mammalian brain generates and uses electro magnetic waves in the lower microwave frequency region as an integral part of the functioning of the central and peripheral nervous systems. This analysis is based primarily upon the potential energy of a protein integral in the neural membrane.

In an article by W. Bise entitled "Low Power Radio-Frequency and Microwave Effects On Human Electroencephalogram and Behavior", Physiol. Chemistry Phys. 10, 387 (1978), it is reported that there are significant effects upon the alert human EEG during radiation by low intensity CW microwave electromagnetic energy. Bise observed significant repeatable EEG effects for a subject during radiation at specific microwave frequencies.

SUMMARY OF THE INVENTION

Results of theoretical analysis of the physics of brain tissue and the brain/skull cavity, combined with experimentally-determined electromagnetic properties of mammalian brain tissue, indicate the physical necessity for the existence of electromagnetic standing waves, called modes in the living mammalian brain. The mode characteristics may be determined by two geometric properties of the brain: these are the cephalic index of the brain (its shape in prolate spheroidal coordinates) and the semifocal distance of the brain (a measure of its size). It was concluded that estimation of brain cephalic index and semifocal distance using external skull measurements on subjects permits estimation of the subject's characteristic mode frequencies, which in turn will permit a mode by mode treatment of the data to simulate hearing.

This invention provides for sound perception by individuals who have impaired hearing resulting from ear damage, auditory nerve damage, and damage to the auditory cortex. This invention provides for simulation of microwave radiation which is normally produced by the auditory cortex. The simulated brain waves are introduced into the region of the auditory cortex and provide for perceived sounds on the part of the subject.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows the acoustic filter bank and mode control matrix portions of the hearing device of this invention.

FIG. 2 shows the microwave generation and antenna portion of the hearing device of this invention.

FIG. 3 shows a typical voltage divider network which may be used to provide mode partition.

FIG. 4 shows another voltage divider device which may be used to provide mode partition.

FIG. 5 shows a voltage divider to be used as a mode partition wherein each of the resistors is variable in order to provide adjustment of the voltage outputs.

FIG. 6 shows a modified hearing device which includes adjustable mode partitioning, and which is used to provide initial calibration of the hearing device.

FIG. 7 shows a group of variable oscillators and variable gain controls which are used to determine hearing characteristics of a particular subject.

FIG. 8 shows a top view of a human skull showing the lateral dimension.

FIG. 9 shows the relationship of the prolate spherical coordinate system to the cartesian system.

FIG. 10 shows a side view of a skull showing the medial plane of the head, section A—A.

FIG. 11 shows a plot of the transverse electric field amplitude versus primary mode number M.

FIG. 12 shows a left side view of the brain and auditory cortex.

FIG. 13 shows the total modal field versus angle for source location.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

This invention is based upon observations of the physical mechanism the mammalian brain uses to perceive acoustic vibrations. This observation is based in part upon neuro anatomical and other experimental evidence which relates to microwave brain stimulation and the perception of sounds.

It is has been observed that monochromatic acoustic stimuli (acoustic tones, or single tones) of different frequencies uniquely stimulate different regions of the cochlea. It has also been observed that there is a corresponding one to one relationship between the frequency of a monochromatic acoustic stimulus and the region of the auditory cortex neurally stimulated by the cochlear nerve under the physiologically normal conditions (tonotopicity).

It is has been observed that for an acoustic tone of a frequency which is at the lower end of the entire acoustical range perceivable by a person, that a thin lateral region ("Line") parallel to the medial axis of the brain and toward the inferior portion of the primary auditory cortex is stimulated. For an acoustic tone whose frequency is toward the high end of the entire perceivable acoustic range, a thin lateral region parallel to the medial axis and toward the superior portion of the primary auditory cortex is stimulated.

Neural stimulation results in the generation of a broad band of microwave photons by the change in rotational energy state of protons integral to the neuron membrane of the auditory cortex. The physical size and shape of the brain/skull cavity, together with the (semiconductor) properties (conductivity and dielectric constant) of the brain tissue provide an electromagnetic resonant cavity. Specific single frequencies are constructively reinforced so that a number of standing electromagnetic waves, each at its own single electromagnetic frequency in the microwave frequency region, are generated in the brain. Each such standing electromagnetic wave is called a characteristic mode of the brain/skull cavity.

Analysis in terms of prolate spheroidal wave functions indicates that transverse electric field components of these modes have maxima in the region of the auditory cortex. This analysis further shows that transverse electric field possess a variation of amplitude with angle in the angular plane (along the vertical dimension of the auditory cortex) and that is dependent only upon the primary mode number.

The auditory cortex in the normally functioning mammalian brain is a source of microwave modes. The auditory cortex generates these modes in accordance with the neural stimulation of the auditory cortex by the cochlear nerve. Mode weighting for any one acoustic tone stimulus is given by the amplitude of each mode along the line region of the auditory cortex which is neurally stimulated by that acoustic tone stimulus. A listing of mode weighting versus frequency of acoustic stimulus is called the mode matrix.

In this invention, the functions of the ear, the cochlear nerve, and the auditory cortex are simulated. Microwaves simulating the mode matrix are inserted directly into the region of the auditory cortex. By this insertion of simulated microwave modes, the normal operation of the entire natural hearing mechanism is simulated.

Referring now to FIG. 1 and FIG. 2 there is shown an apparatus which provides for induced perception of sound into a mammalian brain. This hearing device includes a microphone 10 which receives sounds, an acoustic filter bank 12 which separates the signals from the microphone into component frequencies, and a mode control matrix 14 which generates the mode signals which are used to control the intensity of microwave radiations which are injected into the skull cavity in the region of the auditory cortex.

The acoustic filter bank 12 consists of a bank of acoustic filters F1 through Fk which span the audible acoustic spectrum. These filters may be built from standard resistance, inductance, and capacitance components in accordance with well established practice. In the preferred embodiment there are 24 filters which correspond to the observed critical bandwidths of the human ear. In this preferred embodiment a typical list of filter parameters is given by Table I below:

TABLE I

Filter No.	Center Frequency (Hz)	Bandwidth (Hz)
1	50	less than 100
2	150	100
3	250	100
4	350	100
5	450	110
6	570	120
7	700	140
8	840	150
9	1,000	160
10	1,170	190
11	1,370	210
12	1,600	240
13	1,850	280
14	2,150	320
15	2,500	380
16	2,900	450
17	3,400	550
18	4,000	700
19	4,800	900
20	5,800	1,100
21	7,000	1,300
22	8,500	1,800
23	10,500	2,500
24	13,500	3,500

The rectifier outputs one through K are feed to K mode partition devices. The mode partitioning devices each have N outputs wherein N is the number of microwave oscillators used to generate the microwave radiation. The outputs 1 through N of each mode partition device is applied respectively to the inputs of each gain controlled amplifier of the microwave radiation generator. The function of the mode control matrix 14 is the control of the microwave amplifiers in the microwave amplifier bank 18. In the preferred embodiment thus will be 24 outputs and 24 microwave frequency oscillators.

Connected to each microwave amplifier gain control line is a mode simulation device 16 which receives weighted mode signals from the mode partition devices 14. Each mode simulation device consists of one through k lines and diodes 17 which are each connected to summing junction 19. The diodes 17 provide for isolation from one mode partition device to the next. The diodes 17 prevent signals from one mode partition device from returning to the other mode partition devices which are also connected to the same summing junction of the mode summation device 16. The diodes also serve a second function which is the rectification of the signals received from the acoustic filter bank by way of the mode partition devices. In this way each mode partition device output is rectified to produce a varying DC voltage with major frequency components of the order of 15 milliseconds or less. The voltage at the summation junction 19 is thus a slowly varying DC voltage.

The example mode partition devices are shown in greater detail in FIGS. 3, 4, and 5. The mode partition devices are merely resistance networks which produce 1 through N output voltages which are predetermined divisions of the input signal from the acoustic filter associated with the mode partition device. FIG. 3 shows a mode partitioning device wherein several outputs are associated with each series resistor 30. In the embodiment depicted in FIG. 4 there is an output associated with each series resistor only, and thus there are N series resistors, or the same number of series resistors as there are outputs. The values of the resistors in the mode partition resistor network are determined in accordance with the magnitudes of the frequency component from the acoustic filter bank 12 which is required at the summation point 19 or the gain control line for amplifiers 20.

The microwave amplifier bank 18 consists of a plurality of microwave oscillators 1 through N each of which is connected to an amplifier 20. Since the amplifiers 20 are gain controlled by the signals at summation junction 19, the magnitude of the microwave output is controlled by the mode control matrix outputs F1 through F_n. In the preferred embodiment there are 24 amplifiers.

The leads from the microwave oscillators 1 through N to the amplifiers 20 are shielded to prevent cross talk from one oscillator to the next, and to prevent stray signals from reaching the user of the hearing device. The output impedance of amplifiers 20 should be 1000 ohms and this is indicated by resistor 21. The outputs of amplifiers 20 are all connected to a summing junction 22. The summing junction 22 is connected to a summing impedance 23 which is approximately 50 ohms. The relatively high amplifier output impedance 21 as compared to the relatively low summing impedance 23 provides minimization of cross talk between the amplifiers. Since the amplitude of the microwave signal needed at the antenna 24 is relatively small, there is no need to match the antenna and summing junction impedances to the amplifier 20 output impedances. Efficiency of the amplifiers 20 is not critical.

Level control of the signal at antenna 24 is controlled by pick off 25 which is connected to the summing impedance 23. In this manner, the signal at antenna 24 can be varied from 0 (ground) to a value which is acceptable to the individual.

The antenna 24 is placed next to the subject's head and in the region of the subject's auditory cortex 26. By placement of the antenna 24 in the region of the auditory cortex 26, the microwave field which is generated simulates the microwave field which would be generated if the acoustic sounds were perceived with normal hearing and the auditory cortex was functioning normally.

In FIG. 2A there is shown a second embodiment of the microwave radiation and generator portion of the hearing device. In this embodiment a broad band microwave source 50 generates microwave signals which are feed to filters 52 through 58 which select from the broad band radiation particular frequencies to be transmitted to the person. As in FIG. 2, the amplifiers 20 receive signals on lines 19 from the mode control matrix. The signals on lines 19 provide the gain control for amplifiers 20.

In FIG. 6 there is shown a modified microwave hearing generator 60 which includes a mode partition resistor divider network as depicted in FIG. 5. Each of the mode partition voltage divider networks in this embodiment are individually adjustable for all of the resistances in the resistance network. FIG. 5 depicts a voltage division system wherein adjustment of the voltage partition resistors is provided for.

In FIG. 6, the sound source 62 generates audible sounds which are received by the microphone of the microwave hearing generator 60. In accordance with the operation described with respect to FIGS. 1 and 2, microwave signals are generated at the antenna 10 in accordance with the redistribution provided by the mode control matrix as set forth in FIG. 5.

The sound source 62 also produces a signal on line 64 which is received by a head phone 66. The apparatus depicted in FIG. 6 is used to calibrate or fit a microwave hearing generator to a particular individual. Once the hearing generator is adjusted to the particular individual by adjustment of the variable resistors in the adjustable mode partition portion of the hearing generator, a second generator may be built using fixed value resistors in accordance with the adjusted values achieved in fitting the device to the particular subject. The sound produced by headphone 66 should be the same as a sound from the sound source 62 which is received by the microphone 10 in the microwave hearing generator 60. In this way, the subject can make comparisons between the perceived sound from the hearing generator 60, and the sound which is heard from headphone 66. Sound source 62 also produces a signal on 68 which is feed to cue light 69. Cue light 69 comes on whenever a sound is emitted from sound source 62 to the microwave generator 60. In this manner, if the subject hears nothing, he will still be informed that a sound has been omitted and hence that he is indeed perceiving no sound from the microwave hearing generator 60.

In FIG. 7 there is shown a modified microwave hearing generator which may be used to determine a subject's microwave mode frequencies. In this device, the acoustic filter bank and the mode control matrix have been removed and replaced by voltage level signal generated by potentiometers 70. Also included are a plurality of variable frequency oscillators 72 which feed microwave amplifiers 74 which are gain controlled from the signal generated by potentiometers 70 and pick off arm 76.

This modified microwave hearing generator is used to provide signals using one oscillator at a time. When an oscillator is turned on, the frequency is varied about the estimated value until a maximum acoustic perception by the subject is perceived. This perception however may consist of a buzzing or hissing sound rather than a tone because only one microwave frequency is being received. The first test of perception is to determine the subject's lowest modal frequency for audition $(M=1)$. Once this modal frequency is obtained, the process is repeated for several higher modal frequencies and continued until no maximum acoustic perception occurs.

Another method of determination of a subject's modal frequencies is through anatomical estimation. This procedure is by measurement of the subject's cephalic index and the lateral dimensions of the skull. In this method, the shape is determined in prolate spheroidal coordinance.

Purely anatomical estimation of subject's modal frequencies is performed by first measuring the maximum lateral dimension (breadth) L FIG. 8, of the subject's head together with the maximum dimension D (anterior to posterior) in the medial plane of the subject's head. D is the distance along Z axis as shown in FIG. 10. The ratio L/D, called in anthropology the cephalic index, is monotonically related to the boundary value ξ_0 defining the ellipsoidal surface approximating the interface between the brain and the skull in the prolate spheroidal coordinate system. ξ_0 defines the shape of this interface: ξ_0 and D together give an estimate of a, the semi-focal distance of the defining ellipsoid. Using ξ_0 and a, together with known values of the conductivity and dielectric constants of brain tissue, those wavelengths are found f[illegible] which the radial component of the electric [illegible]s the boundary condition that it is zero at ξ_0.

These wavelengths are the wavelengths associated with the standing waves or modes; the corresponding frequencies are found by dividing the phase velocity of microwaves in brain tissue by each of the wavelengths.

A subject's microwave modal frequencies may also be determined by observing the effect of external microwave radiation upon the EEG. The frequency of the M equal 1 mode may then be used as a base point to estimate all other modal frequencies.

A typical example of such an estimation is where the subject is laterally irradiated with a monochromatic microwave field simultaneous with EEG measurement and the microwave frequency altered until a significant change occurs in the EEG, the lowest such frequency causing a significant EEG change is found. This is identified as the frequency of the $M=1$ mode, the lowest mode of importance in auditory perception. The purely anatomical estimation procedure (FIGS. 8, 9, 10) is then performed and the ratio of each modal frequency to the $M=1$ modal frequency obtained. These ratios together with the experimentally-determined $M=1$ frequency are then used to estimate the frequencies of the mode numbers higher than 1. The prolate spheroidal coordinate system is shown in FIG. 9. Along the lateral plane containing the x and y coordinates of FIG. 9, the prolate spheroidal coordinate variable ϕ (angle) lies FIGS. 9 and 10. Plots of the transverse electric field amplitude versus primary mode number m are shown in FIG. 11. The equation is

$$E_{transverse}(m, \phi) = E_o \sin(m\,\phi)$$

The "elevation view" FIG. 12, of the brain from the left side, shows the primary auditory cortex 10. The isotone lines and the high frequency region are toward the top of 100 and the low frequency region toward the bottom of 100.

The formula I, set forth below is the formula for combining modes from an iso-tone line at $\phi=\phi j$ being excited to obtain the total modal field at some other angular location ϕ. For this formula, if we let $J=1$ (just one iso-tone single frequency acoustic stimulus line), then it can be shown that ALL modes (in general) must be used for any ONE tone.

FORMULA I
RMS TRANSVERSE ELECTRIC FIELD IN ANGULAR PLANE, f(0)

$$f(0) = \left[\sum_{m=1}^{M}\left(\sin(m0)\cdot\sum_{j=1}^{J} e^{-(0-0_j)/\Delta 0_m}\sin(m0_j)\right)^{2}\right]^{\frac{1}{2}}$$

ϕ = ANGLE (0° LATERAL)
ϕ_j = LOCATION OF j-TH SOURCE (TOTAL NUMBER J)
$\Delta\phi_m$ = ATTENUATION LENGTH (IN ANGLE) OF m-TH MODE
m = PRIMARY MODE NUMBER (HIGHEST MODE M)

FIG. 13 shows the resulting total modal field versus angle ϕ for source location ϕ at 5.25°, 12.5°, etc. With reference to the set of curves at the left top of this figure. A spacing of approximately 7.25° in ϕ corresponds to a tonal difference of about 1 octave. This conclusion is based on the side-lobes of pattern coming from $\phi=5.25°$, etc. The total filed (value on y-axis) falls considerably below the top curves for source locations well below 5.25° (toward the high acoustic stimulus end) and also as the source of frequency goes well above 30° (low frequency end). ϕ is plotted positive downward from 0° at lateral location as indicates in FIG. 11.

Resistor weightings are obtained from the $|\sin(m[\phi-\phi j])|$, Formula I. The scale between acoustic frequency and ϕ must be set or estimated from experiment. Approximately 5.25±1° corresponds to a tonal stimulus at about 2 kHz (the most sensitive region of the ear) since this source location gives the highest electric field amplitude.

The apparatus of FIG. 7 may also be used to determine values for a hearing device which are required for a particular subject. Once the modal frequencies have been estimated, the device of FIG. 7 which includes variable microwave oscillators may be used to determine values for the oscillators which match the subject, and to determine resistance values associated with the mode partition devices of the mode control matrix.

In FIG. 7 manual control of the amplifier gain is achieved by potentiometers 76. In this manner the amplifier gains are varied about the estimated settings for an acoustic tone stimulus in the region of two thousand Hertz (2 kHz) until maximum acoustic perception and a purest tone are achieved together. The term purest tone may also be described as the most pleasing acoustic perception by the subject. This process may be repeated at selected frequencies above and below 2 kHz. The selected frequencies correspond to regions of other acoustic filter center frequencies of the subject. When modal frequency (oscillator frequency) and gain set values (setting a potentiometer 76) are noted, it is then possible to calculate fixed oscillator frequencies and control resistor values for the adjusted hearing device for this particular subject.

In the event the subject has no prior acoustic experience, that is deaf from birth, estimated resistor values must be used. Also, a complex acoustic stimulation test including language articulation and pairs of harmonically related tones may be developed to maximize the match of the hearing device parameters for those of this particular subject.

Typical components for use in this invention include commercially available high fidelity microphones which have a range of 50 Hz to 15 kHz with plus or minus 3 dB variation.

The audio filters to be used with the acoustic filter bank 12 are constructed in a conventional manner, and have Q values of about 6. The filters may also be designed with 3 dB down points (½ the bandwidth away from the center frequency) occurring at adjacent center frequency locations.

The diodes 17 in the mode control matrix which provide isolation between the mode partition circuits are commercially available diodes in the audio range.

The microwave oscillators 1 through N and the microwave amplifiers 20 are constructed with available microwave transistors which can be configured either as oscillators or amplifiers. Examples of the transistors are GaAsFET field effect transistors by Hewlitt Packard known as the HFET series or silicone bipolar transistors by Hewlitt Packard known as the HXTR series.

All the cable between the oscillators, the microwave amplifiers, and the antenna should be constructed with either single or double shielded coaxial cable.

The antenna 24 for directing microwave signals to the audio cortex 26 should be approximately the size of the auditory cortex. A typical size would be one and

16

one half CM high and one half to one CM wide. The antenna as shown is located over the left auditory cortex, but the right may also be used. Since the characteristic impedance of the brain tissue at these microwave frequencies is close to 50 ohms, efficient transmission by commercially available standard 50 ohm coax is possible.

The invention has been described in reference to the preferred embodiments. It is, however, to be understood that other advantages, features, and embodiments may be within the scope of this invention as defined in the appended claims.

What is claimed is:

1. A sound perception device for providing induced perception of sound into a mammalian brain comprising in combination:

means for generating microwave radiation which is representative of a sound to be perceived, said means for generating including means for generating a simultaneous plurality of microwave radiation frequencies and means for adjusting the amplitude of said microwave radiation frequencies in accordance with the sound to be perceived; and

antenna means located in the region of the auditory cortex of said mammalian brain for transmitting said microwave energy into the auditory cortex region of said brain.

2. A hearing device for perception of sounds comprising in combination:

means for generating a signal representative of sounds;

means for analyzing said signal representative of said sounds having an output;

means for generating a plurality of microwave signals having different frequencies having a input connected to said output of said means for analyzing said signals, having an output;

means for applying said plurality of microwave signals to the head of a subject, and

whereby the subject perceives sounds which are representative of said sounds.

3. The apparatus in accordance with claim 2 wherein said means for generating a signal is a microphone for detecting sound waves.

4. The apparatus in accordance with claim 2 wherein said means for applying said plurality of microwave signals is an antenna.

5. The apparatus in accordance with claim 4 wherein said antenna is placed in the region of the auditory cortex of the subject.

6. The apparatus in accordance with claim 2 wherein the subject is a human being.

7. The apparatus in accordance with claim 2 wherein said means for analyzing said signal comprises:

an acoustic filter bank for dividing said sounds into a plurality of component frequencies; and

a mode control matrix means for providing control signals which are weighted in accordance with said plurality of component frequencies, having an output connected to said means for generating a plurality of microwave signal inputs.

8. The apparatus in accordance with claim 7 wherein said acoustic filter bank includes a plurality of audio frequency filters.

9. The apparatus in accordance with claim 8 wherein said audio frequency filters provide a plurality of output frequencies having amplitudes which are a function of said signal representative of sounds.

10. The apparatus in accordance with claim 9 wherein said amplitudes are the weighted in accordance with transform function of the signal representative of sounds.

11. The apparatus in accordance with claim 7 wherein said mode control matrix device includes a voltage divider connected to each of said plurality of said audio frequency filters.

12. The apparatus in accordance with claim 11 wherein each of said voltage dividers has a plurality of outputs which are connected in circuit to said means for generating a plurality of microwave signals.

13. The apparatus in accordance with claim 2 wherein said means for generating a plurality of microwave signals comprises a plurality of microwave generators each having a different frequency and means for controlling the output amplitude of each of said generators.

14. The apparatus in accordance with claims 2 wherein said means for generating a plurality of microwave signals comprises a broad band microwave source and a plurality of filters.

15. The apparatus in accordance with claim 13 wherein said generators each comprise a microwave signal source and a gain controlled microwave amplifier.

16. The apparatus in accordance with claim 13 wherein said means for analyzing output is connected to said means for controlling microwave amplifier output amplitudes.

17. The apparatus in accordance with claim 13 wherein analyzing includes K audio frequency filters.

18. The apparatus in accordance with claim 17 wherein there are N microwave generators.

19. The apparatus in accordance with claim 18 including a mode partitioning means which provides N outputs for each of said K audio frequency filters.

20. The apparatus in accordance with claim 19 wherein said N amplifiers each have K inputs from said mode partitioning means.

21. The apparatus in accordance with claim 20 wherein said N amplifiers have K inputs less the mode partitioning means outputs which are so small that they may be omitted.

22. The apparatus in accordance with claim 20 wherein said mode partitioning output device outputs each include a diode connected to each microwave amplifier gain control to provide isolation between all outputs.

23. The apparatus in accordance with claim 20 wherein said K audio frequency filters are chosen to correspond to the critical bandwidths of the human ear.

24. The apparatus in accordance with claim 20 wherein said N microwave generators are each adjustable in frequency output.

25. The apparatus in accordance with claim 18 wherein the frequency of each N microwave generators is determined by anatomical estimation.

26. The apparatus in accordance with claim 18 wherein the frequency of the lowest frequency microwave generator is chosen by determination of the effect of external microwave generation on the EEG of the subject.

27. The apparatus in accordance with claim 18 wherein the frequency of each of said N microwave generators corresponds to the subject's microwave modal frequencies.

17

28. The apparatus in accordance with claim 27 wherein the subject's modal frequencies are determined by measurement of the subject's cephalic index and the lateral dimensions of the skull.

29. The apparatus in accordance with claim 28 wherein the subject's lowest modal frequency is determined by varying the frequency of the lowest frequency microwave generator about the estimated value until a maximum acoustic perception is obtained by the subject.

* * * * *

United States Patent [19]

Malech

[11] **3,951,134**

[45] **Apr. 20, 1976**

---

[54] **APPARATUS AND METHOD FOR REMOTELY MONITORING AND ALTERING BRAIN WAVES**

[75] Inventor: **Robert G. Malech**, Plainview, N.Y.

[73] Assignee: **Dorne & Margolin Inc.**, Bohemia, N.Y.

[22] Filed: **Aug. 5, 1974**

[21] Appl. No.: **494,518**

[52] U.S. Cl. .... **128/2.1 B**

[51] **Int. Cl.²** **A61B 5/04**

[58] **Field of Search** .... 128/1 C, 1 R, 2.1 B, 128/2.1 R, 419 R, 422 R, 420, 404, 2 R, 2 S, 2.05 R, 2.05 V, 2.05 F, 2.06 R; 340/248 A, 258 A, 258 B, 258 D, 229

[56] **References Cited**

UNITED STATES PATENTS

2,860,627	11/1958	Harden et al.	128/2.1 B
3,096,768	7/1963	Griffith, Jr.	128/420
3,233,450	2/1966	Fry	128/2.1 R
3,483,860	12/1969	Namerow	128/2.05 F
3,495,596	2/1970	Condict	128/1 C
3,555,529	1/1971	Brown et al.	128/2.1 R
3,773,049	11/1973	Rabichev et al.	128/1 C
3,796,208	3/1974	Bloice	128/2 S

Primary Examiner—William E. Kamm
Attorney, Agent, or Firm—Darby & Darby

[57] **ABSTRACT**

Apparatus for and method of sensing brain waves at a position remote from a subject whereby electromagnetic signals of different frequencies are simultaneously transmitted to the brain of the subject in which the signals interfere with one another to yield a waveform which is modulated by the subject's brain waves. The interference waveform which is representative of the brain wave activity is re-transmitted by the brain to a receiver where it is demodulated and amplified. The demodulated waveform is then displayed for visual viewing and routed to a computer for further processing and analysis. The demodulated waveform also can be used to produce a compensating signal which is transmitted back to the brain to effect a desired change in electrical activity therein.

11 Claims, 2 Drawing Figures

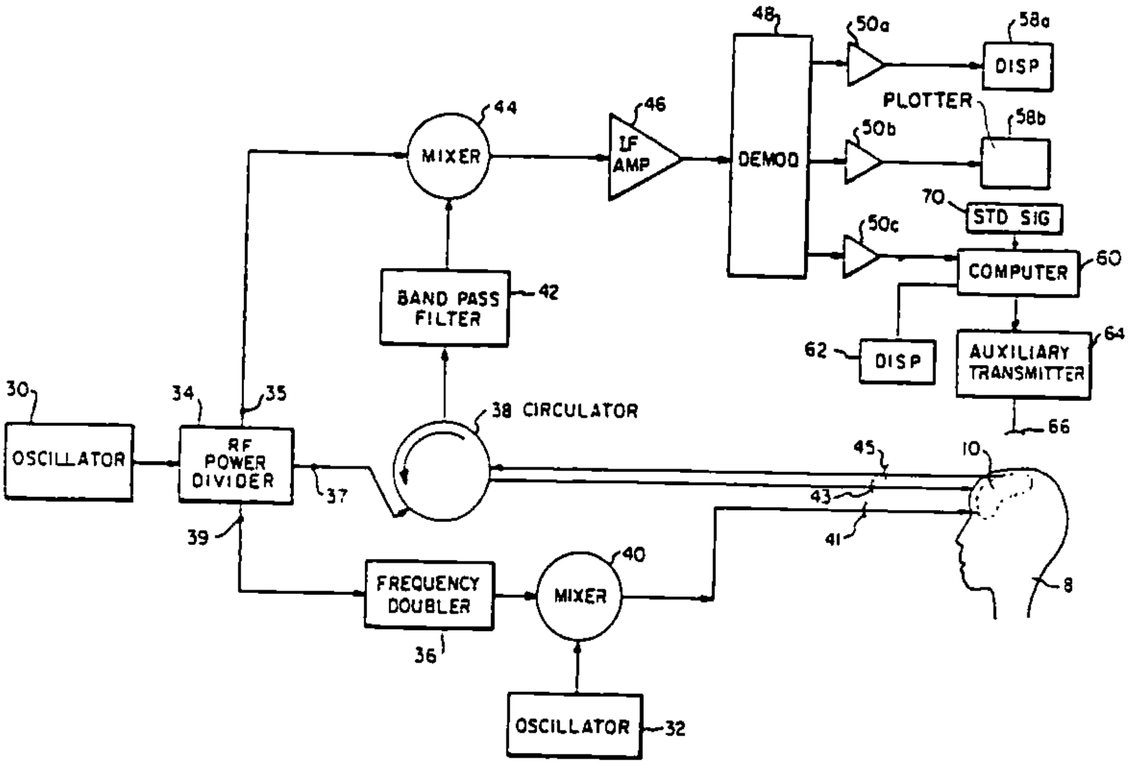
48
50a
58a
DISP
44
46
PLOTTER
58b
MIXER
LF AMP
DEMOD
50b
70
STD SIG
50c
60
COMPUTER
BAND PASS FILTER
42
62
DISP
AUXILIARY TRANSMITTER
64
66
30
34
35
38 CIRCULATOR
45
10
OSCILLATOR
RF POWER DIVIDER
37
43
41
39
40
8
FREQUENCY DOUBLER
MIXER
36
OSCILLATOR
32

19

FIG. 1

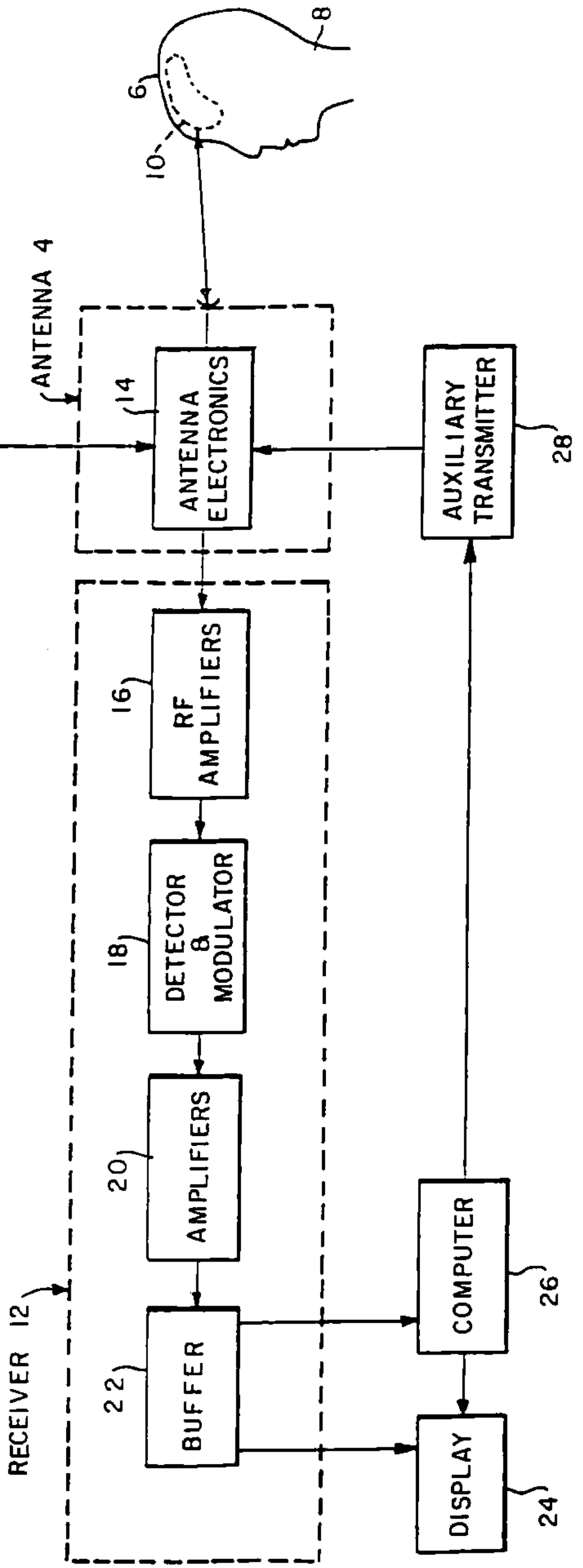
8
6
10
ANTENNA 4
2
TRANSMITTER
14
ANTENNA ELECTRONICS
AUXILIARY TRANSMITTER
28
16
RF AMPLIFIERS
18
DETECTOR & MODULATOR
20
AMPLIFIERS
RECEIVER 12
22
BUFFER
COMPUTER
26
DISPLAY
24

20

FIG. 2

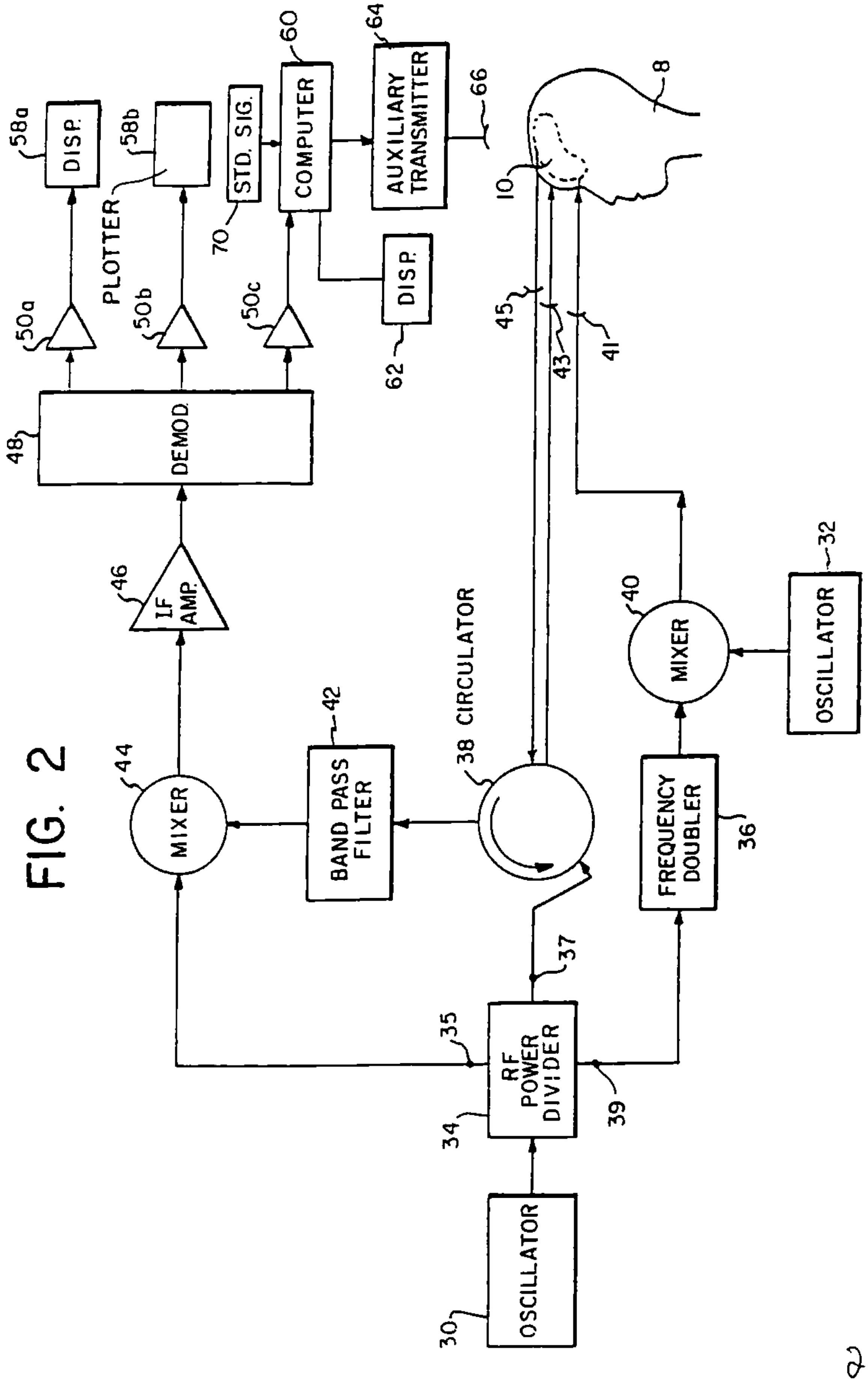
OSCILLATOR
RF POWER DIVIDER
FREQUENCY DOUBLER
38 CIRCULATOR
MIXER
OSCILLATOR
BAND PASS FILTER
MIXER
IF AMP.
DEMOD
DISP.
PLOTTER
STD. SIG.
COMPUTER
DISP.
AUXILIARY TRANSMITTER

21

APPARATUS AND METHOD FOR REMOTELY MONITORING AND ALTERING BRAIN WAVES

BACKGROUND OF THE INVENTION

Medical science has found brain waves to be a useful barometer of organic functions. Measurements of electrical activity in the brain have been instrumental in detecting physical and psychic disorder, measuring stress, determining sleep patterns, and monitoring body metabolism.

The present art for measurement of brain waves employs electroencephalographs including probes with sensors which are attached to the skull of the subject under study at points proximate to the regions of the brain being monitored. Electrical contact between the sensors and apparatus employed to process the detected brain waves is maintained by a plurality of wires extending from the sensors to the apparatus. The necessity for physically attaching the measuring apparatus to the subject imposes several limitations on the measurement process. The subject may experience discomfort, particulary if the measurements are to be made over extended periods of time. His bodily movements are restricted and he is generally confined to the immediate vicinity of the measuring apparatus. Furthermore, measurements cannot be made while the subject is conscious without his awareness. The comprehensiveness of the measurements is also limited since the finite number of probes employed to monitor local regions of brain wave activity do not permit observation of the total brain wave profile in a single test.

SUMMARY OF THE INVENTION

The present invention relates to apparatus and a method for monitoring brain waves wherein all components of the apparatus employed are remote from the test subject. More specifically, high frequency transmitters are operated to radiate electromagnetic energy of different frequencies through antennas which are capable of scanning the entire brain of the test subject or any desired region thereof. The signals of different frequencies penetrate the skull of the subject and impinge upon the brain where they mix to yield an interference wave modulated by radiations from the brain's natural electrical activity. The modulated interference wave is re-transmitted by the brain and received by an antenna at a remote station where it is demodulated, and processed to provide a profile of the suject's brain waves. In addition to passively monitoring his brain waves, the subject's neurological processes may be affected by transmitting to his brain, through a transmitter, compensating signals. The latter signals can be derived from the received and processed brain waves.

OBJECTS OF THE INVENTION

It is therefore an object of the invention to remotely monitor electrical activity in the entire brain or selected local regions thereof with a single measurement.

Another object is the monitoring of a subject's brain wave activity through transmission and reception of electromagnetic waves.

Still another object is to monitor brain wave activity from a position remote from the subject.

A further object is to provide a method and apparatus for affecting brain wave activity by transmitting electromagnetic signals thereto.

DESCRIPTION OF THE DRAWINGS

Other and further objects of the invention will appear from the following description and the accompanying drawings, which form part of the instant specification and which are to be read in conjunction therewith, and in which like reference numerals are used to indicate like parts in the various views;

FIG. 1 is a block diagram showing the interconnection of the components of the apparatus of the invention;

FIG. 2 is a block diagram showing signal flow in one embodiment of the apparatus.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring to the drawings, specifically FIG. 1, a high frequency transmitter 2 produces and supplies two electromagnetic wave signals through suitable coupling means 14 to an antenna 4. The signals are directed by the antenna 4 to the skull 6 of the subject 8 being examined. The two signals from the antenna 4, which travel independently, penetrate the skull 6 and impinge upon the tissue of the brain 10.

Within the tissue of the brain 10, the signals combine, much in the manner of a conventional mixing process technique, with each section of the brain having a different modulating action. The resulting waveform of the two signals has its greatest amplitude when the two signals are in phase and thus reinforcing one another. When the signals are exactly 180° out of phase the combination produces a resultant waveform of minimum amplitude. If the amplitudes of the two signals transmitted to the subject are maintained at identical levels, the resultant interference waveform, absent influences of external radiation, may be expected to assume zero intensity when maximum interference occurs, the number of such points being equal to the difference in frequencies of the incident signals. However, interference by radiation from electrical activity within the brain 10 causes the waveform resulting from interference of the two transmitted signals to vary from the expected result, i.e., the interference waveform is modulated by the brain waves. It is believed that this is due to the fact that brain waves produce electric charges each of which has a component of electromagnetic radiation associated with it. The electromagnetic radiation produced by the brain waves in turn reacts with the signals transmitted to the brain from the external source.

The modulated interference waveform is re-transmitted from the brain 10, back through the skull 6. A quantity of energy is re-transmitted sufficient to enable it to be picked up by the antenna 4. This can be controlled, within limits, by adjusting the absolute and relative intensities of the signals, originally transmitted to the brain. Of course, the level of the transmitted energy should be kept below that which may be harmful to the subject.

The antenna passes the received signal to a receiver 12 through the antenna electronics 14. Within the receiver the wave is amplified by conventional RF amplifiers 16 and demodulated by conventional detector and modulator electronics 18. The demodulated wave, representing the intra-brain electrical activity, is amplified by amplifiers 20 and the resulting information in electronic form is stored in buffer circuitry 22. From the buffers 22 the information is fed to a suitable visual display 24, for example one employing a cathode ray tube, light emitting diodes, liquid crystals, or a mechanical plotter. The information may also be channeled to a computer 26 for further processing and analysis with the output of the computer displayed by heretofore mentioned suitable means.

In addition to channeling its information to display devices 24, the computer 26 can also produce signals to control an auxiliary transmitter 28. Transmitter 28 is used to produce a compensating signal which is transmitted to the brain 10 of the subject 8 by the antenna 4. In a preferred embodiment of the invention, the compensating signal is derived as a function of the received brain wave signals, although it can be produced separately. The compensating signals affect electrical activity within the brain 10.

Various configurations of suitable apparatus and electronic circuitry may be utilized to form the system generally shown in FIG. 1 and one of the many possible configurations is illustrated in FIG. 2. In the example shown therein, two signals, one of 100 MHz and the other of 210 MHz are transmitted simultaneously and combine in the brain 10 to form a resultant wave of frequency equal to the difference in frequencies of the incident signals, i.e., 110 MHz. The sum of the two incident frequencies is also available, but is discarded in subsequent filtering. The 100 MHz signal is obtained at the output 37 of an RF power divider 34 into which a 100 MHz signal generated by an oscillator 30 is injected. The oscillator 30 is of a conventional type employing either crystals for fixed frequency circuits or a tunable circuit set to oscillate at 100 MHz. It can be a pulse generator, square wave generator or sinusoidal wave generator. The RF power divider can be any conventional VHF, UHF or SHF frequency range device constructed to provide, at each of three outputs, a signal identical in frequency to that applied to its input.

The 210 MHz signal is derived from the same 100 MHz oscillator 30 and RF power divider 34 as the 100 MHz signal, operating in concert with a frequency doubler 36 and 10 MHz oscillator 32. The frequency doubler can be any conventional device which provides at its output a signal with frequency equal to twice the frequency of a signal applied at its input. The 10 MHz oscillator can also be of conventional type similar to the 100 MHz oscillator herebefore described. A 100 MHz signal from the output 39 of the RF power divider 34 is fed through the frequency doubler 36 and the resulting 200 MHz signal is applied to a mixer 40. The mixer 40 can be any conventional VHF, UHF or SHF frequency range device capable of accepting two input signals of differing frequencies and providing two output signals with frequencies equal to the sum and difference in frequencies respectively of the input signals. A 10 MHz signal from the oscillator 32 is also applied to the mixer 40. The 200 MHz signal from the doubler 36 and the 10 MHz signal from the oscillator 32 combine in the mixer 40 to form a signal with a frequency of 210 MHz equal to the sum of the frequencies of the 200 MHz and 10 MHz signals.

The 210 MHz signal is one of the signals transmitted to the brain 10 of the subject being monitored. In the arrangement shown in FIG. 2, an antenna 41 is used to transmit the 210 MHz signal and another antenna 43 is used to transmit the 100 MHz signal. Of course, a single antenna capable of operating at 100 MHz and 210 MHz frequencies may be used to transmit both signals. The scan angle, direction and rate may be controlled mechanically, e.g., by a reversing motor, or electronically, e.g., by energizing elements in the antenna in proper synchronization. Thus, the antenna(s) can be of either fixed or rotary conventional types.

A second 100 MHz signal derived from output terminal 37 of the three-way power divider 34 is applied to a circulator 38 and emerges therefrom with a desired phase shift. The circulator 38 can be of any conventional type wherein a signal applied to an input port emerges from an output port with an appropriate phase shift. The 100 MHz signal is then transmitted to the brain 10 of the subject being monitored via the antenna 43 as the second component of the dual signal transmission. The antenna 43 can be of conventional type similar to antenna 41 herebefore described. As previously noted, these two antennas may be combined in a single unit.

The transmitted 100 and 210 MHz signal components mix within the tissue in the brain 10 and interfere with one another yielding a signal of a frequency of 110 MHz, the difference in frequencies of the two incident components, modulated by electromagnetic emissions from the brain, i.e., the brain wave activity being monitored. This modulated 110 MHz signal is radiated into space.

The 110 MHz signal, modulated by brain wave activity, is picked up by an antenna 45 and channeled back through the circulator 38 where it undergoes an appropriate phase shift. The circulator 38 isolates the transmitted signals from the received signal. Any suitable diplexer or duplexer can be used. The antenna 45 can be of conventional type similar to antennas 41 and 43. It can be combined with them in a single unit or it can be separate. The received modulated 110 MHz signal is then applied to a band pass filter 42, to eliminate undesirable harmonics and extraneous noise, and the filtered 110 MHz signal is inserted into a mixer 44 into which has also been introduced a component of the 100 MHz signal from the source 30 distributed by the RF power divider 34. The filter 42 can be any conventional band pass filter. The mixer 44 may also be of conventional type similar to the mixer 40 herebefore described.

The 100 MHz and 110 MHz signals combine in the mixer 44 to yield a signal of frequency equal to the difference in frequencies of the two component signals, i.e., 10 MHz still modulated by the monitored brain wave activity. The 10 MHz signal is amplified in an IF amplifier 46 and channeled to a demodulator 48. The IF amplifier and demodulator 48 can both be of conventional types. The type of demodulator selected will depend on the characteristics of the signals transmitted to and received from the brain, and the information desired to be obtained. The brain may modulate the amplitude, frequency and/or phase of the interference waveform. Certain of these parameters will be more sensitive to corresponding brain wave characteristics than others. Selection of amplitude, frequency or phase demodulation means is governed by the choice of brain wave characteristic to be monitored. If desired, several different types of demodulators can be provided and used alternately or at the same time.

The demodulated signal which is representative of the monitored brain wave activity is passed through audio amplifiers 50 *a*, *b*, *c* which may be of conventional type where it is amplified and routed to displays 58 *a*, *b*, *c* and a computer 60. The displays 58 *a*, *b*, *c* present the raw brain wave signals from the amplifiers

50 *a, b, c*. The computer **60** processes the amplified brain wave signals to derive information suitable for viewing, e.g., by suppressing, compressing, or expanding elements thereof, or combining them with other information-bearing signals and presents that information on a display **62**. The displays can be conventional ones such as the types herebefore mentioned employing electronic visual displays or mechanical plotters **58***b*. The computer can also be of conventional type, either analog or digital, or a hybrid.

A profile of the entire brain wave emission pattern may be monitored or select areas of the brain may be observed in a single measurement simply by altering the scan angle and direction of the antennas. There is no physical contact between the subject and the monitoring apparatus. The computer **60** also can determine a compensating waveform for transmission to the brain **10** to alter the natural brain waves in a desired fashion. The closed loop compensating system permits instantaneous and continuous modification of the brain wave response pattern.

In performing the brain wave pattern modification function, the computer **60** can be furnished with an external standard signal from a source **70** representative of brain wave activity associated with a desired nuerological response. The region of the brain responsible for the response is monitored and the received signal, indicative of the brain wave activity therein, is compared with the standard signal. The computer **60** is programmed to determine a compensating signal, responsive to the difference between the standard signal and received signal. The compensating signal, when transmitted to the monitored region of the brain, modulates the natural brain wave activity therein toward a reproduction of the standard signal, thereby changing the neurological response of the subject.

The computer **60** controls an auxiliary transmitter **64** which transmits the compensating signal to the brain **10** of the subject via an antenna **66**. The transmitter **64** is of the high frequency type commonly used in radar applications. The antenna **66** can be similar to antennas **41, 43** and **45** and can be combined with them. Through these means, brain wave activity may be altered and deviations from a desired norm may be compensated. Brain waves may be monitored and control signals transmitted to the brain from a remote station.

It is to be noted that the configuration described is one of many possibilities which may be formulated without departing from the spirit of my invention. The transmitters can be monostratic or bistatic. They also can be single, dual, or multiple frequency devices. The transmitted signal can be continuous wave, pulse, FM, or any combination of these as well as other transmission forms. Typical operating frequencies for the transmitters range from 1 MHz to 40 GHz but may be altered to suit the particular function being monitored and the characteristics of the specific subject.

The individual components of the system for monitoring and controlling brain wave activity may be of conventional type commonly employed in radar systems.

Various subassemblies of the brain wave monitoring and control apparatus may be added, substituted or combined. Thus, separate antennas or a single multimode antenna may be used for transmission and reception. Additional displays and computers may be added to present and analyze select components of the monitored brain waves.

Modulation of the interference signal retransmitted by the brain may be of amplitude, frequency and/or phase. Appropriate demodulators may be used to decipher the subject's brain activity and select components of his brain waves may be analyzed by computer to determine his mental state and monitor his thought processes.

As will be appreciated by those familiar with the art, apparatus and method of the subject invention has numerous uses. Persons in critical positions such as drivers and pilots can be continuously monitored with provision for activation of an emergency device in the event of human failure. Seizures, sleepiness and dreaming can be detected. Bodily functions such as pulse rate, heartbeat reqularity and others also can be monitored and occurrences of hallucinations can be detected. The system also permits medical diagnoses of patients, inaccessible to physicians, from remote stations.

What is claimed is:

1. Brain wave monitoring apparatus comprising
   means for producing a base frequency signal,
   means for producing a first signal having a frequency related to that of the base frequency and at a predetermined phase related thereto,
   means for transmitting both said base frequency and said first signals to the brain of the subject being monitored,
   means for receiving a second signal transmitted by the brain of the subject being monitored in response to both said base frequency and said first signals,
   mixing means for producing from said base frequency signal and said received second signal a response signal having a frequency related to that of the base frequency, and
   means for interpreting said response signal.

2. Apparatus as in claim 1 where said receiving means comprises
   means for isolating the transmitted signals from the received second signals.

3. Apparatus as in claim 2 further comprising a band pass filter with an input connected to said isolating means and an output connected to said mixing means.

4. Apparatus as in claim 1 further comprising means for amplifying said response signal.

5. Apparatus as in claim 4 further comprising means for demodulating said amplified response signal.

6. Apparatus as in claim 5 further comprising interpreting means connected to the output of said demodulator means.

7. Apparatus according to claim 1 further comprising
   means for producing an electromagnetic wave control signal dependent on said response signal, and
   means for transmitting said control signal to the brain of said subject.

8. Apparatus as in claim 7 wherein said transmitting means comprises means for directing the electromagnetic wave control signal to a predetermined part of the brain.

9. A process for monitoring brain wave activity of a subject comprising the steps of
   transmitting at least two electromagnetic energy signals of different frequencies to the brain of the subject being monitored,
   receiving an electromagnetic energy signal resulting from the mixing of said two signals in the brain modulated by the brain wave activity and retrans-

mitted by the brain in response to said transmitted energy signals, and,

interpreting said received signal.

10. A process as in claim 9 further comprising the step of transmitting a further electromagnetic wave signal to the brain to vary the brain wave activity.

11. A process as in claim 10 wherein the step of transmitting the further signals comprises

obtaining a standard signal,

comparing said received electromagnetic energy signals with said standard signal,

producing a compensating signal corresponding to the comparison between said received electrogagnetic energy signals and the standard signal, and

transmitting the compensating signals to the brain of the subject being monitored.

* * * * *

United States Patent [19]
Zanakis et al.

[11] Patent Number: 4,951,674
[45] Date of Patent: Aug. 28, 1990

[54] BIOMAGNETIC ANALYTICAL SYSTEM USING FIBER-OPTIC MAGNETIC SENSORS

[76] Inventors: **Michael F. Zanakis**, 60 Martin Rd., Livingston, N.J. 07039; **Philip A. Femano**, 69 Alexander Ave., Nutley, N.J. 07110

[21] Appl. No.: 325,942

[22] Filed: Mar. 20, 1989

[51] Int. Cl.^{5} A61B 5/04
[52] U.S. Cl. .... 128/653 R; 128/731; 324/244.1
[58] Field of Search .... 324/244 OP; 128/653 R, 128/639, 630, 731, 732

[56] References Cited

U.S. PATENT DOCUMENTS

4,591,787 5/1986 Hoening 324/260
4,771,239 9/1988 Hoenig .... 128/653 R

OTHER PUBLICATIONS

"Introduction to Magnetoencephalography—A New Window on the Brain", Biomagnetic Technologies, Inc., San Diego, Calif., undated.

Kersey et al., Journal of Lightwave Technology, vol. LT-3, No. 4, Aug. 1985, pp. 836–840.

Enokiharae et al., Journal of Lightwave Technology, vol. LT-5, No. 11, Nov. 1987, pp. 1584–1590.

Mermelstein, Journal of Lightwave Technology, vol. LT-4, No. 9, Sep. 1986, pp. 1376–1380.

Koo et al., Journal of Lightwave Technology, vol. LT-5, No. 12, Dec. 1987, pp. 1680–1684.

Yariv et al., Optics Letters, vol. 5, No. 3, Mar. 1980, pp. 87–89.

Koo et al., Optics Letters, vol. 7, No. 7, Jul. 1982, pp. 334–336.

Koo et al., J. Lightwave Tech., vol. LT-1, No. 3, Sep. 1983, pp. 524–525.

Primary Examiner—Lee S. Cohen
Assistant Examiner—John C. Hanley
Attorney, Agent, or Firm—Michael Ebert

[57] ABSTRACT

A biomagnetic analytical system for sensing and indicating minute magnetic fields emanating from the brain or from any other tissue region of interest in a subject under study. The system includes a magnetic pick-up device constituted by an array of fiber-optic magnetic sensors mounted at positions distributed throughout the inner confines of a magnetic shield configured to conform generally to the head of the subject or whatever other body region is of interest. Each sensor yields a light beam whose phase or other parameter is modulated in accordance with the magnetic field emanating from the related site in the region. The modulated beam from each sensor is compared in an interferometer with a reference light beam to yield an output signal that is a function of the magnetic field being emitted at the related site. The output signals from the interferometer are processed to provide a display or recording exhibiting the pattern or map of magnetic fields resulting from emanations at the multitude of sites encompassed by the region.

8 Claims, 2 Drawing Sheets

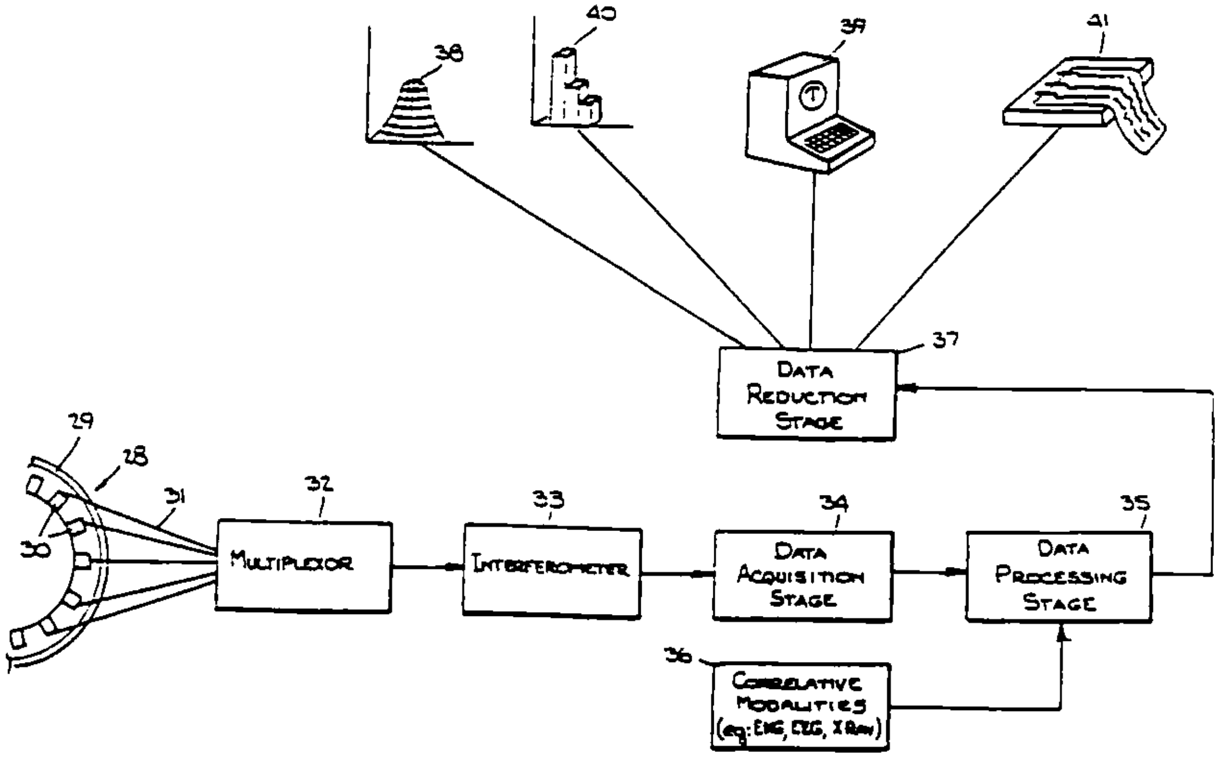

Fig.1.

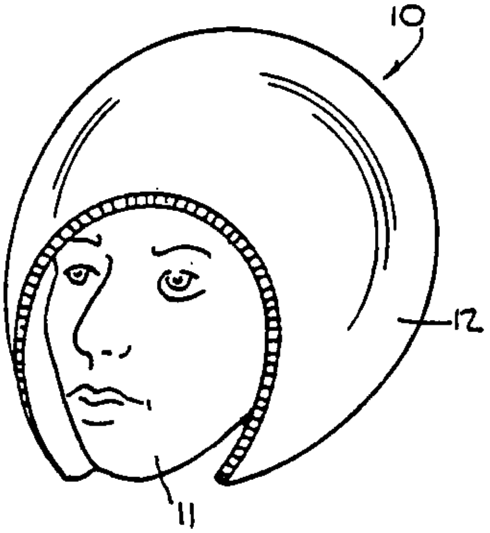

Fig.2.

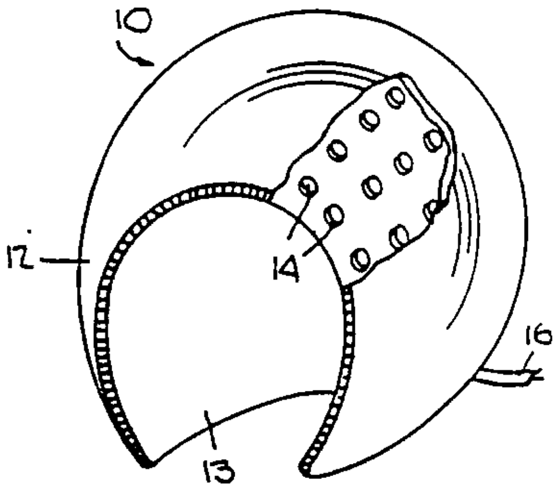
10
12
14
16
13

Fig.3.

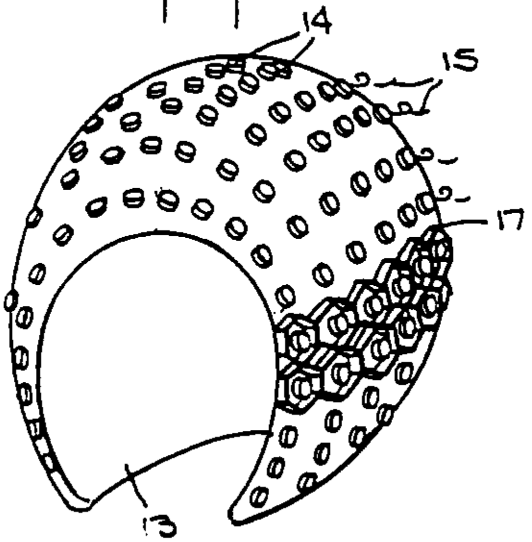

Fig.4.

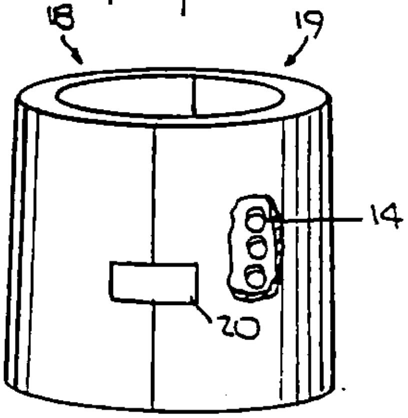

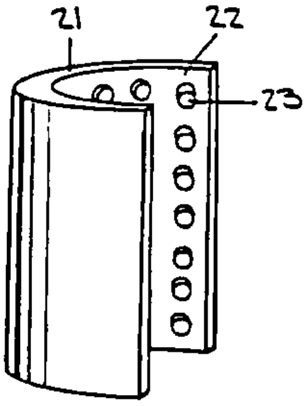

Fig.5.

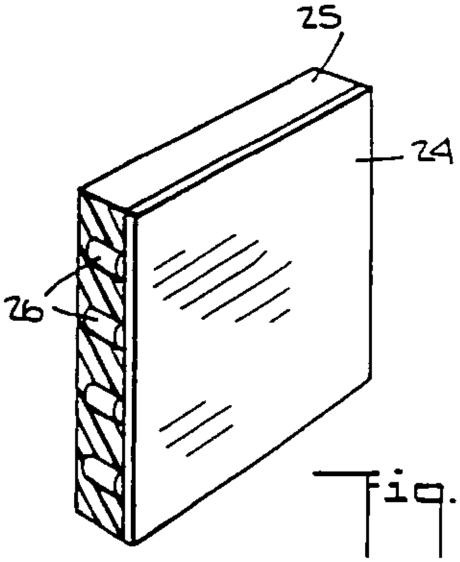

Fig.6.

27

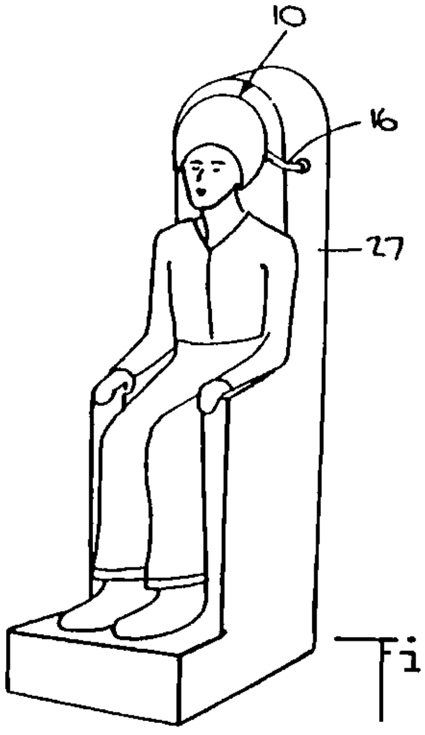

Fig.7.

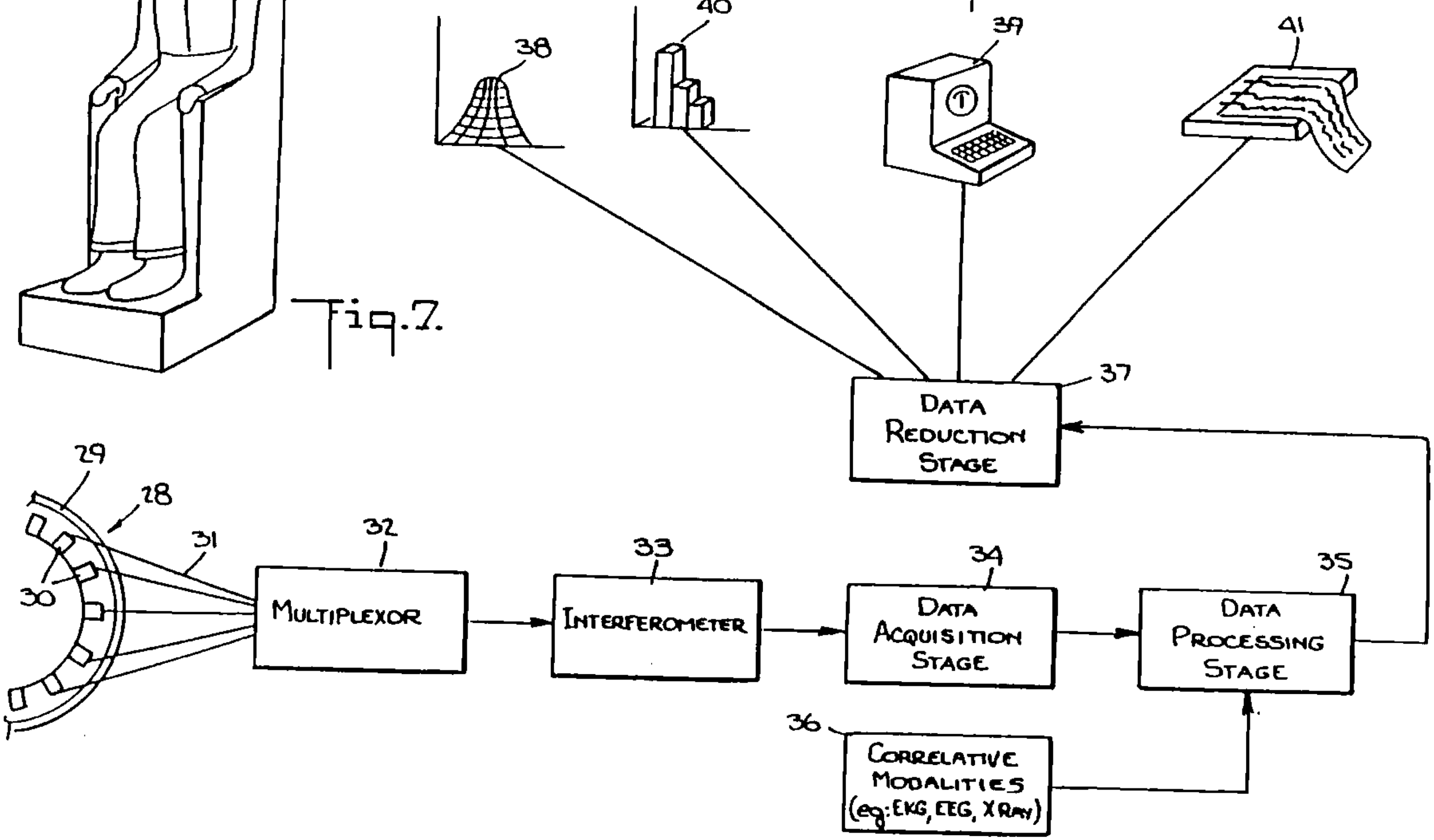
DATA REDUCTION STAGE
MULTIPLEXOR
INTERFEROMETER
DATA ACQUISITION STAGE
DATA PROCESSING STAGE
CORRELATIVE MODALITIES (eg: EKG, EEG, X RAY)

Fig.8.

BIOMAGNETIC ANALYTICAL SYSTEM USING FIBER-OPTIC MAGNETIC SENSORS

BACKGROUND OF INVENTION

1. Field of Invention

This invention relates generally to biomagnetic analytic systems for sensing and indicating minute magnetic fields emanating from the brain and other tissue regions of the human body, and more particularly to a system using fiber-optic magnetic sensor pick-up devices for this purpose.

2. Status of Prior Art

Biomagnetic fields arise from three principal sources, the first being electric currents produced by the movement of ions. The second source is remanent magnetic movement of contaminants, and the third is paramagnetic or diamagnetic constituents of the body.

The first source is of primary significance in human brain activity in which the currents creating the magnetic fields result from signals generated by neurons as they communicate with each other and with sensory organs of the body. The intensity of extracranial magnetic field produced by such currents is extremely minute, having a strength no more than about a billionth of the magnetic field at the earth's surface. It is usually measured in terms of tesla (T) or gauss (G), one T being equal to 10^4 G.

The magnetic field arising from spontaneous brain activity (alpha waves) is about one picotesla (IpT$=10^{-12}$T), whereas the magnetic field at the earth's surface is about 6×10^{-5}T. The magnetic field emanating from the brain has a strength much below that emitted by the heart. Hence monitoring of brain magnetic activity presents formidable difficulties.

A major concern of the present invention is magnetoencephalography (also commonly referred to as MEG). This is the recording of magnetic fields emanating from the brain resulting from neuronal electric currents, as distinguished from an electroencephalogram (EEG) in which electric potentials originating in the brain are recorded. With an EEG measurement, it is difficult to extract the three-dimensional distribution of electrically active brain sites from potentials developed at the scalp. While this difficulty can be overcome by inserting electrodes through apertures bored in the skull, this invasive technique is not feasible in the study of normal brain functions or to diagnose functional brain disorders or brain dysfunctions. Thus ionic currents associated with the production of electrically measurable epileptic seizures generate detectable extracranial magnetic fields, and these can be detected externally without invading the skull.

Non-invasive MEG procedures are currently used in epilepsy research to detect the magnetic field distribution over the surface of the head of a patient with a view to localizing the seizure foci and spread patterns. This analysis serves as a guide to surgical intervention for the control of intractable seizures. (See: "Magnetoencephalography and Epilepsy Research"—Rose et al.; Science—16 Oct. 1987—Volume 238, pp. 329-335.)

MEG procedures have been considered as a means to determine the origin of Parkinson's tremor, to differentiate at the earliest possible stage Alzheimer's disease from other dementias, and to localize the responsible cortical lesions in visual defects of neurological origin. MEG procedures are also of value in classifying active drugs in respect to their effects on specific brain structures, and to in this way predict their pharmaceutical efficacy. And with MEG, one can gain a better understanding of the recovery process in head trauma and strokes by observing the restoration of neurological functions at the affected site.

But while MEG holds great promise in the above-noted clinical and pharmaceutical applications, practical considerations, mainly centered on limitations inherent in magnetic sensors presently available for this purpose, have to a large degree inhibited these applications.

The characteristics of biomagnetic activity that are measurable are the strength of the field, the frequency domain and the nature of the field pattern outside of the body. In magnetoencephalography, measurement of all three of these components are important. Ideally, simultaneous measurement of three orthogonal components of the magnetic field provides a complete description of the field as a function of space and time. Coincident measurement of the magnetic field along the surface of the skull can provide a magnetic field map of the cortical and subcortical magnetic activity. With spontaneous activity, the brain emits magnetic fields of about 10^{-8} to 10^{-9} Gauss, compared with approximately 10^{-6} Gauss emitted by the heart. Thus, monitoring of the brain's magnetic activity places heavy demand upon the required hardware.

In brain activity, the current dipole or source is generated by the current flow associated within a neuron or group of neurons. Volume current is analogous to the extracellular component of the current source. In MEG, the net magnetic field measured depends on the magnetic field generated by the current dipole itself. The contribution from volume conduction is small in which approximations to spherical symmetry are made. However, there are tangential magnetic components originating from secondary sources representing perturbations of the pattern by the volume current at boundaries between regions of different conductivity. Contributions from these secondary sources to the tangential component of the field become relatively more pronounced with distance from the current dipole. But there is no interference from these secondary sources when measurement is confined to the magnetic fields perpendicular to the skull.

In biomagnetic analysis, three types of magnetic sensors are known to have adequate sensitivy and discrimination against ambient noise for this purpose. (See: "Magnetoencephalography"—Sato et al.—Journal of Clinical Neurophysiology—Vol. 2, No. 2—1985.) The first is the induction coil. But because of Nyquist noise associated with the resistance of the windings and its loss of sensitivity at frequencies below a few Herz, the induction coil is rarely used in MEG studies.

The second is the Fluxgate magnetometer; and while this has been used in geophysical studies, it has certain drawbacks when used in MEG applications. It is for this reason that the third type, the SQUID system, is presently used almost exclusively in MEG applications.

A SQUID (Superconducting QUantum Interference Device) comprises a superconducting loop incorporating a "weak link" highly sensitive to the magnetic field encompassed within the area of the loop. While the loop itself can act as a magnetic field sensor, use is made of a detection coil tightly coupled to the superconducting loop, the coil acting as a flux transformer. Both the coil and the loop are immersed in a bath of liquid helium contained within a dewar.

With the advent of so-called high-temperature superconductors operating at liquid nitrogen temperatures, a SQUID magnetometer has been developed using such superconductors. (See: "The Impact of High Temperature Superconductivity on SQUID Magnetometers"—Clarke et al.—Science—Vol. 242—14 Oct. 1988.)

In the booklet published by Biomagnetic Technologies, Inc., of San Diego, Calif., entitled "Introduction to Magnetoencephalography—A New Window on The Brain," there is disclosed a SQUID-type sensor for MEG studies. This SQUID is especially suited to measure magnetic fields in the frequency range from DC to 20 kHz, the magnetic field being converted into a signal that is amplified, filtered and displayed for subsequent analysis.

Because the brain's field falls off sharply with distance from the head, the dewar for the cryogenic liquid, which is inherently bulky, is provided with a tail section of reduced diameter to house the pick-up coil and to minimize the distance of the coil from the head of the patient being studied, thereby maximizing the detected field.

As pointed out in the above-identified booklet, in order to produce a contour map of the brain, the magnetic field must be measured simultaneously at a number of points outside the head. While it is possible with SQUIDS to sample the magnetic field emanating from the brain at one to seven points separated laterally from each other by several centimeters, a complete mapping of the field pattern at a given instant requires forty or more pick-up points. It is proposed, therefore, in the booklet to move SQUID sensors from one point to another to accumulate the required field data. But a measurement taken at a point X will not reveal magnetic brain activity taking place concurrently at a point Y if one has to physically shift the sensor from point X to point Y.

The booklet notes that the ultimate goal of MEG measurement is to simultaneously observe all areas of the brain to produce real-time activity maps responding instantaneously to changes as they occur. However, the booklet concedes that this goal has not yet been realized with SQUID sensors.

The present invention attains this goal by means of fiber-optic magnetometers (FOM). In a FOM sensor, a magnetostrictive alloy is interfaced with an optical fiber to produce a magnetometer whose principle of operation is based on the transference of strain from the magnetostrictive material to the core of the optical fiber via mechanical bonding. This results in modulation of the phase or other parameters of the light propagated in the fiber which is subsequently detected by a fiber-optic interferometer. Integrated fiber-optic magnetometers in which all components are fabricated on or around the optical fibers are now known.

FOM sensors of the type currently available are far less expensive to manufacture and maintain than SQUID sensors; they are considerably more compact, and they operate at room temperature. Their sensitivity to weak magnetic fields, which can be greater than that of a SQUID, renders them suitable for MEG and other applications.

The following publications disclose various forms of FOM sensors:

1. "Single-Mode Fiber-Optic Magnetometer with DC Bias Field Stabilization"—Kersey et al.—Journal of Lightwave Technology—Vol. LT-3, No. 4—August 1985.

2. "Fiber-Optic Polarimetric DC Magnetometer Utilizing a Composite Metallic Glass Resonator"—Mermelstein—Journal of Lightwave Technology, Vol. LT-4, No. 9—September 1986.

3. "Optical Fiber Sensors Using The Method of Polarization-Rotated Reflection"—Enokihara et al.—Journal of Lightwave Technology—Vol. LT-5—No. 11—November 1987.

4. "An Analysis of A Fiber-Optic Magnetometer with Magnetic Feedback"—Koo et al.—Journal of Lightwave Technology—Vol. LT-5—No. 12—December 1987.

The disclosures of these publications are incorporated herein by reference.

SUMMARY OF INVENTION

In view of the foregoing, the main object of this invention is to provide a biomagnetic analytical system which includes a pick-up device employing an array of fiber-optic magnetic (FOM) sensors for measuring and indicating minute magnetic fields emanating from a multitude of sites in the brain or in other tissue regions of interest in a subject being diagnosed.

A significant advantage of a fiber-optic magnetic sensor (FOM) over a SQUID is that the former is a solid-state device that is considerably smaller than the latter and requires no cryogenics, thereby making it possible to distribute a multitude of the sensors (i.e., in excess of forty) around the skull of the patient or about any other tissue region of interest to effect more accurate localization of magnetic activity, as well as a more precise determination of the physiological condition of the region being studied.

More particularly, an object of this invention is to provide a system of the above type for MEG analysis in which the FOM sensors are so distributed in a three-dimensional array as to pick up magnetic fields emanating from a multitude of brain sites simultaneously and to spatially localize the field signals.

Also an object of the invention is to provide a shielded magnetic pick-up device in which the FOM sensors in the array are magnetically shielded from each other to prevent magnetic interaction therebetween, as well as from magnetic fields extraneous to the region of interest, thereby obviating the need for a shielded room to conduct studies on biomagnetic activity.

Yet another object of the invention is to provide a biomagnetic system in which the outputs of the FOM sensors in the array are multiplexed, whereby a common interferometer can be used for the multitude of sensors in the array thereof.

Briefly stated, these objects are attained in a biomagnetic analytical system for sensing and indicating minute magnetic fields emanating from the brain or from any other tissue region of interest in a subject under study. The system includes a magnetic pick-up device constituted by an array of fiber-optic magnetic sensors mounted at positions distributed throughout the inner confines of a magnetic shield configured to conform generally to the head of the subject or whatever other body region is of interest.

Each sensor yields a light beam whose phase or other parameter is modulated in accordance with the magnetic field emanating from the related site in the region. The modulated beam from each sensor is compared in an interferometer with a reference light beam to yield an output signal that is a function of the magnetic field being emitted at the related site. The output signals from the interferometer are processed to provide a display or recording exhibiting the pattern or map of magnetic fields resulting from emanations at the multitude of sites encompassed by the region.

BRIEF DESCRIPTION OF DRAWINGS

For a better understanding of the invention as well as other objects and further features thereof, reference is made to the following detailed description to be read in conjunction with the accompanying drawings, wherein:

FIG. 1 is a perspective view of a magnetic field pick-up device in the form of a helmet which is fitted over the head of a patient and which incorporates an array of FOM sensors for simultaneously detecting magnetic fields emanating from a multitude of sites in the brain;

FIG. 2 shows the helmet partially cut away to expose an inner insulating liner on which the sensors are mounted;

FIG. 3 is a separate view of the inner liner, illustrating the manner in which the FOM sensors are shielded from each other;

FIG. 4 illustrates, in perspective, a cylindrical magnetic pick-up device;

FIG. 5 shows a semi-cylindrical section of the pick-up device;

FIG. 6 shows a flat pick-up magnetic device;

FIG. 7 illustrates a unit for accommodating a patient undergoing a magnetoencepahlographic examination; and

FIG. 8 illustrates schematically a biomagnetic analytical system in accordance with the invention operating in conjunction with a pick-up device that is appropriate to the region being studied.

DETAILED DESCRIPTION OF INVENTION

FOM Pick-Up Devices

Referring now to FIGS. 1 to 3, there is shown one preferred embodiment of a pick-up device 10 in accordance with the invention, adapted to detect magnetic fields emanating from a multitude of sites on the brain of a patient for purposes of magnetoencelographic (MEG) examination.

Pick-up device 10 includes a generally spherical helmet 12 formed of ferromagnetic or superconductive shielding material, the helmet being configured to generally conform to the head of a patient 11 so that the brain therein lies within the confines of the helmet, and the weak magnetic fields emanating from the brain are confined within the helmet which acts to exclude extraneous magnetic fields, including those emanating from external electronic equipment associated with the pick-up device.

Helmet 12 is provided with a conforming inner liner 13 formed of electrical insulating material, such as in synthetic plastic material or an epoxy compound having good dielectric properties. Embedded in liner 13 or otherwise mounted thereon is a three-dimensional array of identical FOM sensors 14, each provided with a fiber-optic light conducting line 15 to supply light from a suitable laser beam source to the sensor and to conduct the light modulated by the sensor in response to the magnetic field detected thereby to an external interferometer. Lines 15 are bundled to form a cable 16 running from the pick-up device to external signal processing apparatus.

The FOM sensors 14 in the three-dimensional array are distributed uniformly throughout the inner confines of helmet 12, so that each sensor acts to pick up a unique magnetic field emanating from the head.

As shown in FIG. 3, FOM sensors 14 are internally shielded from each other but not from the magnetic fields emanating from the brain by an open cell ferromagnetic honeycomb 17 so that there is no magnetic interaction between the sensors.

Fiber-optic magnetometers are well known, as evidenced by references (1) to (4), supra. The operation of FOM sensor is based on the transference of strain from the magnetostrictive material in response to a magnetic field to the core of the optical fiber via mechanical bonding, resulting in a phase modification of the propagated light beam. The modulated light from the sensor is subsequently processed in an interferometer which may take the form of a photodetector which compares the phase-modulated light beam with a reference light beam to provide an output signal that is a function of the phase displacement caused by the magnetic field to which the sensor is exposed.

FOM sensors are available in various configurations. In one such configuration, the optical fiber is bonded to a magnetostrictive element to form a waveguide strip that is then coiled into a spool so that the entire sensor is very small. For alternating-current measurements using, for example, metallic glasses as the sensing material, magnetic sensitivities of the order of 10^{-9} G/m of fiber core are obtainable.

Design improvements such as magnetic feedback nulling (Reference 4) can lead to improvements in the linear dynamic range, high suppression of magnetic hysteresis associated with the magnetic material, and improved long-term stability. Methods of improving fluctuations of the transmission characteristics of the fiber (induced by the surrounding environment) exist such as use of phase-sensitive transducers combined with a single polarization-maintaining fiber. Thus polarization-rotated reflection can be used to enhance the performance of the system.

Of the many configurations the magnetometer can take, all detect an externally-induced optical phase shift. The sensitivity of the system is proportional to the length of the fiber, until such time as the length approaches the point where other optical properties of the fiber interfere with the propagation of light. The measurement of the linear strain in length of the fiber which is bonded to (or coated by) the magnetostrictive material forms part of one arm of a fiber interferometer. Several types of interferometer designs have been employed, among which are the Fabra Perot, Mach Zender, Michelson and Sagnac types. Well designed interferometers can detect induced optical phase shifts below $10^{31\,6}$ rad over the frequency range of 10–10^{4} Hz. Thus, very weak magnetic fields per meter of fiber can be detected.

The sensor configuration can be a fiber bonded to a magnetostrictive tube or mandrel, a metal film deposited on the fiber, a metallic glass strip bonded to fibers or a metallic glass cylinder. The magnetic materials which are sensitive in the range of DC to 50 kHz are nickel, iron-nickel alloys, cobalt-nickel, and metallic glasses. Piezoelectric activity in the jacket of the fiber can be achieved by the use of various types of polymer films.

Sensing of small AC magnetic fields at optimal DC
etic fields for various frequencies can occur.
easuring technique enables the separation of
ic effect (at relatively higher frequencies)

31

from environmental effects (such as temperature or acoustics at lower frequencies) on the fiber interferometer. Thus, the fiber interferometer can be stabilized without losing sensitivity to magnetic fields.

To extend the AC measuring technique to measure DC bias magnetic fields, one can make use of the effect of DC bias magnetic fields to change the interferometer output due to a fixed AC magnetic field drive at a given frequency. In a sense, this technique utilizes an AC approach to measure DC magnetic fields, thereby overcoming both the environmental perturbation and the 1/f noise problem usually associated with low frequency measurements. It is the nonlinear response of the magnetostrictive material that allows the utilization of an AC technique to measure DC magnetic fields.

In the helmet-type pick-up device shown in FIGS. 1 to 3 which is adapted for MEG studies, each FOM sensor 14 in the three-dimensional array thereof is oriented so that the longitudinal axis of the optical waveguide coil or spool is substantially perpendicular to the surface of the skull of patient 11 wearing the helmet. This orientation makes it possible to dispose a multitude of sensors (forty or more) at positions distributed uniformly about the skull. The density of the coils in the array is limited by factors such as the physical diameter of each spool and induced noise from nearby spools such as eddy currents.

In biomagnetic measurements, sensitivity is limited by fluctuations in the ambient fields and not the intrinsic noise of the sensor. Such ambient fields are produced primarily by the sensor itself, motorized machinery and metallic structural components of buildings which distort the earth's geomagnetic field. The earth's geomagnetic field is uniform and steady. The problem arises when a sensing system vibrates, often in the 1–10 Hz range. Also, the subject may produce noise from normal physiological activity. In MEG, the head proper is the source of a significant amount of noise, produced primarily by the cortex. But because the array of sensors lies within a shield and each sensor occupies a position within a cell in a honeycomb shield, the sensors are isolated from ambient noise and magnetic interaction therebetween is prevented.

Each sensor acts as a gradiometer of predetermined order, any three of which can be used to localize biomagnetic sources at any brain site by the use of the computation techniques described hereinafter.

The pick-up device shown in FIGS. 1 to 3 is adapted to be placed over the head of a patient for MEG analysis. But in practice, the pick-up device can be customized to pick up magnetic field activity arising in other body regions of interest. Thus in measuring magnetic fields generated by the heart, the appropriate pick-up device, as shown in FIGS. 4 and 5, is in a cylindrical form composed of a pair of complementary semi-cylindrical sections 18 and 19 which are joined together by Velcro fasteners 20 or similar means which makes it possible to detach the sections from each other. In practice, for heart analysis, the cylinder is positioned around the thorax.

The cylinder is constituted by an outer shell 21 of shielding material having an inner liner 22 of insulating material in which are embedded a cylindrical array of FOM sensors 23. This cylindrical array of sensors surrounds the thorax, each sensor picking up the magnetic field emanating from a respective site in the heart.

The cylindrical pick-up device can also be configured to go around a limb to measure magnetic fields generated by the muscles therein. For other parts of the body, a flat pick-up device may be appropriate. This flat device, as shown in FIG. 6, is provided with an outer metal plate 24 of shielding material laminated to an inner block 25 of electrical insulating material in which a rectangular array of FOM sensors 26 is embedded.

In practice, the cylindrical pick-up device may be applied to any extremity; and a semi-cylindrical pick-up section can be used separately for measurement purposes. The flat pick-up device is useful for small surface measurements.

In practice, the pick-up device may be contoured to conform to any body region of interest. Thus the pick-up device can be used in a broad range of medical applications by making available to the practitioner a family of pick-up devices, each customized for a particular part of the body. The pick-up device is strapped or otherwise attached to the body part when a biomagnetic study is to be conducted.

For MEG procedures, as shown in FIG. 7, the helmet-type pick-up device 10 may be included as the headpiece of a support unit 27 which also provides a seat, a back rest and a foot platform for the patient whose head is received by the helmet. Housed in the unit is electronic equipment for processing the outputs of the FOM array contained in the helmet.

If one wishes to conduct biomagnetic studies on all portions of the body with a single pick-up device, the pick-up device for this purpose (not shown) may take the form of a sarcophagus-like magnetically shielded enclosure with suitable breathing vents, within which enclosure are disposed arrays of FOM sensors to pick up magnetic fields from different regions of the body. In practice, various standard electronic techniques can be used for noise reduction, either separately or in combination with physical shielding.

The Biomagnetic Analytical System

As shown in FIG. 8, one preferred embodiment of a biomagnetic analytical system in accordance with the invention includes a magnetic field pick-up device 28 in a configuration appropriate to the region under study. Device 28 includes an outer shield 29 within which is an array of FOM sensors 30, each coupled by a fiber-optic line 31 to the input of a multiplexer 32 whose output is applied to an interferometer 33.

Thus instead of having a separate interferometer channel for each FOM sensor which would be very costly, a common interferometer now serves sequentially to compare the modulated light beam derived from each sensor in the array with a reference light beam to produce an output signal. This output signal is a function of the sensed magnetic field emanating from the site related to the sensor.

In practice, multiplex transmission of the signals from the interferometer may be either optical or electrical; that is, by optical waveguides or by conductive wires in the case where the interferometer is of the type which applies the modulated light signal to a photodetector to be compared with the reference light beam.

The output signals from interferometer 33 representing the sensed magnetic fields is applied to a data acquisition stage 34.

The output of interferometer 33 is phase information related to the degree of phase rotation resulting from the detection of the incident vector of the magnetic field transient on the fiber optic sensor. This phase information can evolve over time as the biomagnetic

32

signal progresses, and it is of an analog nature. Therefore, it can be acquired, recorded, and processed in either the analog or digital form. For purposes of the present embodiment, reducing the data to, and processing the data in digital form will be discussed. However, the specific means of data handling is not important to the invention.

Time-varying phase information can be represented in the form of an evolving analog voltage which is proportional to the phase shift by a predetermined relationship. This signal can be converted to digital form through a conventional high speed sample-and-hold and analog-to-digital converter (ADC) means whereby the resulting bandwidth of the recorded signal can be limited to that which is determined by the sampling rate of this analog signal according to the Nyquist sampling theorem.

Since the signal from each sensor on the sensor array is time-multiplexed before entering data acquisition stage 34, a single ADC system is sufficient to digitize the signals from the entire sensor array. However, it should be noted that due to the signal bandwidth requirements and the size of the sensor array, it might be possible that the amount of data generated may result in a data rate which is high enough that the conversion rate of a single ADC might be exceeded and thus, a single ADC channel would not suffice. In this case, the multiplexed signal must then be de-multiplexed, the data from each sensor within the array then being directed to a respective ADC channel, there being one ADC channel per sensor, for example.

In practice, several electronic methods may be used to reduce noise, including the use of comb filters, subtracting a reference channel from a signal channel, adaptive filtering of analog or digital format or balancing of detectors. These can quite effectively reduce the ambient noise, even to the point where physical shielding may not be necessary.

The system lends itself to any number of data acquisition and processing means. For example, since the phase information in a magnetoencephalographic signal can be presented in optical form, the data is particularly suited to processing by optical computer technology.

The output of data acquisition stage 34 is applied to a data processing stage 35. Through the applications of standard spatial localization algorithms, such as are addressed by point source theory, a three-dimensional histogram map can be generated to represent the location and relative magnitude of each magnetic dipole source. The processing of the data will depend upon the configuration of the sensor array, since the proper 3-D mapping of the magnetic dipoles mandates a predetermined geometry of the plane of the array as well as the spacing between each sensor element within the array.

Once the data is acquired in digital form, it can be immediately processed by conventional digital computer means for the purposes of relating the data in various formats including those clinically relevant such as (a) time-varying analog and bandwidth information for each channel, and (b) the spatial localization, through the application of point source theory, of the magnetic dipoles which generated the magnetic field transients that were detected by the sensor array. The data can also be directed to a storage medium for the purpose of recording the digitized biomagnetic data for archiving and later retrieval and processing.

Additionally, other correlative modalities may be employed in conjunction with the biomagnetic data in order to obtain more complete information about a particular body tissue or system. Thus, modalities such as EEG, EKG, MRI and X-ray from a source 36 can be combined with the biomagnetic data in the data processing stage 35 and represented according to the particular need of the practitioner.

The output of data processing stage 35 is applied to a data reduction stage 37. The computational techniques necessary for resolution of the image and decoding it into a meaningful "image" for the clinician to interpret requires Fourier analysis of the affected light similar to the analyses utilized by the magnetic resonance imaging (MRI) technology. The FOM-based system in accordance with the invention allows for the detection of a large number of points simultaneously. This feature is most important, since clinicians have long sought a method of actually visualizing the physiological processes of the intact brain (especially deep brain structures) in various states of sleep and consciousness.

Two major problems encountered by the conventional SQUID-based system for MEG is that the resolution is not very good, and that the number of sensors that can be employed is quite limited, thereby limiting the sites in the brain that can be "scanned" for possible pathologies. These two cumbersome problems are overcome with the FOM-based system, making possible many more clinical applications.

With regard to the actual diagnosis of pathologies, the system is targeted toward the same professionals for whom the conventional SQUID-based system is designed. Primarily, these are radiologists, neurologists and surgeons. For example, the practitioner could "see" a functional image of the brain on a CRT, but the image will be a profile of the electromagnetically "active" portions of the brain, as shown by the magnetic pattern 38 derived from data reduction stage 37. The display plane can represent the location of each magnetic dipole source, and one pixel intensity can represent the magnetic dipole magnitude or phase with respect to a reference signal. What these active areas mean, and how they relate to pathologies is currently an area of intense interest in the basic and clinical neurosciences. Such areas of activity might signify anything from a soft-tissue pathology to abnormal behavioral patterns where no morphological or biochemical anomaly can be detected.

The format of data presentation can take the form of a static or time-varying uni- or multi-dimensional display.

Uni-dimensional display conforms to the current standards of SQUID-based magnetometry and of clinical electrophysiology. In this case, data is represented as linear time-varying representations of single detector elements. Multi-dimensional analysis can provide two- or three-dimensional graphic representations of the data. These multi-dimensional constructs are the result of applying the point source theory algorithms referred to above and can represent the dipole data spatially, or in terms of other dimensionalities, such as k-space in the Fourier domain.

Thus, the three-dimensional histogram map which can be representative of the location and relative magnitude of each magnetic dipole source can be rendered on a standard graphic workstation 39. Application-specific labelling must be applied to convey orientation and scaling of the 3-D data matrix. The display may take the graphical form 40 in which the level of each sensed field derived from the FOMs is represented by a column. Or the levels may be separately indicated on a record chart 41.

Advantages

The main advantages of the system are as follows:

I. Each FOM-based sensor is far smaller than a SQUID-based sensor. Thus, a multitude of FCM sensors can be placed over or around the region of interest by the magnetic pick-up device.

II. A system of gradiometers of predetermined order allows for a more accurate localizatior of biomagnetic activity, and a precise determination of the physiological condition.

III. The system affords more freedom to investigate several areas of the tissue region of interest simultaneously, since it provides a means to spatially localize the field signals.

IV. The biomagnetic analyzing system does not use expensive cryogens. The FOM-based pick-up device entails minimal installation time and service requirements.

V. The FOM sensor is relatively easy and inexpensive to construct, since only "solid-state" materials are required.

VI. The magnetic shielding needed for the FOM-based biomagnetic monitoring is limited to the area surrounding the region of interest, thereby eliminating the need for a specially shielded room.

VII. Even if a shielded room or enclosure is preferred for added noise reduction, the enclosure need not be larger than that required to placed the patient in comfortably.

VIII. Multiplexing the output from the sensing array makes it possible for more sensors to be monitored with fewer "decoding" devices at the "back end."

IX. The FOM-based pick-up device produces optical data that lends itself to optical computing available through the field of "photonics" involving optical processing technologies which can drastically decrease data processing times without compromising resolution.

While there has been shown and described a biomagnetic analytical system using various embodiments of FOM pick-up devices in accordance with the invention, it will be appreciated that many changes and modifications may be made therein, without, however, departing from the essential spirit thereof.

We claim:

1. A biomagnetic analytical system for sensing and indicating minute magnetic fields emanating from the brain or any other tissue region of interest in a subject being diagnosed, said system comprising:

(a) a magnetic pick-up device having an outer shell contoured to conform generally to the region of interest, said shell being formed of magnetic shielding material to exclude from its inner confines extraneous magnetic fields, whereby the emitted magnetic fields exist within the confines of the shell, and an array of fiber-optic magnetometer sensors which conforms to the contours of the shell, the sensors being mounted within the shell at positions distributed throughout the inner confines thereof, whereby each sensor is related to a site in the region and yields a light beam modulated in accordance with the magnetic field emanating from this site; and

(b) means including an interferometer to compare the modulated light beam yielded by each sensor in the array with a reference light beam to produce an output signal that is a function of the magnetic field emitted at the related site.

2. A system as set forth in claim 1, wherein said sensors are each disposed within a respective cell of a honeycomb shield supported within the shell to prevent magnetic interaction between the sensors.

3. A system as set forth in claim 1, wherein said sensors in the array are supported on an electrically insulating inner liner conforming to the inner contours of the shell.

4. A system as set forth in claim 1, wherein said shell is configured as a helmet to be worn by the subject for brain magnetic field diagnosis.

5. A system as set forth in claim 1, wherein said shell is configured as a cylinder for heart magnetic field diagnosis.

6. A system as set forth in claim 1, further including means coupled to the output of said interferometer to process the output signals from the interferometer to provide a display exhibiting the pattern of magnetic fields emanating from the sites encompassed by the region.

7. A system as set forth in claim 1, wherein said array has at least forty sensors.

8. A system as set forth in claim 1, wherein the modulated light beams from the sensors are applied sequentially to the interferometer through a multiplexer.

* * * * *

United States Patent [19] [11] **4,408,616**

Duffy et al. [45] **Oct. 11, 1983**

[54] BRAIN ELECTRICAL ACTIVITY MAPPING

[75] Inventors: **Frank H. Duffy**, Brookline, Mass.; **Norman D. Culver**, Spotswood, N.J.

[73] Assignee: **The Children's Medical Center Corporation**, Boston, Mass.

[21] Appl. No.: **264,043**

[22] Filed: **May 15, 1981**

[51] **Int. Cl.**³ **A61B 5/04**
[52] **U.S. Cl.** **128/731**
[58] **Field of Search** 128/731-733, 128/905

[56] References Cited

U.S. PATENT DOCUMENTS

2,928,189	3/1960	Molner et al.	35/22
3,696,808	10/1972	Roy et al.	128/2.1 B
3,705,297	12/1972	John	235/150.53
3,706,308	12/1972	John et al.	128/2.06 R
3,707,147	12/1972	Sellers	128/2.06 G
3,717,141	2/1973	Krohn et al.	128/2.66 R
3,780,724	12/1973	John	128/2.1 B
3,799,146	3/1974	John et al.	128/2.1 B
3,837,331	9/1974	Ross	128/732 X
3,901,215	8/1975	John	128/2.1 B
3,958,563	5/1976	Fernandez et al.	128/731
4,094,307	6/1978	Young	128/2.1 B
4,171,696	10/1979	John	128/731
4,201,224	5/1980	John	128/731
4,214,591	7/1980	Sato et al.	128/731

OTHER PUBLICATIONS

Duffy et al.; *Significance Probability Mapping: An Aid in the Topographic Analysis of Brain Electrical Activity;* EEG and Clin. and Neurophysiology, 1981; pp. 1-8.

Duffy et al.; "Quantification of Focal Abnormalities in Beam Data by Grid Sector Analysis"..

Duffy et al.; "Dyslexia: Automated Diagnosis of Computerized Classif. of Brain Electrical Activity"; *Annals of Neur.*, vol. 7, No. 5, 5-1980, pp. 421-428.

Duffy et al.; "Dyslexia: Regional Diff. in Brain Electrical Activity by Topographic Mapping"; *Annals of Neur.*, vol. 7, No. 5, 5-1980, pp. 412-420.

Ueno et al., Topographic Computer Display of Abnormal EEG Activities in Patients with CNS Diseases, Memoirs of the Faculty of Engineering, Kyushu University, vol. 34, No. 3, (Feb., 1975) pp. 195-209.

Marguerite Zientara (CW Staff) Multiple Personalities 'Mapped' by Computer, Computer World Publication, Jan. 24, 1983, p. 14.

Duffy et al., "Brain Electrical Activity Mapping (BEAM): A Method for Extending the Clinical Utility of EEG and Evoked Potential Data," *Annals of Neurology*, vol. 5, No. 4 (Apr., 1979) pp. 309-321.

Primary Examiner—Lee S. Cohen
Assistant Examiner—Angela D. Sykes

[57] ABSTRACT

Topographic displays of brain electrical activity are produced from matrices of data derived from evoked potential (EP) and steady-state responses of skull transducers. In different aspects, EP responses are displayed at a variable frame rate, the rate of data sampling is sufficient to capture rapid transient events, difference matrices are derived as the difference between matrices corresponding to two different brain conditions, the baseline of the EP responses is zeroed based on the average prestimulus response, and the steady-state response is analyzed by Fourier transforms. In other aspects, statistical comparison matrices representing statistical differences between corresponding elements in two matrices are generated, a coefficient-of-variance matrix is generated, additional display matrices are temporally interpolated, response waveforms are previewed and tagged for elimination from further processing, the topographic maps are displayed on a video monitor with appropriate scaling of the data to the tones of the display, and additional display points are interpolated between the measured data points for display.

82 Claims, 32 Drawing Figures

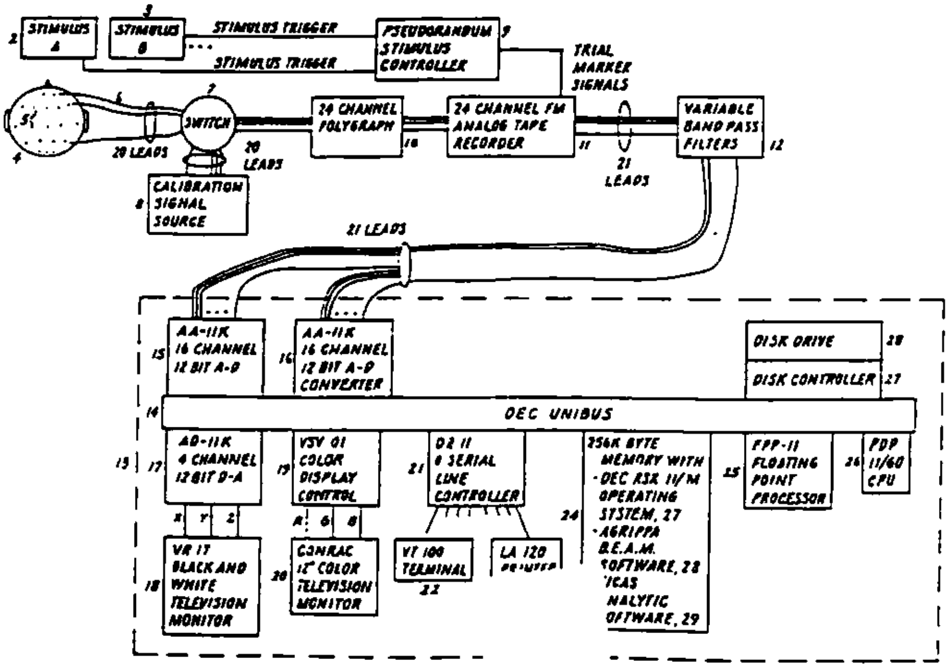
STIMULUS A
2
3
STIMULUS B
STIMULUS TRIGGER
STIMULUS TRIGGER
PSEUDORANDOM STIMULUS CONTROLLER
9
TRIAL MARKER SIGNALS
5
6
4
SWITCH
7
20 LEADS
20 LEADS
CALIBRATION SIGNAL SOURCE
8
24 CHANNEL POLYGRAPH
10
24 CHANNEL FM ANALOG TAPE RECORDER
11
21 LEADS
VARIABLE BAND PASS FILTERS
12
21 LEADS
AA-11K 16 CHANNEL 12 BIT A-D
15
AA-11K 16 CHANNEL 12 BIT A-D CONVERTER
16
DISK DRIVE
28
DISK CONTROLLER
27
14
DEC UNIBUS
13
AD-11K 4 CHANNEL 12 BIT D-A
17
X
Y
Z
VSV 01 COLOR DISPLAY CONTROL
19
R
G
B
DZ 11 8 SERIAL LINE CONTROLLER
21
256K BYTE MEMORY WITH
-DEC RSX 11/M OPERATING SYSTEM, 27
-AGRIPPA B.E.A.M. SOFTWARE, 28
-ICAS ANALYTIC SOFTWARE, 29
24
FPP-11 FLOATING POINT PROCESSOR
25
PDP 11/60 CPU
26
VR 17 BLACK AND WHITE TELEVISION MONITOR
18
CONRAC 12" COLOR TELEVISION MONITOR
20
VT 100 TERMINAL
22
LA 120 PRINTER

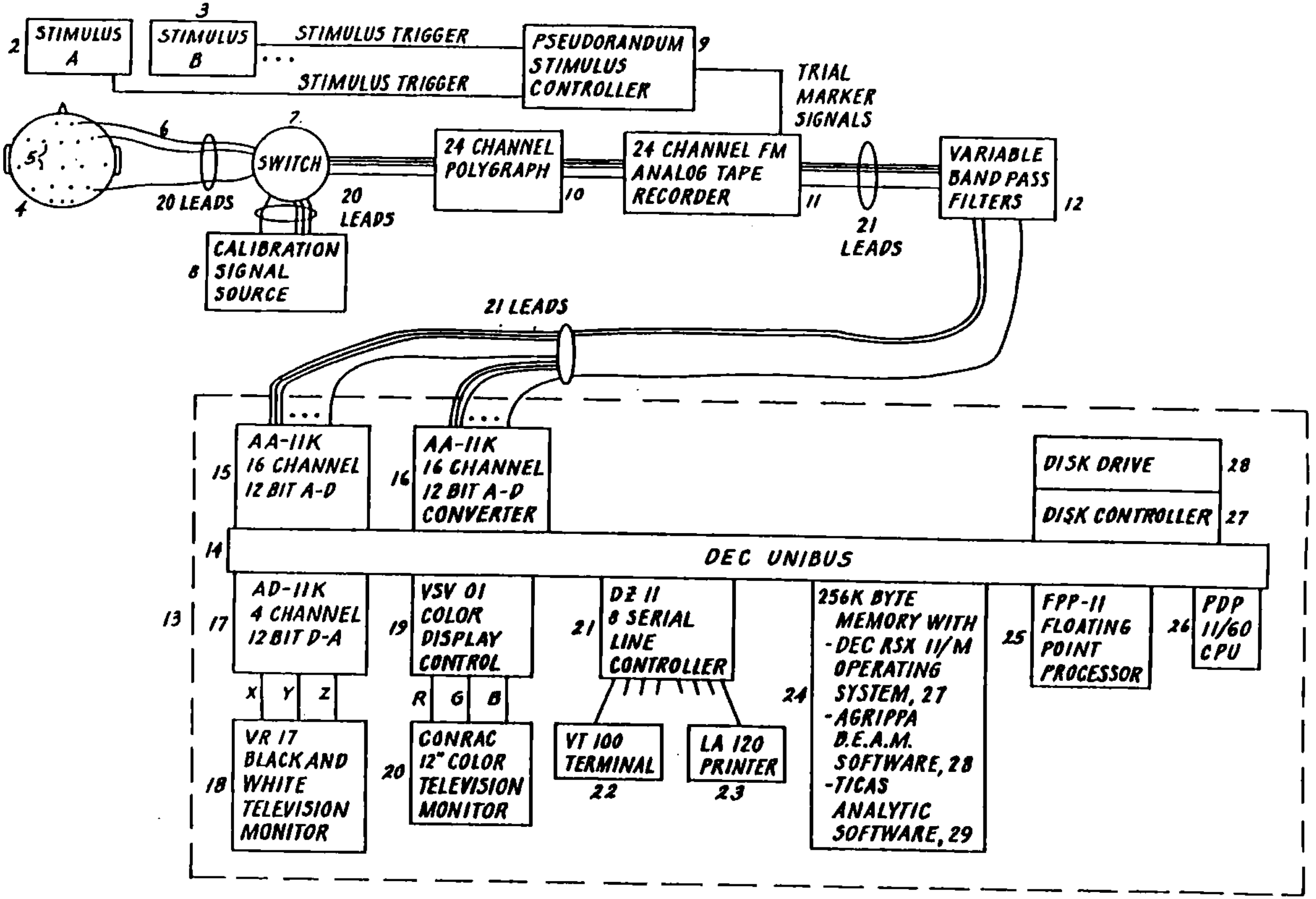
2
STIMULUS A
3
STIMULUS B
STIMULUS TRIGGER
STIMULUS TRIGGER
PSEUDORANDUM STIMULUS CONTROLLER
9
TRIAL MARKER SIGNALS
4
5
6
20 LEADS
7
SWITCH
20 LEADS
8
CALIBRATION SIGNAL SOURCE
24 CHANNEL POLYGRAPH
10
24 CHANNEL FM ANALOG TAPE RECORDER
11
21 LEADS
VARIABLE BAND PASS FILTERS
12
21 LEADS
13
14
15
AA-11K 16 CHANNEL 12 BIT A-D
16
AA-11K 16 CHANNEL 12 BIT A-D CONVERTER
DISK DRIVE
28
DISK CONTROLLER
27
DEC UNIBUS
17
AD-11K 4 CHANNEL 12 BIT D-A
X
Y
Z
18
VR 17 BLACK AND WHITE TELEVISION MONITOR
19
VSV 01 COLOR DISPLAY CONTROL
R
G
B
20
CONRAC 12" COLOR TELEVISION MONITOR
21
DZ 11 8 SERIAL LINE CONTROLLER
VT 100 TERMINAL
22
LA 120 PRINTER
23
24
256K BYTE MEMORY WITH -DEC RSX 11/M OPERATING SYSTEM, 27 -AGRIPPA B.E.A.M. SOFTWARE, 28 -TICAS ANALYTIC SOFTWARE, 29
25
FPP-11 FLOATING POINT PROCESSOR
26
PDP 11/60 CPU

Fig 1

Fig 2

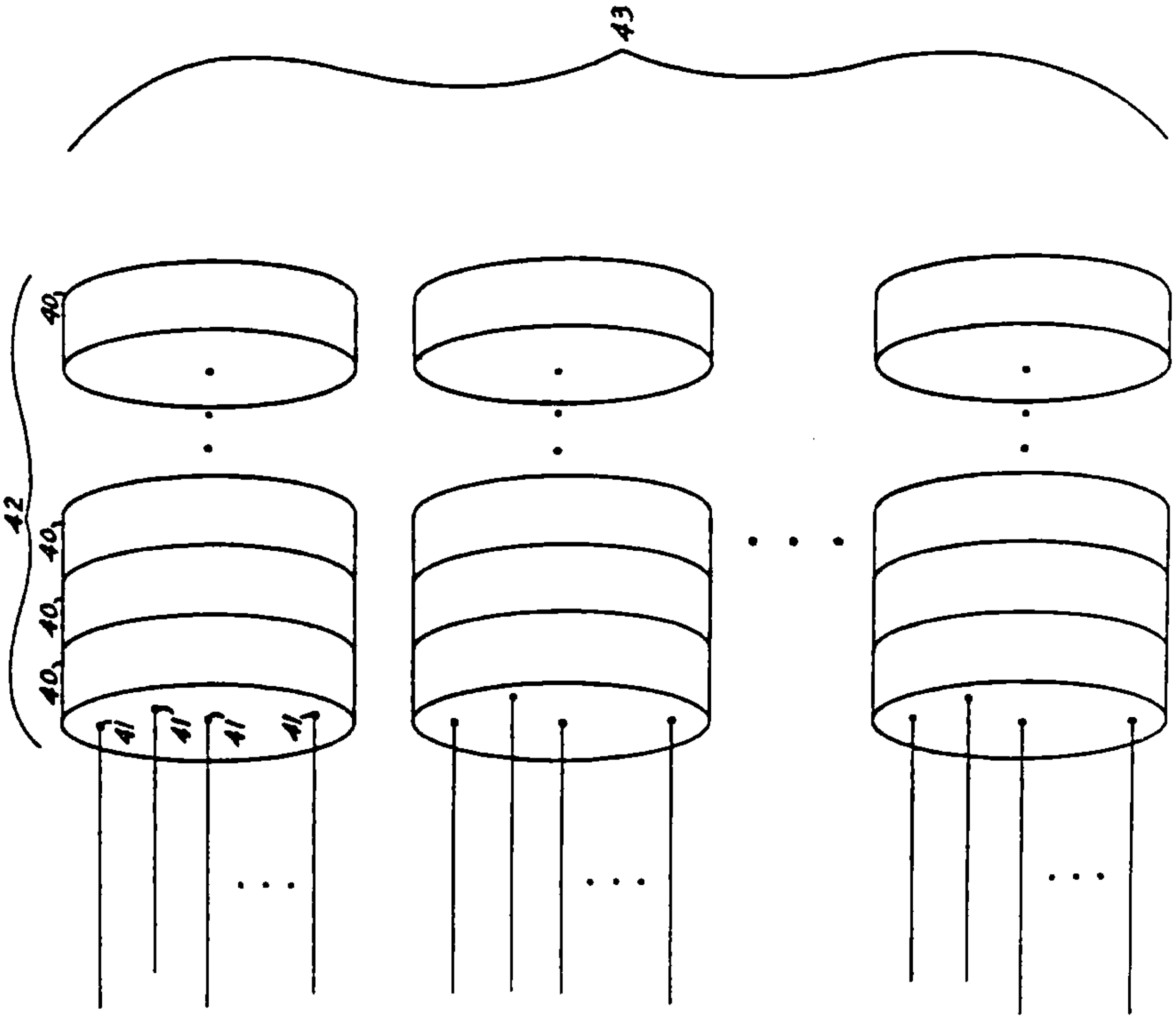
43
42
40
40
40
40
41
41
41
41

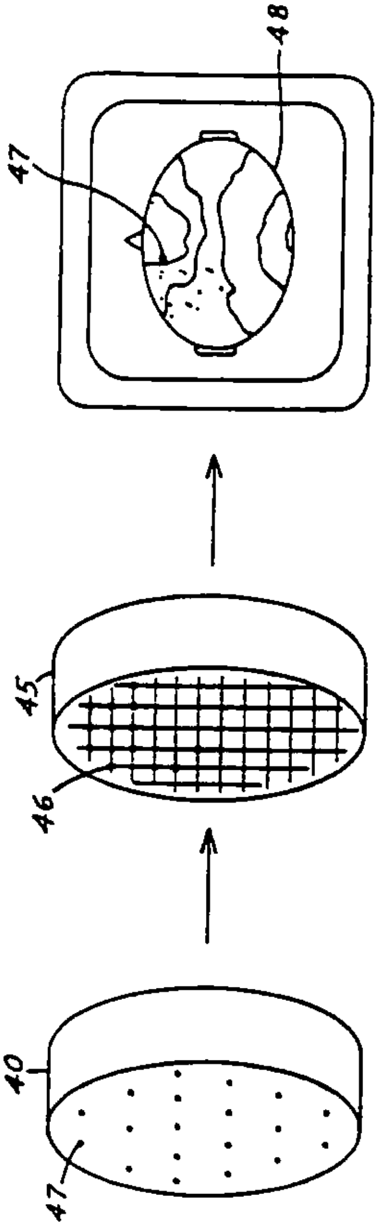
47
48
45
46
40
47

Fig 3

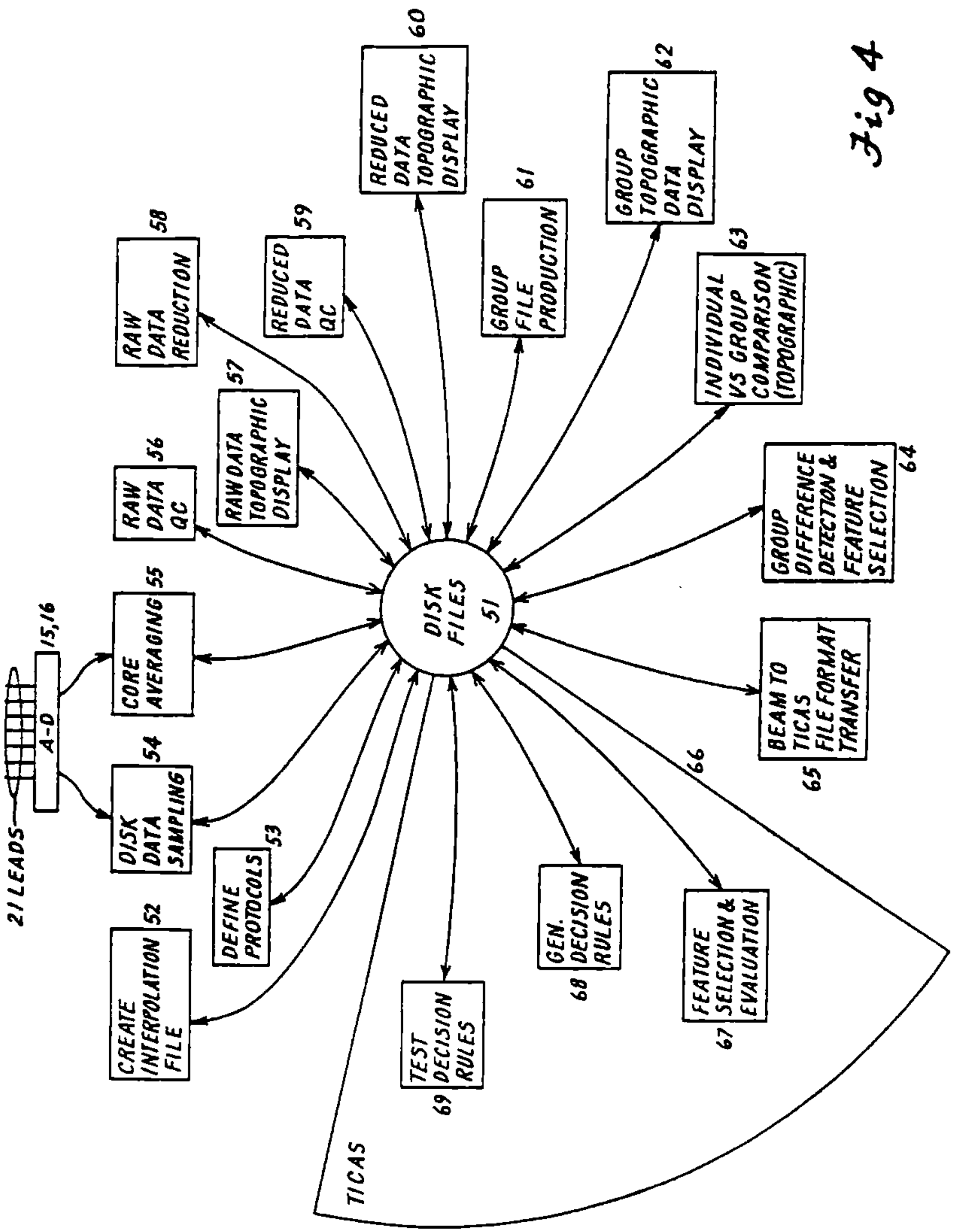
21 LEADS
A-D
15,16
CORE AVERAGING
55
DISK DATA SAMPLING
54
DEFINE PROTOCOLS
53
CREATE INTERPOLATION FILE
52
RAW DATA QC
56
RAWDATA TOPOGRAPHIC DISPLAY
57
RAW DATA REDUCTION
58
REDUCED DATA QC
59
REDUCED DATA TOPOGRAPHIC DISPLAY
60
GROUP FILE PRODUCTION
61
GROUP TOPOGRAPHIC DATA DISPLAY
62
INDIVIDUAL VS GROUP COMPARISON (TOPOGRAPHIC)
63
GROUP DIFFERENCE DETECTION & FEATURE SELECTION
64
BEAM TO TICAS FILE FORMAT TRANSFER
65
DISK FILES
51
66
FEATURE SELECTION & EVALUATION
67
GEN. DECISION RULES
68
TEST DECISION RULES
69
TICAS

Fig 4

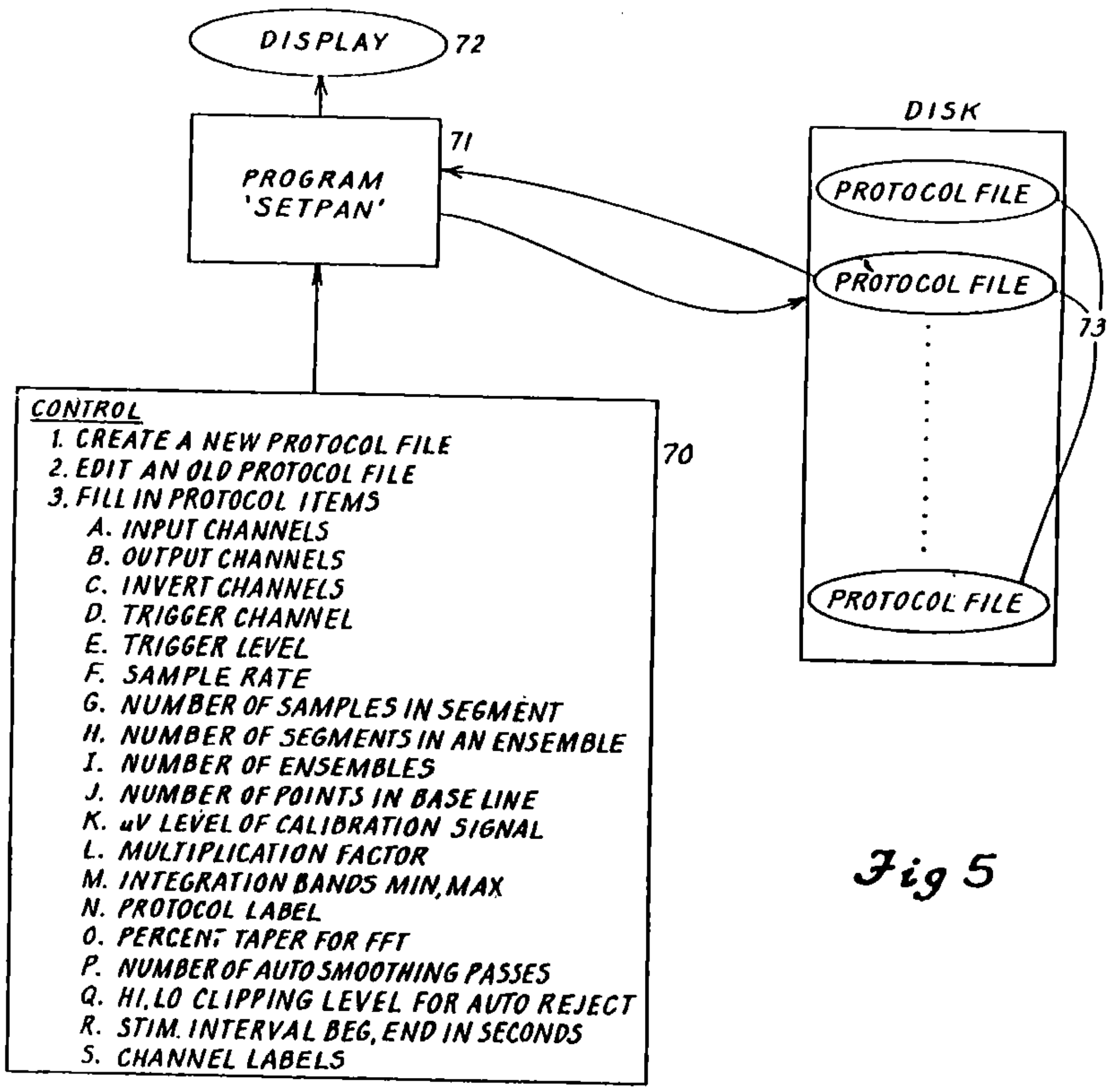
DISPLAY
72
PROGRAM 'SETPAN'
71
DISK
PROTOCOL FILE
PROTOCOL FILE
PROTOCOL FILE
73
70
CONTROL
1. CREATE A NEW PROTOCOL FILE
2. EDIT AN OLD PROTOCOL FILE
3. FILL IN PROTOCOL ITEMS
A. INPUT CHANNELS
B. OUTPUT CHANNELS
C. INVERT CHANNELS
D. TRIGGER CHANNEL
E. TRIGGER LEVEL
F. SAMPLE RATE
G. NUMBER OF SAMPLES IN SEGMENT
H. NUMBER OF SEGMENTS IN AN ENSEMBLE
I. NUMBER OF ENSEMBLES
J. NUMBER OF POINTS IN BASE LINE
K. uV LEVEL OF CALIBRATION SIGNAL
L. MULTIPLICATION FACTOR
M. INTEGRATION BANDS MIN, MAX
N. PROTOCOL LABEL
O. PERCENT TAPER FOR FFT
P. NUMBER OF AUTO SMOOTHING PASSES
Q. HI, LO CLIPPING LEVEL FOR AUTO REJECT
R. STIM. INTERVAL BEG, END IN SECONDS
S. CHANNEL LABELS

Fig 5

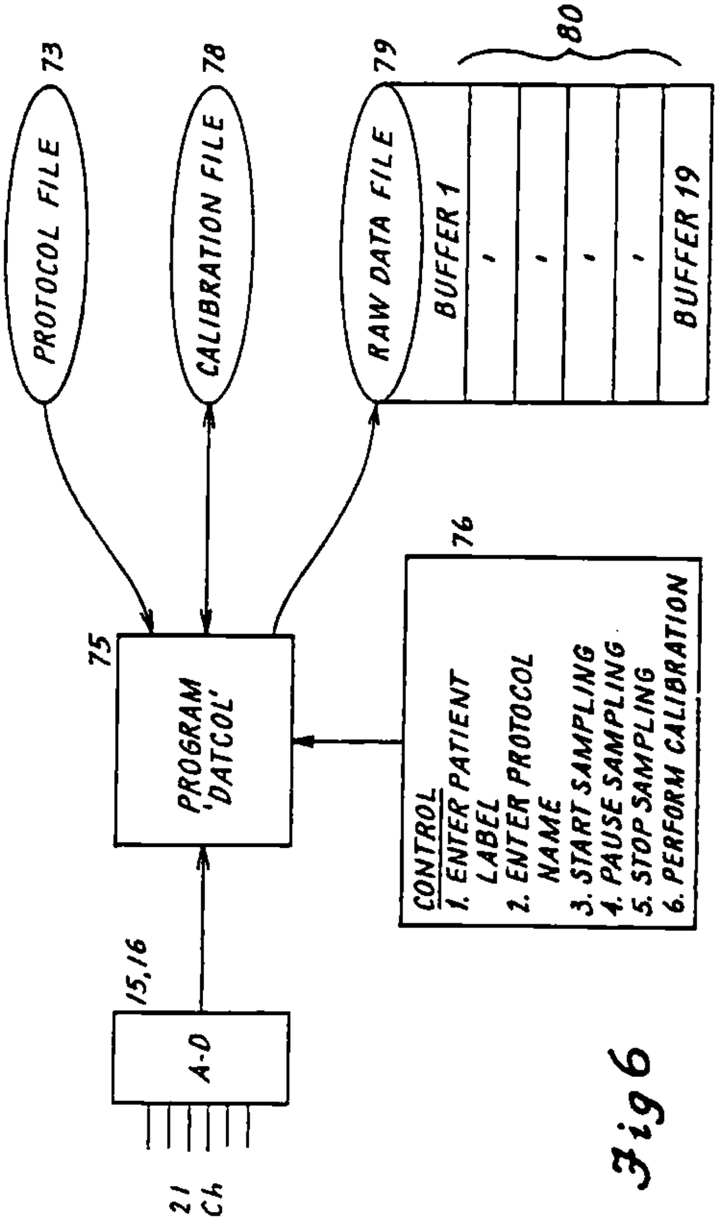
PROTOCOL FILE
73
CALIBRATION FILE
78
RAW DATA FILE
79
BUFFER 1
BUFFER 19
80
PROGRAM 'DATCOL'
75
76
CONTROL
1. ENTER PATIENT LABEL
2. ENTER PROTOCOL NAME
3. START SAMPLING
4. PAUSE SAMPLING
5. STOP SAMPLING
6. PERFORM CALIBRATION
15,16
A-D
21 Ch

Fig 6

41

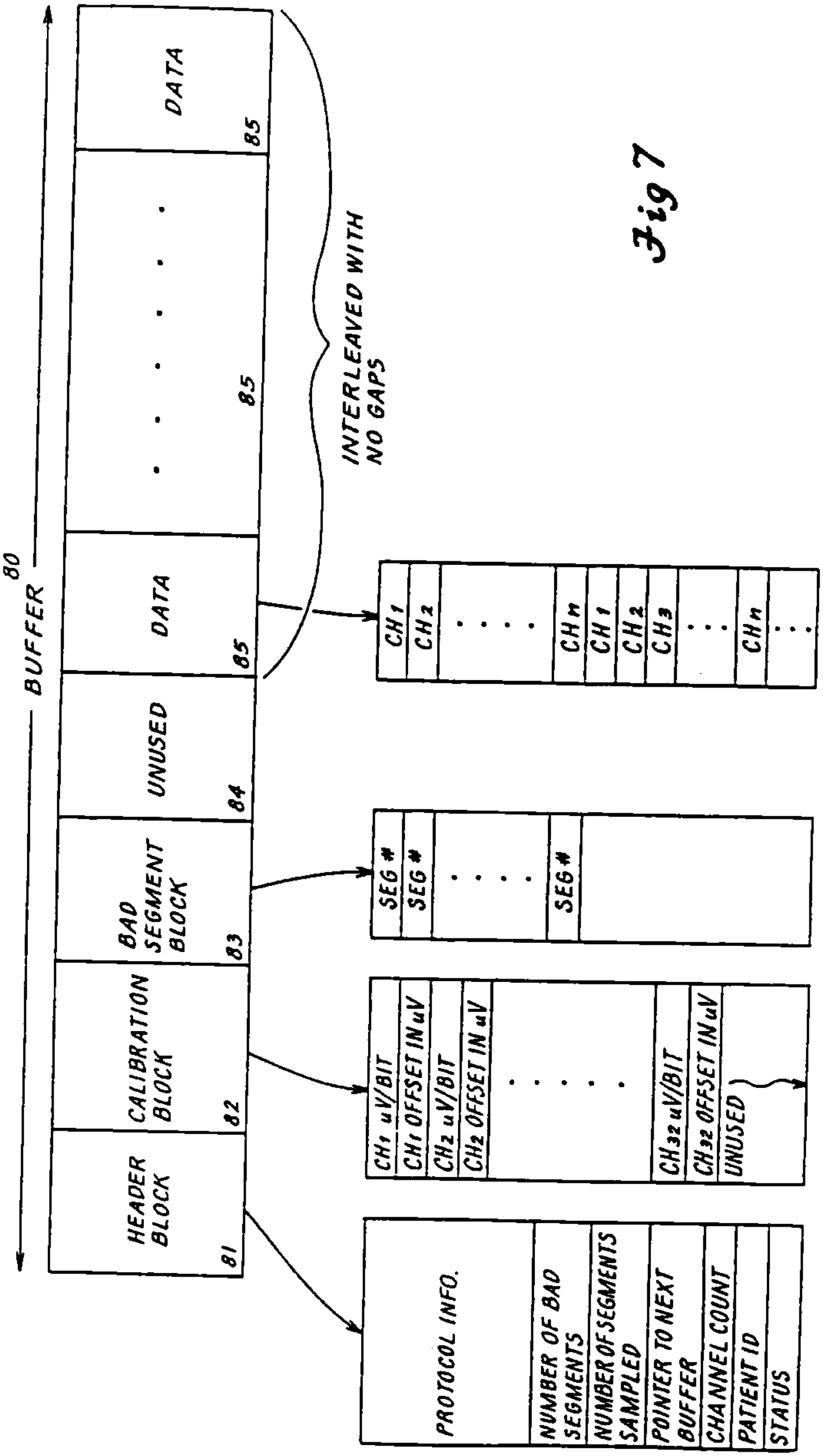
BUFFER
80
HEADER BLOCK
81
CALIBRATION BLOCK
82
BAD SEGMENT BLOCK
83
UNUSED
84
DATA
85
85
DATA
85
INTERLEAVED WITH NO GAPS
PROTOCOL INFO.
NUMBER OF BAD SEGMENTS
NUMBER OF SEGMENTS SAMPLED
POINTER TO NEXT BUFFER
CHANNEL COUNT
PATIENT ID
STATUS
CH1 uV/BIT
CH1 OFFSET IN uV
CH2 uV/BIT
CH2 OFFSET IN uV
CH32 uV/BIT
CH32 OFFSET IN uV
UNUSED
SEG #
SEG #
SEG #
CH1
CH2
CHn
CH1
CH2
CH3
CHn

Fig 7

42

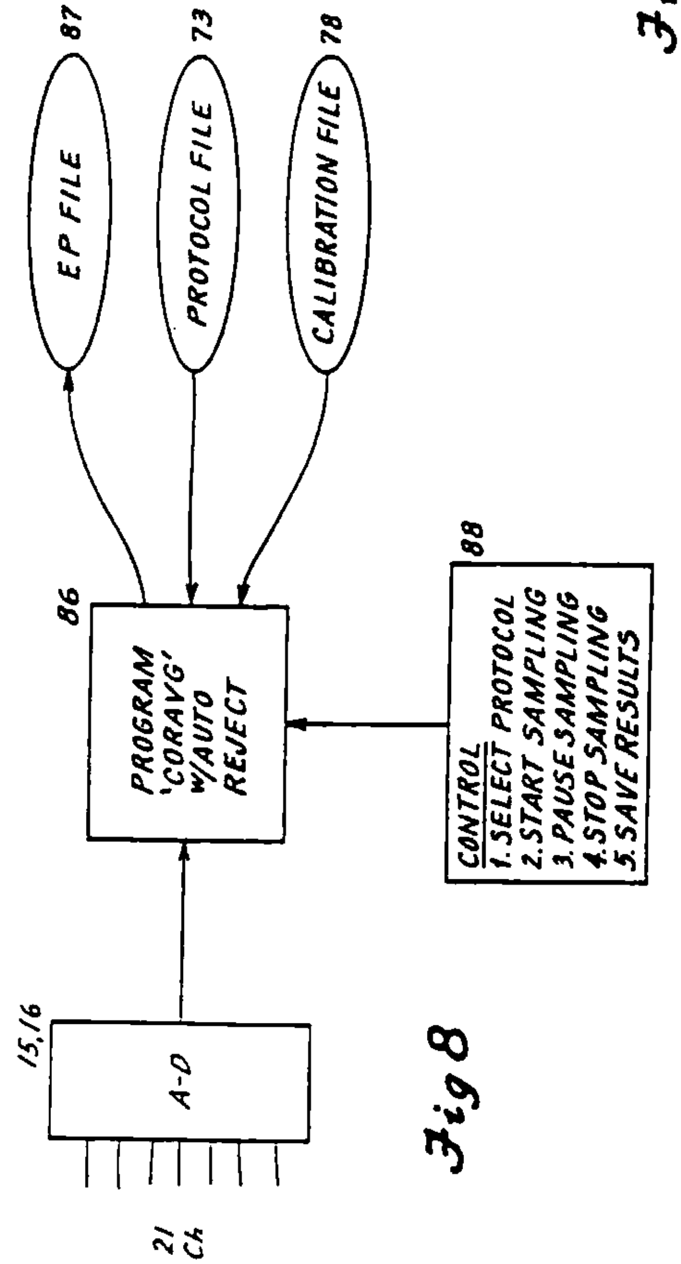
87
EP FILE
73
PROTOCOL FILE
78
CALIBRATION FILE
86
PROGRAM 'CORAVG' W/AUTO REJECT
88
CONTROL
1. SELECT PROTOCOL
2. START SAMPLING
3. PAUSE SAMPLING
4. STOP SAMPLING
5. SAVE RESULTS
15,16
A-D
21 Ch

Fig 8

Fig 9

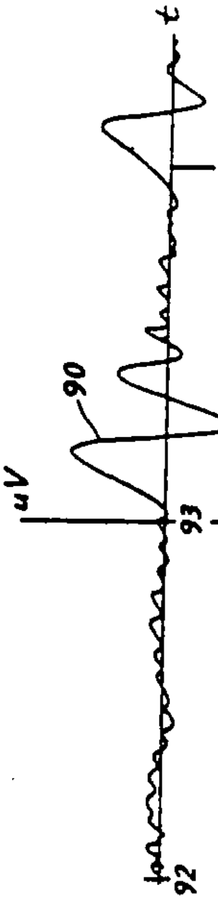
uV
90
93
92
91
t

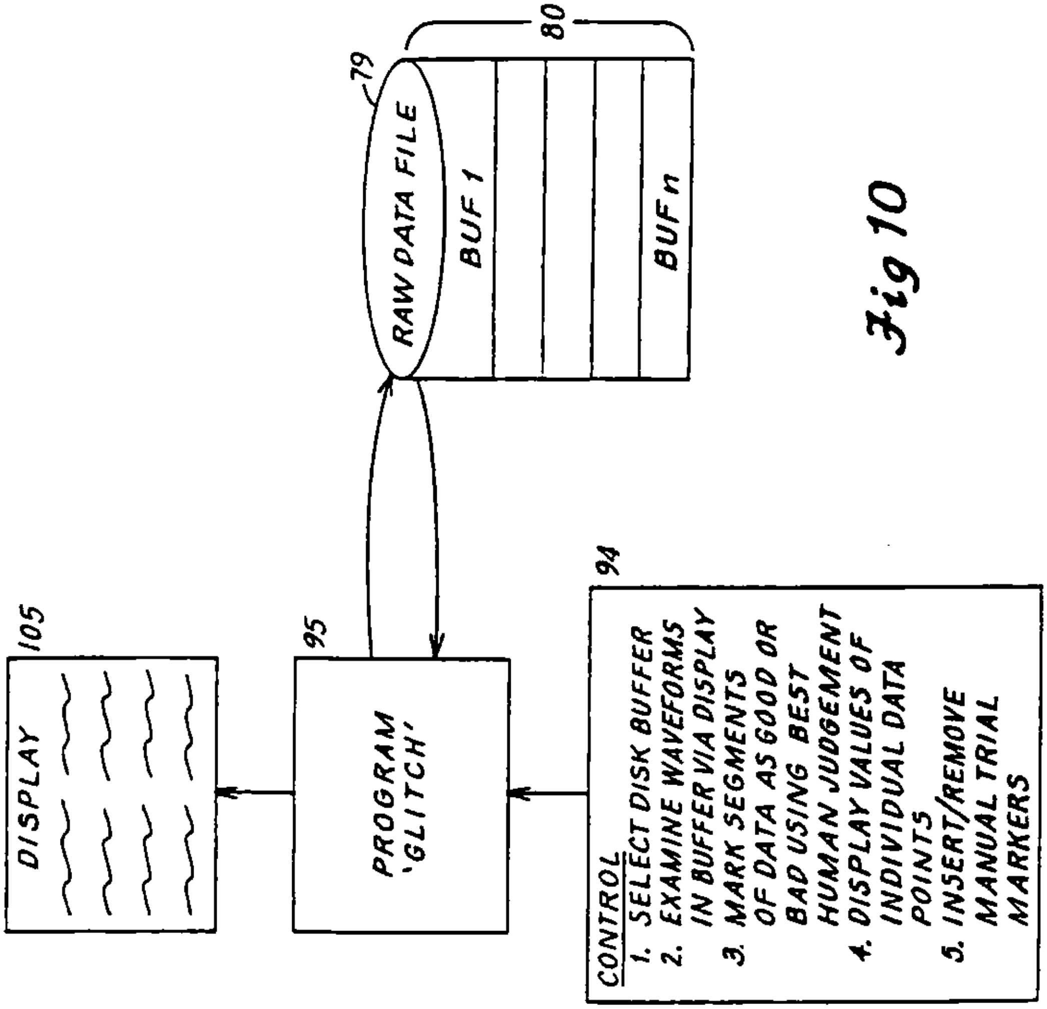
80
79
RAW DATA FILE
BUF 1
BUF n
105
DISPLAY
95
PROGRAM 'GLITCH'
94
CONTROL
1. SELECT DISK BUFFER
2. EXAMINE WAVEFORMS IN BUFFER VIA DISPLAY
3. MARK SEGMENTS OF DATA AS GOOD OR BAD USING BEST HUMAN JUDGEMENT
4. DISPLAY VALUES OF INDIVIDUAL DATA POINTS
5. INSERT/REMOVE MANUAL TRIAL MARKERS

Fig 10

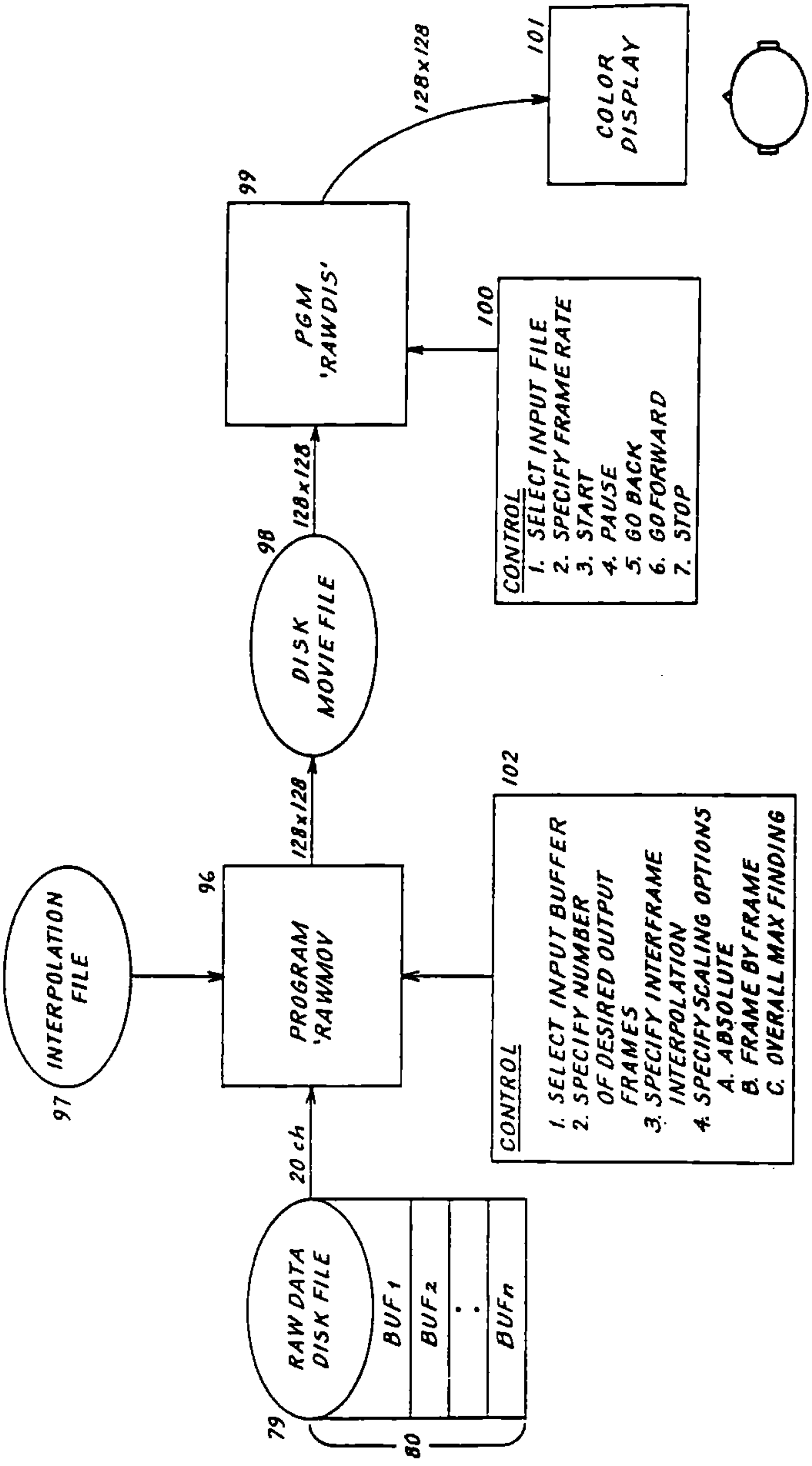
79
RAW DATA DISK FILE
BUF1
BUF2
:
BUFn
80
20 ch
97
INTERPOLATION FILE
96
PROGRAM 'RAWMOV
128x128
102
CONTROL
1. SELECT INPUT BUFFER
2. SPECIFY NUMBER OF DESIRED OUTPUT FRAMES
3. SPECIFY INTERFRAME INTERPOLATION
4. SPECIFY SCALING OPTIONS
A. ABSOLUTE
B. FRAME BY FRAME
C. OVERALL MAX FINDING
98
DISK MOVIE FILE
128x128
99
PGM 'RAWDIS'
100
CONTROL
1. SELECT INPUT FILE
2. SPECIFY FRAME RATE
3. START
4. PAUSE
5. GO BACK
6. GO FORWARD
7. STOP
128x128
101
COLOR DISPLAY

Fig 11

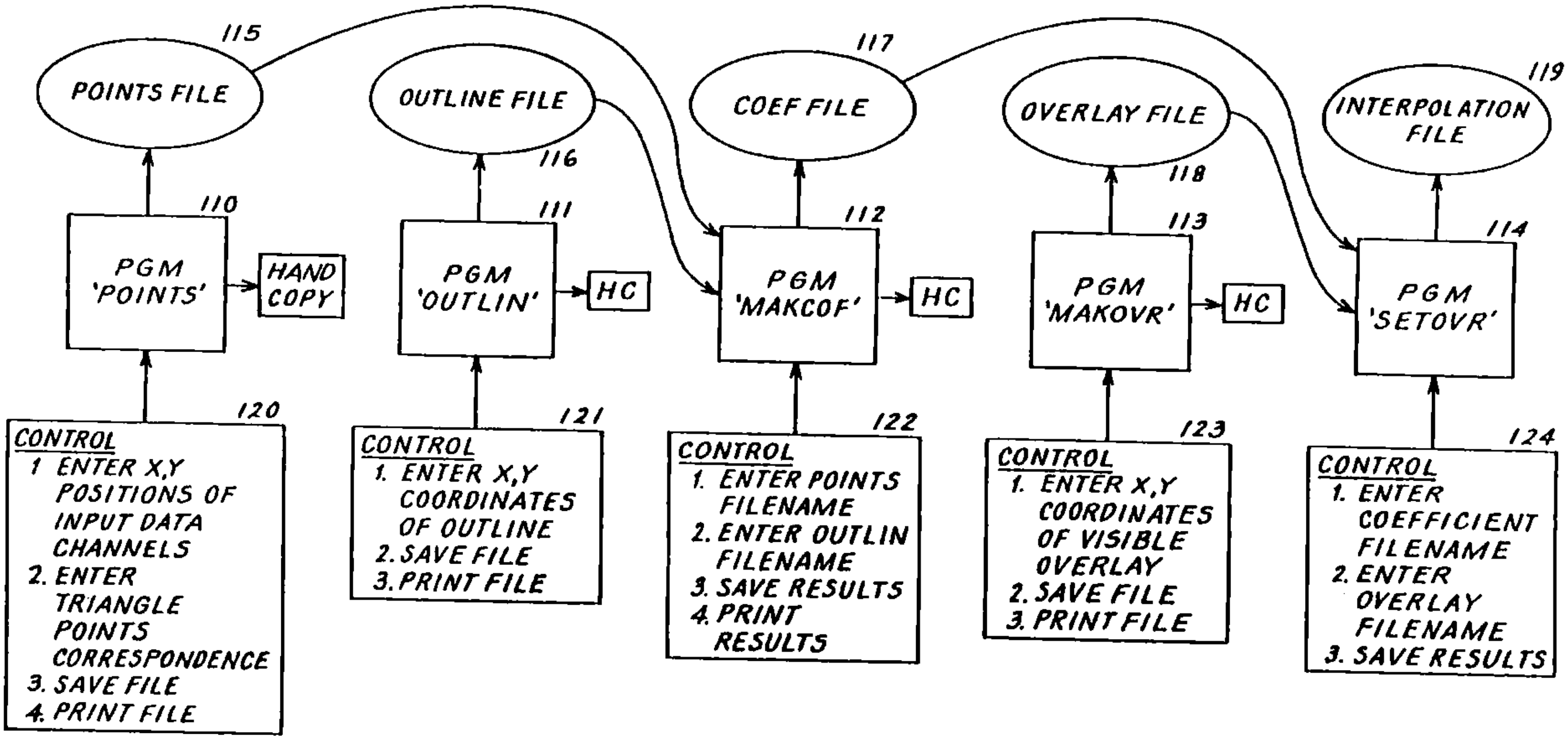
115
POINTS FILE
116
OUTLINE FILE
117
COEF FILE
118
OVERLAY FILE
119
INTERPOLATION FILE
110
PGM 'POINTS'
HAND COPY
111
PGM 'OUTLIN'
HC
112
PGM 'MAKCOF'
HC
113
PGM 'MAKOVR'
HC
114
PGM 'SETOVR'
120
CONTROL
1 ENTER X,Y POSITIONS OF INPUT DATA CHANNELS
2. ENTER TRIANGLE POINTS CORRESPONDENCE
3. SAVE FILE
4. PRINT FILE
121
CONTROL
1. ENTER X,Y COORDINATES OF OUTLINE
2. SAVE FILE
3. PRINT FILE
122
CONTROL
1. ENTER POINTS FILENAME
2. ENTER OUTLIN FILENAME
3. SAVE RESULTS
4. PRINT RESULTS
123
CONTROL
1. ENTER X,Y COORDINATES OF VISIBLE OVERLAY
2. SAVE FILE
3. PRINT FILE
124
CONTROL
1. ENTER COEFFICIENT FILENAME
2. ENTER OVERLAY FILENAME
3. SAVE RESULTS

Fig 12

46

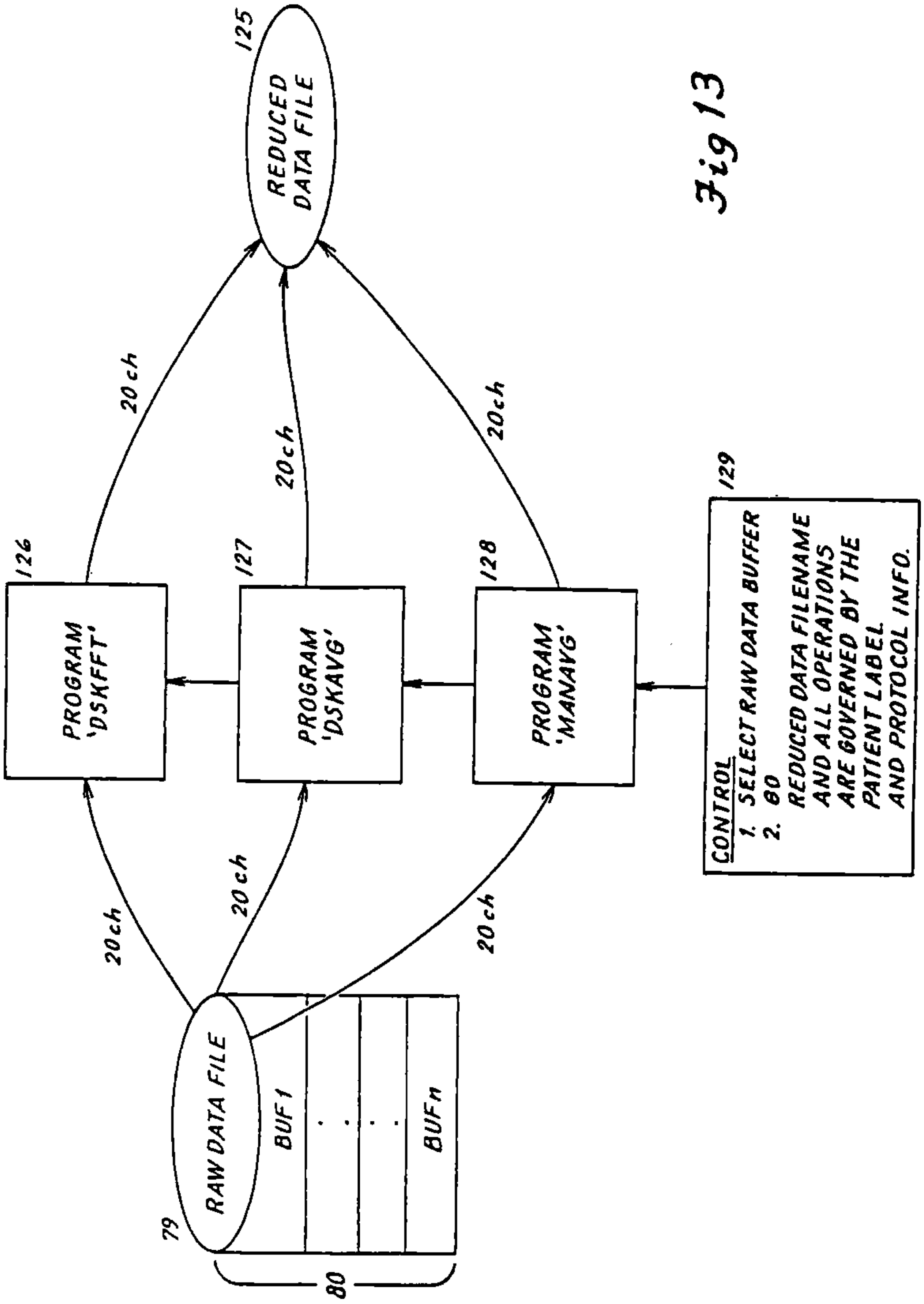
REDUCED DATA FILE
125
20 ch
20 ch
20 ch
126
PROGRAM 'DSKFFT'
127
PROGRAM 'DSKAVG'
128
PROGRAM 'MANAVG'
129
CONTROL
1. SELECT RAW DATA BUFFER
2. GO
REDUCED DATA FILENAME AND ALL OPERATIONS ARE GOVERNED BY THE PATIENT LABEL AND PROTOCOL INFO.
20 ch
20 ch
20 ch
79
RAW DATA FILE
BUF1
BUFn
80

Fig 13

42

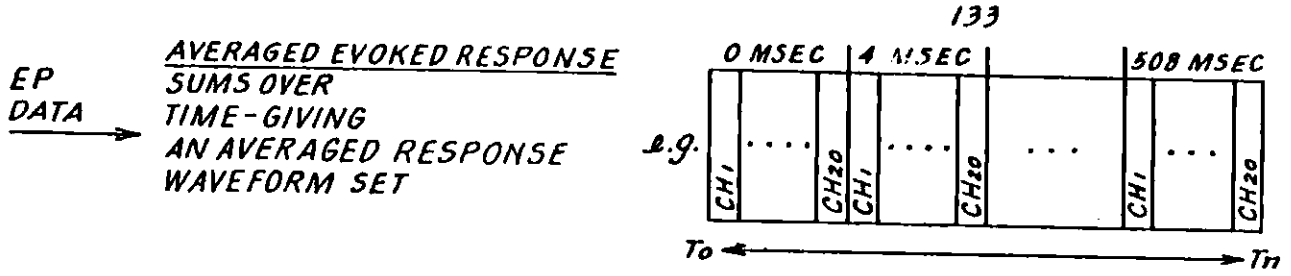
EP DATA
AVERAGED EVOKED RESPONSE
SUMS OVER
TIME-GIVING
AN AVERAGED RESPONSE
WAVEFORM SET
133
0 MSEC
4 MSEC
508 MSEC
e.g.
CH1
CH20
CH1
CH20
CH1
CH20
To
Tn

Fig 14

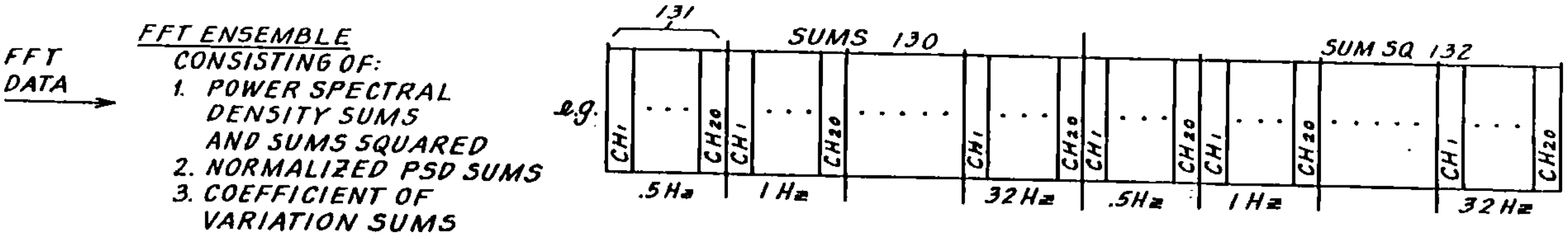
FFT DATA
FFT ENSEMBLE
CONSISTING OF:
1. POWER SPECTRAL
DENSITY SUMS
AND SUMS SQUARED
2. NORMALIZED PSD SUMS
3. COEFFICIENT OF
VARIATION SUMS
131
SUMS 130
SUM SQ 132
e.g.
CH1
CH20
CH1
CH20
CH1
CH20
CH1
CH20
CH1
CH20
CH1
CH20
.5 Hz
1 Hz
32 Hz
.5Hz
1 Hz
32 Hz

Fig 15

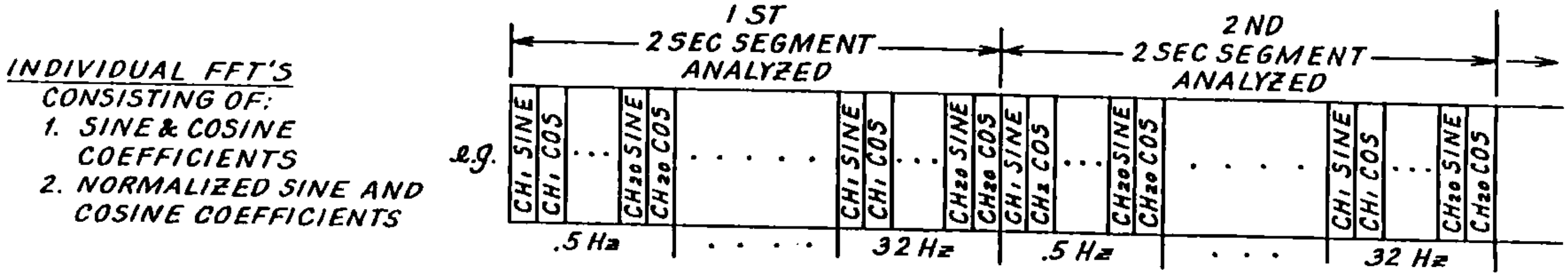
INDIVIDUAL FFT'S
CONSISTING OF:
1. SINE & COSINE
COEFFICIENTS
2. NORMALIZED SINE AND
COSINE COEFFICIENTS
1 ST
2 SEC SEGMENT
ANALYZED
2 ND
2 SEC SEGMENT
ANALYZED
e.g.
CH1 SINE
CH1 COS
CH20 SINE
CH20 COS
CH1 SINE
CH1 COS
CH20 SINE
CH20 COS
CH1 SINE
CH1 COS
CH20 SINE
CH20 COS
CH1 SINE
CH1 COS
CH20 SINE
CH20 COS
.5 Hz
32 Hz
.5 Hz
32 Hz

Fig 16

48

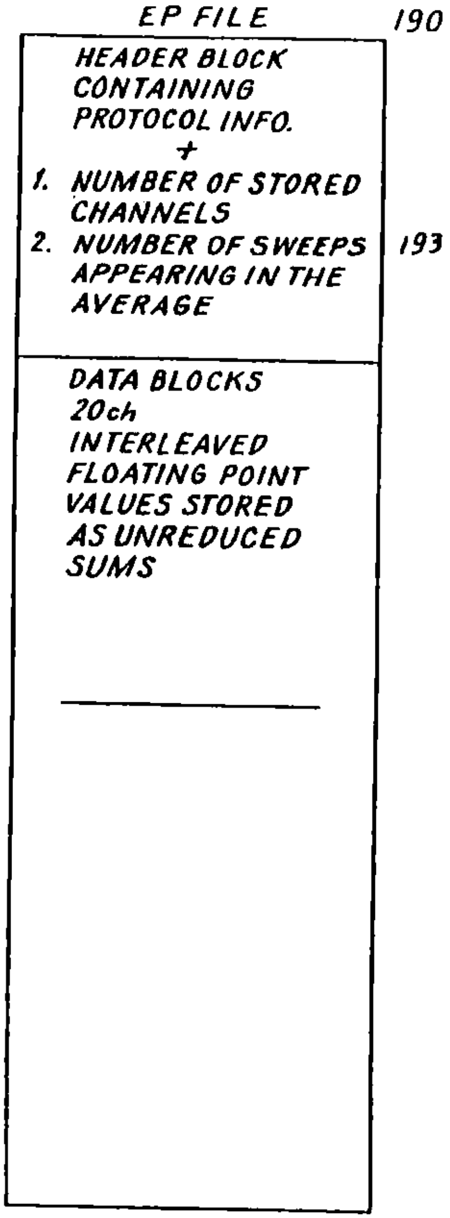
EP FILE
190
HEADER BLOCK
CONTAINING
PROTOCOL INFO.
+
1. NUMBER OF STORED
CHANNELS
2. NUMBER OF SWEEPS
APPEARING IN THE
AVERAGE
193
DATA BLOCKS
20ch
INTERLEAVED
FLOATING POINT
VALUES STORED
AS UNREDUCED
SUMS

Fig 17

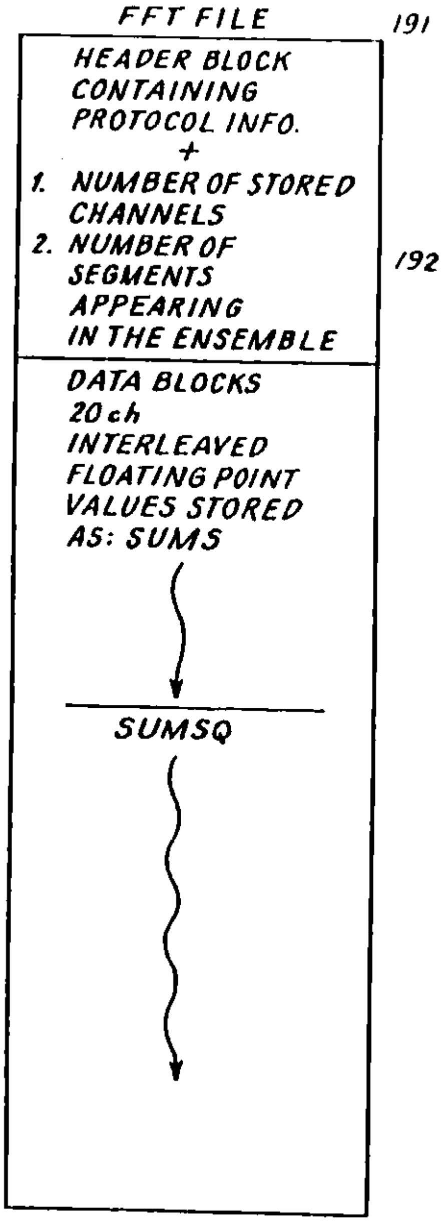
FFT FILE
191
HEADER BLOCK
CONTAINING
PROTOCOL INFO.
+
1. NUMBER OF STORED
CHANNELS
2. NUMBER OF
SEGMENTS
APPEARING
IN THE ENSEMBLE
192
DATA BLOCKS
20ch
INTERLEAVED
FLOATING POINT
VALUES STORED
AS: SUMS
SUMSQ

Fig 18

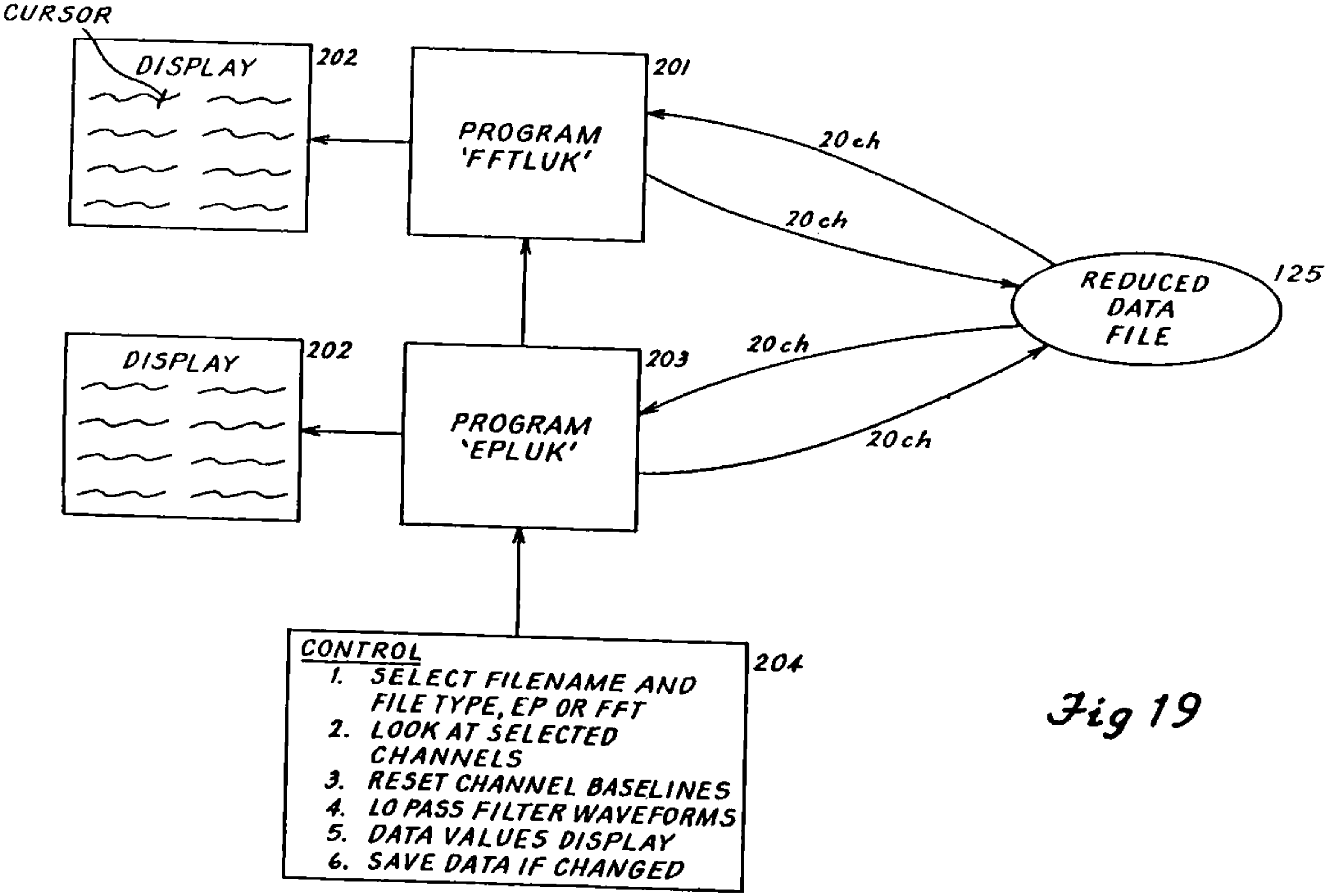
CURSOR
DISPLAY
202
PROGRAM 'FFTLUK'
201
20ch
20ch
REDUCED DATA FILE
125
DISPLAY
202
PROGRAM 'EPLUK'
203
20ch
20ch
CONTROL
1. SELECT FILENAME AND FILE TYPE, EP OR FFT
2. LOOK AT SELECTED CHANNELS
3. RESET CHANNEL BASELINES
4. LO PASS FILTER WAVEFORMS
5. DATA VALUES DISPLAY
6. SAVE DATA IF CHANGED
204

Fig 19

50

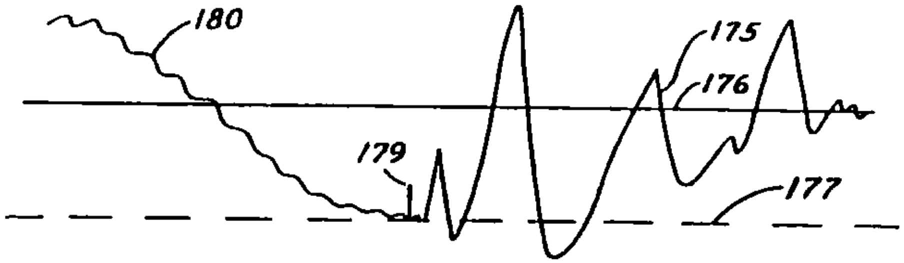
180
175
176
179
177

Fig 20

Fig 26

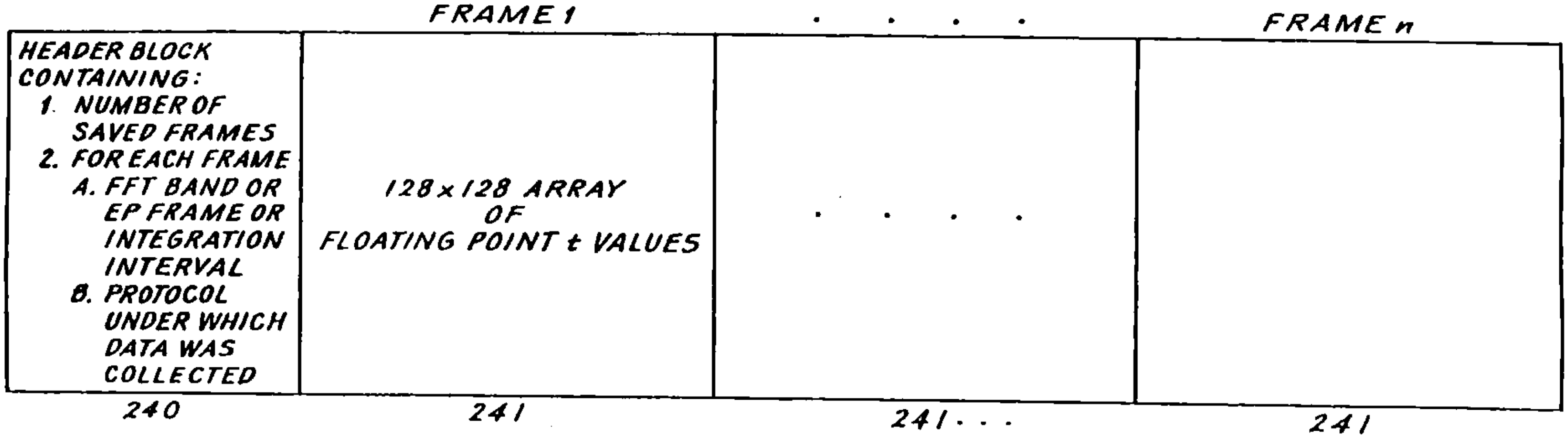
FRAME 1
. . . .
FRAME n
HEADER BLOCK
CONTAINING:
1. NUMBER OF SAVED FRAMES
2. FOR EACH FRAME
A. FFT BAND OR EP FRAME OR INTEGRATION INTERVAL
B. PROTOCOL UNDER WHICH DATA WAS COLLECTED
128 x 128 ARRAY OF FLOATING POINT t VALUES
. . . .
240
241
241 . . .
241

51

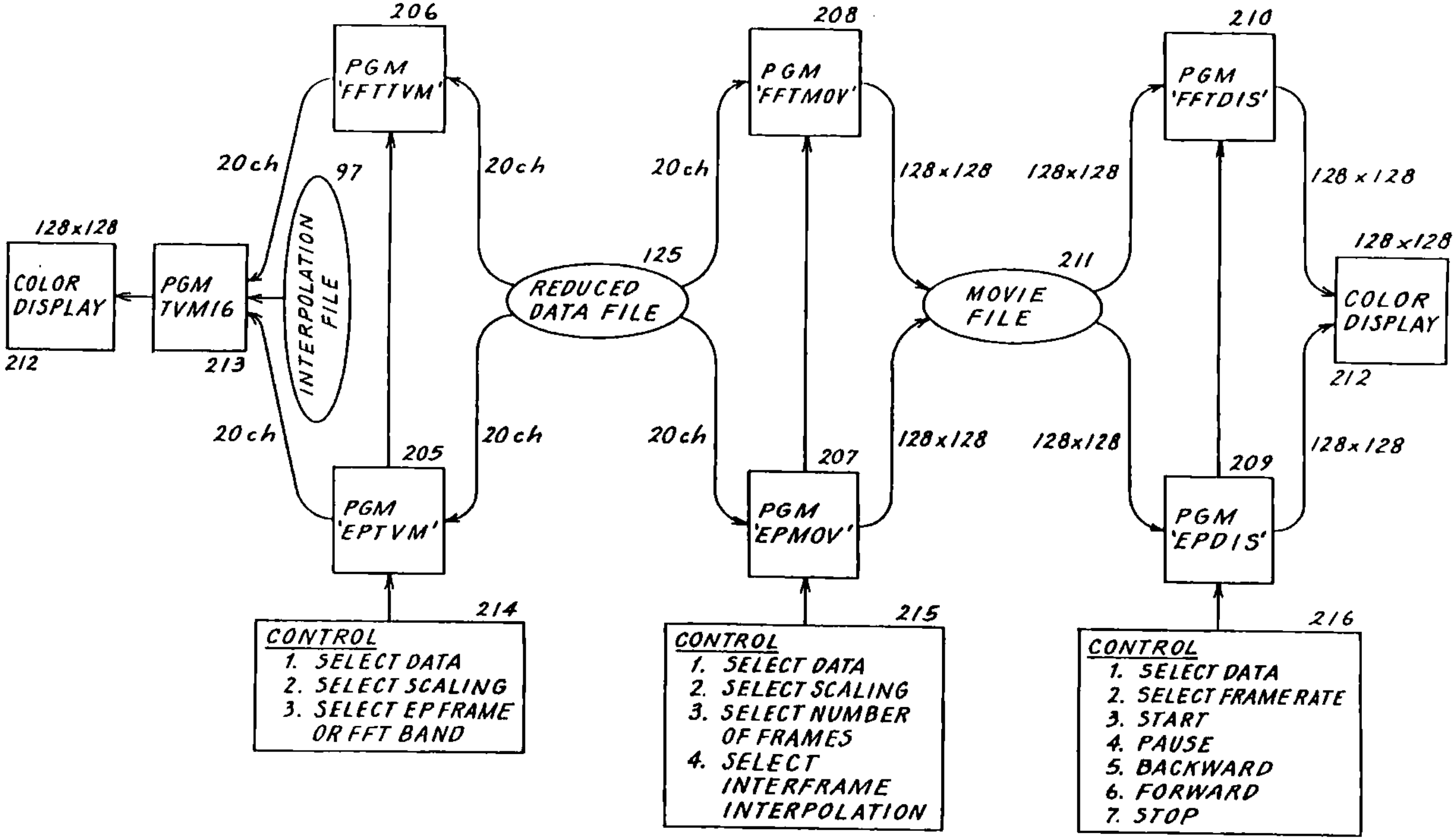
206
PGM 'FFTVM'
208
PGM 'FFTMOV'
210
PGM 'FFTDIS'
20ch
97
20ch
20ch
128x128
128x128
128 x 128
128x128
COLOR DISPLAY
212
PGM TVM16
213
INTERPOLATION FILE
REDUCED DATA FILE
125
MOVIE FILE
211
128x128
COLOR DISPLAY
212
20ch
20ch
20ch
128x128
128x128
128x128
205
PGM 'EPTVM'
207
PGM 'EPMOV'
209
PGM 'EPDIS'
214
CONTROL
1. SELECT DATA
2. SELECT SCALING
3. SELECT EP FRAME OR FFT BAND
215
CONTROL
1. SELECT DATA
2. SELECT SCALING
3. SELECT NUMBER OF FRAMES
4. SELECT INTERFRAME INTERPOLATION
216
CONTROL
1. SELECT DATA
2. SELECT FRAME RATE
3. START
4. PAUSE
5. BACKWARD
6. FORWARD
7. STOP

Fig 21

52

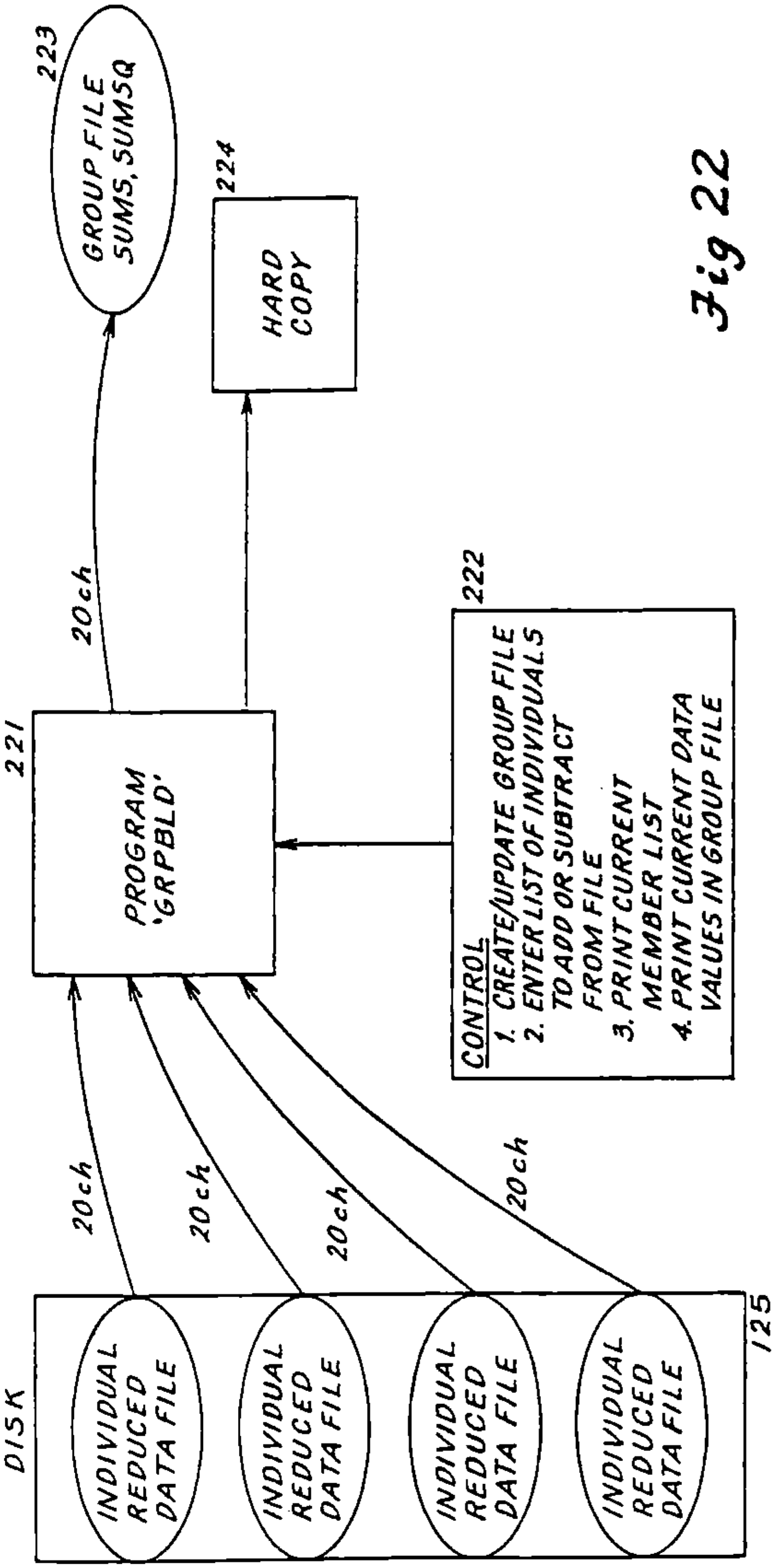
DISK
125
INDIVIDUAL REDUCED DATA FILE
INDIVIDUAL REDUCED DATA FILE
INDIVIDUAL REDUCED DATA FILE
INDIVIDUAL REDUCED DATA FILE
20ch
20ch
20ch
20ch
221
PROGRAM 'GRPBLD'
222
CONTROL
1. CREATE/UPDATE GROUP FILE
2. ENTER LIST OF INDIVIDUALS TO ADD OR SUBTRACT FROM FILE
3. PRINT CURRENT MEMBER LIST
4. PRINT CURRENT DATA VALUES IN GROUP FILE
20ch
223
GROUP FILE SUMS, SUMSQ
224
HARD COPY

Fig 22

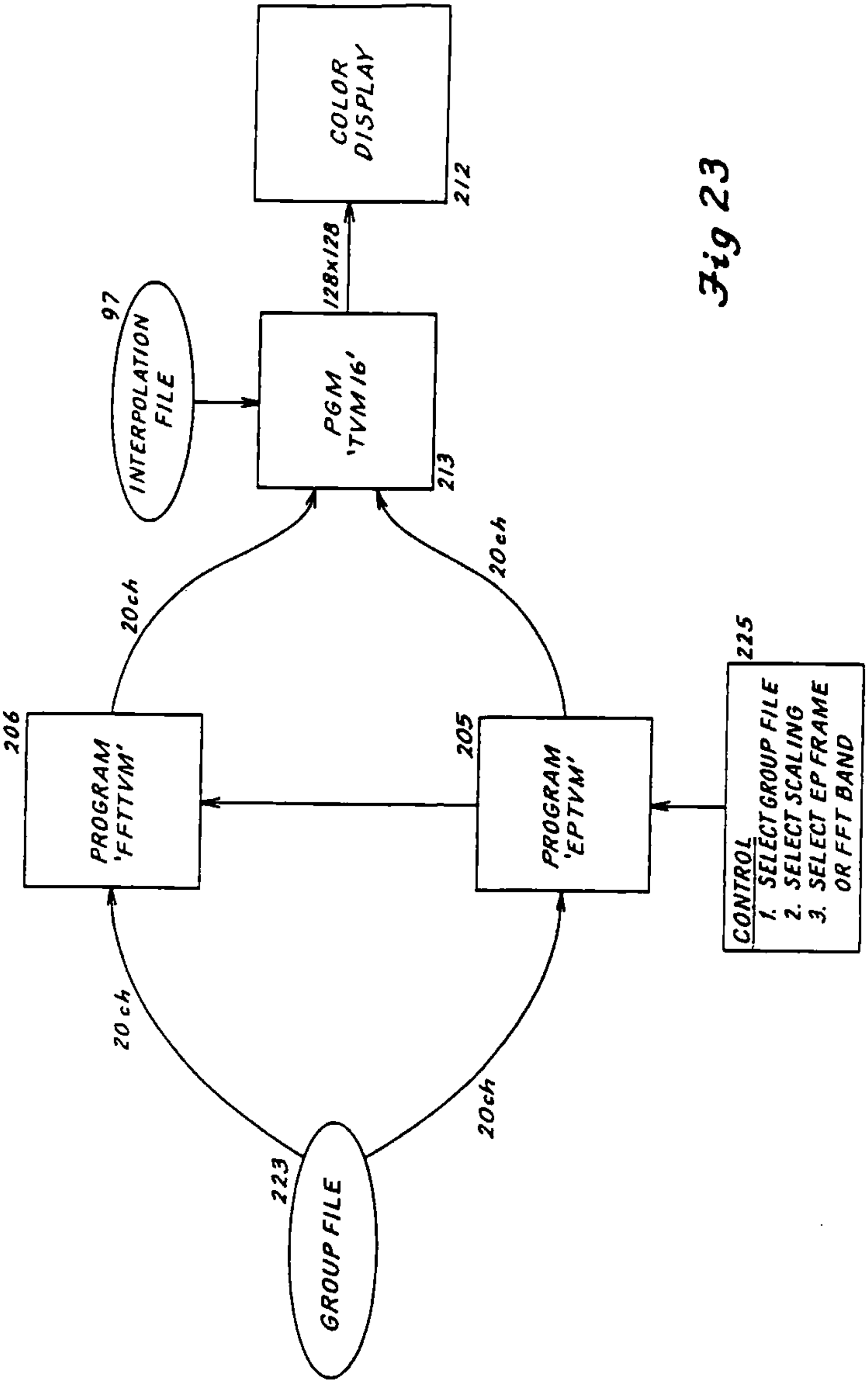
COLOR DISPLAY
212
128x128
PGM 'TVM 16'
213
INTERPOLATION FILE
97
20ch
20ch
PROGRAM 'FFTTVM'
206
PROGRAM 'EPTVM'
205
CONTROL
1. SELECT GROUP FILE
2. SELECT SCALING
3. SELECT EP FRAME OR FFT BAND
225
20ch
20ch
GROUP FILE
223

Fig 23

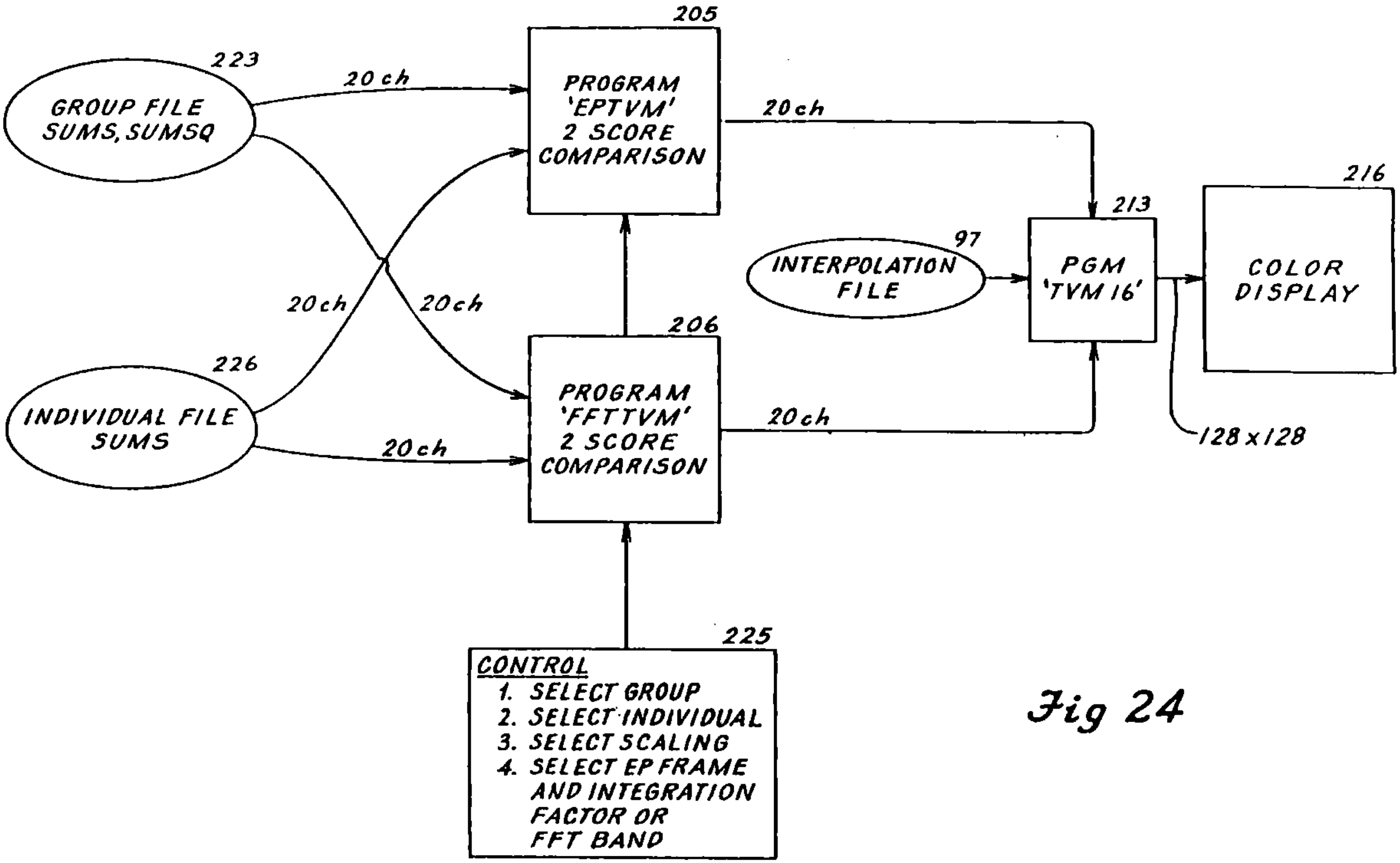
223
GROUP FILE
SUMS, SUMSQ
226
INDIVIDUAL FILE
SUMS
20 ch
20ch
20ch
20ch
205
PROGRAM
'EPTVM'
2 SCORE
COMPARISON
206
PROGRAM
'FFTTVM'
2 SCORE
COMPARISON
20ch
20ch
97
INTERPOLATION
FILE
213
PGM
'TVM 16'
216
COLOR
DISPLAY
128 x 128
225
CONTROL
1. SELECT GROUP
2. SELECT INDIVIDUAL
3. SELECT SCALING
4. SELECT EP FRAME
AND INTEGRATION
FACTOR OR
FFT BAND

Fig 24

55

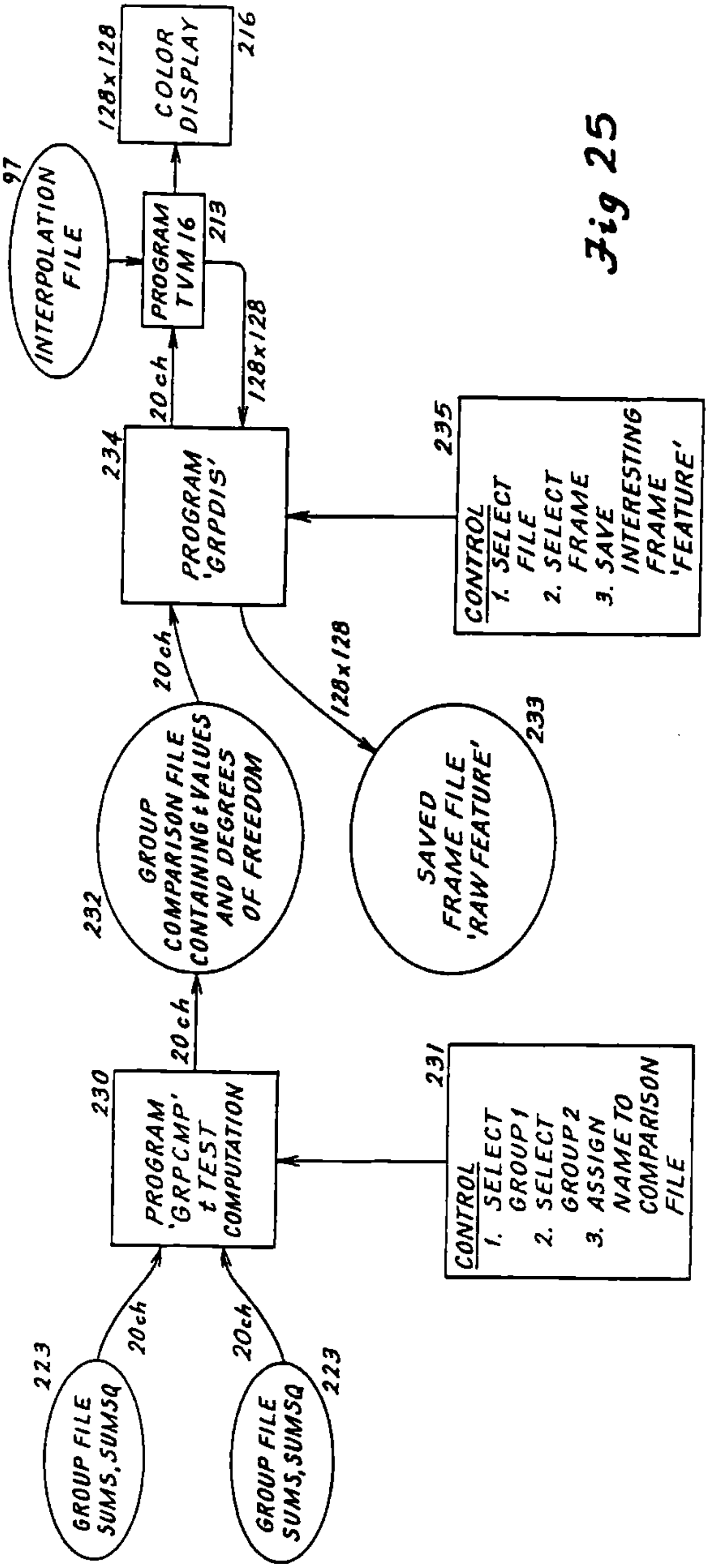
GROUP FILE SUMS, SUMSQ
223
20ch
GROUP FILE SUMS, SUMSQ
223
20ch
PROGRAM 'GRPCMP' t TEST COMPUTATION
230
CONTROL
1. SELECT GROUP 1
2. SELECT GROUP 2
3. ASSIGN NAME TO COMPARISON FILE
231
20ch
GROUP COMPARISON FILE CONTAINING t VALUES AND DEGREES OF FREEDOM
232
20ch
PROGRAM 'GRPDIS'
234
128x128
SAVED FRAME FILE 'RAW FEATURE'
233
CONTROL
1. SELECT FILE
2. SELECT FRAME
3. SAVE INTERESTING FRAME 'FEATURE'
235
20ch
128x128
INTERPOLATION FILE
97
PROGRAM TVM 16
213
128x128
COLOR DISPLAY
216

Fig 25

56

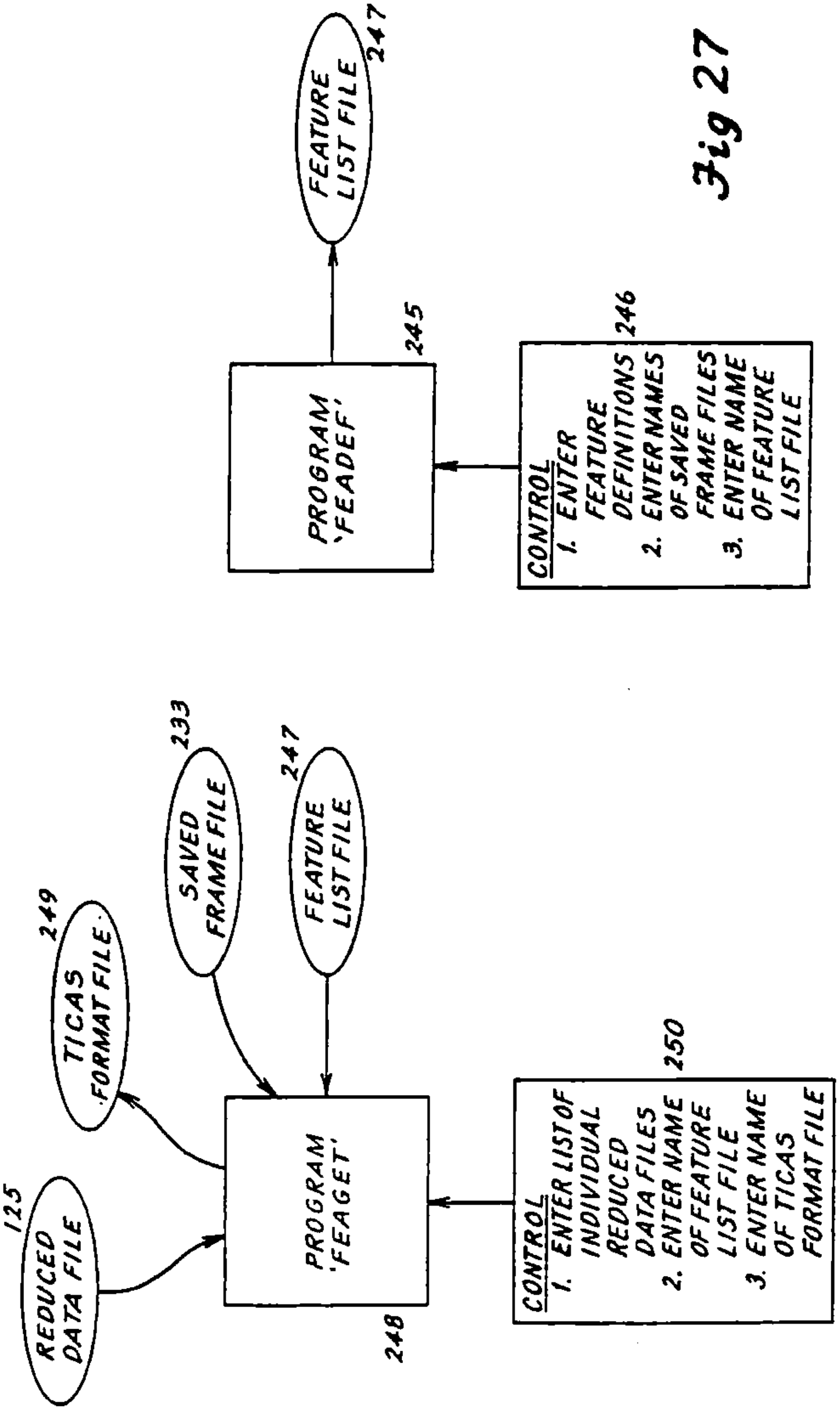
FEATURE LIST FILE
247
PROGRAM 'FEADEF'
245
CONTROL
1. ENTER FEATURE DEFINITIONS
2. ENTER NAMES OF SAVED FRAME FILES
3. ENTER NAME OF FEATURE LIST FILE
246
249
TICAS FORMAT FILE
233
SAVED FRAME FILE
247
FEATURE LIST FILE
125
REDUCED DATA FILE
PROGRAM 'FEAGET'
248
250
CONTROL
1. ENTER LIST OF INDIVIDUAL REDUCED DATA FILES
2. ENTER NAME OF FEATURE LIST FILE
3. ENTER NAME OF TICAS FORMAT FILE

Fig 27

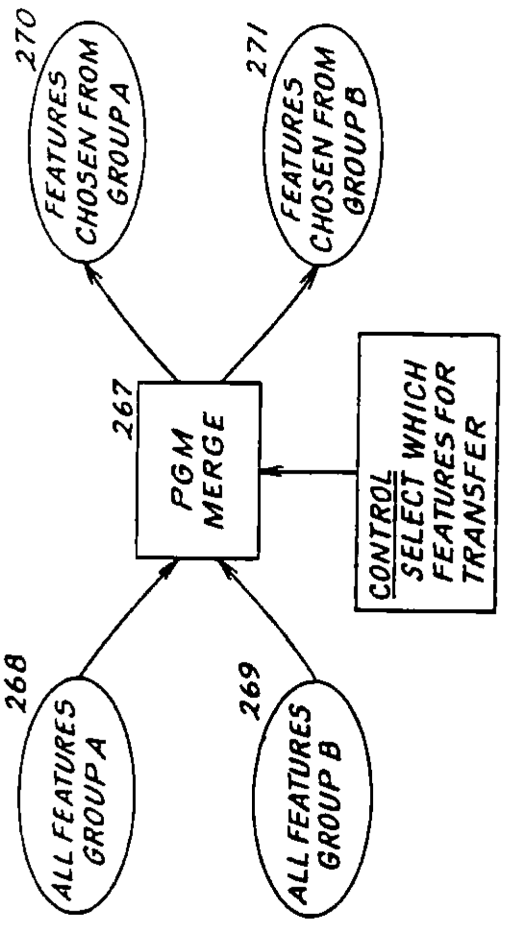
270
FEATURES CHOSEN FROM GROUP A
271
FEATURES CHOSEN FROM GROUP B
267
PGM MERGE
CONTROL
SELECT WHICH FEATURES FOR TRANSFER
268
ALL FEATURES GROUP A
269
ALL FEATURES GROUP B

Fig 28

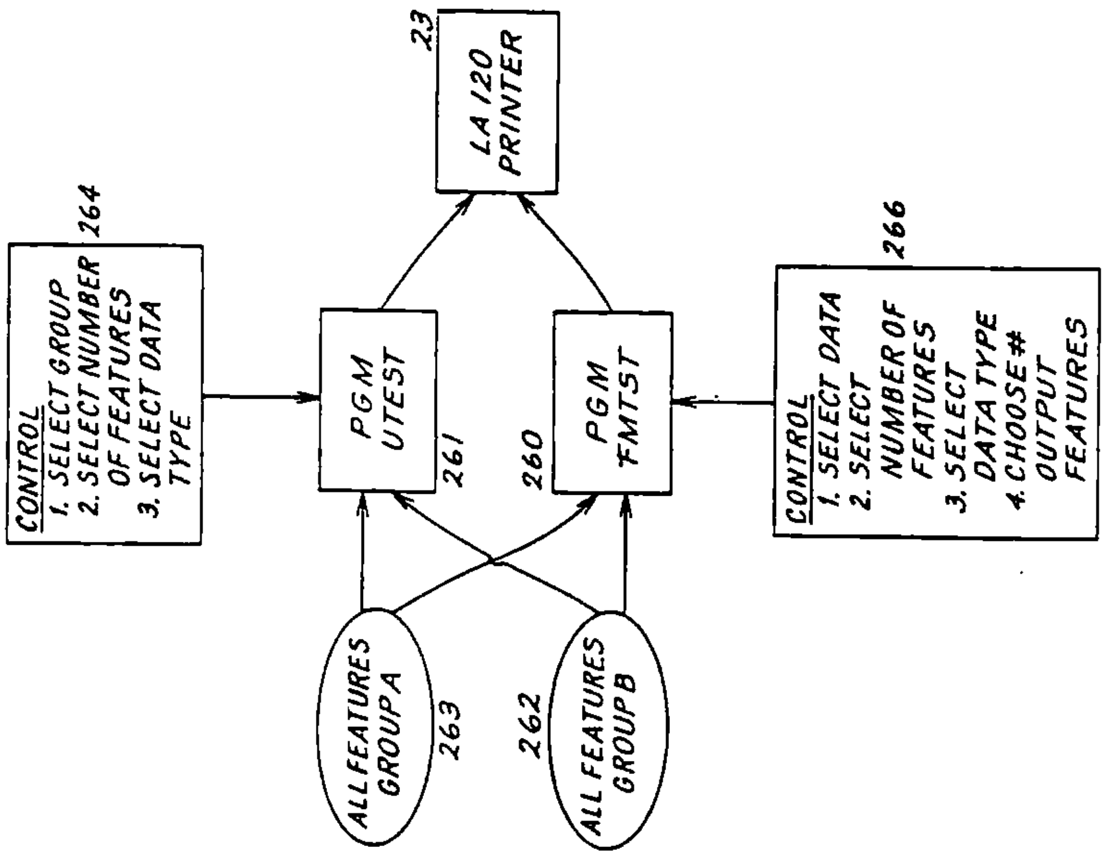
264
CONTROL
1. SELECT GROUP
2. SELECT NUMBER OF FEATURES
3. SELECT DATA TYPE
23
LA 120 PRINTER
PGM UTEST
261
PGM FMTST
260
266
CONTROL
1. SELECT DATA
2. SELECT NUMBER OF FEATURES
3. SELECT DATA TYPE
4. CHOOSE # OUTPUT FEATURES
ALL FEATURES GROUP A
263
ALL FEATURES GROUP B
262

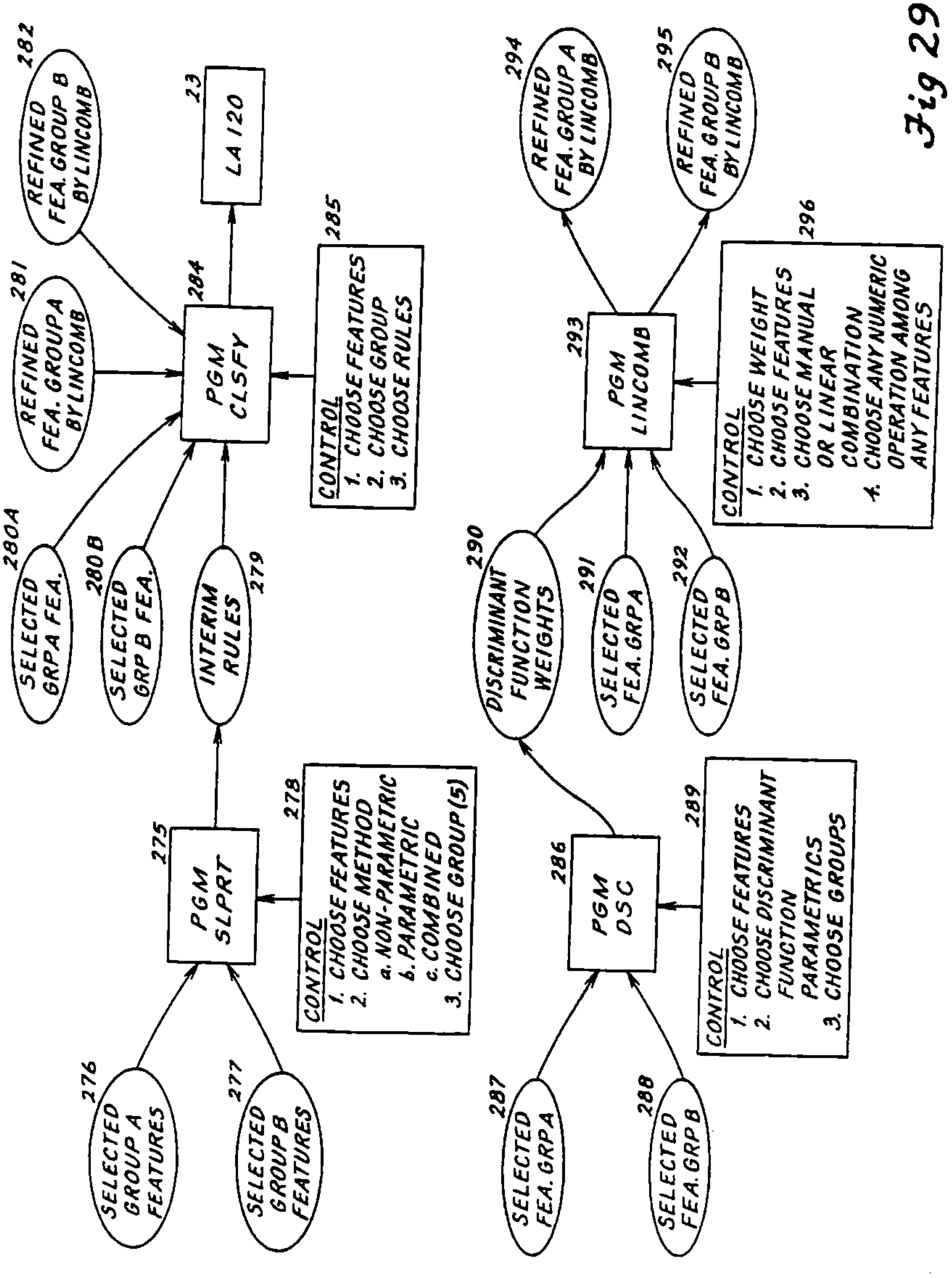

SELECTED GROUP A FEATURES
276
SELECTED GROUP B FEATURES
277
PGM SLPRT
275
CONTROL
1. CHOOSE FEATURES
2. CHOOSE METHOD
a. NON-PARAMETRIC
b. PARAMETRIC
c. COMBINED
3. CHOOSE GROUP(5)
278
INTERIM RULES
279
SELECTED GRPA FEA.
280A
SELECTED GRP B FEA.
280B
REFINED FEA. GROUPA BY LINCOMB
281
REFINED FEA. GROUP B BY LINCOMB
282
PGM CLSFY
284
CONTROL
1. CHOOSE FEATURES
2. CHOOSE GROUP
3. CHOOSE RULES
285
LA 120
23
SELECTED FEA. GRPA
287
SELECTED FEA. GRP B
288
PGM DSC
286
CONTROL
1. CHOOSE FEATURES
2. CHOOSE DISCRIMINANT FUNCTION PARAMETRICS
3. CHOOSE GROUPS
289
DISCRIMINANT FUNCTION WEIGHTS
290
SELECTED FEA. GRPA
291
SELECTED FEA. GRP B
292
PGM LINCOMB
293
CONTROL
1. CHOOSE WEIGHT
2. CHOOSE FEATURES
3. CHOOSE MANUAL OR LINEAR COMBINATION
4. CHOOSE ANY NUMERIC OPERATION AMONG ANY FEATURES
296
REFINED FEA. GROUP A BY LINCOMB
294
REFINED FEA. GROUP B BY LINCOMB
295

Fig 29

59

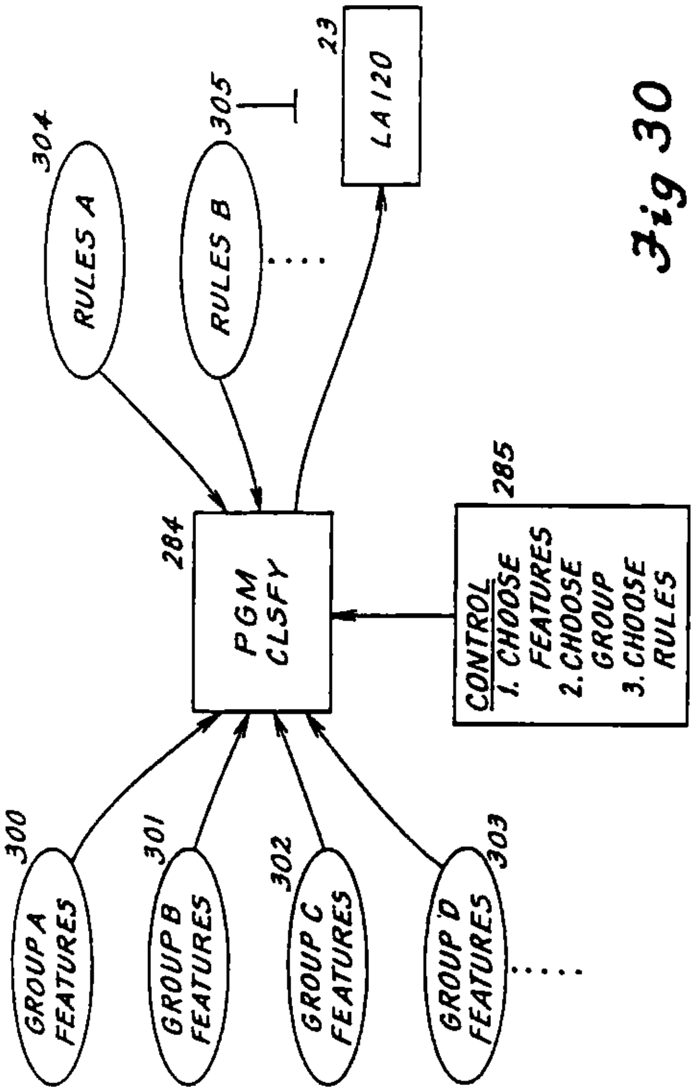
300
GROUP A FEATURES
301
GROUP B FEATURES
302
GROUP C FEATURES
303
GROUP D FEATURES
.....
284
PGM CLSFY
285
CONTROL
1. CHOOSE FEATURES
2. CHOOSE GROUP
3. CHOOSE RULES
304
RULES A
305
RULES B
....
23
LA120

Fig 30

60

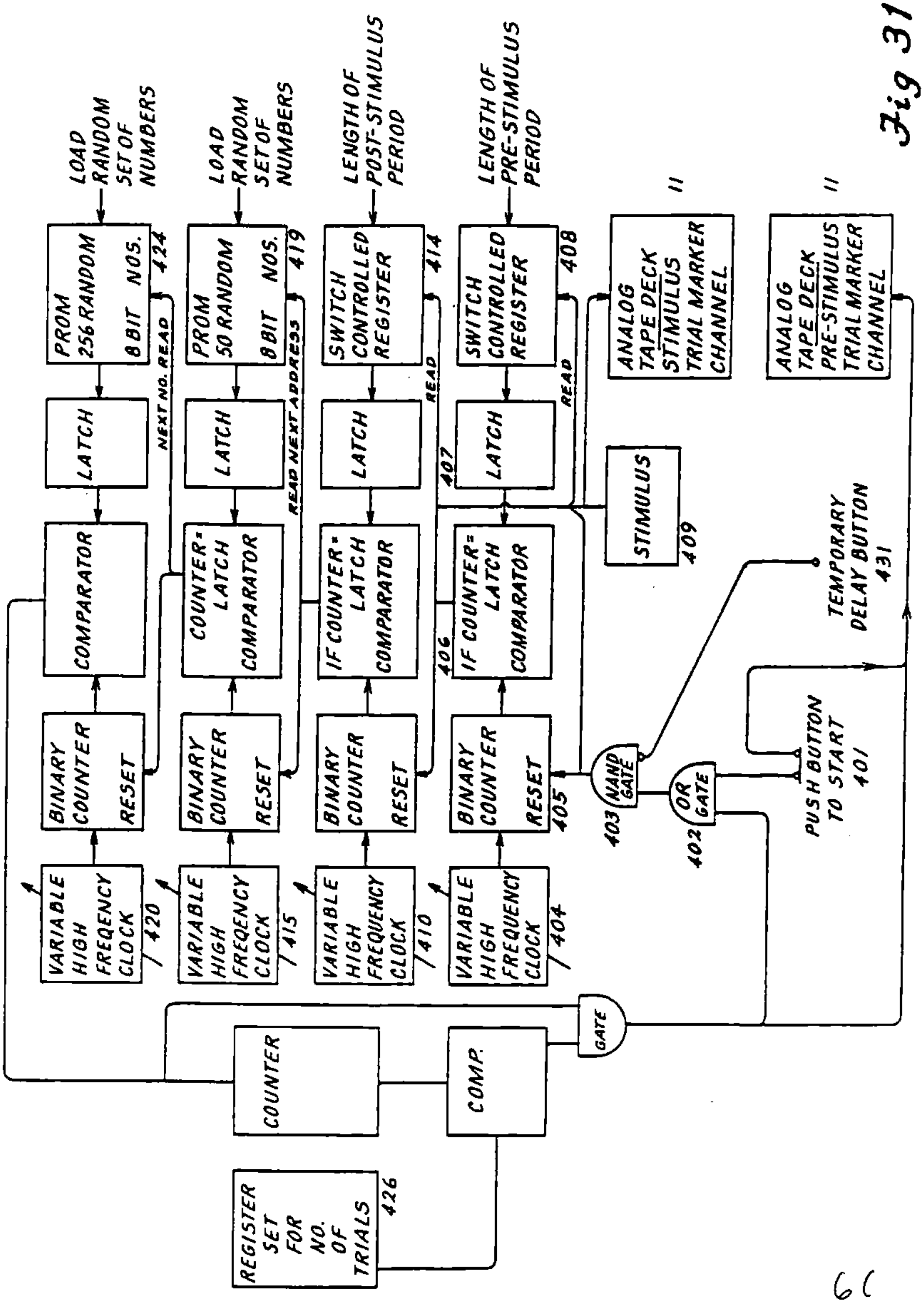
LOAD RANDOM SET OF NUMBERS
PROM 256 RANDOM 8 BIT NOS.
424
LOAD RANDOM SET OF NUMBERS
PROM 50 RANDOM 8 BIT NOS.
419
LENGTH OF POST-STIMULUS PERIOD
SWITCH CONTROLLED REGISTER
414
LENGTH OF PRE-STIMULUS PERIOD
SWITCH CONTROLLED REGISTER
408
ANALOG TAPE DECK STIMULUS TRIAL MARKER CHANNEL
ANALOG TAPE DECK PRE-STIMULUS TRIAL MARKER CHANNEL
LATCH
LATCH
LATCH
LATCH
NEXT NO. READ
READ NEXT ADDRESS
READ
READ
407
STIMULUS
409
COMPARATOR
COUNTER= LATCH COMPARATOR
IF COUNTER= LATCH COMPARATOR
406
IF COUNTER= LATCH COMPARATOR
TEMPORARY DELAY BUTTON
431
BINARY COUNTER RESET
BINARY COUNTER RESET
BINARY COUNTER RESET
BINARY COUNTER RESET
405
NAND GATE
403
OR GATE
402
PUSH BUTTON TO START
401
VARIABLE HIGH FREQENCY CLOCK
420
VARIABLE HIGH FREQENCY CLOCK
415
VARIABLE HIGH FREQUENCY CLOCK
410
VARIABLE HIGH FREQUENCY CLOCK
404
GATE
COUNTER
COMP.
REGISTER SET FOR NO. OF TRIALS
426

Fig 31

6(

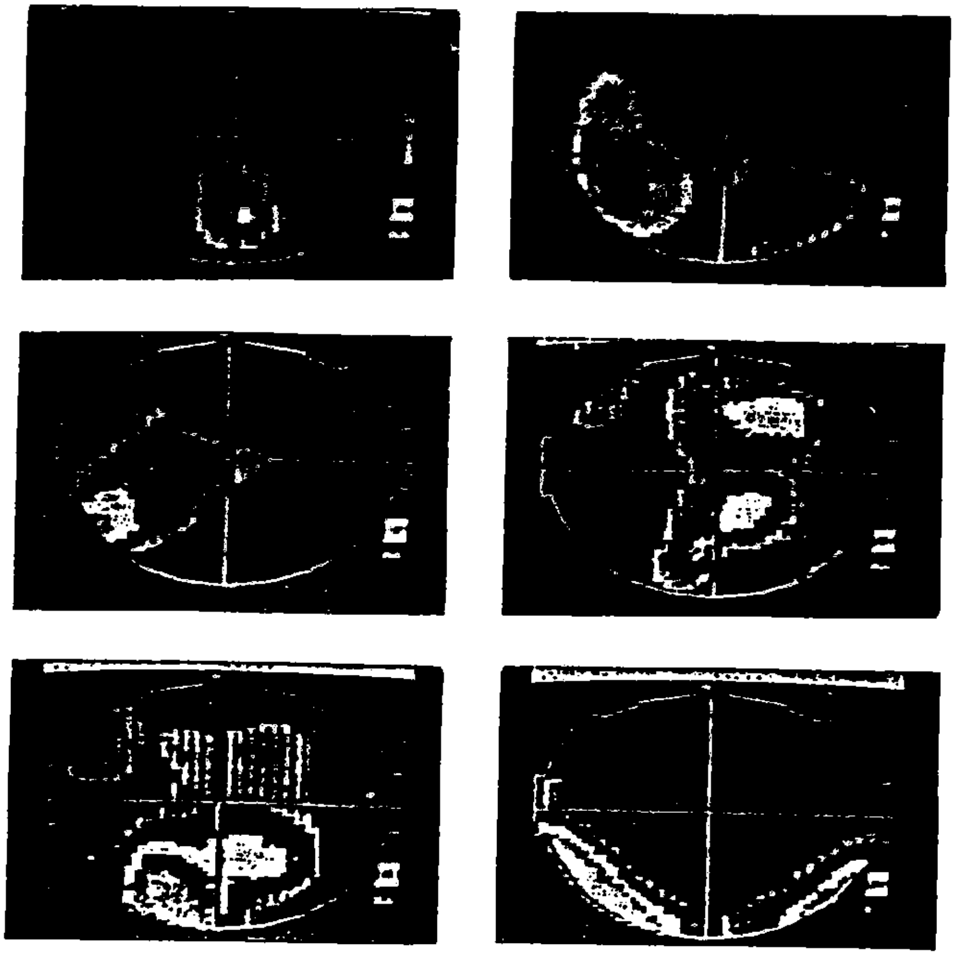

FIG. 32

BRAIN ELECTRICAL ACTIVITY MAPPING

The invention described herein was made in the course of work under a grant or award from the Department of Health and Human Services.

BACKGROUND OF THE INVENTION

This invention relates to analysis of brain electrical activity and diagnosis of brain disorders.

Traditional electro-encephalographic (EEG) techniques of analyzing brain electrical activity to diagnose brain dysfunction require the skilled neurophysiologist to observe and distinguish time and frequency related characteristics of many channels of voltage waveforms derived from an individual's brain and to determine, largely from memory, differences between that individual's waveforms and waveforms characteristic of a normalized population. The process necessarily fails to take account of many subtle but potentially useful pieces of information contained in the analyzed data.

Signal averaged sensory evoked potential (EP) transient responses have also been used as a source for brain electrical activity analysis, but large amounts of useful information contained in such transient response waveforms have traditionally been disregarded because of the difficulty of visualizing the inter-relationship over time of many channels of such information.

SUMMARY OF THE INVENTION

The invention features, in one aspect, displaying time sequences of topographic maps at a variable frame rate. In preferred embodiments, the rate can be selected to display protions of the EP response immediately following the stimulus at a slower rate than later portions; and in preferred embodiments the rate can be varied logarithmically. The variation of display rate permits the operator to give more emphasis to matrices which contain relatively more information, such as the earlier EP response matrices.

In another aspect, the invention features generating topographic displays of information on electrical activity of the brain produced at a plurality of transducers on the skull; generating a time sequence of matrices of electrical activity at successive points in time sufficient in number to capture the onset of a rapid transient event; and displaying the matrices as topographic maps in time sequence at a variable rate. In preferred embodiments, the processor is capable of generating 200 or more matrices for each second of real time. The ability to capture a large number of matrices in a short period of time permits the observation of short-term events such as epileptic spikes.

In another aspect, the invention features generating topographic displays of information on brain electrical activity produced at a plurality of skull transducers; storing the electrical activity of such transducers for two different brain conditions; generating matrices of elements representing electrical activity in the two conditions; forming a difference matrix between corresponding elements of the two matrices; and displaying the difference matrix as a topographic map. In preferred embodiments, the two brain conditions are attained by the use of a patterned light stimulus and a non-patterned light stimulus. The ability to form and display difference matrix enables the operator to identify parts of the brain involved in particular brain states or evoked responses.

In another aspect, the invention features generating topographic displays of information on brain electrical activity produced at a plurality of skull transducers; repeatedly triggering EP responses at the transducers, including pre-stimulus and post-stimulus responses; averaging the responses; setting as a baseline the mean level of the pre-stimulus response; generating matrices from such responses; and displaying topographic maps of the matrices. In preferred embodiments the time of occurrence of each stimulus is stored and the response is divided into pre-stimulus and post-stimulus periods; the matrices generated are a set of time-sequenced frames during the response and are displayed as a sequence of topographic maps; the sequence can be displayed as an endless sequence of maps; the averaging process can be performed using digital words added into summing buffers; and the sampling, storing, response averaging, baseline calculation and subtraction can all be performed digitally; averaging of the pre-stimulus baseline can exclude selected portions of the response; the response can be reviewed to determine the appropriateness of the baseline; there can be calculated the V_{RMS} of the average pre-stimulus and post-stimulus responses and the V_{RMS} can be displayed with the responses, so that the user can determine whether the noise level for any transducer is unacceptably large; the operator can manually adjust the baseline up or down; high-frequency components can be filtered from the post-stimulus response by multipoint interpolation; and the baseline calculation can be repeated until the results are satisfactory. In general, these various features permit the operator to assemble, modify and adjust to an accurate zero level a set of EP responses so that the ultimate topographic display will be accurate and useful. The display of a time-sequence of frames permits the operator to visualize the movement of brain activity in the course of an EP response over the skull. The proper setting at the baseline improves the utility of each response when used in a topographic display, since the relative levels of response at different transducers is more accurately portrayed.

In another aspect, the invention features filtering to remove from the EEG responses frequency components outside the prominent frequency bands of electrical activity; determining, for each transducer, the Fourier transforms and the spectral energy in selected frequency bands, during a period when the brain activity remains in the same state; and processing the results into display matrices for the selected frequency bands. In preferred embodiments, the brain activity can be sampled, stored and Fourier transformed digitally; the filters can remove frequency components below 0.5 Hz and about 50 Hz; the samples can be taken at least 3 times as frequently as the highest frequency in the prominent frequency bands, and particularly at 4 to 5 times that highest frequency; the Fourier analysis can be limited to a period between marked starting and stopping points; the period during which electrical activity is sampled can be limited to avoid interruptions in the subject's brain state, and particularly can be limited to two second sampling periods; the number of samples can be between 20 and 2000; and the frequency bands analyzed can comprise the alpha, beta, delta and theta bands. Removal of irrelevant frequency bands, sampling at high rate, and limiting the sampling period all enable the operator to obtain accurate spectral analyses with minimum interference. The ability to analyze specific frequency bands of interest enables the operator to review information which effectively corresponds to the electrical activity of the brain in various states.

In another aspect, the invention also features generating a statistical comparison matrix from two matrices, each element of the statistical comparison matrix representing a statistical difference between the corresponding elements in the two matrices; and displaying the statistical comparison matrix as a topographic map. In preferred embodiments, the statistical comparison matrix can be interpolated into a display matrix having additional display points; the statistical comparison can be made between two expanded matrices rather than between two unexpanded matrices; the statistical comparison can be a t-statistic analysis, or a z-statistic analysis; and quantitative features useful for diagnosis can be determined from regions of the maps. The ability to perform and topographically display statistical differences between groups and between an individual and a group offers a versatile and effective tool for visualizing brain areas which are connected to particular brain dysfunctions or to particular brain activities, and for neurophysiological diagnosis and research.

In another aspect, the invention also features generating a coefficient-of-variance matrix, each element of which represents the normalized standard deviation at one skull location; and displaying the coefficient-of-variance as a topographic map.

In another aspect, the invention also features temporally interpolating matrices which represent the response at time instants between other matrices; and displaying said interpolated matrices. The temporal interpolation provides a smoother visual transition between the original frames when a time-sequenced display is presented.

In another aspect, the invention also features previewing waveforms and tagging a waveform to indicate whether it should be used in later processing, eliminating a response from further processing, automatically eliminating a response from further processing if a portion of the response exceeds a predetermined threshhold, smoothing a response by eliminating undesired high-frequency components, for adjusting the zero baseline of a response, eliminating selected portions of a response from further processing, and displaying in numerical form the value of a response at a point in time selected by the operator. These waveform quality control procedures enable the operator to improve the quality and accuracy of the topographic displays.

In another aspect, the invention features generating a topographic display of information on the electrical activity produced at a plurality of skull transducers; sampling and storing the information as a series of matrices; viewing the data as a waveform; adjusting or eliminating portions of the data, processing the matrices into processed matrices; interpolating to expand the matrices for viewing; and displaying the matrices as topographic maps in a grey tone scale. In preferred embodiments, the data matrices, processed matrices and expanded matrices can be tagged and stored for later recall and processing; and the data matrix elements can be calibrated to stored calibration signals by calculating a DC offset and gain component for each transducer. The ability to store display matrices for later use enables the operator to accumulate a series of significant matrices derived from diagnostic or research work. The calibration assures that the topographic displays will be accurate.

In another aspect, the invention features generating a topographic display of electrical activity of the brain produced from a plurality of electrical transducers on the skull; generating matrices of elements representing the electrical activity at different points; using a video monitor to display said matrices, each element of which is represented by a discrete point having a gray tone of color lying within a range of gray tones; and scaling the elements to the tones. In preferred embodiments, scaling can be performed so that all elements are linearly interpolated between the maximum gray tone and the minimum gray tone, or between the maximum gray tone and a "zero" gray tone, or to an operator supplied gray one, or so that certain display elements are excluded from the scaling, or that elements falling outside the available gray tone range are assigned to the closest gray tone; the gray tones can be generated in two colors, which can be complementary colors, representing values on either side of a zero tone, the zero tone being an absence of color; the scaling can be performed either on a matrix by matrix basis, or for all matrices taken together; the data which forms the input to the scaling operation can be previewed to by selected lab for exclusion from the scaling operation. The ability to scale the display data to a range of gray tones in a variety of ways improves the utility and visual effect of the display. The variety of scaling options is suitable for the variety of data which may be displayed. The ability to eliminate very large values from the scaling operation assures the most effective scaling for a given set of data.

In another aspect, the invention features normalizing display matrix elements to a selected value which can be assigned to a selected gray tone for purposes of scaling. In preferred embodiments, the normalization can be to a matrix element representing a particular vertex transducer on the skull, or to the root mean square value of the background activity at the transducer being normalized, or to the average root mean square value of the background activity at all skull locations; normalization can be done in connection with apparatus for topographically displaying a sequence of matrices representing averaged EP responses to repeatedly provided stimuli; normalization can be done in connection with apparatus for displaying spectral band matrices derived by Fourier transform analysis of the electrical activity, the normalization being of each element of each spectral band to the total spectral energy at the corresponding transducer on the average total spectral energy at all transducers. The ability to normalize display elements improves the operator's ability to compare different sets of data by normalizing them to the same value.

In preferred embodiments, the invention features interpolating to form additional matrix elements between the transducer points; the interpolation can be three-point interpolation, particularly three-point linear interpolation to the values of the three closest transducers; the number of transducers can be in the range of 10 to 200; and the number of picture elements is at least 5 times the number of transducers. The expansion of a matrix of a small number of points to a display matrix of a large number of points significantly improves the smoothness, readability and utility of the resulting topographic displays.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a block diagram of the BEAM system.

FIG. 2 is a representation of the organization of samples of data in the brain electrical activity mapping system.

FIG. 3 is a representation of the formation of a topographic display from a frame of data in the brain electrical activity mapping system.

FIG. 4 is a block diagram of the functions performed by the BEAM system.

FIG. 5 is block diagram of the define protocols operation.

FIG. 6 is a block diagram of the disk data sampling operation.

FIG. 7 is a data file format diagram of the raw data file.

FIG. 8 is a block diagram of the core averaging operation.

FIG. 9 is a graph of an average EP transient response waveform after automatic baseline zeroing.

FIG. 10 is a block diagram of the raw data quality control operation.

FIG. 11 is a block diagram of the raw data topographic display operation.

FIG. 12 is a block diagram of the create interpolation file operation.

FIG. 13 is a block diagram of the raw data reduction operation.

FIG. 14 is a data file format diagram of signal averaged EP data.

FIG. 15 is a data file format diagram of FFT ensemble data.

FIG. 16 is a data file format diagram of individual FFT data.

FIG. 17 is a data file format diagram of an EP file.

FIG. 18 is a data file format diagram of an FFT file.

FIG. 19 is a block diagram of the reduced data quality control operation.

FIG. 20 is a graph of an average EP transient response waveform after automatic baseline zeroing and after manual baseline readjustment.

FIG. 21 is a block diagram of the reduced data topographic operation

FIG. 22 is a block diagram of the group file production operation.

FIG. 23 is a block diagram of the group topographic display operation.

FIG. 24 is a block diagram of the individual vs. group comparison operation.

FIG. 25 is a block diagram of the group difference detection and feature selection operation.

FIG. 26 is a data file format diagram of a saved frame file.

FIG. 27 is a block diagram of the brain electrical activity mapping to TICAS file transfer operation.

FIG. 28 is a block diagram of the TICAS feature selection and evaluation operation.

FIG. 29 is a block diagram of TICAS generate decision rules operation.

FIG. 30 is a block diagram of TICAs test decision rules operation.

FIG. 31 is a block diagram of pseudorandom stimulus controller.

FIG. 32 is a sample of topographic displays generated by a brain electrical activity mapping system.

DESCRIPTION OF THE PREFERRED EMBODIMENT

We now turn to a description of the preferred embodiment.

System Organization and Software

FIG. 1 illustrates the components of a brain electrical activity mapping system. Twenty electrodes 5 (e.g., Grass gold cup) are attached to subject's skull 4 in a conventional international 10–20 format. Twenty leads 6 from electrodes 5 are connected through switch 7 to conventional 24-channel polygraph 10 (e.g., Grass 8-24D), which contains parallel variable gain differential amplifiers and strip chart recorders. Calibration signal source 8, an A.C. generator, is also connected through switch 7 to polygraph 10. Stimulus A 2 (e.g., Grass Model PS1 strobe light) and stimulus B 3 (e.g., click generator) present stimuli to the subject under the control of pseudorandom stimulus controller 9, which also provides pre-stimulus and stimulus trial marker signals (5 volt spikes) of opposite polarity to one of the input channels to 24-channel FM analog tape recorder 11 (e.g., Honeywell 5600E). In other embodiments, recorder 11 is eliminated and polygraph 10 is connected directly to filter 12 for real-time loading of data. The 21 active outputs of recorder 11 are connected to the inputs of 21 parallel variable band pass filters 12 (e.g., Butterworth filters; EEG Associates Mark 4×24) having variable gain controls. The 21 outputs of filters 12 are connected to 21 of the input terminals of two 16-channel, 12-bit analog-to-digital converters 15, 16 (Digital Equipment Corporation AA-11K), which comprise part of digital computer 13 (Digital Equipment Corporation PDP 11/60). Analog-to-digital converters 15, 16 are attached to data bus 14 (Digital Equipment Corporation Unibus). Also attached to data bus 14 are 4-channel, 12-bit digital-to-analog converter 17 (Digital Equipment Corporation AD-11K) whose three outputs control black and white television monitor 18 (Digital Equipment Corporation VR 17) for waveform displays; color display control 19 (Digital Equipment Corporation VSV 01) whose three outputs control 12" color television monitor 20 (CONRAC) for topographic displays; 8 serial line controller 24 (Digital Equipment Corporation DZ 11) two outputs of which control interactive keyboard and video character display terminal 22 (Digital Equipment Corporation VT 100) and printer 23 (Digital Equipment Corporation LA 120); 256K byte memory 24 containing operating system software 27 (Digital Equipment Corporation RSX 11/M), BEAM software 28 (Agrippa Data Systems), and analytic software 29 (TICAS; University of Arizona); floating point processor 25 (Digital Equipment Corporation FPP-11); central processing unit 26 (Digital Equipment Corporation PDP 11/60); and disk controller 27 controlling at least one disk drive 28.

Software Description

In general, the brain electrical activity mapping system creates color topographic displays reflecting brain electrical activity using, as input, continuous electrical waveforms recorded from a number of points on the skull. The color topographic displays consist of discrete matrices of a large number of display points (also called pixels), each of which has a color or intensity or other visible characteristic which indicates a certain value or values at the location of that point analogous to a point on the skull. In order to generate discrete topographic display matrices having many thousands of display points from continuous analog waveforms at a limited, e.g. 20, number of points on the skull, the brain electrical activity mapping system, as illustrated in FIG. 2, converts the data to digital form and generates discrete sample frames 40, each sample or frame initially comprising 20 recorded values 41 from 20 channels of information. The system treats related groups of samples 40 as segments 42. In the case of EP data, for example, a segment would consist of a series of frames or samples, each 4 milliseconds in length, the series together representing one transient response sequence from the beginning of a pre-stimulus period to the end of the post-stimulus transient response. In the case of steady-state EEG data, a segment would consist of 2 seconds of data divided into 256 samples. A spectral analysis of the EEG data then produces 256 samples, each of which reflects the energy level in a small, e.g. ½ Hz, energy band and a segment consists of the entire series of 256 spectral samples. For signal averaging purposes, the system considers a set of segments together, e.g., 500 segments each representing a transient response to a given stimulus. The 500 segments taken together are known as an ensemble 43. Frames of data can be raw data or data which has been processed or transformed by the system. In any case, as illustrated in FIG. 3, when a frame 40 is to be displayed it is expanded into a matrix 45 consisting of a large number of display points 46 which are determined by an interpolation process from the original frame data points 47. Each point of the matrix is then converted to a visual display point 47 which forms part of the final topographic display 48.

FIG. 4 illustrates the organization of the operations which comprise brain electrical activity mapping software 28 and TICAS analytic software 29. Raw and processed data is stored in disk files 51. Operations 52-65 and 67-69 use data stored in files 51 to perform data manipulation, data display and data storage functions. Operations 54 and 55 also process data from the outputs of converters 15, 16.

FIG. 5 illustrates the function of define protocols operation 53. Protocol files 73 are generated and edited by program 'SETPAR' 71 based on control information 70 provided by the operator through terminal 22, the results of the operation being displayed (block 72) on terminal 22 to the operator. Each protocol file 73 contains information which governs the manner in which other operations are performed on a particular type of data file (e.g., one protocol might apply to the processing of EP transient response data from strobe light stimuli). The protocol information may include the number and identity of input channels, the labeling of the output channels to correspond to specific points on the final display, the identity of the trial marker channel, the voltage level above which to search for the trial markers, the rate in samples per second of sampling of the data, the number of samples in a segment, the number of segments in an ensemble, the number of ensembles, the number of points in a baseline, the microvolt level of the calibration signal (e.g., 100 microvolts at 10 Hz), a multiplication factor, the number (up to 20) and size (width) of integration bands, the label of the protocol, the percent of taper of samples in a segment of data for fast fourier transform processing, the number of automatic smoothing passes, the high and low values for automatic rejection of data during accumulation, the stimulus interval and location in seconds, and channel labels related to electrode positions on the skull.

FIG. 6 illustrates the function of disk data sampling operation 54. Program 'DATCOL' 75 loads raw data from the output of converters 15, 16 into raw data file 79, which is divided into 19 buffers 80 which hold ensembles of data related to particular brain states or stimuli. The operator provides control information 76 designating the patient to whom the data relates and the name of the applicable protocol file 73. Other control information 76 governs the beginning, end, and pauses in data sampling, and the performance of a calibration of signal levels. Calibration data is initially stored in a buffer 80 of raw data file 79. When the operator requests (block 76) a calibration, and designates which buffer 80 contains the raw calibration data, program 'DATCOL' computes the root mean square value and the mean of at least 30,000 points in each channel of calibration data and divides the root mean square value by 0.707 to establish the assumed peak value of the calibration signal. The peak value, representing the level of the original calibration voltage, and the mean value, representing the D.C. offset of the calibration voltage value for each channel, are stored in calibration file 78.

The format of raw data as stored in raw data file 79 is illustrated in FIG. 7. Each buffer 80 contains header block 81, having protocol and other housekeeping information concerning the data stored in the buffer; calibration block 82 containing for each channel of data the calibration value in microvolts per bit and the number of microvolts by which the calibration signal was offset from zero both of which values were found in calibration file 78 at the time raw data was loaded; bad segment block 83 identifying segments of data which the operator will later decide to exclude from subsequent operations; an unused segment 84; and a series of data segments 85, which hold a series of data samples, each containing values for all 20 channels. The data segments 85 are interleaved with no gaps.

FIG. 8 illustrates the function of core averaging operation 55, usually usef for loading and signal averaging raw EP transient responses. Data from converters 15, 16 is read by program 'CORAVG' 86. User provided control information 88 designates the protocol, obtained from protcol file 73, under which the operation is performed, and determines start, end, and pauses of the operation. Program 'CORAVG' 86 samples data beginning at points labeled by the prerecorded trial markers and forms signal averaged EP transient responses from a series of transient responses resulting from repetition of a stimulus. The series of transient response data are accumulated and held in EP file 87, which is a reduced data file as described below. Calibration file 78 holds calibration information accumulated from the data channels in the manner previously described. Program 'CORAVG' 86 automatically rejects as "bad data" any segment which contains values outside of preset limits. Program 'CORAVG' also automatically adjusts the zero baseline with respect to each electrode's average EP transiet response, by subtracting the mean of the pre-stimulus period values for a channel from each point in that channel's transient response curve.

FIG. 9 illustrates a plot of an average EP transient response 90 of microvoltage against time as it could be displayed on monitor 18 following core averaging operation 55. The stimulus was presented at time 93, the transient response includes pre-stimulus period between time 92 and time 93, and the plot shows calculated zero baseline 91.

FIG. 10 illustrates the function of raw data quality control operation 56, which enables the operator interactively to review and eliminate bad segments of raw data before other operations are performed. By means of control information 94 the operator can select for review the contents of any buffer 80 in raw data file 79. The buffer data is displayed (block 105) segment by segment by program 'GLITCH' 95 on television monitor 18 to the operator as an analog waveform. The operator can label any segment of bad data, which causes the bad data segment to be identified on bad segment block 83. Control information 94 can also include the insertion of trial markers indicating a point on a waveform at which subsequent operations should begin, and display to the operator the microvolt value of individual points on a displayed curve.

FIG. 11 illustrates raw data topographic display operation 57, which provides topographic time sequenced displays (cartoons) of raw data frames. Program 'RAWMOV' 96 expands the 20 channels of each data frame into a matrix of 128×128 data points by three-point linear interpolation. The operator provides control information 102 designating the disk buffer 80 on raw data file 79 which contains the data to be displayed; the number of display matrices to be produced; the parameters for interframe interpolation; and the parameters and options (described below) for scaling the data points among the available grey color tones of the display. Program 'RAWMOV' 96 calculates each interpolated data point for the display matrix using three-point linear interpolation from the three closest original channels and scales the data to the available grey color tones of the display. The interpolation is performed using preset coefficients stored in interpolation file 97 by an operation described below.

The display matrices produced by program 'RAWMOV' 96 are stored in sequence in disk movie file 98. Program 'RAWDIS' will display (block 101) the frames stored in disk movie file 98 on monitor 20. Control information 100 permits the operator to designate the file to be displayed, the frame rate, and the starting, stopping and reversing of the display sequence. The displays include labels of information taken from the protocol block, e.g., patient identification.

FIG. 12 illustrates the functions of create interpolation file operation 52. Program 'POINTS' 110 creates points file 115 reflecting the X and Y coordinates of each point in the original electrode layout with respect to the 128×128 grid and associating with each point in the 128×128 display matrix the identity of the three original electrode points with respect to which it should be interpolated. Control information 120 provided by the operator includes the X and Y coordinates of each channel and the identity of the three interpolation points for each display point. Program 'OUTLINE' 111 identifies and stores in outline file 116 the X and Y coordinates of the points which outline the plan view of the skull to be included in the display, based on control information 121. Program 'MAKCOF' 112 generates and stores in coefficient file 117 the coefficients needed to perform the three-point linear interpolation for each matrix display point within the skull outline, using points file 115 and outline file 116 as input. Program 'MAKOVR' 113 stores in overlay file 118 the operator provided (block 123) coordinates of the overlay of the skull, nose and ears outline for the display. Program 'SETOVR' 114 generates interpolation file 119 from overlay file 118 and coefficient file 117. Interpolation file 119 then contains the information required to compute interpolated matrix data points and the skull overlay for display.

FIG. 13 illustrates the function of the raw data reduction operation 58. Three alternative programs can operate on data in buffers 80 to produce reduced data files 125. Program 'DSKFFT' 126 accepts segments of EEG data from raw data file 79, performs a fast fourier transform analysis which produces a new segment of data reflecting the spectral energy in each of a sequence of frequency bands. Program 'DSKFFT' 126 also generates, for each group of segments, an ensemble consisting of the sums (used in a later step to form the average values) and sums squared (used in a later step to form the standard deviations) for each channel across all segments in the group, values reflecting each of the sums as a percentage of the total spectral energy in the segment, and values reflecting the coefficient of variation (the standard deviation divided by the mean) for each channel across an ensemble. FIG. 15 illustrates the format of the resulting ensemble of FFT data stored in reduced data file 125. The sums data 130 is filed in sequence by channel for the first frequency band 131, e.g., 0.5 Hz. Similar sums data follows for the other frequency bands. After all sums data is stored, the sums squared data 132, the normalized power spectral density sums, and the coefficient of variation data are stored in similar fashion. In addition to storing the sums and sums squared data for all segments in the ensemble, program 'DSKFFT' can store spectral information for each segment analyzed. As illustrated in FIG. 16, the data is stored as sine and cosine coefficients for each channel for each frequency band, and as normalized sine and cosine coefficients as a percentage of total spectral energy. As illustrated in FIG. 18, the FFt data file 191 stored on reduced data file 125 also includes a header block 192 housekeeping information.

In FIG. 13, program 'DSKAVG' 127 performs a function similar to core averaging operation 55 in signal averaging EP transient response waveforms, but uses as input raw data stored in raw data file 79 and permits the operator to review each waveform and select those to be used in the averaging process, rejecting others. Program 'MANAVG' 128 permits a similar operator-assisted signal averaging process when the raw data does no contain preset stimulus trial markers, requiring the operator to indicate the point at which averaging is to begin for each waveform. FIG. 14 illustrates the format of signal averaged data produced by programs 'CORAVG' 86, 'DSKAVG' 127 and 'MANAVG' 128. The sums of each channel for all trials for the first time frame 133, e.g., 0–4 milliseconds, are loaded in order, followed by similar information with respect to all subsequent time frames for a given segment. As illustrated in FIG. 17, such EP files 190 are preceded by header block 193.

FIG. 19 illustrates the function of reduced data quality control operation 59, which permits the operator interactively to review and modify data in reduced data file 125. By providing control information 204, the operator can select the file to be reviewed and indicate whether it contains FFT spectral information or EP time-sequenced information. For FFT information, program 'FFTLUK' 201 displays (block 202)selected channels of spectral data as frequency-voltage curves on monitor 18 and permits the operator to low-pass filter the waveforms and display the value of particular data points. For EP data, program 'EPLUK' 203 displays (block 202) selected channels as time-voltage curves and permits the operator to reset the zero baseline, to filter high frequency noise from a channel, and to display the value of any point on a curve. FIG. 20 illustrates the function of manual baseline relocation. Because pre-stimulus period response 180 was not level, automatic baseline 176 set by program 'CORAVG' 86 inaccurately reflects the true zero level for transient response curve 175. The operator can relocate the baseline to a new level 177 by moving cursor 179 to the desired level, causing that voltage value to be subtracted from each point of data along curves 180 and 175.

FIG. 21 illustrates the functions of reduced data topographic display operation 60. For single frame display of FFt data, program 'FFTTVM' 206 reads data from reduced data file 125 as selected by operator control information 214. The data is scaled in accordance with instructions included in control information 214. The selected frame is provided to program 'TVM16' 213 which interpolates a matrix of 128×128 points using the coefficients and other information contained in interpolation file 97, and provides the resulting matrix to color display 212. For single frame EP display, program 'EPTVM' 205 performs an analogous process to that of program 'FFTTVM' 206. Programs 'EPTVM' 205 and 'FFTTVM' 206 also perform compilations of sequences of frames into one display matrix, in accordance with predefined groupings set forth in protocol blocks.

A sequence of FFT matrices or EP matrices can be displayed in rapid time sequence as a cartoon by the use of program 'FFTMOV' 208 and program 'EPMOV' 207, respectively, each of which processes sequences of selected matrices of data from reduced data file 125, using scaling control information 215; interpolates full 128>128 matrices for each frame; interpolates a selected number of additional matrices between the original frames; and stores the resulting matrices in movie file 211. Based on control information 216 specifying data to be displayed, the frame rate of display, start, stop, backward, forward and pause, program 'FFTDIS' 210 and program 'EPDIS' 209 provide cartooned matrices for viewing on color display 212.

FIG. 22 illustrates group file production operation 61. Program 'GRPBLD' 221 creates and updates a composite group file 223 working from selected individual reduced data files 125. Control information 222 provided by the operator indicates the identity of individual reduced data files to be included. The group files 223 consist of the sums and the sums squared for all homologous points in the reduced data files 125 of all individuals in the group. Normalized sums and coefficients of variation may also be produced and stored. Hard copy 224 listing the individuals in each group and the values of group file data are available based on control information 222.

FIG. 23 illustrates the function of group topographic data display operation 62 which is analogous to the operation of the single frame display of reduced data file information, illustrated in FIG. 21, except that the data displayed is from group file 223 instead of reduced data file 125.

FIG. 24 illustrates the function of individual versus group comparison operation 63. Programs 'EPTVM' 205 and 'FFTTVM' 206 are performed respectively on EP and FTT data. In each case, the program generates a frame of points, each of which is the number of standard deviations (z-statistics) by which an individual's point, taken from individual file 226 differs from the average of the group's corresponding points taken from group file 223. The resulting frame is displayed (block 216) by program 'TVM16' 213, which interpolates additional data points to form a 128×128 matrix in the manner previously described.

FIG. 25 illustrates the function of group difference detection and feature selection operation 64, which computes frames of t-statistics reflecting the level of statistical difference between two groups based on the means and standard deviations of homologous data points for the two groups. Program 'GRPCMP' 230 computes t-statistics and degrees of freedom from the sums and sums squared data contained in two different group files 223 designated in control information 231 provided by the operator. The resulting frames are stored in group comparison file 232. Based on file and frame designations set forth by the operator in control information 235, program 'GRPDIS' 234 transmits a selected t-statistic frame for display by program 'TVM16' 213 as previously deascribed. Program 'TVM16' 213 also returns a fully expanded 128×128 matrix back to program 'GRPDIS' 234. Through control information 235, the operator may store such a t-statistic matrix in saved frame file 233. FIG. 26 illustrates the format of saved frame file 233. Header block 240 contains the number of saved frames, and for each frame identifies the time frame length or frequency band and the protocol under which the data was collected. Frames 241 contain 128×128 matrices of floating point t-statistics generated by program 'GRPDIS' 234.

FIG. 27 illustrates the functions of the brain electrical activity mapping to TICAS file transfer operation 65, which converts features of brain electrical activity mapping data to a form usable with TICAS software 29. Program 'FEADEF' 245 generates feature list file 247 from feature definitions and names of saved frame files 234 provided by the operator in control information 246. Program 'FEAGET' 248 then generates TICAS formal file 249 from reduced data files 125, saved frame file 233, and feature list file 247, identified by the operator in control information 250. The feature definitions which may be selected and stored in TICAS format file 249 for subsequent TICAS analysis include nearly all combinations of data found in all individual files, all combinations of integrated bands of frames (e.g., alpha bands) in individual files grouped according to preset protocol specifications, or combinations of individual frame files weighted by the values in a saved t-statistic frame.

FIG. 28 illustrates the functions of TICAS feature selection and evaluation operation 67. In control information 266, 264 the operator designates two groups 262, 263 from TICAS format file 249 whose features are to be analyzed, the type of data to be analyzed, the number of input features to be analyzed, and the number of output features to be produced. Program 'UTEST' 261 peforms a standard two-sample Wilcox-Mann-Whitney U-test and provides a list of U-test scores in rank order which are printed on printer 23 and stored. Program 'FMTST' 260 performs a final merit value (FMV) test and provides and stores a list of values for features in order of final merit values. Program 'FMTST' 260 first determines intermediate merit values as a combination of the standard receiver operating characteristic (ROC) curve or d' value and the results of an ambiguity function analysis. The final merit values for each feature are then determined by correlation of the intermediate merit value with all features ranked higher in intermediate value. Program 'MERGE' 267 selects from Group

A features 268 and group B features 269, based on the FMV values and U values, those features which are most useful, which are then stored in disk files 270 and 271.

FIG. 29 illustrates the functions of TICAS generate discision rules operation 68. Program 'SLPRT' 275 uses an operator controlled (block 278) subset of selected Group A features 276 and selected Group B features 277 to generate disk stored interim rules 279, by a method controlled (block 278) by the operator. The computation method may be a non-parametric d-selection technique, a parametric classification technique, or a combination of the two. Program 'CLSFY' 284 uses operator selected (block 285) interim rules 279 to classify operator selected subjects having operator selected features, from selected group features files 280 or refined group features files 281, 282, printing the results on printer 23, thereby permitting the operator to evaluate the efficacy of each selected interim rule 279. Progran 'DSC' 286 combines operator chosen (289) features from selected Group A features 287 and selected Group B features 288 into new features on the basis of the weighting functions of a standard linear discriminate analysis, subject to the operator's choice (group 289) of discriminant function parameters. Discriminant function weights for the best features are then loaded into a disk file 290. Program 'LINCOM B' 293 uses the original selected Group A features 291 and selected Group B features 292 and the discriminate function weights 29 to create refined Group A features 294 and refined Group B features 295, which are linear or other operator selected (block 296) combinations of the original features. Refined Group A features 294 and refined Group B features 295 can replace the original selected features 276 and 277 as input to program 'SLPRT' 275 to permit an iterative process of decision rule generation.

FIG. 30 illustrates test decision rules operation 69. Program 'CLSFY' 284 uses final sets of decision rules, e.g., rules A 304 and rules B 305, to classify individuals in original Group A 300, original Group B 301 and new unknown groups, e.g., Group C 302 and Group D 303, to determine the efficacy of the final decision rules, the results being provided on printer 23.

Operation

Data Gathering

Accumulation of raw EEG and EP data is accomplished by first attaching 20 electrodes 5 to the scalp of an individual subject in a conventional international 10-20 format. In other embodiments, information from between 10 and 200 electrodes can be gathered and analyzed. Before recording, and if desired during recording, the operator observes the signal levels on the 20 channels of chart recording of polygraph 10 and adjusts the gain on weak signals to produce usable waveforms. A calibration signal of 100 microvolts (10 Hz) from source 8 is recorded on all twenty channels on tape recorder 11 at the beginning of each session and whenever any of the gain levels on polygraph 10 is adjusted.

Data gathering typically begins with a careful administration of a sequence of tests, each of which is intended to establish a particular steady-state electrical condition in the subject's brain. The sequence of tests usually includes instructions to relax and remain still with eyes open, to relax and remain still with eyes closed, to become drowsy, to breathe deeply for hyperventilation, to listen to music and to listen to speech. Other tests rewquire the subject to (1) listen carefully to a story and answer simple questions about its content when completed, (2) remember a set of six abstract figures (often resembling an unknown language alphabet) in black ink on index cards presented by the examiner, (Kimura Figures-Instruction) (3) select the six previously presented figures from a set of 38 figures, verbally indicating yes or no (Kimura Figures-Test), (4) associate each of four abstract figures on index cards with a particular artificial name spoken by the examiner (Paired Associates-Instruction), (5) name each of the four abstract figures when tested by the examiner (Paired Associates-Test), (6) read silently three previously unread paragraphs (e.g., example text from the Gray Oral Reading Test) so as to answer questions subsequently (Reading Test-Instruction), (7) identify whether 34 typed sentences presented by the examiner were previously included in the three paragraphs (Reading Test-Test), and (8) read text upside down. The tests are designed to permit recording of brain electrical signals during simple resting brain activity and during different levels of activation of the left hemisphere, the right hemisphere and both hemispheres of the brain together. This permits the demonstration of pathologies present at rest and those present upon brain activation. The development of specific tests and the choice of tests is determined by the user based in part on the subject being tested and the information eing sought as described in greater detail below. Between twenty seconds and three minutes of steady state brain electrical activity is recorded on all 20 channels during each of the tests. Appropriate records of the tape location of each test are kept. These tests have been used with the brain electrical activity mapping system to demonstrate group differences between normal subjects and those with dyslexia or specific reading disability at the 10-12 year age level and at the six-year age level; to differentiatedemented patients from normals and aged patients from younger patients in clinical settings; to identify patients with an organic basis for sociopathic behavior and other forms of mental illness; to demonstrate epilepsy when the resting background EEG failed to show any abnormalities; to demonstrate abnormalities in EEg and EP data for schizophrenic subjects; and to determine when a brain tumor, previously treated, is about to recur.

With young infants, the brain states tested include sleep, alert and attending to visual and auditory stimuli, alert but not attending to visual and auditory information, and drowsiness. Using these states, it is possible to discriminate among children with poor behavioral scores on a psychological test and those with high psychological scores.

A series of sensory evoked potential (EP) transient responses are then recorded from all electrodes while the subject is repeatedly exposed to a selected stimulus, e.g., a strobe light or a click generator or to a predetermined sequence of two alternate stimuli. Because the EP transient response is weak compared to the background steady-state brain electrical activity, the stimulus must be presented many times (e.g., 500) to the subject for later signal averaging. The total response period of interest is typically 1024 milliseconds, comprising 512 milliseconds before stimulus and 512 milliseconds after stimulus. The process is repeated for different stimuli.

Stimuli presented to the subject can range from simple flash, simple click, simple pattern reversal and simple somatosensory stimulation to those requiring complex decisions. Requiring a subject to discriminate between subtly different auditory stimuli (e.g., the words "tight" and "tyke") is useful in diagnosing dyslexia. This procedure is known as the Tight-Tyke evoked potential phenomic discrimination test. Picking an infrequently different stimulus from among other more frequent stimuli is useful evaluating subjects who have functional brain disorder.

Auditory stimuli generate a set of fast and a set of slow transient responses. The fast responses eminate from the brainstem and have a typical duration of 20 milliseconds. Brainstem responses are normally sampled for a total response period of 40 milliseconds comprising 20 milliseconds before stimulus and 20 milliseconds after stimulus. Filters 12 are adjusted to exclude frequencies below 300 Hz and to include frequencies up to 8000 Hz.

When EP transient responses from such stimuli are averaged to eliminate noise, two types of interference can occur. The first type, known as contingent negative variation (CNV), relates to the connection made by the brain between consecutive equally spaced stimuli when the subject is told to count the stimuli. The D.C. component of the resulting transient response shows a gentle dip and a sharp rise immediately before each stimulus, attributable to the subject's anticipation of the next stimulus. The sharp rise contaminates the evoked potential transient response and makes it difficult to establish a zero baseline. By including as part of the interval between stimuli a first pseudorandom time element which varies from 0 to a period longer than the post-stimulus response and is also a multiple of the wavelength of the interfering frequency described below, the CNV effect is greatly reduced.

The second type of interference results from the existence of background noise at certain characteristic frequencies, e.g., 10 Hz, which reflect prominent bands of steady-state brain wave activity. The major interfering frequency of a given subject may be determined by a spectral analysis of his background EEG signal. The interference problem is especially significant in adults with prominent alpha waves and children with prominent slow brain wave activity. By including in the time interval between stimuli a second pseudorandom time element whose period varies from zero to the wavelength of the major interfering frequency, the background noise can be substantially reduced in the averaging process.

The inclusion of a prestimulus period of recoding for each transient response permits an accurate baseline determination at a later stage of signal processing in order to establish a true zero level for the post-stimulus response and permits a determination of the quality of the signal averaging process.

Pseudorandom stimulus controller 9 measures the interval between stimuli as a combination of the post-stimulus response period, the first pseudorandom period described above, the second pseudorandom period described above, and the pre-stimulus response period.

As illustrated in FIG. 31, pseudorandom stimulus controller 9 comprises a four-stage timer, each stage of which in turn measures one of the four periods included in the interval between stimuli. The first stage comparator 406 measures the pre-stimulus period p_1/f_{c1} where P_1 is a number preset by the operator in register 408 and f_{c1} is the frequency set on variable frequency clock 404. When the first stage timing is completed, the second stage times the post-stimulus period as P_2/f_{c2} where P_2 is a number preset in register 414 and f_{c2} is the preset frequency on clock 410. At the end of the first stage timing period, a stimulus trial marker (5 volt spike) is recorded on the trial marker channel of tape recorder 11 and stimulus 409 is triggered. The selected post-stimulus period is long enough to permit a full decay of the transient response being observed. At the end of the second stage timing period, the third stage measures the first pseudorandom time period P_3/f_{c3}, where P_3 is the next pseudorandom number in programmable read only memory (PROM) 419 and f_{c3} is the preset frequency of variable clock 415. PROM 419 has been preloaded with a pseudorandom sequence of numbers. At the end of the third stage timing, the fourth stage measures a second pseudorandom period in an analogous manner based on a pseudorandom sequence in PROM 424 and a preset frequency on clock 420. When the fourth stage timing period is over, a pre-stimulus trial marker (5 volt spike) is recorded on the trial marker channel of recorder 11. A new timing cycle is then completed and the process is reiterated until the total number of transient responses recorded equals a preset number in register 426 or the process is stopped by the operator. The entire process is begun by pushing button 401, which causes the recording of a pre-stimulus trial marker. Temporary delay button 431 can be used to temporarily delay the continued operation of the timer at the operator's discretion, as for example when the subject is distracted in a manner which would render the transient response useless. A low order bit in PROM 424 or PROM 419 can be set to 0 or 1 by the operator for each number loaded into PROM 24 or PROM 19 so that two different stimuli (e.g., auditory and visual) can be triggered in a preselected order or pattern, with one stimuli being presented more frequently than the other.

The result of the recording session is an analog tape of raw EEG and EP voltage data and calibration voltages on 20 channels with trial markers on a twenty-first channel. The next step is to load the data into computer 13.

As previously described, brain electrical activity mapping software 28 performs data collection, data manipulation, and data display functions in accordance with central information provided by the operator. The various operations can be performed in any sequence and the operator can perform a series of functions iteratively. The operator provides control information through the keyboard of terminal 22 and receives information concerning the various operations on printer 23, waveform monitor 18 and topographic color monitor 20. The flexibility of operation heightens the system's utility as a diagnostic and analytical tool.

Data Loading

Under operator control,EEg data for each brain state and related calibration signals are loaded directly onto disk storage from recorder 11 after passing through filters 12 set to pass frequency components between 0.5 and 50 Hz or between 0.5 and 1000 Hz and epileptic spike data. Gain controls on the filters are adjusted to fully utilize the signal capacity of converters 15, 16. EEG data can be sampled at rates as high as 20,000 samples per second.

Under operator control, EP data is passed through filters 12 set to eliminate frequencies above a selected frequency of between 40 and 100 Hz, or below 300 Hz for brainstem data, and is signal averaged in core memory using core averaging operation 55, which automatically rejects bad data and sets the zero baseline.

Typically EP data is sampled every 4 milliseconds, or every microseconds for brainstem analysis, and 256 sampled are taken, 128 pre-stimulus and 128 post-stimulus. If the operator determines that the EP transient response data is very noisy, he may alternatively record the data as raw data and use raw data reduction operation 58 to average only selected transient response trials. In cases where the transient response data may not contain the necessary stimulus trial markers, such as in recordings of rapid eye movement (REM) sleep, the data can similarly be recorded as raw data and trial markers can be added manually by the operator, using raw data operation 58, before signal averaging is done.

Raw Data Quality Control and Display

To assure the maximum accuracy and utility of raw data, the operator can, using raw data quality control operations 56, display recorded waveforms, accept or reject each waveform for later processing, have mathematical smoothing operations performed, reset baselines or eliminate certain points of data.

The operator views the EP transient responses for the 20 channels for the purpose of evaluating the utility of each curve and specifying modifications which will improve their utility when displayed. The operator may direct a further adjustment of the baseline, which has already been set automatically, by having a constant number added to or subtracted from the value in each frame, or can determine to have the automatic baseline determination redone using a smaller number of the later pre-stimulus frames as the baseline. This procedure is particularly useful when the early pre-stimulus frames are found to contain the tail end of the transient response of the prior stimulus. By reviewing the relative levels of V_{RMS} (pre-stimulus) and V_{RMS} (post-stimulus) for a given channel, the operator can determine whether the background noise level is unacceptably large with respect to the transient response, necessitating another recording session with the subject. The operator may also filter any high frequency noise in the post-stimulus period by three-point interpolation.

As part of the raw data quality control process, if a channel contains spurious values (e.g., voltage spikes) in particular frames, the operator can eliminate those values and substitute values interpolated from the next prior and next later frames. If the voltage levels on one channel are substantially higher than for the other channels, the operator can flag that channel to indicate that the channel should be excluded from the subsequent display scaling procedure (described below). Based on the operator's instructions, the automatic baseline determination may be redone and the results are viewed again until the operator is satisfied that the set of transient responses contain satisfactorily low noise levels and are properly zeroed.

Raw data topographic display operation 57 enables the operator to display a cartoon series of topographic maps of raw data, which has been expanded by interpolation into a matrix of 128 × 128 points. The cartoon can be started or stopped and run forward or reversed at will. When raw EEG data is to be cartooned, the operator can sample the data at a high rate, e.g., 400 frames per second, and then display the information at slower speed or in a series of matrices, each of which is an average of a sequence of frames. The averaging can be done on a running basis, so that the first N frames are averaged and displayed, then the N frames beginning with the second frame are averaged and displayed, and so on.

EEG Data Reduction

Raw EEG data is converted to spectral data using the fast Fourier transform process of raw data operation 58, as previously described. The segments of raw EEG curves whose spectral data is averaged are generally about 2 seconds in length each, which is shorter than the average period between spurious artifact signals. Typically from 15 to 90 segments are spectrally analyzed and the spectra averaged. The spectra usually consist of 128 frequency bands of ½ Hz in width covering the spectrum from 0 to 64 Hz. The ends of each segment can be tapered in accordance with the operator's discretion in connection with the Fourier transform process.

Reduced Data Quality Control and Topographic Display

Working with reduced EP and EEG data, that is sequences of time frames of transient response data and groups of spectral band data, the operator can use reduced data quality control operation 59 to view the waveforms, discard bad data reflecting movement artifact, eye blink, or muscle activity and eliminate high frequency noise. The reduced data can then be topographically displayed on a frame by frame or cartooned basis using the reduced data topographic display operation 60. In either case, the operator can form frames which represent combinations of underlying frames. For example, groups of ½ Hz bandwidth frames can be combined into larger bandwidth frames corresponding with typical spectra of clinical interest, e.g., alpha, beta, delta and theta. Bands of any desired width can be formed. In addition to displaying raw spectral energy information from EEG data, it is possible to display normalized spectral energy in which the points on each display are normalized to the overall spectral energy of each electrode or to the average overall spectral energy at all electrodes. In the case of EP data, it is similarly possible to display each point as a normalized value to the value at one specific electrode, e.g., the vertex electrode designated "C_z", or to a standardized value, or to a selected value, or to the V_{RMS} of the background activity at each electrode, or to the V_{RMS} of all electrodes. Similarly, the 128 frames of an EP transient response can be grouped into frames of greater time duration for display.

Group Data Analysis

By accumulating a number of stored data frames, it is possible for the operator to assemble and display group data files using group file production operation 61 and group topographic data display operation 62.

Significance Probability Mapping (SPM)

Using stored data for various groups and individuals, the operator can perform and display topographically t-statistic comparisons between groups of frames and z-statistic comparisions between an individual frame and a group of frames. Any other statistical group comparison can also be used to form a display matrix to illustrate group differences. This type of analysis, significance probability mapping (SPM), enables the operator to identify significant brain activity features related to various neurophysiological illnesses and conditions.

Grid Sector Analysis

Frames produced by the SPM procedure may be further analyzed by a Grid Sector Analysis (GSA). While the frames produced by the SPM procedure reflect regional abnormalities, the GSA procedure produces numerical measures of the degree of global or focal deviations from normal, which can assist in automatic determination of the existence of regional abnormalities in unknown subjects.

The first step of the GSA process conceptually requires the division of a frame into a number of different grids, each divided into sectors of a uniform size. Within each grid sector, the mean of all values of the data points lying within the sector is determined as the value of that sector. The process is repeated for grids of different fineness. Preferably three grids, of 4000 sectors, 64 sectors, and 16 sectors respectively, are used. Histograms of sector values are then prepared for each grid reflecting sector t-statistic or z-statistic values on the horizontal axis and numbers of sectors having that value on the vertical axis for each grid. Various analyses of the histrograms, which differ for focal and global abnormalities, will indicate whether an abnormality is focal, i.e., localized in one area, or global, i.e., diffused over a large part of the brain. One such analysis would simply be the observation that the peak number of sectors for the coarser grids will be at lower z or t values in the case of a global abnormality than in the case of a focal abnormality.

In the case of focal abnormalities, there is a marked difference in the histograms for the three grids, while in the case of global abnormalities, there is little or no difference for the three grids. A variety of features can be developed from the histograms to serve as possible diagnostic rules. The maximum z-value for each grid, the maximum amount of asymmetry between homologous grid regions, the mean asymmetry between homologous grid regions, and the difference between the absolute values of the sum of all left hemisphere and all right hemisphere values. Also, one can calculate the number of regions above certain criterion levels for each histogram.

A group of spectral maps or a series of EP responses can be analyzed as an ensemble by forming at each matrix point the mean of the values of each map in the ensemble. The grid region on each individual map showing the largest value is given a score of 4, the next largest a score of 2, the third largest a score of 1, and the rest a score of 0. The scores are then summed by region across all maps in the ensemble, and the regions having the three largest scores and their sum are stored as indications of focal features. The same process is repeated for the three regions having the greatest asymmetries in each image between corresponding grid sectors. The resulting information can serve as features which can be processed using TICAS to develop diagnostic rules to classify unknown subjects between a group of normals and a group having a particular dysfunction. The numerical descriptors generated by GSA, when used for statistical analysis are approximately as successful in the identification of patients with brain tumors as visual inspection of EEG data by expert clinicians.

Coefficient of Variation Analysis

Given an ensemble of segments of data, the operator can determine the mean and the standard deviation of each point across the segments. By displaying the standard deviations as percentages of mean at each point, the coefficient of variation (C/V) across the skull can be observed topographically. The normal expect range of C/V values is 40–60% and deviations from that range are immediately evident from the displays. The C/V display is useful in demonstrating head regions where there are wide variations in activity, e.g., epileptogenic regions.

Difference Maps

A display matrix can be formed to represent at each point, the difference in value of corresponding points on two underlying frames. This permits, for example, displays which suggest the regions of the brain activated by patterned light, by comparing the frames corresponding to plain light stimulation and to patterned light stimulation.

Automatic Diagnostic Rules

Working from significant brain activity features and using TICAS-related operations 65, 67, 68 and 69, the operator can develop and test diagnostic rules for accurately classifying unknown subjects between normal and abnormal groups.

Scaling

Several of the available BEAM system operations produce color topographic displays. Video monitor displays typically involve assigning to each point on the display a grey tone of color which represents the value of the point. Sixteen tones of color represent 16 different graduated values. For example, red can be used for positive values and blue can be used for negative values with the grey tone of blue or red indicating the level of positive or negative value. An absence of color represents zero. In order to maximize the visual effectiveness of the display, it is desirable to scale the values of the data points to the available color tone levels in such a way that the useful variations in value are spread among the maximum range of color tones. The scaling can be done according to a variety of options. The data points can be scaled so that the maximum absolute value of the data points over a set of matrices will be equivalent to the maximum positive and negative color tones and all other points will be scaled linearly to the intermediate color tones. Scaling can be done from zero to the maximum absolute value. The same scaling technique can be accomplished with one or more channels excluded by the operator from the scaling process so that unusually high value data points will not skew the scaling process. Scaling can be done on a matrix by matrix basis rather than across a group of matrices. Scaling can be done to a maximum value chosen by the operator. In the scaling process, any data value which is larger than the brightest available grey color will be truncated and displayed as that brightest color.

Three-Point Linear Interpolation

Since the data frames to be presented for display originally contain a relatively small number of points, e.g., 20 points, and the display is preferably of a continuous matrix of 128 × 128 points, expansion of the data by some form of interpolation is required. The expansion is accomplished by three-point linear interpolation, in which each display point is determined as a sum of the values of the three nearest data points on the original data frame, each multiplied by a predetermined coefficient which reflects the precise location of the display point. As an alternative to the software previously described, the calculation of the display point can be done on hardware having an extremely short processing time, making possible "real-time" displays, that is, each display matrix is calculated in a time shorter than the display time for each display matrix. A detailed description of the three-point interpolation technique is contained in U.S. patent application Ser. No. 221,830 (hereby incorporated by reference).

Display Features; Multidimensional Display

As illustrated in FIG. 32, the each topographic display comprises an outline of the skull with an indication of its orientation with respect to the ears and nose. All display points outside the outline are suppressed. Within the skull outline are displayed the grid of data points, each of which reflects a value or values for that point on the skull. The number of dimensions of information which may be represented by a given point varies with the display method. Frequently only one dimension of information is presented at any point in the form of a grey-tone of color on a predetermined grey-tone scale. Alternatively additional dimensions may be reflected at a point as a unique combination of three colors. Three dimensions can be represented by the quantity of each of the original colors which is mixed into the combination and a fourth dimension could be the lightness or darkness of the three. In this manner, for example, spectral EEG data for four frequency bands of brain activity could be displayed simultaneously. A detailed description of this four-dimensional display is contained in U.S. patent application Ser. No. 221,830.

Whenever displays are cartooned, the operator may select the frame rate of display from stationary to ten frames per second minimum. The cartoon can also be displayed legarithmically with time, so that the later matrices are displayed in faster sequence then the earlier ones, which visually compensates for the fact that more EP response information is available just after the stimulus than toward the end of the response period.

Examples of System Use

The brain electrical activity mapping system offers a powerful brain diagnostic and research tool by permitting immediate video display of information about steady-state brain waves, EP transient responses, spectral analyses of EEG signals, and statistical information based on these types of data, and the ability to develop diagnostic rules from selected features of data. The following examples illustrate the versatility and utility of the system.

Suspected Epilepsy

Although the "spike and wave" as seen on routine EEG graphs is virtually diagnostic of epilepsy, over 10% of all true epileptics fail to demonstrate this abnormality. Use of special electrodes, sleep studies, and activating drugs often fails to produce spikes in true epileptics. This means that although an epileptic may have brain cortex that is capable of demonstrating sufficient irritability at times to produce a seizure, that at other times it fails to be sufficiently irritative to produce a spike on the EEG and thus eludes diagnosis. Topographic display is of great assistance in such situations. Such suspected epileptic patients should have eyes open (EO) and eyes closed (EC) topographic studies performed. Irritative cortex presents itself as focal increases of activity over all frequency bands, especially the high frequency beta bands. The visual evoked EP response (VER) topographic study should also be performed. Irritative cortex leads to focal increases of both positive and negative waves. If the epilepsy is associated with an atrophic lesion, a region of reduced EEG and EP activity may be found in close association with the focal irritability.

When spikes are found, displays of their topographic extent are extremely useful in determining their point of origin. In this case, raw EEG data is displayed in cartoon form thus delineating the epileptic dipole.

Suspected Supratentorial Brain Lesion

Patients are often referred for EEG tests in order to rule in or out a lesion of one or both cerebral hemispheres. This includes tumor, stroke, abscess, atrophy, arterio-venous malformation, congenital malformation, hemorrhage, regional encephalitis. These subjects should be subjected to topographic studies in the eyes open and eyes closed brain states, and for the VER and bilateral somotosensory evoked response (BSER) EP situations. In general these lesions may be recognized by the pattern of hypo- or hyperactive cortex that become visible on the brain electrical activity mapping images. For example, tumors show decrease in activity early, excessive activities later, and reduced activity at the vertex. Brain electrical activity mapping greatly adds to the information obtained by radiographic scanning as it is sensitive to the functional disturbances produced by these lesions which usually extend beyond the anatomical limits of the lesion.

To pinpoint abnormalities the technique of significance probability mapping (SPM) should be used. Furthermore, quantification of a lesion by grid sector analysis (GSA) is often useful.

Brain electrical activity mapping is most useful when tests must be applied to a large population for screening purposes or repeatedly to a single person. Such uses would include screening for tumor and stroke, determining whether a lesion is increasing or decreasing, and assessing the effects of treatment on a lesion. Brain electrical activity mapping is completely non-invasive, and not dangerous as radiographic techniques would be in such circumstances. There is also evidence that many lesions produce electrical (functional) disturbances before they can be detected by radiographic means.

Suspected Learning Disabilities

Brain electrical activity mapping studies are most useful in the elucidation of regional abnormalities of brain activity found in dyslexia, hyperactivity, dyscalculia, and combinations of the above. For example, dyslexia reveals abnormalities not just in the classic left temporal lobe speech areas but in the medial frontal lobe bilaterally. To demonstrate these abnormalities, one needs to perform the full test battery which includes: right hemispheric activating tests (the Kimura Figures task and listening to music as described elsewhere); left hemispheric activating (listening to speech and reading Grey Oral passages as described elsewhere); and bi-hemispheral tests (Paired Associates test and the Tight-Tyke evoked potential phenomic discrimination test as described elsewhere).

Automated classification tests to discriminate among these clinical entities can be developed.

Emotional Dysfunction

Many forms of emotional disorder can be caused by the lesions mentioned above. Brain electrical activity mapping can be more useful in the recognition of covert pathology in this patient population than radiographic techniques. In addition, certain forms of psychopathology have recognizable brain electrical activity mapping signatures. For example sociopathic behavior is associated with lack of synchrony between the frontal lobes; e.g., the VER may show different electrical polarity between the right and left frontal lobes. Schizophrenia shows markedly increased EEG slow activity overlying the frontal regions. In this group of subjects, the eyes open and eyes closed EEG and VER studies are most useful.

Infant Competence

Discrimination between babies at risk for future learning and emotional problems is a frequent clinical request. Brain electrical activity mapping has proven useful in accomplishing such discrimination. In addition to studying the EEG and VER in stages 1 and 2 sleep, the EEG should be studied while the babies are brought into the alert state and maintained there as discussed elsewhere. Less competent babies, for example, show paradoxical increases in frontal delta slowing as they are alerted.

Suspected Dementia

Senile and pre-senile dementia represent a major problem for gerontologists and neurologists. Radiograpic evidence of brain abnormality may not be found until the clinical symptom complex is well established. On the other hand, brain electrical activity mapping studies demonstrate early abnormalities in a non-invasive manner. The best battery of tests is similar to those described above for suspected learning disabilities, but generally the tight-tyke EP is replaced by another EP where the subject must discriminate between frequently and infrequently heard tone pips of differing frequency. A difference EP between the response to the two different tone pips is produced. The topographic display of the difference EP shows a marked reduction in dementia and may be used to follow the course of dementia and the response of dementia to pharmacotherapies.

Headache

Headache may be caused by many factors. Brain electrical activity mapping is very useful to screen out serious lesions of the types described as supratentorial lesions above. The specific syndrome of migraine headache has a frequently seen pattern on brain electrical activity mapping of excessive 8-11 Hz occipital oscillations and excessive occipital activity. It is best to use the EO and EC EEG and VER for headache. Occasionally the BSEP is useful.

Comparison of Individual to Group

As described above, the brain electrical activity mapping system is generally able to compare an individual statistically to a group and display the result topographically. In a clinical setting, the individual in question, who may have displayed a normal CT scan, is compared to an age matched/sex matched group of normals, and abnormalities are then displayed in color-mapped form, wherein bright colors show high abnormality and dull colors show insignificant abnormality. This technique provides an effective diagnostic tool.

Comparisons of Groups; Automatic Diagnosis Rules

The result of a group comparision under the system is a topographic display of statistical difference expressed as t-statistics, which when coupled with the number of degrees of freedom available in the calculation, produce a probability level of significant difference between two groups at a particular brain state. For instance, a group of normals could be compared to a group of schizophrenics by the creation of t-statistic displays with respect to a variety of brain states and stimuli. The user looks for displays which exhibit high degrees of coherence and statistical difference. This is normally shown on a screen in color. The larger statistical differences appear as brighter colors. The degree to which the differences are focused at particular points or diffused over the skull is also apparent. Smoothness in the lines dividing areas of different brightness suggests focused differences, while diffuse differences are suggested by ragged edges between dim and bright areas. It is possible for the researcher, upon selection of a particular map that shows something interesting, to save the matrix for later analysis. Such a saved matrix of t-statistics can be used to non-linearly weight the underlying data frames to create features which can be analyzed using TICAS. Once a set of saved frames representing group difference information is accumulated, he then converts all of the saved information, representing features which tend to distinguish the two groups into a file format which is suitable for analysis by TICAS, which is a multi-variate classification system, publicly available from the University of Arizona, courtesy of Dr. Peter H. Bartell.

TICAS is designed to sift through all of the features saved in the course of the inter-group analysis and pick those which prove to be the most discerning mathematically to produce a set of features which succinctly allows automatic diagnosis of a patient.

This procedure has been used to successfully discriminate between normal subjects and those with dyslexia, to discriminate between normal subjects and those with supertentorial brain tumor, and to discriminate between subjects with exposure to organophosphate compounds and nonexposed controls.

Dyslexia Analysis

An article, *Dyslexia Regional Differences in Brain Electrical Activity by Topographic Mapping*, Duffy et al. (Annals of Neurology, Vol. 7, No. 5, May, 1980), hereby incorporated by reference, describes the use of the brain electrical activity mapping system to identify the parts of the brain whose electrical activity differs for individuals suffering from reading disability (dyslexia) as compared with normal individuals, and to establish objective standards for diagnosing dyslexia. The previously described battery of brain state tests were administered to a dyslexic group and a control group. Visual and auditory stimuli were repeatedly presented to both groups and recorded with the appropriate trial markers. The stimuli were offered in pseudorandom fashion. Using the brain electrical activity mapping system, topographic displays of the alpha (8 to 11.75 Hz) and theta (4 to 7.75 Hz) activity at each electrode for each tested brain state for each subject were produced. Similar cartoons of 128 frames (4 milliseconds each) were prepared for each type of EP response for each subject. The resulting brain state frames and EP response frames for the dyslexic group and the control group were then averaged to form mean frames of each group for each state and stimulus. The two groups of mean images were then compared using the t-statistic function. A further transformation produced a matrix of percentile index values (PI) whose value is related inversely to t-values. The PI values permit a graphic localization of regions of maximum difference between the dyslexic group and the control group. By topographically displaying the PI matrices for alpha and theta for each brain state and for each EP stimulus, it was possible to identify the brain regions which differed between the dyslexics and the controls. As a final step, a new display matrix was formed which summarized the differences reflected in all of the PI matrices as indicated by the occurrence of a certain PI level on at least one of the underlying PI matrices. The map of PI differences having a value of at least 2 identified four brain areas related to dyslexia: (1) bilateral medical frontal, (2) left anteriolateral frontal, (3) left mid-temporal and (4) left posterolateral quadrant. Classic concepts of dyslexia had not suggested the involvement of all of these brain areas in dyslexics. The study also indicated that alpha brain activity was involved in dyslexia as well as the theta activity which has previously been viewed as of primary importance.

In *Dyslexia: Automated Diagnosis by Computerized Classification of Brain Electrical Activity,* Duffy et al. (Annals of Neurology, Vol. 7, No. 5, May, 1980) hereby incorporated by reference, specific highly effective diagnostic rules for identifying dyslexics were developed by a rule selection process applying TICAS software to the brain wave data derived in the study described immediately above. Working from displays of brain electrical activity, 183 features were identified for particular regions and brain states in which the strongest differences between the dyslexic group and the normal group occurred. Two of the 183 features were identified as capable of classifying unknown subjects as dyslexic or normal with a success of 80-90%.

Localization of Tumor

In *Brain Electrical Activity Mapping (B.E.A.M.); A Method for Extending the Clinical Utility of EEG and Evoked Potential Data,* Duffy, et al (Annals of Neurology, Vol. 5, No. 4, April, 1979), hereby incorporated by reference, the use of brain electrical activity mapping system topographic displays to identify the location of a brain tumor was discussed. Spectral EEG data in the four classic bands (delta, theta, alpha, and beta) was recorded for various tested brain states. Average EP response data for strobe light stimuli comprising 128 time frames of 4 milliseconds each was also recorded. After three-point linear interpolation to expand the matrix, displays of spectral EEG data, and cartooned EP data were obtained. FIG. 5 of the article illustrates the spectral EEG displays in the four classic bands of brain activity for a patient with a known tumor, which had been located by CT scanning. The assymmetries in the spectral displays also identify the area of the tumor, although the suggested lesion size was larger than indicated by CT scanning. Analysis of 7 tumor patients, whose classic EEG's were normal or non-localizing, showed that brain electrical activity mapping studies were able to define the lesions almost as effectively as CT scan.

Use of Significance Probability Mapping with B.E.A.M. to Compare Groups and Compare Individuals to Groups

In *Significant Probability Mapping: An Aid in the Topographic Analysis of Brain Electrical Activity,* Duffy et al., accepted for publication the authors describe the use of topographical displays of statistical transformations of data. In one application, EP response data was obtained from a group of subjects with brain tumors and a second control group of subjects. The data was broken into sequential frames of 4 milliseconds each. For the control group, new matrices of mean and variance of each electrode over all members of the group were prepared. A z-statistic matrix was formed for each tumor subject to illustrate his deviation from the normal population. Using the z-statistic display a clinical neurophysiologist was able to identify 11 of 12 tumor subjects.

In a second application, discussed in the same article, EEG steady-state signals were recorded for three different brain states (resting but alert with no external stimulation, listening to a tape recording of speech, and listening to a tape recording of music) for individuals in a group of dyslexics and individuals in a group of normal readers. Matrices of alpha band activity were produced for each individual, and mean and variance matrices for each state were prepared for each of the two groups. For each group t-statistic matrices were formed to compare the resting and listening to speech states and the resting and listening to music states. By examining the t-statistic displays for the two groups it was possible to infer the differences in speech-induced and music-induced brain activity between the dyslexics and the normal readers. Those determinations could not have been made from an analysis of the underlying EEG alpha matrices.

Use of Grid Sector Analysis of B.E.A.M. SPM Data to Determine Degree of Focal or Global Deviation From Normal

In an unpublished article, "Quantification of Focal Abnormalities in BEAM Data by Grid Sector Analysis: Automated discrimination Between Normal Subjects and Patients with Supratentorial Brain Tumor", Duffy, et al., describes uses of grid sector analysis as part of the brain electrical activity mapping system for the purpose of automated neurophysiological diagnosis of brain tumor. In this application, EEG and visual EP data were recorded from a group of patients with confirmed supratentarial brain tumor and from a control group. SPM matrices were prepared comparing the tumor subjects to a normal group and comparing the control group to the tumor group. Four 96 millisecond time periods of EP data were analyzed. Grid sector analyses on the data resulted in a set of 1096 combined global and focal features from the combined EEG and EP data. By a process of features selection and rule development and testing, two features were identified as most useful in distinguishing the tumor subjects from the control subjects. When classification rules developed on the initial group of 30 subjects were applied to a new group of 10 subjects, containing 5 normals and 5 subjects with brain tumor, all ten were correctly classified.

Other Embodiments

Other embodiments of the invention are within the following claims. For example, the input data may be obtained from any type of transducer capable of measuring brain electrical activity, and

topographic displays can be prepared from the signals taken from the skull, without interpolation of additional points to form a display matrix.

Related Applications

This application is related to the following applications, each of which is hereby incorporated by reference:

(1) Frank H. Duffy, Brain Electrical Activity Mapping (BEAM)

(2) Norman David Culver, Brain Electrical Activity Mapping (BEAM)

(3) Norman David Culver, Analysis of Brain Electrical Activity

(4) Norman David Culver, Apparatus and Method for Topographic Display of Multichannel EEG Data, U.S. Ser. No. 221,830.

What is claimed is:

1. Apparatus for generating a topographic display of information on the electrical activity of the brain, said apparatus comprising

a plurality of electrical-activity transducers adapted to be placed at spaced apart locations on the skull of a patient,

stimulus means for repeatedly providing a sensory stimulus for activating the brain to produce EP responses to said transducers,

response averaging means connected to be responsive to said transducers, to produce average responses for each transducer,

processing means connected to be responsive to said averaging means for processing said average responses to generate a time sequence of matrices,

display means connected to be responsive to said processing means, for displaying said matrices as a time sequence of topographic maps of the skull, said matrice having elements defining discrete points of said maps,

said display means including means for displaying said topographic maps at a variable frame rate, said maps corresponding to different portions of said average responses being displayed respectively at different selected frame rates.

2. The apparatus of claim 1 wherein said display means includes means for displaying maps corresponding to initial portions of said average responses at a slower rate than maps corresponding to later portions of said average responses.

3. The apparatus of claim 1 wherein said display means includes means to display said topographic maps at a frame rate that varies logarithmically.

4. Apparatus for generating a topographic display of information on the electrical activity of the brain, said apparatus including

a plurality of electrical-activity transducers adapted to be placed at spaced apart locations on the skull,

processing means connected to be responsive to said transducers, for processing responses of said transducers to generate a time sequence of matrices, each of said matrices having elements representing the instantaneous amplitudes of said responses at various locations on the skull and there being a sufficient number of said matrices for a selected time period of actual brain activity for capturing onset of a transient event that occurs with a rapidity on the order of that of an epileptic spike,

said processing means including means to generate from said time sequence of matrices a running-average matrix which at any given time represents the current matrix averaged with a selected number of the matrices preceding it in time,

display means connected to be responsive to said processing means, for displaying said matrices as a time sequence of topographic maps of the skull, said elements defining discrete points of said matrice having maps,

said display means including means for displaying said topographic maps at a variable frame rate and for selectably slowing said frame rate to permit observation of said transient event.

5. The apparatus of claim 4 wherein said processing means includes means for generating 200 or more matrices for each second of real time.

6. Apparatus for generating a topographic display of information on the electrical activity of the brain, said apparatus comprising

a plurality of electrical-activity transducers adapted to be placed at spaced apart locations on the skull of a patient,

storage means connected to be responsive to said transducers for storing responses of said transducers during first and second time periods,

processing means connected to be responsive to said transducers, for processing said responses to generate first and second matrices each having elements representing brain activity at different skull locations, said first matrix representing information on brain activity during said first period, and said second matrix representing information on brain activity during said second period,

different means connected to be responsive to said processing means, for forming a difference matrix having elements each corresponding to the difference between corresponding elements of said first and second matrices,

display means connected to be responsive to said different means, for displaying said different matrix as a topographic map of the skull, said matrix elements each forming a discrete point of said topographic map.

7. The apparatus of claim 6 further comprising stimulus means for providing a sensory stimulus to generate an EP response and wherein at least one of said first and second periods of time are during said EP response.

8. Apparatus for generating a topographic display of information on the electrical activity of the brain, said apparatus comprising

a plurality of electrical-activity transducers adapted to be placed at spaced apart locations on the skull of a patient,

stimulus means for repeatedly generating a sensory stimulus for the brain, to produce at said transducers repeated segments of data, each said segment having a pre-stimulus response and a post-stimulus response,

response averaging means connected to be responsive to said transducers, for averaging said segments to produce average pre-stimulus and average post-stimulus responses for each transducer,

baseline means connected to be responsive to said responsive averaging means, for determining a zero baseline for said average segments from the mean level of at least a portion of the respective average pre-stimulus response, and

subtraction means connected to be responsive to said baseline means, for generating zeroed average segments by subtracting from each average segment the zero baseline determined by said baseline means,

processing means connected to be responsive to said subtraction means, for processing said zeroed average segments to generate one or more matrices, each said matrix having element representing information on the electrical activity of the brain at one location on the skull,

display means connected to be responsive to said processing means, for displaying said one or more matrices as topographic maps of the skull, said matrix elements forming discrete points of said maps.

9. The apparatus of claim 8 further comprising

means connected to be responsive to said stimulus means, for storing the time of occurrence of each said stimulus and

means connected to be responsive to said means for storing, for dividing each said segment into a predetermined pre-stimulus subinterval during which the pre-stimulus response occurs and a predetermined post-stimulus subinterval during which the post-stimulus response occurs, by using said stored times of stimuli occurrence as an indication of the boundary between said subintervals.

10. The apparatus of claim 8 wherein said processing means includes means to generate from said zeroed average segments a time sequence of said matrices, the display elements of each matrix representing the instantaneous amplitude of an EP response at various locations on said skull, and

said display means includes means for displaying said matrices in sequence, to thereby display said EP response as a time-varying topographic map.

11. The apparatus of claim 10 wherein said processing means includes means to generate said sequence of matrices in an endless loop, to thereby produce a cyclical display of said EP response.

12. The apparatus of claim 8 further comprising sampling means connected to be responsive to said transducers, for sampling and storing as a sequence of digital words said segments, and wherein said response averaging, baseline, and subtraction means all include means for performing the respective functions digitally.

13. The apparatus of claim 12 wherein said response averaging means includes

a summing buffer connected to be responsive to said sampling means, and having locations for storing each said sampled digital word for each transducer and

means connected to be responsive to said sampling means, for adding to said buffer locations new digital words corresponding to each new segment, to thereby generate said average pre- and post-stimulus responses.

14. The apparatus of claim 13 wherein said sampling means includes means for taking from 20 to 2000 equally-spaced-in-time samples of said pre-stimulus response and for taking 20 to 2000 equally-spaced-in time samples of said post-stimulus response.

15. The apparatus of claim 13 further comprising means connected to be responsive to said sampling means, for comparing said new digital words to predetermined limits and rejecting digital words falling outside said limits, thereby not adding said words to said buffer locations.

16. The apparatus of claim 12 further comprising means connected to be responsive to said transducers, for filtering high-frequency components from the post-stimulus response for any selected transducer by multipoint interpolation of the digital words for the selected response.

17. The apparatus of claim 8 wherein said baseline means includes means for selectively eliminating from the mean level computation portions of an average pre-stimulus response for a selected transducer, to thereby provide a more accurate baseline for the response of the selected transducer.

18. The apparatus of claim 8 further comprising means connected to be responsive to said baseline means, for displaying said pre-stimulus and post-stimulus responses and said baseline for a selected transducer, to permit an operator to evaluate the appropriateness of said baseline.

19. The apparatus of claim 8 wherein

said baseline means includes means for calculating the root mean square average value of the average pre-stimulus response and of the average post-stimulus response and

said display means includes means for displaying said root mean square average values along with said responses,

whereby such root mean square value can be used in connection with said displayed responses to evaluate whether the noise level in the responses for any particular transducer is unacceptably large and for thereby determining whether new data should be taken.

20. The apparatus of claim 8 or 19 further comprising means connected to be responsive to said baseline means, for an operator to adjust the baseline value up or down for any selected transducer average response.

21. The apparatus of claim 8, 18, or 20 further comprising means connected to be responsive to said baseline means, for repeating the steps of calculating said baseline and viewing said responses for selected transducers, whereby adjustments can be made to said baseline until said responses are satisfactory.

22. Apparatus for generating a topographic display of information on the electrical activity of the brain, said apparatus comprising

a plurality of electrical-activity transducers adapted to be placed at spaced apart locations on the skull of a patient or measure the EEG response at said locations,

sampling means connected to be responsive to said transducers, for sampling said EEG responses during time intervals shorter than the anticipated interval between interruptions in the state of brain activity,

spectral processing means connected to be responsive to said sampling means, for computing, for each of said transducers, the Fourier transforms of the sampled EEG responses and for computing from said Fourier transforms, for each of said transducers, the spectral energy contained in selected frequency bands,

processing means connected to be responsive to said spectral processing means, for processing the output of said spectral processing means to generate a plurality of matrices, one said matrix for each selected frequency band, each said matrix having elements each of which represents the spectral energy within the respective frequency band at one location on the skull,

display means connected to be responsive to said processing means, for displaying said matrices as topographic maps of the skull, said matrix elements forming discrete points of said maps.

23. The apparatus of claim 22 wherein said sampling means is arranged for sampling and storing as a sequence of digital words selected portions of said EEG responses, and wherein said spectral processing means includes means for digitally performing said Fourier transform determination.

24. The apparatus of claim 23 wherein said sampling means includes means to sample said EEG responses at at least 3 times the highest frequency in said selected frequency bands.

25. The apparatus of claim 24 wherein said sampling means includes means to sample said EEG responses at from 4 to 5 times said highest frequency.

26. The apparatus of claim 22 wherein said sampling means includes marking means for storing the start and stop times of a particular brain activity.

27. The apparatus of claim 22 wherein a time maximum interval for sampling said EEG responses is on the order of two seconds.

28. The apparatus of claim 22 wherein said sampling time interval of said EEG responses is from 0.1 to 4.0 seconds long.

29. The apparatus of claim 22 further comprising averaging means connected to be responsive to said spectral processing means, for averaging said Fourier transforms to generate, for each transducer, the average spectral energy contained within said frequency bands.

30. The apparatus of claim 22 wherein said frequency bands comprise the alpha, beta, delta, and theta bands.

31. The apparatus of claim 22 further comprising filtering means connected to be responsive to said transducers, for removing from said EEG responses frequency components outside the prominent frequency bands of brain electrical activity.

32. The apparatus of claim 31 wherein said filtering means includes means for removing from said EEG responses at least frequency components below 0.5 Hz and above 50 Hz.

33. The apparatus of claim 31 wherein said filtering means includes means for removing from said EEG responses at least frequency components below 0.5 Hz and above 1000 Hz.

34. The apparatus of claim 31 wherein said filtering means includes means for removing from said EEG responses at least frequency components below 300 Hz and above 5000 Hz.

35. The apparatus of claim 22 wherein said spectral processing means includes means for tapering the beginning and end portions of said responses prior to determination of the Fourier transform to reduce high frequency artifacts.

36. Apparatus for generating a topographic display of information of the electrical activity of the brain, said apparatus comprising

a plurality of electrical-activity transducers adapted to be placed at spaced apart locations on the skull of a patient:

processing means connected to be responsive to said transducers, for processing electrical responses measured at said transducers, said processing means including means for generating one or more matrices, each matrix containing a plurality of elements, said elements representing information on the electrical activity of the brain at particular skull locations,

statistical processing means connected to be responsive to said processing means, for processing at least two said matrices to generate a statistical comparison matrix having elements, each element being representative of a statistical difference between the corresponding elements in said two matrices,

display means connected to be responsive to said statistical processing means, for displaying said statistical comparison matrix as a topographic map of the skull, said matrix element forming discrete points of said map.

37. The apparatus of claim 36 further comprising interpolation means connected to be responsive to said statistical processing means, for expanding said statistical comparison matrix to a larger matrix prior to display, and larger matrix having additional statistical comparison elements for skull locations intermediate said transducer locations, said additional elements being generated by interpolation from the elements of said statistical comparison matrix.

38. The apparatus of claim 36 further comprising interpolation means connected to be responsive to said processing means, for expanding said matrices to larger matrices prior to said statistical processing, said larger matrices having additional elements for skull locations intermediate said transducer locations, said additional elements being generated by interpolation from the elements of said matrices.

39. The apparatus of claim 36 wherein said statistical processing means includes means to generate as said statistical comparison matrix a matrix of t values representing the statistical difference between a first and a second group of matrices, each group having matrices having a plurality of elements.

40. The apparatus of claim 39 wherein said statistical processing means includes

means connected to be responsive to said means to generate a matrix of t values, for generating a first mean-value matrix having elements which are representative of the mean values of the respective elements the matrices said first group,

means connected to be responsive to said means to generate a matrix of t values, for generating a first standard-deviation matrix having elements which are representative of the standard deviations of the respective elements of the matrices of said first group,

means connected to be responsive to said means to generate a matrix of t values, for generating a second mean-value matrix having elements which are representative of the mean values of the respective elements of the matrices of said second group,

means connected to be responsive to said means to generate a matrix of t values, for generating a second standard-deviation matrix having elements which are representative of the standard deviations of the respective elements of the matrices of said second group,

means connected to be responsive to said means to generate a matrix of t values, for generating said matrix of t values from said first and second mean-value matrices and said first and second standard-deviation matrices.

41. The apparatus of claim 36 wherein said statistical processing means includes means to generate as said statistical comparison matrix a matrix of z values representing the statistical difference between one matrix and a group of matrices.

42. The apparatus of claim 41 wherein said statistical processing means includes

means connected said means to generate a matrix of z values, for generating a first matrix representative of the mean of said group of matrices,

means connected said means to generate a matrix of z values, for generating a second matrix representative of the standard deviation of said group of matrices,

said matrix of z values being generating from said one matrix and said first and second matrices, each of said matrices having a plurality of elements.

43. The apparatus of claim 42 wherein said first matrix is representative of the mean of a normalized reference population, said second matrix is representative of the standard deviation of said population, and the elements of said matrix of z values are thereby representative of the number of standard deviations by which the elements of said one matrix differ from the mean of said population.

44. The apparatus of claim 36 further comprising

means connected to be responsive to said statistical processing means, for producing from said statistical comparison matrix one or more quantitative features which are each determined by processing the elements within selected regions of said statistical matrix,

means connected to be responsive to said means for producing features, for determining a statistical merit value for selected features,

means for determining how to combine said features to form decision rules, and

means connected to be responsive to said means for determining, for classifying individuals using said decision rules.

45. Apparatus for generating a topographic display of information on the electrical activity of the brain, said apparatus comprising

a plurality of electrical-activity transducers adapted to be placed at spaced apart locations on the skull of a patient,

statistical processing means connected to be responsive to said transducers, for processing responses at said electrical-activity transducers to generate a coefficient-of-variance matrix, said matri having element each of which represents the normalized standard deviation of the responses at one skull location, the normalized standard deviation being the standard deviation divided by the mean,

display means connected to be responsive to said statistical processing means, for displaying said coefficient-of-variance matrix as a topographic map of the skull, thereby providing a map of the level of variation in brain activity.

46. Apparatus for generating a topographic display of information on the electrical activity of the brain, said apparatus comprising

a plurality of electrical-activity transducers adapted to be placed at spaced apart locations on the skull of a patient,

processing means connected to be responsive to said transducers, for generating from responses measured at said transducers a time sequence of first matrices, each said first matrices having display elements representing the instantaneous response at various locations on said skull,

temporal interpolation means connected to be responsive to said processing means, for generating by interpolation second matrices interspersed in said sequence of first matrices, said second matrices having elements representing the approximate response at instants of time intermediate the instants of time associated with said first matrices,

display means connected to be responsive to said processing means, for displaying said first matrices and said interspersed second matrices as a time sequence of topographic maps of the skull, said matrix elements forming discrete points of said maps, said second matrices enhancing the visual smoothness of the transitions over time between said first matrices.

47. Apparatus for generating a topographic display of information on the electrical activity of the brain, said apparatus comprising

a plurality of electrical-activity transducers adapted to be placed at spaced apart locations on the skull of a patient,

processing means connected to be responsive to said transducers, for generating one or more matrices from the responses measured at said transducers, each matrix having elements representing information on electrical activity of the brain at one location on the skull,

display means connected to be responsive to said processing means, for displaying said matrices as topographic maps of the skull, said matrix elements forming discrete points of said maps,

waveform quality control means for previewing said transducer responses or processed versions thereof, said quality control means including at least two of the following means

means for tagging a response with information on its waveform quality, including an indication of whether the response should be used in later processing,

means at the discretion of the operator for eliminating a response from further processing,

means for automatically eliminating a response from further processing if a portion thereof exceeds a predetermined threshold,

means for smoothing a response to eliminate undesired high frequency components,

means for adjusting a zero baseline of a response, said zero baseline having been previously set automatically,

means for eliminating selected portions of a response from further processing, and

means for displaying in numerical form the value of a response at a point in time selected by the operator.

48. Apparatus for generating a topographic display of information on the electrical activity of the brain, said apparatus comprising

a plurality of electrical-activity transducers adapted to be placed at spaced apart locations on the skull of a patient,

means connected to be responsive to said transducers, for sampling responses measured at said transducers and storing the sampled responses as a series of data matrices, one matrix for each sampling time,

means connected to be responsive to said means for sampling, for selectively viewing the sampled responses corresponding to a single transducer as a plot of transducer output versus time,

quality control means for adjusting portions of said selectively viewed responses or eliminating said responses from further processing,

processing means connected to be responsive to said means for sampling, for processing said data matrices to generate one or more processed matrices, said processed matrices having elements each corresponding to one transducer location,

interpolation means connected to be responsive to said processing means, for generating expanded matrices by expanding the number of elements in said processed matrices to provide elements corresponding to skull locations intermediate said transducer locations,

display means connected to be responsive to said interpolation means, including a video monitor capable of generating a matrix of grey tones for displaying said expanded matrices as topographic maps of the skull, said expanded matrices having elements defining the grey tones of discrete points on said maps.

49. The apparatus of claim 48 further comprising matrix storage means connected to be responsive to said means for sampling, processing means, and interpolation means, for tagging selected data, processed, or expanded matrices and storing them for later recall and processing.

50. The apparatus of claim 48 further comprising means connected to be responsive to said transducers, for storing calibration responses from said transducers and means for calibrating said responses matrices by determining from said stored calibration responses a DC offset and gain for each said transducer.

51. Apparatus for generating a topographic display of information on the electrical activity of the brain, said apparatus comprising

a plurality of electrical-activity transducers adapted to be placed at spaced apart locations on the skull of a patient,

processing means connected to be responsive to said transducers, for processing responses measured at said transducers to generate one or more matrices, each matrix having elements representing information on brain activity at different skull locations,

display means connected to be responsive to said processing means, for displaying said one or more matrices as topographic maps of the skull, said matrix elements forming discrete points of said maps, said display means including a video monitor which generates a visible grey tone at each said discrete point, said tone being variable from a maximum tone to a minimum tone, and

scaling means connected to be responsive to said processing means, for scaling said matrix elements to the available tones.

52. The apparatus of claim 51 wherein

said video monitor generates a zero tone intermediate said maximum and minimum tone and

said scaling means comprises at least two of the following means

means connected to be responsive to said processing means, for scaling the maximum matrix element to the maximum tone and the minimum matrix element to the minimum tone and linearly interpolating inbetween,

means connected to be responsive to said processing means, for scaling the maximum matrix element to the maximum tone and the minimum matrix element to the zero tone and linearly interpolating inbetween,

means connected to be responsive to said processing means, for scaling the maximum matrix element to an operator supplied tone,

means connected to be responsive to said processing means, for scaling the minimum matrix element to an operator supplied tone,

means connected to be responsive to said processing means, for excluding from the scaling operation selected unusually high or low valued matrix elements and for assigning to matrix elements that fall outside the available tone range the closer of the maximum or minimum tone.

53. The apparatus of claim 51 wherein said video monitor generates colored grey tones of a first and a second color, with grey tones of said first color forming said grey tones between said maximum tone and said zero tone, with grey tones of said second color forming said grey tones between said minimum tone and said zero tone, and with the absence of a color tone forming said zero tone.

54. The apparatus of claim 53 wherein said two colors are complementary colors.

55. The apparatus of claim 53 wherein there are at least 6 grey tones of each of said two colors.

56. The apparatus of claim 51 wherein there are a plurality of said matrices to be displayed and said scaling means includes means for independently performing said tone scaling for each matrix displayed.

57. The apparatus of claim 51 wherein there are a plurality of said matrices to be displayed and said scaling means includes means for scaling with respect to the display elements in all said matrices so that the scaling remains the same for each matrix displayed.

58. The apparatus of claim 51 further comprising preview means connected to be responsive to said transducers, for viewing selected segments of the transducer responses and for tagging selected segments for exclusion from the scaling operation and wherein said scaling means includes means for identifying the presence of said tagging and for excluding data so tagged from said scaling operation.

59. Apparatus for generating a topographic display of information on the electrical activity of the brain, said apparatus comprising

a plurality of electrical-activity transducers adapted to be placed at spaced apart locations on the skull of a patient,

processing means connected to be responsive to said transducers, for processing responses measured at said transducers to generate one or more matrices, each matrix having elements representing information on brain activity at different skull locations,

display means connected to be responsive to said processing means, for displaying said one or more matrices as a topographic map of the skull, said matrix elements forming discrete points of said map, said display means including a video monitor which generates a tone at each said discrete point, said tone varying from a maximum tone to a minimum tone, normalizing means connected to be responsive to said processing means, for normalizing said matrix elements to a selected value, and

scaling means connected to be responsive to said processing means, for scaling said matrix elements to the available tones, said scaling means including means for assigning a selected tone to said selected normalization value.

60. The apparatus of claim 59 wherein said normalizing means includes means for normalizing said matrix elements to the matrix element representing brain electrical activity at the vertex.

61. The apparatus of claim 59 wherein said normalizing means includes means for normalizing each individual matrix element to the root mean square value of the background electrical activity at the skull location represented by the individual matrix element.

62. The apparatus of claim 59 wherein said normalizing means includes means for normalizing said matrix elements to the average root mean square value of the background electrical activity over all skull locations.

63. The apparatus of claim 59, 60, 61 or 62 further comprising

stimulus means for repeatedly providing a sensory stimulus for the brain, to produce repeated EP responses at said transducers,

response averaging means connected to be responsive to said transducers, for averaging said repeated responses for each transducer to produce an average response for each transducer, and

wherein said processing means includes means for generating from said average responses a time sequence of said matrices, and

said display means includes means for displaying said matrices in sequence.

64. The apparatus of claim 59 wherein

there is further provided spectral processing means connected to be responsive to said transducers, for computing the Fourier transforms of the responses measured by said electrical-activity transducers and for computing from said Fourier transforms the spectral energy contained in selected frequency bands,

said processing means includes means for processing the output of said spectral processing means to generate a plurality of matrices, one said matrix for each selected frequency band, each matrix having elements representing the spectral energy within the selected frequency band at various locations on the skull, and

said normalizing means includes means to normalize said matrix elements either to the total spectral energy for all said frequency bands at the skull location corresponding to said matrix element or to the average total spectral energy for all said frequency bands for all matrix elements.

65. The apparatus of claim 1, 4, 6, 8, 22, 45, 46, 47, 48, 51, or 59 wherein said processing means includes interpolation means for generating by interpolation additional matrix elements for skull locations intermediate the locations of said transducers.

66. The apparatus of claim 65 wherein said interpolation means includes means for computing the values of said additional matrix elements by three point interpolation from the values associated with the three closest transducer locations.

67. The apparatus of claim 66 wherein said three point interpolation is linear.

68. The apparatus of claim 1, 4, 6, 8, 22, 36, 45, 46, 47, 48, 51, or 59 wherein the number of said electrical-activity transducers is in the range of 10 to 200 and the number of elements in said topographic maps is at least 5 times the number of transducers.

69. A method for generating a topographic display of information on the electrical activity of the brain, said method comprising the steps of

placing a plurality of electrical-activity transducers at spaced apart locations on the skull of a patient,

storing responses of said transducers during first and second time periods,

processing said responses to generate first and second matrices each having elements representing brain activity at different skull locations, said first matrix representing information on brain activity during said first period, and said second matrix representing information on brain activity during said second period,

forming a difference matrix having elements each corresponding to the difference between corresponding elements of said first and second matrices, and

displaying said difference matrix as a topographic map of the skull, said matrix elements each forming a discrete point of said topographic map.

70. The method of claim 69 further comprising the step of providing as a stimulus for the brain during said first period a source of patterned light and providing as a stimulus for the brain during said second period a source of nonpatterned light.

71. A method of generating a topographic display of information on the EP response of the brain, said method comprising the steps of:

placing a plurality of electrical-activity transducers at spaced apart locations on the skull of a patient,

repeatedly providing a sensory stimulus for the brain so as to produce at said transducers repeated segments of data, each segment including a prestimulus response and a post-stimulus response,

averaging the repeated segments to produce an average segment for each transducer,

determining a zero baseline for each average segment from the mean level of at least a portion of the average pre-stimulus response,

subtracting the zero baselines from the respective average segments to produce zeroed average segments,

processing the zeroed average segments to generate one or more matrices of elements, with each matrix element representing information on the EP response at one location on the skull, and

displaying said one or more matrices as topographic maps of the skull, with each display point forming a discrete point on said maps.

72. A method for generating a topographic display of information on the electrical activity of the brain, said method comprising the steps of

placing a plurality of electrical-activity transducers at spaced apart locations on the skull of a patient,

repeatedly providing a sensory stimulus for activating the brain to produce EP responses at said transducers,

averaging said responses to produce average responses for each transducer,

processing said average responses to generate a time sequence of matrices, displaying said matrices as a time sequence of topographic maps of the skull, said matrices having elements defining discrete points of said maps, and

displaying said topographic maps at a variable frame rate,

maps corresponding to different portions of said average responses being displayed respectively at different selected frame rates.

73. A method for generating a topographic display of information on the electrical activity of the brain, said method including the steps of

placing a plurality of electrical-activity transducers at spaced apart locations on the skull,

processing responses of said transducers to generate a time sequence of matrices, each said matrix having elements representing the instantaneous amplitudes of said responses at various locations on the skull and there being a sufficient number of said matrices for a selected time period of actual brain activity for capturing onset of a transient event that occurs with a rapidity on the order of that of an epileptic spike,

generating from said time sequence of matrices a running-average matrix which at any given time represents the current matrix averaged with a selected number of the matrices preceding it in time,

displaying said matrices as a time sequence of topographic maps of the skull at a variable frame rate, each running-average matrix having elements defiring discrete points of said maps, and

selectably slowing the frame rate at which said topographic maps are displayed so as to permit observation of said transient event.

74. A method for generating a topographic display of information on the electrical activity of the brain, said method comprising the steps of

placing a plurality of electrical-activity transducers at spaced apart locations on the skull of a patient to measure the EEG response at said locations,

sampling the EEG responses during time intervals shorter than the anticipated interval between interruptions in the state of brain activity,

computing, for each of said transducers, the Fourier transforms of the sampled EEG responses and computing from said Fourier transforms, for each of said transducers, the spectral energy contained in selected frequency bands,

processing said computed spectral energy of said selected frequency bands to generate a plurality of matrices, one said matrix for each selected frequency band, said matrices having elements representing the spectral energy within the respective frequency band at one location on the skull, and

displaying said matrices as topographic maps of the skull, said matrix elements forming discrete points of said maps.

75. A method for generating a topographic display of information on the electrical activity of the brain, said method comprising the steps of

placing a plurality of electrical-activity transducers at spaced apart locations on the skull of a patient,

processing electrical responses measured at said transducers, said processing means including means for generating one or more matrices, each matrix containing a plurality of elements, said elements representing information on the electrical activity of the brain at particular skull locations,

processing at least two said matrices to generate a statistical comparison matrix said statistical comparison matrix having elements each of which representative of a statistical difference between the corresponding elements in said two matrices, and

displaying said statistical comparison matrix as a topographic map of the skull, said matrix elements forming discrete points of said map.

76. A method for generating a topographic display of information on the electrical activity of the brain, said method comprising the steps of

placing a plurality of electrical-activity transducers at spaced apart locations on the skull of a patient,

statistically processing responses at said electrical-activity transducers to generate a coefficient-of-variance matrix, said matrix having elements each of which represents the normalized standard deviation of the responses at one skull location, the normalized standard deviation being the standard deviation divided by the mean,

displaying said coefficient-of-variance matrix as a topographic map of the skull, thereby providing a map of the level of variation in brain activity.

77. A method for generating a topographic display of information on the electrical activity of the brain, said method comprising the steps of

placing a plurality of electrical-activity transducers for placement at spaced apart locations on the skull of a patient,

generating from responses measured at said transducers a time sequence of first matrices, each said first matrix having display elements representing the instantaneous response at various locations on said skull,

generating by interpolation second matrices interspersed in said sequence of first matrices, said second matrices having elements representing the approximate response at instants of time intermediate the instants of time associated with said first matrices,

displaying said first matrices and said interspersed second matrices as a time sequence of topographic maps of the skull, said matrix elements forming discrete points of said maps, said second matrices enhancing the visual smoothness of the transitions over time between said first matrices.

78. A method for generating a topographic display of information on the electrical activity of the brain, said method comprising the steps of

placing a plurality of electrical-activity transducers at spaced apart locations on the skull of a patient,

generating one or more matrices from responses measured at said transducers, each matrix having elements representing information on the electrical activity of the brain at one location on the skull,

displaying said one or more matrices as topographic maps of the skull, said matrix elements forming discrete points of said maps,

previewing said transducer responses or processed versions thereof, said previewing including at least two of the following steps

tagging a response with information on its waveform quality, including an indication of whether the response should be used in later processing,

eliminating a response from further processing at the discretion of the operator, automatically eliminating a response from further processing if a portion thereof exceeds a predeterminedthreshold,

smoothing a response to eliminate undesired high frequency components,

adjusting a zero baseline of a response, said zero baseline having been previously set automatically

eliminating selected portions of a response from further processing, and

displaying in numerical form the value of a response at a point in time selected by the operator.

79. A method for generating a topographic display of information on the electrical activity of the brain, said method comprising the steps of

placing a plurality of electrical-activity transducers at spaced apart locations on the skull of a patient,

sampling responses measured at said transducers and storing the sampled data as a series of data matrices, one matrix for each sampling time,

selectively viewing the sampled data corresponding to a single transducer as a plot of transducer output versus time,

adjusting portions of said selectively viewed data or eliminating said data from further processing,

processing said data matrices to generate one or more processed matrices said one or more processed matrices having elements each corresponding to one transducer location,

expanding the number of elements in said one or more processed matrices to provide elements corresponding to skull locations intermediate said transducer locations,

providing a video monitor capable of generating a matrix of grey tones for displaying said expanded matrices as topographic maps of the skull, said expanded matrix elements defining the grey tones of discrete points on said maps.

80. A method for generating a topographic display of information on the electrical activity of the brain, said method comprising the steps of

placing a plurality of electrical-activity transducers at spaced apart locations on the skull of a patient,

processing responses measured at said transducers to generate one or more matrices, each matrix having elements representing information on brain activity at different skull locations,

displaying one or more said matrices as topographic maps of the skull, said matrix elements forming discrete points of said maps,

providing a video monitor which generates a visible grey tone at each said discrete point, said tone being variable from a maximum tone to a minimum tone, and

scaling said matrix elements to the available tones.

81. The method of claim 80 wherein

said providing step comprises providing a said video monitor which generates a zero tone intermediate said maximum and minimum tone, and

said scaling step comprises at least two of the following steps

scaling the maximum matrix element to the maximum tone and the minimum matrix element to the minimum tone and linearly interpolating inbetween,

scaling the maximum matrix element to the maximum tone and the minimum matrix element to the zero tone and linearly interpolating in between,

scaling the maximum matrix element to an operator supplied tone,

scaling the minimum matrix element to an operator supplied tone,

excluding from the scaling operation selected unusually high or low valued matrix elements and assigning to matrix elements that fall outside the available tone range the closer of the maximum or minimum tone.

82. A method for generating a topographic display of information on the electrical activity of the brain, said method comprising the steps of

placing a plurality of electrical-activity transducers at spaced apart locations on the skull of a patient,

processing responses measured at said transducers to generate one or more matrices, each matrix having elements representing information on brain activity at different skull locations,

displaying said one or more matrices as a topographic map of the skull, said matrix elements forming discrete points of said map,

providing a video monitor which generates a tone at each said discrete point, said tone varying from a maximum tone to a minimum tone,

normalizing said matrix elements to a selected value, and

scaling said matrix elements to the available tones, said scaling means including means for assigning a selected tone to said selected normalization value.

* * * * *

US005213562A

United States Patent [19]

Monroe

[11] **Patent Number:** 5,213,562

[45] **Date of Patent:** May 25, 1993

[54] **METHOD OF INDUCING MENTAL, EMOTIONAL AND PHYSICAL STATES OF CONSCIOUSNESS, INCLUDING SPECIFIC MENTAL ACTIVITY, IN HUMAN BEINGS**

[75] Inventor: **Robert A. Monroe**, Nelson County, Va.

[73] Assignee: **Interstate Industries Inc.**, Faber, Va.

[21] Appl. No.: **514,460**

[22] Filed: **Apr. 25, 1990**

[51] Int. Cl.^{5} **A61M 21/00**

[52] U.S. Cl. **600/28**; 128/732

[58] Field of Search 600/26-28; 128/731-732, 905

[56] **References Cited**

U.S. PATENT DOCUMENTS

2,466,054	4/1949	Siebel .	
3,160,159	12/1964	Hoody et al. .	
3,576,185	4/1971	Schulz et al. .	
3,712,292	1/1973	Zentmeyer, Jr. .	
3,753,433	8/1973	Bakerich et al. .	
3,826,243	7/1974	Anderson .	
3,837,331	9/1974	Ross .	
3,884,218	5/1975	Monroe	600/28
4,034,741	7/1977	Adams et al. .	
4,141,344	2/1979	Barbara .	
4,227,516	10/1980	Meland et al. .	
4,335,710	6/1982	Williamson .	
4,573,449	3/1986	Warnke .	
4,834,701	5/1989	Masaki	600/28
5,036,858	8/1991	Carter et al.	128/732

Primary Examiner—Lee S. Cohen
Assistant Examiner—John P. Lacyk
Attorney, Agent, or Firm—Sughrue, Mion, Zinn, Macpeak & Seas

[57] **ABSTRACT**

A method having applicability in replication of desired consciousness states; in the training of an individual to replicate such a state of consciousness without further audio stimulation; and in the transferring of such states from one human being to another through the imposition of one individual's EEG, superimposed on desired stereo signals, on another individual, by inducement of a binaural beat phenomenon.

6 Claims, 5 Drawing Sheets

84

FIG.1A

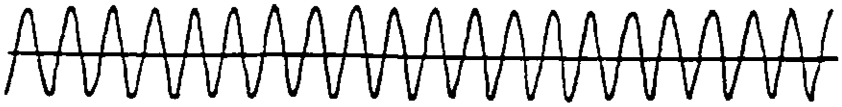

FIG.1B

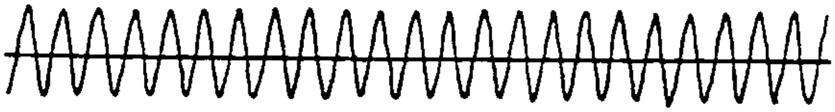

FIG.1C

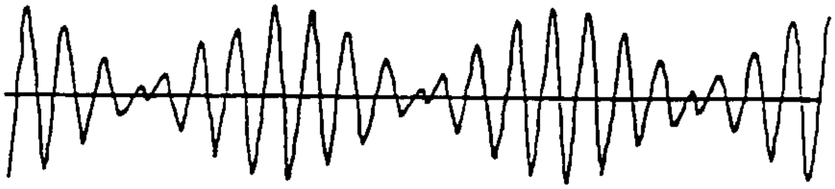

FIG.2A

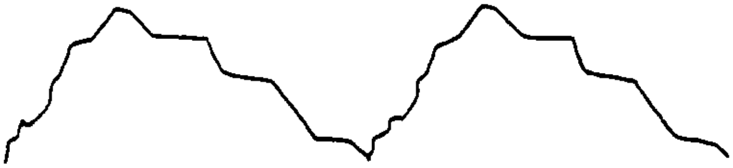

FIG.2B

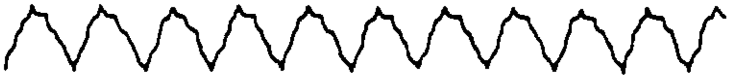

FIG.2C

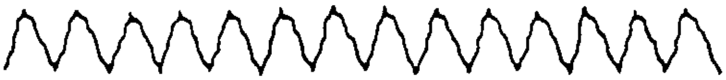

FIG.2D

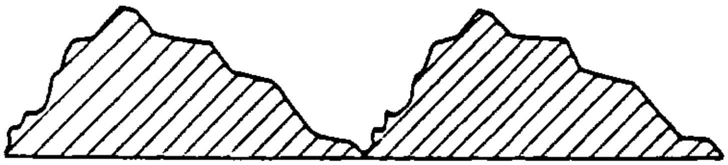

85

FIG.3A

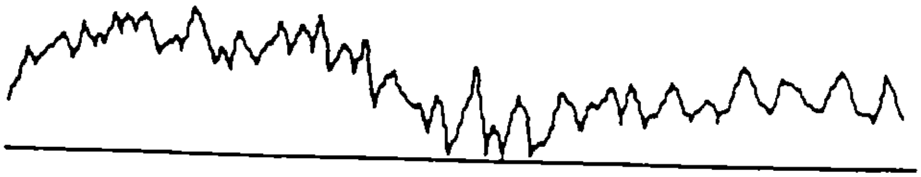

FIG.3B

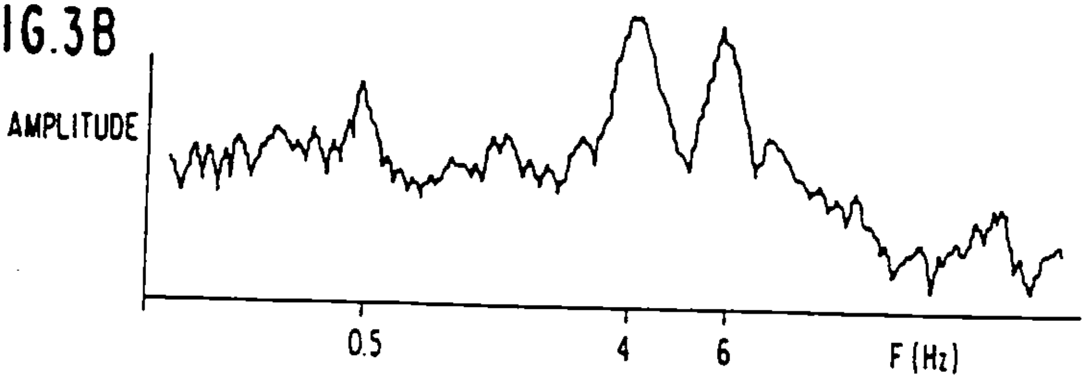

FIG.3C

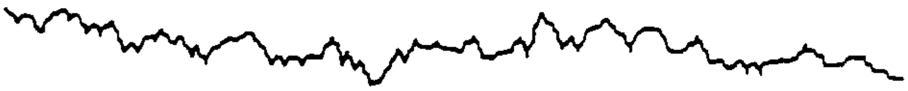

FIG.3D

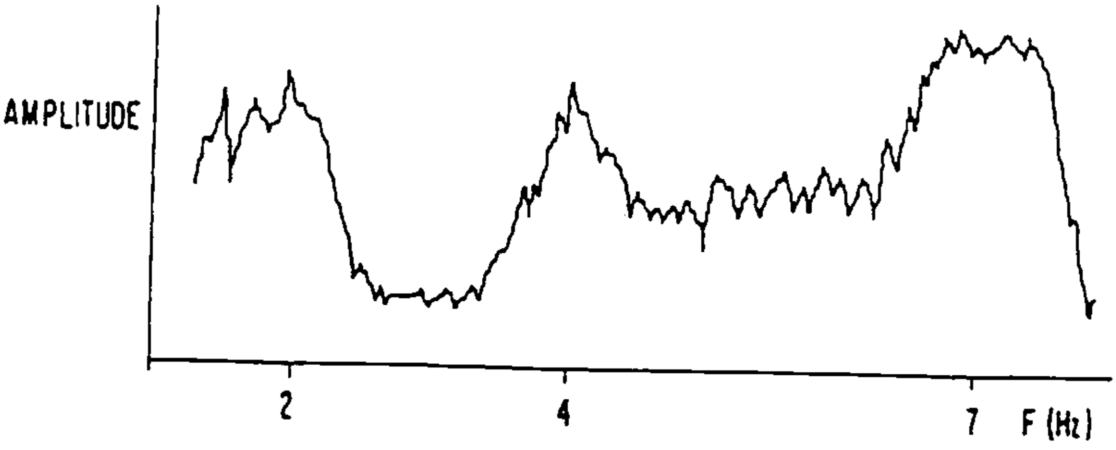

FIG.3E

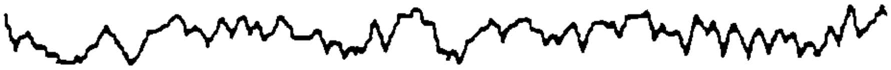

FIG.3F

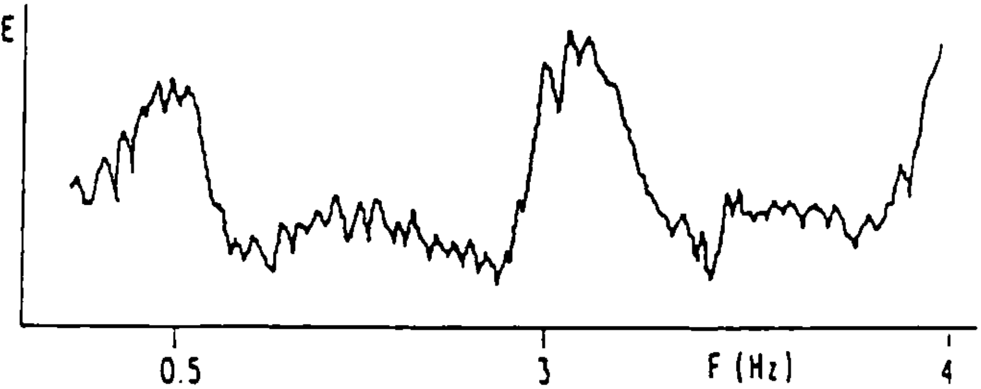

FIG.3G

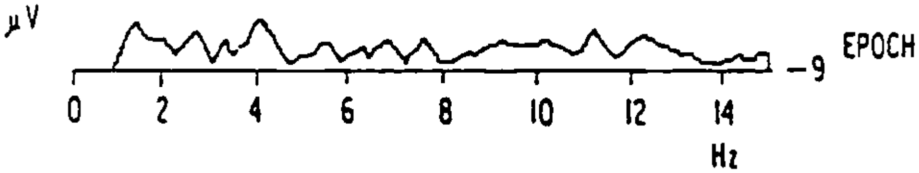

FIG.3H

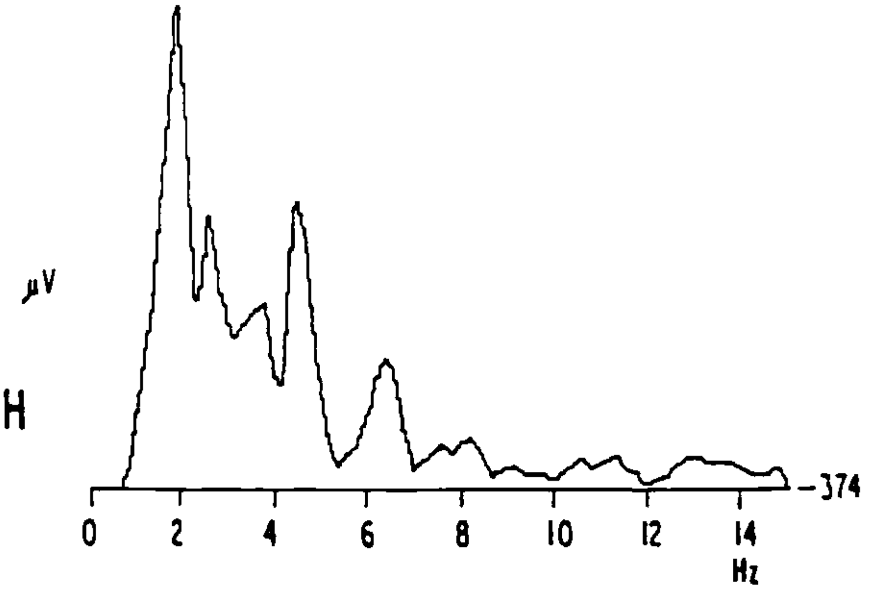

87

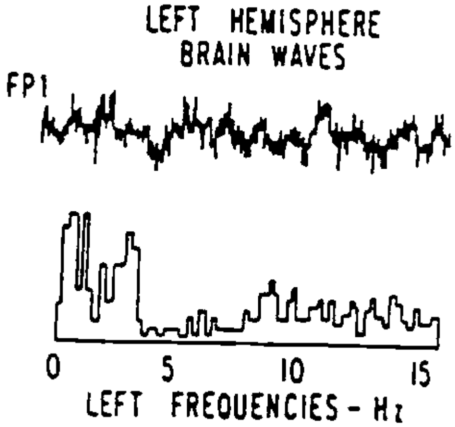

FIG. 4A

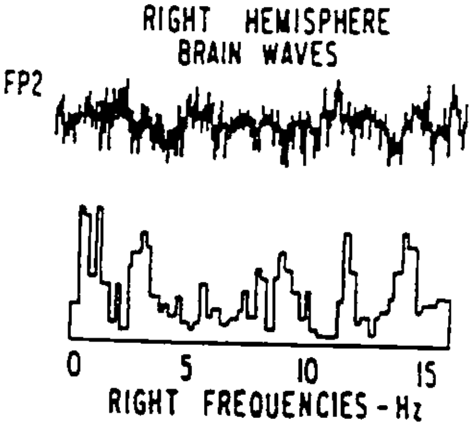

FIG. 4B

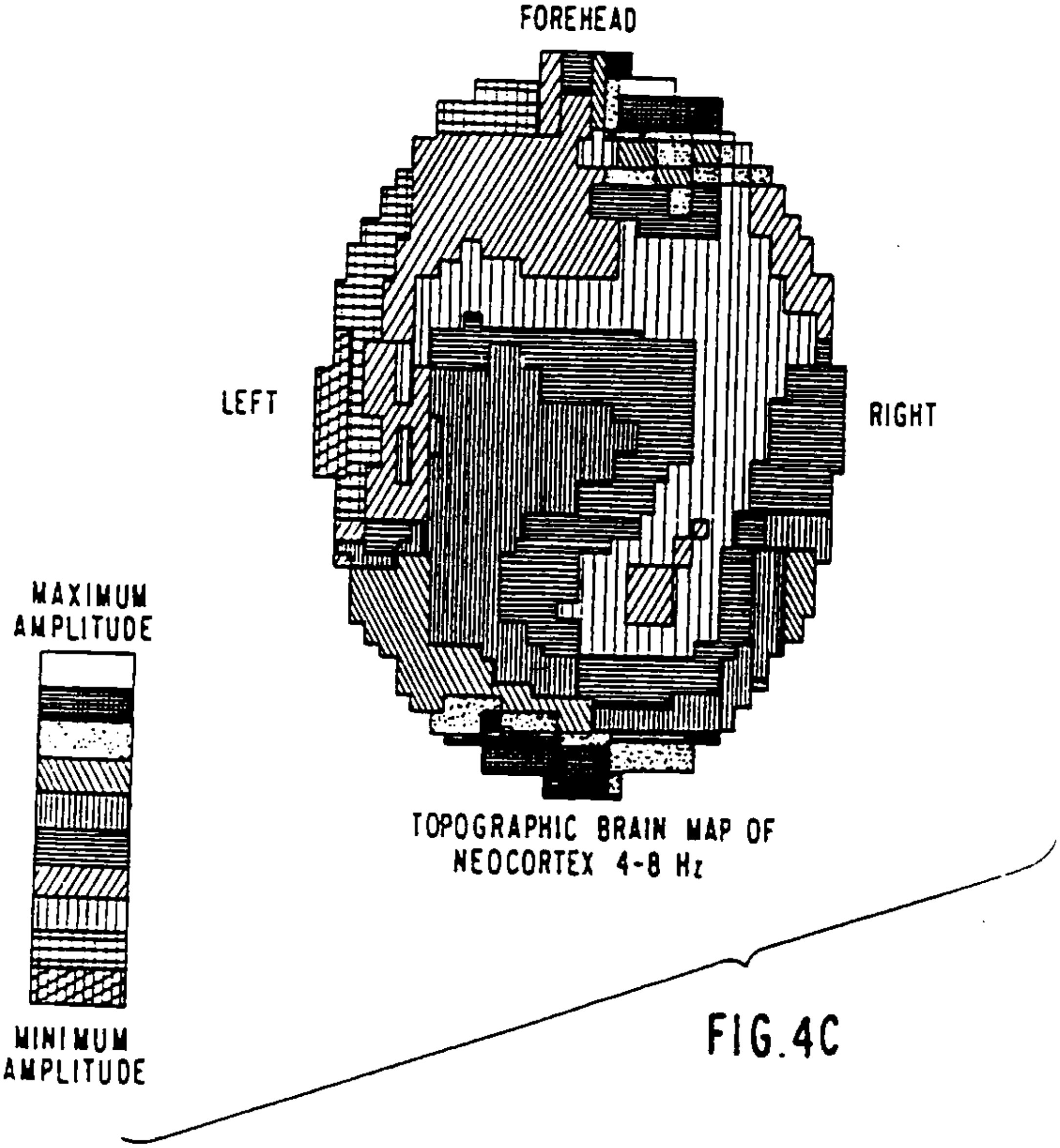
FOREHEAD
LEFT
RIGHT
MAXIMUM AMPLITUDE
MINIMUM AMPLITUDE
TOPOGRAPHIC BRAIN MAP OF NEOCORTEX 4-8 Hz

FIG. 4C

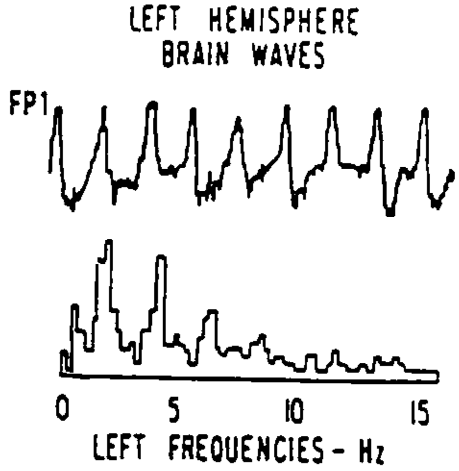

FIG. 4D

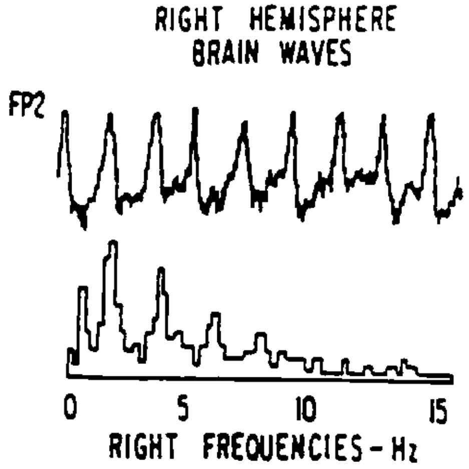

FIG. 4E

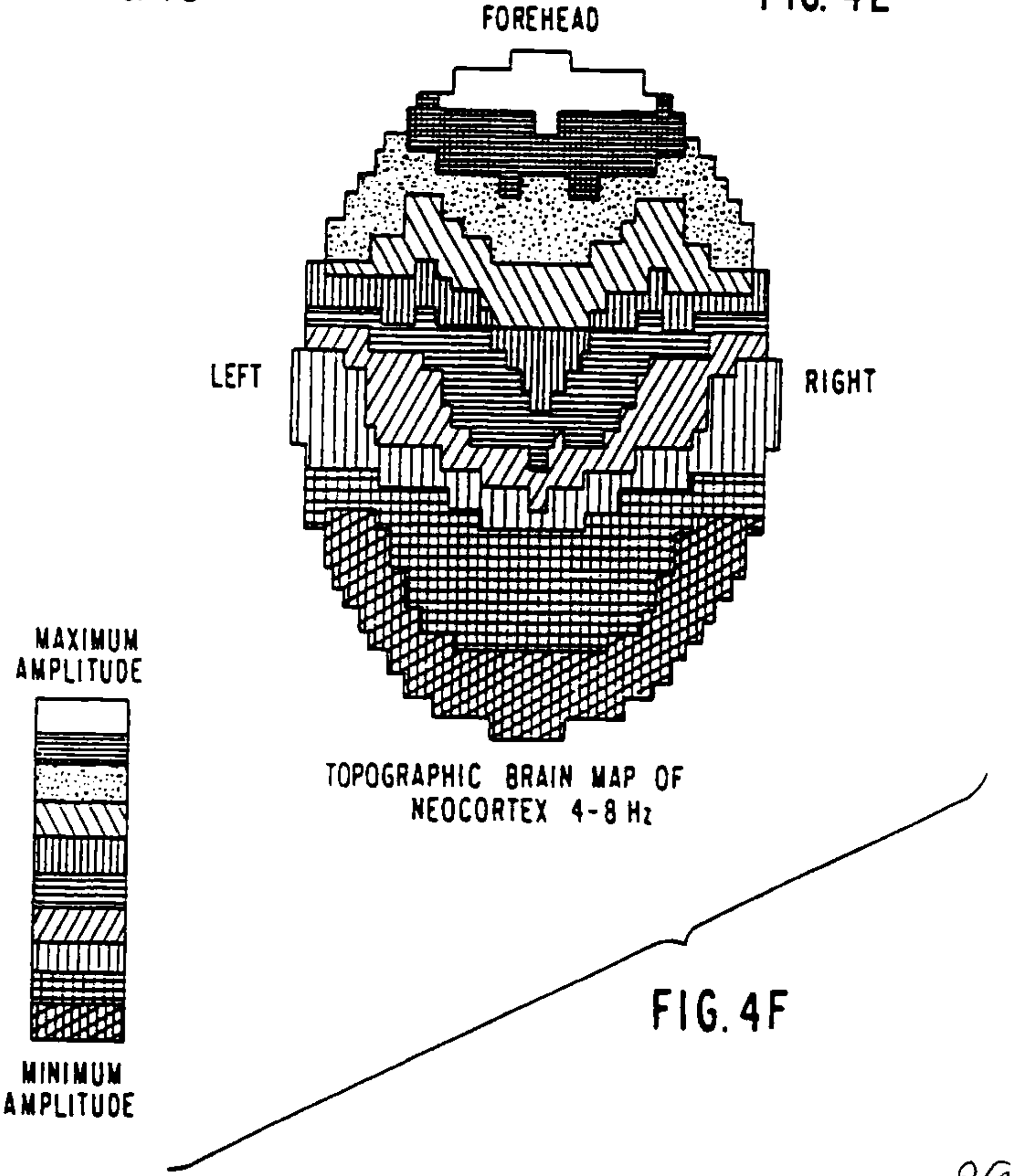
FOREHEAD
LEFT
RIGHT
TOPOGRAPHIC BRAIN MAP OF
NEOCORTEX 4-8 Hz
MAXIMUM
AMPLITUDE
MINIMUM
AMPLITUDE

FIG. 4F

METHOD OF INDUCING MENTAL, EMOTIONAL AND PHYSICAL STATES OF CONSCIOUSNESS, INCLUDING SPECIFIC MENTAL ACTIVITY, IN HUMAN BEINGS

BACKGROUND OF THE INVENTION

The present invention relates to a method of inducing various states of consciousness in human beings. More particularly, the invention relates to a method of inducing such states of consciousness through generation of stereo audio signals having specific wave shapes which act as a carrier of a binaural beat. The resultant binaural beat acts to entrain brain waves into unique waveforms characteristic of identified states of consciousness. The invention is applicable in areas of learning and behavior replication as well as in the area of sleep inducement, and thus represents a significant departure from and improvement over known audio-based sleep inducement techniques, some of which will be discussed below.

The binaural beat phenomenon was discovered in 1839 by H. W. Dove, a German experimenter. Generally, this phenomenon works as follows. When an individual receives signals of two different frequencies, one signal to each ear, the individual's brain detects a phase difference or differences between these signals. When these signals are naturally occurring, the detected phased difference provides directional information to the higher centers of the brain. However, if these signals are provided through speakers or stereo earphones, the phase difference is detected as an anomaly. The resulting imposition of a consistent phase difference between the incoming signals causes the binaural beat in an amplitude modulated standing wave, within each superior olivary nucleus (sound processing center) of the brain. It is not possible to generate a binaural beat through an electronically mixed signal; rather, the action of both ears is required for detection of this beat.

FIGS. 1A and 1B show two superposed waves of different frequencies. FIG. 1C shows the resulting wave, which has a clear beat phenomenon. Assuming the two waves have equal amplitude but different respective frequencies f_1, f_2, the combination of the two waves may be represented mathematically as follows:

$$\begin{aligned} X &= X_1 + X_2 \\ &= a^*\cos(2\pi f_1 t) + a^*\cos(2\pi f_2 t) \\ &= a^*[\cos(2\pi f_1 t) + a^*\cos(2\pi f_2 t)] \\ &= 2a^*\cos\left(2\pi \frac{[f_1 - f_2]}{2}\, t\right)^*\cos\left(2\pi \frac{[f_1 + f_2]}{2}\, t\right) \end{aligned}$$

The beat phenomenon arises from the variation in amplitude of a resulting carrier frequency. Pulses appear every $\frac{1}{2}(f_1-f_2)$, with two maxima occurring each cycle, when $\cos(2\pi)\frac{1}{2}[f_1-f_2]=\pm 1$. That is, the beat frequency is simply f_1-f_2, a result which agrees with experience.

Known consciousness state inducing techniques have not used this binaural beat phenomenon, but have relied on other techniques, as follows. For example the use of audio generators to induce a state of consciousness known as sleep is well known in the prior art, as exemplified by U.S. Pat. No. 2,711,165 and 3,384,074. In one type of technique exemplified in these patents, generated audio signals include pleasing and harmonious study sounds or vibrations, fixed frequency signals which are buried cyclically with respect to amplitude, and repetitive sounds such as the falling of rain on the roof and the sighing wind through the trees.

U.S. Pat. No. 2,304,095 relates to a method of inducing sleep by generation of an audible or tactual signal which is related to the physiological process of heartbeat and respiration. In the disclosed method, the pitch and amplitude of a pleasing audio signal are varied at a rate somewhat slower than either the rate of heartbeat or the rate of respiration. As a result, heartbeat and respiration tend to synchronize with the audio signal, thus lowering heartbeat and respiration rates and inducing sleep.

Of course, there are other naturally-occurring sounds which have been recorded, and which are not varied, but which instead induce a state of relaxation which leads to sleep for a similar reason. For example, the pounding of waves on a shore line occurs at a frequency generally lower than that of heartbeat or respiration, and induces a state of relaxation.

The use of an electroencephalogram (EEG) as a research and diagnostic tool has led to findings that particular brain wave patterns are indicative of different states of consciousness. In 1934, researchers discovered that brain waves, and their associated states of consciousness, could be altered with repetitive visual stimulation at a known frequency, an effect known as entrainment. Scientific interest in entrainment continued throughout the 1960's. In the 1970's, numerous independent studies repeatedly confirmed that rhythmic flashing lights rapidly entrained brain waves.

A sonic equivalent of photic entrainment also is known, as disclosed for example in commonly-assigned U.S. Pat. No. 3,884,218, the inventor of which is the inventor of the present application. This patent discloses a method of inducing sleep in a human being by generating an audio signal which is made up of a familiar pleasing repetitive sound modulated by frequencies usually associated with an EEG sleep pattern. There are different EEG patterns related to various levels or depths of sleep, and it has been found that by modulating the repetitive sound with these different sleep patterns, it is possible to induce various levels of sleep. The inventor has coined the term frequency following response, or FFR, to describe this phenomenon.

Other known techniques for inducing various states of consciousness, or for performing brainwave analysis and related functions, are shown, for example, in the following U.S. patents:

2,466,054	4,034,741	3,160,159	4,141,344
3,576,185	4,227,516	3,712,292	4,335,710
3,753,433	4,573,449	3,826,243	4,834,701
3,837,331.			

The binaural beat phenomenon described above also can create a frequency entrainment effect. If a binaural beat is within the range of brain wave frequencies, generally less than 30 cycles per second, the binaural beat will become an entrainment environment. This effect has been used to study states of consciousness, to improve therapeutic intervention techniques, and to enhance educational environments. However, the modulation of the binaural beat signals with brain waves associated with particular activities has not been attempted previously.

SUMMARY OF THE INVENTION

In view of the foregoing, it is one object of the invention to provide a method of inducing states of consciousness by generating stereo audio signals having specific wave shapes. These signals act as a carrier of a binaural beat. The resulting beat acts to entrain brain waves into unique waveforms characteristic of identified states of consciousness.

The method of the invention extends beyond the confines of the frequency entraining concept, and incorporates waveform entrainment by altering the wave shape of the binaural beat. Conventional binaural beat frequency entrainment previously has been limited to conventional wave shapes, i.e., square triangular sinusoidal, or in some cases, the various musical instruments. For example, it is known that radiant energy, such as sound in this case, may be defined by its frequency, amplitude, and wave shape. A musical note is a particularly suitable example of this. Generally, the musical note A above middle C in the twelve tone diatonic scale is assigned a frequency of 440 cycles per second. The amplitude of that note is expressed as the loudness of the signal. However, the wave shape of that note is related strongly to the instrument used. An A played on a trumpet is quite different from an A played on a violin.

The similarity results from the distinct shapes of the waveforms of each instrument. Similarly, human brain waves also have unique wave shapes, wave contours which are neither sinusoidal, nor square, nor triangular, nor like those of any musical instrument.

In accordance with the invention, human brain waves, in the form of EEGs, are superimposed upon specific stereo audio signals, known as carrier frequencies which are within the range of human hearing. Thus the invention relates not only to techniques of generating the binaural beat, but also to specific waveforms of the binaural beat in frequency, waveshape, and amplitude, and most particularly to the source of the data used to produce such waveforms.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. 1A–1C show two waves at different frequencies, and the resulting binaural beat, respectively;

FIGS. 2A–2D show an input wave, two stereo carrier waves as determined by Fourier analysis, and the resultant binaural beat wave, which matches the contour of the input wave;

FIGS. 3A–3B, 3C–3D, 3E–3F, and 3G–3H are pairs of graphs showing a normal waking EEG and FFR responses in different signal ranges, respectively; and

FIGS. 4A–4F show topographic brain maps of the neocortex of a subject in a normal waking state, and after listening to a binaural beat sound pattern.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

As will be discussed below, different regions of the brain produce distinct electrical waveforms during various physical, mental, and emotional states of consciousness. In the method of the invention, binaural beat audio wave shapes are made to match such particular brain waves as they occur during any mental physical, and emotional human condition of consciousness. Thus, it is possible to convert waveforms from specific brain regions, as well as complete brain surface electrical topography.

In the audio application of the invention, using sampled EEG waveforms from a subject in specific states of consciousness and activity, mental and/or physical, these waveforms are impressed upon multiple sets of sound carrier waves within the human spectrum of hearing. Thus, the waveforms translate into wave amplitude modulations of the carrier to effect what is called a frequency following response, or FFR, as mentioned above.

Some description of the empirical procedure used in the course of developing the invention will be useful, as background. In the 1970s, testing was done on various subjects for effective EEG frequencies using audio signals as a human stimulus. Such frequencies were replicated as amplitude modulation of single-channel audio signals within human hearing ranges, for use in sleep-inducing, attention-focusing, etc.

Where particular subjects responded especially well, those signals were converted to binaural beat patterns. The binaural beat signals were derived by first selecting frequencies of the single-channel audio signals based on the well-known "Oersted Curve", named after the famous 19th century physicist. Using this curve permitted selection of specific audio frequencies to provide the greatest binaural beat frequencies at a much lower range. The effectiveness of the tests were doubled as a result of using binaural beat signals.

In the mid 1980s, EEG waveforms themselves were examined as produced by the binaural signals employed. FFR and entrainment factors thought to be responsible for success were verified. One of the results identified as the probable cause of such effectiveness was the synchronization of the brain hemispheres in such signal frequency ranges (i.e. the induced signals were present simultaneously in major portions of both brain hemispheres).

Experimentation expanded to different subjects in similar states of consciousness. Isolation of EEG patterns in these states of consciousness, and conversion of these patterns to binaural sound, with subsequent reapplication of the binaural sounds produced significantly enhanced results. The effect was especially apparent among naive subjects.

Recently, EEG neuromapping began of subjects with particular talents, where those subjects could utilize those talents (e.g. playing a piano sonata, or solving a mathematical equation) at a mental or visualization level. It was possible to isolate the EEG waveforms related to utilization of those talents, and to convert those waveforms to binaural sound. Subsequent exposure of the subject to such patterns enhanced the individual's ability to replicate the process. Exposing other subjects to the signals produced a learned response through repetition.

Thus, the inventor believes that the inventive process, while not necessarily creating a musician or a mathematician, will set up an EEG ambiance in which learning will be facilitated.

Looking more closely now at the implementation and effects of the invention, FIGS. 2A–2D show a phenomenon wherein an input brain wave signal from a particular brain region is superimposed on stereo carrier waves. FIG. 2D shows the resultant binaural beat wave which matches the contour of the input wave.

The generation and propagation of the binaural beat may be understood from the following series of equations, based on the following.

9(

taking the components from FIGS. 1A-1C, and scaling each component to an appropriate factor (say, α and β).

These components could be recombined to form a beat in accordance with the original components. Linearity and orthogonality principles make these manipulations possible. First, assign the measured wave to be the beat frequency, x.

$$x=\alpha-\beta$$

From the beating waves discussed with respect to FIGS. 1A-1C:

$$\sin(\alpha)+\sin(\beta)=2\cos\left(\frac{\alpha-\beta}{2}\right)\sin\left(\frac{\alpha+\beta}{2}\right)=$$

$$2\cos\left(\frac{x}{2}\right)\sin\left(\frac{2\alpha-x}{2}\right)$$

$$\cos(\alpha)+\cos(\beta)=2\cos\left(\frac{\alpha-\beta}{2}\right)\cos\left(\frac{\alpha+\beta}{2}\right)=$$

$$2\cos\left(\frac{x}{2}\right)\cos\left(\frac{2\alpha-x}{2}\right)$$

Now, let us alter the Fourier series f(x) to produce the beat in the shape of the original wave, f'(x):

$$f(x)=\frac{1}{2}a_0+\frac{1}{2}\sum_{n=1}^{\infty}\cos\left(\frac{nx}{2}\right)\left[a_n\cos\left(n\frac{2\alpha-x}{2}\right)-b_n\sin\left(n\frac{2\alpha-x}{2}\right)\right]\Bigg)$$

$$=\frac{1}{2}a_0+\frac{1}{2}\sum_{n=1}^{\infty}\cos\left(n\frac{\alpha-\beta}{2}\right)\left[a_n\cos\left(n\frac{\alpha+\beta}{2}\right)+b_n\sin\left(n\frac{\alpha+\beta}{2}\right)\right]\Bigg)$$

$$=\frac{1}{2}a_0+\frac{1}{2}\sum_{n=1}^{\infty}\cos\left(n\frac{\alpha-\beta}{2}\right)a_n\cos\left(n\frac{\alpha+\beta}{2}\right)+\cos\left(n\frac{\alpha-\beta}{2}\right)b_n\sin\left(n\frac{\alpha+\beta}{2}\right)\Bigg]\Bigg)$$

$$=\frac{1}{2}a_0+\frac{1}{2}\sum_{n=1}^{\infty}a_n[\cos(\alpha)+\cos(\beta)]+b_n[\sin(\alpha)+\sin(\beta)])$$

$$=\frac{1}{4}a_0+\frac{1}{2}\sum_{n=1}^{\infty}[a_n\cos(\alpha)+b_n\sin(\alpha)]+\frac{1}{4}a_0+\frac{1}{2}\sum_{n=1}^{\infty}[a_n\cos(\beta)+b_n\sin(\beta)]$$

$$=\frac{1}{2}\left(\frac{1}{2}a_0+\sum_{n=1}^{\infty}[a_n\cos(\alpha)-b_n\sin(\alpha)]\right)+\frac{1}{2}\left(\frac{1}{2}a_0+\sum_{n=1}^{\infty}[a_n\cos(\beta)+b_n\sin(\beta)]\right)$$

$$=\frac{1}{2}g(\alpha)+\frac{1}{2}h(\beta)$$

From the foregoing, it can be seen readily that g(α) and h(β) have become two waves, each having half the amplitude of the original wave, the combination of these waves producing a beat which is the input shape f(x).

Thus, using two-channel stereo sound, it is possible to modulate two separate sets of carrier waves so that the replicated EEG waveforms are created as differential beat frequencies between the separate sets. Thus, the method permits the direct application on a frequency base without having to consider the limitation of the spectrum of human hearing. The brain itself synthesizes the signals which cause the effect.

One example may be as follows. If a carrier frequency of 100 Hz were employed in one channel of the audio signal, and a carrier frequency of 104 Hz were employed in the other channel, a binaural beat of 4 Hz would result. In EEG waveform synthesis, as many as 100 separate carrier pairs may be used or a single broad-banded carrier pair may be used to generate a similar number of specific binaural beats that replicate the EEG waveforms in both frequency and amplitude.

A 4 Hz, or a 5 Hz binaural beat would be too low in frequency to hear. Using the Oersted curve mentioned above, the most effective harmonic carrier would be 275 Hz, which is within hearing range. For the multiple waveform situation just discussed, the differential between carrier waves on a single channel also is utilized to produce an FFR.

One type of audio pattern found to be particularly useful in implementing the inventive method is what is known to the inventor as Phased Pink Sound. The full spectrum of audible sound is known commonly as "white" noise. "Pink" sound is known to result from an adjustment in amplitude of white sound to compensate for decline in perception by the human ear at both ends of the audible spectrum.

Phased Pink Sound results from the relative rotational shifting of pink sound from one stereo audio channel to another with cyclic changes in amplitude, frequency, and rate of panning. Such changes generally are synchronous with selected waveforms within the multiple patterns of the binaural beat generating system. Studies have shown that using Phased Pink Sound at a level at least 10 dB lower than the binaural beat signals produces as much as a 30% enhancement in FFR within the EEG waveforms of the listening individual. There is some basis for concluding that Phased Pink Sound provides an audio base that assists the brain in "synthesizing" the binaural beat frequencies normally inaudible in the human hearing process.

Basically, Phased Pink Sound is generated by a digital processor, which converts mathematical sequences, derived from appropriate algorithms, into audible sound. Such digital processors and their operation are well-known in the art, and so are not discussed here. Inherent in such a system is a frequency sensor that synchronizes the phasing with dominant EEG waveforms as those waveforms are introduced from another source.

Examples of suitable algorithms for implementing Phased Pink Sound are as follows:

a2

```
/*********************************************************************
 * Algorithm to generate 8-bit PCM samples in array pink[] of the    *
 * single channel sound that serves as the source for the stereo     *
 * "phased pink" sound                                               *
 *********************************************************************/

#include <math.h>
#include <stdio.h>

#define M2PI   -6.283185307179586
#define SAMPLES_PER_SECOND 10466.5
#define CUTOFF 200.0       /* cutoff frequency for low-pass filter */
#define S 83732            /* number of samples to generate */
#define MINDELAY 60        /* minimum flanging delay (samples) */
#define MAXDELAY 80        /* maximum flanging delay (samples) */

extern short w[];          /* 8192 entry table of 16-bit sine values
                              scaled from 0x8001 to 0x7FFF */
extern double st_entries;/* count of entries in sine table */

long phase;                /* random number generator phase */
long fa;                   /* filter accumulator */
long fc;                   /* filter constant */
long sweep;                /* flanging filter phase */
long sweep0;               /* flanging filter initial phase */
long ds;                   /* flanging filter phase step */
long count;                /* samples remaining in flanging filter cycle */
long count0;               /* samples in flanging filter cycle */
long delay;                /* current flanging filter delay (XXXX.XXXX) */
long delay0;               /* flanging filter delay constant */
long range;                /* flanging filter delay range */
short gainNS;              /* noise sound gain (gain = gainNS/1024) */
short gainFS;              /* flanging sound gain (gain = gainFS/1024) */
short noise[S+MAXDELAY]; /* array to receive noise samples */
short offset;              /* final sample offset to balance values */
short scaleF;              /* final scale factor to range samples */
char pink[S];              /* array to receive "phased pink" samples */


/*********************************************************************
 * Main program                                                      *
 *********************************************************************/

main()
{
   long control_base;     /* initial flanging delay */
   long control_range;    /* range of flanging control */
   int i;                 /* loop index */
   short *np;             /* pointer to filtered noise sample array */
   short *fsnp;           /* pointer to initial/final noise sequence */
   short NoiseGen();      /* next filtered noise sample */
   short Flange();        /* flanging sample */
   short xx;              /* output before final scaling */
   /* Initialize the white noise generator */
   phase = 0x8000;
   /* Initialize low-pass filter */
   fa = 0;
   fc = (1.0 - exp(M2PI * CUTOFF / SAMPLES_PER_SECOND)) * 65536.0;

   /* Initialize flanging filter for 8 second cycle.  Delay sweeps
      sinusodially around 5*PI/2.  Flanginging tone gain is 75%
      of the noise tone */
   sweep = sweep0 = ((long)((.75 * st_entries) * 65536.0 + 0.5))
      & 0x1FFFFFFF;
   control_base = w[sweep0 >> 16];
   control_range = 0x0007FFFL - control_base;
   range = (((double)(MAXDELAY - MINDELAY) * 32767.0) / control_range)
     * 16.0 + 0.5;
   delay0 = (MINDELAY << 16) - control_base * (range >> 3);
   ds = (st_entries / (8.0 * SAMPLES_PER_SECOND)) * 65536.0 + 0.5;
   count =  8.0 * SAMPLES_PER_SECOND + 0.5;
   gainNS = 585;
   gainFS = 439;

   /* Initialize the final offset and scale factor for these filter
      parameters (empirically determined) */
   offset = 153;
   scaleF = 0x245;

   /* Generate an initial sequence of noise samples to provide for
      delayed samples */
   np = fsnp = noise;
   for (i = 0; i < MAXDELAY; i++) *np++ = NoiseGen();

   /* Generate the next S samples of "phased pink" sound */
   for (i = 0; i < S; i++) {

      /* Generate the next colored noise sample.  For looping,
         finish off with the initial noise sequence */
      if (i < S-MAXDELAY) *np = NoiseGen();
      else *np = *fsnp++;

      /* Apply a sweeping cosine comb filter to flange the sound */
      xx = (*np*gainNS + Flange(np)*gainFS) >> 10;
      pink[i] = ((xx + offset) * scaleF) >> 16;
      np++;

   }
}

/************************************************************************
 * NoiseGen -- function to generate a filtered noise sample            *
 ************************************************************************/

short NoiseGen(nsp)
{
   long x;                    /* current noise sample */
   long y;                    /* current filtered noise sample */

   /* Generate sinusodial density noise from white */
   phase = phase << 1;
```

```
      if (phase & 0x10000) phase = phase ^ 0x1DE7;
      phase = phase & 0xFFFF;
      x = w[phase >> 3];

      /* Apply 1st order low-pass digital filter */
      y = (fc*fa) >> 16;
      fa += (x >> 4) - y;
      return((short)(y << 4));
   }


   /************************************************************************
    * Flange — function to generate a flanging noise sample               *
    ************************************************************************/

   short Flange(nsp)
      short *nsp;              /* pointer to current noise sample */
   {
      short f;                 /* flanging noise sample */
      short *dnp;              /* pointer to delayed noise sample */

      /* Apply a sinusodially sweeping comb filter to flange the sound */
      if (count--) sweep = (sweep + ds) & 0x1FFFFFFF;
      else {
         sweep = sweep0;
         count = count0;
      }

      /* Compute the filter delay and linearly interpolate between
         noise samples to simulate a continuously variable delay */
      delay = delay0 + ((w[sweep >> 16] * range) >> 3);
      dnp = nsp - (delay >> 16);
      f = *dnp +
         ((((*(dnp-1) - *dnp) >> 1) * ((delay & 0xFFFF) >> 1)) >> 14);
      return(f);
   }
```

Looking at some results of the inventive method, FIG. 3A shows the EEG of a subject in a normal waking state. FIG. 3B shows an EEG of the individual after listening to binaural beat sounds produced in accordance with the invention. The Figure shows an FFR response in the 1.5, 4, and 6 Hz signal range.

Likewise, FIG. 3C shows the EEG of a subject in a normal waking state, and FIG. 3D shows an EEG of the individual after listening to other binaural beat sounds produced in accordance with the invention. The Figure shows an FFR response in the 2, 4, and 7 Hz signal range.

FIG. 3E shows the EEG of a subject in a normal waking state, and FIG. 3F shows an EEG of the individual after listening to still other binaural beat sounds produced in accordance with the invention. The Figure shows an FFR response in the 0.5, 3, and 4 Hz signal range.

Finally, FIG. 3G shows the EEG of a subject in a normal waking state, and FIG. 3*h* shows an EEG of the individual after listening to still other binaural beat sounds produced in accordance with the invention. The Figure shows FFR response to 1.5, 2, and 4 Hz signals in amplitude, by frequency.

FIGS. 4A–4C shows a typical contour map of a subject in a normal waking state. It should be noted that the map shows a lack of continuity. Note also the lack of significant amplitude patterns ranging between temporal lobes, and the relative lack of intensity within the frontal area.

In contrast, FIGS. 4D–4F shows a contour map of the same individual after listening to binaural beat sound in accordance with the invention. Note the synchronization between hemispheres, and the high amplitude of activity at the frontal portion of the brain. Note also how the left and right hemisphere brain waves exhibit significantly higher amplitudes in the frequencies found in the original sound stimulus.

The application of the binaural beat signals by headphones or other second producing devices causes the following results:

1. When such audio signals are provided simultaneously with the state of being itself, those specific states can be enhanced. The additional pattern superposed upon the original provides a powerful setting to maintain and/or expand the condition.
2. By recording the audio signals and playing them back, an individual may return to an original or

previously-experienced state of consciousness whenever desired.

3. By listing to recordings of these audio signals, an original pattern or condition induced in one individual may be replicated in other individuals.
4. An individual can be trained, based on sufficient repetition of application of these waveforms, to the point that the individual can recall and replicate these waveforms themselves, without further outside stimulation.

The method of the invention has applications in a number of different areas, not the least of which is the inducement of a state of sleep. Other areas of application include inducement of wakefulness of varying degrees; focusing of attention; inducement of mental and physical relaxation; enhancing intellectual performance in various mental disciplines such as mathematics; enhancement of creativity; the reexperience of previous activity; the acquisition of new abilities which others already have; reinforcement and restoration of weak areas in the mind and body; enhancement and strengthening of mental and/or muscular coordination; and development of integration of entire brain function. Human beings have EEG patterns which are unique to the various states of consciousness and mental and/or physical activity just mentioned, so that the imposition of the appropriate stereo audio signals on the desired EEG wave produces the binaural beat which is necessary to induce the state.

While the invention has been described above in detail with reference to a particular specific embodiment, various modifications within the spirit and scope of the invention will be apparent to those of working skill in this technological field. Thus, the invention should be considered as limited only by the scope of the appended claims.

What is claimed is:

1. A method of inducing states of consciousness in human beings, comprising:
   - providing a replicated electroencephalogram (EEG) waveform indicative of a desired state of consciousness;
   - superimposing said EEG waveform on two separate sets of carrier waves using stereo sound;
   - creating differential beat frequencies between said sets of carrier waves in accordance with said superimposing step; and
   - providing the resulting signals in audio form to respective ears of a human being, to induce said state of consciousness.

2. A method as claimed in claim 1, wherein said creating step includes the step of combining pink with said sets of carrier waves by shifting of said pink sound with respect to said EEG waveform from one stereo audio channel to another, with cyclic changes in amplitude, frequency, and rate of panning.

3. A method as claimed in claim 1, wherein all of said steps are performed repeatedly on a particular individual over a period of time so that the individual is able eventually to reproduce said desired state of consciousness without further audio stimulation.

4. A method as claimed in claim 1, wherein all of said steps are performed using the EEG of one individual, but said applying step is carried out with another individual, so as to transfer the desired state of consciousness of one individual to another.

5. A method as claimed in claim 1, wherein said first providing step comprises the step of providing a plurality of EEG waveforms, indicative of different respective states of consciousness, and each of said superimposing, creating, and second providing steps are performed with each of said plurality of EEG waveforms.

6. A method as claimed in claim 1, wherein said second providing step results in substantial synchronization of major portions of both brain hemispheres of said human being.

* * * * *

United States Patent [19]
Monroe

[11] Patent Number: 5,356,368
[45] Date of Patent: * Oct. 18, 1994

[54] METHOD OF AND APPARATUS FOR INDUCING DESIRED STATES OF CONSCIOUSNESS

[75] Inventor: **Robert A. Monroe**, Nelson County, Va.

[73] Assignee: **Interstate Industries Inc.**, Faber, Va.

[*] Notice: The portion of the term of this patent subsequent to May 25, 2010 has been disclaimed.

[21] Appl. No.: **664,176**

[22] Filed: **Mar. 1, 1991**

[51] **Int. Cl.**^{5} **A61M 21/00**
[52] **U.S. Cl.** **600/28**; 128/732
[58] **Field of Search** 600/26-28; 128/731-732

[56] **References Cited**

U.S. PATENT DOCUMENTS

2,466,054	4/1949	Siebel .	
3,160,159	12/1964	Hoody et al. .	
3,576,185	4/1971	Schulz et al. .	
3,712,292	1/1973	Zentmeyer, Jr. .	
3,753,433	8/1973	Bakerich et al. .	
3,826,243	7/1974	Anderson .	
3,837,331	9/1974	Ross .	
3,884,218	5/1975	Monroe .	
4,034,741	7/1977	Adams et al. .	
4,141,344	2/1979	Barbara .	
4,227,516	10/1980	Meland et al. .	
4,335,710	6/1982	Williamson .	
4,573,449	3/1986	Warnke .	
4,834,701	5/1989	Masaki .	
4,883,067	11/1989	Knispel et al.	600/28
5,036,858	8/1991	Carter et al. .	
5,101,831	4/1992	Koyama et al.	600/26

Primary Examiner—Lee S. Cohen
Assistant Examiner—J. P. Lacyk
Attorney, Agent, or Firm—Sughrue, Mion, Zinn, Macpeak & Seas

[57] **ABSTRACT**

Improved methods and apparatus for entraining human brain patterns, employing frequency following response (FFR) techniques, facilitate attainment of desired states of consciousness. In one embodiment, a plurality of electroencephalogram (EEG) waveforms, characteristic of a given state of consciousness, are combined to yield an EEG waveform to which subjects may be susceptible more readily. In another embodiment, sleep patterns are reproduced based on observed brain patterns during portions of a sleep cycle; entrainment principles are applied to induce sleep. In yet another embodiment, entrainment principles are applied in the work environment, to induce and maintain a desired level of consciousness. A portable device also is described.

28 Claims, 21 Drawing Sheets

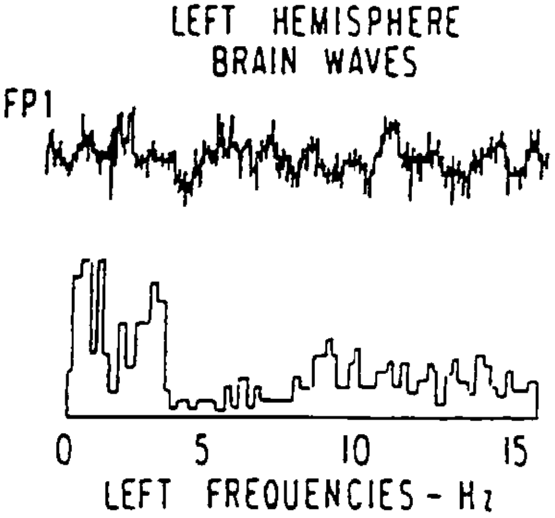

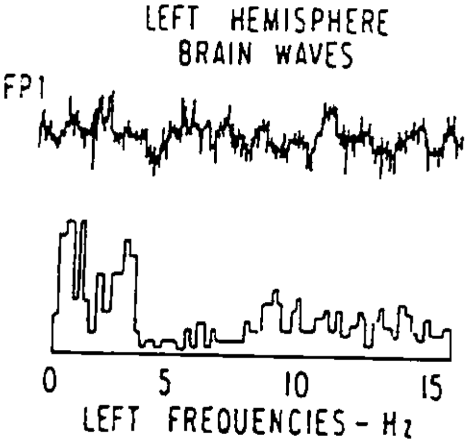

FIG. 1A

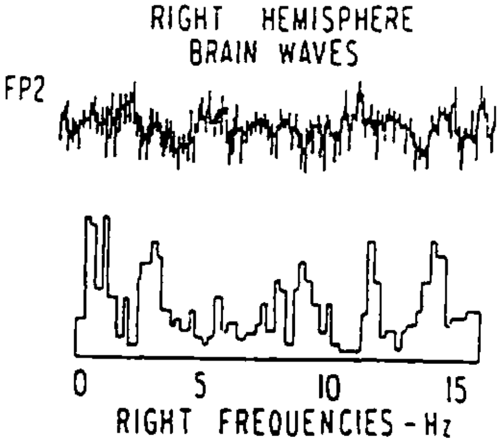

FIG. 1B

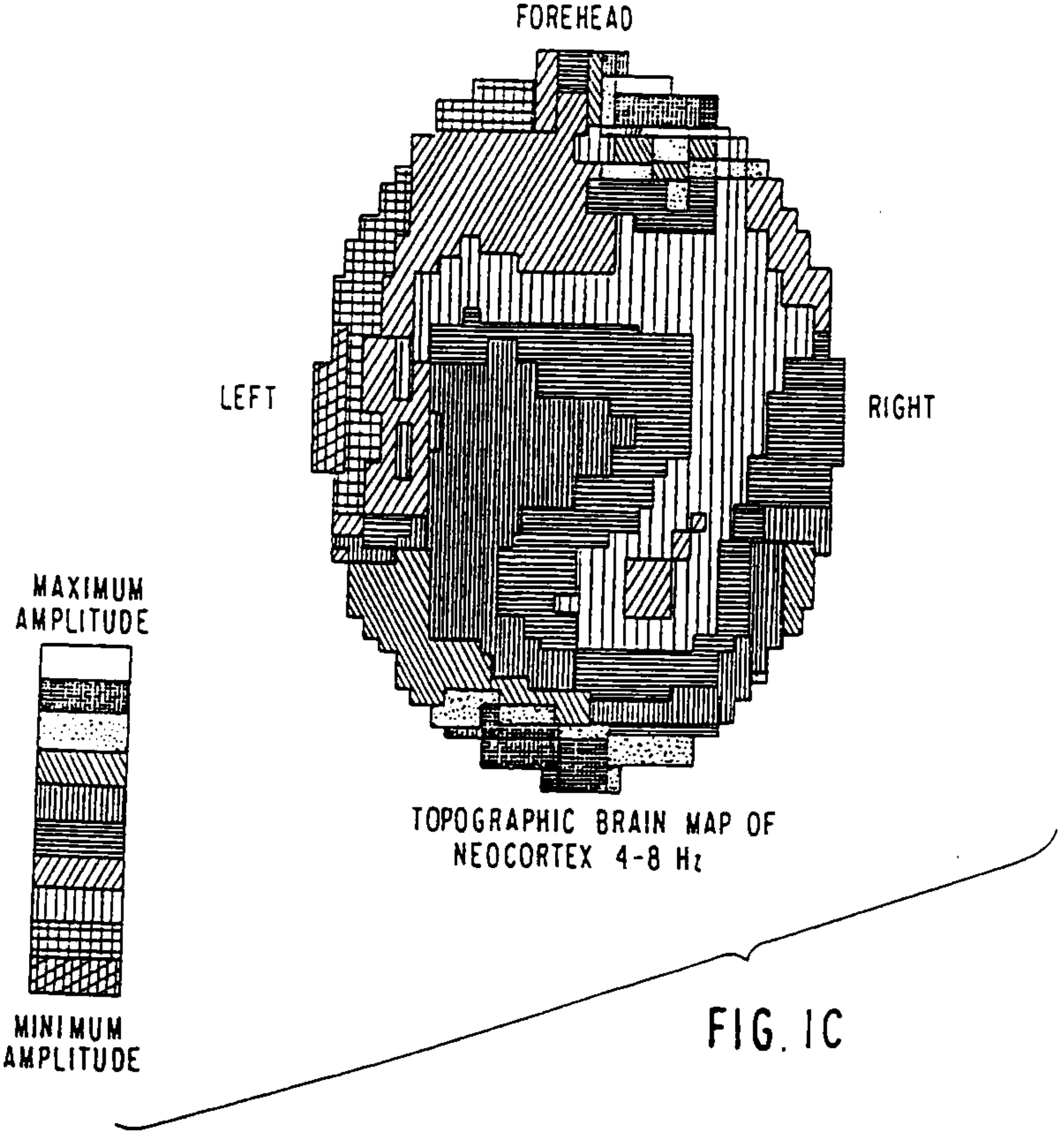
FOREHEAD
LEFT
RIGHT
MAXIMUM
AMPLITUDE
MINIMUM
AMPLITUDE
TOPOGRAPHIC BRAIN MAP OF
NEOCORTEX 4-8 Hz

FIG. 1C

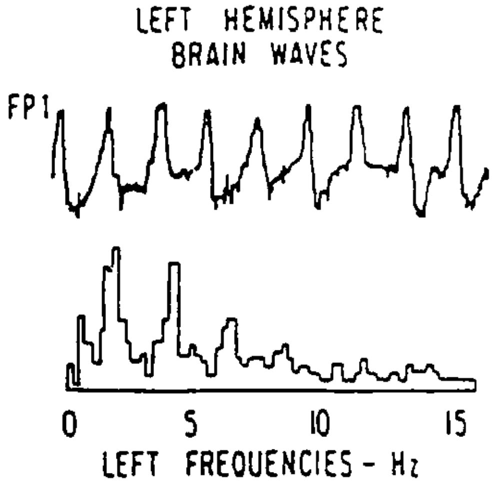

FIG. 1D

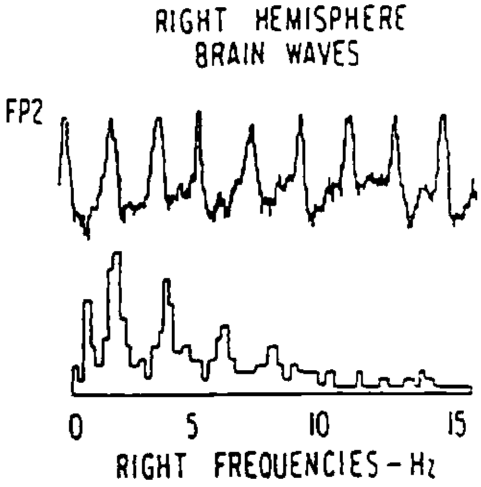

FIG. 1E

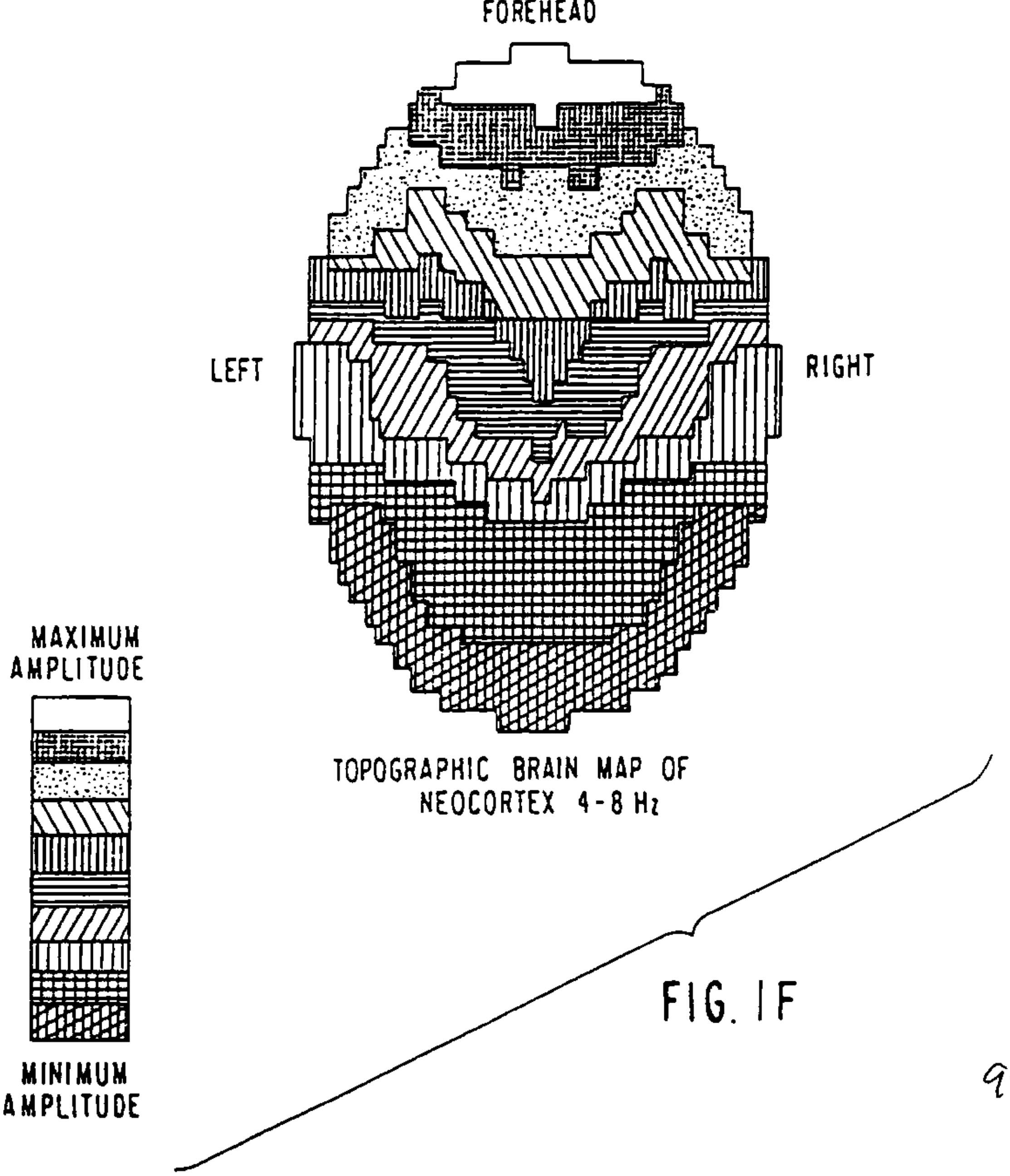
FOREHEAD
LEFT
RIGHT
MAXIMUM
AMPLITUDE
MINIMUM
AMPLITUDE
TOPOGRAPHIC BRAIN MAP OF
NEOCORTEX 4-8 Hz

FIG. 1F

99

FIG. 2

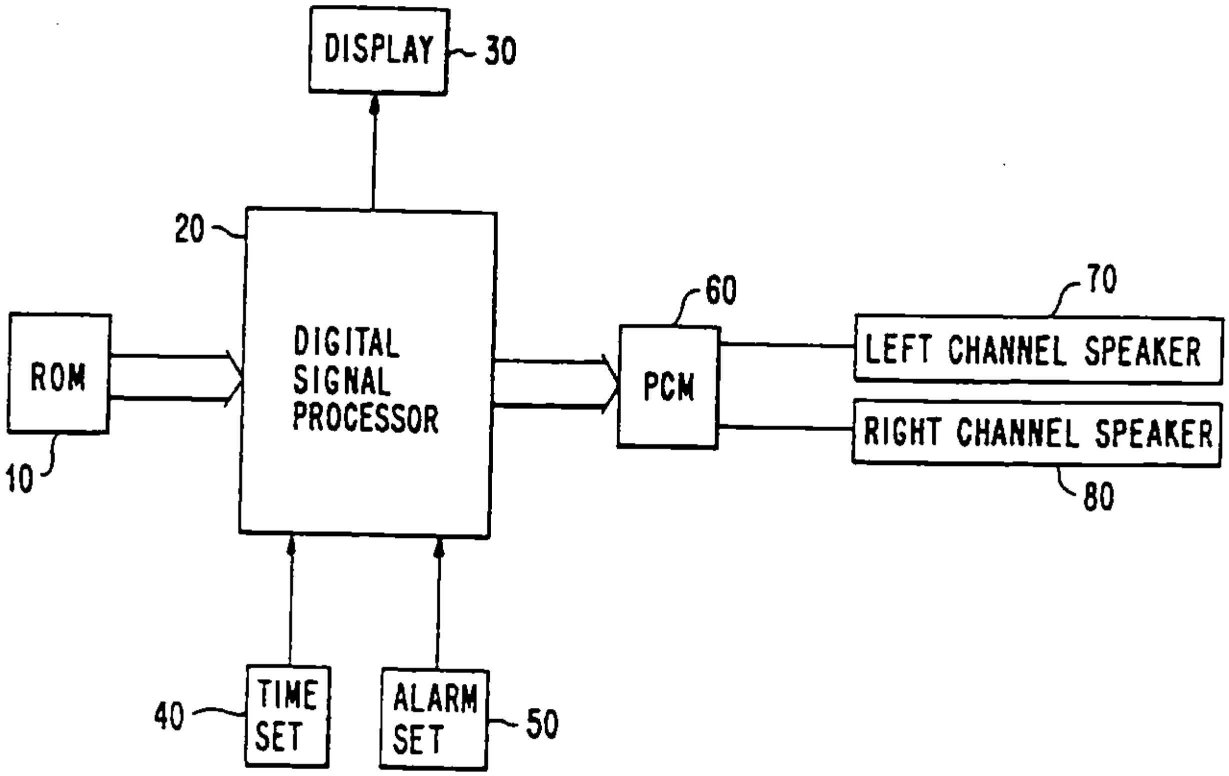
DISPLAY
30
20
DIGITAL SIGNAL PROCESSOR
ROM
10
60
PCM
70
LEFT CHANNEL SPEAKER
RIGHT CHANNEL SPEAKER
80
40
TIME SET
ALARM SET
50

100

FIG. 3

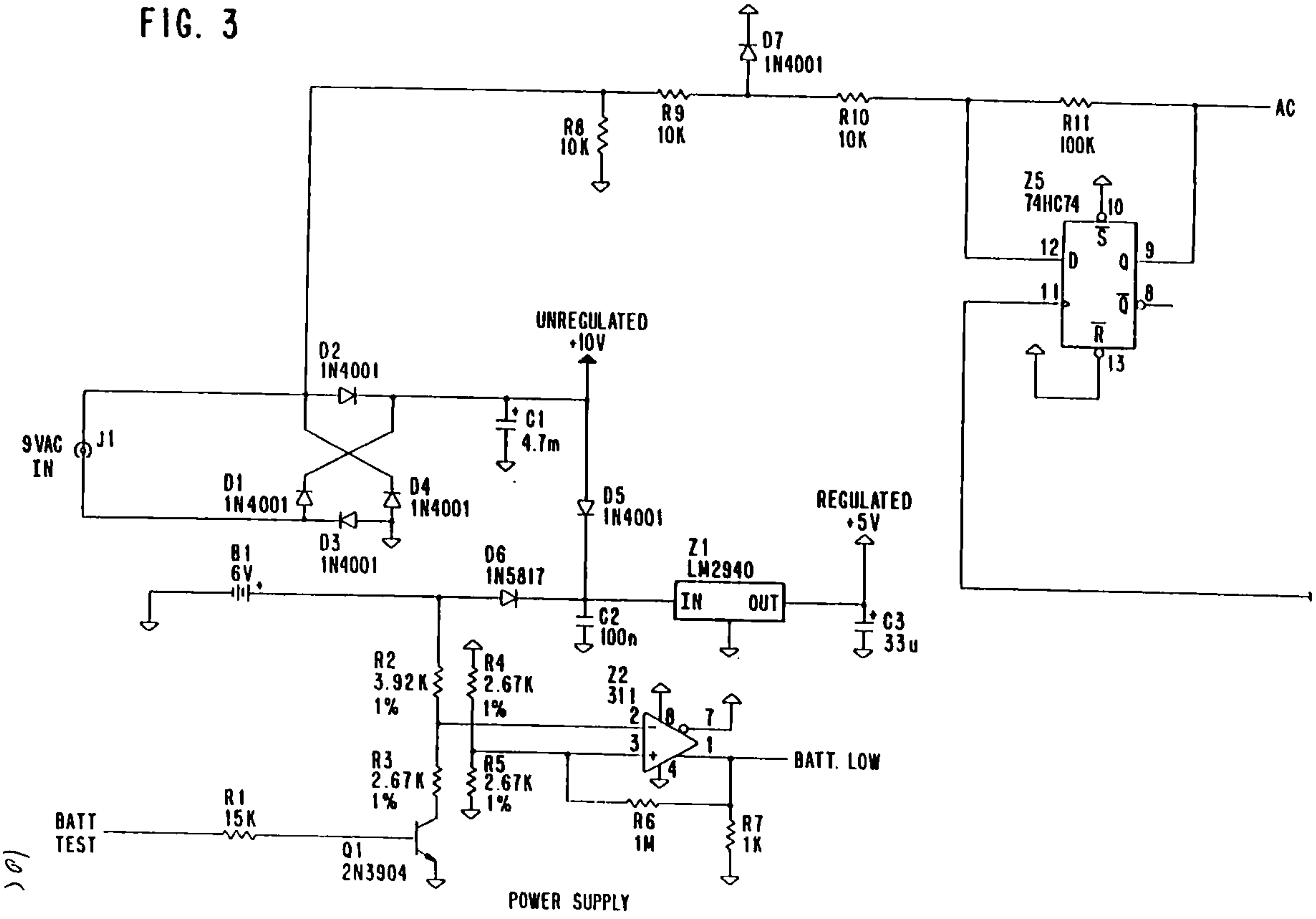
D7
1N4001
R8
10K
R9
10K
R10
10K
R11
100K
AC
Z5
74HC74
10
S̄
12
D
Q
9
11
Q̄
8
R̄
13
UNREGULATED
+10V
D2
1N4001
C1
4.7m
9VAC
IN
J1
D1
1N4001
D4
1N4001
D3
1N4001
D5
1N4001
REGULATED
+5V
B1
6V
D6
1N5817
Z1
LM2940
IN
OUT
C2
100n
C3
33u
R2
3.92K
1%
R4
2.67K
1%
Z2
311
2
8
7
3
1
4
BATT. LOW
R3
2.67K
1%
R5
2.67K
1%
R6
1M
R7
1K
BATT
TEST
R1
15K
Q1
2N3904
POWER SUPPLY

10(

FIG. 3(cont.)

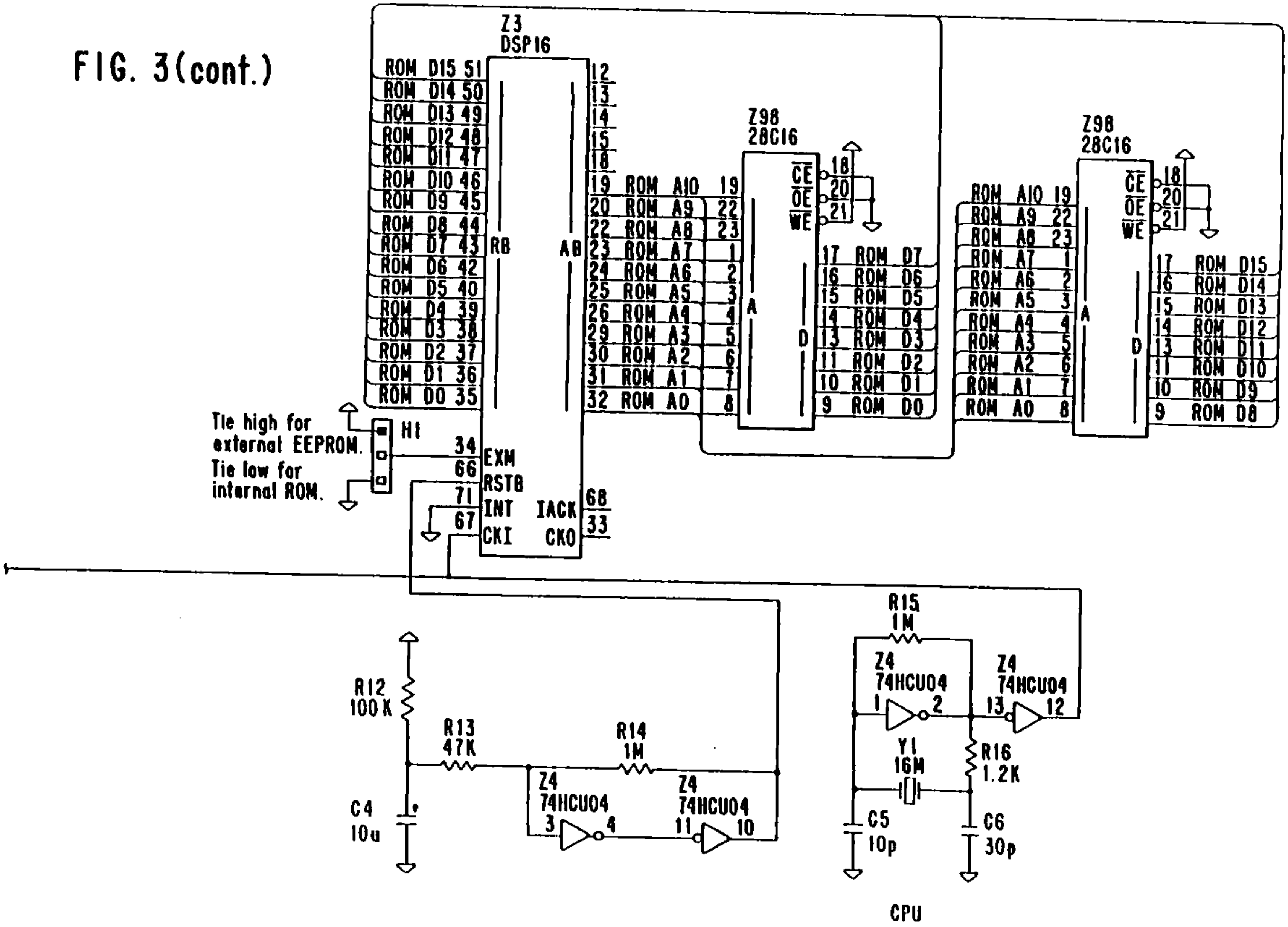
Z3
DSP16
ROM D15 51
ROM D14 50
ROM D13 49
ROM D12 48
ROM D11 47
ROM D10 46
ROM D9 45
ROM D8 44
ROM D7 43
ROM D6 42
ROM D5 40
ROM D4 39
ROM D3 38
ROM D2 37
ROM D1 36
ROM D0 35
RB
AB
12
13
14
15
18
19 ROM A10
20 ROM A9
22 ROM A8
23 ROM A7
24 ROM A6
25 ROM A5
26 ROM A4
29 ROM A3
30 ROM A2
31 ROM A1
32 ROM A0
34 EXM
66 RSTB
71 INT
67 CKI
IACK 68
CKO 33
H1
Tie high for external EEPROM.
Tie low for internal ROM.
Z98
28C16
19
22
23
1
2
3
4
5
6
7
8
A
D
CE 18
OE 20
WE 21
17 ROM D7
16 ROM D6
15 ROM D5
14 ROM D4
13 ROM D3
11 ROM D2
10 ROM D1
9 ROM D0
Z98
28C16
ROM A10 19
ROM A9 22
ROM A8 23
ROM A7 1
ROM A6 2
ROM A5 3
ROM A4 4
ROM A3 5
ROM A2 6
ROM A1 7
ROM A0 8
A
D
CE 18
OE 20
WE 21
17 ROM D15
16 ROM D14
15 ROM D13
14 ROM D12
13 ROM D11
11 ROM D10
10 ROM D9
9 ROM D8
R12
100 K
C4
10u
R13
47K
R14
1M
Z4
74HCU04
3
4
Z4
74HCU04
11
10
R15
1M
Z4
74HCU04
1
2
Z4
74HCU04
13
12
Y1
16M
R16
1.2K
C5
10p
C6
30p
CPU

102

FIG. 4

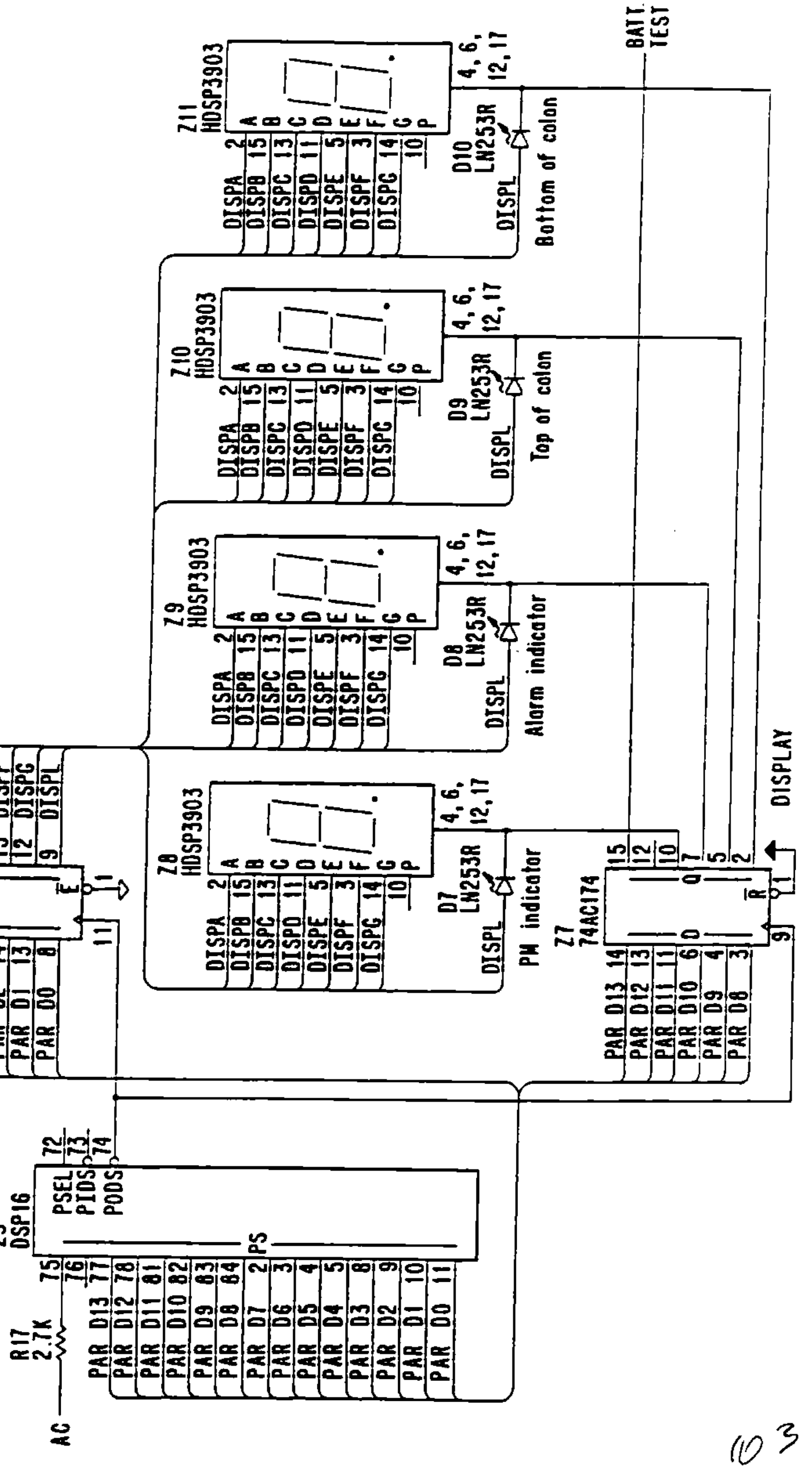
Z3
DSP16
PSEL
PIDS
PODS
72
73
74
PS
AC
R17
2.7K
PAR D13 75
PAR D12 76
PAR D11 77
PAR D10 78
PAR D9 81
PAR D8 82
PAR D7 83
PAR D6 84
PAR D5 2
PAR D4 3
PAR D3 4
PAR D2 5
PAR D1 8
PAR D0 9
10
11
Z6
74AS374
PAR D7 3
PAR D6 18
PAR D5 17
PAR D4 4
PAR D3 7
PAR D2 14
PAR D1 13
PAR D0 8
D
Q
E
11
1
2 DISPA
19 DISPB
16 DISPC
5 DISPD
6 DISPE
15 DISPF
12 DISPG
9 DISPL
Z7
74AC174
PAR D13 14
PAR D12 13
PAR D11 11
PAR D10 6
PAR D9 4
PAR D8 3
D
Q
R
9
1
15
12
10
7
5
2
DISPLAY
BATT. TEST
Z8
HDSP3903
Z9
HDSP3903
Z10
HDSP3903
Z11
HDSP3903
DISPA 2 A
DISPB 15 B
DISPC 13 C
DISPD 11 D
DISPE 5 E
DISPF 3 F
DISPG 14 G
10 P
4, 6, 12, 17
D7
LN253R
PM indicator
D8
LN253R
Alarm indicator
D9
LN253R
Top of colon
D10
LN253R
Bottom of colon
DISPL

103

FIG. 5

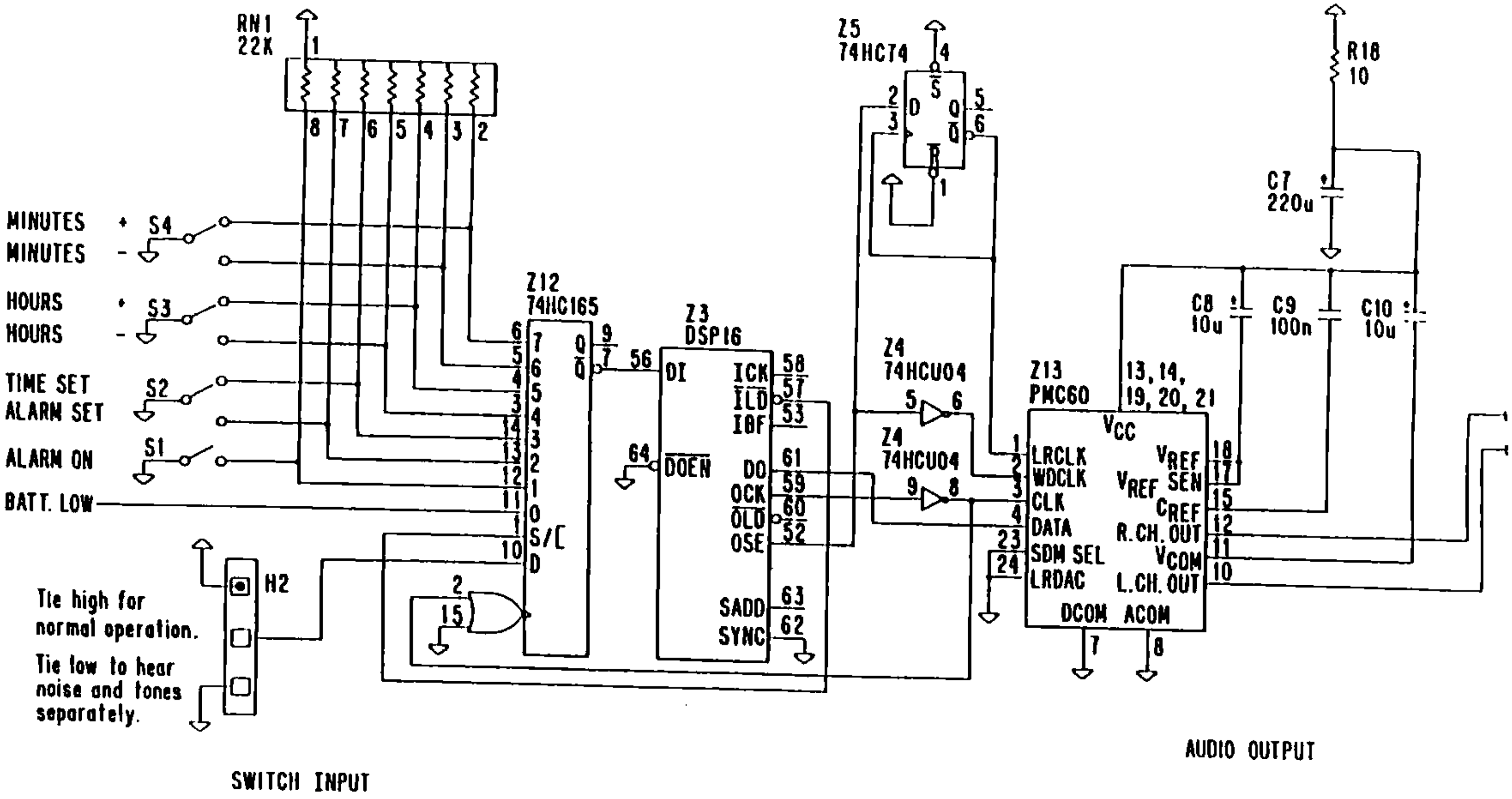

RN1
22K
MINUTES +
MINUTES -
HOURS +
HOURS -
TIME SET
ALARM SET
ALARM ON
BATT. LOW
S4
S3
S2
S1
H2
Tie high for normal operation.
Tie low to hear noise and tones separately.
SWITCH INPUT
Z12
74HC165
S/L
Z3
DSP16
DI
DOEN
ICK
ILD
IBF
DO
OCK
OLD
OSE
SADD
SYNC
Z5
74HC74
Z4
74HCU04
Z4
74HCU04
Z13
PMC60
13, 14,
19, 20, 21
VCC
LRCLK
WDCLK
CLK
DATA
SDM SEL
LRDAC
VREF
VREF SEN
CREF
R.CH. OUT
VCOM
L.CH. OUT
DCOM
ACOM
R18
10
C7
220u
C8
10u
C9
100n
C10
10u
AUDIO OUTPUT

104

FIG. 6A

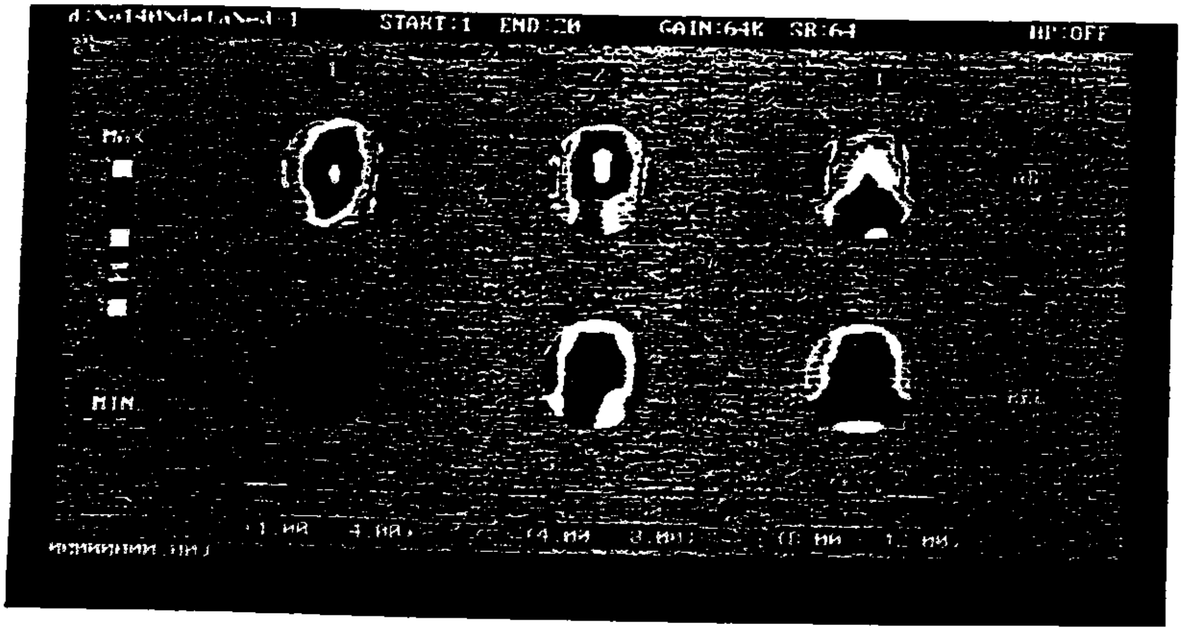
START:1 END:20
GAIN:64K SR:64
HP:OFF
1
MAX
MIN
OD
REL

FIG. 6B

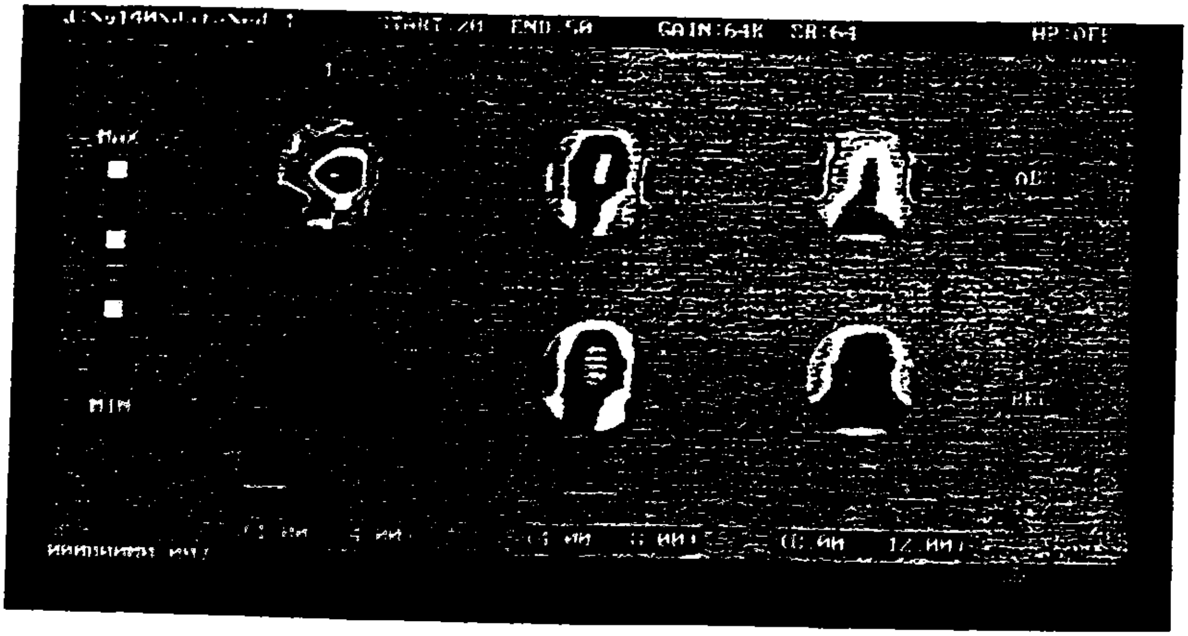
START:20 END:50
GAIN:64K SR:64
HP:OFF
1
MAX
MIN
OD
REL

105

FIG. 6C

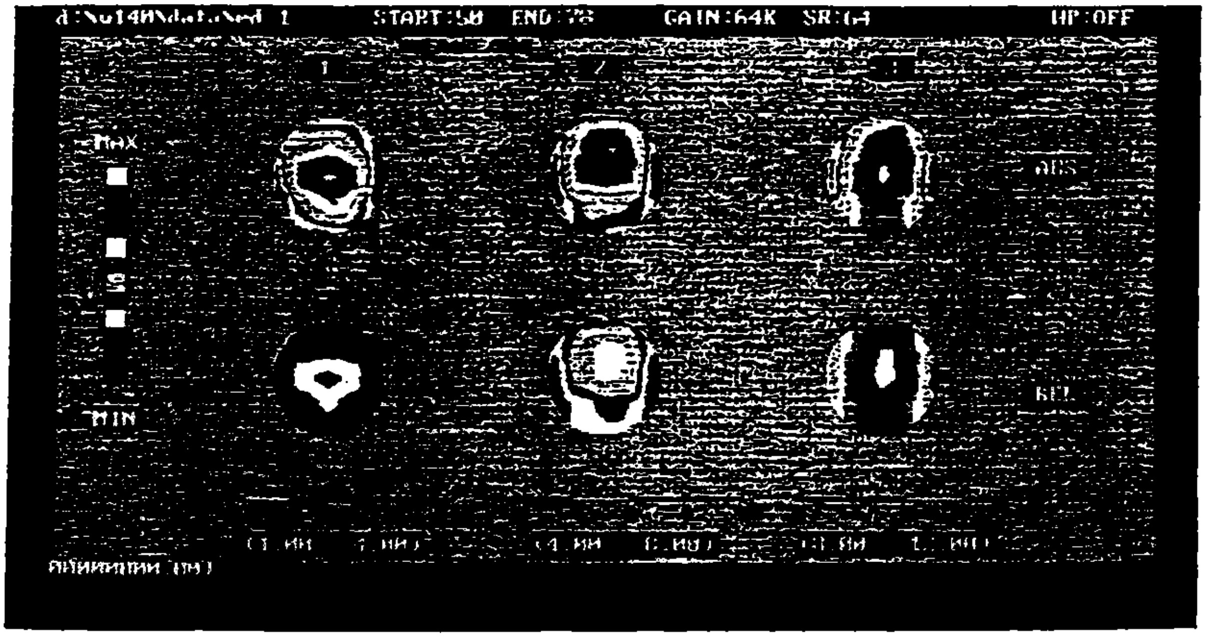
d:\u140\data\ed 1
START:58
END:78
GAIN:64K
SR:64
HP:OFF
1
2
3
MAX
MIN
ABS
REL
(1.00 4.00)
(4.00 8.00)
(8.00 12.00)

FIG. 6D

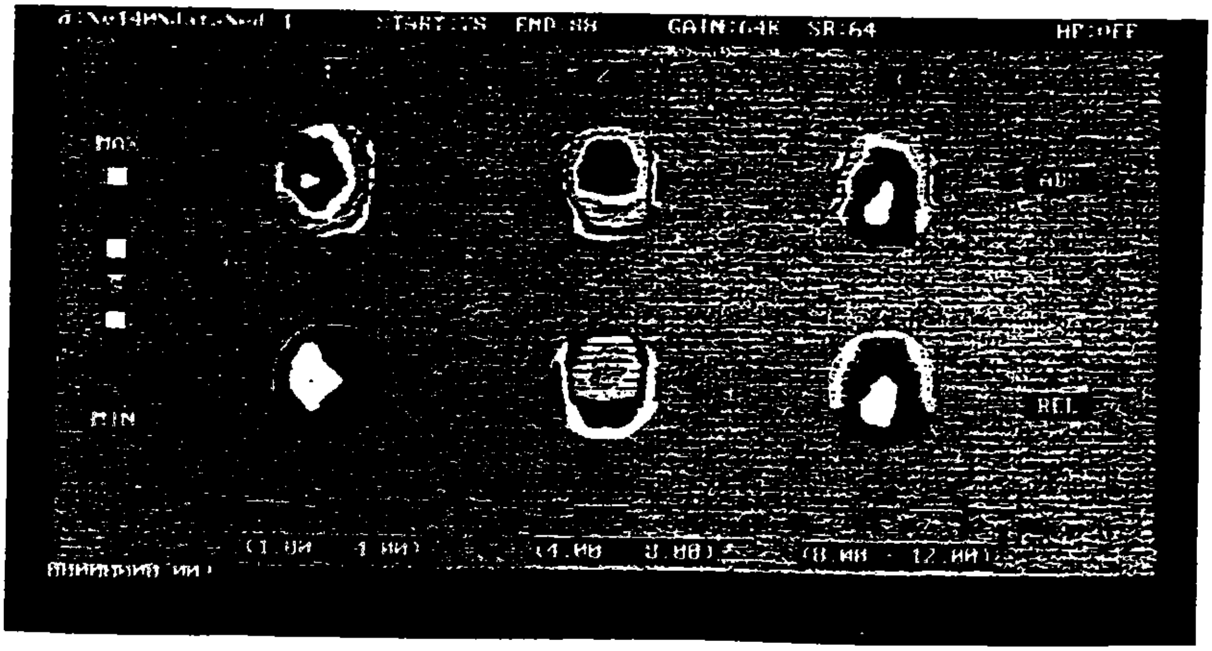
d:\u140\data\ed 1
END:88
GAIN:64K
SR:64
HP:OFF
1
2
3
MAX
MIN
ABS
REL
(1.00 4.00)
(4.00 8.00)
(8.00 12.00)

FIG. 5 (cont.)

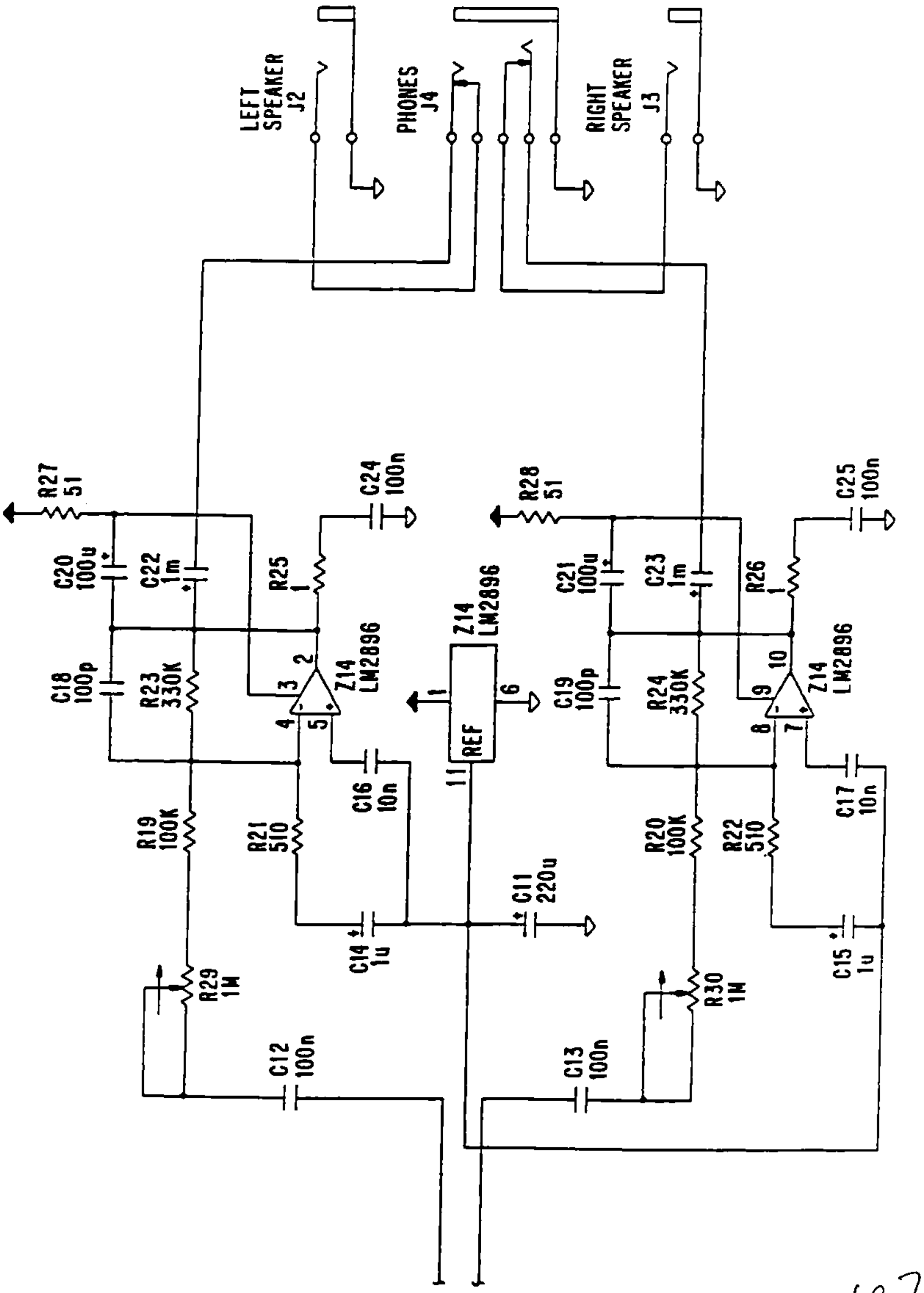
LEFT SPEAKER J2
PHONES J4
RIGHT SPEAKER J3
R27 51
C20 100u
C22 1m
R25 1
C24 100n
C18 100p
R23 330K
Z14 LM2896
2
3
4
5
R19 100K
R21 510
C16 10n
Z14 LM2896
REF
1
6
11
C11 220u
C14 1u
R29 1M
C12 100n
R28 51
C21 100u
C23 1m
R26 1
C25 100n
C19 100p
R24 330K
10
9
8
7
Z14 LM2896
R20 100K
R22 510
C17 10n
C15 1u
R30 1M
C13 100n

107

FIG. 6E

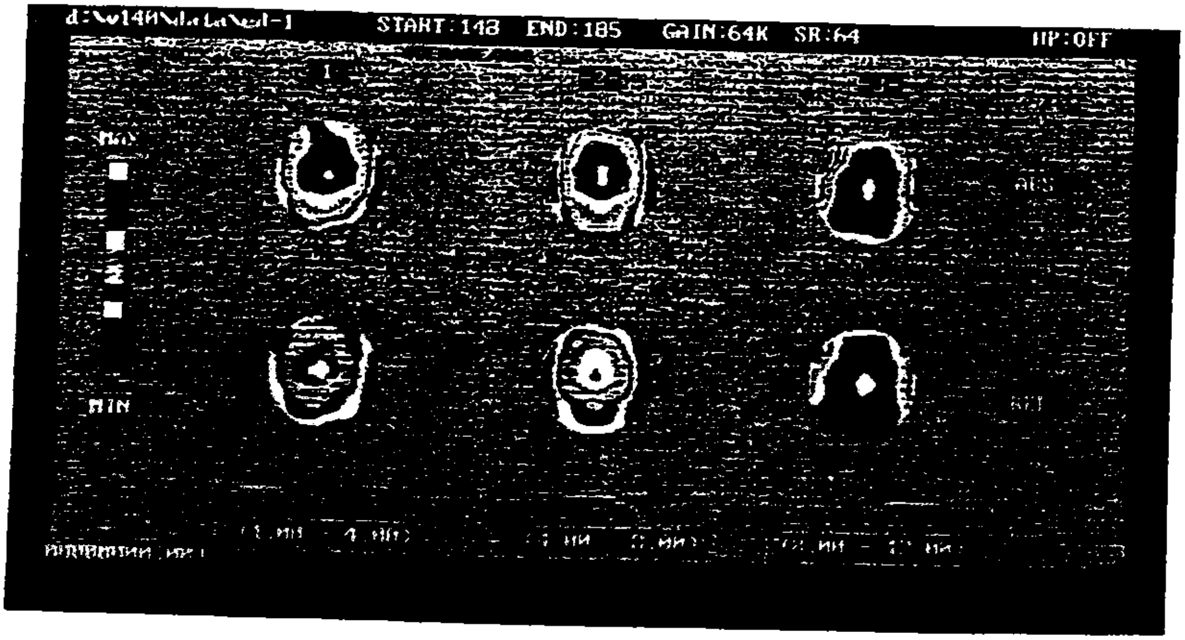
START:148
END:185
GAIN:64K
SR:64
HP:OFF
1
2
MAX
MIN
ABS
REF
(1 MM - 4 MM)

FIG. 6F

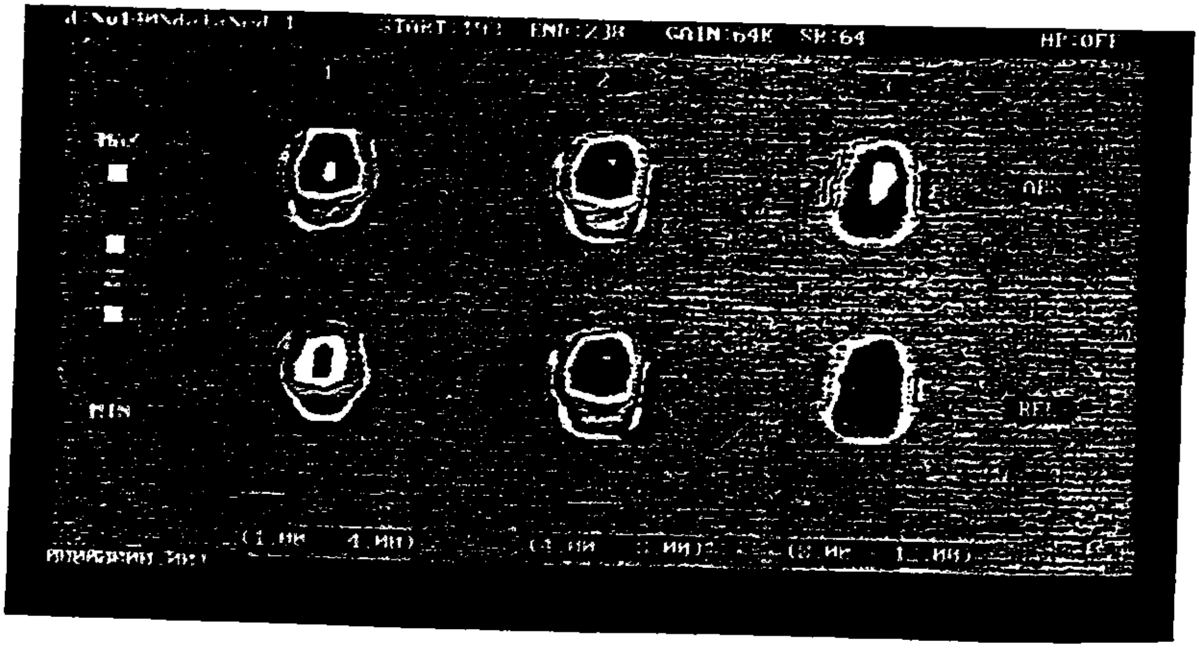
GAIN:64K
SR:64
HP:OFF
1
2
MAX
MIN
ABS
REF
(1 MM - 4 MM)

FIG. 6G

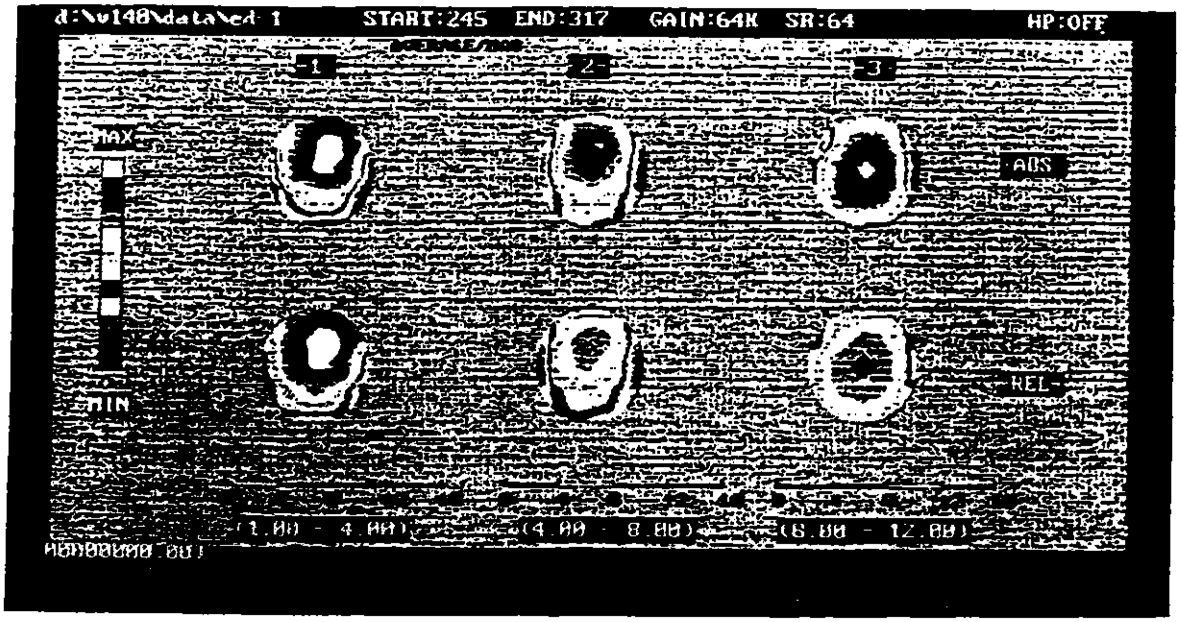
d:\v148\data\ed-1
START:245
END:317
GAIN:64K
SR:64
HP:OFF
1
2
3
MAX
MIN
ABS
REL
(1.00 - 4.00)
(4.00 - 8.00)
(8.00 - 12.00)
00000000.00!

FIG. 6H

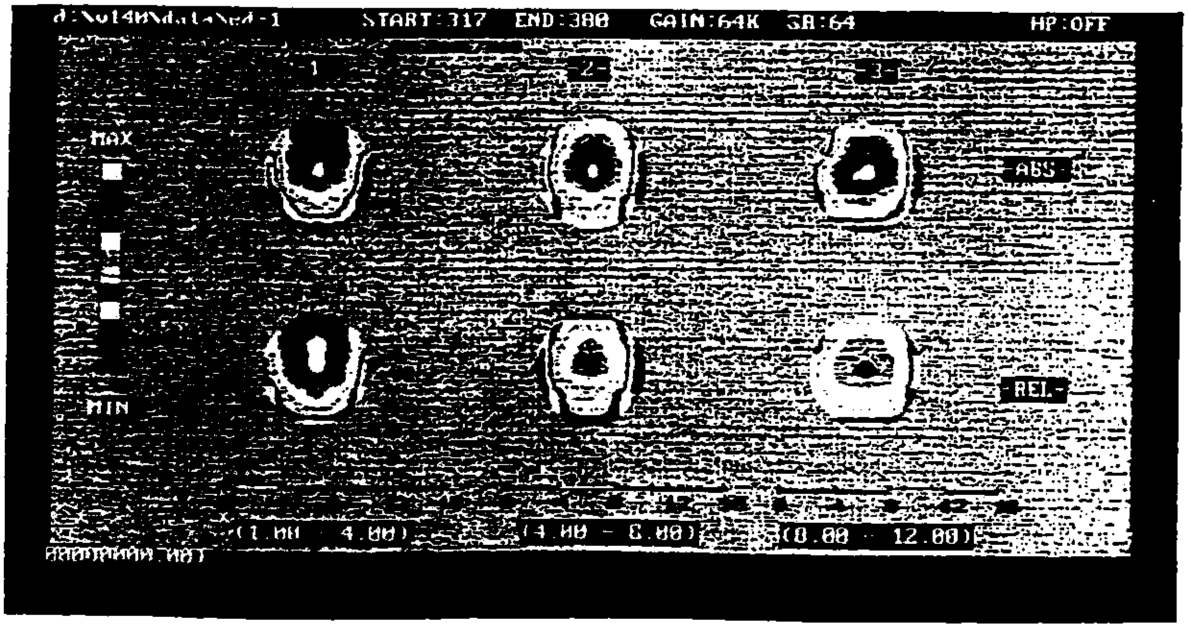
d:\v148\data\ed-1
START:317
END:388
GAIN:64K
SR:64
HP:OFF
1
2
3
MAX
MIN
ABS
REL
(1.00 - 4.00)
(4.00 - 8.00)
(8.00 - 12.00)
00000000.00!

FIG. 6I

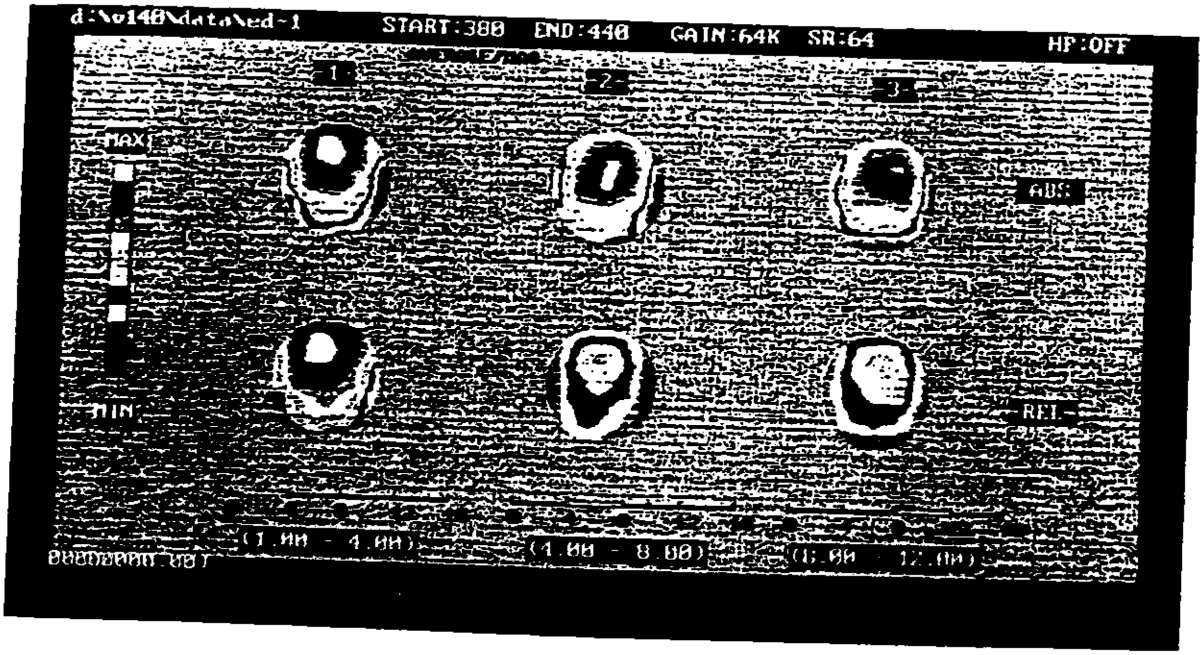
d:\o140\data\ed-1
START:380
END:440
GAIN:64K
SR:64
HP:OFF
1
2
3
MAX
MIN
ABS
REL
(1.00 - 4.00)
(4.00 - 8.00)
(6.00 12.00)
00000000.00T

FIG. 6J

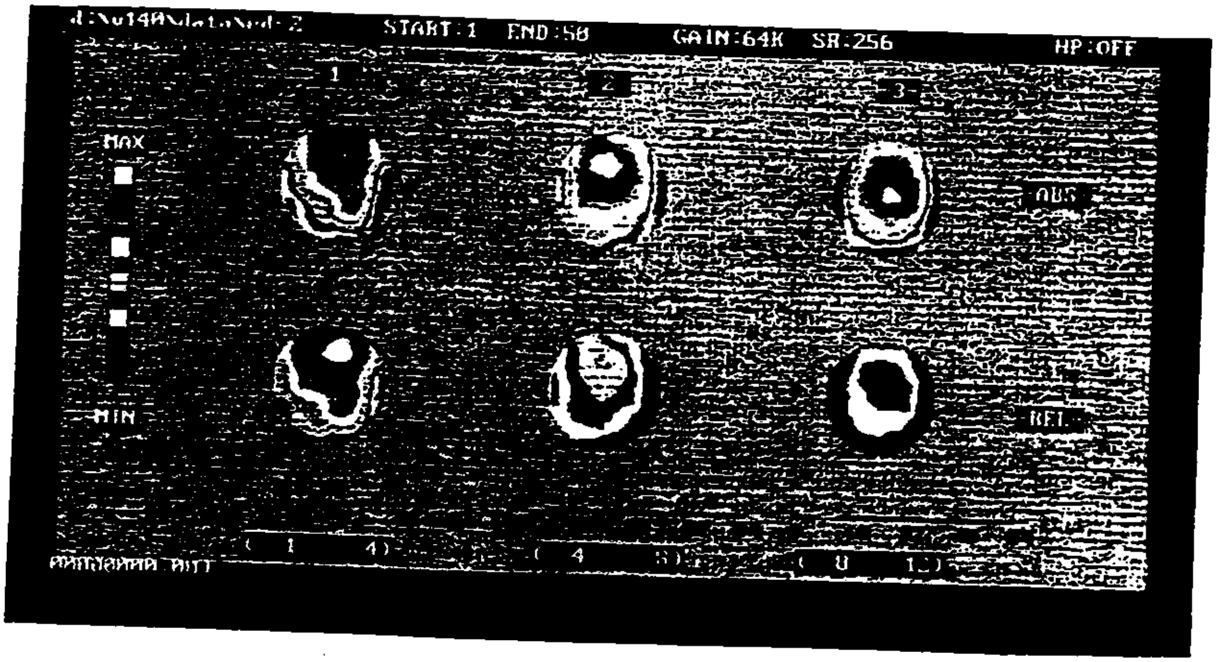
d:\o140\data\ed-2
START:1
END:58
GAIN:64K
SR:256
HP:OFF
1
2
3
MAX
MIN
ABS
REL
(1 4)
(4 8)
(8 12)

FIG. 7

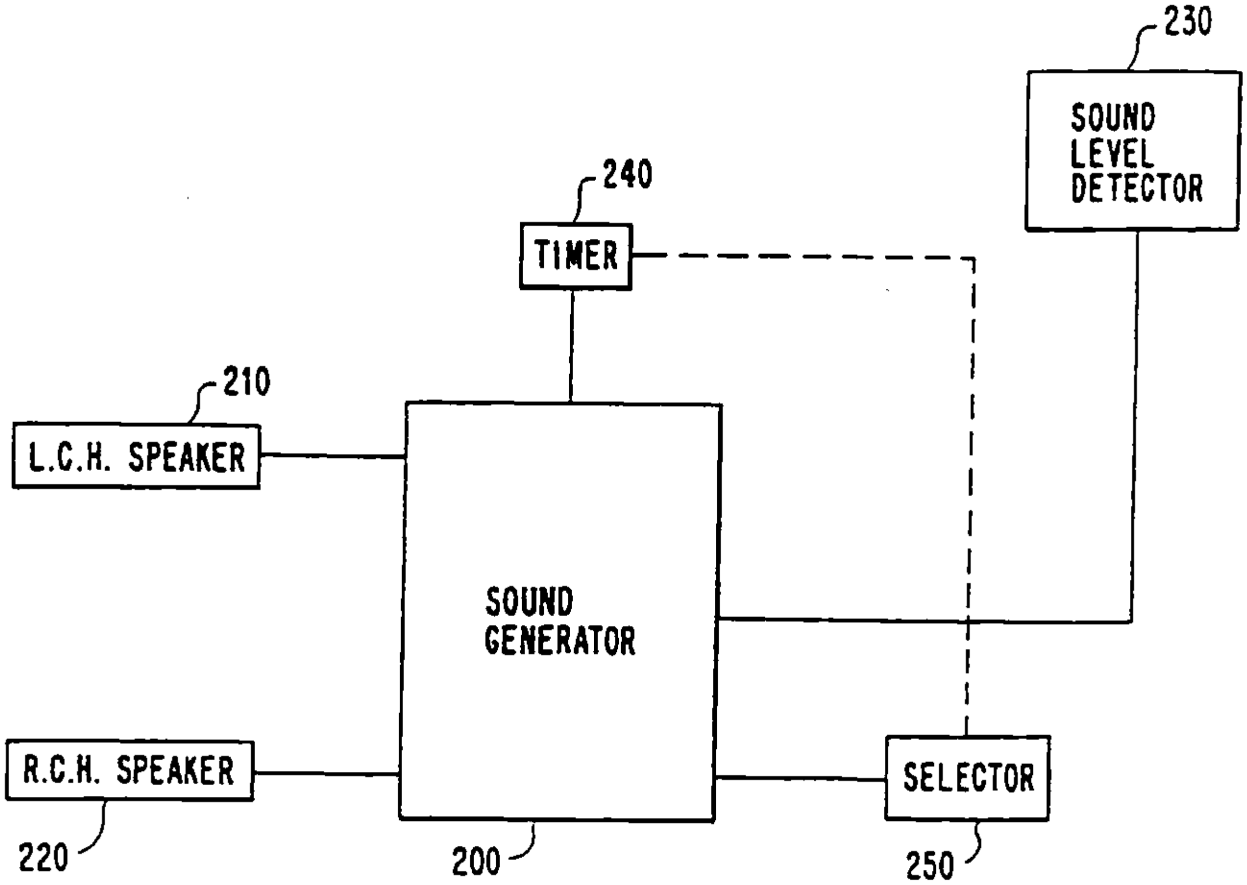
230
SOUND LEVEL DETECTOR
240
TIMER
210
L.C.H. SPEAKER
SOUND GENERATOR
R.C.H. SPEAKER
SELECTOR
220
200
250

FIG. 8A

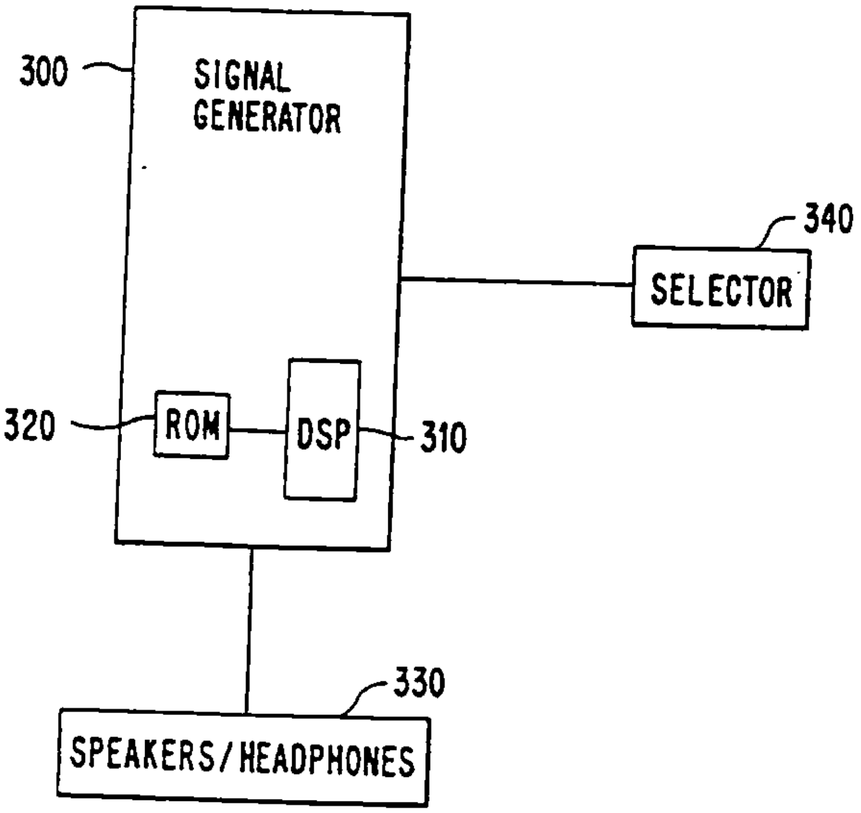
300
SIGNAL GENERATOR
340
SELECTOR
320
ROM
DSP
310
330
SPEAKERS/HEADPHONES

112

FIG. 8B

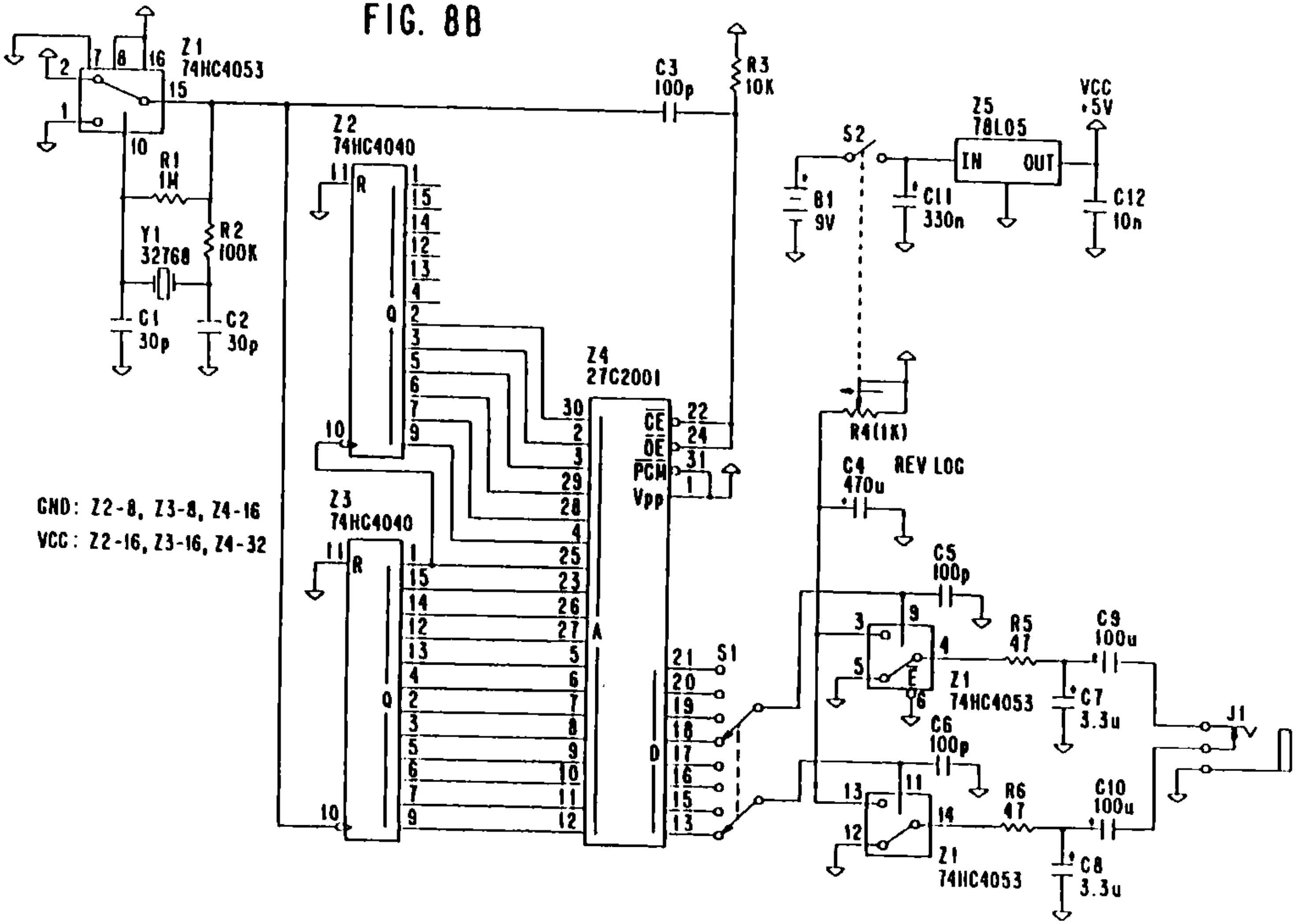
Z1
74HC4053
2
7
8
16
15
1
10
R1
1M
Y1
32768
R2
100K
C1
30p
C2
30p
Z2
74HC4040
11
R
Q
10
1
15
14
12
13
4
2
3
5
6
7
9
Z3
74HC4040
C3
100p
R3
10K
Z4
27C2001
30
2
3
29
28
4
25
23
26
27
5
6
7
8
9
10
11
12
A
D
CE
OE
PGM
Vpp
22
24
31
1
21
20
19
18
17
16
15
13
S1
S2
B1
9V
C11
330n
Z5
78L05
IN
OUT
VCC
+5V
C12
10n
R4(1K)
C4
470u
REV LOG
C5
100p
C6
100p
R5
47
R6
47
C9
100u
C10
100u
C7
3.3u
C8
3.3u
E
J1
GND: Z2-8, Z3-8, Z4-16
VCC: Z2-16, Z3-16, Z4-32

FIG. 9A

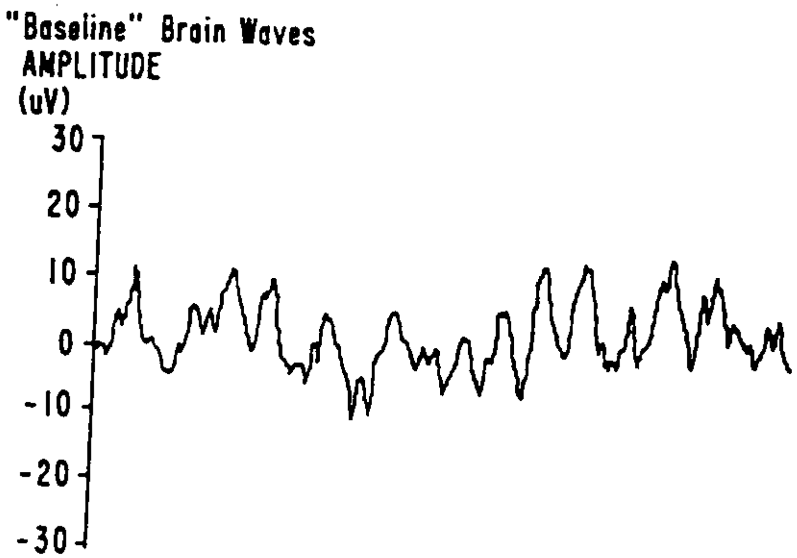

POWER (%uV^2 pp)

0-4Hz=	18.8%
4-8Hz=	9.2%
8-12Hz=	16.2%
12-30Hz=	49.2%

FIG. 9B

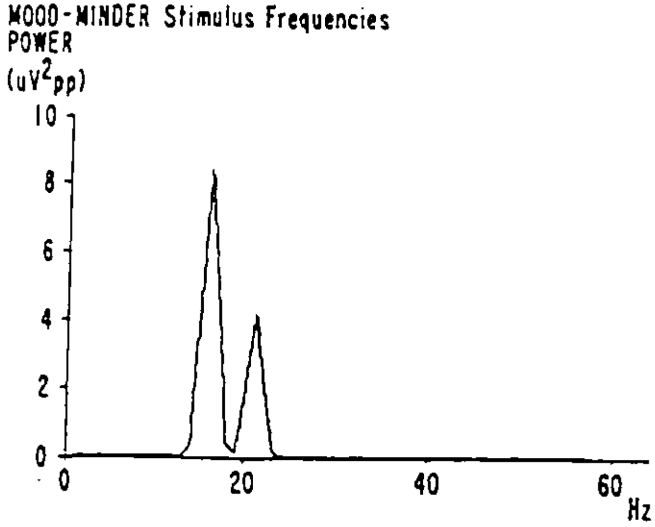

POWER(%uV^{2}pp)
16Hz= 47.6%
21Hz= 23.2%
Awake and Alert

FIG. 9C

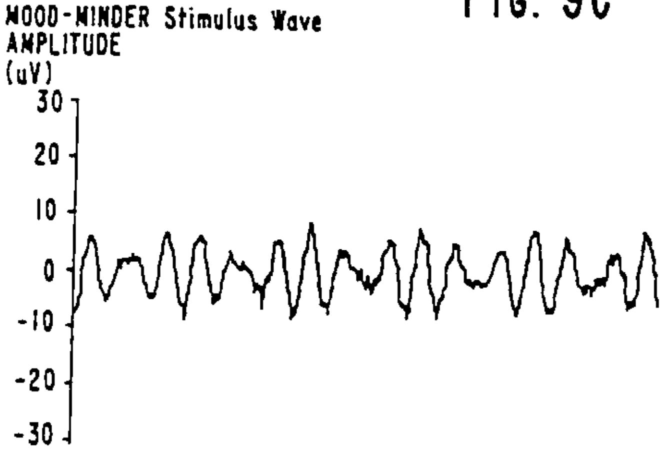

POWER(%uV^{2}pp)
16Hz= 47.6%
21Hz= 23.2%
Awake and Alert

FIG. 9D

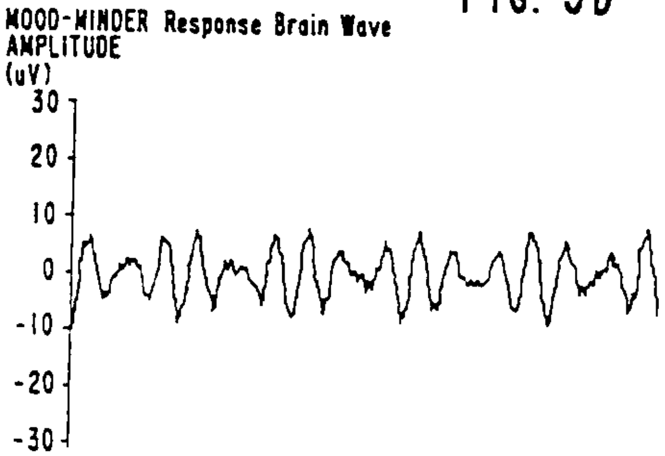

POWER(%uV^{2}pp)
16Hz= 46.5%
21Hz= 23.7%
Awake and Alert

FIG. 9E

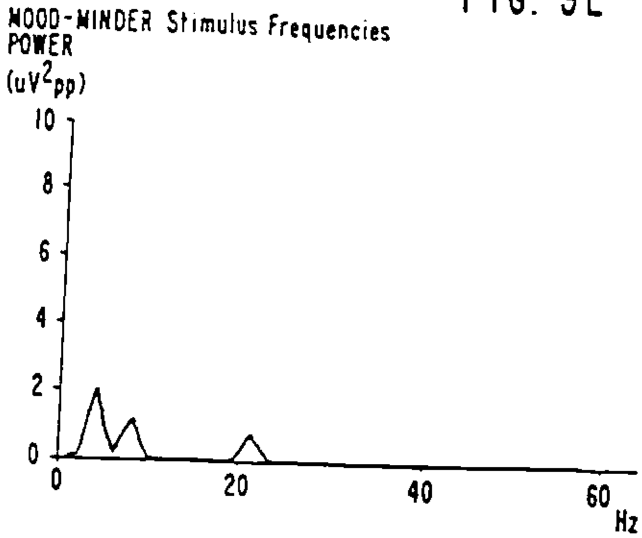

POWER(%uV²pp)
21Hz= 13.3%
8Hz= 22.1%
4Hz= 35.5%
Concentration

FIG. 9F

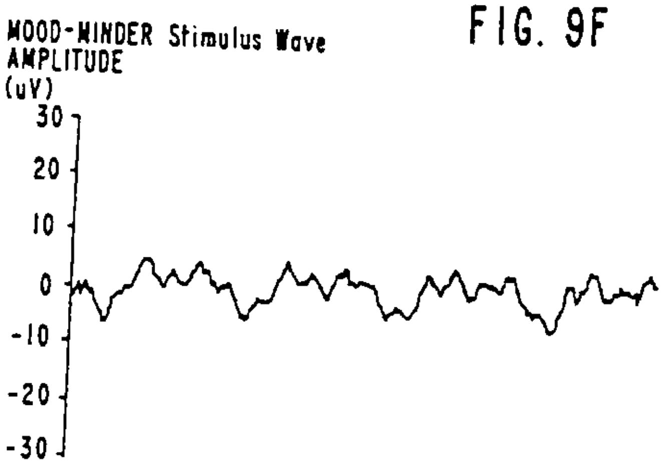

POWER(%uV²pp)
21Hz= 13.3%
8Hz= 22.1%
4Hz= 35.5%
Concentration

FIG. 9G

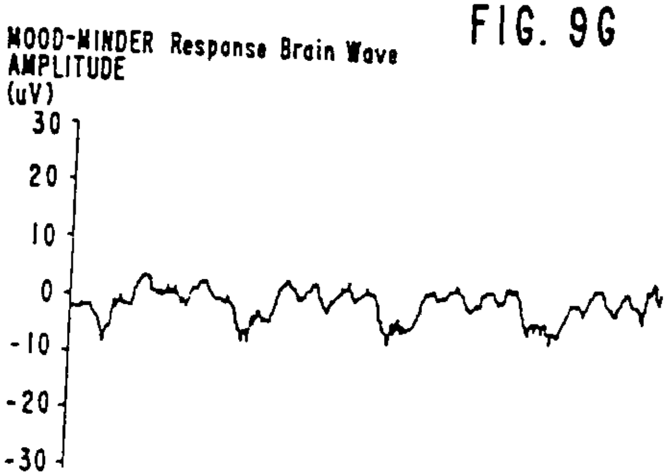

POWER(%uV²pp)
21Hz= 14.3%
8Hz= 21.5%
4Hz= 31.5%
Concentration

116

FIG. 9H

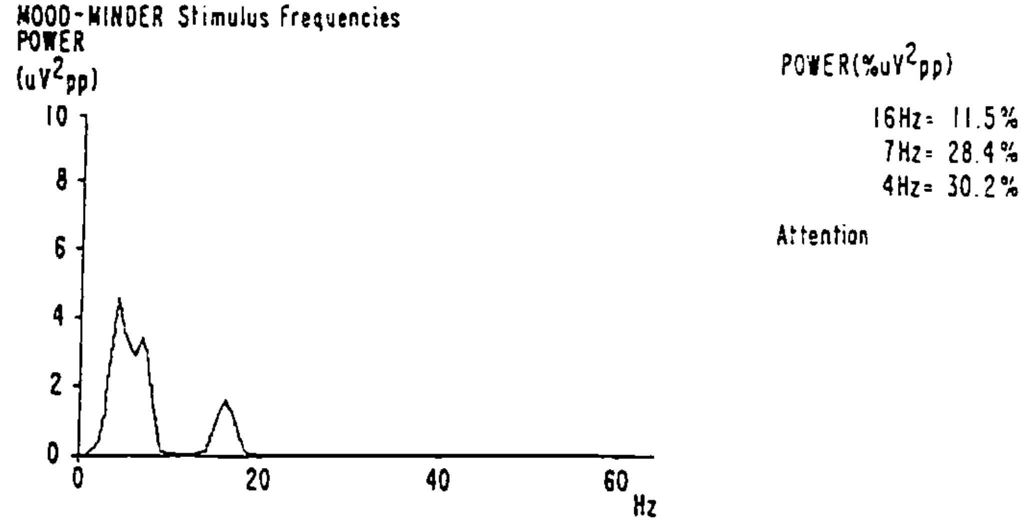

FIG. 9I

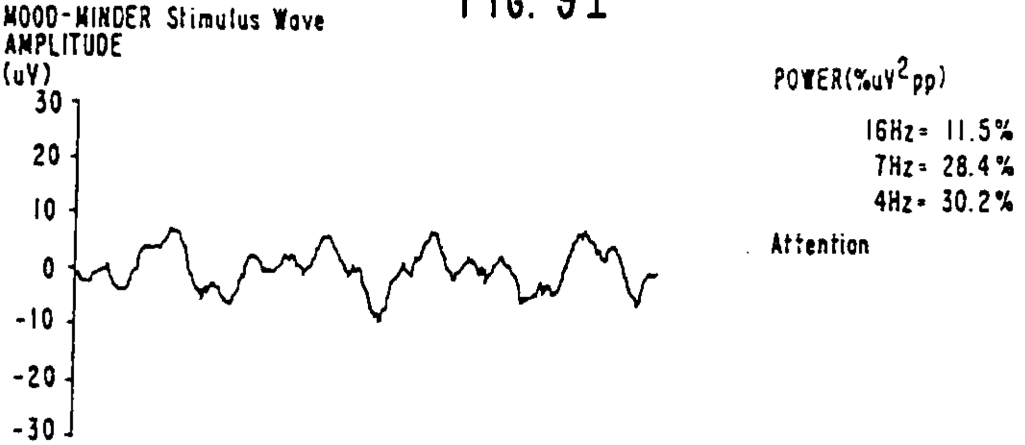

FIG. 9J

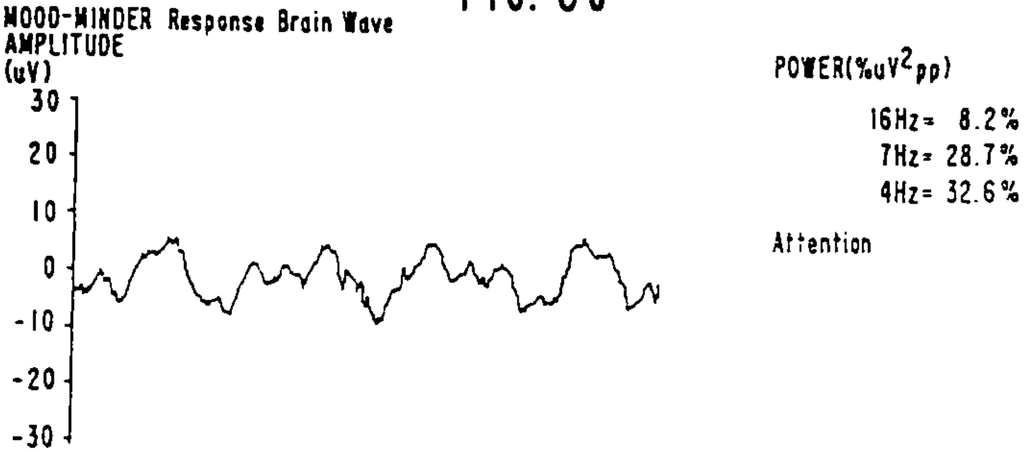

FIG. 9K

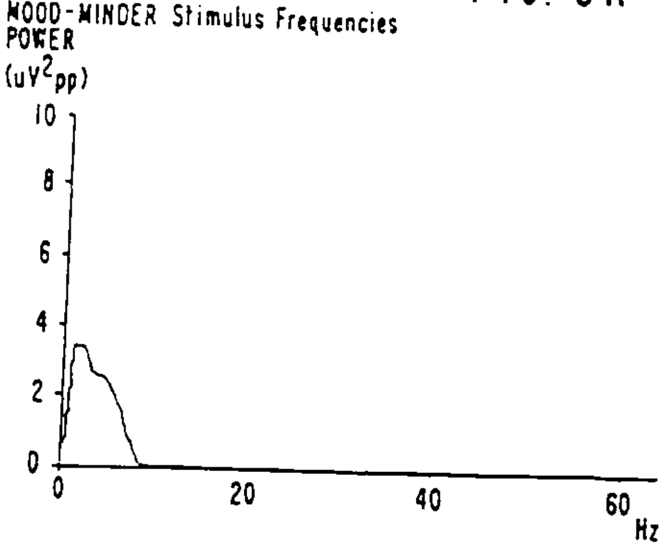

POWER(%uV^2pp)

6Hz= 9.5%
4Hz= 15.4%
1.5Hz= 19.6%

Relaxation

FIG. 9L

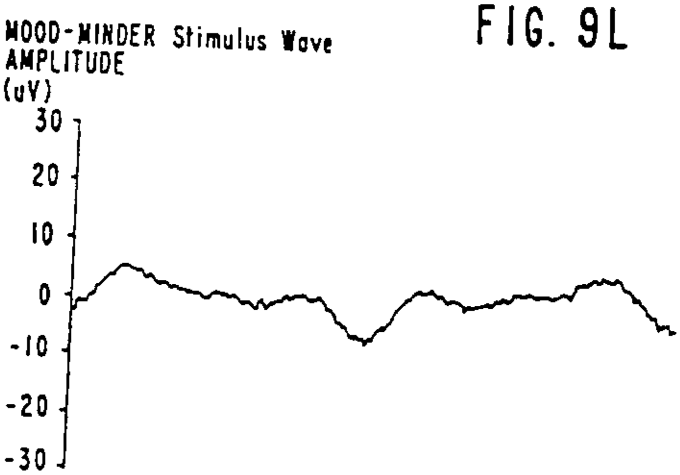

POWER(%uV^2pp)

6Hz= 9.5%
4Hz= 15.4%
1.5Hz= 19.6%

Relaxation

FIG. 9M

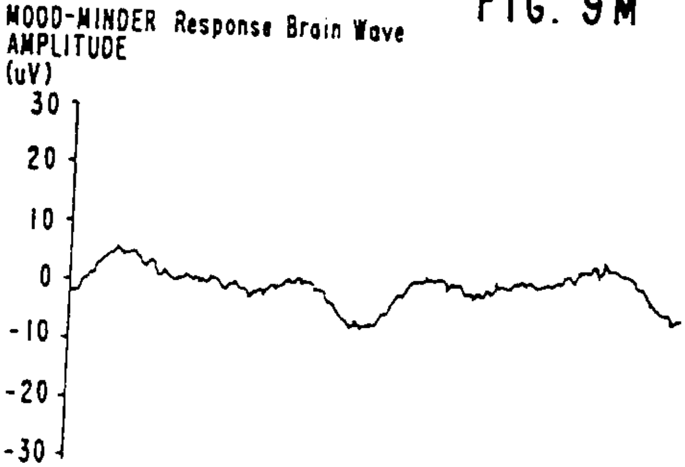

POWER(%uV^2pp)

6Hz= 7.9%
4Hz= 16.8%
1.5Hz= 20.0%

Relaxation

METHOD OF AND APPARATUS FOR INDUCING DESIRED STATES OF CONSCIOUSNESS

CROSS-REFERENCE TO RELATED APPLICATION

The present application is related to copending application No. 07/514,460, filed Apr. 16, 1990 now U.S. Pat. No. 5,213,562.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to an improved method of inducing desired states of consciousness, including different levels of sleep, in human beings, using a technique known as frequency following response (FFR), developed by the present inventor. The invention also relates to apparatus for performing the method. A number of areas of applicability of the invention are described, in accordance with different preferred embodiments.

2. Description of the Background Art

In a prior patent, U.S. Pat. No. 3,884,218, the present inventor described a method of inducing different levels of sleep, using the FFR technique, in which brain waves could be made to follow superimposed frequency patterns. These frequency patterns were provided as sine waves, at frequencies known to correspond to different levels of sleep, such as alpha (exhibiting brain wave activity in the range of 8–12 Hz), theta (6–8 Hz), and delta (1–4 Hz). EEGs exhibiting frequencies between 12 and 30 Hz (known as a beta range) are characteristic of awake individuals, though beta activity at even higher frequencies has been observed in different types of mental activities. Gamma activity has been characterized as all activity above 30 Hz; until recently, it has not been possible to monitor brain activity in the gamma range. (It should be noted that the boundaries between gamma and beta, beta and alpha, alpha and theta, and theta and delta are somewhat arbitrary; the foregoing delineations are intended to be exemplary and not limiting.)

The present inventor discovered that the human brain could be entrained to output brain wave patterns these different frequencies. While frequencies corresponding to these different levels of sleep are not audible, by superimposing those frequencies on some type of sound, such as music, it was determined to be possible to induce desired levels of sleep. The individual listening to the music would "hear" the low frequencies, with the desired effect on brain activity.

An improvement on the inventor's patented technique, to induce varied states of alertness, is the subject of copending Application No. 07/514,460, the contents of which are hereby incorporated herein by reference. This copending application describes a general FFR technique using what is known as a binaural beat phenomenon, details of which are provided in that application. Briefly, a binaural beat is produced by sending signals at different frequencies (some Hz apart, depending on the desired effect) to an individual's left and right ears. The difference between the frequencies defines the frequency of the binaural beat. Using this technique, the desired frequency can be introduced into the individual's brain activity, inducing the desired state of consciousness.

The induction of FFR in the human brain in this manner results in the synchronization of activity in the hemispheres of the brain. FIG. 1A shows brain activity without FFR, and FIG. 1B shows brain activity with FFR. The inventor has coined the term HEMI-SYNC (for Hemisphere Synchronization) to describe this phenomenon.

The copending application describes a technique wherein, in one form, sine waves having a frequency corresponding to a consciousness state are superimposed on two different carrier frequencies to form two different signals to set up the binaural beat. In another form, an actual brain pattern, based on an electroencephalogram (EEG) waveform indicative of that consciousness state is superimposed on the different carrier frequencies to form two different signals. In use, each signal is provided to one ear of a subject. The difference in carrier frequencies sets up the binaural beat.

Another, more limited application of the binaural beat phenomenon is found in U.S. Pat. No. 4,834,701. In contrast to the narrow range of frequencies discussed in that patent, in the above-mentioned copending application, the applicability of the binaural beat phenomenon is investigated over a much wider range of frequencies, spanning the spectrum of brain activity.

Through additional investigation involving mapping of brain activities of different individuals, the present inventor has discovered some significance to the fact that, while brain waves at certain frequencies are characteristic of different levels of sleep, brain patterns of different individuals still vary. The inventor has investigated possible enhancements to the FFR effect by making it more generic among individuals, yet still more specific to brain activity than a simple sine wave, or an EEG of a particular individual.

Another area of investigation being performed by the present inventor relates to human sleep patterns. Based on current knowledge of human sleep patterns, it appears that sleep is composed of a series of 90-minute cycles. As stated earlier, the beta stage is one of alertness. The first sleep state is alpha, or mental and physical relaxation. The second is theta, or light sleep. Next is delta, or deep sleep. The inventor has investigated the possibility of providing FFR waveforms in cyclic patterns, replicating these human sleep patterns, to facilitate sleep. Another possibility is to take advantage of the cyclic nature of sleep patterns to provide a more gentle wake-up for a sleeper.

In considering the need for alertness during activities such as work, the inventor also considered how it might be possible to introduce FFR waveforms into ambient noise in one's surroundings to facilitate maintenance of desired states of consciousness. Particularly in environments such as factories, or in offices where office equipment puts out consistent types of noise, it would be desirable to be able to introduce a binaural beat into that noise at different frequencies, to enhance the degree of alertness of factory or office workers as desired.

SUMMARY OF THE INVENTION

In view of the foregoing, according to one aspect of the invention, EEGs for a number of individuals in different states of consciousness are sampled, and EEG waveforms for the group of individuals, corresponding to each identifiable state of consciousness, are combined. A binaural beat then is generated using the combined EEGs.

According to this aspect of the invention, it has been determined that using groups of EEG waveforms from different individuals and combining them to obtain a representative waveform yields a waveform that a person's brain is more likely to replicate than an individual EEG waveform, or a sine wave representation of the EEG waveform. The combination may be simple averaging, though other combination techniques, such as weighted averaging, for combining different numbers of EEG waveforms as desired, are contemplated. Now that the inventor has discovered that combinations of EEG waveforms provide a particularly effective entrainment environment, it will be seen that various ways of combining these waveforms may yield greater or lesser effects.

In accordance with another aspect of the invention, a method for replicating cyclic sleep patterns for a desired sleep period is provided. In a preferred embodiment according to this aspect of the invention, a subject is led from beta, to alpha, to theta, to delta, then back to theta, then alpha, then a rapid-eye movement (REM) or light dreaming sleep, in a sequence of 90-minute cycles, during a sleep period of desired duration. After the expiration of the period, the subject may wake up voluntarily. Alternatively, the invention can provide a gentle external stimulus to lead the subject to a beta state.

With respect to this aspect of the invention, an apparatus is provided which automatically leads an individual through these cyclic sleep patterns, and enables the individual to set a desired sleep period. This device preferably takes advantage of the techniques to be described relative to the first-mentioned aspect of the invention, but is not so limited. The inventive contributions of this second aspect of the invention are considered to lie in the combination of hardware itself which generates the desired sequence of binaural beats, as opposed to the particular software which determines the nature of those binaural beats. In one form, the invention is constituted by an alarm clock which provides a fade-in theta-alpha signal followed by a strong beta-gamma signal shortly before a desired wake-up time.

According to yet another embodiment of the invention, selectable mind-affecting sound patterns are provided to supplement constant ambient noise in any environment. When the noise is not present, the patterns are not provided. The patterns vary in amplitude in accordance with changes in the environmental noise.

In accordance with still another embodiment of the invention, a portable system is provided to enable the wearer to introduce binaural beat signals of frequencies that are selectable in accordance with a desired level of awareness. Depending on the level of sophistication of the device, the binaural beat may be generated using the combined EEG waveforms of the first aspect of the invention, but this last aspect of the invention is not so limited.

BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing and other aspects of the invention will be understood by those of working skill in this technological field by reference to the following detailed description of the preferred embodiments of the invention, read in conjunction with the accompanying drawings, wherein:

FIGS. 1A-1C and 1D-1F taken from the above-mentioned copending application, show one example of the results which can be achieved using the inventive techniques;

FIG. 2 is a block diagram of the hardware according to a second embodiment of the invention, and FIGS. 3-5 are more detailed schematics therefor;

FIGS. 6A-6J are drawings, similar to FIGS. 1A and 1B, but showing brain activity during various stages of a sleep cycle, using a technique in accordance with the second embodiment of the invention;

FIG. 7 is a block diagram of hardware in accordance with a third embodiment of the invention;

FIG. 8A is a block diagram of hardware in accordance with a fourth embodiment of the invention, and FIG. 8B a schematic of that hardware; and

FIGS. 9A-9M are graphs of different possible effects of the embodiment of FIGS. 8A and 8B, showing a baseline brain pattern, selected stimulus frequencies and corresponding stimulus waves, and associated response waves.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The method according to a first preferred embodiment of the invention, which has been developed through extensive experimentation, derives from the empirically-observed phenomenon that brain patterns of human subjects are entrained more readily to brain patterns which more closely match their own. In prior implementations of the FFR technique, such as in the inventor's prior patent, in which sine waves having frequencies corresponding to desired levels of sleep were superimposed upon a given frequency, entrainment did occur. Use of the binaural beat phenomenon yielded better results, through synchronization of the hemispheres of the brain.

However, simple repetitive frequencies, or even combinations of such frequencies within different ranges, do not represent brain patterns per se, but rather provide entrainment environments for the brain to follow. It has been determined that, the more closely the entrainment environment parallels normal brain function at different levels of consciousness, the more effective the entrainment effect. This phenomenon is what led to the improvement disclosed in the above-mentioned copending application.

As a further improvement on that technique, as mentioned above, the present inventor investigated the possibility of creating more generic models of brain function at different levels of consciousness. As a result of that investigation, it was determined that combinations of EEG waveforms from different individuals functioning at the same identifiable level of consciousness (e.g. alpha sleep, theta sleep, or delta sleep) provided a superior entrainment environment. In the inventive method according to this aspect of the invention, the brain patterns of 40 to 50 individuals were combined to yield the entrainment environment.

One area of applicability of the techniques of the present invention is in the area of sleep therapy. Many individuals suffer from sleep disorders to varying degrees. It is possible to provide a suitable entrainment environment, based on known sleep cycles prevalent in humans, to help individuals to regulate their sleep patterns, and thus help to solve their sleep disorders. One embodiment of the invention, shown in FIG. 2 and also in FIGS. 3-5, implements the inventive techniques in what the inventor calls a Sleep Processor to aid in the regulation of human sleep cycles.

In FIG. 2, a read-only memory (ROM) 10 stores frequency sequences corresponding to different parts of a human sleep cycle. The stored frequency sequences may be in accordance with a predetermined algorithm, or alternatively may provide a less complex entrainment environment, such as simple averaging. A digital signal processor (DSP) 20 selects different ones of these sequences based on the current time and the time to which an alarm is set. The time is displayed on display 30, and is set using time set 40. The alarm is set to a desired wake-up time using alarm set 50.

During operation, the DSP 20 accesses the ROM 10 and provides an output to a pulse code modulator unit (PCM) 60 accordingly. The PCM 60 provides an output to each of left and right channel speakers 70, 80 which are provided in close proximity to the ears of a human subject. Using headphones enhances the effect.

Some additional detail of operation of the DSP 20 in one aspect of this embodiment now will be provided. A serial port in the DSP 20 generates an interrupt at a 50 KHz rate. An interrupt handler in the DSP 20 computes the various sounds, in one form, by generating sine waves using a pair of integrators:

$$\text{cosine} = \text{cosine} + \text{frequency} \times \text{sine}$$

$$\text{sine} = \text{sine} - \text{frequency} \times \text{cosine}$$

The Sleep Processors needs ten frequencies, five for each channel, and all of these frequencies are generated at the same time. The results are multiplied by ten envelopes, most of which are zero at any moment.

Noise is generated by a well-known 16-bit shift-register algorithm. This algorithm generates a noise signal that repeats every 65535 samples, or about every five seconds. The noise is filtered to sound more like pink or red noise, and less like white noise, and is written into a delay line in RAM. For each channel, the filtered noise is averaged with an earlier sample from the delay line, thus imparting a comb filter response to it.

An additional low-frequency sine/cosine pair is generated, to sweep the comb filter delay. 32-bit arithmetic is used here. The approximate sweep rate is about 1/8 Hz. The low-frequency sine wave is used directly to sweep the delay on one channel. The delay on the other channel is controlled by some mix of the sine and cosine waves. By choosing these and other coefficients properly, any phase and amplitude relationship between the left and right sweep can be obtained. The comb filtered noise for each channel is multiplied by a noise envelope value.

The device is operated as follows. A desired wake-up time is set, much like an alarm clock, and the desired volume is selected. A start/stop button then is pressed to start the cycles for the selected sleep period. Throughout the sleep period, the device repeats a 90 minute cycle of sound that leads the subject through alpha, theta, delta, and back to dreaming sleep. Five minutes before the scheduled wake-up time, a beta signal is introduced to bring the subject back to complete physical wakefulness. When the subject wakes up, he/she hits the start/stop button again to stop the sound sequence.

The sounds produced by the DSP 20 include binaural beat carrier sound patterns utilizing both amplitude and frequency modulation, masking pink sound (a known type of sound described in the copending application), and, optionally, occasional single-word voiced affirmations. The binaural beat audio signals may be in the form of appropriate sine waves, or alternatively may be replicas of actual EEG brain waveforms. In the latter case, either the just-described combined EEG waveforms or a single EEG waveform (as described in the copending application) may be used. The entire pattern of sound and control is generated algorithmically.

One aspect of the effectiveness of the device of FIGS. 2-5 is the spacing of sound carriers at related frequencies so as to engender binaural beat signals not only from channel to channel, but also monaurally, in each audio channel. In this preferred embodiment, three binaural beat frequency signals are created between audio carrier channels, and two amplitude beats per channel also are created, yielding a total of seven beat signals. The inventor has coined the term Septon for this set of beat signals. One example of a septon is as follows:

Left Channel		Right Channel
200 Hz carrier (4 Hz monaural beat)	(4 Hz binaural beat)	204 Hz carrier (4 Hz monaural beat)
204 Hz carrier (4 Hz monaural beat)	(4 Hz binaural beat)	208 Hz carrier (4 Hz monaural beat)
208 Hz carrier	(4 Hz binaural beat)	212 Hz carrier

A standard program according to this preferred embodiment would employ the following sound sequence:

0–5 minutes:

Signal Group A (comprised of replicated EEG waveforms having dominant values in the alpha range)

Signal Group B (15 dB below Group A, generated simultaneously with the sounds of Group A, and comprised of replicated EEG waveforms having dominant values in the theta range)

Phased Pink Sound (six seconds, peak-to-peak, on both left and right channels, 20 dB below Group A)

Voice Inserts (repeated at 40 second intervals, 10 dB below Group A, simultaneously with the other sounds, and comprising short sequences of phrases like "relax" "let go", and "sleep")

5–20 minutes:

Signal Group B

Signal Group C (20 dB below Group B, generated simultaneously with Group B, and comprised of replicated EEG waveforms having dominant values in the delta range)

Phased Pink Sound (15 dB below Group B, having a duration as in the first interval)

Voice Inserts (10 dB below Group B, comprised as above)

20–40 minutes:

Signal Group C

Signal Group D (10 dB below Group C, generated simultaneously with Group C, and comprised of replicated EEG waveforms having dominant values in the lower delta range)

Phased Pink Sound (10 dB below Group C, having a duration as in the first interval)

Voice Inserts (20 dB below Group C, comprised as above) 40–65 minutes:

Signal Group D

Phased Pink Sound (10 dB below Group D, having a duration as in the first interval)

Voice Inserts (20 dB below Group D, comprised as above) 65–80 minutes:

Signal Group C

Signal Group D (10 dB below Group C, generated simultaneously with Group C)

Phased Pink Sound (15 dB below Group C, having a duration as in the first interval)

NO voice inserts
80–90 minutes:
Signal Group B
Signal Group C (10 dB below B, generated simultaneously with Group B)
Phased Pink Sound (15 dB below Group B, having a duration as in the first interval)
NO voice inserts

The foregoing sequence is repeated through the sleep period until the wakeup sequence, approximately five minutes before the set wake-up time:

Signal Group AA (a wakeup sequence, comprising replicated EEG waveforms having dominant values in the beta range, or alternatively a 400 Hz/416 Hz envelope yielding frequencies in the beta range)
Voice inserts (10 dB below Group AA, comprised of short phrases such as "waking up", "refreshed", "bright", and repeated at intervals)

One variation of the foregoing embodiment is an alarm clock which, instead of sounding a loud alarm or other jarring noise at wake-up time, starts a gentle sequence of signals some minutes before, to bring an individual up gently through the various levels of sleep to full wakefulness. A fade-in theta-alpha signal may be provided, followed by a stronger beta-gamma signal.

FIGS. 6A to 6J show the effects of the just-described "sleep processor" embodiment. Column 1 shows distribution of delta frequencies; column 2 shows distribution of theta frequencies; and column 3 shows distribution of alpha frequencies. The top row of graphs is the actual pattern observed in the individual, and the bottom row is the baseline pattern.

FIG. 6A corresponds to a normal waking state. Dominant alpha activity is shown in the occipital area of the brain. In FIG. 6B, pink noise has been applied, without any beat frequencies. A narrower focus of waking state is shown.

In FIG. 6C, a signal sequence corresponding to Signal Group A has been applied. Some gain in theta frequencies are seen, with rapid diffusion of alpha frequencies and movement toward the vertex of the head. In FIG. 6D, a signal sequence corresponding to Signal Group B has been applied. There is further diffusion of alpha frequencies, with some movement of delta and theta activity toward the pre-frontal cortex of the brain.

In FIG. 6E, a signal sequence corresponding to Signal Group C has been applied. There is rapid diffusion of alpha frequencies, and increased power of theta and delta frequencies. In FIG. 6F, a signal sequence corresponding to Signal Group D has been applied. Alpha frequencies are diffused further toward the pre-frontal cortex, and there is a marked increase in theta and delta frequencies.

FIG. 6G, continuing application of Signal Group D frequencies, shows a marked increase in delta activity in the pre-frontal cortex, with a steady decrease in alpha activity at the vertex. In FIG. 6H, another binaural beat stimulation has been applied, and characteristics of stage 3 and 4 sleep may be observed. In FIG. 6I, further evidence of the further binaural beat stimulation is observed. Delta is the dominant frequency here. Alpha and theta activity has moved to the prefrontal cortex. Finally, FIG. 6J shows early awakening activity, with a diffusion of delta activity.

FIG. 7 is a block diagram of hardware in accordance with another embodiment of the invention, having application to the work environment, or anywhere a constant source of noise is present, to allow workers, for example, to maintain a desired state of awareness. The device may contain suitable digital signal processor circuitry, as in the preceding embodiment. One difference is that the operation of the device is keyed to the presence of ambient noise, not to a given time duration or selected sleep period.

The device of FIG. 7 includes a sound generator 200 which, as just mentioned, may comprise a digital signal processor. The generator 200 outputs sound patterns via one or more speakers (left and right channel speakers 210, 220 are shown). A sound level detector 230 detects the level of ambient noise in the room, and provides a signal to the sound generator 200, or activates a cut-off switch (not shown), to discontinue output of the sound generator 200 when the ambient noise level drops below a predetermined level.

The sound level detector also preferably provides a signal to the sound generator 200 to boost the sound pattern output when the ambient noise level increases, so that the effect of the provision of the sound pattern is commensurate with the noise level in the room. Alternatively, the user may simply adjust the volume manually, using one or more knobs (not shown) on the sound generator 200.

A timer 240 may be provided to control the duration of provision of the selected sound pattern, or even to change the sound pattern at different times of day by controlling a selector 250 which the user accesses to select a particular sound pattern to be output. The user may select a given sound pattern in the morning, and the timer 240 may change that pattern automatically, based on a need at different times of day for sound patterns providing different states of alertness.

The sound pattern produced by the device of FIG. 7 varies automatically in amplitude in accordance with changes in the ambient noise, and is discontinued when the noise stops. As a result, the sound remains unobtrusive. Depending on the setting, the produced sound pattern can enhance wakefulness, promote relaxation (as, for example, in rest areas in the workplace), reduce anxiety and stress, or focus attention, among other characteristics.

The basic system of FIG. 7 produces and inserts four different sound patterns which are selected manually so as to merge the output into the constant ambient noise. More sound patterns are possible, depending on the desired overall capabilities of the system. Various modifications are possible. For example, a programmable version may be provided, which changes the form of the sound patterns throughout a work day or night, in accordance with the responses desired.

Selectability of patterns may be accomplished differently in a model intended for use in conjunction with a computer system. The computer operator can input a selection, and may vary that selection as desired throughout the work day.

FIG. 8A is a diagram of a portable embodiment of the invention, for use in providing a desired level of consciousness on an individualized basis. A signal generator 300 preferably includes a digital signal processor 310 and a ROM 320 for storing predetermined signals or sequences of signals which correspond to various desired states of awareness. The signal generator 300 may be a simple tone generator or pair of tone generators which provide outputs to speakers or headphones 330 (such as button-sized headphones) to set up a binaural beat. Output of pink sound or phased pink sound by the generator 300 is desirable to facilitate defocusing of the listener and consequent ability to concentrate on the sounds being produced. A selector **340** enables a user to instruct the signal generator **300** to output signals corresponding to the level of consciousness (e.g. focused concentration, relaxation, alertness) that a user desires.

FIG. **8B** shows a schematic of this embodiment, which the inventor calls a "Mood Minder". This embodiment includes a selector for selecting one of four possible types of signals, corresponding to four respective levels of awareness: awake and alert; concentration; attention; and relaxation. However, the invention is not so limited, as the generator **300** may be capable of producing other possible types of signals. Alternatively, pre-set patterns in the generator **300** may vary when specialized use is required. The key to this embodiment is its portability, enabling the user to carry the device everywhere. The device is battery-operated, and is small enough to fit in an upper coat pocket, for example.

FIGS. **9A–9M** show examples of prestored patterns produced by the generator **300**, and of results achieved in use. FIG. **9A** shows baseline brain waves, with relative power output shown at the right for different frequencies. FIGS. **9B**, **9E**, **9H**, and **9K** show the stimulus frequencies produced for four different respective states of consciousness. FIGS. **9C**, **9F**, **9I**, and **9L** show the stimulus waves corresponding to the superposition of the stimulus frequencies on the baseline wave. FIGS. **9D**, **9G**, **9J**, and **9M** show the results achieved in use. As can be seen, the peak-to-peak amplitudes for the response brain waves correspond closely to those of the stimulus waves.

While the present invention has been described in detail with reference to preferred embodiments, various modifications within the scope and spirit of the invention will be apparent to those of working skill in this technological field. Consequently, the invention should be considered as limited only by the scope of the appended claims.

What is claimed is:

1. A method of inducing desired states of consciousness in human beings, comprising the following steps:
   - combining a plurality of replicated electroencephalogram (EEG) waveforms, each indicative of a particular desired state of consciousness, to produce a combined EEG waveform;
   - superimposing said combined EEG waveform on two separate sets of carrier waves using stereo sound;
   - creating differential beat frequencies between said sets of carrier waves based on said superimposing step; and
   - providing the resulting signals in audio form to respective ears of a human being, to induce said state of consciousness.

2. A method as claimed in claim 1, wherein said combining step comprises mathematically averaging said EEG waveforms to produce said combined EEG waveform.

3. A method as claimed in claim 1, further comprising the step of repeating said combining, superimposing, and creating steps for each of a set of desired states of consciousness, and producing a cycle of sets of resulting audio signals, said providing step comprising providing said cycle of sets of resulting audio signals to respective ears of a human being, to induce each of said desired states of consciousness in cyclic fashion.

4. A method as claimed in claim 3, wherein said cycle corresponds to human sleep patterns, said desired states of consciousness comprising wakefulness, alpha sleep, delta sleep, and theta sleep.

5. A method as claimed in claim 3, wherein said cycle corresponds to human sleep patterns, said desired states of consciousness comprising alpha sleep, delta sleep, and theta sleep, said cycle being approximately 90 minutes long.

6. A method as claimed in claim 5, said method further comprising the steps of providing a plurality of repetitions of said cycle, followed by providing a set of audio signals containing a binaural beat at a frequency indicative of beta consciousness.

7. A method as claimed in claim 1, wherein said creating step includes the step of combining pink sound with said sets of carrier waves by shifting of said pink sound with respect to said combined EEG waveform from one stereo audio channel to another, with cyclic changes in amplitude, frequency, and rate of panning.

8. Apparatus for facilitating sleep in a human subject, comprising:
   - means for setting a wake-up time to select a desired sleep duration;
   - means for generating a first sequence of signals in a cycle corresponding to a human sleep pattern, frequencies of said signals in said first sequence being substantially equal to frequencies of human brain patterns at different levels of sleep;
   - means for repeating said cycle a plurality of times based on the selected wake-up time; and
   - means for waking up said human subject at the selected wake-up time.

9. Apparatus as claimed in claim 8, wherein said means for waking up said human subject comprises means for generating a second sequence of signals a predetermined time before the selected wake-up time, frequencies of said signals in said second sequence being substantially equal to frequencies of human brain patterns at or near an awakened state.

10. Apparatus as claimed in claim 9, wherein said predetermined time is approximately five minutes.

11. Apparatus as claimed in claim 8, wherein said first sequence of frequencies comprises, in order, alpha frequencies, theta frequencies, delta frequencies, and theta frequencies.

12. Apparatus as claimed in claim 8, further comprising means for generating phased pink sound in conjunction with said first sequence of frequencies.

13. Apparatus as claimed in claim 8, wherein said first sequence of signals comprises a plurality of sets of combined brainwaves, each of said sets corresponding to a different level of sleep, said combined brainwaves within a given set being constituted by combined electroencephalogram (EEG) waveforms of a plurality of individuals, taken when said individuals had attained a different respective level of sleep.

14. Apparatus as claimed in claim 13, wherein said EEG waveforms are mathematically averaged.

15. Apparatus for awakening an individual using brain pattern entrainment, said apparatus comprising:
   - means for selecting a wake-up time;
   - means for keeping time; and
   - means, operative a predetermined period before said wake-up time as determined by said means for keeping time, for producing a first sequence of signals having frequencies in the theta-alpha range, followed by a second sequence of signals having frequencies in the beta-gamma range.

16. Apparatus as claimed in claim 15, wherein said means for producing said first and second sequences of signals comprises means for producing said second sequence of signals at a higher amplitude than said first sequence of signals.

17. Apparatus as claimed in claim 15, wherein said first sequence of signals comprises a plurality of sets of combined brainwaves, each of said sets corresponding to a different level of consciousness, said combined brainwaves within a given set being constituted by combined electroencephalogram (EEG) waveforms of a plurality of individuals, taken when said individuals had attained a different respective level of consciousness.

18. Apparatus as claimed in claim 16, wherein said EEG waveforms are mathematically averaged.

19. Apparatus for inducing a desired state of consciousness, said apparatus comprising:

means for detecting presence of a predetermined level of ambient noise;

means, responsive to said detecting means, for generating signals having frequencies substantially equal to frequencies of human brain patterns when said ambient noise is present; and

means for selecting said signals in accordance with desired human activity in said areas.

20. Apparatus as claimed in claim 19, further comprising timer means, connected to said generating means, for generating said signals for a predetermined time set by said timer means.

21. Apparatus as claimed in claim 19, wherein said timer means is connected to said selecting means to enable selection of different ones of said signals in accordance with desired human activity at different times of day.

22. Apparatus as claimed in claim 19, wherein said generating means comprises means, responsive to said detecting means, for increasing an amplitude of said signals in response to an increase in amplitude of said ambient noise, and for decreasing an amplitude of said

signals in response to a decrease in amplitude of said ambient noise.

23. Apparatus as claimed in claim 22, wherein said generating means further comprises means for discontinuing said signals when said ambient noise falls below said predetermined level.

24. Apparatus as claimed in claim 19, wherein said generating means comprises a digital signal processor and a read-only memory (ROM) connected to said digital signal processor, said ROM storing a plurality of sets of signals, each of said sets of signals having frequencies substantially equal to human brain patterns at a desired state of consciousness.

25. Apparatus as claimed in claim 24, wherein each of said sets of signals comprises a plurality of sets of combined brainwaves, each of said sets corresponding to a different level of consciousness, said combined brainwaves within a given set being constituted by combined electroencephalogram (EEG) waveforms of a plurality of individuals, taken when said individuals had attained a different respective state of consciousness.

26. Apparatus as claimed in claim 25, wherein said EEG waveforms are mathematically averaged.

27. Apparatus for awakening an individual using brain pattern entrainment, said apparatus comprising:

means for selecting a wake-up time; and

means, operative a predetermined period before said wake-up time, for producing a first sequence of signals having frequencies in a first predetermined range corresponding to a first state of consciousness, followed by a second sequence of signals having frequencies in a second predetermined range corresponding to a second state of consciousness.

28. Apparatus as claimed in claim 27, wherein said first predetermined range is the theta-alpha range, and said second predetermined range is the beta-gamma range.

* * * * *

MIND CONTROL
TECHNOLOGY
PATENTS
II

Technology To Boggle Your Mind…

Part II in a series of research projects focusing on the official US Patent Office patents detailing technologies for mental monitoring, mind alteration and inducing states of consciousness.

This research project is in progress.
IDN is currently obtaining USPTO documents.
Please find abstract of technologies being researched.

TECHNOLOGY TO BOGGLE YOUR MIND...

This is a listing of various technologies available which can be combined for use in direct or subliminal mind-control systems. Please note that these are mainly private inventions intended for positive uses, but could be applied for negative purposes as well.

- Silent Subliminal Presentation System, US Patent #5,159,703, Oliver Lowery, October 27, 1992. A silent communications system in which non-aural carriers in the very low or very high audio-frequency range, or in the adjacent ultrasonic frequency spectrum, are amplitude-modulated with the desired intelligence and propagated acoustically or vibrationally for inducement into the brain.
- Hearing System, US Patent #4,877,027, Wayne Brunkan, October 31, 1989. A method for directly inducing sound into the head of a person, using microwaves in the range of 100 MHz to 10,000 MHz, modulated with a waveform of frequency-modulated bursts.
- Psycho-Acoustic Projector, US Patent #3,568,347, Andrew Flanders, February 23, 1971. A system for producing aural psychological disturbances and partial deafness in the enemy during combat situations.
- Noise Generator and Transmitter, US Patent #4,034,741, Guy Adams and Jess Carden, Jr, July 12, 1977. An analgesic noise-generator.
- Method and System for Altering Consciousness, US Patent #5,123,899, James Gall, June 23, 1992. A system for altering the states of human consciousness involving the use of simultaneous application of multiple stimuli, preferably sounds, having differing frequencies.
- Subliminal Message Generator, US Patent #5,270,800, Robert Sweet, December 14, 1993. A combined subliminal and supraliminal message generator for use with a television receiver; permits complete control of subliminal messages and their presentation. Also applicable to cable television and computers.
- Superimposing Method and Apparatus Useful for Subliminal Messages, US Patent #5,134,484, Joseph Wilson, July 28, 1992.

US Patent #4,717,343, Alan Densky, January 5, 1988. A method of conditioning a person's unconscious mind in order to effect desired change in the person's behaviour, and which does not require the services of a trained therapist.

- Auditory Subliminal Message System and Method, US Patent #4,395,600, Rene Lundy and David Tyler, July 26, 1983. An amplitude-controlled subliminal message may be mixed with background music.
- Auditory Subliminal Programming System, US Patent #4,777,529, Richard Schultz and Raymond Dolejs, October 11, 1988.
- Apparatus for Inducing Frequency Reduction in Brain Wave, US Patent #4,834,701, Kazumi Masaki, May 30, 1989.
- Ultrasonic Speech Translator and Communication System, US Patent #5,539,705, M. A. Akerman, Curtis Ayers, Howard Haynes, July 23, 1996. A wireless communication system, undetectable by radio-frequency methods, for converting audio signals, including human voice, to electronic signals in the ultrasonic frequency range, transmitting the ultrasonic signal by way of acoustic pressure waves across a carrier medium, including gases, liquids and solids, and reconverting the ultrasonic acoustic pressure waves back to the original audio signal. This invention was made with government support under Contract DE-ACO5-840R21400, awarded by the US Department of Energy to Martin Marietta Energy Systems, Inc.
- Non-Audible Speech Generation Method and Apparatus, US Patent #4,821,326, Norman MacLeod, April 11, 1989.
- Apparatus for Electrophysiological Stimulation, US Patent #4,227,516, Bruce Meland and Bernard Gindes, October 14, 1980.
- Method and Recording for Producing Sounds and Messages to Achieve Alpha and Theta Brainwave States and Positive Emotional States in Humans, US Patent #5,352,181, Mark Davis, October 4, 1994.
- Method and Apparatus for Translating the EEG into Music to Induce and Control Various Psychological and Physiological States and to Control a Musical Instrument, US Patent #4,883,067, Knispel et. al., November 28, 1989.
- Method of and Apparatus for Inducing Desired States of Consciousness, US Patent #5,356,368, Robert Monroe, October 18, 1994. Improved methods and apparatus for entraining human brain patterns, employing frequency-following-response (FFR) techniques and facilitating attainment
- Method of Inducing Mental, Emotional and Physical States of Consciousness, including Specific Mental Activity, in Human Beings, US Patent #5,213,562, Robert Monroe, May 25, 1993.
- Device for the Induction of Specific Brain Wave Patterns, US Patent #4,335,710, John Williamson, June 22, 1982. Brainwave patterns associated with relaxed and meditative states in a subject are gradually induced without deleterious chemical or neurologic side effects.
- Method and Apparatus for Repetitively Producing a Noise-like Audible Signal, US Patent #4,191,175, William Nagle, March 4, 1980.
- Apparatus for the Treatment of Neuropsychic and Somatic Diseases with Heat, Light, Sound and VHF Electromagnetic Radiation, US Patent #3,773,049, L. Y. Rabichev, V. F. Vasiliev, A. S. Putilin, T. G. Ilina, P. V. Raku and L. P. Kernitsky, November 20, 1973. Don't let the nice title fool you. This is the patent for *LIDA*, the infamous Soviet brainwashing machine.
- Non-Invasive Method and Apparatus for Modulating Brain Signals through an External Magnetic or Electric Field to Reduce Pain, US Patent #4,889,526, Elizabeth Rauscher and William Van Bise, December 26, 1989.
- Nervous System Excitation Device, US Patent #3,393,279, Gillis Patrick Flanagan, July 16, 1968. A method of transmitting audio information via a radiofrequency signal modulated with the audio info through electrodes placed on the subject's skin, causing the sensation of hearing the audio information in the brain.
- Method and System for Simplifying Speech Waveforms, US Patent #3,647,970, G. Patrick Flanagan, March 7, 1972. A complex speech waveform is simplified so that it can be transmitted directly through earth or water as a waveform and understood directly or after amplification.
- Means for Aiding Hearing, US Patent #2,995,633, Henry Puharich and Joseph Lawrence, August 8, 1961. Means for converting audible signals to electrical signals and conveying them to viable nerves of the facial system.
- Means for Aiding Hearing by Electrical Stimulation of the Facial Nerve System, US Patent #3,170,993, Henry Puharich, February 23, 1965.
- Hearing Device, US Patent #4,858,612, Philip Stocklin, August 22, 1989. A method and apparatus for simulation of hearing in mammals by introduction of a plurality of microwaves into the regions of

RF Energy Harvesting for INTERNET OF THINGS

K.J. Vinoy

ECE Dept
Indian Institute of Science
Bangalore 560012

The world
is the index
Take the world
on line
Take control
of the world
Let the things
talk to each
other
Let thing
become
intelligen

ECE Dept, Indian Institute of Science

- **Vision Statement:** Excellence in Theoretical and Experimental Research in Communications, Signal Processing, Microelectronics and RF/Photonics.
- **Faculty: 24;** Fellows of IEEE: 4; Fellows of INAE: 8
- Active in Publications: Books, Book Chapters, Journal & Conference Papers; Patents; Standardization etc
- Collaborative research

People
Masters Students
[ME, MSc→ Mtech, Mtech(Res)]
PhD Students
Project Staff

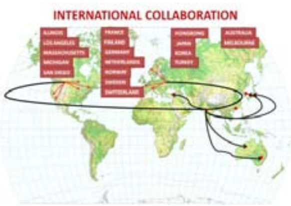
INTERNATIONAL COLLABORATION
ILLINOIS
LOS ANGELES
MASSACHUSETTS
MICHIGAN
SAN DIEGO
FRANCE
FINLAND
GERMANY
NETHERLANDS
NORWAY
SWEDEN
SWITZERLAND
HONGKONG
JAPAN
KOREA
TURKEY
AUSTRALIA
MELBOURNE

NATIONAL COLLABORATION
Major collaborative initiatives

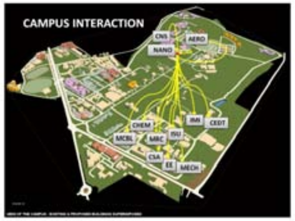
CAMPUS INTERACTION
CNS
AERO
NANO
CHEM
MCBL
MBC
ISU
IMI
CEDT
CSA
EE
MECH

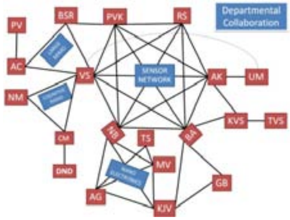
Departmental Collaboration
PV
BSR
PVK
RS
AC
VS
SENSOR NETWORK
AK
UM
NM
KVS
TVS
CM
NB
TS
BA
DND
MV
GB
AG
KJV

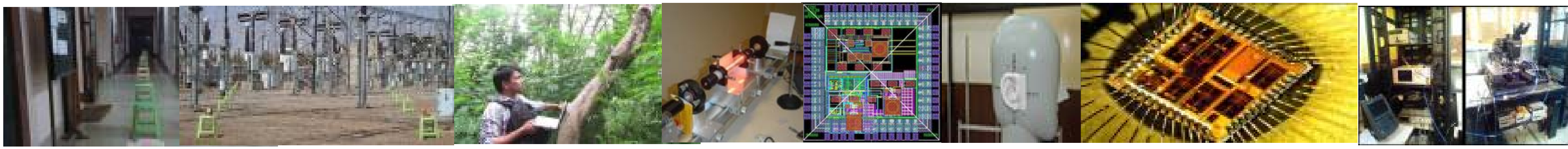

ECE: Microwave Engineering

K.J. Vinoy

- Low-Actuation Voltage Capacitive RF MEMS Switch (<10V)
  - Low-complexity fabrication process to enhance process yield
  - High reliability: no failure even up to 10 million cycles of operation tested
- Meso-scale Electrostatic Phase Shifter on microwave Laminate (MEPL)
  - Utilizes modern printed circuit board fabrication technology.
  - X-band monolithic antenna array system on the microwave laminate board demonstrated.

- Wideband group delay engineering in RF circuits for radar, medical imaging, and spectrum sensing.
  - Demonstration uses two stage All – Pass Networks; can be extended over multiple stages to obtain a higher bandwidth and/or higher group delay slope.
- RF energy harvesting circuits
  - Integrated with RF transmitters and sensors for practical IoT nodes
  - High efficiency RF-DC converter which can operate at input power of -20dBm (10µW) at 2.4GHz using UMC 130nm process MOSFETs.
- FEM based algorithms for Electromagnetic circuits & components (periodic structures such as metamaterials)
  - Fast computation of electromagnetic propagation characteristics
  - Especially suited for evaluation of processes uncertainties

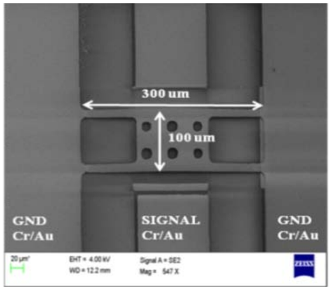

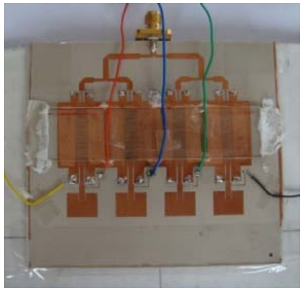

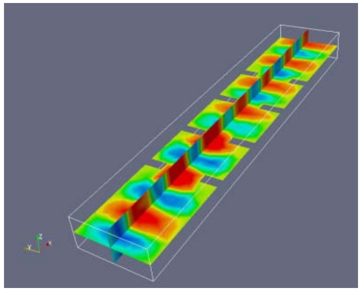

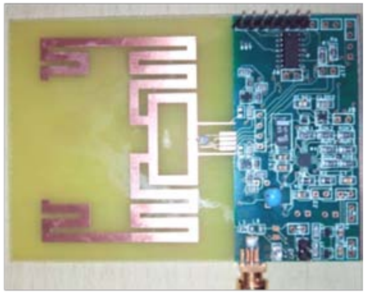

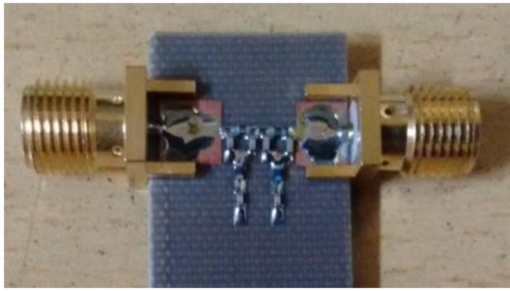

Setting the stage….

- **Introduction**
  - **Wireless Power Transfer**
    - **Energy Harvesting**
  - **Internet of Things**
- Highlights of Recent Development (Hardware)
  - Powering wireless terminals
  - RFID with integrated sensors

Ongoing Research Challenges

Wireless Power Transfer (WPT)

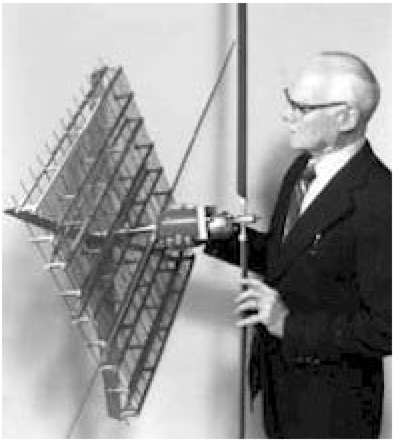

- Indicates transfer of electric energy remotely

- WPT has a long history!!
  - Tesla demonstrated it in 1899 by wirelessly powering fluorescent lamps 40 kms away from the power source.
  - Had multiple patents in early 1900s.

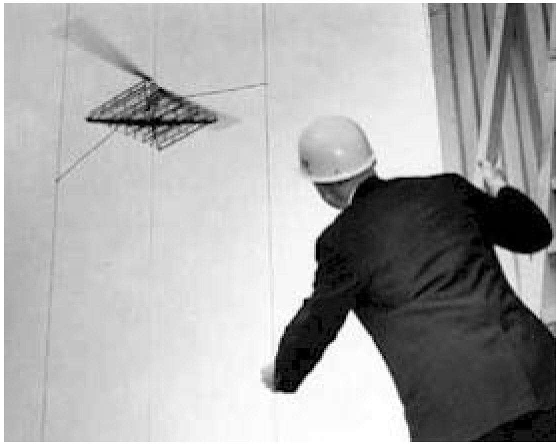

- In 1960s W.C. Brown coined the term Rectenna, which he used to directly converts incoming microwaves to DC.
  - He demonstrated its ability to power a helicopter solely through microwaves for 10 hours continuously.

- These demonstrations involved dedicated sources with **large power to transmit over long distances**.

http://mainland.cctt.org/istf2008/brown.asp

SSPS

- Space Solar Power Satellites

- WPT is widely investigated for putting **solar power generating satellites** into space and transmitting power to Earth stations. (Mainly in Japan)

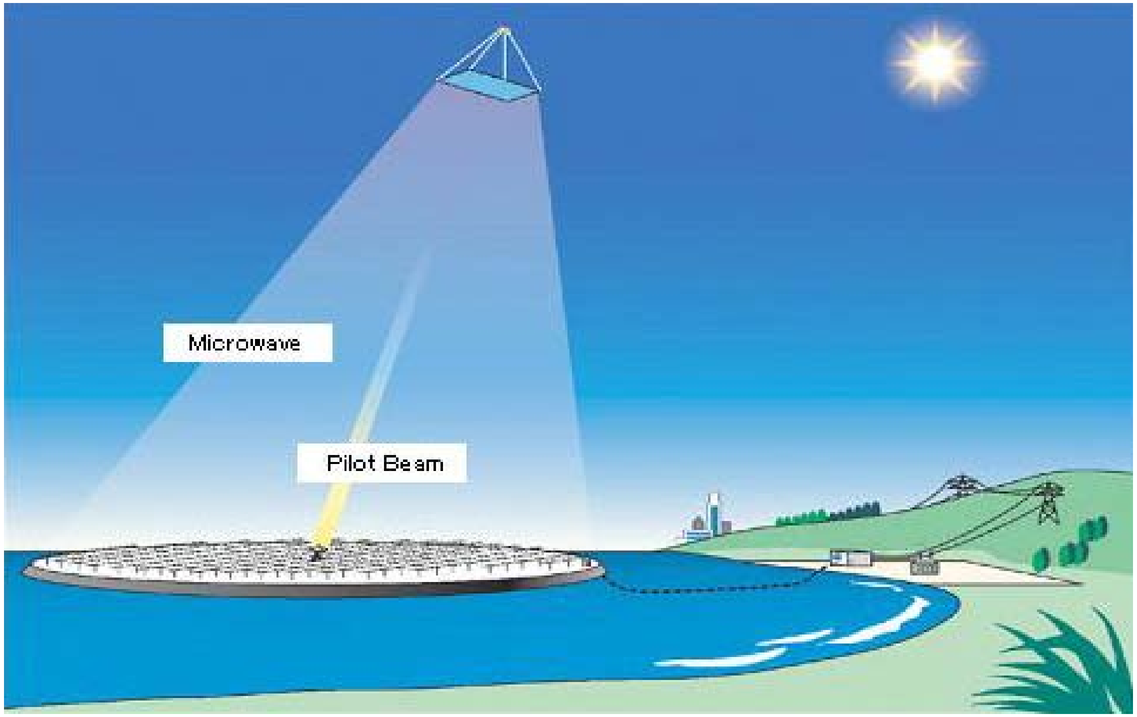

http://www.jspacesystems.or.jp/en_project_ssps/

Near field Wireless Power Transfer

- Recent demonstration by MIT to transfer high RF power (Watts) transferred across meters.
- Resonant coils are used
- Typically at 100 kHz to 10's of MHz
- Many new applications emerged

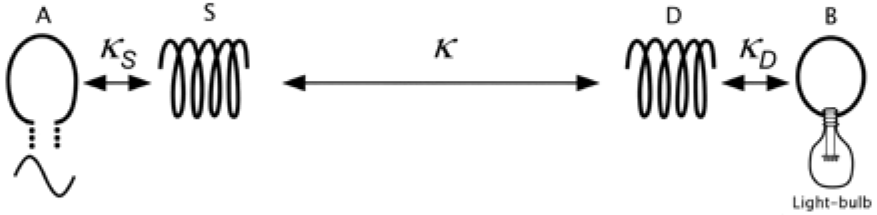

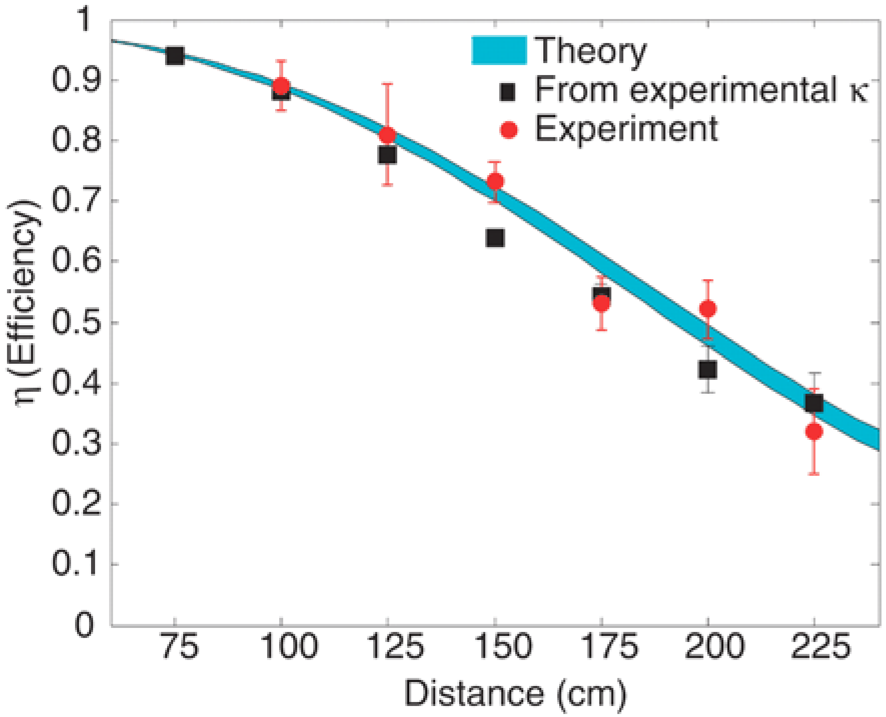

André Kurs et al, Wireless Power Transfer via Strongly Coupled Magnetic Resonances , Science, Vol. 317 no. 5834 pp. 83-86. 6 July 2007.

System Schematic Qi

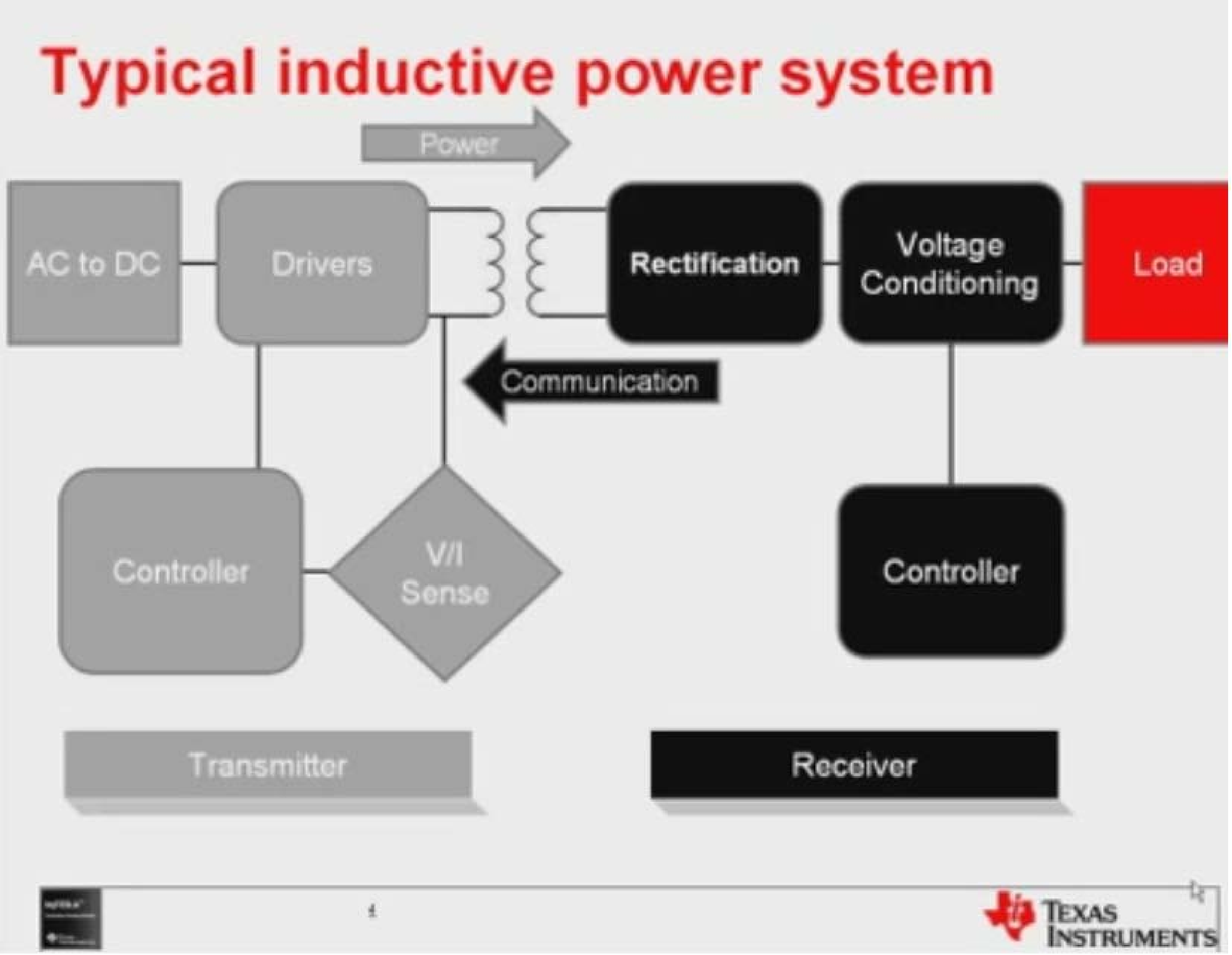

Typical inductive power system

Power
AC to DC
Drivers
Rectification
Voltage Conditioning
Load
Communication
Controller
V/I Sense
Controller
Transmitter
Receiver
Texas Instruments

Source: Texas Instruments Qi Development kit

Far → Near in WPT

- Free space loss factor is a major bottleneck for power transfer at large distances

- Short distance/ Near field options
  - MIT demonstration (2007)
  - Qi Standard
  - Phone charging solutions
  - Vehicles running on wireless power

- Two extremes in WPT
  - mW ← → MW
  - mm ← → 1000s km
  - 100kHz ← → 2.4/5.8GHz
  - 10cm x 10cm ← → km x km
  - Commercial vs bluesky

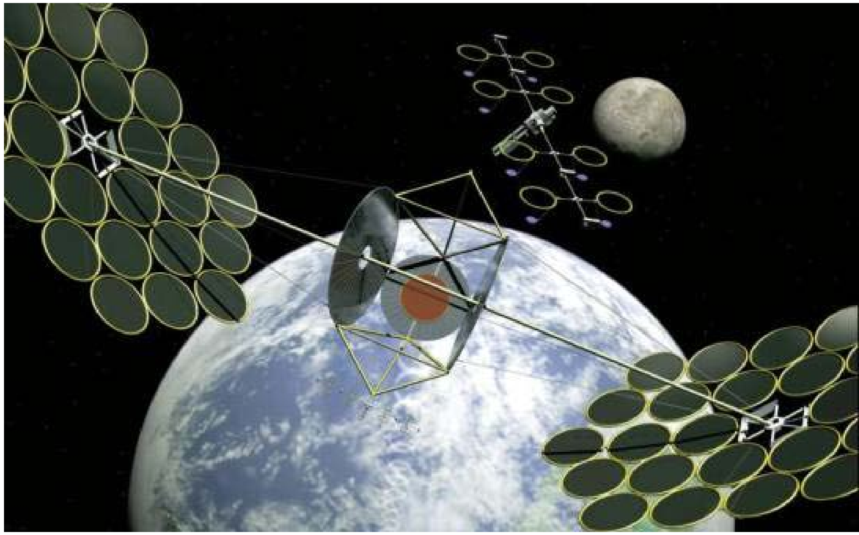

Far Field Transfer of RF Energy

- Focus of this talk
- Applications: RFID tags, Wireless Sensor Network nodes, biomedical equipment, home automation and structural monitoring can benefit from RF energy harvesting.
- Block diagram and a design example:

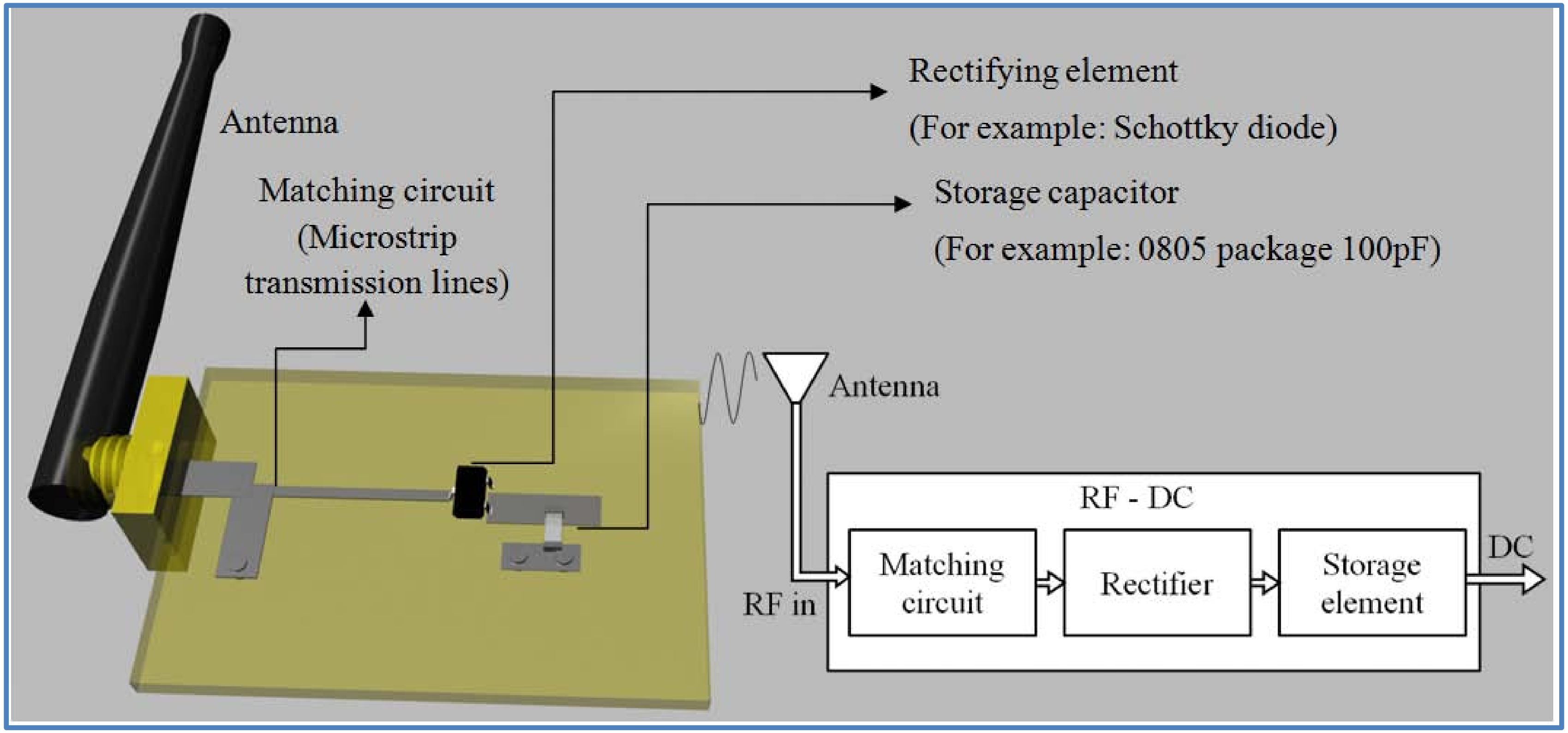
Antenna
Matching circuit
(Microstrip
transmission lines)
Rectifying element
(For example: Schottky diode)
Storage capacitor
(For example: 0805 package 100pF)
Antenna
RF in
RF - DC
Matching
circuit
Rectifier
Storage
element
DC

A New Paradigm: Internet of Things (IoT)

- IoT refers to uniquely identifiable objects and their virtual representations in an Internet-like structure.

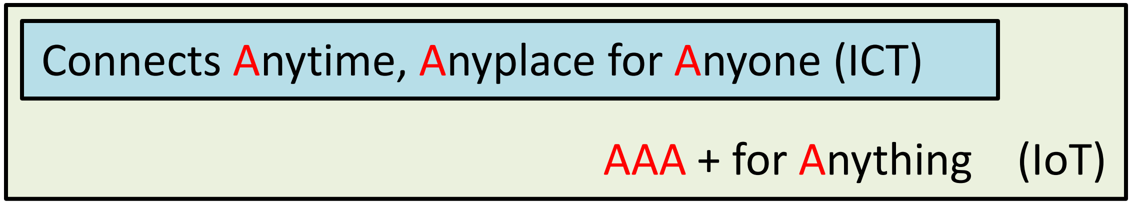

Connects Anytime, Anyplace for Anyone (ICT)

AAA + for Anything (IoT)

- IoT is a scheme for connecting things: sensors, actuators, and other smart technologies, thus enabling person-to-object and object-to-object communications.

- **Continuous availability of power is crucial for their deployment**

IoT Applications

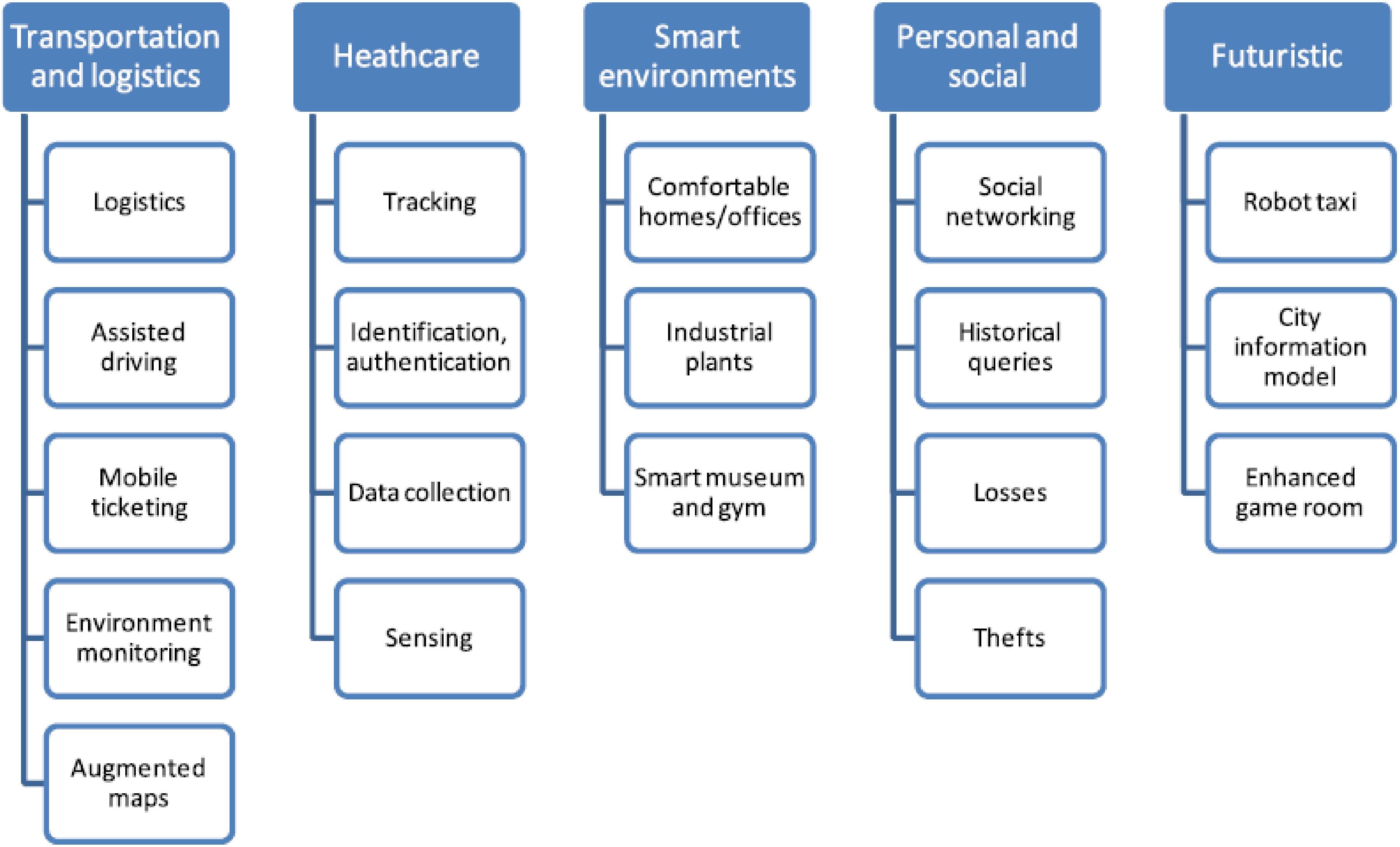

- Transportation and logistics
  - Logistics
  - Assisted driving
  - Mobile ticketing
  - Environment monitoring
  - Augmented maps
- Heathcare
  - Tracking
  - Identification, authentication
  - Data collection
  - Sensing
- Smart environments
  - Comfortable homes/offices
  - Industrial plants
  - Smart museum and gym
- Personal and social
  - Social networking
  - Historical queries
  - Losses
  - Thefts
- Futuristic
  - Robot taxi
  - City information model
  - Enhanced game room

L. Atzori , A. Iera, G. Morabito, The Internet of Things: A survey Computer Networks 54 (2010) 2787–2805

Comparison of different wireless protocols

Today, a lot can be done at low power!!

Characteristics of key 2.4GHz ISM Band Radios studied :

	BLE	ANT	Zigbee	WLAN
Topologies	P2P , Star	P2P , Star, tree, mesh	P2P , Star, mesh	P2P , Star
Modulation	GFSK	GFSK	OQPSK	DSSS (802.11b)
Max data rate	1Mbps	12.8-60 Kbps	250Kbs (@2.4Ghz)	1- 11Mbps (802.11b)
Throughput	305 kbps	20Kbps	100Kbps	6Mbps (802.11b)
Range (in m)	10-100(0-10dBm)	30 (@ 0dBm)	10-100 (0-20dBm)	100+(20dBm)
Max nodes in piconet	7	65533	Star-65536	32-64
Battery life	1-2 years (coin cell)	1-2years (coin cell)	100-1000 days	0.5-5days

Key Aspects :

BLE is robust and has lowest power consumption but cannot natively form mesh networks,

Zigbee can support large mesh networks, power consumption is higher than BLE and throughput is lower: It is suitable for low data rate, low power, large size networks.

WLAN is primarily suitable for transferring bulk data at high speeds, Not suitable for low power applications.

Power Requirements in Common WSN

	Crossbow MICAz	Intel IMote2	Jennic JN5139
Radio standard	IEEE802.15.4/ZigBee	IEEE802.15.4	IEEE802.15.4/ZigBee
Typical range	100m (outdoor), 30m (indoor)	30m	1 km
Data rate (kbps)	250 kbps	250 kbps	250 kbps
Sleep mode (deep sleep)	15 μA	390 μA	2.8 μA (1.6μA)
Processor only	8mA active mode	31–53mA$*$	2.7+0.325mA/MHz
RX	19.7mA	44mA	34mA
TX	17.4mA (+0dbm)	44mA	34mA (+3 dBm)
Supply voltage (minimum)	2.7V	3.2V	2.7V
Average	2.8mW	12mW	3mW

JM. Gilbert$*$ F. Balouchi, Comparison of Energy Harvesting Systems for Wireless Sensor Networks, International Journal of Automation and Computing 05(4), October 2008, 334-347

	Jennic JN5148	**TI- CC430**	**BLE**	**Zarlink ZL70250**
Active mode current at 16MHz [mA]	6	4	6.7	3.2
Deep sleep current [nA]	100	1000	400	20
Transmission current [mA]@Tx-power [dBm]	15@2.5	18@0	36@2	2@-10
Transmit frequency	2.4 GHz	2.4 GHz	2.4 GHz	868 MHz
Wakeup time [ms]	1	3	0.12	0.16
Energy consumption for a transmission cycle of 2ms [µJ]	183	300	196	32
Power supply voltage [V]	2.2 – 3.6	1.8 -3.6	2-3.6	1.2 – 1.8

In perspective

- Energy requirements in different devices/systems
- 6 orders of magnitude variation!!!
- Energy requirements in WSN
  - Depends on the complexity/ standard/ range
  - eg 90 μW to power a pulse oxymeter sensor, to process data and to transmit them at intervals of 15 s

Device type	**Power consumption**
Smartphone	1W
MP3 decoder chip	58 mW
Hearing aid	1 mW
Wireless sensor node	100 μW*
RF receiver chip	24 mW
GPS receiver chip	15 mW
6D motion sensor	14.4 mW
Cell phone (standby)	8.1 mW
PPG sensor	1.473 mW
Humidity	1 mW
Pressure	0.5 mW
3D accelerometer	0.324 mW
Temperature	27 μW
Cardiac pacemaker	50 μW
Wristwatch	7 μW
Memory R/W	2.17 μW
A-D conversion	1 μW

J Yun, S. Patel, M.Reynolds, G. Abowd “A Quantitative Investigation of Inertial Power Harvesting for Human-powered Devices,” UbiComp’08, September 21-24, 2008, Seoul, Korea.

R.J.M. Vullers, et.al, Micropower energy harvesting, Solid-State Electronics 53 (2009) 684–693

Power Density from Various Harvesters

Ambient RF	< 1 μW/cm^2
Ambient light	100 mW/cm^2 (directed toward bright sun)
	100 μW/cm^2 (illuminated office)
Thermoelectric	60 μW/cm^2
Vibrational	4 μW/cm^3 (human motion ~Hz)
microgenerators	800 μW/cm^3 (machines ~kHz)
Ambient airflow	1 mW/cm^2
Push buttons	50 μJ/N
Hand generators	30 W/kg
Heel strike	Up tp 7 W for 1 cm deflection

Power requirements in conventional sensor network nodes may not be met by harvesting alone!

J Yun, S. Patel, M.Reynolds, G. Abowd "A Quantitative Investigation of Inertial Power Harvesting for Human-powered Devices," UbiComp'08, September 21-24, 2008, Seoul, Korea.

Electromagnetic: Solar → RF

- PV is a good source of energy. Recall,

Ambient light	100 mW/cm^2 (directed toward bright sun)
	100 µW/cm^2 (illuminated office)

- However, it light is not always available.
  - At night
  - Specific scenarios: in a closed chamber, or mine.

- RF is an alternative
  - Unlike others, RF sources may be ambient or intentional.
  - Ambient sources such as base stations or broadcast stations
  - Special sources: RF ID reader, Phone charger, special beacons

Demand and Supply

- The peak currents needed during transmit and receive operation is not achievable using the harvester alone.
- Buffering is also needed to ensure continuous operation during times without power generation.
- The combination of an energy harvester with a small-sized storage is the best approach to enable energy autonomy of the network over the entire lifetime.
  - Rechargeable battery
  - Thin film batteries
    - can be integrated directly in Integrated Circuit (IC) packages in any shape or size,
    - Flexible when fabricated on thin plastics
    - Thin film batteries have high impedance;
    - Low discharge efficiency compared to Li-ion batteries
  - super capacitor
    - Leakage in super capcitors depends on the voltage. Low at low voltage

Internet of Things (IoT)

Embedding short-range mobile transceivers into a wide array of gadgets and everyday items, enabling new forms of communication between people and things, and between things themselves.

RFID + **Sensor** + **Smart Tech** + **Nano Tech** = IoT

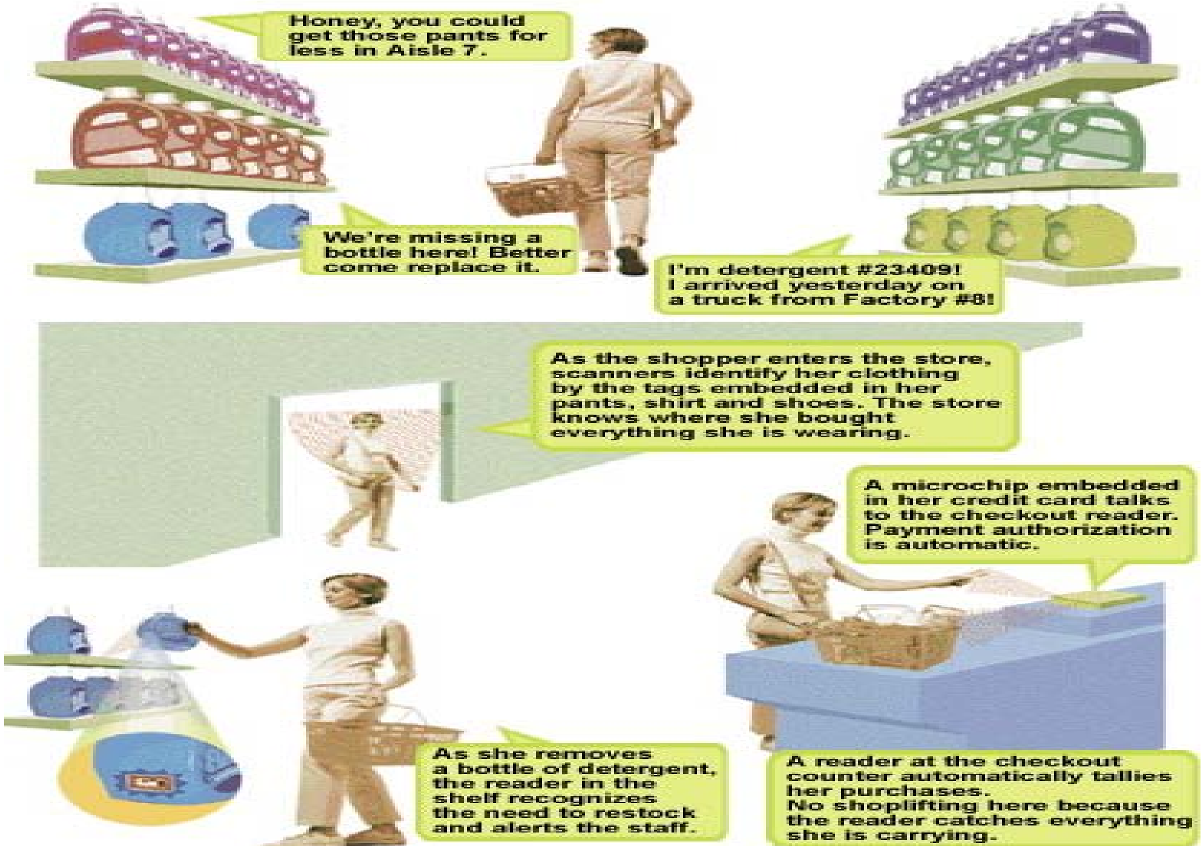
Honey, you could get those pants for less in Aisle 7.
We're missing a bottle here! Better come replace it.
I'm detergent #23409! I arrived yesterday on a truck from Factory #8!
As the shopper enters the store, scanners identify her clothing by the tags embedded in her pants, shirt and shoes. The store knows where she bought everything she is wearing.
A microchip embedded in her credit card talks to the checkout reader. Payment authorization is automatic.
As she removes a bottle of detergent, the reader in the shelf recognizes the need to restock and alerts the staff.
A reader at the checkout counter automatically tallies her purchases. No shoplifting here because the reader catches everything she is carrying.

When shopping in the market, the goods will introduce themselves.

When entering the doors, scanners will identify the tags on clothing.

When paying for the goods, the microchip of the credit card will communicate with checkout reader.

When moving the goods, the reader will tell the staff to put a new one.

Introduction to RFID

- The reader converts incident field and returns useful data
- In passive RFID systems reader transmits EM energy that "wakes up“ the tag and provides power for the tag to respond to the reader.

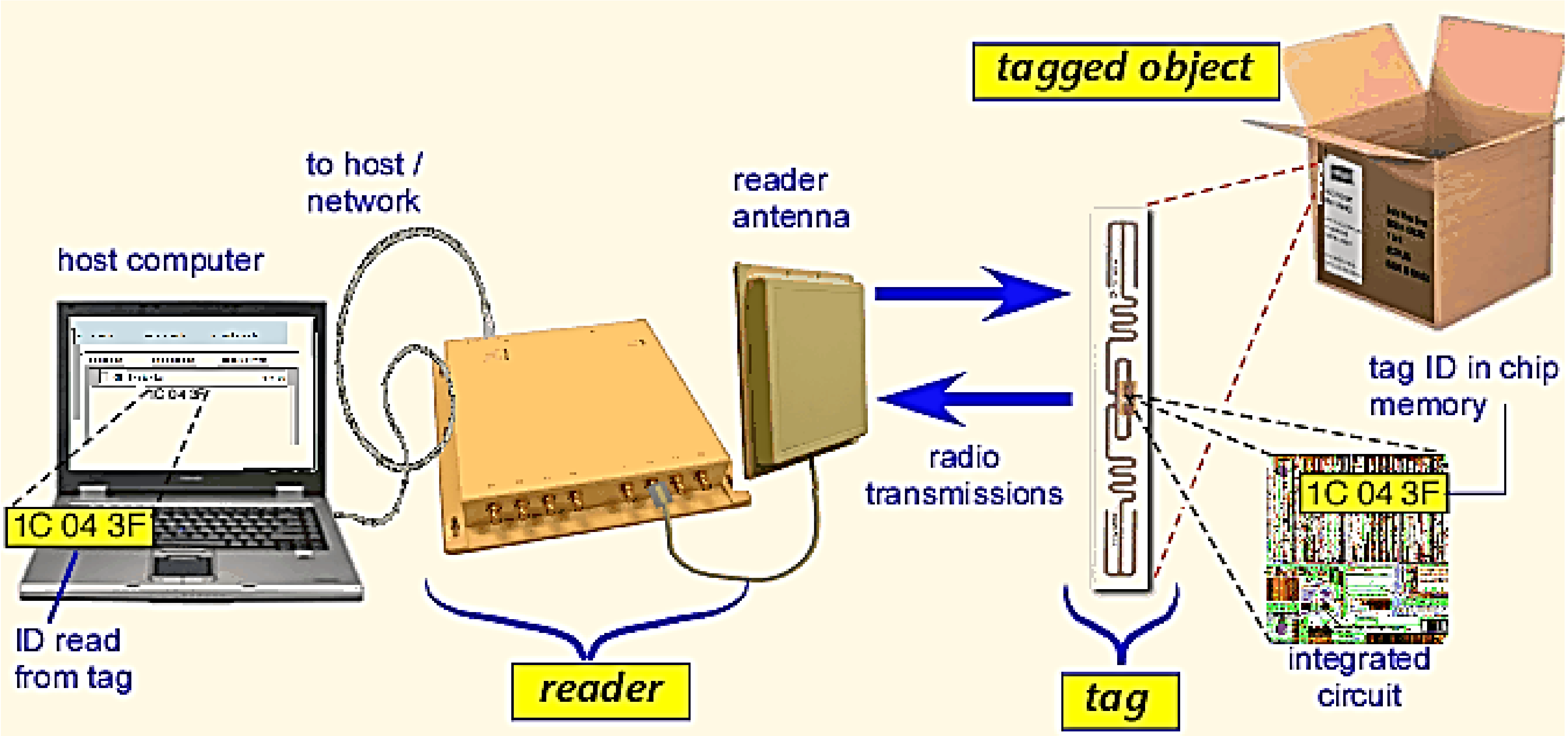
tagged object
to host / network
reader antenna
host computer
tag ID in chip memory
radio transmissions
1C 04 3F
1C 04 3F
ID read from tag
reader
tag
integrated circuit

Backscatter Communication

- **Backscatter** is the reflection of signals back towards their source.
  - In this scheme, two devices communicate using incident (or ambient) RF as the source of power.
  - Backscattering is achieved by changing the impedance of a receiver in the presence of an incident signal.
  - When waves encounter a new media that have different impedances, a part of the wave is reflected.
  - The reflection depends on the difference in the impedances.
- By modulating the impedance at the receiver port, one can control the scattered RF energy, hence enabling information transmission.

abc.cs.washington.edu/files/comm153-liu.pdf

RFID → IoT

- RFID
  - Uses radio waves for identifying or tracking the object.
  - Proven to be a simple and cost effective system
  - Tags are very cheap and is possible to be attached to everyday objects.

- RFID is considered a prerequisite of **Internet of Things.**

- **Example: RFID tags can be integrated with sensors**
  - When a reader reads a tag, the sensor information will be sent to the reader along with the identity of the object.

Some Examples

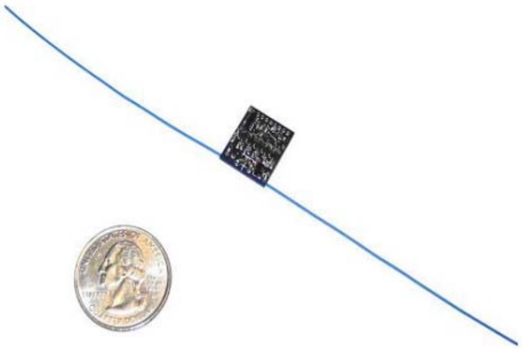

Discrete Element Based

WISP (Wireless Identification Sensing Platform)

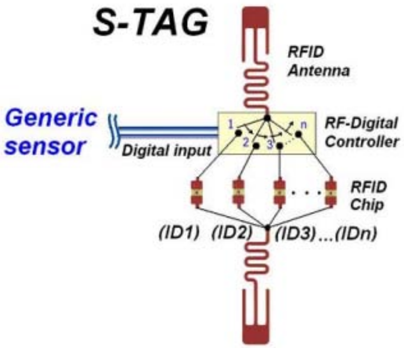
S-TAG
RFID Antenna
Generic sensor
Digital input
RF-Digital Controller
1
2
3
n
RFID Chip
(ID1) (ID2) (ID3)...(IDn)

Multi-Chip Based S-tag

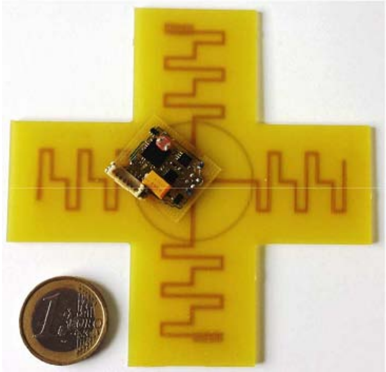

Chips with I2C / SPI

SPARTACUS / RAMSES (Self-Powered Augmented RFID Tag for Autonomous Computing and Ubiquitous Sensing / RFID Augmented Module for Smart Environmental Sensing)

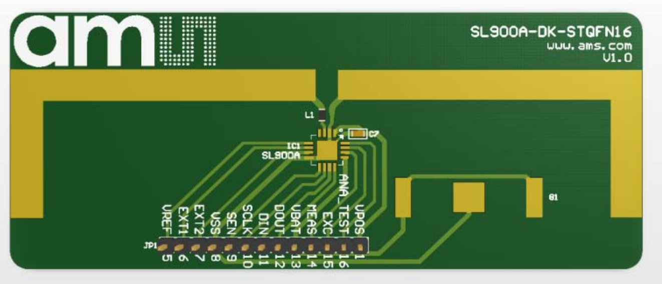
ams
SL900A-DK-STQFN16
www.ams.com
V1.0
L1
IC1
SL900A
C7
B1
JP1
VREF 5
EXT1 6
EXT2 7
VSS 8
SEN 9
SCLK 10
DIN 11
DOUT 12
VBAT 13
MEAS 14
EXC 15
TEST 16
VPOS 1
ANA

Single IC Based Sensing Tags

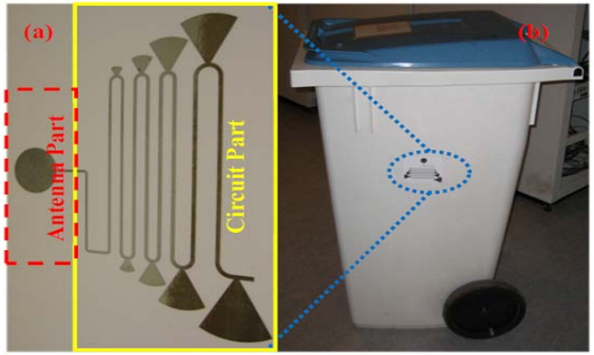
(a)
(b)
Antenna Part
Circuit Part

Printed Chipless RFID Tag

WISP

- Wireless Identification Sensor Platform (2009)
  - WISPs are a wireless, battery-free sensing and computation platform, powered by harvested energy from off-the-shelf UHF RFID readers.
  - To a RFID reader, a WISP is a EPC gen1 or gen2 tag; but inside the WISP, the harvested energy is operating a 16-bit general purpose microcontroller.
  - The microcontroller can perform computing tasks, including sampling sensors, and communicate to the RFID reader.
  - WISPs have been built with various sensors, WISPs can write to flash and perform cryptographic computations.

A collaboration between Intel Research Seattle and the University of Washington.

WISP 4.1DL
Intel Research Seattle

http://wisp.wikispaces.com/Wisp+4.1+DL

RFID Sensors (Products)

- ID operation is passive; yet most sensors require power sources

- Powercast has a wireless sensor that is battery-less. Uses RF energy harvesting.
- Harvesting schemes works at power as low as -12dBm. (RF-DC conversion efficiency above 40% only above -8dBm)
- Harvested power >0.4mW for RF in of >-1dBm.
- Multiple custom ICs and discretes

- Other suppliers include
  - Phase IV
  - RFID sensor systems
  - etc

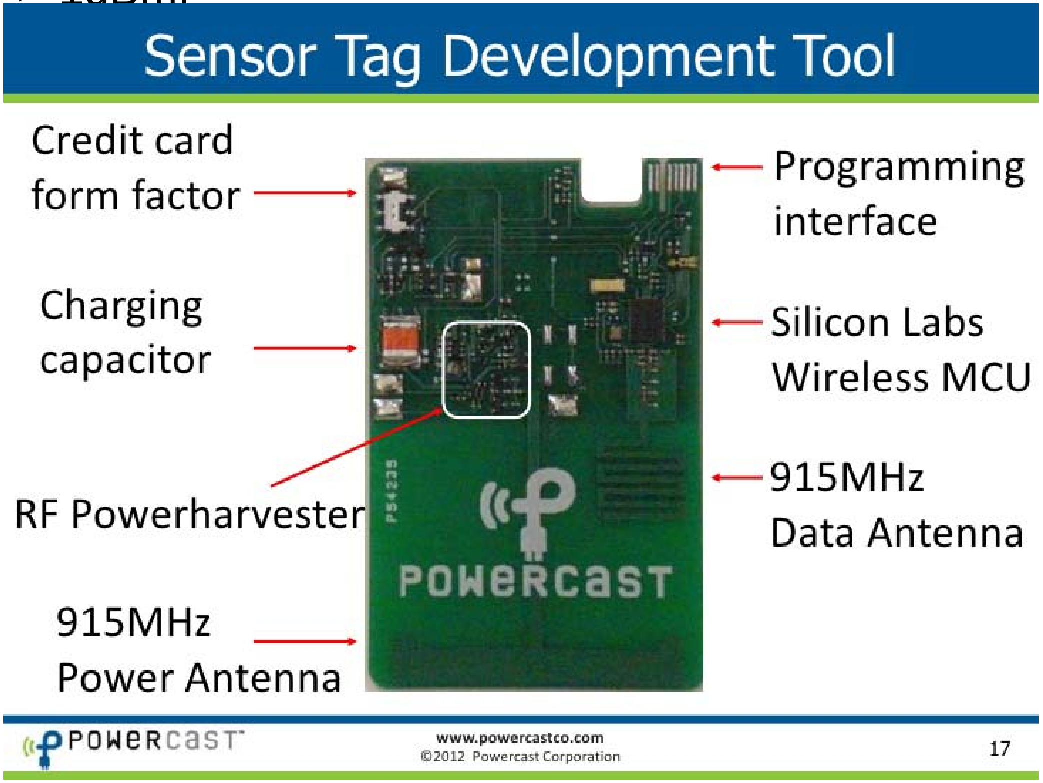

Sensor Tag Development Tool

Credit card form factor

Charging capacitor

RF Powerharvester

915MHz Power Antenna

Programming interface

Silicon Labs Wireless MCU

915MHz Data Antenna

POWERCAST

www.powercastco.com
©2012 Powercast Corporation

17

http://www.powercastco.com/uhf-rfid-sensing-passive-rfid-wireless-sensor-tags/

Battery-less Wireless Terminals

- Most of our work in this direction was towards battery-less terminals
- Long life terminals without wiring
- These are useful when
  - Terminals are embedded within structures (or body)
  - Devices to be deployed in hostile environments
  - Use of battery is not allowed (potential cause for explosion)
- Other factors
  - Cost, weight, etc.
- Primary focus: use of radio frequencies (ambient/intentional)

Ambient RF Sources

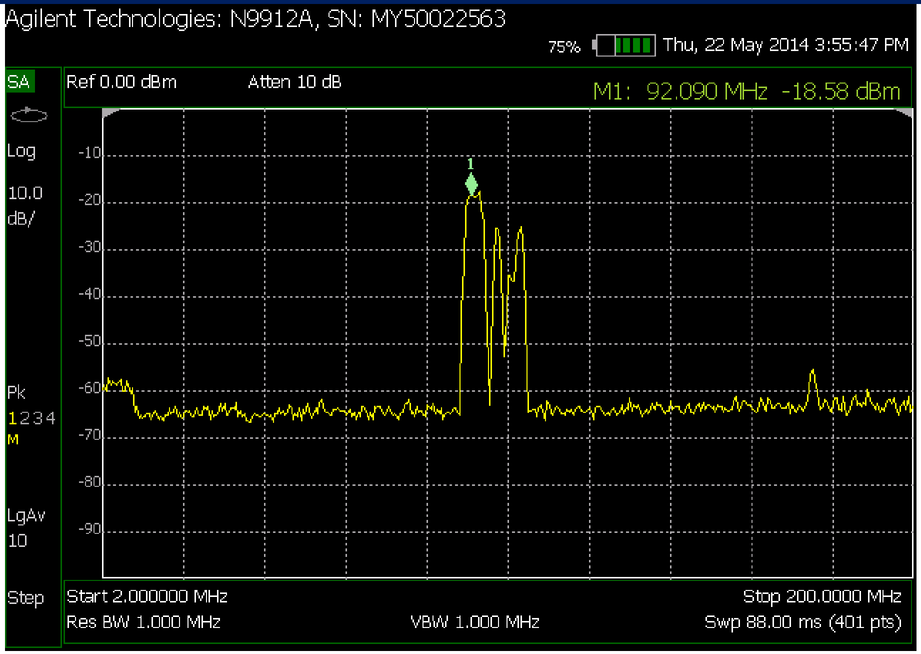
Agilent Technologies: N9912A, SN: MY50022563
75%
Thu, 22 May 2014 3:55:47 PM
SA
Ref 0.00 dBm
Atten 10 dB
M1: 92.090 MHz -18.58 dBm
Log
10.0
dB/
-10
-20
-30
-40
-50
-60
-70
-80
-90
1
Pk
1234
M
LgAv
10
Step
Start 2.000000 MHz
Res BW 1.000 MHz
VBW 1.000 MHz
Stop 200.0000 MHz
Swp 88.00 ms (401 pts)

- Several sources:
  - WiFi Access points (mW) [2.4/.6GHz]
  - Cellular Tower (W) [900/1800 MHz]
  - TV Broadcast (MW) [150-450MHz]
  - FM broadcast (kW) [90-108MHz]
  - AM Radio broadcast (kW) [<1MHz]

- In general
  - Lower frequencies help non-line-of sight propagation
  - Power availability from ambient sources is limited and varies from place to place.

- Note
  - Unlike other sources, most practical RF harvesters (eg in RF ID) depend on **intentionally generated energy**.
  - This is called **wireless power transfer (WPT)** in the conventional RF/Microwave parlance.

Wireless Communication System

- Power transfer scheme is no different!!

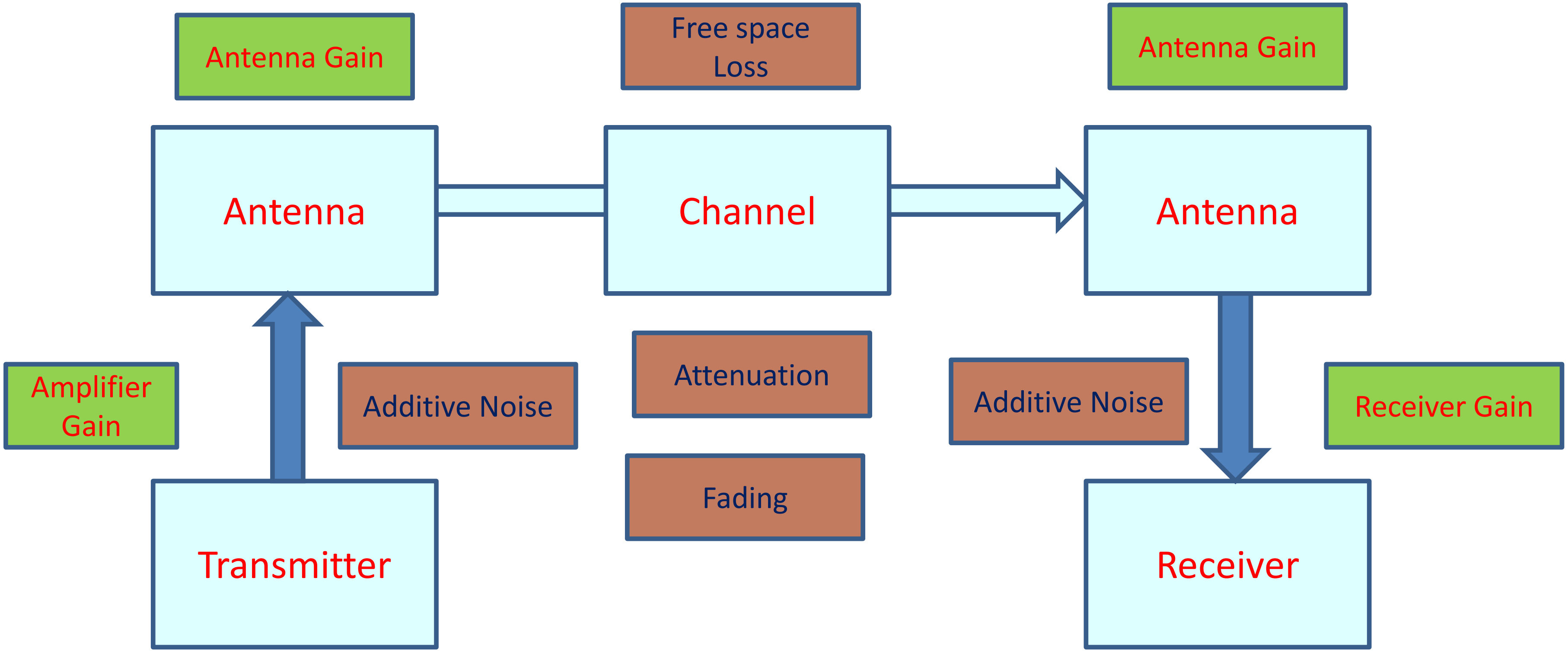

Antenna Gain

Free space Loss

Antenna Gain

Antenna

Channel

Antenna

Amplifier Gain

Additive Noise

Attenuation

Additive Noise

Receiver Gain

Fading

Transmitter

Receiver

Antenna Fundamentals: Directivity

This is the ratio of the radiation intensity in a given direction to the radiation intensity averaged over all direction

Average radiation intensity, $U_0 = \dfrac{P_{rad}}{4\pi}$

Directivity,

$$D(\theta,\phi) = \frac{U(\theta,\phi)}{U_0} = \frac{4\pi U(\theta,\phi)}{P_{rad}}$$

- If direction is not specified, it implies the direction of maximum radiation intensity

$$D_{\max} = \frac{4\pi U_{\max}}{P_{rad}} \qquad D_{dB} = 10\log D$$

Maximum directivity and Maximum effective area

The radiated power density by a transmitter at a distance R

$$W_t = \frac{P_t D_t}{4\pi R^2}$$

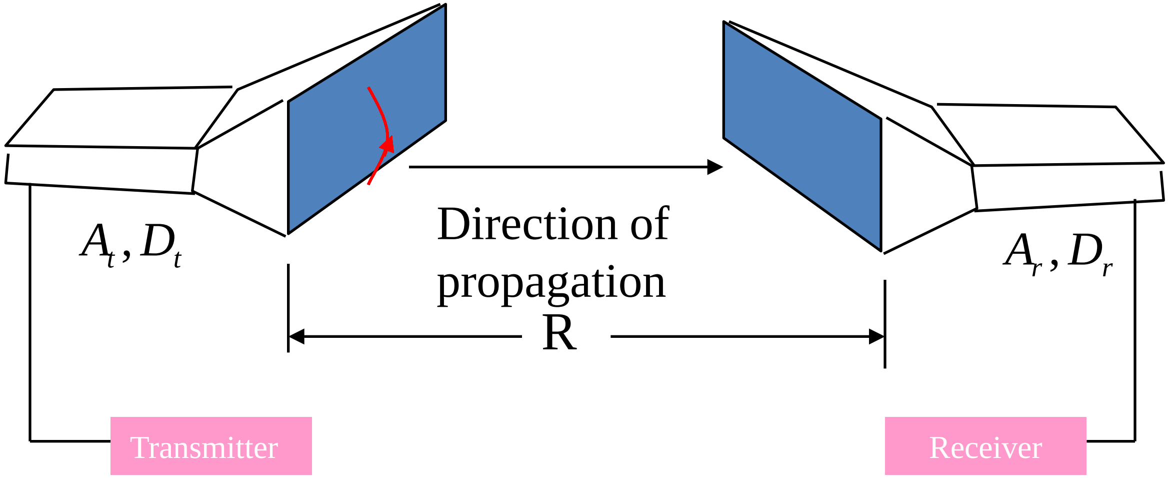
A_t, D_t
Direction of propagation
R
A_r, D_r
Transmitter
Receiver

Power received

$$P_r = W_t . A_r = \frac{P_t D_t A_r}{4\pi R^2}$$

$$or, \; D_t . A_r = \frac{P_r}{P_t} . 4\pi R^2$$

By reversing the transmission direction

$$D_r A_t = \frac{P_r}{P_t} \cdot 4\pi R^2$$

$$\therefore \frac{D_t}{A_t} = \frac{D_r}{A_r}$$

this can be generalized by

$$\frac{D_1}{A_1} = \frac{D_2}{A_2} = k = \frac{4\pi}{\lambda^2}$$

$if, \; D_0 = \text{max. directivity}$

$A_{em} = \text{max. effective area}$

$$D = \frac{4\pi A_{em}}{\lambda^2}$$

Frii's Transmission Equation

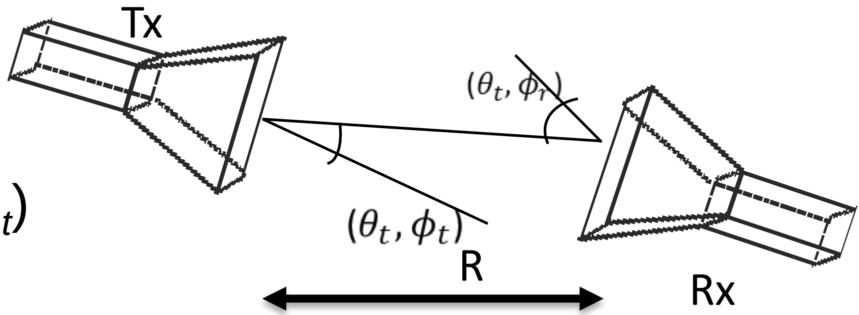

- The radiation intensity for an isotropic radiator is $W_0 = \frac{P_t}{4\pi R^2}$
- For an antenna of gain G_t (or directivity D_t)

$$W_t = \frac{P_t G_t(\theta_t,\phi_t)}{4\pi R^2} = e_t \frac{P_t D_t(\theta_t,\phi_t)}{4\pi R^2}$$

- The effective aperture of a receiving antenna is given by $A_r = e_r D_r(\theta_r,\phi_r)\frac{\lambda^2}{4\pi}$
- Therefore,

$$P_r = e_r D_r(\theta_r,\phi_r)\frac{\lambda^2}{4\pi} W_t = e_t\, e_r D_t(\theta_t,\phi_t) D_r(\theta_r,\phi_r)\frac{\lambda^2}{(4\pi R)^2}|\hat{\boldsymbol{\rho}}_t\cdot\hat{\boldsymbol{\rho}}_r|^2$$

$$\frac{P_r}{P_t} = e_t e_r D_t(\theta_t,\phi_t) D_r(\theta_r,\phi_r)\frac{\lambda^2}{(4\pi R)^2}|\hat{\boldsymbol{\rho}}_t\cdot\hat{\boldsymbol{\rho}}_r|^2$$

$$\frac{P_r}{P_t} = e_{cdt} e_{cdr}(1-|\Gamma_t|^2)(1-|\Gamma_r|^2) D_t(\theta_t,\phi_t) D_r(\theta_r,\phi_r)\frac{\lambda^2}{(4\pi R)^2}|\hat{\boldsymbol{\rho}}_t\cdot\hat{\boldsymbol{\rho}}_r|^2$$

- When the antennas are pointing towards each others' peak radiation direction,

$$\frac{P_r}{P_t} = G_{0t} G_{0r}\left(\frac{\lambda}{4\pi R}\right)^2$$

Note that includes a loss factor (usually called **Free space Loss factor)**
Does not include dissipation/attenuation in medium; caused by spreading

Some numbers on Radiative form of WPT…

- Practical systems will have
  - Operational frequencies in ISM bands.
  - Most terminals are compact.
  - Antenna efficiency is compromised.
  - Nearly isotropic radiations expected.

- Main bottleneck is the physical limits in transmission.

$$P_r = P_t{*}G_t{*}G_r{*}(\lambda/4\pi r)^2$$

  - At 1 GHz (λ=30cm) r=1m; Antenna gain @0dBm, free space loss factor is about **0.06%**
  - Even with a moderate gain transmitter antenna (6dBi) power received @1m for 1W transmission, is just 2mW.
    - Drops to 23μW at 10m !!
    - **The voltage of the signal is low!!**

- In radiative power transfer, Distance from transmitter is a major concern.

Some questions addressed in our work

- **Harvesting of ambient radiations or Radiative transfer of energy addressed**
  - Is it possible to harvest the RF energy from base stations
  - Are there other viable sources of RF energy

- Can low power communication systems be designed to operate entirely from harvested energy
  - Integrate sensors, control, etc

- Can we use RF EH/ WPT to increase the range of backscatter communication (RFID scenario)

Design of Rectifiers

Required for converting incoming RF into DC power.

The challenge lies in maximizing the power conversion efficiency for low input power and minimizing the dimensions.

- RF to DC conversion by rectification of the incident RF signal by a ***Schottky diode***
  - Most diodes have a finite cut-in voltage
  - Diode is a non-linear device (performance depends on current or load)
  - Impedance matching required between antenna and diode
  - In most cases, the input voltage needs to be boosted

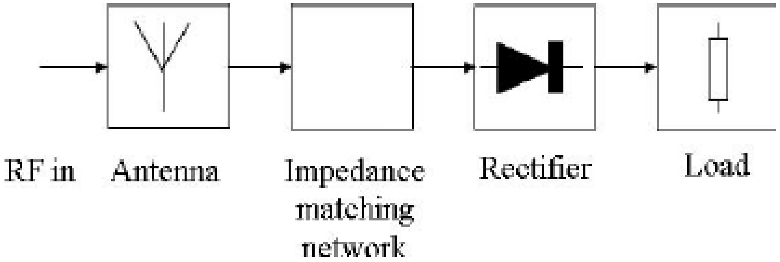
RF in
Antenna
Impedance matching network
Rectifier
Load

Conceptual diode based Rectenna

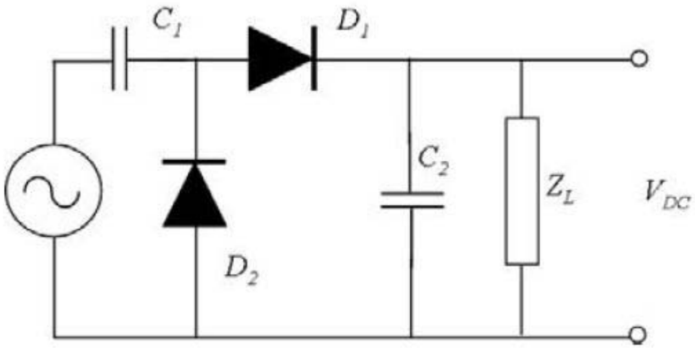
C_1
D_1
C_2
Z_L
V_{DC}
D_2

Voltage doubler

Voltage magnification in Matching circuit

- Matching circuit is required to provide impedance match between antennas (50Ω typical) to diode terminated with high impedance load (capacitor and/or high R in parallel).
  - LC matching networks provide voltage magnification.
  - This helps the diode conduct a good fraction of half cycle.
- The higher the voltage across the diodes, the more efficient the rectifier gets.
  - In practice Q is limited
  - Applications requiring higher voltages, a voltage multiplier configuration is used.

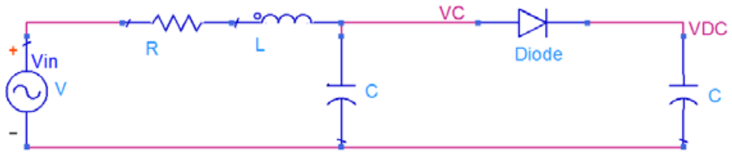

$$V_C = V_{in} \times \frac{\frac{1}{j\omega C}}{R + j\omega L + \frac{1}{j\omega C}}$$

At resonance,

$$V_C = V_{in} \times \frac{1}{j\omega_0 CR} = -jQV_{in}$$

Tuned Rectifier at RF

- A tuned rectifier implemented using discrete components

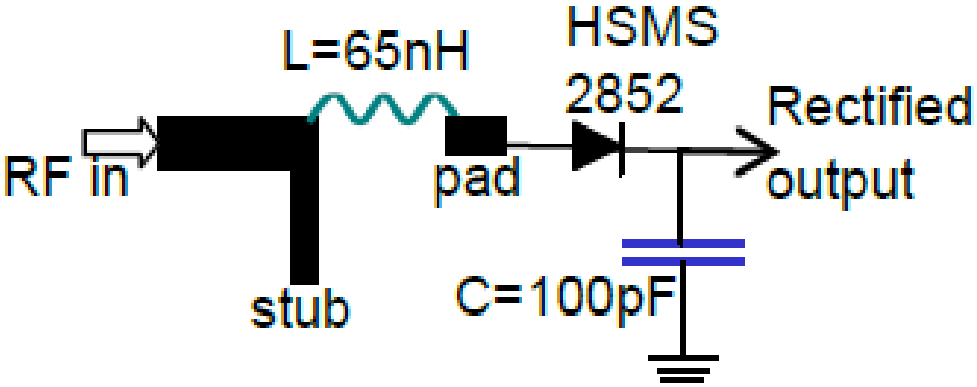

P_in→	**-10dBm**	**-13dBm**	**-16dBm**	**-20dBm**	**-25dBm**
Freq. ↓					
930MHz	917mV	664mV	469mV	281mV	131mV
945MHz	1016mV	736mV	515mV	300mV	132mV
955MHz	1038mV	747mV	513mV	289mV	122mV
960MHz	1032mV	736mV	499mV	276mV	114mV
Peak efficiency	51%	47%	39%	33%	20%

K. J. Vinoy, T. V. Prabhakar, A Universal Energy Harvesting Scheme for Operating Low-Power Wireless Sensor Nodes Using Multiple Energy Resources, pp. 453-466, Micro and Smart Devices and Systems, Springer 2014.

Typical Performance of Rectifier

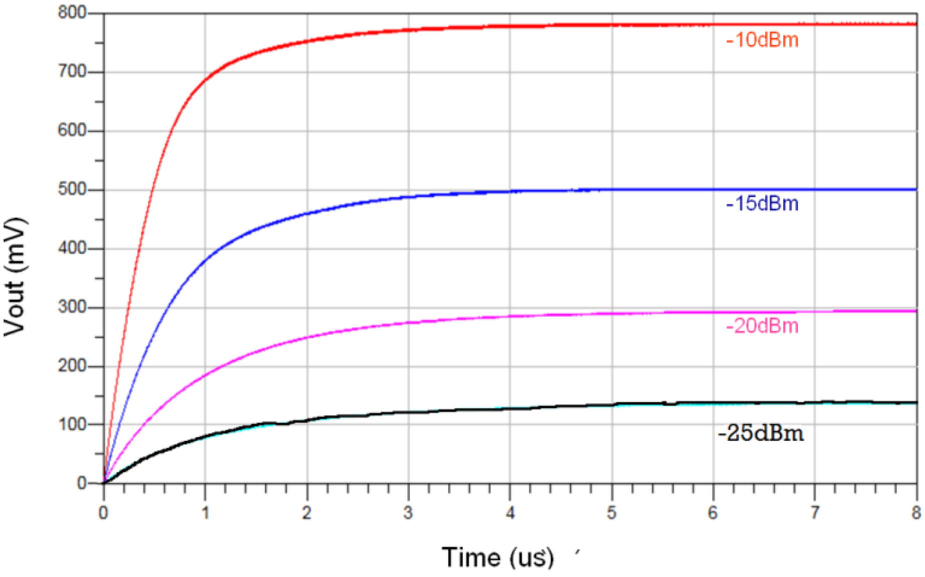

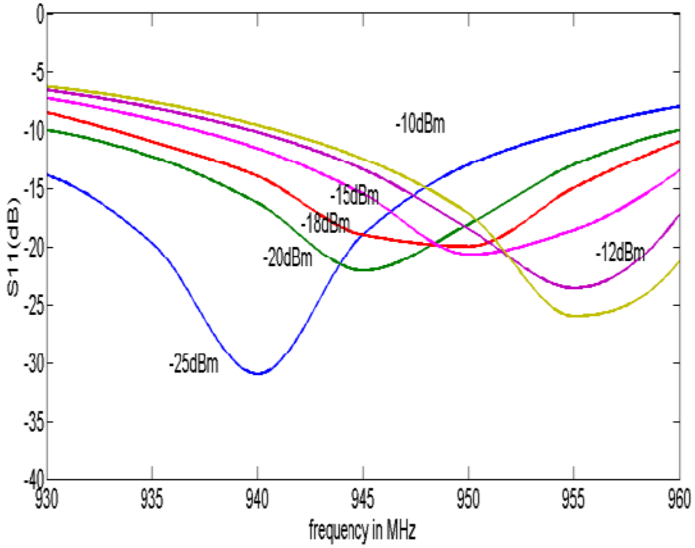

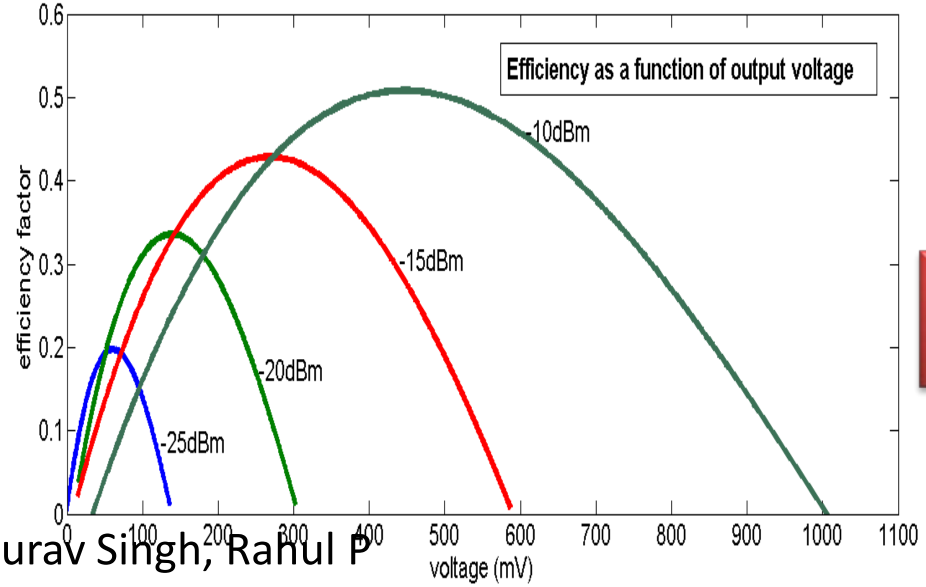

RF-DC Conversion efficiency depends on various conditions

Rectifier Circuit using 4 diodes

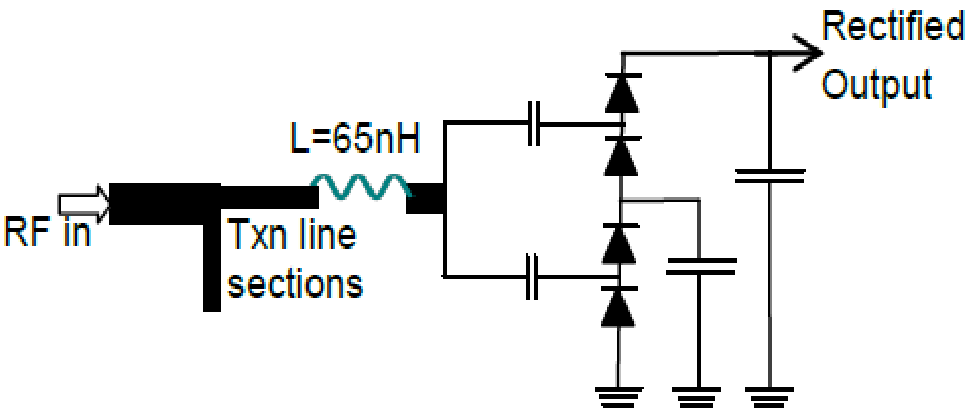
Rectified
Output
L=65nH
RF in
Txn line
sections

Power level (dBm)	Charging time (ms)	Efficiency (%)
0	40	64.11
-2	55	63.77
-3	67.5	63.5
-5	90	63
-7	230	59.89
-10	370	56.78
-12	500	53.89
-15	900	45
-18	2000	20.56

K. J. Vinoy, T. V. Prabhakar, A Universal Energy Harvesting Scheme for Operating Low-Power Wireless Sensor Nodes Using Multiple Energy Resources, pp. 453-466, Micro and Smart Devices and Systems, Springer 2014.

Aditya Mitra, Chaithanya (2011-12)

1: Scavenging Mobile Tower Radiations

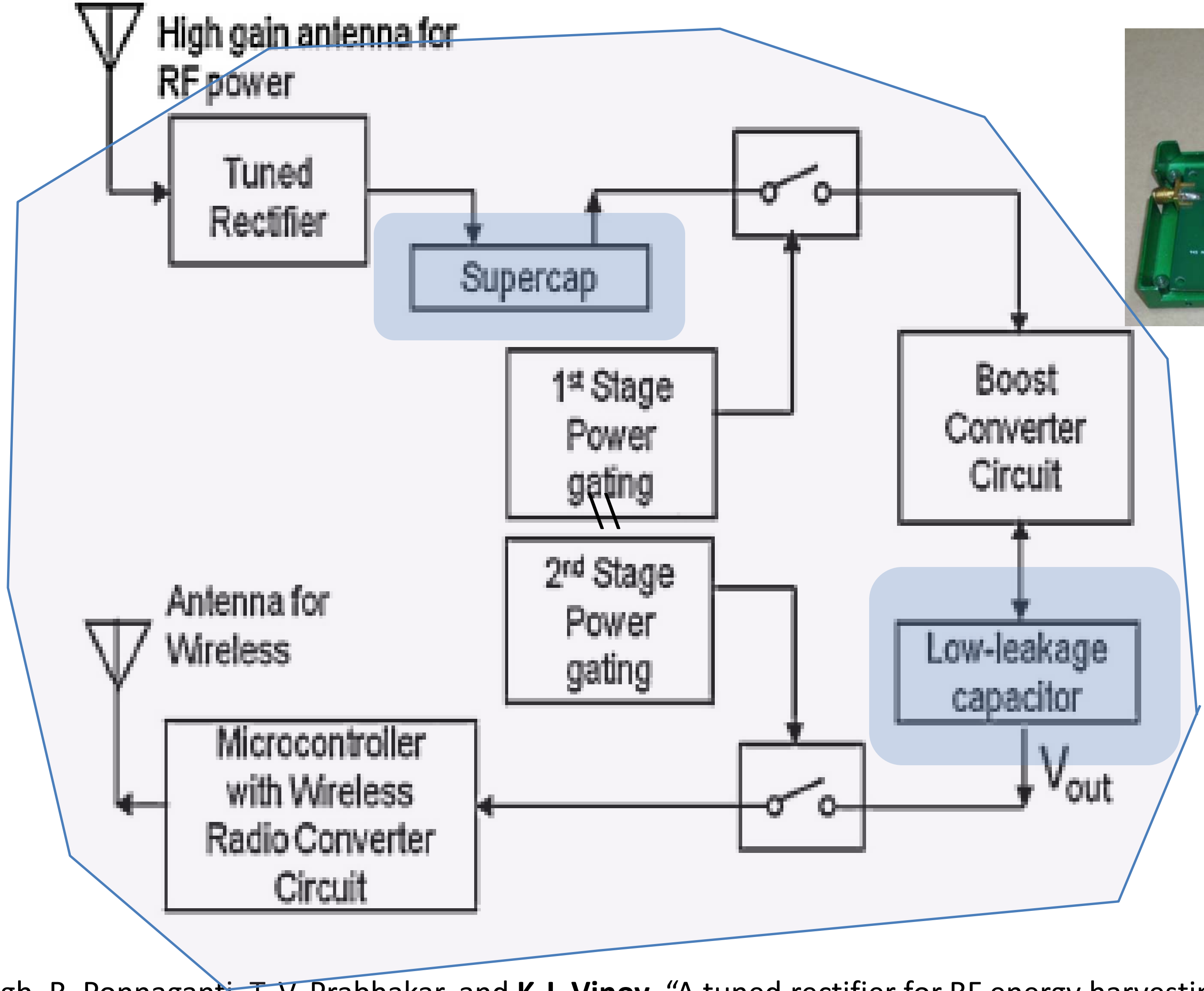
High gain antenna for RF power
Tuned Rectifier
Supercap
1st Stage Power gating
2nd Stage Power gating
Boost Converter Circuit
Low-leakage capacitor
V_{out}
Antenna for Wireless
Microcontroller with Wireless Radio Converter Circuit

G. Singh, R. Ponnaganti, T. V. Prabhakar, and **K.J. Vinoy**, “A tuned rectifier for RF energy harvesting from ambient radiations,” Int. J. Electronics & Communications, vol. 67, no. 7, pp. 564-569, July 2013

Gaurav Singh, Rahul P (2010-11)

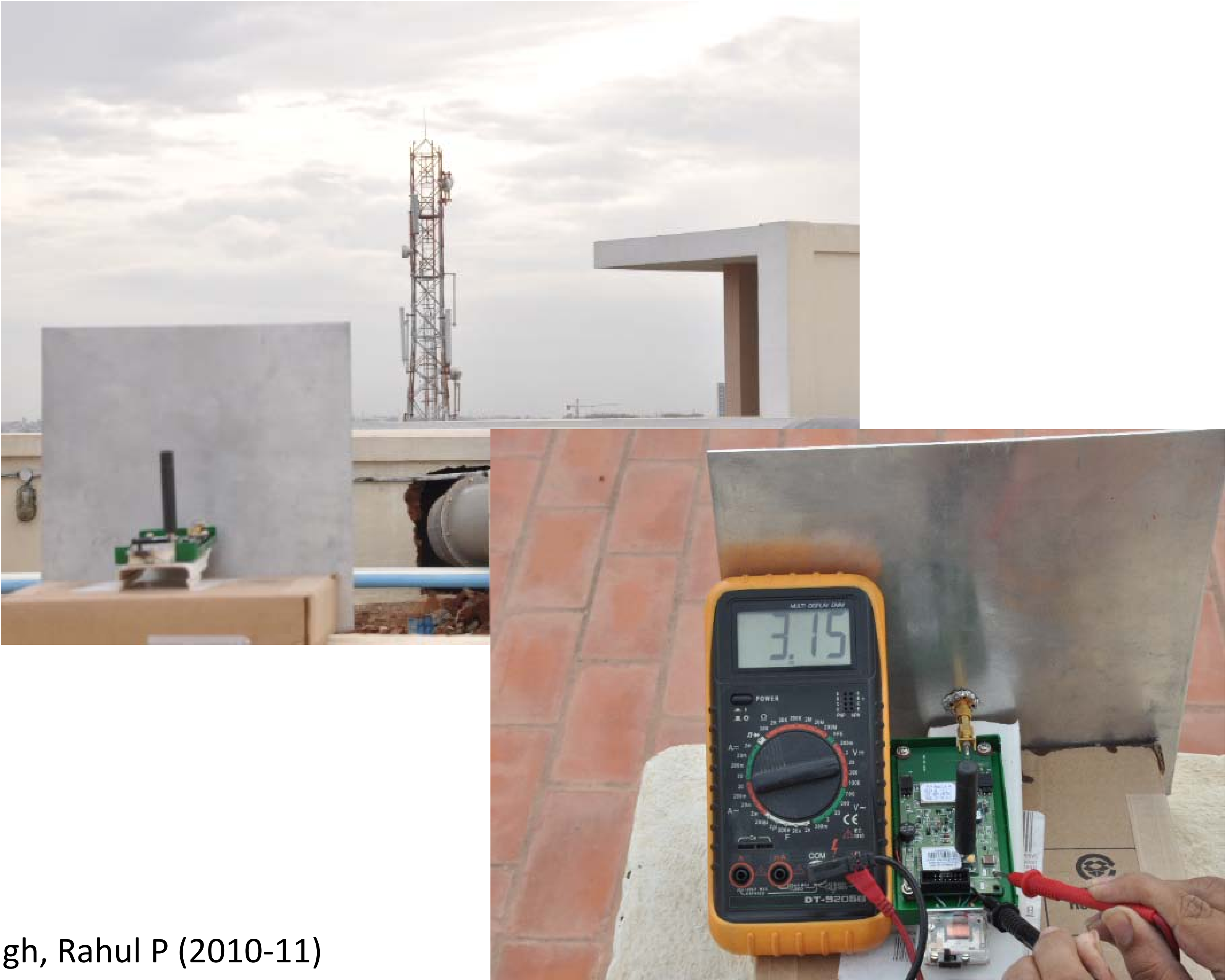
3.15
POWER
DT-9205B

Gaurav Singh, Rahul P (2010-11)

Characterization in Lab

- Using various antennas

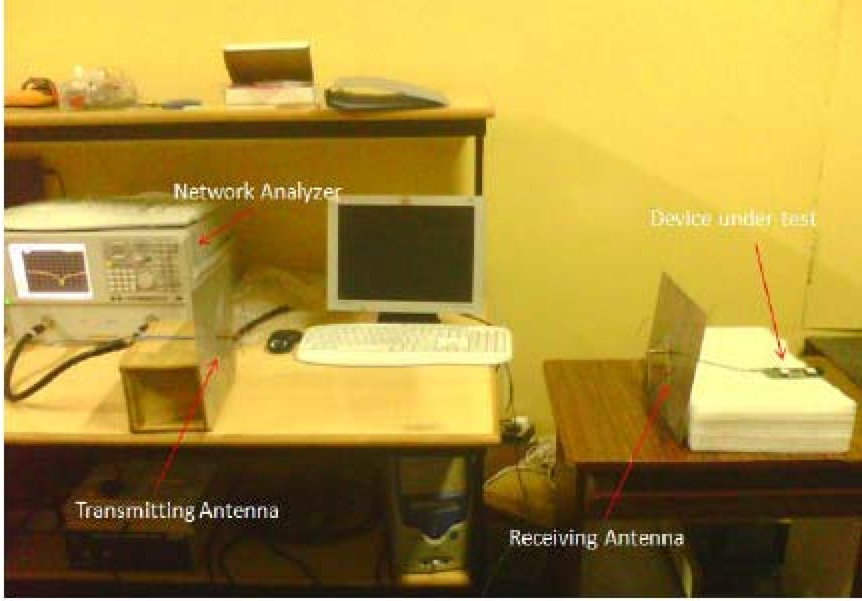
Network Analyzer
Device under test
Transmitting Antenna
Receiving Antenna

Distance from Transmitter [m]	1.5	2	2.5	3
Power received by dipole antenna [dBm]	-20.5	-22.1	-23.9	-25.2
Calculated power density [uW/cm^2]	0.078	0.055	0.035	0.03
Power received by patch antenna [dBm]	-15.1	-16.1	-17.6	-19.2
Transmit interval [mm:ss]	07:26	12:13	25:00	never
Power received by biquad antenna [dBm]	-11.8	-13.2	-14.9	-15.9
Transmit interval [mm:ss]	02:20	03:25	7:10	10:33

G. Singh, R. Ponnaganti, T. V. Prabhakar, and **K.J. Vinoy**, “A tuned rectifier for RF energy harvesting from ambient radiations,” Int. J. Electronics & Communications, vol. 67, no. 7, pp. 564-569, July 2013

Gaurav Singh, Rahul P (2010-11)

Incident power above -18dBm required for operation

2. Universal Energy Harvesting Platform

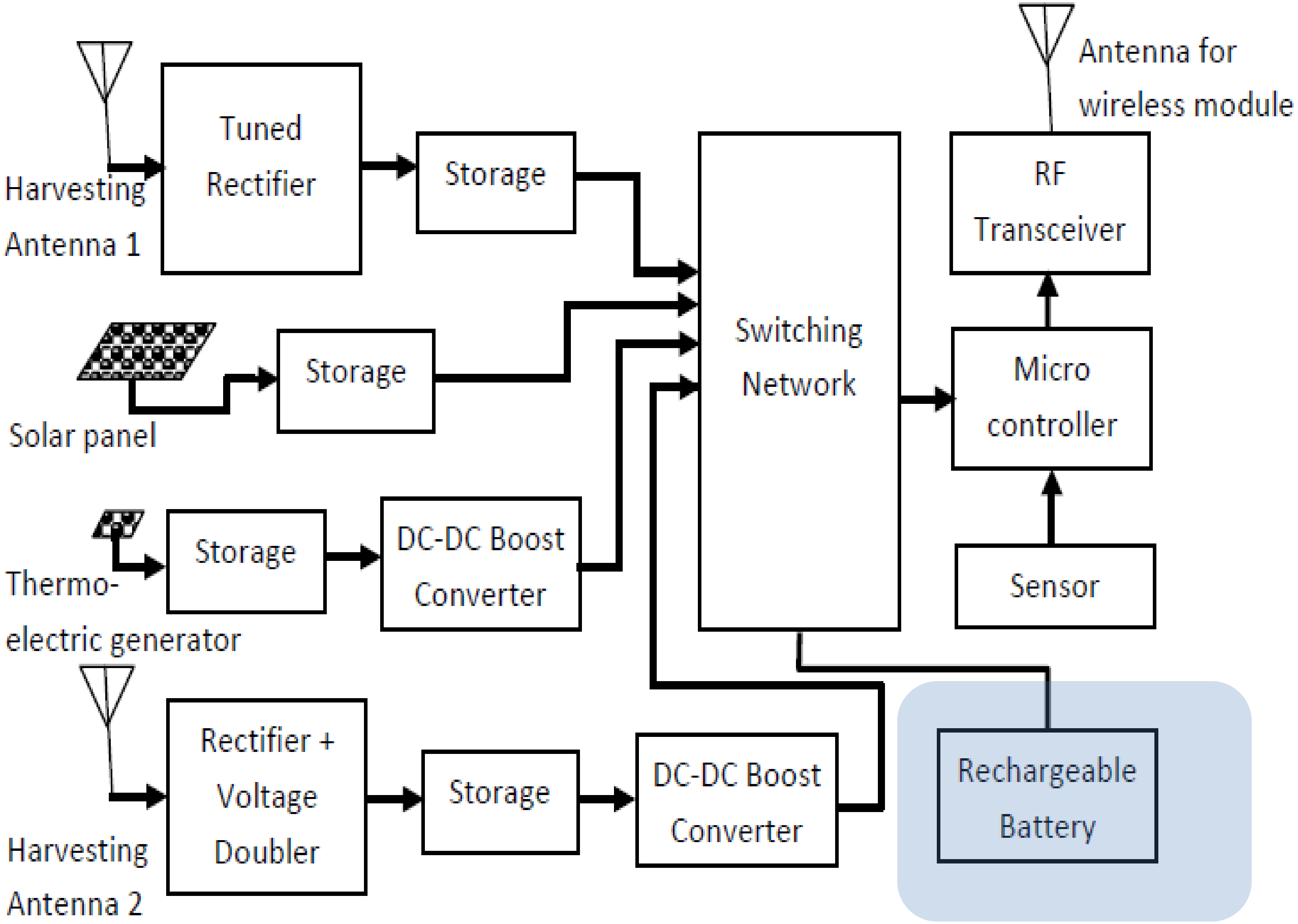
Harvesting Antenna 1
Tuned Rectifier
Storage
Solar panel
Storage
Thermo-electric generator
Storage
DC-DC Boost Converter
Harvesting Antenna 2
Rectifier + Voltage Doubler
Storage
DC-DC Boost Converter
Switching Network
Antenna for wireless module
RF Transceiver
Micro controller
Sensor
Rechargeable Battery

Only a small thin film battery is used

UEHP: Performance with different sources

Solar		**RF**		**TEG**	
Light Intensity (Lux)	Duty Cycle of operation (s)	Power Level (dBm)	Duty Cycle of Operation (s)	Temperature Differential (°C)	Duty Cycle of Operation (s)
1000	7	0	3	55	9
300	11	-5	6	45	13
200	**20**	**-7**	20	35	240
100	42	-10	50	-	-
-	-	-12	240	-	-

An incident RF power of -7dBm (~0.2mW) performs similarly as at low light PV.

An appropriately oriented 20mW source with a high gain antenna (~10dB) can reach this RF power at a low gain rectenna (eg using PIFA) at 1 m distance.

Power levels within emission guidelines…

K. J. Vinoy, T. V. Prabhakar, A Universal Energy Harvesting Scheme for Operating Low-Power Wireless Sensor Nodes Using Multiple Energy Resources, pp. 453-466, Micro and Smart Devices and Systems, Springer 2014.

Aditya Mitra, Chaithanya (2011-12)

Other Possibilities using Wireless Power Transfer

- Power transfer by radiation is not efficient
- Waveguiding systems can ensure better transmission of power
  - Loss in waveguide is a small fraction of a dB/m (~0.2dB/m)
  - Metal ducts may carry higher order modes with higher losses
  - Extended to conducting ducts, Tunnels, mine shafts etc with some compromise
- Other possibilities
  - Surface wave
  - Focusing of fields

- Empty enclosures with metallic walls
  - Containers, tanks, airplane cabin, trains, etc
  - Other objects in the path may reduce the efficiency!

3: Antenna for RFID sensors

- Work involved design of antenna for RF harvesting sensors
  - These fuel level sensors to be deployed in a fuel tank of aircrft.
  - Optimization of design should focus on efficiency
  - High gain or directivity is not required.

- EH platform to be used with RFID sensors deployed inside fuel tank

- Requirements/Assumptions:
  - Incident energy is of random polarity and direction.
  - Operating frequency is 902MHz-928MHz.
  - Antenna must operate in air (relative permittivity = 1) and fluid (relative permittivity = 2.1)
  - Dimensions of planar antenna board:
    - Target dimensions: 3 in. x 2 in.
    - Maximum dimensions: 6 in. x 4 in.

Antenna Design dimensions

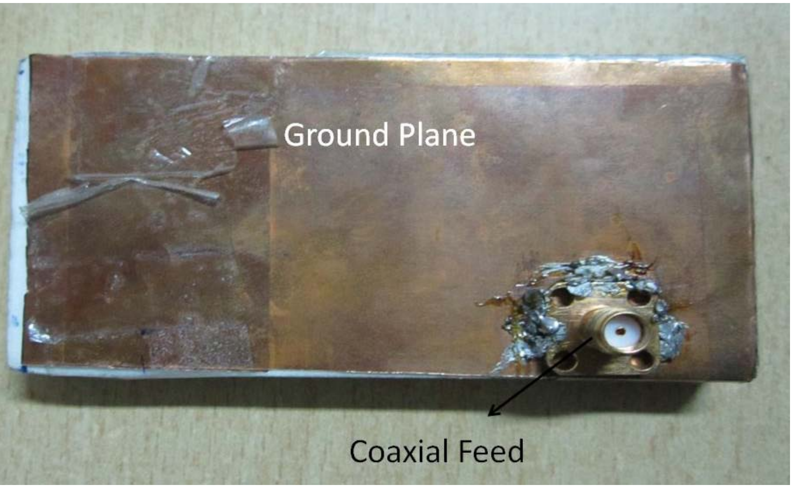
Ground Plane
Coaxial Feed

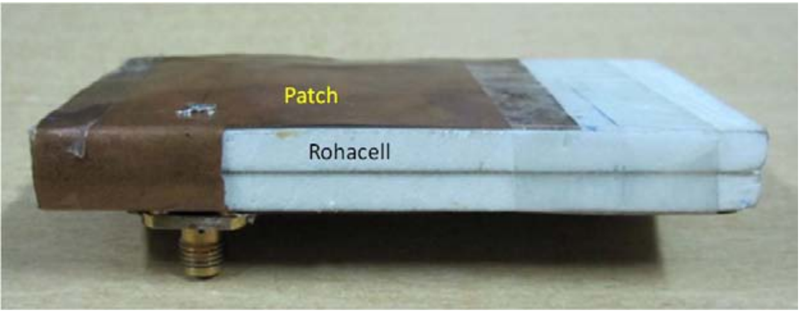
Patch
Rohacell

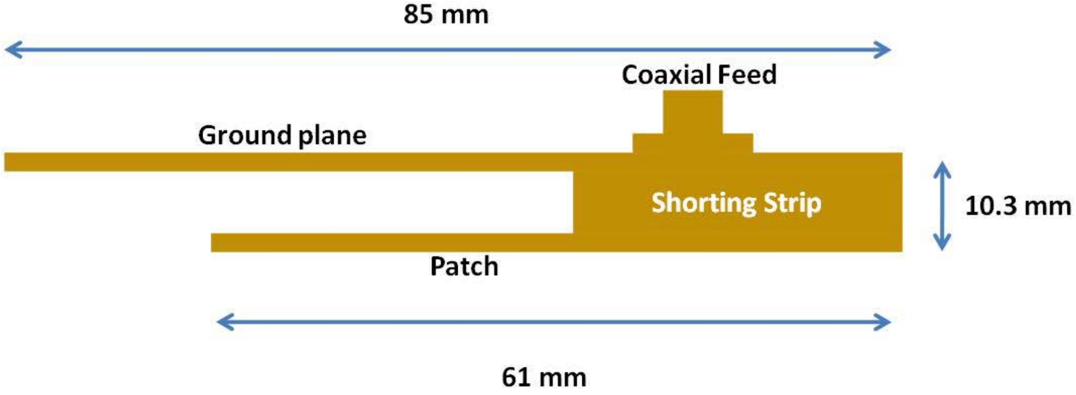
85 mm
Coaxial Feed
Ground plane
Shorting Strip
10.3 mm
Patch
61 mm

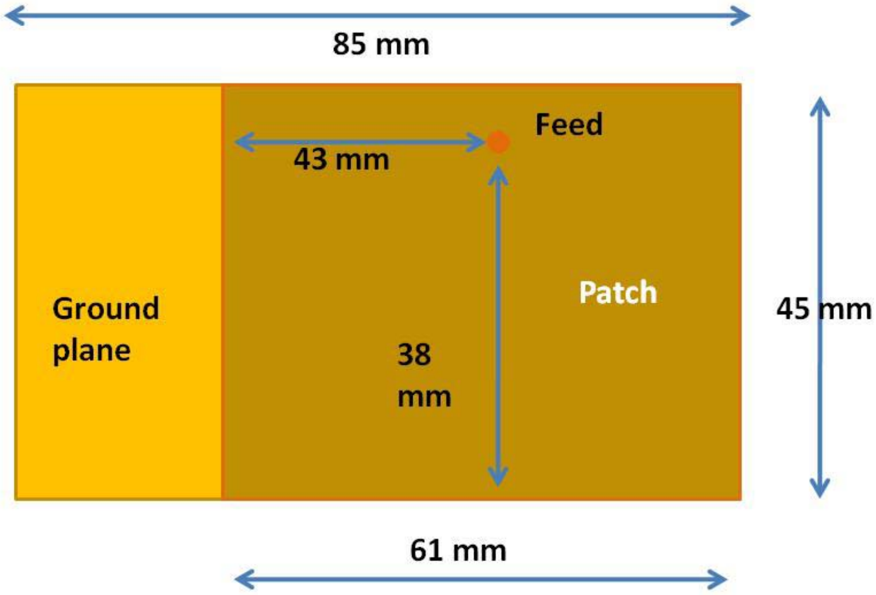
85 mm
Feed
43 mm
Ground plane
Patch
38 mm
45 mm
61 mm

Vivekanand M, Harikiran M

Measurements at Boeing (Nov 2013)

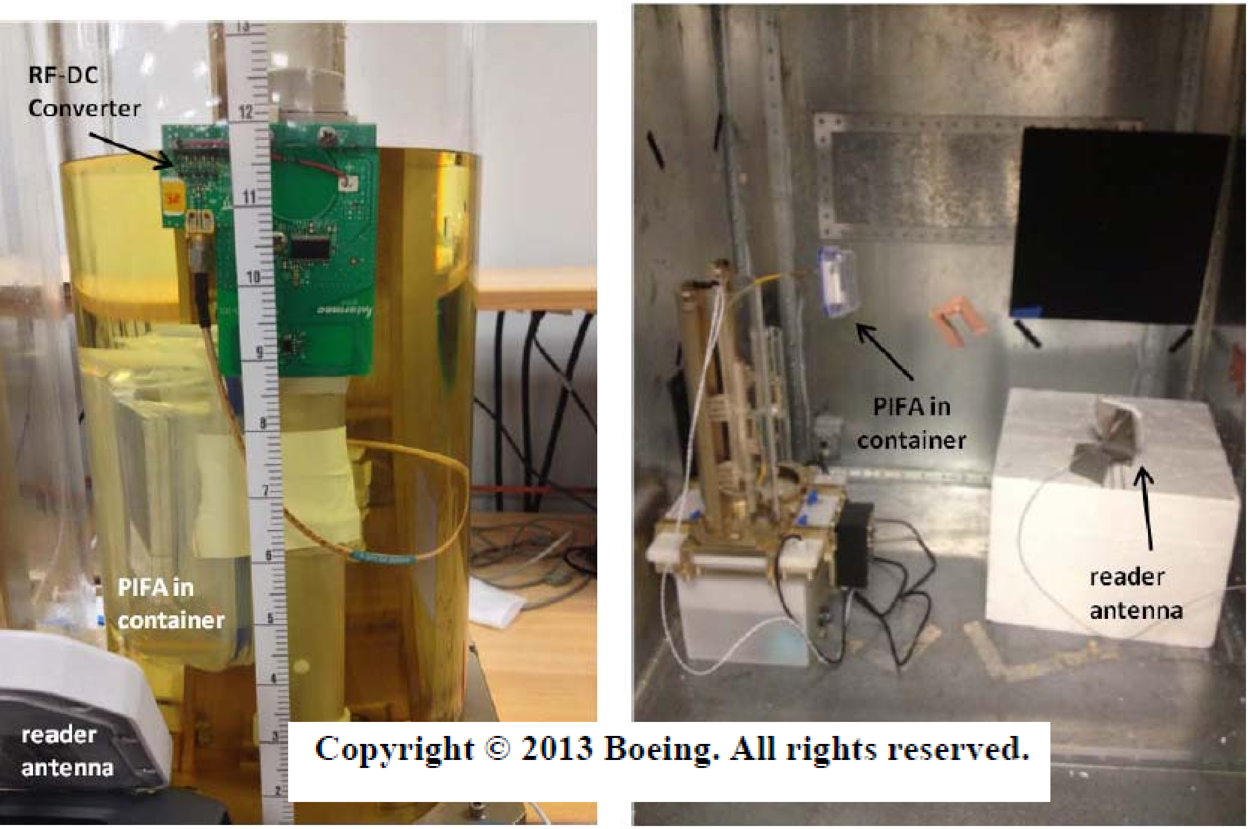
RF-DC
Converter
PIFA in
container
reader
antenna
PIFA in
container
reader
antenna

Copyright © 2013 Boeing. All rights reserved.

After Integrating with Sensor and RFID board;

Measured in a room and reverberation chamber

ANRC

Resolution of fluid height measurement to within 0.25”.
1W maximum transmit power.
Uses a modified reader protocol.

A. Robb, J. Bommer, R. Martinez, J. Harrigan, S. Ramamurthy , H. Muniganti, V. Mannangi, and KJ Vinoy, “Wireless Aircraft Fuel Quantity Indication System,” 2014 IEEE Sensors Applications Symposium - , Feb 18-20, 2014, Queenstown, Newzealand.

4. RFID Integrated with Sensor

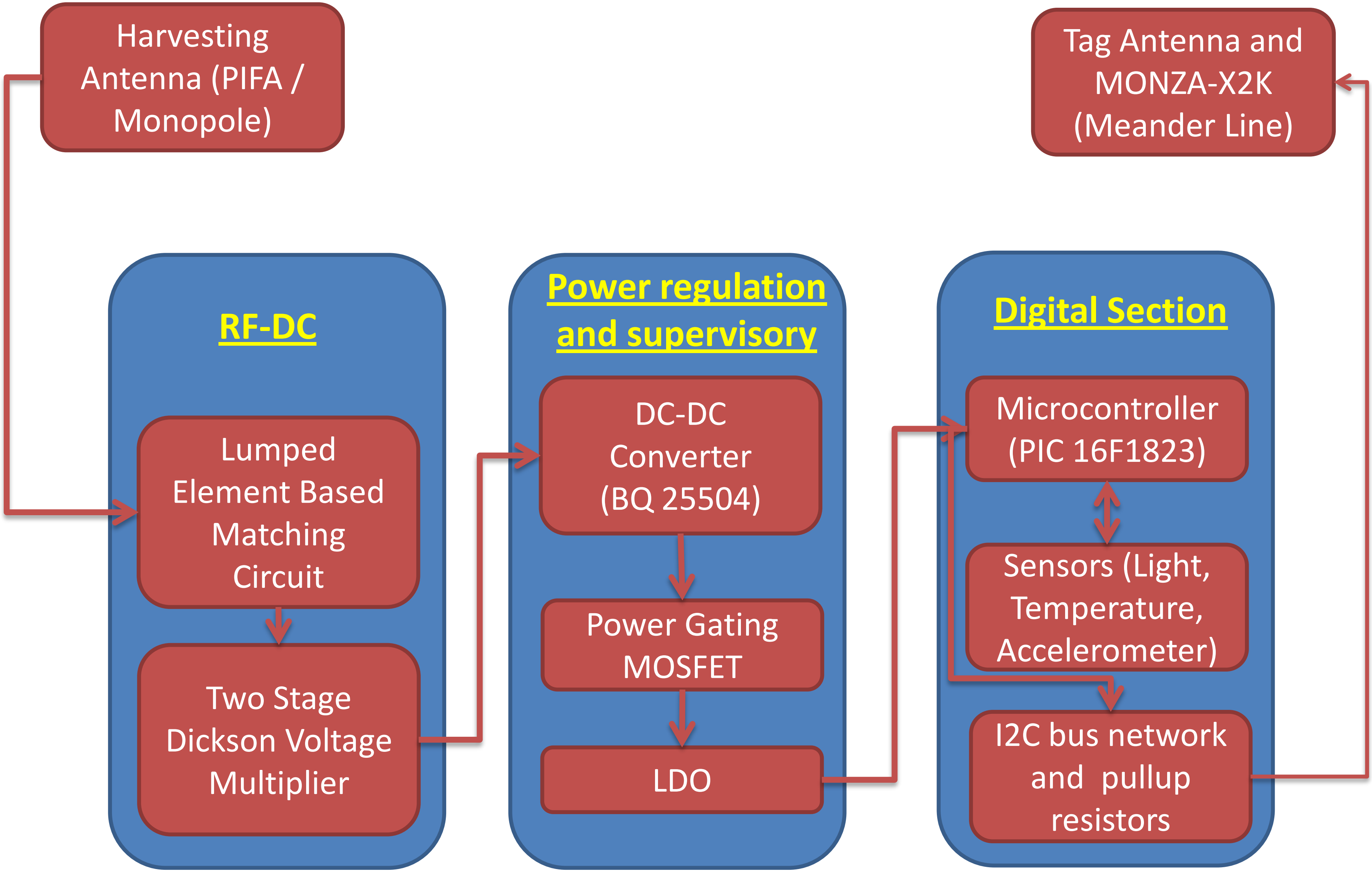

Harvesting Antenna (PIFA / Monopole)

Tag Antenna and MONZA-X2K (Meander Line)

RF-DC

Lumped Element Based Matching Circuit

Two Stage Dickson Voltage Multiplier

Power regulation and supervisory

DC-DC Converter (BQ 25504)

Power Gating MOSFET

LDO

Digital Section

Microcontroller (PIC 16F1823)

Sensors (Light, Temperature, Accelerometer)

I2C bus network and pullup resistors

Parts of Fabricated System

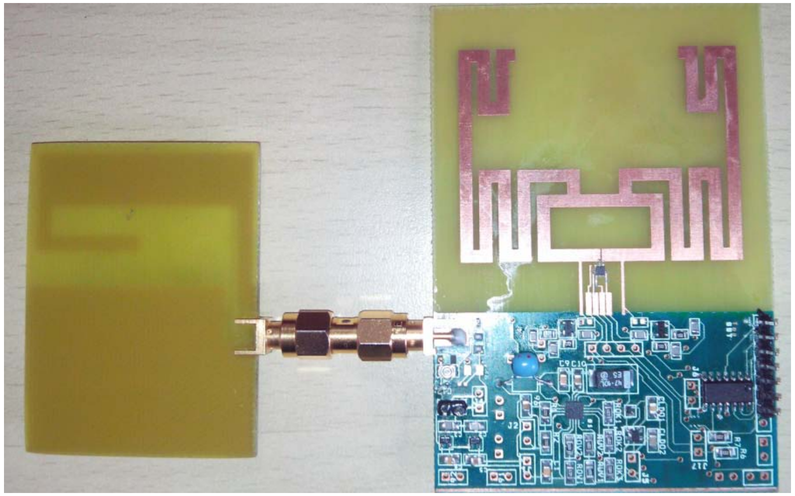

PIFA CONNECTED SENSING TAG

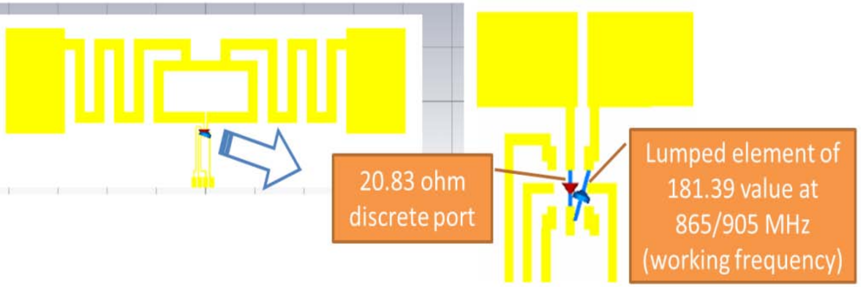

20.83 ohm discrete port

Lumped element of 181.39 value at 865/905 MHz (working frequency)

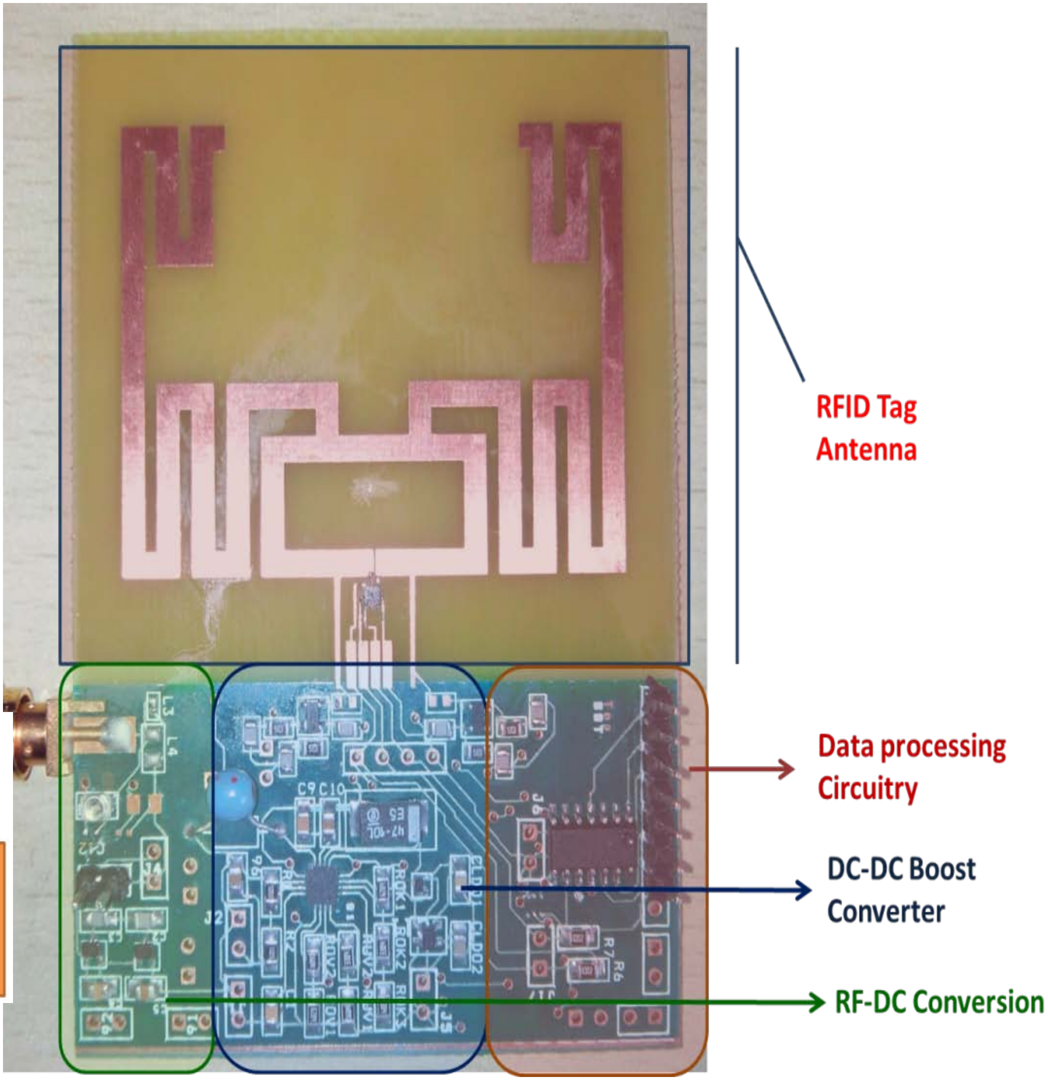

RFID Tag Antenna

Data processing Circuitry

DC-DC Boost Converter

RF-DC Conversion

Sandeep Rana 2014-15

Sandeep Rana, TV Prabhakar, KJ Vinoy, An Efficient Architecture for Battery-less Terminals for Internet of Things, Applied Computational Electromagnetic Conference, Guwahati, Dec 28-21, 2015

Characterization of Performance

Source	Power
RFID Reader	30 dbm
Circularly polarized antenna	8 dbi
Polarization loss	3 dbi
Monopole Antenna	5 dbi
PIFA antenna	1 dbi
Meander Antenna	0.4 dbi

Tag Antenna	EIRP (dbm) + Gr	Rg Expected (-10 dbm) and 50 % overall efficiency	Range Achieved
Monopole	40	7.5 mtr	7 mtr
PIFA	36	5 mtr	5.5 mtr
Meander Line	35.4	4.5 mtr	4.5 mtr

$$\lambda = 34.5\ cm \qquad \text{Prx} = \text{EIRP} * \text{Gr} * \left(\frac{\lambda}{4*\pi*r}\right)^2$$

Efficiency worked out for -10 dbm
RF-DC efficiency – 20 % and DC-DC efficiency – 80%
Overall efficiency – 16%

Sandeep Rana 2014-15

5. Harvesting at 2.4GHz

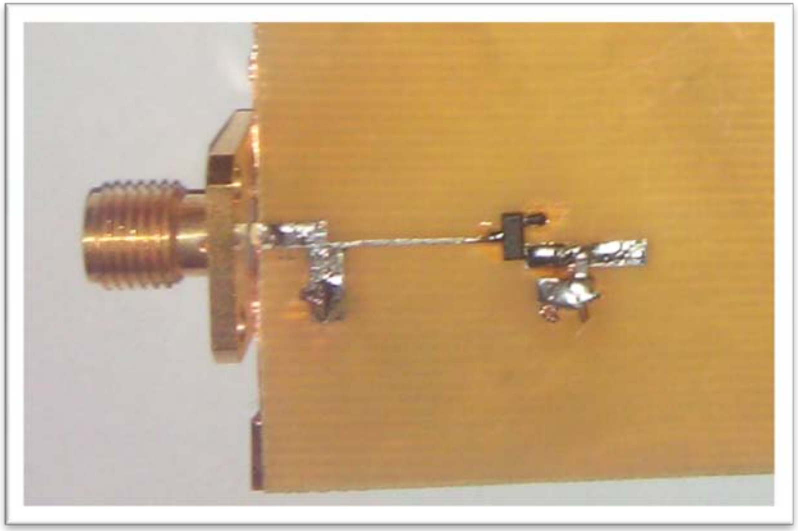

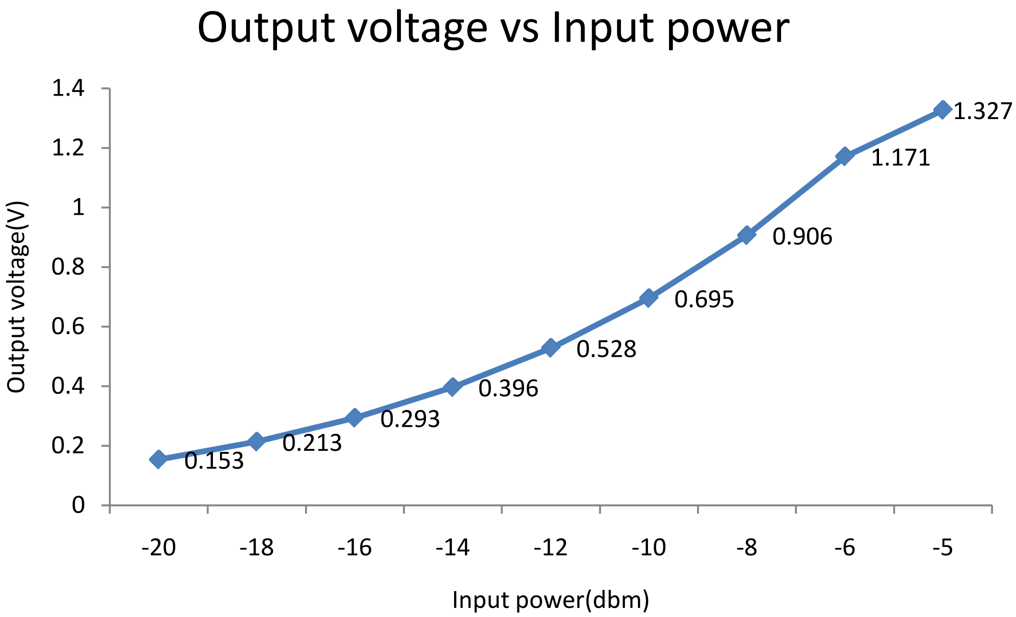

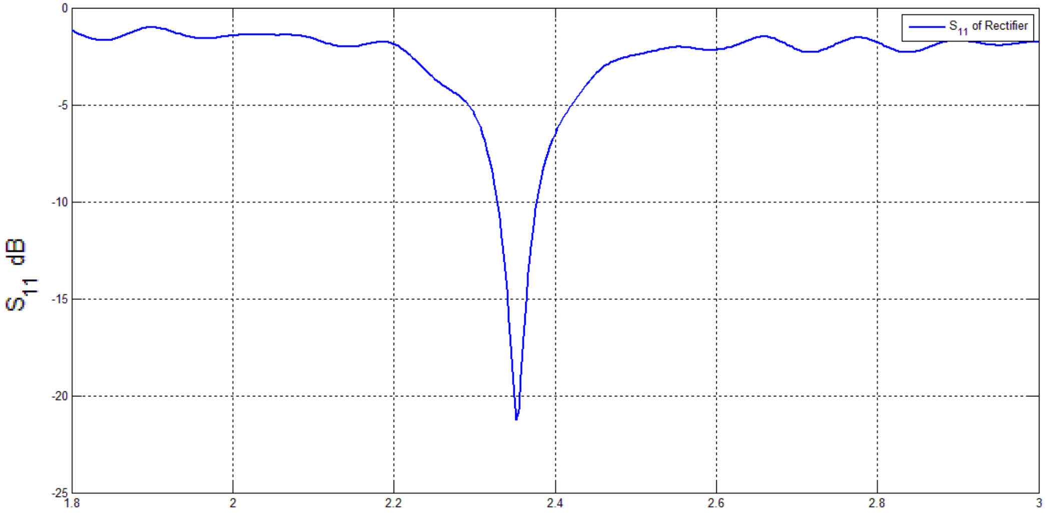

Sanjeev K. 2014-15

Comparison of Efficiencies

Schottky diode

	900MHz	2400MHz
-15dbm	Efficiency=47% Output voltage=0.29V Load resistor=6K	Efficiency =17% Output voltage=0.2V Load resistor=6K
-20dbm	Efficiency=31% Output voltage=0.15V Load resistor=6K	Efficiency=5% Output voltage=0.1V (0.153V unloaded) Load resistor=6K
	Used HSMS 2852	Used HSMS 2862

Diode connected MOS with high Q matching

	900MHz	2400MHz
-14dbm	Efficiency=6.2% Output voltage=1.1V Load resistor=500K	Efficiency=2% Output voltage=0.632V Load resistor=500K
-20dbm	Efficiency=1.8% Output voltage=0.3V Load resistor=500K	Efficiency=0.2% Output voltage=0.118V Load resistor=500K

Zero V_{TH} CMOS

	900MHz	2400MHz
-15dbm	Efficiency=4.6% Output voltage=0.86V Load resistor=500K	Efficiency=3.79% Output voltage=0.774V Load resistor=500K
-25dbm	Efficiency=2.4% Output voltage=0.198V Load resistor=500K	Efficiency=1.7% Output voltage=0.165V Load resistor=500K

← Power

← Technologies

Frequencies →

Cross-coupled Rectifiers for Low Power

- CMoS integration requires diodes using MoSFETs.
- Simple diode connected configurations are not effective at low power/voltage levels
- In Cross Coupled Rectifiers
  - Biasing of MOSFETs by charge stored in capacitors. This is a way of threshold compensation.
  - Low ON resistance due to high overdrive voltage.
  - In both cycles of input, output capacitor is charged. Although DCP uses both cycles, only alternate cycles charges the output capacitor and the other cycle charges the input capacitor.

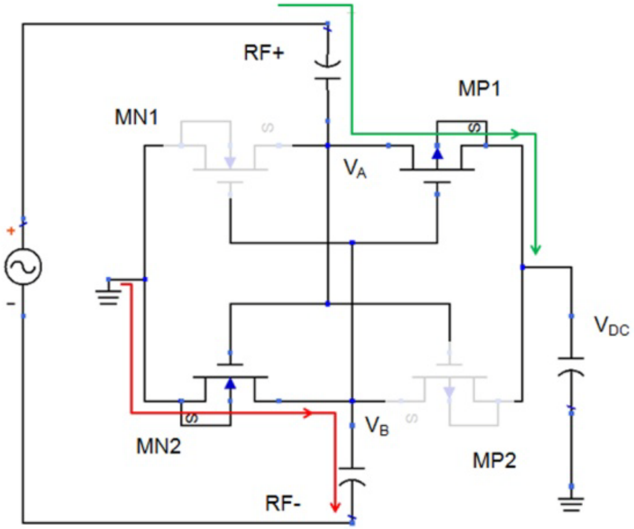

DC-DC converter

- Low loss switched capacitor DC-DC converter:

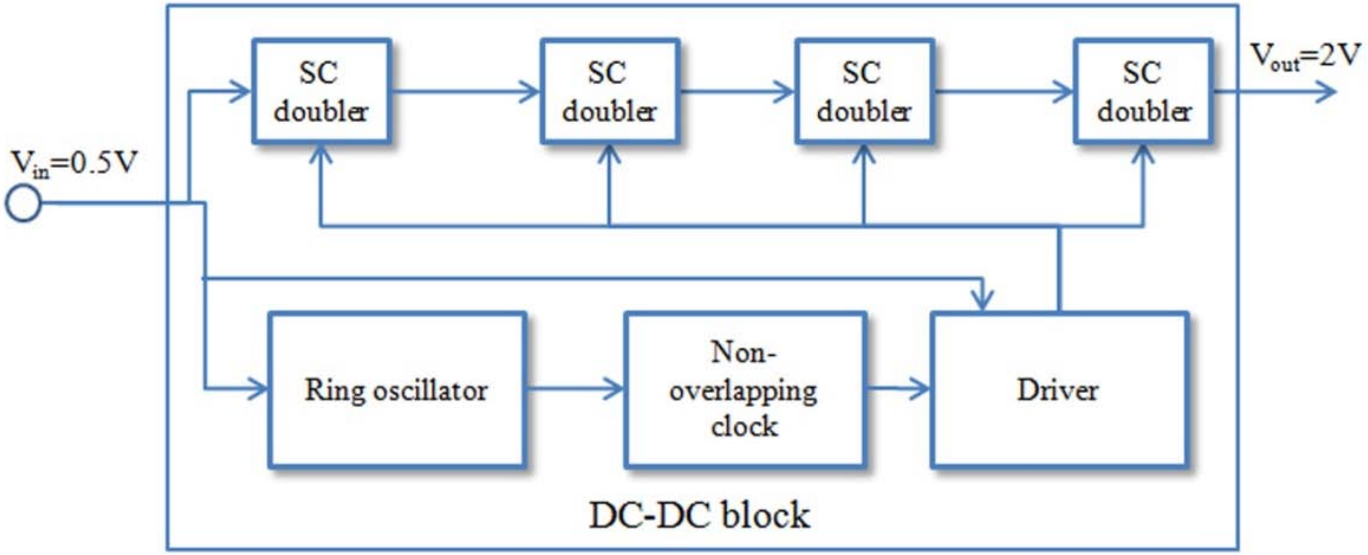
SC doubler
SC doubler
SC doubler
SC doubler
V_{out}=2V
V_{in}=0.5V
Ring oscillator
Non-overlapping clock
Driver
DC-DC block

Full system block diagram

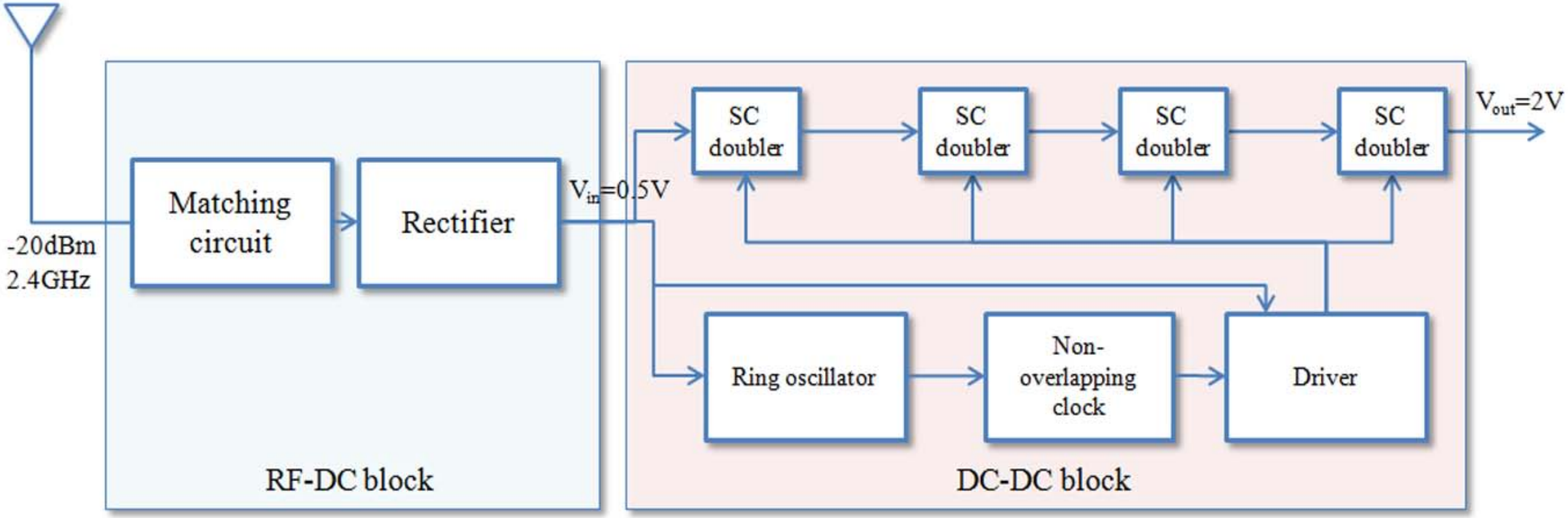

-20dBm
2.4GHz
Matching circuit
Rectifier
V_{in}=0.5V
RF-DC block
SC doubler
SC doubler
SC doubler
SC doubler
V_{out}=2V
Ring oscillator
Non-overlapping clock
Driver
DC-DC block

1. Output capacitor of RF-DC supplies DC-DC
2. Enable generator logic constructed using back to back inverters
3. 5 MOSFETs added to limit supply voltage to ring oscillator

RICOH
imagine. change.

Sanjeev K, Manjunath M, TV Prabhakar, KJ Vinoy, Some Practical Considerations of RF to DC Converter using Low Vth CMOS Rectifier, Applied Computational Electromagnetic Conference, Guwahati, Dec 28-21, 2015

Full system simulations

- Output capacitor of RF-DC loaded heavily when clocks transition, so ripples exist in RF-DC output.
- Below 0.5V at clock transitions, above 0.5V between clock transitions
- Time step is 8ps for RF-DC simulation and DC-DC has to run for hundreds of µs or few ms, so simulations times are large.

Output voltage=2V across a load of 1.6MΩ, which gives

$$Efficiency = \frac{2^2/1.6\times10^6}{10^{-5}} = 25\%$$

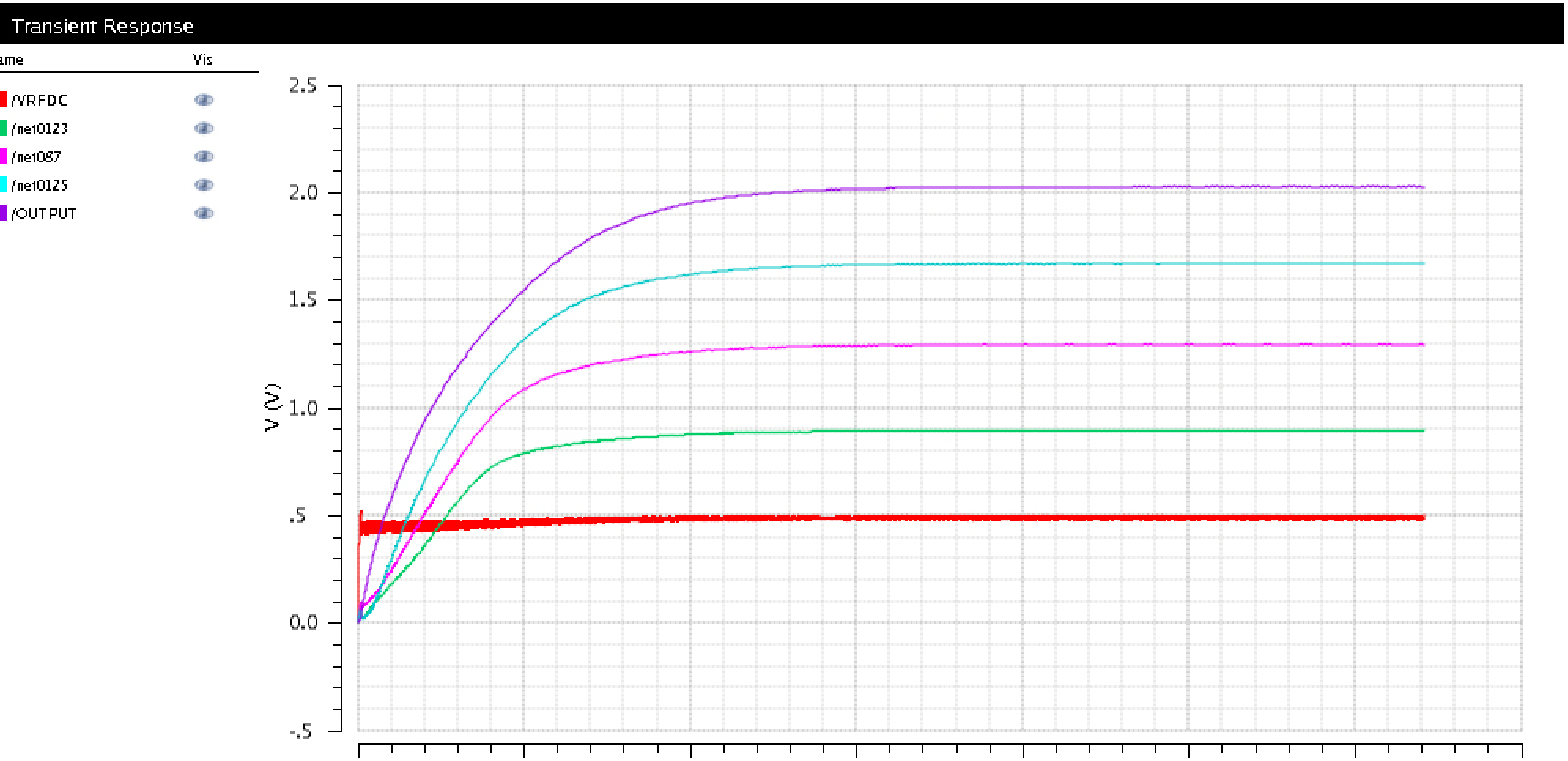

Summary with IC Design

- Low Vth NMOS based DCP gives 42.3% efficiency.
- FGCCR gives 55% efficiency.
- The overall system efficiency is 25%.
- Higher than efficiency reported in literature for RF-DC converter operating at -20dBm, 2.4GHz in 130nm technology.

Reference	This work	[5]	[7]	[6]
Power level	-20dBm	-25.7dBm	-22.6dB	-20dBm
Frequency	2.4GHz	2.45GHz	906MHz	2.4GHz
Efficiency	55%(simulated)	37%(measured)	10%(measured)	36%(simulated)
Rectifier	FGCCR in UMC 130nm CMOS	DCP in 0.5µm Silicon on Sapphire	DCP with floating gate transistors in 0.25µm CMOS	FGCCR in 130nm CMOS

6. ASIC Design for IOT

- Working on a 3-chip architecture
  - Our chip to enable sensing, and control functions
  - Communication using an external Monza chip
- Fabricated chip using commercial services!!

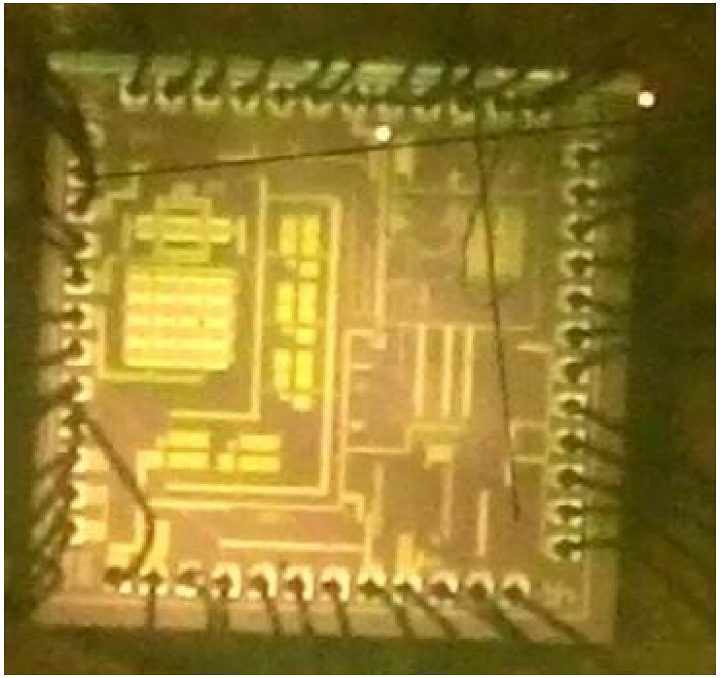

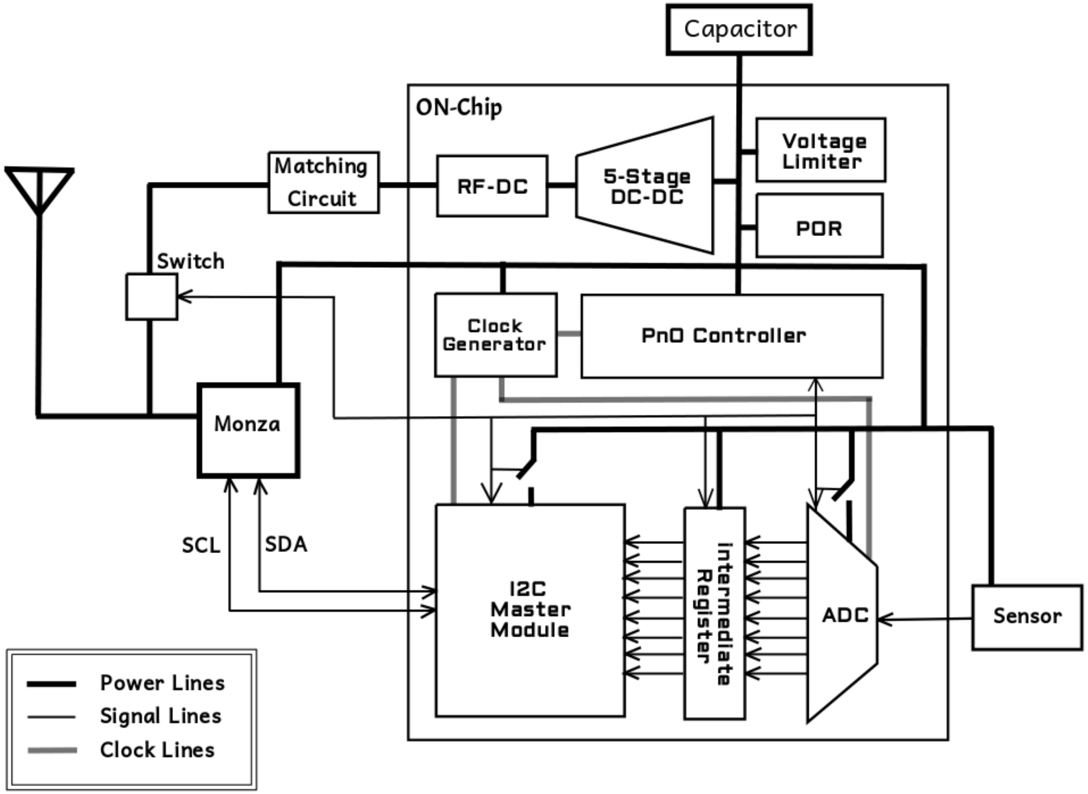
Capacitor
ON-Chip
Matching Circuit
RF-DC
5-Stage DC-DC
Voltage Limiter
POR
Switch
Clock Generator
PnO Controller
Monza
SCL
SDA
I2C Master Module
Intermediate Register
ADC
Sensor
Power Lines
Signal Lines
Clock Lines

Bharat Hegde & Syed Younis 2015-16

Battery-less Sensor node for BLE

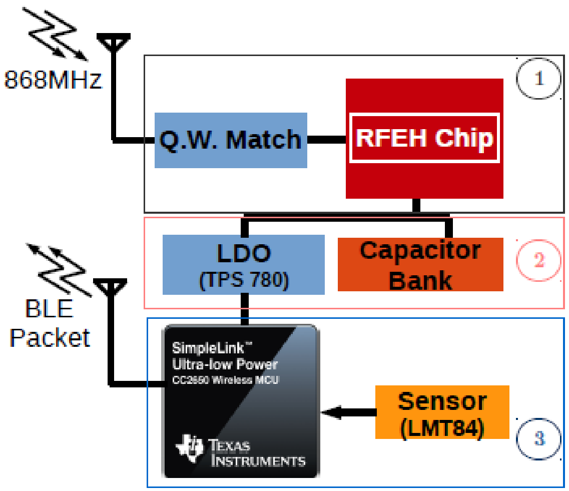
868MHz

Q.W. Match

RFEH Chip

①

LDO
(TPS 780)

Capacitor Bank

②

BLE Packet

SimpleLink™ Ultra-low Power CC2650 Wireless MCU

Texas Instruments

Sensor
(LMT84)

③

① Energy Harvesting Block

② Energy Storing Block

③ Sensing & Communication Block

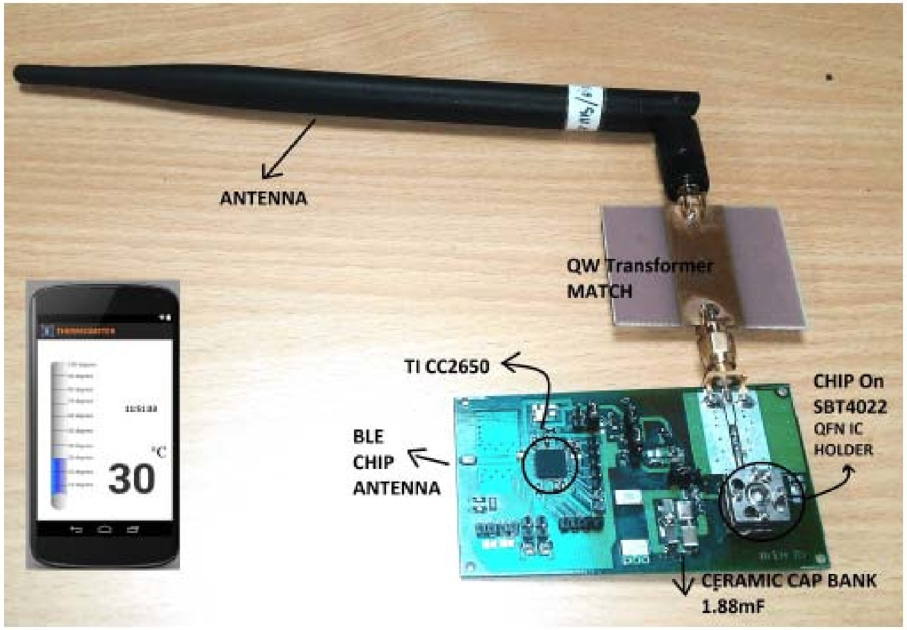
ANTENNA
QW Transformer MATCH
TI CC2650
BLE CHIP ANTENNA
CHIP On SBT4022 QFN IC HOLDER
CERAMIC CAP BANK 1.88mF
30 °C

Distance from Reader (m)	Input Power (dBm)	Cold Start Time (min)	Packet Time (min)	Efficiency (%)
3	-2	15	1.75	2.2
4	-5	35	7	1.7
5	-8	120	32	1

Eficiency(%) vs Load (5stage DC-DC)

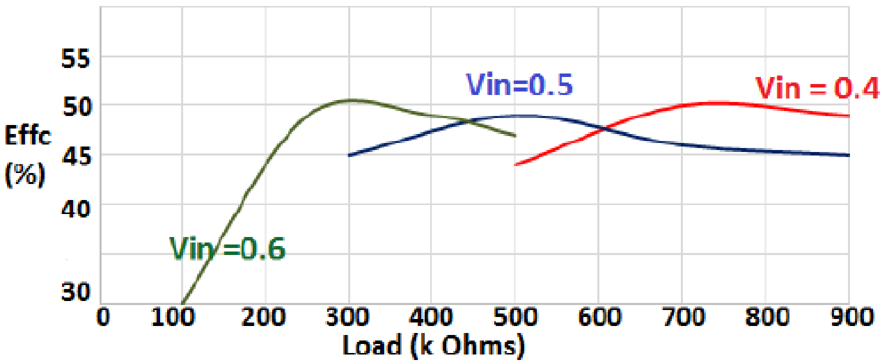

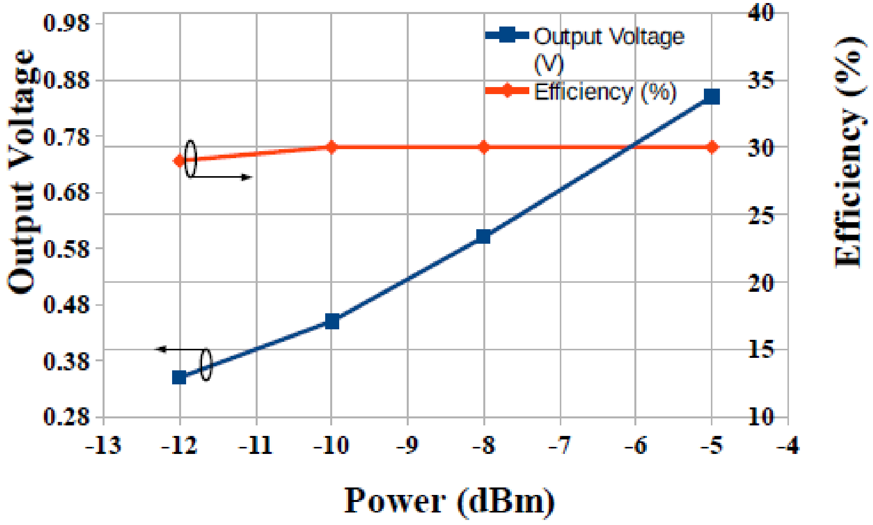

Bharat Hegde & Syed Younis 2015-16

Summary

- Most low power wireless terminals operate intermittently
- These require anywhere 50uW to about 10mW for their operation.
  - Batteries limited: cost, size, stored energy
  - Solar: not dependable through

- WPT and RF EH can enable wide use of IoT
  - Main challenges in the design is the low incident energy/power/voltage
  - High Quality factor components may help

- Several fabricated examples discussed here: All can transmit data to an aggregator wirelessly
  - Different standards implemented.

Acknowledgements

- Dr. TV Prabhakar, DESE, IISc
  - Gaurav Singh
  - Rahul P
  - Chaithanya C
  - Aditya Mitra
  - Prashuk Jain
  - Prashanth Raja
  - Nirmal John
  - Nithin Jose
  - Syed Younis
  - Bharat Hegde
  - Niharika Thakuria
- Prof Bharadwaj Amrutur
  - Uday S
  - NS Sreeram
- Others
  - Vivekanand M
  - Harikiran M
  - Manjunath M
  - Sanjeev K
  - Sandeep Rana
- Partial Funding From
  - ANRC (Boeing, Wipro, HCL)
  - Ricoh Research, India

ANRC

RICOH
imagine. change.

Thank YOU

Take the world
on line
Take control
of the world
Let the things
talk to each
other
Let thing
become
intelligen

Atmospheric Electrostatics

Lars Wåhlin
Colutron Research Corporation
Boulder, Colorado, USA

RESARCH STUDIES PRESS
Letchworth, Hertfordshire, England

JOHN WILEY & SONS INC.
New York, Chichester, Toronto,Brisbane, Singapore

Preface

Static electricity became a fashionable science in the early 1700 century and several investigators drew a parallel between the sparks produced in the laboratories to that of lightning and thunder produced during foul weather. It is surprising, however, that even today with our highly advanced technology in electronics and space science we still do not know what causes thunderclouds to charge. The purpose of this book is to give a general overview on atmospheric electricity and to discuss several proposed charging mechanisms including resent important discoveries in atmospheric electrochemistry. The aspects of atmospheric electrochemistry becomes important when we realize that the atmosphere, due to the constant bombardment of cosmic rays, is ionized and behaves very much like an electrolyte. Electrochemical potentials are produced on material surfaces that are exposed to our ionized atmosphere and are as common as contact potentials generated when dissimilar conducting materials touch each other.

This book is not a review of the most current publications on atmospheric electricity but serves as overview of the basic problems still at large and the purpose is to try to inspire new fresh blood into the oldest field of electricity. Two excellent textbooks are recommended for those interest in a detailed picture of the electrical structure of our atmosphere: H. Israel, Atmospheric Electricity Vol. 1, (1970), and Vol.2 1973. J.A. Chalmers, Atmospheric Electricity, (1957).

The author is thankful to the Burndy Library; The High Voltage Research Institute, Uppsala University, Sweden and the High Voltage

Laboratory, T.U. Munchen for supplying historic illustrations.

Thanks is also due to Dr. John Hughes for initiating this work and for editing and facilitating the publishing of this book.

Boulder, January 1985. Lars Wåhlin

CHAPTER 1

HISTORICAL BACKGROUND

1.1 PRIMITIVE BELIEFS

An electric storm is one of nature's most spectacular phenomena, and its display of lightning and thunder has fascinated and frightened man throughout time. In ancient times it was believed that the great gods were responsible for hurling thunderbolts that could kill, upturn boulders, split trees or kindle fires. Man's early steps towards civilization began when he learned how to control fires started by lightning and use it against predatory animals. Tales from American Indians explain how the world was a cold place before the first fire was started by the thunderbird god. As recently as the 1870's German soldiers were convinced that magic power from the **donnerkeil**, or lightning stone, would protect them from French bullets. Lightning stones, believed to be spearheads of lightning bolts, could be found buried where lightning had struck. The lightning stones and their fragments were sold throughout Europe for many hundreds of years and were thought to protect from illness and evil. We know now that most of the stones and fragments found were old relics and artifacts from the Stone Age (Lundquist 1969).

Today we smile at old superstitions, yet we too bend to the mystery of lightning and its atmospheric implications, for modern man has yet to understand it.

1.2 EARLY ELECTROSTATICS

The first person on record to have suggested a relationship between electricity and lightning was an Englishman named D. William Wall (1708). He noted a similarity between lightning and the crackling sparks produced by the rubbing of amber. S. Gray (1735) and A.G Rosenberg (1745) both mention the similarity between lightning phenomena and electric fire produced by electricity machines in the laboratory; and in a book published in Leipzig 1746, J.H Winkler describes several resemblances between lightning and electricity. During this time improved electricity machines and Leyden jars became readily available and a new era of electrical science was born. Many more scientists, among them Benjamin Franklin, also questioned the nature of lightning, recognizing its similarity to the snapping sparks produced in the laboratory. "How loud must be the crack of 10,000 acres of electrified cloud!" exclaimed Franklin. In a letter to Dr. John Mitchel of the Royal Society in England, he enclosed a treatise, "The Sameness of Lightning and Electricity". According to Mitchel, the paper was read by the Society amidst laughter from its professed experts on electricity. Another paper dated July 1750, was sent to the Royal Society through a friend, a Mr. Collinson. In this paper Franklin described how electric "fluid" is attracted to pointed conductors. "Might not the same principles be of use to man in teaching him to protect houses, churches, ships and other structures from damage occasioned by lightning?" he asked. Thus the idea of the lightning rod was born.

As yet no experimental tests had been performed to prove that lightning was an electrical phenomenon. Franklin therefore proposed an experiment to answer the question once and for all. On a high tower or steeple a sentry-box was to be erected large enough to contain a man and an electrically-insulating platform. A long pointed rod or antenna would be attached to the sentry-box by means of insulators and would be connected to the insulated platform inside (Fig. 1). A man standing on the insulated platform would thus become charged from the rod when a thundercloud passed overhead. At will he might then draw sparks from his fingertips to the surrounding grounded wall.

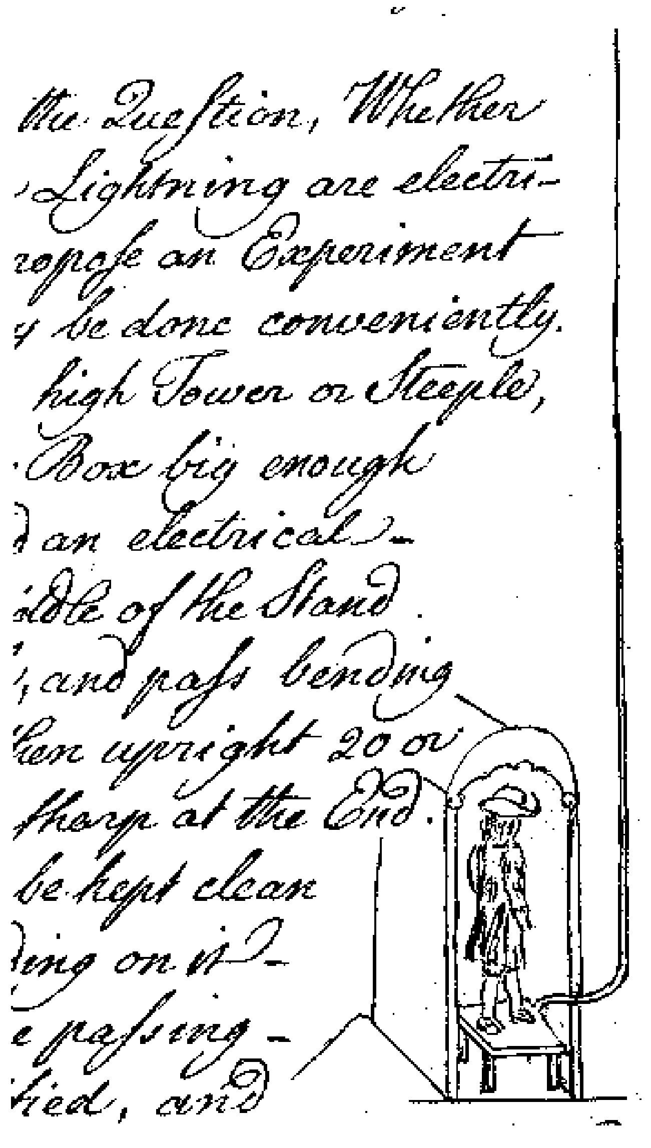
the Queſtion, Whether
Lightning are electri-
ropoſe an Experiment
y be done conveniently.
high Tower or Steeple,
Box big enough
d an electrical
ddle of the Stand
, and paſs bending
hen upright 20 or
ſharp at the End.
be kept clean
ding on it
e paſsing
fied, and

Fig. 1 Franklin’s proposed sentry-box experiment.

In principle it was just another lightning rod. From laboratory experiments Franklin knew that charge could be drawn from a nearby charged body by means of a pointed rod, which attracted electric "fire"

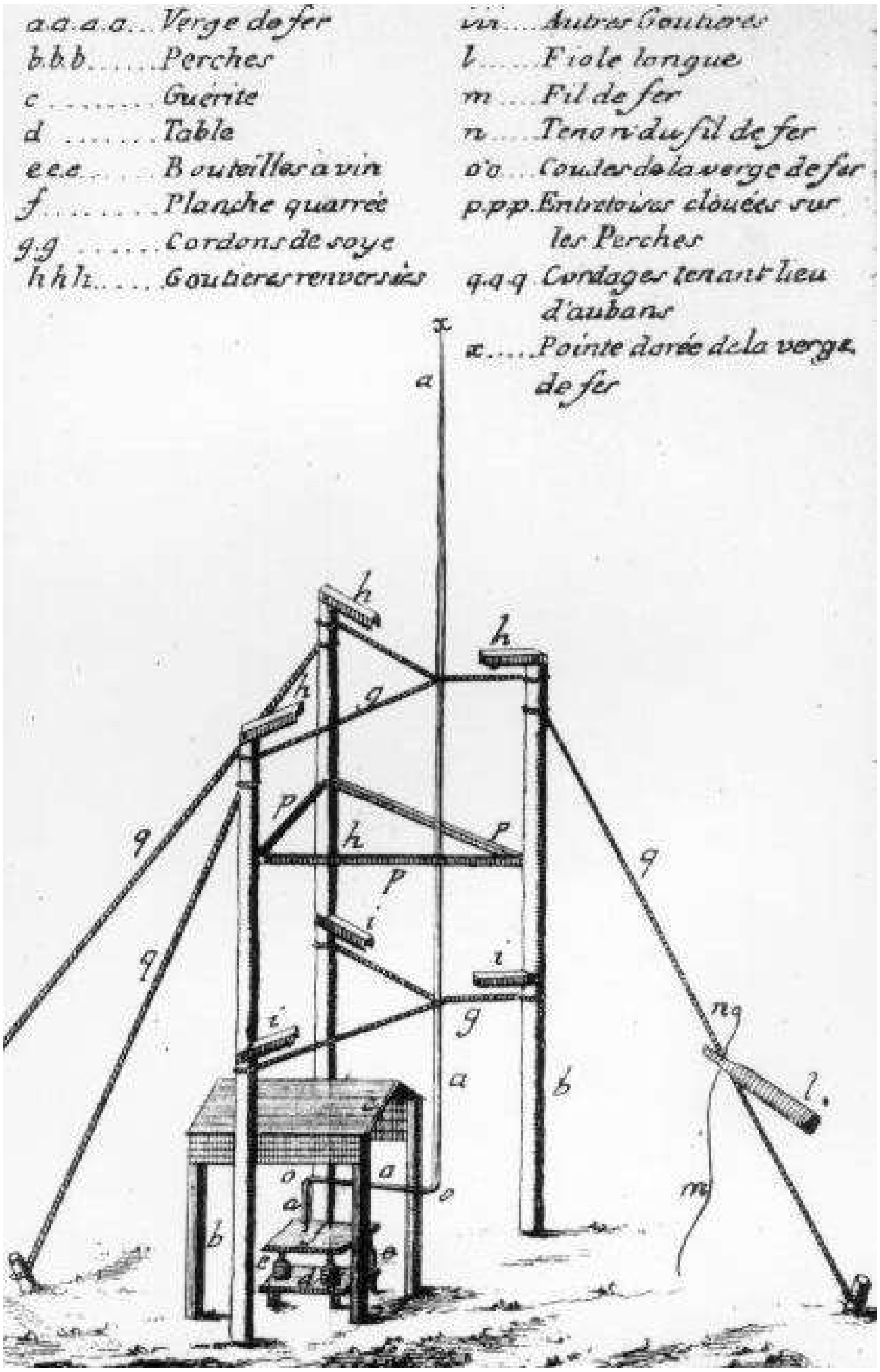
a.a.a.a... Verge de fer
b.b.b...... Perches
c .... Guérite
d .... Table
e.e.e...... Bouteilles à vin
f .... Planche quarrée
g.g ...... Cordons de soye
h.h.h...... Goutieres renversées
i.i.i.... Autres Goutieres
l..... Fiole longue
m ... Fil de fer
n..... Tenon du fil de fer
o.o ... Couder de la verge de fer
p.p.p. Entretoises clouées sur les Perches
q.q.q. Cordages tenant lieu d'aubans
x..... Pointe dorée de la verge de fer
x
a
h
h
h
g
p
p
h
p
i
i
g
i
a
b
o
a
a
b
e
d
e
c
q
q
q
q
n
l
m

Fig. 2 D'Alibard's version of Franklin's sentry-box experiment.

The experiment was never to be performed by Franklin, due to the lack of financial aid, yet with time Franklin's ideas gained increasing approval from the Royal Society. In 1751 several of his papers were published in England in book form and soon thereafter translated into

French by the naturalist D'Alibard. So intrigued was D'Alibard by the sentry-box experiment that he decided to put it to the test himself. An experimental structure slightly different in design was erected outside Paris at Marly. (Fig. 2). By the 10th of May, 1752 D'Alibard (1752) had successfully determined that thunderclouds are indeed electrically charged.

In America a few weeks later Franklin, unaware of D'Alibard's success, performed his famous kite experiment. It was a poor man's experiment, simple and brilliant. It demonstrated that lack of financial help is an insufficient deterrent to genius. (One wonders if Dr. Franklin was advised, when looking for financial support, to go fly a kite). In his kite experiment Franklin not only confirmed the electrical character of lightning but also, more importantly, found clouds to be negatively charged at the base and positively charged on top, thus forming giant electric dipoles floating around in our atmosphere.

1.3 FRANKLIN VERSA NOLLET

It is interesting to note that Franklin spoke of positive and negative electricity. This is a product of his own theory on electricity and the concept of positive and negative charge is still being used today. Franklin envisaged electricity as a kind of fluid where a body could have either an excess or a deficit of fluid. For example, if a person standing on an insulated stool were to touch the glass cylinder of an electricity machine he would lose some of the fluid. A deficit or loss of fluid represents a minus state and excess fluid equals a plus state. The sum of positive and negative electricity is always nil. The plus state of the fluid was determined by Franklin as follows: A positively charged pith ball would snuff out the flame on a candle i.e. the flame will retreat from the ball while the ball is trying to rid itself of excess fluid. A modern explanation is that the flame, which contains a great number of positive ions produced by the heat and combustion, will be repelled by any other positive charge in its vicinity. One consequence of Franklin's theory is the convention of labelling the direction of electric current. Although in metals the charge is carried by electrons flowing from minus to plus, we are now forced to assume that electric current or charge is flowing from plus to minus in order to keep things algebraically consistent.

The Abbè Nollet (1700-1770) of Paris was the authority on electricity in Europe during the time of Franklin's discoveries. Nollet was aware of two different kinds of electricity, namely glass electricity and resin electricity. He believed that the two different kinds of electricity were liquid streams flowing in and out of electrified bodies in opposite directions, effluence and affluence. Nollet maintained a very negative attitude towards Franklin's ideas on electricity and especially to the lightning rod. This is perhaps one reason why it took so long for the lightning rod to become established in Europe. The Abbè Nollet, however, was a master of experimental electricity. He often performed electricity demonstrations before Louis XV and his court. A typical demonstration would be to pass a high voltage static charge from a Leyden jar, through seven hundred monks holding hands, causing them to jump in perfect unison, to the King's delight. Nollet was often criticized for activities such as this. His theories on electricity were eventually abandoned in favour of Franklin's ideas which, to a certain extent, still are used today.

1.4 THE LIGHTNING ROD

A year after the famous tests of D'Alibard and Franklin a Russian professor named Richmann attempted to repeat the sentry-box experiment. The result is pictured in Fig. 3. Professor Richmann was killed instantly when lightning struck his antenna on the roof of the laboratory in St. Petersburg. As a result of this accident the general public grew skeptical of the lightning rod as a protective device.

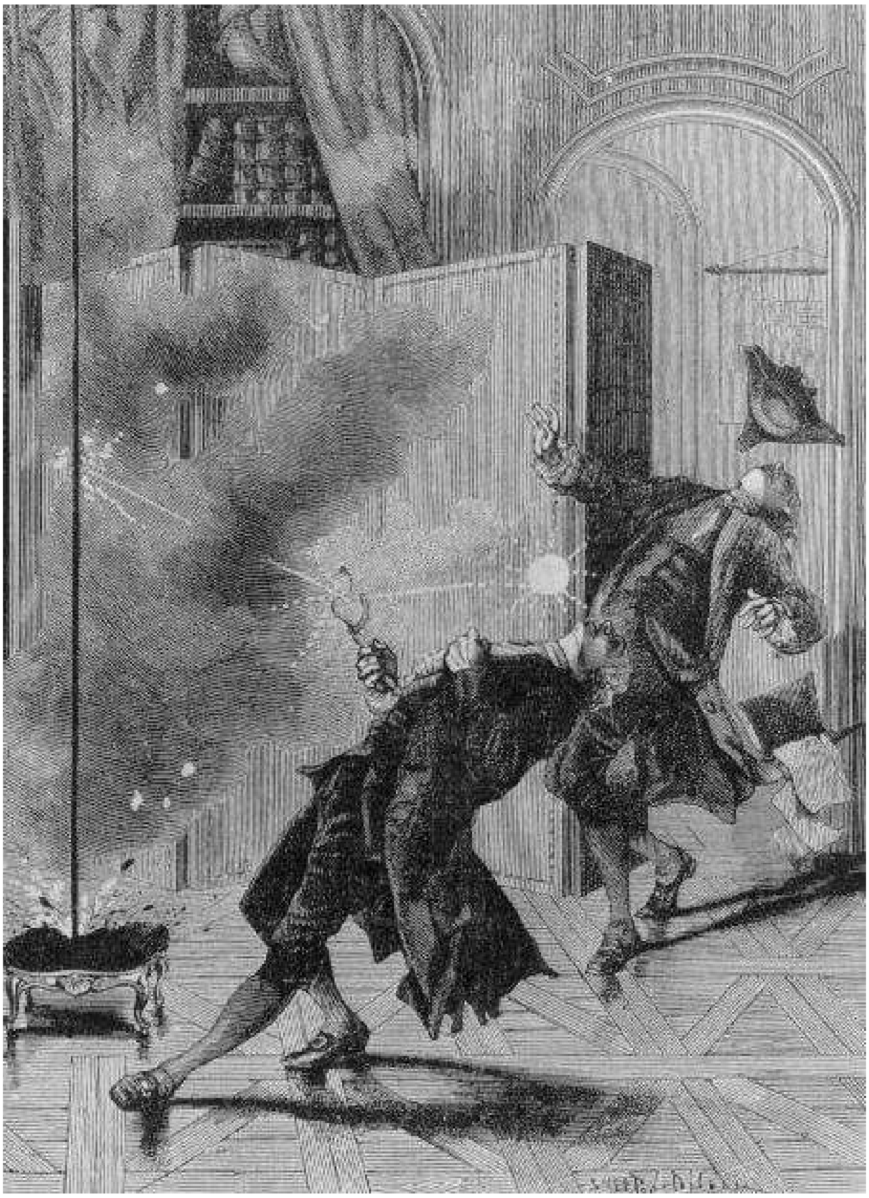

Fig. 3 Professor Richmann is killed when lightning strikes his experimental antenna in St. Petersburg, 1753.

Did not the death of Professor Richmann prove that lightning rods attract lightning? Some twenty years later Europeans finally accepted Franklin's invention, but the arguments against it were many. One such argument, advanced by a number of scientists spearheaded by Nollet, ran as follows:

If a large structure like a church steeple is not spared by lightning,

> how much larger must the lightning rod itself be to withstand the devastating power of a severe bolt?

The argument seemed valid at the time, for church steeples literally exploded when struck by lightning (Ohm's Law had not yet been formulated). The electric resistance of a church steeple is perhaps one megohm, and since the average lightning bolt carries about 25,000 amperes, it dissipates a peak power of $R \times I^2 = 6.25 \times 10^8$ megawatts. Although the duration of the lightning bolt is short, its total energy is considerable. The function of the lightning rod is, of course, to lower the resistance across the structure a million or more times and thus to permit the energy to dissipate into the ground.

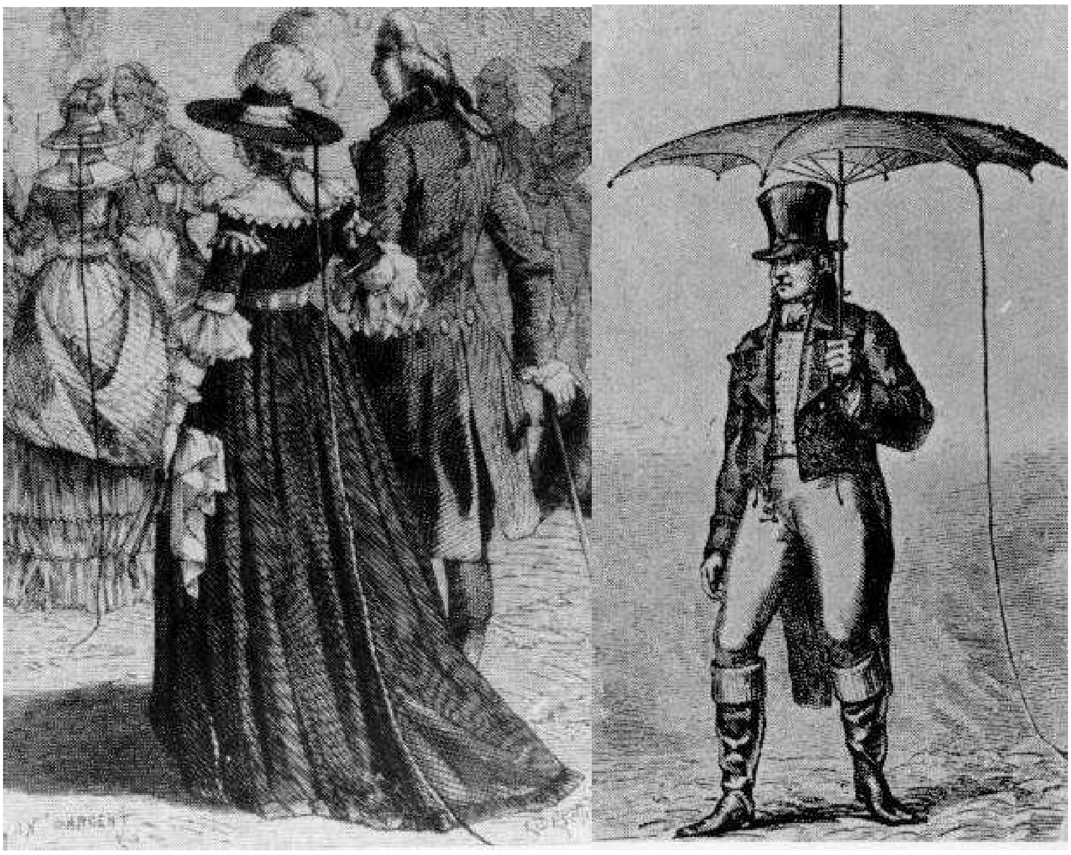

Fig. 4 Hat pin and umbrella lightning rods (Paris, 1778).

Eventually the lightning rod became a common sight on both sides of the Atlantic. Europeans, who had once strongly rejected the notion, embraced it with an enthusiasm that led to extravagant measures for protecting life and limb from the perils of lightning (Fig. 4).

Ironically, Europeans today are occasionally amused by the different shapes and designs of American lightning rods which they feel are costly and only of psychological importance, see Fig. 5. (Muller-Hillebrand 1963).

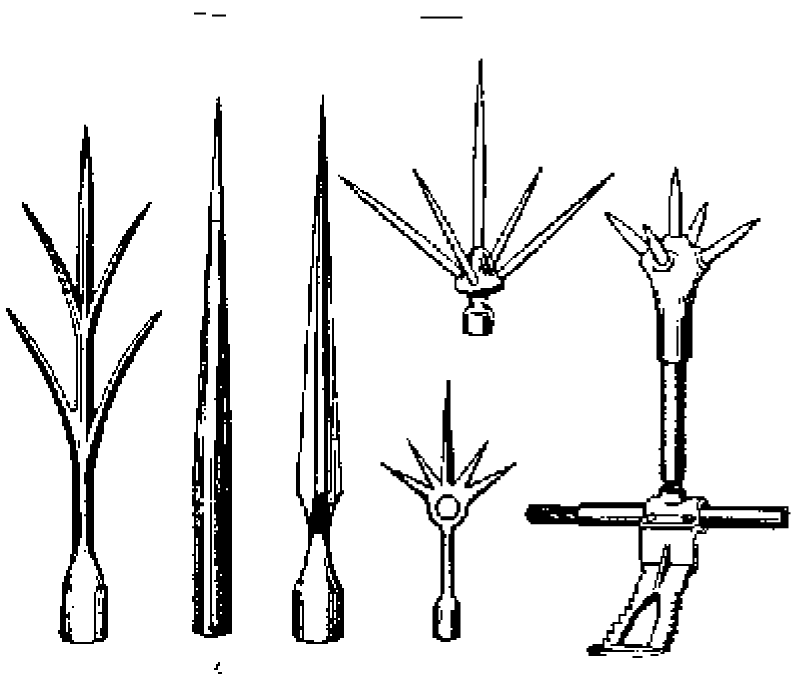

Fig. 5 Points according to American standards (1959) not cheap and only of psychological importance.

1.5 LATER DEVELOPMENTS

The D'Alibard-Franklin experiments were repeated by many investigators and most noteworthy is perhaps L. Lemonnier (1752) who, with his more sensitive apparatus, discovered that weak electrical charges could be detected in the atmosphere in the absence of clouds. He also noticed a difference in electric intensity during night and day. The discovery of Lemonnier is important because it gave birth to a new field of research in atmospheric physics, namely "Fairweather Electricity".

In 1775 the Italian scientist Beccaria (1775) confirmed Lemonnier's findings with one important addition, that pointed rods and insulated wires suspended in the atmosphere became positively charged relative to the earth's surface during fine weather, under cloud-free skies. This meant that an electric field is present in the atmosphere even when no thunderclouds are near. Beccaria also noticed a polarity reversal when a thundercloud passed overhead, indicating the introduction of negative charge from above. Thus he substantiated Franklin’s earlier finding that thundercloud are generally charged negatively at the bottom and positively at the top, an observation that can be readily confirmed today.

The first notable attempt to explain the origin of fairweather electricity was made by Lord Kelvin (1860). He was first to envisage the fairweather field as electric field lines which must stretch from the earth's surface and terminate on charges in the atmosphere. The direction of the field is towards the earth's surface. He theorized that the charged atmosphere forms a giant capacitor with the earth's surface. The capacitor model was later expanded by Kennely and Heavyside (1902) who suggested that the conducting layer in the atmosphere is the ionosphere which forms a spherical capacitor with the earth's surface where each "plate" of the capacitor carries a charge of nearly one million coulombs. A charge of this magnitude will, of course, produce an electric field of a few hundred volts per metre at the earth's surface. This is the normal value of the persisting fairweather field and Lord Kelvin, not knowing what process had charged the capacitor plates, at least seemed to have a reasonable idea explaining the structure of the fairweather field. Unfortunately for the attractive theory, later discovery of ions and cosmic rays proved the atmosphere to be electrically conducting. Lord Kelvin's capacitor model would therefore lose its charge in less than 10 minutes at a rate of nearly 2000 coulombs per second (=2000 amperes). It should be mentioned that C.A. Coulomb, the father of the electrostatic force law, discovered in 1785 that air is slightly conducting, an observation that was not understood at the time.

Where does the atmospheric fairweather charge come from? How is it replenished at the rate of 2000 coulombs per second? Many answers have been proposed, the most popular being that of C.T.R. Wilson 1n 1925. Wilson suggested that all thunderstorms around the world are electrical generators which by their violent discharges continusally supply electricity to the earth-ionosphere system. More recent evidence, however, weighs heavily against this evocative idea. Insufficient charge is available from worldwide thunderstorms to drive such a global electric circuit. Furthermore, data presented by Imyanitov and Chubarina, among others, demonstrate that annual variations of the fairweather field are not in phase with typical thunderstorm activity throughout the world for the same period. Despite the argument against Wilson's proposal and the lack of evidence to support it, many investigators still favour the idea since, until very recently, there has been no other explanation available. Now there is a new theory in competition with Wilson's concept based on the electrochemical properties of the atmosphere. This theory is discussed in Chapters 2 and 3.

Serious research on lightning and thunderclouds started late. With the birth of electric power in the early 20th century and the many power failures due to lightning, a better understanding of lightning and thundercloud charging was necessary. Charles Proteus Steinmetz, a German immigrant to America who worked as an engineer at General Electric Laboratories to develop lightning arrestors, might be considered as one of the early pioneers of modern lightning research. His work led to the construction of high-power high-voltage generators which could simulate lightning flashes. The machines consisted primarily of a large capacitor bank which was charged by a high voltage transformer via rectifiers or diodes. For the first time research could be performed on large electric discharges under controlled conditions. Much information was gathered by Steinmetz and his lightning machine which has helped us understand electric transients and has aided in the design of lightning protectors. The invention of the oscilloscope increased our knowledge on the lightning discharge because of its fast response to electrical transients. H. Norinder at the University of Uppsala High Voltage Research Institute was first to obtain such oscilliograms from lightning surges in 1925. Parallel with electric research during the first part of the 20th century was the development of high speed cameras which would record the optical properties of lightning. The first camera suitable for recording the rapid changes in lightning flashes was created by Sir Charles Boys who himself, unfortunately, did not obtain any satisfactory pictures of lightning flashes. It was not until 1933 that the Boys camera became a main contributor to what we now know about lightning discharges. It was B.F.J. Schonland and his team in South Africa who discovered the different sequences of a lightning flash with the aid of a Boys camera (Fig. 6) and revealed the initial process of the stepped leader. Boys' pictures show that a bright-tipped leader works its way down from the cloud in steps. When the leader gets near ground it is met by the main return stroke which carries the main discharge current through the ionized conducting path provided by the leader. Normally the stepped leader is invisible to the naked eye and only the main stroke can be seen. The return stroke therefore, appears to start from ground explaining the saying that lightning travels from ground up.

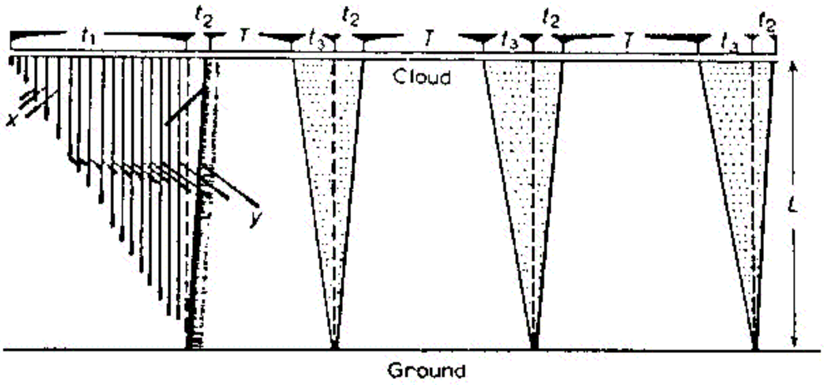

Fig. 6 Boys' camera photograph of cloud to ground discharge.

At the present time much is still unknown about atmospheric electricity and its origin. What mechanism is responsible for the build up of cloud charges and what generates the electric fairweather field? How is charge drained from a cloud or how does the lightning bolt connect up to all the myriads of charged drops in the cloud? These are still valid questions and, in the author's opinion, is the reason why Atmospheric Electricity is one of the most challenging fields in science today.

CHAPTER 2

THE ELECTRIC STRUCTURE OF THE ATMOSPHERE

2.1 IONS

The existence of ions in the atmosphere is the fundamental reason for atmospheric electricity. An absence of ions would mean zero electric field in the atmosphere and most probably no thunderstorms or lightning. The concept of positive and negative ions as charge carriers in the atmosphere was first put forward by J. Elster and H. Geitel (1899) in order to explain the electric conductivity of air. Much work has since been done on ions and their role in atmospheric electric phenomena. Today we know there are mainly three classes of ions, namely small ions, intermediate ions and large ions. Most important are the small ions since their higher mobility allows them to take a more active part in the transfer of charge throughout the atmosphere. The mobility of ions can be measured in metres per second per volt per metre which signifies the velocity that an ion will reach when subjected to an electric field of one volt per metre. For small ions the mobility is of the order of 0.0001 with a slight edge of the negative ion over its positive counterpart. In fact, the negative to positive mobility ratio of small ions is about 1.25 (Wåhlin 1985) which is a paradox since negative ions are believed to be more massive than positive ions. One explanation (Papoular, 1965) is that for part of its lifetime a negative ion is really an electron jumping from molecule to molecule. Molecules such as NO and NO_2 are believed to dominate the negative small ion population while oxonium and water might make up the positive small ions in the atmosphere. Their true molecular structure and mass are not well known because it is difficult to get spectroscopic mass analysis of small ions in the lower atmosphere. The problem is their relative short life time, about 100 seconds, which is much shorter than the transit time required for molecules or ions to reach the source end of a mass analyser.

The ionization in the lower atmosphere is mostly caused by cosmic rays and natural radioactivity. Ions are also produced in and near thunderclouds by lightning and corona processes. Cosmic rays originate from solar flares and other galactic objects such as supernovas and exploding stars. One interesting thought is; do stellar events affect our lives here on earth? We know that cosmic rays are by far the major ion producers in the lower atmosphere and if thunderstorms need ions to feed on in order to charge, we certainly would not have thunderstorms if there were no cosmic rays. Ancient man would not have had access to fires and the many thousands deaths each year from lightning strokes would have been avoided.

Cosmic rays originate from deep space and usually consist of very high-velocity atoms that have been stripped of their orbiting electrons. There are also electrons present in space that travel with near-light velocities, but such particles are usually absorbed at very high altitudes in the earth's atmosphere. However, heavy cosmic rays penetrate the atmosphere quite far and often reach the earth's surface. During such an encounter numerous secondary electrons are produced (electron showers) along its track from ionizing collisions with atmospheric molecules. The secondary electrons in turn might ionize a fair amount of molecules themselves before they slow down and attach themselves to atmospheric molecules to form negative ions. The result is that one cosmic particle could be responsible for the creation of as many as one billion ion pairs. Fig. 7 shows the rate of ion production by cosmic rays as a function of altitude (Wåhlin 1994). The average production

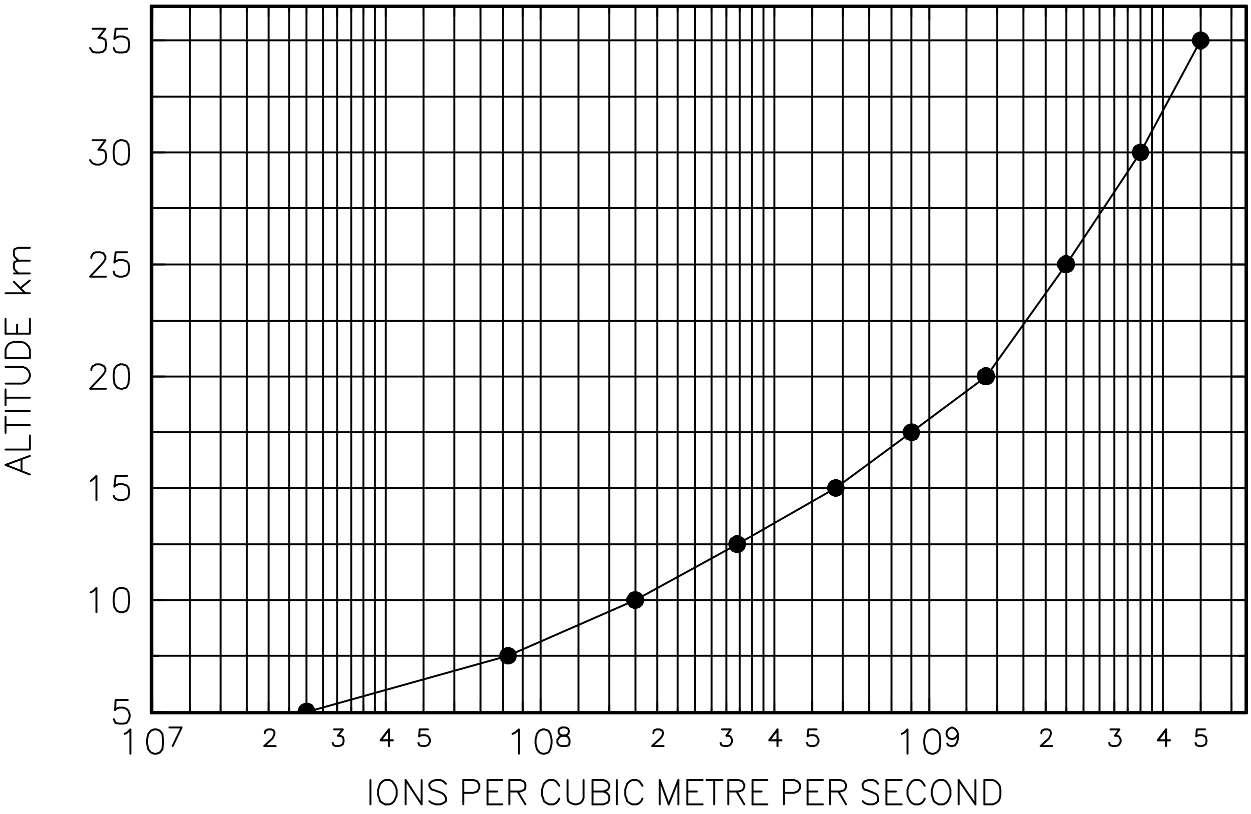

Fig. 7 Ion production as function of altitude.

rate at sea level is about ten million ion pairs per cubic metre per second. However, the average ion population at any given time is nearly one hundred times more, and ion mobility with altitude as air gets thinner has a drastic effect on the electric structure of the atmosphere. Typical conductivity data as a function of altitude are shown in Fig. 8. The conductivity is the inverse of specific resistance and is usually measured with a Gerdien cylinder (Gerdien 1905). The Gerdien apparatus consists of a cylinder with a coaxial mounted electrode (see Fig. 9). Air is drawn through the cylinder at a velocity of a few metres per second by means of a fan. The centre electrode is biased to a few volts via a sensitive electrometer. The amount of current registered on the electrometer relates to the amount of ions per unit volume of air. Care must be taken not to over-bias the inner electrode in order to avoid saturation currents. When properly calibrated, the Gerdien instrument can read both the positive and

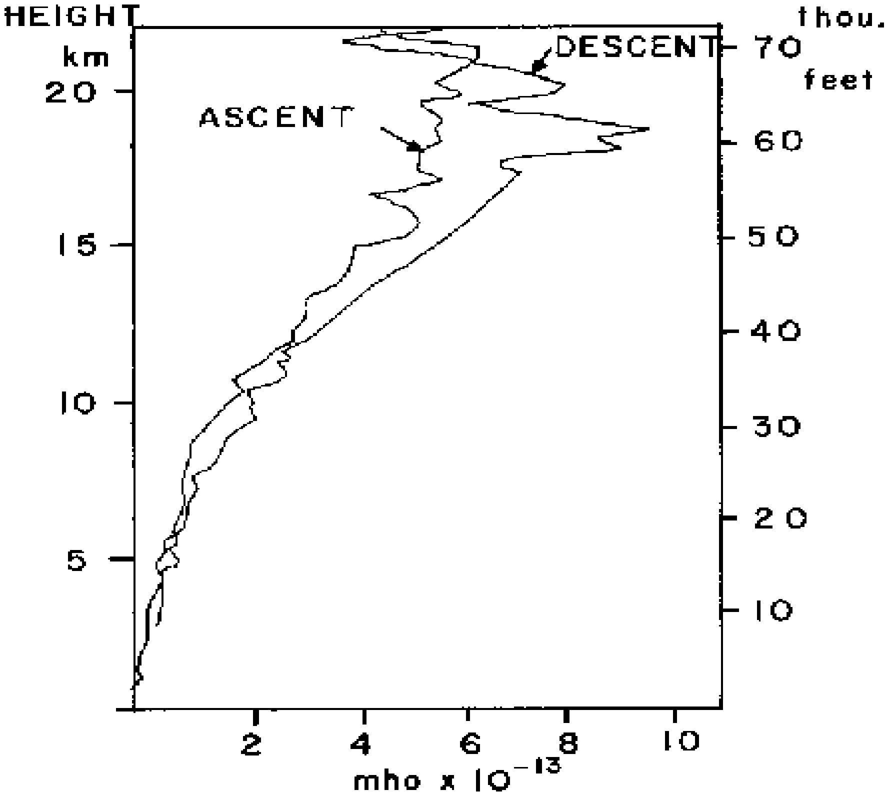

Fig. 8 Conductivity as a function of altitude (Rocket data).

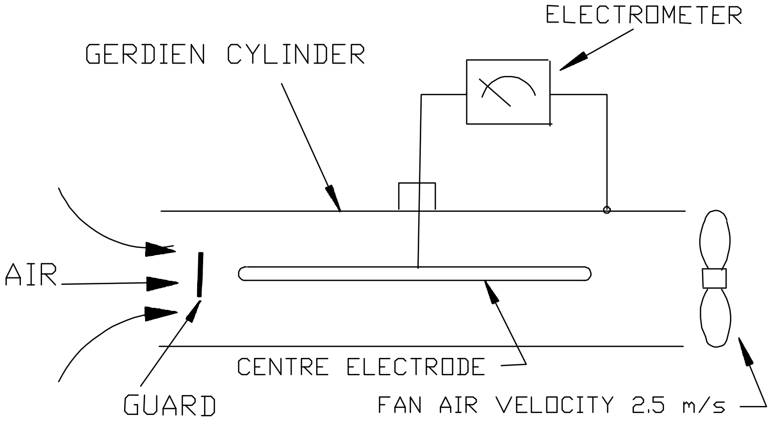
ELECTROMETER
GERDIEN CYLINDER
AIR
GUARD
CENTRE ELECTRODE
FAN AIR VELOCITY 2.5 m/s

Fig. 9 Gerdien conductivity apparatus.

negative conductivity of air depending on the polarity applied to the centre electrode. A typical ion current plot is shown in Fig. 10 where two identical Gerdien cylinders were tested, the only difference being that one was made of stainless steel and the other of aluminium. Three important features appear from the results shown in Fig. 10. First, the slopes of the two curves are different for negative and positive ion currents. This is due to the difference in ion mobility between negative and positive ions; therefore, the ratio of the slopes equals the ratio of the ion mobilities. Secondly, there is still a negative ion current going to the centre electrode when the electrode is at zero volt bias. Thirdly, the negative ion current at zero bias is higher for a Gerdien instrument made of aluminium than stainless steel. In order to obtain a zero ion current on the electrometer one needs to bias the stainless steel at 0.4 volts and the aluminium at about 1 volt. These results led to the discovery of electrochemical potentials in the atmosphere (Wåhlin 1971) which appear on material surfaces in contact with ionized air. The

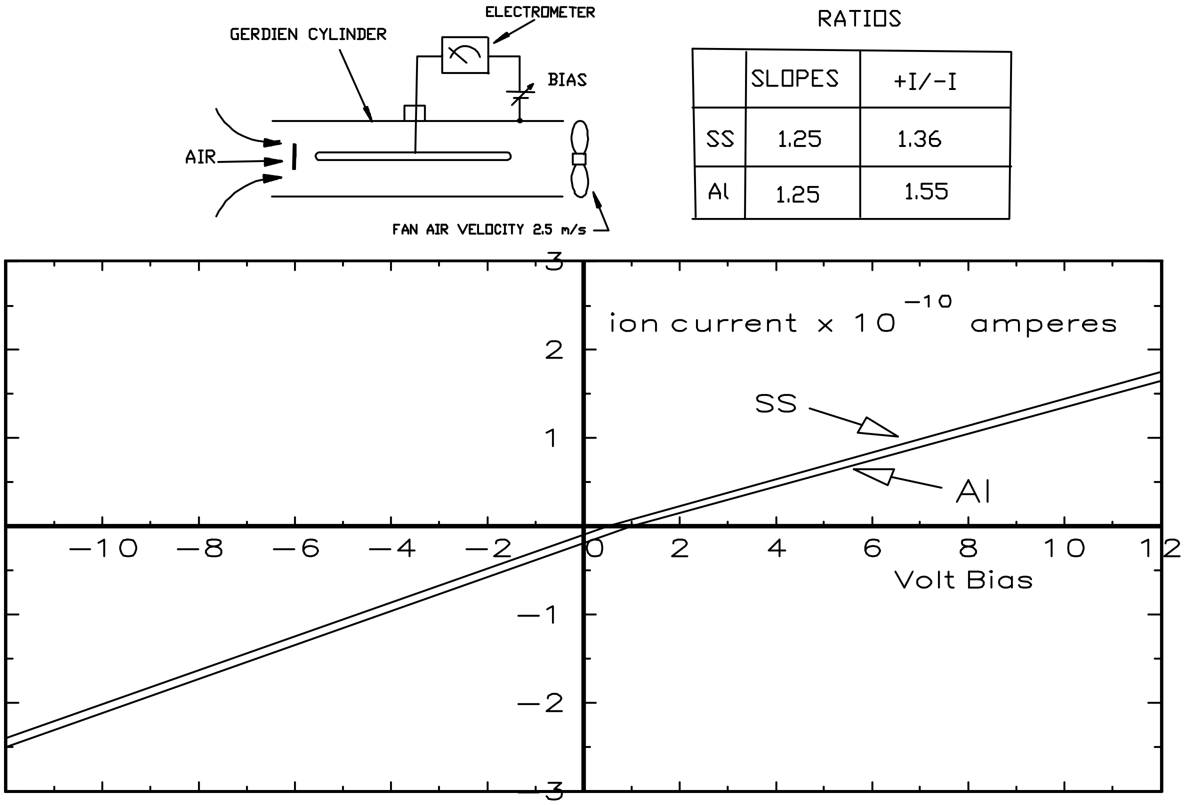

RATIOS

	SLOPES	+I/−I
SS	1.25	1.36
Al	1.25	1.55

Fig. 10 Typical ion current readings from a Gerdien cylinder.

reason why stainless steel and aluminium have to be biased at different positive potentials, to achieve zero ion current, is to cancel out

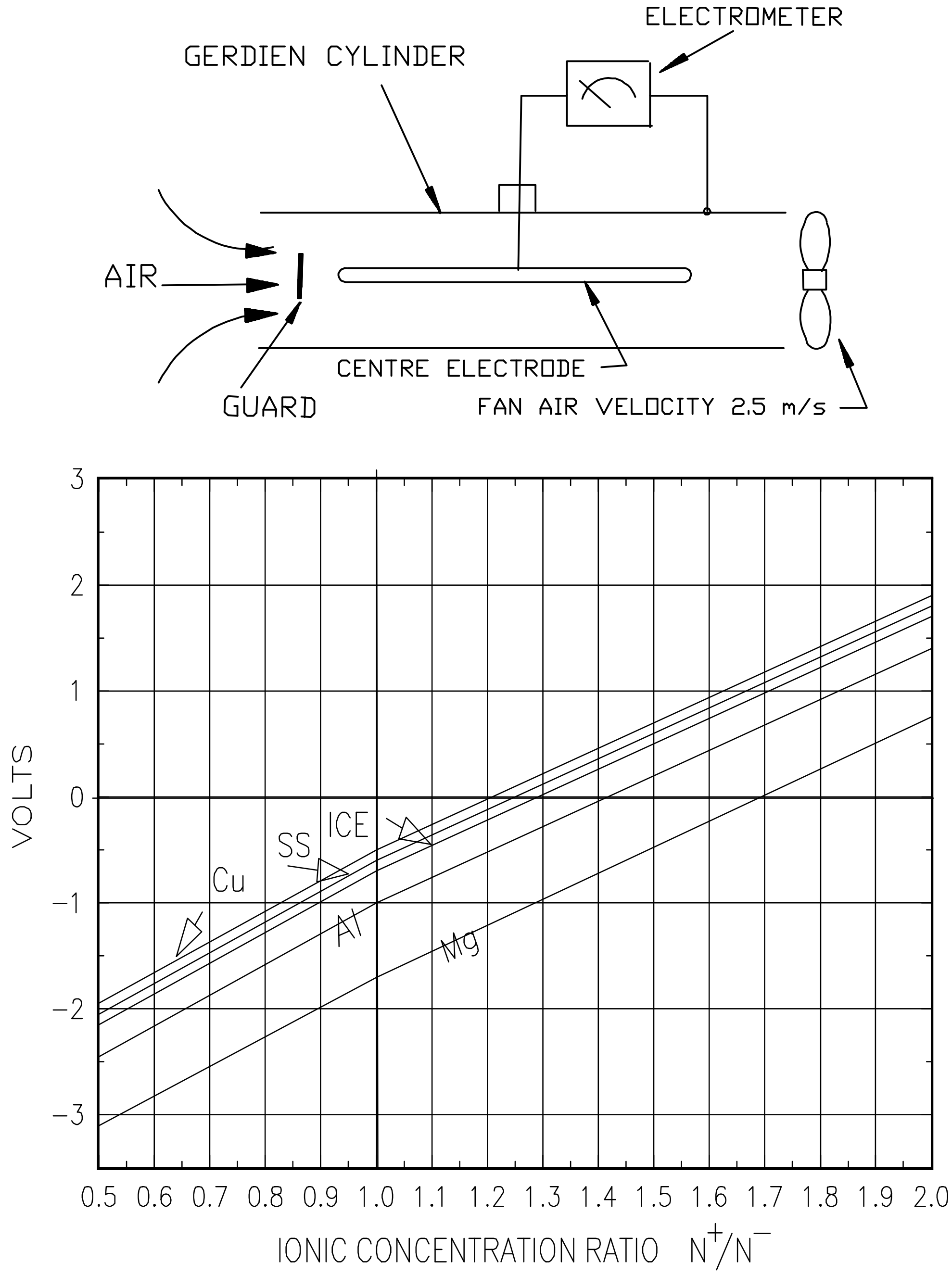

Fig. 11 Electrochemical equilibrium potentials for different materials as a function of ion concentration ratios.

the electrochemical or oxidaion-reduction potentials which are characteristic of each material and appear when exposed to an ionized environment. The effects of contact potentials are eliminated since both the inner and outer electrodes of the Gerdien instrument are made of the same material. The graphs in Fig. 11 show electrochemical potentials on different materials as a function of positive to negative ion concentration ratio. Electrochemical and contact potentials will be discussed further in Chapter 3.

2.2 THE FAIRWEATHER ELECTRIC FIELD

The fairweather electric field discovered by Lemonnier and Beccaria (see Chapter 1.2) is almost entirely due to the excess of positive ions over negative ions in the atmosphere.

The fairweather field is best understood if we assume that the earth's surface has absorbed a certain number of negative ions from the atmospheric ion pair population. It will create a slight excess of negative charge on the earth's surface with an equal excess of opposite charge in the form of positive ions left behind in the atmosphere. If we imagine that each captured charge on the earth's surface will produce

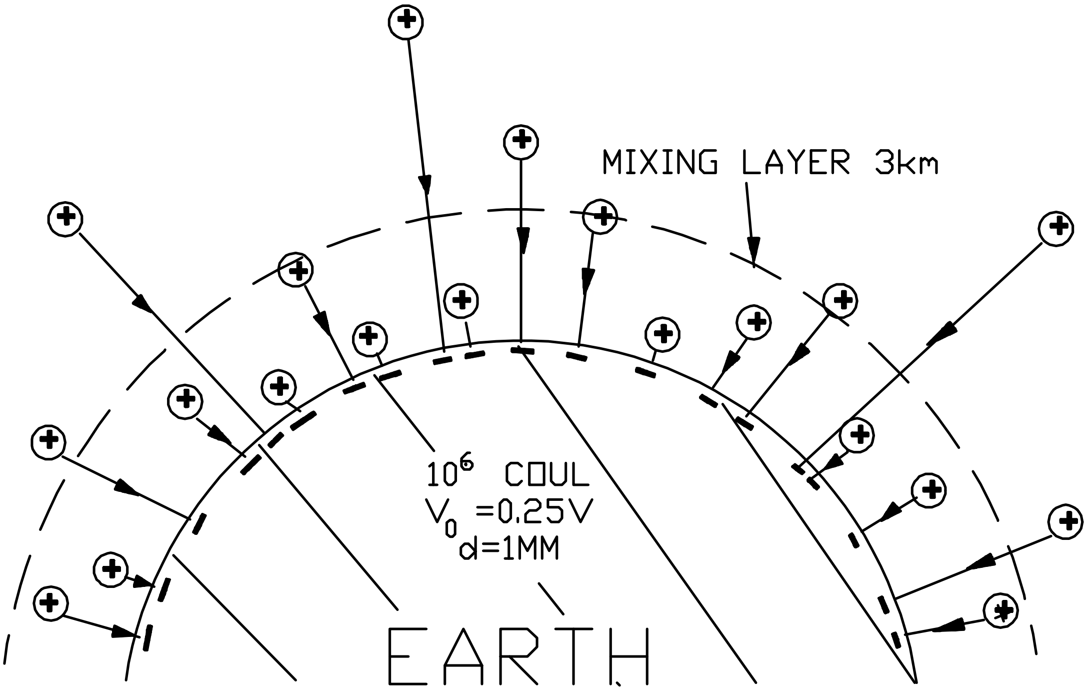

Fig. 12 The fairweather electric field in the atmosphere.

an electric field line which must terminate on a positive excess ion left behind in the atmosphere one obtains a fairly accurate picture of the electric fairweather field in the atmosphere, such as shown in Fig. 12. The excess positive ions are more or less uniformly mixed in the lower 3 km of the atmosphere which, to the meteorologist, is known as the "Austauch" or mixing region. The mixing is produced by convection and eddy-diffusion and the ionic distribution follows the mixing patterns of other constituents in the atmosphere such as radon for example. Radon is a radioactive gas emitted by the earth's surface and is constantly released into the atmosphere. Fig. 13 shows the vertical distribution profile of excess positive ions compared to that of the radon gas. The positive ion distribution is determined from electric field measurement at varying altitudes and by applying Poisson's equation. The radon profiles are obtained from airborne radioactive counters that detect the daughter products of the decaying radon gas.

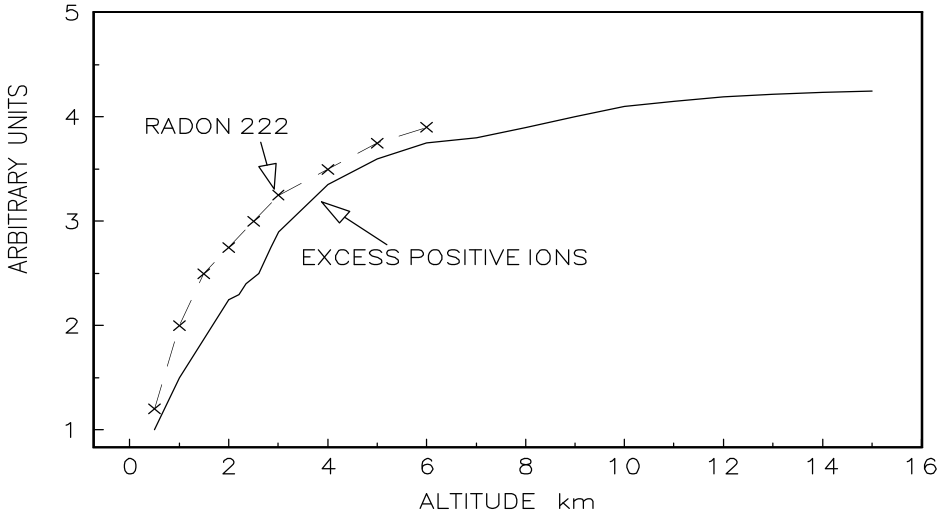

Fig. 13 Relative distribution of radon atoms and excess positive ions as a function of altitude.

The number of field lines per unit surface area produced by the positive charge or space charge above the earth's surface is also a measure of the electric field strength. Obviously the electric field strength reaches a maximum at the earth's surface since it contains the largest number of field lines per unit area. The average field strength at the earth's surface is on the average 100 volts per metre and decreases to less than 10 volts per metre at an altitude of 3 km. If one integrates the electric field as a function of altitude one obtains the total potential difference V at different heights. A typical value of V at 3 km is 200 kV with respect to the earth's surface. The total charge Q on the earth's surface is

$$Q = AE\varepsilon_0 , \tag{1}$$

where A is the surface area of the Earth, E the electric field strength at the surface and ε_0 the permittivity of free space ($\varepsilon_0 = 8.85 \times 10^{-12}$ Farad m^{-1}). The total energy of the fairweather field is

$$W = \tfrac{1}{2} VQ . \tag{2}$$

Fig. 14 shows the total electric energy, charge and potential in the atmosphere as a function of altitude. More than 90% of the energy is confined to an altitude below 3 km which together with the charge distribution curve in Fig. 13 seems to indicate that convection and eddy diffusion play a predominant part in the distribution of the fairweather electric field and that the bulk of its energy is distributed throughout the mixing region by the so called "Austauch Generator" (Kasemir 1950).

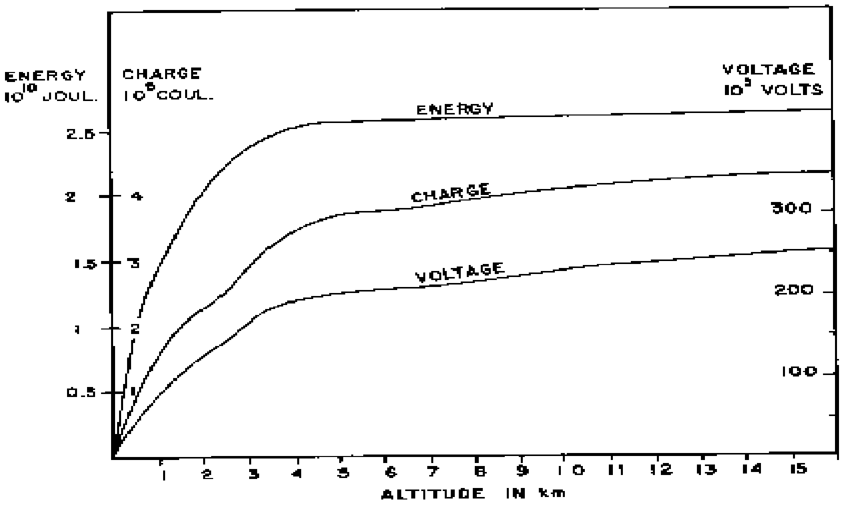

Fig. 14 Electric energy, charge and potential vs. altitude.

One crucial question still remains to be answered. What causes the positive space charge in the atmosphere and how is the opposite negative charge maintained on the earth's surface? As mentioned before there are two schools of thought on this one in which all thunderstorms around the world are believed to charge the earth-atmosphere system (Wilson 1929) and a more recent theory proposed by the author (1973) which considers the electrochemical effect as a charging mechanism where negative atmospheric ions are preferentially captured by the earth's surface leaving a space charge of positive ions behind in the surrounding atmosphere. Both theories might be supported by the evidence of a small systematic diurnal variation in the fairweather field, which is believed to be related to the world-wide atmospheric convection activity. The effect was first discovered in Lappland 1905 by Simpson whose findings were later augmented by Hoffmann (1923) and Mauchly (1923). The effect is

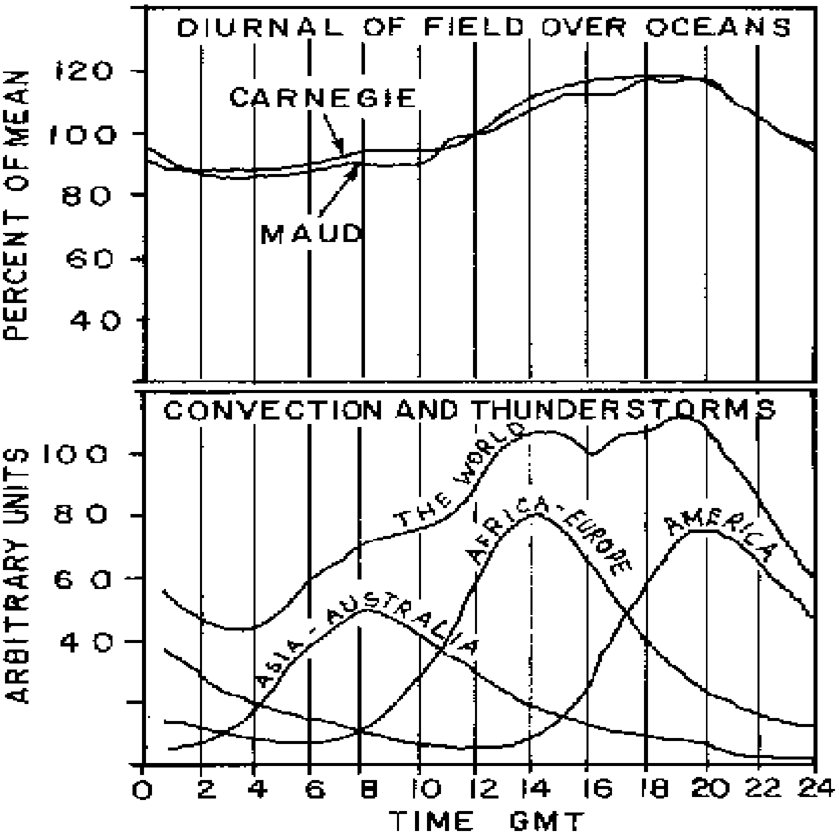

Fig. 15 Diurnal variations in the fairweather field compared to world-wide convection.

illustrated in Fig. 15 where the average variation in the world-wide potential gradient is compared to the estimated world-wide convection activity at different times of day (by GMT). The top graph shows the global variations in the electric field measured at sea in the absence of local disturbances such as pollution, fog, etc. The top graph seems to coincide with the lower graph which gives an estimate of the world-wide convection activity produced by the heat of the sun during a diurnal period. The steady convection over oceans, however, is thought to smooth out the electric field variations as is evident from the top graph. Before discussing the electrochemical and global thunderstorm circuits as possible generators of the fairweather field, it is necessary to examine the global leakage current and its implications.

2.3 THE AIR TO EARTH CURRENT

As already mentioned, the atmosphere is conducting and the earth's electric potential or field must cause a current to flow in the atmosphere. Since there is an excess of positive ions residing in the atmosphere and an opposite negative charge bound on the earth's surface, charge must flow to earth in the form of a positive ion current.

Direct measurements of electric currents in the atmosphere are difficult if not impossible. Therefore, ion current values at different altitudes are almost always computed indirectly from conductivity and electric field data by the use of Ohm's law. Direct current measurements can be made, however, at ground level by isolating a portion of the earth's surface and measure the charge collected over a given time. Several methods can be used (Wilson 1906, 1916, Simpson 1910, Mühleisen 1953 and Kasemir 1951) but in almost all cases the indirect current gives a value often twice as large as the direct method (Lutz 1939, Israel 1954). Whipple (1932) pointed out that the discrepancy in currents can be explained by the fact that there is always convection and eddy diffusion in the atmosphere which will mechanically move charges upwards in the atmosphere thus generating a mechanical or convection current in the opposite direction of the leakage current (the Austauch generator). As later explained, the question whether or not the convection and leakage current on the average are equal is crucial to the electrochemical charging theory and is a problem which has not yet been settled.

From direct current measurements it is possible to estimate the total fairweather current over the whole earth to be nearly 2000 amperes which corresponds to a current density of about 4×10^{-12} amperes per square metre. Other charge transfer mechanisms in the atmosphere of importance are point discharges, precipitation currents and lightning discharges.

2.4 POINT DISCHARGE CURRENTS

It is difficult to determine the total charge brought to the earth's surface by means of point discharge currents under electrified clouds. Wormell (1930) has made some estimates from the amount of charge brought down by a single point over a period of 4 years. He made a guess that the total point discharge current around the world brings negative charge to the surface at a rate of about 1500 amperes which would supply about 75% of the total fairweather leakage current. Other investigators give slightly lower values for the average point discharge current but not less than 25% of the fairweather current. The source of point discharge currents are the electrified clouds which of course also bring charge to ground by lightning. The point discharge current is, to a certain extent, canceled by the large amount of positive lightning flashes to ground and through positive charge reaching the earth's surface by precipitation.

2.5 PRECIPITATION CURRENTS

The electricity of precipitation has played an important role in atmospheric research due to the belief that charging of precipitation particles in some way must relate to whatever charging mechanism is active in clouds. Paradoxically, this is not always true because the final charge on a cloud drop is determined in the space between the cloud base and ground and is usually of opposite sign to the charge of the cloud base where it came from. This peculiar phenomenon is called the mirror-image effect and is demonstrated in Fig. 16 by the two curves which show the change in electric field strength and amount of precipitation charge reaching the earth's surface as a function of time.

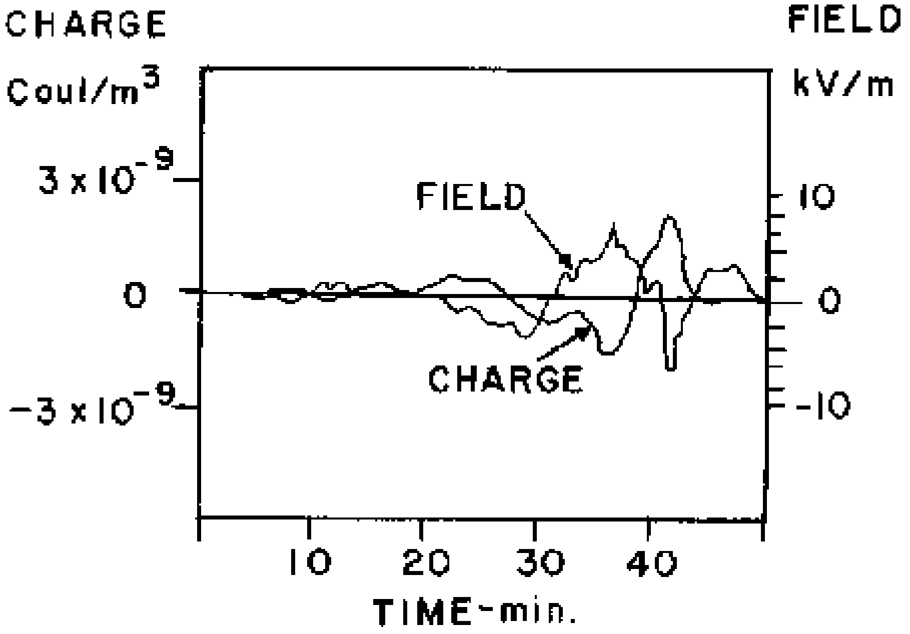

Fig. 16 The mirror-image effect.

One can easily see that when the electric field goes negative (negative charge in the cloud base) the precipitation current becomes positive and vice versa. As pointed out by Chalmers, a drop must take several minutes to fall from the cloud base to ground. Since the precipitation charge changes with the potential gradient below the cloud, it must mean that the drops also obtain their final charge below the cloud or very near ground. The electrochemical charging process can possibly explain the mirror-image effect if one assumes that the positive to negative ion concentration ratio near ground is affected by the strong electric field under the cloud. For example, a positive charge on the earth's surface, caused by a strong negative cloud charge above, would attract and remove part of the negative ion population near the surface. The result would be a higher than normal positive to negative ion concentration ratio at lower levels. When the positive to negative ion ratio exceeds 1.2 (see Fig. 11) it will produce a positive electrochemical potential on water drops falling through such a region as demonstrated by the Gerdien apparatus experiments in section 2.1. On the other hand, a positive cloud charge above would reverse the effect because drops now fall through an environment containing a higher negative to positive ion concentration ratio which will generate negative electrochemical charges on their surfaces. Other explanations of the mirror-image effect take the Wilson charging mechanism into consideration. This charging mechanism is based on the idea that rain drops become electrically polarized when immersed in an electric field such as under an electrified cloud. A negative cloud charge above will induce a positive charge on the top surface of a drop and the bottom surface will acquire a negative charge induced by the positive charge on the earth's surface. The total net charge on the drop, however, would remain zero. As the drop falls through the ionized region below a cloud it would preferentially sweep up positive ions by its negatively-charged bottom. Calculations, however, show that the Wilson mechanism is too feeble to account for the amounts of charge normally collected by drops (the Wilson charging mechanism is discussed further in Chapter 3). In contrast to rain, precipitation currents carried to ground by snow are usually always negative under potential gradients between ±800 V/m (Chalmers 1956). The total precipitation current around the earth is estimated to be about +340 amperes.

2.6 LIGHTNING CURRENTS

The charge brought to earth by lightning is estimated to average –340 amperes which would cancel the precipitation current. It must be remembered that a mean current of –340 amperes represents the excess of negative charge over positive charge reaching ground by lightning and that the ratio of negative to positive ground strokes equals about 10:1. The average current in a negative lightning stroke to ground is about 25,000 amperes but the total charge averages only 25 coulomb. Positive ground strokes usually carry as much as 10 times more charge and current than do negative strokes although they are outnumbered by 10:1. The ratio of negative to positive ground strokes seems to vary with global location.

It is believed that about 2,000 thunderstorms are active at one time around the earth which amounts to a total number of 50,000 thunderstorms per day.

2.7.1 THE ELECTRIC BUDGET

Where does the energy of nearly 200 million watts come from that is required to maintain the earth-atmosphere electric fairweather field? Are thunderstorms generating the fairweather field by leaking off positive charge from cloud tops to the conducting ionosphere and by bringing negative charge to earth in the form of negative ground strokes and point discharge currents? Or is the electric charge on the earth's surface maintained by the electrochemical charging mechanism in close collaboration with convection and eddy diffusion? These are some of the basic questions that are still in need of answers. Both mechanisms are, in the author's opinion, certainly capable of supplying enough charge and energy to the earth-atmosphere system, but new ideas and more sophisticated measuring techniques are needed in order to find the right answers.

2.7.2 THE GLOBAL THUNDERSTORM CIRCUIT

The concept that all thunderstorms around the world generate charge to the earth-ionosphere system was first suggested by C.T.R. Wilson in 1920. The diagram in Fig. 17 shows the global thunderstorm system as interpreted by Wilson. Electric field lines extend from the top of a cloud to the highly conducting upper layers of our atmosphere (50km and higher). Charge is presumed to leak along the field lines from the top of the cloud to the surrounding ionosphere. Note that field lines also go from the top of the cloud to ground thus leaking positive charge to ground. Negative charge is brought to ground mainly by lightning strokes and point discharge currents. The number of field lines between the cloud top and ionosphere compared to the number between cloud top and ground is an unanswered but crucial question which will determine the amount of charge supplied to the global fairweather circuit. For example, if the number of field lines going from the top of the cloud to earth would equal the number of field lines returning from earth to the bottom of the cloud, no current can flow to the ionosphere.

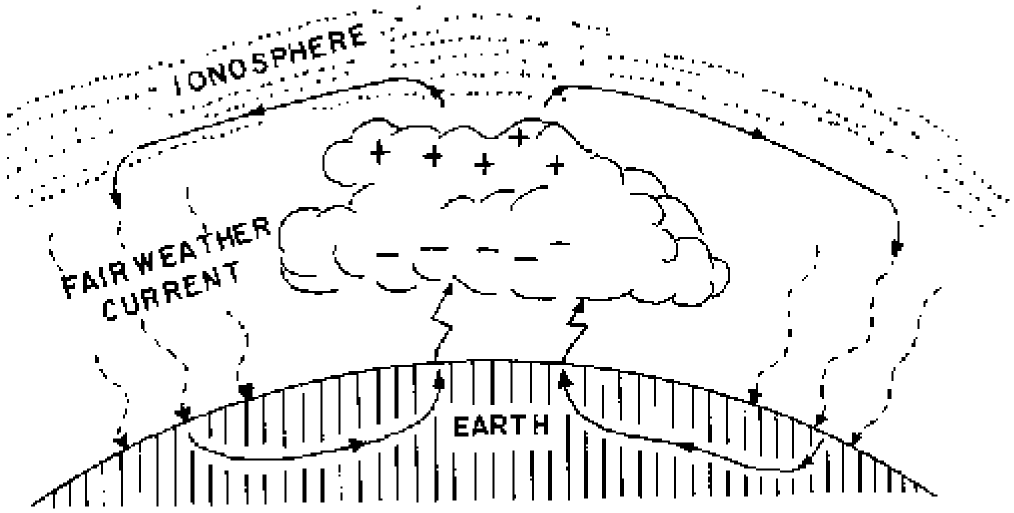

Fig. 17 The global electric circuit by C.T.R. Wilson.

Another problem presents itself when one examines the charge distribution of the fairweather field. Figs. 12 and 14 illustrate that about 90% of the fairweather field and charge is confined within an altitude of 2 km which is far below the conducting ionosphere. The situation is usually explained as follows: consider two conducting surfaces such as the ionosphere and the earth's surface carrying opposite charges at a potential difference of several hundred kilovolts, the earth's surface being negative and the ionosphere positive (see Fig 17). A current driven by thunderstorm generators is flowing in the form of negative ions towards the ionosphere and positive ions towards the earth's surface. Since conductivity and ion mobility increase with altitude, it is believed that negative ions (which flow upwards) will disappear faster on the positive electrode, the ionosphere, than positive ions can disappear on the earth's surface. Positive ions which face an increase in resistance are believed to slow down and congregate in a space charge cloud near the earth's surface. This is believed to produce an excess of positive ions near the earth's surface and could

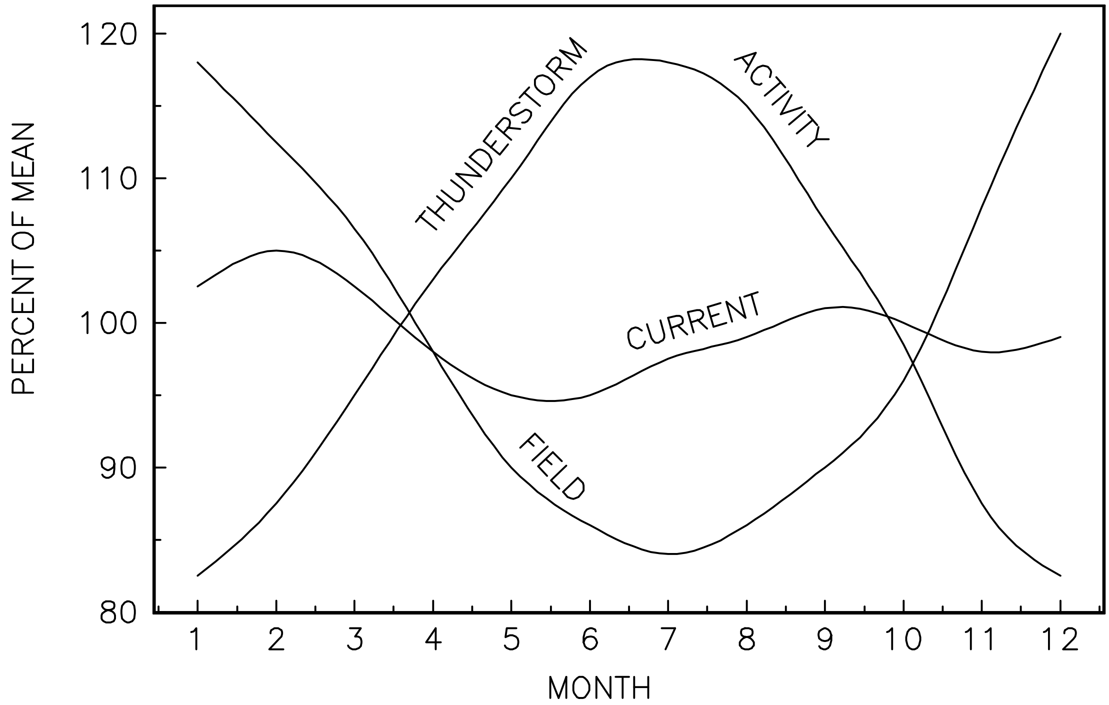

Fig. 18 Variation in fairweather field and current compared to thunderstorm activity.

explain the observed positive space charge distribution of the fairweather electric field. Plasma physics does not allow ions of one kind to disappear faster on one of the electrodes because it would mean that one of the electrodes or conductors in question would carry more current than the other, which is impossible. Theory and experiments require that an opposite and equal amount of space charge must build up near the other electrode as well (see Papoular 1965). In the case of the fairweather electric field, such a negative space charge near the ionosphere or near tops of thunderstorm generators, have never been found.

Data presented by Imyanitov and Chubarina (1967) provides little support for the closed circuit idea since they show that annual variations in the fairweather field are not in phase with typical thunderstorm activity throughout the world for the same period (see Fig. 18). Furthermore, Kasemir has pointed out that the curve showing diurnal variations of the fairweather field in Fig. 15 is much too smooth to fit the world-wide thunderstorm activity data, because recent satellite data show that thunderstorms are mainly active over continental land masses rather than over oceans (Turman, 1978, Turman and Edgar, 1982). The smooth curve in Fig. 15 might be more representative of the world-wide atmospheric convection and follows the field variations of the Austauch generator, a situation which would favour the electrochemical charging theory.

2.7.3 THE ELECTROCHEMICAL MECHANISM

The electrochemical charging mechanism considers the earth as an electrode immersed in a weak gaseous electrolyte, the naturally ionized atmosphere. The earth's surface will adsorb negative ions and achieve an electrode potential of about -0.25 volts which will appear at the earth-atmosphere interface in form of an electric double layer. This is analogous to a battery cell where an electrode becomes charged relative to an electrolyte. The thickness of the electric double layer is very small, about 1mm in the normally ionized atmosphere. The double layer can be pictured as containing field lines which connect each captured negative ion on the earth's surface with an excess positive ion left behind in the atmosphere. As convection and eddy diffusion lift the excess positive ions to higher elevations the field lines stretch thus increasing the potential with altitude. The result is an electric field build-up which will follow a pattern completely dictated by the mixing and diffusion mechanisms in the atmosphere such as shown in Figs. 12 and 13. The field strength at the earth's surface will equal the electrochemical potential divided by the double layer thickness $(0.25\ \text{volts}/\text{mm} = 250\ V/\text{m})$ which is in close agreement with measurements. With a few exceptions (Griffiths and Vonnegut (1975), Moore and Vonnegut (1977) and Willet (1980)) no serious criticism has yet been directed against the electrochemical charging mechanism perhaps because it is relatively new theory. The above investigators believe that contact potentials rather than electrochemical potentials are responsible for the results reported by the author and that such a charging mechanism is too insignificant to play any major part in atmospheric electricity.

CHAPTER 3

CHARGING MECHANISMS

3.1 SUMMARY

There are several basic charging processes proposed which in one way or another might contribute to atmospheric electric phenomena. Some of the more important are contact electrification, electrochemical charging, influence charging, diffusion charging and mechanisms involving freezing and splinting of ice particles. Most of the above charging mechanisms were devised to explain the charging of thunder clouds. Some theories, such as those involving influence charging for example, will not operate in the earlier stages of cloud growth, and others, which involve freezing and ice, cannot be considered dominating since warm tropical thunderstorms exist which do not contain ice. There are numerous other theories which will not be mentioned here and there are even theories which argue that a combination of all charging mechanisms might be at play at one and the same time. The situation is very challenging. The latest theory to be proposed is that of the electrochemical process put forward by the author. The electrochemical charging mechanism has the advantage of being able to explain both thunderstorm charging and fairweather electricity. So far it is the only charging mechanism that can be readily demonstrated by working laboratory models. Contact electrification involves mechanical contact between solids where electrons from a lower work function material spill over to a higher work function material. Contact potentials are of the order of a few tenths of a volt and might occur in the atmosphere when solid precipitation particles of different temperatures collide or when solid precipitation particles bounce off material surfaces either on the ground or in the atmosphere.

Any process in which charge is captured or transferred by ions is by definition an electrochemical process. Charge transfer by ions can be referred to as oxidation-reduction reactions. An ion which gains an electron is reduced and an ion that loses an electron is oxidized. Electrochemical potentials are encountered in everyday life and can be found in batteries and dry cells for example, and has been known to chemists and included in their text books for centuries. One problem, however, is that the language and conventions used by chemists are not exactly tailor-made for physicists who therefore, in the author's opinion, seem to shun the electrochemical effect and often confuse it with contact electrification. An attempt will be made later to explain the difference between contact potentials and electrochemical potentials as seen by a non-chemist.

Influence charging deals with charges that appear on material surfaces which are exposed to an electric field. For example, dust resting on the earth's surface will be negatively charged during normal fairweather conditions since the electric field lines from the positive space charge in the atmosphere above must terminate on the negatively charged earth's surface (the surface in this example being covered with dust). During strong winds the negative charged dust particles can become airborne and form highly electrified dust clouds. Dust and sandstorms are most often found to be negatively charged. The same explanation can be applied to waterfall electricity where the negative charge, induced by the fairweather field on a surface of water, is being carried over the edge of a waterfall. As the electrified water falls over the edge it breaks up into small drops and forms a mist of negative space charge, referred to as the Lenard effect or waterfall electricity, first discovered by Tralles of Bern in 1786. Induction charging in thunder clouds has been considered by several investigators and is based on the idea that cloud drops, which are polarized while subject to the electric fairweather field, preferentially capture negative charge from smaller drops. As large drops fall and collide with smaller drops a transfer of charge is believed to occur where the upper half or negative pole of a smaller drop gives up its charge during the encounter. Negative charge collected in this manner would descend and occupy the lower portion of a cloud while the smaller drops with excess positive charge would remain behind to form an upper positive region.

3.2 CONTACT CHARGING

There are many excellent books and papers on contact potentials or Volta potentials, such as Loeb (1958) and Lord Kelvin's famous paper presented at the Bakerian lecture to the Royal Society (1898). But ever since Voltas original observations there has been confusion between contact potentials and electrochemical potentials, a situation which persists even today. For example, Lord Kelvin became very upset when Professor Lodge presented his paper "On the Seat of the Electromotive Force in a Voltaic Cell" (1885) and later in a private letter to Lord Kelvin he expressed his belief that contact potentials are related to the difference in oxidation energies of different materials. Loeb in his book warns, "Volta potentials must never be confused with electrolytic potentials". With due respect to both Kelvin and Loeb (the latter was a colleague of the author's) the author believes that Professor Lodge was also right.

Kelvin's condenser system which is shown in Fig. 19 demonstrates how volta or contact potentials are measured. It consists of two large capacitor plates made of dissimilar metals (copper and zinc) which are separated by an adjustable air gap. A switch and an electrometer are connected in parallel with the capacitor plates. The plates are connected to an electrometer. At first the switch k is momentarily closed which will allow electrons to flow from the zinc plate to the copper plate because zinc has a lower work function than copper. An electric field with a total potential of V_c, which is the difference in work function between the two metals, will appear between the two surfaces. When the spacing between the plates is increased from a to b the field lines will extend and the potential across the electrometer will increase to $V = V_c(a+b)/a$. In the above example the effects of fringing fields and stray capacitance were neglected. The diagram in Fig. 20 illustrates the relationship between work functions and contact potentials. Fig. 20a shows a typical potential well diagram for a metal

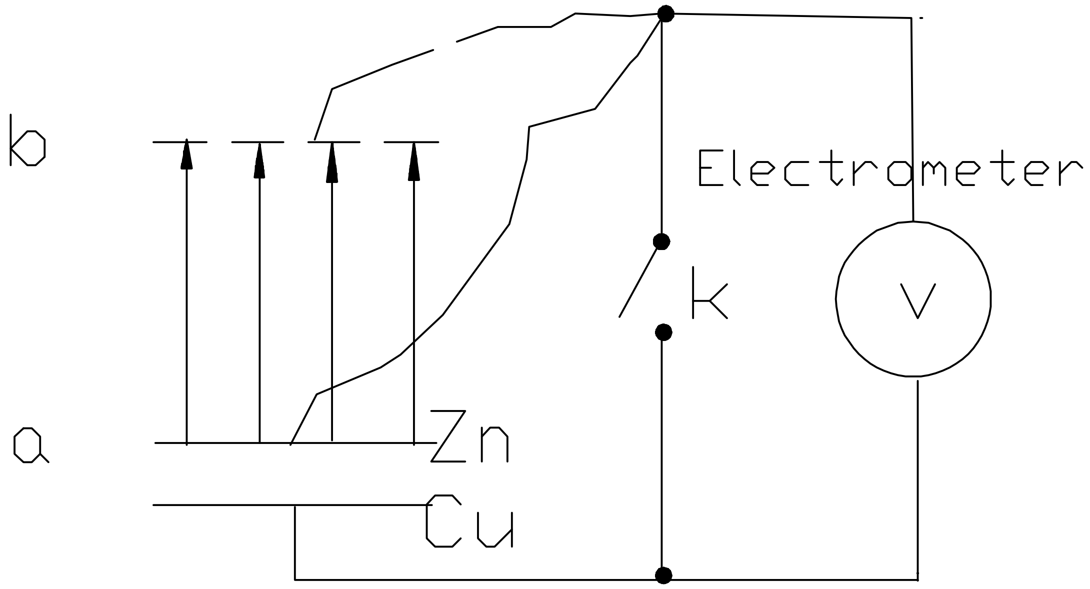

Fig. 19 Lord Kelvin's condenser system.

surface where the conduction and valence electrons are trapped at an energy level which is equal to the work function of the material in question. The potential barrier set up by the work function is represented in Figs. 20b and 20c by small electric cells between the conduction band of the material and its outermost surface. At first, before the switch k has been closed, there is no field between the zinc and copper surfaces. On closing the switch electrons will spill over from the zinc, having the lower workfunction to the

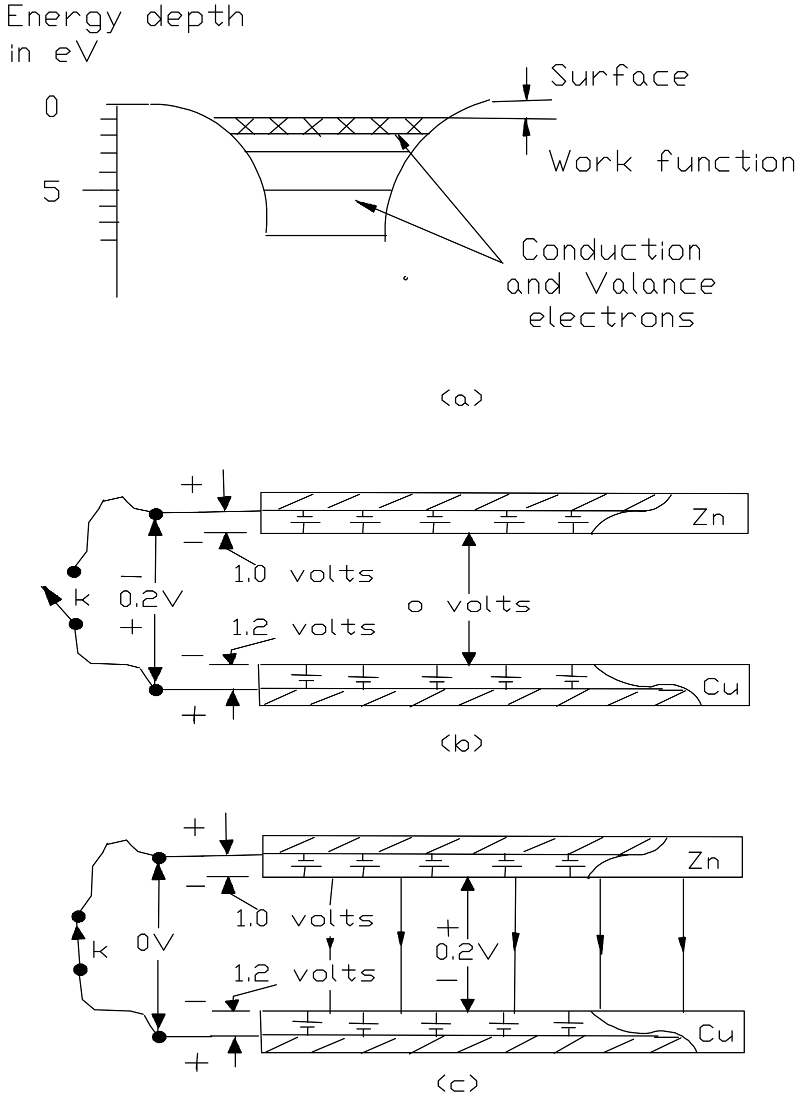
Energy depth
in eV
0
5
Surface
Work function
Conduction
and Valance
electrons
(a)
+
−
1.0 volts
k
0.2V
1.2 volts
o volts
Zn
Cu
(b)
+
−
1.0 volts
k
0V
1.2 volts
0.2V
Zn
Cu
(c)

Fig. 20 Contact potential between zinc and copper. (a) Potential well diagram of a metal surface. (b) Electric field before contact is made. (c) Electric field after contact is made.

higher or deeper potential well of the copper. Bringing the conduction bands of both materials in electric contact will cause the difference in work function potential to appear across the gap of the plates. The amount of charge transferred is $V_c C$, where C is the capacitance of the plates. One characteristic feature of contact charging is that the current ceases as soon as the capacitance C is fully charged; i.e. contact potentials are not current driving sources such as galvanic cells, for example. The total amount of energy dissipated in the contact charging process is

$$W = \tfrac{1}{2}V^2C \tag{3}$$

It is interesting to note that in his lecture to the Royal Society, Kelvin mentioned one experiment in which he placed a drop of cold water between the plates in his condenser apparatus and found the electrometer swing toward the opposite direction to that of the contact potential, but with the same magnitude. He attributed this effect to electrolytical conduction and not to contact potential charging. Experiments by Maclean and Goto in Glasgow in 1890 proved that zinc and copper, with fumes from flames passing up between them, gave, when connected to an electrometer, deviations in the same direction, as if cold water had been in place of the flame. Kelvin also mentioned that beside the wonderful agency in fumes from flames, there were reports by other investigators that ultraviolet light and x-rays traversing the gap between the plates caused the same effect as that of cold water. These effects, Kelvin thought, would to some degree fulfill Professor Lodge's idea of some potentially oxidizing process, but, "each one fails wholly or partially to maintain electric force or voltaic potential difference in the space between them". Further communication between Kelvin and Professor Lodge broke down when Lodge, in a letter, said that Kelvin was unrepentant.

Are contact potentials related to the chemical properties of the materials in question or are they purely a physical phenomenon? To clarify this problem one first needs to ask; where does the energy come from that causes contact potentials to build up? The contact potential which equals the difference in work function between two materials is also equal to the binding energy difference between the electrons in the materials. For example, the electrons in copper are more tightly bound to the atoms in the surface lattice than the electrons in zinc. Therefore, it will require more work to remove an electron from the copper surface than from the zinc surface, thus the name "work function". The chemical binding forces between the metal atoms can be pictured as hooks bonding the atoms together in all directions. However, at the surface there will be dangling bonds because the outermost atoms will have nothing to attach to outside the surface boundary. These dangling bonds make up for the surface energy or work function potentials which are simulated by the electric cells in Fig. 20 b and c. Since the energy of dangling bonds are chemical in nature then any transfer of charge, due to this energy, could technically be classified as an electrochemical process.

What will happen if the space between the plates in Fig 20c is filled with an equal amount of positive and negative ions either in form of a liquid or an ionized gas such as air? Will the negative ions go to the positive plate and the positive ions to the negative plate of the condenser and will there be a continuous current flowing in the circuit as long as the ion supply lasts? The answer is yes. But does the energy that drives the current in the circuit come from the contact potential (potential difference in work function between the materials)? The answer is no. First, the energy available from the contact potential is too minute (see Equation (3)) and cannot sustain a current for very long. It is therefore very clear that the energy driving the current must come from the ions themselves as they interact and recombine with the surfaces. Recombination here means the neutralization of an ion as it loses its charge to a material surface. The electric field between the plates, set up by the contact potential, see Fig. 21a, will draw negative ions to the zinc and positive ions to the copper. Once negative ions enter the zinc, charge is brought to the surface in the form of electrons.

These new electrons will immediately spill over from the zinc to the deeper potential well of the copper work function. However, no constant current can flow unless other electrons are allowed to leave the copper electrode at the same rate new electrons spill over from the zinc; i.e. the rate of negative charge brought to the zinc by negative ions must equal the rate of negative charge leaving the copper to combine with the positive ions drawn to the copper plate.

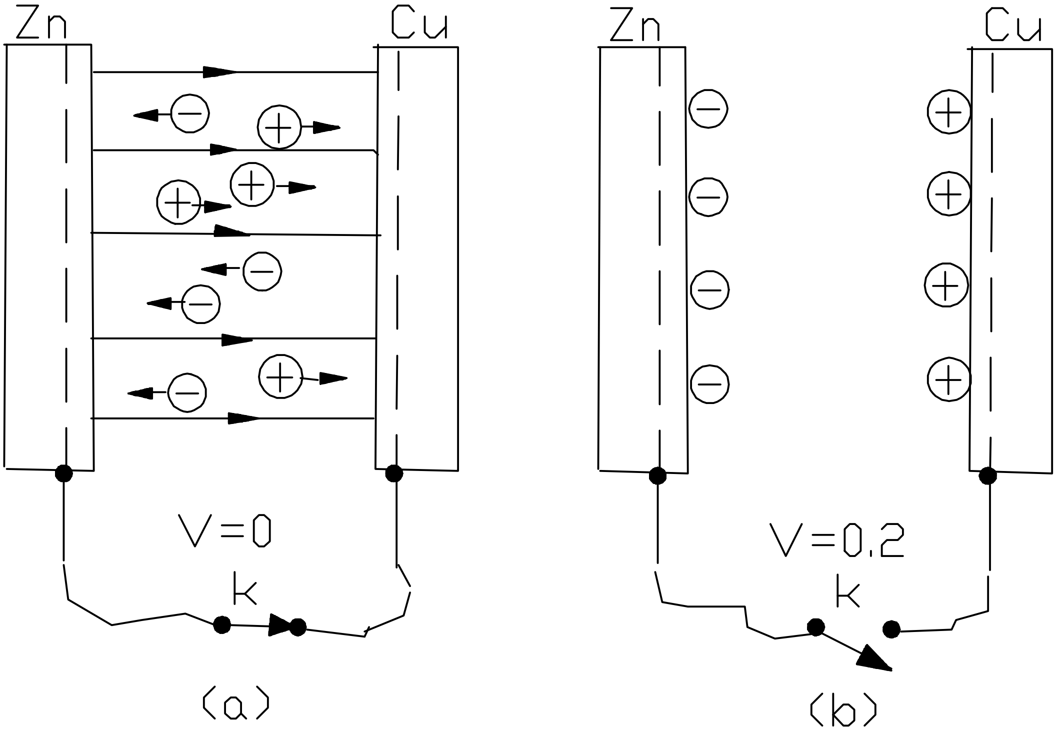

Fig. 21 Contact potentials in an ionized environment. (a) Ion current with switch k closed. (b) Potential build-up with switch k open.

What is the source of energy that is capable of lifting the electrons back out of the potential well of the copper work function? It is obviously not the work function or contact potential itself because that would be analogous to lifting oneself by the hair. The energy supplied must come from the ions residing in the spacing between the plates. The ionization or recombination energies of the ions will provide the necessary energy and a current will flow as long as the supply of ions lasts unless switch k in Fig. 21a is opened. If the current is interrupted by the switch, ions will continue to flow and charge the electrodes until a back-emf is built up that cancels the original field of the contact potential across the plates, at which point the flow of ions ceases. A potential equal to the contact potential will now appear across the switch and a situation such as shown in Fig. 21b will be reached.

The presence of ions between two electrodes of dissimilar materials and their ability to generate a steady current is nothing less than a galvanic cell. The primary source of energy is the ionization agency which might be radioactivity or cosmic rays as in the case of our ionized atmosphere. Ions are also produced in liquids where one or both electrode materials might slowly go into solution in the form of ions. The dissolved ions supply the energy that drives galvanic currents through circuits and usually at the cost of the lower work function material itself. A typical example is the flashlight battery or dry cell. Most charging processes involving ions are electrochemical processes.

3.3 ELECTROCHEMICAL CHARGING

Electrochemistry has played a major part in both industry and science for the last two hundred years. It might come as a surprise, however, to discover that the function of the familiar dry cell, such as used in portable radios, is not yet perfectly understood. The chemical reactions taking place are believed to be as follows: zinc metal from the outside casing is dissolved by the acidic electrolyte and leaves the container wall as positive metal ions. The charge removed by the positive ions going into the solution will cause a back- emf to build up between the zinc and electrolyte. When the back-emf has reached the same value as the solvation energy, the process stops because the electric field of the back-emf will prevent more ions from going into the solution. The result is an electric half cell with the positive metal ions in the solution forming a tightly bound electric double layer with the negative charged zinc. The other half of the cell is the carbon rod which has the function of supplying electrons to the electrolyte during operation but, the exact chemical reaction involved is not known. The voltage of the cell is determined by the difference in potential between the two half cells. The absolute energy or voltage of each half reaction is not known and no method has yet been devised to measure half cell potentials separately. The problem is how to electrically connect a voltmeter across the double layer without introducing another half cell reaction. Chemists have therefore settled for a compromise method whereby all types of half cells are compared to the voltage produced by a standard cell, the hydrogen half cell. This arbitrary method considers the potential of the hydrogen half cell as equal to zero and the difference in potential between the hydrogen half cell and any other half cell can be found in tables under the heading of Electromotive Force Series. One difficulty is that the potentials listed in the Electromotive Force Series refer to electrodes which are immersed in their own individual solutions containing their own ions, whereas in the dry cell both electrodes are in the same solution; i.e. a solution which contains zinc ions but no carbon ions. Another important question is what part does contact potential play in the electrochemical cell? The contact potential is often equal to or very near the potential of the cell itself and contact potentials are hardly ever mentioned in electrochemistry. Are electrochemical and contact potentials so closely related that when Professor Lodge argued with Lord Kelvin 100 years ago, he was right stating that they are of the same nature? It is now known for certain that the valence electrons, which determine the magnitude of contact potentials in metals, also determine the energy involved in electrochemical reactions.

In order to understand the electrochemical charging mechanism, especially as applied to atmospheric electricity, it might be helpful to describe a few laboratory and field experiments which were carried out in an attempt to clarify some of the above problems. Consider the following tests, which can readily be performed in the laboratory, and which are shown in Figs 22 a and 22 b:

1. Lower into an empty glass beaker two electrodes of differing materials (e.g. gold and magnesium) and connect an electrometer across the two electrodes. When a suitable radio-active source (e.g. Po 210, 500μCu) is placed at the bottom of the beaker, a potential of 1.2 volts will be registered between the electrodes, magnesium being negative with respect to gold.

2. Remove the radioactive source and fill the beaker with sufficient water to partially immerse both electrodes. The electrometer will again register a potential of 1.2 volts and the same polarity.

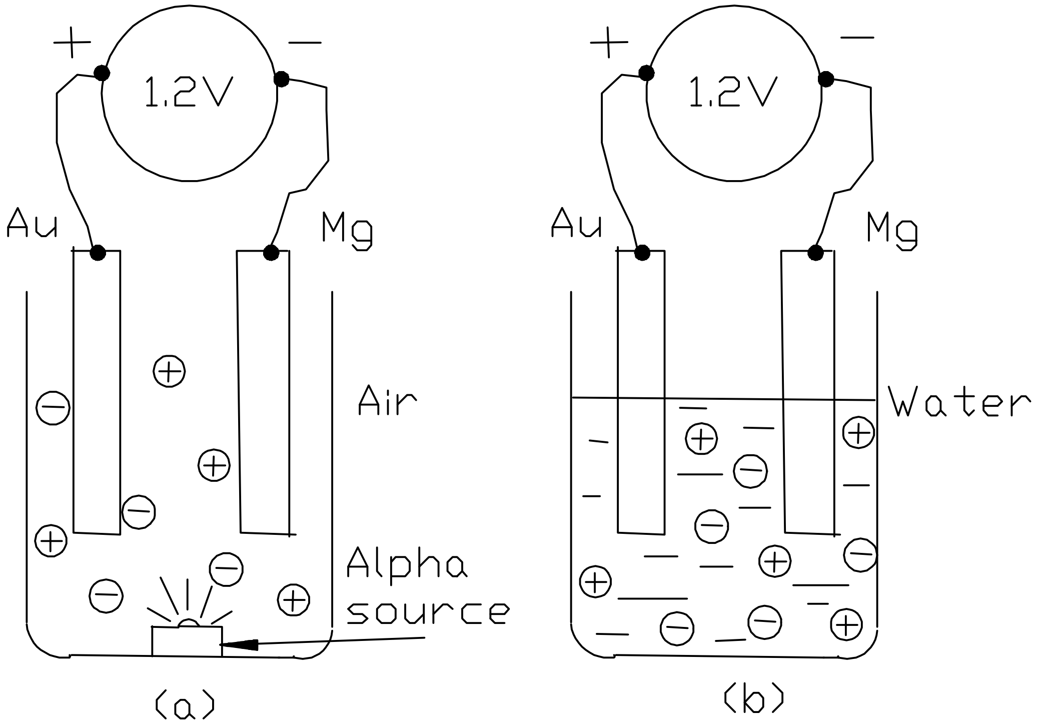

Fig. 22 Experiment demonstrating electrochemical charging. (a) Galvanic cell with ionized air as electrolyte. (b) Galvanic cell with water as electrolyte.

Both experiments deal with an electrochemical cell, using water for electrolyte in one case and air in the other. The fact that water contains ion pairs makes it an electrically conducting electrolyte which is vital for an electrochemical cell. In the first case, Fig. 22a, the beaker contains air instead of water, but the air is slightly ionized by the radioactive source, and although the number of ion pairs is relatively low, the ionized air is electrically conducting and exhibits the same properties characteristic of an electrolyte.

The average number of ions in the atmosphere is of the order of one billion pairs per cubic metre. The atmosphere is thus very much like an electrolyte. Its ion pairs are produced mainly by the constant bombardment of cosmic rays that reach our atmosphere from solar and stellar sources. Other ions are produced by radiation from radioactive materials in the atmosphere and in the earth's crust (see section 2.1.).

That material surfaces in contact with the ionized atmosphere are subject to electrochemical charging can also be shown by simply probing the atmosphere with electrodes and it can be easily demonstrated that in the absence of external electric fields negative ions rather than positive ions have a tendency to adhere to material surfaces. It can also be observed that when the probes are ventilated by strong winds, more ions are supplied and the negative ion current to the probes increases accordingly. At first, when such measurements were performed it was not at all clear what caused this type of charging although it was believed at the time that work function potentials might play an important part in the charging process. Experiments were therefore set up in the laboratory in an attempt to duplicate the field measurements described. Surfaces of diverse materials were ventilated by artificially ionized air. The results revealed a marked difference in activity among various materials in contact with the ionized air. For example, a small sheet of magnesium metal will absorb negative ions and continue to charge until it reaches a potential of -1.6 volts with respect to its surroundings. Each material tested was found to reach its own characteristic potential. An abridged list of these materials with their potentials appears in Table 1. The potentials in Table 1 agree with the potentials measured by the Gerdien cylinder shown in Fig. 11 section 2.1.

Material	Work function	Equilibrium Potential Volts
Magnesium	3.66	-1.6
Wet Filter Paper	-	-1.05
Aluminium	4.24	-1.0
Cadmium	4.22	-1.0
Tantalum	4.25	-0.65
Molybdenum	4.50	-0.6
Copper	4.65	-0.45
Stainless Steel	4.75	-0.4
Ice	-	-0.4
Gold	5.35	-0.22

Table 1. Work function and electrochemical equilibrium potentials for different materials.

At first it was disappointing to discover that the potentials in Table 1 were not at all proportional but rather inversely proportional to the work function potentials of the different materials in question. However, Table 1 has a very familiar appearance to the chemist; it resembles a chemical activity series in which the electrical potentials are values related to the oxidation energies of the materials in question. To further prove the electrochemical effect, the same materials were immersed in distilled water, two at a time. Differences of potential between the two materials were then measured and found to be directly proportional to the respective differences between their values in Table 1. From the experiments and the results of Table 1 it became clear that we were dealing with an electrochemical effect and the following conclusions were drawn: negative ions in air are formed from acidic molecules such as O_2 and NO_2. These molecules are electrophilic (electron seekers) and capture free electrons produced in air by the various ionization processes. The molecular ion, with its valence slightly reduced by a captured electron, still remains acidic and very active as an oxidizer. When the molecular ion with its piggy back electron oxidizes a surface material, negative charge is transferred by the captured electron to the surface. As more ions reach the surface, a negative back-emf will build up, eventually repelling any incoming ions until no more can reach the surface. Just as in the case of the dry cell an electrical equilibrium potential is achieved. The electrochemical reaction halts when the back-emf has reached a value equal to the oxidation potential, or the energy of the chemical oxidation-reduction reaction involved. Oxidation- reduction reactions can be described as follows: if two elements combine where one wants to share one or more of its electrons with another element which is an electron acceptor, then an oxidation-reduction reaction has taken place. The electron donor is said to be oxidized and the electron acceptor is reduced. It is interesting to note that in Table 1 the elements on top of the activity list are the elements with the lowest work functions. These elements freely share their valence electrons and become easily oxidized because of their lower electron binding energies or lower work functions. Electronegative elements or compounds are the electron seeking oxidizers that want to react with and oxidize material surfaces. The oxidizers or electron seekers appear most commonly as negative ions in the atmosphere and solutions because of their tendency to pick up free electrons in the surrounding environment. When the oxidizer reacts with a material surface it brings a captured electron along. The electron will charge the surface and the chemical reaction can therefore be classified as an electrochemical reaction.

The fact that oxidizers in general appear as negative ions forms the basis for electrochemistry in the atmosphere and in solutions. It is very important to remember, especially in the case of atmospheric electrochemistry, that the process which brings the negative ion to a material surface is the chemical reaction and not the electric field of the ionic charge itself. In an electrochemical reaction the electron simply enjoys a piggy-back ride to the surface and the image force produced by its charge is too feeble and short in range to compete with chemical processes. Also, the strength of the image forces are equal for both positive and negative ions so that no preferential charging of either sign can be expected. Nevertheless, a theory put forward by Phillips and Gunn (1954) considers the difference in mobility between negative and positive ions as possible mechanism for preferential charging. This theory will be discussed in section 3.4.

Numerous experiments in ionized air revealed that all metals and conductors tested proved to be oxidized by negative ions and achieved negative equilibrium potentials. No positive equilibrium potentials were ever encountered. In cases where two electrodes are immersed in an ionized medium, both electrodes become negatively charged relative to the electrolyte, the lower work function material being more negative. The potential between the electrodes equals the difference in oxidation potentials between the materials which also seems to equal their difference in work function except for the reversal of sign. The results were the same for electrodes immersed in ionic solutions where the electrodes did not go into solution. No tests were performed involving solvation energies.

One striking feature of the electrochemical charging process is the formation of the electric double layer. The oxidiation potential and the thickness of the double layer determines the charge density on the oxidized surface. In normally ionized air the double layer thickness is about 1 mm which means that a surface of water, for example, which has an oxidation potential of -0.25 volts, will charge to

$$Q / \mathrm{m}^2 = V_0 \varepsilon_0 / d \tag{4}$$

where d is the double layer thickness and V_0 the oxidation potential. The value of d was derived from laboratory experiments which proved that the average thickness of the double-layer in air, to the accuracy of the measurements, is equal to the average distance between the ions. It appears that each time a negative ion reaches the surface, it leaves behind a positive ion at a distance equalling the average ion spacing. This distance for normally ionized atmosphere is about one millimetre. In an experiment that followed, isolated spheres (to simulate cloud drops) were ventilated by ionized air for the purpose of studying the charge collection on their surfaces. Negative ions reacting with the surface of a drop or a sphere leave behind positive ions in the surrounding air, thus building up a diffuse space-charge cloud around the sphere. This cloud of positive ions forms an electric double-layer with the charged surface of the sphere. The double-layer, formed by the charged surface of a sphere surrounded by its diffusion cloud of opposite charge, constitutes a spherical capacitor with a capacitance of

$$C = 4\pi\varepsilon_0 R(1 + R / d) \tag{5}$$

R is the radius of the sphere, and d is the mean thickness of the double-layer. The charge on the sphere can then be calculated from the electrostatic expression

$$Q = V_0 C = V_0 4\pi\varepsilon_0 R(1 + R / d) \tag{6}$$

where the double-layer voltage, V_0, is also the oxidation- reduction potential of the reaction between the ions and material in question. The sphere with its attached double- layer is, of course, electrically neutral with regard to its surroundings. No drastic charge separation has yet occurred.

The fact that the sphere or drop is ventilated (by falling through the atmosphere for example) is of the utmost importance because the flow of air will partially strip away and remove the outer positive layer of the double-layer, a process that not only increases the potential of the sphere but also makes it appear to be negatively charged relative to its general surroundings. For example, removing the outer-charged layer to infinity will increase the potential on the sphere to

$$V = \frac{Q}{C} = V_0(1 + R/d) \tag{7}$$

Common experimental values for charged water drops are shown in Figure 23. These values were obtained from actual measurements in the atmosphere by several investigators and are compared here to the predicted values represented by the solid curve. The curve is constructed from Equation (6), using a double-layer thickness of one millimetre, a distance that equals average ion spacing in the normally ionized atmosphere. An oxidation potential of $V_0 = -0.26$ volts for water was chosen from the experimental results of Chalmers and Pasquill (1937) who, in their laboratory, measured the equilibrium potentials on water drops. The oxidation potential for water in Table 1 is -0.4 volts

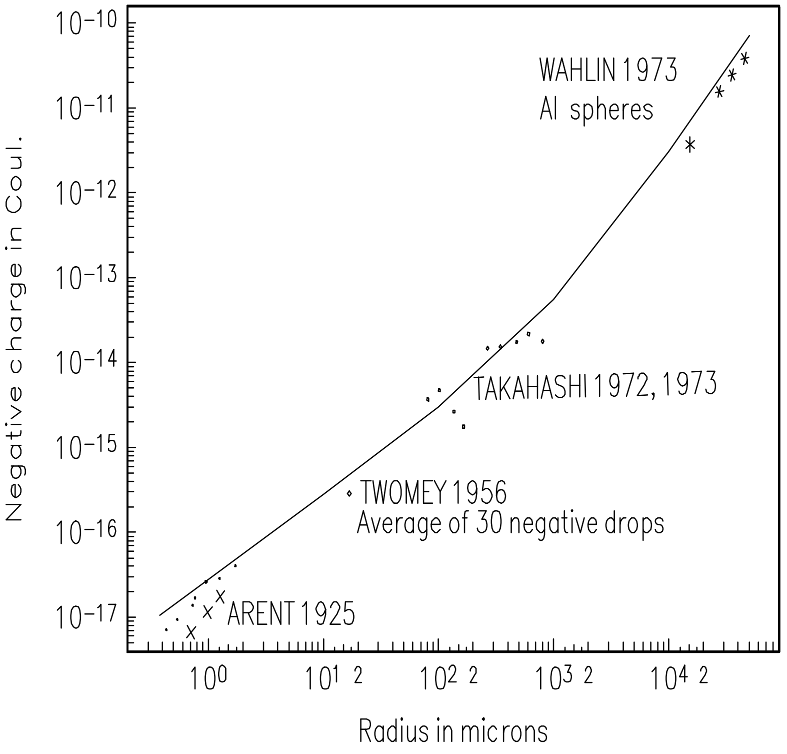

Fig. 23 Charge-size measurements of rain drops compared to values predicted by the electrochemical charging process.

which is the value obtained for water when ventilated by an equal amount of positive and negative ions. In the normally ionized atmosphere there are more positive than negative ions. A typical ratio is $N^+ / N^- = 1.2$ which will lower the equilibrium potential to about one third (see the results in Fig. 11) and which agrees with Chalmers and Pasquill's results.

3.4.1 OTHER CHARGING PROCESSES

Except for the well established effects of contact and electrochemical potentials there have been several other charging mechanisms proposed in order to explain atmospheric electric phenomena. Some involve influence charging in combination with colliding drops such as the Elster-Geitel theory, or influence charging and the capture of ions as proposed by Wilson. There are charging mechanisms which consider ice splintering and freezing of water drops to be important in producing charge on precipitation in clouds. The process of evaporation and recondensation of water has intrigued many investigators, including Volta, as a possible source of positive and negative charge in clouds. Takahashi (1973) has recently carried out work along these lines. The validity of the above charging mechanisms are difficult to verify since they lack rigid experimental proof. Recent years have seen many sophisticated computer models which will work if the right parameters are plugged in. This is especially true for influence charging mechanisms involving collision between polarized drops in strong vertical electric fields. Some mathematical computer models are often conceptually difficult. Since mathematics is only a tool of science it is more desirable to have a conceptual theory that is mathematical sound than a mathematical theory that is not conceptually sensible.

3.4.2 THE ELSTER-GEITEL PROCESS

The Elster-Geitel (1885) process deals only with the charging of precipitation particles such as drops in clouds. It is based on the assumption that cloud drops, which are polarized in the electric fairweather field, collide with each other and exchange surface charges in a manner that will enhance the fairweather field to a magnitude found in thunderstorms. The proposed charging mechanism is shown in Fig. 24 where a smaller drop which is still light enough to be swept upwards by the updraft winds in the cloud, collides with a much larger and heavier drop falling down through the cloud. Since the large drop is polarized in the electric fairweather field its upper surfaces will attain a small excess of negative charge while the same amount of positive charge will appear on the lower surface. A small drop bouncing off the lower surface will pick up some of the positive charge at the moment of contact, and bring it along to the top of the cloud while the heavier and larger drop, now negative from the loss of positive charge, will continue to fall and bring negative charge to the lower region of the cloud. It can easily be seen that as

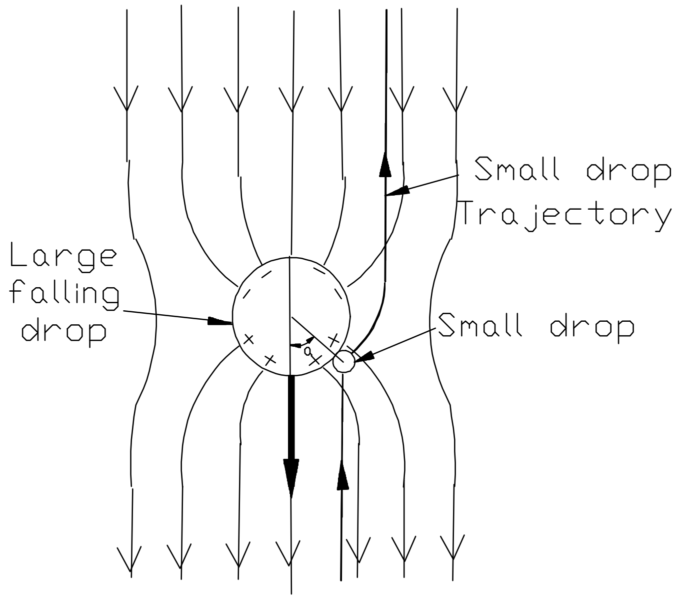

Fig. 24 The Elster-Geitel charging process of cloud drops.

charge is being separated the electric field strength will increase which in turn intensifies the charging process.

Although sophisticated computer models of the Elster-Geitel theory have been studied (Sartor (1954), Levin (1975) and (1983)) the general feeling is that the charging mechanism is too weak. The theory might possibly work if the initial electric field strength was equal to that found in thunderstorms in which case some other process must prevail (Pathak 1980). Unfortunately there is no laboratory support for the Elster-Geitel theory and some of the unanswered questions are: do small drops really bounce off larger drops at a reasonable rate or do they simply coalesce? How important is the grazing angle of the impact?

3.4.3 THE WILSON EFFECT

The Wilson (1929) effect is similar to the Elster-Geitel process with the exception that instead of small drops colliding with larger polarized drops negative ions are believed to be swept up by the larger drops as they fall through the normally ionized air. The positive charge induced on the bottom half of a large drop is believed to preferentially collect negative ions as the drop falls down to lower altitudes thus leaving an excess positive charge behind at higher levels. One serious objection to the Wilson theory is that there are not enough ions produced inside a thunder cloud to account for the amount of charge separated.

3.4.4 DIFFUSION CHARGING

Gunn (1957) considered a charging mechanism which is based on the difference in mobility between positive and negative ions. He assumed that since negative ions display a higher mobility in an electric field than do positive ions they must be less massive and therefore also have a greater diffusion coefficient in the absence of electric fields.

The expected result is that negative ions will diffuse on to cloud drops at a higher rate than positive ions can. Cloud drops will therefore become more negatively charged and as they fall down to the lower region of the cloud they leave an excess of positive charge behind in the upper region. Experiments with metal spheres ventilated by ionized air were carried out by Phillips and Gunn (1954), see Fig. 25. They confirmed that chrome plated metal spheres exposed to ionized air indeed charged negatively and reach certain equilibrium potentials.

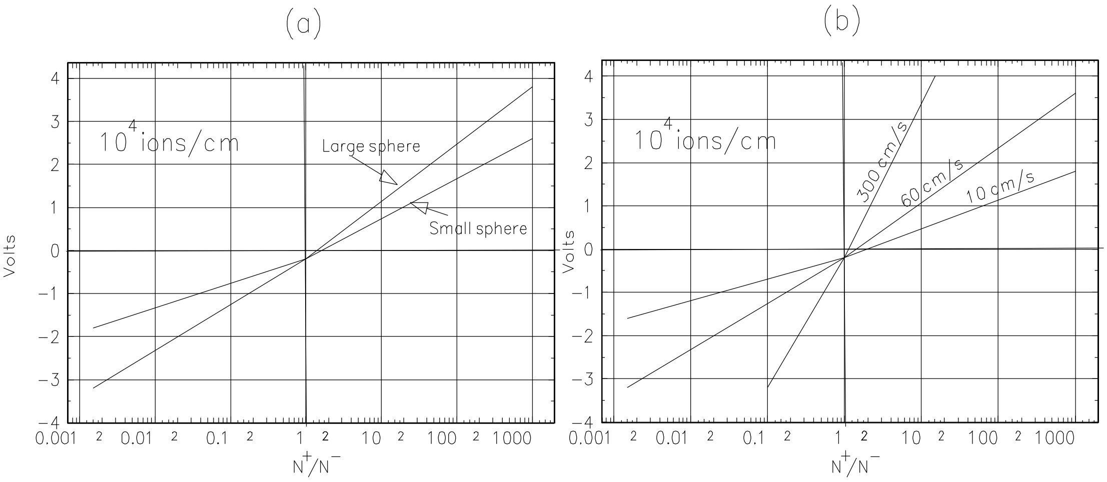

Fig. 25 Phillip's and Gunn's Experiment with metal spheres. (a) Equilibrium potentials for different dia. spheres. (b) Equilibrium potentials for different air velocities.

What is interesting about Phillips' and Gunn's experiment is its similarity to the experiments which demonstrate the electrochemical charging mechanism by the use of Gerdien cylinders, see Figs. 10 and 11. In fact, some investigators argue that the equilibrium potentials in Table 1 and the results shown in Figs. 10 and 11 can be explained by the diffusion theory. But this is hard to believe since the diffusion theory does not predict that different materials charge to different equilibrium potentials as shown in Table 1. Also, the energy of the diffusion process relates to the thermal agitation of the ions in their atmospheric environment and is determined by the air temperature. This means that the equilibrium potential, using Boltzmann's constant k and a maximum air temperature of T = 20 C, cannot exceed

$\frac{3}{2}kT \times 0.25$ or 0.008 volts where 0.25 represents the 25 percent higher mobility of negative ions over positive ions. This is only 3% of the values shown in Table 1 and the graphs of Phillips and Gunn (Fig. 25).

3.4.5 FREEZING POTENTIALS

Although it is known that warm clouds (clouds that do not contain ice or frozen precipitation) can charge to considerable potentials and occasionally produce lightning, there is still a great deal of attention devoted to charging processes that might involve icing. One reason for this is the observation by Simpson and Scrase (1937) who noticed that separation of charge in thunderclouds seems to occur in regions where the temperature is below freezing. They thought that colliding ice particles might charge negatively and leave positive charge behind in the air in the form of positive ions. The idea was further developed by Reynolds (1954). Workman and Reynolds (1950, 1953) have also discussed a theory in which glazing processes were assumed to produce charge. Many theories have been put forward which involve freezing of water, but again there are, unfortunately, no convincing laboratory experiments.

CHAPTER 4

THE THUNDERCLOUD

4.1 THERMAL DYNAMICS

On a warm and sunny day heat is absorbed by the earth's surface causing both water vapour and hot air to rise to higher atmospheric levels. The warmer the air the more water vapour it can hold before saturation. The secret of a thunderstorm and its devastating power lies in the amount of water vapour present. The latent energy of water vapour is the key factor in thunderstorm formation and electrification. At the beach most of us have experienced the sensation of cold after getting out of the water. The explanation is that water takes away heat from the body surface at a rate of about 540 calories or 2250 joules per gramme when it evaporates. When water condenses to form drops, the same amount of energy, in the form of heat, is released. When a cloud is formed by warm humid air reaching higher and cooler levels, condensation will create small drops between 5 - 10 microns dia. During condensation latent heat is released warming the surrounding air which will rise further (the hot air balloon effect) to considerable heights pulling more humid air up from below to take its place. A chain reaction starts when humid air is fed in from below pumping more and more energy into the cloud in the form of heat and convection. A cloud might grow up to 60,000 feet or 20 km in height in very narrow columns. Updraft velocities can easily reach 30 m/s. Each column is called a cell and a thundercloud usually consists of several cells. Each cell has an average life time of 30 minutes. When a cell reaches a maximum altitude it is in a so-called mature stage, at which time the cloud top usually flattens out to form the typical anvil shape often associated with thunder heads. The flat top is believed to occur when a cloud reaches the stratosphere or the boundary in the atmosphere where air temperatures begin to rise with altitude which will cause the cloud to evaporate again. In the growing stage of a cloud, drops increase in size as they reach higher levels and eventually become heavy enough to overcome the updraft velocities of the convection. At this point rain or precipitation starts to fall down and the cell has reached its dissipating stage. At greater latitudes, thundercloud precipitation without exception always reach temperatures below freezing. Hail and sleet are formed which often fall down and reach ground before melting. The cold precipitation will cool the air in the cell which becomes heavy and begins to move downward as well. The cold heavy air might cause considerable down drafts which, when reaching ground, supply the outward rush of cool air often felt as a relief to most of us after a hot day. When precipitation and down drafts occur the cell is said to be in its dissipating state and although some lightning activity is still going on its life is nearly over. New cells are formed adjacent to the old ones and often while observing their growth one can see small bubbles or turrets developing above and around each cell. The thunder cloud described is a typical heat cloud or isolated storm. Also common are lines of thunderstorms or squall lines. Lines of thunderstorms are formed when a cold front wedges in under the warm humid air mass along a warm front and squall lines can extend for several hundred miles. The energy source is the same, namely condensation of water and release of latent energy. Thunderstorms along a squall line often seem to lean over on their sides because of the heavier and cooler air wedging in underneath the warm front. Although there is considerable entrainment of air through the sides of such storms, the main energy is still being supplied by the updraft winds caused by condensation and convection.

4.2 THE ENERGY OF THUNDERSTORMS

The energy of a thunderstorm is determined by the amount of water vapour present. Typically, a single cell contains 8×10^{8} kg of water (Israel 1973) of which about 60% is converted into heat and precipitation. The total energy dissipated by a cell is therefore 1.4×10^{14} joules per cell. If the average life time of a cell is 30 minutes then the average power per cell equals 7.8×10^{10} watts. Since it is estimated that about 2,000 thunderstorms are active at the same time around the earth then the total energy dissipated must equal about 1.6×10^{14} watts which is about 0.1% of the energy reaching the earth's surface from the sun.

4.3 THE ELECTRICAL ENERGY IN THUNDERSTORMS

The electric energy of a thunderstorm can be determined in several ways. Schonland (1950, 1953) for example, estimated the potential difference involved in a typical lightning flash to be about 10^{8} and 10^{9} volts while the average charge transferred is about 20 coul. The total amount of free charge separated in a thundercloud is about 1400 coul according to Wormell (1953). From the above information we can determine the total electric energy involved to be of the order of $\frac{1}{2}VQ = 7 \times 10^{11}$ joules. A more realistic method presented by Israel (1973) yields a larger value of 1.7×10^{13} joules. In Israel's approach the average thunderstorm cell is compared to an electric circuit diagram (Kasemir 1965) such as shown in Fig. 26. According to Wait (1950), on the other hand, a thunderstorm cell generates on the average 2.5 amperes of current and from Ohm's law one then finds that the total power output of a cell is $RI^{2} = 9.4 \times 10^{9}$ watts, which when integrated over the duration of a cell (30 min.) amounts to 1.7×10^{13} joules.

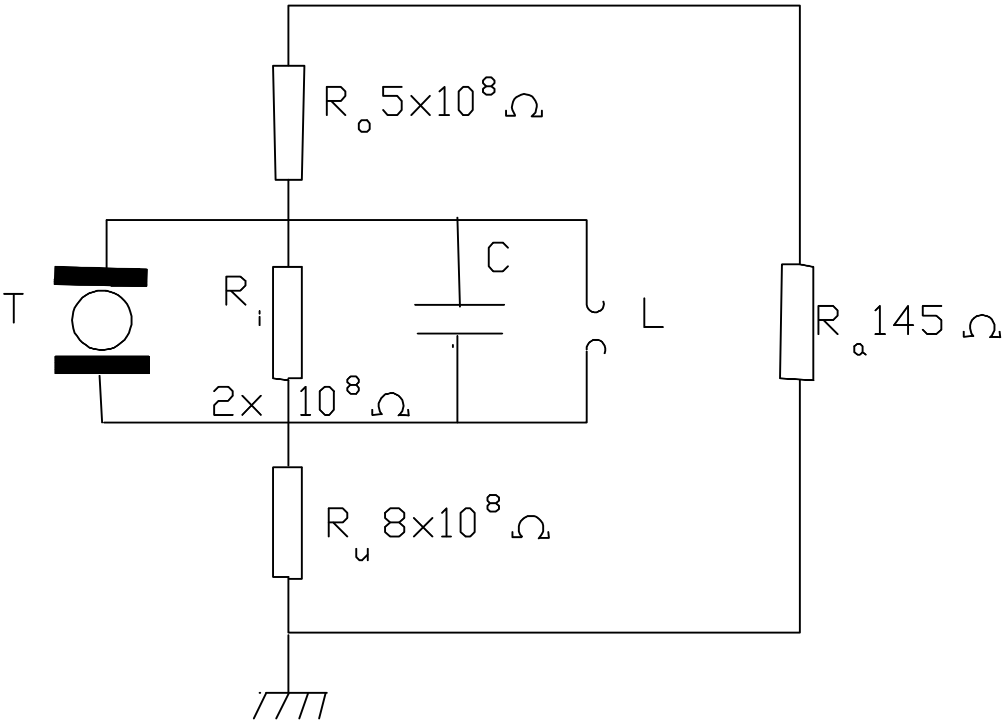

Fig. 26 Electric circuit diagram of a thunderstorm (Kasemir).

There is at least one serious objection to the use of Ohm's law in determining the electric power of a thunderstorm. There is no doubt that the internal resistance of a thundercloud, such as represented in Fig. 26, is an accurate measurement since it has been determined in situ by conductivity measurements. The problem is that the conductivity or resistance depends on the number of ions available in the cloud. Ions are the charge carriers in the cloud and solely responsible for the flow of electric current. Considering that the normal ion production in the atmosphere is about ten ion pairs or 1.6×10^{-18} coul of charge per cubic centimetre per second and that the average thunder cloud cell measures fifty cubic kilometres in volume, then the total charge production within the cell is only 0.08 coul per second which translates to a meagre saturation current of 80 mA.

Ohm's law will not hold once a state of saturation has been reached. An ohmic current of 2.5 amperes used by Israel might, therefore, not be a representative value. If the charging current in thunderstorms is non-ohmic the only other alternative is a mechanical transfer of charge by means of convection and updraft currents or by the process of gravitation in which charge is transported by falling precipitation particles. The author has reason to believe that when Israel calculated the average electric power involved in charging of clouds he not only considered the ohmic current but also took the mechanical currents into account.

An interesting thought is that if one could harness a lightning bolt there would be 100 kW hours of energy available, enough to keep a 100 watt light bulb burning for a month and a half. There are nearly 50,000 thunderstorms active around the world per day, and if each storm produced 100 lightning bolts, there would be close to 20,000 megawatts, which is only enough power to satisfy the needs of a large city such as New York.

4.4 THE LIGHTNING DISCHARGE

There are two types of lightning discharges, namely cloud to ground lightning and intracloud lightning or cloud to cloud discharges. As the reader recalls from section 1, Benjamin Franklin observed that thunderclouds tend to form gigantic electric dipoles in the atmosphere where the lower portion of the cloud is predominantly negative and the top positively charged. Occasionally a small pocket of positive charge can be found in the lower cloud section during the clouds dissipating stage in the vicinity of the precipitation shaft. There is about three times as many intracloud discharges as cloud to ground flashes, a ratio which seems to vary somewhat with geographical location. Normally negative charge is brought to ground by lightning that originates from the lower section of a cloud. However, as more and more negative charge is drained by successive lightning bolts to ground, the more the positive charge in the upper half of the dipole becomes dominant. The imbalance of the dipole will eventually cause a positive discharge to appear from the upper region of the cloud to ground. The ratio of positive to negative ground strokes is about 1 to 10 but a positive discharge from the upper region of the cloud is as a rule ten times more powerful and carries ten times more charge.

A lightning flash to ground might consist of several ground strokes which appear in such a rapid sequence not distinguishable to the naked eye but easily observable with the aid of a Boys camera (see Fig. 6). In fact, pictures taken with a Boys camera reveal quite a complicated discharge mechanism of lightning. It appears that lightning is triggered by faint pilot streamers (Schonland, 1938) which provide the initial ionized path from cloud to ground by weak corona discharges.

The pilot streamer is usually followed by a stepped leader which can be described as small current surges catching up with the pilot streamer in small steps averaging 50 metres in length. As the pilot streamer works its way down to ground, followed by the stepped leader, charge is brought down at a rate of 600 to 2,600 amperes (Hodges, 1954). As soon as the pilot streamer and stepped leader reach ground a conducting path between cloud and earth is established and the main stroke starts. The main stroke will transfer charge between the cloud and earth at a rate of 20,000 or sometimes 400,000 amperes. One peculiar behaviour of the main stroke is that it starts from ground and travels upwards which means that large amounts of charge are transferred from the earth's surface to the cloud. This is why it is often said that lightning travels up to the cloud rather than down. Although the main stroke only lasts for a thousandth of a second, it is visible to the naked eye and the sound or acoustic shock wave that follows is unmistakably that of a lightning bolt. The main stroke often breaks up into several subsequent strokes usually three or four per lightning flash. Lightning flashes having as many as 42 subsequent strokes have been reported. There are usually no pilot streamers or stepped leaders present between ensuing strokes in a multi-stroke flash since an ionized path has already been established after the first stroke. The stepped leader is replaced by a so-called dart leader which works its way down from the cloud in one single jump.

One of the most challenging problems in lightning research is the mystery of how lightning bolts drain charge from the thunder cloud. How does the lightning bolt connect up to the myriads of charged drops involved, and how can this be accomplished in a small fraction of a second? This is a question that has not yet been answered satisfactory. One of the best solution so far, in the author's opinion, is offered by Kasemir (1950b) who believes that the mechanism involved

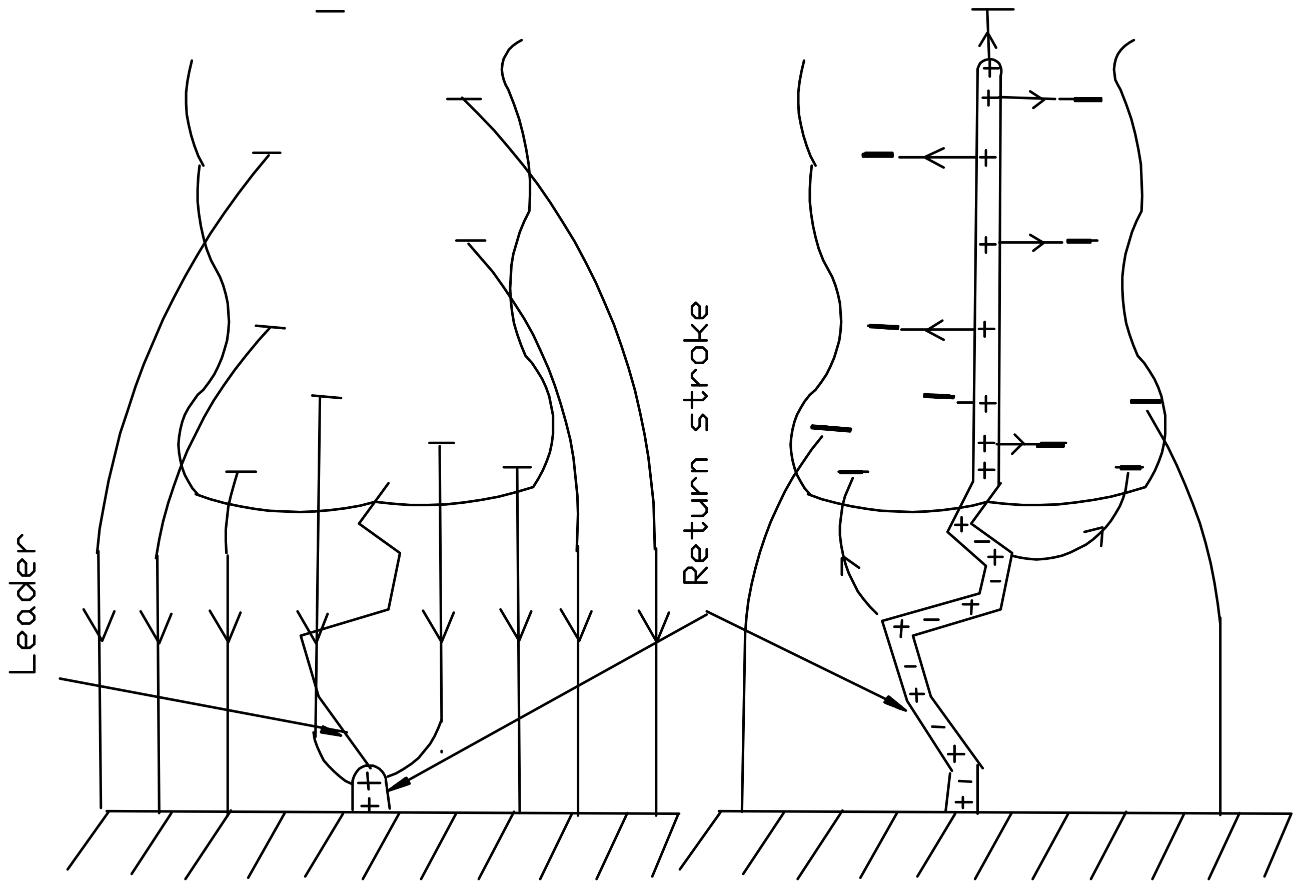

Fig. 27. Lightning discharge theory by Kasemir. (a) Start of return stroke. (b) Charge transfer by induction.

is that of induction. For example, consider a negative charge centre in the lower portion of a cloud (see Fig. 27a.) from which a leader stroke has just reached ground. Not much charge is brought down because the negative charge in the cloud is bound on small drops or precipitation particles. On the other hand the earth's surface, which is a good conductor, can easily supply positive charge up to the cloud through the ionized channel provided by the stepped leader. When the positive charge in the return stroke reaches the cloud it will penetrate the charge centre like a giant lightning rod which causes the cloud potential to drop drastically since many field lines now are much shorter (see Fig. 27b). In this way much of the electric energy is drained without contact or charge being removed from the drops in the cloud. Another way to look at Kasemir's lightning mechanism is to compare the electric potential and capacitance of the cloud before and after a lightning bolt has penetrated the cloud. A 25 coul charge centre measuring 1 km in diameter and located several kilometres above ground has a capacitance of roughly

$$C = 4\pi\varepsilon_0 R \approx 6 \times 10^{-8} \text{ Farad}, \tag{8}$$

which yields an electrical potential of

$$V = \frac{Q}{C} \approx 4 \times 10^{8} \text{ volts} \tag{9}$$

If one assumes that the charge centre is cylindrically shaped and extends $\ell = 1$ km in height then the capacity between the charge centre and the ionized channel of the lightning bolt can be estimated from

$$C = \frac{2\pi\varepsilon_0 \ell}{\ln(b/a)} \approx 9 \times 10^{-9} \text{ Farad}, \tag{10}$$

where a is the radius of the lightning channel and b the average distance of the charges in the charge centre with respect to the lightning channel. The radius of the lightning channel was taken as 1.5 m. The result is that the capacity is lowered by about 15 percent which corresponds to a charge transfer of approximately 3.5 coulombs. Since there are on average three or four main strokes per lightning flash, then the total charge transferred by a lightning flash can easily be between 10 to 15 coulombs. In fact, Kasemir's capacitor model does not only explain why the bulk of charge is moving upwards, from ground to cloud under the influence of induction, it also presents us with a $R\ C$ circuit which might explain the occurrence of multiple strokes. The time constant of the circuit (average 200 micro seconds) is the product of (9) and (10) divided by the average discharge current of 20,000 amperes. Once the capacitor is charged the lightning current will cease and the time elapsed is of course determined by the time constant. After an average length of 0.07 s (Bruce and Golde, 1942) the next stroke will follow but this time triggered by a dart leader. The ensuing stroke will form a branch which will penetrate the cloud in a different direction. More strokes might follow until the negative potential of the entire region has been drastically lowered. The induced charge transferred in this manner must slowly leak off to the individual drops by means of corona and quiet discharges that will produce numerous ion pairs inside the cloud.

In an intracloud discharge a lightning bolt starts between two charge centres where the electric field is high and where electric breakdown of air first occurs. After a small parcel of air becomes conducting from the electric break-down of the air, it will grow rapidly in length under the influence of induction. As the lightning rod type discharge grows and becomes more and more polarized, positive charge will rush towards the end facing the negative charge centre while negative charge collects at the end nearest the positive charge centre. Once the lightning bolt penetrates both charge centres, enough charge is transferred by the polarization effect to account for the observed electric field changes.

One important implication of Kasemir's discharge mechanism is that the electric breakdown of air, and formation of small ionized air parcels in the strong field region between two charge centres, can trigger lightning bolts. This means that aircraft flying in the high field region between two charge centres can become electrically polarized and trigger lightning as well. The risk of damage is therefore much greater than previously thought when it was believed that lightning would only strike aircraft in a random fashion. Pilots usually refer to two kinds of lightning strokes; the "static discharge" when lightning seems to originate from the aircraft when it is trying to rid itself of excess charge or the direct hit when lightning is believed to come from a nearby charged cloud centre. The "static discharge" is controversial in that the charge on an aircraft, as pointed out by many scientists, is by far to small (about 0.1 coul) to create a lightning bolt. Kasemir's induction mechanism, however, explains that the strongly polarized electric field on the aircraft, which causes corona discharges and electric break down of air, will supply all the charge necessary for the lightning bolt. For example, although the net charge on the aircraft is small, the amount of ionic charge produced in the surrounding air by corona is enough to feed a lightning bolt by the inductive mechanism. One cubic centimetre of completely ionized air contains as much as 5 coulomb of ionic charge.

The size of a lightning bolt or the diameter of the lightning channel is not well known. A rough idea of the channel diameter can be obtained if one uses Gauss's law to determine the maximum radius of electric break down of air due to the strong field and corona produced by the charged lightning channel. For example, an average lightning bolt which carries a current of I = 25000 amperes and propagates at a typical velocity of $v = 10^8\ \mathrm{m/s}$ will contain a charge of

$$\lambda = I / V \approx 2.5 \times 10^4\ \mathrm{coul/m}, \tag{11}$$

which by the use of Gauss's law equals a channel radius of

$$R = \lambda / (2\pi\varepsilon_0 E), \qquad (12)$$

where $E = 3 \times 10^6$ volts / meter is the breakdown field of air. It is believed that the bulk of the lightning current is confined within a small core having a cross section of about 0.15 m in diameter. The lightning channel can be pictured as narrow current core surrounded by a large ionized envelope.

4.5 PROTECTION AGAINST LIGHTNING

Modern lightning research has proved that the function of a lightning rod is not necessarily to attract lightning and lead it to ground in order to dissipate its power. Careful studies show that large amount of corona discharges occur at the point of a lightning rod before lightning strikes. The corona discharges produce numerous ions which often develop a protective dome around structures fitted with lightning rods. A large conductive dome or sphere has a smoothing effect over protruding surfaces and will lower the electric field strength in its vicinity. Experiments by Vonnegut and Moore have shown that a conducting wire supported between two mountain tops did not get struck by lightning as often as the mountain sides. In fact, the lightning seemed to avoid the wire which can be explained by the above effect that corona must produce a large diameter ion cloud around wire. This has the effect of lowering the field strength in its vicinity.

The lightning rod is an important device for protecting houses and other structures from lightning. This is accomplished in two ways. First, corona from the sharp tip of a lightning rod creates a protective ion cloud above the structure and secondly, in the event of a direct hit by lightning, the lightning rod will force the electric power to dissipate into ground and prevent damage to the building. When protecting a house or structure from lightning damage, some fundamental rules should be followed. Basically, there are two types of protection internal and external. External protection concerns the safety of the roof and sidings of a house while interior protection deals with sparks between metal object inside the house including appliances, radios and telephone lines. An effective exterior lightning protection can be described as follows:

1. Lightning rods and guard wires must be symmetrically placed over the roof to ensure uniform distribution of the discharge current.

2. Lightning rods and guard wires should be symmetrically attached to vertical conductors which distribute the discharge current to ground.

3. A good ground connection, far away from under-ground cables and pipes, must be established in order to let the energy dissipate into the earth's surface as effectively as possible.

4. No point on a roof should be further than 10 m (30 feet) away from a lightning rod or guard wire.

5. All large metal objects on the roof should be electrically connected to the system.

6. Any large protruding non-conducting structures or objects on the roof should be fitted with an individual lightning rod.

7. A building should have at least two vertical conductors to ground.

8. Sharp bends or curves in the conductors should be avoided since they might cause unnecessary induction and build-up of sparks.

9. Large metal objects on the outside walls such as balconies and fire escapes must be connected to the system.

The recommended cross sectional area of guard wires and vertical conductors should be at least 2.5 mm^2 (0.04 in^2)and twice that area for the buried wires that serve as ground connections. Fig. 28 shows several different ways of protecting houses and similar structures. Fig. 29 demonstrates a simple lightning protection for sail boats. Thousands of people around the world are killed by lightning every year. Most fatalities occur in open fields and usually when lightning strikes within a range of 50 to 100 metres of a victim. The danger is that when lightning strikes ground a tremendous amount of current radiates out in all directions from the point of impact. Since the current travels on the surface, which offers a certain electric resistance, a voltage drop will develop which can easily reach 1,000 volts per metre 100 metres from the point of impact. A person walking might bridge a lethal voltage in one step or could easily be thrown off his feet by muscular contractions from the electric shock. It has often been said

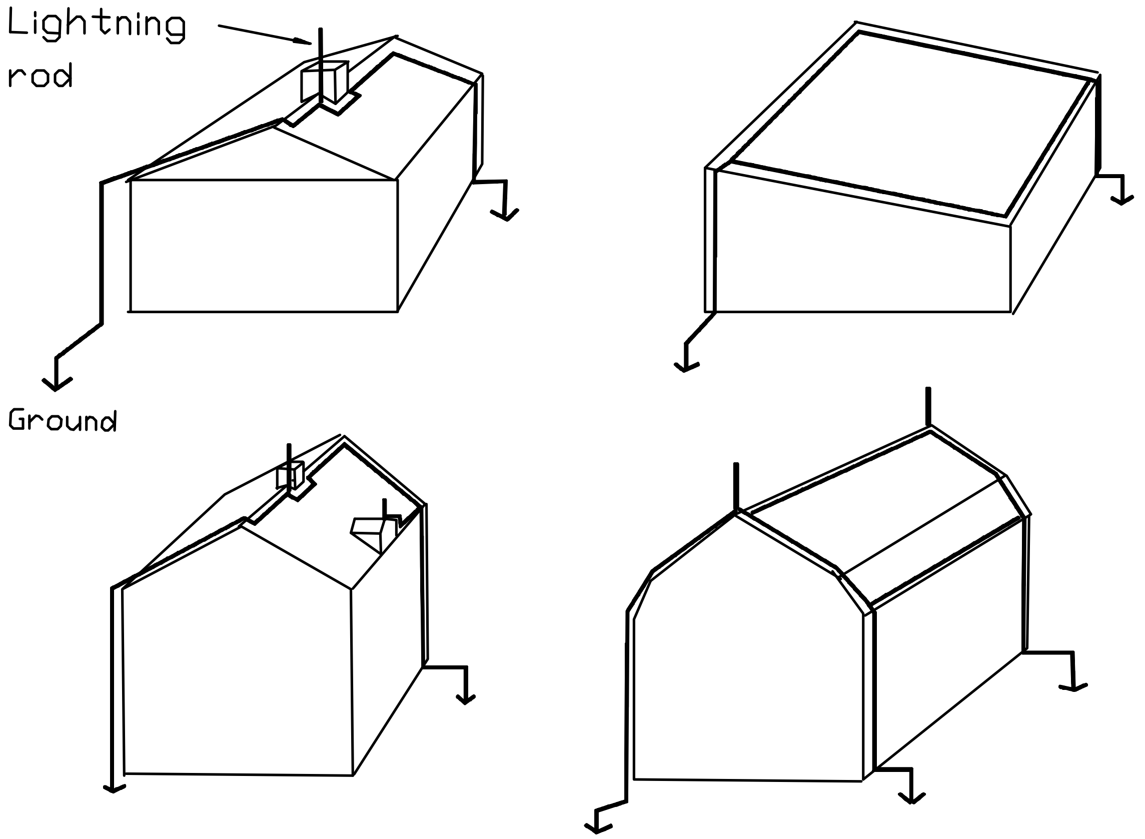

Fig. 28 Lightning protection for houses.

that it is safer to run than to walk across a field during an overhead lightning storm. This might not be far from the truth since only one foot would be in contact with ground in case lightning strikes. Lightning often strikes trees, consequently one should never take shelter under trees, near power lines, *etc*.

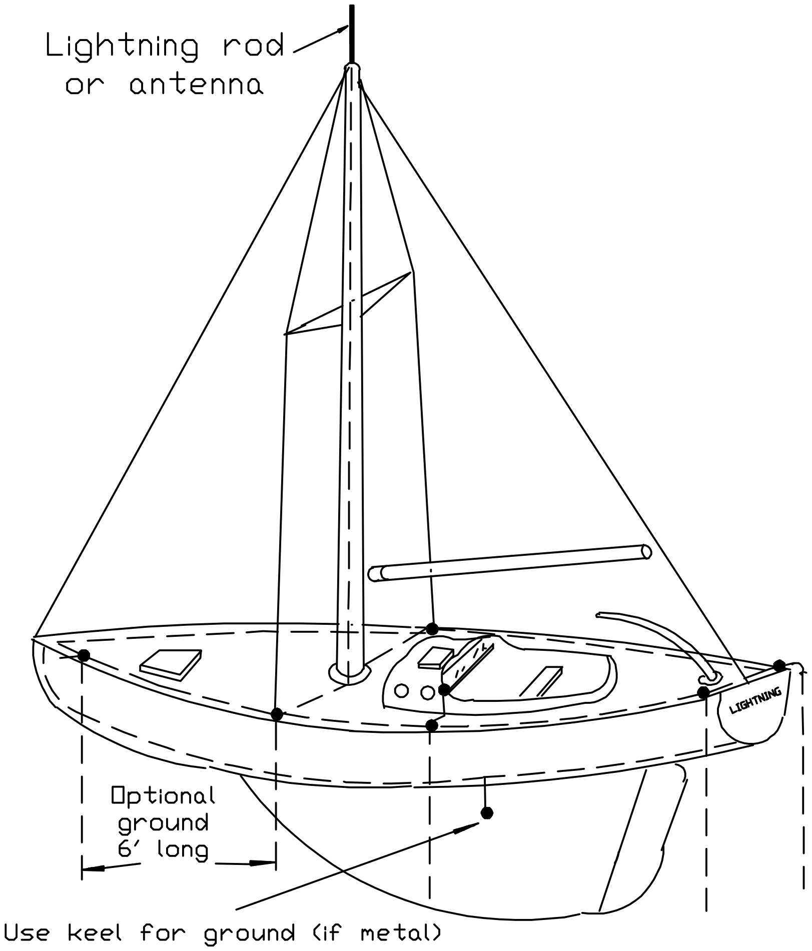
Lightning rod or antenna
LIGHTNING
Optional ground 6′ long
Use keel for ground (if metal)

Fig. 29 Lightning protection for sail boats.

4.6 THE ELECTRIC CHARGING OF CLOUDS

Numerous observations and measurements have shown that a thundercloud theory must satisfy certain conditions:

1. It must explain a positive charge in the upper region of the cloud and a negative charge in its lower portion and during the

later stage in the cloud's life cycle it must justify the appearance of a small pocket of a positive charge near its base.

2. The theory must clarify why it takes approximately 20 min. to generate enough charge for the first lightning flash and only a few second for the ensuing flashes.

3. The process must be capable of separating a charge of at least 1,000 coulomb per thunder cloud during an average life time of 30 minutes. A typical diagram of a thunder cloud is shown in Fig. 30.

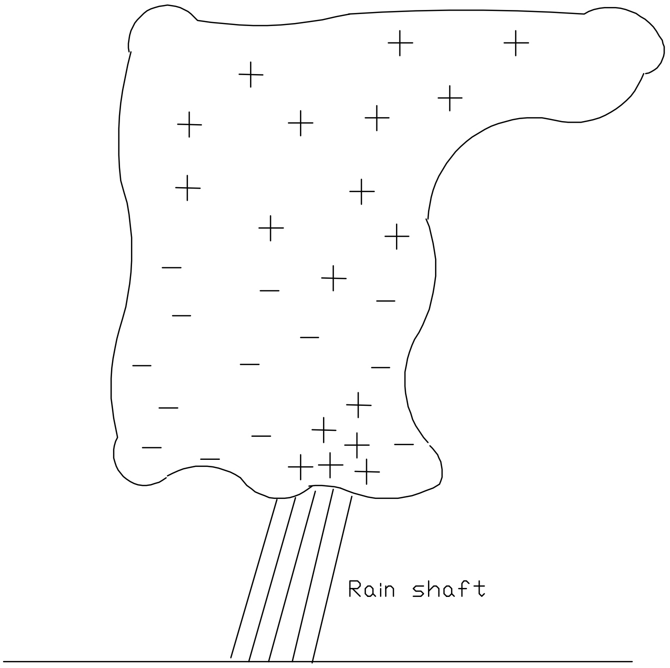

Fig. 30 Typical charge distribution in a thundercloud.

Many theories have been developed ever since Franklin's and D'Alibard's discoveries over 200 years ago. A theory must resort to some kind of energy source and most thundercloud theories seem to fall within two such categories, namely that of gravitational energy and energy provided by convection. Gravitational mechanisms rely on the gravitational potential energy available as precipitation particles fall from the upper part of a cloud to its lower region. Convection mechanisms make use of energy supplied by the intense updraft winds in thunder clouds.

Most theories such as the Wilson mechanism, influence charging, icing and ice splintering mechanisms all fall under the gravitational process. Charge is produced on precipitation particles by one of the above processes in which negative charge is preferentially attached to larger and heavier drops which fall down and separate from the smaller drops causing a large build up of the electric field within the cloud. One serious problem with gravitational mechanisms is that according to the data presented by Israel (see sec. 4.2 and 4.3) an average thunderstorm cell which contains 8×10^8 kg of water must produce 1.7×10^{13} joules of electric energy, which is not possible. Simple calculations show that the gravitational energy available in a cell which contains 8×10^8 kg of water has a potential gravitational energy of $0.6\,mgh / 2 = 1.2 \times 10^{13}$ joules (g is the gravitational acceleration, h the average vertical separation of roughly 5 km between the charge centres and m *is* the total water content). This means that there is not enough gravitational energy available to account for the electric energy present in the cell. If the above reasoning holds true then all charging mechanisms which rely on gravitational potential energy can be ruled out. However, more careful evaluation of existing data and more measurements are needed before such a decision can be made.

Convection mechanisms on the other hand can easily provide the energy and charge needed through the continuous ventilation of near stationary drops. There are essentially only two theories developed so far which are based on convection, the theory by Grenet (1947) and the electrochemical mechanism (Wåhlin 1973). Grenet and later Vonnegut (1955) suggested that natural positive space charge is brought in from below the cloud and carried to the top by convective updraft currents. The positive charge at the top of the cloud will attract negative ions from the outside air volume which are captured by down draft currents and thus assumed to be brought to the bottom of the cloud. As the process continues, the increased electric field will produce more positive ions under the cloud by corona point discharges at the earth's surface which in turn are carried through to the top by updrafts. One objection (Chalmers, 1957) is that it is difficult to understand how the updraft and downdraft currents can distinguish between negative and positive ions in a way that only negative ions can be transported by down drafts and positive ions by updrafts.

In the electrochemical charging mechanism convection currents simply bring natural ions of both signs in from outside and ventilate them through the drop population of the cloud. However, negative ions which are extremely reactive attach themselves to drop surfaces at the bottom of the cloud while the inert positive ions sieve through and are carried by the updrafts to the upper regions. The electrochemical charging mechanism, which is described in Chapter 3, is very effective and despite criticism by Vonnegut, Moore, Griffith and Willet (see section 2.7.3) is the only process which can easily be modelled in the laboratory. The diagram in Fig. 31 shows such a model constructed from two clumps of steel wool. The two sections of steel wool, one representing the bottom and the other the top of a thunder cloud, are electrically insulated from each other. A fan blows air through the cloud model to simulate the updraft winds inside a thunder cloud. A radioactive alpha particle source (Po 210, 500 micro Cu) near the fan feeds positive and negative ions at equal number into the cloud model. Negative ions are immediately captured by the electrochemical process in the bottom section while the inert positive ions pass through and reach the upper steel wool section where they stop together with the air flow. In a model the size of a large football over one hundred volts of charge separation is readily achieved. A small neon light, connected across the cloud model, demonstrates the lightning discharge in a realistic manner. The polonium 210 is readily available from several mail order suppliers and does not require a license.

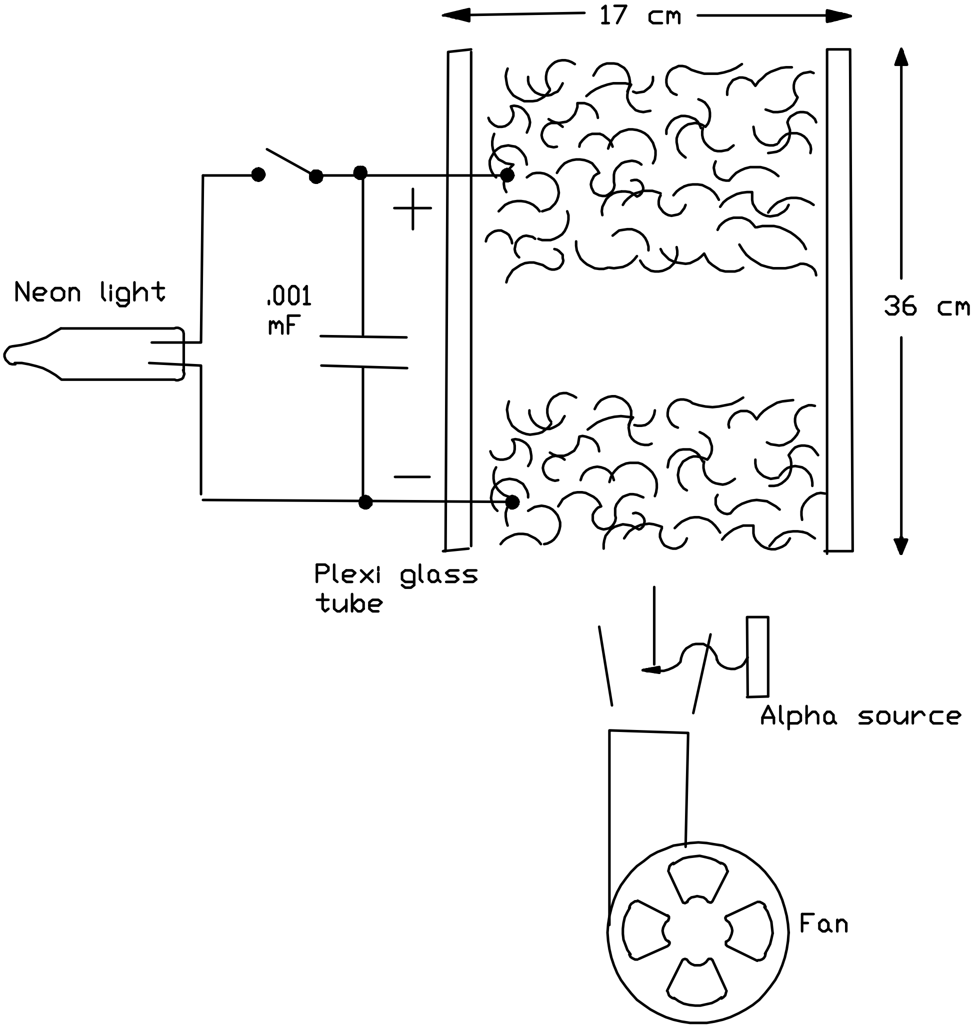

Fig. 31 Laboratory model of thundercloud.

The electrochemical charge separation in clouds has been theoretically scrutinized by Pathak, Rai and Varshneya (1980) who determined that enough charge can be separated from atmospheric ions within 10 minutes to trigger the first lightning discharge. Once lightning has occurred numerous ion pairs are created by the discharge itself which will increase the charge rate drastically to allow for ensuing lightning discharges to occur within seconds of each other. In the dissipating stage when down drafts begin to appear in an organized way, the charging process can reverse and form a small positive charge centre near bottom of the cloud. This explains the small pocket of positive charge near the cloud base which is often observed by investigators (see Fig. 30).

4.7.1 BALL LIGHTNING

Ball lightning is a controversial subject and such a rare phenomenon that many investigators in the field still doubt its existence. Many sightings have been reported dating back hundreds of years, but no convincing photographs or scientific records have ever been produced. Nevertheless, ball lightning has intrigued a great number of reputable scientists and has resulted in many exotic theories. Certain fundamental characteristics of ball lightning has been agreed on which can be described as follows: The geometrical shape of ball lightning is that of spheroid with a diameter of 10 - 20 cm although diameters as small as 2 cm and as large as 150 cm have been reported. The colour of ball lightning varies from white, bluish-white, greenish-white, red or orange-red. The light or glow seems to be steady throughout its duration which might be a few seconds to several minutes. Ball lightning might decay slowly and faint out or disappear in a sudden explosion often causing severe damage. Ball lightning has been reported to pass through window panes and walls, a phenomenon which is difficult to conceive. In most cases it moves silently above ground or along fence wires and power lines. Occasionally it is accompanied by small sparks and corona discharges which indicate the presence of electrostatic charges. Ball lightning appears only during severe thunderstorms that produce intense lightning discharges. It is believed that ball lightning might originate from the ionized lightning channel of super bolts.

Small electric plasma balls have often been observed in submarines when battery banks have been accidentally shorted out. The high short-circuit currents cause intense arcs to appear across contacts of reverse current relays which are fitted with magnetic blow-out coils. Occasionally a fire ball is blown out which will float off into the engine room and remain visible for a short time while decaying by the normal electron-ion recombination process.

As previously mentioned, there are many theories proposed in order to explain ball lightning (Barry, 1980). One such theory is the ring current plasmoid model which is derived from observations and experiments performed with submarine batteries (see Silberg, 1965). The ring current model does not mention what mechanism creates a ball lightning but deals more with the decay time of the plasma itself. Other more exotic theories are the standing wave model which is based on microwave resonance and the crystal model which treats the problem from a quantum mechanical point of view.

4.7.2 THE QUANTUM MODEL

The quantum model is interesting in that it tries to explain how a completely ionized plasma is maintained at very low temperatures, less than 340 C. Neugebauer (1937) used quantum mechanical arguments, derived from well-known free-electron gas models, to show that once a plasma ball has been created it will require very little energy too sustain itself. For example, from the black body radiation of such a plasma it can be determined that a completely ionized ball of 10 cm in diameter would require less than 24 watts and a 20 cm ball about 90 watts to remain completely ionized. The quantum model does not explain where the power comes from that generates the heat or how the plasma was created in the first place.

4.7.3 THE STANDING WAVE MODEL

Cerrillo (1943) and Kapista (1955) advanced a model of ball lightning in which the ionized plasma is produced and maintained by radio frequency oscillations. It is believed that a standing wave is created inside a plasma where the size of the plasma is equal to one quarter wavelength of its cavity frequency. If the radio frequency is suddenly cut off, the ball will produce a weak shock wave when collapsing. If the

RF energy is gradually diminishing then the ball will slowly faint away without any drastic after effects. How to explain the mechanism that is responsible for producing such a high intensity RF field remains a problem.

4.7.4 THE RING CURRENT MODEL

Fire balls formed in engine rooms of submarines have led to the speculation that ring currents might be responsible for the production of plasmoids with reasonable long life times. The long life time is based on the idea that a plasmoid must have a very low ohmic resistance R and a certain amount of induction L, due to the magnetic field of the ring current. The time constant $t = L/R$ is therefore a measure of the plasmoids life time. Fig. 32 shows a diagram of the plasmoid itself. Also, it is not clear how such a ring current might be produced and how it can obtain a spherical shape.

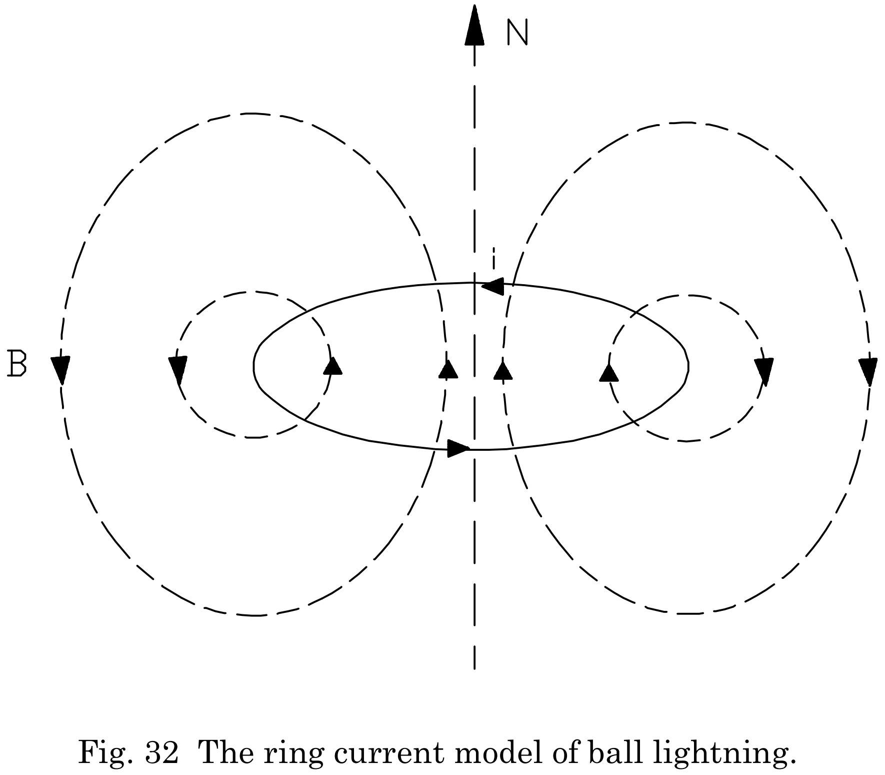

Fig. 32 The ring current model of ball lightning.

4.7.5 THE PINCH EFFECT

The pinch effect is a well known phenomenon. It explains how lightning bolts sometimes break up into small balls called bead lightning or pearl lightning, an event which has been observed by the author in both Africa and Sweden. The strong magnetic field produced by the lightning bolt current can force the flow of charge to bend inward towards the centre of the lightning channel to such an extent that it will choke and pinch itself off. The pinch effect can be better understood if one examines Ampere's law, see Fig. 33, which states that the attractive force between two parallel conductors, which carry currents i' and i" respectively, is equal to

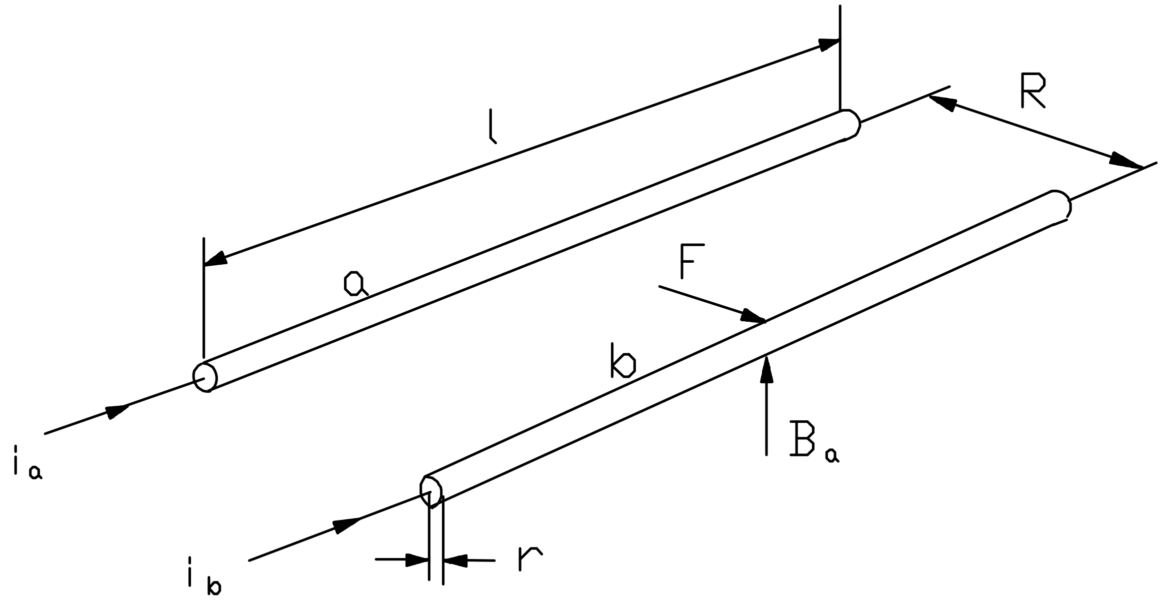

Fig. 33 Two parallel wires which carry currents in the same direction attract each other.

$$F = \frac{\ell i' i'' \mu_0}{2\pi R} = \ell i' B, \tag{13}$$

where ℓ is the length of the current elements which are separated by a distance R. The magnetic field generated by one current at the site of the other is B and μ_0 is the permeability constant.

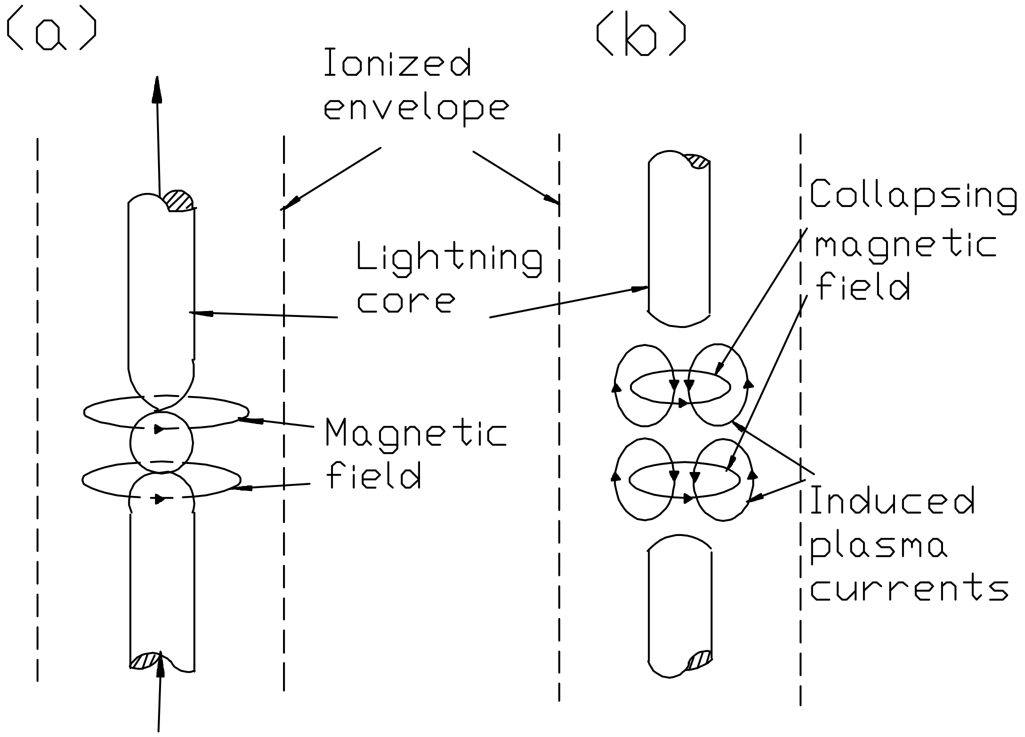

Fig. 34 The pinch effect causes the lightning core to break up into small segments (bead or pearl lightning). permeability constant.

In a lightning channel, where numerous charges or current elements move in the same direction, a mutual attractive force is developed which will cause the lightning channel to contract, see Fig. 34 a. The force of contraction can reach a magnitude where the lightning channel gets pinched off and becomes separated into small spherical sections (beads or pearls) see Fig 34 b. Once the flow of charge is pinched off, the electric current will cease which causes the magnetic field to collapse. The collapse of the magnetic field, on the other hand, will induce eddy currents in the ionized plasma, isolated by the electromagnetic pinch effect, which will maintain the electric discharge and prevent the magnetic field from rapid decay. The plasma ball formed can be pictured as a magnetic ring field trapped inside a toroid electric plasma current, see Fig. 35. One of the conditions required is that the plasma must be completely ionized and super-conducting in order to offer minimal resistance or ohmic loss. Any presence of neutral particles will cause recombination and rapid decay. A collision with a conductive object such as an iron stove, for example, could drastically neutralize the plasma ball in an explosive manner. The theoretical model described implies that that a ball 10 cm in diameter, completely ionized, would contain an ionization energy o 7 kilo joules and eight times that amount for a 20 cm diameter plasma ball. The field of induction set up by the plasma current is

$$B = \mu_0 I, \tag{14}$$

where I is equal to the lightning current at the moment of pinch-off. The energy density of the trapped magnetic field is

$$U_B = \tfrac{1}{2}\frac{B^2}{\mu_0}. \tag{15}$$

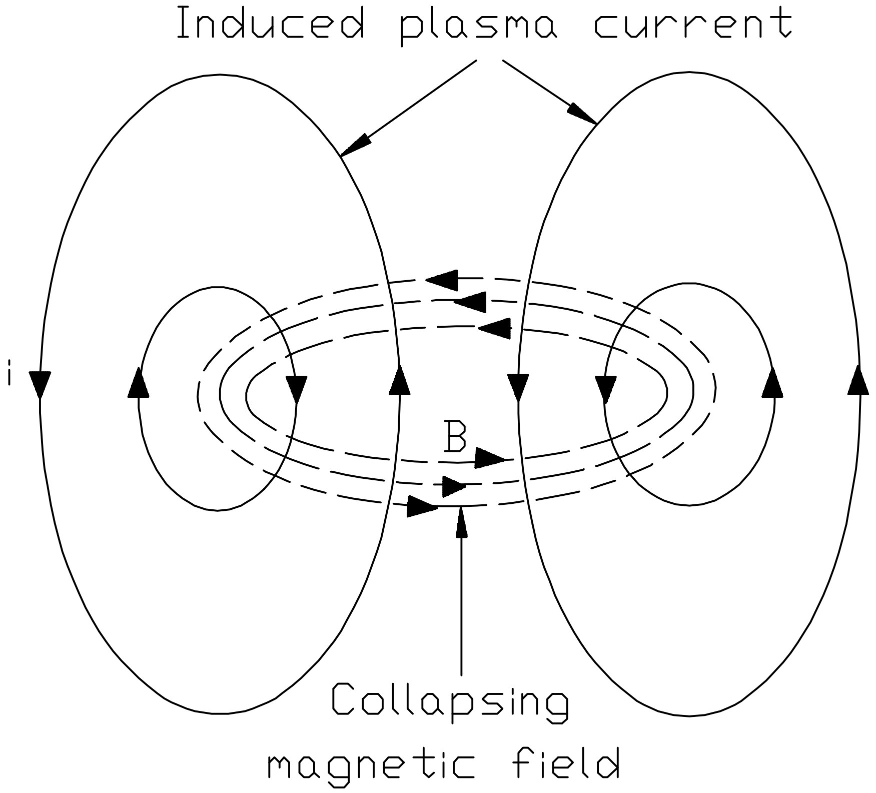

Fig. 35 Magnetic ring field model of ball lightning.

The energy of the magnetic field in a 20 cm diameter ball, produced by a super bolt delivering 400 kA peak current, is equal to about 400 joules. Using Neugebauer's argument that 20 watts is sufficient to maintain a 20 cm plasma ball (see 4.7.2) means that 400 joules will make the plasma ball last for about 20 seconds.

CHAPTER 5

FAIRWEATHER PHENOMENA

5.1 EXPERIMENTS AND OBSERVATIONS

After Lemonnier and Beccaria discovered the fairweather electric field by probing the atmosphere with poles and stretched wires, the first considerable advance was made by de Saussure (1779) who introduced several new methods of measurements. De Saussure constructed an electrometer consisting of two elder pith balls suspended from silver wires inside a glass jar with metallic shielding. When the balls were charged to the same polarity they repelled one another and the amount of deflection was a measure of the electric intensity from a source brought in contact with the silver wires. By raising and lowering an antenna which was connected to the electrometer, de Saussure could show that the electric potential in the atmosphere increased with altitude. Another clever experiment by de Saussure was to throw a lead ball, which was connected to the electrometer by a thin wire, straight up in the air braking the wire and electric connection. The ball, after reaching a certain potential, would take its charge with it and leave an opposite charge behind on the electrometer. De Saussure realized that the electrostatic charge induced on his antennas was caused by a positive static charge in the atmosphere above. He is also believed to have been the first to discover the systematic annual change in the fair weather field, see section 1.5.

One interesting experiment involving the use of kites was performed by Mahlon Loomis, a Washington D.C. dentist, during the period 1862 - 1873. Loomis used two kites at the end of stranded bronze wires about 600 feet in length. The wires were both grounded through sensitive galvanometers but the ground connections could be broken by means of telegraph keys. Loomis was able to transmit morse code messages from one antenna to the other over a distance of up to 16 miles without the use of any external electric source. In the early 1870's Loomis was supposed to have achieved voice communication using a Berliner's microphone. He applied for government funds for a classified military project, but since the Civil War was over, Congress would not appropriate the money. What is important is that Loomis must have been first to achieve wireless communication but obviously not by radio but by disturbing the electric fairweather field. A member of the family, Thomas Appleby, has published a book on Loomis' experiments and copies of patents issued (Appleby 1967).

In the beginning of this century it was discovered that ions cause the atmosphere to be conductive and new theories were needed to explain how charges on the earth-atmosphere system are maintained. It has been suggested that the earth's rotation in its own magnetic field could produce a magneto effect which would drive the electric fairweather circuit, or that evaporation of water from the earth's surface might charge the earth negatively and the water vapour positively by some unknown process. In fact, Volta (1800), advanced an electrification theory based on vapourization and condensation of water in order to explain charge production in thunderclouds. Today, although the question still remains open, there are primarily two theories dealing with the problem of the origin of the electric fairweather field. One is the closed circuit hypothesis which asserts that all thunderstorms around the world recharge the earth-ionosphere system. The other theory is the electrochemical process which is an electrostatic mechanism based on the preferential capture of negative ions by the earth's surface. The amount of charge produced by all thunderstorms around the earth at any given time is certainly adequate for replenishing the fairweather charge and the variations of the fairweather field as a function of GMT seems, to a certain extent to follow the variation in the world-wide convection pattern as shown in Fig. 15. The electrochemical charging mechanism, on the other hand, suggests that if a sphere the size of our earth is ventilated by ionized air then, by extending the curve in Figure 23, it will charge to approximately one million coulombs, a value that conforms with the measured charge on earth. The curve, of course, shows the charge on water surfaces, but it should be remembered that two-thirds of the earth is covered by water. Also, the earth's surface is constantly being ventilated by winds carrying natural ions which are supplied at a steady rate by cosmic and radioactive decay. The excess positive ions left behind near the earth's surface are carried aloft by convection and by eddy diffusion, thereby mixing through the atmosphere to heights of several kilometres. Lifting the positive ions from near ground to higher altitudes stretches the electric field lines and increases the potential. This is analogous to the contact potential experiments performed by Lord Kelvin (see section 3.2) in which the potential is increased by separating the two metal plates in his condenser apparatus. The electrochemical charging mechanism is strongly dependent on convection and is also expected to follow the world-wide convection pattern as shown in Fig. 15.

About 90% of the electric fairweather field is confined to altitudes of less than two kilometres, or that part of the atmosphere referred to by meteorologists as "the mixing region". Integrating electric field strengths up to a variety of heights will give the values of electric potentials at varying altitudes. At an altitude of two kilometres, for example, this potential is normally about +200 kilovolts with respect to earth. At the University of West Virginia a team of physicists have managed to build an electric motor driven by the electric fairweather field utilizing an antenna extended high into the atmosphere (Jefimenco, 1971).

Since the electric fairweather field is enhanced by convection and mixing in the fine weather atmosphere, measurements of electric field strength at varying altitudes and at diverse locations may give important information about the atmospheric mixing processes in terms of local weather forecasting. Large scale forecasts might also be possible because the electric fairweather field exhibits a slight variation every twenty-four hours that is synchronous for all parts of the world. The diurnal fluctuation is linked to the fact that continental land masses of various sizes are exposed by the daily rotation of the earth to the heat of the sun and this results in convection and enhancement of the electric field (see Fig. 15).

Many other phenomena in the atmosphere can be explained by the electrochemical effect. For example, airplanes in flight, are subject to ventilation by atmospheric ions and thus charge accordingly. A Boeing 747 should reach a potential of over -100 kV by rough estimation based on Equation (7). A large body such as a zeppelin could reach a potential of several hundred thousand volts in the naturally ionized atmosphere, making a voyage in a hydrogen airship a highly dangerous venture, as proved by the mysterious disintegration of the Hindenburg in 1937.

Charging also occurs inside combustion chambers of engines where ions are formed at high temperatures. By means of the electrochemical process, ions of one sign (normally negative) are preferentially captured by engine walls, permitting those of opposite charge to escape in the exhaust stream. Exhaust charging is a common phenomenon in our modern world. Not only do aircraft charge by this process but motor vehicles on rubber tyres become electrostatically charged even at stand still while idling. Exhaust gas from engines and smoke stacks produce excessive positive space charge over heavily populated areas. Aircraft flying through such areas during landing approaches become positively charged in contrast to what is normally experienced during regular fairweather flights. The explanation is that if an aluminium fuselage is subject to ventilation of an abnormally high positive to negative ion ratio, where $N^{+} / N^{-} = 1.3$ or more then, according to Fig. 11, the aircraft must charge positively.

Exhaust charging also plays an important role in geophysical phenomena such as volcanic eruptions which often produce lightning and during earthquakes where a glow or earthquake light is emitted along fissures and cracks.

It is difficult to say whether atmospheric electricity has any practical applications at the present time. The structure of the fairweather field and how it is linked to convection and turbulence in the atmosphere might be of meteorological interest. The fairweather field as an energy source is too feeble for practical purposes and will only yield about 1μwatt per m^2 at the earth's surface, or approximately one billion times less than solar power.

Also worthy of consideration is the fact that for the past 50 years scientists have been looking for physiological effects caused by ions and electric fields in the earth's atmosphere. Many experiments have already been performed that seem to show such a relationship. However, one still lacks concrete physical evidence and much more work is needed before one can determine the significance of atmospheric electricity with respect to one's state of health. There is also a growing interest in planetary atmospheric electricity as a consequence of data sent back from space probes.

5.2.1 EXHAUST CHARGING

Lord Kelvin was aware of the "wonderful agency in flames and fumes" causing deviations in his electrometer readings when passing through his condenser plates. Earlier (c. 1800) Volta discovered that static charges on conductors changed rapidly when exposed to candle flames or lit fuses. Modern science explains that ions produced from combustion increase the conductivity of the surrounding air causing potential gradients to change more rapidly. An increase in ion population also enhances the electrochemical charging effect. In combustion chambers of engines, where numerous ions are formed, the electrochemical charging effect often becomes a nuisance and in some cases a hazard. For example, fuel might ignite while attempting to refuel a charged vehicle or aircraft. During combustion negative ions are usually captured by the engine walls while the positive ions are carried away by the exhaust stream. The result is an accumulation of negative charge on the vehicle. Fig. 36 shows an experiment performed by the author with a diesel car blocked up on insulators. A charging current of 2 micro amperes was registered at fast idle of the 2.25 litre engine. Similar experiments were carried out on small aircraft and negative charging currents of up to 10 micro

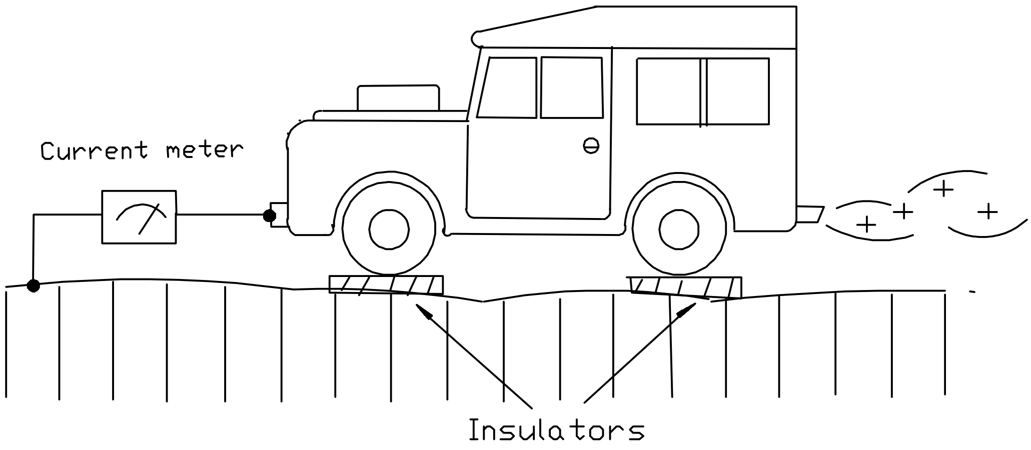

Fig. 36 Exhaust charging of a motor car.

amperes were measured. Exhaust charging can easily produce static potentials of several thousand volts. Severe electric shocks can be obtained from large aircraft not fitted with static eliminators or by touching metal hooks lowered from hovering helicopters.

5.2.2 VOLCANIC ERUPTIONS

Volcanic plumes charge to high potentials and, just like thunderstorms are accompanied by a display of lightning discharges. Not much is known about the charging mechanism involved and very few measurements are available that can shed light on the electric structure of such plumes. Space charge measurements at ground level near volcanic eruptions reveal negative charge on precipitating ash particles which, according to the mirror-image effect (sec. 2.5), means that there is positive charge on the plume above. Opposite signs have been reported and the conclusion is that ash particles from different

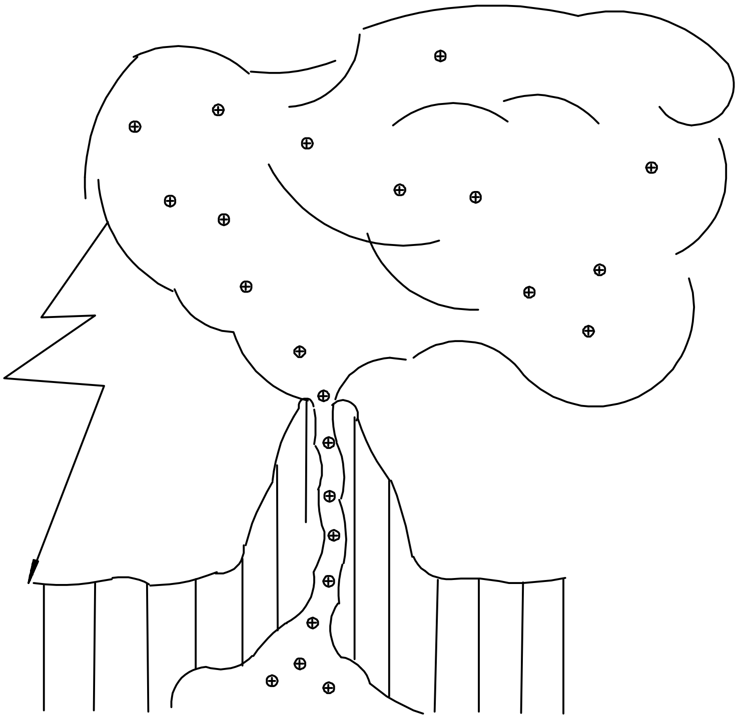

Fig. 37 Electrostatic charging of volcanic plume.

volcanoes charge differently. It is difficult if not, impossible to determine the dominating sign of charge in plumes from ground measurements because sign reversals, due to lightning discharges and mirror-image effects, are always present.

It has been suggested that volcanic plumes charge by frictional effects between small and large ash particles and that separation of charge is enhanced when larger particles fall out of the plume. Other charging effects to be considered are convection mechanisms and exhaust charging. Convection and condensation of water vapour as in regular thunderstorms can also explain the build-up of charge in the plume. Exhaust charging might play a dominating role since numerous ions must be present in the hot interior of the volcano. Negative ions can be expected to be adsorbed to the inside walls before reaching the earth's surface while the more inert positive ions are carried up with the plume, see Fig. 37.

Electric measurements during volcanic eruptions are obviously very difficult to perform if not dangerous. Also, volcanic eruptions are not as common as ordinary thunderstorms and data is, therefore, very sparse.

5.2.3 EARTHQUAKE LIGHT

Earthquake light is a rare phenomenon and its existence is still being questioned by some scientists. It was not before 1965, during the Matsushiro earthquakes in Japan, that actual photos were obtained of earthquake light (Yasui, 1968 and 1971). Earthquake light has been observed as a faint red and white glow for hundreds of miles around just before the earth begins to shake and crack along a fault zone. There has been much speculation on what might cause earthquake light. One fact is certain, that the light must come from atoms that have been ionized or excited by some mechanism triggered by the earthquake. The light is emitted when a detached or excited electron falls back again into its stable orbit around the atomic nucleus. Ionizing collisions in air, between atoms and particles or between atoms and photons, by far outnumber exciting collisions. Ionization of air molecules, to the extent where the sky will glow for miles around, could be caused by corona discharges near the earth's surface or by ionized exhaust gas escaping from the hot interior below the earth's crust. It has been suggested that frictional heating of a shear zone will occur during an earthquake and that the frictional heat will lead to vaporization of water in and near the shear zone (Lockner *et al*, 1979). The result would be a drastic increase in the electric resistivity of the rock throughout the shear zone coupled with some sort of charge separation due to the evaporation of water. After enough charge has been collected in the shear zone, corona discharges along its top edge would stream into the atmosphere like St. Elmo's fire. One problem, however, is that corona discharges draw a great deal of current ($100\mu A$ per cm^2) and it is difficult to understand how such a current can be supplied over a highly insulating rock surface. Also, corona is quite a noisy phenomenon unlike silent earthquake light.

Other theories take into account the piezoelectric effect in which strong electric fields are presumed to form in quartz rich crustal rocks subject to high mechanical pressures (Mitzutani et al 1976). Again high electric fields along the fault zone would create corona and St. Elmo's fire. One interesting idea is that the sparks and crackle from piezoelectric discharges in the interior would create electromagnetic radiation, similar to that of old spark transmitters used before the invention of vacuum tubes and transistors, to the extent that the air above the earth's surface would ionize and glow. The power of the electromagnetic radiation must be high enough to ionize air and make it glow. One would expect that such a powerful electromagnetic radiation, that will cause electric breakdown of air, would cook everything else in sight and heat the earth’s surface to extreme temperatures.

Perhaps the explanation of earthquake light has nothing to do with electrostatic charging but simply the escape of gases such as the radioactive element radon. According to the Handbook of Chemistry and Physics there is about 1 gramme of Radium per square mile of soil to a depth of 6 inches which releases radon into the air. During an earthquake, radon is released from considerable depths and brought to the earth's surface in large amounts. Each radon atom is capable of producing one million ion pairs. The handbook states that when radon is frozen below its melting point it will exhibit a brilliant phosphorescence which becomes yellow as the temperature is lowered and orange-red at the temperature of liquid air. The light is from ionizing collisions between alpha particles and atoms. It can be expected that when large quantities of radon escape it might be accompanied by other foreign gases trapped far beneath the earth's surface. The gases, unnoticed by humans, could easily be detected by animals and the unfamiliar scent might have a disturbing effect on their behaviour. It is believed that the Chinese are able to predict earthquakes and it is well known that they are monitoring animal behaviour for this purpose.

5.3 BIOLOGICAL EFFECTS

The exposure of human and animals to ions and strong electric fields has only very recently become an important issue in many countries. The Soviet Union, for example, has very rigorous rules and regulations concerning the exposure of workers to electric fields in power stations. In Sweden studies have been carried out on the effects of ions and strong electric fields on humans under different working environment (Lövstrand et al 1978 and Backman, 1979). Many other countries are and have been involved in similar research. The exposure to electric fields is in itself not considered important, but the production and movement of ions in electric fields are of chief concern. Ions carry charge to the body; they are also inhaled and trapped in the respiratory system. The question is; do ions effect our physiological or psychological well being? So far, numerous tests have been performed in different environments where ion concentrations and electric fields have been changed under controlled conditions. ECG and EEG recordings, blood and urine tests as well as blood pressure and neurasticthenics (headache, nausea, insomnia, tiredness) have been monitored by many investigators but no conclusive results have ever been reported. There are many who believe that negative ions affect them in a pleasant way and that an excess of positive ions have an adverse effect. The author knows of physicists who installed negative ion generators in their cars and which are turned on during long trips in the event their wives and children become rowdy. One atmospheric electrician confided that occasionally he turns on a positive ion generator, which puts his wife to sleep, so that he can watch a TV programme of his own choice! It has been established that ion currents produced on the human body as a consequence of the electrostatic fields, are too insignificant to

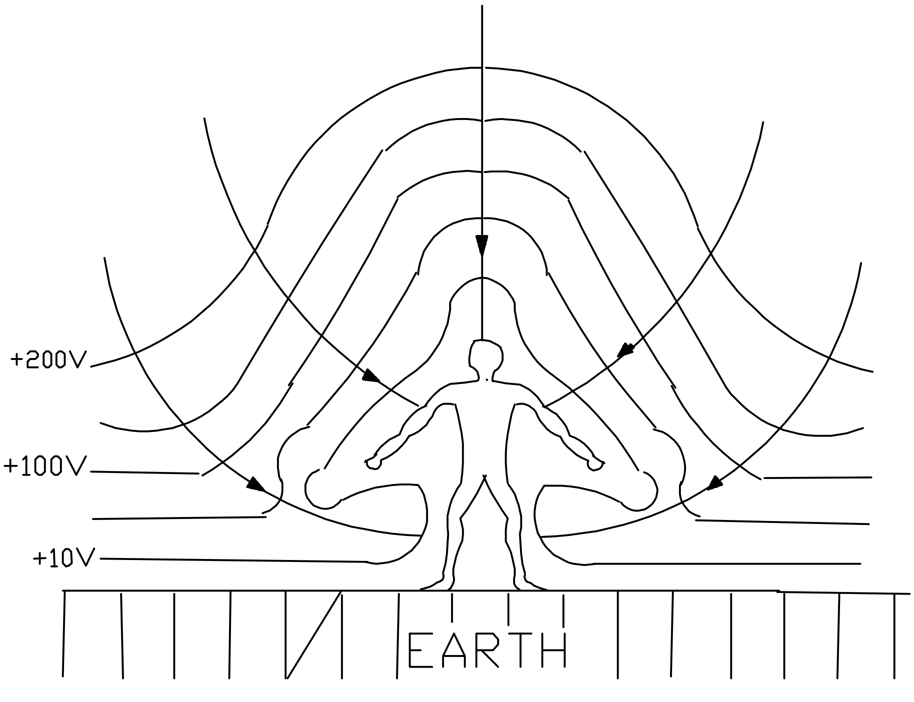

Fig. 38 A person exposed to the electric fairweather field.

cause any noticeable effects (Backman, 1979). A person exposed to the normal fairweather field, see Fig. 38, will perhaps collect a positive ion current of 10^{-13} ampere on his head. Internal brain currents are millions of times larger and it is hard to believe that the external ion current will have any effect on the function of the brain. The effect of inhaled positive and negative ions on the respiratory system is unknown. It is possible that negative ions might have a stimulating effect on the cilia in the alveoli. The amount of ions inhaled and the sign, whether negative or positive, is strongly dependent on the electrostatic charge carried by the person in question. For example, a person walking across a carpet in an office might charge negatively to several hundred volts. The strong electric field will repel negative ions and allow only positive ions to enter the respiratory system. In fact Bach (1963) has demonstrated this effect using an insulated mannequin head biased at different electric

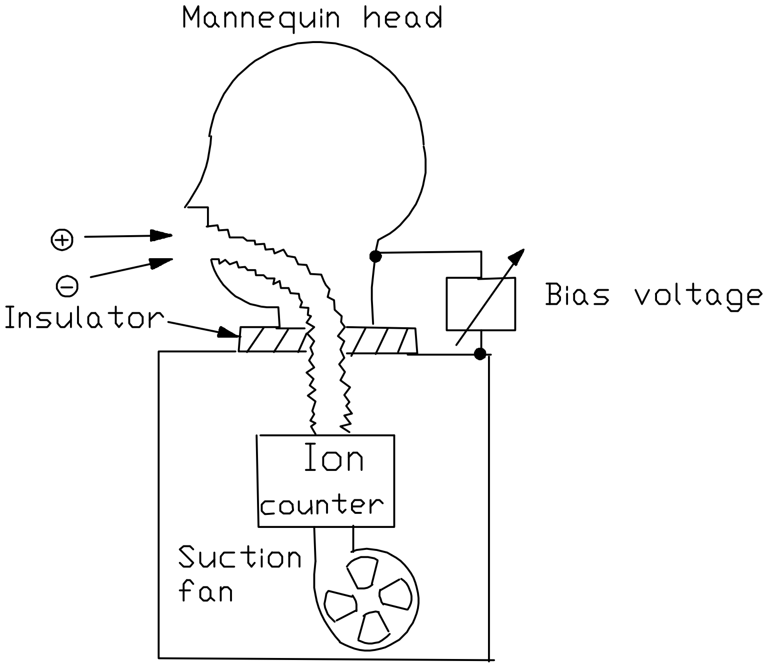

Fig. 39 Apparatus intake through the respiratory system.

potentials, see Fig. 39. The result is shown in Fig. 40 where N' represents the increase in either positive or negative ions as compared to the normal ion ratio at zero bias.

The belief that ions might have a bioelectric effect on humans and animals probably started in the Alps of Europe. It has always been noticed that during the föhn when dry winds sweep down the mountainside some people become depressed or develop head-aches and that dogs and chicken tend to hide away. It has also been observed that the positive space charge or positive ion concentration in the atmosphere during the föhn is very high which is probably one reason why positive ions are blamed for causing discomfort amongst mountain villager. Another observation is that warm mountain winds, like the föhn, are very dry which can cause considerable discomfort as explained by Tromp (1969).

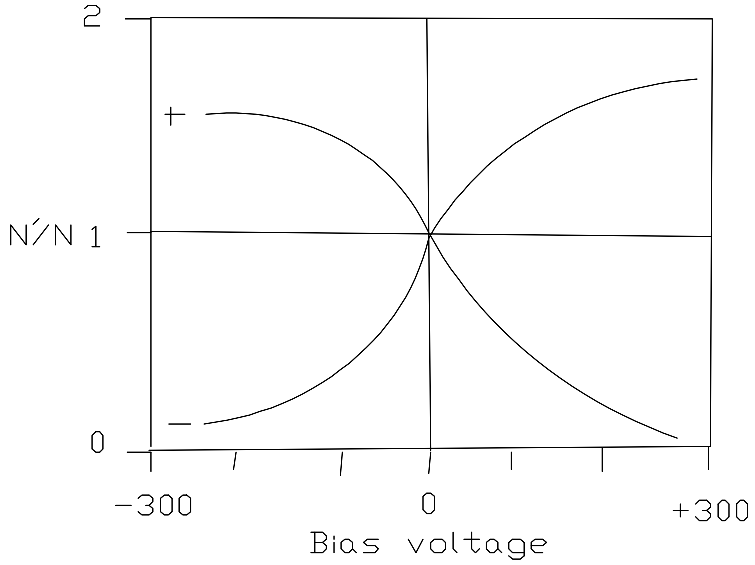

Fig. 40 Ratio of charged ions inhaled at different bias voltages.

Warm mountain winds occur when cold air masses pass over mountain ranges, and when the cold heavy air reaches the other side it falls down, compresses, and heats up again. When cold air warms up it is capable of holding more water vapour and its relative humidity decreases. The warm dry air will take up water from the surrounding environment and dry everything in sight. The Indians in the Rocky Mountain region call the dry mountain wind "the Chinook", or snow-eater, because of the rapid disappearance or vaporization of the snow in its path.

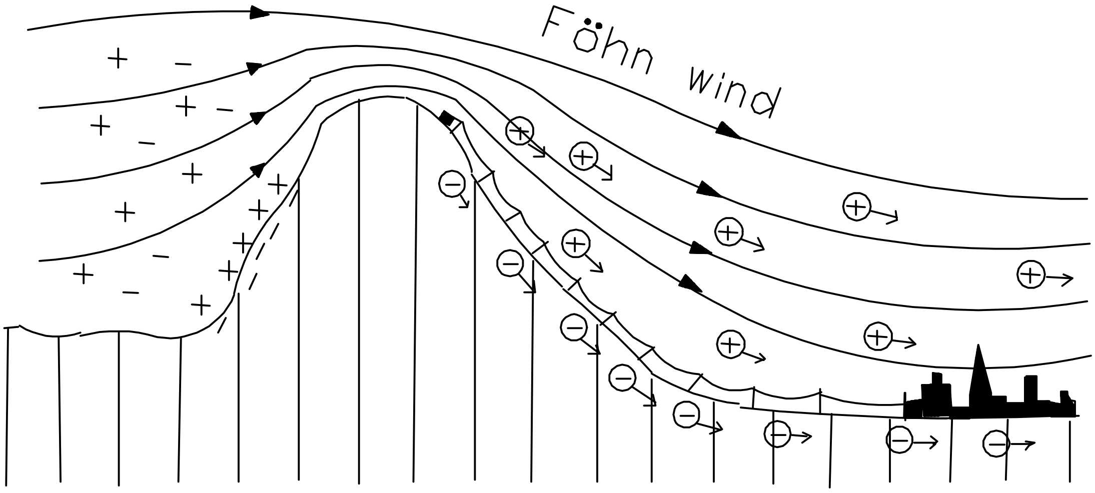

Fig. 41 Dry mountain winds (föhn or Chinook) and positive ions.

The high positive ion concentration can be explained by the fact that winds ventilating the other side of the mountain loose negative ions which are depleted by the electrochemical effect while the excess inert positive ions will follow the air stream over the top, see the diagram in Fig. 41. As the positive space charge moves down the mountain side it is followed by an induced negative image charge along the earth's surface. The induced negative charge is particularly noticeable on telegraph lines going up and down the mountain side. A protruding telegraph line will carry induced negative charge down the mountain which is registered as an electric current in the line flowing up the mountain.

5.4 THE ELECTRODE EFFECT

The electrostatic fairweather field near the earth's surface exhibits some very peculiar variations ascribed the so called electrode effect.

For example, the electric fairweather field can be slightly higher just a few metres above the earth's surface than at the earth's surface itself. This is due to a sudden increase of negative ions in a thin layer above the earth's surface. From Fig. 12, section 2.2, it was explained that the fairweather field can be thought of as electric field lines which connect the excess positive ion population in the atmosphere with trapped negative charges on the earth's surface. If, however, a cloud of negative ions was suddenly formed near the earth's surface it would mean that some of the field lines in Fig. 12 do not reach ground but will terminate on those negative ions. The number of field lines per unit surface area would, therefore, be less at

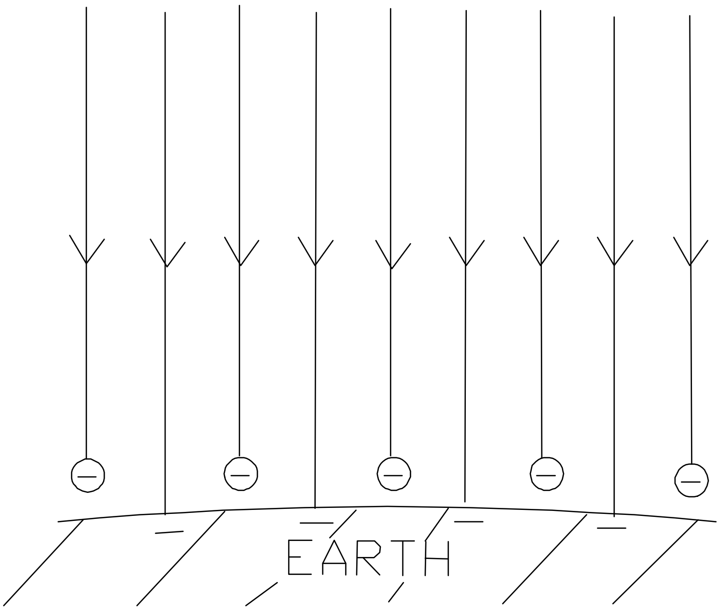

Fig. 42 The electrode effect.

the surface and higher just above the thin negative ion cloud see Fig. 42. A sudden increase in negative ion concentration near the earth's surface could be caused by the passage of a cloud above with a strong positive charge centre. The passage of a positive charge above would temporarily force positive ions to the earth's surface leaving an excess of negative ions behind in a thin layer above ground. The reverse effect is also possible when a negative charge centre passes overhead in which case there is a considerably stronger than normal electric field very near the earth's surface. It is also believed that an electrode effect exists even in the absence of clouds. The idea is that, since positive ions congregate towards the earth's surface because of the normal fairweather field and since there is no supply of negative ions leaving the surface, an imbalance must occur with a higher than normal positive ion concentration near the earth's surface. Numerous measurements, however, do not clearly show such an effect.

5.5 INTERPLANETARY STATIC ELECTRICITY

That static electricity is present on other planets has been established by instrumentation carried on board unmanned space probes. For example, lightning discharges have been observed on Jupiter by the Voyager spacecraft (Smith et al, 1979, Cook et al, 1979 and Gurnett et al, 1979). There is also a strong belief that lightning might occur on Venus.

The rate of current carried by cosmic ray particles to the earth is about 0.2 amperes. This means that the earth will charge 280 volts per second positively unless there is a flow of electrons to the earth at the same rate. The electrons might come from the sun, in which case the sun must charge positively. The existence of strong electrostatic fields in interplanetary space is an untouched field of research. There are problems with orbiting spacecraft building up electrostatic charges which can have adverse effects on measuring instruments and data collected. Extreme charging might occur on space vehicles when returning to earth through the highly ionized layers of the ionosphere. The ventilation of ionospheric ions in combination with the electrochemical charging effect might cause the plasma discharge observed during re-entry, This makes radio communication to from the space vehicles impossible.

CHAPTER 6

INSTRUMENTATION AND MEASUREMENTS

6.1 MEASUREMENTS

Atmospheric electricity is a field that is very easy to get into because it does not require a large capital investment for measuring equipment. It is a difficult field, however, when it comes to the understanding and interpretation of the various measurements. For example, despite vast data collected over two centuries, there is still no agreement on how thunderclouds charge or how the earth-atmosphere fairweather field is maintained.

Today's advanced technology offers eloquent and accurate instruments with fast response time. It is now possible to construct simple electrometers and electromagnetic radiation detectors using chip circuitry. Video recording equipment is becoming a household item that also has great potential for lightning research, making it possible to view flights of lightning bolts in micro-second intervals.

6.2 ELECTRIC FIELD MEASUREMENTS

Most common are measurements of the electric field at or near the earth's surface. The potential gradient at the earth's surface varies in many ways but usually averages about +100 volts per metre during fine weather in the absence of clouds. Except for the normal diurnal variations of the fairweather field (see sec. 2.2), strong rapid changes are often registered which might be due to electrically charged dust or heavy pollution. Marked changes in the fairweather field have been observed before the onset of fog or during temperature inversions in the lower atmosphere. Many of the above field effects and their causes are still not fully understood. The electric field below thunderclouds are usually very strong and easy to detect. Rapid field changes occur during lightning discharges and field reversals are common. Field measurements below thunderclouds reveal such features as charging rates, electric polarity and discharge rates to mention a few. There are several ways one can measure the electric field in the atmosphere. A simple approach is to connect an electrometer to an antenna of some sort such as shown in Fig. 43. The antenna can be a conducting disc placed at a predetermined height

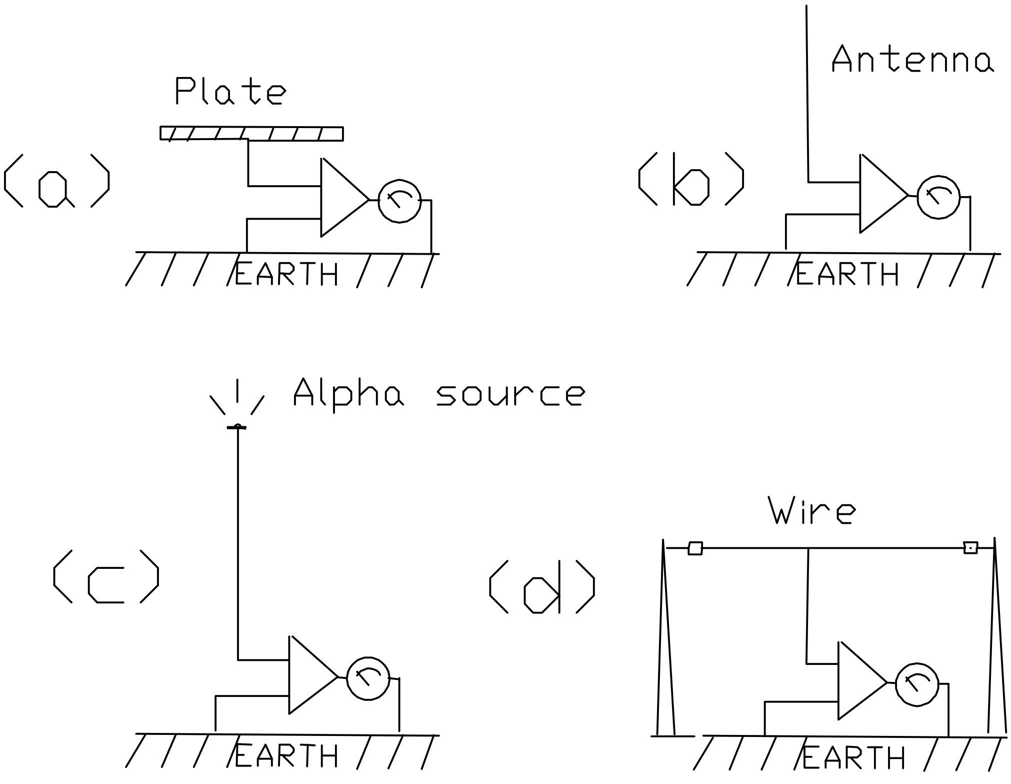

Fig. 43 Different antenna configurations for electric field measurements in the atmosphere. (a) Disc antenna. (b) Whip antenna. (c) Whip antenna with radioactive probe. (d) Long wire antenna.

above ground (Fig. 43a). The disc will charge and reach a potential which is equal to or very near the potential of the atmosphere at that height. The diagram in Fig. 43b shows an electrometer connected to a short whip antenna which gives voltage readings that are difficult to calibrate since the antenna whip will protrude through many levels of electric potential. An ion-producing radioactive alpha source at the tip of the whip antenna, see Fig. 43c, will increase the electric conductivity in the air near the tip and ensure a better accuracy of potential measurement as a function of height. A long wire, suspended above ground at predetermined levels, will give accurate readings of atmospheric potentials as a function of height, see Fig. 43d.

6.3 THE ELECTROMETER

There are many excellent electrometers commercially available, but since the electronic market of today can offer a variety of sophisticated integrated circuits, it makes simple home made devices nearly as

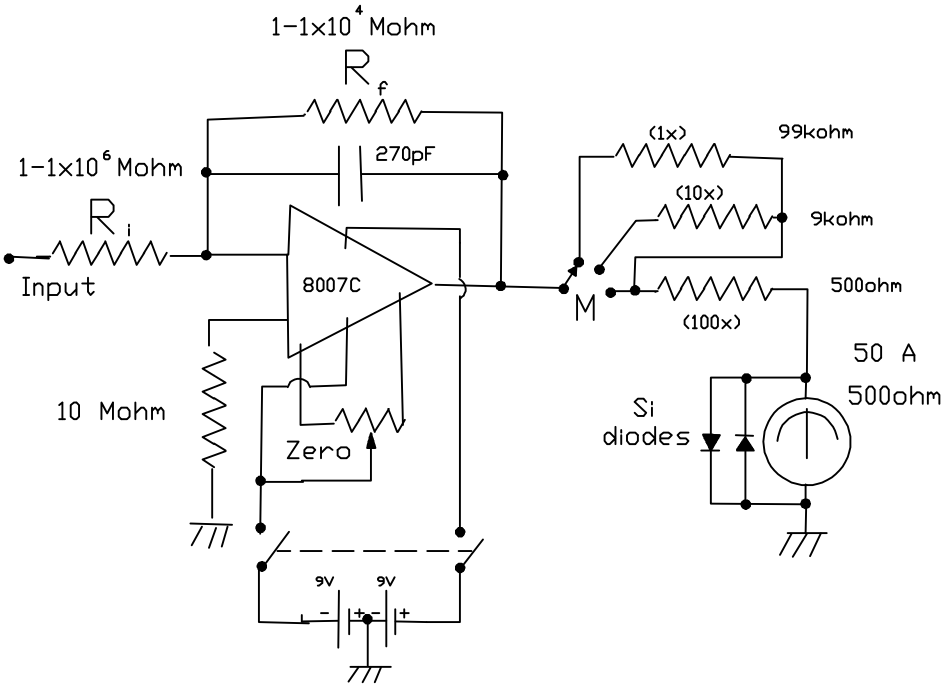
1-1x10^4 Mohm
R_f
270pF
1-1x10^6 Mohm
R_i
Input
8007C
10 Mohm
Zero
9V
9V
(1x)
99kohm
(10x)
9kohm
M
500ohm
(100x)
50 A
500ohm
Si
diodes

Fig. 44 Electrometer circuit.

effective. For example, the circuit in Fig. 44 has been used by the author in many experiments including the Gerdien cylinder measurements described in section 2.1. It uses two 9 volt batteries and is portable, which makes it convenient for experiments in the field. The electrometer is mounted in an electrically shielded metal box.

Measuring electric fields and field changes during thunderstorms will produce many interesting results. Lightning discharges cause rapid field changes and polarity reversals which can be registered on a chart recorder. The recovery time of the electric regeneration process in clouds is also a very interesting feature that can be studied with the help of a recorder. During thunderstorm measurements it is recommended that the electrometer in Fig. 44 be fitted with an R C network which increases the time constant of the circuit to several seconds. Because of the strong fields produced by thunderstorms, it is also advisable to decrease the sensitivity of the circuit, by changing R_i and R_f. A typical circuit modification for thunderstorm measurements is shown in Fig. 45.

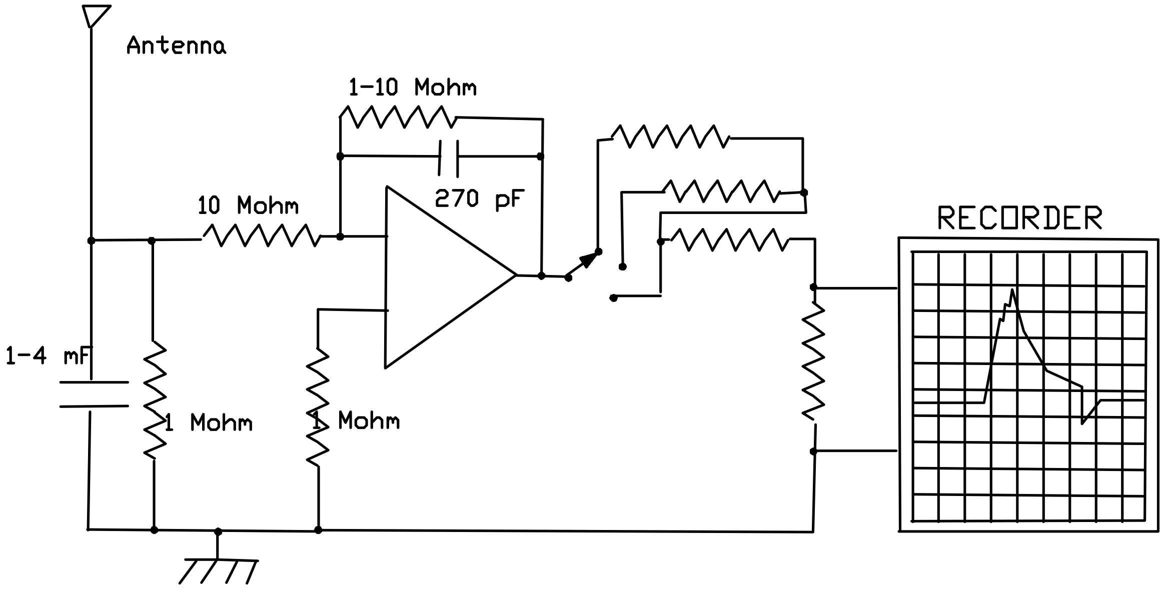

Fig. 45 Circuit for electric field measurements of thunderclouds

6.4 THE FIELD MILL

The electric field mill is a device based on electrostatic induction. It consists of one or two electrodes which either rotate in an electrostatic field or become periodically exposed to a field by rotating vanes. Fig. 46 illustrates a cylindrical field mill which consists of two cylinder halves that are electrically insulated from each other. An electric motor rotates the two halves in the electric field to be measured so that they become alternately exposed to both the positive and negative direction of the field. The result is that an alternating (ac) signal is generated across the two halves which can be easily amplified. The rotating shutter field mill, on the other hand, comprises a stationary electrode, which becomes periodically exposed to the external electric field through a rotating grounded disc (see Fig. 47). A variation on this type of field mill is a stationary grounded cover plate with a rotating

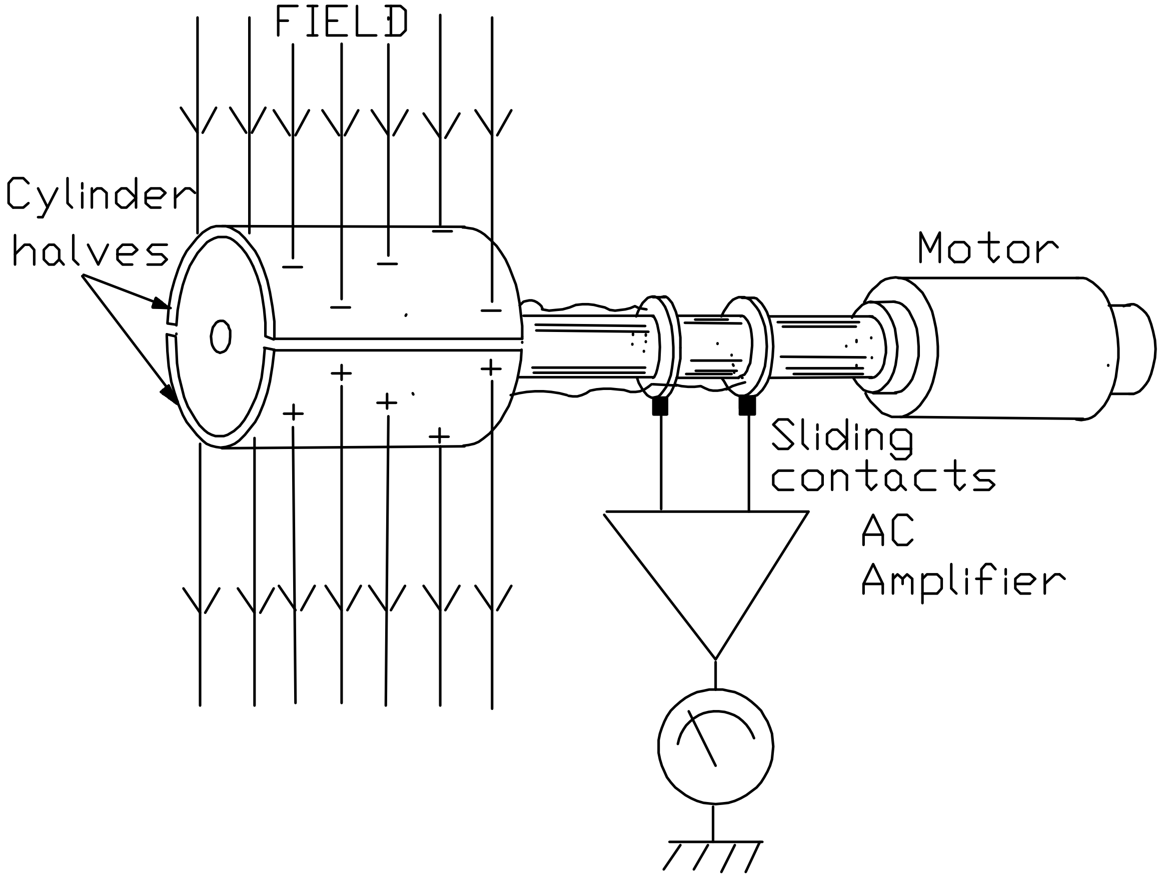

Fig. 46 Cylindrical electric field mill.

disc electrode. Although not commonly used, the cylindrical field mill has the advantage that when mounted in a fixed position it can also indicate the direction of the field. This is accomplished by measuring the phase shift of the ac signal relative to the orientation of electric fields need to be measured. The rotating shutter type field mill has to be pointed towards the source of the field in order to obtain a maximum reading. This feature is useful when searching for static charges that might cause problems in the laboratory or industrial plant. The rotating shutter field mill is commercially available in several countries.

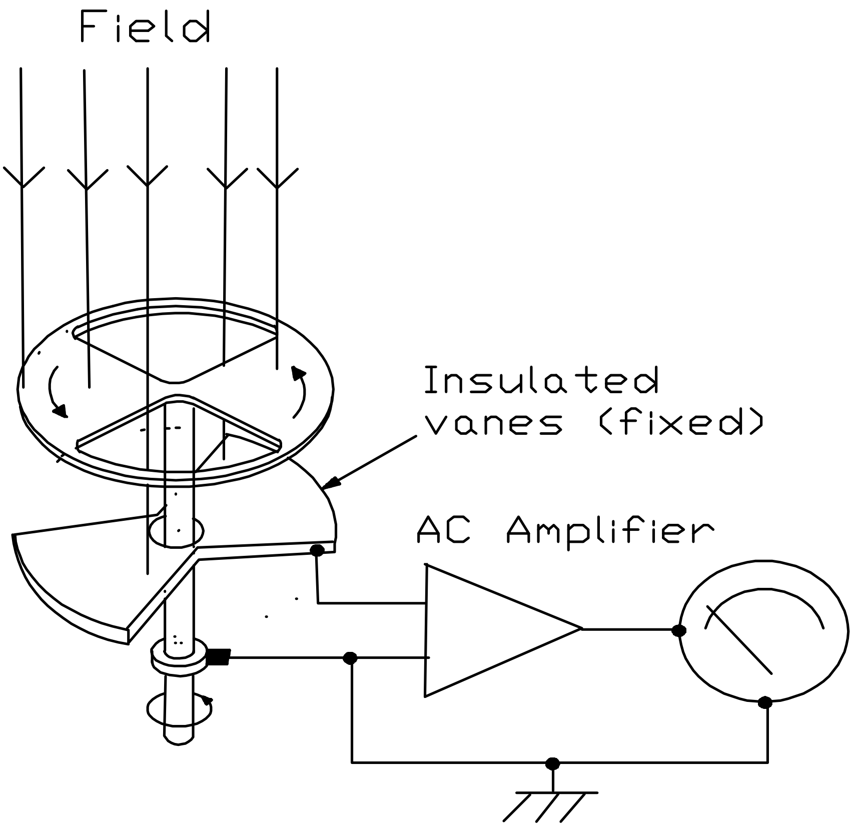

Fig. 47 Rotating shutter electric field mill.

6.5 ELECTROMAGNETIC DETECTION

The crackling sound of atmospherics generated by lightning is familiar to everyone who listens to radio. "Sferics" interference is usually stronger in the low frequency band (10 - 500 kHz) were the electromagnetic energy spectrum of lightning is most powerful. There is also an interest in the higher frequency bands of lightning radiation, around 10 - 30 MHz, relating to corona discharges and fast field changes. An attractive feature of atmospherics is that it enables long distance detection of lightning storms. Radio receivers with directional antennae are used to determine the position and movement of thunderstorms. Several stations can accurately pin-point storms by triangulation. Loop or frame antennae, such as seen on ships, are used and a typical electric block diagram is shown in Fig. 48. Experience has shown that the shape of electromagnetic pulses received from lightning changes with distance, which makes it possible to estimate the distance from the antenna to the discharge. Some aircraft are fitted with "stormscopes" utilizing the above principles to determine the direction and distance of nearby intracloud discharges in order to avoid dangerous confrontation with lightning. Radar echoes of lightning bolts can be seen as reflections superimposed on echoes of heavy precipitation in clouds. It is believed that the ionized channel, produced by lightning, is large enough to cause reflection of radio waves. Radar observations of antenna for

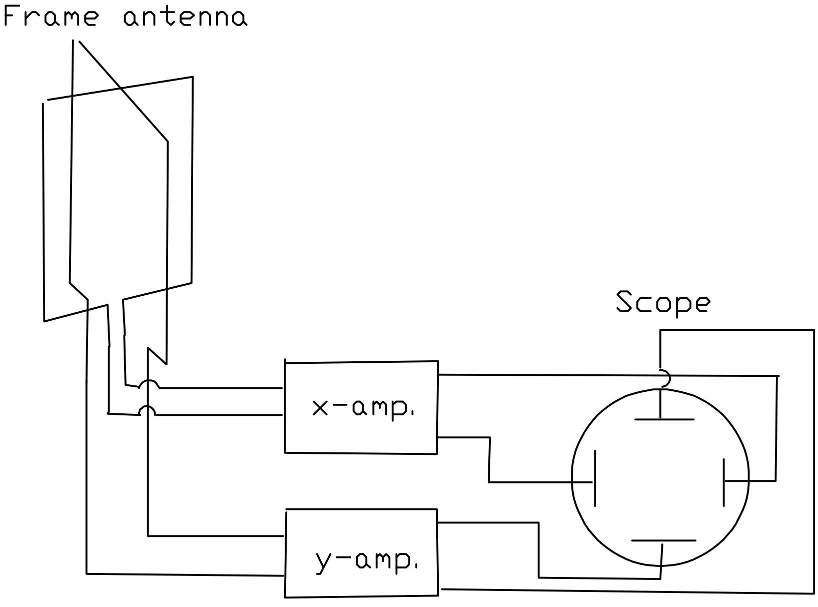

Fig. 48 Loop or frame lightning detection.

thunderstorms bring out many interesting features which might aid in the understanding of the physical processes at work in clouds. For example, radar reflections from precipitation appear several minutes before any visible electrical activity. The radar reflections rise at first together with the cloud top but start to descend as soon as lightning occurs. At the same time lightning commences, precipitation becomes visible and starts to fall out of the cloud. One question often asked is whether lightning induces gushes of precipitation; or vice versa.

6.6 THE FUTURE

There are still many problems outstanding in the field of atmospheric electrostatics. New and better measuring devices will help solve some of the problems still at large, but equally important is the development of new ideas and techniques.

Electrostatic fields and ions in the atmosphere and their effect on our well being is still an open question. There are no conclusive results yet available and more research is needed. It is known, however, that both positive and negative ions reduce the life of bacteria (Serratia marsescens) residing in aerosols and that high negative ion concentrations seem to offer relief to persons suffering from bronchial disorders.

Static electricity in the atmosphere might not offer much practical value but it has intrigued many of the best theoreticians and experimentalists throughout time. There is no shortage of thunderstorms and lightning but the problem is how to safely penetrate a cloud and study its micro structure. One alternative is to use large cloud chambers capable of producing convective clouds. Some of the problems might be solved with the aid of orbiting space laboratories or unmanned satellites. The world wide thunderstorm activity and lightning distribution can be monitored from space and correlated with electric field measurements at earth. This could help

DESIGNING A DX CRYSTAL SET

by Mike Tuggle
46-469 Kuneki St.
Kaneohe, HI 96744-3536
E-Mail: tuggle@apacs.com

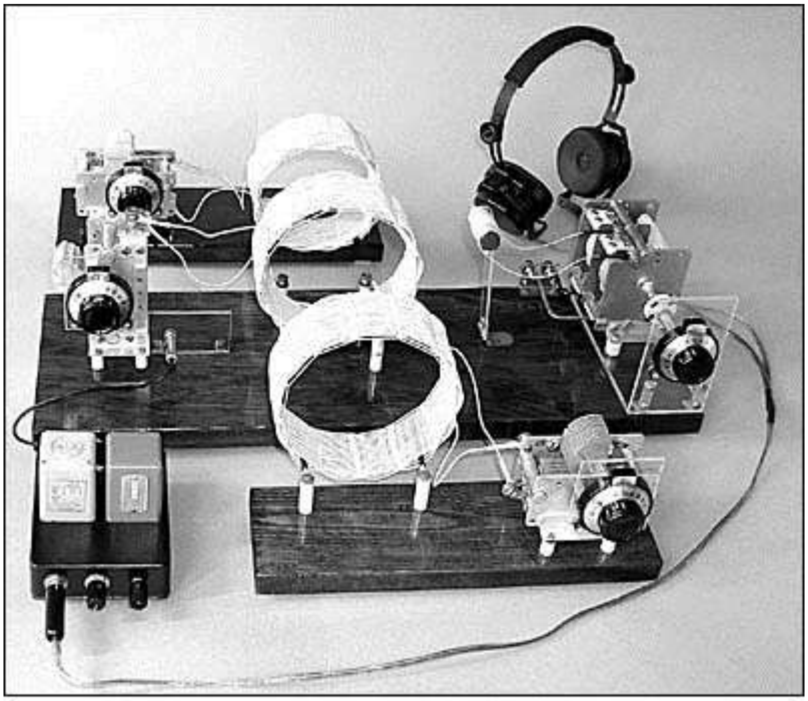

Fig. 1. The 'Lyonodyne-17,' an example of a well-designed DX crystal set.

Long distance reception with crystal radios is once again becoming a serious pursuit among hams and other radio hobbyists. They are discovering that the DX capabilities of these receivers have been greatly underrated. I've been particularly interested in crystal sets since 1959, when I first discovered you could actually DX with them. The Internet has aided crystal set DX activity by making exchange of ideas between like-minded enthusiasts much, much easier. This article only touches the surface from a personal perspective. The reader is encouraged to pursue the online resources, described below, covering all aspects and providing further examples of this fascinating hobby.

In this article I'll cover general design considerations for building DX crystal sets, but will leave the actual construction specifications up to you.

Equipment

Figures 1 and 2 show an example of a DX crystal set. I call it the "Lyonodyne-17." The design evolved from an earlier version described in the November, 1978 OTB. As you can see, this is not your grandfather's crystal set.

Antenna (right) and detector (left) tuners are mounted on the middle board. Wave traps are located fore and aft on separate boards, making it possible to adjust the coupling by moving the boards. The antenna coil (L1) is wound on a short ferrite rod. The matching transformer unit for my RCA "Big Cans" sound powered phones is at front, left.

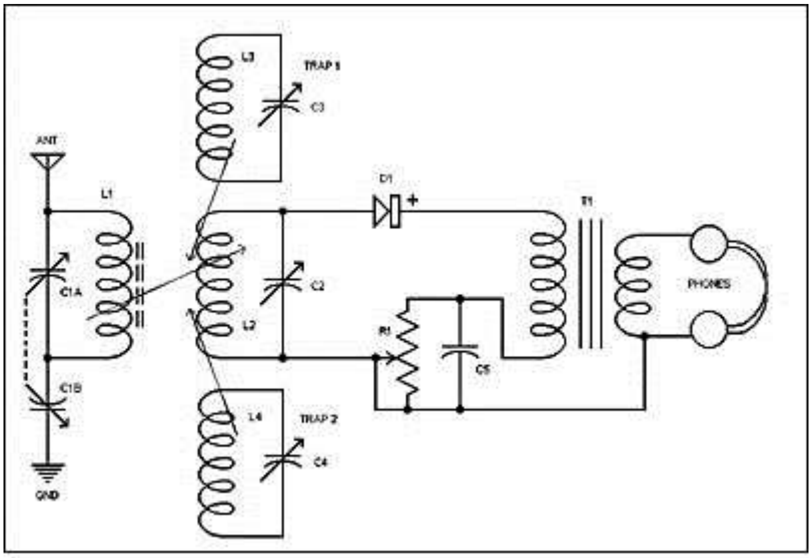

Fig. 2. Schematic of the 'Lyonodyne-17.'

So, what makes a crystal radio a "DX" set? Well, this one is double-tuned (L1-C1 and L2-C2) for selectivity to tune weak DX stations in the RF jungle we now live in. Some DX sets resort to triple tuning for even more selectivity -- depends on how many hands you have to do the tuning. The wavetraps (L3-C3 and L4-C4) can be tuned to reject strong unwanted stations in the manner of the notch filter on a communications receiver.

Coils L2 through L4 are basket-wound litz wire for low loss and high Q. Variable capacitors C1 through C4 are highest-quality, silver-plated, ceramic-insulated units--as perhaps only the military could specify. All these components are isolated from the mounting board by ceramic standoff insulators. Overkill maybe, but why take chances?

Design Principles

Two fundamental principles underlie DX crystal set design and construction: use of low-loss components and proper impedance matching between stages.

Avoid the temptation to construct the set with vintage components. The result may be a handsome set that is only a fair performer. I have such sets, and they sit up on the shelf and look nice. However, for ultimate performance, one should rely on top quality (usually modern) components and materials.

This is especially important for tuning capacitors, coil wire and forms, and detector diodes. Layout and construction, particularly in the RF-carrying sections (antenna through detector), should follow good HF practices: short direct leads, careful insulation of components, and avoidance of switches, taps and other trappings in the 'hot', RF-carrying sections.

The crystal set, like any other radio, consists of a series of stages--each with a function, each coupled to the next. The stages of a simplified crystal receiver are diagrammed in Figure 3. The resistances shown are used for detailed circuit analysis that will not discussed here.

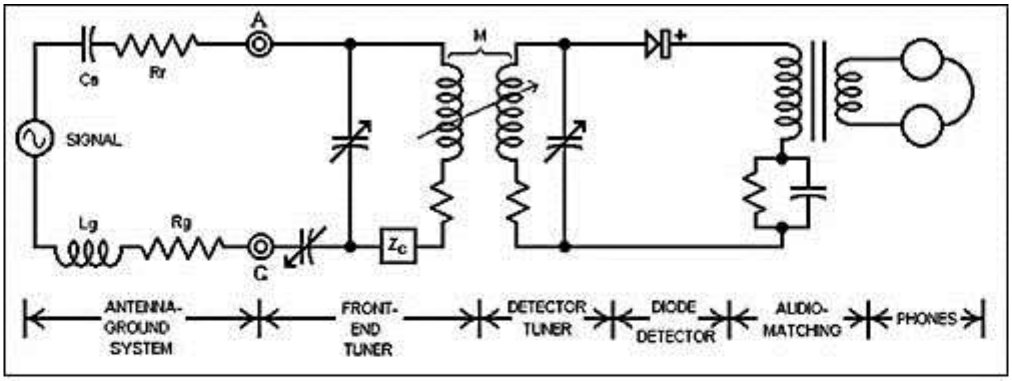

Fig. 3. The stages of a typical DX crystal set (see text).

We have, from left to right: antenna-ground system, front-end (antenna) tuner, secondary (detector) tuner, diode detector, audio matching and finally, phones. Maximum transfer of available signal power from one stage to the next happens when the impedances of these stages are matched to each other--from antenna to phones--and indeed, from phones to one's ears.

Components

A good outside antenna-ground system is essential for DXing. Loop antennas do not have enough pickup to be effective. An inverted L longwire 20 to 30 feet high and 50 feet long is a good start. The one truth to antenna design is, "higher is better." Any effort expended to raise the antenna, even a few feet, will be amply rewarded. In the most recent crystal set DX contest a third design principle emerged: A huge antenna can overcome shortcomings in the first two principles. The winner suspended 140 feet of litz wire near vertically using helium-filled balloons. Another high-scorer just happened to have a 140-foot tall base-insulated tower and four 1000-foot Beverage antennas to complement his junk box set. Most of us have neither the real estate nor the rigging capability to put up one of these 'mega' antennas. So it behooves us to heed principles 1 and 2.

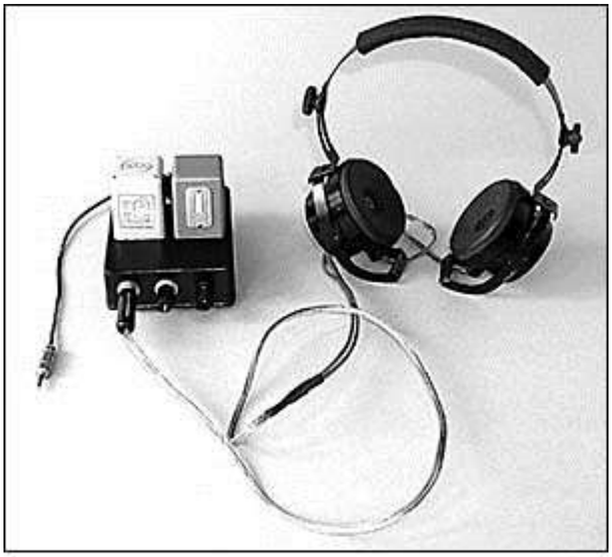

Fig. 4. RCA balanced-armature phones, with matching transformer.

The front-end (antenna tuner) design shown here is one of several that can be used. This particular design tunes a wide variety of antenna-ground systems. Other DX crystal sets use a simple series tuning circuit effectively--especially on long antennas.

The diode detector provokes more mystery, controversy and debate than any other component. Some folks favor low-resistance germanium diodes like the 1N34A or rock stands with galenas. Others swear by high-tech Schottky diodes. The emerging truth is, there is no universally perfect diode.

The 'best' diode depends specifically

upon the set it is used in. Per the second principle, the diode needs to match the tank circuit feeding it and the transformer and phones that it feeds. With some care, a high-Q tank with litz wire coil and a modern, military-grade variable capacitor can attain a resistance of from several hundred kilohms up to a megohm. Only Schottky diodes and a few modern germanium diodes have resistances this high. The old catswhisker-rock stand detectors have far lower resistances.

The most practical approach to selecting the right diode is to apply an A-B listening test to a number of them. Two diodes are mounted in a test stand arranged to quickly switch between them. Using a fairly weak station, one can test a pile of diodes, pair-wise, keeping the winner after each test, until the ultimate one is found. But it's most important to realize that this diode is 'best' for the particular set it was tested in. It may be a quite poor performer in a different set.

Surplus, sound-powered (more properly, balanced-armature) phones have become the industry standard for DX crystal sets. Baldwin Type C's were an early example. Now, post-WWII surplus units made by RCA and US Instruments Corp. (USI) are preferred.

These phones are low impedance and must be matched to the high-impedance tank (L2-C2) and detector diode by an audio transformer having, typically, 50- to 600-ohm and 50- or 100-kilohm windings (Figure 4). The quality of the transformer is very important, so that its insertion loss is small. UTC input transformers have a good reputation for low loss. The final touch is a comfortable set of headphone cushions for good acoustical coupling to the ears and exclusion of outside noise.

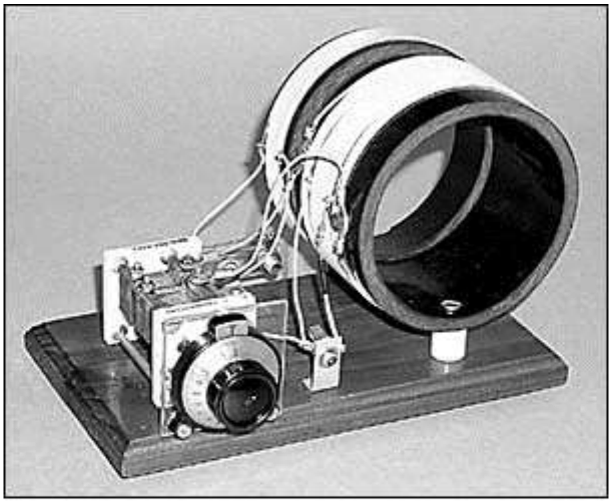

Fig. 5. A '12,000-mile' short-wave receiver.

DX Experiences

Under favorable conditions, medium wave or broadcast band DX crystal sets can receive hundreds of stations--some of them thousands of miles distant. In fact, DX crystal set performance is comparable to any other radio (powered or not) short of a full-blown communications receiver with its own outside antenna. The rule of thumb is, if you can hear them on a radio, you can hear them on a crystal set.

The best times to listen are at sunrise and sunset, when stations are signing on or off, raising or dropping their powers, and changing their antenna patterns. These circumstances make for a jumble of regional stations ripe for the picking. Deep night is usually the best time for flat-out DX.

In 15-year stints at two locations in Maryland, I accumulated logs of over 600 and 400 stations, respectively. The most distant receptions, at 1800 to 2200 miles, were a few powerful stations in the Caribbean area and adjacent South America. I was aided, no doubt, by the largely water path between them and me. The farthest overland station was in Denver at about 1500 miles.

When I moved to Hawaii, I wasn't sure what to expect. The local Honolulu stations were givens. But what about the outer islands? And all-importantly the next stations, mainlanders, the closest some 2400 miles away? As it turns out, Hawaii is an ideal DX location--for distance if not for sheer numbers of stations. With no regional stations, sunrise-sunset activity is non-existent.

West coast 500-watt stations, 2500 miles distant, have been heard here. High-power stations in Cuba (4800 mi.) and the Caicos Islands (5500 mi.) have also been heard, thanks to a mostly water path. Stations from the "interior" also make it over. Last year, it was neat to hear KRVN in snowbound Nebraska (3600 mi.) using a homemade cat's whisker mineral detector. This must be what it was like in the old days.

With low station powers and crowding, medium-wave broadcast band DXing represents the greatest challenge. On short wave there's no limit to the distance--reception is truly worldwide. I occasionally hear the South African broadcasts nearly 12,000 miles away; Johannesburg and Hawaii are nearly at antipodes. From the mainland, Australia (11,000 miles) was a routine catch. Figures 5 and 6 show my "12,000-mile" crystal set.

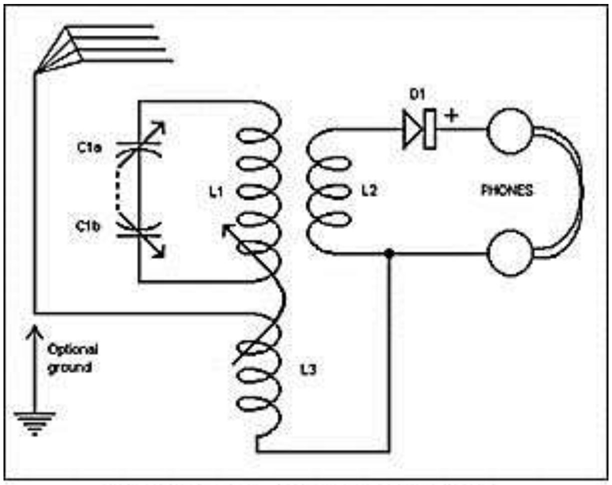

Fig. 6. Schematic of the '12,000 mile-er.'

Crystal Set DX Activities

Crystal set DX activities include discussion forums sharing ideas and results, an annual DX listening contest now into its sixth year, and the occasional set-building contest where some remarkably high-caliber craftsmanship comes to the fore. All of these activities are served up on the Internet.

The following site is a highly recommended grand portal to the wonderful world of crystal set DXing: Owen Pool's Crystal Radio Resources at http://www.thebest.net/wuggy. The comprehensive set of links here covers all aspects of the hobby. This site is headquarters for the annual crystal set DX (XSDX) contests open to everyone. The contest usually takes place in late January. Watch for announcement of specific dates and rules.

My own experiences over the years have led me to shed my skepticism of the incredible DX reports made by old-timers back in the early days. Component technology was not what it is today, but circuit technology was, and the sparsely populated bands back then had to be a lot more DX-friendly. You may want to pull an antique crystal set off the shelf and give it a spin. I'd encourage it.

Some of these old sets are extremely well built. The construction of the military BC-14A/SCR-65 never ceases to amaze me--and others were built just as well. But to start out, I'd recommend conceding to modern technology by using a good, modern diode and sound-powered phones with matching transformer. With just a little luck, prepare to be amazed all over again!

(No Model.)

R. VOSS.

INFLUENCE MACHINE.

No. 410,053. Patented Aug. 27, 1889.

Fig. 1.

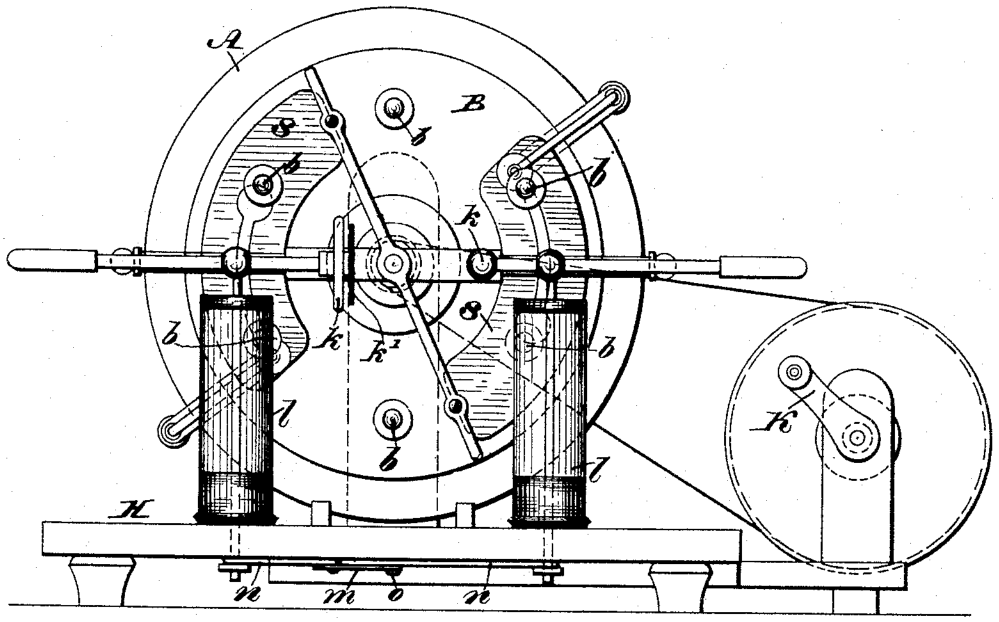
A
S
B
b
b
b
k
b
k
k1
S
b
l
b
l
K
H
n
m
o
n

Fig. 2.

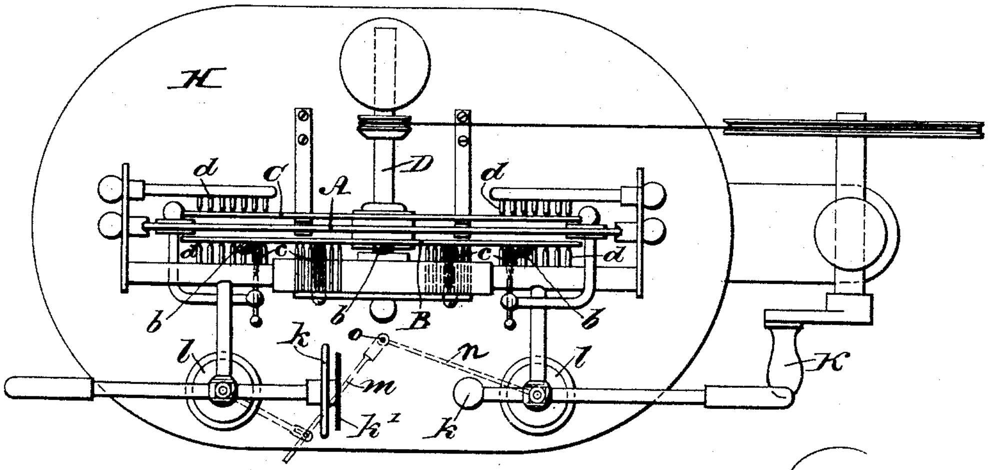
H
D
d
C
A
d
d
c
c
d
b
b
B
b
l
k
o
n
l
m
k1
k
K

WITNESSES:

L. Douville,

P. F. Nagle.

INVENTOR:

Robert Voss.

BY Wiedersheim & Kintner

ATTORNEYS.

N. PETERS, Photo-Lithographer, Washington, D. C.

UNITED STATES PATENT OFFICE.

ROBERT VOSS, OF BERLIN, GERMANY, ASSIGNOR TO JAMES W. QUEEN & CO., OF PHILADELPHIA, PENNSYLVANIA.

INFLUENCE-MACHINE.

SPECIFICATION forming part of Letters Patent No. 410,053, dated August 27, 1889.

Application filed January 21, 1889. Serial No. 297,066. (No model.)

To all whom it may concern:

Be it known that I, ROBERT VOSS, a subject of the King of Prussia, German Emperor, residing at the city of Berlin, in the Kingdom of Prussia, German Empire, have invented certain new and useful Improvements in Influence-Machines; and I do hereby declare that the following is a full, clear, and exact description of the invention, which will enable others skilled in the art to which it appertains to make and use the same.

My invention has reference to improvements in double-acting electrical influence-machines; and it consists in an arrangement of three circular glass disks, two of which are rotary and on the same axis, one of these receiving any electricity that may escape on its side from the fixed or third disk and also generating electricity by friction with the air.

It also consists in certain details of construction hereinafter described, and particularly pointed out in the claims which follow this specification.

I carry out my invention in the way shown in the accompanying drawings, in which—

Figures 1 and 2 are respectively front and top views of a machine provided with my improvements.

Similar letters of reference indicate like parts in both of the figures.

A is a fixed glass disk having on it two paper sectors *s s*, placed diametrically opposite each other.

B and C are the two rotary glass disks, B being set out at equal distances from its center with small metal knobs *b b*, which, in conjunction with small metal brushes *c c*, serve to generate the electricity to be originally communicated to the influence-machine.

d d are metallic points serving as conductors.

k k are electrodes, which are for the purpose of producing a continuous spark, and they are in communication with two Leyden jars *l l*. I provide either one or both of these electrodes *k k* with a flat circular or disk-like attachment *k'* of insulating material, as rosin, gum, or hard rubber, which is secured to the metal electrode by a central stem, as shown, leaving an air-space between the two parts. With such a surface having this air-space between it and the metal electrode I obtain marked efficiency in the dispersion of the spark and avoid the troublesome brush-discharge.

Underneath the base-plate H is an arrangement by which the two Leyden jars *l l* may be placed in electric communication with each other. It consists of a metal bar *m*, turning on a point *o*, and of the metal wires *n n*, each connected to one of the Leyden jars. The bar *m*, as shown in the drawings, may effect a connection between the two wires *n n*, which may again be disconnected by turning the bar *m* on point *o*. The disk C is uncovered and serves for the reception of any electricity escaping on the rear side from the disk A, and also for generating a small quantity of electricity by its friction with the air when in rotation. This central disk A, being stationary and located in close proximity to the rotary disk on either side, acts after the manner of a plate in a condenser, tending to hold the electricity bound on both surfaces. The electricity thus held in tension appears to react upon the rotary plates to give increased effect. The two disks B C are mounted on one common axis D, receiving rotary motion from the crank K.

I am aware that for a long time past influence-machines have been produced having for their object to obtain by influence a larger quantity of electricity than that obtainable with the Holz machine. While this latter consisted mainly of a pair of circular varnished glass disks, one fixed and the other rotating, the improved construction had a number of such pairs of glass disks. By this arrangement the machine was very large and expensive.

By my machine a relatively larger quantity of electricity is obtained in a very simple way by influence through the back second wholly uncovered disk serving both as a protecting-body against any escape of electricity and as a generator of electricity by reason of its generating a small quantity of electricity by friction when rotating with the air present between it and the fixed disk, such electricity being then re-enforced by influence.

I am aware that influence-machines have

been constructed in which two rotary disks are adapted to rotate on opposite sides of two centrally-located fixed disks, said fixed disks having space between them. In this construction, however, of two fixed disks the space between said fixed disks is not utilized, and much of the influence of the intervening bound of electricity is lost. I therefore do not claim a double arrangement of such mechanism collectively; but

What I do claim as my invention, and desire to secure by Letters Patent, is—

1. In an influence-machine, the combination of a central fixed disk and two rotary disks attached to a shaft extending through the fixed disk, with two electrodes, one of which consists of a non-conducting surface attached to a conducting-surface, but separated therefrom by an air-space, the other electrode being of usual form, and both electrically connected to the influence-machine, substantially as described.

2. In an influence-machine, an electrode consisting of a non-conducting surface attached to the conducting-electrode, said parts being separated by an air-space, substantially as shown and described.

3. An electrode for an influence-machine, consisting of a flat or disk-like conductor, in combination with a flat or disk-like conducting-surface connected to the conducting part by a central stem, as shown and described.

In testimony whereof I signed this specification in the presence of two subscribing witnesses.

ROBERT VOSS.

Witnesses:
B. Roi,
Adolph Schwabe.

Dec. 12, 1961 C. H. GOLDIE 3,013,201

SELF-EXCITED VARIABLE CAPACITANCE ELECTROSTATIC GENERATOR

Filed May 31, 1960

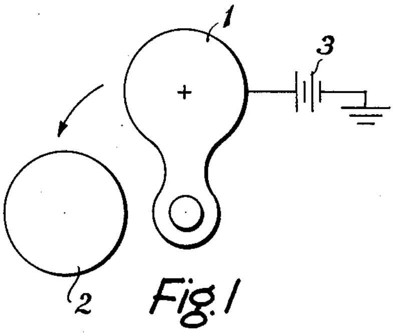

Fig. 1

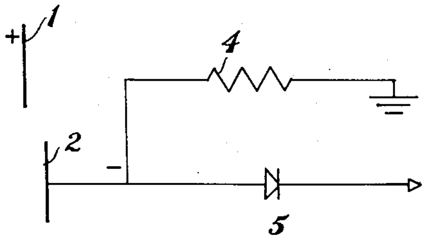

Fig. 2

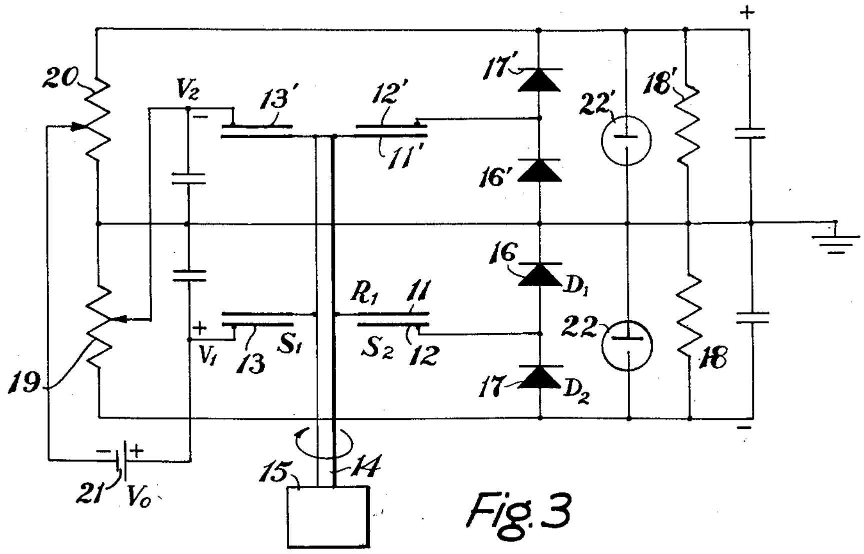
+
20
V₂
−
13'
12'
11'
17'
22'
18'
16'
16
D₁
R₁
11
22
18
+
V₁
13
S₁
S₂
12
19
17
D₂
−
+
21
V₀
15
14

Fig. 3

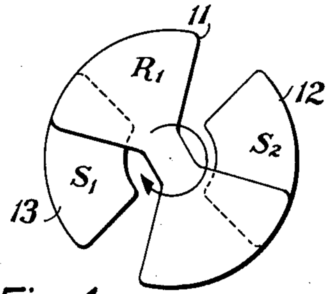
11
R₁
12
S₂
S₁
13

Fig. 4

3,013,201

SELF-EXCITED VARIABLE CAPACITANCE ELECTROSTATIC GENERATOR

Charles H. Goldie, Bedford, Mass., assignor to High Voltage Engineering Corporation, Burlington, Mass., a corporation of Massachusetts

Filed May 31, 1960, Ser. No. 32,821
4 Claims. (Cl. 322—2)

This invention relates to electrostatic generators wherein the charge transfer mechanism makes use of capacitive effects between charged bodies and electrodes upon which charge is to be induced, and in particular to the self-excitation of such generators.

The invention may best be understood from the following detailed description thereof, having reference to the accompanying drawings in which:

FIG. 1 is a diagram illustrating one form of a variable-capacitance electrostatic generator;

FIG. 2 is another diagram further illustrating the arrangement of FIG. 1;

FIG. 3 is a diagram illustrating a modification of the apparatus of FIG. 1 and embodying the invention; and

FIG. 4 is a diagram illustrating a portion of the apparatus of FIG. 3.

Referring to FIG. 1 of the drawings, therein is shown a rotor **1** and a stator **2**. The rotor **1** has a charge thereon and in FIG. 1, it is charged positively. A power supply **3** is indicated for maintaining the charge on the rotor **1**.

Referring now to FIG. 2, it will be seen that the stator **2** is connected to ground by a resistive path **4** which is connected in parallel with at least one rectifier **5** so that current can flow in only one direction between the stator **2** and ground.

In the device shown in FIGS. 3 and 4 the rotor serves as a capacitive link between a stationary charging electrode and a stator. Referring thereto, a rotor **11** periodically passes close to a stator **12**. However, instead of carrying its own charge as in the device of FIGS. 1 and 2, the necessary inducing charge on the rotor **11** is itself induced on the rotor **11** by the charging electrode **13**. Thus when one end of the rotor **11** is in the vicinity of the stator element **12** and the opposite end of the rotor element **11** is in the close vicinity of the charging electrode **13**, a negative charge is induced on that portion of the rotor **11** which is near the charging electrode **13** with the result that positive charge appears at the opposite end of the rotor **11** since the rotor **11** must itself be electrically neutral.

The invention will now be described with reference to the apparatus shown in FIGS. 3 and 4, but it will be clear from the foregoing that the invention could equally well be used with a device of the type shown in FIGS. 1 and 2. The principal advantage of the device shown in FIGS. 3 and 4 is the fact that it eliminates the necessity for a brush on the rotor to supply charge thereto, since the charge on the rotor is induced capacitively.

Referring back briefly to FIGS. 1 and 2, it will be recalled that the invention is concerned with the construction of a self-excited generator with no external energy other than shaft power supplied. That is to say, the objective is a completely self-charging generator. Referring to FIGS. 1 and 2, it will be seen that direct feedback of the output to the induction plate is ruled out by the polarity switch inherent in this type of machine. The difficulty is avoided by the two section unit shown in FIGS. 3 and 4. Here a fraction of the output voltage of each section is fed to the induction plate of the other.

The action of the circuit is best understood by looking first at the lower section only. The rotor **11** is an electrically isolated flat fan-like rotor mounted on an insulating shaft **14** driven by an appropriate motor or other prime mover **15**. As it rotates, the rotor **11** periodically covers the stator **2** and the charging electrode **13**, which are each insulated and sector-shaped. The charging electrode **13** is maintained at a positive D.C. potential with respect to ground. As the rotor **11** rotates into a position covering the charging electrode **13** and the stator **12**, the capacitance between the charging electrode **13** and the stator **12** increases and negative charge is induced from ground through a rectifier **16** to the stator **12**. As the rotor **11** uncovers the stator **12** and the charging electrode **13**, the capacitance decreases and the stator **12** rises in potential negative with respect to ground. The rectifier **16** blocks any flow of negative charge from the stator **12** and as the negative potential of the stator **12** rises, current flows through a second rectifier **17** to the load **18**. A fraction V2 of this negative load voltage is applied to the charging electrode or induction plate **13'** of the second or upper section by means of the potentiometer **19**. The charging action in this second section is identical to that in the first section except for reversed polarities and the output of the second section is correspondingly fed-back to the first section induction plate **13** via a second potentiometer **20**. The variable resistors of the potentiometers **19**, **20** control the output voltage and power by adjusting the feedback ratios.

The circuit is regenerative; it remains only to supply a small reference voltage to start in the right direction. The use of stator and rotor materials with suitable contact potentials is one solution to this; a small battery **21**, which might be, for example, of the nuclear type, in one induction line as shown is another solution. Suitable voltage limiting devices such as corona tubes **22** would be added either across the induction plate **13** or, as shown in FIG. 3, across the load **18** as checks to prevent excessive voltage build-up.

A number of features inherent with this circuit are of interest.

The usual slip rings and brushes are eliminated by the use of the stator **12** and the charging electrode **13**, and an isolated rotor **11** rather than a single stator **2** and a rotor **1** maintained at constant induction potential as shown in FIGS. 1 and 2.

In spite of the fact that the second section was added for the purpose of self-excitation, both output voltages are added together across the load **18**, **18'**. The arrangement pictured is merely for explanation of the self-excitation principle. The ground shown, for example, is not required; and for higher voltages, the units may be cascaded with a common insulating drive shaft **14**. For a given total voltage output, the number of rotor stator sets is the same for this self-charge generator as for a generator requiring auxiliary charging supplies.

Two poles are shown for simplicity; in an actual machine a greater number would be more practical. The system also applies equally well to other geometrical arrangements, such as the concentric or segmented drum design mentioned in U.S. Patent Application Serial No. 829,823.

Having thus described the principles of the invention, together with an illustrative embodiment thereof, it is to be understood that, although specific terms are employed, they are used in a generic and descriptive sense and not for purposes of limitation, the scope of the invention being set forth in the following claims.

I claim:

1. A self-excited variable-capacitance D.-C. generator comprising in combination a first and a second electrode, a third and a fourth electrode capacitively coupled to said first and second electrodes respectively, means for vary-

ing the capacitance between said first and third electrodes and between said second and fourth electrodes, a first rectification means between said third and fourth electrodes, a load one end of which is connected to said third electrode through a second rectification means and the other end of which is connected to said fourth electrode through a third rectification means, said rectification means being thus connected in series and with the same orientation, and means for utilizing a portion of the voltage across said load to charge said first and second electrodes in mutually opposite senses.

2. A self-excited variable-capacitance D.-C. generator comprising in combination a first and a second rotor, a first and a second stator capacitively coupled to said first and second rotor respectively, means for rotating said rotors so as to vary the capacitance between said first rotor and said first stator and between said second rotor and said second stator, a first rectification means between said stators, a load one end of which is connected to said first stator through a second rectification means and the other end of which is connected to said second stator through a third rectification means, said rectification means being thus connected in series and with the same orientation, and means for utilizing a portion of the voltage across said load to charge said rotors in mutually opposite senses.

3. A self-excited variable-capacitance D.-C. generator comprising in combination a first and a second charging electrode, a first and a second stator, a first rotor capacitively coupling said first charging electrode and said first stator, a second rotor capacitively coupling said second charging electrode and said second stator, means for rotating said rotors so as to vary the capacitance between said charging electrodes and said stators, a first rectification means between said stators, a load one end of which is connected to said first stator through a second rectification means and the other end of which is connected to said second stator through a third rectification means, said rectification means being thus connected in series and with the same orientation, and means for utilizing a portion of the voltage across said load to charge said charging electrodes in mutually opposite senses.

4. A self-excited variable-capacitance D.-C. generator comprising in combination a first and a second charging electrode, a first and a second stator, a first rotor capacitively coupling said first charging electrode and said first stator, a second rotor capacitively coupling said second charging electrode and said second stator, means for rotating said rotors so as to vary the capacitance between said charging electrodes and said stators, a first rectification means between said stators, a load one end of which is connected to said first stator through a second rectification means and the other end of which is connected to said second stator through a third rectification means, said rectification means being thus connected in series and with the same orientation, means for utilizing a portion of the voltage across said load to charge said charging electrodes in mutually opposite senses, and voltage limiting means across said load.

No references cited.

United States Patent [19] [11] **4,127,804**

Breaux [45] **Nov. 28, 1978**

[54] **ELECTROSTATIC ENERGY CONVERSION SYSTEM**

[75] Inventor: **Onezime P. Breaux,** Trotwood, Ohio

[73] Assignee: **The United States of America as represented by the Secretary of the Air Force,** Washington, D.C.

[21] Appl. No.: **689,272**

[22] Filed: **May 24, 1976**

[51] **Int. Cl.**^{2} **H02N 1/00**

[52] **U.S. Cl.** .... **322/2 A;** 310/308; 320/1

[58] **Field of Search** .... 322/2 A; 310/5, 10, 310/308, 309; 320/1

[56] **References Cited**

U.S. PATENT DOCUMENTS

2,540,327	2/1951	Felici	322/2 A X
3,094,653	6/1963	LeMay et al.	322/2 A
3,971,938	7/1976	O'Hare	310/5 X
4,054,826	10/1977	Wahlstrom	322/2 A

Primary Examiner—Robert J. Hickey
Attorney, Agent, or Firm—Joseph E. Rusz; Robert Kern Duncan

[57] **ABSTRACT**

Two inversely ganged variable capacitors having a common movable element varying the capacities such that when one capacitor is at maximum capacitance, the other is at minimum capacitance, have, after initially charging, a substantially constant contained or trapped charge distributed between the two capacitors. As the capacities of the capacitors are varied, a potential difference is developed between the fixed plate of one capacitor and the fixed plate of the other. When these two potential points are connected to a load, charge is transferred from one capacitor to the other as a current flow through the load. The total charge is not diminished, the energy supplied the load being the energy expanded on moving the movable charged plates. As the movable plates are moved in a cyclic manner an alternating current is provided to the load.

8 Claims, 19 Drawing Figures

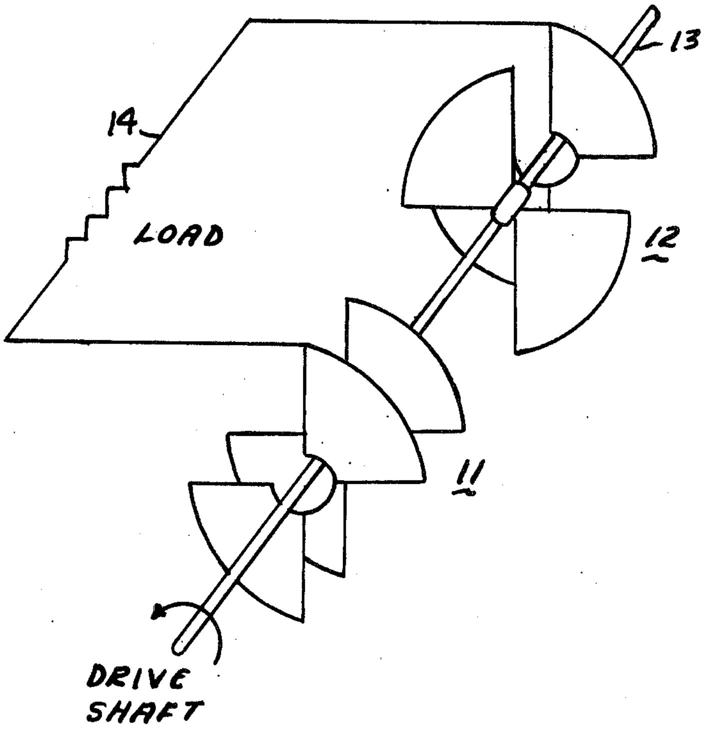

Fig-1

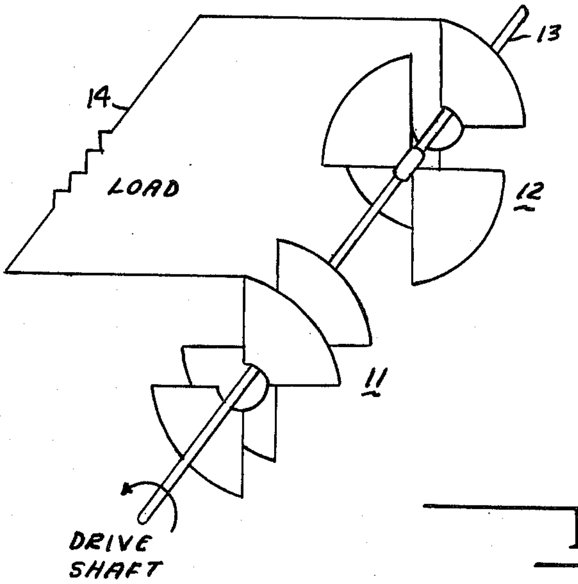
13
14
LOAD
12
11
DRIVE SHAFT

Fig-2

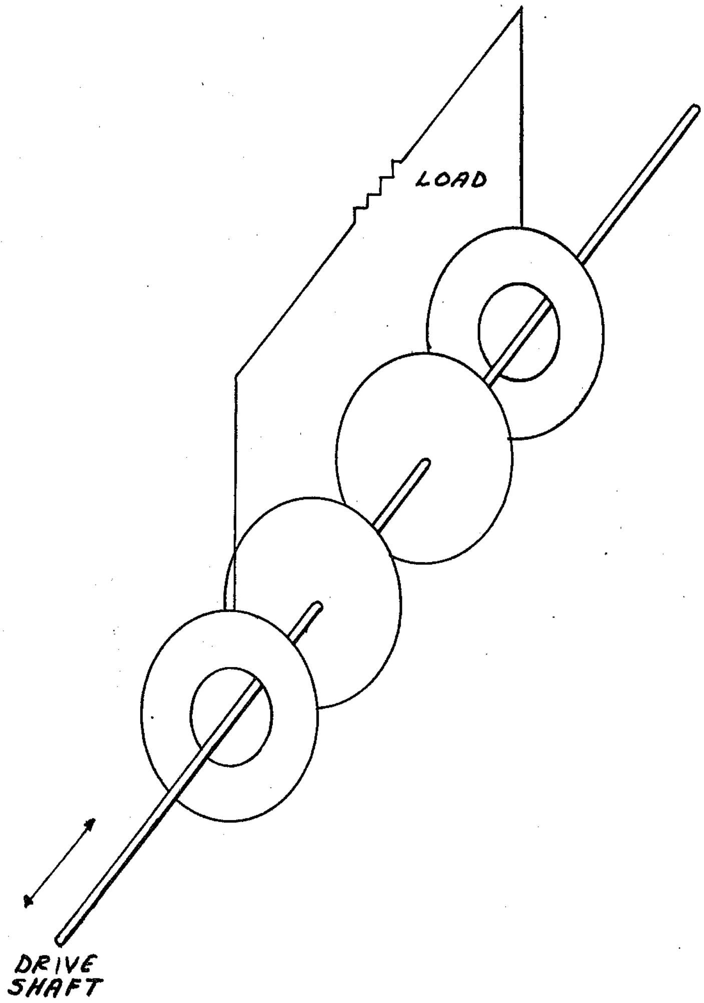
LOAD
DRIVE SHAFT

Fig-3

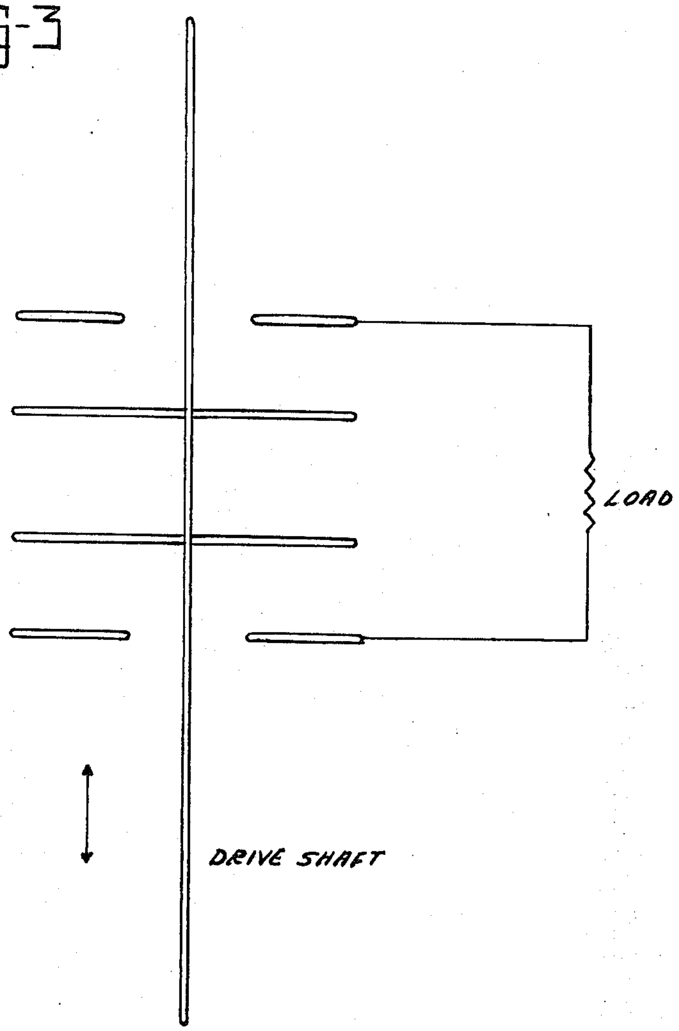
LOAD
DRIVE SHAFT

Fig-5

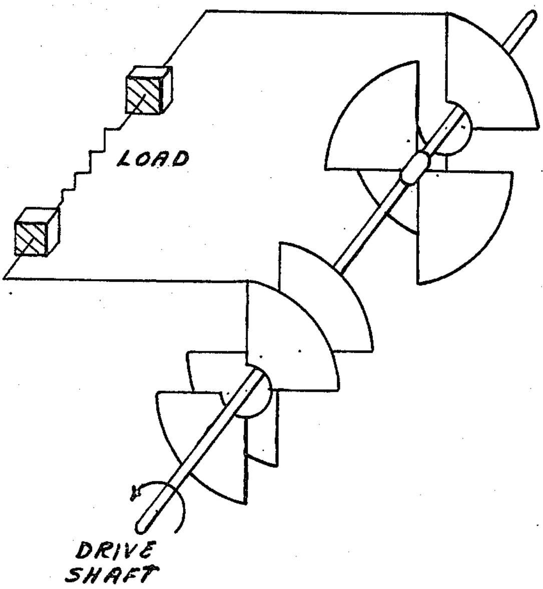
LOAD
DRIVE SHAFT

Fig-4

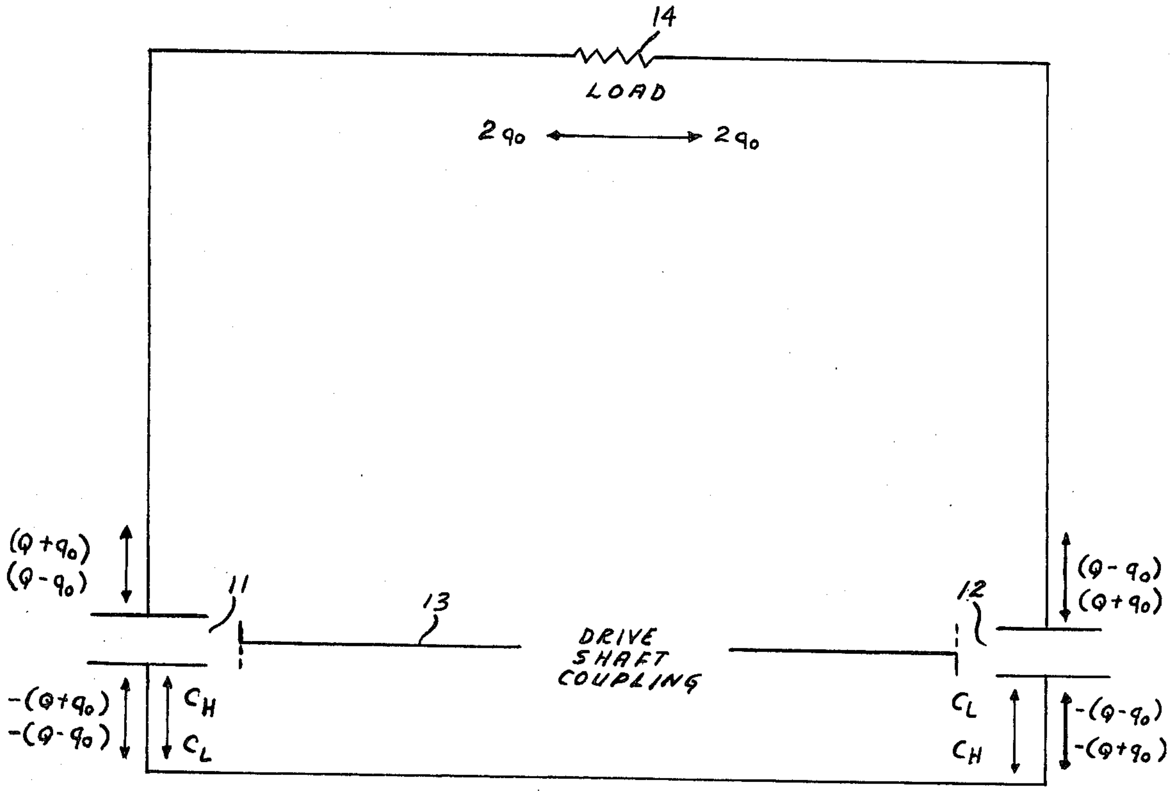
14
LOAD
$2q_0$
$2q_0$
$(Q+q_0)$
$(Q-q_0)$
11
13
DRIVE SHAFT COUPLING
12
$(Q-q_0)$
$(Q+q_0)$
$-(Q+q_0)$
$-(Q-q_0)$
C_H
C_L
C_L
C_H
$-(Q-q_0)$
$-(Q+q_0)$

Fig-6

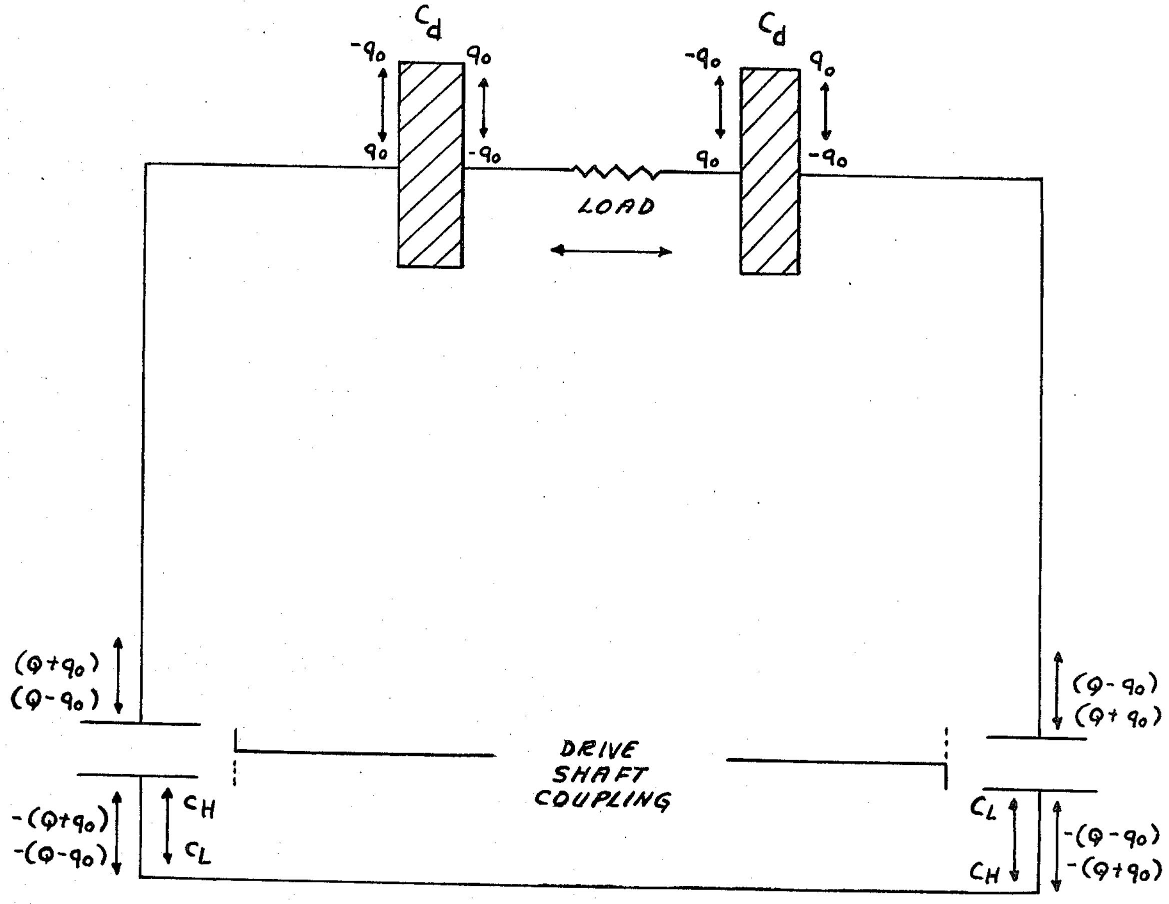
C_d
$-q_0$
q_0
q_0
$-q_0$
C_d
$-q_0$
q_0
q_0
$-q_0$
LOAD
$(Q+q_0)$
$(Q-q_0)$
$-(Q+q_0)$
$-(Q-q_0)$
C_H
C_L
DRIVE SHAFT COUPLING
$(Q-q_0)$
$(Q+q_0)$
C_L
C_H
$-(Q-q_0)$
$-(Q+q_0)$

Fig-7

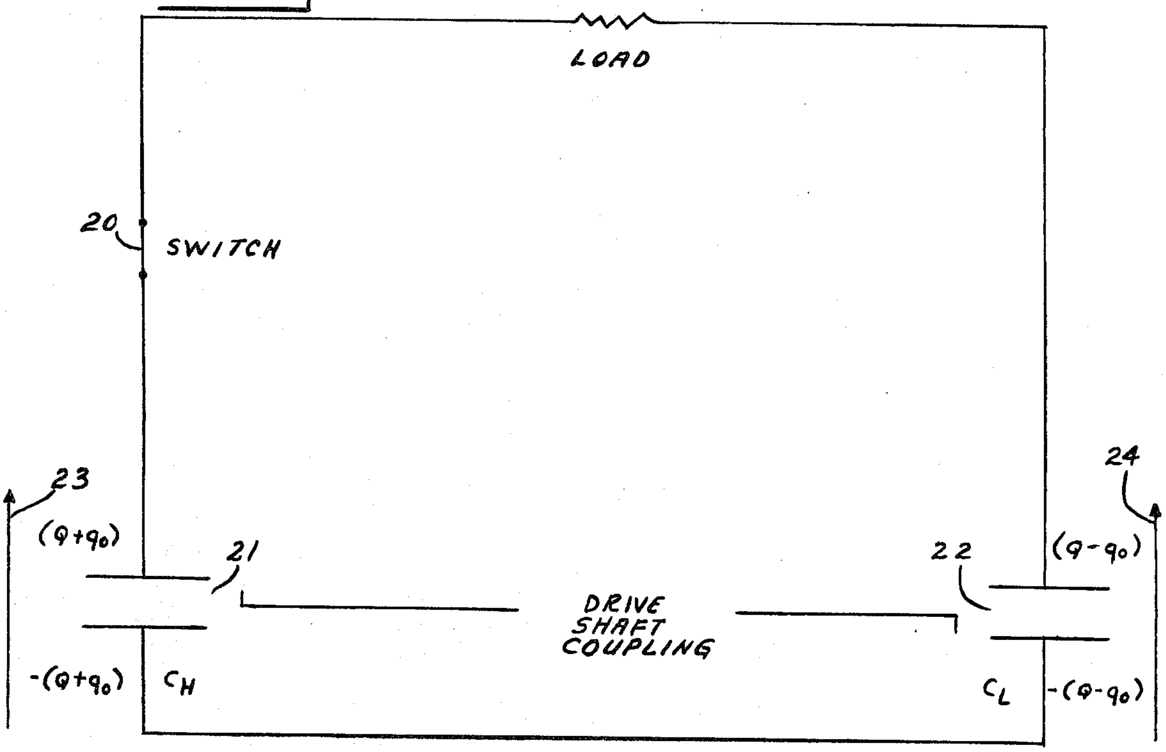
LOAD
20
SWITCH
23
$(Q+q_0)$
$-(Q+q_0)$
21
C_H
DRIVE SHAFT COUPLING
22
24
$(Q-q_0)$
$-(Q-q_0)$
C_L

Fig-8

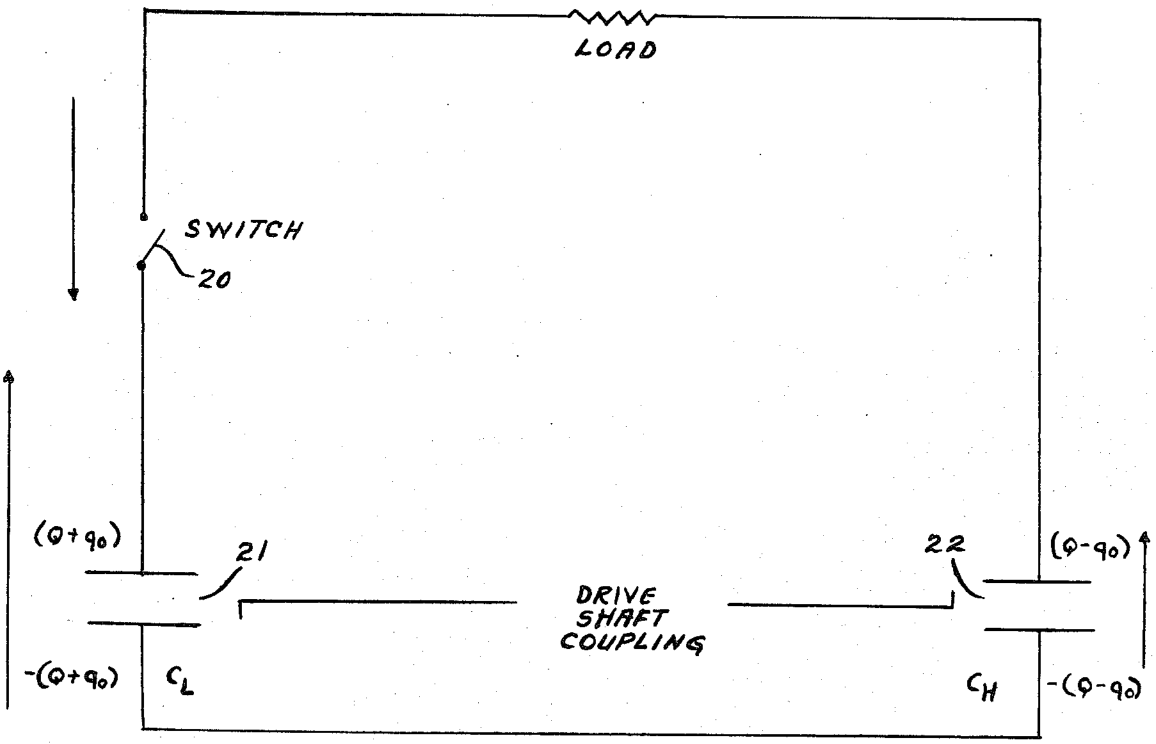
LOAD
SWITCH
20
$(Q+q_0)$
$-(Q+q_0)$
21
C_L
DRIVE SHAFT COUPLING
22
$(Q-q_0)$
$-(Q-q_0)$
C_H

Fig-9

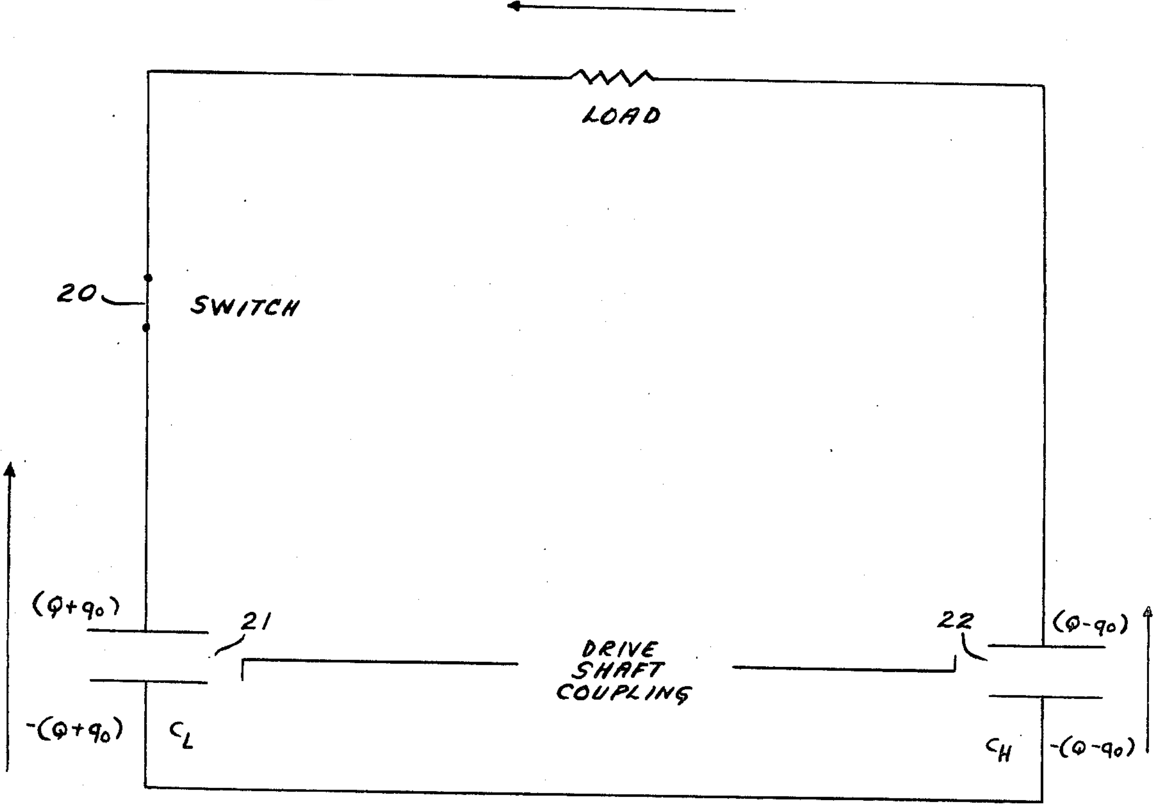
LOAD
20
SWITCH
(φ+q₀)
21
-(φ+q₀)
C_L
DRIVE
SHAFT
COUPLING
22
(φ-q₀)
C_H
-(φ-q₀)

Fig-10

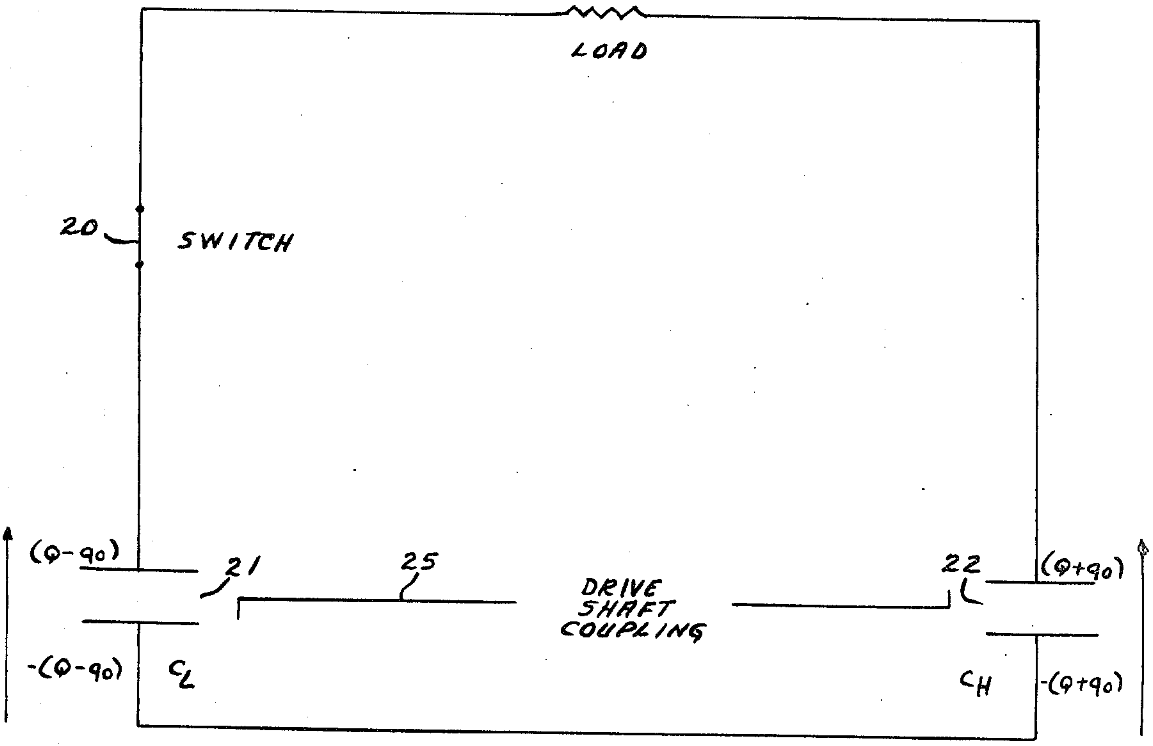
LOAD
20
SWITCH
(φ-q₀)
21
25
-(φ-q₀)
C_L
DRIVE
SHAFT
COUPLING
22
(φ+q₀)
C_H
-(φ+q₀)

Fig 11

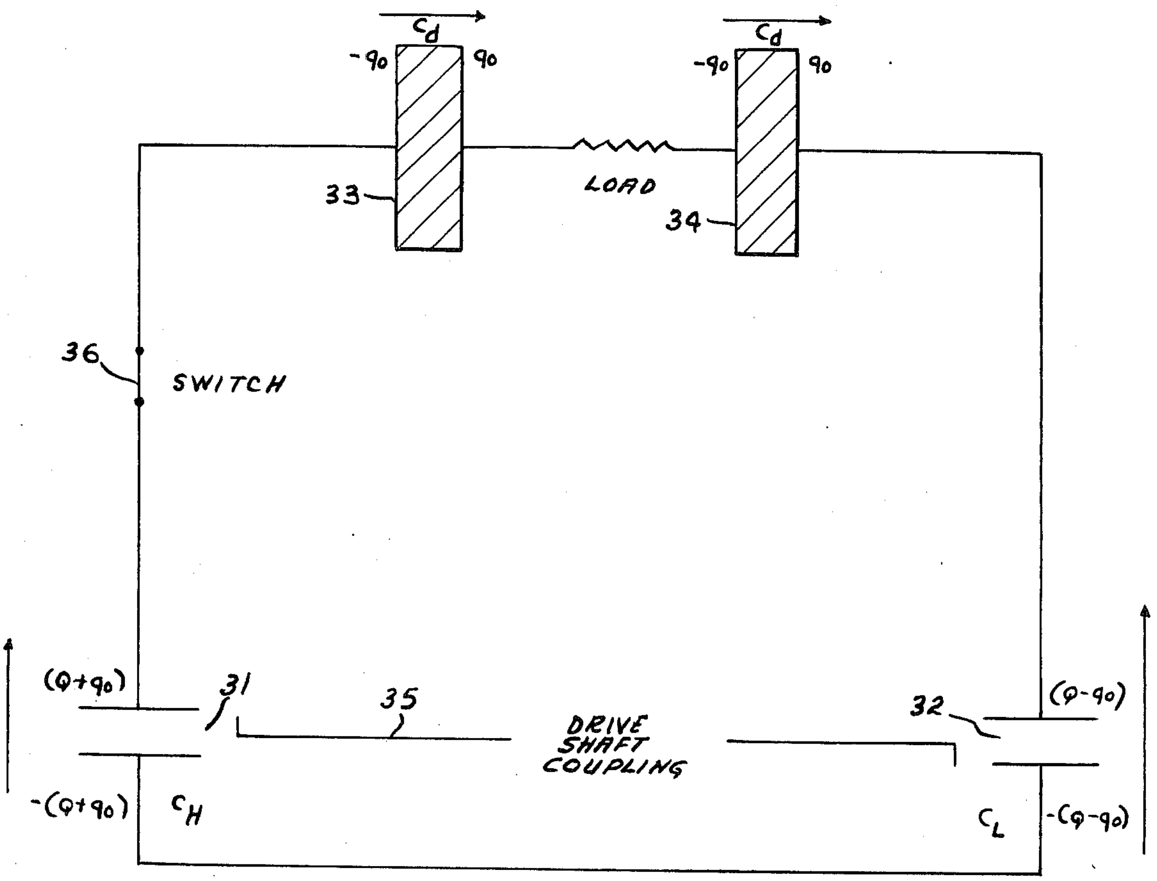
C_d
-q0
q0
C_d
-q0
q0
33
LOAD
34
36
SWITCH
(Q+q0)
31
35
DRIVE SHAFT COUPLING
32
(Q-q0)
-(Q+q0)
C_H
C_L
-(Q-q0)

Fig-12

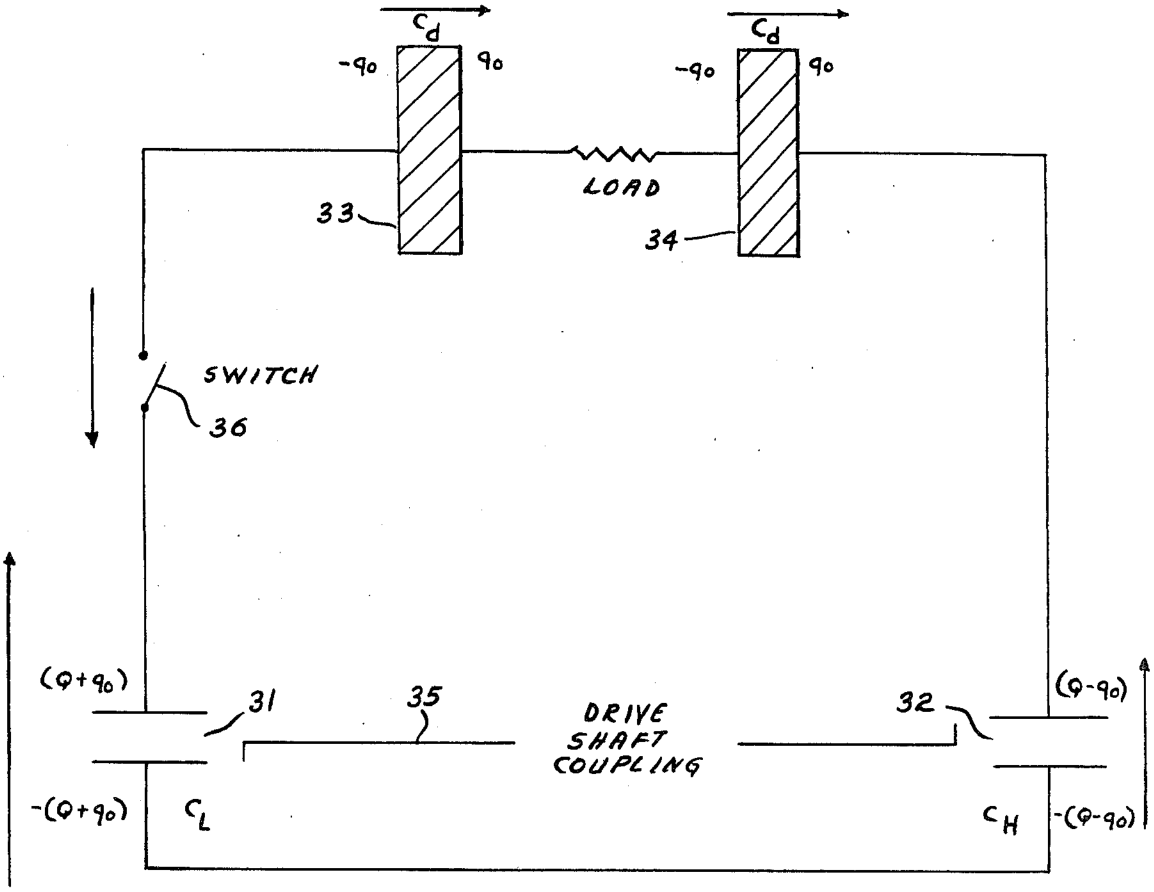

C_d
$-q_o$
q_o
C_d
$-q_o$
q_o
LOAD
33
34
SWITCH
36
$(Q+q_o)$
31
35
DRIVE SHAFT COUPLING
32
$(Q-q_o)$
$-(Q+q_o)$
C_L
C_H
$-(Q-q_o)$

Fig-13

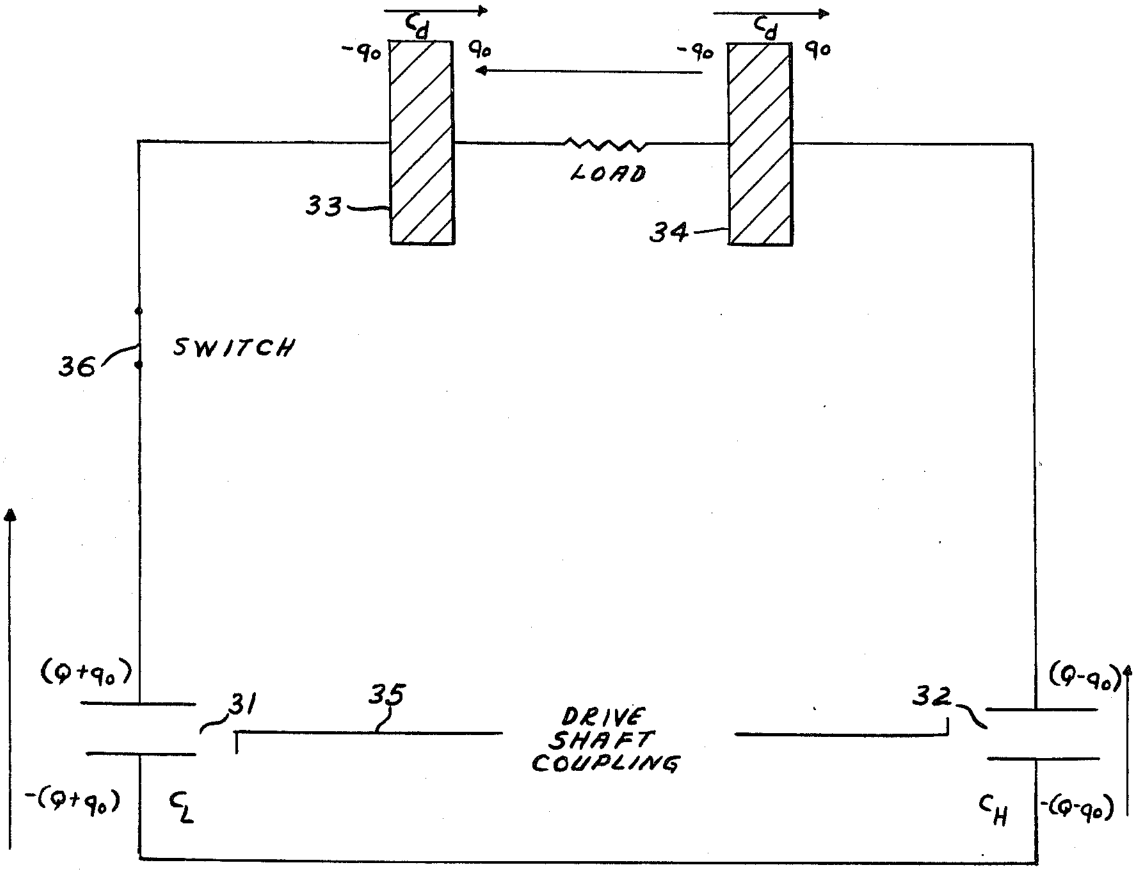
C_d
$-q_0$
q_0
C_d
$-q_0$
q_0
LOAD
33
34
36
SWITCH
$(q+q_0)$
$-(q+q_0)$
31
35
DRIVE SHAFT COUPLING
32
$(q-q_0)$
$-(q-q_0)$
C_L
C_H

Fig-14

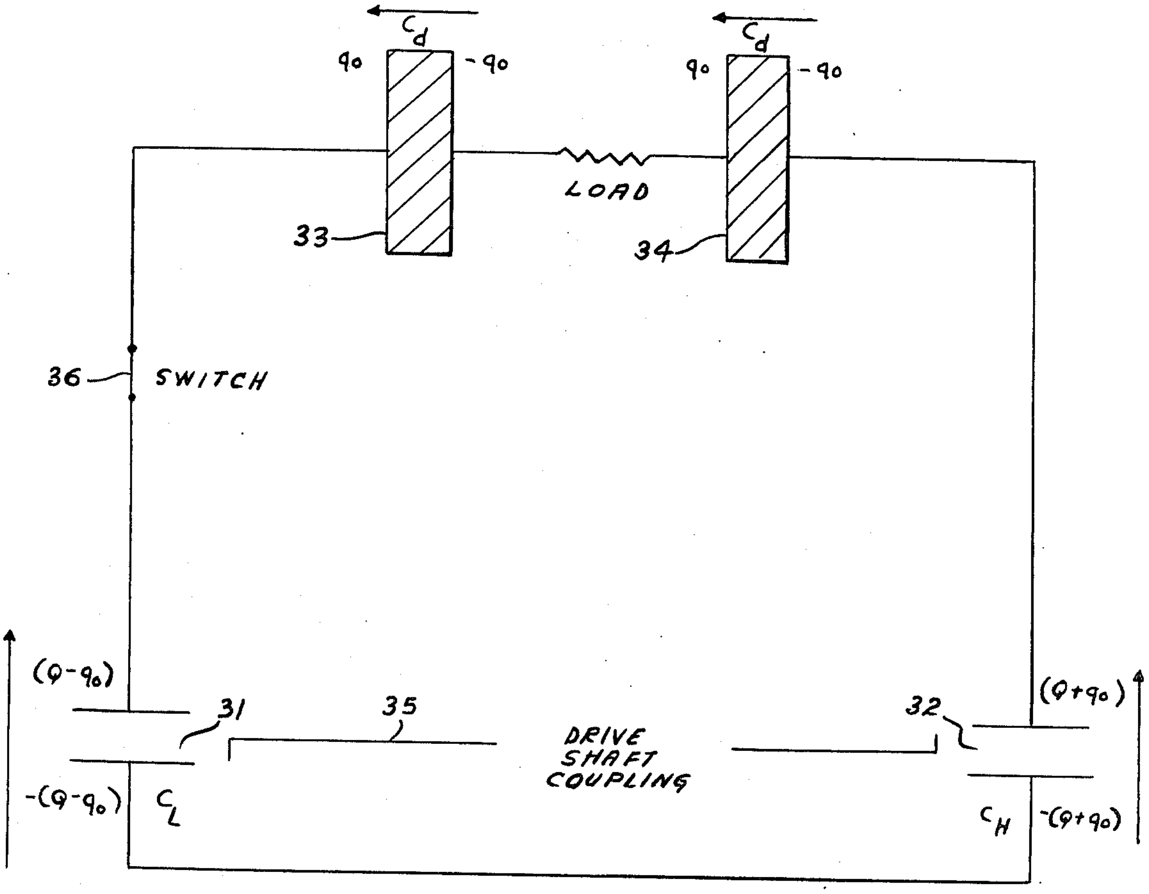
C_d
q_0
$-q_0$
C_d
q_0
$-q_0$
LOAD
33
34
36
SWITCH
$(q-q_0)$
31
35
DRIVE SHAFT COUPLING
32
$(q+q_0)$
$-(q-q_0)$
C_L
C_H
$-(q+q_0)$

Fig-15

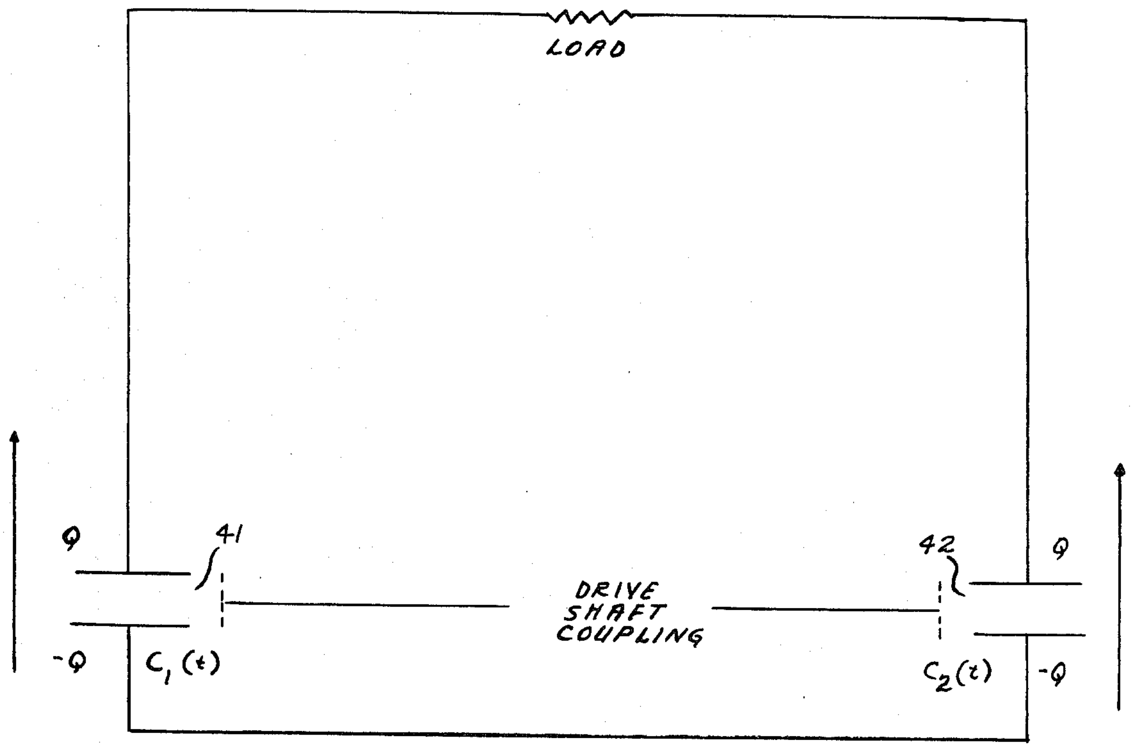
LOAD
Q
-Q
41
C₁(t)
DRIVE SHAFT COUPLING
42
Q
-Q
C₂(t)

Fig-16

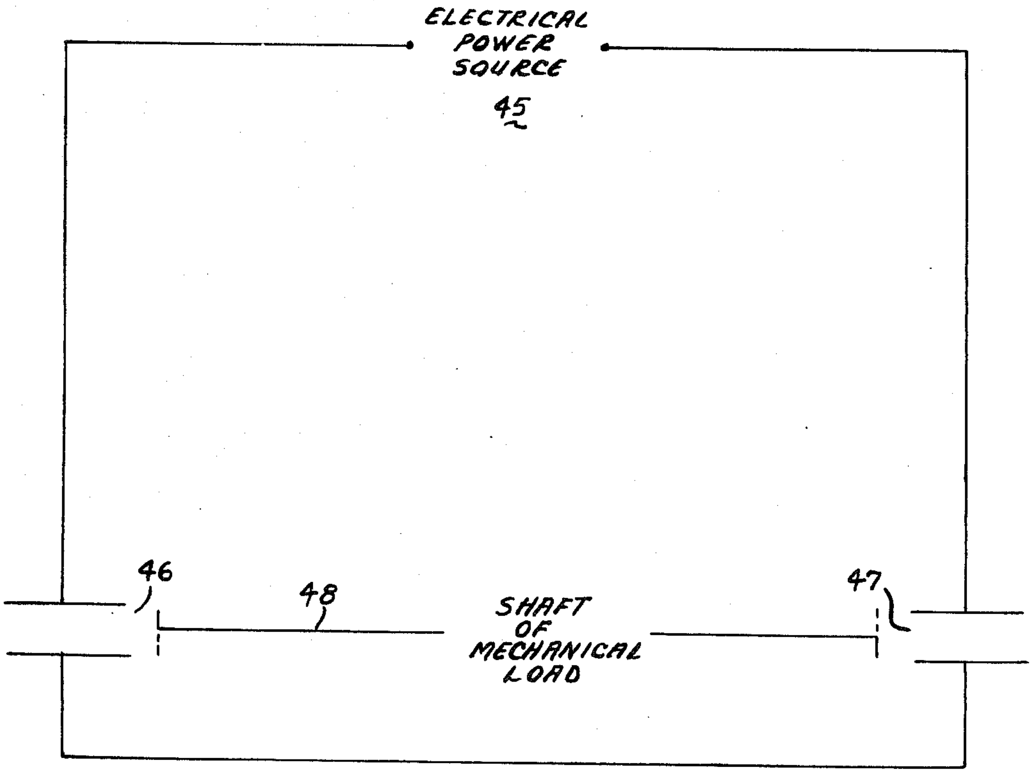
ELECTRICAL POWER SOURCE
SHAFT OF MECHANICAL LOAD

Fig 17

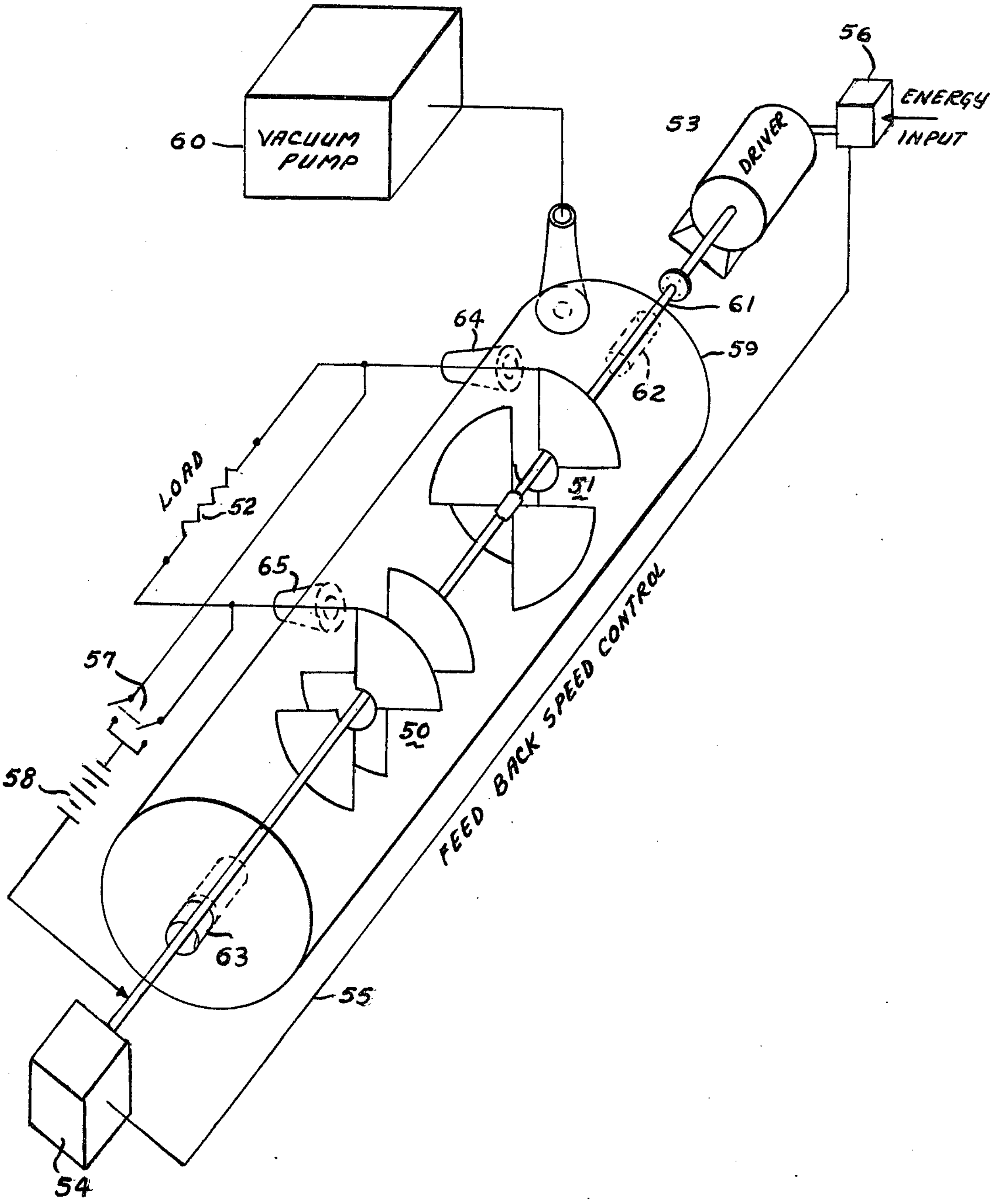
VACUUM PUMP
ENERGY INPUT
DRIVER
LOAD
FEED BACK SPEED CONTROL

Fig-18

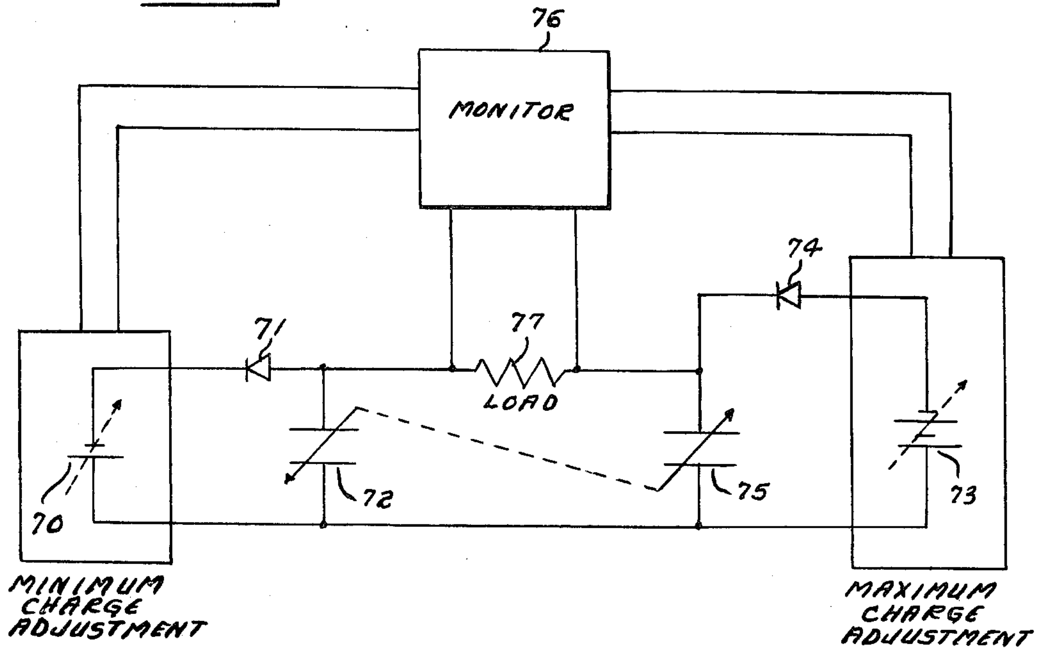
MONITOR
LOAD
MINIMUM CHARGE ADJUSTMENT
MAXIMUM CHARGE ADJUSTMENT

Fig-19

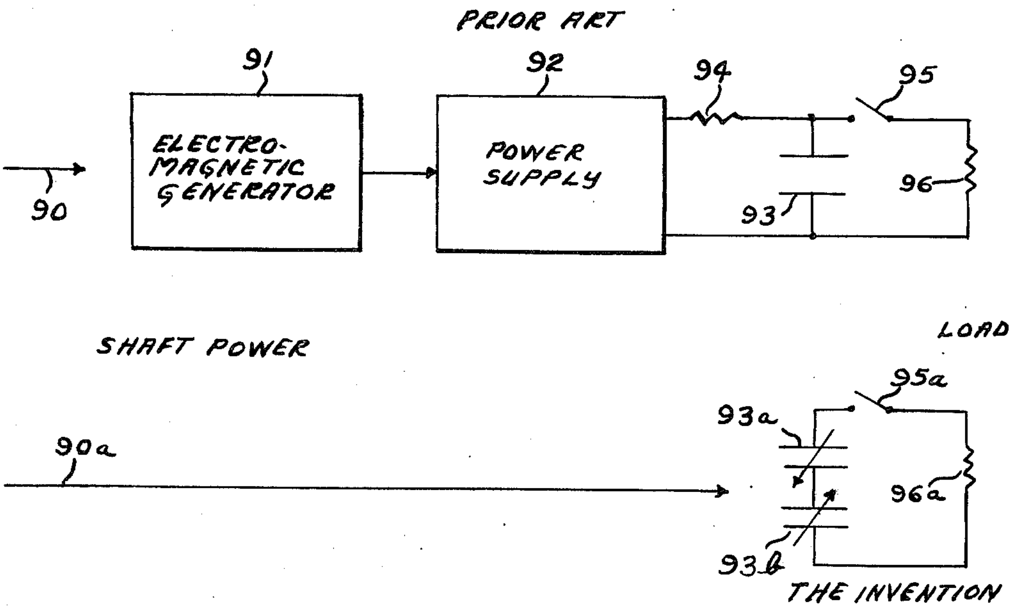
PRIOR ART
ELECTRO-MAGNETIC GENERATOR
POWER SUPPLY
SHAFT POWER
LOAD
THE INVENTION

ELECTROSTATIC ENERGY CONVERSION SYSTEM

RIGHTS OF THE GOVERNMENT

The invention described herein may be manufactured and used by or for the Government of the United States for all governmental purposes without the payment of any royalty.

BACKGROUND OF THE INVENTION

The field of the invention is in electrostatic machines.

Electrostatic machines have been well known for many years. The Toepler-Holtz and the Wimshurst machines were early electrostatic generators. Now their usage is mainly for demonstrations in the Physics Laboratory. The only remaining electrostatic machine to remain in extensive usage is the Van de Graaff type electrostatic generator. It is presently used to generate extremely high voltage potentials for atomic research. Substantially all of the prior art devices are direct current type machines.

The following publications may be helpful in further understanding the art. Electrostatic Sources of Electric Power, John G. Trump, Electrical Engineering, Vol. 66, No. 6, June 1947, pages 525-534; U.S. Pat. No. 2,810,878 to patentee N. Felici, and U.S. Pat. No. 3,094,653 to patentees D. B. LeMay et al.

SUMMARY OF THE INVENTION

An ac electrostatic variable capacity energy conversion system utilizing a fixed contained charge functions as an electrical generator when mechanically driven or as a motor providing mechanical shaft power when electrically driven.

BRIEF DESCRIPTION OF THE DRAWING

FIG. 1 is a simplified pictorial-schematic diagram representative of a rotational generator embodiment of the invention;

FIG. 2 is a simplified pictorial schematic diagram representative of a reciprocally driven generator embodiment of the invention;

FIG. 3 is an electrical schematic diagram of the embodiment illustrated in FIG. 2;

FIG. 4 is a schematic diagram illustrative of the electrical charges and capacitances of the embodiment physically illustrated in FIG. 1;

FIG. 5 is a pictorial-schematic representation of an embodiment of the invention wherein the load is between dielectric materials;

FIG. 6 is a schematic diagram of the system shown in FIG. 5 illustrative of the charges and capacitances;

FIG. 7 is a schematic diagram of a pulsed energy input embodiment of the invention in State A;

FIG. 8 is the same embodiment except in State B;

FIG. 9 is the same embodiment except in State C;

FIG. 10 is the same embodiment except in State D;

FIG. 11 is a schematic representation of an embodiment of the invention having pulsed energy input with the generators indirectly connected to the load through dielectric material, in State A;

FIG. 12 is the same embodiment of FIG. 11 except in State B;

FIG. 13 is the same embodiment of FIG. 11 except in State C;

FIG. 14 is the same embodiment of FIG. 11 except in State D;

FIG. 15 is a schematic diagram of an embodiment of the invention having continuous mechanical input providing a continuous wave (cw) electrical output with the generators directly connected to the load;

FIG. 16 is a schematic diagram of an energy conversion system of the invention wherein the system functions as a motor receiving electrical energy and delivering mechanical energy;

FIG. 17 is a pictorial-block-schematic diagram of an embodiment of the invention for receiving rotational energy from a mechanical source and delivering electrical energy to a load;

FIG. 18 is a schematic drawing shown a control system for maintaining a given maximum and minimum charge on the generating capacitors; and

FIG. 19 is a block-schematic diagram showing the improvement of the disclosed invention over a typical prior art system.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

In embodiments of the invention, energy conversion is obtained for the Variable Capacitance Electrostatic Energy Conversion System by operation of variable capacitance electrostatic generators, involving contained charge, in paired opposition to one another across the load (or electrical source of power), either directly or indirectly through dielectric material, to provide for conversion of mechanical energy to electrical energy (or electrical energy to mechanical energy).

For a variable capacitance electrostatic generator, the electrical power input to the capacitor, P_e (watts), is equal to the increase in capacitor electrostatic field energy, W_{ef} (joules), per unit time, t (seconds), and the mechanical power output, P_M (watts):

$$P_e = (dW_{ef}/dt) + P_M$$

with

$P_e = Vdq/dt$

V = capacitor potential (volts)

q = charge on capacitor (coulombs)

and $W_{ef} = \frac{1}{2} CV^2$

C = capacitor capacitance (farads)

Also, $q = CV$

Therefore, $P_M = \frac{1}{2} V^2 (dC/dt)$

and the time rate of conversion of mechanical to electrical energy, dW_{ME}/dt, equal to $-P_M$, is given as

$$dW_{ME}/dt = -\frac{1}{2} V^2 (dC/dt)$$

Conversion of mechanical shaftpower to electrical power to a load by variable capacitance electrostatic generators operating in paired opposition to one another across the load (and involving contained charge) shall be considered first.

Examples of two specific physical arrangements of electrostatic generators operating in paired opposition to one another connected directly to a load R (ohms) are shown in FIGS. 1 and 2. Capacitance change is by rotary motion in the embodiment of FIG. 1 and reciprocating motion for the embodiment represented by FIG. 2. FIG. 2 is shown in partial schematic form in FIG. 3 and the electrical schematic diagram for the device of FIG. 1 is shown in FIG. 4. Referring to FIGS. 1 and 4, the capacitors 11 and 12 each range in value from high capacitance C_H to low capacitance C_L and are coupled by the drive shaft 13 so that they operate in opposition to each other; that is, when one capacitor has value C_H, the other capacitor has value C_L, and vice versa. The capacitors each are initially charged to possess a charge Q, and hence there exists a contained (or "trapped") charge of 2Q for the system. As variable capacitor **11** varies from C_H to C_L, its charge varies from $(Q + q_o)$ to $(Q - q_o)$; as variable capacitor **12** varies from C_L to C_H, its charge varies from $(Q - q_o)$ to $(Q + q_o)$. A charge $2q_o$ is directly transferred across the load **14**. As capacitor **11** is decreasing in capacitance mechanical energy is converted by it to electrical energy and electrical energy is converted to mechanical energy by variable capacitor **12** which is increasing in capacitance. However, since capacitor **12** is drive-shaft coupled to capacitor **11**, this mechanical energy may be considered as simultaneously reconverted to electrical energy. With rotation, capacitor **11** reaches the state initially occupied by capacitor **12** and capacitor **12** is now in the state initially occupied by capacitor **11**. As the drive shaft continues to rotate, the capacitors return to their respective initial states with another transfer of charge $2q_o$ across the load, and a cycle is completed.

A schematic-pictorial physical arrangement of the electrostatic generators operating in paired opposition to one another, indirectly connected to the load through dielectric (insulating) material, is shown in FIG. **5**; an electrical schematic diagram for this embodiment is shown in FIG. **6**. The advantages of using a dielectric as electrode material is well known as exemplified by U.S. Pat. No. 3,745,481 to patentees G. F. Erickson et al. In this embodiment for each cycle a charge of $4q_o$ is transferred by displacement across the load; there is charge q_o associated with the dielectric material capacitors of capacitance c_d.

The simplified embodiments presented above were for initial explicit exposition of the concept involved for the Energy Conversion System. The following embodiments, presented in more detail, will explain the operation of the system and aid in the comprehension of the invention.

The pulsed energy input systems with the generators directly connected to the load are perhaps the easiest to comprehend. The electrical schematic diagram of these embodiments is shown in FIG. **7**. It is similar to the electrical schematic diagram of previous FIG. **4** except for the added switch **20**. The pictorial diagram of FIG. **1** is also a suitable physical representation if a switch is placed in the line going to the load. The operation will be explained considering four states or positions of the capacitors after a movement from their prior state. It is assumed that the capacitors have been initially charged and that operation through previous cycles has taken place. The switch is triggered cyclicly as a function of drive shaft position.

State A (FIG. **7**): Capacitor **21** has capacitance C_H and a charge $(Q + q_o)$; capacitor **22** has capacitance C_L and a charge $(Q - q_o)$; the potentials (volts) across the capacitors are equal. (NOTE: The magnitude of the potentials associated with the various circuit elements are indicated by the relative length of the arrows in the schematics as presented in the figures, e.g., arrows **23** and **24**, with the tip of the arrow indicating the direction of increasing potential for the charge; the arrows, representing the potential, are not necessarily drawn to scale.) The switch is in closed position; the switch, for the purpose of this exposition, shall be considered as an ideal switch (that is, it has negligible inductance, infinitely small capacitance relative to other circuit elements, infinite resistance when open, and zero resistance when closed). The potential across a capacitor is equal to the charge on the capacitor divided by its capacitance, therefore;

$$(Q + q_o)/C_H = (Q - q_o)/C_L$$

The energy associated with the electrostatic field of a capacitor is given by one-half the value of its charge squared, divided by its capacitance. Therefore, the total electrical energy of the electric circuit in State A, W_A, is given as

$$W_A = \tfrac{1}{2}(Q + q_o)^2/C_H + \tfrac{1}{2}(Q - q_o)^2/C_L$$

The switch is opened and, as the drive shaft rotates, the capacitances of capacitors **1** and **2** change to State B.

State B (FIG. **8**): Capacitor **21** now has capacitance C_L and capacitor **22** has capacitance C_H. The total electrical energy for the electric circuit in State B, W_B, is given as

$$W_B = \tfrac{1}{2}(Q + q_o)^2/C_L + \tfrac{1}{2}(Q - q_o)^2/C_H$$

The potential rise across capacitor **21** is equal to the potential drop across the switch **20** and across capacitor **22**.

State C (FIG. **9**): The switch **20** is now closed. The potential rise across capacitor **21** is now equal to the potential drop across the load and the potential drop across capacitor **22**. It is assumed, for the purpose of this exposition, that the transient response time of the electric circuit is much faster than the time required for one complete cycle of the variable capacitors (that is, the time required for the variable capacitors, starting at respective initial states, to return to those respective states, as the drive shaft rotates). With q now representing the charge being transferred, the equation for the electrical circuit response is given as

$$(Q + q_o - q)/C_L = R\,(dq/dt) + (Q - q_o + q)/C_H$$

When $t = 0$, $q = 0$; when $t = \infty$, $q = 2q_o$.
The solution of the previous equation is as follows:

$$q = 2q_o[1 - \exp[-(1/R)\,(1/C_L + 1/C_H)t]]$$

And from the earlier equation $(Q + q_o)/C_H = (Q - q_o)/C_L$

$$q_o = \frac{(1/C_L - 1/C_H)}{(1/C_L + 1/C_H)}\,Q$$

State D (FIG. **10**): After charge rearrangement by a flow of current through the load, capacitor **22** is in the same state as capacitor **21** in State A; capacitor **21** is in the same state as capacitor **22** in State A; and the total electrical energy for the electrical circuit in State D, W_D, is given as

$$W_D = \tfrac{1}{2}(Q + q_o)^2/C_H + \tfrac{1}{2}(Q - q_o)^2/C_L$$

By subtracting this equation from the one in State B, it is seen that:

$$W_B - W_D = 2q_o^2\,(1/C_L + 1/C_H) = 4[\tfrac{1}{2}\,q_o^2(1/C_L + 1/C_H)]$$

-continued

$$ {}_{o}^{\infty} R\left(\frac{dq}{dt}\right)^2 dt = 2q_o^2(1/C_L + 1/C_H) = W_B - W_D $$

Electrostatic energy is delivered to the load. As the drive shaft **25** rotates, capacitors **21** and **22** will again go through States B and C, with their roles interchanged, until the capacitors return to their respective states as given in State A. For this second half of the cycle, the same amount of energy is delivered to the load as was delivered in the first half of the cycle; hence, the energy delivered to the load per cycle, W, is given as follows:

$$ W = 2(W_B - W_D) = 8[\tfrac{1}{2}\, q_o^2(1/C_L + 1/C_H)] $$

The limit on energy transfer may be determined as follows. The breakdown potential for the variable capacitors, V_{CB} (volts), may be related to the charge on the variable capacitors:

$$ (Q + q_o)/C_L = V_{CB} $$

By use of this equation and the one given at the end of State C, the limiting value of q_o may be given as follows:

$$ q_o = (C_L/2)\,(1 - C_L/C_H)V_{CB} $$

By substitution of the above value of q_o into the foregoing equation for W, the limit on W, W_{CB}, is given as

$$ W_{CB} = (1 + C_L/C_H)\,(1 - C_L/C_H)^2 C_L V_{CB}^2 $$

The range of W is then given as follows: $0 \leqq W < W_{CB}$

For non-negligible circuit inductance, such as, for example, a series inductance L (henries), the equation for the electrical circuit response is given as

$$ (Q + q_o - q)/C_L = R(dq/dt) + L(d^2q/dt^2) + (Q - q_o + q)/C_H $$

When $t = 0$, $q = 0$; when $t = \infty$, $q = 2q_o$; the solutions are given as follows.

For over damped response, $(R/2L)^2 > (1/L)\,(1/C_L + 1/C_H)$,

$$ q = 2Q[\frac{C_H - C_L}{C_H + C_L}]\,[1 - \exp\left(-\frac{Rt}{2L}\right)[\frac{R}{2La}\sinh(at) + \cosh(at)]] $$

With

$$ a = \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{L}\left(\frac{1}{C_L} + \frac{1}{C_H}\right)} $$

For critically damped response, $(R/2L)^2 = (1/L)\,(1/C_L + 1/C_H)$,

$$ q = 2Q[\frac{C_H - C_L}{C_H + C_L}]\,[1 - \exp\left(-\frac{Rt}{2L}\right) - \frac{Rt}{2L}\exp\left(-\frac{Rt}{2L}\right)] $$

For under damped response, $(R/2L)^2 < (1/L)\,(1/C_L + 1/C_H)$,

$$ q = 2Q[\frac{C_H - C_L}{C_H + C_L}]\,[1 - \exp\left(-\frac{Rt}{2L}\right)[\frac{R}{2Lb}\sin(bt) + \cos(bt)]] $$

With b being given by a, as defined above, divided by the square root of minus one. The relation between Q and q_o is

$$ q_o = \frac{(1/C_L - 1/C_H)}{(1/C_L + 1/C_H)}\,Q, $$

as previously stated.

An embodiment of the invention for operation with pulsed energy input and the electrical load connected through dielectric material, such as barium titanate, is schematically illustrated in FIG. **11**. This type of system is ideally suited for the production of molecular excitation and volume ionization of high-pressure lasing media. As in the previous embodiment, four states or end positions of arcs of rotational movement will be illustrated.

State A (FIG. **11**): Capacitor **31** has capacitance C_H and a charge $(Q + q_o)$; capacitor **32** has capacitance C_L and a charge $(Q - q_o)$. The potential rise across capacitor **32** is equal to the sum of the potential drops across each dielectric electrode (**33** and **34**, each of capacitance c_d) and capacitor **31**:

$$ (Q - q_o)/C_L = 2q_o/c_d + (Q + q_o)/C_H $$

The total electrical energy of the electric circuit in State A is given as follows:

$$ W_A = \tfrac{1}{2}\,(Q + q_o)^2/C_H + q_o^2/c_d + \tfrac{1}{2}\,(Q - q_o)^2/C_L $$

Switch **36** is opened and, as drive shaft **35** rotates, the capacitance of capacitors **31** and **32** changes, and state B is reached.

State B (FIG. **12**): Capacitor **31** now has capacitance C_L and capacitor **32** has capacitance C_H. The total electrical energy for the electric circuit in State B is given as follows:

$$ W_B = \tfrac{1}{2}\,(Q + q_o)^2/C_L + q_o^2/c_d + \tfrac{1}{2}\,(Q - q_o)^2/C_H $$

The potential rise across capacitor **31** and the two capacitances c_d equals the potential drop across the switch and capacitor **32**.

State C (FIG. **13**): Switch **36** is now closed. The potential rise across capacitor **31** and the two capacitances c_d equals the potential drop across the load and capacitor **32**. The equation for the electrical circuit response is given as follows:

$$ (Q + q_o - q)/C_L + 2(q_o - q)/c_d = R(dq/dt) + (Q - q_o + q)/C_H $$

When $t = 0$, $q = 0$; when $t = \infty$, $q = 2q_o$.

Solution of the equation is as follows:

$$ q = 2q_o[1 - \exp[-(1/R)\,(1/C_L + 2/c_d + 1/C_H)t]] $$

With, as given previously,

$$ q_o = \frac{(1/C_L - 1/C_H)Q}{(1/C_L + 2/c_d + 1/C_H)} $$

State D (FIG. **14**): After charge rearrangement, capacitor **32** is in the same state as capacitor **31** in State A; capacitor **31** is in the same state as capacitor **32** in State A; and the polarity of the charge on the dielectric capacitances has been reversed. The total electrical energy for the electric circuit in State D is given as follows:

$$W_D = \tfrac{1}{2}(Q + q_o)^2/C_H + q_o^2/c_d + \tfrac{1}{2}(Q - q_o)^2/C_L$$

By combining equations, it is seen that:

$$W_B - W_D = 2q_o^2(1/C_L + 2/c_d + 1/C_H) = 4[\tfrac{1}{2} q_o^2 (1/C_L + 2/c_d + 1/C_H)]$$

And that:

$$\int_0^\infty R\left(\frac{dq}{dt}\right)^2 dt = 2q_o^2(1/C_L + 2/c_d + 1/C_H) = W_B - W_D$$

Electrostatic energy is delivered to the load. As drive shaft **35** rotates, capacitors **31** and **32** will again go through States B and C, with their roles interchanged, until the capacitors return to their respective states as given in State A. For this second half of the cycle, the same amount of energy is delivered to the load as was delivered in the first half of the cycle; hence, the energy delivered to the load per cycle, W, is given as follows:

$$W = 2(W_B - W_D) = 8[\tfrac{1}{2} q_o^2](1/C_L + 2/c_d + 1/C_H)$$

Assuming q_o limited by the breakdown potential of the dielectric material, V_{dB}, $q_o = c_d V_{dB}$. Then, the limit on W, W_{dB}, is as follows:

$$W_{dB} = 4(c_d/C_L + 2 + c_d/C_H)c_d V_{dB}^2$$

The range of W is then given as follows: $0 \leqq W < W_{dB}$

FIG. **15** schematically illustrates an embodiment of the invention having continuous mechanical energy input and providing a continuous wave output to the loads. It is similar to previous FIG. **4** except for designations of charge and capacitance.

Consider the system schematic as shown in FIG. **15**; and, at time $t = 0$, the variable capacitors **41** and **42** have the same capacitance, the same charge Q, and hence are at the same potential. Capacitor **42** is out of phase with capacitor **41** by half a period; the capacitors vary periodically with period T (seconds).

The circuit equation for the transported charge q is as follows: assuming negligible circuit inductance:

$$(Q - q)/C_1 = R(dq/dt) + (Q + q)/C_2$$

with $q = 0$, for $t = 0$.

With sinusoidal input to the load, for the following assumed capacitance:

$$C_1 = \frac{1}{\frac{1}{C_H} + \left(\frac{1}{C_L} - \frac{1}{C_H}\right)\left[\frac{1 - \cos\frac{2\pi}{T}\left(t + \frac{T}{4}\right)}{2}\right]}$$

$$C_2 = \frac{1}{\frac{1}{C_H} + \left(\frac{1}{C_L} - \frac{1}{C_H}\right)\left[\frac{1 + \cos\frac{2\pi}{T}\left(t + \frac{T}{4}\right)}{2}\right]}$$

$$q = \frac{\frac{Q}{R}\left[\frac{1}{C_L} - \frac{1}{C_H}\right]}{\sqrt{\left(\frac{2\pi}{T}\right)^2 + \left[\frac{1}{R}\left(\frac{1}{C_L} + \frac{1}{C_H}\right)\right]^2}} X$$

$$\left[\begin{array}{l} \cos\left[\frac{\pi}{2} - \arctan\left(\left[\frac{2\pi}{T}\right] / \left[\frac{1}{R}\left(\frac{1}{C_L} + \frac{1}{C_H}\right)\right]\right)\right] \exp\left[-\frac{1}{R}\left(\frac{1}{C_L} + \frac{1}{C_H}\right)t\right] \\ -\cos\left[\frac{2\pi}{T}\left(t + \frac{T}{4}\right) - \arctan\left(\left[\frac{2\pi}{T}\right] / \left[\frac{1}{R}\left(\frac{1}{C_L} + \frac{1}{C_H}\right)\right]\right)\right] \end{array}\right]$$

with average power, P (watts), given, for the electrical circuit time constant,

$$1/\left[\frac{1}{R}\left(\frac{1}{C_L} + \frac{1}{C_H}\right)\right],$$

much less than T, and after transient decay, as follows:

$$P = \frac{1}{T}\int_t^{t+T} R\left(\frac{dq}{dt}\right)^2 dt = \frac{2\pi^2 R Q^2 (C_H - C_L)^2}{T^2 (C_H + C_L)^2}$$

and with Q limited by the breakdown potential V_{CB} of the capacitors as follows:

$$\left[\frac{(Q - q)}{C_1}\right]_{maximum} = \frac{Q(C_H + C_L)}{2C_H C_L} = V_{CB}$$

By use of these last two equations, the limit on P, P_{CB}, is:

$$P_{CB} = \frac{8\pi^2 C_H C_L (C_H - C_L)^2 V_{CB}^2}{T^2 (C_H + C_L)^3 \left(\frac{1}{R}\right)\left(\frac{1}{C_L} + \frac{1}{C_H}\right)}$$

The range of P is then given as follows: $0 \leqq P < P_{CB}$

As with the pulsed embodiments the load on the *cw* embodiments may be an electrostatic field between dielectric electrodes. It is also to be understood that the electrical outputs of generating embodiments of the invention, while generally of quite high electrical potential, may be rectified (and filtered), or run through a transformer effecting an impedance change, or any of the other conventional electrical changes and usages that are common practice with continuous wave or pulsed alternating current voltages.

It is also to be understood that for simplicity, only one movable and one fixed plate has been shown for a variable capacitor, and that embodiments may be constructed that have capacitors with many fixed and movable plates. A plurality of embodiments may be connected in series for providing higher output voltages and better voltage breakdown characteristics. The polarity of the charge placed on the capacitors, i.e., the common polarity of all the fixed plates of an embodiment with respect to the movable plates is immaterial.

The electrostatic energy conversion system of the invention may also be operated as a motor. FIG. **16** is a simplified schematic representation of an embodiment of the invention for receiving electrical energy from a power source **45** and providing torque to rotate shaft **48**. Variable capacitors **46** and **47** are similar to those in the generating device, with the capacitance of capacitor **46** being a maximum when the capacitance of capacitor **47** is a minimum. Shaft **48** is driven by the forces of the capacitor plates to rotate in synchronism with the frequency of source **45**. Thus, unless the frequency of source **45** is relatively slowly increased from dc to the desired operating frequency, other conventional means such as are normally used to get synchronous motors up to synchronous speed may be used to bring shaft **48** up to substantially synchronous speed with the line frequency of the source of electrical power before connection is made.

It is to be understood that the devices of the invention, generally like other electrostatic devices from a practical consideration, generally function with relatively high voltages compared to conventional electromagnetic equipment. As pointed out by Trump in the previously referenced AIEE (now IEEE) paper, two metallic plates of 100 square inches area which are parallel and insulated from each other, with an electric field of 300 volts per centimeter between them have an electrostatic force of attraction of about 0.0005 pound. If the active gradient is increased to 30,000 volts per centimeter, which is the highest voltage that can be insulated in atmospheric air, the electrostatic force between the plates is increased to little more than 0.5 pound. When the plates are immersed in a high vacuum and the gradient increased to 3 million volts per centimeter, the electrostatic force then becomes approximately 5,700 pounds.

From the foregoing, it is readily apparent that for embodiments of the invention to be practical, the variable capacitors should be operated in a high vacuum. For embodiments used in outer space, this is generally not a problem, but for terrestial embodiments, it is generally preferred that a sealed chamber having a high vacuum be used. Insulative gases such as sulphur hexafluoride or air at high pressures may be used instead of the high vacuum to provide the necessary insulation at the high voltages required for practical operation. Another reason for generally preferring the high vacuum environment for embodiments of the invention is the elimination of drag on the moving members. A complete system for the cw generation of electrical power under normal terrestial conditions is shown schematically in FIG. **17**. Variable capacitors **50** and **51**, with rotation, pass a contained charge back and forth between them causing a current to flow through load **52** delivering power to it. Driver **53** may be any suitable device that can supply the torque necessary to rotate the movable plates of the capacitors such as an electric motor, a turbine, or even a hand crank. Generally, it is desirable to provide conventional speed sensor **54** with conventional feedback control **55**, actuating energy input controller **56**, so that the frequency of the generated electrical energy supplied to load **52** is relatively constant. An initial charge is placed on variable capacitors **50** and **51** by momentarily closing switch **57** causing conventional high potential source **58** to place an equal potential charge of the same polarity on each capacitor. It is generally desirable to position the movable plates of the capacitors such that they have equal capacity when initially charged so that the initial charge magnitudes will also be equal. The capacitors are operated in a high vacuum in conventional sealed chamber **59**. Conventional vacuum pump **60** provides the high vacuum in the chamber (approximately 10^{-8} Torr). Mechanical drive shaft **61** extends through the chamber walls in magnetically sealed rotary motion feedthroughs **62** and **63**. These conventional, commercially available, sealed bearings having ferrofluid seals provide substantially zero leakage (at speeds up to 10,000 rpm) to a high vacuum on one side of the bearing seal and atmospheric pressure on the other side of the bearing. Conventional capacitor plate material such as stainless steel is suitable and generally preferable. Conventional high voltage feedthroughs **64** and **65** provide for transfer of charge from the stator capacitor plates to the load.

As previously stated, once the charge is placed on the capacitors it is not diminished or dissipated in supplying power to the load. It is a contained or trapped charge, and is not run down by the load. If the insulation within the chamber were perfect, the charge would never have to be replenished. Since nothing is perfect the charge will, over a long period of time, decrease. It may be replenished by momentarily actuating switch **57** as in the initial charging of the capacitors or the charge may be automatically held within predetermined maximum and minimum values on the capacitor plates by an electrical circuit as shown schematically in FIG. **18**. For simplicity of explanation, maximum voltage monitoring is shown on one side of the capacitor pair and minimum voltage monitoring on the other. It is to be understood that both a maximum and a minimum monitoring may be on the same side, or simultaneously on both sides, if desired. Potential source **70** and diode **71** add charge when the voltage potential of capacitor **72** is below that of source **70** when capacitor **72** is at maximum capacitance. Potential source **73** in cooperation with diode **74** removes charge when the voltage of capacitor **75** (when going through its minimum capacitance) is greater than the potential of **73**. By having potential sources **70** and **73** adjustable and controllable by conventional voltage monitor **76**, the correct value of trapped charge for the system is maintained for desired voltage (or power) for load **77**. The regulating system shown in FIG. **18** may also be used to compensate for leakage losses.

FIG. **19** illustrates in block schematic form the advantages of the disclosed invention in supplying a high potential to a load such as a laser. In the conventional prior art systems, shaft power **90** turns conventional electrical generator **91** generating relatively low voltage, high current energy, to operate conventional high voltage power supply **92** which changes capacitor **93** through resistor **94**. Capacitor **93** discharges when switch **95** is closed to supply the load **96** with energy. In the present invention, shaft power **90*a*** directly actuates inversely ganged capacitors **93*a*** and **93*b*** providing an electrostatic energy charge equivalent to that placed on capacitor **93** by power supply **92**. Thus, the closing of switch **95*a*** provides the same energy to the identical load **96*a*** as in the prior art system. In the illustration bipolar pulses flow to load **96*a***. If unipolar pulses are desired for the load **96*a*** as for load **96** in the prior art circuit, conventional rectification of the bipolar pulse may be used.

I claim:

1. An alternating current electrostatic energy converter comprising:

a. a first and a second variable capacitor, each having a maximum and a minimum capacitance, connected in paired opposition and inversely ganged such that one capacitor passes through a maximum capacitance as the other capacitor is going through a minimum capacitance;

b. means for placing an initial and contained electrical charge on the said first and second capacitors;

c. means for actuating the said capacitors through their maximum and minimum capacitance passing contained charge between them; and

d. means for extracting power from the said capacitors.

2. An electrostatic alternating current generating system comprising:

a. a first and a second variable capacitor, each having a maximum and a minimum capacitance, inversely ganged on a common shaft such that one capacitor passes through a maximum capacitance as the other capacitor is going through a minimum capacitance;

b. means for placing an initial and contained electrical charge on the said first and second capacitors;

c. means for actuating the said shaft moving capacitors through their maximum and minimum capacitance;

d. an electrical load; and

e. means for connecting the said first and second variable capacitors in paired opposition to one another across the said load whereby contained charge is passed back and forth between the said capacitors through the said load.

3. An electrostatic alternating current generator comprising:

a. a first variable capacitor having a fixed plate and a movable plate providing a maximum and a minimum capacitance between the said fixed plate and the said movable plate;

b. a second variable capacitor having a fixed plate and a movable plate providing a maximum and a minimum capacitance between the said fixed plate and the said movable plate;

c. means for mechanically coupling the said movable plate of the said first capacitor to the movable plate of the said second capacitor such that the capacity of one variable capacitor is at a maximum capacitance when the other variable capacitor is at a minimum;

d. means for electrically connecting the said movable plates;

e. means for placing a contained electrical charge on the said variable capacitors such that the said fixed plates have like polarity charge;

f. means for actuating the said variable capacitors through a change of capacitance; and

g. means for connecting an electrical load between the said fixed plates of the variable capacitors, whereby with change of capacitance of the said capacitors, contained charge is passed between the said capacitors through the said load.

4. The apparatus as claimed in claim 3 wherein the said means for placing a charge on the said variable capacitors includes means for placing equal initial charges on the said variable capacitors.

5. The apparatus as claimed in claim 4 wherein a means for providing an evacuated enclosure and a means for positioning the said first and second variable capacitors within the said evacuated enclosure are provided.

6. The apparatus as claimed in claim 5 wherein the said means for actuating the said variable capacitors is rotational.

7. The apparatus as claimed in claim 5 wherein the said means for actuating the said variable capacitors is reciprocating.

8. A synchronous alternating current electrostatic motor comprising:

a. a first and a second variable capacitor each having a maximum and a minimum capacitance, connected in paired opposition inversely ganged on a common shaft such that one capacitor passes through a maximum capacitance as the other capacitor is going through a minimum capacitance;

b. means for placing an initial, contained, electrical charge on the said first and second capacitors;

c. means for placing an alternating potential voltage on the said first and second capacitors; and

d. means for connecting a mechanical load to the said shaft.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 4,127,804
DATED : Nov. 28, 1973
INVENTOR(S) : Onezime P. Breaux

It is certified that error appears in the above-identified patent and that said Letters Patent are hereby corrected as shown below:

On the front page delete:

--/73/ Assignee: The United States of America as represented by the Secretary of the Air Force, Washington, D.C.

Signed and Sealed this

Twenty-fourth Day of *April 1979*

[SEAL]

Attest:

RUTH C. MASON
Attesting Officer

DONALD W. BANNER
Commissioner of Patents and Trademarks

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 4,127,804
DATED : Nov. 28, 1973
INVENTOR(S) : Onezime P. Breaux

It is certified that error appears in the above-identified patent and that said Letters Patent are hereby corrected as shown below:

On the front page delete:

--/73/ Assignee: The United States of America as represented by the Secretary of the Air Force, Washington, D.C.

Signed and Sealed this

Twenty-fourth **Day of** *April 1979*

[SEAL]

Attest:

RUTH C. MASON
Attesting Officer

DONALD W. BANNER
Commissioner of Patents and Trademarks

July 7, 1953 N. H. LEHRER 2,644,903

SELF-STARTING ELECTROSTATIC GENERATOR

Filed Sept. 25, 1952

FIG. 1

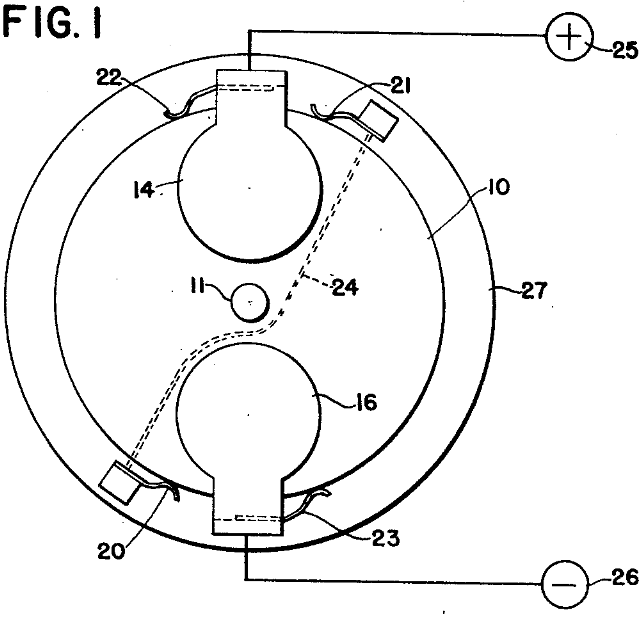
25
22
21
14
10
11
24
27
16
20
23
26

FIG. 2

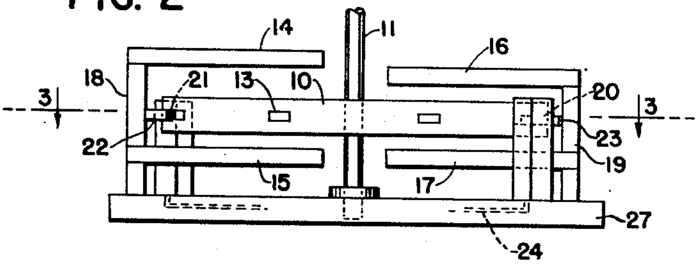
14
11
16
18
21
13
10
20
3
3
22
23
19
15
17
27
24

FIG. 3

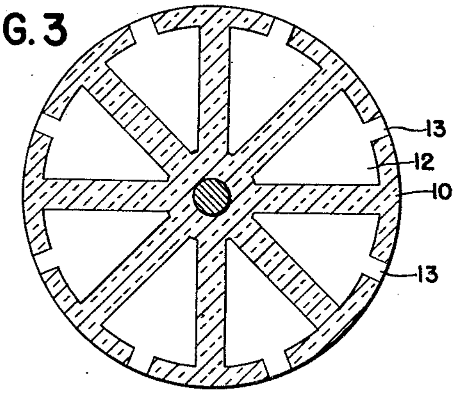

NORMAN H. LEHRER
INVENTOR.

BY Ralph E. Bitner

ATTORNEY

Patented July 7, 1953 **2,644,903**

UNITED STATES PATENT OFFICE

2,644,903

SELF-STARTING ELECTROSTATIC GENERATOR

Norman H. Lehrer, Bloomfield, N. J., assignor to Chatham Electronics Corporation, Newark, N. J.

Application September 25, 1952, Serial No. 311,360

5 Claims. (Cl. 310—6)

The subject matter of this patent may be used by or for the Government for governmental purposes without the payment of any royalties to me.

This invention relates to influence type electrostatic generators, specifically those employing feedback, and has particular reference to an arrangement of components which provides reliable starting action and which always starts with the same polarity.

Prior art generators of the induction type are not self-starting and a separate source is generally used to apply a charge to one of the induction plates to start the generator. The separate source can be a friction generator, a battery, or any other source of direct current power.

The present machine is an improvement over prior art machines in that it provides reliable starting action with the desired polarity without requiring additional parts. Reversal of polarity caused by polarization of a dielectric covering the stator plates is prevented by leaving the stator plates uncovered.

One of the objects of the present invention is to provide an improved electrostatic generator which avoids one or more of the disadvantages and limitations of prior art generators.

Another object of the invention is to provide an electrostatic influence generator which always starts its generating action without separate excitation.

Another object of the invention is to provide an electrostatic generator which always starts with the same polarity.

Another object of the invention is to reduce the cost and complexity of self-starting generators by using only the parts generally employed in generators that are not self-starting.

The invention comprises an electrostatic generator which includes a disk shaped rotor containing a number of conducting sectors imbedded in the rotor insulating material. Brush contact with the conducting sectors is made at the edge of the rotor disk where small portions of the sectors project. One of the collecting brushes is made of molybdenum or similar material and, in addition to making contact with the sector portions, rubs against the edge of the insulator disk, causing a charge to be generated due to frictional phenomena. The usual set of stator influence plates is mounted close to the disk surface for inducing a charge in the rotating sectors. One of the stator plates is separated from the disk surface a greater distance than the other plates. One set of brushes is connected to a diametral conductor which isolates charges on the sectors as they pass through the field of the stator plates. Another set of brushes, connected to the stator plates, keeps the plates charged and produces the voltage for the load circuit.

For a better understanding of the present invention, together with other and further objects thereof, reference is made to the following description taken in connection with the accompanying drawings.

Fig. 1 is a plan view of an electrostatic influence generator shown partly in schematic representation.

Fig. 2 is a side view of the generator showing the arrangement of the brushes.

Fig. 3 is a cross sectional view of the rotor taken along line 3—3 of Fig. 2.

Referring now to the figures, a rotor **10** is made of non-conducting material such as Lucite and is mounted for rotation on a central shaft **11**. This shaft may be turned by any convenient means, by hand, or by an electric motor. Imbedded in the body of rotor **10** are a number of conducting sectors **12**, formed in a wedge-shaped outline with a small tab portion **13** which extends to the periphery of the disk to make contact with a brush or collector electrode.

The conducting sectors **12** may be positioned in the rotor body as indicated in Fig. 3, in a single plane, or they may be mounted in a dual plane arrangement in overlapping positions in order to increase the current output.

In order to induce an electric charge on the sectors, four induction stator plates **14**, **15**, **16**, and **17** are mounted in the positions shown, close to the rotor face. These plates are insulated from the rotor only by an air gap. Plate **14** is mounted a greater distance from the rotor face than the other plates. The stator plates are secured by conductive brackets **18** and **19** which also serve as supports for two of the brushes which make contact with the rotor edge.

Four brushes **20**, **21**, **22**, and **23** are mounted around the periphery of the disk to make contact with the sector tabs as they revolve. Two of these brushes **20** and **21** are connected by the well known diametral conductor which is generally a component of all electrostatic induction generators and which segregates the charges on the sectors as they pass through the electric field of the stator plates.

Brushes **22** and **23** are connected to stator plates **14**, **15** and **16**, **17** respectively. They collect part of the charge of the rotating sectors to maintain the stator plates at maximum potential.

One of the diametral brushes **20** is made of molybdenum and is mounted a little out of line so that part of the brush rubs on the edge of the rotor insulation. This rubbing action generates sufficient frictional charge to start the machine. Any one of the brushes **20**, **21**, **22**, and **23** can be positioned so as to function as a brush and also to generate the frictional electric charge needed to start the machine. Also a separate brush can be added to generate this charge if desired.

Two terminals **25** and **26** are provided for connecting to a load circuit. These terminals are connected to the stator plate combinations **14**, **15** and **16**, **17**. An insulating base **27** acts as a support for all the components including the rotor shaft **11**. Either terminal may be grounded.

The operation of the generator is as follows: at the start let it be assumed that all parts of the machine are at zero potential and no accumulated charge exists on any of the plates or sectors. Power is applied to the rotor and as it revolves the rubbing of brush **20** along the surface of the rotor insulator generates an electrical charge, some of which resides at the upper edge of the rotor. This charge, which is positive in the case of molybdenum against Lucite, induces negative charges on both stator plates **14** and **16**. Since plate **16** is closer to the charged surface on which the positive charge resides it has a negative charge of higher voltage induced on its lower surface than the negative charge on the lower surface of plate **14**. When the rotating sectors are brought within the range of influence of these stator plates, the sectors under plate **16** are charged with a greater positive potential than the sectors under plate **14** and when the diametral conductor **24** makes and breaks the connection between the sectors, the sector under plate **14** is charged negatively and the sector under **16** is charged positively.

When the sectors move away from the stator plates, their potential increases as the capacity decreases in accordance with the well known law which states that the quantity of electricity is equal to the product of the capacity and the potential. When the sectors arrive at the opposite plate their charge is collected by one of the brushes **22** or **23** and the stator plates are charged to a higher potential. This cumulative process continues until leakage or a load limits the potential to a maximum value.

The potential of an isolated conductor in an electric field is proportional to the work done upon it to move it from one portion of the field to another. Moving the sectors from an oppositely charged plate such as **16** increases the potential in proportion to the decrease in capacity and in a similar manner, moving the sectors toward a similarly charged plate such as **14** also increases the potential in proportion to the increase in capacity.

The above is believed to be a correct explanation of the principles underlying applicant's invention. However, further investigation may lead to a modification of this theory. It is to be understood that the invention is independent of any theory that may be advanced to account for the results obtained.

While there have been described and illustrated specific embodiments of the invention, it will be obvious that various changes and modifications

may be made therein without departing from the field of the invention which should be limited only by the scope of the appended claims.

I claim:

1. A self-starting electrostatic influence generator comprising; a rotor made of insulating material; a plurality of conducting sectors imbedded in the rotor material, each having a connection to the surface of said rotor; a pair of conducting stator plates mounted adjacent to the rotor, one of said plates mounted at a greater distance from the rotor than the other; a diametral conductor connecting brushes on opposite sides of the rotor, said brushes mounted to make contact with said sector connections as the sectors pass through the electric field of the stator plates; a pair of collector brushes, each connected to a stator plate and mounted to make contact with said sector connections as the sectors pass into the electric field of the stator plates; and a brush mounted so as to rub on the rotor insulating material.

2. A self-starting electrostatic influence generator comprising; a rotor made of insulating material mounted for rotation on a shaft; a plurality of conducting sectors imbedded in the rotor material, each having a connection to the surface of said rotor; a pair of conducting stator plates mounted adjacent to the rotor, one of said plates mounted at a greater distance from the rotor than the other; a diametral conductor connecting brushes on the opposite sides of the rotor, said brushes mounted to make contact with said sector connections as the sectors pass through the electric field of the stator plates; a pair of collector brushes, each connected to a stator plate and mounted to make contact with said sector connections as the sectors pass into the electric field of the stator plates; one of said brushes mounted so as to rub on the rotor insulating material.

3. A self-starting electrostatic influence generator comprising; a rotor disk made of insulating material mounted for rotation on a shaft; a plurality of conducting sectors imbedded in the rotor material equally spaced and insulated from each other, each of said sectors having a connection to the surface of said rotor; a pair of conducting stator plates mounted adjacent to the rotor, one of said plates mounted at a greater distance from the rotor than the other; a diametral conductor connecting brushes on the opposite sides of the rotor disk, said brushes mounted to make contact with said sector connections as the sectors pass through the electric field of the stator plates; a pair of collector brushes, each connected to a stator plate and mounted to make contact with said sector connections as the sectors pass into the electric field of the stator plates; one of said brushes mounted so as to rub on the rotor insulating material and generate frictional charges thereon.

4. A self-starting electrostatic influence generator comprising; a rotor disk made of insulating material and mounted for rotation about an axis normal to the disk faces; a plurality of flat conducting sectors imbedded in the rotor material, insulated from each other and parallel to the disk faces; each of said sectors having a connection to the periphery of the rotor; a pair of conducting stator plates mounted parallel to the rotor faces, one of said plates mounted at a greater distance from the rotor face than the other; a diametral conductor, terminated by brushes on opposite sides of the rotor disk, said brushes mounted so as to make contact with the sector connections

as the sectors pass through the electric field of the stator plates; a pair of collector brushes, each connected to a stator plate and mounted so as to make contact with said sector connections as the sectors pass into the electric field of the stator plates; one of said brushes mounted so as to rub on the rotor insulating material and generate frictional charges thereon.

5. A self-starting electrostatic influence generator comprising; a rotor disk made of insulating material and mounted for rotation about an axis normal to the disk faces; a plurality of flat conducting sectors imbedded in the rotor material, insulated from each other and parallel to the disk faces; each of said sectors having a connection to the periphery of the rotor where contact can be made with one or more stationary brushes; a pair of conducting stator plates mounted adjacent to the rotor faces and parallel thereto, one of said plates mounted at a greater distance from the rotor face than the other; a diametral conductor, terminated by brushes on opposite sides of the rotor disk, said brushes mounted so as to make contact with the sector connections as the sectors pass through the electric field of the stator plates; a pair of collector brushes, each connected to a stator plate and mounted so as to make contact with said sector connections as the sectors pass into the electric field of the stator plates; one of said diametral or collector brushes mounted so as to rub on the rotor insulating material as well as to make contact with the sector connections.

NORMAN H. LEHRER.

References Cited in the file of this patent

UNITED STATES PATENTS

Number	Name	Date
2,523,689	Felici	Sept. 26, 1950
2,577,466	Bosch	Dec. 4, 1951

1

Energy Harvesting Suits Remote Low-Power Devices

Wed, 02/02/2011 - 10:08am by Jon Titus, Senior Technical Editor

Grabbing "free" energy involves more engineering than buying an off-the-shelf transducer.

Jon Titus

Contrary to what you might think, the awareness of "green power" didn't spawn the drive to harvest energy. Low-power electronic fabrication technologies did the trick. They cut the power needs of small monitoring devices to the point where energy harvesting has started to make engineering and economic sense.

When engineers start to think about using energy-harvesting techniques to power a device, they must consider the power path. "You have an energy-harvesting device followed by an energy-conversion device, an energy-storage component, and ultra-low-power energy-management circuitry," said Steve Grady, vice president of marketing at Cymbet. "Although there is a lot of talk about 'battery-less' devices, energy-harvesting-based systems still need something to store energy for times when the energy transducer is inactive."

"A piezo-electric device can generate 10's to 100's of Volts while thermo-electric device or small solar panels generate microamps and millivolts," said Adrian Valenzuela, product marketing engineer at Texas Instruments. "So your power-management ICs must work with power across a wide range of currents and voltages. Say you need well-regulated 3-Volt power for an MCU," noted Valenzuela. "For some types of energy storage, you also need a low-power circuit to produce a well-regulated 4.2-Volt output to charge an energy-storage device."

"Energy-harvesting equipment needs an energy-storage device that has a high energy density, small size, and supports thousands of charge/discharge cycles" said Grady. "An energy-storage device also should exhibit a flat voltage profile and very low self discharge."

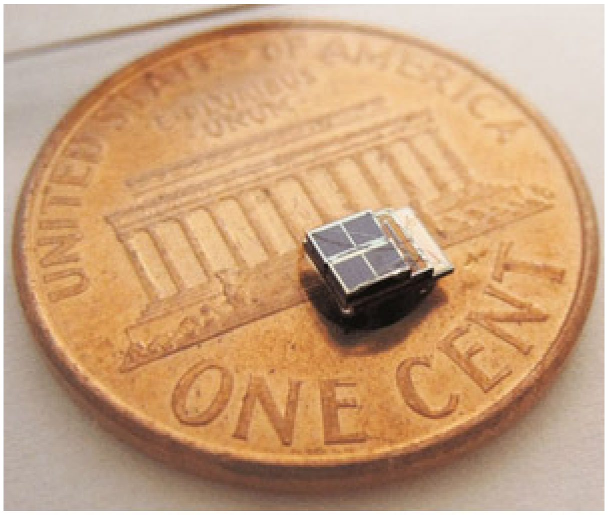

Figure 1. A Cymbet EnerChip placed on a penny shows the relative size of the energy-storage and -management circuits available now for energy-harvesting devices.

Figure 1. A Cymbet EnerChip placed on a penny shows the relative size of the energy-storage and -management circuits available now for energy-harvesting devices.

Cymbet manufactures rechargeable EnerChip solid-state batteries that offer energy storage in small SMT devices that can go through a reflow oven with other SMT ICs and components. The company's CBC3112, for example, provides a 3.3-V output at 12 uAh with an integrated power-management IC in a 7-by-7-mm SMT package. An EnerChip can go through more than 5000 charge/discharge cycles at a 10-percent depth of discharge.

In addition to figuring out how to gather, manage, and store energy, engineers face another challenge: money. "A battery costs 15 cents, so an energy-harvesting device cannot compete on cost," said TI's Valenzuela. "For now, engineers must work with exotic materials and devices, and they pass their high material costs on to consumers. In the majority of applications where engineers have a technical need for energy-harvested power the final decision almost always rests on the total cost of ownership."

"We also need to teach engineers how to think about power efficiency," noted Valenzuela. "Getting them to write energy-aware software is not easy, particularly when their application includes a wireless link. A wireless transceiver requires several orders of magnitude higher power than an MCU. So you must know how to compress data so it takes less time to transmit. And you must choose an antenna that produces the most efficient wireless signal for your equipment."

"We put a lot of effort into 'low-power firmware,' for a solar-energy-harvesting kit we created with Microchip," said Steve Grady of Cymbet. "When a harvester generates only a few microwatts or milliwatts, every Joule is precious. Inefficient use of a wireless transmitter at system start-up, for example, could completely drain the power source as a node searches for other wireless devices."

"We always have to coach our customers interested in energy harvesting about changing some of their code," continued Grady. "They just want to replace their existing power source with an energy-harvesting device, but that approach won't work. Sometimes success of a product depends as much on optimized firmware as on the harvesting device, energy storage and power-management circuitry."

Today the market for energy-harvesting devices seems a mile wide and an inch deep. "Every application is different," said Chris Ludlow, director of mechanical engineering, at Midé, a manufacturer of piezo-electric energy harvesters and electronics. "There's no silver-bullet product we can make to sell to 1000 customers. Vibration-type energy harvesters for devices on the the Golden Gate Bridge in San Francisco won't be the same as those for the Zakim Bridge in Boston because they have different vibration characteristics. We couldn't make a single product that would work on every bridge."

Don't feel discouraged by the challenges of harnessing "free" energy. Companies have created interesting and innovative harvesting technologies, power-management ICs, energy-

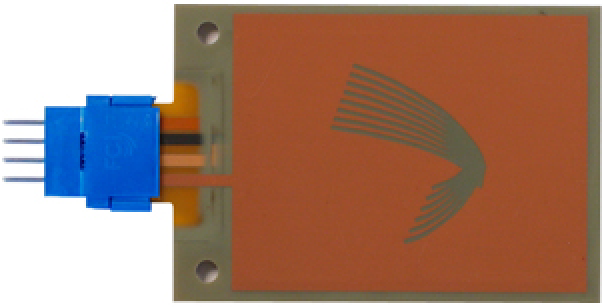

Figure 2. This image from Midé shows a "raw" Volture energy harvester engineers can use in a prototype or in an end product to convert vibrations into electrical energy. In this form, the harvester does not include any power-management circuits or energy-storage capability.

Titus-Figure 2

storage devices, as well as low-power MCUs and wireless transceivers.

Midé, for example, sells a line of standard harvesters that pick up vibrational energy with a piezo-electric transducer and convert it to electric power. "Piezos are brittle materials and they generally don't do well in harsh environments," said Ludlow. "But our packaging process protects the piezos and makes them easy to connect so you can put them into a product and have a reliable energy source."

"Many industries employ AC motors that create vibrations between about 50 Hz and 240 Hz, so you could use a standard piezo harvester on a motor to gather energy," noted Ludlow. "Then your device could report motor conditions to other equipment. Our main focus is on helicopters or rotary-wing aircraft where the vibration frequency range goes from about 10 Hz to 40 Hz. We also create under-train railway devices that operate in the 100-Hz range."

As Ludlow noted earlier, no one vibration harvester fills all needs. To help engineers understand vibration frequencies available on site, Midé created the SlamStick, a rechargeable USB-stick data logger that measures acceleration in all three axes. Configuration software lets users tailor the device to their specific needs. "Within 60 seconds we can characterize the vibration profile of a pump, for example, to find the best frequency for energy harvesting," said Ludlow. "That's the first type of analysis engineers should do for a vibration energy-harvesting application."

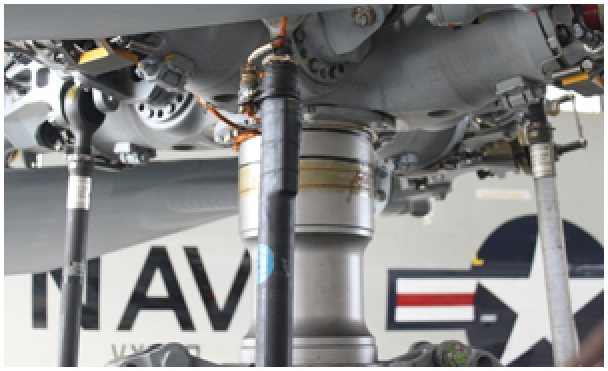

Figure 3. This photo shows a MicroStrain energy-harvesting pitch-link node installed on an Sikorsky MH-60S used for flight testing. The US Navy installed a connector on the pitch link to let people collect slip-ring data from a separate, hard-wired strain gauge as they also collected data from a MicroStrain wireless pitch-link load sensor.

Titus-Figure 3

MicroStrain also produces piezo-electric energy-harvesting devices for use in helicopters. "We're now involved with flight tests on MH-60S Navy Sea Hawk and UH-60A Army Black Hawk helicopters and V-22 Osprey tilt-rotor aircraft," said company President, Steve Arms. "Our harvesters convert the operational strains and vibrations of a helicopter into power that lets us monitor loads on rotating components. And we transmit the loads' information in real time."

"You might be surprised at the energy available in a device where you didn't think it existed, particularly at higher frequencies," noted Arms. In helicopter gearboxes, for example, we see a small displacement but because it shakes at such high frequencies, we can get a lot of energy. Even though the gearbox doesn't feel like it moves much, it is a rich environment for energy harvesting."

MicroStrain makes a product that can help engineers explore energy harvesting. The EH-Link Energy Harvesting Wireless Node lets people use piezo-electric, inductive, electromagnetic-field, solar, thermoelectric and other harvesting devices as power sources. The module includes an embedded triaxial accelerometer, relative-humidity and temperature sensor, and a single-channel differential input for an external sensor. Communications use a standard IEEE 802.15.4 transceiver.

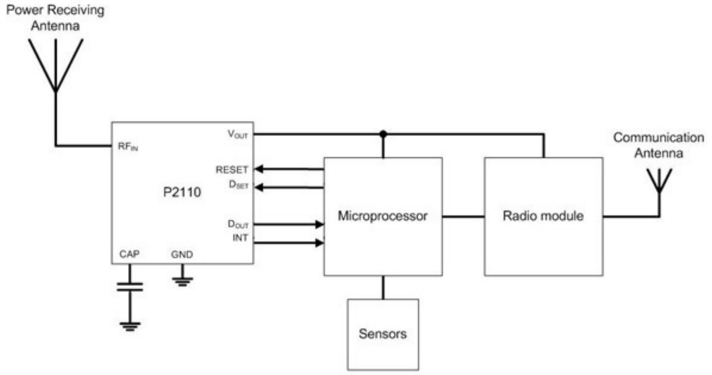

Figure 4. This block diagram shows the Powercast P2110 Powerharvester receiver IC provides RF energy harvesting and power management for battery-free, micro-power devices. A microprocessor can communicate with a P2110 and help improve system efficiency.

Titus-Figure 4

Not all energy you harvest comes from "free" vibrations, temperature differences, or sunlight. Sometimes you must pay a little extra to create the energy in the first place. Powercast takes that approach with its radio-frequency energy-harvesting devices, the P1110 and P2110 receivers that operate between 850 MHz and 950 MHz.

The wireless-harvesting technology provides controllable and reliable power that doesn't rely on intermittent sources. "You can turn on a Powercaster transmitter and provide power for as long as needed," said Harry Ostaffe, vice president of marketing and business development at Powercast. "Once charged, the receiver nodes can transmit their data or send a message to say 'Turn off the power, we’re fully charged.'" The TX91501 transmitter produces a 3-Watt output effective isotropic radiated power (EIRP) at 915 MHz and the companion receivers can pick up that energy from 40 to 50 feet (12 to 15 m) away.

"Along with power, the transmitter sends fixed data such as an identification code," noted Ostaffe. "And an MCU at the receiver could interpret that data for device authentication and location tracking. The data could also provide timestamps for remote devices that lack a real-time clock."

"Adopting the Microchip MiWi protocol let us use less power for each data transmission so the sensor node can operate more frequently while still maintaining its energy balance," said Ostaffe. "That's a benefit for a control application such as an HVAC system. And because RF power is available at all times and the receiver needs no battery, people can place sensors between walls and in toxic or dirty areas where there is no wind or light, or that would be inaccessible or dangerous for battery replacement."

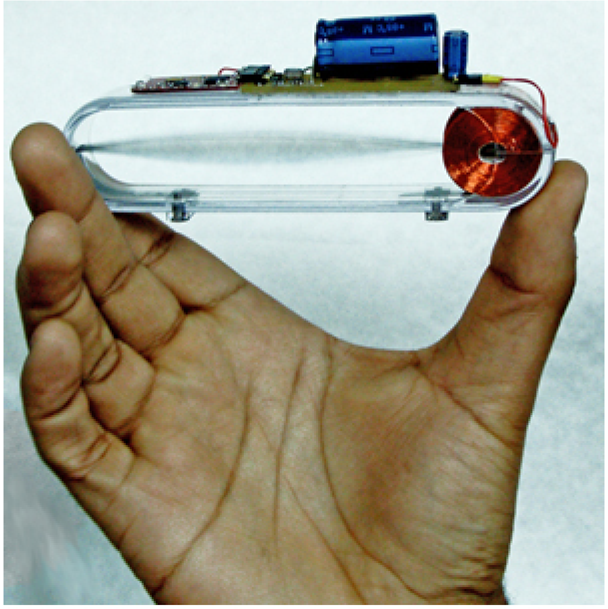

Figure 5. The tensioned membrane in this prototype from Humdinger Wind Energy moves a magnetic field through coils to create energy from wind at speeds of 6 meters/sec. or greater.

Titus-Figure 5

Some energy-harvesting devices show promise but haven't yet found a large market. Look, for example, at the approach taken by Humdinger Wind Energy. The company invented the Windbelt technology that harnesses wind energy in devices that can range from palm-size up to several meters across.

"We use a tensioned membrane that has magnets mounted on it," said Shawn Frayne, Humdinger's president and co-founder. "As air flows across the membrane it undergoes aeroelastic flutter. It's the same effect that ripped apart the Tacoma Narrows Bridge in 1940. When the membrane reaches a limit--a stable positive-feedback loop between competing lift and tension forces in the structure--we can extract power from a coil as the magnets change the field through it. With a wind speed of six meters per second (about 13 or 14 miles per hour), we can obtain from a milliwatt up to several milliwatts in a pack-of-gum sized system.

According to Frayne, Humdinger makes custom units for people who want to explore energy harvesting on different scales; however the technology is still very much "in the lab." But Frayne sees many potential uses wherever people have enough fluid flow they can tap for energy.

From the business direction, Frayne seeks to work with partners who can create complete products. "I don't see a situation where someone buys a wireless sensor node from Company A and then says, 'Let's power this with a windbelt.' They probably want to buy a little block that has all the wireless sensors and includes a windbelt."

"There's always interest in energy harvesting," said Frayne. "Everyone's just waiting for the market to blossom."

For more information:

Cymbet: www.cymbet.com.

Humdinger: www.humdingerwind.com.

MicroStrain: www.MicroStrain.com/eh-link.aspx.

Midé: www.mide.com/products/volture/volture_catalog.php.

Powercast: www.powercastco.com.

Microchip: www.microchip.com/stellent/idcplg?IdcService=SS_GET_PAGE&nodeId=2042&param=en551270.

Texas Instruments: www.ti.com/ww/en/apps/energy-harvesting/index.shtml.

Figure captions

Figure 1. A Cymbet EnerChip placed on a penny shows the relative size of the energy-storage and -management circuits available now for energy-harvesting devices.

///Images in file: 02ECN-Titus ES EH Fig 1.jpg///

Figure 2. This block diagram shows the Powercast P2110 Powerharvester receiver IC provides RF energy harvesting and power management for battery-free, micro-power devices. A microprocessor can communicate with a P2110 and help improve system efficiency.

Figure 3. This Texas Instruments MSP430 Solar Energy Harvesting Development Tool (EZ430-RF2500-SEH) uses a high-efficiency solar panel (2.25-by-2.25 in; 4.8-by-4.8 cm) optimized for indoor low-intensity fluorescent lights. The wireless-sensor system manages and stores energy in a pair of thin-film rechargeable EnerChips from Cymbet.

Figure 4. The tensioned membrane in this prototype from Humdinger Wind Energy moves a magnetic field through coils to create energy from wind at speeds of 6 meters/sec. or greater.

Figure 5. This photo shows a MicroStrain energy-harvesting pitch-link node installed on an Sikorsky MH-60S used for flight testing. The US Navy installed a connector on the pitch link to let people collect slip-ring data from a separate, hard-wired strain gauge as they also collected data from a MicroStrain wireless pitch-link load sensor.

Figure 6. The SlamStick portable and rechargeable vibration recorder from Midé measures acceleration in all three axes. The device uses a USB port to download data and to configure and charge the "stick." The vibration data helps equipment designers characterizes a vibration profile for energy-harvesting devices.

Figure 7. This image from Midé shows a "raw" Volture energy harvester engineers can use in a prototype or in an end product to convert vibrations into electrical energy. In this form, the harvester does not include any power-management circuits or energy-storage capability.

RELEVANT PRODUCT

EXPERIMENTAL
DEPARTMENT
ELECTRIC
X-RAYS

Experimental Crystal Set Receivers

Hints, Tips & Secrets

An anthology of articles extracted from 1911
and 1912 issues of Modern Electrics Magazine

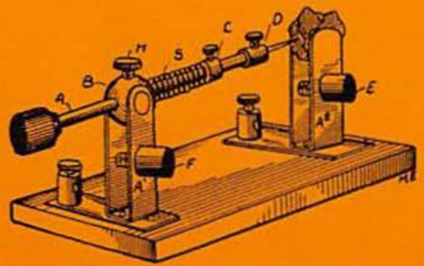
A
B
H
S
C
D
E
F
A¹
A²

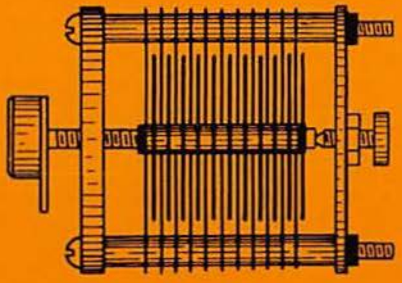

reprinted by Lindsay Publications Inc

EXPERIMENTAL
DEPARTMENT
ELECTRIC
X-RAYS

Experimental Crystal Set Receivers

Hints, Tips & Secrets
submitted to the publisher of

MODERN
ELECTRICS

Originally published by
Hugo Gernsback
New York

original articles appeared
in 1911 & 1912 issues of
Modern Electrics Magazine

Reprinted by
Lindsay Publications Inc
Bradley IL 60915

Design © 2005

ISBN 1-55918-327-6

1 2 3 4 5

A GALENA DETECTOR

Herewith is a photo of my detector. It works well, and needs no descriptive matter to work by.

The whole surface of the Crystal "Galena" can be easily searched, by gently lifting the spring with a match and letting it down again.

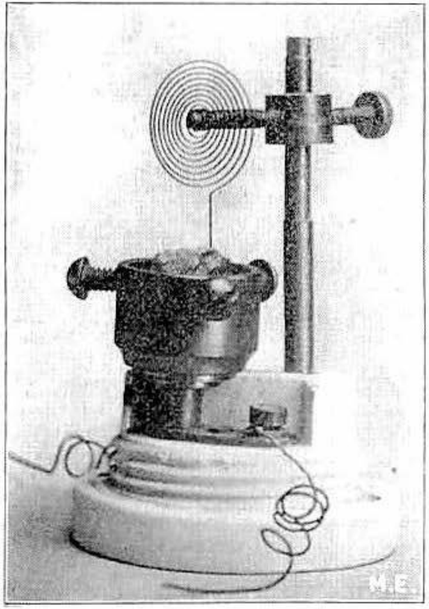

Modern Electrics has a good circulation in this colony, and is very popular.

Contributed by

W. F. HALL, Australia.

COMBINATION GALENA DETECTOR.

This detector is used in case a quick change is necessary. If one of the points of this detector happens to be knocked out of adjustment by static or other means the other detector can be easily thrown in the circuit by making use of the two point switch. The arm of this switch is connected to one side of the circuit while the points are connected to B1 and B2. The base of this detector, or A1, A2, A3, are all connected together and led to the other side of the circuit. You will find that this detector can be adjusted to the highest degree of sensitiveness, and is easy to construct, its dimensions being made according to your liking; but I would suggest, for the uprights, ½ inch square brass rod, each 1 inch high. The distance between A1 and A2 1¾ inches, and 1 inch between A1 and B1. The springs between X and A1 will hold the mineral in tight adjustment, while the check nut in A2 holds the other rod from slipping.

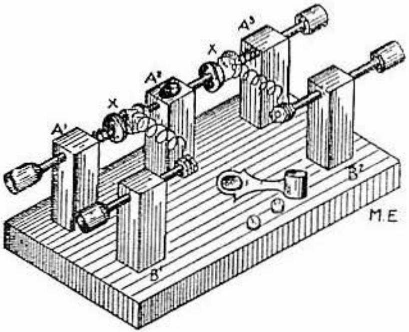

Contributed by

JESSE SAUTER.

CONTACT FOR LOOSE COUPLER

While building a loose coupler of the style described in "How to Make Wireless Instruments," the idea of placing two springs on my secondary to make contact with the rods did not seem very good to me, so I bought two brass curtain rods, 3/16 inches diameter, of the kind that has a small tube sliding into a slightly larger one. I used the small tubes for the rods on which the secondary slides, and after cutting two pieces of the larger tube the length of my secondary, I soldered one end of the wire wound on the secondary to one large tube and the wire from the switch to the other tube. I then placed the tubes in holes bored one inch above and below the center of the end pieces of the secondary. I find that it is possible to obtain a better contact by this means than by the springs, because the whole weight of the secondary lies upon the brass tubes, and I also find that the secondary slides much easier this way.

Contributed by

T. E. JONES.

A TUNING COIL SLIDER.

The rod is made of two No. 14 brass wires, D.

The slider is made of a brass block, B, three-eighths inch square and one-eighth inch thick. Two holes are drilled one-sixteenth inch from the edge, F, so that the wires will just fit the holes.

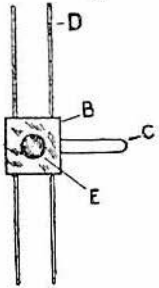

The spring, C, is to make contact with the tuning coil wire, and is fastened to the block, B, by means of the battery binding post, G, and the hard rubber knob, E.

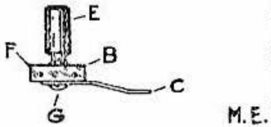

The brass wires are fastened to the tuning coil ends by means of brass screws.

Contributed by
BERTHEL CARLSON.

AN ADJUSTABLE CONDENSER.

The condenser consists of 10,000 square inches of tin foil and wax paper, made up in seven sections as per sketch; 14 pieces of brass, 1 inch by 1 inch by 3-16 inch, and 2 pieces ½ inch by 3-16 inch by 7¾ inches; 14 binding posts; and 12 plugs to fit holes as per sketch. The condenser is made up in the regular way, first a sheet of wax paper, then a sheet of tin foil, until seven condensers have been made. It will be seen by the sketch, which is very plain, that there is practically no limit to the different capacities that can be obtained. This condenser can be used for primary circuits or tuning, and is a very useful instrument around the laboratory for experimenting purposes. The tinfoil sheets can be cut any size to suit the maker. It is understood that this condenser cannot be used for secondary purposes.

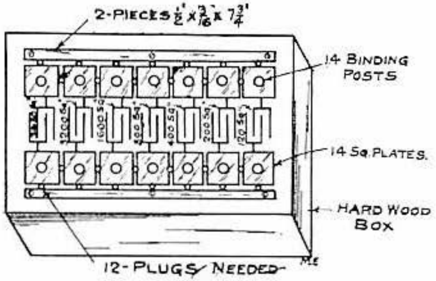
2-PIECES ½" x 3/16" x 7¾"
14 BINDING POSTS
1200 Sq
1600 Sq
400 Sq
400 Sq
200 Sq
120 Sq
14 Sq. PLATES.
HARD WOOD BOX
12-PLUGS NEEDED
ME

Contributed by JAMES G. FITCHETT.

HOW TO MAKE AN EXHAUSTED COHERER.

Procure 44 inches of glass tubing such as is used in a chemical laboratory, about one-quarter of an inch in diameter. Take a dime and cut from it two disks, B, about one-quarter of an inch in diameter. Next get two inches of copper or brass rod one-quarter of an inch in diameter and cut it into two pieces each one inch long. Solder the silver disks to one end of each of these rods, E, neatly. Next get two pieces of No. 22 platinum wire, each one and a half inches in length; drill a small hole in the other end of the rods and solder the wire in these as in E.

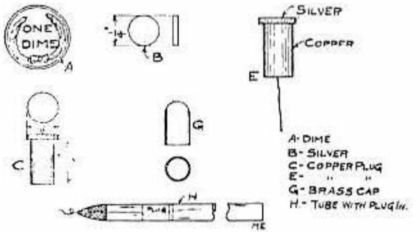
ONE DIME
A
¼
B
SILVER
COPPER
E
C
G
H
PLUG
ME
A- DIME
B- SILVER
C- COPPER PLUG
E- " "
G- BRASS CAP
H.- TUBE WITH PLUG IN.

The next is a delicate job. Both of these plugs are filed to fit the tube

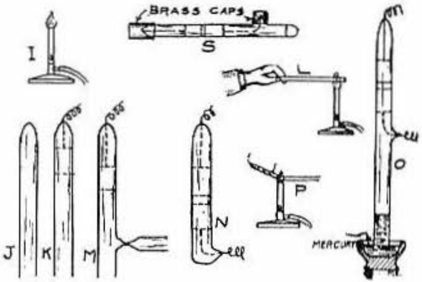

snugly and also must be airtight. Extreme care should be taken not to break the tube, as it must be over forty inches long. Next seal the plugs in the tube. You will need a small bowl, some mercury, and burner, also a helper. Hold one end of the tube over the flame until it is red-hot as in L. Then seize with the fingers and draw it out to a point; break this off about one-quarter of an inch from the long end and the end will look like J. Slip one plug through the open end and draw the platinum wire through the small end; hold this over the flame, and the glass will adhere to the platinum. Measure about four inches from the sealed end and hold it in the flame; also hold another piece of tubing in the flame at the same time; when both are hot touch the small piece to the long piece at the four-inch mark and draw out a small tip, as shown at M. Break off the short piece and leave the tip. Place the filings in the tube, made of 94 per cent. nickel and 6 per cent. silver. These can be obtained from coins, using a clean new file. Slip the other plug into position and pull the platinum wire through the tip, as shown at N. Space the plugs one-sixteenth of an inch apart, and seal off this tip as was the first. The tube is now filled with mercury, using a funnel, tap the tube so as to expel all the air bubbles. Fill this tube completely and place your thumb on the end. Invert, and place the open end under the surface of the mercury, then release your thumb. Then with the aid of a helper seal off the tube above the mercury level, as the mercury will drop to the thirty-inch level as in P. Place brass caps over the sealed ends, making contact with the platinum wire. The instrument is now ready for use.

Contributed by

FANNON BEAUCHAMP.

LEYDEN JAR CONDENSER.

The materials needed for the construction of this condenser are: two wire spools, one about six inches in diameter, the other three inches, the larger being four inches high, the smaller three inches in height; six test tubes, with corks to fit; enough tinfoil to cover both outer and inner sides of the tubes; two binding posts, and four or five feet of green flexible wire.

To begin construction, smaller spool a is glued to bottom of larger spool b, to form a stand. Then six ¾-inch holes H are bored straight through both butts. Care must be taken to get these holes spaced evenly apart.

The test tubes, after being covered with the tinfoil by gluing, are slipped into the holes in the stand forming the Leyden jars. The corks have a hole c bored through their middle, in which is inserted a wire S, connecting the inner covering of tinfoil to binding post d. Then wires are wrapped tightly around the outside covering of the tubes w, connecting to binding post e.

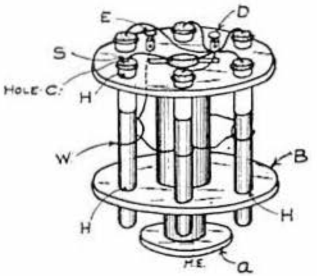

After varnishing, the device is ready for work.

Contributed by

JOHN B. BRADY.

A PORTABLE RECEIVING OUTFIT.

Below is described a portable wireless outfit smaller than a small camera, in which sensitiveness is not sacrificed to saving of space.

It consists of a variometer, a condenser contained within the inner coil, and a detector of any type preferred by the maker.

The variometer consists of a cardboard cylinder, 1⅞x3½ inches in diameter, with the inner cylinder 3 inches outside diameter. Both cylinders are wound with an equal amount of No. 24 B. & S. gauge enameled wire. In placing the inner coil in position, a brass piece as shown in Fig. 1 may be found convenient to screw into the brass rod, if the outer coil is not sufficiently flexible to allow the rod to go through both holes.

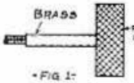
BRASS
RUBBER KNOB

-FIG 1-

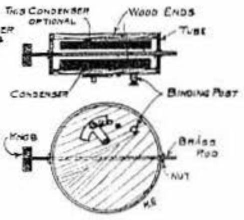
THIS CONDENSER OPTIONAL
WOOD ENDS
TUBE
CONDENSER
BINDING POST
KNOB
BRASS ROD
NUT
H.E.

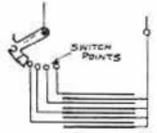
SWITCH POINTS
CONDENSER

-FIG 2-

The condenser is built up of a number of sheets of tinfoil and paraffined paper or some similar dielectric, as shown. The leads are made to a switch as shown in Fig. 2, after which the whole condenser is embedded in paraffin. Two condensers would be an improvement.

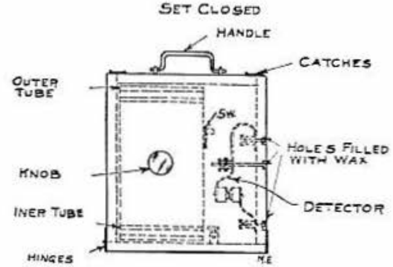
SET CLOSED
HANDLE
CATCHES
OUTER TUBE
SW.
HOLES FILLED WITH WAX
KNOB
DETECTOR
INER TUBE
HINGES
H.E.

The mineral detector, either perikon or silicon, is used. The type shown in the drawing is very satisfactory.

The details of the construction of the box containing the instruments are left to the maker. The over-all dimensions should be at the most a trifle more than 4¼x4¼x3¼ inches. Black stain and a coat of varnish will give a finished appearance.

-SET OPEN-

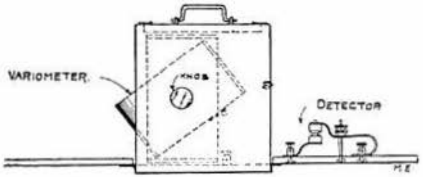
VARIOMETER
KNOB
DETECTOR
H.E.

Contributed by

LEWIS C. MUMFORD.

ROTARY LOOSE COUPLER.

The following gives data for construction of a loose coupler. The primary is made of wood, and the secondary of hose.

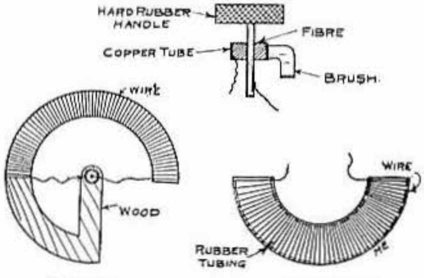
HARD RUBBER HANDLE
FIBRE
COPPER TUBE
BRUSH.
WIRE
WOOD
WIRE
RUBBER TUBING
H.E.

-PRIMARY-

-SECONDARY-

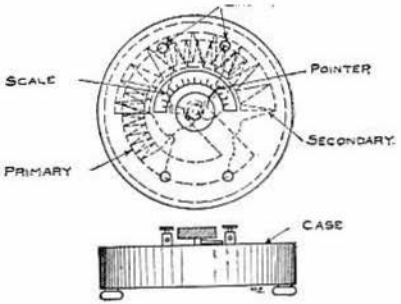
SCALE
POINTER
SECONDARY.
PRIMARY
CASE

I think the diagram explains all details.

Contributed by

WM. T. REBENHAULT.

A ROTARY TUNING COIL.

The following is a description of my double slide, rotary, and very accurate tuner. The circle and base are of 1 inch stock, the circle being wound with No. 24 enameled wire. The novel part is that each slider works independently, and that they are insulated from one another. It is connected into the circuit the same as any other double slide tuner. Drawing is self-explanatory.

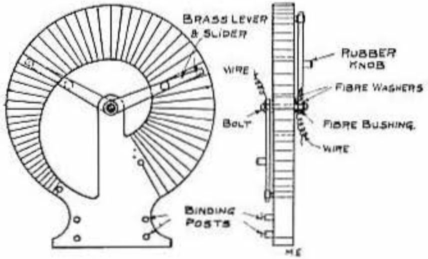
BRASS LEVER & SLIDER
RUBBER KNOB
WIRE
FIBRE WASHERS
BOLT
FIBRE BUSHING.
WIRE
BINDING POSTS
H.E.

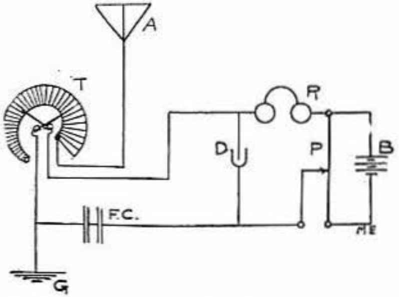
A
T
R
D
P
B
F.C.
G
H.E.

Contributed by

C. J. SEDLAK.

TUNING TRANSFORMER.

As anybody knows who has tried it, the making of a transformer for tuning is a difficult job, and also a very clumsy one. Another objection is the expense of material. A very efficient substitute can be made as follows: Cut out a disk of wood one inch thick and about four inches in diameter. Then cut a groove on the circumference about half an inch wide, and a quarter inch deep, as per illustration. Then wind about a hundred turns of No. 36 insulated copper wire on this disk, leaving the ends for attaching to binding posts. Over this wind twenty turns of No. 22 copper wire in the same direction, having insulated it well from the No. 36 wire, also leaving leads for two binding posts. Then on positions indicated in diagram, screw

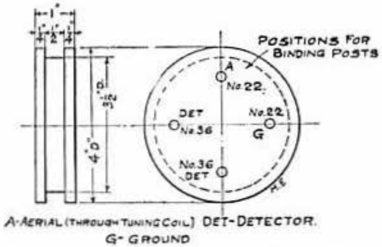
1"
1/4
1/2
1/4
POSITIONS FOR BINDING POSTS
A
No. 22.
DET No. 36
No. 22 G
No. 36 DET
4.0"
H.E.
A-AERIAL (THROUGH TUNING COIL) DET-DETECTOR.
G-GROUND

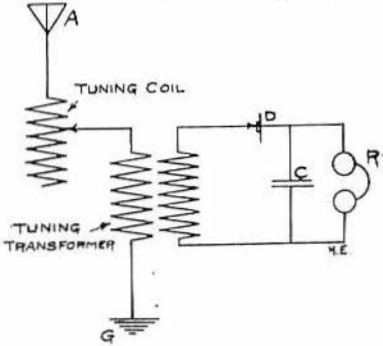
A
TUNING COIL
D
C
R
TUNING TRANSFORMER
H.E.
G

in binding posts, connecting the ends of the wires, one to each binding post, and mark them. Here we now have a condensed form of tuning transformer, with a very short wave length. A tuning coil is now connected in series with this instrument, and we have an arrangement which tunes sharper and with much less trouble than the ordinary tuning transformer, and costing a great deal less, as almost every amateur possesses a tuning coil. The No. 22 wire is connected in the aerial circuit, and the No. 36 in the receiver circuit. I have found this to work as good as a regular tuning transformer.

Contributed by

WALLACE ELLS.

ADJUSTABLE CONDENSER—FOR SENDING ONLY.

First, select six or eight thin table glasses, free from bubbles and cracks. It is best to wax the tops of each glass to prevent leakage. (See Fig. 1.)

Second, get a metal pan which will hold all of the glasses, and which is of the same height as the glasses.

The pan should be mounted on a base, provided with two binding posts. To one of the binding posts a wire is fastened, which, in turn, is soldered to the pan.

After the pan is fastened to the base, two uprights are provided, which are fastened to the base, about four inches from the pan. Between the uprights, a metal rod is suspended, from which wires are dropped down into each glass; and which is connected to the other binding post. (See Fig. 2.)

The glasses are now filled with a solution of salt and water.

The capacity of this condenser is varied according to the amount of solution in the pan; or if a large capacity is required, the pan is filled with this solution up to the top. If a low capacity is required, the solution in the pan may be but a few inches. Thus the capacity of the condenser is varied according to the amount of solution in the pan.

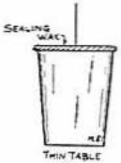

-FIG. 1-

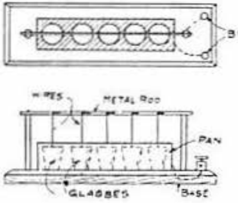

Contributed by

JAY JAKOWSKY.

HOW TO MAKE A SENSITIVE GALVANOMETER.

Finding the need of a good galvanometer, and one that would be very sensitive, without paying a high price for one, I succeeded in working out the following:

Make a base 4x4x⅝ inches thick, of mahogany or walnut and bevel the edges as shown in Fig. 1. Make two standards or sides, 2¼x3¼x½ inches. Fasten these to the base, with some good glue, 2¼ inches apart (inside measurements).

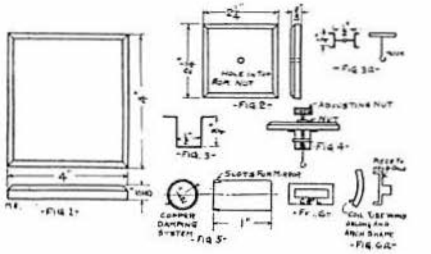

Now make a top of the same material as the base, size 2¾x2¼x⅝ inches, and bevel the top edges of this. Drill a hole in the center, large enough to admit a battery nut, see Fig. 2. Shape a piece of copper or brass as shown in Fig. 3. Make a cross piece to fit this, Fig. 3a. Fasten the support to the under side of the top, Fig. 4. This part of the apparatus works the same as the adjustment of the E. I. Co.'s electrolytic detector. Examine Fig. 4 carefully and all will be clear. Make a thumb screw as described in November, 1910, issue of *Modern Electrics,* page 452. The top may now be put in place with glue, and is shown complete in Fig. 4.

Next we must make a copper cylinder one inch long by one-half inch diameter. Prick or punch two small holes one-quarter of an inch from the top, and diametrically opposite; also two one-sixteenth of an inch from the bottom. These are for the magnetic needles. Cut two slats in the top one-eighth of an inch deep and (wide enough to receive a small mirror, size ¾x½ inch), diametrically opposite.

Magnetize two needles or small bits of steel about one to three-quarters of an inch long. Place one of these in the top holes of the copper cylinder with its north pole to the left. Now place the other needle in the bottom holes, with the north pole to the right. This is to prevent the earth's magnetic field from affecting the readings, and is called the astatic system. The needles may now be fastened with a drop of glue or shellac. Now cut a mirror, ¾x½ inches, and fasten in the slots (on the top of the cylinder), with glue. Fasten a small silk thread to the center of the mirror with a drop of glue, and the other end of the thread must be fastened to the small hook on the plunger, Fig. 7. The spring on the bottom of Fig. 7 is made of No. 40 copper wire which has been annealed, one end being fastened to the copper cylinder and the other end to the base with a small tack.

Now make two coils, each two inches long by one inch wide by three-quarters of an inch thick. Wind them in the form of an arc as described by Fig. 6. They should be wound on pasteboard forms, and wound with fifty feet each of No. 36 wire. The resistance will then be approximately 100 ohms. They should be fastened to the sides of the box, with two pieces of wood as shown in 6a, and connected in series. The other ends are brought on the outside, to binding posts; see Fig. 7.

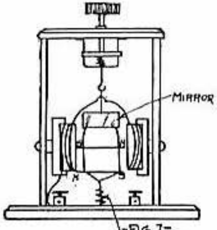

-FIG. 7-

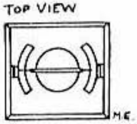

-FIG. 8-

Now make a back and front from a piece of plain glass and fasten the front so that it will be removable. Fasten a dial on two pieces of wire so that it will be adjustable and finish instrument with several coats of shellac.

Fig. 7 shows a side view of the instrument, and Fig. 8 shows a view from the top.

Contributed by

FRANK L. BUNKER.

HOW TO MAKE A GALVANOSCOPE.

A galvanoscope for detecting small currents of electricity can be made from a coil of wire, A; a glass tube, B, full of water; a core, C; and a base, D; with binding posts as shown. The core, C, which is made of iron and cork is a trifle lighter than the water it displaces, and will therefore remain in the tube normally; but as soon as a current of electricity passes through the coil, the core is drawn down out of sight. The current required is very small as the core is so nearly balanced that the least attraction will cause it to sink. The glass tube may be a test-tube as shown in Fig. 2, or any empty

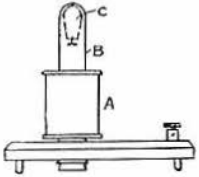

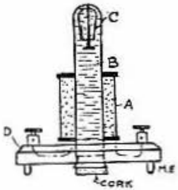

developer tube. If one has neither a test-tube nor a developer tube, an empty pill bottle may be used. The washers at the end of the tube may be fiber, hard rubber, or wood. The base is made of wood, or any other insulating material, and should have four legs on bottom. Make the coil of about No. 18 wire, and connect ends to binding posts as shown at Fig. 2. The core is made by pushing a small nail through a piece of cork. It should be made so as to rise slowly when placed under water. Some filing may be necessary to get the weight just right, but it should be remembered that the buoyancy of the core can be adjusted after the parts are assembled by pressing the cork in the bottom of the test-tube. This causes compression in the water space and especially of the upper cork, reducing its displacement and causing it to sink. The lower cork is then slowly withdrawn, by twisting until the core slowly rises. The instrument is then ready for use.

Connect the binding posts to a single cell of battery—any kind will do, as a slight current will answer. On completing the circuit the core will descend; or put in a switch or push button on one of the battery wires. If the button be concealed where the operator can reach it, the core will obey his command to rise or fall, according to his control of the current. This is a mysterious-looking instrument, the core being moved without visible connection to any other part. If this instrument is connected up as per directions, one will have a very nice looking galvanoscope.

Contributed by

LYLE BRITT.

PLATE GLASS CONDENSER FOR RECEIVER.

Procure a thin glass plate 20 inches long and fasten securely between suitable standards and with corresponding base board, glass being covered with tinfoil to within one-quarter inch of edge and suitable leads

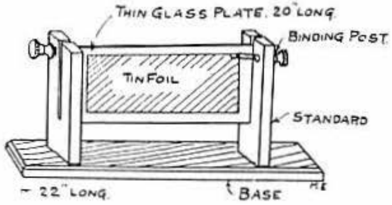

to two binding posts. This condenser has been used to great advantage by maker and can be used in many different hook-ups.

Contributed by

ETHAN CLARKE.

A SIMPLE LOOSE COUPLER FRAME.

As a great many experimenters wish to build their apparatus as simply, cheaply, and with as few parts as possible, I will endeavor to outline a coupler that is simply a "fool killer."

We will start with the base: Procure a well seasoned cherry board 8 inches wide, 20 inches long, and three-quarter inch thick, and cut a slot one inch wide and 16 inches long in the middle as in Fig. 2 A. File this slot perfectly smooth and get two strips of aluminum or brass, 15 in long, one-half inch wide, and one-eighth inch thick as at B. Drill five holes in each and countersink for No. 6 screws. Fasten these strips, one on each side of base and about two inches from slot, and one-half inch from back end of board. After this is done, make two pieces of wood 8x½x1 inches and screw them on the bottom of base at each end as at B, B, (Fig. 2 B). These will keep base from splitting or warping.

Make two pieces 7x7 inches and one-half inch thick, as at a Fig. 3; center them, and cut a hole five inches in diameter through each (or the diameter of your primary tube). Glue one of these boards on each end of tube, and wind on your wire.

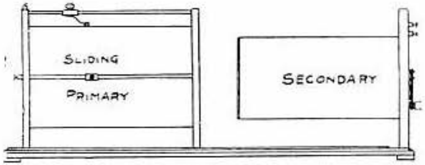

-FIG 1-

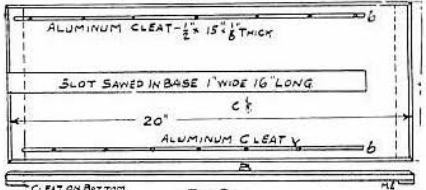
ALUMINUM CLEAT-½"x 15"·⅛" THICK
b
SLOT SAWED IN BASE 1" WIDE 16" LONG
c
20"
ALUMINUM CLEAT X
b
B
CLEAT ON BOTTOM
HL

-FIG 2-

After your winding is completed, fasten a cleat on bottom of primary heads, one inch wide, three-quarter inch deep, and eight inches long; this is to slide in groove of base.

For your secondary, make a board 7x7x½ inches, and then cut out a piece four inches in diameter, (or to fit inside secondary tube), and fasten with a screw and some glue in center of head board. Now glue the secondary tube to the head board and make windings, taking off taps so as to use a switch on head of secondary. Fasten the secondary head with screws one-half inch from end of base, X. The instrument may be finished to taste.

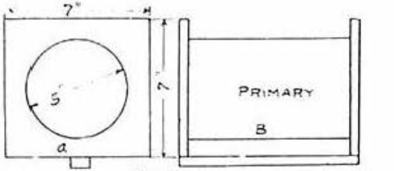

-FIG 3-

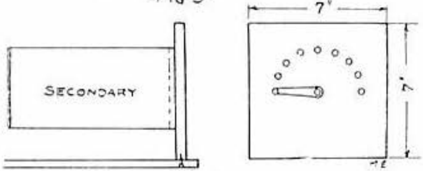

-FIG 4-

I think that Fig. 1 will make all other points clear.

Contributed by

F. W. DOTY.

A LOOSE COUPLER.

First procure a piece of maple four feet long and seven inches wide. Work this until it is 7½ inches wide and three-quarter inch thick. From this piece cut one fourteen inches long. This is to be the base. Plane a three-eighth inch bevel on this piece. Now cut two pieces; one to be 6½ inches square, the other to be a little less than 6½ inches on two edges and 6½ inches on the other two sides. Now have a piece of white pine turned on a lathe until it is five inches long and five inches in diameter. Hollow this out for a distance of four inches and leave a rim one-quarter inch thick around it. Turn out another piece of pine four inches in diameter and five inches long. Glue these pieces to the ends which are 6½ inches and 6½ by 6 7-16 inches respectively. Now get a steel rod 11¼ inches long and pass it through both cylinders until it reaches the end of the large cylinder. Now bore a one-quarter inch hole in the small piece "a" until it is one-half inch

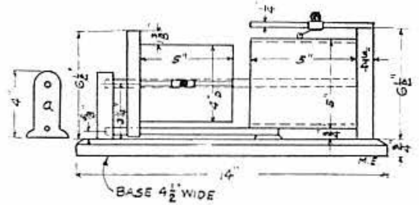

deep and 3¼ inches from the base. Fit the large rod into this. Now take a smaller piece of rod six inches long and bore a hole in the end of the small cylinder, five-eighth inch from the base. Bore another hole in the end "a," one-half inch deep. This rod is to keep the end of the small cylinder from sliding and rubbing the base.

Now screw the piece "a" to the base and assemble the parts. Then screw the end of the large cylinder also to the base. Now insert a small piece of wood under the smaller rod to hold it in place. Wind the large cylinder with No. 22 enameled wire and the small cylinder with No. 28 enameled wire. If a standard E. I. Co.'s slider and rod is used, the secondary slider and rod will clear the primary coil and allow adjustment while the secondary coil is in the primary. Procure two binding posts for both primary and secondary; each one being connected to a slider and the end of the wire which is nearest to the end of the core, respectively on both coils.

This instrument, when given a mahogany stain and polished, not only makes a useful piece of apparatus, but an ornament to any wireless experimenter's table.

Contributed by

WALTER LEAN.

MINERAL CLASP FOR DETECTORS.

The accompanying sketch shows a clasp to be used in connection with a wireless detector mineral cup.

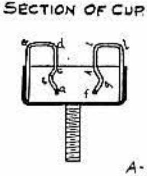

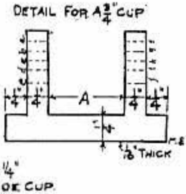

Contributed by

HAROLD W. THOMPSON.

MULTIPLE DETECTOR STAND.

While looking over June, 1910, issue of *Modern Electrics* I saw a detector designed by E. E. Ely.

I started to make one, but changed it so I could use it for more than one stand, by putting two carbon cups on two brass springs, an equal distance from the center bolt L (as in diagram). By this means you can bring either silicon, carborundum, tantalum or electrolytic into use without the aid of a detector switch.

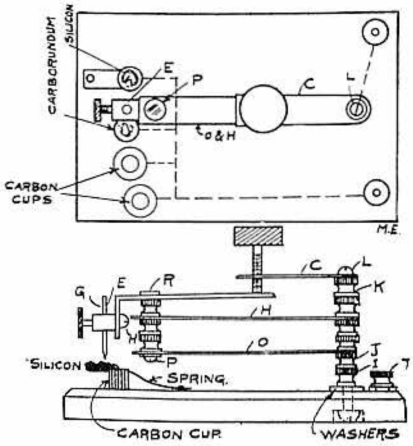

This saves making a switch and more than one stand.

RECEIVING CONDENSER.

A cheap, yet neat and efficient, receiving condenser, can be made from the following:

Two pieces of wood (preferably hardwood), 1¾x1¾x½ inches (Fig.

2); nickel-plated shaving-stick box, 1⅜ inch diameter, (Fig. 3); two screw-bottom binding posts, (Fig. 5); section of broomstick 2½ inches long; five sheets of tinfoil, 1¾x12 inches between six sheets of waxed paper that are free from flaws, 2¼x13½ inches.

Cut out the wooden ends, centre them, and after having drilled a hole one-eighth inch diameter clear through and rounded the two upper corners, draw a circle with a diameter of 1⅜ inches on the inside of each end, and with a bit or a gauge cut away the wood inside the circle half way through each of the ends; then give them a couple of coats of shellac or varnish (Fig. 2).

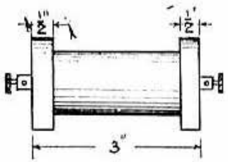

-FIG. 1-

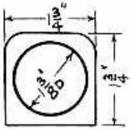

-FIG 2-

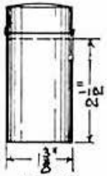

-FIG. 3-

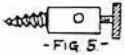
-FIG 5-

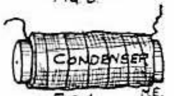

-FIG. 4

Cut the bottom out of the shaving stick box, and make another cut 2½ inches above, leaving a nickel-plated tube 2½ inches long; then from a broomstick cut a section 2½ inches long; center the ends and start a hole with a gimlet.

Heat the waxed papers containing the tinfoil and then wrap around the wooden core, securing with a few wraps of cotton, being sure that a terminal of the condenser is free at each end (Fig. 4). Insert the condenser into the nickel-plated tube, place the terminal that protrudes from each end into the small holes in the ends of the wooden core; then fit the wooden ends on the tube, and through the holes bored through these ends insert the binding posts and screw into the wooden core to hold the wooden ends in place and make contact with the condenser terminals. (Fig. 1).

Contributed by

HOLLY STARTZMAN.

A SILICON DETECTOR.

Silicon is generally a very hard substance to get into adjustment, and once there, is easy to get out of adjustment again. The following will give an idea of my own, how to get rid of both of these troublesome faults of Silicon. This will make an efficient detector at a very low cost.

For the base, procure a piece of fibre 2½x2x⅜ inches. Polish it with a file and sand paper and bevel the edges, just for appearance's sake. Then drill two holes, 1½ inches apart the long way of the base and in the center of the distance across.

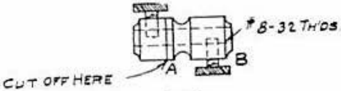

-FIG. 1-

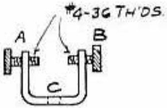

-FIG. 2-

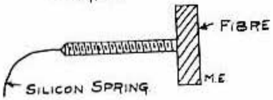

-FIG. 3-

Now get an old double connector (Fig. 1) and cut off about one-third of it at A (Fig. 1). Tap the hole in the center for 8-32 thread. Of course, one of the holes at the end, where the set screw was, (B Fig. 1), should be left.

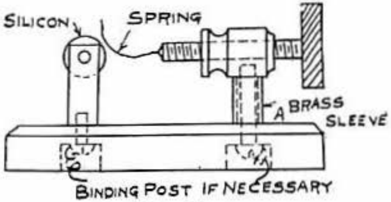

-FIG. 4-

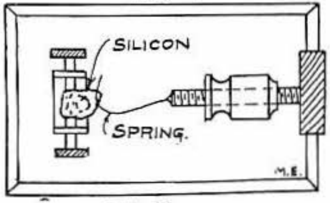

-FIG. 5-

Now take a brass sleeve that fits the 8-32 thread snugly and place it over an 8-32 screw which passes up through one of the holes in the base (A, Fig. 4). This is to hold part shown in Fig. 1, one-half inch off the base.

For the holder for the silicon, take a piece of brass and bend it into shape (Fig. 2) and thread it at A and B (Fig. 2) with 4-36 threads, to hold the silicon. It is held to the base by a screw through the bottom of the holder at C (Fig. 2).

Now get a brass machine screw 1½ inches long and make a suitable handle for the end of it. Then get a piece of silver or brass (silver preferred), and get a jeweller to roll it down to No. 33 and cut a piece 1-16x1½ inches and solder it on the end of the machine screw (Fig. 3). Then bend it as shown.

The silicon can be placed in the holder and this spring set up against the silicon so as to form a medium contact with it and thus get your adjustments. You should try to use a sharp or fairly sharp edge of the silicon to set your spring up against.

I have found that this beats all the kinds of detectors I have tried. This detector ranks even above the perikon or electrolytic, I think. And with it and suitable apparatus I have heard Honolulu, 2,100 miles away, as well as 1,000 miles over land every night.

Other minerals such as ferron, pyron, etc., can be used in it to good advantage.

Contributed by

CLARENCE BALLARD.

A Silicon Ticker.

ANY wireless operator living near one of the New Poulsen Wireless stations, doubtless has trouble in reading them, and perhaps he can not receive their signals at all.

If you live near such a station and have a detector known as a "Silicon Cat-whisker Detector," gently pull the steel wire off and then let it fall back on the silicon; and as it vibrates it will give forth a beautiful, clear, humming sound; like picking a guitar string. This only lasts for an instant; as long as the wire vibrates against the silicon. From this experiment I worked out the following instrument which gives the result which the Poulsen people say can only be accomplished by their instruments, i. e., the Poulsen "Ticker."

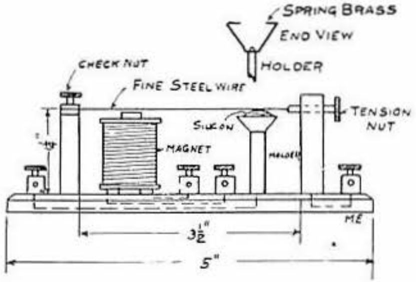

The base is made of hard wood or rubber one-half inch in thickness, five inches long and 3 inches wide.

The small bell magnet can be taken from any electric bell of convenient size; generally they are one one-eighth inches long; the dimensions on the diagram being for this size. If you have a different size the dimensions must be changed so the magnet will always be one-eighth inch away from the steel wire. This wire is the E string of a mandolin and must be steel. Make two brass standards one one-half inches long and drill holes in them one one-quarter inches from bottom, as in diagram, then thread in from the top for the check-nut on one of them, and on the side of the other for the tension nut. The tension nut can be taken from the spring of an old telegraph sounder; this is to keep the wire from breaking by twisting around.

Distance between standards is three and one-half inches. The electro-magnet is mounted exactly one-eighth inch under the wire. The holder is also mounted under the wire, so the wire will strike the top surface of the silicon. End view of holder is shown in Fig. 1.

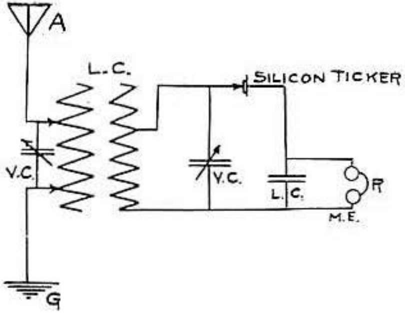

Binding posts should be screwed on the ends of the ticker. One goes to the standards and one to the silicon holder; also two in the middle for the magnet.

The magnet is connected to the alternating current supply, preferably 110 V. A. C. through a resistance which can be varied. Not more than four volts should pass through the magnet or it will heat up. A small water rheostat works very well in this connection. A small switch should be mounted near the ticker to open the circuit when you are not using it.

The ticker is connected up as in Fig. 2. The condenser which shunts the phones is three times as large as an ordinary condenser used in wireless work. The variable condenser should also be used to get good results.

Now adjust the tension nut and vary the current in the magnet, and tune very closely for what is known as the "Conversation Wave" and you will be a surprised person.

These stations send out two separate trains of waves. The arc sends out a sound which sounds like spilling milk, but you must tune for the waves which are generated in the closed oscillation circuit, arc, condenser and inductance. The station in question has an immense fan aerial composed of thirty-seven wires stretched between two 300-foot towers, 400 feet apart. They talk daily with a station in Texas.

Contributed by

STANLEY HYDE.

A SMALL VARIABLE CONDENSER.

Procure a chemist's test-tube six by seven-eighths of an inch in diameter. Coat the outside with shellac, and when nearly dry apply one turn of tinfoil. This forms one surface.

For the other surface take a wooden rod a trifle smaller than the inside of the test tube and about seven inches long. Cut a knob on one end to facilitate the sliding. Build this rod up with tinfoil until it barely slides into the test-tube.

For the clip use a piece of aluminum one-half inch wide and three and a half inches long. This will serve to hold the test-tube to the base, which is 8x2x½ inches, and also to make connection to the tinfoil.

By referring to the diagram you can see how the other surface is connected

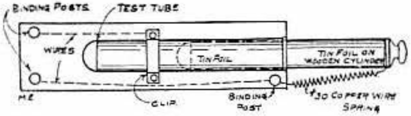

to the wires. The wire is scraped bare for eight inches and then wrapped around the tinfoil.

This makes a very efficient condenser and has a capacity of about .001 M. F.

Contributed by

N. E. DORRINGTON.

THREE SLIDE TUNER.

If one is making a three slide tuning coil and does not know how to operate it, or for some reason would like to change from a three slide tuning coil to a one or two slide, one can do so by adding the following:

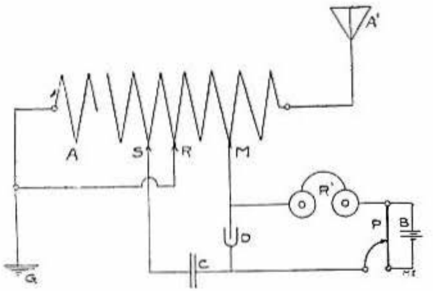
A'
A
S
R
M
R'
D
P
B
C
G
H E

When your coil is almost finished winding cut the wire about one inch from the end, then add a separate winding, A, and ground as shown in drawing.

To operate as a two slide tuning coil move slider, S, to winding, A, and then only slider, R and M, are in use.

To operate as a single slide tuning coil, move sliders, S and R, to winding, A, thereby using only slider, M.

To operate as a three slide tuning coil move sliders off from winding, A, thereby putting them all in use.

This will help one to adjust a three slide tuning coil also to experiment with single and double slide coils.

Contributed by

WILLIAM S. WILDER.

VARIABLE CONDENSER.

The accompanying sketch represents a variable condenser of high capacity, which I recently built, and which gives fine results. I used thirteen glass plates six inches by 3 inches, enclosing all in a varnished box, with switch and binding posts on top. The switch was made of an ordinary one except a circular extension of brass. The drawings I believe are self-explanatory.

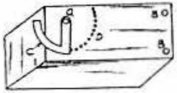

- DIAGRAM 1 -

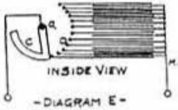
a
C
a
H E
INSIDE VIEW

- DIAGRAM E -

Contributed by

LELAND A. REINHOLD.

A VARIABLE SENDING CONDENSER.

The base is made of a dry piece of wood about 10x8x1 inches. A piece of tinfoil about 6x6inches is fastened on in the centre with mucilage, and connected to binding post. A small hole is drilled in each end, and a nut fastened over each one, threaded to take thumb screws as shown in drawing. A thin spring, about 1¼ inches long, is fastened at each corner. These serve to raise the glass and upper

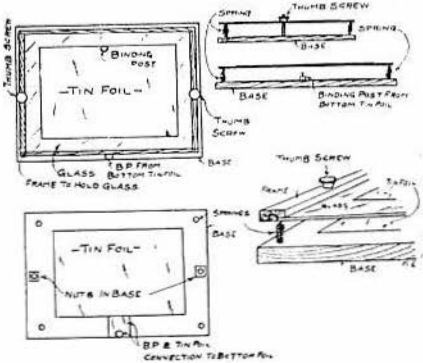
THUMB SCREW
BINDING POST
-TIN FOIL-
GLASS
B.P. FROM BOTTOM TINFOIL
FRAME TO HOLD GLASS
BASE
SPRING
THUMB SCREW
SPRING
BASE
BASE
BINDING POST FROM BOTTOM TIN FOIL
THUMB SCREW
-TIN FOIL-
NUTS IN BASE
BP & TIN FOIL CONNECTION TO BOTTOM FOIL
THUMB SCREW
FRAME
TINFOIL
GLASS
SPRINGS
BASE
BASE

tinfoil when thumb screws are loosened. A frame is made to hold the glass, with room enough at each end to put thumb screws through, which must screw into nuts in base. A piece of tinfoil 6x6 inches is fastened on upper side of glass, and a binding post fastened to it. With this condenser you can get a very fine adjustment, and I have used it with great success.

Contributed by

WILLIAM McGUIRE.

AN EASILY CONSTRUCTED VARIABLE CONDENSER.

THE following is a description of my variable condenser which was constructed at home, very cheaply and easily.

First procure twelve glass test-tubes, six of them 6x1 inch, and the other six of a size that will just slip inside these. Cover all the tubes with tinfoil, on the outside only, and make the inside ones fit very snugly, as the closer the tinfoil surfaces are together, the more efficient the condenser will be. Next get a box, eight or nine inches long and five inches wide by one and one-half inches deep. A cigar box with the lid nailed down and one side taken out will do very well.

Connect the large tubes with a piece of bare wire, making two or three turns around each one, to insure good connections. Space them about one and one-half inches apart. Leave four or five inches of wire to connect with the binding-post.

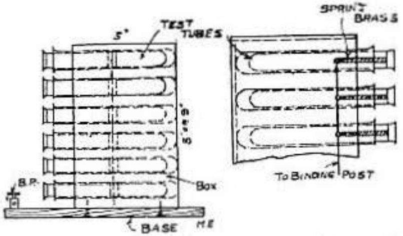

Now place the tubes in the box and run the wire through a hole in the bottom. Fill the space around them with melted paraffine and space them carefully before it hardens.

The connections to the movable tubes are made of strips of thin spring brass. They should bear pretty hard on the tubes. If the box is stained black and given a coat of shellac, it will be a very good looking instrument.

Secure this on end, to a base about 3x8 inches, and connect the binding-posts. It is then ready for use.

Vary the capacity by pushing jars in or out.

Contributed by

JAMES BITLER.

A SIMPLE ADJUSTABLE DETECTOR.

Any boy with a set of stocks and dies or a machine shop where he can use them can make this detector. It needs no explanation as by referring to the illustration the parts can be easily constructed.

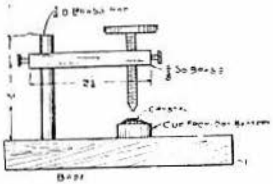

Contributed by

THORNTON KEARFOTT.

ROTARY DETECTOR.

The case is made of ⅝-inch oak. It is 2 inches deep, 6 inches wide and 4 inches high. A small window is cut in the front at the top to inspect the contact. On top a large thumbscrew turns in a brass plate, to make contact. On the front there is a knob and pointer which will indicate what mineral is being used.

The wheel which the minerals are placed on is made of a strip of copper (Fig. 2). Small brass cups which hold the mineral are soldered to this strip. This is then tacked around a wooden disc, 3 inches in diameter. Through this disc the shaft is placed which turns in the front and back of case for bearings. A copper strip is tacked to the bottom of the case, inside, which makes contact with the strip on the wheel. Connections are made from this strip and from the brass plate on top. The drawing will make the description more clear.

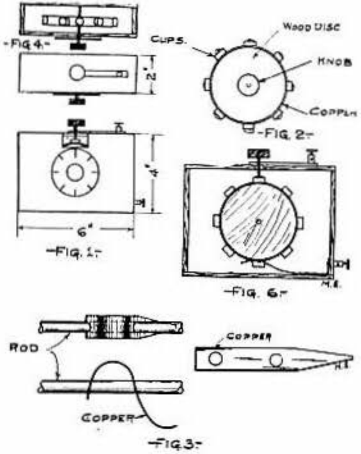

Contributed by
RAYMOND RUTHERFORD.

VARIABLE CONDENSER.

The first thing to do is to obtain four pairs of handles from a medical coil. These may be obtained from almost any electrical house or from your friends who are equipped with medical batteries.

A piece of hard wood or fibre seven inches by six, and about half an inch thick will do for the base. Another piece of wood seven inches by two inches and half an inch thick will do for the upright shown in Fig. 1.

Now bore four holes in the upright, 1½ inches from the bottom and 1 3-16, 2 13-16, 4 5-16 and 5 15-16 inches from the end. These holes should take an 8-32 machine screw.

Screw the upright to the base and we are ready to use the electrodes. Remove the screws from each end and insert 8-32 battery screws through the holes left.

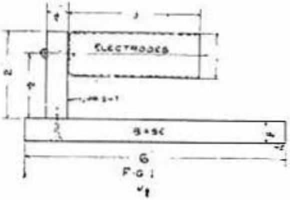

FIG 1

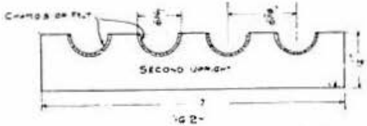

FIG 2

The four larger electrodes may now be clamped to the upright and the latter fastened to the base as in sketch. The four nuts may be connected by a piece of sheet brass or thin wire.

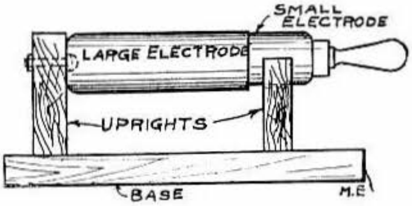

-FIG. 3-

Dip the four smaller handles in thick shellac and let stand till dried. A brass bar six inches long three-quarter inches wide and one-quarter inch thick is now clamped to the four smaller electrodes

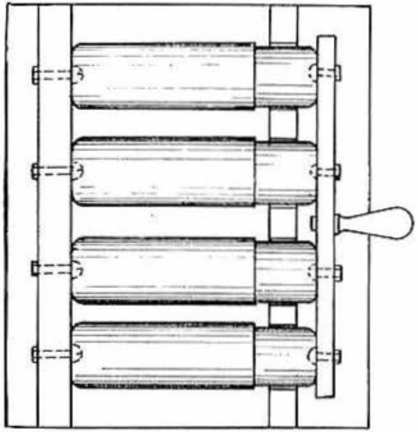

in the same manner as before. A handle may be fastened to the bar to facilitate its use. An upright is made of wood as in Fig. 2 and fastened to the base. Two binding posts complete the instrument. The movable electrodes should be connected to the binding posts and instrument should look like Fig. 3. A high polish on all wood and brass parts gives it a very handsome appearance and any amateur may well be proud of it on his instrument table.

Contributed by

STANLEY F. PATTEN.

A CLOSE COUPLED TUNER.

We have herewith the plans for a very close tuning coil.

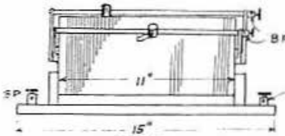

-FIG 1-

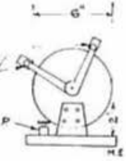

-FIG 2-

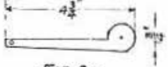

-FIG 3-

The coil is wound with No. 22 bare copper wire, and in such manner that the wire does not touch the next wire in any instance. (This can be done by winding a thread between the wires.) The brackets are put on so that they will move, but not enough that they may fall of their own weight. The connections are the same as any other double-slide tuner.

With this coil very sharp tuning can be done and it also replaces the small tuner usually used for sharp tuning.

Contributed by

HAROLD WOOD.

GALENA DETECTOR

The drawings herewith are intended to illustrate a detector I constructed recently, which, after testing, I think will be welcomed by the amateur who still adheres to the mineral type of detector.

One great fault I find with the average detector is that the up and down adjustment of the contact is not sufficient to find the sensitive spot in the mineral; and if the adjustment is made by sliding the cup around it is liable to get out of adjustment at the least jar.

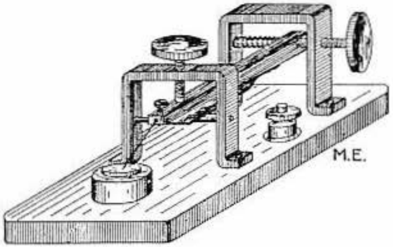

The adjustment of my detector is very delicate; and to use it, first adjust vertical thumb screw until wire just touches mineral, then turn side thumb screw, drawing the contact wire sidewise over the mineral until a spot is found. If all goes well, and a good piece of galena is used, a large station will come in like thunder.

The contact wire may be No. 30 to 36 platinum or copper (I prefer platinum). Manipulating the top thumb screw gives a front and back adjustment to the contact point. This, together with the side thumb screw, permits

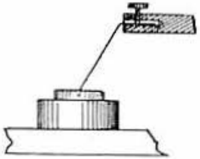

Detail of Contact with Crystal.

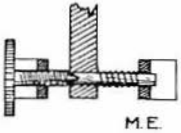

Construction of Rear Pivot.

any point on the surface of the crystal to be reached. The cup is fixed in position, but may be revolved.

Contributed by

J. F. ARNOLD.

A FLAME AUDION.

The Audion as a receptor of Hertzian waves seems to have lost a place in the heart of the wireless experimenter. However, the following may interest a few, because it provides a good field for investigation. The object of all audions, is to get one of two terminal plates to throw off ions, which act as a relay to high-frequency oscillations that pass between the plates.

Bunsen burner is used to furnish the flame, A, and salt or any alkaline metal in the flame produces the ions. The salt is contained in a small trough of platinum, which is easily made by flattening and shaping a piece of platinum wire, since the trough is only three-eighths of an inch long and one-sixteenth of an inch wide. About one-sixteenth of an inch above the trough is placed another piece of platinum wire. The trough is the cathode of the telephone circuit and the wire the anode.

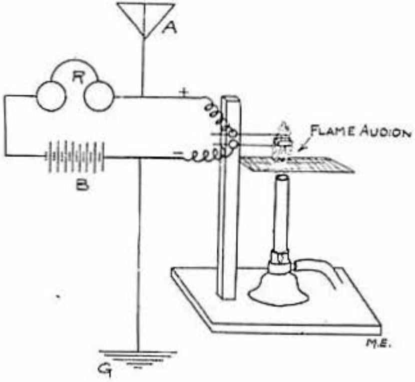

This detector is so sensitive that the least flicker in the flame is registered as a sound in the receiver. So in order to produce a steady flame, a wire gauze is placed just above the burner, then, the flame produced is extremely steady. However, it should be provided with a mica chimney to protect it from the air currents.

The sketch explains the arrangement and construction. The base and pillar are made out of wood, or better still hard rubber. Binding posts are fastened to the pillar to take the trough and wire.

Eight dry cells, connected to a multiple point switch so that any adjustment from 4 to 12 volts may be secured, are placed in series with the detector and receiver. By keeping plenty of salt in the trough and carefully adjusting the voltage, a marvelously sensitive detector may be made.

Contributed by

MEARLE MELLINGER.

I have been an experimenter in wireless for about three years. During that time I have been bothered with a continual hum in my phones.

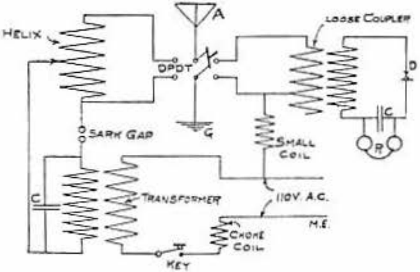

I have finally, after trying many ways, succeeded in stopping this hum without decreasing the signals in the least, and if anything, increasing them.

The drawing will explain itself. The coil does not have to be of any particular design, although I used an old tuner, wound with No. 24 S. C. cotton wire. I also find that it will cut out static to a certain extent.

Contributed by

PHILIP T. BROWN.

A VARIABLE SENDING CONDENSER FOR ANY SIZE COIL.

Illustration is self-explanatory; full details may be obtained from sketch.

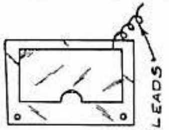

-FIG. 1-

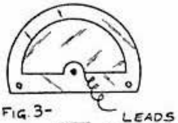

-FIG. 3-

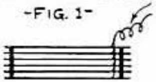

-FIG 2-

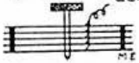

-FIG 4-

This is a good condenser for changing from a large coil to a small one. Dimensions and number of plates can be varied.

Contributed by

HARRY KING.

ELIMINATION OF INDUCTANCE DISTURBANCES.

A clever method has been suggested and found to be successful in practice, for the elimination of sounds in telephone receivers, produced by neighboring power lines. This method, set forth by S. G. Brown, the well-known British wireless investigator, is shown in the diagram which is self-explanatory. A five-slide tuning coil, three adjustable condensers, detector, and telephone receivers are necessary. Probably a three-slide tuning coil will fulfill the purpose if the end connections on the coil are used in place of two of the sliders.

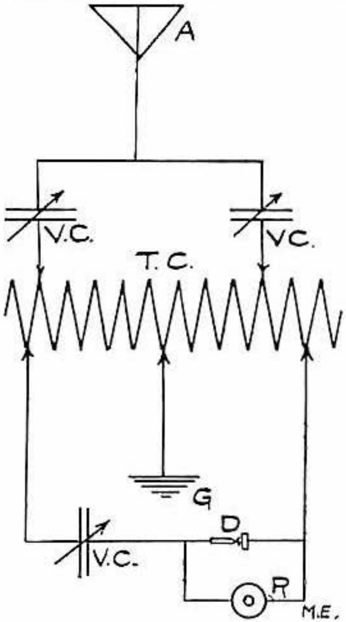

UNIVERSAL DETECTOR.

An improvement over Mr. Hutchinson's detector in the May issue of *Modern Electrics* is shown in the illustration.

Instead of fixing the slider rod on two brass uprights it is pivoted on a bar, A, at one end, and at the other end is provided with an adjusting mechanism. This consists of an aluminum bracket, B, the kind E. I. Co. use on the "Electro"-lytic detector, and a thumbscrew, C. Before screwing the bracket, B, to the base, a spring, D, is cut from a one-sixteenth

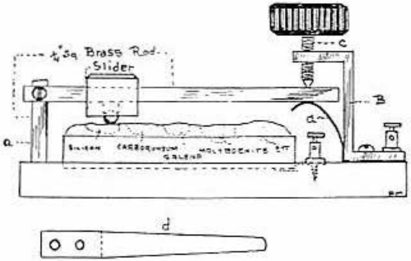

inch sheet brass, and drilled as shown. It is then fastened under bracket, B, as shown.

This spring is used to keep the slider rod up against the screw, C.

Contributed by

P. MERTZ.

ANOTHER VARIABLE CONDENSER.

Enclosed herewith are some drawings that perhaps will interest some of your readers, on a variable condenser which I am using with good results. It is a fine looking apparatus when polished and finished, besides easily operated.

The wood is made out of oak, three-eighth-inch stock, and polished or stained with mahogany dye. The contacts are made of German silver and the knobs are from a typewriter which can be gotten at any typewriter store. The condenser is made of tinfoil and paraffined paper, being sealed in wax after all connections are made in the base. The brass tube can be cut with tin shears and the edges filed until the contacts slide on easily.

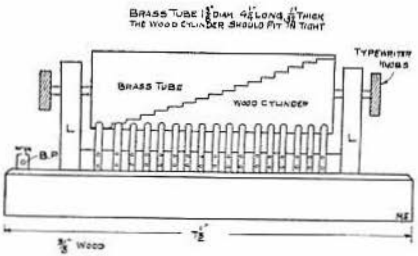
BRASS TUBE 1 5/8" DIAM 4 1/2" LONG 1/32" THICK
THE WOOD CYLINDER SHOULD FIT IN TIGHT
TYPEWRITER KNOBS
BRASS TUBE
WOOD CYLINDER
L
L
B.P
M.E.
7 1/2"
3/8" WOOD

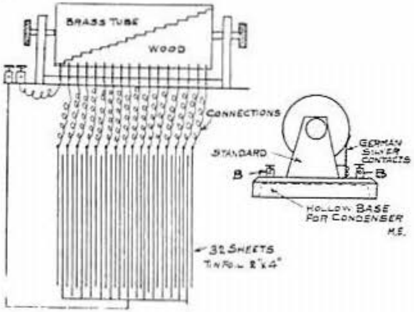
BRASS TUBE
WOOD
CONNECTIONS
STANDARD
GERMAN SILVER CONTACTS
B
B
HOLLOW BASE FOR CONDENSER
M.E.
32 SHEETS TINFOIL 2"x4"

The wires may be soldered onto the contacts and then run through the base. I think the rest is self-explanatory.

Contributed by

RUDOLPH W. ALSING.

IMPROVED SILICON TICKER.

The following apparatus is an improvement on the ticker described by the writer on page 88, May issue of *Modern Electrics.* I have been waiting to see an improvement appear in *Modern Electrics* on this piece of apparatus but am disappointed; evidently not many experimenters having used it.

The improvement consists of soldering an inch or more spring phosphor bronze one-eighth of an inch wide to an ordinary Gillette safety razor blade which is very flexible and has sharp edges to make contact with the silicon. The blade should be about one-eighth of an inch above the bell magnet. It should rest gently on the silicon. The rest of the ticker is constructed the same as the first one described in the May issue.

The advantage in this form is that it is very much easier to adjust and holds its adjustment indefinitely. It can also be used to receive ordinary spark signals by simply cutting off the A. C. current. Using it this way it is very much better than a so-called "cat whisker detector." The Poulsen people

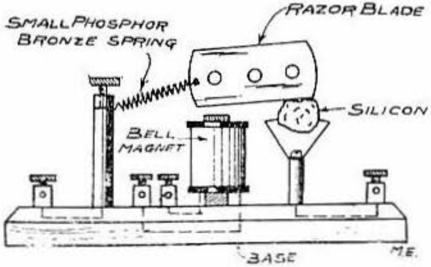
SMALL PHOSPHOR BRONZE SPRING
RAZOR BLADE
SILICON
BELL MAGNET
BASE
M.E.

have certainly "got one over" on the spark systems.

The station in this city keeps up unbroken communication with El Paso when the United Wireless and amateurs are sending within a few blocks' radius. They cannot hear spark stations unless they connect on an ordinary receiving outfit, which they sometimes use.

Contributed by

STANLEY E. HYDE.

A NOVEL CONDENSER.

This variable condenser is quite convenient to use in connection with a semivariable condenser where the glass plates are very large, as after cutting in or out the proper number of plates, the adjustment may be brought to a very fine point by moving the movable glass plate up or down.

The construction may be clearly seen in sketch.

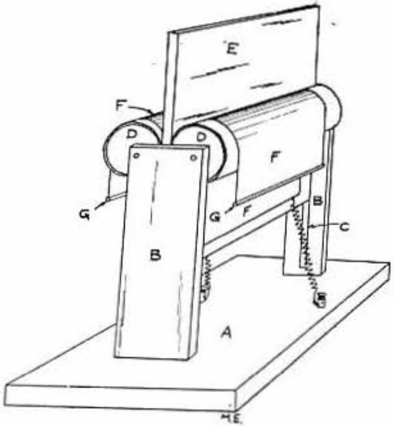

A is suitable wood base. B. B. are wooden uprights with grooves, C, on the inner sides. D. D. are wooden rollers—pieces cut from a curtain pole are excellent—covered with felt or several layers of soft cloth. E is glass plate of proper size to move freely in grooves, C. F. F. F. is tinfoil an inch less in diameter than the glass used. G. G. are metal rods to counterbalance the glass plate.

Fasten the rods used for counterweights to one side of each piece of tinfoil and then glue the opposite side of each piece along a strip of one-half inch on opposite sides of glass plate, the right distance from edge, and next give the rest of the plate that is to come in contact with the tinfoil a coat of petroleum jelly and the plate is ready to insert between the rollers, as shown in sketch.

The rollers must work smoothly and press tinfoil and plate firmly together. After adjusting plate connect a piece of fine wire, coiled spring fashion, to tinfoil on each side and opposite ends of plate and bring the other ends down to two binding posts on base and the condenser is complete.

To operate push plate down or pull up, thereby causing tinfoil to adhere to or strip from plate, till the point for best results is found.

Contributed by

R. S. TRUE.

A NEW TYPE OF VARIABLE CONDENSER.

Here is a new kind of variable condenser which is very effective in cutting out static and high voltage line interference.

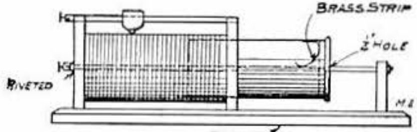

-FIG 1-

Materials required: One brass rod ¼x8 inches, 1 thin brass strip ¼x2½ inches, 1 slider to fit, 1 mailing tube 3x7½ inches, 1 brass mailing tube 2¾ x7 inches, 1 fibre end for small tube ½ inch thick, 2 block ends 4x4 inches, 1 base 5x16x½ inches, 1 brass tube 14 x¼ inches, 1 brass rod ⅜x1⅞ inches,

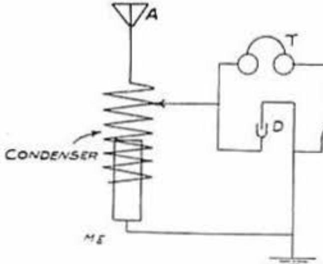

-FIG 2-

1 brass rod machine screw about 14-20 inch round head, ½ pound No. 18 bare copper wire.

Assemble the parts as shown in diagram. Connections are also shown in Fig. 2. A quarter inch hole is drilled in one of the binding posts to fit the tube.

Contributed by

LAWRENCE KREIDLER.

AN INDUCTANCE OR LOADING COIL.

Sometimes when you want to listen to a high powered station like Glace Bay you find that the wave length of the station is greater than the combined wave length of your antenna and tuning coil. If, however, an extra coil of wire is placed in the aerial circuit, the wave length is considerably increased.

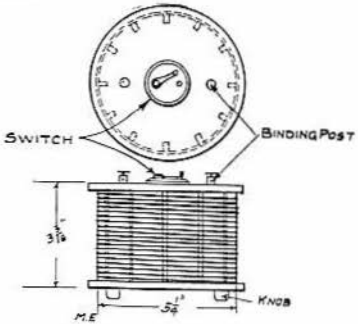

To make this extra inductance coil, procure several wooden spools such as annunciator wire comes on, several hundred feet of enameled wire No. 22, B. & S. gauge, and five or six large binding posts. The spools will be found to measure about 5¼ inches in diameter by 3 3-16 inches high; this will allow about 55 turns of wire 18 inches long, the turns being nearly 1-16 of an inch apart. Cut 16 pieces of hard wood 3 3-16 inches by ⅛ inch by ¼ inch, round the corners with a knife, and fasten around the circumference of the spool with 1 inch brads equal distances apart. The completed frame should be given a couple of coats of varnish and allowed to dry before winding on the wire. Wind 55 turns of wire on, with the turns about 1-16 inch apart. Fasten two binding posts on the top of the coil and connect the two ends of wire to them. Over the hole in the center of the coil head place a battery switch (1 point) to be used for shorting the inductance when not needed; connect one pole of the switch to each binding post. Give the whole coil a coat of varnish and while still wet place a piece of thin rubber around the wound part. This gives a finished appearance to the coil and helps to keep the wire in place. Mount on small porcelain knobs.

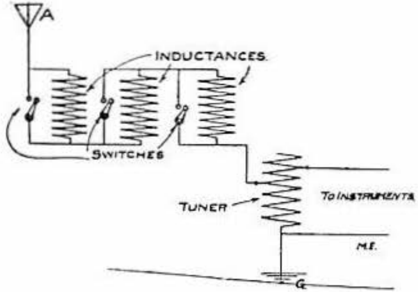

If your aerial and tuner are small several inductances may be required, and all must be connected in series.

For details and wiring diagram see the drawings.

Contributed by

R. S. CRAWFORD.

ENCLOSED COMBINATION DETECTOR.

The detector described in this article has been tried out on long distance receiving and holds its adjustment admirably. All parts can be made on a small lathe with little trouble.

Fig. 1, is a sectional view of the detector. The hard rubber case is turned from a piece of hard rubber 1¾ inches in diameter. The inside diameter is 1⅛ inches. The top edge is beveled so it will not have such a square appearance, the bevel is about 45 degrees. The drawing shows a sectional view of the detector and is approximately full-size, so all the dimensions are not given as you can vary the dimensions to suit yourself.

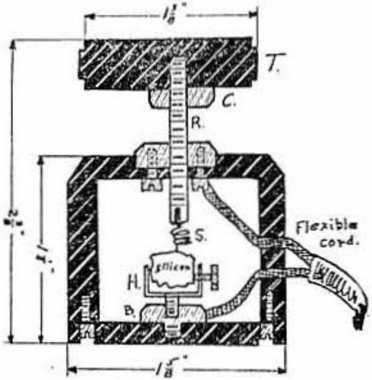

Cross Section of Detector.

The thumb screw and the case are turned from hard rubber. You will find the hardest part in turning out the inside of the case; the walls of which should be left at least one-quarter inch in thickness, so there will be room for the small screws that hold the bottom cover on. The cover should be fitted to the case after it has been turned out to the proper size. Rod (R) is threaded with 10-24 die, as well as the bottom part of the crystal holder. Check-nut (C) is screwed on to the rod to help out the appearance and also brace it on tighter to the thumb-nut.

Great care should be taken in threading any of the rubber parts as the tap is apt to break the threads if screwed in too fast. On the end of the rod is soldered a comparatively fine steel spring wire that fits in a small hole bored in the rod. Holder (H) is made so it can hold any mineral by the use of the screw which clamps it in place.

Finished Detector.

Two holes are bored in the side of the case so as to let two leads of a flexible cord pass through, which are soldered, one to the small screw which holds the rod in place, and the other to the bottom of the crystal holder. The finished rubber case can be then buffed up and made to look very attractive.

The photograph gives a much better idea of the appearance which this detector gives when it is finished. It is capable of very fine adjustment and can be used for any crystal. The length of the cord is made to suit the needs of the user, being in this case one foot long.

Contributed by

STANLEY E. HYDE.

FIXED CONDENSER.

Herewith find a sketch of a fixed condenser, which I think will be very efficient in wireless work.

It will require a one quart fruit jar and a strip of tinfoil 3 inches wide and long enough to go around the jar, and a wet battery zinc. Remove the glass from under the can cover and bore a hole in the top of the cover, next fill with water up to dotted line on sketch, add a teaspoonful of salt to the water. The sketch will explain the rest.

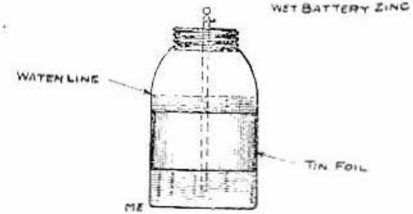

Contributed by

JOHN HOGAN

AN IMPROVED POULSEN TICKER.

For all operators who are able to hear any of the various Poulsen stations in the United States, an efficient "Ticker" will prove of considerable pleasure and interest. The "Ticker" to be described is very simple of installation and operation.

A low resistance telegraph relay comprises this "Ticker." This may be purchased new at any supply house from three to four dollars, and second hand, for very much less.

The magnets of the relay are connected to any alternating current source of electricity in series with a 16 c.p. lamp and suitable means for making and breaking the circuit. The other two binding posts on the relay, which are connected to the platinum contact points, are connected to the two sides of the fixed condenser in the receiving circuit. Two different hookups are shown below.

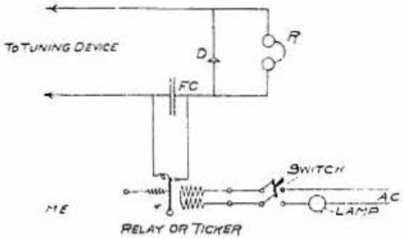

-FIG. 1-

Whenever the familiar frying of the Poulsen arc is heard, the "Ticker" should be switched onto the A.C. mains, and the signals may be read as easily as with the spark system of wireless telegraphy.

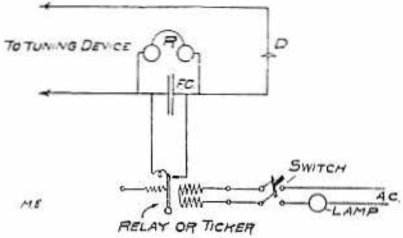

-FIG. 2-

This "Ticker" will prove much more efficient than those which are directly connected to the rectifying element in a detector. The moment that the "Ticker" is disconnected from the A. C. current, the set is receptive for spark system signals.

Contributed by

ELLERY W. STONE,
California.

A SIMPLE LOADING COIL.

A simple and efficient loading coil for receiving long waves may be made as follows:

Make a hollow tube of cardboard eight inches in diameter and eight inches in height. Boil the tube in paraffine until the bubbles cease to come from the cardboard. Procure several hundred feet of No. 24, B. & S. gauge enameled wire.

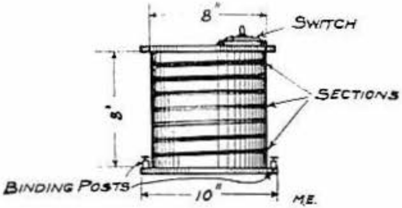

One half inch from the top of the tube start winding the wire and continue winding until a section one-half inch wide has been completed. Break off the wire allowing about one foot for lead and fasten. Measure one-half inch from completed section and wind another section. Continue making these sections until the tube has been filled but a space of one-half inch should separate each section. The lead from each section should be led to a switch and connected as per diagram.

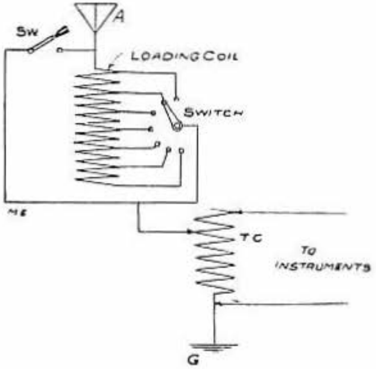

This coil makes a very neat instrument when constructed with care and is capable of getting those high-powered stations with ease.

The coil will be greatly improved if end pieces are added and the switch mounted on one end.

Contributed by
ALLEN DAHLQUIST.

AN OIL CONDENSER.

An authority states that more sending sets prove inefficient through losses brought about by a poor condenser than from any other one thing, losses in some stations amounting to as much as 33 1-3 per cent. of the total energy. Why not save this lost energy by building the condenser right in the first place?

From any large photographer procure 40—8x10 inch used plates. He will give you all that you can carry away. These plates are particularly suited for this purpose as they are absolutely free from flaws of any kind. Clean the plates in hot water and dry them thoroughly. These plates may seem small but they allow the use of six-inch tinfoil, which is much easier to procure than wider sizes.

Between two and three pounds of tinfoil will be needed as the lugs and plates are to be cut in one piece. Cut the pieces 6x8 inches with a lug 1 inch wide and 3 inches long on one end. Alternate these lug ends so that on adjacent plates the lugs come out at opposite ends of the pile of plates. The first lug should be placed as at A. (Fig. 1); the next two as at B (Fig. 1); the next three as at C (Fig. 1); the next four as at D (Fig. 1); the next ten as at E (Fig. 1), and the remaining twenty as at F (Fig. 1).

The glass plates should now be thinly coated with vaseline; the plates of tinfoil placed on the glass so as to clear the edges by one inch all around, and the foil rolled tight to the plate. The vaseline makes an excellent adhesive for this purpose, and as it does

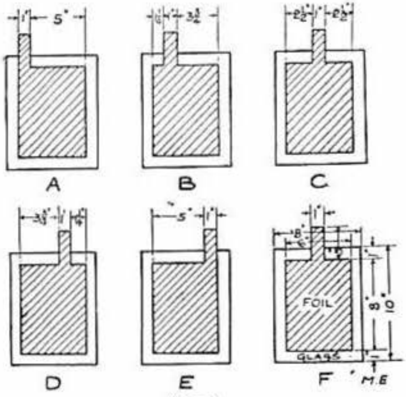

-FIG. 1-

not harden all of the air can be gotten out from under the tinfoil. When the foil is all rolled tight coat the whole plate with from one-sixteenth of an inch to one-eighth of an inch of vaseline. Treat the remainder of the plates in a similar manner. Now pile up the plates so that at one end 20 lugs are at the center, while at the other end there are five bunches of 1, 2, 3, 4, and 10 lugs each respectively. These bunches of lugs must come from consecutive plates and be on the same part of the plate. The pile should look as in Fig. 2. Compress the pile and tape it together.

Build a box of oak (⅝ inch stock) 10x14x5 inches (outside measure). Miter or rabbet the corners, glue the bottom on, and fasten the top on with nickeled round headed brass screws. The box may be filled and varnished or just shellacked as preferred. It should be well shellacked inside, however. On one end of this box a hard-rubber or fiber plate ¼x4x9 inches should be fastened. On this plate mount five 15 ampere S.P.S.T. knife switches and two large binding posts.

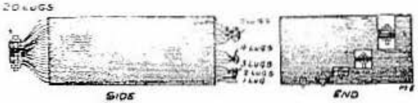

-FIG. 2-

One end of every switch should be connected to one of the binding posts by a copper bus-bar. The five other ends are connected to the five bunches of lugs 1, 2, 3, 4, 10, in order. The lugs at the other end of the plates are all connected together and to the other binding post.

This about completes the construction. By this arrangement any number of plates from one to twenty can be cut in by closing the proper switches. For instance, if 17 plates were desired, the switches connected to the bunches of 3, 4, and 10 plates would be closed, and so on. The thick vaseline cuts out the brush discharge almost entirely and furthers the general insulation. This condenser was designed for use on any coil or transformer up to and including ½ kw., but any larger size can be built accordingly.

Contributed by

L. W. TELLER.

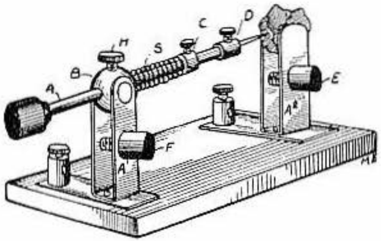

-FIG. 1-

UNIVERSAL DETECTOR.

The parts of the detector are mounted on a hard wood base 2½x3½x½ inches. The edges may be beveled to give a better appearance.

Next make two U-shaped standards (A^1 and A^2) of spring brass, using material ½ inch wide and 1-16 inch thick. They should stand about two inches high, thus requiring 4½ inches of material for each standard. One of these clamps the mineral while the other supports a ½ inch brass ball (B), which is held in place by practically the same method used on the back of wireless receivers, and allows the rod A to be moved in various directions.

By studying the drawings one will easily be able to understand how the detector works. Suppose you want to use a mineral that requires a rather heavy pressure on its surface. Place the mineral in standard A^2 using the knob E to clamp it tight. Thumbscrew H must be loosened so that rod A is able to slide back and forth through ball B. The pressure of the point on the mineral is regulated by the collar C, by changing the tension of the spring S.

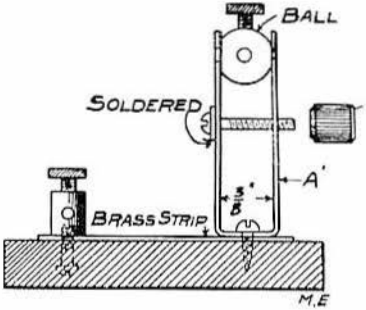

-FIG 2:-

All that is necessary now is to place the point on a sensitive spot of the mineral. Silicon, carborundum, and pyron are generally used with the set screw H loose, using the spring S to push the point against the mineral. Galena is an excellent mineral to use.

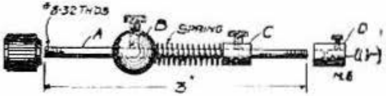

-FIG.3-'

The detector may be used as a perikon type, by clamping the zincite in A^2 and for the other mineral solder a piece of bornite in a cup having a shank on it, and clamping the shank in the tip D.

The writer has made two of these detectors and mounted them on his receiving cabinet, using a double point switch to change from one to the other. By connecting two detectors this way they may be easily adjusted to the highest degree of sensitiveness, which they possess.

Contributed by

R. COWDEN.

VARIABLE CONDENSER.

In the September issue of this magazine appeared a Novel Variable Condenser contributed by R. S. True, one of which I made and found to be alright, except for the cloth covered wood rolls. I found it almost impossible to get them to run true. When the glass plate was moved up or down rolls would move away from plate just enough to prevent the tinfoil from making a good contact with plate.

So I made a device as shown in enclosed sketch which is composed of four brass strips two on each end of plate ½x1-16x5 inches long, the strips having 3-16 inch holes bored ½ inch from each end, and two small coil

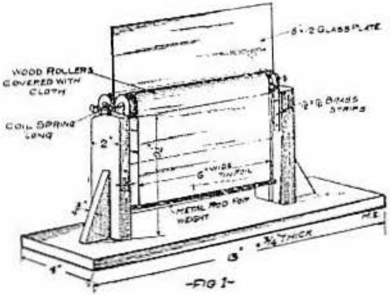

-FIG 1-

springs, one for each end of the condenser. These are slipped over the screws together with the brass strips before the screws are put into the wood rolls. I have found this to give fine results and it works very smoothly when glass plate is moved up or down. As the brass springs hold rolls firmly against plate nearly perfect contact between the tinfoil and the glass is secured. The enclosed sketches will fully describe everything.

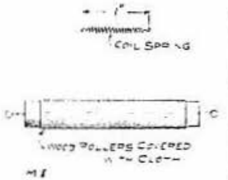
1"
COIL SPRING
C
C
WOOD ROLLERS COVERED WITH CLOTH
M E

-FIG 3-

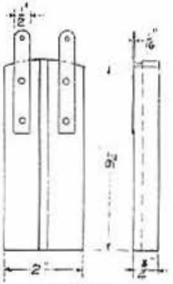
1/2"
1/16"
9 1/2"
2"
1/8"

-FIG 2-

Contributed by
CLARENCE E. CHAMBERLIN.

A NOVEL VARIABLE CONDENSER.

A variable condenser of almost any desired capacity may be made as shown in the sketch, by tacking the fixed plate to a broad hard wood base board, allowing about 3 inches of base board vacant for tacking the movable plate on.

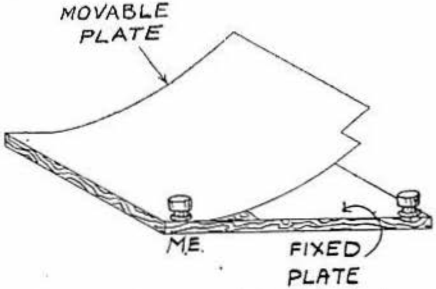
MOVABLE PLATE
M.E.
FIXED PLATE

Insulate the top of the fixed plate.

The capacity is varied by raising and lowering the top plate.

Contributed by
CHARLES A. WILSON.

PORTABLE RECEIVING OUTFIT FOR LONG DISTANCE.

The long distance receiving outfit described below cannot be carried in a "Cane" or stuck in your hat. If you are thinking of making an instrument of that kind don't read any further.

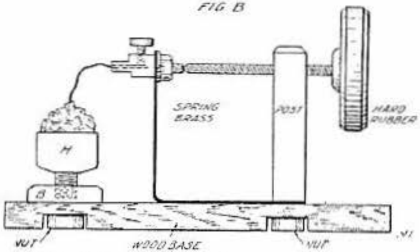
SPRING BRASS
POST
HARD RUBBER
H
B
NUT
WOOD BASE
NUT

FIG B

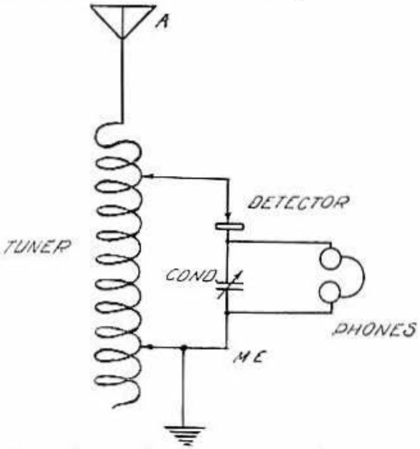
A
DETECTOR
TUNER
COND.
PHONES
M E

The first thing to procure is the box for holding the tuning coil, variable condenser and the detector. In this case the box was formerly a medical battery case measuring about 9x5x5½ inches. Excellent boxes of this kind can be purchased from the Electro Importing Co., for a very reasonable price.

The tuning coil is wound with No. 24 enameled wire and is 1½ inches in diameter. It just slips into the case and is held in place either by screws or carpenters' glue, which is very strong. The sliders are the regular hard rubber ball bearing sliders, which are made to fit a ¼ inch brass rod. Care should be taken in scraping off the insulation from the wire where the sliders make contact with it, as the wire is not very large in cross section.

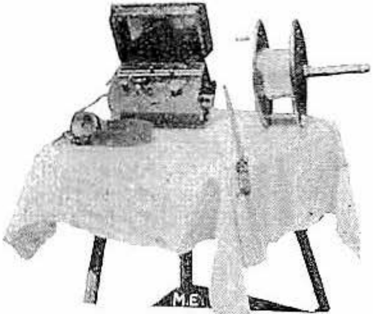

Portable Wireless Outfit.

Th condenser is made up in sections and each one is twice as large as the one preceding, the first one being 10x4 inches. A good way to get these condensers is from a local telephone company. They can be purchased for about 25c. and about 10 receiving condensers can be made from a single telephone condenser. There are five of these and taps are brought out on the case with a switch for cutting in the different capacities, as shown in the picture.

At the left of the condenser is the detector which is shown in Fig. B. The crystal holder (H) is ½ inch in diameter and in this is soldered the silicon crystal or other mineral that is preferred. It is threaded with 10:24 die to fit the holder base (B). It is well to make a number of these holders so that any crystals can be used. An excellent solder that will cool off but will not harden for about fifteen minutes can be made by melting one part mercury with three parts of solder.

Two binding posts are provided on each end of the case, one pair for the telephones and the other for the ground and aerial. The phone is a single one of 1,000 ohms and can be carried in the pocket along with the short head-band.

The spool for carrying the aerial wire is a large magnet-wire spool and measures 9 inches in diameter. It is 3½ inches deep and will carry about 2,000 feet of No. 18 annunciator wire. This wire is very flexible and does not kink easily.

With the connections shown at Fig. (C) the use of a ground connection is determined by the length of the aerial wire. If the wire is very long the use of a ground actually hinders the receiving ability of the set. This is not always the case, but for every 7 stations out of 10 this rule will hold true.

So that if you are out in the wilderness some where, you don't need to worry about a ground connection if you can tie a rock to the end of the wire and heave it up into some high tree and stretch the wire out for about 300 feet.

To show the sensitiveness of small sets like this, the writer one morning some months ago, stretched a wire from a tree 15 feet high for about 100 feet across a vacant lot and without any ground heard the (S.O.S.) sent out from the ill fated S.S. Santa Rosa, which went on the rocks near Pt. Arguello during one of the most severe storms ever experienced on the Pacific Coast. Contributed by

STANLEY E. HYDE.

CHEAP POTENTIOMETER.

Below is description of an efficient potentiometer.

The base is made of wood (preferably hard wood), size about 8x8 inches, ¾ inch thick.

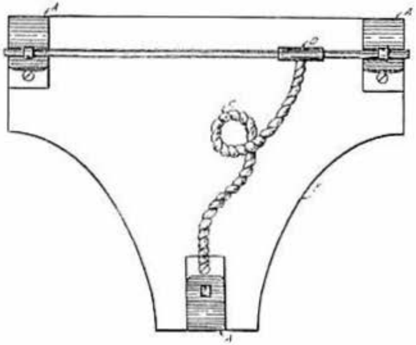

The graphite rod is taken from a hard lead pencil, it can be taken out by soaking pencil in hot water for some time.

The ends of rod are held in place by inserting ends in spring clips at ends. The contact is made with rod by a short piece of brass tubing, just large enough to slide along easily on graphite rod. To this tubing is soldered a piece of flexible cord, which is fastened to binding post at bottom.

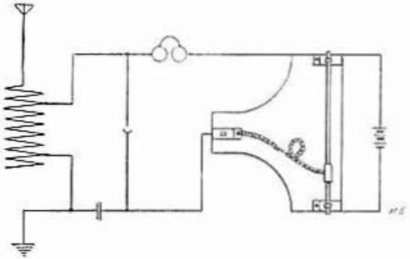

Drawing shows how to connect up for use.

Contributed by

JOHN BURLEIGH.

PERIKON DETECTOR STAND.

The following is a description of a "Perikon Detector," which I have been using for some time with excellent results. It cost me exactly $.40, and was made, excluding the base, in twenty minutes. All the material necessary can be purchased from the Electro Imp. Co., and will conform nicely.

The base is of mahogany, 3x5½ inches, and beveled to improve its appearance. If this cannot be procured, a S. P. D. T. switch base will do nicely. The two standards, are double binding posts. The one supporting the zinc crystal is set on two 3-16 inch washers. When fastening it down it should be let a little loose, and a drop of oil be placed between the washers so as to act as a swivel. The other one is set directly on the base. A large brass washer or pillar is next procured; with a hole in it to pass an 8-32 machine screw. It should be ⅝ inch high and ½ inch in diameter. A hole is drilled in the side, just deep enough to get a few 8-32 threads in it. A small piece of threaded brass rod fastens it on the binding post. However, it should clear the top by about 1-16 inch so as to revolve in a semicircle freely. Both the rods are of 5-32 inch brass, threaded 8-32 at both ends. The spring should be rather weak, fine phosphor-bronze wire is best. I think the drawing explains the other points. Fig. 2, is the small piece which allows the copper pyrites cup to revolve in a small arc. It is cut out of 1-16 inch brass sheeting, and the holes bored about ¼ inch apart. The size is dependent on the size of the cup used. The pressure on the minerals can be varied by moving the zincite mineral backwards or forwards. It may take a little while to get used to this de-

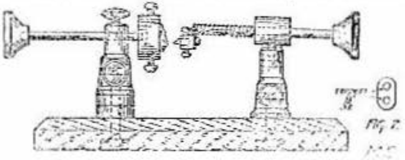

Fig. 2

tector, but it is very sensitive, and there is no possible adjustment which cannot be made; it being possible to reach every point on the zincite. Also, when sending, instead of shorting the detector, the minerals are disengaged, thereby saving them from being burnt out.

Contributed by

GEO. W. KELBY, JR.

CRYSTAL MOUNTING.

As most experimenters know, soldering takes a good deal of the sensitiveness away from the mineral, and tinfoil wads no matter how tightly packed, do not hold the crystal firm. Wads of fine copper wire (34-36 B. S.) jammed between the crystal and the cup hold it firm and make a good contact.

Contributed by

VAN ALLEN TREAT.

IMPROVED ELECTROLYTIC DETECTOR.

In no wireless instrument have makeshift methods been more often resorted to than in the electrolytic detector. Designs for such detectors by men admittedly experts in the field display most amateurish crudity. The stumbling block has been, of course, the attachment for holding the delicate Wollaston wire. Amateurs become disgusted with fingering minute screws and discard a most excellent detector for a more convenient, if not so efficient, type.

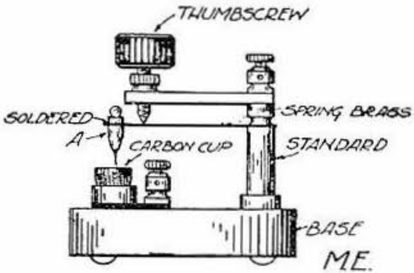

The detector shown overcomes the difficulty in a novel way. *A* in the drawing is nothing more nor less than a tie-pin guard which may be obtained at any jewelry store, slipped through a suitable hole in a piece of spring brass and soldered thereto. By this method, putting in the Wollaston wire is made an easy matter and the principal objection to the electrolytic detector is removed.

Contributed by

LEWIS C. MUMFORD.

A FIXED CONDENSER.

Make a board three-quarters inch thick by 5¾ inches long by 3 inches wide.

Make another board three-quarters inch wide, by 5 inches high by 4¾ inches wide, nail or screw them together as shown in Figs. 1 and 2. Procure two test tubes seven-eighths inch by 6 inches, coat the inside and outside with tinfoil 4 inches wide. When this is done, procure two pieces of No. 14 copper wire about 6 or 7 inches long, solder to one

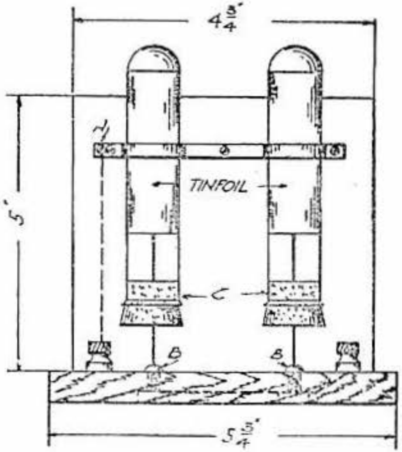

FIG. 1

end of each wire some fine copper wire taken from lamp cord. Then run them through corks that just fit the test tube, C, Fig. 1. The fine wires are to make contact with the inside foil, Fig. 4.

The tubes are held to the base by the battery binding posts, B, and the strip of brass, A. I think the drawing will explain any other questions.

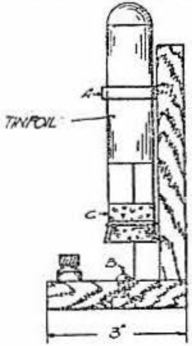

FIG. 2

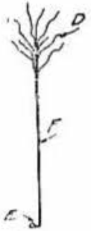

FIG. 3

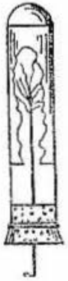

FIG. 4

Contributed by B. CARLSON.

TRANSFORMER FOR POULSEN TICKER.

Since a 15 k.w. Poulsen station will soon operate in Chicago the amateurs will want a ticker detector. Those not having A. C. cannot very well have one.

I have made one using an induction coil to get the A. C.

A core of iron wire ½ inch in diameter and six inches long is wound with two layers of No. 22 enameled copper wire. This is insulated and over this is wound two more layers of the same wire. This is the secondary, using a vibrator to interrupt the primary. When the primary circuit is closed the induced current flows in one direction and when the primary circuit is broken the induced current flows in the opposite direction. This gives an alternating current, the frequency depending on the speed of the vibrator. With 2 or 3 dry cells this will give good results.

Contributed by

ALFRED HOLMSTROM.

A SENSITIVE GALVANOMETER.

Here is a galvanometer which is very simple in construction and if properly made is only surpassed by the high resistance phone.

The base board may be of any good piece of wood of about 3x3 inches with holes drilled for binding posts. Now make a reel of cardboard or wood (the dimensions are clearly given in Fig. 1), and drill a hole for a brass tube at the top. Wind about one-fourth pound of No. 30 D.C.C. wire on the reel. Be sure that the same amount of wire is wound on either side of tube.

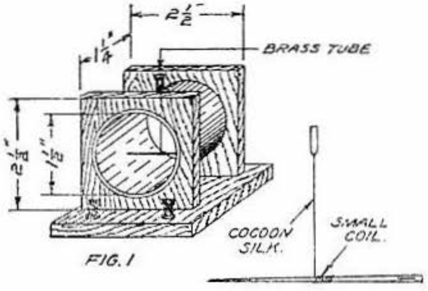

FIG. 1

FIG. 2

Now magnetize the finest needle you can find. Take a small piece of the wire and wrap three or four turns around the needle as shown at A in Fig. 2. Now place two or three inches of black silk thread on a sheet of writing paper. With one or two needles unravel one end of the thread until you have obtained a silk fibre that is scarcely visible to the naked eye. Put a drop of glue on the coil at A, Fig. 2, and carefully fasten the silk fibre to it. After the glue has thoroughly set balance the needle by pushing the coil toward or away from the eye. All the superfluous silk may be rolled upon the upper part of the thread.

Place the needle inside the coil; bring the end of the silk thread up through the brass tube and carefully pull the needle up till it is exactly in the center of the coil and fasten the thread with glue, to a wooden plug in the top of the tube. Next securely glue the coil to baseboard and connect to the binding posts. To exclude disturbing air currents, enclose with a glass case such as a lamp chimney. If the experimenter desires to compare currents with this galvanometer, place a graduated scale, drawn upon a piece of cardboard, just below the needle. Of course the coil must in any case be parallel with the magnetic meridian.

This instrument could be wonderfully improved by fastening a very small mirror to the needle and by causing a beam of light from a lamp to fall on the mirror, the reflected light could be focused on a scale placed on a suitable frame just in front of the lamp.

The author finds this apparatus so sensitive, that the tiny current generated by dipping a steel needle and piece of zinc (connected by wires to galvanometer) into a drop of salt water will set the needle swinging. Of course this instrument may be greatly improved in appearance, but the writer will leave that to the constructor.

Contributed by

ERNEST BARNHARD.

A SILICON DETECTOR

Here is a silicon detector which I am using in connection with my wireless set

The parts are mounted on a hard wood base 4½x1¼ inches and ⅜ inch thick.

Two pieces of brass, B, and D, are bent as shown in sketch. A, and C, are made of spring brass. All these are ⅜ inch wide, and of any convenient thickness. A small brass spring is placed under the end of C, supporting the cup.

The contact point, made of a small brass nail with the head flattened is soldered to A, as shown. All holes for screws are No. 19 drill and all tapped holes are made with a No. 29 drill and tapped 8-32 thread.

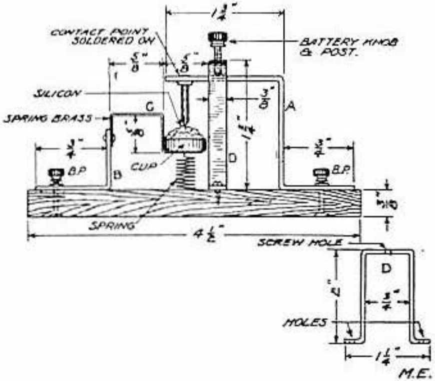

The sketch gives all necessary dimensions and shows how the detector is assembled.

Contributed by

C. W. SEAMAN.

A CIRCULAR POTENTIOMETER.

The potentiometer here described is one which will, and does, give excellent results, and may be made by any one familiar with the use of tools in an hour's time at the cost of a few cents.

The dimensions are of small account, but the ones shown will be found to be very convenient. The base is turned from any hard wood, or hard rubber, if preferred, ½-inch thick and four inches in diameter. A groove is then turned in one side 3-16 inch wide by ⅛-inch deep and with a center diameter of 3½ inches. The holes are drilled nearly all around the groove 5-16 inch apart and large enough to admit 8-32 brass screws rather tightly. Then file the heads flat on enough screws to fill the holes and screw them in place. These screw heads should all be the same height, and the shanks should be a little less than ½-inch long. In the two holes at the end of the circle place two binding posts.

The groove is then filled with a paste made of lamp black and shellac. This will dry in a few days; and the resistance is about 500 ohms. If this is not enough it may be increased by adding a little powdered slate to the paste. Another excellent substance to use is the graphite which may be purchased for greasing bicycle chains. This may be melted and poured into the groove.

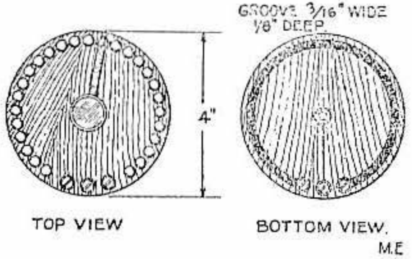

A switch lever is then attached and connection made with the middle binding post. The other two being placed at the ends of the groove, make connections with the paste when it is applied.

The chief advantage of this potentiometer is that it overcomes the trouble caused by imperfect contact when a graphite rod is used with a sliding contact.

Contributed by
SIDNEY E. ANDERSON.

A COMPACT LOADING COIL.

Finding that the usual type of loading coil, such as described by Mr. Allen Dahlquist in the February issue of *Modern Electrics,* occupied too much space I devised a more compact one. Here it is:

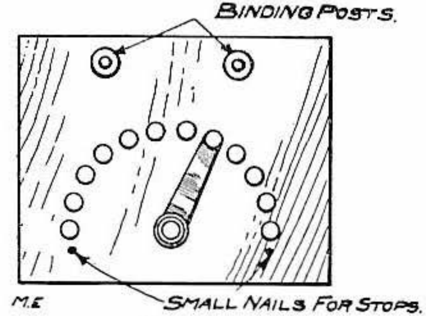

-FIG.1-

Make a box of some hard wood, preferably oak, having inside dimensions of 7x6x2½ inches.

On top of this mount a twelve point switch and two hard rubber binding posts at points indicated in Fig. I.

Next, from some ½ inch material cut ten pieces 2¼x6½ inches. Wind on each of these sixty feet of No. 24 enameled wire and place them in the box standing them side by side with a little space between each of them. A good way to insure against having them touch each other would be to slip a piece of cardboard between each one.

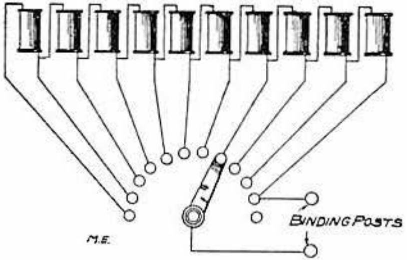

-FIG. 2-

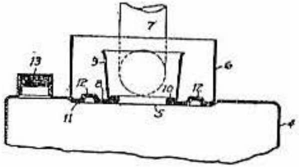

Figure II shows how the connections are made.

When all wiring is completed fill the box, containing the sections, with an insulating compound.

Screw the cover on with round headed brass screws.

If this coil is carefully made it will make an attractive addition to any amateur's set.

Contributed by HARRY HAAS.

THE CONSTRUCTION OF A ROTARY VARIABLE CONDENSER.

The following is a description of a rotary variable condenser that can be easily constructed by anyone without the use of many tools, and makes an efficient addition to any wireless station.

The accompanying diagrams give the necessary dimensions of the various parts. This condenser is built of a rather large number of small metal plates. This is done in order that it will not take up much room on the wireless table and still have a high capacity.

The condenser is chiefly used in connection with a loosely coupled tuning coil, as it's capacity is just about right for this.

Figure 1 shows a plan and elevation of the complete instrument. The top, shown in Fig. 2, is of hard rubber ¼-inch thick and 6 inches in diameter. Six holes are bored through it as shown. F is ¼-inch out of center and is ¼-inch in diameter. The three holes A, B, and C, are 90 degrees apart and ⅛-inch from the edge. The two holes, D and E, are for binding posts. All the holes except F are 3-16 inch in diameter.

The stationary plates, Fig. 3, are made of sheet brass 1-32 inch thick and are cut out to the shape of a semi-circle 6 inches in diameter. They have 3 holes bored in them to correspond with the holes A, B and C, in the top. Twelve of these plates will be needed.

The rotating plates are shown in Fig. 4. These are cut from 1-32 inch sheet brass, and have a radius of 2½ inches. A projecting lug, J, is left on and has a 3-16 inch hole in it. This hole must be ¼-inch from the straight edge of the plate. Eleven of these are needed.

Fig. 5 shows the bolts and separators. The bolt, K, 8-32 thread, is of brass, has a round head and is 3¾ inches long. Three will be needed. L, is a piece of brass tube that will slip over K and has an inside diameter 3-16 inch and an outside diameter of ¼-inch and is ⅜-inch long. Six of these will be needed. M, is a separator, which is 3-32 inch thick and ¼-inch outside diameter. It must

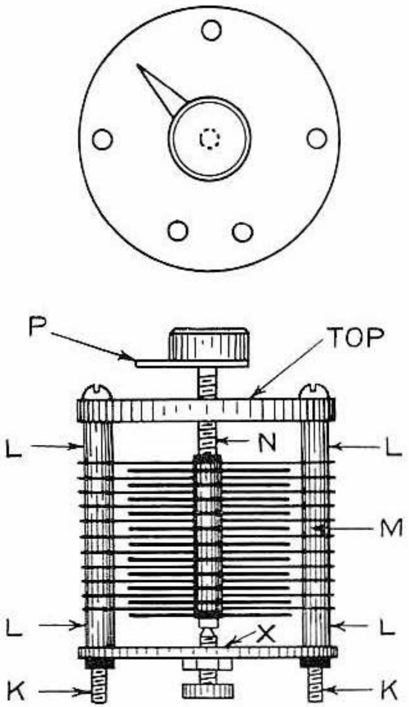

FIG. I

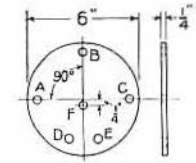

DETAILS OF TOP
FIG. 2

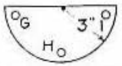

STATIONARY PLATE
FIG. 3 M.E.

be able to just make a firm fit on K Forty-three of these will be needed.

Fig. 6 shows the rotating pieces. N, is a piece of 8-32 threaded brass rod, 4 inches long and fitted with a hard rubber knob. The lower end of the rod should be hollowed out so as to fit over the point on the adjusting screw. O, is a piece of brass tube that is similar to L, in Fig. 5. P, is the pointer and is 2½ inches long, but its shape can be altered to suit the builder.

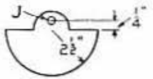

ROTATING PLATE
FIG. 4

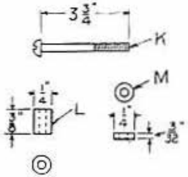

FIG. 5 M.E.

Fig. 7 shows the adjusting screw. It has a sharp point on the end and is threaded with an 8-32 thread. It is 1 inch long and has a locknut. The piece of hard rubber, X, is similar to the stationary plate; but has a lug, similar to that on the rotating plates. A hole is drilled and tapped in the lug with 8-32 thread.

threaded with an 8-32 thread. It is 1

The assembling is done as follows. First take the stationary part. Force the piece, O, in the hole, F, of the top. Then place the three bolts, K, in the holes, A, B, and C, of the top. Slip a piece, L, on each bolt, then a stationary plate, then a separator, one on each bolt, then another plate, separators, etc. After all the plates are on, put on the three remaining pieces, L, then the hard rubber plate, X, and fasten all tight with nuts on the ends of each bolt, K.

To assemble the rotating parts, first take the rod, N, and screw a nut to within ½-inch of the upper end. Put on a plate, separator, another plate, etc., until all the plates are on. Then fasten all firm with a locknut. Put the end of N up through the hole, F, of the top, and turn the plates so that they interlock. Put the adjusting screw in place and fit it into the hollow end of N. Adjust the screw until the plates swing clear. Connect the bushing, O, to one binding post and the bolt, K, to the other post. Put on the pointer and hard rubber knob.

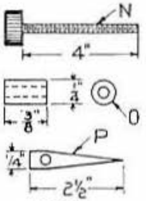

FIG. 6

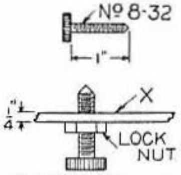

ADJUSTMENT SCREW.
FIG. 7 M.E.

The top may be marked off in 180 degrees. At 0, the capacity will be nil, and at 180 will be the maximum.

The condenser can be mounted in a case with the top exposed, or left open, as desired.

Contributed by

B. F. DASHIELL.

CAT WHISKER DETECTOR.

I enclose diagram of a wireless detector which I wish to enter in your prize contest. It is easily made and will give good service. Almost any kind of mineral may be used. Following are dimensions and directions for making:

Base is of hard wood, three and one-half inches by two and one-half inches by three-eighths inch.

First procure three large binding

posts. In one of these tap a hole for the screw, G.

In a piece of brass, three-eighths inch wide and about two inches long, drill a hole in center for fastening to base, bend to shape shown, and file the edges like a saw, to hold crystal. Fasten binding posts and brass clip in positions as shown.

Thread a brass rod about one inch long, so as to fit the hole previously tapped. To one end of this, fasten a small rubber knob. Screw the rod into the binding post. To the other end of the rod solder a fine silver, or silver-plated, wire, so as to be in line with center of crystal holder. Connect one post to the clip and the other to the wire holder.

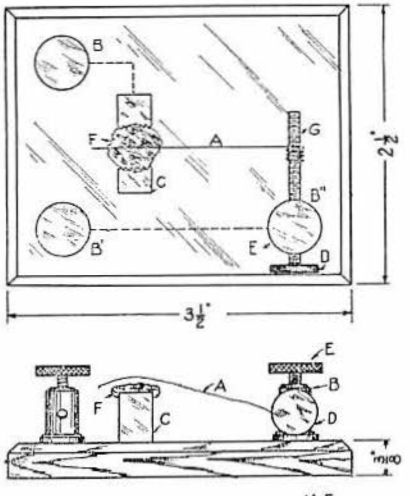

A, is the silver wire; B, B′, B″ are binding posts; C, is crystal holder; F, is crystal; E, is a lock screw on wire holder; D, is rubber knob; and G, the threaded rod.

Pressure of contact is adjusted by means of rubber knob and screw.

Contributed by

EARL ZANDER.

BALANCED DETECTOR STAND.

A is the base 5x2 inches; B is a double binding post, with top set screw removed; C is hole where lead wire goes in; D is hole in binding post for the balance rod, and the rod should fit easily in the hole, or the hole may be drilled to fit one-eighth inch rod; E is balance rod (one-eighth inch copper). The part, N, in Fig. 2 should be made first and put through the hole in B and then bent as shown. **F** is a counter weight and should be one-half inch in diameter and one-quarter inch thick with one-eighth inch hole in center to slip over balance rod E, and should have a small set screw, M. A small hole should be drilled in rod, E, at N, to receive the No. 30 steel or phosphor bronze wire, G, which should be bent as shown in Fig. 1, and be three-quarters inch from the end of balance rod to the point where it bends down to the mineral. Cup, H, should be not less than one-half inch in diameter, and should be soldered to plate, I, and drilled to receive small screw, J, and binding post, K, as in Fig. 1.

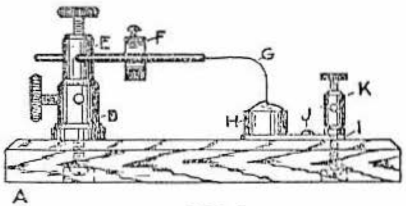

FIG. 1

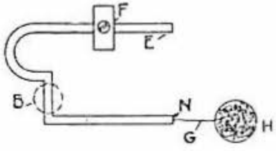

FIG. 2

Contributed by

M. E. ARMSTRONG.

A WIRELESS OUTFIT FOR A DOLLAR.

Recently I finished constructing a wireless outfit, with which, while it cost me only one dollar to make, I am able to send and receive to the extent of over a quarter of a mile. My idea in building it was to see just how cheaply an outfit could be built.

The materials I bought are as follows, together with the prices:

1 Phone	$0.35
1 Buzzer	0.35
¼ Lb. No. 24 bare copper wire..	0.10
1/3 Lb. No. 14 bare aluminum wire	0.20
Total	$1.00

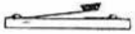

FIG. I

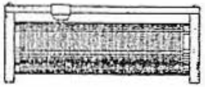

FIG. 2

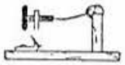

FIG. 3

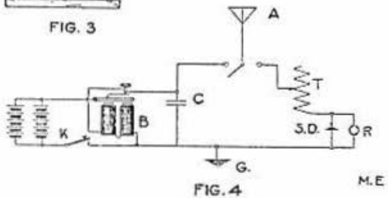

FIG. 4

The rest of the material I found lying around the house and was merely scrap.

The sending battery consists of (12) twelve dry cells connected in multiple-series (two rows of six cells in series). These batteries I got for nothing at different garages in the city, being considered no longer fit for use; but I found I could get quite a respectable current from them when a large number was used.

The key is made of strips of brass, (or other metals may be used) 1-32 inch by ½ inch. Figure 1 shows the key. The condenser is made of a 4 x 5 inch photograph plate with a sheet of tin foil 3 x 4 inches on each side and having a lug ½ inch wide extending out from the glass. Wrap the condenser with some thick brown paper and glue or bind it. The lugs should be wrapped around two (2) pieces of copper wire about six (6) inches long, and soldered or bound tightly with thread.

For the receiving set you will need a small tuner made as shown in Figure 2. The No. 24 bare copper wire is wound on with a thread between the turns. The design for the detector is shown in Figure 3. The contact on the silicon is made by a fine silver, gold or copper wire, but I find silver is the best.

The set is wired up on a baseboard as shown in Figure 4.

The aluminum wire was used in making the aerial.

Contributed by

LANCE BOYD.

A NEW TYPE OF VARIABLE CONDENSER.

Here is an easily made variable condenser.

It consists of a cylinder, A, fixed on a shaft, B, and arranged so that it can be made to rotate. A semi-cylindrical

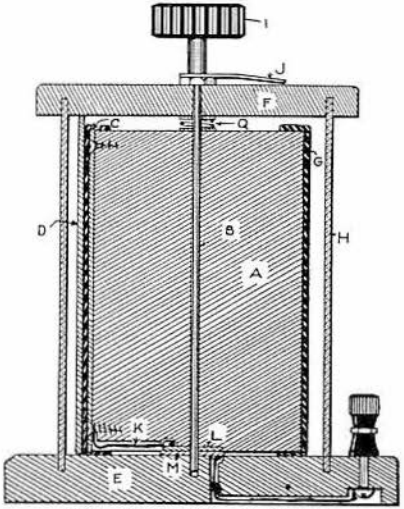

FIG. I. SECTIONAL ELEVATION

piece of sheet aluminum, C, is screwed to the cylinder, A, in a depression in the latter so that both surfaces lie flush. This is clearly shown in Fig. 2.

Over the cylinder, A, is glued a thin sheet of insulating material, G. In this instance it was paraffined paper but a coating of shellac, asphaltum, mica, etc., would be just as good. The seam, if it is made of paper or mica, is *not* over the metal plate.

Over this cylinder is made to bear another semi-cylindrical aluminum plate, D. This has four ears (two at the top and two at the bottom), projecting, and bent over four corresponding pieces of spring-brass wire. These wires are inserted in holes made in the base, E, and the cap, F, and are so bent that the plate, D, is made to bear upon the cylinder, A, with a light pressure.

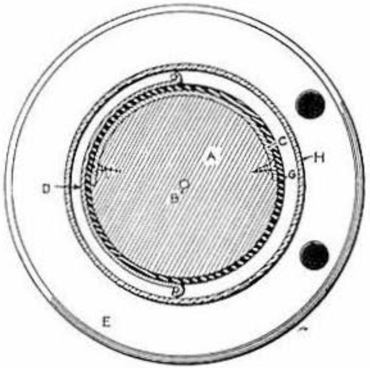

FIG. 2 CROSS SECTION

The base, E, and cap, F, are turned out of hardwood, to the shape shown, with grooves to admit the covering cylinder, H. This is made of aluminum, but it might as well be made of a sheet of black fibre, the seam being a lap joint. The reason a metal case was chosen was because it is more durable than fibre. It should have a tight fit in the grooves, as it is the only thing holding the base and cap together.

A large knob, I, is screwed on the shaft, B, and, beneath it, an aluminum index, J, is clamped with a hexagon nut.

The connections to the binding-posts are made as follows:

A wire, K, is caught under one of the screws holding the plate, C, to the cylinder, A. This wire is passed through a groove in the cylinder, as shown, to a washer, L, to which it is soldered. This washer makes contact with another washer, M, to which a wire is soldered. This wire is passed through a groove underneath the base to one of the binding posts. The other binding post is connected to one of the ears of the stationary plate, D, by a wire, running through another groove underneath the base.

Mention was made above of the cylinder, H, being of metal. The reader may have, at first thought, believed that this was not good, inasmuch as a short-circuit could accidentally be made. If the drawings are carefully observed, however, it will be seen that contact *might* be made between it and the plate, D, but a contact with the plate, C, is absolutely impossible.

A scale, H, for the index, J, to indicate the comparative capacity is made from a sheet of German silver and screwed to the cap, F. In my case the numbers and lines were etched upon it as follows:

It was covered with a thin coating of paraffin wax, and the characters scratched on it, so as to leave the surface of the German silver bare at those points. Then it was dipped in a bath of dilute nitric acid and left there for a while. When taken out it was washed in water and then dipped in boiling water to melt off the wax. The engraved surface was then painted over with asphaltum varnish, and when the latter had dried, it was washed off with a rag wet with gasoline. The varnish in the grooves made by the engraving remained, so that a neat and attractive, as well as durable, scale was the result. It will be noticed that German silver was used, instead of aluminum. This is because aluminum cannot be etched easily, as nitric acid has no effect upon it.

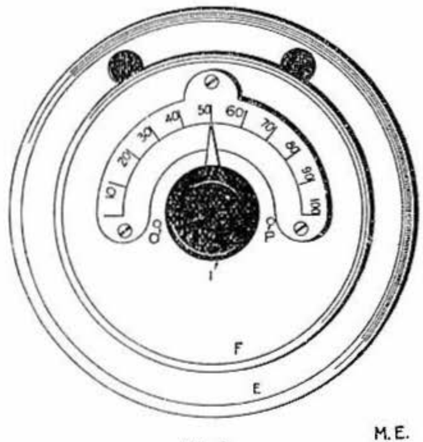

FIG. 3

Two pins, O and P, are inserted in the cap, F, at the points shown, to prevent the knob from being turned more than 180 degrees.

As the condenser may be operated at any angle, a small spring, Q, was inserted between two washers at the place shown, to keep the washers, L and M, in contact.

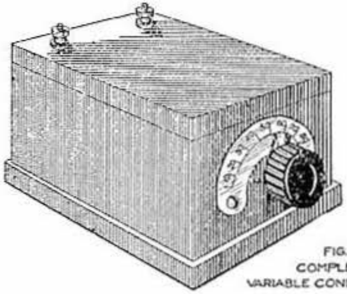

FIG. 4
COMPLETE
VARIABLE CONDENSER

From the description above, it would not seem that a condenser of this type could have a high capacity in a small space, as there are only about twenty-four square inches of active surface with a cylinder three and one-quarter inches in diameter and four and three-quarters inches long, the size which I adopted.

However, it must be borne in mind that the two plates, C and D, may be very close to each other, by making G very thin. In my case it is only 0.001 in. thick, resulting in a capacity of about 0.02 mfd.! The total space occupied by the condenser is seven and one-quarter inches by six inches by six inches.

The results were so gratifying that a smaller one was made, having a capacity of 0.005 mfd., and occupying a space of less than three and three-quarter inches by three inches by three inches (Fig. 4)! This also shows a different form of case.

Although not absolutely necessary, I found it to greatly increase the efficiency of the condenser if all the sharp edges of the plates, etc., were rounded with a file.

Contributed by

P. MERTZ.

SIMPLE POULSEN TICKER.

The present is a diagram of a Poulsen Ticker, which, I think, will be of interest to some amateurs. It is not necessary to have a telegraph relay, as shown in the February issue, in the Experimental Department. The fact that a relay was required

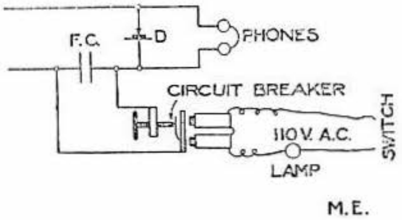

would prevent some from attempting to construct the ticker.

Instead of a relay I used a buzzer. Disconnect the magnets from the contact points and connect them as in the relay. Now connect to the contact points just as you would connect the relay contacts. You now have a ticker that is just as good as the one constructed from the relay, and it costs very much less.

Contributed by

JAS. LEROY HODGES.

A SEMIVARIABLE CONDENSER.

The following described variable condenser, which is built in sections, is very useful when one of the rotary, or plate type, cannot be obtained; or it may be used as an auxiliary. The condenser is of paraffined paper and tin foil, built in sections of whatever capacity is desired, and the method of construction is shown in sketch.

Procure a piece of one-quarter inch rod, seven inches long. Also a piece of five-sixteenths inch (outside) tubing, three inches long, to fit the rod. Tap the rod at both ends for an eight-thirty-two screw. Also procure two brass pillars, one-half inch high, five-eighths inch thick, with holes for an eight-thirty-two screw, an electrose knob, and some thin sheet brass.

Cut up the sheet brass into pieces and bore holes in them, as shown in Fig. 2. Cut as many as there are sections of the condenser. Then, assemble on the condenser case, or receiving cabinet, as shown in Fig. 1. Bend the brass contacts into the shape shown and fasten to the wood with small screws, connecting a section of the condenser to each one. Bend them up until each one nearly touches the one-quarter inch rod. Now, when the five-sixteenths inch tube slides over the rod it should touch all contacts. By pushing back and forth the number of sections, and hence the capacity, may be varied. The one-quarter inch rod is fastened to the board by eight-thirty-two screws from below, through the pillar, or by wood screws passing down from the top. A flat-headed eight-thirty-two screw is soldered to the tube and the electrose knob placed on it.

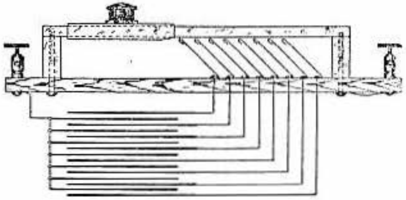

FIG. 1

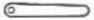

FIG. 2 M.E.

If one side of each section be one sheet of tinfoil, having ten square inches of surface, and using paraffined linen paper, the capacity of each will be approximately .0001 mfd. If desired a pointer or hand from a clock may be fastened under the electrose knob, and a scale made, showing the capacity in circuit.

This condenser may be used wherever a variable condenser, or a fixed condenser of several capacities, is needed. It may be shunted across one of the rotary or slide plate type. Rough adjustment for capacity is made on this while finer adjustment is obtained by the plate type condenser. Sharp tuning is thus made possible.

Contributed by

HENRY L. DAHM.

COMBINATION DETECTOR.

Enclosed please find a blue-print of a detector stand which I have worked out. This detector has a new idea as to the fine adjustment. Instead of having the fine adjustment screw between the contact and the standard, I have placed it on the other side of the contact, and therefore obtain a much finer adjustment. If you turn the screw down one-sixteenth of an inch the contact point will only go down one-thirty-second of an inch or less, where, as in the other kinds, the points would go down farther than the screw.

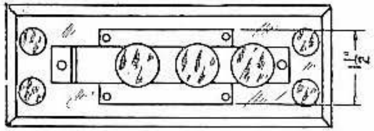

PLAN

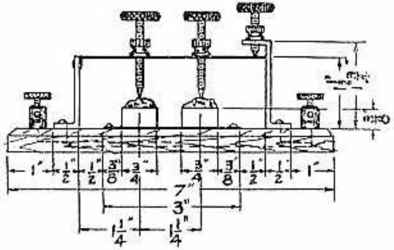

SIDE VIEW

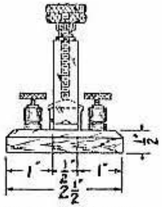

M. E. END VIEW

Another scheme is in the metal plate, on which the mineral cups rest. The plate is connected to a binding post, and the cups just rest on the plate. They can be moved around until the right part of the mineral is under the contact point.

I do not know whether you could use two minerals at the same time, but you can make a quick change from one to the other in this model. Some of the dimensions are not needed, but I think the drawing is clear.

Contributed by

ROBERT GIBSON.

INCREASED RECEIVING EFFICIENCY.

In most books dealing with wireless subjects, very little is said about three-slide tuners, so a word here would not be out of place.

Experimenters using single or double slide tuning coils will find that the selectivity of their station is increased from 25% to 50% by using a three-slide tuner in conjunction with a loop aerial, connecting instruments up as in Fig. 1. Excellent results can also be had from Fig. 2.

Any single or double slide coil can be converted into a triple slide coil at a cost of from twenty-five to fifty cents. After this has been done, get a piece of sheet copper, about 1-64th inch thick, and fasten it to the body of the coil, directly under one outside slider, so that when the slider is moved off the wire, it will slide onto the copper. Then connect this piece of copper to the other outside slider. Using the hookup shown in Fig. 2, if the slider is moved from the wire to the sheet copper, the loop aerial is converted to a straightaway aerial.

The writer believes that the method of connecting 'phones and detector (in series instead of in parallel), which was shown to him by a United Wireless operator, is new to most amateurs;

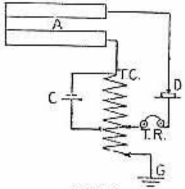

FIG. 1

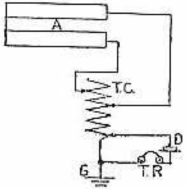

FIG. 2

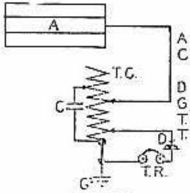

FIG. 3

A AERIAL
C CONDENSER (FIXED OR VARIABLE)
D CRYSTAL DETECTOR
G GROUND
T.C. TUNING COIL
T.R. TELEPHONE RECEIVERS.

M.E.

series instead of in parallel), which was shown to him by a United Wireless operator, is new to most amateurs; and, although it does not seem possible for the high frequency oscillations received by the aerial to go through the 'phones, whose inductance is very high, still such is the case.* Using crystal detectors, without a battery, such as perikon, galena, molybdenite, ferron, carborundum, etc., the signals come in fully 25% louder than when the 'phones are in parallel with the detector. When using a battery, the 'phones must necessarily be connected in parallel, but otherwise all connections are the same.

In case the experimenter has no loop aerial, and does not wish to convert his straightaway aerial into one of the loop type (which would amply repay him for the trouble and expense), the hook up shown in Fig. 3 gives very good results.

All of these "hook ups" were given several trials at a commercial station, as well as at the writer's station, and gave most gratifying results.

Contributed by
EDWARD B. WOOD.

**Note.*—The high frequency current does not pass through the 'phones, as stated. What does go through, in this case, is a rectified pulsating current. If the 'phones be shunted by a small fixed condenser, of about 0.002 mfd., still better results will be obtained. —Ed.

FASTENING COILS OF LOOSE COUPLERS.

The fastening of the primary and secondary coils to the wooden supports seems to trouble almost everyone who attempts to make a loose coupler. The following is a simple yet substantial method of overcoming this difficulty.

Cut out of a piece of one-half inch wood two circular pieces, one which will fit in the inside of the primary tube and the other in the inside of the secondary tube. This can be easily done with a knife or chisel. The primary can now be laid aside, as it is ready to fasten to the tube and also to the supports. The method of fastening is shown in Fig. 2.

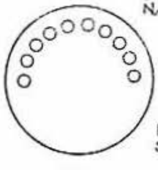

FIG. 1

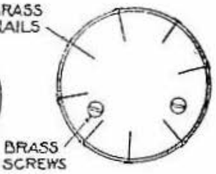

FIG. 2

M.E

Bore in the secondary as many holes as you are to have switch points on the secondary. If a slider is used on the secondary, it will not be necessary to bore any holes. Fig. 1 explains how the holes should be bored. They should be just wide enough for wire to pass through, as the wires from the various sections of the secondary pass through them to the switch points.

The switch points on the secondary support should be fastened so that the nuts will be below the surface of the wood, in order to allow the tube to be fastened on.

Brass nails and screws should be used as steel screws, or nails will become magnetized.

I have used this method with great success and found it to be much stronger than gluing the tubes in a groove.

Contributed by
ELMER S. LINDMARK.

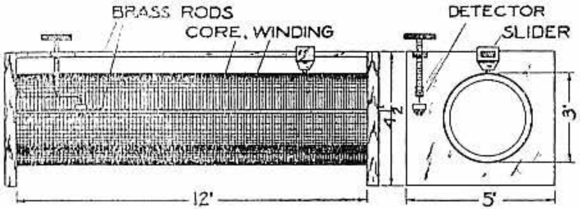

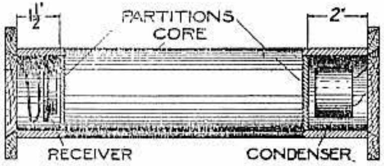

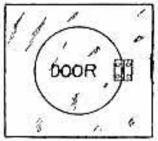

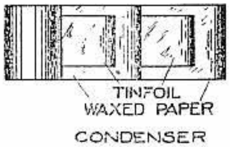

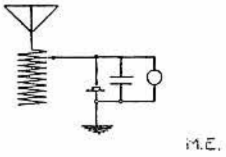

COMPACT PORTABLE RECEIVING SET.

The accompanying diagram shows a compact portable wireless receiving outfit.

The core of the tuning coil contains the condenser and is arranged to con- running from end to end of the coil serve as a support for the detector. The upper rod is tapped for the adjustment screw. The lower rod is tapped to receive a brass cup, containing the mineral. It is well to have several cups, each containing a different mineral. To change minerals, unscrew the cup and replace it with another.

The condenser is made by rolling long strips of waxed paper and tinfoil tightly around a wooden core. After it is rolled, tie it up and boil in paraffine. It is then placed in the tuning coil core and paraffine poured in around it.

A long piece of No. 18 annunciator wire will do for the aerial. It can be attached to a large box kite to raise it in the air. When not in use it can be wound on a wire spool.

Contributed by
N. CLINTON YOUNGSTROM.

IMPROVED FORMS OF DETECTORS.

An improved form of detector, that is more convenient, and at the same time much simpler than the so-called "sensitive" detectors, which contain a mutiplicity of springs and hard-rubber knobs, and adjusting screws, is shown in the illustrations herewith.

It consists of a double binding-post, A, through the upper connection of which is passed a shaft, B. On one side of this shaft is fastened a hard-rubber adjusting knob, D. On the other side is soldered a length of spring-brass wire, C, which is wound a few turns spirally, as shown in the drawing of the perikon detector; or

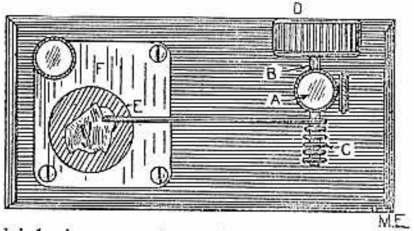

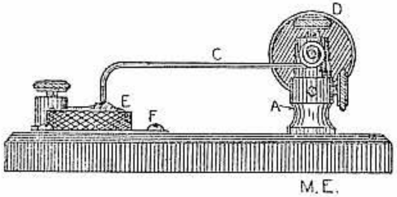

helically, as shown in that of the galena detector. The end of the wire is then brought out and bent as required.

For the silicon detector it is bluntly pointed at the end. For the galena, a small length of No. 30 copper wire is soldered, as shown. For the perikon detector, the end of the wire is wound around a piece of bornite, or chalcopyrite, mounted in metal, in the manner about to be described.

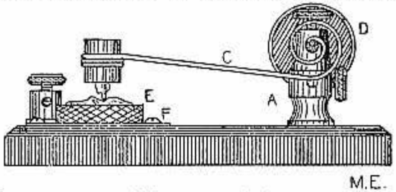

The lower mineral, E, is mounted in an alloy made of 3 parts solder to 1 part mercury. This is moulded into a "button," in a mold made by boring an inch hole through a piece of ¼ inch wood, and screwing this over another flat piece of wood. The crystals are set in while the metal is yet liquid, and, when it has thoroughly hardened, the edges of the metal button are knurled by rolling under a coarse file.

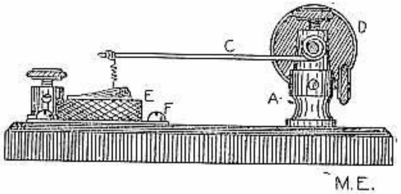

This button rests on a piece of sheet brass, or zinc, F, which is screwed to the base, and has a binding-post screwed over one corner.

The advantage of mounting the mineral this way, over mounting it in a cup, is clearly evident. Any part of the mineral can be reached quickly and conveniently, and a sensitive spot, when found, can be kept, because the point does not spring away.

The writer has found this type of detector far more convenient and sensitive than the so-called "sensitive" detectors, spoken of before, besides being much more easily and cheaply constructed.

Contributed by

P. MERTZ.

WIRELESS RECEIVING TUNER.

This is a very compact, loose coupled tuner, composed of two coils wound on hollow cores bent in the form of semicircles. The secondary coil is arranged to telescope into the primary, by turning a handle, which causes the secondary to swing round a central pivot. The primary is provided with a sliding contact, mounted on the curved rod, as shown. The diagram shows the tuner connected through a loading coil and with an ordinary detector circuit.

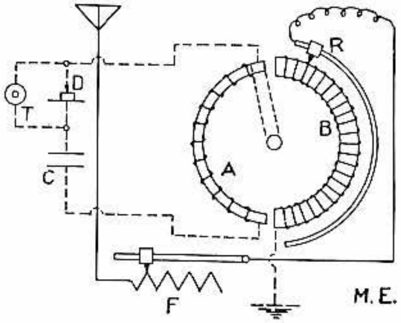

A NEW INDUCTIVE TUNER.

If carefully made, I believe the instrument here described will quickly prove its superiority over loose couplers of the sliding tube type, and, as an interference preventer, it is in a class by itself.

The wire is wound not unlike the ribbon of a pancake helix. Great care should be taken not to let the wire overlap. This may be accomplished by using a small amount of thick shellac which also adds to the insulation.

The drawings are practically self-explanatory, except for the taps. Taps should be taken out about every five feet on the primary, and every 30 feet on the secondary. Multiple point switches are placed on the backs of the coils, and the taps led to the switch points.

In Fig. 1, A and A¹, are blocks of wood 1 x 1 x 4 inches, screwed to the base, and which hold the ¼ inch brass rods, on which the secondary slides. The primary is screwed to A. B is a block 1 x 1 x 1 inch, screwed to the secondary, with a hole large enough to easily admit a ¼ inch rod drilled half way through it. B is a block 1 x 1 x 1 way through it. B is a block 1 x 1 x 1 inch, screwed to the secondary and has a ¼ inch hole drilled through it. C is a ¼ inch round brass rod 6½ inches long, upon which the secondary is mounted. D is a hardwood board 1 x 2 x 4 inches. E are four blocks 1 x 1 x 1½ inches, screwed to D, and have holes drilled through them to admit the ¼ inch rods, F. Binding posts, and the necessary switches, may be mounted on the base in the manner most convenient.

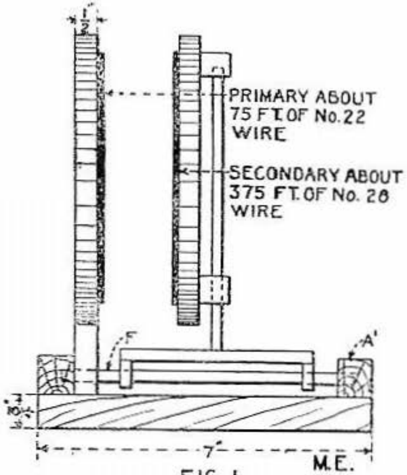
PRIMARY ABOUT 75 FT. OF No. 22 WIRE
SECONDARY ABOUT 375 FT. OF No. 28 WIRE
F
A'
7"
M.E.

FIG. 1

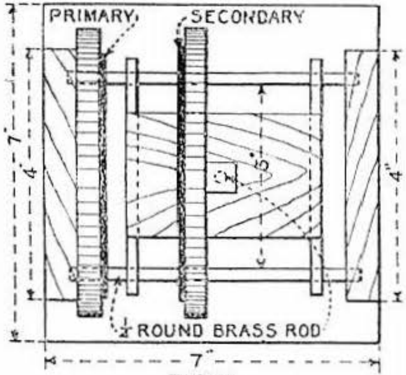

FIG. 1A

In the hook-up diagram, m is a 2-point switch, which is used to cut in the tuning coil, w, if a long wave length is desired; n is a 2-point switch allowing signals to enter either the center or outside of primary; x a 2-point switch allowing signals to be discharged through either the center or outside of secondary; s a 2-point switch to cut in the variable condenser if needed; l and k are the multiple point switches used to vary the amount of wire used on primary and secondary. The windings of the primary and sec-

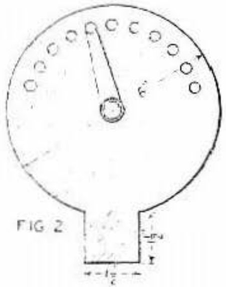

FIG. 2

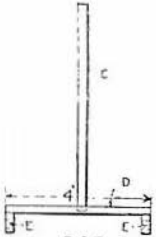

FIG. 3

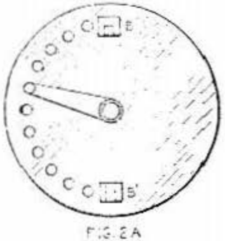

FIG. 2A

M.E.

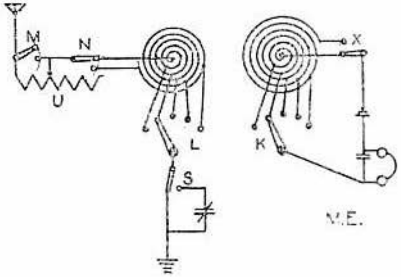

ondary should be made in opposite directions.

Contributed by

R. TREWEEKE.

A COMPACT RECEIVING SET.

I designed this receiving set with the view of taking up as little room as possible. It can be screwed to the wall or to a window sill.

The wooden core of the tuning coil is 3½ inches in diameter and 9 inches long and is wound with No. 24 B. & S. enameled wire. The end blocks are 4x4½x½ inches and are screwed to the core. The slider rods are screwed on top of the ends.

The detector, Fig. 2, is mounted on one end of the tuner and on the other end is the condenser.

The detector is made of brass strip, battery screws, a battery cup, and other odds and ends as shown in Fig. 2.

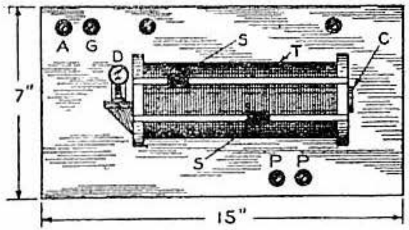

FIG. 1

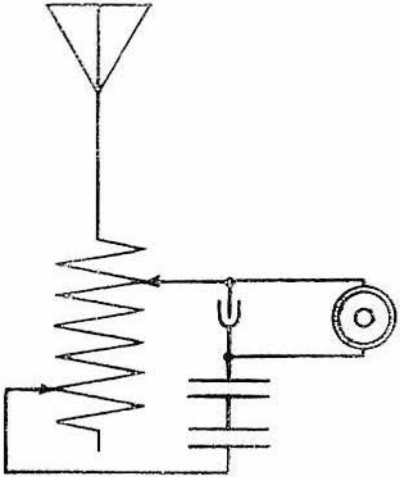

FIG. 3

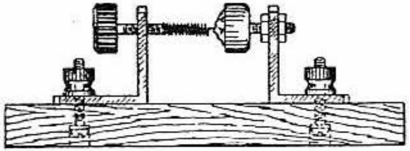

FIG. 2

The condenser is of the ordinary tinfoil and paper construction.

All the parts are fastened on the base, as in Fig. 1, which when screwed up lies flat against the wall.

The parts are connected as in Fig. 3.

Contributed by

ALEX. T. McCONE.

MODERN ELECTRICS

For a fascinating selection of the highest quality books for experimenters, inventors, tinkerers, mad scientists, and a very few normal people... visit

www.lindsaybks.com

ISBN 1-55918-327-6
90000
9 781559 183277

A field mill for measuring atmospheric electricity

by Frank Thompson F.Thompson@manchester.ac.uk

Introduction

Every time we walk outside, we are bathed in force fields that are produced by nature. They are gravitational, magnetic and electric fields.

In the on-line text *Motion Mountain* – Volume 3, on page 209 – Christopher Schiller makes the following statement about electric fields:

"…every person should know that there is a vertical electric field of between 100 and 300 V/m on a clear day, as discovered in 1752; the earth is permanently negatively charged and, on a clear day, current flows downwards (electrons flow upwards) trying to discharge our planet."

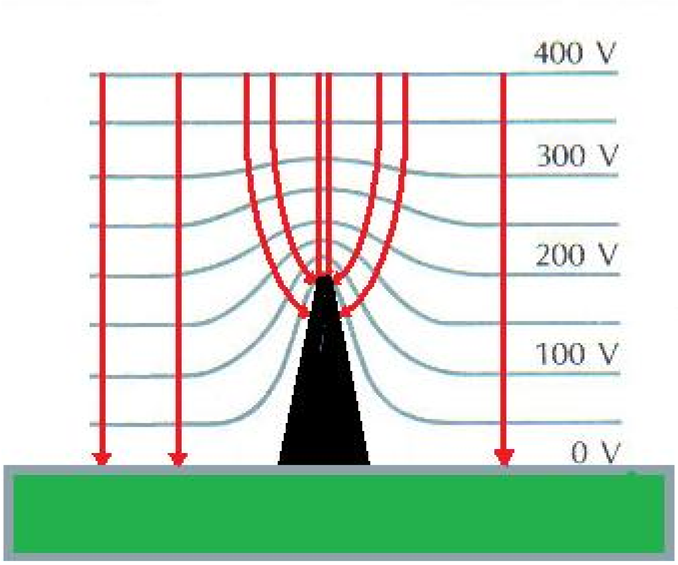

Figure 1 Distortion of field lines round a conducting object (grey - potential, red - field)

Now, we can all "feel" the gravitational force (just try jumping up and down) and we can easily detect magnetic fields with a compass needle. Indeed, these forces are very easily measured [1], [2]. The electric field, however, is more of a mystery and this is largely to do with the fact that it does not come indoors. This is a major obstacle to detection. An illustration is given in Figure 1:

Any conducting object at ground level will distort both potential and field lines of an electric field as shown. Houses, trees and people all have relatively good conductivity when compared to the surrounding air and therefore electric fields will not be present *inside* buildings at ground level.

It was taken as a challenge to measure this field and in doing so it was hoped that students would be made more aware of the vast topic of Atmospheric Physics.

It may be worthwhile visiting "University of Reading METFiDAS weather observations". On this web page, note that PG (potential gradient) denotes electric field.

Graphical data, listed under the heading "Atmospheric Electricity" at the bottom of the web page, shows the electric field values for the particular location of Reading, UK.

Figure 2 - Diagram of a typical Field Mill with electronic circuitry

Apparatus

A little background reading showed that a Field Mill was the most common arrangement for this field measurement and a typical layout is shown in Figure 2:

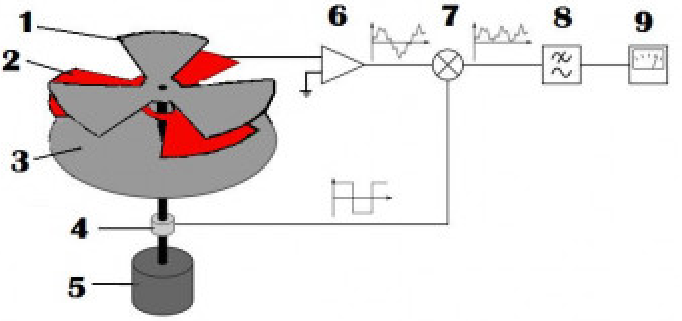

The key for Figure 2 is as follows:
1 - **Rotor blade** at earth potential, turned by the **motor** (item **5**)
2 - **Fixed plate** isolated from earth (this is also called the **sensor plate**)
3 - **Ground plate**, also held at earth potential
4 - **Tachometer** with three equally spaced black and white partitions for its optical switch
5 - **Motor** providing power to the **rotor blade** (item **1**)
6 - **Charge amplifier**
7 - **Mixing device** and in the present apparatus it is a **multiplier circuit** AD 633
8 - **Low pass filter**
9 - **Centre-zero meter**

Figure 3 - Photograph showing the motor plus tachometer

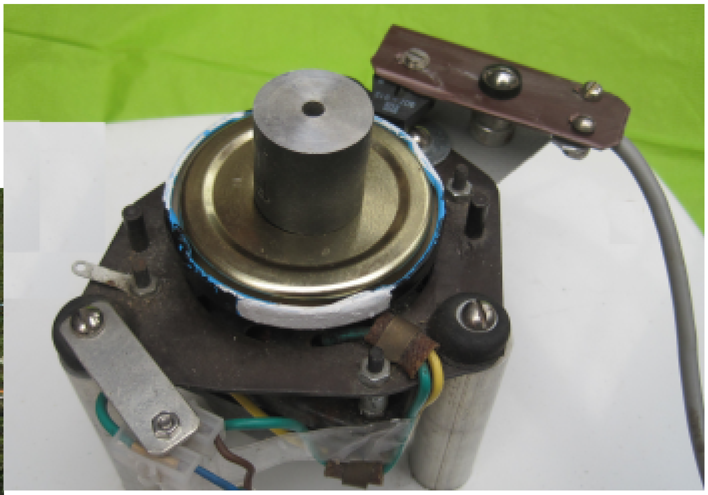

In the present work, item **5** is a small motor with a disc attached to the spindle as shown in Figure 3:

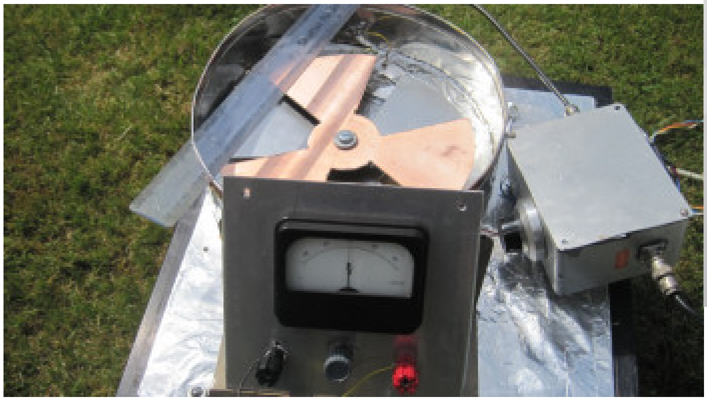

Figure 4 - Field Mill with separate charge amplifier (item 6) and centre-zero meter (item 9) resting on the case (ruler indicates the size of the apparatus)

Three black and three white stripes were painted on the rim so that the reflective optical-switch (RS Components Ltd, RS 307-913 or Optek OPB704W-Z supplied by Amazon) gave a square wave output signal.

The motor also turns the rotor blade (item **1**).

The ground plate (item **3)** is fabricated from a steel biscuit tin and the upper part of the spindle passes through.

The charge sensor plate (item **2**) is supported by this tin on insulated pillars and is alternately ***exposed to*** or ***shielded from*** the incident field by the rotor blade plate (item **1**).

The instrument/apparatus, positioned at gound level, is shown in Figure 4. Both the stationary and rotating blades are enclosed in a biscuit tin so allowing the instrument to be zeroed when the tin lid is placed on the tin. It also provided electrical shielding as extraneous noise from the motor was quite high.

Figure 5 - Multiplier circuit (item 7) for mixing the tacho and charge signals (X and Y)

The charge amplifier, encased in a separate metallic box, rests on the case of the apparatus and is connected to the fixed plate (item **2**). If the signal from this amplifier were to be rectified then only the magnitude of the field would be measured and whether the field was incoming or outgoing would not be determined. But, by using ***a mixer*** (item **7**). we have synchronous demodulation which gives the sign of the field and a significant noise reduction. Commercial instruments are available for these measurements (e.g. Model 410 Scitech Instruments Ltd) and Figure 5 shows a low-cost alternative:

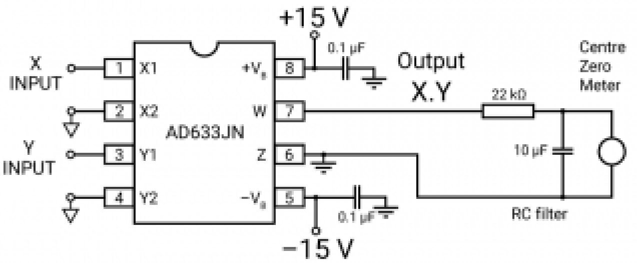

Basic multiplier connections

Figure 6 - Signals obtained when a charged PTFE rod is positioned over the Field Mill. (a) Signals from the tachometer (red) and the charge amplifier (blue)

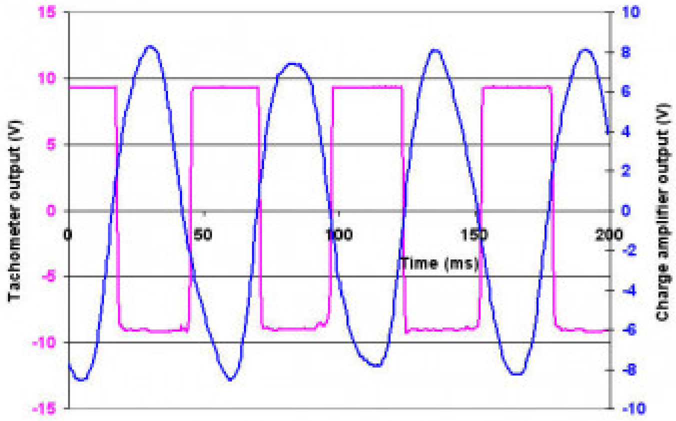

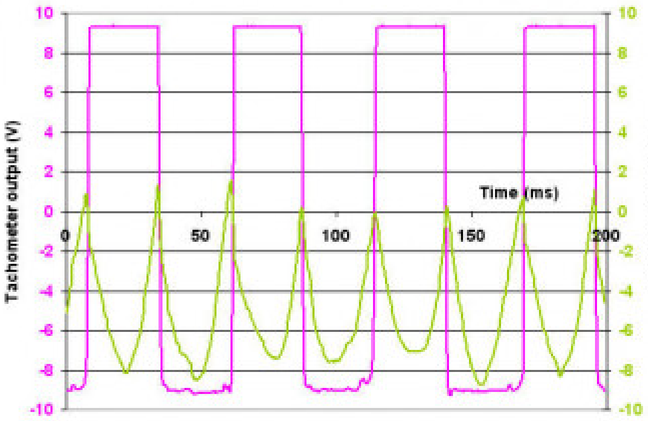

6(b) Signals from the tachometer (red) and the mixer (green)

In setting up the field mill one can place a charged object over the sensor plate (item 2) and observe the following signals. See Figure 6 (a), (b):

The output from the multiplier circuit is then passed through a low-pass filter and displayed on a centre-zero meter (item 9). The above readings (in Figure 6) were taken with a PicoScope 5242A oscilloscope and copied to Excel spreadsheets.

Preliminary results

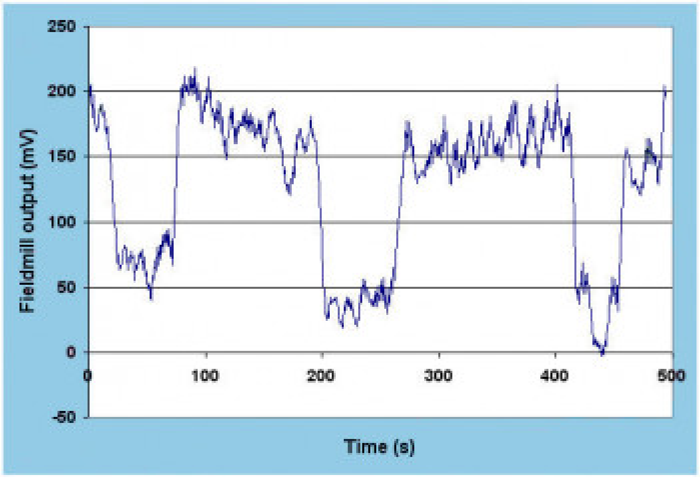

After the setting up, we can be confident that the output voltage from the filtered signal is directly proportional to the earth’s field and a typical, fine day, outdoor graph is given in Figure 7:

As a very approximate calibration (carried out at the meteorological department of Reading University) the output readings in mV could be converted directly into field readings in V/m. Thus, the field mill gives the earth’s electric field as 120 (±20) Vm^-1 as the covered reading of approximately 50 mV is subtracted from the uncovered reading of 170 mV. Further measurements were made with the field mill on 22 June 2017. A thunderstorm had been forecast so there was a possibility of variability in the earth's electric field.

Figure 8 - Earth's electric field changes as a thunderstorm was approaching (rotor covered for periods 30–70, 107–140, 230–260 and 350–380 seconds).

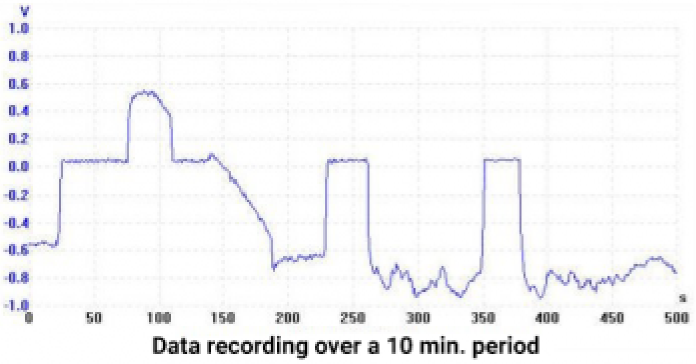

The apparatus was placed in an open space at about 8:15 am and the trace (Figure 8) was obtained. At approximately 8:30 am both the meter and the trace were off-scale, indicating a large negative signal. Adjustment of the gain by a factor of 10x restored an on-scale signal and, for several periods a zero signal condition was tested by placing the tin lid over the rotor. The changes in the signal were quite rapid as Figure 8 shows:

In Figure 8 we see that the earth's electric field changes from negative to positive in about 30 seconds and then the positive signal changes to negative in about 100 seconds.

The instrument had to be moved indoors after 600s as rainfall had commenced. Thunder and lightning occurred shortly afterwards. The electric field is seen to vary between +600 V/m and –1 kV/m.

Figure 9 - Trace showing the atmospheric electric field on a fine day

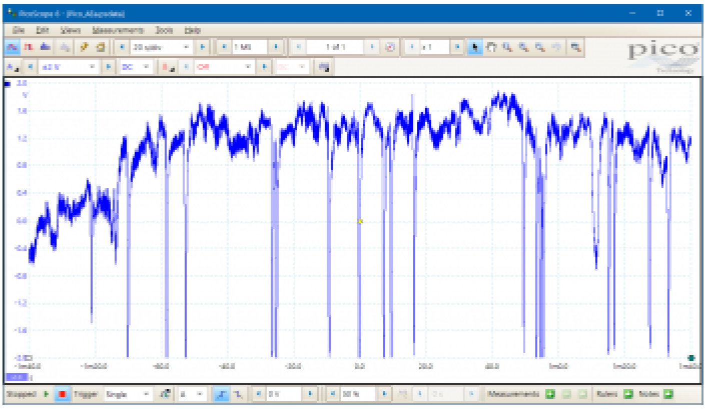

The PicoScope 5242A is ideal for data logging as the time scale can be set as long as 5000 seconds per division. Figure 9 shows a typical fine weather trace for about 5 mins:

For the initial minute the rotating blade was covered with the tin lid and then it was uncovered for the rest of the time. The observed spikes were assumed to be spurious but further measurements need to be carried out to see if they are caused by an instrumental defect.

Conclusions

On a fair-weather day, the Earth's electric field was found to be close to 120 V/m but the accuracy for these preliminary results was at best ±20%. This agrees with the Schiller statement. The offset error of approximately 50 mV is likely to result from imperfect blade dimensions and/or the black and white tachometer divisions as these were hand-painted.

The variability of the Earth's electric field (Reading University - graphical data showing potential gradient) does pose a problem so it is advisable to carry out measurements on a fine day and then the field will be close to 120 V/m.

As a final point, the signal-to-noise performance of the apparatus is very low even with blade diameters close to 20 cm. If one could replace the simple mixer with a phase sensitive detector (PSD) then improvements would be possible. The PSD would greatly simplify the setting up procedure of the apparatus as both gain and phase of the charge signal could then be adjusted.

US006246561B1

(12) United States Patent
Flynn

(10) Patent No.: US 6,246,561 B1
(45) Date of Patent: Jun. 12, 2001

(54) METHODS FOR CONTROLLING THE PATH OF MAGNETIC FLUX FROM A PERMANENT MAGNET AND DEVICES INCORPORATING THE SAME

(75) Inventor: **Charles J. Flynn**, Greenwood, MO (US)

(73) Assignee: **Magnetic Revolutions Limited, L.L.C**, St. Louis, MO (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/127,056**

(22) Filed: **Jul. 31, 1998**

(51) **Int. Cl.**^{7} **H02K 21/12**
(52) **U.S. Cl.** **361/147**; 361/210
(58) **Field of Search** 361/143, 147, 361/206, 210

(56) **References Cited**

U.S. PATENT DOCUMENTS

1,161,819	*	11/1915	Grob .	
3,135,880	*	6/1964	Olson et al. .	
3,184,651	*	5/1965	Albosta .	
3,451,055	*	6/1969	Pihl .	
3,634,735	*	1/1972	Komatsu	335/234
3,670,189		6/1972	Monroe	310/46
3,683,239	*	8/1972	Sturman	335/170
3,743,898	*	7/1973	Sturman	335/254
3,949,250		4/1976	Walker et al.	310/36
4,279,008	*	7/1981	Del Picchia et al.	361/194
4,479,162	*	10/1984	Offutt et al.	361/210
4,528,533		7/1985	Montagu	335/230
4,688,138	*	8/1987	Nagata et al.	361/154
4,747,010	*	5/1988	Bayer	361/210
4,965,695	*	10/1990	Baumann	361/142
5,015,903		5/1991	Hancock et al.	310/168
5,155,651	*	10/1992	Yoda et al.	361/144
5,254,925		10/1993	Flynn	318/696
5,268,662		12/1993	Uetsuhara et al.	335/229
5,304,881		4/1994	Flynn et al.	310/156
5,455,474		10/1995	Flynn	310/181
5,463,263		10/1995	Flynn	310/181
5,677,580		10/1997	Huang	310/44
5,710,493		1/1998	Flynn	318/254
5,719,543		2/1998	Berling	335/229
5,753,990		5/1998	Flynn et al.	310/156
5,818,680	*	10/1998	Schmitz et al.	361/160
5,822,167	*	10/1998	Schmitz	361/143

* cited by examiner

Primary Examiner—Fritz Fleming
(74) *Attorney, Agent, or Firm*—Haverstock, Garrett and Roberts LLP

(57) **ABSTRACT**

A permanent magnet device includes a permanent magnet having north and south pole faces with a first pole piece positioned adjacent one pole face thereof and a second pole piece positioned adjacent the other pole face thereof so as to create at least two potential magnetic flux paths. A first control coil is positioned along one flux path and a second control coil is positioned along the other flux path, each coil being connected to a control circuit for controlling the energization thereof. The control coils may be energized in a variety of ways to achieved desirable motive and static devices, including linear reciprocating devices, linear motion devices, rotary motion devices and power conversion.

42 Claims, 32 Drawing Sheets

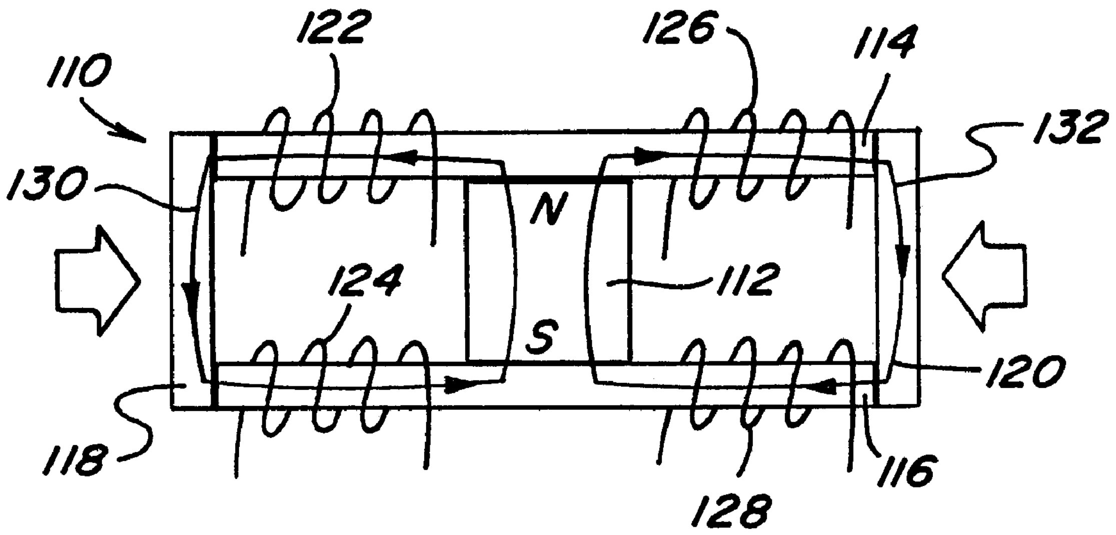

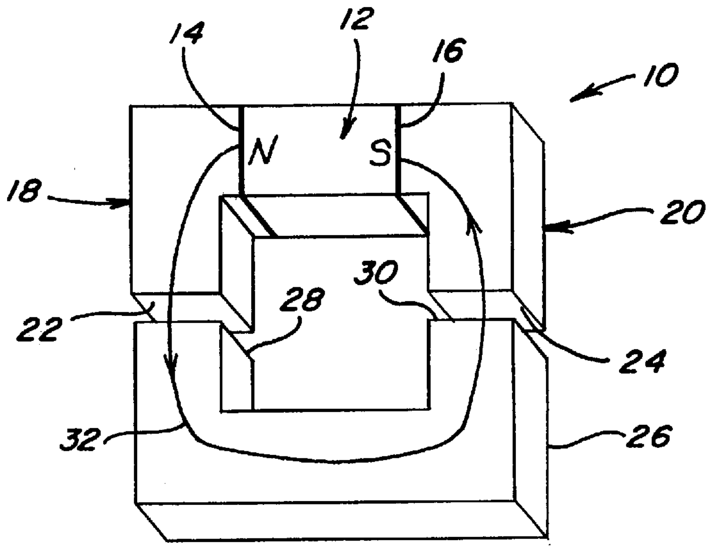
N
S

Fig. 1

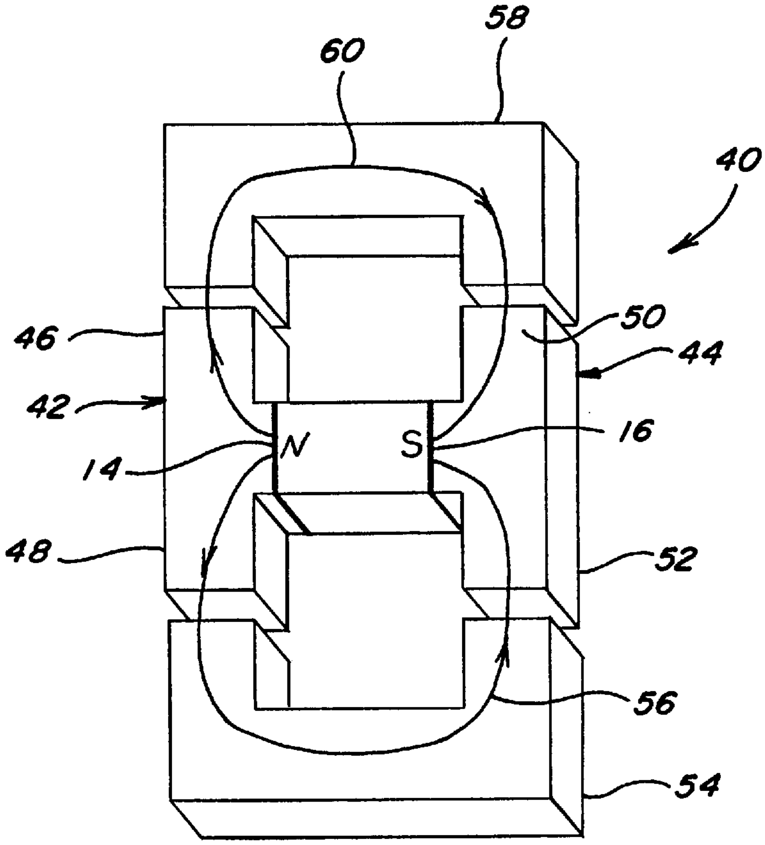
N
S

Fig. 2

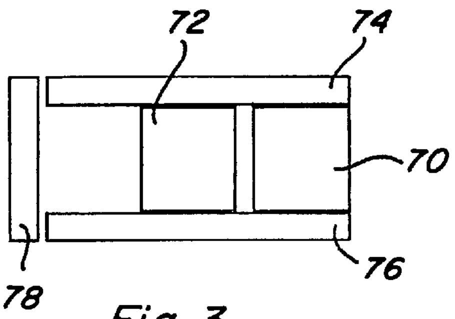
72
74
70
76
78

Fig. 3

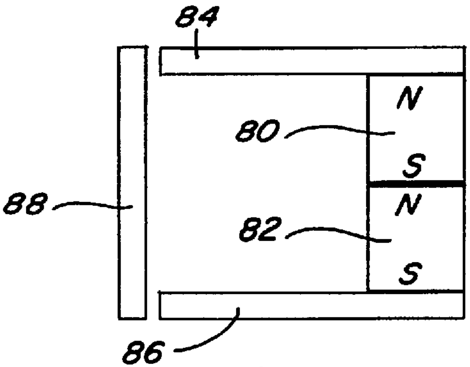
84
N
80
S
N
88
82
S
86

Fig. 4

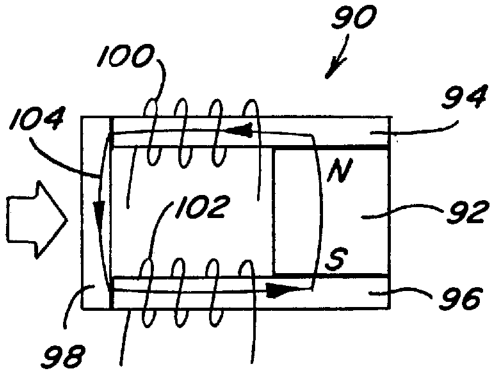
90
100
94
104
N
102
92
S
96
98

Fig. 5

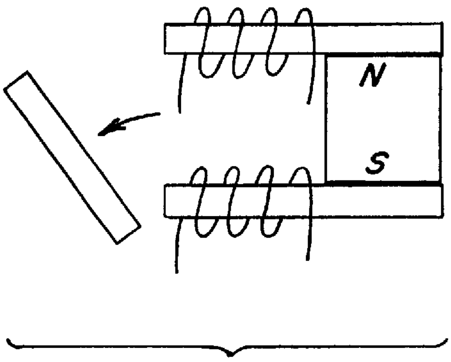

Fig. 6

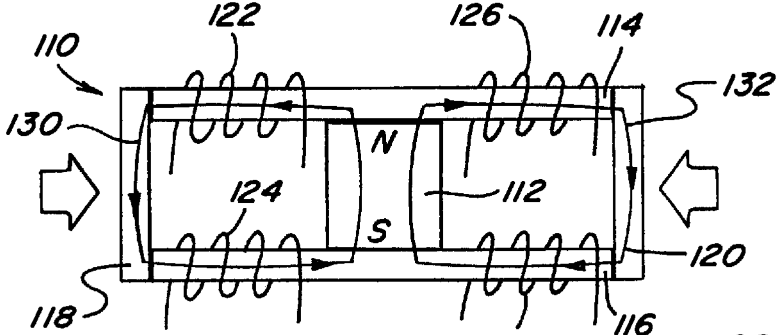
122
126
114
110
132
130
N
124
112
S
120
118
116
128

Fig. 7

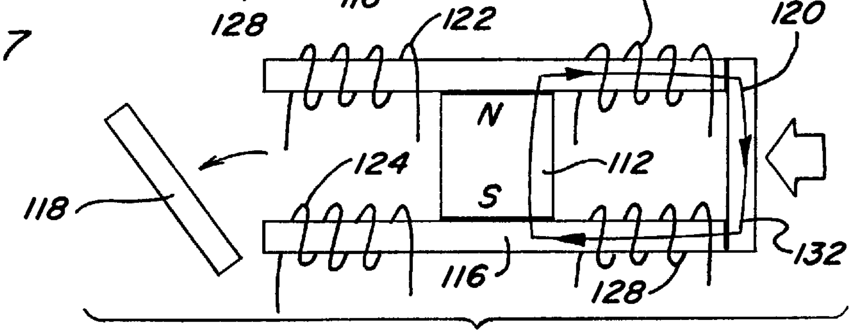
126
122
120
N
124
112
118
S
132
116
128

Fig. 8

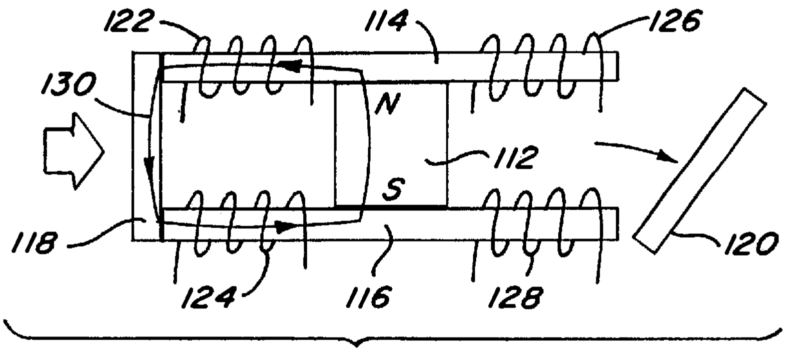
122
114
126
130
N
112
S
118
120
124
116
128

Fig. 9

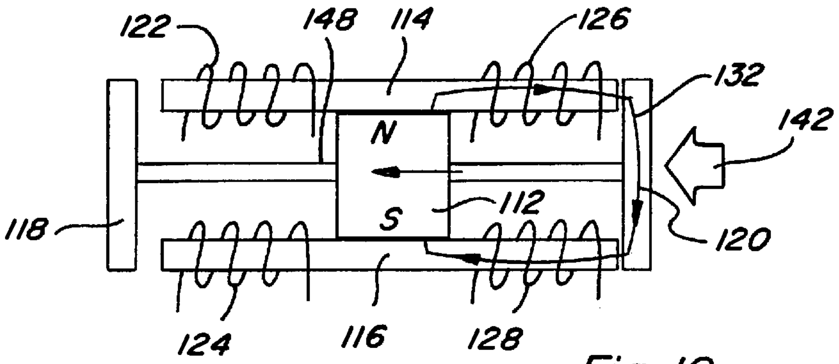
122
148
114
126
132
142
N
112
S
118
120
124
116
128

Fig. 10

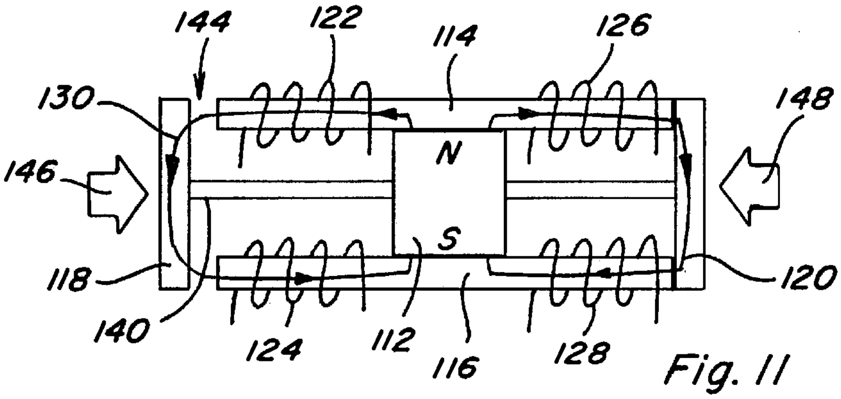
144
122
114
126
130
148
146
N
S
118
120
140
124
112
116
128

Fig. 11

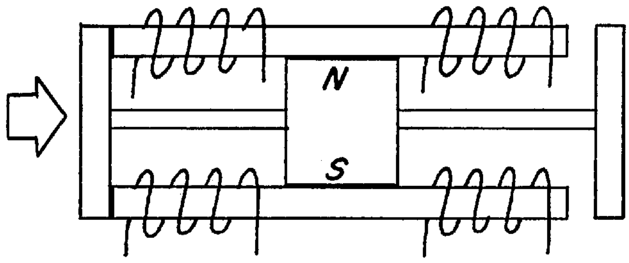

Fig. 12

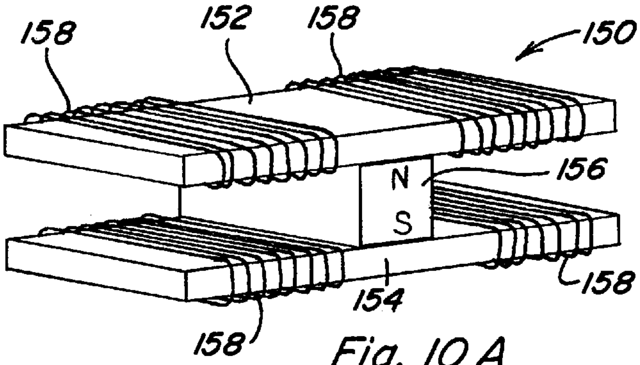
158
152
158
150
156
N
S
158
154
158

Fig. 10A

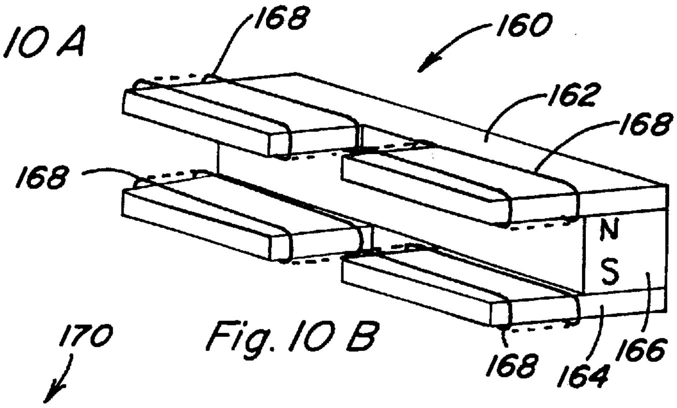
168
160
162
168
168
N
S
168
164
166

Fig. 10B

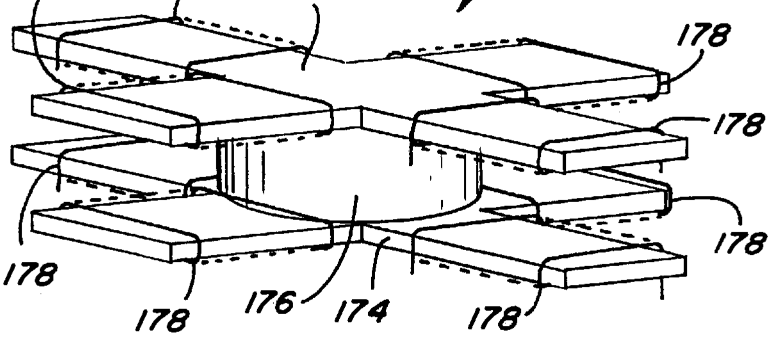
178
178
172
170
178
178
178
178
178
176
174
178

Fig. 10C

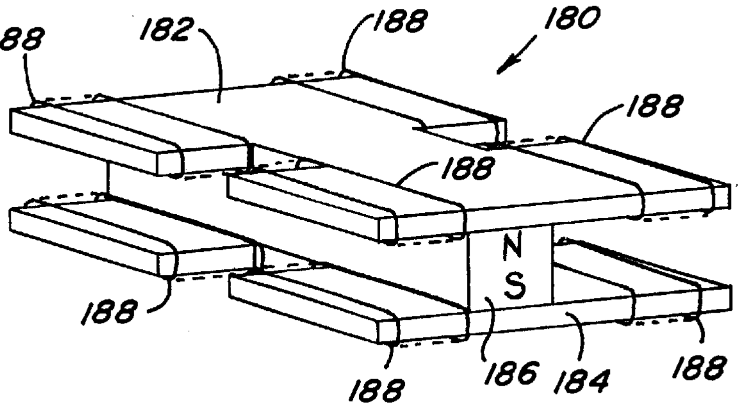
188
182
188
180
188
188
N
S
188
188
186
184
188

Fig. 10D

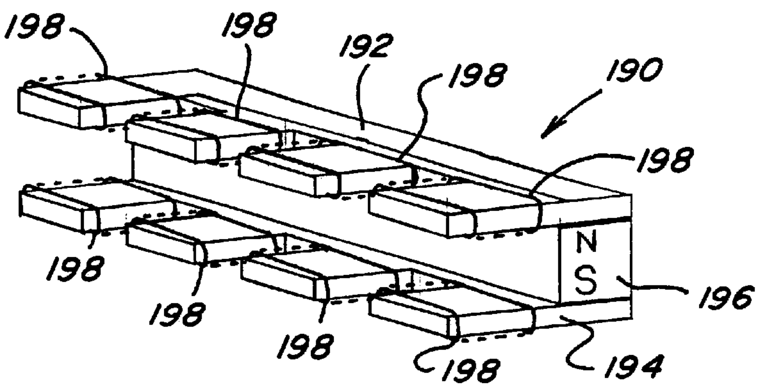
198
198
192
198
190
198
N
S
196
198
198
198
198
194

Fig. 10E

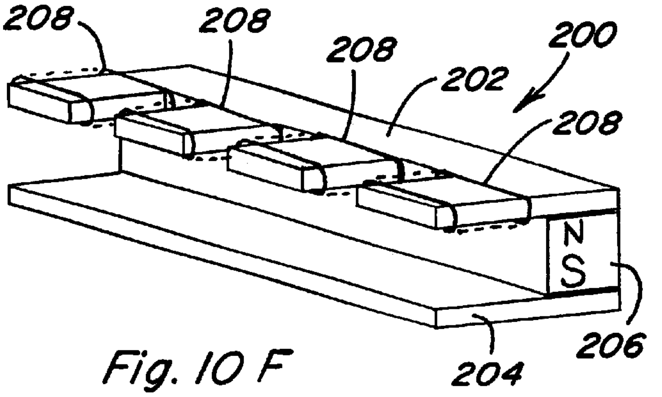
208
208
208
200
202
208
N
S
206
204

Fig. 10F

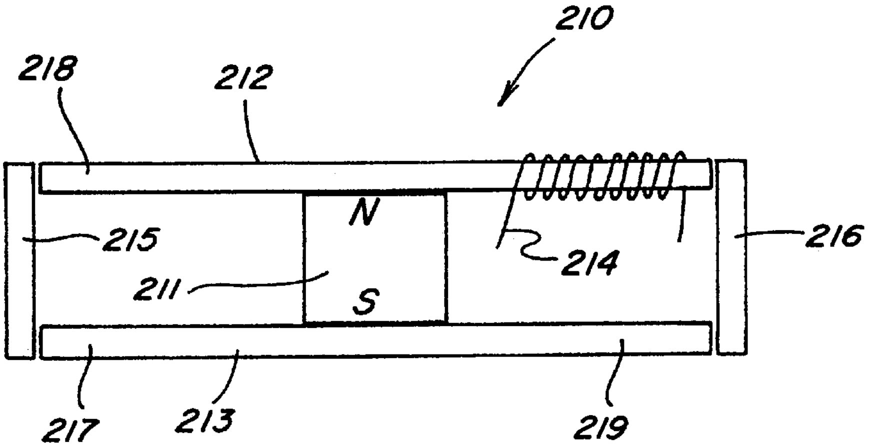
210
218
212
N
215
216
211
214
S
217
213
219

Fig. 10 G

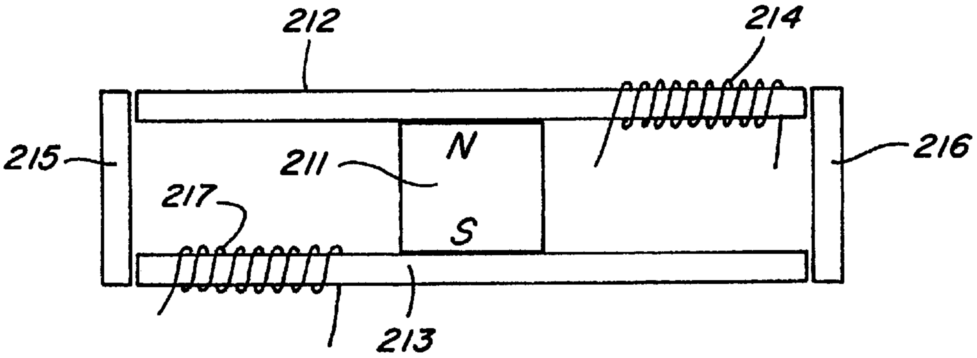
212
214
215
211
N
216
217
S
213

Fig. 10 H

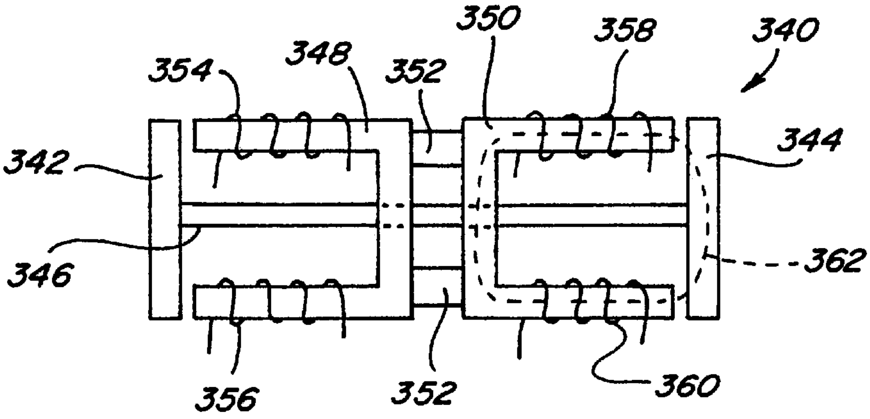
350
358
340
354
348
352
344
342
346
362
356
352
360

Fig. 14

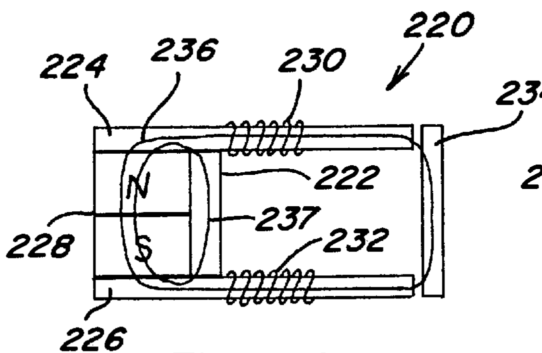
224
236
230
220
234
N
222
237
228
S
232
226

Fig. 11A

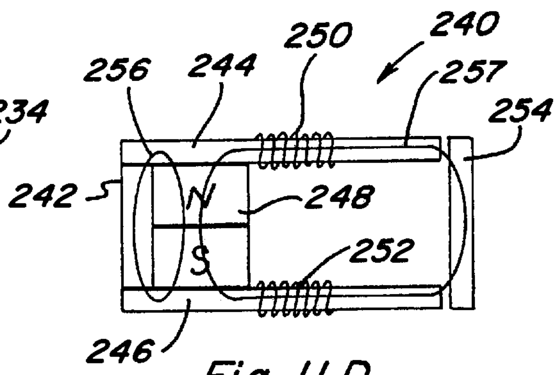
256
244
250
240
257
254
242
N
248
S
252
246

Fig. 11D

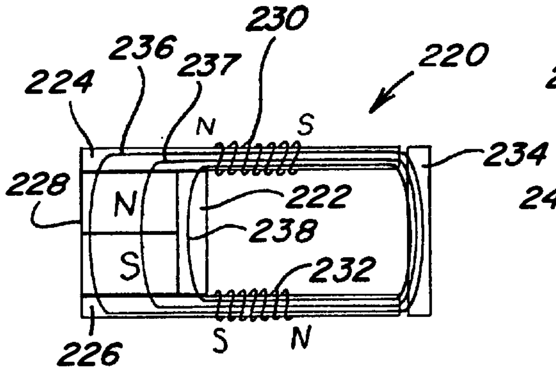
224
236
237
230
220
228
N
S
N
222
238
S
232
234
226
S
N

Fig. 11B

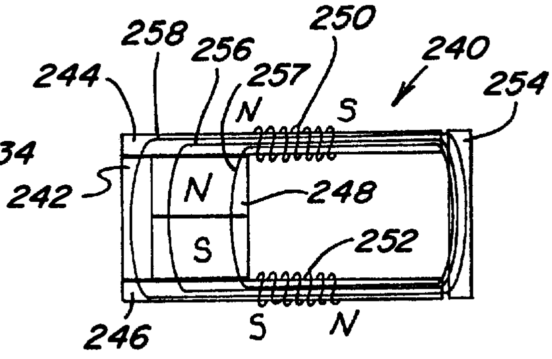
244
258
256
257
250
240
254
242
N
S
N
248
S
252
246
S
N

Fig. 11E

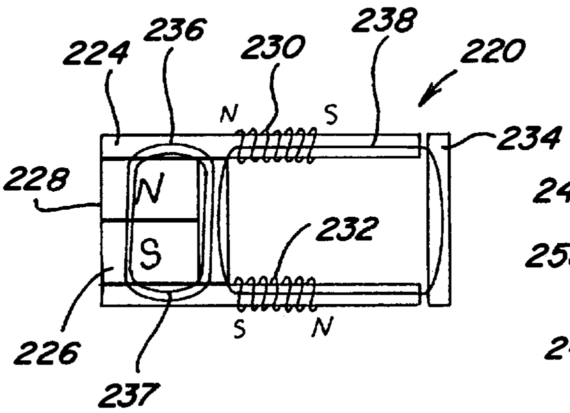
224
236
230
238
220
N
S
234
228
N
S
232
226
S
N
237

Fig. 11C

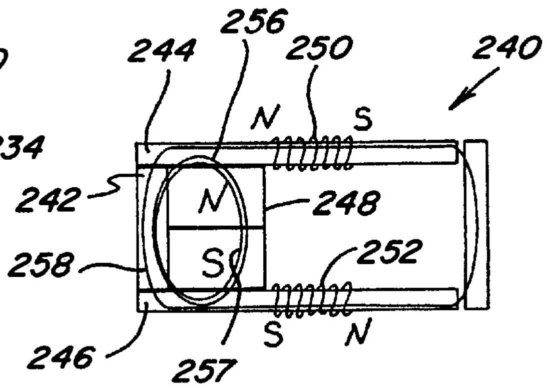
244
256
250
240
N
S
242
N
248
258
S
252
246
257
S
N

Fig. 11F

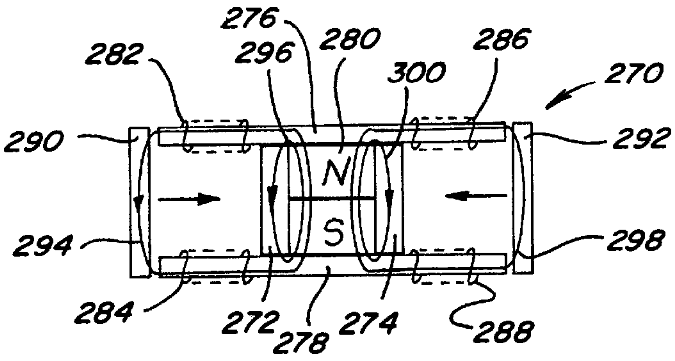
276
282
296
280
286
300
270
290
292
N
S
294
298
284
272
278
274
288

Fig. 12A

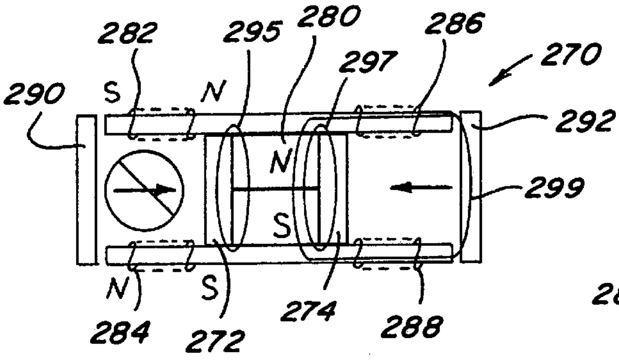
282
295
280
286
297
270
290
S
N
292
N
299
S
N
S
284
272
274
288

Fig. 12B

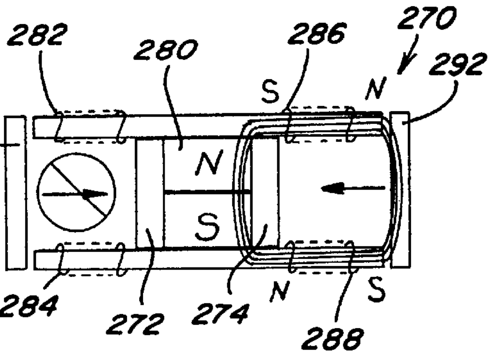
270
282
280
286
292
S
N
290
N
S
284
272
274
N
S
288

Fig. 12D

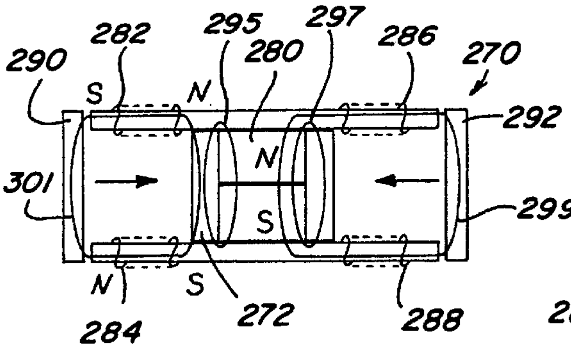
295
297
282
280
286
290
270
S
N
292
N
301
S
299
N
S
284
272
288

Fig. 12C

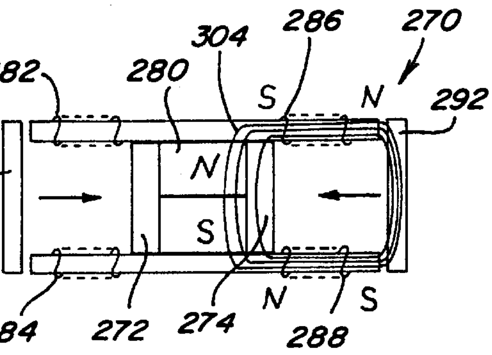
286
270
304
282
280
S
N
292
290
N
S
284
272
274
N
S
288

Fig. 12E

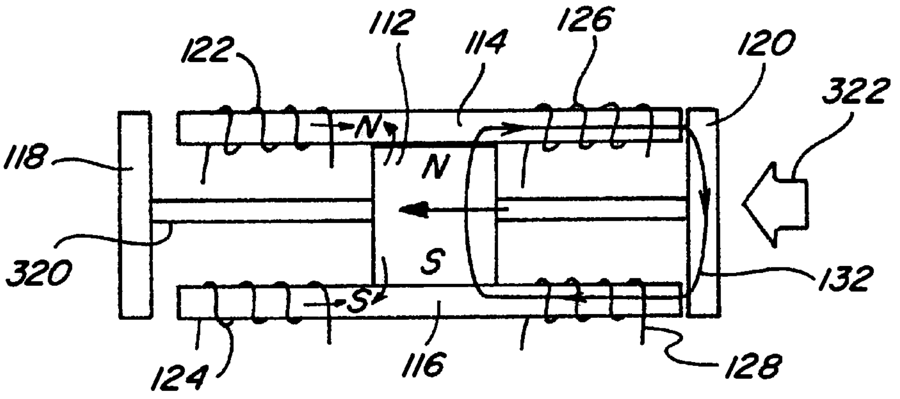
122
112
114
126
120
322
118
N
N
320
S
132
S
124
116
128

Fig. 13A

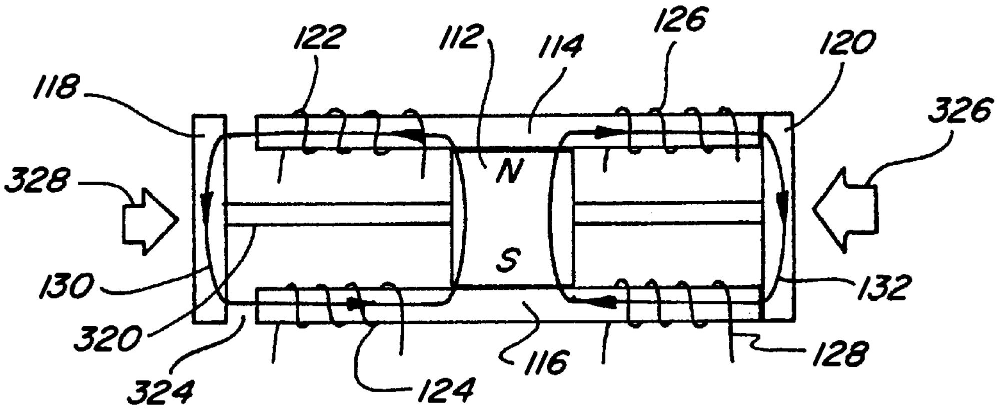
122
112
114
126
120
118
326
328
N
S
130
132
320
324
124
116
128

Fig. 13B

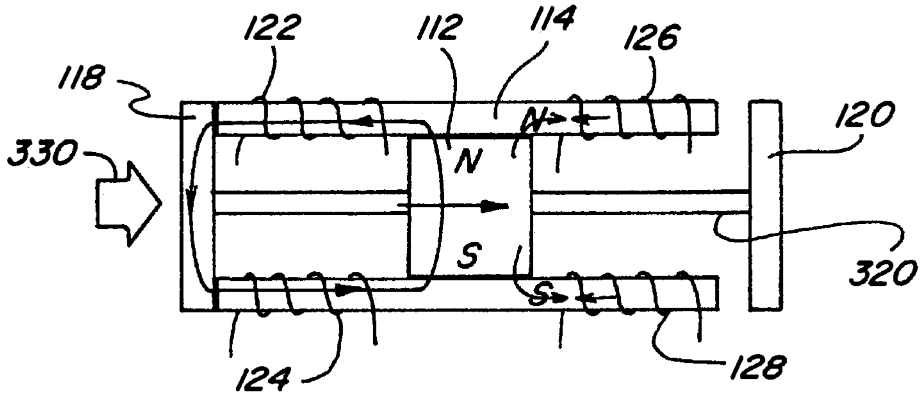
122
112
114
126
118
N
120
330
N
S
320
S
124
128

Fig. 13C

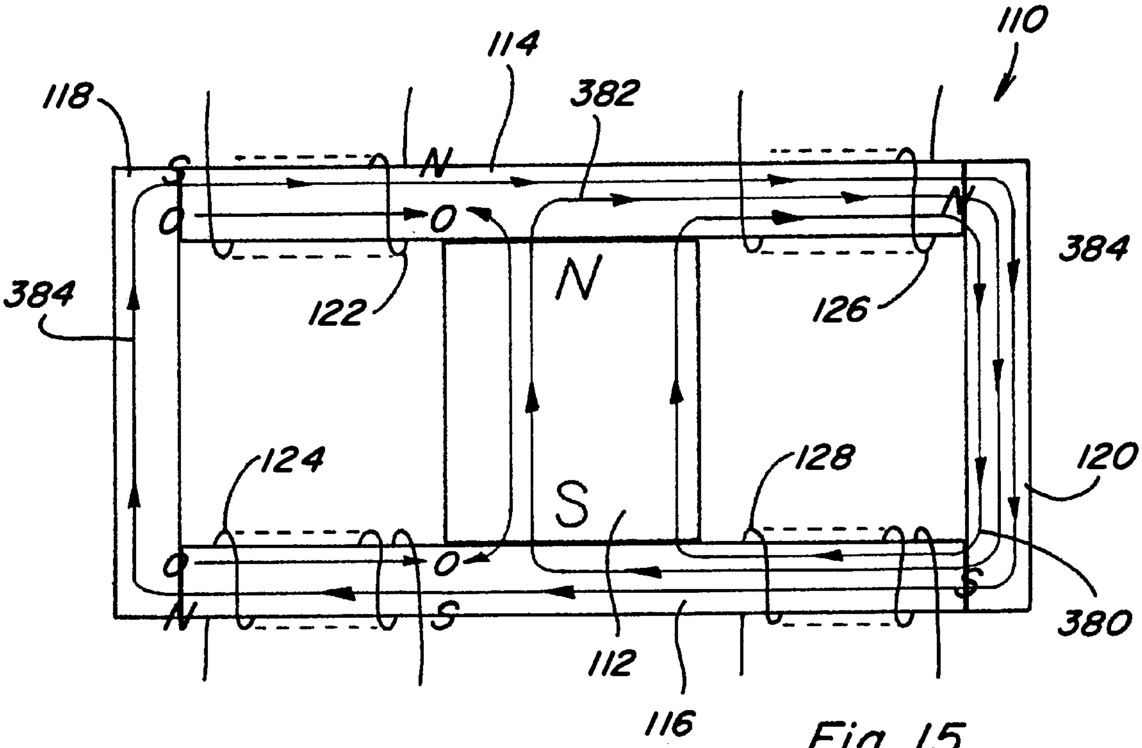
110
114
118
382
S
N
O
O
N
384
384
122
N
126
124
128
120
S
O
O
S
N
S
380
112
116

Fig. 15

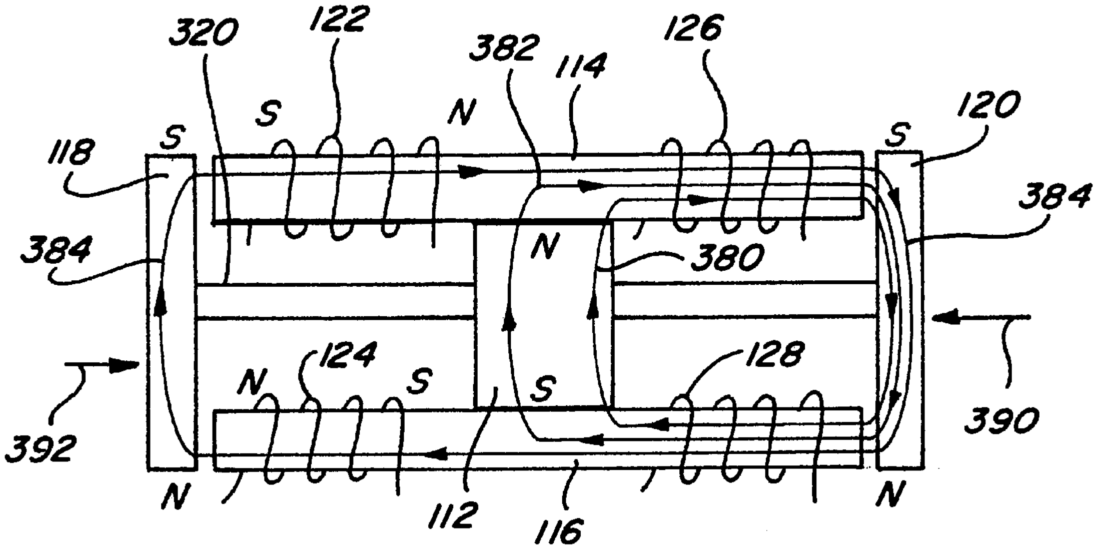
320
122
382
114
126
120
118
S
S
N
S
384
384
N
380
390
392
N
124
S
S
128
N
N
112
116

Fig. 16 A

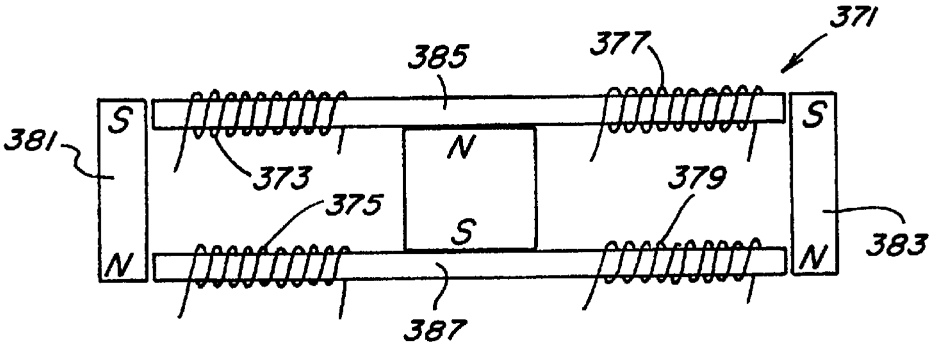
371
385
377
S
381
N
373
379
375
S
383
N
N
N
387

Fig. 16 B

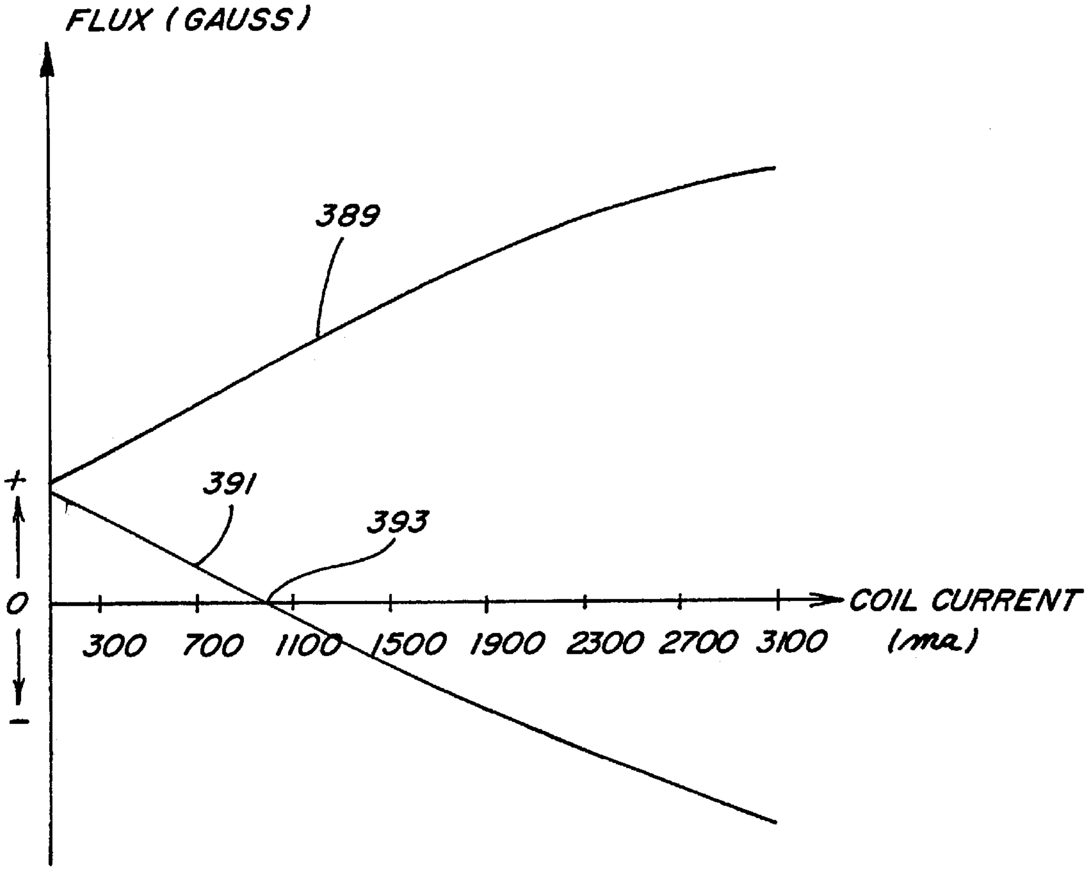

Fig. 16 C

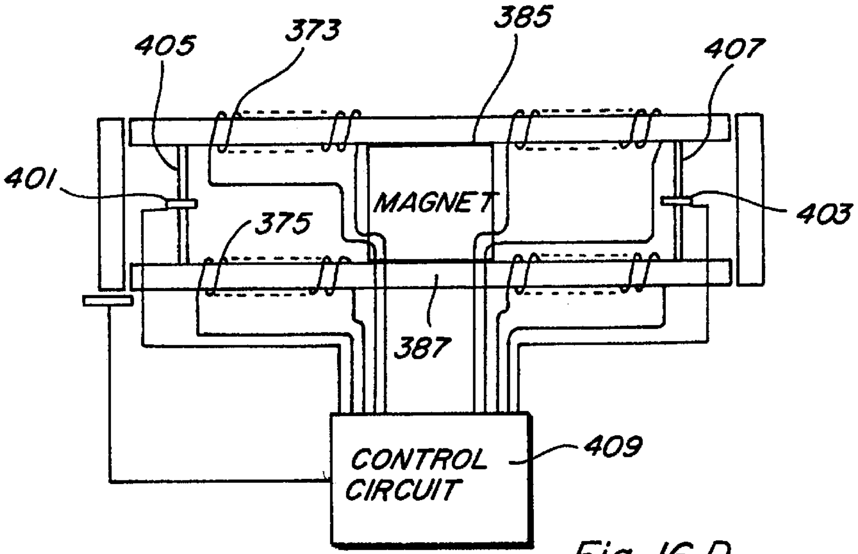
405
373
385
407
401
403
MAGNET
375
387
CONTROL CIRCUIT
409

Fig. 16 D

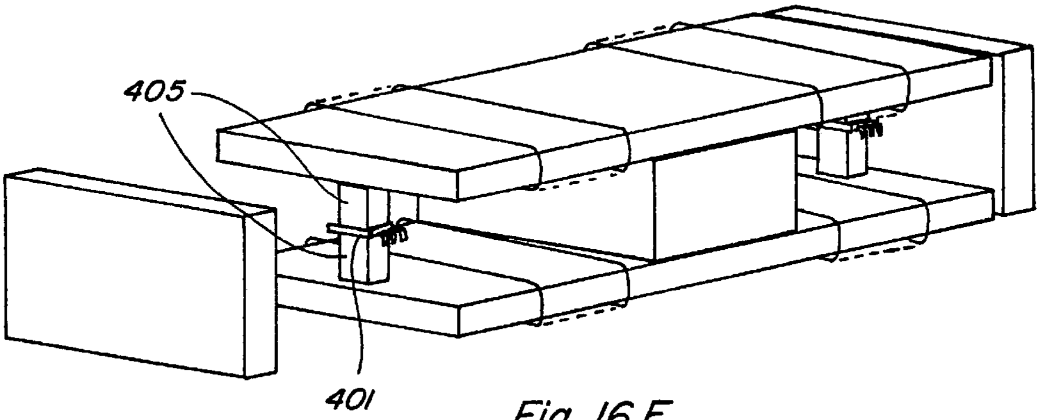

Fig. 16 E

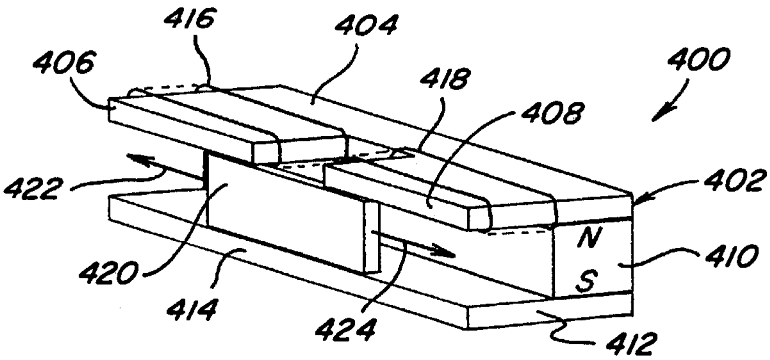
416
404
406
400
418
408
422
402
N
410
S
420
414
424
412

Fig. 17A

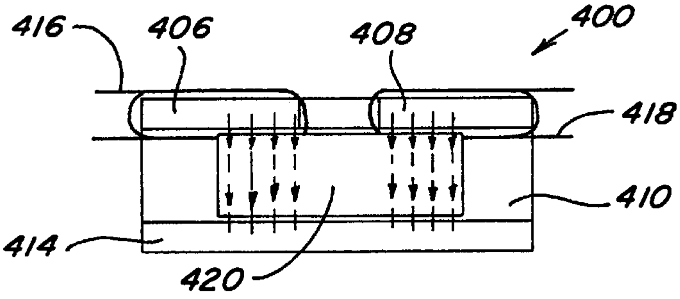
416
406
408
400
418
410
414
420

Fig. 17B

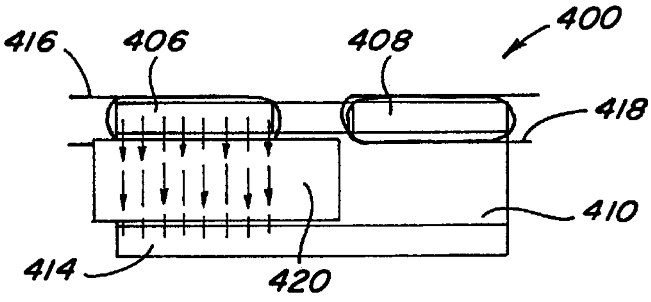
416
406
408
400
418
410
414
420

Fig. 17C

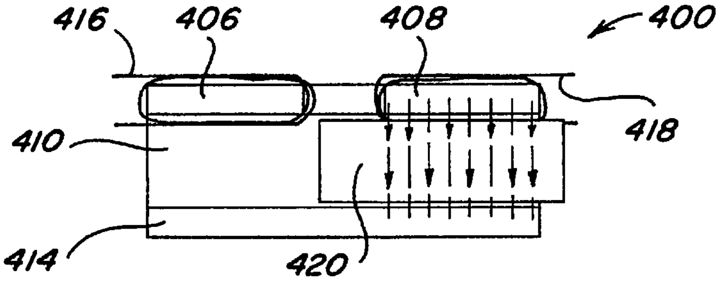
416
406
408
400
410
418
414
420

Fig. 17D

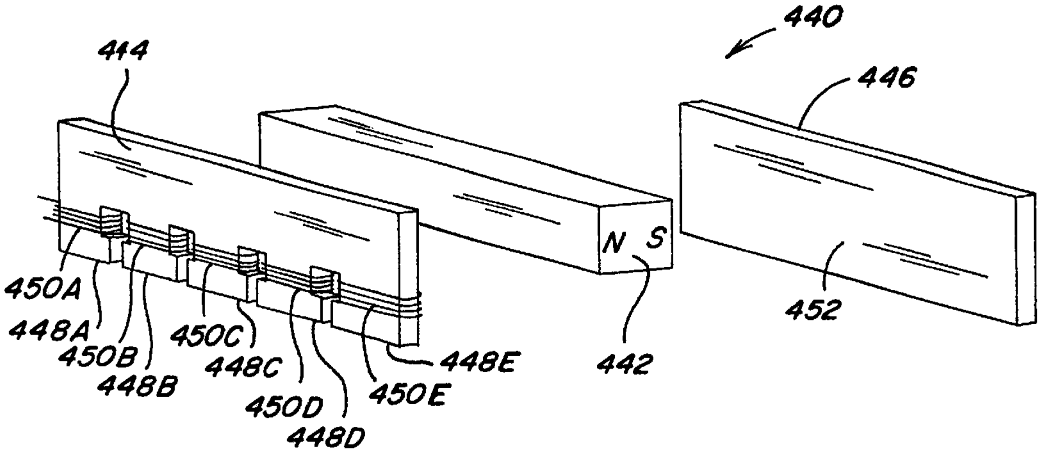
440
414
446
N
S
452
442
450A
448A
450B
448B
450C
448C
450D
448D
450E
448E

Fig. 18A

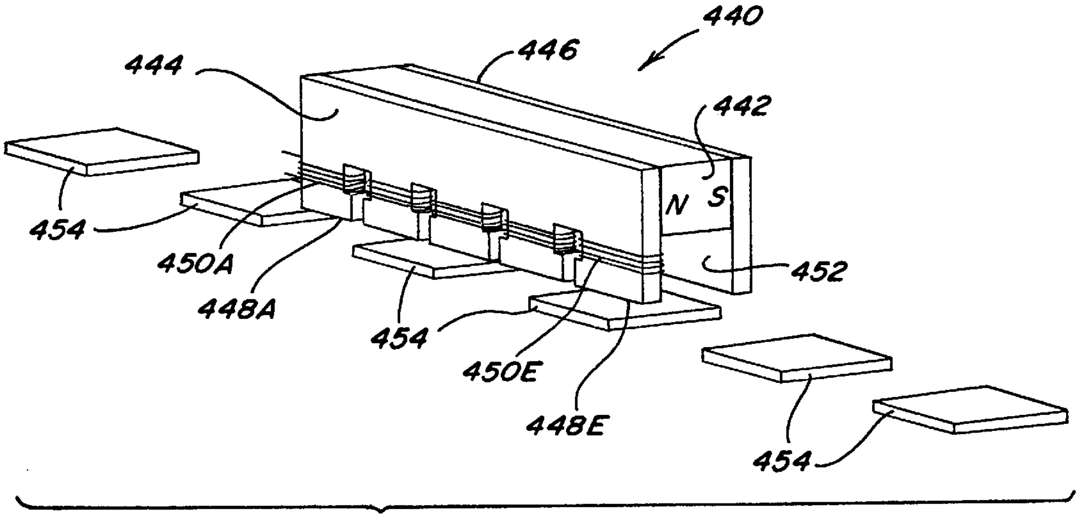
440
444
446
442
N
S
452
454
450A
448A
454
450E
448E
454

Fig. 18B

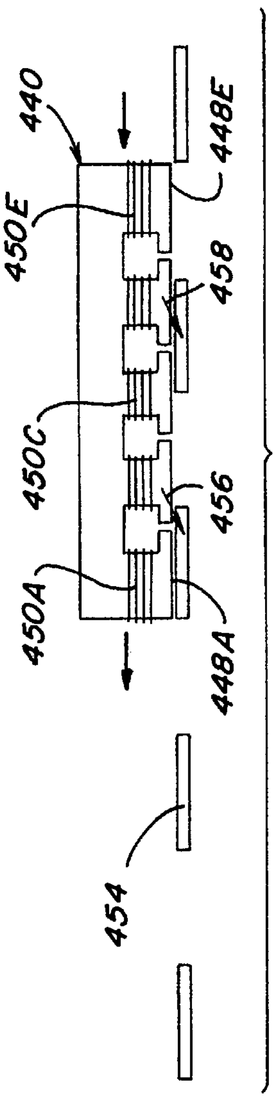
440
450E
448E
458
450C
456
450A
448A
454

Fig. 18C

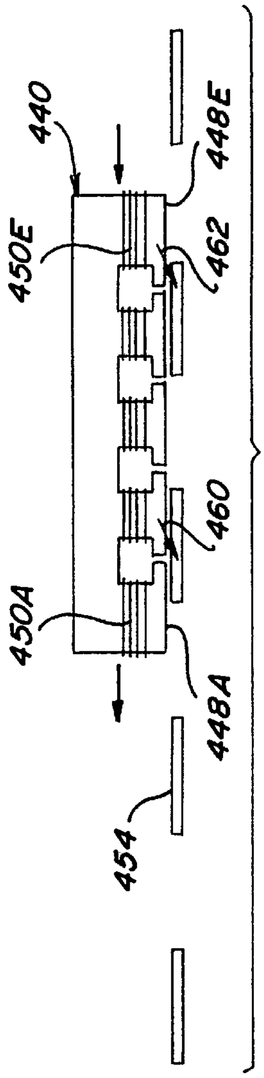
440
450E
448E
462
460
450A
448A
454

Fig. 18D

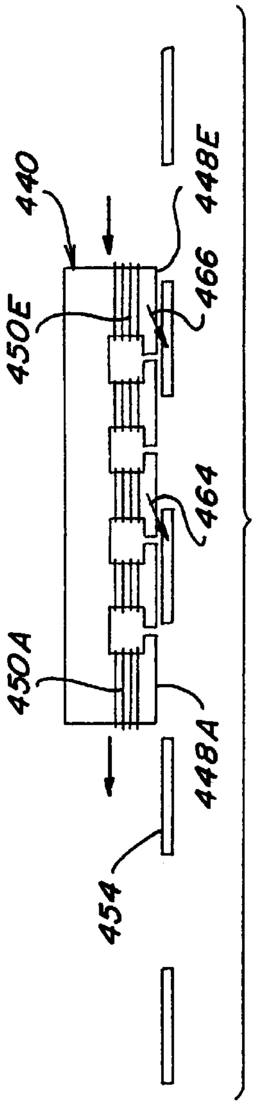
440
450E
448E
466
464
450A
448A
454

Fig. 18E

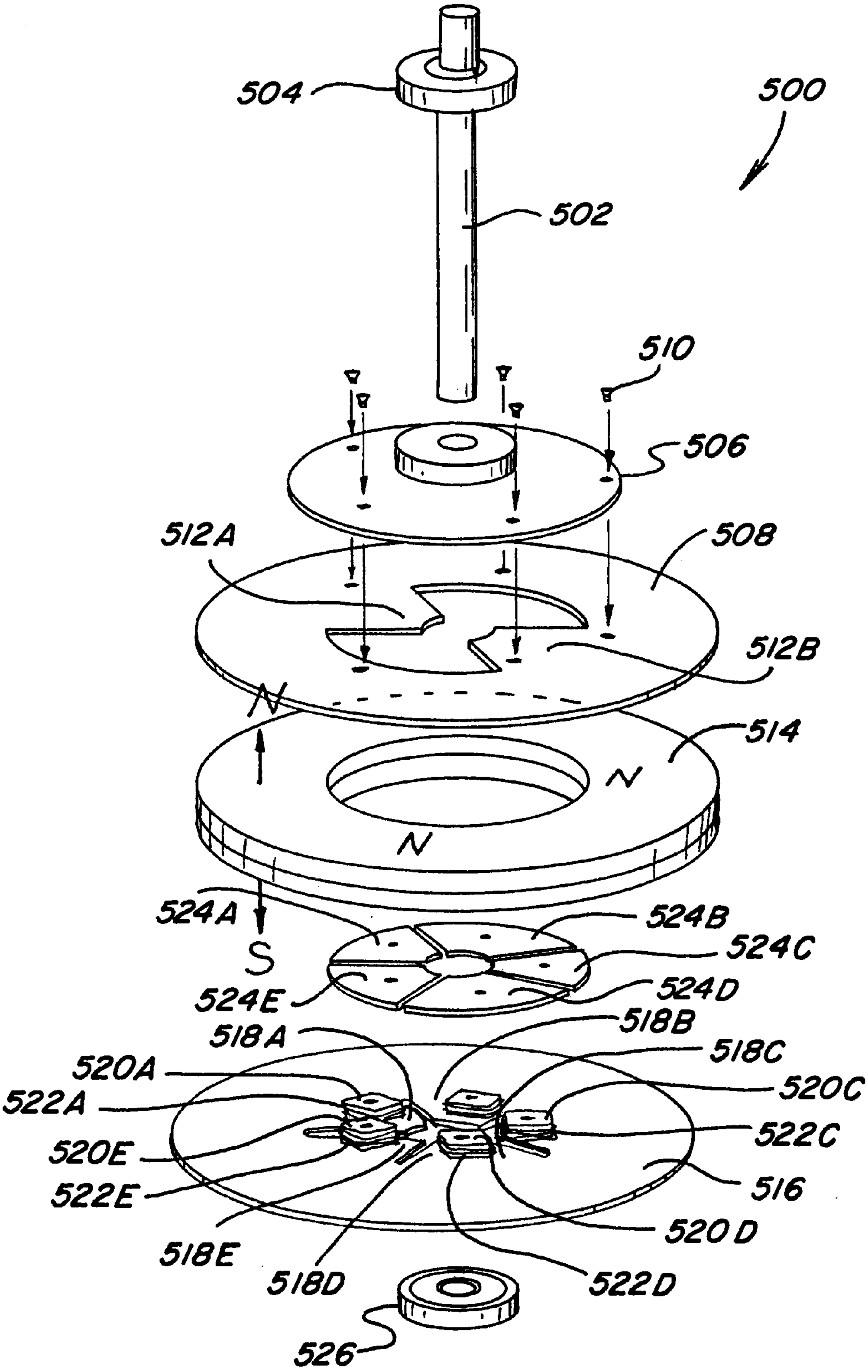
N
N
N
S

Fig. 19

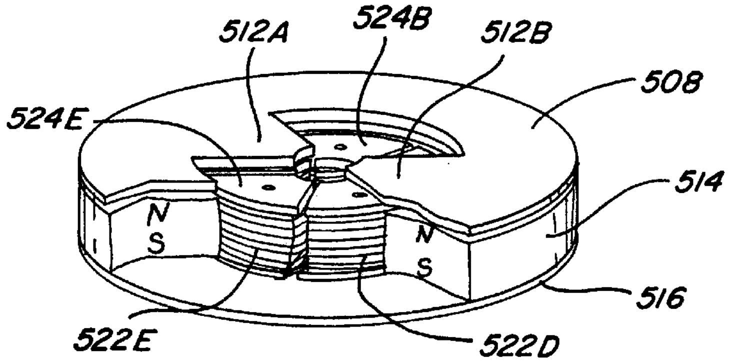
512A
524B
512B
508
524E
514
N
S
N
S
516
522E
522D

Fig. 20

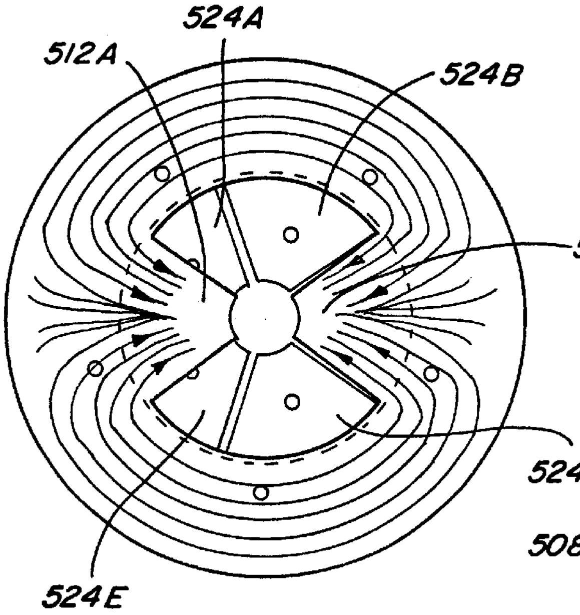
524A
512A
524B
512B
524D
524E

Fig. 21A

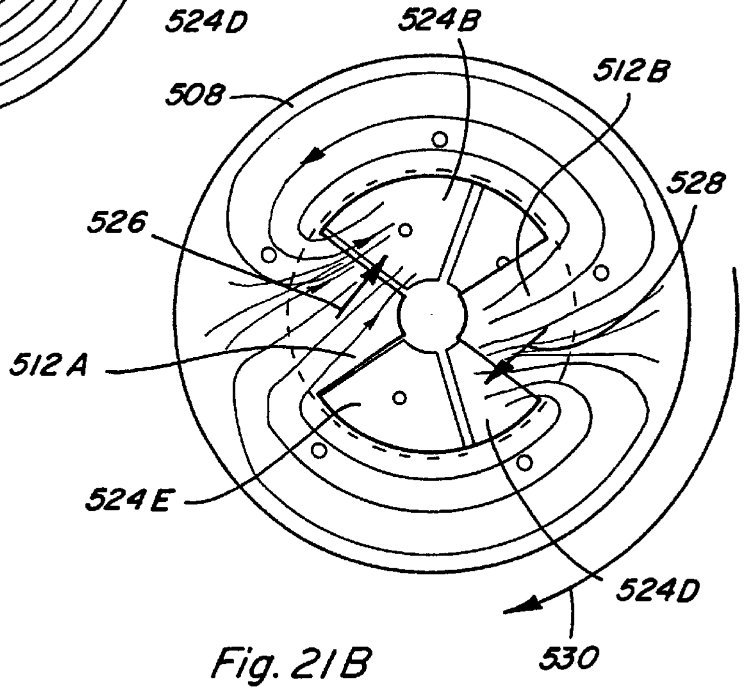
524B
512B
508
528
526
512A
524E
524D
530

Fig. 21B

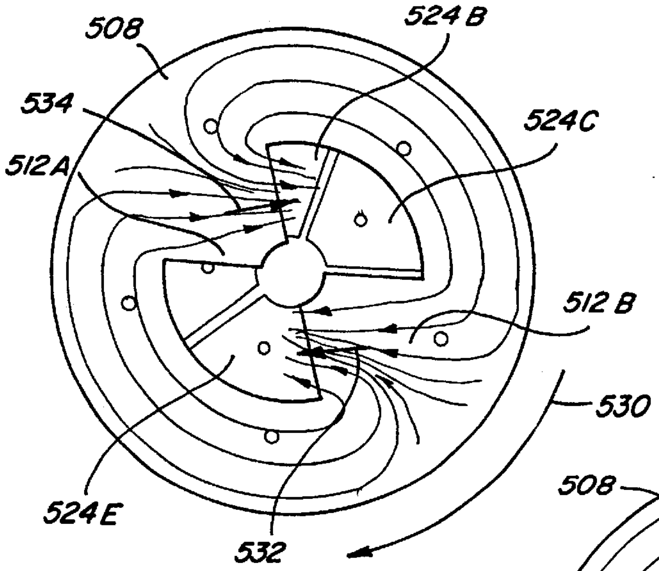
508
524B
534
524C
512A
512B
530
524E
532

Fig. 21C

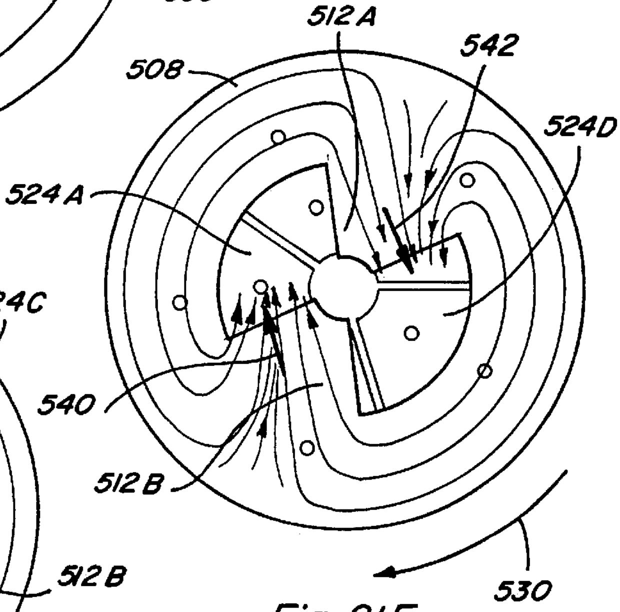
512A
542
508
524D
524A
540
512B
530

Fig. 21E

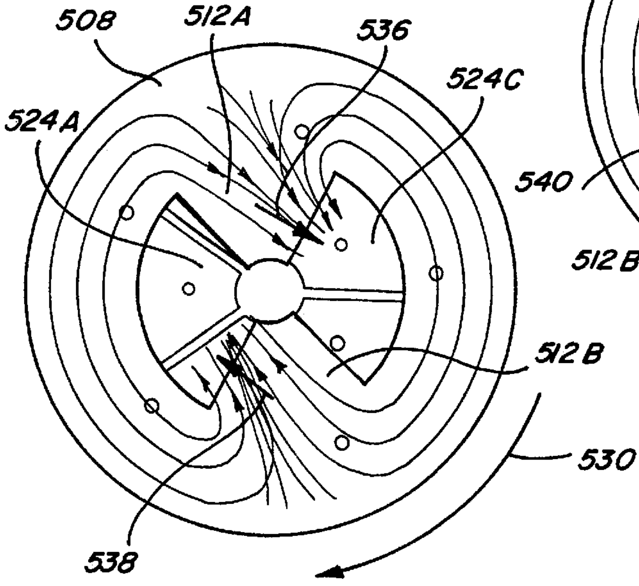
508
512A
536
524C
524A
512B
530
538

Fig. 21D

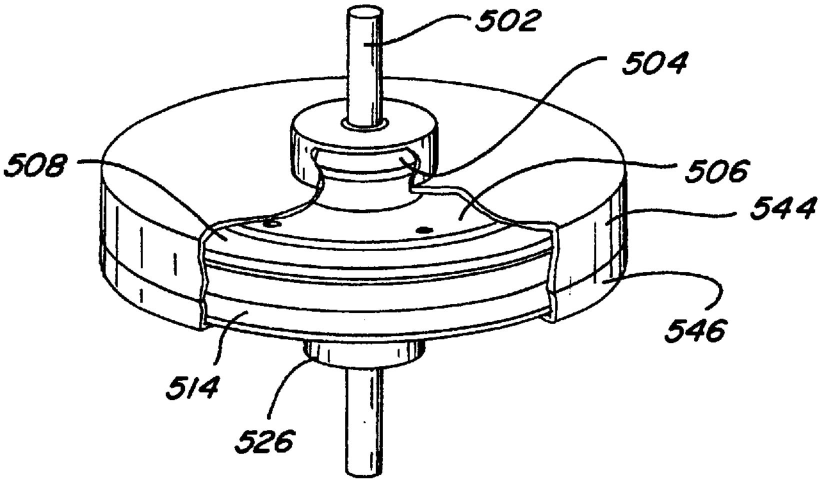

Fig. 22

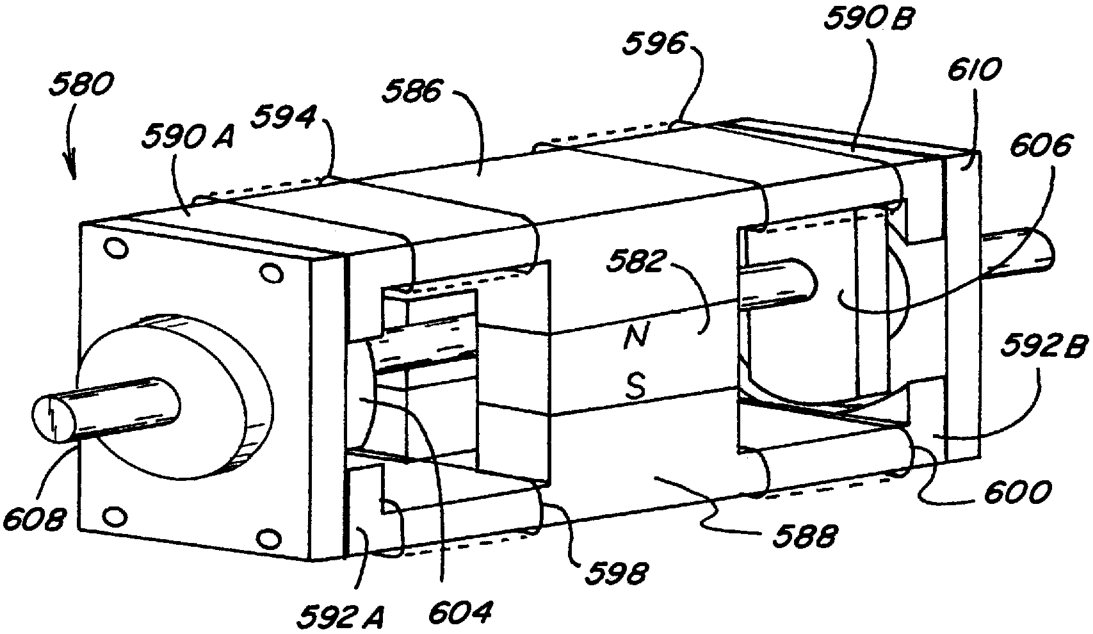
590 A
590 B
N
S
592 B
592 A

Fig. 24

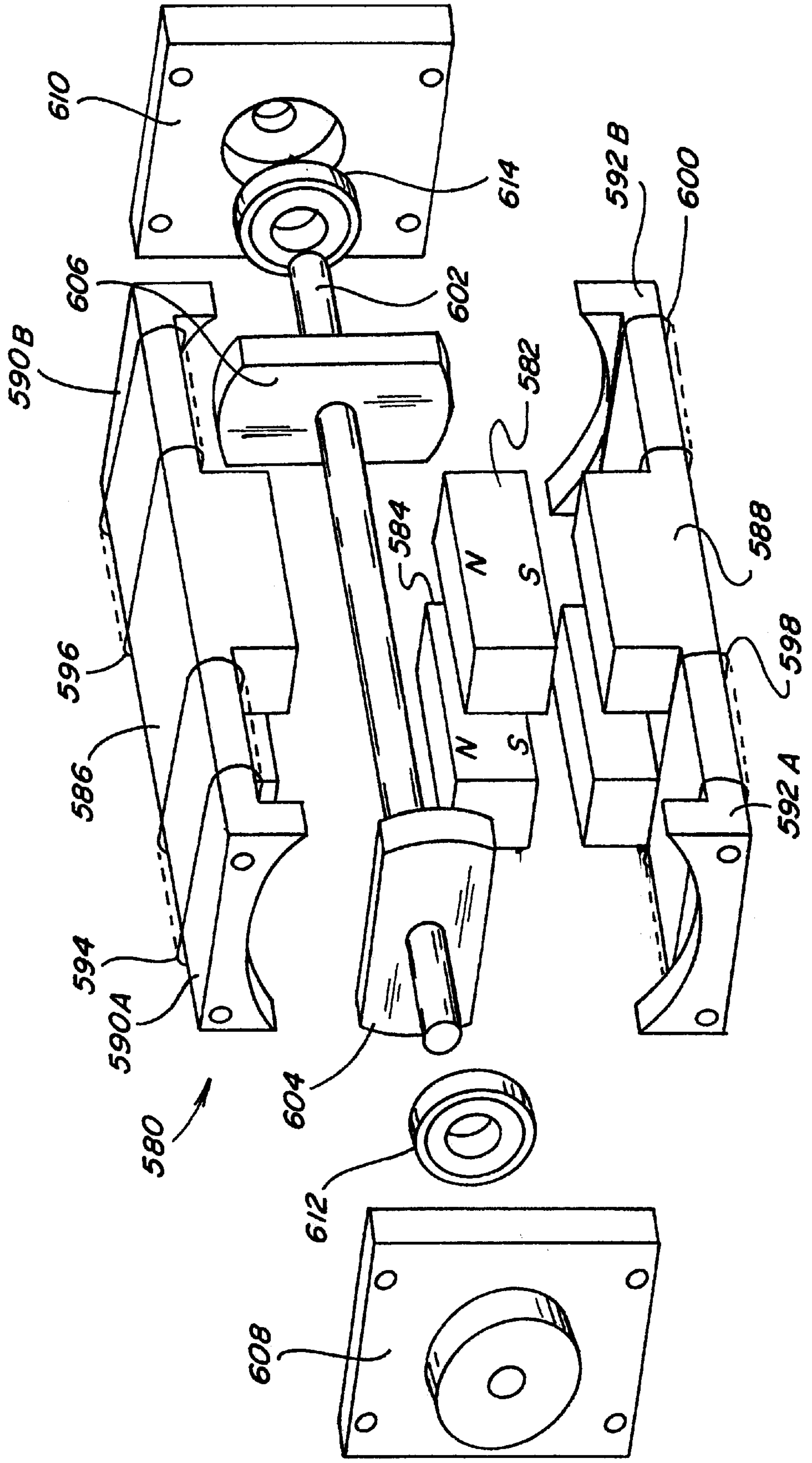
N
S
N
S

Fig. 23

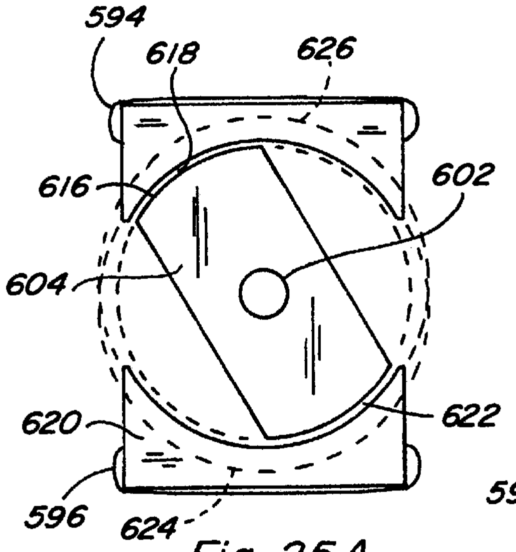
594
618
626
616
602
604
620
622
596
624

Fig. 25A

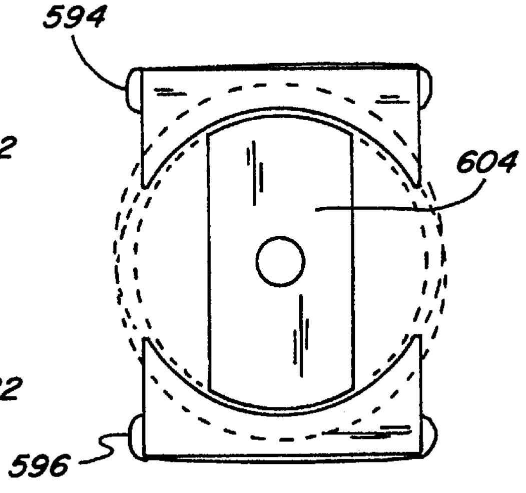

Fig. 25B

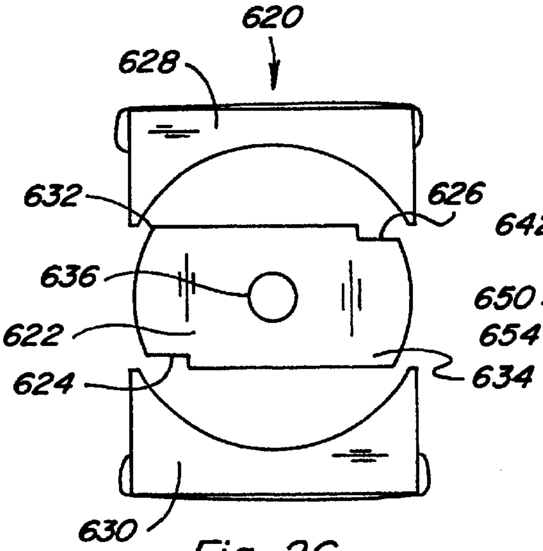
620
628
632
626
636
622
624
634
630

Fig. 26

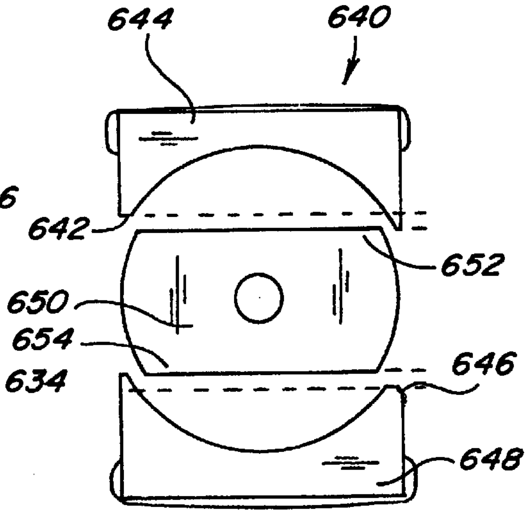
644
640
642
652
650
654
646
648

Fig. 27

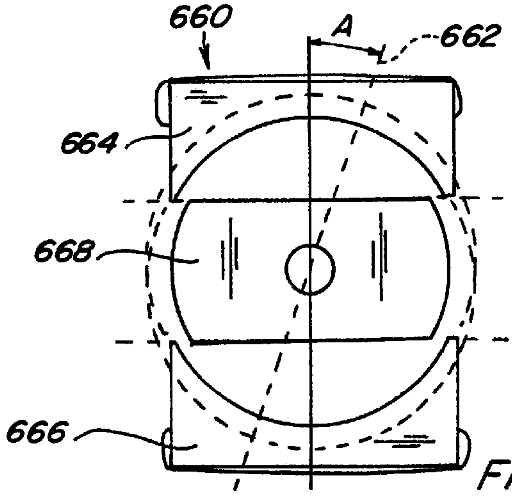
660
A
662
664
668
666

Fig. 28

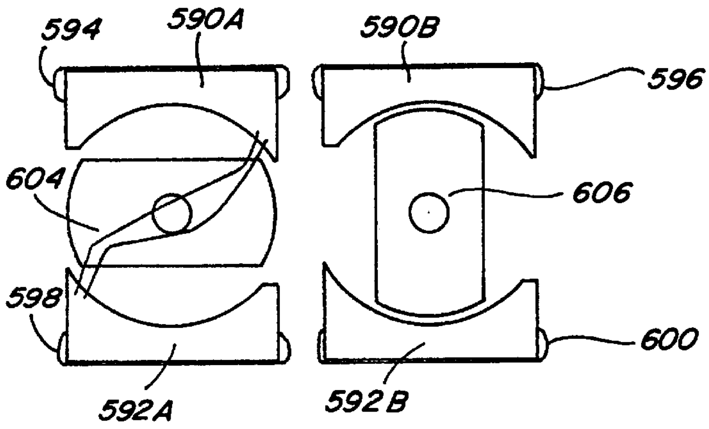
594
590A
590B
596
604
606
598
600
592A
592B

Fig. 29A

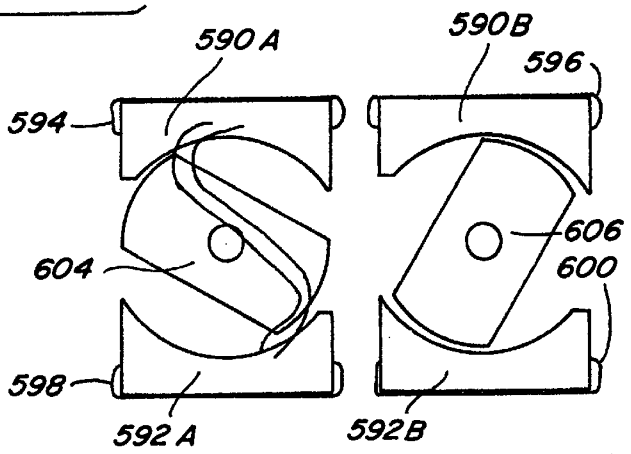
590A
590B
596
594
606
604
600
598
592A
592B

Fig. 29B

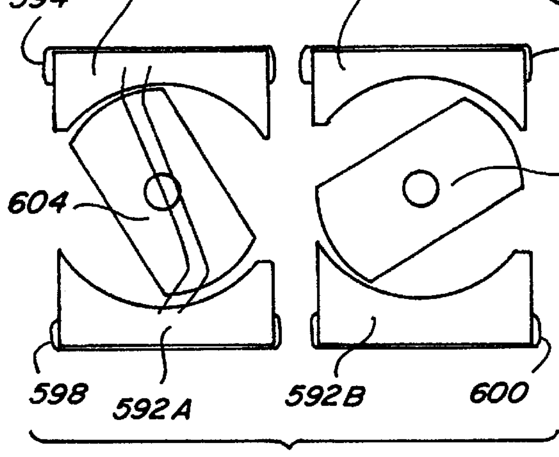
590A
590B
594
596
606
604
598
592A
592B
600

Fig. 29C

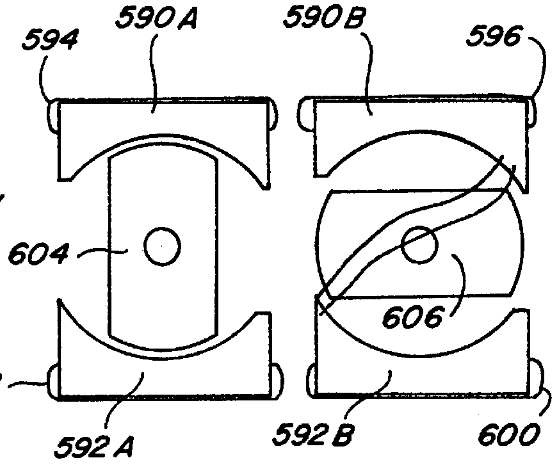
594
590A
590B
596
604
606
598
592A
592B
600

Fig. 29D

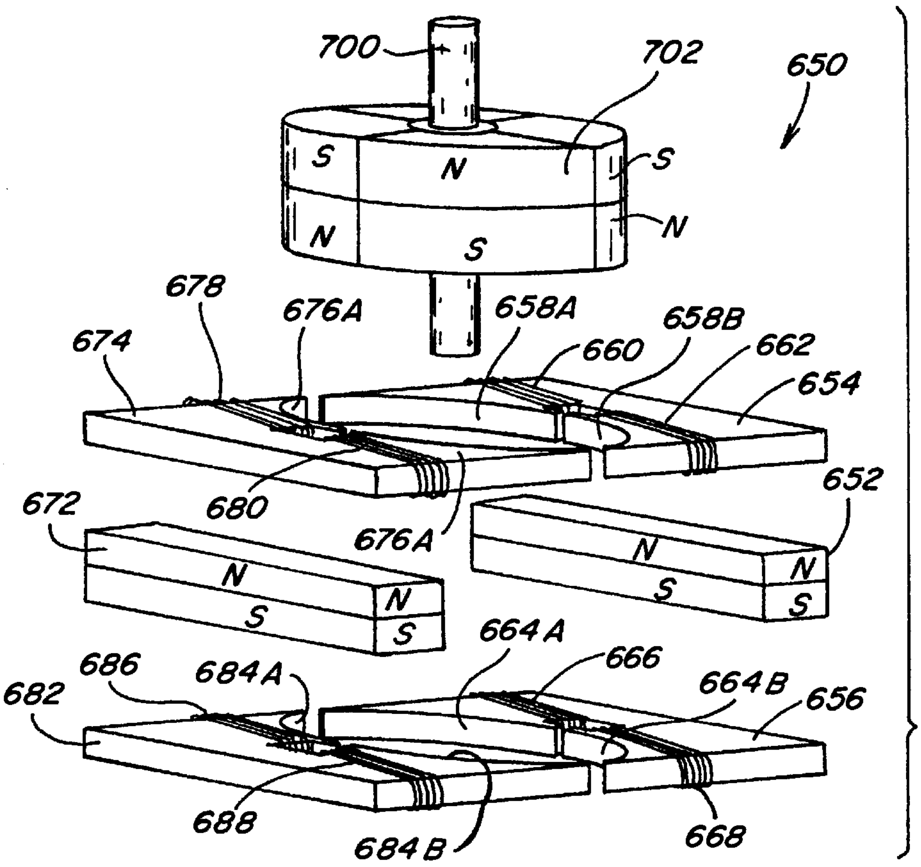
S
N
S
N
N
S
N
N
N
S
S
S
N
S

Fig. 30

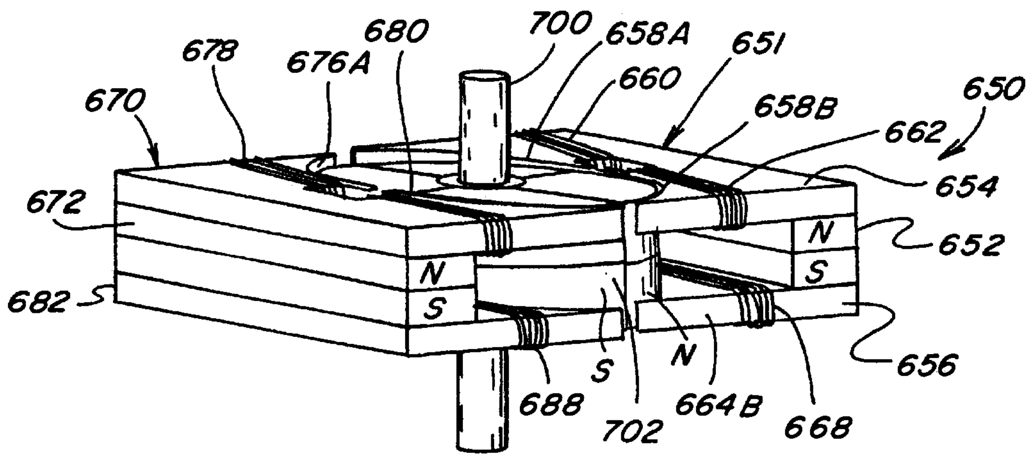
N
N
S
S
S
N

Fig. 31

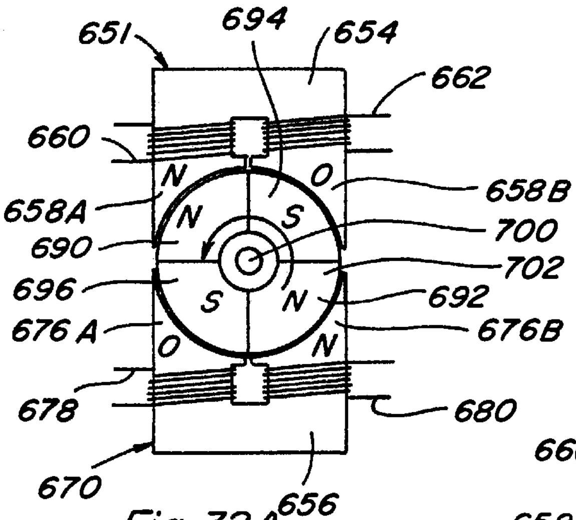
651
694
654
662
660
N
O
658A
658B
690
N
S
700
702
696
S
N
692
676A
676B
O
N
678
680
670
656

Fig. 32A

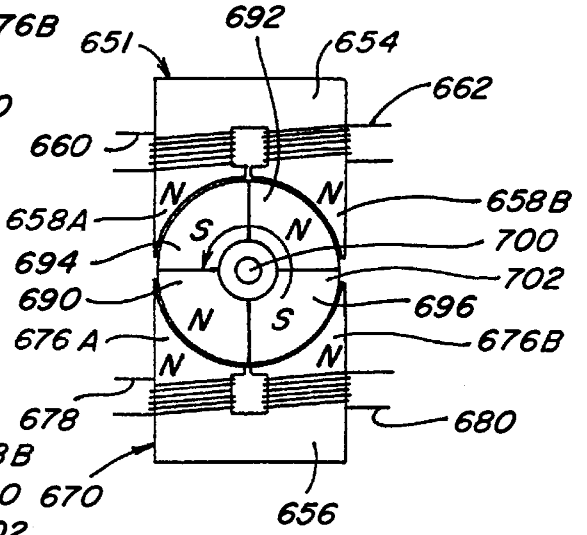
692
651
654
662
660
N
N
658A
658B
694
S
N
700
690
702
696
N
S
676A
676B
N
N
678
680
670
656

Fig. 32B

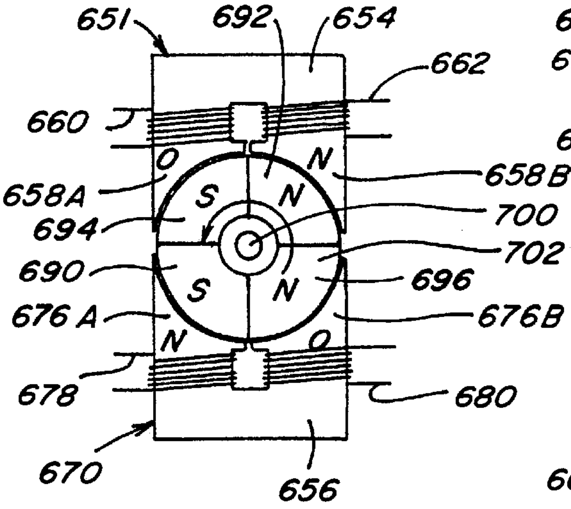
651
692
654
662
660
O
N
658A
658B
S
N
694
700
702
690
696
S
N
676A
676B
N
O
678
680
670
656

Fig. 32C

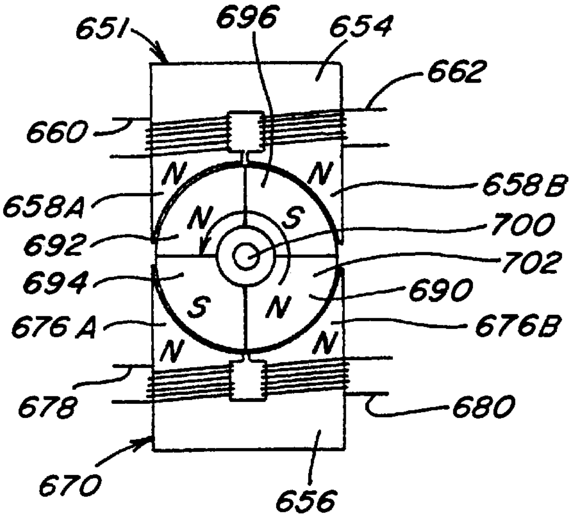
651
696
654
662
660
N
N
658A
658B
N
S
692
700
702
694
690
S
N
676A
676B
N
N
678
680
670
656

Fig. 32D

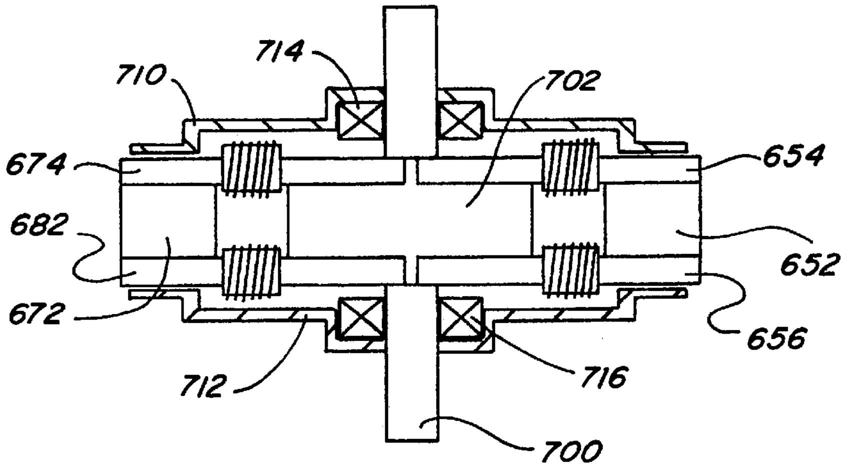

Fig. 33

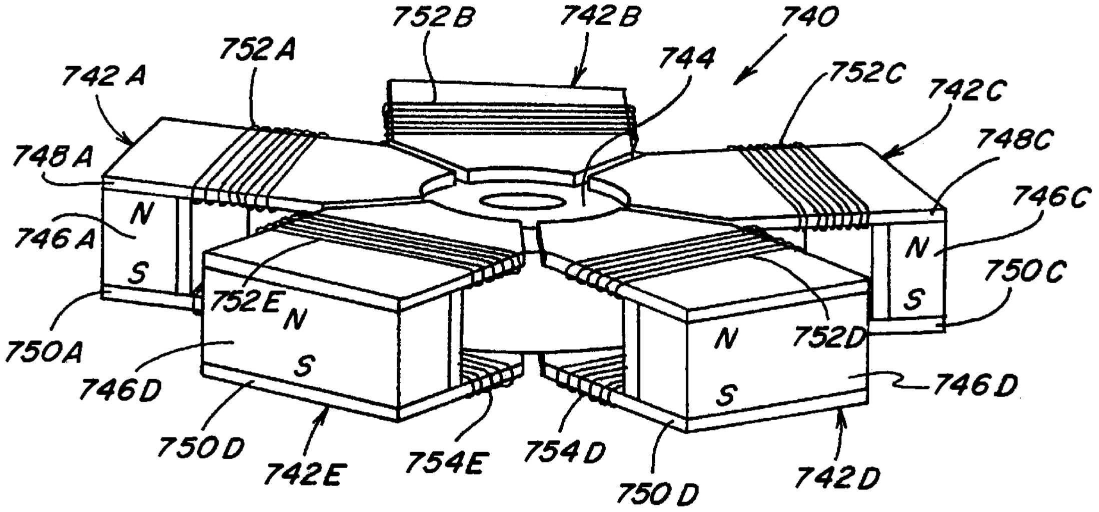
N
S
N
S
N
N
S
S

Fig. 34

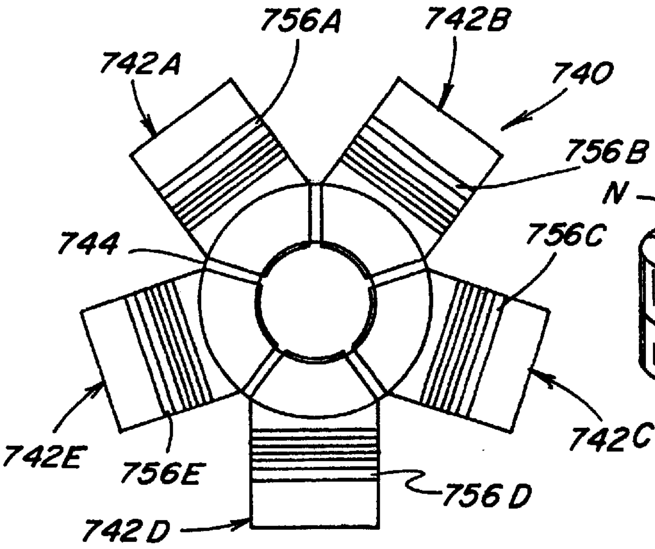
756A
742B
742A
740
756B
744
756C
742E
756E
756D
742C
742D

Fig. 35

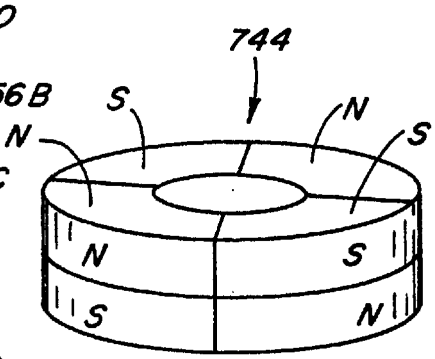
744
S
N
S
N
N
S
S
N

Fig. 36

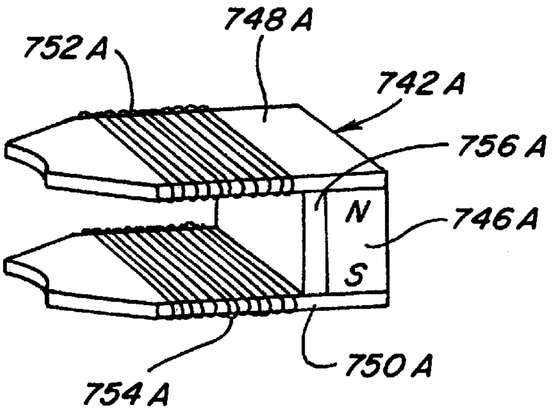
752A
748A
742A
756A
N
746A
S
754A
750A

Fig. 37

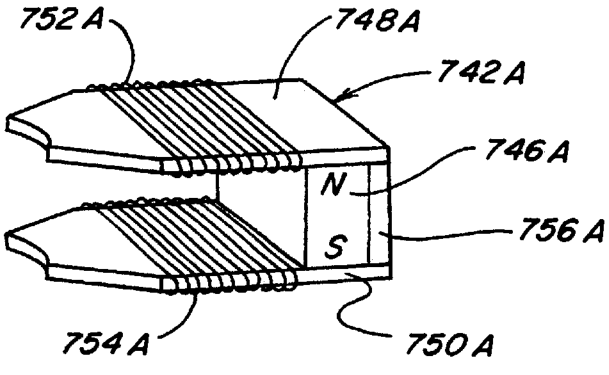
752A
748A
742A
746A
N
756A
S
754A
750A

Fig. 38

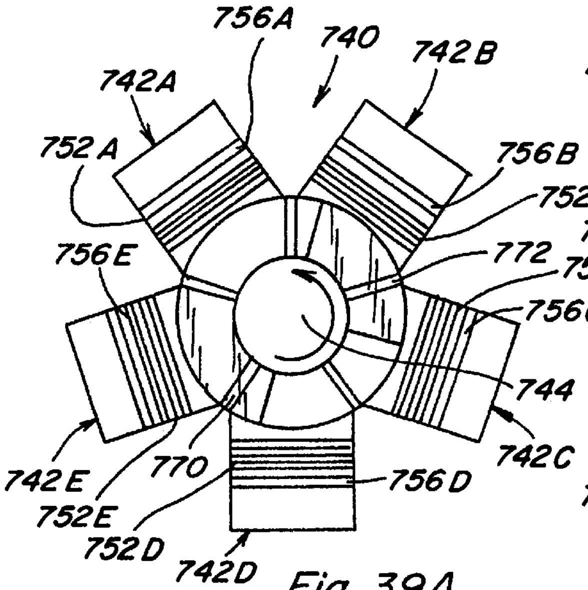
756A
740
742B
742A
752A
756B
752B
756E
772
752C
756C
744
742C
742E
770
752E
756D
752D
742D

Fig. 39A

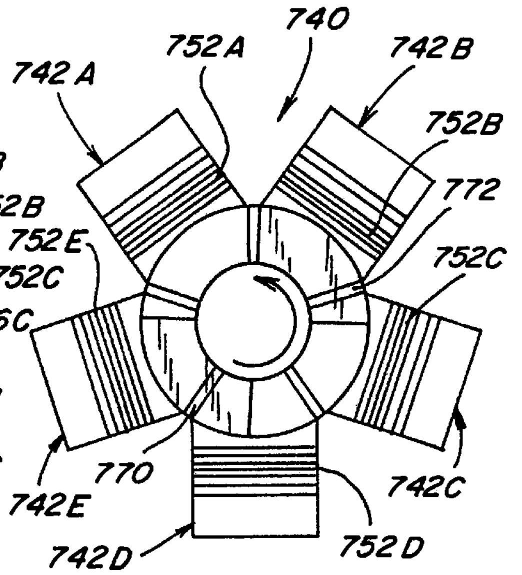
740
752A
742B
742A
752B
772
752E
752C
770
742C
742E
742D
752D

Fig. 39B

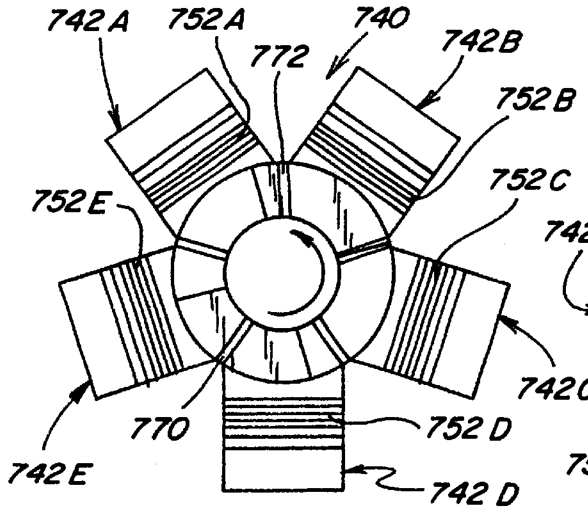
742A
752A
740
742B
772
752B
752E
752C
742C
770
752D
742E
742D

Fig. 39C

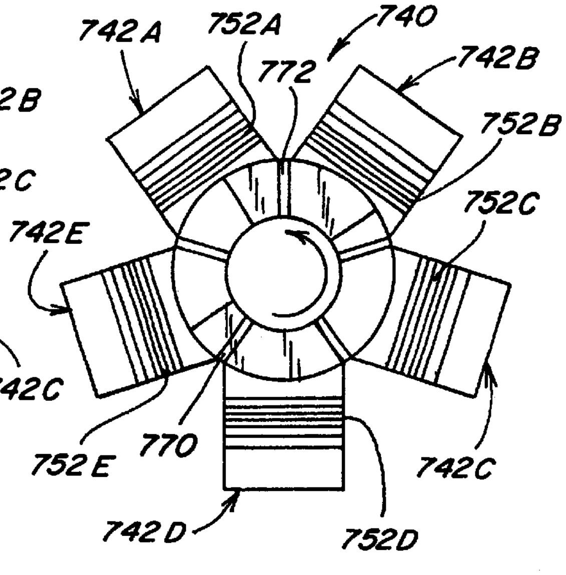
742A
752A
740
772
742B
752B
742E
752C
752E
770
742C
742D
752D

Fig. 39D

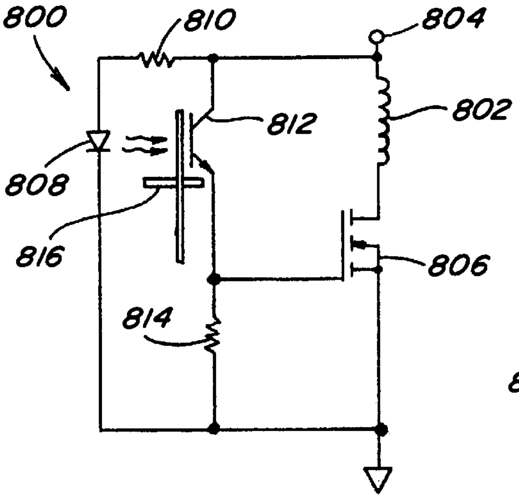
800
810
804
812
802
808
816
806
814

Fig. 40

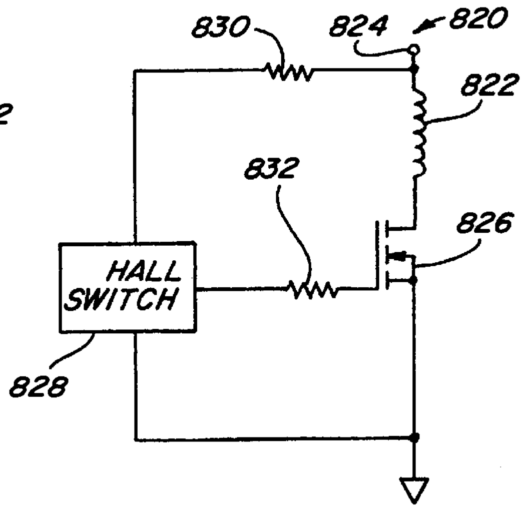
830
824
820
822
832
826
HALL SWITCH
828

Fig. 41

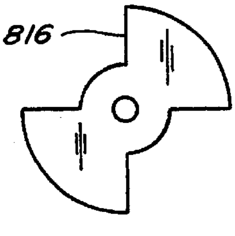

Fig. 42

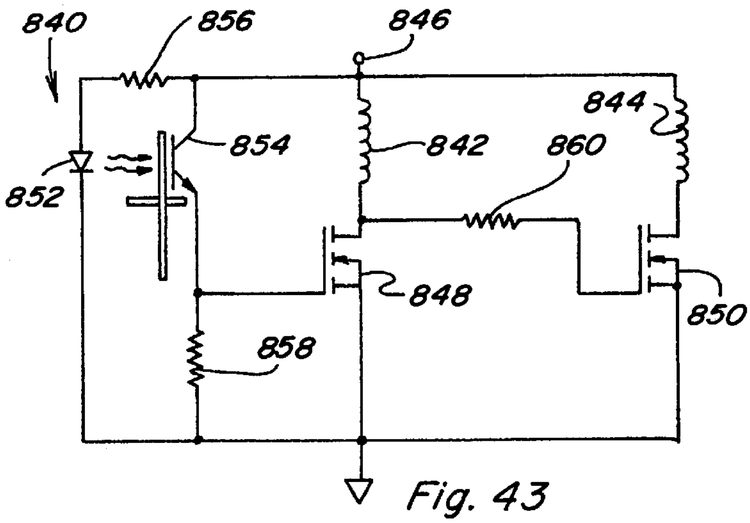
840
856
846
844
854
842
860
852
848
850
858

Fig. 43

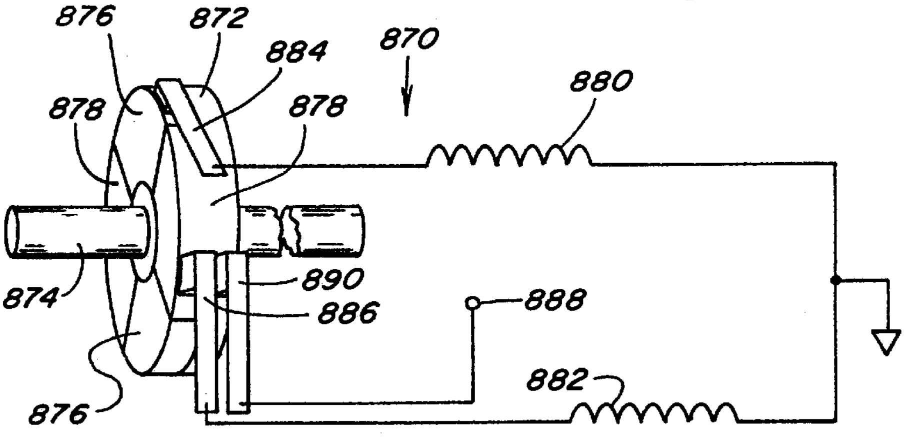
876
872
884
870
878
878
880
874
890
886
888
882
876

Fig. 44

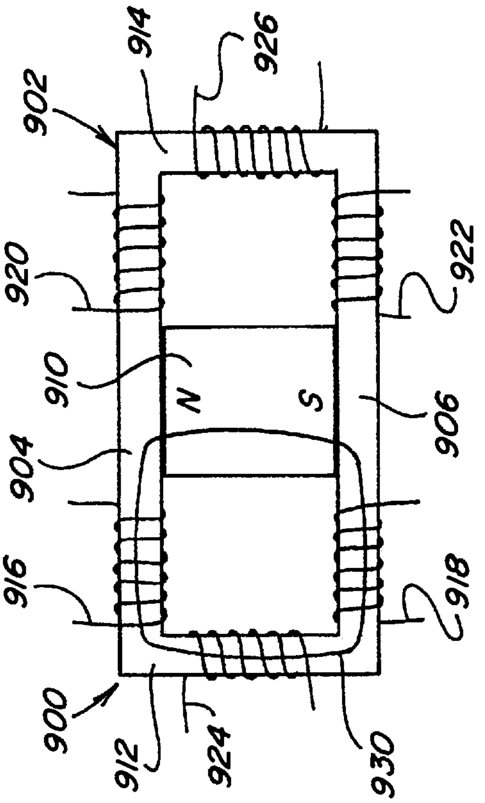
902
914
926
920
922
910
N
S
906
904
916
918
900
912
924
930

Fig. 45B

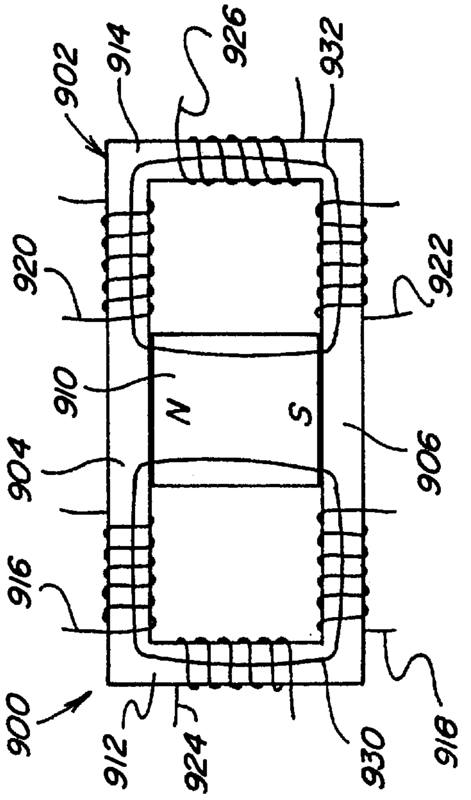
902
914
926
932
920
922
910
N
S
904
906
916
900
912
924
930
918

Fig. 45A

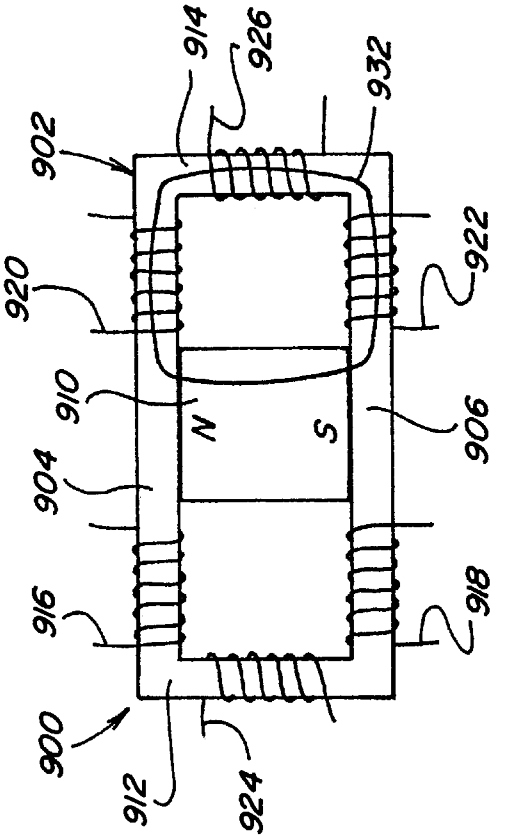
914
926
932
902
920
922
910
N
S
906
904
916
918
900
912
924

Fig. 45C

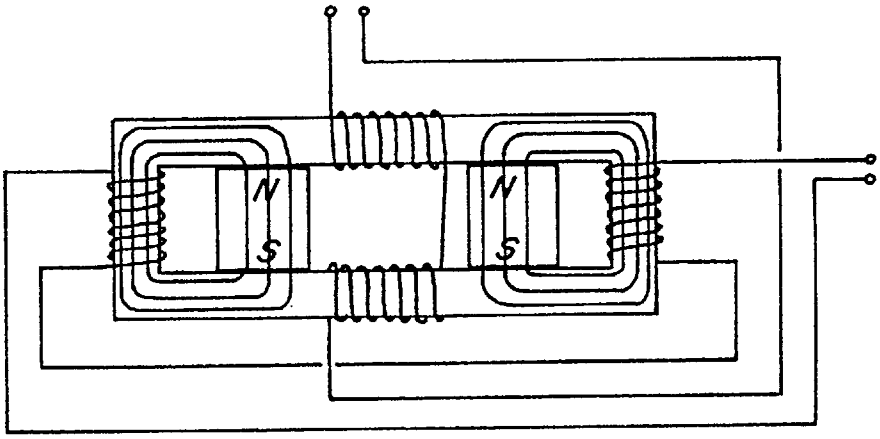

Fig. 45 X

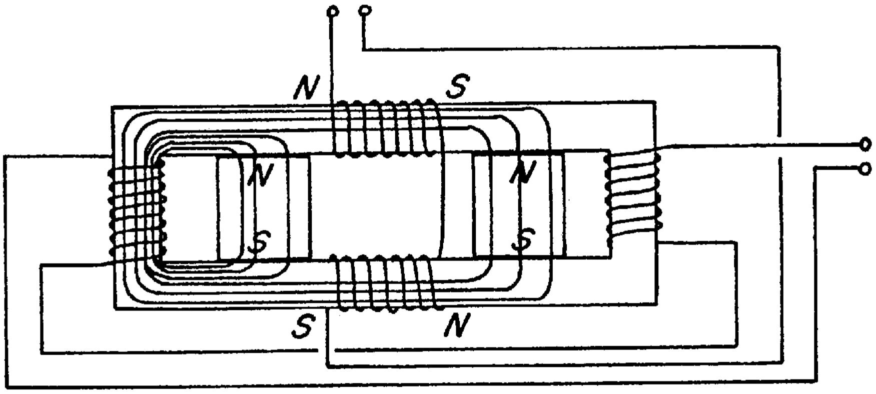

Fig. 45 Y

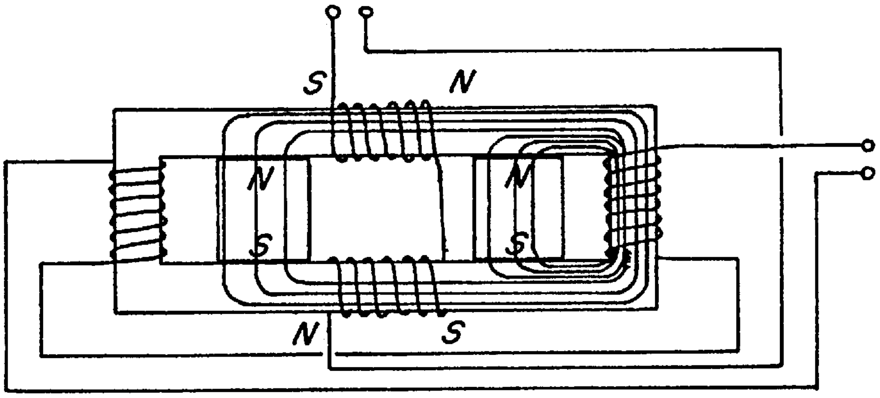

Fig. 45 Z

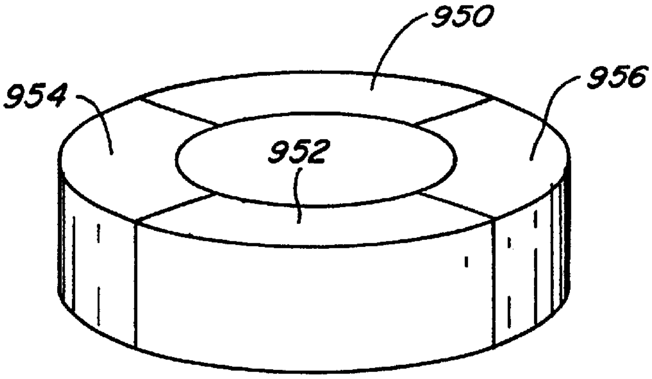
950
954
956
952

Fig. 46

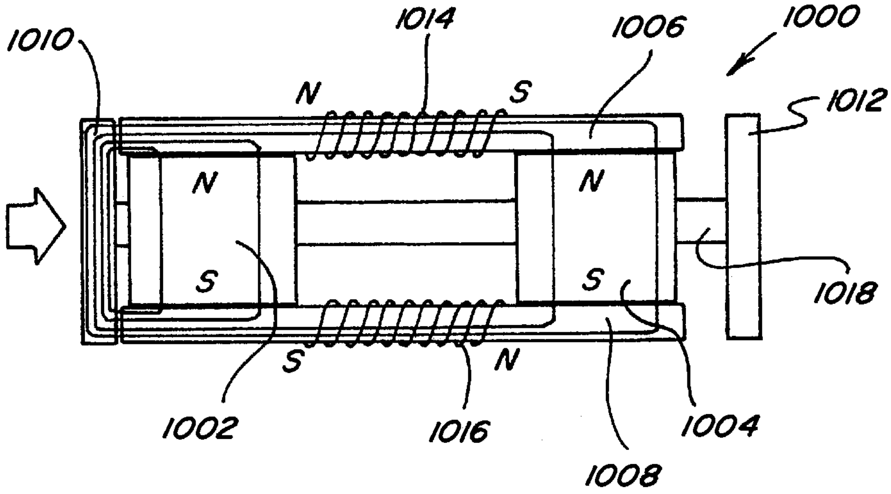
1010
1014
1006
1000
N
S
1012
N
N
S
S
1018
S
N
1002
1016
1004
1008

Fig. 47

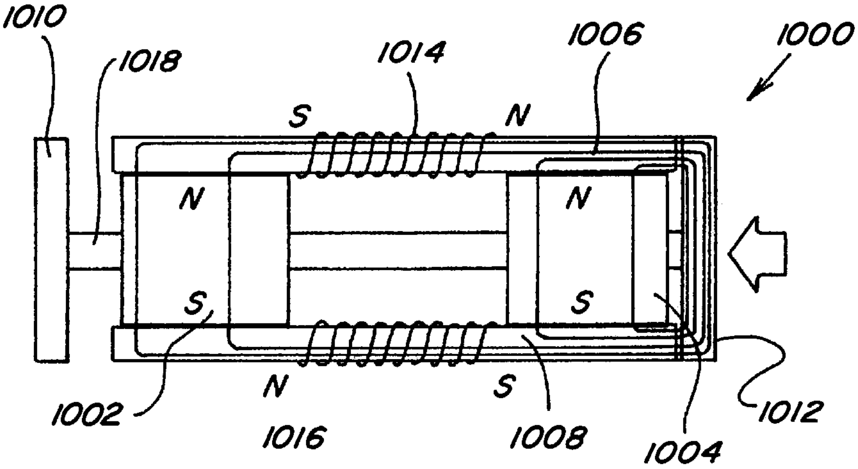
1010
1018
1014
1006
1000
S
N
N
N
S
S
1002
N
S
1016
1008
1012
1004

Fig. 48

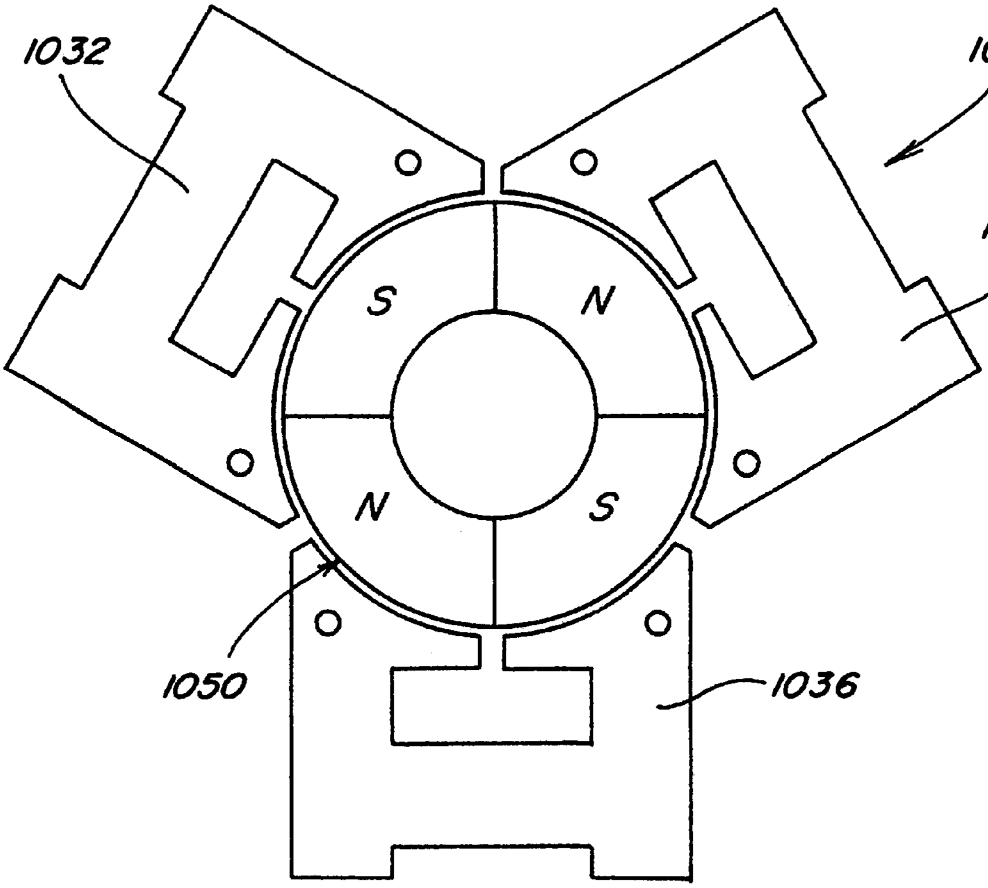
1032
1030
1034
S
N
N
S
1050
1036

Fig. 49

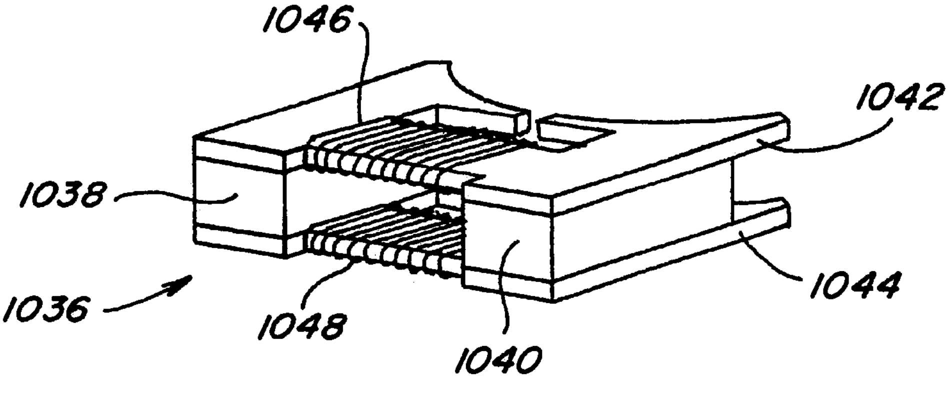

Fig. 50

METHODS FOR CONTROLLING THE PATH OF MAGNETIC FLUX FROM A PERMANENT MAGNET AND DEVICES INCORPORATING THE SAME

FIELD OF THE INVENTION

This invention relates generally to permanent magnet devices and more particularly, to a permanent magnet control component in which the flow of flux from a permanent magnet is controlled between two or more flux paths by utilizing timed delivery of electrical signals through one or more coils placed along at least one of the flux paths. Such permanent magnet control components may take on a variety of configurations facilitating use of such components in a variety of applications including applications involving the production of reciprocating, linear, and rotary motion and power conversion. Several novel permanent magnet rotary motion devices of motor constructions which operate by controlling the path of magnetic flux from one or more permanent magnets are described, such permanent magnet rotary motor constructions having increased efficiency and more desirable torque characteristics as compared to many currently used motors.

BACKGROUND OF THE INVENTION

Magnetic force of attraction is commonly used in a variety of types of permanent magnet devices including both linear and rotary motors. In the field of such permanent magnet devices there is a continuous pursuit of increased efficiency and reduced complexity.

Accordingly, an object of the present invention is to provide a permanent magnet control component in which the path of a given level of permanent magnet flux can be controlled by a lesser level of electromagnetic flux.

Another object of the present invention is to provide a permanent magnet control component in which substantially all of the flux from a permanent magnet can be switched between at least two different flux paths of the permanent magnet control component so as to enable useful work in the form of linear, reciprocating, and rotary motion.

Still another object of the present invention is to provide permanent magnet control components and motor constructions in which flux path control is provided by energizing an **1 0** electromagnet to oppose the magnetic flux of one or more permanent magnets.

Another object of the present invention is to provide permanent magnet control components and motor constructions in which flux path control is provided by energizing an electromagnet to aid the magnetic flux of one or more permanent magnets.

Yet another object of the present invention is to provide permanent magnet motor **15** constructions with improved operating characteristics.

SUMMARY OF THE INVENTION

These and other objects of the invention are attained by an apparatus which, in one aspect, is a permanent magnet device, comprising a permanent magnet having north and south pole faces, a first pole piece, a second pole piece, a first control coil, a second control coil, and circuit means, the first pole piece positioned adjacent the north pole face of the permanent magnet and including a first path portion, a second path portion and a third portion, the first path portion extending beyond a perimeter of the north pole face and the second path portion extending beyond the perimeter of the north pole face to define first and second flux paths for magnetic flux emanating from the north pole face of the permanent magnet, the first path portion of the first pole piece connected to the second path portion of the first pole piece by the third portion which extends across the north pole face of the permanent magnet, the second pole piece positioned adjacent the south pole face and including a first path portion and a second path portion, the first path portion extending beyond a perimeter of the south pole face and substantially aligned with the first path portion of the first pole piece, the second path portion extending beyond the perimeter of the south pole face and substantially aligned with the second path portion of the first pole piece, the first control coil positioned around the first path portion of the first pole piece, the second control coil positioned around the second path portion of the first pole piece, the circuit means connected to each of the first control coil and the second control coil to alternatingly energize the first coil and the second coil in a timed sequential manner.

Another aspect of the present invention provides a method for controlling the path of magnetic flux from a permanent magnet which involves placing a first pole piece adjacent a first pole face of the permanent magnet so as to have at least first and second path portions extending beyond a perimeter of the first pole face. A second pole piece is placed adjacent a second pole face of the permanent magnet so as to include at least one portion which substantially aligns with the first and second path portions of the first pole piece. A first control coil is placed along and around the first path portion of the first pole piece and a second control coil is placed along and around the second path portion of the first pole piece. The first control coil is repeatedly energized in a permanent magnet magnetic flux opposing manner so as to prevent magnetic flux of the permanent magnet from traversing the first path portion of the first pole piece, and the second control coil is repeatedly energized in a permanent magnet magnetic flux opposing manner so as to prevent magnetic flux of the permanent magnet from traversing the second path portion of the first pole piece.

Yet another aspect of the present invention provides a method for controlling the path of magnetic flux from a permanent magnet by placing a first pole piece adjacent a first pole face of the permanent magnet so as to have at least first and second path portions extending beyond a perimeter of the first pole face. A second pole piece is placed adjacent a second pole face of the permanent magnet so as to include at least one portion which substantially aligns with the first and second path portions of the first pole piece. A first control coil is placed along and around the first path portion of the first pole piece, and a second control coil is placed along and around the second path portion of the first pole piece. The following steps are alternatingly performed in a repeated manner:

(i) energizing the first control coil in a permanent magnet magnetic flux aiding manner so as to couple with substantially all magnetic flux of the permanent magnet such that substantially no magnetic flux of the permanent magnet traverses the second path portion of the first pole piece when the first control coil is so energized; and

(ii) energizing the second control coil in a permanent magnet magnetic flux opposing manner so as to couple with substantially all magnetic flux of the permanent magnet such that substantially no magnetic flux of the permanent magnet traverses the first path portion of the first pole piece when the second control coil is so energized.

A further aspect of the present invention provides method for controlling the path of magnetic flux from a permanent magnet by placing a first pole piece adjacent a first pole face of the permanent magnet so as to have at least first and second path portions extending beyond a perimeter of the first pole face, and placing a second pole piece adjacent a second pole face of the permanent magnet so as to include at least one portion which substantially aligns with the first and second path portions of the first pole piece. A first control coil is placed along and around the first path portion of the first pole piece, and a second control coil is placed along and around the second path portion of the first pole piece. The following steps are alternatingly performed in a repeated manner:

(i) energizing the first control coil in a permanent magnet magnetic flux aiding manner so as to couple with substantially all magnetic flux of the permanent magnet such that substantially no magnetic flux of the permanent magnet traverses the second path portion of the first pole piece when the first control coil is so energized; and

(ii) energizing the second control coil in a permanent magnet magnetic flux opposing manner so as to couple with substantially all magnetic flux of the permanent magnet such that substantially no magnetic flux of the permanent magnet traverses the first path portion of the first pole piece when the second control coil is so energized.

BRIEF DESCRIPTION OF THE INVENTION

For a better understanding of the present invention reference may be made to the accompanying drawings in which:

FIG. **1** is a perspective view of a magnetic device in which the magnetic flux from a magnetic member traverse a single path to produce a coupling force;

FIG. **2** is a perspective view of a magnetic device in which the magnetic flux from a magnetic member splits between two paths;

FIG. **3** is a side view of two magnetic members arrange in parallel between pole pieces;

FIG. **4** is a side view of two magnetic members arranged in series between pole pieces;

FIGS. **5**–**6** are side views of a permanent magnet device including a permanent magnet having pole pieces positioned against the pole faces thereof and including a movable armature;

FIGS. **7**–**9** are side views of a permanent magnet device including a permanent magnet having pole pieces positioned against the pole faces thereof to provide two magnetic flux paths and including a movable armature positionable along each magnetic flux path;

FIGS. **10**, **10**A–**10**H are perspective views of various embodiments of permanent magnet **5** control components which include two or more magnetic flux paths;

FIGS. **11**, **11**A–**11**F are side views of a permanent magnet device including a permanent magnet having pole pieces positioned against the pole faces thereof and including a movable armature and a permanent bypass extending between the pole pieces;

FIGS. **12**, **12**A–**12**E are side views of a two path permanent magnet device including two bypasses;

FIGS. **13**A–**13**C are side views of a permanent magnet linear reciprocating device;

FIG. **14** is a side view of an electromagnetic linear reciprocating device;

FIG. **15** is a side view of a two path permanent magnet device showing control coils energized in an exceeding manner;

FIGS. **16**A–E are a side view of a linear reciprocating device with control coils energized in an exceeding manner;

FIGS. **17**A–**17**D depict another embodiment of a linear reciprocating device;

FIGS. **18**A–**18**E show a linear motion device;

FIG. **19** is an exploded perspective view of a rotary motion device;

FIG. **20** is a partial assembled and cut away view of the rotary motion device of FIG. **19**;

FIGS. **21**A–**21**E are top views of the partial assembly of FIG. **20**, which views depict rotational motion thereof,

FIG. **22** is an assembled, cut-away view of the rotary motion device of FIG. **19** including a housing;

FIG. **23** is an exploded perspective view of another embodiment of a rotary motion device;

FIG. **24** is a perspective view of the rotary motion device of FIG. **23** as assembled;

FIGS. **25**A–**25**B are end views of the rotary motion device of FIG. **24** with the end cap removed to expose the rotor member;

FIGS. **26**–**28** show end views of various configurations for skewing the direction of rotation in the rotary motion device of FIG. **24**;

FIGS. **29**A–**29**D are end views of the rotary motion device of FIG. **24** illustrating a sequence of rotational movement thereof;

FIG. **30** is an exploded partial perspective view of another embodiment of a rotary motion device;

FIG. **31** is a perspective view of the rotary motion device of FIG. **30** as assembled

FIGS. **32**A–**32**D are top views of the rotary motion device of FIG. **31** illustrating rotational movement thereof;

FIG. **33** is a side view of the rotary motion device of FIG. **31** as assembled and including a housing;

FIG. **34** is a perspective view of another embodiment of a rotary motion device;

FIG. **35** is a top view of the rotary motion device of FIG. **34**;

FIG. **36** is a perspective view of the permanent magnet rotor member of the rotary motion device of FIG. **34**;

FIGS. **37** and **38** show alternative configurations for the control component incorporated into the rotary motion device of FIG. **34**;

FIGS. **39**A–**39**D are top views of the rotary motion device of FIG. **34** and depict rotational movement thereof;

FIGS. **40**–**44** are alternative variations of circuit means for controlling the timed energization of control coils in the various devices of the present invention;

FIGS. **45**A–**45**C and **45**X–**45**Z are side views of two path power conversion devices;

FIG. **46** is a schematic view of the permanent magnet portion of a rotor for use in some embodiments of the present device;

FIGS. **47** and **48** show other embodiments of a linear motion device;

FIG. **49** is a top view of another embodiment of a rotating motor like construction; and

FIG. **50** is a schematic view of one of the three stator portions of the device shown in FIG. **49**.

DETAILED DESCRIPTION OF THE DRAWINGS

Referring now to the drawings, FIGS. **1**–**4** are provided to facilitate an understanding of various aspects or features of the technology utilized in the present invention. FIG. **1** depicts a device **10** having a magnetic flux producing member **12** which may be a permanent magnet or electromagnet with magnetic poles **14** and **16** as shown. Pole pieces **18** and **20** are positioned adjacent respective poles **14** and **16** to provide a path for the magnetic flux of member **12**. Each pole piece **18** and **20** includes a respective pole piece end face **22** and **24**. As used throughout this specification it is understood that a pole piece, regardless of its shape or size, is preferably formed of soft iron, steel or some other magnetic material, with the preferred material being one which provides low reluctance, exhibits low hysterisis, and has a high magnetic flux density capability. Accordingly, the various pole pieces disclosed and described herein could likewise be of laminate type construction. Referring again to FIG. **1** an armature **26**, also formed of magnetic material, is shown with end faces **28** and **30** which are positioned and sized for being placed adjacent pole piece end faces **22** and **24**, such that when so positioned a substantially continuous low reluctance path **32** is provided for magnetic flux from north pole **14**, through pole piece **18**, through armature **26**, through pole piece **16**, and to south pole **16**. The magnetic flux traveling along such path **32** results in a force which tends to hold armature **26** in position adjacent pole piece end faces **22** and **24**. The resulting magnetic coupling or holding force F provided between adjacent pole piece end face **22** and armature end face **28**, and between adjacent pole piece end face **24** and armature end face **30**, can be approximated by the following equation:

$$F=B^2A/2\mu_0,$$

where B is the magnetic flux density passing through the adjacent end faces and where A is the surface area of the adjacent end faces. Assuming B uniform throughout flux path **32** and the area A of all end faces **22**, **24**, **28**, and **30** to be the same, the total holding force F_{T26} of armature **26** against pole pieces **18** and **20** will be:

$$F_{T26}=B^2A/\mu_0.$$

In FIG. **2** a device **40** having the same magnetic flux producing member **12** with magnetic poles **14** and **16** is shown. Pole pieces **42** and **44** are positioned adjacent respective pole faces **14** and **16** to provide two paths, as opposed to one above, for the magnetic flux of member **12**. In particular, pole piece **42** includes a first path portion **46** extending beyond a perimeter of north pole face **14** in one direction and a second path portion **48** extending beyond the perimeter of north pole face **14** in another direction. Similarly, pole piece **44** includes a first path portion **50** extending beyond the perimeter of south pole face **16** in one direction and a second path portion **52** extending beyond the perimeter of south pole face **16** in another direction. Each pole piece path portion **46**, **48**, **50**, **52** includes a respective end face. A first armature **54** is positionable adjacent the end faces of pole piece path portions **48** and **52** to provide a first magnetic flux path **56** and a second armature **58** is positionable adjacent the end faces of pole piece path portions **46** and **50** to provide a second magnetic flux path **60**. If the flux carrying area along flux paths **56** and **60** is the same as the flux carrying area along flux path **32** of FIG. **1**, the magnetic flux density along each flux path **56** and **60** will be one-half the magnetic flux density along flux path **32** of FIG. **1** because the same amount of flux is split between two like paths. The effect of dividing a given amount of magnetic flux along two like flux paths instead of along one flux path can be seen by examining the holding force on armature **54** as compared to the holding force on armature **26** of FIG. **1**. As already noted the magnetic flux density along path **56** will be one-half that along flux path **32** and thus the total holding force F_{T54} can be determined as:

$$F_{T54}=(B/2)2A/,\mu_0=B^2A/4\mu_0=F_{T26}/4.$$

It is therefore seen that dividing the same amount of magnetic flux along two flux paths rather than along one flux path reduces the magnetic holding or coupling force on an armature by one-fourth rather than one-half as might have been expected. This unexpected magnetic holding or coupling force differential, resulting from multiple flux paths, can provide advantageous properties in linear, reciprocating, and rotary motion devices.

Referring now to FIGS. **3**–**4**, the behavior of multiple magnetic flux sources arranged in parallel and series is described as compared to a single flux source. When identical flux sources or magnetic flux producing members **70** and **72** are positioned in parallel as shown in FIG. **3** with pole pieces **74** and **76** positioned adjacent the poles thereof to provide a flux path through armature **78**, the flux density B through armature **78** is double what the flux density would be if only one magnetic flux producing member were present. However, the field intensity H resulting from the two members **70** and **72** remains unchanged. This result holds true regardless of whether members **70** and **72** are both permanent magnets, are both electromagnets, or are a combination of one permanent magnet and one electromagnet. On the other hand, the properties resulting from magnetic flux producing members **80** and **82** arranged pole-to-pole in series between pole pieces **84** and **86**, with armature **88**, as shown in FIG. **4**, will vary depending on the nature of the members **80** and **82**.

In a first case, if both members **80** and **82** are permanent magnets, the magnetic field intensity H resulting from the two permanent magnets will be double that of one permanent magnet and the flux density B through armature **88** will be the same as what the flux density would be if only one permanent magnet type member were present.

In a second case, if both members **80** and **82** are electromagnets, the field intensity H again doubles and the flux density B increases according to the B/H curve or relationship of the pole piece **84**, **86** and armature **88** materials.

In a third case, if member **80** is a permanent magnet and member **82** is an electromagnet, the field intensity H again doubles, but, since the permanent magnet is near flux density saturation B_r the flux density can only be increased from B_r to B_{max} of the permanent magnet. At the point where electromagnet-type member **82** contacts permanent magnet-type member **80** the flux from the electromagnet-type member **82** couples with the flux of the permanent magnet-type member **82** until the flux density through permanent magnet-type member **80** reaches B_{max}. At that point additional flux from electromagnet-type member **82** does not contribute to the flux density along the flux path unless a bypass path around the permanent magnet-type member is provided. Use of such bypass paths will be described hereinbelow.

Controlling the flow of flux along both one and multiple flux paths is best described with reference to FIGS. **5**–**9**. In FIGS. **5** and **6** a permanent magnet device **90** including a permanent magnet **92** having pole pieces **94** and **96** positioned adjacent the pole faces thereof and an armature **98** completing a low reluctance path **104** from pole to pole is shown. Control coils **100, 102** are positioned along path **104**. When control coils **100, 102** are not energized, the magnetic flux of permanent magnet **92** follows path **104** as shown and armature **98** is held in place against pole pieces **94, 96** due to the resulting magnetic coupling forces. However, if coils **100, 102** are energized to provide an equal but opposing magnetic flux to that of permanent magnet **92**, the result is that the magnetic flux of permanent magnet **92** is blocked and no magnetic flux traverses the path which includes armature **98** and therefore no magnetic coupling forces act on armature **98** allowing it to fall away as shown in FIG. **6**. The permanent magnet device **90** is useful, although as will become apparent below, it is more advantageous to provide multiple flux paths rather than one.

In this regard, in FIG. **7** a permanent magnet device **110** includes a permanent magnet **112** having pole pieces **114, 116** positioned adjacent the pole faces thereof with armatures **118, 120** completing two low reluctance paths **130, 132** from pole to pole thereof. Control coils **122, 124** are positioned along path **130** and control coils **126, 128** are positioned along path **132**. The two paths provided are assumed to be of equal reluctance. With no coils energized, the magnetic flux of permanent magnet **112** divides equally along flux path **130** and along flux path **132** such that both armatures **118, 120** are subjected to a magnetic coupling force which holds them in place against pole pieces **114, 116**.

If coils **122, 124** are energized to provide a magnetic flux equal to but opposing the magnetic flux which travels along flux path **130** from permanent magnet **112** when no coils are energized, the result is that the magnetic flux of permanent magnet **112** is blocked and no magnetic flux traverses the path which includes armature **118** and therefore no magnetic coupling forces act on armature **118** allowing it to fall away as shown in FIG. **8**. Further, the magnetic flux traversing path **132** will be double that of when no coils are energized and therefore the magnetic coupling force on armature **120** will be about four (4) times that of when no coils are energized. By energizing coils **126, 128** in an opposing manner a similar result would be achieved such that armature **120** would fall away and such that the magnetic coupling force on armature **118** would be increased.

If coils **122, 124** are energized to provide a magnetic flux equal to and aiding the magnetic flux which travels along flux path **130** when no coils are energized, the result is that the control coils couple completely with the magnetic flux of permanent magnet **112** and no magnetic flux traverses the path which includes armature **120** and therefore no magnetic coupling forces act on armature **120** allowing it to fall away as shown in FIG. **9**. Further, the magnetic flux traversing path **130** will be double that of when no coils are energized and therefore the magnetic coupling force on armature **118** will be about four (4) times that when no coils are energized. By energizing coils **126, 128** in an aiding manner a similar result would be achieved such that armature **118** would fall away and the magnetic coupling force on armature **120** would be increased.

Based on the foregoing it is seen that the full magnetic coupling force available from the permanent magnet **112** can be switched from one path to another path by the application of one half the power it would require for a coil alone to produce the same magnetic flux along one path. The ability to easily switch the full magnetic coupling force from one path to another allows for efficient reciprocating, linear, and rotary motion and power conversion to be achieved.

The basic device utilized to achieve permanent magnet flux division and to control such permanent magnet flux division is defined herein as a "permanent magnet control component," various configurations of which are shown by way of example only, and not by way of limitation, in FIGS. **10**A–**10**F. FIG. **10**A depicts a permanent magnet control component **150** in which pole pieces **152** and **154** are positioned adjacent the pole faces of permanent magnet **156** to provide two magnetic flux paths extending from opposite sides of permanent magnet. Control coils **158** are positioned along each path. FIG. **10**B depicts a permanent magnet control component **160** in which pole pieces **162** and **164** are positioned against the pole faces of permanent magnet **166** to provide two spaced, adjacent magnetic flux paths extending from the same side of permanent magnet **166**. Control coils **168** are positioned along each path. FIG. **10**C depicts a permanent magnet control component **170** in which pole pieces **172** and **174** are configured so as to be positioned adjacent the pole faces of permanent magnet **176** so as to provide four flux paths, each flux path extending in a respective direction from permanent magnet **176**. Control coils **178** are also positioned along each path. FIG. **10**D depicts another four path configuration of a permanent magnet control component **180** in which pole pieces **182, 184** are configured and positioned to provide four flux paths for permanent magnet **186**, with a pair of spaced, adjacent flux paths extending from each side of permanent magnet **186**. Control coils **188** are positioned along each path. FIG. **10**E depicts another four path configuration of a permanent magnet control component **190** in which all four flux paths formed by pole pieces **192, 194** extend from one side of permanent magnet **196**. Again, control coils **198** are positioned along each flux path. FIG. **10**F still further depicts a four path configuration of a permanent magnet control component **200** in which pole pieces **202, 204** extend to one side of permanent magnet **206**, with pole piece **202** defining four flux paths and with pole piece **204** including a continuous return path. Control coils **208** are positioned along each path of pole piece **202**. Many other variations are possible.

Accordingly, it is seen that a variety of different configurations of permanent magnet control components in accordance with the present invention are possible. The important considerations for division of permanent magnet flux in such permanent magnet control components include extending each pole piece to or beyond the outer perimeter of the pole face of the permanent magnet in each region where a flux path is intended and assuring that the pole face of the permanent magnet intersects each of the flux paths. It is not necessary for each pole piece to include the same number of path portions extending beyond the perimeter of the respective permanent magnet pole face as noted with reference to permanent magnet control component **200**. Although two control coils are shown along each of the flux paths in FIGS. **10**A–**10**E it is apparent from component **200** in FIG. **10**F that one control coil positioned along a flux path is generally sufficient for purposes of the present invention. Further, although in the illustrated configurations each pole piece is positioned to contact a respective pole face of the permanent magnet, a small spacing between a pole piece and its adjacent permanent magnet pole face could be provided, particularly in applications where relative movement between the subject pole piece and the permanent magnet will occur.

In its simplest form a two path permanent magnet control component only requires one control coil positioned along one of the control paths to permit the magnetic flux of a permanent magnet to be switched between the two paths. In particular, a side view of such a two path component **210** is shown in FIG. **10**G and includes a permanent magnet **211** pole pieces **212** and **213**, and control coil **214** which may be connected to a suitable control circuit. By alternating energizing control coil **214** in an opposing manner and an aiding manner the magnetic flux of permanent magnet can be switched between the path including armature **215** and the path including armature **216**. When control coil **214** is energized in an opposing manner the magnetic flux will traverse the path including armature **215** and when control coil **214** is energized in an aiding manner the magnetic flux will traverse the path including armature **216**. Control coil **214** could also be placed at anyone of positions **217, 218**, or **219** to achieve the flux path switching. Further, in the two coils embodiment shown in FIG. **10**H control coil **217** is added. In such a device flux switching can be achieved by simultaneously energizing control coil **214** in a flux aiding manner and control coil **217** in a flux opposing manner, and by then simultaneously reversing the energization of the respective control coils **214** and **217**.

Reference is made to FIGS. **11**A–**11**F which depict devices similar to that of FIGS. **5**–**6** except that a bypass, formed of magnetic material, is provided in each case. In device **220** of FIGS. **11**A–**11**C a bypass **222** is provided from pole piece **224** to pole piece **226** and is located between permanent magnet **228** and control coils **230, 232**, with armature **234** located adjacent the ends of pole pieces **224, 226**. In FIG. **11**A with no coil energization, magnet flux components **236** and **237** travel as shown. When coils **230** and **232** are energized in an aiding or adding manner as in FIG. **11**B, the result is permanent magnet magnetic flux components **236** and **237** traveling as shown, and with the added magnetic flux component **238** from coils **230** and **232** also traveling as shown. Thus, in device **220** energizing the coils in an aiding manner results in an increased magnetic coupling force on armature **234**. In FIG. **11**C coils **230, 232** are energized in an opposing exceeding manner which results in permanent magnetic flux components **236** and **237** traveling as shown and excess magnetic flux component **238** traveling as shown. Thus, in device **220** energizing the coils in an opposing exceeding manner results in magnetic coupling force on armature **234**, albeit smaller than that in the aiding exceeding case.

In device **240** of FIGS. **11**D–**11**F a bypass **242** is provided between pole piece **244** and pole piece **246** but is located on an opposite side of permanent magnet **248** as compared to control coils **250, 252** and armature **254**. Permanent magnet flux components **256** and **257** are shown for no coil energization in FIG. **11**D. In FIG. **11**E the paths of permanent magnet flux components **256** and **257**, as well as excess coil magnetic flux **258**, are shown when coils **250, 252** are energized in an aiding exceeding manner. In FIG. **11**F the path of each magnetic flux component **256, 257**, and **258** is shown when coils **230, 232** are energized in an opposed exceeding manner.

FIGS. **12**A–**12**E depict a device **270** similar to that shown in FIGS. **7**–**9** except that bypasses **272** and **274** are provided from pole piece **276** to pole piece **278**. Bypass **272** is located between permanent magnet **280** and control coils **282, 284** and bypass **274** is located between permanent magnet **280** and control coils **286, 288**. Armatures **290** and **292** are also provided. When no coils are energized permanent magnet magnetic flux components **294, 296, 298**, and **300** travel as shown in FIG. **12**A.

If coils **282, 284** are energized in an opposing manner permanent magnet flux components **295, 297**, and **299** travel as shown, with no flux component traversing the path which includes armature **290** and therefore no magnetic coupling force acting thereon. This would be the case for when coils **282, 284** are energized to the level where the coils magnetic flux just blocks, but does not exceed, the magnetic flux component **294** (FIG. **12**A) from permanent magnet **280**. If, however, coils **282, 284** are energized in an opposed exceeding manner an excess coil magnetic flux component **301** is produced which travels a path including armature **290** and bypass **272** results as shown in FIG. **12**C.

Coils **286, 288** may be energized in an aiding manner such that all permanent magnet magnetic flux travels along the path which includes armature **292** as shown in FIG. **12**D. If coils **286, 288** are energized in excess of the level of FIG. **12**D then the excess magnetic flux component **304** traverses the path which includes armature **292** and bypass **274** as shown in FIG. **12**E, thereby increasing the magnetic coupling force on armature **292** as compared to FIG. **12**D. The advantage of incorporating such bypasses into permanent magnet control components in certain applications will become apparent below.

Reciprocating Motion

As mentioned above, controlling the path of magnetic flux from a permanent magnet can be useful in a variety of applications such as achieving reciprocating motion. In this regard, if the device **110** of FIGS. **7**–**9** is modified such that armatures **118** and **120** are fixed to a sliding shaft **320** as shown in FIGS. **13**A–**13**C, and if the distance between the armatures is greater than the end to end length of pole pieces **114, 116**, limited linear motion in two directions (left and right in FIGS. **13**A–**13**C), and therefore linear reciprocating motion, can be achieved by the timed, alternate delivery of electrical signals to control coils **122, 124** and control coils **126, 128**. By way of example, FIG. **13**A represents the position of shaft connected armatures **118, 120** when coils **122, 124** are energized in an opposing manner to block the flux of permanent magnet **112** such that all magnetic flux traverses path **132** as shown and such that the resulting magnetic coupling force acts to the left as indicated by arrow **322**. As shown in FIG. **13**B when coils **122, 124** are de-energized the magnetic flux from permanent magnet **112** can again travel along path **130** through armature **118**. However, due to the air gap **324** between armature **118** and pole pieces **114, 116** the reluctance along path **130** will be significantly greater than the reluctance along path **132**. Accordingly, the amount of magnetic flux which flows along path **130** will be less than the amount of magnetic flux which flows along path **132** such that the magnetic coupling force on armature **118** acting to the right will be significantly less than the magnetic coupling force on armature **120** acting to the left as shown by arrows **326** and **328**, which arrows are sized to represent the strength of the respective directional force. FIG. **13**C represents the position of shaft connected armatures **118, 120** after coils **126, 128** are energized in a manner to oppose the flux of permanent magnet **112** such that all flux traverses path **130** and the resulting magnetic coupling force on armature **118**, depicted by arrow **330**, moves the shaft **10** connected armatures **118, 120** to the right.

Control coils **122, 124** and **126, 128** could also be energized in a flux aiding manner to achieve the same result. In such a device, FIG. **13**A would represent coils **126, 128** energized to aid magnetic flux along path **132**, FIG. **13**B would again represent no coils energized, and FIG. **13**C would represent coils **122, 124** energized to aid magnetic flux along path **130**.

Thus, by alternatingly energizing and de-energizing control coils **122, 124** and **126, 128** a linear reciprocating motion of shaft connected armatures **118, 120** may be achieved. Further, such reciprocating motion may be achieved by energizing the coils in either an opposing or aiding manner. The magnetic coupling force exerted on a given armature when 20 the control coils are energized to establish all magnetic flux along a single path which includes that armature is significantly greater than the magnetic coupling force which would be exerted on such armature by an identical energization of the control coils in the absence of the permanent magnet. This is demonstrated with reference to FIG. **14** which depicts a reciprocating device **340** in which only coils or electromagnets are utilized. As shown armatures **342** and **344** are connected by shaft **346**, and each armature **342, 344** includes a respective U-shaped pole path piece **348, 350** which pole path pieces are mechanically connected by a non-magnetic material **352**. Each pole path piece **348** and **350** has respective control coils **354, 356** and **358, 360** positioned therealong. By comparison with the device of FIGS. **13**A–**13**C, if coils **358, 360** of device **340** are energized to cause magnetic flux flow in either direction, clockwise or counterclockwise, along path **362**, the amount of electrical energy which would be required in order to achieve the same magnetic coupling force on armature **344** as achieved on armature **120** above in FIG. **13**A would be twice that delivered to coils **122, 124** or **126, 128** in FIG. **13**A. It is therefore demonstrated that by controlling or switching the flow of magnetic flux from a permanent magnet between at least two different paths results in greater coupling forces per unit of input electrical energy, and therefore that such control or switching will enable more work to be achieved per unit of input electrical energy.

As described above, if a coil is energized beyond the point where the magnetic flux produced by the coil aiding the amount of the permanent magnet's flux that is either opposed or aided, the extra magnetic flux needs a low reluctance path between the poles of the coil that produces the excess magnetic flux. If a complete low reluctance path is not provided for the excess magnetic flux there is little potential for taking advantage of the excess magnetic flux in terms of producing additional magnetic coupling forces. The path for such excess flux cannot be through a permanent magnet member. In assemblies which include an armature on each path, the armature will provide the necessary low reluctance path. Referring to FIG. **15**, various components of the magnetic flux in device **110** (FIGS. **7**–**9**) are depicted by numerals **380, 382**, and **384** for the case when coils **122, 124** are energized to oppose the magnetic flux of permanent magnet **112** in an amount which exceeds the level of magnetic flux which permanent magnet **112** would cause to flow through armature **118** when no coils are energized. FIG. **15** is likewise representative of the case when coils **126, 128** are energized to aid the magnetic flux of permanent magnet **112** in an amount which exceeds the level of magnetic flux which permanent magnet **112** would cause to flow through armature **118** when no coils are energized. In particular, magnetic flux component **380** represents the magnetic flux of permanent magnet **112** which normally flows through the path including armature **120**; magnetic flux component **382** represents the magnetic flux of permanent magnet **112** which is diverted by the opposing field of coils **122, 124** so as to traverse the path which includes armature **120**; and magnetic flux component **384** represents the magnetic flux produced by coils **122, 124** which is in excess of the diverted magnetic flux **382**. As shown, the excess magnetic flux **384** produced by coils **122, 124** traverses the path which includes armature **120** and bypasses permanent magnet **112** so as to also traverse the path which includes armature **118**. Thus, the excess magnetic flux produced by coils **122, 124** adds to the permanent magnet flux traversing the path which includes armature **120**, thus increasing the magnetic coupling force on armature **120**, while at the same time providing a magnetic coupling force on armature **118**.

In a reciprocating device where armatures **118** and **120** are connected by shaft **320** as shown in FIGS. **13**A–**13**C and again in FIG. **16**, excess magnetic flux **384** will increase magnetic coupling force **390** on armature **120** acting to the left. However, because such excess flux **384** also traverses the path which includes armature **118**, such excess magnetic flux **384** also results in a magnetic coupling force **392** on armature **118** which acts to the right. Even though excess magnetic flux **384** traversing the path which includes an armature **118** has an opposite polarity to that which would traverse the path due to permanent magnet **112**, the magnetic coupling force on armature **118** still acts to the right because armature **118** is not polarity sensitive, that is, armature **118** will be attracted regardless of the direction of the magnetic flux traversing the path. The overall effect is that a resultant force which is the difference between force **390** and force **392** will act on the shaft-connected armatures **118, 120**. However, if armatures **118** and **120** were formed by permanent magnets having polarities as shown at the top and bottom of such armatures, the force acting on each armature would be in the same direction and therefore additive.

In this regard reference is made to FIG. **16**B in which a two path device **371** having four control coils **373, 375, 377** and **379** is shown with the illustrated armatures being formed by permanent magnets **381** and **383** having polarities as shown. With no coils energized both permanent magnet armatures **381** and **383** are attracted to the ends of pole pieces **385** and **387**. With coils **373 375** energized in an opposing manner and coils **377, 379** energized in an aiding manner, the attractive force on permanent magnet armature **383** will generally increase and the attractive force on permanent magnet armature **381** will generally decrease. This is demonstrated with reference to the graph of FIG. **16**C which depicts a graph of the current flowing in the control coils on the x-axis verses the magnetic flux in gauss on the y-axis with line **389** representing the flux along the aiding side of device **371** and line **391** representing the flux along the opposing side of device **371**. As shown the magnetic flux on the coil opposing side decreases as the coil current increases and passed through zero at point **393**. After point **393** reverse magnetic flux begins to be produced and would result in a repelling force on permanent magnet armature **381**. In some applications particularly those where permanent magnet armatures and rotors are not utilized, it will be critical to recognize point **393** so that reverse magnetic flux is not produced. In this regard reference is made to FIGS. **16**D and **16**E in which use of Hall Effect switches **401** and **403** is incorporated to enable controlling the coil energizing current in situations where it is desirable to prevent reverse magnetic flux. As shown small bypasses **405** and **407** are provided with Hall Effect switches **401** and **403** disposed in gaps therealong, the switches being connected to control circuit **409**. As the flux traveling along the bypass path falls to zero the Hall Effect switch can be utilized to prevent further energization of the control coils so that no reverse flux is created.

Another embodiment of a device **400** which would provide reciprocating motion is shown in FIGS. **17**A–**17**D in which a permanent magnet control component **402** having two flux paths may is provided. A first pole piece **404** has two spaced, adjacent path portions **406** and **408** extending beyond the perimeter of the pole face of permanent magnet

410, and a second pole piece **412** includes only one continuous portion **414** extending beyond the perimeter of the pole face of permanent magnet **410**, each path portion **406** and **408** of pole piece **404** being substantially aligned with at least a part of portion **414** of pole piece **412**. Control coil **416** is positioned along pole piece path portion **406** and control coil **418** is positioned along pole piece portion **408**. An armature **420** is positioned in the region between pole piece path portions **404**, **406** and pole piece portion **414** and is free to slide from side to side as shown by arrows **422** and **424**. A front view of component device **400** with no coils energized and armature **420** at a mid-point depicts flux flowing from the north pole face of permanent magnet **410**, through each of pole piece path portions **406** and **408**, through armature **420**, and returning to the south pole face through pole piece portion **414**. Thus, the magnetic flux divides equally along two paths. If coil **416** is energized in an aiding manner, or if coil **418** is energized in an opposing manner, all or a majority of the permanent magnets'magnetic flux can be made to flow through pole piece portion **406** such that a resulting magnetic coupling force on armature **420** causes it to move to the left as shown in FIG. **17**C.

Likewise, if control coil **416** is energized in an opposing manner, or if control coil **418** is energized in an aiding manner, all or a majority of the permanent magnet flux can be made to flow through pole piece path portion **408** such that a resulting magnetic coupling force on armature **420** causes it to move to the right as shown in FIG. **17**D. Accordingly, by alternately energizing and de-energizing coils **416** and **418** a reciprocating motion of armature **420** may be achieved.

Linear Motion

Referring now to FIGS. **18**A–**18**E, linear motion in accordance with the present invention is described. In particular, a permanent magnet control component **440** including a permanent magnet **442** with a pole piece **444** positioned against the north pole face thereof and a pole piece **446** positioned against the south pole face thereof is shown in an exploded view in FIG. **18**A and assembled in FIG. **18**B. Pole piece **444** includes five path portions **448**A–**448**E which extend beyond the perimeter of the north pole face of permanent magnet **442** to one side thereof and at respective positions along the length thereof, each path portion **448**A–**448**E including a respective control coil **450**A–**450**E positioned therearound. Pole piece **446** includes one portion **452** extending beyond the perimeter of the south pole face of permanent magnet **442** to the one side thereof, which portion **452** extends along the entire length of permanent magnet **442**. A plurality of armatures **454** define a path of relative movement between permanent magnet control component **440** and such armatures **454**, and by providing timed energization of given control coils **450**A–**450**E such relative movement can be achieved. The sequence of side views depicted in FIGS. **18**C–**18**E illustrate such relative movement, with coils **450**A, **450**C and **450**E being energized in an opposing manner simultaneously in FIG. **18**C, with coils **450**A and **450**D being energized simultaneously in an opposing manner in FIG. **18**D, and with coils **450**B and **450**D being energized simultaneously in an opposing manner in FIG. **18**E. In FIG. **18**C, magnetic flux will only flow along path portions **448**B and **448**C of pole piece **444** causing resultant magnetic coupling forces depicted by arrows **456**, **458** which act to move permanent magnet control component **440** to the left, assuming armatures **454** are fixed. Similarly, due to the timing of subsequent coil energization resultant magnetic forces depicted by arrows **460**, **462** in FIG. **18**D and arrows **464**, **466** in FIG. **18**E act to continue movement of permanent magnet control component **440** to the left. Thus, if permanent magnet control component **440** were fixed to a device or structure, controlled movement of the device or structure along the path defined by armatures **454** could be achieved. Conversely, if permanent magnet control component **440** were fixed and armatures **454** were located on a device or structure, controlled movement of the device or structure could also be achieved. It is also easily recognized that by varying the coil energization sequence and timing relative movement in the opposite direction can be achieved. Further, if the permanent magnet was doughnut shaped and the armatures were arranged in a circumferential pattern, rotary motion would likewise be achievable.

Rotary Motion

One embodiment of a rotary motion device or motor **500** which incorporates various permanent magnet flux control aspects of the present invention is shown in the exploded view of FIG. **19** and in the partial assembled view of FIG. **20**. Motor **500** includes a rotor assembly which includes a shaft **502** and associated upper bearing **504**, a non-magnetic disk member **506** mounted for rotation with shaft **502**, and a rotor pole piece **508** which is mounted for rotation with disk member **506** such as by the use of screws **510**. Rotor pole piece **508** includes a ring-shaped portion having two inwardly extending magnetic flux path portions **512**A and **512**B. A stator assembly of motor **500** includes a doughnut or ring-shaped permanent magnet **514** having an upwardly directed north pole face positioned adjacent and in close proximity to rotor pole piece **508**, and a downwardly directed south pole face positioned adjacent and in contact with a stator pole piece **516**. Stator pole piece includes a ring-shaped portion having five inwardly projecting path portions **518**A–**518**E. Each path portion includes a respective winding post **520**A–**520**E extending therefrom and having a respective control coil **522**A–**522**E wound thereon. Stator pole piece faces **524**A–**524**E are positionable on respective winding posts **518**A–**518**B and, as shown in the partial assembly of FIG. **20**, are substantially aligned with the top surface of permanent magnet **514** so as to be positionable adjacent rotor path portions **512**A and **512**B when aligned therewith. Each of winding posts **518**A–**518**E and stator pole piece faces are formed of magnetic material, and although shown as separate pieces, an integral, one piece stator could be formed with similar winding posts and pole piece faces machined thereon. Lower bearing **526** is also shown.

FIGS. **21**A–**21**E illustrate top views of the partial assembly of FIG. **20** with magnetic flux shown. In FIG. **21**A magnetic flux travel when none of coils **522**A–**522**E are energized is depicted. Disregarding leakage flux, due to the low reluctance path provided by rotor pole piece path portions **512**A and **512**B, the majority of magnetic flux from the north pole face of permanent magnet **514** will travel radially inward along one of such path portions before passing downward through the stator assembly and returning to the south pole face of permanent magnet **514**. It is noted that rotor pole piece **508** includes two path portions and stator pole piece **516** includes five path portions such that rotor pole piece path portions **512**A and **512**B will always be skewed relative to the stator pole piece faces **524**A–**524**E. Only one rotor pole piece path portion can directly align with a stator pole piece face at a given time. By alternatingly energizing the control coils of each of the stator pole piece paths, rotary motion of the rotor may be achieved.

In particular, referring to FIGS. **21B–21D**, an energizing sequence which results in such rotary motion is described. In FIG. **21B**, control coils **522A** and **522C** are energized in a permanent magnet flux opposing manner. Permanent magnet magnetic flux traveling along rotor pole piece path portion **512A** tends to traverse to stator pole piece face **524B** causing a magnetic coupling force indicated by arrow **526**. Likewise, permanent magnet flux traveling along rotor pole piece path portion **512B** tends to traverse to stator pole piece face **524D** causing a magnetic coupling force indicated by arrow **528**. The result is rotation of rotor pole piece **508** in a clockwise direction as indicated by arrow **530**.

Referring to FIG. **21C**, just after rotor pole piece path portion **512B** is no longer aligned with stator pole piece face **524D**, control coil **522C** is de-energized and control coil **522D** is energized in an opposing manner such that the permanent magnet flux traveling along rotor pole piece path **512B** tends to traverse to stator pole piece face **524E** resulting in magnetic coupling force indicated by arrow **532**. Control coil **522A** remains energized such that a magnetic coupling force indicated by arrow **534** results. Accordingly, clockwise rotation of rotor pole piece **508** is continued.

In FIG. **21D**, just after rotor pole piece path portion **512A** is no longer aligned with stator pole piece face **524B**, control coil **522A** is de-energized and control coil **522B** is energized in a permanent magnet magnetic flux opposing manner such that the permanent magnet magnetic flux traveling along rotor pole piece path **512A** tends to traverse to stator pole piece face **524C** such that a magnetic coupling force indicated by arrow **536** results. Control coil **522D** remains energized such that a magnetic coupling force indicated by arrow **538** results, and clockwise rotation of rotor pole piece **508** is continued.

As shown in FIG. **21E**, just after rotor pole piece path portion **512B** is no longer aligned with stator pole piece face **524E**, control coil **522D** is de-energized and control coil **522E** is energized in a permanent magnet magnetic flux opposing manner such that the permanent magnet magnetic flux traveling along rotor pole piece path **512B** tends to traverse to stator pole piece face **524A** such that a magnetic coupling force indicated by arrow **540** results. Control coil **522B** remains energized such that a magnetic coupling force indicated by arrow **542** results, and clockwise rotation of rotor pole piece **508** is continued.

Thus, by alternating energizing and de-energizing control coils **522A–522E**, in a predetermined timed sequence based upon rotation of the rotor assembly, continued rotation movement of rotor pole piece **508** may be achieved. Such an energization/de-energization scheme can be achieved utilizing circuitry common in the art, such as the control circuitry described in Applicant's U.S. Pat. Nos. 5,463,263 and 5,455,474, as well as various of the circuit configurations described herein below. Referring now to FIG. **22**, an assembled view of rotary motor **500** is shown including a housing or cover formed by an upper housing member **544** and a lower housing member **546**, with portions of each housing member cut away to expose motor structure described above. It is recognized that such housing members **544** and **546** should be formed of a non-magnetic material, and likewise that motor shaft **502** and bearings **504**, **526** should be formed of a non-magnetic material.

In another embodiment, a rotary motion device or motor **580** in accordance with the present invention is shown in an exploded perspective view in FIG. **23** and in an assembled perspective view in FIG. **24**. Two spaced permanent magnets **582** and **584** are positionable between stator pole pieces **586** and **588**. Stator pole piece **586** includes two path portions **590A** and **590B** extending away from permanent magnets **582**, **584** in opposite directions. Likewise, stator pole piece **588** includes two path portions **592A** and **592B** extending away from permanent magnets **582**, **584** in opposite directions and alignable respectively with stator pole piece path portions **590A** and **590B**. Control coils **594**, **596**, **598**, and **600** are each positioned along a respective stator pole piece path portion as shown. A non-magnetic shaft **602** includes a pair of like elongated rotor members **604** and **606**, formed of magnetic material, mounted at spaced locations thereon and being angularly oriented with respect to each other, shaft **602** passing between spaced permanent magnets **582** and **584**. Non-magnetic end cap members **608** and **610** are attachable to the ends of stator pole pieces **586** and **588** and are configured for receiving shaft **602** and respective bearings **612** and **614**.

The ends of respective stator pole pieces **506** and **508** are configured for a given desired coupling relationship with rotor members **604** and **606**. For example, as shown in the exemplary end views of FIGS. **25A** and **25B**, with end cap **608** removed, the end of stator pole piece **586** may include an arcuate portion **616** which is configured to create a variable reluctance air gap **618** with elongate rotor member **604**. The end of stator pole piece **588** includes an arcuate portion **620** which is likewise configured to create a variable reluctance air gap **622** with rotor member **604**. In particular, portion **618** includes a circumferential curvature which has a center point offset below the axis of rotation of shaft **602** and rotor member **604** as indicated by circle **624** shown in shadow. Similarly, portion **620** includes a circumferential radius of curvature which has a center point offset above the axis of rotation of shaft **602** and rotor member **604**. When magnetic flux is passing along the path which includes a given end of the assembly, maximum coupling between the rotor member and stator pole pieces occurs when the rotor is positioned as shown in FIG. **25B**. Accordingly, the illustrated rotor member and stator pole piece configurations in and of themselves do not provide any skewing to the direction of rotation of the rotor assembly.

In this regard, various configurations for the rotor and ends of the stator pole pieces are shown in the end views of FIGS. **26–28**, which configurations provide skewing the direction of rotation. In particular, in device **620** of FIG. **26** a rotor member **622** having notches **624** and **626**, which notches provide for greater magnetic coupling with the stator pole pieces **628** and **630** at corners **632** and **634** such that rotation is skewed in the clockwise direction. If notches were instead located at corners **632** and **634**, skewed rotation in the counterclockwise direction would be the result. In device **620** such counterclockwise rotation could also be achieved by removing rotor **622** from shaft **636**, flipping it end to end, and replacing it on shaft **636**.

In the device **640** of FIG. **27**, a portion **642** of the arcuate end portion of stator pole piece **644** is removed and a portion **646** of the arcuate end portion of stator pole piece **648** is removed. This configuration results in greater magnetic coupling between rotor member **650** and stator pole piece **644** at corner **652**, and greater magnetic coupling between rotor member **650** and stator pole piece **648** at corner **654**, such that rotation is skewed in the counterclockwise direction. Clockwise rotation could be achieved by instead modifying the opposite side of stator pole pieces **644** and **648**.

FIG. **28** depicts an end view of a device **660** in which the axis **662** of the arcuate end portion of upper stator pole piece **664** and lower stator pole piece **666** is placed at an angle A as shown. This configuration creates an unequal variable reluctance air gap where opposite corners of rotor member **668** are closer to stator pole pieces **664** and **666**. Further, the angle at which maximum magnetic coupling between rotor member **668** and stator pole pieces **664** and **666** occurs is retarded by angle A. Rotation would be in the counterclockwise direction for the illustrated configuration.

Referring again to motor **580** of FIGS. **23–25**, rotary motion of such device is depicted in the end views of FIG. **29**A–**29**D. In each end view the end cap has been removed to show rotation of the rotor members and in each of FIGS. **29**A–**29**D an end view depicting rotor member **604** and an end view depicting rotor member **606** are shown side-by-side. In FIG. **29**A, rotor member **604** is defined as being at zero degrees and rotor member **606** is defined as being at ninety degrees. Control coils **594**, **598** are energized in a permanent magnet magnetic flux aiding manner such that no magnetic flux passes through stator pole piece path portions **590**B and **592**B. This allows rotor member **606** to move out of its ninety degree position and the magnetic coupling between rotor member **604** and stator pole piece path portions **590**A and **592**A will cause rotation to the position shown in FIG. **29**B and then FIG. **29**C. When rotor member **604** reaches the ninety degree position shown in FIG. **29**D control coils **594**, **598** are de-energized and control coils **596**, **600** are energized in a permanent magnet magnetic flux aiding manner causing rotation to continue due to the magnetic coupling between rotor member **606** and stator pole piece path portions **590**B and **592**B. Thus, by alternatingly energizing the control coils of each path with every ninety degree rotation of rotor members **604** and **606**, continuous rotary motion is achieved.

The initial direction of rotation can be controlled by the circuit means used to energize control coils **594**, **598** and **596**, **600**, which circuit means includes circuitry for detecting the angular position of the rotor members. In particular, if rotor members **604** and **606** are at rest in the position shown in FIG. **29**A, and coils **594**, **598** are energized in an aiding manner, rotation may be clockwise or counterclockwise. If the desired direction is clockwise but upon energization of coils **594**, **598** the rotor members begin to move counterclockwise, the detection circuitry will immediately de-energize coils **594**, **598** and energize coils **596**, **600** so that the counterclockwise direction is achieved.

Further, bypasses around permanent magnets **582** and **584** could be provided in rotary motion device **580**, such as those shown in FIG. **12**, and rotor members **604** and **606** could be formed by permanent magnets so as to take advantage of energizing the control coils in an exceeding manner.

A third embodiment of a rotary motion device or motor **650** is shown in the exploded partial perspective view of FIG. **30** and in the assembled partial perspective view of FIG. **31**. In motor **650** the stator assembly includes a control component **651** including a permanent magnet **652** having a stator pole piece **654** positioned adjacent one pole face thereof and a stator pole piece **656** positioned adjacent the opposite pole face thereof. Stator pole piece **654** includes a path portion **658**A extending to one side of permanent magnet **652** and a path portion **658**B extending to the one side thereof and spaced from first path portion **658**A. Control coils **660** and **662** are positioned along respective stator pole piece path portions **658**A and **658**B. Likewise, stator pole piece **656** includes path portions **664**A and **664**B which extend in a similar manner therefrom so as to be aligned with stator path portions **658**A and **658**B respectively. Control coils **666** and **668** are positioned along respective stator pole piece path portions **664**A and **664**B. Positioned opposite and facing control component **651** is a like control component **670** including permanent magnet **672**, stator pole piece **674** with path portions **676**A and **676**B having respective control coils **678** and **680**, and stator pole piece **682** with path portions **684**A and **684**B having respective control coils **686** and **688**. The end of each of the pole piece path portions **658**A, **658**B, **664**A, **664**B, **676**A, **676**B, **684**A, and **684**B is of a generally arcuate configuration.

A rotor assembly of motor **650** includes a non-magnetic shaft **700** having a permanent magnet rotor member **702** mounted thereon for rotation therewith. Permanent magnet rotor member **702** is generally ring-shaped and segmented to include distinct north and south pole faces which reverse about every ninety degrees therearound. When assembled the top and bottom surfaces of permanent magnet rotor member **702** align with pole pieces **654**, **656**, **674**, and **682** of the stator assembly and are preferably configured such that a minimal gap between the outer surface of permanent magnet rotor member **702** and the arcuate surfaces of the pole piece path portions is provided.

Rotation of device **650** can be achieved by controlled, timed energizing and de-energizing of control coils **660**, **662**, **666**, **668**, **678**, **680**, **686**, and **688**. Exemplary rotation is demonstrated with reference to the top views of FIGS. **32**A–**32**B which depict counterclockwise rotation of permanent magnet rotor member **702** through one-hundred eighty degrees. In FIG. **32**A stator pole piece path portion **658**A of component **651** is active and stator pole piece path portion **658**B is not active, which may be achieved by energizing control coil **660** in a permanent magnet magnetic flux aiding manner or by energizing control coil **662** in a permanent magnet magnetic flux opposing manner. Stator pole piece path portion **676**B of component **670** is active and stator pole piece path portion **676**A is not active, which may be achieved by energizing control coil **680** in a permanent magnet magnetic flux aiding manner or by energizing control coil **678** in a permanent magnet magnetic flux opposing manner. Thus, portions **690** and **692** of permanent magnet rotor member **702**, which both have a north magnetic polarity, will be repelled by the north polarity of stator pole piece path portions **658**A and **676**B aligned therewith. Portions **694** and **696** of permanent magnet rotor member **702**, both of which have a south magnetic polarity, will be attracted to the active path portions **658**A and **676**B. At the instant that rotor member portion **694** becomes aligned with stator pole piece path portion **658**A, as shown in FIG. **32**B, all coils are de-energized such that all pole piece path portions will be active as shown. Pole piece path portions **658**B and **676**A are then kept active while pole piece path portions **658**A and **676**B are made inactive. This is achieved by energizing control coils **662** and **678** in a permanent magnet magnetic flux aiding manner or by energizing control coils **660** and **680** in a permanent magnet magnetic flux opposing manner. Rotor member portions **690** and **692** will again be repelled by the north polarity of path portions **658**B and **676**A aligned therewith such that rotation of permanent magnet rotor **702** is continued. In FIG. **32**D all coils are shown de-energized when rotor portion **692** aligns with pole piece path portion **658**A. By continuing this timed sequence of energization and de-energization of the control coils, continued rotary movement is achieved. As explained above, the initial direction of rotation can be controlled by circuit means which detects the initial direction of permanent magnet rotor **702** and immediately alters the coil energization scheme if the initial direction is incorrect.

A side view of assembled motor **650** is shown in FIG. **33** and includes an upper housing or enclosure portion **710**, a bottom housing portion **712**, upper bearing **714**, and a lower bearing **716**.

A fourth embodiment of a rotary motion device or motor **740** is illustrated in FIGS. **34**–**39**. Motor **740** includes five stator control components **742**A–**742**E positioned around a ring shaped permanent magnet rotor member **744** (FIG. **36**). As shown with reference to component **742**A in FIG. **37** each stator component **742**A includes a permanent magnet **746**A with an upper pole piece **748**A positioned adjacent one pole face thereof and a lower pole piece **750**A positioned adjacent the opposite pole face thereof. Control coils **752**A, **754**A are positioned along respective pole pieces **748**A, **750**A. A bypass **756**A extends from pole piece **748**A to pole piece **750**A and is positioned between permanent magnet **746**A and control coils **752**A, **754**A. Alternatively, bypass **756**A could be provided on the opposite side of permanent magnet **746**A as shown in FIG. **38**. Although not shown, it is anticipated that permanent magnet rotor member **744** would be mounted on an axis for rotation therewith and that a motor housing or enclosure could be provided, such as shown in relation to motor **650** of FIG. **33**.

Referring to the top views of FIGS. **39**A–**39**D, rotary motion of rotor member **744** is depicted by the sequence of views. Regions **770** and **772** in FIGS. **39**A–**39**D represent the magnetic north regions of the top of permanent magnet rotor **744**. In FIG. **39**A control coils **752**E and **752**C are energized in a permanent magnet aiding and exceeding manner such that regions **770** and **772** of permanent magnet rotor **744** are repulsed by components **742**E and **742**C while permanent magnet motor regions **774** and **776** are attracted by components **742**E and **742**C. The resultant coupling forces act to move permanent magnet rotor in a counterclockwise direction to the location shown in FIG. **39**B. Just after permanent magnet rotor region **772** passes the point shown in FIG. **39**C, control coil **752**B is energized in a permanent magnet aiding and exceeding manner, while control coils **752**E and **752**C also remain energized, and counterclockwise rotation of permanent magnet rotor **744** is continued. Just after permanent magnet rotor region **772** passes by control component **742**C control coil **752**C is de-energized, while control coils **752**E and **752**B remain energized, so as to continue counterclockwise rotation. Then, just after permanent magnet rotor region **770** reaches the location shown in FIG. **39**D control coil **752**D is energized in a permanent magnet flux aiding and exceeding manner, while coils **752**E and **752**B remain energized, so as to continue counterclockwise rotation. Thus, as in the other embodiments, repeated and timed energization and de-energization of the control coils produces the desired rotational movement.

In terms of controlling the energization of coils in the devices described above, various electronic control circuit/switching means and electromechanical control circuit/switching machines are depicted in FIGS. **40**–**44**. In circuit **800** of FIG. **40** a given coil **802** is placed in series between an electrical energy source **804** and a power mosfet **806**. An LED **808** is connected to electrical energy source **804** through resistor **810** and is positioned to impinge upon a phototransistor **812** which is connected in series with resistor **814**. A control input of mosfet **806** is connected between phototransistor **812** and resistor. Accordingly, when LED **808** activates phototransistor **812** the voltage drop across resistor **814** activates, or turns ON, mosfet **806** and coil **802** is energized. Timed energization of coil **802** is provided by mounting an interrupter **816**, such as shown in FIG. **42**, to the shaft **816** of the motor device to be controlled, such that as interrupter **814** rotates with shaft **816** coil **802** is alternately energized and de-energized. In a device with a plurality of coils a corresponding plurality of LED/photoresistor pairs may be provided.

In circuit **820** of FIG. **41** a coil **822** is positioned between electrical energy source **824** and power mosfet **826**. A hall switch **828** is connected in series with resistor **830**. Hall switch **828** is also connected to the control input of mosfet **826** through resistor **832**. In a given device hall switch **828** would be positioned to react to a change in magnetic flux so as to control the ON/OFF switching of mosfet **826**, and thus the alternate energization and de-energization of coil **822**.

In FIG. **43** a circuit **840** for controlling two coils in an opposite manner is provided such that when coil **842** is energized coil **844** is de-energized, and such that when coil **842** is de-energized coil **844** is energized. Both coils **842** and **844** are connected in series between electrical energy source **846** and respective power mosfets **848** and **850**. An LED **852** and phototransistor **854** arrangement is provided, LED connected in series with resistor **856** and phototransistor connected in series with resistor **858**. When LED **852** turns phototransistor **854** ON the voltage drop across resistor **858** turns mosfet **848** ON and coil **842** is energized. At that time the voltage applied at the control input of mosfet **850** will be low and therefore mosfet **850** will be OFF and coil **844** will be de-energized. When interrupter **814** blocks LED **852**, phototransistor **854** is turned OFF and mosfet **848** is likewise turned OFF. The control input of mosfet **850** is therefore pulled high through resistor **860** and mosfet **850** is turned ON such that coil **844** is energized.

In a FIG. **44** a system **870** including member **872** mounted on rotating shaft **874** is provided, with the left side of member **872** being alternately conductive at **876** and non-conductive at **878**. Coils **880** and **882** are connected to respective brushes **884** and **886** which are positioned to contact member **872** during rotation thereof. Member **872** is connected through brush **890** to power supply **888**. Thus, coils **880** and **882** will alternatingly be energized and de-energized as the respective brushes thereof contact the conductive and non-conductive portions of member **872**.

Any of such circuit means, variations thereof, or other circuit means may be used to provide the timed energization of the control coils in the various embodiments of the present invention.

From the preceding description of the illustrated embodiments, it is evident that the objects of the invention are attained. Although the invention has been described and illustrated in detail, it is to be clearly understood that the same is intended by way of illustration and example only and is not to be taken by way of limitation.

For example, although the magnetic flux control techniques of the present invention have been discussed as applicable mainly to various motive applications, such magnetic flux control techniques are also useful in static applications.

Power Conversion

Referring to FIGS. **45**A–**45**C there is shown the permanent magnet device **900** of FIGS. **45**A–**45**C that has two magnetic flux paths provided by rectangular pole piece **902** which includes upper portion **904** and lower portion **906** each positioned against a respective pole face of permanent magnet **910**. Unlike the device of FIGS. **7**–**9**, fall away armatures are not provided. Rather, fixed armatures in the form of integral pole piece portions **912** and **914** extend from upper portion **904** to lower portion **906** completing the two flux paths in a permanent manner. Control coils **916**, **918** are provided along one flux path and control coils **920**, **922** are provided along the other flux path, such control coils acting as primary windings in device **900**. One coil **924** is positioned around pole piece portion **912** and another coil **926** is positioned around pole piece portion **914**, such coils **924**, **926** acting as secondary windings in device **900**.

In FIG. **45**A coils **916**, **918** are energized in a permanent magnet magnetic flux aiding manner so as to couple with all the magnetic flux of permanent magnet **910**. All magnetic flux flows along path **930** as shown and thus couples with coil **924**. In FIG. **45**B no coils are energized and the permanent magnet magnetic flux splits evenly between paths **930** and **932**, coupling with both coil **924** and coil **926**. In FIG. **45**C coils **920**, **922** are energized in a permanent magnet magnetic flux aiding manner such that all magnetic flux traverses path **932** and couples with coil **926**. In FIG. **45**D no coils are energized as in FIG. **45**B. By continuously alternatingly energizing and de-energizing coils **916**, **918** and **920**, **922** in such a manner energy conversion is achieved due to the coupling with coils **924** and **926**. The magnetic flux in the integral pole piece portions **912** and **914**, and thus the flux coupling with respective coils **924** and **926**, varies by a factor of twice the amount of magnetic flux generated by energizing coils **916**, **918** and **920**, **922**.

The construction shown in FIGS. **45**A and **45**X are similar to the construction shown in FIGS. **7** and **47**. The difference in both cases relates to replacing the two flux paths and armatures with one continues flux path. The arrangement in FIG. **7** has one permanent magnet and four coils and the arrangement in FIG. **47** has two permanent magnets and two coils. Although the physical aspects of the two arrangements and the details of the flux control vary, the control method for varying the permanent magnets flux are similar and will be described simultaneously and only differences will be pointed out.

With continuous flux paths the static flux from the permanent magnet or magnets is useless. However, if the static flux of the permanent magnet confined to the flux paths were modified to be time varying it would have utility for electromagnetic induction devices for power conversion like transformers and power inverters. However, the same basic method for controlling the flux of a permanent magnet to provide linear and rotary motion can also be applied to time varying the static flux from the permanent magnetic. The construction shown in FIG. **45**X utilizes four control coils and a single permanent magnet and the construction shown in FIG. **45**A uses two control coils and two permanent magnets. The flux that would normally be supplied by a primary winding is supplied by the static flux of the permanent magnet or magnets and the control coils convert this static flux into a time varying flux in a novel way. Both arrangements use two secondary coils, the secondary coils are placed in the region of the continuous flux path that would be occupied by an armature or rotor in the linear or rotary arrangements. The regions of the flux paths that perform work are the same in all cases.

In all cases the control coils can either be wired in series or parallel and the secondary coils can be either wound in series or parallel. More than one secondary coil or secondary coils with multiple taps can be placed in the working regions and further multiple flux paths can be utilized with one or more secondary coils placed in each of the working regions. This is made obvious by the disclosures of the linear and rotary devices herein and based on the fact that the working regions of the flux paths are identical.

FIGS. **45**X and **45**A also show the paths of the static flux of the permanent magnet or magnets when no current is flowing in the control coils. In the arrangement shown in FIG. **45**X the flux from the single permanent magnet divides between the two working areas of the flux path. In the arrangement of FIG. **45**A all of the flux of one of the permanent magnets passes through one of the working regions and all of the flux of the second permanent magnet passes through the other working region. Each of the working regions in both cases are occupied by secondary coils.

FIGS. **45**Y and **45**B show the control coils energized with the polarity shown with respect to the polarity of the permanent magnet or magnets included. In FIG. **45**Y the opposing coil blocks the passage of flux from the permanent magnet and the aiding coil couples with the flux of the permanent magnet and therefore all of the flux of the permanent magnet passes through one working region as shown. In FIG. **45**A the opposing side of the coil blocks the passage of flux from the permanent magnet on the opposing side of the coil and the aiding side of the coil couples with the flux of the other permanent magnet and therefore all of the flux of both the permanent magnets passes through the working region as shown.

FIGS. **45**Z and **45**C show the control coils energized with a polarity opposite of that shown in FIGS. **45**Y and **45**B. The same action occurs and results in all of the permanent magnet or magnets path flux passing through the opposite working regions.

By alternating the polarity of the control coils during one cycle, one working region experiences an increasing flux and the opposite region experiences a decreasing flux and during the next cycle the opposite occurs. This results in the induction of a voltage in the secondary coils that is decided by the magnitude of the change in flux in the working region and the time in which this change occurs. The novelty of this discovery is that the primary flux inducing the voltage in the secondary coils is supplied by the permanent magnet or magnets and is far greater than the flux supplied by the control coils.

Further, in the rotary motion devices of FIGS. **31** and **34**, it is not necessary that respective rotor members **702** and **744** be formed of permanent magnets. Each could take the form shown in FIG. **46** where sections **950** and **952** are formed of magnetic material such as soft iron and sections **954** and **956** are formed by a non-magnetic filler material.

FIGS. **47** and **48** show another embodiment **1000** of the subject device. The embodiment **1000** includes two spaced permanent magnets **1002** and **1004** each of which has its north pole adjacent the upper surface and its south pole adjacent the lower surface. A magnetizable bridging member **1006** extends across and makes contact with the north magnetic poles of the magnets **1002** and **1004** and another magnetizable bridging member **1008** makes contact with the south magnetic poles of the two permanent magnets **1002** and **1004**. The members **1006** and **1008** extend slightly beyond the opposite sides of the respective permanent magnets **1002** and **1004** and a pair of spaced armature members **1010** and **1012** are positioned to move into and out of engagement with the ends of the members **1006** and **1008**. Coils **1014** and **1016** are mounted respectively on the members **1006** and **1008** in the space between the permanent magnets **1002** and **1004**, and the armatures **1010** and **1012** are shown connected together by a rod **1018** which enables them to move backwards and forwards into engagement with the respective members **1006** and **1008** when different voltages are applied to the respective coils **1014** and **1016**. In FIGS. **47**, the coils **1014** and **1016** are energized as shown with the coil **1014** having its north magnetic end to the left and its south magnetic end to the right and the opposite is true of the coil **1016**. In FIG. **48**, the voltage applied to the respective coils **1014** and **1016** is reversed so that the polarity of the left end of coil **1014** is south and the polarity of the opposite end of the same coil **1014** is a north magnetic pole. The reverse is true of the coil **1016**. It is to be noted in FIGS. **47** and **48** that the relationship of aiding and opposing is indicated on the figures to indicate the relationship when the coils are energized. For example, in FIG. **47** when the coils are energized as shown the relationship is opposing for the permanent magnet **1002** and is aiding with respect to the permanent magnet **1004**. The reverse is true when the voltage on the coils is reversed as shown in FIG. **48**. By properly timing the voltage on the respective coils the movement of the armature is controlled. The same principles can be applied to produce rotating movement as shown in FIG. **42**.

FIG. **49** shows another embodiment **1030** of the subject invention using principles similar to those described in connection with FIGS. **47** and **48**. The embodiment **1030** includes a plurality, three being shown, of stationary members **1032**, **1034** and **1036**. The details of these members are better shown in FIG. **50** which shows the details of the member **1036**. This member includes a pair of permanent magnets **1038** and **1040**, each of which has magnetizable members mounted adjacent opposite sides thereof as in the previous construction. The members **1042** and **1044** also have coils **1046** and **1048**, respectively, and the coils are energized as described in connection with FIGS. **47** and **48** to produce aiding and opposing magnetism. The construction shown in FIG. **49** may have three stator portions as shown or it may have more stator portions as desired. The rotor **1050** is positioned in the space between the members **1032**, **1034** and **1036** and includes a permanent magnet portion part of which has its north magnetic pole on the surface as shown and the other parts has its south magnetic pole in the same surface as shown. The permanent magnets **1038** and **1040** on the stators interact with the permanent magnets on the rotor to produce the rotating motion and is controlled by the energizing of the coils.

Other applications and advantages of the devices and methods of the present invention exist and various modifications are possible, and therefore the present invention is not intended to be limited to the specific examples disclosed herein. Accordingly, the spirit and scope of the invention are to be limited only by the terms of the appended claims.

What is claimed is:

1. A permanent magnet device, comprising a permanent magnet having north and south pole faces, a first pole piece, a second pole piece, a first control coil, a second control coil, and circuit means, the first pole piece positioned adjacent the north pole face of the permanent magnet and including a first path portion, a second path portion and a third portion, the first path portion extending beyond a perimeter of the north pole face in one direction and the second path portion extending beyond the perimeter of the north pole face in another direction to define first and second flux paths for magnetic flux emanating from the north pole face of the permanent magnet, the first path portion of the first pole piece connected to the second path portion of the first pole piece by the third portion which extends across the north pole face of the permanent magnet, the second pole piece positioned adjacent the south pole face and including a first path portion and a second path portion, the first path portion extending beyond a perimeter of the south pole face and substantially aligned with the first path portion of the first pole piece, the second path portion extending beyond the perimeter of the south pole face and substantially aligned with the second path portion of the first pole piece, the first control coil positioned around the first path portion of the first pole piece, the second control coil positioned around the second path portion of the first pole piece, the circuit means connected to each of the first control coil and the second control coil to alternatingly energize the first coil and the second coil in a timed sequential manner.

2. The permanent magnet device as set forth in claim **1**, wherein the first control coil and the second control coil are alternatingly energized in a permanent magnet magnetic flux aiding manner.

3. The permanent magnet device as set forth in claim **1**, wherein the first control coil and the second control coil are alternatingly energized in a permanent magnet magnetic flux opposing manner.

4. The permanent magnet device as set forth in claim **1**, further comprising a rotor member mounted on a shaft for rotation therewith, the rotor member sized, shaped, and positioned to extend substantially from the first path portion of the first pole piece to the first path portion of the second pole piece during at least some part of its rotation.

5. The permanent magnet device as set forth in claim **4**, wherein the rotor member is formed by at least one permanent magnet.

6. The permanent magnet device as set forth in claim **1**, wherein the second path portion of the first pole piece and the second path portion of the second pole piece are positioned alongside the first path portion of the first pole piece and the first path portion of the first pole piece.

7. The permanent magnet device as set forth in claim **1**, further comprising a first bypass extending from the first path portion of the first pole piece to the first path portion of the second pole piece, one end of the first bypass positioned adjacent the first path portion of the first pole piece and between the permanent magnet and the first control coil.

8. The permanent magnet device as set forth in claim **6**, further comprising a second bypass extending from the second path portion of the first pole piece to the second path portion of the second pole piece, one end of the second bypass positioned adjacent the second path portion of the first pole piece and between the permanent magnet and the second control coil.

9. The permanent magnet device as set forth in claim **1**, further comprising a plurality of armatures arranged to define a path of movement, wherein the second path portion of the first pole piece and the second path portion of the second pole piece are positioned alongside the first path portion of the first pole piece and the first path portion of the second pole piece, and wherein all of such pole piece path portions include an end face positioned adjacent the path of movement defined by the plurality of armatures.

10. The permanent magnet device as set forth in claim **1**, wherein the first control coil and the second control coil are simultaneously energized one in a permanent magnet magnetic flux aiding manner and one in a permanent magnet magnetic flux opposing manner.

11. The permanent magnet device as set forth in claim **1**, further comprising two shaft connected armatures positionable adjacent the ends of the first and second pole pieces, wherein each of the armatures is formed by a permanent magnet.

12. The permanent magnet device of claim **1** further comprising a first fixed armature extending between the first path portion of the first pole piece to the first path portion of the second pole piece and a second fixed armature extending between the second path portion on the first pole piece to the second path portion of the second pole piece.

13. The permanent magnet device of claim **12** where a first secondary coil is wrapped around the first fixed armature and a second secondary coil is wrapped around the second fixed armature.

14. The permanent magnet device of claim **13** including circuit means connected to the control coils to control the energizing thereof to produce a varying flux in the armatures and to induce voltage in the secondary coils.

15. The permanent magnet device of claim **1** wherein there are at least two permanent magnets each having north and south pole faces, the first pole piece being positioned extending between the north pole faces of the permanent magnets and the second pole piece positioned extending between adjacent south pole faces of the permanent magnets.

16. A method for controlling the path of magnetic flux from a permanent magnet, the method comprising the steps of:

(a) placing a first pole piece adjacent a first pole face of the permanent magnet so as to have at least first and second path portions extending beyond a perimeter of the first pole face;

(b) placing a second pole piece adjacent a second pole face of the permanent magnet so as to include at least one portion which substantially aligns with the first and second path portions of the first pole piece;

(c) placing a first control coil along and around the first path portion of the first pole piece;

(d) placing a second control coil along and around the second path portion of the first pole piece;

(e) repeatedly energizing the first control coil in a permanent magnet magnetic flux opposing manner so as to prevent magnetic flux of the permanent magnet from traversing the first path portion of the first pole piece; and

(f) repeatedly energizing the second control coil in a permanent magnet magnetic flux opposing manner so as to prevent magnetic flux of the permanent magnet from traversing the second path portion of the first pole piece.

17. The method as set forth in claim **16** wherein the energization of steps (e) and (t) take place in a simultaneous manner.

18. A method for controlling the path of magnetic flux from a permanent magnet, the method comprising the steps of:

(a) placing a first pole piece adjacent a first pole face of the permanent magnet so as to have at least first and second path portions extending beyond a perimeter of the first pole face;

(b) placing a second pole piece adjacent a second pole face of the permanent magnet so as to include at least one portion which substantially aligns with the first and second path portions of the first pole piece;

(c) placing a first control coil along and around the first path portion of the first pole piece;

(d) placing a second control coil along and around the second path portion of the first pole piece; and

(e) alternatingly performing the following steps in a repeated manner:

(i) energizing the first control coil in a permanent magnet magnetic flux aiding manner so as to couple with substantially all magnetic flux of the permanent magnet such that substantially no magnetic flux of the permanent magnet traverses the second path portion of the first pole piece when the first control coil is so energized; and

(ii) energizing the second control coil in a permanent magnet magnetic flux opposing manner so as to couple with substantially all magnetic flux of the permanent magnet such that substantially no magnetic flux of the permanent magnet traverses the first path portion of the first pole piece when the second control coil is so energized.

19. A method for controlling the path of magnetic flux from a permanent magnet the method comprising the steps of:

(a) placing a first pole piece adjacent a first pole face of the permanent magnet so as to have at least first and second path portions extending beyond a perimeter of the first pole face;

(b) placing a second pole piece adjacent a second pole face of the permanent magnet so as to include at least one portion which substantially aligns with the first and second path portions of the first pole piece;

(c) placing a first control coil along and around the first path portion of the first pole piece;

(d) placing a second control coil along and around the second path portion of the first pole piece; and

(e) alternatingly performing the following steps in a repeated manner:

(i) simultaneously energizing the first control coil in a permanent magnet magnetic flux aiding manner and the second control coil in a permanent magnet flux opposing manner; and

(ii) simultaneously energizing the first control coil in a permanent magnet flux opposing manner and the second control coil in a permanent magnet magnetic flux aiding manner.

20. A rotary motion device, comprising a rotor assembly including a shaft which defines an axis of rotation of the assembly, a rotor pole piece mounted for rotation with the shaft, the rotor pole piece including an outer ring portion having at least two path portions extending inwardly from a periphery of the outer ring portion;

a stator assembly including a permanent magnet having a generally ring-shaped configuration, a first pole face of the permanent magnet positioned adjacent the outer ring portion of the rotor pole piece, the stator assembly further comprising a stator pole piece including an outer ring portion positioned adjacent a second pole face of the permanent magnet and having a plurality of path portions extending inwardly from the periphery, each path portion further including a respective portion which extends toward a plane defined by the first pole face of the permanent magnet and alignable with each of the rotor pole piece path portions at certain rotational positions of the rotor pole piece, each path portion including a control coil positioned therealong;

and circuit means connected to each of the coils and including a source of electrical energy and switch means for energizing respective ones of the control coils in a predetermined timed sequence based upon rotation of the rotor assembly.

21. A rotary motion device, comprising:

a rotor assembly including a shaft which defines an axis of rotation of the assembly, a pair of spaced elongated rotor members mounted on the shaft at spaced locations thereon and angularly oriented with respect to each other, each of the elongated rotor members formed of a magnetic material;

a stator assembly including a permanent magnet having opposed first and second pole faces, a first pole piece positioned adjacent the first pole face and a second pole piece positioned adjacent the second pole face, each pole piece including a respective first path portion extending beyond a perimeter of its adjacent pole face and having an arcuate shaped end portion, the first path portion of the first pole piece aligned with the first path portion of the second pole piece, each pole piece further including a respective second path portion extending beyond the perimeter of its adjacent pole face in a direction opposite to that of the first path portions and having an arcuate shaped end portion, the second path portion of the first pole piece aligned with the second path portion of the second pole piece, at least one of the first path portions of the first pole piece and the first path portion of the second pole piece including a control coil mounted on at least one of the pole pieces, at least one of the second path portions of the first pole piece and the second path portion of the second pole piece including a control coil mounted on at least one of the pole pieces,

wherein the rotor assembly extends from end to end of the stator assembly such that the elongate members are aligned with the arcuate shaped end portions of the path portions of the pole pieces;

and circuit means connected to each of the coils and including a source of electrical energy and switch means for energizing respective ones of the control coils in a predetermined timed sequence based upon rotation of the rotor assembly.

22. A rotary motion device comprising:

a rotor assembly including a shaft which defines an axis of rotation of the assembly, a ring-shaped rotor member mounted for rotation with the shaft, the ring-shaped rotor member including a plurality of distinct circumferential regions;

a stator assembly including a first permanent magnet, a first pole piece positioned against a first pole face and a second pole piece positioned against a second pole face, the first pole piece including at least a first path portion extending beyond a perimeter of the first pole face, the second pole piece including at least a first path portion extending beyond a perimeter of the second pole face, the first path portion of the first pole piece aligned with the first path portion of the second pole piece, at least a portion of the ring-shaped rotor member positioned between the first path portion of the first pole piece and the first path portion of the second pole piece, at least one of the first path portions of the first pole piece and the first path portion of the second pole piece including a first control coil positioned at a point intermediate the first permanent magnet and the ring-shaped rotor member;

and circuit means connected to the first control coil and including a source of electrical energy and switch means for energizing the first control coil in a predetermined timed manner based upon rotation of the rotor assembly.

23. The rotary motion device as set forth in claim **22**, wherein the ring-shaped rotor member is formed by a permanent magnet having distinct circumferential regions of opposite polarity.

24. The rotary motion device as set forth in claim **23**, wherein the first pole piece includes a second path portion spaced from and extending adjacent to the first path portion, the second pole piece including a second path portion spaced from and extending adjacent to the first path portion such that the second path portion of the first pole piece is aligned with the second path portion of the second pole piece, at least a portion of the ring-shaped permanent magnet rotor member positioned between the second path portion of the first pole piece and the second path portion of the second pole piece, at least one of the second path portions of the first pole piece and the second path portion of the second pole piece having a second control coil mounted on at least one of the pole pieces at a point intermediate the first permanent magnet and the ring-shaped permanent magnet rotor member, the second control coil connected to the circuit means so as to be energized in a predetermined timed manner based upon rotation of the rotor assembly.

25. The rotary motion device as set forth in claim **22**, wherein the stator assembly further comprises a second permanent magnet, a third pole piece positioned adjacent a first pole face of the second permanent magnet and a fourth pole piece positioned adjacent a second pole face of the second permanent magnet, the third pole piece including at least a first path portion extending beyond a perimeter of the second permanent magnet first pole face, the fourth pole piece including at least a first path portion extending beyond a perimeter of the second permanent magnet second pole face, the first path portion of the third pole face aligned with the first path portion of the fourth pole piece, at least a portion of the ring-shaped permanent magnet rotor member positioned between the first path portion of the third pole piece and the first path portion of the fourth pole piece, at least one of the first path portions of the third pole piece and the first path portion of the fourth pole piece including a third control coil mounted on at least one of the pole pieces at a point intermediate the second permanent magnet and the ring-shaped permanent magnet rotor member, the third pole piece including a second path portion spaced from and extending adjacent to the first path portion the fourth pole piece including a second path portion spaced from and extending adjacent to the first path portion thereof such that the second path portion of the third pole piece is aligned with the second path portion of the fourth pole piece, at least a portion of the ring-shaped permanent magnet rotor member positioned between the second path portion of the third pole piece and the second path portion of the fourth pole piece, at least one of the second path portions of the third pole piece and the second path portion of the fourth pole piece including a fourth control coil mounted on at least one of the pole pieces at a point intermediate the second permanent magnet and the ring-shaped permanent magnet rotor member, wherein each of the third and fourth control coils are connected to the circuit means so as to be energized in a predetermined timed manner based upon rotation of the rotor assembly.

26. A device for producing rotary motion comprising:

a rotor assembly including a shaft which defines an axis of rotation for the assembly, a ring-shaped rotor member mounted for rotation with the shaft, the ring-shaped rotor member having a plurality of distinct circumferentially positioned regions extending around the axis, a stator assembly including a first permanent magnet, a first pole piece positioned against the first pole face of the first pole piece and a second pole piece positioned against a second pole face of the first pole piece, the first pole piece including at least a first path portion extending beyond a perimeter of the first pole face, the second pole piece including at least a first path portion extending beyond the perimeter of the second pole face, the first path portion of the first pole piece aligned with the first path portion of the second pole piece, at least a portion of the ring-shaped rotor member positioned between the first path portion of the first pole piece and the first path portion of the second pole piece, at least one of the first path portions of the first pole piece and the first path portion of the second pole piece including a first control coil mounted on at least one of the pole pieces at a point intermediate the first permanent magnet and the ring-shaped rotor member; and circuit means connected to the first control coil and including a source of electrical energy and switch means for energizing the first control coil in a predetermined timed manner based upon position of the rotor assembly during rotation of the rotor assembly.

27. The device for producing rotary motion of claim **26** wherein the circuit means includes means for timing the energizing of the first control coil includes means for adjusting the timing thereof.

28. The device for producing rotor motion of claim **26** including means to vary the flux generated in the first and second pole pieces.

29. A device for handling the flux between two separate permanent magnets each of which has a north magnetic pole adjacent one side face and the south magnetic pole adjacent to the opposite side face, the north and south side pole faces respectively of both magnets being substantially in alignment, a first member in surface-to-surface contact with the north magnetic faces of the spaced permanent magnets, a second member in surface-to-surface contact with the south magnetic faces of the spaced permanent magnets, first and second armatures each positioned adjacent opposite ends of the first and second permanent magnets and adjacent to opposite ends of the spaced members, a coil mounted on each of the members in the space between the adjacent permanent magnets, and means for applying voltages of predetermined polarities across the respective coils to change the magnetic coupling between the permanent magnets and between the armatures.

30. A device for producing rotational movement comprising:

a rotor having a shaft rotatable about the axis thereof, a member constructed of permanent magnets mounted on the shaft, said member having circumferential portions some of which have a north magnetic pole and others a south magnetic pole adjacent to the same side thereof, the opposite surface of the permanent magnet member having north magnetic poles opposite the south magnetic poles and south magnetic poles opposite the north magnetic poles, a stator having a plurality of circumferentially spaced portions each of which includes at least one permanent magnet and a pair of members mounted adjacent opposite sides of the permanent magnets, the members being positioned adjacent to the periphery of the rotor permanent magnet member and means on the member adjacent each opposite side of the stator permanent magnet for mounting a coil, and means for energizing the coil on each stator portion in sequence to produce magnetic coupling force between the stator and the rotor in a direction to produce rotating motion of the rotor.

31. A device including a rotating member and a stationary member, each having a permanent magnet portion positioned to produce magnetic coupling force therebetween in predetermined positions thereof, the rotor including a shaft rotatable about its axis and the permanent magnet extending around the shaft and formed by a plurality of adjacent portions of permanent magnet material whereby adjacent portions have their north and south magnetic pole faces on opposite sides of the rotor permanent magnet, a plurality of stator members each stator member having at least one permanent magnet having a north magnetic pole adjacent one side and a south magnetic pole adjacent to the opposite side, a pair of members positioned adjacent respective opposite sides of the stator permanent magnet in position to extend to adjacent the rotor permanent magnet whereby a flux path is formed between the members and the stator and rotor permanent magnets, a coil mounted on each member of the stator and means for applying a voltage of predetermined polarity to each of said coils to control the flux through a path between the permanent magnets and to control the coupling force between the permanent magnets on the stator and the permanent magnets on the rotor.

32. A motion producing device comprising at least one permanent magnet having a north pole opposite and spaced from a south pole, a pair of spaced substantially parallel members adjacent respectively the north and south poles of the at least one permanent magnet and extending outwardly to substantially aligned opposite edges, a flux supporting member positioned adjacent the respective opposite edges of each pair of parallel members, a coil on selected ones of the parallel members, and a source of electrical energy connected to each of the coils for energizing the coils to change the flux in the parallel members and in the flux supporting members.

33. The motion producing device of claim **32** wherein there are at least two spaced permanent magnets extending between the parallel members.

34. The motion producing device of claim **32** wherein one of said pair of parallel members is subdivided into a plurality of sidewardly extending portions extending to one of said opposite side edges, at least one of said coils being positioned on at least one of said sidewardly extending portions.

35. The motion producing device of claim **34** wherein there are coils on a plurality of respective ones of the sidewardly extending portions.

36. The motion producing device of claim **32** wherein the permanent magnet and the parallel members are annular in shape.

37. The motion producing device of claim **32** including a by-pass member extending between the pair of spaced substantial parallel members adjacent one side of the permanent magnet.

38. A permanent magnet device comprising at least two permanent magnets each having north and south pole faces, a first pole piece, a second pole piece, a first control coil, a second control coil and circuit means, the first pole piece positioned adjacent the north pole faces of the at least two permanent magnets and including a first path portion, a second path portion and a third path portion, the first path portion extending beyond the perimeter of the north pole faces and the second path portion extending beyond the perimeter of the north pole faces to define first and second flux paths for magnetic flux emitting from the north pole faces of the at least two permanent magnets, the first path portion of the first pole piece connected to the second path portion of the first pole piece by a third portion which extends across the north pole face of the at least two permanent magnets, the second pole piece positioned adjacent to the south pole faces of the at least two permanent magnets and including a first path portion and a second path portion, the first path portion extending beyond a perimeter of the south pole faces and substantially aligned with the first path portion of the first pole piece, the second path portion extending beyond the perimeter of the south pole faces and substantially aligned with the second path portion of the first pole piece, the first control coil positioned around the first path portion of the first pole piece, the second control coil positioned around the second path portion of the first pole piece, and the circuit means connected to each of the first control coil and the second control coil to alternatingly energize the first coil and the second coil in a timed sequential manner.

39. The permanent magnet device of claim **38** further comprising a first fixed armature extending between the first path portion of the first pole piece to the first path portion of the second pole piece and a second fixed armature extending between the second path portion of the first pole piece to the second path portion of the second pole piece.

40. The permanent magnet device of claim **39** where a first secondary coil is wrapped around the first fixed armature and a second secondary coil is wrapped around the second fixed armature.

41. The permanent magnet device of claim **40** including circuit means connected to the control coils to control the energizing thereof to produce a varying flux in the armatures and to induce voltage in the secondary coils.

42. The permanent magnet device of claim **38** wherein there are at least two permanent magnets each having north and south pole faces, the first pole piece being positioned extending between the north pole faces of the permanent magnets and the second pole piece positioned extending between the south pole faces of the permanent magnets.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,246,561 B1
DATED : June 12, 2001
INVENTOR(S) : Charles J. Flynn

Page 1 of 1

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 3,
Line 53, "Figs 10, 10A-H" should be -- Figs. 10A-H --.
Line 56, "Figs 11, 11A-F" should be -- Figs. 11A-F --.
Line 61, "Figs 12, 12A-E" should be -- Figs. 12A-E --.

Column 4,
Line 4, "Figs. 16A-E" should be -- Fig. 16 --.

Signed and Sealed this

Twenty-ninth Day of January, 2002

Attest:

JAMES E. ROGAN
Director of the United States Patent and Trademark Office

Attesting Officer

28th June 2013

AM/FM/Loop Crystal Radio

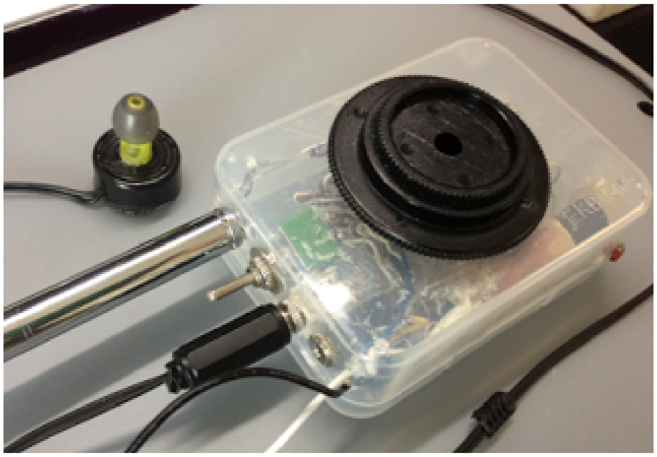

[http://3.bp.blogspot.com/-W5MoCILOqOU/Uc2WU1CTbbI/AAAAAAAABNA/hH64_FNxMpo/s436/Screen+Shot+2013-06-23+at+12.06.23+PM.png]

Build a multifunction battery free crystal radio to listen to either AM or FM radio.
You can also use the same radio to connect to a bigger loop coil to get stronger signal for the AM radio instead of relying on the smaller AM coil inside the radio.

Parts you will need:
A Diode that works in high frequency (200Mhz or above, e.g. AA112, 1N34A, 1N48, Iss86 or Iss106.
A Diode that work in AM frequency e.g 1N34A, 1N60, BAT85.
A quadruple PVC Variable capacitor with one end for 2x 20pf, the other end for 2x 140pf.
50pf and 180pf fixed ceramic capacitor
100K 1/4W resistor
33uH Choke coil.
AM radio coil - ferrite rod
3.5mm mono headphone jack
3.5mm stereo headphone jack with internal switches (for selecting whether to use the external big loop or the internal coil for AM.
A 2 by 3 switch to select AM or FM.
A high impedance 4K Ohm earphone or a crystal earphone.

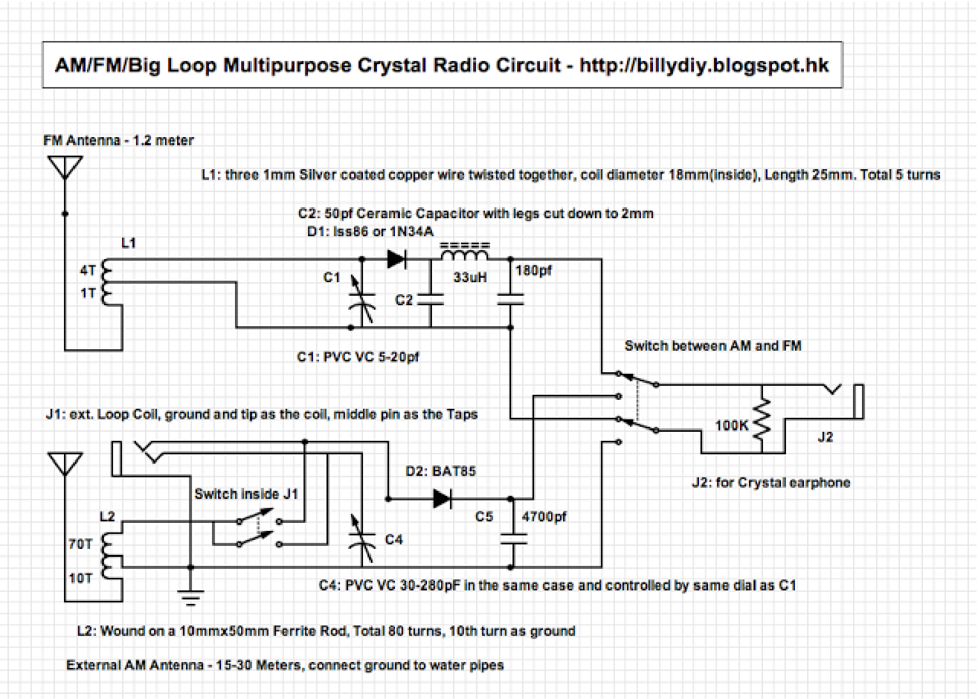
AM/FM/Big Loop Multipurpose Crystal Radio Circuit - http://billydiy.blogspot.hk
FM Antenna - 1.2 meter
L1: three 1mm Silver coated copper wire twisted together, coil diameter 18mm(inside), Length 25mm. Total 5 turns
C2: 50pf Ceramic Capacitor with legs cut down to 2mm
D1: Iss86 or 1N34A
L1
4T
1T
C1
C2
33uH
180pf
C1: PVC VC 5-20pf
Switch between AM and FM
J1: ext. Loop Coil, ground and tip as the coil, middle pin as the Taps
100K
J2
D2: BAT85
J2: for Crystal earphone
Switch inside J1
L2
70T
10T
C4
C5
4700pf
C4: PVC VC 30-280pF in the same case and controlled by same dial as C1
L2: Wound on a 10mmx50mm Ferrite Rod, Total 80 turns, 10th turn as ground
External AM Antenna - 15-30 Meters, connect ground to water pipes

[http://2.bp.blogspot.com/--ZAhsy2SRko/Uc2WUu2Gz6I/AAAAAAAABM8/czIHQF1iPzI/s855/Screen+Shot+2013-06-23+at+11.22.21+AM.png]

To connect to external Loop coil for AM reception instead of using the internal AM coil, connect 3-color crocodile camp to the 3.5mm Stereo jack - white for the tuning, black for ground, yellow for the decoder diode.

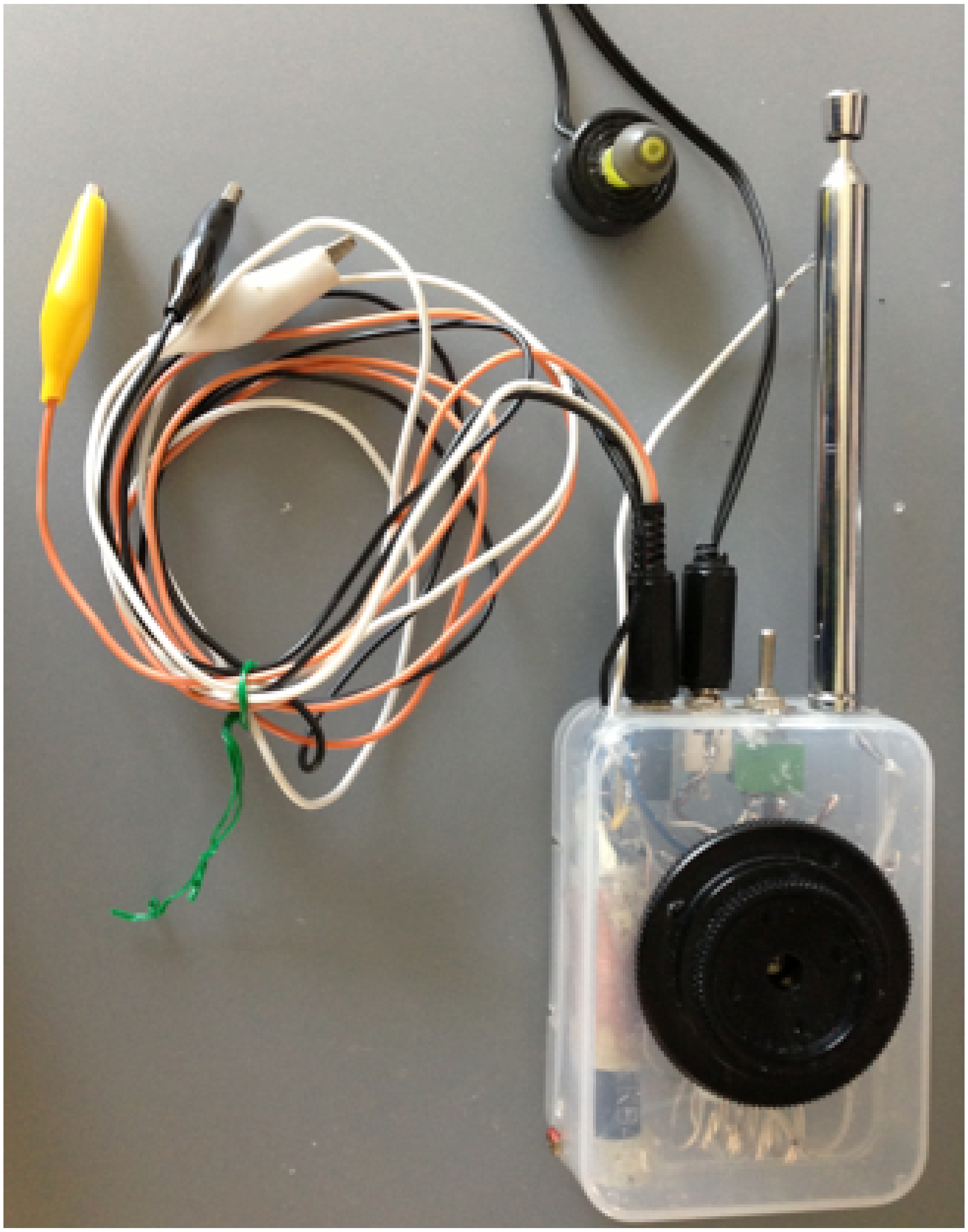

[http://3.bp.blogspot.com/-df0Gv1TiXME/Uc2WTw3VL8I/AAAAAAAABMs/hJaPbNYh4uQ/s415/Screen+Shot+2013-06-23+at+1.19.26+PM.png]

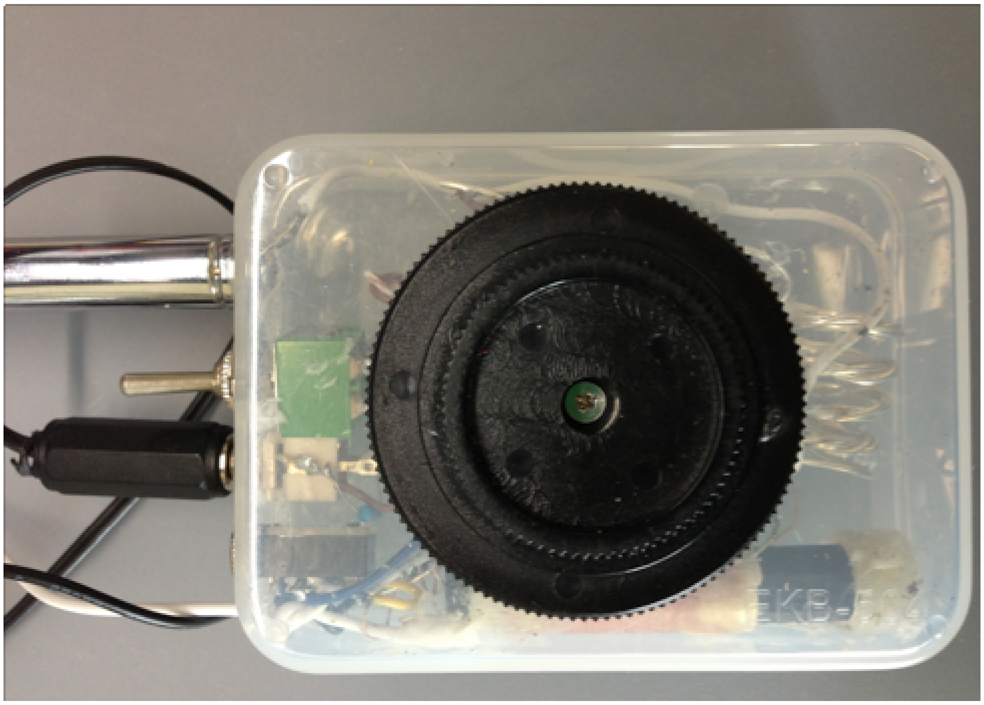

[http://2.bp.blogspot.com/-966mXwGZozE/Uc2WVmB2lLI/AAAAAAAABNY/P2KdoMFVIR0/s489/Screen+Shot+2013-06-23+at+12.07.13+PM.png]

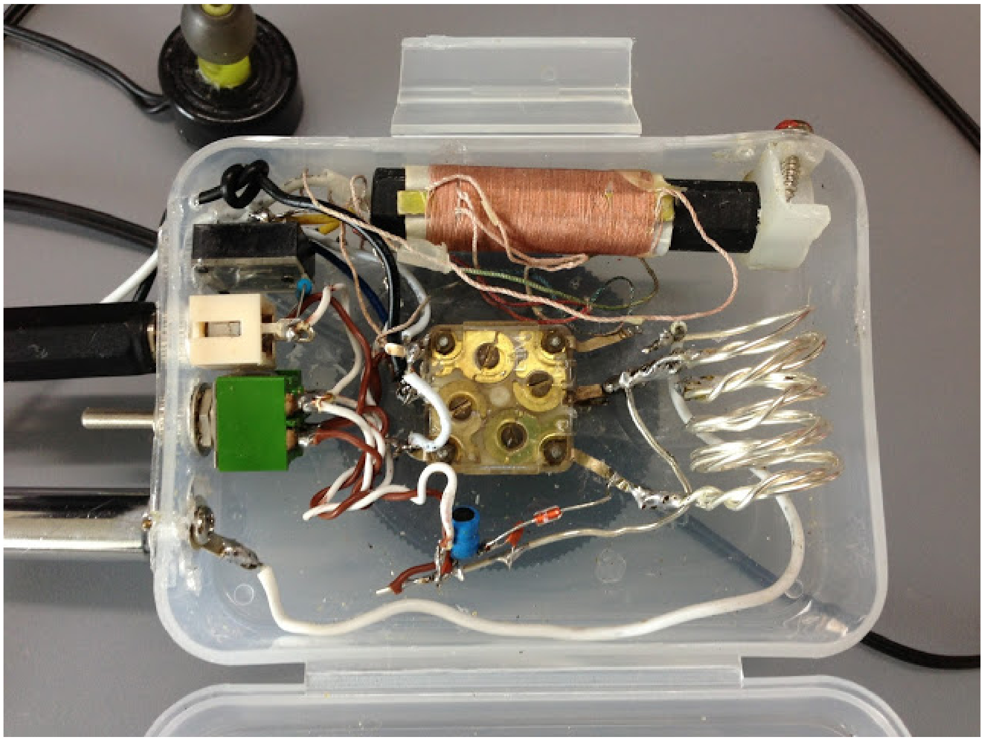

[http://1.bp.blogspot.com/-UURJiV8VfBY/Uc2WTLqcLFI/AAAAAAAABMc/q9_Y7C47Zb8/s1600/IMG_2521.JPG]

Posted 28th June 2013 by Billy

Precompliance Antenna
HighEnd Log-Periodic Wideband EMC Antennas from 1Hz to 40GHz
Go to aaronia.com/antennas

0 Add a comment

Enter your comment...

Comment as: Unknown (Goo ▼ Sign out

Publish **Preview** ☐ Notify me

Amplifier for crystal radio earphone

Related links:
[Crystal radios], [Build Van de Graaff]

Home
Renewable Energy
Science/Electronics
 Projects
  Build Van de Graaff
  Wimshurst machine
  Kelvin water dropper
  Electroscope
  Corona motors
  Ball bearing motor
  Crystal radios
   Tips and tricks
   Crystal earphones
   Using earbuds
   Razor blade diode
   Amplifier for speaker
   Amplifier for earphone
   Loop antenna radio
  AM Radio Transmitter
  Fresnel lens
  Electrets
  Piezoelectric
  Electrolytic capacitor
  Joule thief
  Peltier/Seebeck effect
  Franklins bell
  Can Stirling
  Big Stirling
  555 timer music
  TEA laser
  Laser diode fun
  How Ion Prop. Works
  Lifter/ionocraft
  Star Trek Enterprise
  Smoke precipitator
  Moving with sound
  Straw music instruments
  Gravity light - DIY
  Nitinol wire
  Pyramid hologram
  Arduino speech/Talkie
  Arduino controlled Skull
  Pinhole camera
  Wireless electricity
  Fly swatter powered CFL
  Chlorophyll/fluorescence
  Small SG Tesla coil
  Photophone
  Laser communicator
  Hero's steam engine
  Alcohol stove
  Sunflower heat engine
  BB-8 Droid
  Ball cyclotron
  Atmos. electricity
  Photoelectric effect
  Bottle rocket
  Neural networks
  Ornithopter
  Curie temp. experiment
  Lightning in bottle
  Gyroscope - vinyl records
  Raspberry Pi
  Pepper's ghost phone
  Potato chip circuit
  Ion wind rotor
  LED flashlight
 Resources
 Power supplies
 HV Probes/measuring
 Miscellaneous
Efficiency & Conservation
Composting
Food
Non-conv. Energy
Non-conv. Propulsion
Space
Stories

Donate

1 **Handheld Analyzer - Made in Germany - up to 20GHz** Portable RF & EMF Spectrum Analyzer. From 1Hz to 20GHz. Weighs less than 500g. aaronia.com/spect_analyzer

2 **Google AdWords** Reach Potential Customers In The Moments That Matter. Start Today. adwords.google.com

3 **PCB Fab-Assembly-Turnkey** U.S. manufacturer with most advanced technology in the Industry. arrowpcb.com

This is a crystal radio amplifier for amplifying the output to a crystal earphone or telephone receiver earphone. It uses a single transistor and is very simple to make. This circuit is by Charles Wenzel and originated from his website here.

Amplifier circuit. The crystal radio used for example purposes is the crystal radio made from scraps here.

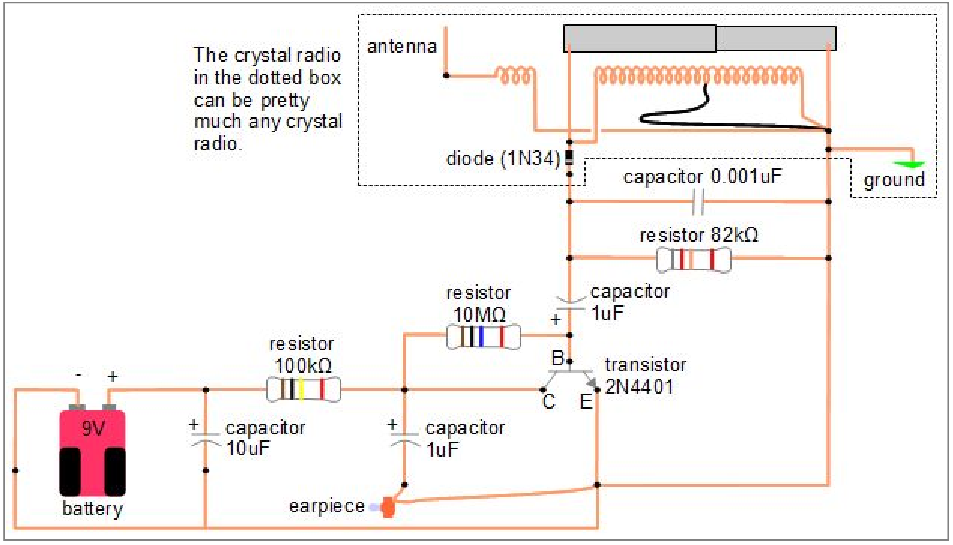

The main components mounted on a breadboard.

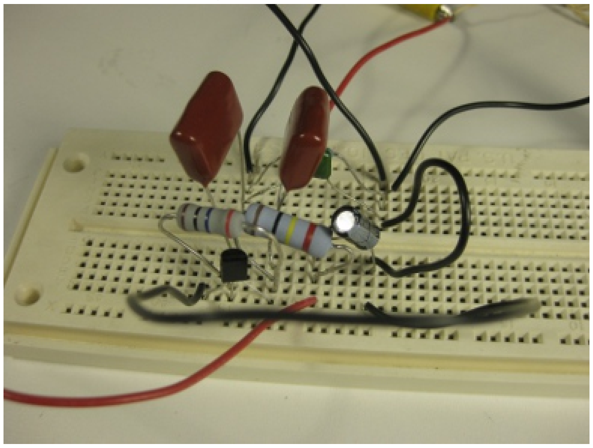

The complete amplifier.

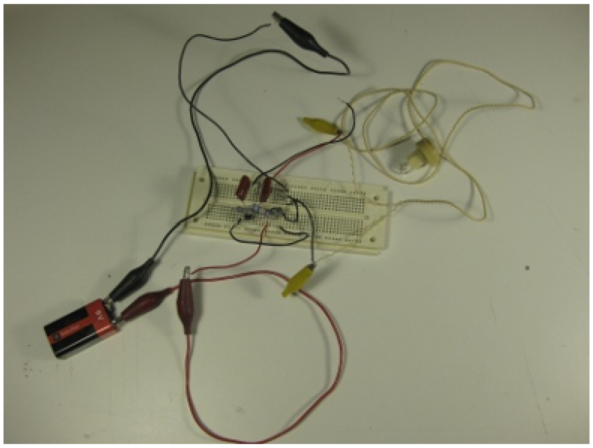

The amplifier attached to a crystal radio.

AdChoices
Crystal Radio
Building Kits
Handheld Analyzer - Made in Germany - up to 20GHz
Portable RF & EMF Spectrum Analyzer. From 1Hz to 20GHz. Weighs less than 500g.
aaronia.com

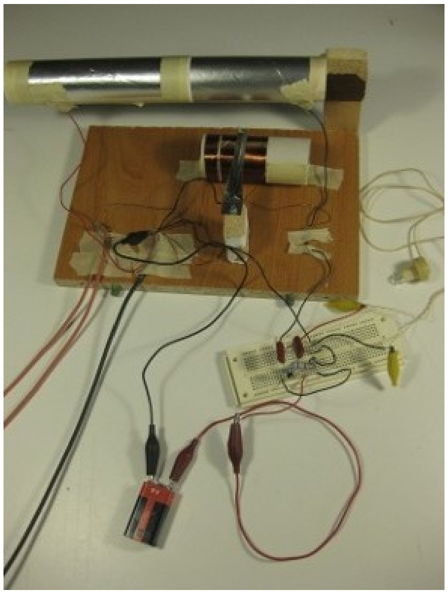

If you did want to try making the above circuit on a breadboard as in the above photo, then the following is a diagram showing how I arranged all the parts. The videos below use this same arrangement.

Diagram for the ampifier circuit on the above breadboard.

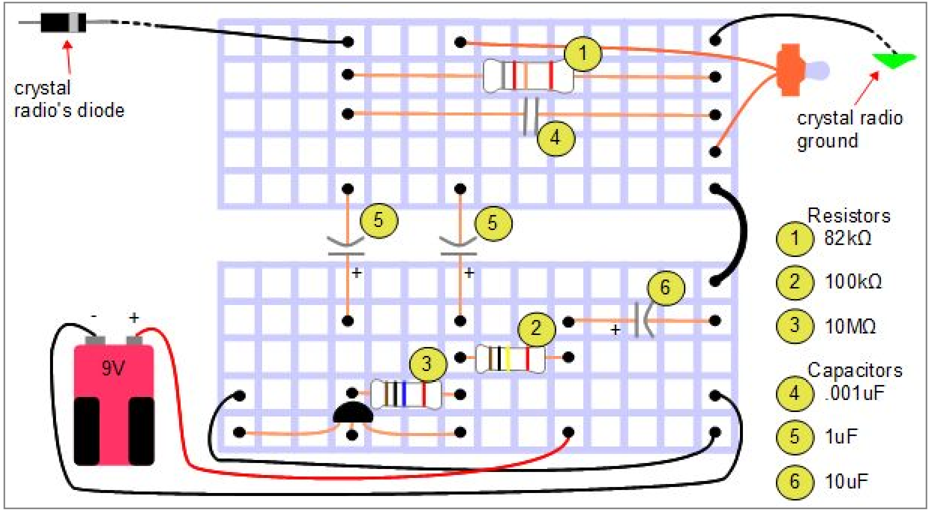
crystal radio's diode
crystal radio ground
1
4
5
5
6
2
3
-
+
9V
+
+
+
Resistors
1 82kΩ
2 100kΩ
3 10MΩ
Capacitors
4 .001uF
5 1uF
6 10uF

Videos with details of the amplifier construction

The following are two construction videos of the making of this amplifier.

This first one includes:

- discussion about the parts,
- how to use a breadboard,
- circuit diagram and how to use it to fill in the breadboard,
- a little about electrolytic capacitors,
- a little about transistors,
- demonstration of it working, and
- a little about how the circuit works.

rexresearch.com

George SQUIER

Tree Antennas

2018 Admissions Now Open
Inquire About The Enrollment Requirements & Process At Saint Isidore's Today. stisidore.org

Tesla Generator - Build Your Own Generator
Eliminate Your Power Bill Easy. Do It Yourself. Great Discount Now! Just $49. tesla-energy.org/Generator

Tesla Generator - Build Your Own Generator
Eliminate Your Power Bill Easy. Do It Yourself. Great Discount Now! Just $49. tesla-energy.org/Generator

General George O. Squier

***Scientific American* (July 14, 1919, p. 624) ~**

"With Trees For Ears"

A Wireless Station Within the Reach of Everybody

With a pair of receives to his ears, an amazed visitor to a certain radio station heard a high-toned hum which changed to a low growl, then skied to the upper reaches of the musical scale in a

faint, very faint buzz, as if some microscopic mosquito had had his song made audible. The operator rapidly rapidly turning the knobs on his couplers and condensers, raised his hand: suddenly, through the changing radio signals which were clamoring for attention together in the receivers came his voice; "There --- the loud, easily heard one is New Brunswick; the fainter, lower one is Nauen, in Germany".

If all this had taken place in the great Arlington station one would not have wondered, save perhaps at the inability to tune out all radio but Nauen. But it was a little portable house erected in thick woods near the edge of the District of Columbia and the signals were received through an oak tree for an antenna.

It is not a joke nor a scientific curiosity, this strange discovery of Gen. George O. Squire, Chief Signal Officer, that trees --- all trees, of all kinds and all heights, growing anywhere --- are nature's own wireless towers and antenna combined. The matter first came to his attention in 1904, through the use of trees as grounds for Army buzzer and telegraph and telephone sets, which, in perfectly dry ground and in a dry season, functioned poorly or not at all with ordinary grounds. Right then he began experiments with a view to seeing what possibilities, if any, the tree had as an aerial. But in 1904 radiotelegraphy was far more undeveloped than at present, and vacuum amplifying tubes were not thought of.

During the war the Signal Corps established a chain of special receiving stations in different localities to copy and record enemy and allied radio messages. Some of those stations were instructed to test the efficiency of growing trees as receiving antennae.

With the remarkably sensitive amplifiers now available, it was not only possible to receive signals from all the principle European stations through a tree, but it has developed beyond a theory and to a fact that a tree is as good as any man-made aerial, regardless of the size or extent of the latter, and better in the respect that it brings to the operator's ears far less static interference.

This is a rather broad statement, yet there beyond the Capital of the nation stands a little portable house, the oak tree, a small receiving set and a couple of enlisted men and an officer on duty; and the curious may, with permission, hear for themselves that the signals so received are neither faint not interrupted, but strong, full-toned dots and sashes even when they come from far-off Nauen. Page after page is copied daily from the propaganda material which Nauen sends out by the ream. Lyons, Poldhu, ships at sea, even the NC-4 on her way, are heard plainly. As for New Brunswick or nearby Arlington --- they deafen the listener if he is unwise enough to try to "take" them otherwise than with the phones lying on the tables.

It will puzzle the amateur as it has puzzled the experts, how a tree, which is certainly well grounded, can also be an insulated aerial. The method of getting the disturbances in potential from treetop to instrument is so simple as to be almost laughable. One climbs a tree to two-thirds of its height, drives a nail a couple of inches into the tree, hangs a wire therefrom, and attaches the wire to the receiving apparatus as if it were a regular lead-in from a lofty copper or aluminum aerial. Apparently some of the etheric disturbances passing from treetop to ground through the tree are diverted through the wire --- and the thermionic tube most efficiently does the rest.

It is interesting to learn that the tree behaves very much like any other aerial; it receives better in dry clear weather than in muggy, damp weather. It plucks messages from the ether more clearly at night than in the day. It is affected very little by rain. It is affected not at all by the presence of other trees; so far as has yet been ascertained it makes little difference whether one drives his nail in a tree in the forest or a lone tree on the plain. Certainly it makes no difference that amounts to anything whether the tree be just an ordinary tree or a giant; it was a 60-foot oak over which the very awe-struck correspondent heard Nauen telling a waiting world what good people the Germans really are. And to prove that it made no particular difference what kind of tree was used the officer in charge switched to a pine tree, which received equally well.

A dead tree will not do, and a tree not in leaf is not so sensitive as one in full foliage. It makes much difference where the nail is driven. General Squier calls the proper place the optimum point, and experimentally it has been determined that two-thirds of the distance from ground to top is the best place -- in a 60-foot tree, 40 feet from the ground.

One nail is sufficient, and it may be any kind of nail; but copper is preferred as not rusting. In practice, if a tree station is to be at all permanent, several nails would be driven and connected to the same wire, each additional nail up to 6 or 8 making the diverted current a little stronger. But 40 nails apparently produce no clearer signals than half a dozen.

The tree may serve as a receiving station for several sets, either connected in series with the same material or from separate terminals.

Some skeptics have expressed the belief that it was not the tree, but the wire leading to the nail in the tree which was the real aerial. The absurdity of thinking a 40-foot wire could receive the widely differing wave lengths which come through the tree station is obvious, but to set any doubt at rest, the wire to the tree has been hung to the nail by means of an insulator, when the signals immediately cease, only to come in as strong as ever just as soon as the connection is again established.

Just what will the tree do as a transmitting station for radio telegraphic messages has not been determined in the Signal Corps Experimental Laboratory. As those in charge express it, "The fact has been demonstrated, but the matter is still in laboratory stage only. What remains to be done now is to develop the best methods of using the demonstrated fact".

But it has already been shown that the tree can be used in wireless telephony and for short distances it has been shown that two-way telephonic communications is easily established through trees with remarkably low values of transmitting antenna current.

If a tree may be used to send wireless telephonic waves it seems not unreasonable to suppose that it will do so as easily with the telegraphic waves. At present the Signal Corps is at work on apparatus to test the possibilities of the tree as a transmitting station.

Just what this development of the art of radio telegraphy may mean has not yet been worked out. It is the history of most discoveries that their potentialities are hardly dreamed of when they are first made --- for instance the telephone, the electromagnet, the vacuum tube amplifier. But it seems fairly obvious that in war, at least, the tree receiving station opens up great possibilities.

True enough there are few trees which remain intact under shell fire, and doubtless with this possibility in mind the armies of the future (if there be such) will in action consider all trees as dangerous enemy aerial stations. But there will always be trees behind the lines and not all actions will be fought on bare ground. What would it have meant to the "lost battalion" to have had a tree wireless set along by which it could have heard that every effort was being made to find and relieve it, or by which it might have sent back messages supplementing that carried by the pigeon?

The greatest development, however, of the tree as the foundation for a receiving and possibly a sending station will come in peace uses. General Squier has written:

"In view of what has been accomplished in space telegraphy, it is difficult to predict to what extent this means of communication may ultimately be developed. If, as indicated in these experiments, the earth's surface is already generously provided with efficient antennae, which we have but to utilize for such communication, even over short distances, it is a fascinating thought to dwell upon in connection with the future development of the transmission of intelligence.

"Since a transmitting station is a central point for electromagnetc waves sent out in all directions over the surface of the earth, a large class of information, such as meteorological reports, crop reports, and general news items of interest to all, may in time be sent from central points, to be received in many places within the radius of influence of the signal station, and this, too, by the simplest form of apparatus."

The amateur wireless world will unquestionably take an intense interest in the tree radio work. At present, while the government has lifted the ban upon amateur aerials, it has not removed the structures against sending.

The aerial is always the greatest problem for the amateur. Lack of both money and material prevents him from erecting anything very large or of very great capacity. If any lad with a receiving set and some thermionic tubes can hook to a tree and take in any wavelength he can tune to, will not tree radio vastly increase the devotees of this particular variety of indoor sport? The matter is one of some importance, inasmuch as many valuable recruits to the radio world have come from amateur ranks, and many a radio engineer had got his first taste for the fascinating art through a homemade tuning coil and detector, under the attic roof. The greater the amateur wireless world, the quicker the development of the art as a whole.

Explorers, discoverers, engineers in far places, the forest service, the woodsman, all have use for the new development. Moreover the tree as an antenna offers unusual possibilities for the investigation of atmospheric phenomena and for what may be called the physics of botany (or the botany of physics) and perhaps is the road by which the unsolved puzzle of growth may be studied.

Meanwhile, it is a thought not without great power to move the sensitive imagination that every tree, growing everywhere, is a wireless tower and antenna and that, as General Squier says, "It is significant that a tree, possessing utility and natural strength, architectural beauty of design and endurance far superior to artificial structures prepared by man, should be able yet further to minister to his needs".

***Electrical Experimenter* (July 1919), p. 204**

Talking Through The Trees

by

Major-General George O. Squier

(Chief Signal Officer, US Army)

How Transatlantic Radio messages Are Copied Via Tree Antenna

As long ago as 1904, the author conducted some experiments with a view to utilizing growing trees as antennae for radio-telegraphy and discovered the efficacy, in a general way, of using a direct metallic contact to certain trees (principally Eucalyptus) to increase the audibility of radio signals. My attention was first called to this phenomenon during the course of summer maneuvers of the Army at Camp Atascadero, CA, where, due to the prevalence of the dry season of the soil, it was found that the regular Army buzzer telephone and telegraph sets were inoperative with any ordinary ground or earth but became operative when connected to a metallic nail driven in the trunk or roots of a live tree. This incident led the author to pursue the subject experimentally in the autumn of 1904 continuing the experiments to the range of frequencies than employed in radio-telegraphy.

Tree Antennae

In connection with the organization and development of Transatlantic radio reception, which was carried out during the period of the war to provide against the possibility of the interruption of the submarine cable system, the Signal Corps established a chain of special receiving stations in different parts of the United States to copy and record enemy and Allied radio messages from European stations for the information of our Army General Staff.

In the prosecution of this work, directions were given to the Signal Corps Laboratory at Camp Alfred Vail, Little Silver, NJ, and also to the experimental staff in Washington to test the efficiency of growing trees as receiving antenna, in connection with this service, using the vastly superior technique and facilities now represented in the radio Art as compared with the crude apparatus with which the discovery was made in 1904. With a collection of apparatus representing the most advanced state of the radio Art, the problem, as a war measure, was attacked anew and has now reached a point where a very brief outline of some of the physical results obtained should be presented in the interests of the development of the Art in general. Since the phenomena involved embrace a variety of physical problems rather than strictly engineering ones, these data are presented in the hope that our scientists may see in the experiments some points of departure for further research.

It was immediately discovered that with the sensitive amplifiers now in use it was possible to receive signals from the principal European stations by simply laying a small wire netting on the ground beneath the tree and connecting an insulated wire to a nail driven in the tree well within the outline of the tree top. See Figure 1. Messages having been received from England, France, Germany and Italy.

This encouraging first result justified a more careful examination of the phenomena and the most suitable arrangement of circuits for the purpose.

And Now For the Floraphone

The messages carried over this tree telephone and telegram system have been named by the writer. They are to be ‘floragrams’. The tree telephone is to be a ‘floraphone’; the tree telegraph a ‘floragraph’.

The discovery is now announced after experiments covering 15 years, beginning in California and continuing intermittently until the outbreak of the war, when they went forward with vigor as an emergency means of communication. The system was utilized during the war in listening-in on the German radio communication.

The final development took place in a small portable laboratory, purchased from a mail order house, capable of being carried to any place in the woods. A group of soldiers, taking turns at copying messages, assisted the writer in the development of the apparatus.

Without entering into the details of these preliminary experiments here it may be said that one fo the best receiving arrangements is found to be an elevated tree earth-terminal in the upper part of the tree top, and an earth consisting practically of several short pieces of insulated wire, sealed at the outer end, radiating out from a common center, and buried a few inches beneath the surface of the ground in the neighborhood of the tree. See Figure 2.

It was soon found that a tree-antenna could be used efficiently as a multiple radio receiving set over widely different wavelengths, see Figure 3A, receiving either from separate terminals at the same (shown dotted in Figure 3A) or different heights of the tree, or in series from the same terminal. See Figure 3B.

This same type of circuit was employed in an inverse manner for radio-telephonic transmitting purposes, although the experiments thus far have been limited to short distances. It was found that 2-way radio-telephonic communication was easily established with remarkably low values of transmitting antenna current. See diagram of this test illustrated at Figure 4.

The flexibility of this arrangement is very striking. The linking up of wire and wireless methods was found to be both convenient and efficient. Radio-telephonic messages from airplanes were readily received by the tree-antenna arrangement and transferred thence to the wire system of the city of Washington and finally received at any point desired.

Furthermore, radio-telephonic transmission through the tree-antenna was received by another tree-antenna, and automatically returned to the sender on a wire system, thus making a complete circuit. Illustrated at Figure 6.

Long distance reception on any wave length from all the larger European stations and from our ships at sea was easily accomplished and traffic copied on a 24 hour schedule by the regular enlisted operators of the Signal Corps. A small portable house serving as a field laboratory was erected in the midst of the forest area on Grant Road, Washington DC, and here was assembled a collection of amplifiers from the Army, navy, the British and French, and American manufacturers. With these facilities it was a matter of a few days to test out, at least in a superficial manner, a large number of proposed arrangements using trees as antenna.

Figure 7 shows how a number of trees can be connected up to give the effect of one large antenna.

In France officers of the Signal Corps, by the simple device of driving a spike into a tree-trunk to which connect the audion set which wireless operators use to make up wireless messages, it was found possible to listen in on communications between German airplanes and the German lines. Messages were thus intercepted in spots into which it would have been impossible to transport a field wireless apparatus. See Figure 8.

The Physics of Tree Wireless

We may regard the metallic electrode rigidly driven into the living organism of a tree, as described above, as a potential earth-terminal for the study of the potential distribution on the surface of the earth itself. It has been shown in these experiments that this metallic terminal intimately connected to the earth itself and a part thereof is subject to changes of potential representing the innumerable frequencies required by modern radio-telephony and telegraphy, as well as any other electrical disturbances which may occur on the surface of the earth or the atmosphere above the earth.

It has also been shown, as expected, that we can select from this composite one or more of the different frequencies by tuned electrical loop circuits suitably connected to this electrode and study each in turn, at will, just as color screens can select a particular component of white light. We may, indeed, by means of a highly insulated conductor bring this terminal directly to the laboratory and connect it immediately to the modern thermionic tube and amplify almost at will the particular effects we are studying…

We can consider that trees have been pieces of electrical apparatus from their beginning and with their manifold chains of living cells are absorbers, conductors and radiators of the long electromagnetic waves as used in the radio Art.

For our present purpose we may consider, therefore, a growing tree as a higly organized piece of living earth, to be used in the same manner as we now use the earth as a universal conductor for telephony and telegraphy and other electrical purposes…

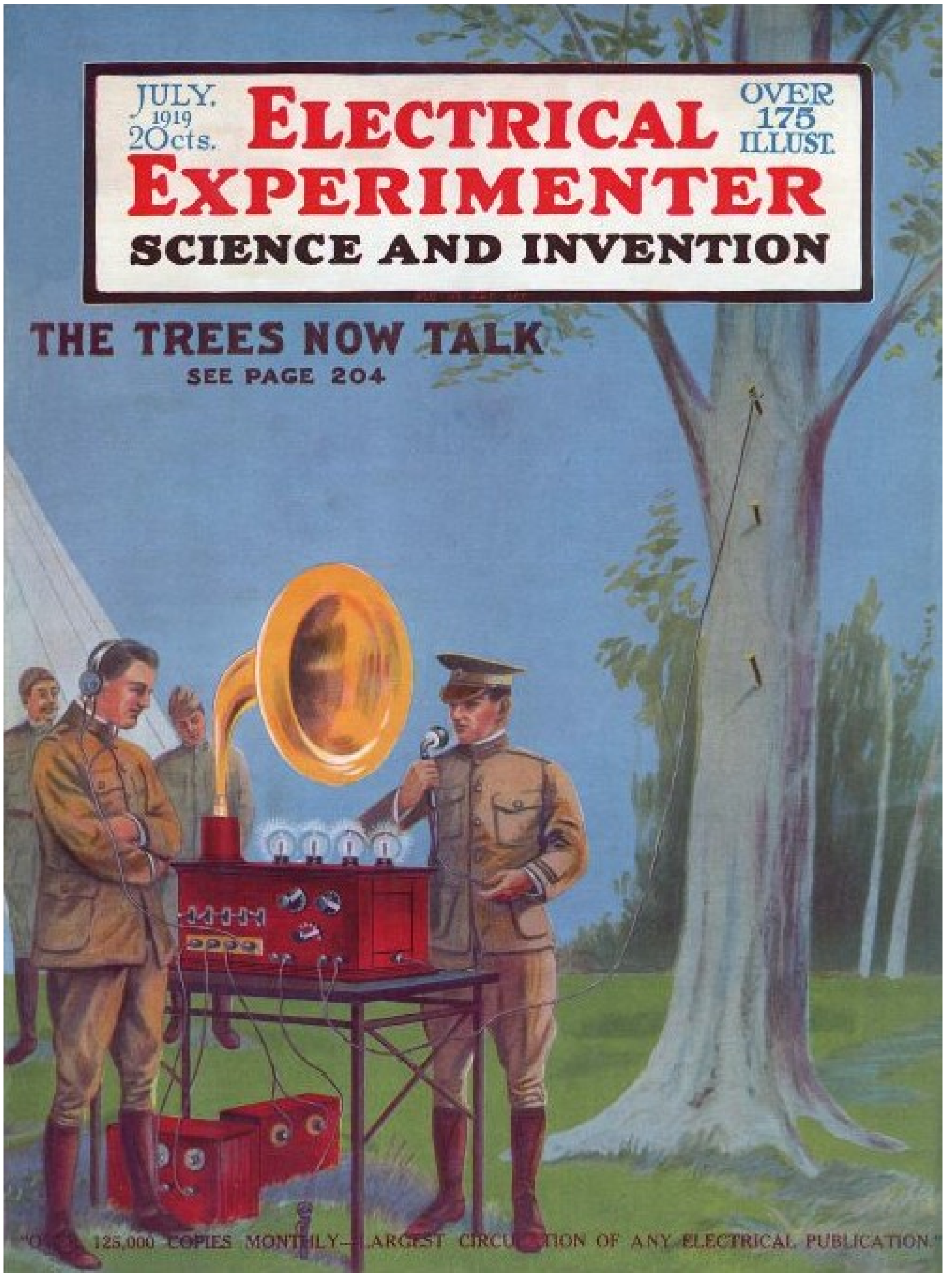
JULY, 1919 20cts.
ELECTRICAL EXPERIMENTER
OVER 175 ILLUST.
SCIENCE AND INVENTION
THE TREES NOW TALK
SEE PAGE 204
"O 125,000 COPIES MONTHLY—LARGEST CIRCU TION OF ANY ELECTRICAL PUBLICATION."

Signal Corps Photo

The U. S. Signal Corps Laboratory Near Washington, D. C., Where Remarkable Results Were Obtained Using Living Trees as Radio Antennae. Signals from Europe and Other Countries Were Easily Copied. Trees Have Been Successfully Used Also for Radiophone Transmission. Insert View: Major General George O. Squier, Inventor of "Tree Wireless."

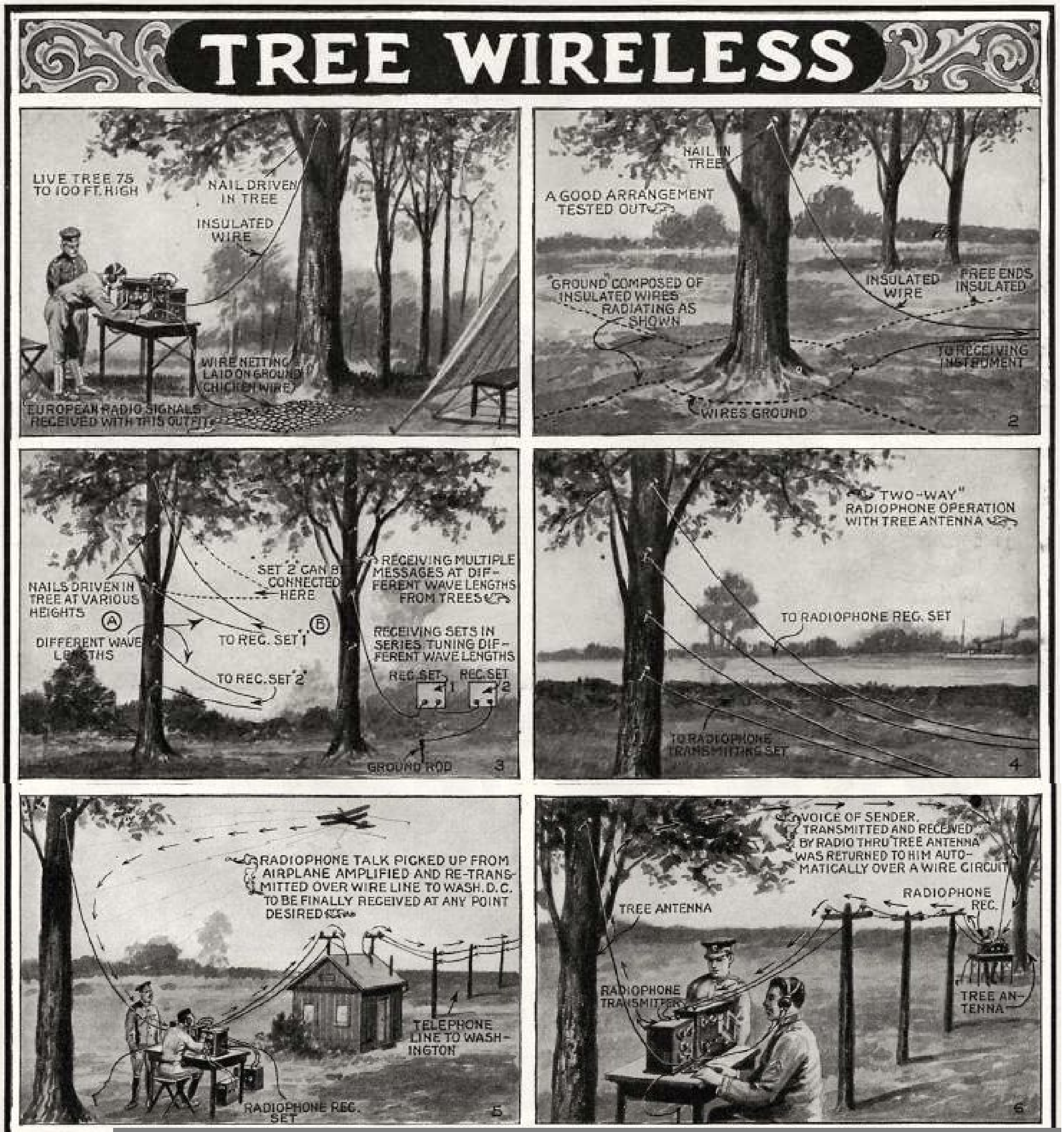
TREE WIRELESS
LIVE TREE 75 TO 100 FT. HIGH
NAIL DRIVEN IN TREE
INSULATED WIRE
WIRE NETTING LAID ON GROUND (CHICKEN WIRE)
"EUROPEAN RADIO SIGNALS" RECEIVED WITH THIS OUTFIT
NAIL IN TREE
A GOOD ARRANGEMENT TESTED OUT
"GROUND" COMPOSED OF INSULATED WIRES RADIATING AS SHOWN
INSULATED WIRE
FREE ENDS INSULATED
TO RECEIVING INSTRUMENT
WIRES GROUND
2
NAILS DRIVEN IN TREE AT VARIOUS HEIGHTS
A
DIFFERENT WAVE LENGTHS
SET "2" CAN BE CONNECTED HERE
TO REC. SET "1"
TO REC. SET "2"
B
RECEIVING MULTIPLE MESSAGES AT DIFFERENT WAVE LENGTHS FROM TREES
RECEIVING SETS IN SERIES, TUNING DIFFERENT WAVE LENGTHS
REC. SET 1
REC. SET 2
GROUND ROD
3
"TWO-WAY" RADIOPHONE OPERATION WITH TREE ANTENNA
TO RADIOPHONE REC. SET
TO RADIOPHONE TRANSMITTING SET
4
RADIOPHONE TALK PICKED UP FROM AIRPLANE AMPLIFIED AND RE-TRANSMITTED OVER WIRE LINE TO WASH. D.C. TO BE FINALLY RECEIVED AT ANY POINT DESIRED
TELEPHONE LINE TO WASHINGTON
RADIOPHONE REC. SET
5
VOICE OF SENDER, TRANSMITTED AND RECEIVED BY RADIO THRU "TREE ANTENNA" WAS RETURNED TO HIM AUTOMATICALLY OVER A WIRE CIRCUIT
TREE ANTENNA
RADIOPHONE REC.
RADIOPHONE TRANSMITTER
TREE ANTENNA
6

British Patent Specification # 149,917

Improvements in & Relating to Radio Communication Systems

George O. Squier

This invention relates to radio communication systems. More particularly, the invention relates to radio transmission and reception through the use of living vegetable organisms such as trees, plants, and the like.

As disclosed in the Specification of my prior Letters Patent No. 25,610 (1904), I have discovered heretofore that tall trees and like growing vegetation possessed electrical conductivity of a certain nature adapted for the reception of signaling electromagnetic waves and capable of forming a part of an antenna or serial with the use of a direct earth "ground" or equivalent point of connection with the tree, in the potential node region.

I have recently discovered that living vegetable organisms generally are adapted for transmission and reception of radio or high frequency oscillations, whether damped or undamped, with the use of a suitable counterpoise. I have further discovered that such living organisms are adapted for respectively transmitting or receiving a plurality of separate trains of radio or high frequency oscillations simultaneously, in the communication of either or both telephonic or telegraphic messages.

In such use of an antenna comprising living vegetable organism and a counterpoise, I have discovered that optimum results are obtained upon arranging the point of connection of the transmitting or the receiving set within a certain range of the tree, or other living vegetable organism; for trees indigenous in the temperate zones, I have discovered that such optimum point of connection is in a region approximately two-thirds of the height of the tree above the exposed surface of the earth.

In accordance with my invention for receiving radio trains of telephonic or telegraphic high frequency oscillations, a plurality of receiving sets attuned respectively to any desired frequency, may be connected by the same leading-in wire to the same tree, whereby such tree serves as the common receiving antenna for the respective individual receiving sets; and, similarly, for transmitting radio trains of telephonic or telegraphic high frequency oscillations, a plurality of transmitting sets respectively resonant to any desired frequency may be connected by a common lead-in wire to the same tree serving as the common antenna.

In the more preferred forms of my invention, I have devised suitable means for selective directional reception of radio oscillations of any desired definite frequency, by the use of a coil having the turns of its windings disposed in substantial parallelism and mounted to be rotated in a horizontal plane, whereby the received oscillations are restricted to the direction coincident with or parallel to a plane lying normal to any plane passing through any turn of the windings of the coil.

The counterpoise comprises preferably a suitable extent of conductor in the form of one or more lengths of wire disposed below the surface of the earth and suitable insulated therefrom; it is desirable to arrange a number of individual counterpoises, each extending rectilinearly in different directions to one another and to employ in directional selection the particular counterpoise extending substantially in the selected direction.

Further features and objects of the invention will be more fully understood from the following description and the accompanying drawings, in which:

Figure 1 is a diagrammatic vertical elevation of a preferred form of the invention;

Fig 1

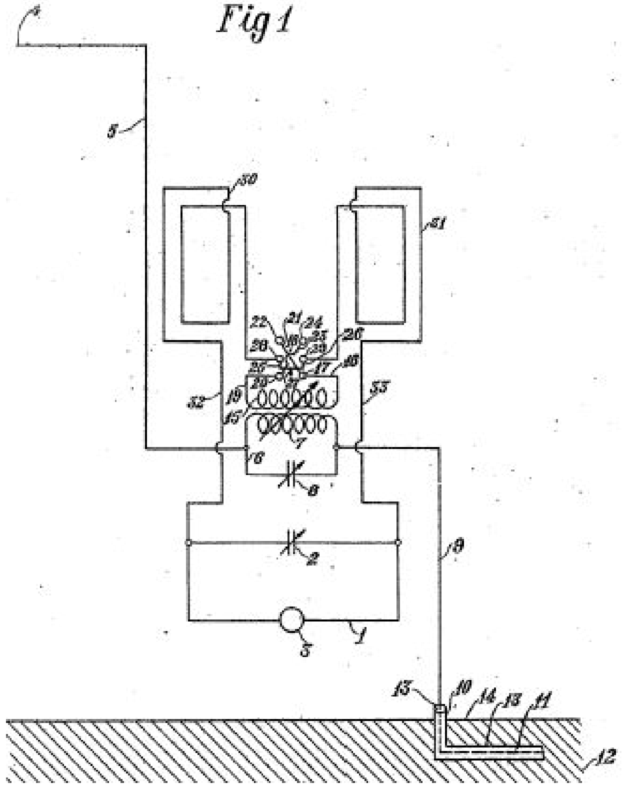
5
30
31
22
21
24
20
16
25
23
26
27
17
28
18
52
19
53
15
7
6
8
2
9
3
1
13
10
14
13
11
12

Figure 2 is a diagrammatic view of one preferred form of an arrangement of receiving sets, embodying my invention; and

Fig 2

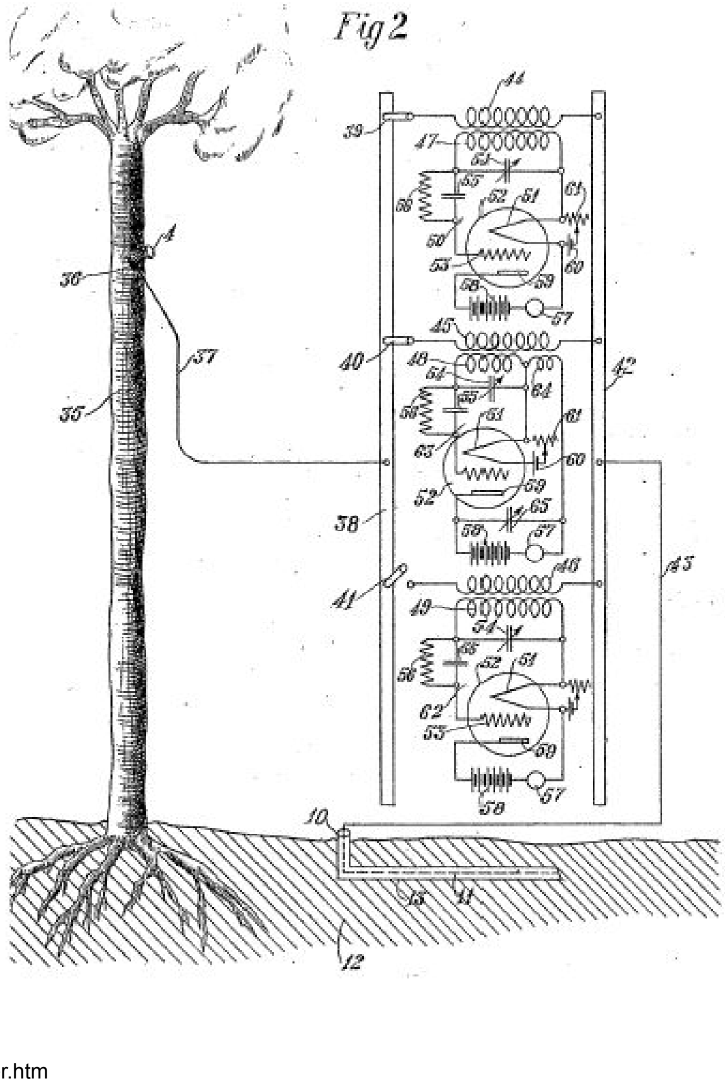
44
39
47
54
66
55
52
51
61
50
60
53
58
59
57
45
40
48
54
55
64
56
51
61
63
60
52
59
65
58
57
46
41
49
54
55
52
51
56
62
53
59
57
58
4
36
37
35
38
42
43
10
13
11
12

Figure 3 is a diagrammatic view of a preferred form of an arrangement of transmission sets embodying my invention.

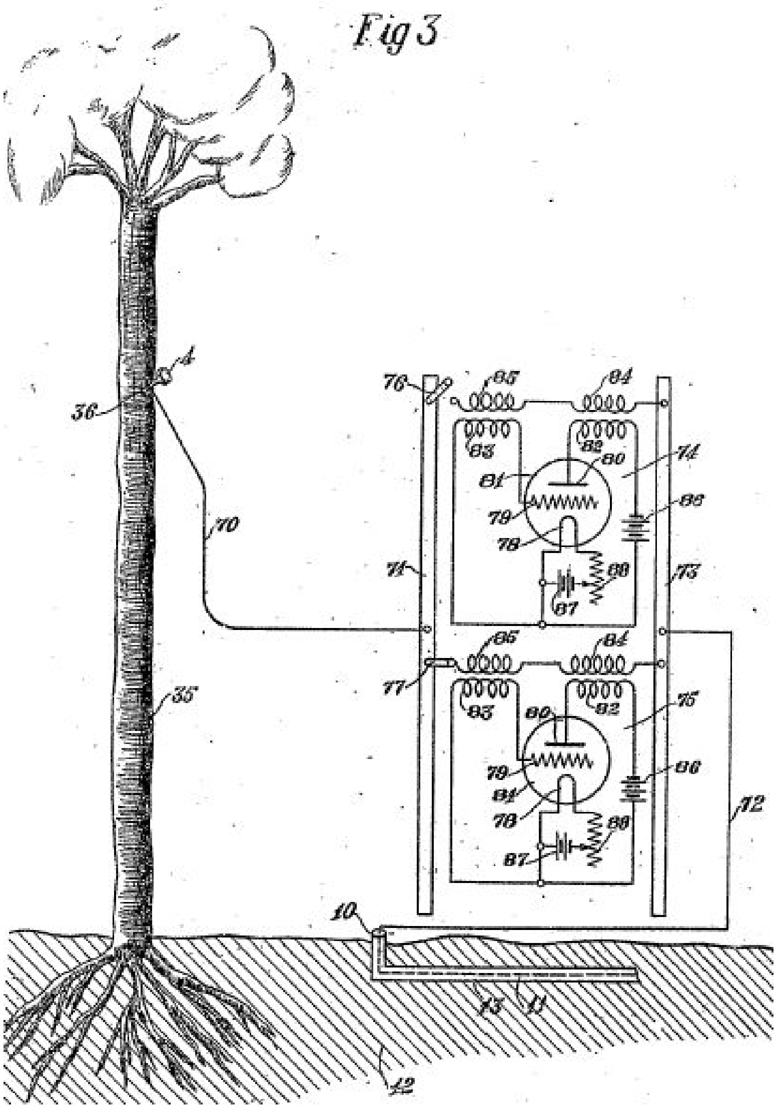

Referring to Figure 1, the circuit designated 1 represents any approved form of resonant receiving circuit comprising the variable condenser 2, and detector, rectifier or amplifier 3, arranged with a telephone or any suitable recording device in the usual manner. Such attuned receiving circuit 1 is suitably coupled through transformer windings or otherwise with proper regulatable reactance with the tree or other living vegetable organism and in Figure 1 I have indicated the location 4 as representing the location of a metallic nail, or equivalent, affixed to and extending within the body of such tree. I have discovered that optimum results for any particular tree are dependent upon the physical contour of the space occupied by the branches and leaves as well as the subterranean portions of the tree such as the roots and that generally the optimum results are attained by affixing the nail 4 within the region of the tree substantially two-thirds of the height of the tree above the surface of the earth.

To such nail 4 is connected one end of the lead-in wire 5, preferably enclosed in suitable insulation, and connected at its other end with the resonant circuit 6 comprising the adjustable inductance coil 7 and the adjustable condenser 8. The lead 9 connects the circuit 6 with the counterpoise 10, preferably comprising one or more metallic wires, 11, extending substantially rectilinearly and electrically insulated from the earth 12 by means of suitable insulation 13, as by enclosing such wire or wires 11 within a suitable insulating coating. Such coating 13 extends preferably an appreciable distance above the surface 14 of the earth 12. In the use of my invention with living antennae, I have arranged the counterpoise 10-13 extending in the rectilinear direction substantially parallel to the plane passing through the transmitting station, the receiving station and the center of the earth. It is advantageous to arrange a plurality of individual counterpoises 10 radiating in different rectilinear directions, with a suitable switching device for connecting the lead 9 with the particular counterpoise parallel to or most closely parallel to the selected direction.

Such coil 7 is preferably a primary coil suitably spacially coupled with an adjustable secondary coil 15 connected in the circuit comprising the variable condenser 2 and 3. In the specific form of the invention shown in Figure 1, such secondary coil 15 is provided with the lead wire 16 connected to the terminal 17 of the reversing switch 18 (shown diagrammatically) the other end of the secondary coil 15 being connected by the cross wire 20 of the reversing switch 18. Such terminal 17 is connected by the cross wire 21 with the terminal 24. The pivoted knife blades 25, 26, having the handle 27, are mounted in the usual manner on the central terminals 28, 29. The coil 30 represents diagrammatically a coil comprising a desired number of windings or successive turns of exposed or insulation covered wire, each turn lying substantially in a common vertical plane and the coil indicated at 31 represents a similar plurality of turns of windings mounted in substantially a common vertical plane. One terminal of the coil 30 is connected with one of the central terminals of the reversing switch 18, say terminal 28, and one terminal of the coil 31 is connected with the other central terminal 29, of the reversing switch 18. The other terminal of the coil 30 is connected by the wire 32 with the leads of the variable condenser 2 in shunt with the detector or amplifier 3, while the other terminal of the coil 31 is connected by the wire 33 to the opposite lead of the variable condenser 2 and the detector or amplifier 3.

Such set of coils 30, 31 are suitably mounted to be rotated horizontally, whereby all turns of the coils 30, 31, at any given position lie substantially in a common vertical plane. Such coils 30, 31 may be arranged on a common wooden or like insulating frame of cylindrical, rectangular or other desired contour on cross section.

By virtue of such an arrangement, the primary and secondary circuits are respectively directly exposed to and simultaneously electromagnetically acted upon by the incoming oscillations, and by means of the reversing switch 18 the oscillations generated in the coil 30 may be neutralized or accumulated relatively to the oscillation generated in the coil 31 to produce either the "barrage" or the "amplifying" effect.

By means of such "barrage" effect, the oscillations of any given wavelength directed in any absolute direction are cut out from the detector or amplifying circuit and enable oscillations of similar or the same wave length received in any definite direction to be conducted to the detector or amplifying circuit. By means of such accumulating effect of the coils 30, 31, and the switch 18, the waves of any definite wavelength received in any definite absolute direction are amplified and enable long distance reception at any wave length.

In practice, I have carried out the invention for reception by the use of either a single receiving set or a plurality of receiving sets. Such receiving set or sets may be arranged either to receive damped high frequency oscillations or undamped high frequency oscillations.

In Figure 2 I have illustrated diagrammatically one form of the application of the invention for the simultaneous reception of a plurality of different trains of high frequency oscillations.

The nail 4 is indicated as located in electrical connection with the tree 35 at an optimum point 36, the lead-in wire 37 being connected at its upper end with the nail 4 and at its lower end with the incoming connector bar 38. The connector 38 is provided with the individual switches 39-40-41, etc., corresponding to the number of individual receiving sets. The outgoing connector bar 42 is connected by the lead 43 with the counterpoise 10, preferably constructed and arranged as set forth hereinabove.

In suitable relation with the incoming connector bar 38 and outgoing connector bar 42 and the switches 39, 40, 41 are arranged the respective primary coils 44, 45, 46, etc., respectively coupled in any approved arrangement with the secondary coils 47, 48, 49.

The receiving set 50 is arranged for the reception of undamped waves and is shown of the vacuum tube valve type. The terminals of the secondary 47 are suitably connected in the input circuit including the filament cathode 51 of the vacuum valve 52 and the grid 53 and comprises the variable condenser 54 in shunt with the secondary 47, the condenser 55 and the grid leak resistance 56. The output circuit comprises the battery 58, the filament cathode 51, and the telephone receiver or other audible, or any visual indicator 57. The heating circuit of the filament cathode includes the battery 60 and the variable resistance 61.

The receiving set 62 is shown of a similar undamped oscillation receiving type and like elements are designated by like reference numbers.

The receiving set 63 is indicated of the damped oscillation receiving type and comprises the additional inductance turns 64 having one terminal connected to one terminal of the filament cathode 51 and its other terminal connected to the variable condenser 65 in shunt with the telephone receiver 57 and the storage battery 58. The remaining elements of the receiving set 63 correspond to like elements of the receiving set 50 and are designated by like reference numbers.

In Figure 2 I have shown the switch 39 in closed position with the primary inductance 44 of the receiving set 50, the switch 40 in closed position with the primary inductance 45 of the receiving set 63 and the switch 41 in open position with the primary inductance 46 of the receiving set 62, and accordingly enabling the reception and detection of undamped oscillations of the frequency to which the receiving set 50 is tuned and simultaneously the reception and detection of damped oscillations of the frequency to which the receiving set 53 is tuned.

Figure 3 shows one form of practical application of my invention for transmitting simultaneously radio oscillations, either telephonic or telegraphic, either damped or undamped, or both. The nail 4, lodged at an optimum point 36 in the tree 35, is connected by the lead 70 to the connector bar 71, and the counterpoise 10 is connected by the lead 72 to the connector bar 73.

The transmitting sets 74, 75, etc., of the desired number, are suitably arranged for single or multiple transmission, for which purpose the switches 76, 77, etc., are provided. The transmitting sets 74, 75 are shown of any approved type, such as the oscillating vacuum tube type comprising the filament cathode 78, the grid 79 and anode 80, arranged in the vacuum tube 81, suitably connected with the primary inductance 84, 85. The primary inductances 84, 85 are suitably connected to the connector bar 73 and the connector bar 71 through the respective switches 76, 77. The battery or other source of electric current is indicated at 86. The adjustable heating circuit of the cathode filament 79 is shown comprising the battery 87 and the variable resistance 88.

In Figure 3, the switch 77 is in closed position thus placing the transmitting set 75 in operative connection with the tree serving as the antenna. Upon closing the switch 76 the transmitting set 74 is similarly placed in operative connection with the tree antenna 35. Upon closure of both switches 76, 77 the tree serves as the antenna for the transmission of simultaneous trains of oscillations emitted by the respective transmitting sets 74, 75, modified by a key or telephone transmitter, or other suitable modulator, for the transmission of telegraphic or telephonic messages as desired.

In the use of trees or other living vegetable organisms serving as the antenna or a art thereof, I have discovered from tests that such tree possess impedance consisting of two components, resistance and reactance, the latter being usually condensive, for oscillation within the range of present day wavelengths. The tests also show that the apparent capacity of a tree serving as an antenna is substantially proportional to the height of contact of the lead wire connecting the transmission set with the tree.

The tests also show that the apparent resistance of a tree serving as an antenna is appreciably greater than the resistance of the ordinary metallic antenna, and it is accordingly desirable to modify the resistance or alter the design of the receivers, if it desired to increase the effectiveness of the energy of the electromagnetic radiation impinging on the tree; however, the ordinary receivers of present design may be used and are responsive to any wave length of radiation employed in present day practice.

Whereas I have described my invention by reference to specific forms thereof, it will be understood that many changes and modifications may be made without departing from the spirit of the invention as defined by the appended claims.

An alternative electric power generating system that draws energy from a seemingly unlikely yet abundant, eminently renewable and virtually free power source has been submitted for patenting by MagCap Engineering, LLC, Canton, Mass., in collaboration with Gordon W. Wadle, an inventor from Thomson, Ill. Wadle has invented a way to capture the energy generated by a living non- animal organism --- such as a tree.

Source: KeelyNet / Jerry Decker email (12/21/05)

"Unlimited Electric Energy from the Environment?"

Chris Lagadinos, president of MagCap, developed circuitry that converts this natural energy source into useable DC power capable of sustaining a continuous current to charge and maintain a battery at full charge.

"As unbelievable as it sounds, we've been able to demonstrate the feasibility of generating electricity in this manner," said Wadle.

"While the development is in its infancy, it has the potential to provide an unlimited supply of constant, clean energy without relying on fossil fuels, a power generating plant complex or an elaborate transmission network."

Wadle likened the invention to the discovery of electricity over 200 years ago when charged particles were harnessed to create an electric current. "Now we've learned that there is an immense, inexhaustible source of energy literally all around us that can be harnessed and converted into usable electric power," he said. Ultimately, it should prove to be more practical than solar energy or wind power, and certainly more affordable than fuel cells, he added.

Wadle said he got the original idea of harnessing a tree for electrical energy from studying lightening, more than 50 percent of which originates from the ground. This prompted him to develop the theories resulting in a method to access this power source.

Lagadinos then designed circuitry that filtered and amplified these energy emanations, creating a useable power source.

Basically, the existing system includes a metal rod embedded in the tree, a grounding rod driven into the ground, and the connecting circuitry, which filters and boosts the power output sufficient to charge a battery.

In its current experimental configuration, the demonstration system produces 2.1 volts, enough to continuously maintain a full charge in a nickel cadmium battery attached to an LED light. "Think of the environment as a battery, in this case," said Lagadinos, "with the tree as the positive pole and the grounding rod as the negative."

Lagadinos said the system could be enhanced enough to generate 12 volts and one amp of power, "a desirable power level that could be used to power just about anything," he said. It is enough power to charge batteries for any type of vehicle, including hybrids and electric cars, or to use with an AC converter to produce household power, he added. The LED industry is a prime example of a potential user of this power source.

While the basic concept of this invention -- using a tree to generate electric power -- seems too incredible to be true, Lagadinos said it can be demonstrated quite simply. "Simply drive an aluminum roofing nail through the bark and into the wood of a tree -- any tree -- approximately one half inch; drive a copper water pipe six or seven inches into the ground, then get a standard off-the-shelf digital volt meter and attach one probe to the pipe, the other to the nail and you'll get a reading of anywhere from 0.8 to 1.2 volts of DC power," he said.

"You can't do anything with it in that form because it is 'dirty' -- i.e. highly unstable and too weak to power anything," he added. In order to properly harness this potential energy source, MagCap devised two test circuits: one with three capacitors that were connected in parallel by means of a switch and charged to 0.7 volts each.

When fully charged they are switched to a series mode, multiplying the voltage to 2.1 volts and flashing an LED to show that sufficient power could be generated to produce a useable result. The second circuit included a filtering device to stabilize and "clean" the current so it could be used to charge and maintain a NiCad battery.

The battery then could be connected to the LED to keep the LED lit continuously. Wadle pointed out that there seems to be no limit to the amount of power that can be drawn from an individual tree, no matter how many "taps" are inserted -- each produces the same amount of energy, an average of 0.7 - 0.8 volts. Size of the tree also seems not to matter.

Interestingly, while conventional wisdom would seem to indicate that the tree draws much of its energy from photosynthesis via its leaves, the voltage output actually increases to 1.2-1.3 volts in the winter after the leaves have fallen.

Patents by George O. Squire:

Apparatus for using telephone-subscribers' loops for transmission and reception of broadcasting
SQUIER GEORGE O
EC: H04H1/08 IPC: H04H1/08; H04H1/04
US1742422
1930-01-07

Apparatus for effecting electrical signaling
SQUIER GEORGE O; LOUIS COHEN
EC: H04L27/06 IPC: H04L27/06; H04L27/06
US1608252
1926-11-23

Aerial attachment device for radiofrequency signal systems
SQUIER GEORGE
EC: H01Q1/46 IPC: H01Q1/46; H01Q1/44
Publication info: US1584197
1926-05-11

Electrical signaling
SQUIER GEORGE O; MAUBORGNE JOSEPH O
EC: H01F21/02 IPC: H01F21/02; H01F21/02
US1641608
1927-09-06

Apparatus for electrical signaling
SQUIER GEORGE
EC: H04B1/034 IPC: H04B1/034; H04B1/02
US1531629
1925-03-31

Improvements in the Art of Telegraphy
SQUIER GEORGE OWEN; AUSTIN LOUIS WINSLOW
EC: H04J1/00 IPC: H04J1/00; H04J1/00
GB108763
1917-08-23

Improvements in the Art of Electrical Signalling
SQUIER GEORGE; AUSTIN LOUIS WINSLOW
EC: H04J1/00 IPC: H04J1/00; H04J1/00
GB108230
1917-08-02

Improvements in Duplex and Multiplex Telegraphy
SQUIER GEORGE; COHEN LOUIS
EC: H04J1/00 IPC: H04J1/00; H04J1/00

GB103771
1917-02-08

Combined sound and light distributing apparatus
SQUIER GEORGE
EC: F21V33/00; H05K11/00 IPC: F21V33/00; H05K11/00; F21V33/00 (+1)
US1999579
1935-04-30

Radio signaling apparatus
SQUIER GEORGE
EC: H04B3/56 IPC: H04B3/56; H04B3/54
US1791541
1931-02-10

Improvements in or relating to Receiving Systems for Radio Telegraphy or Telephony
SQUIER GEORGE
EC: H01Q1/46; H04H1/04 IPC: H01Q1/46; H04H1/04; H01Q1/44 (+1)
GB191417487
1914-11-19

Improvements in or relating to Radio-telegraphic and Radio-telephonic Receiving Systems
SQUIER GEORGE
EC: H04M11/06 IPC: H04M11/06; H04M11/06
GB191403191
1915-02-08

Improvements in or connected with Multiplex Telephony and Telegraphy
SQUIER GEORGE
EC: H04J1/00 IPC: H04J1/00; H04J1/00
GB191030003
1911-12-27

Improvements in or relating to Receiving Systems for Radio Telegraphy or Telephony
SQUIER GEORGE
EC: H01Q1/46; H04H1/04 IPC: H01Q1/46; H04H1/04; H01Q1/44 (+1)
GB191417487
1914-11-19

Improvements in or relating to Radio-telegraphic and Radio-telephonic Receiving Systems
SQUIER GEORGE
EC: H04M11/06 IPC: H04M11/06; H04M11/06
GB191403191
1915-02-08

Improvements in or connected with Multiplex Telephony and Telegraphy
SQUIER GEORGE
EC: H04J1/00 IPC: H04J1/00; H04J1/00
GB191030003
1911-12-27

Improvements in or connected with Telegraphic Apparatus
CREHORE ALBERT CUSHING (US); SQUIER GEORGE
GB189919867
1900-10-03

Improvements in Telegraphic Apparatus
CREHORE ALBERT CUSHING (US); SQUIER GEORGE

GB189918980
1900-09-20

Improvements in Telegraphic Apparatus
CREHORE ALBERT CUSHING (US); SQUIER GEORGE
GB189917722
1900-09-01

Improvements in Submarine Cable Telegraphs
CREHORE ALBERT C (US); SQUIER GEORGE
GB189823556
1900-02-08

Method and apparatus for using telephone subscribers' loops for transmission and reception of broadcasting
SQUIER GEORGE
EC: H04B3/56; H04H1/04 IPC: H04B3/56; H04H1/04; H04B3/54 (+1)
FR644877
1927-10-16

TELEPHONE AND BROADCASTING SYSTEM
SQUIER GEORGE
CA285479
1928-12-11

HIGH FREQUENCY CURRENT SIGNALLING
SQUIER GEORGE
CA244330
1924-11-11

ELECTRIC SIGNALLING SYSTEM
SQUIER GEORGE
CA233299
1923-08-07

TREE TELEPHONY AND TELEGRAPHY
SQUIER GEORGE
CA232487
1923-07-03

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

rexresearch.com

Anthony SUTERA, *et al.*

Spray-On Antenna

Ad

Bluebeam® Revu® Official Site - Create, Markup & Edit In PDF

Eliminate Repetitive Processes So You Can Focus On What Really Matters bluebeam.com

OPEN

Save up to $50 + free delivery
on your first order*
*See terms & conditions.
ARROWHEAD
SHOP NOW >
Ad

Save up to $50 + free delivery
on your first order*
*See terms & conditions.
ARROWHEAD
SHOP NOW >
Ad

Patents :

SUTERA -- MATERIAL USED FOR PROPAGATION, EMISSION & ABSORPTION OF EM RADIATION -- WO2012078362 // US2012146855

Nano-Copper

Nano-Antennas

---

http://news.cnet.com/8301-17938_105-57376903-1/spray-on-antenna-wireless-in-a-can/
February 13, 2012

Spray-on antenna: Wireless in a can

by

Amanda Kooser

CHAMTECH
SPRAY ON ANTENNA SYSTEM
DANGER!
NET WT. 10 oz /283 g

Chamtech's spray-on antenna uses a nano material to provide a low-power boost to antenna range. The wireless-in-a-can product may some day bring an end to unsightly cell towers. The antenna can, seen next to a closeup of the nano material.
(Credit: Video screenshot by Amanda Kooser/CNET)

It sounds like a particularly suspicious late-night infomercial: Spray your way to a better wireless signal! Improve your range! Save battery! Transmit over great distances under water!

But Chamtech's spray-on antenna is a real product with some impressive claims. It can be sprayed on almost any surface, even trees and orange barrels. It doesn't suck up power. It works in a mysterious nanotech way.

Here's how I imagine the antenna process goes:

Step 1: Spray antenna material on surface.
Step 2: Connect phone to material.
Step 3: ????
Step 4: Make a phone call to the moon.

Chamtech co-founder Anthony Sutera imagines a world where wireless antenna towers are replaced with nano-paint on walls, and issues like iPhone Antennagate are a thing of the past.

"We have come up with a material that when you spray it on, it lays out just in the right pattern and all of these little capacitors charge and discharge extremely quickly in real time and they don't create any heat," Sutera says in a video presentation about the product.

One of Chamtech's tests turned an RFID chip with a 5-foot range into an RFID chip with a 700-foot range. The company lists a spray antenna kit on its site, but pricing for the public is not revealed. The U.S. government is reportedly already playing with the new material.

If all these claims bear out, then I can see everybody wanting to get their hands on a fresh can full of antenna. My only question is where in the grocery store it will be stocked: with the spray cheese or with the gold food paint?

---

http://www.neowin.net/news/the-amazing-spray-on-antenna
12 February 2012

The amazing spray-on antenna

by

Tyler Holman

Wouldn't it be great if you could just pop out a can of spray-on antenna and boost your signal whenever it was running low? Chamtech Enterprises is hoping to make that dream a reality, and if you happen to work for the government, it already is, according to Chamtech's Solve for X presentation.

Chamtech took to Google's Solve for X to show off the new technology. The company plans to turn its focus towards mobile phones and medical devices, offering a quick solution to boosting signals from existing antennas, or creating new ones. Chamtech's 'antenna in a can' is far more efficient than traditional antenna models, offering energy savings equal to 12 times the amount of energy generated by solar and wind in the US annually - and it even works great underwater.

A traditional antenna would require thousands of watts to send out a signal with a one mile range underwater. Chamtech's can do that with only three watts, and have a stronger signal to boot. So how does all of this work?

The truth is, we don't really know. According to Chamtech's co-founder Anthony Sutera, he and his team came up with it in his living room two years ago. It works by manipulating magnetic and radio signals through mysterious organic materials, and you can spray it on any virtually any surface and hook into it with a flexible circuit cable.

According to Sutera, the US government has had a lot of success with the technology, getting better performance out of it than their existing portable antennas, which he described as some of the best around. With the efficiency and mobility offered, it could even be used to rapidly deploy new infrastructure in disaster areas.

Sutera and chief technology officer Rhett Spencer have coated their car antennas with the stuff, boasting that they can now listen to radio stations in Salt Lake City fifty miles away, with 10,000-foot mountain range in between. Within a few months, they hope to be thinking about financing the company, and are looking at venture capital options to help them bring their technology to you.

For his part, Spencer can't wait. “Can you imagine the infrastructure side of things? Telecomm under the oceans, Internet infrastructure, ships and satellite communications in the sea– they can do it out under the water.” Check out video of Chamtech's Solve for X presentation below.

---

http://www.youtube.com/watch?feature=player_embedded&v=4efE_gO9lFo

---

http://chamtechops.com/news/

Chamtech / Sutera Patents

MATERIAL USED FOR AT LEAST ONE OF PROPAGATION, EMISSION AND ABSORPTION OF ELECTROMAGNETIC RADIATION
WO2012078362

Also published as: US2012146855

An antenna system and method for fabricating an antenna are provided. The antenna system includes a substrate (310) and an antenna. The antenna includes a conductive particle based material (320) applied onto the substrate (310). The conductive particle based material (320) includes conductive particles and a binder. When the conductive particle based material (320) is applied to the substrate (310), the conductive particles are dispersed in the binder so that at least a majority of the conductive particles are adjacent to, but do not touch, one another.

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit under 35 U.S.C. $119(e) of a U.S. provisional patent application filed on Nov. 22, 2010 in the U.S. Patent and Trademark Office and assigned Ser. No. 61/416,093, a U.S. provisional patent application filed on Apr. 8, 2011 in the U.S. Patent and Trademark Office and assigned Ser. No. 61/473,726, a U.S. provisional patent application filed on Apr. 20, 2011 in the U.S. Patent and Trademark Office and assigned Ser. No. 61/477,587, and a U.S. provisional patent application filed on Aug. 2, 2011 in the U.S. Patent and Trademark Office and assigned Ser. No. 61/514,435, the entire disclosure of each of which is hereby incorporated by reference.

BACKGROUND OF THE INVENTION

[0002] 1. Field of the Invention

[0003] The present invention relates to techniques for a material used for at least one of propagation, emission and absorption of electromagnetic radiation. More particularly, the present invention relates to techniques for a conductive particle based material used for at least one of propagation, emission and absorption of electromagnetic radiation.

[0004] 2. Description of the Related Art

[0005] A conventional antenna is a device with an arrangement of one or more conductive elements that are used to generate a radiating electromagnetic field in response to an applied alternating voltage and the associated alternating electric current, or can be placed in an electromagnetic field so that the field will induce an alternating current in the antenna and a voltage between its terminals. The conductive elements employed in the conventional antenna are typically fabricated from solid metallic conductors. However, the use of solid metallic conductors is limiting.

[0006] Therefore, a need exists for an improved material used for at least one of propagation, emission and absorption of electromagnetic radiation, and implementations of the improved material.

SUMMARY OF THE INVENTION

[0007] An aspect of the present invention is to address at least the above-mentioned problems and/or disadvantages and to provide at least the advantages described below. Accordingly, an aspect of the present invention is to provide techniques for a conductive particle based material used for at least one of propagation, emission and absorption of electromagnetic radiation.

[0008] In accordance with an aspect of the present invention, an antenna system is provided. The antenna system includes a substrate and an antenna. The antenna includes a conductive particle based material applied onto the substrate. The conductive particle based material includes conductive particles and a binder. When the conductive particle based material is applied to the substrate, the conductive particles are dispersed in the binder so that at least a majority of the conductive particles are adjacent to, but do not touch, one another.

[0009] In accordance with another aspect of the present invention, an antenna enhancer system is provided. The antenna enhancer system includes an antenna and an antenna enhancer. The antenna enhancer includes a conductive particle based material. The antenna enhancer is disposed adjacent to and offset from the antenna. The conductive particle based material comprises conductive particles and a binder. When the conductive particle based material is disposed adjacent to and offset from the antenna, the conductive particles are dispersed in the binder so that at least a majority of the conductive particles are adjacent to, but do not touch, one another.

[0010] In accordance with yet another aspect of the present invention, a method for fabricating a conformable antenna is provided. The method includes selecting a substrate on which to fabricate an antenna, selecting a template corresponding to an antenna design, the template comprising one or more cut out portions, applying a conductive particle based material, through the one or more cutout portions of the template, and onto the substrate to form the antenna, and fixing a coupler of a feed line to the antenna. The conductive particle based material comprises conductive particles and a binder. When the conductive particle based material is applied to the substrate, the conductive particles are dispersed in the binder so that at least a majority of the conductive particles are adjacent to, but do not touch, one another.

[0011] Other aspects, advantages, and salient features of the invention will become apparent to those skilled in the art from the following detailed description, which, taken in conjunction with the annexed drawings, discloses exemplary embodiments of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] The above and other aspects, features, and advantages of certain exemplary embodiments of the present invention will be more apparent from the following description taken in conjunction with the accompanying drawings, in which:

[0013] FIG. 1 is a captured image of a conductive particle based material according to an exemplary embodiment of the present invention;

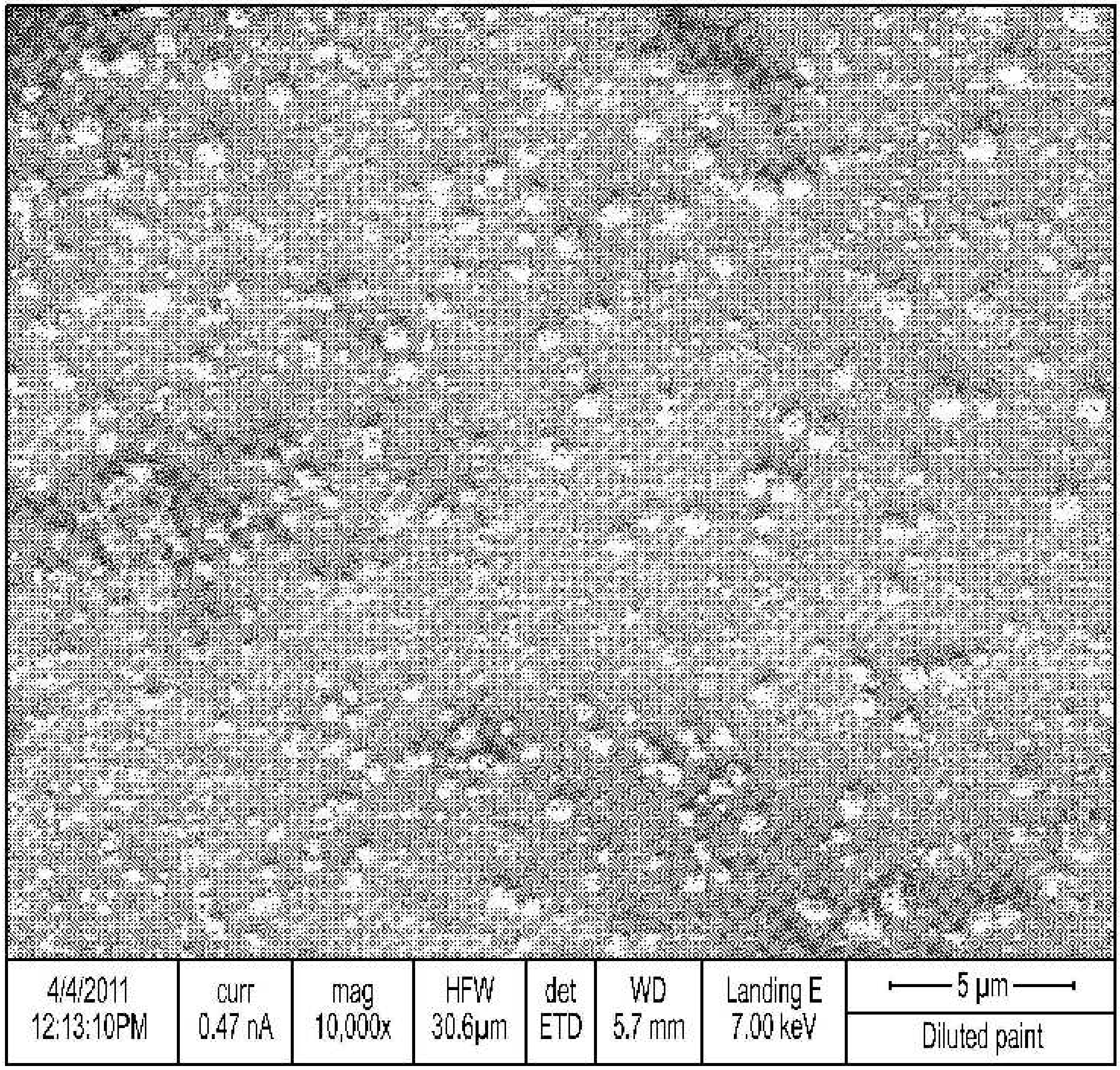
4/4/2011
12:13:10PM
curr
0.47 nA
mag
10,000x
HFW
30.6μm
det
ETD
WD
5.7 mm
Landing E
7.00 keV
5 μm
Diluted paint

FIG. 1

[0014] FIG. 2 illustrates a conductive particle based antenna according to an exemplary embodiment of the present invention;

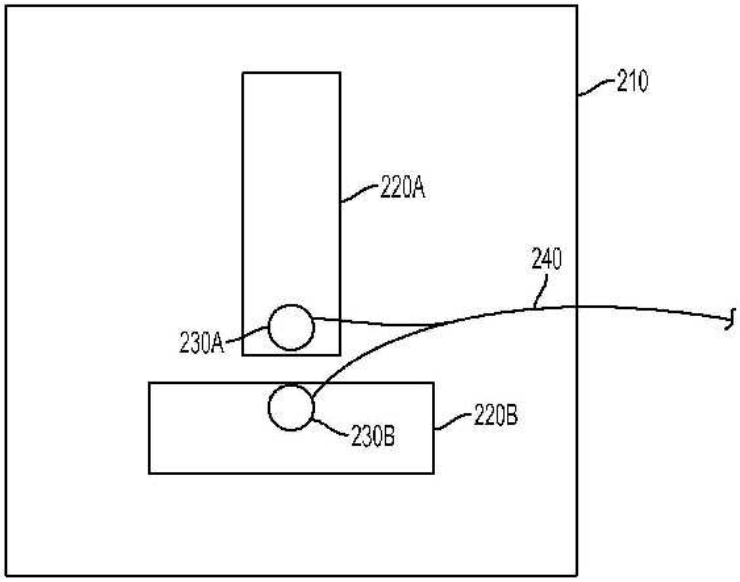

FIG. 2

[0015] FIG. 3 illustrates a structure of a conductive particle based antenna according to an exemplary embodiment of the present invention;

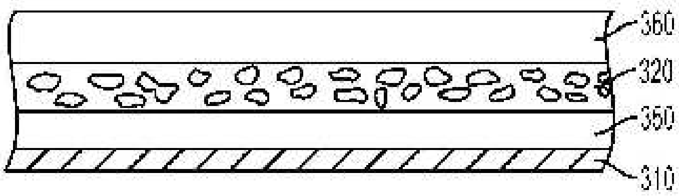

FIG. 3

[0016] FIG. 4 illustrates an implementation of a conductive particle based antenna enhancer according to an exemplary embodiment of the present invention;

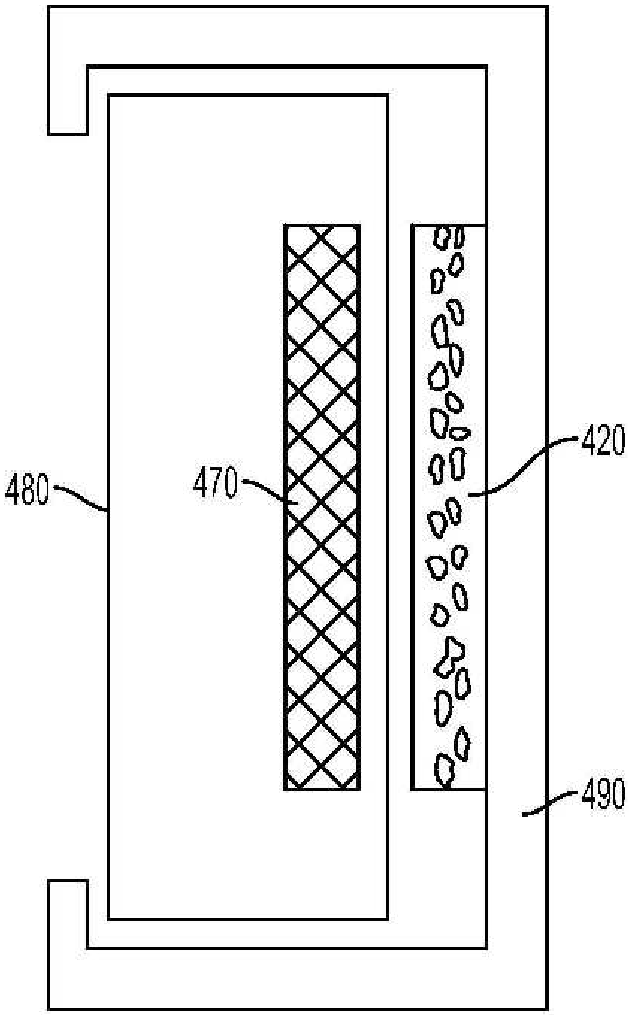

FIG. 4

[0017] FIG. 5 illustrates a structure of a coated conductive particle based antenna enhancer according to an exemplary embodiment of the present invention;

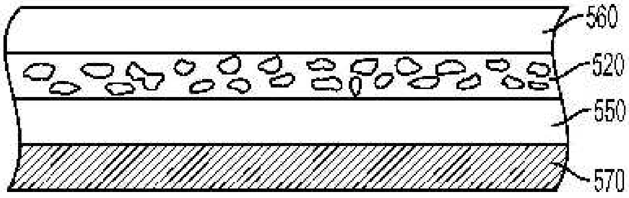

FIG. 5

[0018] FIG. 6 illustrates an antenna partially coated with a conductive particle based antenna enhancer according to an exemplary embodiment of the present invention;

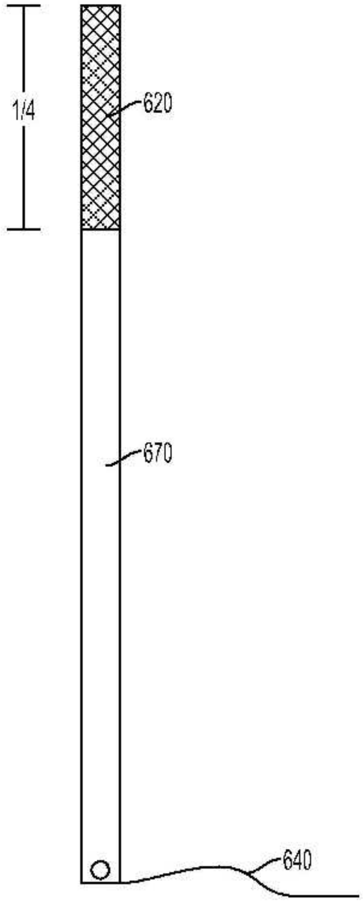

FIG. 6

[0019] FIG. 7 illustrates a template used to fabricate a conductive particle based conformable antenna according to an exemplary embodiment of the present invention;

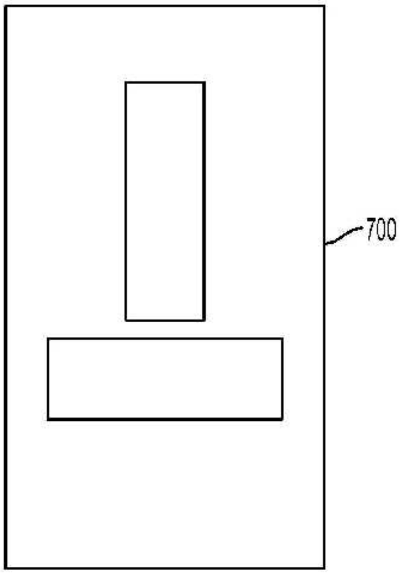

FIG. 7

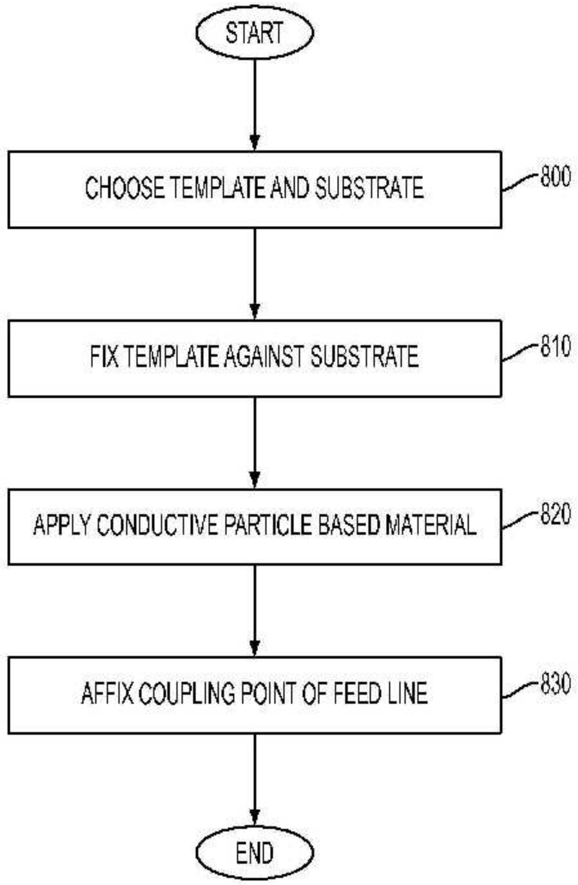
START
CHOOSE TEMPLATE AND SUBSTRATE
800
FIX TEMPLATE AGAINST SUBSTRATE
810
APPLY CONDUCTIVE PARTICLE BASED MATERIAL
820
AFFIX COUPLING POINT OF FEED LINE
830
END

FIG. 8

[0020] FIG. 9 illustrates a method for fabricating a conductive particle based conformable antenna using a computerized device according to an exemplary embodiment of the present invention; and

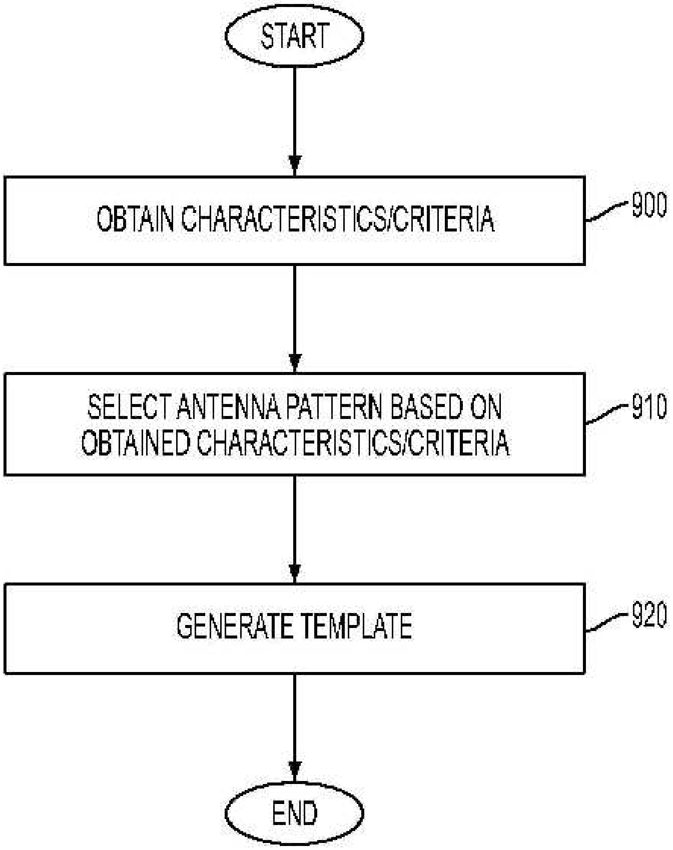
START
OBTAIN CHARACTERISTICS/CRITERIA
900
SELECT ANTENNA PATTERN BASED ON OBTAINED CHARACTERISTICS/CRITERIA
910
GENERATE TEMPLATE
920
END

FIG. 9

[0021] FIG. 10 illustrates a structure of computerized device used for fabricating a conductive particle based conformable antenna according to an exemplary embodiment of the present invention.

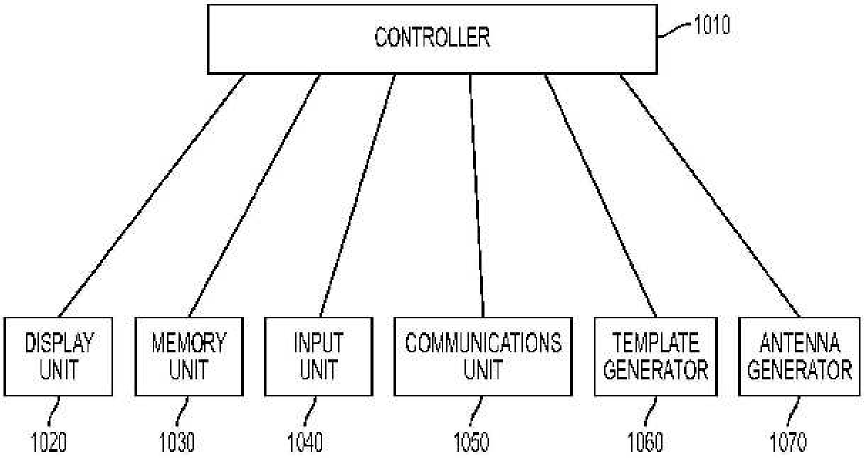
CONTROLLER
1010
DISPLAY UNIT
MEMORY UNIT
INPUT UNIT
COMMUNICATIONS UNIT
TEMPLATE GENERATOR
ANTENNA GENERATOR
1020
1030
1040
1050
1060
1070

FIG. 10

[0022] Throughout the drawings, like reference numerals will be understood to refer to like parts, components, and structures.

DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS

[0023] The following description with reference to the accompanying drawings is provided to assist in a comprehensive understanding of exemplary embodiments of the invention as defined by the claims and their equivalents. It includes various specific details to assist in that understanding but these are to be regarded as merely exemplary. Accordingly, those of ordinary skill in the art will recognize that various changes and modifications of the embodiments described herein can be made without departing from the scope and spirit of the invention. In addition, descriptions of well-known functions and constructions are omitted for clarity and conciseness.

[0024] The terms and words used in the following description and claims are not limited to the bibliographical meanings, but, are merely used by the inventor to enable a clear and consistent understanding of the invention. Accordingly, it should be apparent to those skilled in the art that the following description of exemplary embodiments of the present invention are provided for illustration purpose only and not for the purpose of limiting the invention as defined by the appended claims and their equivalents.

[0025] It is to be understood that the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a component surface" includes reference to one or more of such surfaces.

[0026] As used herein, the term "substantially" refers to the complete or nearly complete extent or degree of an action, characteristic, property, state, structure, item, or result. For example, an object that is "substantially" enclosed would mean that the object is either completely enclosed or nearly completely enclosed. The exact allowable degree of deviation from absolute completeness may in some cases depend on the specific context. However, generally speaking the nearness of completion will be so as to have the same overall result as if absolute and total completion were obtained. The use of "substantially" is equally applicable when used in a negative connotation to refer to the complete or near complete lack of an action, characteristic, property, state, structure, item, or result.

[0027] As used herein, the term "about" is used to provide flexibility to a numerical range endpoint by providing that a given value may be "a little above" or "a little below" the endpoint.

[0028] As used herein, the term "antenna" refers to a transducer used to transmit or receive electromagnetic radiation. That is, an antenna converts electromagnetic radiation into electrical signals and vice versa. Electromagnetic radiation is a form of energy that exhibits wave-like behavior as it travels through space. In free space, electromagnetic radiation travels close to the speed of light with very low transmission loss. Electromagnetic radiation is absorbed when propagating through a conducting material. However, when encountering an interface of such a material, the electromagnetic radiation is partially reflected and partially transmitted there-though. Herein, exemplary embodiments of the present invention described below are directed toward techniques that allow for a more efficient interface by reducing the reflections at the interface.

[0029] In addition, exemplary embodiments of the present invention described below relate to techniques for a conductive particle based material used for at least one of propagation, emission and absorption of electromagnetic radiation. While the techniques for the conductive particle based material may be described below in various specific implementations, the present invention is not limited to those specific implementations and is similarly applicable to other implementations.

[0030] An initial overview of the conductive particle based material is provided below and then specific implementations in which the conductive particle based material is employed are described in detail further below. This initial overview of the conductive particle based material is intended to aid readers in understanding the conductive particle based material that is the basis of various exemplary implementations, but is not intended to identify key features or essential features of those various exemplary implementations, nor is it intended to limit the scope of the claimed subject matter.

Conductive Particle Based Material

[0031] In one exemplary embodiment, a conductive particle based material is employed. The conductive particle based material includes at least two constituent components, namely conductive particles and a binder. However, the conductive particle based material may include additional components, such as at least one of graphite, carbon (e.g., carbon black), titanium dioxide, etc.

[0032] The conductive particles may be any conductive material, such as silver, copper, nickel, aluminum, steel, metal alloys, carbon nanotubes, any other conductive material, and any combination thereof. For example, in one exemplary embodiment, the conductive particles are silver coated copper. Alternatively, the conductive particles may be a combination of a conductive material and a non-conductive material. For example, the conductive particles may be ceramic magnetic microspheres coated with a conductive material such as any of the conductive materials described above. Furthermore, the composition of each of the conductive particles may vary from one another.

[0033] The conductive particles may be any shape from a random non-uniform shape to a geometric structure. The conductive particles may all have the same shape or the conductive particles may vary in shape from one another. For example, in one exemplary embodiment, each of the conductive particles may have a random non-uniform shape that varies from conductive particle to conductive particle.

[0034] The conductive particles may range in size from a few nanometers up to a few thousand nanometers. Alternatively, the conductive particles may range in size from about 400 nanometers to 30 micrometers. The conductive particles may be substantially similar in size or may be of various sizes included in the above identified ranges. For example, in one exemplary embodiment, the conductive particles are of various sizes in the range of about 400 nanometers to 30 micrometers. Herein, when a range of sizes of the conductive particles are employed, the distribution of the sizes may be uniform or non-uniform across the range. For example, 75% of the conductive particles may be a larger size within a given range while 25% of the conductive particles are a smaller size.

[0035] An effective amount of conductive particles are included relative to the binder so that the conductive particles are dispersed in the binder. The conductive particles may be randomly or orderly dispersed in the binder. The conductive particles may be dispersed at uniform or non-uniform densities. The conductive particles may be dispersed so that at least a majority of the conductive particles are closely adjacent to, but do not touch, one another.

[0036] The binder is used to substantially fix the conductive particles relative to each other and should be a non-conductive or semi-conductive substance. Any type of conventional or novel binder that meets these criteria may be used. The non-conductive or semi-conductive material of the binder may be chosen to function as a dielectric with a given permittivity.

[0037] The conductive particle based material may be formed as a rigid or semi-rigid structure. For example, the conductive particle based material may be a plastic sheet having the conductive particles dispersed therein. The conductive particle based material may be clear or opaque, and may include any shade of color.

[0038] In addition, the conductive particle based material may be a liquid, paint, gel, ink or paste that dries or cures. Here, the binder may include distillates, hardening agents, or solvents such as a Volatile Organic Compound (VOC). In this case, the conductive particle based material may be applied to a substrate. Also, when the conductive particle based material is a liquid, paint, gel, ink or paste that dries or cures, the binder may adhere to the substrate. The conductive particle based material may be spayed on, brushed on, rolled on, ink-jet printed, silk screened, etc. onto the substrate. The use of the conductive particle based material that is a liquid, paint, gel, ink or paste that dries or cures is advantageous in that the conductive particle based material may be thinly applied to a substrate and conform to the surface of the substrate. This allows the conductive particle based material to occupy very little space and, in effect, blend into the substrate.

[0039] The substrate may be the surface of any conductive, non-conductive or semi-conductive substance. The substrate may be rigid, semi-flexible or flexible. The substrate may be flat, irregularly shaped or geometrically shaped. The substrate may be paper, cloth, plastic, polycarbonate, acrylic, nylon, polyester, rubber, metal such as aluminum, steel and metal alloys, glass, composite materials, fiber reinforced plastics such as fiberglass, polyethylene, polypropylene, fiberglass, textiles, wood, etc.

[0040] The substrate may have a coating applied thereto. The coating may be a conductive, non-conductive or semi-conductive substance. The coating may be a paint, gel, ink, paste, tape, etc. The coating may be chosen to function as a dielectric with a given permittivity.

[0041] At least one of a protective and concealing (or decorative) coating may be applied over the conductive particle based material once it has been applied to a substrate.

[0042] An example of the conductive particle based material is described below with reference to FIG. 1.

[0043] FIG. 1 is a captured image of a conductive particle based material according to an exemplary embodiment of the present invention.

[0044] Referring to FIG. 1, the conductive particle based material includes conductive particles and a binder. The conductive particles are randomly shaped, sized and located. However, conductive particles are dispersed so that at least a majority of the conductive particles are closely adjacent to, but do not touch, one another.

[0045] Herein, without intending to be limiting, for a conductive particle based material of a given density of conductive particles, the conductive particle based material may be applied at a thickness such that the conductive particles are dispersed in the binder so that at least a majority of the conductive particles are closely adjacent to, but do not touch, one another. Herein, without intending to be limiting, it has been observed that a conductive particle based material has a resistance of about 3-17 ohms across any given two points on the surface.

[0046] Herein, without intending to be limiting, it has been observed that when the conductive particle based material is formulated such that the conductive particles are dispersed in the binder so that at least a majority of the conductive particles are closely adjacent to, but do not touch, one another, the conductive particle based material exhibits properties that enable it to at least one of efficiently propagate electromagnetic radiation, efficiently absorb electromagnetic radiation from space, and efficiently emit electromagnetic radiation into space. Moreover, it has been observed that those properties may be either supplemented or enhanced by including an effective amount of carbon, such as carbon black, in the conductive particle based material. For example, an effective amount of carbon black may be an amount that corresponds to about 1-7% of the conductive particles included in the conductive particle based material.

[0047] Without intending to be limiting, it is believed that when electromagnetic radiation is introduced into the conductive particle based material, electromagnetic radiation may pass from conductive particle to conductive particle via at least one of capacitive and inductive coupling. Here, the binder may function as a dielectric. Thus, it is believed that the conductive particle based material may act as an array of capacitors, which may be at least part of the reason why the conductive particle based material at least one of efficiently propagates electromagnetic radiation, efficiently absorbs electromagnetic radiation from space, and efficiently emits electromagnetic radiation into space.

[0048] Alternatively or additionally, and without intending to be limiting, it is believed that the properties that enable the conductive particle based material to at least one of efficiently propagate electromagnetic radiation, efficiently absorb electromagnetic radiation from space, and efficiently emit electromagnetic radiation into space, may be explained by quantum theory at the atomic level.

[0049] Herein, without intending to be limiting, it has been observed that the conductive particle based material generates electrical energy when exposed to sunlight.

[0050] Herein, without intending to be limiting, it has been observed that the resistance of the conductive particle based material continuously changes over time. Herein, without intending to be limiting, it has been observed that, when energized with a radio signal, the conductive particle based material has infinitely low resistance to that signal.

[0051] Herein, while the present disclosure is described in the context of electromagnetic radiation, without intending to be limiting, it is believed that the present invention is equally applicable to bioelectromagnetic energy. Thus, any disclosure herein that refers to electromagnetic radiation equally applies to bioelectromagnetic energy.

Conductive Particle Based Antenna

[0052] In one exemplary embodiment, the conductive particle based material is employed to implement a conductive particle based antenna. When used as a conductive particle based antenna, the conductive particle based antenna is fabricated using the conductive particle based material. Here, the conductive particle based material may be formed into a shape that conforms to the desired characteristics of the antenna. For example, the shape and size of the antenna may vary depending on the frequency and/or polarization of the electromagnetic radiation to be communicated. The conductive particle based antenna is at least one of electrically, capacitively, and inductively coupled to at least one of a receiver, a transmitter, and a transceiver at a coupling point of the conductive particle based antenna. The coupling point of the conductive particle based antenna may substantially be an end point of the conductive particle based antenna. The coupling point of the conductive particle based antenna may be coupled to a coupling point of a feed line electrically connected to the receiver, transmitter, or transceiver. When capacitively or inductively coupled, the coupling may occur through a distance that includes an air gap or that has a substance, such as glass, disposed therein.

[0053] When a conductive particle based antenna is fabricated using the conductive particle based material, the conductive particle based antenna may exhibit a broad bandwidth self-tuning characteristic by using only a small section of the conductive particle based antenna to emit the electromagnetic radiation into space.

[0054] In addition, when the conductive particle based antenna is fabricated using the conductive particle based material, there may be no or little I<2>R losses due the small practical size and the majority of the particles not contacting each other. In addition, there may be no or little Radio Frequency (RF) skin effect losses due to the small practical size. Once the signal is coupled to the conductive particle based antenna, the conductive particle based antenna provides little to no resistance to the transmission signal and it is emitted without significant loss into space. The same may happen in reverse for receiving. That is, the received signal may be absorbed and delivered with little to no loss to the coupling device and is then propagated down a feed line to a receiver.

[0055] An example of the conductive particle based antenna is described below with reference to FIG. 2.

[0056] FIG. 2 illustrates a conductive particle based antenna according to an exemplary embodiment of the present invention. The particular structure of the conductive particle based antenna 200 shown in FIG. 2 is merely an example used for explanation and is not intended to be limiting. The conductive particle based material used to fabricate the conductive particle based antenna 200 of FIG. 2 is assumed to be formulated as a liquid, paint, gel, ink, or paste that dries or cures.

[0057] Referring to FIG. 2, the conductive particle based antenna 200 includes a substrate 210, a first antenna segment 220A, a second antenna segment 220B, a first coupler 230A, a second coupler 230B, and a feed line 240.

[0058] The substrate 210 is a rigid flat sheet of a non-conductive material, such as plexiglass. However, any other surface may be chosen as substrate 210. For example, the surface of a vehicle, the wall of a building, the casing of a wireless device, glass, a tree, cloth, a rock, a plastic sheet, etc., may be chosen as the substrate. When a conductive material is chosen as the substrate 210, an insulative coating of a non-conductive or semi-conductive material may be applied to the area of the substrate 210 where the conductive particle based antenna 200 is to be applied. Examples of the insulative coating of the non-conductive or semi-conductive material include plastic tape, paper tape, paint, etc. Also, when the substrate 210 is a conductive material, the substrate may be utilized as a ground plane. In addition, a surface preparation coating may be applied to the substrate 210 that allows for better adhesion of the conductive particle based material to the substrate 210. The insulative coating may serve the same function as the surface preparation coating. Also, the surface preparation coating may be applied beneath or on top of the insulative coating. Furthermore, the surface preparation coating may be used when the insulative coating in not applied.

[0059] The first antenna segment 220A and the second antenna segment 220B are applied to the substrate 210 according to a desired design. Here, the first antenna segment 220A is functioning as an active antenna element and the second antenna segment 220B is functioning as a ground plane. When the substrate 210 is functioning as a ground plane or an earth ground is employed, the second antenna segment 220B may be omitted. Here, the first antenna segment 220A and the second antenna segment 220B are formed using a conductive particle based material formulated as a liquid, paint, gel, ink, or paste that dries or cures. The non-conductive material may be sprayed on, brushed on, rolled on, silk screened, ink jet printed, etc.

[0060] The first coupler 230A and the second coupler 230B at least one of electrically, capacitively, and inductively couple to the first antenna segment 220A and the second antenna segment 220B, respectively. In addition, the first coupler 230A and the second coupler 230B adhere to, or are otherwise in a fixed relationship with, the first antenna segment 220A and the second antenna segment 220B. The first coupler 230A and the second coupler 230B are electrically connected to respective portions of the feed line 240.

[0061] The feed line 240 is electrically connected to first coupler 230A and the second coupler 230B. Also, the feed line 240 is electrically connected to at least one of a receiver, a transmitter, and a transceiver.

[0062] An example of a structure of a conductive particle based antenna is described below with reference to FIG. 3.

[0063] FIG. 3 illustrates a structure of a conductive particle based antenna according to an exemplary embodiment of the present invention. The particular structure of the conductive particle based antenna shown in FIG. 3 is merely an example used for explanation and is not intended to be limiting. The conductive particle based material used to fabricate the conductive particle based antenna of FIG. 3 is assumed to be formulated as a liquid, paint, gel, ink, or paste that dries or cures.

[0064] Referring to FIG. 3, the conductive particle based antenna includes a substrate 310, first coating 350, conductive particle based material coating 320, and a second coating 360. One or more of the substrate 310, the first coating 350, and the second coating 360 may be omitted. In addition, one or more additional coatings may be utilized.

[0065] The substrate 310 may be any surface of any object, regardless of what material(s) the object is constructed of. For example, the surface of a vehicle, the wall of a building, the casing of a wireless device, glass, a tree, cloth, a rock, a plastic sheet, etc., may be chosen as the substrate. When the substrate 310 is a conductive material, the substrate 310 may function as a ground plane.

[0066] The first coating 350 is applied on top of the substrate 310. The first coating 350 may be at least one of an insulative coating and a surface preparation coating. As an insulative coating, the first coating 350 may be a non-conductive or semi-conductive material. Examples of the insulative coating of the non-conductive or semi-conductive material include plastic tape, paper tape, paint, etc. As a surface preparation coating, the first coating 350 may be any material that allows for better adhesion of the conductive particle based material coating 320 to the substrate 310. The same coating may serve as both the insulative coating and a surface preparation coating. Alternatively, separate insulative and a surface preparation coatings may be utilized either together or individually. The first coating 350 may be formulated as a liquid, paint, gel, ink, or paste that dries or cures. In this case, the first coating 350 may be sprayed on, brushed on, rolled on, silk screened, ink jet printed, etc. The first coating 350 may be omitted.

[0067] The conductive particle based material coating 320 is applied on top of the first coating 350, if present. Otherwise, the conductive particle based material coating 320 is applied on top of the substrate

320. Alternatively, the conductive particle based material coating 320 may be an independent structure. The conductive particle based material coating may be formulated using any formulation of the conductive particle based material described herein. For example, the conductive particle based material coating 320 may be formulated as a liquid, paint, gel, ink, or paste that dries or cures. In this case, the non-conductive material may be sprayed on, brushed on, rolled on, silk screened, ink jet printed, etc.

[0068] The second coating 360, if utilized, is applied on top of the conductive particle based material coating 320. The second coating 360 may serve to protect and/or conceal the conductive particle based material coating 320. The second coating 360 may be any material or structure that protects and/or conceals the conductive particle based material coating 320. The same coating may serve as both the protective coating and the concealment coating. Alternatively, separate protective and concealment coatings may be utilized either together or individually. In one exemplary embodiment, the second coating 360 is formulated as a liquid, paint, gel, ink, or paste that dries or cures. In this case, the second coating 360 may be sprayed on, brushed on, rolled on, silk screened, ink jet printed, etc. The second coating 360 may be omitted.

[0069] Tests were conducted to compare the conductive particle based antenna to a conventional antenna. The conductive particle based antenna was formed using the conductive particle based material whereas the conventional copper antenna was formed using solid copper strips. Both the conductive particle based antenna and the conventional copper antenna were fabricated with the same shape (i.e., the shape shown in FIG. 2) of the same size so that the effect of the particular structure, if any, is equal to both antennas. A non-conductive plexiglass substrate was used to fix both antennas. The same transmit power and frequency were used for the test. The frequency selected was in the range of about 460 MHz. Testing equipment included a Yeasu FT 7900 Dual band FM transceiver, a Telewave Model 44 Wattmeter, and a FieldFox Model N9912A Portable Network Analyzer operated in SA mode used with a Yeasu Model Rubber Duck Antenna that was located 160 feet from the test antennas. The test data for the conventional copper antenna and the conductive particle based antenna are provided below in Table 1.

[0000]

TABLE 1

	Conventional Copper Antenna	**Conductive Particle Based Antenna**
Forward Power	22 watts	41 watts
Reverse Power	12 watts	1 watt
Relative Signal Strength	-35 decibels	-26 decibels

[0070] As can be seen in Table 1, the conductive particle based antenna exhibits a significantly higher forward power (i.e., 41 watts) than the forward power of the conventional copper antenna (i.e., 22 watts). This can be explained by the conductive particle based antenna exhibiting a significantly lower reverse power (i.e., 1 watt) than the reverse power of the conventional copper antenna (i.e., 12 watts). Accordingly, the resulting relative signal strength of the conductive particle based antenna is higher (-26 decibels) than the resulting relative signal strength of the conventional copper antenna (-35 decibels).

[0071] As can be gleaned from the test, for a given antenna structure, the conductive particle based antenna is more efficient at emitting electromagnetic radiation into space than the conventional copper antenna. Therefore, the conductive particle based antenna has a higher effective gain than the conventional copper antenna. Also, since there is less reverse power, less of the electromagnetic radiation input to the conductive particle based antenna may be converted into heat. Thus, the antenna may operate at a lower temperature for a given input power and therefore may have a higher power rating.

[0072] The added gain by using the conductive particle based antenna is well suited to any application in which higher gain and/or lower transmit power for a given antenna structure is desired.

[0073] It has been observed that the transmission performance of the conductive particle based antenna varies depending on the type of amplifier used to drive the antenna. For example, the transmitter used in the Yeasu FT 7900 Dual band FM transceiver in the above test is a class C amplifier. When a linear class A amplifier is employed, the transmission performance of the conductive particle based antenna is reduced and approaches that of the conventional copper antenna. Thus, the performance of the conductive particle based antenna is greater when used with an amplifier that operates for less than the entire input cycle, such as the class C amplifier. While a class C amplifier is referred to herein for convenience in explanation, the use of any amplifier that operates for less than the entire input cycle is equally applicable.

[0074] Herein, power constrained devices typically employ a class C amplifier in order to take advantage of their efficiency so as to conserve power. Similarly, the use of the conductive particle based antenna in power constrained devices that employ a class C amplifier takes advantage of the efficiency of the conductive particle based antenna so as to further conserve power. The power conservation gained by the power constrained devices by using the conductive particle based antenna may allow for longer operational times and/or smaller power source (e.g., batteries) (and thereby smaller devices and/or a lower cost).

Conductive Particle Based Antenna Enhancer

[0075] In one exemplary embodiment, the conductive particle based material is employed to implement a conductive particle based antenna enhancer. When used as a conductive particle based antenna enhancer, the conductive particle based antenna enhancer is fabricated using the conductive particle based material. Here, the conductive particle based antenna enhancer is disposed in an adjacent offset relationship to a conventional antenna with a non-conductive or semi-conductive material disposed there between. Alternatively or additionally, an air gap between the conventional antenna and the conductive particle based antenna enhancer may be employed. Here, the conventional antenna is electrically coupled to at least one of a receiver, a transmitter, and a transceiver.

[0076] In this configuration, the conductive particle based antenna enhancer is at least one of capacitively and inductively coupled to the conventional antenna. Herein, the electromagnetic radiation that is capacitively and inductively coupled from the conventional antenna to the conductive particle based antenna enhancer is efficiently radiated into space by the conductive particle based antenna enhancer.

[0077] The conductive particle based antenna enhancer may be fabricated and positioned so as to be adjacent and offset from the conventional antenna. For example, the conductive particle based antenna enhancer may be added or built into a structure that places it in an adjacent and offset relationship to the conventional antenna.

[0078] For example, the structure may create an air gap between the conventional antenna and a surface onto which the conductive particle based material is applied. The structure may be constructed of a nonconductive material. Alternatively, the structure may be constructed of a conductive material and at least partially coated with a nonconductive material. If the structure is constructed of a conductive material, the conductive particle based material may be applied on top of the nonconductive material coating the structure. Herein, the conductive particle based material may be applied to a side of the structure closest to the conventional antenna or a side of the structure furthest from the conventional antenna. The conductive particle based material may be coated with a layer of the nonconductive material or another material. Examples of the structure include a housing of a device (e.g., a housing of a wireless device), an enclosure placed over the existing antenna, and a case placed over a housing of a device (e.g., a protective cover for a wireless device). The conductive particle based material is at least one of capacitively and inductively coupled to the conventional antenna and thereby increases the performance of the conventional antenna. Here, the thickness the nonconductive material and/or air gap directly affects the performance gain of the conductive particle based antenna enhancer and if the nonconductive thickness and/or air gap is too large, performance may decrease. The thickness of the air gap and/or nonconductive material is very small in relationship to the wavelength of the frequency the conventional antenna is designed for. In a specific example of the exemplary implementation described above, a conventional bumper case for an iPhone, which is manufactured by Apple, may have the conductive particle based material applied to a portion thereof that is adjacent to the antenna of the iPhone (the surface that is concealed when the iPhone is installed therein). Here, the conductive particle based material may have a layer of nonconductive material applied on top.

[0079] Another example of an implementation of a conductive particle based antenna enhancer is described below with reference to FIG. 4.

[0080] FIG. 4 illustrates an implementation of a conductive particle based antenna enhancer according to an exemplary embodiment of the present invention. The particular structure of the conductive particle based antenna shown in FIG. 4 is merely an example used for explanation and is not intended to be limiting. The conductive particle based material used to fabricate the conductive particle based antenna enhancer of FIG. 4 is assumed to be formulated as a liquid, paint, gel, ink, or paste that dries or cures.

[0081] Referring to FIG. 4, a wireless device 480 and a protective cover 490 are shown. The wireless device 480 includes an internal antenna 470. The protective cover 490 includes a conductive particle based antenna enhancer 420 that is disposed so as to be adjacent to the internal antenna 470 when the wireless device 480 is disposed in the protective cover 490.

[0082] While the conductive particle based antenna enhancer 420 is shown to correspond to the size of the internal antenna 470, the conductive particle based antenna enhancer 420 may be smaller or larger than the internal antenna 470. In addition, while the conductive particle based antenna enhancer 420 is shown as being disposed immediately adjacent to the internal antenna, the conductive particle based antenna enhancer 420 may be disposed at a different location on the protective cover 490.

[0083] While the conductive particle based antenna enhancer 420 is shown as being applied to an inner surface of the protective cover 490, the conductive particle based antenna enhancer 420 may be applied to an outer surface of, or may be disposed within, the protective cover 490. When the conductive particle based antenna enhancer 420 is disposed within the protective cover 490, the material used to construct the protective cover 490 may serve as the binder for the conductive particle based material. When, the conductive particle based antenna enhancer 420 is disposed at an inner or outer surface of the conductive particle based material, one or more of an insulative coating, a surface preparation coating, a protective coating, and a concealment coating may be used. In addition, the conductive particle based antenna enhancer 420 may be formed as an independent structure (with or without a substrate) that is fixed to the protective cover 490.

[0084] The conductive particle based antenna enhancer may be added to an existing conventional antenna or may be added at the time the conventional antenna is fabricated.

[0085] In one exemplary embodiment, the conductive particle based antenna enhancer is used to coat a conventional antenna that has been coated with a non-conductive material. The coating of the non-conductive material may be implemented as a liquid, paint, gel, ink, or paste that dries or cures. Herein, the non-conductive material may be sprayed on, brushed on, rolled on, silk screened, ink jet printed, etc. Alternatively, the coating of the non-conductive material may be a film or tape that is applied to the conventional antenna. Layers of other materials may be disposed between the conventional antenna and the non-conductive material and/or between the non-conductive material and the conductive particle based material. Here, depending on the configuration, the conductive particle based material may be coated with a layer of the nonconductive material and/or another material. Here, the thickness the non-conductive material may directly affect the performance gain of the conductive particle based material and if the thickness of the non-conductive material is too large, performance may decrease. The thickness of the non-conductive material is very small in relationship to the wavelength of the frequency the conventional antenna is designed for.

[0086] An example of a structure of a coated conductive particle based antenna enhancer is described below with reference to FIG. 5.

[0087] FIG. 5 illustrates a structure of a coated conductive particle based antenna enhancer according to an exemplary embodiment of the present invention. The particular structure of the conductive particle based antenna shown in FIG. 5 is merely an example used for explanation and is not intended to be limiting. The conductive particle based material used to fabricate the conductive particle based antenna of FIG. 5 is assumed to be formulated as a liquid, paint, gel, ink, or paste that dries or cures.

[0088] Referring to FIG. 5, the coated conductive particle based antenna includes a conventional antenna 570, a first coating 550, a conductive particle based material coating 520, and a second coating 560. One or more of the first coating 550, and a second coating 560 may be omitted. In addition, one or more additional coatings may be utilized.

[0089] The conventional antenna 570 may be any surface of any conventional antenna, which in this example, is assumed to be constructed of a conductive material such as metal.

[0090] The first coating 550 is applied on top of the conventional antenna 570. The first coating 550 may be at least one of an insulative coating and a surface preparation coating. As an insulative coating, the first coating 550 may be a non-conductive or semi-conductive material. Examples of the insulative coating of the non-conductive or semi-conductive material include plastic tape, paper tape, paint, etc. As a surface preparation coating, the first coating 550 may be any material that allows for better adhesion of the conductive particle based material coating 520 to the conventional antenna 570. The same coating may serve as both the insulative coating and a surface preparation coating. Alternatively, separate insulative and a surface preparation coatings may be utilized either together or individually. The first coating 550 may be formulated as a liquid, paint, gel, ink, or paste that dries or cures. In this case, the first coating 550 may be sprayed on, brushed on, rolled on, silk screened, ink jet printed, etc. The first coating 550 may be omitted.

[0091] The conductive particle based material coating 520 is applied on top of the first coating 550, if present. Otherwise, the conductive particle based material coating 320 is applied on top of the conventional antenna 570. The conductive particle based material coating may be formulated using any formulation of the conductive particle based material described herein. For example, the conductive particle based material coating 520 may be formulated as a liquid, paint, gel, ink, or paste that dries or cures. In this case, the non-conductive material may be sprayed on, brushed on, rolled on, silk screened, ink jet printed, etc.

[0092] The second coating 560, if utilized, is applied on top of the conductive particle based material coating 520. The second coating 560 may serve to protect and/or conceal the conductive particle based material coating 520. The second coating 560 may be any material or structure that protects and/or conceals the conductive particle based material coating 520. The same coating may serve as both the protective coating and the concealment coating. Alternatively, separate protective and concealment coatings may be utilized either together or individually. In one exemplary embodiment, the second coating 560 is formulated as a liquid, paint, gel, ink, or paste that dries or cures. In this case, the second coating 560 may be sprayed on, brushed on, rolled on, silk screened, ink jet printed, etc. The second coating 560 may be omitted.

[0093] The conductive particle based antenna enhancer may be fabricated and positioned so as to be adjacent and offset from all or a portion of the conventional antenna. For example, the conductive particle based antenna enhancer may be fabricated and positioned so as to be adjacent to a portion of the conventional antenna corresponding to half or a quarter of the desired wavelength.

[0094] An example of an antenna partially coated with a conductive particle based antenna enhancer is described below with reference to FIG. 6.

[0095] FIG. 6 illustrates an antenna partially coated with a conductive particle based antenna enhancer according to an exemplary embodiment of the present invention. The particular structure of the antenna partially coated with the conductive particle based antenna enhancer shown in FIG. 6 is merely an example used for explanation and is not intended to be limiting. The conductive particle based material used to fabricate the conductive particle based antenna of FIG. 6 is assumed to be formulated as a liquid, paint, gel, ink, or paste that dries or cures.

[0096] Referring to FIG. 6, an antenna 670 that is connected to a feed line 640 is shown. The antenna 670 is partially coated with a conductive particle based antenna enhancer 620. As can be seen, the conductive particle based antenna enhancer 620 coats about a quarter of the antenna 670.

[0097] Tests were conducted to compare a conventional copper antenna to the conventional copper antenna with the conductive particle based antenna enhancer. In particular, the same equipment and testing conditions as the test described above with respect to the conductive particle based antenna were performed. Here, insulative tape was applied to the entirety of the conventional copper antenna and the conductive particle based material was then applied onto the insulative tape.

[0098] The test data for the conventional copper antenna and the conventional copper antenna that has been enhanced with the conductive particle based antenna enhancer are provided below in Table 2.

[0000]

TABLE 2

Conventional Copper Antenna with Conventional Conductive Particle Based Antenna Copper Antenna Enhancer

Forward Power	22 watts	28 watts
Reverse Power	12 watts	10 watts
Relative Signal Strength	-35 decibels	-27 decibels

[0099] As can be seen in Table 2, the conventional copper antenna with the conductive particle based antenna enhancer exhibits a significantly higher forward power (i.e., 28 watts) than the forward power of the conventional copper antenna alone (i.e., 22 watts). This can be explained by the conventional copper antenna with the conductive particle based antenna enhancer exhibiting a significantly lower reverse power (i.e., 10 watts) than the reverse power of the conventional copper antenna alone (i.e., 12 watts). Accordingly, the resulting relative signal strength of the conventional copper antenna with the conductive particle based antenna enhancer is higher (-27 decibels) than the resulting relative signal strength of the conventional copper antenna (-35 decibels).

[0100] As can be gleaned from the above identified test, the conventional copper antenna with the conductive particle based antenna enhancer is more efficient at emitting electromagnetic signals into space than the conventional copper antenna alone. Therefore, the conventional copper antenna with the conductive particle based antenna enhancer has a higher effective gain than the conventional copper antenna alone. Also, since there is less reverse power, less of the electromagnetic radiation input to the conventional copper antenna with the conductive particle based antenna enhancer will be converted into heat. Thus, the conventional copper antenna with the conductive particle based antenna enhancer may operate at a lower temperature for a given input power and therefore may have a higher power rating.

[0101] Accordingly, the conductive particle based material may be used to enhance a conventional antenna.

Conductive Particle Based Transmission Line

[0102] The conductive particle based material may be used to form a conductive particle based transmission line. To implement a conductive particle based transmission line, a transmission line is formed in any of the various ways described herein for forming an object using the conductive particle based material. Herein, at least some of the properties that enable the conductive particle based material to efficiently radiate electromagnetic radiation into space allow the conductive particle based material to efficiently radiate electromagnetic radiation down the transmission line formed using the conductive particle based material. The use of the conductive particle based material as a transmission line is beneficial due to its lower resistance and heat generation.

Conductive Particle Based Electromagnetic Radiation Harvester

[0103] The conductive particle based material may be used as an electromagnetic radiation harvester. The high efficiencies of the conductive particle based material in at least one of propagating and absorbing electromagnetic radiation make it ideally suited for use in collecting electromagnetic radiation. While such collected electromagnetic radiation may be electromagnetic radiation that was transmitted with the intention of being harvested by the electromagnetic radiation harvester, the collected electromagnetic radiation may be background electromagnetic radiation. Herein, the electromagnetic radiation harvester may be coupled to a receiver that collects the energy absorbed by the electromagnetic radiation harvester. The electromagnetic radiation harvester is formed in any of the various ways described herein for forming an object using the conductive particle based material.

Conductive Particle Based Conformable Antenna

[0104] The conductive particle based material may be used to construct a conductive particle based conformable antenna. The benefit of the conductive particle based conformable antenna may be easily appreciated when considered in the context of an exemplary use case, which is described below.

[0105] According to the exemplary use case, the conductive particle based conformable antenna may use used in a military setting. The Special Operations community has a major logistical and safety issue when it comes to communications in the theater. The US Department of Defense (DoD) has rapidly expanded its communications capabilities within the radio spectrum. In the past, two way radios in a variety of form factors where used for conventional Push-To-Talk (PTT) communications. The use of these systems has now evolved into a true "Digital Battlefield" consisting of a multitude of communications platforms. Vast arrays of data networks came into reality. The scope of radios used today varies widely from conventional voice to Satellite, mesh networks, to Unmanned Aerial Vehicles (UAVs) and unattended ground sensors.

[0106] The reason this wide variety of systems is mentioned is to give an understanding of why the conductive particle based conformable antenna may be beneficial to the mission of soldiers. Every RF device utilized by the military operates on a wide range of frequencies and a different type of transmission (Amplitude Modulation (AM), Frequency Modulation (FM), Satcom, Single Side band, etc.).

[0107] However, conventional antenna systems are designed and tuned for a limited range of frequencies and are generally designed to work with only one of the hundreds of types of radio devices on the market. The other major downsides to these conventional antenna systems are the logistics of getting them into battle. They are heavy, bulky, expensive, and difficult to transport. Accordingly, there is a need to address the shortcomings of the conventional antenna systems.

[0108] The conductive particle based conformable antenna addresses the shortcomings of the conventional antenna systems by being operable with any and all of the radios currently deployed and being developed. As opposed to being an antenna of fixed form, the conductive particle based conformable antenna may instead be constructed on an as needed basis.

[0109] For example, the conductive particle based conformable antenna may be constructed on site using the conductive particle based material. In this case, the conductive particle based material is a liquid, paint, gel, ink or paste that dries or cures. Herein, the conductive particle based conformable antenna may be applied to a substrate. In particular, the conductive particle based material may be sprayed on, brushed on, rolled on, silk screened, ink jet printed, etc.

[0110] The conductive particle based conformable antenna may be designed based on typical antenna design, theory, and formulas. The antenna design may be generated in advance or at the time the antenna is needed based on desired characteristics.

[0111] The conductive particle based material is applied to the substrate to form the conductive particle based conformable antenna based on the desired antenna design.

[0112] The substrate may be any surface of any material, such as acrylic, ABS, structural foams, solvent sensitive materials such as polycarbonate and polystyrene, and non-porous surfaces including primed wallboard, wood and clean metals, etc.

[0113] When the substrate is a conducting material, a non-conductive or semi-conductive coating may first be applied to the substrate. In this case, the conducting material may serve as a ground plane. When the substrate is a non-conducting material, a ground plane can be accomplished by using the earth's natural ground. Alternatively, the ground plane can be accomplished by fabricating an independent ground plane.

[0114] Once the antenna is fabricated, a feed line is coupled to the conductive particle based conformable antenna and an RF communications device. The conductive particle based conformable antenna is at least one of electrically, capacitively, and inductively coupled to a coupling point of the feed line. The conductive particle based conformable antenna may be coupled to the coupling point of the feed line at an end point of the conductive particle based conformable antenna. When capacitively or inductively coupled, the coupling may occur through a distance that includes an air gap or a substance, such as glass.

[0115] To fabricate the conductive particle based conformable antenna, a template of the desired antenna design may be used. The template may be a sheet formed of any rigid or semi-rigid material in which the desired design of the antenna is cut out.

[0116] An example of a template used to fabricate a conductive particle based conformable antenna is described below with reference to FIG. 7.

[0117] FIG. 7 illustrates a template used to fabricate a conductive particle based conformable antenna according to an exemplary embodiment of the present invention.

[0118] Referring to FIG. 7, a template 700 is shown. The template 700 may be any material that may be used to form a template or stencil. For example, the template 700 may be a sheet formed of a rigid or semi-rigid material. The cut out of the template 700 may be at least one of a positive and a negative of a desired design of an antenna. The template 700 may be an image displayed on a surface showing where conductive particle based material should or should not be applied. The template 700 may be an image displayed on a display or in a guide book that shows a desired design of an antenna. Herein, the template 700 shown in FIG. 7 corresponds to the antenna design shown in FIG. 2.

[0119] Examples of various cutout designs for the template 700 are found in U.S. Design patent application Ser. No. 29/390,425, filed on Apr. 25, 2011, and entitled "ANTENNA"; U.S. Design patent application Ser. No. 29/390,427, filed on Apr. 25, 2011, and entitled "ANTENNA"; U.S. Design patent application Ser. No. 29/390,432, filed on Apr. 25, 2011, and entitled "ANTENNA"; U.S. Design patent application Ser. No. 29/390,435, filed on Apr. 25, 2011, and entitled "ANTENNA"; U.S. Design patent application Ser. No. 29/390,436, filed on Apr. 25, 2011, and entitled "ANTENNA"; U.S. Design patent application Ser. No. 29/390,438, filed on Apr. 25, 2011, and entitled "ANTENNA"; and U.S. Design patent application Ser. No. 29/390,442, filed on Apr. 25, 2011, and entitled "ANTENNA", the entire disclosure of each of which is hereby incorporated by reference.

[0120] An exemplary method for fabricating a conductive particle based conformable antenna using a template is described below with reference to FIG. 8.

[0121] FIG. 8 illustrates a method for fabricating a conductive particle based conformable antenna using a template according to an exemplary embodiment of the present invention. Herein, the conductive particle based material used to fabricate the conductive particle based conformable antenna is assumed to be formulated as a liquid, paint, gel, ink, or paste that dries or cures.

[0122] Referring to FIG. 8, a template and substrate is chosen in step 800. In step 810, the chosen template may be fixed against the chosen substrate. In step 820, the conductive particle based material may then be applied on the template such that the conductive particle based material passes through at least one cut out portion of the template so as to be applied to the corresponding portion of the substrate. The conductive particle based material may be applied until its particle density reaches a certain threshold. This may be determined by measuring the resistance of the material across the length of the antenna (or antenna segment). Here, the threshold may correspond to a predefined resistance or range of resistances (e.g., 11-15 ohms).

[0123] The template may then be removed leaving the conductive particle based material to dry or cure on the chosen substrate according to the desired design. In step 830, one or more coupling points of a feed line may be affixed to the conductive particle based conformable antenna. Herein, step 830 may be omitted. In addition, additional steps may be included, such as applying at least one of an insulative coating, a surface preparation coating, a protective coating, and a concealment coating. Any or all of this fabrication technique may be automated, as will be described below.

[0124] While a conductive particle based conformable antenna is described herein, any disclosure related to a conductive particle based conformable antenna is equally applicable to a conductive particle based conformable antenna enhancer.

Fabrication Techniques for Conductive Particle Based Conformable Antenna

[0125] In one exemplary embodiment, techniques for constructing a conductive particle based conformable antenna are described. Herein, a computerized device is used to generate a template that is used to construct a conductive particle based conformable antenna.

[0126] The computerized device may be any of a desktop computer, a laptop computer, a netbook, a tablet computer, a Personal Data Assistant (PDA), a Smartphone, a portable media device, a specialized mobile device, etc. The computerized device may include one or more of a display, an input unit, a control unit, a printer, memory, a communications unit, and a projection unit.

[0127] The conductive particle based conformable antenna that is constructed using the template may be formed using the conductive particle based material that is sprayable, rollable or brushable. The conductive particle based material may be applied directly onto any substrate. The conductive particle based conformable antenna, once fabricated onto a surface, may be painted over with a paint in order to conceal the antenna, provide protection to the antenna, or provide the antenna with desired aesthetics.

[0128] According to an exemplary embodiment of the present invention, to create and install an antenna, the computerized device may be used to generate the template. The computerized device may include a graphical user interface that queries a user regarding certain characteristics/criteria or otherwise allows a user to enter certain characteristics/criteria. Based on the input characteristics/criteria, the computerized device generates the template. Herein, the user may input less than all of the characteristics/criteria. In this case, the characteristics/criteria not input by the user may be obtained via a formula, or a local or remote database. In addition, assumed values for the characteristics/criteria not input by the user may be used.

[0129] Examples of the characteristics/criteria include one or more of a substrate on which the antenna will be disposed, frequency of operation, aperture or antenna pattern, whether a space saving design is desired, velocity factor, resonant frequency, Q factor, impedance, gain, polarization, efficiency, bandwidth, heat characteristics, type of amplifier, environment, etc. Further, one or more of the characteristics/criteria may include a number of preset options for a given characteristic/criteria. For example, the options for the substrate on which the antenna will be disposed may include one or more of wood, metal, glass, plastic, etc. For another example, the options for the desired antenna pattern include one or more of an omni-directional antenna pattern, a directional antenna pattern, a circular antenna pattern, a phased array antenna pattern, etc.

[0130] The computerized device may guide a user in inputting at least one of the one or more the characteristics/criteria and may request additional information from the user.

[0131] Based on the input one or more characteristics/criteria, the computerized device determines an antenna pattern using a pattern determination algorithm. The antenna pattern may be a preset antenna pattern or an antenna pattern formed based on an algorithm and the input one or more characteristics/criteria. In addition, the computerized device may determine one or more of a scaling factor of the antenna pattern, dimensions of the antenna pattern or elements of the antenna pattern, grain direction, application notes, etc. Alternatively, or additionally, the characteristics/criteria may not be preset.

[0132] The computerized device may determine more than one antenna pattern and may allow a user to select a desired antenna pattern from among the determined more than one antenna pattern.

[0133] Once the antenna pattern is determined, as well as one or more of the scaling factor of the antenna pattern, dimensions of the antenna pattern or elements of the antenna pattern, grain direction, application notes, etc., a resulting template may be at least one of displayed on the display of the computerized device, projected onto a surface using the projection unit of the computerized device, and printed using one of an external and an integrated printed. When a projection unit is employed, the computerized device may further include a device that adjusts the scale of the projected template based on at least the distance between the projection unit and the surface on which the antenna is to be constructed. Further, when a projection unit is employed, the computerized device may further include a device that adjusts the location of the projected template so that the projected template remains on the same location of the surface regardless of the movement of the computerized device. The template may then be used to construct the antenna.

[0134] Also, the template may correspond to digital data that is stored in a storage device or communicated to another device that applies the antenna material based on the digital data.

[0135] In one exemplary embodiment, the computerized device communicates the input characteristics/criteria to a remote computerized device which determines one or more of the antenna pattern, the scaling factor of the antenna pattern, dimensions of the antenna pattern or elements of the antenna pattern, grain direction, application notes, etc., which is then communicated to the computerized device.

[0136] In one exemplary embodiment, the antenna patterns may be stored remotely from the computerized device and communicated to the computerized device before or after the antenna pattern is determined. The antenna patterns may be communicated to the computerized device in response to a request by the computerized device or another entity.

[0137] An exemplary method for fabricating a conductive particle based conformable antenna using a computerized device is described below with reference to FIG. 9.

[0138] FIG. 9 illustrates a method for fabricating a conductive particle based conformable antenna using a computerized device according to an exemplary embodiment of the present invention.

[0139] Referring to FIG. 9, in step 900, the characteristics/criteria are obtained by the computerized device as described above. In step 910, an antenna pattern is selected by the computerized device based on the obtained characteristics/criteria, as described above. In step 920, a template is generated as described above.

[0140] An example of the computerized device described above is described below with reference to FIG. 10.

[0141] FIG. 10 illustrates a structure of computerized device used for fabricating a conductive particle based conformable antenna according to an exemplary embodiment of the present invention.

[0142] Referring to FIG. 10, the computerized device includes a controller 1010, a display unit 1020, a memory unit 1030, an input unit 1040, a communications unit 1050, a template generator 1060, and an antenna generator 1070. One or more of the components of the computerized device shown in FIG. 10 may be omitted. Also, the functions of one or more of the components of the computerized device shown in FIG. 10 may be performed by a combined component. In addition, additional components may be included with the computerized device.

[0143] The controller 1010 controls the overall operations of the computerized device. More specifically, the controller 1010 controls and/or communicates with the display unit 1020, the memory unit 1030, the input unit 1040, the communications unit 1050, the template generator 1060, and the antenna generator 1070. The controller 1010 executes code to have performed or perform any of the functions/operations/algorithms/roles explicitly or implicitly described herein as being performed by a computerized device. The term "code" may be used herein to represent one or more of executable instructions, operand data, configuration parameters, and other information stored in the memory unit 1030.

[0144] The display unit 1020 is used to display information to a user. The display unit 1020 may be any type of display unit. The display unit 1020 may be integrated with or separate from the computerized device. The display unit 1020 may be integrated with the input unit 1040 to form a touch screen display. The display unit 1020 performs any of the functions/operations/roles explicitly or implicitly described herein as being performed by a display.

[0145] The memory unit 1030 may store code that is processed by the controller 1010 to execute any of the functions/operations/algorithms/roles explicitly or implicitly described herein as being performed by a computerized device. In addition, one or more of other executable instructions, operand data, configuration parameters, and other information may be stored in the memory unit 1030. Depending on the exact configuration of the computerized device, the memory unit 1030 may be volatile memory (such as Random Access Memory (RAM)), non-volatile memory (e.g., Read Only Memory (ROM), flash memory, etc.) or some combination thereof.

[0146] The input unit 1040 is used to enable a user to input information. The input unit 1020 may be any type or combination of input unit, such as a touch screen, keypad, mouse, voice recognition, etc.

[0147] The communications unit 1050 transmits and receives data between one or more entities. The communications unit 1050 may include any number of transceivers, receivers, and transmitters of any number of types, such as wired, wireless, etc.

[0148] The template generator 1060 may perform any of the functions/operations/algorithms/roles explicitly or implicitly described herein as being performed when generating a template. For example, the template generator 1060 may be a printer, a cutter, a projector, a display, etc.

[0149] The antenna generator 1070 may perform any of the functions/operations/algorithms/roles explicitly or implicitly described herein as being performed when generating an antenna. For example, the template generator 1060 may be a sprayer that sprays the conductive particle based material onto a substrate.

[0150] Herein, the functionality described above of the computerized device may result from an application installed on and being executed by the computerized device.

[0151] At this point it should be noted that the present exemplary embodiment as described above typically involve the processing of input data and the generation of output data to some extent. This input data processing and output data generation may be implemented in hardware, or software in combination with hardware. For example, specific electronic components may be employed in a mobile device or similar or related circuitry for implementing the functions associated with the exemplary embodiments of the present invention as described above. Alternatively, one or more processors operating in accordance with stored instructions (i.e., code) may implement the functions associated with the exemplary embodiments of the present invention as described above. If such is the case, it is within the scope of the present disclosure that such instructions may be stored on one or more non-transitory processor readable mediums. Examples of the non-transitory processor readable mediums include ROM, RAM, Compact Disc (CD)-ROMs, magnetic tapes, floppy disks, and optical data storage devices. The non-transitory processor readable mediums can also be distributed over network coupled computer systems so that the instructions are stored and executed in a distributed fashion. Also, functional computer programs, instructions, and instruction segments for accomplishing the present invention can be easily construed by programmers skilled in the art to which the present invention pertains.

[0152] While the invention has been shown and described with reference to certain exemplary embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the spirit and scope of the invention as defined by the appended claims and their equivalents.

ChamTech Patents

Antenna
USD652029

Antenna
USD652410

Antenna
USD652028

Antenna
USD652027

Antenna
USD652409

NEAR FIELD COMMUNICATIONS SYSTEM HAVING ENHANCED SECURITY
WO2009021220

SYSTEM AND METHOD FOR NEAR FIELD COMMUNICATIONS HAVING LOCAL SECURITY
WO2009042977

Nano Copper

NANO-COPPER PATENTS

Surface plasmon resonance rectenna and preparation method therefor

CN102544182

The invention provides a surface plasmon resonance rectenna and a preparation method therefor. The surface plasma resonance rectenna adopts a three-layer structure, wherein the lower layer is made from metal Ti; a $TiO2$ nanotube array layer is generated on one surface of the metal Ti in an oxidation manner; a Cu nano particle metal layer is arranged on the surface of the $TiO2$ nanotube array layer through photodeposition; and the micro-surface appearance of the Cu nano particle metal layer is nano particles. In the invention, as photodeposition replaces ultrahigh vacuum electron beam evaporation technology, and inexpensive metal copper Cu replaces precious metal Au to prepare the rectenna adopting the Ti/TiO2NT/Cu structure, the difficult problem that the conventional metal layer can not be used in large-scale industrial production due to high cost in deposition technology, equipment investment and precious metal, and green low-cost development of solar energy technology is facilitated.

Method for preparing copper nanowires and copper nano pointed cones

CN102776469

The invention discloses a method for preparing copper nanowires and copper nano pointed cones. The method comprises the steps of heating a copper substrate in oxygenated atmosphere to form a copper oxide thin film or a copper oxide nanowire thin film, placing the films into a vacuum chamber, performing bombarding by using a argon ion source, and controlling the energy and the time of the ion bombarding to obtain nanowires or copper nano pointed cone arrays. According to the method, no catalyst is used, copper nanowires and copper nano pointed cones, which have different density and sizes are prepared on the substrate conveniently, and the prepared copper nanowires and copper nano pointed arrays can be applied to photoelectric devices such as a display device, a lithium battery, a super capacitor and the like.

Method for preparing ultrafine copper powder for electronic paste

CN102764898

The invention belongs to the field of electronic paste and particularly relates to a method for preparing a ultrafine copper powder for the electronic paste, comprising the step of adding a dispersing agent and a reducing agent into a copper salt aqueous solution to reduce the elemental copper. The method is characterized in that nano-copper particles are added into the copper salt aqueous solution before the dispersing agent and the reducing agent are added, wherein the nano-copper particles have the size of 5-20nm, and the addition amount of the nano-copper particles is 0.3-0.6 per mill of the mass of copper salt in the copper salt aqueous solution. The method is simple in process and flexible to operate. The raw materials are nontoxic, environmentally friendly and suitable for industrial production. The ultrafine copper powder prepared by the adopting the method has uniform and controllable particle size, a narrow distribution range and high purity.

Preparation method of nano porous copper capable of being patterned

CN102766893

The invention discloses a preparation method of nano porous copper capable of being patterned, comprising the following steps: (1), sputtering a Cr-Cu seed layer on a sheet glass, spin coating positive photo-resist, baking the photo-resist, exposing, and developing in turn, so as to pattern the nano porous copper photo-resist; (2), depositing a copper stress buffer layer and a Cu-Zn alloy layer by the electro-deposition technique, to obtain a patterned precursor alloy thin film; and (3), performing de-alloying treatment on the patterned Cu-Zn patterned precursor alloy thin film in an acidic solution, removing the positive photo-resist, finally implementing the patterned nano porous copper so as to obtain a nano porous copper array. The preparation method and the micro-processing technique are compatible; specifically, the nano porous copper arrays with various patterns are obtained by the photo resist patterning technique, the Cu-Zn alloy co-deposition technique, and the de-alloying technique; the preparation method has the advantages of simple process, low cost, easy control, and good compatibility to the micro-processing technique.

Method for preparing nano copper lubricating material by wire electrical explosion method

CN102744414

The invention relates to a method for preparing a nano copper lubricating material by a wire electrical explosion method. On the aspect of preparing the nano lubricating material, the method effectively solves the agglomeration problem in the process of preparing nano metal powder by the existing electrical explosion method and fundamentally eliminates the problems of derivatives and impurities in the process of preparing the nano lubricating material by a chemical method. The method comprises the specific steps of: by a device comprising parts such as a high voltage discharge system, a wire throwing system, a liquid circulating system, a vacuum system, an explosion chamber and the like, after vacuumizing to the pressure of minus 0.1MPa and filling argon to the pressure of 0.1 to 0.2MPa, implementing effective crushing on a copper wire section with the specification of phi 0.2*(35 to 60)mm under a voltage of 7 to 13KV and simultaneously infiltrating the obtained product into lubricating oil; and in the electrical exploding process, directly generating a modified thin film on the surface of nano copper powder so as to form nano lubricating oil.

NANOMETER-SIZED COPPER-BASED CATALYST, PRODUCTION METHOD THEREOF, AND ALCOHOL PRODUCTION METHOD USING THE SAME THROUGH DIRECT HYDROGENATION OF CARBOXYLIC ACID
EP2561928

Also published as: US2013030224 (A1) WO2011132957 (A2) WO2011132957 (A9) WO2011132957 (A3) CN102946994 (A)

Disclosed is a nano-sized Cu based catalyst and a method of preparing the same including dissolving, in an aqueous solution, a first component comprising a Cu precursor, a second component precursor comprising one or more selected from the group consisting of a transition metal, an alkaline earth metal and a Group IIIb metal, and a third component precursor comprising one or more selected from the group consisting of alumina, silica, silica-alumina, magnesia, titania, zirconia and carbon and then performing stirring; precipitating the stirred mixture solution using Na2CO3 and NaOH to form a catalyst precursor precipitate; and washing and filtering the formed catalyst precursor precipitate. Also a method of preparing alcohol is provided, including reacting hydrogen with carboxylic acid including a single acid or an acid mixture of two or more acids derived from a microorganism fermented solution, using the nano-sized Cu based catalyst.

Copper nano powder and preparation method of copper nano powder
CN102601381

The invention discloses copper nano powder and a preparation method of the copper nano powder. The powder is prepared by the following methods: preparing copper salt solution from copper salt, adding a drag reduction agent and an inhibitor into the copper salt solution to prepare copper salt mixing solution and pre-heating the copper salt mixing solution; preparing reducing agent solution, adjusting the pH value of the reducing agent solution, and pre-heating the reducing agent solution; uniformly mixing and agitating the copper salt mixing solution and the reducing agent solution, heating the mixture to the reaction temperature to enable the mixture to react for a certain time so as to obtain nano copper suspension solution; aging the nano copper suspension solution; performing vacuum filtration on the suspension solution obtained from the above steps to obtain the copper nano powder in which the impurity is not removed; processing the copper nano powder to obtain the copper nano powder cladded by organic solvent; performing vacuum filtration on the copper nano powder to obtain wet copper nano powder, and performing low-temperature vacuum drying on the copper nano powder obtained from the above steps so as to obtain the copper nano powder product. The powder is small in particle size and narrow in particle size distribution. The method is controllable in reaction process, simple in technology and suitable for industrial production.

Water-soluble nano-copper and preparation method thereof
CN102554217

The invention belongs to the technical field of nanometer materials, and particularly relates to water-soluble nano-copper and a preparation method of the water-soluble nano-copper, wherein the water-soluble nano-copper is nano-copper clusters which are surface-modified by stable organic single molecules formed in a way that organic compound surface modifier containing sulfydryl is bonded on the surface of copper nanoparticles. The invention can obtain copper nanoparticles which can be effectively dispersed in water phases and can exist stably, and is simple in preparation process and preparation devices, low in raw material cost, low in the production cost, high in yield, and suitable for large-scale industrial production, and the raw materials are easily accessible.

New method of preparing powder nano material
CN1436626

The new nano powder material preparing process of the present invention is suitable for preparing carbide, nitride and particle metal powder. The preparing process includes pumping the reactor into vacuum, filling protecting gas or reaction gas, setting graphite electrode inside the reactor, applying voltage between the graphite electrode and the metal inside the crucible inside the reactor to produce high temperature carbon arc and metal vapor. When protecting gas is filled, nano nickel powder, copper powder, aluminum powder and other metal powder coated by graphite atoms with less agglomeration may be prepared. When reaction gas is filled, nano carbide or nitride powder may be prepared. Compared with available technology, the produced nano powder has less agglomeration, the production process has low cost and great arc power and is suitable for large scale production and is widely applicable.

METHOD FOR PREPARING NANO-PARTICLES

EP1550633

Also published as: EP1550633 (A4) EP1550633 (B1) US2004197884 (A1) US7204999 (B2) JP3703479 (B2) more

The method of the production of a nanoparticle of the present invention includes a step of forming a nanoparticle including a compound of a metal ion in a cavity part of a protein, in a solution containing the protein having the cavity part therein, the metal ion, and a carbonate ion and/or a hydrogen carbonate ion. Examples of the aforementioned compound include e.g., a hydroxide. The aforementioned metal ion is preferably any one of a nickel ion (Ni<2+>), a chromium ion (Cr<2+>) or a copper ion (Cu<2+>). According to the aforementioned method, nanoparticles having a uniform particle diameter can be produced.

Method for manufacturing nano-copper

CN1557589

The present invention is nano copper preparing process and belongs to the field of nano material preparing technology in radiochemistry. The technological process includes adopting copper sulfate as copper source, isopropanol as oxidant clearer in water solution, hydrophilic suitable for PVA for controlling the crystal kernel growth speed and grain size and electronic accelerator to produce electron beam for irradiation treatment; washing irradiated solution, centrifugal separation and stoving to obtain claret nano copper powder. The present invention has simple technological process, short production period, no pollution and high safety.

Method for preparing nano copper particle

CN1709617

Also published as: US2006053972 // US7422620

The invention discloses a method for preparation of nm copper grains with very perfect dispersitivity. It comprises copper hydrosol prepared by reduction method, using organic liquor which contains specifical extraction solution to extract copper colloidal granules, and altering polarity of the organic phase to separate nm copper granules. The preparation has very many characteristics, for example, cheap and handy raw material, simplicity and convenience, low cost, high productivity and so on. It fits large-scale industrial production and the diameter of manufactured nm copper granules is between one and ten nm. Also it stabilizes in the air and is able to scatter in the industrial non-polarity organic liquor and kinds of lube. In conclusion, it has extensive industrial purposes.

POWDERS OF NANO CRYSTALLINE COPPER METAL AND NANO CRYSTALLINE COPPER ALLOY...

WO2005092541

A nano crystalline copper metal powder, which comprises aggregates of nano crystal grains of copper metal or a copper alloy, wherein the nano crystal grain has a size of 2 to 1000 nm; a bulk material of a nano crystal copper or copper alloy exhibiting high hardness, high strength, high electric conductivity and high toughness, which comprises a great number of above nano crystal grains being firmly bound with one another; and a method for producing the above bulk material of a nano crystal copper or copper alloy exhibiting high hardness, high strength, high electric conductivity and high toughness, which comprises subjecting the above copper powder or copper alloy powder to a solidification forming such as a vacuum hot solidification forming or an explosive forming, for example, a spark plasma sintering at a temperature of 250 to 700 DEG C, hot pressing, sheath rolling, hot forging, extruding or hot isotropic pressure forming (HIP).

A method and apparatus for the production of copper nanofluids by using chemical method
TWI262111
A method for the production of copper nanofluids uses chemical reduction method. For synthesis of nanofluids containing Cu nanoparticles, copper precursor and reactants (e.g. reducing agent) are mixed uniformly. By controlling the reaction rate and environment, copper particle in nano scale is obtained through nucleation and growth process. The copper nanoparticle can increase the thermal conductivity of nanofluids with its higher thermal conductivity compared with that of fluids.

Preparation of nano copper fluid
CN101264525
The invention relates to a preparation process, belonging to the technical field of the preparation processes for inorganic materials, which comprises the following steps: metallic copper salt and polyethylene pyrrolidone are dissolved into ethylene glycol to form settled solution, then reducing agent is added into the solution and mixed evenly; the mixed liquor is put into a microwave stove for heating and reaction, after cooling, Cu/ ethylene glycol nanometer fluid is obtained. Therefore, the preparation process has the advantages of simple process, low cost, high yield, and facility for industrial production; moreover, the product acquired by use of the preparation process has the advantages of good dispersibility, long-term stability with a plurality of better performances than prior products, and wide popularization and application prospects.

Production of nano-catalyst
CN101028600
A process for preparing the nano-catalyst used to selectively adsorb and convert harmful substances includes such steps as adding copper sulfate and manganese sulfate to deionized water, heating to 80-90 deg.C, stirring, adding strong alkali solution, stirring, filtering, washing the filtered cake until it becomes neutral, drying, calcining, cooling and sieving.

ELECTRODE, AND APPARATUS AND METHOD FOR MANUFACTURING METALLIC FINE PARTICLE
JP2007270184
PROBLEM TO BE SOLVED: To provide a method for manufacturing copper nano particles by which granular nano sized metal fine particles are efficiently manufactured without growing a reduced metal like dendrite (tree branch) and without enlarging the particle size in the mass production of the particles. ; SOLUTION: In the method of manufacturing the copper nano particles, the copper nano particles are manufactured by applying current between an anode comprising copper and a cathode comprising aggregate of many acicular projections of platinum and electrically insulated from the anode. The platinum acicular projections are, for example, cylindrical platinum having <=1 [mu]m diameter or rectangular platinum having <=1 [mu]m one side length and electrolytically deposited on a conductive electrode.

Method for preparing nano copper powder and copper slurry
CN101077529
The present invention is process of preparing nanometer copper powder and copper slurry with high antioxidant performance. The process includes solvent replacement, the first reduction, the second reduction, separation, drying and other steps. The present invention has the features of simple preparation process at normal temperature and normal pressure, low production cost, the effective protection of produced nanometer copper particle in organic phase, small copper powder size, environment friendly preparation process, etc. The prepared nanometer copper powder and copper slurry may be applied in producing large scale PCB, conducting ink, multilayer ceramic capacitor, etc.

Method for preparing nano copper powder

CN101372037

The invention discloses a method for preparing copper nanoparticle. The preparation method is as follows: cupric salt is taken as raw materials; hydroborate serves as reducer; reducer aqueous solution is poured into cupric salt aqueous solution which complexes with ammonia and is stirred quickly, when the cupric salt aqueous solution becomes colorless, the copper nanoparticle is prepared. When the stirring is stopped, the copper nanoparticle sinks quickly, the supernatant fluid is dumped. The copper nanoparticle is washed with water repeatedly, stands and is clarified, and the supernatant fluid is dumped. Finally being filtered, the copper nanoparticle with good dispersibility is prepared by being dried in the flow argon atmosphere. The preparation method has simple production technology, is easy to operate. The copper nanoparticle can be produced on a large scale; and an appropriate covering layer can be covered on the surface of the copper nanoparticle.

COPPER MICROPARTICLE, METHOD FOR PRODUCTION OF COPPER MICROPARTICLE

WO2008041780

Disclosed is a method for producing a copper microparticle, which comprises: a first reduction step for mixing an unsaturated fatty acid solution containing a copper ion with an aldose (a reducing monosaccharide) solution to form an emulsion; a second reduction step for adding an aqueous ascorbic acid solution to the emulsion; and a copper microparticle separation step. The method can produce a copper microparticle which has an average particle diameter of 100 nm or smaller and whose surface is partly or entirely modified with a carboxyl terminal group derived from the unsaturated fatty acid. The copper microparticle produced by the methodshows high dispersibility in spite of having a nano-order size.

METHOD FOR MANUFACTURING METAL NANO PARTICLE SOLUTION

US2009176875

A method for manufacturing a stabilized metal nano particle solution is disclosed. This method manufactures a metal nano particle solution so as to make a metal substance such as silver, gold, copper, zinc or cobalt into ultra-capsular nano particles. That is to say, this new method is simple and suitable for mass production without requiring any separate reducer putting process at a room temperature while a transition metal nano particle solution is produced. In this method, an alcohol solution including a metal salt solution and a soluble polymer is mixed at a room temperature to make a nano metal particle solution with a particle size of 100 nm or less.

Method for preparing metallic simple substance nano-crystal material

CN101279374

A method for preparing metal single substance nanocrystalline materials is applicable to the preparation of the metal nanocrystalline materials of copper, silver, lead, palladium, tin, antimony and so on. A melted composite alkali metal hydroxide solvent is used for carrying out the chemical reaction synthesis under the atmospheric pressure and the temperature of 100 to 300 DEG C, the used raw materials are soluble inorganic metal salts and zinc powder or iron powder, the cost during the synthesis process is low, various parameters in the reaction process are easy to be monitored and controlled, the environmental pollution is less, the evenness of a reaction system is good, the process is simple and the production is easy to be enlarged; ; furthermore, the obtained metal crystal has good crystallization, clean surface and even size, which is applicable to the research of intrinsic properties and the maximization of the functions of the nanocrystalline materials. The metal single substance nanocrystalline materials of the metal nanocrystalline materials are characterized by metal properties, electric and electronic properties, magnetic properties, chemical properties, thermal properties and luminous properties etc., which can be widely used in superconductive, chemical, medical, optical, electronic, electric appliance and other industries.

Surface finish nano copper/copper alloy particles and preparation thereof

CN101259531

The invention relates to a nano-copper/copper alloy particle for surface modification, which is prepared by the following method that: water solution of copper salt or copper salt and alloy component salt with the concentration of 0.001 to 1 mol/L is prepared, then the water solution is fully mixed with the mixed solution of a reductant, a surface modifier and a weak polar or non-polar organic solvent, the still placement is carried out for 30 to 180min after the reaction for 30 to 180min under the alkaline environment with pH of 8 to 13, the organic phase is obtained by separation, and the nano-copper/copper alloy particle for surface modification is obtained after the concentration.; The copper/copper alloy nanoparticle for surface modification which is provided by the invention has controllable particle size and distribution, difficult clustering, good antioxidant property, monodisperse property and good dispersion stability in the organic solvent. The preparation process and a device of the nano-copper/copper alloy particle for surface modification are simple, the raw materials are cheap and easy to obtain, the cost is low, and the yield is high, thus being applicable to the large-scale industrial production.

Method for preparing metallic copper nano particle
CN101337277

The invention relates to a method for preparing metallic copper nanoparticles, and belongs to the technical field of preparing nanomaterial by liquid-phase chemical reduction. The method adopts sodium hydrosulfite as a reducer to reduce cupric hydroxide under alkaline conditions, and adopts sodium dodecyl sulfate as a dispersant to generate copper nanoparticles with the grain diameter of 30 to 90nm. The method has the characteristics of simple technological process, low production cost and suitability for industrialized production.

Preparation method of whole continuous nano-porous copper
CN101596598

The invention relates to a preparation method of whole continuous nano-porous copper, comprising the following steps: (1) heating the pure metallic aluminum and copper to the molten state, stirring and mixing the mixture to form Cu-Al alloy liquid; (2) fast blowing the alloy liquid out with inert gas to solidify the molten liquid metal on the copper rollers which rotate in high speed and prepare alloy strips, or solidify the molten liquid metal in copper molds to prepare alloy plates or alloy bars; and (3) performing dealloying process to the obtained alloy, cleaning the alloy in the distilled water to be neutral and airing the alloy to obtain nano-porous copper. The beneficial effects of the invention are as follows: (1) the produced nano-porous copper is whole continuous and whole continuous nanosize bulk materials can also be prepared; (2) low concentration of corrosion solution is used to prepare nano-porous copper in the method of the invention, and the operation process is simple and the method is suitable for large-scale industrialized production; and (3) according to the components of the mother alloy and the types of corrosion solution, the structure and size of nanoporous copper can be regulated.

Preparation method of nano-copper
CN101607317

The invention belongs to the nanometer material technical field, in particular to a preparation method of nano-copper, comprising the following steps: dissolving copper salt and organic protective agents in a solvent, heating the solution to 30-100 DEG C, simultaneously adding reducing agent in the reaction system to react while stirring for 20-30min, then cooling gradually; standing the cooled solution, and then centrifuging and finally washing the solution with ethanol and acetone repeatedly to obtain pure nano-copper. The invention has simple process, mild reaction conditions and short reaction time; in the reaction process, two protective agents are both adopted to prevent the growth and oxidation of nano-copper; the product performance is good, the particle size of nano-copper is less than 20nm, the nano-copper can not be oxidized for one month in the air; the production cost is low and no hazardous waste can be generated, thus meeting the demand of 'green production'.

PROCEDURE FOR PRODUCTION OF NANO DISPERSED COPPER POWDER
RU2009147519

FIELD: metallurgy. ^ SUBSTANCE: procedure for production of nano dispersed copper powder by reduction consists in mixing copper salt with solution of glucose, in dissolving salt at heating, in introduction of sodium hydroxide, in conditioning under isothermal mode and in successive extraction of metal copper in form of nano dispersed powder. Also, sulphate of copper is used as copper salt. Copper sulphate is mixed with solution of glucose at mole ratio of glucose to copper equal to (1.0-2.5):1.0. Dissolving is carried out at 50-60C. Sodium hydroxide is introduced upon complete copper sulphate dissolving and solution heating to temperature 70C.; It is carried out gradually at several stages for maintaining pH equal to 6-11 in process of reduction reaction, preferably, to 8-9, first to formation of oxide of univalent copper, and further to metal copper. ^ EFFECT: simplified, with reduced prime cost process of production of nano dimension particles of copper due to reduced number of process operations of synthesis.

Method for preparing copper zinc tin sulfur selenium nano particles
CN101830444

The invention relates to a method for preparing copper zinc tin sulfur selenium nano particles, which belongs to the technical field of photoelectric materials. The method comprises the following steps of: 1, mixing divalent zinc salt, divalent tin salt and monovalent or divalent copper salt, adding surfactant into the mixture, and heating the mixture; 2, mixing sulfur powder with selenium powder, adding the surfactant into the mixture, and heating the mixture; and 3, injecting solution obtained by the step 2 into solution obtained by the step 1 for warming, and purifying a product to prepare the copper zinc tin sulfur selenium nano particles. The preparation method has the advantages of no pollution, mild and simple reaction condition and low cost, and is suitable for large-scale production. The forbidden band width of the copper zinc tin sulfur selenium nano particles can change with the proportion change of sulfur selenium.

Method for preparing flaky nano copper powder
CN101890504

The invention discloses a method for preparing flaky nano copper powder, comprising the following steps: compounding a copper salt aqueous solution with a surfactant to obtain a mixed solution A, wherein the water in the copper salt aqueous solution and the surfactant form a surfactant compound system; (2) compounding a reducing agent aqueous solution with the surfactant which is the same as that used in step (1) to obtain a mixed solution B, wherein the water in the reducing agent aqueous solution and the surfactant form the surfactant compound system; (3) mixing the mixed solution B obtained in step (2) and the mixed solution A obtained in step (1) by stirring at the temperature of 20 DEG C to 80 DEG C, and reacting completely with stirring; and (4) carrying out solid and liquid separation on the product, washing the solid, and carrying out vacuum drying to obtain the flaky nano copper powder. The method of the invention has mild condition, simple process, short production period, low equipment investment and low preparation cost, and is easy to realize industrial production. The prepared flaky nano copper powder has consistent appearance, uniform size and good dispersibility.

Method for preparing nano-copper
CN101890506

The invention relates to a method for preparing nano-copper. The method comprises the following steps of: dispersing a copper precursor into the liquid paraffin serving as a solvent and a reducing agent; adding a surfactant into the liquid paraffin with stirring; performing heat treatment for 2 to 4 hours in an atmosphere of N2 protective gas or no protective gas; standing; centrifuging to obtain a product; and alternately washing the product by using petroleum ether, distilled water or absolute ethyl alcohol to obtain the nano-copper. The nano-copper powder prepared by using the method has the advantages of particle sizes of less than 50 nm, uniform distribution, difficult oxidation, high dispersibility in polar solvents or non-polar solvents, cheap raw materials, low energy consumption, simple process equipment and easy large-scale production.

Preparation method of nano-copper powder
CN102586800

The invention discloses a preparation method of nano-copper powder, which comprises the following steps of: in a self-made special electrolytic cell, taking metal copper as an anode, taking conducting material as a cathode, taking electrolyte as organic alcohol, taking ammonium salt which is soluble in the electrolyte as electrolyte, forbidding the ammonium salt to participate chemical reaction, and combining the cathode copper with the organic alcohol, so that the precursor of the nano-copper can be generated; mixing the precursor of the nano-copper, the organic alcohol or the other saturated hydrocarbon with the unsaturated hydrocarbon liquid, and sealing the precursor of the mixed nano-copper in a high-pressure kettle to be subjected to reductive treatment, so that the nano-copper powder can be obtained; and separating the nano-copper powder from the mixture of the reduced nano-copper powder and the organic solvent by an industrial centrifugal machine device, and washing with industrial alcohol, so that wet nano-copper powder can be obtained, putting the nano-copper powder into a vacuum drying oven, and treating, so that the nano-copper powder which accords with the national standard can be obtained. The preparation method has the advantages of being free of pollution in reaction process, small in investment, low in cost, good in product dispersibility, even in distribution, and capable of realizing different-quantity production.

Water-soluble nano-copper and preparation method thereof
CN102554217

The invention belongs to the technical field of nanometer materials, and particularly relates to water-soluble nano-copper and a preparation method of the water-soluble nano-copper, wherein the water-soluble nano-copper is nano-copper clusters which are surface-modified by stable organic single molecules formed in a way that organic compound surface modifier containing sulfydryl is bonded on the surface of copper nanoparticles. The invention can obtain copper nanoparticles which can be effectively dispersed in water phases and can exist stably, and is simple in preparation process and preparation devices, low in raw material cost, low in the production cost, high in yield, and suitable for large-scale industrial production, and the raw materials are easily accessible.

Copper nano powder and preparation method of copper nano powder
CN102601381

The invention discloses copper nano powder and a preparation method of the copper nano powder. The powder is prepared by the following methods: preparing copper salt solution from copper salt, adding a drag reduction agent and an inhibitor into the copper salt solution to prepare copper salt mixing solution and pre-heating the copper salt mixing solution; preparing reducing agent solution, adjusting the pH value of the reducing agent solution, and pre-heating the reducing agent solution; uniformly mixing and agitating the copper salt mixing solution and the reducing agent solution, heating the mixture to the reaction temperature to enable the mixture to react for a certain time so as to obtain nano copper suspension solution; aging the nano copper suspension solution; performing vacuum filtration on the suspension solution obtained from the above steps to obtain the copper nano powder in which the impurity is not removed; processing the copper nano powder to obtain the copper nano powder cladded by organic solvent; performing vacuum filtration on the copper nano powder to obtain wet copper nano powder, and performing low-temperature vacuum drying on the copper nano powder obtained from the above steps so as to obtain the copper nano powder product. The powder is small in particle size and narrow in particle size distribution. The method is controllable in reaction process, simple in technology and suitable for industrial production.

Method for preparing single copper with porous micro/nano hierarchical structure
CN102672198

The invention discloses a method for preparing single copper with a porous micro/nano hierarchical structure. The method comprises the following steps of adding phenol and vitamin C into a mixed culture of copper ion and ammonia water under the condition of one atmosphere at the temperature of 10 to 30 DEG C; quickly reacting; and preparing the single copper with the porous micro/nano hierarchical structure. By adoption of the preparation method, special equipment and reaction conditions are not required, and heating and calcination are not required, or a porous matrix serving as a template is not required to be added. The method has the characteristics that the process is simple, the energy consumption is low, the yield is high and the like, and is environment-friendly and suitable for large-scale industrial production.

Method for preparing ultrafine copper powder for electronic paste
CN102764898
The invention belongs to the field of electronic paste and particularly relates to a method for preparing a ultrafine copper powder for the electronic paste, comprising the step of adding a dispersing agent and a reducing agent into a copper salt aqueous solution to reduce the elemental copper. The method is characterized in that nano-copper particles are added into the copper salt aqueous solution before the dispersing agent and the reducing agent are added, wherein the nano-copper particles have the size of 5-20nm, and the addition amount of the nano-copper particles is 0.3-0.6 per mill of the mass of copper salt in the copper salt aqueous solution. The method is simple in process and flexible to operate. The raw materials are nontoxic, environmentally friendly and suitable for industrial production. The ultrafine copper powder prepared by the adopting the method has uniform and controllable particle size, a narrow distribution range and high purity.

Nano Antennas

RADAR ANTENNA
WO9214277

SENSORS BASED ON NANO-STRUCTURED COMPOSITE FILMS
CA2105869

ANTENNA
JP7221533

Magnetic core element for antenna, thin-film antenna, and card equipped with thin-film antenna
US5567537

PHASED ARRAY ANTENNA
JP11127020

GPS RECEPTION SECTION GRAVITY CENTER REGULATOR (NANOBALANCER)
JP2000174534

ANTENNA AND COMMUNICATION DEVICE
JP2003298338

MAGNETIC COMPOSITE MATERIAL FOR ANTENNA TAG
JP2004179270

SMALL-SIZED HIGH-SENSITIVITY ANTENNA
JP2005057444

Antenna for shielding electromagnetic radiation emission
TWI241739

APPLICATIONS OF NANO-ENABLED LARGE AREA MACROELECTRONIC SUBSTRATES INCORPORATING NANOWIRES AND NANOWIRE COMPOSITES
WO2004032191

DEVICE FOR MANUFACTURING CARBON NANO-CAPSULE ENCAPSULATING DNA, METHOD FOR MANUFACTURING CARBON NANO-CAPSULE ENCAPSULATING DNA,

AND CARBON NANO-CAPSULE ENCAPSULATING DNA
JP2005230970

Selective production of light-emitting ruthenium(II) ligand and mixed ligand complexes, used e.g. as molecular or nano-sensor, switch or wire, antenna, dendrimer, energy converter or photocatalyst, uses solvent and microwaves in first stage
DE102004009551

NANO-ANTENNA APPARATUS AND METHOD
WO2005119838

Ultra wide band signal receiver for ultra wide band data transmission system, has two sampling cell sets to sample waveform of signal received with preset delay, and correlator delivering detected signal
FR2881588

NANO STRUCTURE, MAGNETIC STORAGE MATERIAL USING IT, WIRING BOARD AND ANTENNA BASE MATERIAL
JP2006247795

Composition For Coating Organic Electrode And Method Of Manufacturing Organic Electrode Having Excellent Transparency Using The Composition
US2007200099

Microwave non reciprocal effect circulator for e.g. motor vehicle's radar antenna, has resonant disc comprising circular shaped matrix with nanowires whose density depends on distance from center of matrix
FR2884971

Transmitting Antenna Arrangement For Emitting A Longwave Wake-Up Signal For An Id Transmitter In A Keyless Motor Vehicle Access System
US2007279300

UHF ACTIVE ANTENNA AMPLIFIER FOR DIGITAL VIDEO BROADCAST (DVB) RECEIVER AND RELATED DVB DEVICE
US2007024373

Preparation method of carbon nano-tube array and its application in preparing antenna array
CN1948142

METHOD FOR CONCEALED SUPPRESSION OF EAVESDROPPING DEVICE, CONTAINING LOGICAL ELECTRONIC COMPONENTS
RU229265

Antenna structure using nana carbon tube and mfg. method
CN1979948

GLASS FIBER REINFORCED NANO COMPOSITES FOR OUTDOOR ANTENNA HAVING MICROWAVE SHIELDING PROPERTY
KR20070096299

Never charging, radiation-proof harmless mobile telephone
CN101064879

ANTENNA EQUIPMENT
JP2007006465

Applications of nano-enabled large area macroelectronic substrates incorporating nanowires and nanowire composites
US2006237537

Easy-manufacturing antenna with dielectric structure
EP1798812

ANTENNA APPARATUS
JP2007089232

RFID environmental manipulation
US2008197144

RFID antenna gain and range enhancement
US2008217309

TRANSMISSION ANTENNA AND TRANSMITTER USING THE SAME
JP2008219305

RFID silicon antenna
US2008217560

ANTENNA
JP2008228227

Nanostructured, magnetic tunable antennas for communication devices
US2008238779

NANO ANTENNA
WO2007118163

Body Chip
US2008266107

RFID chip
CN101359668

Manufacturing method of RFID chip
CN101359633

NON-CONTACT DATA TRANSMITTING/RECEIVING UNIT
JP2009093333

FILM ANTENNA, RECEIVER AND VEHICLE
JP2009177471

Process for preparing silicon nitride nano wave-pervious material
CN101239826

CONDUCTIVE POLYMER USING CARBON NANO TUBE AND METHOD FOR PRODUCING THEREOF
KR20090097031

DESIGN TECHNOLOGY OF HIGH TEMPERATURE TAG ANTENNA FOR RFID
KR20090103240

FIBER RFID TAG APPARATUS PRINTED BY USING CONDUCTIVE NANO INK AND A

MANUFACTURING METHOD THEREOF
KR20090114901

Preparation of magnetic conductive polyaniline nanometer composite material
CN101280106

Combination lightwave antenna and spectral analyzer and methods
US2009153871

Antenna useful in an antenna arrangement for mobile radio sector, comprises flat composite structure with the layers connected with one another, where the layers are electrically conductive lattice structure and nanostructure
DE102008002900

MAGNETIC MATERIAL, ARTIFICIAL MEDIUM AND MAGNETIC MATERIAL MANUFACTURING METHOD
JP2010010237

SOLAR PHOTOVOLTAIC STRUCTURE COMPRISING QUANTIZED INTERACTION SENSITIVE NANOCELLS
US2010193017

METHODS, SYSTEMS AND APPARATUS FOR LIGHT CONCENTRATING MECHANISMS
WO2009038791

CARBON NANOTUBE BASED VARIABLE FREQUENCY PATCH-ANTENNA
US2009231205

Mobile phone radiation-protection shielding device
CN201365281

Radiation-resistant shielding device for mobile phone
CN201345420

NANO AND MICRO BASED ANTENNAS AND SENSORS AND METHODS OF MAKING SAME
US2010097273

ANTENNA SUBSTRATE AND METHOD FOR MANUFACTURING THE SAME
KR20090099504

DYNAMICALLY DISTRIBUTABLE NANO RFID DEVICE AND RELATED METHOD
WO2010005953

LIGHT AMPLIFYING DEVICES FOR SURFACE ENHANCED RAMAN SPECTROSCOPY
WO2011005253

DYNAMICALLY TRIGGERABLE NANO RFID DEVICE AND RELATED METHOD
WO2010006332

Composite electromagnetic medium material containing dicyclopentadienyl iron phthalocyanine metal organic magnetic body and preparation method thereof
CN101692366

PRINTED RF LABEL FOR DISPLAY
KR100968578

Antenna and wireless communication device using same
CN102025018

Antenna and wireless communication device with the antenna
TW201112495

Antenna assembly, manufacturing method thereof and electronic device shell with antenna assembly
CN102035064

NANO-ANTENNA ENHANCED IR UP-CONVERSION MATERIALS
US2010103504

Solar battery with nano-sized antenna
CN101714837

ANTENNA MODULE AND WIRELESS COMMUNICATION DEVICE USING THE SAME
US2011050510

Method for manufacturing antenna and antenna manufactured by same
CN101719583

WIRELESS POWER TRANSFER DEVICE
KR20110075105

Composition for manufacturing antenna base and method for preparing base
CN101775195

NANO-ANTENNA FOR WIDEBAND COHERENT CONFORMAL IR DETECTOR ARRAYS
WO2010126641

NANO MEMORY, LIGHT, ENERGY, ANTENNA AND STRAND-BASED SYSTEMS AND METHODS
US2010155692

Radio frequency identification (RFID) tag antenna and manufacturing method thereof
CN102195128

Method for making a nano-scaled optical antenna array
TW201136825

Energy conversion new scheme
CN101877562

A method of fabricating RFID antenna substrates from non-woven slag fiber paper
TW201210119

METHOD FOR MAKING A NANO-OPTICAL ANTENNA ARRAY
US2011195201

METHOD OF MANUFACTURING CONDUCTIVE PATTERN TRANSFER FILM METHOD OF TRANSFERRING CONDUCTIVE PATTERN USING THE FILM
KR20120024177

Electronic scanning antenna composed of a network of two-dimensional radiating nano elements
EP2341579

ANTENNA SYSTEM USING CARBON NANO MATERIALS
KR101125135

Preparation and post-treatment device of multi-parameter accurate and adjustable multi-system microwave assisted nano-material
CN102092680

Post-processing device for multi-parameter precisely adjustable microwave-assisted nano material
CN201971629

A STRUCTURE COMPRISING NANO ANTENNAS AND METHOD FOR PREPARING THE SAME
KR20120088462

Quantum broadband antenna
US2012212375

Semiconductor nano-wire antenna solar cells and detectors
US2011284723

Fe-BASED NANO CRYSTAL ALLOY THIN BAND LAMINATE, MAGNETIC CORE FOR ANTENNA AND ANTENNA
JP2011254066

Device for detecting e.g. hazardous gas to measure hydrogen concentration in fuel cell field, has circular-shaped nano-structure arranged in plasmonic near-field or in nano-focus of triangular-shaped nano-structure
DE102011101698

MOBILE PHONE CASE PRINTING ANTENNA AND MANUFACTURING METHOD THEREOF
KR20120018722

SENSOR SYSTEM WITH PLASMONIC NANO-ANTENNA ARRAY
US2012050732

ENHANCEMENT OF MOLECULAR EMISSION USING OPTICAL-ANTENNA STRUCTURES
US2012091365

Rotary-insertion type RF mobile phone user identification card
CN2024221835

LIGHT MODULATORS AND OPTICAL APPARATUSES INCLUDING THE SAME
US2012170097

Nano integrated chip
CN202393916

NANOPHOTONIC PRODUCTION, MODULATION AND SWITCHING OF IONS BY SILICON MICROCOLUMN ARRAYS
US2012153142

Resonance-principle-based nano film half-wave plate
CN102721993

Manufacturing method of high-efficiency nano antenna solar battery
CN102709399

AUTONOMOUS LIGHT AMPLIFYING DEVICE FOR SURFACE ENHANCED RAMAN SPECTROSCOPY

US2012300202

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

Copyright 1996 - 2003

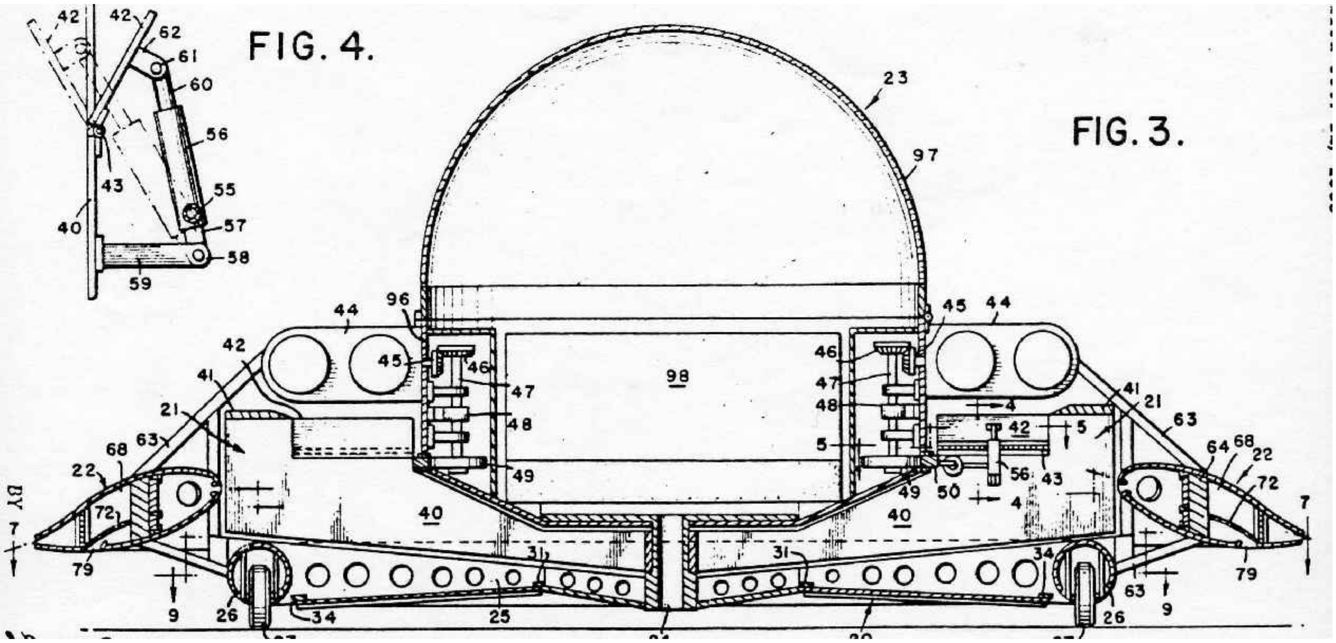

FIG. 4.

FIG. 3.

42
42
62
61
60
56
43
55
40
57
58
59
23
97
44
96
45
44
42
45
46
47
48
98
46
47
48
41
21
63
68
22
5
49
50
56
43
42
5
4
4
41
21
63
64
68
22
72
7
40
40
72
79
7
9
26
34
25
31
31
34
26
63
9
79
BY

AIRCRAFT

Filed Oct. 12, 1965

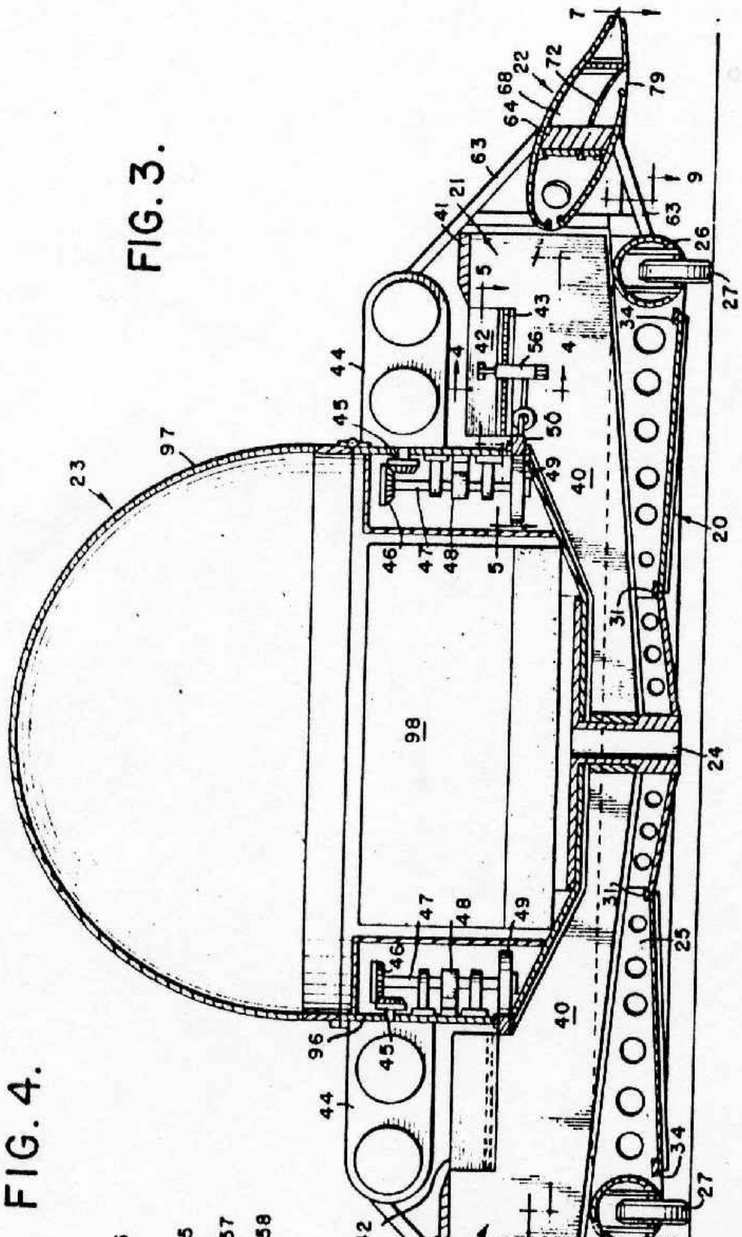
7
72
22
68
64
79
63
21
41
9
26
27
34
5
43
42
56
4
44
45
50
49
40
97
23
46
47
48
20
31
98
24
25
96
34
26
9
41
42
21
63
68
72
22
79
BY

FIG. 3.

FIG. 4.

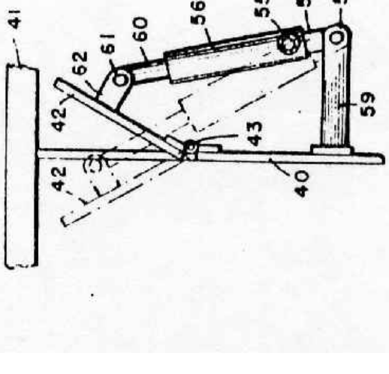
41
62
61
60
56
55
57
58
59
42
43
42
40

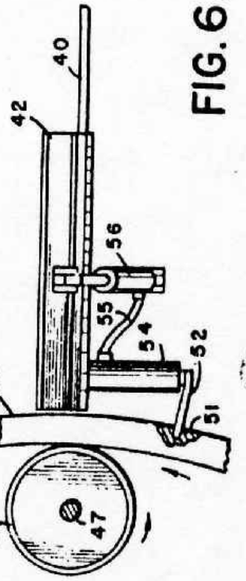
40
42
56
55
54
52
50
51
49
47

FIG. 6.

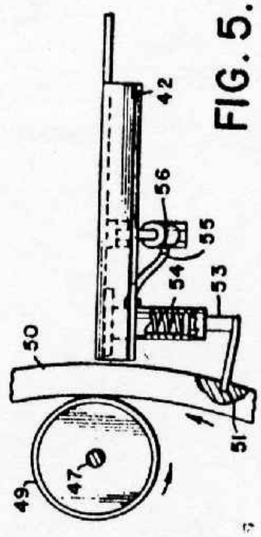
42
56
55
54
53
50
51
49
47

FIG. 5.

INVENTORS
CLARENCE D. LENNON
ROBERT G. VARNER

BY Garvey & Garvey

ATTORNEYS

FIG. 1.

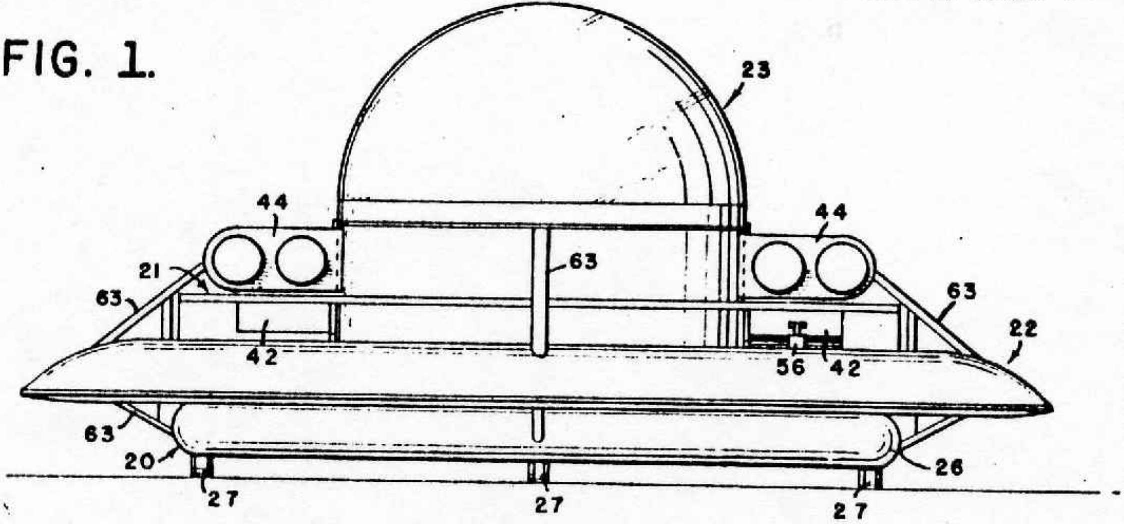
23
44
44
21
63
63
63
22
42
56
42
63
20
26
27
27
27

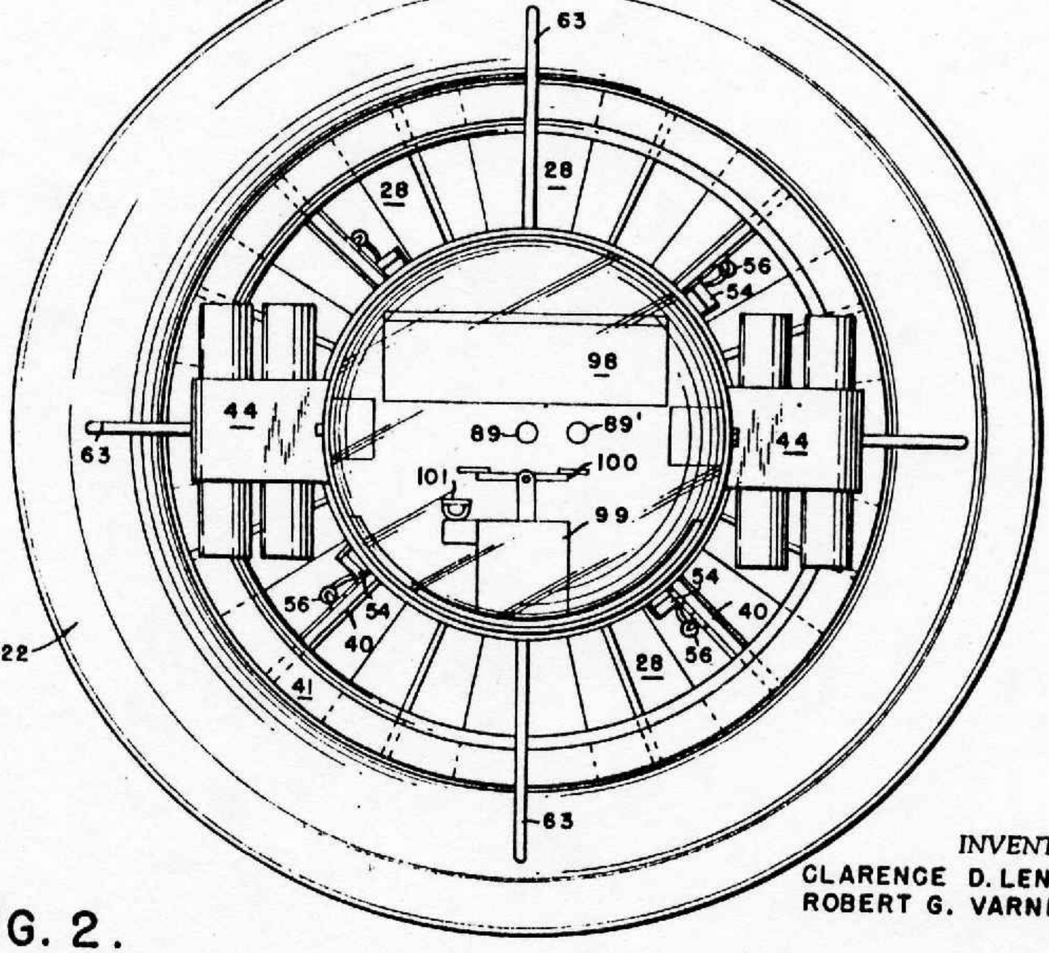
63
28
28
56
54
98
44
44
63
89
89'
100
101
99
56
54
40
54
40
22
41
28
56
63

FIG. 2.

INVENTORS
CLARENCE D. LENNON
ROBERT G. VARNER

BY Garvey & Garvey
ATTORNEYS

United States Patent Office

3,312,425
Patented Apr. 4, 1967

3,312,425
AIRCRAFT

Clarence D. Lennon, 1014 12th Ave., St. Petersburg, Fla. 33705, and Robert G. Varner, Box 823, Rte. 1, Palm Harbor, Fla. 33563

Filed Oct. 12, 1965, Ser. No. 495,172
10 Claims. (Cl. 244—12)

This invention relates to aircraft, and more particularly, to aircraft adapted for vertical take-off and landing, an object of which is to effect superior lift characteristics and high lateral speed capability.

Another object is to provide a disc-shaped aircraft including a cabin, a circular, horizontally-disposed, centrifugal impeller assembly around the cabin, and an airfoil embodying a lift and control device mounted circumferentially about, and immediately adjacent to, the impeller assembly, for directing air over and through the airfoil.

Other objects are to provide an airfoil of predetermined shape set at an angle to the horizontal, for effecting optimum lift; to provide an airfoil including a lift and control device having spaced, stationary and movable vanes forming a plurality of air chambers for effecting control, the movable vanes being adjusted to produce directional heading of the aircraft; and to provide an airfoil of the character described having a plurality of orifices selectively opened and closed to produce vertical and lateral thrust components with resultant vertical and lateral movement of the aircraft.

A further object is to provide an aircraft including an impeller assembly and airfoil of the character described, wherein the upper face of the impeller assembly is open and the lower face thereof is normally closed, closure being effected by a system of shutters which are adapted to be opened in the event of failure of the aircraft engines, producing auto-rotation of the impeller, to permit a safe emergency landing.

Other objects of the invention will be manifest from the following description of the present preferred form of the invention, taken in connection with the accompanying drawings, wherein:

FIG. 1 is a front elevational view of an aircraft constructed in accordance with the present invention;

FIG. 2 is a top plan view of the same;

FIG. 3 is a transverse sectional view of the aircraft of the present invention, showing to advantage details of construction;

FIG. 4 is an enlarged sectional view taken along the lines 4—4 of FIG. 3, looking in the direction of the arrows, and showing the movable blade of the impeller assembly forming a part of the present invention;

FIG. 5 is an enlarged sectional view taken on the lines 5—5 of FIG. 3, looking in the direction of the arrows, showing the movable blade of the impeller assembly in inoperative position;

FIG. 6 is a view similar to FIG. 5, showing the movable blade in operative position;

FIG. 7 is a sectional view taken along the lines 7—7 of FIG. 3, looking in the direction of the arrows and showing to advantage the airfoil lift and control assembly forming a part of the present invention;

FIG. 8 is an enlarged horizontal sectional view taken through the airfoil forming a part of the present invention;

FIG. 9 is a sectional view taken on the line 9—9 of FIG. 3, looking in the direction of the arrows, showing to advantage the base of the present aircraft;

FIG. 10 is an enlarged fragmentary plan view of a portion of the base showing to advantage the emergency shutter system forming a part of the present invention;

FIG. 11 is a sectional view taken along the lines 11—11 of FIG. 10, looking in the direction of the arrows, showing in dotted lines, the alternate position of the shutters;

FIG. 12 is a schematic showing of the control system for the movable vanes of the airfoil lift and control device; and

FIG. 13 is a schematic showing of the control mechanism for a sliding valve of the airfoil lift and control assembly forming a part of the present invention.

Referring now in greater detail to the drawings, the aircraft of the present invention comprises a base **20** of disc shape on which is centrally mounted a circular, centrifugal impeller assembly **21**. An airfoil lift and control assembly **22** is mounted circumferentially about, and immediately adjacent impeller assembly **21** for coaction therewith. A cabin **23** is centrally mounted above base **20** and impeller assembly **21**.

As shown to advantage in FIGS. 3 and 9, base **20** includes a central hub **24** from which a plurality of equispaced girders **25** of equal length extend radially. An annular flotation ring **26** is fixed in any suitable manner to the outer limits of girders **25** for buoying the aircraft in water. Base **20** further includes wheel assemblies **27** fixed to, and pending from, flotation ring **26** at 90° intervals for supporting the aircraft on land surfaces.

Base **20** is further provided with a plurality of shutters **28** hingedly connected at **29** to girders **25** intermediate flotation ring **26** and hub **24**. As shown in FIG. 11, shutters **28** are preferably hinged to both faces of each girder **25**, each shutter extending to a point midway adjacent girders **25**, so that a pair of shutters lies between the girders. Each shutter **28** is provided with an extension tab or abutment **30**, the tabs of adjacent shutters **28** being so located that they are contiguous when in the lowered position, as shown in FIG. 10. An inboard slidable ring **31** is supported by girders **25** superjacent extensions **30** of shutters **28**, said ring being provided with a series of recesses or notches **32** normally non-aligned with shutter extension tabs **30** but adapted to be moved into superjacent relationship with the tabs when it is desired to open shutters **28** to the position shown in dotted lines in FIG. 11. Each shutter **28** is additionally provided with an outer extension tab or abutment **33**, which tabs, as shown in FIG. 10, are also adapted to be in contiguous relationship when the shutters are in lowered or closed position. An outboard ring **34** is also slidably supported by girders **25**, which ring is also provided with recesses or notches **35** adapted to register with extensions **33** synchronously with the alignment of inboard ring recesses **32** with extensions **30** to permit opening of the shutters.

In accordance with the objects of the present invention, shutters **28** are opened only in case of emergency to effect safe landing of the aircraft, and the opening is effected manually by means of a shutter cable **36** trained over pulleys **37** and **38** to outboard ring **34**. A supplemental shutter cable **39** affixed to cable **36** is trained over pulley **37** into engagement with movable inboard ring **31**. Cable **36** extends upwardly into cabin **23** for ready access to the pilot. Shutters **28** are normally left in the closed position shown in FIGS. 3 and 9 and are only raised to the open position under emergency conditions to be hereinafter more fully discussed.

Impeller assembly **21** is of the centrifugal type and includes a plurality of vertically-disposed blades **40** extending radially from hub **24** in spaced relationship, a portion of each blade being reduced for interpositioning between base **20** and cabin **23**. The height of each blade **40** is increased laterally of cabin **23** and to the outermost portion thereof is affixed a ring **41** in engagement with the upper edge of each blade and extending completely around the aircraft, for shaping and directing the centrifugal flow of air produced by the rotating blades. Adjacent ring **41**, a portion of blade **40** is hingedly connected at **43** to the main portion thereof to provide a mov-

#707

April 4, 1967 C. D. LENNON ETAL **3,312,425**

AIRCRAFT

Filed Oct. 12, 1965 5 Sheets-Sheet 5

FIG. 12.

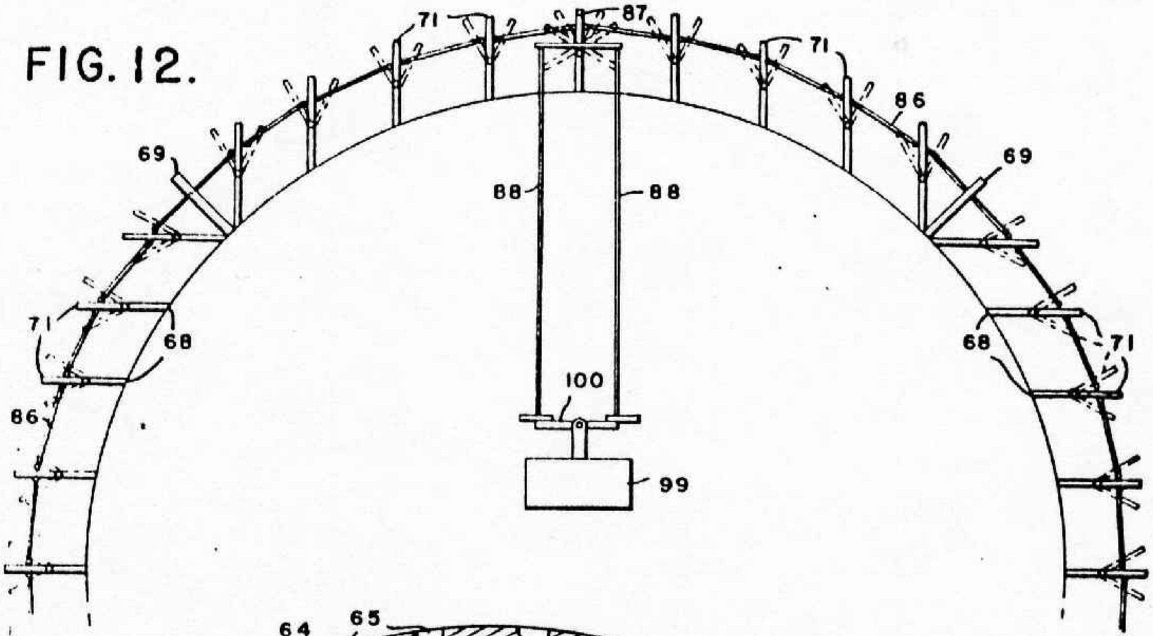

71
87
71
86
69
69
88
88
71
68
68
71
86
100
99

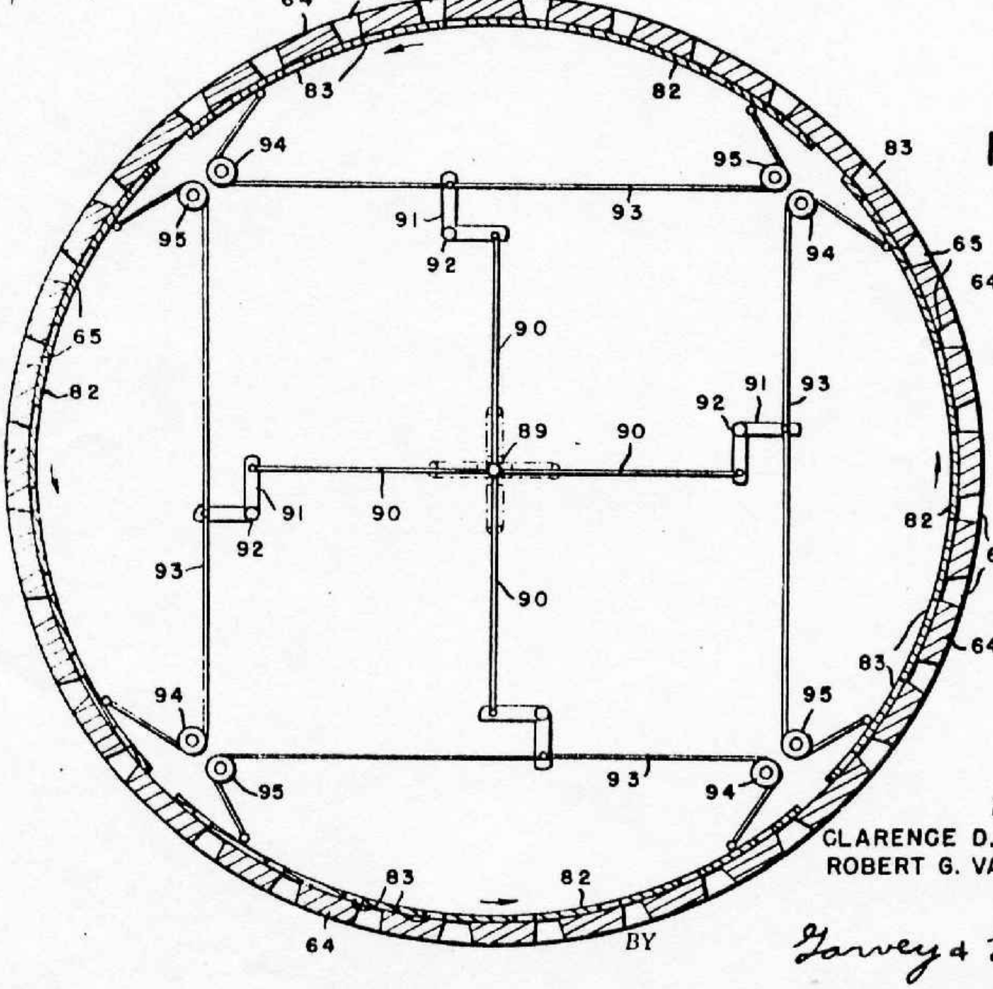

64
65
83
82
94
95
83
95
93
91
92
94
65
64
65
82
90
91
93
92
89
90
91
90
82
92
65
93
90
64
83
94
95
95
93
94
83
82
64

FIG. 13.

INVENTORS
CLARENCE D. LENNON
ROBERT G. VARNER

BY Garvey & Garvey
ATTORNEYS

April 4, 1967 C. D. LENNON ETAL **3,312,425**

AIRCRAFT

Filed Oct. 12, 1965 5 Sheets-Sheet 4

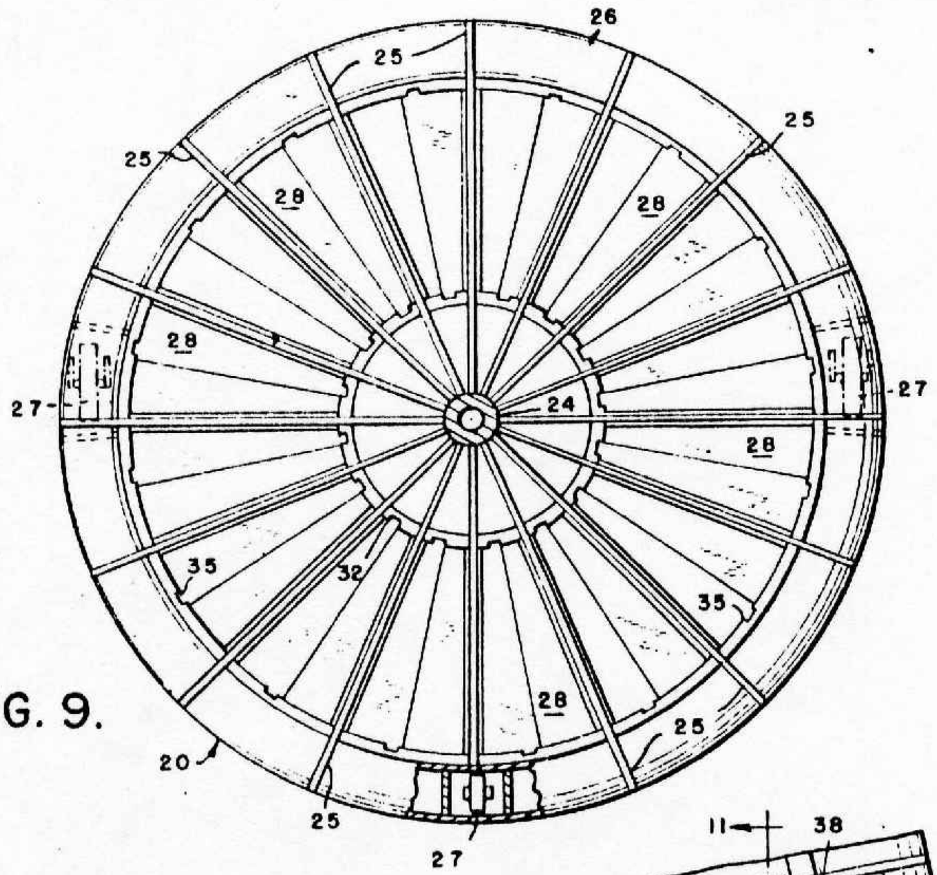
25
26
25
25
28
28
28
27
24
27
28
35
32
35
28
25
20
25
27

FIG. 9.

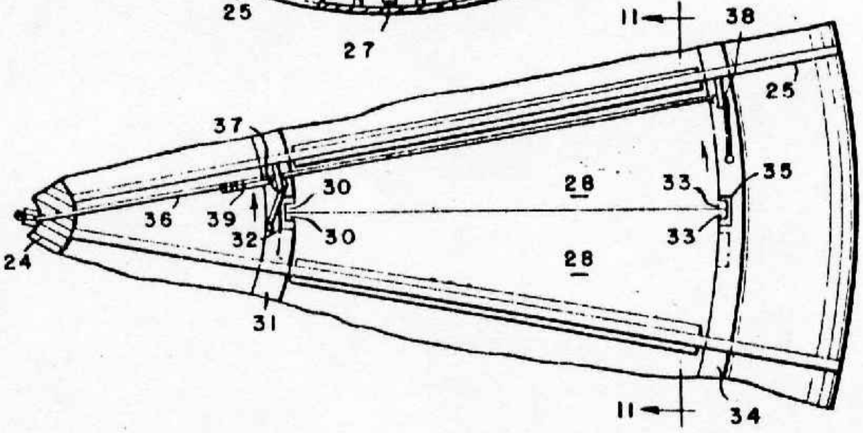
11
38
25
37
33
35
30
28
36
39
30
33
32
24
28
31
11
34

FIG. 10.

FIG. 11.

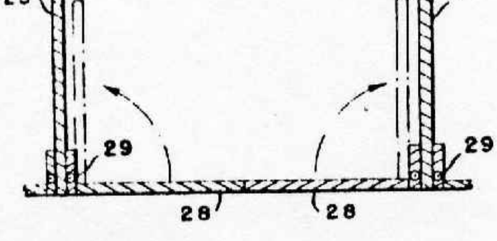
25
25
29
29
28
28

INVENTORS
CLARENCE D. LENNON
ROBERT G. VARNER

BY Garvey & Garvey

ATTORNEYS

April 4, 1967 C. D. LENNON ETAL 3,312,425

AIRCRAFT

Filed Oct. 12, 1965 5 Sheets-Sheet 3

FIG. 7.

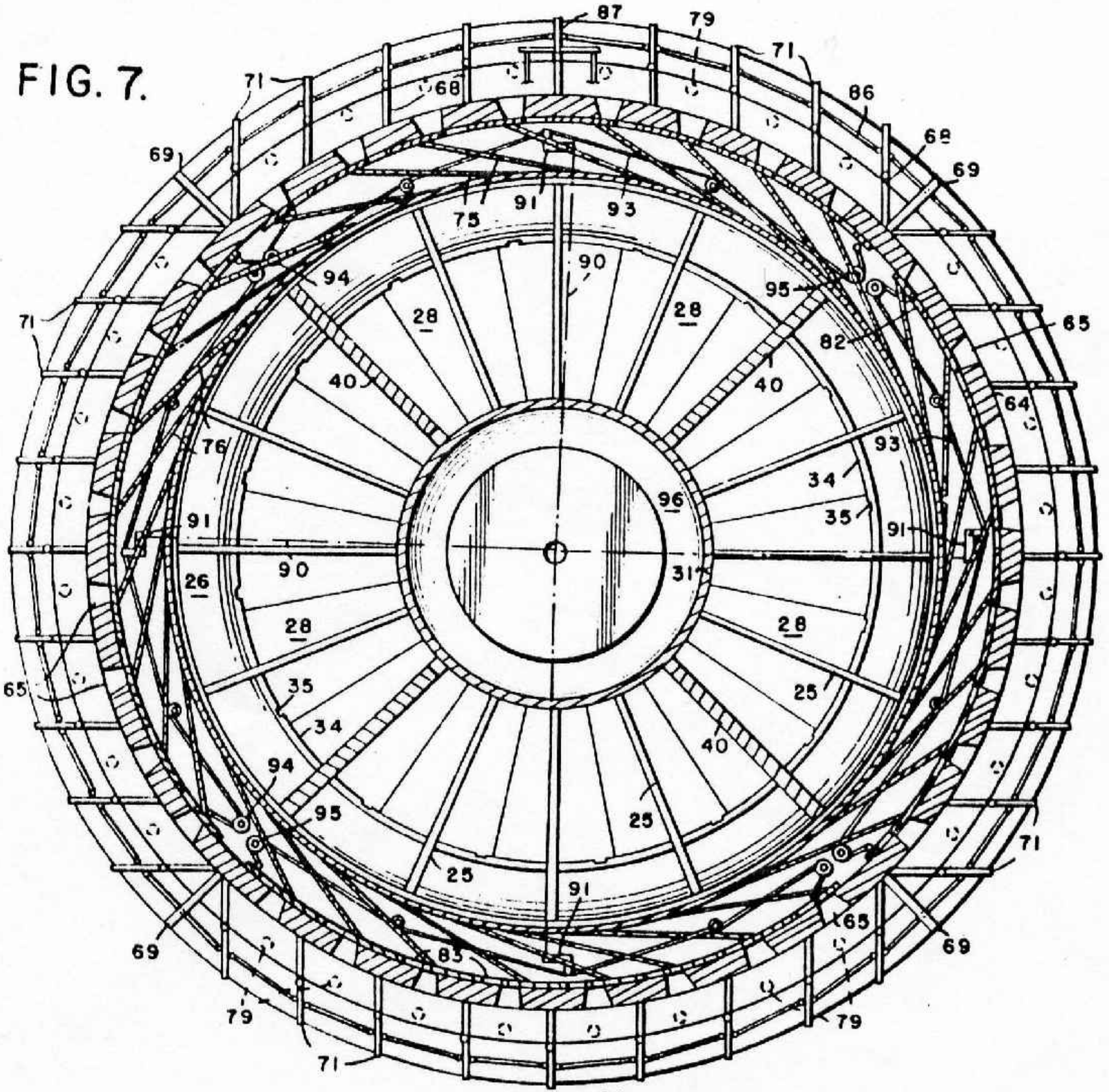

87
79
71
68
86
69
71
68
69
75
91
93
90
94
95
28
28
82
65
71
40
40
64
76
93
34
96
35
91
91
90
31
26
28
28
65
35
25
34
40
94
95
25
71
25
91
65
69
69
83
79
79
71

FIG. 8.

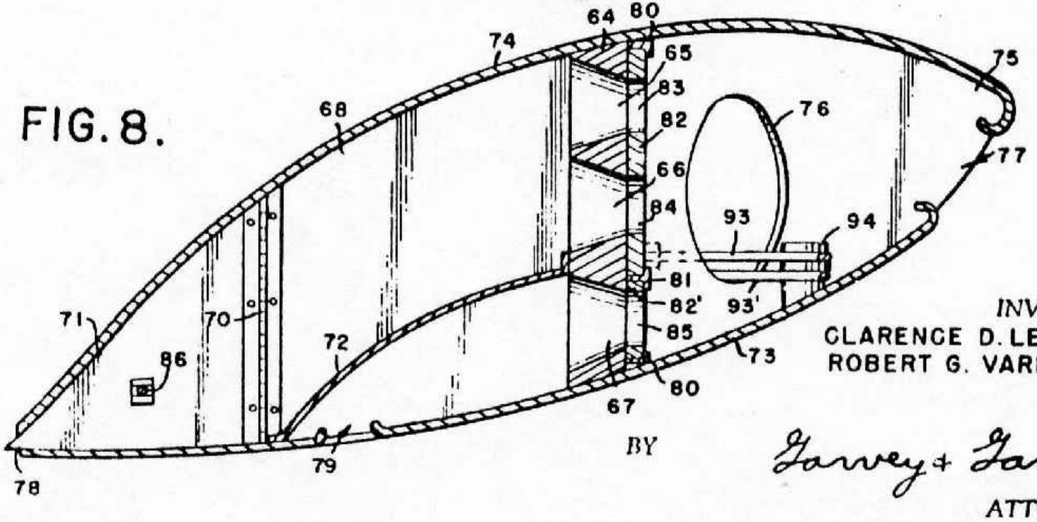

64
80
74
65
75
83
68
82
76
77
66
84
93
94
81
82'
93'
71
70
72
85
73
86
80
67
79
78

INVENTORS
CLARENCE D. LENNON
ROBERT G. VARNER

BY Garvey + Garvey
ATTORNEYS

BeFreeTech.com

[Back] [Next]

Suppressed Inventions
Suppressed Inventions 2
Atlantis Rising

Free Electricity
Subscribe
Make Money
Products
Videos & Audios
Links
Contact

Suppressed Inventions

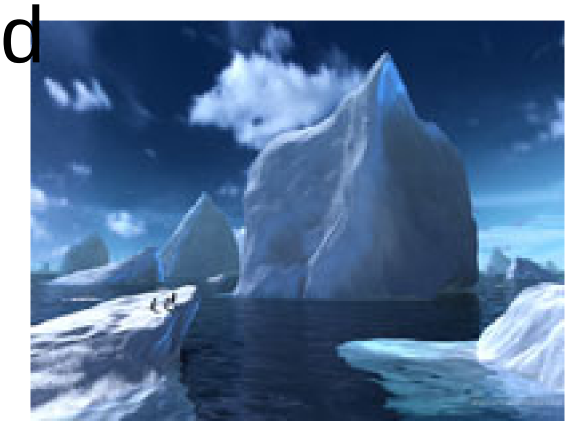

by LeRoy Pea (click to send an email) (updated Jan. 18, 2005)

GOVERNMENT INVOLVEMENT IN SUPPRESSED INVENTIONS: CHRONOLOGY COPYRIGHT APRIL, 1989 by PEA RESEARCH (PEA Research site is now at http://pea-research.50megs.com/)

Format for PEA RESEARCH article.

Each incident of "Suppressed Inventions" is presented in the following format and separated by a horizontal line:

> Date of incident   Reference
> Description of incident

(Note: US Patent drawings require an TIFF viewer attachment to Windows Explorer and can be obtained from: Internetiff.com for Explorer & Netscape. The FREE version

works just fine.)

Unable to Date: B1-B,3 CARBURETION, MOORE The most productive in the field of carburetion was George Arlington Moore. Out of 17000 Patents, 250 of them were related to carburetion. As far back as the mid 1920's, Moore was capable of virtually eliminating carbon monoxide pollution with his carburetors. People in the automotive field considered Moore to be a genius and couldn't understand why the automobile industry ignored his inventions. *Want to know more?...*Article Refers to "*"The Works of George Arlington Moore." "* Thanks Christopher Bird. (deceased 1996?) *Want to know more?...*Article US02123485 drawing. (Requires TIFF viewer.)

1924 CANCER CURE, Georges LakhovskyLakhovsky reported curing "tumors similar to those of cancer in animals" with his Multiple Wave Oscillator (MWO). *Want to know more?...*Article

1925 B1-A,31 GENERATOR, FREE ENERGY, HENDERSHOT Lester Hendershot successfully builds and test a FREE ENERGY device. (see 1961) *Want to know more?...*Article Free Energy Generators (#1) Thanks Bob Nelson. mailto:alchemy618@earthlink.net

1930's B1-A,37 GENERATOR, FREE ENERGY, MORAY Henry T. Moray is refused a Patent on his cold semiconductor cathode because the patent examiner couldn't understand how it could emit electrons. 20 years later the development of the transistor proved that his device could actually have worked! (see 1940) *Want to know more?...*Article "The Sea of Energy in Which the Earth Floats" FREE eBook. Thanks Bob. mailto:alchemy618@earthlink.net

---

1930's CANCER CURE, Royal Rife The cure for cancer was covered up, and the treatment suppressed. *Want to know more?...*Article "The Cancer Cure that Worked" report published in 1987. Thanks Planet Newspaper. *Want to know more?...*Article "Rife's Long-Suppressed Cure For Cancer and 50-Plus Other Diseases is Out of the Closet and Becoming Available." *Want to know more?...*Article "Suppression of the Rife Technology." Thanks Mark Lester. email

---

Jan. 1935 ANTI-GRAVITY, HENRY BULL. Popular Science publishes article about a

"Reactionless drive" for airplanes. Said to be "impossible", he demonstrates an easy to build model. *Want to know more?...*Article *Want to know more?...*

END

Jan. 3, 1935 J1-A,1,21 CARBURETION, POGUE, Pat.#353538 (Canadian) Charles N. Pogue is issued a Canadian Patent for a High Mileage Carburetor. (see 1/7/36) (See 1981, Ultra-Lean carburetors)

Apr. 9, 1935 J1-A,1,19 CARBURETION, POGUE, Pat.#US01997497 Charles N. Pogue is issued a Patent for a High Mileage Carburetor. (see 1/7/36) (See 1981, Ultra-Lean carburetors) *Want to know more?...*Article US01997497 drawing.

Jan. 7, 1936 J1-A,1,16 CARBURETION, POGUE, Pat.#US02026798 Charles N. Pogue is issued a Patent for his newer High Mileage Carburetor. Pogue used the carburetor for about ten years on his car and produced about 200 carburetors thru the Economy Carburetor Co. (see 4/30/36;

8/10/36) (See 1981, Ultra-Lean carburetors) *Want to know more?...*Article US02026798 drawing.

early 1936 J1-A,18 CARBURETION, POGUE BREEN MOTOR CO., Winnipeg, Manitoba, Canada: Tests the POGUE Carburetor on a Ford V-8 Coupe and got 26.2 miles on one pint of gasoline. The performance of the car was 100% in every way. Under 10 mph the operation much smoother than a standard carburetor. T.G. Green, President of the BREEN MOTOR CO., did the tests. (see 1/7/36) (See 1981, Ultra-Lean carburetors)

Apr. 30,1936 J1-A,18 CARBURETION, FORD MOTOR CO. FORD MOTOR CO., Winnipeg, Manitoba, Canada: Tests the POGUE Carburetor and is at "a loss to understand" how the carburetor got "25.7 miles on one pint of gasoline"! (That's approximately 205 mpg). Mr. W.J. Holmes and Mr. Purdy conducted the test for FORD MOTOR CO. (see 1/7/36)

Aug. 10,1936 J1-A,18 CARBURETION, POGUE S. Stockhammer tested the POGUE Carburetor on a 1934 Ford V-8 Coupe and got 28 miles per pint of gas. "I can say the performance was all anyone could desire in every shape of form." (see 1/7/36) (See 1981, Ultra-Lean carburetors)

Dec. 12, 1936 J1-A,26 CARBURETION,

POGUE Canadian Automotive Magazine states that the standard carburetor gets about 25 mpg at only 9% efficiency. Therefore the POGUE carburetor is 72% efficient overall at 200 mpg. (see 1/7/36) (See 1981, Ultra-Lean carburetors)

1940 B1-A,37 GENERATOR, FREE ENERGY, MORAY Henry T. Moray demonstrates before members of the Public Utilities Commission (PUC) his FREE ENERGY generator that gave a continuous output of 250,000 volts with no apparent input. The next day he was found shot in his lab and all of his notes and device were stolen. Currently, the Cosray Research Inst. of Salt Lake City has been collecting Moray's papers from old acquaintances and attempting to duplicate the device. (see 1930's)

Oct. 22, 1940 J1-A,14 CARBURETION, NEWBERY, Pat.#US02218922 Ivor B. Newbery gets a Patent for a high mileage carburetor that he says delivers a truly dry vapor to the engine and produces very little, if any, exhaust pollution. *Want to know more?...*Article US02218922 drawing.

1945 Carburetors marked "POGUE CARBURETOR, DO NOT OPEN" were used on unnamed source American Army tanks throughout WWII but were removed from circulation after the war ended. (see 1/7/36)

(See 1981, Ultra-Lean carburetors)

1946 B1-A,40 GENERATOR, FREE ENERGY, COLER, von UNRUH CLASSIFIED BRITISH INTELLIGENCE REPORT: A German researcher by the name of Hans Coler, was visited and interrogated about his "Magneto-tromapparat" FREE ENERGY generator. The inventor claimed to get electrical energy without the use of a chemical or mechanical source of power. This device consisted of only permanent magnets, copper coils, and condensers in a static arrangement and delivered 450mv for a period of some hours. Coler also claimed his "Stromerzouger" could deliver 6kw of power from the few watts delivered by a dry cell. (see also, Joseph Newman). *Want to know more?...*Article "Alleged New Source of Power". Thanks Bob Nelson. mailto:alchemy618@earthlink.net

---

1947 J3-C,43 FLYING SAUCER, T. BROWN Townsend Brown constructed Flying condensers shaped like Saucers. A high positive charge on the leading edge caused thrust to be generated in the direction of the region of low flux density. His Saucers required no jets or propellers, no moving parts at all. They modified the gravity field around themselves to produce, in effect, the incline of a "hill" under the craft, which they slid down like a surfboard on a wave. (aka Biefield-Brown effect) *Want to know more?...*Article "The Flying Saucer" Thanks Mason Rose! Optical Multimedia *Want to know more?...*Article US02949550 1960 Brown flying saucer drawings.

1947 FLYING SAUCER, J.R.R. SEARL. Searl, an military electrical engineer, builds a composite motor driven disc. After starting the device, the force of the disc breaks the motor drive shaft. The disc jumps up into the air, glows blue, and disappears. When the English inventor tries to get a patent on the disc, it is "secretized" by the Queen of England. *Want to know more?...*Article Levitation disc photos. Thanks Bob Nelson! mailto:alchemy618@earthlink.net *Want to know more?...*Article Thanks John Searl!

---

late 1950's J1-A,1 CARBURETION, FISH/KENDIG B1-B,3 Variable Venturi The Fish/Kendig Variable Venturi carburetors get some very interesting mileage figures. The Fish carburetor even got into production on a small scale, but for some unknown reason, these carburetors fade away in the early 1960's. (see early 1970's) *Want to know more?...*Article US03752451 1973 Kendig variable venturi carburetor drawing. *Want to know more?...*Article US02801086 1957 J. R. Fish carburetor drawing.

1953 J1-A,26 CARBURETION, POGUE CARS Magazine states that in the opening months of 1936, Pogue panicked the Toronto stock exchange and threw fright into the major oil companies. Stock exchange offices & brokers were swamped with orders to dump all oil stock immediately. This same article states that Winnipeg's largest automobile dealers tested the POGUE carburetor and got results of up to 216.8 mpg! (see 1/7/36) (See 1981, Ultra-Lean carburetors)

Mid 1950's B1-B,3 CARBURETION, BASCLE, Pat.#US04257378Joseph Bascle got a Patent for an improved carburetor which was used on every car in the Yellow Cab fleet in Baton Rouge, NY. It raised mileage by 25% and decreased pollution by 45%. *Want to know more?...*Article US04257378 1981 drawing.

1953 J3-E,39 INTEGRATON, VAN TASSEL UFO Contact, George W. Van Tassel begins the construction project on the INTEGRATON machine in Yucca Valley, CA. Its' purpose is to rejuvenate the human physical frame.

(Tom McGraw, deceased, told of sending a camera thru time at the Integratron and photographing dense vegetation of the PAST where, now, is only desert.) *Want to know more?...*Article "Introduction to the Integratron" theintegratron@aol.com

Aug. 23, 1958 J3-A,91 FLYING SAUCER, OTC-X1, CARR Spacecraft Convention, Pleasonton, CA: Panel discussion between Daniel W. Fry and the inventor of the OTC-X1 Flying Saucer, Otis T. Carr and Major Wayne S. Aho, formerly Army Combat Intelligence Officer during WWII. (see 5/5/60)

Mar. 9, 1959 J3-A,49 FLYING SAUCER, OTC-X1, CARR Immediate News Release to ALL Media: OTC Enterprises announces plans to market kit plans to build a prototype model of the firm's projected spacecraft, the OTC-X1, said to be under construction now for launching later this year. Mr. Carr plans to launch a six-foot operating model in April in Oklahoma where it is now being built in field laboratories. (see 4/19/59;5/5/60)

Apr. 19,1959 J3-A,53 FLYING SAUCER, OTC-X1, CARR The OTC-X1 didn't fly for several reason today: Mr. Carr was in the hospital for eight days with lung hemorrhage. There is no indications of cancer, TB or other lung ailments. Also, in the preliminary pre-flight test the accumulator developed a leak which sprayed mercury all over the inside mechanism. "We don't plan to announce the next pre-flight test in advance", said Mr. Wayne S. Aho. (see 5/5/60)

Nov. 10, 1959 J3-A,19 FLYING SAUCER, OTC-X1, CARR Pat.#US02912244Otis T. Carr gets a Patent for an "amusement device" (Flying Saucer). He claims that this actually demonstrates a workable space craft. *Want to know more?...*Article US02912244 drawing.

Dec. 10, 1959 J2-D, J3-A,58 FLYING SAUCER, OTC-X1, CARR Mr. Wayne Aopes to say, "take me to your leader, moon-man!" Mr. Aho announces plans to take the OTC-X1 to the moon at 12:01 a.m. on Dec. 7, 1959. He will get there in 5 hours and orbit it for 7 days, then he will return to earth. The Carr space ship is 45 ft. in diam. and weighs 30 tons and is powered by electricity. Maj. Aho was formerly an Army Intelligence Officer who says he landed on Omaha Beach in Europe on D-Day and won the Bronze Star and Purple Heart for wounds received in the Battle of the Bulge. He claims to have seen 13 Flying Saucers in flight. (see 5/5/60)

May 5, 1960 J3-A,54 FLYING SAUCER, OTC-X1, CARR Pat.#US02912244 Victorville, CA: First time ever in any newspaper: Otis Carr tells how his Flying Saucer, the OTC-X1, works. Carr developed his power plant for the saucer at the Osbrink Plant in Apple Valley.

The OTC- X1 is described "as being two tops that spin in opposite directions around a single axis...The entire circular-foil craft that is Mr. Carr's most profound development, is an assembly of only two major moving mechanical parts." (see 3/9/59;1/61)

*Want to know more?...*Article "Otis T. Carr & his Flying Machine" Thanks Jerry Decker. Jerry Decker

---

June 2, 1960 J3-A,57 FLYING SAUCER, OTC-X1, CARR Otis T. Carr told an "open house" audience of 300 people on Memorial Day that it is a "treacherous misstatement of fact to say or infer that we (OTC Enterprises) are coming to California to raise money in stock sales". Carr appeared to be planning to sell rights to build, sell or operate the OTC-X1 Flying Saucer he claims he'll be producing at the former Osbrink Plant. Thus Carr wouldn't be violating a federal injunction secured by the U.S. Securities Exchange Commission ordering him to stop selling unregistered stock. (see 5/5/60)

Jan., 1961 J2-B,38 FLYING SAUCER, OTC-

X1, CARR True Magazine labels Mr. Otis Carr and his OTC-X1 a hoax and a con-artist. Attorney General J. Lefkowitz is investigating him and Lefkowitz said that Mr. Carr has bilked $50,000 from various people in New York. (see 5/5/60)

1961 B1-A,31 GENERATOR, FREE ENERGY, HENDERSHOT, Dr. Ed Skilling, from Columbia University, successfully builds and test a HENDERSHOT free energy device, out of which he got 300 watts. Skilling had been associated with Hendershot and learned of the device thru him. The generator was self-resonant at 500 khz. (see 1925) *Want to know more?...*Article "Assembly of the Hendershot basket weave Power Capture Unit." Thanks Craig Svenson.

May 2, 1961 J1-A,12 CARBURETION, SHELTON, Pat.#US02982528 Robert S. Shelton gets a Patent for a high mileage carburetor that he claims gives 8 to 10 times the mileage of a standard carburetor. *Want to know more?...*Article US02982528 images.

July 20,1961 J3-A,102 GENERATOR, FREE ENERGY, COLTON IMMEDIATE PRESS RELEASE: I, Norman Evans Colton (Director of Sales, O. T. Carr co.), do hereby declare under oath that I have perfected a machine to draw electricity from the atmosphere without the use of any fuel. The machine is operated entirely by environmental gravitic forces. A newly constructed "Colton-Gravity-Electric-Engine" has been installed at the Claremont Hotel in Berkeley, CA, where it may be seen, examined, evaluated and photographed.

1963 J3-B,51 BINDING METER, W.B. SMITH Wilbert B. Smith, founder and editor of "Topside" magazine, claims to have had contact with Aliens and to have found out the reason why there space ships have crashed. There are certain areas of this planet that are afflicted with "reduced binding". The 3 basic fields of nature that affect "binding" are tempic, electric and magnetic, all of which are non-linear in there reactions to each other. He claims to have developed a "binding meter" to measure these forces. *Want to know more?...*Article Thanks Grant Cameron. presidentialUFO@canada.com

---

Dec. 26, 1966 J1-A,10 CARBURETION, SCHWARTZ, Pat.#US03294381 Harold Schwartz gets a Patent for a high mileage Vaporizer Carburetor that he claims to be 100% efficient with virtually no exhaust pollution. *Want to know more?...*Article US03294381 drawings.

1966 CANCER CURE, Dr. Wlm. F. Koch, and Dr. F. M. Eugene Blass Confronted by the FDA, both doctors are found murdered in 1966. *Want to know more?...*Article Thanks Mark Lester. email

---

1968 J1-A,1 CARBURETION, FORD MOTOR CO. FORD MOTOR CO. begins experiments with an engine that has a new kind of combustion chamber. A 430 cub. inch is tested and runs very well on a 26:1 air/gas mixture. Estimate possible production by 1985. (see 1985)

Early 1970's B1-B,3 CARBURETION, KENDIG A young college student (name unknown), entered his Mercury "gas hog" in the California Air Pollution car race and won hands down. He was using the KENDIG Variable Venturi carburetor that a small company (name unknown), manufactured in Los Angeles for racing cars. Within a week the student was told to remove the carburetor since it wasn't approved by the Air Resources Board (ARB). His car had reduced pollution and doubled the gas mileage (for that model). (see late 1950's) *Want to know more?...*Article Kendig Technologies, Inc. Thanks Marty!

1970's B1-B,6 ALCOHOL The U.S. Army's Nalick Laboratories in Massachusetts discovered and developed certain fungi which could convert a wide variety of cellulose into the sugars necessary for producing alcohol. Using the fungi to reduce paper scrap, they could produce 65 gals. of high grade alcohol from one ton of scrap. *Want to know more?...*Article DOE 1982 publication #5046267.

1971 ANTI-GRAVITY, Henry Wallace US03626605, US03626606, US03823570Patents awarded to Henry Wm Wallace for "anti-gravity" technology. *Want to know more?...*Article Thanks Robert Stirniman. *Want to know more?...*Article US03626605 diagrams.

1972 J3-E,16 OTEC, NASA NASA Report on OTEC (Ocean Thermal Energy Conversion): "Tapping the energy of the Gulf Stream could supply all the electrical energy needs of the US as far ahead as 1985 with only three-tenths of a degree reduction in the temperature of the great warm ocean river. This reduction may be beneficial as it would slightly offset the ocean's thermal pollution due to other uses." *Want to know more?...*Article What is Ocean Thermal Energy Conversion?

Apr. 4, 1972 J1-A,6 CARBURETION, TUCKER, Pat.#US03653643 Oliver M. Tucker gets a Patent for a high mileage Vaporizer Carburetion System. This is the basic type that La Pan used as mentioned in Mechanics Illustrated. (see 1974) *Want to know more?...*Article US03653643 drawings.

1973 J1-A,1 CARBURETION, SHELL OIL CO. At the Annual SHELL OIL Companies employees contest, a highly modified 1959 Opel sets the record with 376.59 mpg at

Wood River, IL.

1974 B1-B,3 CARBURETION, DRESSERATOR Dresser Company SELLS OUT to Holley Carburetor and FORD MOTOR CO. Lester Berriman spent 5 years designing and testing the "Dresserator" carburetor for Dresser Co. which used a super-accurate mixture control to obtain a 22:1 mixture. Pollution standards were passed along with a typical gain of 18% mpg.

May, 1974 B1-B,6 GAS/WATER ADDITIVE John Andrews, a Portugese chemist, demonstrated a water to gas additive before Navy officials which allowed ordinary water to be added to gasoline without decreasing the combustability of the gas and would drive the cost of gasoline down to 2 cents per gallon. When Navy officials finally went to his lab to negotiate for the formula, they found Andrews missing and his lab ransacked.

Mid 1970's B1-B,3 CARBURETION, FRANCH, KRAFT Guido F. Franch demonstrates a water-to-gas miracle to the chemists at HAVOLINE Chemical Co., MI, and to the University of Michigan. Using a "green powder" starting agent derived from soaking coal in a liquid (unnamed), he cold make water operate a car better than gasoline. He said he learned this process from a coal miner named Alexander Kraft over 50 years before.

1974 J1-A,1 CARBURETION, La Pan An article from Mechanics Illustrated about "Humidifier type fuel systems" tells of a man by the name of La Pan who claims to get from 60 to 100 mpg with his system. Mechanics Illustrated

---

July, 1974 J3-C,43 GENERATOR, FREE ENERGY, EVGRAY Pat.#3890548 The L.A. District Attorney, acting under orders from the "Top", raids Ed Gray's assembly plant and confiscated his prototype generator as well as his plans and records, brought numerous false charges against him, and defied all attempts by Gray's lawyers to get the confiscated materials returned. Gray's opponents have driven him into bankruptcy. (see 6/17/75)

Apr. 22, 1975 B1-A,26 GENERATOR, FREE ENERGY, ECKLIN, Pat.#3879622 John W. Ecklin, of Alexandria, VA, gets a patent for a FREE ENERGY generator that is powered by 2 kw and delivers 12 kw output.

June 17,1975 B1-A,4-23 ENGINE, FREE ENERGY, EDWIN GRAY, Pat.#3890548 Edwin V. Gray, of Evgray Enterprises, Van Nuys, CA, gets a Patent for a FREE ENERGY automobile engine. The engine starts off of two car batteries and once it gets going, it recharges the batteries and powers the car. Unfortunately, Evgray Ent. has become embroiled in a controversy with the

Securities Exchange Commission (SEC), which is unrelated to the engine itself but has halted the development of the engine. (see 1975)

1975 B1-A,4 ENGINE, FREE ENERGY Governor Ronald Reagan, of California, gave a special commendation to Edwin Gray, of Van Nuys, CA, for his Patent of a FREE ENERGY automobile engine. (see 6/17/75)

Aug. 12, 1975 B1-A,51 GENERATOR, FREE ENERGY, KINNISON, Pat.#US03899703 Robert W. Kinnison of Fort Collins, CO, gets a Patent for a FREE ENERGY generator which offers a hope for a fuelless economy. His design resembles ECKLIN'S generator. (see 4/22/75)

Sep. 18, 1975 J2-F, J3-F,48 Joseph C. Yater invents an Heat-to-Electricity converter that he says will cost the consumer approx. $200 and would be up to 90% efficient. The device operates by capturing "fluctuation voltage" (static noise heard on radios and amplifiers). The rooftop device would be heated by the sun and use millions of microcircuits to tap the freed electrons from heated molecules. Yater took his device to the Government, which declared that his device had "real potential". After being promised a working model within 6 months by the Gov't., they came back to him with the reply that the device would be impractical.

1976 J1-A,1 CARBURETION A modified Ford Pinto equipped with a turbocharged Nissan diesel engine is tested and gets up to 80 mpg.

1977 J1-A,1 CARBURETION Pat.#US04177779 Tom Ogle of El Paso, TX, claims to have gotten 100 mpg on a 1970 Ford Galaxy with a V-8 engine. The Ford used only 2 gals. of gas on a trip from El Paso, TX, to Demming, NM, roundtrip! His Vapor Fuel System eliminates the carburetor and has no exhaust pollution emission! (see 1981) *Want to know more?...*Article US04177779 drawing.

July 26, 1977 J3-D,29 WEATHER CONTROL, CONSTABLE Trevor James Constable aims his Weather Control Machine at the cloud system near Palm Springs, CA, and created a coastal cooling during an excessive heat spell. He also produced record rains for Utah in the worst drought stricken area of the United States. He successfully completed other Weather Divergence plans and notified the Federal Government of these plans as required by Federal Regulations. *Want to know more?...*Article "Loom of the Future, Weather Engineering."

1978 CANCER CURE, Rene Caisse Rene cured patients of cancer, using herbs. After her death in 1978, all of her records in her house were destroyed. *Want to know more?...*Article Thanks Mark Lester. email

---

Apr. 2, 1978 J3-E,21 ELF WAVES A man-made standing columnar ELF wave collapsed at Bell Island, Newfoundland, Canada, today. This conclusion was reached after an analysis done by the Dept. of Geology at Memorial University of Newfoundland and with senior officials of Newfoundland and Labrador Hydro. Affects of this wave collapse were: Power lines were vaporized. A power surge was reported. A voltage regulator was fused. All fuses leading to the local power company's transformer melted, yet the transformer escaped unscathed. Ball lightning was seen, rumbling and humming below 60 hz was felt and heard. *Want to know more?...*Article (ELF Waves and) Total Population Control. *Want to know more?...*Article Waveguide. Thanks R. Bedard. rbedard@cruzio.com

---

1978 J1-A,1 CARBURETION The Flex Gas Vaporizer (as advertised in national magazines) and sells for a reasonable price, claims to double gas mileage on most any car and gets 110 mpg on some!

Sept. 3, 1978 US04188580J3-E, 32 SCRAMBLER DEVICE, Patent applied for but REFUSED. U.S. Government Agency Restricts CB Radio & Telephone Privacy device: Four inventors, Carl Nicolai, William

Raike, Carl Quale and David L. Miller of Seattle, WA, developed a device to prevent eavesdropping on CB radio and telephone transmissions for nonmilitary, nongovernment use. The NSA was quick to issue a secrecy order against them. This means they may not talk about the invention in any detail without risking a jail term! The order was received six months after a patent was filed for the device. "Your application...has been found to contain subject matter, the unauthorized disclosure of which might be detrimental to national security," wrote the Patent Office. (This appears to be part of a general plan to limit the privacy of the American people.) *Want to know more?...*Article US04188580 drawings.

1979 J1-A,1 CARBURETION Ralph Moody, Jr., of Oak Hill, FL, gets 84 mpg from his modified Ford Capri which has been equipped with a turbocharged 4 cylinder Perkins diesel engine.

Apr. 24, 1979 B1-A,59 MOTOR, FREE ENERGY, JOHNSON, Pat.#US04151431 Howard R. Johnson gets a Patent for a FREE ENERGY motor that starts itself and would deliver 5 kw from permanent magnets. (see Spring, 1980) *Want to know more?...*Article US04151431 drawings.

Spring, 1980 B1-A,59 MOTOR, FREE ENERGY, JOHNSON Science and Mechanics magazine interview regarding the JOHNSON FREE ENERGY motor: When asked if he (Johnson) thought that the load on his motor might drain the magnetic spin (of the electrons) he replied, "I didn't start the electron spins and I don't know anyway to stop them - do you?" The Johnson Permanent Motor is one of the most promising possibilities in the world of "Free Energy". It is simple to construct and yields high output. (see 4/24/79) *Want to know more?...*Article Science & Mechanics, 1980. *Want to know more?...*Article Open sourcing the Howard Johnson Magnetic Motor technology.

Sep. 11, 1980 J-3J,4 RADIONICS A-9 CHARLES WHITEHOUSE Norfolk, VA: Lawyer Charles E. Poston decided he had the "right" to "judge" Dr. Charles Whitehouse, of Virginia Beach, VA and revoke his chiropractic license on the grounds he was practicing medicine outside the scope of his chiropractic profession. Guilty?... He was helping many people FREE OF CHARGE, and especially helping mentally retarded children with color therapy and radionics. Whitehouse is on the Board of Directors of the U.S. Psychotronics Assoc. and is an expert Radionics Operator. Even the U.S. Navy has purchased one of his radionics inventions for tracking submarines. Whitehouse was giving away approx. $200,000 in treatments a year. (see 12/80) (see 1/7/81) *Want to know more?...*Article Section: *United States Psychotronic Association.*

1980 J1-A,27 CARBURETION, POGUE Mr. Arthur C. Sgrignoli, after 45 yrs., has built a legendary POGUE carburetor by hand and is said to have achieved an efficiency of 86%. Contact was made thru his brother, William J. Sgrignoli. (see 1/7/36) (See 1981, Ultra-Lean carburetors)

Dec., 1980 J-3J,5;A-9 RADIONICS Norfolk Circuit Court, VA: Dr. Whitehouse won his appeal and was awarded his chiropractic license back. (see 9/11/80)

Dec. 1980 ANTI-GRAVITY, COOK USO4238968. Robert Cook gets a patent for "converting centrifugal force to linear force." *Want to know more?...*Article USO4238968 drawings. *Want to know more?...*Article CIP home page by Robert Cook. RCook2@bak.rr.com or bob@flakmag.com

---

Jan. 7, 1981 J-3J,5;A-9 RADIONICS Virginia State Supreme Court, VA: The lower courts' decision is reversed and Dr. Charles Whitehouses' chiropractic license is again revoked. Whitehouse will again be going to court in March, 1981. (see 1/7/81) (see 12/80)

Jan. , 1981 J1-A,29 POGUE, Charles N. As of January, Charles N. Pogue is still alive at age 81 and is living in a rest home in Winnipeg, Canada. He refuses to talk to anyone or to receive visitors from outside his own family. (see 1/7/36) (See 1981, Ultra-Lean carburetors)

1981 J-3J,8 TESLA EXPERIMENTER, ROBERT GOLKA Wendover, UT: Robert Golka, a well-know Tesla researcher, has been victimized by an attack on his workshop which has, for years, been in a deactivated Air Force hanger. His high powered energy tower outside the hanger has been partially destroyed. A witness said insulators and sections of the tower were laying around the base. In the mean time the rent on the Air Force hanger has been raised 1000 percent! These new difficulties now threaten to bring an end to his research with Tesla devices.

1981 J-3J,8 SUPER PAINT Research funded by Naval Air Systems Command has come up with a polished film paint that will reflect 98% of the energy directed at it by a high energy laser beam weapon thus making Star Wars obsolete. [So why haven't we channeled the money to more worth while causes?]

1981 J-3J,16 CANCER CURE DOCTOR ARRESTED Buenos Aires, Argentina: Carlos Jerez has been jailed because he had healed 13,000 cases of cancer, including terminal cases. Jerez used a maser whose electric output was polarized through an artificial blue sapphire crystal. He also invented a computerized tomographer that detected and indicated the missing chemical elements in the diseased body. His equipment and properties were seized by the secret police and totally destroyed. All of his medical records were confiscated by the Servicio de Informaciones del Estado (SIDE) and his records are being kept by the Secret Police. He is incarcerated at the Prison of San Pedro, 250 kilometers north of Buenos Aires, Argentina. *Want to know more?...*Article (in Spanish)

1981 J1-A,29 CARBURETION, POGUE ULTRA-LEAN CARBURETORS of Northridge, CA, is selling a set of plans for the POGUE carburetor for $50. (see 1/7/36) *Want to know more?...*Article FREE from REX RESEARCH. Thanks Bob! mailto:alchemy618@earthlink.net

---

1981 J1-A,28-29 CARBURETION, WALLACE Allan Wallace's experimental carburetor gets 30 mpg at a speed of 55 mph for 60 miles. *Want to know more?...*Article "The Secret Super High Mileage Report" FREE eBook. Thanks Bruce! bmcburne@becon.org

---

1981 J1-A,29 CARBURETION, SHELL MOTOR OIL CO. At Sell Motor Oil Mileage Marathon, Great Britain, a new world record is set at 1368 mpg on a specially built three-wheeler with a 90cc engine using the "sprint & coast" method. *Want to know more?...*Article Finnish Mileage Marathon Clubs.

1981 J1-A,29 CARBURETION, OGLE Tom Ogle SOLD OUT his high mileage carburetor to ADVANCE FUEL SYSTEMS of Seattle, WA, for $250,000 while still retaining 20% interest in his device that got 100 mpg on a 1970 Ford. (see 1977) *Want to know more?...*Article 100 mpg article from April 1979. 100mpg@earthlink.net

---

1983 WATER CARBURETOR, CELLA Carl Cella builds an water carburetor from surplus aviation parts and mounts in his 1979 Cadillac Coupe de Ville. Runs fine, with no exhaust pollution. He is told by DOE agents if he tried to sell pre-built units, he'd have a lot of "problems". *Want to know more?...*Article

Carl Cella hydrogen car. Thanks Josef Hasslberger. sepp@lastrega.com

---

1985 J1-A,1 CARBURETION Well, the 1968 FORD MOTOR CO. high-mileage carburetor (26:1) isn't on the market yet. I wonder what happened? (see 1968)

1993, AIDS CURE, Steven Kaali US05188738 *Want to know more?...*Article "HIV & Virus attenuation by blood electrification." US05188738 drawings. *Want to know more?...*Article "HIV & Virus attenuation by blood electrification."

Nov. 1997 ANTI-GRAVITY, RICHARD FOSTER US05685196. Foster gets patent for inertial propulsion device. *Want to know more?...*Article US05685196 drawings.

May 1998 ANTI-GRAVITY, DAVID COWLISHAW. Dave publishes the "GIT Theory of Gravity." *Want to know more?...*Article Thanks David Cowlishaw!
David Eugene Cowlishaw

Misc. research & rumors:

1. The mixed chemical stone found in England and secretized by the Queen.? It generated its' own electrical sparks.

2. The Alexander Patent (#US03913004) claims 3 watts out for 1 watt in. J-3J,15

3. Demonstrations by S. Seike (now deceased) at the Germany Symposium of Gravitational Field J-3J,13 SEIKE Energy. (Nov. 27, 1980)

*Want to know more?...*Article "Unconventional", suppressed, dormant, or emerging sciences, technologies, inventions, theories, therapies. Thanks Bob Nelson. mailto:alchemy618@earthlink.net (1982? Yep, I was getting articles in the snail-mail from Bob before there were home computers!)

*Want to know more?...*Article Energy 21: Suppressed and unusual technologies links contents page for free and alternative energy, renewable, sustainable power generation. Thanks Geoff Egel.

*Want to know more?...*Article More history on Suppressed Inventions. Thanks Byron! byronw@erols.com

*Want to know more?...*Article Purposely Wasteful - Automobile mpg's. Thanks Free Republic!

*Want to know more?...*Article Government, Big Business Prevent Marketing of Super

High Mileage Carburetors. Thanks Free Republic!

*Want to know more?...*Article Action to Declassify Secret Energy Patents. Thanks Drunvalo Melchizedek. webmaster@spiritofmaat.com

*Want to know more?...*Article Energy Suppression, an Invisible Galaxy of Inventions. Thanks Christopher Bird. (deceased 1996?)

*Want to know more?...*Article Suppressed Science.

*Want to know more?...*Article Energy Technology. Thanks Josef Hasslberger. sepp@lastrega.com

*Want to know more?...*Article Suppressed Science/ No Ice Age/ Tesla/ Air-Powered Cars. Thanks Tom Brown. mailto:thomas@thomasbrown.org

*Want to know more?...*Article Where in the World is all that Free Energy? Thanks Peter Lindemann. support@free-energy.ws

*Want to know more?...*Article "The World of Free Energy" by Peter Lindemann.

*Want to know more?...*Article Suppressed

Science & Technology. Thanks NEXUS.

*Want to know more?...*Article "Secret Government Technology, Reverse Engineering of Alien Artifacts." Thanks Edgar Fouche. Fouche@connecti.com

*Want to know more?...*Article Prescription Drugs: 3rd leading cause of death in North America! Thanks Warren Wichman. WRWichman@yahoo.com.au

*Want to know more?...*Article The World has been pulled over your eyes, to blind you from the truth. Things are not as they appear to be. Thanks Vitamin City.

*Want to know more?...*Article Dark Government News Site. Alternative and Suppressed News.

*Want to know more?...*Article Strategic Overview of Suppressed Technologies.

*Want to know more?...*Article Suppression from Higher Up. Inventors Beware! Thanks Erik Masen.

*Want to know more?...*Article The goal of Energy Technology Alliance (ETA) is to create a cooperative synergy and an alliance with present research and development from several groups.

W. I. PENNOCK.

APPARATUS FOR COLLECTING ATMOSPHERIC ELECTRICITY.

APPLICATION FILED JUNE 26, 1907.

911,260. Patented Feb. 2, 1909.

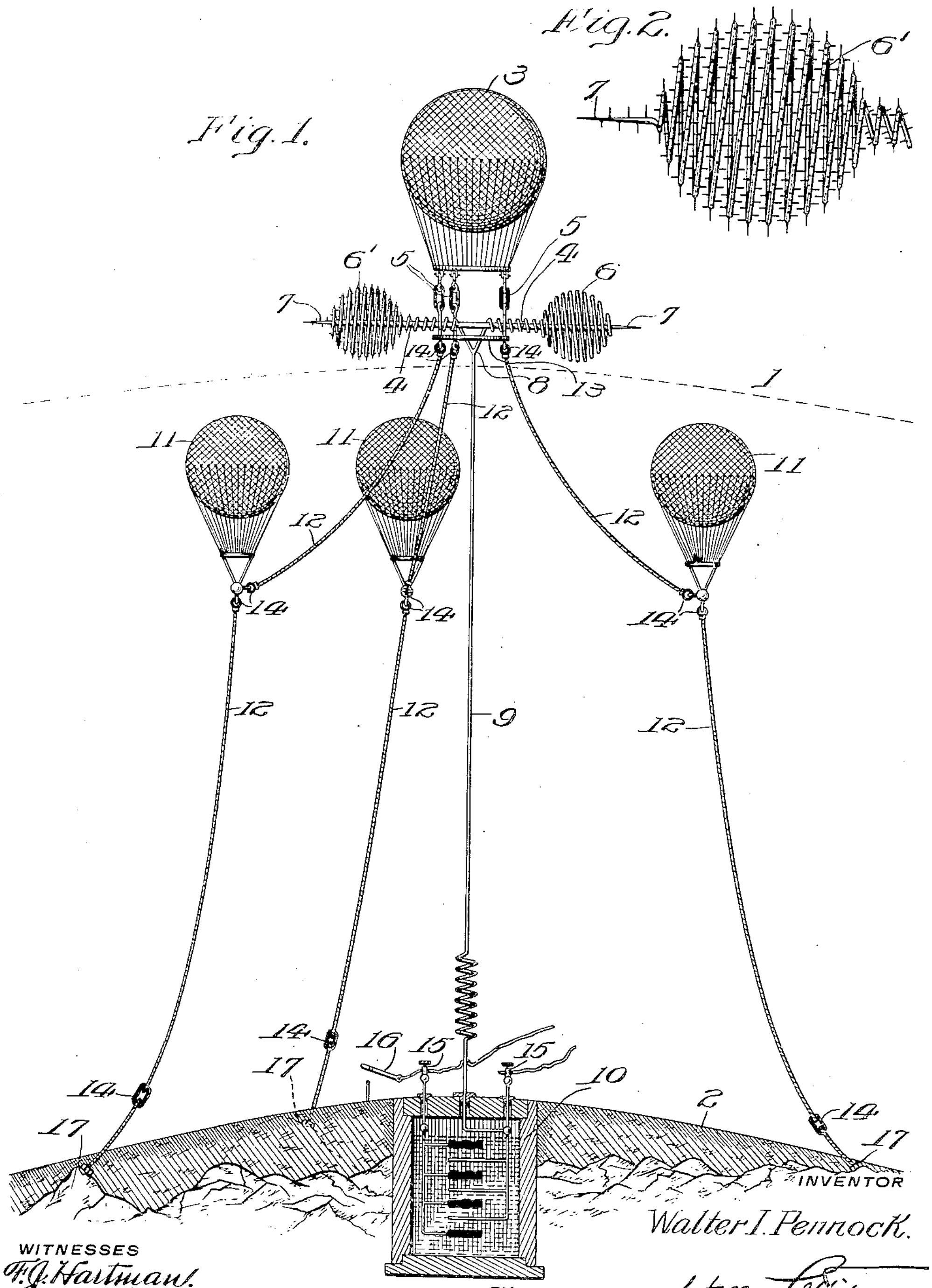
Fig. 2.
6'
7
Fig. 1.
3
5
4
6'
5
6
7
7
14
14
14
4
8
13
1
12
11
11
11
12
12
14
14
14
12
12
9
12
14
16
15
15
10
17
14
17
2
14
17

WITNESSES

F. J. Hartman.

Alston B. Moulton

INVENTOR

Walter I. Pennock.

BY

ATTORNEY

UNITED STATES PATENT OFFICE.

WALTER I. PENNOCK, OF PHILADELPHIA, PENNSYLVANIA.

APPARATUS FOR COLLECTING ATMOSPHERIC ELECTRICITY.

No. 911,260. Specification of Letters Patent. Patented Feb. 2, 1909.

Application filed June 26, 1907. Serial No. 380,907.

To all whom it may concern:

Be it known that I, Walter I. Pennock, a citizen of the United States, residing at Philadelphia, in the county of Philadelphia and State of Pennsylvania, have invented certain new and useful Improvements in Apparatus for Collecting Atmospheric Electricity, of which the following is a specification.

My invention relates to a method of collecting electricity from a strata laden with electricity at high altitudes in the atmosphere, through the medium of a wire cable suspended from one or more balloons and in conveying this electricity to the earth's surface.

The object of my invention is to provide a conveyance of the electro motive force to be found in the upper strata of the atmosphere to the earth's surface; where it may be utilized for commercial and other purposes.

A further object of my invention is to provide a device or mechanism by which a suitable collector for the electrical energy in the upper strata of the earth's atmosphere may be elevated in the said strata and by which the said electric energy may be transmitted to and collected at the earth's surface, from which point it may be conducted to any place where it is desired to use the same.

A further object of my invention is to support and anchor said device in any desired position.

A further object of my invention is to provide an improved form of collector through which the energy of the said upper strata of the earth's atmosphere may be collected and transmitted for various purposes to the earth's surface.

Other objects of my invention will appear in the specification and claims below.

For a further full, clear and complete disclosure of my invention, reference may be had to the following description and accompanying drawings, in which like reference characters refer to corresponding parts.

Figure 1 is an elevational view of one form or embodiment of my invention and Fig. 2 is a detailed view of one form of my improved collector.

The passage of the electrical current to the earth under ordinary conditions is prevented by an obstruction afforded by the dense lower strata of the atmosphere, which is a bad conductor of electricity, as shown by the electrodes of an ordinary static machine. When the electrodes are placed close together, the atmosphere is seen to carry the current across from one electrode to another, but if placed far apart the current is obstructed by the intervening air. The dense lower strata of the atmosphere affords one of the best non-conductors of electricity, as shown in the conduction of the electric current by telegraph, or trolley wires on the earth's surface, where only a small quantity of the electric current escapes through the atmosphere; while rarefied atmosphere affords a good conducting media, as shown by the vacuum tube. The upper strata of the atmosphere being rare in proportion to the altitude, and being a good conductor of electricity while the lower strata of the atmosphere being dense and affording a non-conducting media for the electric current, thus causes an obstruction to the electric current, or power in its passage to the earth's surface from the electric strata of the atmosphere under ordinary conditions. When, however, the cumulus clouds of a thunder storm pass over the surface of the earth, these clouds being of very great height, the moisture in said clouds forms a better conductor of electricity than does the dry air, with the result that the electricity in the upper strata of the atmosphere breaks through the said cloud as a streak of lightning and in that form reaches even to the earth's surface, while the thin or shallow strata of clouds, observed in the so-called "settled rain" storm, do not extend upward to a sufficient height to form a conducting medium for the electricity from the electric strata to the earth's surface. For this reason there is usually no lightning during the said latter variety of rain storm.

By means of my invention, I have provided a mechanism for collecting the electrical energy or power created by nature and stored in the upper strata of rarefied air of the earth's atmosphere and have provided a conductor for said electric energy to the earth's surface.

Referring now to the drawings 1 indicates what may be called the lower limits or boundaries of the strata of electric energy above the surface 2 of the earth.

3 indicates a balloon which is elevated to a high altitude until it enters the said stra-

tum. The said balloon 3 carries a bar 4 of wood or any other suitable non-conducting material suspended by insulating links 5 or any other suitable form of insulation from the balloon 3. Upon either end of the said bar 4 I mount collectors, one form of which I have illustrated as spheres of coiled wire 6—6′ the outer ends 7 of which terminate in sharp points. The inner turns of the said spheres 6 and 6′ are wrapped around the wooden bar 4 and united as at 8 to a conductor of electric energy, preferably a large copper wire 9. This wire or conductor 9 extends to the earth's surface and may have its end suitably attached to an electric accumulator or other piece of electrical apparatus. I have illustrated one form of my invention in which the conductor 9 is connected to one pole of a storage battery 10 on the earth's surface.

The spiral spheres 6, 6′ are preferably provided with a polished metallic surface to form a good conductor of electricity and the material of said spheres should also be of such a character that it will not rust or corrode or tarnish. A polished copper wire or a copper wire plated with platinum or gold or a solid platinum or gold wire may be used for this purpose, inasmuch as these materials are least affected by moisture and the atmosphere. The said spheres or collectors may be made of smooth wire as shown in the sphere designated by the numeral 6 or of barbed wire, as is shown at 6′ and illustrated on a larger scale in Fig. 2 of the drawings. The latter form is preferable inasmuch as it provides a large number of points through which the electricity may flow to the wire from the surrounding air in the said upper strata of the earth's atmosphere.

In order that the supporting balloon 3 may be held in a relatively fixed position, it should be suitably anchored to the earth's surface. Inasmuch as the balloon 3 must be elevated to a very high position, the weight of the anchoring cables forms an important consideration, and if desired or necessary one or more supplemental balloons 11, 11, 11 may be attached to each of the anchoring cables 12, 12, 12 as illustrated in Fig. 1, to relieve the balloon 3 of such weight as would prevent it from ascending into the said electrical strata. In order that the electricity from the said upper strata of the earth's atmosphere may not be conducted down the anchoring cables 12, 12, 12, I attach them to the supporting balloon 3 and to the supplemental balloons 11 and to the earth's surface through suitable insulating devices 14.

In the form of my invention illustrated in Fig. 1, below the bar 4, I suspend a ring 13 of any suitable material from the balloon 3, and attach the anchoring cables 12 thereto by means of insulating rings 14, 14, 14. I may also provide additional insulating rings 14 between that portion of the anchoring cables 12 between the balloon 3 and the supplemental balloons 11, and also between that part of the cable 12 between the supplemental balloons 12 and the earth's surface. I may also provide, near the earth's surface and at the lower end of the anchoring cables 12, similar insulating links 14. While I have described links as forming a convenient form of insulating device for the purposes above set forth, I do not wish to be construed as being limited to the same, inasmuch as any suitable non-conducting connection may be used in place of the links 14.

The terminals 15, 15 of the storage battery 10 may be connected to any piece of electric apparatus which it is desired to run or operate and if desirable one of the poles of the battery is adapted to be connected by the switch 16 with the earth's surface. The lower ends of the anchoring cables 12 are securely anchored to the earth's surface as at 17.

With the apparatus arranged and connected in the manner illustrated in Fig. 1, the electric energy in the high strata of the earth's atmosphere passes to and through the conductive spiral spheres 6 or 6′ to the conductor 9 and is suitably stored or used at the earth's surface, while the balloons 11 support a part of the weight of the anchoring cables 12, and permit the balloon 3 to ascend as high as is possible, or necessary for it to enter the said electrical strata of the earth's atmosphere. By arranging the anchorage 17 of the cables 12 symmetrically or in any other position than in a straight line, the balloon 3 may be held in a substantially fixed position with relation to the earth.

While I have illustrated in the drawings and have described in the specification a form of apparatus in which my invention may be carried out, it is obvious that the drawings are more or less diagrammatic drawings, that is to say, that the proportions of the various parts are not necessarily those which would operate to the best advantage, inasmuch as certain portions have been shown as greatly enlarged in the drawings for the sake of clearness, and that it is likely that more than one supplemental balloon would be requird for each cable in order to support the weight of the same, and to relieve the supporting balloon 3 of such weight, as would prevent it from ascending into the high electrical strata of the earth's atmosphere, but such changes in form, proportion and arrangement I regard as being fully within the aim and scope of my invention, so long as such forms or modifications fall within the scope of the appended claims. It is also to be understood that the storage battery or accumulator which I have shown as being connected to my collector, is only a type of apparatus which can be operated by the current collected by the spheres 6, 6′ and transmitted to the earth through the wire 9, and when I use the word "accumulator", I mean any piece of useful apparatus which is operated by the current transmitted thereto through the wire or conductor 9.

Having thus described my invention, what I claim and desire to protect by Letters Patent of the United States is:

1. The combination with an electrical collector comprising a bar of non-conducting material, and an open spherical conductor carried by said bar, of means to support said collector in the high electrical strata of the earth's atmosphere.

2. The combination with a balloon, of an electrical collector supported thereby and insulated therefrom, comprising a bar of non-conducting material, and a conductor wound spirally around said bar.

3. The combination with an electrical collector comprising a non-conducting bar, and a conductive wire wound thereon to form an open substantially spherical body, and means to support said collector in the high electrical strata of the earth's atmosphere.

4. The combination with a balloon, of an electrical collector carried thereby and comprising a non-conducting bar, and a polished wire wound spirally thereon to form an open substantially spherical body.

5. The combination with an electrical collector comprising a non-conducting bar, and a wire wound spirally thereon to form an open substantially spherical body, said wire being provided with pointed conducting projections, and means to support said collector in the high electrical strata of the earth's atmosphere.

6. The combination with a balloon, of an electrical collector carried thereby comprising a non-conducting bar, and a conducting wire wound spirally thereon to form open substantially spherical bodies upon the opposite ends thereof.

7. The combination with a balloon, of an electrical collector carried thereby comprising a non-conducting bar, a conducting wire wound spirally thereon to form open substantially spherical bodies upon the opposite ends thereof, an electrical accumulator, and an electrical connection between the said collector and said accumulator.

8. The combination with a balloon, and means to anchor said balloon, of an electrical collector supported by said balloon and insulated therefrom, an electrical accumulator, and a conductor connecting said collector and said accumulator.

9. The combination of an electrical collector, means to support said collector at a high elevation from the earth's surface and within the electrical strata of the earth's atmosphere, an electrical accumulator at the earth's surface, an electrical connection between said collector and said accumulator, and means to insulate said supporting means from said collector and from the earth.

In testimony whereof, I have hereunto set my hand this 25th day of June, 1907.

WALTER I. PENNOCK.

Witnesses:
HUGH F. QUINN,
WM. G. GLENN.

OVERUNITY ELECTRICITY

Atmospheric Electric Generator

https://overunity-electricity.blogspot.com/2022/04/atmospheric-electric-generator.html?m=0

Atmospheric Electric by Nikola Tesla

Nikola Tesla is known for his invention of the rotating magnetic field, which is the principle behind alternating current (AC) power. He also developed a version of radiant energy he called "Ether Power" or "Atmospheric Electric Generator" (AEG). AEG extracts electrostatic potential from the atmosphere and converts it into useful electrical energy. It is a free-energy device that does not require any external power source to operate.

There are several versions of the AEG design, but all of them extract energy from the atmosphere by using an electrically charged metal plate. The plate is charged by either an external power source or by capturing energy from natural lightning storms. When the plate is exposed to the atmosphere, it creates an electrical field that extracts electrical potential from the air. This potential is then converted into useful electrical energy by one or more electrical generators.

The original version of Nikola Tesla, radiant energy generator, both resonant and non-resonant, can be made in many different versions.

A practical guide to the electrical (non-resonant) calculation of radiant energy: How to Build Tesla's Radiant Energy Receiver

A famous version is Moray's (with resonance to generate arbitrary energy):

Practical guide - in just one box - engineer Moray's way
☆* **Revealed At Last**: Ancient Invention Generates Energy-On-Demand
The design includes:

- Harnessing electricity from the Earth: Neither is Schumann Resonance, nor is it known by Electromagnetism. It's Energy from Ether in Which the Earth Floats
- Extracted from ordinary electricity by the method called “fractionation.”
- Reverse Tesla coil - "Back to Back" mechanism
- Combination of **radiant energy** and negative resistance to amplify electricity
- And many other plans for Free Energy.

MORAY GENERATOR

Moray Generator - FREE Energy Systems - How Close Are We Into Generating Free and Cheap Electricity?

Moray Generator - FREE Energy Systems - How Close Are We Into Generating Free and C...

Update articles on radiant energy according to mainstream science: Radiant Energy - Moray Generator

Next we introduce another radiant energy version, in the style of capturing energy from the sky, like Tesla and Moray.

GUILLOT, Jules: Atmospheric Electric Generator -- A powerful proven design from the 1920s

The Guillot device generated about 2.5-3 Kilowatts with antenna height of ~ 20 meters. Power depends on the total collector surface and height of the vertical antenna. The apparatus in the photo produced ~300 watts with a collector 2 meters tall.

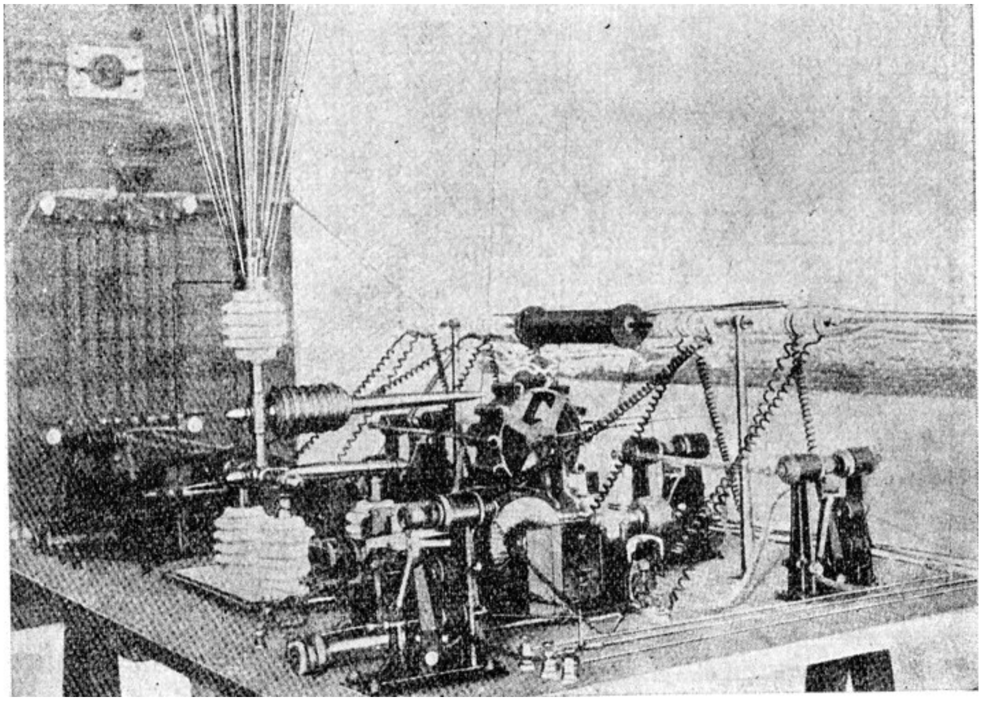

"Practical Utilization of Atmospheric Electricity"

by Gillbert Darida in *The Invention Encyclopedia,* pp. 204-207 (1930, Geo. Constantinescu, Ed.)

"The Earth has its own negative electricity, in the soil, and the atmosphere which surrounds the Earth is charged with positive electricity. The electric potential (the voltage) increases with the altitude, so we can say that the electric force is proportional to the atmospheric altitude (after Franklin, Quetelet, Lord Kelvin, Mascart, Joubert and other scientists).

The recent observations demonstrate that the air at 6000-7000 meters in altitude, is very highly charged with positive electricity, which could be explained by the friction between the external photosphere and the upper atmosphere of Earth, which rotates at a speed of more than 100,000 Kilometer per hour.

In that way, the Earth works like an electrostatic generator with electric charging by influence --- the upper atmosphere is positively charged by influence and the Earth crust obtains the negative polarity.

Between the two environments, the air and the soil, and inside the low atmosphere, in conditions of good weather, there are sbout 800 positive ions and also, 680 negative ion (and electrons) in just one square-centimeter of normal ionised air.

The Earth behaves like an huge electric armature negatively charged, which repels the electrons and attracts the positive ions. That positive ions' attraction determinates an electric current, also called "convective current". That's like an invisible continous bombardment, subject to daily and seasonal variations, which could be aproximated at **3 x 10 ^ -16 Amperes per square-centimeter,** and that is a total value of 1500 Amperes for the entire surface of the Earth.
The question is --- how does this current always maintains the same direction ? We can suppose that the natural radioactive emissions of the soil is responsable for this. We also know that radioactive emissions of the Earth works usually near the soil surface, and that explains the ionisation phenomenona inside caves.

The ionisation of the low atmosphere could also be the effect of the radioactive emissions of the Earth, especially when the X radiation works. Also, the Earth atmosphere is ionisated by the external radiation proceeding from the Sun and from the space environment, especially the action of ultraviolet radiation and the electron fascicles emitted by the Sun surface, at the temperature of 6000 deg. Celsius.

The capture of atmosphere electricity has been used in France, with aerial cables mounted on the Mont Blanc, and also in Germany --- with conductive cables carried by the captive baloons.

The atmosphere electricity collect system invented by eng. Jules Guillot is most ingenious and it relies on "the **electric siphon**" [*m.n. --- the discharging devices or spark-gap used today from the protection of aerial electric cables against the atmospheric electricity ; Jules Guillot has thought to recuperate and to use that electricity*].

His method consisted in the direct "pumping" of the atmosphere electricity using a collecting device which had two antennae and several collecting rods.

One antenna is **vertical** and it has a lot of **rods** scattered like an opened **fan**, with the tips against the **zenith**, for collecting the **negative** electricity which comes from the air ; the **horizontal** antenna is orientated against the **South** and its role is to collect the **positive** electricity.

The air electricity seems to have the double-polarity, as we can see at two electrised clouds when between them apears the electric discharge. The air could have different electric charging and the ionisation processus of atmosphere is very heterogenous. The inventor Guillot used **two separated and insulated armatures with the positive armature against the South** (more precisely, against the **Equator**...) and the **negative** armature against the **zenith**.

We can see in the scheme :

1,- the device P is the "**collector apparatus**" having **15-20 meters height** ; it also has a lot of **rods** which are scattered **fan**-like, all this on top of an **iron pillar** ; and also an **horizontal antenna** S oriented against the **South** ; the two antennae are disposed at **90 deg. angle** forming the two armatures where the electrical tension must apear. As the **height** of the pillar **increases**, the **voltage** increases more and more ;

2,- a system of **lightning-rods** (p1, p2, p3) for protection against discharges of atmospheric electricity during storms;

3,- the regulator system (**voltage regulator**, etc) R ;

4,- the electric **resistors** (Cn) for absorbing the parasite currents ;

5,- the "electric siphons" which have the role of making a magnetic field as a good environment for the transport of the ionized fluid, absorbing the electricity from the air ; the E1 and E2 are their regulator devices ;

6,- an **auxilliary continous current power supply** which send the current against an special **excitation coil**, used as starting device ; it works only when the system is starting.
This collector device works very good for lighting and heating. The device made by Guillot had
20 meters height on the total surface of **3 square-meters**. Also, J.Guillot used and electrical transformer for the industrial utilization of this "collector" --- as power supply for industrial electric engines."

French Patent # 551,882

Apparatus for Capture of Electric Currents in the Atmosphere

16 April 1923

Considering the terrestrial globe like the inductor of a dynamo, where the extraterrestrial ether is the inductor of electric currents circulating in the atmosphere; the apparatus of the invention described here selects through the device described below two perfectly distinct currents and eliminates others.

The device includes:

1, The air sensor; 2, a series of lightning rods; 3, vacuum; 4, an array of resistances; 5, vacuum; 6, witnesses.

1. An aerial sensor mounted on a pole is composed of **magnetic steel** fixed and sealed by a porcelain **insulator**, and surrounded a base by a **bronze ring** which are screwed **32 points of soft iron**, all forming antenna a pole. level beside it and isolated from the first, forming the 2nd pole, is fixed a sharp point of magnetic steel fixed in the horizontal position and directed towards the south and **movable to an angle of 45** is sealed in a porcelain insulator. This point is also circled a **ring of copper** notches figure. 1.

2. A series of 6 lightning surge arrestors at corners between each pole and the earth and the various gauges.

3. A first **regulator** form of a device similar absolutely similar to the aerial amtenna, but the two poles are superposed and opposed to the vertical point, and a **copper disk connects to ground**. Opposite the horizontal point, a **ring of tin** to which are welded tubes alternatively **16 tubes** composed each composed of **copper-lead** and **iron-lead** . The ring connects to the **ground** as shown in figure. 2.

A second regulator form of **automatic breakers** balanced on 3 poles and two poles also equally balanced.

4. A panel of **resistances** composed of **wire mesh** in bunches from different sections of **glass tubes** containing **copper** dust, **coal** and flowers of **sulfur**.

Vacuum cleaners consists of a **wooden box** on each pole, containing a **porcelain vase** in which **layers** isolated with **mica** is made up of **mercury, tin, coal, copper** and **sulfur**, all contaied in a **copper tube**.

6. The witnesses are comprised of ordinary incandescent lamps.

Resume

By the point at the zenith and the point to the south we channel two currents forming the two poles. We also protect from lightning. It regularizes the flow by a regulator and similar devices by a controlling each devise of adequate strength of the current harmful nature of these devices do not have loads. The refined current is conducted by ordinary copper wires**.**

French Patent # 565,395

Combined Apparatus for Capture of Atmospheric Electric Currents with Immediate Implementation

25 January 1924

We know that earth with its constitution, its rotation and movement in space, provides the electricity in the atmosphere. The electrical currents escape into space or largely accumulate towards the equator, as a result of the greater periphery of the globe.

In the atmosphere, there are two clearly distinct poles, i.e., clouds that can be electrically positive or negative; everything in nature shows this, otherwise it would be difficult to explain the lightning that occurs between the clouds and which are none other than contrary cloud electrical charges discharged by too close proximity.

The invention relates to a set of devices capable of capturing the atmospheric electricity.

So far, all searches made for this purpose has been to capture the atmosphere, i.e., that brought together into one system to capture two poles and opposes effect has been to destroy or cancel each other leaving the amount as the difference of the strongest to weakest.

It is therefore easy to understand why, considering the two poles of air as about equal, it is almost impossible to measure some potential with the land, the highest on the lowest remaining which can be positive or negative, copper being the strongest of one or other of these polarities.

It is on this basis that all research until now have been made, and that is why we must consider that all the time, between the two poles captured in the atmosphere and not in any mixture of systems uptake, it is possible to obtain tension and intensity, a considerable and measurable power captured separately between the two poles and remaining isolated from one another until their utilization .

In accordance with this invention, this collection is made by attraction of primary electrical currents in teh atmosphere, either positively or negatively charged, by an antenna that has two points absolutely isolated from one another, where one points to the zenith to attract negative electricity, and the other turned south, toward the equator to draws in the masses grouped in this area.

This shows and proves the existence of atmospheric electrical currents, and that it is possible to capture them with specially designed equipment primarily acting as regulators, as the persistent difficulty encountered so far in realizing their capture lies with large variations of tension in which the current atmosphere is present in space, for each of the poles, and an overload could inevitably fatal.

The patent in its present form of presentation does not cover the regulators or devices that are anticipated in the system, because they can be designed in different ways and give the same result, but covers the application and grouping of equipment operating as automatic relays with variable influence to limit current tension to that chosen for utilization.

These devices have a role associated with an overflow reservoir outside carrying excess liquid; these regulators will divert the excess voltage flows to ground.

We can still absorb these regulators having filters because they have the sole purpose of diverting the post being used , currents that are not yet known but likely that we will call for simplicity, abnormal currents. [sic...]

The patent also covers systems for regulating currents, applied to each of the poles because it is recognized that the shape and tension of the currents of the two poles are not equivalent.

The method of capture is by special antenna has directed two points in the atmosphere, as will be stated after this, with Boot ecoulement prior to the current atmosphere is also of great importance.

Finally, the current atmospheric reception is adequate even with the installation of a post, which can be expected at any location without the need for special altitude, as has been attempted in previous efforts.

We refer to the attached drawings:.

Figure 1 represents in elevation and side view an antenna that only has two points, to overcome an elevation ;

Figure 2 is the end of one of the peaks, the horizontal:

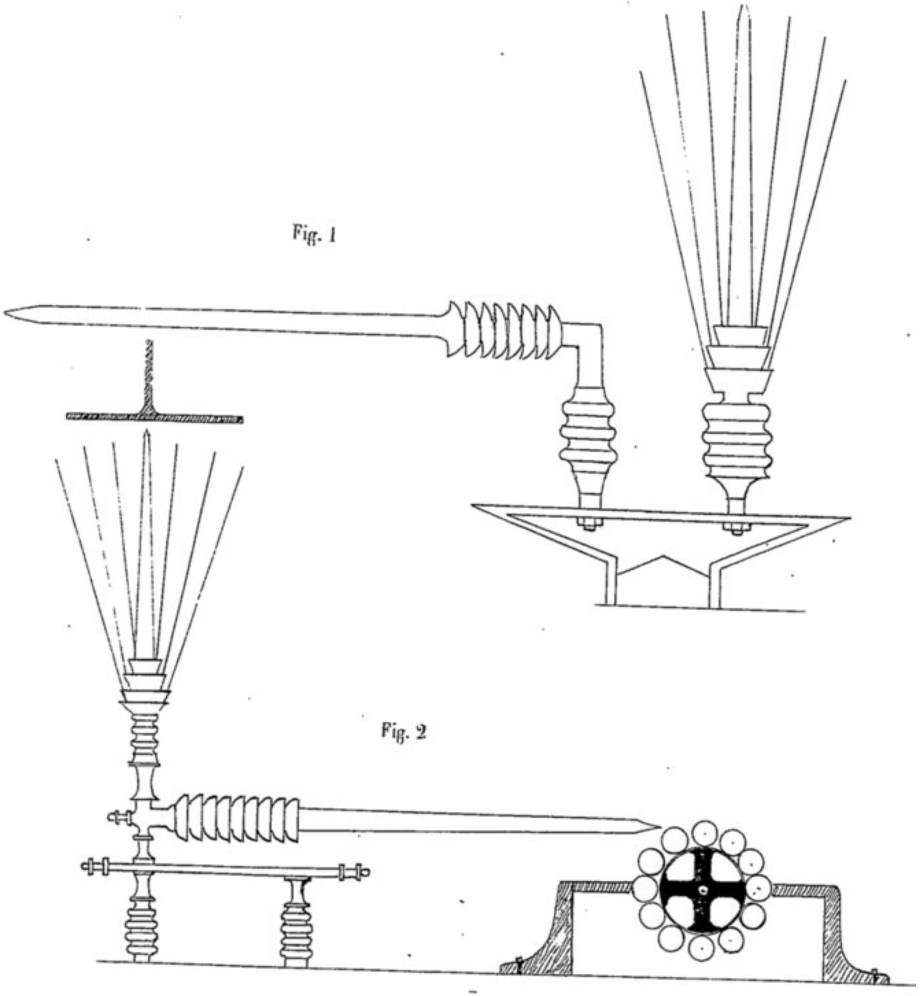

Figure 3 shows schematically the table on which are mounted various devices or accessories that act as regulators of current tensions.

Fig.3.

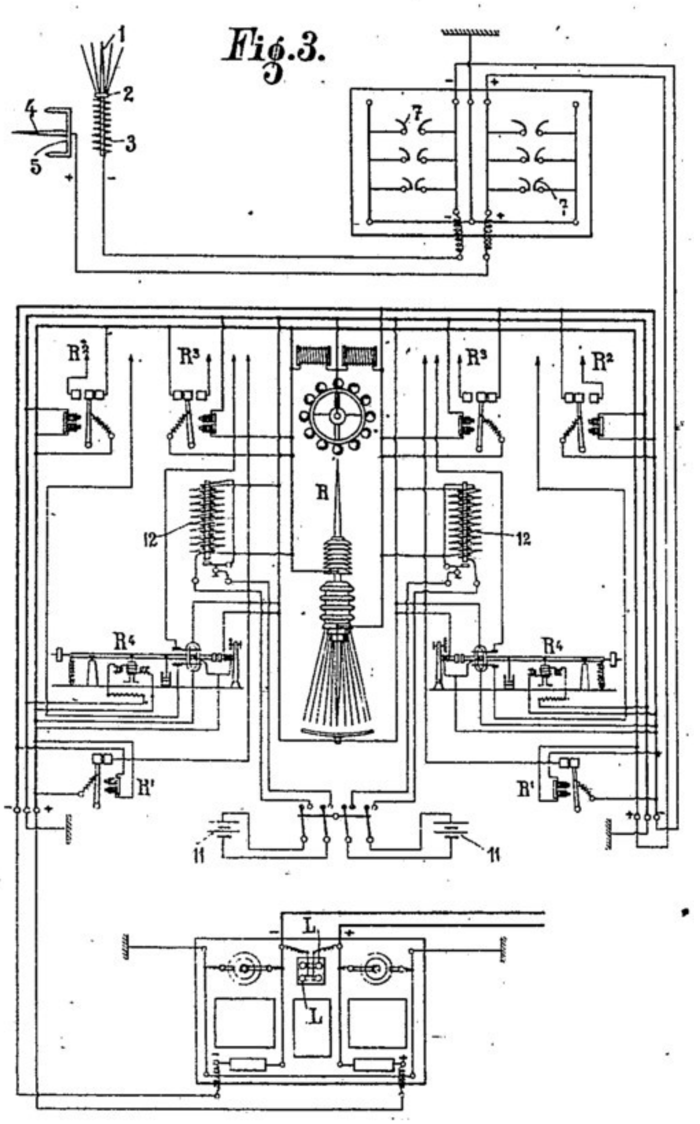
1
2
3
4
5
7
7
R²
R³
R³
R²
R
12
12
R⁴
R⁴
R¹
R¹
11
11
L
L
+
-

Figure 4 is a scheme of surge arrestors, and

Figure 5 a table of resistances.

Fig. 4.

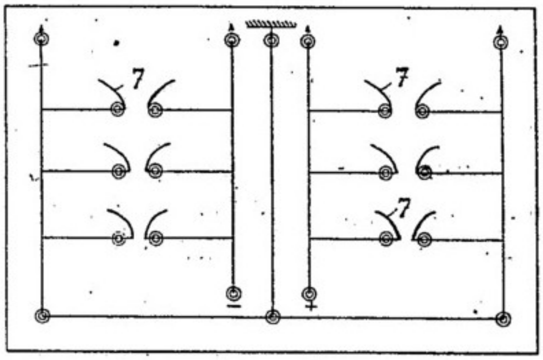

Fig. 5.

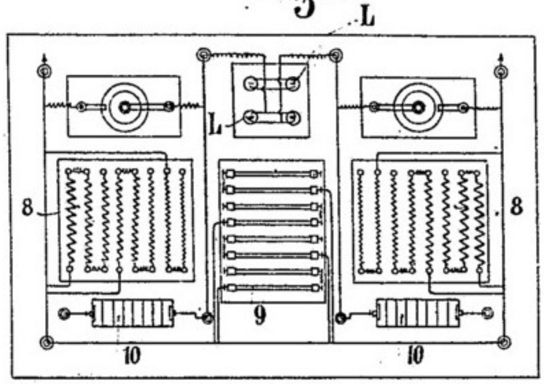

The necessary arrangements to capture atmospheric current form is as follows:

1. The **antenna** in figure 1 is formed with two peaks: the 1 is **steel**, supported by a **bronze ring** 2, surrounded by a **spiral form** 3; this point is oriented for attracting the **zenith** layers of **negative electricity** and the other edge 4 contains **copper coils** 6, and is oriented **south** to the **equator** to draw in the masses grouped in this area. To use this antenna it suffices to be raised on a mast or **pole** to a height a little higher than that of surrounding houses.

2. On the table **lightning arrestors**, contained 4, forms of 7 **copper horns** used to absorb large atmospheric discharges by certain time, as it would be dangerous to allow movement in the devices regulators.

3. On a table of **two** symmetrical systems **regulators**, Figure 3, one for each pole, to obtain and allow a regularization of large excesses in the current tensions, by automatically absorbing surges before passing over the surfacea chosen for the desired use. These regulators are represented on the drawing by the references R, R1, R2, R3, R4.

4. In the **resistance** panel in communication with regulators, these resistances are formed as **spiral coils or screens** 8, **tubes 9 of glass filled with fine coal dust , aluminum powder, sulfur and fine copper dust**. Finally, **two** special 10 **processors** and constitutes **layers** of metal forming an **absorbant** for harmful or abnormal currents.

5. As a source of energy formed **auxiliary batteries** 11 and **Ruhmkorff coils** 12 to get into the antenna sufficient attraction by means of an adjustment of intensity, which once it begins, allows the constant flow of atmospheric current in the system.

In considering Figure 3, which represents a sort of an installation scheme of the device, the reception of atmospheric electrical current will, as set out above, producing a priming of these currents. This boot is made in launching the system which is double to answer each of the poles, the current batteries 11 through 12 and reels going to the antenna.

This boot will continue until the disposal of electrical current atmosphere is evident by the one or more table lamps L of resistance shown in Figure 5.

Once the lights illuminate it is advisable to charge the batteries to boot. Natural flow and constant currents will be captured by the regulators , Figure 3; they will regularize in tension and in amperage and then directed to use in table lamps.

It has been said above that regulators R, etc., are intended and designed to automatically eliminate earth surge currents capture and abnormal currents that cannot yet be level but that these regulators could be replaced by devices with the same function.

There is therefore no need to describe in detail and especially the claims. We only specify that these regulators must be proper protection for the purposes described above.

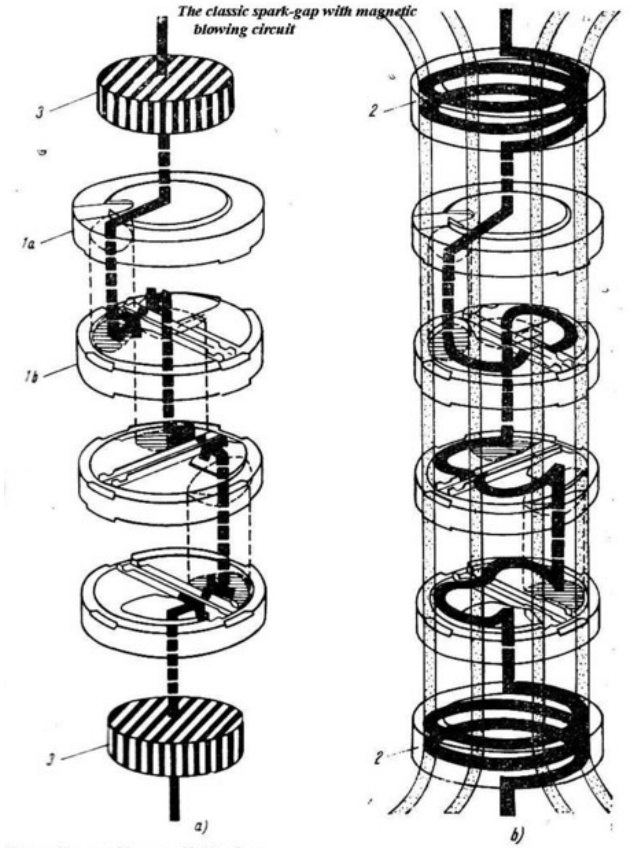

The classic spark-gap with magnetic blowing circuit

The spark-gap with magnetic blowing :
a,- the initial stage ; the impulse current send in soil through the shunt-resistences 3, because the coils 2 have an impedance too big for the impulse of 10/20 µs ;
the spark-gap 1 has two elements disc-shaped 1a and 1b, the inferior disc have one electrode and the upper disc the other electrode of the spark-gap circuit.
b,- the second stage ; the accompanying current discharge - after the impulse current ;
this accompanying current pass through the coils 2 which produce the magnetic field with the role to extend the electric arc into the slit between insulator discs 1a and 1b ; the thickness of this slit is smaller and smaller in accord with the extension of the electric arc, in this manner producing the magnifing of its electric resistance and finally, the electric arc extinction.

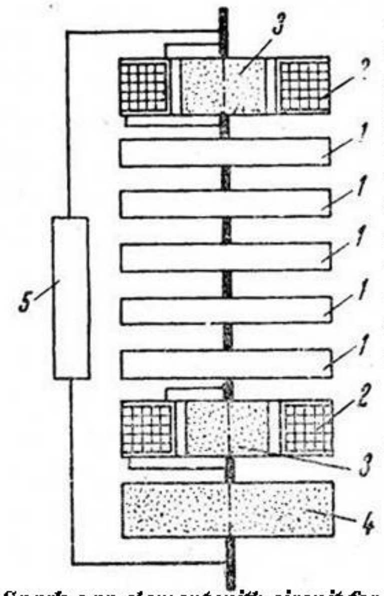

1,- the spark-gap individual devices which work in the manner to extend the electric arc of the accompanying electric current ;
2,- the coils which produce the magnetic field for the magnetic blowing ;
3,- the electric shunt-resistances in parallel connection on the blowing coils, with the role of reducing the current inside this coils
4,- the disc with unlinear resistance ;
5,- electrical resistance for the homogenous distribution of the electric field in all the elements of the spark-gap.

In Jules Guillot scheme, that kind of device could be used for leading the current.

Spark-gap element with circuit for the magnetic blowing of the electric fascicle

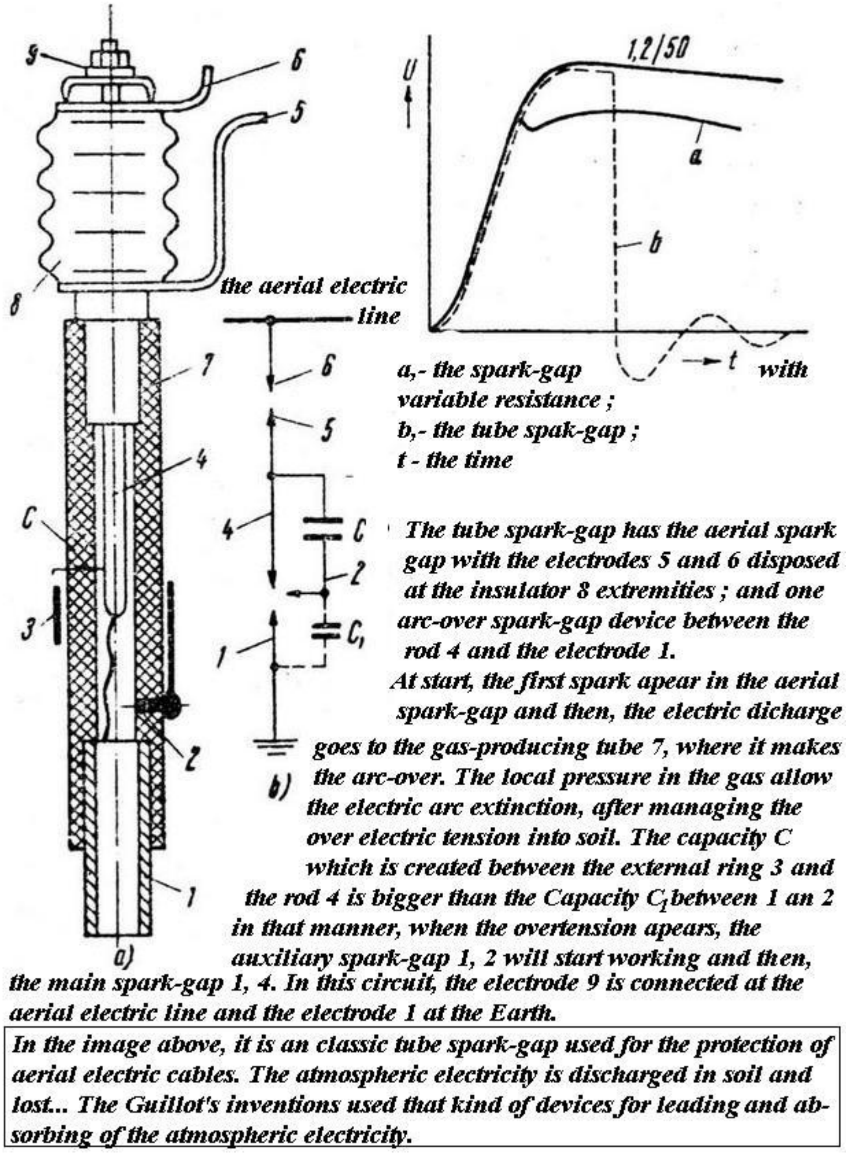

a,- the spark-gap with variable resistance ;
b,- the tube spak-gap ;
t - the time

The tube spark-gap has the aerial spark gap with the electrodes 5 and 6 disposed at the insulator 8 extremities ; and one arc-over spark-gap device between the rod 4 and the electrode 1.
At start, the first spark apear in the aerial spark-gap and then, the electric dicharge goes to the gas-producing tube 7, where it makes the arc-over. The local pressure in the gas allow the electric arc extinction, after managing the over electric tension into soil. The capacity C which is created between the external ring 3 and the rod 4 is bigger than the Capacity C_1 between 1 an 2 in that manner, when the overtension apears, the auxiliary spark-gap 1, 2 will start working and then, the main spark-gap 1, 4. In this circuit, the electrode 9 is connected at the aerial electric line and the electrode 1 at the Earth.

In the image above, it is an classic tube spark-gap used for the protection of aerial electric cables. The atmospheric electricity is discharged in soil and lost... The Guillot's inventions used that kind of devices for leading and absorbing of the atmospheric electricity.

DARIDA, Gillbert : "Practical Utilization of Atmospheric Electricity";
(in *The Invention Encyclopedia,* pp. 204-207 ; 1930, Geo. Constantinescu, Ed.) --

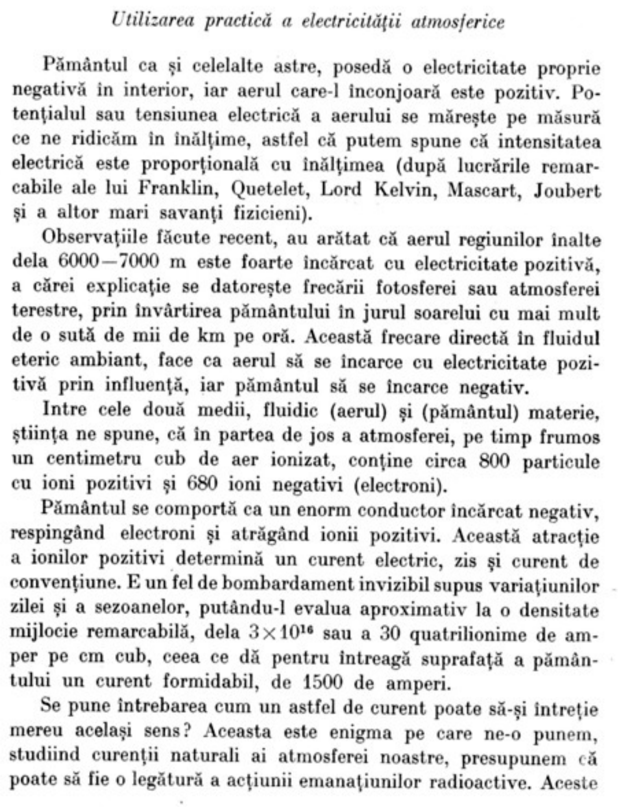

Pământul ca și celelalte astre, posedă o electricitate proprie negativă în interior, iar aerul care-l înconjoară este pozitiv. Potențialul sau tensiunea electrică a aerului se mărește pe măsură ce ne ridicăm în înălțime, astfel că putem spune că intensitatea electrică este proporțională cu înălțimea (după lucrările remarcabile ale lui Franklin, Quetelet, Lord Kelvin, Mascart, Joubert și a altor mari savanți fizicieni).

Observațiile făcute recent, au arătat că aerul regiunilor înalte dela 6000—7000 m este foarte încărcat cu electricitate pozitivă, a cărei explicație se datorește frecării fotosferei sau atmosferei terestre, prin învârtirea pământului în jurul soarelui cu mai mult de o sută de mii de km pe oră. Această frecare directă în fluidul eteric ambiant, face ca aerul să se încarce cu electricitate pozitivă prin influență, iar pământul să se încarce negativ.

Intre cele două medii, fluidic (aerul) și (pământul) materie, știința ne spune, că în partea de jos a atmosferei, pe timp frumos un centimetru cub de aer ionizat, conține circa 800 particule cu ioni pozitivi și 680 ioni negativi (electroni).

Pământul se comportă ca un enorm conductor încărcat negativ, respingând electroni și atrăgând ionii pozitivi. Această atracție a ionilor pozitivi determină un curent electric, zis și curent de convențiune. E un fel de bombardament invizibil supus variațiunilor zilei și a sezoanelor, putându-l evalua aproximativ la o densitate mijlocie remarcabilă, dela 3×10^{16} sau a 30 quatrilionime de amper pe cm cub, ceea ce dă pentru întreagă suprafață a pământului un curent formidabil, de 1500 de amperi.

Se pune întrebarea cum un astfel de curent poate să-și întreție mereu același sens? Aceasta este enigma pe care ne-o punem, studiind curenții naturali ai atmosferei noastre, presupunem că poate să fie o legătură a acțiunii emanațiunilor radioactive. Aceste

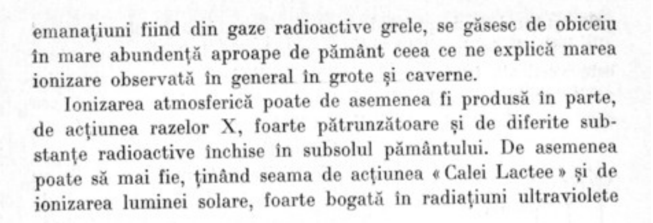

emanațiuni fiind din gaze radioactive grele, se găsesc de obiceiu în mare abundență aproape de pământ ceea ce ne explică marea ionizare observată în general în grote și caverne.

Ionizarea atmosferică poate de asemenea fi produsă în parte, de acțiunea razelor X, foarte pătrunzătoare și de diferite substanțe radioactive închise în subsolul pământului. De asemenea poate să mai fie, ținând seama de acțiunea «Calei Lactee» și de ionizarea luminei solare, foarte bogată în radiațiuni ultraviolete

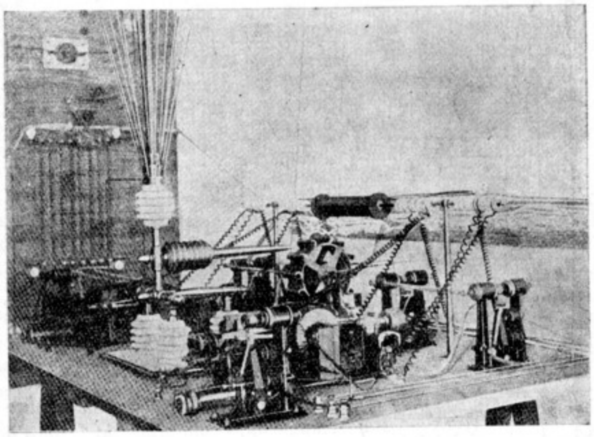

Fig. 108

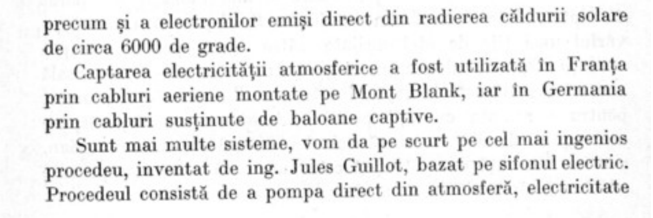

precum și a electronilor emiși direct din radierea căldurii solare de circa 6000 de grade.

Captarea electricității atmosferice a fost utilizată în Franța prin cabluri aeriene montate pe Mont Blank, iar în Germania prin cabluri susținute de baloane captive.

Sunt mai multe sisteme, vom da pe scurt pe cel mai ingenios procedeu, inventat de ing. Jules Guillot, bazat pe sifonul electric. Procedeul consistă de a pompa direct din atmosferă, electricitate

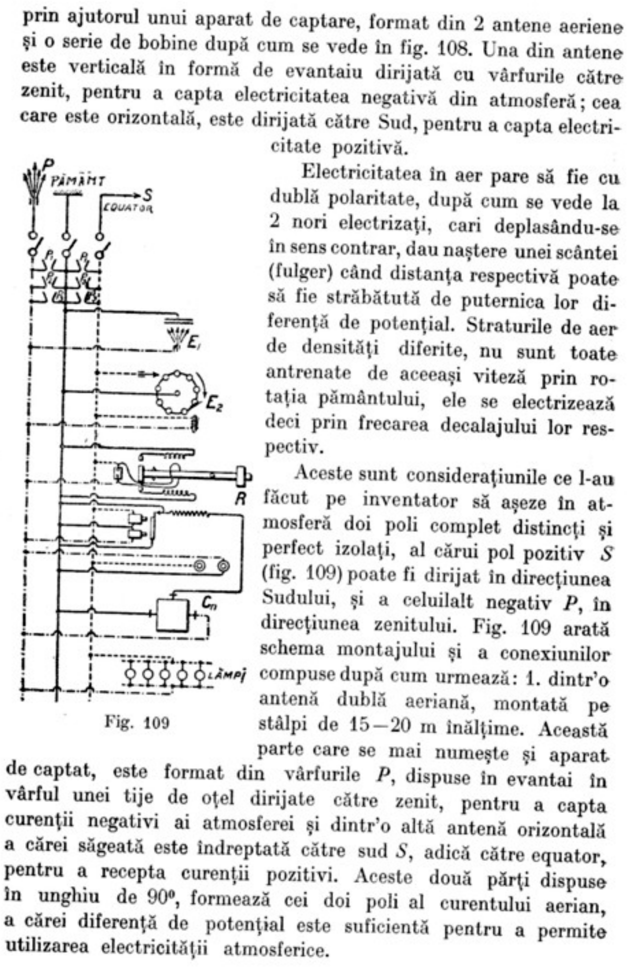

prin ajutorul unui aparat de captare, format din 2 antene aeriene şi o serie de bobine după cum se vede în fig. 108. Una din antene este verticală în formă de evantaiu dirijată cu vârfurile către zenit, pentru a capta electricitatea negativă din atmosferă; cea care este orizontală, este dirijată către Sud, pentru a capta electricitate pozitivă.

Electricitatea în aer pare să fie cu dublă polaritate, după cum se vede la 2 nori electrizaţi, cari deplasându-se în sens contrar, dau naştere unei scântei (fulger) când distanţa respectivă poate să fie străbătută de puternica lor diferenţă de potenţial. Straturile de aer de densităţi diferite, nu sunt toate antrenate de aceeaşi viteză prin rotaţia pământului, ele se electrizează deci prin frecarea decalajului lor respectiv.

Aceste sunt consideraţiunile ce l-au făcut pe inventator să aşeze în atmosferă doi poli complet distincţi şi perfect izolaţi, al cărui pol pozitiv *S* (fig. 109) poate fi dirijat în direcţiunea Sudului, şi a celuilalt negativ *P*, în direcţiunea zenitului. Fig. 109 arată schema montajului şi a conexiunilor compuse după cum urmează: 1. dintr'o antenă dublă aeriană, montată pe stâlpi de 15—20 m înălţime. Această parte care se mai numeşte şi aparat de captat, este format din vârfurile *P*, dispuse în evantai în vârful unei tije de oţel dirijate către zenit, pentru a capta curenţii negativi ai atmosferei şi dintr'o altă antenă orizontală a cărei săgeată este îndreptată către sud *S*, adică către equator, pentru a recepta curenţii pozitivi. Aceste două părţi dispuse în unghiu de 90°, formează cei doi poli al curentului aerian, a cărei diferenţă de potenţial este suficientă pentru a permite utilizarea electricităţii atmosferice.

Fig. 109

French Patent # 551,882

Appareil Capteur de Courants Electriques dans L'Atmosphere

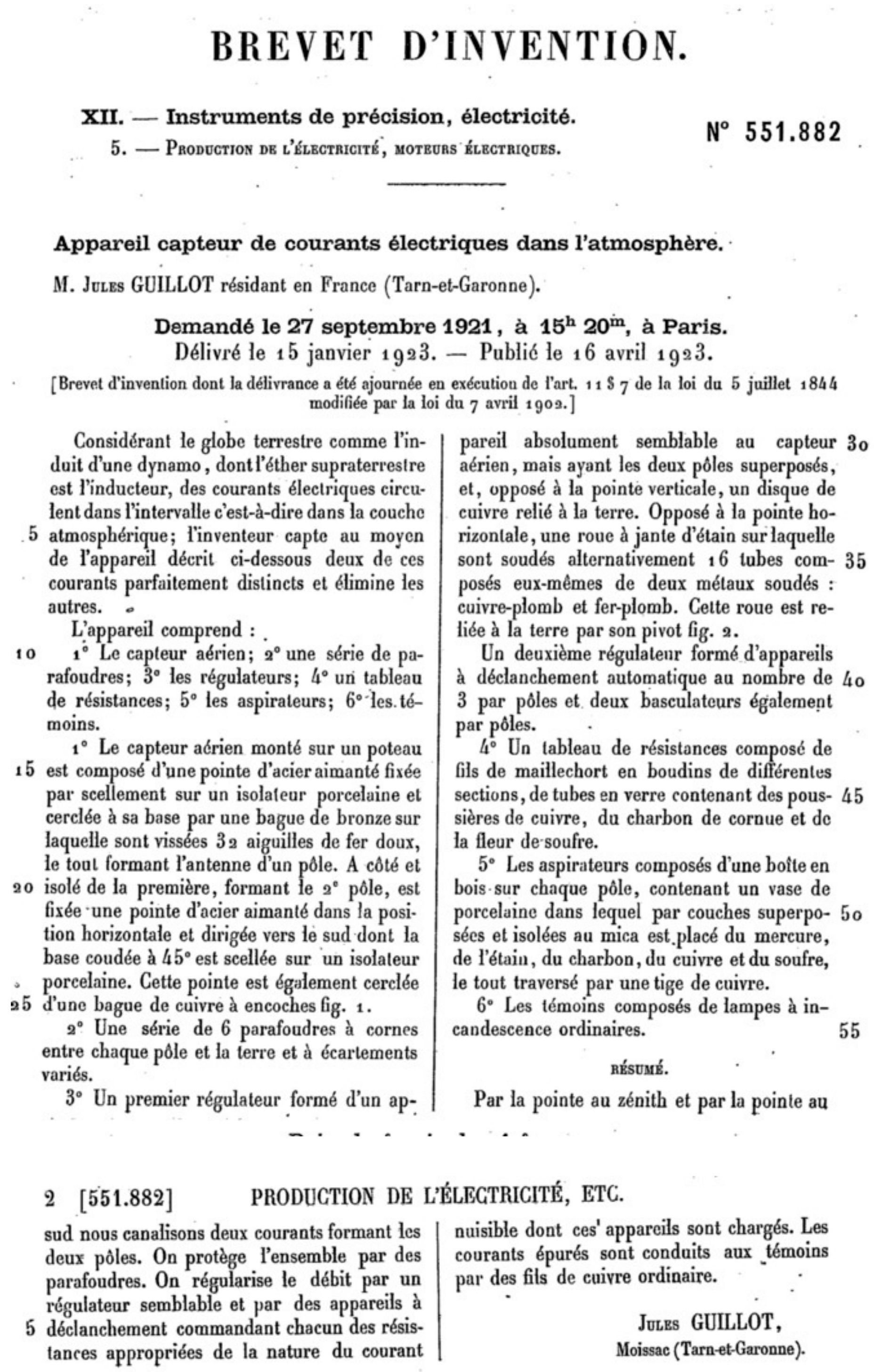

BREVET D'INVENTION.

XII. — Instruments de précision, électricité.

5. — PRODUCTION DE L'ÉLECTRICITÉ, MOTEURS ÉLECTRIQUES.

N° 551.882

Appareil capteur de courants électriques dans l'atmosphère.

M. JULES GUILLOT résidant en France (Tarn-et-Garonne).

Demandé le 27 septembre 1921, à 15ʰ 20ᵐ, à Paris.
Délivré le 15 janvier 1923. — Publié le 16 avril 1923.
[Brevet d'invention dont la délivrance a été ajournée en exécution de l'art. 11 § 7 de la loi du 5 juillet 1844 modifiée par la loi du 7 avril 1902.]

Considérant le globe terrestre comme l'induit d'une dynamo, dont l'éther supraterrestre est l'inducteur, des courants électriques circulent dans l'intervalle c'est-à-dire dans la couche atmosphérique; l'inventeur capte au moyen de l'appareil décrit ci-dessous deux de ces courants parfaitement distincts et élimine les autres.

L'appareil comprend :

1° Le capteur aérien; 2° une série de parafoudres; 3° les régulateurs; 4° un tableau de résistances; 5° les aspirateurs; 6° les témoins.

1° Le capteur aérien monté sur un poteau est composé d'une pointe d'acier aimanté fixée par scellement sur un isolateur porcelaine et cerclée à sa base par une bague de bronze sur laquelle sont vissées 32 aiguilles de fer doux, le tout formant l'antenne d'un pôle. A côté et isolé de la première, formant le 2ᵉ pôle, est fixée une pointe d'acier aimanté dans la position horizontale et dirigée vers le sud dont la base coudée à 45° est scellée sur un isolateur porcelaine. Cette pointe est également cerclée d'une bague de cuivre à encoches fig. 1.

2° Une série de 6 parafoudres à cornes entre chaque pôle et la terre et à écartements variés.

3° Un premier régulateur formé d'un appareil absolument semblable au capteur aérien, mais ayant les deux pôles superposés, et, opposé à la pointe verticale, un disque de cuivre relié à la terre. Opposé à la pointe horizontale, une roue à jante d'étain sur laquelle sont soudés alternativement 16 tubes composés eux-mêmes de deux métaux soudés : cuivre-plomb et fer-plomb. Cette roue est reliée à la terre par son pivot fig. 2.

Un deuxième régulateur formé d'appareils à déclanchement automatique au nombre de 3 par pôles et deux basculateurs également par pôles.

4° Un tableau de résistances composé de fils de maillechort en boudins de différentes sections, de tubes en verre contenant des poussières de cuivre, du charbon de cornue et de la fleur de soufre.

5° Les aspirateurs composés d'une boîte en bois sur chaque pôle, contenant un vase de porcelaine dans lequel par couches superposées et isolées au mica est placé du mercure, de l'étain, du charbon, du cuivre et du soufre, le tout traversé par une tige de cuivre.

6° Les témoins composés de lampes à incandescence ordinaires.

RÉSUMÉ.

Par la pointe au zénith et par la pointe au sud nous canalisons deux courants formant les deux pôles. On protège l'ensemble par des parafoudres. On régularise le débit par un régulateur semblable et par des appareils à déclanchement commandant chacun des résistances appropriées de la nature du courant nuisible dont ces' appareils sont chargés. Les courants épurés sont conduits aux témoins par des fils de cuivre ordinaire.

JULES GUILLOT,
Moissac (Tarn-et-Garonne).

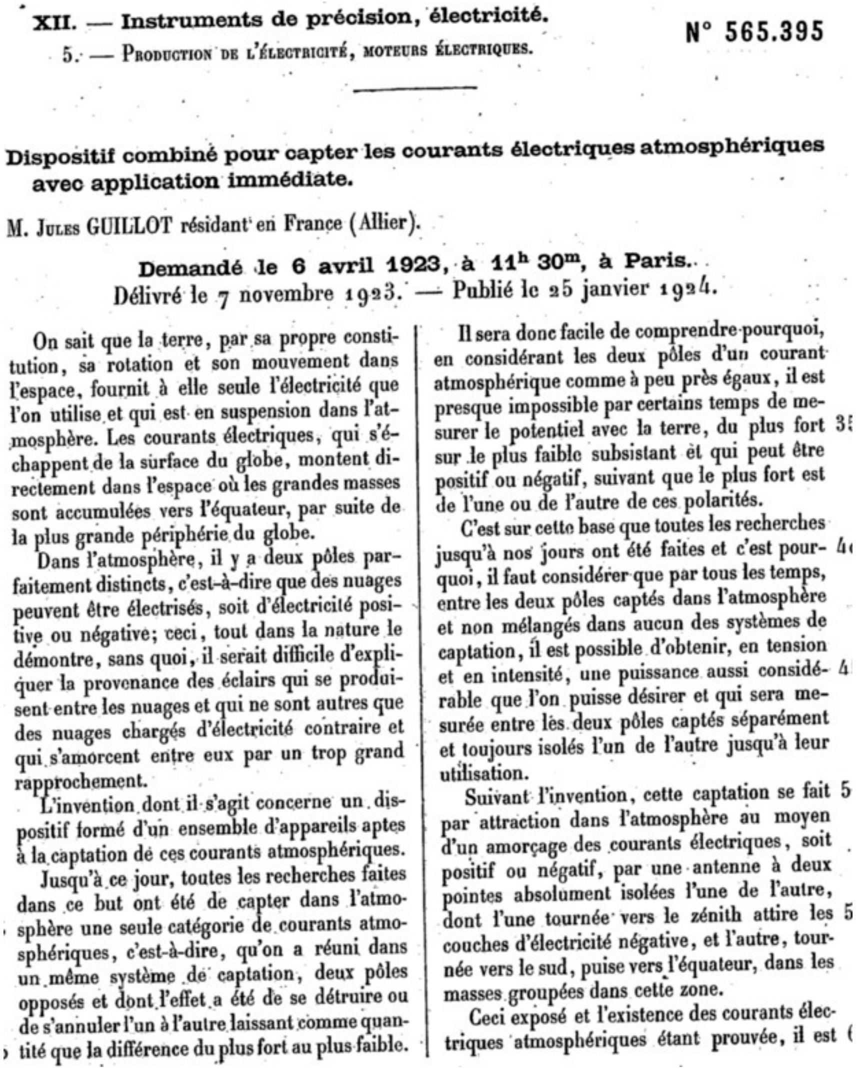

XII. — Instruments de précision, électricité.

5. — Production de l'électricité, moteurs électriques.

N° 565.395

Dispositif combiné pour capter les courants électriques atmosphériques avec application immédiate.

M. Jules GUILLOT résidant en France (Allier).

Demandé le 6 avril 1923, à 11^{h} 30^{m}, à Paris.
Délivré le 7 novembre 1923. — Publié le 25 janvier 1924.

On sait que la terre, par sa propre constitution, sa rotation et son mouvement dans l'espace, fournit à elle seule l'électricité que l'on utilise et qui est en suspension dans l'atmosphère. Les courants électriques, qui s'échappent de la surface du globe, montent directement dans l'espace où les grandes masses sont accumulées vers l'équateur, par suite de la plus grande périphérie du globe.

Dans l'atmosphère, il y a deux pôles parfaitement distincts, c'est-à-dire que des nuages peuvent être électrisés, soit d'électricité positive ou négative; ceci, tout dans la nature le démontre, sans quoi, il serait difficile d'expliquer la provenance des éclairs qui se produisent entre les nuages et qui ne sont autres que des nuages chargés d'électricité contraire et qui s'amorcent entre eux par un trop grand rapprochement.

L'invention dont il s'agit concerne un dispositif formé d'un ensemble d'appareils aptes à la captation de ces courants atmosphériques.

Jusqu'à ce jour, toutes les recherches faites dans ce but ont été de capter dans l'atmosphère une seule catégorie de courants atmosphériques, c'est-à-dire, qu'on a réuni dans un même système de captation, deux pôles opposés et dont l'effet a été de se détruire ou de s'annuler l'un à l'autre laissant comme quantité que la différence du plus fort au plus faible.

Il sera donc facile de comprendre pourquoi, en considérant les deux pôles d'un courant atmosphérique comme à peu près égaux, il est presque impossible par certains temps de mesurer le potentiel avec la terre, du plus fort sur le plus faible subsistant et qui peut être positif ou négatif, suivant que le plus fort est de l'une ou de l'autre de ces polarités.

C'est sur cette base que toutes les recherches jusqu'à nos jours ont été faites et c'est pourquoi, il faut considérer que par tous les temps, entre les deux pôles captés dans l'atmosphère et non mélangés dans aucun des systèmes de captation, il est possible d'obtenir, en tension et en intensité, une puissance aussi considérable que l'on puisse désirer et qui sera mesurée entre les deux pôles captés séparément et toujours isolés l'un de l'autre jusqu'à leur utilisation.

Suivant l'invention, cette captation se fait par attraction dans l'atmosphère au moyen d'un amorçage des courants électriques, soit positif ou négatif, par une antenne à deux pointes absolument isolées l'une de l'autre, dont l'une tournée vers le zénith attire les couches d'électricité négative, et l'autre, tournée vers le sud, puise vers l'équateur, dans les masses groupées dans cette zone.

Ceci exposé et l'existence des courants électriques atmosphériques étant prouvée, il est

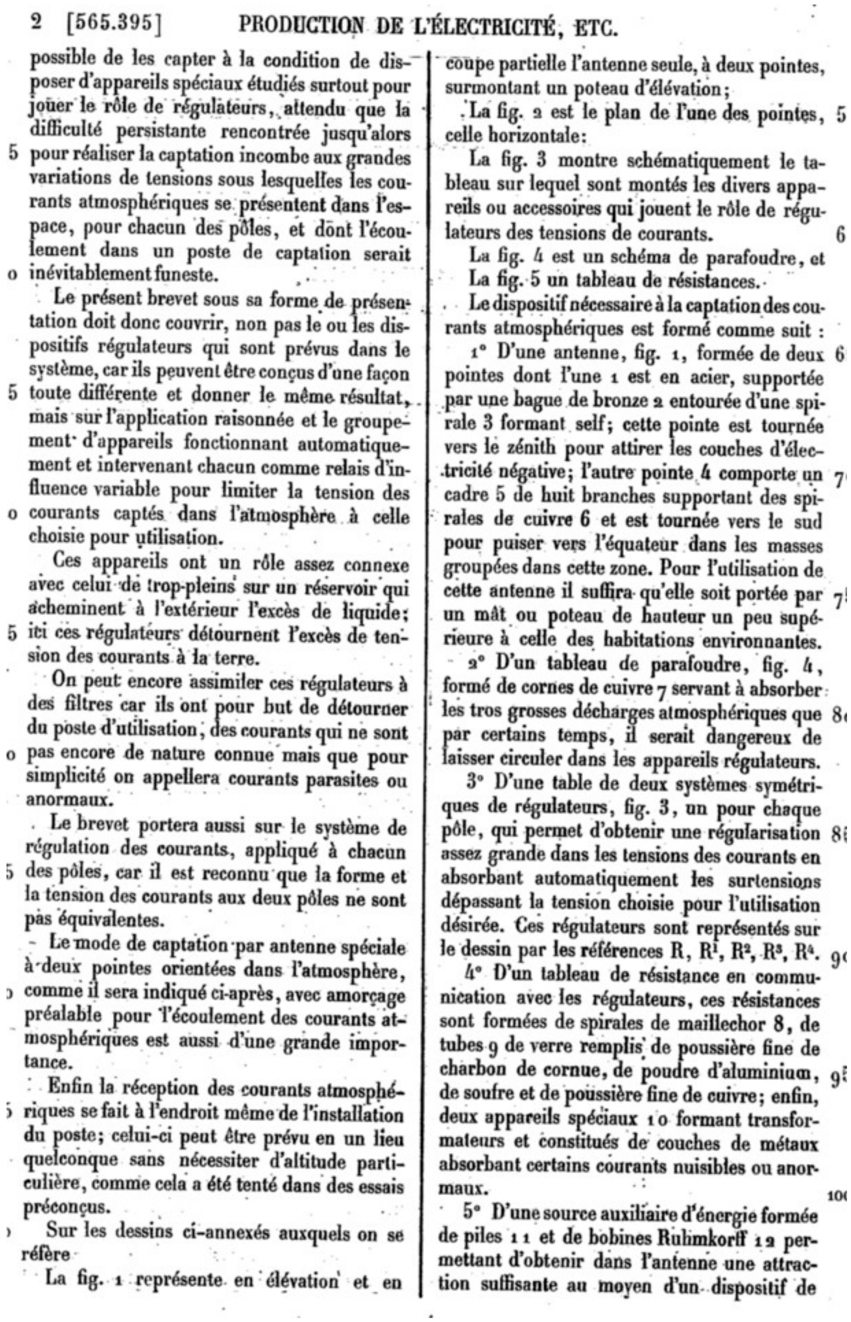

possible de les capter à la condition de disposer d'appareils spéciaux étudiés surtout pour jouer le rôle de régulateurs, attendu que la difficulté persistante rencontrée jusqu'alors pour réaliser la captation incombe aux grandes variations de tensions sous lesquelles les courants atmosphériques se présentent dans l'espace, pour chacun des pôles, et dont l'écoulement dans un poste de captation serait inévitablement funeste.

Le présent brevet sous sa forme de présentation doit donc couvrir, non pas le ou les dispositifs régulateurs qui sont prévus dans le système, car ils peuvent être conçus d'une façon toute différente et donner le même résultat, mais sur l'application raisonnée et le groupement d'appareils fonctionnant automatiquement et intervenant chacun comme relais d'influence variable pour limiter la tension des courants captés dans l'atmosphère à celle choisie pour utilisation.

Ces appareils ont un rôle assez connexe avec celui de trop-pleins sur un réservoir qui acheminent à l'extérieur l'excès de liquide; ici ces régulateurs détournent l'excès de tension des courants à la terre.

On peut encore assimiler ces régulateurs à des filtres car ils ont pour but de détourner du poste d'utilisation, des courants qui ne sont pas encore de nature connue mais que pour simplicité on appellera courants parasites ou anormaux.

Le brevet portera aussi sur le système de régulation des courants, appliqué à chacun des pôles, car il est reconnu que la forme et la tension des courants aux deux pôles ne sont pas équivalentes.

Le mode de captation par antenne spéciale à deux pointes orientées dans l'atmosphère, comme il sera indiqué ci-après, avec amorçage préalable pour l'écoulement des courants atmosphériques est aussi d'une grande importance.

Enfin la réception des courants atmosphériques se fait à l'endroit même de l'installation du poste; celui-ci peut être prévu en un lieu quelconque sans nécessiter d'altitude particulière, comme cela a été tenté dans des essais préconçus.

Sur les dessins ci-annexés auxquels on se réfère

La fig. 1 représente en élévation et en coupe partielle l'antenne seule, à deux pointes, surmontant un poteau d'élévation;

La fig. 2 est le plan de l'une des pointes, celle horizontale:

La fig. 3 montre schématiquement le tableau sur lequel sont montés les divers appareils ou accessoires qui jouent le rôle de régulateurs des tensions de courants.

La fig. 4 est un schéma de parafoudre, et

La fig. 5 un tableau de résistances.

Le dispositif nécessaire à la captation des courants atmosphériques est formé comme suit :

1° D'une antenne, fig. 1, formée de deux pointes dont l'une 1 est en acier, supportée par une bague de bronze 2 entourée d'une spirale 3 formant self; cette pointe est tournée vers le zénith pour attirer les couches d'électricité négative; l'autre pointe 4 comporte un cadre 5 de huit branches supportant des spirales de cuivre 6 et est tournée vers le sud pour puiser vers l'équateur dans les masses groupées dans cette zone. Pour l'utilisation de cette antenne il suffira qu'elle soit portée par un mât ou poteau de hauteur un peu supérieure à celle des habitations environnantes.

2° D'un tableau de parafoudre, fig. 4, formé de cornes de cuivre 7 servant à absorber les tros grosses décharges atmosphériques que par certains temps, il serait dangereux de laisser circuler dans les appareils régulateurs.

3° D'une table de deux systèmes symétriques de régulateurs, fig. 3, un pour chaque pôle, qui permet d'obtenir une régularisation assez grande dans les tensions des courants en absorbant automatiquement les surtensions dépassant la tension choisie pour l'utilisation désirée. Ces régulateurs sont représentés sur le dessin par les références R, R^1, R^2, R^3, R^4.

4° D'un tableau de résistance en communication avec les régulateurs, ces résistances sont formées de spirales de maillechor 8, de tubes 9 de verre remplis de poussière fine de charbon de cornue, de poudre d'aluminium, de soufre et de poussière fine de cuivre; enfin, deux appareils spéciaux 10 formant transformateurs et constitués de couches de métaux absorbant certains courants nuisibles ou anormaux.

5° D'une source auxiliaire d'énergie formée de piles 11 et de bobines Ruhmkorff 12 permettant d'obtenir dans l'antenne une attraction suffisante au moyen d'un dispositif de

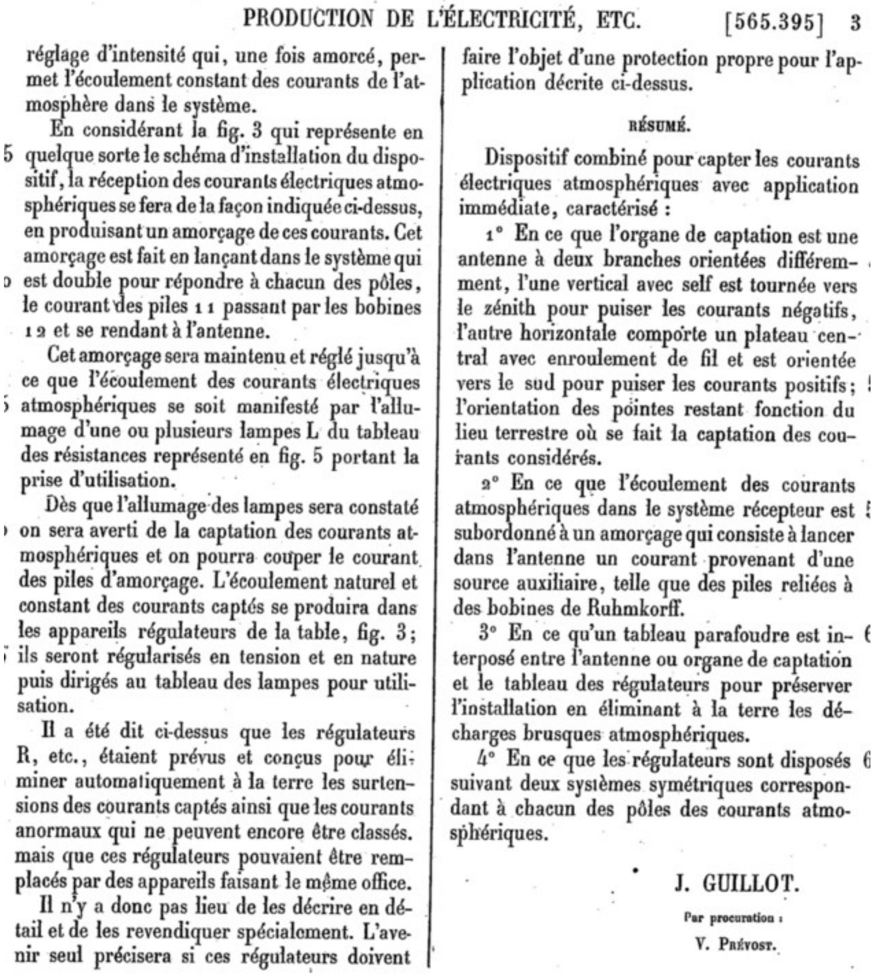

réglage d'intensité qui, une fois amorcé, permet l'écoulement constant des courants de l'atmosphère dans le système.

En considérant la fig. 3 qui représente en quelque sorte le schéma d'installation du dispositif, la réception des courants électriques atmosphériques se fera de la façon indiquée ci-dessus, en produisant un amorçage de ces courants. Cet amorçage est fait en lançant dans le système qui est double pour répondre à chacun des pôles, le courant des piles 11 passant par les bobines 12 et se rendant à l'antenne.

Cet amorçage sera maintenu et réglé jusqu'à ce que l'écoulement des courants électriques atmosphériques se soit manifesté par l'allumage d'une ou plusieurs lampes L du tableau des résistances représenté en fig. 5 portant la prise d'utilisation.

Dès que l'allumage des lampes sera constaté on sera averti de la captation des courants atmosphériques et on pourra couper le courant des piles d'amorçage. L'écoulement naturel et constant des courants captés se produira dans les appareils régulateurs de la table, fig. 3; ils seront régularisés en tension et en nature puis dirigés au tableau des lampes pour utilisation.

Il a été dit ci-dessus que les régulateurs R, etc., étaient prévus et conçus pour éliminer automatiquement à la terre les surtensions des courants captés ainsi que les courants anormaux qui ne peuvent encore être classés. mais que ces régulateurs pouvaient être remplacés par des appareils faisant le même office.

Il n'y a donc pas lieu de les décrire en détail et de les revendiquer spécialement. L'avenir seul précisera si ces régulateurs doivent faire l'objet d'une protection propre pour l'application décrite ci-dessus.

RÉSUMÉ.

Dispositif combiné pour capter les courants électriques atmosphériques avec application immédiate, caractérisé :

1° En ce que l'organe de captation est une antenne à deux branches orientées différemment, l'une vertical avec self est tournée vers le zénith pour puiser les courants négatifs, l'autre horizontale comporte un plateau central avec enroulement de fil et est orientée vers le sud pour puiser les courants positifs; l'orientation des pointes restant fonction du lieu terrestre où se fait la captation des courants considérés.

2° En ce que l'écoulement des courants atmosphériques dans le système récepteur est subordonné à un amorçage qui consiste à lancer dans l'antenne un courant provenant d'une source auxiliaire, telle que des piles reliées à des bobines de Ruhmkorff.

3° En ce qu'un tableau parafoudre est interposé entre l'antenne ou organe de captation et le tableau des régulateurs pour préserver l'installation en éliminant à la terre les décharges brusques atmosphériques.

4° En ce que les régulateurs sont disposés suivant deux systèmes symétriques correspondant à chacun des pôles des courants atmosphériques.

J. GUILLOT.

Par procuration :

V. Prévost.

CONTACT OVERUNITY ELECTRICITY:

- Address: 9981, Overunity, Tartaria, Colorado, United States
- Phone: +1 303 779 8899
- Email: overunity.electricity.blogspot@gmail.com
- News & FAQs about Energy
- Generates Energy-On-Demand: Easy Power Plan Will Change Our World Forever
- Download the guide to learn how to build your own DIY power generator: Easy DIY Power Plan
- Tesla's radio technology and its application in electricity generation: Radiant Energy

POST A COMMENT

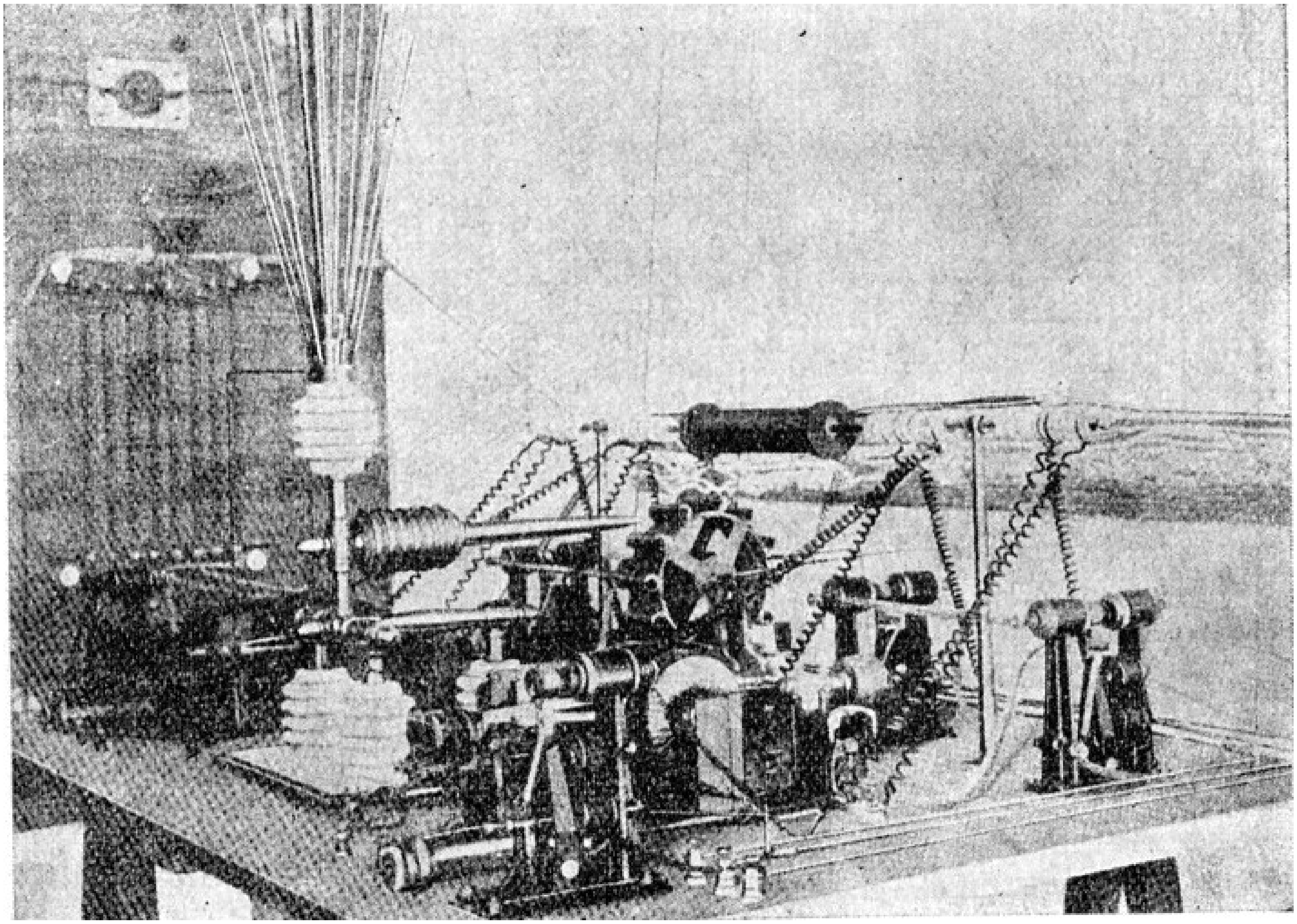

Shared by L. Cozma (Thanks !) --

The Guillot device generated about 2.5-3 Kilowatts with antenna height of ~ 20 meters. Power depends on the total collector surface and height of the vertical antenna. The apparatus in the photo produced ~300 watts with a collector 2 meters tall.

"Practical Utilization of Atmospheric Electricity"

by

Gillbert Darida

in *The Invention Encyclopedia,* pp. 204-207 (1930, Geo. Constantinescu, Ed.)

"The Earth has its own negative electricity, in the soil, and the atmosphere which surrounds the Earth is charged with positive electricity. The electric potential (the voltage) increases with the altitude, so we can say that the electric force is proportional to the atmospheric altitude (after Franklin, Quetelet, Lord Kelvin, Mascart, Joubert and other scientists).

The recent observations demonstrate that the air at 6000-7000 meters in altitude, is very highly charged with positive electricity, which could be explained by the friction between the external photosphere and the upper atmosphere of Earth, which rotates at a speed of more than 100,000 Kilometer per hour.

In that way, the Earth works like an electrostatic generator with electric charging by influence --- the upper atmosphere is positively charged by influence and the Earth crust obtains the negative polarity.

Between the two environments, the air and the soil, and inside the low atmosphere, in conditions of good weather, there are sbout 800 positive ions and also, 680 negative ion (and electrons) in just one square-centimeter of normal ionised air.

The Earth behaves like an huge electric armature negatively charged, which repels the electrons and attracts the positive ions. That positive ions' attraction determinates an electric current, also called "convective current". That's like an invisible continous bombardment, subject to daily and seasonal variations, which could be aproximated at **3 x 10 ^ -16 Amperes per square-centimeter,** and that is a total value of 1500 Amperes for the entire surface of the Earth.
The question is --- how does this current always maintains the same direction ? We can suppose that the natural radioactive emissions of the soil is responsable for this. We also know that radioactive emissions of the Earth works usually near the soil surface, and that explains the ionisation phenomenona inside caves.

The ionisation of the low atmosphere could also be the effect of the radioactive emissions of the Earth, especially when the X radiation works. Also, the Earth atmosphere is ionisated by the external radiation proceeding from the Sun and from the space environment, especially the action of ultraviolet radiation and the electron fascicles emitted by the Sun surface, at the temperature of 6000 deg. Celsius.

The capture of atmosphere electricity has been used in France, with aerial cables mounted on the Mont Blanc, and also in Germany --- with conductive cables carried by the captive baloons.

The atmosphere electricity collect system invented by eng. Jules Guillot is most ingenious and it relies on "the **electric siphon**" [*m.n. --- the discharging devices or spark-gap used today from the protection of aerial electric cables against the atmospheric electricity ; Jules Guillot has thought to recuperate and to use that electricity*].

His method consisted in the direct "pumping" of the atmosphere electricity using a collecting device which had two antennae and several collecting rods.

One antenna is **vertical** and it has a lot of **rods** scattered like an opened **fan**, with the tips against the **zenith**, for collecting the **negative** electricity which comes from the air ; the **horizontal** antenna is orientated against the **South** and its role is to collect the **positive** electricity.

The air electricity seems to have the double-polarity, as we can see at two electrised clouds when between them apears the electric discharge. The air could have different electric charging and the ionisation processus of atmosphere is very heterogenous. The inventor Guillot used **two separated and insulated armatures with the positive armature against the South** (more precisely, against the **Equator**...) and the **negative** armature against the **zenith**.

We can see in the scheme :

1,- the device P is the "**collector apparatus**" having **15-20 meters height** ; it also has a lot of **rods** which are scattered **fan**-like, all this on top of an **iron pillar** ; and also an **horizontal antenna** S oriented against the **South** ; the two antennae are disposed at **90 deg. angle** forming the two armatures where the electrical tension must apear. As the **height** of the pillar **increases**, the **voltage** increases more and more ;

2,- a system of **lightning-rods** (p1, p2, p3) for protection against discharges of atmospheric electricity during storms;

3,- the regulator system (**voltage regulator**, etc) R ;

4,- the electric **resistors** (Cn) for absorbing the parasite currents ;

5,- the "electric siphons" which have the role of making a magnetic field as a good environment for the transport of the ionized fluid, absorbing the electricity from the air ; the E1 and E2 are their regulator devices ;

6,- an **auxilliary continous current power supply** which send the current against an special **excitation coil**, used as starting device ; it works only when the system is starting.
This collector device works very good for lighting and heating. The device made by Guillot had **20 meters height** on the total surface of **3 square-meters**. Also, J.Guillot used and electrical transformer for the industrial utilization of this "collector" --- as power supply for industrial electric engines."

French Patent # 551,882

Apparatus for Capture of Electric Currents in the Atmosphere

16 April 1923

Considering the terrestrial globe like the inductor of a dynamo, where the extraterrestrial ether is the inductor of electric currents circulating in the atmosphere; the apparatus of the invention described here selects through the device described below two perfectly distinct currents and eliminates others.

The device includes:

1, The air sensor; 2, a series of lightning rods; 3, vacuum; 4, an array of resistances; 5, vacuum; 6, witnesses.

1. An aerial sensor mounted on a pole is composed of **magnetic steel** fixed and sealed by a porcelain **insulator**, and surrounded a base by a **bronze ring** which are screwed **32 points of soft iron**, all forming antenna a pole. level beside it and isolated from the first, forming the 2nd pole, is fixed a sharp point of magnetic steel fixed in the horizontal position and directed towards the south and **movable to an angle of 45** is sealed in a porcelain insulator. This point is also circled a **ring of copper** notches figure. 1.

2. A series of 6 lightning surge arrestors at corners between each pole and the earth and the various gauges.

3. A first **regulator** form of a device similar absolutely similar to the aerial amtenna, but the two poles are superposed and opposed to the vertical point, and a **copper disk connects to ground**. Opposite the horizontal point, a **ring of tin** to which are welded tubes alternatively **16 tubes** composed each composed of **copper-lead** and **iron-lead** . The ring connects to the **ground** as shown in figure. 2.

A second regulator form of **automatic breakers** balanced on 3 poles and two poles also equally balanced.

4. A panel of **resistances** composed of **wire mesh** in bunches from different sections of **glass tubes** containing **copper** dust, **coal** and flowers of **sulfur**.

Vacuum cleaners consists of a **wooden box** on each pole, containing a **porcelain vase** in which **layers** isolated with **mica** is made up of **mercury, tin, coal, copper** and **sulfur**, all contaied in a **copper tube**.

6. The witnesses are comprised of ordinary incandescent lamps.

Resume

By the point at the zenith and the point to the south we channel two currents forming the two poles. We also protect from lightning. It regularizes the flow by a regulator and similar devices by a controlling each devise of adequate strength of the current harmful nature of these devices do not have loads. The refined current is conducted by ordinary copper wires**.**

French Patent # 565,395

Combined Apparatus for Capture of Atmospheric Electric Currents with Immediate Implementation

25 January 1924

We know that earth with its constitution, its rotation and movement in space, provides the electricity in the atmosphere. The electrical currents escape into space or largely accumulate towards the equator, as a result of the greater periphery of the globe.

In the atmosphere, there are two clearly distinct poles, i.e., clouds that can be electrically positive or negative; everything in nature shows this, otherwise it would be difficult to explain the lightning that occurs between the clouds and which are none other than contrary cloud electrical charges discharged by too close proximity.

The invention relates to a set of devices capable of capturing the atmospheric electricity.

So far, all searches made for this purpose has been to capture the atmosphere, i.e., that brought together into one system to capture two poles and opposes effect has been to destroy or cancel each other leaving the amount as the difference of the strongest to weakest.

It is therefore easy to understand why, considering the two poles of air as about equal, it is almost impossible to measure some potential with the land, the highest on the lowest remaining which can be positive or negative, copper being the strongest of one or other of these polarities.

It is on this basis that all research until now have been made, and that is why we must consider that all the time, between the two poles captured in the atmosphere and not in any mixture of systems uptake, it is possible to obtain tension and intensity, a considerable and measurable power captured separately between the two poles and remaining isolated from one another until their utilization .

In accordance with this invention, this collection is made by attraction of primary electrical currents in teh atmosphere, either positively or negatively charged, by an antenna that has two points absolutely isolated from one another, where one points to the zenith to attract negative electricity, and the other turned south, toward the equator to draws in the masses grouped in this area.

This shows and proves the existence of atmospheric electrical currents, and that it is possible to capture them with specially designed equipment primarily acting as regulators, as the persistent difficulty encountered so far in realizing their capture lies with large variations of tension in which the current atmosphere is present in space, for each of the poles, and an overload could inevitably fatal.

The patent in its present form of presentation does not cover the regulators or devices that are anticipated in the system, because they can be designed in different ways and give the same result, but covers the application and grouping of equipment operating as automatic relays with variable influence to limit current tension to that chosen for utilization.

These devices have a role associated with an overflow reservoir outside carrying excess liquid; these regulators will divert the excess voltage flows to ground.

We can still absorb these regulators having filters because they have the sole purpose of diverting the post being used , currents that are not yet known but likely that we will call for simplicity, abnormal currents. [sic...]

The patent also covers systems for regulating currents, applied to each of the poles because it is recognized that the shape and tension of the currents of the two poles are not equivalent.

The method of capture is by special antenna has directed two points in the atmosphere, as will be stated after this, with Boot ecoulement prior to the current atmosphere is also of great importance.

Finally, the current atmospheric reception is adequate even with the installation of a post, which can be expected at any location without the need for special altitude, as has been attempted in previous efforts.

We refer to the attached drawings:.

Figure 1 represents in elevation and side view an antenna that only has two points, to overcome an elevation ;

Figure 2 is the end of one of the peaks, the horizontal:

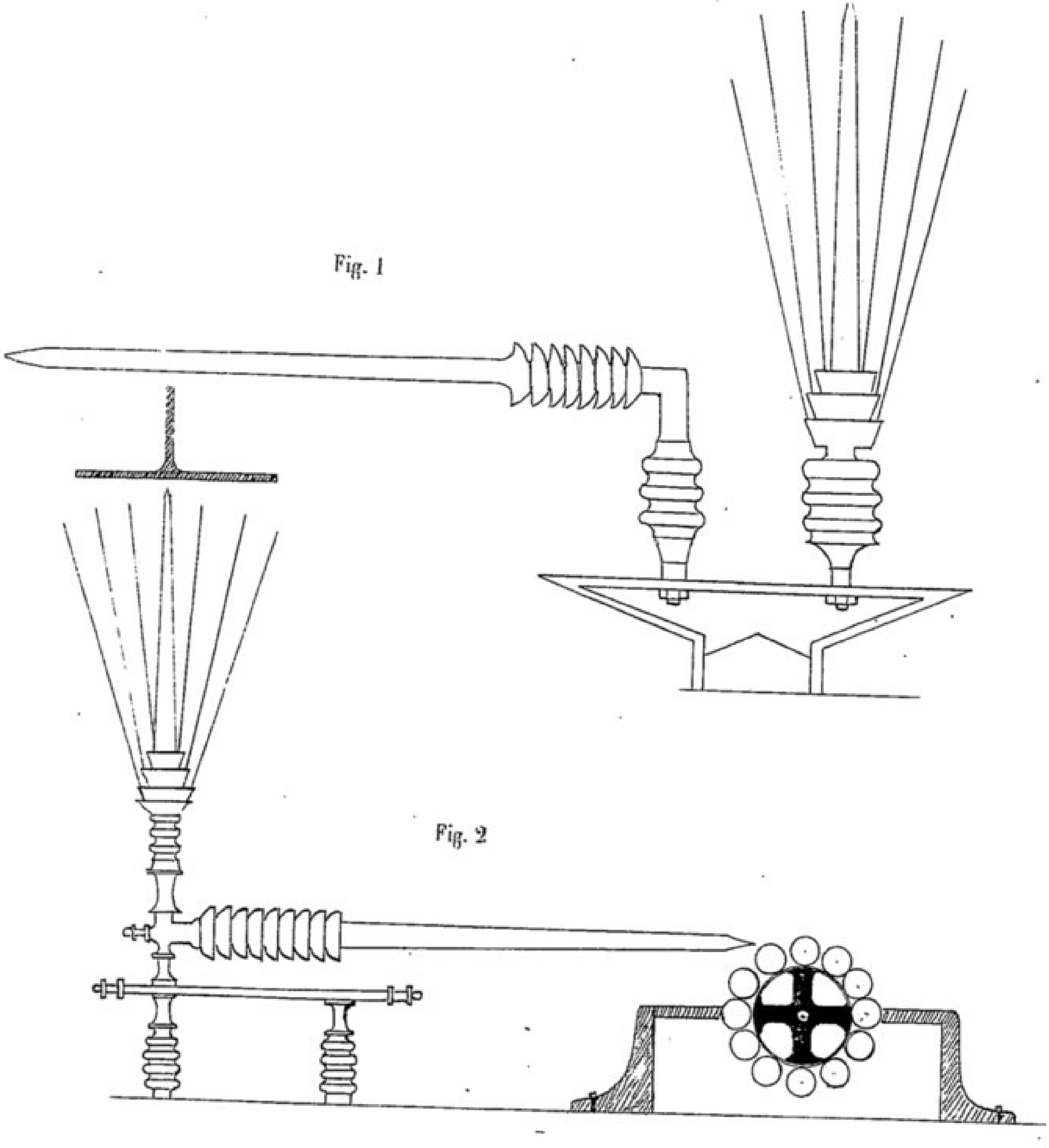

Fig. 1

Fig. 2

Figure 3 shows schematically the table on which are mounted various devices or accessories that act as regulators of current tensions.

Fig. 3.

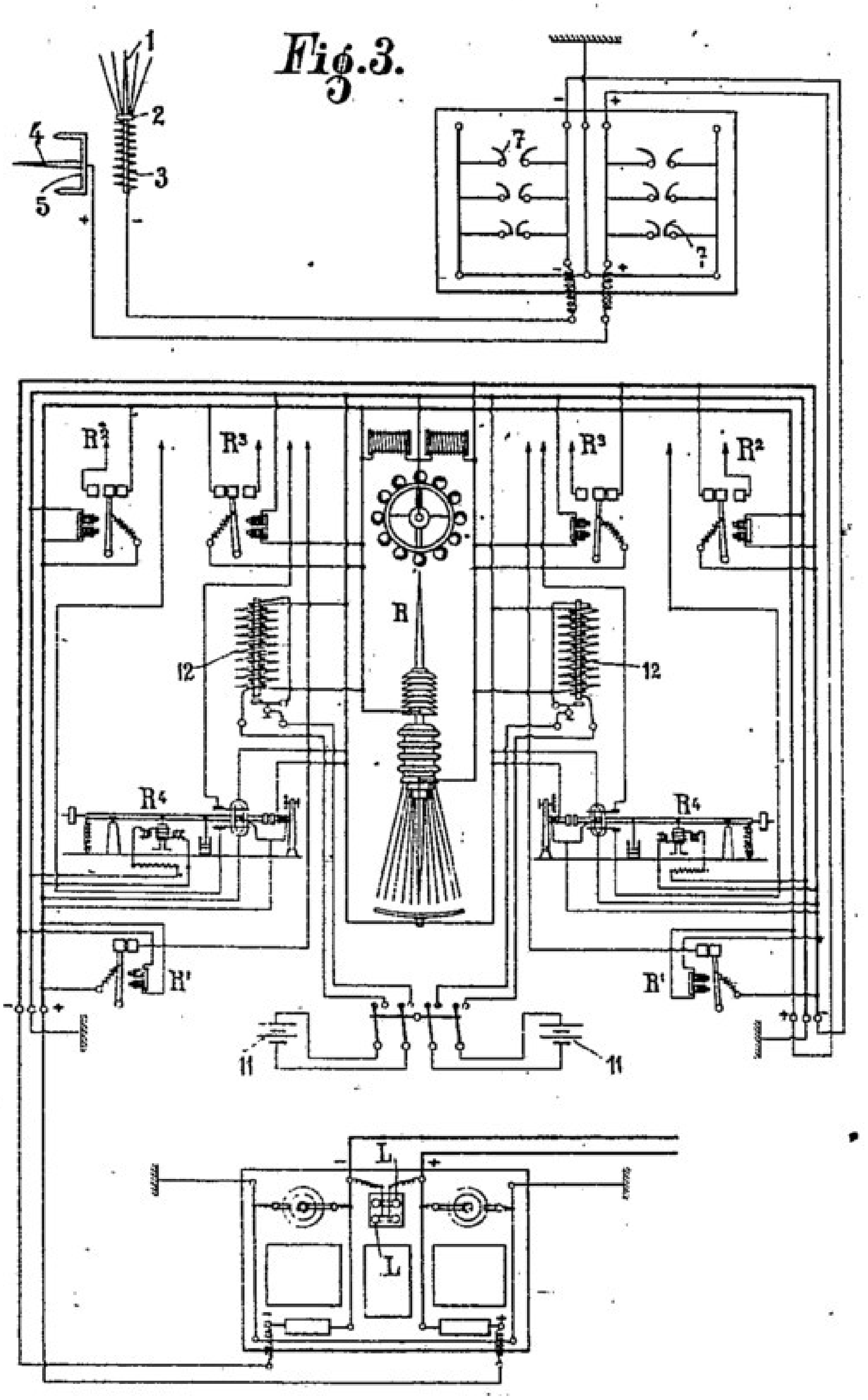
1
2
3
4
5
+
-
7
7
R²
R³
R³
R²
R
12
12
R⁴
R⁴
R¹
R¹
11
11
L
L

Figure 4 is a scheme of surge arrestors, and

Figure 5 a table of resistances.

Fig. 4.

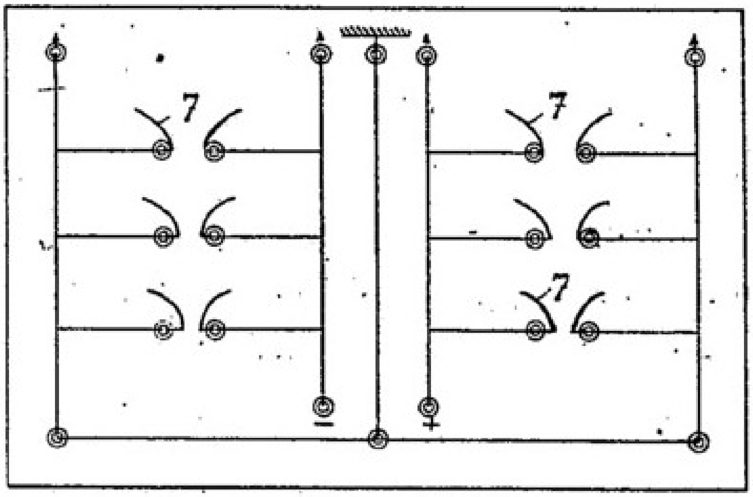

Fig. 5.

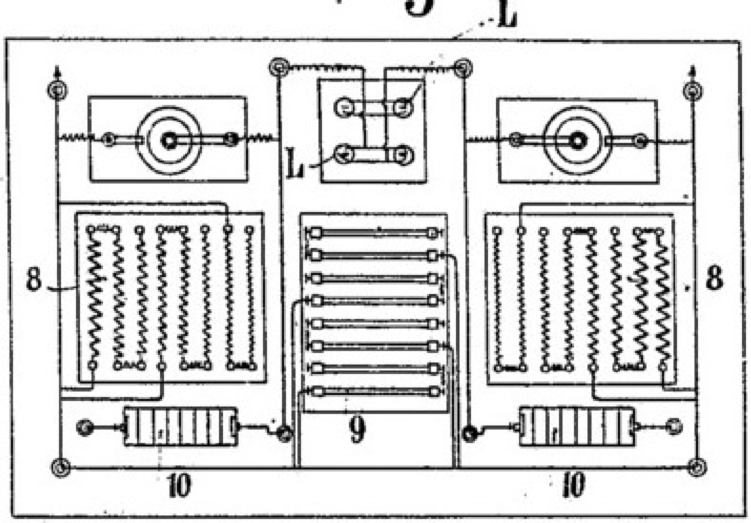

The necessary arrangements to capture atmospheric current form is as follows:

1. The **antenna** in figure 1 is formed with two peaks: the 1 is **steel**, supported by a **bronze ring** 2, surrounded by a **spiral form** 3; this point is oriented for attracting the **zenith** layers of **negative electricity** and the other edge 4 contains **copper coils** 6, and is oriented **south** to the **equator** to draw in the masses grouped in this area. To use this antenna it suffices to be raised on a mast or **pole** to a height a little higher than that of surrounding houses.

2. On the table **lightning arrestors**, contained 4, forms of 7 **copper horns** used to absorb large atmospheric discharges by certain time, as it would be dangerous to allow movement in the devices regulators.

3. On a table of **two** symmetrical systems **regulators**, Figure 3, one for each pole, to obtain and allow a regularization of large excesses in the current tensions, by automatically absorbing surges before passing over the surfacea chosen for the desired use. These regulators are represented on the drawing by the references R, R1, R2, R3, R4.

4. In the **resistance** panel in communication with regulators, these resistances are formed as **spiral coils or screens** 8, **tubes 9 of glass filled with fine coal dust , aluminum powder, sulfur and fine copper dust**. Finally, **two** special 10 **processors** and constitutes **layers** of metal forming an **absorbant** for harmful or abnormal currents.

5. As a source of energy formed **auxiliary batteries** 11 and **Ruhmkorff coils** 12 to get into the antenna sufficient attraction by means of an adjustment of intensity, which once it begins, allows the constant flow of atmospheric current in the system.

In considering Figure 3, which represents a sort of an installation scheme of the device, the reception of atmospheric electrical current will, as set out above, producing a priming of these currents. This boot is made in launching the system which is double to answer each of the poles, the current batteries 11 through 12 and reels going to the antenna.

This boot will continue until the disposal of electrical current atmosphere is evident by the one or more table lamps L of resistance shown in Figure 5.

Once the lights illuminate it is advisable to charge the batteries to boot. Natural flow and constant currents will be captured by the regulators , Figure 3; they will regularize in tension and in amperage and then directed to use in table lamps.

It has been said above that regulators R, etc., are intended and designed to automatically eliminate earth surge currents capture and abnormal currents that cannot yet be level but that these regulators could be replaced by devices with the same function.

There is therefore no need to describe in detail and especially the claims. We only specify that these regulators must be proper protection for the purposes described above.

The classic spark-gap with magnetic blowing circuit

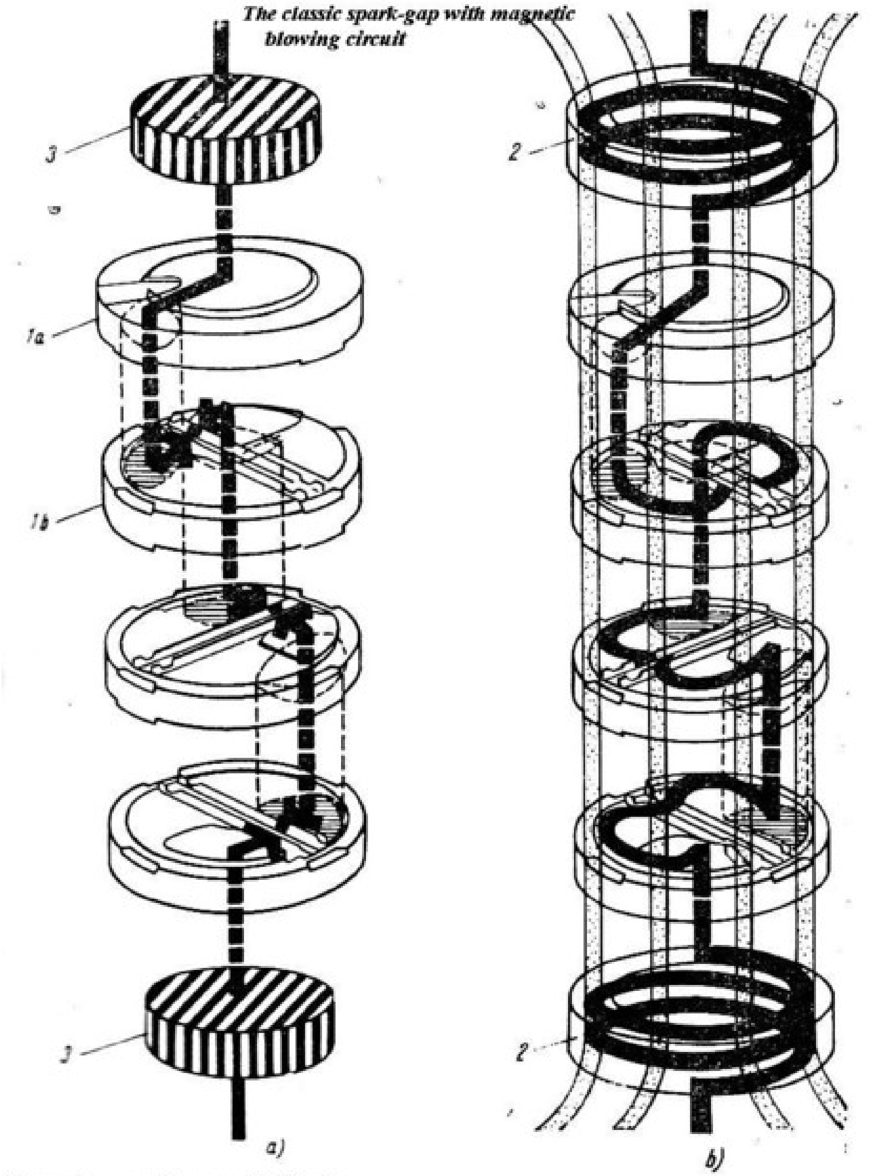

The spark-gap with magnetic blowing :
a,- the initial stage ; the impulse current send in soil through the shunt-resistences 3, because the coils 2 have an impedance too big for the impulse of 10/20 µs ;
the spark-gap 1 has two elements disc-shaped 1a and 1b, the inferior disc have one electrode and the upper disc the other electrode of the spark-gap circuit.
b,- the second stage ; the accompanying current discharge - after the impulse current ;
this accompanying current pass through the coils 2 which produce the magnetic field with the role to extend the electric arc into the slit between insulator discs 1a and 1b ; the thickness of this slit is smaller and smaller in accord with the extension of the electric arc, in this manner producing the magnifing of its electric resistance and finally, the electric arc extinction.

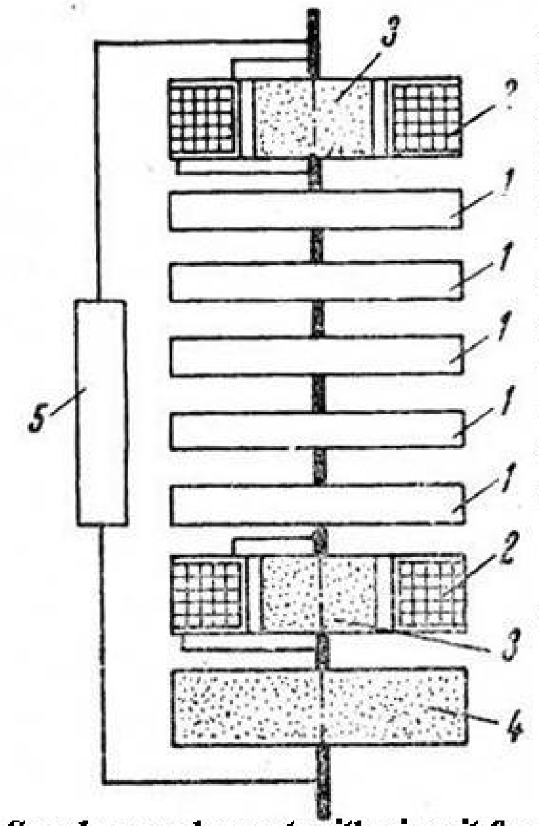

1,- the spark-gap individual devices which work in the manner to extend the electric arc of the accompanying electric current ;
2,- the coils which produce the magnetic field for the magnetic blowing ;
3,- the electric shunt-resistances in parallel connection on the blowing coils, with the role of reducing the current inside this coils
4,- the disc with unlinear resistance ;
5,- electrical resistance for the homogenous distribution of the electric field in all the elements of the spark-gap.

In Jules Guillot scheme, that kind of device could be used for leading the current.

Spark-gap element with circuit for the magnetic blowing of the electric fascicle

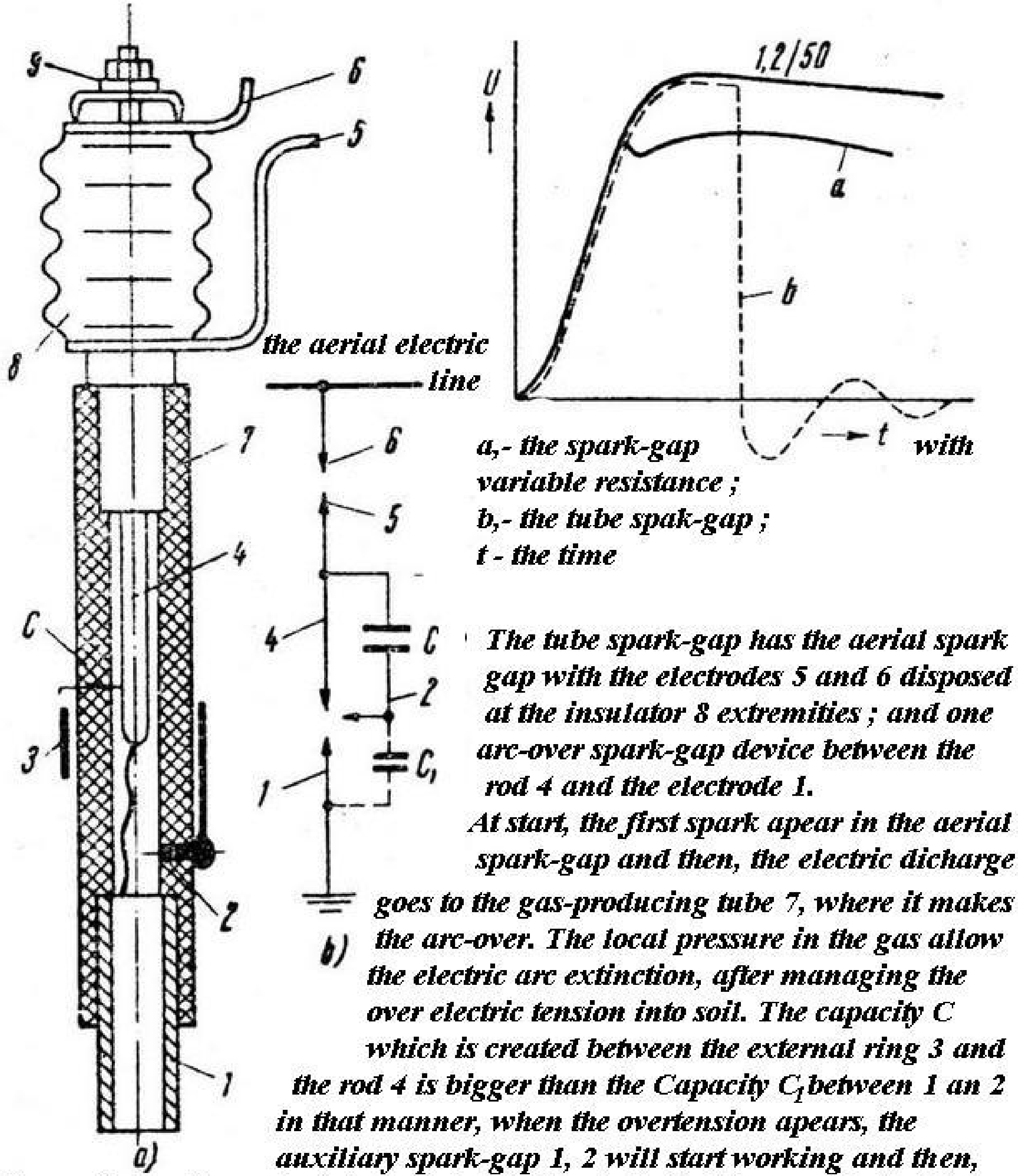

a,- the spark-gap with variable resistance ;
b,- the tube spak-gap ;
t - the time

The tube spark-gap has the aerial spark gap with the electrodes 5 and 6 disposed at the insulator 8 extremities ; and one arc-over spark-gap device between the rod 4 and the electrode 1.

At start, the first spark apear in the aerial spark-gap and then, the electric dicharge goes to the gas-producing tube 7, where it makes the arc-over. The local pressure in the gas allow the electric arc extinction, after managing the over electric tension into soil. The capacity C which is created between the external ring 3 and the rod 4 is bigger than the Capacity C_1 between 1 an 2 in that manner, when the overtension apears, the auxiliary spark-gap 1, 2 will start working and then, the main spark-gap 1, 4. In this circuit, the electrode 9 is connected at the aerial electric line and the electrode 1 at the Earth.

In the image above, it is an classic tube spark-gap used for the protection of aerial electric cables. The atmospheric electricity is discharged in soil and lost... The Guillot's inventions used that kind of devices for leading and absorbing of the atmospheric electricity.

DARIDA, Gillbert : "Practical Utilization of Atmospheric Electricity";
(in *The Invention Encyclopedia,* pp. 204-207 ; 1930, Geo. Constantinescu, Ed.) --

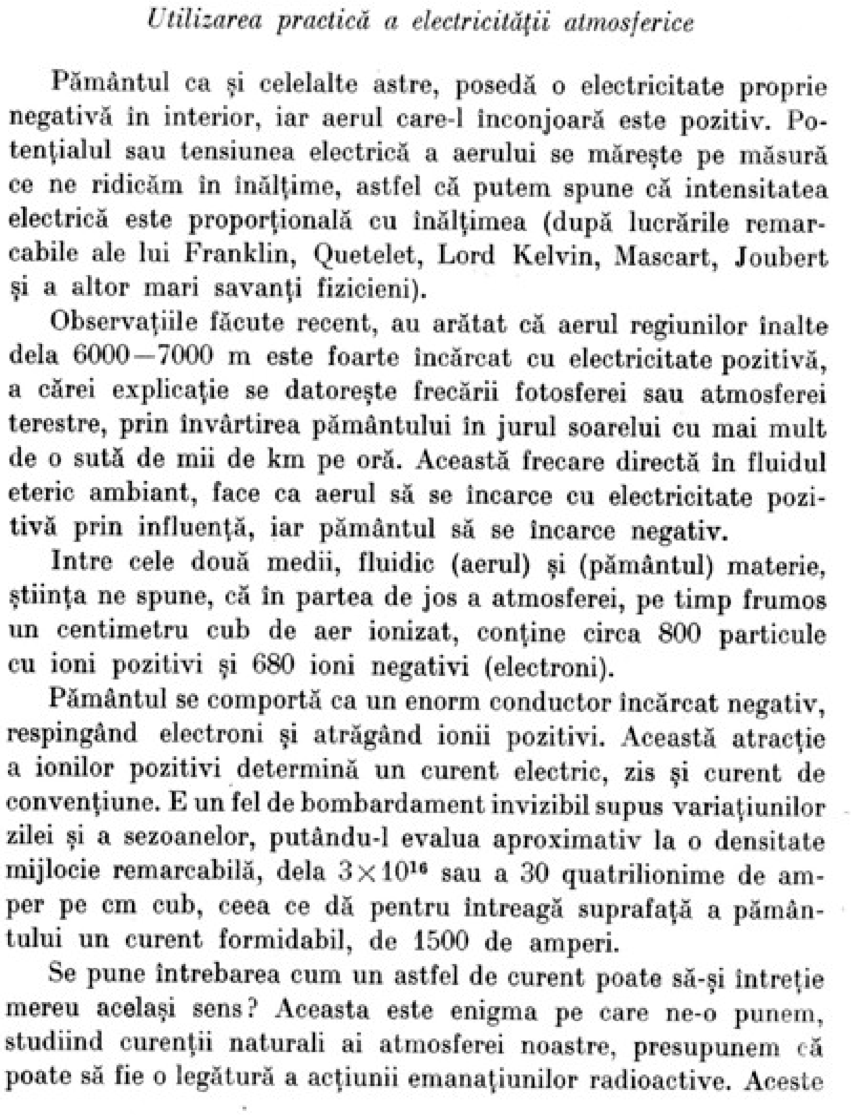

Utilizarea practică a electricității atmosferice

Pământul ca și celelalte astre, posedă o electricitate proprie negativă în interior, iar aerul care-l înconjoară este pozitiv. Potențialul sau tensiunea electrică a aerului se mărește pe măsură ce ne ridicăm în înălțime, astfel că putem spune că intensitatea electrică este proporțională cu înălțimea (după lucrările remarcabile ale lui Franklin, Quetelet, Lord Kelvin, Mascart, Joubert și a altor mari savanți fizicieni).

Observațiile făcute recent, au arătat că aerul regiunilor înalte dela 6000—7000 m este foarte încărcat cu electricitate pozitivă, a cărei explicație se datorește frecării fotosferei sau atmosferei terestre, prin învârtirea pământului în jurul soarelui cu mai mult de o sută de mii de km pe oră. Această frecare directă în fluidul eteric ambiant, face ca aerul să se încarce cu electricitate pozitivă prin influență, iar pământul să se încarce negativ.

Intre cele două medii, fluidic (aerul) și (pământul) materie, știința ne spune, că în partea de jos a atmosferei, pe timp frumos un centimetru cub de aer ionizat, conține circa 800 particule cu ioni pozitivi și 680 ioni negativi (electroni).

Pământul se comportă ca un enorm conductor încărcat negativ, respingând electroni și atrăgând ionii pozitivi. Această atracție a ionilor pozitivi determină un curent electric, zis și curent de convențiune. E un fel de bombardament invizibil supus variațiunilor zilei și a sezoanelor, putându-l evalua aproximativ la o densitate mijlocie remarcabilă, dela 3×10^{16} sau a 30 quatrilionime de amper pe cm cub, ceea ce dă pentru întreagă suprafață a pământului un curent formidabil, de 1500 de amperi.

Se pune întrebarea cum un astfel de curent poate să-și întreție mereu același sens? Aceasta este enigma pe care ne-o punem, studiind curenții naturali ai atmosferei noastre, presupunem că poate să fie o legătură a acțiunii emanațiunilor radioactive. Aceste

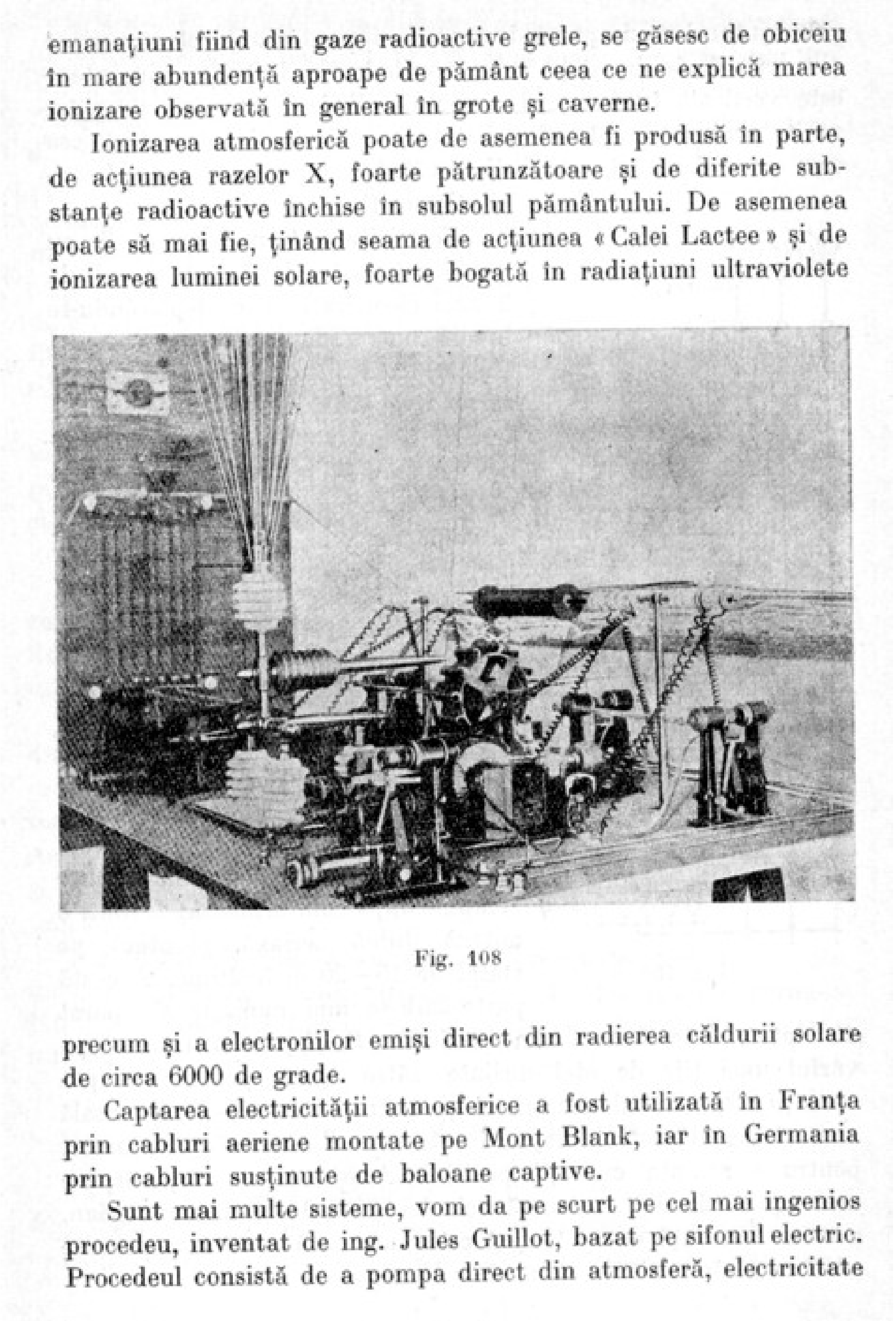

emanaţiuni fiind din gaze radioactive grele, se găsesc de obiceiu în mare abundenţă aproape de pământ ceea ce ne explică marea ionizare observată în general în grote şi caverne.

Ionizarea atmosferică poate de asemenea fi produsă în parte, de acţiunea razelor X, foarte pătrunzătoare şi de diferite substanţe radioactive închise în subsolul pământului. De asemenea poate să mai fie, ţinând seama de acţiunea «Calei Lactee» şi de ionizarea luminei solare, foarte bogată în radiaţiuni ultraviolete

Fig. 108

precum şi a electronilor emişi direct din radierea căldurii solare de circa 6000 de grade.

Captarea electricităţii atmosferice a fost utilizată în Franţa prin cabluri aeriene montate pe Mont Blank, iar în Germania prin cabluri susţinute de baloane captive.

Sunt mai multe sisteme, vom da pe scurt pe cel mai ingenios procedeu, inventat de ing. Jules Guillot, bazat pe sifonul electric. Procedeul consistă de a pompa direct din atmosferă, electricitate

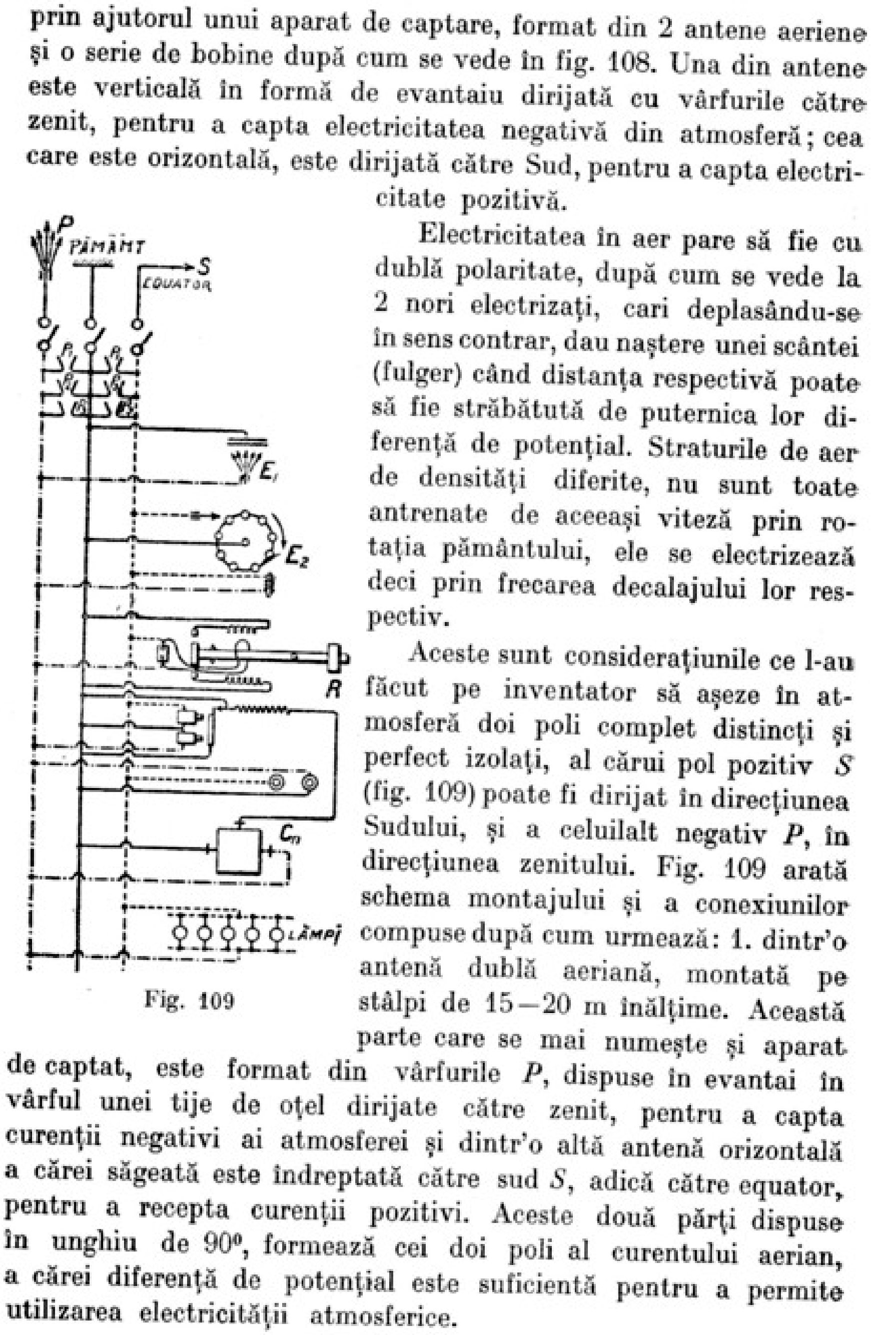
P
PĂMÂNT
S
EQUATOR
E₁
E₂
R
Cₙ
LĂMPI

prin ajutorul unui aparat de captare, format din 2 antene aeriene şi o serie de bobine după cum se vede în fig. 108. Una din antene este verticală în formă de evantaiu dirijată cu vârfurile către zenit, pentru a capta electricitatea negativă din atmosferă; cea care este orizontală, este dirijată către Sud, pentru a capta electricitate pozitivă.

Electricitatea în aer pare să fie cu dublă polaritate, după cum se vede la 2 nori electrizaţi, cari deplasându-se în sens contrar, dau naştere unei scântei (fulger) când distanţa respectivă poate să fie străbătută de puternica lor diferenţă de potenţial. Straturile de aer de densităţi diferite, nu sunt toate antrenate de aceeaşi viteză prin rotaţia pământului, ele se electrizează deci prin frecarea decalajului lor respectiv.

Aceste sunt consideraţiunile ce l-au făcut pe inventator să aşeze în atmosferă doi poli complet distincţi şi perfect izolaţi, al cărui pol pozitiv *S* (fig. 109) poate fi dirijat în direcţiunea Sudului, şi a celuilalt negativ *P*, în direcţiunea zenitului. Fig. 109 arată schema montajului şi a conexiunilor compuse după cum urmează: 1. dintr'o antenă dublă aeriană, montată pe stâlpi de 15—20 m înălţime. Această parte care se mai numeşte şi aparat de captat, este format din vârfurile *P*, dispuse în evantai în vârful unei tije de oţel dirijate către zenit, pentru a capta curenţii negativi ai atmosferei şi dintr'o altă antenă orizontală a cărei săgeată este îndreptată către sud *S*, adică către equator, pentru a recepta curenţii pozitivi. Aceste două părţi dispuse în unghiu de 90°, formează cei doi poli al curentului aerian, a cărei diferenţă de potenţial este suficientă pentru a permite utilizarea electricităţii atmosferice.

Fig. 109

French Patent # 551,882

Appareil Capteur de Courants Electriques dans L'Atmosphere

BREVET D'INVENTION.

XII. — Instruments de précision, électricité.

5. — Production de l'électricité, moteurs électriques.

N° 551.882

Appareil capteur de courants électriques dans l'atmosphère.

M. Jules GUILLOT résidant en France (Tarn-et-Garonne).

Demandé le 27 septembre 1921, à 15h 20m, à Paris.
Délivré le 15 janvier 1923. — Publié le 16 avril 1923.

[Brevet d'invention dont la délivrance a été ajournée en exécution de l'art. 11 § 7 de la loi du 5 juillet 1844 modifiée par la loi du 7 avril 1902.]

Considérant le globe terrestre comme l'induit d'une dynamo, dont l'éther supraterrestre est l'inducteur, des courants électriques circulent dans l'intervalle c'est-à-dire dans la couche atmosphérique; l'inventeur capte au moyen de l'appareil décrit ci-dessous deux de ces courants parfaitement distincts et élimine les autres.

L'appareil comprend :

1° Le capteur aérien; 2° une série de parafoudres; 3° les régulateurs; 4° un tableau de résistances; 5° les aspirateurs; 6° les témoins.

1° Le capteur aérien monté sur un poteau est composé d'une pointe d'acier aimanté fixée par scellement sur un isolateur porcelaine et cerclée à sa base par une bague de bronze sur laquelle sont vissées 32 aiguilles de fer doux, le tout formant l'antenne d'un pôle. A côté et isolé de la première, formant le 2e pôle, est fixée une pointe d'acier aimanté dans la position horizontale et dirigée vers le sud dont la base coudée à 45° est scellée sur un isolateur porcelaine. Cette pointe est également cerclée d'une bague de cuivre à encoches fig. 1.

2° Une série de 6 parafoudres à cornes entre chaque pôle et la terre et à écartements variés.

3° Un premier régulateur formé d'un appareil absolument semblable au capteur aérien, mais ayant les deux pôles superposés, et, opposé à la pointe verticale, un disque de cuivre relié à la terre. Opposé à la pointe horizontale, une roue à jante d'étain sur laquelle sont soudés alternativement 16 tubes composés eux-mêmes de deux métaux soudés : cuivre-plomb et fer-plomb. Cette roue est reliée à la terre par son pivot fig. 2.

Un deuxième régulateur formé d'appareils à déclanchement automatique au nombre de 3 par pôles et deux basculateurs également par pôles.

4° Un tableau de résistances composé de fils de maillechort en boudins de différentes sections, de tubes en verre contenant des poussières de cuivre, du charbon de cornue et de la fleur de soufre.

5° Les aspirateurs composés d'une boîte en bois sur chaque pôle, contenant un vase de porcelaine dans lequel par couches superposées et isolées au mica est placé du mercure, de l'étain, du charbon, du cuivre et du soufre, le tout traversé par une tige de cuivre.

6° Les témoins composés de lampes à incandescence ordinaires.

RÉSUMÉ.

Par la pointe au zénith et par la pointe au

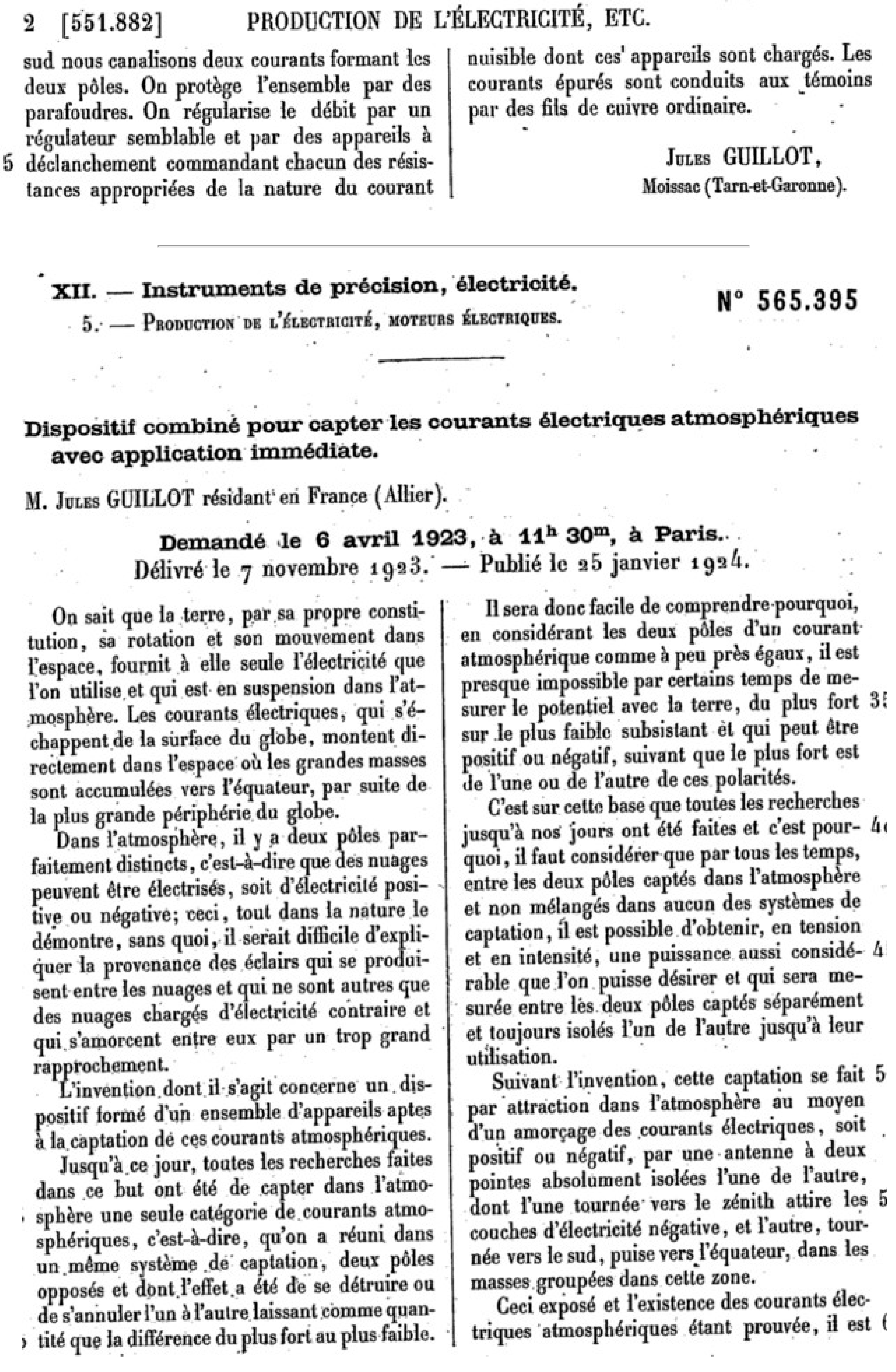

sud nous canalisons deux courants formant les deux pôles. On protège l'ensemble par des parafoudres. On régularise le débit par un régulateur semblable et par des appareils à déclanchement commandant chacun des résistances appropriées de la nature du courant nuisible dont ces appareils sont chargés. Les courants épurés sont conduits aux témoins par des fils de cuivre ordinaire.

Jules GUILLOT,
Moissac (Tarn-et-Garonne).

XII. — Instruments de précision, électricité.

5. — Production de l'électricité, moteurs électriques.

N° 565.395

Dispositif combiné pour capter les courants électriques atmosphériques avec application immédiate.

M. Jules GUILLOT résidant en France (Allier).

Demandé le 6 avril 1923, à 11h 30m, à Paris.
Délivré le 7 novembre 1923. — Publié le 25 janvier 1924.

On sait que la terre, par sa propre constitution, sa rotation et son mouvement dans l'espace, fournit à elle seule l'électricité que l'on utilise et qui est en suspension dans l'atmosphère. Les courants électriques, qui s'échappent de la surface du globe, montent directement dans l'espace où les grandes masses sont accumulées vers l'équateur, par suite de la plus grande périphérie du globe.

Dans l'atmosphère, il y a deux pôles parfaitement distincts, c'est-à-dire que des nuages peuvent être électrisés, soit d'électricité positive ou négative; ceci, tout dans la nature le démontre, sans quoi, il serait difficile d'expliquer la provenance des éclairs qui se produisent entre les nuages et qui ne sont autres que des nuages chargés d'électricité contraire et qui s'amorcent entre eux par un trop grand rapprochement.

L'invention dont il s'agit concerne un dispositif formé d'un ensemble d'appareils aptes à la captation de ces courants atmosphériques.

Jusqu'à ce jour, toutes les recherches faites dans ce but ont été de capter dans l'atmosphère une seule catégorie de courants atmosphériques, c'est-à-dire, qu'on a réuni dans un même système de captation, deux pôles opposés et dont l'effet a été de se détruire ou de s'annuler l'un à l'autre laissant comme quantité que la différence du plus fort au plus faible.

Il sera donc facile de comprendre pourquoi, en considérant les deux pôles d'un courant atmosphérique comme à peu près égaux, il est presque impossible par certains temps de mesurer le potentiel avec la terre, du plus fort sur le plus faible subsistant et qui peut être positif ou négatif, suivant que le plus fort est de l'une ou de l'autre de ces polarités.

C'est sur cette base que toutes les recherches jusqu'à nos jours ont été faites et c'est pourquoi, il faut considérer que par tous les temps, entre les deux pôles captés dans l'atmosphère et non mélangés dans aucun des systèmes de captation, il est possible d'obtenir, en tension et en intensité, une puissance aussi considérable que l'on puisse désirer et qui sera mesurée entre les deux pôles captés séparément et toujours isolés l'un de l'autre jusqu'à leur utilisation.

Suivant l'invention, cette captation se fait par attraction dans l'atmosphère au moyen d'un amorçage des courants électriques, soit positif ou négatif, par une antenne à deux pointes absolument isolées l'une de l'autre, dont l'une tournée vers le zénith attire les couches d'électricité négative, et l'autre, tournée vers le sud, puise vers l'équateur, dans les masses groupées dans cette zone.

Ceci exposé et l'existence des courants électriques atmosphériques étant prouvée, il est

possible de les capter à la condition de disposer d'appareils spéciaux étudiés surtout pour jouer le rôle de régulateurs, attendu que la difficulté persistante rencontrée jusqu'alors pour réaliser la captation incombe aux grandes variations de tensions sous lesquelles les courants atmosphériques se présentent dans l'espace, pour chacun des pôles, et dont l'écoulement dans un poste de captation serait inévitablement funeste.

Le présent brevet sous sa forme de présentation doit donc couvrir, non pas le ou les dispositifs régulateurs qui sont prévus dans le système, car ils peuvent être conçus d'une façon toute différente et donner le même résultat, mais sur l'application raisonnée et le groupement d'appareils fonctionnant automatiquement et intervenant chacun comme relais d'influence variable pour limiter la tension des courants captés dans l'atmosphère à celle choisie pour utilisation.

Ces appareils ont un rôle assez connexe avec celui de trop-pleins sur un réservoir qui acheminent à l'extérieur l'excès de liquide; ici ces régulateurs détournent l'excès de tension des courants à la terre.

On peut encore assimiler ces régulateurs à des filtres car ils ont pour but de détourner du poste d'utilisation, des courants qui ne sont pas encore de nature connue mais que pour simplicité on appellera courants parasites ou anormaux.

Le brevet portera aussi sur le système de régulation des courants, appliqué à chacun des pôles, car il est reconnu que la forme et la tension des courants aux deux pôles ne sont pas équivalentes.

Le mode de captation par antenne spéciale à deux pointes orientées dans l'atmosphère, comme il sera indiqué ci-après, avec amorçage préalable pour l'écoulement des courants atmosphériques est aussi d'une grande importance.

Enfin la réception des courants atmosphériques se fait à l'endroit même de l'installation du poste; celui-ci peut être prévu en un lieu quelconque sans nécessiter d'altitude particulière, comme cela a été tenté dans des essais préconçus.

Sur les dessins ci-annexés auxquels on se réfère

La fig. 1 représente en élévation et en coupe partielle l'antenne seule, à deux pointes, surmontant un poteau d'élévation;

La fig. 2 est le plan de l'une des pointes, celle horizontale:

La fig. 3 montre schématiquement le tableau sur lequel sont montés les divers appareils ou accessoires qui jouent le rôle de régulateurs des tensions de courants.

La fig. 4 est un schéma de parafoudre, et

La fig. 5 un tableau de résistances.

Le dispositif nécessaire à la captation des courants atmosphériques est formé comme suit :

1° D'une antenne, fig. 1, formée de deux pointes dont l'une 1 est en acier, supportée par une bague de bronze 2 entourée d'une spirale 3 formant self; cette pointe est tournée vers le zénith pour attirer les couches d'électricité négative; l'autre pointe 4 comporte un cadre 5 de huit branches supportant des spirales de cuivre 6 et est tournée vers le sud pour puiser vers l'équateur dans les masses groupées dans cette zone. Pour l'utilisation de cette antenne il suffira qu'elle soit portée par un mât ou poteau de hauteur un peu supérieure à celle des habitations environnantes.

2° D'un tableau de parafoudre, fig. 4, formé de cornes de cuivre 7 servant à absorber les tros grosses décharges atmosphériques que par certains temps, il serait dangereux de laisser circuler dans les appareils régulateurs.

3° D'une table de deux systèmes symétriques de régulateurs, fig. 3, un pour chaque pôle, qui permet d'obtenir une régularisation assez grande dans les tensions des courants en absorbant automatiquement les surtensions dépassant la tension choisie pour l'utilisation désirée. Ces régulateurs sont représentés sur le dessin par les références R, R^1, R^2, R^3, R^4.

4° D'un tableau de résistance en communication avec les régulateurs, ces résistances sont formées de spirales de maillechor 8, de tubes 9 de verre remplis de poussière fine de charbon de cornue, de poudre d'aluminium, de soufre et de poussière fine de cuivre; enfin, deux appareils spéciaux 10 formant transformateurs et constitués de couches de métaux absorbant certains courants nuisibles ou anormaux.

5° D'une source auxiliaire d'énergie formée de piles 11 et de bobines Ruhmkorff 12 permettant d'obtenir dans l'antenne une attraction suffisante au moyen d'un dispositif de

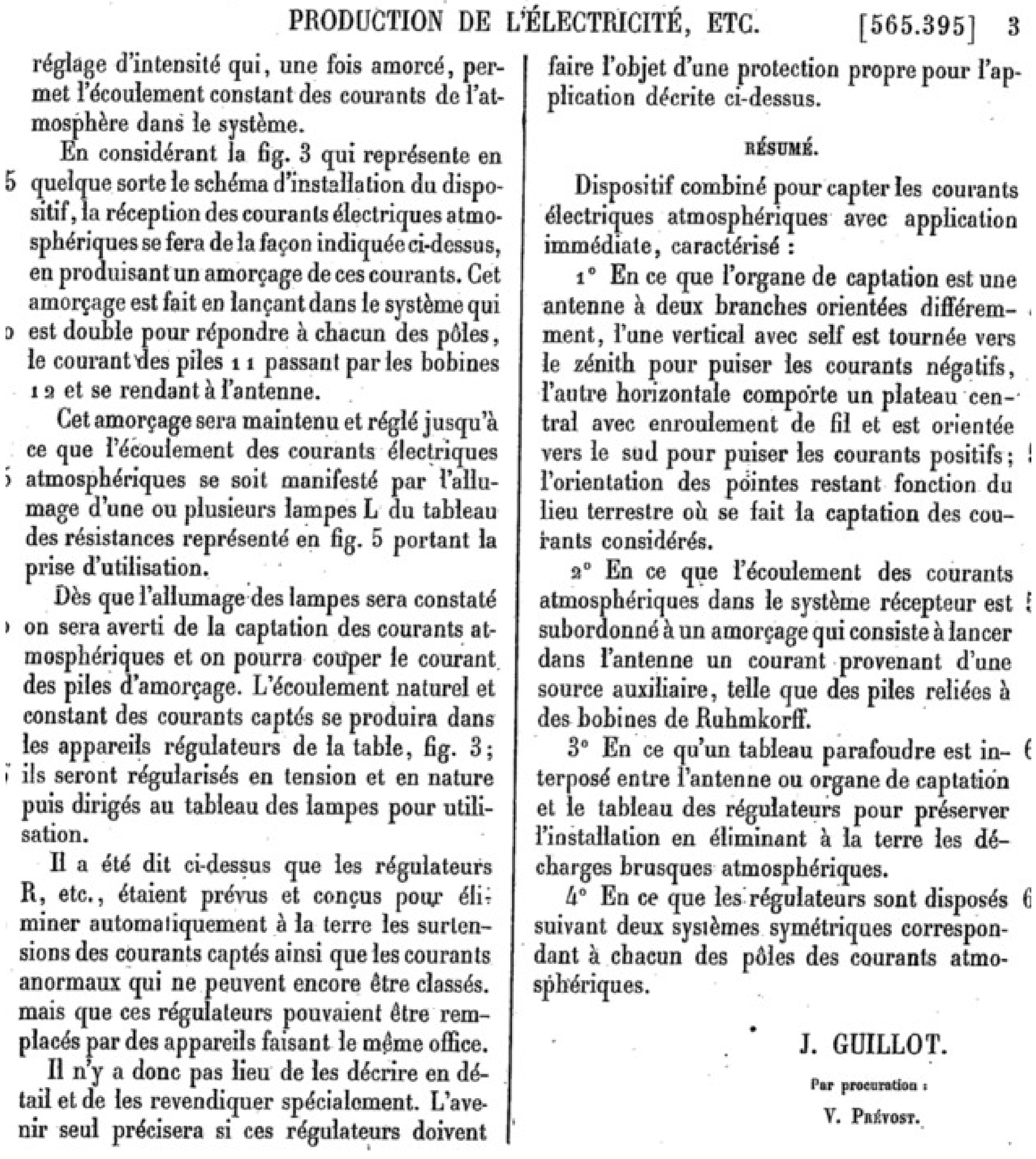

réglage d'intensité qui, une fois amorcé, permet l'écoulement constant des courants de l'atmosphère dans le système.

En considérant la fig. 3 qui représente en quelque sorte le schéma d'installation du dispositif, la réception des courants électriques atmosphériques se fera de la façon indiquée ci-dessus, en produisant un amorçage de ces courants. Cet amorçage est fait en lançant dans le système qui est double pour répondre à chacun des pôles, le courant des piles 11 passant par les bobines 12 et se rendant à l'antenne.

Cet amorçage sera maintenu et réglé jusqu'à ce que l'écoulement des courants électriques atmosphériques se soit manifesté par l'allumage d'une ou plusieurs lampes L du tableau des résistances représenté en fig. 5 portant la prise d'utilisation.

Dès que l'allumage des lampes sera constaté on sera averti de la captation des courants atmosphériques et on pourra couper le courant des piles d'amorçage. L'écoulement naturel et constant des courants captés se produira dans les appareils régulateurs de la table, fig. 3; ils seront régularisés en tension et en nature puis dirigés au tableau des lampes pour utilisation.

Il a été dit ci-dessus que les régulateurs R, etc., étaient prévus et conçus pour éliminer automatiquement à la terre les surtensions des courants captés ainsi que les courants anormaux qui ne peuvent encore être classés. mais que ces régulateurs pouvaient être remplacés par des appareils faisant le même office.

Il n'y a donc pas lieu de les décrire en détail et de les revendiquer spécialement. L'avenir seul précisera si ces régulateurs doivent faire l'objet d'une protection propre pour l'application décrite ci-dessus.

RÉSUMÉ.

Dispositif combiné pour capter les courants électriques atmosphériques avec application immédiate, caractérisé :

1° En ce que l'organe de captation est une antenne à deux branches orientées différemment, l'une vertical avec self est tournée vers le zénith pour puiser les courants négatifs, l'autre horizontale comporte un plateau central avec enroulement de fil et est orientée vers le sud pour puiser les courants positifs; l'orientation des pointes restant fonction du lieu terrestre où se fait la captation des courants considérés.

2° En ce que l'écoulement des courants atmosphériques dans le système récepteur est subordonné à un amorçage qui consiste à lancer dans l'antenne un courant provenant d'une source auxiliaire, telle que des piles reliées à des bobines de Ruhmkorff.

3° En ce qu'un tableau parafoudre est interposé entre l'antenne ou organe de captation et le tableau des régulateurs pour préserver l'installation en éliminant à la terre les décharges brusques atmosphériques.

4° En ce que les régulateurs sont disposés suivant deux systèmes symétriques correspondant à chacun des pôles des courants atmosphériques.

J. GUILLOT.

Par procuration :

V. Prévost.

Parts List

Steel Rods, magnetic, 2
Insulator, porcelain
Ring, bronze
Rods, soft iron, 32
Ring, copper
Surge Protector
Disk, copper
Ring, tin
Tubes, steel
Rx Cu-Pb, Fe-Pb
Breakers, automatic
Resistances : Wire Mesh, Glass Tube. Cu, C, S powder
Vacuum Cleaners : Box, Jar, Mica &

Tube, copper;
Rx : Hg, Sn, C, Cu, S
Battery
Ignition Coil

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

Rex Research, POB 19250, Jean, NV 89019 USA

US007855476B2

(12) **United States Patent**
Ogram

(10) **Patent No.: US 7,855,476 B2**
(45) **Date of Patent: Dec. 21, 2010**

(54) **ATMOSPHERIC ELECTRICAL GENERATOR**

(76) Inventor: **Mark Ellery Ogram**, 780 S. Freeman Rd., Tucson, AZ (US) 85748

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 254 days.

(21) Appl. No.: **12/218,297**

(22) Filed: **Jul. 14, 2008**

(65) **Prior Publication Data**

US 2010/0007218 A1 Jan. 14, 2010

(51) **Int. Cl.**
H02G 11/00 (2006.01)

(52) **U.S. Cl.** **307/145**; 307/149; 361/218; 361/230

(58) **Field of Classification Search** 307/149, 307/145; 244/30, 31; 174/2; 361/212, 215–218, 361/230, 231
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427	A *	5/1901	Palencsar	307/149
911,260	A *	2/1909	Pennock	307/149
1,014,719	A *	1/1912	Pennock	307/149
1,540,998	A *	6/1925	Plauson	310/309
3,121,196	A *	2/1964	Kasemir	324/464
4,234,167	A *	11/1980	Lane	254/291
4,486,669	A *	12/1984	Pugh	290/44
4,842,219	A *	6/1989	Jakubowski et al.	244/31
5,203,542	A	4/1993	Coley et al.	
6,254,034	B1 *	7/2001	Carpenter	244/153 R
6,829,911	B2	12/2004	Davis et al.	
6,961,662	B2	11/2005	Murphy	
7,016,785	B2	3/2006	Makela et al.	
7,200,418	B2	4/2007	Kaikuranta et al.	
7,330,366	B2	2/2008	Lee et al.	
7,375,492	B2	5/2008	Calhoon et al.	
7,378,181	B2	5/2008	Skinlo	
7,388,350	B1	6/2008	Wright	
7,394,671	B2	7/2008	Fukumoto et al.	
7,397,220	B2	7/2008	Uchida et al.	

* cited by examiner

Primary Examiner—Fritz M Fleming
(74) *Attorney, Agent, or Firm*—Mark Ogram

(57) **ABSTRACT**

A mechanism to tap an electrical source which uses an aircraft (preferably a lighter than air balloon) tethered by a conductive line. The conductive line is extended/withdrawn by a winch motor to adjust the altitude of the aircraft. The conductive line is isolated from the ground and an electrical conductor is connected to the conductive line and to an electrical load. In this manner, static electricity generated in the atmosphere is gathered for use.

15 Claims, 4 Drawing Sheets

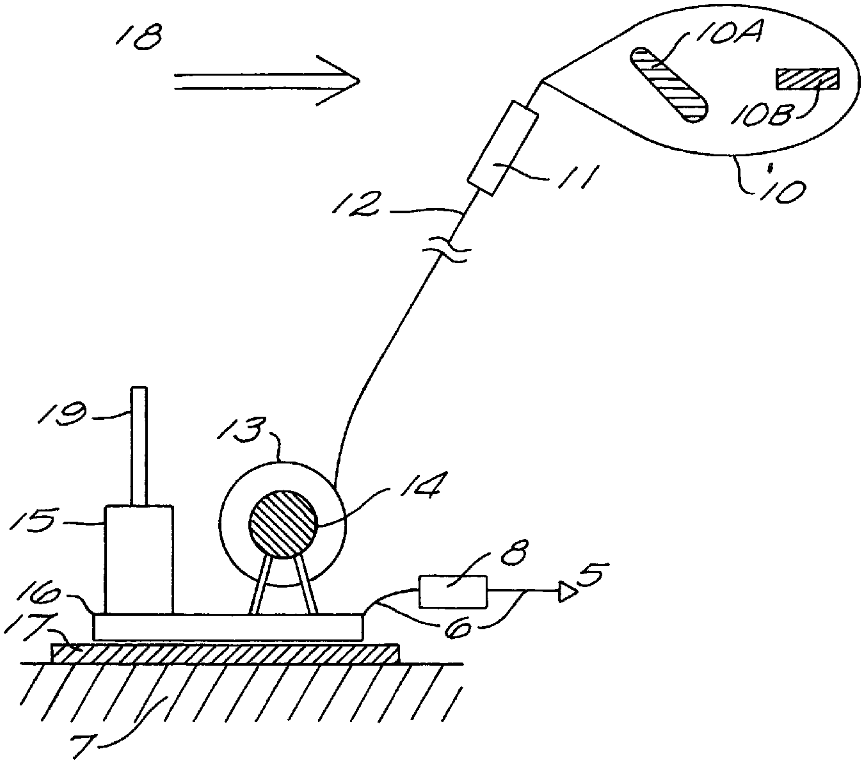

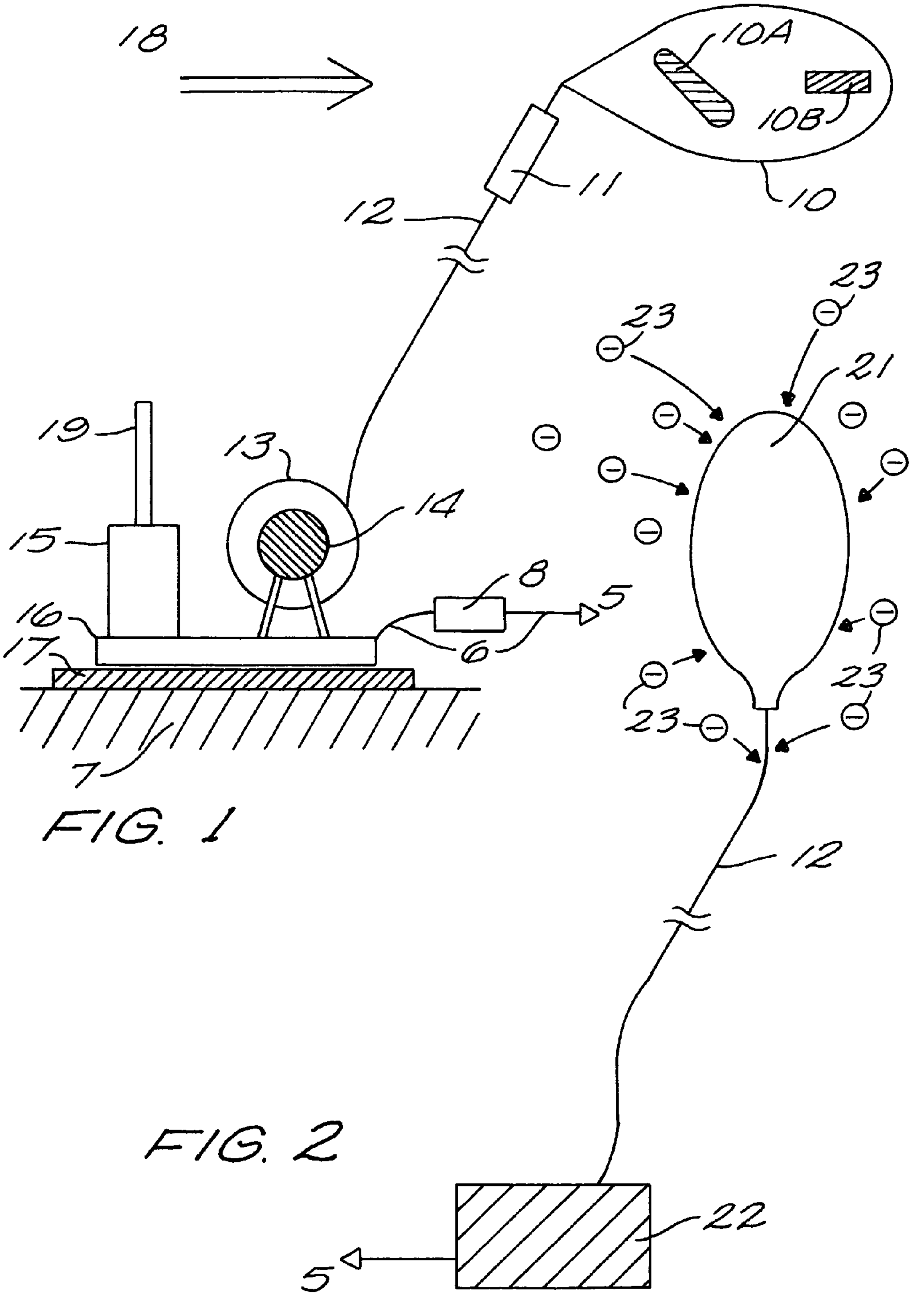
18
10A
10B
10
11
12
19
13
14
15
8
5
16
6
17
7
23
23
21
23
23
23
12
22
5

FIG. 1

FIG. 2

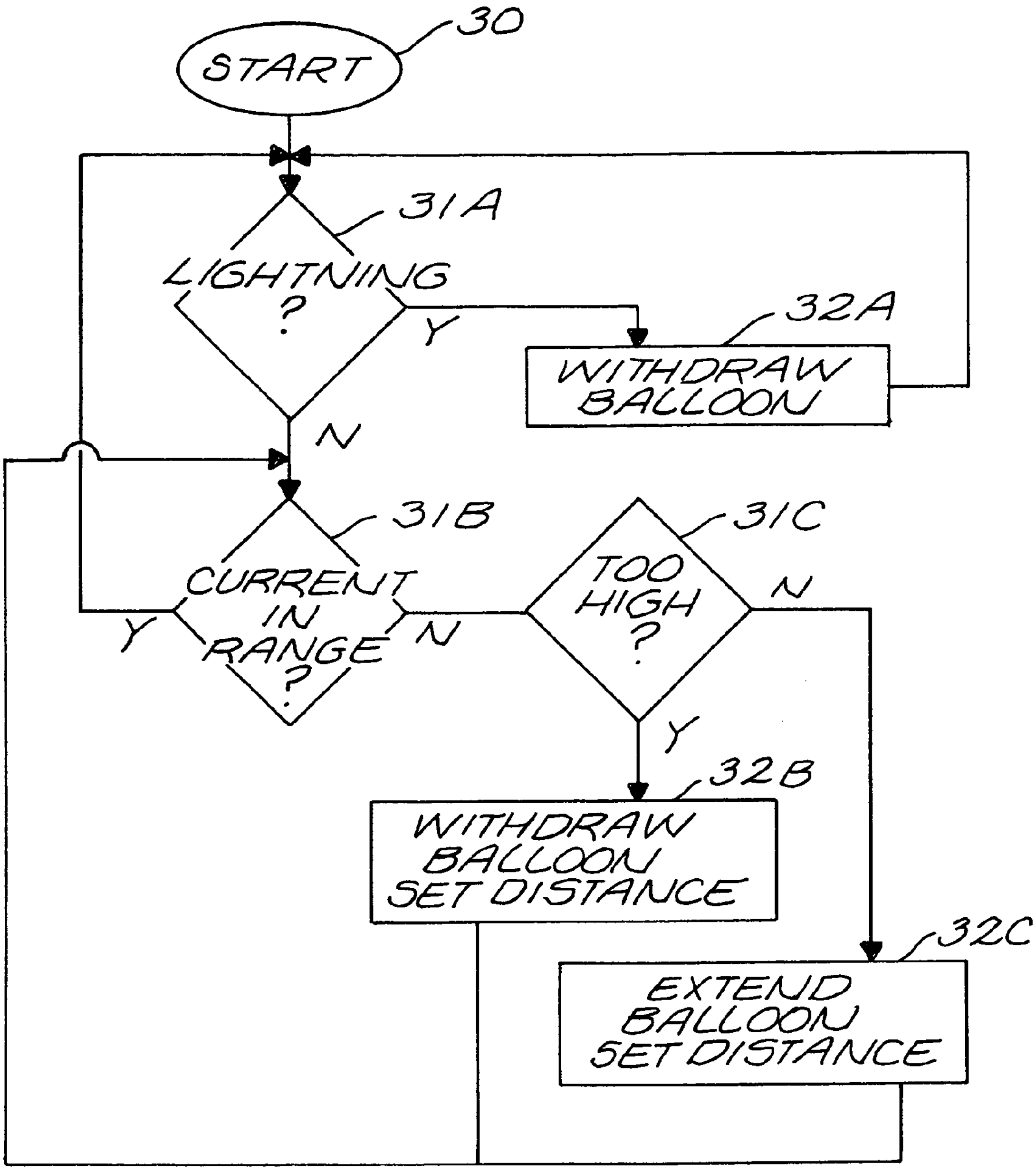
30
START
31A
LIGHTNING ?
Y
32A
WITHDRAW BALLOON
N
31B
CURRENT IN RANGE ?
Y
N
31C
TOO HIGH ?
N
Y
32B
WITHDRAW BALLOON SET DISTANCE
32C
EXTEND BALLOON SET DISTANCE

FIG. 3

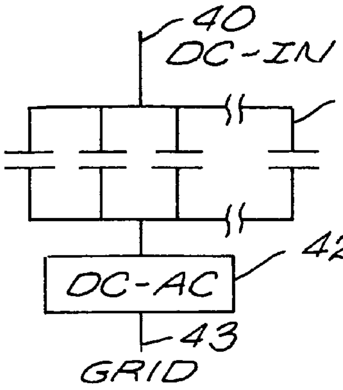
40
DC-IN
41
42
DC-AC
43
GRID

FIG. 4A

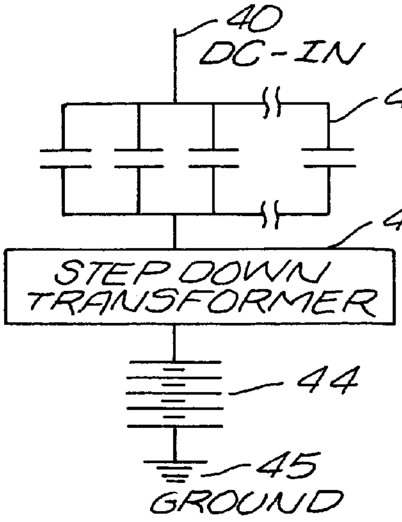
40
DC-IN
41
43
STEP DOWN TRANSFORMER
44
45
GROUND

FIG. 4B

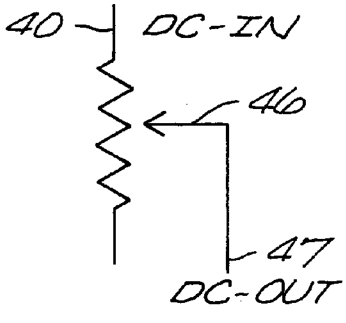
40
DC-IN
46
47
DC-OUT

FIG. 4C

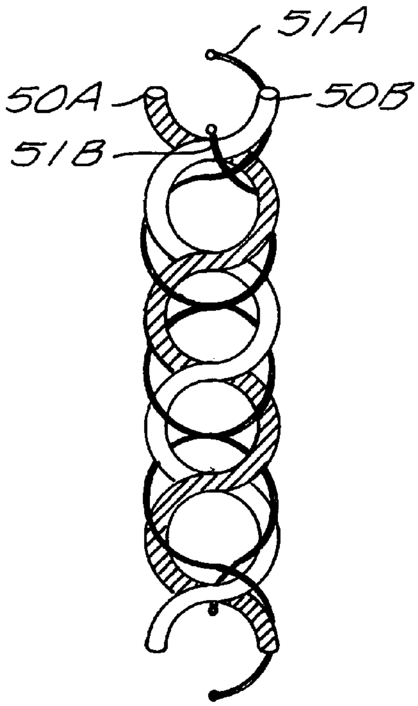
51A
50A
50B
51B

FIG. 5

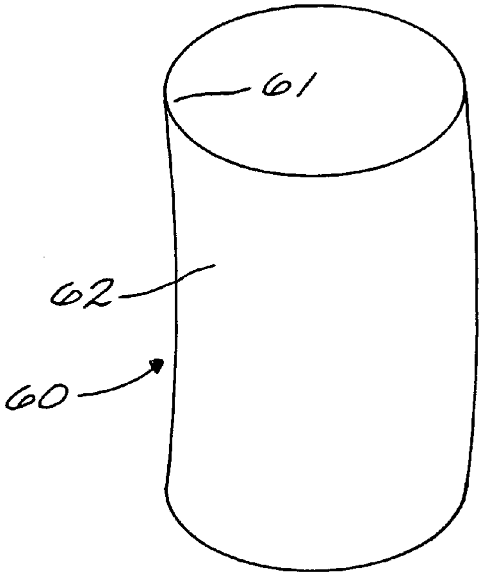

FIG. 6A

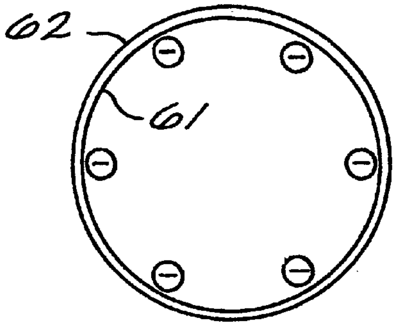

FIG. 6B

REPLACEMENT

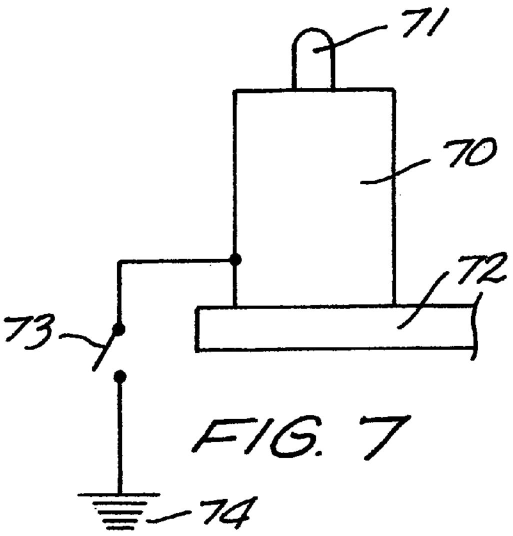
71
70
72
73
74

FIG. 7

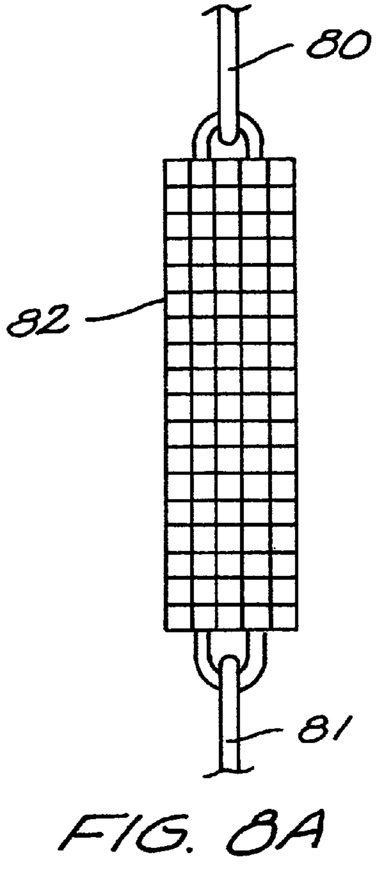

FIG. 8A

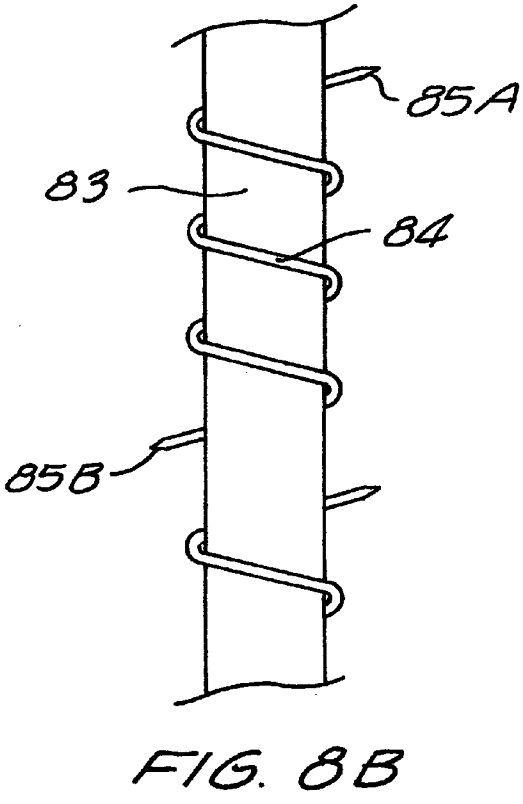
85A
83
84
85B

FIG. 8B

ATMOSPHERIC ELECTRICAL GENERATOR

BACKGROUND OF THE INVENTION

This invention relates to the production of electrical energy and more particularly to the production of electrical energy from the atmosphere.

Everyone is familiar with Benjamin Franklin's kite experiment of 1752. Using a kite whose string had become wet, negative charges from the passing clouds flowed into the string, down to the suspended key, and then into a Leyden jar via a thin metal wire. Franklin was protected by a dry silk string; but, when Franklin's knuckle came too close to the key, he received a strong shock. Fortunately, Benjamin Franklin was not killed, others who tried this same experiment were not so lucky.

Since then, the formation of lightning has remained something of a mystery. Lightning bolts are triggered when a negatively charged cloud base induces a positive charge from the ground, thereby forming a "pathway" for the discharge of the collected electrical energy.

Lightning travels up to 60,000 miles per hour with a flash that is brighter than ten million 100-watt lightbulbs. This wattage is as much power as is produced by all of the electricity plants in the United States and with a voltage of up to 300 million volts.

It is this very fact, the power within lightning is immense, that has prevented any successful collection of the electrical energy from lightning. The electricity in lightning is far too extreme for current technology to harness.

While lightning has attracted a energy starved industrial world, no one has developed any technique to harness this naturally occurring electrical source.

It is clear there is a continuing need for an electrical source other than carbon-based fuels and that the naturally occurring electricity in the atmosphere is being ignored.

SUMMARY OF THE INVENTION

The invention is a mechanism which taps into the naturally occurring static electricity in the atmosphere. Whereas heretofore, the attempt to garner electricity from the atmosphere has focused exclusively on capturing lightning, the present invention syphons off the static electricity which is generated from any agitated air and avoids lightning.

Lightning is only the final discharge of the static electricity, whether that lightning is intra-cloud lightning, cloud-to-ground lightning, or inter-cloud lightning. Other types of final discharges are known as heat lightning, summer lightning, sheet lightning, ribbon lightning, silent lightning, ball lightning, bead lightning, elves, jets, and sprites. Well before these discharges are observed, as the atmosphere becomes agitated by wind or thermal activity, static electricity is being generated.

The present invention recognizes that this static electricity is being formed and creates a mechanism to capture it.

The mechanism of this invention utilizes an aircraft such as a lighter than air balloon. While the preferred embodiment uses a foil balloon, a variety of other aircraft are obvious to those of ordinary skill in the art, including, but not limited to: gliders, rubber balloons (such as weather balloons), biaxially-oriented polyethylene terephthalate polyester film balloons, and latex balloons.

Within this discussion, the balloon is referenced, but, the invention is not intended to be limited solely to balloons.

The balloon is sent aloft and is tethered by a conductive line. In this context, the conductive line may be any obvious to those of ordinary skill in the art. For the preferred embodiment, the conductive line is a generically referred to as a "poly-rope" and is commercially available through a variety of sources. A suitable conductive line is described in U.S. Pat. No. 5,203,542, entitled "Apparatus for Improved Electric Fence Wire Construction for use with Intensive Grazing" issued Apr. 20, 1993, to Coley, et al. and incorporated hereinto by reference.

The conductive line is played out of a winch to control the altitude of the balloon. The motor controlling the winch is able to reverse direction to both extend and withdraw the conductive line which is wrapped around a spool on the winch. The winch/spool combination are part of a base unit.

In some embodiments of the invention, the spool is constructed of rubber so as to insulate the conductive line from the winch assembly. In this embodiment, only the conductive line is charged by the atmospheric static electricity while the winch remains neutral.

In yet another embodiment, the winch/spool are part of a base unit which is itself isolated from the ground by an insulator. In this embodiment, the entire base unit is charged by the atmospheric static electricity.

A conductor, such as an insulated wire, is electrically connected to the conductive line. In one embodiment, where the conductive line is electrically isolated from the spool and winch motor, the conductor is connected to the conductive line. In the embodiment where the conductive line is electrically connected to the base unit, then the conductor is connected anywhere on a metallic base unit.

The other end of the conductor is connected to a load. The load in this case can be any of a variety of electrical loads well known to those of ordinary skill in art, including, but not limited to a motor, a battery system, or the electrical grid for the system.

In the preferred embodiment, a sensor array is used to monitor the activities both at the base unit (such as electrical flow within the conductor) and in the surrounding locale.

A sensor monitoring the electrical flow (i.e. voltage and/or current) within the conductor is used to monitor the electrical activity within the conductor.

In the preferred embodiment, a lightning sensor monitors for lightning activity within the locale. As noted earlier, the electrical characteristic of lightning is so extreme that ideally this discharge is avoided as it might damage the mechanism of this invention.

The sensor array is utilized by a controller, such as microprocessor, programmed to operate the mechanism as outlined herein.

The controller operates the winch motor to extend or withdraw the conductive line and by extension the altitude of the balloon. The controller is programmed to operate the winch by monitoring the electrical characteristics of the conductor and adjusting the balloon's altitude to maintain these characteristics within the conductor within a preset range.

This preset range is established either in the base programming of the controller or is established by an operator of the system.

As example, by controlling the amount of current being withdrawn from the atmosphere, the mechanism operates within a safe range and also provides a relatively stable current flow from which a variety of activities can take place (such as DC-AC conversion).

The controller also utilizes the lightning sensor to protect the mechanism from a lightning strike. Should lightning be detected within a pre-determined range (as established by the software or defined by an operator), then the balloon is pulled down to minimize the risk of damage from a lightning strike.

The invention, together with various embodiments thereof will be more fully explained by the following description of the accompanying drawings.

DRAWINGS IN BRIEF

FIG. **1** diagrams the preferred embodiment of the invention.

FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

DRAWINGS IN DETAIL

FIG. **1** diagrams the preferred embodiment of the invention.

Balloon **10** is an aircraft which, in this illustration, is a lighter than air balloon. Wings **10**A, extending from the body of balloon **10**, provide additional lift in air flow **18**. Tail **10**B helps to stabilize balloon **10**.

Balloon **10** is tethered to the ground via conductive line **12**. As noted earlier, a variety of configurations and materials are available to serve as conductive line **12**. In this illustration, a poly-wire is used Poly-wire is commercially available through a variety of vendors, including, but not limited to: Jeffers Livestock and Zareba Systems, Inc, of Ellendale, Minn.

In this embodiment, located proximate to balloon **10**, is an electrical collection enhancement lead **11** which assists in the collection of the static electrical charge in the atmosphere. Electrical collection enhancement lead **11** is configured to attract the static charge and conduct the charge into the conductive line **12**.

The electricity flows down the conductive line into spool **13**, where the conductive line **12** is collected and either withdrawn or dispensed through operation of winch motor **14**.

Winch motor **14** and spool **13** are mounted onto base unit **16** which is electrically isolated from ground **7** using insulator **17**. Note, in this embodiment of the invention, when electricity is being collected from the atmosphere, the entire base unit **16** becomes charged. In another embodiment of the invention, spool **13** is constructed of rubber, thereby preventing base unit **16** from becoming charged, thereby restricting the charging from the atmosphere to only conductive line **12**.

In this embodiment, conductor **6** is connected to base unit **16** (since the entire base unit **16** is charged and the base unit is metallic) to communicate the electrical current to load **5**. Conductor **6** is ideally an insulated wire.

The electrical current through conductor **6** is measured using sensor **8**.

In the alternative embodiment discussed above, where only the conductive line **12** is charged, then conductor **6** is connected to conductive line **12**.

Controller **15**, located in this embodiment on base unit **16**, operates winch motor **14** in response to signals from sensor **8** (measuring the current being discharged to load **5**) to maintain the current flow within a pre-defined range. As the current flow diminishes, then the conductive line **12** extended from spool **13** to increase the altitude of balloon **10** to that more static charge from the atmosphere is gathered; as the current flow falls exceeds a preset level, conductive line **12** is withdrawn onto spool **13** to decrease the static charge being collected from the atmosphere.

The range of current flow through conductor **6** is ideally set by the program, although some embodiments of the invention permit an operator to establish this range of operation.

In an alternative embodiment, the sensor monitoring conductor **6** monitors the voltage therein.

In the preferred embodiment of the invention, controller **15** is also equipped with a lightning sensor **19**. In this embodiment, when lightning is sensed within a preset range, then substantially all of conductive line **12** is wound onto spool **13** to pull balloon **10** near the ground and protect the entire mechanism from being damaged from a lightning discharge.

In the preferred embodiment, the "safe" distance form lightning is set in the programming of controller **15** and is ideally two miles; other embodiments permit the operator to "safe" distance.

There are a variety of lightning sensors well known to those of ordinary skill in the art, including, but not limited to those described in: U.S. Pat. No. 7,016,785, entitled "Lightning Detection" issued to Makela, et al. on Mar. 21, 2006; U.S. Pat. No. 6,829,911, entitled "Lightning Detection and Prediction Alarm Device" issued to Jones, et al. on Dec. 7, 2004; U.S. Pat. No. 7,200,418, entitled "Detection of Lightning" issued to Karikuranta, et al. on Apr. 3, 2007; and U.S. Pat. No. 6,961,662, entitled "Systems and Methods for Spectral Corrected Lightning Detection" issued to Murphy on Nov. 1, 2005; all of which are incorporated hereinto by reference.

In another embodiment of the invention, controller **15** is not located on base unit **16**, rather it is remote and communicates its control signals to winch motor **14** using radio waves.

FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

Static charges **23** are generated in the atmosphere by agitated air. These static charges are often collected at the bottom of clouds, but exist in other environments as well.

Balloon **21** is extended into this strata of static charges **23** which are then attracted to conductive line **12** to flow to base unit **22** and then onto load **5**.

By increasing or decreasing the altitude of balloon **21** (defined by the length of the extended conductive line **12**), conductive line **12** is selectively exposed to varying densities and levels of the static charge strata, and by extension, the current flow or voltage is increased or decreased.

FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

Once the program starts **30**, the lightning sensor is checked to determine if lightning has occurred within the unsafe range **31**A, if it has, then the balloon is lowered **32**A, and the program continues monitoring the status of lightning until no lightning is detected.

When the lightning status is acceptable, then the current within the conductor is checked to see if the current is within the prescribed range **31**B. If the current is acceptable (within range) the program returns to check the lightning status **31**A; otherwise a determination is made to see if the current is above the prescribed range **31**C.

If the current is above the prescribed range, then the altitude of the balloon is withdrawn a set amount **32**B (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

If the current is below the prescribed range, then the altitude of the balloon is extended a set amount **32**C (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

In this manner of feed-back and minor adjustments in the altitude of the balloon, the current is maintained within a prescribed range which can be handled by the downstream electrical system.

As noted earlier, some embodiments of the invention monitor the voltage instead of the current.

FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

By maintaining the voltage being collected in a prescribed range, an electrical conversion system is easily designed. While FIGS. **4**A, **4**B, and **4**C illustrate some electrical configurations, those of ordinary skill in the art readily recognize a variety of other configurations which will serve the same function.

Referencing FIG. **4**A, Direct Current In (DC IN) **40** is buffered by a gang of capacitors **41** before being communicated to a DC/AC converter **42**. The DC/AC converter converts the direct current into a an alternating current suitable for placement over an existing electrical grid **43** such as normally found from a power-plant.

Those of ordinary skill in the art readily recognize a variety of DC/AC converters, including, but not limited to: U.S. Pat. No. 7,394,671, entitled “Controller IC, DC-AC Conversion Apparatus, and parallel running system of DC-AC Conversion Apparatuses” issued to Fukumoto, et al. on Jul. 1, 2008; and, U.S. Pat. No. 7,330,366, entitled “DC-AC Converter” issued to Lee, et al. on Feb. 12, 2008; all of which are incorporated hereinto by reference.

FIG. **4**B illustrates an electrical arrangement suitable for use in charging a battery. DC IN **40** is buffered by capacitor bank **41** before entering into a step down transformer **43**. Step down transformer **43** reduces the voltage so that the voltage can safely be introduced into battery **44** which is connected to ground **45** at the battery’s other pole.

Those of ordinary skill in the art readily recognize a variety of batteries which will work in this capacity, including, but not limited to those described in: U.S. Pat. No. 7,378,181, entitled “Electric Storage Battery Construction and Manufacture” issued to Skinlo on May 27, 2008; U.S. Pat. No. 7,388,350, entitled “Battery with Electronic Compartment” issued to Wright on Jun. 17, 2008; U.S. Pat. No. 7,397,220, entitled “Connection Member and Battery Pack” issued to Uchida, et al. on Jul. 8, 2008; and, U.S. Pat. No. 7,375,492, entitled “Inductively Charged Battery Pack” issued to Calhoon, et al. on May 20, 2008; all of which are incorporated hereinto by reference.

In FIG. **4**C, DC IN **40** is fed into an adjustable rheostat **46** which is controlled by the controller so that the DC OUT **47** falls within a specified range.

FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

This type of conductive line is commonly called poly-wire and consists of multiple interwoven strands of plastic **50**A and **50**B woven into a cord or rope arrangement having intertwined therein exposed metal wires **51**A and **51**B. While this illustration shows two plastic strands and two metal wires, any number of possible combinations is possible.

The exposed metal wires **51**A and **51**B attract the atmospheric static charge and transmit the charge down to the base unit (not shown).

FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

This conductive line utilizes a tube **60** having an outer layer **62** of PET Film (Biaxially-oriented polyethylene terephtalate polyester film) which provides exceptionally high tensile strength and is chemically and dimensionally stable. The tube has an ideal diameter of between two and three inches.

An interior metal coating **61** provides an initial conduit for the flow of static charge. The static charge through the metal forces the tube to expand due to the repulsion experienced by like charges. Further, the flow of electricity causes the interior of the tube **60** to become ionized to provide an additional pathway for the atmospheric static charges to the base unit (not shown).

Because outer layer **62** provides a gas barrier, the resulting ionization is not dissipated by air currents, thereby providing a highly stable pathway.

FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

In this embodiment, controller box **70**, resting on insulating pad **72**, is in communication with the sensors as described above. Using the input from these sensors, when there is flow of electricity through the base unit, warning flashing light **71** is illuminated. To electrically neutralize the mechanism, switch **73** is activated to pass any existing current into the ground **74**.

FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

Referencing FIG. **8**A, enhanced electrical collection lead **82** is a wire mesh which is in electrical communication with conductive line **81** and balloon **80**. Because of the significant amount of metal exposed by enhanced electrical collection lead **82**, more static electricity from the atmosphere is drawn to the collection lead **82**, and then down conductive line **81** to the base unit (not shown).

Conductive lead **82** is positioned proximate to balloon **80**.

In FIG. **8**B, poly-wire **83** has enhanced electrical collection leads **84** wrapped therearound. Collection leads **84** have pointed ends **85**A and **85**B which have a propensity to attract more electricity than rounded ends do.

It is clear from the foregoing that the present invention captures an entirely new source of electrical energy.

What is claimed is:

1. A mechanism to tap an electrical source comprising:
- a) a lighter than air balloon suspended in the atmosphere;
- b) a base unit having a spool of conductive line on a winch motor, one end of said conductive line secured to said lighter than air balloon, a portion of said conductive line collecting electricity in the atmosphere, said winch motor capable of selectively extending or withdrawing said conductive line from said spool;
- c) an insulator electrically isolating said conductive line from ground;
- d) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load being powered by collected electricity from said conductive line;
- e) an electrical flow sensor monitoring electrical flow through said conductor and generating an electrical flow indicia indicative of said electrical flow in said conductor; and,

f) a controller receiving said electrical flow indicia and selectively operating said winch motor such that said electrical flow indicia remains within a selected operating range.

2. The mechanism to tap an electrical source according to claim **1**, wherein said selected operating range is established by an operator.

3. The mechanism to tap an electrical source according to claim **2**, further including a warning light activated when said electrical flow indicia is non-zero.

4. The mechanism to tap an electrical source according to claim **1**,

a) further including a lightning sensor generating a lightning presence indicia indicative of lightning within a prescribed range, said presence indicia being communicated to said controller; and,

b) wherein said control mechanism, in response to said lightning presence indicia, operates said winch motor to withdraw substantially all of said conductive line onto said spool.

5. The mechanism to tap an electrical source according to claim **4**, wherein said prescribed range is established by an operator.

6. The mechanism to tap an electrical source according to claim **1**, further including an electrical collection enhancement lead in electrical contact with a first end of said conductive material, said electrical collection enhancement lead configured to attract static electricity.

7. The mechanism to tap an electrical source according to claim **6**, wherein said electrical collection enhancement lead is positioned proximate to said lighter than air balloon.

8. The mechanism to tap an electrical source according to claim **7**, wherein said electrical collection enhancement lead includes at least two pointed electrical conductors.

9. The mechanism to tap an electrical source according to claim **1**, wherein said insulator electrically isolates said conductive line from said winch motor.

10. The mechanism to tap an electrical source according to claim **1**, wherein said insulator electrically isolates said base unit from ground.

11. A mechanism comprising:

a) an airborne aircraft having a conductive line secured to a winch capable of extending or withdrawing said conductive line from a spool, when said airborne aircraft is aloft, said conductive line collecting electricity from the atmosphere;

b) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load such that collected electricity from said conductive line powers said load;

c) an electrical flow sensor monitoring electrical flow through said conductor and generating an electrical flow indicia indicative of said electrical flow in said conductor; and,

d) a controller receiving said electrical flow indicia adjusting an altitude of said airborne aircraft via said winch such that said electrical flow indicia remains within a selected operating range.

12. The mechanism to tap an electrical source according to claim **11**,

a) further including a lightning sensor communicating with said controller; and,

b) wherein said controller, in response to selected signals from said lightning sensor withdraws substantially all of said conductive line.

13. A mechanism to tap an electrical source comprising:

a) a lighter than air balloon suspended in the atmosphere;

b) a base unit having a spool of conductive line on a winch motor, one end of said conductive line suspended by said balloon, a portion of said conductive line collecting electricity from the atmosphere, said winch motor capable of adjusting an altitude of said lighter than air balloon by selectively extending or withdrawing said conductive line from said spool;

c) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load being powered by said collected electricity from said conductive line;

d) a sensor array having,

1) an electrical flow sensor monitoring electrical flow through said conductor and generating an electrical flow indicia indicative of said electrical flow in said conductor, and,

2) a lightning sensor monitoring existence of proximate lightning; and,

e) a controller responsive to said electrical flow indicia from said sensor array to selectively operate said winch motor.

14. The mechanism to tap an electrical source according to claim **13**,

a) further including a warning light; and,

b) wherein said controller activates said warning light when said electrical flow in said conductor is non-zero.

15. The mechanism to tap an electrical source according to claim **14**, further including an electrical collection enhancement lead in electrical contact with said conductive material, said electrical collection enhancement lead configured to attract static electricity and positioned proximate to said lighter than air balloon.

* * * * *

US 20130093261A1

(19) **United States**

(12) **Patent Application Publication**

Allen et al.

(10) **Pub. No.: US 2013/0093261 A1**

(43) **Pub. Date: Apr. 18, 2013**

---

(54) **SYSTEM FOR HARVESTING ATMOSPHERIC ELECTRICITY**

(75) Inventors: **Edward H. Allen**, Lancaster, CA (US); **Sergey Macheret**, Lancaster, CA (US)

(73) Assignee: **LOCKHEED MARTIN CORPORATION**, Bethesda, MD (US)

(21) Appl. No.: **13/276,259**

(22) Filed: **Oct. 18, 2011**

Publication Classification

(51) **Int. Cl.**
H05F 3/06 (2006.01)

(52) **U.S. Cl.**
USPC **307/149**

(57) **ABSTRACT**

Systems, methods and apparatus for harvesting atmospheric electricity are provided. The system includes a laser configured to form a plasma filament and a collector configured to collect electricity flowing along the plasma filament. The plasma filament comprises an electrically conducting plasma filament. Atmospheric electricity may be collected by having the plasma filament form at least a part of a conducting path: (1) between ground and a cloud, (2) between differently charged regions of the same cloud, (3) between differently charged regions of different clouds, and (4) between different regions of atmosphere, where there is a vertical voltage gradient. When the plasma filament is not long enough to form the entire conducting path, a lightning may be triggered to complete the conducting path needed to collect atmospheric electricity.

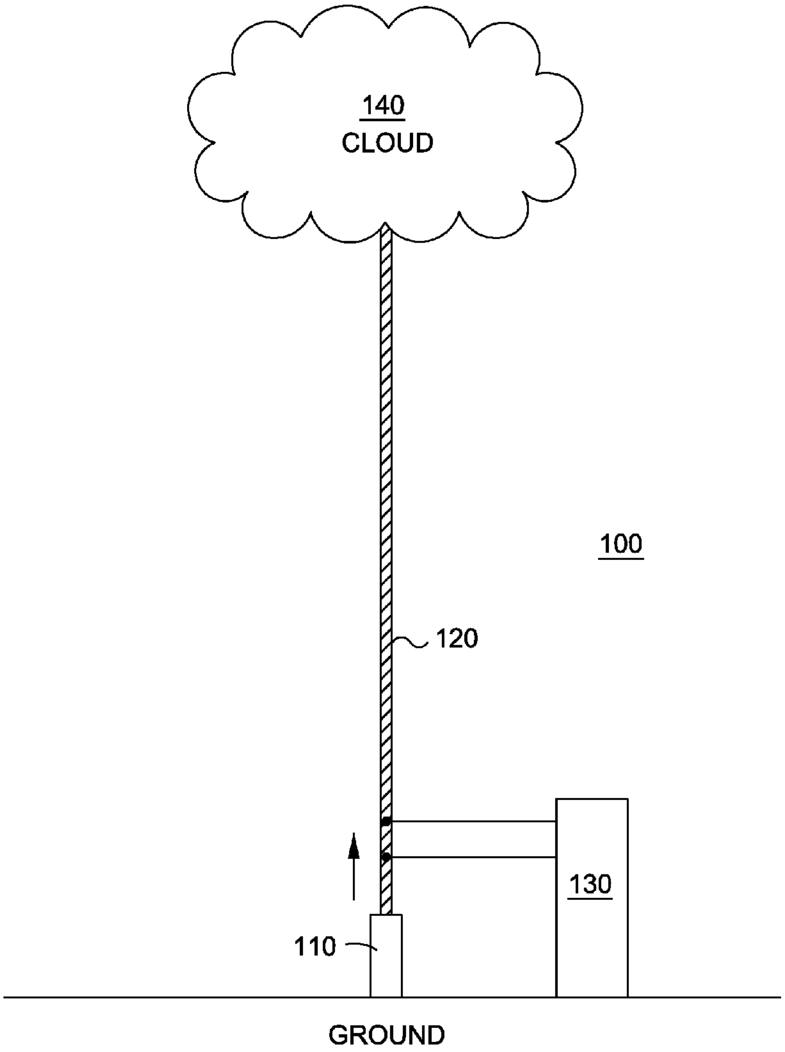

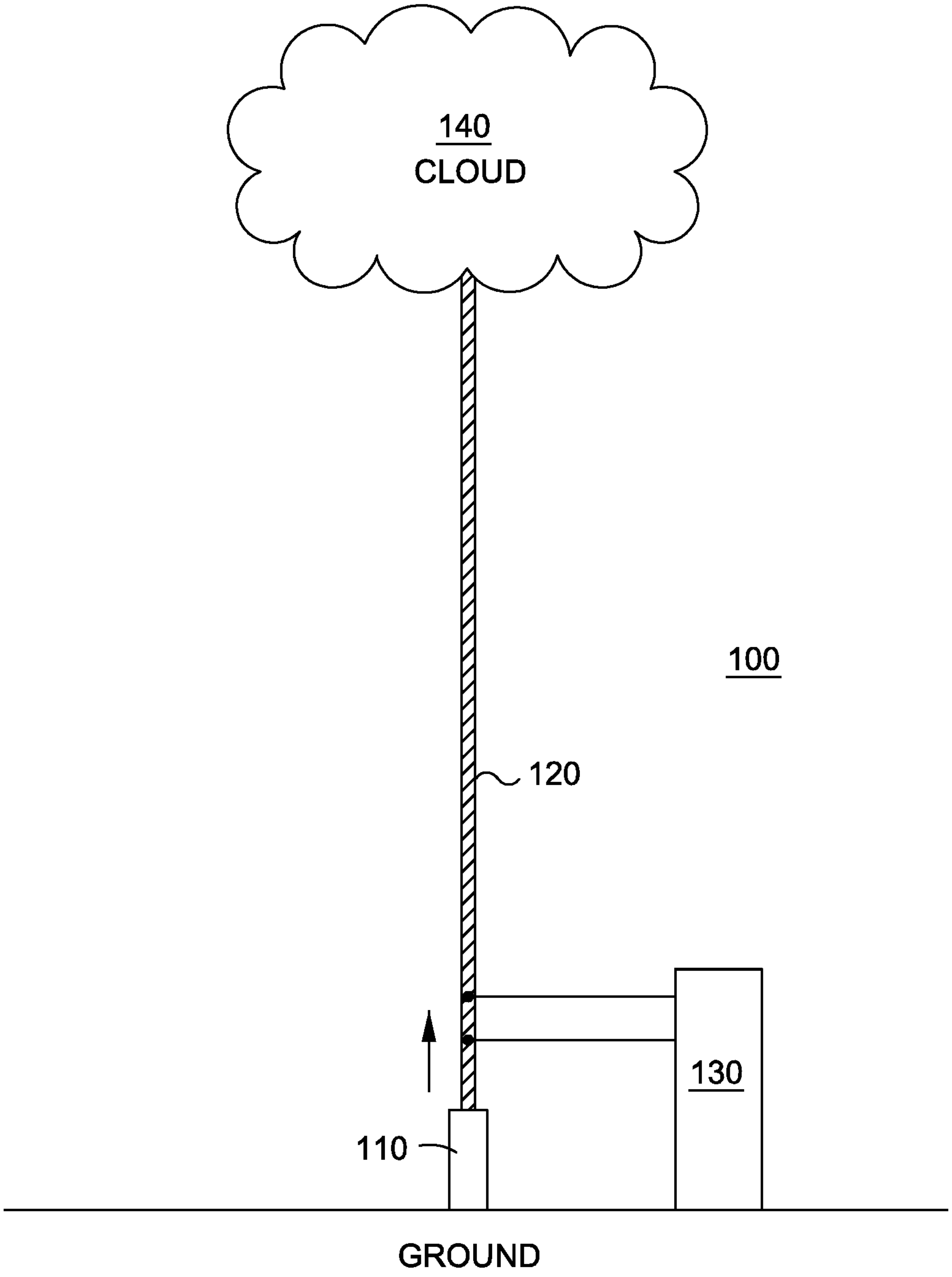
140
CLOUD
100
120
130
110
GROUND

FIG. 1

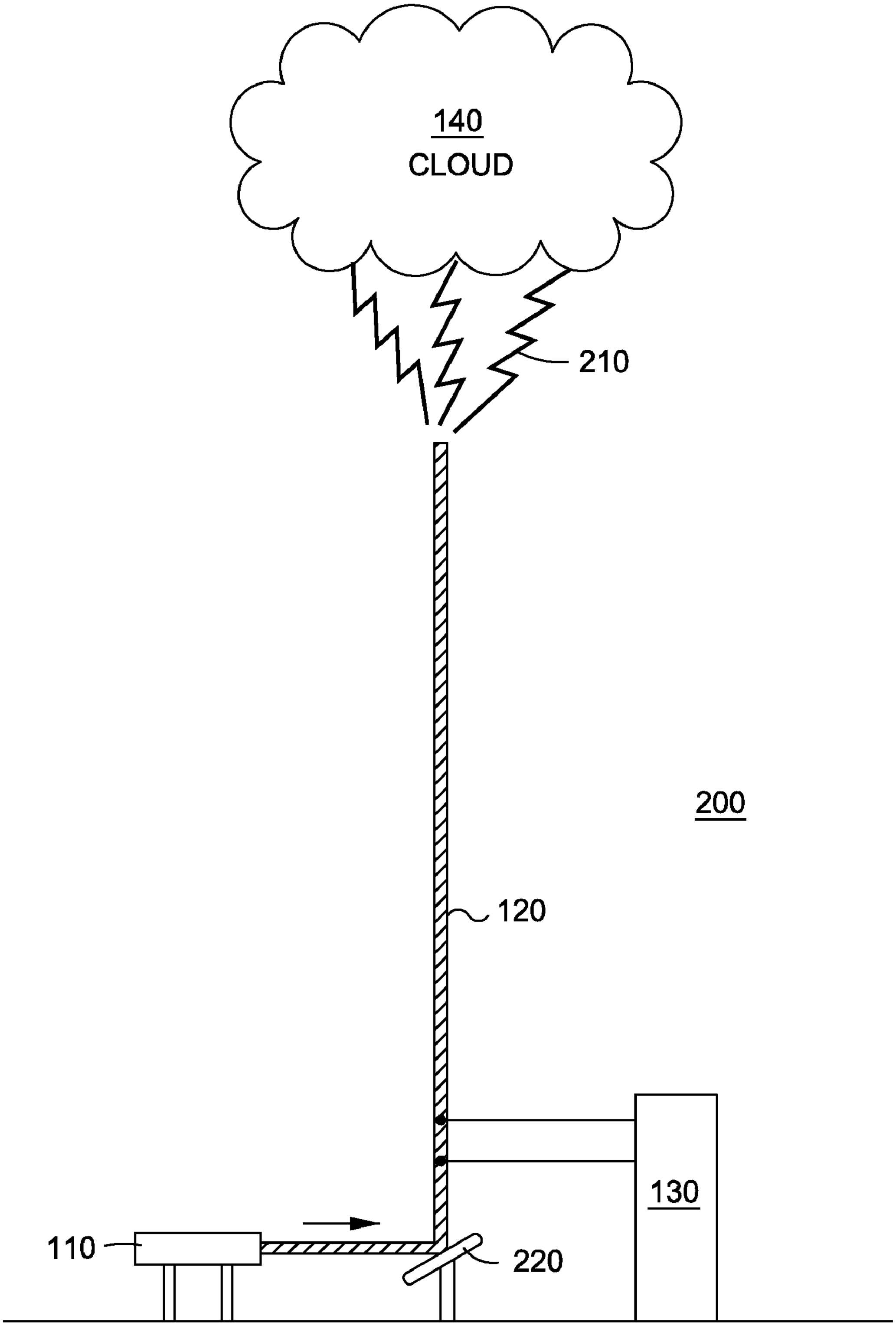
140
CLOUD
210
200
120
130
110
220
GROUND

FIG. 2

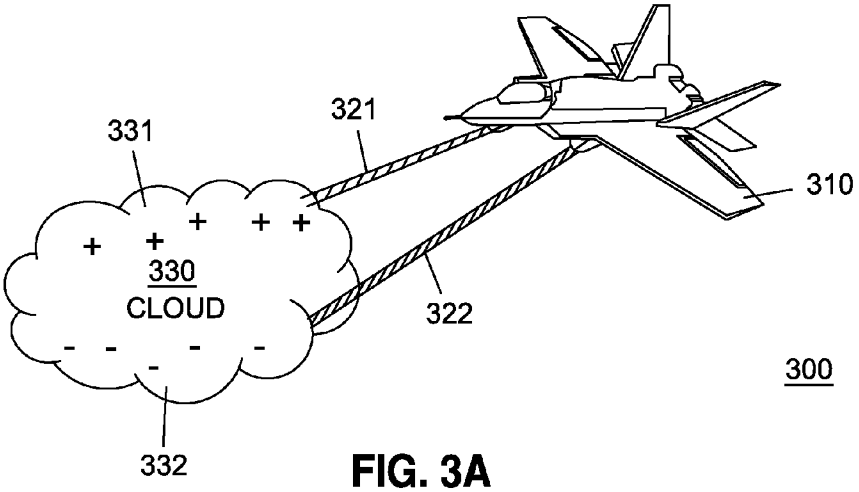
CLOUD

FIG. 3A

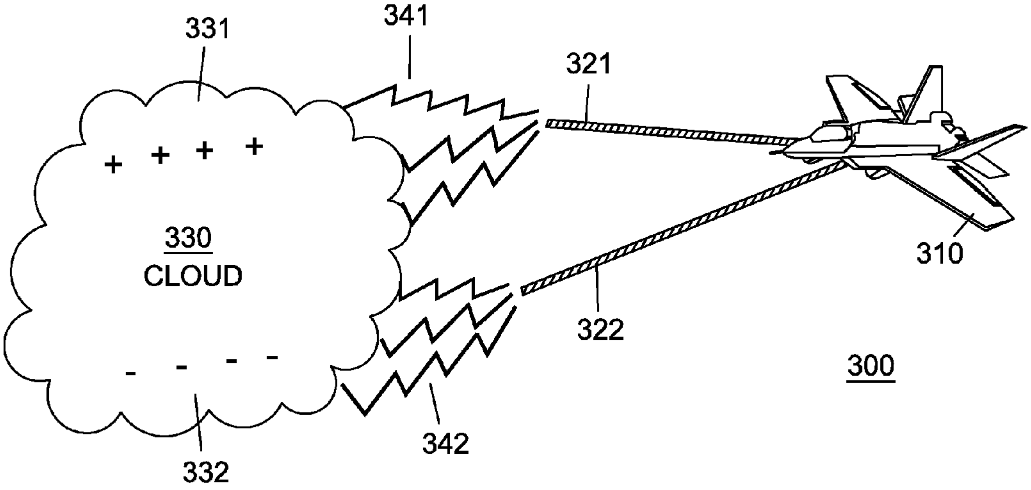
CLOUD

FIG. 3B

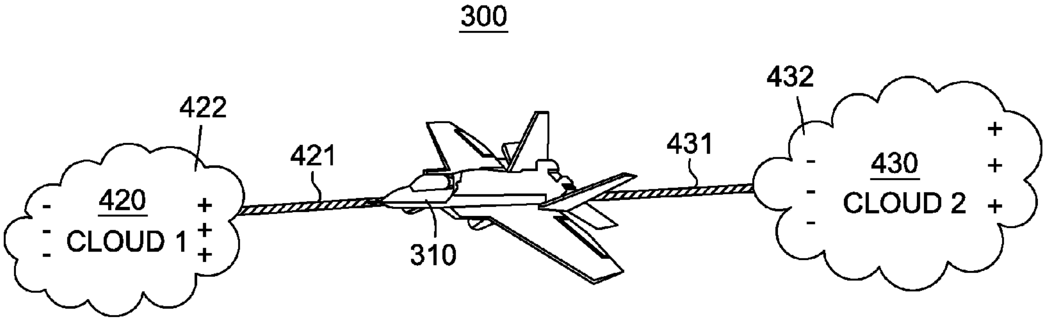
300
422
421
431
432
420
CLOUD 1
430
CLOUD 2
310
- - -
+ + +
- - - -
+ + + +

FIG. 4

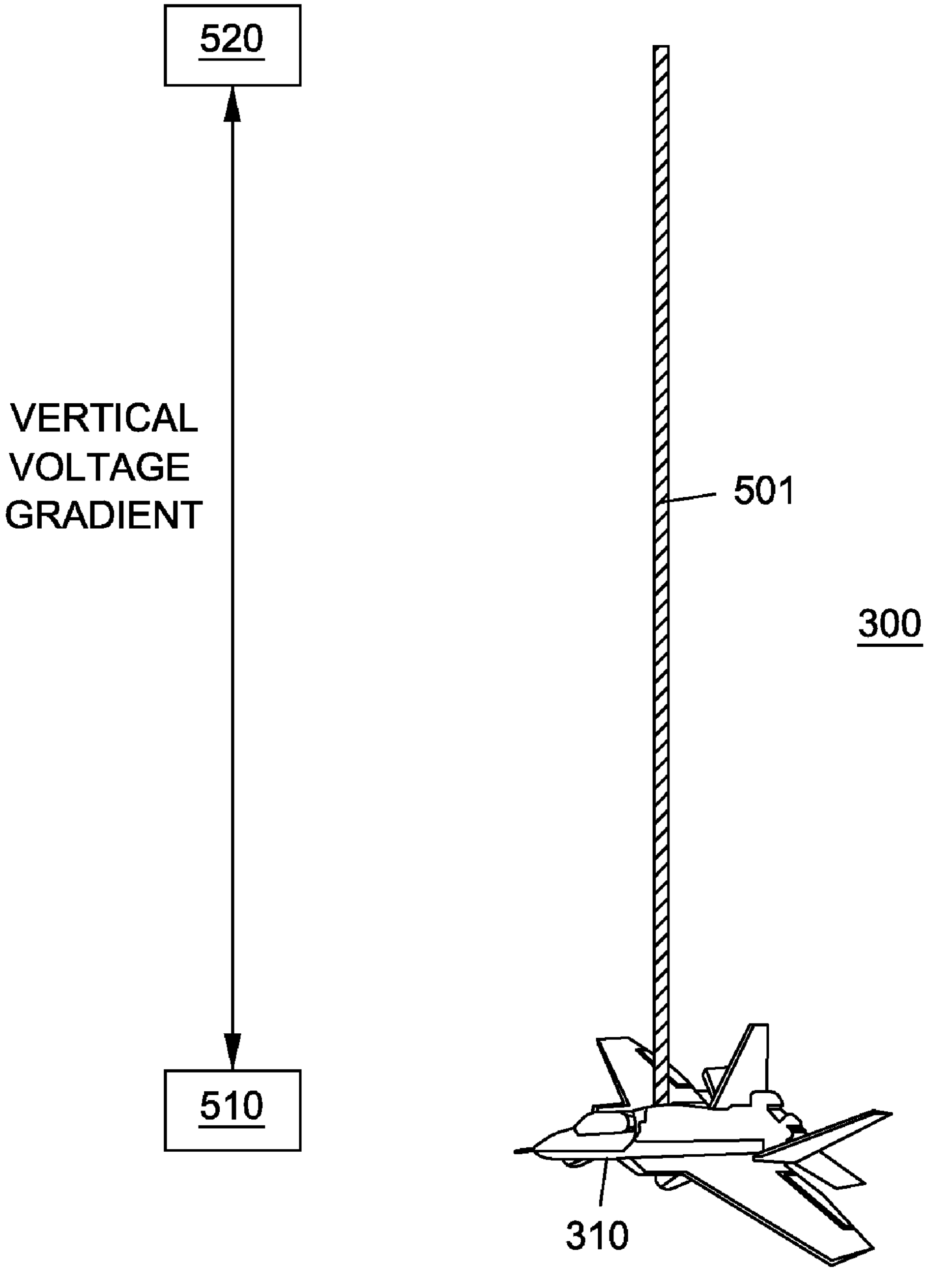
520
VERTICAL
VOLTAGE
GRADIENT
510
501
300
310

FIG. 5

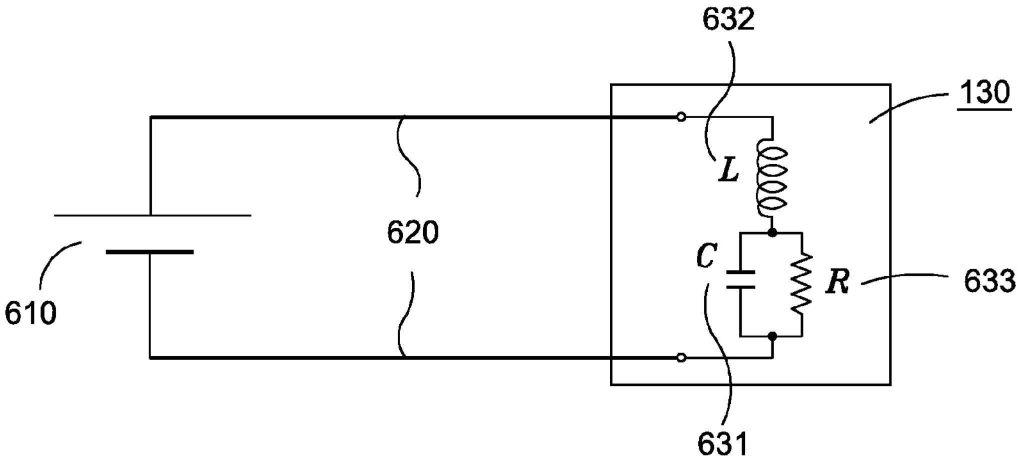
632
130
L
620
C
R
633
610
631

FIG. 6

SYSTEM FOR HARVESTING ATMOSPHERIC ELECTRICITY

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0001] Not applicable.

FIELD OF THE INVENTION

[0002] The subject technology relates in general to a system for harvesting atmospheric electricity, and more particularly to a system that uses a laser to form an electrically conducting plasma filament for harvesting atmospheric electricity.

BACKGROUND

[0003] It is well known that large quantities of electrical energy are present in the atmosphere and in lightning. Lightning was one of the first forms of electricity harnessed in the modern age by Benjamin Franklin in his famous kite experiment.

[0004] A lightning discharge typically contains on the order of le Joules of energy. It has been estimated that the total electrical power of lightning across the Earth is of the order of 10^{12} watts. Collection of electrical energy from clouds associated with lightning is a problem that needs to be resolved.

[0005] Lightning is but a small part of the total electrical activity of the atmosphere. When a local build up of charge above the Earth exceeds the local breakdown potential of the atmosphere a lightning discharge occurs. However, there is a continual invisible flow of charge from Ionosphere to Earth occurring day and night over the entire surface of the globe, which exceeds the global lightning power output by many times. The source of this flow of charge from the atmosphere to Earth is the Sun. Radiation from the Sun helps form most of the ions found in the Ionosphere, a highly charged region above the atmosphere, through ionization of atmospheric molecules. Solar radiation, particularly in the UV and soft x-ray bands, consist of photons whose energies are well suited to ionization. The Sun also emits continuously a solar wind of positively charged particles. These are captured by the Earth's magnetic field and further contribute to the Ionosphere. The positively charged region in the Ionosphere in turn induces (by electrostatic induction) a negative charge on the surface of the Earth. The Earth becomes in effect an enormous spherical capacitor. A potential gradient or electric field is thus established between the two "plates" of this capacitor, the Ionosphere (or Electrosphere) and the surface of the Earth. While the upper strata of the atmosphere conduct electricity reasonably well, the lower levels act as an insulator or dielectric. Near the surface of the Earth, this electrostatic potential gradient is on the order of about 100 Vm^{-1} in summer, rising to 300 Vm^{-1} in winter. This flow of charge can be tapped and directed to provide useable electrical power. This source of atmospheric electricity has the following advantages: (1) Simple and robust technology; (2) Low cost technology—much cheaper than photovoltaics or wind turbines; (3) Available day and night in all weather conditions—in fact, more power is produced at night than during the day; and (4) Available at any point on the Earth's surface. Collection of this source of atmospheric electricity is another problem that needs to be resolved.

SUMMARY

[0006] According to various aspects of the subject technology, a system for harvesting atmospheric electricity is provided that solves some or all of the foregoing problems. In some aspects, the system may be used to harvest atmospheric electricity arising from an electrostatic potential gradient: (1) between ground and a cloud, (2) between differently charged regions of the same cloud, (3) between differently charged regions of different clouds, and (4) between different regions of atmosphere, where there is a vertical voltage gradient. In other aspects, the system may be used to harvest atmospheric electricity arising from an electrostatic potential gradient between the Ionosphere (or Electrosphere) and the surface of the Earth.

[0007] According to various aspects of the subject technology, a system for harvesting atmospheric electricity is provided. The system comprises a laser configured to form a plasma filament. The system also comprises a collector configured to collect electricity flowing along the plasma filament.

[0008] According to various aspects of the subject technology, a method for harvesting atmospheric electricity is provided. The method comprises forming a plasma filament with a laser. The method also comprises collecting electricity flowing along the plasma filament.

[0009] According to various aspects of the subject technology, an apparatus for harvesting atmospheric electricity is provided. The apparatus comprises means for forming a plasma filament with a laser. The system also comprises means for collecting electricity flowing along the plasma filament.

[0010] It is understood that other configurations of the subject technology will become readily apparent to those skilled in the art from the following detailed description, wherein various configurations of the subject technology are shown and described by way of illustration. As will be realized, the subject technology is capable of other and different configurations and its several details are capable of modification in various other respects, all without departing from the scope of the subject technology. Accordingly, the drawings and detailed description are to be regarded as illustrative in nature and not as restrictive.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] The accompanying drawings, which are included to provide further understanding of the subject technology and are incorporated in and constitute a part of this specification, illustrate aspects of the subject technology and together with the description serve to explain the principles of the subject technology.

[0012] FIG. **1** illustrates a first example of a system for harvesting atmospheric electricity. The system comprises a laser configured to form a plasma filament and a collector configured to collect electricity flowing along the plasma filament. The plasma filament forms a conducting path between ground and a cloud.

[0013] FIG. **2** illustrates a second example of a system for harvesting atmospheric electricity. A plasma filament and a lightning together form a conducting path between ground and a cloud. A condensing mirror directs the beam of laser towards the cloud.

[0014] FIG. **3**A illustrates a third example of a system for harvesting atmospheric electricity. A laser system situated on an airplane forms two plasma filaments that together complete a conducting path between a first region of a cloud and a second region of the cloud. The first region of the cloud and the second region of the cloud carry opposite electric charges. A collector configured to collect electricity flowing along the two plasma filaments is also situated on the same airplane as the laser system.

[0015] FIG. **3**B illustrates a fourth example of a system for harvesting atmospheric electricity. It is similar to FIG. **3**A, except two lightnings complete the conducting path between a first region of a cloud and a second region of the cloud (i.e., a first lightning electrically connects the first region of the cloud to one of the plasma filaments and a second lightning electrically connects the second region of the cloud to the other plasma filament).

[0016] FIG. **4** illustrates a fifth example of a system for harvesting atmospheric electricity. A laser system situated on an airplane forms two plasma filaments that together complete a conducting path between a first cloud and a second cloud. The region of the first cloud that is connected to the first plasma filament and the region of the second cloud that is connected to the second plasma filament carry opposite electric charges.

[0017] FIG. **5** illustrates a sixth example of a system for harvesting atmospheric electricity. A laser situated on an airplane forms a plasma filament that completes a conducting path between a first region of atmosphere and a second region of atmosphere. A collector situated on the same airplane as the laser can collect electricity flowing along the plasma filament, because a vertical voltage gradient exists between the first region of atmosphere and the second region of atmosphere.

[0018] FIG. **6** illustrates an example of a collector configured to collect electricity flowing along the plasma filament. The collector may be a charge collection circuit characterized by capacitance, inductance, and resistance. The electricity may be collected by charging a capacitor.

DETAILED DESCRIPTION

[0019] The detailed description set forth below is intended as a description of various configurations of the subject technology and is not intended to represent the only configurations in which the subject technology may be practiced. The appended drawings are incorporated herein and constitute a part of the detailed description. The detailed description includes specific details for the purpose of providing a thorough understanding of the subject technology. However, it will be apparent to those skilled in the art that the subject technology may be practiced without these specific details. In some instances, well-known structures and components are shown in block diagram form in order to avoid obscuring the concepts of the subject technology. Like components are labeled with identical element numbers for ease of understanding.

[0020] FIG. **1** illustrates a first example of a system **100** for harvesting atmospheric electricity. The system **100** comprises a laser **110** configured to form a plasma filament **120** and a collector **130** configured to collect electricity flowing along the plasma filament **120**. The plasma filament **120** forms a conducting path between ground and a cloud **140**.

[0021] The plasma filament **120** comprises an electrically conducting plasma filament. The plasma comprises of electrons and positive ions. The electrically conducting plasma filament may be formed in atmosphere using an ultra short pulse laser (USPL). The USPL forms a plasma filament by raising the conductivity of a very small diameter channel (~100 microns or less) for a very brief time. The propagation of high power (~10^{14} W/cm^2) pulses from the USPL is accompanied by filamentation—self-channeling of femtosecond laser pulses in stable high-intensity light filaments with ~100 microns diameters (so that the Kerr effect focusing balances defocusing due to plasma formation). This filamentation keeps the beam virtually free of diffraction divergence. The long plasma strings have high electron densities (10^{16}/cm^3) which support high conductivities. The ionization of air at these intensities results from multi-photon processes. Filamentation has been produced and detected at altitudes as high as 13-20 km. Almost no beam scattering occurs in filamentation phenomenon, and virtually all the pulse energy is invested in ionization of air molecules. With the typical electron density and channel diameter, a 160 mJ pulse is sufficient to generate a 1 km long channel.

[0022] Once initiated, plasma filaments cannot go on forever, and typically only propagate a kilometer or so before energy depletion and exhaustion. However, by using two coaxial USPL projection optics having different time-lensing parameters, a near and far filament could be initiated in a "daisy-chain" to extend the total length of the conductive channel. The onset of filamentation in the far filament is time-lensed to coincide with the exhaustion range of the near filament. Additionally, using negative chirp, femtosecond laser pulses can propagate almost without loss until different frequency components of the wave packet, propagating at different speeds due to dispersion in air, come together at a certain desired location so that the laser intensity there exceeds the filamentation threshold, resulting in long ionized filaments created at a distance of up to several miles from the laser source. In other words, negative chirp is used to essentially focus a beam of the laser at a desired distance to form the plasma filament there. Accordingly, this technique may be used to move the starting point and to extend the length of the plasma filament.

[0023] The size, electrical conductivity, and lifetime of the plasma filaments are enhanced by an auxiliary radio-frequency electromagnetic waves or microwaves. The plasma filament may act like an antenna to absorb the auxiliary radio-frequency (RF) electromagnetic waves or microwaves.

[0024] The pulsed collection of electric charge and energy is repeated with an optimal (possibly very high) repetition rate of the pulsed laser and of the auxiliary RF waves or microwaves. The energy collected per one pulse should exceed the energy spent on creating and enhancing the plasma filaments, so that the net energy collected is positive. The optimal pulse repetition rate (PRR) depends upon atmospheric conditions which can vary widely over even short time durations. As a general rule, one tries to tune the repetition rate to the plasma relaxation time (the ion-electron recombination time) so that the conductivity of the filament remain as high as is optimal given the atmospheric conditions at the time. A feedback control system that varies the PRR as well as the rest of the pulse-shaping parameters may be employed to optimize the net power production of the system. For more information regarding pulse shaping and tailoring, which is the central theme of the field called "quantum control", please see A. Assion, T. Baumert, M. Bergt, T. Brixner, B. Kiefer, V. Seyfried, M. Strehle, G. Gerber (30 October). "Control of Chemical Reactions by Feedback-Optimized

Phase-Shaped Femtosecond Laser Pulses". Science 282 (5390): 919-922. doi:10.1126/science.282.5390.919. PMID 9794756.

[0025] FIG. **2** illustrates a second example of a system **200** for harvesting atmospheric electricity. A plasma filament **120** and a lightning **210** together form a conducting path between ground and a cloud **140**. A condensing mirror **220** directs the beam of laser **110** towards the cloud **140**. A collector **130** is configured to collect electricity flowing along the plasma filament **120**.

[0026] In the case of electricity collection from the cloud **140**, the initial and RF-enhanced plasma filaments can be relatively short, not reaching all the way from ground to the cloud **140**. The relatively short plasma filament can be optimized so that it would trigger propagation of the so-called leader (as in natural lightning), which would effectively use the energy and potential of the cloud to eventually create an electrically conducting path between ground and the cloud **140**. The final conducting path would comprise of a plasma filament **120** and a lightning **210**.

[0027] It is not shown in the figures, but alternatively both the laser **110** and the collector **130** may be placed on a mobile device, which moves together with the cloud. One such mobile device may be a vehicle moving on the ground, but another such mobile device may be an airplane flying in the air. In the case of an airplane flying in the air, plasma filaments may be formed by the laser to reach both ground and the cloud at the same time. In the event the plasma filament directed at ground is not long enough to reach ground, a lightning may also be formed to complete the conducting path to ground.

[0028] FIG. 3A illustrates a third example of a system **300** for harvesting atmospheric electricity. A laser system situated on an airplane **310** forms two plasma filaments (**321**, **322**) that together complete a conducting path between a first region **331** of a cloud **330** and a second region **332** of the cloud **330**. The first region **331** of the cloud **330** and the second region **332** of the cloud **330** carry opposite electric charges. In the example shown here, the first region **331** of the cloud **330** carries positive electric charges, while the second region **332** of the cloud **330** carries negative electric charges. A collector configured to collect electricity flowing along the two plasma filaments is also situated on the same airplane **310** as the laser. Positive and negative charges are naturally separated within the cloud, so this example makes use of the electrostatic potential gradient between different regions of the same cloud.

[0029] FIG. 3B illustrates a fourth example of the system **300** for harvesting atmospheric electricity. It is similar to FIG. 3A, except two lightnings (**341**, **342**) complete the conducting path between a first region **331** of a cloud **330** and a second region **332** of the cloud **330** (i.e., a first lightning **341** electrically connects the first region **331** of the cloud **330** to one of the plasma filaments **321** and a second lightning **342** electrically connects the second region **332** of the cloud **330** to the other plasma filaments **322**). The lightning formation is again because the initial and RF-enhanced plasma filaments can be relatively short, not reaching all the way from the collector situated on airplane **310** to the cloud **330**. These relatively short plasma filaments can be optimized so that they would trigger propagation of the so-called leader (as in natural lightning), which would eventually create an electrically conducting path between the two regions of cloud **330** via a connection through airplane **310**. The final conducting path would comprise of two plasma filaments (**321**, **322**) and two lightnings (**341**, **342**). Of course, one of the lightnings may not be formed, if one of the plasma filament can reach all the way from airplane **310** to cloud **330**.

[0030] FIG. **4** illustrates a fifth example of the system **300** for harvesting atmospheric electricity. A laser system situated on an airplane **310** forms two plasma filaments (**421**, **431**) that together complete a conducting path between a first cloud **420** and a second cloud **430**. The region **422** of the first cloud **420** that is connected to the first plasma filament **421** and the region **432** of the second cloud **430** that is connected to the second plasma filament **431** carry opposite electric charges. In the example shown here, region **422** of cloud **420** carries positive electric charges, while region **432** of cloud **430** carry negative electric charges. Positive and negative charges are naturally separated within the cloud, so this example makes use of the electrostatic potential gradient between different regions of different clouds. Once again, in case one of the plasma filament cannot reach all the way from airplane **310** to any one of the clouds, then lightning may be formed to complete an electrically conducting path from airplane **310** to any one of the clouds. But these embodiments are not shown in the figures.

[0031] FIG. **5** illustrates a sixth example of the system **300** for harvesting atmospheric electricity. A laser system situated on an airplane **310** forms a plasma filament **501** that completes a conducting path between a first region **510** of atmosphere and a second region **520** of atmosphere. A collector situated on the same airplane **310** as the laser system can collect electricity flowing along the plasma filament **501**, because a vertical voltage gradient exists between the first region **510** of atmosphere and the second region **520** of atmosphere. Because no clouds are needed for this embodiment, atmospheric electricity collection is possible under fair weather conditions (relying on the naturally existing vertical voltage gradient in the atmosphere) as well as under foul weather conditions.

[0032] FIG. **6** illustrates an example of a collector **130** configured to collect electricity flowing along a plasma filament **620**. The collector may be a charge collection circuit characterized by capacitance **631**, inductance **632**, and resistance **633**. The electricity may be collected by charging a capacitor associated with capacitance **631**. Battery **610** represents these naturally existing voltage sources: (1) between ground and a cloud, (2) between differently charged regions of the same cloud, (3) between differently charged regions of different clouds, and (4) between different regions of atmosphere, where there is a vertical voltage gradient. The plasma filament **620** can, for the purpose of analysis and optimization of the charge collection process, be viewed as electromagnetic transmission lines characterized by their effective capacitance, inductance, and resistance per unit length. Accordingly, the parameters of the charge collection circuit and of the effective transmission line representing the plasma filaments are optimized, so as to maximize the net charge and energy collected (minus the energy spent on generating and enhancing the plasma filaments).

[0033] The foregoing description is provided to enable a person skilled in the art to practice the various configurations described herein. While the subject technology has been particularly described with reference to the various figures and configurations, it should be understood that these are for illustration purposes only and should not be taken as limiting the scope of the subject technology.

[0034] There may be many other ways to implement the subject technology. Various functions and elements described herein may be partitioned differently from those shown without departing from the scope of the subject technology. Various modifications to these configurations will be readily apparent to those skilled in the art, and generic principles defined herein may be applied to other configurations. Thus, many changes and modifications may be made to the subject technology, by one having ordinary skill in the art, without departing from the scope of the subject technology.

[0035] It is understood that the specific order or hierarchy of steps in the processes disclosed is an illustration of exemplary approaches. Based upon design preferences, it is understood that the specific order or hierarchy of steps in the processes may be rearranged. Some of the steps may be performed simultaneously. The accompanying method claims present elements of the various steps in a sample order, and are not meant to be limited to the specific order or hierarchy presented.

[0036] A phrase such as "an aspect" does not imply that such aspect is essential to the subject technology or that such aspect applies to all configurations of the subject technology. A disclosure relating to an aspect may apply to all configurations, or one or more configurations. An aspect may provide one or more examples of the disclosure. A phrase such as an "aspect" may refer to one or more aspects and vice versa. A phrase such as an "embodiment" does not imply that such embodiment is essential to the subject technology or that such embodiment applies to all configurations of the subject technology. A disclosure relating to an embodiment may apply to all embodiments, or one or more embodiments. An embodiment may provide one or more examples of the disclosure. A phrase such an "embodiment" may refer to one or more embodiments and vice versa. A phrase such as a "configuration" does not imply that such configuration is essential to the subject technology or that such configuration applies to all configurations of the subject technology. A disclosure relating to a configuration may apply to all configurations, or one or more configurations. A configuration may provide one or more examples of the disclosure. A phrase such as a "configuration" may refer to one or more configurations and vice versa.

[0037] Furthermore, to the extent that the term "include," "have," or the like is used in the description or the claims, such term is intended to be inclusive in a manner similar to the term "comprise" as "comprise" is interpreted when employed as a transitional word in a claim.

[0038] The word "exemplary" is used herein to mean "serving as an example, instance, or illustration." Any embodiment described herein as "exemplary" is not necessarily to be construed as preferred or advantageous over other embodiments.

[0039] A reference to an element in the singular is not intended to mean "one and only one" unless specifically stated, but rather "one or more." The term "some" refers to one or more. Underlined and/or italicized headings and subheadings are used for convenience only, do not limit the subject technology, and are not referred to in connection with the interpretation of the description of the subject technology. All structural and functional equivalents to the elements of the various configurations described throughout this disclosure that are known or later come to be known to those of ordinary skill in the art are expressly incorporated herein by reference and intended to be encompassed by the subject technology. Moreover, nothing disclosed herein is intended to be dedicated to the public regardless of whether such disclosure is explicitly recited in the above description.

What is claimed is the following:

1. A system for harvesting atmospheric electricity, comprising:

a laser configured to form a plasma filament; and

a collector configured to collect electricity flowing along the plasma filament.

2. The system of claim **1**, wherein the plasma filament comprises an electrically conducting plasma filament.

3. The system of claim **1**, wherein the laser is a pulsed laser.

4. The system of claim **3**, wherein the pulsed laser is pulsed at an optimal repetition rate.

5. The system of claim **1**, wherein the plasma filament is enhanced by a plurality of auxiliary electromagnetic waves.

6. The system of claim **5**, wherein the plurality of auxiliary electromagnetic waves comprises at least one of radio-frequency waves or microwaves.

7. The system of claim **1**, wherein the plasma filament forms at least a part of a conducting path between ground and a cloud.

8. The system of claim **7**, wherein at least one lightning forms another part of the conducting path between ground and the cloud.

9. The system of claim **1**, wherein the plasma filament forms at least a part of a conducting path between a first region of a cloud and a second region of the cloud.

10. The system of claim **9**, wherein at least one lightning forms another part of the conducting path between the first region of the cloud and the second region of the cloud.

11. The system of claim **1**, wherein the plasma filament forms at least a part of a conducting path between a first cloud and a second cloud.

12. The system of claim **1**, wherein the plasma filament completes a conducting path between a first region of atmosphere and a second region of atmosphere.

13. The system of claim **12**, wherein a vertical voltage gradient exists between the first region of atmosphere and the second region of atmosphere.

14. The system of claim **1**, wherein the collector comprises a charge collection circuit.

15. The system of claim **1**, wherein the laser is situated on ground.

16. The system of claim **1**, wherein the laser is situated on a mobile device.

17. The system of claim **16**, wherein the mobile device is an airplane.

18. The system of claim **1**, wherein negative chirp is used to move the starting point of formation of the plasma filament to a desired location.

19. A method for harvesting atmospheric electricity, the method comprising:

forming a plasma filament with a laser; and

collecting electricity flowing along the plasma filament.

20. An apparatus for harvesting atmospheric electricity, the apparatus comprising:

means for forming a plasma filament with a laser; and

means for collecting electricity flowing along the plasma filament.

* * * * *

Capacitors we used are electrolytic rated at 400 volts x 47 uF put in series to equal 6,000 volts, the diodes we used were silicon 1000 volt 2 amp placed in series to equal 6,000 volts......... Ground was connected to laboratory wall out let ground.

If you decide to try Tesla's experiment by pumping DC into the ground be careful, I tried this and it does work but is very dangerous to you or your neighbors. If someone is taking a shower or using water they can get killed or shocked. do this experiment far away from humans and animals. you can get far more energy out than you put in. I will not tell you much more because it is such a dangerous experiment.

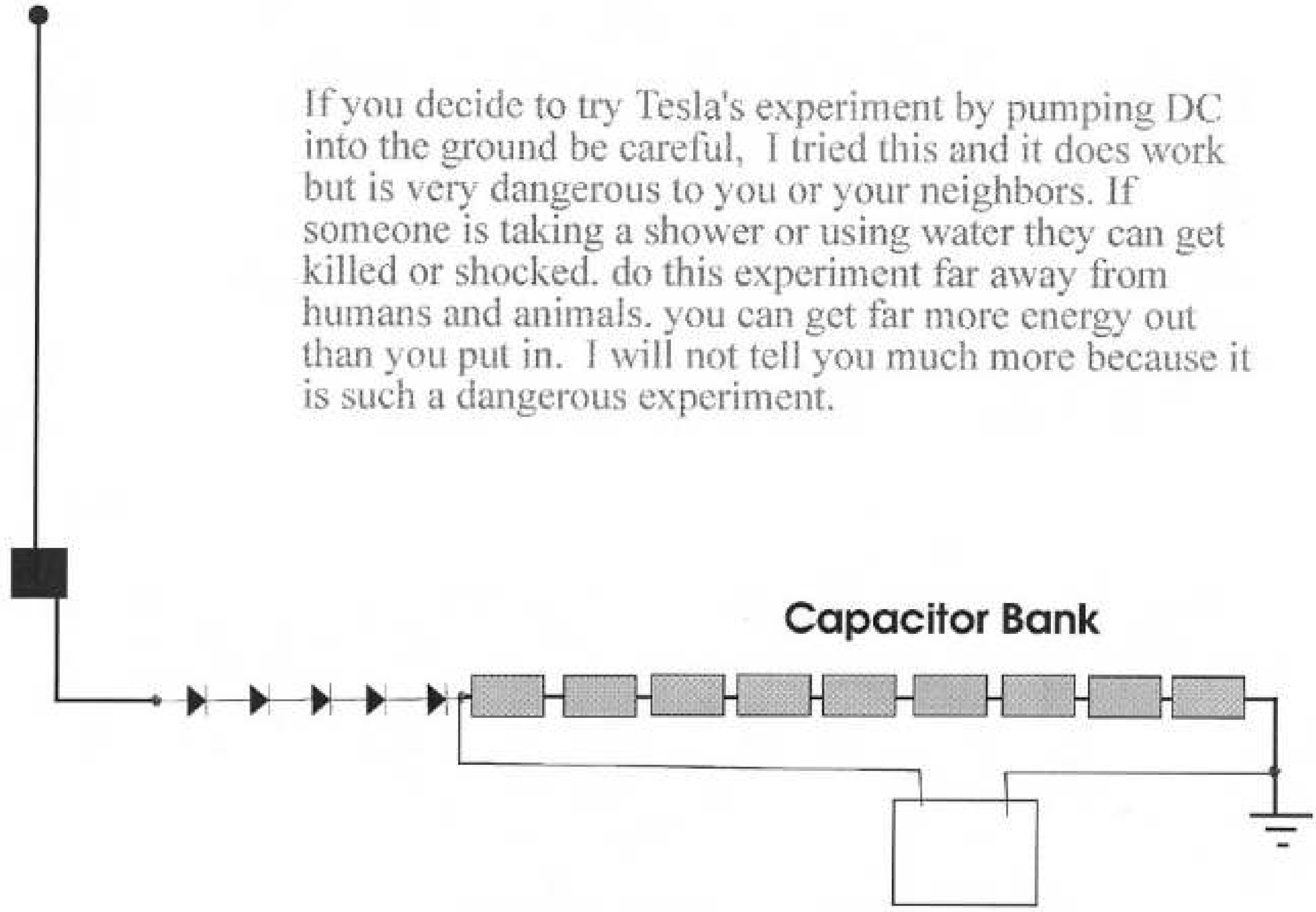

Meter or Load

Nov. 12, 1957

L. R. CRUMP

POWERING ELECTRICAL DEVICES WITH ENERGY
ABSTRACTED FROM THE ATMOSPHERE

Filed March 12, 1954

2,813,242

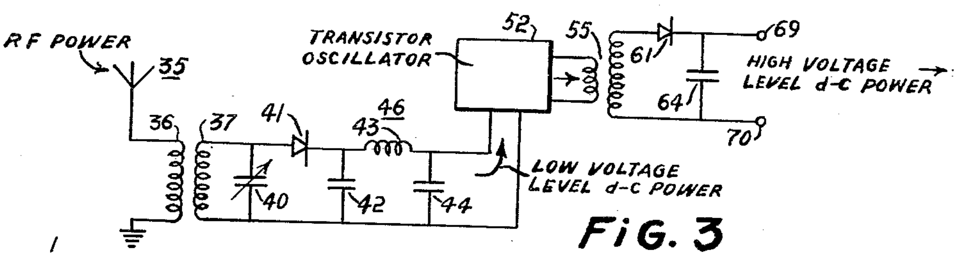
RF POWER
35
36
37
41
46
43
40
42
44
TRANSISTOR OSCILLATOR
52
55
61
64
69
70
HIGH VOLTAGE LEVEL d-c POWER
LOW VOLTAGE LEVEL d-c POWER

FIG. 3

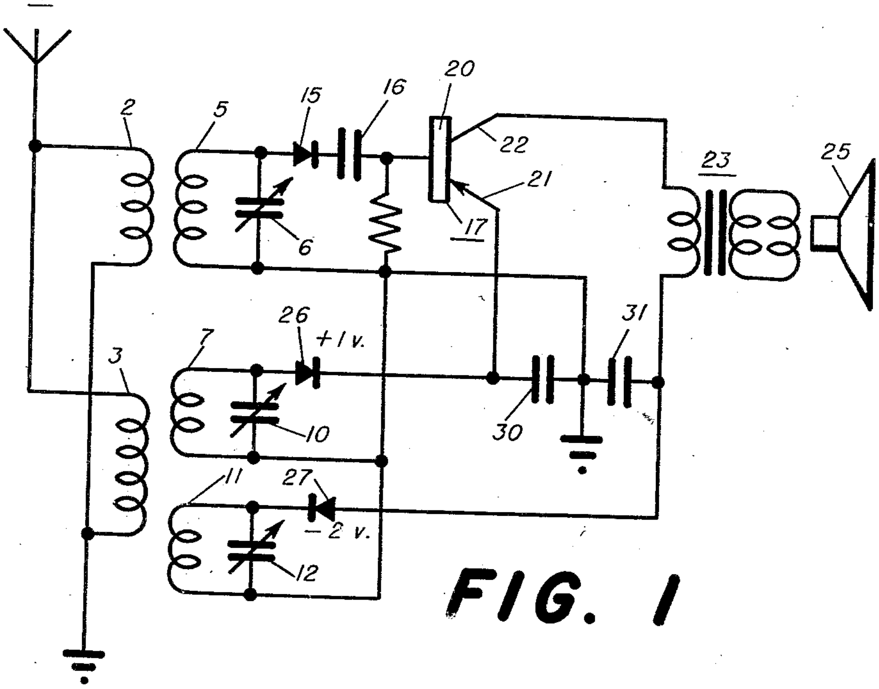
1
2
5
15
16
20
22
23
25
21
17
6
26
+1 v.
31
7
3
10
30
11
27
−2 v.
12

FIG. 1

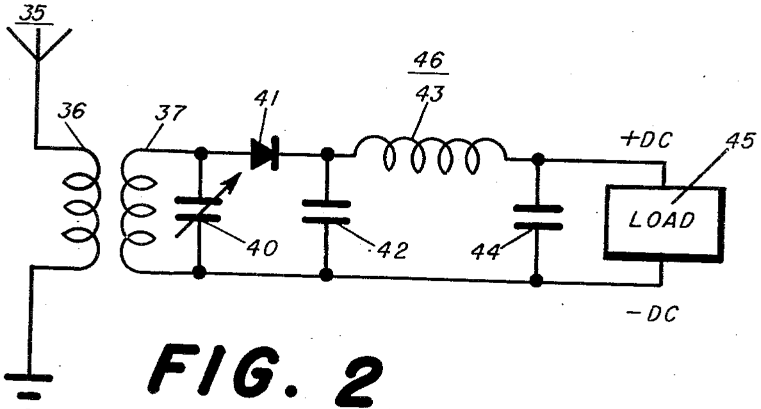
35
36
37
41
46
43
+DC
45
LOAD
40
42
44
−DC

FIG. 2

INVENTOR
Lloyd R. Crump

BY W.E.Thibodeau + A.W. Dew

ATTORNEYS

United States Patent Office

2,813,242

Patented Nov. 12, 1957

2,813,242

POWERING ELECTRICAL DEVICES WITH ENERGY ABSTRACTED FROM THE ATMOSPHERE

Lloyd R. Crump, Silver Spring, Md.

Application March 12, 1954, Serial No. 415,986
1 Claim. (Cl. 321—2)
(Granted under Title 35, U. S. Code (1952), sec. 266)

The invention described herein may be manufactured and used by or for the Government for governmental purposes without payment to me of any royalty thereon.

This invention relates to the convenient and economical provision of power for the operation of electronic circuits and devices using transistors, and of other electrical devices having modest power requirements.

A great advantage of transistors, and a major reason for their enthusiastic reception since their introduction a few years ago, is the fact that they will operate satisfactorily with very low supply voltages and currents. One milliwatt or even less is sufficient to power a transistor in many applications. Various batteries have been developed to provide, in a minimum of space, the relatively minute amounts of power needed by transistors.

My invention provides methods and means that permit transistor circuits, and also other low-powered electrical devices, to be economically and conveniently operated without any batteries whatever, and indeed without any power supply whatever as power supplies are ordinarily conceived.

The invention centers around my discovery that it is practicable to construct operative transistor circuits that are able to abstract from the atmosphere sufficient electromagnetic energy to provide all necessary supply voltages and currents for their own operation. Circuits and devices powered according to my invention will operate indefinitely without any local power source whatever.

I have successfully constructed and demonstrated such circuits. For example, I have constructed a batteryless transistor radio receiver on which I have listened to either nearby or distant broadcast stations as desired, using either headphones or a loudspeaker; this receiver has been powered entirely by electromagnetic energy abstracted from the atmosphere.

From the successful operation of this receiver, and from other experimental work, it becomes clear that, by the methods and means of the invention, a great variety of practical and useful transistor circuits can be powered entirely by energy abstracted from the atmosphere.

Furthermore, as will become apparent below, my invention is applicable to the powering of other electrical devices requiring relatively small amounts of power.

An object of the present invention is to provide methods and means for powering transistor circuits entirely from radiofrequency energy abstracted from the atmosphere.

Another object is to provide methods and means for powering remote radio receivers, low-powered radio transmitters, and other low-powered electrical devices, with energy received by radio from a master station, so that no local power supplies are needed by the devices and so that the powering or non-powering of the remote device is under the control of the master station.

A further object is to provide methods and means for powering transistor circuits and other low-powered electrical devices with radiofrequency energy received from one or more remote radio transmitters.

Other objects, aspects, uses, and advantages of the invention will become apparent from the following description and from the drawing.

Figure 1 is a schematic diagram of a transistor radio receiver in which all necessary power is supplied by energy abstracted from the atmosphere in accordance with the invention.

Figure 2 is a schematic diagram showing a general application of the invention to provide direct-current power to a load.

Figure 3 is a schematic diagram of a system for obtaining a high energy D.-C. source at a high voltage level using energy abstracted from the atmosphere.

Referring to Figure 1, a receiving antenna **1** is connected to antenna coupling coils **2** and **3**, the other ends of which are connected to ground. A parallel resonant circuit consisting of coil **5** and capacitor **6** is coupled to coil **2**. A second parallel resonant circuit consisting of coil **7** and capacitor **10** is coupled to coil **3**. A third parallel resonant circuit consisting of coil **11** and capacitor **12** is also coupled to coil **3**.

Coil **5** and capacitor **6** are tuned to the frequency of a radio transmitter from which it is desired to receive information—for instance, an amplitude-modulated standard broadcast station. The signal received from this transmitter need not be strong. The signal is detected by diode **15** to obtain an audio-frequency information signal. This audio signal is coupled through a capacitor **16** and is amplified by a circuit that includes a transistor **17** having a base **20**, an emitter **21**, and a collector **22**. The amplified audio output of the transistor is coupled through an audio transformer **23** to an electroacoustical transducer, preferably a permanent-magnet dynamic loudspeaker **25** as shown.

The novelty of the invention lies largely in the method and means by which the necessary direct-current power is supplied to the emitter and collector circuits of transistor **17**. This method and means will now be described.

Coil **7** and capacitor **10**, and also coil **11** and capacitor **12**, are tuned to receive radio signals of relatively high strength. It does not matter whether these signals contain information. These power signals are rectified by diodes **26** and **27** to provide direct-current power that is filtered by capacitors **30** and **31**. The D.-C. power thus obtained is utilized to power the transistor **17**.

In the circuit shown, two tuned circuits (coil **7** and capacitor **10**, and coil **11** and capacitor **12**) are tuned to power signals and the D.-C. voltages obtained from each are connected in series. The tuned power circuits may be tuned to the same or different power signals. Under certain circumstances it may be desirable to use more than two tuned power circuits and to tune them to more than two power signals; in this way power can be obtained from several signals and combined. On the other hand, if a strong power signal is available, a single tuned power circuit may suffice to give the needed D.-C. power.

Even weak information signals can be received successfully. A plurality of transistor amplifier stages can be used if desired, or other circuits such as superheterodyne circuits can be used. It is merely necessary that a sufficiently strong power signal or signals be available to provide the small amount of D.-C. needed to power the transistors.

If the information signal happens to be strong, it can be used as the power signal; all of the tuned circuits (coil **5** and capacitor **6**, coil **7** and capacitor **10**, coil **11** and capacitor **12**) are tuned to the information signal.

Engineers who have observed my invention in operation have been surprised at the unexpectedly good results obtained, even with readily available power signals of quite moderate strength. For instance, sufficient power for satisfactory operation of a loudspeaker at low volumes is

readily obtained from a 5-kilowatt standard broadcast station 5 miles away, using only an indoor antenna to pick up the power signal as well as information signals. In typical operation under these conditions a D.-C. voltage of about 2.5 to 3 volts is obtained between the emitter and the collector, at a current of about 250 microamperes; D.-C. power input to the transistor is thus of the order of 0.5 to 1 milliwatt. So far as I am aware, no one has ever before discovered and demonstrated the practicability of this method of powering a radio receiver.

Because existing broadcast stations within a radius of a number of miles provide adequate power signals, the invention is readily practicable with existing power signals in almost any location in or near any city in the United States.

Although I have described a transistor radio receiver powered by my invention, it will be readily apparent that the invention is applicable to the powering of any transistor circuit using one or a number of transistors, and to the powering of other devices requiring relatively small amounts of power. For instance, sensitive electromechanical, electrochemical, or electrothermal devices can be operated by the method of the invention.

Referring to Figure 2, which shows a more general embodiment of my invention, an antenna **35** picks up radiofrequency energy from the atmosphere. This energy flows through coil **36**, which is coupled to a tuned circuit consisting of coil **37** and capacitor **40**. The radiofrequency voltage across capacitor **40** is rectified by diode **41** and filtered by a low-pass filter **46** consisting of capacitors **42** and **44** and choke coil **43**. The resulting D.-C. voltage is applied to a load **45**.

In the practice of my invention, larger amounts of power can be obtained for short periods of time by storing received energy in a suitable energy storage device. Stored energy may then be withdrawn at intervals at a more rapid rate than that at which it was received and put into the storage device. In this way the invention can be used to provide short pulses of relatively very high electrical energy. This result can be readily obtained by charging a relatively large capacitor with direct current and then discharging the capacitor rapidly into a load when desired. This rapid discharge can be initiated automatically when the voltage across the capacitor reaches a certain level, or it can be initiated when a transistor radio receiver receives a certain information signal.

Higher voltages can be obtained with the invention by means of well known devices for raising D.-C. voltages as shown in Figure 3. The D.-C. voltage output from the capacitor **44** can be used to power a low frequency transistor oscillator **52** whose A.-C. output is raised to a higher voltage level by the transformer **55**. This relatively high A.-C. voltage can then be rectified by a diode **61** and fed to a capacitor **64** to provide a high energy D.-C. source at a relatively high voltage level at the terminals **69** and **70**. If desired, energy can now be withdrawn from the capacitor **64** at intervals in short pulses of high energy at a high voltage level. Pulsed radio transmission is one of the possible uses for this form of the invention. Other uses would be to provide a single relatively powerful pulse needed to actuate an electrothermal or electromechanical device.

As has been indicated above, in many locations and particularly anywhere in or near most American cities, power signals normally present in the atmosphere are readily available for the easy and convenient practice of the invention. However, the invention also has important applications in systems in which the necessary power signal is generated and transmitted specifically for the operation of the particular system. Such systems can, for example, comprise a master station transmitting all the power that is needed for hundreds or thousands of fixed or mobile transistor receivers or other remote devices over a range of many miles. This eliminates the need for hundreds or thousands, as the case may be, of local power supplies. At the same time, such a system has the advantage that all of the remote devices can be simultaneously activated or deactivated at the will of the master station, simply by starting or stopping the transmission of the power signal. In such systems it will often be advantageous to use power signals of frequencies sufficiently high to permit the use of resonant receiving antennas of small physical dimensions for signal pickup at the remote devices. In addition to the power signal, the master station may transmit an information signal on the same or a different carrier.

Certain types of devices powered entirely by received radio waves are of course well known. The well-known "crystal set" of the early days of radio, which used a diode rectifier to demodulate an amplitude-modulated radiofrequency signal, is an outstanding example of such a device. My invention is readily distinguishable from such prior devices, however. In typical prior devices a modulated radiofrequency signal is applied to a diode to obtain unidirectional half-wave pulses whose amplitudes vary with modulation. These pulses are integrated by means of a capacitor to obtain a unidirectional signal the amplitude of which follows the audiofrequency modulation envelope. If the radiofrequency signal is received with sufficient strength the audio signal may have sufficient power to operate headphones or similar utilization device without power amplification; but the signal is utilized for its information content, rather than to supply non-information-containing power.

My invention, on the other hand, entails the utilization of received radiofrequency energy to supply power to at least one pair of circuit points (across capacitor **31** in Fig. 1, for example), such circuit points requiring power solely for its power content and not for any information or modulation it may contain. In other words, my invention entails the utilization of radiofrequency energy to supply power that would otherwise have to be supplied by batteries, generator, or other local power source.

It will be apparent that the embodiments shown are only exemplary and that various modifications can be made in construction and arrangement within the scope of the invention as defined in the appended claim.

I claim:

An electrical device for obtaining a high energy D.-C. source at a high voltage level using energy abstracted from the atmosphere, said device comprising in combination: resonant means for receiving radio waves, first rectifier means for converting said radio waves into first direct current energy, first capacitor means for storing said first direct current energy, an oscillator powered by said direct current energy, said oscillator producing an A.-C. output, transformer means for raising said A.-C. output to an increased voltage level, second rectifier means for converting the A.-C. output of increased voltage level from said transformer into second direct current energy, and second capacitor means for storing said second direct current energy.

References Cited in the file of this patent

UNITED STATES PATENTS

1,540,998	Plauson	June 9, 1925
2,056,824	Cawley	Oct. 6, 1936
2,553,734	Adler	May 22, 1951
2,569,687	Pan	Oct. 2, 1951
2,647,957	Mallinckrodt	Aug. 4, 1953

OTHER REFERENCES

Turner: "A crystal receiver with transistor amplifier," Radio and Television News, January 1950.

Creative SCIENCE & Research

459

Information Big Brother & the Oil Companies do not want you to know!

Copyright 2002 - 2004 Creative Science & Research

PO BOX 557 New Albany, IN. 47151 USA

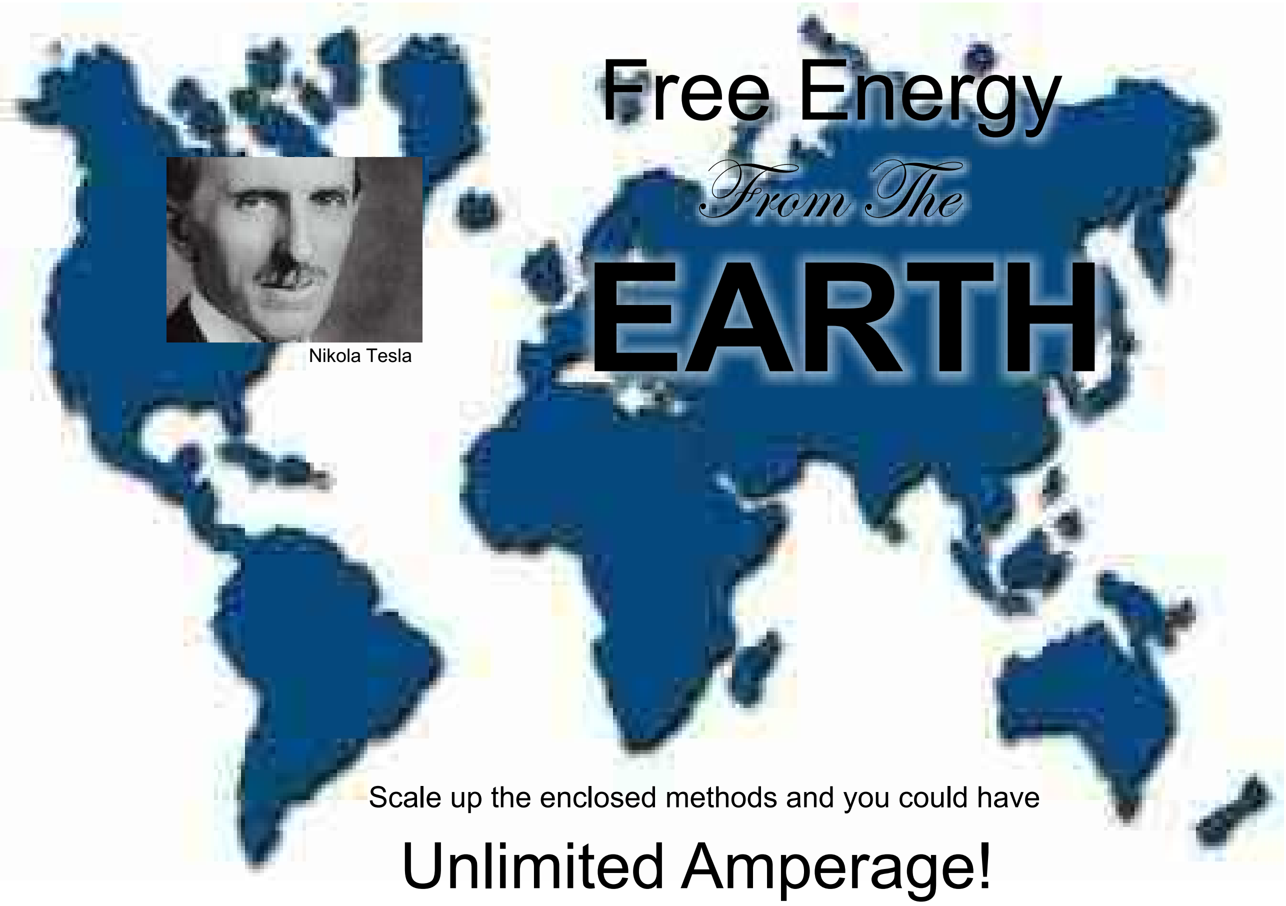

Nikola Tesla

Free Energy *From The* EARTH

Scale up the enclosed methods and you could have

Unlimited Amperage!

By Unlimited we mean you can keep adding more earth cells to your earth power station as you can afford it year by year adding more and more amp power.

Aluminum

Copper

There are many ways to pick up on the free energy in the earth, Just a few are described in these plans. These devices if made right will pick up earth energy that many call ether or static currents as well as acting as high efficient earth battery's. Ether travels through plastic wood etc... Study and research these plans closely there is more to these plans than what you may think. You can learn to get enough power to run your home!

Learning The basic’s of our invention

Experiment #1 How to assemble a 12 vdc system

1. You will have 12- 4" 3/4" copper piping, or If you did not purchase the kit then cut your copper to this size. KIT: take each piece of copper and tape the tops with 1" masking tape.

2. Now using RED LACQUER spray paint (Which you can buy at any hardware store), Paint the outside of the copper all around. (DO NOT PAINT THE INSIDE.) let air dry for 24 hrs. Now remove 1" masking tape.

3. Take the copper pipe outside to the area that you are going to use. use a hammer and hammer each one Into the ground up to the 1" copper exposed top. do not let the exposed copper touch the earth in the final placement. Dirt should be moist. Not real dry.

4. Now remove each pipe and place a thick piece of plastic on bottom of copper pipe to cap it. LET IT BE LOSE, DO NOT TAPE IT. YOU WANT RAIN WATER TO BE ABLE TO ESCAPE FROM EACH PIPE, Do not let exposed outside copper touch the earth.

5. Now hammer each pipe back Into the same holes. place each 2" piece of zinc rod (or a long bolt, most bolts are zinc plated) into the center of each pipe. The zinc Is the NEGATIVE and the copper is the + POSITIVE.

6. Now connect them in series to get 12 volts, use clip wire or solder them. If you are going to leave It outside for a long time then make sure connections are soldered and then painted to protect them from the weather. The rain will replenish your earth batteries. Top get more amaperage with this type simply add more copper and zinc rod cells USE 3' PIECES OF COPPER PIPE INSTEAD OF 4" PIPE. THEN CONNECT EVERY ROW OF 12 PIPES IN PARALLEL.. ADD AS MANY ROWS AS YOU NEED UNTIL YOU GET THE DESIRED AMPERAGE YOU WANT. **(THINK BIG!)** This can be a very powerful free energy system for your home or other. Not only will you get an earth battery effect, but you will also be collecting earth currents from ether /static and under ground radio waves as well.

7. The deeper you go with the copper pipe, the more amperage and voltage you can get.

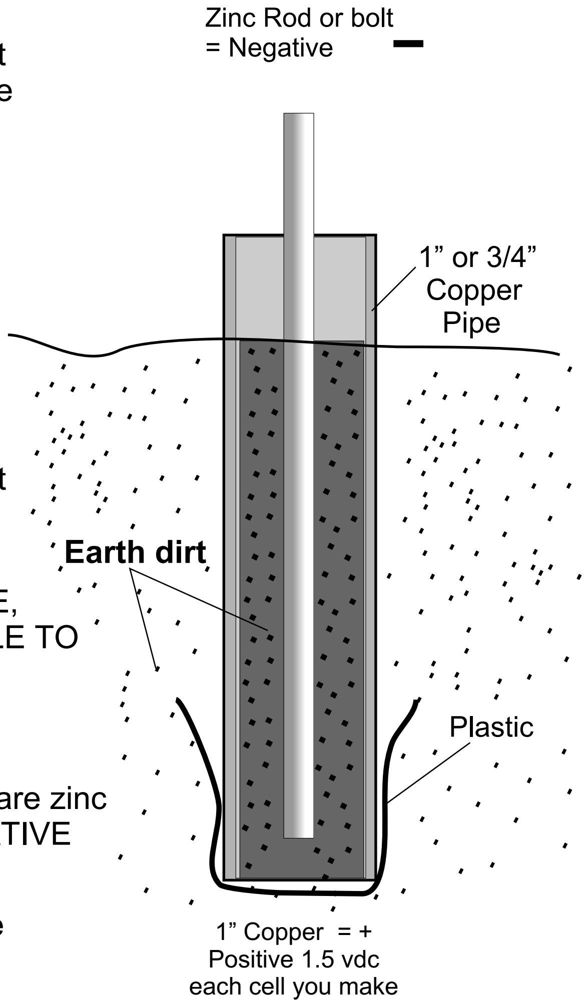

This represents one cell. A qty of 12 to 14 will give you 12 - 14 volts dc when connected in series.

Cut 12 - 4” L x 3/4” copper pipe

Copyright 2002 - 2004 Creative Science & Research

Learning The basic's of our invention

Experiment #2 How to assemble a 12 vdc system

This is just an example to go by and is not the best way but is low in efficiency! See our capacitor earth cells, coil type or flat sheet type. They produce far more amperage and voltages than this type.

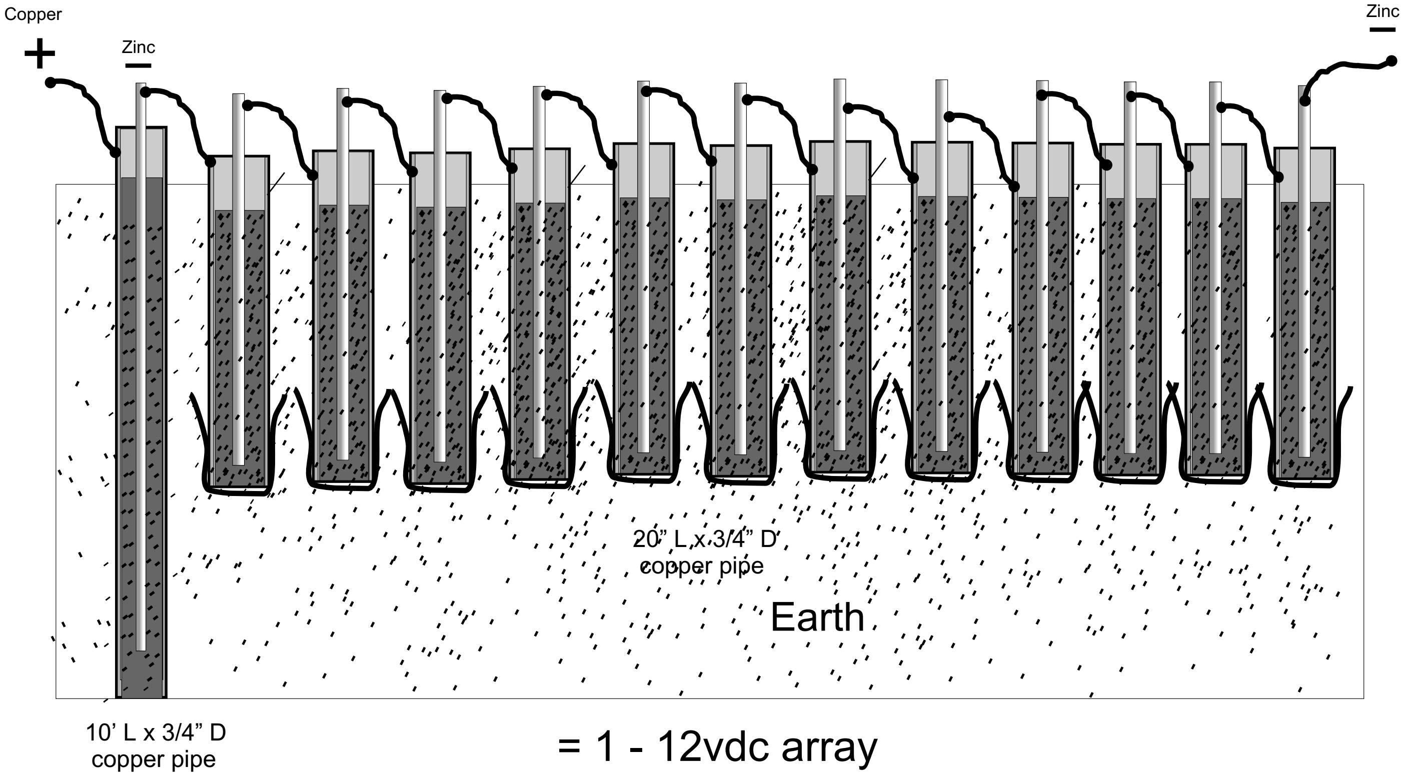

= 1 - 12vdc array

NOTICE: Do not paint the 10 foot copper cell. Keep all copper exposed to the earth. The more exposed copper the better. For higher amperage output, use a zinc or aluminum rod 10/16" Diameter which will leave a space of more than 1/16" to the copper.

Advantages of earth powered batteries

1. Free Energy
2. Long Life
3. The collection of aether energy.
4. Cells are also replenished by outside weather. Rain as well as lightning charges.

Copyright 2002 - 2004 Creative Science & Research

Learning The basic's of our invention

You can purchase Zinc rods or zinc plated bolts at any hardware store, most all nuts and bolts they sell are zinc coated and will work just fine. They also carry aluminum sheeting. Most hardware stores can order you copper sheeting as well.

Experiment #3 How to assemble a 12 vdc system

This is a simple way to produce more amperage but is not practical, we are only showing you this to educate you. The closer the zinc rod is to the copper the more amperage you will get as well as going deeper into the earth. If you do decide to build this basic set up. It is suggested that you solder all leads very well and dig deep enough into the earth to hide the tops of each cell. By covering the tops with dirt the cells can short out so what you want to do is to paint the exposed tops with lacquer or enamel paint then cover with plastic, cover with dirt and allow both the negative and the positive leads to stick out of the ground. But like I said this is not practical, what your goal should be is to make a very strong uf earth capacitor / battery, This will allow you to pick up and collect earth radiant waves of energy as well as lighting energy during storms, the energy you can collect in a strong lighting storm will amaze you! You must use an antenna on the zinc rod or copper rod. **WARNING! Be careful when collecting lighting it can kill you, be sure the positive and negative wires are not coming into your home. During a storm it is advised to have the lead wires going into a large uf cap bank via a one way diode on the positive lead to the positive of the cap or battery. The earth battery's can be dangerous at times and must be completely buried in the earth so a child, adult or animal can not touch or get to. Our earth capacitor batteries can store and hold thousands of volts. So be careful. Use rubber gloves and clothing to protect yourself. We are not responsible for anything in these plans you build at your own risk.**

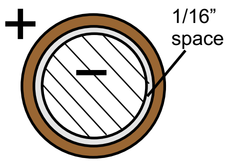

Example of a perfect Zinc rod copper pipe cell

Top view of a 12 volt dc array high amp system

Connect the series arrays in parallel to each other as shown. Or for a high voltage system connect all in series! Pulse the dc into a 10x step up transformer, 15,000 v Neon transf.

Zinc rod can be replaced by using rolled up aluminum sheeting or Zinc sheeting, Use 100% cotton cloth wrapped around the rolled sheeting or rod to give a small space in between the aluminum sheet and copper.

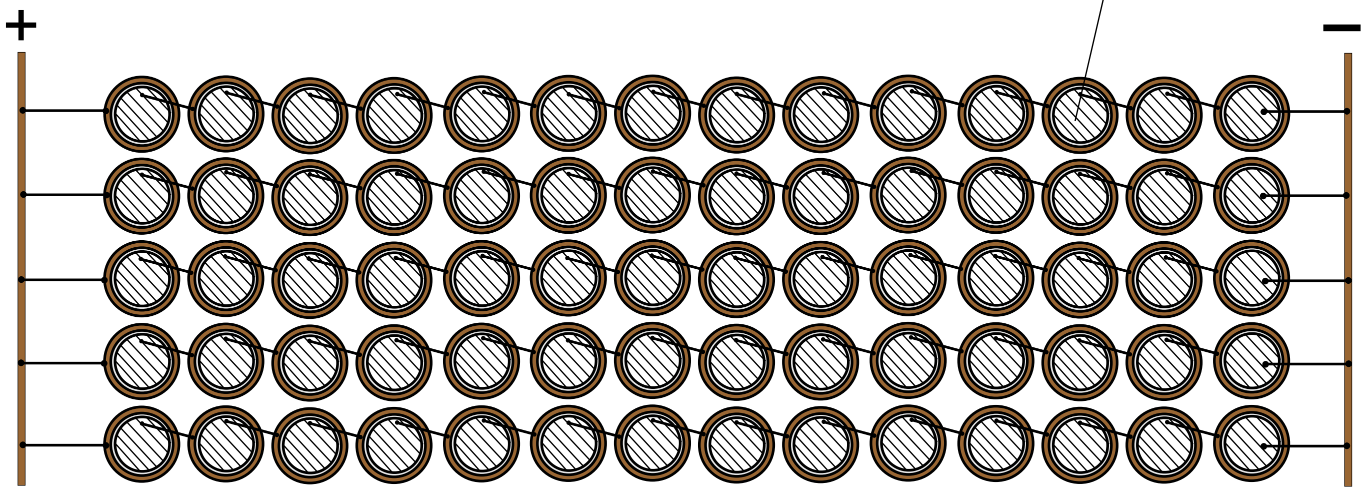

Copyright 2002 - 2004 Creative Science & Research

Learning The basic’s of our invention

Experiment #4 The 6 foot spacing method (old way)

There are many US Patents that were Issued back in the 1800's, one of them was issued to a Mr. Deckman. Deckman found that if you take several small pieces of Zinc and Carbon rod and stick them into the ground next to each other and connect them in series (much like a battery) you get very liittle gain at all.

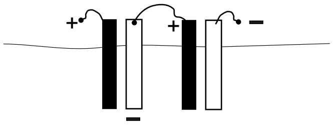

But if you put them 6 feet apart you will get a gain In voltage and they will not cancel one another out. so by doing this you can put them In series to increase your voltage and your volt amps. The theory is that there Is some sort of natural energy VORTEX that takes up about that much space for each unit or cell.

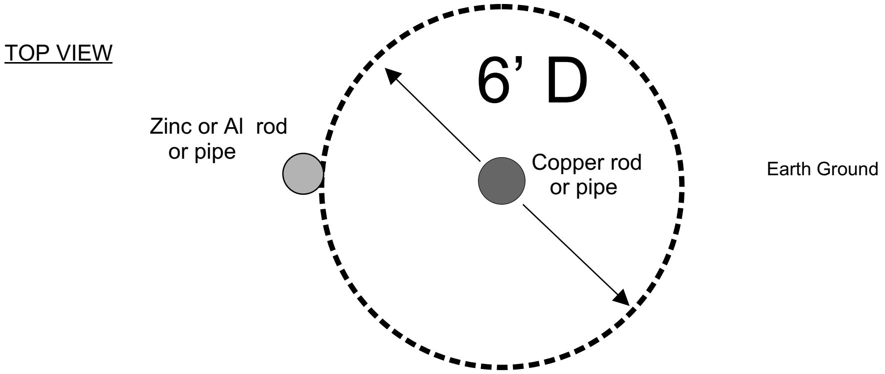

By using this method it will take a lot of ground space, ground space that many people just do not have with the exception of farmers. If you do have the space then think BIG! dig deep and have as many cells as you can get up to 2 to 4 acres. There are much better ways - vs - the old method as you will see. but by using Deckmans old way or our new methods, you can get as much free energy as you want with voltages or amperages as high as you want or desire. The higher the desired amperage the more costly it can get. We are always working to better our inventions to keep the cost down.

Copyright 2002 - 2004 Creative Science & Research

Learning The basic’s of our invention

By using our new system you can get as much voltage or amperage as you want. Just size It to meet your needs or your budget.

How to get AC current for house hold use.

You can convert the incoming dc current to ac by using a home made inverter (which we sell the plans to, **5 kw inverter plans $40 order #579**) or you can purchase a 12 v to 115 vac inverter from a store or solar panel supply house.. If you are planing to make a 115 VAC system this Is the only way you can do it. You can not get AC current from the ground. You must use our system to charge a deep cycle 12 v marine battery, and then connect our 5,000 watt 120 vdc to ac inverter to the battery to run your home. (we do not sell kits or inverters assembled, we only sell the plans at this time.)

More on the basic pipe method

Please note: the dirt level inside each pipe should be up to 1/4" from the top. each pipe should be tightly fliied with dirt. Which will automatically happen when you hammer the pipe into the ground the first time.

Each copper pipe (or cell) should be spaced 1 Inch to 1 1/2 inch apart. Doing this will give you more amperage and voltage.

How to dig a simple deep hole for 8' pipe: There are many ways in which you can do this. you can rent a ground drill bit and rig it up with a electric 2 hp motor or you can use 1" copper pipe.

First: prepare the ground with water to make it moist but not muddy, 2nd: hammer a 1" x 5' copper pipe into the moist ground about 5 inches deep/ then remove it from the ground and take a stick or a water hose to remove the dirt from inside the pipe. Then just repeat the process over and over until you get the desired depth that you want. you can get about 10 feet or more using this method.

WARNINGI BE CAREFUL OF ELECTIC POWER LINES WHEN YOU PLACE OR PULL THE COPPER PIPE FROM THE GROUND, IT WILL KILL YOU IF YOU TOUCH A POWER LINE WITH THE END OF THE COPPER PIPE.

We are dealing with a new type of energy as well as conventional energy < Ether energy can go through rubber/ plastic. Iron, etc....

Copyright 2002 - 2004 Creative Science & Research

What is ether or Radiant energy?

The understanding comes from a very deep study of etheric or static forces, this energy that we are getting out of the ground is what some call LIGHT ETHER, when you separate the light ether from the warm ether you can get these superconductive events. It appears like electricity but it has no heat events related to It.

The point is that the best polarity's of things to get a differential in the light ether is to have a silvered coloured metal and then something that is totally absorbitive/ Like black Carbon.

So now you have something that's the perfect reflector and something that is the perfect absorber, and once you understand what your pulling out of the ground (and that it Is not classic electricity as we get out of a normal type of chemical battery), then your going to be farther down the road to understanding how to get much more power out of this stuff.....

There are researchers now that are getting 1 /2 volt to 2 volts dc on a "Quote" volt meter, but it will shock you through an inch of rubber shoes this is not normal electricity!

Q: Is this energy coming from the sun and being absorbed In the earth or from the center of the earth?

A: Yes, but the primary source of light ether is the sun light, yes the earth does have an economy of light ether which does follow the sun around. BUT IT IS ALSO IN THE GROUND! AND ITS IN THE AIR!

Q: Did NokolaTesla know about this in the 1900's?

A: Yes, if you read into his writings he was always saying things like, there is more energy moving than just the electricity in this stuff.

Example: If you take a large Tesla coil and fire it up, and you drive it with a van de graph generator and you put the output of that into a Tesla coil. you can throw sparks across the room! What these electrostatic machines are gathering is light ether.

Copyright 2002 - 2004 Creative Science & Research

A Very Strange and New Discovery!

While working outside on another experiment I discovered a very strange generator effect, much to my surprise, I found that by using 2 aluminum 8" disks connected to my DC volt meter and by placing both disks in a vertical position while touching the bare earth dirt, and moving one of the disks and allowing the other to be stationary I get a dc voltage positive output! This effect works with any size or shape of aluminum, we used 8" diameter disk x 1/8" thick, we found that size to work best. Also by stacking them like a capacitor you will find that the current output is far greater, For Example: see figure 1, Disk A would have 4 more disks connected in parallel of the same size separated by paper die electrics. The same goes for Disk   B .

First find a spot any where on the ground where there is no grass (best when wet). Please note, it's not just creating an acid battery effect, you are also getting a good contact with the ground as a conductor to pick up ether energy.

Now place the disks side by side, attach your DC voltmeter probes onto the disks, one negative and one positive. DC meter should be on the lowest setting, now place both disks on the ground very softly and move **disk B** and **do not move disk A** (positive.) You will notice you are getting a small amount of voltage and amperage. Now Move **disk A** and do not move **disk B.** THE POLARITY CHANGES.

The rule here is; The disk that moves is always positive. since we can get a polarity change just by moving the disks at opposite times. You can make a very simple mechanical device to get AC current from on top of the earth. Again, this is not just a chemical reaction we have done many test to prove this. You can also get a small voltage reading in dry dirt. The reason for the smaller voltage in using dry dirt is because a low conductance resistance is taking place here, keeping the system from being in contact with the earth and collecting the earth **ether energy.**

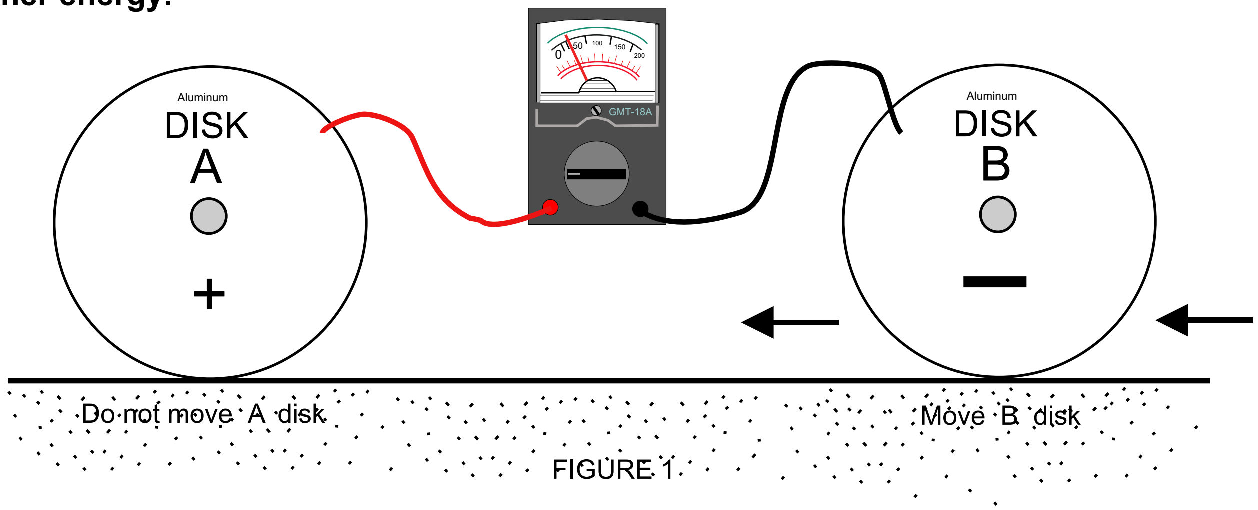

FIGURE 1

Copyright 2002 - 2004 Creative Science & Research

Capacitor Sheet Method

This method is far better than using pipes or rods. By using copper and zinc or aluminum foil sheets you will get much more amperage out of your system!

The amperage you will be collecting will be coming from 3 different sources,

1. **The acid in the ground and water**
2. **Energy that is being transmitted from the earth itself**
3. **Energy that is being transmitted from the sky and space.**

This may all seem unbelievable but it's true and is a fact! The more plates that you add the more power you will get! The copper sheet(s) is the positive, it must be facing down toward the earth. The aluminum or zinc sheet(s) is the negative and must be facing toward the sky! Between each sheet place 1- cotton or course screen printing mesh or other plastic course mesh cloth material. you can use a hole puncher to punch exact holes in each sheet, use 2 holes per sheet copper and aluminum. Use a 8 1/2" x 11" paper puncher one that is adjustable. next use a 3/4 wood base and drill hole for wood dowel rods to fit to the holes you punched in the sheets, you will have a total of 4 wood dowel rods, 2 on the right for the copper sheets and 2 on the left for the aluminum sheets to fit over, cloth die electric can be done the same. You then need to assemble dry and use another 3/4" wood board for top holding plate, which will fit over the wood dowel rods on top, sandwhich and press the assembly together and tape in place with duct tape, drill 2 long bolt holding holes on bottom wood base and top wood base then attach long bolts and nuts and tighten down. Remove tape. and attach wires, water the entire capacitor assembly and bury in the ground.

Copper sheet

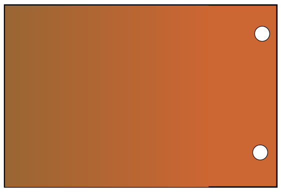

Top View

Aluminum sheet

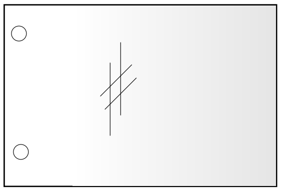

Top View

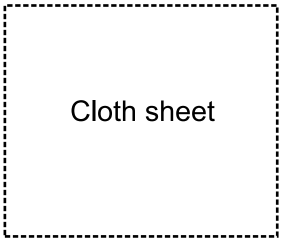

Top View

Place Aluminum sheet on top of copper sheet

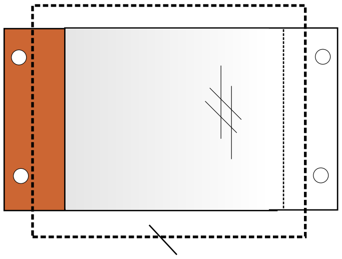

Dotted lines indicate where cloth goes

Top View

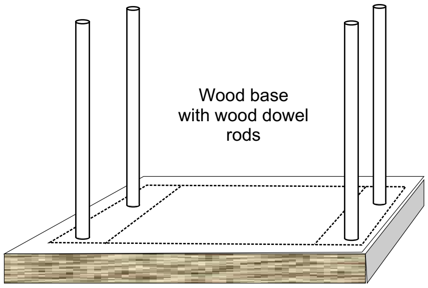

Side View

Copyright 2002 - 2004 Creative Science & Research

Capacitor Sheet Method

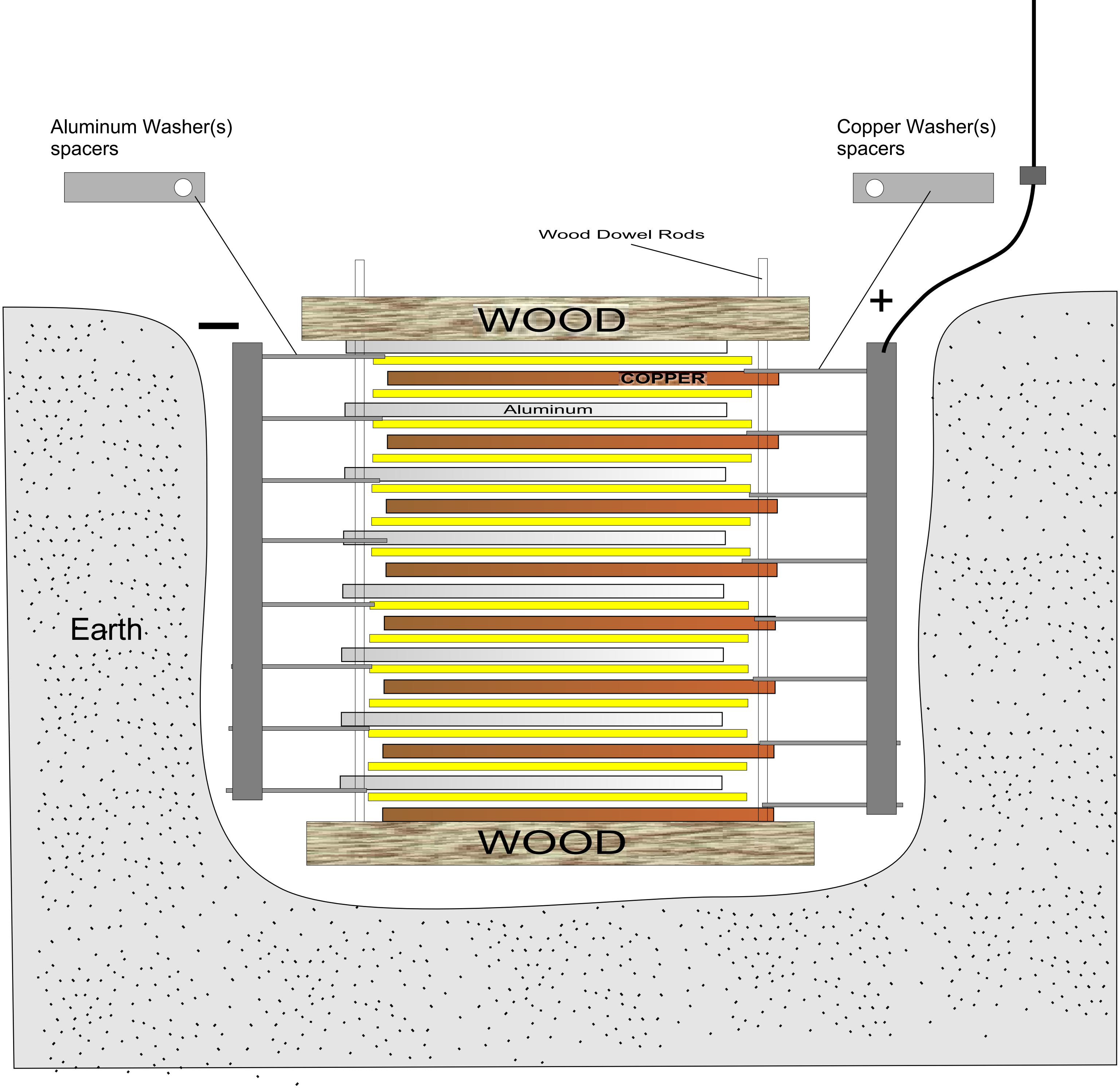
10' Antenna
Aluminum Washer(s) spacers
Copper Washer(s) spacers
Wood Dowel Rods
+
—
WOOD
COPPER
Aluminum
Earth
WOOD

Copyright 2002 - 2004 Creative Science & Research

Capacitor Sheet Method

Again, the more plate sheets of metal you add (or cells) the more power you will get! Remember it's all about capacitance. You are collecting more than just simple battery type power. The copper plate sheet is the Positive + and the Aluminum sheet is the negative. There are many ways in which one can build this type of earth energy collector, below is simply an example of another shape and form. In this type we use water holes. water holes must all be in exact spots.

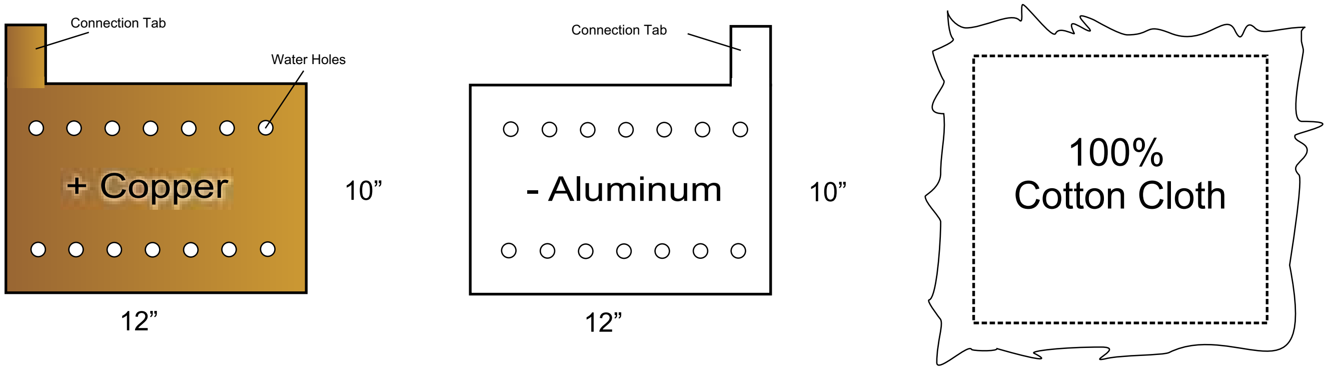
Connection Tab
Water Holes
+ Copper
10"
12"
Connection Tab
- Aluminum
10"
12"
100%
Cotton Cloth

Figure #1 Top View of separate pieces

Figure #1 Side View

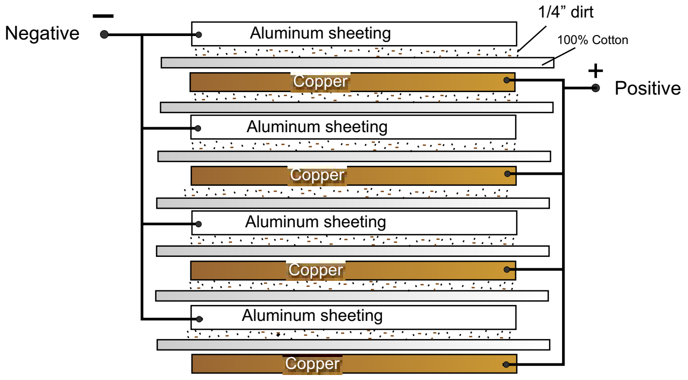
Negative
−
1/4" dirt
100% Cotton
+
Positive
Aluminum sheeting
Copper
Aluminum sheeting
Copper
Aluminum sheeting
Copper
Aluminum sheeting
Copper

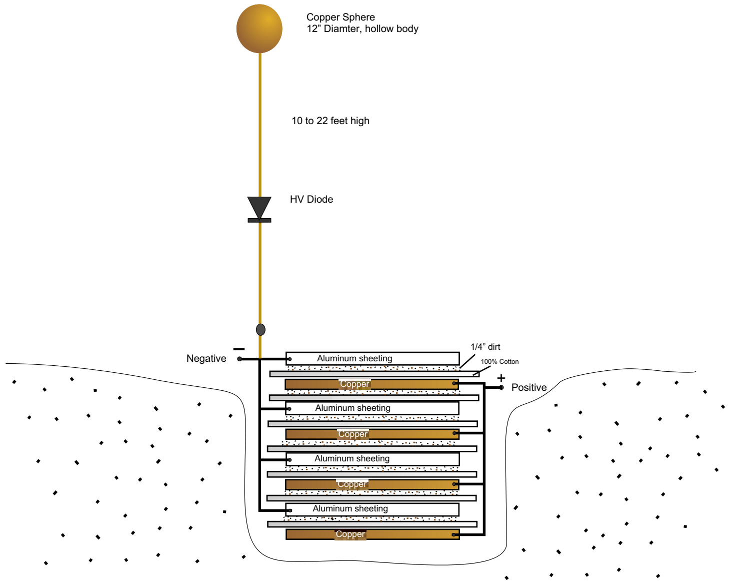
Copper Sphere
12" Diamter, hollow body
10 to 22 feet high
HV Diode
Negative
−
1/4" dirt
100% Cotton
+
Positive
Aluminum sheeting
Copper
Aluminum sheeting
Copper
Aluminum sheeting
Copper
Aluminum sheeting
Copper

Copyright 2002 - 2004 Creative Science & Research

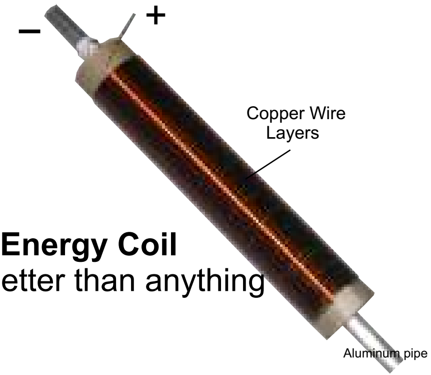

OUR MORE ADVANCED EARTH CAPACITOR!

(EECC) Earth Energy Coil Capacitor!

To collect more ether earth energy atoms, build the following **Earth Energy Coil Capacitor**, **Designed & Created by Creative Science & Research**. This is far better than anything we have found yet! Much more power and is easier to build.

Start with our EECC Model #1, Take a 5/16” zinc rod or aluminum rod, cut to 7.5” long. Now use #27 bare copper wire, secure it in a drill press (which you will then turn by hand). Cover zinc rod with paper one layer, use very tiny bits of tape to secure. Now apply the copper wire over the paper, remember to use copper that is not coated! Tape one end of the copper wire to the end of the zinc or aluminum rod and begin to slowly wind. Each wind must be perfectly fitted side by side, paper should be 5.5” long, total length of winds should be 4” long centered on paper. Once you are done with the first layer, apply another piece of paper to the top layer of the first copper layer, do not use to much tape to secure paper in place! Now begin your 2nd layer, repeat this same process until you have about 10 layers, (the more layers the better!) Copper must not touch the zinc or aluminum. When done hot glue or epoxy a small end piece of the copper wire so it will not come unwound. This is one complete cell, keep in mind this is a small test version, to get more power you have to THINK BIG! Build larger cells, using larger size copper. Then place them in salt water to test, 2 cells can be connected in series, after that you must connect them by use of capacitors, diodes and electronic or hand wound spring switches. If you build 20 of these and try to connect them all in series without a cap/ diode switching method, the cells will cancel each other out. Keep in mind you want water to get to each layer as well as the zinc or aluminum. You can also use Aluminum or zinc pipe.

Now concerning the paper Die Electric, Paper will not last very long, test and find a water absorbing plastic die electric,(mesh type). We suggest that you use Plastic screen mesh that you can buy at any hardware store, if all you are interested in is collecting small voltages then use one layer in between each layer of copper, if you are interested in collecting higher voltages such as lightning charges then you must use more layers of plastic screen mesh to thicken the die electric so the higher voltages will not short out your system. Paper may not last longer than 1 year, Plastic screen mesh or polyester screen printing mesh will last for many years.

EECC Model #2 Same as #1, but different. In this design we use #27 copper COATED wire, instead of bare wire. since the copper wire is coated and not bare, each layer must be sanded, that is, the top portion of the copper coated layer must be sanded off so it will be exposed to the water and acid, a die electric should still be used in this case. This type makes for a very high efficient earth cap!

EECC Model #3 Same as #1 but reversed! You will use aluminum wire and a copper rod or pipe. Use a 3/4” copper pipe and wind the die electric paper or plastic mesh and Aluminum bare wire around the copper pipe as explained in Model #1

Copyright 2002 - 2004 Creative Science & Research

OUR MORE ADVANCED EARTH CAPACITOR!

(EECC) Earth Energy Coil Capacitor!

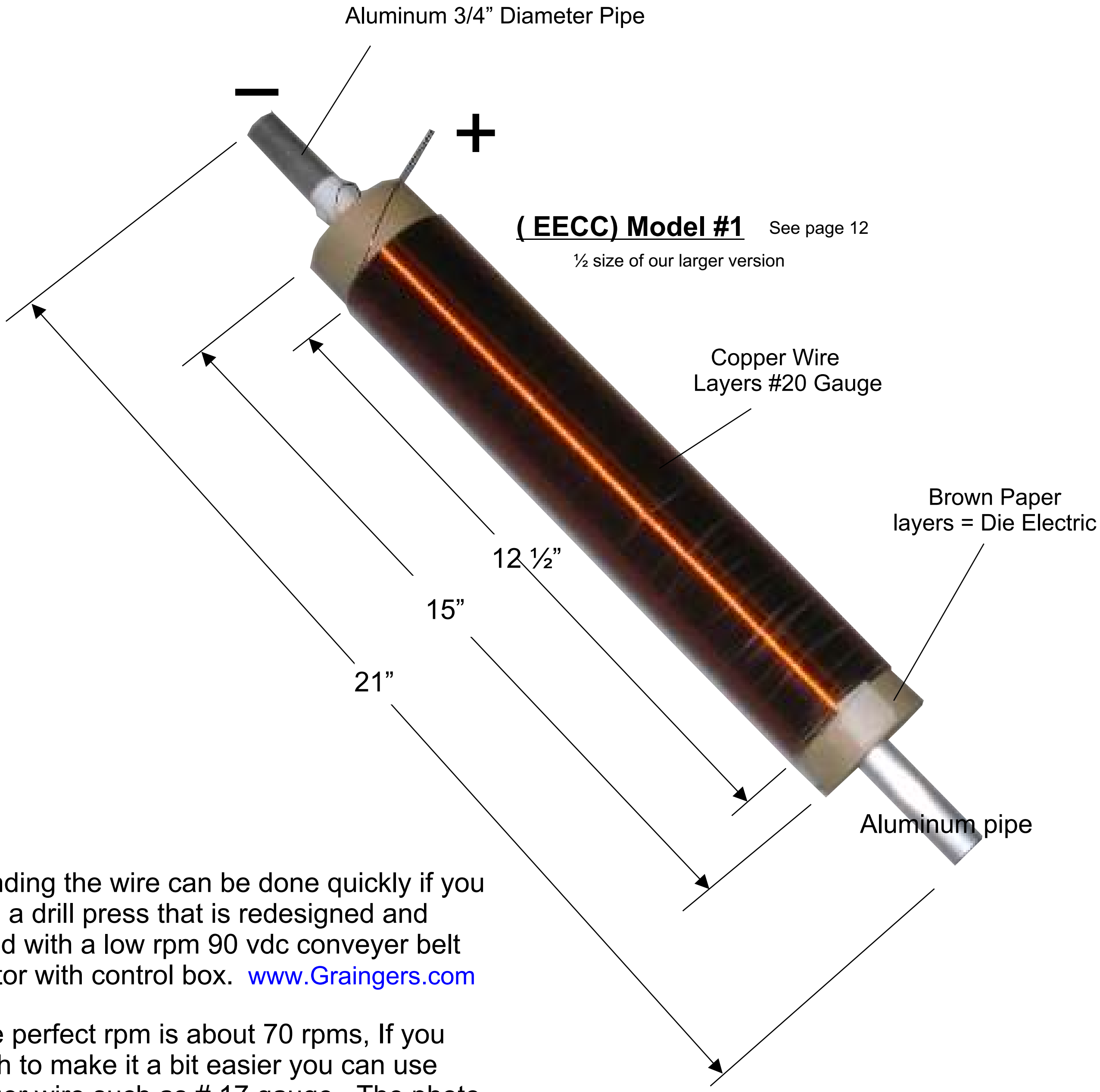
Aluminum 3/4” Diameter Pipe
–
+
Copper Wire
Layers #20 Gauge
Brown Paper
layers = Die Electric
12 ½”
15”
21”
Aluminum pipe

(EECC) Model #1 See page 12

½ size of our larger version

Winding the wire can be done quickly if you use a drill press that is redesigned and fitted with a low rpm 90 vdc conveyer belt motor with control box. www.Graingers.com

The perfect rpm is about 70 rpms, If you wish to make it a bit easier you can use larger wire such as # 17 gauge. The photo is an example of a mid size cell and can be very powerful! This is ½ size of a large cell.

Copyright 2002 - 2004 Creative Science & Research

OUR MORE ADVANCED EARTH CAPACITOR!

(EECC) Earth Energy Coil Capacitor!

Our small test model #1, we used #22 bare copper wire wrapped around a zinc rod, each layer had a paper die electric, tested in tap water. 5 layers. Output = 89 vdc x 3.5 ma, this is low. more layers will = more capacitance which ='s more amperage and voltage.

Small Drill press turned into a Coil Winder

The more layers, the more free earth energy you can collect!

Small Drill press turned into a Coil Winder. We turned this by hand. Larger coils is best to use low rpm motor at about 70 rpms, In this picture you see that we wound each wind side by side, this makes for a more high efficient earth cap. You can wild wind if you like if you use smaller wire like #31 to #34.

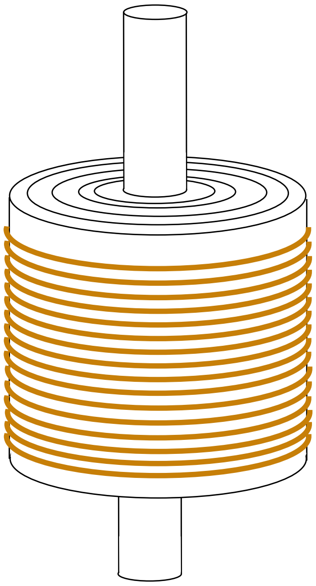

Side View
Ideal size for
experimenting

Copyright 2002 - 2004 Creative Science & Research

(EECC) Earth Energy Coil Capacitor!

Rolled Earth Capacitor

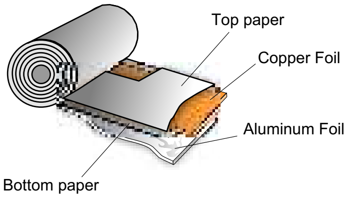
Top paper
Copper Foil
Aluminum Foil
Bottom paper

MODEL #4 Rolled Earth Capacitor

Copper foil is placed between 2 pieces of paper strips and Aluminum is placed on bottom. Use 400 foot strips 2” wide. This method is very hard to do. But is very high eff.

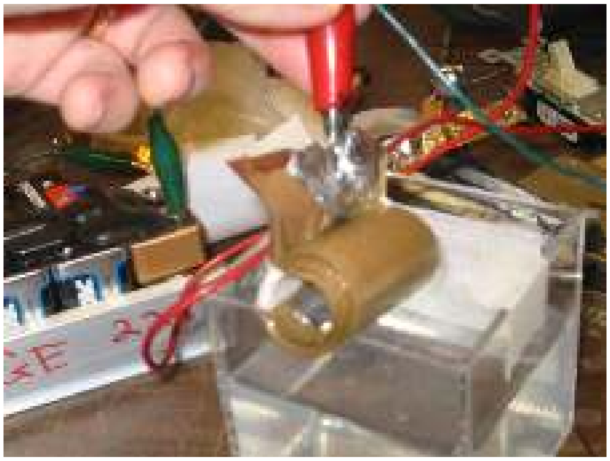

We used salt water to test rolled caps before placing them in the ground.

Copyright 2002 - 2004 Creative Science & Research

(EECC) Earth Energy Coil Capacitor!

MODEL #4 Rolled Earth Capacitor

Rolled Earth Capacitor

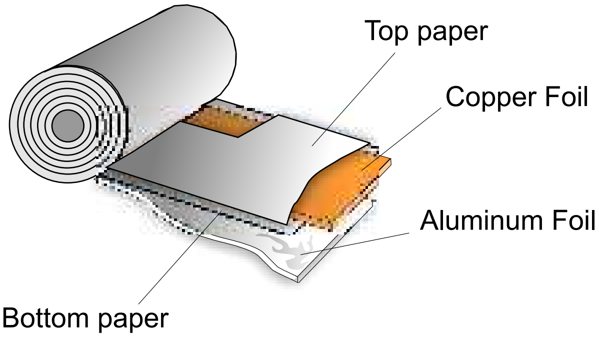

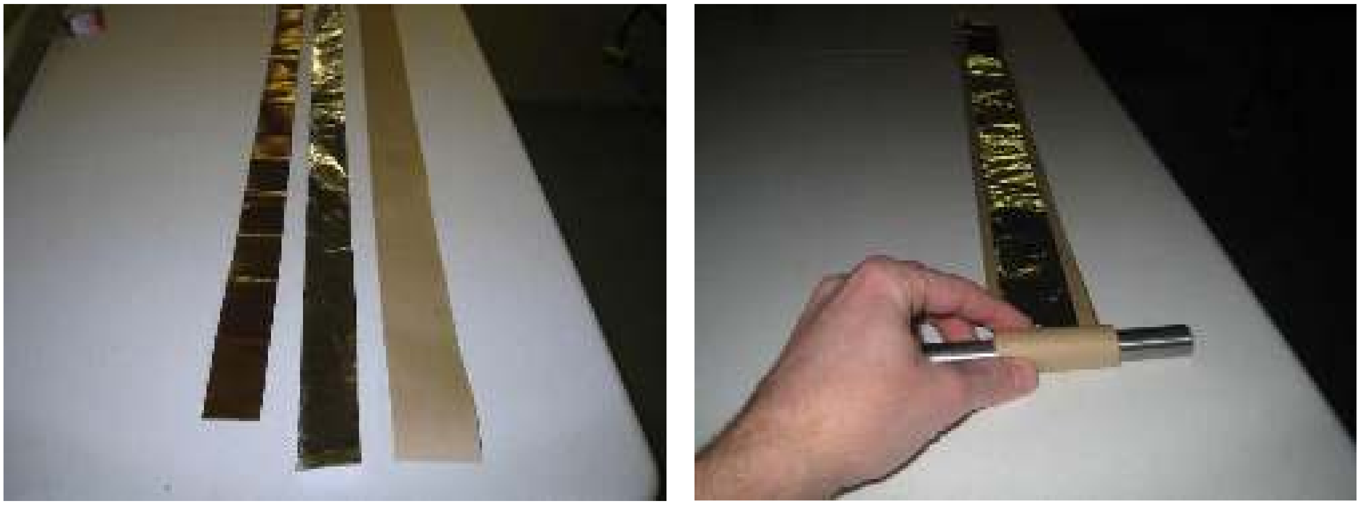

Again rolling this type of earth capacitor is very hard, Everything must be kept straight and inline with one another, Manufacturers of rolled electrolytic capacitors use computers and machinery. I am sure a simple home made roller with pin registration could be designed very easily. This type of earth battery is very high efficient. Paper is not a good die electric for high voltages! If you are wanting to collect earth and storm HV Currents you must use a thicker die electric. Metals must be spaced properly for storing high voltage.

(EECC) Earth Energy Coil Capacitor!

Rolling a capacitor is the best way but not the easiest,
Which makes this pipe coil model much more desirable to build

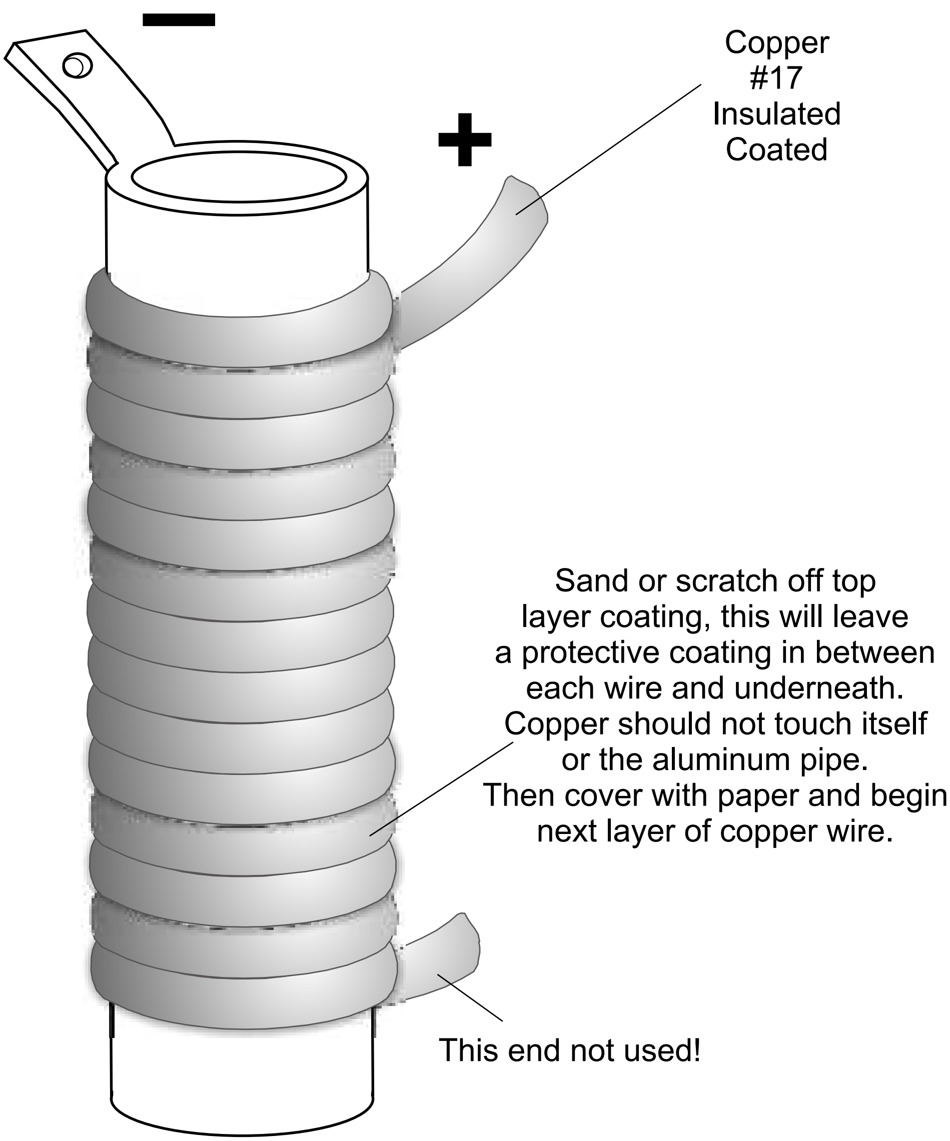

Sand or scratch off top layer coating, this will leave a protective coating in between each wire and underneath. Copper should not touch itself or the aluminum pipe. Then cover with paper and begin next layer of copper wire.

This end not used!

Symbols

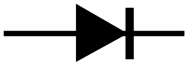

= One way DIODE

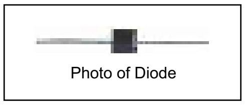

Photo of Diode

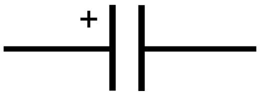

Capacitor (Electrolytic)

Photo of one type of Capacitor

Copyright 2002 - 2004 Creative Science & Research

CONNECTING THE CELLS IN SERIES

How to connect the cells in series to increase the output voltage without allowing the cells to cancel each other out. Figure **A** The Copper pipe and zinc rod method without the use of plastic on bottom of pipe. **Figure B** is the coil earth cap method.

Figure A

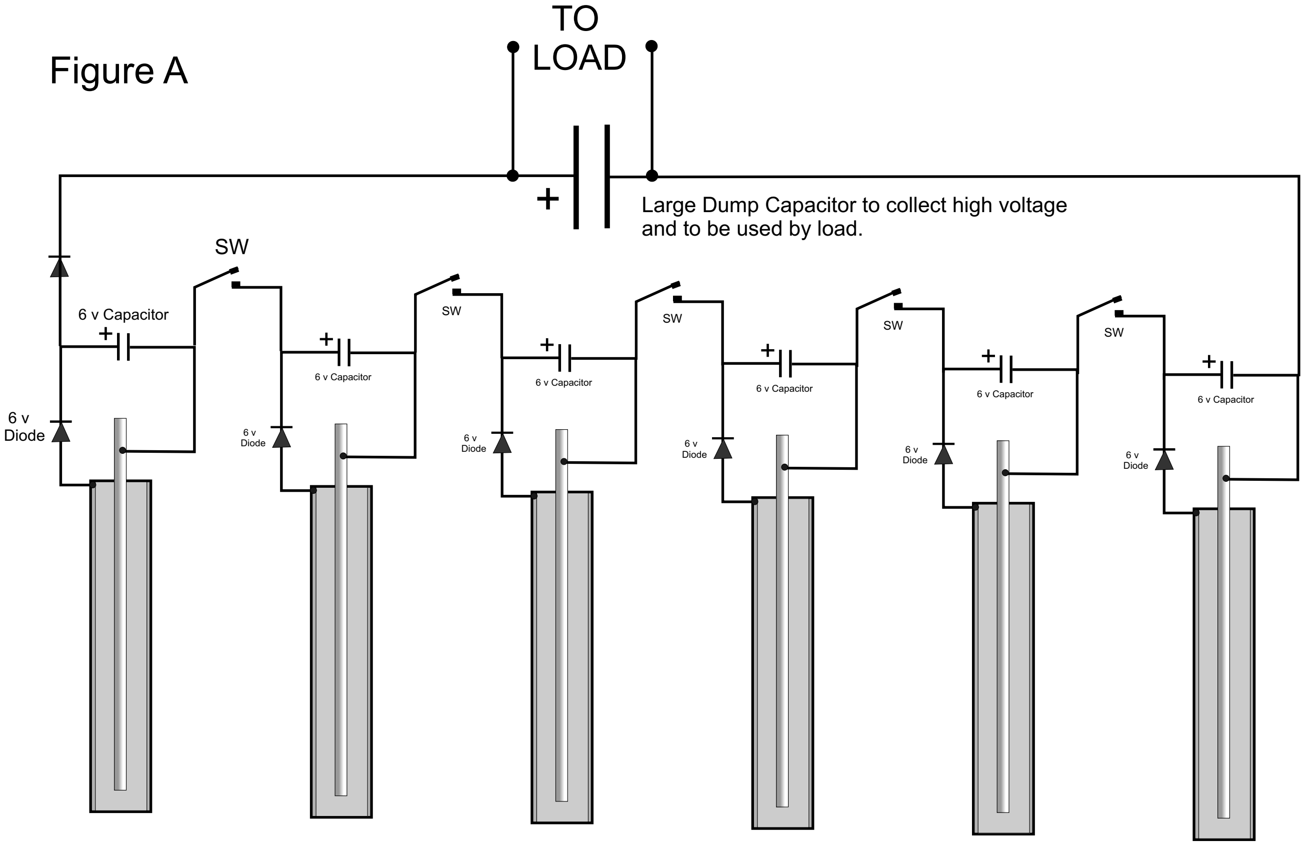
TO LOAD
+
Large Dump Capacitor to collect high voltage and to be used by load.
SW
6 v Capacitor
+
6 v Diode
SW
SW
SW
SW
SW
6 v Capacitor
6 v Capacitor
6 v Capacitor
6 v Capacitor
6 v Capacitor
6 v Diode
6 v Diode
6 v Diode
6 v Diode
6 v Diode

Use electrolytic capacitors to store the charge coming from each cell, then connect them in series using Commutator Sw's connected to a small rotor shaft, connected to a milliamp solar cell hobby motor or build a small fuelless engine electric motor to run on and off switch's. You can also build a very low milliamp electronic switching circuit and use the earth energy to power it. Or another more simple way would be to use a hand spring type device (timer type) glue small magnets on a long rotating rotor arm connected to the inner spring. Then place small reed relay magnetic switches, on each timing magnet ON position. should have multi reed sw's, as the unit turns if gives time for the caps to charge and then comes around and connects caps all in series which in turn discharges into main cap dump.

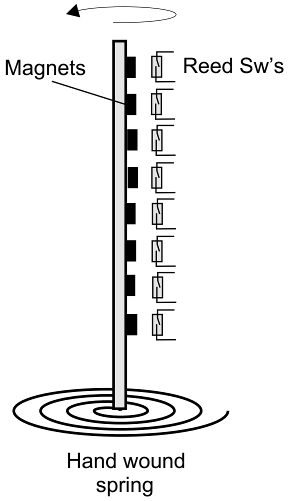
Magnets
Reed Sw's
Hand wound spring

Copyright 2002 - 2004 Creative Science & Research

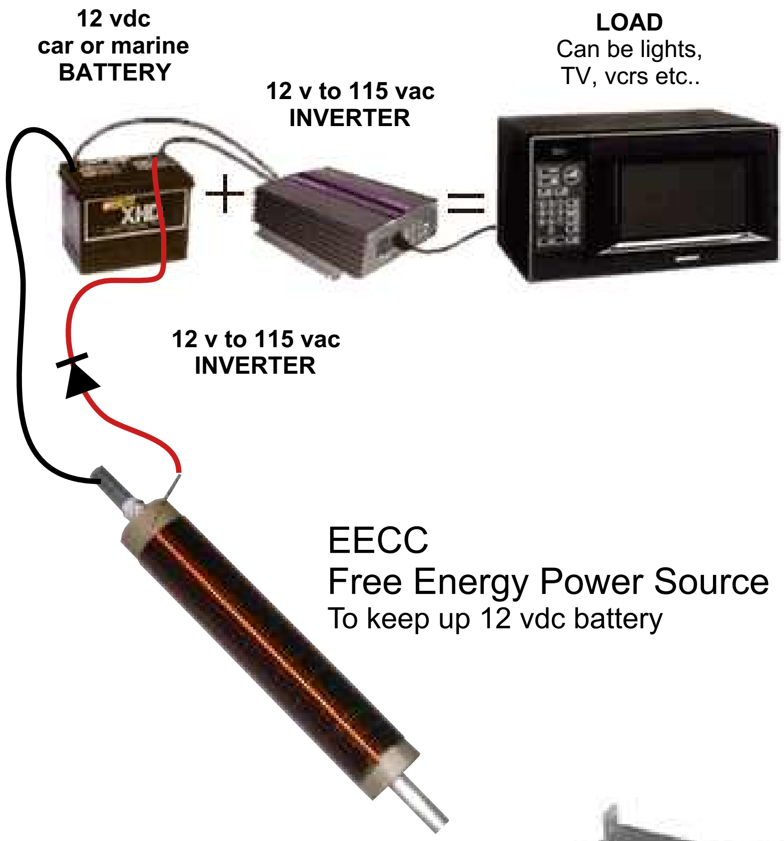
12 vdc
car or marine
BATTERY
12 v to 115 vac
INVERTER
LOAD
Can be lights,
TV, vcrs etc..
+
=
12 v to 115 vac
INVERTER
EECC
Free Energy Power Source
To keep up 12 vdc battery

Copper

Symbols

= One way DIODE

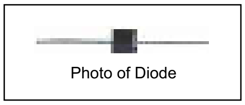

Photo of Diode

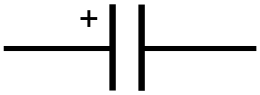

Capacitor (Electrolytic)

Photo of one type
of Capacitor

Copyright 2002 - 2004 Creative Science & Research

Suppliers List

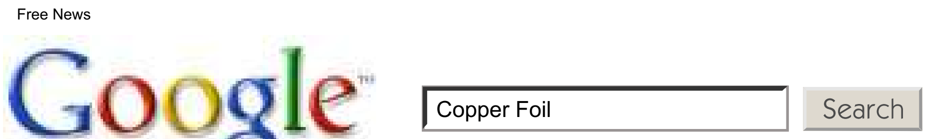

See www.google.com

1. Http://www.delphiglass.com/

Use there search engine to find copper foil tape

2. http://www.glassmart.com/foil_menu.htm
3. http://www.globalsources.com

Search Google.com for:

1. Copper flashing
2. Inverters
3. Batteries
4. Copper coated wire
5. Copper wire
6. Electronic parts and supplies
7. Diodes
8. Capacitors
9. Free energy from the earth
10. Zinc rods

Copyright 2002 - 2004 Creative Science & Research

EECC update

Step One

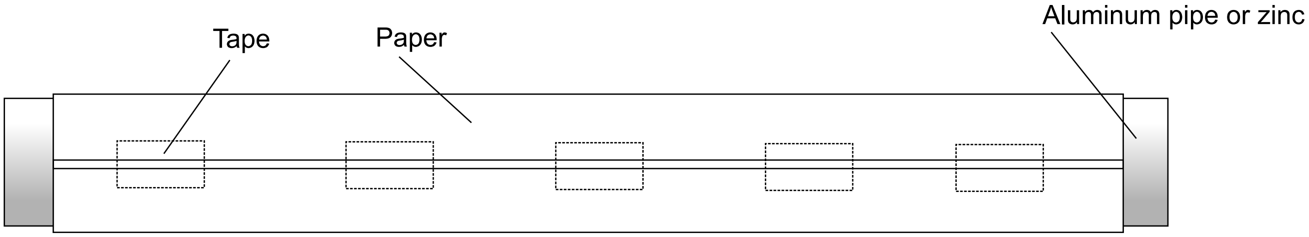
Tape
Paper
Aluminum pipe or zinc

Step two

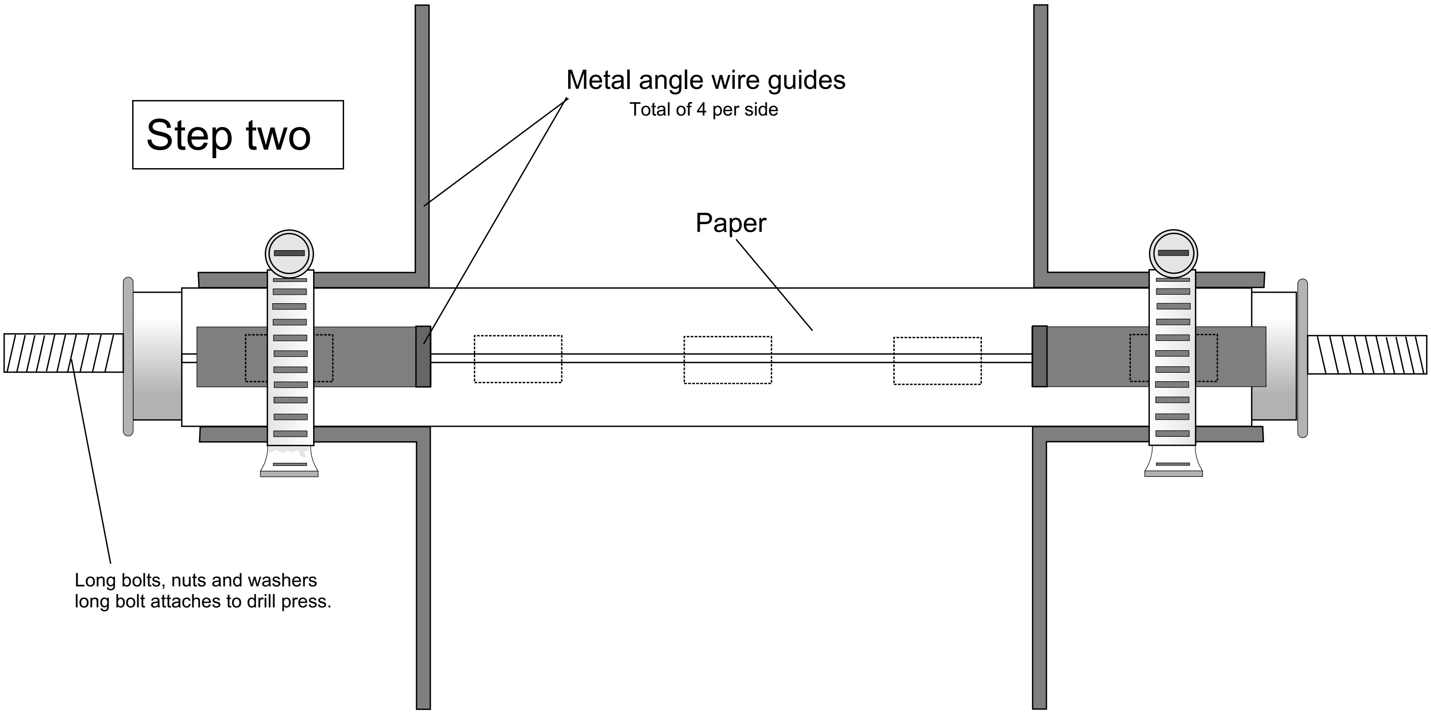
Metal angle wire guides
Total of 4 per side
Paper
Long bolts, nuts and washers
long bolt attaches to drill press.

Apply first layer of wire, remove, metal wire guides, apply more paper, then repeat over and over again until you have about a 5 to 6 inch diameter roll of wire and paper. Makes a very powerful earth capacitor!

Copyright 2002 - 2004 Creative Science & Research

EECC update

These are just a few extra idea's that we threw together. I hope they are helpful to you. Let us know how you like these plans. Send us photo's of what you have done we love to hear from our customers. If you got any new free energy idea's let us know we can research them for you. Many customers do not have the extra research time that we have.

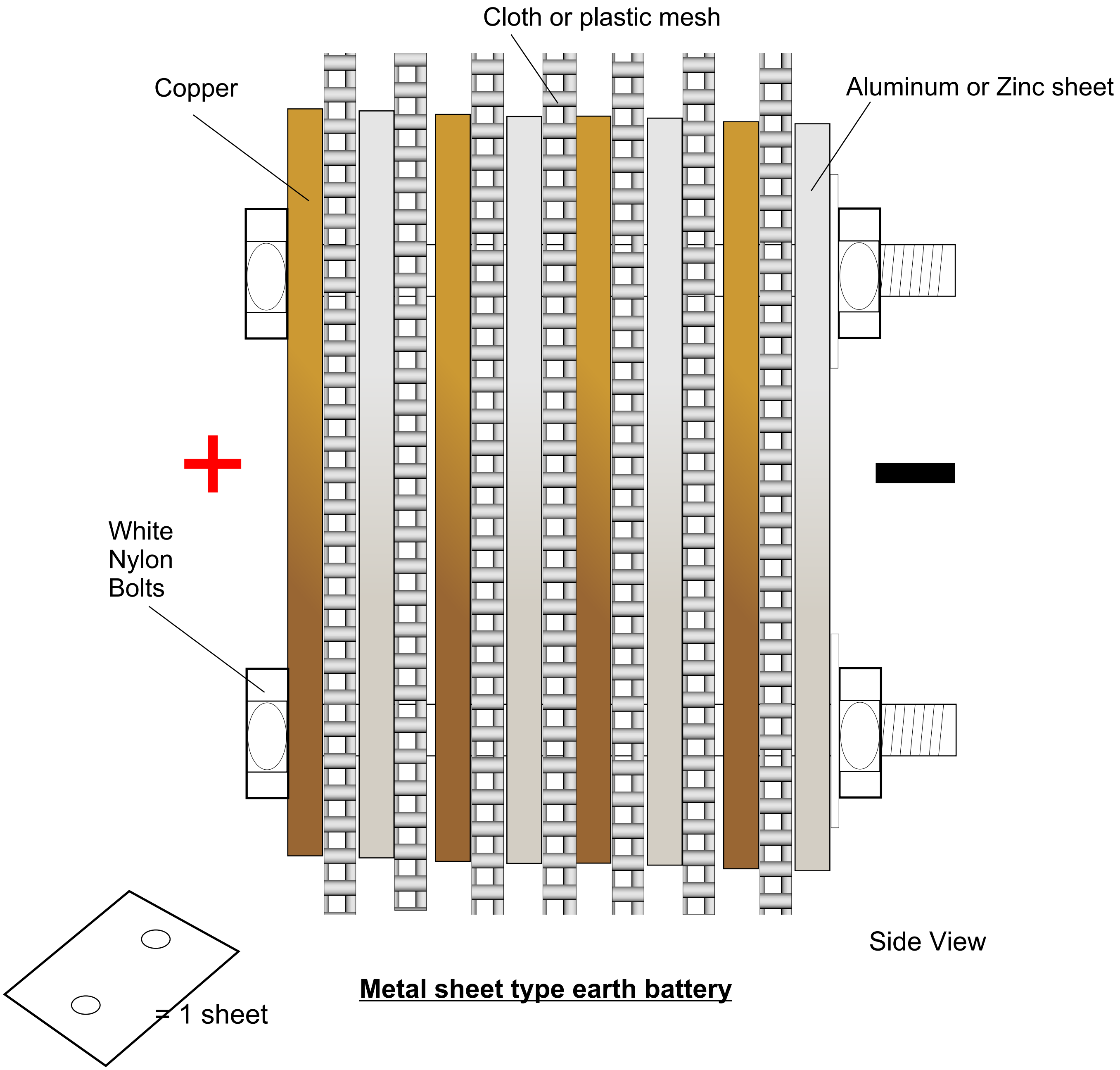
Cloth or plastic mesh
Copper
Aluminum or Zinc sheet
+
−
White Nylon Bolts
Side View
= 1 sheet

Metal sheet type earth battery

US007439712B2

(12) **United States Patent**
McCowen

(10) **Patent No.: US 7,439,712 B2**
(45) **Date of Patent: Oct. 21, 2008**

(54) **ENERGY COLLECTION**

(76) Inventor: **Clint McCowen**, 1902 Rue La Fontaine, Navarre, FL (US) 32566

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 509 days.

(21) Appl. No.: **11/358,264**

(22) Filed: **Feb. 21, 2006**

(65) **Prior Publication Data**

US 2007/0195481 A1 Aug. 23, 2007

(51) **Int. Cl.**
H02N 1/00 (2006.01)

(52) **U.S. Cl.** .... **322/2 A**; 307/149; 361/225; 361/117; 174/2; 320/101

(58) **Field of Classification Search** .... 361/220, 361/117, 212, 233, 221, 222, 225; 322/2 A, 322/2 R; 307/149; 174/2, 3; 320/101
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427	A *	5/1901	Palencsar	307/149
911,260	A *	2/1909	Pennock	307/149
1,014,719	A *	1/1912	Pennock	307/149
4,104,696	A	8/1978	Cochran, Jr.	
5,047,892	A	9/1991	Sakata et al.	
6,735,830	B1	5/2004	Merciel	
6,974,110	B2	12/2005	Grandics	
2003/0107869	A1	6/2003	Pan et al.	
2004/0160711	A1	8/2004	Stumberger	
2004/0212945	A1	10/2004	Sprenger et al.	

* cited by examiner

Primary Examiner—Mark T Le
(74) *Attorney, Agent, or Firm*—Thomas, Kayden, Horstemeyer & Risley, LLP

(57) **ABSTRACT**

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support wire system supported by poles. The support wire system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but carbon and graphite are preferred. Diodes may be used to restrict the backflow or loss of energy.

28 Claims, 10 Drawing Sheets

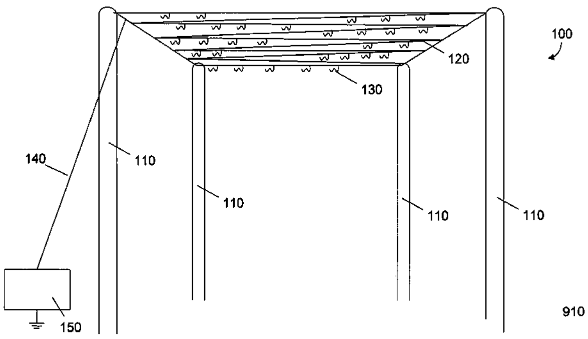
100
120
130
140
110
110
110
110
150

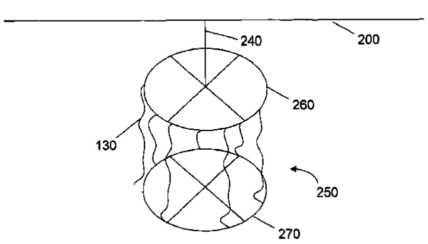
240
200
260
130
250
270

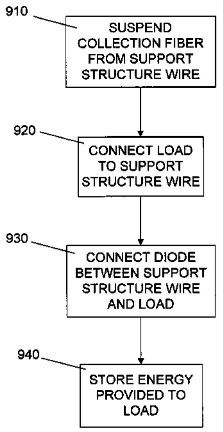
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

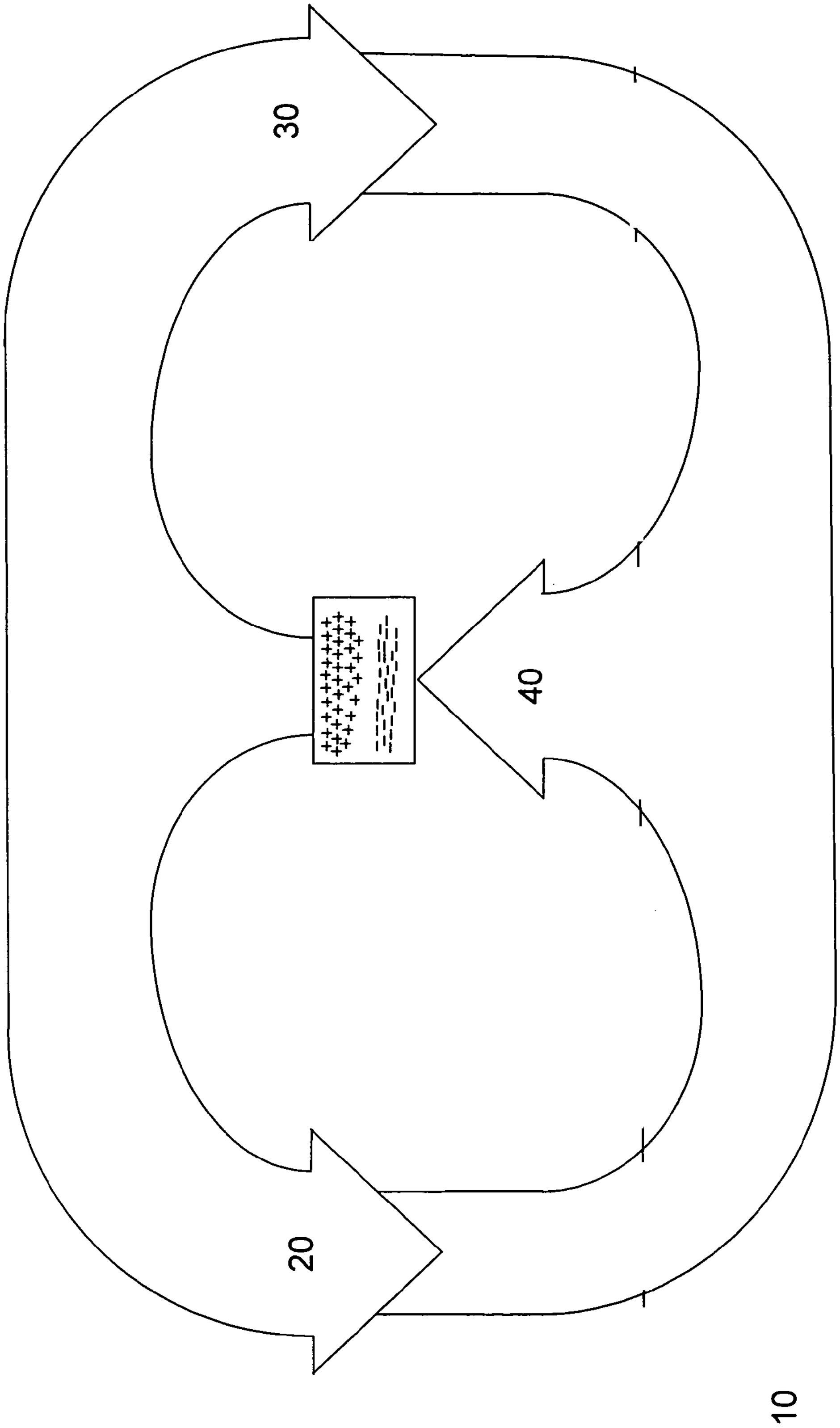

FIGURE 1

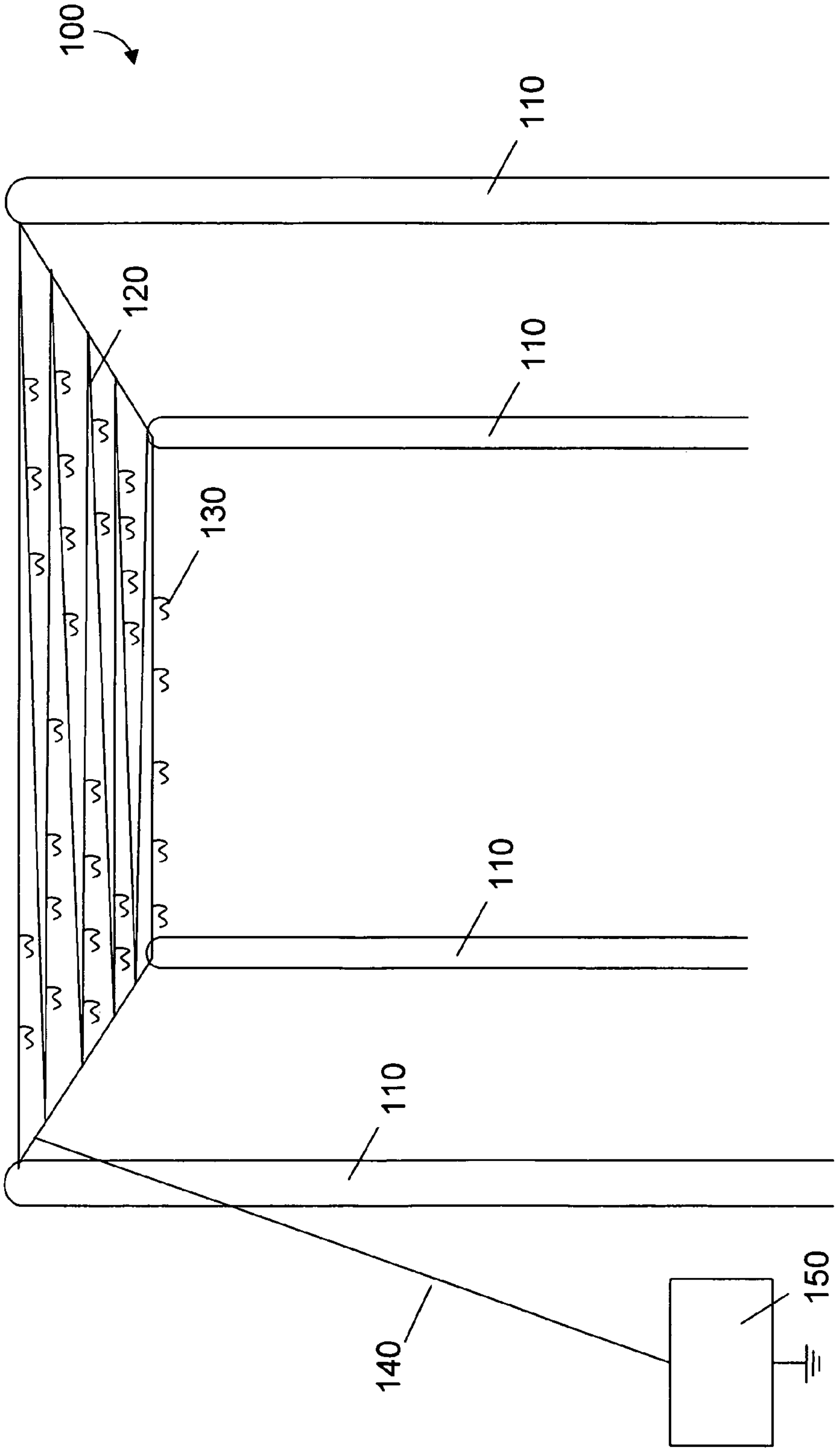
100
110
110
110
110
120
130
140
150

FIGURE 2

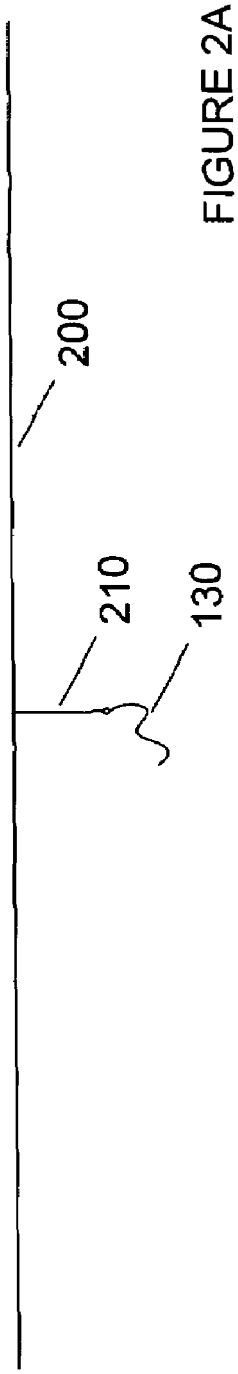

FIGURE 2A

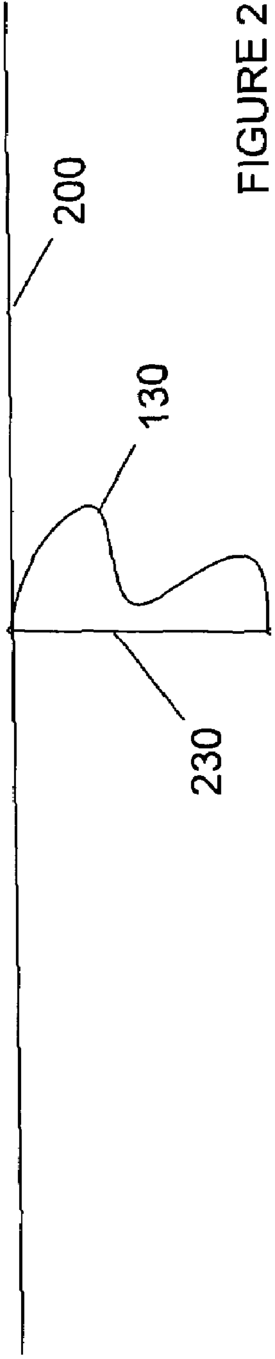

FIGURE 2B

FIGURE 2C

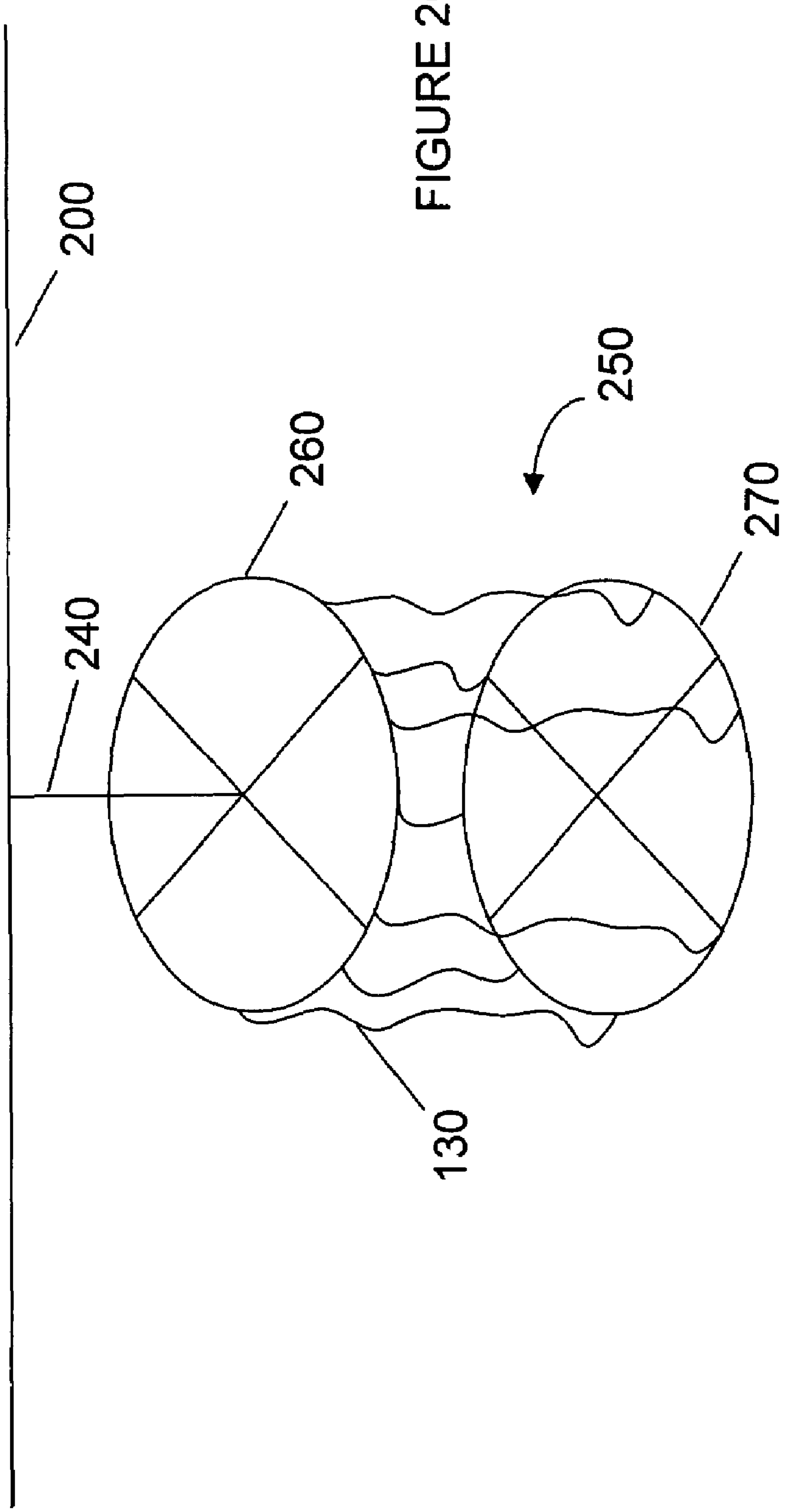
200
250
260
270
240
130

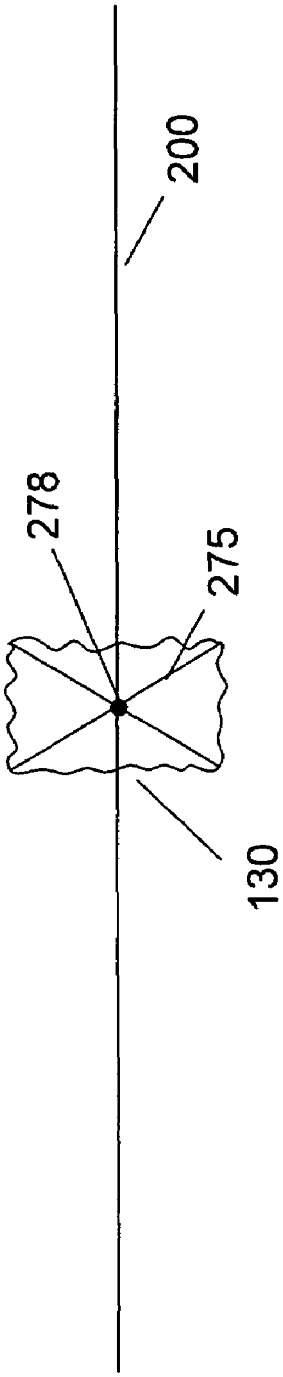
200
278
275
130

FIGURE 2D

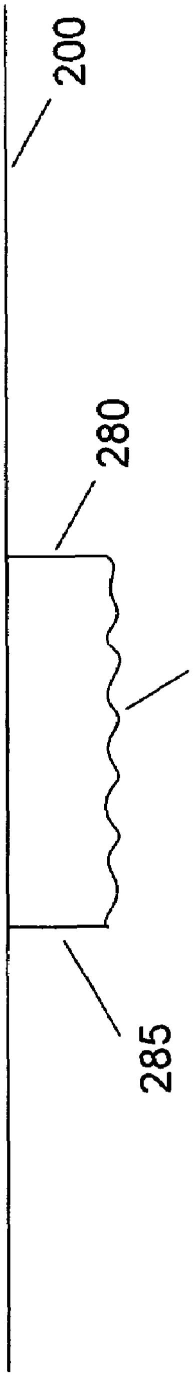
200
280
130
285

FIGURE 2E

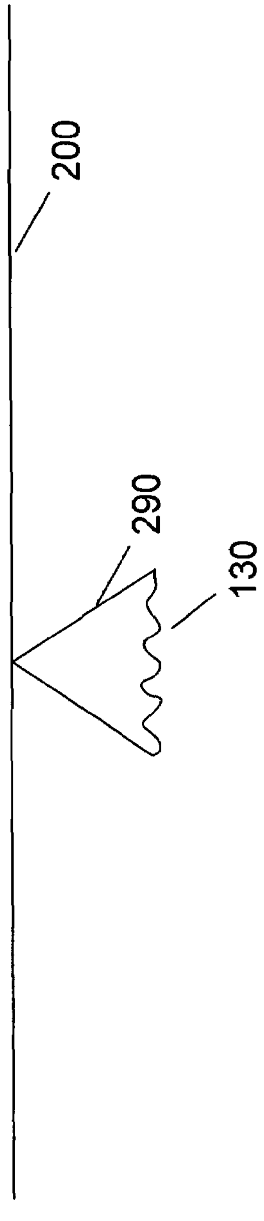

FIGURE 2F

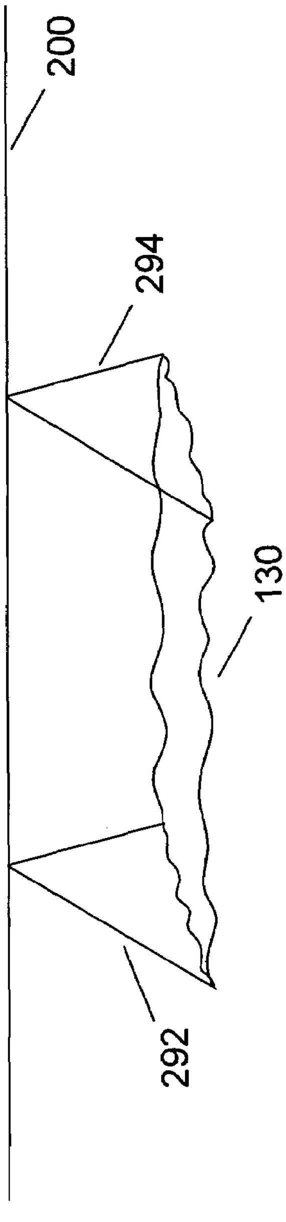
200
294
130
292

FIGURE 2G

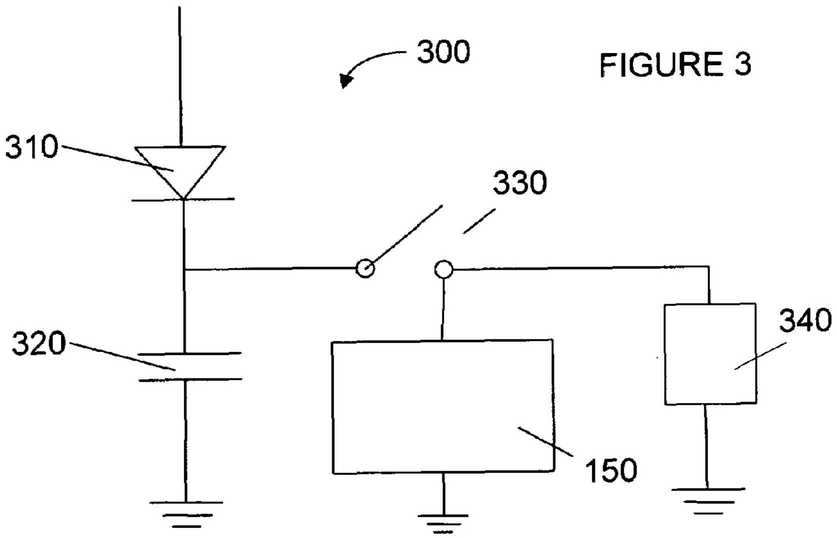
300
310
330
320
340
150

FIGURE 3

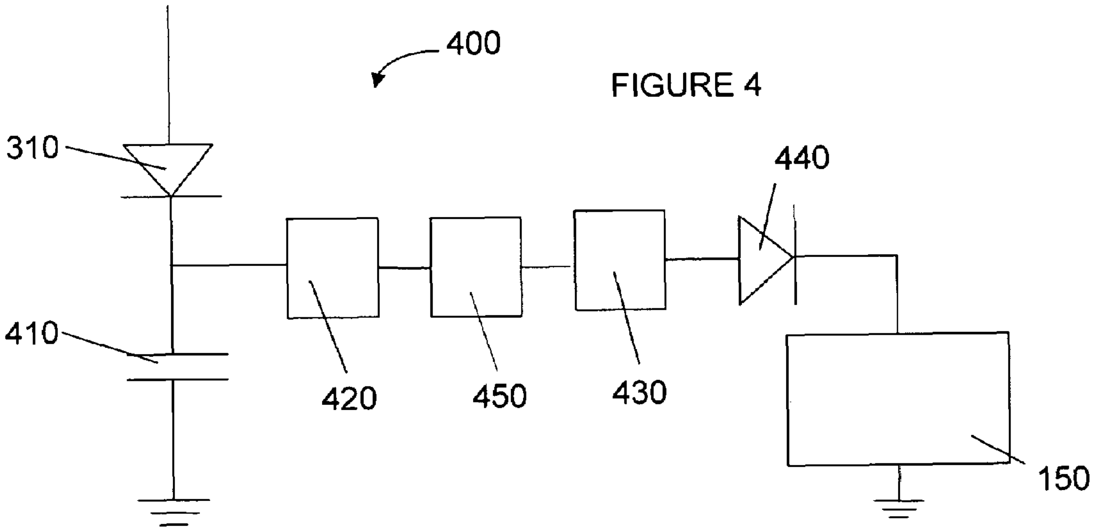
400
310
440
410
420
450
430
150

FIGURE 4

FIGURE 5

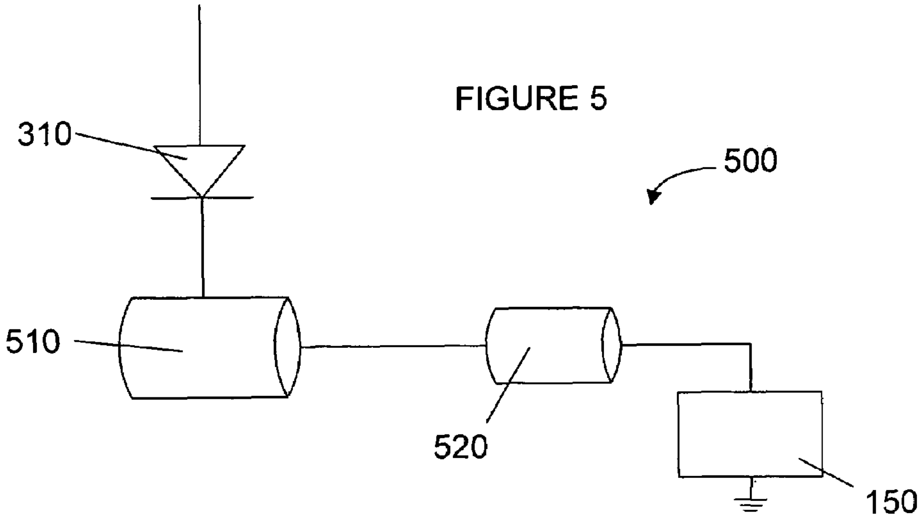
310
500
510
520
150

FIGURE 6

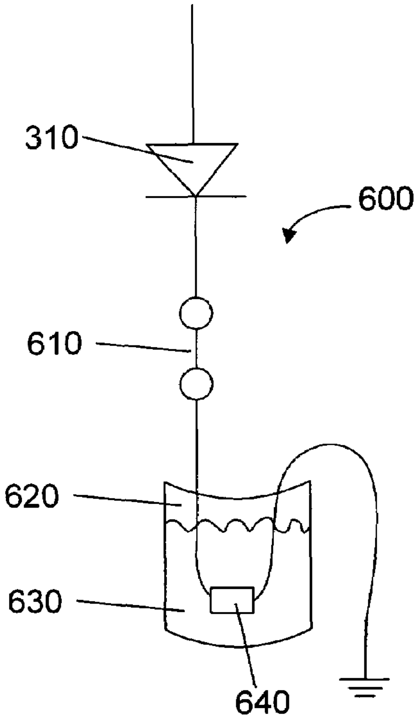
310
600
610
620
630
640

FIGURE 7

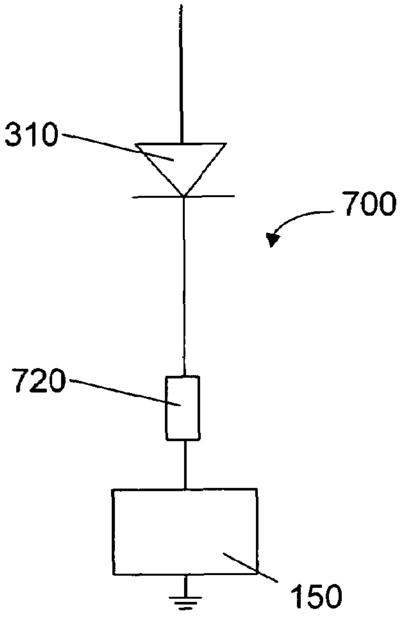
310
700
720
150

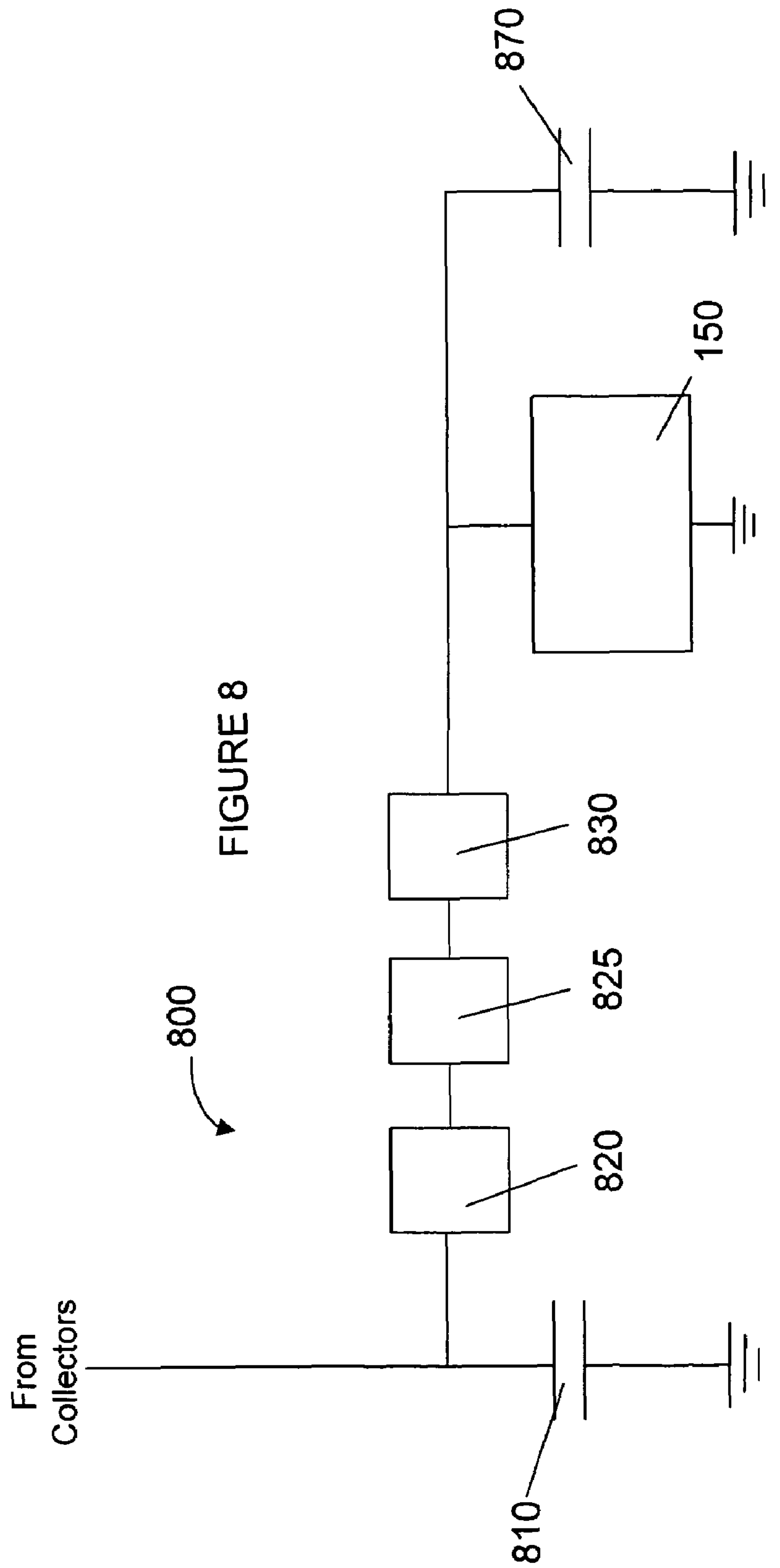
From Collectors
810
800
820
825
830
150
870

FIGURE 8

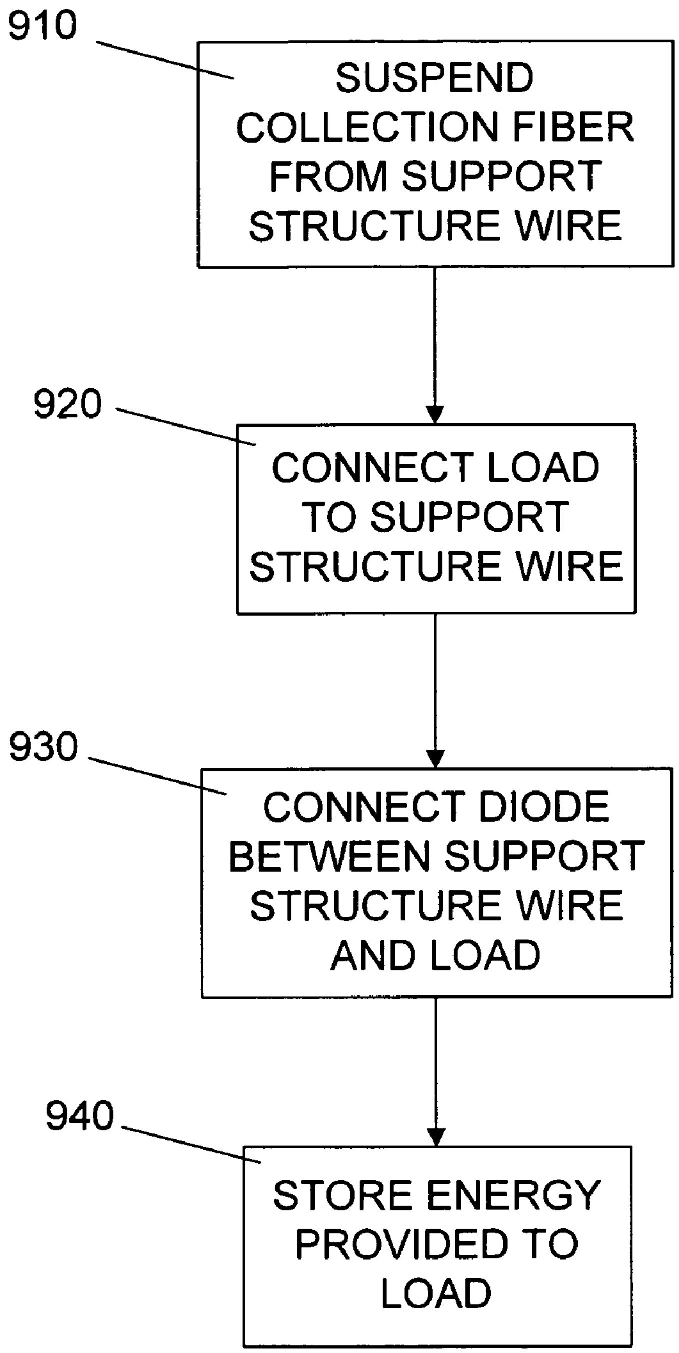
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

ENERGY COLLECTION

TECHNICAL FIELD

The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure wire elevated above a ground level, at least one collection fiber electrically connected to the support structure wire; a load electrically connected to the support structure wire; and a diode electrically connected between the load and at least one collection fiber.

Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection fiber from a support structure wire elevated above ground level, the fiber electrically connected to the support structure wire; providing a load with an electrical connection to the support structure wire to draw current; and providing a diode electrically connected between the collection fiber and the load.

Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. **1** is a circuit diagram of a weather energy circuit.

FIG. **2** is a perspective view of an exemplary embodiment of many energy collectors elevated above ground by a structure.

FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

FIG. **2**B is a side view of an exemplary embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

FIG. **2**D is a side view of an exemplary embodiment of a support structure for multiple energy collection fibers.

FIG. **2**E is a side view of a support structure for an energy collection fiber.

FIG. **2**F is a side view of an exemplary embodiment of a support structure for an energy collection fiber.

FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

FIG. **3** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **4** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **5** is a circuit diagram of an exemplary embodiment of an energy collection circuit for powering a generator and motor.

FIG. **6** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

FIG. **7** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

FIG. **8** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy.

FIG. **9** is a flow diagram of an exemplary embodiment of collecting energy with a collection fiber.

DETAILED DESCRIPTION

Electric charges on conductors reside entirely on the external surface of the conductors, and tend to concentrate more around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An exemplary embodiment of this disclosure takes advantage of this

property, among others, to collect and use the energy generated by an electric field in the atmosphere. Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An exemplary embodiment utilizes carbon or graphite fibers for static electricity collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including aluminum or steel, but most notably, copper. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

An exemplary embodiment of the collection fibers as collection device **130** includes graphite or carbon fibers. Graphite and carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, or graphite fibers are examples that demonstrate exemplary collection ability.

In at least one exemplary embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than **100** volts per meter under some conditions. When support wire **120** is suspended in the air at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An exemplary embodiment of the collection device comprises the diode with no collection fiber. A preferred embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restricts energy that may be collected through one fiber from escaping through another fiber.

Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2A-2G** using a collection fiber embodiment. FIG. **2A** presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**.

Likewise, FIG. **2B** shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

FIG. **2C** presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

FIG. **2D** presents another exemplary embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2E** provides another exemplary embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

FIG. **2F** presents another exemplary embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2G** provides two supports as provided in FIG. **2F** such that at least two support members **292**, **294** may be connected to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and between multiple support structures. A plurality of diodes may be

placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**.

FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load.

Capacitor **320** may be connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

In an exemplary embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel. Non-limiting examples of spark gap **420** include mercury-reed switches and mercury-wetted reed switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground.

The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes may be placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches and mercury-wetted reed switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel.

FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit. Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

FIG. **8** presents exemplary circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches and mercury-wetted reed switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

It should be emphasized that the above-described embodiments of the present disclosure, particularly, any “preferred” embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A method of collecting energy comprising:

suspending at least one collection device from a plurality of support structure wires elevated above a ground level, the collection device including a carbon or graphite fiber that collects electrical enemy from the air and is electrically connected to the support structure wire; and

providing a load with an electrical connection to the plurality of support structure wires to draw current.

2. The method of claim **1**, wherein the collection device comprises a diode.

3. The method of claim **1**, wherein the collection device comprises a diode electrically connected between the fiber and the load.

4. The method of claim **1**, further comprising storing energy provided to the load.

5. The method of claim **4**, wherein storing energy provided to the load comprises storing energy in a capacitor or an inductor.

6. The method of claim **2**, wherein the diode is elevated relative to the ground level.

7. A system of energy collection comprising:

a plurality of support structure wires elevated above ground level; at least one collection device including at least one carbon or graphite fiber that collects electrical enemy from the air and is electrically connected to the plurality of support structure wires; and

a load electrically connected to the plurality of support structure wires.

8. The system of claim **7**, wherein the collection device comprises a diode.

9. The system of claim **7**, wherein the collection device comprises a diode electrically connected between the load and the fiber.

10. The system of claim **8**, wherein the diode is elevated relative to the ground level.

11. The system of claim **7**, further comprising a diode electrically connected between the at least one collection device and the plurality of support structure wires.

12. The system of claim **7**, wherein the fiber comprises a first end and a second opposing end, and wherein the system further comprises a support structure connected to both ends of the fiber.

13. The system of claim **7**, comprising:

said at least one fiber including a plurality of collection fibers; a support frame having a top side and a bottom side; and a conducting connecting wire between the plurality of support structure wires and the support frame, wherein one end of each of a plurality of collection fibers is connected to the top side of the support frame, and the opposing end of each of a plurality of collection fibers is connected to the bottom side of the support frame.

14. The system of claim **7**, further comprising a rigid structure, the rigid structure comprising multiple supports extending outward from a single point of the support structure, wherein each end of the fiber is connected to an end of the multiple supports.

15. The system of claim **7**, further comprising a rigid structure, the rigid structure comprising multiple supports extending outward from multiple points of the support structure, wherein each end of the fiber is connected to an end of the multiple supports.

16. The system of claim **14**, further comprising a plurality of rigid structures wherein said at least one fiber includes collection fibers connected between a plurality of the rigid structures.

17. The system of claim **7**, further comprising:

a switch connected in series between the plurality of support structure wires and the load; and

a capacitor connected in parallel with the switch and the load.

18. The system of claim **17**, wherein the switch comprises an interrupter connected between the plurality of support structure wires and the load.

19. The system of claim **18**, wherein the interrupter comprises one of a fluorescent tube, a neon bulb, an AC light, or a spark gap.

20. The system of claim **18**, further comprising a transformer connected between the interrupter and the load.

21. The system of claim **7**, further comprising:

a motor for providing power the motor connected between the plurality of support structure wires and the load; and

a generator powered by the motor.

22. The system of claim **7**, wherein the load comprises a spark gap in a container of fluid, and the load is used to produce a chemical reaction.

23. The system of claim **22**, wherein the fluid comprises water and the chemical reaction comprises the production of hydrogen and oxygen.

24. The system of claim **7**, further comprising a fuel cell between the plurality of support structure wires and the load.

25. The system of claim **7**, wherein the load comprises a fuel cell.

26. The system of claim **24**, wherein the fuel cell produces hydrogen and oxygen.

27. The system of claim **24**, further comprising a diode connected between the plurality of support structure wires and the fuel cell.

28. A system of collecting energy comprising:

means for suspending carbon or graphite collection fibers, the means elevated above a ground level, the collection fibers collecting electrical enemy from the air and electrically connected to the means for suspending collection fibers;

means for inducing current flow, the means for inducing current flow electrically connected to the means for suspending collection fibers; and

means for restricting the backflow of charge carriers, the means for restricting the backflow of charge carriers electrically connected between the collection fibers and the means for inducing current flow.

* * * * *

No. 674,427. Patented May 21, 1901.

A. PALENCSÁR.

APPARATUS FOR COLLECTING ATMOSPHERIC ELECTRICITY.

(Application filed July 10, 1900.)

(No Model.) 3 Sheets—Sheet 1.

Fig:1.

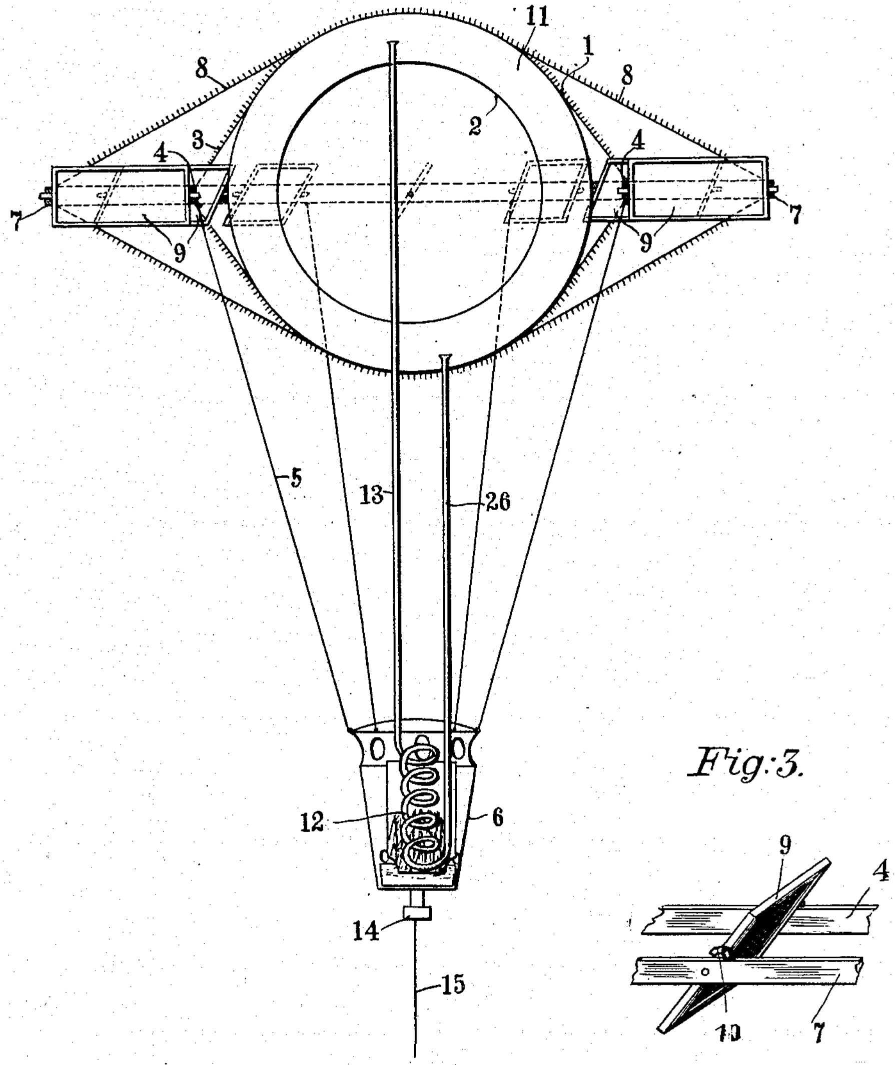

11
1
8
8
2
3
4
4
7
7
9
9
9
9
5
13
26
12
6
14
15

Fig:3.

9
4
10
7

WITNESSES:
Ella L. Giles
Otto Munk

INVENTOR
Andor Palencsár
BY Richards
ATTORNEYS

THE NORRIS PETERS CO., PHOTO-LITHO., WASHINGTON, D. C.

No. 674,427. Patented May 21, 1901.

A. PALENCSÁR.

APPARATUS FOR COLLECTING ATMOSPHERIC ELECTRICITY.

(Application filed July 10, 1900.)

(No Model.) 3 Sheets—Sheet 2.

Fig: 2.

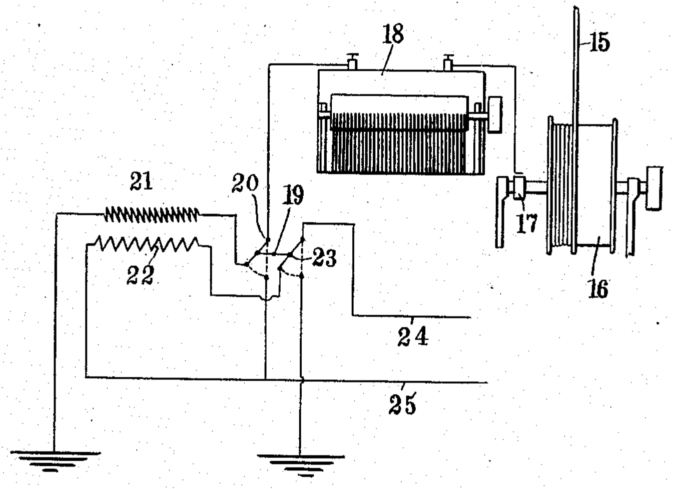

WITNESSES:
Ella L. Giles
Otto Munk

INVENTOR
Andor Palencsár
BY
Richards & Co
ATTORNEYS

THE NORRIS PETERS CO., PHOTO-LITHO., WASHINGTON, D. C.

No. 674,427. Patented May 21, 1901.

A. PALENCSÁR.

APPARATUS FOR COLLECTING ATMOSPHERIC ELECTRICITY.

(Application filed July 10, 1900.)

(No Model.) 3 Sheets—Sheet 3.

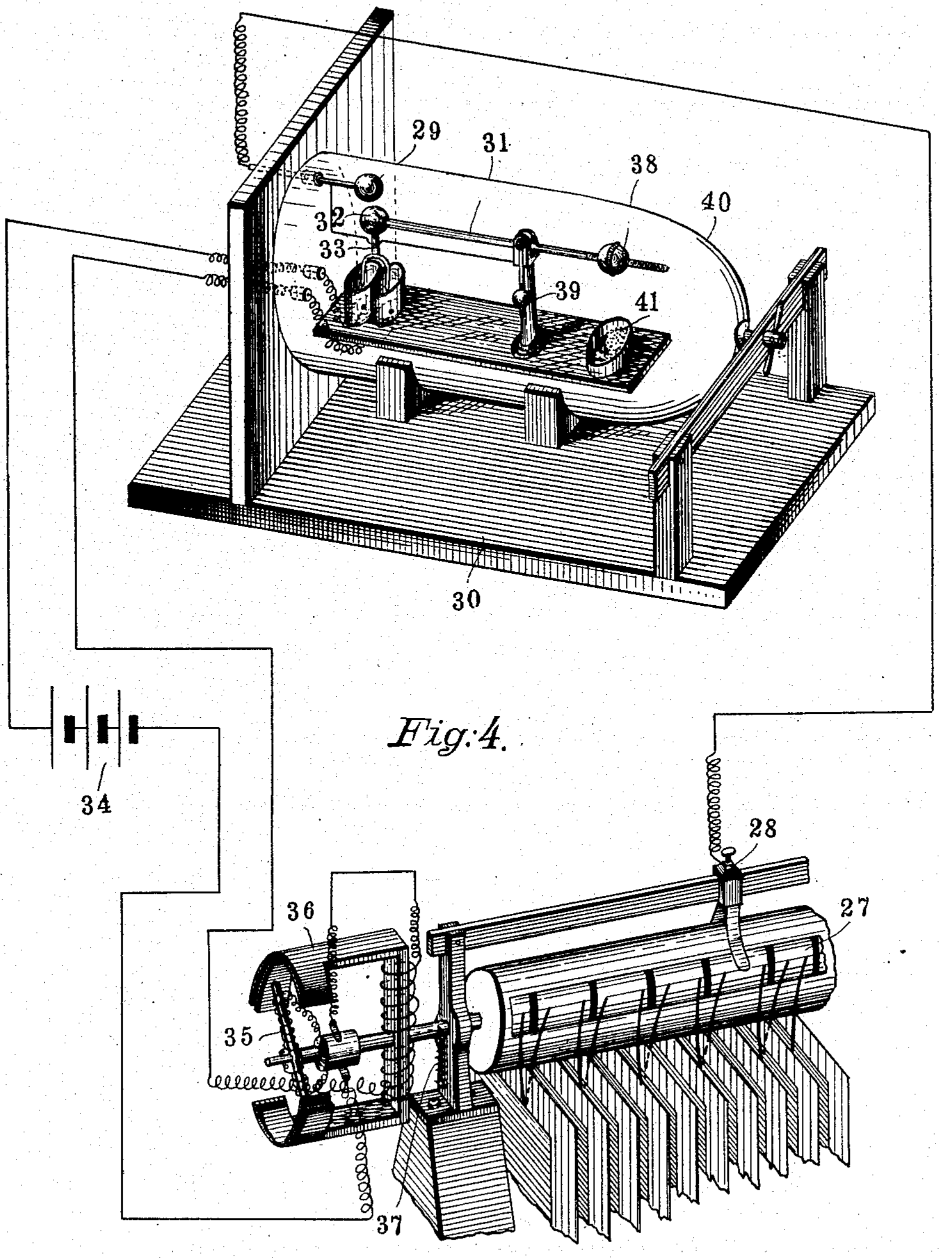
29
31
38
40
32
33
39
41
30
34
28
36
27
35
37

Fig. 4.

WITNESSES:
Ella L. Giles
Otto Munk

INVENTOR
Andor Palencsár
BY
Richards & Co
ATTORNEYS

THE NORRIS PETERS CO., PHOTO-LITHO., WASHINGTON, D. C.

UNITED STATES PATENT OFFICE.

ANDOR PALENCSÁR, OF BUDA-PESTH, AUSTRIA-HUNGARY.

APPARATUS FOR COLLECTING ATMOSPHERIC ELECTRICITY.

SPECIFICATION forming part of Letters Patent No. 674,427, dated May 21, 1901.

Application filed July 10, 1900. Serial No. 23,102. (No model.)

To all whom it may concern:

Be it known that I, ANDOR PALENCSÁR, a subject of the King of Hungary, residing at Buda-Pesth, Austria-Hungary, have invented a certain new and useful Apparatus for Collecting Atmospherical Electricity, of which the following is a full, clear, and exact specification.

The present invention concerns an apparatus for collecting and driving or conducting atmospherical electricity which renders it practicable to obtain material quantities of the same in serviceable form.

The experiments made hitherto have been limited to deviating or conducting the electricity by simple lightning-rods without making allowance for the progress of the theory of atmospherical electricity. It is obvious that only extremely small quantities of electricity can be collected in this manner, and, moreover, it was obtained in a form which entirely precluded its utilization. It is evident that a source of electricity of irregular yield can only be utilized by means of accumulators and for charging accumulators only a current of constant potential can be employed. The potential of electricity derived or deviated by means of a lightning-rod varies within wide limits, and, moreover, it is so high that it cannot be used at all for charging accumulators. All these drawbacks are remedied by the present invention, by which the atmospheric electricity is obtained in larger quantities with as low and constant a potential as may be desired.

The idea underlying the invention is based on the modern theory of atmospheric electricity, according to which it is produced by the condensation of steam or aqueous vapors, and that the increase of potential is effected by the concentration of the small drops of water into larger ones, as the proportion of the surface of the drops to the volume of same is materially reduced thereby. According to this theory the water-droplets floating in the layers of air are considered as vehicles or carriers of the electricity, and a rational system of the deviation of the atmospheric electricity must derive it from the water-drops. This is attained by the present invention in the following manner: A collecting-body of as large a surface as possible is provided with sharp needles. This is moved in the higher layers of air, while being continually heated by a suitable heating device. Owing to the heat the water-drops immediately ambient to the collecting-body will be evaporated, their capacity is gradually reduced, while the potential of the charge grows until it reaches an infinite height with infinite smallness of the drops. It is readily apparent that the whole charge of the drops as soon as they have been evaporated will have passed to the collector or collecting-body, from which it can be conducted. For the purpose of replacing the evaporated drops, which have been deprived of their charge by new-charged drops, the collector is moved in relation to the ambient air.

One way of carrying out the invention is represented in diagram in the accompanying drawings, in which—

Figure 1 is a longitudinal section of the collecting-balloon. Fig. 2 shows the deviating and conversion device, and Fig. 3 a detail. Fig. 4 is a diagrammatic perspective view of a form of automatic regulator for the rheostatic machine.

The apparatus consists of a balloon having two walls and covered with a light wire net, preferably of aluminium wire, said net being studded with needles. Besides, the balloon carries the net 3, on which the ring 4, made of a solid but light material, (wood, cane, &c.,) is fixed. This ring carries the basket 6 by means of the cords or ropes 5. On a level with the ring 4 is the ring 7, which is kept spread by the blades or wings 9, which are journaled in a manner to rotate easily. The ring 7 is kept fixedly in position by the cords or ropes 8. The blades or wings consist of a frame covered with light material, and their rotation in either direction is limited by the stops or ledges 10. The blades or wings form advantageously an angle of sixty to seventy degrees with the vertical line.

All wire nets, ropes, rings, and blades or wings may be covered with small metal needles which are electrically connected with each other.

From the space 11, formed between the double walls of the balloon, the pipe 26 leads from the lowest point of the same to the serpentine 12, arranged in the basket 6, which in turn is

connected with the pipe 13, ending in the upper part of the space 11 between the two walls 12. The serpentine is heated by means of a suitable source of heat, whereby a warm current of gas or air circulates continually between the double walls of the balloon.

Under the basket the ball-bearing 14 is arranged in electrical connection with the wire nets, and its stud is electrically connected with the carefully-insulated light though sufficiently strong cable 15.

On the earth's surface is a winch 16, Fig. 2, by means of which the balloon can be made to ascend or descend as soon as the interior space of the balloon is filled with illuminating gas or hydrogen.

The end of the cable-core is soldered to a collector arranged upon and insulated from the axle of the winch, and the electricity is conducted from this collector by means of a sliding contact. The collecting of the electricity takes place by moving the balloon continuously up and down by means of the winch. In this movement the balloon is turned by means of the wings or blades 9, which are adapted to turn the balloon always in the same direction whether ascending or descending, as in the change from ascending into descending of the balloon, or vice versa, the blades are turned over by the aerial resistance, and thus impart the rotating motion to the balloon in the same direction. In order to avoid torsion of the cable, the ball-bearing 14 is provided. This up-and-down motion and rotating of the balloon accomplishes the purpose of bringing it into contact with as many water particles floating in the air as possible.

As the electricity conducted from the collector-wing 17 possesses a much too high and varying potential for making its direct application practicable, and as it is usual with an irregular source of electricity to first charge accumulators and to further utilize the easily-regulated current of the same only, it becomes necessary to seek to maintain the electricity conducted from the collector 17 for the charging of the accumulators at a constant potential and convert the potential to a much lower one; but as we deal in this case with a direct current ordinary converters cannot be used for this purpose. Moreover, the electricity possesses in this case a much too high potential, so that with the employment of ordinary converters the greatest part of the collected electricity would be lost again. The only practical method for this purpose is the converting by means of the Planté rheostatic machine, by which this high-voltage electricity can be transformed almost without any loss whatever. Thus the question of conversion would be solved, and only the question of maintenance of constant potential remains. This is obtained in the following manner: The rheostatic machine 18 or only a part of the plates is connected with an electrometer of any construction, the movable part of which closes a contact which eventually actuates an electromagnet which effects the switching or reversion of the rheostatic machine. After the switching of the rheostatic machine it is discharged, the potential falls to zero, (*nil*,) and the electrometer resumes its initial position, whereby the current of the electromagnet which effects the reversion is interrupted and the plates of the rheostatic machine are reconnected to potential. The machine is then ready for renewed charge and is again discharged when the determined potential is reached. This action is continually repeated as long as the apparatus is in operation.

A form of the automatic regulator for the rheostatic machine is represented in Fig. 4. 27 is the contact-cylinder of the rheostatic machine, on which, for clearness sake, only the contacts for the charging position of the condenser-plates are shown, while the contacts for the discharging position, which come into action after the contact-cylinder has been turned, are omitted. The coating of the condenser-plates is connected in electrical circuit with the stationary ball 29 and the movable ball 32 of the charge-meter 30. When the charge of the rheostatic machine rises, the ball 32, arranged on the one extremity of the lever 31, is repelled and at a certain stroke actuated by means of the fork 33, which is fastened on the ball 32, and by dipping into a mercury-cup it closes the circuit of the source of current 34. This current passes through the windings of the anchor or armature 35, fixed on the shaft of the contact-cylinder, and it also passes through the electromagnet 36. Thereby the contact-cylinder is turned by a certain angle and the rheostatic machine is thus reversed. Then if the tension decreases by discharge and the repulsion of the balls 29 32 declines so far that the contact is interrupted at 33 the spring 37 turns the contact-cylinder into its normal position and the rheostatic machine is again switched to tension. The regulation of the electrometer is effected by the adjustable weights 38 39. When larger quantities of electricity shall be derived or deviated, two rheostatic machines may operate alternately so, that while the one is being discharged the other can be charged. It is readily apparent that if the capacity of the rheostatic machine is not changed the switching or reversion by the electrometer will always take place at the same potential of the rheostatic machine, and as the number of plates, and hence the proportion of conversion, remains the same the current impulses derived or deviated from the rheostatic machine will also have the same potential.

The irregularities of the source of electricity change the interval of time in which the charges follow each other; but as long as the potential remains constant this has no injurious effect on the charge of the accumulators. The current derived. or deviated from the

rheostatic machine can be further transformed by an ordinary converter 21 22, and whenever it is sufficiently constant it can be utilized without the intervention of the accumulators. The converters can be connected or disconnected by means of the double switch 19 20 23.

24 and 25 are the conducting-wires, which run either directly to the place of consumption or to an accumulator-battery.

Having now described my invention, what I claim as new, and desire to secure by Letters Patent, is—

1. An apparatus for collecting atmospheric electricity comprising a collecting-body adapted to be kept in motion, heating means for said body, and a rheostatic machine and a converter connected with the said body electrically, substantially as described.

2. An apparatus for collecting atmospheric electricity for storage comprising a collecting-body adapted to be kept in motion in the ambient air, heating means for said body, a conductor leading from said body, a rheostatic machine connected with the said conductor, an electrometer connected with said rheostatic machine electrically, a contact controlled by the electrometer, and an electromagnet controlled by said contact, said electromagnet controlling the reversing of the rheostatic machine, substantially as described.

3. In combination, the balloon-like collecting-body, having the collecting-points, means for heating the interior space of said body, a conductor leading from the balloon-like body, and electrical devices for receiving the current therefrom, substantially as described.

4. In combination, the balloon-like collecting-body, means for turning the same constantly in one direction in both ascending and descending, conducting means leading from the balloon and electrical devices for receiving the current therefrom, substantially as described.

In witness whereof I have hereunto set my hand in presence of two witnesses.

ANDOR PALENCSÁR.

Witnesses:
EUGÈNE HARRÄNJO,
PAUL BÖLESKEY.

Graphene/Graphite Atmospheric Electricity Collectors – Plus Horrific Hexacopter Crash!

by lasersaber | Jun 2, 2014 | Atmospheric Electricity

While on family vacation I was able to swing by and check out this atmospheric energy test site. They have some amazing energy harvesting technology and encourage replication. Since the visit I have had a non-stop stream of new ideas running through my head, but...

3D Printed Wireless Motor

Mini AtmoMotor

by lasersaber | Mar 10, 2014 | Atmospheric Electricity, Motors

This is the most sensitive AtmoMotor design yet. The motor has been designed with easy construction in mind. It makes an excellent wireless demonstration motor.

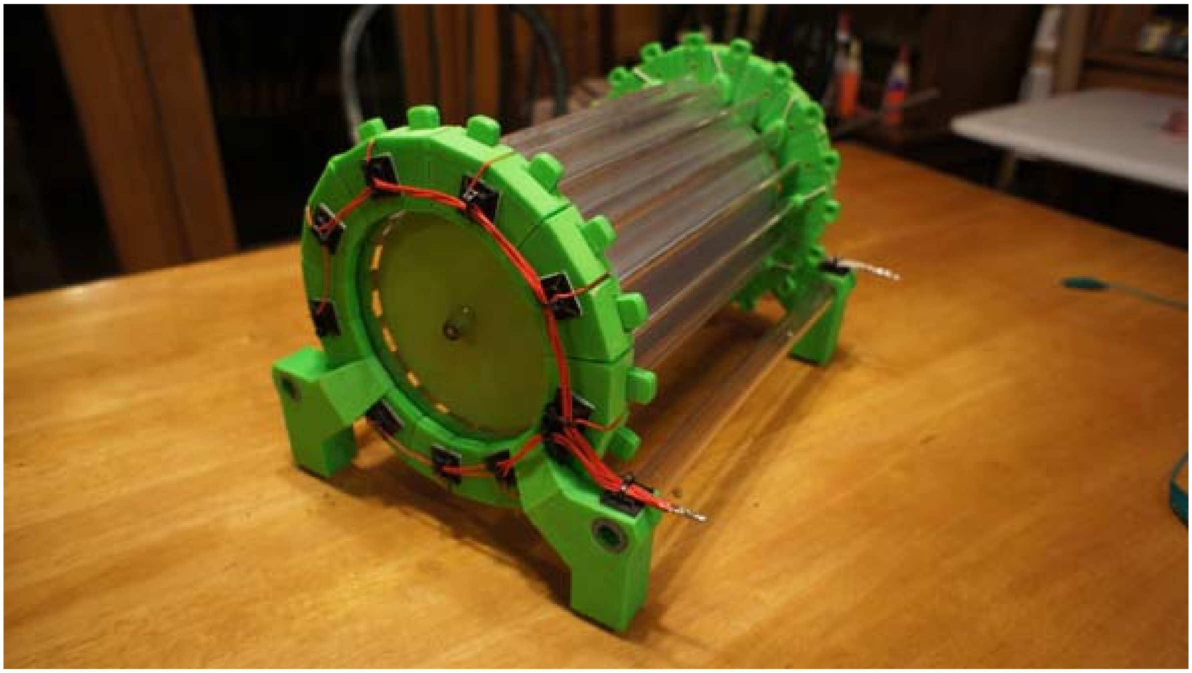

AtmoMotor HV Atmospheric Motor

by lasersaber | Dec 24, 2013 | Atmospheric Electricity, Motors

A high torque atmospheric motor. Tests underway…

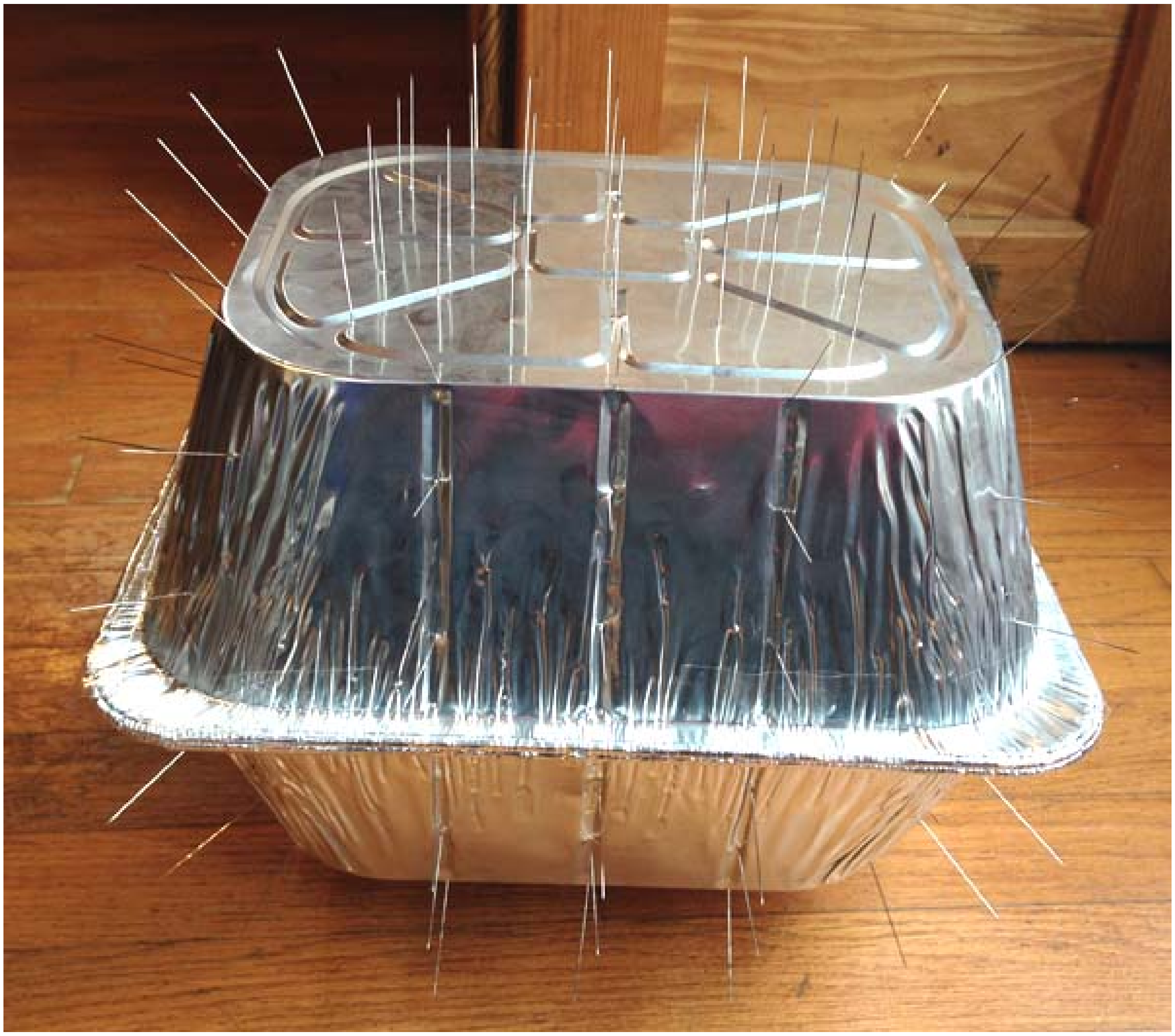

Atmospheric Electricity Antenna Testing

by lasersaber | Dec 4, 2013 | Atmospheric Electricity

Testing to determine the best type of antenna for my experiments.

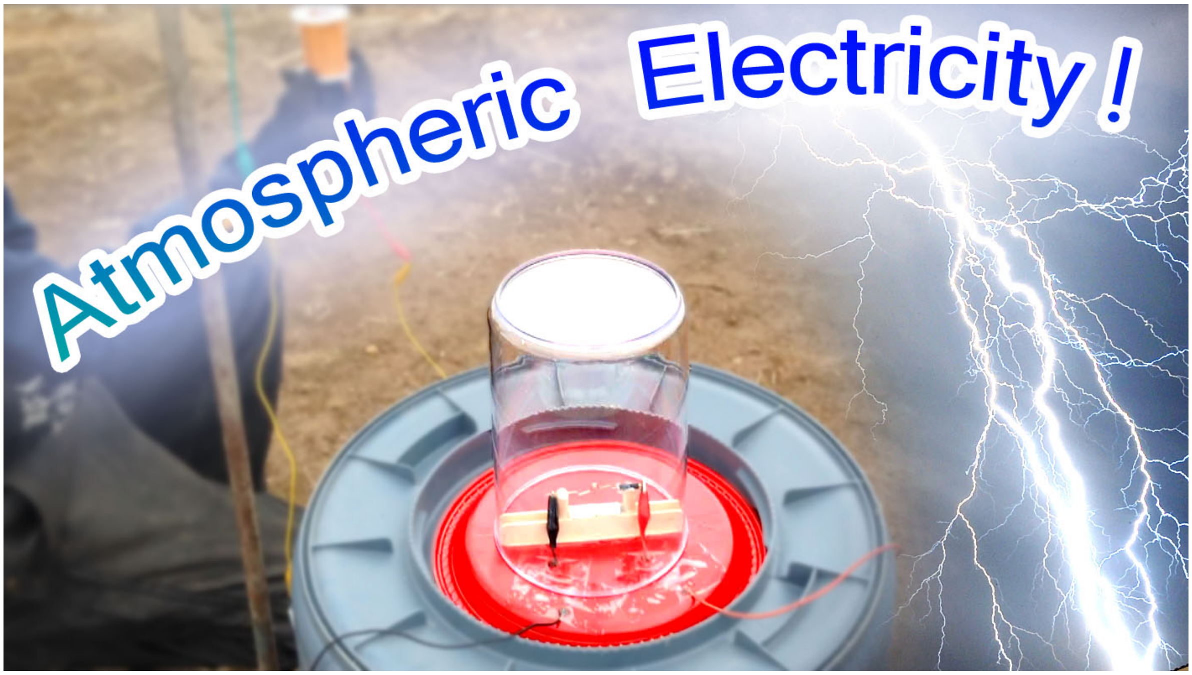

Free Atmospheric Electricity Powers Small Motor – Tesla Radiant Energy

by lasersaber | Nov 26, 2013 | Atmospheric Electricity, Featured Projects

I have always wanted to test a radiant energy motor. This little motor is the beginning of my research in that direction. In the near future I will be making larger more powerful motors. The precision required will probably require my 3D printer.

–

amazon

Tiny House Engineers...

$14.99 Prime

Shop now

–

rexresearch.com

Atmospheric & RF Electricity Collectors

GB 157263

Process and Apparatus for Converting Static Atmospheric Electrical Energy into Dynamic Electrical Energy of any Suitable High Periodicity

Abstract -- Atmospheric electricity, utilizing; influence machines.-Static atmospheric charges are converted into alternating-currents of any desired frequency by rotary machines, the stator and rotor elements of which form condensers having a capacity which varies during the revolution as described in Specincation 157,262. In the simplest form, Fig. 2, the curved stator plates 2, 1 are connected to earth E2 and the aerial A which collects the atmospheric charges and is earthed through a safety spark gap F. The rotor is motor-driven and consists of concentric plates 3, 4 joined through slip-,rings (not shown) to an inductance 9 shunted by a condenser 5. With a given aerial polarity, rotation of the plates 3, 4 causes current reversals on the circuit A, 1, 9, 2, E@, whereby oscillations are set upon the circuit 9, 5, these being tapped by leads 11, 12. The condenser 5 may consist of concentric tubes rotating with the rotor or of two juxtaposed wound spirals having capacity. The stator and rotor condenser plates may extend over nearly a semicircle in place of the quadrants shown in Fig. 2 and the rotor may comprise two consecutive cylindrical plates or coil condensers each divided into halves, the inner half of each being connected to the outer half of the other. Fig. 7 shows a form in which rings 1, 2, mounted in a casing comprising upper and lower parts insulated from each other, are connected to the stator condenser plates at intervals. The rotor plates 3, in equal number, are connected to slip-rings 5, 6 from which alternating-current is taken by brushes when the rotor is motor-driven. The rings 1, 2 are respectively connected to the aerial collector and earth E2. Several methods of grouping the stator-rotor condenser pairs and the components of the oscillating circuit are described. The oscillatory current in the rotor circuit may react through a transformer on the main charging current and two pairs of stator and rotor plates 1, 2, 3, 4, Fig. 13, may be employed arranged on diameters at right-angles, the oscillatory circuit in this case consisting of a mutual inductance 5 having capacity and a portion of an adjustable inductance 9 in the stator earth connector E2. By suitably proportioning the stator and rotor plates, 'resonance may be established, whereby undamped waves may be produced. In the form shown in Fig. 15, for high-frequency oscillations, the aerial A and earth wire are tapped at points between the stator plates 1, 2 and condensers 16, 18 which are joined through an inductance 9. The rotor current is conveyed to the power circuit 11, 12 through transformers 10, 10<a>. The members composing the stator and rotor plates may be slotted, as shown in Fig. 8, with holes 1 for attaching them to the frame, or they may have spiral slots in their surfaces. Alternatively, they may be formed by embedding spiral wires 3 in a vulcanite mass 1 so as to give a smooth surface as shown in Fig. 11.

Fig. 1

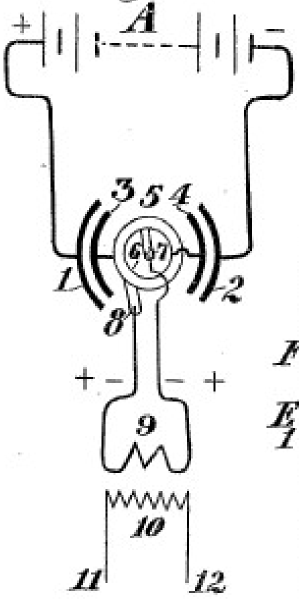
A
+
−
3 5 4
6 7
1
2
8
+ − − +
9
10
11
12

Fig. 3

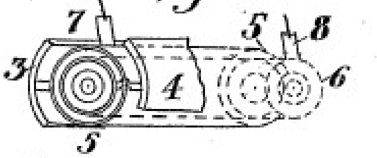

Fig. 2

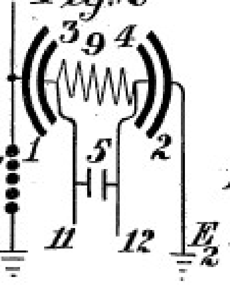
3 9 4
1
2
5
F
E1
11
12
E2

Fig 4

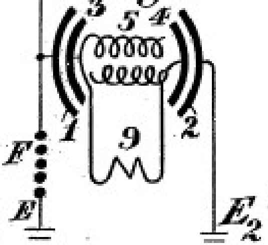
3 5 4
1
9
2
F
E
E2

Fig. 5.

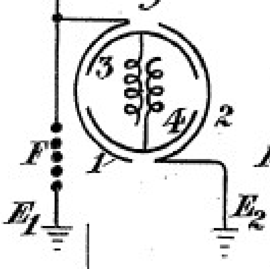

Fig. 6.

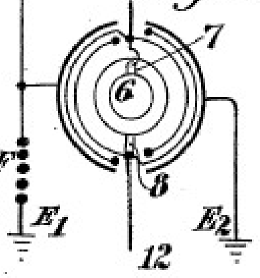
7
6
8
F
E1
12
E2

Fig. 7.

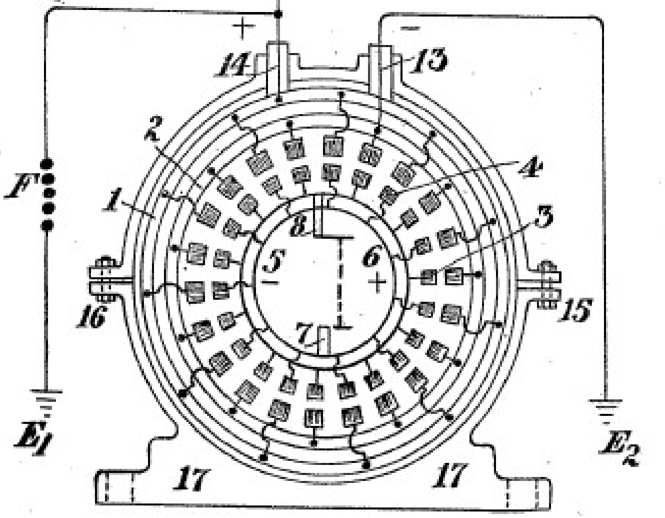
+
−
14
13
2
4
1
F
8
3
5 −
6 +
16
15
7
E1
E2
17
17

Fig.8

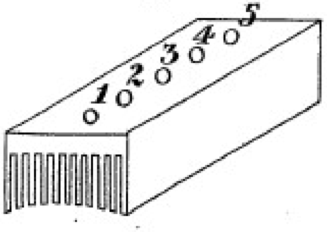

Fig.10

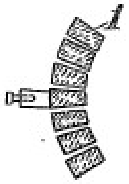

Fig.9.

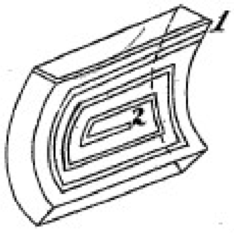

Fig.11

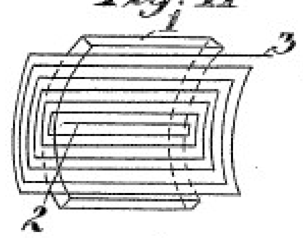

Fig.12

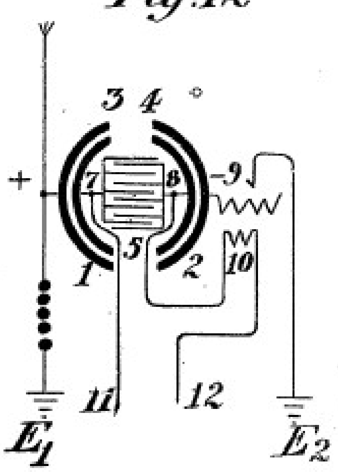
+
3
4
7
8
9
1
5
2
10
11
12
E_1
E_2

Fig.13.

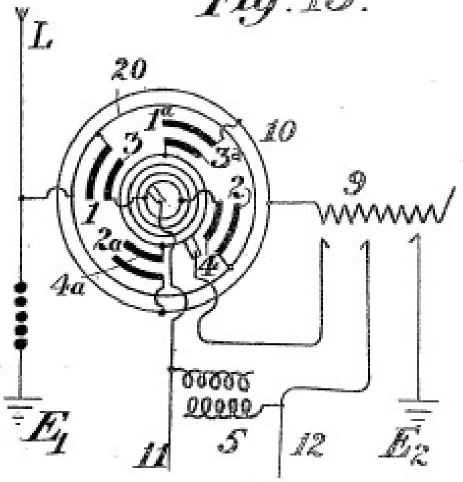
L
20
1^a
3
3^a
10
2
9
1
2a
4
4a
5
11
12
E_1
E_2

Fig.14

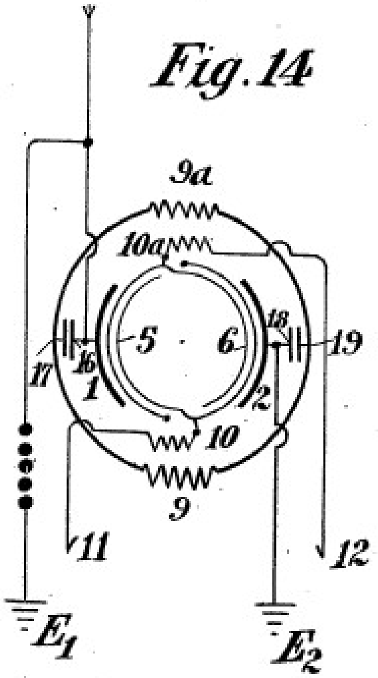
9a
10a
5
6
18
19
17
16
1
2
10
9
11
12
E1
E2

Fig.16

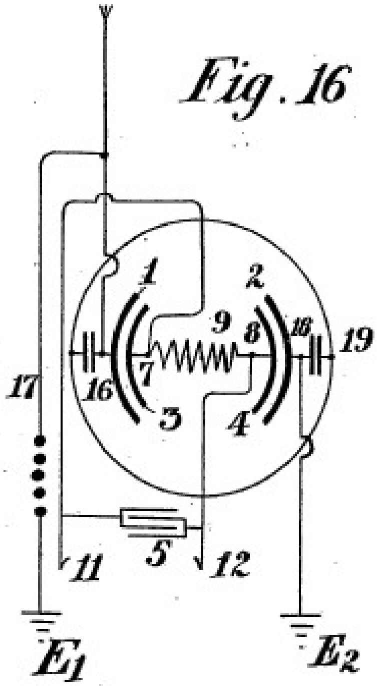
1
2
9
8
18
19
7
17
16
3
4
5
11
12
E1
E2

Fig.15

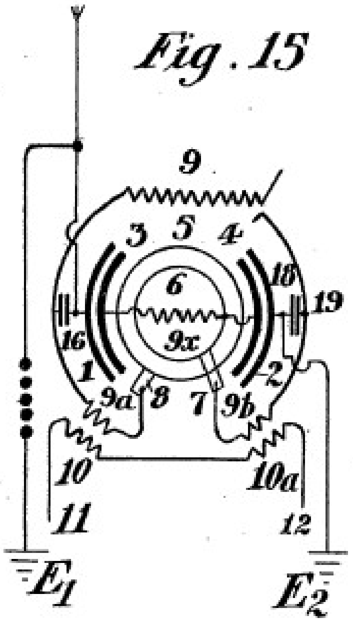
9
3
5
4
18
6
19
16
9x
1
2
9a
8
7
9b
10
10a
11
12
E1
E2

Fig.17.

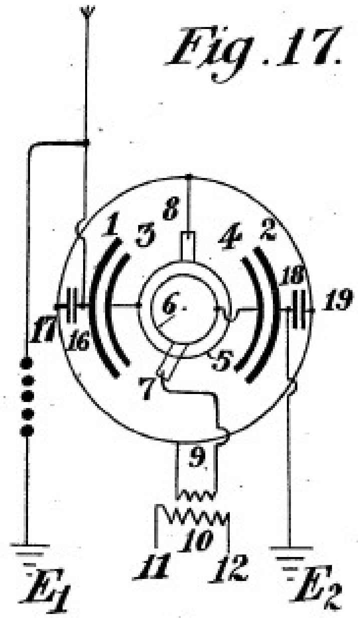
8
1
3
4
2
18
19
17
16
6
5
7
9
10
11
12
E1
E2

French Patent # 551,882

Apparatus for Capture of Electric Currents in the Atmosphere

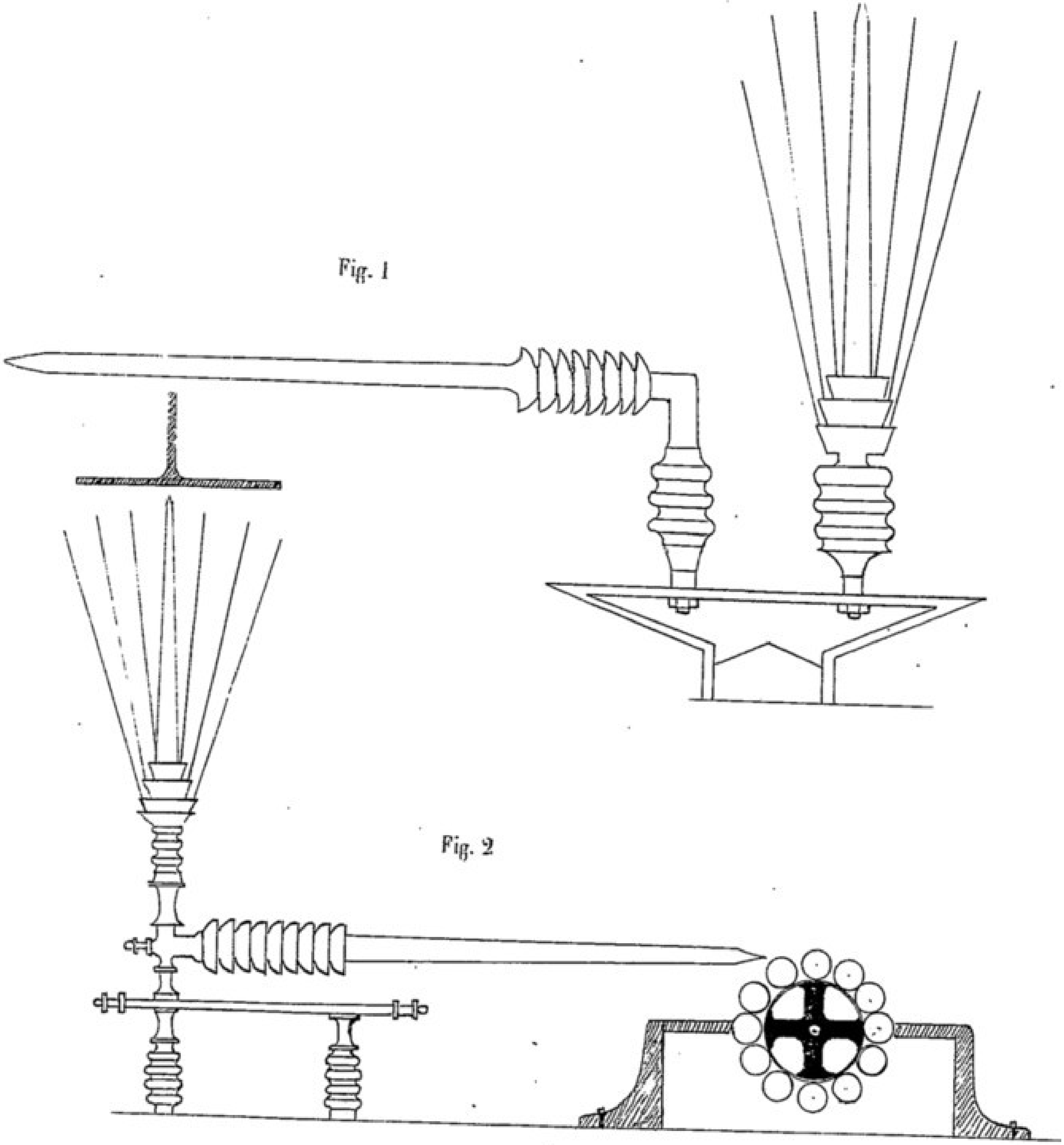

Fig. 1

Fig. 2

French Patent # 565,395

Combined Apparatus for Capture of Atmospheric Electric Currents with Immediate Implementation

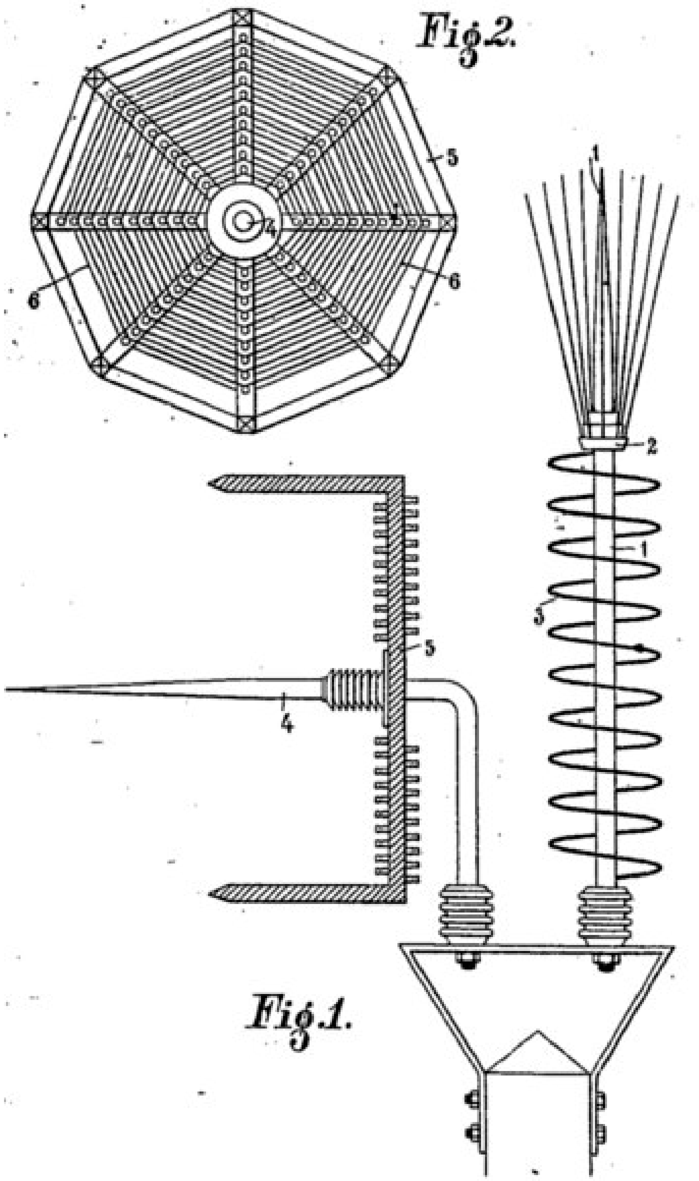

Fig. 2.

Fig. 1.

Fig. 3.

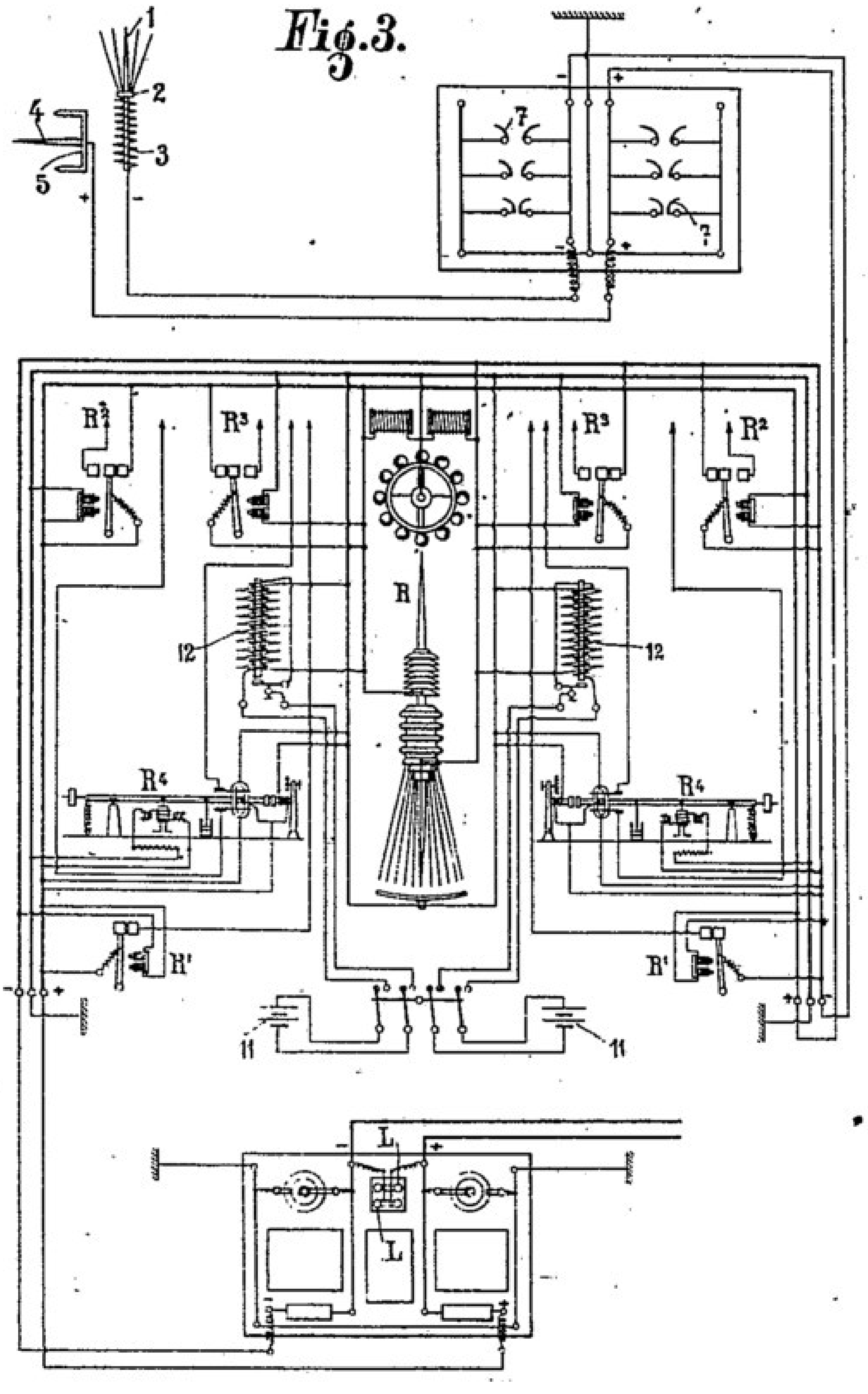

1
2
3
4
5
+
-
7
7
R²
R³
R³
R²
R
12
12
R⁴
R⁴
R¹
R¹
11
11
L
L

Fig. 4.

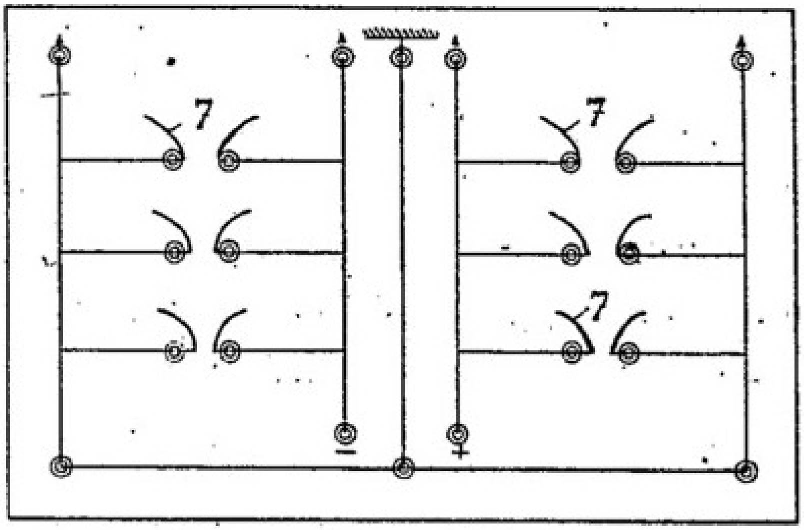

Fig. 5.

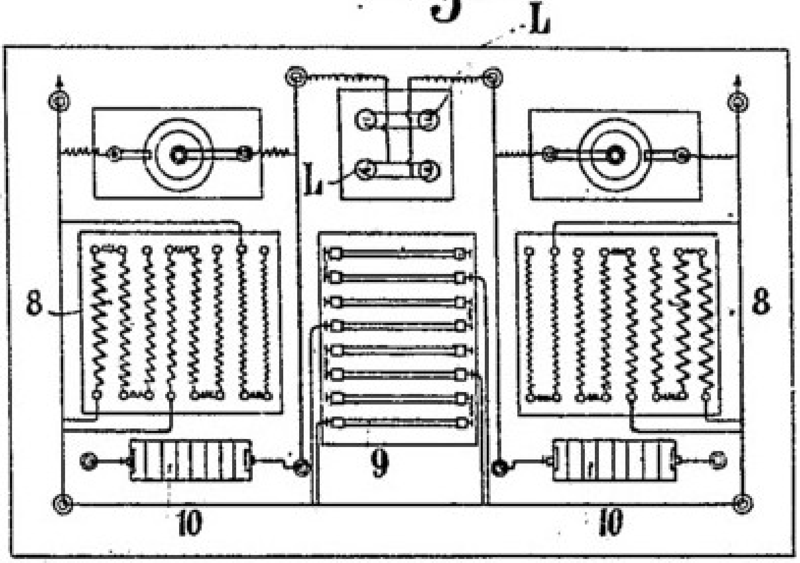

US 28793
Electric Apparatus

Abstract -- An aerostat for obtaining atmpspheric and terrestrial electricity...

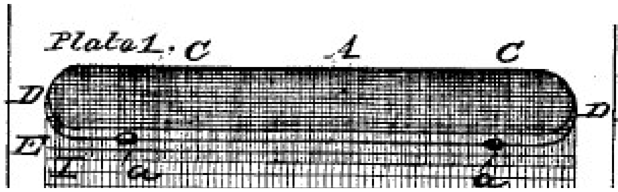
Plate 1.
C
A
C
D
D
E
a
a

Plate 2
A
B
D
B
E
A
D
B

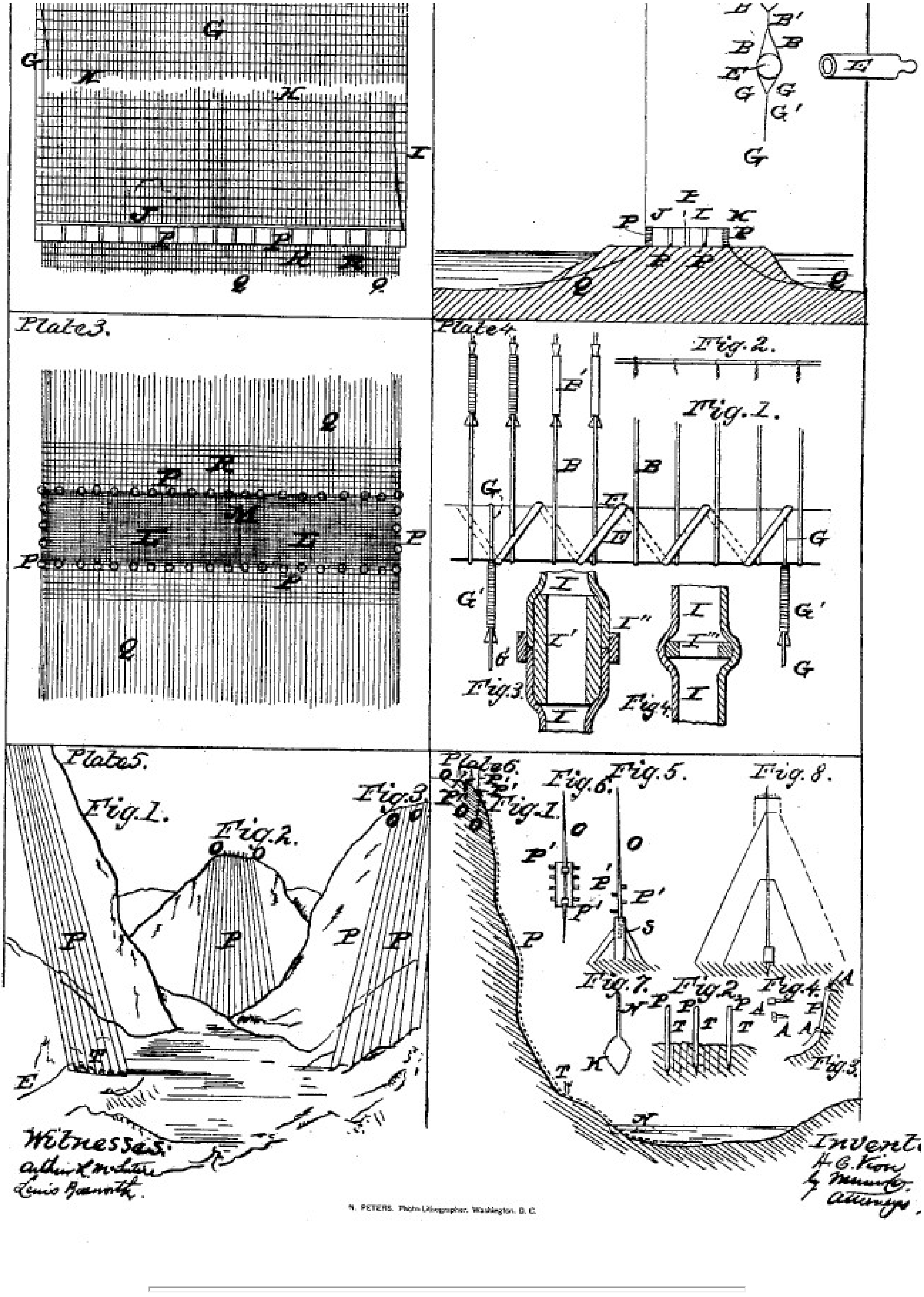
G
K
I
J
P
R
Q
B
B'
E
G'
F
L
Plate 3.
M
Plate 4
Fig. 2.
Fig. 1.
E'
D
I''
I'
I'''
Fig. 3.
Fig. 4.
Plate 5.
Fig. 1.
Fig. 2.
Fig. 3.
Plate 6.
Fig. 6.
Fig. 5.
Fig. 8.
Fig. 7.
Fig. 2.
Fig. 4.
Fig. 3.
O
P'
S
N
T
A
K
Witnesses:
Inventor:
H. G. Vion
by Munn & Co.
Attorneys.
N. PETERS, Photo-Lithographer, Washington, D. C.

US 126356

Improvement in Collecting Energy for Telegraph

Abstract -- A tower for the purpose of receiving and imparting natural electricity, so as to be in constant contact with the upper stratum of electricity which surrounds the earth, by tapping which a never-failing supply is formed when brought into contact with the earth.

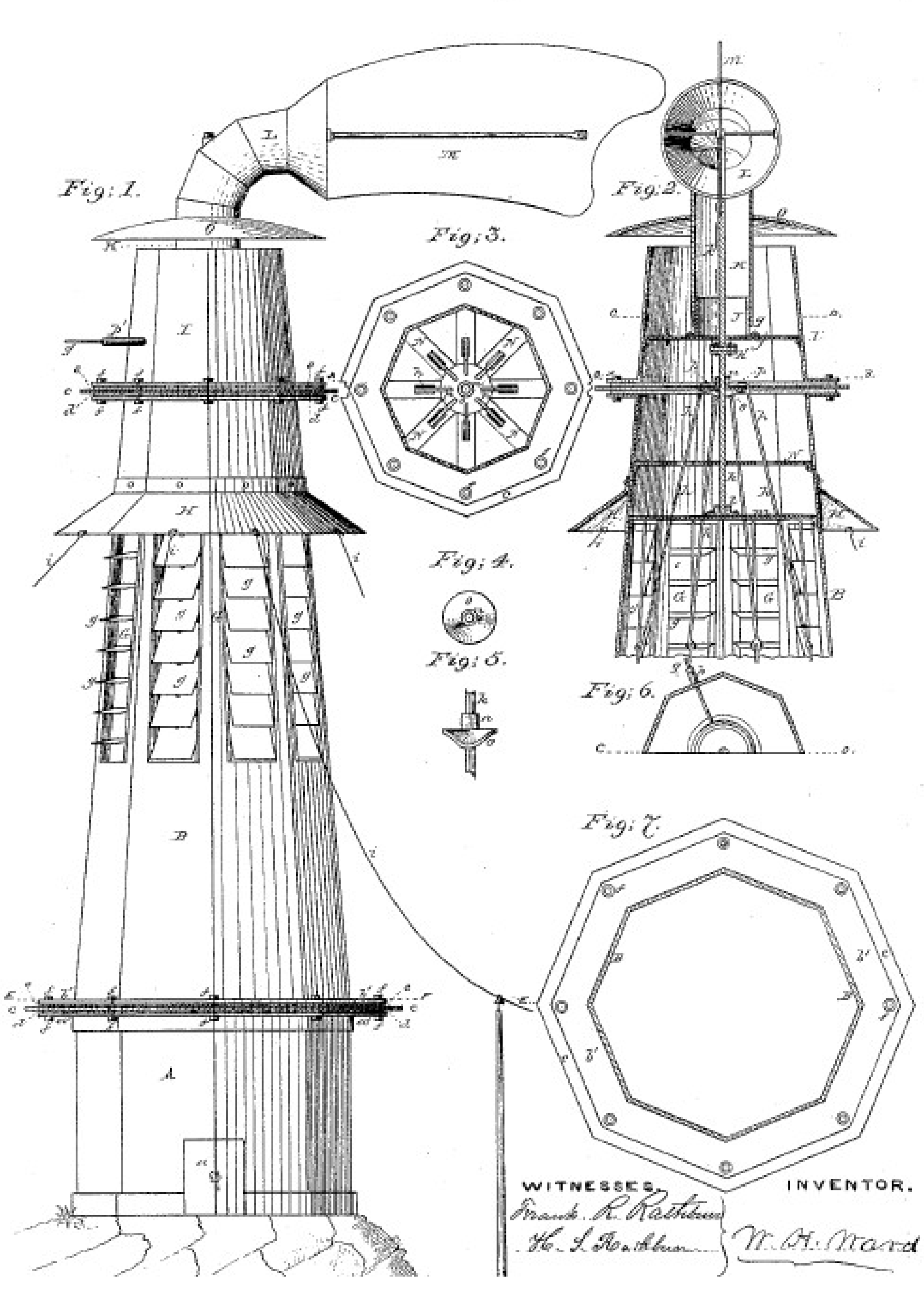
Fig: 1.
Fig: 2.
Fig: 3.
Fig: 4.
Fig: 5.
Fig: 6.
Fig: 7.
WITNESSES.
INVENTOR.
Frank R. Rathbun
W. H. Ward

US 129971
Telegraph Power

Abstract -- Utilizing natural electricity and establishing an electrical current without the aid of wires, batteries, or cables...

US 674427
Collecting Atmospheric Electricity

Abstract -- An appartus for colecting and driving or conducting atmospheric electricity which renders it practicable to obtain material quantities of teh same in serviceable form.

Fig:1.

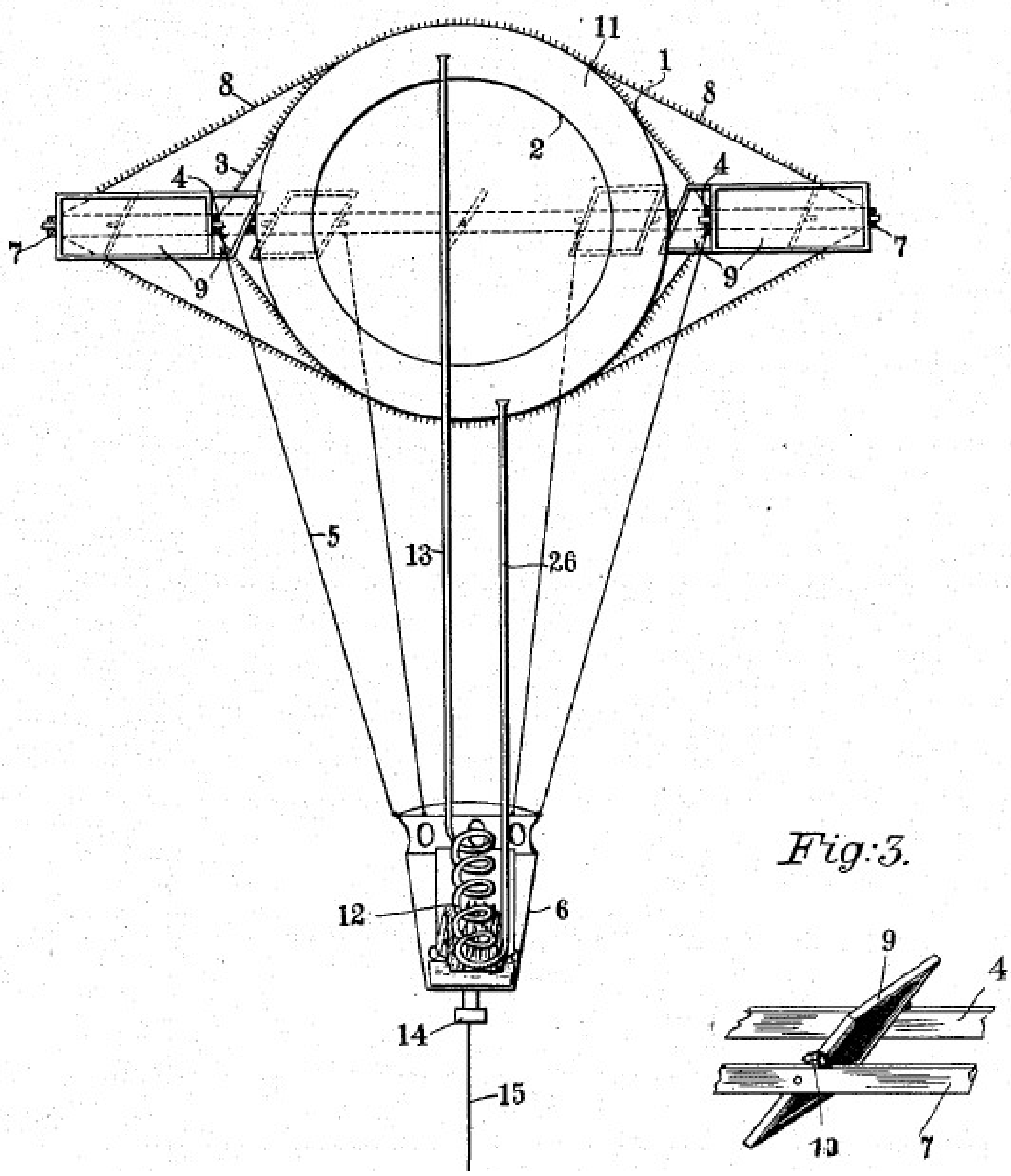

Fig:3.

Fig: 2.

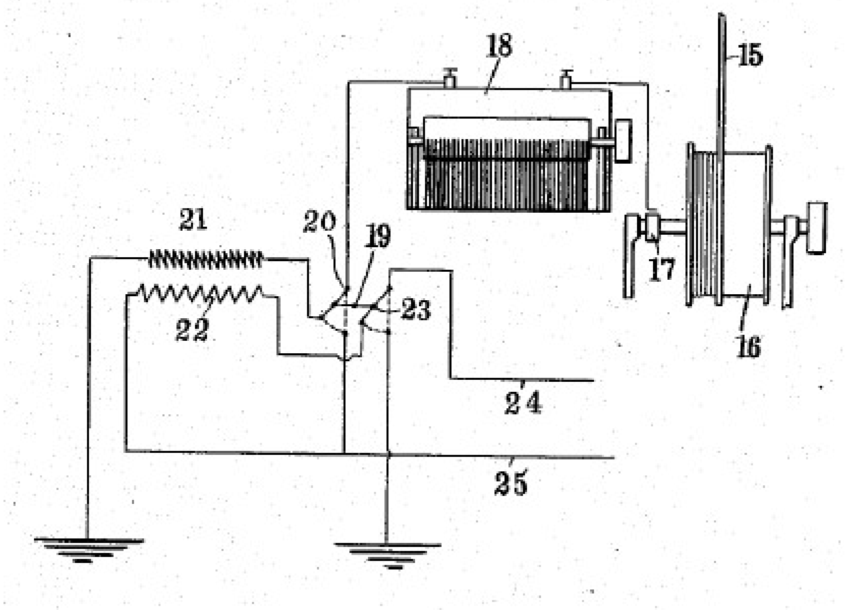
18
15
21
20
19
17
23
22
16
24
25

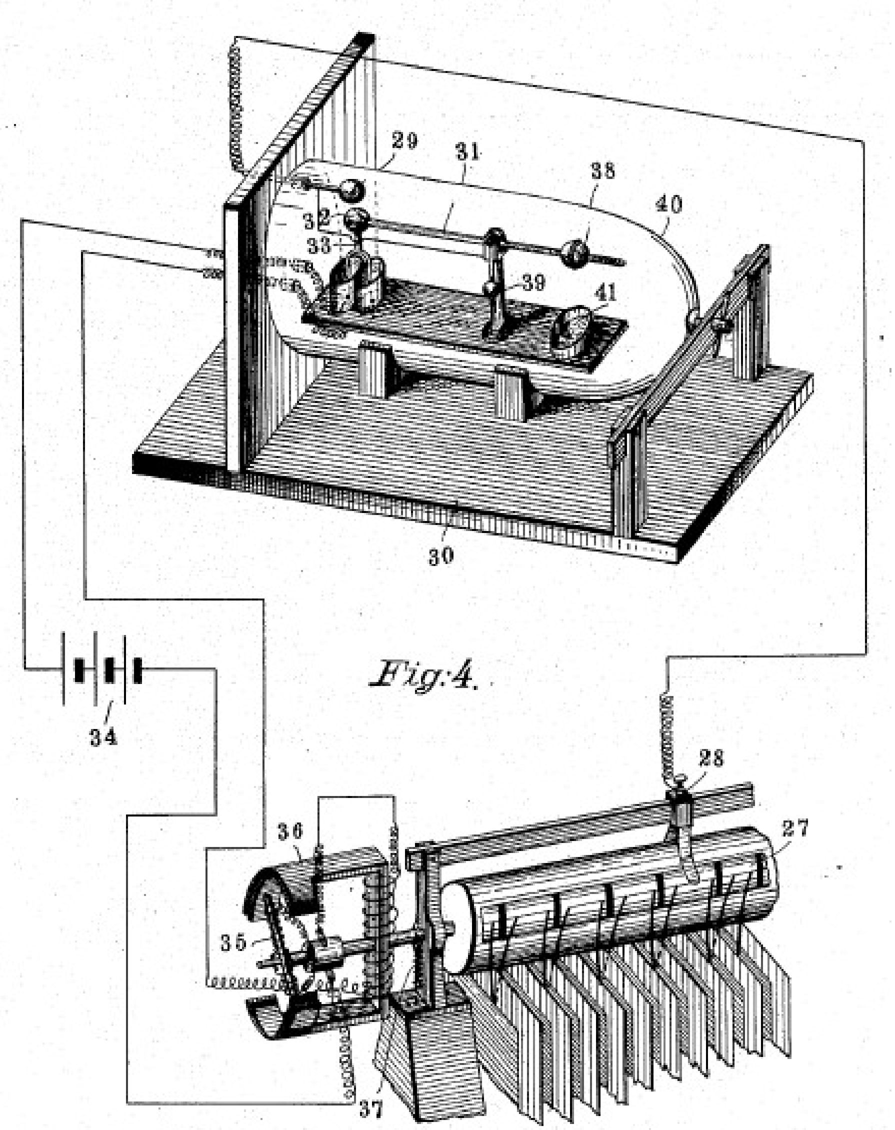

Fig:4.

US 911260
Collecting Atmospheric Electricity

Abstract -- A method of colecting electricity from a strata laden with electricity at high altitudes in the atmoshere, through the medium of a wire cable suspended from one or more balloons ...

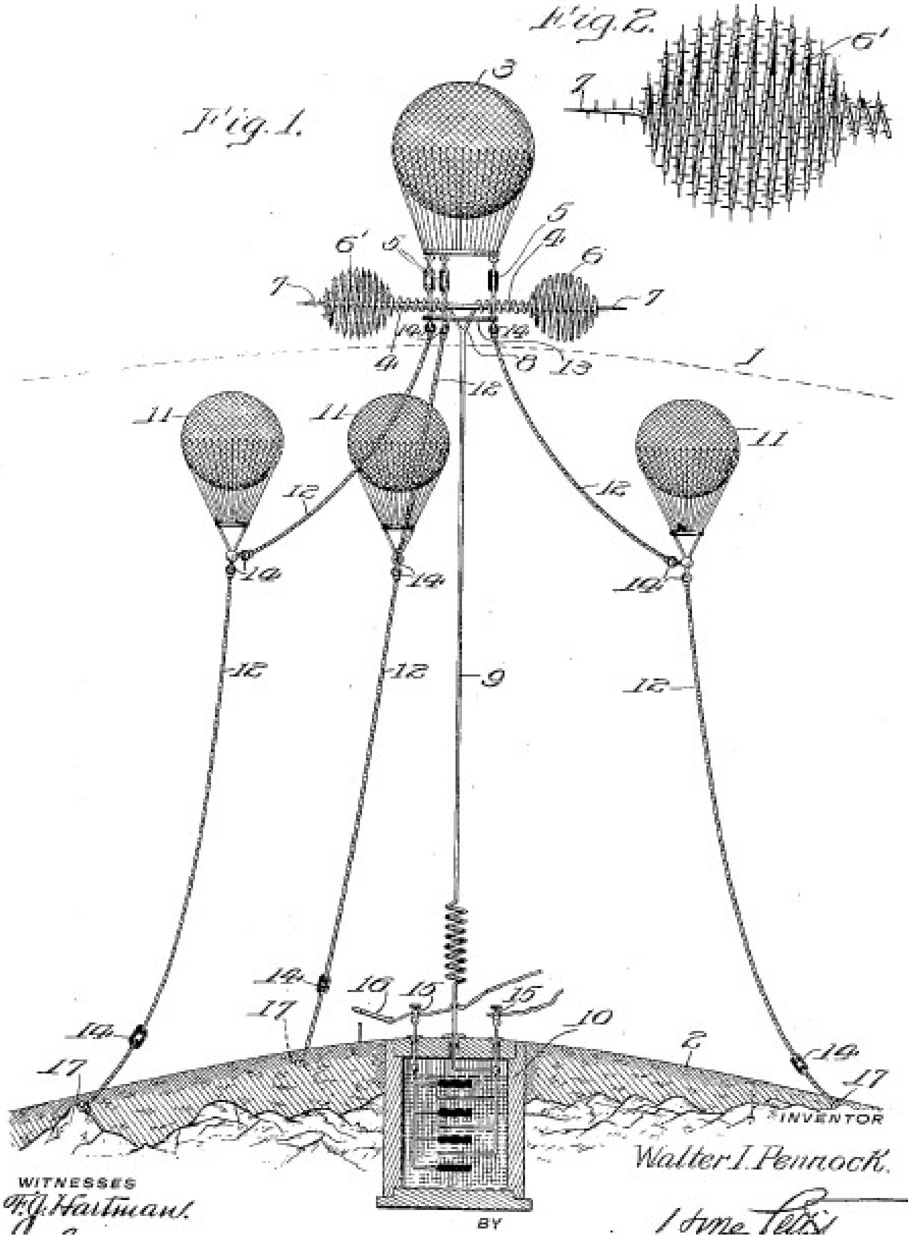
Fig. 2.
7
6'
Fig. 1.
3
5
6' 5
4
6
7
7
14 14
8 13
1
11
11
18
11
12
12
14
14
14
12
12
9
12
14
16 15
15
17
10
2
14
17
17
INVENTOR
Walter I. Pennock.
WITNESSES
F.J. Hartman
BY

US 1014719

Collecting Atmospheric Electricity

Abstract -- A method of colecting electricity from a strata laden with electricity at high altitudes in the atmoshere, through the medium of a wire cable suspended from one or more balloons ...

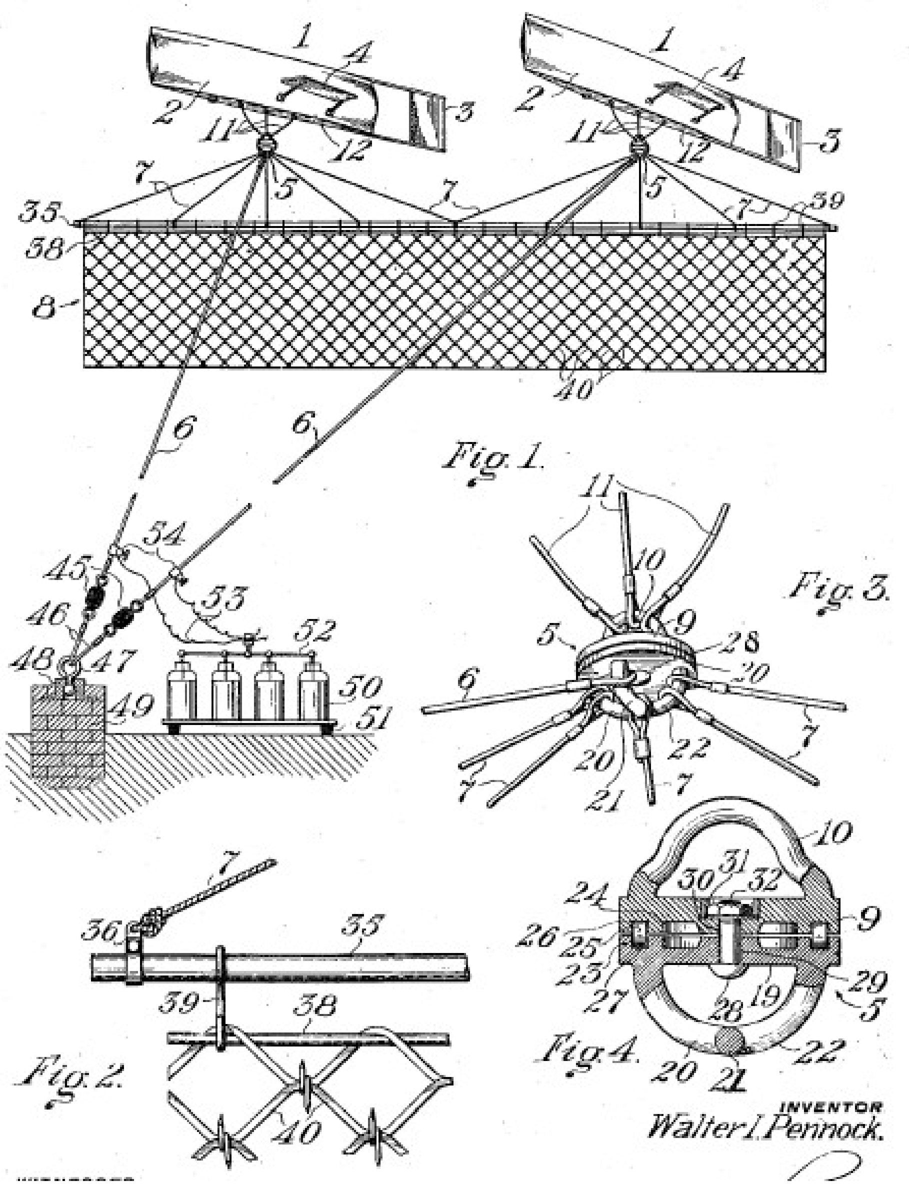
1
4
2
11
12
3
1
4
2
11
12
3
7
5
7
5
7
35
39
38
8
40
6
6
54
45
33
52
46
48
47
50
49
51
11
10
9
5
28
20
6
7
22
20
21
7
7
7
36
35
39
38
40
10
30
31
32
24
9
26
25
23
29
27
19
5
28
22
20
21

Fig. 1.

Fig. 3.

Fig. 2.

Fig. 4.

INVENTOR
Walter I. Pennock.

Fig. 5.

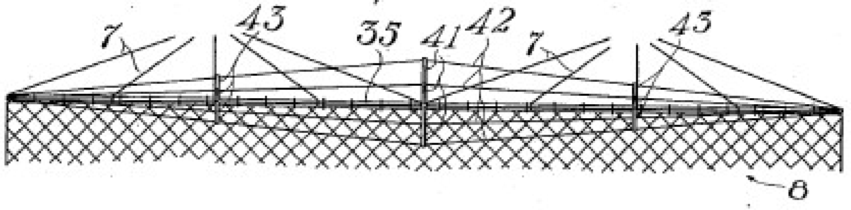
7
43
35
41
42
7
43
8

Fig. 6.

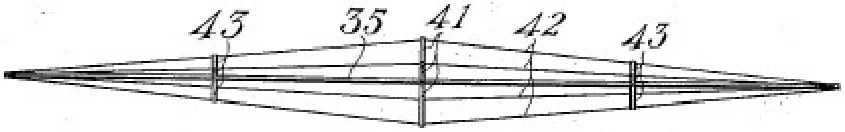
43
35
41
42
43

Fig. 7.

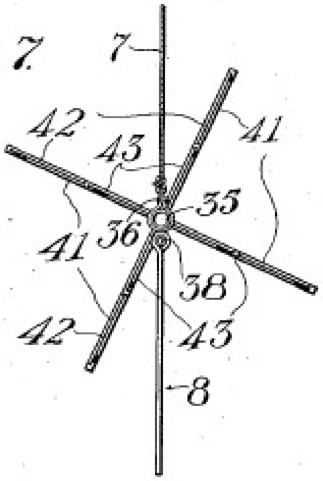
7
42
41
43
35
41
36
38
42
43
8

INVENTOR
Walter I. Pennock.

US 1540998
Conversion of Atmospheric Electric Energy

Abstract -- Method of obtaining atmospheric electricity by means of metallic ballon collectors...

Filed Jan. 15, 1921 12 Sheets-She

Fig. 1.

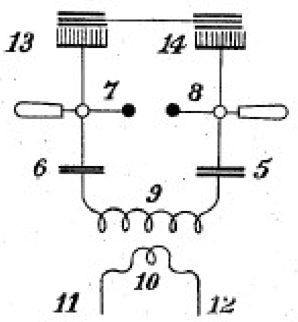
13
14
7
8
6
5
9
10
11
12

Fig. 2.

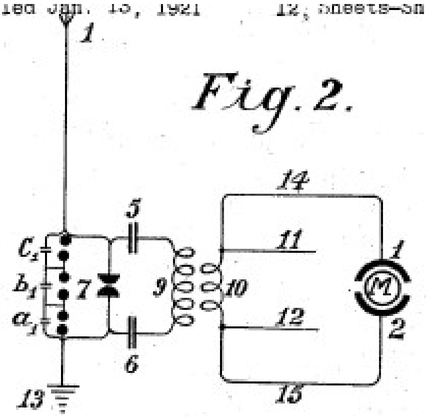
1
14
5
C1
b1
a1
7
9
10
11
12
1
M
2
6
13
15

Fig. 3.

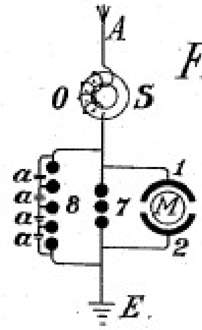
A
O
S
a
a
a
a
8
7
1
M
2
E.

Fig. 4.

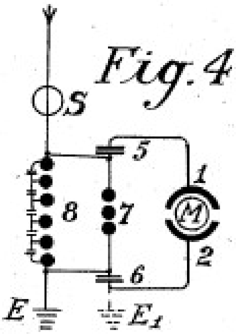
S
5
8
7
1
M
2
6
E
E1

Fig. 5.

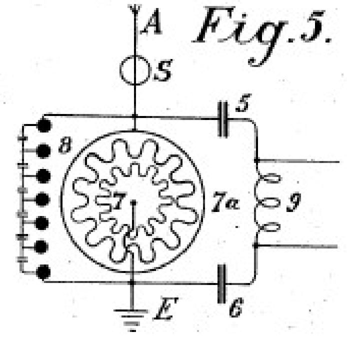
A
S
5
8
7
7a
9
E
6

Fig. 6

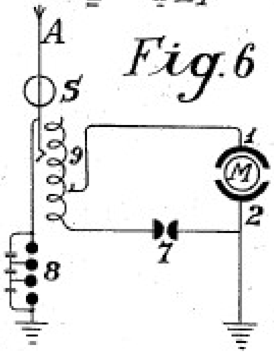

Filed Jan. 13, 1921 12 Sheets-Sheet 2

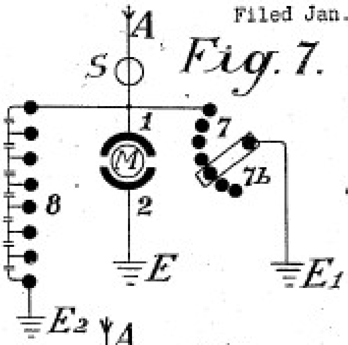
A
S
1
M
2
7
7b
8
E
E1
E2

Fig. 7.

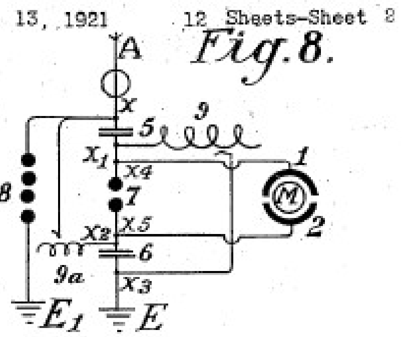
A
X
5
9
X1
X4
1
M
7
8
X5
X2
6
9a
X3
2
E1
E

Fig. 8.

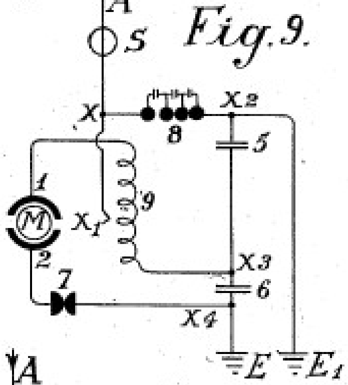
A
S
X
8
X2
5
1
M
X1
9
2
7
X3
6
X4
E
E1

Fig. 9.

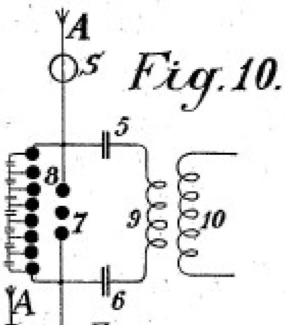
A
S
5
8
7
9
10
6
E

Fig. 10.

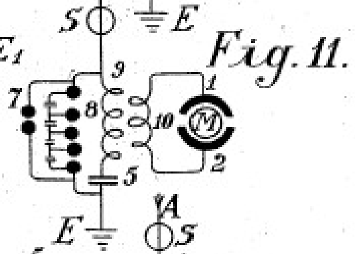
A
S
9
7
8
10
1
M
2
5
E

Fig. 11.

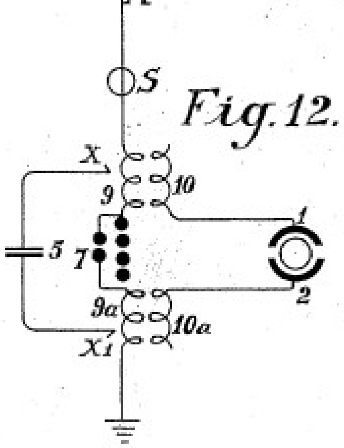
A
S
X
9
10
7
5
1
2
9a
10a
X1

Fig. 12.

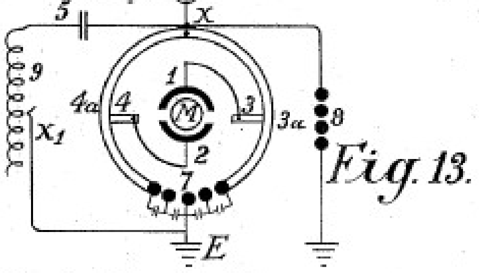
A
S
5
X
9
1
4a
4
M
3
3a
8
X1
2
7
E

Fig. 13.

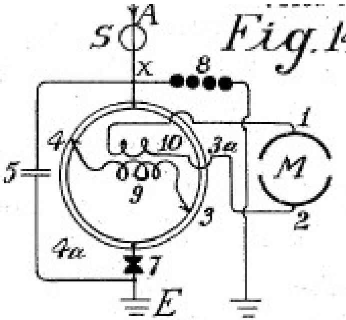
Fig. 14.
S
A
X
8
1
M
2
10
3a
4
5
9
3
4a
7
E

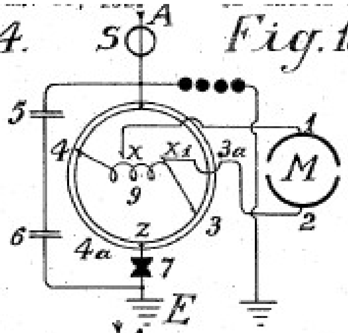
Fig. 15.
S
A
5
1
M
2
4
X
X1
3a
9
3
6
Z
4a
7
E

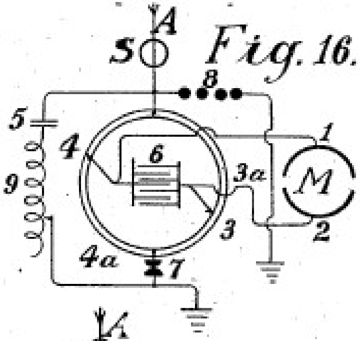
Fig. 16.
S
A
8
5
4
6
3a
1
M
2
9
3
4a
7

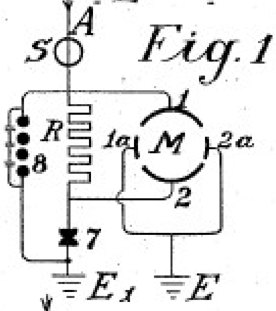
Fig. 17.
S
A
R
1
1a
M
2a
8
2
7
E1
E

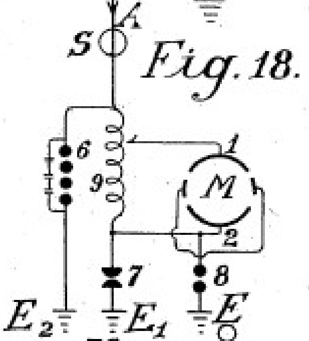
Fig. 18.
S
A
6
9
1
M
2
7
8
E2
E1
E

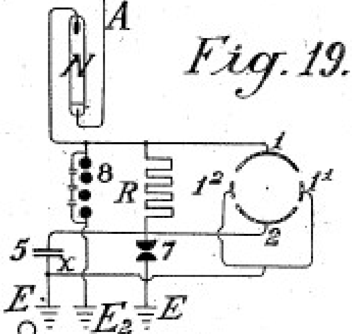
Fig. 19.
A
N
8
R
1
1²
1¹
2
5
X
7
E
E2
E

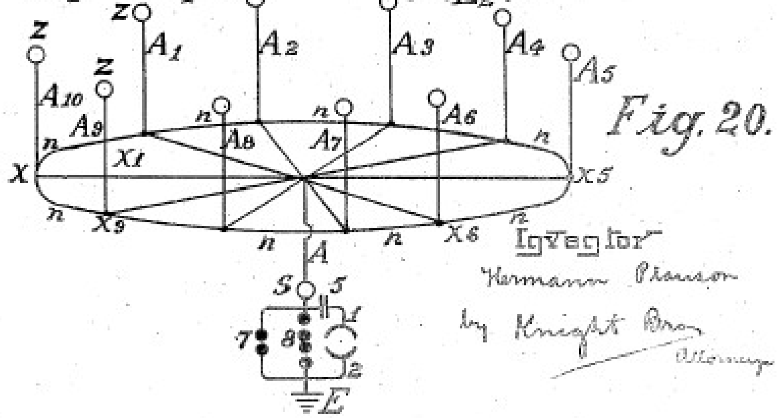
Fig. 20.
Z
Z
Z
A1
A2
A3
A4
A5
A10
A9
A8
A7
A6
n
X
X1
X5
X9
X6
A
S
5
1
7
8
2
E

Inventor
Hermann Plauson
by Knight Bros
Attorneys

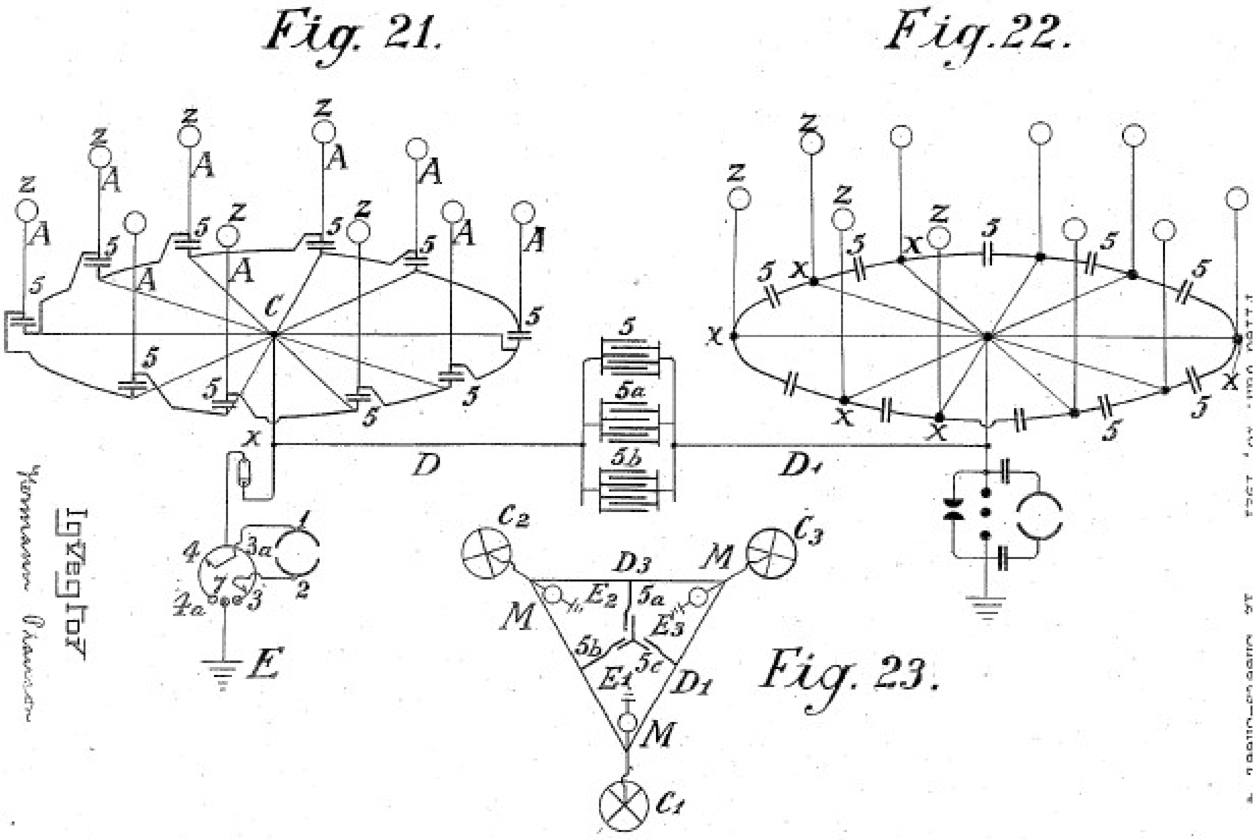
Z
A
5
C
X
D
D1
5a
5b
1
2
3
3a
4
4a
7
E
C2
C3
C1
M
D3
E2
E3
E1
5c

Fig. 21.

Fig. 22.

Fig. 23.

Inventor

Hermann Plauson

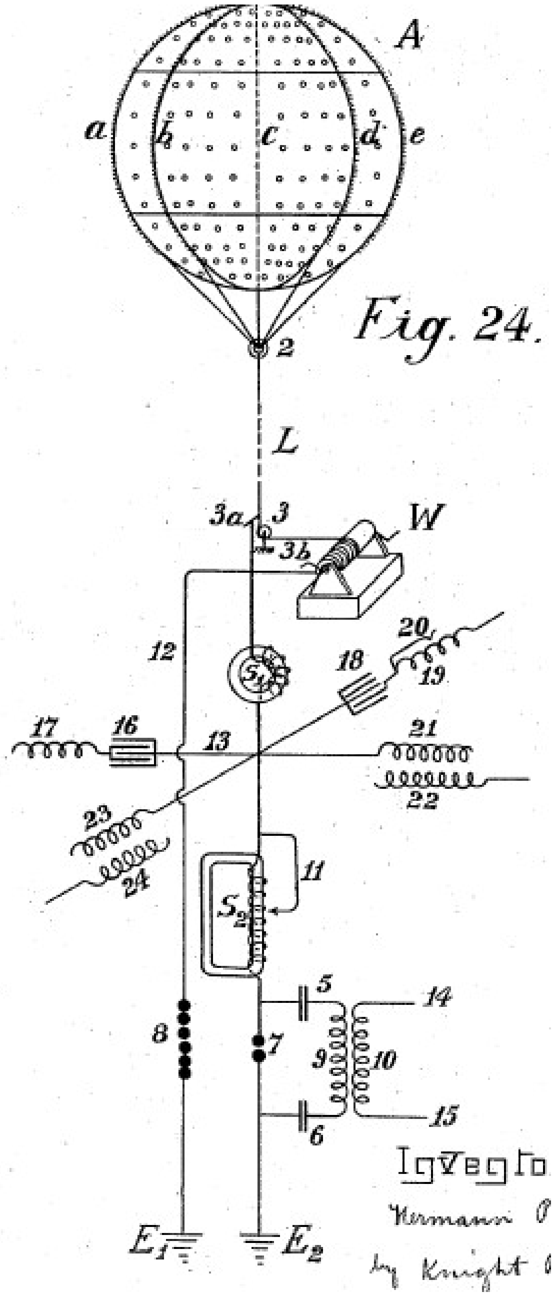

Fig. 24.

A
a
b
c
d
e
2
L
3a
3
3b
W
12
20
19
18
17
16
13
21
22
23
24
11
S1
S2
5
14
8
7
9
10
15
6
E1
E2
Invento
Hermann P
by Knight

Fig. 25.

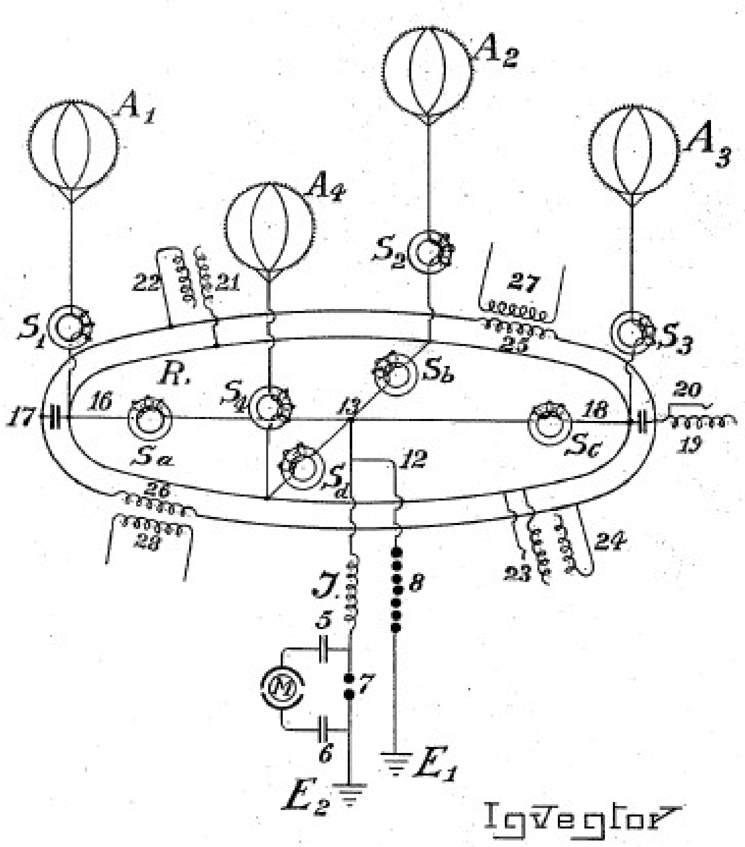
A1
A2
A3
A4
S1
S2
S3
S4
Sa
Sb
Sc
Sd
R.
12
13
16
17
18
19
20
21
22
23
24
25
26
27
28
J.
5
6
7
8
M
E1
E2
Inventor

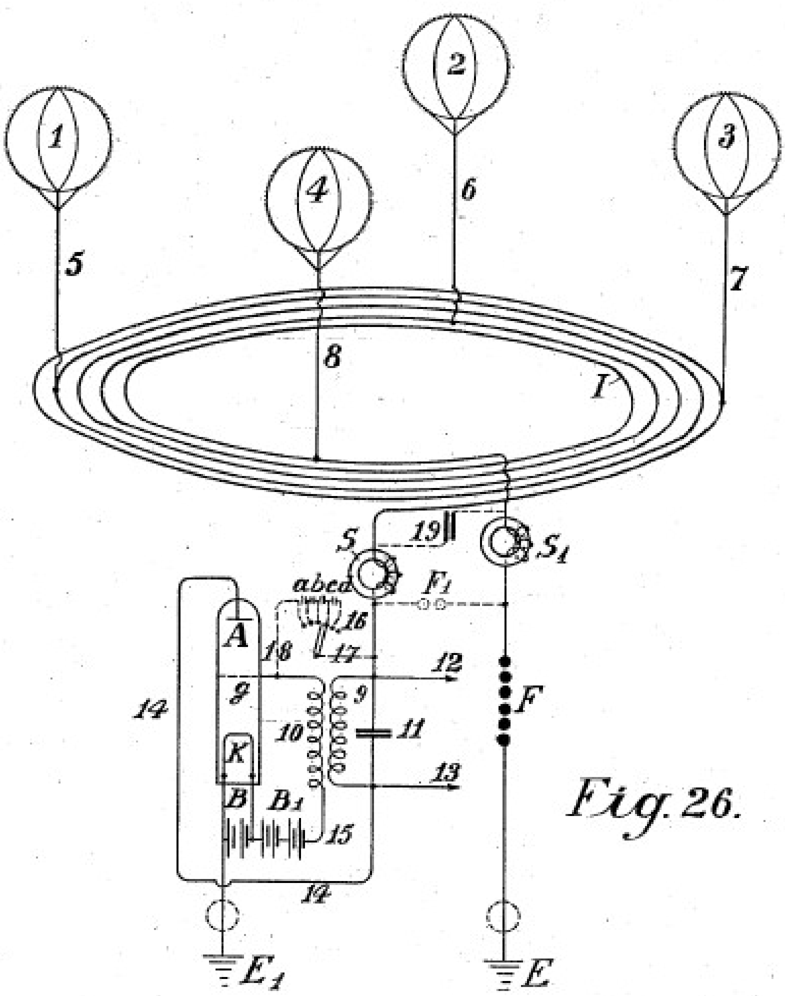
I
S
S1
abcd
F1
A
K
B
B1
F
E1
E

Fig. 26.

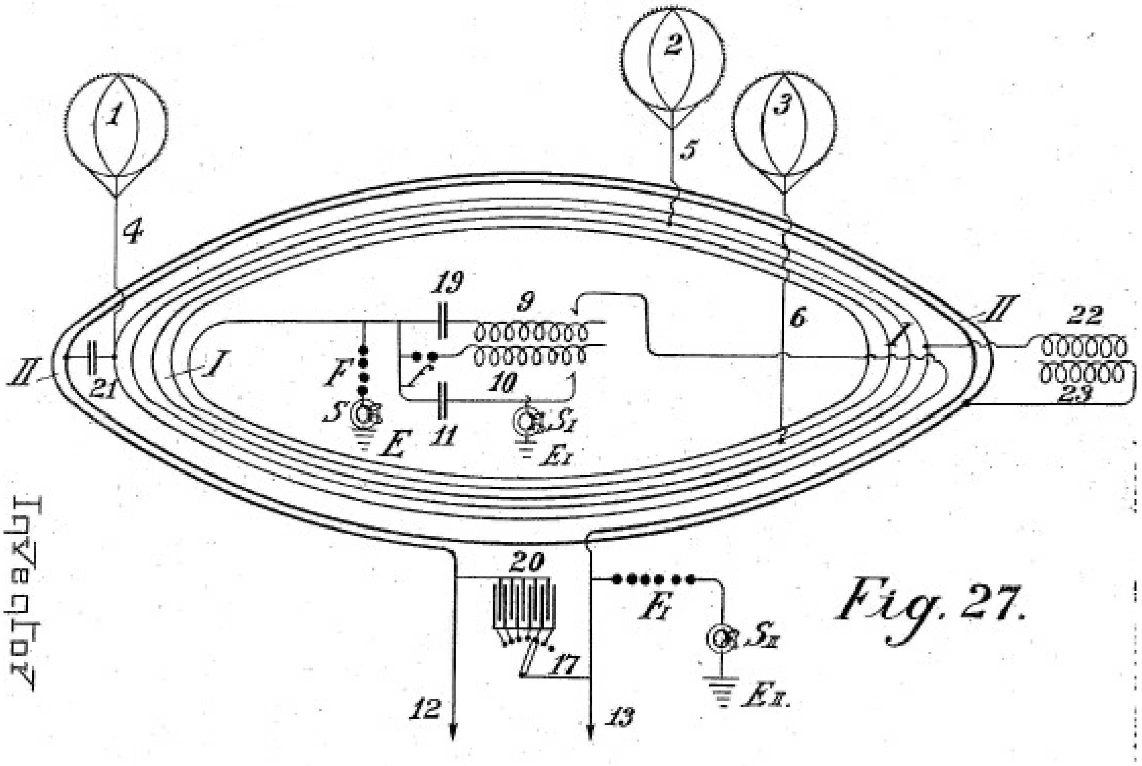
1
2
3
4
5
6
II
21
I
19
9
F
f
10
11
S
E
S1
E1
II
22
23
20
F1
SII
17
EII.
12
13

Fig. 27.

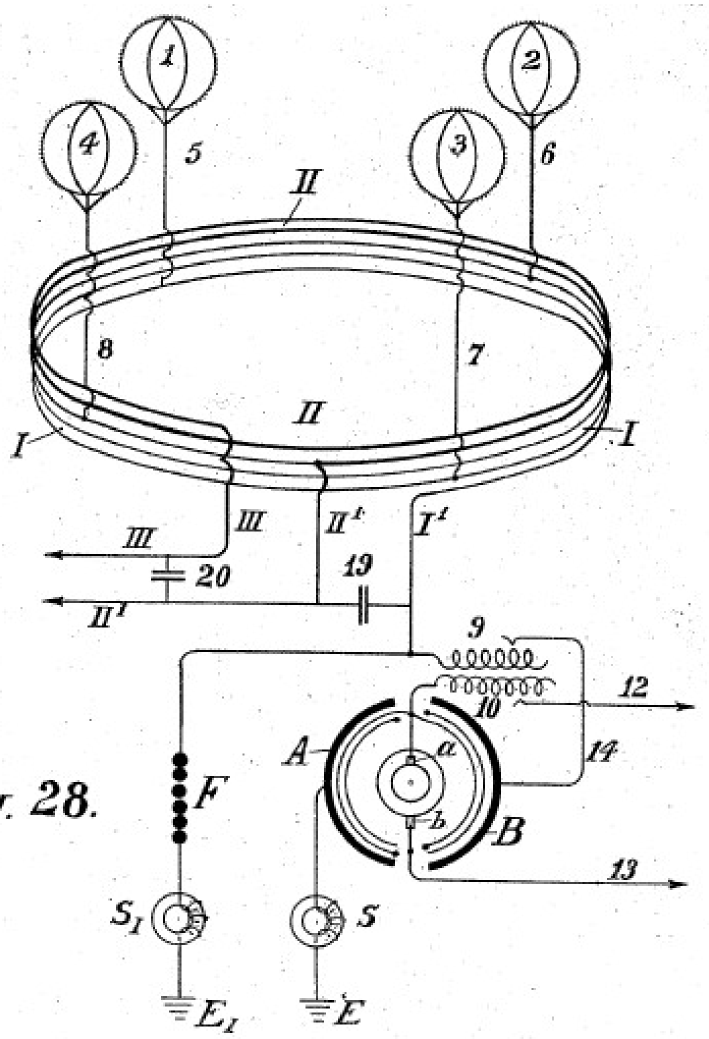
1
2
4
5
3
6
II
8
7
II
I
I
III
II¹
I¹
III
20
19
II¹
9
12
10
A
a
14
F
B
b
13
S₁
S
E₁
E

Fig. 28.

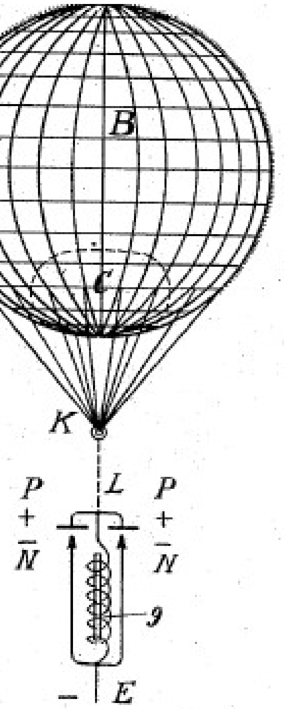
B
C
K
L
P
+
N
P
+
N
9
–
E

Fig. 29.

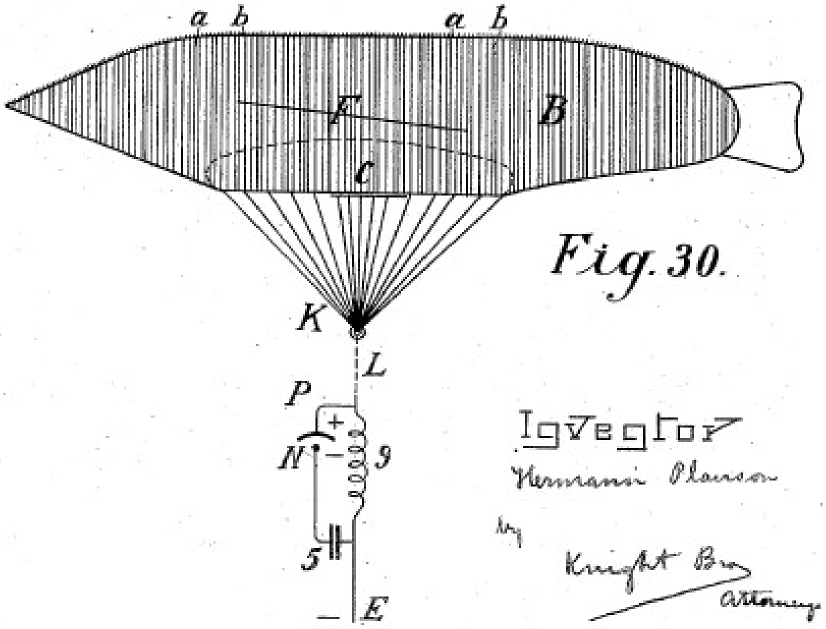
a
b
a
b
F
B
C
K
L
P
+
N
–
9
5
–
E

Fig. 30.

Inventor
Hermann Plauson
by
Knight Bros
Attorneys

Filed Jan. 13, 1921 12 Sheets-Sheet 11

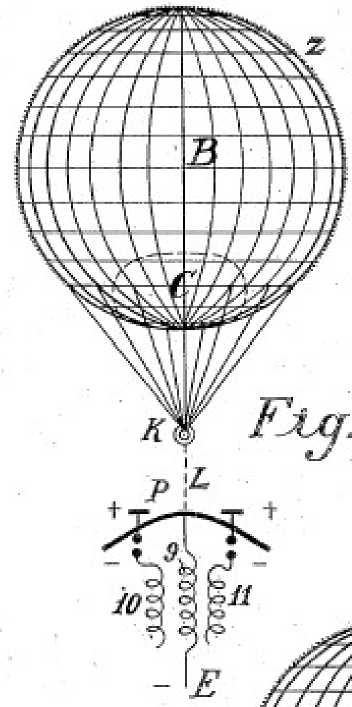
z
B
C
K
L
P
+
+
-
-
9
10
11
- E

Fig. 31.

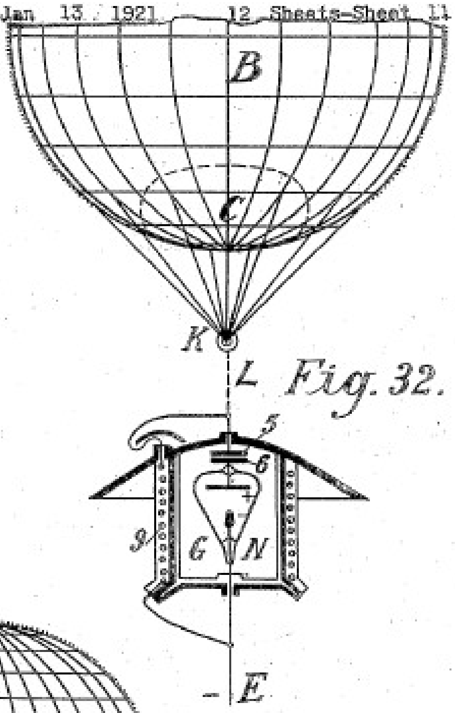
z
B
C
K
L
5
6
9
G
N
- E

Fig. 32.

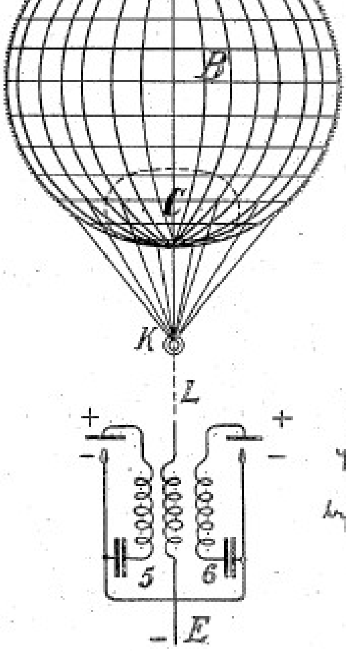

Fig. 33.

Inventor
Hermann Plauson
by Knight Bros
attorneys

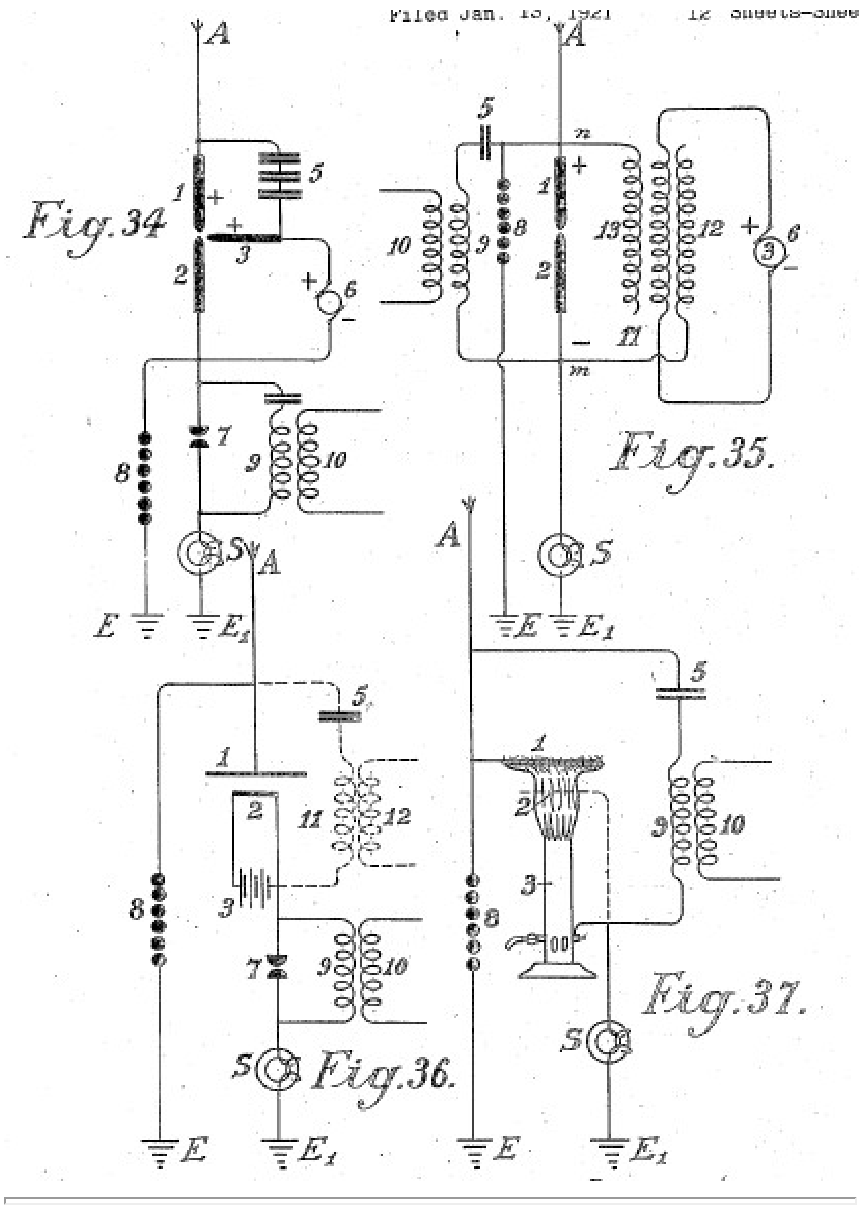
Filed Jan. 13, 1921 12 Sheets-Shee
A
1
+
2
3
5
6
7
8
9
10
S
E
E_1
n
m
11
12
13
−

Fig. 34

Fig. 35.

Fig. 36.

Fig. 37.

GB 149,917
Improvements in & Relating to Radio Communication Systems

Abstract -- Radio transmission and reception through the use of living vegetable organisms such as trees, plants, and the like.

Fig 1

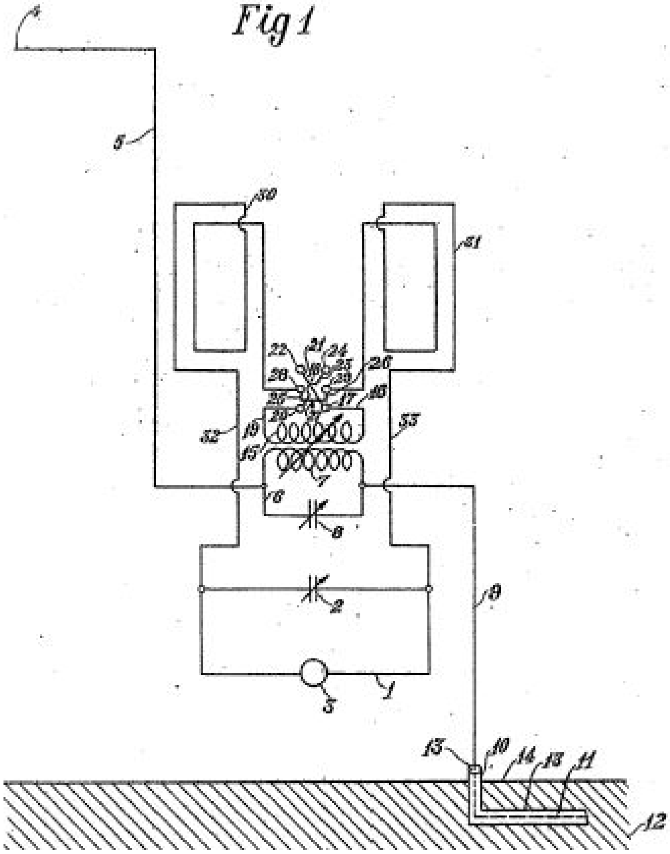
5
30
31
22
21
24
23
28
26
25
17
16
29
19
32
33
15
7
6
8
2
9
1
3
13
10
14
13
11
12

US 1826727
Radio Apparatus

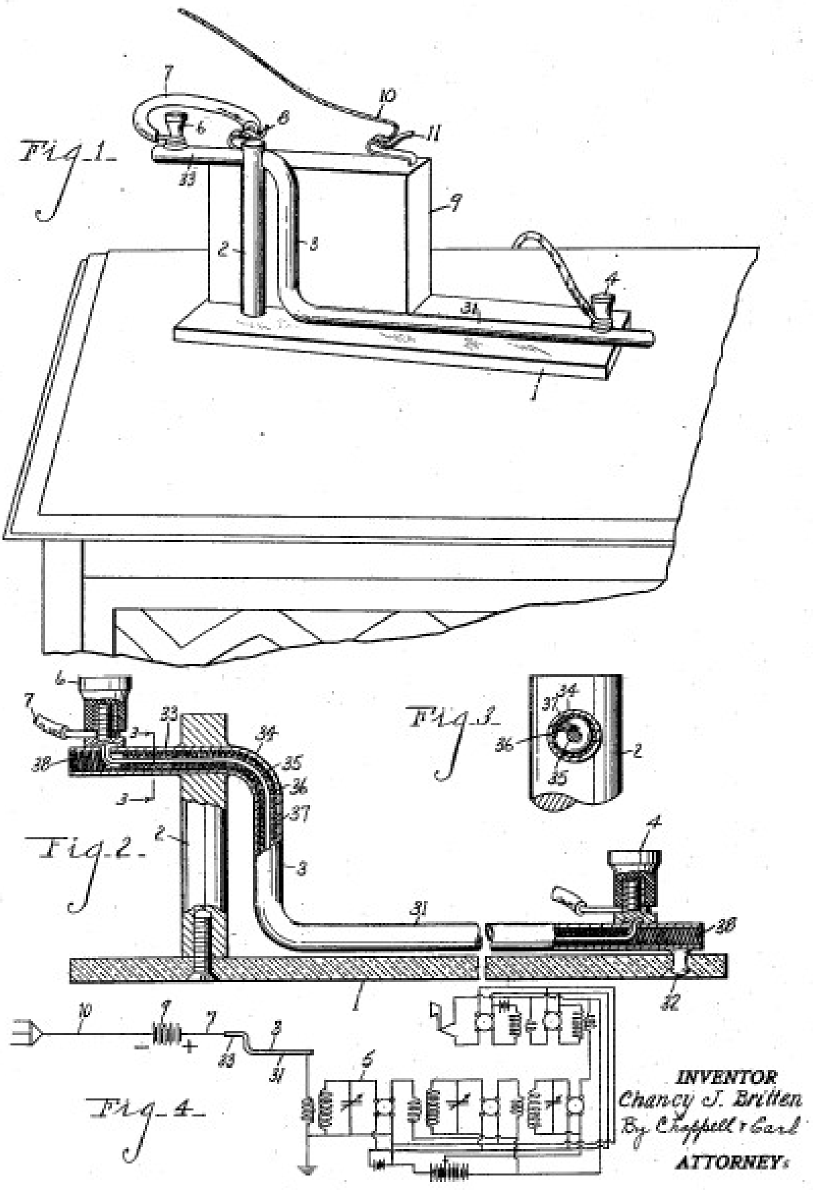
Fig_1_
Fig_2_
Fig_3_
Fig_4_
1
2
3
4
6
7
8
9
10
11
31
32
33
34
35
36
37
38
INVENTOR
Chancy J. Britten
By Chappell & Carl
ATTORNEYs

US 2143437
Radiant Energy System

Abstract -- An apparatus for receiving and transmitting radiant energy within a wide range of frequencies without necessity of a source of local energy.

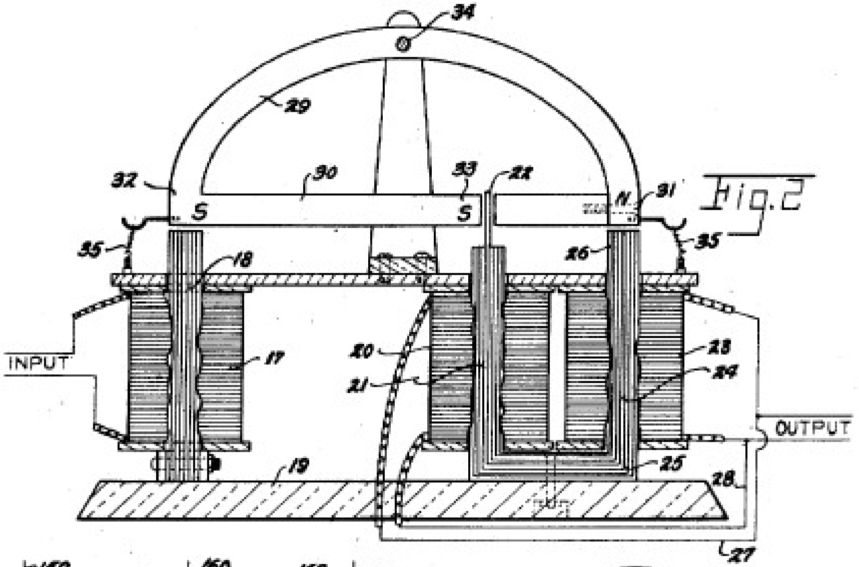
34
29
32
30
33
22
31
S
S
N
35
35
18
26
INPUT
17
20
21
23
24
OUTPUT
25
19
28
27

Fig. 2

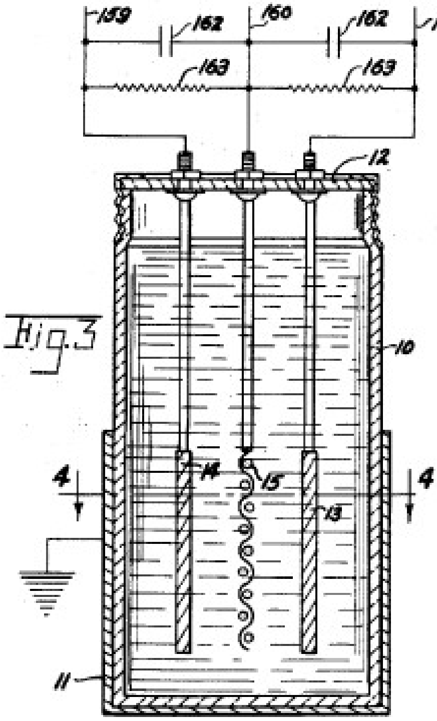
159
162
160
162
161
163
163
12
10
14
15
13
4
4
11

Fig. 3

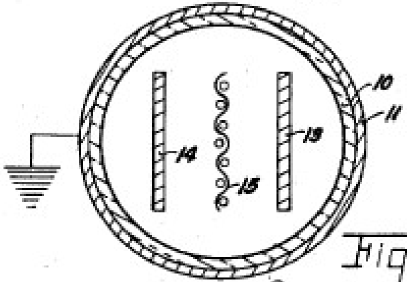
10
11
14
15
13

Fig. 4

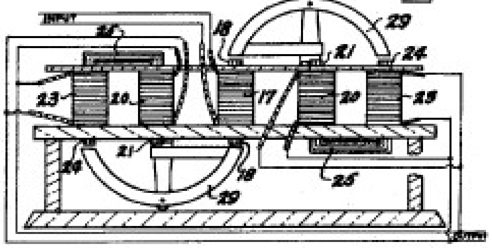
INPUT
22
18
29
21
24
23
20
17
20
23
21
24
18
29
25

Fig. 9

Inventor

James Marion Fisher

By Murray & Zugelter

Attorney

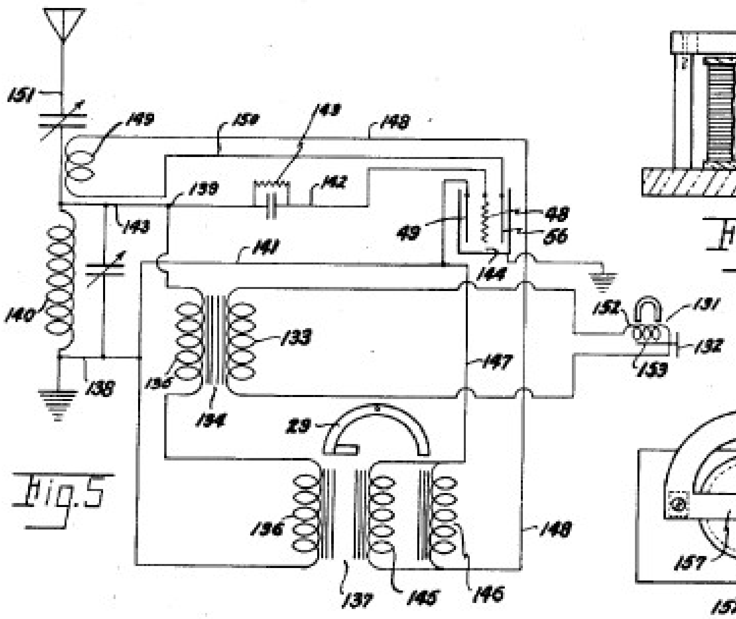
151
149
150
143
148
139
142
143
141
49
48
56
144
140
133
152
131
153
132
135
134
147
138
29
136
137
145
146
148

Fig.5

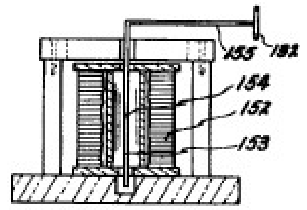
155
182
154
152
153

Fig.7

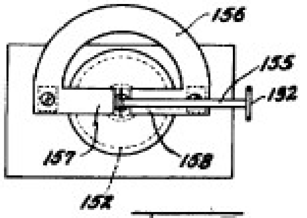
156
155
152
157
158
152

Fig.8

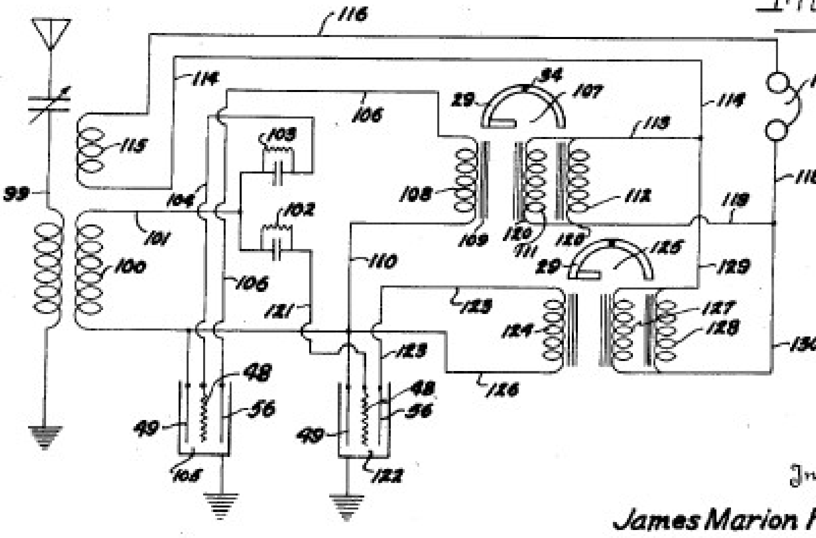
116
114
117
34
29
107
106
113
114
105
115
118
99
104
103
108
112
119
102
101
100
109
120
111
128
110
125
29
129
106
121
123
124
127
128
130
123
126
48
56
49
105
48
56
49
122

Inventor

James Marion Fisher

US 2813242
Powering Electrical Devices with Energy Abstracted from the Atmosphere

Abstract -- A convenient and economical provision of power for the operation of electronic circuits and devices using transistors, and of other electrical devices hving modest power requirements.

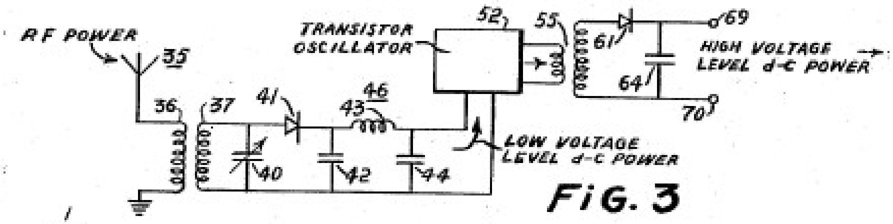
RF POWER
35
36
37
41
46
43
40
42
44
TRANSISTOR OSCILLATOR
52
55
61
64
69
70
HIGH VOLTAGE LEVEL d-c POWER
LOW VOLTAGE LEVEL d-c POWER

FiG. 3

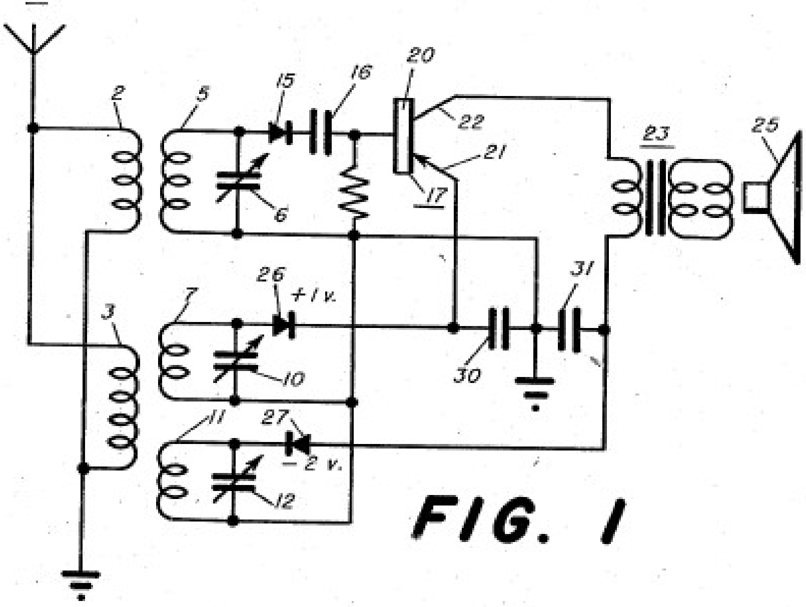
1
2
5
15
16
20
22
21
17
6
23
25
3
7
26
+1 v.
10
31
30
11
27
-2 v.
12

FIG. 1

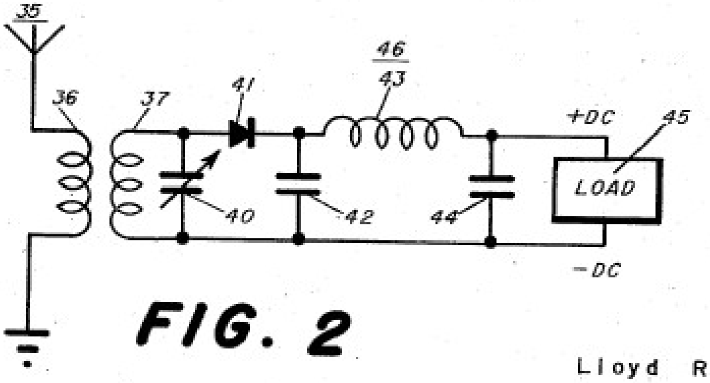
35
36
37
41
46
43
+DC
45
LOAD
40
42
44
-DC

FIG. 2

INVENTOR
Lloyd R. Crump

US 3205381
Ionospheric Battery

Abstract -- A method and apparatus for tapping the enormous reservoir of energy existing in the ionosphere...

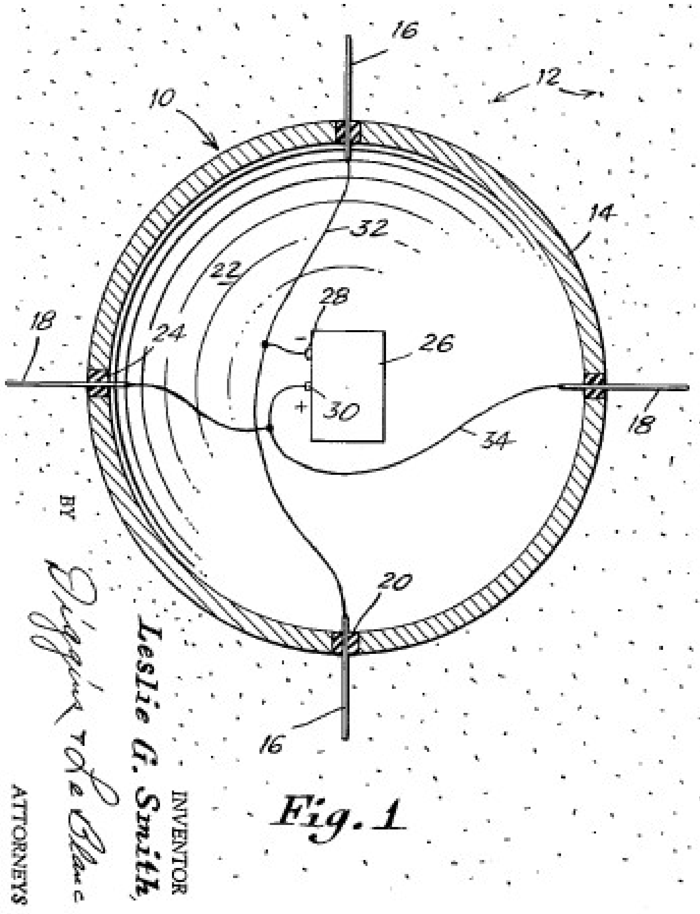
16
10
12
14
32
22
28
26
18
24
30
34
18
20
16
INVENTOR
Leslie G. Smith,
BY
ATTORNEYS

Fig. 1

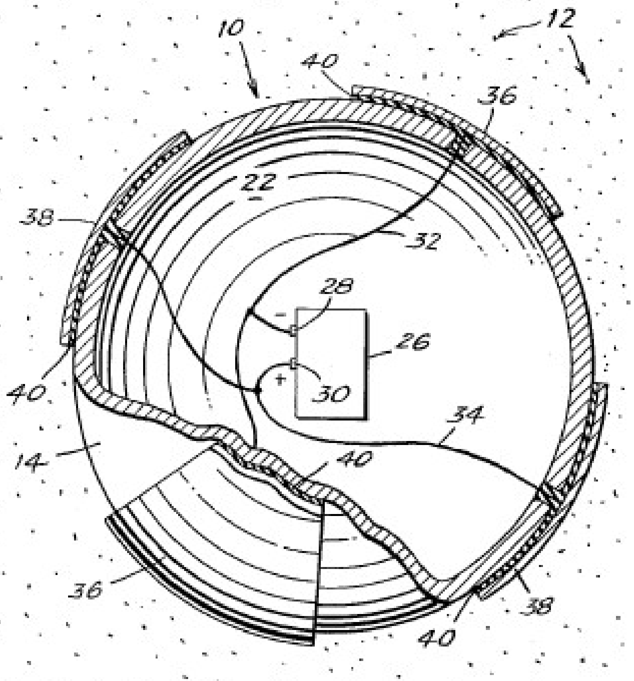
10
12
40
36
22
32
38
28
26
40
30
34
14
40
36
38
40

Fig. 2

US 3273066
Apparatus for Detecting Changes in the Atmospheric Electric Field

Abstract -- A cloud detector and or an apparatus for sensing electrostatic conditions in the earth's atmosphere...

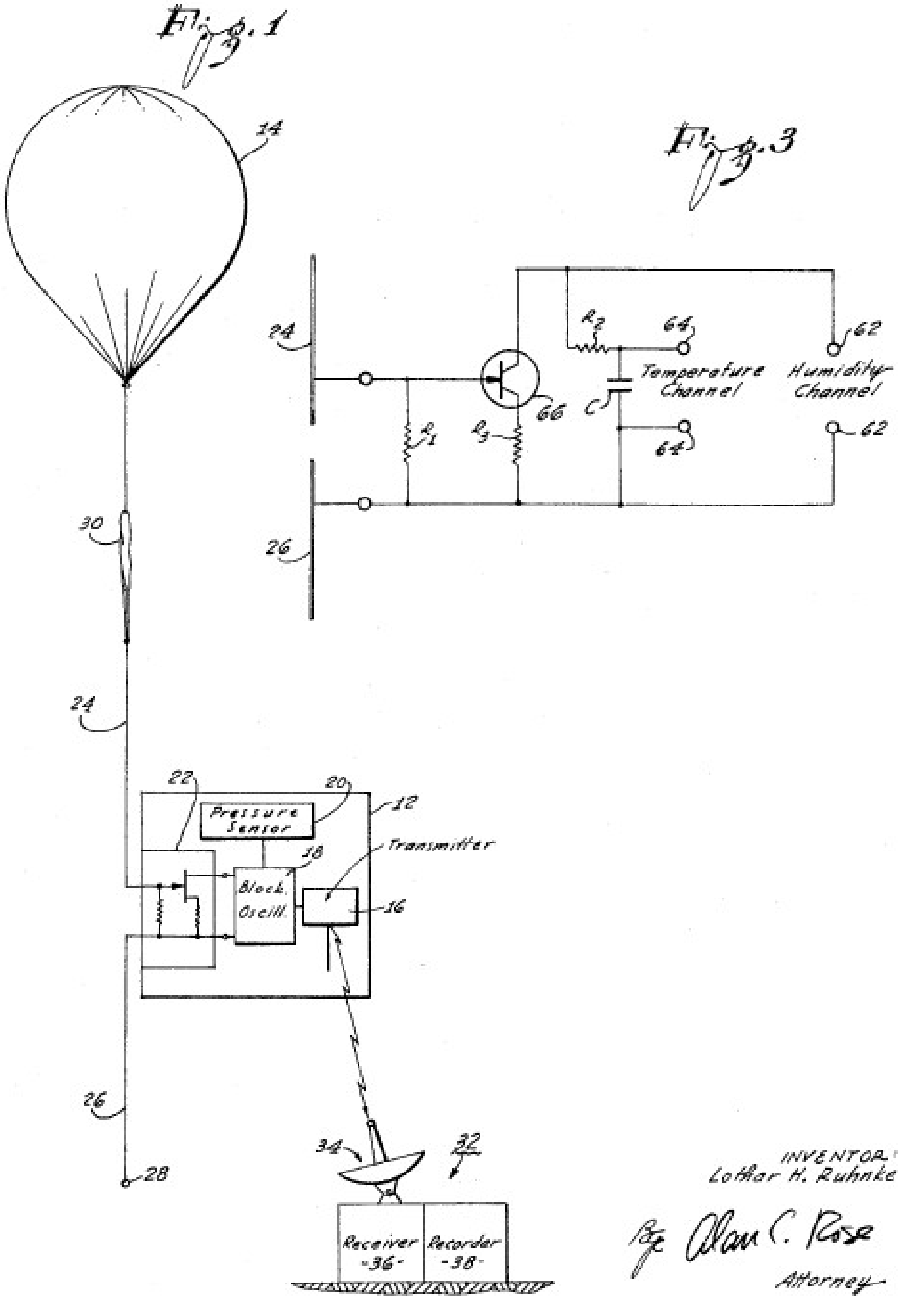
Fig. 1
14
30
24
22
20
12
Pressure Sensor
Transmitter
18
Block. Oscill.
16
26
28
34
32
Receiver -36-
Recorder -38-
Fig. 3
24
26
R1
R3
66
R2
64
C
Temperature Channel
Humidity Channel
62
64
62

INVENTOR:
Lothar H. Ruhnke

By Alan C. Rose
Attorney

US 3760257
EM Wave Energy Converter

Abstract -- Electromagnetic wave energy is converted into electric power with an array of mutually insulated electromagnetic wave absorber elements each responsive to an electric field component of the wave as it impinges thereon. Each element includes a portion tapered in the direction of wave propagation to provide a relatively wideband response spectrum. Each element includes an output for deriving a voltage replica of the electric field variations intercepted by it. Adjacent elements are positioned relative to each other so that an electric field subsists between adjacent elements in response to the impinging wave. The electric field results in a voltage difference between adjacent elements that is fed to a rectifier to derive d.c. output power. The element pairs may be arranged in a two-dimensional

array to provide power conversion of randomly polarized electromagnetic waves, such as sunlight.

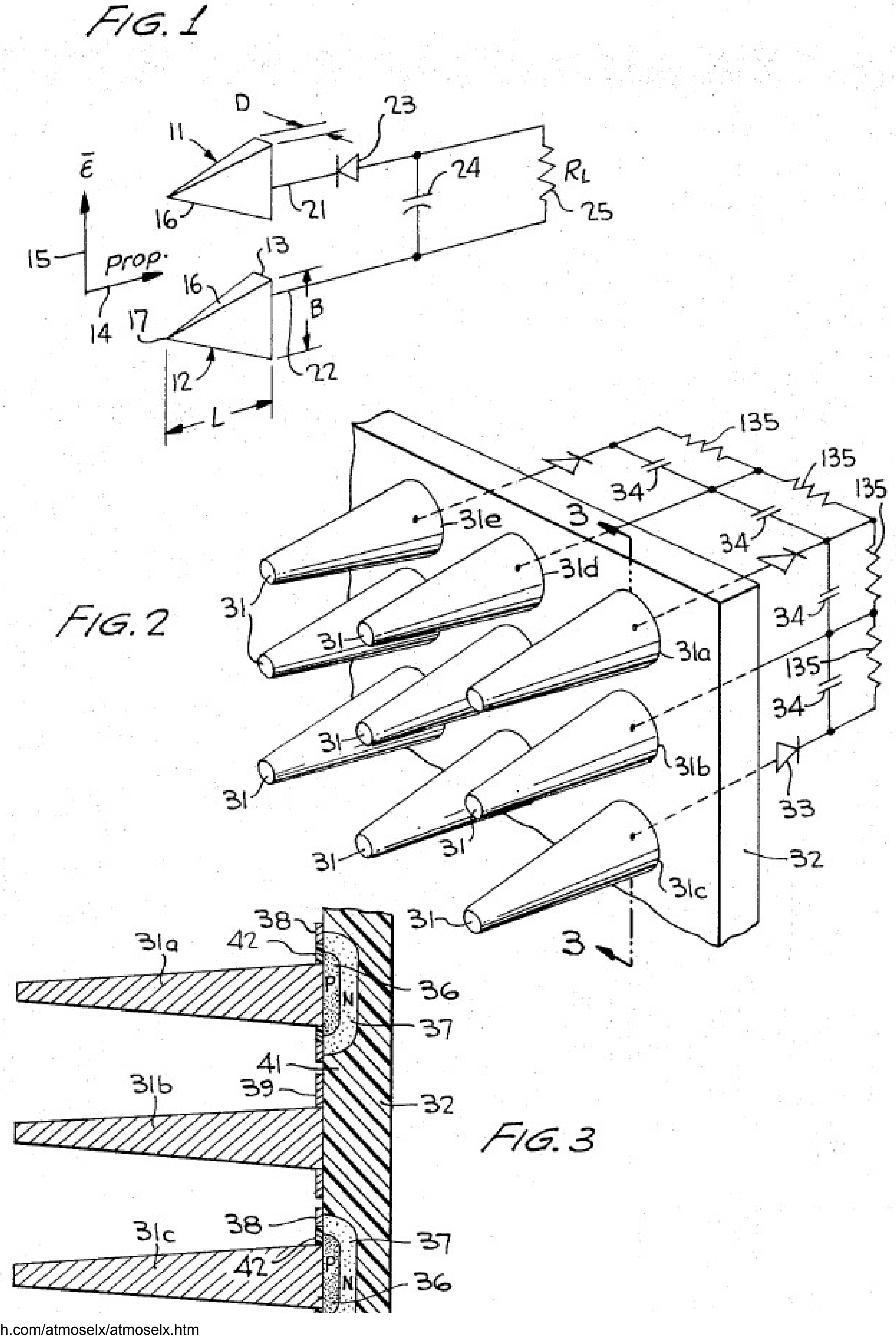

FIG. 1

FIG. 2

FIG. 3

D
23
11
$\bar{E}$
24
R_L
16
21
25
15
13
Prop.
16
B
14
17
12
22
L
135
135
135
34
34
3
31e
31d
34
31
31
31a
135
34
31
31
31b
33
31
31
31
32
31c
31
3
38
42
31a
P
N
36
37
41
31b
39
32
38
31c
37
42
P
N
36

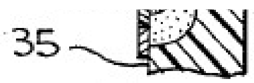

US 3931577
Radio Receiver Protection Arrangement

Abstract -- A receiver protection arrangement in which a voltage dependent on aerial voltage and a voltage dependent on aerial current are aggregated and applied to means responsive to a predetermined aggregate level to effectively short circuit the aerial lead. Additional means may be provided to dissipate energy in the input stage of a protected receiver to cause a reduced impedance to be reflected across the output of the protective arrangement.

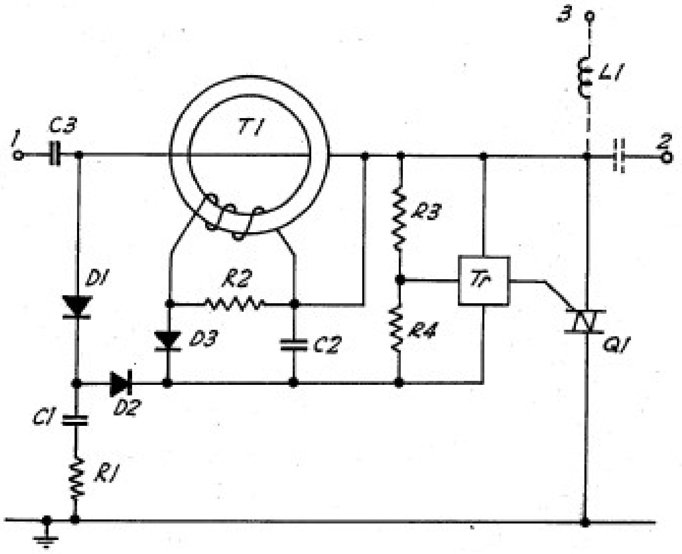

Fig. 1.

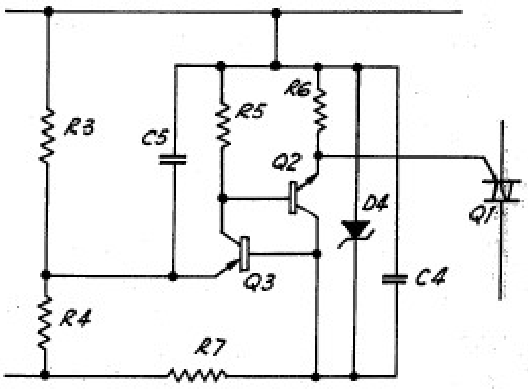

Fig. 2.

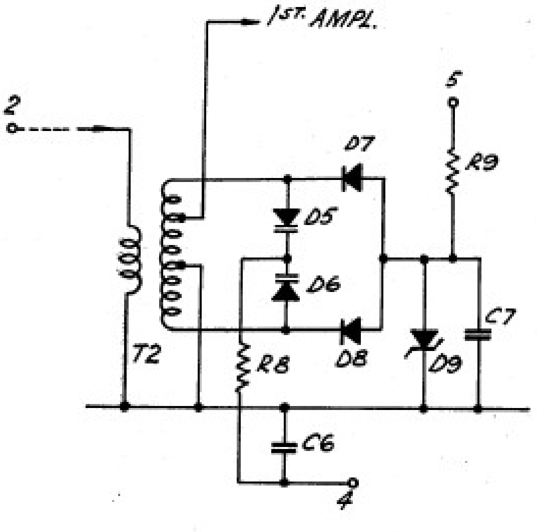
1ST. AMPL.
2
5
D7
R9
D5
D6
C7
T2
R8
D8
D9
C6
4

Fig. 3.

US 4,628,299
Seismic Warning System Using RF Energy Monitor

Abstract -- The ambient broadband radio frequency field strength from broadcast stations is monitored (Figure 4) by periodic sampling (50, 52). A warning indication is provided if the field strength drops significantly. Drops in such field strength have been correlated empirically with the occurrence of seismic activity, usually several days later. Thus the indication serves as an early warning of an impending earthquake. In one preferred embodiment, a broadband, horizontal, very long monopole antenna (40) was connected to a rectifying and smoothing circuit (Figure 3) to provide a dc output proportional to the ambient rf field...

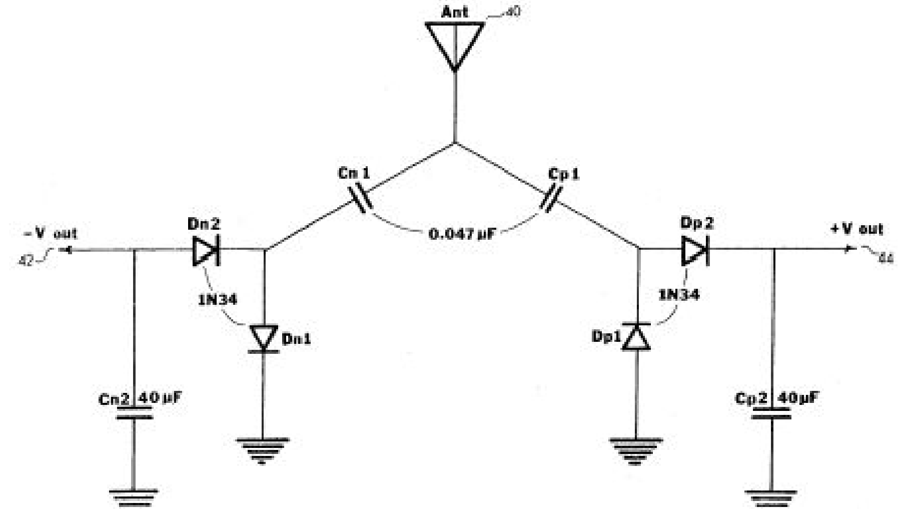
Ant
40
Cn 1
Cp 1
0.047 µF
-V out
42
Dn2
Dp2
+V out
44
1N34
Dn1
1N34
Dp1
Cn2 40 µF
Cp2 40 µF

FIG 3

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

Fig. 1

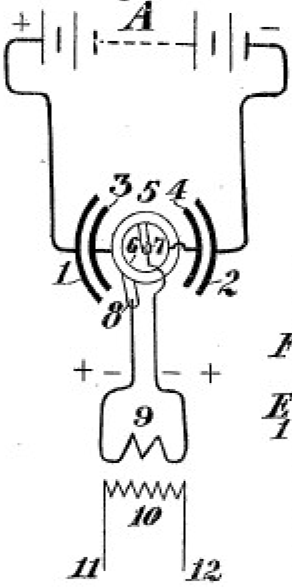
A
+
−
3
5
4
6
7
1
2
8
+
−
−
+
9
10
11
12

Fig. 3

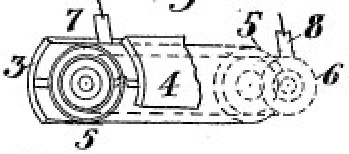

Fig. 2

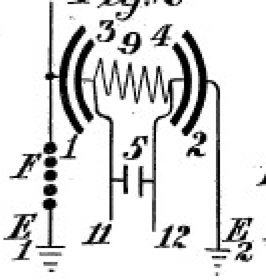
3
9
4
1
2
F
5
E
1
11
12
E
2

Fig 4

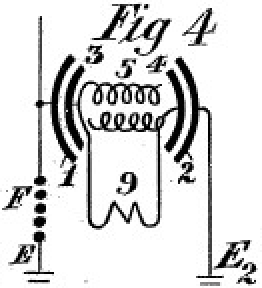
3
5
4
1
2
F
9
E
E
2

Fig. 5.

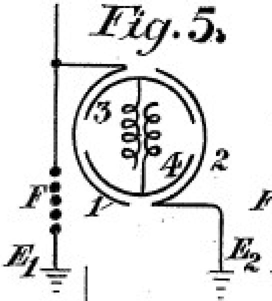

Fig. 6.

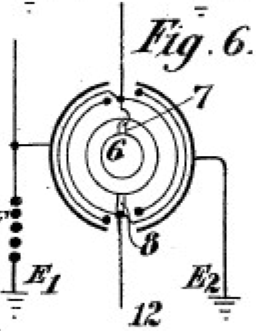
7
6
F
8
E
1
E
2
12

Fig. 7.

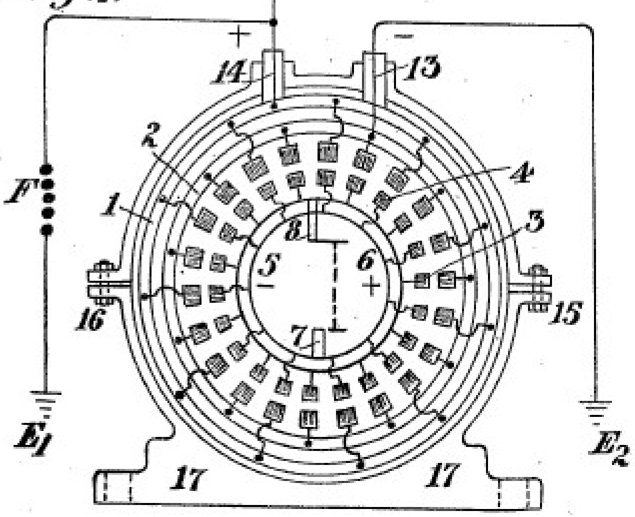
+
−
14
13
2
4
F
1
3
8
5
6
−
+
16
15
7
E
1
E
2
17
17

Fig.8

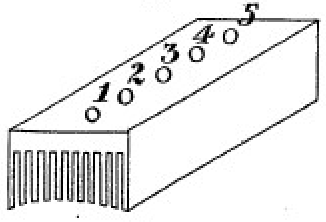

Fig. 10

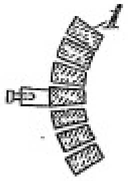

Fig. 9.

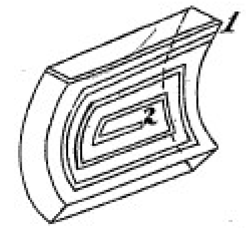

Fig. 11

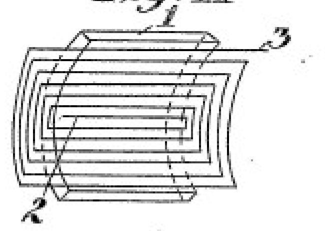

Fig.12

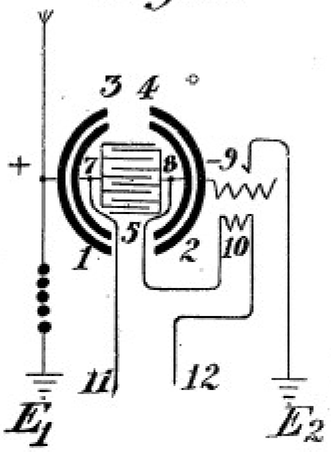
+
3
4
7
8
9
1
5
2
10
11
12
E_1
E_2

Fig. 13.

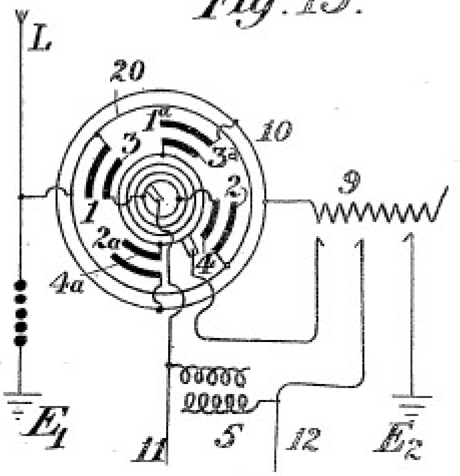
L
20
1^a
10
3
3^a
2
9
1
2^a
4
4^a
11
5
12
E_1
E_2

Fig. 14

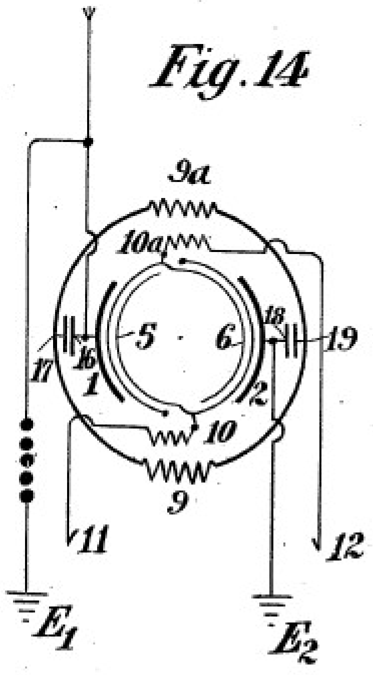
9a
10a
5
6
18
19
17
16
1
2
10
9
11
12
E1
E2

Fig. 16

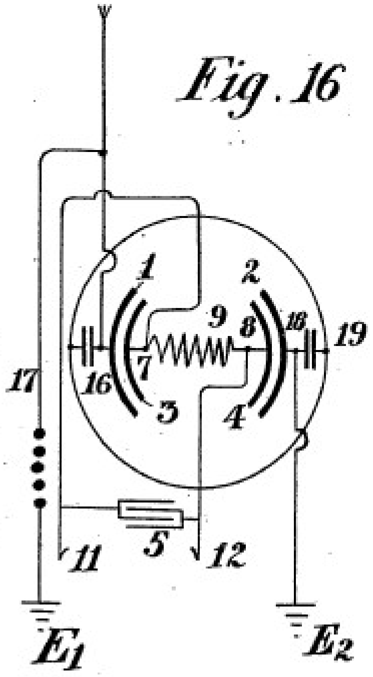
1
2
9
8
18
19
17
16
7
3
4
5
11
12
E1
E2

Fig. 15

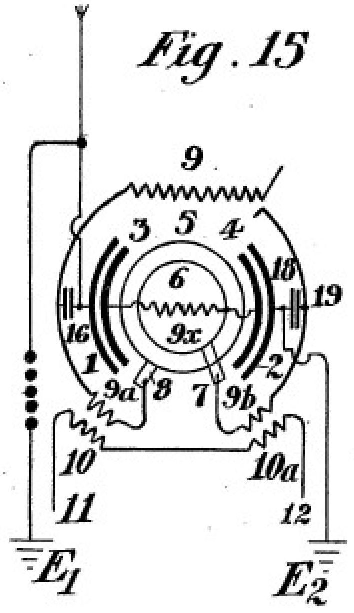
9
3
5
4
6
18
19
16
9x
1
2
9a
8
7
9b
10
10a
11
12
E1
E2

Fig. 17.

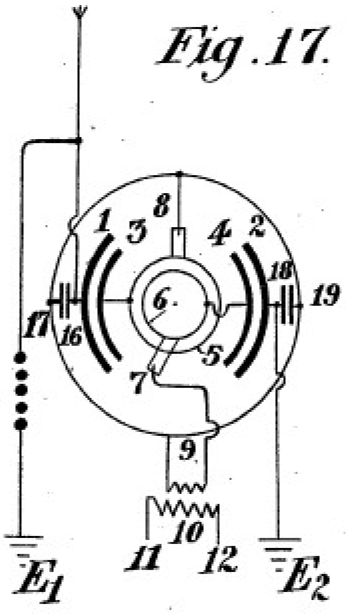
8
1
3
4
2
18
19
17
16
6
5
7
9
10
11
12
E1
E2

French Patent # 551,882

Apparatus for Capture of Electric Currents in the Atmosphere

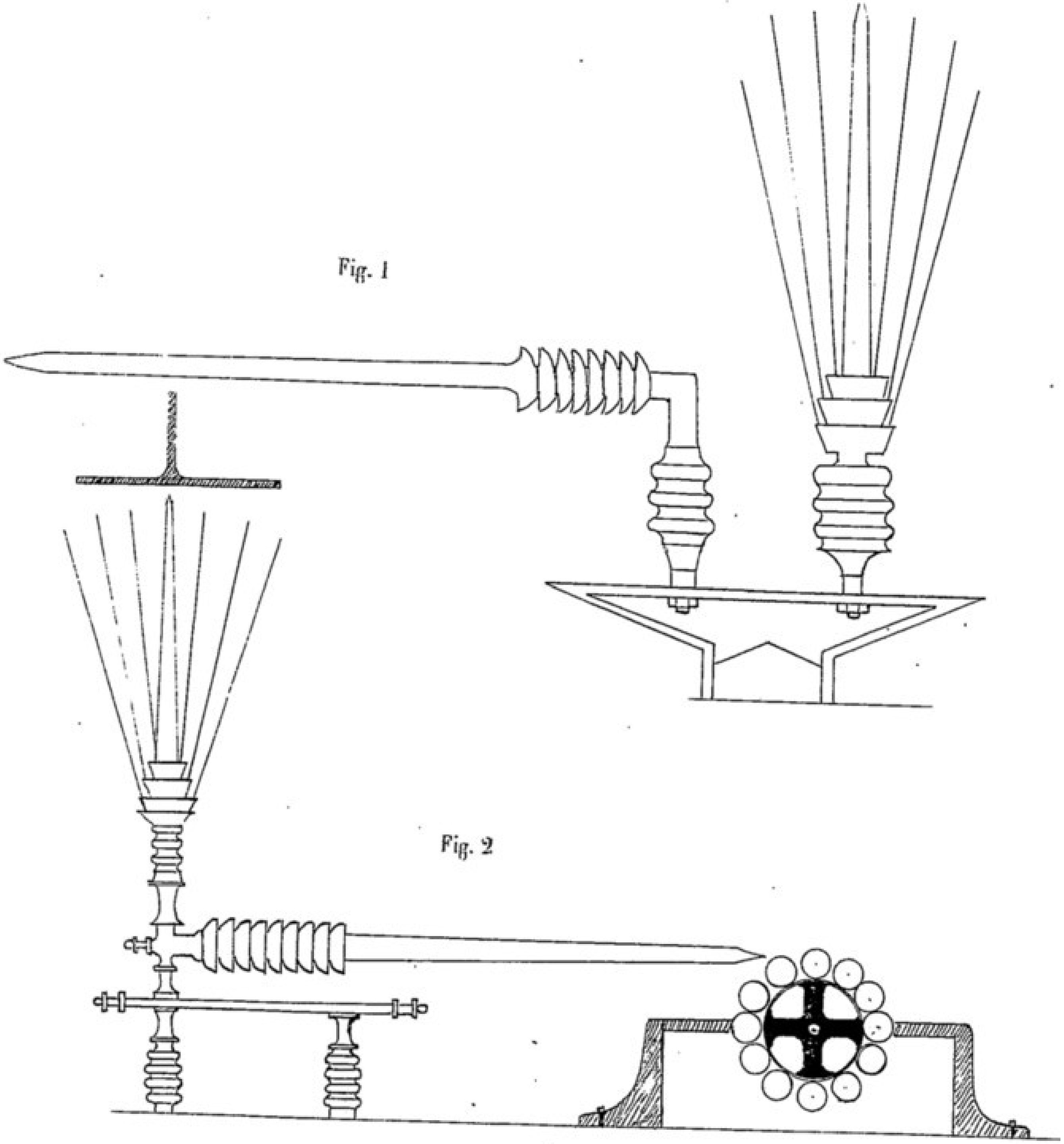

Fig. 1

Fig. 2

French Patent # 565,395

Combined Apparatus for Capture of Atmospheric Electric Currents with Immediate Implementation

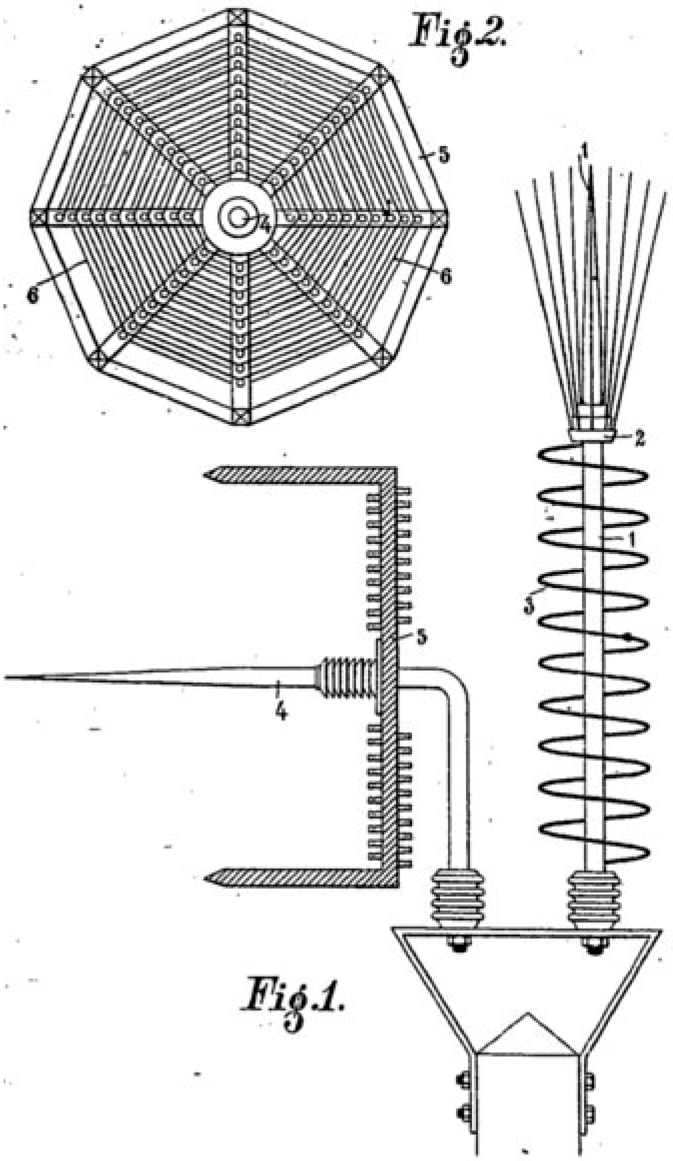
5
4
1
6
6
1
2
1
3
5
4

Fig.2.

Fig.1.

Fig. 3.

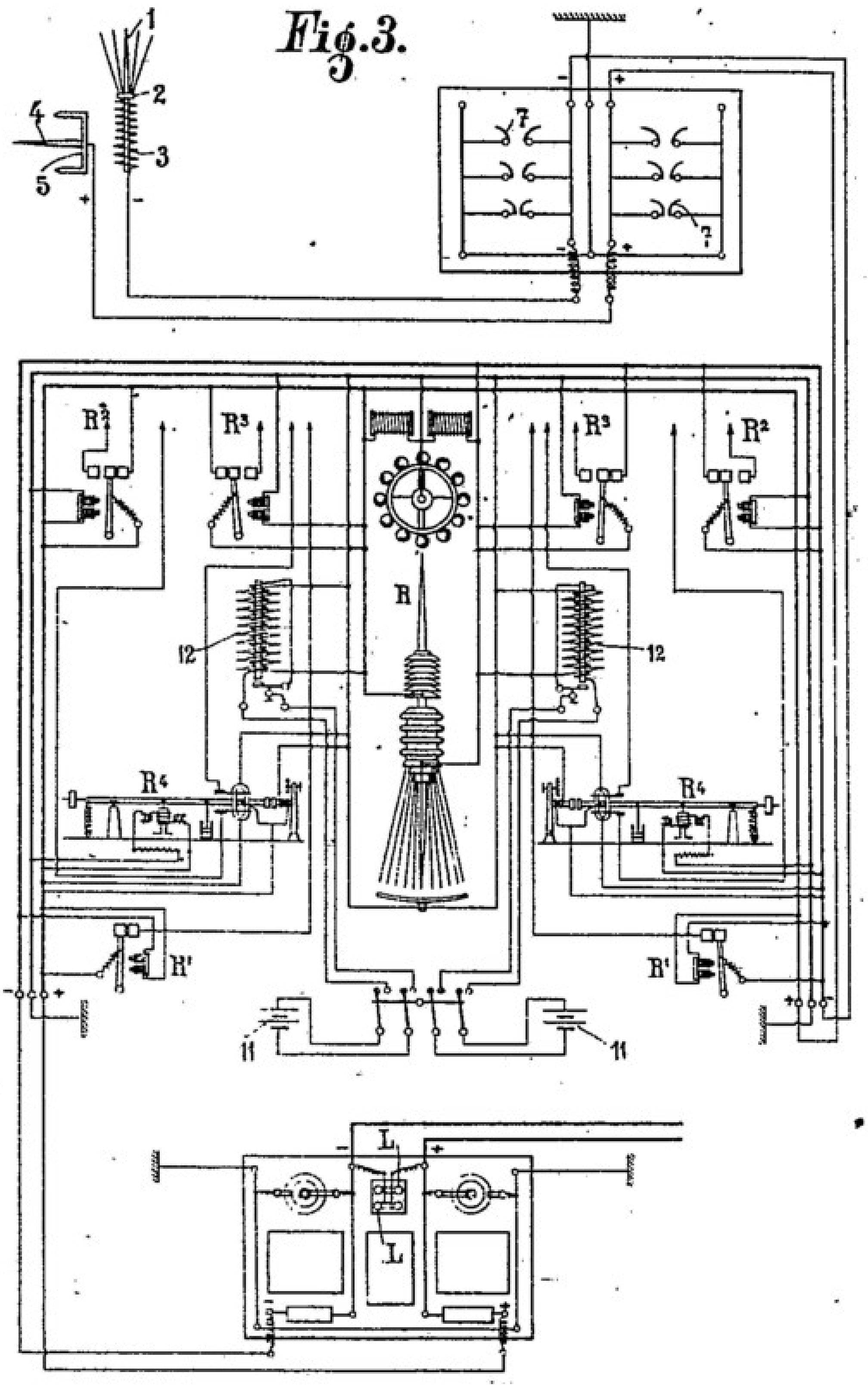
1
2
3
4
5
+
-
7
7
-
+
-
+
R²
R³
R³
R²
R
12
12
R⁴
R⁴
R¹
R¹
-
+
+
-
11
11
L
L
-
+

Fig. 4.

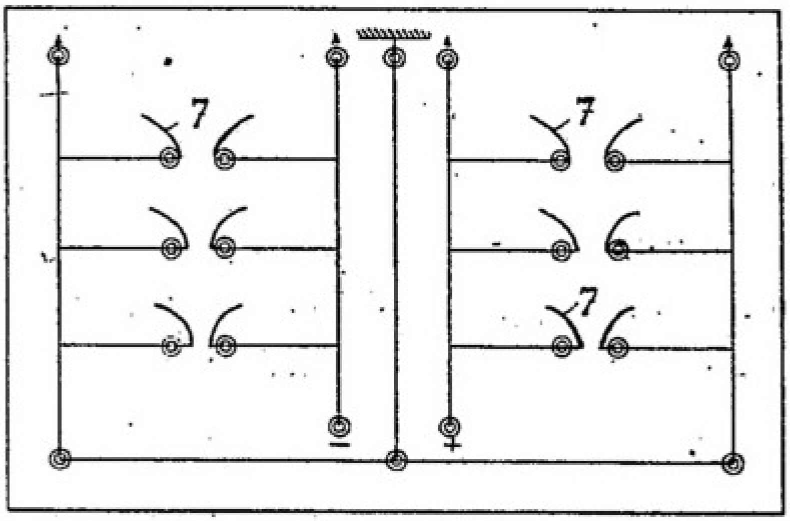

Fig. 5.

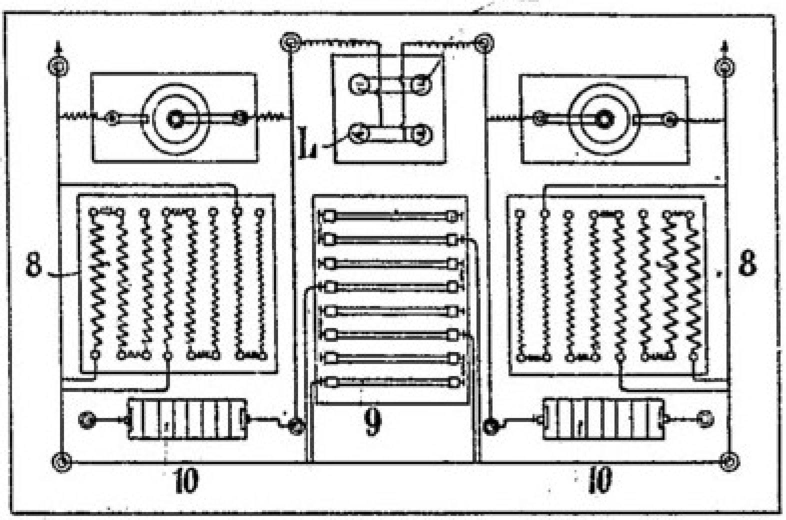

US 28793
Electric Apparatus

Abstract -- An aerostat for obtaining atmpspheric and terrestrial electricity...

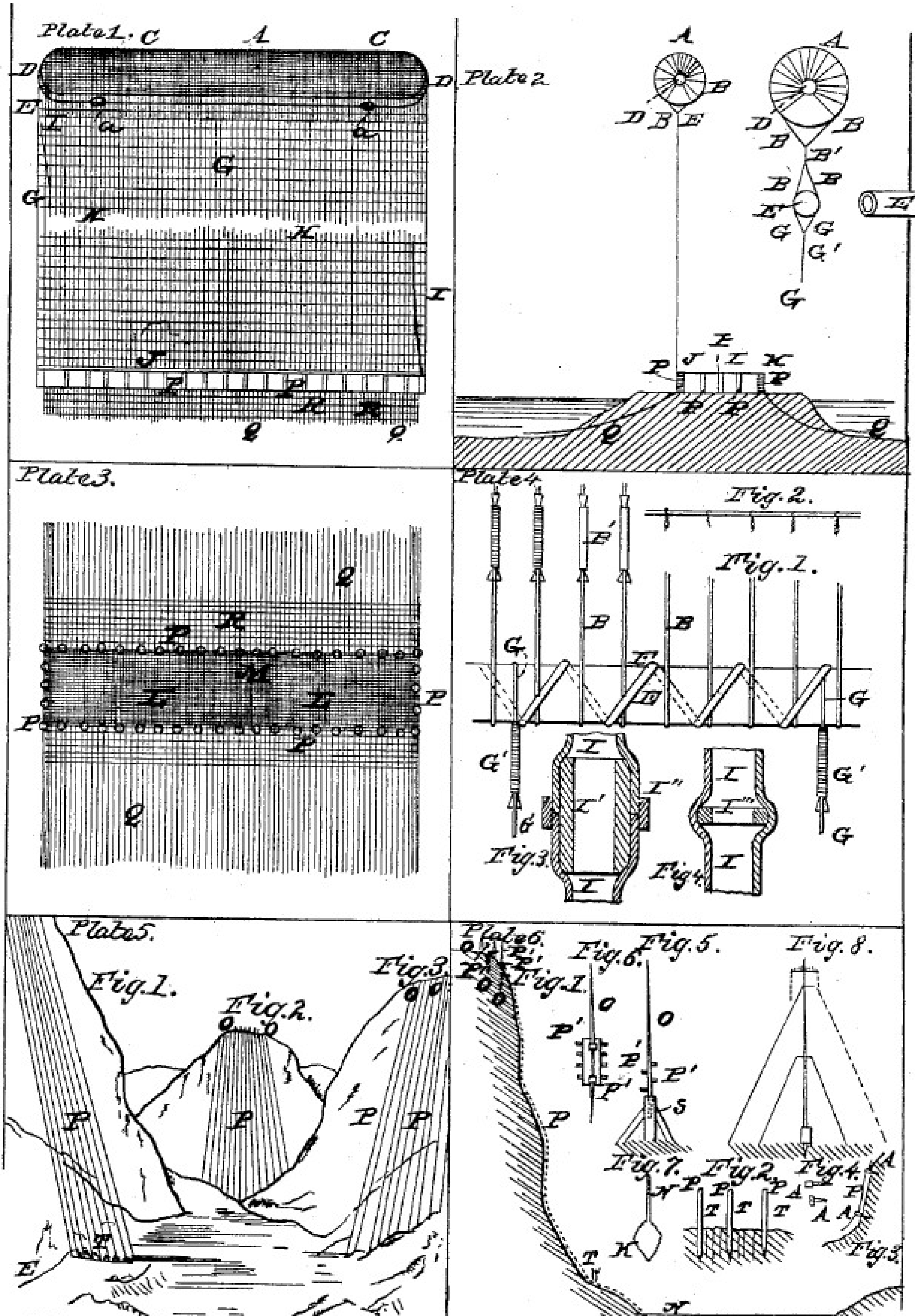
Plate 1.
A
C
D
E
G
H
I
J
P
R
Q
a
Plate 2.
A
B
B'
D
E
G
G'
I
J
K
P
Q
Plate 3.
Q
R
P
M
L
Plate 4.
Fig. 2.
Fig. 1.
B
E'
E
D
G
G'
I
I'
I''
I'''
Fig. 3.
Fig. 4.
Plate 5.
Fig. 1.
Fig. 2.
Fig. 3.
P
E
Plate 6.
Fig. 1.
Fig. 5.
Fig. 6.
Fig. 8.
O
P
P'
S
Fig. 7.
Fig. 2.
Fig. 4.
Fig. 3.
N
K
T
A
P'

Witnesses:
Arthur L. McIntire
Louis Bosworth.

Inventor
H. G. Kon
by Munn & Co.
Attorneys.

N. PETERS. Photo-Lithographer, Washington, D. C.

US 126356
Improvement in Collecting Energy for Telegraph

Abstract -- A tower for the purpose of receiving and imparting natural electricity, so as to be in constant contact with the upper stratum of electricity which surrounds the earth, by tapping which a never-failing supply is formed when brought into contact with the earth.

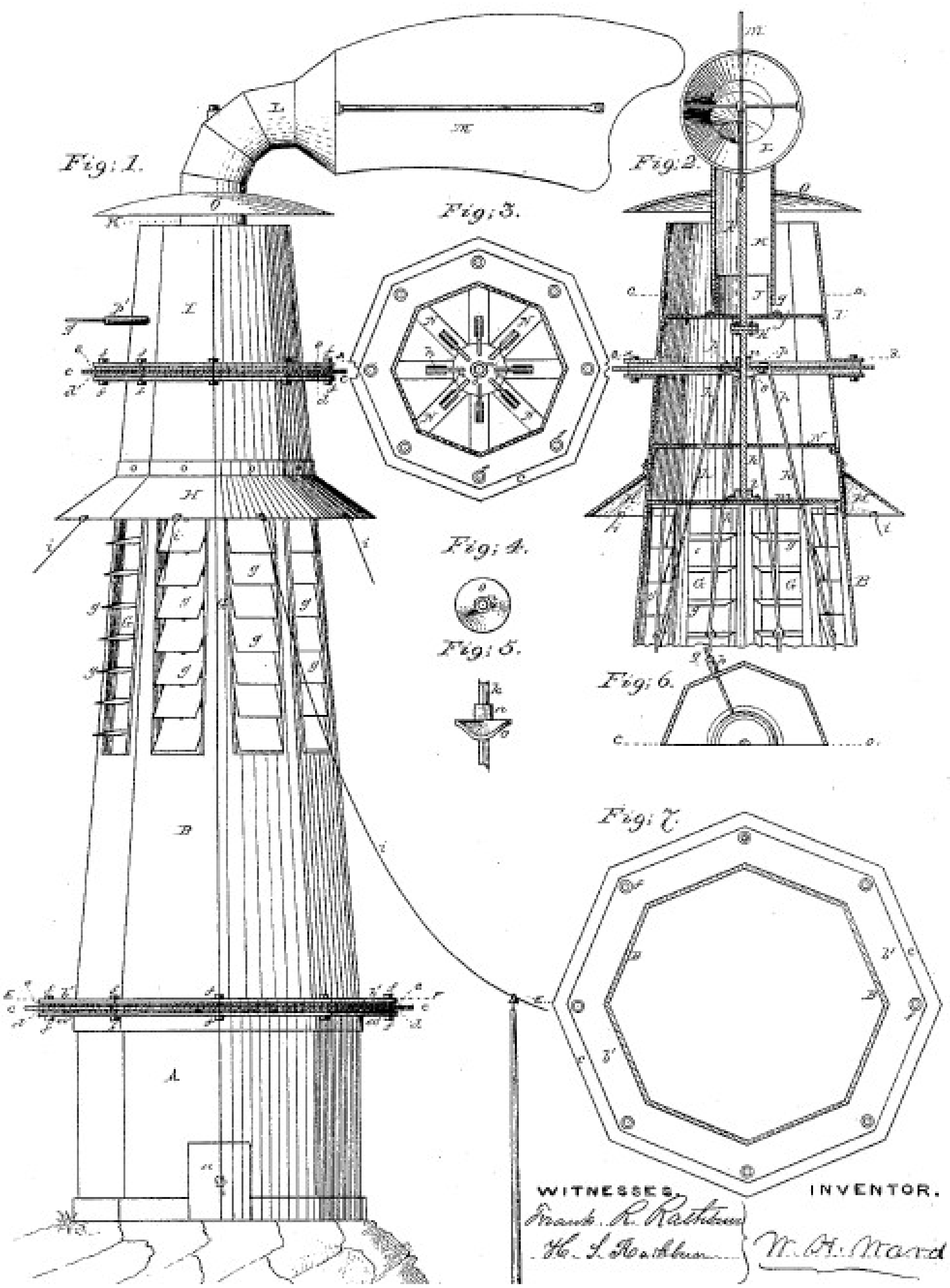
Fig: 1.
Fig: 2.
Fig: 3.
Fig: 4.
Fig: 5.
Fig: 6.
Fig: 7.
WITNESSES.
INVENTOR.

US 129971
Telegraph Power

Abstract -- Utilizing natural electricity and establishing an electrical current without the aid of wires, batteries, or cables...

US 674427
Collecting Atmospheric Electricity

Abstract -- An appartus for colecting and driving or conducting atmospheric electricity which renders it practicable to obtain material quantities of teh same in serviceable form.

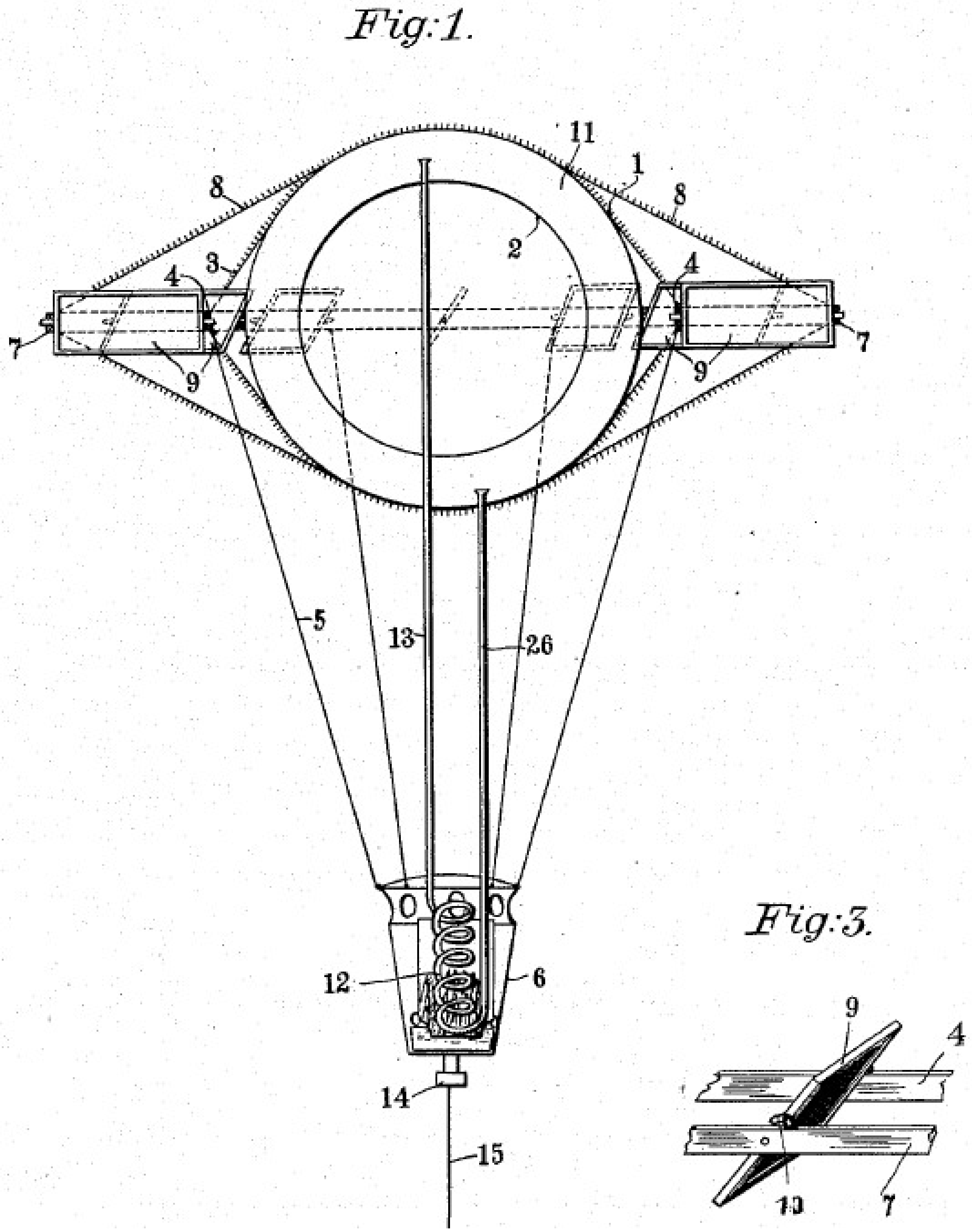

Fig:1.

Fig:3.

Fig: 2.

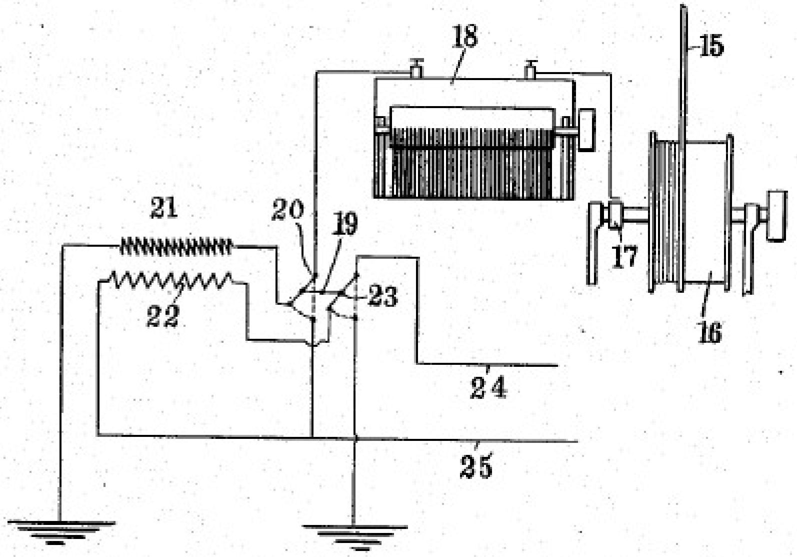

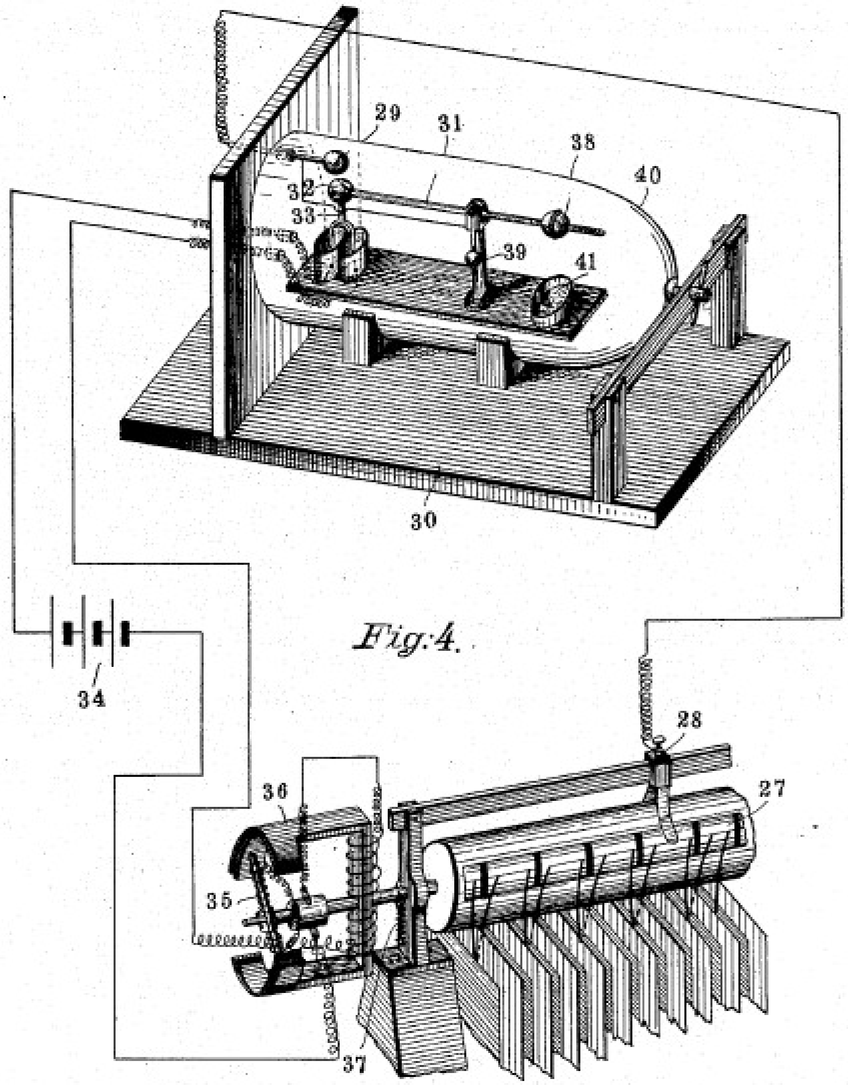
Fig: 4.

US 911260
Collecting Atmospheric Electricity

Abstract -- A method of colecting electricity from a strata laden with electricity at high altitudes in the atmoshere, through the medium of a wire cable suspended from one or more balloons ...

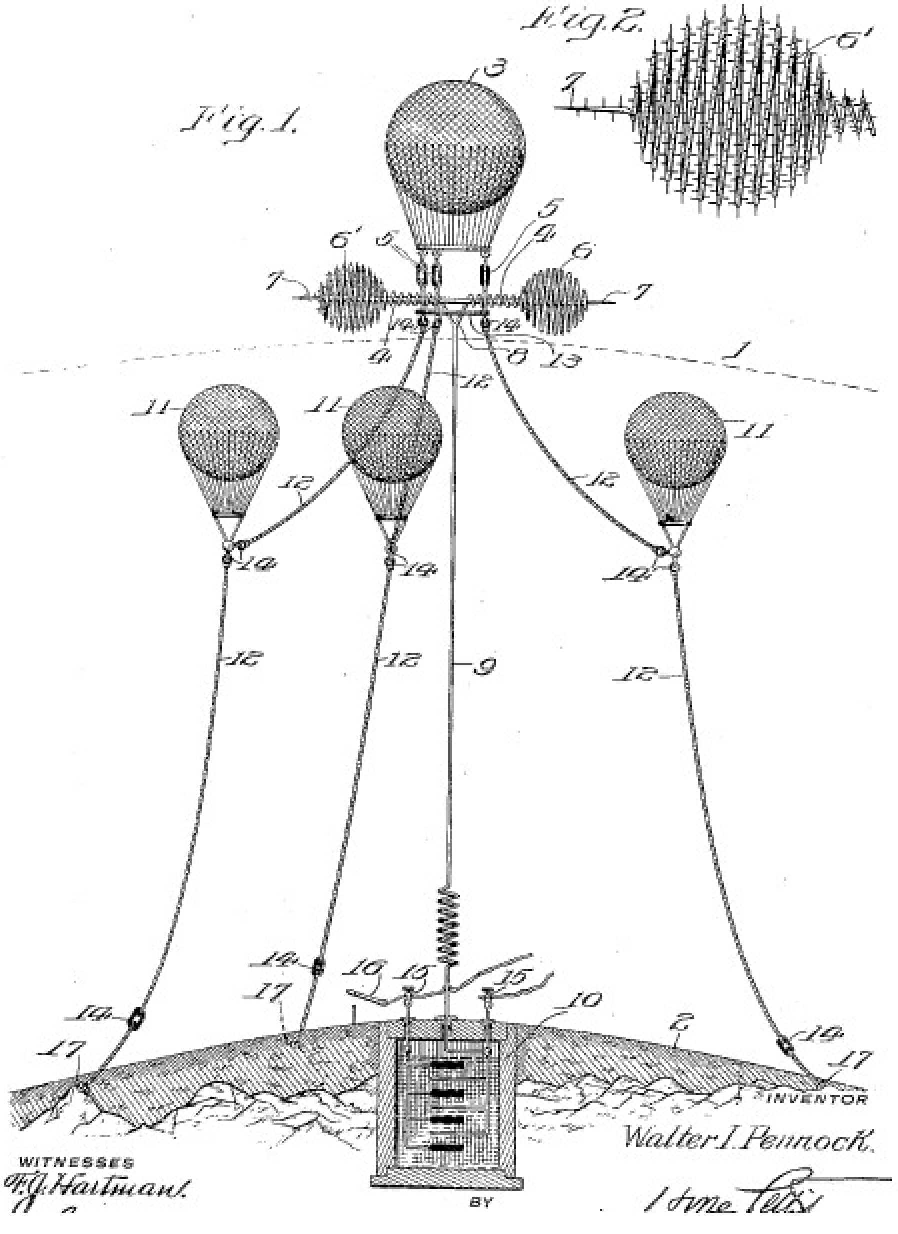
Fig. 2.
6'
7
Fig. 1.
3
5
6'
5
4
6
7
7
14
8
13
1
11
11
11
12
12
12
14
14
14
12
12
9
12
14
16
15
15
10
17
2
14
14
17
17
INVENTOR
Walter I. Pennock.
WITNESSES
F.J. Hartman
BY

US 1014719
Collecting Atmospheric Electricity

Abstract -- A method of colecting electricity from a strata laden with electricity at high altitudes in the atmoshere, through the medium of a wire cable suspended from one or more balloons ...

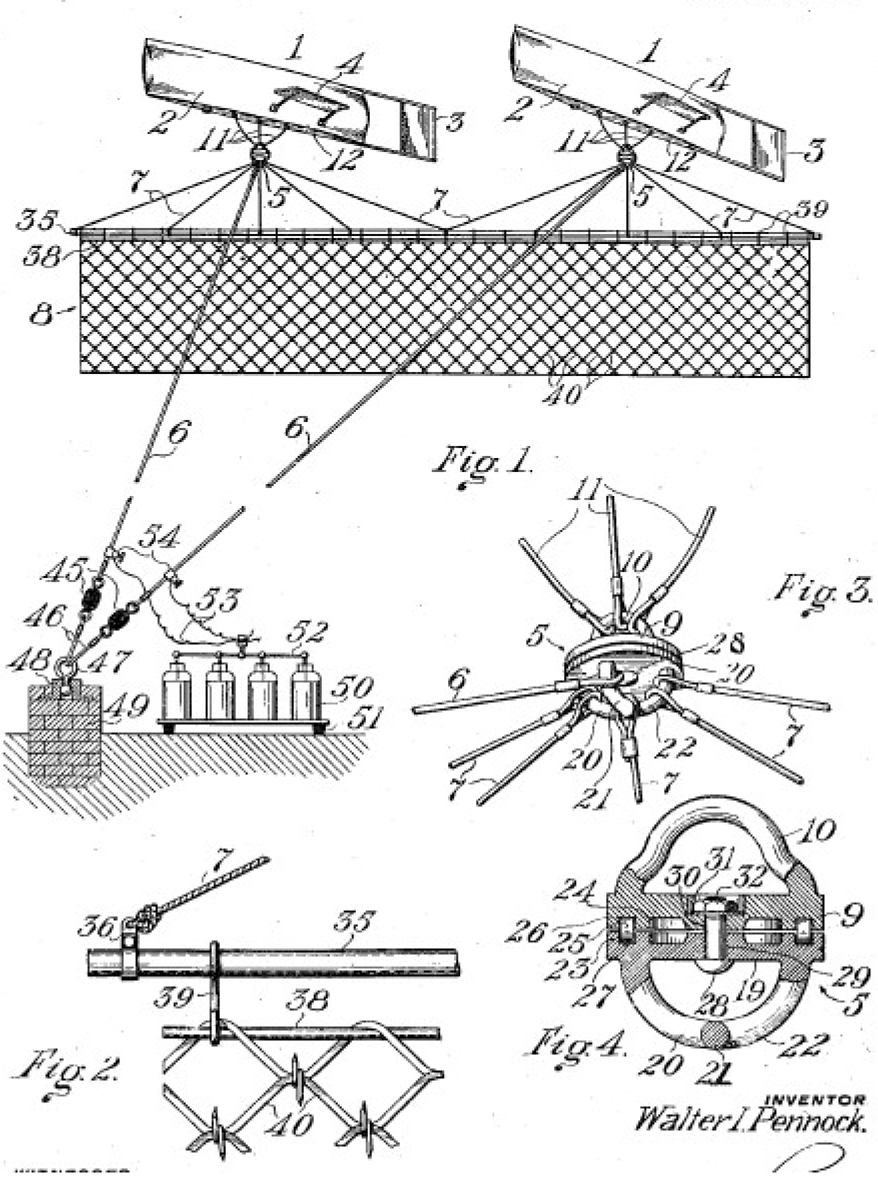
1
4
2
3
11
12
5
7
35
38
39
8
40
6
54
45
33
46
52
48
47
50
49
51
11
10
9
5
28
20
6
22
7
20
21
7
10
24
30
31
32
9
26
25
23
27
19
28
29
5
20
21
22
7
36
35
39
38
40

Fig. 1.

Fig. 2.

Fig. 3.

Fig. 4.

INVENTOR
Walter I. Pennock.

Fig. 5.

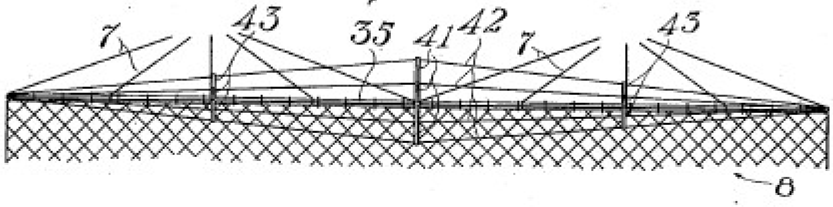
7
43
35
41
42
7
43
8

Fig. 6.

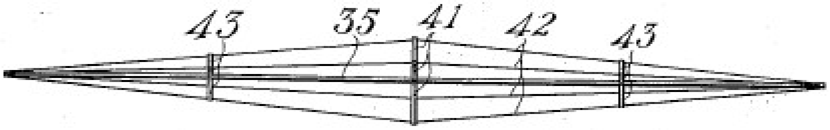
43
35
41
42
43

Fig. 7.

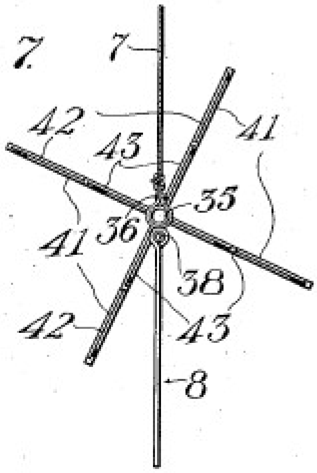
7
42
41
43
35
36
41
38
42
43
8

INVENTOR
Walter I. Pennock.

US 1540998
Conversion of Atmospheric Electric Energy

Abstract -- Method of obtaining atmospheric electricity by means of metallic ballon collectors...

Filed Jan. 15, 1921 12 Sheets-She

Fig. 1.

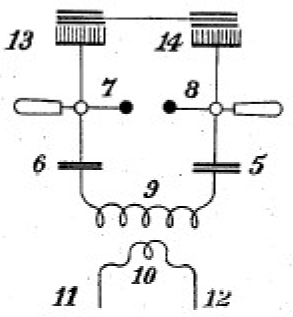
13
14
7
8
6
5
9
10
11
12

Fig. 2.

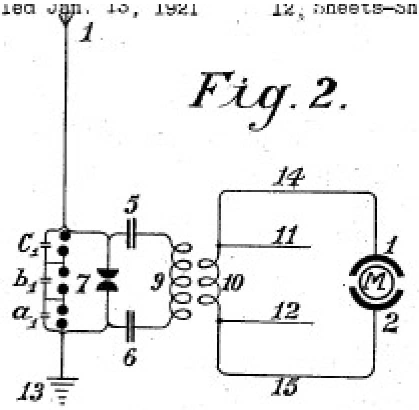
1
14
5
11
C_1
b_1
7
9
10
M
a_1
12
2
6
13
15

Fig. 3.

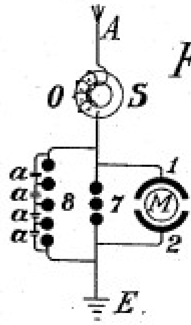
A
O
S
a
a
a
a
8
7
1
M
2
E.

Fig. 4.

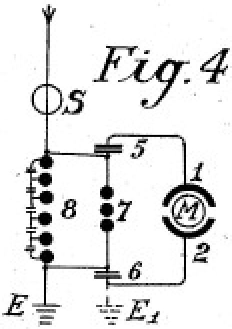
S
5
1
8
7
M
2
6
E
E_1

Fig. 5.

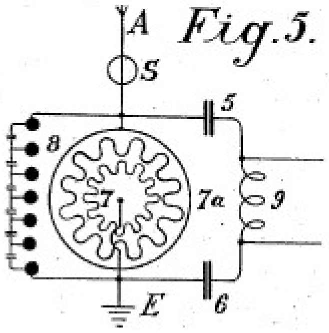
A
S
5
8
7
7a
9
E
6

Fig. 6

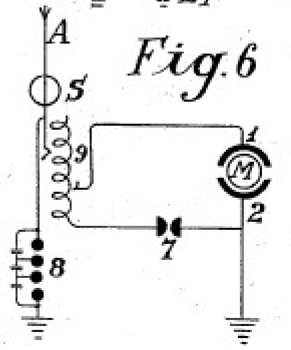

Filed Jan. 13, 1921 12 Sheets-Sheet 2

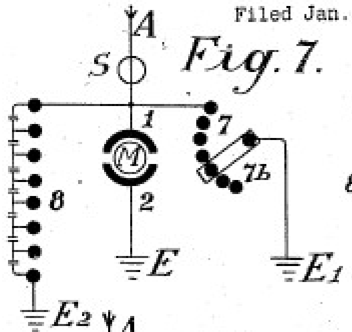
Fig. 7.
A
S
1
M
2
7
7b
8
E
E1
E2

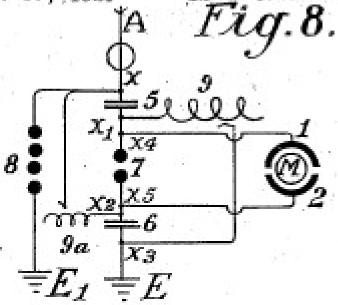
Fig. 8.
A
X
5
9
X1
X4
7
1
M
2
8
X5
X2
6
9a
X3
E1
E

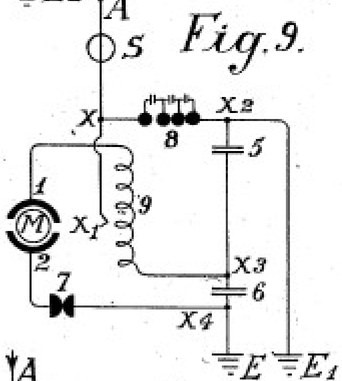
Fig. 9.
A
S
X
8
X2
5
1
M
X1
9
2
7
X3
6
X4
E
E1

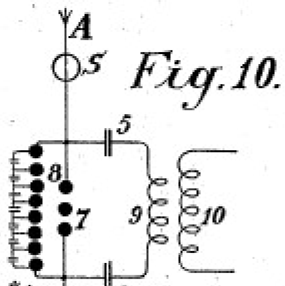
Fig. 10.
A
S
5
8
7
9
10
6
E

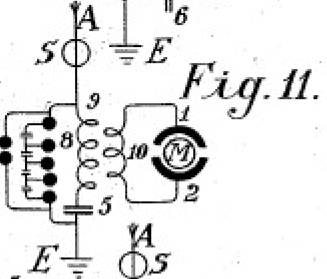
Fig. 11.
A
S
9
7
8
10
1
M
2
5
E

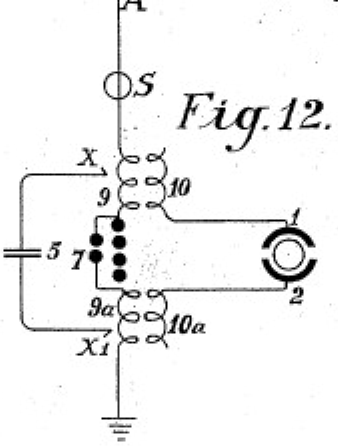
Fig. 12.
A
S
X
9
10
1
5
7
2
9a
10a
X1

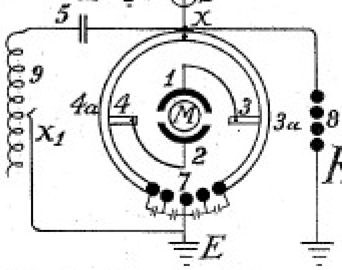
Fig. 13.
A
S
5
X
9
1
M
4a
4
3
3a
8
X1
2
7
E

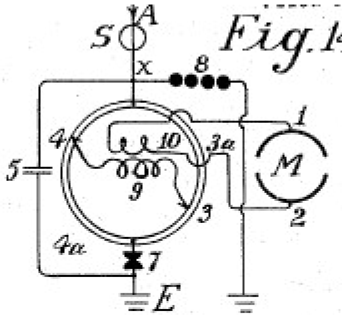
A
S
X
8
1
4
10
3a
M
5
9
3
2
4a
7
E

Fig. 14.

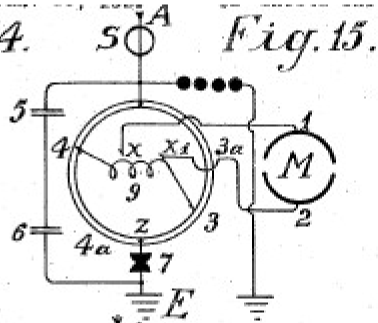
A
S
5
1
4
X
X1
3a
M
9
2
Z
3
6
4a
7
E

Fig. 15.

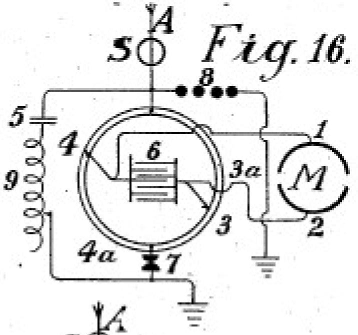
A
S
8
5
1
4
6
3a
M
9
3
2
4a
7

Fig. 16.

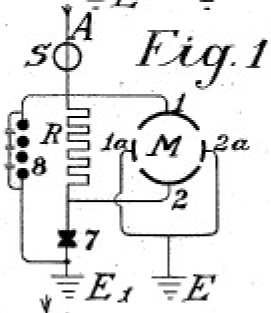
A
S
1
R
1a
M
2a
8
2
7
E1
E

Fig. 17.

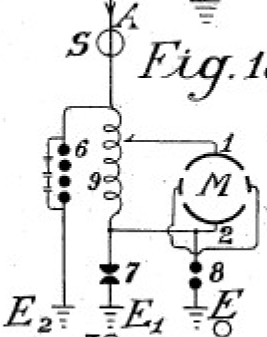
A
S
1
6
9
M
2
7
8
E2
E1
E

Fig. 18.

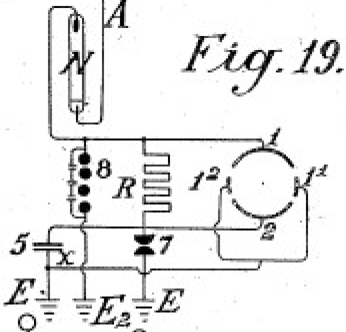
A
N
1
8
R
1²
1¹
2
5
X
7
E
E2
E

Fig. 19.

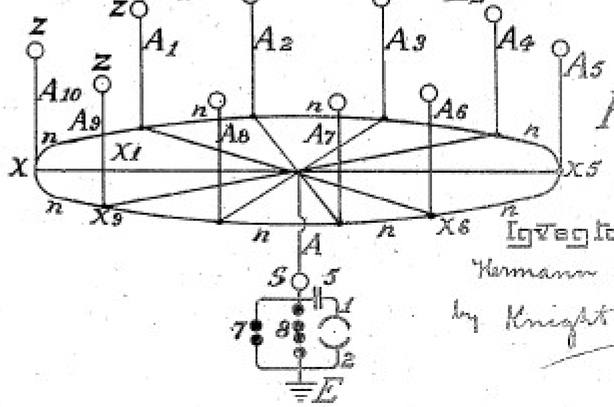
Z
Z
Z
A1
A2
A3
A4
A5
A10
A9
n
A8
n
A7
A6
n
n
X
X1
X5
n
X9
n
A
n
X6
n
S
5
1
7
8
2
E

Fig. 20.

Inventor
Hermann Plauson
by Knight Bros
Attorneys

Fig. 21. Fig. 22.

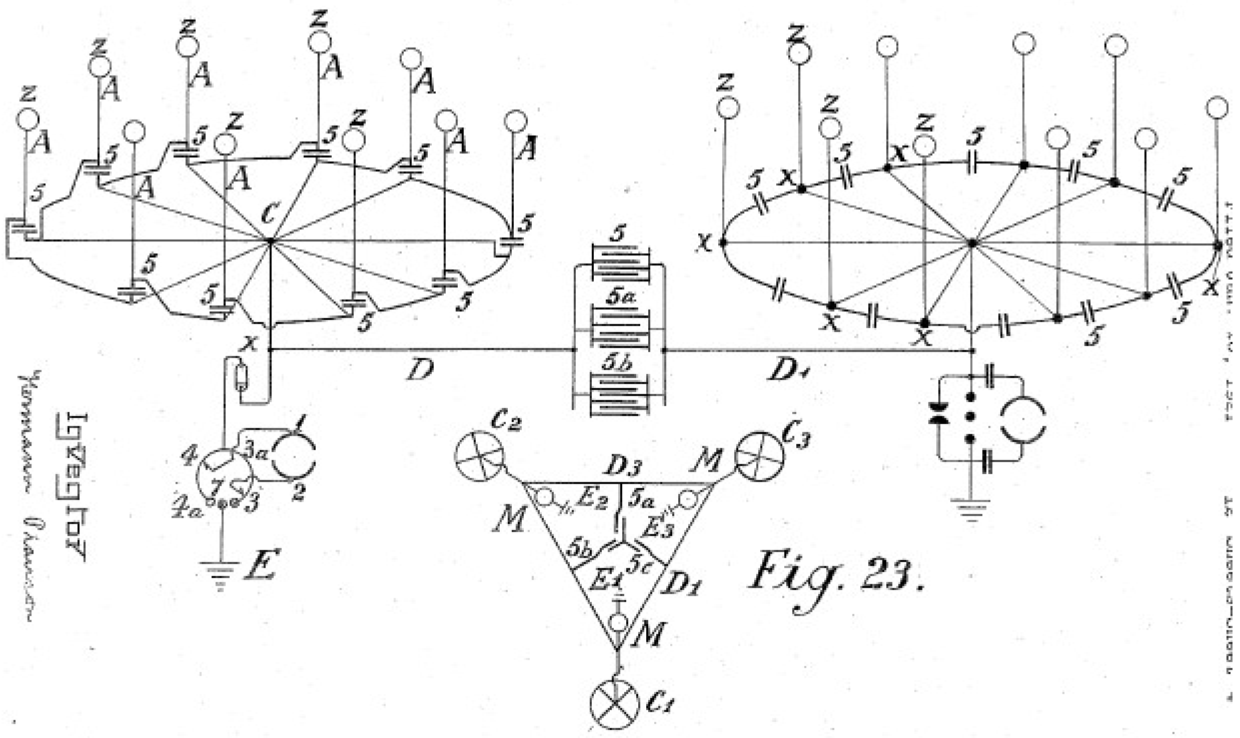
Z
A
5
C
x
D
D1
5a
5b
X
C2
C3
D3
M
E2
E3
5c
E1
D1
C1
1
2
3
3a
4
4a
7
E
Inventor
Hermann Plauson

Fig. 23.

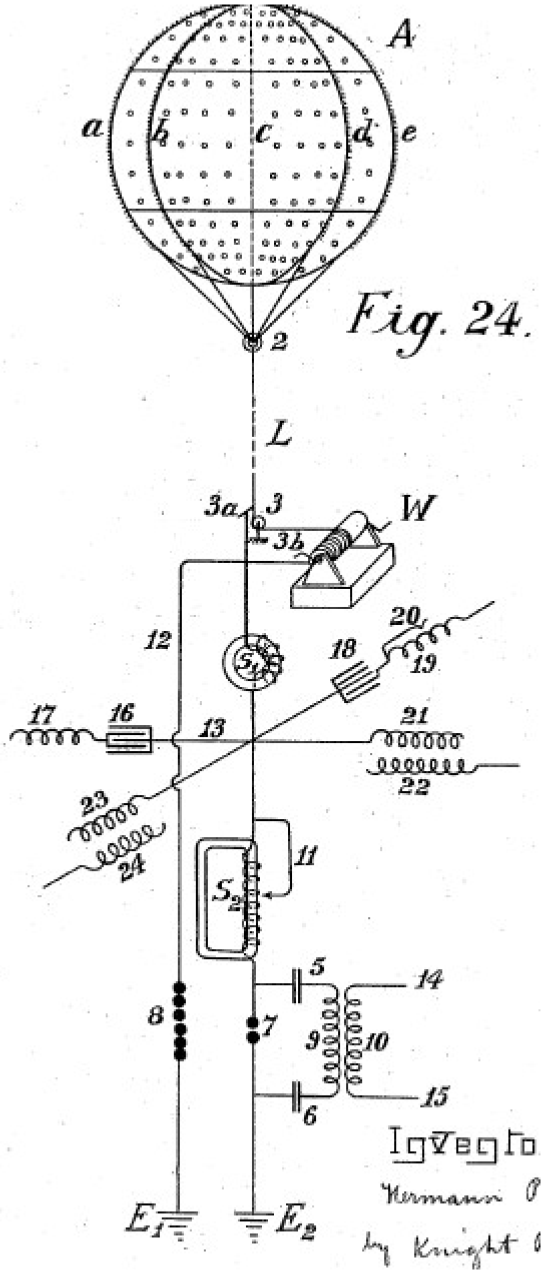

Fig. 24.

A
a
b
c
d
e
2
L
3a
3
3b
W
12
S1
18
20
19
17
16
13
21
22
23
24
11
S2
5
14
8
7
9
10
6
15
E1
E2
Invento
Hermann
by Knight

Fig. 25.

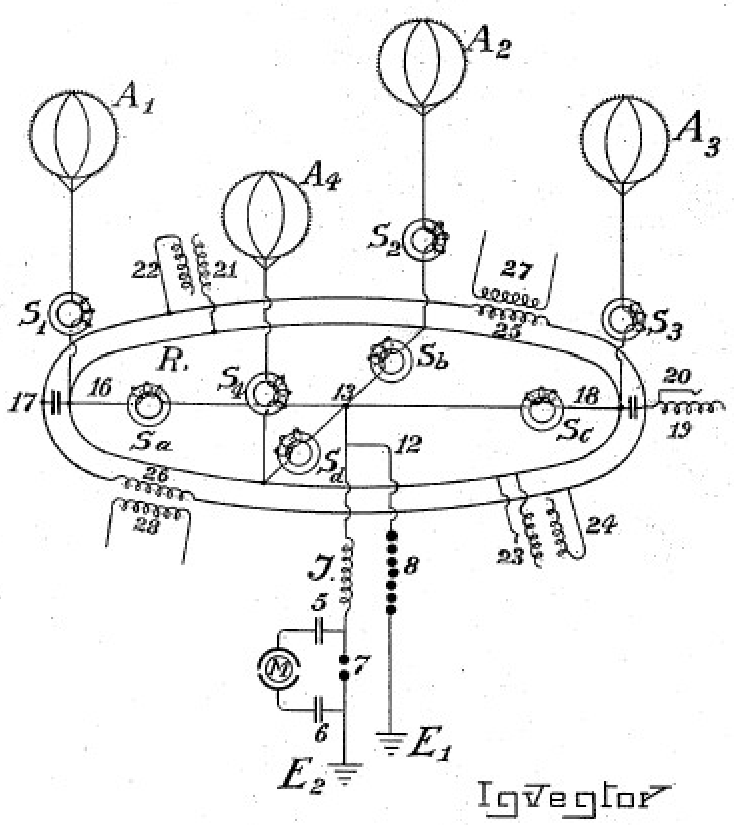
A1
A2
A3
A4
S2
22
21
27
S1
25
S3
R.
Sb
20
17
16
S4
13
18
Sa
Sd
12
Sc
19
26
28
24
23
7.
8
5
M
7
6
E1
E2
Inventor

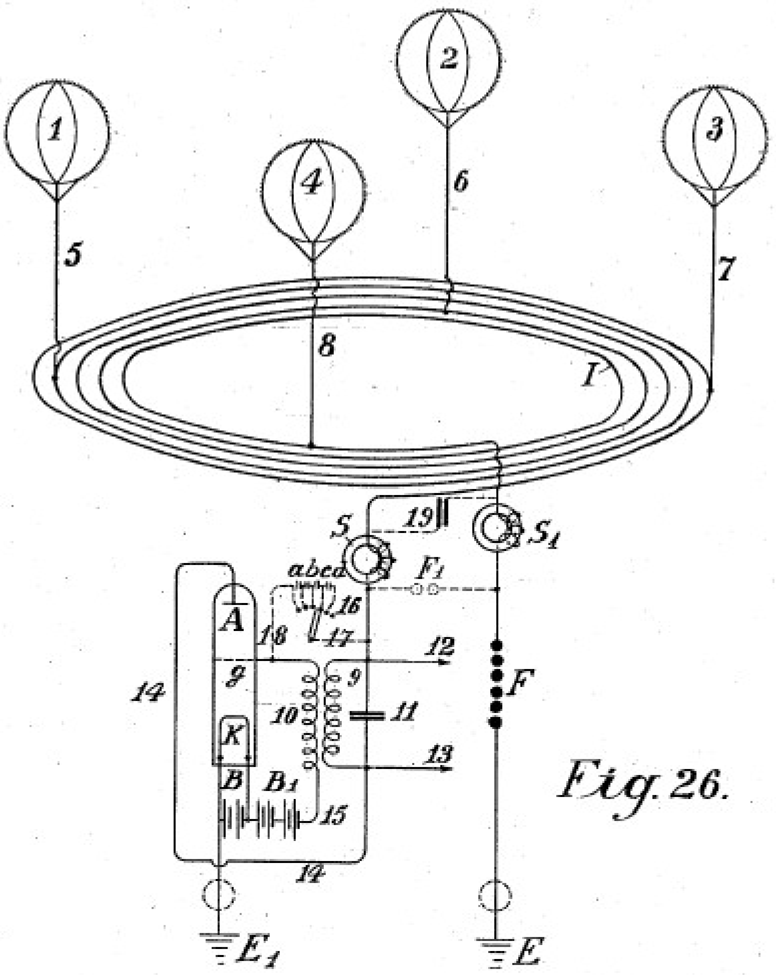
I
S
S1
abcd
F1
A
g
K
F
B
B1
E1
E

Fig. 26.

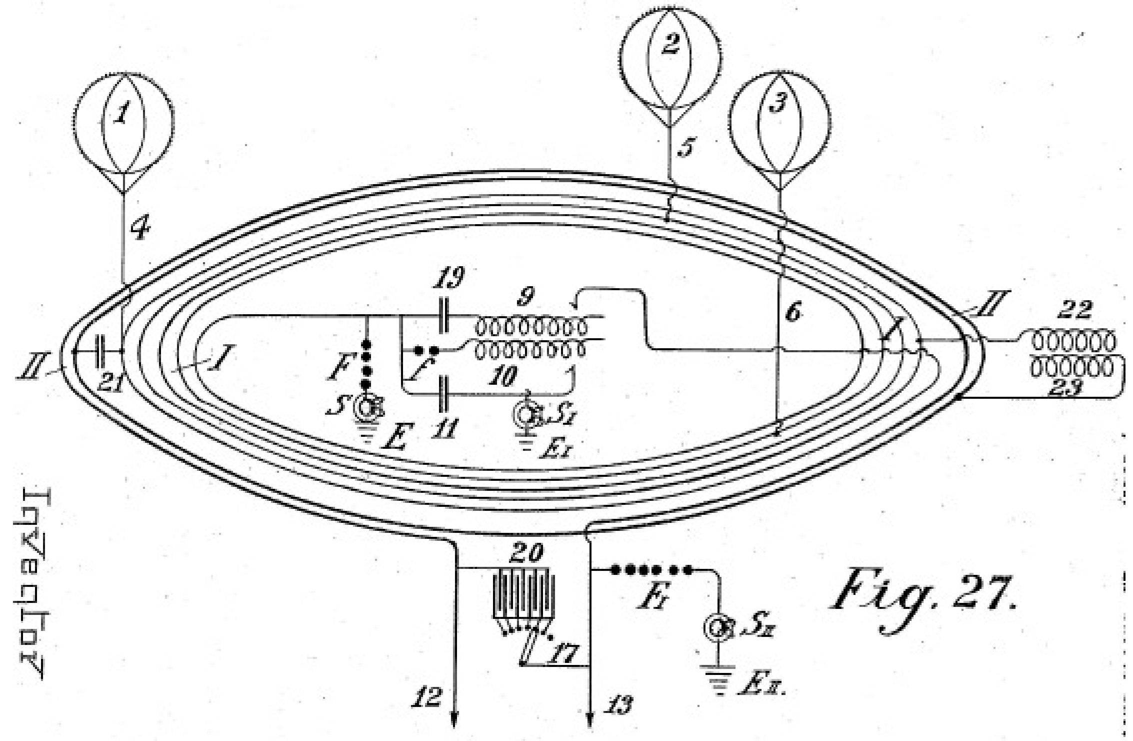
1
2
3
5
4
6
II
21
I
19
9
F
f
10
S
E
11
S1
E1
II
22
23
20
F1
SII
17
EII.
12
13
Inventor

Fig. 27.

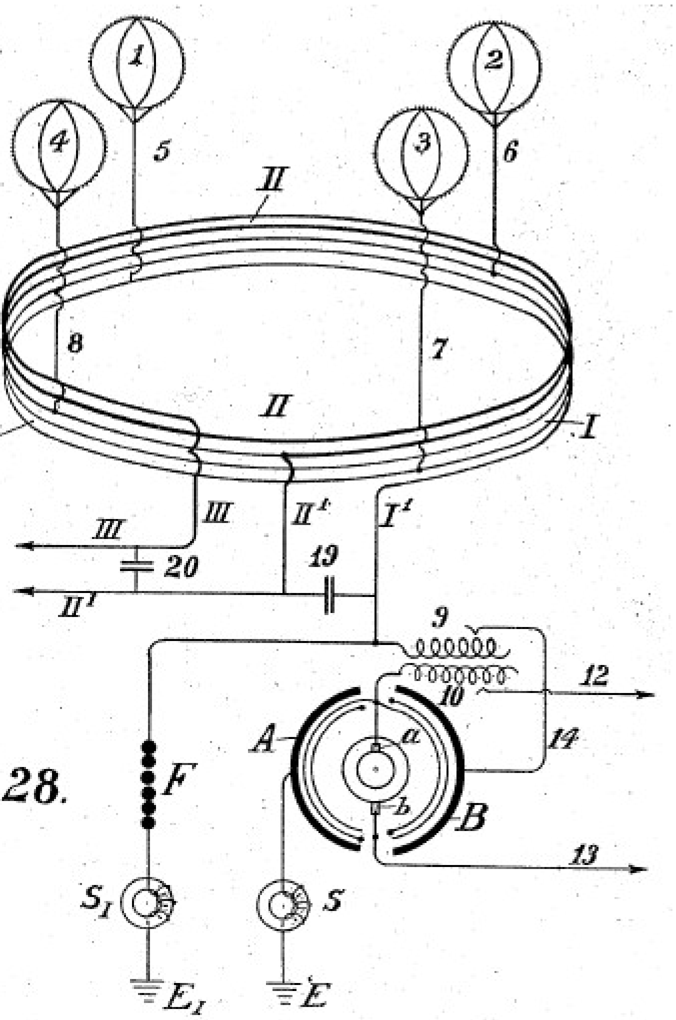
1
2
3
4
5
6
II
8
7
I
II
I
III
II'
I'
III
20
19
II'
9
12
10
A
a
14
F
b
B
13
S1
S
E1
E

Fig. 28.

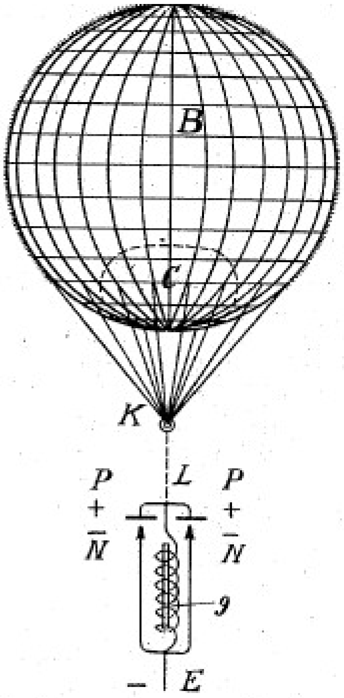
B
C
K
P
L
P
+
+
N
N
9
–
E

Fig. 29.

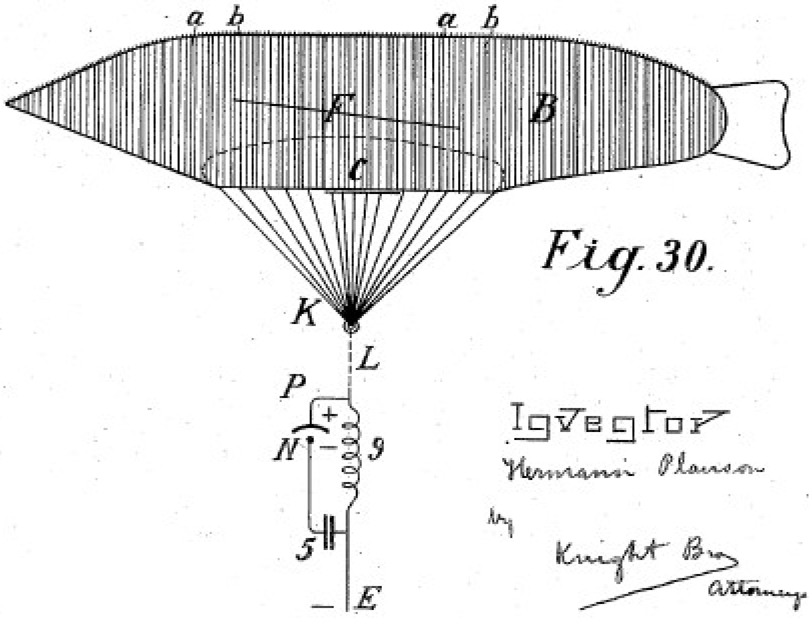
a
b
a
b
F
B
C
K
L
P
+
N
–
9
5
–
E

Fig. 30.

Inventor
Hermann Plauson
by
Knight Bros
Attorneys

Filed Jan. 13, 1921 12 Sheets-Sheet 11

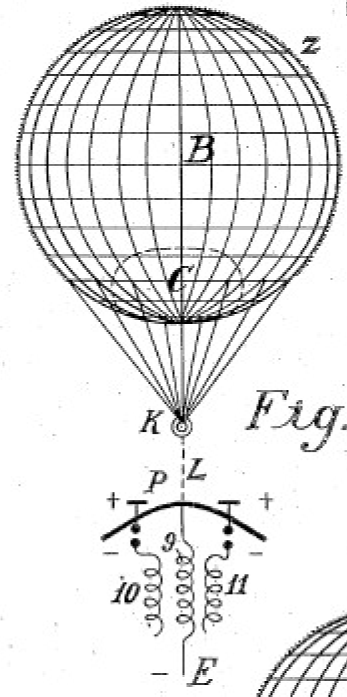
z
B
C
K
P
L
+
+
−
−
9
10
11
−
E

Fig. 31.

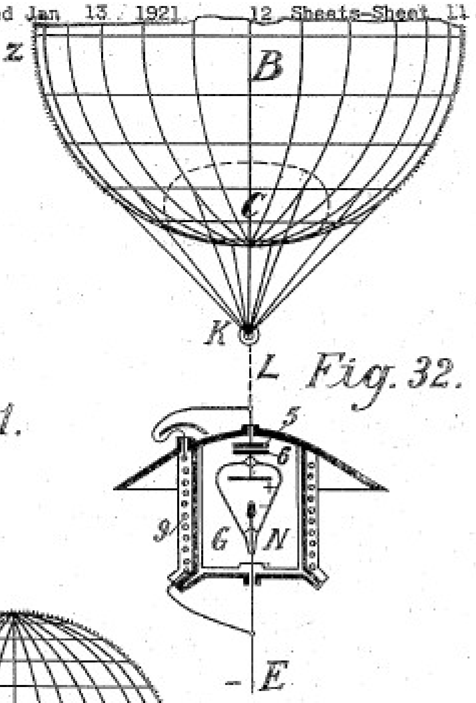
z
B
C
K
L
5
6
+
9
G
N
−
E

Fig. 32.

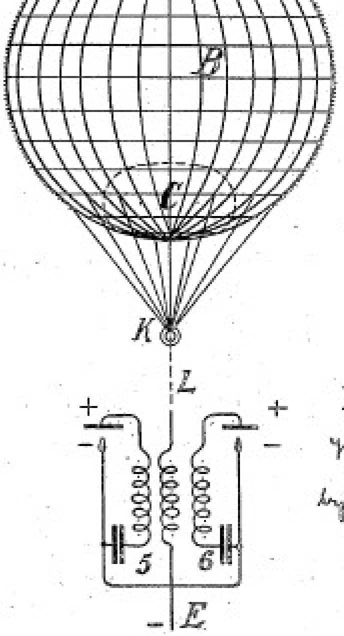

Fig. 33.

Inventor
Hermann Plauson
by Knight Bros
attorneys

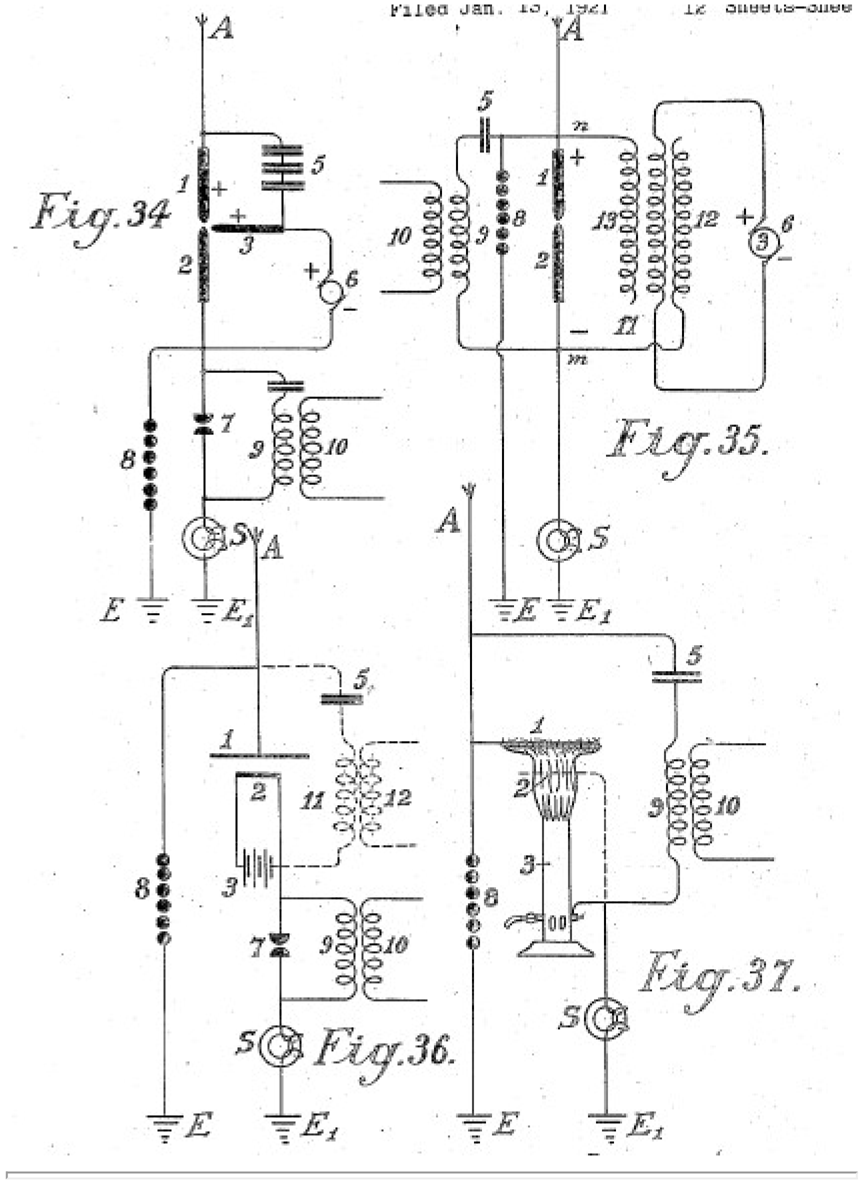
Filed Jan. 13, 1921 12 Sheets-Shee
Fig. 34
Fig. 35.
Fig. 36.
Fig. 37.
A
1
2
3
5
6
7
8
9
10
11
12
13
S
E
E_1
+
−
n
m

GB 149,917
Improvements in & Relating to Radio Communication Systems

Abstract -- Radio transmission and reception through the use of living vegetable organisms such as trees, plants, and the like.

Fig 1

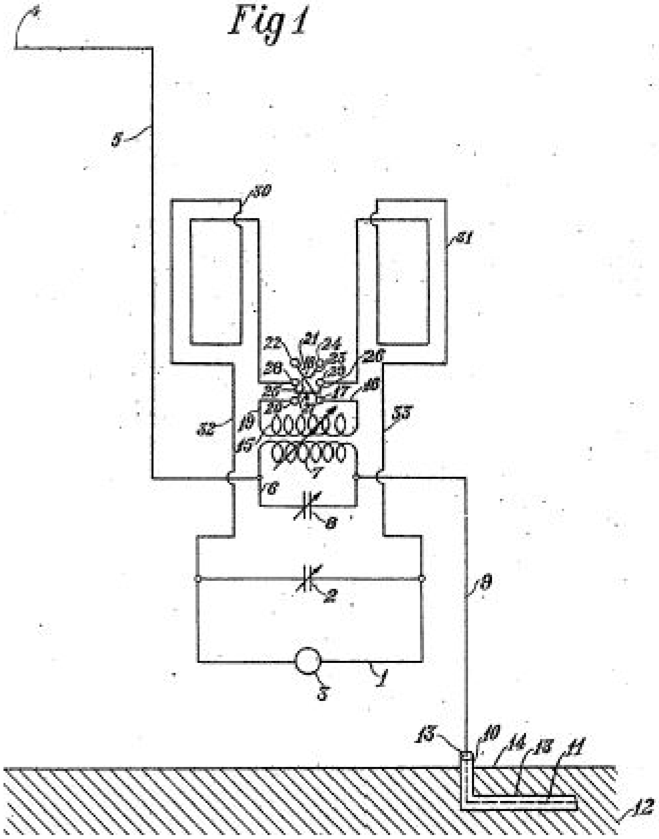
4
5
30
31
22
21
24
18
25
26
28
29
25
17
16
20
19
32
33
15
7
6
8
2
9
3
1
13
10
14
13
11
12

US 1826727
Radio Apparatus

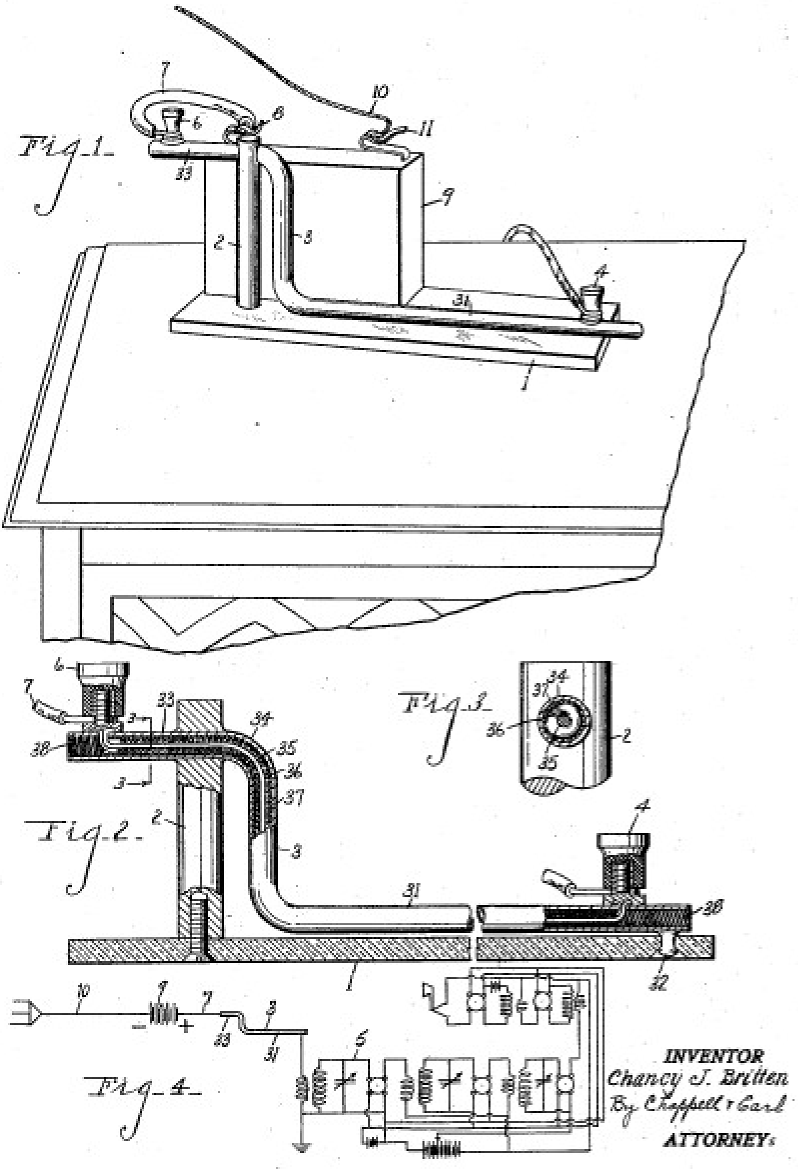
Fig. 1.
7
6
8
10
11
33
9
2
3
4
31
1
Fig. 2.
6
7
38
3
33
34
35
36
37
2
3
31
4
28
1
32
Fig. 3.
37
34
36
35
2
Fig. 4.
10
9
7
3
33
31
5
INVENTOR
Chancy J. Britten
By Chappell & Earl
ATTORNEYS

US 2143437
Radiant Energy System

Abstract -- An apparatus for receiving and transmitting radiant energy within a wide range of frequencies without necessity of a source of local energy.

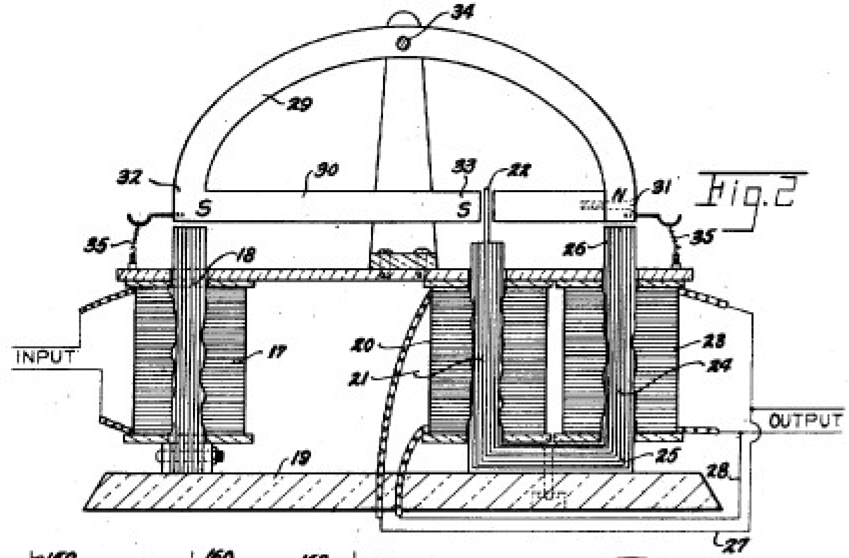
34
29
32
30
33
22
31
S
S
N
35
18
26
35
INPUT
17
20
21
23
24
OUTPUT
25
19
28
27

Fig.2

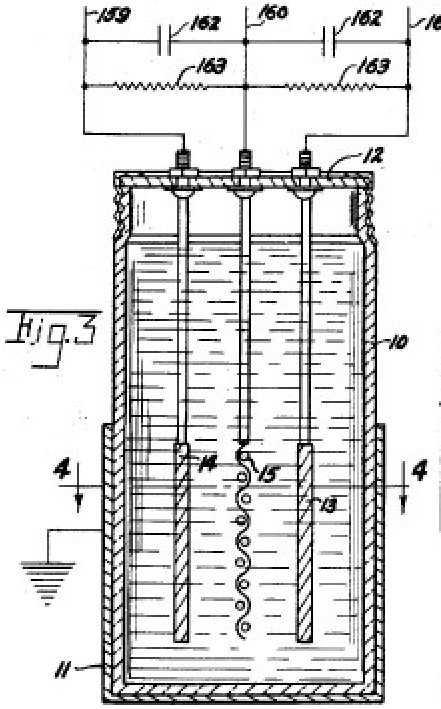
159
162
160
162
161
163
163
12
10
14
15
13
4
4
11

Fig.3

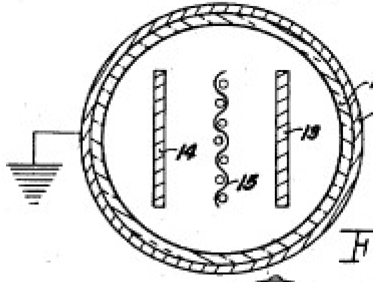
10
11
14
15
13

Fig.4

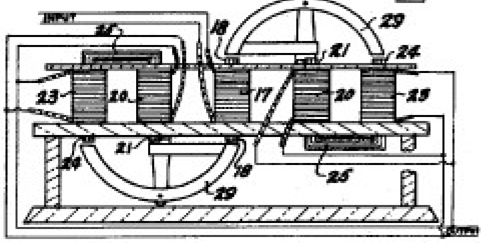
INPUT
22
18
29
21
24
23
20
17
20
23
21
24
18
29
25

Fig.9

Inventor

James Marion Fisher

By Murray & Zugelter Attorney

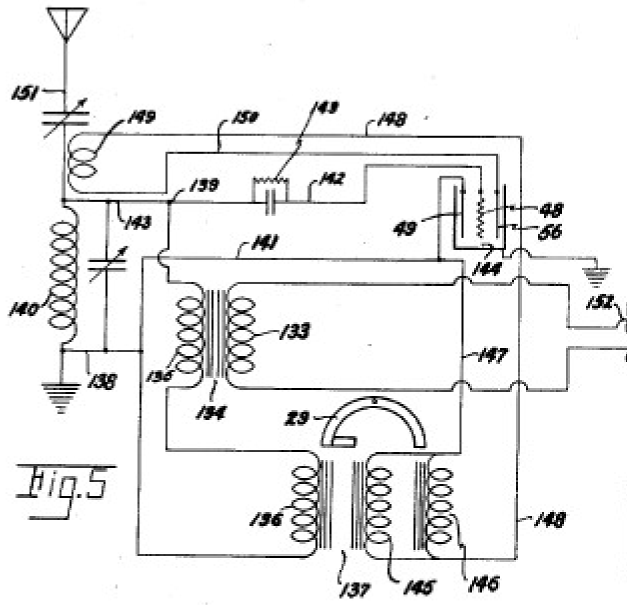
151
149
150
143
148
139
142
143
141
49
48
56
144
140
133
136
147
138
134
29
136
137
145
146
148

Fig.5

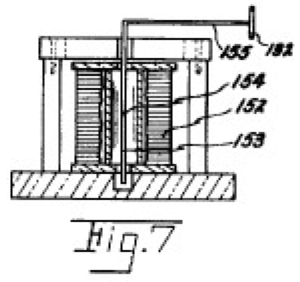
155
132
154
152
153

Fig.7

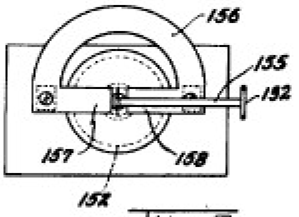
156
155
132
157
158
152
152
131
153
132

Fig.8

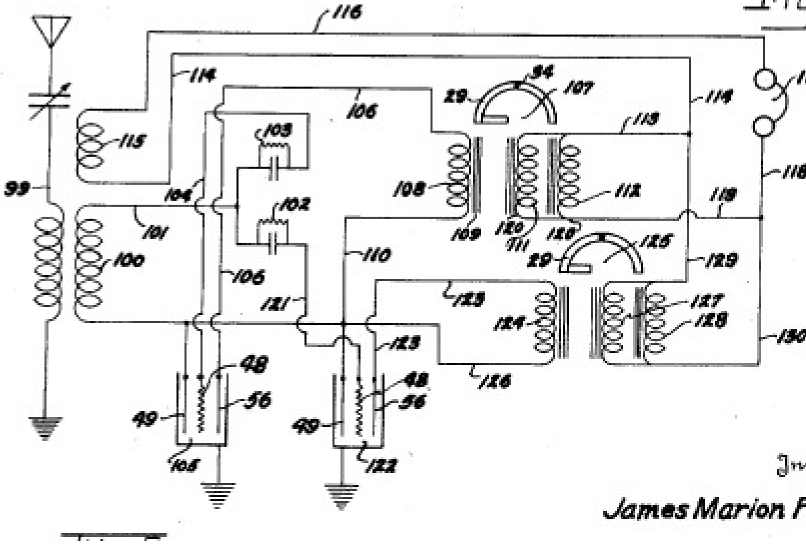
116
114
117
115
34
29
107
106
113
114
118
99
104
103
108
112
119
101
102
109
120
111
126
129
100
106
110
29
121
123
124
127
128
130
48
56
49
125
48
56
126
105
49
122

Inventor

James Marion Fisher

US 2813242

Powering Electrical Devices with Energy Abstracted from the Atmosphere

Abstract -- A convenient and economical provision of power for the operation of electronic circuits and devices using transistors, and of other electrical devices hving modest power requirements.

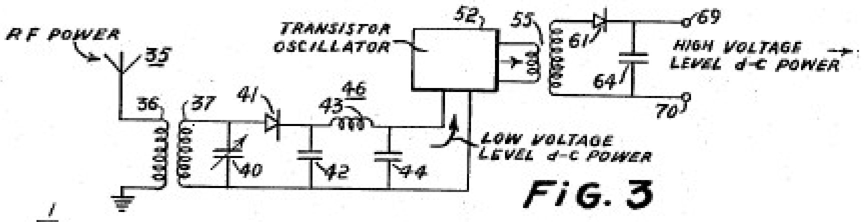
RF POWER
35
36
37
41
46
43
40
42
44
TRANSISTOR OSCILLATOR
52
55
61
64
69
70
HIGH VOLTAGE LEVEL d-C POWER
LOW VOLTAGE LEVEL d-C POWER

FiG. 3

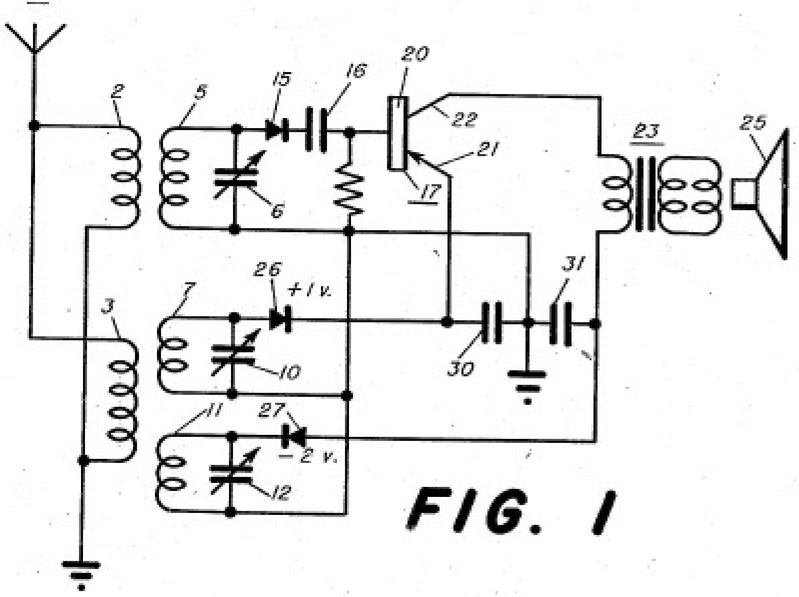
1
2
3
5
6
7
10
11
12
15
16
17
20
21
22
23
25
26
+1 v.
27
-2 v.
30
31

FIG. 1

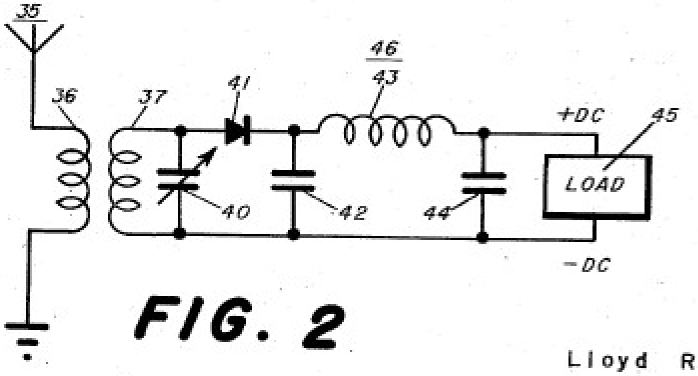
35
36
37
41
46
43
40
42
44
+DC
45
LOAD
-DC

FIG. 2

INVENTOR
Lloyd R. Crump

US 3205381
Ionospheric Battery

Abstract -- A method and apparatus for tapping the enormous reservoir of energy existing in the ionosphere...

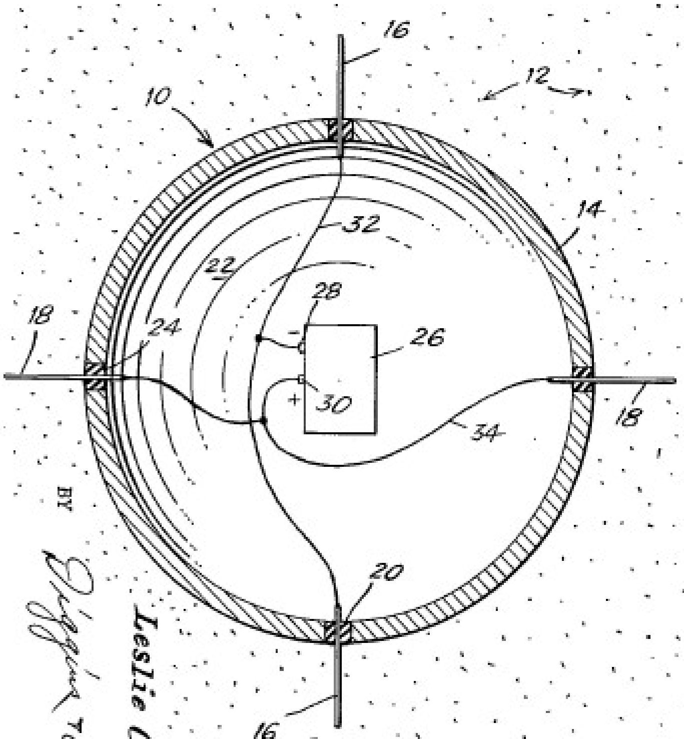
16
10
12
14
32
22
28
26
18
24
30
34
18
20
16

Fig. 1

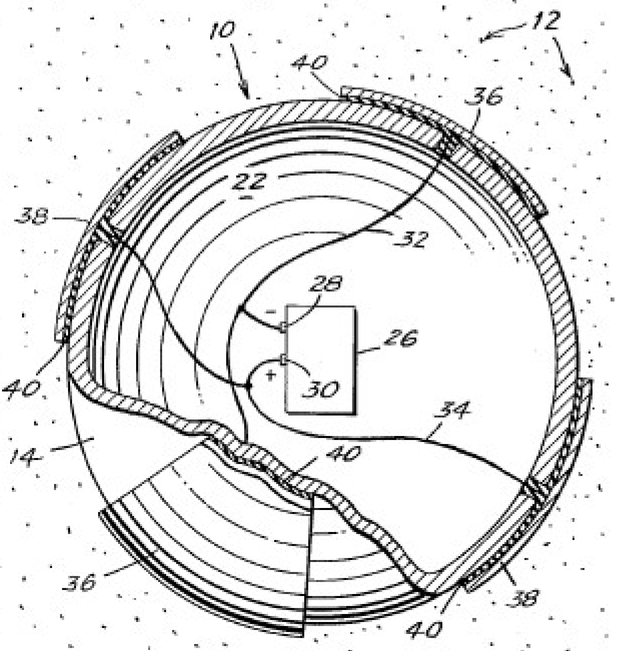
10
40
12
36
22
38
32
28
26
40
30
34
14
40
36
38
40

Fig. 2

INVENTOR
Leslie G. Smith,
BY
Higgins & Le Blanc
ATTORNEYS

US 3273066
Apparatus for Detecting Changes in the Atmospheric Electric Field

Abstract -- A cloud detector and or an apparatus for sensing electrostatic conditions in the earth's atmosphere...

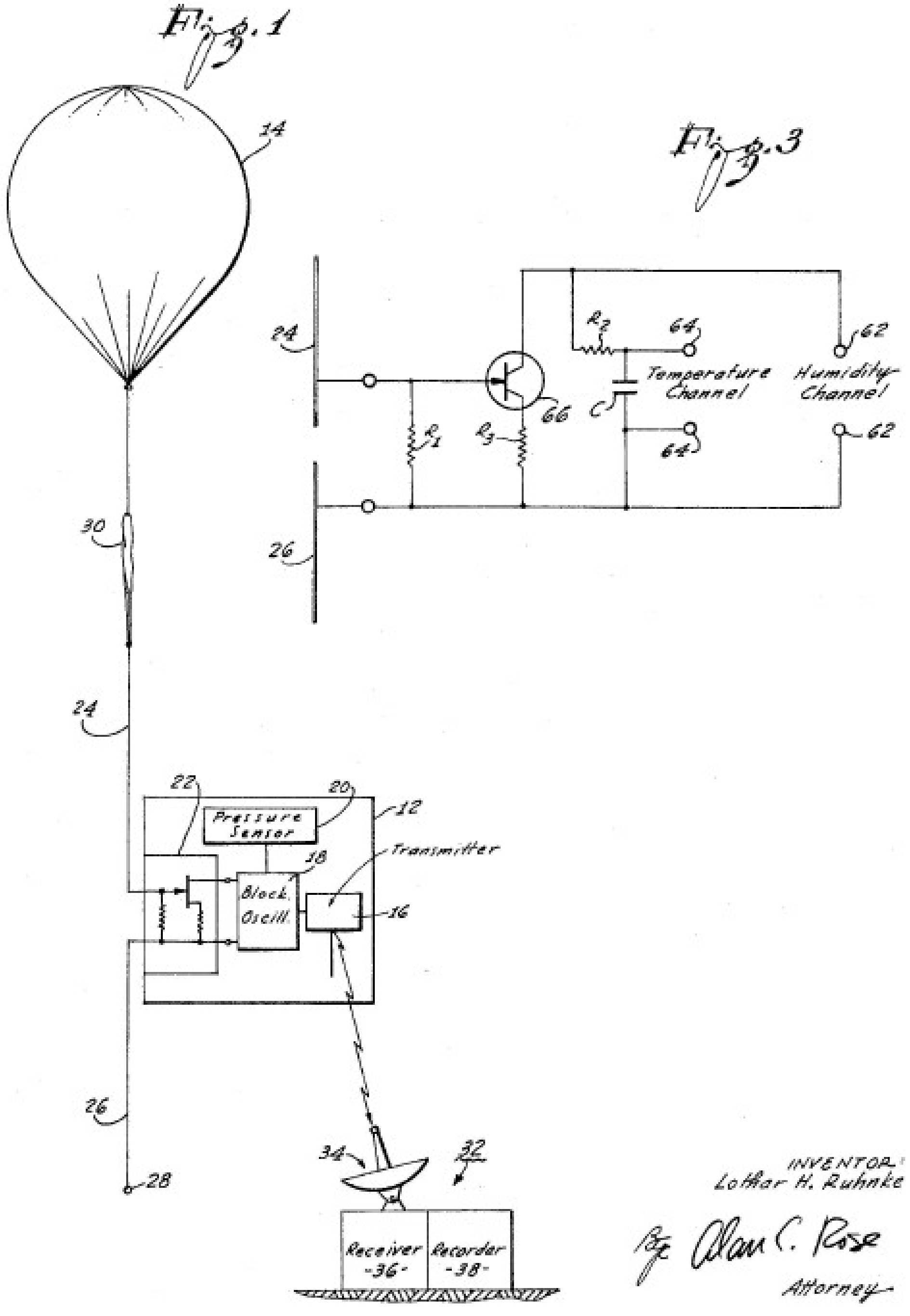
Fig. 1
14
30
24
22
20
12
Pressure Sensor
18
Transmitter
Block. Oscill.
16
26
28
34
32
Receiver -36-
Recorder -38-
Fig. 3
24
26
R1
R2
R3
66
C
64
64
62
62
Temperature Channel
Humidity Channel
INVENTOR:
Lothar H. Ruhnke
By Alan C. Rose
Attorney

US 3760257
EM Wave Energy Converter

Abstract -- Electromagnetic wave energy is converted into electric power with an array of mutually insulated electromagnetic wave absorber elements each responsive to an electric field component of the wave as it impinges thereon. Each element includes a portion tapered in the direction of wave propagation to provide a relatively wideband response spectrum. Each element includes an output for deriving a voltage replica of the electric field variations intercepted by it. Adjacent elements are positioned relative to each other so that an electric field subsists between adjacent elements in response to the impinging wave. The electric field results in a voltage difference between adjacent elements that is fed to a rectifier to derive d.c. output power. The element pairs may be arranged in a two-dimensional array to provide

power conversion of randomly polarized electromagnetic waves, such as sunlight.

FIG. 1

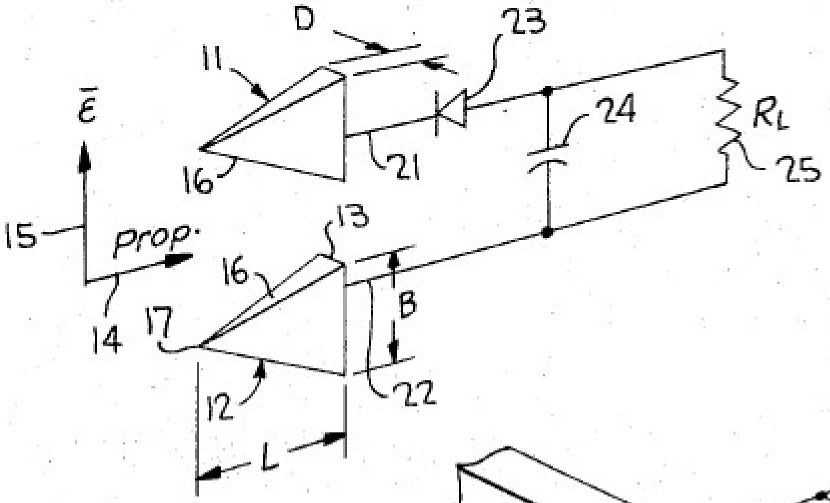
D
23
11
Ē
24
R_L
25
16
21
15
Prop.
13
16
B
14
17
12
22
L

FIG. 2

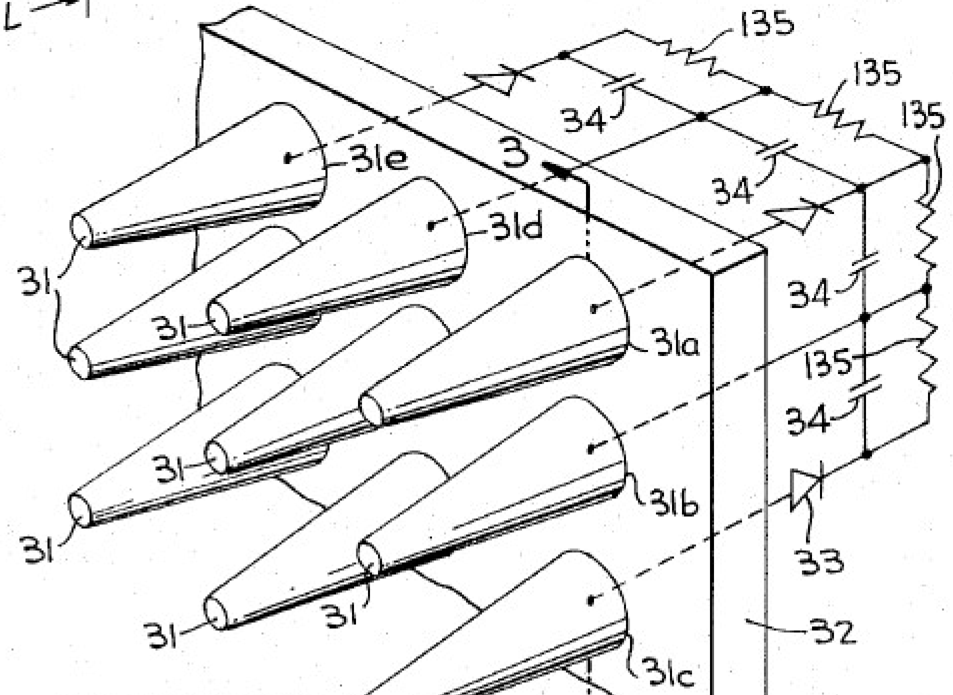
135
135
135
34
34
3
31e
31d
31
31
31
31a
34
135
34
31
31
31b
33
31
31
31
32
31c
31
3

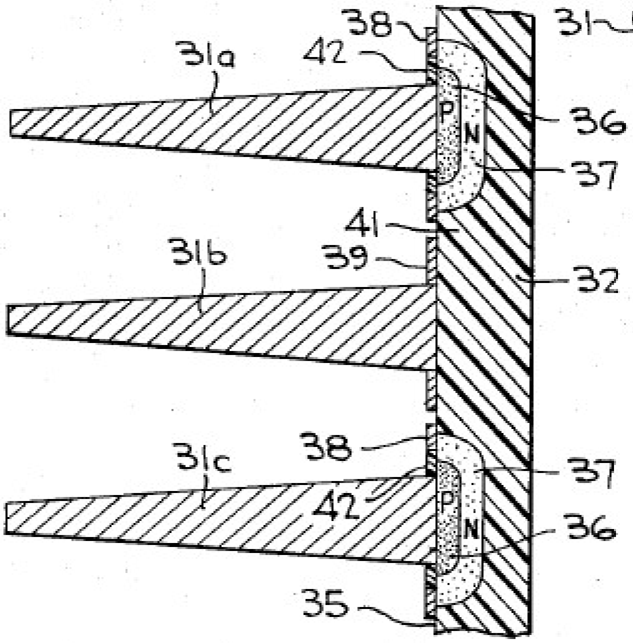
38
42
31a
P
N
36
37
41
39
31b
32
38
31c
37
42
P
N
36
35

FIG. 3

US 3931577
Radio Receiver Protection Arrangement

Abstract -- A receiver protection arrangement in which a voltage dependent on aerial voltage and a voltage dependent on aerial current are aggregated and applied to means responsive to a predetermined aggregate level to effectively short circuit the aerial lead. Additional means may be provided to dissipate energy in the input stage of a protected receiver to cause a reduced impedance to be reflected across the output of the protective arrangement.

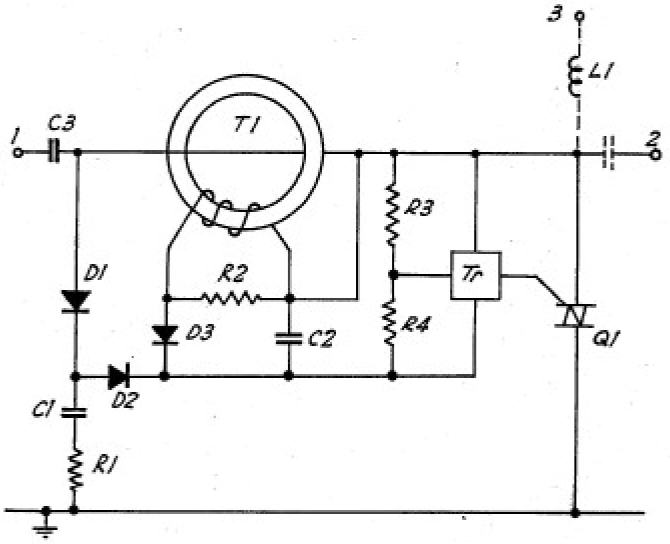
3
L1
1
C3
T1
2
R3
D1
R2
Tr
D3
C2
R4
Q1
C1
D2
R1

Fig. 1.

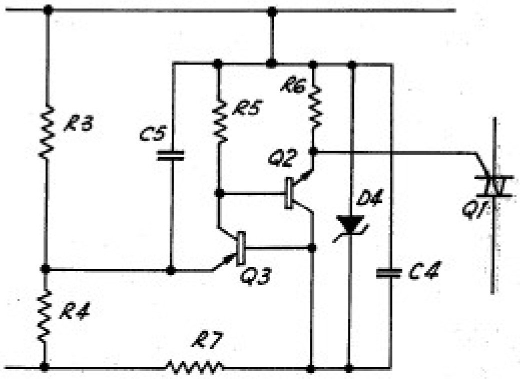
R6
R3
R5
C5
Q2
D4
Q1
Q3
C4
R4
R7

Fig. 2.

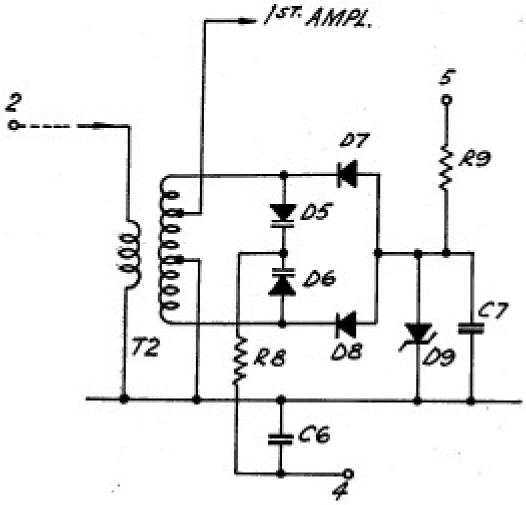
1ST. AMPL.
5
2
D7
R9
D5
D6
C7
T2
R8
D8
D9
C6
4

Fig. 3.

US 4,628,299
Seismic Warning System Using RF Energy Monitor

Abstract -- The ambient broadband radio frequency field strength from broadcast stations is monitored (Figure 4) by periodic sampling (50, 52). A warning indication is provided if the field strength drops significantly. Drops in such field strength have been correlated empirically with the occurrence of seismic activity, usually several days later. Thus the indication serves as an early warning of an impending earthquake. In one preferred embodiment, a broadband, horizontal, very long monopole antenna (40) was connected to a rectifying and smoothing circuit (Figure 3) to provide a dc output proportional to the ambient rf field...

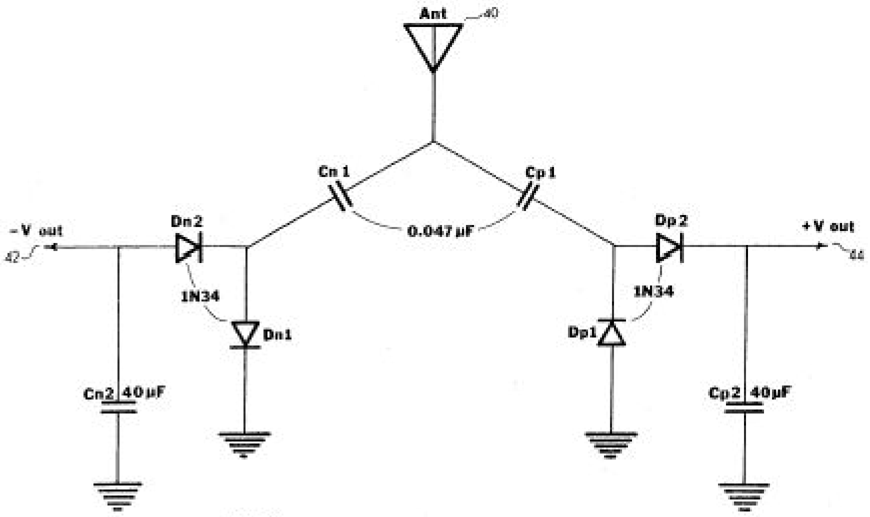

FIG 3

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

Nov. 12, 1957

L. R. CRUMP

POWERING ELECTRICAL DEVICES WITH ENERGY ABSTRACTED FROM THE ATMOSPHERE

Filed March 12, 1954

2,813,242

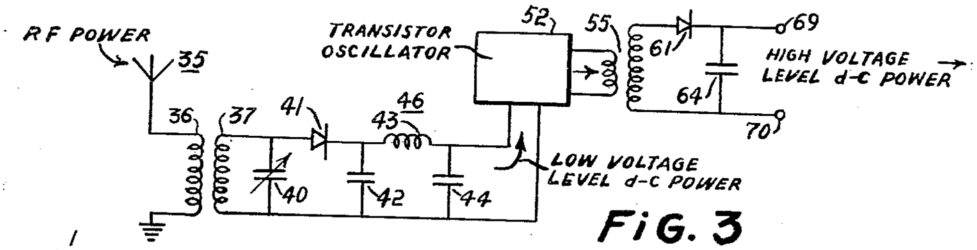
RF POWER
35
36
37
41
46
43
40
42
44
TRANSISTOR OSCILLATOR
52
55
61
64
69
70
HIGH VOLTAGE LEVEL d-c POWER
LOW VOLTAGE LEVEL d-c POWER

FIG. 3

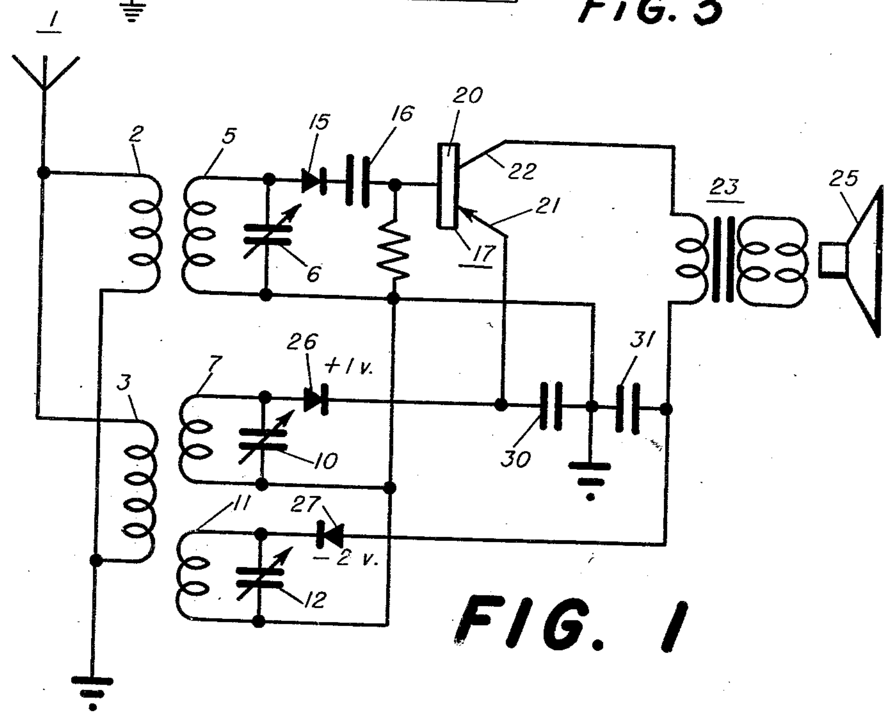
1
2
5
15
16
20
22
21
17
6
23
25
3
7
26
+1 v.
10
30
31
11
27
−2 v.
12

FIG. 1

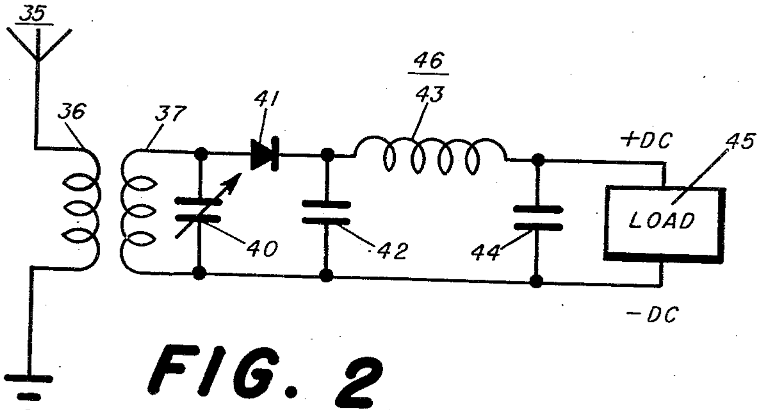
35
36
37
41
46
43
40
42
44
+DC
45
LOAD
−DC

FIG. 2

INVENTOR

Lloyd R. Crump

BY W.E. Thibodeau + A.W. Dew

ATTORNEYS

United States Patent Office 2,813,242
Patented Nov. 12, 1957

2,813,242

POWERING ELECTRICAL DEVICES WITH ENERGY ABSTRACTED FROM THE ATMOSPHERE

Lloyd R. Crump, Silver Spring, Md.

Application March 12, 1954, Serial No. 415,986
1 Claim. (Cl. 321—2)
(Granted under Title 35, U. S. Code (1952), sec. 266)

The invention described herein may be manufactured and used by or for the Government for governmental purposes without payment to me of any royalty thereon.

This invention relates to the convenient and economical provision of power for the operation of electronic circuits and devices using transistors, and of other electrical devices having modest power requirements.

A great advantage of transistors, and a major reason for their enthusiastic reception since their introduction a few years ago, is the fact that they will operate satisfactorily with very low supply voltages and currents. One milliwatt or even less is sufficient to power a transistor in many applications. Various batteries have been developed to provide, in a minimum of space, the relatively minute amounts of power needed by transistors.

My invention provides methods and means that permit transistor circuits, and also other low-powered electrical devices, to be economically and conveniently operated without any batteries whatever, and indeed without any power supply whatever as power supplies are ordinarily conceived.

The invention centers around my discovery that it is practicable to construct operative transistor circuits that are able to abstract from the atmosphere sufficient electromagnetic energy to provide all necessary supply voltages and currents for their own operation. Circuits and devices powered according to my invention will operate indefinitely without any local power source whatever.

I have successfully constructed and demonstrated such circuits. For example, I have constructed a batteryless transistor radio receiver on which I have listened to either nearby or distant broadcast stations as desired, using either headphones or a loudspeaker; this receiver has been powered entirely by electromagnetic energy abstracted from the atmosphere.

From the successful operation of this receiver, and from other experimental work, it becomes clear that, by the methods and means of the invention, a great variety of practical and useful transistor circuits can be powered entirely by energy abstracted from the atmosphere.

Furthermore, as will become apparent below, my invention is applicable to the powering of other electrical devices requiring relatively small amounts of power.

An object of the present invention is to provide methods and means for powering transistor circuits entirely from radiofrequency energy abstracted from the atmosphere.

Another object is to provide methods and means for powering remote radio receivers, low-powered radio transmitters, and other low-powered electrical devices, with energy received by radio from a master station, so that no local power supplies are needed by the devices and so that the powering or non-powering of the remote device is under the control of the master station.

A further object is to provide methods and means for powering transistor circuits and other low-powered electrical devices with radiofrequency energy received from one or more remote radio transmitters.

Other objects, aspects, uses, and advantages of the invention will become apparent from the following description and from the drawing.

Figure 1 is a schematic diagram of a transistor radio receiver in which all necessary power is supplied by energy abstracted from the atmosphere in accordance with the invention.

Figure 2 is a schematic diagram showing a general application of the invention to provide direct-current power to a load.

Figure 3 is a schematic diagram of a system for obtaining a high energy D.-C. source at a high voltage level using energy abstracted from the atmosphere.

Referring to Figure 1, a receiving antenna **1** is connected to antenna coupling coils **2** and **3**, the other ends of which are connected to ground. A parallel resonant circuit consisting of coil **5** and capacitor **6** is coupled to coil **2**. A second parallel resonant circuit consisting of coil **7** and capacitor **10** is coupled to coil **3**. A third parallel resonant circuit consisting of coil **11** and capacitor **12** is also coupled to coil **3**.

Coil **5** and capacitor **6** are tuned to the frequency of a radio transmitter from which it is desired to receive information—for instance, an amplitude-modulated standard broadcast station. The signal received from this transmitter need not be strong. The signal is detected by diode **15** to obtain an audio-frequency information signal. This audio signal is coupled through a capacitor **16** and is amplified by a circuit that includes a transistor **17** having a base **20**, an emitter **21**, and a collector **22**. The amplified audio output of the transistor is coupled through an audio transformer **23** to an electroacoustical transducer, preferably a permanent-magnet dynamic loudspeaker **25** as shown.

The novelty of the invention lies largely in the method and means by which the necessary direct-current power is supplied to the emitter and collector circuits of transistor **17**. This method and means will now be described.

Coil **7** and capacitor **10**, and also coil **11** and capacitor **12**, are tuned to receive radio signals of relatively high strength. It does not matter whether these signals contain information. These power signals are rectified by diodes **26** and **27** to provide direct-current power that is filtered by capacitors **30** and **31**. The D.-C. power thus obtained is utilized to power the transistor **17**.

In the circuit shown, two tuned circuits (coil **7** and capacitor **10**, and coil **11** and capacitor **12**) are tuned to power signals and the D.-C. voltages obtained from each are connected in series. The tuned power circuits may be tuned to the same or different power signals. Under certain circumstances it may be desirable to use more than two tuned power circuits and to tune them to more than two power signals; in this way power can be obtained from several signals and combined. On the other hand, if a strong power signal is available, a single tuned power circuit may suffice to give the needed D.-C. power.

Even weak information signals can be received successfully. A plurality of transistor amplifier stages can be used if desired, or other circuits such as superheterodyne circuits can be used. It is merely necessary that a sufficiently strong power signal or signals be available to provide the small amount of D.-C. needed to power the transistors.

If the information signal happens to be strong, it can be used as the power signal; all of the tuned circuits (coil **5** and capacitor **6**, coil **7** and capacitor **10**, coil **11** and capacitor **12**) are tuned to the information signal.

Engineers who have observed my invention in operation have been surprised at the unexpectedly good results obtained, even with readily available power signals of quite moderate strength. For instance, sufficient power for satisfactory operation of a loudspeaker at low volumes is

readily obtained from a 5-kilowatt standard broadcast station 5 miles away, using only an indoor antenna to pick up the power signal as well as information signals. In typical operation under these conditions a D.-C. voltage of about 2.5 to 3 volts is obtained between the emitter and the collector, at a current of about 250 microamperes; D.-C. power input to the transistor is thus of the order of 0.5 to 1 milliwatt. So far as I am aware, no one has ever before discovered and demonstrated the practicability of this method of powering a radio receiver.

Because existing broadcast stations within a radius of a number of miles provide adequate power signals, the invention is readily practicable with existing power signals in almost any location in or near any city in the United States.

Although I have described a transistor radio receiver powered by my invention, it will be readily apparent that the invention is applicable to the powering of any transistor circuit using one or a number of transistors, and to the powering of other devices requiring relatively small amounts of power. For instance, sensitive electromechanical, electrochemical, or electrothermal devices can be operated by the method of the invention.

Referring to Figure 2, which shows a more general embodiment of my invention, an antenna **35** picks up radiofrequency energy from the atmosphere. This energy flows through coil **36**, which is coupled to a tuned circuit consisting of coil **37** and capacitor **40**. The radiofrequency voltage across capacitor **40** is rectified by diode **41** and filtered by a low-pass filter **46** consisting of capacitors **42** and **44** and choke coil **43**. The resulting D.-C. voltage is applied to a load **45**.

In the practice of my invention, larger amounts of power can be obtained for short periods of time by storing received energy in a suitable energy storage device. Stored energy may then be withdrawn at intervals at a more rapid rate than that at which it was received and put into the storage device. In this way the invention can be used to provide short pulses of relatively very high electrical energy. This result can be readily obtained by charging a relatively large capacitor with direct current and then discharging the capacitor rapidly into a load when desired. This rapid discharge can be initiated automatically when the voltage across the capacitor reaches a certain level, or it can be initiated when a transistor radio receiver receives a certain information signal.

Higher voltages can be obtained with the invention by means of well known devices for raising D.-C. voltages as shown in Figure 3. The D.-C. voltage output from the capacitor **44** can be used to power a low frequency transistor oscillator **52** whose A.-C. output is raised to a higher voltage level by the transformer **55**. This relatively high A.-C. voltage can then be rectified by a diode **61** and fed to a capacitor **64** to provide a high energy D.-C. source at a relatively high voltage level at the terminals **69** and **70**. If desired, energy can now be withdrawn from the capacitor **64** at intervals in short pulses of high energy at a high voltage level. Pulsed radio transmission is one of the possible uses for this form of the invention. Other uses would be to provide a single relatively powerful pulse needed to actuate an electrothermal or electromechanical device.

As has been indicated above, in many locations and particularly anywhere in or near most American cities, power signals normally present in the atmosphere are readily available for the easy and convenient practice of the invention. However, the invention also has important applications in systems in which the necessary power signal is generated and transmitted specifically for the operation of the particular system. Such systems can, for example, comprise a master station transmitting all the power that is needed for hundreds or thousands of fixed or mobile transistor receivers or other remote devices over a range of many miles. This eliminates the need for hundreds or thousands, as the case may be, of local power supplies. At the same time, such a system has the advantage that all of the remote devices can be simultaneously activated or deactivated at the will of the master station, simply by starting or stopping the transmission of the power signal. In such systems it will often be advantageous to use power signals of frequencies sufficiently high to permit the use of resonant receiving antennas of small physical dimensions for signal pickup at the remote devices. In addition to the power signal, the master station may transmit an information signal on the same or a different carrier.

Certain types of devices powered entirely by received radio waves are of course well known. The well-known "crystal set" of the early days of radio, which used a diode rectifier to demodulate an amplitude-modulated radiofrequency signal, is an outstanding example of such a device. My invention is readily distinguishable from such prior devices, however. In typical prior devices a modulated radiofrequency signal is applied to a diode to obtain unidirectional half-wave pulses whose amplitudes vary with modulation. These pulses are integrated by means of a capacitor to obtain a unidirectional signal the amplitude of which follows the audiofrequency modulation envelope. If the radiofrequency signal is received with sufficient strength the audio signal may have sufficient power to operate headphones or similar utilization device without power amplification; but the signal is utilized for its information content, rather than to supply non-information-containing power.

My invention, on the other hand, entails the utilization of received radiofrequency energy to supply power to at least one pair of circuit points (across capacitor **31** in Fig. 1, for example), such circuit points requiring power solely for its power content and not for any information or modulation it may contain. In other words, my invention entails the utilization of radiofrequency energy to supply power that would otherwise have to be supplied by batteries, generator, or other local power source.

It will be apparent that the embodiments shown are only exemplary and that various modifications can be made in construction and arrangement within the scope of the invention as defined in the appended claim.

I claim:

An electrical device for obtaining a high energy D.-C. source at a high voltage level using energy abstracted from the atmosphere, said device comprising in combination: resonant means for receiving radio waves, first rectifier means for converting said radio waves into first direct current energy, first capacitor means for storing said first direct current energy, an oscillator powered by said direct current energy, said oscillator producing an A.-C. output, transformer means for raising said A.-C. output to an increased voltage level, second rectifier means for converting the A.-C. output of increased voltage level from said transformer into second direct current energy, and second capacitor means for storing said second direct current energy.

References Cited in the file of this patent

UNITED STATES PATENTS

1,540,998	Plauson	June 9, 1925
2,056,824	Cawley	Oct. 6, 1936
2,553,734	Adler	May 22, 1951
2,569,687	Pan	Oct. 2, 1951
2,647,957	Mallinckrodt	Aug. 4, 1953

OTHER REFERENCES

Turner: "A crystal receiver with transistor amplifier," Radio and Television News, January 1950.

Energy Projects

Electrodynamic Tether Photos

Deployment of the Tether Satellite

Boom Deploy

Initial Deploy

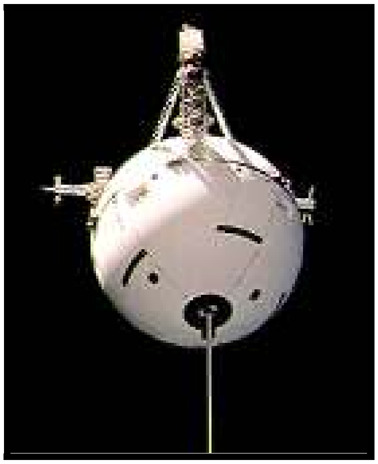

Satellite Deployed

Closeup of boom showing tether

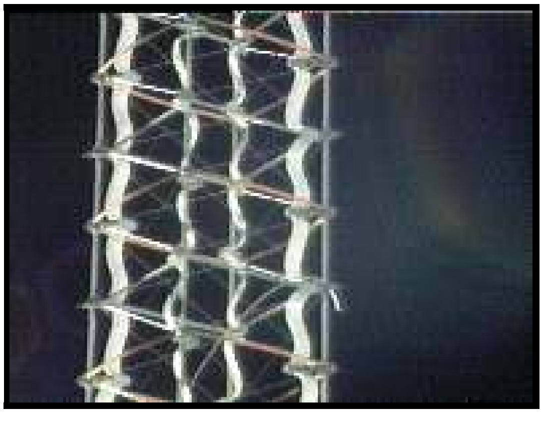

Close up of end of boom

Tether Break

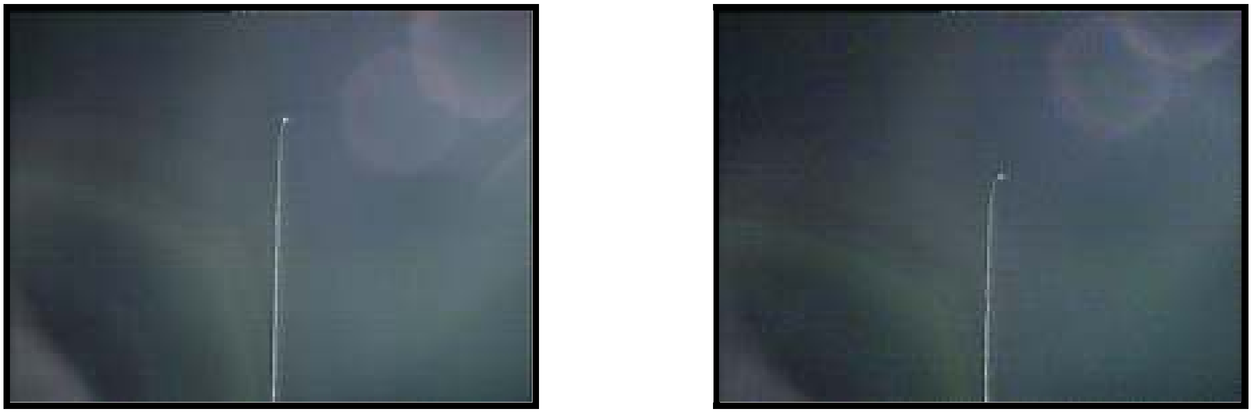

The tether deployed to one mile

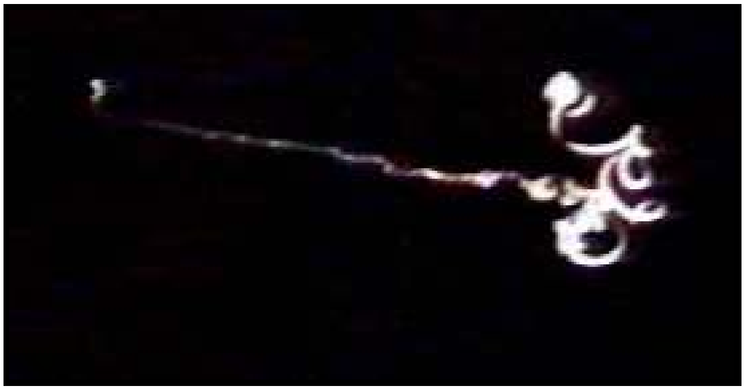

Shortly after the breakreak

The tether coils as it moves away from the Shuttle

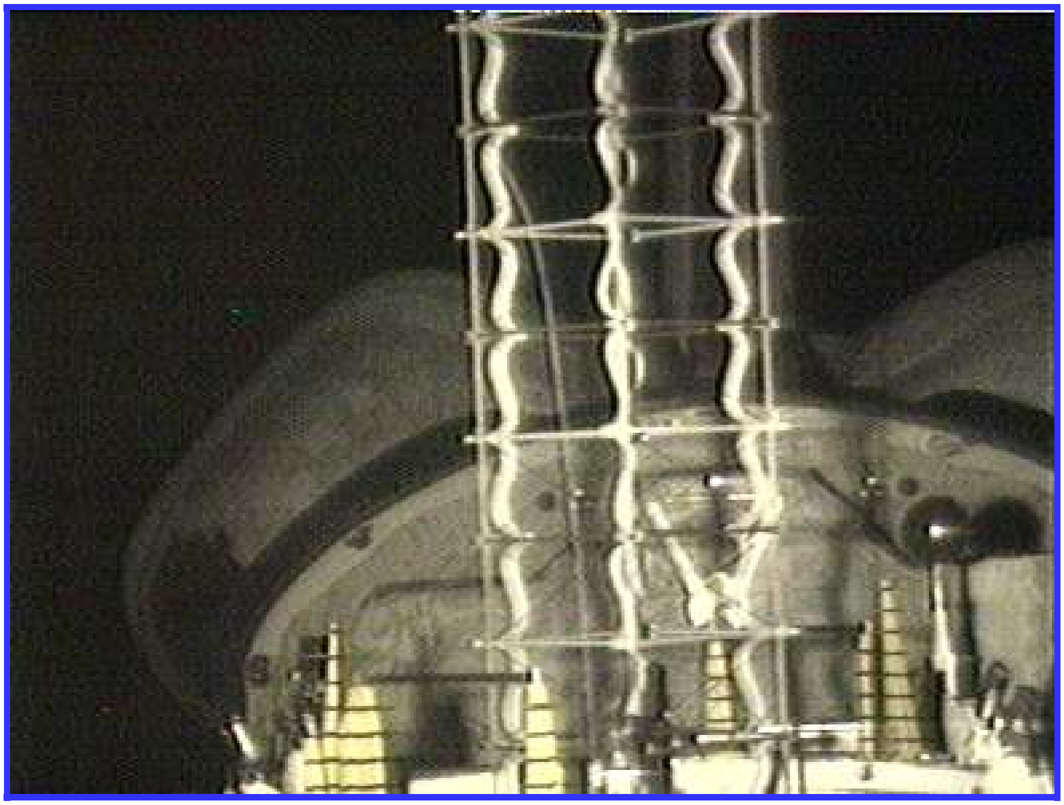

The base of the tether boom. The area looks more like a Tesla Laboratory than the space shuttle hanger bay.

After the Break

The following three images are those displayed for the public showing the white specks that appearall around the tether after it has broken free. NASA calls these "dust and debris" particles, though Mission control in the audio portion of the videos state that they see a "lot of stuff swimming around" and that the tether is wider than expected.

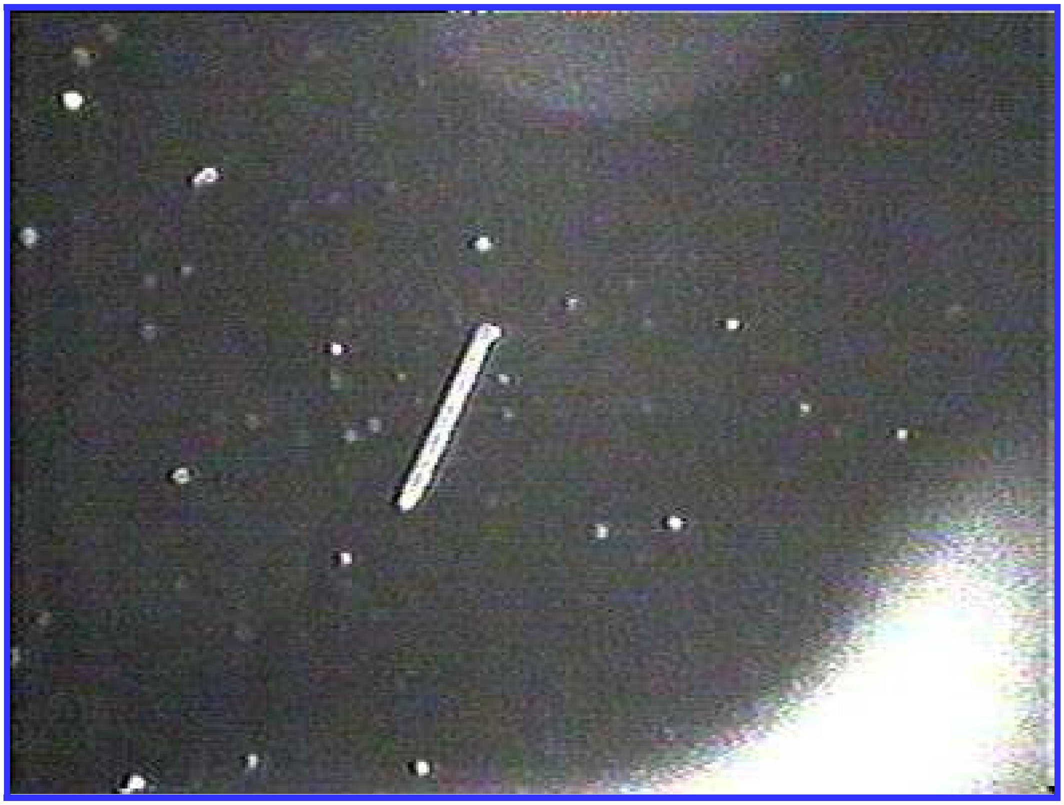

If you compare these images to the ones we captured from a hard to find high resolution copy of the video on this page you will see a remarkale difference. The video is a must see.

Video Clip
STS-75 'The Tether Incident' Feb 1996
Editors Note: This video has been removed several times by Youtube. Currently it is available again. If it does get removed again, email me and I can send you a copy as we have it on disk now. Contact me at webmaster
This one is from the hand held Infrared Camera and is not as clear as the other one was STS-75 Alternate Copy

Comments:

What is most interesting to note is that most of the video clips above are taken from this video, yet NASA does not show the video itself. The copy on Youtube is a transmission intercept copy. However the fact that the NASA info site DOES use those clips gives credibility to the video that it is indeed a NASA film - Zorgon

Tethered Satellite System (TSS-1R)
Source of Clips: NASA STS-75 Mission Pictures

Kennedy Space Center STS-75 Mission Fact Sheet

"Critters" from STS-75

The Report After the Fact

NASA/TP—2003-21228
Low Earth Orbit Spacecraft Charging Design Guidelines

Below are a few keynote excerpts from the PDF file. The entire 367 page document is available below for those with technically savvy who want to read the entire paper...

Excerpt 1:

In the case of the TSS-1R tether, its 20 km length produced a maximum of about 3500 V potential between its most positive and negative ends, since it wasn?t oriented perfectly perpendicular to the velocity vector and the Earth?s magnetic field. A satellite at its upper end collected electrons, and an electron gun at the lower end emitted electrons to complete the circuit. When the electron gun was not in operation, a large resistance prevented the Shuttle from being biased thousands of volts negative of its surrounding plasma. However, there remained a large voltage between the tether lower end and the Shuttle orbiter. This enormous bias eventually led to a continuous arc on the tether (see The Continuous Arc, section 4.2.3.1 below), which broke, freeing the satellite and ending the experiment. During the arc, the satellite collected over 1 Amp of electron current to keep the arc going. Probe theory (Cohen et al, 19870010625 N) is usually used to calculate the total current collected by a wire with distributed potentials. However, before the break, TSS-1R demonstrated that a satellite at a high positive potential could collect an anomalously large electron current. See Zhang, et al (20000110580), Stone and Raitt (19990084046 and 20000025437), and Stone, et al (19980202347)

Excerpt 2:

Sustained arcs (continuous arcs) - These are the events that have been attributed with the destruction of on-orbit solar arrays. Generally, the process begins with a fast transient (a so-called trigger arc). Under some conditions, the transient develops into an arc that is fed directly by the entire array, effectively becoming a short-circuit. Such events invariably involve large quantities of energy and can be severely damaging to cells, interconnects or power traces.

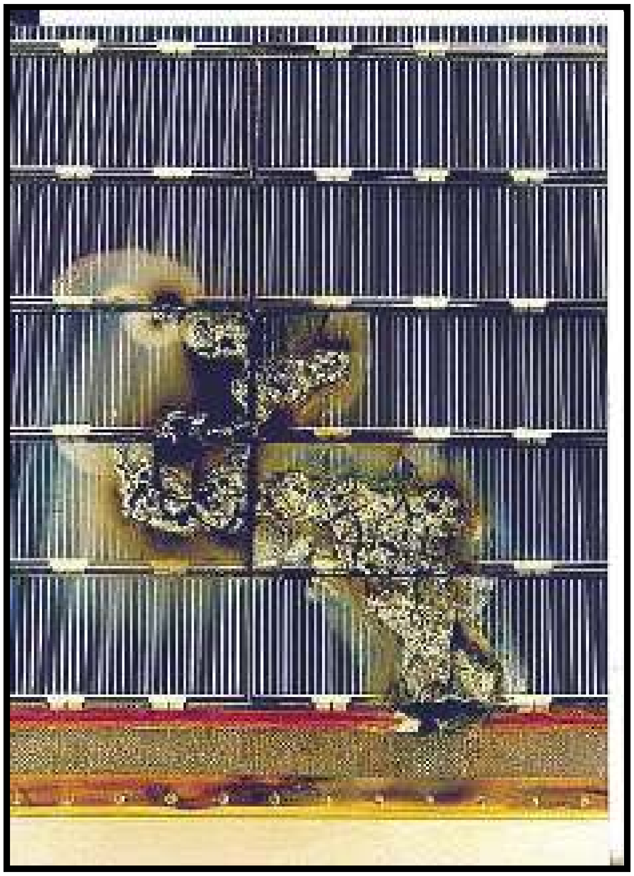

Excerpt 3:

When the structure or array capacitance electrically connected to the arc site is sufficiently large, the initial transient arcs themselves can be large enough to produce significant damage. In Figure 9, we see an anodized aluminum plate that has undergone repeated arcing in the laboratory with the ISS structure capacitance attached. Its thermal properties have been completely destroyed, along with most of the insulating surface layer of aluminum oxide.

Excerpt 4:

The most famous sustained arc event of all led to the breakage of the TSS-1R electrodynamic tether, and the loss of the attached satellite. The image below shows the burned, frayed and broken tether end still attached to the Shuttle after the break. Incidentally, the tether continued arcing long after it and its satellite were drifting free, until finally it went into night conditions where the electron density was insufficient to sustain the arc. Noel Sargent (2002) has investigated whether the TSS-1R arc was seen to disrupt Shuttle communications. Although he has found no record of disturbed communications during the event, for most of the time the arc was shielded by metallic structures from the communications antennas, and when the tether broke, the arc was many meters from the receiving antennas. It remains to be seen whether sustained arcs produce radio noise severe enough to be a communications problem.

Comments:

This is official confirmation that the tether continued to produce plasma energy long after it broke free, accounting for the "fluorescent bulb" glowing effect viewed after. We believe this concentrated collection of plasma energy is what attracted the "swarm" of "critters" to a "feeding frenzy"

To get the full PDF file you can download LEO_Charging_Guidelines_v1.3.1.zip

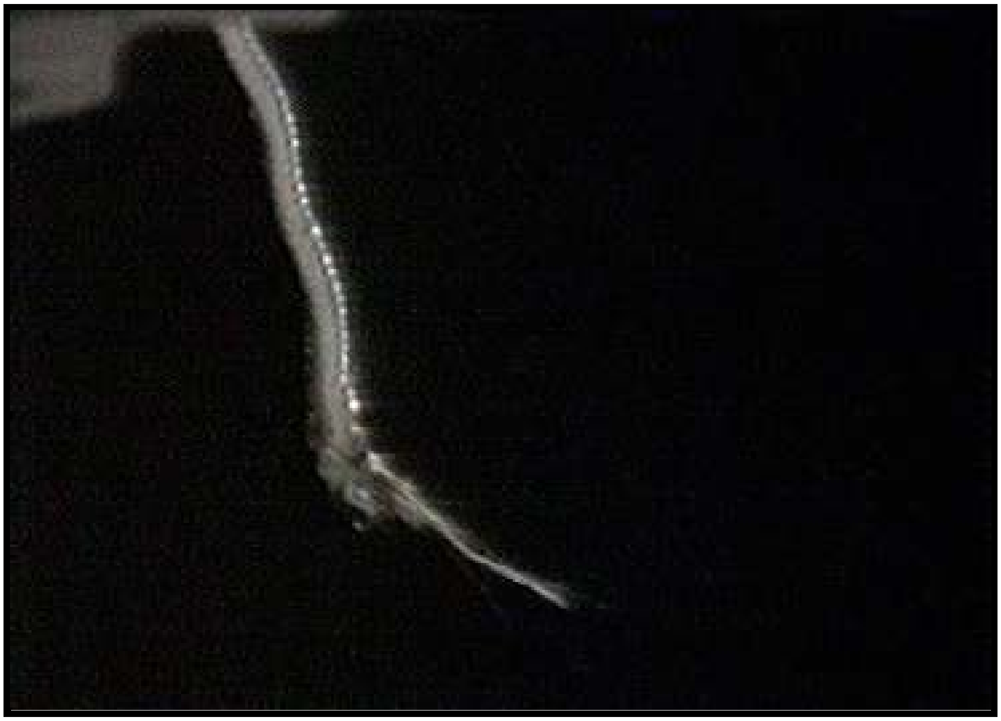

Tether Optical Phenomena Experiment (TOP)

Using a hand-held camera system with image intensifiers and special filters, the TOP investigation will provide visual data that may allow scientists to answer a variety of questions concerning tether dynamics and optical effects generated by TSS-1R. In particular, this experiment will examine the high-voltage plasma sheath surrounding the satellite...

In one mode of operation, the current developed in the Tethered Satellite System is closed by using electron accelerators to return electrons to the plasma surrounding the orbiter. The interaction between these electron beams and the plasma is not well understood...

Associate Investigator: Stephen Mende, Lockheed Martin

SOURCE

The Report After the Fact

...

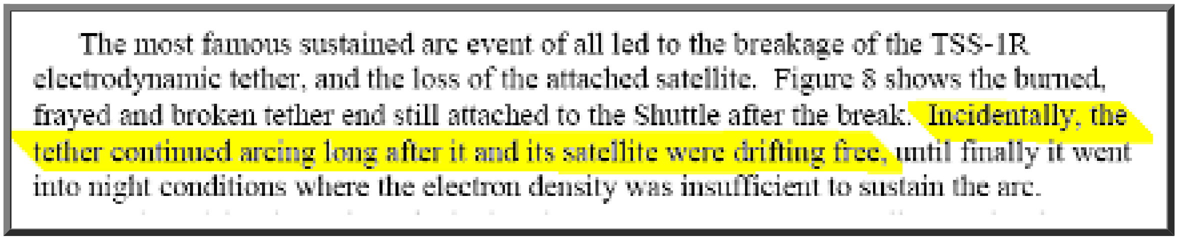

The most famous sustained arc event of all led to the breakage of the TSS-1R electrodynamic tether, and the loss of the attached satellite. Figure 8 shows the burned, frayed and broken tether end still attached to the Shuttle after the break. Incidentally, the tether continued arcing long after it and its satellite were drifting free, until finally it went into night conditions where the electron density was insufficient to sustain the arc.

.....

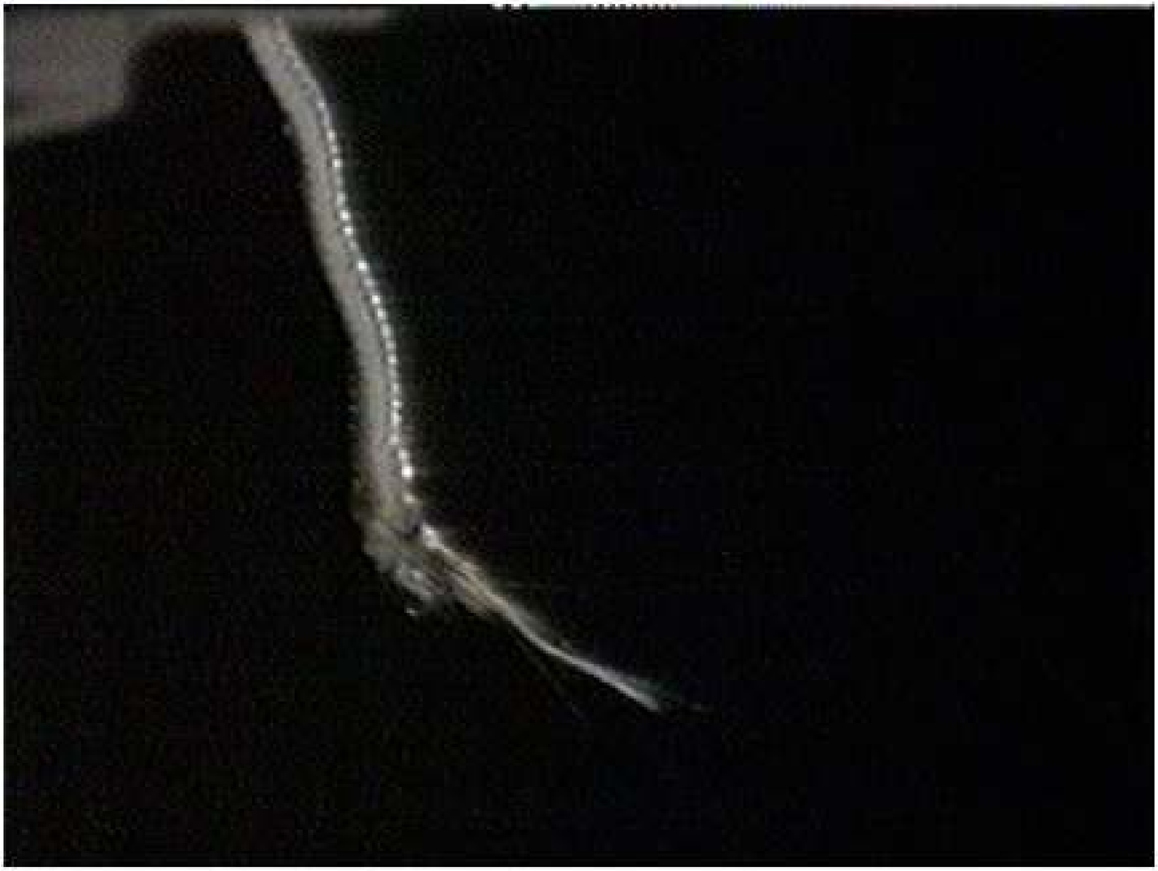

Figure 8 – The end of the remaining TSS-1R tether

…

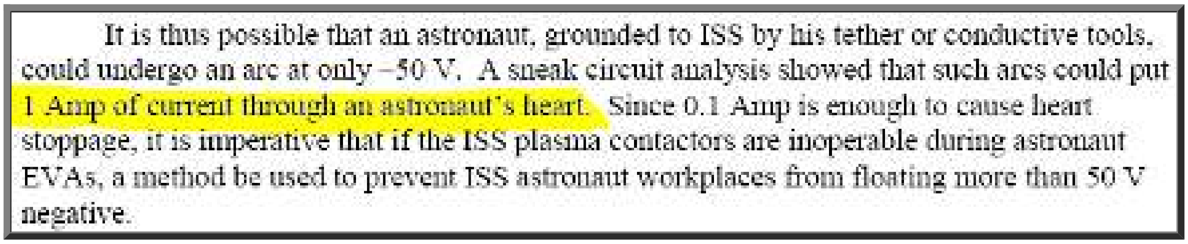

It is thus possible that an astronaut, grounded to ISS by his tether or conductive tools, could undergo an arc at only −50 V. A sneak circuit analysis showed that such arcs could put 1 Amp of current through an astronaut's heart. Since 0.1 Amp is enough to cause heart stoppage, it is imperative that if the ISS plasma contactors are inoperable during astronaut EVAs, a method be used to prevent ISS astronaut workplaces from floating more than 50 V negative.

…

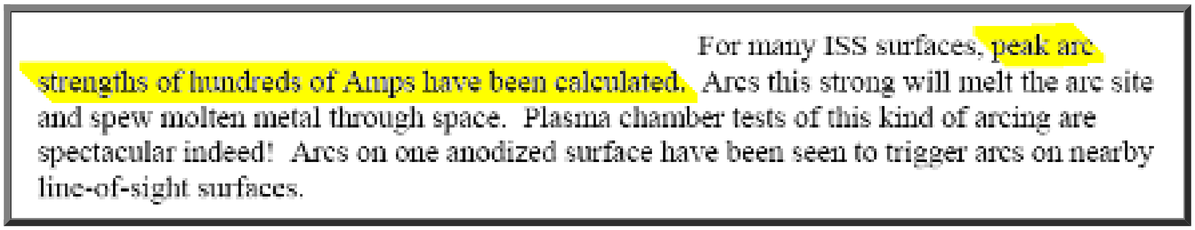

For many ISS surfaces, peak arc strengths of hundreds of Amps have been calculated. Arcs this strong will melt the arc site and spew molten metal through space. Plasma chamber tests of this kind of arcing are spectacular indeed! Arcs on one anodized surface have been seen to trigger arcs on nearby line-of-sight surfaces.

…

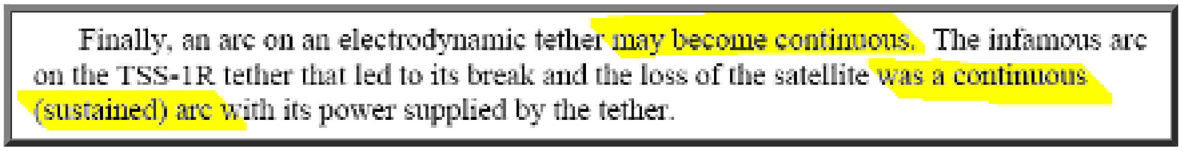

Finally, an arc on an electrodynamic tether may become continuous. The infamous arc on the TSS-1R tether that led to its break and the loss of the satellite was a continuous (sustained) arc with its power supplied by the tether.

…

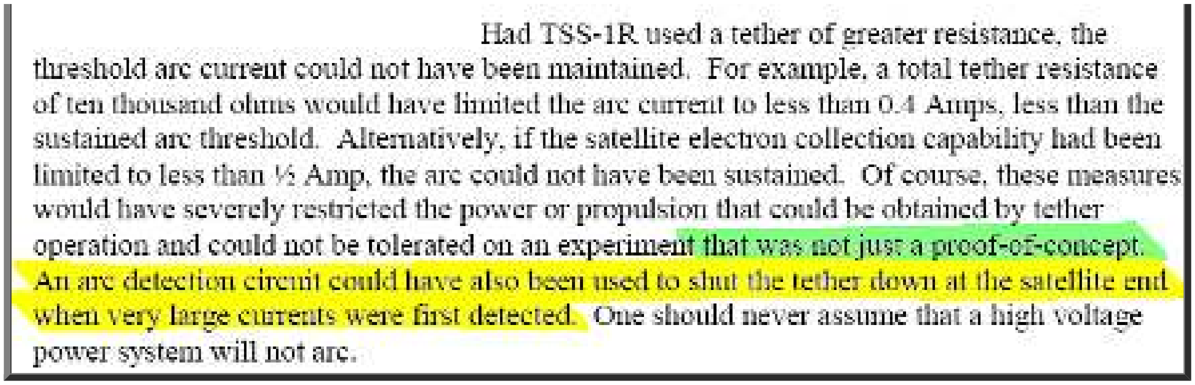

Had TSS-1R used a tether of greater resistance, the threshold arc current could not have been maintained. For example, a total tether resistance of ten thousand ohms would have limited the arc current to less than 0.4 Amps, less than the sustained arc threshold. Alternatively, if the satellite electron collection capability had been limited to less than ½ Amp, the arc could not have been sustained. Of course, these measures would have severely restricted the power or propulsion that could be obtained by tether operation and could not be tolerated on an experiment that was not just a proof-of-concept. An arc detection circuit could have also been used to shut the tether down at the satellite end when very large currents were first detected. One should never assume that a high voltage power system will not arc.

…

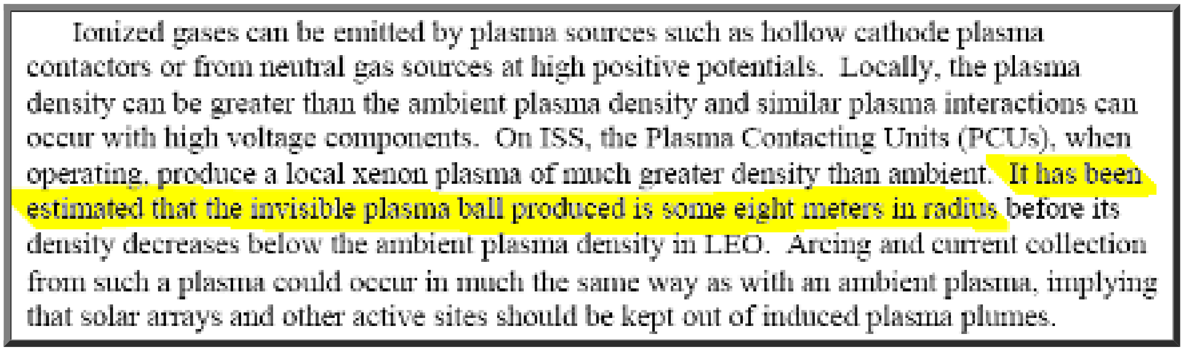

Ionized gases can be emitted by plasma sources such as hollow cathode plasma contactors or from neutral gas sources at high positive potentials. Locally, the plasma density can be greater than the ambient plasma density and similar plasma interactions can occur with high voltage components. On ISS, the Plasma Contacting Units (PCUs), when operating, produce a local xenon plasma of much greater density than ambient. It has been estimated that the invisible plasma ball produced is some eight meters in radius before its density decreases below the ambient plasma density in LEO. Arcing and current collection from such a plasma could occur in much the same way as with an ambient plasma, implying that solar arrays and other active sites should be kept out of induced plasma plumes.

…

Electrodynamic Tethers ~ Menu
Critters ~ Critters 01 ~ Critters 02 ~ Critters 03

Copyright Notice

All pictures and photos on this page, unless otherwise noted, have been gathered from pubic domain sources or are available under GNU License. Further documentation is available on our Copyright Page and our Legal Department.

Article Quotes all have links to their appropriate source and are noted. All other text not so noted is available under the terms of the GNU Free Documentation License

All information on this page is used only for non profit educational purposes. All reasonable attempts have been made that no credits are missed, but with a work of this size omissions may occur. If you see anything of yours that has not been properly credited or wish to have removed please contact the Webmaster

Menu

Webpages © 2001-2006
Blue Knight Productions

STS 75 TETHER INCIDENT

-

Tether Deployed

Tether Satellite being deployed from the Shuttle Bay of STS-75

- -

219506148 DAY:056

Tether Activated

Tether Satellite begins to glow as it gathers electricity from the Ionosphere

213809089 DAY:056

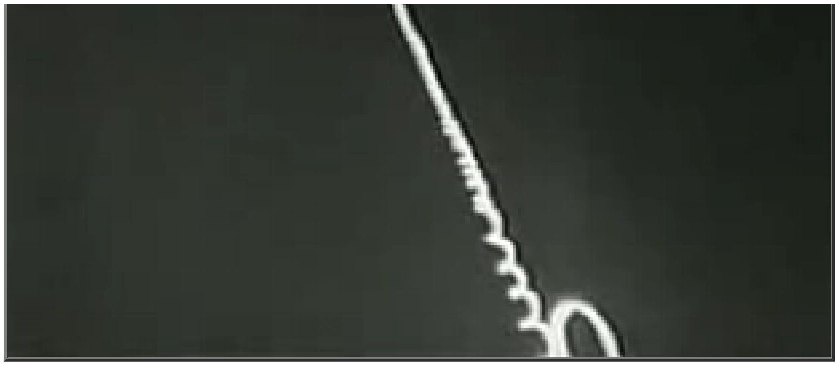

Tether Breaks Free

Tether Satellite cable fries the connection and breaks free... still producing energy. The glow increases and the cable gets wider. (More on this later)

-

Tether Broke Free

Tether Satellite worth 100 million has overloaded the circuits within minutes of being activated because the NASA Scientists under estimated the potential. The tether in just a few moments of operation produced 10 times the anticipated power and they forgot to install a circuit breaker.

-

NOTES:
Add "Critter" Portion
Add Video
Add FLV Viewer

TETHER BROKEN FREE AND DRIFTING AWAY

..

Source: Original Image NASA Cassini/ Huygens N00008771.jpg

~ MENU ~

US006116544A

United States Patent [19]

Forward et al.

[11] **Patent Number: 6,116,544**

[45] **Date of Patent: Sep. 12, 2000**

---

[54] **ELECTRODYNAMIC TETHER AND METHOD OF USE**

[75] Inventors: **Robert L. Forward**, Houston, Tex.; **Robert P. Hoyt**, Seattle, Wash.

[73] Assignee: **Tethers Unlimited, Inc.**, Seattle, Wash.

[21] Appl. No.: **08/929,271**

[22] Filed: **Sep. 12, 1997**

[51] **Int. Cl.**^{7} **B64G 1/00**

[52] **U.S. Cl.** .... **244/158 R**; 244/166; 244/172

[58] **Field of Search** .... 244/158 R, 164, 244/166, 172

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,097,010	6/1978	Colombo et al.	244/158 R
4,824,051	4/1989	Engelking	244/158 R
4,923,151	5/1990	Roberts	244/158 R

OTHER PUBLICATIONS

"Tethers Open New Space Options" by Ivan Bekey, *Astronautics & Aeronautics*, 1983.

Primary Examiner—J. Woodrow Eldred
Attorney, Agent, or Firm—Arthur M. Dula

[57] **ABSTRACT**

The present invention comprises an electrodynamic tether structure and a method of use. The structure of the tether taught by the present invention is a short, wide, interconnected multiwire (compared to the long, narrow single wires of the prior art) conductive tether whose area maximizes electrodynamic drag while simultaneously minimizing the Area-Time-Product swept by the tether during its operating life. The preferred tether length is two kilometers to five kilometers. The preferred tether mass is one percent to five percent of the spacecraft mass. The method of operation comprises orienting the tether structure at an angle to the local vertical to maximize electrodynamic drag on the host spacecraft and minimize tether instability. The angle of 35.26 degrees is preferred.

26 Claims, 12 Drawing Sheets

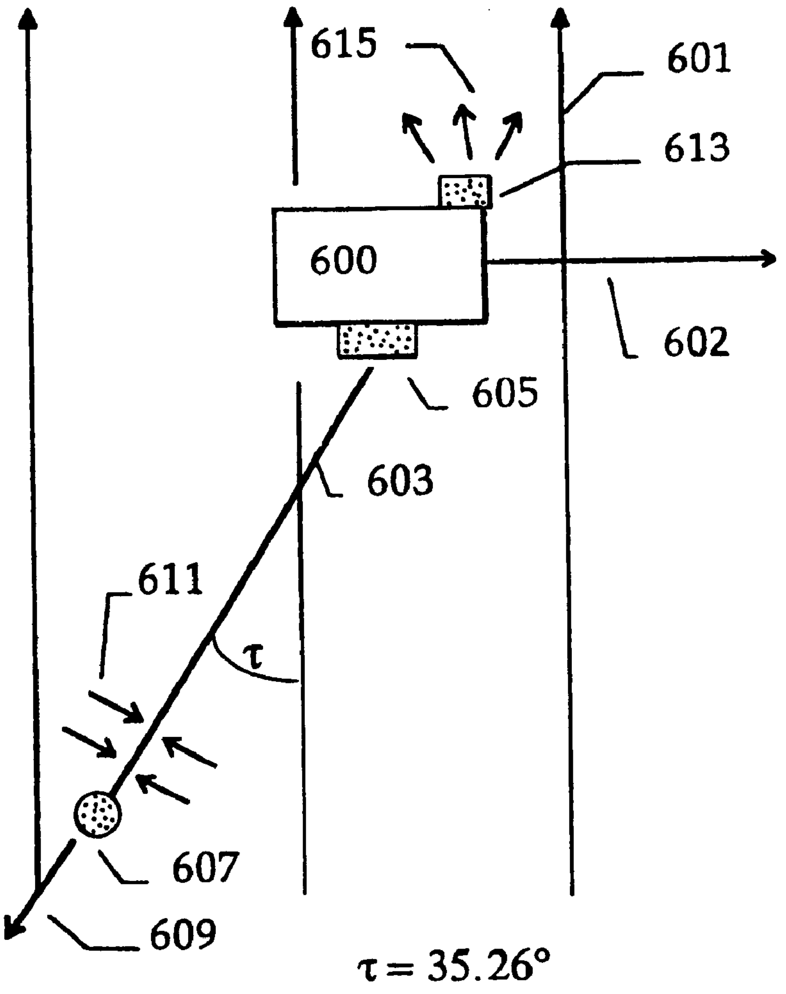

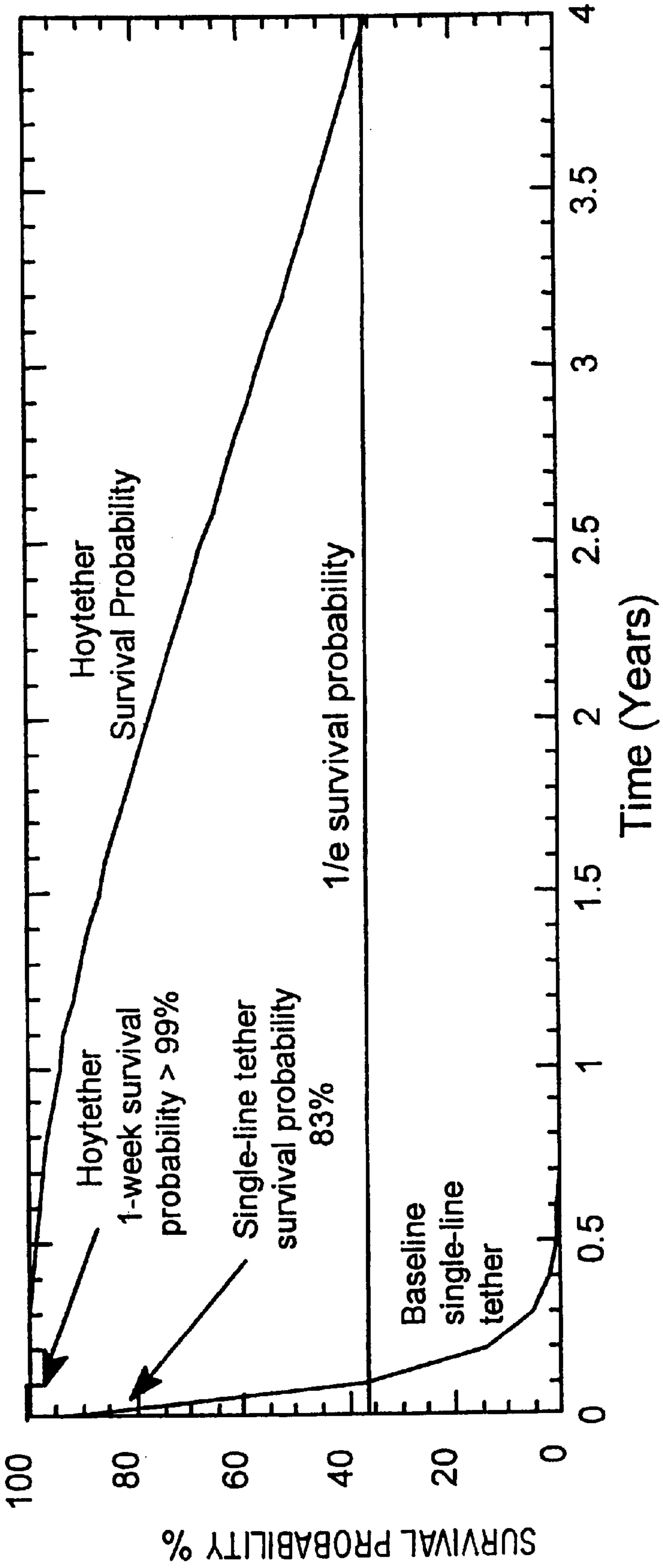
SURVIVAL PROBABILITY %
100
80
60
40
20
0
Time (Years)
0
0.5
1
1.5
2
2.5
3
3.5
4
Hoytether Survival Probability
1/e survival probability
Hoytether 1-week survival probability > 99%
Single-line tether survival probability 83%
Baseline single-line tether

FIG. 1a

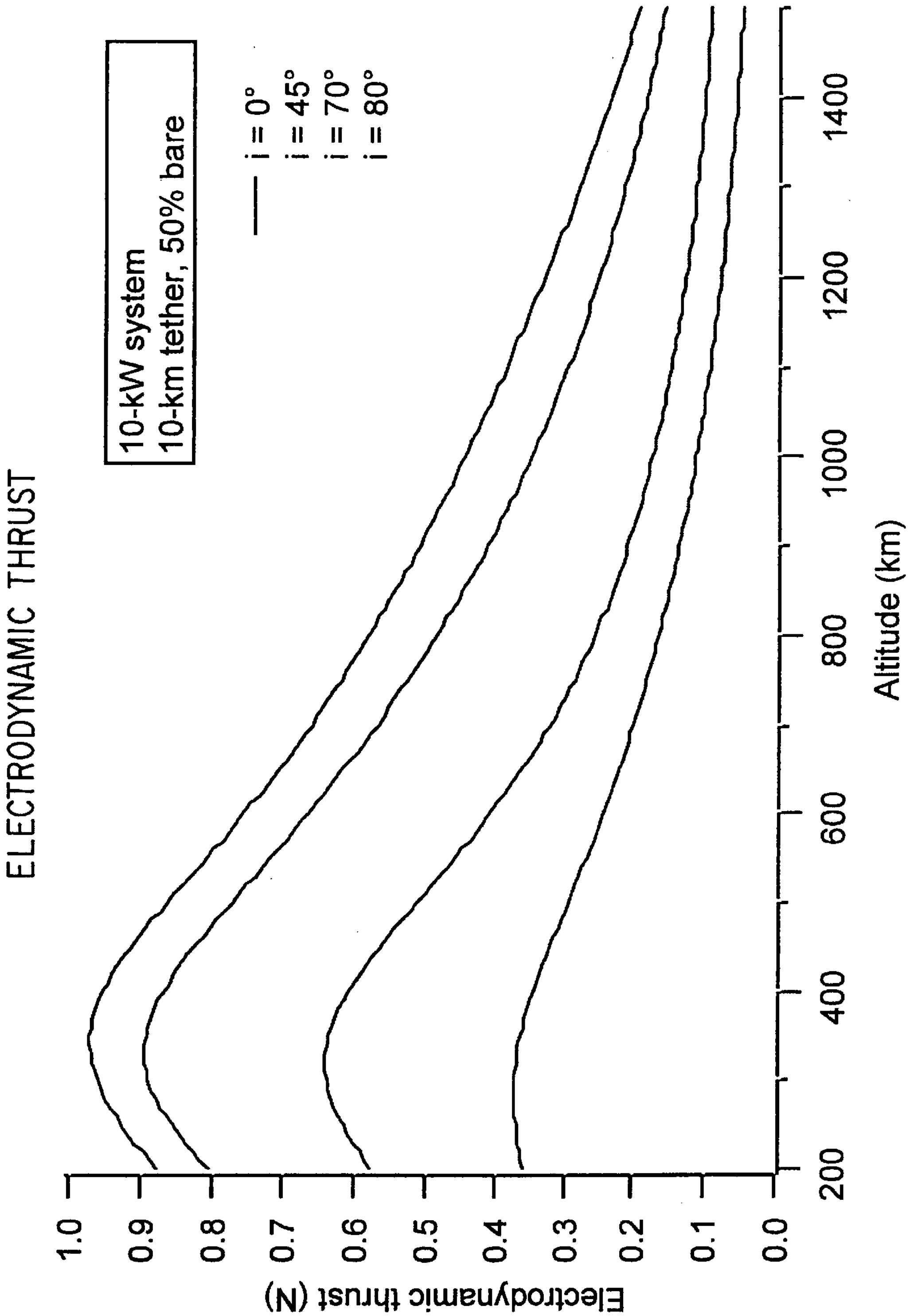
ELECTRODYNAMIC THRUST
10-kW system
10-km tether, 50% bare
i = 0°
i = 45°
i = 70°
i = 80°
Electrodynamic thrust (N)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
200
400
600
800
1000
1200
1400
Altitude (km)

FIG. 1b

Reentry Time Sensitivity
Unpowered 5 km ProSEDS Tether

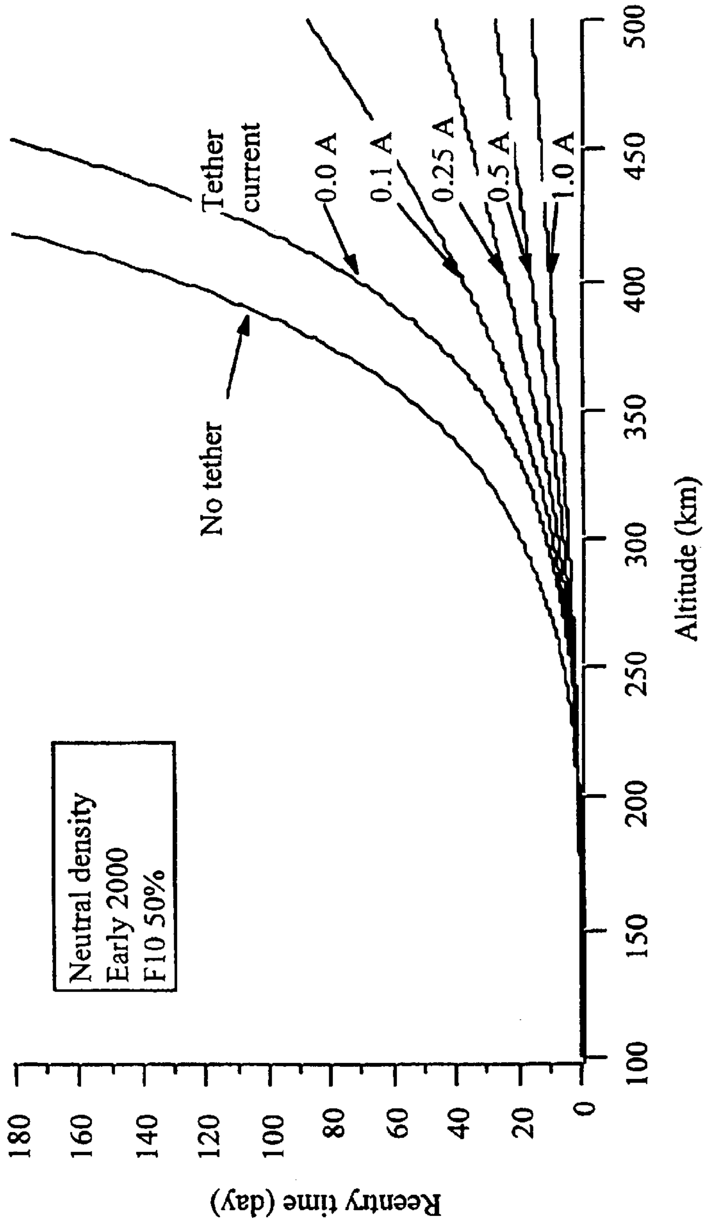

Enrico Lorenzini, SAO, NASA ProSEDS Mission (July 1997),
NASA/MSFC Grant NAG8-1303

FIG. 1c

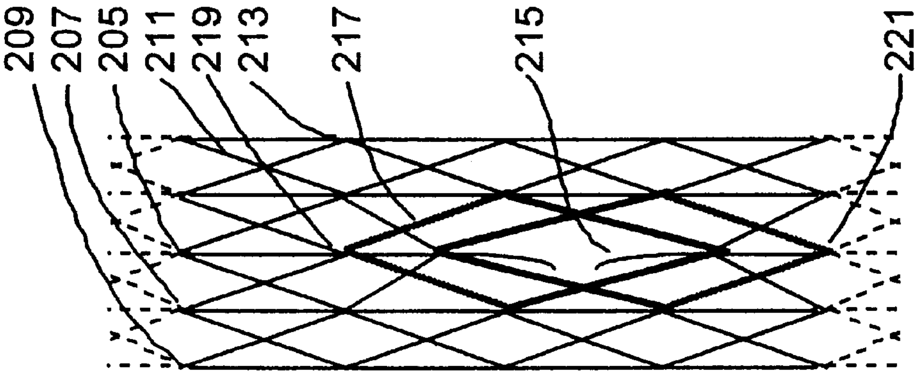
209
207
205
211
219
213
217
215
221

FIG. 2c

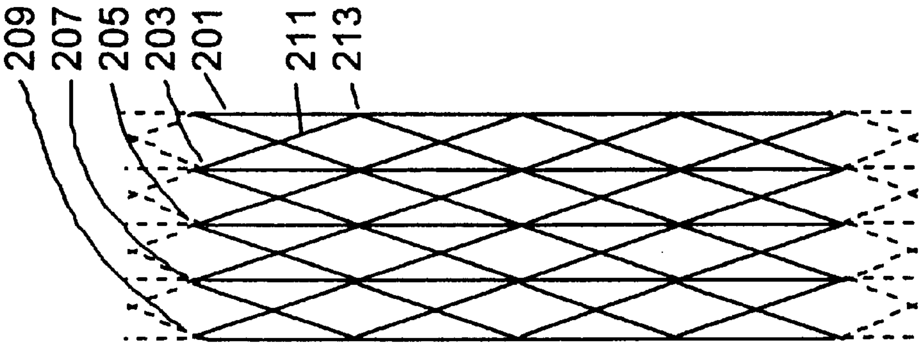
209
207
205
203
201
211
213

FIG. 2b

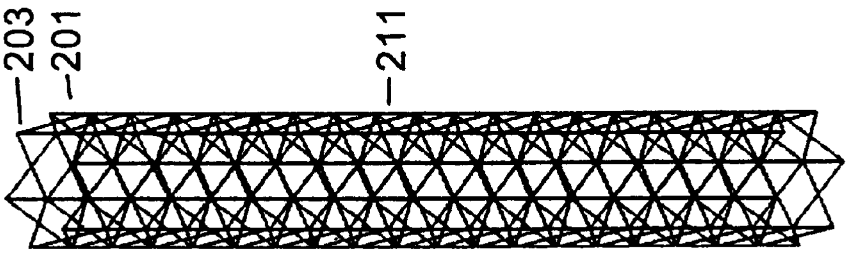

FIG. 2a

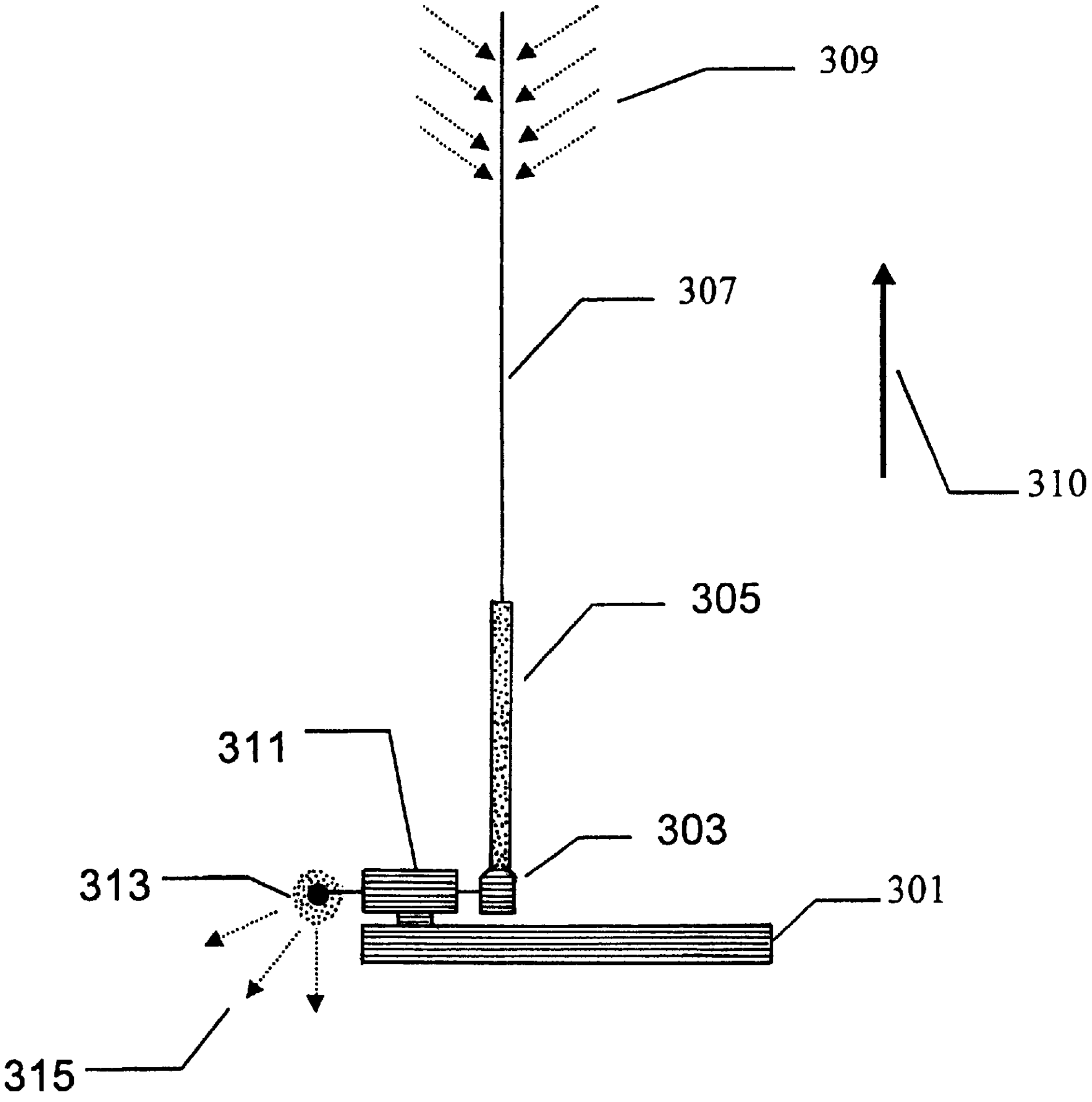
309
307
310
305
311
303
313
301
315

FIG. 3

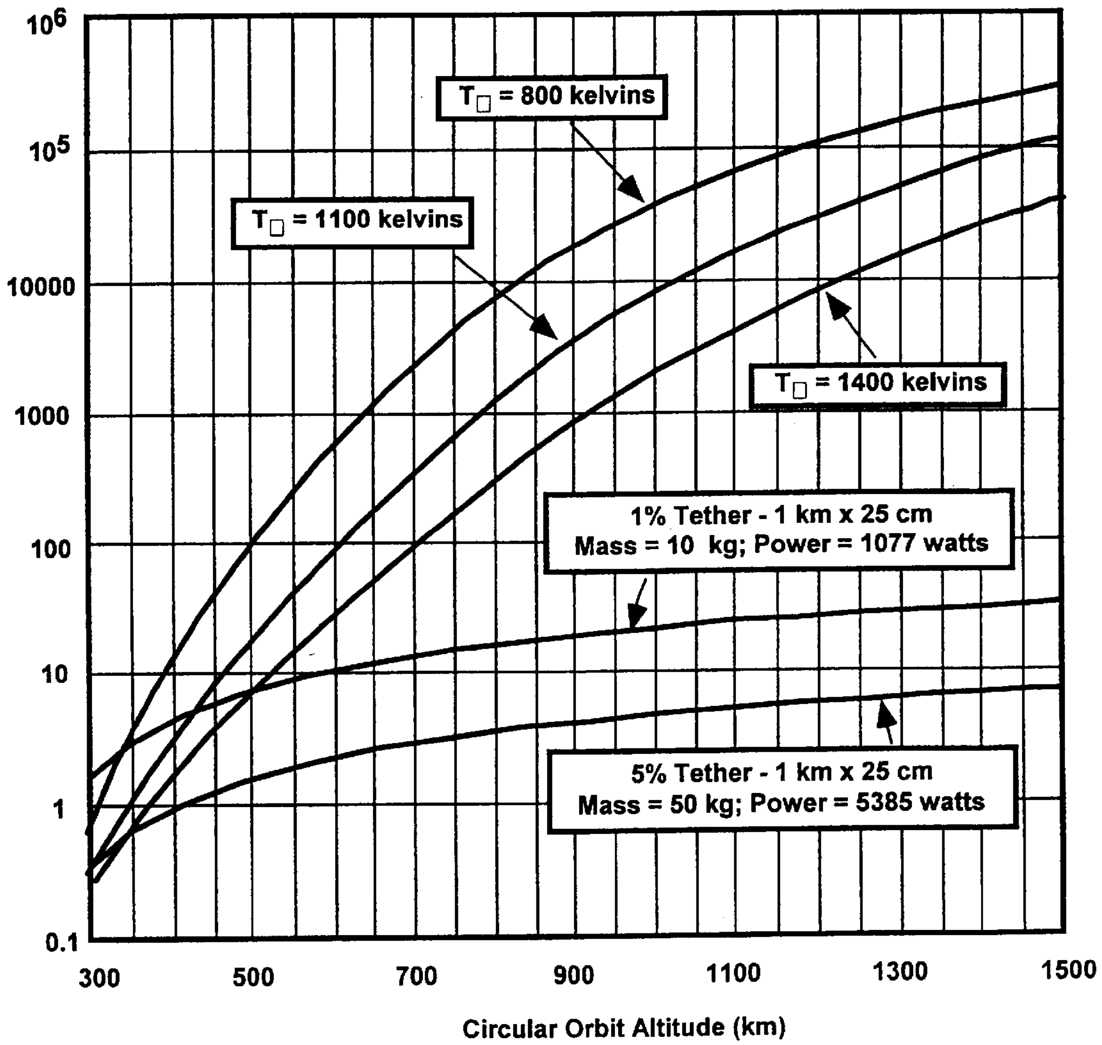

Area-Time Product for Mean and Extremes of Exospheric Temperature

(m = 1000 kg, Cd = 2.0 ; 1% Tether, 1 km x 25 cm; power = 1077 watts)

(5% Tether, 1 km x 25 cm; power =5385 watts)

FIG. 4

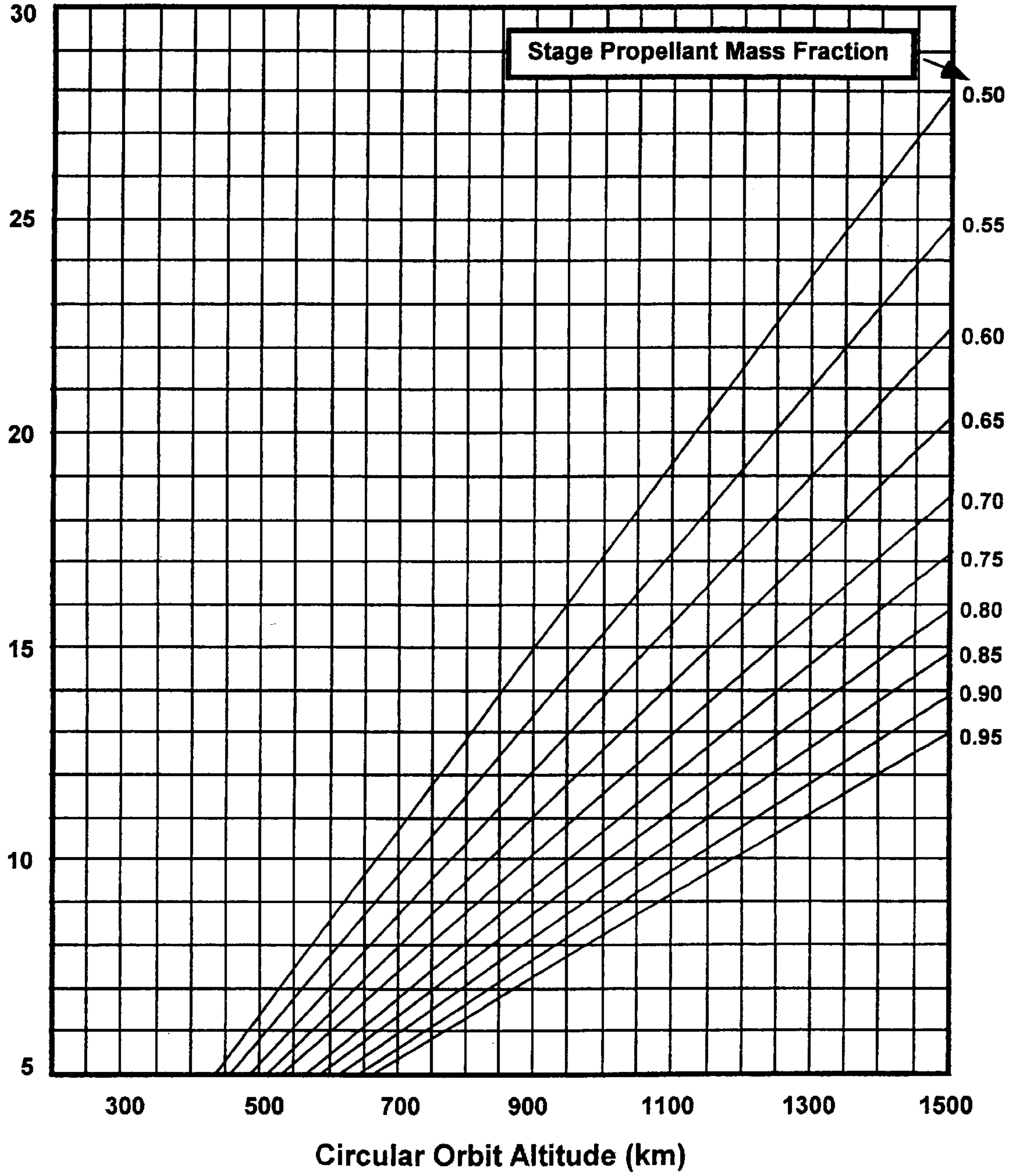

Conventional Solid Rocket Motor System

Percent Increase in Orbital Mass vs. Altitude and I' (Isp = 288 sec)

FIG. 5

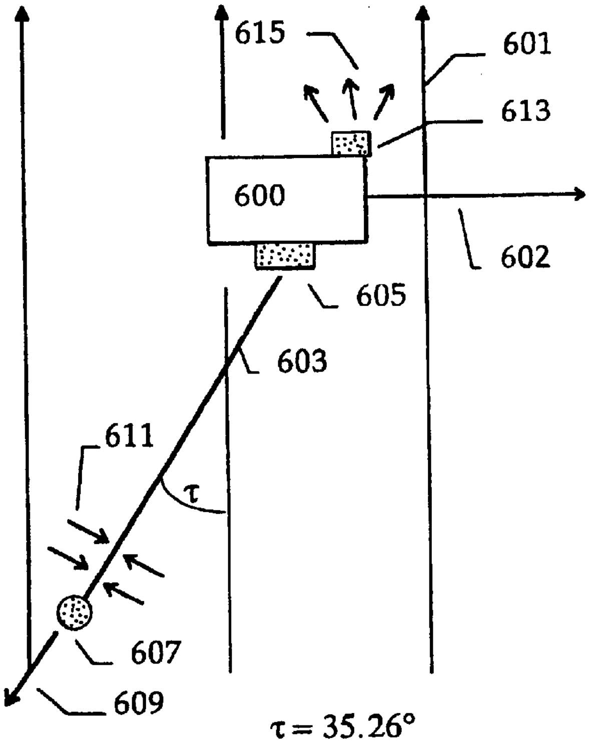
τ
τ = 35.26°

FIG. 6

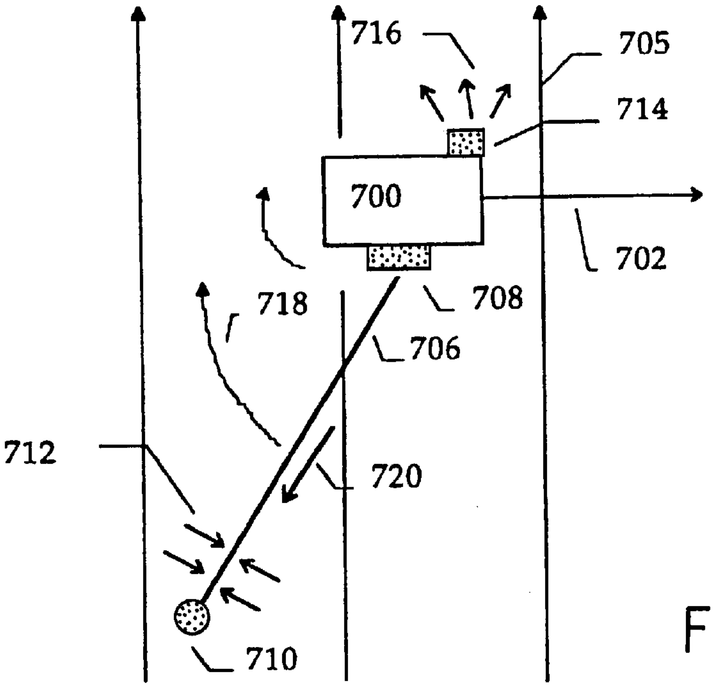

FIG. 7

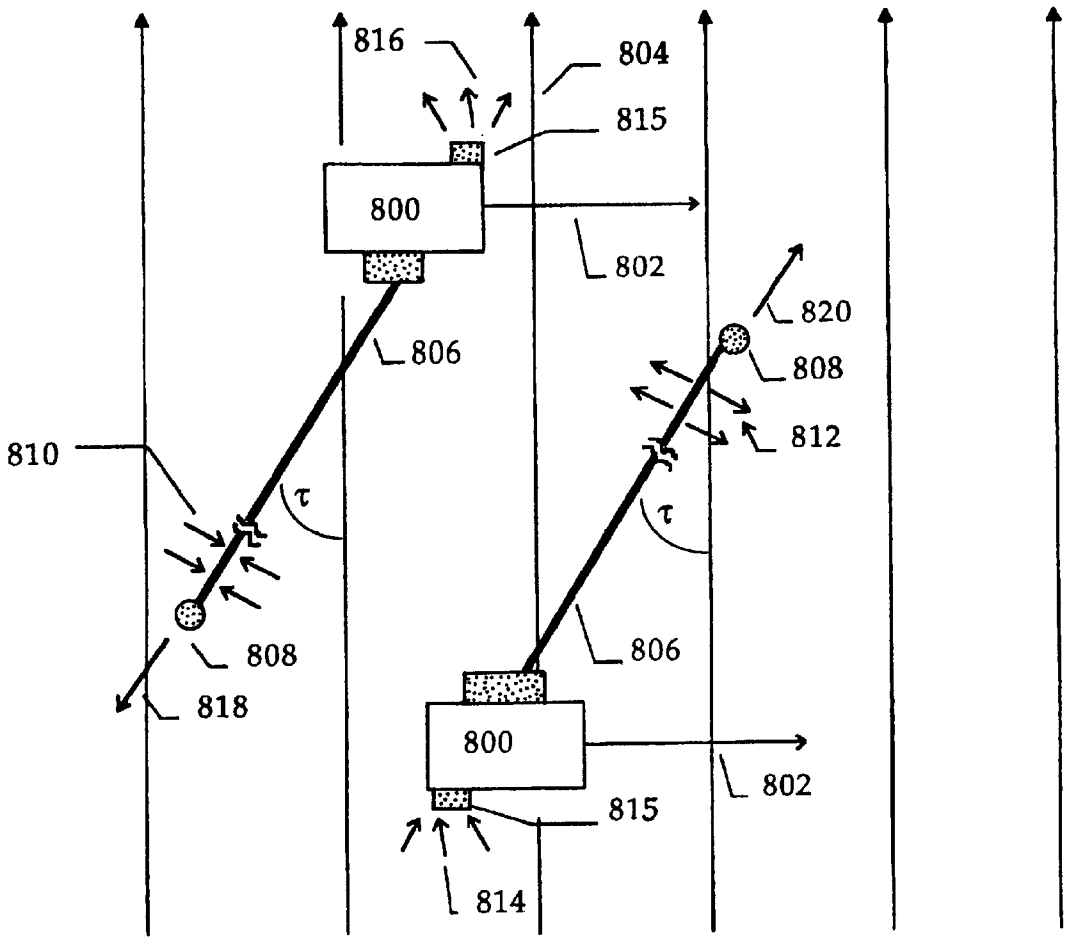
τ
τ

FIG.8

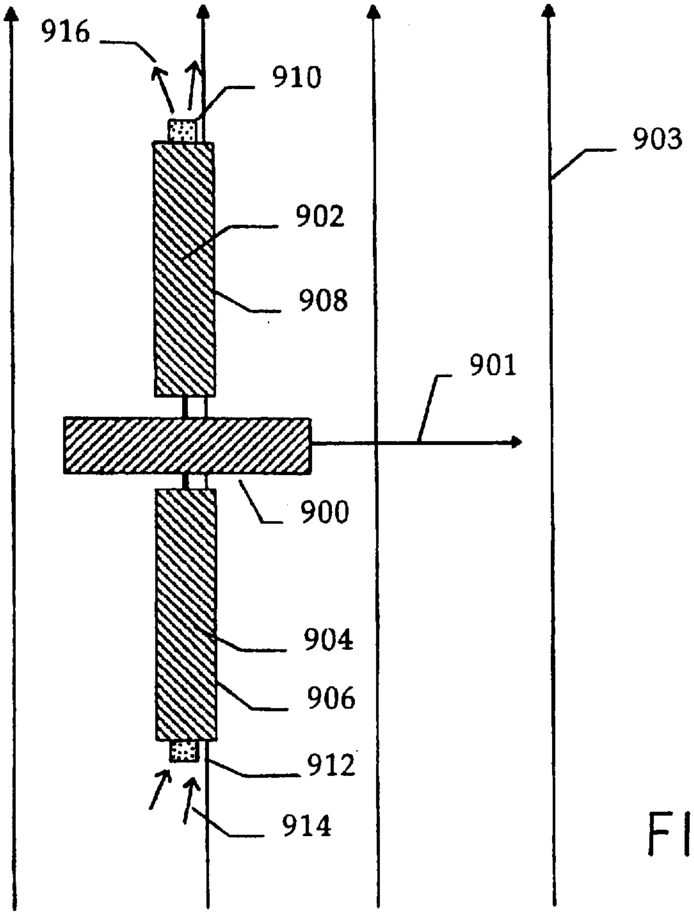

FIG. 9

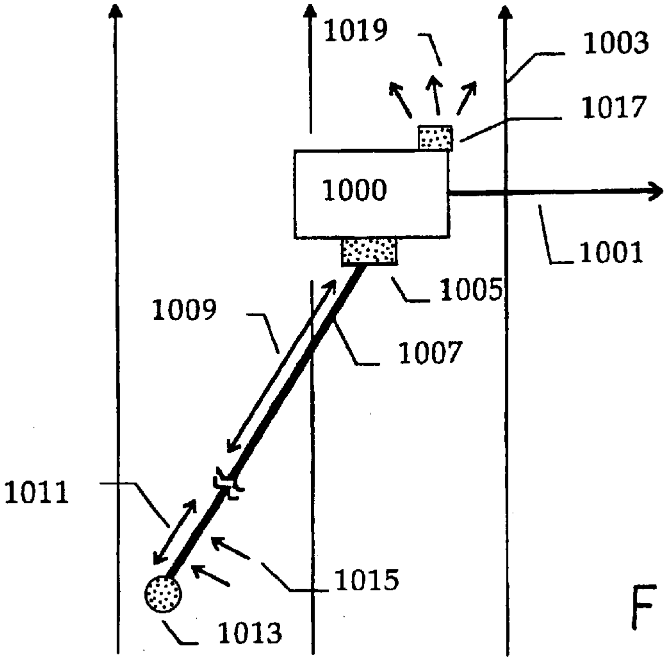

FIG. 10

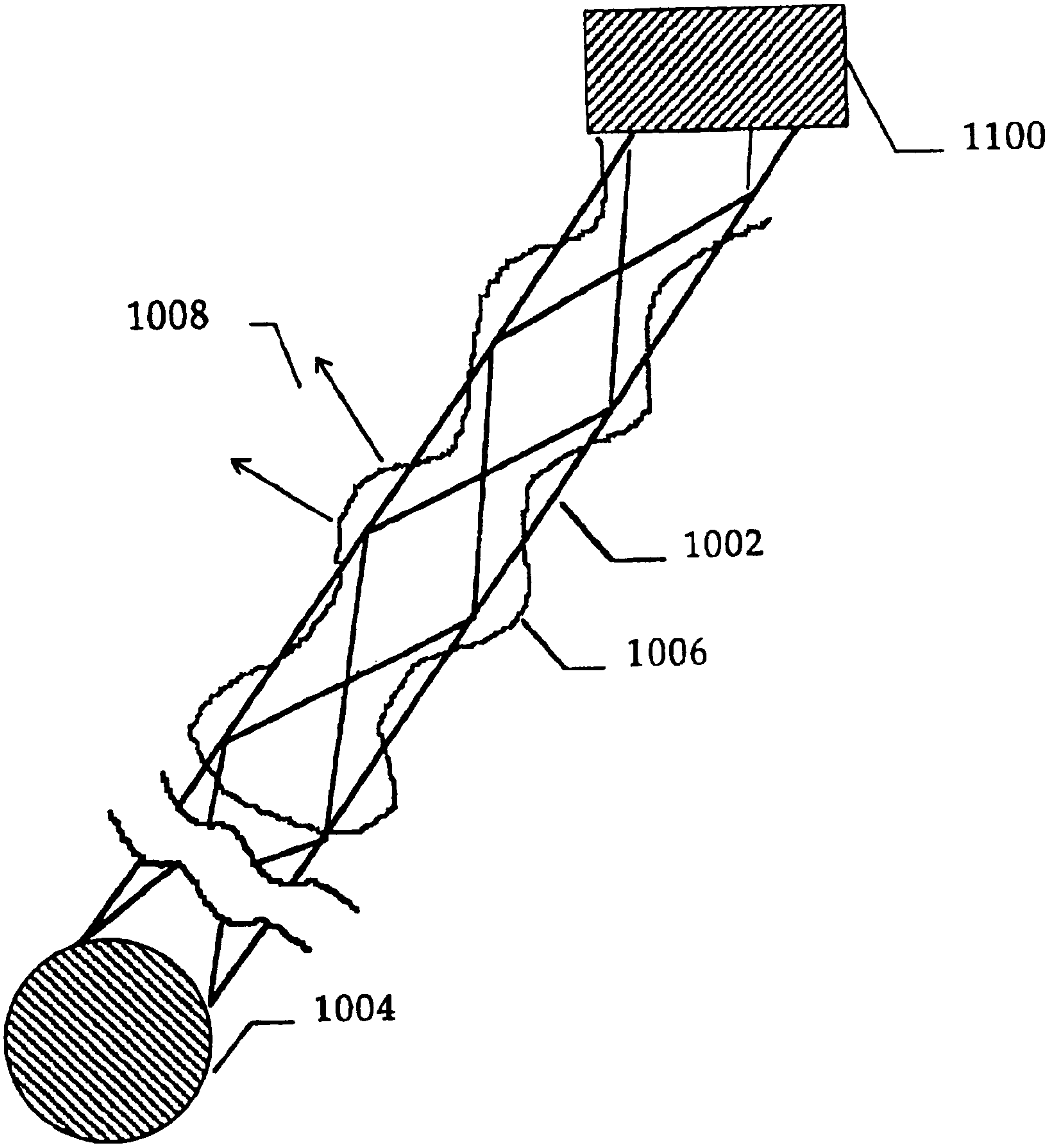
1100
1008
1002
1006
1004

FIG. 11

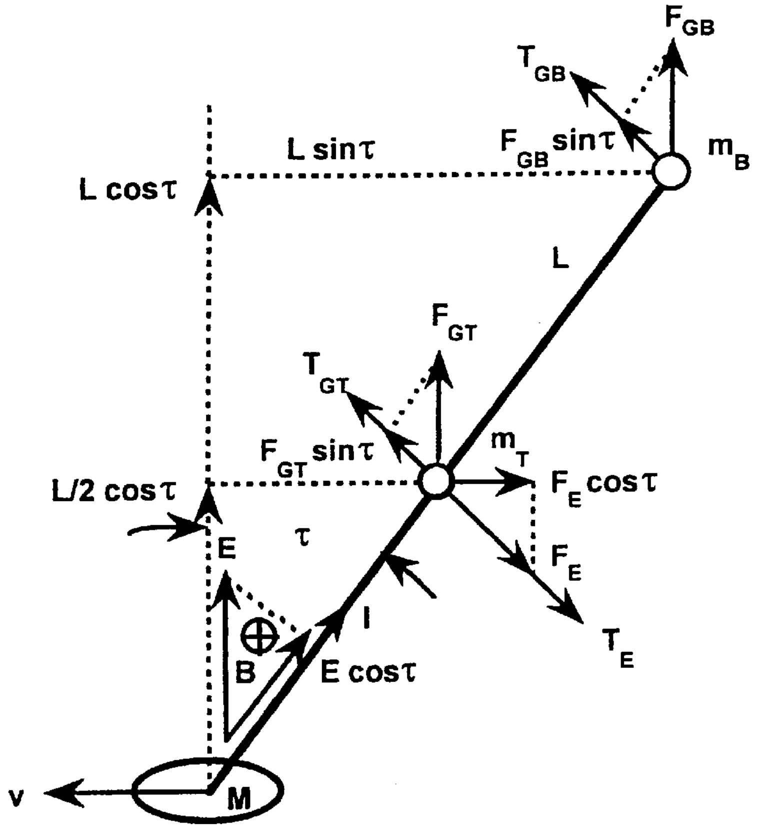
F_{GB}
T_{GB}
$F_{GB}\sin\tau$
m_B
$L\sin\tau$
$L\cos\tau$
L
F_{GT}
T_{GT}
$F_{GT}\sin\tau$
m_T
$F_E\cos\tau$
$L/2\cos\tau$
E
τ
F_E
I
T_E
B
$E\cos\tau$
v
M

FORCES AND TORQUES ON AN UPWARD DEPLOYED CONDUCTING TETHER DUE TO THE MOTION OF THE HOST SPACECRAFT THROUGH THE MAGNETIC FIELD OF THE EARTH.

FIG. 12

ELECTRODYNAMIC TETHER AND METHOD OF USE

TECHNICAL FIELD

This invention relates generally to apparatus and methods useful for changing the state vector of a space object when the space object is moving relative to a magnetic field. More specifically, the present invention relates to an apparatus and method of using a conducting tether to produce an electrodynamic force to deorbit a satellite from its orbit around a celestial body, such as the Earth, which has an associated magnetic field.

RESERVATION OF RIGHTS

This application is subject to certain rights of the U.S. government as a result of contracts between the U.S. government and the inventors.

BACKGROUND ART

The present invention has its principal utility in outer space, primarily for deorbiting satellites at the end of their useful life to mitigate the harm and reduce the liability created by the proliferation of space debris. In order to obtain a better understanding of the present invention it is helpful to understand the prior art of space tethers, especially tether dynamics and tether electrodynamics. The present invention may be more readily understood through a review of the experimental prior art and a mathematical analysis of electrodynamic space tethers.

Prior Art Tethers:

A tether was originally a rope or chain used to fasten an animal so that it grazed only within certain limits. Tethers have been used for decades in space to attach astronauts to their spacecraft.

In 1974 Professor Guiseppe Colombo, holder of the Galileo chair of astronomy at the University of Padua in Italy, proposed using a long tether to support a satellite from an orbiting platform. U.S. Pat. No. 4,097,010, which issued to Professor Colombo and Mario Grossi on Jun. 27, 1978, teaches a satellite connected by means of a long tether to a powered spacecraft. Colombo actively pursued the design of a tethered satellite system.

Several NASA experiments, such as the two Small Expendable Deployer System (SEDS 1 & 2) experiments and the Plasma Motor Generator (PMG) experiment used tethers in space. SEDS used a nonconducting tether. The PMG used a 500-meter conducting tether. The Tethered Satellite System flights in 1992 and 1996 (TSS-1 & 1R) used a 20,000-meter conducting tether.

On the TSS-1 mission the tether deployed only 260 meters (853 feet) before the deployer failed. On the TSS-1R the tether was deployed 19,500 meters. In the SEDS-2 flight, a 0.8-mm diameter, 20,000-meter long braided single-line tether was deployed to study tether dynamics and lifetime. Orbital debris or a meteoroid severed this tether in less than four days.

In the TSS-1R flight, the conducting single-line tether was severed after five hours of deployment. This failure was caused by an electric arc produced by the 3,500 volts of electric potential generated by the conductive tether's movement through the Earth's magnetic field.

The Tether Physics and Survivability (TiPS) satellite consists of two end masses connected by a 4,000-meter long non-conducting tether. This satellite was deployed on Jun. 20, 1996 at an altitude of 1,022 kilometers (552 nautical miles). Its tether is an outer layer of Spectra™ 1000 braid over a core of acrylic yarn. The yarn will "puff" its outer braid to two millimeters to "give it a larger cross section to improve its resistance to debris and small micrometeoroids", according to the National Reconnaissance Office (NRO), which is a sponsor of the TiPS mission. As of Jun. 21, 1997 the TiPS tether had survived one year.

References:

1. Joseph A. Carroll, "SEDS Deployer Design and Flight Performance", paper WSEDSA-1 at the 4th International Conference on Tethers in Space, Washington, D.C., April 1995.
2. James E. McCoy, et. al. "Plasma Motor-Generator (PMG) Flight Experiment Results", pp.57–84, Proceedings of the 4th International conference on Tethers in Space, Washington, D.C., April 1995.
3. W. John Raitt, et. al. "The NASA.ASI-TSS-1 Mission, Summary of Results and Reflight Plans, pp. 107–118, Proceedings of the 4th International conference on Tethers in Space, Washington, D.C., April 1995.
4. Joseph C. Anselmo, "NRO Orbiting Spacecraft Studies Tether Survivability", Aviation Week, page 24, Jul. 1, 1996.

These experiments all used single line tethers.

The following reference is illustrative of the current state of the art in space tethers: Paul A. Penzo and Paul W. Ammann. Tethers in Space Handbook—Second Edition. NASA Office of Space Flight, NASA Headquarters, Washington, D.C. 20546. See also the hundreds of references in the 33 page bibliography at the end of the handbook.

The "Hoytether"™, an Improved, High-Reliability Tether:

In 1991, one of the present inventors, Robert Hoyt, invented a lightweight net-like structure that provides many redundant load-bearing paths. A number of primary load bearing lines running the length of the structure are connected periodically by diagonal secondary lines. The disclosed embodiment of this invention has the secondary lines firmly connected by knots to the primary lines. The secondary lines are connected only to the primary lines. At either end of the disclosed structure, a support ring enforces the cylindrical spacing between the primary lines. The secondary lines are designed with a small amount of slack. These secondary lines are only put under load if a primary line fails. This specific tether structure was disclosed to the public in 1992 (Forward, R. L., "Failsafe Multistrand Tether Structures for Space Propulsion", AIAA paper 92–3214, 28th Joint Propulsion Conference, Nashville, Tenn., 1992 (hereinafter "1992 Hoytether structure"). This structure was named a "Hoytether". The term "Hoytether" is used throughout the remainder of this specification for this type of tether structure.

The present invention uses an improved Hoytether, which was invented by the same inventors as the present invention. This improved Hoytether is the subject of a copending PCT application. The Hoytether is discussed briefly in this specification to aid understanding of the present invention.

The 1992 Hoytether design teaches that the normally slack secondary lines have half the cross-section (0.707 the diameter) of the primary lines. There are twice as many secondary lines as primary lines, thus the mass of the secondary lines is equal to the mass of the primary lines. In an undamaged Hoytether, the primary lines carry the entire load, while none of the secondary lines are under load.

While the survival probability of a single-line tether decreases exponentially with time, the Hoytether can maintain a high, i.e. greater than 99 percent, survival probability for periods of months or years (forward and Hoyt, "Failsafe Multiline Hoythther Lifetimes", Paper AIAA 95–2890, 31^{st} Joint Propulsion Conference, July 1995).

References:

1. Robert L. Forward, Failsafe Multistrand Tethers for Space Propulsion, Forward Unlimited, P.O. Box 2783, Malibu, Calif. 90265, July 1992, Final Report on NASA Contract NAS8-39318 SBIR 91-1 Phase I.
2. Robert L. Forward and Robert P. Hoyt, Failsafe Multistrand Tether SEDS Technology Demonstration, Final Report on NAS8-40545 with NASA/MSFC (Jun. 14, 1995).
3. Robert L. Forward and Robert P. Hoyt, "High Strength-to-Weight Tapered Hoytether for LEO to GEO Payload Transfer" Final Report on contract number NAS8-40690 with NASA/MSFC (Jul. 10, 1996).

The Hoytether is essentially a tri-axial net structure, with 'primary' lines running along the length of the tether and two sets of 'secondary' lines connecting these primaries diagonally. They can be made by hand and connected with knots as is taught by the 1992 Hoytether structure. Because knotted connections severely limit the strength of a structure, it is desirable to use a knotless fabrication technique to achieve interconnections that have strengths approaching the limits of the constituent material. As these tethers may be many kilometers long; fast and inexpensive mechanical methods are required for their practical fabrication.

Hoytethers may be made by mechanical braiding, i.e. three-dimensional braiding, such as 3-D rotation braiding using braiding machines such as those developed by the Herzog Company in Germany (August Herzog Maschinenfabrik GmbH & Co., Postfach 2260.26012, Oldenburg, Germany. The specialized loom developed by the Nichimo Company of Japan (Nichimo Company Ltd., 2-6-2 Ohtemachi, Chiyoda-Ku, Tokyo, Japan) is used to produce "Ultracross" knotless fishing nets in which the individual strands are braided as a 4-braid line, and the strands are interbraided where they cross. This produces netting that has slipless interconnections that are very strong, approaching the maximum capability of the fiber. Such a loom could, with some modifications, produce the present invention's structure. Only two such machines exist, one in Japan, the other in Washington State. Unfortunately neither can work with the small line diameters needed to practice the preferred embodiment of the present invention. See generally, Ko, F. K., "Braiding", in *Engineered Materials Handbook, Vol. 1., Composites*. ASM International, Metals Park, Ohio, 1957. Pp. 519–528.

The most common 3-dimensional braiding machines are 4-step braiders based upon the designs of Maistre (German Patent P230-16986, issued 1973) and Forentine (U.S. Pat. No. 4,312,261, issued 1982). Braiding is accomplished by using pneumatics or solenoids to push the parts of the braiding machine to the proper positions. This is a slow process and making a Hoytether kilometers long with these machines would be very time consuming and expensive. The composites division of Albany International (Albany International Research Company, 777 West Street, Mansfield, Mass.) also produces a 3-D braiding machine. This machine uses modular braiding components that arc assembled breadboard fashion on a large wall.

Although braiding is the preferred technique, alternate fabrication methods such as Raschel knitting and crocheting can be used successfully. Multikilometer long Hoyththers are presently being produced for the inventors by the vendors Culzean Fabrics and Flemings Textiles using an electronically controlled crochet machine produced by Comez in Italy.

Space Tether Systems:

The prior art teaches the use of tethers in space applications. U.S. Pat. No. 5,163,641, issued on Apr. 9, 1990 to Yasaka, teaches the use of a powered spacecraft connected by a tether to a satellite. This tether is disconnected to change the state vector of the satellite. The state of the art of energy and momentum transfer using space tethers is discussed in Ivan Beckey's article "Tethering, a new Technique for Payload Deployment", Aerospace America, March 1997, at pages 36–40. Beckey concludes, "Tethers can perform the same functions as propulsive upper stages of direct payload injection, but at lower weight and cost per pound." U.S. Pat. No. 4,923,151, issued Mar. 1, 1988 to Roberts, Wilknson and Webster, teaches a tether power generator for earth orbiting satellites. U.S. Pat. No. 4,580,747, issued Mar. 15, 1983 to Pearson, teaches use of a long tether extending downward into the atmosphere from a satellite. The state vector of the satellite is changed by forces acting on a lifting body connected to the end of the tether. U.S. Pat. No. 4,824,051, issued Jan. 12, 1987 to Engelking, teaches passing an electric current through a conductive tether attached to a satellite to provide propulsive force to alter the orbit of the satellite. U.S. Pat. No. 5,082,211, issued Jan. 21, 1992 to Werka, teaches use of a tether to deorbit space debris. U.S. Pat. No. 4,727,373, issued Mar. 31, 1986 to Hoover, teaches an orbiting stereo imaging radar system having two spacecraft in synchronous parallel orbits connected by a tether.

Tether Dynamics:

In order to understand the forces that cause a tethered satellite to move upward and away from an orbiting satellite, for example, it is first necessary to explain briefly how a satellite remains in orbit. An orbiting satellite is acted on by the force of gravity which pulls it toward Earth, and by a centrifugal force, which pushes it away from Earth. The centrifugal force" (actually inertia) results from the motion of the satellite around its circular orbit. This is the same force that one can experience by swinging a ball around on the end of a string. A satellite is maintained in its orbit when it travels at the natural speed for its altitude and, as a result, the centrifugal force is equal to the gravitational force.

At the typical orbital altitude of 250 kilometers for a low-Earth orbit satellite, for example, a speed of approximately 7.6-km per second is required to create sufficient centrifugal force to balance gravitational attraction on the satellite. If the altitude is changed, the two opposing forces will no longer be in balance unless the satellite also changes its speed. A higher orbital altitude requires a slightly lower speed so the satellite will take longer to complete an orbit. Because of this, if two free-flying satellites are in orbits at different altitudes, the lower satellite will circle the Earth in less time than the satellite in the higher orbit.

If two satellites, at different altitudes, are connected to each other by a tether, they are forced to travel around their orbits together—in the same period of time, which is longer than the natural period of the lower satellite but shorter than that of the upper satellite. The lower satellite will, therefore, slow down below the natural speed for its orbit and will tend to fall to a lower orbit because the centrifugal force will now be less than the gravitational attraction of the Earth. An upward force in the tether that makes up the difference between centrifugal and gravitational forces holds it in place, however.

Correspondingly, the upper satellite will be accelerated above its natural orbiting speed (increasing its centrifugal force above the gravitational attraction) and will tend to move to a higher orbit. It, too, is held in place by an additional force (downward) in the tether. In other words, the net force downward on the lower satellite is balanced, through the tether, by the net force upward on the upper satellite. The effect of unbalanced forces on the two satellites is, therefore, to create tension in the tether. During the TSS-1 & 1R experiments, the inertia of the tethered satellite causes the satellite to rise above the orbiter as the tether is reeled out. Very close to the orbiter, there is little difference in the two orbits, and the tension force is insufficient to overcome friction in the deployer mechanism; therefore, until the satellite reaches a separation of approximately 1000-meters, the tension is augmented by small tether-aligned thrusters on the satellite. Beyond this point, the tension in the tether is the only force required.

By experimenting with a ball hung on a piece of elastic cord (a paddleball, for example) it is possible to simulate all the different types of oscillations that are possible on a space-based tether system. The elastic cord, representing the tether, may compress and stretch, causing the ball to bounce up and down (longitudinal oscillation). It also may move in a circular (skip-rope) motion or may develop wave-like motions (transverse oscillations). Even if the string itself remains straight, it is possible to get the ball swinging back and forth about its attachment point on the paddle like a child on a swing rope (pendulous motion).

Each type of motion occurs with a particular frequency, which depends on the length and tension of the tether. When the frequencies are different, the motions do not interact; however, at some tether lengths, the frequencies of two or more types of oscillation can become very close. At this point, energy can be transferred from one typc of motion to another, a phenomenon known as resonance. For instance, the transverse oscillations in the tether may cause the satellite to swing back and forth in pendulous motion.

Many different factors may cause oscillations; the movements of the satellite or Shuttle are but two of these. For an electrodynamic tether, the skip-rope and pendulous oscillations are of particular interest. If a current is passed through a tether, the current will interact with Earth's magnetic field, resulting in a force that may produce skip-rope and pendulous oscillations. Because it is necessary to maintain control of the satellite, much study has gone into identifying the different types of possible motions and the methods used to control them.

One way to control the magnitude of those motions that cause a change in tension or transverse motion at the end of the tether is to have an end mass connected to the Hoytether that maintains a controlled tension on the tether, working much like a spring-loaded 'dog leash'. This may be as simple as a coiled spring, or as complex as an active control system that measures the tension and transverse forces on the tether and adjusts the applied tension according to a local or remotely operating algorithm.

Electrodynamic Effects of Conductive Tethers:

Electric potential is generated across a conductive tether by its motion through the Earth's magnetic field. Electromagnetic forces acting on a conductive tether in orbit can make the tether system behave like an electric motor or generator, thereby exerting a useful force to alter the state vector of the tether and any mass attached to it.

Electrodynamic tether propulsion is unlike most other types of space propulsion in use or being developed for space application today—there is no hot gas expelled to provide thrust. Instead, the environment of near-Earth space is being utilized to propel a spacecraft or upper stage via electrodynamic interactions.

A charged particle moving in a magnetic field experiences a force that is perpendicular to its direction of motion and the direction of the field. When a current flows through a long, conducting tether the electrons flowing through the tether experience this force due to the fact that they are moving along the wire in the presence of Earth's magnetic field. This force is transferred to the tether and to whatever the tether is attached (like a spacecraft, satellite, space station or upper stage). It can be an orbit-raising thrust force or orbit-lowering drag force, depending upon the direction of current flow. Operation in one mode allows boost from LEO to higher orbit while reversing the current flow provides negative thrust for deboost. The principle is much the same for an electric motor; reverse its operation and it acts as a generator. The current passing through the tether is returned through the ionosphere to complete the current loop with collection and emission occurring on opposite ends of the tether.

The PMG experiment demonstrated that a conducting tether can be used as both a motor and a generator. The TSS experiments, especially TSS-1R showed that very large voltages (about 3500 volts) can be generated by a sufficiently long tether.

Uses of an electrodynamic tether as an orbit raising and lower propulsion system has many advantages over competing systems:

a. It is nearly propellantless. Most other systems expel hot gases and require extensive resupply. To emit current, the electrodynamic tether propulsion system may use plasma contactors developed as a part of the International Space Station Program. These contactors consume less than 20 kg of xenon gas per year with a 50% duty cycle. The electrodynamic tether propulsion system can also use field emitter arrays, which emit electrons without the use of any gas.

b. It can change both altitude and inclination. The Earth's magnetic field is non-uniform and can therefore provide both in- and out-of-plane forces for inclination changes as well as altitude changes. This is of particular interest to payloads requiring polar orbits in that they can be launched on a small launch vehicle into a lower inclination orbit and have it raised in space by the proper phasing of current through the tether.

A demonstration of the propulsive capabilities of electrodynamic tethers was recently approved for a flight test in 1999. The Propulsive SEDS or ProSEDS mission, will fly as a secondary payload on a Delta II launch vehicle and deploy a 5-km conducting tether using the existing SEDS deployer concept. The ProSEDS experiment will be followed by the Electrodynamic Tether Upper Stage (EDTUS) experiment that will demonstrated the use of electrodynamic forces to change both the altitude and inclination of the experimental spacecraft. FIGS. **1B** and **1C** show the calculated electrodynamic thrust at several inclinations and the reentry time sensitivity of the ProSEDS tether, respectively.

One application for long-life conducting electrodynamic tethers is as a "Terminator Tether™" for removing from orbit unwanted Earth orbiting spacecraft at the end of their useful lives. When the mission of the satellite is completed, the Terminator Tether™, weighing a small fraction of the mass of the satellite, would be deployed. The electrodynamic interaction of the conducting tether with the Earth's magnetic field will induce current flow in the tether conductor. The resulting energy loss from the heat generated by the current flowing through the ohmic resistance in the conducting tether will remove energy from the spacecraft, eventually causing it to deorbit, thus reducing the amount of orbital space debris that must be coped with in the future.

In the following analysis, it is shown that the amount of energy loss generated by an electrodynamic tether is essentially independent of its length or area, and instead is primarily proportional to the tether mass and the physical properties of the conductor metal chosen. In the typical example calculated, a 1000-kg spacecraft can be deorbited from a 1000-km high Earth orbit by a 10-kg mass tether in a month, while a 1-kg tether can deorbit a 1000-kg spacecraft in less than a year.

To the knowledge of the inventors, Joseph P. Loftus of NASA/JSC first proposed the general concept of using an electrodynamic tether to deorbit spent satellites. (Joseph P. Loftus <JLoftus@ems.jsc.nasa.gov>, personal communication via email to Robert Forward, Monday Jun. 10, 1996 15:50:10.) In order to show that the Loftus deorbit concept was not obvious to those skilled in the art of electrodynamic tethers, Forward contacted the leading expert, Joseph Carroll, of Chula Vista, Calif., who built and participated in the flight test of the PMG. After being told of the Loftus concept in a telephone conversation, his reply in an Email message dated Aug. 5, 1996, was "such a system would be feasible . . . by it is still not obvious to me that it would be useful . . . "

Loftus was considering the use of electrodynamic drag from a conducting tether to achieve this goal of bringing the unwanted spacecraft down from its high orbit (where atmospheric drag is negligible) to a 200-km orbit, where atmospheric drag would rapidly finish off the task of removing the unwanted spacecraft from orbit. The tether Loftus was considering was a single-line, conducting tether, typically 1-mm in diameter, 1-km long, and, if made of aluminum, 2-kg in mass. He would include means at the ends of the tether to contact the ambient space plasma around the Earth to complete the current loop.

Unfortunately it is probable that space impactors would sever the 1-mm diameter, 1-km long single-line tether proposed by loftus within a 1/e lifetime of four months. This would produce orbital debris rather than removing it. The motivation for this work is the NASA Safety Standard NSS 1740.14 "(Guidelines and Assessment Procedures for Limiting Orbital Debris." The relevant portion of the Standard starts on page 6-3: General Policy Objective-Postmission Disposal of Space Structures. Item 6-1: "Disposal for final mission orbits passing through LEO: A spacecraft or upper stage with perigee altitude below 2000 km in its final orbit will be disposed of by one of three methods." The method of interest is the atmospheric reentry option, Option a: "Leave the structure in an orbit in which, using conservative projections for solar activity, atmospheric drag will limit the lifetime to no longer than 25 years after completion of mission. If drag enhancement devices are to be used to reduce the orbit lifetime, it should be demonstrated that such devices will significantly reduce the area-time product of the system or will not cause the spacecraft or large debris to fragment if a collision occurs while the system is decaying from orbit."

The NASA standard applies only to NASA spacecraft and even then only to completely new spacecraft designs. New versions of existing designs are to make a "best effort" to meet the standard, but will not be required to change their design to do so. The Department of Defense has adopted the NASA standard with the same provisos. An Interagency Group report has recommended that the NASA standard be taken as a starting point for a national standard. It is NASA's recommendation to the Interagency Group that the safety requirement be phased in only as spacefaring nations reach consensus internationally, which is being done through the International Debris Coordination Working Group whose members are Russia, China, Japan, ESA, UK, India, France, Italy, and the US.

Thus, although the NASA Safety Standard in its present form is not the "Law", the existence of the standard means that some time in the future a similar requirement may be imposed on all spacecraft. This could result in major growth in future space tether business, with a sale to every non-geostationary spacecraft being "mandated" by government safety regulations, somewhat as the sale of seat belts and airbags for every car are mandated.

In fact, three of the companies planning to set up "constellations" of low to medium orbit communications: Teledesic, Iridium and Odessey have committed their companies to abide by the spirit of NASA Safety Standard 1740.14 by using one means or another to deorbit their spacecraft before they reach end of life.

Problems with Prior Art Tethers:

All electrodynamic tether designs proposed by the prior art teach that the tether should be operated at a right angle to the magnetic field through which the tether is moving. This is a problem because the electrodynamic force acting on the tether causes the tether to align itself with the magnetic field force lines. To overcome this problem the prior art teaches the use of a large ballast mass attached to the end of the tether and/or use of a very long (tens to hundreds of kilometers) tether. The large ballast mass is expensive to take to orbit because it replaces useful payload. The long tether sweeps a larger Area-Time-Product during its useful life and thus is more likely to impact other space objects, either debris or another spacecraft.

Another problem common to all proposed prior art tethers is tether instability. If the tether produces a large electrodynamic drag force, which is desirable because a large drag force will cause the satellite to deorbit quickly, then the tether will be dynamically unstable. This instability can cause the tether to lose its effectiveness, act uncontrollably and even wrap around the satellite or otherwise malfunction. Experts skilled in the art of tether design have opined that this dynamic instability is inherently unavoidable in any electrodynamic tether system. The prior art solution, such as that presently being used in the ProSEDS experiment, has been to use a large ballast mass to increase the stabilizing gravity-gradient force and/or to limit the electrodynamic drag of the tether to less that the maximum that could be produced. In the ProSEDS experiment, the conducting electrodynamic tether is five kilometers long. To insure stability, it will be augmented by a 20–35 kilometer long non-conducting tether, which to further have stability will have a 40 kilogram ballast end mass.

Yet another problem of all proposed prior art electrodynamic tether systems is how to radiate away the energy produced by the tether's operation. A satellite moving at an orbital velocity of 18,000 miles per hour has a kinetic energy of over 10,000 calories per gram. To put this amount of energy in an understandable perspective, it may be noted that when nitroglycerine explodes it produces about 1,500 calories per gram. Prior art designs of electrodynamic drag tethers teach the use of the electrical energy generated by the tether to charge batteries or operate electronics, with the excess energy being converted into heat by a resistive load. This excess heat must be radiated to the space environment or it will melt the resistive load. Thus the resistive load, and/or its associated radiator structures, must be massive and replace useful payload.

DISCLOSURE OF THE INVENTION

The present invention comprises an electrodynamic tether structure and a method of use. The principal industrial utility of the present invention is to deorbit satellites in Earth orbit at the end of their useful life. This embodiment of the present invention is sometimes referred to in this specification as a "Terminator Tether™" because it terminates the orbital lifetime of the host spacecraft. The structure of the tether taught by the present invention is a short, wide, interconnected-multiwire (compared to the long single wires of the prior art) conductive Hoytether whose area maximizes electrodynamic drag while simultaneously minimizing the Area-Time-Product swept by the tether during its operating life. The preferred tether length is two to five kilometers. The preferred tether mass is one to five percent (1%–5%) of the spacecraft mass. The method of operation comprises orienting the tether structure at a 35.26-degree trailing angle to the local vertical to maximize electrodynamic force on the tether while avoiding tether instability and allowing use of a small tether end mass.

The present invention also teaches that the satellite-tether system may be rotated around its common center of mass to centrifugally produce tension force in the tether structure to oppose forces causing tether instability. The angle of the conductive tether structure of the present invention with respect to the velocity vector of the host spacecraft may be controlled by the method of the present invention so it interacts with the encountered magnetic field to induce a maximum current flow in the tether. This produces maximum electrodynamic drag. All or a portion of this electric power may be stored and then controllably applied to the conductive tether to produce an induced electrodynamic force. This induced electrodynamic force may by used to enhance the drag force, to rotate the tether-satellite system and/or to provide satellite propulsion, i.e. to change the state vector of the satellite for any useful purpose, e.g. to avoid collision or to change the host spacecraft's orbit to an orbit more favorable for more rapid deorbiting.

The present invention also teaches a tether structure that also functions as a thermal radiator and/or plasma contactor. An embodiment of the present invention using conducting elements of the satellite, e.g. the solar arrays, as electrodynamic tether structures is also disclosed.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention may be better understood with reference to the following drawings:

FIG. **1***a* is a graph comparing the survival probability of a single line tether as taught by the prior art to the Hoytether used by the present invention.

FIG. **1***b* is a graph showing the level of electrodynamic thrust for a 10 kilometer, 10 kilowatt tether at various altitudes and inclinations.

FIG. **1***c* is a graph showing reentry time sensitivity of a five kilometer ProSEDS tether for several conditions of current flow in the electrodynamic tether.

FIGS. **2***a*, **2***b* and **2***c* show the Hoytether structure used by the preferred embodiment of the present invention.

FIG. **2***d* shows the Hoytape structure used by the preferred embodiment of the present invention.

FIG. **3** is a block electrical diagram of an electrodynamic tether system.

FIG. **4** is a graph showing a graph of the Area-Time product for three separate levels of exospheric temperature (for neutral drag calculations) and for two separate assumptions regarding tether mass and power drawn from the ambient plasma (for the Terminator Tether™ calculations).

FIG. **5** is the percent additional mass required in orbit to drop the perigee of a circular orbit at altitude, at, to a value of 200 kilometers using standard chemical rocket as taught by the prior art for deorbiting unwanted spacecraft.

FIG. **6** is a deorbit tether system shown operating with its electrodynamic tether at a trailing 35.26-degree angle to the encountered magnetic field, as is taught by the preferred embodiment of the present invention.

FIG. **7** shows an embodiment of the present invention in which the electrodynamic tether—satellite system is rotated about its center of mass to avoid tether instability.

FIG. **8** shows an embodiment of the present invention wherein the tether is powered to provide induced electrodynamic force.

FIG. **9** shows an embodiment of the present invention wherein the solar power system structure of the satellite is used as an electrodynamic tether.

FIG. **10** shows the use of a portion of the tether structure as a thermal radiator and plasma contactor.

FIG. **11** shows the braiding of an ohmic resistive load into the tether structure.

FIG. **12** is a force diagram showing the forces and torques on an upward deployed conducting tether due to the motion of the host spacecraft through the magnetic field of the Earth.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

The principal industrial utility of the present invention is the deorbiting of satellites in Earth orbit, which necessarily can happen only in outer space. Until the present invention is reduced to actual practice by use with satellites in outer space and a body of practical experience is thereby obtained, the present invention can only be constructively reduced to practice, such as by this written specification and its associated drawings, diagrams and graphs and by reference to electric measurements made on conducting tethers in space by the PMG and TSS experiments. The inventors provide herein a detailed discussion of the theory of the present invention to help those skilled in the art of aerospace and tether engineering to understand the present invention and to make and use the best embodiment of the present invention known to the inventors at the time this specification was prepared, without undue experimentation.

Although the present invention is discussed in this specification in its preferred embodiment as a means of deorbiting satellites, it must be emphasized that discussion of this specific use in no way limits the broad scope of the present invention. The present invention can be used to change the direction and speed, i.e. the state vector, of any space object providing that object is in motion relative to any magnetic field. This magnetic field may be that of the Earth or of any other celestial body, for example Jupiter or the sun.

In this specification the material forming the structure of the electrodynamic tether is specified as being a conductor. For example, a metal, such as copper or aluminum wire could be used. Likewise, nonmetallic conductor, such as carbon nanotubes, or a conductive polymer could form the conductive structure of the tether.

FIG. **1***a* shows the survival curve for a Hoytether as used by the present invention vs. the survival curve of a prior art single line tether. The survival probability curve of the Hoytether as a function of time does not conform to the well known "1/e decay" shape of a single-line tether. The Hoytether can maintain a high level of survival probability, about 99 percent, until it nears its 'lifetime'. Its survival probability then drops rapidly to zero. A detailed mathematical analysis of the difference between Hoytether and single tether survival probabilities is given in Appendix E "Small Impactor Survival Probabilities of Hoytethers" and Appendix F "Large Orbital Debris Survival Probabilities of Hoytethers", both contained in the Final Report of NASA Contract NAS8-40545. The resulting cut probability with time for the Hoytether has a "bingo curve" shapes. In a bingo game, at least five numbers must be called before anyone can win, and usually many numbers have to be called before one of the bingo cards gets five in a row. In the Hoytether at least four cuts must happen at the same level before any failure occurs, and many cuts have to be made before any one of the levels has all four lines cut. The bingo curve has the property that the probability of survival stays very high for periods short compared to the lifetime. The probability of survival is greater than 99% for periods shorter than 10% of the lifetime. This is much better performance than the 1/e curve of a single line tether, where the probability of survival is only 90% at 10% of the 1/e lifetime.

FIG. **1***b* shows the calculated level of electrodynamic thrust, in Newtons, produced by a ten kilowatt, ten kilometer single conductor tether at altitudes between 200 and 1400 kilometers and at orbital inclinations between zero degrees and eighty degrees.

FIG. **1***c* is a graph showing the reentry time sensitivity calculated for a five kilometer ProSEDS single conductor tether at altitudes between 200 and 500 kilometers with current flow in the tether between on flow and one ampere.

FIGS. **1***b* and **1***c* are the work of Enrico Lorenzini of the Smithsonian Astrophysical Observatory from his study "Performance Evaluation of the Electrodynamic Tether Tug", NASA/MSFC Grant NAG8-1303 (Nov. 12, 1966). They show that experts in the field believe that the present invention will have industrial utility, as the current flow of one ampere in FIG. **1***c* decreases the deorbit time for the calculated case from more than 180 days to less than 20 days.

FIG. **2***a* is an isometric drawing showing the generally cylindrical structure of the Hoytether. In FIG. **2***a*, primary lines **201** and **203** are shown connected via secondary lines **211**. In FIG. **2***b*, which is a schematic illustration of the structure shown in FIG. **2***a*, primary lines **201**, **203**, **205**, **207** and **209** are connected, each to their respective adjacent primary lines, by a plurality of secondary lines **221**. These connections, for example as shown at interconnection **213**, are made by knotless, slipless interconnections, such as Soutache braiding of twisted connections.

FIG. **2***c* shows the Hoytether structure of FIG. **2***b* but with a severed primary line **205** at breakpoint **215**. The second level of secondary lines **217**, shown as bold lines in FIG. **2***c*, redistribute the load from the severed primary line **215** at point **217** back to line **215** at points **219** and **221**, above and below the break, respectively.

For the electrodynamic tether, a preferred embodiment would be to unroll the tubular Hoytetherr™ into a flat Hoytape™ as shown in FIG. **2***d*.

FIG. **3** generally shows the arrangement of functional elements of an electrodynamic tether system. Of course such a system must have a means of attachment to the spacecraft, a deployer, the tether itself, some form of space plasma cathode contactor, some form of space plasma anode contactor, a power dissipation system, and a communication and control system. Each of these individual elements are known in some form to those skilled in the art of aerospace engineering. Thus the present inventors do not feel it is necessary to describe them in detail in this specification.

In FIG. **3**, a space object **301**, i.e. a satellite in Earth orbit, or any other space object either natural or man made, is physically connected to the tether system. The tether system comprises a deployer **303** from which a conductive Hoytether **305** having a bare segment **307** extends upward from space object **301**. The positively biased anode end **309** of tether **305** collects electrons from the ionosphere as space object **301** moves in direction **310** across the Earth's magnetic field. These electrons flow through the conductive structure of the Hoytether to the power system interface **311**, where it supplies power to an associated load, not shown. The electrons then flow to the negatively biased cathode **313** where electrons are ejected into the space plasma **315**, thus completing the electric circuit.

In order to allow those skilled in the art to better appreciate the broad scope of the present invention, the inventors will now provide an example using specific tether and system values to compare the Terminator Tether™ to prior art means, ie a rocket, for deorbiting a satellite.

Terminator Tethers™:

When an Earth orbiting conducting space tether moves through the magnetic field of the Earth, an electric field is generated in the tether that is proportional to the velocity of the conductor, the magnetic field strength of the Earth, and the angle between the conducting tether and the magnetic field lines. From data obtained during the various electrodynamic experiments that have been conducted in space to date, such as the PMG, TSS-1, and TSS-1R experiments, a typical value of the generated emf per unit length of tether of E=100–200 Volts/kilometer can be assumed. The electric potential V developed at the ends of a tether of length L is then V=EL. For a tether of length L=10 km, the electric potential developed is V=1000 V. For calibration, the 20-km long TSS-1R tether, at the moment of failure, was developing a potential of 3500 volts, which is 175 volts/km.

The mass of a conducting tether of length L, cross-sectional area A, and densin, d is given by m=dLA. Typical values for the density are d=2700 kg per cubic meter for aluminum and 8900-kg per cubic meter for copper. For a typical aluminum tether of mass m=10 kg and length L=10 km, the cross-sectional area will be A=0.37 mm squared. If this were a solid-wire single-line tether, the diameter of the wire would be D=0.69 mm (21.5 gauge). If this were an 18 line tubular Hoytether, the diameter of the lines in the Hoytether would be D=0.16 mm (34 gauge).

The resistance of a conducting tether of length L and cross-sectional area A is given by R=rL/A, where r is the resistivity of the conductor in nano-ohm-meters (nΩ-m). Typical values for the resistivity are r=27.4nΩ-m for aluminum and 17.0 -nΩ-m for copper. For the m=10 kg aluminum tether of length L=10 km and cross-sectional area A=0.37 mm squared at 20 degrees C, the end-to-end resistance is 750 Ω. This value of resistance is the essentially the same whether the conductor area is concentrated into a single-line tether or divided up into many lines as in a multiline Hoytether.

The current I generated in the conducting tether by the electric potential V between the ends of the tether applied across the tether resistance R is I=V/R. For the 10 kg mass aluminum tether of length L=10 km, electric potential V=EL=1000 volts, and resistance R=750 Ω the current is I=V/R=1.33 Amps. Currents near these values were measured in the TSS-1R experiment at the time of failure. At the time of failure of the TSS-1R tether, none of the plasma contactors on the Space Shuttle or the Italian Satellite were operating. Thus the current was being collected by the conductive surface area of the Space Shuttle and the Italian

Satellite at the two ends of the tether. The Space Shuttle area is quite large, so it was not the limiting factor in current collection. The diameter of the Italian Satellite was 1.6 m, which would give it an effective plasma contact area of about 8 square meters. This shows that, if the plasma contact area of the ends of a conductive space tether can be make large enough, then ampere level currents can be extracted from the ambient space plasma without the use of space plasma contactors.

The power dissipated as ohmic heating in the tether is given by P=IE. For an aluminum tether of mass m=10 kg, resistivity r=27 nΩ-m, and density d=2700 kg/cubic meter, subjected to an electric field of E=100 V/m, the power dissipated in the ohmic losses of the conductor is P=IV=1330 Watts. For a 1 kg mass tether, it would be still a considerable 133 Watts of dissipation.

There will no doubt be additional dissipation of energy in plasma ohmic losses, plasma wave generation, and plasma ion acceleration, but the ohmic losses in the conducting tether alone are sufficient for the task of deorbiting an unwanted spacecraft massing 100 to 1000 times more than the tether.

The decay time of a metric ton spacecraft moving from a 1000 km altitude orbit to a 200 km altitude orbit with an energy difference of dU=3.3GJ, when its energy is being dissipated at a power of P=1330 W by an aluminum tether massing just 10 kg, or 1% the mass of the spacecraft, is about one month. This is a remarkably short time, and indicates that the concept of using a conductive tether to deorbit a spacecraft is indeed feasible. If the aluminum tether massed only 1 kg, or 1/1000th the mass of the spacecraft it was deorbiting, then the decay time would rise to 10 months, still a reasonable value.

In reality, of course, the actual decay time will be longer than this. If the electrodynamic drag force is very large, and becomes larger than the gravity gradient forces pulling on the ends of the tether (which force is proportional to the mass of the tether), then the tether will tend to align itself along the magnetic field lines instead of across them, and the drag force will decrease because of the small angle between the conductor length and the magnetic field lines. The tether will then settle into an angle determined by the balance between these two forces.

Is the Terminator Tether™ theoretically a better means than atmospheric drag or a rocket engine for deorbiting satellites at the end of their useful life?

In the formula for da/dt (change in altitude per change in time), using the assumptions of near-circular spiral trajectories, the Area-Time-Product (Z), the criteria by which NASA judges compliance with Safety Standard 1740.14, is given by:

$$Z = A\int dt = -\frac{m}{C_D}\int\frac{da}{\rho(a)\sqrt{\mu a}},$$

where ρ(a) is simply the density as a function of semi-major axis. Thus, for a static atmosphere, the problem of area-time product is reduced to quadrature. Note that the area-time product depends only on the density profile and the ratio of spacecraft mass to drag-coefficient and is linearly related to that ratio.

The basic altitude-dependence of the atmospheric density, represented here by ρ(a) can be expressed in a low-order power series of ln(density)=f(ln(altitude)), so that it will be possible to develop a representation of the log—log relationship between altitude and density. The three static levels of atmospheric density are modeled as 5th order polynomials in the natural log of the altitude and are representative of well known atmosphere models, for example, Jacchia, L. G., "Thermospheric Temperature, Density, and Composition: New Model," SAO Special Report 375, March 1977. 2. Anon., Marshall Space Flight Center, "Long-Range Statistical Solar Activity Estimation," Atmospheric Sciences Division, 1989.) The values of exospheric temperature are reasonable (about 2-sigma) high and low solar activity values of 1400 and 800 kelvins respectively. The mean value is taken to be 1100 kevins.

FIG. **4** shows a graph of the Area-Time product for three separate levels of exospheric temperature (for neutral drag calculations) and for two separate assumptions regarding tether mass and power drawn from the ambient plasma (for the Terminator Tether™ calculations). The density profiles for the three values of exospheric temperature represent the extremes and mean values to be expected during the next several decades. The exospheric temperature goes through a cycle of about an 11 year period with maximum about 1400 kelvins and minimum about 800 kelvins.

The tether calculations include an inherent assumption that the Terminator Tether™ transfers all the energy extracted from the ambient space plasma into drag that decreases the orbital energy of the spacecraft at a rate given by the power drawn by the tether. Thus, the change in energy for a spacecraft of mass m at an initial altitude h is given by the expression:

$$\Delta E = -m\left\{\frac{\mu}{2(R_e+h)} - \frac{\mu}{2(R_e+250)}\right\},$$

and the time required to effect this change in energy of the mass, m, is just

$$\Delta t=\Delta E/P,$$

where P is the power drawn by the tether from the ambient plasma charge.

The curves for 1% and 5% tethers do not include the effects of atmospheric drag because the electrodynamic drag is orders of magnitude greater than the neutral atmospheric drag. At altitudes greater than about 700 km, the electrodynamic drag is 200 to 3000 times greater than the neutral drag forces.

It is clear, from FIG. **4** that the Terminator Tether™ concept is far superior to neutral drag in removing spacecraft from orbit, no matter how much additional area is added to the passive spacecraft to increase the atmospheric drag. The analyst should note that FIG. **4** is proportional to mass; that is, the Area-Time-Product values should be multiplied by the mass of the spacecraft and divided by 1000 kg. This is true for the neutral drag and Terminator Tether™ curves alike. The power levels assumed for the 1% and 5% tethers are only 80% of their theoretical values. This is done to provide a 20% margin on the power available from the ambient plasma and electrodynamic gradient.

Conventional rocket mechanisms can remove spacecraft from orbit, but this mechanism is apparently not viable from mass considerations when these factors are compared with the capabilities of the Terminator Tether™. A satellite owner may decide to satisfy the NASA safety requirements by adding a small solid motor and the associated hardware, software, sensors and structure to make the package independent, to be used to deorbit the spacecraft in case the main spacecraft power, attitude, or propulsion system fails.

The requirements of such a rocket motor system are more stringent than those attributed to ordinary spacecraft. The rocket motor deorbit system must operate when some or all other systems of the spacecraft have failed. These more stringent requirements are balanced by lesser requirements of performance. The backup system must simply deorbit the spacecraft; it does not have to perform all the other duties of the spacecraft. But the backup system must know when to fire under all kinds of anomalous situations, including tumbling, offset of center of mass (because of loss of parts due to collisions), and lack of knowledge of the orbital position.

FIG. **5** shows the percent additional mass required by the rocket motor system to drop the perigee of a circular orbit at some initial perigee altitude to a value of 200 km. The atmospheric drag at this 200 km perigee altitude will remove any spacecraft (in the range considered) from orbit in a few revolutions. The contours of constant stage propellant mass fraction, λ, range from low values of 0.5 through reasonable values of 0.65 to 0.7, up to the values associated with the best solid motors ($\lambda\nabla 0.93$) without adding any extra hardware to the emergency stage. An effective, independent stage to provide a retro deltaV of from 50 to 325 m/s will almost certainly have a λ of the order of 0.6 to 0.75. If the emergency stage is required to perform its own attitude determination, the stage propellant mass fraction may be as low as 0.55 or 0.50. Note that the additional mass, shown in FIG. **5**, must also be lofted to orbit in the first place, to provide the mass on orbit for the originally intended service. The stage propellant mass fraction, λ, is a key stage performance parameter that describes the ratio of the mass of propellant to the mass of the stage. That is $\lambda=m_p/(m_p+m_I)$, where m_p is the mass of the propellant and m_I is the mass of everything else in the stage (not including the payload or any stages above or below the stage being considered). Typical independent small rocket stages have values of λ from 0.60 to 0.75. The curves of stage propellant mass fraction are shown as straight lines (a minor approximation) and are included only from 0.5 to 0.95, the extremes of reasonable design practice.

These calculations that show, beyond any reasonable doubt, that the Terminator Tether™ concept is far superior to conventional mechanisms such as drag enhancement devices or small rocket deorbit propulsion systems. The superiority is measured in terms of Area-Time-Product, NASA's measure of the likelihood of collision with other spacecraft in the path of the descending spent member of a constellation. Tether calculations were made using conservative assumptions that the power extractable from the ambient plasma and electrodynamic gradient is only 80% of the theoretical power available to a perfect tether crossing the magnetic field lines at a right angle, i.e. normally.

The following analysis is presented by the inventors to help those skilled in the art to better understand the present invention when it is used as a Terminator Tether™ for removing from orbit unwanted non-geostationary Earth-orbiting spacecraft at the end their useful lives. The primary result of that analysis is that the electrical power P in the tether that is converted into heat by the resistance of the tether and radiated away into space will remove energy from the spacecraft, causing it to rapidly deorbit, thus reducing the amount of orbital space debris that must be coped with in outer space. Tether electric power P is given by:

$$P=(vB)^2m/2rd$$

where m is the mass of the conducting tether, r and d are the resistivity and density of the conducting material, and v is the velocity of the spacecraft's motion through the Earth's magnetic field B. For a m=10 kg tether of aluminum with resistivity of r=27.4 nΩ-m and density d=2700 kg/m^3, moving at a velocity v=7037 m/s relative to the Earth's horizontal magnetic field B=26.5 μT, the power dissipated is P=2350 W! This energy loss in the form of heat must necessarily come out of the kinetic energy of the host spacecraft. For a typical example, a 1000 kg spacecraft in a 1000 km high orbit subjected to an energy loss of 2350 J/s from a 10 kg tether (1% the mass of the host spacecraft) will be deorbited in a few weeks.

Power levels of the magnitude estimated in the previous paragraph have been measured in a real orbital space experiment, the TSS-1R mission carried out on the Shuttle Orbiter in 1995. In that experiment, a large Italian spacecraft, 1.6 m in diameter, was deployed upward from the Shuttle Orbiter at the end of a conducting copper wire tether covered with electrical insulation. As the tether was slowly deployed upwards, a series of measurements were made of the open circuit voltage induced in the tether by its motion through the Earth's magnetic field. The voltage between the end of the tether and the Orbiter ground varied from zero volts at the start to 3500 V when the amount of tether deployed approached its maximum length of 20 km. Periodically, the end of the tether was connected either to one of two different electron guns, which supplied contact to the surrounding space plasma, or to the Orbiter ground. The bare surfaces of the Shuttle Orbiter proved to be a surprisingly good plasma contactor via a combination of ton collection and secondary electron emission. The current flow through the tether was deliberately limited by control circuits and the current capacity of the electron guns to about 0.5 amperes, but power levels of 1800 Watts were reached.

The tether was intended to have a fully deployed length of 20 km, but at a deployed length of 19.5 km, when about 3500 V was being induced at the end of the tether inside the Orbiter reel mechanism, a flaw in the insulation allowed an electrical spark to jump in an uncontrolled manner from the tether to the Orbiter ground. With no control circuits to keep the current level down to 0.5 amperes, the current flow jumped to 1.1 amperes, and the total power generated was P=3850 Watts Most of this energy went into the electrical arc, which burned through the tether, causing it to break and halting the experiment. This experiment showed that large areas of bare conducting material, such as that provided by the 8 square meter area of the Italian spacecraft at one end of the tether and the very large surface area of the Shuttle Orbiter spacecraft at the other end of the tether, can collect and emit amperes of current, while thousands of volts of potential can be generated by sufficiently long tethers moving at orbital speeds.

Thus, both theory and experimental data collected in the space environment indicate that significant amounts of electrodynamic drag force can be obtained from a low mass conducting tether attached to a host spacecraft, provided the ends of the conductor can exchange sufficient numbers of electrons with the surrounding space plasma.

Experimental data from the TSS-1R data also produced the amazing result that the efficiency of a bare metal surface in "contacting" the space plasma is many times better than the standard theory would predict. The 8 square meters of bare surface area of the Italian spacecraft were sufficient to collect the 1.1 A of electron current. This amount of area is easily replicated by a few hundred meters of bare wire, considering that the effective collection diameter around the wire is the Debye length, which is a few centimeters at the typical values for space plasma density and temperature.

Because of this result, that a bare wire can easily collect electrons, Les Johnson, Nobie Stone, Chris Rupp, and others at NASA Marshall Space Flight Center have formed a team, which includes the present inventors, which is embarked on a new flight experiment. The experiment is scheduled for a piggy-back flight on a Delta II launch of an AF Global Positioning Satellite in early 2000. The goal of the experiment is to demonstrate that electrodynamic drag from a wire moving at orbital speeds through the Earth's magnetic field will create a large enough electrodynamic drag force to deorbit the Delta II second stage, whose mass is greater than 1000 kilograms, in a few weeks. This is essentially a demonstration of the Loftus electrodynamic drag deorbit concept and the first step in the development of a Terminator Tether™.

The ProSEDS (Propulsion Small Expendable-tether Deployer System) mission will use a 5 km long copper wire conductive tether massing 18 kg connected to a 20–35 km long nonconducting tether, which is in turn connected to a 25–40 kg ballast mass. The total of 25–40 km of tether length and the 25–40 kg ballast mass on the end will provide enough gravity gradient force to keep the tether aligned near the zenith, so that the direction of the current in the tether is at right angles to both the direction of the spacecraft motion in the nominal EW direction and the Earth's near-equatorial magnetic field in the nominal NS direction.

An important feature of the ProSEDS experiment is that it is designed to be completely self-powered. It uses a battery to initiate deployment and to power up the plasma contactor, but once current is flowing through the tether, some of the power is tapped off and used to recharge the battery. The battery, in turn, powers the current control electronics, the telemetry, system, and the plasma contactor. The ProSEDS mission will not be designed to allow ground control changes in operation, primarily because of the increase in complexity and cost associated with that option.

The present invention is the use of a small, low-mass deployer/controller package containing a large collecting area, short length, multiline space tether, such as a Hoytape mesh made of aluminum wire, as a "Terminator Tether™" for a constellation spacecraft. The Terminator Tether™ would be deployed when the host spacecraft is no longer working or no longer wanted. The electrodynamic drag from the Terminator Tether™ would rapidly remove the unwanted spacecraft from the constellation and a few weeks later complete the deorbit of the host spacecraft from space by burnup in the upper atmosphere of the Earth. For a Terminator Tether™ to be of maximum usefulness for constellation spacecraft, it would be desirable to minmize the mass and the length of the tether. A lower added mass means more mass for revenue producing transponders, while a shorter tether length means a lower collision cross-section Area-Time Product during deorbit.

FIG. **6** shows a spacecraft **600** having a state vector **602** that causes the spacecraft to move across magnetic field lines **601**. Spacecraft **600** has a conductive tether **603** attached to it by tether deployer and control system **605**. The outer end of tether **603** is attached to tether end mass **607**. Spacecraft **600** also has an attached electron emitter means **613**.

In FIG. **6**, spacecraft **600** has a velocity and direction defined by state vector **602**. As spacecraft **600** moves it causes conductive tether **603** to cut magnetic field lines **601**. As this happens the distal end of the conductive tether collects electrons, shown as arrows **611**, from the space plasma. Electrons **611** move through the conductive tether **603**, are passed through a resistive load, not shown, in control system **605**, and are emitted back into the space plasma as electrons **615** from electron emitter **613**. This produces an electrodynamic drag in the direction shown by arrow **609**.

As is discussed in detail in the electrodynamic drag analysis below, especially in that sections of the analysis titled "Optimization of Tether Angle", the maximum electrodynamic drag of the tether is achieved when the angle, shown as the Greek letter τ in FIG. **6**, is 35.26 degrees, trailing, to the direction of the magnetic field lines. This angle may be maintained by a feedback control circuit that maximizes the electrodynamic drag of the tether system, as is discussed in detail below. The tether structure **603** is a Hoytether having a length of about 2 to 5 kilometers. The use of the 35.26 degree angle to the magnetic field reduces the tether instability, allowing the use of a short tether which will have a small Area-Time-Product. This will minmize the possibility that the tether will impact another space object during its use.

In FIG. **7**, spacecraft **700** has a state vector **702** across magnetic field lines **704**. Conductive tether **706** is attached to spacecraft **700** by tether deployer and control system **708**. The distal end of tether **706** is attached to a tether end mass **710**. Spacecraft **700** has an electron emitter means **714**.

Functionally, electrons **712** flow from the space plasma into conductive tether **706**, are passed through a resistive load in control unit **708** and are emitted into the space plasma by electron emitter **714** as free electrons **716**. As is discussed in detail in the analysis below, this produces an electrodynamic drag on satellite **700**, which causes it to deorbit.

In FIG. **7**, spacecraft **700** is rotating with an angular velocity indicated by arrow **718**. This rotation causes a centrifugal force **720** to place tension on conductive tether **706**. This force **720** places tension on the tether **706** in a direction that counters tether instability. The Tether **706** may be at any angle to the magnetic field while the rotation of the tether-satellite system is producing this useful tension. FIG. **7** shows the angle to be 35.26 degrees, which is optimum, but any angle will work.

FIGS. **8***a* and **8***b* show a powered Terminator Tether™. In FIG. **8***a* the spacecraft **800** has a state vector **802** that causes it and its associated conductive tether **806** to move across magnetic field **804**. Tether **806** has an end mass **808** that is proximate a plasma contactor **810**. Just as was discussed in the embodiment of the present invention shown in FIG. **6** and FIG. **7**, an electric current moves through the circuit formed by the space plasma, the plasma contactor **810**, the conductive tether **806**, and the electron emitter **815**, thence back into the space plasma as electrons **816**. This creates an electrodynamic drag on the tether in the direction shown by arrow **818**.

In FIG. **8***b*, similar numbers indicate similar structures. In FIG. **8***b*, however, an electric power source, not shown, in spacecraft **800** provides electrons to conductive tether **806**. The electrons are emitted from the end **812** of the tether **806**. Electrons are collected to from the space plasma **814** by contactor **815**, thus completing the electric circuit. The result is an electrodynamic force on tether **806**, and therefore on spacecraft **800**, in the direction show by arrow **820**.

Spacecraft **800** in FIG. **8***b* may be rotated as is taught in the embodiment of the invention shown in FIG. **7**, above, to allow the control system to time the application of electric current to the tether **806** to cause force **820** to be exert force in any desired direction along the circle made by the rotating tether. This allows the state vector of the spacecraft to be modified to raise or lower the spacecraft's orbital altitude, or to otherwise usefully change its orbital elements.

FIG. **9** show an embodiment of the present invention wherein the conductive structure of a satellite is used as the tether. In FIG. **9**, a spacecraft **900** which has a state vector

901 and is moving across magnetic field **903**, has two attached solar panels, **902** and **904**. Solar panels **902** and **904** have conductive aluminum frames **908** and **906**, respectively. Frame **906** is electrically and mechanically connected to plasma contactor **912**. Frame **908** is electrically and mechanically connected to electron emitter **910**. As spacecraft **900** moves across magnetic field **903**, electrons from the space plasma **914** flow through contactor **912** and frame **906** to the body of spacecraft **900**, where the current flows through a resistive load, not shown, and then into frame **908** and electron emitter **910** to the space plasma **916**, thus completing the circuit. This creates an electrodynamic drag on spacecraft **900**. If the power output from solar panels **902** and **904** was directed to flow into the frames, **906** and **908**, then the electrodynamic interaction with the space plasma would cause a useful force to be impressed on spacecraft **900**, as was discussed in connection with FIG. **8** above. The spacecraft may be rotated and/or oriented to a desired angle with the magnetic field lines, as was also discussed above.

FIG. **10** shows an embodiment of the present invention wherein the plasma contactor and the resistive load are structurally part of the conductive tether. In FIG. **10** spacecraft **1000** has a state vector **1001** that causes it to pass through magnetic field **1003**. Spacecraft **1000** has a tether control and deployment system **1005** that is connected to a conductive tether **1007** having an end mass **1013**. A contactor portion **1015** of the conductive tether **1007** near end mass **1013** is adapted to contact the space plasma so as to receive or emit electrons. This may be the bare wire strands of a wide Hoytether, or it may be any other plasma contactor means that can be conveniently made electrically part of the Hoytether structure, such as points or metal fuzz. A second resistive portion **1009** of tether structure **1007** is adapted to be a resistive load. This may be done by making a portion of the tether structure from a material, such as nichrome steel wire, that has suitable electrical resistance The large surface area of the Hoytether provides an excellent radiator structure into the three degree Kelvin radiation sink of outer space.

FIG. **11** shows an alternative embodiment of the load resistance/radiator structure of the present invention. In FIG. **11**, a biline Hoytether **1002** is attached to its control and deployer system **1100** on a spacecraft, not shown, and to a tether end mass **1004**. A resistive load, **1006** is woven into the Hoytether structure, whereby this load **1006** and the section of the Hoytether **1002** into which it is woven act together as a thermal radiator allowing photons **1008** to radiate into the thermal sink of outer space.

The following detailed analysis of the present invention will help those skilled in the art to better understand and use the present invention.

Electrodynamic Drag Analysis:

The following is an optimization analysis of the electrodynamic drag produced by a conducting tether deployed from a host spacecraft that it is to deorbit. The force and torque diagram used in the analysis is shown in FIG. **12**, which shows the forces and torques on an upward deployed conducting tether due to the motion of the host spacecraft through the magnetic field of the Earth.

Physical Constants and Assumptions:

The analysis that follows will use the following physical constants and assumptions:

Newton's gravitational constant $G=6.67\times10^{-11}$ $m^3/kg\text{-}s^2$.

Mass of Earth $M_e=5.976\times10^{24}$ kg.

Radius of Earth near equator $R_e=6378$ km.

Assumed host spacecraft altitude h=622 km.

Assumed host spacecraft orbital radius a=7000 km.

Host spacecraft orbital velocity $w=\omega a=[GM_e/a\ ]^{1/2}=7546$ m/s.

Vertical gravity gradient at spacecraft $2\Gamma=2GM_e/a^3=2.32\times10^{-6}\ s^{-2}$.

Centrifugal gradient at spacecraft $\omega^2=GM_e/a^3=1.16\times10^{-6}\ s^{-2}$.

Combined gradient at spacecraft $3\Gamma=3GM_e/a^3=3\omega^2=3.49\times10^{-6}\ s^{-2}$.

Magnetic field of Earth (tilted dipole approximation):

Magnitude $B_o=35\ \mu T=0.35$ gauss.

Angle between magnetic pole and spin pole 11.5°.

Field components at spacecraft with altitude a and angle β between radius vector of spacecraft and magnetic equatorial plane.

Horizontal $B_H=B_o(R_e/a)^3\cos\beta=26.5\ \mu T\cos\beta$.

Vertical $B_V=2B_o(R_e/a)^3\sin\beta=52.9\ \mu T\sin\beta$.

Electrodynamic Tether Constants and Assumptions:

The electrodynamic tether is assumed to be made of a conducting metal, and have a length L, density d, resistivity r, and cross-sectional area A that is constant along the length of the tether. If the tether is a single round wire of diameter D, then the cross-sectional area is $A=\pi D^2/4$. Because of the micrometeorite and space debris hazard, however, it is likely the tether will be made up of redundantly interconnected multiple lines whose individual cross-sectional areas add up to A. Given these assumptions, the tether mass is then $m_T=dLA$, while the end-to-end tether resistance is $R_T=rL/A=rdL^2/m_T$.

Specific Conductivity Parameter:

The choice of the metal conductor to be used in a space tether is determined by a combination of low resistivity (high conductivity) and low density, with cost, strength, and melting point as secondary considerations for certain applications. Copper has a resistivity r=17.0 nΩ-m, a density d=8933 kg/m, and a "specific conductivity" of 1/rd=6,585 m^2/Ω-kg. Aluminum has a resistivity r=27.4 nΩ-m, which is significantly greater than that of copper, but it has a much lower density of d=2700 kg/m^3. As a result, aluminum's "specific conductivity" of 1/rd=13,500 m^2/Ω-kg is twice the conductivity per unit mass of copper. Silver, because of its higher density and higher cost, is not competitive as an electrodynamic space tether even though its resistivity of 16.1 nΩ-m is slightly better than that of copper. An alternate candidate material would be beryllium, with a resistivity r=32.5 nΩ-m, density d=1850 kg/m^3, and a "specific conductivity" of 1/rd=16,630 m^2/Ω-kg, slightly better than that of the much cheaper aluminum. Beryllium also has a higher melting point at 1551 K than aluminum at 933 K, so some of its alloys may be a preferred material for some electrodynamic applications despite its higher materials cost. Unfortunately, despite decades of metallurgical research by the nuclear power industry, highly ductile alloys of beryllium have not been found, so it is difficult to pull beryllium into wire. As a result, because of its high specific conductivity, low cost, and ready availability in ductile wire form, it will be assumed for this analysis that the electrodynamic tether will be made of aluminum wire.

Typical Resistance Values:

To be economically competitive, the mass of the tether needs to be a small fraction of the mass of the host spacecraft it is required to deorbit. Since a typical constellation spacecraft has a mass of about 1000 kg, a typical Terminator Tether™ with a mass that is 2% of the host spacecraft mass would consist of a deployer/controller package with a mass $m_D=10$ kg, containing an aluminum tether with a mass $m_T=10$ kg with a volume of $LA=m_T/d=3.70\times10^{-3}\ m^3$. If this 10 kg of aluminum were formed into a tether with a length of $L=2$ km and a cross-sectional area of $A=1.85$ mm^2, then the end-to-end resistance of the tether would be $R=rL/A=rdL^2/m_T=29.6\ \Omega$ A longer tether would have a proportionately smaller cross-sectional area and a higher resistance; for example, a 5 km long tether with the same mass would have a resistance of 185 Ω.

Orbit Inclination Assumption:

In order to make the remainder of the analysis mathematically tractable, it will be assumed that the orbit of the host spacecraft is circular and above the magnetic equator, so that the angle between the radius vector to the spacecraft and the magnetic equatorial plane is $\beta=0$ degrees. In this orbit, the velocity v of the spacecraft with respect to the magnetic field is the orbital velocity w less the rotational velocity of the magnetic field at the orbital radius a due to the once per day rotation of the Earth, or:

$$v=W-2\pi a/1\text{day}=(7546-509)\text{m/s}=7037\text{m/s}$$

This equation also shows why electrodynamic drag will not be useful for removing geostationary spacecraft from orbit. At the geostationary orbital radius of a=42,200 km, the relative velocity of the spacecraft and the rotating magnetic field of the Earth is zero.

With this assumption of an orbit above the magnetic equator, the vertical component of the Earth's magnetic field is zero and the horizontal component of the Earth's magnetic field is at right angles to both the local vertical and the direction of motion of the spacecraft. The total magnetic field seen by the host spacecraft and its Terminator Tether™ is then orthogonally horizontal and has the magnitude:

$$B=B_H=B_o[R_e/a]^3\cos\beta=26.5\mu T.$$

Where for simplicity the subscript H will be dropped for the remainder of this analysis.

This orbit, with its inclination of 11.5° with respect to the spin equator, will not stay in the plane of the magnetic equator, but will vary ±11.5° above and below it as the Earth rotates, causing a variation in magnetic field strength of ±0.5 μT, as well as a slight variation in angle. This 2% variation is negligible in terms of the other uncertainties in this analysis. It is not until the orbital inclination of the host spacecraft orbit reaches 60°, where cos 60°=0.5, that there is a significant drop in the expected magnitude of the electrodynamic drag forces calculated in this analysis.

Electromagnetic Drag Effects in Polar Orbit:

In many medium Earth orbit communication satellite constellations, there are a significant number of spacecraft at high inclinations and in nearly polar orbits. The high inclination spacecraft, with inclinations between 60° and 78.5°, will all have orbits that stay between the magnetic poles. Although the amount of electrodynamic drag will be significantly less than that experienced by spacecraft with orbits at lower inclinations, the direction of the induced electric fields in the tether will always be in the proper direction. If the tether is vertically upward, the outer tip of the tether will be positively charged and the bare wire in the tether will pull electrons out of the surrounding space plasma, while the electron emitter at the host spacecraft end will eject the electrons back into the space plasma to complete the circuit.

For spacecraft in near polar orbits with inclinations between 78.5° and 90°, however, there will be much more drastic variations. First of all, for a spacecraft in a 90° orbit that happens to pass directly over the magnetic poles, there is no horizontal component of the magnetic field when it is passing over one of the magnetic poles, so no voltage is generated in the tether during that part of the orbit, while the horizontal component of the magnetic field near the magnetic equator, while strong in magnitude, is along the spacecraft velocity vector, so no voltage is generated in the tether in that part of the orbit either. As a result, there will be negligible electrodynamic drag experienced by the host spacecraft during that particular orbit. The Earth is rotating, however, and the magnetic pole is rotating with it. A few orbits later, the host spacecraft will be passing over the Earth's spin pole at a point where the horizontal component of the magnetic field is exactly at right angles to the direction of motion of the spacecraft, so the full voltage is generated in the tether. When passing over the spin pole, the spacecraft radius vector is at an angle of 11.5° from the magnetic pole and at an angle $\beta=78.5°$ away from the magnetic equatorial plane. With these assumptions, the magnitude of the horizontal component of the magnetic field at the Earth's spin pole is a respectable:

$$B=B_o[R_e/a]^3\cos 78.5°=5.28\mu T$$

or 20% of the maximum value experiences by spacecraft orbiting above the magnetic equator. This value will drop slightly as the orbit continues, then build back up as the spacecraft passes over the opposite pole. As the Earth continues to rotate, bringing the magnetic pole again under the orbit, and the interaction of the tether with the horizontal component of the magnetic field again drops to zero. The average coupling of a tether to the Earth's magnetic field over all polar trajectories has been calculated to be 12.3 percent.

A problem experienced only by spacecraft with orbit inclinations greater than 78.5°:

The spacecraft will no longer be traveling from west to east with respect to the magnetic field axis, but will have a retrograde motion as it moves through the magnetic field. As a result, the voltage generated by the motion of the tether through the Earth's magnetic field will switch direction. The outer tip will be negatively charged and will attempt to collect ions, which is a much less efficient process than collecting electrons.

There are a number of solutions to this problem. The first is to increase the mass and length of the tether supplied to a spacecraft assigned to a polar orbit, so that higher voltages, currents, and drag are generated during the limited times the spacecraft is passing over the poles in the right direction. The second is to supply a tether with plasma contactors at both ends that can emit electrons from either end, allowing the current to flow either way, depending upon which direction the spacecraft is passing around the magnetic pole. The third is to utilize the first number of passes to torque the orbit of the spacecraft until the orbit inclination has been shifted below 60°, then turn on the electrodynamic drag full time to deorbit the spacecraft from this more favorable orbit inclination. This orbit torqueing maneuver is accomplished by activating the electrodynamic drag mechanism only when the magnetic field orientation is such that a strong out-of-plane component of force is created. This orbit torqueing maneuver can be augmented by switching to a propulsion mode, where power saved in the batteries during the drag force mode is pumped back into the tether when the magnetic field is in the opposite direction, applying electrodynamic propulsion to torque the orbit even further. With the tether at a large and stable trailing orientation, the coupling to the magnetic field can be significant to a vertical tether.

In summary, spacecraft in near-polar orbits might take longer to bring down, and might have to utilize specially designed Terminator Tethers™ that might cost and mass more than the simpler Terminator Tethers™ usable in lower inclination orbits, but a Terminator Tether™ can still remove a spacecraft from a polar orbit when desired.

The inventors now present a detailed analysis and optimization of a typical Terminator Tether™ attached to a typical host spacecraft in a typical low inclination orbit.

Electromotive Generation of Voltage and Current in the Tether:

When an object is moved at a velocity v through a magnetic field B, an electric field is generated in the frame of reference of the moving object given by:

$$E = v \times B = vB$$

where the magnetic field B of the Earth, being mostly tangent to the Earth's surface in the north-south direction, is at right angles to the velocity vector v of the spacecraft, assumed to be orbiting in a generally west-east direction. The direction of the electric field E will be at right angles to both v and B, or along the local vertical. It should be noted that this electric field exists in the moving frame of reference because a moving magnetic field creates an electric field. No object actually has to be there, but if it is, then the relative motion of the magnetic field of the Earth will not only apply magnetic forces to whatever material the object is made out of, but electric forces too.

Note also that the velocity used in this equation is the relative velocity between the object and the magnetic field. Because the Earth's magnetic field rotates with the Earth, the motion of the magnetic field must be subtracted from the orbital velocity of the object to obtain the relative velocity.

Voltage generation in a Conducting Tether:

If the moving object is a long conducting wire of length L, the electric field E, generated in the wire produces a voltage V between the opposite ends of the wire given by:

$$V = E \bullet L = EL \cos \tau = vBL \cos \tau$$

where τ is the angle between the length vector L of the tether and the electric field vector E, assumed to be in the vertical direction at right angles to the velocity vector v in the plane of FIG. **12** and the magnetic field vector B out of the plane of FIG. **12**. A typical value for the voltage level built up in a vertically oriented tether with length L=5 km and $\tau=0$, moving at a velocity of v=7037 m/s through the Earth's horizontal magnetic field of strength 26.5 μT, is 932 V, or 0.186 V/m. Spacecraft in higher inclination orbits would experience somewhat smaller electric fields. For calibration, the voltage measured between the ends of the 19.5 km long TSS-1R tether was 3500 V, or 0.175 V/m.

Contacting the Space Plasma:

Although a voltage will build up between the ends of the conducting tether, no current will flow unless the circuit is completed. The circuit cannot be completed with another wire, for it too will have a similar voltage generated in it by the moving magnetic field. Fortunately, empty space is not empty, and in near-Earth regions not too distant from the Earth's atmosphere there exists highly electrically conductive space plasma, kept partially ionized by radiation from the Sun. The electron and ion density varies from the dark to light side, with altitude, with season, with sunspot cycle, with contamination level, etc. but it typically varies from 10^{11} to 10^{13} electrons/cc.

Fortunately, it has been found that if a bare conductive surface such as the spherical Italian spacecraft in the TSS-1R experiment, or the long bare wire to be used in the ProSEDS experiment, or a multiline conducting wire Hoytape mesh, is charged to a few hundred volts, the bare conductor will readily pull electrons out of the space plasma. So all that is needed to complete a connection to the space plasma at the positively charged (upper) end of the tether is a sufficient large area of uninsulated conductor.

It is more difficult to eject electrons from a wire or to collect positive ions from the space plasma. Although it is conceivable that a very large area at the other end of the tether could collect enough ions to complete the circuit as was demonstrated during the TSS-1R mission when the Shuttle orbiter was found to be an adequate plasma contactor for over an amp of current, the present method chosen is to use an electron emitter of some sort, either a hot cathode, a plasma cathode or contactor, a field-emission device, or something similar. Once provisions have been made at both ends of the tether to allow the flow of electrons out of one end of the tether and into the other end, and the altitude of the host spacecraft is not too high, then there will be sufficient conductivity in the space plasma surrounding the host spacecraft to allow current to flow through the tether.

Current Flow In The Tether:

The amount of current I flowing through the tether depends upon the total resistance R in the circuit. This resistance will consist of three components, the effective resistance of the plasma, the resistance of the tether, calculated earlier as $R_T = rL/A = rdL^2/m_T$, and a control resistor R_C, which will be varied as needed to optimize the Terminator Tether™ performance. There will also need to be a parasitic load on the current in the form of a charging device to charge some batteries. The batteries in turn will be used to power the control and communication circuits, and drive the electron emitting devices at the negative end of the tether. A well-designed Terminator Tether™ will thus be completely self-powered, except for an initial charge in the batteries to provide electrical power for the deployment and startup procedure. For simplicity of analysis it will be assumed that this battery charging load, which absorbs power like a resistor, but which stores it and uses it later instead of dissipating it immediately as heat, is included in the control resistor R_c. Normally, this load would act as a "base resistance" below which the control resistor could not be lowered, but since the charging circuit can be turned off when desired, and operations continued without interruption using the power stored in the batteries, it will be assumed that the control resistor can be taken to near zero value in those circumstances where the space plasma conductivity is low, or the magnetic field is in the wrong orientation and the voltage being generated in the tether is not large. Under these circumstances, lowering the control resistor to near zero allows a much higher current to flow for a given generated voltage, thus increasing the power being dissipated in the tether and maintaining a high level of electrodynamic drag on the host spacecraft.

A properly designed Terminator Tether™ will have plenty of bare metal area for electron collection at the positive end, while the electron emitters at the other end are efficient in terms of emitting large electron currents at low voltages and therefore low powers, while at the same time the mass, length, and area of the conducting tether have been made such that the resistance of the tether is moderately high. Under these conditions, the effective resistance of the space plasma will be much less than the design resistance of the tether. To make the mathematics more tractable it will be assumed that the plasma resistance is can be neglected and that the voltage available to drive the tether is the full voltage generated by the moving magnetic field. Although voltage will be needed to power the electron emitter, whatever its form, it will be assumed that the required voltage and power will be supplied by the batteries. The batteries in turn will be powered by the battery charging circuit, whose effective resistance is included in the control resistor (except for those short periods when the plasma or magnetic field interactions are weak).

Given all these assumptions concerning the total resistance in the circuit, the current I flowing through the tether is then given by:

$$I=V/R=vBL\cos\tau/(R_C+rdL^2/m_T)$$

A typical current level can be estimated by assuming that the control resistance R_C should be about the same size as the tether resistance R_T, which for an aluminum tether with resistivity r=27.4 nΩ-m, density d=2700 kg/m^3, length L=5 km, and mass m_T=10 kg, is R_T=185 Ω, giving a total resistance, including the control resistance R_C, of R=370 Ω. Thus the current flowing though a tether of length L=5 km, generating a voltage of V=932 V, would be I=V/R=2.52 A.

Power Dissipation Analysis:

The induced current I flowing through the tether resistance R_T will generate heat in the wire, which will be radiated into space and lost. That radiated energy must come from somewhere. There is also power taken out of the current flow by the battery charging circuit and ultimately dissipated. In order to conform to energy conservation laws, this power and energy must come from a decrease in the total kinetic and potential energy of the host spacecraft, causing it to deorbit. Set forth below is a detailed, optimized force analysis which will calculate the drag force on the host spacecraft. In this section, by using some simplifying assumptions, it is possible to produce a general argument based on energy conservation laws, which will illustrate the broad scope of the present invention.

For this energy conservation analysis it will be assumed that the plasma resistance is small compared with the tether resistance, and that the control resistance is equal to the tether resistance:

$$R_C=R_T=rL/A=rdL^2/m_T$$

Assume that the ballast mass at the end of the tether is a large piece of the defunct host spacecraft, such as a solar panel, antenna, or battery pack, so that the gradient force is large and the tether is always oriented along the vertical so that the angle τ=0°. Also assume the spacecraft is in an orbit above the magnetic equator and thus is moving at right angles to the Earth's magnetic field, so that the angle between the orbital velocity vector and the magnetic field vector is 90°, and the electric field vector is in the vertical direction, aligned perfectly with the tether length vector. Under these ideal conditions, the voltage between the ends of the tether is given by:

$$V=E\cdot L=(v\times B)\cdot L=vBL$$

The current in the tether is then just:

$$I=V/(R_C+R_T)=vBm_T/2rdL$$

The power dissipated as ohmic heating in the tether is then given by:

$$P=IV=(vB)^2m_T/2rd$$

This equation shows the interesting result that the power dissipated does not depend upon the length or the area of the tether (within reason), but only the mass of the tether m_T, the resistivity r and density d of the tether material, and the velocity v of the conducting tether through the Earth's magnetic field B.

For a typical aluminum tether of mass m_T=10 kg, the power dissipated is an impressive P=2350 W. Even if only a small fraction of this power is dissipated in a real system, the kinetic energy of the host spacecraft would suffer a significant energy loss with time.

Orbital Energy Extraction Analysis:

The total energy U of a spacecraft of mass M in orbit around the Earth consists of two components, its positive orbital kinetic energy and its negative gravitational potential energy. If a circular orbit of altitude h and radius a=R_e+h is assumed, then the total energy of the spacecraft moving at a velocity w=$(GM_e/a)^{1/2}$ is:

$$U=+Mw^2/2-GMM_e/a=-GMM_e/2a$$

The system will be dissipating energy in the tether to decrease the energy U of the unwanted spacecraft from its relatively low negative value in a high Earth orbit to a greater negative value in a lower Earth orbit. To give a specific example, an unwanted spacecraft with a mass of M=1000 kg (a metric ton) in an orbit with an altitude of h=1000 km and a high orbital radius of a(H)=R_e+h=7371 km, then its (negative) total energy is:

$$U(H)=-GMM_e/2a(H)=-27.1\text{ GJ or }-27.1\text{ MJ/kg}$$

It is desired to lower the spacecraft to an orbit with an altitude just outside the atmosphere of h=200 km or a lower orbital radius of a(L)=R_e+h=6571 km, where atmospheric drag will finish the job. The total (negative) energy of a metric ton spacecraft at 200 km altitude is:

$$U(L)=-GMM_e/2a(L)=-30.4\text{ GJ}$$

Thus, the amount of energy needed to be dissipated in order to move the 1000 kg spacecraft from a 1000 km orbit with a total energy U(H)=−27.1 GJ down to a 200 km orbit with energy U(L)=−30.4 CJ is:

$$dU=U(H)-U(L)=3.3\text{ GJ.}$$

The decay time T of a metric ton spacecraft moving from a 1000 km altitude orbit to a 200 km altitude orbit with an energy difference of dU=3.3 GJ, when its energy is being dissipated at a power of P=2350 W by a typical aluminum tether massing just 10 kg or 1% the mass of the spacecraft, is found to be T=dU/P=16 days.

The actual decay time will be longer than this. If the electrodynamic drag force is very large, and becomes larger than the gravity gradient forces pulling on the ends of the tether, then the tether will tend to align itself along the magnetic field lines instead of across them, and the drag force will decrease because of the small angle between the conductor length and the magnetic field lines. The tether current will need to be controlled until the angle of the tether settles into an angle determined by the balance between these two forces. These electrodynamic torque vs. gradient torque dynamic interaction effects are discussed in the next section.

Force and Torque Balance Analysis:

It is now possible to calculate the forces and torques on the tether. It is also possible to use the fact that the electrodynamic and gravity forces and torques on the tether must balance each other out to calculate and teach some optimum values for some of the Terminator Tether™ parameters.

Electrodynamic Force and Torque:

As discussed above, both theory and experiment show that: provided the conducting tether is moved rapidly through the Earth's magnetic field in order to generate a voltage across it, and provided good contact is made with the space plasma, the conducting tether will have a current flowing through it. When a wire (moving or not) carrying a current I is embedded in a magnetic field B, there will be an electrodynatnic force F_E generated on each element of the wire. The electrodynamic force will be at right angles to the magnetic field vector and the length vector of the wire element, with a magnitude given by:

$$F_E=J\times B=ILB$$

where B is the horizontal component of the magnetic field, which is perpendicular to the assumed magnetic equator orbit plane, while the tether length vector L is assumed to lie in the orbit plane. If, as will be the case most of the time, the electron current is leaving the space plasma and entering the tether along the length of the tether, then IL needs to be replaced with an integral of the current along the length of the tether.

Note that the electrodynamic force amplitude or direction is not directly dependent on the motion of the tether through the Earth's magnetic field. The electrodynamic force would be the same if the tether were not moving and the current was being supplied by a battery. Since the current I is a function of the orbital velocity, however, there is a secondary dependence of the electrodynamic force magnitude on the orbital velocity, but not its direction. The electrodynamic force is always at right angles to the conductor, and stays at right angles to the conductor as the angle τ varies, as shown in FIG. **12**.

Assuming that the electrodynamic drag force is applied uniformly along the length of the tether, it is possible to make the simplifying assumption that the integrated force is effectively applied at right angles to the center of mass of the tether at the point L/2 as shown in FIG. **12**. The electrodynamic torque on the tether is then:

$$T_E=F_EL/2=IBL^2/2=vB^2L^3\cos\tau/2(R_C+rdL^2/m_T)=vB^2L^3\cos\tau/2R$$

Gravity Gradient Forces and Torques:

When a tether and its ballast end mass are deployed from a host spacecraft, the gravity gradient force field of the Earth, combined with the orbital centrifugal gradient force field, will cause the tether to deploy either up or down from the host spacecraft. The direction desired depends on which end of the tether is connected to the electron emitter. Normally, the electron emitter will be on the end attached to the host spacecraft, in which case the desired direction of deployment will be upward so that the induced voltage in the tether will produce an excess of electrons at the electron emitter end of the tether. The desired upward direction is chosen by having the deployer eject the ballast mass in the upward direction. Once the ballast mass has been started in that direction, the centrifugal force due to the orbital motion of the ballast mass will cause the ballast mass to continue to accelerate in the upward direction until it is brought to a halt by the full deployment of the tether.

If there were no electrodynamic or atmospheric drag, the equilibrium direction of the tether would be exactly along the vertical, since the combined gradient field is a maximum in that direction. Because a significant amount of electromagnetic drag is expected, the actual angle of the tether with respect to the local vertical will be at some angle τ, lagging behind the spacecraft motion in the plane of the orbit, as shown in FIG. **12**. In the following analysis shows that there is an optimum angle for τ that produces the largest electrodynamic drag force on the host spacecraft without producing tether instability.

The combined vertical gravity gradient and centrifugal gradient field 3Γ acting on the ballast mass m_B at the end of the tether of length L will produce a gradient force F_{GB} given by:

$$F_{GB}=-3_\Gamma m_BL\cos\tau$$

The strength of the force depends not only on the ballast mass m_B and the strength of the gradient field 3Γ, but also the component along the radial direction of the distance of the ballast mass from the center of mass, which is L cos τ. This force acts in the vertical direction along the radius vector leading from the ballast mass away from the center of the Earth. The amount of gradient force is not large. For a ballast mass of m_B=10 kg and a tether length of L=5 km, the gradient force is about 0.175 N.

As can be seen in FIG. **12**, the component of this gradient force that is at right angles to the tether, given by $F_{GB}\sin\tau$, will produce a torque T_{GB} on the tether that tends to restore the tether toward the vertical, lessening the angle τ.

$$T_{GB}=LF_{GB}\sin\tau=-3_\Gamma m_BL^2\sin\tau\cos\tau$$

The tether mass m_T also contributes to the gradient force and torque. If it is assumed that the tether has a uniform cross section, then it is possible to replace the distributed mass of the tether with an equivalent point mass m_T placed at the center of mass of the tether, which is the point L/2 along the tether, and a distance L/2 cos τ in the radial direction. The gradient force due to the tether mass is then:

$$F_{GT}=-3/2_\Gamma m_TL\cos\tau$$

While the gradient torque is:

$$T_{GT}=-L/2F_{GT}\sin\tau=-3/4_\Gamma m_BL^2\sin\tau\cos\tau$$

The total gradient torque attempting to restore the tether to its vertical orientation is then:

$$T_G=T_{GB}+T_{GT}=3_\Gamma(m_B+m_T/4)L^2\sin\tau\cos\tau$$

It is important to notice the variation of the total gravity gradient torque as the tether angle τ is varied. Since the gradient force is always in the radial or vertical direction, there is no torque on the tether when the tether is vertical, as is the case when there are no aerodynamic or electromagnetic drag forces. Once the drag forces become important and start to apply torque to the tether, increasing the tether angle τ, those drag torques causing an increase in tether angle τ will be opposed by a rising gradient torque which will attempt to decrease the tether angle. For small tether angles, the gradient torque increases first linearly with τ, then as sin τ, since cos τ is near unity for small τ, then at τ=45° the gradient torque reaches its maximum, where sin τ=cos τ=0.707 and sin τ cos τ=0.50. When this angle is reached, the tether is a point of catastrophic instability, for if there is a further increase in the electrodynamic drag force, causing the angle τ to become greater than 45°, the gradient torque, instead of growing stronger to counteract the increased drag torque, will become weaker and the tether angle will go rapidly to 90°.

To restore control to the tether angle if the instability occurs, it will be necessary to turn off the electrodynamic drag forces by shutting off the current flow through the tether. The τ=90° position for the tether and ballast mass is a gravitationally unstable orientation. After a time, slight fluctuations in the gravity field will allow the gradient force to slowly take over and restore the tether to the vertical orientation, which, unless it can be controlled in some way, is equally likely to be up or down. It would therefore be desirable to maintain control of the tether angle so as to avoid the tether angle getting into the region of instability.

Torque Balance on the Tether:

The angle τ of the tether is determined by a balance between the electrodynamic torque T_E attempting to increase the angle τ and the gradient torque T_G attempting to decrease the angle τ. Balance is achieved when the two torques are equal:

$$T_E=T_G=T_{GB}+T_{GT}$$

or:

$$vB^2L^3\cos\tau/2R=3\Gamma(m_B+m_T/4)L^2\sin\tau\cos\tau$$

Simplifing and rearranging yields an equation giving us the angle τ at which torque balance occurs:

$$\tau=\arcsin[vB^2L/6\Gamma(m_B+m_T/4)R]$$

The physical interpretation of this equation is that the maximum electrodynamic force that can be sustained on the tether is limited by the gradient force on the tether. It is possible to increase the electrodynamic force by decreasing the total resistance R of the tether (for example, by using a tether with a larger mass and a lower tether resistance), but if the tether resistance is too low, then the quantity in brackets becomes greater than unity and this equation has no solution, indicating that the tether has reached an angle where instability sets in.

To maintain control of the tether angle, it will be necessary to vary the control resistance of the tether to compensate for variations in magnetic field strength and direction, plasma density (which affects the plasma resistance), and other factors, and thereby maintain the tether at an intermediate angle where both the electrodynamic and gradient forces are at an appreciable level and balance each other.

Optimization of Tether Angle:

At first glance, it might seem that the optimum angle for the tether would be 45°, since at that angle the gradient torque is largest and therefore can counteract a larger electrodynamic drag force. However, since the 45° angle is the point where instability sets in, it is desirable to set the tether angle at some value below 45°. The optimum angle is that which maximizes the horizontal or drag component of the electrodynamic force. This optimum angle τ is derived from the equation for the horizontal component of the electrodynamic force, or the electrodynamic drag force, since it opposes the host spacecraft motion:

$$F_{ED}=F_E\cos\tau=6[m_B+m_T/4]\Gamma L\sin\tau\cos^2\tau$$

By setting the partial derivative$\square F_{ED}/\square\tau=0$ and solving, it is possible to calculate that the optimum angle for the tether that gives the maximum electrodynamic drag force F_{ED}, while still keeping the tether torques balanced and under control, is $\tau=\arctan(0.707)=35.26°$. With this angle selected and maintained by varying the control resistor R_C to compensate for variations in plasma contact resistance and variatins in the strength and direction of the Earth's magnetic field B seen at the spacecraft, the tether experience the maximum stable value for the electromagnetic drag force of:

$$F_{ED}(\max,\tau=35.26°)=2.31[m_B+m_T/4]\Gamma L$$

There are many ways to provide the sensing information needed to provide the feedback signals to the control resistor, but the simplest is to merely measure the drag acceleration on the host spacecraft with a set of accelerometers, and maximiize the force in the direction opposite to the host spacecraft motion. Another method would be to measure the current in the tether, and knowing the tether resistance and the amount of control resistance, calculate the power being extracted and maximize that value. Alternate methods would be to use an optical position sensor or GPS receivers at both ends of the tether to measure the angle of the tether or the position of the ballast mass with respect to the host spacecraft.

To make an estimate of the magnitude of drag force attainable, assume that a typical aluminum tether with a length of L=5 km, a tether mass of m_T=10 kg and a ballast mass of m_B=10 kg is being used. Then the maximum gradient force limited electrodynamic drag force is F_{ED}= 0.167 N. The power being dissipated in the canoical tether at the angle τ, with $R_C=R_T=rdL^2/m_T=185\ \Omega$ is given by:

$$P=IR=[vBL\cos\tau]^2(R_C+rdL^2m_T)$$

which for an angle $\tau=35.26°$ results in a power dissipation of 1570 W. Thus, the thrust level per unit power obtained by the typical Terminator Tether™ would be about 0.106 N/kW. This number is comparable to the value of 0.148 N/kW estimated for the much heavier and longer TSS-1R tether.

Reconciliation of Energy and Force Analysis:

By a force and torque balance analysis an optimum angle for the tether has been discovered at which it is possible to obtain a maximum in the drag component of the electrodynamic force. There is also an additional component of the electrodynamic force, the component in the vertical direction, $F_{EV=FE}\sin(\tau)$, which is downward for an upwardly deployed tether. This component of force combines with the gravity force of the Earth to effectively allow the host spacecraft to orbit a little faster than normal for that orbital altitude. It does not contribute to the deorbiting of the host spacecraft. But since this vertical component of force is created by current running through the tether, and that current is creating heat and dissipating energy as it passes through the tether, there might be some concern that the force and torque balance analysis above does not conform to the law of conservation of energy.

The inventors will now show, in a very general manner, that no matter what the tether angle, the electrical power being dissipated in the tether is exactly equal to the power being lost by the slowing of the host spacecraft.

The "deorbit power" P_D that must be removed from a spacecraft moving at a velocity v when that motion is opposed by a drag force F_{ED} is:

$$P_D=v\bullet F_{ED}=vF_E\cos\tau=vILB\cos\tau=I[vBL\cos\tau]$$

But, since the voltage V induced across the tether of length L and tilt angle τ moving at a velocity v through a horizontal magnetic field B is

$$V=E\bullet L=EL\cos\tau=vBL\cos\tau$$

Therefore:

$$P_D=IV=P_E$$

And the deorbit power P_D extracted from the slowing of the spacecraft by the drag component of the electrodynamic force is always exactly equal to the electrical power P_E being dissipated as heat in the tether circuits, independent of the tether angle.

Optimization of Tether Mass Distribution:

If it is not possible to use a piece of the host spacecraft as ballast mass, then the mass of the ballast must be included in the Terminator Tether™ mass. It would be desirable to minimize the total Terminator Tether™ mass, since every kilogram saved means that another kilogram's worth of revenue-producing transponders can be added to the working payload of the host spacecraft. Given a total mass for the Terminator Tether™ and the mass of the deployer/controller unit, it is possible to optimize the mass distribution between the ballast mass and the tether mass to obtain a minimum total Terminator Tether™ mass. A well-designed Terminator Tether™ will also have most, if not all, of the deployer mass incorporated into the ballast mass.

Assume that the total Terminator Tether™ mass consists of the deployer/controller mass m_D, the tether mass m_T, and the ballast mass m_B, with $m_B > m_T > m_D$. Of the three mass components in the Terminator Tether™, two of them affect the electrodynamic drag performance. If it is assumed that the ballast mass is a factor X larger than the tether mass, or $m_B = Xm_T$, the maxinum drag force that can be obtained is now:

$$F_{ED}(\max)=2.31 m_T (X+0.25)\Gamma L$$

If it is assumed that the control resistor has been adjusted so that this maximum value for the electrodynamic drag force is maintained as the motion of the spacecraft along its orbit moves the spacecraft into regions with different magnetic field strengths and plasma densities, then for the angle τ=35.26°, cos(35.26°)=0.817, and the maximum electrodynamic drag force in terms of the electrodynamic parameters will be:

$$0.817 vB^2L^2/(R_C+rdL^2/m_T)=F_{ED}(\max)=2.31 m_T (X+0.25)\Gamma L$$

or canceling out terms and rearranging:

$$(X+0.25)=0.353 vB^2L/\Gamma m_T(R_C+rdL^2/m_T)$$

It is now possible to make the further simplifying assumption that to maintain control of the tether, the control resistor needs to be roughly the same size as the tether resistance or $R_C \approx rdL^2/m_T$. Using this assumption produces an equation for the optimum value for the ratio X of the ballast mass to the tether mass:

$$\begin{aligned} X = vB^2/\Gamma rdL - 0.25 &= -0.25 + \frac{vB^2}{12\sin\gamma\cos\gamma\Gamma rdL} \\ &= -0.25\frac{\Lambda}{L} \end{aligned}$$

Where Λ is the "effective electrodynamic length".

If the tether is short, so that L is small, this equation indicates that the ballast mass must be increased to increase the gradient force, since it is the gradient force which determines the upper limit to the amount of electrodynamic drag force that can be generated without losing control of the tether.

Assuming the usual values for the Earth's magnetic and gradient fields, the velocity of the host spacecraft through the magnetic field, and the resistivity and density of aluminum, this equation becomes:

$$X=10.9/L-0.25$$

where L is in kilometers.

For a typical tether of length L=5 km, X=1.93. If the total mass for the Terminator Tether™ is 20 kg, or 2% of the host spacecraft mass, and the mass of the deployer is 10 kg, then the remaining 10 kg should be distributed so that 6.59 kg is in the ballast mass while 3.41 kg is in tether mass.

Optimization of Tether Length:

If the ballast mass consists of a piece of the host spacecraft, then the mass of the ballast does not have to be considered in the optimization of the Tether system. In this case, the optimum distribution of the Tether mass is to put as much mass as possible into the tether, with a minimum in the deployer/controller package. Under this assumption, the optimum length of the tether is determined by the desire to keep the Area-Time Product of the host spacecraft plus Tether to a minimum.

In a long tether of length L, the effective "collision" cross-sectional area is not the area of the tether, but the larger area produced by multiplying the length of the tether by sum of the width of the tether plus the width of the "target" spacecraft that would be damaged by the tether hitting it. Thus, to decrease the "area" portion of the Area-Time Product, it would be desirable to shorten the tether, making it thicker, and perhaps slightly wider, at the same time, thus keeping the same tether mass and electron collection area, and thereby maintaining the drag force and keeping the decay time at the same level. The result will be a shortening of the Area-Time Product.

If the tether is shortened too much, however, the smaller voltage generated across the shorter tether will leave less voltage margin for the operation of the plasma contactors and the battery recharge system. Since the maximum voltage that can be generated is about 200 V/km, and typical plasma contactors and battery, chargers take 25–100 V to operate, a minimum length for a Tether would be roughly 2 km. Host spacecraft operating in polar orbits, where the conditions for voltage generation are not as good, may require a tether length of 5 km or more.

Since many watts of heat power will be dissipated in the control resistor, means must usually be provided to radiate the heat away into space. One low-mass method of accomplishing this is to make the control resistor (or resistors) in the form of a long, electrically insulated, high resistivity wire similar to those used in electric blankets, but designed to operate at a higher temperature, and during the fabrication of the electrodynamic tether, incorporate the high resistivity wire (or wires) into the end of the tether closest to the control circuit by weaving or braiding the insulated high resistivity wire in with the uninsulated aluminum wire of the electrodynamic tether. The surface area of a long wire is very large, so in this extended wire configuration, exposed to the space all around it, the hot control resistor wire can self-radiate its power into cold space without incurring the mass penalty of a separate radiating surface.

Implementation:

The basic optimum structure for a Terminator Tether™ would be one of the many types of Hoytethers. A multiline (6–12 primary line) Hoytape™ will provide the largest contact area with the plasma, since both sides of the tape would be able to pass current to the plasma. If the spacing between the primary lines is chosen to be larger than twice the Debye length of the plasma, then the effective current collection area per unit length of the Hoytape is proportional to the width of the Hoytape mesh, not the diameter of the wires in the mesh. Thus, a Hoytape not only provides an assured longer life for the Terminator Tether™, but very short lengths will also provide a very large current collection areas.

The deployer for the tether can deploy the Terminator Tether™ either down or up or both. The deployer can stay attached to the spacecraft as was done in the SEDS missions, or perhaps a better alternative would have the deployer ejected from the spacecraft, with one end of the tether still attached to the spacecraft, reeling out tether as it leaves. The empty deployer would then act as a ballast mass at the end of the Terminator Tether™, which would improve the performance.

The "Remora Remover"

In addition to attaching Terminator Tethers to spacecraft before launch, it is possible to consider a missile-like application. This "Remora Remover" missile would use a Terminator Tether™ carried by a seeker missile similar to the small "hit-to-kill" missiles developed by the Space Defense Initiative Office that deployed a net loaded with oriented metal rods. The Remora Remover missile would hunt down a spacecraft that needs to be removed from space, but instead of hitting the spacecraft, the missile would be programmed to rendezvous with the spacecraft and attach itself to the host spacecraft using a hooked net, harpoon, or adhesive "sucker". The Remora Remover missile would then deploy the Terminator Tether™, which would bring down both the spacecraft and the missile.

The inventors have presented two analyses, backed up with data obtained from space flight experiments. One analysis was based on generalized energy conservation laws while the other analysis use force and torque balance arguments. Both analyses show that optimized aluminum wire tethers 2 to 5 km in length and massing just 1% to 5% of the mass of the host spacecraft can deorbit the host spacecraft in the order of a few weeks, thus mitigating the long-lived orbital debris hazard created by a constellation spacecraft after their end-of-life.

Power Augmented Operation of the Tether:

It is well known in the aerospace literature that if a spacecraft has a power supply and a conducting tether hanging from it, and the spacecraft is in an orbit or trajectory that takes it near a body, such as the Earth or Jupiter, which has both a magnetic field and a region of moderate density electron-ion plasma, and electrical current is pumped from the power supply through the tether and back through the plasma, that the current flowing through the tether will interact with the magnetic field of the body, producing forces on the tether and thence on the spacecraft, changing its orbit or trajectory. These forces can be used to increase or decrease the spacecraft altitude and/or inclination. The amount of altitude or inclination change is proportional to the ratio of the power available from the power supply divided by the mass of the spacecraft and varies with altitude. The unpowered Terminator Tether™ is a simple example of such a system, in which only altitude decrease is possible, although small amounts of both inclination increase or decrease are possible at the same time. Some typical references to this prior art are: Les Johnson, "Propulsive Small Expendable Deployer System Mission (ProSEDS)", Proceedings of the OAST 8th Advanced Propulsion Workshop, JPL, Pasadena, Calif., 20–22 May 1997; and Les Johnson, Joe Carroll, Robert D. Estes, Enrico Lorenzini, Brian Gilchrist, Manuel Martinez-Sanchez, Juan Sanmartin, and Irwin Vas, "Electrodynamic Tethers for Reboost of the International Space Station and Spacecraft Propulsion, AIAA Paper 1996.

Since the Terminator Tether™ of the invention is normally associated with a satellite that has a power supply in the form of a solar panel combined with a storage battery, the Terminator Tether™ can be operated in the "powered propulsion" mode if desired. Such a mode would be useful when attempting to avoid a collision between the Terminator Tether™ and another spacecraft. In such cases, the power from the solar panel, augmented by the power stored in the battery, can be used to temporarily operate the Terminator Tether™ in the propulsion mode instead of the normal drag mode. This would enable to Terminator Tether™ to raise its altitude and/or increase or decrease its inclination to avoid a collision.

In addition, the Terminator Tether™ could be operated in the "power-augmented drag" mode. In this mode, there will be a current induced in the tether by the motion of the spacecraft through the magnetic field, which will cause a drag force on the spacecraft. This induced current would then be augmented by additional current in the same direction generated by the solar panel and the storage batteries in the power supply. The increased current will cause an increased drag force. If there is a significant vertical component of the magnetic field there will also be an increased force normal to the orbital plane, which will cause an increased rate of change of inclination.

When the Terminator Tether™ is operated in the "powered propulsion" or "power-augmented drag" mode, the conditions for stable optimum operation will be essentially the same as in the "unpowered drag" mode. The reason the conditions stay the same is that the maximum tether force than can be effectively utilized is limited by the strength of the gradient force, which has not changed, since the length and mass of the tether and the ballast mass has not changed for these different modes of operation. As a result the optimum operational angle for the tether is still 35.26 degrees, and the optimum ratio of ballast mass to tether mass for different tether lengths and tether material is still determined by the analysis of the invention as is set forth above.

In the analysis given above, where the optimum tether angle was found, the tether was assumed to be in the plane of the orbit and lagging behind the spacecraft motion, since in the "unpowered drag" mode, that would be its normal position, although if the component of the magnetic field along the orbit were high, there would be some tilt of the tether out of the plane of the orbit, producing a force tending to change the inclination of the orbit. In either of the "powered" modes of operation of the Terminator Tether™, however, the tilt of the tether could be forward toward the motion of the spacecraft or strongly tilted to one side or the other of the orbital plane. In all these cases, if the electrodynamic force is allowed to become too large, and the tether angle exceeds 45 degrees, the restoring force of the gradient field will drop off and the tether angle will go unstable. As a result, the optimum angle for the tether that will give the maximum stable force, whether it is a maximum drag force, and maximum propulsion force, or a maximum inclination change force, will be 35.26 degrees, and the optimum ratio for the mass of the tether versus the mass of the ballast will be the same as for the unpowered drag case.

The analysis that gave the optimum angle as 35.26 degrees assumed that the tether would be oriented like a rigid rotor. In reality, variations in forces along the tether will probably cause the tether to hang in a slightly curved shape, where the optimum angle may not be exactly 35.26 degrees from the local vertical. The optimum angle of the tether as it leaves the host spacecraft in such a case is where the drag force is largest.

The present invention is discussed in this disclosure in terms of its space applications as a Terminator Tether™ useful for the removal of unwanted satellites. It should be understood, however, that the present invention is useful in any application where a space object can use a conductive tether to produce electrodynamic force through interaction with a magnetic field. The invention, therefore, should be limited not by this description, but only by the appended claims and their equivalents.

What is claimed is:

1. A method of operating a conductive tether attached to a space object, said conductive tether being oriented at an angle to magnetic field lines of an external magnetic field and said conductive tether moving across the magnetic field lines of said external magnetic field at said angle to producing an electric current in said conducting tether and a resulting electrodynamic force acting on said tether and said attached space object comprising,

Controlling the amount of electric current flowing in said conducting tether to vary the angle between the conductive tether the field lines of the external magnetic field and the electrodynamic force acting on said tether and attached space object.

2. A method as in claim **1** wherein said angle is less than 45 degrees.

3. A method as in claim **1** wherein said angle is 35.26 degrees.

4. A method as in claim **1** wherein said angle is controlled to average 35.26 degrees over time.

5. A method as in claim **1** wherein said angle is controlled to average 35.26 degrees over the length of the tether.

6. A method as in claim **1** wherein the space object and the tether connected to it are rotated about their center of mass while the conductive tether interacts with the external magnetic field.

7. A method as in claim **1** wherein the angle of the conductive tether to the external magnetic field is controlled by sensing a measurable parameter of the space object/conductive tether system.

8. A method as in claim **7** wherein the parameter is current flow in the conductive tether.

9. A method as in claim **7** wherein the parameter is the electrical power dissipated in the conductive tether and the current control system.

10. A method as in claim **7** wherein the parameter is the relative position of the space object and the position of the far end tip of the tether.

11. A method as in claim **7** wherein the parameter is the acceleration of the space object.

12. A method as in claim **6** wherein the parameter is the state vector of the space object as measured by the Global Positioning System.

13. A method as in claim **1** including the step of applying electric power to the conductive tether to change the state vector of the space object.

14. A method of operating a conductive tether attached to a space object, said conductive tether being oriented at an angle to the local vertical between said space object and a celestial body having a magnetic field having magnetic field lines external to said conductive tether; said conductive tether moving across said magnetic field lines of said external magnetic field at an angle to the local vertical to produce an electric current in said conducting tether and a resulting electrodynamic force acting on said tether and said attached space object comprising,

Controlling the amount of electric current flowing in said conducting tether to vary the angle between the conductive tether the local vertical between said space object and said celestial body to vary the electrodynamic force acting on said tether and attached space object.

15. A method as in claim **14** wherein said angle is less than 45 degrees.

16. A method as in claim **14** wherein said angle is controlled to average 35.26 degrees over time.

17. A method as in claim **14** wherein the space object and the tether connected to it are rotated about their center of mass while the conductive tether interacts with the external magnetic field.

18. A method as in claim **14** wherein the angle of the conductive tether to the external magnetic field is controlled by sensing a measurable parameter of the space object/conductive tether system.

19. A method as in claim **18** wherein the parameter is current flow in the conductive tether.

20. A method as in claim **18** wherein the parameter is the acceleration of the space object caused by the electrodynamic force acting on the space object.

21. A method as in claim **18** wherein the parameter is the change in the state vector of the space object as measured by the Global Positioning System.

22. A method as in claim **14** including the step of applying electric power to the conductive tether to change the state vector of the space object.

23. A method as in claim **14** wherein the celestial body is the Earth and the magnetic field is the Earth's magnetic field.

24. A method as in claim **14** wherein the celestial body is the sun and the magnetic field is the Sun's magnetic field.

25. A method as in claim **14** wherein the magnetic field is the vector sum of the magnetic fields from all celestial bodies in the solar system.

26. A method as in claim **14** wherein the conductive tether is a Hoytether.

* * * * *

US 20010032905A1

(19) United States

(12) Patent Application Publication

Grandics

(10) Pub. No.: US 2001/0032905 A1

(43) Pub. Date: Oct. 25, 2001

(54) METHOD AND APPARATUS FOR CONVERTING ELECTROSTATIC POTENTIAL ENERGY

(76) Inventor: **Peter Grandics**, Carlsbad, CA (US)

Correspondence Address:
OPPENHEIMER WOLFF & DONNELLY LLP
Suite 3800
2029 Century Park East
Los Angeles, CA 90067-3024 (US)

(21) Appl. No.: **09/732,110**

(22) Filed: **Dec. 7, 2000**

Related U.S. Application Data

(63) Non-provisional of provisional application No. 60/169,378, filed on Dec. 7, 1999.

Publication Classification

(51) Int. Cl.^{7} **B64C 15/00**; B64C 27/82

(52) U.S. Cl. **244/51**

(57) ABSTRACT

A new method is described to produce directional currents in the space lattice, the fundamental fabric of the universe. The currents may be unidirectional or vortexual in nature and are suitable for propulsion or the generation of electric power. For propulsion, the unidirectional currents are induced by charging capacitors possessing suitable geometries. This will allow the manufacture of vehicles capable of levitation and flight. The vortexual space lattice currents are produced in a suitable pyramid when electrostatic and magnetic fields interact transversally. The vortex causes charge separation and current in the coil wrapped around the pyramid.

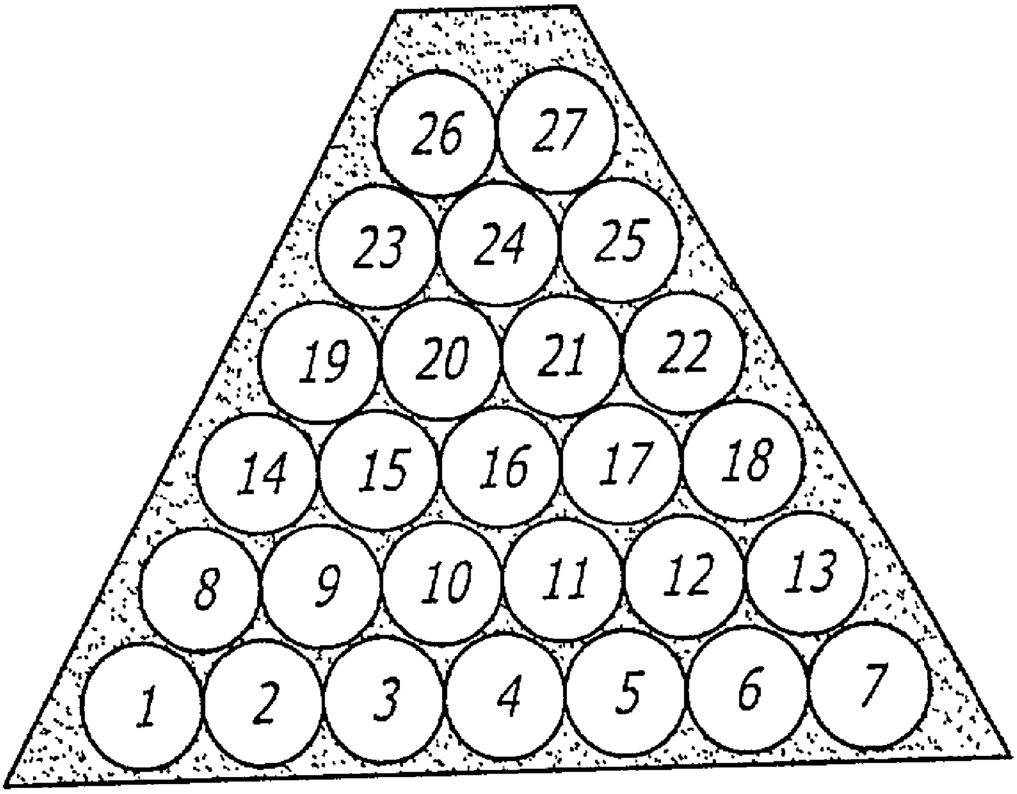

FIG. 1

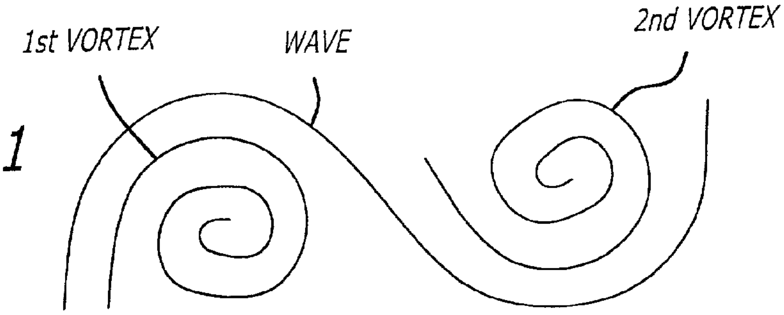
1st VORTEX
WAVE
2nd VORTEX

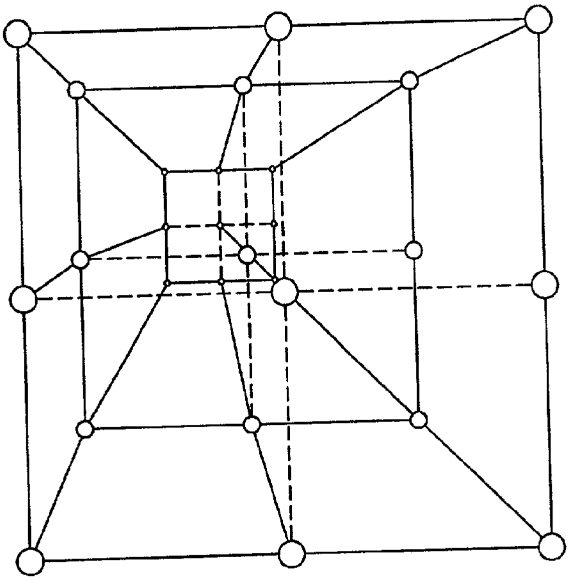

FIG. 2

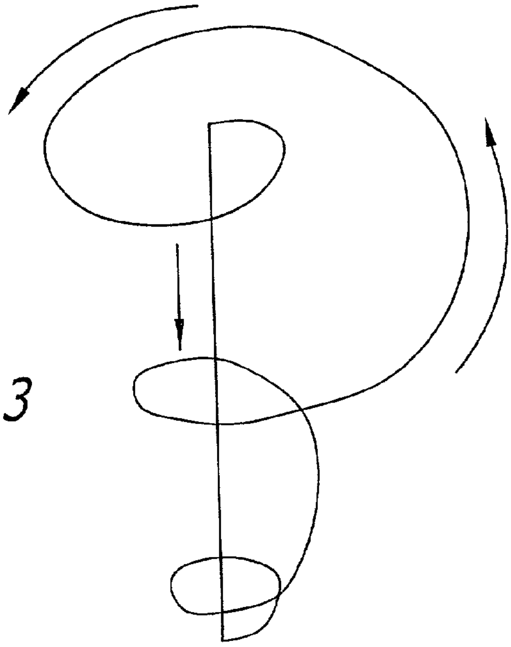

FIG. 3

FIG. 4

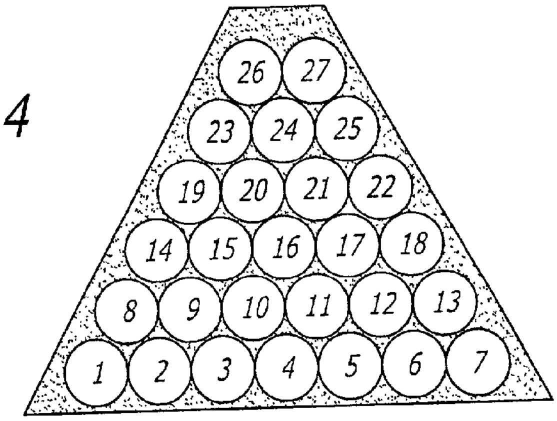
26
27
23
24
25
19
20
21
22
14
15
16
17
18
8
9
10
11
12
13
1
2
3
4
5
6
7

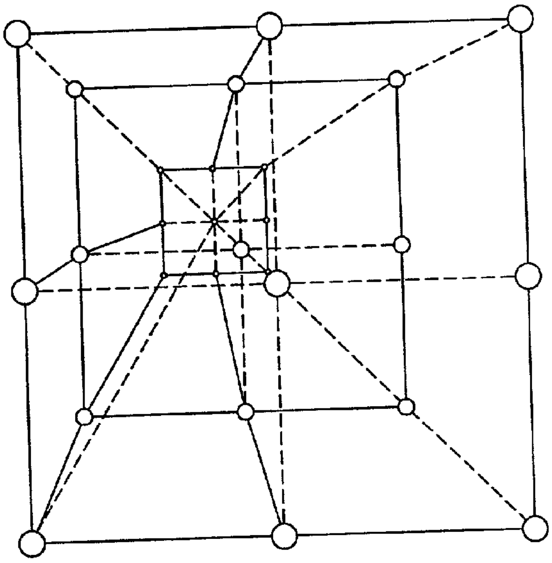

FIG. 5

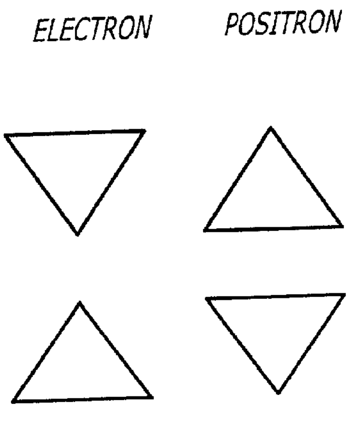
ELECTRON
POSITRON

FIG. 6

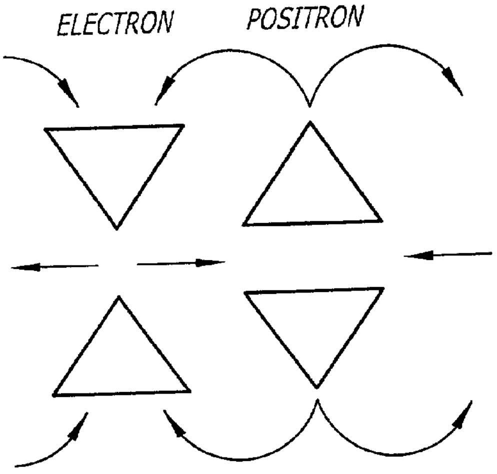
ELECTRON
POSITRON

FIG. 7

FIG. 8

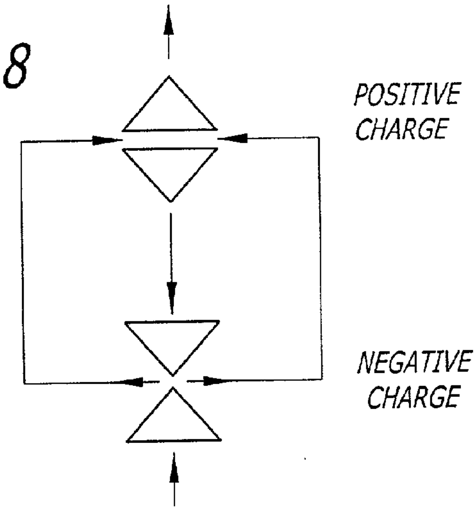
POSITIVE CHARGE
NEGATIVE CHARGE

FIG. 9

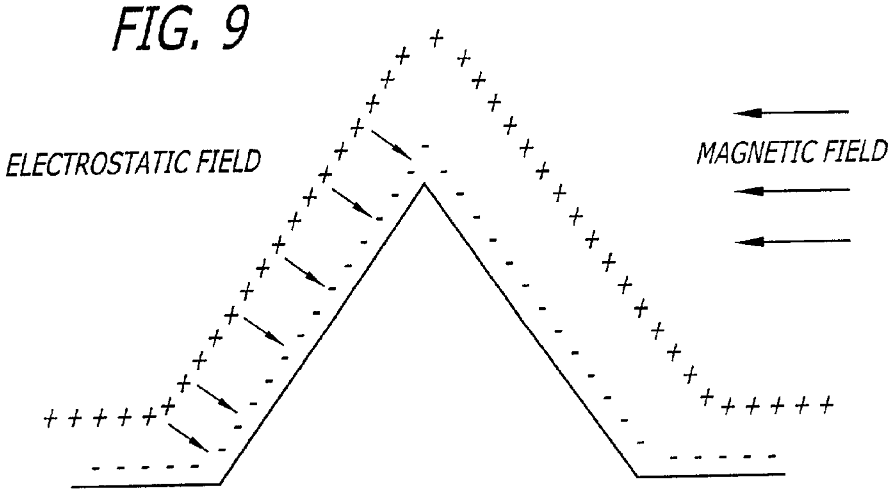
ELECTROSTATIC FIELD
MAGNETIC FIELD

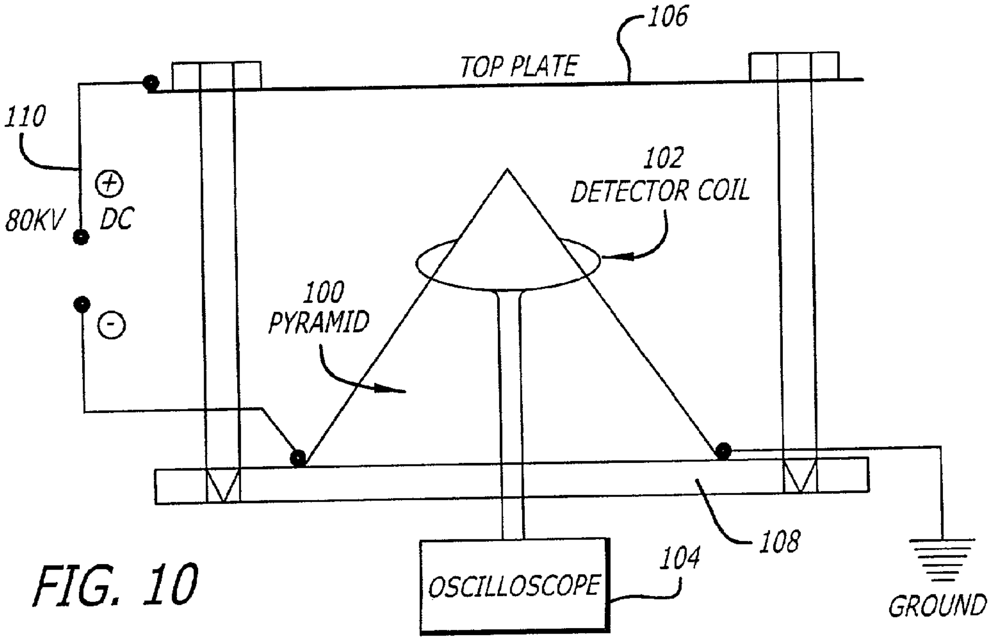
106
TOP PLATE
110
80KV
DC
102
DETECTOR COIL
100
PYRAMID
108
104
OSCILLOSCOPE
GROUND

FIG. 10

METHOD AND APPARATUS FOR CONVERTING ELECTROSTATIC POTENTIAL ENERGY

CROSS-REFERENCES

[**0001**] This application claims priority from Provisional Application Ser. No. 60/169,378, by Peter Grandics, filed Dec. 7, 1999, also entitled "Method and Apparatus for Converting Electrostatic Potential Energy," the contents of which are hereby incorporated by this reference.

BACKGROUND OF THE INVENTION

[**0002**] This invention is directed to methods and apparatus for conversion of electrostatic energy. Specifically, this invention describes methods to cause directional movements in the space lattice.

[**0003**] To fully comprehend the theoretical basis for these methods, first the genesis of the elemental particles electron and positron needs to be discussed based on geometric or space relationships, and their interactions at the sub-elemental particle level must be clarified. The advantage of such an approach is that it can be visualized, and not just expressed mathematically. by definition, the word "structure" implies a geometric relationship, and such relationship must be of great importance in atomic structures even though they are hidden from direct view.

[**0004**] The study of geometric relationships of macroscopic structures of matter may provide important insights into the properties of matter at the microscale. Astrophysics suggests the universe probably started with an energy burst from one point. This big bang theory is supported by visual and electronic evidence that the galaxies are expanding away from a single center.

[**0005**] In stellar formation, gravitational pull and velocity generate a rotary vortex. Even in the relative absence of such factors, all objects spiral at some given rate peculiar to their special influences. A spiral is created when an object moves forward while rotating. Earth's movement in space is an example of this process. The Earth orbits the sun, while the sun pulls it along towards the direction of Vega in the constellation Hercules. the combined circular and forward motion creates a spiral. Our sun has the same motion in relation to the galaxy center. Our galaxy, the Milky Way, also spirals away from the Big Bang center.

[**0006**] Water going down a drain demonstrates some of the special influences affecting spiral formation. The spiral, caused by Coriolis forces, changes dynamically under the effects of gravitational pull, drain diameter, obstructions, temperature, pressure, volume, viscosity, and stirring. The spiral changes shape and acceleration but maintains the universal shape of a spiral. The water flow is responsive to all possible factors, and so is the spiral.

[**0007**] Remarkably, the spiral vortex has a "memory" of itself. When a vortex is distorted to become elliptical, it spontaneously returns to its original circular form. The vortex is a self-sustaining type of motion; its resiliency is comparable to that of atomic bodies.

[**0008**] Spirals condense energy and sine waves transmit it along a frictional line of force between them (**FIG. 1**). Spirals and waves could be considered illusions of each other. Two opposed spirals form a wave, or a wave produces two spirals. A sine is the producer of waves of spiral forces. Any fluid capable of supporting wave motion can also support vortex motion.

[**0009**] The spiral is the prominent form of organization of matter. The large proportion of spiral galaxies among celestial bodies visibly demonstrates this. On the microscale, even the building block of life, the DNA uses this structure and spirals are likely to be dominant at the level of the atom and below. This is the subject of the present theory. I propose that all atoms, all stellar formations use the mutable spiral to adapt to their spiral environment.

[**0010**] Another proposition is that all mediums of matter can be considered a type of crystal. Crystallinity is readily recognizable in the mineral world, but it is also a more general state of matter. Besides solid crystals, liquid crystals also exist. by definition, a crystal is a regularly repeating atomic arrangement, such as a chemical element, a compound or an isomorphous mixture. Therefore, the term crystal applies to material expressions where crystallinity is not so obvious, e.g., gases, complex biologics and various life forms, including viruses, bacteria and higher organisms. Air and water could be considered loose crystals subject to fast molecular drift. At low temperatures where molecular drift is reduced, gases form solid or liquid crystals. Soil and stone and metal are opaque cryptocrystals with slower rates of molecular drift. On the macroscale, the Earth could be seen as faceted crystal with its mountain ridges are the ridges of a geodesic sphere. Biopolymers such as DNA, proteins or polysaccharides fit well the definition of a crystal and are commonly made into crystals. All life can be seen as a crystal. From fish to humans, we are liquid crystals on a skeletal lattice.

[**0011**] Crystals are the shape of discrete units of matter and notably the channelized direction of energy, that is, the direction energy that flows unforced. Crystals form the basis for corpuscle-wave conversion. Crystals create resonance and conduct the flow of energy between states. Crystals can also be considered lenses. By definition, a lens is a device capable of refracting, or bending light. Light is an energy flow, so on a more general term, a lens can be defined as any object capable of changing the direction of energy flow. By this broader definition, even an electric wire is a lens as it is capable of changing the direction of flow of electrical energy. lenses communicate energy as part of the principle that all matter vibrate, all matter transmit, and all matter receives energy. The universe changes energy states with lenses. Following this line of reasoning, the universe could be viewed as a resonant crystal lens.

[**0012**] This observation is important because the universe is considered to be largely empty, the largest component of which is the so-called vacuum space. Since Nature seems to use the same geometrical organizing principles from macro to micro, I suggest that the vacuum space must also be "crystalline". Since the term crystalline is associated with material of which the vacuum space is substantially devoid, I will use an extended meaning of "structured" when discussing crystallinity of vacuum space.

[**0013**] The theory of "crystalline" vacuum space was introduced by Simhony (Simhony, M. (1990) The Electron-Positron Lattice Space, Cause of Relativity and Quantum Effects, Physics Section 5, The Hebrew University, Jerusalem., and Simhony, M. (1994) Invitation to the Natural

Physics of Matter, Space, and Radiation, World Scientific Publishing Co., London, UK.). Simhony reasoned that three-dimensional physical phenomena must have three-dimensional physical causes and explanations.

[**0014**] He demonstrated that physical reality can be described by the laws of classical physics supplemented by the presence of a space lattice. This led to the development of the theory of an electron-positron lattice space (epola for short). In the epola, bound electron-positron pairs reside at the lattice sites of a face-centered cubical “crystal” structure similar to that of NaCl crystals (Kennon, N. F. (1978) Patterns in Crystals, John Wiley & Sons, New York.).

[**0015**] The epola theory allowed a physical explanation of all yet unexplained postulates of quantum mechanics and relativity including the particle-wave duality, the quantized nature of electron orbits in the atom, the electromagnetic radiation, the photon, and gravitational interactions as well as the relationship of electrostatic, magnetic and gravitational interactions. He demonstrated gravity to be a derivative of electromagnetism.

[**0016**] There is a large body of evidence suggesting that the vacuum space is not empty at all. Experiments verify that the vacuum space contains an enormous residual background energy (some examples are: Boyer, T. H. (1985) The classical vacuum (zero-point energy) Scientific American 70-78, Haisch, B., Rueda, A. and Puthoff, H. E. (1994) Beyond $E=mc^2$, The Sciences, 34, No. 6, 26-31, and B.

[**0017**] Haisch and A. Rueda, (1999).On the relation between zero-point-field-induced inertial mass and the Einstein-de Broglie formula, Physics Letters A, in press), called zero-point energy (ZPE). The ZPE manifests as a pervasive and vast electromagnetic field called the zero-point field (ZPF). a dynamic field, ZPF is a virtual plasma, with particles arising and disappearing of a background energy field serving as a baseline, or zero point, for all physical processes. The ZPE remains even at absolute zero. Simhony has described zero-point energy fluctuations as analogous to brownian motion of epola particles around their lattice sites. A potential alternative term for the lattice space is zero-point-field (ZPF) described in 1994 by Haisch, Rueda and Puthoff (Haisch, B., Rueda, A. and Puthoff, H. E. (1994) Beyond E=mc2, The Sciences, 34, No. 6, 26-31). This area remains an active field of research.

[**0018**] Simhony also suggested that all particles of matter are directly formed from the lattice space, the mechanism of which remains undetermined. The present theory intends to answer this question in terms of lattice space.

[**0019**] I propose that the vacuum space has a lattice structure similar to that introduced by Simhony. this structure is a face-centered cube having 27 lattice sites (**FIG. 2**). on the lattice sites reside the elemental “particles” forming all particles of matter. These are “particles of energy” rather than particles of matter.

[**0020**] I propose that the “particle of energy” of the space lattice, called here the subtle electron, is a spiral energy vortex tied into itself in the form of a circumvolution cissoid. The circumvolution cissoid is a spiral turn around an axis converging into an apex, in a self-imploding, self-sustaining vortex motion (**FIG. 3**). The vortex pulsates and its vibration is a function of **(2_Ø)x, where Ø is the Fibonacci series number and x**=0,1,2,3, the number of turns the circumvolution cissoid makes. Once started up, such a vortex would run practically indefinitely inside the space lattice.

[**0021**] The internal friction of such a space lattice must be so low that it would only be noticeable as a red shift in the spectra of distant galaxies. The space lattice, like its constituent energy vortices, must also be a resilient structure with only limited compressibility. At the same time, it must have fluidity since it is capable of transmitting waves with transverse displacement. In the absence of atomic oscillators, the space lattice would be incapable of dissipating energy in the form of heat.

[**0022**] Vortices maintain their circular forms as well as their proportions and dimensions. The adjacent vortices have a coordinating effect that establishes axial alignment and rolling contact between vortices within the space lattice. In that sense, there is a great deal of similarity between crystals of material bodies and the structure of the space lattice. I postulate that the space lattice is an incompressible, frictionless fluid made out of cubes of units cells of subtle electron vortices.

[**0023**] We can consider the space lattice equivalent to motion. A space lattice in motion necessarily has inertia. Although inertia is generally attributed to moving material bodies, it is actually a property of motion. Inertia is just the continuity of motion. In the case of vortex motion the inertia is localized. Localized inertia can also be called momentum. To account for the elemental particles of matter, we need a space lattice that is capable of moving.

[**0024**] Anderson discovered in 1932, that when 1.02 MeV photon energy is absorbed into the vacuum space, an electron-positron pair can appear. this observation is explained here as a glimpse into the formation of elemental particles of matter. To become matter, energy must become more angular. This hypothesis is derived from the observation that all material expression is a type of crystal, and is the compound and derivative of a fundamental triangular form which all the seven crystal systems can be derived (Kennon, N. F. (1978) Patterns in Crystals, John Wiley & Sons, New York.). The mechanics of this expression at the level of the space lattice are explained as follows: As the energy of electromagnetic radiation propagates through the space lattice, it polarizes the subtle electron energy vortices (for analogy see **FIG. 1**). Subsequently, the “energy particles” of the space lattice, the subtle electrons, undergo a phase transition similar to the condensation of gases. This happens at the resonant frequency of the 1.02 MeV gamma radiation.

[**0025**] The 27 subtle electrons of the unit cell of space lattice rearrange into a pyramidal segment of the cube (**FIG. 5**) on six levels each forming circles of vortices. An open-flat presentation of the rearrangement is shown in **FIG. 4**. This structure is the postulated smallest unit of matter. note that the cube is composed of six interlocking pyramids so the cube and the pyramid are resonant structures. Inside the pyramid, the subtle electron rings form a vortex capable of circulating the fluid space lattice. The pumping action is driven by the self-sustaining, pulsating vortex motion of its constituent subtle electrons (**FIG. 3**). The overall shape is a cone fitting inside the pyramid. The formation of matter follows the geometry of the space lattice (**FIG. 5**) and thus we may conclude that the space lattice provides the blueprint for matter.

[**0026**] An electron is produced when two such cones are joined in a tip-to-tip conformation (**FIG. 6**). The positron is made out of two vortices (cones) facing base-to-base. The energy vortices which make up the electrons have a resilient vibratory structure which should be capable of vibrating at various frequencies and modes. This would allow us to account for the series of spectral lines, a signature of absorbing and emitting energy at various frequencies.

[**0027**] These complex vortex particles have a pumping effect and circulate the fluid space lattice. For the electron, the space lattice is drawn in polarly and expelled equatorially. For the positron, the space lattice is drawn in equatorially and expelled polarly (**FIG. 7**). The direction of circulation of fluid space lattice determines the positive or negative polarities. When in close proximity, a specific flow coupling occurs between the electron and positron (**FIG. 7**). It is quite probable that a similar flow coupling exists between a proton and an electron inside the atom, although such coupling is stable in the case of the proton. This is likely a result of the different vibratory structure of the vortex proton.

[**0028**] The instability of the electron-positron pair is also derived from this model. The perfectly fitting, counter-rotating cones of energy extinguish each other with the release of a combined energy of 1.02 Mev. This results in the reconstitution of the respective unit cells of the space lattice along with the release of the phase transition energy.

[**0029**] The size of the subtle electron is estimated based on the “nuclear radius” of 0.1 fm for the electron, to be approximately 0.005 fm. The lattice constant for the unit cell of space lattice is approximately half of the “nuclear radius” of the electron, i.e,.0.05 fm. This indicates that the space lattice is quite dense compared to atomic bodies.

[**0030**] The model explains the attraction of free electric charges as follows (**FIG. 8**): Opposite charges move towards each other. This movement is due to the equatorial circulation of fluid space lattice from the negative charge to the positive charge and the polar circulation of space lattice from the positive to the negative charge. The attraction of the vortices pull the two partners together. The electrostatic field between separated charges is defined as the flow of the space lattice from the protons to the electrons and the outside return flow of space lattice from the electrons to the protons (see **FIG. 8**). This explanation will eventually lead to a hydromechanical theory of electricity.

[**0031**] Interestingly, there is a net unidirectional flow of fluid space lattice in the axial direction of the free electric charges (**FIG. 8**) flowing in at the negative pole and flowing out at the positive pole. if the charges have a steady parallel orientation, e.g., as in an electric condenser, then a pressure differential must arise in the space lattice around the opposite poles of the condenser. To fully appreciate the significance of this conclusion, we must first examine how material bodies accelerate through the space lattice.

[**0032**] Accelerating objects encounter resistance facing an increased pressure of the space lattice at the front end and a reduced pressure at the rear end of the object. This situation is common to all propulsion methods which apply mechanical force on the physical object. Therefore, it is logical to suggest that a pressure differential of space lattice at the opposite sides of material bodies is always accompanied by a change in the rest or motion of such bodies. To obtain a propulsion force, instead of exerting force on the physical object we should transfer the space lattice that controls the behavior of the object from the rear end of the object to its front end.

[**0033**] Therefore, a sufficiently charged object in which the charges have a steady parallel orientation should behave as an accelerating object, i.e., it should move toward the direction of its positive pole. In fact, this happens due to its own generated space lattice pressure differential Now that the fundamentals are developed, the effect may provide the basis for a new propulsion and gravity cancellation method, and may give us insights into the mechanism of gravitation. A craft utilizing such propulsion method could exhibit inertialess acceleration as it would meet no resistance from the surrounding space lattice.

[**0034**] The theory also allows us to develop strategies for tapping into the energy of the space surrounding us. Separated charges in an electric condenser cause a directional flow of the fluid space lattice. Conversely, if we could induce a directional flow of the space lattice, it would cause a separation of charges in material objects. To explain how such an effect could be produced, first the physical basis of magnetism must be explained.

[**0035**] I have described the physical basis for the electrostatic field as a flow of a space lattice current from the positive charges to the negative charges and back on the outside to the positive charges. I shall now explain how these space lattice currents flow in an electric wire. The space lattice flow which connects the electrons to the protons of the atoms in the wire becomes extended in the length of the wire. This is the same direction as the direction of movement of electrons which, in the current, flow in the outer shell of the wire. The external return flow of the space lattice will be in the opposite direction in the space around the wire. This flow constitutes the magnetic field. In a solenoid, the surrounding space lattice flows in the opposite direction relative to the path of the electrons. The magnetic effect will appear as either N or S magnetic poles. The magnetic poles are mirror images of each other. This suggests that a single isolated magnetic pole cannot exist.

[**0036**] In the electrostatic field, the outside space lattice currents flow between electrically charged particles. The magnetic field, on the other hand exhibits a closed circuit flow of the space lattice along the path of a solenoid or a circuit. Electric and magnetic forces are produced by movement of the space lattice in spirals or whirls.

[**0037**] An electric charge and a magnetic pole do interact even though they do not apply any force to each other. A magnetic pole will bring about a directional orientation of the electrons in a charged object, while an electric charge will cause the electric polarization of the adjacent surface of a magnetic pole.

[**0038**] The final question remains is the superimposition of electrostatic and magnetic fields on each other. The practical significance of the interaction of axially oriented electrostatic and magnetic fields is likely to be small compared to their interaction at an angle. A transversal superimposition of a magnetic field on an electrostatic field will cause the distortion of both fields. The axial flow of the space lattice between the electric poles, which is the physical basis for the electrostatic field, will be distorted, but its circulation will continue. Similarly, the rotating space lattice, which is the basis for the magnetic field, will also be distorted, flowing in irregular circuits.

[**0039**] To study the transverse superimposition of electrostatic and magnetic fields, a shaped electrostatic field must be created (**FIG. 9**). When a magnetic field interacts with the shaped flow of the space lattice, it could result in the formation of a space lattice vortex at the points where the two flows intersect. This is based on previous hypothesis that the space lattice is an incompressible fluid, wherein the intersecting currents cannot penetrate but distort each other's flow patterns.

[**0040**] To expand this shaped electrostatic field into three dimensions, a pyramid is required. I have already discussed the importance of the pyramidal geometry in the formation of matter. The pyramidal geometry may also be important for producing a space lattice perturbation, which is the equivalent of a magnetic field when it follows a rotational symmetry (e.g. a vortex).

[**0041**] If with transversally superimposed electrostatic and magnetic fields a pyramid did indeed produce a vortex in the space lattice, then the rotational flow of the space lattice should cause the separation of charges and an electric current in material bodies, as predicted by this theory. This means that under appropriate conditions, a pyramid could become a power generator.

[**0042**] For power generation in a pyramid, the source of the electrostatic and magnetic fields could potentially be the Earth. The earth has a high voltage electric field such that the earth's surface is negatively charged while the atmosphere is positively charged (Feynman, R. P. (1964) Lectures on Physics, v.2, 9-2, Addison Wesley, Inc., Palo Alto, Calif.). The voltage field is believed to be maintained by the Sun's radiation energy and/or cosmic ray energy and extends from the ionosphere to the surface creating a potential difference of about 400,000 V. The Earth's electrostatic field never seems to get depleted despite all the storm activities and discharging of energy. The voltage gradient is estimated to be 200 V/m around the surface of the earth. The total energy reaching earth at any moment is about 700 MW (400,000 V multiplied by a current of 1800 A). The magnetic field strength of the earth is about 0.5 Gauss and exhibits continuous pulsations with main frequencies of 8 to 16 Hz.

[**0043**] Therefore, the earth's electrostatic and magnetic fields could potentially be utilized to create a space lattice vortex in and around a sufficiently sized pyramid. This hypothesis has been tested experimentally. The earth's electrostatic field was modeled as described in the Examples.

[**0044**] Therefore, there is a need for an improved method and apparatus for the conversion of electrostatic energy as described herein.

SUMMARY OF THE INVENTION

[**0045**] Pursuant to this invention new techniques are described to produce directional currents in the space lattice. The currents are induced either by separated charges, properly positioned magnetic fields or by the interaction of electrostatic and magnetic fields at an angle, specified by the side angle of a pyramid. The induced space lattice currents can be utilized to generate specific effects which may vary from propulsion, to gravity cancellation and power generation.

[**0046**] In particular, an aspect of the present invention is a method producing directional currents in the space lattice utilizing:

[**0047**] (1) separated electric charges; and

[**0048**] (2) the interaction of electrostatic and magnetic fields.

[**0049**] In one preferred embodiment of the invention, electric charges are separated in an oriented fashion. The orientation can be a parallel orientation. The orientation can be obtained in a capacitor. The capacitor can be used for the propulsion of a vehicle. The capacitor can provide lift and lateral movement control for said vehicle.

[**0050**] The interaction of electrostatic and magnetic fields can be set up in a pyramid.

[**0051**] The interaction can produce a vortex in the space lattice.

[**0052**] The vortex can be tapped to produce electric power, which can be obtained by wrapping a coil around the pyramid.

BRIEF DESCRIPTION OF THE DRAWINGS

[**0053**] The following invention will become better understood with reference to the specification, appended claims, and accompanying drawings, where:

[**0054**] **FIG. 1** is a diagram showing the relationship between a spiral and a wave;

[**0055**] **FIG. 2** is a diagram demonstrating the geometry of the space lattice;

[**0056**] **FIG. 3** is a diagram demonstrating the circumvolution cissoid;

[**0057**] **FIG. 4** is a diagram demonstrating the smallest unit of matter;

[**0058**] **FIG. 5** is a diagram showing the rearrangement of the unit cell;

[**0059**] **FIG. 6** is a diagram exhibiting the structures of electron and positron;

[**0060**] **FIG. 7** is a diagram demonstrating the flow coupling of electron and positron;

[**0061**] **FIG. 8** is a diagram showing the attraction of electric charges;

[**0062**] **FIG. 9** is a diagram showing the interaction of a shaped electrostatic field with a magnetic field;

[**0063**] **FIG. 10** is a diagram displaying the arrangement of the capacitors in the craft; and

[**0064**] **FIG. 11** is a diagram demonstrating the pyramid experimental setup.

DESCRIPTION OF THE PREFERRED EMBODIMENT

[**0065**] Pursuant to this invention, novel methods are described to induce directional currents in the space lattice which is postulated to be part of the boundary conditions of the Universe. These methods are useful in providing for the conversion of electrostatic energy into a useful form. The first method relates to generating an unidirectional flow of the space lattice which could be useful in cancelling gravity and generating a propulsion force. A possible configuration for keeping the charges separated and stored is in the form of a capacitor. The shape of the capacitor and the body of such device is important for optimal performance. A pyramidal or conical shape is preferred for the alignment of the capacitors (**FIG. 10**). Optimally, the shape of the capacitor follows the shape of the craft.

[**0066**] In **FIG. 10**, the pyramid **100** is surrounded by a detector coil **102** which is in turn connected to an oscilloscope **104**. In the experimental setup shown, the field is established by a top plate **106** and a ground **108** connected to a source of DC electrical energy **110**.

[**0067**] The conductive elements of the capacitor are layered into the pyramidal or conical segment of the hull of such craft. Optimally, there should be several layers of negatively charged conductors on the outside while a single layer of positively charged conductor in the inside of the craft. All conductive elements should be carefully insulated in order to keep the charges inside the capacitor.

[**0068**] In the center of the craft, a conductive, insulated pole is mounted such that the pole extends out into the exterior at the top of the craft. The tip of that pole has a metal ball mounted on which is also positively charged. When all the capacitor plates are charged, they induce an unidirectional, upward flow of the space lattice. This means that the space lattice pressure above the craft will be higher than the space lattice pressure at the bottom of the craft. This pressure differential is unstable and progresses towards an equilibrium. It will grab the craft and propels it forward in order to eliminate the pressure gradient. This requires that a sufficiently high voltage be applied. My estimation is that about 5 kV is needed to lift every gram of an object. The capacitors incorporated into the hull will give control over lift.

[**0069**] To obtain directional control over all three axes of movement, three capacitors need to be mounted at the bottom of the craft apart from each other. These capacitors, could take the shape of a ball and would be mounted on retractable shafts. The top half of the capacitors would be positively charged while the bottom half is negatively charged. When in flight, the capacitors are retracted and charged based on the needs of directional control. To achieve forward movement, the charges on one capacitor facing the direction of the desired path should be reduced relative to the other two capacitors. This would make the craft tilt forward and start moving in that direction. An alternative method of lateral directional control could be obtained by dividing the capacitor bank in the hull into segments which segments could be charged individually.

[**0070**] Upon landing, the three ball capacitors would be protracted on their shafts and serve as landing gear. On landing, they would be discharged because the earth is the greater ground. On launching, the main capacitors of the body would be activated causing a lift-off. The three capacitor landing gear would then be retracted and charged up so that they can function for lateral directional control. Onboard static generators would supply the voltage required for the operation of the craft.

[**0071**] The most significant aspect of this type of propulsion system would be the loss of inertia during acceleration. Inertial forces pose great obstacles to improving the performance of current propulsion techniques. Since we are now in control of the space lattice surrounding the craft, it will encounter no resistance from the space lattice. The passengers of such craft would feel no forces acting upon their bodies even if the craft achieved enormous accelerations.

[**0072**] This overall design is suitable for the development of a wide range of transportation vehicles, from automobile-type to aircraft and spacecraft. The impact of such a technology on our economy would be enormous. We could eliminate a major factor in environmental pollution, the generation of greenhouse gases which place a great stress on our ecosystem. In the United states, 66% of the greenhouse gases are produced by transportation vehicles and just 34% by power generation. The signs of global warming and the deterioration of the ecosystem are visible from every direction.

[**0073**] The developing countries are bent on establishing the same economical model evolved in the developed countries, representing about 16% of the world population. Worldwide mass deployment of our current energy technologies would most likely cause an environmental catastrophe which may terminate higher life-forms on this planet.

[**0074**] This makes it very important that the focus of our energy sector shifts towards renewable, non-polluting sources of energy. The subject invention offers a solution in this regard. The transversal interaction of electrostatic and magnetic fields in and around a pyramid offers the potential to tap into the electromagnetic field of our planet. I have found that a pyramid becomes a power generator when the electrostatic voltage field gradient of the earth is reproduced around it. Tapping the energy field was as simple as placing a coil around the pyramid.

[**0075**] The optimal size production pyramid is likely going to be large. The optimal material composition may be alternative layers of metal and dielectric to maximize the total amount of charges the pyramid can hold. This is important for increasing the performance of the pyramid. A coil is then wrapped around the pyramid to tap into the generated field. The orientation of the pyramid to the magnetic North pole does not seem to be a prerequisite for operation. Such a pyramid generator will be called the pyramitron.

[**0076**] The following Examples illustrate the features and advantages of the subject invention. Accordingly, it is to be understood that the description in this disclosure is to facilitate comprehension of the invention and should not be construed to limit the scope thereof as persons skilled in the art can, in light of this disclosure, generate additional embodiments without exceeding the scope or departing from the spirit of the claimed invention.

EXAMPLE 1

[**0077**] Demonstration of the pyramitron: For the experiments, I have randomly selected a one-foot base length foam pyramid from a pyramid vendor (The Pyramid Project, Ft. Wayne, Ind.). The outside of the pyramid was covered with aluminum foil. The pyramid was placed on a 2'x2' insulating polyethylene platform equipped with an adjustable height 2'x2' size aluminum top plate, 1/16" thick (**FIG. 10**). The height of the aluminum plate was adjusted as needed and a gap of 1 ¼" between the plate and the tip of the pyramid was used in the experiments.

[**0078**] A high voltage (HV) CRT power source producing 30 kV DC was taken from a television set. I have assumed that an actual energy producing pyramid should be relatively high in order to obtain a large voltage drop from its tip to the ground. Therefore, assuming a height of 150 m for a life-size pyramid and a voltage drop of 200 V/m near the surface of the Earth, the 30 kV voltage is in the range of the voltage-drop expected for the height of a practical size pyramid.

[**0079**] The positive pole was attached to the top aluminum plate. This simulated the positive charge of the atmosphere. One corner of the pyramid was attached to the negative pole of the high voltage power source, while the opposite corner of the pyramid was grounded. This setup served as a model for the electrostatic field distribution around a potential life sized pyramid.

[**0080**] As controls, either a 1'x1' sheet of aluminum foil or an aluminum foil-covered box, having the main dimensions of the test pyramid (1'x1'x⅝"), was used as a negative pole. The detector coils were made by winding up a 24 gauge enamel-coated magnetic wire (20 turns, approximately 8 cm diameter). A Tektronix high-frequency oscilloscope, model no.2236 was used for signal acquisition and analysis.

[**0081**] The first set of experiments were control measurements with the box of the same height and base lengths as the test pyramid. The detector coil was placed on the top of the box. Measurements were taken with and without the high voltage applied. One corner of the box was attached to the HV power source (negative pole) and the opposite corner to the ground. The same arrangement was used for the flat square (1'x1') foil. The peak-to-peak signal amplitude for the box was 8 mV and the signal frequency was 2 MHz. For the flat foil sheet, the signal amplitude was 12 mV with a frequency of 1.43 MHz. The signal form was of a decaying sine wave.

[**0082**] When high voltage was applied to these shapes, a signal amplitude of 14 mV was obtained for the flat sheet and a 16 mV for the box. The signal frequency was 1.54 MHz for the flat sheet and 2 MHz for the box. The wave forms were of decaying sine waves in all these experiments.

[**0083**] When the pyramid was tested without HV, the peak-to-peak signal amplitude was measured at 60 mV with a frequency of 2 MHz. The wave form was different from those of the control experiments; it was a modulated, decaying sine. When the high voltage was applied, the signal amplitude increased up to 180-200 mV, while the frequency remained at 2 MHz. The pyramid produced a signal intensity significantly higher than the controls. When a metal (aluminum) pyramid of the same size (wall thickness 1/16" inch) was tested in the same high voltage field using the same detection coil, a voltage of 1-1.5 V was detected at the frequency of 2 MHz.

[**0084**] The experiments confirmed theoretical assumptions about the pyramidal shape's ability to cause perturbations in the space lattice when the pyramid is placed in transverse electrostatic and magnetic fields. The resulting space lattice vortex causes a separation of charges in the detection coil, producing a voltage signal greatly exceeding control levels. Control values for the flat sheet and the box in the electrostatic field were at the level of the background noise. Thus, when the electrostatic and magnetic fields intersected at 90° or 0°, no effect was observed. However, when the fields intersected at an angle defined by the pyramidal shape, a large signal was produced. The pyramid is clearly instrumental in generating the signal. This finding provides the second experimental proof for the present theory of electric polarities. It also demonstrates that potentially, we may be able to tap into the electromagnetic field of our planet and draw electric energy. The practical implications of this finding are enormous.

I claim:

1. A method producing directional currents in the space lattice utilizing:

(a) separated electric charges; and

(b) the interaction of electrostatic and magnetic fields.

2. The method of claim 1 wherein said electric charges are separated in an oriented fashion.

3. The method of claim 2 wherein said orientation is a parallel orientation.

4. The method of claim 3 wherein said orientation is obtained in a capacitor.

5. The method of claim 4 wherein said capacitor is used for the propulsion of a vehicle.

6. The method of claim 5 wherein said capacitor is providing lift and lateral movement control for said vehicle.

7. The method of claim 1 wherein interaction of electrostatic and magnetic fields is set up in a pyramid.

8. The method of claim 1 wherein said interaction is producing a vortex in the space lattice.

9. The method of claim 8 wherein said vortex can be tapped to produce electric power.

10. The method of claim 9 wherein said electric power is obtained by wrapping a coil around the pyramid.

* * * * *

US009160156B2

(12) **United States Patent**
Allen et al.

(10) **Patent No.: US 9,160,156 B2**
(45) **Date of Patent: Oct. 13, 2015**

(54) **SYSTEM FOR HARVESTING ATMOSPHERIC ELECTRICITY**

(75) Inventors: **Edward H. Allen**, Lancaster, CA (US); **Sergey Macheret**, Palmdale, CA (US)

(73) Assignee: **Lockheed Martin Corporation**, Bethesda, MD (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 357 days.

(21) Appl. No.: **13/276,259**

(22) Filed: **Oct. 18, 2011**

(65) **Prior Publication Data**

US 2013/0093261 A1 Apr. 18, 2013

(51) **Int. Cl.**

H05F 7/00	(2006.01)
H02G 13/00	(2006.01)
B64D 45/02	(2006.01)

(52) **U.S. Cl.**
CPC ***H02G 13/20*** (2013.01); ***B64D 45/02*** (2013.01); *H05F 7/00* (2013.01)

(58) **Field of Classification Search**
CPC H02G 13/20; B64D 45/02; H05F 7/00
USPC 361/231, 230
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,719,829	A *	3/1973	Vaill	307/149
4,017,767	A *	4/1977	Ball	361/117
4,950,978	A *	8/1990	Govaert	324/72
5,726,855	A *	3/1998	Mourou et al.	361/213
8,045,314	B2 *	10/2011	Ibok	361/231
2014/0231679	A1 *	8/2014	Kremeyer et al.	250/505.1

OTHER PUBLICATIONS

Tzzortzakis et al. "Concatenation of plasma filaments created in air by femtosecond laser infrared laser pulses", 2003, Applied Physics B, vol. 767, pp. 609-612.*

Zhou et al. , Revival of femtosecond laser plasma filaments in air by a nanosecond laser, Jun. 24, 2009, Optic Express, vol. 17, No. 14, pp. 1 -7.*

Assion et al., "Control of Chemical Reactions by Feedback-Optimized Phase-Shaped Femtosecond Laser Pulses", Science, Oct. 30, 1998, pp. 919-922, vol. 282.

* cited by examiner

Primary Examiner — Zeev V Kitov
(74) *Attorney, Agent, or Firm* — Baker Botts L.L.P.

(57) **ABSTRACT**

Systems, methods and apparatus for harvesting atmospheric electricity are provided. The system includes a laser configured to form a plasma filament and a collector configured to collect electricity flowing along the plasma filament. The plasma filament comprises an electrically conducting plasma filament. Atmospheric electricity may be collected by having the plasma filament form at least a part of a conducting path: (1) between ground and a cloud, (2) between differently charged regions of the same cloud, (3) between differently charged regions of different clouds, and (4) between different regions of atmosphere, where there is a vertical voltage gradient. When the plasma filament is not long enough to form the entire conducting path, a lightning may be triggered to complete the conducting path needed to collect atmospheric electricity.

21 Claims, 6 Drawing Sheets

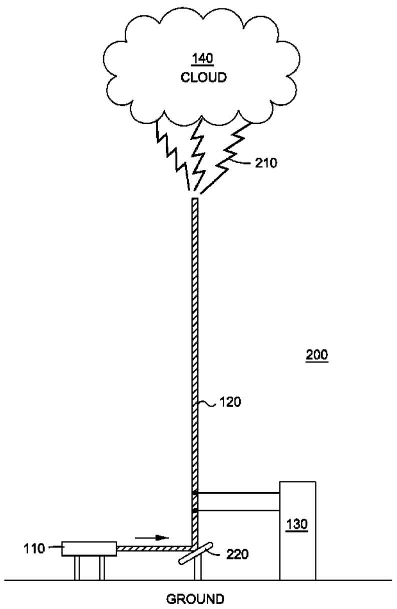

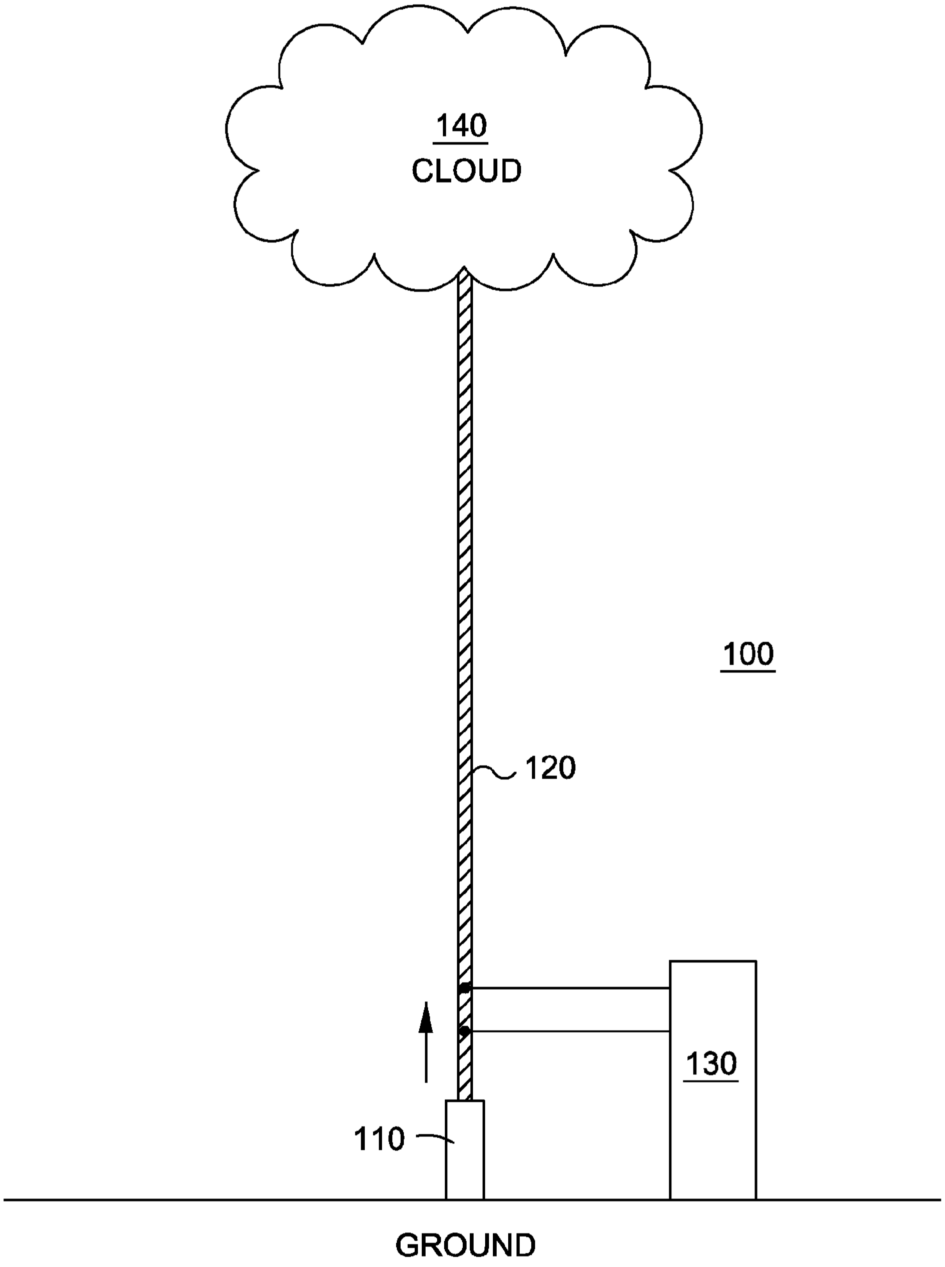
140
CLOUD
100
120
130
110
GROUND

FIG. 1

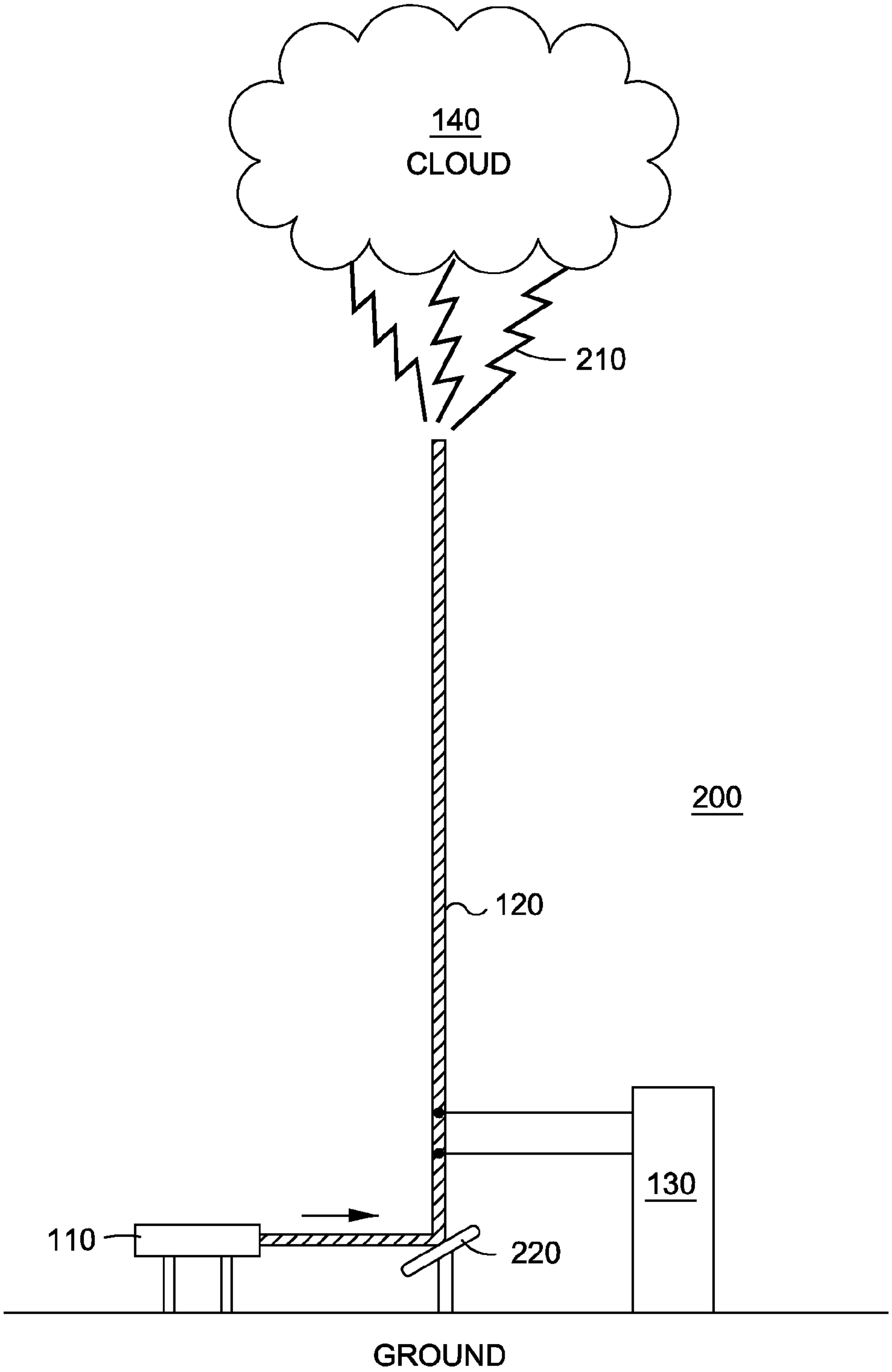
140
CLOUD
210
200
120
130
110
220
GROUND

FIG. 2

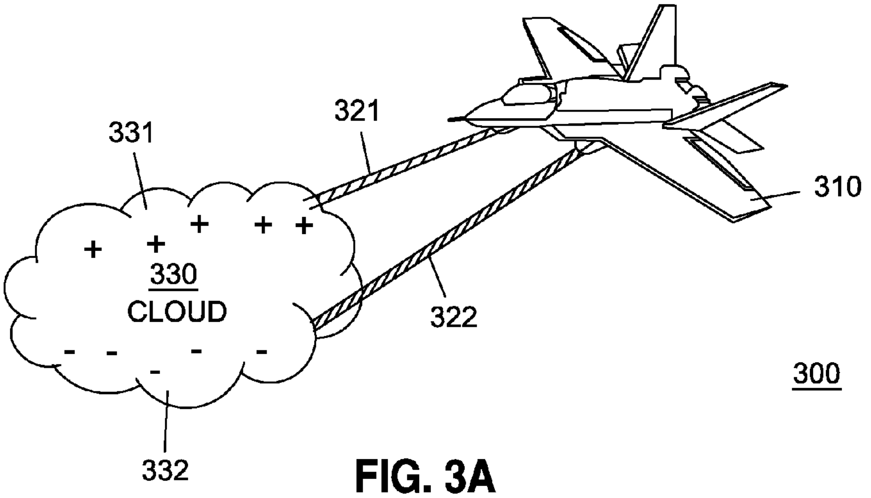
CLOUD
+ + + + + +
- - - - -

FIG. 3A

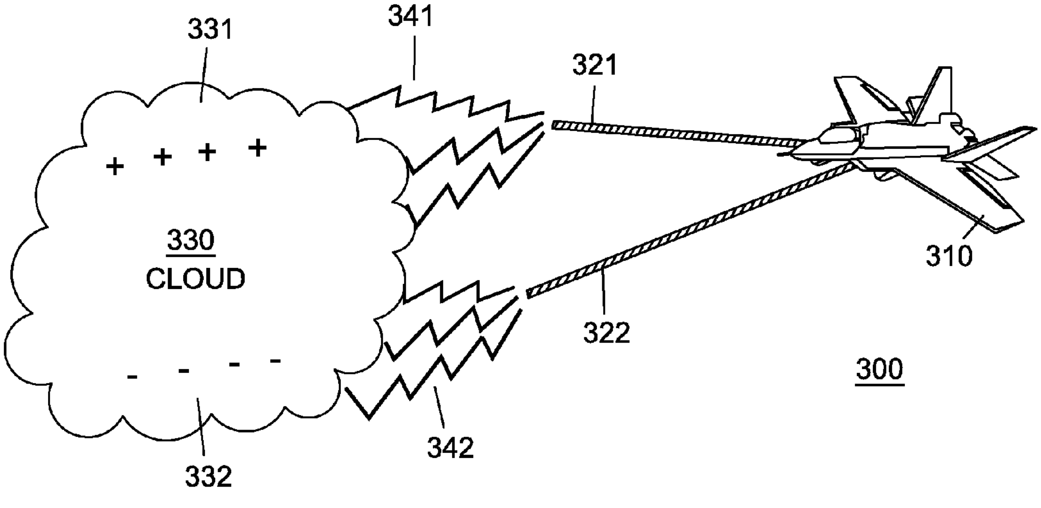
CLOUD
+ + + +
- - - -

FIG. 3B

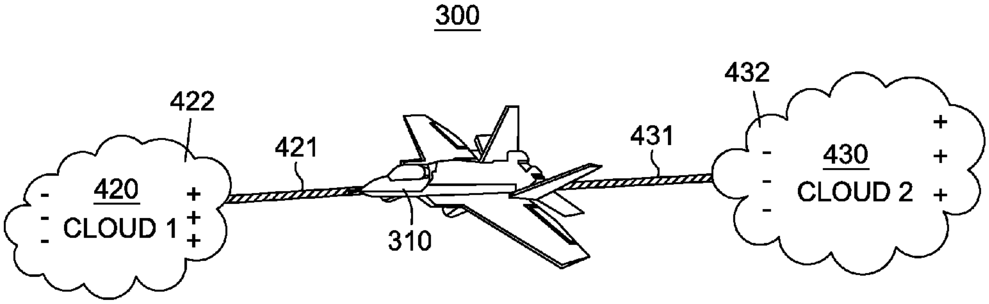
300
422
421
420
CLOUD 1
-
-
-
+
+
+
310
431
432
430
CLOUD 2
-
-
-
-
+
+
+
+

FIG. 4

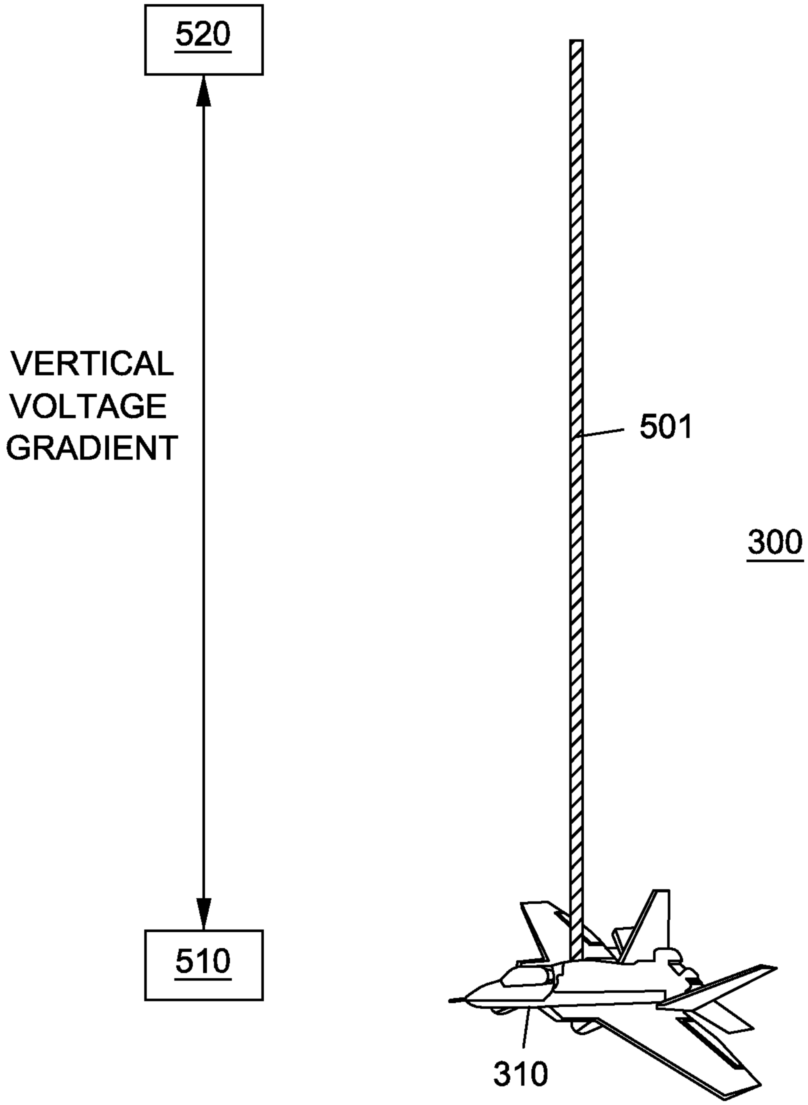
520
VERTICAL
VOLTAGE
GRADIENT
510
501
300
310

FIG. 5

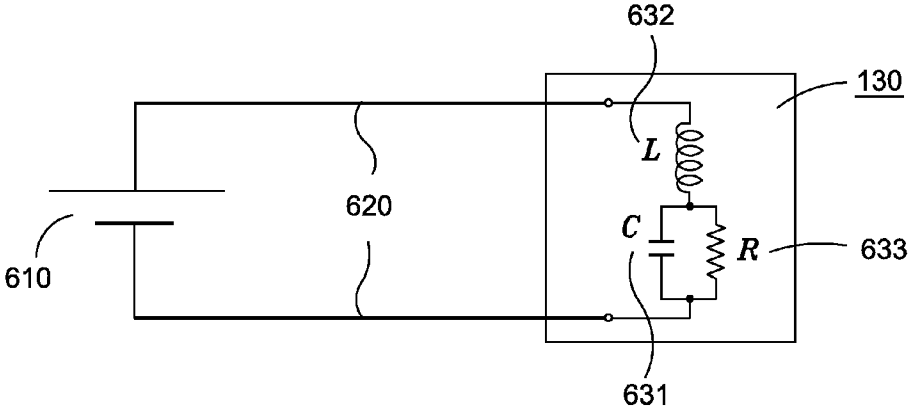
632
130
L
620
C
R
633
610
631

FIG. 6

SYSTEM FOR HARVESTING ATMOSPHERIC ELECTRICITY

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

Not applicable.

FIELD OF THE INVENTION

The subject technology relates in general to a system for harvesting atmospheric electricity, and more particularly to a system that uses a laser to form an electrically conducting plasma filament for harvesting atmospheric electricity.

BACKGROUND

It is well known that large quantities of electrical energy are present in the atmosphere and in lightning. Lightning was one of the first forms of electricity harnessed in the modern age by Benjamin Franklin in his famous kite experiment.

A lightning discharge typically contains on the order of 10^{10} Joules of energy. It has been estimated that the total electrical power of lightning across the Earth is of the order of 10^{12} watts. Collection of electrical energy from clouds associated with lightning is a problem that needs to be resolved.

Lightning is but a small part of the total electrical activity of the atmosphere. When a local build up of charge above the Earth exceeds the local breakdown potential of the atmosphere a lightning discharge occurs. However, there is a continual invisible flow of charge from Ionosphere to Earth occurring day and night over the entire surface of the globe, which exceeds the global lightning power output by many times. The source of this flow of charge from the atmosphere to Earth is the Sun. Radiation from the Sun helps form most of the ions found in the Ionosphere, a highly charged region above the atmosphere, through ionization of atmospheric molecules. Solar radiation, particularly in the UV and soft x-ray bands, consist of photons whose energies are well suited to ionization. The Sun also emits continuously a solar wind of positively charged particles. These are captured by the Earth's magnetic field and further contribute to the Ionosphere. The positively charged region in the Ionosphere in turn induces (by electrostatic induction) a negative charge on the surface of the Earth. The Earth becomes in effect an enormous spherical capacitor. A potential gradient or electric field is thus established between the two "plates" of this capacitor, the Ionosphere (or Electrosphere) and the surface of the Earth. While the upper strata of the atmosphere conduct electricity reasonably well, the lower levels act as an insulator or dielectric. Near the surface of the Earth, this electrostatic potential gradient is on the order of about 100 Vm^{-1} in summer, rising to 300 Vm^{-1} in winter. This flow of charge can be tapped and directed to provide useable electrical power. This source of atmospheric electricity has the following advantages: (1) Simple and robust technology; (2) Low cost technology—much cheaper than photovoltaics or wind turbines; (3) Available day and night in all weather conditions—in fact, more power is produced at night than during the day; and (4) Available at any point on the Earth's surface. Collection of this source of atmospheric electricity is another problem that needs to be resolved.

SUMMARY

According to various aspects of the subject technology, a system for harvesting atmospheric electricity is provided that solves some or all of the foregoing problems. In some aspects, the system may be used to harvest atmospheric electricity arising from an electrostatic potential gradient: (1) between ground and a cloud, (2) between differently charged regions of the same cloud, (3) between differently charged regions of different clouds, and (4) between different regions of atmosphere, where there is a vertical voltage gradient. In other aspects, the system may be used to harvest atmospheric electricity arising from an electrostatic potential gradient between the Ionosphere (or Electrosphere) and the surface of the Earth.

According to various aspects of the subject technology, a system for harvesting atmospheric electricity is provided. The system comprises a laser configured to form a plasma filament. The system also comprises a collector configured to collect electricity flowing along the plasma filament.

According to various aspects of the subject technology, a method for harvesting atmospheric electricity is provided. The method comprises forming a plasma filament with a laser. The method also comprises collecting electricity flowing along the plasma filament.

According to various aspects of the subject technology, an apparatus for harvesting atmospheric electricity is provided. The apparatus comprises means for forming a plasma filament with a laser. The system also comprises means for collecting electricity flowing along the plasma filament.

It is understood that other configurations of the subject technology will become readily apparent to those skilled in the art from the following detailed description, wherein various configurations of the subject technology are shown and described by way of illustration. As will be realized, the subject technology is capable of other and different configurations and its several details are capable of modification in various other respects, all without departing from the scope of the subject technology. Accordingly, the drawings and detailed description are to be regarded as illustrative in nature and not as restrictive.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are included to provide further understanding of the subject technology and are incorporated in and constitute a part of this specification, illustrate aspects of the subject technology and together with the description serve to explain the principles of the subject technology.

FIG. **1** illustrates a first example of a system for harvesting atmospheric electricity. The system comprises a laser configured to form a plasma filament and a collector configured to collect electricity flowing along the plasma filament. The plasma filament forms a conducting path between ground and a cloud.

FIG. **2** illustrates a second example of a system for harvesting atmospheric electricity. A plasma filament and a lightning together form a conducting path between ground and a cloud. A condensing mirror directs the beam of laser towards the cloud.

FIG. **3**A illustrates a third example of a system for harvesting atmospheric electricity. A laser system situated on an airplane forms two plasma filaments that together complete a conducting path between a first region of a cloud and a second region of the cloud. The first region of the cloud and the second region of the cloud carry opposite electric charges. A collector configured to collect electricity flowing along the two plasma filaments is also situated on the same airplane as the laser system.

FIG. **3**B illustrates a fourth example of a system for harvesting atmospheric electricity. It is similar to FIG. **3**A, except two lightnings complete the conducting path between a first region of a cloud and a second region of the cloud (i.e., a first lightning electrically connects the first region of the cloud to one of the plasma filaments and a second lightning electrically connects the second region of the cloud to the other plasma filament).

FIG. **4** illustrates a fifth example of a system for harvesting atmospheric electricity. A laser system situated on an airplane forms two plasma filaments that together complete a conducting path between a first cloud and a second cloud. The region of the first cloud that is connected to the first plasma filament and the region of the second cloud that is connected to the second plasma filament carry opposite electric charges.

FIG. **5** illustrates a sixth example of a system for harvesting atmospheric electricity. A laser situated on an airplane forms a plasma filament that completes a conducting path between a first region of atmosphere and a second region of atmosphere. A collector situated on the same airplane as the laser can collect electricity flowing along the plasma filament, because a vertical voltage gradient exists between the first region of atmosphere and the second region of atmosphere.

FIG. **6** illustrates an example of a collector configured to collect electricity flowing along the plasma filament. The collector may be a charge collection circuit characterized by capacitance, inductance, and resistance. The electricity may be collected by charging a capacitor.

DETAILED DESCRIPTION

The detailed description set forth below is intended as a description of various configurations of the subject technology and is not intended to represent the only configurations in which the subject technology may be practiced. The appended drawings are incorporated herein and constitute a part of the detailed description. The detailed description includes specific details for the purpose of providing a thorough understanding of the subject technology. However, it will be apparent to those skilled in the art that the subject technology may be practiced without these specific details. In some instances, well-known structures and components are shown in block diagram form in order to avoid obscuring the concepts of the subject technology. Like components are labeled with identical element numbers for ease of understanding.

FIG. **1** illustrates a first example of a system **100** for harvesting atmospheric electricity. The system **100** comprises a laser **110** configured to form a plasma filament **120** and a collector **130** configured to collect electricity flowing along the plasma filament **120**. The plasma filament **120** forms a conducting path between ground and a cloud **140**.

The plasma filament **120** comprises an electrically conducting plasma filament. The plasma comprises of electrons and positive ions. The electrically conducting plasma filament may be formed in atmosphere using an ultra short pulse laser (USPL). The USPL forms a plasma filament by raising the conductivity of a very small diameter channel (~100 microns or less) for a very brief time. The propagation of high power (~10^{14} W/cm^2) pulses from the USPL is accompanied by filamentation—self-channeling of femtosecond laser pulses in stable high-intensity light filaments with ~100 microns diameters (so that the Kerr effect focusing balances defocusing due to plasma formation). This filamentation keeps the beam virtually free of diffraction divergence. The long plasma strings have high electron densities (10^{16}/cm^3) which support high conductivities. The ionization of air at these intensities results from multi-photon processes. Filamentation has been produced and detected at altitudes as high as 13-20 km. Almost no beam scattering occurs in filamentation phenomenon, and virtually all the pulse energy is invested in ionization of air molecules. With the typical electron density and channel diameter, a 160 mJ pulse is sufficient to generate a 1 km long channel.

Once initiated, plasma filaments cannot go on forever, and typically only propagate a kilometer or so before energy depletion and exhaustion. However, by using two coaxial USPL projection optics having different time-lensing parameters, a near and far filament could be initiated in a "daisy-chain" to extend the total length of the conductive channel. The onset of filamentation in the far filament is time-lensed to coincide with the exhaustion range of the near filament. Additionally, using negative chirp, femtosecond laser pulses can propagate almost without loss until different frequency components of the wave packet, propagating at different speeds due to dispersion in air, come together at a certain desired location so that the laser intensity there exceeds the filamentation threshold, resulting in long ionized filaments created at a distance of up to several miles from the laser source. In other words, negative chirp is used to essentially focus a beam of the laser at a desired distance to form the plasma filament there. Accordingly, this technique may be used to move the starting point and to extend the length of the plasma filament.

The size, electrical conductivity, and lifetime of the plasma filaments are enhanced by an auxiliary radio-frequency electromagnetic waves or microwaves. The plasma filament may act like an antenna to absorb the auxiliary radio-frequency (RF) electromagnetic waves or microwaves.

The pulsed collection of electric charge and energy is repeated with an optimal (possibly very high) repetition rate of the pulsed laser and of the auxiliary RF waves or microwaves. The energy collected per one pulse should exceed the energy spent on creating and enhancing the plasma filaments, so that the net energy collected is positive. The optimal pulse repetition rate (PRR) depends upon atmospheric conditions which can vary widely over even short time durations. As a general rule, one tries to tune the repetition rate to the plasma relaxation time (the ion-electron recombination time) so that the conductivity of the filament remain as high as is optimal given the atmospheric conditions at the time. A feedback control system that varies the PRR as well as the rest of the pulse-shaping parameters may be employed to optimize the net power production of the system. For more information regarding pulse shaping and tailoring, which is the central theme of the field called "quantum control", please see A. Assion, T. Baumert, M. Bergt, T. Brixner, B. Kiefer, V. Seyfried, M. Strehle, G. Gerber (30 October). "Control of Chemical Reactions by Feedback-Optimized Phase-Shaped Femtosecond Laser Pulses". Science 282 (5390): 919-922. doi: 10.1126/science.282.5390.919. PMID 9794756.

FIG. **2** illustrates a second example of a system **200** for harvesting atmospheric electricity. A plasma filament **120** and a lightning **210** together form a conducting path between ground and a cloud **140**. A condensing mirror **220** directs the beam of laser **110** towards the cloud **140**. A collector **130** is configured to collect electricity flowing along the plasma filament **120**.

In the case of electricity collection from the cloud **140**, the initial and RF-enhanced plasma filaments can be relatively short, not reaching all the way from ground to the cloud **140**. The relatively short plasma filament can be optimized so that it would trigger propagation of the so-called leader (as in natural lightning), which would effectively use the energy and potential of the cloud to eventually create an electrically conducting path between ground and the cloud **140**. The final conducting path would comprise of a plasma filament **120** and a lightning **210**.

It is not shown in the figures, but alternatively both the laser **110** and the collector **130** may be placed on a mobile device, which moves together with the cloud. One such mobile device may be a vehicle moving on the ground, but another such mobile device may be an airplane flying in the air. In the case of an airplane flying in the air, plasma filaments may be formed by the laser to reach both ground and the cloud at the same time. In the event the plasma filament directed at ground is not long enough to reach ground, a lightning may also be formed to complete the conducting path to ground.

FIG. **3**A illustrates a third example of a system **300** for harvesting atmospheric electricity. A laser system situated on an airplane **310** forms two plasma filaments (**321**, **322**) that together complete a conducting path between a first region **331** of a cloud **330** and a second region **332** of the cloud **330**. The first region **331** of the cloud **330** and the second region **332** of the cloud **330** carry opposite electric charges. In the example shown here, the first region **331** of the cloud **330** carries positive electric charges, while the second region **332** of the cloud **330** carries negative electric charges. A collector configured to collect electricity flowing along the two plasma filaments is also situated on the same airplane **310** as the laser. Positive and negative charges are naturally separated within the cloud, so this example makes use of the electrostatic potential gradient between different regions of the same cloud.

FIG. **3**B illustrates a fourth example of the system **300** for harvesting atmospheric electricity. It is similar to FIG. **3**A, except two lightnings (**341**, **342**) complete the conducting path between a first region **331** of a cloud **330** and a second region **332** of the cloud **330** (i.e., a first lightning **341** electrically connects the first region **331** of the cloud **330** to one of the plasma filaments **321** and a second lightning **342** electrically connects the second region **332** of the cloud **330** to the other plasma filaments **322**). The lightning formation is again because the initial and RF-enhanced plasma filaments can be relatively short, not reaching all the way from the collector situated on airplane **310** to the cloud **330**. These relatively short plasma filaments can be optimized so that they would trigger propagation of the so-called leader (as in natural lightning), which would eventually create an electrically conducting path between the two regions of cloud **330** via a connection through airplane **310**. The final conducting path would comprise of two plasma filaments (**321**, **322**) and two lightnings (**341**, **342**). Of course, one of the lightnings may not be formed, if one of the plasma filament can reach all the way from airplane **310** to cloud **330**.

FIG. **4** illustrates a fifth example of the system **300** for harvesting atmospheric electricity. A laser system situated on an airplane **310** forms two plasma filaments (**421**, **431**) that together complete a conducting path between a first cloud **420** and a second cloud **430**. The region **422** of the first cloud **420** that is connected to the first plasma filament **421** and the region **432** of the second cloud **430** that is connected to the second plasma filament **431** carry opposite electric charges. In the example shown here, region **422** of cloud **420** carries positive electric charges, while region **432** of cloud **430** carry negative electric charges. Positive and negative charges are naturally separated within the cloud, so this example makes use of the electrostatic potential gradient between different regions of different clouds. Once again, in case one of the plasma filament cannot reach all the way from airplane **310** to any one of the clouds, then lightning may be formed to complete an electrically conducting path from airplane **310** to any one of the clouds. But these embodiments are not shown in the figures.

FIG. **5** illustrates a sixth example of the system **300** for harvesting atmospheric electricity. A laser system situated on an airplane **310** forms a plasma filament **501** that completes a conducting path between a first region **510** of atmosphere and a second region **520** of atmosphere. A collector situated on the same airplane **310** as the laser system can collect electricity flowing along the plasma filament **501**, because a vertical voltage gradient exists between the first region **510** of atmosphere and the second region **520** of atmosphere. Because no clouds are needed for this embodiment, atmospheric electricity collection is possible under fair weather conditions (relying on the naturally existing vertical voltage gradient in the atmosphere) as well as under foul weather conditions.

FIG. **6** illustrates an example of a collector **130** configured to collect electricity flowing along a plasma filament **620**. The collector may be a charge collection circuit characterized by capacitance **631**, inductance **632**, and resistance **633**. The electricity may be collected by charging a capacitor associated with capacitance **631**. Battery **610** represents these naturally existing voltage sources: (1) between ground and a cloud, (2) between differently charged regions of the same cloud, (3) between differently charged regions of different clouds, and (4) between different regions of atmosphere, where there is a vertical voltage gradient. The plasma filament **620** can, for the purpose of analysis and optimization of the charge collection process, be viewed as electromagnetic transmission lines characterized by their effective capacitance, inductance, and resistance per unit length. Accordingly, the parameters of the charge collection circuit and of the effective transmission line representing the plasma filaments are optimized, so as to maximize the net charge and energy collected (minus the energy spent on generating and enhancing the plasma filaments).

The foregoing description is provided to enable a person skilled in the art to practice the various configurations described herein. While the subject technology has been particularly described with reference to the various figures and configurations, it should be understood that these are for illustration purposes only and should not be taken as limiting the scope of the subject technology.

There may be many other ways to implement the subject technology. Various functions and elements described herein may be partitioned differently from those shown without departing from the scope of the subject technology. Various modifications to these configurations will be readily apparent to those skilled in the art, and generic principles defined herein may be applied to other configurations. Thus, many changes and modifications may be made to the subject technology, by one having ordinary skill in the art, without departing from the scope of the subject technology.

It is understood that the specific order or hierarchy of steps in the processes disclosed is an illustration of exemplary approaches. Based upon design preferences, it is understood that the specific order or hierarchy of steps in the processes may be rearranged. Some of the steps may be performed simultaneously. The accompanying method claims present elements of the various steps in a sample order, and are not meant to be limited to the specific order or hierarchy presented.

A phrase such as “an aspect” does not imply that such aspect is essential to the subject technology or that such aspect applies to all configurations of the subject technology. A disclosure relating to an aspect may apply to all configurations, or one or more configurations. An aspect may provide one or more examples of the disclosure. A phrase such as an "aspect" may refer to one or more aspects and vice versa. A phrase such as an "embodiment" does not imply that such embodiment is essential to the subject technology or that such embodiment applies to all configurations of the subject technology. A disclosure relating to an embodiment may apply to all embodiments, or one or more embodiments. An embodiment may provide one or more examples of the disclosure. A phrase such an "embodiment" may refer to one or more embodiments and vice versa. A phrase such as a "configuration" does not imply that such configuration is essential to the subject technology or that such configuration applies to all configurations of the subject technology. A disclosure relating to a configuration may apply to all configurations, or one or more configurations. A configuration may provide one or more examples of the disclosure. A phrase such as a "configuration" may refer to one or more configurations and vice versa.

Furthermore, to the extent that the term "include," "have," or the like is used in the description or the claims, such term is intended to be inclusive in a manner similar to the term "comprise" as "comprise" is interpreted when employed as a transitional word in a claim.

The word "exemplary" is used herein to mean "serving as an example, instance, or illustration." Any embodiment described herein as "exemplary" is not necessarily to be construed as preferred or advantageous over other embodiments.

A reference to an element in the singular is not intended to mean "one and only one" unless specifically stated, but rather "one or more." The term "some" refers to one or more. Underlined and/or italicized headings and subheadings are used for convenience only, do not limit the subject technology, and are not referred to in connection with the interpretation of the description of the subject technology. All structural and functional equivalents to the elements of the various configurations described throughout this disclosure that are known or later come to be known to those of ordinary skill in the art are expressly incorporated herein by reference and intended to be encompassed by the subject technology. Moreover, nothing disclosed herein is intended to be dedicated to the public regardless of whether such disclosure is explicitly recited in the above description.

What is claimed is the following:

1. A system for harvesting atmospheric electricity, comprising:

a laser configured to form a plasma filament comprising at least a first filament and a second filament, wherein onset of formation of the second filament coincides with an exhaustion range of the first filament;

an auxiliary source of electromagnetic energy operable to enhance the plasma filament by applying electromagnetic energy to the plasma filament; and

a collector configured to collect electricity flowing along the plasma filament.

2. The system of claim **1**, wherein the plasma filament comprises an electrically conducting plasma filament.

3. The system of claim **1**, wherein the laser is a pulsed laser.

4. The system of claim **3**, wherein the pulsed laser is pulsed at an optimal repetition rate.

5. The system of claim **1**, wherein the electromagnetic energy comprises at least one of radio-frequency energy or microwave energy.

6. The system of claim **1**, wherein the plasma filament forms at least a part of a conducting path between ground and a cloud.

7. The system of claim **6**, wherein at least one lightning forms another part of the conducting path between ground and the cloud.

8. The system of claim **1**, wherein the plasma filament forms at least a part of a conducting path between a first region of a cloud and a second region of the cloud.

9. The system of claim **8**, wherein at least one lightning forms another part of the conducting path between the first region of the cloud and the second region of the cloud.

10. The system of claim **1**, wherein:

the laser is further configured to form a second plasma filament,

the plasma filament is connected to a region of a first cloud, and

the second plasma filament is connected to a region of a second cloud.

11. The system of claim **1**, wherein the plasma filament completes a conducting path between a first region of atmosphere and a second region of atmosphere.

12. The system of claim **11**, wherein a vertical voltage gradient exists between the first region of atmosphere and the second region of atmosphere.

13. The system of claim **1**, wherein the collector comprises a charge collection circuit.

14. The system of claim **1**, wherein the laser is situated on ground.

15. The system of claim **1**, wherein the laser is situated on a mobile device.

16. The system of claim **15**, wherein the mobile device is an airplane.

17. The system of claim **1**, wherein negative chirp is used to move the starting point of formation of the plasma filament to a desired location.

18. A method for harvesting atmospheric electricity, the method comprising:

forming a plasma filament with a laser, wherein the plasma filament comprises at least a first filament and a second filament, and wherein onset of formation of the second filament coincides with an exhaustion range of the first filament;

applying electromagnetic energy to the plasma filament to enhance the plasma filament; and

collecting electricity flowing along the plasma filament.

19. An apparatus for harvesting atmospheric electricity, the apparatus comprising:

means for forming a plasma filament with a laser, wherein the plasma filament comprises at least a first filament and a second filament, and wherein onset of formation of the second filament coincides with an exhaustion range of the first filament;

means for applying electromagnetic energy to the plasma filament to enhance the plasma filament; and

means for collecting electricity flowing along the plasma filament.

20. The system of claim **1**, wherein the laser is an ultra short pulse laser (USPL).

21. The system of claim **1**, wherein the laser comprises two coaxial USPL projection optics having different time-lensing parameters.

* * * * *

http://www.ava.fmi.fi/~tjt/fairw.html Go MAR MAY JUL
18 captures 23
5 Feb 2005 - 30 Jun 2017 2005 2007 2011 About this capture

logo Maan keski-ilmakehä / Geofysiikan tutkimus / Ilmatieteen laitos

---

Fair-weather atmospheric electricity

Basic facts about atmospheric electricity

Atmospheric electricity involves phenomena which are connected with the separation of electric charges in the sub-ionospheric atmosphere (below about 100 km height). In the ionosphere and magnetosphere there occur strong electric currents originating directly from the solar-terrestrial interaction; in the lower atmosphere, there flows a much weaker electric current in the so-called global circuit, which is maintained by the thunderstorm activity. **Charge separation** takes place in three ways:

thermodynamically
In a thundercloud, small ice crystals collide with rime-growing graupels; the crystals gain positive charge, the graupels negative (the microscopic mechanism is not yet well known).
Convection in the thundercloud carries the ice crystals to the cloud top, the heavier graupels staying in the mid-cloud: a macroscopic dipole structure forms.

by radiation ionization
Cosmic and radioactive radiation ionize air, and equal numbers of molecular-size positive and negative small ions are formed; air becomes (weakly) electrically conductive.
Small ions are also attached to airborne dust (aerosol), which thus regularizes the number of small ions.

by collision ionization
Lightning and other discharges in the thundercloud ionize air temporarily into electrically conducting channels.

The mobility of electrically charged particles depends strongly on their size and on the density of the medium. The mobility of small ions in the lower atmosphere corresponds to a (terminal) speed of about 1 cm/s in an electric field of 100 V/m. Dust particles are slower by a few orders of magnitude, so that the electric conductivity of air is practically that due to the small ions. In the stratosphere and mesosphere (10 - 90 km) air becomes thinner and cosmic radiation intensifies so that the conductivity increases exponentially with height: while the electric conductivity of air near the ground is of the order of 20 fS/m (femtosiemens/metre), in the ionosphere it is already as high as that of the ground. (In the ionosphere, the electric conductivity becomes also anisotropic: the geomagnetic field has an increasing influence on the ion motion.)

The thundercloud charge centres, accumulating tens of coulombs of electricity, are discharged mainly by lightning: cloud flashes (most abundant) cause mutual neutralization of the centres; the lower centre is also discharged to the ground - by negative ground flashes - and charges up the earth (the positive centre is discharged similarly, but by a smaller amount). An excess charge will be left in the upper positive centre, and it leaks by conduction to the surrounding air, about one ampere per thunderstrom cell. Because of the exponentially increasing conductivity, most of this leak current is guided to the ionosphere, where it is distributed over the globe and charges the upper atmosphere to a potential of about 300 kV with respect to the ground. This "ionospheric potential" maintains the so-called fair-weather current, whose density is about 2 pA/m^2 (picoamperes per square metre). According to Ohm's law, the fair-weather current density and the electric conductivity are associated with a downward electric field, about 100 V/m near the ground. The number of simultaneously active thunder cells ("thunderstorms") over the globe is about 1000-2000, so the whole circuit carries a current of about 1000 amperes.

http://www.ava.fmi.fi/~tjt/fairw.html Go MAR MAY JUL
18 captures 23
5 Feb 2005 - 30 Jun 2017 2005 2007 2011
About this capture

threat comes from a lightning strike.

The ion balance near the ground can most simply be described by the equation

$q = an^2 + bnN$

where q is ionization, n is the density of small ions (both polarities separately, assumed equal), N is the density of aerosol, and a and b are recombination coefficients, which describe the neutralization of ions with each other or with aerosol particles. a and b are both about 1.6×10^{-12} m^3s^{-1}. Ionization in southern Finland is about 6.6×10^6 $m^{-3}s^{-1}$, one fifth or sixth coming from radioactivity. The small-ion densities are both about 5×10^8 m^{-3} and the aerosol density is more than tenfold, almost 10^{10} m^{-3}.

It should be noted that the notion "dirt increases electric conductivity" is valid for surfaces, for example insulators, but the matter is contrary for air: dirty (dusty) air has poorer conductivity. However, if the "dirt" is a radioactive pollutant, the higher ionization increases the electric conductivity. Such an episode happened in May 1986, when an iodine cloud from the Chernobyl emission arrived at southern Finland: the electric conductivity grew tenfold, but returned to a lower level in a few days, because the half-life of radioactive iodine is 8 days.

Information on international atmospheric electricity research (including lightning research) can be found at the address Atmospheric Electricity HomePage, which among others has a link to a comprehensive English glossary of atmospheric electricity.

US008810049B2

(12) **United States Patent**
McCowen

(10) **Patent No.: US 8,810,049 B2**
(45) **Date of Patent: Aug. 19, 2014**

(54) **ENERGY COLLECTION**

(75) Inventor: **Clint McCowen**, Navarre, FL (US)

(73) Assignee: **Ion Power Group, LLC**, Navarre, FL (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 88 days.

(21) Appl. No.: **13/569,133**

(22) Filed: **Aug. 7, 2012**

(65) **Prior Publication Data**

US 2012/0299559 A1 Nov. 29, 2012

Related U.S. Application Data

(63) Continuation-in-part of application No. 12/255,130, filed on Oct. 21, 2008, now Pat. No. 8,686,575, which is a continuation of application No. 11/358,264, filed on Feb. 21, 2006, now Pat. No. 7,439,712.

(51) **Int. Cl.**

F02B 63/04	(2006.01)
F03G 7/08	(2006.01)
H02K 7/18	(2006.01)
F03D 9/00	(2006.01)
H02P 9/04	(2006.01)

(52) **U.S. Cl.**
USPC .... **290/1 R**; 290/44; 290/55; 310/300; 310/310; 416/95

(58) **Field of Classification Search**
CPC Y02E 10/721; Y02E 10/70; Y02E 10/722; B32B 2262/101; C01B 31/00
USPC .... 290/1 R, 44, 55
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427	A		5/1901	Palencsar	307/149
911,260	A		2/1909	Pennock	307/149
1,014,719	A		1/1912	Pennock	307/149
2,473,819	A		6/1949	Pittman	315/36
3,532,959	A		10/1970	Erickson	320/166
3,780,722	A		12/1973	Swet	126/680
3,946,227	A		3/1976	Bingham	250/281
4,079,225	A		3/1978	Warner	219/110
4,092,250	A		5/1978	Sano et al.	210/500.29
4,104,696	A		8/1978	Cochran, Jr.	361/229
4,146,800	A	*	3/1979	Gregory et al.	290/44
4,180,056	A		12/1979	Schnabel et al.	126/660
4,201,197	A	*	5/1980	Dismer	126/600
4,206,396	A	*	6/1980	Marks	322/2 A
4,224,496	A		9/1980	Riordan et al.	219/110
4,307,936	A		12/1981	Ochiai	385/120

(Continued)

FOREIGN PATENT DOCUMENTS

WO	2007098341	8/2007

OTHER PUBLICATIONS

U.S. Appl. No. 12/637,720, Dec. 14, 2009, McCowen.

Primary Examiner — Pedro J Cuevas
(74) *Attorney, Agent, or Firm* — Benjamin A. Balser; Balser & Grell IP Law

(57) **ABSTRACT**

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support wire system supported by poles. The support wire system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but carbon and graphite are preferred. Diodes may be used to restrict the backflow or loss of energy.

20 Claims, 11 Drawing Sheets

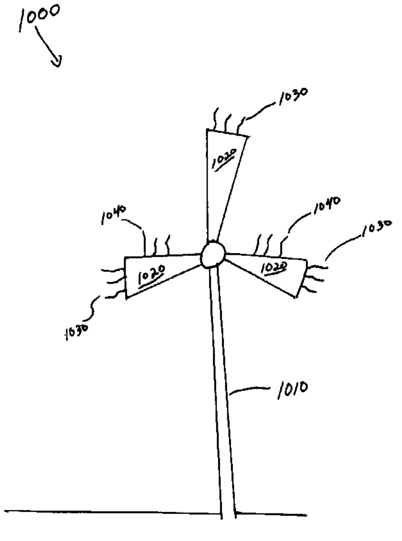

(56) References Cited

U.S. PATENT DOCUMENTS

4,314,192	A	2/1982	Caro	322/2 R
4,346,478	A	8/1982	Sichling	398/107
4,425,905	A	1/1984	Mori	126/578
4,433,248	A	2/1984	Marks	290/44
RE31,678	E	9/1984	Ochiai	385/115
4,483,311	A	11/1984	Whitaker	126/602
4,489,269	A	12/1984	Edling et al.	322/2 R
4,512,335	A	4/1985	Mori	126/678
4,653,223	A	3/1987	Mori	47/17
4,653,472	A	3/1987	Mori	126/683
4,676,226	A	6/1987	Mori	126/680
4,809,675	A	3/1989	Mori	126/578
4,852,454	A	8/1989	Batchelder	89/1.11
4,943,125	A	7/1990	Laundre' et al.	385/35
4,988,159	A	1/1991	Turner et al.	385/33
5,047,892	A	9/1991	Sakata et al.	361/231
5,114,101	A	5/1992	Stern et al.	244/172.8
5,145,257	A	9/1992	Bryant et al.	374/131
5,379,103	A	1/1995	Zigler	356/73
5,851,309	A	12/1998	Kousa	136/248
5,942,806	A	8/1999	Veliadis	290/1 R
6,025,591	A	2/2000	Taylor et al.	250/292
6,038,363	A	3/2000	Slater et al.	385/147
6,116,544	A	9/2000	Forward et al.	244/158.2
6,173,922	B1	1/2001	Hoyt et al.	244/158.2
6,226,440	B1	5/2001	Lyons	385/119
6,419,191	B1	7/2002	Hoyt et al.	244/158.2
6,425,391	B1	7/2002	Davoren et al.	126/683
6,457,943	B1 *	10/2002	Olsen et al.	416/230
6,735,830	B1	5/2004	Merciel	250/324
6,765,212	B2	7/2004	Faus et al.	250/339.11
6,846,447	B2	1/2005	Mbachu et al.	264/406
6,853,447	B2	2/2005	Goetz	356/237.1
6,894,772	B2	5/2005	Goetz et al.	356/237.1
6,925,852	B2	8/2005	Susko	73/23.2
6,974,110	B2	12/2005	Grandics	244/171.5
7,109,597	B1	9/2006	Bose	290/43
7,128,867	B2	10/2006	Mbachu et al.	264/406
7,231,809	B2	6/2007	Susko	73/31.05
7,259,475	B2	8/2007	Hong et al.	307/64
7,303,373	B2 *	12/2007	Viertl	416/1
7,439,712	B2 *	10/2008	McCowen	322/2 A
7,449,668	B2	11/2008	Schutten et al.	250/214 SW
7,478,712	B2 *	1/2009	McCowen	191/45 R
7,532,819	B1	5/2009	Triebes et al.	398/121
7,637,462	B2 *	12/2009	Pal	244/199.2
8,115,333	B2 *	2/2012	Parsche et al.	290/55
8,118,559	B2 *	2/2012	Llorente Gonzalez et al.	416/226
8,342,805	B2 *	1/2013	Mendez Hernandez et al.	416/146 R
8,519,596	B1 *	8/2013	Kim et al.	310/310
8,629,570	B1 *	1/2014	Kamenov	290/44
8,662,853	B2 *	3/2014	Vasudeva et al.	416/226
8,686,575	B2 *	4/2014	McCowen	290/1 R
2002/0026933	A1	3/2002	Gottlieb	126/628
2002/0108892	A1	8/2002	Goetz et al.	209/576
2002/0109094	A1	8/2002	Goetz et al.	250/339.11
2002/0109835	A1	8/2002	Goetz	356/237.1
2003/0125630	A1	7/2003	Furnish	600/461
2003/0193319	A1	10/2003	Wood et al.	322/2 A
2004/0011925	A1	1/2004	Grandics	244/62
2004/0083793	A1	5/2004	Susko	73/31.05
2004/0094853	A1	5/2004	Mbachu et al.	264/40.1
2004/0160711	A1	8/2004	Stumberger	361/20
2004/0212945	A1	10/2004	Sprenger et al.	361/212
2005/0101024	A1	5/2005	Mbachu et al.	436/85
2005/0136311	A1	6/2005	Ueda et al.	700/297
2005/0270525	A1	12/2005	Susko	356/326
2006/0051207	A1 *	3/2006	Becerra et al.	416/5
2007/0098551	A1 *	5/2007	Viertl	416/61
2007/0107765	A1	5/2007	Schutten et al.	136/243
2007/0195481	A1 *	8/2007	McCowen	361/233
2007/0217918	A1 *	9/2007	Baker et al.	416/227 R
2007/0218323	A1	9/2007	Sudo et al.	429/413
2007/0223171	A1	9/2007	Guy et al.	361/126
2007/0252047	A1 *	11/2007	Pal	244/35 R
2007/0273206	A1 *	11/2007	McCowen	307/9.1
2007/0297892	A1	12/2007	Kildegaard	415/1
2008/0145714	A1	6/2008	Kagami	429/413
2009/0040680	A1	2/2009	McCowen	361/233
2009/0107842	A1	4/2009	Park et al.	204/554
2009/0114495	A1	5/2009	McCowen	191/45 R
2009/0216292	A1	8/2009	Pless et al.	607/33
2010/0090562	A1	4/2010	McCowen	310/309
2010/0329881	A1 *	12/2010	Mendez Hernandez et al.	416/230
2012/0134826	A1 *	5/2012	Arocena De La Rua et al.	416/146 R
2012/0312918	A1 *	12/2012	Heppe	244/30
2013/0028739	A1 *	1/2013	Erichsen et al.	416/146 R
2013/0032671	A1 *	2/2013	Giles et al.	244/134 R
2013/0037650	A1 *	2/2013	Heppe	244/2
2013/0136598	A1 *	5/2013	Loewe et al.	416/95
2013/0170992	A1 *	7/2013	Loewe et al.	416/95
2013/0334824	A1 *	12/2013	Freda	290/55
2014/0084748	A1 *	3/2014	Wang et al.	310/300
2014/0086748	A1 *	3/2014	Peltola et al.	416/95
2014/0099208	A1 *	4/2014	Janiuk	416/210 R

* cited by examiner

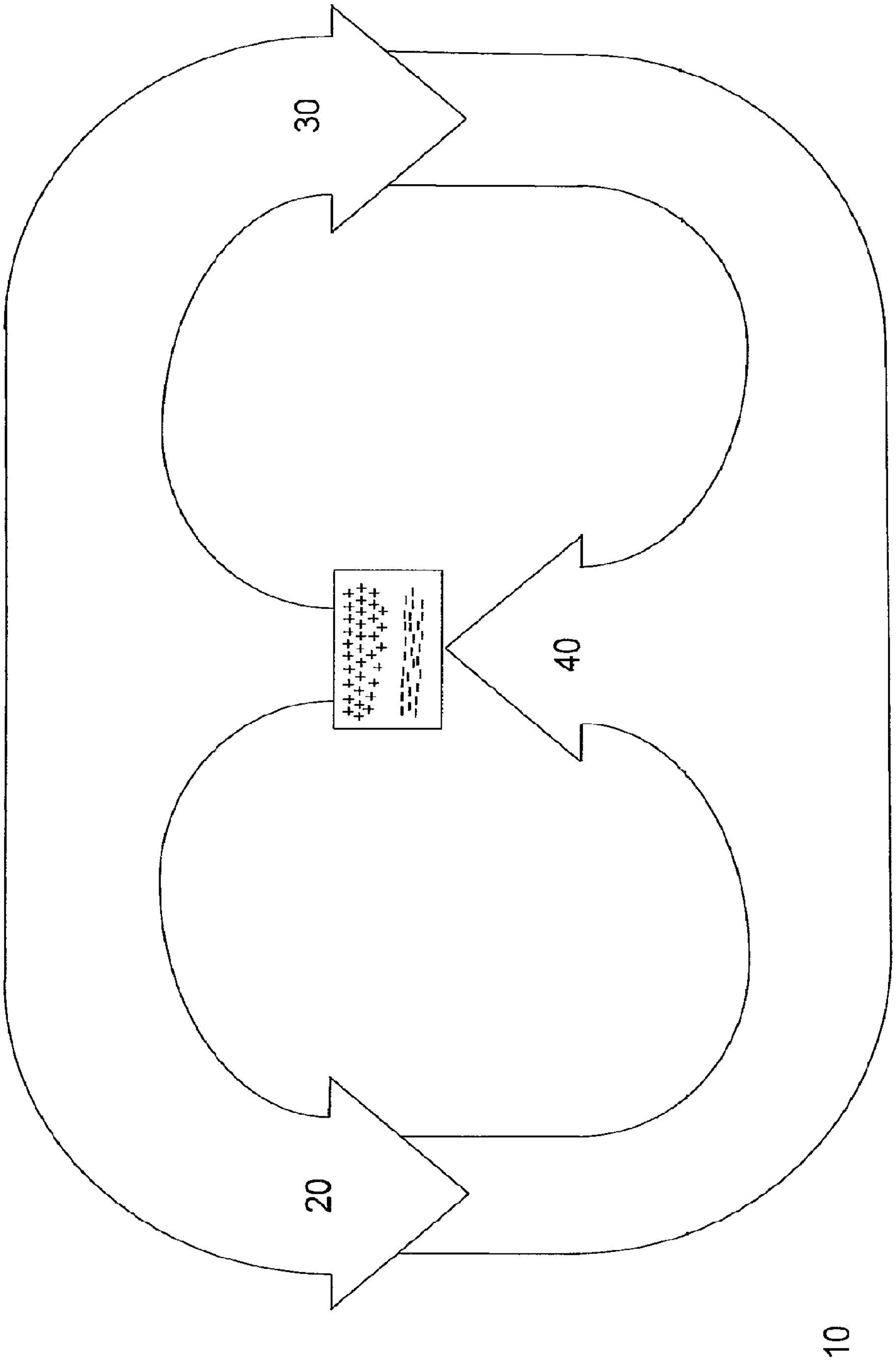

FIGURE 1

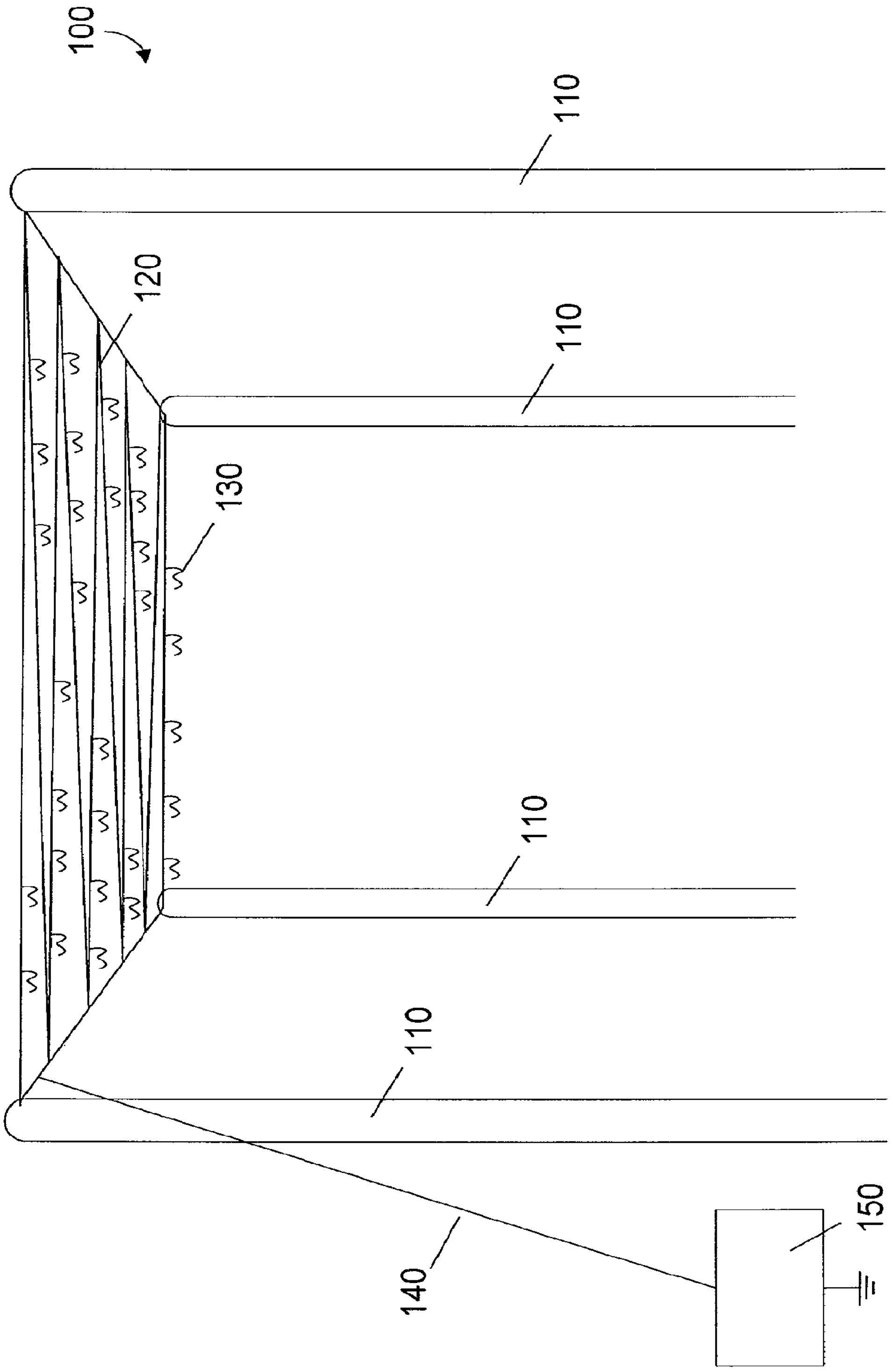
100
110
110
110
110
120
130
140
150

FIGURE 2

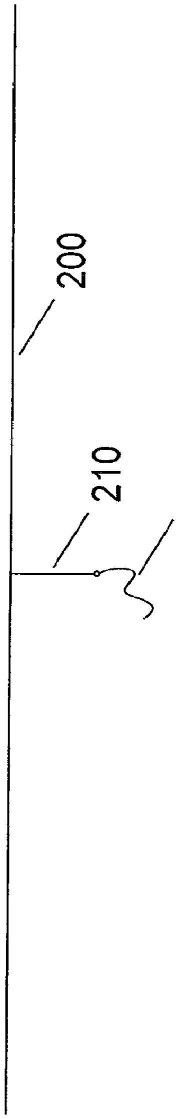

FIGURE 2A

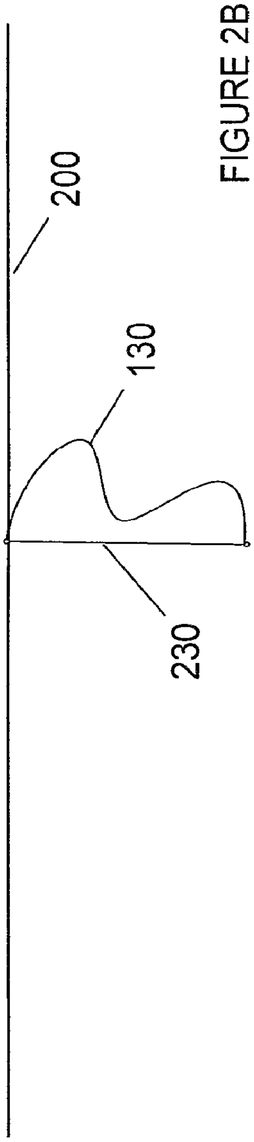

FIGURE 2B

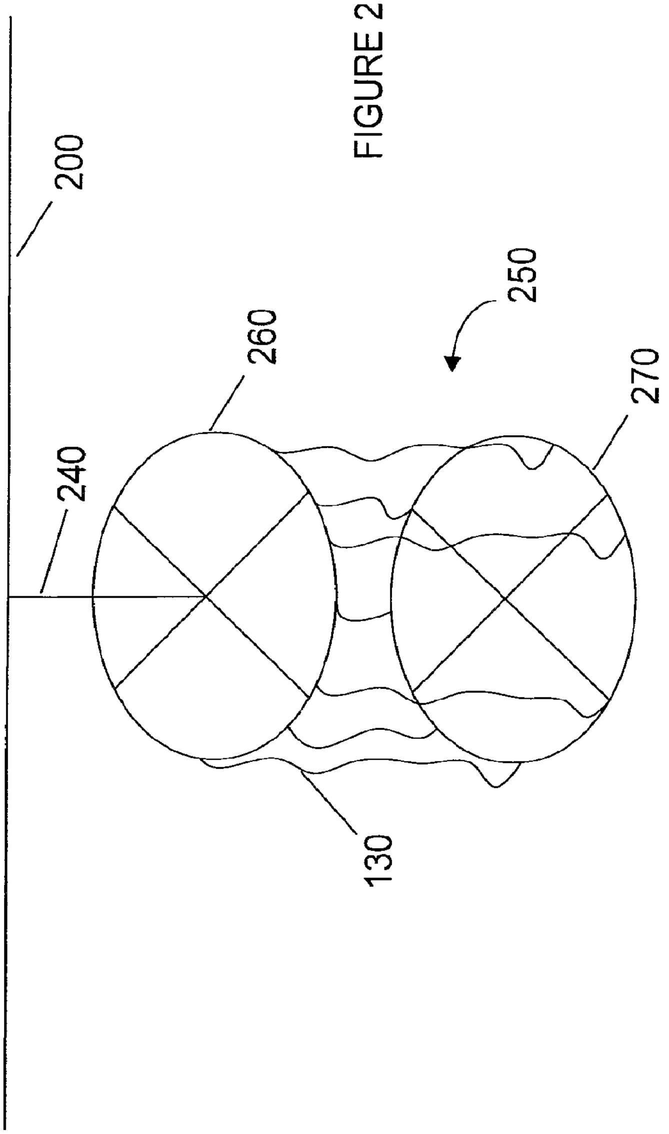
200
250
260
270
240
130

FIGURE 2C

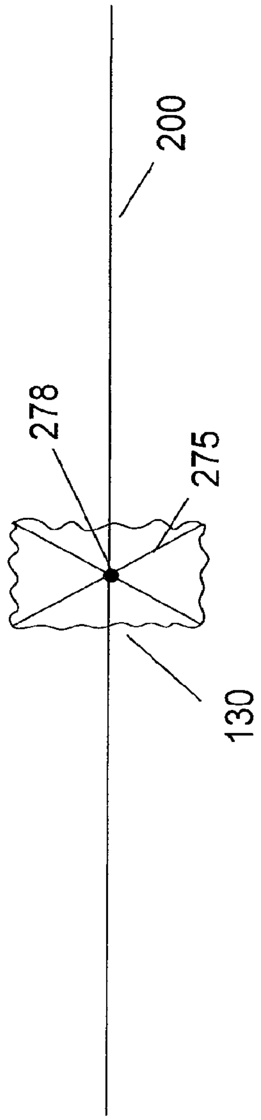
200
278
275
130

FIGURE 2D

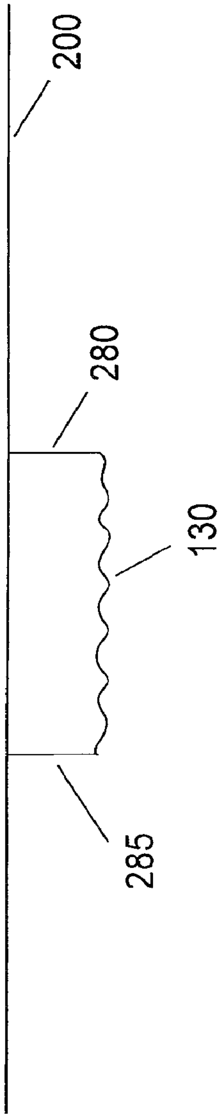
200
280
130
285

FIGURE 2E

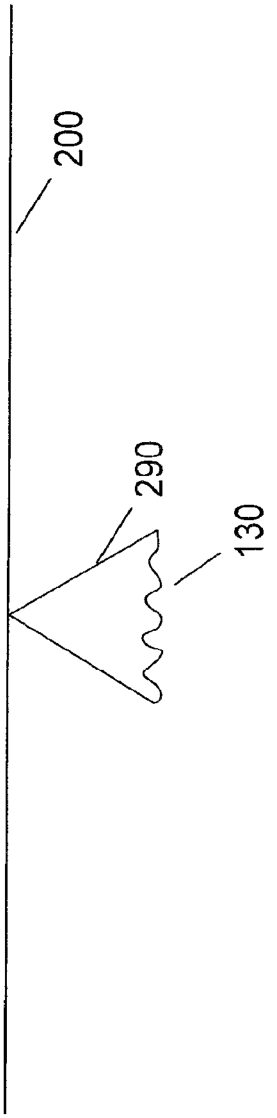

FIGURE 2F

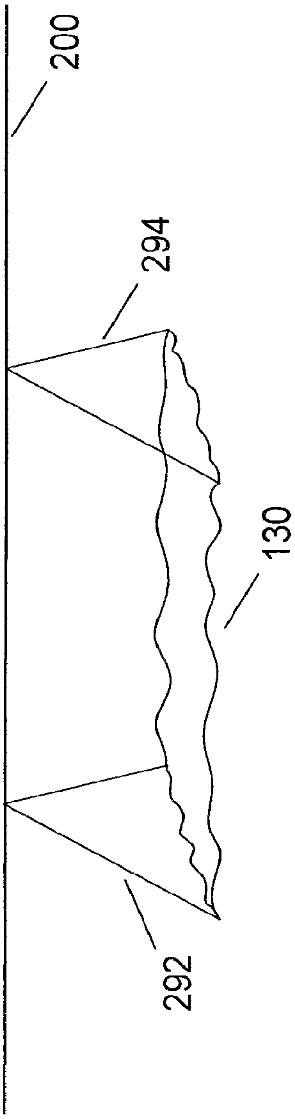
200
294
130
292

FIGURE 2G

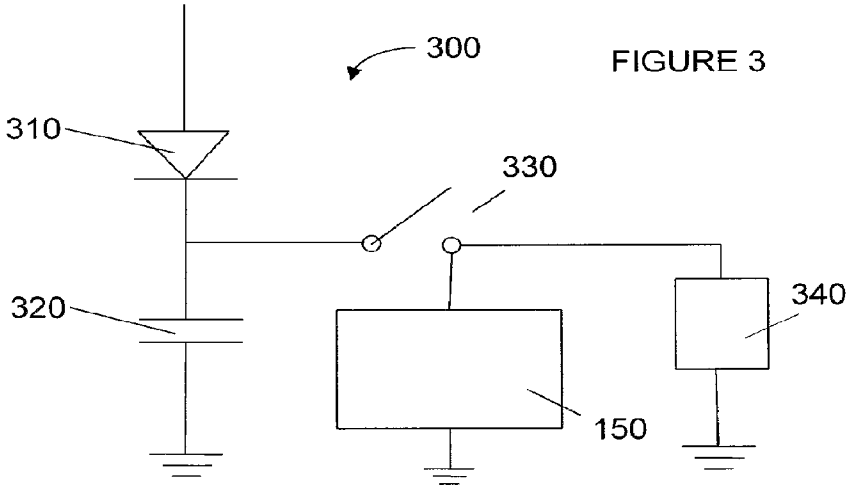
300
310
330
320
340
150

FIGURE 3

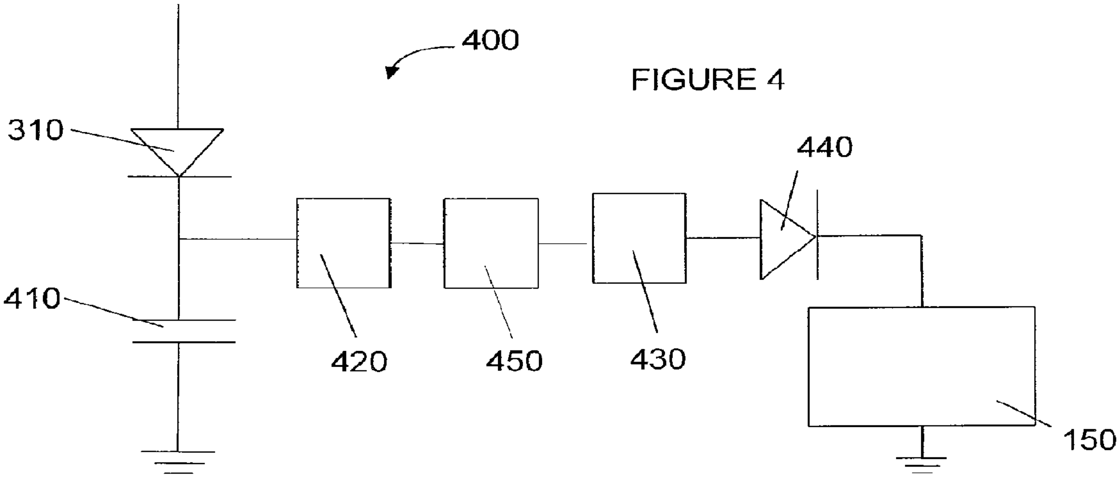
400
310
440
410
420
450
430
150

FIGURE 4

FIGURE 5

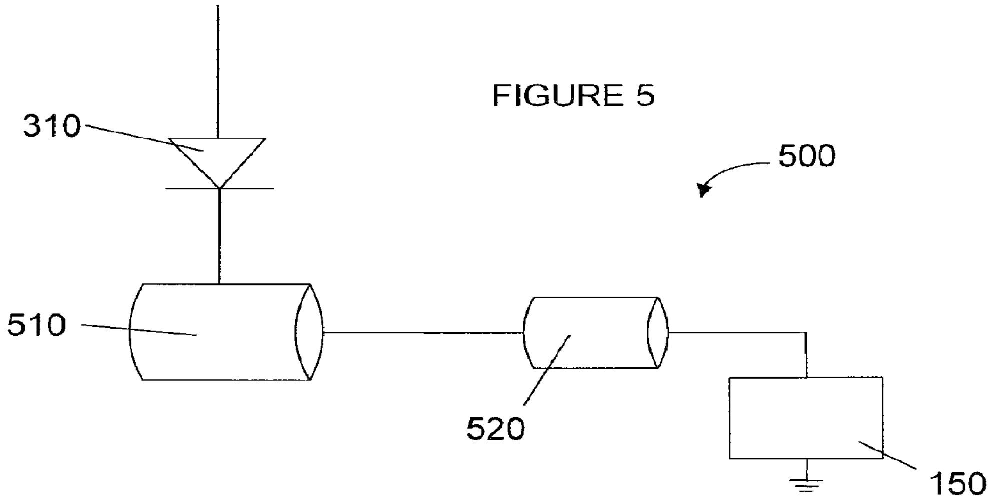
310
500
510
520
150

FIGURE 6

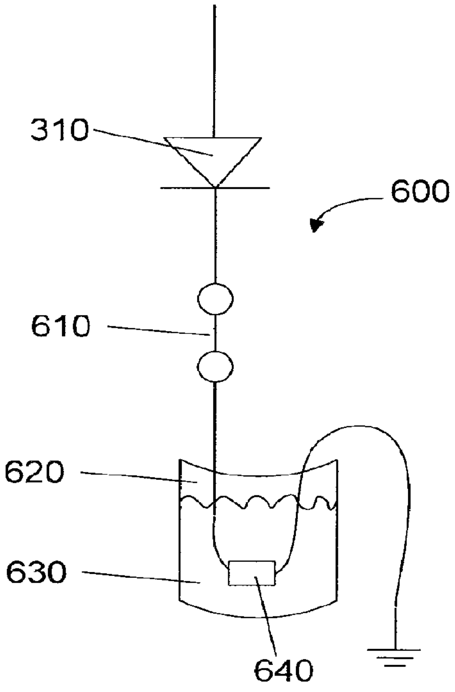
310
600
610
620
630
640

FIGURE 7

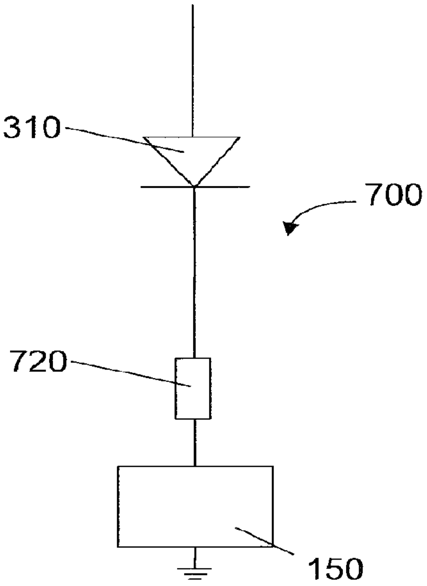
310
700
720
150

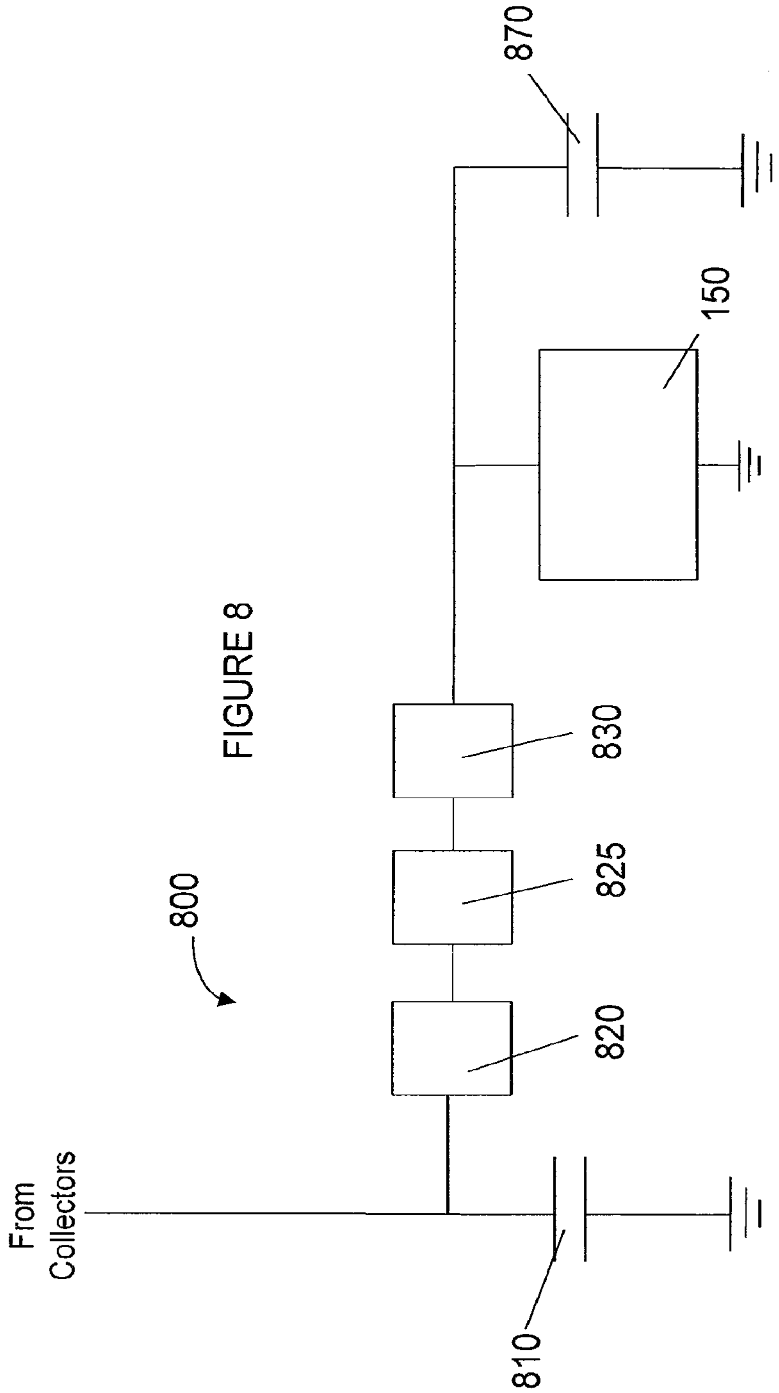
From Collectors
810
820
825
830
800
150
870

FIGURE 8

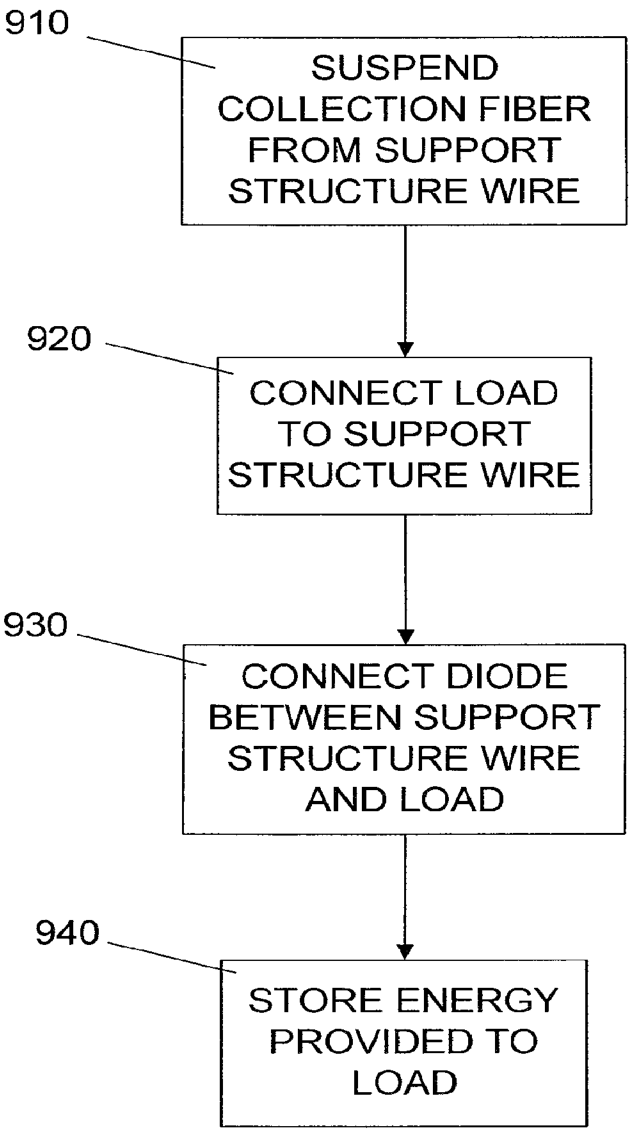
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

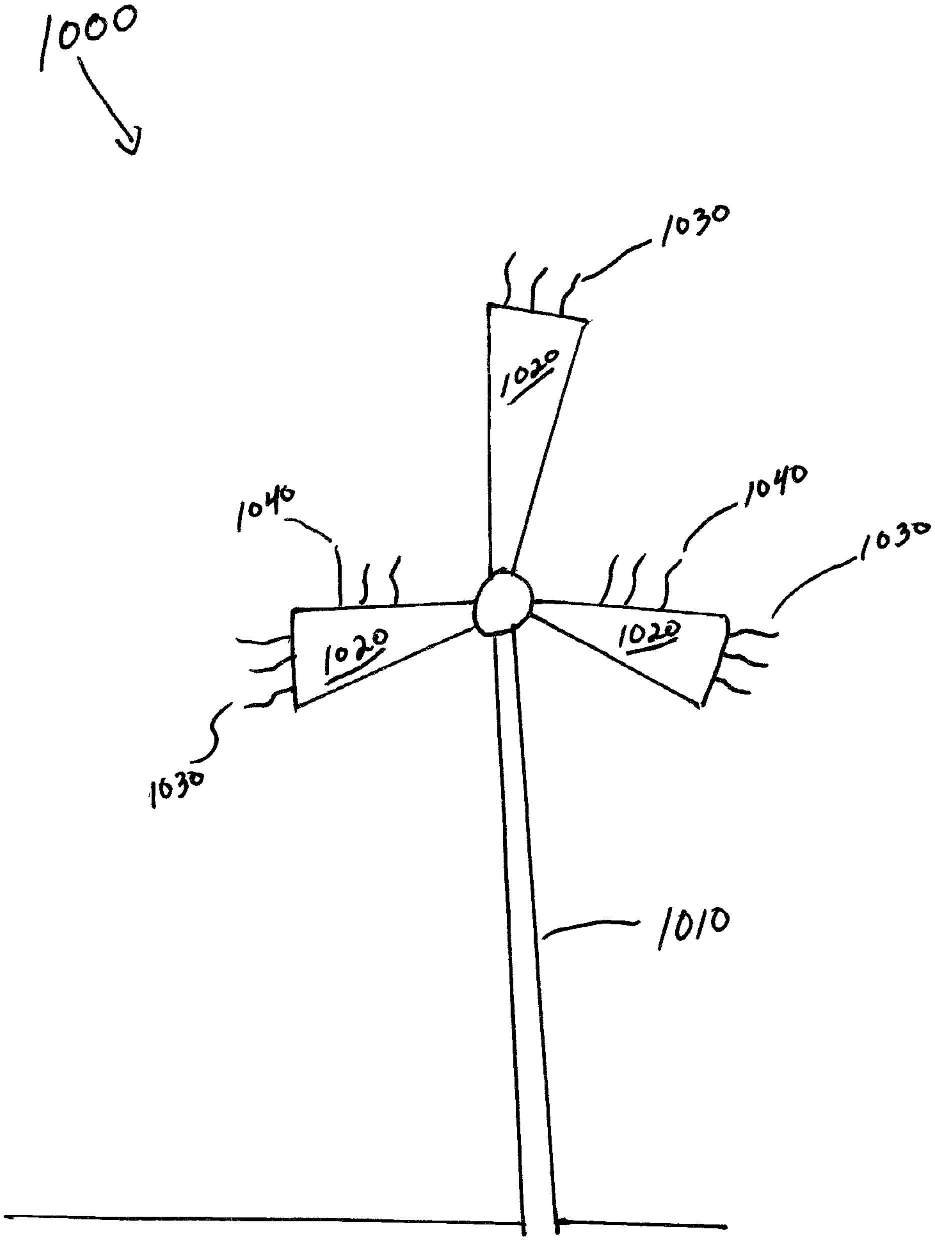
1000
1030
1020
1040
1040
1030
1020
1020
1030
1010

FIG. 10

ENERGY COLLECTION

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation in part application of U.S. patent application Ser. No. 12/255,130 filed on Oct. 21, 2008, which is a continuation application of U.S. patent application Ser. No. 11/358,264, filed on Feb. 21, 2006, which are both incorporated by reference herein.

TECHNICAL FIELD

The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure wire elevated above a ground level, at least one collection fiber electrically connected to the support structure wire; a load electrically connected to the support structure wire; and a diode electrically connected between the load and at least one collection fiber.

Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection fiber from a support structure wire elevated above ground level, the fiber electrically connected to the support structure wire; providing a load with an electrical connection to the support structure wire to draw current; and providing a diode electrically connected between the collection fiber and the load.

Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. **1** is a circuit diagram of a weather energy circuit.

FIG. **2** is a perspective view of an exemplary embodiment of many energy collectors elevated above ground by a structure.

FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

FIG. **2**B is a side view of an exemplary embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

FIG. **2**D is a side view of an exemplary embodiment of a support structure for multiple energy collection fibers.

FIG. **2**E is a side view of a support structure for an energy collection fiber.

FIG. **2**F is a side view of an exemplary embodiment of a support structure for an energy collection fiber.

FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

FIG. **3** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **4** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **5** is a circuit diagram of an exemplary embodiment of an energy collection circuit for powering a generator and motor.

FIG. **6** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

FIG. **7** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

FIG. **8** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy.

FIG. **9** is a flow diagram of an exemplary embodiment of collecting energy with a collection fiber.

FIG. **10** is a system diagram of an exemplary embodiment of a windmill with energy collectors.

DETAILED DESCRIPTION

Electric charges on conductors reside entirely on the external surface of the conductors, and tend to concentrate more around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An exemplary embodiment of this disclosure takes advantage of this property, among others, to collect and use the energy generated by an electric field in the atmosphere. Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An exemplary embodiment utilizes carbon or graphite fibers for static electricity collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including aluminum or steel, but most notably, copper. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

An exemplary embodiment of the collection fibers as collection device **130** includes graphite or carbon fibers. Graphite and carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, or graphite fibers are examples that demonstrate exemplary collection ability.

In at least one exemplary embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than 100 volts per meter under some conditions. When support wire **120** is suspended in the air at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An exemplary embodiment of the collection device comprises the diode with no collection fiber. A preferred embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restricts energy that may be collected through one fiber from escaping through another fiber.

Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2**A-**2**G using a collection fiber embodiment. FIG. **2**A presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**.

Likewise, FIG. **2**B shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

FIG. **2**C presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

FIG. **2**D presents another exemplary embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2**E provides another exemplary embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

FIG. **2**F presents another exemplary embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2**G provides two supports as provided in FIG. **2**F such that at least two support members **292**, **294** may be connected to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and between multiple support structures. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**.

FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load. Capacitor **320** maybe connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

In an exemplary embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel. Non-limiting examples of spark gap **420** include mercury-reed switches and mercury-wetted reed switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground. The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes maybe placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches and mercury-wetted reed switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel.

FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit. Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

FIG. **8** presents exemplary circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches and mercury-wetted reed switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

A windmill is an engine powered by the energy of wind to produce alternative forms of energy. They may, for example, be implemented as small tower mounted wind engines used to pump water on farms. The modern wind power machines used for generating electricity are more properly called wind turbines. Common applications of windmills are grain milling, water pumping, threshing, and saw mills. Over the ages, windmills have evolved into more sophisticated and efficient wind-powered water pumps and electric power generators. In an example embodiment, as provided in FIG. **10**, windmill tower **1000** of suitable height and/or propeller **1020** of windmill tower **1000** may be equipped with energy collecting fibers **1030**, **1040**. Collecting fibers **1030**, **1040** may turn windmill **1000** into a power producing asset even when there is not enough wind to turn propellers **1020**. During periods when there is enough wind to turn propellers **1020**, collecting fibers **1030**, **1040** may supplement/boost the amount of energy the windmill produces.

Windmill **1000**, properly equipped with ion collectors **1030**, **1040**, such as non-limiting example carbon fibers, can produce electricity: 1) by virtue of providing altitude to the carbon fiber to harvest ions, and 2) while the propeller is turning, by virtue of wind blowing over the carbon fiber producing electricity, among other reasons, via the triboelectric effect (however, it is also possible for the triboelectric effect to occur, producing electricity, in winds too weak to turn the propeller).

There are at least two ways that energy collectors may be employed on or in a windmill propeller to harvest energy. Propellers **1020** may be equipped with energy collectors **1030**, **1040** attached to, or supported by, propeller **1020** with wires (or metal embedded in, or on propeller **1020**) electrically connecting energy collectors **1030**, **1040** to a load or power conversion circuit. There may be a requirement to electrically isolate energy collectors **1030**, **1040**, which is added to propeller **1020**, from electrical ground, so that the energy collected does not short to ground through propeller **1020** itself or through support tower **1010**, but rather is conveyed to the load or power conversion circuit. Energy collectors may be connected to the end of propellers **1020** such as collectors **1030**. Alternatively, energy collectors may be connected to the sides of propellers **1020** such as collectors **1040**.

Alternatively, propeller **1020** may be constructed of carbon fiber or other suitable material, with wires (or the structural metal supporting propeller **1020** may be used) electrically connecting to a load or power conversion circuit. In the case of propeller **1020** itself being constructed of carbon fiber, for example, the fiber may be 'rough finished' in selected areas so that the fiber is "fuzzy." For example, small portions of it may protrude into the air as a means of enhancing collection efficiency. The fuzzy parts of collectors **1030**, **1040** may do much of the collecting. There may be a requirement to electrically isolate carbon fiber propeller **1020** from electrical ground, so that the energy it collects does not short to ground through metal support tower **1010**, but rather is conveyed to the load or power conversion circuit. Diodes may be implemented within the circuit to prevent the backflow of energy, although diodes may not be necessary in some applications.

In an alternative embodiment, windmill **1000** may be used as a base on which to secure an even higher extension tower to support the energy collectors and/or horizontal supports extending out from tower **1010** to support the energy collectors. Electrical energy may be generated via ion collection due to altitude and also when a breeze or wind blows over the collectors supported by tower **1010**.

Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

It should be emphasized that the above-described embodiments of the present disclosure, particularly, any "preferred" embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A method of collecting energy comprising:

suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment from a windmill, the at least one collection device electrically connected to the windmill; and

providing a load with an electrical connection to the at least one collection device to draw current.

2. The method of claim **1**, wherein the collection device collects energy by triboelectric effect.

3. The method of claim **1**, wherein the collection device comprises a collection fiber.

4. The method of claim **1**, wherein the collection device comprises a diode and a collection fiber and the diode is electrically connected between the collection fiber and the load.

5. The method of claim **1**, further comprising storing energy provided to the load.

6. The method of claim **5**, wherein storing energy provided to the load comprises storing energy in a capacitor or an inductor.

7. The method of claim **3**, wherein the collection fiber comprises carbon fiber or graphite fiber.

8. A system of energy collection comprising:

a windmill;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment electrically connected to the windmill; and

a load electrically connected to the at least one collection device.

9. The system of claim **8**, wherein the collection device collects energy by triboelectric effect.

10. The system of claim **8**, wherein the collection device comprises a collection fiber.

11. The system of claim **8**, wherein the collection device comprises a collection fiber and a diode electrically connected between the load and the collection fiber.

12. The system of claim **9**, wherein the diode is elevated relative to the ground level.

13. The system of claim **10**, wherein the collection fiber comprises a carbon fiber or a graphite fiber.

14. The system of claim **8**, further comprising a diode electrically connected between the at least one collection device and the support structure.

15. The system of claim **8**, further comprising: a switch connected in series between the at least one collection device and the load; and a capacitor connected in parallel with the switch and the load.

16. The system of claim **15**, wherein the switch comprises an interrupter connected between the plurality at least one collection device, and wherein the interrupter comprises at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

17. The system of claim **8**, further comprising: a motor for providing power, the motor connected between the at least one collection device and the load; and a generator powered by the motor.

18. The system of claim **8**, further comprising a fuel cell between the support structure and the load.

19. The system of claim **18**, wherein the fuel cell produces hydrogen and oxygen.

20. A system of collecting energy comprising:

means for suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment from a windmill structure, the at least one collection device electrically connected to the means for suspending;

means for inducing current flow, the means for inducing current flow electrically connected to the means for suspending; and

means for restricting the backflow of charge carriers, the means for restricting the backflow of charge carriers electrically connected between the at least one collection device and the means for inducing current flow.

* * * * *

FIP8102 XR 3,273,066

X5264R

Sept. 13, 1966 L. H. RUHNKE SD 3,273,066

APPARATUS FOR DETECTING CHANGES IN THE ATMOSPHERIC ELECTRIC FIELD

Filed Dec. 20, 1963

2 Sheets-Sheet 1

Fig. 1

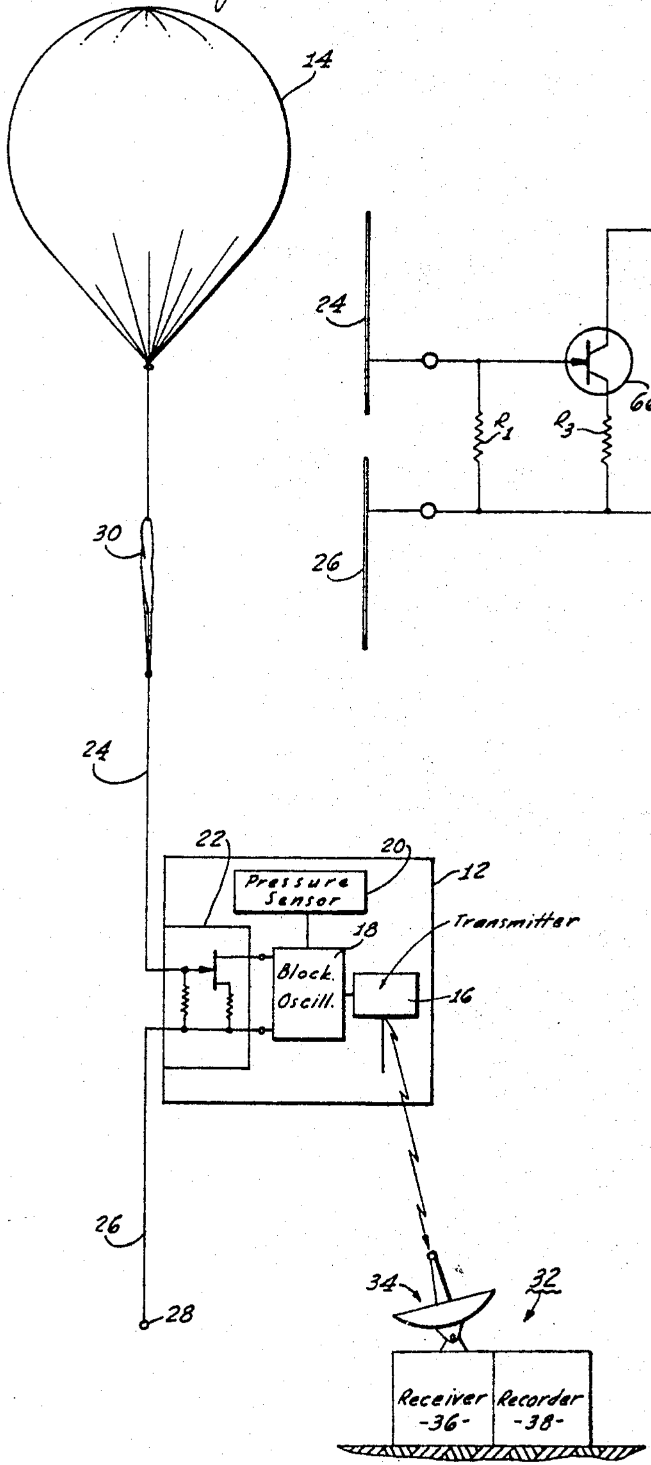
14
30
24
22
20
Pressure Sensor
12
Transmitter
18
Block. Oscill.
16
26
28
34
32
Receiver -36-
Recorder -38-

Fig. 3

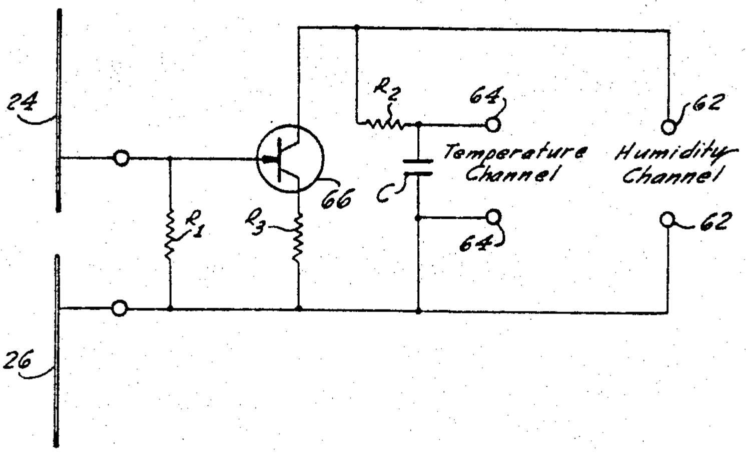
24
R_2
64
62
Temperature Channel
Humidity Channel
66
C
R_1
R_3
64
62
26

INVENTOR:
Lothar H. Ruhnke

By Alan C. Rose
Attorney

Sept. 13, 1966

L. H. RUHNKE

3,273,066

APPARATUS FOR DETECTING CHANGES IN
THE ATMOSPHERIC ELECTRIC FIELD

Filed Dec. 20, 1963

2 Sheets-Sheet 2

Fig. 2

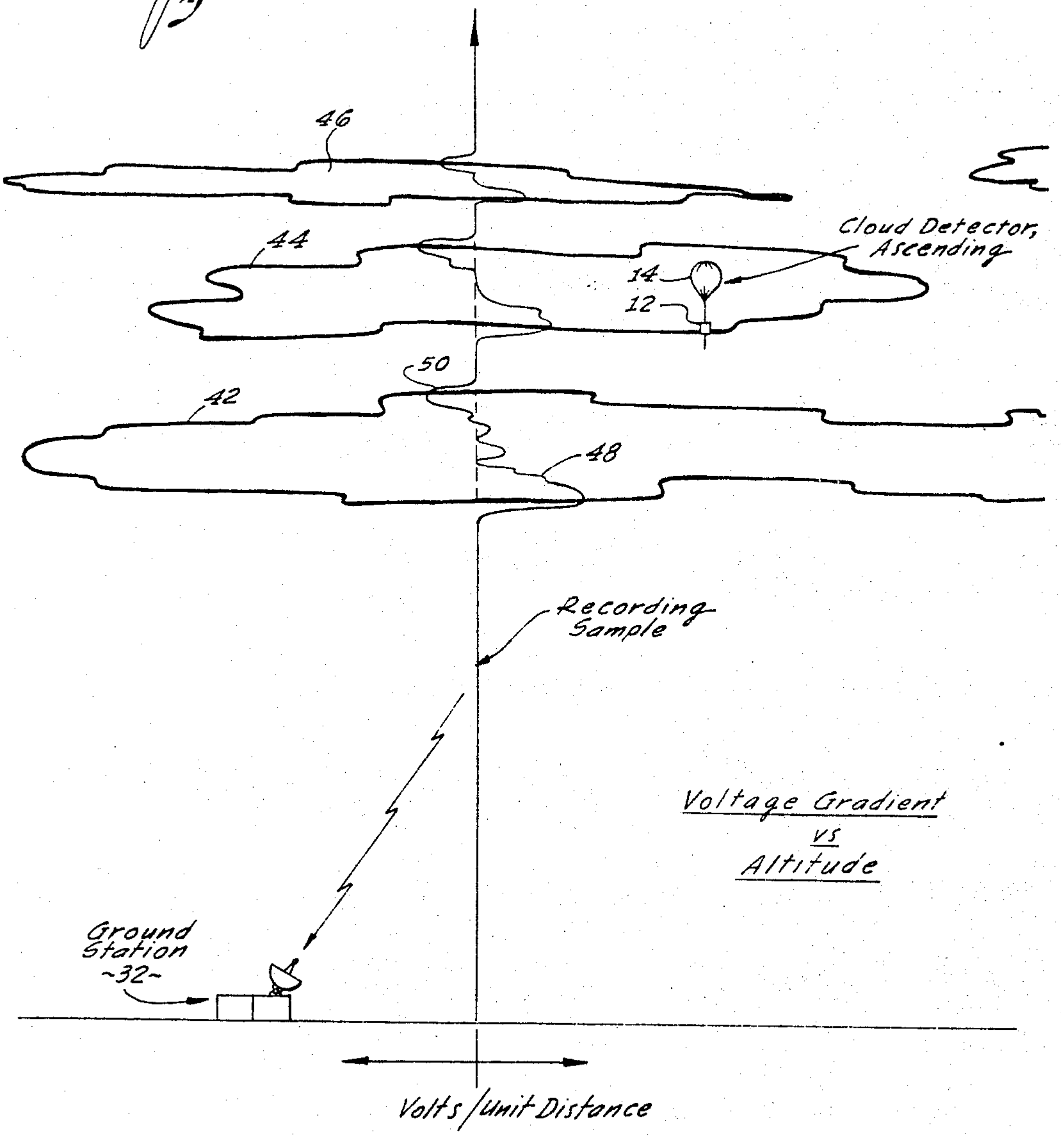
46
44
42
50
48
14
12
Cloud Detector,
Ascending
Recording
Sample
Voltage Gradient
vs
Altitude
Ground
Station
~32~
Volts/Unit Distance

INVENTOR:
Lothar H. Ruhnke

By Alan C. Rose
Attorney

United States Patent Office

3,273,066

Patented Sept. 13, 1966

3,273,066

APPARATUS FOR DETECTING CHANGES IN THE ATMOSPHERIC ELECTRIC FIELD

Lothar H. Ruhnke, Minneapolis, Minn., assignor to Litton Systems, Inc., Beverly Hills, Calif.

Filed Dec. 20, 1963, Ser. No. 332,186
4 Claims. (Cl. 325—113)

This invention relates to a cloud detector or an apparatus for sensing electrostatic conditions in the earth's atmosphere.

Cloud detection apparatus which is presently employed is usually of the "ceilometer" type which merely measures the height of the lower surface of the clouds to indicate the ceiling conditions. One object of the present invention is to determine the location of both the top and the bottom of one or more cloud formations of the atmosphere.

While electrostatic sensors of one type or another have been proposed heretofore for use in the atmosphere, they have been so complex and expensive that they have not enjoyed widespread use. In addition, cloud radars have recently been developed, but their usefulness is limited by failures when different zones of the atmosphere contain droplets of different sizes, as discussed more fully below. Accordingly, it is another object of the present invention to simplify and reduce the cost of cloud detectors, while maintaining reliability, compatibility with existing equipment, and adequate sensitivity for all practical purposes.

In accordance with an illustrative embodiment of the present invention, a conventional U.S. Weather Bureau or U.S. Army meteoroligical radiosonde transmitter may be provided with a modulator which is controlled by a field effect transistor. The input to the field effect transistor is a dipole antenna. Each of the arms of the dipole antenna may be five to fifty feet in length, for example, and one of the dipole elements may trail the radiosonde package while the other element may extend forward in the direction of movement of the unit. When a balloon is employed to carry the cloud detector unit aloft, a weight may be attached to one of the antenna elements so it hangs below the instrument package, while the other antenna element can extend along, or even form the flexible connection between the radiosonde unit and the balloon. A receiving unit is provided to pick up the modulated signals from the radiosonde apparatus, demodulate these signals and record them. While it is contemplated that the receiver may normally be situated on the ground and receive signals from a rising radiosonde unit, the radiosonde unit could well be employed by an aircraft to determine conditions below a cloud layer. Under these conditions the receiver would be mounted on the plane and the radiosonde unit could be parachuted to the ground.

In accordance with a feature of the present invention, therefore, a system for detecting electrostatic field conditions in the atmosphere includes a transmitter, arrangements for moving the transmitter through the atmosphere, an antenna extending a substantial distance from the transmitter, and a field effect transistor connected to receive electrostatic input signals from said antenna and to modulate the output of the transmitter in accordance with voltages detected by the antenna.

In accordance with one aspect of the present invention, simplicity of circuitry and stability of operation are achieved by the use of dynamic circuitry which measures the instantaneous difference in potential between two antenna elements. Thus, a high input resistor may be connected between the input antenna elements. With this type of arrangement the instrument package must be continuously moved through the atmosphere in order to

give meaningful results. Under these conditions, the recorded output indicates changes in electric fields. A constant electric field, corresponding to a linear change in potential produces no substantial output signal. It has been determined that a balloon velocity of about five meters per second will produce good results with a dipole antenna shunted by a 1,000 megohm resistor and connected to a field effect transistor as discussed in detail in the present specification. With lower velocities, a higher input impedance or greater amplification would be required, with higher velocities, such as rocket propulsion or the like, permitting even lower input impedances and simpler input circuits.

Accordingly, it is a feature of the invention that apparatus is provided for moving a radiosonde unit through the atmosphere at a predetermined rate, that two element input antenna preferably of the dipole type is provided, and that the antenna is connected to the radiosonde by simple dynamic amplification circuitry, which requires continuous movement of the apparatus to give useful output information.

The novel features which are believed to be characteristic of the invention both as to its organization and method of operation, together with further objects and advantages thereof, will be better understood from the following description considered in conjunction with the accompanying drawings in which a typical embodiment of the invention is illustrated by way of example. It is to be expressly understood, however, that the drawings are for the purpose of illustration and description only and are not intended as a definition of the limits of the invention.

In the drawing:

FIG. 1 is a schematic block diagram of a radiosonde equipment for sensing changes in the electric field of the atmosphere, and thus detecting clouds, in accordance with the present invention;

FIG. 2 is a diagrammatic showing of the record obtained under typical cloud conditions; and

FIG. 3 is a circuit diagram indicating a typical arrangement of connections of a dipole antenna, a field effect transistor, and the input terminals of a standard meteorological radiosonde apparatus.

With reference to the drawings, FIG. 1 shows a radiosonde package **12** supported by a balloon **14**. The radiosonde unit includes a transmitter **16**, blocking oscillator **18**, pressure sensor **20**, and a field effect transistor unit **22**. A dipole antenna having an upper element **24** and a lower element **26** is connected to the input of the field effect transistor unit **22**. A small weight **28** is secured to the lower end of the antenna element **26** to hold it in the extended position. Similarly, the upper antenna element **24** extends in the other direction, and it may either form the support for the radiosonde unit **12** or may be closely associated with the supporting cable. A parachute **30** is provided for safe return of the instrument package to ground.

A ground or receiving station **32** may include an antenna **34**, receiver and demodulator **36**, and a recording apparatus **38**.

FIG. 2 shows a typical recording of the electrostatic field obtained as the ballon **14** and instrument package **12** rise through the successive layers of clouds **42**, **44**, and **46**, along the center line of the drawing. It is known that the electrostatic potential in the atmosphere increases from ground potential to plus 300,000 volts at elevations of about 30 miles. It is also known that the atmospheric electric field, which is the gradient of the potential, is substantially increased within clouds. As the package enters the lower edge of a cloud, the sudden increase in electric field causes a pulse of current to flow through resistor R_1 in one direction, and upon leaving the cloud, the sudden reduction in electric field causes an oppositely directed current pulse. This produces the positive swing shown at **48** in FIG. 2 as the unit enters the cloud and the negative peak **50** as the apparatus is rising through the upper surface of the cloud stratus **42**. A similar pattern is observed as the unit goes through the higher cloud formations **44** and **46**. With a knowledge of the speed of ascent of the balloon and instrument package, the location of the upper and lower surfaces of successive cloud layers may be determined by the present apparatus. Accordingly, despite the presence of a relatively low cloud layer which completely obscures other cloud formations, ground personnel may readily determine the meterological conditions relating to clouds above this low ceiling.

Returning to FIG. 1, the radiosonde equipment represented by blocks **16**, **18** and **20** is well known. The meterological radiosonde system which will be taken as typical is designated by the military by the numbers AN/AMT 4B. The Army technical manual which describes this radiosonde system is identified as TM 11–2432A–TO 31M4–2AMT–11, dated June 26, 1958. The radiosonde transmitter operates at a frequency of 1680 megacycles per second. This transmitter is modulated by a blocking oscillator **18**, the frequency of the blocking oscillator being controlled by the meteorological conditions which are detected. Normally, a thermistor for measuring the temperature and a humidity sensor are alternately switched to the input terminals of the blocking oscillator **18**. In the present case, the cloud detector circuitry is connected to both of these terminals. The barometric pressure sensor **20** may also be retained in the unit, and it may be connected to control the frequency of the blocking oscillator **18** by a switching action as described below.

With reference to FIG. 3, the humidity channel input terminals **62** and the temperature channel input terminals **64** of the standard radiosonde unit are shown at the right-hand side of the circuit. Atmospheric electric signals received by the dipole antenna elements **24** and **26** are applied to the field effect transistor **66** and are then coupled by a suitable circuit to the input terminals **62** and **64** of the radiosonde. While other field effect transistors may be employed, type C 610 of Crystallonics, Inc., was found to be suitable from a sensitivity and reliability standpoint for the present purposes. The input to the field effect transistor is shunted by a 1,000 megohm resistor R_1 to stabilize the input characteristic. Suitable biasing of the transistor **66** is provided by resistor R_3. Because the impedance of the transistor **66** is substantially lower than that of the temperature and humidity sensors which are normally employed, a 110,000 ohm resistor in the blocking oscillator was removed and a capacitor C was connected in parallel to the input terminals. The capacitor was selected to produce a modulation frequency of approximately 100 cycles per second in the absence of an input signal. Positive and negative voltages will then raise or lower this frequency.

In the radiosonde apparatus, signals from the pressure sensor **20** are normally employed to switch from the temperature channel terminals **64** to the humidity channel terminals **62**. This switch is well known in the art as a baroswitch. The barometric pressure switching action provides an indication of altitude, so it is useful to preserve this information. Accordingly, the resistor R_2 is employed between the transistor **66** and one terminal **64** to reduce the output signal slightly, thus changing the modulation level and preserving the pressure and thus altitude information.

With regard to the length of the dipole elements **24** and **26**, the size of resistor R_1 and the amplification of the field effect transistor, these are fixed by several factors. First, the antenna elements should be large enough to provide a substantial input signal, but should be relatively small as compared with normal cloud thicknesses. Lengths of 5 to 50 feet for each dipole would be of the proper order of magnitude. The resistor R_1 was chosen to reduce the input impedance between dipoles **24** and **26** below the point where fog and dampness and the resultant shunt impedance will introduce instability into the transmitted signals. With these parameters, and a speed of ascent of the order of 5 meters per second, it was determined that a single field effect transistor provided the necessary amplification for operation with the standard radiosonde unit.

As mentioned generally above, cloud radars are subject to failure under certain atmospheric conditions. Thus, the radar return is proportional to the sixth power of the droplet or particle size, while the human eye and optical systems respond to the second power of the droplet size. With the radar return proportional to such a high power of the droplet diameter, if a radar is adjusted to pick up a cloud having large particles, other clouds having small size particles are not detected. The present electrostatic system, however, responds in a manner which is nearly the same as the optical response, and therefore does not fail under atmospheric conditions where different size droplets or particles are present in successive clouds.

It is to be understood that the above described arrangements are illustrative of the application of the principles of the invention. Numerous other arrangements within the scope of the invention may be devised by those skilled in the art. Thus, by way of example and not of limitation, a somewhat different shape of antenna could be employed to pick up electrostatic field variations and apply them to the field effect transistor, an inherently stabilized transistor may be used, and the instrument package may be carried by rocket or other suitable propulsion arrangements than the balloon shown in the drawings. Accordingly, it is evident that various changes may be made in the present invention without departing from the spirit of the invention as defined in the present claims.

What is claimed is:

1. An apparatus for sensing electrostatic conditions in the atmosphere comprising:

a radiosonde apparatus including means for modulating the output frequency thereof, said means including an R-C circuit;

a pair of probes positioned in respectively different locations in the atmosphere with respect to said radiosonde;

a field effect transistor having its drain and gate terminals connected across said pair of probes and having its source and drain terminals connected in series with the resistance of said R-C circuit to modulate said radiosonde; and

a stabilizing high resistance connected across the two probes and across said gate and drain terminals of said transistor.

2. In an apparatus for sensing atmospheric electrostatic field conditions;

a balloon;

an instrument package connected to and spaced from said balloon, said instrument package including a transmitter, means including an R-C circuit for modulating said transmitter, and a field effect transistor having source and drain terminals connected in series with the resistance in said R-C circuit to modulate said transmitter;

antenna means including an antenna and resistance means connected across said antenna, said resistance means also being connected across the gate and the drain of said field effect transistor, said antenna extending upward from said package toward said balloon and downwardly from said package for applying signals to said field effect transistor in response to said field conditions to render said modulating means responsive to said field conditions.

3. An apparatus for sensing changes of the atmospheric electric field at cloud boundaries, which comprises:

a blocking oscillator-type modulated radiosonde including fixed resistance means and capacitance means connected in parallel with said fixed resistance means, said capacitance means being effective to discharge through said resistance means to control the modulation of said radiosonde;

dipole antenna means responsive to changes of the atmospheric electric field for producing a signal indicative of said changes;

resistance means connected across said dipole antenna means; and

a field effect transistor having source and drain terminals connected in series with said fixed resistance means and a gate terminal connected to said dipole antenna means for reception of said signal, said field effect transistor being responsive to said signal for varying the rate of discharge of said capacitance means to cause the blocking oscillator to modulate said radiosonde in accordance with the second changes in the atmospheric electric field.

4. An apparatus for sensing changes of the atmospheric electric field at cloud boundaries, which comprises:

radiosonde means including a blocking oscillator having fixed resistance means and capacitance means connected in parallel with said fixed resistance means, said capacitance means being effective to discharge through said resistance means to control the frequency of oscillation of said blocking oscillator;

dipole antenna means responsive to changes of the atmospheric electric field for producing a signal indicative of said changes;

resistance means connected across said dipole antenna; and

a field effect transistor having source and drain terminals connected in series with said fixed resistance means and a gate terminal connected to said dipole antenna means for reception of said signal, said field effect transistor being responsive to said signal for varying the rate of discharge of said capacitance means to control the frequency of oscillation of said blocking oscillator in accordance with the sensed changes in the atmospheric electric field.

References Cited by the Examiner

UNITED STATES PATENTS

2,463,527	3/1949	Dunmore	325—113
3,038,154	6/1962	Zworykin et al.	325—113 XR
3,056,888	10/1962	Atalla	307—88
3,062,043	11/1962	Marsh et al.	340—193.2

DAVID G. REDINBAUGH, *Primary Examiner.*

JOHN W. CALDWELL, *Examiner.*

US 20100007218A1

(19) **United States**
(12) **Patent Application Publication**
Ogram
(10) **Pub. No.: US 2010/0007218 A1**
(43) **Pub. Date: Jan. 14, 2010**

(54) **ATMOSPHERIC ELECTRICAL GENERATOR**

(76) Inventor: **Mark Ellery Ogram**, Tucson, AZ (US)

Correspondence Address:
Mark E. Ogram (30343)
Ste. 110-186, 6990 E. 22nd Street
Tucson, AZ 85710-5192 (US)

(21) Appl. No.: **12/218,297**

(22) Filed: **Jul. 14, 2008**

Publication Classification

(51) **Int. Cl.**
H02G 11/00 (2006.01)
(52) **U.S. Cl.** **307/145**

(57) **ABSTRACT**

A mechanism to tap an electrical source which uses an aircraft (preferably a lighter than air balloon) tethered by a conductive line. The conductive line is extended/withdrawn by a winch motor to adjust the altitude of the aircraft. The conductive line is isolated from the ground and an electrical conductor is connected to the conductive line and to an electrical load. In this manner, static electricity generated in the atmosphere is gathered for use.

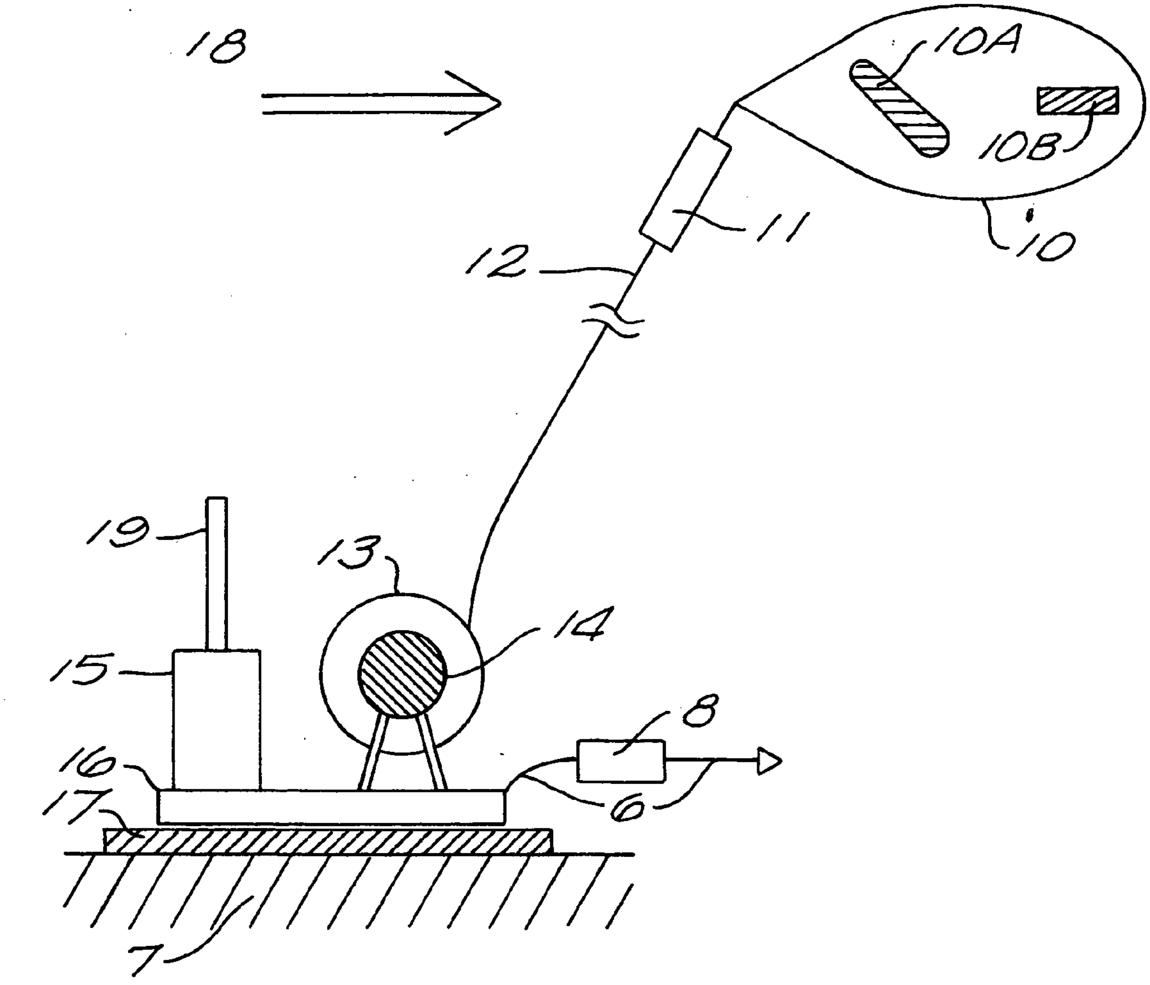

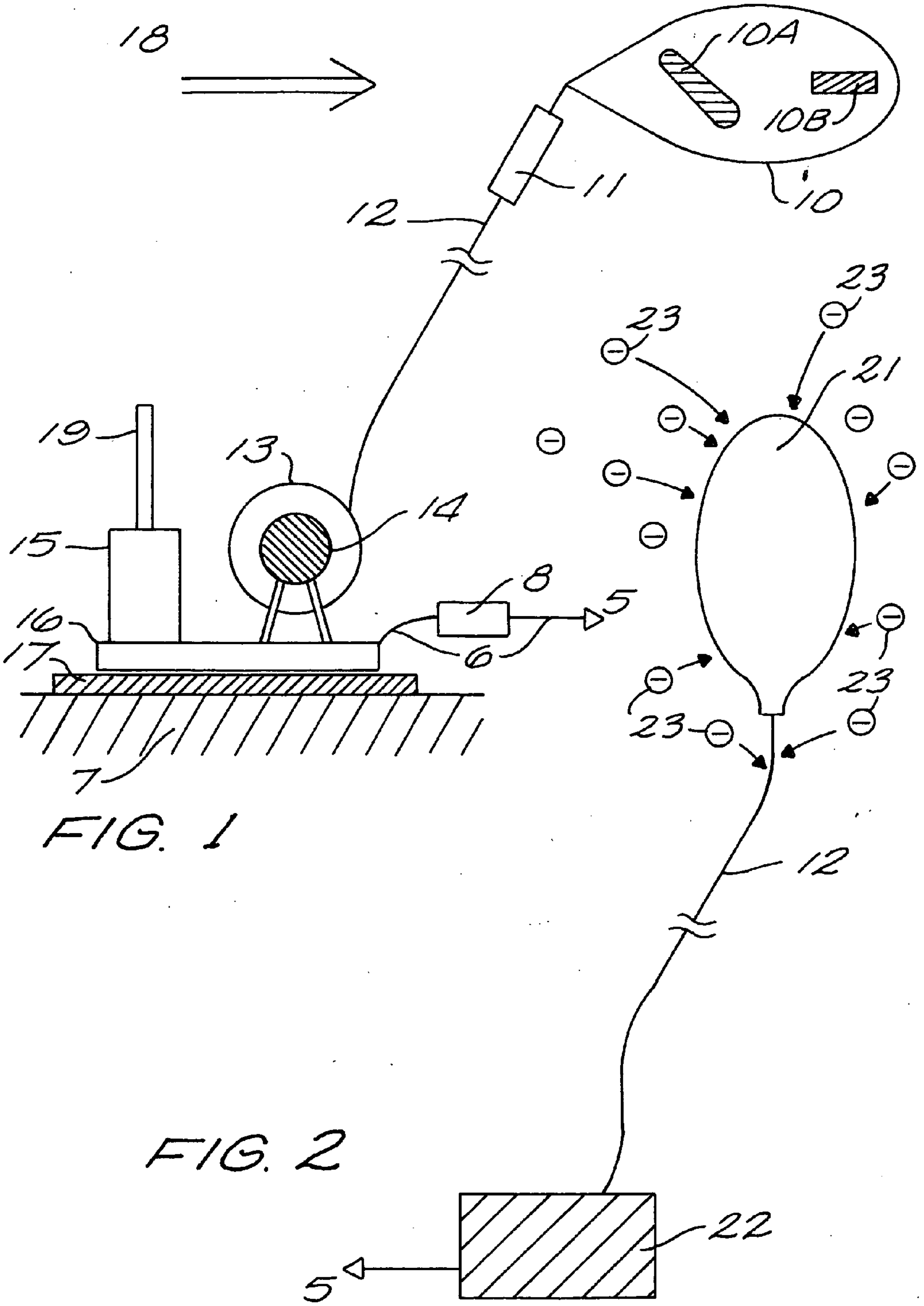
18
10A
10B
10
11
12
19
13
14
15
8
5
16
6
17
7
23
23
21
23
23
23
12
22
5

FIG. 1

FIG. 2

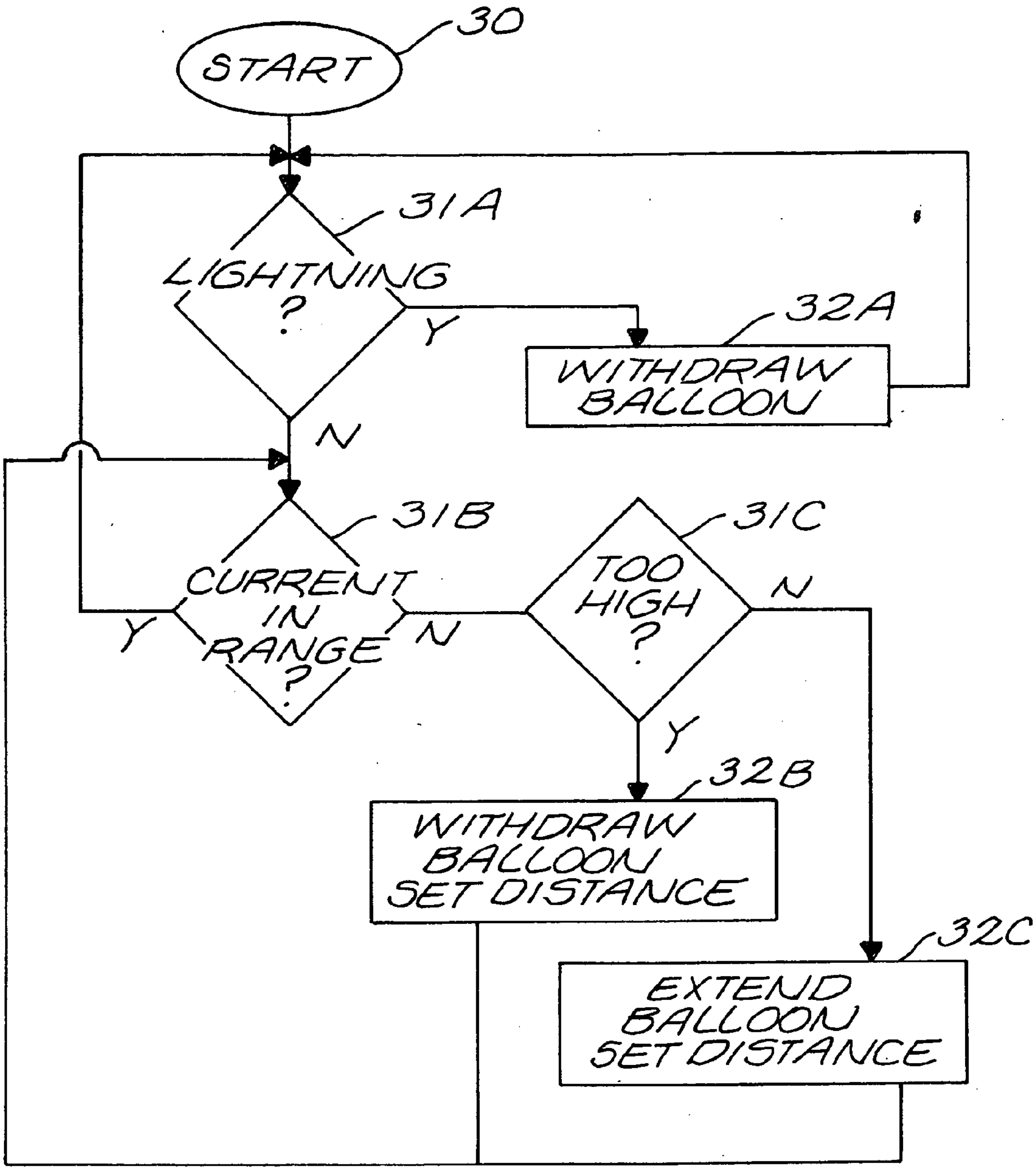
30
START
31A
LIGHTNING ?
Y
32A
WITHDRAW BALLOON
N
31B
CURRENT IN RANGE ?
Y
N
31C
TOO HIGH ?
N
Y
32B
WITHDRAW BALLOON SET DISTANCE
32C
EXTEND BALLOON SET DISTANCE

FIG. 3

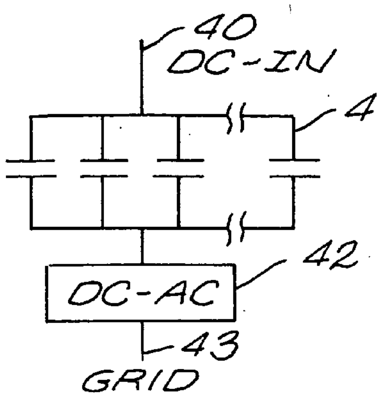
40
DC-IN
41
42
DC-AC
43
GRID

FIG. 4A

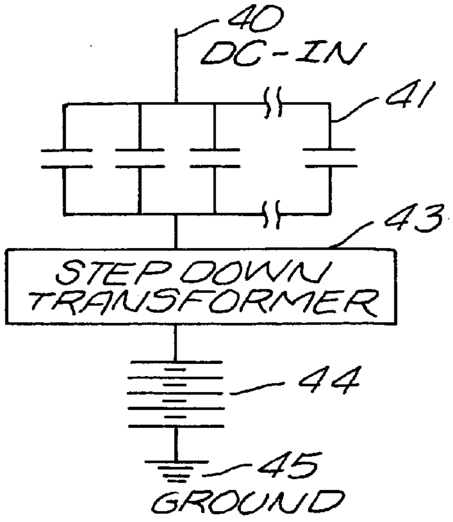
40
DC-IN
41
43
STEP DOWN TRANSFORMER
44
45
GROUND

FIG. 4B

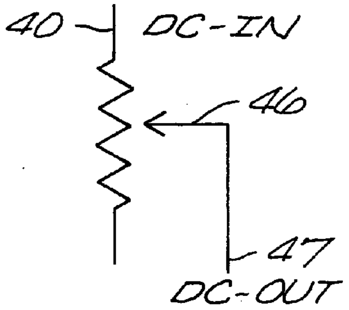
40
DC-IN
46
47
DC-OUT

FIG. 4C

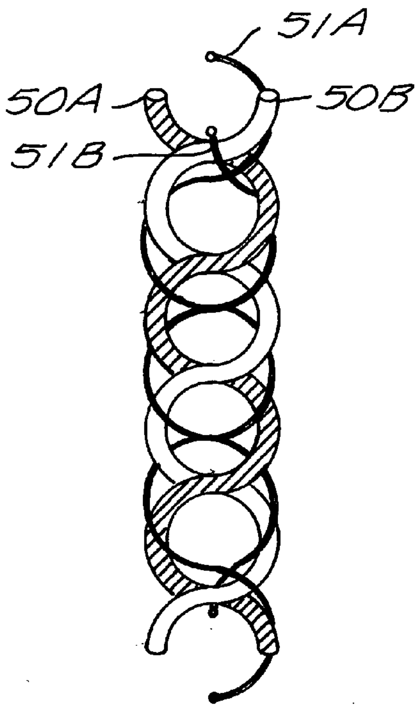
51A
50A
50B
51B

FIG. 5

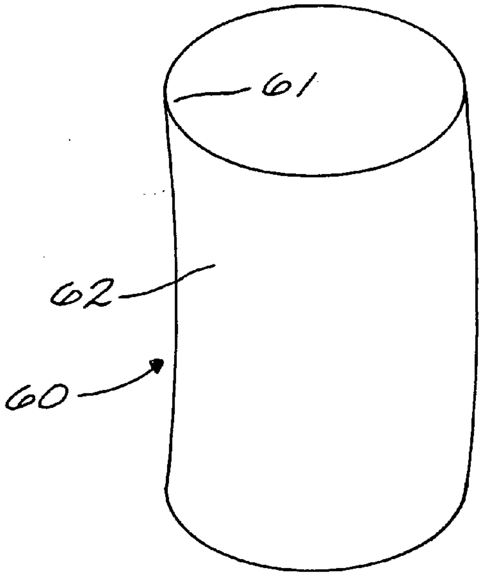

FIG. 6A

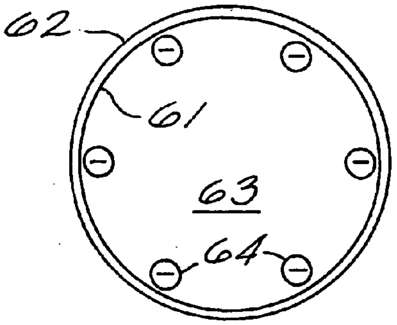

FIG. 6B

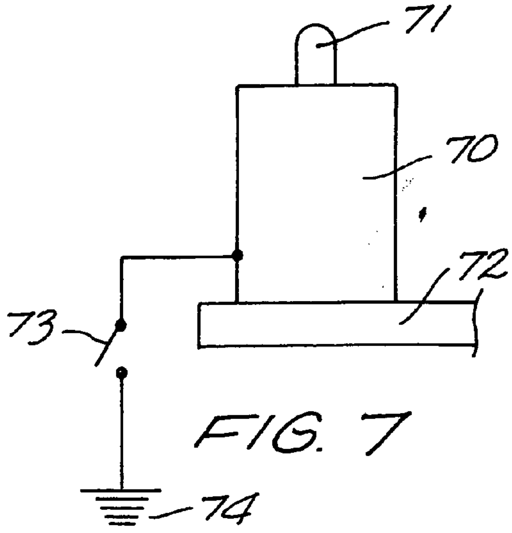
71
70
72
73
74

FIG. 7

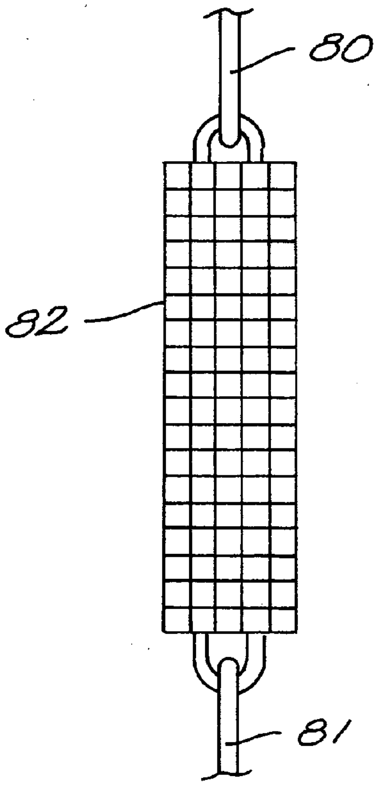

FIG. 8A

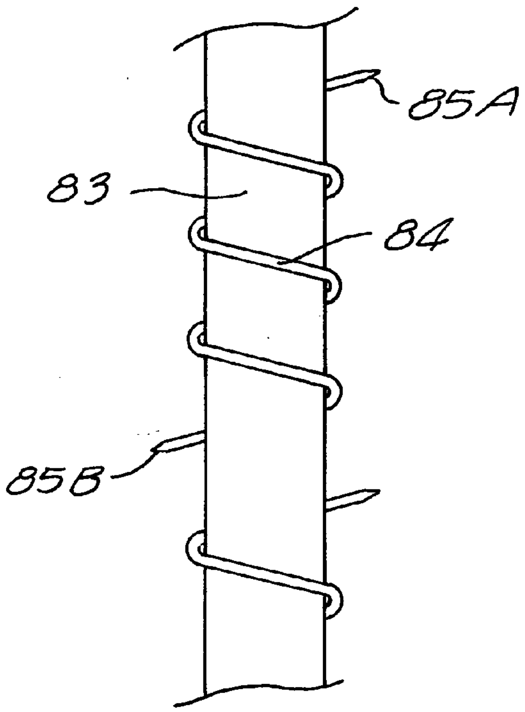
85A
83
84
85B

FIG. 8B

ATMOSPHERIC ELECTRICAL GENERATOR

BACKGROUND OF THE INVENTION

[0001] This invention relates to the production of electrical energy and more particularly to the production of electrical energy from the atmosphere.

[0002] Everyone is familiar with Benjamin Franklin's kite experiment of 1752. Using a kite whose string had become wet, negative charges from the passing clouds flowed into the string, down to the suspended key, and then into a Leyden jar via a thin metal wire. Franklin was protected by a dry silk string; but, when Franklin's knuckle came too close to the key, he received a strong shock. Fortunately, Benjamin Franklin was not killed, others who tried this same experiment were not so lucky.

[0003] Since then, the formation of lightning has remained something of a mystery. Lightning bolts are triggered when a negatively charged cloud base induces a positive charge from the ground, thereby forming a "pathway" for the discharge of the collected electrical energy.

[0004] Lightning travels up to 60,000 miles per hour with a flash that is brighter than ten million 100-watt lightbulbs. This wattage is as much power as is produced by all of the electricity plants in the United States and with a voltage of up to 300 million volts.

[0005] It is this very fact, the power within lightning is immense, that has prevented any successful collection of the electrical energy from lightning. The electricity in lightning is far too extreme for current technology to harness.

[0006] While lightning has attracted a energy starved industrial world, no one has developed any technique to harness this naturally occurring electrical source.

[0007] It is clear there is a continuing need for an electrical source other than carbon-based fuels and that the naturally occurring electricity in the atmosphere is being ignored.

SUMMARY OF THE INVENTION

[0008] The invention is a mechanism which taps into the naturally occurring static electricity in the atmosphere. Whereas heretofore, the attempt to garner electricity from the atmosphere has focused exclusively on capturing lightning, the present invention syphons off the static electricity which is generated from any agitated air and voids lightning.

[0009] Lightning is only the final discharge of the static electricity, whether that lightning is intra-cloud lightning, cloud-to-ground lightning, or inter-cloud lightning. Other types of final discharges are known as heat lightning, summer lightning, sheet lightning, ribbon lightning, silent lightning, ball lightning, bead lightning, elves, jets, and sprites. Well before these discharges are observed, as the atmosphere becomes agitated by wind or thermal, static electricity is being generated.

[0010] The present invention recognizes that this static electricity is being formed and creates a mechanism to capture it.

[0011] The mechanism of this invention utilizes an aircraft such as a lighter than air balloon. While the preferred embodiment uses a foil balloon, a variety of other aircraft are obvious to those of ordinary skill in the art, including, but not limited to: gliders, rubber balloons (such as weather balloons), biaxially-oriented polyethylene terephthalate polyester film balloons, and latex balloons.

[0012] Within this discussion, the balloon is referenced, but, the invention is not intended to be limited solely to balloons.

[0013] The balloon is sent aloft and is tethered by a conductive line. In this context, the conductive line may be any obvious to those of ordinary skill in the art. For the preferred embodiment, the conductive line is a generically referred to as a "poly-rope" and is commercially available through a variety of sources. A suitable conductive line is described in U.S. Pat. No. 5,203,542, entitled "Apparatus for Improved Electric Fence Wire Construction for use with Intensive Grazing" issued Apr. 20, 1993, to Coley, et al. and incorporated hereinto by reference.

[0014] The conductive line is played out of a winch to control the altitude of the balloon. The motor controlling the winch is able to reverse direction to both extend and withdraw the conductive line which is wrapped around a spool on the winch. The winch/spool combination are part of a base unit.

[0015] In some embodiments of the invention, the spool is constructed of rubber so as to insulate the conductive line from the winch assembly. In this embodiment, only the conductive line is charged by the atmospheric static electricity while the winch remains neutral.

[0016] In yet another embodiment, the winch/spool are part of a base unit which is itself isolated from the ground by an insulator. In this embodiment, the entire base unit is charged by the atmospheric static electricity.

[0017] A conductor, such as an insulated wire, is electrically connected to the conductive line. In one embodiment, where the conductive line is electrically isolated from the spool and winch motor, the conductor is connected to the conductive line. In the embodiment where the conductive line is electrically connected to the base unit, then the conductor is connected anywhere on a metallic base unit.

[0018] The other end of the conductor is connected to a load. The load in this case can be any of a variety of electrical loads well known to those of ordinary skill in art, including, but not limited to a motor, a battery system, or the electrical grid for the system.

[0019] In the preferred embodiment, a sensor array is used to monitor the activities both at the base unit (such as electrical flow within the conductor) and in the surrounding locale.

[0020] A sensor monitoring the electrical flow (i.e. voltage and/or current) within the conductor is used to monitor the electrical activity within the conductor.

[0021] In the preferred embodiment, a lightning sensor monitors for lightning activity within the locale. As noted earlier, the electrical characteristic of lightning is so extreme that ideally this discharge is avoided as it might damage the mechanism of this invention.

[0022] The sensor array is utilized by a controller, such as microprocessor, programmed to operate the mechanism as outlined herein.

[0023] The controller operates the winch motor to extend or withdraw the conductive line and by extension the altitude of the balloon. The controller is programmed to operate the winch by monitoring the electrical characteristics of the conductor and adjusting the balloon's altitude to maintain these characteristics within the conductor within a preset range.

[0024] This preset range is established either in the base programming of the controller or is established by an operator of the system.

[0025] As example, by controlling the amount of current being withdrawn from the atmosphere, the mechanism oper-

ates within a safe range and also provides a relatively stable current flow from which a variety of activities can take place (such as DC-AC conversion).

[0026] The controller also utilizes the lightning sensor to protect the mechanism from a lightning strike. Should lightning be detected within a pre-determined range (as established by the software or defined by an operator), then the balloon is pulled down to minimize the risk of damage from a lightning strike.

[0027] The invention, together with various embodiments thereof will be more fully explained by the following description of the accompanying drawings.

DRAWINGS IN BRIEF

[0028] FIG. **1** diagrams the preferred embodiment of the invention.

[0029] FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

[0030] FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

[0031] FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

[0032] FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

[0033] FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

[0034] FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

[0035] FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

DRAWINGS IN DETAIL

[0036] FIG. **1** diagrams the preferred embodiment of the invention.

[0037] Balloon **10** is an aircraft which, in this illustration, is a lighter than air balloon. Wings **10**A, extending from the body of balloon **10**, provide additional lift in air flow **18**. Tail **10**B helps to stabilize balloon **10**.

[0038] Balloon **10** is tethered to the ground via conductive line **12**. As noted earlier, a variety of configurations and materials are available to serve as conductive line **12**. In this illustration, a poly-wire is used. Poly-wire is commercially available through a variety of vendors, including, but not limited to: Jeffers Livestock and Sareba Systems, Inc. of Ellendale, Minn.

[0039] In this embodiment, located proximate to balloon **10**, is an electrical collection enhancement lead **11** which assists in the collection of the static electrical charge in the atmosphere. Electrical collection enhancement lead **11** is configured to attract the static charge and conduct the charge into the conductive line **12**.

[0040] The electricity flows down the conductive line into spool **13**, where the conductive line **12** is collected and either withdrawn or dispensed through operation of winch motor **14**.

[0041] Winch motor **14** and spool **13** are mounted onto base unit **16** which is electrically isolated from ground **7** using insulator **17**. Note, in this embodiment of the invention, when electricity is being collected from the atmosphere, the entire base unit **16** becomes charged. In another embodiment of the invention, spool **13** is constructed of rubber, thereby preventing base unit **16** from becoming charged, thereby restricting the charging from the atmosphere to only conductive line **12**.

[0042] In this embodiment, conductor **6** is connected to base unit **16** (since the entire base unit **16** is charged and the base unit is metallic) to communicate the electrical current to load **5**. Conductor **6** is ideally an insulated wire.

[0043] The electrical current through conductor **6** is measured using sensor **8**.

[0044] In the alternative embodiment discussed above, where only the conductive line **12** is charged, then conductor **6** is connected to conductive line **12**.

[0045] Controller **15**, located in this embodiment on base unit **16**, operates winch motor **14** in response to signals from sensor **8** (measuring the current being discharged to load **5**) to maintain the current flow within a pre-defined range. As the current flow diminishes, then the conductive line **12** extended from spool **13** to increase the altitude of balloon **10** to that more static charge from the atmosphere is gathered; as the current flow falls exceeds a preset level, conductive line **12** is withdrawn onto spool **13** to decrease the static charge being collected from the atmosphere.

[0046] The range of current flow through conductor **6** is ideally set by the program, although some embodiments of the invention permit an operator to establish this range of operation.

[0047] In an alternative embodiment, the sensor monitoring conductor **6** monitors the voltage therein.

[0048] In the preferred embodiment of the invention, controller **15** is also equipped with a lightning sensor **19**. In this embodiment, when lightning is sensed within a preset range, then substantially all of conductive line **12** is wound onto spool **13** to pull balloon **10** near the ground and protect the entire mechanism from being damaged from a lightning discharge.

[0049] In the preferred embodiment, the "safe" distance form lightning is set in the programming of controller **15** and is ideally two miles; other embodiments permit the operator to "safe" distance.

[0050] There are a variety of lightning sensors well known to those of ordinary skill in the art, including, but not limited to those described in: U.S. Pat. No. 7,016,785, entitled "Lightning Detection" issued to Makela, et al. on Mach 21, 2006; U.S. Pat. No. 6,829,911, entitled "Lightning Detection and Prediction Alarm Device" issued to Jones, et al. on Dec. 7, 2004; U.S. Pat. No. 7,200,418, entitled "Detection of Lightning" issued to Karikuranta, et al. on Apr. 3, 2007; and U.S. Pat. No. 6,961,662, entitled "Systems and Methods for Spectral Corrected Lightning Detection" issued to Murphy on Nov. 1, 2005; all of which are incorporated hereinto by reference.

[0051] In another embodiment of the invention, controller **15** is not located on base unit **16**, rather it is remote and communicates its control signals to winch motor **14** using radio waves.

[0052] FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

[0053] Static charges **23** are generated in the atmosphere by agitated air. These static charges are often collected at the bottom of clouds, but exist in other environments as well.

[0054] Balloon **21** is extended into this strata of static charges **23** which are then attracted to conductive line **12** to flow to base unit **22** and then onto load **5**.

[0055] By increasing or decreasing the altitude of balloon **21** (defined by the length of the extended conductive line **12**),

conductive line **12** is selectively exposed to varying densities and levels of the static charge strata, and by extension, the current flow or voltage is increased or decreased.

[0056] FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

[0057] Once the program starts **30**, the lightning sensor is checked to determine if lightning has occurred within the unsafe range **31**A, if it has, then the balloon is lowered **32**A, and the program continues monitoring the status of lightning until no lightning is detected.

[0058] When the lightning status is acceptable, then the current within the conductor is checked to see if the current is within the prescribed range **31**B. If the current is acceptable (within range) the program returns to check the lightning status **31**A; otherwise a determination is made to see if the current is above the prescribed range **31**C.

[0059] If the current is above the prescribed range, then the altitude of the balloon is withdrawn a set amount **32**B (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

[0060] If the current is below the prescribed range, then the altitude of the balloon is extended a set amount **32**C (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

[0061] In this manner of feed-back and minor adjustments in the altitude of the balloon, the current is maintained within a prescribed range which can be handled by the downstream electrical system.

[0062] As noted earlier, some embodiments of the invention monitor the voltage instead of the current.

[0063] FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

[0064] By maintaining the voltage being collected in a prescribed range, an electrical conversion system is easily designed. While FIGS. **4**A, **4**B, and **4**C illustrate some electrical configurations, those of ordinary skill in the art readily recognize a variety of other configurations which will serve the same function.

[0065] Referencing FIG. **4**A, Direct Current In (DC IN) **40** is buffered by a gang of capacitors **41** before being communicated to a DC/AC converter **42**. The DC/AC converter converts the direct current into a an alternating current suitable for placement over an existing electrical grid **43** such as normally found from a power-plant.

[0066] Those of ordinary skill in the art readily recognize a variety of DC/AC converters, including, but not limited to: U.S. Pat. No. 7,394,671, entitled "Controller IC, DC-AC Conversion Apparatus, and parallel running system of DC-AC Conversion Apparatuses" issued to Fukumoto, et al. on Jul. 1, 2008; and, U.S. Pat. No. 7,330,366, entitled "DC-AC Converter" issued to Lee, et al. on Feb. 12, 2008; all of which are incorporated hereinto by reference.

[0067] FIG. **4**B illustrates an electrical arrangement suitable for use in charging a battery. DC IN **40** is buffered by capacitor bank **41** before entering into a step down transformer **43**. Step down transformer **43** reduces the voltage so that the voltage can safely be introduced into battery **44** which is connected to ground **45** at the battery's other pole.

[0068] Those of ordinary skill in the art readily recognize a variety of batteries which will work in this capacity, including, but not limited to those described in: U.S. Pat. No. 7,378,181, entitled "Electric Storage Battery Construction and Manufacture" issued to Skinlo on May 27, 2008; U.S. Pat. No. 7,388,350, entitled "Battery with Electronic Compartment" issued to Wright on Jun. 17, 2008; U.S. Pat. No. 7,397,220, entitled "Connection Member and Battery Pack" issued to Uchida, et al. on Jul. 8, 2008; and, U.S. Pat. No. 7,375,492, entitled "Inductively Charged Battery Pack" issued to Calhoon, et al. on May 20, 2008; all of which are incorporated hereinto by reference.

[0069] In FIG. **4**C, DC IN **40** is fed into an adjustable rheostat **46** which is controlled by the controller so that the DC OUT **47** falls within a specified range.

[0070] FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

[0071] This type of conductive line is commonly called poly-wire and consists of multiple interwoven strands of plastic **50**A and **50**B woven into a cord or rope arrangement having intertwined therein exposed metal wires **51**A and **51**B. While this illustration shows two plastic strands and two metal wires, any number of possible combinations is possible.

[0072] The exposed metal wires **51**A and **51**B attract the atmospheric static charge and transmit the charge down to the base unit (not shown).

[0073] FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

[0074] This conductive line utilizes a tube **60** having an outer layer **62** of PET Film (Biaxially-oriented polyethylene terephtalate polyester film) which provides exceptionally high tensile strength and is chemically and dimensionally stable. The tube has an ideal diameter of between two and three inches.

[0075] An interior metal coating **61** provides an initial conduit for the flow of static charge. The static charge through the metal forces the tube to expand due to the repulsion experienced by like charges. Further, the flow of electricity causes the interior of the tube **60** to become ionized to provide an additional pathway for the atmospheric static charges to the base unit (not shown).

[0076] Because outer layer **62** provides a gas barrier, the resulting ionization is not dissipated by air currents, thereby providing a highly stable pathway.

[0077] FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

[0078] In this embodiment, controller box **70**, resting on insulating pad **72**, is in communication with the sensors as described above. Using the input from these sensors, when there is flow of electricity through the base unit, warning flashing light **71** is illuminated. To electrically neutralize the mechanism, switch **73** is activated to pass any existing current into the ground **74**.

[0079] FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

[0080] Referencing FIG. **8**A, enhanced electrical collection lead **82** is a wire mesh which is in electrical communication with conductive line **81** and balloon **80**. Because of the significant amount of metal exposed by enhanced electrical collection lead **82**, more static electricity from the atmosphere is drawn to the collection lead **82**, and then down conductive line **81** to the base unit (not shown).

[0081] Conductive lead **82** is positioned proximate to balloon **80**.

[0082] In FIG. **4**B, poly-wire **83** has enhanced electrical collection leads **84** wrapped therearound. Collection leads **84** have pointed ends **85**A and **85**B which have a propensity to attract more electricity than rounded ends do.

[0083] It is clear from the foregoing that the present invention captures an entirely new source of electrical energy.

What is claimed is:

1. A mechanism to tap an electrical source comprising:

a) a lighter than air balloon;

b) a base unit having a spool of conductive line on a winch motor, one end of said conductive material secured to said lighter than air balloon, said winch motor capable of selectively extending or withdrawing said conductive line from said spool, one end of said conductive line connected to said lighter than air balloon;

c) an insulator electrically isolating said conductive line from ground; and,

d) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load.

2. The mechanism to tap an electrical source according to claim **1**, further including a controller, said controller selectively operating said winch motor based upon a voltage level in said conductor.

3. The mechanism to tap an electrical source according to claim **2**,

a) further including an electrical flow sensor monitoring electrical flow through said conductor and generating an electrical flow indicia indicative of said electrical flow in said conductor, said electrical flow indicia communicated to said controller; and,

b) wherein said controller operates said winch motor such that said electrical flow indicia remains within a selected operating range.

4. The mechanism to tap an electrical source according to claim **3**, wherein said selected operating range is established by an operator.

5. The mechanism to tap an electrical source according to claim **4**, further including a warning light activated when said electrical flow indicia is non-zero.

6. The mechanism to tap an electrical source according to claim **3**,

a) further including a lightning sensor generating a lightning presence indicia indicative of lightning within a prescribed range, said presence indicia being communicated to said controller; and,

b) wherein said controller, in response to said lightning presence indicia, operates said winch motor to withdraw substantially all of said conductive line onto said spool.

7. The mechanism to tap an electrical source according to claim **6**, wherein said prescribed range is established by an operator.

8. The mechanism to tap an electrical source according to claim **3**, further including an electrical collection enhancement lead in electrical contact with a first end said conductive material, said electrical collection enhancement lead configured to attract static electricity.

9. The mechanism to tap an electrical source according to claim **8**, wherein said electrical collection enhancement lead is positioned proximate to said lighter than air balloon.

10. The mechanism to tap an electrical source according to claim **9**, wherein said electrical collection enhancement lead includes at least two pointed electrical conductors.

11. The mechanism to tap an electrical source according to claim **1**, wherein said insulator electrically isolates said conductive line from said winch motor.

12. The mechanism to tap an electrical source according to claim **1**, wherein said insulator electrically isolates said base unit from ground.

13. A mechanism comprising:

a) an airborne aircraft tethered by a conductive line secured to a winch capable of extending or withdrawing said conductive line from a spool, thereby adjusting an altitude of the airborne aircraft;

b) an insulator electrically isolating said conductive line from ground; and,

c) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load.

14. The mechanism according to claim **13**, further including a controller, said controller selectively operating said winch motor based upon a voltage level in said conductor.

15. The mechanism to tap an electrical source according to claim **14**,

a) further including a lightning sensor communicating with said controller; and,

b) wherein said controller, in response to selected signals from said lightning sensor withdraws substantially all of said conductive line.

16. A mechanism to tap an electrical source comprising:

a) a lighter than air balloon;

b) a base unit having a spool of conductive line on a winch motor, one end of said conductive material secured to said balloon, said winch motor capable of adjusting an altitude of said lighter than air balloon by selectively extending or withdrawing said conductive line from said spool;

c) an insulator electrically isolating said conductive line from ground; and,

d) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load;

e) a sensor array having,

  1) an electrical flow sensor monitoring electrical flow through said conductor, and,

  2) a lightning sensor monitoring existence of proximate lightning; and,

f) a controller responsive to signals from said sensor array to selectively operate said winch motor.

17. The mechanism to tap an electrical source according to claim **16**,

a) further including a warning light; and,

b) wherein said controller activates said warning light when said electrical flow in said conductor is non-zero.

18. The mechanism to tap an electrical source according to claim **17**, further including an electrical collection enhancement lead in electrical contact with said conductive material, said electrical collection enhancement lead configured to attract static electricity and positioned proximate to said lighter than air balloon.

* * * * *

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 1

Fig. 1.

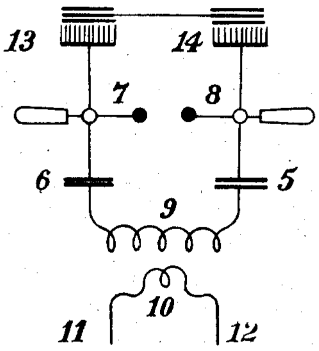
13
14
7
8
6
5
9
10
11
12

Fig. 2.

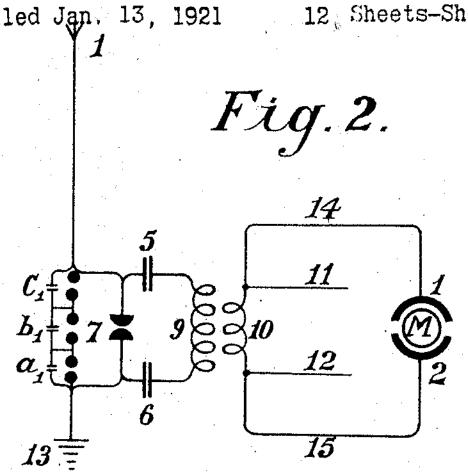
1
5
14
11
1
C1
b1
7
9
10
M
a1
12
2
6
13
15

Fig. 3.

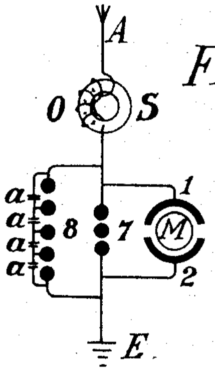
A
0
S
a
a
a
a
8
7
1
M
2
E.

Fig. 4.

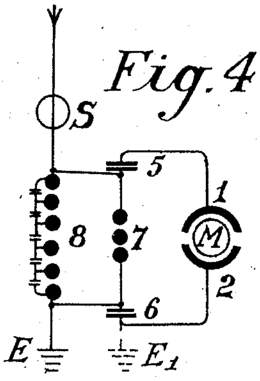
S
5
1
8
7
M
2
6
E
E1

Fig. 5.

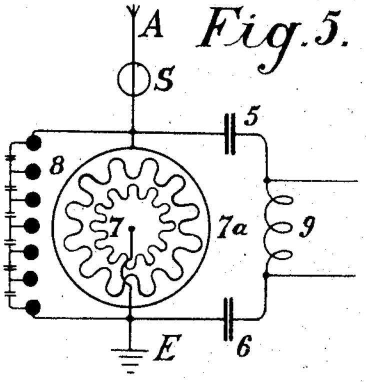
A
S
5
8
7
7a
9
E
6

Fig. 6

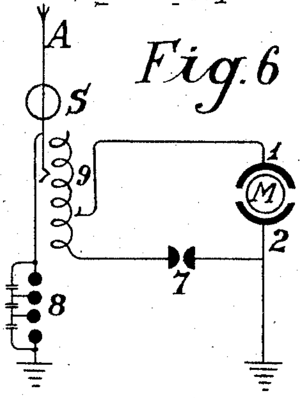

Inventor
Hermann Plauson

by Knight Bros
Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 2

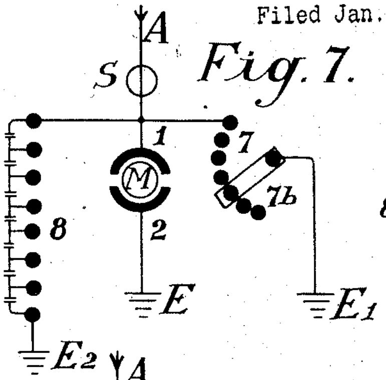
A
S
1
M
2
7
7b
8
E
E1
E2

Fig. 7.

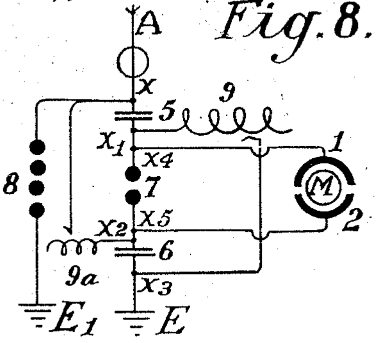
A
x
5
9
x1
x4
1
M
7
8
x5
2
x2
6
9a
x3
E1
E

Fig. 8.

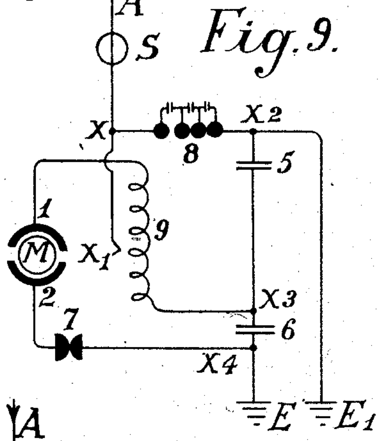
A
S
x
8
x2
5
1
M
9
x1
2
x3
7
6
x4
E
E1

Fig. 9.

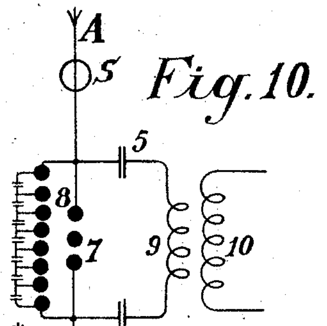
A
S
5
8
7
9
10
6
E

Fig. 10.

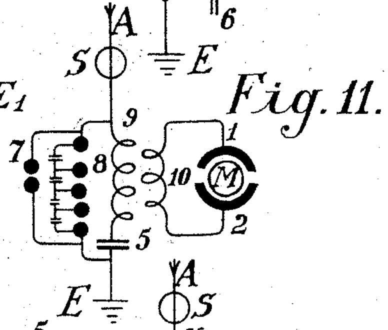
A
S
9
1
7
8
10
M
2
5
E

Fig. 11.

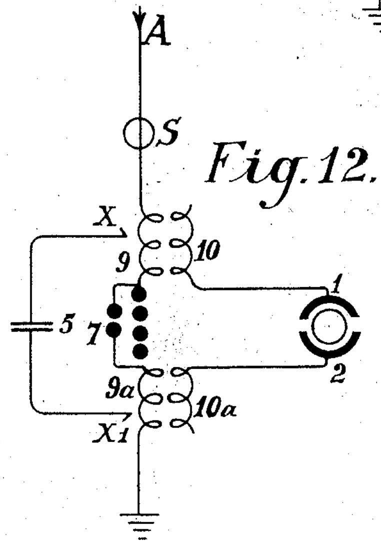
A
S
x
9
10
1
5
7
2
9a
10a
x1

Fig. 12.

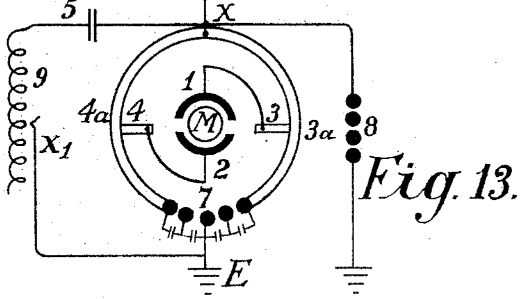
A
S
5
x
9
1
4a
4
M
3
3a
8
x1
2
7
E

Fig. 13.

Inventor
Normann Plauson

by Knight Bros
attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 3

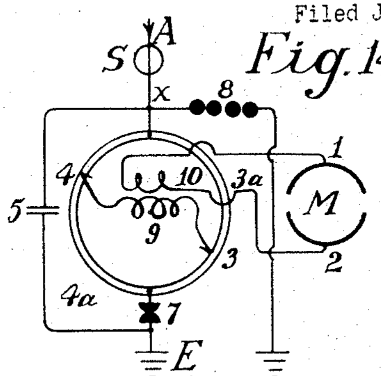
Fig. 14.
S
A
x
8
1
4
10
3a
5
M
9
3
2
4a
7
E

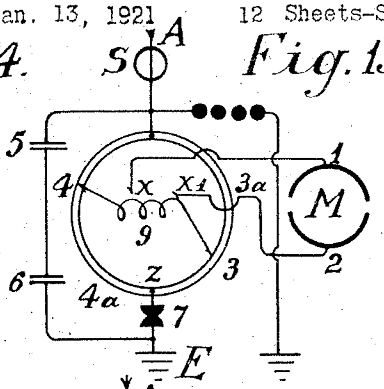
Fig. 15.
S
A
5
1
4
x
X1
3a
M
9
3
2
6
z
4a
7
E

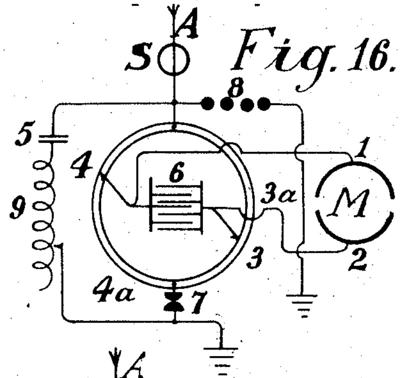
Fig. 16.
S
A
8
5
1
4
6
3a
M
9
3
2
4a
7

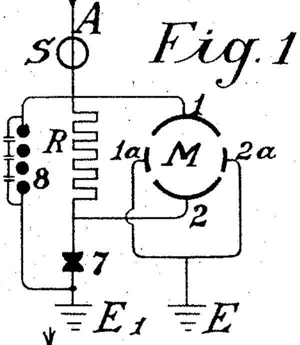
Fig. 17.
S
A
1
R
1a
M
2a
8
2
7
E1
E

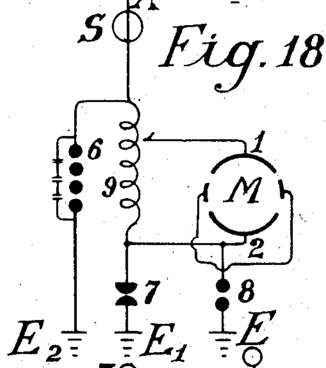
Fig. 18.
S
A
1
6
9
M
2
7
8
E2
E1
E

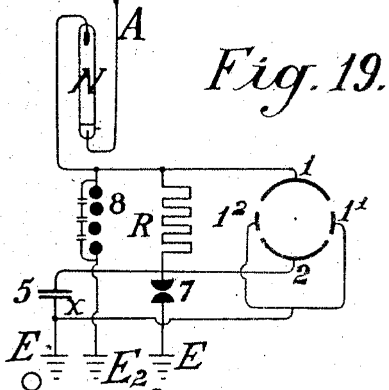
Fig. 19.
A
N
1
8
R
1²
1¹
2
5
x
7
E
E2
E

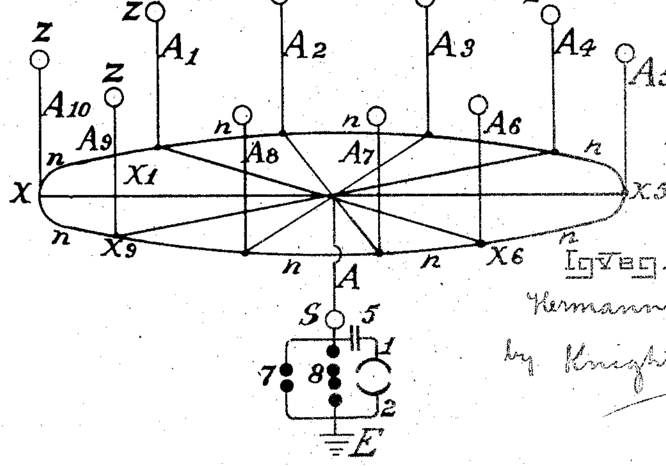
Fig. 20.
z
z
z
A1
A2
A3
A4
A5
A10
A9
n
A8
n
A7
A6
n
n
X1
X
X5
n
X9
n
n
X6
n
A
S
5
1
7
8
2
E

Inventor
Hermann Plauson
by Knight Bros
Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 4

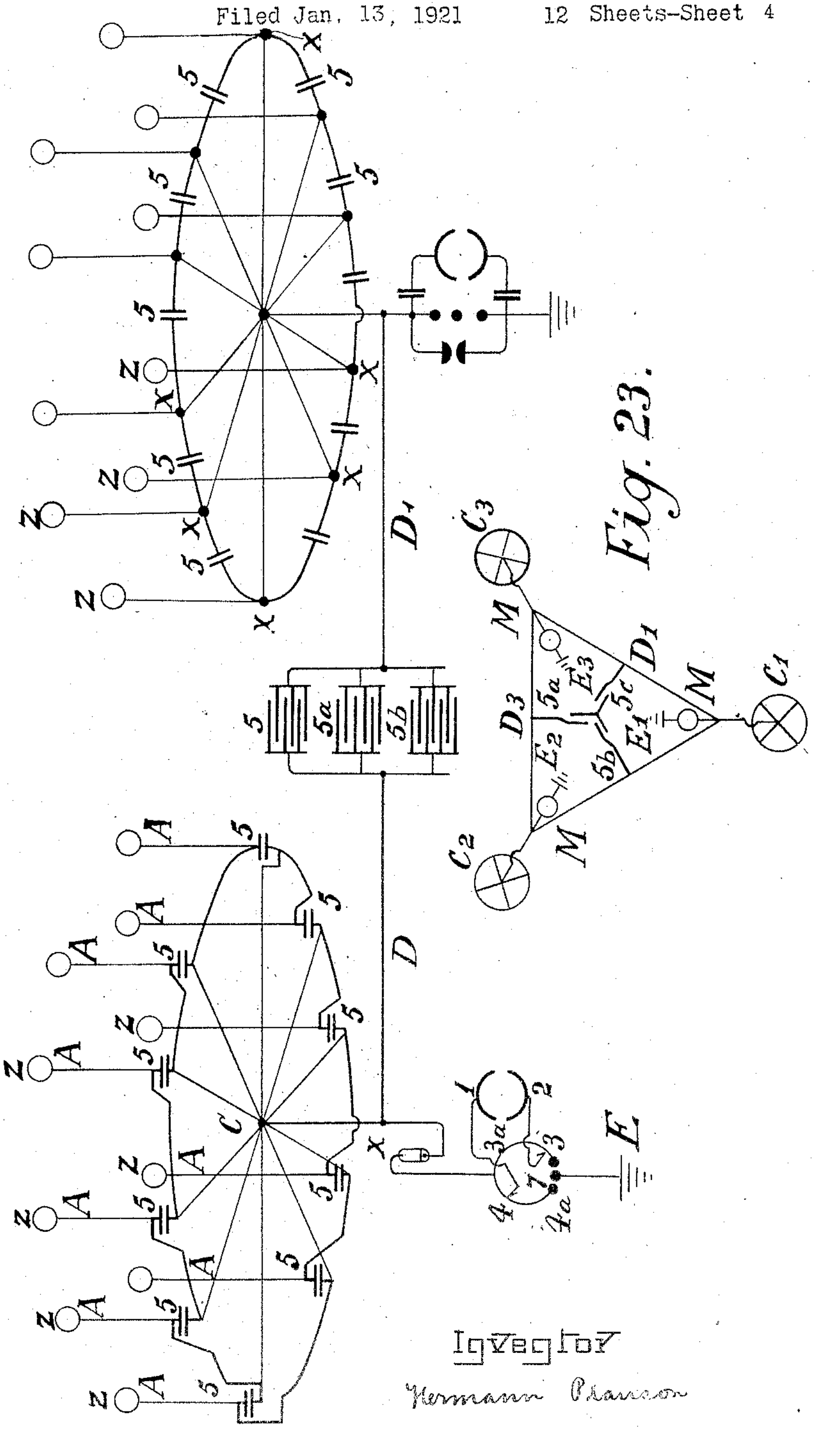
Fig. 22.
Fig. 21.
Fig. 23.
5
5a
5b
X
Z
A
C
D
D1
E
1
2
3
3a
4
4a
7
C1
C2
C3
M
D1
D3
E1
E2
E3
5c

Inventor
Hermann Plauson
by Knight Bros
Attorneys

June 9, 1925. H. PLAUSON 1,540,998

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 5

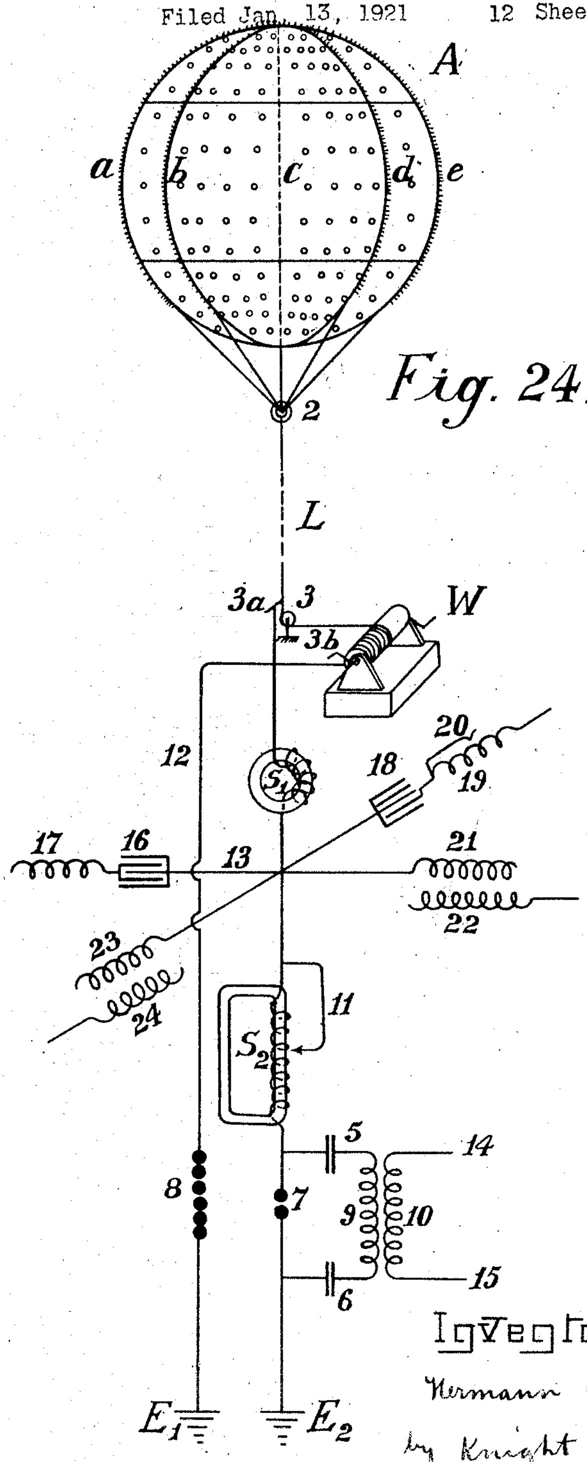
A
a
b
c
d
e
2
L
3a
3
3b
W
12
S1
20
18
19
17
16
13
21
22
23
24
11
S2
5
14
8
7
9
10
15
6
E1
E2

Fig. 24.

Inventor

Hermann Plauson

by Knight Bros

Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 6

Fig. 25.

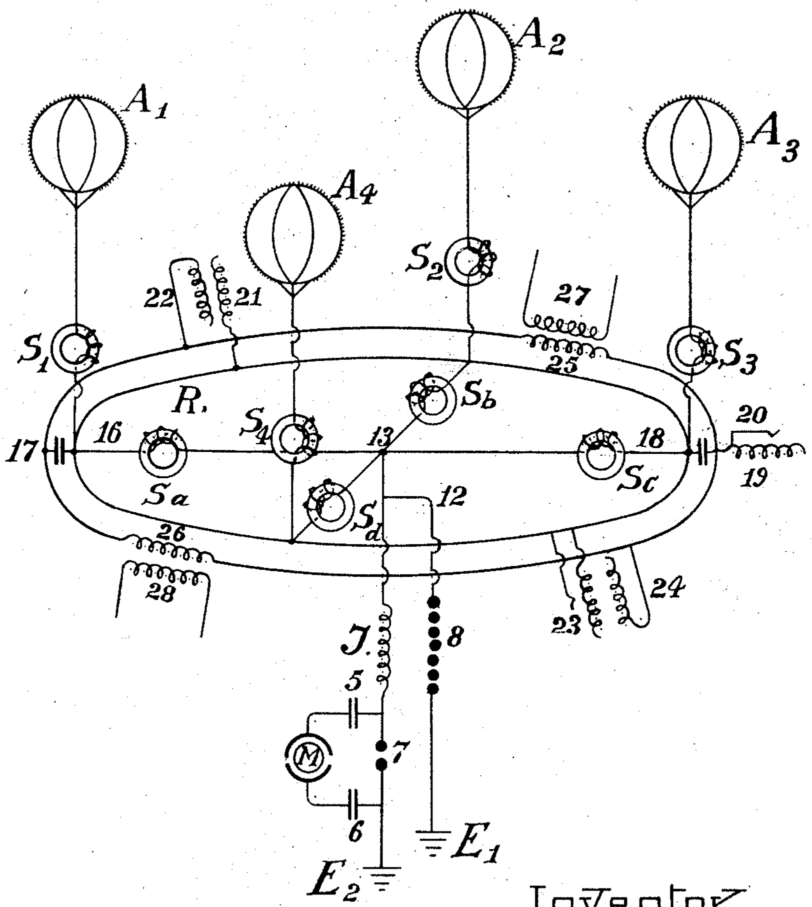

A_2
A_1
A_3
A_4
S_2
22
21
27
S_1
25
S_3
R.
S_b
S_4
16
13
18
20
17
19
S_a
S_c
S_d
12
26
28
24
23
J.
8
5
M
7
6
E_1
E_2

Inventor

Hermann Plauson

by

Knight Bros

attorneys

June 9, 1925. **1,540,998**

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 7

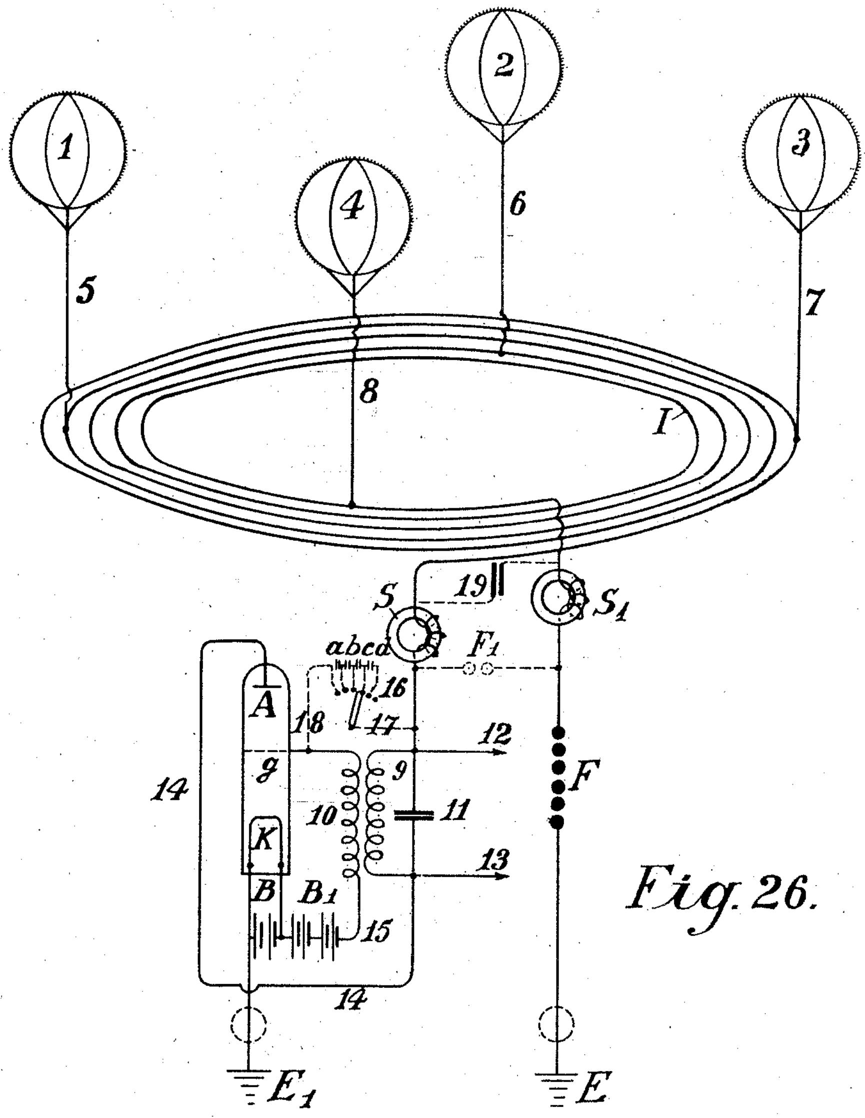
2
1
3
4
6
5
7
8
I
19
S
S_1
abcd
F_1
16
A
18
17
12
14
g
9
10
11
K
13
B
B_1
15
14
F
E_1
E

Fig. 26.

Inventor
Hermann Plauson
by
Knight Bros
Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921. 12 Sheets-Sheet 8

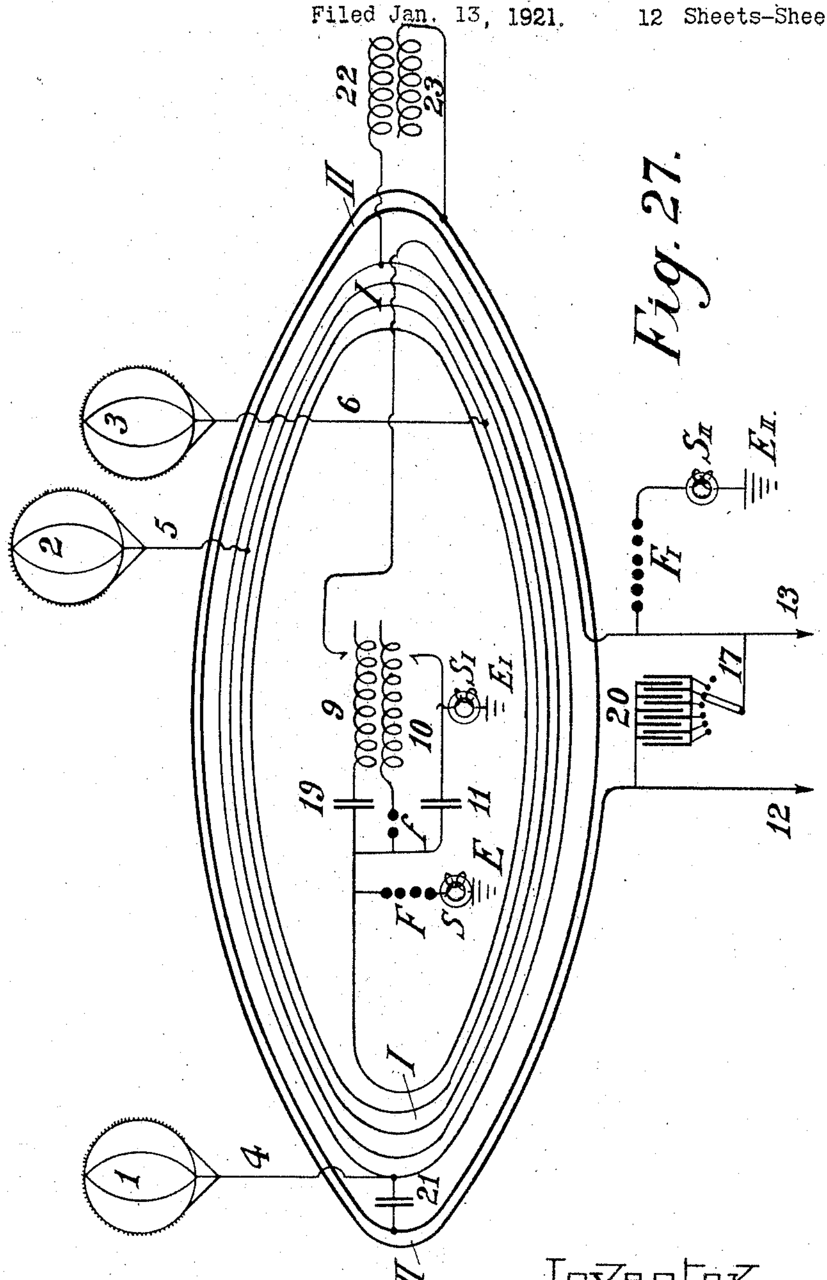
22
23
II
I
6
3
2
5
1
4
21
II
I
9
10
19
11
f
F
S
E
S_I
E_I
F_I
S_{II}
E_{II}
13
12
20
17

Fig. 27.

Inventor

Hermann Plauson

by Knight Bros

Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 9

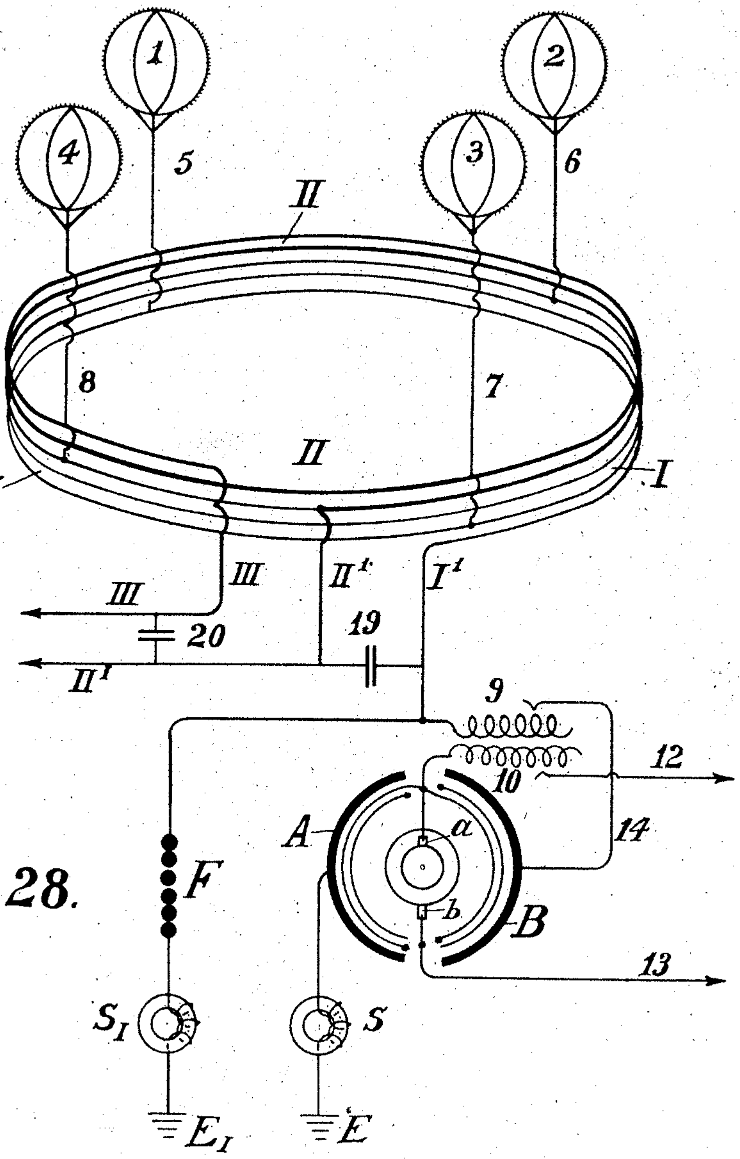
1
2
4
5
3
6
II
8
7
II
I
I
III
II1
I1
III
20
19
II1
9
12
10
A
a
14
b
B
13
F
S1
S
E1
E

Fig. 28.

Inventor
Hermann Plauson
by
Knight Bros
attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 10

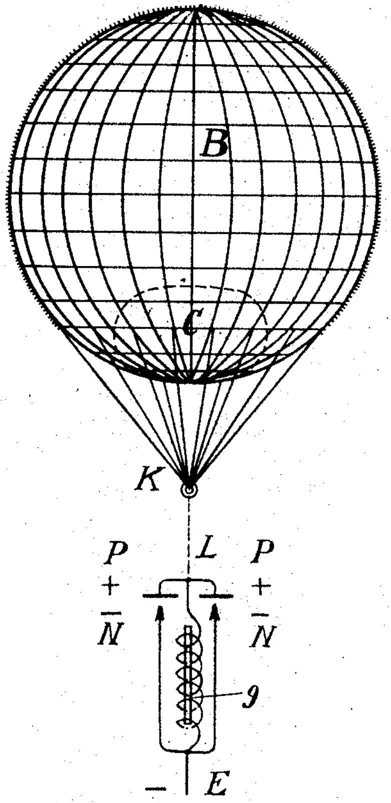
B
C
K
L
P
+
$\overline{N}$
P
+
$\overline{N}$
9
−
E

Fig. 29.

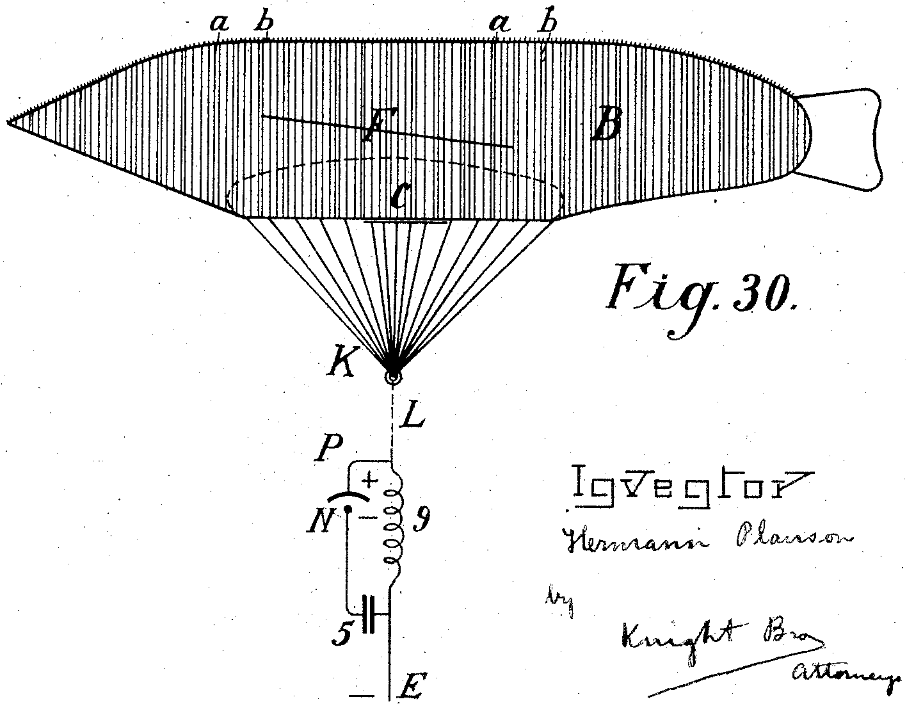
a
b
a
b
F
B
C
K
L
P
+
N
−
9
5
−
E

Fig. 30.

Inventor
Hermann Plauson
by
Knight Bros
Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan 13, 1921 12 Sheets-Sheet 11

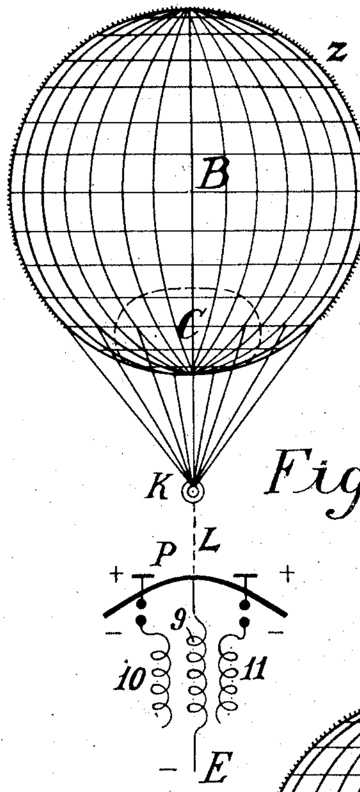
Z
B
C
K
L
P
+
+
−
−
9
10
11
−
E

Fig. 31.

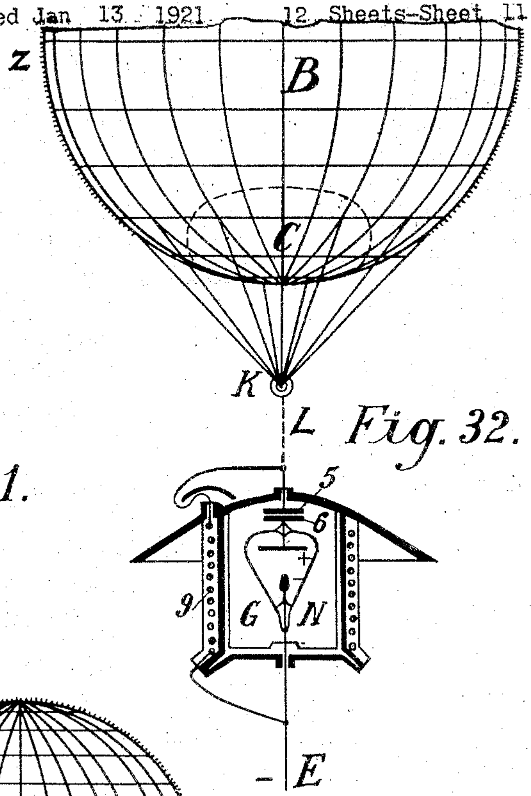
Z
B
C
K
L
5
6
+
−
9
G
N
−
E

Fig. 32.

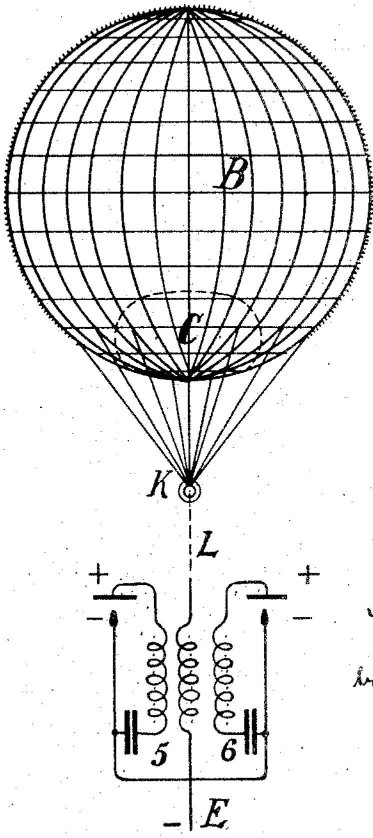

Fig. 33.

Inventor
Hermann Plauson
by Knight Bros
attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 12

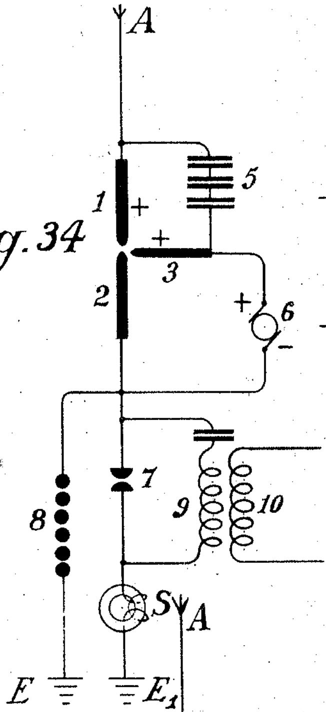
A
1
+
+
5
3
2
+
6
−
7
8
9
10
S
E
E1

Fig. 34

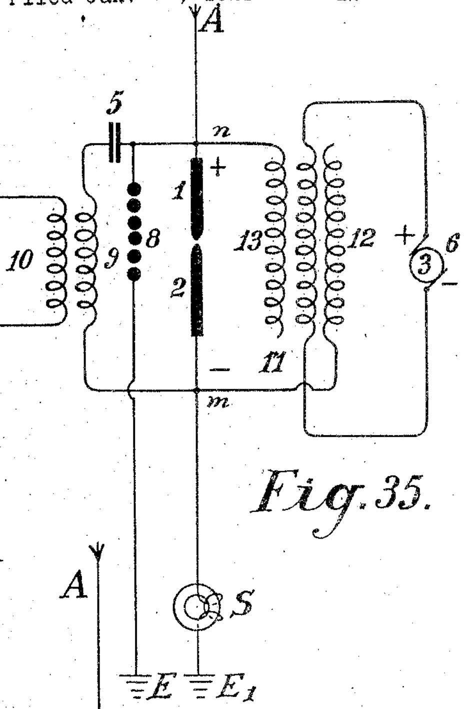
A
5
n
+
1
8
9
10
13
12
+
3
6
−
2
−
11
m
S
E
E1

Fig. 35.

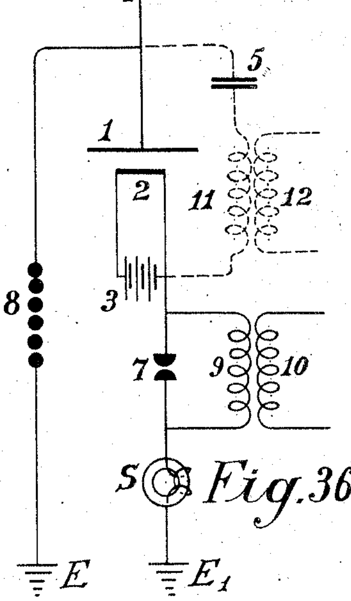
A
5
1
2
11
12
8
3
7
9
10
S
E
E1

Fig. 36.

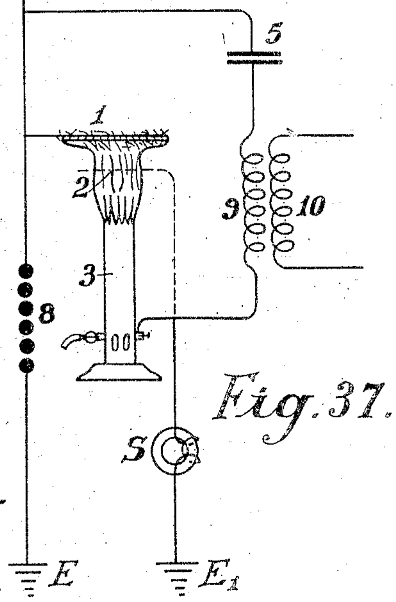
A
5
1
2
3
9
10
8
S
E
E1

Fig. 37.

Inventor

Hermann Plauson

by Knight Bros

Attorneys

Patented June 9, 1925. **1,540,998**

UNITED STATES PATENT OFFICE.

HERMANN PLAUSON, OF HAMBURG, GERMANY.

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY.

Application filed January 13, 1921. Serial No. 437,107.

To all whom it may concern:

Be it known that I, Hermann Plauson, Esthonian subject, residing in Hamburg, Germany, have invented certain new and useful Improvements in the Conversion of Atmospheric Electric Energy, of which the following is a specification.

Methods of obtaining atmospheric electricity by means of metallic nettings set with spikes which are held by means of ordinary or anchored kite balloons made of fabrics and filled with hydrogen, are in theory already known. Atmospheric electricity obtained in this way has been suggested to be used in the form of direct current for the charging of accumulators. This knowledge however is at present only theoretical as the conversion in practice has hitherto been a failure. No means are known of protecting the apparatus from destruction by lightning. The balloons used for collecting the charge must also be made of very large size in order to be able to support the weight of the metallic netting and the heavy cable connections.

Instead of using heavy metallic netting as collectors attached to single air balloons of non-conducting materials which are liable to be torn and are permeable to the gas, it is proposed to use metallic balloon collectors which have the following important advantages—

(*a*) The metallic cases are impenetrable to helium and hydrogen; they also represent large metallic weather-proof collecting surfaces.

(*b*) Radio active means and the like may be easily applied internally or externally; whereby the ionization is considerably increased and therewith also the quantity of atmospheric electricity capable of being collected.

(*c*) Such balloon collectors of light metal do not require to be of large size as they have to carry only their own moderate weight, and that of the conducting cable or wire.

(*d*) The entire system therefore offers little surface for the action of storm and wind and is resistant and stable.

(*e*) Each balloon can be easily raised and lowered by means of a winch so that all repairs, recharging and the like can be carried out without danger during the operation.

It is further proposed to use a collecting aerial network of several separate collectors spread out in the air above the earth, which collectors are interconnected by electrical conductors.

According to this invention charges of atmospheric electricity are not directly converted into mechanical energy, and this forms the main difference from previous inventions, but the static electricity which runs to earth through aerial conductors in the form of direct current of very high voltage and low current strength is converted into electro-dynamic energy in the form of high frequency vibrations. Many advantages are thereby obtained and all disadvantages avoided.

The very high voltage of static electricity of a low current strength can be converted by this invention to voltages more suitable for technical purposes and of greater current strength. By the use of closed oscillatory circuits it is possible to obtain electromagnetic waves of various amplitude and thereby to increase the degree of resonance of such current. Such resonance allows various values of inductance to be chosen whereby again the governing of the starting and stopping of machines driven thereby by simply tuning the resonance between coils of the machine and the transformer circuit forming the resonance can easily be obtained. Further, such currents have the property of being directly available for various uses, even without employing them for driving motors, of which there may be particularly mentioned, lighting, production of heat and use in electro-chemistry.

Further, with such currents a series of apparatus may be fed without direct current supply through conductors and also the electro-magnetic high frequency currents may be converted by means of special motors adapted for electro-magnetic oscillations into mechanical energy, or finally converted by special machines into alternating current of low frequency or even into direct current of high potential.

The invention is more particularly described with reference to the accompanying diagrams in which:—

Figure 1 is an explanatory figure.

Figure 2 is a diagrammatic view of the simplest form.

Figure 3 shows a method of converting atmospheric electrical energy for use with motors.

Figure 4 is a diagram showing the use of protective means.

Figure 5 is a diagram of an arrangement for converting large current strengths.

Figure 6 is a diagram of an arrangement including controlling means.

Figure 7 shows means whereby the spark gap length can be adjusted.

Figure 8 shows a unipolar connection for the motor.

Figure 9 shows a weak coupled system suitable for use with small power motors.

Figures 10, 11 and 12 show modified arrangements.

Figure 13 shows a form of inductive coupling for the motor circuit.

Figure 14 is a modified form of Figure 13 with inductive coupling.

Figure 15 is an arrangement with non-inductive motor.

Figure 16 is an arrangement with coupling by condenser.

Figures 17, 18 and 19 are diagrams of further modifications.

Figure 20 shows a simple form in which the aerial network is combined with special collectors.

Figure 21 shows diagrammatically an arrangement suitable for collecting large quantities of energy.

Figure 22 is a modified arrangement having two rings of collectors.

Figure 23 shows the connections for three rings of collectors.

Figure 24 shows a collecting balloon and diagram of its connection of condenser batteries.

Figures 25 and 26 show modified collector balloon arrangements.

Figure 27 shows a second method of connecting conductor for the balloon aerials.

Figure 28 shows an auto-transformer method of connection.

Figure 29 shows the simplest form of construction with incandescent cathode.

Figure 30 shows a form with cigar shaped balloon.

Figure 31 is a modified arrangement.

Figure 32 shows a form with cathode and electrode enclosed in a vacuum chamber.

Figure 33 is a modified form of Figure 32.

Figure 34 shows an arc light collector.

Figure 35 shows such an arrangement for alternating current.

Figure 36 shows an incandescent collector with Nernst lamp.

Figure 37 shows a form with a gas flame.

Figure 1 illustrates a simple diagram for converting static electricity into dynamic energy of a high number of oscillations. For the sake of clearness in the drawings an influence machine is assumed to be employed and not an aerial antenna. 13 and 14 are combs for collecting the static electricity of the influence machine. 7 and 8 are spark discharging electrodes, 6 and 5 condensers, 9 an inductive primary coil, 10 secondary coil, 11 and 12 ends of conductors of the secondary coil 10. When the disc of the static influence machine is rotated by mechanical means, the combs collect the electric charges one the positive and the other the negative, and charge the condensers 5 and 6 until such a high potential is formed across the spark gap 7—8, that the spark gap is jumped. As the spark gap 7—8 forms a closed circuit with condensers 6 and 5, and inductive resistance 9, as is well known, waves of high frequency electromagnetic oscillations will pass in this circuit.

The high frequency of the oscillations produced in the primary circuit induces waves of the same periodicity in the secondary circuit. Thus in the primary circuit electromagnetic oscillations are formed by the passage of the spark over the spark gap and these waves are maintained by fresh charges of static electricity.

By suitably selecting the ratio between the number of the coils in the primary and secondary circuits with regard to a correct application of the co-efficients of resonance (capacity, inductance, and resistance) the high voltage of the primary circuit may be suitably converted into low voltage and high current strength.

When the oscillatory discharges in the primary circuit becomes weaker or entirely cease, the condensers are charged again by the static electricity until the accumulated charge again breaks down the spark gap. All this is repeated as long as electricity is produced by the static machine by employing mechanical energy.

An elementary form of the invention is shown in Figure 2 in which two spark gaps in parallel are used one of which may be termed the working gap 7 in Figure 2, whilst the second serves as a safety device for excess voltage and consists of a larger number of spark gaps than the working section, which gaps are arranged in series and are bridged by very small capacities as is illustrated in a_1, b_1, c_1, Figure 2 which allow of uniform sparking in the safety section.

In Figure 2 A is the aerial antenna for collecting charges of atmospheric electricity, 13 is the earth connection of the second part of the spark gap, 5 and 6 are condensers, 9 a primary coil. Now when through the aerial A the positive atmospheric electricity seeks to combine with the negative charge to earth, this is prevented by (the air gap between) the spark gaps. The resistance of the spark gap 7 is, as shown in the drawings, lower than that of the other safety section which consists of three spark gaps connected in

series, and consequently a three times greater air resistance is offered by the latter.

So long therefore, as the resistance of the spark gap 7 is not overloaded, so that the other spark gaps have an equal resistance with it the discharges take place only over spark gap 7. Should however the voltage be increased by any influences so that it might be dangerous for charging the condensers 5 and 6 or for the coil insulation 9 and 10 in consequence of break down, by a correct regulation of this spark gap the second spark gap can discharge free from inductive effects direct to earth without endangering the machine.

Without this second spark gap, arranged in parallel having a higher resistance than the working spark gap it is impossible to collect and render available large quantities of electrical energy.

The action of this closed oscillation circuit consisting of spark gap 7, two condensers 5 and 6, primary coil 9, and also secondary coil 10 is exactly the same as the one described in Figure 1 with the arrangement of the static induction machine with the only difference that here the second spark gap is provided. The electromagnetic high frequency alternating current obtained can be tapped off from the conductors 11 and 12 for lighting and heating purposes. Special kinds of motors adapted for working with these peculiar electrical charges may be connected at 14 and 15 which can work with static electricity charges or with high frequency oscillations.

In addition to the use of spark gaps in parallel a second measure of security is also necessary for taking off the current. This precaution consists according to this invention, in the introduction of and method of connecting certain protective electromagnets or choking coils in the aerial circuit as shown by S in Figure 3.

A single electromagnet only having a core of the thinnest possible separate laminations is connected with the aerial.

In the case of high voltages in the aerial network or at places where there are frequent thunder storms, several such magnets may however be connected in series.

In the case of large units or plants several electromagnets can be employed in parallel or in series parallel.

The windings of these electromagnets may be simply connected in series with the aerials. In this case the winding preferably consists of several thin parallel wires, which make up together, the necessary section.

The winding may be made of primary and secondary windings in the form of a transformer. The primary winding will be then connected in series with the aerial network, and the secondary winding more or less short-circuited over a regulating resistance or an induction coil. In the latter case it is possible to regulate to a certain extent the effect of the choking coils. In the further description of the connecting and constructional diagrams the aerial electromagnet choke coil is indicated by a simple ring S.

Figure 3 shows the simplest way of converting atmospheric electricity into electromagnetic wave energy by the use of special motors adapted for high oscillatory currents or static charges of electrical energy. Recent improvements in motors for working with static charges and motors working by resonance, that is to say, having groups of tuned electromagnetic cooperating circuits render this possible but such do not form part of the present invention.

A motor adapted to operate with static charges will for the sake of simplicity be diagrammatically indicated by two semicircles 1 and 2 and the rotor of the motor by a ring M. (Figure 3.) A is a vertical aerial or aerial network. S the safety choke or electromagnet with coil O as may be seen is connected with the aerial A. Adjacent the electromagnet S the aerial conductor is divided into three circuits, the circuit 8 giving the safety spark gap, the circuit 7 with the working spark gap, and then a circuit including the stator terminal 1, the rotor and stator terminal 2 at which a connection is made to the earth wire. The two spark gaps are also connected metallically with the earth wire. The method of working these diagrams is as follows:

The positive atmospheric electric charge collected tends to combine with the negative electricity (or earth electricity) connected with the earth wire. It travels along the aerial A through the electromagnet S without being checked as it flows in the same direction as the direct current. Further, its progress is arrested by two sparks gaps placed in the way and the stator condenser surfaces. The stator condenser surfaces are charged until the charge is greater than the resistance of the spark gap 7, whereupon a spark springs over the spark gap 7 and an oscillatory charge is obtained as by means of the motor M, stator surfaces 1 and 2, and spark gap 7, a closed oscillation circuit is obtained for producing the electromagnetic oscillations. The motor here forms the capacity and the necessary inductance and resistance, which, as is well known, are necessary for converting static electricity into electromagnetic wave energy.

The discharges formed are converted into mechanical energy in special motors and can not reach the aerial network by reason of the electromagnet or choke. If, however, when a spark springs over the spark gap 7 a greater quantity of atmospheric electricity tends to flow to earth, a counter voltage is

induced in the electromagnet, which is greater the more rapidly and strongly the flow of current direct to the earth is. By the formation of this opposing voltage a sufficiently high resistance is offered to the flow of atmospheric electricity direct to earth to prevent a short circuit with the earth.

The circuit containing spark gap 8 having a different wave length which is not in resonance with the natural frequency of the motor, does not endanger the motor and serves as security against excess voltage, which, as practical experiments have shown, may still arise in certain cases, but can be conducted direct to earth through this spark gap.

In the diagram illustrated in Figure 4 the spark gap 7 is shunted across condensers 5 and 6 from the motor M. This construction affords mainly a better insulation of the motor against excess voltage and a uniform excitation through the spark gap 7.

In Figure 5 a diagram is illustrated for transforming large current strengths which may be employed direct without motors, for example, for lighting or heating purposes. The main difference is that here the spark gap consists of a star shaped disc 7 which can rotate on its own axis and is rotated by a motor opposite similarly fitted electrodes 7^a. When separate points of stars face one another, discharges take place, thus forming an oscillation circuit over condensers 5 and 6 and inductance 9 for oscillatory discharges. It is evident that a motor may also be directly connected to the ends of the spiral 9.

The construction of the diagram shown in Figure 6 permits of the oscillation circuit of the motor being connected with an induction coil. Here a regulating inductive resistance is introduced for counter-acting excess voltages in the motor. By cutting the separate coils 9 (coupled inductively to the aerial) in or out the inductive action on the motor may be more or less increased or variable aerial action may be exerted on the oscillation circuit.

In Figure 7 the oscillation circuit is closed through the earth (E and E_1). The spark gap 7 may be prolonged or shortened by more or fewer spark gaps being successively connected by means of a contact arm 7^b.

Diagram 8 shows a unipolar connection of the motor with the aerial network. Here two oscillation circuits are closed through the same motor. The first oscillation circuit passes from aerial A through electromagnet S, point x, inductance 9^a to the earth condenser 6 and further, over spark gap 7 to the aerial condenser 5 and back to x. The second oscillation circuit starts from the aerial condenser 5 at the point x^1 over the inductance 9 to the earth condenser 6 at the point x^3 and through the condenser 6 over the spark gap 7 back to x^1. The motor itself is inserted between the two points of the spark gap 7. From this arrangement slightly damped oscillation wave currents are produced.

In the diagram illustrated in Figure 9 a loosely coupled system of connections is illustrated which is assumed to be for small motors for measuring purposes. A indicates the aerial conductor, S the electromagnet in the aerial conductor, 9 the inductance, 7 the spark gap, 5 and 6 condensers, E the earth, M the motor, and 1 and 2 stator connections of the motor. The motor is directly metallically connected with the oscillation circuit.

In Figure 10 a purely inductive coupling is employed for the motor circuit. The motor is connected with the secondary wire 10 as may be seen in Figure 11 in a somewhat modified diagram connection. The same applies to the diagram of Figure 12.

The diagrams hitherto described preferably allow of motors of small and medium strength to be operated. For large aggregates, however, they are too inconvenient as the construction of two or more oscillation circuits for large amounts of energy is difficult; the governing is still more difficult and the danger in switching on or off is greater.

A means of overcoming such difficulties is shown in Figure 13. The oscillation circuit here runs starting from the point x over condenser 5, variable inductance 9, spark gap 7 and the two segments (3^a and 4^a) forming arms of a Wheatstone bridge, back to x. If the motor is connected by brushes 3 and 4 transversely to the two arms of the bridge as shown in the drawings, electromagnetic oscillations of equal sign are induced in the stator surfaces 1 and 2 and the motor does not revolve. If however, the brushes 3 and 4 are moved in common with the conducting wires 1 and 2 which connect the brushes with the stator poles a certain alteration or displacement of the polarity is obtained and the motor commences to revolve.

The maximum action will result if one brush 3 comes on the central sparking contact 7 and the other brush 4 on the part x. They are however, usually in practice not brought on to the central contact 7 but only held in the path of the bridge segments 4^a and 3^a in order not to connect the spark gaps with the motor oscillation circuit.

As however, the entire oscillation energy can thereby not act on the motor it is better to carry out the same system according to the diagram 14. The diagram 14 differs from the foregoing only by the motor not being directly metallically connected with the segments of the commutator, but only a primary coil 9 which induces in a secondary coil 10, current which feeds the motor M and takes the place of the rotor. By this

arrangement a good transforming action is obtained, a loose coupling and also an oscillation circuit without a spark gap.

In Figure 15 the motor is not purely inductively as in 14, but directly metallically branched off from the primary coil (at x and x^1) after the principle of the auto-transformer.

In Figure 16 instead of an inductance a condenser 6 is in similar manner, and for the same object inserted between the segments 3^a and 4^a. This has the advantage that the segments 3^a and 4^a need not be made of solid metal but may consist of spiral coils whereby a more exact regulation is possible and further motors of high inductance may be employed.

The arrangements of Figures 17, 18 and 19 may be employed for use with resonance and particularly with induction condenser motors; between the large stator induction condenser surfaces, small reversing pole condensers are connected, which, as may be seen from Figures 17, 18 and 19 are led together to earth. Such reversing poles have the advantage that with large quantities of electrical energy the spark formation between the separate oscillation circuits ceases.

Figure 19 shows a further method which prevents electromagnetic oscillations of high number of alternations formed in the oscillation circuit striking back to the aerial conductor. It is based on the well known principle that a mercury lamp, one electrode of which is formed of mercury, the other of solid metal such as steel allows an electric charge to pass in only one direction from the mercury to the steel and not vice versa. The mercury electrode of the vacuum tube N is therefore connected with the aerial conductor and the steel electrode with the oscillation circuit. From this it results that charges can pass only from the aerial through the vacuum tube to the oscillation circuit, but not vice versa. Oscillations which are formed on being transformed in the oscillation circuit cannot pass to the aerial conductor.

In practice these vacuum tubes must be connected behind an electromagnet as the latter alone affords no protection against the danger of lightning.

As regards the use of spark gaps, all arrangements as used for wireless telegraphy may be used. Of course the spark gaps in large machines must have a sufficiently large surface. In very large stations they are cooled in liquid carbonic acid or better still in liquid nitrogen or hydrogen; in most cases the cooling may also take place by means of liquefied low homologues of the metal series or by means of hydrocarbons the freezing point of which lies at between —90° C. and —40° C. The spark gap casing must also be insulated and be of sufficient strength to be able to resist any pressure which may arise. Any undesirable excess super-pressure which may be formed must be automatically let off. I have employed with very good results mercury electrodes which were frozen in liquid carbonic acid, the cooling being maintained during the operation from the outside through the walls.

Figure 20 is one of the simplest forms of construction of an aerial network in combination with collectors, transformers and the like illustrated diagrammatically. E is here the earth wire, 8 the safety spark gap, 7 the working spark gap, 1 and 2 the stator surfaces of the motor, 5 a condenser battery, S the protective magnet which is connected with the coil in the aerial conductor, A^1 to A^{10} aerial antennæ with collecting balloons, N horizontal collecting or connecting wires from which, to the centre a number of connections run.

The actual collectors consist of metal sheaths preferably made of an aluminium magnesium alloy, and are filled with hydrogen or helium and are attached to copper plated steel wires. The size of the balloon is selected so that the actual weight of the balloon and the weight of the conducting wire is supported thereby. On the top of the balloon aluminium spikes, made and gilded in a special manner hereinafter described, are arranged in order to produce a conductor action. Small quantities of radium preparations, more particularly polonium-ionium or mesothorium preparations considerably increase the ionization, and therewith the action of these collectors.

In addition to metal balloons, fabric balloons which are superficially metal coated according to Schoop's metal spraying process, may however also be employed. A metallic surface may also be produced by lacquering with metallic bronzes, preferably according to Schoop's spraying process or lacquering with metallic bronze powders in two electrical series of widely different metals, because thereby the collecting effect is considerably increased.

Instead of the ordinary round balloons, elongated cigar shaped ones may be employed. In order also to utilize the frictional energy of the wind, patches or strips of non-conducting substances which produce electricity by friction, may be attached to the metallized balloon surfaces. The wind will impart a portion of its energy in the form of frictional electricity, to the balloon casing, and thereby the collecting effect is substantially increased.

In practice however, very high towers (up to 300 metres is fully admissible) may be employed as antennæ. In these towers copper tubes rise freely further above the top of the tower. A gas lamp secured

against the wind is then lit at the point of the copper tube and a netting is secured to the copper tube over the flame of this lamp to form a collector. The gas is conveyed through the interior of the tube up to the summit. The copper tube must be absolutely protected from moisture at the place at which it enters the tower and also rain must be prevented running down the walls of the tower which might lead to a bad catastrophe. This is done by bell shaped enlargements which expand downwards, being arranged in the tower in the form of high voltage insulators of Siamese pagodas.

Special attention must be devoted to the foundations of such towers. They must be well insulated from the ground, which may be obtained by first embedding a layer of concrete in a box form to a sufficient depth in the ground and inserting in this an asphalt lining and then glass bricks cast about 1 or 2 metres in thickness. Over this in turn there is a ferro-concrete layer in which alone the metal foot of the tube is secured. This concrete block must be at least 2 metres from the ground and be fully protected at the sides by a wooden covering, from moisture. In the lower part of the tower a wood or glass house for the large condenser batteries or for the motors may be constructed. In order to lead the earth connection to the ground water, a well insulated pit constructed of vitreous bricks, must be provided. Several such towers are erected at equal distances apart and connected with a horizontal conductor. The horizontal connecting wires may either run directly from tower to tower or be carried on bell shaped insulators similar to those in use for high voltage conductors. The width of the network may be of any suitable size and the connection of the motors can take place at any suitable places.

In order to collect large quantities of electricity with few aerials it is well to provide the aerial conductor with batteries of condensers as shown in two methods of construction in Figures 21 and 22. In Figure 21 the batteries of condensers 5 are connected on the one hand with the aerial electricity collectors Z by the aerial conductor A, and on the other hand interconnected in series with an annular conductor from which horizontal conductors run to the connecting points C to which the earth wire is connected.

Figure 22 shows a similar arrangement. Should two such series of antennæ rings be shown by a voltmeter to have a large difference of potential (for example, one in the mountains and one in the plain) or even of different polarity these differences may be compensated for by connecting sufficiently large condenser batteries (5, 5^a, 5^b) by means of Maji star conductors D and D^1. In Figure 23 a connection of three such rings of collectors to form a triangle with a central condenser battery is illustrated.

The condenser batteries of such large installations must be embedded in liquefied gases or in liquids freezing at very low temperatures. In such cases a portion of the atmospheric energy must be employed for liquefying these gases. It is also preferable to employ pressure. By this means the condenser surfaces may be diminished, and still allow for large quantities of energy to be stored, secure against breakdown. For smaller installations the immersing of the condensers in well insulated oil or the like, suffices. Solid substances on the other hand cannot be employed as insulators.

The arrangement in the diagrams hitherto described was always such that the condenser batteries were connected with both poles directly to the aerial conductors. An improved diagram of the connections for obtaining atmospheric electricity for the condenser batteries has however, been found to be very advantageous, this arrangement consists in that they are connected by only one pole (unipolar) to the collecting network. Such a method of arrangement is very important, as by means of it a constant current and an increase of the normal working pressure or voltage is obtained. If for example a collecting balloon aerial which is allowed to rise to a height of 300 metres, shows 40,000 volts above earth voltage, in practice it has been found that the working voltage (with a withdrawal of the power according to the method hereinbefore described by means of oscillating spark gaps and the like) is only about 400 volts. If however, the capacity of the condenser surfaces be increased, which capacity in the above mentioned case was equal to that of the collecting surface of the balloon aerials, to double the amount, by connecting the condenser batteries with only one pole, the voltage rises under an equal withdrawal of current up to and beyond 500 volts. This can only be ascribed to the favourable action of the connecting method.

In addition to this substantial improvement it has also been found preferable to insert double inductances with electromagnets and to place the capacities preferably between two such electromagnets. It has also been found that the useful action of such condensers can be further increased if an induction coil be connected as inductive resistance to the unconnected pole of the condenser, or still better if the condenser itself be made as an induction condenser. Such a condenser may be compared with a spring which when compressed carries in itself accumulated force, which it again gives off when released. In charging, a charge with reversed sign is formed at the

other free condenser pole, and if through the spark gap a short circuit results, the accumulated energy is again given back since now new quantities of energy are induced at the condenser pole connected with the conductor network, which in fact charges with opposite signs to that at the free condenser pole. The new induced charges have of course the same sign as the collector network. The whole voltage energy in the aerial is thereby however increased. In the same space of time larger quantities of energy are accumulated than is the case without such inserted condenser batteries.

In Figures 24 and 25 two different diagrams of connections are more exactly illustrated, Figure 24 shows a collecting balloon and the diagram of the connections to earth. Figure 25 four collecting balloons and the parallel connection of the condenser batteries belonging thereto.

A is the collecting balloon made of an aluminium magnesium alloy (electron metal, magnalium) of a specific gravity of 1.8 and a thickness of plate 0.1 to 0.2 mm. Inside there are eight strong vertical ribs of T shaped section about 10 to 20 mm. in height and about 3 mm. in thickness with the projecting part directed inwards (indicated by *a*, *b*, *c*, *d* and so forth); they are riveted together to form a firm skeleton and are stiffened in a horizontal direction by two cross ribs. The ribs are further connected with one another internally and transversely by means of thin steel wires, whereby the balloon obtains great power of resistance and elasticity. Rolled plates of 0.1 to 0.2 mm. in thickness made of magnalium alloy are then either soldered or riveted on this skeleton so that a fully metallic casing with smooth external surface is obtained. Well silvered or coppered aluminium plated steel wires run from each rib to the fastening ring 2. Further, the coppered steel hawser L preferably twisted out of separate thin wires (shown in dotted lines in Figure 24) and which must be long enough to allow the balloon to rise in the desired height, leads to a metal roller or pulley 3 and from thence to a winch W, well insulated from the earth. By means of this winch, the balloon, which is filled with hydrogen, or helium, can be allowed to rise to a suitable height (300 to 5,000 metres) and brought to the ground for recharging or repairs.

The actual current is taken directly through a friction contact from the metal roller 3 or from the wire, or even from the winch or simultaneously from all three by means of brushes (3, 3^a and 3^b). Beyond the brushes the conductor is divided, the paths being:—firstly over 12 to the safety spark gap 8, from thence to the earth conductor E^1, and secondly over electromagnet S^1, point 13, to a second loose electromagnet having an adjustable coil S^2, then to the spark gap 7 and to the second earth conductor E^2. The actual working circuit is formed through the spark gap 7, condensers 5 and 6, and through the primary coil 9; here the static electricity formed by oscillatory discharges is accumulated and converted into high frequency electromagnetic oscillations. Between the electromagnets S^1 and S^2 at the crossing point 13, four condenser batteries are introduced which are only indicated diagrammatically in the drawings each by one condenser. Two of these batteries (16 and 18) are made as plate condensers and prolonged by regulating induction coils or spirals 17 and 19 while the two others (21 and 23) are induction condensers. As may be seen from the drawings each of the four condenser batteries 16, 18, 21, 23 is connected only by one pole to the aerial or to the collector conductor. The second poles 17, 19, 22, 24 are open. In the case of plate condensers having no inductive resistance an induction coil is inserted. The object of such a spiral or coil is the displacement of phase of the induction current by ¼ periods, whilst the charging current of the condenser poles which lie free in the air, works back to the collector aerial. The consequence of this is that in discharges in the collector aerial the back inductive action of the free poles allows a higher voltage to be maintained in the aerial collecting conductor than would otherwise be the case. It has also been found that such a back action has an extremely favourable effect on the wear of the contacts. Of course the inductive effect may be regulated at will within the limits of the size of the induction coil, the length of the coil in action being adjustable by means of wire connection without induction (see Fig. 24, No. 20).

S^1 and S^2 may also be provided with such regulating devices in the case of S^2 (illustrated by 11). If excess voltage be formed it is conducted to earth through the wire 12 and spark gap 8 or through any other suitable apparatus, since this formation would be dangerous for the other apparatus.

The action of these condenser batteries has already been hereinbefore described.

The small circles on the collector balloon indicate places at which zinc amalgam or gold amalgam or other photoelectric acting metals in the form of small patches in extremely thin layers (.01 to .05 mm. in thickness) are applied to the balloon casing of light metal. Such metallic patches may also be applied to the entire balloon as well as in greater thickness to the conducting network. The capacity of the collector is thereby considerably strengthened at the surface. The greatest possible effect in collecting may be obtained by polonium amalgams and the like. On the surface of the

collector balloon metal points or spikes are also fixed along the ribs, which spikes serve particularly for collecting the collector charge. Since it is well known that the resistance of the spikes is less the sharper the spike is, for this purpose it is therefore extremely important to employ as sharp spikes as possible. Experiments made as regards these have shown that the formation of the body of the spike or point also plays a large part, for example, spikes made of bars or rollers with smooth surfaces, have a many times greater point resistance as collector accumulator spikes than those with rough surfaces. Various kinds of spike bodies have been experimented with for the collector balloons hereinbefore mentioned. The best results were given by spikes which were made in the following way. Fine points made of steel, copper, nickel, or copper and nickel alloys, were fastened together in bundles and then placed as anode with the points in a suitable electrolyte (preferably in hydrochloric acid or muriate of iron solutions) and so treated with weak current at 2 to 3 volts pressure. After 2 to 3 hours according to the thickness of the spikes or pins the points become extremely sharp and the bodies of the spikes have a rough surface. The bundle can then be removed and the acid washed off with water. The spikes are then placed as cathode in a bath consisting of solution of gold, platinum, iridium, paladium or wolfram salts or their compounds and coated at the cathode galvanically with a thin layer of precious metal, which must however be sufficiently firm to protect them from atmospheric oxidation.

Such spikes act at a 20 fold lower voltage almost as well as the best and finest points made by mechanical means. Still better results are obtained if polonium or radium salts are added to the galvanic bath when forming the protective layer or coating. Such pins have a low resistance at their points and even at one volt and still lower pressures have an excellent collector action.

In Figure 24 the three unconnected poles are not connected with one another in parallel. That is quite possible in practice without altering the principle of the free pole. It is also preferable to interconnect in parallel to a common collector network, a series of collecting aerials.

Figure 25 shows a diagram for such an installation. A^1, A^2, A^3, A^4 are four metal collector balloons with gold or platinum coated spikes which are electrolytically made in the presence of polonium emanations or radium salts, which spikes or needles are connected over four electro-magnets S^1, S^2, S^3, S^4, through an annular conductor R. From this annular conductor four wires run over four further electromagnets S^a, S^b, S^c, S^d, to the connecting point 13. There the conductor is divided, one branch passing over 12 and the safety spark gap 8 to the earth at E^1, the other over inductive resistance J and working spark gap 7 to the earth at E^2. The working circuit, consisting of the condenser 5 and 6 and a resonance motor or a condenser motor M, such as hereinbefore described, is connected in proximity round the sparking gap section 7.

Instead of directly connecting the condenser motor of course the primary circuit for high frequency oscillatory current may also be inserted.

The condenser batteries are connected by one pole to the annular conductor R and can be either inductionless (16 and 18) or made as induction condensers as shown by 21 and 23. The free poles of the inductionless condensers are indicated by 17 and 19, those of the induction condensers by 22 and 24. As may be seen from the drawings all these poles 17, 22, 19, 24 may be interconnected in parallel through a second annular conductor without any fear that thereby the principle of the free pole connection will be injured. In addition to the advantages already set forth the parallel connection also allows of an equalization of the working pressure in the entire collector network. Suitably constructed and calculated induction coils 25 and 26 may also be inserted in the annular conductor of the free poles, by means of which a circuit may be formed in the secondary coils 27 and 28 which allows current produced in this annular conductor by fluctuations of the charges or the like appearances to be measured or otherwise utilized.

According to what has been hereinbefore stated separate collector balloons may be connected at equidistant stations distributed over the entire country, either connected directly with one another metallically or by means of intermediate suitably connected condenser batteries through high voltage conductors insulated from earth. The static electricity is converted through a spark gap into dynamic energy of a high number of oscillations and may in such form be coupled as a source of energy by means of a suitable method of connecting, various precautions being observed and with special regulations. The wires leading from the collector balloons have hitherto been connected through an annular conductor without this endless connection, which can be regarded as an endless induction coil, being able to exert any action on the whole conductor system.

It has now been found that if the network conductor connecting the aerial collector balloons with one another is not made as a simple annular conductor, but preferably short circuited in the form of coils over a

condenser battery or spark gap or through thermionic tubes or valves or audions, then the total collecting network exhibits quite new properties. The collection of atmospheric electricity is thereby not only increased but an alternating field may be easily produced in the collector network. Further, the atmospheric electrical forces showing themselves in the higher regions may also be directly obtained by induction. In Figures 26 and 28 a form of construction is shown on the basis of which the further foundations of the method will be more particularly explained.

In Figure 26 1, 2, 3, 4 are metal collector balloons, 5, 6, 7, 8 their metallic aerial conductors and I the actual collector network. This consists of five coils and is mounted on high voltage insulators in the air, on high voltage masts (or with a suitable construction of cable embedded in the earth). One coil has a diameter of 1 to 100 km. or more. S and S¹ are two protective electromagnets, F the second safety section against excess voltage, E its earth conductor and E¹ the earth conductor of the working section. When an absorption of static atmospheric electricity is effected through the four balloon collectors, the current in order to reach the earth connection E¹ must flow spirally through the collector network over the electromagnet S, primary induction coil 9, conductor 14, anode A of the audion tube, incandescent cathode K, as the way over the electromagnet and safety spark gap F offers considerably greater resistance. Owing to the fact that the accumulated current flows in one direction, an electromagnetic alternating field is produced in the interior of the collector network coil, whereby the whole free electrons are directed more or less into the interior of the coil. An increased ionization of the atmosphere is therefore produced. In consequence of this the points mounted on the collector balloon show a considerably reduced resistance and therefore increased static charges between the points on the balloon and the surrounding atmosphere are produced. The result of this is a considerably increased collector effect.

A second effect which could not be obtained otherwise is obtained by the electromagnetic alternating field which running parallel to the earth surface, acts more or less with a diminishing or increasing effect on the earth magnetic field, whereby in the case of fluctuations in the current a return induction current of reversed sign is always produced in the collector coil by earth magnetism. Now if, however, a constantly pulsating continuous alternating field is produced as stated in the above collector network I, an alternating current of the same periodicity is produced also in the collecting network coil. As the same alternating field is further transmitted to the aerial balloon, the resistance of its points is thereby considerably reduced, whilst the collector action is considerably increased. A further advantage is that positive electrons which collect on the metal surfaces during the conversion into dynamic current produce a so-called drop of potential of the collector area. As an alternating field is present, the negative ions surrounding the collector surfaces, when discharge of the collector surfaces takes place produce by the law of induction, an induction of reversed sign on the collector surface and so forth (that is to say again a positive charge). In addition to the advantages hereinbefore set forth, the construction of connecting conductors in coil form when of sufficiently large diameter, allows of a utilization of energy arising in higher regions also in the simplest way. As is well known electric discharges frequently take place at very great elevations which may be observed, such as St. Elmo's fires or northern lights. These energy quantities have not been able to be utilized up to now. By this invention all these kinds of energy, as they are of an electromagnetic nature and the direction of the axis of the collector coils stands at right angles to the earth's surface, can be more or less absorbed in the same way as a receiver in wireless telegraphy absorbs waves coming from a far distance. With a large diameter of the spiral it is possible to connect large surfaces and thereby to take up also large quantities of energy.

It is well known that large wireless stations in the summer months, and also in the tropics are very frequently unable to receive the signals in consequence of interruptions which are caused by atmospheric electricity, and this takes place with vertical coils of only 40 to 100 m. diameter. If on the contrary horizontal coils of 1 to 100 km. diameter be employed very strong currents may be obtained through discharges which are constantly taking place in the atmosphere. Particularly in the tropics or still better in the polar regions where the northern lights are constantly present, large quantities of energy may probably be obtained in this way. A coil with several windings should act the best. In similar manner any alteration of the earth magnetism should act inductively on such a coil.

It is not at all unlikely that earthquakes and spots on the sun will also produce an induction in such collector coils of sufficient size. In similar manner this collector conductor will react on earth currents more particularly when they are near the surface of the earth or even embedded in the earth. By combining the previous kind of current collectors so far as they are adapted for

the improved system with the improved possibilities of obtaining current the quantities of free natural energy which are to be obtained in the form of electricity are considerably increased.

In order to produce in the improved collector coil uniform current oscillations of an undamped nature so-called audion high vacuum or thermionic tubes of suitable connection are employed instead of the previously known spark gaps (Fig. 26, Nos. 9–18). The main aerial current flows through electromagnet S (which in the case of a high number of alternations is not connected here but in the earth conductor E^1) and may be conveyed over the primary coils in the induction winding through wire 14 to the anode A of the high vacuum grid tube. Parallel with the induction resistance 9 a regulating capacity of suitable size, such as condenser 11 is inserted. In the lower part of the vacuum grid tube is arranged the incandescent filament or the cathode K which is fed through a battery B. From the battery B two branches run, one to the earth conductor E^1 and the other through battery B^1 and secondary coil 10 to the grid anode *g* in the vacuum tube. By the method of connections shown in dotted lines, a desired voltage at the grid electrode *g* may also be produced through the wire 17 which is branched off from the main current conductor through switches 16 and some small condensers (*a*, *b*, *c*, *d*) connected in series, and conductor 18, without the battery B^1 being required.

The action of the entire system is somewhat as follows:—

On the connecting conductor of the aerial collector network being short circuited to earth, the condenser pole 11 is charged and slightly damped oscillations are formed in the short circuited existing oscillation circuit formed of the condenser 11 and self inductance 9. In consequence of the coupling through coil 10, fluctuations of voltage take place in the grid circuit 15 with the same frequency, which, fluctuations in turn influence the strength of the electrode current passing through the high vacuum amplyfying tube and thus produce current fluctuations of the same frequency in the anode circuit. A permanent supply of energy to the oscillation circuits 9 and 10 consequently takes place, until a condition of balance is set up, in which the consumed oscillation energy is equal to that absorbed. Thereby constant undamped oscillations are now produced in the oscillation circuits 9–11.

For regular working of such oscillation producers high vacuum strengthening tubes are necessary and it is also necessary that the grid and anode voltages shall have a phase difference of 180° so that if the grid is negatively charged, then the anode is positively charged and vice versa. This necessary difference of phase may be obtained by most varied connections, for example, by placing the oscillation circuit in the grid circuit or by separating the oscillation circuit and inductive coupling from the anodes and the grid circuit and so forth.

A second important factor in this way of converting static atmospheric electricity into undamped oscillations is that care must be taken that the grid and anode voltages have a certain relation to one another; the latter may be obtained by altering the coupling and a suitable selection of the self induction in the grid circuit, or as shown by dotted lines 18, 17, 16 by means of a larger or smaller number of condensers of suitable size connected in series; in this case the battery B^1 may be omitted. With a suitable selection of the grid potential a glow discharge takes place between the grid *g* and the anode A, and accordingly at the grid there is a cathode drop and a dark space is formed. The size of this cathode drop is influenced by the ions which are emitted in the lower space in consequence of shock ionization of the incandescent cathodes K and pass through the grid in the upper space. On the other hand the number of the ions passing through the grid is dependent on the voltage between the grid and the cathode. Thus if the grid voltage undergos periodic fluctuations (as in the present case) the amount of the cathode drop at the grid fluctuates and consequently the internal resistance of the tube correspondingly fluctuates, so that when a back coupling of the feed circuit with the grid circuit takes place, the necessary means are afforded for producing undamped oscillations and of taking current, according to requirements from the collecting conductor.

The frequency of the undamped oscillations produced is with a suitably loose coupling equal to the self frequency of the oscillation circuits 9 and 10. By a suitable selection of the self induction of the coil 9 and capacity 11 it is possible to extend from frequencies which produce electromagnetic oscillations of only a few metres wave length down to the lowest practical alternating current frequency. For large installations a suitable number of frequency producing tubes in the form of the well known high vacuum transmission tubes of .5 to 2 kw. in size may be connected in parallel so that in this respect no difficulty exists.

The use of such tubes for producing undamped oscillations, and also the construction and method of inserting such transmission tubes in an accumulator or dynamo circuit is known and also that such oscillation producing tubes only work well at voltages of 1,000 up to 4,000 volts, so that on the contrary their use at lower voltages is

considerably more difficult. By the use of high voltage static electricity this method of producing undamped oscillations as compared with that through spark gaps must be regarded as an ideal solution particularly for small installations of outputs of from 1 to 100 kw.

By the application of safety spark gaps, with interpolation of electro-magnets, not only is short circuiting avoided but also the taking up of current is regulated. Oscillation producers inserted in the above way form a constantly acting electromagnetic alternating field in the collector coil, whereby as already stated, a considerable accumulating effect takes place. The withdrawal wire or working wire is connected at 12 and 13, but current may be taken by means of a secondary coil which is firmly or movably mounted in any suitable way inside the large collector coil, i. e. in its electromagnetic alternating field, so long as the direction of its axis runs parallel with that of the main current collecting coil.

In producing undamped oscillations of a high frequency (50,000 per second and more) in the oscillation circuits 9 and 11, electromagnets S and S^1 must be inserted if the high frequency oscillations are not to penetrate the collector coil, between the oscillation producers and the collector coil. In all other cases they are connected shortly before the earthing (as in Figs. 27 and 28).

In Figure 27 a second method of construction of the connecting conductor of the balloon aerials is illustrated in the form of a coil. The main difference consists in that in addition to the connecting conductor I another annular conductor II is inserted parallel to the former on the high voltage masts in the air (or embedded as a cable in the earth) but both in the form of a coil. The connecting wire of the balloon aerials is indicated as a primary conductor and also as a current producing network; the other is the consumption network and is not in unipolar connection with the current producing network.

In Figure 27 the current producing network I is shown with three balloon collectors 1, 2, 3 and aerial conductors 4, 5, 6; it is short circuited through condenser 19 and inductance 9. The oscillation forming circuit consists in this diagram of spark gap *f*, inductance 10, and condenser 11; the earth wire E, is connected to earth over electromagnet S^1. F is the safety spark gap which is also connected to earth through a second electromagnet S at E. On connecting up the condenser circuit 11 this is charged over the spark gap *f* whereby an oscillatory discharge is formed. This discharging current acts through inductance 10 on the inductively coupled secondary 9, whereby in the producing network a modification of the potential of the condenser 19 is produced. The consequence of this is that oscillations arise in the coil shaped producer network. These oscillations induce a current in the secondary circuit II, which has a smaller number of windings and a less resistance, the voltage of which, according to the proportion of the number of windings and of the ohmic resistance, is considerably lower whilst the current strength is greater.

In order to convert the current thus obtained into current of an undamped character, and to tune its wave lengths, a sufficiently large regulatable capacity 20 is inserted between the ends 12 and 13 of the secondary conductor II. Here also current may be taken without an earth conductor, but it is advisable to insert a safety spark gap E^1 and to connect this with the earth over an electromagnet S^2.

The producer network may be connected with the working network II over an inductionless condenser 21 or over an induction condenser 22, 23. In this case the secondary conductor is unipolarly connected with the energy conductor.

In Figure 28 the connecting conductor between the separate accumulator balloons is carried out according to the autotransformer principle. The collecting coil connects four aerial balloons 1, 2, 3, 4, the windings of which are not made side by side but one above the other. In Figure 28 the collector coil I is shown with a thin line, the metallically connected prolongation coils II with a thick line. Between the ends I^1 and II^1 of the energy network I a regulating capacity 19 is inserted. The wire I^1 is connected with the output wire and with the spark gap F.

As transformer of the atmospheric electricity an arrangement is employed which consists in using rotary pairs of condensers in which the one stator surface B is connected with the main current, whilst the other A is connected with the earth pole. Between these pairs of short circuited condensers are caused to rotate from which the converted current can be taken by means of two collector rings and brushes, in the form of an alternating current, the frequency of which is dependent on the number of balloons and the revolutions of the rotor. As the alternating current formed in the rotor can act, in this improved method of connection described in this invention, through coils 10 on the inductance 9, an increase or diminution of the feed current in I can be obtained according to the direction of the current by back induction. Current oscillations of uniform rhythm thereby result in the coil shaped windings of the producer net-work.

As the ends of this conductor are short cir-

cuited through the regulatable condenser 19 these rhythms produce short circuited undamped oscillations in the energy conductor, the periodicity and wave lengths of which oscillations can be adjusted according to desire by altering the capacity 19 to a given wave length and therewith also to a given frequency. These currents may also be employed in this form directly as working current through the conductors II^1 and III. By inserting the condenser 20 a connection between these conductors may also be made, whereby harmonic oscillations of desired wave length are formed. By this means quite new effects as regards current distribution are obtained. The withdrawal of current can even take place without direct wire connection if, at a suitable point in the interior of the producing network (quite immaterially whether this has a diameter of 1 or 100 km.) a coil tuned to these wave lengths and of the desired capacity is firmly or movably mounted in the aerial conductor in such a way that its axial direction is in parallel with that of the collector coil. In this case a current is induced in the producing network, the size of which is dependent on the total capacity and resistance and also on the periodicity employed. A possibility is thereby afforded in future, of taking energy from the producer network by wireless means. As thereby in addition to atmospheric electricity also magnetic earth currents and the energy from the higher atmosphere (at least partially) may be simultaneously obtained, this last system for collecting the atmospheric energy is of particular importance for the future.

Of course everywhere instead of spark gaps suitable grid vacuum tubes may be employed as producers for undamped oscillations. The separate coils of the producer net-work with large diameters may be connected with one another through separate conductors all in parallel or all in series or in groups in series. By regulating the number of oscillations and also the extent of the voltage more or less large collector coils of this kind may be employed. The coils may also be divided spirally over the entire section. The coils may be carried out in annular form or also in triangular, quadrangular, hexagonal or octagonal form.

Of course wires may be carried from a suitable place to the centre or also laterally which serve the current waves as guides. This is necessary when the currents have to be conducted over mountains and valleys and so forth. In all these cases the current must be converted into a current of suitable periodicity.

As already hereinbefore mentioned separate collecting balloons may be directly metallically interconnected at equidistant stations distributed over the entire country or may be connected by interpolation of suitable condenser batteries by means of high voltage conductors. The static electricity is converted through a spark gap into dynamic energy of a high number of oscillations, and could then in such form, with a suitable arrangement of the connections, observing various measures of precaution, be employed as source of energy after separate or special regulation.

According to this invention in order to increase the collecting effect of the balloon in the aerial collector conductor or in the earth wire, radiating collectors are employed. These consist either of incandescent metal or oxide electrodes in the form of vacuum grid tubes, or electric arcs (mercury and the like electrodes) Nernst lamps, or finally flames of various kinds may be simply connected with the respective conductor.

It is well known that energy can be drawn off from a cathode consisting of an incandescent body opposite an anode charged with positive electricity (vacuum grid tube). Hitherto however, a cathode was always first directly placed opposite an anode, and secondly the system always consisted of a closed circuit.

Now if we dispense with the ordinary ideas in forming light or flame arcs in which a cathode must always stand directly opposite an anode, and if we place an incandescent cathode opposite an anode charged to a high potential or another body freely floating in the air, or regard the incandescent cathode only as a source of unipolar discharge (which represent group and point discharges in electro-static machines similar to unipolar discharges), it may be ascertained that incandescent cathodes and less perfectly all incandescent radiators, flames and the like admit of relatively large current densities and allow large quantities of electric energy to radiate into the open space in the form of electron streams as transmitters.

The object of this invention is as described below, if such incandescent oxide electrodes or other incandescent radiators or flames are not freely suspended in space but connected metallically with the earth so that they can be charged with negative terrestrial electricity, these radiators possess the property of absorbing the free positive electrical charges contained in the air space surrounding them (that is to say of collecting them and conducting them to earth). They can therefore, serve as collectors and have, in comparison to the action of the spikes, or points, a very large radius of action R; the effective capacity of these collectors is much greater than the geometrical capacity (R_0–) calculated in an electro-static sense.

Now as our earth is surrounded as is well

known, with an electro-static field and the difference of potential

$$\frac{\delta V}{\delta h}$$

of the earth field according to the latest investigations, is in summer about 60 to 100 volts and in winter 300 to 500 volts per metre of difference in height (δh), a simple calculation gives the result that when such a radiation collector or flame collector is arranged for example on the ground, and a second one is mounted vertically over it at a distance of 2,000 metres and both are connected by a conducting cable, there is a difference of potential in summer of about 2,000,000 volts and in winter even of 6,000,000 volts and more.

According to Stefan Boltzmann's law of radiation, the quantity of energy which an incandescent surface (temperature T) of 1 sq. cm. radiates in a unit of time into the open air (temperature T_o) is expressed by the following formula:

$$S=\sigma\ (T^4-T^4_o)\ \text{watt/sq. cm.}$$

and the universal radiation constant σ is according to the latest researches of Ferry (Annales de Chimie et de physique 17 page 267 (1909)) equal to 6.30×10^{-12} watt/sq. cm.

Now if an incandescent surface of 1 sq. cm. shows, as compared with the surrounding space a periodic fall of potential δV it radiates (independent of the current direction, that is to say of the sign) in accordance with the above formula, for example at a temperature of 3725° C. an energy of 1.6 kw. per sq. cm. per second. As for the radiation the same value can be calculated for the collection of energy, but reversed. Now as carbon electrodes at the temperature of the electric arc support on the current basis a current density up to from 60 to 65 amperes per sq. cm. no difficulties will result in this direction in employing radiating collectors as accumulators.

If the earth be regarded as a cosmically insulated condenser in the sense of geometrical electro-statics *x* there results from the geometric (compare Ewald Rasch, "das elektrische Bogenlicht" (The electric arc light) page 169) capacity of the earth according to Chwolson:

For negative charging 1.3×10^6 Coulomb

For negative potential $V=10\times10^8$ volts. From this there results however, $EJT\cong24.7\times10^{24}$ watt/Sec. Now if it is desired to make a theoretic short circuit through an earthed flame collector this would represent an electric total work of about 79,500 10^{10} kilowatt years. As the earth must be regarded as a rotating mechanism which is thermo-dynamically, electromagnetically, and also kinematically coupled with the sun and stars system by cosmic radiations and gravitation a diminution of the electric energy of the earth field is not to be feared. The energies which the incandescent collectors would withdraw from the earth field can only cause by the withdrawal of motor work a lowering of the earth temperature (temperature $T_E=300$) and reduce this to that of the world space ($T=O$) by using the entire energy. This is however not the case as the earth does not represent a cosmically entirely insulated system. On the contrary there is conveyed to the same according to the recent value corrected by Ferry for the solar constants through the radiation from the sun an energy of $18{,}500\times10^{10}$ kw. Accordingly any lowering of the earth temperature (T_E) without a simultaneous lowering of the sun's temperature (T_S) would contradict Stefan Boltzmann's law of radiation.

$$S=\sigma\,(T_S^4-T_E^4).$$

From this it must be concluded that if the earth temperature (T_E) sinks, the total radiation S absorbed by the earth increases, and further also that the secular speed of cooling of the earth is directly dependent on that of the sun and the other radiators cosmically coupled with the sun and is connected most closely with these.

The incandescent radiation collectors may, according to this invention, be employed for collecting atmospheric electricity if they (1) are charged with the negative earth electricity (that is to say when they are directly connected by means of a metallic conductor with the earth) and (2) if large capacities (metal surfaces) charged with electricity are mounted opposite them as positive poles in the air. This is regarded as the main feature of the present invention as without these inventive ideas it would not be possible to collect with an incandescent collector, sufficiently large quantities of the electrical charges contained in the atmosphere as technology requires; the radius of action of the flame collectors would also be too small, especially if it be considered that the very small surface density (energy density) (σ about$=2\times7\ .\ 10^9$ St. E. per sq. cm.) does not allow of large quantities of charge being absorbed from the atmosphere.

x) Calculated according to Poisson's calculation;

$\Delta V=-4\pi\delta$; as here the alteration of the potential or potential gradients only takes place in the direction of the normal, this calculation assumes the simple form

$$\delta=\frac{1}{4\pi}\times\frac{\delta^2V}{\delta n^2}$$

It has indeed already been proposed to employ flame collectors for collecting atmospheric electricity and it is known that

their collecting effect is substantially greater opposite the points. It is however, not known that the quantities of current which could hitherto be obtained are too small for technical purposes. According to my experiments the reason for this is to be found in the too small capacities of the collector conductor poles. If such flame or radiating collectors have no or only small positive surfaces, their radius of action for large technical purposes is too small. If the incandescent collectors be constantly kept in movement in the air they may collect more according to the speed of the movement, but this is again not capable of being carried out in practice.

By this invention the collector effect is considerably increased by a body charged with a positive potential and of the best possible capacity being also held floating (without direct earth connection) opposite such an incandescent collector which is held floating in the air at a desired height. If for example, a collecting balloon of sheet metal or of metalized balloon fabric be caused to mount to 300 up to 3,000 metres in the air and as positive pole it is brought opposite such a radiating collector connected by a conductor to earth, quite different results are obtained.

The metallic balloon shell (with a large surface) is charged to a high potential by the atmospheric electricity. This potential is greater the higher the collecting balloon is above the incandescent collector. The positive electricity acts concentratedly on the anode floating in the air as it is attracted through the radiation shock ionization, proceeding from the incandescent cathode. The consequence of this is that the radius of action of the incandescent cathode collector is considerably increased and thereby also the collecting effect of the collecting balloon surface. Further the large capacity of the anode floating in the air plays therefore an important part because it allows of the taking of large charges, and thereby a more uniform current is obtained even when there is a large consumption: this cannot be the case with small surfaces.

In the present case the metallic collecting balloon is a positive anode floating in the air and the end of the earth conductor of this balloon serves as positive pole surface opposite the surface of the radiating incandescent cathode, which in turn is charged with negative earth electricity being conductingly connected to earth.

The process may be carried out by two such contacts (negative incandescent cathode and anode end of a capacity floating in the air) a condenser and an inductive resistance being switched on in parallel, whereby simultaneously undamped oscillations may be formed.

In very large installations it is advisable to connect two such radiating collectors in series. Thus an arc light incandescent cathode may be placed below on the open ground and an incandescent cathode which is heated by special electro-magnetic currents be located high in the air. Of course for this the special vacuum Liebig tubes with or without grids may also be employed. An ordinary arc lamp with oxide electrodes may be introduced on the ground and the positive pole is not directly connected with the collecting balloon, but through the upper incandescent cathode or over a condenser. The method of connecting the incandescent cathode floating in the air may be seen in Figs. 29–33.

B is the air balloon, K a Cardan ring (connection with the hawser) C the balloon, L a good conducting cable, P a positive pole, N negative incandescent cathode, and E earth conductor.

Fig. 29 represents the simplest form of construction. If electric oscillations are produced below on the ground by means of a carbon arc lamp or in other suitable way a considerably greater electric resistance is opposed to that in the direct way by inserting an electrical inductive resistance 9. Consequently between P and N a voltage is formed, and as, over N and P only an inductionless ohmic resistance is present, a spark will spring over so long as the separate induction co-efficients and the like are correctly calculated. The consequence of this is that the oxide electrode (carbon or the like) is rendered incandescent and then shows as incandescent cathode an increased collecting effect. The positive poles must be substantially larger than the negative in order that they may not also become incandescent. As they are further connected with the large balloon area which has a large capacity and is charged at high voltage, an incandescent body which is held floating in the air and a positive pole which can collect large capacities is thereby obtained in the simplest way. The incandescent cathode is first caused to become incandescent by means of separate energy produced on the earth, and then maintained by the energy collected from the atmosphere.

Fig. 30 only shows the difference that instead of a round balloon a cigar shaped one (of metal or metalized fabric) may be employed and also a condenser 5 is inserted between the incandescent cathode and the earth conductor so that a short circuited oscillation circuit over P. N 5 and 9 is obtained. This has the advantage that quite small quantities of electricity cause the cathode to become incandescent and much larger cathode bodies may be rendered incandescent.

In this form of construction both the in-

candescent cathode and also the positive electrode may be enclosed in a vacuum chamber as may be seen in Fig. 32. A cable L is carried well insulated through the cover of a vessel and ends in a condenser disc 5. The cover is arched in order to keep off the rain. The vessel is entirely or partially made of magnetic metal and well insulated inside and outside. Opposite the disc 5 another disc 6 and on this again a metallic positive pole of the vacuum tube *g* with the incandescent cathode (oxide electrode) N is arranged. The negative electrode is on the one hand connected with the earth conductor E, and on the other hand with the inductive resistance 9 which is also connected with the cable L with the positive pole and wound round the vessel in coils. The action is exactly the same as that in Fig. 29 only instead of an open incandescent cathode one enclosed in vacuo is employed. As in such collectors only small bodies can be brought to incandescence in large installations a plurality of such vacuum tubes must be inserted in proximity to one another. According to the previous constructions Figs. 31 and 33 are quite self evident without further explanations.

Figs. 34–37 represent further diagrams of connections over radiating and flame collectors, and in fact, how they are to be arranged on the ground.

Fig. 34 shows an arc light collector with oxide electrodes for direct current and its connection; Fig. 35 a similar one for alternating current, Fig. 36 an incandescent collector with a Nernst lamp and Fig. 37 a similar one with a gas flame.

The positive pole 1 of the radiating collectors is always directly connected to the aerial collecting conductor A. In Fig. 34 this is further connected over the condenser battery 5 with a second positive electrode 3. The direct current dynamo *b* produces current which flows over between the electrodes 3 and 2 as an arc light. On the formation of an arc the negative incandescent electrode 2 absorbs electricity from the positive poles standing opposite it and highly charged with atmospheric electricity and conveys the same to the working circuit. The spark gap 7, inductive resistance 9 and induction coil 10 are like the ones previously described. The protective electromagnet S guards the installation against earth circuiting, the safety spark gap 8 from excess voltage or overcharging.

In Fig. 35 the connection is so far altered that the alternating current dynamo feeds the exciting coil 11 of the induction condenser. 12 is its negative and 13 its positive pole; if the coil 3 on the magnet core of the dynamo is correctly calculated and the periodicity of the alternating current is sufficiently high an arc light can be formed between the two poles 1 and 2. As the cathode 2 is connected with the negatively charged earth, and therefore always acts as a negative pole, a form of rectification of the alternating current produced by the dynamo 3 is obtained, the second half of the period is always suppressed. The working circuit may be carried out in the same way as in Fig. 34; the working spark gap 7 may however be dispensed with, and instead thereof between the points *n* and *m* a condenser 5 and an induction resistance 9 may be inserted from which the current is taken inductively.

Fig. 36 represents a form of construction similar to Fig. 34 only that here instead of an arc lamp a Nernst incandescent body is employed. The Nernst lamp is fed through the battery 3. The working section is connected with the negative pole, the safety spark gap with the + poles. The working spark gap 7 may also be dispensed with and the current for it taken at 12 over the oscillation circuit 5, 11 (shown in dotted lines).

Flame collectors (Fig. 37) may also be employed according to this invention. The wire network 1 is connected with the aerial collector conductor A and the burner with the earth. At the upper end of the latter, long points are provided which project into the flame. The positive electrode is connected with the negative over a condenser 5 and the induction coil 9 with the earth.

The novelty in this invention is firstly, the use of incandescent cathodes opposite positive poles which are connected with large metallic capacities as automatic collecting surfaces, (2) the connection of the incandescent cathodes with the earth whereby, in addition to the electricity conveyed to them from the battery or machine which causes the incandescing, also the negative charge of the earth potential is conveyed, and (3) the connection of the positive and negative poles of the radiating collectors over a condenser circuit alone or with the introduction of a suitable inductive resistance, whereby simultaneously an oscillatory oscillation circuit may be obtained. The collecting effect is by these methods quite considerably increased.

I declare that what I claim is:—

1. An electrical energy generating system, comprising a conducting surface for static charges, means to support same at a distance above the earth, a conductor leading to the earth level, a spark gap associated with said conductor to convert electrostatic charges into electromagnetic high frequency oscillations means to supply said electromagnetic energy to a net work, and a spark gap of greatly increased relative resistance in parallel therewith.

2. An electrical energy generating system

comprising a conductor, means to support same above the earth level, an inductance therein, a spark gap associated with said conductor, a second spark gap of much higher relative resistance in parallel therewith and an energy receiving circuit coupled with the spark gap of lesser resistance.

3. An electrical energy generating system comprising a collecting surface, means to support same above the earth level, a conductor connecting said collecting surface with the earth level, a choke in said conductor, an electromagnetic resistance converting electrostatic energy to electromagnetic energy, a safety higher resistance in parallel therewith and a net work coupled with the conversion resistance of lesser value.

4. An electrical energy generating system comprising electric conductors spaced above the earth to form electromagnetic oscillating circuits, conductors connecting to earth level, electrostatic to electromagnetic energy conversion means therein, a safety high electrostatic resistance in parallel therewith and means to alter the electromagnetic characteristic of the circuits.

5. An electrical energy generating system comprising in combination a static collecting surface arranged above the earth, conductors connecting to earth level, a pair of spark gaps in parallel of different electrostatic resistance, a utilization net work shunted across the spark gap of lesser resistance and an electromagnetic choke in said conductors.

6. An electrical energy generating system comprising an open circuit energy collecting aerial, a pair of sparking gaps in parallel of widely different resistance, connected thereto and a closed electric oscillation circuit in shunt across the gap of lesser resistance.

7. An electrical energy generating system comprising an open circuit energy collecting aerial, a pair of sparking gaps in parallel of widely different resistance connected thereto, a closed electric oscillation circuit in shunt across the gap of lesser resistance, a plurality of electrostatic collecting surfaces, means to connect said collecting surfaces in parallel in groups and means to connect said groups symmetrically with said aerial.

In witness whereof, I have hereunto signed my name this 30 day of Dec., 1920, in the presence of two subscribing witnesses.

HERMANN PLAUSON.

Witnesses:
H. F. Armstrong,
W. H. Beeston.

W. I. PENNOCK.

APPARATUS FOR COLLECTING ELECTRICAL ENERGY.

APPLICATION FILED JAN. 4, 1911.

1,014,719. Patented Jan. 16, 1912.

2 SHEETS—SHEET 1.

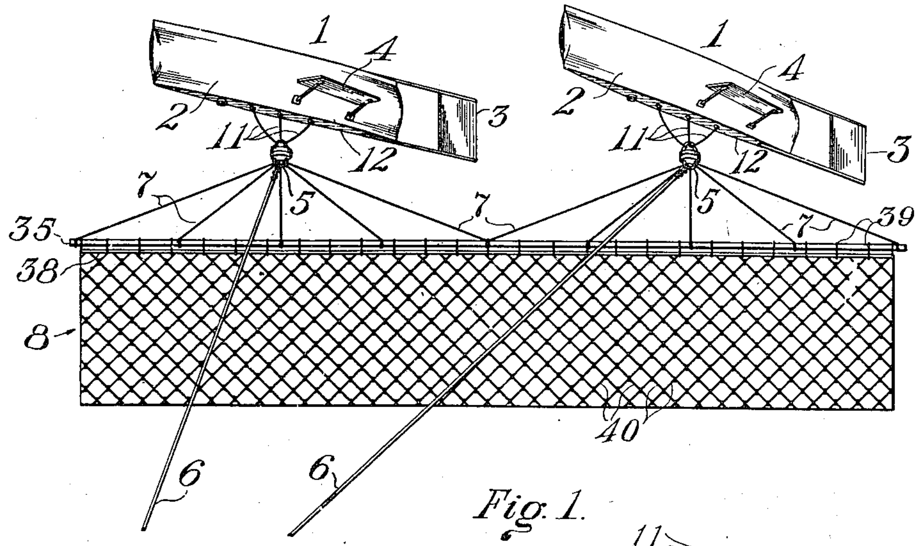
1
4
2
3
11
12
5
7
35
39
38
8
40
6

Fig. 1.

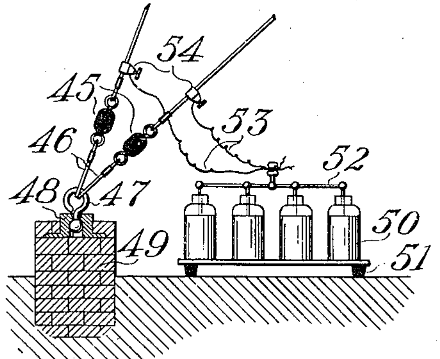
54
45
53
46
52
48
47
50
49
51

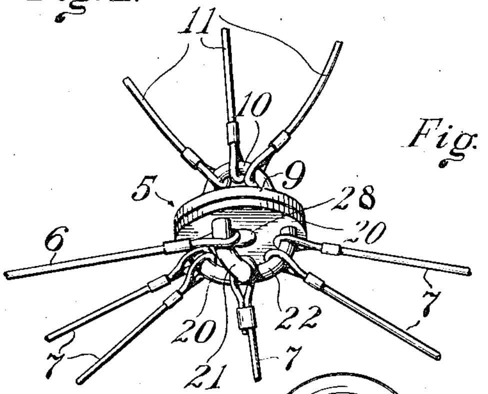
11
10
9
5
28
20
6
7
20
22
21
7

Fig. 3.

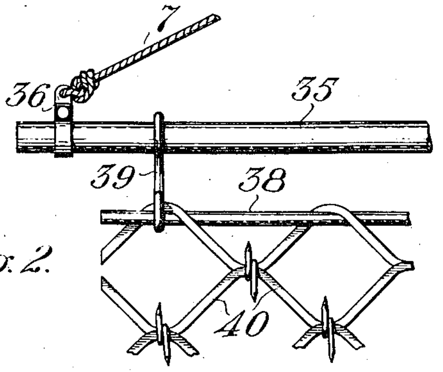
7
36
35
39
38
40

Fig. 2.

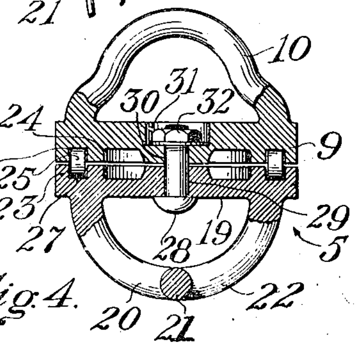
10
31
30
32
24
9
26
25
23
29
27
19
28
5
20
21
22

Fig. 4.

INVENTOR

Walter I. Pennock.

WITNESSES

Elizabeth L. Smith

J. Stuart Freeman.

BY

ATTORNEY

W. I. PENNOCK.

APPARATUS FOR COLLECTING ELECTRICAL ENERGY.

APPLICATION FILED JAN. 4, 1911.

1,014,719. Patented Jan. 16, 1912.

2 SHEETS—SHEET 2.

Fig. 5.

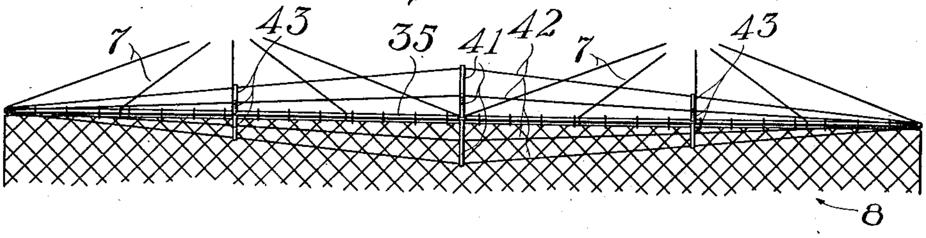
7
43
35
41
42
7
43
8

Fig. 6.

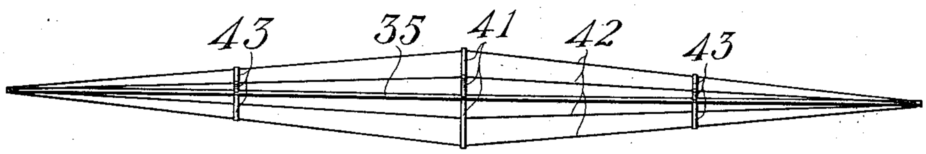
43
35
41
42
43

Fig. 7.

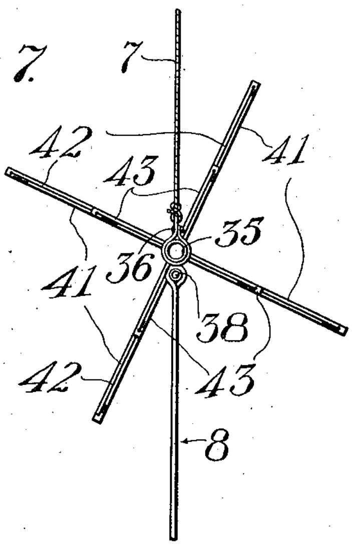
7
42
41
43
35
41
36
38
42
43
8

INVENTOR

Walter I. Pennock.

WITNESSES

Elizabeth L. Smith,

J. Stuart Freeman.

BY

ATTORNEY

UNITED STATES PATENT OFFICE.

WALTER I. PENNOCK, OF PHILADELPHIA, PENNSYLVANIA.

APPARATUS FOR COLLECTING ELECTRICAL ENERGY.

1,014,719. Specification of Letters Patent. Patented Jan. 16, 1912.

Application filed January 4, 1911. Serial No. 600,777.

To all whom it may concern:

Be it known that I, Walter I. Pennock, a citizen of the United States, residing at Philadelphia, county of Philadelphia, and State of Pennsylvania, have invented certain new and useful Improvements in Apparatus for Collecting Electrical Energy, of which the following is a full, clear, and exact disclosure.

The present invention relates to an improved means for collecting the charges of electricity from the upper atmosphere and more particularly to that form which consists in one or more captive balloons from which is suspended a suitable form of metallic conductor.

The principal objects of the device are: to provide a collector for atmospheric and static electricity, which when in operative position will present a large surface to currents of air, but which will offer comparatively little resistance thereto, to provide a collector of such material and construction as will be more efficient in its operation than any previously constructed for a similar purpose, to provide a means for maintaining such a collector suspended in the air and at right angles to opposing currents thereof, and to provide a suitable anchorage for holding said means captive.

With these principal objects in view, the present invention consists in further advantages which are brought out in the following specification and accompanying drawings, in both of which like numerals refer to like parts, and in which drawings—

Figure 1 is a perspective view of the complete device in operative position, Fig. 2 is an enlarged detail view of the wire mesh and the manner in which it is attached to the supporting balloons, Fig. 3 is a detail of the manner of securing the collector-supporting and anchor cables to the balloons, Fig. 4 is an enlarged cross section of the swivel connection shown in Fig. 3, Fig. 5 is top plan view of the reinforcing braces on the screen, Fig. 6 is an elevation of the same and Fig. 7 is an end view of the structure shown in Figs. 5 and 6.

Referring to the drawings, in Fig. 1 thereof, a plurality of balloons 1 of any suitable type is shown, each of which embodies hollow metallic elongated gas tanks 2, extending from the rear of which are single, rigidly affixed rudders 3, while on the sides of the tanks are secured stationary lifting planes 4.

To the bottom and slightly to the rear of the center of the tanks 2 is secured a suitable swivel 5, by which the anchor ropes 6 and the suspension ropes 7 for the metallic conductor 8 are secured to the balloons 1. A suitable form of swivel joint is illustrated in Figs. 3 and 4, but any type can be used that embodies the essential features shown therein.

The swivel joint illustrated consists in the base plate 9 having a looped portion 10 integral therewith and projecting from the upper face thereof. Secured to the loop 10 is a set of three light electrically conductive supporting ropes or cables 11 which extend upwardly and are secured at intervals to the bottom 12 of the balloon above.

The lower or revoluble member 19 of the swivel joint preferably comprises three upwardly directed curved arms, 20, 21, and 22, respectively, forming at their junction a T-shape as shown, said arms at their upper extremities being integral with the plate 23. The member 19 is revoluble below and concentric with the plate 9, and the two members are lightened in weight by opposed concentric grooves as shown at 24. Contact between said members is made through the roller bearings 25, which are carried in the opposed concentric grooves 26 and 27 of the respective upper and lower plates. Furthermore, said plates are maintained in coöperative relation with each other by means of a bolt 28 passing through centrally drilled holes 29 and 30 in the respective lower and upper plates, the drilled hole 30 opening upwardly into an enlarged recess 31, in which is sunken the nut 32 on the bolt 28. Two of the arms 20 and 22 of the lower revoluble member 19 extend in diametrical alinement, while from the central point thereof extends the third arm 21 at right angles thereto, and upward to the plate 23.

Suspended below the plurality of balloons is a hollow rod 35, of any suitable material, connected at regular intervals such as at points 36, by metallic ropes 7, to the alined arms 20 and 22 of the swivel joint 5. Below and parallel to the rod 35 is a similar but smaller rod 38 suspended therefrom by means of suitable couplings 39. From the rod 38 hangs a wire mesh 40 of any suitable form, such as can be extended over a considerable area and which will offer comparatively little resistance to the passage of air therethrough. The metal in such a screen is preferably rough, sharp or jagged and a convenient form embodying these characteristics is expanded sheet metal such as is used for lathing and for reinforcing concrete construction. Instead of the single hollow rod 35 alone, applicant contemplates the use of reinforcing means to prevent the buckling of the rod in a stiff wind, such, for instance, as the arrangement of wire bracing shown in Figs. 5, 6 and 7 in which cross rods 41, secured to the rod 35 at the center thereof are attached to the ends thereof by means of stretched wires 42. Between the rods 41 and the ends of the rod 35 are secured sets of cross rods 43 which support the wires 42 and strengthen the rod 35 at as many points as they may be placed.

For the purpose of maintaining the balloons and apparatus suspended therefrom captive, the light metallic cables 6 terminating downwardly in insulators 45 are employed. These insulators in turn are connected by means of short sections of rope 46 to the eye 47 of a suitable swivel 48, embedded in the anchorage 49.

When the balloons with the metallic screen suspended therefrom are allowed to rise into one of the higher altitudes, the entire apparatus being of metallic construction and uninsulated will become energized by contact with the surrounding natural charges of electricity. From the above description it is evident then that, while the screen 8, on account of its great extent, will be the greatest collecting agent, it will be seen that the balloons themselves and the suspending wires will also coöperate as one large collector, since no parts of which are insulated from any of the neighboring parts thereof. Consequently, when the apparatus described has reached an altitude or strata of the atmosphere abounding in static charges of electricity, an amount of the said charges proportionate to the surface area of the metal exposed will collect upon the apparatus as a whole and will be conducted downwardly toward the earth by means of the various anchor ropes 6, but will not pass into the ground on account of the interposition of the insulators 45.

To use the electrical charges thus acquired, a plurality of Leyden jars 50, or other suitable collectors are supported above the surface of the earth and insulated therefrom by any suitable means as represented by the blocks 51. Either the inner or outer conducting surfaces of the jars may be connected together and energized by the accumulated charge. In the present instance, the inner surfaces of the accumulators are shown to be connected, and the connecting means 52 is in turn connected to the ropes 6 by means of wires 53. These wires are secured to the ropes mentioned by means of suitable binding posts 54.

In the device described the anchor ropes 6 are of substantially the same length, and when the apparatus is raised to the desired altitude and is being blown by the currents of air, the balloons are turned by means of the vanes 3 to parallel relation with each other, and furthermore, from the manner in which the device is held captive and the collecting net 40 is suspended from the balloons, it is obvious that said net will at all times readily swing into a plane substantially perpendicular to any current of air acting upon the balloons above. The purpose of the swivel joint shown in Figs. 3 and 4 is principally for allowing the balloons to readily aline themselves with any new direction of the wind before the apparatus, including the suspended screen, can swing about the swivel 48, and said joint will also prevent the twisting of the ropes 7, when any rapid shifting of the air currents may occur.

While applicant has shown a set of Leyden jars as the accumulators in the accompanying drawings, it is obvious that any other suitable form may be used, and furthermore, that although not illustrated, any suitable apparatus may be run thereby, such as for instance, wireless telegraphic instruments.

Furthermore, although but one embodiment of the invention has been described, it is to be understood that various modifications may be made therein, and in fact several are contemplated by applicant that are of such structure as fall well within the scope of the appended claim.

Having thus described my invention, what I claim and desire to protect by Letters Patent of the United States, is:

A collector for charges of electricity, comprising a plurality of supporting means, a metallic gauze sustained thereby to lie in a plane and substantially equidistant from each of said supporting means, anchoring means emanating from a common point to each of said supporting means, means operative to maintain said supporting means in their normal positions, and means between each of said supporting means and said gauze to permit each of said supporting means to readily and independently aline itself to accord with any alteration in the direction of opposing air currents.

In witness whereof I have hereunto set my hand this 28 day of December, A. D. 1910.

WALTER I. PENNOCK.

Witnesses:
MILDRED S. TEMPLE,
E. EUGENIA PENNOCK.

Иоффе Александр Юльевич: другие произведения.

Transmission of electricity on one conductor to an isolated object

Журнал "Самиздат": [Регистрация] [Найти] [Рейтинги] [Обсуждения] [Новинки] [Обзоры] [Помощь]

Реклама: Новинки на КНИГОМАН Реклама: Прода на ZELLULOZA

LingvoAnalyse

- Оставить комментарий
- © Copyright Иоффе Александр Юльевич (ioffe.shura@Yandex.ru)
- Размещен: 17/04/2015, изменен: 20/08/2017. 90k. Статистика.
- Статья: Foreign+Translat
- Иллюстрации/приложения: 20 шт.
- Скачать FB2

Ваша оценка: не читал ▼ OK

- **Аннотация:**

In this article possibility of transfer of electric energy and/or information to the object isolated from Earth or from other carrying-out body on one conductor, and short history of this question is discussed. Some electric circuits of such electricity transmission are given and it is defined to what physical mechanism (a single-wire electricity transmission), according to the author, corresponds. Attempt of calculation of the transferred power and relative loss of energy is made. The author gives possible ways of increase in power and reduction of losses, and also protection against an adverse effect of such electricity transmission on a human body. In article there is a description of the conducted experiments. Also in article it is possible to read what, according to the author, are possible areas of practical application, advantage and shortcomings of transfer of the electric power on one conductor. The author apologizes, that not all declared chapters are published. Subject very extensive, in essence, rather new area of electrical equipment while there is a work, he decided to take out that on court of readers that is already ready.

TABLE OF CONTENTS.

- Foreword.
- Chapter 1: The mechanism of single-wire transmission.
- Chapter 2. Calculation of the transmitted power.
- Chapter 3. * Efficiency single-wire transmission and possible protection against radiation.
- Chapter 4 Ways of increasing the transmission power.
- Chapter 5. A third method of single-wire power transmission
- Chapter 6. * Our experiments.
- Chapter 7. Possible uses a single-wire power.
- Literature.

* Chapters not ready (at work)

CHAPTER 1. MECHANISM OF A SINGLE-WIRE ELECTRICITY TRANSMISSION.

The first experiments on transmission of the electric power on one conductor were made at a dawn of formation of the theory of electricity at the end of the 18th century. It is possible to find the description of experience in which the charged full-sphere (transmitter) connects to the uncharged full-sphere (receiver) one conductor in any solid textbook of physics. On the uncharged full-sphere charges flow, it becomes the owner of electric potential and consequently also energies. But except as contribution to the theory of potentials and condensers, for development of the theory of a single-wire electricity transmission and practical use this experience then brought nothing.

In 1899 Nikola Tesla in laboratory about the town of Colorado Springs made brilliant experiment on transfer of the electric power on one conductor, he used Earth as such conductor. The lamps thrust to Earth were lit. On my understanding, he made it using, in some measure, the above described experience. Results were reflected in article in the 'Century Magazine', magazine under Robert Johnson's edition.

But, unfortunately, demonstrations of experiences are the limit, futher affairs did not progress. And it occurred, owing to circumstances, that not depend in him. Further this Titan put a problem more widely. He tried to find a way, and probably found, how to transfer the electric power to any given square of space in general without wires by means of resonance "swing" of an ionosphere. Made successful experiments in Uordenkliff. Therefore, he is a father of wireless transfer of the electric power which now too goes uphill.

Somewhere around 1981 I, then a young engineer working in one of the Soviet research institutes, in connection with the above experience from the textbook, came up with this idea. And what if you take 4 sphere and one conductor? See Figure 1.

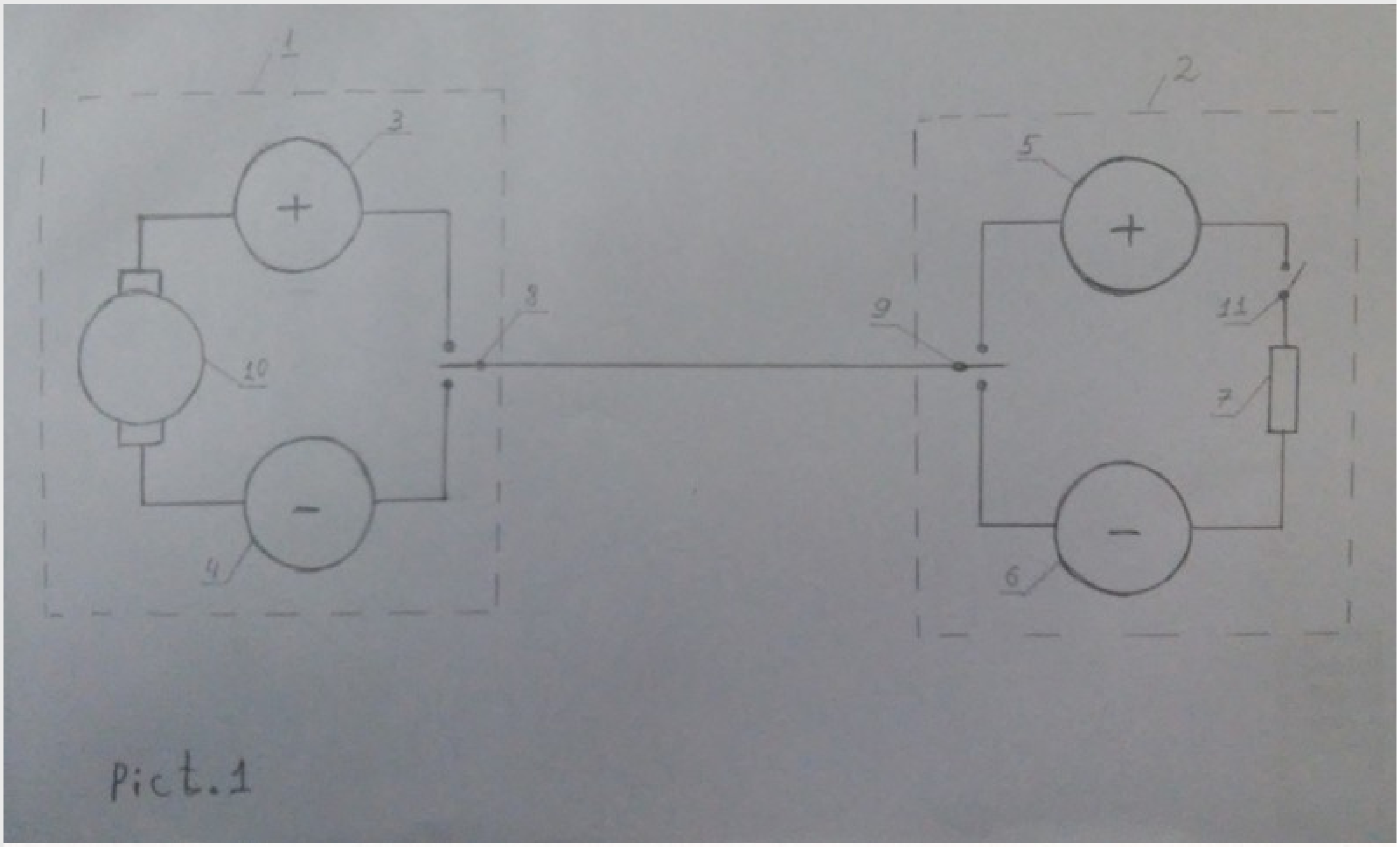
1
2
3
5
+
+
10
8
9
11
7
4
6
-
-

Pict.1

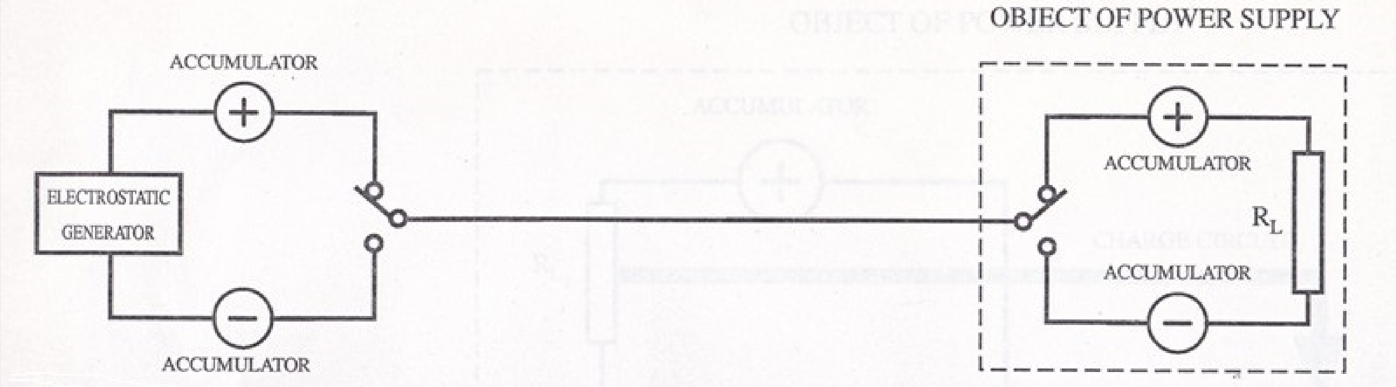
OBJECT OF POWER SUPPLY
ACCUMULATOR
ELECTROSTATIC GENERATOR
ACCUMULATOR
ACCUMULATOR
ACCUMULATOR
RL

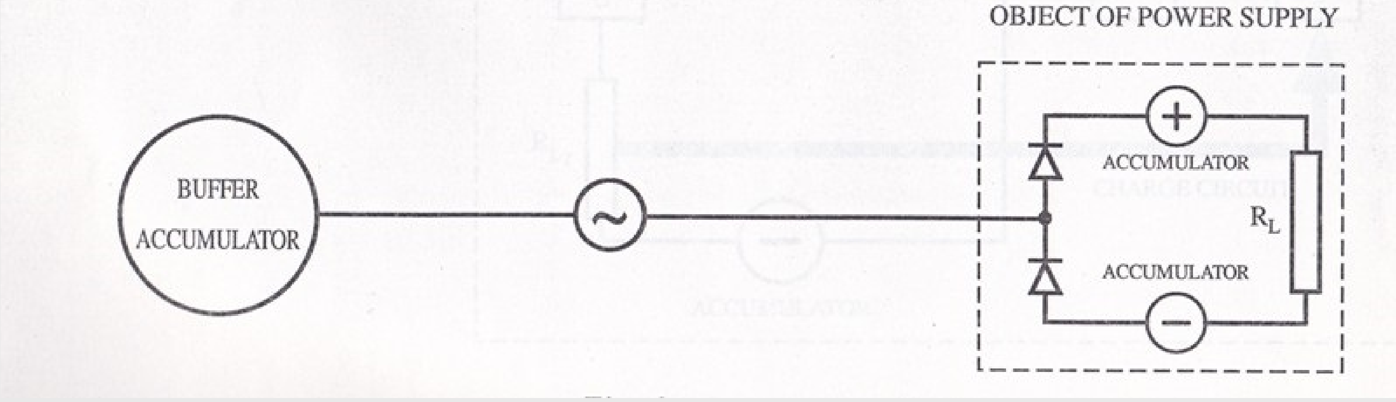
OBJECT OF POWER SUPPLY
BUFFER ACCUMULATOR
~
ACCUMULATOR
ACCUMULATOR
RL

Two spheres (5 and 6) won't be loaded yet and to be on the receiver (2), and two loaded (3 and 4) one plus charges, and the second the minus - to belong to the transmitter (1). Then, working cyclically with switches (8 and 9), it is possible to connect in the first half of a cycle the same conductor the first loaded sphere (3) with not loaded (5), and in the second - the second loaded sphere (4) with another not loaded (6). Thus, at the end of a cycle we will receive two loaded spheres on the receiver, and charges can be removed from them through the payload (7), having involved the switch (11). We will feed transmitter spheres with charges all the time, having connected them, for example, to the battery, the generator of a direct current (10) or an influence machine. It is the first way of single-wire transfer. Subsequently, in operating time over this subject, also the second way, and simpler appeared. See Fig. 2.

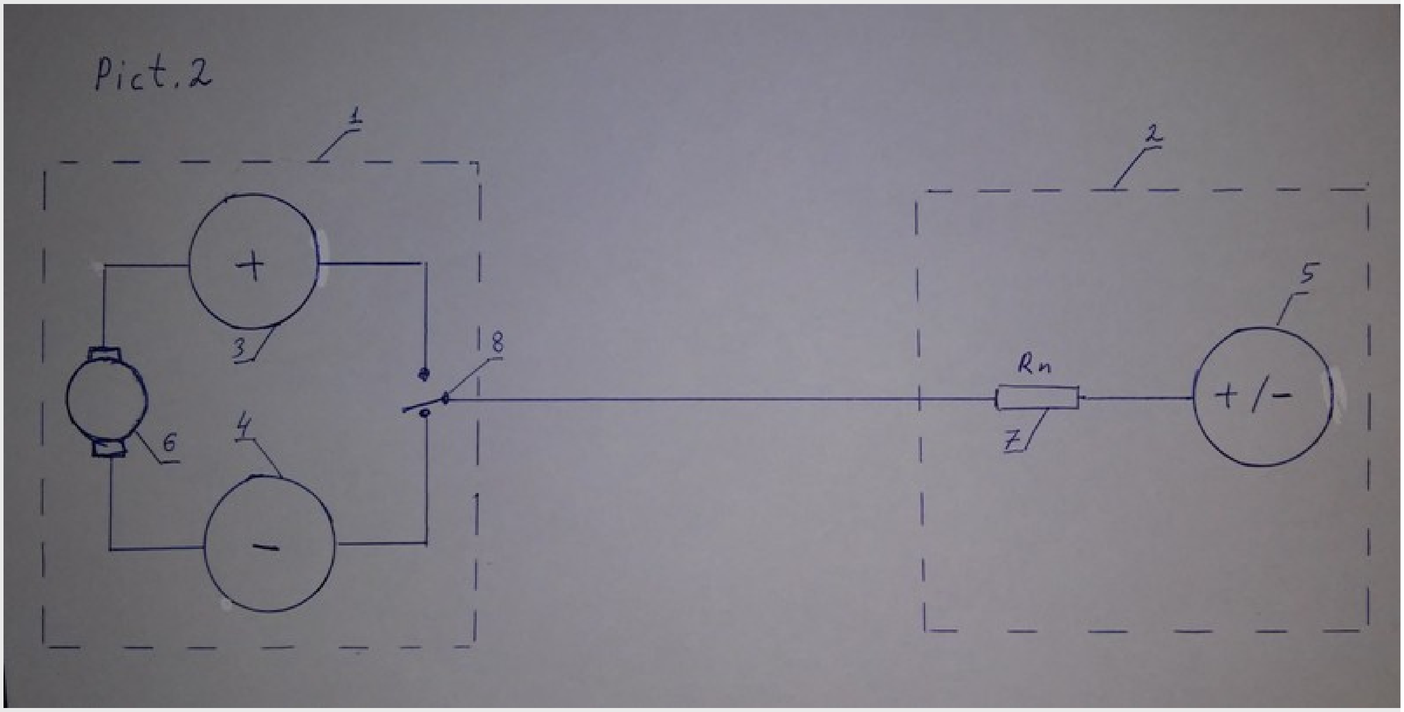

On the receiver (2) it is possible to leave only one store of electric charges(5) (in the text we will call spheres - stores of electric charges or it is simple - stores especially as stores aren't obliged to be a spherical form), but (7) we will place payload between it and the conductor going from the transmitter (1) now. Working with the switch (8), as well as at the first way, we connect alternately stores of the transmitter (3 and 4) to one conductor going to the receiver. The receiver store in each half-cycle of a cycle will receive charges of a different sign. And passing charges, every time a way from source stores to the receiver store, will make useful work, passing as well through load. Transmitter stores are recharged from the generator of a direct current (6) all the time. The continuity of transfer of current is provided to these. From the further text of article it will be clear that for the second way as load it is best of all to use, a winding of the transformer or an electric motor/electromagnet with the rectifier as thermal work won't be made. But if we want to receive, nevertheless, on the receiver heat on resistance, this resistance should be connected to a secondary winding of the transformer, and primary as it was told above, then to use as load.

Subsequently, for increase of overall performance and convenience, the scheme 1 and 2 underwent big changes. The transmitter in both schemes began to be the following device*.See Fig.3.

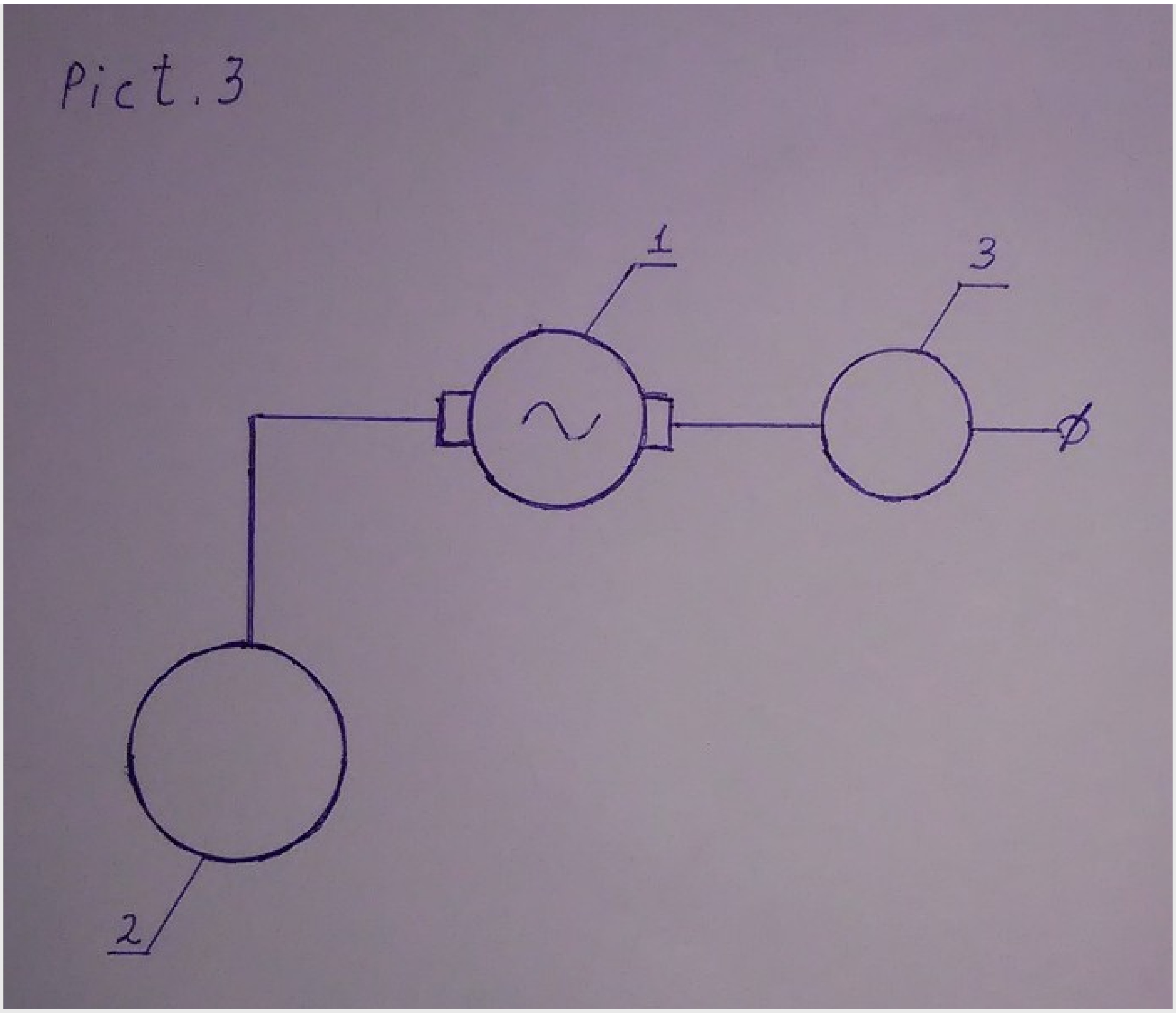

It is necessary to emphasize that here the generator of alternating current (1), but not constant already works. The store (3) transmitters from which charges for the conductor going to object of current supply not necessarily are removed has to possess big capacity, if receipt of charges from the generator continuously. In this case it can represent simply thin conductor. In case of faltering giving of charges from a source, its capacity has to be such that for a half-cycle not to allow essentially to fall to potential on it at "fence" of charges through the conductor on the receiver and interruptions in giving of charges from the generator. At such option its capacity and respectively the geometrical sizes, have to be at least 10 times more receiver store parameters. If interruptions small - that and the sizes modest. And here other store of the transmitter called in this article "buffer" (2), connected to its other conclusion even during the continuous operation of the generator has to possess rather big capacity and corresponding the big geometrical sizes. Its capacity has to be such that it at the set transfer of current potential (generator parameter) during a half-cycle could store so many charges of an opposite sign transferred, how many perceives the receiver store. That is in any way it isn't less than a capacity and according to the geometrical sizes of the store of the receiver. If it is more, it is only better. On Earth, a very good fit Earth itself - the spherical capacitor with huge geometric dimensions. That is, the second terminal of the generator, if conditions permit, the easiest way to ground **. The switch (9) in the first scheme (see Fig. 1) it was convenient to replace with a two diode switch 'diode fork' (4) (see Fig. 4).

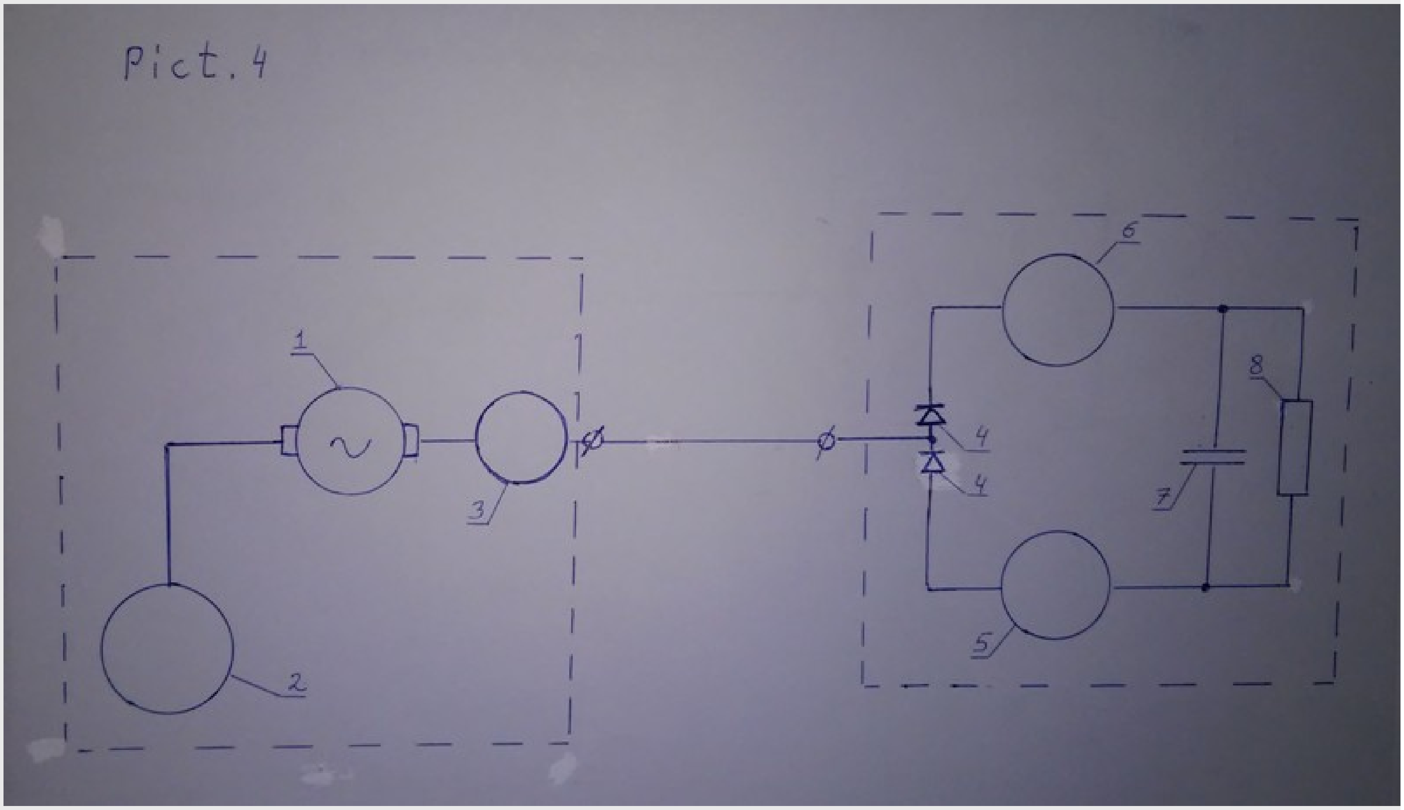

The switch (11) (see Fig. 1) in the same scheme was not necessary.

So, we will consider the modified schemes. The scheme for the first way of an electricity transmission is submitted in Fig.4, and for the second in Fig.5a. In the first way, for the smoothing proceeding through loading (8) (see Fig.4) current, it is useful to connect the condenser (7) (see Fig. 4) parallel to this load. And as for the second way if we want that through load current of one direction began to flow, it is possible to use the diode bridge (6) (see Fig. 5b) and, besides, for smoothing to connect parallel to load (4) (see Fig.5b) the condenser (7) (see Fig.5b). This scheme is submitted in Fig.5b.

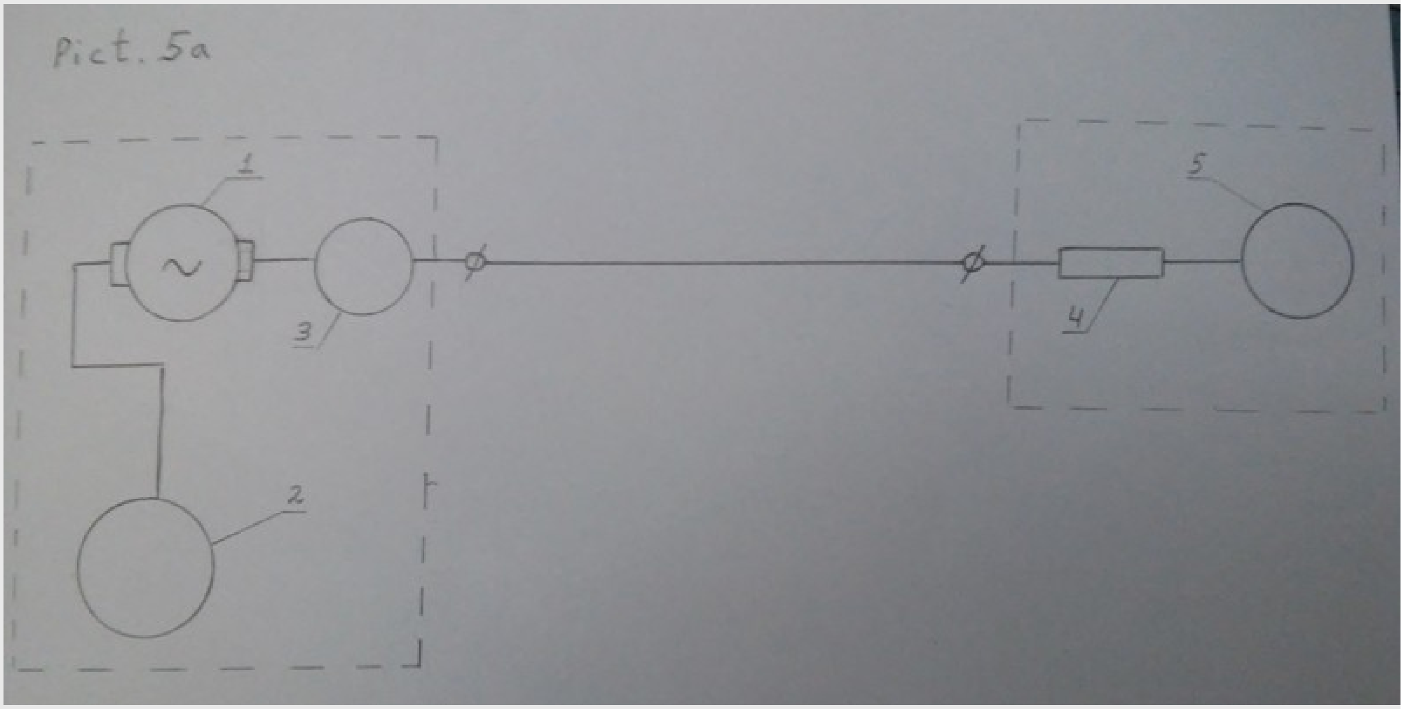

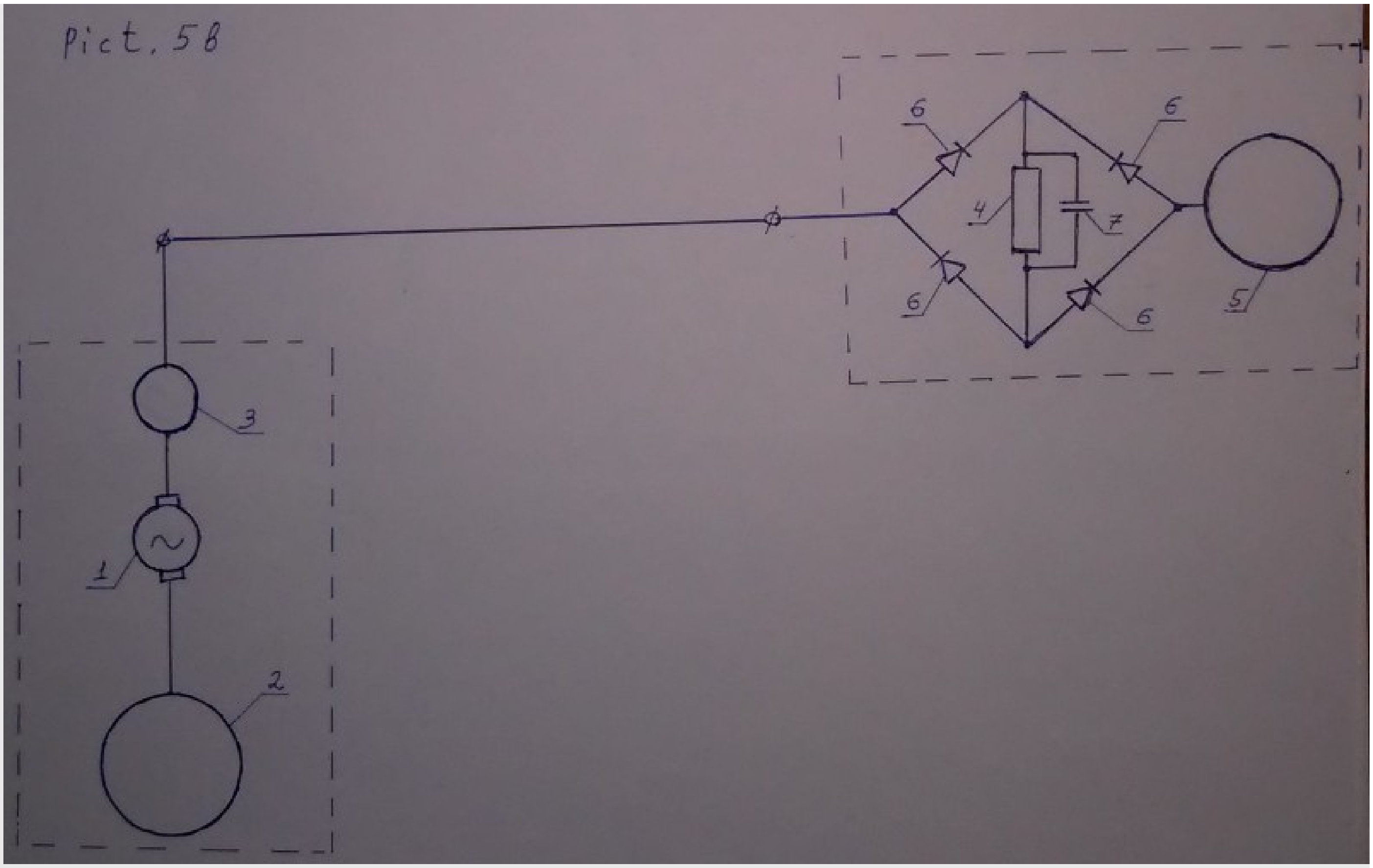

It is necessary to tell that as work on this subject didn't include in the plan of works of my scientific research institute, and I saw off it as a hobby, it moved ahead extremely slowly. During the researches I often went on consultation to Candidate of Physical Sciences Kirsanov Boris Petrovich and he made a big contribution to development of schemes and theories. In 1981/3 years I with Kirsanov B.P. submitted applications for inventions [5,6](photocopies at the end of the article). But experts of VNIIGPE didn't grant to us certificates, having referred to the low transferred power in our pilot unit executed according to the scheme of Fig. 4. But we experimental installation collected for the only purpose - to prove course of current in load, possibility of that was denied by experts absolutely at the first stage of ours with them communication. Then (in 1989) I with assistance of Kirsanov B. P. wrote article to the Inventor and Rationalizer magazine [1](photocopy at the end of the article). Soon after publication of my article it became clear that in parallel with us this subject is developed also by other people.

Group of inventors under the leadership of S. V. Avramenko under the auspices of Moscow Scientifically - Research Electrotechnical Institute collected and experienced installation on transfer of the electric power on one conductor [2,3]. This installation is very similar on thought up by us with Kirsanovy B. P. and published in the magazine [1]. In this article this installation is represented in Pict.4. They claimed that they made it before date of issue of our article. And that submitted the article to the same magazine, but didn't publish it. I won't argue with it. I will tell only, as at us before the publication installations too were collected and applications for inventions are submitted. I do not know how things are going with the priority, but that they are in the works, after our meeting and reading my article about something we have not mentioned, I think it ugly

As for their works, I think that they placed accents not there. Too great value was attached to a two diode switch 'diode fork' though the same results can be received and without it with synchronous switches. They worked (their installation) at very high frequency. Subsequently, from the text of this article it will be visible that it reduces efficiency of transfer and in general is harmful. But they collected very impressive pilot unit , I myself personally observed - it was possible to light the glow lamp.

Great merit of this group of people I see that they first drew attention to the fact that the single-wire transmission of electricity resistance of a conductor does not affect the transmit power. Furthermore, the transfer of electricity and do not affect the small gaps in the conductor cm. [4]. And I do not understand why these inventors have not submitted an application for discovery in the field of physics. As far as the physical process, it can be assumed that we are faced with a different mechanism of passing an electric current through a conductor. The charges are on the outer contour of the conductor, seeking to redistribute equipotentially throughout, even very distant surface. They do not go into the conductor and not get hit by the vibrating atoms of its crystal lattice. Nothing however retards the electrons and the conductor resistance is negligible regardless of its thickness and the material from which it is made. In experiments carried out in the writing of this article, such a property is supported by a series of experiments ... In addition, if the free electrons are facing a narrow "abyss" - a break in the conductor, the repulsion "accumulates" neighboring electrons helps them to "jump" through the "cleft" and "land" to "sparsely" surface.

As for later works, Kosinov N. V. article is interesting [2]. I consider this work among works of other authors the most advanced, it is possible to see the scheme very similar on presented on Pict.5b. Here for the first time, if to speak about works of other authors, in an explicit form there is a receiver store called for some reason "antenna". However, at such big frequency at which as I suspect, experiments and about which they for some reason "modestly are silent" were made, most likely it already radio frequencies, such name arises by itself. And that is "still interesting", about its (store) of a role in this work, for some reason, it is told nothing. And in general, looking through literature on this subject I see that as stores connecting wires, and authors of works, it seems, even serve don't guess their such function. Stores with small own capacity, like wires also will be suitable for high frequencies, but whether here it is worth working with high frequencies is a big question.

NOTES.

* For the elementary experiments as the transmitter it is possible to use one conclusion of the socket of the household power supply network, i.e. to insert one conductor into the right opening of the socket (it is a phase). But, as in a household network low frequency, current will go very small and it is necessary "to arm" for its registration with the microampermeter. It is possible to connect instead of the microampermeter the condenser / (plus the rectifier) and to load it some minutes then to pull together its connections wires and to observe a spark.

** Just the household network is grounded or connected to big electric capacitance.

CHAPTER 2. CALCULATION OF THE TRANSFERRED POWER.

Calculation is made for low frequencies of an electricity transmission. For high there will be differences. The main is that the geometrical sizes of stores of the receiver will cease to influence the accepted power as these conductive bodies won't manage to be filled for a half-cycle with charges. It is confirmed in work[10].

CALCULATION OF POWER FOR THE FIRST WAY OF A SINGLE-WIRE ELECTRICITY TRANSMISSION.

We will take an interval of time (t) equal to a half-cycle (t=T/2) of the frequency (nu) of switchings of the direction of a current of charges at a single-wire electricity transmission. This time throughout which the store of the receiver is connected by the conductor to the transmitter store possessing at this stage charges of some one sign, or "+" or "-". And these charges are redistributed so that to form an equipotential surface, i.e. charge the receiver store until electropotential on it isn't equal to transmitter store electropotential.

Stores, in all schemes provided in article are condensers which one facing is a surface of their physical body, and other- infinity. Therefore, according to a well-known physical formula for the condenser, we can write that the charged store possesses energy (W1), calculated on a formula (2.1).

(2.1) W1 = C x Phi squared/2

Where C - the electric capacity, Phi - potential.

While capacitor charging time (tau) is defined by (2.2).

(2.2) Tau = r x C, where r - resistance of the conductors going to the condenser.

As it was stated above, according to a series of experiments of N. (the same can be read and in works of other authors [2, 3]) follows that at a single-wire electricity transmission resistance of the conductors going to the store is insignificant a little. With what physical process it was possible to explain it, it is specified in Chapter 1. Therefore, it is insignificant a little and time of charging of the store (condenser). According to it, it is possible to fill a surface of the store which is on the receiver almost instantly. But what follows for calculation of the transferred power from here? And that we can reduce somehow time of a half-cycle (t) considered by us at calculations follows from this and respectively increase the frequency of an electricity transmission (nu), after all they are connected by a formula (2.3):

(2.3) nu = 1/T = 1/2t, or t = 1/2nu

the store of a receiver will always manage to be loaded in time = t.

The basic formula for the power (P1) says:

(2.4) P1 = W1/t Reducing (t) we increase the transferred power. We will substitute in a formula (4) expression for (t) from a formula (2.3). We will receive:

(2.5) P1 = 2 x nu x W1

So it is possible to draw a conclusion. The transferred power is directly proportional to the frequency of an electricity transmission (nu). Now question. To what level we can increase frequency? Restrictions "from above" following. At increase of frequency losses on the electromagnetic radiation of a wire grow, and they grow in proportion to not the first degree of frequency, as the power received by the receiver, and the second. We will deal with this issue in more detail later when we consider electricity transmission K.P.D. in Chapter 3. Besides, the high frequency and respectively strong electromagnetic radiation, can negatively affect in the medical plan the population of areas where the similar electricity transmission will be carried out. Too high frequency will be included into area of radio frequencies and will create hindrances of broadcasting and a special radio communication. Well and technical capabilities of the transmitter of the electric power, especially powerful, have some restrictions at creation of too high frequency.

Now we will consider other aspect. Influence of the geometrical sizes of the store , on the power received by this receiver. The best, in respect of the maximum electric capacitance at the minimum size, a form of the single store is a form of a sphere. After all at such form charges don't flow down from acute angles. But also other bodies with the carrying-out surface will accumulate charges. In calculations, nevertheless, as they have the general character, we will accept that we have a store of a spherical form. According to a well-known formula, the capacity of the spherical lonely condenser (C) it is equal:

(2.6) C = 4 x Pi x epsilon zero x R

Where epsilon zero - electric constant, and R - sphere radius. We will substitute this expression in a formula (1). We will receive:

(2.7) W1 = 4 x Pi x epsilon zero x R x Phi a square / 2, or after reduction: W = 2 x Pi x epsilon zero x R x Phi square

We will unite formulas (2.5) and (2.7) now and we will find expression for power.

(2.8) P1 = 4 x Pi x epsilon zero x nu x R x Phi square

We see that the transferred power P is proportional to the geometrical sizes of the store of the receiver (to sphere radius - R).

We will consider one more factor - electricity transmission potential. From the received formula (2.8) it is visible that the power of an electricity transmission P is proportional to the potential of Phi on the receiver store in the second degree. Now restrictions on potential. They are connected with safety of the people dealing with such electricity transmission and with loss of energy on the "crown" and leakage of electrons that we will consider later in Chapter 3. Later we will consider intake option where high potentials are only in its isolated part, and the current transfer conductor intended for it are under the usual not alternating (plus / minus) life-threatening potentials. Here it should be noted that though Nikola Tesla also opened that current of high frequency isn't dangerous to the person, if frequency is higher than 700 Hz in case of any potential, but there are sudden interruptions in electrical power supply, and also switching on and switch-offs. The pulses resulting from it can be very dangerous. Therefore, I believe ,that insulation is all the same necessary even if we will decide to work at high frequency.

One more aspect. The formula (2.8) is fair for vacuum and is only approximate for the air environment (where dielectric permeability is very close to unit). After all the formula (2.2) on which we base calculation is written for a lonely sphere in vacuum. If we surround the receiver store with a cover, the formula for power will be already another. Store capacity,and trasmited power, will raise if to surround it with the isolating cover with a big dielectric permeability (epsilon). We will remember a formula for the spherical condenser:

(2.9) C = (4 x Pi x an epsilon x an epsilon zero) / (1/a - 1/b),

where a - the radius of the internal sphere, b - the radius of the external sphere. Here, for example, a ferroelectric material has coefficient of dielectric permeability about 10 in the fourth degree. But it doesn't mean that power will increase in the corresponding number of times. It would be fair only for the two-superficial condenser where all space between facings is filled with dielectric. In our case, one facing is a surface of the store, and the second - infinity, and we can't fill all infinite space with dielectric. Therefore also dependence will be weaker.

Series of experiments of N... shows that the cover from material with big dielectric permeability positively influences amount of the energy transferred to the store. These researches aren't finished yet and within my laboratory conditions hardly in general will be qualitatively executed. But so far all of us can accept that the transferred power depends on thickness of a cover of the store of the receiver and on dielectric permeability of material of which it is made. We will reflect it in a new formula.

(2.10) P1 = K1 x f (d, epsilon) x nu x R x Phi a square

Where:

K1 - coefficient , approximately equal (as not all stores of receivers will be the correct spherical form) 4 x Pi x epsilon zero. K1 is approximately equal = 1.11 x 10 to the minus tenth degree.

f - function, with parameters: d - receiver store cover thickness, an epsilon - dielectric permeability of material of which the cover is made. f > 1.

So, we "will sound" a final formula. Power of single-wire transfer grows in proportion to a potential square on the line, transmission frequency and to the sizes of stores in the first degree and it is influenced positively by the isolating cover of stores (the more thickness and dielectric permeability of material, the better).

It was calculation for the first way of a single-wire electricity transmission. We will consider as affairs with a power for the second way are.

CALCULATION OF POWER FOR THE SECOND WAY OF A SINGLE-WIRE ELECTRICITY TRANSMISSION.

We will define the energy received by object for a half of the period of fluctuations of potential on a power line. It is natural that all this energy is spent in the load located between a transmission line and the store. Work (A) of movement charges (Q) in electric field is defined by a formula:

(2.11) A = Q x deltaPhi ,

where deltaPhi - a potential difference of the store of object and the line. To start any half-cycle of transfer, on the line and on the store there will be potentials of a different sign, each of which on the module is equal to Phi. Potential difference

deltaPhi = Phi-(-Phi) =2Phi.

But, in process of leakage of charges from the store and their passing through the load, potential on the store, we will call it (Phi2) changes therefore deltaPhi is function of quantity of charges on the store. We will write down it so:

deltaPhi = F(Q)

Potential on the line at the movement of charges won't change, as annihilation of charges which now on the line, from revenues to the line of charges of other sign, all the time is compensated by giving to the line charges from the generator.

From the point of view of power, process of passing of electrons through load can be divided into two stages. The first stage where work (A1) will be made - this movement of all charges from the receiver store (we will assume negative charges) through the load to a power line. This process happens to a total disappearance of charges of this sign on the store and, according to it, transformation of its potential into the zero. The second stage where work (A2) - this passing through load of charges from the line will be made (in our example positive) to the receiver store before establishment of potential on it (positive) equal to the potential on the line. Thus, the general work or energy (that in this case the same) which is obtained for a half of the period on the load:

(2.12) W2 = A1 + A2

Despite the fact that we are talking here about the two stages of the process, in fact, in terms of movement of electrons - it is continuous, and the electrons will flow in one direction only. It's just a matter of determining the sign of the charge.

Let's try to calculate the work A1.

Elementary work (its differential) will be equal to the multiplication of the potential difference deltaPhi on the elementary charge (differential Q).

(2.13) dA1 = deltaPhi (Q) x dQ = F(Q) x dQ

As already mentioned above, the potential difference is a function of the charge on the store F(Q). Let's see how this function can be expressed. With the leakage of charges through the load, it decreases. The potential of the store (Phi2) is determined by the following formula.

(2.14) Phi2 = Q / C, then

(2.15) F(Q) = Phi - (-Fi2) = Phi + Fi2 =Phi + Q / C

Substituting this expression into (2.13), we obtain:

(2.16) dA1 = (Phi + Q / C) x dQ

In order to define work A1, we have to take the integral of the expression within the following limits: lower = 0, and the top, equal to the maximum (initial) charge that is Qmax = C x Phi. After the corresponding calculations which are excessive for bringing here, we will receive:

(2.17) A1 = Phi square x C + Phi square x C / 2

Now let's calculate the A2.

Work or energy of movement of charges, as well as for A1, is determined by a formula (2.11). The potential difference between the line and the receiver store (deltaPhi) is also a function of the charge on this store. Let's see how in this case is defined F (Q). During all process, including at its beginning, potential on the line = Phi. Phi2 store potential is equal in the beginning 0, but in process of receipt on it charges which pass a way from the line through load, potential grows. Phi2 is defined by a formula (2.14). Its final value will be Phi. For a potential difference or F(Q) the following formula has to be fair.

(2.18) F(Q) = Phi - Phi2 = Phi - Q/C

Elementary work (dA2) is defined by a formula (2.13). We will substitute in it the expression which is just received by us for a potential difference, i.e. a formula (2.18). We will receive:

(2.19) dA2 = (Phi - Q/C) x dQ

In order to define work A2, we have to take the integral of the expression within the following limits: lower = 0, and the top, equal to the maximum (final) charge that is Qmax = C x Phi. After the corresponding calculations which are excessive for bringing here, we will receive:

(2.20) A1 = Phi square x C - Phi square x C / 2

To find the general work or energy emitted on load for a half of the period of an electricity transmission (W2) we substitute A1 and A2 values in a formula (2.12). We will receive:

(2.21) W2 = Phi square x C + Phi square x C/2 + Phi square x C - Phi square x C/2 = 2 x Phi square x C

We will compare this formula to a formula (2.1) for the first way and we will see that the energy for a half period received by the object of a current supply at the second way grew by 4 times. We can write down.

(2.22) W2 = 4 x W1

As all the rest in calculation of power remains the same, it is possible to write down that the power (P2) too grew by 4 times. Besides, at the second way we receive "bonus" - it isn't necessary to have two stores on the receiver, only one and the same size.

(2.23) P2 = 4 x P1

We can write for the second method, the formula similar to the formula (2.10), only take into account the increase in the power coefficient. K2 = K1 x 4 or approximately equal to 4.44 x 10 in -10 degrees. So, power for the second way is expressed by a formula:

(2.24) P2 = K2 x f (d, epsilon) x nu x R x Phi square

CHAPTER 3. EFFICIENCY OF SINGLE-WIRE TRANSMISSION AND POSSIBLE PROTECTION AGAINST RADIATION.

Efficiency of power transmisson can be defined as follows:
Efficiency = (received power / power line input) x 100%
We introduce the notation:
Pr - received power of the receiver.
Pl - losses in the power transmission process.
Power line input= Pr + Pl
Now the expression for the efficiency can be rewritten as follows:
(3.1) Efficiency = (Pr / (Pr + Pl)) x 100%
For ease of calculation, let's divide the numerator and denominator of this formula on the Pr.
(3.2) Efficiency = (1 / (1 + Pl / Pr)) x 100%
Let's denote the ratio Pl / Pr as the alfa,
alfa = Pl / Pr ,
Then formula (3.2) is as follows:
(3.3) Efficiency = (1 / (1 + alfa)) x 100%

We see that change of all function completely depends on changes of a ratio of Pl / Pr= alfa. For ensure of high efficiency, "alfa" shall aim to 0, then efficiency will be hundred percent. How to calculate the "Pr" (the power received by the object receiver) we set in chapters 2 and 5, it is possible to look at formulas (2.10), (2.24) and (5.26). As for "Pl", here we can be guided by the following provisions.

a) Losses on corona discharge and electron leakage. Because of the relatively low voltage on the line (we have introduced this limit for safety reasons), the corona discharge is not possible, and electron leak is very small, and these losses cannot be taken into account.

b) Heating the wire losses - are virtually absent, since it has no resistance to the motion of electric charges. This has been repeatedly described above.

c) Losses on the mechanical oscillation of the conductor. (Repulsion from the magnetic field of the Earth). These losses are due to their insignificance cannot be taken into account (after all, small currents).

g) Losses from radiation of electromagnetic waves.

This is the main power losses in this method of transmission.

Every conductor with alternating current radiates electromagnetic waves, and with increasing frequency energy emitted waves increases sharply (in proportion to the square of the frequency). Electromagnetic waves permanently move away from the wire, and therefore the energy consumption of the radiation waves are the transmission system losses. The radiated energy depends on the amplitude and frequency of current as well as the length of the conductor. The total average power "Nav" emitted by segment conductor with current (short antenna), in which the length "L" is much smaller than the wavelength, is equal to [12]:

(3.4) Nav = Mu zero x L square x Omega in a square x I in a square / (12 x Pi x C) where
Mu zero - a magnetic constant
Mu zero = 4 x Pi x 10 in minus 7 degrees
Pi - constant = 3.14
Omega - frequency
I - maximum current
C - light velocity.

Substituting in this formula the values of the coefficients and the speed of light, and taking into account that Nav = Pl , and the frequency in our article was usually designated as "nu", it is possible to re-write it in the following form:
(3.5) Pl = 1.11 x 10 at minus 16 degrees x L square x nu in square x I in the square

LET'S DERIVE THE FORMULA FOR "ALFA" IN THE SECOND TRANSMISSION METHOD.

Here we give the power received by the receiver of the object "Pr". For the second transmission method, we derived a formula for it in Chapter 2, this is formula (2.24)
(2.24) P2 = K2 x f (d, epsilon) x nu x R x Phi square
Where
f - function, with parameters: d - thickness of the shell of the collector of the receiver, epsilon - dielectric permeability of the material from which the shell is made. f > 1.
R - is the radius of the receiver store
Phi - is the potential on the transmission line
K2 - is approximately equal to 4.44 x 10 at minus 10 degrees. (K2 = 4 x K1)
We replace "P2" by "Pr" , then the formula (2.24) takes the following form:
(3.6) Pr = K2 x f (d, epsilon) x nu x R x Phi square

In order to write the relation of formulas (3.5) and (3.6), we must reduce them to uniformity. For this, we express the current "I" in (3.5) in terms of other parameters.
(3.7) I = Q / (T / 2),

Where "T" is the power transmission period (of the line potential change). Since the frequency nu = 1 / T, formula (3.7) can be rewritten in the following form.

(3.8) I = 2 x nu x Q

"Q" can be expressed through the capacity of the receiver store and the potential on the power line.

(3.9) Q = C x Phi

here C = K1 x f (d, epsilon) x R (watch Gl.2 formulas (2.9) and (2.10))

we will rewrite a formula (3.8).

(3.10) I = 2 x K1 x nu x f (d, epsilon) x R x Phi

Substitude the derived value for "I" in a formula (3.5). We will receive

(3.10) I = 2 x K1 x nu x f (d, epsilon) x R x Phi

Substitude the derived value for "I" in a formula (3.5). We will receive

(3.11) Pl = 1.11 x 10 at minus 16 degrees x L square x nu in the fourth degrees x 4 x K1 x K1 x f (d, epsilon) x f (d, epsilon) x R x R x Phi x Phi

Recalling that K2 = 4 x K1, we give here the formula for the received power, the transformed formula (3.6):

(3.12) = 4 x K1 x f (d, epsilon) x nu x R x Phi square

Now we can write a formula for "alfa", which is the ratio of the right-hand sides of formulas (3.11) and (3.12).

(3.13) alfa = Pl / Pr = K1 x 1.11 x 10 in minus 16 degrees x L square x nu in the third degrees x f (d, epsilon) x R

We substitute here the value for K1 = 1.11 x 10 in minus 10 degrees.

(3.14) alfa = 1.23 x 10 in minus 26 degrees x L square x nu in the third degrees x f (d, epsilon) x R

WE WILL DERIVE THE FORMULA FOR "ALFA" IN CASE OF THE THIRD METHOD OF THE ELECTRICITY TRANSMISSION.

Here 'I' for the formula (3.5) will be defined as follows.

(3.15) I = 2 x nu x Q

this formula is similar to formula (3.8) for the second method, but 'Q' for the third method is defined differently.

(3.16) Q = C x (Phi m + Phi l),

Where 'Phi m' is the potential of the charging device (an electrostatic machine), And 'Phi l' is the potential on the transmission line (corresponding to 'Phi' for the second method).

The formulas for capacity are similar

C = K1 x f (d, epsilon) x R

We write the expression for the current:

(3.17) I = 2 x K1 x nu x f (d, epsilon) x R x (Phi m + Phi l)

We substitute 'I' in the formula for the lost power (3.5) with this expression. We get:

(3.18) Pl = 1.11 x 10 at minus 16 degrees x L square x nu in the fourth degrees x 4 x K1 in the square x f (d, epsilon) in the square x R in the square x (Phi m + Phi l) in the square

Now we give here the power received by the receiver of the object 'Pr'. For the third electricity transmission method, we derived a formula for it in Chapter 5, this is formula (5.26)

(5.26) Pr = 2 x nu x K1 x f (d, epsilon) x R x Phi l x (Phi m + Phi l)

Now we can write a formula for alfa, which will be the ratio of the right-hand sides of formulas (3.18) and (5.26).

(3.19) alfa = Pl / Pr = 2 x K1 x 1.11 x 10 at minus 16 degrees x L square x nu in the third degrees x f (d, epsilon) x R x (Phi m + Phi l) / Phi l

We substitute here the value for 'K1' = 1.11 x 10 in minus 10 degrees.

(3.20) alfa = 2.46 x 10 in minus 26 degrees x L square x nu in the third degrees x f (d, epsilon) x R x (Phi m + Phi l) / Phi l

Since the third method of single-wire transmission requires an electrostatic machine, a battery and a DC motor in the receiver, we must also take into account their efficiency. And since the efficiency of the battery and the motor is rather low, on the average 85 and 83 percent respectively, then they, and not the calculated efficiency on the basis of radiation losses, will determine the efficiency of the third method of electricity transmission. We will see this, considering the calculation of the example given in this chapter.

Mean values:

DC motor efficiency = 83%

Battery efficiency = 85%

Efficiency of the electrostatic machine - it was not possible to find the value in the literature. However, what I want to note: if you do not take into account the friction of the brushes (in a large machine, this is a miser), then the main losses will go to the friction of the discs about the air. Leakage charges will be small due to the intended isolation of the stores. There are no braking force fields in the machine, so the efficiency will be high.

Recall the preliminary condition of these calculations: 'L' must be much smaller than the length of the radiated wave.

We will provide the table of lengths of waves "VL" and the frequencies "Fr" corresponding to them here. VL and Fr are related by the following relationship: Fr = c / VL, where C is the speed of light.

Fr, Hz----------VL, m
100 Hz------ - 3 000 000 m
1000 Hz (1 KHz)--------------300 000 m
10000 Hz (10 KHz)----------------30 000 m
100000 Hz (100 KHz)-------------------3 000 m
1000000 Hz (1 MHz)-------------------300 m

EXAMPLES.

EXAMPLE N1

Suppose we have a working room with robots that receive electricity for their work. We will use the third method of electricity transmission. The room is equipped with an electrically conductive floor, and the robots themselves contain electric receiver and built- in electric intakes that comes into contact with the floor. The floor measures 100 x 100 m. Transmitters will be on the edges of the floor. Then L max = 100m. In order to comply with the above condition, i.e. L (antenna length) should be much smaller than the wavelength of the radiated wave-VL. Assume that VL min = 1000m (ten times more than L max). Let's calculate Fr max , it will be 300 kHz. Let the geometric dimensions of receiver's stores in robots be real and amount to 1 m in diameter. We use the existing Vimschurst electrostatic machine. The electrical potential on the receiver's stores (with insulating sheath) will be equal the potential on this machine and will be 300 kV. The voltage on the power line (and this is the floor), we agreed to use only safe for life and equal to 220 V.

Let's answer the following questions:
1) What maximum power can be transmitted to each robot ?
2) What is the efficiency of the power line ?

To answer the first question, let us use the formula derived by us in Ch. 5, which determines the power transmitted by the third method of transmission. This is the formula (5.26)

(5.26) P3 = 2 x Nu x K1 x f (d, epsilon) x R x Fl x (FM + FL)

Where:

K1 - coefficient , approximately equal (as not all stores of receivers will be the correct spherical form) 4 x Pi x epsilon zero. K1 is approximately equal = 1.11 x 10 to the minus tenth degree.

f - function, with parameters: d - receiver store cover thickness, an epsilon - dielectric permeability of material of which the cover is made. f > 1.

R- is the radius of the store. In our example it is 0.5 m.

Fl is the potential on the transmission line (its absolute value), in this case, equal to 220V.

Fm is the potential on an electrostatic machine or some other device of an object receiver that recharges its drives (also its absolute value). In this example, equal to 300kV.

Nu is the frequency equal to Fr max = 300 kHz.

We substitute these values in the formula. We get P3 = 2199.4 W or about 3 horsepower. Thus, we will be able to give the robots enough power.

Now let's compare what would have happened if we had used the second method of electricity transmission. The formula for the transmitted power for this method was derived in Ch.2. This is the formula (2.24).

(2.24) P2 = K2 x f (d, epsilon) x nu x R x Phi square

Where:

f - function, with parameters: d - thickness of the shell of the collector of the receiver, epsilon - dielectric permeability of the material from which the shell is made. f > 1.

R- is the radius of the reciever store. In our example it is 0.5 m.

Phi - is the potential on the transmission line

Phi = Fl = 220 V

K2 - is approximately equal to 4.44 x 10 at minus 10 degrees. (K2 = 4 x K1)

Nu is the frequency equal to Fr max = 300 kHz.

We substitute these values in the formula. We obtain P2 = 3.22 W, i.e. Incomparably less transmitted power.

CHAPTER 4. WAYS OF INCREASING THE TRANSMISSION POWER.

Goal is - to create a single-wire electricity transmission, powerful, safe for a human body, at the small sizes of a current collector and at high efficiency and low cost.

Chapter 2 has a formula of power. Frequency and potential increase power, but as it was already specified, there are restrictions. High frequency reduces efficiency creates strong electromagnetic radiation and hindrances, and a lot of potential on the line may cause a fatal electric shock. In each case it is necessary to define the maximum potential on the line, and the maximum transmission frequency. In some cases, we will be ready to sacrifice high efficiency for increasing transmission power. For example, if the line goes through the taiga or in the desert, and the potential can be increased by high-frequency interference and electromagnetic radiation will not cause much harm, if you set the warning signs. But in the city it's different. In some cases, for example for electric transport transmission line should be no isolated because of the current collectors, respectively, and the potential to be safe. If the transmission line very long, too high frequency can lead to that charges for a half of a half-cycle won't manage to reach the store the receiver as a long wire has inductance which will work for delay. It can be explained from the point of view of the theory of an electrical oscillating circuit. In this case we have a sequential oscillating circuit with the certain natural frequency determined by values of capacity and inductivity. Capacity is a capacity of the receiver store (depends on the geometrical sizes), and inductivity is defined by length of a wire (line). It is known that the largest amplitude will be when the stimulated frequency (transmitter generator frequency) is equal to the natural frequency of a circuit determined by Thomson's formula. When the stimulated frequency is greater than the natural frequency, the value of the transmitted current decreases. A long wire has a large inductivity and will decrease natural frequency of the circuit. With the usual current transmission for small distances and with small geometric dimensions of the drives, we are always "not reaching" to the natural frequency of the circuit, therefore, as the frequency increases, the transmitted power increases, as shown earlier in Chapter 2.

Store capacity and consequently also transfer power, will increase if to surround it with the isolating cover with a big dielectric permeability (epsilon). In each case we define what material, firm or liquid will suit us depending on high cost, technical capabilities of performance and the demanded isolating properties.

Possible to use different physical effects and devices to increase the amount of charge received at halftime (of frequency period). According to the author, suitable Hall effect (the author has even experimental design in this area) and "Stop revolving coil" - as it was first used in the experiments L.I.Mandelstam and N.D.Papaleksi and later in experiments R.CH.Tolman T.D.Stewart.

The author has planned some more improvements (for example poly-cascade electric circuits) of an electricity transmission (increase in power) which need the experimental verification.

Besides, there is one more option - to apply "the third way of an electricity transmission", though it and more difficult for technical realization in devices, than the first two.

CHAPTER 5. A THIRD WAY OF A SINGLE-WIRE ELECTRICITY TRANSMISSION.

The idea is that for small (comfortable) geometric dimensions of the store of the receiver it would be placed at the half of period a large number of charges, much bigger, than it can place at safe potential on the line. And, potentials on object can be and much more, than on the line. It is admissible from the point of view of safety measures. As the small device of the receiver can be insulated carefully is a small space. But all of a transmission line is very difficult and in some applications as it was stated above, it just also has to be without isolation. This large amount of the charges will be "pumped" through the load toward the transmission line carrying the charges with opposite sign, and despite the fact that we have to expend energy in the beginning (to charge the receiver store) the total received by the load energy will be more than in the first two methods.

We will define this way so - giving of charges to object at safe potential on the line to the object stores, which are previously charged to higher potential, through the loads (which are on object), and charges on stores have to be an opposite sign, in relation to transferred.

This method has been presented long time ago in [1], but was then not well explained.In order to understand how it works, let's solve the small problem in physics, it is shown in Fig.6.

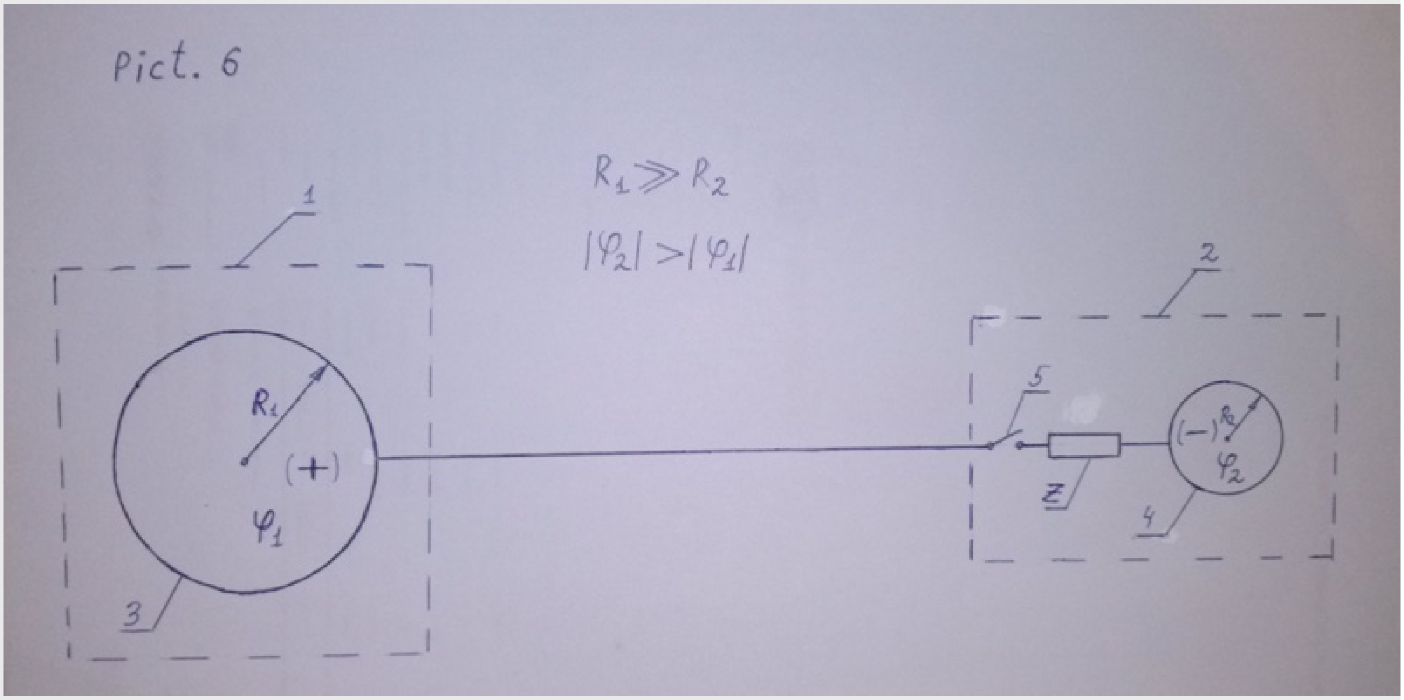

TERMS AND CONDITIONS:

In the transmitter (1) there is a electrically conductive sphere (3) very big sizes which radius we will designate (R1). On it there are many electric charges, and it is loaded to the potential (Phi1). It is small potential and it isn't deadly to a human body. Further, the transmitter is connected by one conductor with the receiver (2) in which too there is a electrically conductive sphere (4), but already very modest sizes convenient, say, for the moving object of a current supply. It via the switch (5) and through load (Z) is connected with the line. Radius of this sphere we will designate (R2). We can write that the radius (R1) is much more (R2). The switch is meanwhile opened. The sphere of the receiver is previously loaded to the potential (Phi2). And this potential is very high much more on the module, than the potential (Phi1). I think that makes sense to load this sphere (4) by means of an electrostatic machine*. Charges on a sphere of the receiver (4) are opposite on a sign to charges of a sphere of the transmitter (3). Previously having loaded a sphere of the receiver (4) we spent energy (Wzero). Now we close the switch (5). What will occur? The author believes that all charges from a sphere (4) through load (Z) will flow on a transmission line on a sphere (3). And, potential on a sphere (3) and on the line will practically not change as a sphere big and has many charges. After all condenser potential, as we know, is defined by a formula:
(5.1) Phi = Q / C ,
where Q - amount of charge, and C - capacitance.
Since we chose sphere large geometric dimensions (large R), it has a large capacity (because C = 4 x Pi x epsilon x epsilon zero x R) and has a large number of charges. Therefore, a small increase or decrease of the charge, which is necessary for the "work" with the sphere (4) will not greatly affect the potential of the sphere (3) and, consequently, the line. So it will have a real transmitter, as the annihilation of the charges that are currently on the line by entering the line of charges of opposite sign, all the time is offset by applying to the line charges from the generator. Then, part of the charges from the sphere (3), again, through a transmission line and through the load (Z) will flow to the sphere (4) to establish an equipotential surface. Ie until the potential on the sphere (4) will not be equal to the potential on the sphere (3) and on the line. The potential on the sphere (3) and on the line as before - will not change. Despite the fact that I am talking here about the two stages of the process, in fact, in terms of movement of electrons - it is continuous, and the electrons will flow in one direction only. It's just a matter of determining the sign of the charge. Time we take a half of the period and it more than enough for the described process. On load (Z), as a result of electron flow through it , some energy will be emitted **. We will designate it (Wz).

QUESTION N1: Will the released energy to the load (Wz) more than originally consumed at a charge storage facility (Wzero)? And if so, how much? The energy difference is denoted (deltaW).

QUESTON N2: To find (W3)- it is the energy received by object for a half of cycle at such way of an electricity transmission minus originally spent (Wzero), and to correlate it to the energy received for the same time at the second way of an electricity transmission (W2) (it more powerful, than the first way) on the store of the same size (R2), i.e. the same capacity (C2), and with the same potential on the line (Phi1). It all the same what to speak about comparison of power.

AUTHORS` SOLUTION:

The calculation given below, is very similar to the calculation of power for the second way of an electricity transmission given in Ch.2.

Work (A) of movement charges (Q) in electric field is defined by a formula:

(5.2) A = Q x deltaPhi ,

where deltaPhi - a potential difference of the sphere (4) and the line. At the beginning on the line and on a sphere (4) potentials of the different sign Phi1 and Phi2. Potential difference

deltaPhi = Phi1-(-Phi2) =Phi1 + Phi2.

But, in process of leakage of charges from a sphere (4) and their passing through the load (Z), potential on a sphere (Phi2) changes, therefore deltaPhi is function of quantity of charges on a sphere (4). We will write down it so:

deltaPhi = F(Q)

From the point of view of power, process of passing of electrons through the load can be divided into two stages. The first stage where work (A1) will be made - this movement of all charges from the receiver sphere (4) (we will assume negative charges) through the load (Z) to a power line. This process lead to a total disappearance of charges of this sign on the sphere (4) and, according to it, transformation of its potential into the zero. The second stage where work (A2) - this passing through the load of charges from the line will be made (in our example positive) to the receiver sphere (4) before establishment of potential on it (positive) equal to the potential on the line (Phi1). Thus, the general work or energy (that in this case the same) which is obtained for a half of the period on the load:

(5.3) Wz = A1 + A2

Let's try to calculate the work A1.

Elementary work (its differential) will be equal to the multiplication of the potential difference deltaPhi on the elementary charge (differential Q).

(5.4) dA1 = deltaPhi (Q) x dQ = F(Q) x dQ

As already mentioned above, the potential difference is a function of the charge on the sphere F(Q). Let's see how this function can be expressed. With the leakage of charges through the load, it decreases. The potential of the store (Phi2) is determined by the following formula.

(5.5) Phi2 = Q / C, then

(5.6) F(Q) = Phi1 - (-Phi2) = Phi1 + Phi2 =Phi1 + Q / C

Substituting this expression into (5.4), we obtain:

(5.7) dA1 = (Phi1 + Q / C) x dQ

In order to define work A1, we have to take the integral of the expression within the following limits: lower = 0, and the top, equal to the maximum (initial) charge that is Qmax = C2 x Phi2.

After the corresponding calculations which are excessive for bringing here, we will receive:

(5.8) A1 = Phi1 x Phi2 x C2 + Phi2 square x C2 / 2

Now let's calculate the A2.

Work or energy of movement of charges, as well as for A1, is determined by a formula (5.2). The potential difference between the line and the receiver sphere (deltaPhi) is also a function of the charge on this sphere. Let's see how in this case is defined F (Q). During all process, including at its beginning, potential on the line = Phi1. Phi2 sphere (4) potential is equal in the beginning 0, but in process of receipt on it charges which pass a way from the line through the load, potential grows. Phi2 is defined by a formula (5.5). Its final value will be Phi1. For a potential difference or F(Q) the following formula has to be fair.

(5.9) F(Q) = Phi1 - Phi2 = Phi1 - Q/C

Elementary work (dA2) is defined by a formula (5.4). We will substitute in it the expression which is just received by us for a potential difference, i.e. a formula (5.9). We will receive:

(5.10) dA2 = (Phi1 - Q/C) x dQ

In order to define work A2, we have to take the integral of the expression within the following limits: lower = 0, and the top, equal to the maximum (final) charge that is Qmax = C2 x Phi1.

After the corresponding calculations which are excessive for bringing here, we will receive:

(5.11) A2 = Phi1 square x C2 - Phi1 square x C2 / 2 = Phi1 square x C2 / 2

To find the general work or energy emitted on loading for a half of the period of an electricity transmission (Wz) we substitute A1 and A2 values in a formula (5.3). We will receive:

(5.12) Wz = Phi1 x Phi2 x C2 + Phi2 square x C2 / 2 + Phi1 square x C2 /2

Now we will try to find the answer to the question N1

According to the formula of calculating the energy of the charged capacitor determine Wzero:

(5.13) Wzero = Phi2 square x C2/2

(5.14) deltaW = Wz - Wzero

We will substitute in a formula (5.14) expressions for Wz and for Wzero, we will receive:

(5.15) deltaW = Phi1 x Phi2 x C2 + Phi1 square x C2 / 2

Thus, we see that we won power, and quite significantly, because the potential Phi2 can be very high, since the compact means of generating (eg electrostatic machines ***) - are available, and the means of isolation**** -the same.

LET'S ANSWER THE QUESTION N2.

We will consider energy which will receive object of a current supply in our task (at the third way of an electricity transmission) for a half of the period of an electricity transmission. We designated it as W3 above. We already found one component of W3 - it is deltaW - the energy emitted on load minus originally spent energy on store charging. Besides, the receiver store, having received a charge from the line of transfers (at the end of process) and therefore having got the potential Phi1, receives some power status, i.e. potential energy Wp which we too have to consider at calculation of a total energy W3 received by the receiver. Later, by consideration of a question of providing a continuity of an electricity transmission at the third way, it will be explained how to use this energy. So,

(5.16) W3 = deltaW + Wp

We will find Wp on a formula of calculation of energy of the charged condenser:

(5.17) Wp = Phi1 square x C2 / 2

We will substitute the found expressions for deltaW and for Wp in a formula (5.16). We will receive:

(5.18) W3 = Phi1 x Phi2 x C2 + Phi1 square x C2 / 2 + Phi1 square x C2 / 2 and after transformation:

(5.18) W3 = Phi1 x Phi2 x C2 + Phi1 square x C2

We will compare this expression to the energy received by the receiver at the second way of an electricity transmission for a half of the period - W2. The formula for W2 we learn in Chapter.2 It is a formula - (2.21). Phi in it it is possible to replace with Phi1, and C with C2. So: (2.21) W2 = 2 x Phi1 square x C2

In order to compare the energy W3 and W2, introduce the measure - n. This value shows how many times the potential of previously charged receiver store - Phi2 is more than a potential on the line - Phi1. We can write:

(5.19) Phi2 = n x Phi1

We will substitute Phi2 with the found now expression for Phi2 in a formula (5.18). We will receive:

(5.20) W3 = n x Phi1 square x C2 + Phi1 square x C2

and after transformation:

(5.20) W3 = (n+1) x Phi1 square x C2

Now you can see what will be equal to the ratio of W3 and W2?

(5.21) W3/W2 = [(n+1) x Phi1 square x C2] / [2 x Phi1 square x C2]

and after reduction:

(5.21) W3 / W2 = (n+1) / 2

ANSWER TO A QUESTION N1:
deltaW = Phi1 x Phi2 x C2 + Phi1 square x C2 / 2

ANSWER TO A QUESTION N2:
W3 / W2 = (n+1) / 2 ,
where n = Phi2 / Phi1

CONCLUSION :

At the third way of single-wire transfer, the power received by the receiver will increase almost in one thousand times in comparison with the first two ways. So, if to take the potential produced by electrostatic machine of Vimshurst = 300 kV - it there will be Phi2 and potential on the line = 200 V - it will be Phi1 (potential, safe for human life, on the standard canons), n will be 1.5 x 1000, and W3 relation to W2 on a formula (5.21) will be equal about 750. But Vimshurst's electrostatic machine - it yet not a limit of perfection. Such task - to create the electrostatic machine developing the maximum voltage at the minimum sizes wasn't set for mankind simply yet. So, in ONE THOUSAND TIMES it is quite possible to reach increase in power.

PROVIDING CONTINUITY OF THE ELECTRICITY TRANSMISSION.

Pict. 7

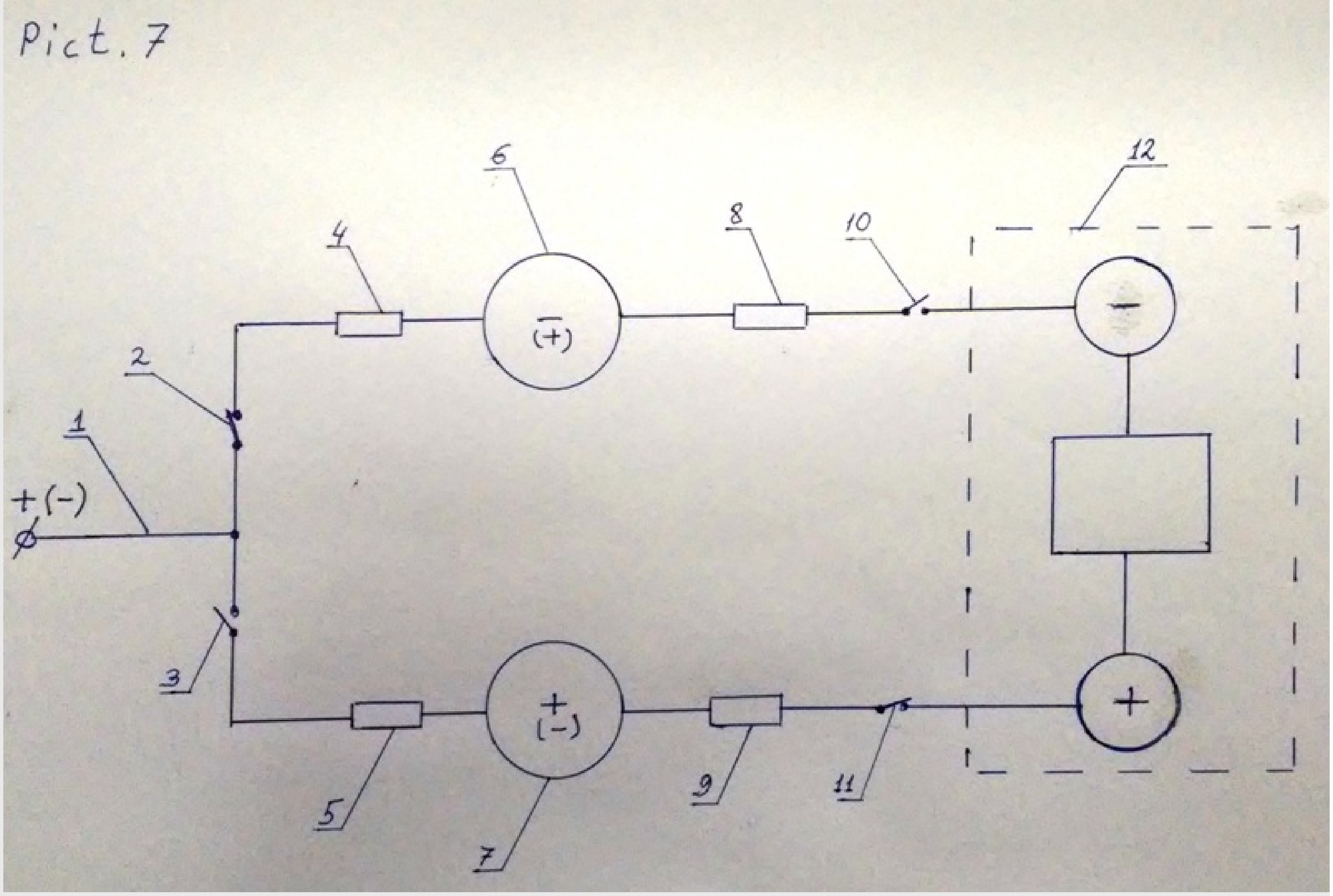

The diagram (Pict. 7) displays only the object of current supply and appropriate to it single-wire transmission line segment (1).The transmitter remains same what it was at the first and second ways of an electricity transmission (see Pict. 1-5).

In this scheme it is impossible to put a two diode switch 'diode fork' on an entrance of the object of a current supply as at the first way (see Pict. 4). Because of the fact, that previously charged object stores (6 and 7) will electrically connect with each other. However, through loads (4 and 5), but so they won't save up charges, and will lose the potentials, and to us, at this way of transmission, it is necessary that they had many charges at high potentials. Therefore we will use the operated switches (2 and 3).

Loads (8 and 9) between terminals of the electrostatic machine (12) and stores of object (6 and 7) are necessary for purpose to use the energy Wp formed as a result of a charge of stores by charges of an opposite sign (in relation to output terminals of the electrostatic machine) at the end of time of connection of each of them to a power line. This energy will be used when the store is switched-off from a transmission line and connected for charging to the electrostatic machine by means of the operated switch (10 or 11).

What happens in the first period of the power supply we have already discussed in detail above in the solution of the task (see. also Fig. 6).

For providing a continuous electricity transmission, on object of a current supply (see Fig. 7) there is a second store (the best form - sphere) which is previously charged with electric charges of other sign, than the first, and in the second half of the period we switch a transmission line to it through other load (4 or 5). And on the transmission line itself already the charge sign too will be changed, because of connection of other sphere of the transmitter (or operation of the generator). And in this half-cycle too everything will occur as in the considered task. In that part of the period when any store of object is switched-off from the line, it will be loaded on the object with the help, we will tell, the electrostatic machine or other device, but in the beginning of this half-cycle it will give the charge, which remained from the previous half-cycle, through the corresponding load (8 or 9) connecting it to the charger. Thus the power status received from the line of transfers at the previous half-cycle (Wp) will be used through this load.

If to compare this process to charging of the neutral store, yes, it is valid, the part of charges going from the electrostatic machine, equal to "spaciousness" of the store of the receiver Q = C x Phi at the potential of line = Phi1 will be annihilated, and must be "restored" by electrostatic machine. This of course would require additional energy (compared to charging a neutral store) from the electrostatic machine, but it will be "returned to us"(the truth with small losses which we don't consider in this calculation) i.e. will turn in "useful" as it will be allocated on load, and from loads we take energy for return to the electrostatic machine and on performing of useful work.

Algorithm of operation of the switches operated (naturally automatically) the following: in the first half-cycle switches (2) and (11) are closed, and opened (3) and (10). And in the second - are closed (3) and (10), and are opened (2) and (11). In the third half-cycle - as in the first, and in the fourth - as in the second etc.

When selecting receiver stores shells material at the third way, it is necessary to pay attention not only to dielectric permeability of material, as at the first two, but also to its electric durability as we will deal with high and life-threatening voltage. Besides, it is necessary to look how these properties of material will behave with the planned electricity transmission frequency (it concerns also the first two ways). Cost too should be considered.

ELECTRO-RECEIVER OF OBJECT OF A ELECTRIC POWER SUPPLY, WORKING AT THE THIRD WAY OF ELECTRICITY TRANSMISSION.

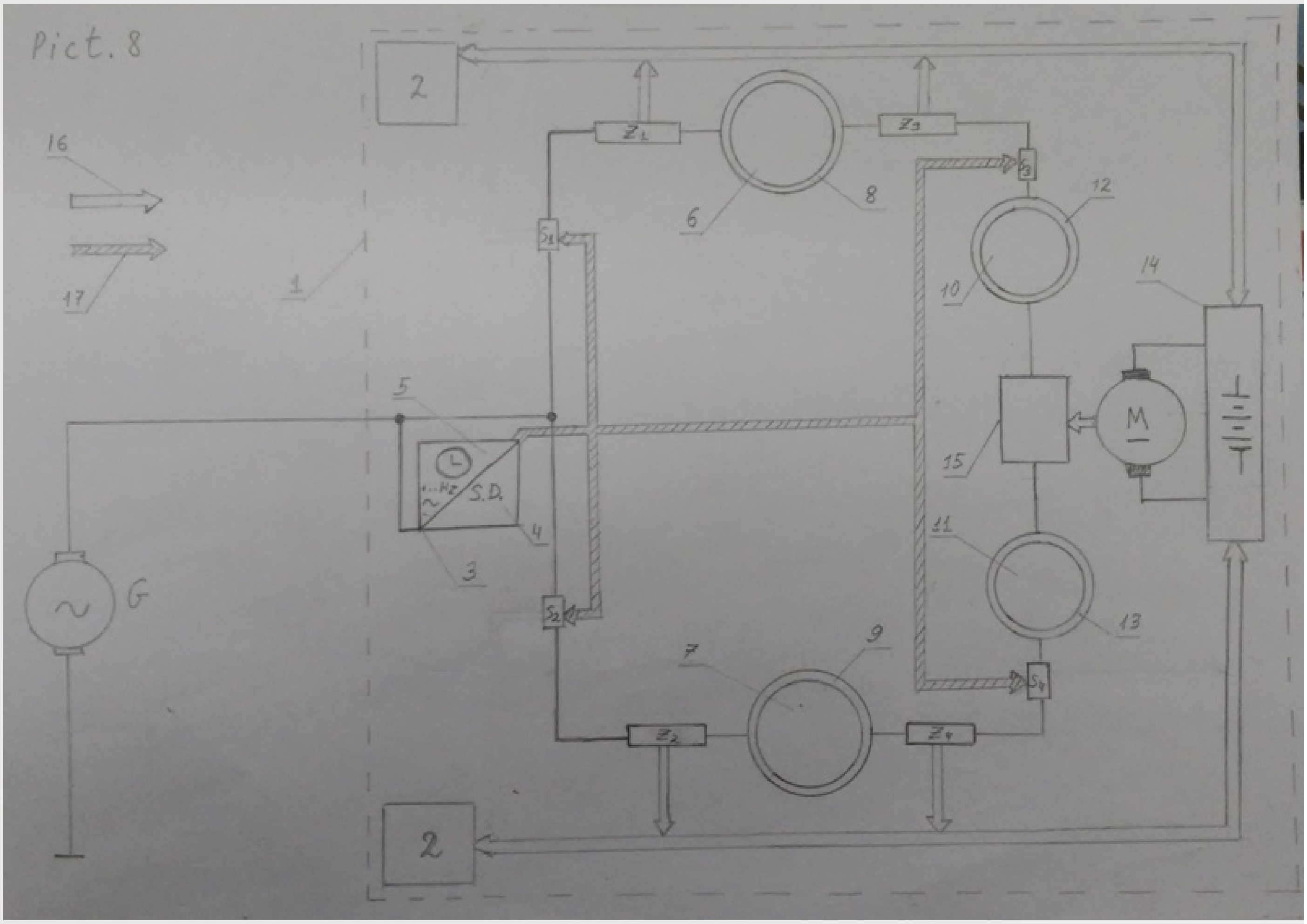

Pict. 8 SHEME OF THE ELECTRO-RECEIVER OF OBJECT OF A ELECTRIC POWER SUPPLY, WORKING AT THE THIRD WAY OF ELECTRICITY TRANSMISSION.

Signatures to Pict. 8

1-Object of a electric power supply.

2-The consumer of the electric power on object.

3-The block defining sign of charges on a power line.

4-The device defining a sign of charges on the line by means of the scheme of a single-wire electricity transmission.

5- The operating electric clock.

6,7-Stores of charges of object (main).

8,9-Isolating covers of the main stores.

10,11- Stores of charges of the electrostatic machine.

12,13-Isolating covers of stores of the electrostatic machine.

14-Rechargeable battery.

15-Electrostatic machine.

16-Energy stream.

17-Flow of managing signals.

Z1, Z2, Z3, Z4-Loads**.

S1, S2, S3, S4- The Operated switches.

TRANSFER OF THE ELECTRIC POWER ON OBJECT IN THE PRESENCE OF A SIMILAR COLLECTOR IS CARRIED OUT AS FOLLOWERS.

The generator (G) delivers on a power line charges of a different sign, in each half-cycle different, in one half-cycle - plus, in others-minus. And as he is connected by one terminal to the buffer store (in this example it is Earth - the grounding displayed on the scheme), these charges are "free", i.e. capable to fill a surface of conductors. You can object to this term, but how then to distinguish charges that could put a battery or generator in an electrical circuit connecting the two their terminals from the charges that capable to fill surface of free conductors and to interact with the same "free" charges? In any battery or generator "outside forces" divide charges. But from a terminal of the ordinary battery or the generator they are able to flow only on other terminal of the same device. Even to the terminal of other battery with an opposite sign they won't begin to flow. And in the Earth they won't begin to flow just like that (with an empty second terminal). Otherwise the situation is with the source given in the scheme and with electrostatic machines in which "outside forces" work too. It turns out that presence of "outside forces" for development of "free charges" - the necessary condition, but insufficient as mathematics say.

On a current collector 1 to which lead a line there is a Unit 3. A purpose of this unit - to define what charges in this half-cycle are on the line (plus or minus). If the unit is connected to the line, then it with it is helped by the device switch 4 (the device 4 is given as an example, also other options, say, the device consisting of diodes, electroscopes and optical sensors are actually possible - petals of electroscopes will cross beams), which operation we will consider later. But it can not be connected, then the sign of charges is defined by means of Clock 5 which are synchronized with the generator. The second purpose of the Unit 3 is development and a message of the operating signals on the operated switches (S1, S2, S3, S4). Passing of the operating signals on the scheme is conditionally designated by a stream 17 (the shaded arrows). These signals will depend on what charges (on a sign) in this half-cycle on the line of an electric supply.

On a current collector 1 are the electrostatic machine 15 and its stores of charges (10, 11) with the isolating covers (for safety) (12,13), the accumulator 14, the DC motor (M) for rotating the disk of electrostatic machine. All this "housekeeping" is necessary to receive more 'free' charges from the power line which is under the low potential, safe for life. The motor takes the electric power from the accumulator, puts the electrostatic machine in action and that charges the stores, one (we will tell 10) with positive charges, and another (we will tell 11) negative. This process goes constantly to all the time of an electricity transmission, irrespective of the fact which in this half-cycle charges (positive or negative) on the line going to a current collector. It is thought that the energy received on loads (Z1, Z2, Z3, Z4) quite will be enough both for recharge of the accumulator and for providing consumers of energy 2 objects. Streams of energy 16 are designated on the scheme by not shaded arrows.

The basic working elements of a current collector are stores of charges (6 and 7) and above-mentioned loads (Z1, Z2, Z3, Z4). Stores are surrounded with covers (8 and 9). These covers have double function: to increase the storage capacitor capabilities of stores, dielectrics of which they consist if surround the conductor, then with success it is done, and to isolate stores for safety reasons as on stores there will be sometimes a high voltage. By the way, the case of all current collector has to be made of the good insulating stuff capable to protect the person from high voltage.

We will notice that at the beginning of any half-cycle one of stores of object (6 or 7) has to be completely charged from the store of the electrostatic machine (10 or 11) and have the high potential, equal with him, and another has to begin to be loaded from other store through one of the loads (Z3 or Z4).

We will consider a "negative" half-cycle in the beginning. On the line the negative charge, the store 6 is loaded positively to the potential of the electrostatic machine (it 'inheritance' of the previous "positive" half-cycle), the S1 switch is open, the S3 switch is closed, the S2 switch is closed, the S4 switch is open. What is going on? The store 6 and the line interact through Z1 load, this process has been described above, on load energy is emitted, she is shown by not shaded arrow departing from her. Energy goes to the consumer of energy 2 objects and for accumulator charging 14. At the end of this half-cycle, the store 6 will get a negative charge and potential of the line and will transfer this "inheritance" to following, already "positive", half-cycle of an electricity transmission . The store 7 in the same time interacts with the store of the electrostatic machine 11 through S4 load. On the store 7 a positive charge and potential of the line (which has got to it "in inheritance" from the previous "positive" half-cycle) and at the end of this "negative" half-cycle it will be charged from the store 11 negatively and will get the potential of the electrostatic machine. From load Z4 energy will go to the same purposes.

Now, on the line "comes" the "positive" half-cycle. The Unit 3 "feels" it and "sends telegrams" on the shaded arrows on the operated switches. As a result S1 and S4 are closed, and S2 and S3 - open. The store 7, having high negative potential from last half-cycle, interacts with the line through Z2 load, at the end of a half-cycle he will get a positive charge and potential of the line. And the store 6, having a negative charge and potential of the line, interacts with the "positive" store 10 of the electrostatic machine through Z3 load. At the end of this half-cycle he will get the high positive potential equal to him.

When on the line again "come" "negative" half-cycle, object stores again, will have, the "inheritance", described right at the beginning. And the Unit 3, "having understood a situation", will open S1 and S4 and will close S2 and S3.

And so on everything will cyclically occur.

SWITCH *****,OPERATING DEPENDING ON A SIGN OF CHARGES ON A POWER LINE.

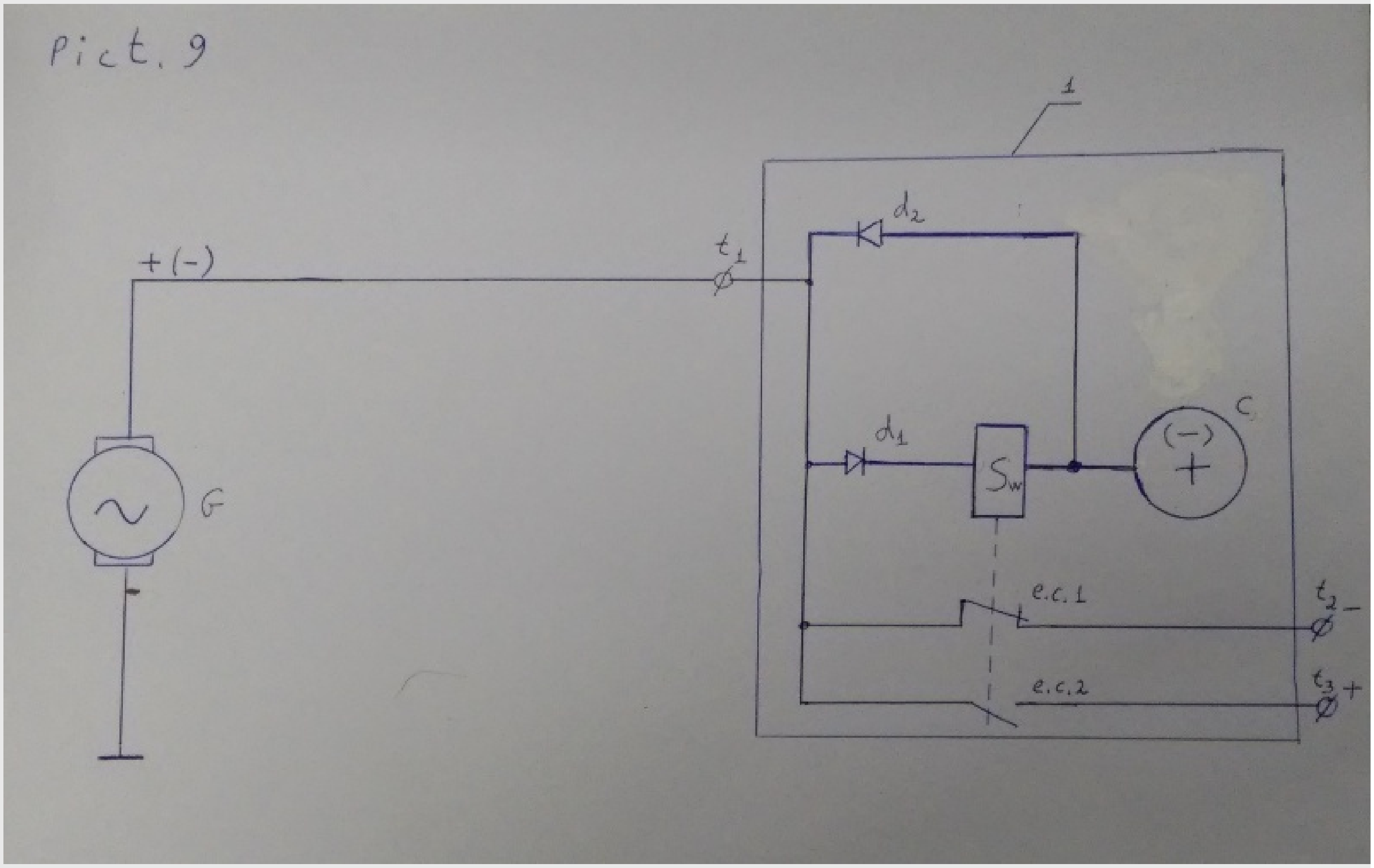

Pict. 9. THE SWITCH *****, ACTING IN ACCORDANCE WITH THE SIGN OF THE CHARGES ON THE POWER LINE.

This switch - unit 4 from the previous diagram (Figure 8).

1- the housing.

Sw- relay.

C-store of charges.

e.c.1- normally closed contact.

e.c.2- normally open contact.

Operation of this device is as follows.

As already mentioned above, the apparatus 4, part of unit 3 (scheme 8) should provide recognition charge sign which currently are in the single-wire transmission line as well as production and send control signals to the switches (S1, S2, S3, S4). Charges from a transmission line enter on the t1 plug. After it they are waited by the diode fork of d2 - d1. If positive charges, they, seeking to fill the store C, pass through the d2 diode and a winding of the Sw relay, having turned on this relay. If charges on the line the negative, then positive charges from the store C via the d1 diode go to the line (and are annihilated with the negative charges of the line) and pass a Sw relay winding. The Sw relay in this half-cycle will be in a silent state - there is no current through a winding.

Thus, depending on the operation or non-operation switch Sw charges recognition occurs on the line.

So:

When "+" charges on power line running to the object, the relay Sw is triggered (current passes through its winding) closes its contact e.s.2 that supplies currents (control signals) to the relay coil switches S2 and the S3, and, thus opens them. Normally closed contact e.s.1 thus opens and stops the currents flow (control signals) via a relay coil S1 and S4 and thus disables them.

If '-' charges on the transmission line, the relay Sw is de-energized , contact e.s.1 closed. He sends control currents to the relay S1 and S4 switches and opens them. e.c.2 - open. He stops for the currents in the windings of relay controlled switches S2 and S3, and disables them.

Pict. 9 SWITCH, OPERATING DEPENDING ON A SIGN OF CHARGES ON A POWER LINE.

This device includes the device 4 and the operated S1 and S2 switches from the previous scheme (Pict. 8).

Signatures to Pict.9

1- housing.

Sw- relay.
C- storage charges.
e.c.1- normally closed contact.
e.c.2- normally open contact.

FORMULA FOR POWER AT THE THIRD WAY OF THE SINGLE - WIRE ELECTRICITY TRANSMISSION.

Let's consider energy which will be received by object of an electric power supply at the third way of an electricity transmission for a half of the period of an electricity transmission. We already found it above and wrote for it a formula (5.18) having designated as W3.

(5.18) W3 = Phi1 x Phi2 x C2 + Phi1 square x C2

Now for larger simplicity and clarity we will change designations and we will explain them. W3 is replaceable on A, Phi1 to Phil, Phi2 to Phim. C2 can be replaced just with the simple C.

Let's rewrite taking into account new designations a formula (5.18), now it will become a formula (5.22).

(5.22) A = Phil x Phim x C + Phil square x C,

or

(5.22) A = C x (Phil x Phim x C + Phil square),

or

(5.22) A = C x Phil x (Phim + Phil)

Here :

A - the energy transferred to the object for half frequency period.
C - the capacity of one of receiver stores (there are two identical stores).
Phil - potential on a transmission line (its absolute value).
Phim - the potential of an electrostatic machine or some other device , which charges receiver stores (also its absolute value).

To find the power (denoted by P3), it is necessary to divide the work (A) on the time for which we have found it . We shall take an interval of time (t) equal to a half-cycle (t=T/2) of the frequency (nu) of switchings of the direction of a current of charges in case of a single-wire electricity transmission. Write a formula.

(5.23) P3 = A / t = A / (T / 2), but the T = 1 / nu, so, finally, this formula is as follows:

(5.23) P3 = 2 x nu x A

Substitute in this formula the expression for A (formula 5.22). We get: (5.24) P3 = 2 x nu x C x Phil x (Phim + Phil).

Now let's express storage capacity through the other parameters, as we did for the counting capacity of the first and second methods in the Chapter 2.

(5.25) C = K1 x f (d, epsilon) x R

Where:

K1 - coefficient , approximately equal (as not all stores of receivers will be the correct spherical form) 4 x Pi x epsilon zero. K1 is approximately equal = 1.11 x 10 to the minus tenth degree.
f - function, with parameters: d - receiver store cover thickness, an epsilon - dielectric permeability of material of which the cover is made. f > 1.
R- is the radius of the store

Substituting this formula in the expression for power. We get:

(5.26) P3 = 2 x K1 x f (d, epsilon) x nu x R x Phil x (Phim + Phil).

We will compare this formula to the formula derived by us in Chapter 2 for the power transferred in case of the second method of an electricity transmission (formula 2.24).

(2.24) P2 = K2 x f (d, epsilon) x nu x R x Fi square,

where Fi is a potential on the line, i.e. it is Phil in the present designations, and K2 = 4 x K1

We can rewrite a formula (2.24) in the following look now

(2.24) P2 = 4 x K1 x f (d, epsilon) x nu x R x Phil x Phil

and to compare it to a formula (5.26). We see that in a formula for the third method, one of multiplicands (Phil) is replaced on (Phim + Phil). And as Phim is a potential of the electrostatic machine, and it much above Phil, so delivered power for the third method will be much higher.

There is a question why in the case of count of power for the second method of an electricity transmission we didn't consider the potential energy Wp received by the store at the end of a half-cycle? But for the third - consider? - Because, this potential energy, we "transfer" without using in the following half-cycle (only on the next time span it brought benefit)- second way. And here (third way) we its, truth, transfer too without using in the present half-cycle, but we receive it (not its, it is more correct correct to say - equal to it) on the other store in the same interval of time, i.e. in a "estimated" half-cycle. Energy is received in the beginning of charging from the electrostatic machine of the parallel store on the appropriate load (Z3 or Z4 - see the Diagram 8).

We will remind that calculation is made for low frequencies of an electricity transmission. For high there will be differences. The main is that the geometrical sizes of stores of the receiver will cease to influence the accepted power as these conductive bodies won't manage to be filled for a half-cycle with charges. It is confirmed in work[10

NOTES :

* The author has an idea how to make an electrostatic machine with a miniature mechanism, i.e. does not take up much space on the receiver of the object, and without rubbing brushes - collectors. In fact, this idea belongs Kirsanov B.P., but the author of this article it creatively developed and perfected.

** The load, as noted in Chapter 1, should be in the form of a winding motor or transformer because of the three manifestations of the electric current (electromagnetic, thermal, chemical) with single-wire electric power transmission guaranteed only one thing - electromagnetic.

*** Wimshurst electrostatic machine, for example, with relatively small size, gives us 300 kV.

**** So, for example, electric durability of mica =135 kV/mm, polyethylene makes 40 kV/mm, and ebonite and porcelain = 25 kV/mm. Therefore , the insulation thickness of only a few cm can protect from breakdown at voltages of above-stated electrostatic machine.

***** If we want to define a sign of the plug of the battery or other device which isn't (like a given above transmission line) a source of 'free' charges, it is necessary to connect the second plug of the battery to the 'buffer' store of charges, i.e. to the big conducting body or to the Earth.

CHAPTER 6. OUR EXPERIMENTS .

The test report of the our pilot unit, executed according to the scheme of Pict. 4.

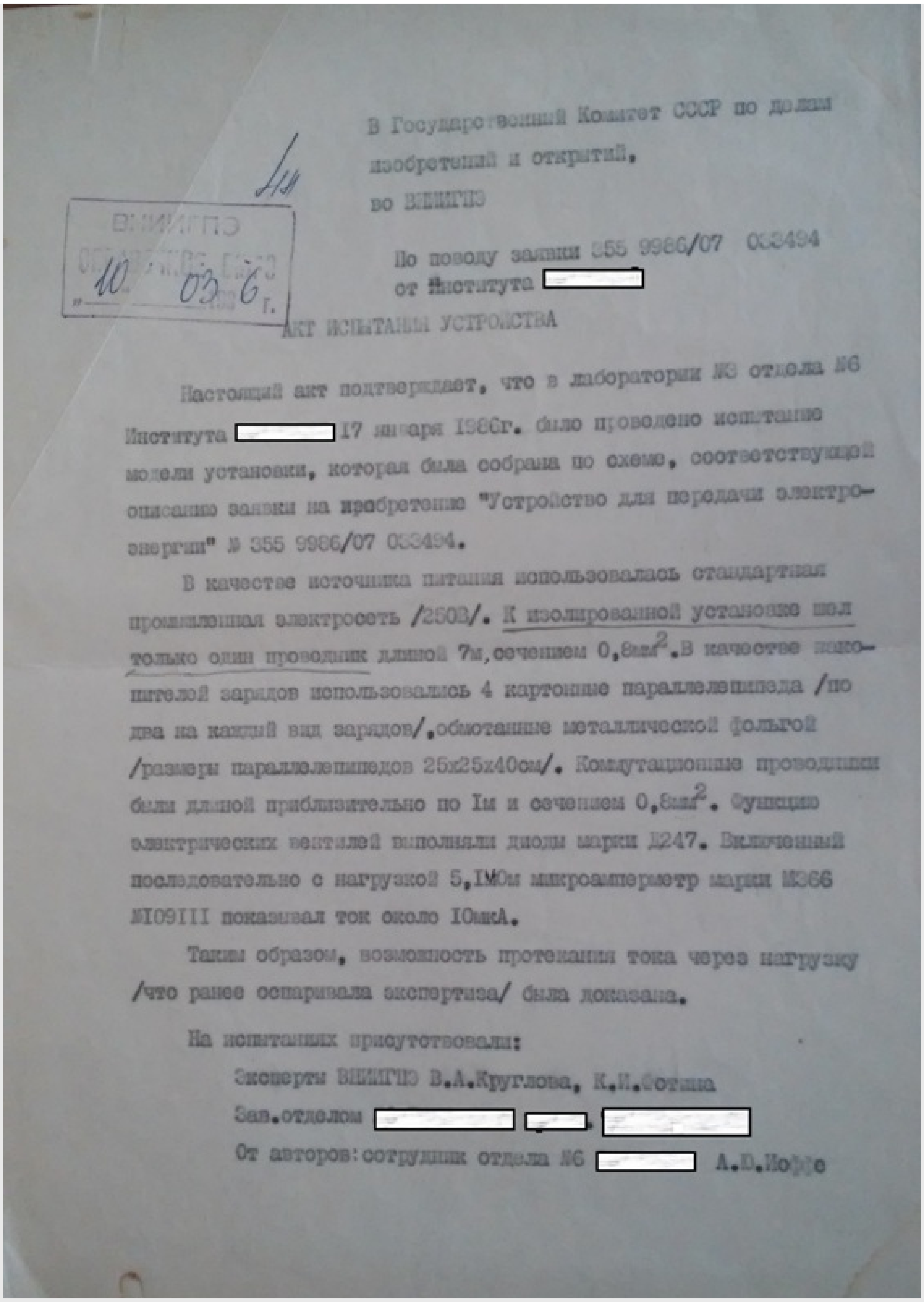

В Государственный Комитет СССР по делам
изобретений и открытий,
во ВНИИГПЭ

ВНИИГПЭ
[illegible]
« 10 » 03 6 г.

По поводу заявки 355 9986/07 033494
от Института [redacted]

АКТ ИСПЫТАНИЯ УСТРОЙСТВА

Настоящий акт подтверждает, что в лаборатории №3 отдела №6 Института [redacted] I7 января I986г. было проведено испытание модели установки, которая была собрана по схеме, соответствующей описанию заявки на изобретение "Устройство для передачи электроэнергии" № 355 9986/07 033494.

В качестве источника питания использовалась стандартная промышленная электросеть /250В/. К изолированной установке шел только один проводник длиной 7м, сечением 0,8мм^2. В качестве накопителей зарядов использовались 4 картонные параллелепипеда /по два на каждый вид зарядов/, обмотанные металлической фольгой /размеры параллелепипедов 25x25x40см/. Коммутационные проводники были длиной приблизительно по Iм и сечением 0,8мм^2. Функцию электрических вентилей выполняли диоды марки Д247. Включенный последовательно с нагрузкой 5,IМОм микроамперметр марки М366 №I09III показывал ток около I0мкА.

Таким образом, возможность протекания тока через нагрузку /что ранее оспаривала экспертиза/ была доказана.

На испытаниях присутствовали:

Эксперты ВНИИГПЭ В.А.Круглова, К.И.Фотина
Зав.отделом [redacted] [redacted], [redacted]
От авторов: сотрудник отдела №6 [redacted] А.Ю.Иоффе

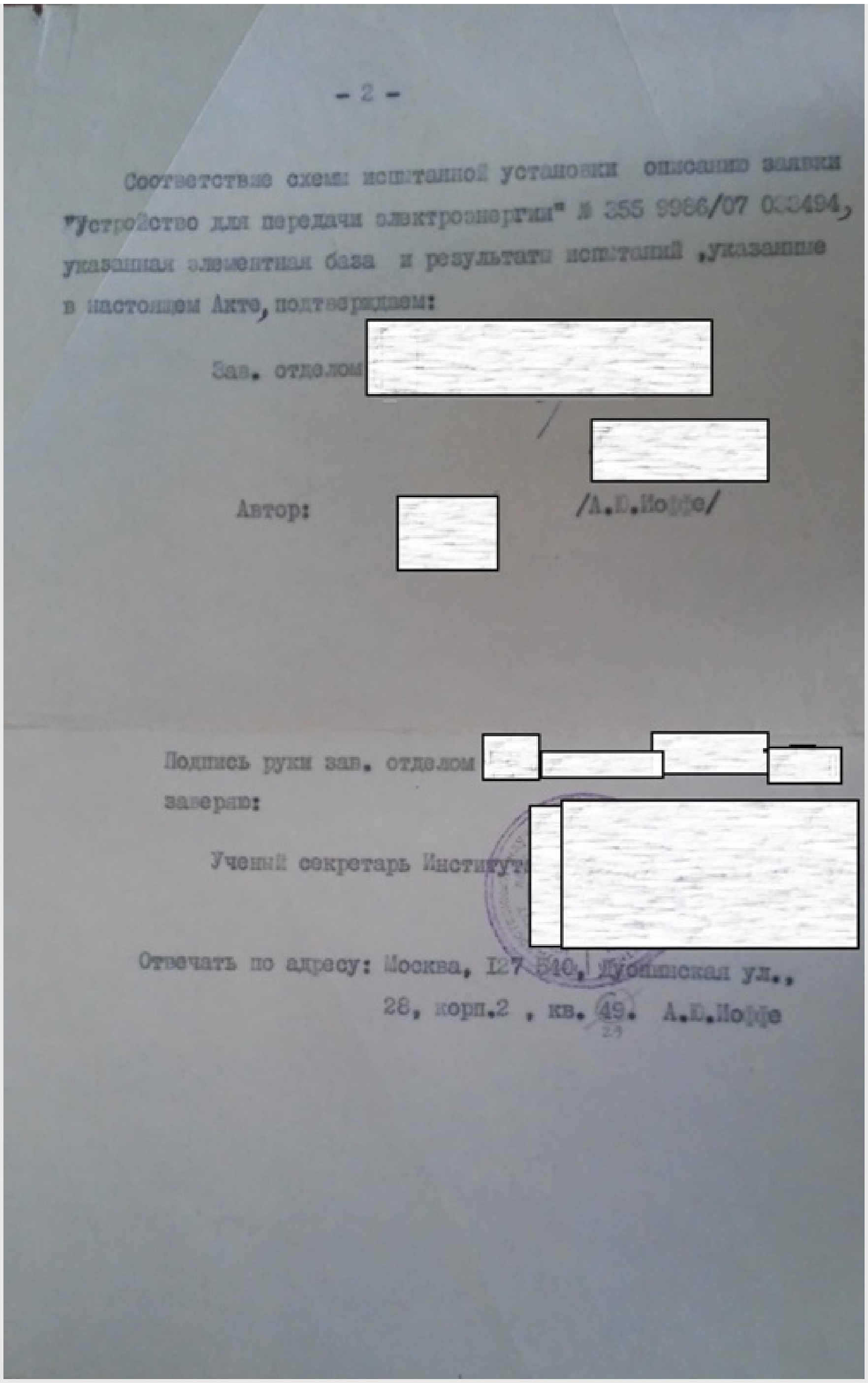

- 2 -

Соответствие схемы испытанной установки описанию заявки "Устройство для передачи электроэнергии" № 355 9986/07 033494, указанная элементная база и результаты испытаний ,указанные в настоящем Акте, подтверждаем:

Зав. отделом [redacted]

Автор: [redacted] /А.Ю.Иоффе/

Подпись руки зав. отделом [redacted] [redacted] [redacted] [redacted]
заверяю:

Ученый секретарь Институт [redacted] [redacted]

Отвечать по адресу: Москва, I27 540, Дубнинская ул.,
28, корп.2 , кв. 49. А.Ю.Иоффе
23

Translation :

CHAPTER 7. A SINGLE-WIRE POWER TRANSMISSION APPLICATIONS.

So we see that the ability to transmit electricity at one conductor to an isolated object, still weak, but still there. But there is a question, and for what purpose in general these should be engaged?

The author according to his fancy may suggest the following applications.

1. Power supply of the mobile objects contacting to the conductive firm surface / the liquid environment. It can be, for example, robots with the built-in intakes, and the floor of the machine hall will be carrying out. The single-wire electricity transmission won't be, unlike two-wire, to limit their mobility. Charging or change of accumulators won't be required. Still it can be vehicles on an electricity conductive roadbed. I will remind that I had long ago, in the USSR, article on this matter [1]. Now, when everywhere in big cities there are separate road strips for public transport, it can be real. It is only necessary to make these strips electricity conductive (for the person if only he isn't grounded" contact with such strip won't constitute big danger). It is optional to do conductive this strip on all width, for the maneuvers which aren't interrupting electricity contact of a slip ring will be in it a narrow strip enough. It is natural that these objects of a current transfer have to have the corresponding current collectors where as stores their conducting cases or part them can serve. In addition, for ensuring electric contact there have to be slip rings representing, for example, freely turning conducting castors. Water

crafts in the electrolytic environment, including salty water can also be mobile objects of a current transfer. We will imagine an artificial pond with the sea or added some salt water and with the electricity isolated bottom. On it the boats with electric motors possessing the corresponding current collectors [4] can move.

2. Information electricity transmission when it is difficult to ground or duplicate the line carrying current and information (providing the second conductor). Such situations are rare, but, nevertheless, sometimes, can meet. Then, having the corresponding equipment, by means of such electricity transmission we can transmit signals modulating its parameters. For example, it is possible to transfer information and even energy on the metal pipe which isn't contacting to the earth.

3. It is possible to try to use such way of transfer of the electric power and information on monorail roads.

4. Power supply zeppelin having conductive body (huge storage charges) by means of thin cables coming from the ground. The cable for each zeppelin will be only one. Also will be a thin wire, with the coaxial carrying-out cover intended for prevention of an exit outside of the electromagnetic radiation which can be harmful to all live and, besides, will create a radio noise. It is possible to imagine and the return electric energy when on the zeppelin solar batteries are installed on all its surface, and the cable transferring the developed energy to the earth there will be a single thin wire with a coaxial cover as such cable will be easier and won't pull the zeppelin strongly down, owing to that that the wire can be very thin at a single-wire electricity transmission.

5. Transfer of the electric power on such single-wire lines to the isolated objects will save metal of electricity cables, and not because a wire only one, after all single-wire lines, are already applied to the grounded objects, and because wires can be made thin owing to other mechanism of conductivity which hypothesis of action is given above in Chapter 1. However, as reception objects it will be necessary to build buildings of the big sizes with the electricity conductive cases which, however, can be used in economy, say, as warehouses. Wires, as it was noted above, it is necessary to envelope in the shielding cover. And, even despite it, in view of the small diameter of such coaxial cables, the economy of metal is possible.

6. Transfer of electric power / information on the conductor or the channel having gaps. It can be both the metal conductor, and the interrupting stream of electrolyte. About possibility of a single-wire electricity transmission on not continuous conductor it is possible to read in work [4]. Besides, it is confirmed by a series of experiments... in this work.

7. Transmission of electricity / information between two crafts / vehicles upon contact of their bodies. The shells must be conductive, but not over the entire surface, otherwise we will encounter the "Faraday cage" effect.

8. Single-wire transmission of electricity in the plasma channel formed in the air by means of a laser beam. For information, energy and military purposes.

9. For medicine. For diagnostic devices.

10. Disinfection conductive surfaces. Nikola Tesla discovered that high frequency currents and high voltages have a bactericidal effect. For some conductive surfaces, especially hard-moving and rotating, bactericidal transfer currents may be more convenient for just one conductor and not two.

LITERATURE.

1. Ioffe A.Yu. "electric highway". Magazine "Inventors and Innovators." Number 10 for the year 1989. Moscow.
2. Article on the Internet: Kosinov NV "Experiments on wireless power transmission: CONFIRMATION revolutionary ideas Tesla' http://kosinov.314159.ru/kosinov31.htm
3. SV Avramenko , VN Lisin, Zaev NE Single-wire transmission line (fb2) Librusec http://lib.rus.ec/b/209819/read
4. http://loricsin.livejournal.com/70594.html
5. A.Yu.Ioffe, B.P.Kirsanov "method of transmitting electricity" invention application number 3351879/11 159 868 filed in the USSR VNIIGPE November 9, 1981.
6. A.Yu.Ioffe, B.P.Kirsanov "An apparatus for transmitting electricity" invention application number 3559986/07 033,494 filed in the USSR VNIIGPE February 28, 1983.
7. Strebkov DS, Nekrasov A.M. Resonance methods of electric energy transmission. - M .: GNU VIESH, 2004. - 188 p.
8. SV Avramenko 1-Wire Power Transmission - fantasy or reality? // Electricity and life. - 2000. - N 3. - S.18-27. - Bibliography .: 8 titles.
9. http://zaryad.com/tag/avramenko/
10. https://geektimes.ru/post/81513/
11. http://www.vitanar.narod.ru/AVR/AVR.htm

Bellow are photocopies of:
requests for registration of inventions and the article in the Magazine "Inventors and Innovators"

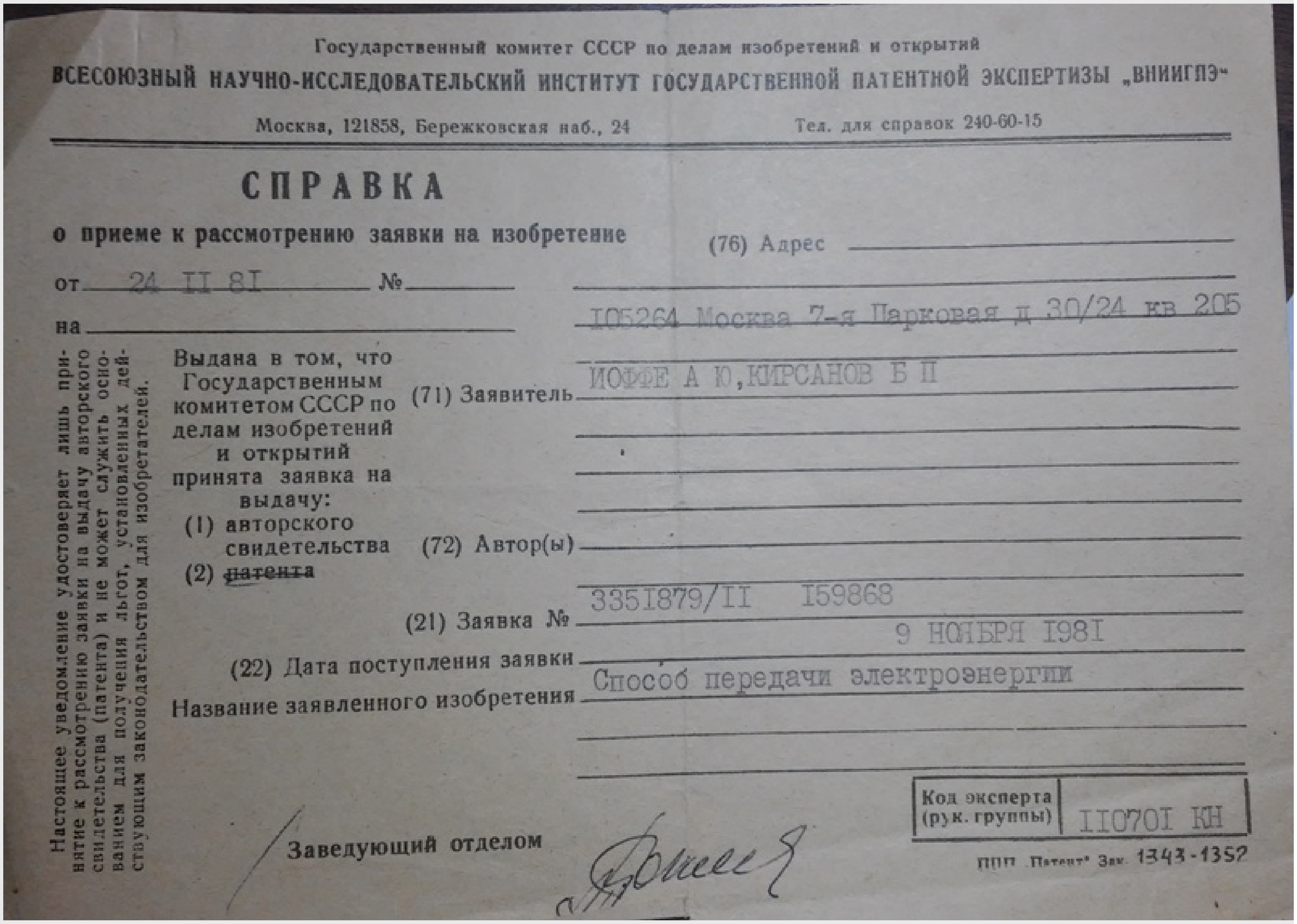

Государственный комитет СССР по делам изобретений и открытий

ВСЕСОЮЗНЫЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ ГОСУДАРСТВЕННОЙ ПАТЕНТНОЙ ЭКСПЕРТИЗЫ „ВНИИГПЭ“

Москва, 121858, Бережковская наб., 24 Тел. для справок 240-60-15

СПРАВКА

о приеме к рассмотрению заявки на изобретение

от 24 II 8I №

на

Настоящее уведомление удостоверяет лишь принятие к рассмотрению заявки на выдачу авторского свидетельства (патента) и не может служить основанием для получения льгот, установленных действующим законодательством для изобретателей.

Выдана в том, что Государственным комитетом СССР по делам изобретений и открытий принята заявка на выдачу:

(1) авторского свидетельства

(2) ~~патента~~

(76) Адрес I05264 Москва 7-я Парковая д 30/24 кв 205

(71) Заявитель ИОФФЕ А Ю,КИРСАНОВ Б П

(72) Автор(ы)

(21) Заявка № 3351879/II I59868

(22) Дата поступления заявки 9 НОЯБРЯ I98I

Название заявленного изобретения Способ передачи электроэнергии

Код эксперта (рук. группы)	II070I КН

Заведующий отделом

ППП „Патент“ Зак. 1343-1352

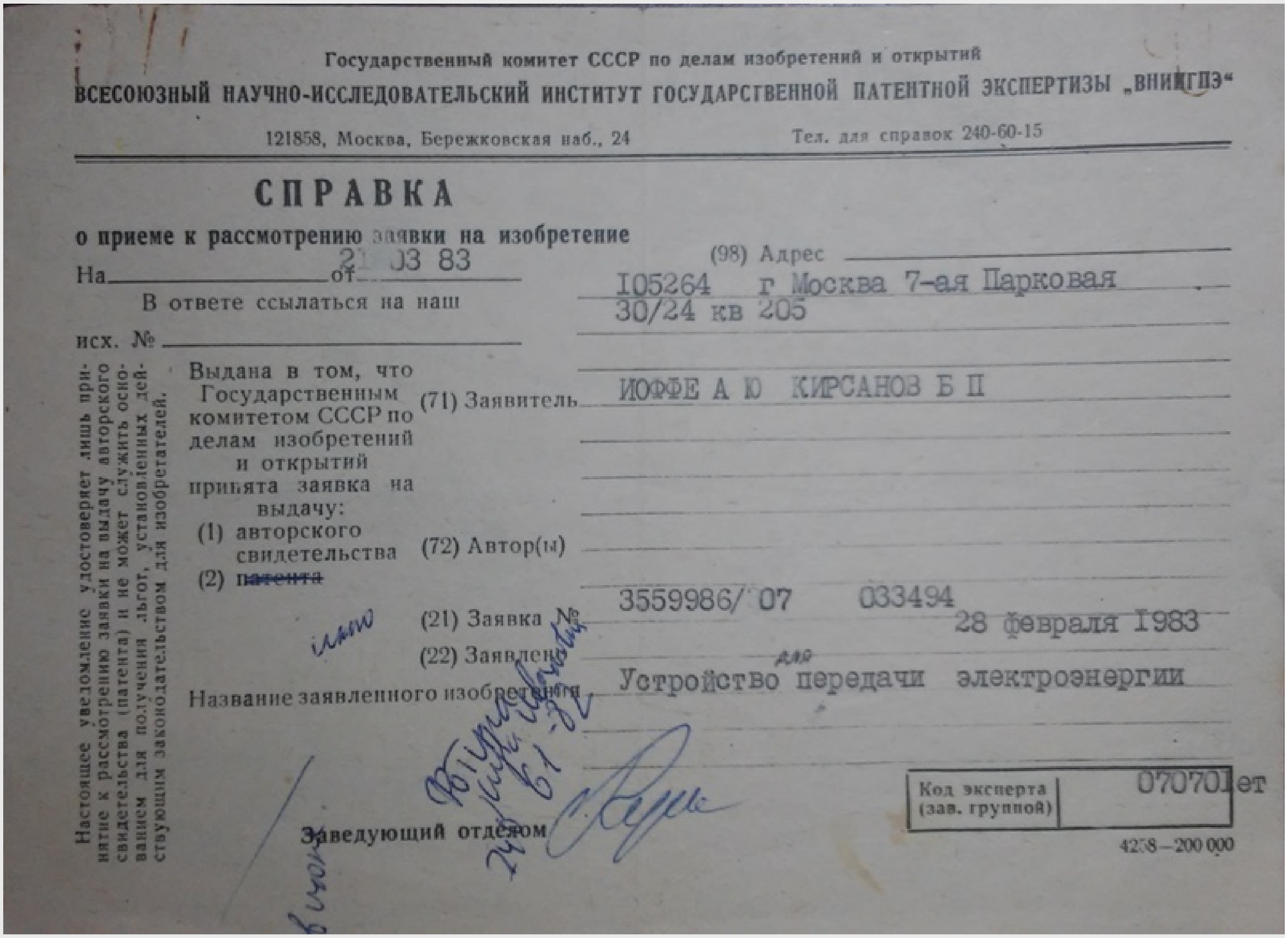

Государственный комитет СССР по делам изобретений и открытий

ВСЕСОЮЗНЫЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ ГОСУДАРСТВЕННОЙ ПАТЕНТНОЙ ЭКСПЕРТИЗЫ „ВНИИГПЭ“

121858, Москва, Бережковская наб., 24 Тел. для справок 240-60-15

СПРАВКА

о приеме к рассмотрению заявки на изобретение

На ______ от 2I 03 83

В ответе ссылаться на наш

исх. № ______

(98) Адрес I05264 г Москва 7-ая Парковая 30/24 кв 205

Выдана в том, что Государственным комитетом СССР по делам изобретений и открытий принята заявка на выдачу:

(1) авторского свидетельства

(2) п~~атента~~

(71) Заявитель ИОФФЕ А Ю КИРСАНОВ Б П

(72) Автор(ы)

(21) Заявка № 3559986/07 033494

(22) Заявлено 28 февраля I983

Название заявленного изобретения Устройство для передачи электроэнергии

Код эксперта (зав. группой) 07070Iет

Заведующий отделом

4258—200 000

Настоящее уведомление удостоверяет лишь принятие к рассмотрению заявки на выдачу авторского свидетельства (патента) и не может служить основанием для получения льгот, установленных действующим законодательством для изобретателей.

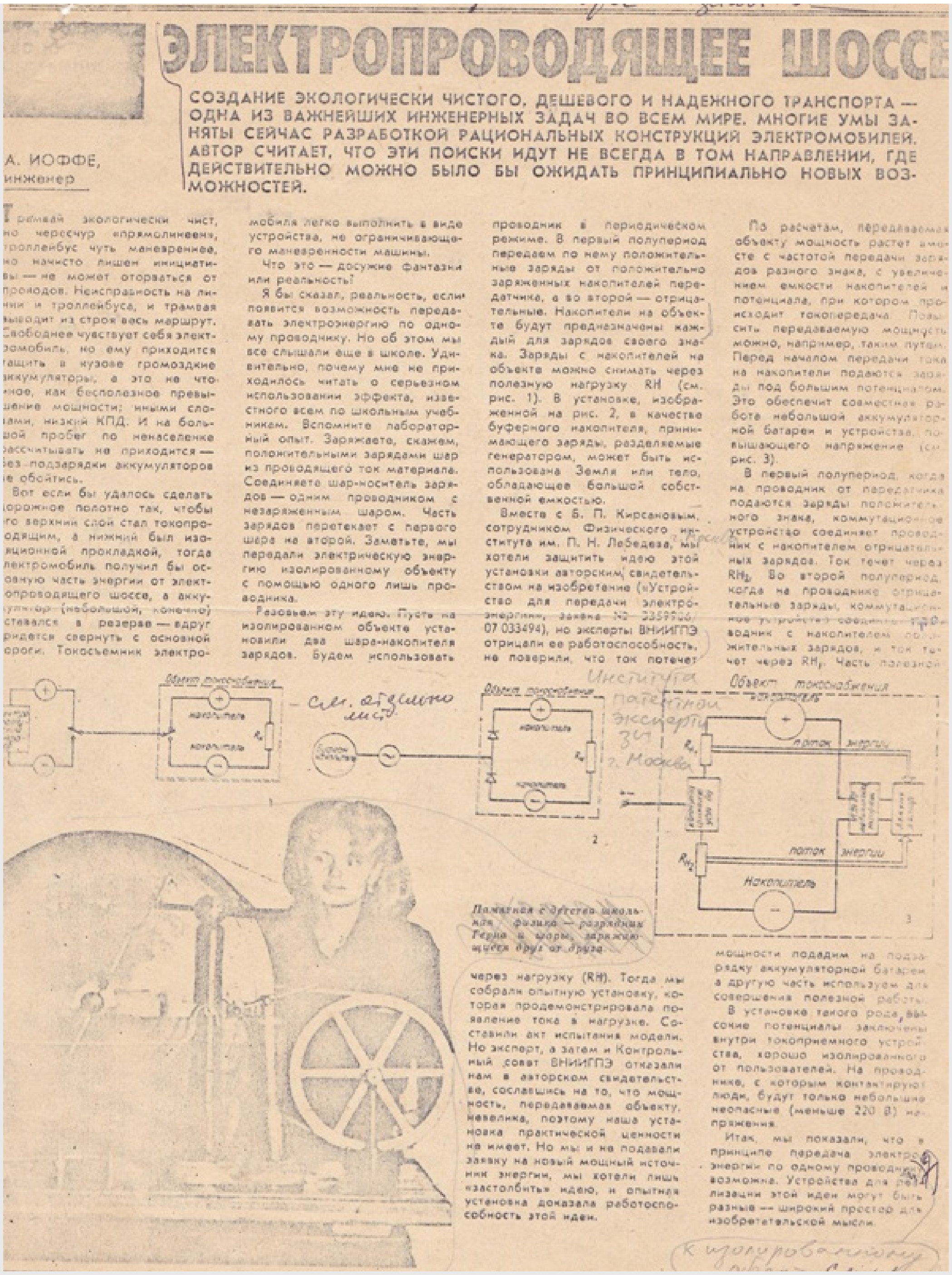
Объект токосъёмщения
накопитель
накопитель
R_1
— см. отдельно лист
Института патентной экспертизы 34 г. Москва
2
Объект токосъёмщения
поток энергии
RH_2
Накопитель
3
К изолированному

ЭЛЕКТРОПРОВОДЯЩЕЕ ШОССЕ

СОЗДАНИЕ ЭКОЛОГИЧЕСКИ ЧИСТОГО, ДЕШЕВОГО И НАДЕЖНОГО ТРАНСПОРТА — ОДНА ИЗ ВАЖНЕЙШИХ ИНЖЕНЕРНЫХ ЗАДАЧ ВО ВСЕМ МИРЕ. МНОГИЕ УМЫ ЗАНЯТЫ СЕЙЧАС РАЗРАБОТКОЙ РАЦИОНАЛЬНЫХ КОНСТРУКЦИЙ ЭЛЕКТРОМОБИЛЕЙ. АВТОР СЧИТАЕТ, ЧТО ЭТИ ПОИСКИ ИДУТ НЕ ВСЕГДА В ТОМ НАПРАВЛЕНИИ, ГДЕ ДЕЙСТВИТЕЛЬНО МОЖНО БЫЛО БЫ ОЖИДАТЬ ПРИНЦИПИАЛЬНО НОВЫХ ВОЗМОЖНОСТЕЙ.

А. ИОФФЕ,
инженер

Трамвай экологически чист, но чересчур «прямолинеен», троллейбус чуть маневреннее, но начисто лишен инициативы — не может оторваться от проводов. Неисправность на линии и троллейбуса, и трамвая выводит из строя весь маршрут. Свободнее чувствует себя электромобиль, но ему приходится тащить в кузове громоздкие аккумуляторы, а это не что иное, как бесполезное превышение мощности; иными словами, низкий КПД. И на большой пробег по ненаселенке рассчитывать не приходится — без подзарядки аккумуляторов не обойтись.

Вот если бы удалось сделать дорожное полотно так, чтобы его верхний слой стал токопроводящим, а нижний был изоляционной прокладкой, тогда электромобиль получил бы основную часть энергии от электропроводящего шоссе, а аккумулятор (небольшой, конечно) оставался в резерве — вдруг придется свернуть с основной дороги. Токосъемник электромобиля легко выполнить в виде устройства, не ограничивающего маневренности машины.

Что это — досужие фантазии или реальность?

Я бы сказал, реальность, если появится возможность передавать электроэнергию по одному проводнику. Но об этом мы все слышали еще в школе. Удивительно, почему мне не приходилось читать о серьезном использовании эффекта, известного всем по школьным учебникам. Вспомните лабораторный опыт. Заряжаете, скажем, положительными зарядами шар из проводящего ток материала. Соединяете шар-носитель зарядов — одним проводником с незаряженным шаром. Часть зарядов перетекает с первого шара на второй. Заметьте, мы передали электрическую энергию изолированному объекту с помощью одного лишь проводника.

Разовьем эту идею. Пусть на изолированном объекте установили два шара-накопителя зарядов. Будем использовать проводник в периодическом режиме. В первый полупериод передаем по нему положительные заряды от положительно заряженных накопителей передатчика, а во второй — отрицательные. Накопители на объекте будут предназначены каждый для зарядов своего знака. Заряды с накопителей на объекте можно снимать через полезную нагрузку RH (см. рис. 1). В установке, изображенной на рис. 2, в качестве буферного накопителя, принимающего заряды, разделяемые генератором, может быть использована Земля или тело, обладающее большой собственной емкостью.

Вместе с Б. П. Кирсановым, сотрудником Физического института им. П. Н. Лебедева, мы хотели защитить идею этой установки авторским свидетельством на изобретение («Устройство для передачи электроэнергии», заявка № 3359506, 07 033494), но эксперты ВНИИГПЭ отрицали ее работоспособность, не поверили, что ток потечет через нагрузку (RH). Тогда мы собрали опытную установку, которая продемонстрировала появление тока в нагрузке. Составили акт испытания модели. Но эксперт, а затем и Контрольный совет ВНИИГПЭ отказали нам в авторском свидетельстве, сославшись на то, что мощность, передаваемая объекту, невелика, поэтому наша установка практической ценности не имеет. Но мы и не подавали заявку на новый мощный источник энергии, мы хотели лишь «застолбить» идею, и опытная установка доказала работоспособность этой идеи.

По расчетам, передаваемая объекту мощность растет вместе с частотой передачи зарядов разного знака, с увеличением емкости накопителей и потенциала, при котором происходит токопередача. Повысить передаваемую мощность можно, например, таким путем. Перед началом передачи тока на накопители подаются заряды под большим потенциалом. Это обеспечит совместная работа небольшой аккумуляторной батареи и устройства, повышающего напряжение (см. рис. 3).

В первый полупериод, когда на проводник от передатчика подаются заряды положительного знака, коммутационное устройство соединяет проводник с накопителем отрицательных зарядов. Ток течет через RH_2. Во второй полупериод, когда на проводнике отрицательные заряды, коммутационное устройство соединяет проводник с накопителем положительных зарядов, и ток течет через RH_1. Часть полезной мощности подадим на подзарядку аккумуляторной батареи, а другую часть используем для совершения полезной работы.

В установке такого рода высокие потенциалы заключены внутри токоприемного устройства, хорошо изолированного от пользователей. На проводнике, с которым контактируют люди, будут только небольшие неопасные (меньше 220 В) напряжения.

Итак, мы показали, что в принципе передача электроэнергии по одному проводнику возможна. Устройства для реализации этой идеи могут быть разные — широкий простор для изобретательской мысли.

Памятная с детства школьная физика — разрядник Герца и шары, заряжающиеся друг от друга.

- Оставить комментарий
- © Copyright Иоффе Александр Юльевич (ioffe.shura@Yandex.ru)
- Размещен: 17/04/2015, изменен: 20/08/2017. 90k. Статистика.
- Статья: Foreign+Translat

Ваша оценка: не читал ▼ OK

РЕКЛАМА: популярное на LitNet.com **Э.Ридлин** "Сердце подскажет" (Любовное фэнтези) | | **К.Лестова** "Без меня на мне женились" (Городское фэнтези) | | **Р.Вольная** "Одна из тысячи звезд" (Современный любовный роман) | | **K.Favea** "21 ночь" (Романтическая проза) | | **К.Невестина** "Брачная охота на главу тайной канцелярии" (Юмористическое фэнтези) | | **Д.Мар** "Когда я тебя найду" (Приключенческое фэнтези) | | **С.Шавлюк** "Я с тобой не останусь" (Романтическая проза) | | **Л.Свадьбина** "Попаданка в академии драконов 4" (Приключенческое фэнтези) | | **К.Ши** ""Муж" на час" (Короткий любовный роман) | | **А.Тарасенко** "Демон для попаданки" (Попаданцы в другие миры) | |

Связаться с программистом сайта.

Новые книги авторов СИ, вышедшие из печати:
Д.Смекалин "Ловушка архимага" Е.Шепельский "Варвар,который ошибался" В.Южная "Холодные звезды"

ATMOSPHERIC ELECTRICITY—B. F. J. SCHONL

9515·06

METHUEN'S MONOGRAPHS ON PHYSICAL SUBJECTS

ATMOSPHERIC ELECTRICITY

B. F. J. SCHONLAND

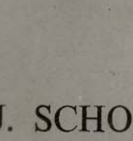

METHUEN'S MONOGRAPHS ON PHYSICAL SUBJECTS

General Editor: B. L. WORSNOP, B.Sc., Ph.D.

F'cap 8vo. **Illustrated**

THIS series is intended to supply science students at University level with a compact statement of the modern position in each subject. The Honours student and the research worker in other branches of physics, those engaged on work in related sciences, and those who are no longer in contact with active scientific work, will find here a series of expositions by authors who are actively engaged in research on the subject of which they write.

For list of series, see back of this jacket

METHUEN & C NDON

8/6
NET

Cat. No. 4053/U . net

(£1 =

CHRIST'S COLLEGE
LIBRARY

Methuen's Monographs on Physical Subjects
General Editor : B. L. WORSNOP, B.Sc., Ph.D.

ATMOSPHERIC ELECTRICITY

ATMOSPHERIC ELECTRICITY

BY

B. F. J. SCHONLAND

C.B.E., F.R.S.
DIRECTOR OF THE BERNARD PRICE INSTITUTE OF GEOPHYSICAL RESEARCH
UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

WITH 20 DIAGRAMS

LONDON: METHUEN & CO. LTD.
NEW YORK: JOHN WILEY & SONS, INC.

First published, November 1932
Second edition, revised and reset, 1953

WITHDRAWN
CHRIST'S COLLEGE
LIBRARY
ACCESSION NO.
14097
CLASS NO. LIHE STACK
910·5156
Cat Off
2 10 69

CATALOGUE NO. 4053/U (METHUEN)

PRINTED IN GREAT BRITAIN

TO
C. T. R. WILSON

CONTENTS

CHAP.		PAGE
I.	THE IONISATION OF THE ATMOSPHERE	1
II.	ELECTRIC FIELDS AND ELECTRIC CURRENTS IN THE ATMOSPHERE	19
III.	THE ELECTRIFICATION OF THUNDERCLOUDS	46
IV.	THE LIGHTNING DISCHARGE	65
V.	CHARGE GENERATION IN CLOUDS	80
	GENERAL REFERENCES	90
	SUBJECT INDEX	91
	INDEX OF NAMES	94

CHAPTER I

THE IONISATION OF THE ATMOSPHERE

The electrical conductivity of the air, first established by Linss in the year 1887, plays an important part in all the phenomena included under the title Atmospheric Electricity. Since 1901, when Elster and Geitel and C. T. R. Wilson independently discovered that this conductivity is due to the presence of small carriers of positive and negative charge called ions, the nature and the origin of the ions have been the subject of many investigations. Two different paths have been followed, the one dealing with the number and the nature of the atmospheric ions, the other with the radiations producing them. The two paths converge on the problem of the ionisation-balance of the atmosphere, the manner in which an equilibrium is maintained between the rate of creation and the rate of disappearance of the ions. It is only of recent years that an answer to this problem has emerged.

1. The Atmospheric Ions, Small and Large. A gas molecule is ionised when some external agency supplies the energy necessary to remove an electron from it, whereupon the positively charged residue of the molecule and the ejected electron, both of which under ordinary pressures very soon attach themselves to one or more uncharged molecules, form what is called an ion-pair. Such ions are known as "normal", "small", or "fast", and when found in the atmosphere have mobilities of the same order as those determined for pure dry air in laboratory experiments, about 1·5 cm./sec. in a field of gradient 1 v./cm.

The presence in the air of minute particles of dust and salt, of the products of combustion, and of extremely small drops of water even when the air is unsaturated, leads very often to the capture of a normal ion by one of these condensation or Aitken nuclei, as they are termed, and so to the formation of another type of ion of low mobility,

CHRIST'S CO

ranging from 0·008 to 0·0003 cm./sec., known as "large", "slow", or "Langevin" ions. Roughly speaking, one-third of the condensation nuclei present in the air is electrically neutral, the other two-thirds carry, nearly equally, positive and negative charges captured from small ions. The interplay between the large and the small ions and the uncharged nuclei is of considerable importance.

Besides these two main types of atmospheric ion, a group of "intermediate" ions whose mobilities lie between 0·2 and 0·01 cm./sec. has been found to be present under special meteorological conditions such as low relative humidity. Some of these are known to be associated with molecular clusters of sulphuric acid from industrial processes.[1]

It has been found that the great majority of the ions, large and small, carry a single elementary charge, either positive or negative.[2] Over land areas the large ions considerably outnumber the small ones, often in the ratio of ten to one. Thus while the numbers, n_+ and n_-, of small ions of each sign per c.c. of the air over land usually range from 300 to 1000, those of the large ions, N_+ and N_-, amount to from 1000 to 80,000 per c.c. Since practically all the large ions are formed from the capture of small ones by nuclei, an increase in the number of these nuclei, such as is due to the fires of an industrial area, will increase the proportion of the large ions in the air at the expense of the small ions.

Over the oceans this state of affairs is reversed; the air away from land areas is but poorly supplied with nuclei and the large ions are in the minority. The mean values of n_+ and n_- at sea lie between 500 and 700 per c.c., while N_+ and N_- are about 200 per c.c. In view of their comparative freedom from capture by nuclei, it is at first sight rather surprising that the number of small ions per c.c. in mid-ocean differs little from the same quantity over the land, where the average values of n_+ and n_- are also about 600 per c.c. This is, however, to be explained (§ 12) by the fact that the decreased rate of disappearance

of the small ions over the sea is balanced by a decrease in the rate at which they are produced.

2. The Determination of the Number of Ions per c.c. To determine the numbers of the small ions, positive and negative, per c.c. (n_+ and n_-) an arrangement known as an Ebert ion-counter is usually employed. This consists in essentials of a long earthed metal tube A (Fig. 1) on the axis of which is mounted a charged insulated rod B, connected to a quartz fibre electroscope. Air is sucked through the tube for about five minutes by means of a clock-work driven fan F and the speed of the air-stream is arranged to be slow enough for all the small ions entering A with charges opposite in sign to that on the central rod to be drawn to B before reaching the bottom of the tube. The speed of flow must at the same time be great enough to make it impossible for any appreciable number of large ions to have time to reach the rod. If v is the volume of the air sucked through in the course of the measurement (indicated by an anemometer C registering its revolutions on a dial), the quantity of electricity gathered in by the central electroscope system is ven_-, where e is the elementary charge on a single ion and the rod

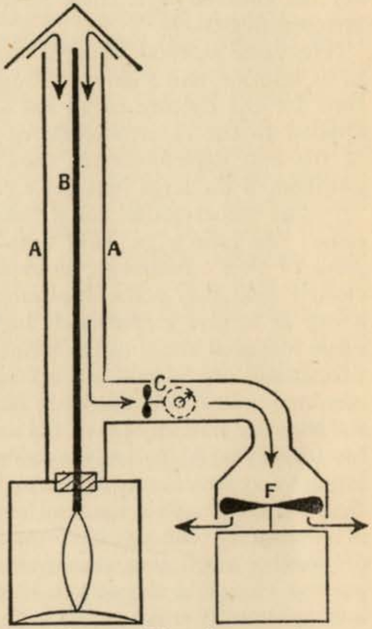

Fig. 1.—Ebert ion-counter.

is supposed to have been positively charged. If then C is the capacity of the whole central system, rod and electroscope, and V and V' are the potentials indicated by the electroscope before and after the experiment,

$$C(V - V') = ven_-,$$

an equation from which n_- can be found. To determine n_+, the measurement must be repeated with the rod B charged negatively.

The same method, using a longer tube, a stronger field in its interior, and a slower draught of air, and thus giving time for the capture of all the slow ions as well, can be applied to the determination of $N + n$. A combination of the two determinations then yields N_+ and N_-, the numbers of the large ions per c.c. of each sign.

The values found for n_+ and n_- are not usually the same; the ratio n_+/n_- over both land and sea has a mean value of 1·22. Although, as explained in § 22, the earth's electric field may sometimes affect the apparatus in such a way as to give a spuriously high ratio, Swann, who has made a special study of the point, concludes that when all precautions are taken a real excess of positive ions must be admitted. Probably the main factor causing this excess is the superior mobility (k) of the small negative ion; Nolan has found that k_-/k_+ has the value 1·16. A lower mobility implies a lower velocity of thermal agitation for the positive than for the negative ion, and consequently less likelihood of capture by condensation nuclei. This general excess of positive small ions does not necessarily involve a net positive charge in the air; it is easily compensated for by a comparatively small excess in the much larger number of large negative ions.

An instrument known as the Aitken dust-counter is often used to find the total number of nuclei, charged and uncharged, per c.c. It consists of a small expansion chamber filled with a sample of the air and saturated with water-vapour. The sudden movement of a piston gives an adiabatic expansion which causes drops of water to condense upon the nuclei; the resulting cloud falls upon a glass plate divided into squares and the number of droplets is counted under a microscope. Though called a dust-counter it does not record the coarser dust in the air.[3]

In a recent improvement on the Aitken counter (Nolan [3]) a much larger sample of saturated air is expanded in a cylindrical vessel. The opacity of the cloud then gives a relative measurement of the concentration of nuclei.

3. The Polar Conductivities and the Ionic Mobilities. A simple relation holds between the conductivity of the air and the numbers per c.c. and the mobilities of the ions responsible for conduction. Consider a point in the atmosphere at which the electric field intensity is F and the specific conductivity, the reciprocal of the specific resistance, is λ. The conduction current, i, through unit area drawn perpendicular to the field is then equal to $F\lambda$, for F is numerically equal to the potential difference, and $1/\lambda$ is the resistance, between the ends of a unit cube of the air. In this equation we assume that conditions are such that Ohm's Law can be applied to the air. Since the current is carried by positive and negative streams of ions moving in opposite directions we may write

$$i = F(\lambda_+ + \lambda_-), \quad . \quad . \quad . \quad (1.1)$$

where λ_+ and λ_- are called the polar conductivities.

Actually the existence of the two main types of ion, large and small, leads to the flow of four different ion-streams, two up and two down if the field is vertical, in each of which the velocity is different, being given by $F \times k$, where k is the corresponding mobility. The total current is thus

$$i = Fe(n_+k_+ + n_-k_- + N_+K_+ + N_-K_-), \qquad (1.2)$$

where k_+ and k_-, K_+ and K_-, are the mobilities of the positive and the negative ions, large and small respectively, and e is the elementary charge. From (1.1) and (1.2), if the oppositely moving streams are separated,

$$\lambda_+ = e(k_+n_+ + K_+N_+) \text{ and } \lambda_- = e(k_-n_- + K_-N_-). \quad (1.3)$$

In clear country air over land, n_+ and n_- number about 600 per c.c., while N_+ and N_- are about 2000 per c.c.;

k_+ and k_- are 1·5 approximately and K_+ and K_-, 0·0004. Here, therefore, the ratio of the first to the second term in each bracket above is 1100 : 1 and most of the conductivity of the air is due to the small ions. Near large towns, however, n is usually of the order of 100 per c.c. and N, on account of the heavy pollution of the atmosphere may be 50,000. Under these conditions the second term would be one-seventh of the first and the large ions would play a part in the transfer of electricity through the air.*

Nolan and Nolan have obtained the polar conductivities by separate measurement of the n's and k's in eqn. (1.3) above. It is, however, more usual to find λ directly, from measurements of the rate which the charge on an insulated body exposed to the free air is dissipated by the flow of ions to it. Consider a negatively charged body exposed to moving air and let σ be the density of the electrification on any portion of area dS. Positive ions will move to the body, and since the field F just outside the surface is $4\pi\sigma$, the current per unit area tending to dissipate the charge will be $F\lambda_+ = 4\pi\sigma\lambda_+$. The total dissipation current, obtained by integrating over the whole exposed surface, is $4\pi\lambda_+\iint \sigma dS$ or $4\pi\lambda_+ Q$, where Q is the charge exposed to the air. Since this current represents the rate at which the body losses charge, we have $-dQ/dt = 4\pi\lambda_+ Q$, a relation which provides a means of determining λ_+. A similar equation, with λ_- substituted for λ_+, holds for the case of a positively charged body.[4]

An arrangement used for this purpose is due to Gerdien (Fig. 2); it resembles the Ebert apparatus already described, but the air-stream is now made so rapid and the field in the interior of the tube so weak, that only a very small fraction of the total number of ions passing through A has time to reach the central rod. In this case the numbers of ions per c.c. are not appreciably altered by the experiment and Ohm's Law applies.

---

* The methods used to determine ionic mobilities are described in text-books on the conduction of electricity through gases.

If the potential V of the negatively charged central electroscope system is observed to rise at a rate dV/dt and if C is the total capacity of the central system, rod and electroscope, we have $-dQ/dt = -CdV/dt$. Also, if C' is the capacity of that portion of the central system exposed to the stream of air, the rod B and its support, $Q = C'V$, consequently

$$-CdV/dt = 4\pi\lambda_+ C'V, \quad . \quad . \quad (1.4)$$

a relation from which λ_+ can be determined. A similar experiment with the rod positively charged yields λ_-. The use made of the values found for the polar conductivities and an adaptation of the Gerdien method due to Schering are described in Chapter II, § 22.

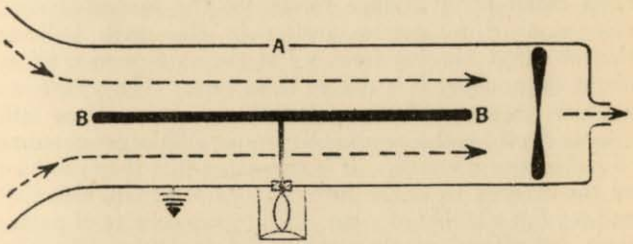

FIG. 2.—Gerdien conductivity apparatus.

4. The Ionising Agencies. Three principal agencies are responsible for the ionisation of the lower atmosphere: radiations from radioactive substances in the surface of the earth, radiations from radioactive matter present in the air itself, and the cosmic rays. For a discussion of their relative importance it is convenient to use the symbol q to represent the number of "pairs of ions produced per c.c. per second in air at N.T.P.", a quantity which will be referred to as the ionising power of the radiation. This phrase in inverted commas is usually represented by the symbol I. Thus the statement $q = 4I$ means that, at the point in question, the radiations create four pairs of ions per c.c. per second, if the medium there is air at N.T.P. The contributions of the three different agencies just mentioned to the total ionising power, q, are distinguished by the symbols q_e, q_a, and q_h.

5. The Effect of Radioactive Matter in the Earth. Uranium, thorium, and their products are widely distributed through the earth's crust, whose surface layers therefore emit α, β, and γ rays into the air. Since the α rays can emerge only from a very thin layer and can ionise only the first few cm. of the air, their effect is inconsiderable. The β rays can come from greater depths and penetrate much further into the atmosphere. Calculations based upon average values for the measured concentration of radioactive matter in the earth indicate that an effect ranging from 1 I at the surface to 0·1 I at about 10 m. above it is due to these rays. The γ rays are a more important factor, for they can come from still greater depths and consequently from a still larger quantity of radioactive material. It is estimated that they produce on the average 3 I at the surface, 1·5 I at 150 m. above it, and 0·3 I at a height of 1 km. The γ ray activity of potassium introduces another source of ionisation, for the potassium content of the earth's crust is considerable. No very detailed survey of this question has yet been made, but measurements indicate that the contribution of potassium may exceed that of the combined uranium-radium and thorium families. The average value for the ionising power of the earth-radiations near the ground has been found directly to lie between 2 and 10 I, which is in accord with the measurements of the radioactive content of the earth's crust. Local surface and geological conditions naturally have a marked effect upon the value of q_e at any given spot.

The radioactive content of sea-water is found to be very small in comparison with that of soil and rocks, and over the oceans the earth-radiation effect is negligible.

6. The Effect of Radioactive Matter in the Air. The atmosphere contains a considerable quantity of the radioactive emanations, radon and thoron, and their successive products. A wire charged negatively to a potential of a few hundred volts and exposed to the air

collects an easily detectable quantity of radium A and thorium A by recoil from the disintegration of these emanations. The gases themselves arise from the decay of radium and thorium within the earth, from which they escape by diffusion, by thermal convection, and as a result of decreases in the external atmospheric pressure. The greater portion of the ionisation thus produced is due to the α rays from the emanations themselves and from their A and C products, for here, unlike the case of the radiations from the earth, there is no absorbing layer between the disintegrating atom and the air. The radioactive gases and their products are fairly evenly distributed by atmospheric turbulence through the lower air. This "air-radiation" effect can be calculated from radioactive data if the emanation content of the air is determined. The emanations may be removed from a measured volume of air by drawing it through tubes cooled with liquid air or filled with coconut charcoal. The quantity of emanation condensed or absorbed is then determined by the effect produced in an ionisation electroscope, corrected for the decay of the material during the time of the experiment. A more direct method is to fill a closed ionisation chamber first with ordinary air and then with air which has been freed by the above method from the emanations and from their products, and to determine the saturation currents in the two cases. If these are i_1 and i_2 respectively, v the volume of the vessel and e the elementary charge, the value of the air-radiation ionising power, q_a, is given by $i_1 - i_2 = q_a ve$.

The results of such measurements agree with estimates based upon observations of the average emanation content of the air in placing the air-radiation effect at about 2 I in the neighbourhood of the earth's surface. Of this about 55 per cent. is due to radon and its products and the remainder to the thoron series. It varies somewhat with locality and with those meteorological conditions which determine the rate of escape of the emanations. Apart from the effect of strong upward winds in elevating the richer surface layers, the emanation content of the air should decrease fairly rapidly with height. Direct information on this point is difficult to obtain, but what exists indicates that the air-radiation effect falls to less than 2 per cent. of its surface value at a height of 5 km.

The effect, like that of the earth-radiation, is extremely small over the oceans, where the emanation content of the air is only about 1 per cent. of its value over land areas.

7. The Effect of the Cosmic Rays. The third factor in the ionisation of the atmosphere is a radiation of peculiar interest, whose origin and nature are still the subject of much investigation. It has a penetrating power which is considerably greater than that of the most penetrating γ rays from radioactive bodies, and it travels to the earth's surface in a downward direction, undergoing a certain amount of absorption on the way. Its ionising power at sea-level over both land and ocean areas when measured in closed metal vessels varies from about $2{\cdot}0\, I$ in low geomagnetic latitudes[6] to $1{\cdot}5\, I$ near the magnetic equator, the earth's field preventing the slower ionising particles from reaching the equatorial belt.

Until the year 1911 it was generally thought that the ionisation inside a closed vessel was entirely due to radiations from the walls and from radioactive matter in the earth and the outside air. It was therefore to be expected that this ionisation would decrease when the vessel was raised above the surface of the earth. This was tested in balloon ascents by Hess in 1911, and Kolhörster in 1913, who found that the diminution with height was appreciable only during the first kilometre, after which an increase occurred. These observations led Hess to postulate an ionising radiation coming from above and suffering partial absorption in the atmosphere.

The ionisation due to this "cosmic" radiation (so-called because it is now known to originate outside the atmosphere and to have little connection with the sun) increases rapidly with height above sea-level, rising tenfold in the first 5 km.

8. The Ionisation Inside a Closed Vessel. It is a matter of considerable difficulty to determine separately the ionising powers of the earth-radiation and of the cosmic rays. Calculations of the former must be based upon measurements of the radioactive content of the earth's crust, and these are necessarily tedious and rough. To proceed further, one must examine the ionisation produced in the air inside a sealed metal chamber under conditions which are modified from those existing in the free air by the presence of the metal walls. This involves correction to free-air conditions. A further disadvantage is that a new ionising agency is introduced, the radioactive matter in and on the inside walls. The ionising power of this "wall-effect", or the "residual ionisation" of the vessel as it is called, is a constant depending on the material of which the walls are made and the cleanliness of the inner surface. It is usually represented by the symbol q_0.

The total ionising power of all the radiations operating upon the gas in a closed vessel may be called q' and the modified effects of the three main ionising agencies q'_e, q'_a and q'_h; we then have

$$q' = q'_e + q'_a + q'_h + q_0. \qquad \cdot \quad \cdot (1.5)$$

We may suppose the vessel to take the form of a sealed ionisation chamber filled with air at N.T.P. which has been freed from radioactive matter. The saturation current carried by the ions to the central collecting electrode is then given by $i = q've$, where v is the volume of the vessel and e the elementary charge on an ion.

Various methods have been employed to separate out the four terms on the right-hand side of eqn. (1.5) from one another. The most satisfactory of these begins by determining q_0. The instrument is taken to a large snow-fed lake whose radioactive content, like that of the sea, is negligible. When it is sunk to the depth of a few feet in such a lake, the effects of earth and air-radiations are non-existent and eqn. (1.5) becomes $q' = q'_h + q_0$; on lowering the vessel still further, q'_h diminishes as the cosmic radiation becomes absorbed in the increasing thickness of water, and it is possible to reach depths at which the saturation current is practically constant and unaffected by further sinking of the instrument. At this stage $q' = q_0$, and the residual ionisation is determined. By combining this observation with that made close to the surface of the lake, q'_h may be found.

Since the effects of the cosmic radiation and of the walls of the vessel are both independent of local conditions other than height above sea-level, they remain the same when the instrument is transported to a site on land at the same level. A measurement of the saturation current in the chamber under land conditions gives q' as stated by eqn. (1.5), and hence the sum of the earth and air effects, q'_e and q'_a can be found. In actual practice the air-radiation effect within a closed vessel is extremely small; the walls exclude α and β rays coming from outside and reduce q'_a to the ionising power of the γ rays emitted by the B and C products of the radio-active emanations. This amounts to but 4 per cent. of the effect produced by the air-radiation in the free atmosphere, in the case of a vessel with steel walls a few millimetres thick. It is therefore sufficient to determine the air-radiation effect, q'_a, by calculation based upon a knowledge of the emanation content of the air and of the nature of the walls of the vessel. The fourth factor, the effect of the earth-radiation, q'_e, is then obtained by substitution in eqn. (1.5).

To pass from these closed vessel determinations of the effect of the radioactive matter in the earth to the effect to be expected in the free air, experiments may be made with screens of different thicknesses placed around the ionisation chamber and the values thus obtained for the earth-radiation effect extrapolated to zero thickness. In doing so account has to be taken of the effect of the screens upon q'_h.

To derive the true ionising effect of the cosmic rays in the free air from observations inside a closed vessel, account must be taken of the ionisation due to secondary radiation from the walls of the vessels. For a brass ionisation chamber with walls 2·5 mm. thick Hess has shown that the secondary effect is only 0·09 I.[6] It is therefore likely that the free-air value of q_h in temperate latitudes is not far from 1·90 I.

The greater part of the residual ionisation arises from α rays and takes place close to the walls; q_0 is therefore the effect of the residual ionisation near the walls divided by the volume of the vessel.

9. Summary of Ionising Powers of Various Agencies. The table on page 14 gives a rough idea of the effects of the three main factors in the ionisation of the air at sea-level, both in closed vessels of steel, a few millimetres thick, and in the free atmosphere. The values given refer to measurements at about a metre from the surface of the earth; they are averages and, except in the case of the cosmic radiation, are subject to considerable fluctuations with local conditions.

It is noteworthy that the rate of production of ions over land areas is about four times as great as over the oceans, where the ionisation is practically all due to the action of the cosmic rays. Over land and close to the surface of the earth, these rays are much less important, but at greater heights the effects of the earth and the air-radiations fall off rapidly as we have seen, so that from a height of one or two kilometres upwards, the cosmic radiation is the chief agent in making the air a conductor.

10. The Rate of Disappearance of the Small Ions. The number of ions produced by the agencies we have discussed will not increase indefinitely as time goes on. Even in pure filtered air, free from nuclei, the small ions disappear by diffusion to the walls and by recombination with one another. Experiments have shown that when the diffusion loss can be neglected, the equation $dn/dt = q - \alpha n^2$ applies to such filtered air, n being as before the number of positive or negative small ions per c.c., q the number of pairs of ions produced per c.c. per second, and α the so-called recombination constant of small ions. In the course of time a state of equilibrium will be reached, in which the processes of creation and disappearance of ions balance one another; then $dn/dt = 0$

and the equilibrium equation $q = \alpha n^2$ holds. Applying this equation to the free air, taking $q = 7{\cdot}4\, I$ from the preceding section and $\alpha = 1{\cdot}6 \times 10^{-6}$ cm.^{3}/sec. from laboratory experiments, we find $n = 2145$. This is very much larger than the number of small ions actually observed, which is round about 600 per c.c. The discrepancy is due to neglect of the important part played by the Aitken or condensation nuclei in capturing the small ions.

TABLE I

Ionising Powers at 1 m. above the Earth's Surface, Sea-level

Ionising Agent.	Ionising Power in Ion-pairs per c.c. per sec. in Air at N.T.P.			
	Unscreened Closed Vessel.		Free Air.	
	Over Land.	Over Sea.	Over Land.	Over Sea.
Radioactive matter in the earth . .	3·5	0	3·5	0
Radioactive matter in the air . . .	0·2	0	2·0	0
Cosmic Radiation .	2·0	2·0	1·9	1·9
	Total ionising power for the free air .		7·4	1·9

This question has been examined over many years, principally by Nolan and his co-workers in Dublin, and it is found to be extremely complex.[3] Separate considerations must be given to the capture of small ions by charged nuclei of opposite sign, which leads to neutralisation, and by uncharged nuclei, which leads to the formation of large ions.

For the small ions, making no distinction of sign, we write :

$$dn/dt = q - \alpha n^2 - \eta_0 n N_0 - \eta n N \qquad . \quad (1.6)$$

where N_0 and N are the numbers of uncharged and of oppositely charged nuclei per c.c. and η_0 and η are the coefficients for the two types of capture. For the large ions :

$$dN/dt = \eta_0 n N_0 - \eta n N \qquad . \quad . \quad (1.7)$$

disregarding the small loss by recombination of large ions with one another.

For equilibrium both dn/dt and dN/dt are zero, hence $\eta_0 N_0 = \eta N$ and

$$q - \alpha n^2 - 2\eta n N = 0 \qquad . \quad . \quad (1.8)$$

From experiment it is known that η is of the same order of magnitude as α and since $N \gg n$ as a rule, eqn. (1.8) reduces to

$$q = 2\eta n N = 2\eta_0 n N_0 \qquad . \quad . \quad . \quad (1.9)$$

If the total number of nuclei, charged and uncharged, is Z, so that $Z = N_0 + 2N$, eqn. (1.9) reduces to

$$q = \left\{\frac{2\eta_0\eta}{2\eta_0 + \eta}\right\} Zn \qquad . \quad . \quad (1.10)$$

This is of the same form as an equation given originally by von Schweidler, $q = \beta n$, though β is constant only if the concentration of nuclei, Z, is constant.[5] It states a linear law of recombination, as opposed to the quadratic law which holds for filtered air. If the action of the ionising agents could be removed, the small ions would disappear at a rate $dn/dt = \beta n$, and $n = n_0 e^{-\beta t}$ would be the number present at any subsequent time. The quantity $\theta = 1/\beta$ can, by analogy with the equation for radioactive decay, be termed the *average life* of a small ion. Eqn. (1.10) then becomes

$$q\theta = n \qquad . \quad . \quad . \quad (1.11)$$

in which θ should vary inversely as the concentration of nuclei in the air. Though this simple relation is only an approximation, it is useful for general purposes.

The average life θ is found over land areas to lie between 10 and 60 sec. according to the purity of the air, the higher values being associated with observations made in the country. Over the oceans it may be as much as 230 sec., according to the measurements of Hess[7] and of Mathias[8] on the island of Heligoland. The table below, taken from Hess, illustrates the great difference in the equilibrium between small and large ions brought about when the wind direction changes; the first set of observations was made when the wind blew from the nuclei-laden European mainland, the second when it blew from the polar regions:

TABLE II

OBSERVATIONS OF THE STATE OF ATMOSPHERIC IONISATION (HESS, HELIGOLAND)

Wind.	No. of Small Ions per c.c.(n).		No. of Large Ions per c.c.(N).		Total No. of Nuclei.	Average Life of Small Ion, secs.	
	n_+	n_-	N_+	N_-	$Z = N_0 + N_+ + N_-$	θ_+	θ_-
S.W. (Land)	220	267	1480	1510	6300	22	22
N.N.W. (Arctic)	794	843	200	200	1100	115	204

These authors find that for values of θ lying between 36 and 100 sec., Schweidler's linear eqn. (1.11) is valid; in purer air, where θ is greater than 100 sec., the simple equation no longer holds, for the quadratic recombination term αn^2 then becomes significant.

Measurements with city air have, however, shown that the sizes of the nuclei vary over a wide range and that the coefficients η_0 and η in eqns. (1.6) and (1.7) show a corresponding wide range of variation.[3]

When, as in a city, the air is continually being provided with fresh nuclei, they do not usually reach a final state of equilibrium with the small ions. This would require a time of the order of 15 min. and stable conditions such as are found only in mist and fog. A further complicating factor is the discovery that the meteorological conditions which cause large variations in the nature and the number of the nuclei also alter the value of q by altering the radioactive content of the air.

11. Methods of Determining the Average Life of a Small Ion. The average life, θ, of the small ions can be determined for a sample of air enclosed in an ionisation chamber by two methods suggested by Schweidler.[5] The first involves the separate measurement of the ionising power, q, of the ionising radiations, and of the number of small ions per c.c., n, and substitution in the equation $q\theta = n$. q is obtained from the saturation current, i, using the relation $i = qev$, where v is the volume of the vessel. The corrections to be applied have already been discussed (§ 8). The quantity n is found by allowing the air to stand undisturbed for a sufficient time for equilibrium between small ions and nuclei to be attained and then applying a strong field for a few seconds to sweep the small ions on to the collecting electrode. The second method involves the measurement of the currents flowing in an ionisation vessel for various values of the applied potential difference.[9]

12. The Ionisation-balance in the Lower Atmosphere. It is of considerable importance to decide to what extent the various factors in the creation and destruction of ions account for the state of ionisation actually observed in the air over land and sea. Any failure in this respect would suggest that new processes must be sought.

Although the variable nature of the process of capture by heavy nuclei prevents an accurate comparison of the rate of production of small ions with their rate of disappearance, the available evidence suggests that the chief factors in the ionisation-balance have been discovered. Over land areas where nuclei are usually plentiful, the Schweidler equilibrium equation, $q\theta = n$, applies approximately. An average value of 60 sec. for θ, the mean life of a small ion, and of 650 per c.c. for the small ion concentration, requires that the rate of creation, q, should be 650/60 or 10·8 ions/c.c./sec. in reasonable agreement with direct measurements of q, shown in Table I. No direct measurements of θ have yet been made over the oceans. The concentration of nuclei in mid-ocean has been found to be from 400 to 800 per c.c. and the mean life-time of a small ion should be about 300 sec.[10] Since n was found on the *Carnegie* to amount to some 550 small ions/c.c., the ionising power at sea should amount to 550/300 or 1·83 I, in agreement with the value of q in the last column of Table I.

The problem of the ionisation of the lower atmosphere cannot be said to be completely solved until much more has been done to settle the manner in which the size and numbers of the nuclei affect their interaction with small ions. The general nature of the factors in the ionisation-balance has, however, been established, and an explanation can be given of the paradox that the conductivity of the air undergoes no large change as one passes from mid-ocean to land, despite a four-fold increase in the rate at which ions are formed. The increased birth-rate of the small ions is counterbalanced by an increased death-rate, owing to the greater concentration of condensation-nuclei over the land areas.

REFERENCES TO PAPERS

[1] A. R. Hogg, *Proc. Phys. Soc.*, **51**, p. 1014 (1939).

[2] J. J. Nolan, R. K. Boylan and G. P. de Sachy, *Proc. Roy. Irish Acad.*, A, **37**, p. 1 (1925).

[3] J. J. Nolan, *Archiv. f. Met., Geophys. u. Biok.*, A, III, p. 17 (1950).

[4] C. T. R. Wilson, *Dict. of Appd. Physics, Art. Atmospheric Electricity*, Vol. 3, p. 84.

[5] E. v. Schweidler, *Wien. Sitz. Ber.*, 2a, **127** (1918), **128** (1919), **133** (1924).

[6] V. F. Hess and R. P. Vancour, *Journ. Atm. and Terr. Phys.*, **1**, p. 13 (1950).

[7] V. F. Hess, *Gerland's Beit. z. Geophysik.*, **22**, p. 256 (1929).

[8] O. Mathias, *ibid.*, **27**, p. 360 (1930).

[9] C. O'Brolchain, *ibid.*, **29**, p. 1 (1931).

[10] V. F. Hess and J. D. Roll, *Phys. Rev.*, **73**, p. 916 (1948).

CHAPTER II

ELECTRIC FIELDS AND ELECTRIC CURRENTS IN THE ATMOSPHERE

13. The Fine-weather Field. It was first found in the year 1752 by Lemonnier that the air above the earth is the seat of a persistent electric field during fine weather. The direction of this field shows that the earth carries a negative charge and the upper layers of the air a positive one. The state of affairs close to a flat portion of the surface of the earth can be expressed in three ways: (*a*) the earth carries a charge of surface density σ per unit area; (*b*) there is a vertical field of strength $F = 4\pi\sigma$, just above the ground; or (*c*) between two horizontal planes close to the ground there is a difference of potential

$$dV = V_{h+dh} - V_h = - Fdh = - 4\pi\sigma dh,$$

where $h + dh$ and h are the heights of the planes.

The quantity dV/dh is called the potential gradient and is positive, since σ is found to be negative. For the same reason, the field F is directed downwards, and this downward direction is by convention adopted as the positive direction for electric fields in the atmosphere. In these definitions it is assumed that the portion of the ground considered is flat and far removed from projections, such as trees and buildings, which would disturb the distribution of charge and concentrate the lines of force at certain points.

The average value of the fine-weather potential gradient is about 100 v./m.; the corresponding value for the average charge density is $2{\cdot}7 \times 10^{-4}$ electrostatic units/sq. cm. or 0·0009 coulomb/sq. km. The total fine-weather charge on the earth is of the order of 500,000 coulombs. Measurements of the fine-weather field are made by determining either σ or dV/dh; they are always expressed in terms of the latter, in volts per metre.

14. Direct Measurement of the Surface Density of the Earth's Charge. The charge on such a limited portion of the ground as can be isolated for the determination of σ is so small that a sensitive electrometer must be used to measure it. Fig. 3 illustrates the "Universal portable electrometer" devised by C. T. R. Wilson for this purpose. A flat circular plate P, called the test-plate, is surrounded by a guard-ring and mounted flush with the surface of the earth but insulated from it. This plate is joined to a gold-leaf electroscope G, whose leaf moves inside a positively charged case and is observed through a telemicroscope. C is a variable cylindrical condenser,

called a compensator, with its inner plate joined to the gold-leaf system, and its outer plate maintained at a constant potential, $-V$.

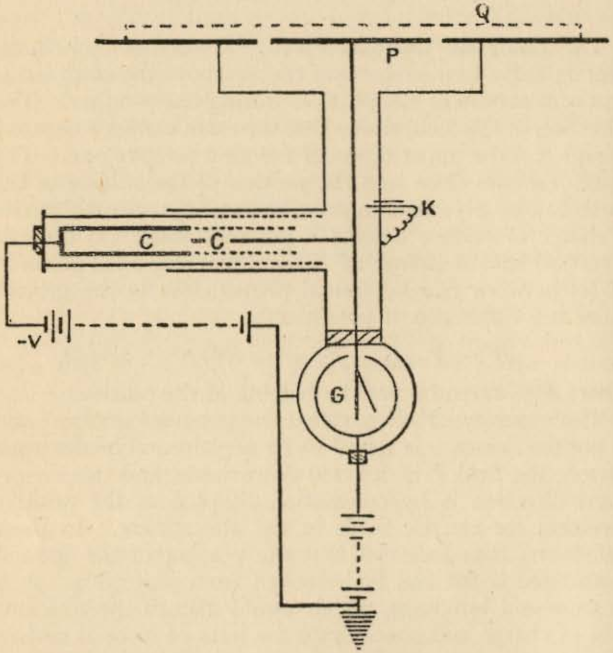

FIG. 3.—Universal portable electrometer (Wilson).

To make a measurement, the instrument is initially shielded from the earth's field by placing over it the earthed metal cover Q, indicated by the dotted lines. The reading of the leaf corresponding to earth or zero potential is found by momentarily earthing the plate P and the gold-leaf system with the key K. This earthing key is then withdrawn, the cover removed and the plate exposed to the field, with the compensator capacity zero. A negative charge thus appears by induction upon P, an equal positive charge is set free upon the central system, and the leaf moves inwards. It can be restored to its zero position if a negative charge is induced on the central system by increasing the capacity of the compensator to a value C'. The charge on the test-plate is then $-C'V$ and the surface density, σ, is $-C'V/A$, where A is the area of the test-plate.

In its original form the apparatus is small and portable, and the two batteries shown in the figure are replaced by small Leyden jars of silvered quartz, which keep their charge for a considerable time. For absolute measurements, with the plate mounted at ground level, a pit must be provided to house the observer; very often relative measurements are made with the instrument mounted on a tripod, and the factor necessary to reduce the observations to standard conditions determined from simultaneous observations in a pit or with a potential gradient method. Since the lines of force are concentrated upon any projection above the earth, the factor is greater than unity; in the case of a tripod 1 m. high it amounts to 3 or 4. A useful method of calibrating the instrument is to place it in artificial electric fields of known strength. Two large flat metal sheets, one flush with the test-plate and the other above it, are charged to a measured difference of potential by means of a battery of small cells.[1]

15. Measurement of the Potential Gradient. The potential gradient dV/dh may be found by measuring the difference of potential between two insulated conductors at different heights above the ground. Various forms of electrometer are employed, and one conductor may be the earth itself, while the other is a stretched horizontal wire about a metre above ground-level. Sometimes two wires at different heights are used. It is of course essential that the supports of the wires and the presence of the observer and his instruments should not materially alter the field to be measured. In the arrangement in Fig. 4 (*a*), the wire is hung between amber or sulphur insulators at the ends of two vertical rods; its length should exceed five times its height above ground.

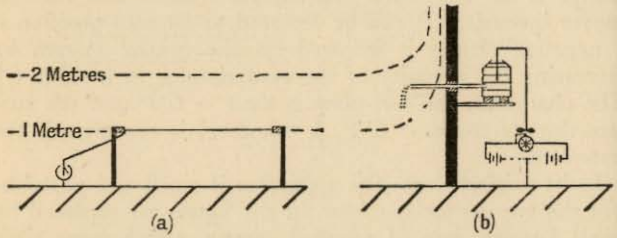

FIG. 4.—Measurement of the potential gradient.
(*a*) Absolute determination in open field.
(*b*) Continuous recording with the Kelvin water-dropper.

The wire when first set up is at a different potential to that of the air in its neighbourhood, and a current will flow between them which tends to remove this potential difference. The ordinary conductivity of the air is, however, too small for this process of equalisation to take place very rapidly and the electrometer reading will always lag behind a fluctuating field. It is necessary for this reason to place at the centre of the wire some active ionising agent which speeds up the process by increasing the local conductivity of the air. This may take the form of a glowing fuse of filter-paper, impregnated with a 5 per cent. solution of lead nitrate, or of a disc or spiral of metal coated with a deposit of polonium or radio-thorium giving off α rays. Such "collectors", as they are called, have the advantage of causing the wire and the electrometer to acquire the potential of the air at the centre of the stretched length, where the electric field is least disturbed by supports and observer. The time taken for equalisation of potential depends upon the type and activity of the collector and varies from 1 to 60 sec. There is always some uncertainty as to the exact position at which the ionising agent exerts its effect, especially if a wind is blowing, but this can be avoided by making observations at different heights.

16. Continuous Recording of the Potential Gradient. An adaptation of the arrangement just described is employed to obtain continuous records of the potential gradient, the electrometer being housed in a building and the collector carried on an insulated rod projecting from the wall. The earth's field is very much distorted by the presence of the building, and a reduction factor for the installation must occasionally be determined by making simultaneous measurements in an open field. Fig. 4 (*b*) shows the nature of the distortion of the equipotential surfaces due to a building.

The Kelvin water-dropper is still sometimes used at recording stations as a potential equaliser. A jet of water from an insulated cistern within the building escapes from the end of a pipe passing through and insulated from the wall, and breaks up into fine drops at the point where the potential is required (Fig. 4). On starting the jet, the potential of the insulated system will in general be somewhat lower than that of the air at the end of the pipe; this end will carry an induced negative charge and the cistern a positive one. Each drop carries a negative charge away and leaves the system slightly higher in potential than before. Ultimately the negative charge at the end of the pipe disappears and the system is at the potential of the air at the point at which the drops break away. In practice this takes a time of the order of 30 sec.

Radioactive collectors are also employed, but whatever the type of collector, the insulated system is usually joined to the needle of a quadrant electrometer of low sensitivity, and opposite pairs of quadrants connected to a battery whose centre is earthed. The record can be made intermittently, by means of a clockwork arrangement which depresses an inked pointer attached to the needle (Benndorf) or continuously, by receiving the spot of light from a small mirror on a moving strip of bromide paper.

17. The Electric Fluxmeter or Field Mill. If the cover Q in Fig. 3 were rapidly and periodically swung away from and over the plate P, the bound charge on P would alternately appear and disappear. If P is earthed through a high resistance R, an alternating potential difference of amplitude proportional to the field will appear across R. This can be amplified and if necessary rectified so that an output meter will show the magnitude of the field, an auxiliary device indicating its sign. Calibration can be carried out by comparison with a stretched wire in the open (§ 15).

Developments of this principle which was first employed by Russelvedt in Norway, have been used to measure electric fields during aircraft flights and under thunderclouds. In a recent form, capable of recording rapidly-varying fields, P is a ring of separate metal studs and Q is a rotating disc with a corresponding set of holes.[2]

18. Variation of the Potential Gradient with Height above the Ground : Space-charge. A good many observations of the potential gradient have been made with balloons fitted with collectors of the radioactive or the glowing fuse pattern ; they all indicate a rapid decrease in the field with increasing height above the ground (Table III, § 25). Though no great accuracy can be claimed for these measurements—for they are subject to errors arising from the distortion of the field by free and induced charges on the balloon and are but momentary observations of a quantity subject to considerable variation—it is well established that at a height of 18 km. the field has fallen to less than 1/100 of its value at ground level and is still diminishing slowly.

This diminution in the potential gradient with height indicates that a free positive charge, practically equal to the negative charge on the surface of the earth, resides within the lower 18 km. of the atmosphere. Integration of the values of dV/dh obtained in balloon ascents gives a total potential difference between the earth and the air at a height of 18 km. of about 4×10^5 v. ; at this height, and above it, the conductivity of the air is so great that this figure may be taken to represent the potential of the highly conducting "Kennelly-Heaviside layer" some 80 km. above the ground.

The density, ρ, of the positive "space-charge" referred to can be determined from Poisson's equation, which, if the lines of force are vertical, takes the form

$$\frac{d}{dh}(dV/dh) = -4\pi\rho.$$

Several investigators have examined the nature of the space-charge quite close to the ground, using collectors at different heights. The results obtained are conflicting as to the magnitude, and even the sign, of ρ in this region ; evidently local conditions are of main importance.

19. Variations of the Potential Gradient with Locality, Time and Season. Before any discussion of the more important variations of the potential gradient, mention should be made of the fact that incessant fluctuations are observed over short periods of time ; these are due to local changes in the space-charge and the conductivity of the air in the vicinity of the collector.

For many years continuous records of the potential gradient have been taken at certain stations in the Northern Hemisphere. Less extensive information is available for the Southern Hemisphere, but observations have been made over considerable periods in the Arctic and Antarctic regions. Some very important measurements over the oceans have been carried out by the survey ship *Carnegie* of the Carnegie Institute of Washington. It appears that the sign of the potential gradient in fine weather is positive all over the earth's surface, but that over land areas its value varies considerably with local conditions. The mean of dV/dh at Kew is 317 v./m., while that at Davos in Switzerland is but 64 v./m. Over the oceans a value of

126 v./m., which varies little if at all with geographical position, has been found. The average value for the whole earth is not far from 120 v./m.

It would seem that the potential gradient observed on land is not a quantity of much fundamental significance, its value depending primarily upon local conditions of atmospheric conductivity. This is evident from the fact that a simple periodic variation according to *local* time is the basis of all fine-weather land records. The minimum values are obtained in the early morning at about 4 a.m. and the maxima in the evening between 6 and 8 p.m. In many places a second maximum at 8 a.m. and a second minimum at midday are found. The amplitude of these daily variations at land stations sometimes reaches 50 per cent. of the mean value for the day.

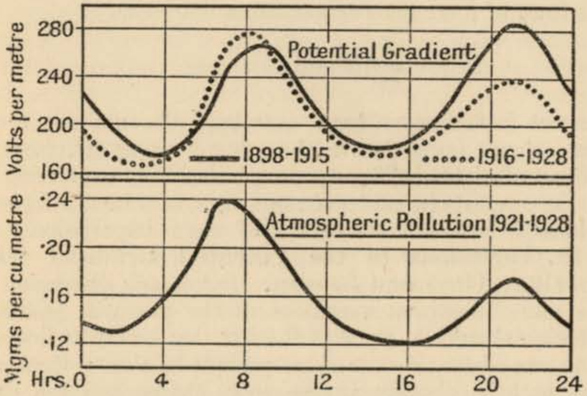

FIG. 5.—Relation between potential gradient and atmospheric pollution (Whipple).

The local time variation has been shown by Whipple [3] to be closely correlated with the amount and distribution of the smoke pollution of the atmosphere near large cities, where these continuous records have generally been made. Fig. 5 shows the variation at Kew of the potential gradient during the summer months for the two periods 1898-1915 and 1916-28, together with the variation in the amount of pollution (determined by aspirating the air through filter-paper) for the period 1921-28. The figure shows not only the close relation which exists between potential gradient and pollution but also the striking effect of the introduction of "summer time" into Great Britain in the year 1916, after which the morning minimum and maximum moved back by approximately one real hour. The atmospheric pollution curve has maxima and minima which have been explained by Simpson as due to the combined effect of variations in the amount of smoke produced in the city and variations in the stability of the atmosphere, that is in the mixing action of surface winds and general turbulence of the lower air.

As regards the annual variation of the potential gradient, land stations in both hemispheres show a maximum in the local winter and a minimum in the local summer. The only significant exception is the Antarctic region, where a reversal of phase occurs, giving rise to a maximum in the local summer and a minimum in the winter.

In view of the local origin of the variations of the gradient found at land stations near towns, observations over the oceans, where no pronounced local effects are to be expected, are of very great importance. Mauchly [4] analysed the measurements of the *Carnegie* cruises and found that there is a well-marked diurnal variation at sea, with maxima and minima occurring at the same moment in all parts of the globe. Thus the potential gradient in all oceans is found to be about 15½ per cent. below the mean at 5 hours, G.M.T., and about 20 per cent. above it at 19 hours, G.M.T. It is remarkable that the same 24-hour wave, progressing according to universal solar time, had been observed by Simpson in the Antarctic and Lapland. Fig. 6 (*a*) (after Whipple) illustrates these results, which were confirmed by the last cruise of the *Carnegie*. In each case shown in the figure, the mean potential gradient has been reduced to 100 v./m. No appreciable annual variation has so far been observed over the sea.

Mauchly's discovery of a variation of potential gradient according to universal time is generally accepted as being of fundamental significance. The variation must clearly be ascribed to an alteration in the difference of potential between the upper conducting layers of the atmosphere and the earth; it affords an important means of testing any theory of the origin and maintenance of this potential difference.

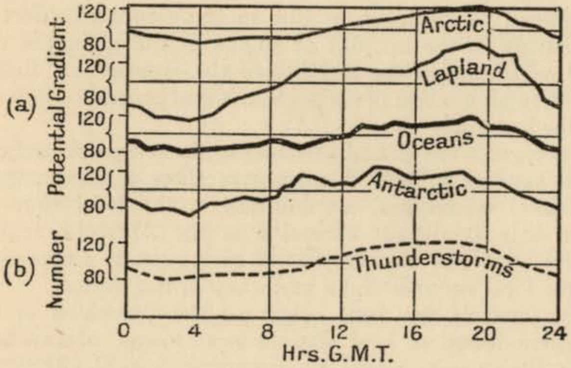

Fig. 6.

(*a*) Variation of potential gradient with universal time.
(*b*) Variation of thunderstorm activity over the earth's surface (Whipple).

20. The Fine-weather Current: Conduction and Convection Currents. In the ordinary fine-weather field positive ions are driven towards the earth and negative ions away from it; the motion of these ions constitutes a downwardly directed conduction current. The ionic conduction current passing through 1 sq. cm. of a horizontal plane can be written as $I_0 = F(\lambda_+ + \lambda_-)$, where F is the strength of the electric field and λ_+ and λ_- are the polar conductivities of the air at the point in question (§ 3).

There is, however, a fine-weather current of a different nature which plays a part in the transfer of electricity by the atmosphere. If the air should contain at any point an excess of ions of one sign (a space-charge), movement due to wind or ordinary turbulence will give rise to a mechanical transference of electric charge. Thus if v is the upward vertical component of the velocity of the air and ρ the space-charge per c.c., the upward convection current due to this cause will be $v\rho$ per sq. cm. The real current-density in fine weather is thus the resultant of the downward conduction and the upward convection currents, and is given by $I = I_0 - v\rho$.

The mean value of the total fine-weather current over the whole globe is not far from 2×10^{-16} amp./sq. cm. or 2 microamp./sq. km., so that the total current flowing in this way between the upper atmosphere and the whole earth is about 1000 amp. It would seem, from measurements which have been made in various parts of the world, that this current varies much less with changes in geographical position, in time of day and season of the year, than does the potential gradient. Although in the past it has not received a great deal of attention, it would appear to be a more useful quantity than the potential gradient and less affected by purely local conditions.

21. Direct Determination of the Conduction Current at the Surface of the Earth. Direct measurements of the magnitude of the conduction current which flows into the earth's surface have been made with the Wilson universal electrometer, with the test-plate mounted flush with the surface of the ground, as described in § 14. The curve of Fig. 7 illustrates the procedure followed in the measurements. At time t_1 the test-plate is exposed to the earth's field by removing the cover, and the capacity of the compensating condenser is adjusted to the value C necessary to bring the gold leaf back to the zero position. The charge on the plate is then $-CV$ and the field strength $F = 4\pi\frac{CV}{A}$, where, as before, A is the area of the plate and $-V$ the potential of the outside of the compensator. The test-plate is now left exposed until time t_2 and kept continuously at earth potential by increasing the compensator capacity to balance the charge received from the positive ions driven from the air on to the plate. Owing to variations in the strength of the field, and so in the induced charge on the plate, this increase is not a regular one. At time t_2 the plate is again shielded with the cover and the compensator capacity is reduced from C_1 to C_2 to maintain the shielded system at earth potential. At this moment the field strength is $4\pi(C_1 - C_2)V/A$. Since at the beginning of the measurement the compensator capacity was zero, it is evident that during the interval, $t_2 - t_1$, the plate received a positive charge C_2V

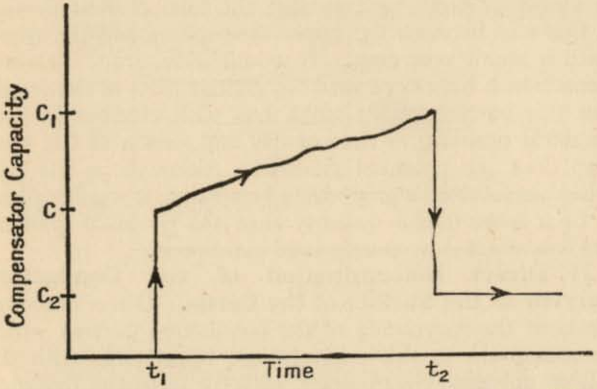

FIG. 7.—Measurement of conduction current with the Wilson universal electrometer.

and the average value of the conduction current per square centimetre was $C_2V/A(t_2 - t_1)$. For routine observations the instrument is mounted on a tripod above the ground, and the readings are reduced to standard conditions by occasional determinations of the reduction factor, from observations in a pit.

22. Indirect Determination of the Conduction Current at a Point in the Air. The conduction current at a point in the air can be determined also by separate measurements of the field strength, F, and the polar conductivities, λ_+ and λ_-. The methods used to determine

F, or the potential gradient, have already been described in this chapter; the conductivities are usually found by the Gerdien method sketched in § 2. In another form of the Gerdien method, due to Schering, the cylindrical condenser is replaced by a large earthed wire cage, along the axis of which is a charged insulated wire connected to an electrometer. The rate at which the charge is dissipated by ions of opposite sign is determined in exactly the same manner as in the Gerdien method, and the same equations apply if certain conditions are fulfilled. The cage must be large enough, and the charge on the wire small enough, to ensure that no considerable alteration in the number of ions is caused by the flow of current. The arrangement must be set up in the open air to ensure that ordinary atmospheric circulation prevents the development of the electrode space-charges to be discussed in the next section. And finally, the whole cage must be shielded from the earth's field by a rough roof or by trees, otherwise it will carry an induced negative charge on the outside and rob the entering air of negative ions.

23. The Comparison of the Direct and Indirect Methods: Electrode Space-charge. The direct method due to Wilson measures the conduction current entering the earth; the indirect method measures the current at a certain height above the earth. In the latter case the current is carried by two oppositely moving streams of ions of opposite sign; in the former it consists of a single stream, for while positive ions enter the earth, negative ions do not pass from it into the atmosphere. The two methods, therefore, do not necessarily measure the same quantity. One determines the true conduction current at a height, and the other its positive component at ground-level.

Consider a vertical cylinder of perfectly still air whose height is h and cross-section 1 sq. cm., with its base on the ground (Fig. 8). Let F' and F be the electric field strengths at top and bottom of this cylinder, and λ'_+, λ'_- and λ_+, λ_- the polar conductivities. The conduction current through the top is $F'(\lambda'_+ + \lambda'_-)$; in each second a positive charge $F'\lambda'_+$ enters, and a negative charge $F'\lambda'_-$ leaves. Through the bottom there is only a downward flow of positive ions, carrying away a positive charge $F\lambda_+$ per sec. The cylinder as a whole thus loses a positive charge $F\lambda_+ - F'\lambda'_+$ per sec., and a negative charge $F'\lambda'_-$ per sec. This may be expressed as a development of a positive space-charge at a rate

$$F'\lambda'_- - (F\lambda_+ - F'\lambda'_+) = F'(\lambda'_+ + \lambda'_-) - F\lambda_+$$

per sec. The growth of this "electrode" space-charge should automatically decrease the field F' and should cease only when F' has such a value that as many negative as positive ions leave the cylinder in each second, i.e. when

$$F'(\lambda'_+ + \lambda'_-) = F\lambda_+ . \quad (2.1)$$

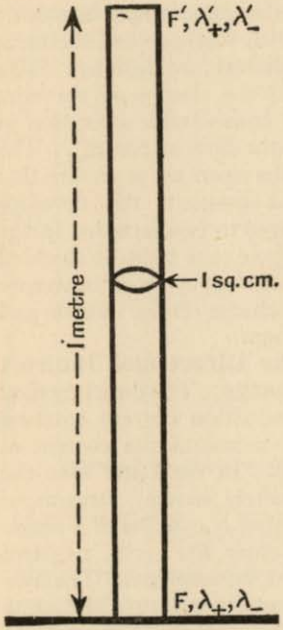

FIG. 8.—Electrode space-charge.

The two sides of this last equation are respectively the conduction currents at height *h* and at ground level. Thus, for perfectly still air, the two methods we have described should give the same numerical values, though they measure different quantities.

It was thought for some time that λ_+ must necessarily be considerably smaller than $(\lambda'_+ + \lambda'_-)$ and hence that F must be about double F'. Scrase[5] has, however, found practically no difference between F and F' and very little sign of electrode space-charge. A suggestion by Whipple that the space-charge is removed by atmospheric turbulence as rapidly as it is formed is unable to explain these results.

The difficulty appears to have been removed by the discovery by Hogg[6] that, owing to the local effect of α and β rays from radioactive matter in the soil, λ_+ is much larger than was thought and approximately equal to $(\lambda'_+ + \lambda'_-)$. Since $F \cong F'$, eqn. 2.1 is satisfied and the problem of electrode space-charge does not arise.

Values of the air-earth current in fine-weather obtained by the direct method in various parts of the world range from 1·0 to $4{\cdot}0 \times 10^{-16}$ amp./cm.^{2} Those by the indirect method vary from 0·4 to $3{\cdot}2 \times 10^{-16}$ amp./cm.^{2} Though as indicated above and found in practice, the two methods should give the same result, the variability of the ionising agents producing λ_+ and of convection currents suggests that the direct method is the more reliable of the two.

24. Variations in the Fine-weather Current. Both direct and indirect methods of studying the air-earth current indicate that in fairly fine weather it is less subject to variations than is the potential gradient. Simultaneous measurements of the field and of the total conductivity of the air, $\lambda = \lambda_+ + \lambda_-$, show that these quantities vary to some extent in inverse ratio, with the result that their product, the conduction current, is less variable than either. Factors such as an increase in the number of condensation nuclei or the presence of fog particles, which decrease λ by raising the proportion of the slow ions present, cause a corresponding increase in the potential gradient. The simple form of the daily variation in the earth's field has its counterpart in an inverse variation of the conductivity, with a maximum in the early morning between 3 and 4 a.m. The extreme effect of atmospheric pollution near large cities, which increases the potential gradient and diminishes the conductivity, has already been discussed (§ 19). In general however the reciprocal relation between λ and F is not exact enough to keep the conduction current invariable. At land stations in Europe the current is found to have an annual variation which coincides in phase with that of the potential gradient but has a smaller amplitude. At other stations its daily variation follows the changes in the conductivity.

CHRIST'S COLLEG

25. The Conduction Current at Great Heights. It has been found from simultaneous determinations of potential gradient and atmospheric conductivity during balloon ascents that the rapid decrease in F with height above the ground is accompanied by an increase in λ, of such magnitude as to keep the conduction current approximately constant. Table III shows the results of an ascent by Wigand,[7] from which it will be seen that when the field had fallen to 1/15 of its value at the ground, the conductivity had risen eleven-fold.

TABLE III

Conduction Current during Balloon Ascent (Wigand)

Height (Metres).	F (volts/metre).	λ (e.s.u.).	I (amps./sq. cm.).
0	136	$1{\cdot}1\ 10^{-4}$	$1{\cdot}7\ 10^{-16}$
2500	27	$4{\cdot}8$,,	$1{\cdot}4$,,
4400	18	$8{\cdot}2$,,	$1{\cdot}6$,,
6500	$8{\cdot}8$	$12{\cdot}6$,,	$1{\cdot}2$,,

These results have been confirmed and extended by conductivity measurements made during the flight of the United States army balloon Explorer II in 1935.[8] They showed that up to a height of 18 km., the ionisation of the air increases by roughly 500 ion-pairs per km. and that the conductivity at 18 km. was about 100 times the usual value at the ground.

26. The Significance of the Air-earth Current. The general constancy of the air-earth current with height above the ground and in different parts of the world suggests that, at any rate to a first approximation, it has its origin in a constant potential difference between the conducting layers of the upper atmosphere and the earth, and that its actual value depends upon the resistance of the air between the two regions. The conductivity of the upper portion of this air-path is entirely due to the cosmic radiation * and is unlikely to alter appreciably with time of day, locality or season. But below a height of about 2 km. over land areas, an additional ionising influence is exercised by radioactive matter in the air and in the earth, and this, as we have seen (§ 5), varies very much. Moreover, the conductivity of the air in this region is dependent upon the relative concentrations of the large and the small ions, and these alter with the number of nuclei present. The current-bearing column of air above the point of observation thus consists of an upper constant resistance and a lower variable one in series with each other. The extent to which the current carried by the air-column alters with local conditions will depend on the ratio which the lower variable resistance bears to the total resistance.

Both the results of Wigand in Table III and those from the Explorer II flight[8] indicate that the total resistance of a column of air of cross-sectional area 1 sq. cm. is about $1{\cdot}0 \times 10^{21}$ ohms. The conductivity between 14 and 18 km. is so high that this upper region contributes only a small fraction to the resistance and higher regions would be still less important. If the variable part of the column be the first 100 m. above the earth, where the specific resistance is usually 8×10^{15} ohm-cm., this portion will contribute 8×10^{19} ohms to the total. We see that even if the conductivity of the first 100 m. were halved, the effect upon the air-earth current would amount to but 4 per cent. The potential gradient near the ground, on the other hand, would be doubled, for from this point of view it represents the fall of potential per unit length of the lowest part of the current-bearing column. It is evident that potential gradient measurements over land, and in particular near cities, cannot be expected to yield information which bears directly upon the question of the total potential difference between upper atmosphere and earth. In the case of observations at sea and in polar regions the position is different; practically all the ionisation throughout the whole column is then due to the cosmic radiation, and the only local effect which can influence the fine-weather gradient is a change in the number of nuclei available for the capture of small ions.

---

* This refers to the regions from 15 to 30 kilometres above the ground. The E layer 80 kilometres up has a solar origin.

27. The Electric Field during Disturbed Weather. The normal downwardly directed field, due to a positively charged upper atmosphere and a negatively charged earth, is frequently disturbed when the weather ceases to be fine. During fog, as explained in § 24, it is very much increased, and may reach ten times its normal value; during dust-storms, of the type common in semi-arid regions and deserts, powerful reversed or negative fields are usual and reach 10,000 v./m. The effects found during cloudy weather and rain are variable, and range from gradients of the order of a few hundred volts per metre in a fine drizzle to as much as 50,000 v./m. under thunderclouds.

Generally speaking, negative fields predominate when light or steady rain is falling, though occasional positive excursions are observed. In the case of heavy rainstorms and thunderclouds, the sign of the field depends upon the portion of the cloud passing over the point of observation, but here, too, there is evidence that negative gradients are the more frequent.

The field during thunderstorms fluctuates a great deal if the cloud is active in producing lightning. Its value at the ground does not as a rule exceed 10,000 to 20,000 v./m. except for very brief intervals occupying fractions of a second. Equally high fields are observed during heavy rainstorms unaccompanied by lightning. During a snowstorm the field is usually positive and in a heavy storm may attain 10,000 v./m. The methods employed for the study of the large and rapidly varying fields of thunderclouds are described in Chapter III.

28. The Charge on Rain and Snow; the Precipitation Current. The apparatus required to examine the charge carried to the earth by precipitation—rain, hail, snow, and sleet—is comparatively simple, and consists essentially of an insulated metal vessel to catch the rain, etc., and an electrometer to measure the charge received. Great precautions must, however, be taken to eliminate the disturbing effect of the negative charge given to the air as the drops splash on the container (Lenard effect). If some of this is lost a spurious positive charge will be indicated. For a similar reason, the entry of spray due to the splashing of rain falling outside the receiving vessel must be prevented. The receiver must be screened electrostatically from the electric field outside. The rate of rainfall, and sometimes the size of individual drops, is also observed. By amplifying the induction effect of a drop as it passes through an insulated cylinder the charges carried by individual drops may be measured. As the charges brought down vary considerably, it is necessary in all such work to carry out a long series of observations.

Since the year 1908, when the modern series of observations began, all observers have found a preponderance of positive charge on rain in general. The average ratio of the quantities of positive and negative electricity brought down by all kinds of precipitation varies according to the situation of the observer. Some of the values found for this ratio are: Potsdam (Schindelhauer), 1·4 : 1; Puy-en-Velay (Baldit), 2·4 : 1; Dublin (McClelland and Nolan), 4·8 : 1; Simla (Simpson), 2·4 : 1; Otago, New Zealand (Marwick), 3·2 : 1. As exceptions to the general rule, it may be noted that snow usually carries a considerable negative charge, and that the fine drops of a drizzle are also negative.

The subject of the charge on rain and its relation to the prevailing electric field has recently received special attention because of its connection with the electrification of thunderclouds. This work is discussed in § 45.

The charge per c.c. in the case of heavy thunderstorm rains reaches average values of 30 to 40 e.s.u., while single drops may carry 100 to 200 e.s.u. (Gschwend) and attain potentials as great as 300 v. The mean values reported by different observers for the net charge per c.c. on all forms of precipitation during a long series of observations range from + 0·029 e.s.u. at Potsdam (Schindelhauer) to + 0·176 e.s.u. at Simla (Simpson).

The transfer of electricity from atmosphere to earth in this manner constitutes a "precipitation" current which appears on the whole to be in the same direction as the fine-weather conduction current, since it carries more positive than negative charge to the earth. In the case of thunderstorms this current sometimes reaches values as high as 2×10^{-11} amp./sq. cm., and the total convection current carried by the rain in such a storm may amount to more than 0·1 amp. Before it is possible to frame a trustworthy estimate of the amount of the excess of positive over negative electricity conveyed to the whole earth in this way, much more information is required from tropical and equatorial regions.

29. Currents due to Point-discharge. It is well known that an electric glow or brush discharge takes place from a pointed conductor when placed in a sufficiently strong field. It consists of a stream of ions of the same sign as the charge upon the point. Discharges of this kind must, therefore, be expected from the ends of exposed conductors on the surface of the earth whenever the potential gradient and the height and sharpness of the ends are sufficiently great. Indeed the earliest investigators, Franklin, Dalibard and Lemonnier with high metal rods and kites detected electric fields in thundery and in fine weather, by means of this current. Under favourable conditions, afforded by mountain peaks or the masts of ships exposed to the intense fields of a thundercloud, the glow discharge may become very conspicuous (St. Elmo's fire).

It was first pointed out by C. T. R. Wilson [9] that point-discharge currents of much smaller intensity must play an important part in the interchange of electricity between the earth and the atmosphere. The fields prevailing under rainstorms and thunderclouds very frequently reach values sufficiently great for considerable currents to be discharged from an assemblage of exposed conductors, such as trees, bushes, housetops and even fields of grass. A conductor need not end in a sharp point nor project to a great height in order that it should begin to act as a discharger. For example, an earth-connected sphere of radius 1 cm. need be raised only to a height of 3 m. in the not unusual field of 10,000 v./m. for the electric intensity at its surface to reach the critical value of 30,000 v./cm. at which brush discharge begins. A practical instance of the correctness of this suggestion is afforded by some experiments made by the author, in which a small tree, 4 m. high, was cut off at its base, mounted on insulators and connected to earth through a galvanometer. Exposure to the fields due to nearby thunderstorms yielded the average point-discharge currents shown below: [10]

TABLE IV

POINT-DISCHARGE CURRENTS FROM A SMALL TREE (SCHONLAND)

Field (volts/metre).	Current (microamperes).
− 3,500	0·07
− 5,500	0·20
− 11,000	1·00
− 16,000	4·00

The contribution of a single tree such as this must be multiplied by a very large factor to obtain the total current flowing between the earth and a thundercloud; in the case quoted, the tree was typical of the exposed natural conductors for many miles around, and a rough surface integration over the area affected by the cloud indicated that the total point-discharge effect was of the order of 2 amp. in an upward direction.

Since the direction and magnitude of the large fields associated with disturbed weather vary considerably, it is important to know the relative amounts of positive and negative electricity discharged from an exposed point over a long period of time. In a series of observations by Wormell,[11] a sharp point carried on insulators at the top of a pole was connected by cable to one electrode of a voltameter filled with dilute sulphuric acid, the other electrode being earthed. The voltameter, shown, with some minor but significant details omitted, in Fig. 9, was constructed of fine capillary tubing (diameter 0·8 mm.) and the gases evolved were separately collected, their volumes being determined from the lengths of the bubbles formed. If v_1 and v_2 are the volumes of the mixed gases collected at the earthed and the point-connected electrodes respectively, the quantities of positive and negative electricity, q_1 and q_2 which have been discharged from the point can be separately determined in the following way: If one unit of quantity liberates in electrolysis c c.c. of oxygen and $2c$ c.c. of hydrogen, we have

$$q_1c + 2q_2c = v_1 \quad \text{and} \quad 2q_1c + q_2c = v_2,$$

from which we may determine q_1 and q_2.

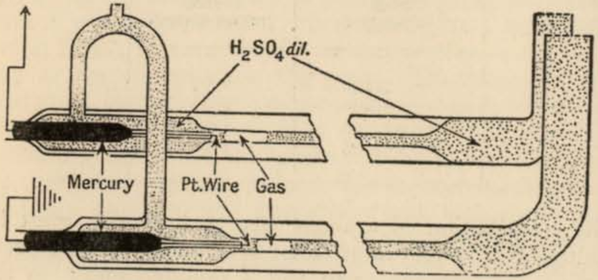

FIG. 9.—Water micro-voltameter (Wormell).

The quantities discharged in this way have been examined by Wormell at Cambridge over a period of three years. Some of his results are shown in the Table V.

Each year shows a considerable net loss of positive electricity from the discharger.

In a more detailed discussion it is noted that of 147 separate periods of disturbed weather accompanied by precipitation, 103 showed an excess of upward current (+ ve discharge), 34 an excess of downward current (− ve discharge), and in 10 the net discharge was zero. This is what would be expected from the predominance of negative potential gradients in disturbed weather.

Similar measurements have been made in other parts of the world (§ 30).

TABLE V

INTEGRATED EFFECTS OF POINT-DISCHARGE (WORMELL)

Year	1927.	1928.	1929.
Positive quantity discharged, q_1 (coulombs)	0·28	0·24	—
Negative quantity discharged, q_2 (coulombs)	0·14	0·11	—
Net positive discharge, $q_1 - q_2$ (coulombs) .	0·14	0·13	0·11

30. The Interchange of Electricity Between the Earth and the Atmosphere; the Maintenance of the Earth's Charge. The negatively charged earth and the positively charged upper layers of the atmosphere form two plates of a spherical condenser with the lower air as the dielectric. Although the conductivity of the air between the plates is small, the applied potential difference is great enough to make the leakage of charge through it very considerable. The average value of the charge per unit area of the earth's surface is 9×10^{-14} coulombs/sq. cm., and that of the fine-weather conduction current is 2×10^{-16} amp./sq. cm. Left to itself, a condenser of this kind would be discharged by internal leakage in a time of the order of 10 min. But this is not all; to the dissipation of the earth's charge by the fine-weather current we must add the effect of the charge conveyed by rain and snow, for it has been shown that on the whole the net charge conveyed to the earth by precipitation is positive in sign. Thus to the conduction current of 1000 amp. over the whole surface of the earth must be added a precipitation current in the same direction, estimated by Wigand[12]—though any such estimate can only be provisional—at 400 amp., making a total of 1400 amp. tending to dissipate the 500,000 coulombs with which the earth is charged.

In spite of the continuous operation of these two factors, the earth's charge remains practically constant, so it is clear that there must exist some reverse or compensatory process which neutralises their effect. Some agency must be continuously at work replenishing the charges on the earth and in the upper air. The discovery of the mechanism responsible for this replenishment has been one of the chief problems of atmospheric electricity, and many proposals have been made to this end.

It was first suggested by C. T. R. Wilson[13] that it is to regions of disturbed weather that we must look for the mechanism of replenishment, that under rain and thunderclouds, where the potential gradient is more often reversed than in the normal direction, there are two processes at work which convey considerable quantities of negative charge to the earth. The first of these is the action of point-discharge from conductors projecting from the ground. This has already been discussed, and it has been shown that experimental tests indicate that it is of great importance. The second process is the charge conveyed to the ground by lightning flashes from thunderclouds. Every second, as will be seen in Chapter III, some 100 lightning discharges occur over the whole surface of the earth, each involving the passage of a charge of the order of 30 coulombs. Only a fraction of these, perhaps one in four on the average, strike the ground. The effectiveness of this process of transference of charge will depend upon whether more charges of one sign than another are conveyed to the earth. If, for example, practically all the charges transferred were negative, the process would be equivalent to a continuous compensation current of about 500 amp. Though there is evidence from a number of regions (§ 33) that this is actually the case, the question is still under investigation.

An estimate of the annual electrical "balance-sheet" of a square kilometre of ground at Cambridge in England has been made by Wormell,[11] on the basis of his point-discharge measurements and observations on lightning discharges. The air-earth conduction current loss and the effect of rain and snow were estimated from average values of these quantities. The results were:

	Coulombs/sq. km./annum.
By natural point-discharge, gained . . . —	100
By lightning, gained . . . . . —	20
By atmospheric conduction, lost . . . —	60
By precipitation, lost . . . . . —	20
Net gain of negative charge . . . . —	40

Though such an estimate is only a very rough one, it appears to be quite possible that in this locality the four processes approximately balance one another, or even that the earth gains a negative charge.

These results have been supported and extended by similar measurements at Kew Observatory, at Durham, at Pretoria and in Nigeria.

Over the oceans, where thunderstorms are infrequent and elevated points are absent, the balance-sheet may be expected to show a net loss of negative charge, so that the position for the whole globe would be one of approximate equilibrium.

This suggestion of Wilson involved more than the interchange of electricity between the bases of cumulonimbus clouds, showers and thunderstorms, and the ground; he regarded such clouds as equally active in supplying positive electricity to the upper air. The existence of this upward current has been established by recent high-altitude flights above thunderclouds. Gish and Wait,[14] by measuring both the electric field and the conductivity of the air, have shown that an average current of 0·8 amp. flows in the expected direction. The cloud thus acts as an electrical generator which removes positive electricity from the earth and supplies it to the conducting layers of the atmosphere above, in which it is rapidly distributed in such a way as to maintain them at a constant potential of about 4×10^5 v., in spite of losses.

We should therefore expect the potential difference, and so the potential gradient, to show a maximum when the thundery regions of the earth are at their maximum activity. The diurnal variation of the gradient discovered by Mauchly (§ 19) should agree in phase with the diurnal variation of thunderstorm activity over the earth. It has been shown by Whipple[15] that this appears to be the case. Whipple's examination of the diurnal variation of the world's thunderstorms is represented by the bottom curve of Fig. 6 (*b*), where it will be seen that the parallelism with the variation of the potential gradient over the oceans and the polar caps is indeed very close.

Although the evidence discussed above makes it very probable that Wilson's replenishment theory is correct, it must be admitted that further quantitative measurements in different parts of the globe are needed to establish it.

A rather revolutionary extension of Wilson's ideas has been put forward by Frenkel,[16] who suggests that the generators chiefly responsible for the circulation of electricity between the earth and the upper air, in the reverse direction to the fine-weather current, are not thunderstorms but ordinary clouds. According to this theory all fog-like clouds carry a positive charge above and a negative charge below (§ 44) and thus create a conduction current which is opposite to that of fine weather. The fields due to these cloud dipoles (of which there are estimated to be a hundred thousand at any moment over the earth's surface) will, in Frenkel's view, induce and maintain the negative charges observed on the distant " serene sky " regions of the earth and positive charges in the air above them.

It has been shown, however, in § 24, that the enhanced fields below ordinary fog-like clouds are due to an alteration in the conductivity of the air and are in the normal positive direction. Such clouds do not give reversed electrical fields of the kind required by Frenkel's theory unless they are shower-clouds from which rain is falling (§ 45).

REFERENCES TO PAPERS

[1] R. E. Watson, *Met. Office Geophys. Memoir*, No. 45 (1929).

[2] D. J. Malan and B. F. J. Schonland, *Proc. Phys. Soc.*, B, **43**, p. 402 (1950).

[3] F. J. Whipple, *Quart. Journ. Roy. Met. Soc.*, **55**, p. 351 (1929).

[4] S. J. Mauchly, *Terr. Mag.*, **28**, p. 61 (1923).

[5] F. Scrase, *Met. Office Geophys. Memoir*, **67** (1935).

[6] A. Hogg, *Mem. Comm. Solar Obs., Canberra*, **7** (1939).

[7] A. Wigand, *Phys. Zeit.*, **26**, p. 81 (1925).

[8] O. H. Gish, *Terrestrial Magnetism and Atmospheric Electricity* (ed. J. A. Fleming), Chap. IV, New York, Dover Publications (1949).

[9] C. T. R. Wilson, *Proc. Phys. Soc.*, **37**, p. 32*d* (1925).

[10] B. F. J. Schonland, *Proc. Roy. Soc.*, A, **118**, p. 252 (1928).

[11] T. W. Wormell, *ibid.*, A, **115**, p. 443 (1927) ; **117**, p. 567 (1929).

[12] A. Wigand, *Ann. d. Physik*, **66**, p. 81 (1921).

[13] C. T. R. Wilson, *The Observatory*, **44** (1922).

[14] O. H. Gish and G. R. Wait, *Journ. Geophys. Res.*, **55**, p. 473 (1950).

[15] F. J. Whipple, *Quart. Journ. Roy. Met. Soc.*, **55**, p. 1 (1929).

[16] J. Frenkel (*a*) *Journ. Franklin Inst.*, **243**, p. 287 (1947) ; (*b*) *Ac. Sc. U.S.S.R., Journ. of Physics*, **8**, p. 285 (1944).

CHAPTER III

THE ELECTRIFICATION OF THUNDERCLOUDS

31. The Number of Thunderstorms over the Globe. Although a thunderstorm is a comparatively rare event at any one station in temperate latitudes, the number of thunderstorms per day experienced by the earth as a whole is very large. Even in such a limited area as France there are comparatively few days in the year on which thunder is not reported from some part of the country. According to a survey made by Brookes,[1] the earth experiences 16,000,000 thunderstorms per annum or 44,000 per day. Taking one hour as the average duration of a single storm there will on the average be 1800 thunderstorms in progress in different parts of the world at any one moment. As has been shown in § 19, Fig. 6, world thunderstorm activity alters regularly with the sun, reaching a maximum when the afternoon storms of equatorial Africa are ending and those of equatorial America are beginning. Brookes has estimated that about 100 flashes of lightning occur every second or 360,000 every hour.

The most thundery region of the earth is the island of Java, where thunder is heard at any one place on 61 per cent. of the days in a year. This percentage, the *isoceraunic level* of the place concerned, is 41 in parts of Central Africa, 39 to 37 in Southern Mexico and Panama, 29 in Brazil and 26 in Madagascar. In North America the isoceraunic level rises from 3 per cent. for Southern Canada to 25 per cent. for Florida. On the Californian coast it is about 1 per cent.

Over most of Europe thunder is heard on the average on about 11 days in the year; near the Alps the figure rises to 18 days. The figure for Johannesburg in South Africa is 50-60 days in the year. In using these figures account must be taken of the fact that on about 1 in 7 of thunderstorm " days " there is more than one storm within audible distance.

Over the oceans thunderstorms are infrequent except off the south-east coasts of Brazil and South Africa and in the seas surrounding Madagascar and the East Indian islands.

32. Methods used to Measure the Electric Fields of Thunderclouds. (*a*) *The capillary electrometer method.* This was devised by C. T. R. Wilson [2] for use in his classic investigations of thunderstorms and has also been employed by Wormell [3] and by the writer.[4] The general principle is similar to that of the test-plate or induced charge method of studying the fine-weather field (§ 14), in which an exposed conductor is kept at zero potential by alteration in the capacity of a compensating condenser, when the charge induced upon the conductor by the field is equal to that given by the compensator to the system. In the measurement of rapidly changing thunderstorm fields, however, compensation is produced automatically and very quickly by means of an ingenious form of capillary electrometer. As shown in Fig. 10, this consists of a small bubble of dilute sulphuric acid enclosed between mercury threads in a narrow capillary tube. The threads lead to end-cups of mercury, and one of these, A, is connected to a test-plate T insulated from, but flush with, the surrounding ground, while the other is connected to the earth. Between the mercury and the glass there is everywhere a thin film of dilute acid, so that an electrical " double layer " surrounds the mercury and makes each half of the electrometer a condenser charged to the potential difference of the double layer, about 0·90 v., the mercury being positively charged with respect to the acid. If a positive charge is given to the left-hand side, it will increase the charge on the left-hand condenser and the left-hand thread will move forward a certain distance so as to absorb this charge, by increasing the area of the acid-mercury surface and the capacity of the left-hand condenser. A corresponding movement of the right-hand thread causes a reduction in area on this side and the passage to earth of a quantity of electricity exactly equal to that originally supplied to the electrometer. The distance moved, *y*, is easily seen to be related to the charge passed through, q, by the equation

$$q = 2\pi rCV'(1 - r/R)y,$$

where r is the radius of the capillary and R that of the end-tube, and V' is the potential difference, and C the capacity per unit area, of the double layer. Thus $q = ky$, where k is an instrumental constant which can be determined by discharging a known quantity of electricity through the instrument. The arrangement, therefore, gives a linear relation between the movement of the acid bubble and the quantity discharged. In spite of the large value of C, about 30 microfarads/sq. cm., the electrometer is extremely sensitive. By using a microscope focused upon one end of the bubble, quantities of the order of 10^{-9} coulombs may be measured. A permanent record is best obtained by placing a slit and a moving photographic plate in the focal plane of the objective.

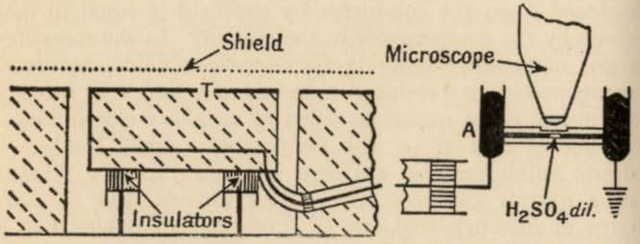

FIG. 10.—Measurements of the electric fields due to thunderstorms (C. T. R. Wilson).

When connected to a test-plate the electrometer automatically maintains it at earth potential and records the quantities of electricity passing to and from the earth as a result of changes in the external field and of the induced charges on the plate. The type of record obtained from this arrangement during a thunderstorm is illustrated by the thick line of Fig. 11. An earthed cover is swung over the test-plate for a few seconds at the beginning of each minute, causing the movements A′AA″, B′BB″, etc., as the induced charge flows away to earth and then returns. The points A, B, C, D, E would lie on a horizontal line if there were no actual transfer of charge between the plate and the atmosphere taking place, that is to say, no conduction current, point-discharge, or transport of charge by rain. The first 2 min. in the figure show the slight slope due to the conduction current only, the last two show the effect of a fall of positively charged rain. The charge, for example, conveyed to the plate by rain and conduction during the third minute is $k(y_D - y_C)$ and the mean value of the current per square centimetre $k(y_D - y_C)/60A$, where A is the area of the plate and k the constant of the electrometer. The dotted line joining the points A, B, C, D, E gives the reading of the instrument when the test-plate is shielded at any time during the record, and this line serves as the zero from which the field and field-changes are measured. The field at the end of the first minute, for instance, is given by

$$F = 4\pi k(y_B - y_{B'})/A.$$

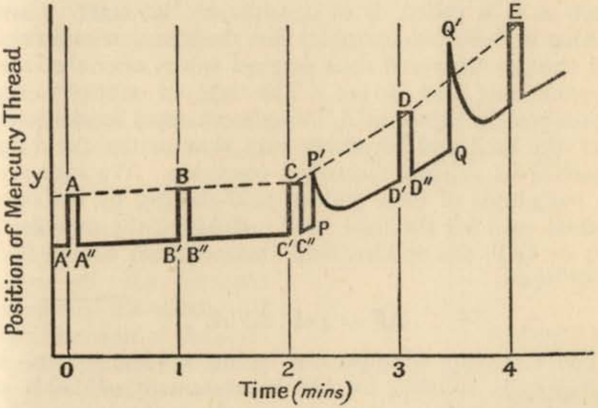

FIG. 11.—Capillary electrometer record during a thunderstorm.

Two sudden interruptions in the record, such as are caused by lightning flashes, are shown at P and Q. At P the record suddenly touches the dotted line, indicating that for a moment the field was completely destroyed and the original induced charge ran to earth through the electrometer. At Q the movement extends beyond the dotted line, showing that the field was actually reversed in sign by the discharge and reached a reverse value equal to half its original strength. The shape of the record immediately after the passage of these lightning flashes, the recovery curve as it is called, is of considerable interest; it will be seen in these two examples that the thundercloud fields and charges recovered their original values after the lapse of something like 20 sec. The rate of recovery was approximately exponential, being most rapid immediately after the flash and becoming very slow as the cloud approached its original electrified condition. We may find the magnitude of these sudden field-changes by the same method used for the field itself; if Δy be the movement PP′, or QQ′, the sudden field-change which caused it is given by:

$$\Delta F = 4\pi k \,.\, \Delta y/A.$$

The test-plate arrangement, using a plate 50 cm. in diameter, is suitable for the measurement of fields of strength exceeding 1000 v./m., such as are caused by thunderstorms when they come within a distance of about 8 km. An inverted test-plate of smaller dimensions mounted on a pole and with a cover periodically and automatically swung below it for a short time has been used by Wormell for measurements of the large fields due to near storms. This is of great use during heavy rain. For more distant storms Wilson employed a different form of exposed conductor, a copper ball 1 ft. in diameter mounted on insulators at the end of an iron pipe 5 m. in length (Fig. 12). When the arrangement is in use the pipe is held vertical; by turning it about a horizontal axis through its base, the pipe is lowered and the ball enters an earthed metal case which shields it from the field. Connection to the capillary electrometer is made by an insulated wire passing down the centre of the pipe.

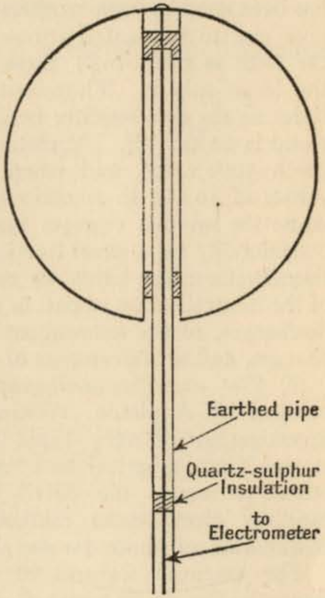

FIG. 12.—Elevated sphere for measuring fields of distant thunderstorms.

The field producing a given movement of the electrometer is calculated as follows: The charge q induced upon the ball when raised to a height h above the ground must be such as to maintain it at earth potential, for the ball is earth-connected through the electrometer. Thus if V is the undisturbed potential of the air at the point occupied by the centre of the sphere of radius r, we have

$$V + q/r - q/2h = 0,$$

the third term allowing for the effect of the electrical image of the charged sphere in the earth. If the movement of the electrometer caused by raising the ball from its case to the height h is y, we have

$$q = -ky$$

and $V = k(1/r - 1/2h)y$.

This observation therefore gives for the mean potential gradient between the ground and the height h,

$$F = V/h = (k/h)(1/r - 1/2h)y.$$

The same equation relates the sudden changes of field ΔF to the corresponding sudden movements Δy of the electrometer. Since these relations, like those derived for the test-plate, are derived on the assumption of a perfectly plane earth, small corrections must be applied to allow for the effect of the concentration of the lines of force upon the hut used to house the recording instrument.

The effects obtained from the elevated sphere become too small for accurate measurement when the storm is more than 20 km. away; for still greater distances a wireless aerial has been used by some workers. Such a wire can, however, give rise to unwanted corona (point-discharge) effects if the field is fairly big; these are avoided by the use of the large sphere. Thunderstorm fields are of the same order as the fine-weather field, about 100 v./m., when the cloud is 20 km. off. At distances of 6 to 8 km. they often reach 5000 v./m., and when the storm passes overhead values of 10,000 to 20,000 v./m. are commonly observed, while the sudden changes may exceed these figures very considerably for a small fraction of a second. In an active thunderstorm the fields are constantly altering as a result of the neutralisation of part or all of the charge by lightning discharges, of the subsequent rebuilding of the destroyed charges, and of movements of the cloud as a whole.

(*b*) *The amplifier-oscillograph method.* This was developed by Appleton, Watson-Watt and Herd[5] for the investigation of very rapid field-changes from distant storms (atmospherics) and has been used by them and others to record the details of field-changes which the capillary electrometer method, which has a time discrimination of about 0·1 sec., is unable to resolve.

The essential features of the arrangement used are shown in Fig. 13, in which C_0 is the capacity of an exposed conductor such as the ball of Fig. 12 (or for very distant storms, a horizontal wire aerial) and R_1 is a resistance of some 50,000 ohms inserted to render the aerial circuit aperiodic. C_1 is a condenser which is variable in steps to suit the field to be measured and R_2 is a high-resistance selected so that the time-constant C_1R_2 is large compared with the duration of the field-change to be recorded.

If the field alters by ΔF as a result of a change ΔV of the potential at the effective height h of the exposed conductor, a change $-\Delta q$ must take place in the charge on the conductor and the charge on the upper plate of C_1 must increase by $+\Delta q$.

This will produce a potential difference Δv across C_1 given by

$$\Delta v = \Delta q/C_1 = \Delta V - \Delta q/C_0,$$

since both these expressions give the change of potential at the point X.

Hence $$\Delta v = \frac{C_0}{C_0 + C_1} \cdot \Delta V = \frac{C_0}{C_0 + C_1} \cdot h\Delta F.$$

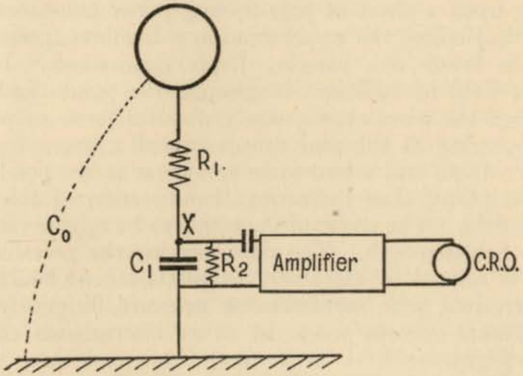

FIG. 13.—Input circuit for amplifier-oscillograph method (Appleton, Watson-Watt and Herd).

The arrangement thus acts as a condenser potentiometer and the amplified value of Δv causing the deflection of the oscillograph spot is proportional to the field-change. The oscillograph holds this deflection momentarily; its deflection drifts back to zero in a time set by the quantity C_1R_2 provided the time-constants of the amplifier are larger than this. C_1R_2 can usually be made long enough for the faithful recording of all but comparatively slow changes of field.

CHRIST'S COLLEGE

If C_1 and R_1 are removed from the circuit, the arrangement of Fig. 13 can be used to record the rate of change of field, dF/dt. For then

$$hdF/dt = dv/dt + (dq/dt)/C_0 = dv/dt + v/C_0R_2,$$

since the current i in the resistance R_2 is $v/R_2 = dq/dt$. If C_0R_2 is small enough, the second term is much greater than the first and $v \simeq hC_0R_2dF/dt$.

(*c*) *The Alti-electrograph.* This instrument was devised by Simpson and Scrase [6] for attachment to a small balloon to measure directly the electric fields at points near and within active thunderclouds. Two separate iron pins resting upon a sheet of pole-finding paper are connected to vertical wires, the upper one on a bamboo framework and the lower one hanging freely downwards. In an electric field of sufficient magnitude for point discharge to charge the wires appropriately, the difference of potential appearing at the pins causes a small current to flow between them and a blue stain to appear at the positively charged point, thus indicating the direction of the prevailing field. The width of the stain can be related roughly to the field-strength. The disc carrying the pole-finding paper is rotated by clockwork to give a record which can be correlated with simultaneous pressure, humidity and temperature records made by other instruments carried on the flight.

(*d*) *The electric fluxmeter method.* The principle of this method has been described in § 17. When the device is carried in aircraft inside clouds, care has to be taken to eliminate the " autogenous " field produced by the charge collected by the aircraft itself. This is done (Ross Gunn) by employing two fluxmeters, one on the upper surface of a wing and the other, facing downwards, on the lower surface. For one of these the external and the autogenous field will be in the same direction and for the other they will be in opposite directions. The deformation of the field due to the aircraft itself can be corrected for by model experiments. Owing to icing difficulties this device cannot be used at levels where the temperature is much below 0° C.

A high-resolution form of this instrument devised by Malan and the writer has been employed in observations at ground-level to fill the gap between method (*a*) which has a poor time-resolution for field-changes and method (*b*) which is more suitable for rapid than for slow field-changes. Records of the output of the fluxmeter are made on a magnetic tape recording machine and those portions of interest are played back and recorded photographically later on, thus providing economical high-speed recording of transient field-changes.

(*e*) *The point-discharge galvanometer method.* If a fine point erected on a tall pole is insulated and connected to earth through a recording galvanometer, the current I due to point-discharge is roughly related to the prevailing field F by the equation $I = a(F^2 - M^2)$ where a and M are constants for the particular arrangement used. This method, which has been examined in detail by Whipple and Scrase,[8] provides a useful general record of the sign and approximate value of F for near storms but its time-resolving power is low and the above equation does not hold with much accuracy when the usual thunderstorm winds are blowing.

33. The Distribution of Electricity in Thunderclouds. All the methods described in § 32 have contributed to the solution of the problem of the distribution of electricity in the thundercloud. After some controversy it is now generally agreed that the main distribution is one in which an upper positive charge lies several kilometres above a lower negative one. This pattern may be repeated in different parts (cells) of a large cloud. The arguments of C. T. R. Wilson [2] on which this conclusion was first based, are set out below.

Consider a charge Q which occupies a spherical region within the cloud, its centre at height H above the ground.

The field at P (Fig. 14) is that due to an electric doublet of moment $2QH$, the opposite charge induced on the earth by Q being replaced by the image-charge $-Q$. This field is then $F = 2QH(H^2 + L^2)^{-3/2}$ and is directed vertically downwards if Q is a positive charge.

A lightning discharge involving Q may be of two kinds. If it is a discharge within the cloud, Q may most simply be considered to move vertically from a height H_2 to a height H_1, where it neutralises an opposite charge $-Q$. If it is a discharge to ground, Q disappears and F becomes zero. The net field-change produced by a cloud discharge is then

$$\Delta F_C = F_1 - F_2 = 2Q[H_1(H_1^2 + L^2)^{-3/2} - H_2(H_2^2 + L^2)^{-3/2}] \quad . \quad (3.1)$$

and that due to a discharge to ground is

$$\Delta F_G = O - F = -2QH(H^2 + L^2)^{-3/2} \quad . \quad (3.2)$$

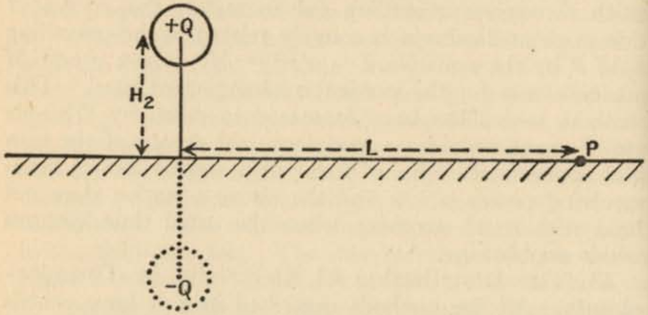

FIG. 14.—Thundercloud charge above the earth.

From observations of the sign of these field-changes one can infer the polarity of the cloud dipole responsible for them. If Q is a positive upper charge ($H_2 > H_1$) eqn. (3.1) shows that ΔF_C will be positive for all values of L less than a critical value L_0 and negative for all values of L greater than L_0. L_0 will be given by

$$H_1(H_1^2 + L_0^2)^{-3/2} = H_2(H_2^2 + L_0^2)^{-3/2}.$$

Again, if Q is a negative lower charge moving from height H to ground, eqn. (3.2) shows that ΔF_G should be positive at all distances.

Observations of ΔF_C and ΔF_G at various distances have verified these deductions and hence indicate that the majority of thunderclouds carry positive charges above and negative charges below. Table VI gives the results of observations of ΔF_G which show predominantly positive field-changes at all distances.

TABLE VI

SIGN OF FIELD-CHANGES DUE TO FLASHES TO GROUND AT ALL DISTANCES

Positive.	Negative.	Observer.	Place.
67	4	Schonland	Somerset East, S. Africa
267	16	Halliday	Johannesburg, S. Africa
77	13	Jensen	Nebraska, U.S.A.
411	33		

As concerns ΔF_C, eqn. (3.1) and (3.2) show that both cloud and ground flashes should produce positive field-changes when the storm is near but that the former should reverse to negative changes when the storm is at a distance greater than L_0. Since in storms in tropical and semi-tropical regions cloud discharges considerably outnumber ground discharges, we should expect to find in such storms a preponderance of negative field-changes at a distance and of positive ones when the cloud is near. That this is the case is shown by the data from South Africa in Table VII.

The South African observations in the Table offer strong evidence for the reversal effect. Those from Great Britain support them in showing a considerable preponderance of positive field-changes at small distances but in this case field-changes of either sign are about equal in number at distances exceeding 15 km. The reason for this is that in Britain cloud and ground discharges are approximately equally frequent whereas in South Africa the former considerably outnumber the latter.

Observations of actual zero values for ΔF_C at the reversal distance ($L = L_0$) are rare. In practice the flash does not usually take a vertical path between H_2 and H_1. For this reason and because H_2 and H_1 vary from one flash to another the concept of a reversal distance must be replaced by that of a reversal zone, in which small field-changes of either sign are equally probable. This zone may extend over a range of from 7 to 15 km. from the point of observation.[4]

TABLE VII

Signs of Discharge Field-changes of all Types

$L_{\text{kms.}}$	$\Delta F_{+\text{ve.}}$	$\Delta F_{-\text{ve.}}$	Observer.	Country.
Distant Storms.				
> 15	250	2375	Schonland	S. Africa
> 10	113	644	Halliday	S. Africa
> 15	1699	1546	Wormell	England
Near Storms.				
< 5	54	16	C. T. R. Wilson	England
< 7	563	70	Schonland	S. Africa
< 6	86	18	Halliday	S. Africa
< 5	143	63	Wormell	England

Further strong evidence that the lower charge is usually a negative one has been given by direct observations of the polarity of lightning discharges to earthed metal structures (§ 41). These have been made in Europe, Russia, Japan and the United States and show that 90 per cent. or more of such flashes discharge negative electricity. The total number of direct observations available in 1941 was 2505 and 2422 indicated negative discharges.

In conformity with the general picture of the charge distribution outlined above, it is generally observed that the electric field from a nearby storm just before it discharges (the pre-discharge field) is negative, since the lower pole is nearer to the point of observation, while at a considerable distance it is found to be positive, the effect of the upper pole being then predominant. In these observations there is, however, a complicating factor of some interest. It is frequently observed that just before a very close discharge to ground the pre-discharge field becomes small and may even become positive.

This effect is explicable if there exists below the lower negative pole of the cloud an additional small " pocket " of positive charge either in the base of the cloud or in the air just below the base. Direct evidence for this pocket is discussed in the next section. Indirect evidence is afforded by the observation of Wormell [3] that there is a type of cloud (or air) discharge which reverses in sign at a much shorter distance than the normal and higher discharge between the main poles of the cloud and which is responsible for some of the apparently anomalous negative values of ΔF from near discharges quoted in Table VII. The type β leader process discussed in § 36 leads to the same conclusion.

All these deductions as to the distribution of charge in the thundercloud have been very beautifully confirmed and extended by the alti-electrograph balloon observations of Simpson, Scrase and Robinson.[6, 9] These have shown that in a balloon ascent near to or through the active centres of a thundercloud, the field is frequently positive on entering the cloud-base, after which it becomes negative for about 1 km., then positive for as much as 4 km. and thereafter it remains negative to the top of the cloud and beyond. The results from a typical sounding of this type, in which the width of the looped lines indicates the order of magnitude of the field and negative and positive fields are shown as open and hatched loops respectively, are given in Fig. 15 (*a*). The figure also shows the distribution of the three charges, *P*, *N* and *p* which were suggested to account for the observations. The values found for the heights of these charges, and so for the temperature levels at which they lie vary considerably from storm to storm. It is not possible to be certain what path the balloon has followed in its fairly slow flight through the cloud nor whether the charge distribution derived from the results is that existing at the moment a flash takes place, but typical values for large thunderclouds in England are *P*, 7 km. (– 27° C.), *N*, 2·5 km. (– 4° C.) and *p*, 1·5 km. (+ 2° C.).

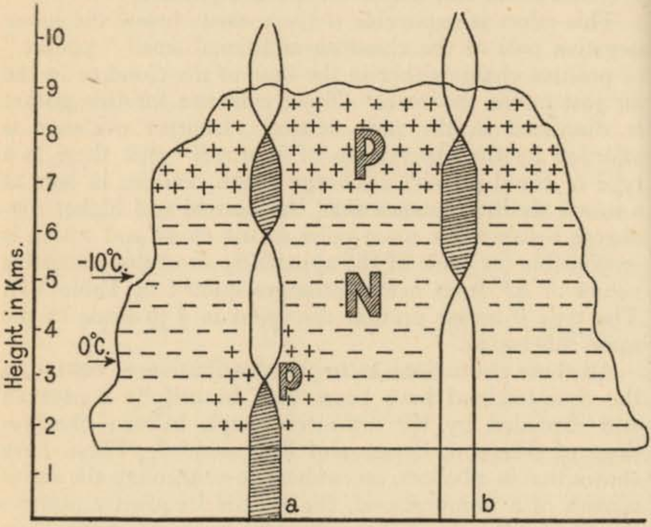

FIG. 15.—Typical alti-electrograph soundings (after Simpson, Scrase and Robinson)

The small *p* charge (which is called *Q* by Simpson, Scrase and Robinson) is not always observed, either because it is absent or because its small magnitude and dimensions cause the balloon to miss it. In such cases the record obtained is as in Fig. 15 (*b*).

Further confirmation of the general $P - N$ dipole distribution in thunderclouds has been obtained by Gish and Wait (§ 30) using the electric fluxmeter in aircraft flights above the tops of thunderclouds.

34. Magnitudes of the Electrical Quantities Involved in Thunderclouds. When the distance L is large compared with the heights H_1 and H_2 of the main P and N cloud-charges, eqn. (3.1) takes the form

$$\Delta F_C = 2Q(H_1 - H_2)/L^3$$

or

$$2Q(H_2 - H_1) = -\Delta F_C \times L^3 \quad . \quad . \quad (3{\cdot}3)$$

The quantity on the left is the difference between the electric moments of the thundercloud charges taking part in the discharge and is called the electric moment, M, of the discharge dipole. To determine it, the field-change due to a distant cloud flash and the horizontal distance L are required. The most frequent value of M determined in Great Britain[3] is 110 coulomb-km. though values as high as 1200 coulomb-km. have been observed. The average value for South Africa is 90 coulomb-km.[4]

Since $H_2 - H_1$ is approximately 2 km. the average quantity of electricity, Q, neutralised in a cloud discharge is about 20 coulombs. The range of variation of Q in individual discharges is from 2 to 200 coulombs if the vertical length of the discharge is supposed constant. If this length could be measured in every case, the variation would probably be much smaller.

The average electric moment of discharges to ground, which usually involve the lower negative pole only, is found to be of the same order as that of discharges within the cloud.

A single cloud-flash appears to discharge most of the section of the cloud which is producing lightning at the time. A fairly active thunderstorm produces one flash every 20 sec. To feed these flashes alone, it would have to separate out charges of opposite sign at an average rate of 1 coulomb/sec. and the generating mechanism would have to provide an average current of 1 amp. Reference has been made in § 32 to the rapid rate at which the initial regeneration of the field takes place after a flash. In the case of cloud discharges it is usually such as to be able to restore the whole field in about 5 sec. if it continued for that time. The exponential form of the recovery curves suggests that this initial rate of supply of charge remains constant but that an opposite neutralising effect occurs which tends to diminish the rate of increase of the charges and which is proportional to the quantity of charge present at any time.

That neutralising or dissipating effects of this nature exist below the cloud is shown by measurements of the mean current from positive point-discharge flowing upwards from the earth to the *N* charge (§ 29). This current is of the order of 2 amp. A similar dissipating current bringing negative charge down from the upper air to neutralise the *P* charge has been observed (§ 30). Its average value is about 1 amp. but it has been found on occasion to reach 6·5 amp.

The electrical energy generated by the cloud is thus only partially spent in the production of lightning discharges; just before a discharge most of the power of the generating mechanism is being used to overcome leakage effects opposing the building up of the main dipole. It may happen that the field, owing to this leakage load, never reaches the sparking value and the cloud, though electrified, does not produce lightning.

An estimate of the maximum potential reached by the two main cloud charges can be found from an argument given by C. T. R. Wilson.[10] Consider first the simple case of a charge Q contained within a sphere of radius R. The electric field at the surface is then given by $F = Q/R^2$. The maximum value of F for a discharge to occur is known from experiments by Macky on the behaviour of water drops in strong fields to be about 10,000 v./cm. or 33 e.s.u. This figure is one-third of that in normal air; the reduction is due to the fact that in strong electric fields a water-drop becomes unstable and pointed filaments are drawn out from opposite ends. For an average value of Q of 20 coulombs or 6×10^{10} e.s.u. the above equation gives $R = 427$ m. and the dimensions of the charged region are of the order of 1 km. The potential at the surface of the sphere is given by

$$V = Q/R = FR = 4{\cdot}2 \times 10^8 \text{ volts.}$$

If the radial electric force everywhere within the sphere were such as to produce sparking, the potential at the centre would be $2FR$ or about 8×10^8 v. Two such spheres carrying opposite charges would therefore be at a difference of potential of $1{\cdot}6 \times 10^9$ v. A similar result follows from the very different assumption that the two poles of the cloud are distributed in horizontal layers with a neutral region of thickness D, between them. If F is 10,000 v./cm. and D is 2 km., the difference of potential between them is 4×10^9 v.

Recent work on the manner of development of the lightning discharge (Chapter IV) suggests, however, that the average value of the field at the surfaces of the charged regions is considerably lower than 10,000 v./cm. Once a short leader streamer (§ 36) has been produced in a limited region where the field is momentarily 10,000 v./cm., its further development may be possible in fields which are only one-tenth as large. Actual measurements of fields inside thunderclouds by Ross Gunn[7] support this conclusion, for the highest "undistorted field" value reported was 1600 v./cm. just before the observing aircraft was struck by a discharge and the average maximum field for nine different thunderclouds was 1300 v./cm. If F is one-ninth of the value used above, the diameter of the spherical charged region ($2R$) becomes about 2·5 km., the potential reached before discharge $1{\cdot}4 \times 10^8$ v. and that between the poles $2{\cdot}8 \times 10^8$ v. These are likely to be lower limits for the potentials.

The disappearance of 20 coulombs in a lightning discharge is thus estimated to involve the dissipation of at least 6×10^{16} ergs or $1{\cdot}6 \times 10^9$ calories, most of which is converted into heat along the path of the lightning flash. A cloud giving one flash every 20 sec. is dissipating electrical energy in the form of lightning at an average continuous rate of at least 3×10^5 kw. To maintain in addition currents of the order of 2 amp. above and below it involves the provision of a further 6×10^5 kw.

REFERENCES TO PAPERS

[1] C. E. P. Brookes, *Met. Office Geophys. Memoir*, No. 24 (1925).

[2] C. T. R. Wilson, *Proc. Roy. Soc.*, A, **92**, p. 555 (1916); *Phil. Trans.*, A, **221**, p. 73 (1921).

[3] T. W. Wormell, *Phil. Trans.*, A, **238**, p. 249 (1939).

[4] B. F. J. Schonland, *Proc. Roy. Soc.*, A, **118**, p. 233 (1928); A, **114**, p. 229 (1927) (with J. Craib).

[5] E. V. Appleton, R. A. Watson-Watt and J. F. Herd, *Proc. Roy. Soc.*, A, **111**, p. 654 (1926).

[6] G. C. Simpson and F. J. Scrase, *Proc. Roy. Soc.*, A, **161**, p. 309 (1937).

[7] R. Gunn, *Phys. Rev.*, **71**, p. 181 (1947).

[8] F. J. Whipple and F. J. Scrase, *Met. Office Geophys. Memoir*, **68** (1936).

[9] G. C. Simpson and G. D. Robinson, *Proc. Roy. Soc.*, A, **177**, p. 281 (1940).

[10] C. T. R. Wilson, *Journ. Franklin Inst.*, **208**, p. 1 (1929).

CHAPTER IV

THE LIGHTNING DISCHARGE

35. General Features. The most frequent form of lightning is the *cloud discharge*. Though most cloud flashes are known from field-change records to be discharged between the main *P* and *N* charges of the cloud dipole (Fig. 15), there is evidence that others pass between the *N* charge and a lower positive charge, either the small *p* pocket of Fig. 15 (*N-p discharge*) or a positive space-charge in the air below the cloud (*air-discharge*).

A similar type of flash between the *P* charge and negative charge drawn down from the upper air to the top of the cloud or to the air immediately above it, would account for many cloud-flashes which are seen at night to be restricted to the cumulus heads of thunderclouds and the regions above them.

Most ground-discharges, as has been shown in § 33, remove negative charge from the *N* region. There is a good deal of evidence to show that they are usually triggered off by the strong field between the *N* and *p* charges or the positive space-charge in the air, thus extending an *N-p* or an air discharge all the way to the ground. The first strokes of such discharges thus bring to ground the residue of the negative charge left after the smaller and lower positive charge has been neutralised.

A general account of the various forms of the lightning discharge and their effects will be found in reference 1.

Many lightning flashes, both within the cloud and to ground, are observed to flicker. Early camera studies by Walter and others showed that this is due to the fact that the discharge consists of a number of separate components or strokes. Though the whole flash often occupies more than a second, the luminosity of the component strokes is generally over in a few milliseconds. After each stroke the ionisation along its channel, though much reduced by electron-attachment and by recombination and diffusion of the ions, usually persists in sufficient measure to provide the following stroke with a ready-made path which it follows. If, however, the time-interval is exceptionally long, the channel ionisation is too much weakened and the following stroke may take a different path.

The channel is often shifted several metres by the wind in the intervals between strokes ; the discontinuous nature of the discharge is then shown as *ribbon lightning*.

Observations with moving cameras show that the branching of the lightning discharge to ground is almost always confined to the first stroke of a series.

The maximum number of strokes so far observed from a single flash to ground is 42. The most frequent number is 3 or 4 and single-stroke discharges are quite common. The time-interval between strokes is variable, the most frequent value lying between 0·03 and 0·07 sec.

36. Luminous Streamer Processes in the Discharge. (*a*) *Experimental methods.* To investigate the nature of the luminous processes in the lightning discharge much use has been made of a method of high-speed photographic recording devised by Sir Charles Boys. If a camera is so constructed as to provide relative motion of lens and film, the recorded image of any series of luminous events in two-dimensional space will be distorted in such a way that later events are progressively displaced in position relative to earlier ones. Comparison of the distorted picture of a lightning discharge with one taken on a fixed camera would then show the order in which light was emitted at different parts of the channel. In practice, however, this is very difficult unless the distortion is large, because both fixed and moving cameras require an accurately placed reference line for the comparison. The difficulty is avoided in the Boys camera by the use of two lenses rotating in a circle ; these give two pictures with distortions in opposite directions and a reference line provided by a diameter of the circle. This principle has been used to follow the processes in the discharge with a time resolution of 1 microsec. Full details of the Boys camera and of later forms of camera which have been used by the writer and his associates in South Africa are given in references 1 and 2.

(*b*) *General results.*[1, 3] Each separate stroke of a discharge to ground has been found to be of a dual nature. A downward-moving luminous process, the *leader* streamer, is followed upon arrival near the surface of the ground by a much more rapid and intensely luminous main or *return* streamer which re-traverses the channel in the reverse direction. The leader to the first stroke is of a different nature to the leaders to subsequent strokes ; it takes about ten times as long to reach the ground and its luminosity is intermittent instead of continuous.

Fig. 16 (*b*) (after Schonland, Malan and Collens) illustrates these points in the case of a discharge of three strokes which has been photograped by a single-lens camera in which the lens has moved in a straight line from left to right. It is actually a picture in two dimensions of space and one of time since horizontal displacements with reference to the fixed picture (*a*) represent time as measured from the point P. These displacements are not shown on a uniform scale in the diagram. The Boys camera gives two simultaneous pictures from two lenses moving in opposite directions and thus avoids the use of a fixed picture.

The figure shows that the leader to the first stroke proceeds in a series of steps. Each of these is about 50 m. in length (range 10 to 206 m.) with fairly regular pauses (31 to 91 microsec.) between the steps. After each pause there is almost always a fresh direction of travel, and these changes in direction are responsible for the tortuous form of the lightning channel. The branches of the leader, which are usually found in the first stroke only, are similarly developed in steps and move outwards from the main channel. Each step is the bright tip of a long streamer stretching down from the cloud, the stem of which is faintly illuminated.

When the most advanced branch of the stepped leader system approaches within from 15 to 50 m. of the ground,

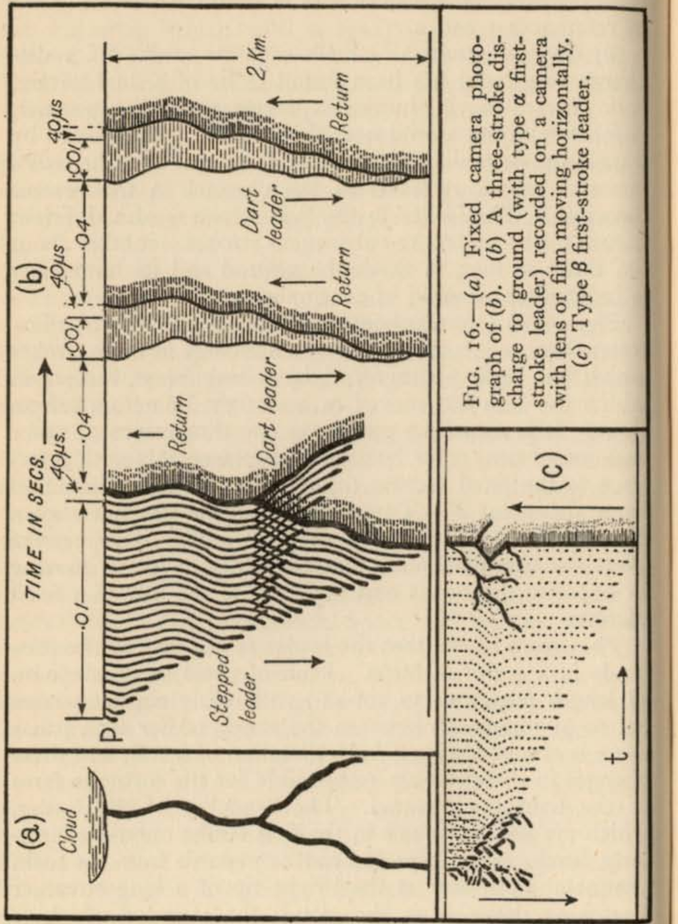

FIG. 16.—(*a*) Fixed camera photograph of (*b*). (*b*) A three-stroke discharge to ground (with type α first-stroke leader) recorded on a camera with lens or film moving horizontally. (*c*) Type β first-stroke leader.

the return streamer jumps the gap and passes rapidly back from ground to cloud, running outwards along each branch. This process takes microseconds to traverse parts of the channel originally created in milliseconds by the stepped leader.

Stepped leaders are of two types. In the type α leader (Fig. 16 (*b*)) the steps are short and are often difficult to photograph because they are not very bright. In the type β leader (Fig. 16 (*c*)) the upper part of the channel is formed by longer and brighter steps than the lower portion which has very weak type α steps. This difference is associated with extensive branching in the upper part of the channel. It suggests that there is a stronger electric field, due to positive space-charge in the air below the cloud, at the top of the leader path than later on. Electric field studies of the leader process show that although photographic records of events in the air below the cloud indicate that only 10 per cent. of leaders are of type β, this type is actually the most frequent one. The β stage in the leader is usually hidden within the cloud, the high field regions required to produce it presumably lying between the *N* and *p* charges of Fig. 15.

In the case of strokes subsequent to the first, the leader is not stepped but travels continuously to ground. As shown in the figure, it records as a continuous bright line followed (in the time dimension) by fainter illumination. This indicates that the leader takes the form of a fast-moving javelin, or dart, with a bright point about 40 m. long and a long, faintly-illuminated stem extending back to the cloud. For this reason it is termed a *dart* leader. The only exceptions to the continuous movement occur when an unusually long time-interval has elapsed since the preceding stroke ended. The dart leader is then much slower; frequently it shows the stepped method of progression over part of its length and takes a path different from that of its predecessors.

A typical first-stroke return streamer record is shown in Fig. 17. Times are shown in microseconds after the start at O.

The streamer covers the distance of about 2 km. between ground and cloud in 60 microsec. The figure shows that it continues to remove charge from the cloud for a further 600 microsec. The brightness and velocity of the streamer decrease as it travels upwards but are revived suddenly

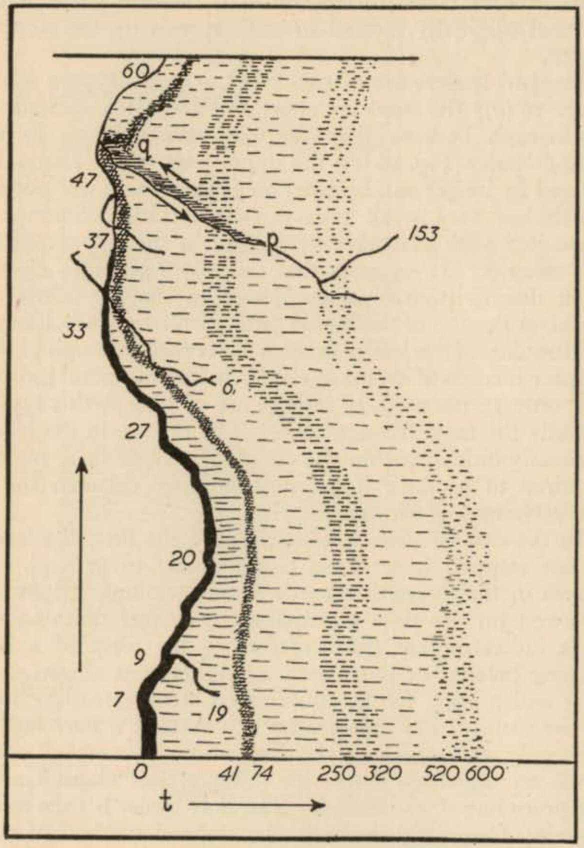

FIG. 17.—Return streamer time-table (Malan and Collens). (Times in μ-secs.)

as each of the branching points is reached, indicating, as has been shown by the detailed studies of Malan and Collens,[4] that the charge on a branch calls into existence a fresh flow of current from the ground. The current at the base of the channel is subject to violent fluctuations (*M* components) from these effects. Further fluctuations, (like the last two in Fig. 17), are produced by processes which are hidden in the cloud. In the case of subsequent strokes these latter fluctuations are the only ones observed because there are usually no branches to be traced by the return streamer.

The average velocities of the three processes which have been described, the dart, return and step streamers, are 2,000, 20,000 and more than 10,000 km./sec. respectively. Although individual steps are pushed forward at > 10,000 km./sec., the step-streamer process as a whole, as Fig. 16 shows, is slow, for its numerous pauses cause it to have an *effective* velocity which seldom exceeds 300 km./sec. and can fall as low as 100 km./sec.

In the case of very tall structures like the Empire State Building, 1250 ft. high, McEachron and Hagenguth[5] have shown that the lightning discharge frequently begins with an *upwardly* directed stepped leader which travels from ground to cloud and is not followed by a return process. The current continues to flow for some time in the channel thus formed, after which a normal downward dart leader, followed by an upward return streamer, creates the second and later strokes.

37. Electrostatic Field-changes associated with Lightning Processes. The processes described in § 36 give rise to electrical field-changes which are best studied by the amplifier-oscillograph method of § 32 (*d*) or the electric fluxmeter of § 17.

Fig. 18 (A) shows the general form of the field-changes produced by a ground discharge of three strokes when near (3-5 km.), but not directly over, the recording instrument.[6, 7] Fig. 18 (B) shows the form taken by the same field-changes when the flash is fairly distant (12-20 km.). In this case, as described in § 40, induction

CHRIST'S COLLEGE

and radiation field-changes would be superimposed on the electrostatic field record but for the sake of clarity they have been omitted from Fig. 18 (B).

The manner in which various sections of the field-change record have been identified with the various streamer-processes observed with the Boys camera is shown in the figure. The *l* portions of the record correspond to the downward movement of leader streamers and the two *r* portions, r_b and r_c, with the upward return streamer. The remaining *j* portion corresponds to a streamer process which cannot be photographed since it takes place within the cloud. From a detailed study of records of this kind the following conclusions can be drawn :

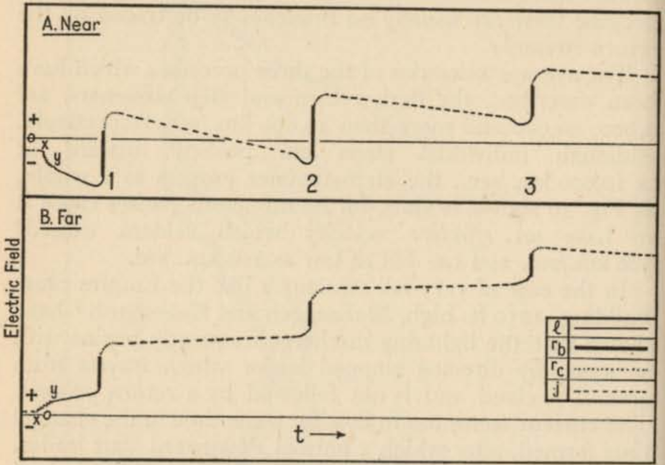

FIG. 18.—Electrostatic field-changes due to a discharge to ground (Schonland and Malan).

(1) *Leader* (*l*) *field-changes.* These, which produce hook-shaped negative changes when nearby and parabolic positive changes when at a distance, indicate clearly the lowering of negative charge towards the ground by the leader streamers. Since they remove charge from an initial height, H_2, to a lower effective height, H_1 (§ 33, eqn. 3.1) when the charge is distributed along the channel, their net field-changes reverse in sign with distance.

The leader to the first stroke in Fig. 18 shows a field-change which has a much larger slope over the first section *xy* than later. This is because it is a type β leader (§ 36 *b*), and lowers charge more rapidly in its initial than in its final stage.

As Fig. 18 shows, dart leaders to subsequent strokes occupy a much shorter time than the first stepped leader but show a similar form of field-change.

It is of some importance that the durations of leader field-changes are often much longer than the durations of the corresponding leader processes recorded by cameras operating at the same time. For this implies that a considerable length of the leader channel, sometimes of the same order as the visible length below the cloud, is hidden inside it and that the *N* charge creating the leader is at a considerable height.

(2) *Return streamer* (r_b *and* r_c) *field-changes.* Since these involve the removal to ground of the negative charge on the leader (r_b) and remaining in the cloud (r_c) their field-changes are positive at all distances. The rapid r_b change occupies a time of the order of 50 microsec. and has a complex fine structure which is not shown in Fig. 18. The slower r_c change, which occupies 1 or 2 millisec., is considered to be due not only to the final mopping-up of the charge remaining in the region which produced the leader, but also to an upward flow of positive charge induced on the top of the channel by *N* regions as yet untapped by the discharge.

(3) *Junction* (*j*) *field-changes.* These changes occupy the intervals between strokes. Since they are negative when near and positive when far, they must be created by slow movements of charge within the cloud, in which positive charge moves upwards or negative charge moves downwards. There is some evidence [7] that in most cases they are due to the upward movement of positive streamers

from the top of the channel to new regions of negative charge, which they render conducting.

(4) *The mechanism responsible for successive strokes.* It has been shown by Malan and the writer[7] that a number of lines of evidence leads to the conclusion that each stroke of a discharge to ground comes from an *N* region which is higher than that producing its predecessor. From this and other information it is deduced that the *N* regions creating a discharge of many strokes are not haphazardly distributed within the cloud but form a single and approximately vertical column which is intermittently discharged and which extends upwards to considerable heights. A ground flash of many strokes may come from a column 6 km. long. These authors have shown that the pauses between strokes represent the time required for the junction-streamer process to move upwards and render the next section of the charged column sufficiently conducting for it to discharge down the channel as a leader.

38. The Mechanism of Streamer Development.[8] It is known from laboratory studies of gas-discharge processes that an electrical streamer possesses a conducting stem filled with ions and electrons and that the motion of these charged particles in the field prevailing inside the stem constitutes a current which keeps the streamer surface, and its tip in particular, supplied with charge as it moves forward. For forward movement the electric field at the tip must be sufficiently great to cause free electrons in the air in front of the advancing streamer, some of which are created photoelectrically by ultraviolet light from the streamer itself, to ionise by collision. This ionisation is cumulative and leads to the production of electron avalanches which extend the streamer tip.

The streamers give out light at and behind their tips because excitation as well as ionisation occurs in the strong field region. Since the excited atoms and molecules return a short time later to their ground-states, and since the tip moves forward during this interval, the region of intense luminosity extends backwards for a limited distance behind the tip.

In the case of the dart leader, the negatively charged streamer is guided along a pre-existing channel because ionisation created by previous strokes exists in front of the advancing tip. It can be shown that the velocity attained by the streamer in this case should be proportional to the cube root of the pre-existing electron density, n, in front of it.[8] When the interval between strokes is short, n is large and high dart-leader velocities are possible, but when the interval is long and n has been reduced by electron attachment to atoms and molecules and by diffusion processes, the dart leader velocity is low.

The fast return-streamer also travels along a previously ionised channel, more heavily ionised and fresher than that available to the dart leader. In this case, however, as was first shown by Simpson, the tip of the streamer is positively charged and the electron avalanches in front of it move inwards in a strongly converging field. These factors account for the high velocity of the return streamer.

The stepped leader, which is the first of all the discharge processes, appears from the photographic records to involve advance into virgin air and to follow no pre-existing trail of ionisation. Its slow effective velocity, which is of the same order as the minimum velocity at which an electron can ionise by collision, coupled with the very high velocity of the step streamers themselves suggests, however, that it too follows an ionised path prepared by an unrecorded *pilot* streamer, or a cone of such streamers, which precedes it. On this view the steps are secondary effects superimposed upon the pilot streamer and catching up at intervals with it.

Such pilot streamers would be controlled in direction by purely local variations in the electric field in front of their tips and their paths will be tortuous and branching, especially if the air contains concentrations of positive ion space-charge.

A fuller discussion of these streamers and of the unsettled question of the mechanism responsible for the fairly regular pauses observed in the stepped process will be found in references 9, 10 and 11. It is generally agreed that these pauses owe their origin to a property of the streamer itself[8] and not to any intermittent feature in the discharge from the cloud but whether this property involves a periodic transition from a "glow" to an "arc" form of streamer or is simply due to the streamer overreaching itself and requiring a pause to renew the field-strength and the conductivity of its stem is not yet certain.

39. Discharges which do not Pass to Ground. Discharges which do not reach the ground, whether they are air-discharges or take place entirely within the cloud are found to be of the leader type, without the usual return process. Air discharges when photographed show complex branched type β step-leader streamers with dart leaders superimposed upon them. They travel towards local concentrations of positive space-charge below the cloud and weak "recoil" streamers moving back from such concentrations have been observed. An example of recoil is shown along p, q in Fig. 17.

The study of discharges which are entirely within the cloud has so far been possible by field-change methods only. These show that cloud discharges are of the same character as air discharges and can occupy times of the order of milliseconds.

40. Induction and Radiation Field-changes. The form of the electrostatic field-changes associated with lightning discharge processes has already been described. The complete electric field from a small dipole of moment M is actually the sum of the components represented by

$$E = M/r^3 + (dM/dt)/cr^2 + (d^2M/dt^2)/c^2r,$$

expressed in e.s.u. where c is the velocity of light and r the radius vector to the centre of the dipole. The first term corresponds with eqn. (3.1) for the electrostatic field, the second is the induction or magnetic term and the third the well-known expression for radiation from a Herzian dipole. Within 10 km. of a lightning flash the first term is much larger than the other two and the field-changes caused by lightning produce mainly electrostatic effects. At 25 km., however, the last term shows up strongly and at 50 km. and beyond, it is predominant.

The radiation field is proportional to d^2M/dt^2 and so to the rate of change of the current in the portion of the discharge concerned at time $t - r/c$. Radiation effects from lightning are therefore most prominently associated with processes which involve the rapid starting and stopping of streamers carrying large currents. These are (*a*) the stepped processes found in first stroke leaders and in cloud and air discharges and (*b*) the return processes in each stroke.

The radiation effects of these processes are found to constitute the majority of atmospheric (static) disturbances on the longer wavelengths used in radio communication and can be observed more than 4000 miles away from the parent discharge. They have been of practical value in the location of thunderstorms for meteorological purposes (sferics) and the manner in which they are modified by successive reflections from the ionosphere has been used to give information about the height and nature of its lower D layer.[1, 12]

41. Currents and Current Variations. Much work has been done to determine directly the magnitudes of lightning currents to ground and their time-variations.[9, 10] One instrument used for determining peak currents through elevated conductors and towers struck by lightning is a simple recording voltmeter, known as a Klydonograph, placed across one of a string of insulators acting as a potential divider and connected between the conductor and the earth. This device makes use of the fact that the diameter of the "Lichtenberg" figure produced around a positive point resting on a flat photographic film backed by an earthed plate is proportional over a certain range (up to 25,000 v.) to the potential difference between point and plate. To record discharges of both polarities two instruments, oppositely connected, are necessary.

The Klydonograph method requires a knowledge of the surge impedance between the conductor and the earth and this is not always easy to determine. More use has recently been made of "magnetic links," two or three bundles of short unmagnetised steel strips of high retentivity which are mounted at different known distances from the conductor. The peak currents passing in the discharge can then be found from the remanent magnetism in the links, calibration being effected with known currents.

In a development of this device called the Fulchronograph (Wagner and McCann)[14] the links are carried on an aluminium wheel which rotates between coils through which the lightning discharge passes. The remanent magnetism of successive links then gives the time variation of the current.

The time variation has also been determined by shunting special types of oscillograph across resistances of high current capacity. The data are, however, as yet insufficient for a generalised statement of the manner in which the current in a discharge varies with time in the successive strokes of a discharge. In first strokes it rises to a peak of average value 20,000 amp. (maximum observed value 250,000 amp.) in from 1 to 20 microsec., with a most frequent " time to peak " of 2 microsec. Its manner of decay is variable but it is usually stated to fall to half its peak value in about 24 microsec. (range 7 to 115 microsec.).

Continuing currents of considerable magnitude and of long duration have been found by McEachron[13] to follow strokes to the Empire State Building but in discharges to lower structures the current is usually found to have fallen to very low values ($< 0{\cdot}1$ amp.) a few milliseconds after the start of the return portion of the stroke.[14]

The quantities of electricity found by integrating the current-time curves of direct discharges are of the same order of magnitude (average 30 coulombs, maximum 164 coulombs) as those determined by the method of C. T. R. Wilson (§ 34).

REFERENCES TO PAPERS

[1] B. F. J. Schonland, *The Flight of Thunderbolts*, Oxford Univ. Press (1950).

[2] D. J. Malan, *Revue d'Optique*, **29**, p. 513 (1950).

[3] B. F. J. Schonland, D. J. Malan and H. Collens, *Proc. Roy. Soc.*, A, **152**, p. 595 (1935) ; *ibid.*, **168**, p. 455 (1938).

[4] D. J. Malan and H. Collens, *Proc. Roy. Soc.*, A, **162**, p. 175 (1937).

[5] K. B. McEachron, *Journ. Franklin Inst.*, **227**, p. 149 (1939).

[6] D. J. Malan and B. F. J. Schonland, *Proc. Roy. Soc.*, A, **191**, p. 487 (1947).

[7] D. J. Malan and B. F. J. Schonland, *Proc. Roy. Soc.*, A, **206**, p. 145 (1951) ; *ibid.*, **209**, p. 158 (1951).

[8] B. F. J. Schonland, *Proc. Roy. Soc.*, A, **164**, p. 132 (1938).

[9] *Reports on Progress in Physics*, Vol. X, art. by J. M. Meek and F. R. Perry, The Physical Society (1946).

[10] C. E. Bruce, *Proc. Roy. Soc.*, A, **183**, p. 228 (1944) ; C. E. Bruce and R. H. Golde, *Surge Phenomena*, Chap. I (1941), E.R.A., London.

[11] L. B. Loeb and J. M. Meek, *The Mechanism of the Electric Spark*, Stanford Univ. Press (1941).

[12] B. F. J. Schonland, J. S. Elder, D. B. Hodges, W. E. Phillips and J. W. van Wyk, *Proc. Roy. Soc.*, A, **176**, p. 180 (1940).

[13] K. B. McEachron, *Trans. Amer. Inst. Elec. Eng.*, **60**, p. 885 (1941).

[14] G. D. McCann, *Trans. Amer. Inst. Elec. Eng.*, **63**, p. 1157 (1944) ; *see also* C. F. Wagner, G. D. McCann and E. Beck, *ibid.*, **60**, p. 1222 (1941).

CHRIST'S COLLEGE LIBRARY

CHAPTER V

CHARGE GENERATION IN CLOUDS

42. Introduction. Many suggestions have been made as to the manner in which electric fields can be built up in clouds. Some of these have had to be discarded as facts on which to test them have accumulated. Most of the mechanisms left in the field may well be in operation at the same time and it is not easy to decide which are predominant in creating the electrification of thunderstorms.

An acceptable theory must (*a*) offer an explanation of polarity of the main *P-N* dipole; (*b*) give *P* and *N* charges each of about 30 coulombs under reasonable conditions of temperature, water-content and updraught; (*c*) operate at a temperature at least as low as − 10° C., and (*d*) be shown experimentally to work. Subsidiary mechanisms may be necessary to account for the lower *p* charge and for observations to be described in § 45 on the relation of the charge on thunderstorm rain to the prevailing field.

Of the theories which have been rejected because they did not meet all these conditions the best-known is that of Simpson, which proposed to account for the main electrification of the cloud by the Lenard effect, the breaking-up of water-drops in ascending air-currents. This mechanism, which left the larger fragments with a positive charge can satisfy neither the temperature condition (*c*) above nor condition (*a*), since it requires the lower portion of the cloud to be positively charged. It may, however, as Simpson and Scrase have suggested, be responsible for the small *p* charge of Fig. 15.

43. The Separation of the Opposite Charges once they are Formed. All the mechanisms which are at present under consideration satisfy condition (*a*) because they all suggest two entirely different types of carrier for positive and for negative charge and because all give a

negative charge to the heavier of the two types. Polarisation of the cloud in the requisite direction then follows automatically by the action of gravity in pulling the negative charges down (Fig. 19). When an updraught in the cloud is capable of supporting or even of driving upwards the heavier negative group of carriers (raindrops or snow or large ice-particles) the lighter ones (droplets or ice-fragments) are driven upwards still faster, so that the relative velocity of separation of the two groups is the same as if no updraught were present.

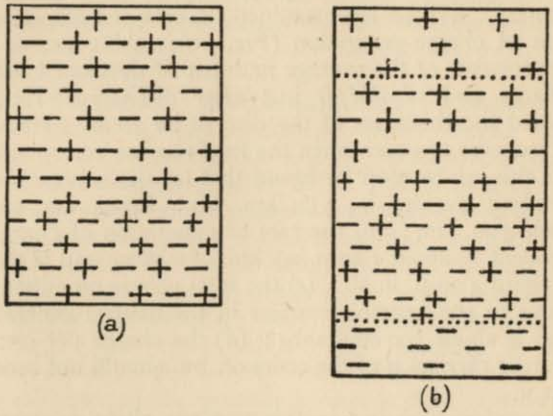

FIG. 19.—Gravitational separation of opposite charges (C. T. R. Wilson).

The shape of the field-changes immediately after a cloud-discharge (§ 32) suggests that the process of separation of the oppositely-charged carriers is one involving initially free fall under gravity and that this free fall is progressively retarded and finally reduced to nearly zero by the attractive forces between the separating charges. The rate of fall of the heavier carriers in this final stage must be considered just sufficient to maintain the charges constant in the face of dissipation currents within the cloud and directed to it from outside. Ultimately the field between the separated charges may become strong enough for a discharge to pass. If the charges are equal, this leaves the cloud in its initial state and the larger carriers in the neutral region of Fig. 19 once more separate out under the action of gravity. The relative velocity of separation has been taken by C. T. R. Wilson[1] to be of the order of 6 m./sec., the terminal velocity of a water-drop 0·15 cm. in radius. Relative to such a drop, the much slower positive droplets may be considered as stationary. Wilson has examined the case of a disc-shaped region of charge-generation (Fig. 19) making use of his measurements of the average moment M developed before discharge and of dM/dt just after discharge. He has assumed the thickness of the disc to be about 1 km. and the discharge to occur when the field reaches 10,000 v./cm. From this information he found that (*a*) the volume of the generating region is 4 cu. km. and hence its cross-section 4 sq. km.; (*b*) the rate of generation of charge is equivalent to about 1 amp./sq. km. of surface and is therefore about 4 amp. in all; (*c*) the total charge on either the positive or the negative carriers in the neutral generating region is about 500 coulombs; (*d*) the charge per c.c. on the larger carriers is of the order of, but should not exceed, 30 e.s.u.

The values deduced in this way are all in conformity with what is known of the quantities concerned. They are not seriously changed if modifications are made in the postulates to bring these into closer agreement with more recent information (thickness 3 km., breakdown field 1000-10,000 v./cm.).

A different type of cloud polarisation by gravity has been discussed by Frenkel in connection with fog-like clouds of small droplets. These are considered by him to be capable of capturing negative ions from the air (§ 44 (*b*)) which is then left with an excess positive charge.[2] The charged droplets are not held up completely by the internal field but drift slowly downwards at such a rate that the polarisation charge developed per unit area is equal to the charge neutralised by the internal conduction or dissipation current in the same time. The internal field, E_s, in this "stationary" condition is related to the equilibrium field, E_q in which the drops are fully supported, by the equation $E_s/E_q = \Delta/\lambda$ where λ is the ionic conductivity of the air and Δ is the much smaller contribution to the conductivity provided by the droplets. E_s is found to be of the order of 100 v./cm., a hundred times less than Eq. In clouds where the water content is greater, the positive ions will attach themselves to larger particles and the development of polarisation in the thundercloud then appears to become the same as that considered by C. T. R. Wilson.

44. Suggested Mechanisms of Thundercloud Electrification. (*a*) *Mechanisms which depend on the selective capture of negative ions by the larger carriers.* A mechanism suggested by C. T. R. Wilson depends upon the selective capture of negative ions from the air by large falling water-drops or ice-particles, the residual excess positive charge in the air being absorbed by very small droplets or ice-fragments in the upper part of the cloud.

For the mechanism to work, the ions concerned must be large ones (§ 1). Since their mobility is about 0·0003 cm./sec., the velocity with which they move under the influence of the prevailing electric field is much less than the velocity of fall of the larger carriers.

Consider a large drop or ice-particle inside a cloud in which a downwardly-directed field already exists. Owing to this field the drop will have a negative induced charge on its upper surface and a positive charge below (Fig. 20). In the case of an ice-particle, Chalmers has shown that dielectric polarisation will produce a similar distribution of charge.[3] As these larger carriers move downwards, they will approach positive ions moving down more slowly and negative ions moving up; encounters with other drops or ice-particles are considered to be relatively rare. The drops will behave differently in encounters with the two ion-streams. At the under-surface of a drop the negative ions will be attracted and the positive ions repelled and the former, if within a suitable distance, will be captured, giving the drop a net negative charge. The negatively-charged upper surface of the drop does not reverse the process, because here it is receding from the ion-streams and positive ions have been repelled away from the drop in the course of their passage past it. Innumerable episodes of this kind can cause the drop to acquire a considerable net negative charge, without appreciably altering the chance of further captures. A water-drop 1 mm. in radius, falling in a field of 10,000 v./cm. carries at each end an induced charge of 0·25 e.s.u. If it has collected a net negative charge equivalent to the very reasonable value of 12 e.s.u./c.c., the increase in the upper negative charge and the decrease in the lower positive charge will be 10 per cent. only. Whipple and Chalmers have shown that the maximum charge which can be collected in this way if ions of both signs are in equal concentrations is $-0\cdot52Fr^2$ when F is the field-strength and r is the radius of the drop. This gives 40 e.s.u./c.c. for the case just considered. If negative ions only are present the maximum charge collected is $-3Fr^2$, six times as great.

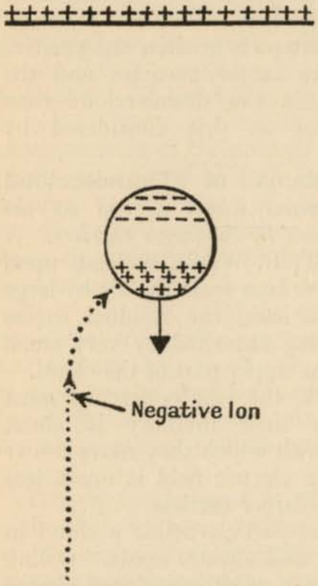

FIG. 20.—Capture of slow negative ions by large falling drops.

The tiny droplets and ice and snow particles, which fall relative to the air with a velocity of the same order as that of the ion-streams, will be unable to exercise the selective action described. They will therefore absorb a net positive charge, the complement of the negative charge collected by the larger carriers.

Wilson's mechanism operates in such a way as to enhance enormously a pre-existing field. This could be the normal field of fine weather or a field produced initially by one of the other mechanisms to be described. The Wilson process has been shown by Gott to work under laboratory conditions but it is not generally agreed that it is the chief agency in building up the fields of thunderclouds. One criticism which has been raised against it is that the concentration of large ions necessary for it to be effective is not likely to be present until the electric field is very strong.

The theory of Frenkel[5] is based upon a selective adsorption of small negative ions by water-drops or wet ice-crystals. Owing to the polar character of the water molecule, the skin of such carriers is considered to consist of an electrical double-layer with positive poles outwards. This tends to capture negative ions from the air until a difference of potential v_0 is created between the drop and the cloud of positive ions which surround it. The equilibrium negative charge acquired in this way is v_0r where r is the radius of the drop and v_0 is 0·25 v. This mechanism gives a considerable negative charge density to small drops but is not very effective in the case of large ones ($r = 0\cdot1$ cm., charge density 0·02 e.s.u./c.c.) nor does it give very large fields. To carry it further, Frenkel considers that adjacent downward-moving droplets coalesce into larger drops as a result of hydrodynamical forces between them. These larger drops in their downward flight sweep up and collect the charge from slower-moving droplets, obtaining in this way a charge which varies as r^3 and creating much larger fields. The theory does not appear to have been tested experimentally.

(*b*) *Mechanisms which depend upon electrification as a result of change of state.* In its earliest stage of development the cumulonimbus cloud consists almost entirely of droplets of water. Some of the droplets in the portion of the cloud above the freezing level remain unchanged in state and supercooled. As the cloud top rises, an increasing proportion of these droplets are converted into ice-particles, either directly or by the action of special "sublimation-nuclei". The effectiveness of these processes increases with decreasing temperature and hence with increasing height. Ice-particles formed in this way grow very rapidly at the expense of surrounding supercooled drops if the temperature is low. As these grow in size they fall and the mature thundercloud contains a complicated mixture of hard and soft hail, snow particles, rime-covered ice (graupel) and water drops. All these hydrometeors are in various stages of growth, or of disappearance, depending upon the temperature-levels to which they have risen or fallen and all of them are constantly colliding and coalescing with one another. The changes of state, and the collision and coalescence processes thus involved, offer a variety of opportunities for the generation of electric charge and a number of possible mechanisms have been based upon them, most of which have been tested experimentally. These are listed in Table VIII below.

Mechanism No. 7 is not one of change of state but is included in the Table because it involves ice; experimental test of it has so far yielded conflicting results. Mechanism No. 6 would not create a dipole of the correct polarity and could only operate in the lower levels of the cloud where like the Simpson breaking-drop mechanism, it may contribute to the formation of the lower *p* charge and the strengthening of the *N* charge above it (Fig. 15). Experiments by Kramer indicate that mechanisms Nos. 2 and 3 of Findeisen must be regarded as of minor significance compared with No. 1.

The remaining mechanisms, Nos. 1, 4 and 5, are all based upon electrification resulting from the freezing of supercooled drops, either directly (5) or by contact with falling ice-pellets (1 and 4). If one or other of them is responsible for the formation of the main cloud dipole, the first occurrence of lightning should be closely related to the first development of the ice-phase in the cloud. This in turn is bound up with the height and so with the temperature-level to which supercooled water-drops are raised by the updraught within the cloud.

CHRIST'S COLLEGE LIBRARY

TABLE VIII

	Proposer.	Charge Carriers.		Generating Process.	Sign of Charge on Heavy Carrier.	Remarks.
		Heavy.	Light.			
1	Findeisen and Kramer	Large rime-coated ice-particles	Small ditto	Coalescence of super-cooled drops with ice to form rime	−	Small rime fragments +vely charged
2	Findeisen	Ice	Ice-splinters	Growth of ice by sublimation from drops	+	Found experimentally to be of minor significance (Kramer)
3	Findeisen	Ice	Ice-splinters	Evaporation of ice	+	As for (2) above
4	Workman and Reynolds	Ice	Fragments of super-cooled drops	Glaze-ice formation on hail	−	Effectiveness increased by traces of $CaH_2(CO_3)_2$ in air
5	Chalmers	Ice	Water-drops	Sudden freezing of super-cooled drops	− ?	Not tested experimentally
6	Dinger and Ross Gunn	Water	?	Melting of ice at low levels	+	Presence of dissolved gases necessary before freezing if effect is to be appreciable.
7	Simpson and Scrase	Ice	?	Frictional electrification by collision	−	Inferred from Antarctic blizzard observations

Evidence as to the height of the larger droplets within the cloud at various stages has recently been made available by microwave radar studies of the so-called " cloud echo ". These suggest that full-scale electrification leading to fields sufficiently great to produce lightning does not occur until the supercooled droplets have been raised to a temperature level, variously given as between − 28° C. and − 40° C., where ice-crystal formation on an appreciable scale can occur.

These observations, together with the conclusions of Malan and the author (§ 37 (4)) that the *N* charge occupies a column of considerable vertical extent, whose top reaches to very low temperature-levels, make it probable that the main dipole electrification of the thundercloud is initiated as a consequence of freezing processes of one kind or another as soon as the ice-phase forms from super-cooled drops. The larger carriers of charge will then fall downwards and build up a field which may of itself be sufficient to cause spark break-down or may have to be enhanced by the selective charging mechanism of C. T. R. Wilson. Another method of enhancement, by droplets bouncing off melting hail in the lower portion of the cloud, has been discussed by Workman and Reynolds.[6]

45. The Charge on Rain from Electrified Clouds.

Although the predominant sign of the electric field below a thundercloud is negative (§ 33) it has long been known that the charge on the rain from it is mainly positive in sign. This can be qualitatively explained by C. T. R. Wilson's selective ion-capture mechanism (§ 44 (*a*)), which would operate as in Fig. 20 with all the signs in the diagram reversed and might be expected to be very effective since point-discharge at the ground creates an abundance of positive ions most of which will be converted by capture-nuclei of various kinds into heavy ions during their passage upwards to the cloud.

Simpson,[6] in an important series of observations made at Kew, has compared the current, *i*, carried down by rain to a square centimetre of the ground with simultaneous records of the rate of rainfall, *R*, and of the point-discharge current, *I*, from an elevated point. For potential gradients exceeding 2000 v./m. he finds the relation $i \propto -I \times R^{0\cdot 57}$ from which it is clear that (i) the sign of the charge on the rain in these fields is always opposite to that of the prevailing field (mirror-image effect); (ii) with a given rate of rainfall the rain-current is directly proportional to the point-discharge current, and (iii) with a constant point-discharge current the rain-current varies with the rate of rainfall, increasing or decreasing approximately as $\sqrt{R}$.

It is clear from these results that the rain derives its charge from point-discharge ions. Ultimately, in very heavy rain, all the point-discharge ions are swept up and brought back again to earth, the rain current being found equal to the estimated point-discharge current per square centimetre of the ground. Whether the Wilson mechanism alone can account quantitatively for these results is still under discussion; it is not impossible that more than one process is at work in charging the falling drops. The rain current by itself gives a statistical mean for the charge on the drops and there is evidence that individual drops can carry charges of opposite sign.

In light steady rain and in some showers, the electric field at the ground is found by Simpson to be generally negative and as high as − 1000 v./m. When it is positive it seldom exceeds the fine weather gradient of about + 400 v./m. Such rain is found to be positively charged, the charge on it being directly proportional to the departure of the field from its fine weather value. No explanation of this effect has as yet been suggested.

REFERENCES TO PAPERS

[1] Ref. 10, Chap. III.

[2] Ref. 16*b*, Chap. II.

[3] J. A. Chalmers, *Quart. Journ. Roy. Met. Soc.*, **73**, p. 324 (1947).

[4] F. J. Whipple and J. A. Chalmers, *ibid.*, **70**, p. 103 (1944).

[5] Ref. 16, Chap. II.

[6] E. J. Workman and S. E. Reynolds, *Phys. Rev.*, **78**, p. 254 (1950).

[7] Sir G. Simpson, *Met. Office Geophys. Mem.*, **84** (1949).

LIBRARY

GENERAL REFERENCES

Victor F. Hess, *The Electrical Conductivity of the Atmosphere and its Causes*, English translation by Codd, gives a detailed account of the matter treated in Chapter I, up to the year 1927.

The Dictionary of Applied Physics, vol. 3, contains a valuable article on "Atmospheric Electricity," by C. T. R. Wilson.

Encyclopædia Britannica, 1910, art. "Atmospheric Electricity," by C. Chree.

K. Kähler, *Einführung in die Atmosphärische Elektrizität*, gives details of instruments and the technique of measurement.

E. von Schweidler, *Einführung in die Geophysik*, II, art. "Luftelektrizität."

K. Kähler, *Die Elektrizität der Gewitter*.

A. Gockel, *Die Luftelektrizität*.

B. Chauveau, *Électricité Atmosphérique*.

E. Mathias and collaborators, *Traite d'Électricité Atmosphérique et Tellurique*.

J. A. Chalmers, *Atmospheric Electricity*, Clarendon Press, 1949.

Terrestrial Magnetism and Atmospheric Electricity, ed. J. A. Fleming, Dover Publications, New York (1949).

T. W. Wormell, *Atmospheric Electricity*, Quart. J. Roy. Met. Soc., 79, 11, 1953.

For general references on lightning, see Chapter II, refs. 1, 9, 10.

SUBJECT INDEX

Air-earth current, *see* Fine-weather current
Air-radiation, 8
Aitken counter, 4; nuclei, *see* Nuclei
Alti-electrograph, 54
Amplifier-oscillograph method, 52
Atmospherics, 77

Boys' camera, 66

Capillary electrometer, 47
Charge on the earth, 19; maintenance of, 41
Collectors, 22
Condensation nuclei, *see* Nuclei
Conduction current, *see* Fine-weather current
Conductivity of the atmosphere–
  effect of pollution, 26, 33; of large ions, 6
  measurement, 6
  polar, 5
  variation with height, 34; with time and season, 33
Convection current, 29
Cosmic radiation, 7, 10

Dissipation of charge, 6
Drops, deformation in electric fields, 62
  electrification by breaking, 80; by change of state, 85; by Wilson process, 83
Dust-counter, 4
Dust-storm fields, 36

Earth-radiation, 8
Ebert ion-counter, 3
Electric fluxmeter, 24, 54
Electrode space-charge, 31
Electrometer: Benndorf, 24; capillary, 47; universal portable, 20
Emanation in air, 8
Equilibrium of atmospheric ionisation, 13, 17

Fast ions, *see* Ions, small
Field-mill, 24, 54
Fields in disturbed weather, 36; from thunderclouds, 52; in fog, 33, 36
Fine-weather current, 28 *et seq.*
  measurement, 29
  significance, 34
  variation with height, time and season, 33
Fine-weather field, 19 *et seq.*
  continuous recording, 23
  measurement, 21
  origin, 42
  variation with height, time and season, 24, 25
Fulchronograph, 78

Gerdien conductivity apparatus, 7

Intermediate ions, 2
Ionisation balance, 17
Ionisation, in closed vessel, 11; over oceans, 16, 18; residual, 11

Ionising power of radiations, 7, 14
Ion-counter, 3
Ions, atmospheric, 1 *et seq.*
Ions, intermediate, 2
Ions, large, 2
  charge on, 2
  effect on conductivity, 5, 33
  formation, 2, 15
  mobility, 2
  number per cc., 2
Ions, small, 2
  average life, 15, 17
  capture of, 3, 15
  charge, 2
  effect on conductivity, 5, 33
  mobility, 1, 4
  number per cc., 2
Isoceraunic level, 46

Kelvin water-dropper, 22
Klydonograph, 77

Lightning, Chapter IV
  Air-discharges, 65
  Charge carried, 61, 78
  Cloud-discharges, 65
  Currents, 77
  Effect on earth's charge, 42
  Electric moments, 61
  Empire State Building, 71
  Field changes; electrostatic, 72; induction and radiation, 76
  Frequency of flashes, 46
  Junction streamers, 73
  Leader, 67 *et seq.*; dart, 69; mechanisms, 74; pilot, 75; stepped, 67, 75
  M components, 71
  Photography, 66
  Polarity, 58
  Quantities in discharge, 61, 78
  Recoil streamers, 76
  Recovery of field after, 50, 61, 82
  Reversal distance, 57, 58
  Ribbon, 66
  Streamers, 66, 73
  Streamer velocities, 71
  Successive strokes, 65, 74

Magnetic links, 78

Normal ions, *see* Ions, small
Nuclei, Aitken or condensation, 2
Nuclei; counting, 4; effect on life of small ions, 15; effect on field, 27; and smoke pollution, 27; sublimation, 86

Point discharge currents, 38, 55; and earth's charge, 41; and thunderclouds, 62
Polar conductivity, *see* Conductivity
Potassium, gamma rays, 8
Potential gradient, *see* Field, *passim*
Precipitation current, 38, 41

Radar cloud-echoes, 88
Radioactivity of air and earth, 8, 33
Rain, charge on, 36, 88
Residual ionisation, 11

St. Elmo's fire, *see* Point discharge
Slow ions, *see* Ions, large
Snow; charge on, 36; field during, 36
Space-charge, 25, 31
Static, 77
Sublimation nuclei, 86

Thunderclouds, Chapters III and V
  Cells, 55
  Charge on precipitation, 88
  Distribution of charge, 55

  Electric fields; outside, 52; inside, 42, 63
  Energy involved, 63
  Field-changes, 57
  Generation mechanisms, 80 *et seq.*
  Number, 46
  Polarity, 55
  Potentials in, 63
  Recovery of charge, 50, 61, 82
  Separation of charges, 81

Universal electrometer, 20
Upper atmosphere; potential, 25; currents to, 43

Voltameter, water, 40

LIBRARY

INDEX OF NAMES

Aitken, 1, 4
Appleton, 52, 64

Baldit, 37
Beck, 79
Benndorf, 24
Boylan, 18
Boys, 66
Brookes, 46, 64
Bruce, 79

Carnegie, s.s., 18, 25, 27
Chalmers, 83, 84, 87, 89, 90
Chauveau, 90
Chree, 90
Collens, 67, 70, 79
Craib, 64

Dalibard, 38
De Sachy, 18
Dinger, 87

Ebert, 3
Elder, 79
Elster, 1
Explorer II, 34, 35

Findeisen, 87
Fleming, 90
Franklin, 38
Frenkel, 44, 45, 82, 85

Geitel, 1
Gerdien, 6
Gish, 43, 45, 60
Gockel, 90
Golde, 79
Gott, 85
Gschwend, 37
Gunn, 54, 63, 64, 87

Hagenguth, 71
Halliday, 57, 58
Herd, 52, 64
Hess, 10, 16, 18, 90
Hodges, 79
Hogg, 18, 33, 45

Jensen, 57

Kähler, 90
Kelvin, 23
Kolhörster, 10
Kramer, 86, 87

Langevin, 2
Lemonnier, 19, 38
Lenard, 80
Loeb, 79

Macky, 62
Maclelland, 37
Malan, 45, 55, 67, 70, 74, 78, 88
Marwick, 37
Mathias, E., 90
Mathias, O., 16, 18
Mauchly, 27, 45
McCann, 78, 79
McEachron, 71, 78, 79
Meek, 79

Nolan, 4, 5, 6, 14, 18, 37

O'Brolchain, 18

Perry, 79

Reynolds, 87, 88, 89
Robinson, 59, 64
Roll, 18
Russelvedt, 24

Schindelhauer, 37
Schonland, 39, 45, 55, 57, 58, 64, 67, 74, 78, 88
Scrase, 45, 54, 55, 59, 64, 80, 87
Simpson, 27, 37, 54, 59, 64, 75, 80, 87, 88, 89
Swann, 4

Vancour, 18
Von Schweidler, 15, 17, 18, 90

Wagner, 78
Wait, 43, 45, 60
Watson, 45
Watson-Watt, 52, 64
Whipple, 26, 32, 44, 45, 55, 64, 84, 89
Wigand, 34, 41, 45
Wilson, 18, 20, 38, 42, 45, 47, 55, 58, 62, 64, 81, 83, 90
Workman, 87, 88, 89
Wormell, 39, 43, 45, 58, 59, 64

COLLEGE
LIBRARY

PRINTED IN GREAT BRITAIN AT
THE UNIVERSITY PRESS
ABERDEEN

RIST'S COLLEGE LIB
WOOLT

FLUORESCENCE AND PHOSPHORESCENCE. By E. HIRSCHLAFF

AN INTRODUCTION TO VECTOR ANALYSIS. By B. HAGUE

THE CYCLOTRON. By W. B. MANN

THE SPECIAL THEORY OF RELATIVITY. By HERBERT DINGLE

WAVE GUIDES. By R. L. LAMONT

WAVE FILTERS. By L. C. JACKSON

HIGH FREQUENCY TRANSMISSION LINES. By WILLIS JACKSON

HEAVISIDE'S ELECTRIC CIRCUIT THEORY. By H. J. JOSEPHS

PHOTOELECTRIC TUBES. By A. SOMMER

MAGNETIC MATERIALS. By F. BRAILSFORD

X-RAY OPTICS. By A. J. C. WILSON

FUNDAMENTALS OF DISCHARGE TUBE CIRCUITS. By V. J. FRANCIS

AN INTRODUCTION TO THE LAPLACE TRANSFORMATION. By J. C. JAEGER

ATMOSPHERIC TURBULENCE. By O. G. SUTTON

PHOTONS AND ELECTRONS. By K. H. SPRING

ELECTROMAGNETIC PROBLEMS OF MICROWAVE THEORY. By H. MOTZ

AN INTRODUCTION TO SERVO-MECHANISMS. By A. PORTER

INTEGRAL TRANSFORMS IN MATHEMATICAL PHYSICS. By C. J. TRANTER

SEMI-CONDUCTORS. By D. A. WRIGHT

AN INTRODUCTION TO ELECTRON OPTICS. By L. JACOB

THE MEASUREMENT OF RADIO ISOTOPES. By DENIS TAYLOR

PHYSICAL FORMULAE. By T. S. E. THOMAS

ATOMIC SPECTRA. By R. C. JOHNSON

WAVE MECHANICS. By H. T. FLINT

THE PHYSICAL PRINCIPLES OF WIRELESS. By J. A. RATCLIFFE

THE CONDUCTION OF ELECTRICITY THROUGH GASES. By K. G. EMELÉUS

X-RAY CRYSTALLOGRAPHY. By R. W. JAMES

THE COMMUTATOR MOTOR. By F. J. TEAGO

APPLICATIONS OF INTERFEROMETRY. By W. EWART WILLIAMS

THERMIONIC VACUUM TUBES. By W. H. ALDOUS and Sir EDWARD APPLETON

PHOTOCHEMISTRY. By D. W. G. STYLE

THERMODYNAMICS. By A. W. PORTER

ATMOSPHERIC ELECTRICITY. By B. F. J. SCHONLAND

THE METHOD OF DIMENSIONS. By A. W. PORTER

COLLISION PROCESSES IN GASES. By F. L. ARNOT

PHYSICAL CONSTANTS. Selected by W. H. J. CHILDS

ELECTROMAGNETIC WAVES. By F. W. G. WHITE

THE GENERAL PRINCIPLES OF QUANTUM THEORY. By G. TEMPLE

THE KINETIC THEORY OF GASES: Some Modern Aspects. By Professor MARTIN KNUDSEN

LOW TEMPERATURE PHYSICS. By L. C. JACKSON

HIGH VOLTAGE PHYSICS. By L. JACOB

RELATIVITY PHYSICS. By W. H. MCCREA

X-RAYS. By B. L. WORSNOP and F. C. CHALKLIN

HYPERFINE STRUCTURE IN LINE SPECTRA AND NUCLEAR SPIN. By S. TOLANSKY

INFRA-RED AND RAMAN SPECTRA. By G. B. B. M. SUTHERLAND

THERMIONIC EMISSION. By T. J. JONES

ELECTRON DIFFRACTION. By R. BEECHING

MAGNETISM. By EDMUND C. STONER

THE EARTH'S MAGNETISM. By S. CHAPMAN

MERCURY ARCS. By F. J. TEAGO and J. F. GILL

MOLECULAR BEAMS. By R. G. J. FRASER

ALTERNATING CURRENT MEASUREMENTS. By DAVID OWEN

(Continued on back flap)

ATMOSPHERIC ELECTRICITY—B. F. J. SCHONL

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/258381209

Electrodynamical Coupling of Earth's Atmosphere and Ionosphere: An Overview

Article *in* International Journal of Geophysics · December 2011
DOI: 10.1155/2011/971302

CITATIONS
17

READS
254

4 authors:

Prof. Ashok K. Singh
University of Lucknow
112 PUBLICATIONS **840** CITATIONS
SEE PROFILE

Devendraa Siingh
Indian Institute of Tropical Meteorology
149 PUBLICATIONS **1,211** CITATIONS
SEE PROFILE

R. P. Singh
Inter University Accelerator Centre
807 PUBLICATIONS **4,831** CITATIONS
SEE PROFILE

Sandhya Mishra
7 PUBLICATIONS **20** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project The study of equatorial plasma bubble during January to April 2012 over Kolhapur (India) View project

Project Study of Physical, Optical and Chemical properties of Atmospheric Aerosols over Indo-Gangetic Basin View project

All content following this page was uploaded by Devendraa Siingh on 24 November 2014.

The user has requested enhancement of the downloaded file.

Hindawi Publishing Corporation
International Journal of Geophysics
Volume 2011, Article ID 971302, 13 pages
doi:10.1155/2011/971302

Review Article

Electrodynamical Coupling of Earth's Atmosphere and Ionosphere: An Overview

A. K. Singh,[1] Devendraa Siingh,[2] R. P. Singh,[3] and Sandhya Mishra[1]

[1] *Department of Physics, University of Lucknow, Lucknow 226007, India*
[2] *Indian Institute of Tropical Meteorology, Pune 411-008, India*
[3] *Physics Department, Banaras Hindu University, Varanasi 221005, India*

Correspondence should be addressed to A. K. Singh, aksphys@gmail.com

Received 24 March 2011; Revised 31 July 2011; Accepted 9 September 2011

Academic Editor: Lucilla Alfonsi

Copyright © 2011 A. K. Singh et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Electrical processes occurring in the atmosphere couple the atmosphere and ionosphere, because both DC and AC effects operate at the speed of light. The electrostatic and electromagnetic field changes in global electric circuit arise from thunderstorm, lightning discharges, and optical emissions in the mesosphere. The precipitation of magnetospheric electrons affects higher latitudes. The radioactive elements emitted during the earthquakes affect electron density and conductivity in the lower atmosphere. In the present paper, we have briefly reviewed our present understanding of how these events play a key role in energy transfer from the lower atmosphere to the ionosphere, which ultimately results in the Earth's atmosphere-ionosphere coupling.

1. Introduction

The atmosphere of the Earth is a layer of gases surrounding the Earth that is retained by Earth's gravity. The atmosphere protects life on the Earth by absorbing the ultraviolet solar radiation, warming the surface through heat retention (greenhouse effect), and reducing temperature extremes between day and night. Based on temperature distribution, atmosphere is divided into the troposphere, stratosphere, mesosphere, and thermosphere. The temperature in the thermosphere remains almost constant (Figure 1) [1]. The stratosphere and mesosphere regions are also grouped as the middle atmosphere. The region above the middle atmosphere is called the upper atmosphere where solar radiation and other sources ionize the neutral constituents forming plasma of ions and electrons. The region extending from the mesosphere to the thermosphere is called ionosphere where plasma dynamics is controlled by the collisions between the ionized particles and neutrals as well as between the ionized particles themselves. The region above the ionosphere is known as the magnetosphere. In this region, charged particles dynamics is controlled by the Earth's magnetic field because the density collision frequency is very low. There is no sharp boundary between the upper ionosphere and the lower magnetosphere region.

The ionosphere system is mainly controlled by various external sources of forcing and number of mechanisms operative in the system to convert, transport, and redistribute the input energy. Solar extreme ultraviolet (EUV) radiation and particle energy from the sun in the form of precipitating solar wind plasma energetic particle influence from the above, while tides, planetary waves, gravity waves, electromagnetic waves in wide frequency range, turbulence, convection, and so forth from the below. Even processes taking place below/on/above the surface of the Earth also affect the ionosphere and its processes. In fact, lower atmosphere/middle atmosphere/upper atmosphere (ionosphere)/magnetosphere acts as a multi-coupled system. The coupling occurs mainly through the dynamical, chemical, and electrical processes.

The ionosphere reacts to various phenomena such as lightening discharges, functioning of high-power transmitters, high-power explosion, earthquakes, volcano eruptions, and typhoons through a chain of interconnected processes in the lithosphere-atmosphere-ionosphere interaction system. Thunderstorms play a major role in transferring energy from the atmosphere to the ionosphere [3] and in establishing

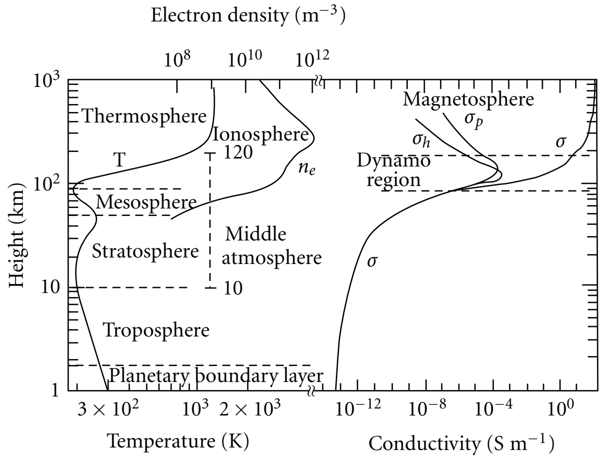

Figure 1: Variation of temperature, electron density and electrical conductivity of various atmospheric layers, after [1].

electrical coupling between atmosphere and ionosphere through the global electric circuit (GEC). The Earth's surface has a net negative charge, and there is an equal and opposite positive charge distributed throughout the atmosphere above the surface. The electrical structures of the lower atmosphere, GEC, and conductivity profile shown in Figure 1 are deeply influenced by cosmic ray flux [4], high-power transmitted waves [5–7], and earthquakes [8]. Lightening-generated whistler mode waves scatter radiation belt trapped electrons which precipitate into the lower ionosphere and result into the additional ionization and formation of ionospheric inhomogeneities [9, 10]. The powerful high-frequency transmitted waves produce ionospheric heating which, in turn, causes generation of ultra-low-frequency (ULF) and extremely low-frequency (ELF) waves [11], the formation of very low-frequency (VLF) ducts and other types of inhomogeneities [12, 13], the acceleration of ions, and the excitation of atmospheric emissions in different spectral bands [14]. The effects of these processes on the GEC (including the ionosphere) are not yet fully understood.

The convective activity in thunderstorms produces gravity waves which propagate to the lower ionosphere and deposit energy and momentum through the process of breaking and absorption and may help in the initiation of sprites and other transient luminous events (TLEs) [15–17]. The converse problem is the generation of acoustic gravity waves by sprites and other TLEs. The thermal energy deposited in the neutral atmosphere within a sprite volume ($\sim 10^4$ km^3) during short (<1 ms) duration may produce an impulsive pressure pulse that propagates laterally outward as an acoustic/gravity wave [18]. The energy deposited by a large sprite could be as large as ~1 GJ [19]. The acoustic gravity wave causes spatial and temporal modulation of plasma density and electric conductivity in the ionospheric E layer. The associated wave instability causes generation of related field-aligned currents and plasma density irregularities in the upper ionosphere [20]. The satellite observations anomalous DC electric field, ULF magnetic pulsations, small scale plasma inhomogeneities, and correlated ELF emissions [21] may be considered as experimental evidence for these processes. However, the full knowledge of the interconnection between gravity waves and TLEs remains unresolved.

The purpose of this paper is to review briefly the present understanding of the link between most of the processes operative in the lower atmosphere and their electrodynamic coupling with the ionosphere. After this introduction, Section 2 deals with a more detailed description of physics of the atmosphere-ionosphere coupling. Various coupling processes including AC and DC phenomena are discussed in Section 3. Brief summary is given in Section 4 with some discussions.

2. Physics of Earth's Atmosphere-Ionosphere (AI) Coupling

Thunderstorms and lightning discharges are the main sources of electromagnetic and electrostatic energy in the Earth's atmosphere. Electrostatics involves the buildup of charge on the surface of objects due to contact with other surfaces. Although charge exchange happens whenever any two charged surfaces contact and separate, the effects of charge exchange are usually only noticed when at least one of the surfaces has high resistance to electrical flow. In thunderstorms, a rapid vertical rearrangement of deep air layers takes place. Large processes promote, shape vertical and horizontal air motions, and processes within storms control the development of rain and strong local winds. Vigorous vertical air currents and thunderstorms are a consequence of excessive warmth and moisture at low altitudes.

Thunderstorm consists of a cluster of clouds. Electrical charges generate and separate during the developmental stage of thunderstorm and are neutralized during lightning discharges. Numerous processes operating within the environment of a mature convective cloud with varying effectiveness and time dependencies affect cloud electrification [22, 23]. The charging of thunderstorms can be discussed as inductive charging or noninductive charging. An inductive process requires preexisting electric fields to induce charges on a particle so that when it rebounds from another particle, charge is separated and the field is enhanced. In the atmosphere, the fair-weather electric field resulting from positive charges in the atmosphere and negative charges on the ground could be considered as the preexisting field. However, there are experimental results from airborne instruments which require some other processes of charging [22]. Noninductive processes are independent of the presence of an external electric field. In an ordinary thundercloud, the smaller ice crystals are charged positively and move upward, whereas larger graupel particles charged negatively descend relative to the smaller particles under the action of gravity. This is the normal situation, depending on the prevailing conditions of temperature, liquid water content, and mixing in the thunderstorm. Situation in thunderstorm is quite variant, and charging process may not remain as discussed above [24]. To better understand these phenomena, further investigations

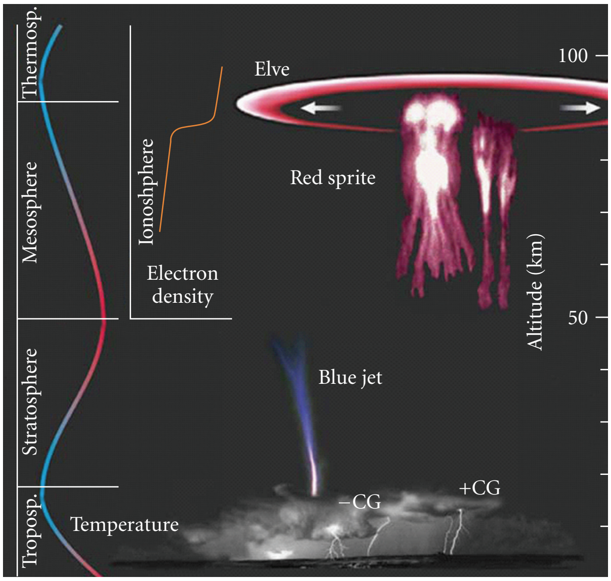

Figure 2: Depiction of various optical events in the atmosphere and the altitude at which they occur, credit [2].

involving both precise measurements of time development of electric field inside the cloud and simulations are required.

The charge buildup during the active period of a thunderstorm ultimately results into lightning discharge either through the processes of dielectric breakdown or runaway breakdown [29–32]. The electric field computation favours runaway breakdown mechanism [33]. However, discharge processes are not fully understood, and further observations and works are required. Lightning discharge creates charge imbalance in the cloud/thunderstorm, and as a result upward propagating quasi-electrostatic field initiates from the upper charge-layer of thunderstorm. This quasi-electrostatic field connects troposphere with the stratosphere and the lower ionosphere for a very short duration. There are about 40 to 100 lightning discharges occurring every second. Apart from the upward propagating quasi-electrostatic fields, lightning discharge current generates intense electromagnetic pulses of ~20 GW peak powers [34] which may cause significant ionospheric disturbances because of the heating and acceleration of electrons, production of ionization, optical emissions, and so forth.

The upward propagating quasi-electrostatic causes transient luminous events (sprites, blue jets/gigantic jets, elves, etc.) in the stratosphere and the mesosphere which influence the lower ionosphere with additional ionization and formation of ionospheric inhomogeneities and has motivated scientists all over the world to reexamine our understanding of the electrical processes and properties of the atmosphere. Figure 2 shows the variety of transient luminous events and the typical altitude at which they occur. Sprites may occur in cluster of short-lived (~50 ms) two, three, or more carrot-shaped emissions of ~1 km thickness over a horizontal distance of 50–100 km, with the separation between sprite elements of ~10 km [2, 35]. The upper portion of the sprites is red, with wispy, faint blue tendrils extending to 40 km or at time as low as the cloud top and bright region is in the altitude range 65–85 km [35]. They typically last for 5–50 ms and may take the form of one or more vertical columns of a few hundred meter radius for the smaller column sprites or large jellyfish-shaped structures of tens of kilometers of radius and extending from the ionosphere D-region almost down to the thunderstorm cloud tops [2, 35]. The knowledge gained from laboratory experiment of gas discharges at subatmospheric pressure has been used to understand sprite spectroscopy and associated phenomena [25]. Figure 3 shows broad identical spectral characteristics of light of positive column of the laboratory tube and sprites [25]. The difference in spectral characteristics may be due to difference in applied electric field, gas pressure, and gas composition in the mesosphere and gas tube discharges. Physical processes associated with sprites and other optical events are also associated with thunderstorm activity in the troposphere and are thought to result in the gradual build-up of conductivity changes in the lower ionosphere [36]. Liszka [37] suggested the generation of infrasound waves by sprites, whose signatures were detected by a network sensors in Sweden [38]. The shape of the chirp signature in the spectrograms of infrasound can be explained by the horizontal size of the sprite [39]. Neubert et al. [40] have concluded that sprite detection by infrared is an attractive alternative to optical detection, because it is not limited by clear viewing condition or by the absence of daylight.

Elves are concentric rings of optical emissions propagating horizontally outwards at the bottom edge of ionosphere between 75–95 km altitudes caused by the electromagnetic pulse radiated by the cloud to ground lightning discharge current of either polarity. Elves produce average electron density enhancements of 210–460 electrons cm^{-3} over large (165–220 km diameter) circular regions having an assumed 10 km altitude extent [41, 42] and hence produce perturbation in electrical conductivity of the mesosphere which can be evaluated by measuring changes in the amplitude of VLF waves [43, 44] from transmitters and propagating in the Earth-ionosphere waveguide. Mika [45] has discussed experimental data in which the incident VLF transmitter signal seems to be scattered from horizontally extended diffuse regions of electron density enhancements, most likely associated with halos or diffuse regions of the upper part of carrot sprites, rather than small-scale streamers observed at lower altitudes. The electromagnetic pulse (EMP) generating elves also create ionization [46, 47], which depends on its intensity. The individual contributions of different components of constituent gas to the optical emissions of sprite/elves are not known. An attempt should be made to investigate lightning chemistry and energetics in the region of sprites/elves. This may require future studies using a large number of narrow band filters in optical observations and kinetic modeling in theoretical analyses.

The transient luminous events provide a link between tropospheric processes in the thunderstorms and mesospheric processes in the upper atmosphere, and their studies also promise to improve our understanding of the elusive mesosphere, perhaps the least understood layer of Earth's atmosphere. Hiraki et al. [48] suggested that sprites may

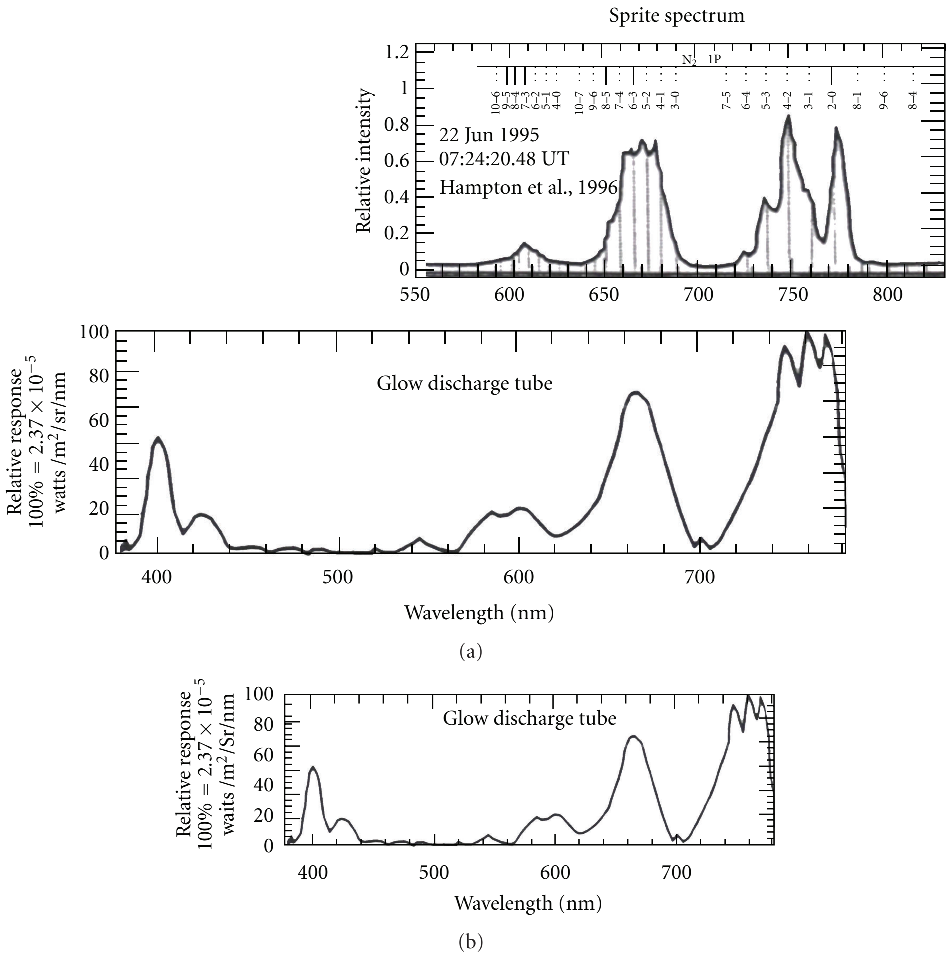

FIGURE 3: Comparison of optical spectra for (a) sprite in the mesosphere and for (b) the positive column of large glow discharge tube, after [25].

change the concentration of NOx and HOx in the mesosphere and lower atmosphere. These chemical changes have impact on the global cooling or heating in the middle atmosphere [49–51]. Nitrogen oxides are critical components of the troposphere which directly affect the abundance of ozone and hydroxyl radicals [52]. Ozone absorbs solar ultraviolet radiation and controls the dynamic balance of the atmosphere. NOx creates ozone in the troposphere and destroys it in the stratosphere and mesosphere and thus affects the climate. This area of investigation remains unexplored.

Lightning generated whistler mode waves propagate along geomagnetic field lines without appreciable attenuation and thus connects troposphere with the ionosphere and the magnetosphere. Whistler mode waves while propagating through the equatorial region of the magnetosphere interact with the Earth's radiation belt trapped electrons and cause their precipitation in the form of plasma inhomogeneities in the high-latitude lower ionosphere. Such a type of coupling has been confirmed in ground-based observations [53], balloon measurements [54], rocket experiments [55, 56], and satellite observations [57, 58].

Thunderstorms and lightning discharges form the major current source in GEC, essential features of which are shown in Figure 4 [26]. Solar wind interactions with the Earth's magnetic generate additional current [59], and potential gradient modulation may arise from coupling of geomagnetically-induced changes in the magnetospheric dynamo through the global circuit. Precipitations from electrified clouds are also current driver [60]. The total current flowing in the atmospheric GEC is ~1-2 kA [27]. A realistic model of equivalent circuit with capacitors, resistors, and switches is shown in Figure 5 [28]. A switch is closed for a short time if a certain type of discharge occurs. For example, switch S_2 (S_3) closes for ~1 ms when positive (negative) cloud to ground discharge occurs for a particular storm. The switch S_1 closes for a few ms when a sprite occurs above a particular thunderstorm and fair-weather time constant, $r\ C \sim 2$ min ($r \sim 200$ ohms, $C \sim 0.7$ F). Generators act over <1% of Earth's surface, and remaining ~99% of the Earth's surface region acts as a load on the circuit. The total current (I) flowing in the circuit is ~1 kA [28]. The GEC is based on the concept that the quantity of electric charge has to be conserved in the Earth's atmosphere-ionosphere-magnetosphere

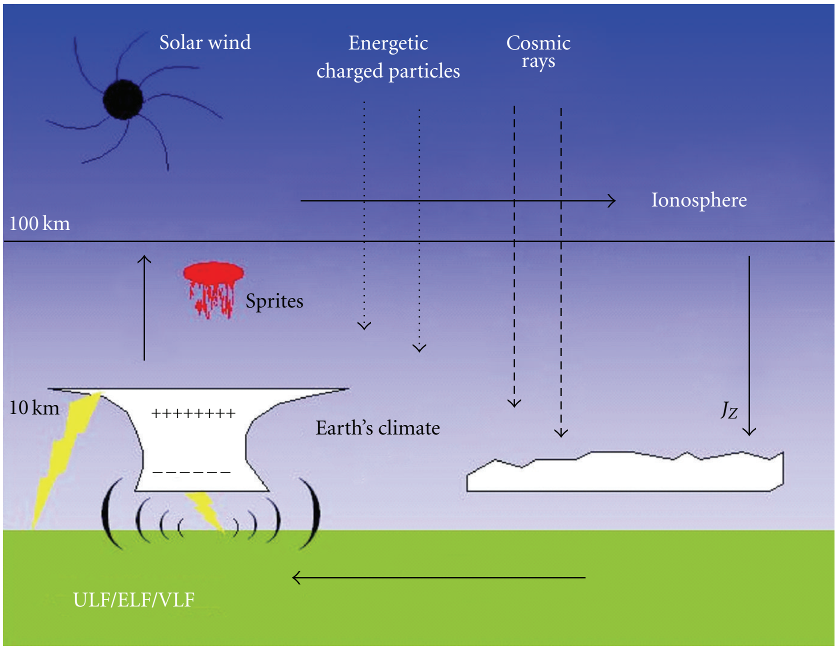

Figure 4: Essential features of the global atmospheric electric circuit [26, 27].

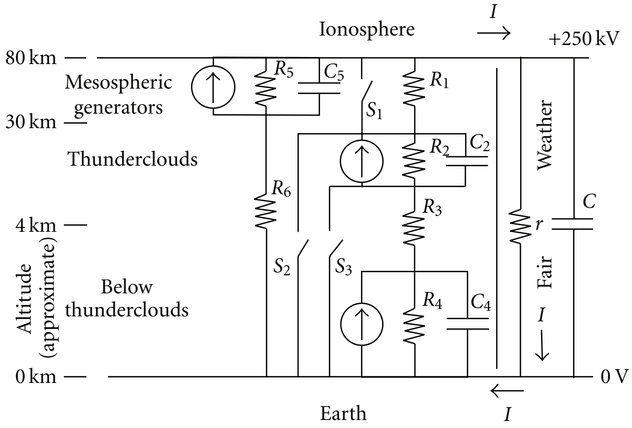

Figure 5: Diagram showing a schematic equivalent circuit for global electric circuits, credit [28].

system, that electric currents redistribute charge and that the electric currents are continuous.

The power supplied by thunderstorms is insufficient to maintain a field of magnitude observed in fair-weather regions. Rycroft et al. [61] included the generator associated with electrified clouds in the GEC model; this was found to be of the same magnitude as that due to thunderstorms. The optical phenomena occurring in the upward branch of the GEC above the thunderstorms are likely to influence only the upper atmosphere conductivity. Since they occur much less frequently (only one sprite out of 200 lightning) because of their association with intense lightning discharges [50, 62], their contribution to the ionospheric potential is very small [61]. The gigantic jets transport large quantities of negative charge discharging the atmospheric capacitor [63–66] whose effects on the ionosphere and GEC have not yet been modeled. The role of sprite/TLE events on the flow, charging/discharging of GEC, modification of electric fields near the Earth's surface remains unanswered. Since optical emissions could change electrical properties of the atmosphere and influence processes related with weather and climate, intense research activity in this area is required.

The earthquakes affect the electrodynamics of the atmosphere through the generation of electric and magnetic fields with crustal deformation, fault-failure-related piezomagnetism, stress/conductivity, electrokinetic effects, charge generation processes, thermal remagnetization, and demagnetization effects, and so forth [67]. These processes in the Earth's lithosphere relate with disturbances in the atmosphere and ionosphere. Sorokin et al. [8] discussed the processes forming the electrodynamic model of the effect of seismic and meteorological phenomena on the ionosphere. Radioactive substances and charged aerosols injected into the atmosphere modify the altitude profile of conductivity, generation of external currents, perturbation of electric field, and current in the ionospheric layer. As a result, Joule heating of the ionosphere and instability of acoustic gravity waves take place, which manifests in the formation of horizontal inhomogeneities of ionospheric conductivity. Finally, excitation of plasma density fluctuations and ULF/ELF emissions in the ionosphere, generation of field aligned currents and plasma layers, upward plasma transport, and modification of F_2-layer, and change in the ion composition of the upper ionosphere take place [68, 69]. These changes may also affect the GEC and the Earth's climate which remains a challenging problem to be solved. Figure 6 shows a schematic diagram which can be used to

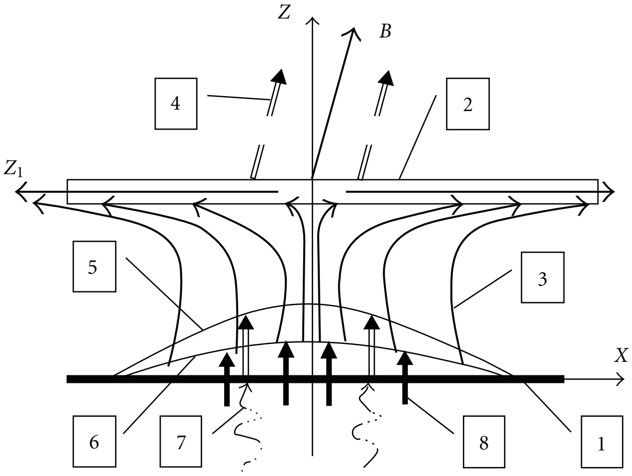

Figure 6: Schematic model used for the calculation of electric field due to the injection of charged aerosols in the atmosphere-ionosphere circuit. 1—Earth's surface, 2—ionosphere, 3—conducting current in the atmosphere and the ionosphere, 4—field-aligned current, 5—zone of upward convection of charged aerosols and external electric current formation, 6—zone of perturbation of atmospheric conductivity induced by radioactive elements emanation, 7—charged aerosols moving upward with soil gases, 8—radioactive elements emanation, credit [3].

compute external currents and enhancement of electrical conductivity in the lower atmosphere that causes an increase in the electric field in the ionosphere due to the injection of charged aerosols and radioactive elements from the solid Earth during earthquakes [3].

There are some other complex processes showing multiple scales of variability. Turbulent mixing and eddy diffusion are few other ways of transport from below. They control the relative abundances of atmospheric species in the mesosphere whose distribution tends to be independent of altitude. As part of a physical mechanism, the influence of internal atmospheric waves may be considered. The upward propagation of internal atmospheric waves (planetary waves, tides, and gravity waves) from the troposphere and stratosphere is an essential source of energy and momentum for the thermosphere and ionosphere. Weaker mesospheric gravity waves will lead to stronger winds in the upper atmosphere, which in turn will affect the ability of planetary waves to propagate, producing an additional feedback route [70]. Of course, the study of internal waves is the province of meteorology, a discipline that has enjoyed a long and independent development of its own and has its own complicated problems, sufficiently different from ionospheric physics that the two are regarded as separate but neighbouring disciplines. However, the internal waves launched by weather fronts or any other sources in the troposphere and stratosphere sometimes appear to be capable of penetrating into the ionosphere, where they dissipate most of their energy. The leakage of wave energy from the troposphere and stratosphere at least up to 100–115 km was introduced as coupling from below [71] and is considered as a mechanism of the meteorological influence on the ionosphere.

3. Various Coupling Processes

Thunderstorms directly couple the atmosphere and the ionosphere. Potential difference between the ionosphere and the Earth is maintained by thunderstorms' pumping action of lightning discharges and electrified clouds. In the Earth-ionosphere cavity, the electric field and the conduction current in the lower atmosphere are primarily controlled by ions. Ions have the characteristic parameters such as mobility, lifetime, and generation rate that vary with altitude. In recent years, some progress has been made towards understanding thunderstorm and lightning [22, 72, 73]. The essential factors in marinating the lightning discharge is the gas breakdown caused by the stored electrostatic charge in the leader head. This extends the channel into new regions. General models of the cloud charge distribution are based on electric field measurements inside a thundercloud and on the ground below the thunderstorm [75]. Earlier, an oversimplified tripolar charge structure consisting of a negative charge in the middle of the cloud, a positive charge above it, and a smaller positive charge below were widely used. Later on, a screening negative charge on the upper cloud boundary was considered [76] which are depicted in Figure 7. In the convective updraft zone of a matured thunderstorm, usually four charge layers are found and outside updraft convection region at least six charge layers are seen. Forward and rearward anvils typically contain positive charges screened with negatively charged layers. However, this model does not represent super cell storms which usually have dominant positive charge at mid-levels in place of the typical main negative charge region [77], that is, super cell storms have "inverted polarity" charge structure, and they produce mainly positive cloud-to-ground lightning. The proposed model (Figure 7) is based on balloon soundings of convective region of storms that occurred in a limited portion of the lower mid-latitudes. Its validity for storms in other mid-latitude regions, the tropics, and high latitudes are yet to be tested because of scarce *in situ* data.

The global electric circuit (GEC) links the electric field and current flowing in the lower atmosphere, ionosphere, and magnetosphere forming a giant spherical condenser [50, 59, 78], which is charged by thunderstorms and other generators to a potential of several hundred thousand volts [79] and drives a vertical current through the fair weather atmosphere's columnar resistance ($\sim 1.3 \times 10^{17}\,\mathrm{ohm\,m^{-2}}$). Recently, Stolzenburg et al. [30] examined electric field data of 32 balloon soundings through MCS (mesoscale convective system) stratiform regions and inferred that for 15 cases the stratiform cloud was charging the GEC, while in the other 17 cases the cloud was discharging the circuit. Thus, the overall effect of MCS on GEC could be very large or very small. Davydenko et al. [80] modeled the quasi-electrostatic electric fields and currents inside and near an isolated MCS producing positive discharges and also sprites and showed that the large (~25 A) cloud-to-ground, quasi-DC current discharges the GEC. Their full impact on the GEC and that of the positive cloud-to-ground flashes which they initiate remains to be quantified. The vertical current causes weak electrification of stratified clouds [81] and produces a vertical

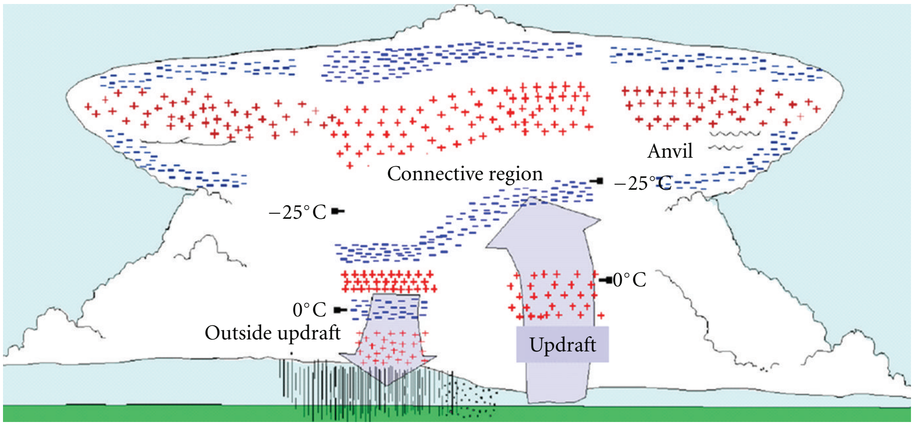

Figure 7: Conceptual model of the charge structure within an idealized isolated, mature thunderstorm based on 49 balloon sounding through different clouds, credit [22, 76].

potential gradient in the atmosphere near the Earth's surface. The circuit is closed by a horizontal current flowing through the highly conducting Earth and the ionosphere, and by vertical currents from the ground into the thunderstorm and from the top of the thunderstorm to the ionosphere.

3.1. AC Phenomena. Thunderstorm is the main source of both AC and DC coupling of the lower atmosphere and the ionosphere through lightning discharges. Return stroke of lightning discharge produces a pulse of radio energy containing waves from a few Hz up to several MHz. These signals propagate in the Earth-ionosphere wave guide over long distances (several thousand kms).The lowest frequencies of the discharge spectrum are ELF (3–3000 Hz) waves which excite the Earth-ionosphere cavity at 8, 14, 20, 26,...Hz, commonly known as Schumann resonances [82]. Resonances are being continuously excited because lightening is always occurring all over the world. Recent interest in Schumann resonance studies arose because the charge movement change (CMC) of the parent lightening discharge can be evaluated by measuring the contribution of each mode in the resonance pattern which has been used in TLEs studies [83–85]. Thus, coupling between the atmosphere near the surface of the Earth and the ionosphere can be studied on the time scale ranging from a fraction of second to less than 1 μs.

Some of the wave energy from lightning generated wave spectra propagates into the ionosphere and magnetosphere system, where it interacts with ambient plasma particles and affects the global ionospheric/magnetospheric energy budget. Waves in the very-low-frequency (VLF, 3–30 kHz) range propagate in the ducted mode along the geomagnetic field lines [86] and cause magnetospheric electron precipitation into the lower ionosphere, leading to D-region conductivity modification [87, 88]. Unducted VLF wave propagation can also take place [89]. Precipitating charged particles may participate in dissociation and ionization processes. There are VLF signatures of ionospheric disturbances associated with sprites [29, 43, 90, 91].

Hydromagnetic waves in the ultra-low-frequency (ULF, <3 Hz) propagate through the magnetosphere in one of several different modes at speeds considerably less than the speed of light. These waves interact with the ions in the ionosphere and magnetosphere. ULF waves sometimes are called magnetospheric analogues of seismic waves, which enable remote sensing of volumes from point measurements [92]. Waves in the frequency range $\sim$0.1–5 Hz also excite the ionospheric Alfven resonator (IAR) [93, 94] which is centered on the maximum of the ionospheric F_2-region where the refractive index for Alfven waves has a maximum value. The variation of spectral resonance structure of the IAR over a solar cycle has also been studied [95]. The boundaries of IAR are at $\sim$1000 km altitude in the topside ionosphere and in the E-region ionosphere.

The ULF-VLF electric fields were detected up to topside ionosphere over powerful Pacific Ocean typhoons [96]. However, details of the mechanisms are not yet known. Ionospheric heating due to lightning electromagnetic pulses was considered and modeled [97, 98]. Large amplitude radio waves from ground-based transmitters at VLF [99] and high frequency (HF, 3–30 MHz) [100] have been used to study ionospheric heating and modification. The associated phenomena include perturbations of electron density and temperature (increase of ionization at E-region, E-sporadic occurrence, increase of temperature, and electron density in F-region), excitation of electrostatic wave turbulence, and enhanced optical emissions.

The heating of the lower thermosphere by lightening-generated electromagnetic pulses may result into electrical breakdown and production of sprites and other optical emissions [101, 102]. The more exotic relativistic runaway (avalanche) breakdown mechanism, which can produce bremsstrahlung and gamma-rays radiations, has also been proposed and discussed [31, 103, 104]. Rowland [47] had reviewed both mechanisms. The two processes are coupled in the sprite streamer: the thermal discharge produces a beam of relativistic seed electrons that allow ignition of runaway relativistic discharge [105]. The thundercloud electric field at the lower thermosphere altitude is very transient, $\sim$1–10 ms, consequently the observed sprite and halo phenomena are also of transient nature. The conventional breakdown mechanism explains most observations of sprites and their associated phenomena. Gerken and Inan [106] indicated that relatively faint and diffuse sprites confined to lower altitude could be associated with very high-peak currents and short-duration lightning discharges. Adachi et al. [107] reported that the number of columns in a sprite event was proportional to the peak current intensity of positive cloud-to-ground discharges, which was not corroborated by the results of van der Velde et al. [108]. This suggests a complex relationship between sprite morphology and in-cloud lightning processes, which are not yet fully understood.

Streamers associated with sprites are needle-shaped filaments of ionization and are driven by enhanced electric fields due to charge separation in their head, and the sign of the charge in head determines polarity of streamers. Liu and Pasko [109] have modeled preionization ahead of the streamer by ionizing ultraviolet photon originating in the streamer head. Thus, they consider photo ionization to be important in determining the propagation and branching of sprite streamer. The presence of inhomogeneities in the background density could also be helpful in the formation of streamers [110]. They have also explained the polarity asymmetry in triggering sprite streamer. Blue jets sometimes appear at the top of an energetic thundercloud and propagating upwards with a vertical velocity of ~100 km s^{-1} to a maximum altitude of ~40 km or so. These are believed to be a positive streamer formed in the upward quasi-electrostatic field above the cloud [63, 101]. Many streamers originate from the surface of the leader head (both positive and negative).

Pasko and George [111] have modeled the branching streamers as fractal tree growing in three dimensions. Pasko et al. [112] reported an unusual blue jet above a relatively small thunderstorm cell in Puerto Rico, and propagating up to ~70 km altitude. Its characteristics above 42 km resembled those of sprites but having longer duration. This shows that an electrical contact was made between the top of the thunderstorm and the lower ionosphere. Gigantic jets extend from the cloud top to the ionosphere at 90 km altitude and can last up to ~0.8 sec [64, 113–115]. The observed gigantic jets were of negative polarity, produced by a normally electrified storm and therefore resembled to negative cloud-to-ionosphere discharges. Out of five gigantic jets, four events were found to be associated with ELF radio waves, while no cloud-to-ground lightning was observed to trigger these waves. This indicates that ELF waves were generated by negative cloud-to-ionosphere discharges. Each event reduces the ionospheric potential by removing 30 coulombs from the ionosphere [64], and their effect needs to be included in the GEC model.

3.2. DC Phenomena. In the atmospheric GEC, thunderstorms are thought to act as a giant battery distributed over the globe causing electric potentials up to ±100 MeV [74, 116], which are discharged in short (~1 ms) time scales by cloud-to-ground or intracloud lightning. Even upward discharges like gigantic jets also discharge the GEC. Such large potentials could produce energetic photons, X-rays [117], and even gamma rays [118], which are absorbed in the short distance in the lower atmosphere but could travel appreciable distance before being absorbed in the middle and upper atmosphere.

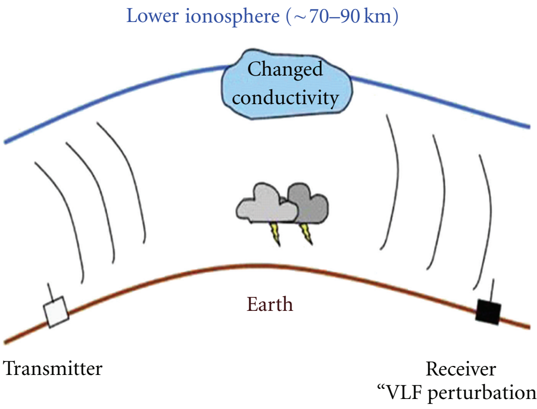

Figure 8: Diagram showing how the bottom side ionosphere and the mesosphere above a thunderstorm are probed by ground based transmitter, credit [40].

The inherent current sources [119, 120] in the mesosphere may generate large DC electric fields in the 50–70 km region [121], which has been observed via rocket measurements [122, 123] and remotely via a series of MF radar measurements [124–127]. The large DC electric fields increase the electron temperature and effective collision frequency in D-region. The enhanced electron temperature over neutral temperature violates the local thermodynamic equilibrium. During the tropospheric disturbances, tropospheric conductivity increases and the electrical coupling between the troposphere and the mesosphere results in grounding the mesospheric current source and in a significant decrease in the intensity of large mesospheric electric fields. The electron temperature and effective collision frequency decrease, which results into the D-region cooling [128, 129].

The upward electric currents associated with the charge separation process for a single active thunderstorm are ~1 A. Assuming ~1000 thunderstorms operate at a time over the entire globe, there is an upward current of ~1000 A. Theses currents charge the highly conducting ionosphere (~80 km altitude) to a potential of ~250 kV with respect to the Earth. The total charge on the plates of the Earth-ionosphere capacitor is ~2 × 10^5 C [59]. The electric currents (~2 pAm^{-2}) flowing from the ionosphere to the Earth's surface remote from thunderstorms depend on the electrical conductivity of the medium. The conductivity may change due to the incident flux of cosmic rays, precipitation of energetic charged particles from the magnetosphere, changes in aerosol distribution, changes in the pressure, and the temperature distribution in the troposphere. The upward quasi-electrostatic fields also change the atmospheric/lower ionospheric conductivity. Figure 8 shows the measuring principles of the change in the electrical conductivity above thunderstorms in the mesosphere and the lower ionosphere by measuring changes in the amplitudes of signals radiating from narrow band radio transmitters passing over thunderstorms region [40]. Global warming affects distribution of aerosol, temperature, and concentration of gases in the troposphere and stratosphere and hence modifies the dynamics of the lower and the middle atmosphere [130]. These changes result into changes in thunderstorm and lightening properties. It would be interesting to study the changes produced in the GEC by the global warming phenomena.

In the study of electrical coupling of the atmosphere and the ionosphere, the latter is considered to be equipotential surface, which is not exactly true. During very active auroral conditions, the potential difference across the narrow (1° or 2° wide) auroral electrojet region may become ∼100 kV. Even ionospheric dynamos may create potential differences of ∼15–20 kV between low and midlatitudes. These potential differences contradict the concept of the ionosphere as an equipotential surface. Ogawa et al. [131] have studied the variation of horizontal electric field in the middle atmosphere. The issue of equipotential further complicates because spatial and temporal variations of these potentials have a variety of scale lengths and various time scales. Intense atmospheric disturbances caused by earthquakes, volcanoes, tropical storms, typhoons, and so forth also affect electrical properties of the lower ionosphere. Some definite progresses in experimental and theoretical studies of man-made and natural effects on the ionosphere have been recently made [3]. However, there is significant gap in the understanding of an origin and interconnection of many processes involved in the atmosphere-ionosphere coupling.

4. Summary

The electrodynamic coupling between the Earth's atmosphere and the ionosphere is very complex and described by the global electric circuit. Currently, many aspects are not well understood. Recent measurements show that coupling influences both the electron density and electrical conductivity. Thunderstorms are the main generators situated in the troposphere. Measurements and theoretical studies have led to some understanding of charging mechanisms and charge structures of thunderstorms, however various aspects remain unanswered. Thunderstorms are large, complex, and short-lived phenomena; the charge generation and separation mechanism and charge structure are not known, and it is difficult to model them. The electrical processes are intimately related to the cloud dynamics or motions and to the microphysics of the clouds which are incompletely known or understood. Their detailed comprehension is essential as it controls climate and its variability. For proper understanding, it is essential to make simultaneous measurements of electric parameters from different parts of thunderstorms at small time intervals.

The role of transient luminous events on the local ionospheric potentials, modification of the electric field near the Earth's surface, and charging/discharging of GEC are yet to be quantified. Even there is dispute about the breakdown mechanism involved in the mesospheric discharges generating sprites, blue jets, elves, and so forth. Lightning-generated whistler mode waves cause precipitation of Van Allen belt electrons which modify D-region conductivity. The relative importance of different types of whistler mode waves is not known.

The generation of transient mesospheric electric field needs detailed study as it provides a basis for developing coupled troposphere-mesosphere-ionosphere electrodynamic models under disturbed conditions. In fact, numerous phenomena that occur in the upper atmosphere of the Earth are caused by the sources located in the lower atmosphere and on the ground such as thunderstorms, typhoons, dust storms, earthquakes, volcanic eruptions, and radioactive emissions from the nuclear power plants. All these phenomena affect the electrical conductivity from the Earth's surface to the lower ionosphere. Variation of conductivity and external current in the lower atmosphere lead to the perturbation of electric current flowing in the GEC and to the associated DC electric field perturbations both on the Earth's surface and in the ionosphere and hence affect the electrodynamic coupling of the Earth's atmosphere and the ionosphere.

Acknowledgments

A. K. Singh is thankful to DST for financial support. A. K. Singh, D. Siingh are also thankful to Indian Space Research Organization (ISRO) for partial financial support under CAWSES program. R. P. Singh acknowledges the facilities provided by the Head, Department of Physics, BHU, Varanasi, India.

References

[1] D. K. Singh, R. P. Singh, and A. K. Kamra, "The electrical environment of the earth's atmosphere: a review," *Space Science Reviews*, vol. 113, no. 3-4, pp. 375–408, 2004.

[2] T. Neubert, "Atmospheric science: on sprites and their exotic kin," *Science*, vol. 300, no. 5620, pp. 747–749, 2003.

[3] V. M. Sorokin and V. M. Chmyrev, "Atmosphere-Ionosphere Electrodynamic coupling," in *The Atmosphere and Ionosphere*, Physics of Earth and Space Environment, pp. 97–104, Springer, New York, NY, USA, 2010.

[4] E. S. Kazimirovsky, "Coupling from below as a source of ionospheric variability: a review," *Annals of Geophysics*, vol. 45, no. 1, pp. 1–30, 2002.

[5] T. B. Jones, K. Davies, and B. Wieder, "Observations of D-Region modifications at low and very low frequencies," *Nature*, vol. 238, no. 5358, pp. 33–34, 1972.

[6] R. J. Gamble, C. J. Rodger, M. A. Clilverd et al., "Radiation belt electron precipitation by man-made VLF transmissions," *Journal of Geophysical Research A*, vol. 113, no. 10, Article ID A10211, 2008.

[7] V. O. Rapoport, V. L. Frolov, G. P. Komrakov et al., "Some results of measuring the characteristics of electromagnetic and plasma disturbances stimulated in the outer ionosphere by high-power high-frequency radio emission from the "Sura" facility," *Radiophysics and Quantum Electronics*, vol. 50, no. 8, pp. 645–656, 2007.

[8] V. M. Sorokin, V. M. Chmyrev, and A. K. Yaschenko, "Electrodynamic model of the lower atmosphere and the ionosphere coupling," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 63, pp. 1681–1691, 2001.

[9] H. C. Chang and U. S. Inan, "Lightning-induce energetic electron precipitation from the magnetosphere," *Journal of Geophysical Research*, vol. 90, pp. 4531–4539, 1985.

[10] U. S. Inan, D. C. Shafer, W. Y. Yip, and R. E. Orville, "Subionospheric VLF signatures of nighttime D-region perturbations in the vicinity of lightning discharges," *Journal of Geophysical Research*, vol. 93, pp. 11455–11467, 1988.

[11] R. C. Moore, U. S. Inan, T. F. Bell, and E. J. Kennedy, "ELF waves generated by modulated HF heating of the auroral electrojet and observed at a ground distance of 4400 km," *Journal of Geophysical Research A*, vol. 112, no. 5, Article ID A05309, 2007.

[12] V. L. Frolov, V. O. Rapoport, G. P. Komrakov et al., "Density ducts formed by heating the Earth's ionosphere with high-power HF radio waves," *JETP Letters*, vol. 88, no. 12, pp. 790–794, 2008.

[13] G. M. Milikh, K. Papadopoulos, H. Shroff et al., "Formation of artificial ionospheric ducts," *Geophysical Research Letters*, vol. 35, no. 17, Article ID L17104, 2008.

[14] N. V. Dzhordzhio, M. M. Mogilevskii, V. M. Chmyrev et al., "Acceleration of ions in the plasma environment of the Earth by the radiation from a low-frequency transmitter on the ground," *JETP Letters*, vol. 46, pp. 405–409, 1987.

[15] H. L. Rowland, R. F. Ferseler, and P. F. Bernhardt, "Breakdown of the neutral atmosphere in the D region due to lightning driven electromagnetic pulses," *Journal of Geophysical Research*, vol. 101, pp. 7935–7945, 1996.

[16] V. P. Pasko, U. S. Inan, T. F. Bell, and Y. N. Taranenko, "Sprites produced by quasi-electrostatic heating and ionization in the lower ionosphere," *Journal of Geophysical Research A*, vol. 102, no. 3, Article ID 96JA03528, pp. 4529–4561, 1997.

[17] S. Fadnavis, D. Siingh, and R. P. Singh, "Mesospheric inversion layer and sprites," *Journal of Geophysical Research D*, vol. 114, no. 23, Article ID D23307, 2009.

[18] D. D. Sentman, E. M. Wescott, R. H. Picard et al., "Simultaneous observations of mesospheric gravity waves and sprites generated by a midwestern thunderstorm," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 65, no. 5, pp. 537–550, 2003.

[19] M. J. Heavner, *Optical spectroscopic observations of sprites, blue jets, and elves: Inferred microphysical processes and their macrophysical implications*, Ph.D. dissertation, University of Alaska Fairbanks, 2000.

[20] V. M. Sorokin, V. M. Chmyrev, and N. V. Isaev, "A generation model of small-scale geomagnetic field-aligned plasma inhomogeneities in the ionosphere," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 60, no. 13, pp. 1331–1342, 1998.

[21] V. M. Chmyrev, V. M. Sorokin, and O. A. Pokhotelov, "Theory of small scale plasma density inhomogeneities and ULF/ELF magnetic field oscillations excited in the ionosphere prior to earthquakes," in *Atmospheric and Ionospheric Electromagnetic Phenomena Associated with Earthquakes*, M. Hayakawa, Ed., pp. 759–776, Terrapublication, Tokyo, Japan, 1999.

[22] M. Stolzenburg and T. C. Marshall, "Charge structure and dynamics in thunderstorms," *Space Science Reviews*, vol. 137, no. 1-4, pp. 355–372, 2008.

[23] Y. Yair, "Charge generation and separation processes," *Space Science Reviews*, vol. 137, no. 1-4, pp. 119–131, 2008.

[24] C. Saunders, "Charge separation mechanisms in clouds," *Space Science Reviews*, vol. 137, no. 1-4, pp. 335–353, 2008.

[25] E. Williams, M. Valente, E. Gerken, and R. Golka, "Calibrated radiance measurements with an air-field glow discharge tube: application to sprites in the mesosphere," in *Sprites, Elves and Intense Lightning Discharges*, M. Fullekrug et al., Ed., vol. 225 of *NATO Science Series*, pp. 237–251, Springer, New York, NY, USA, 2006.

[26] M. J. Rycroft and M. Füllekrug, "The initiation and evolution of SPECIAL," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 66, no. 13-14, pp. 1103–1113, 2004.

[27] K. L. Aplin, R. G. Harrison, and M. J. Rycroft, "Investigating earth's atmospheric electricity: a role model for planetary studies," *Space Science Reviews*, vol. 137, no. 1-4, pp. 11–27, 2008.

[28] M. J. Rycroft, "Electrical processes coupling the atmosphere and ionosphere: an overview," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 68, no. 3-5, pp. 445–456, 2006.

[29] T. C. Marshall, M. Stolzenburg, C. R. Maggio et al., "Observed electric fields associated with lightning initiation," *Geophysical Research Letters*, vol. 32, no. 3, pp. 1–5, 2005.

[30] M. Stolzenburg, T. C. Marshall, W. D. Rust, E. Bruning, D. R. MacGorman, and T. Hamlin, "Electric field values observed near lightning flash initiations," *Geophysical Research Letters*, vol. 34, no. 4, Article ID L04804, 2007.

[31] A. V. Gurevich and K. P. Zybin, "Runaway breakdown and electric discharges in thunderstorms," *Physics-Uspekhi*, vol. 44, no. 11, pp. 1119–1140, 2001.

[32] A. V. Gurevich and K. P. Zybin, "Runaway breakdown and the mysteries of lightning," *Physics Today*, vol. 58, no. 5, pp. 37–43, 2005.

[33] A. V. Gurevich, G. G. Mitko, V. P. Antonova et al., "An intracloud discharge caused by extensive atmospheric shower," *Physics Letters Section A*, vol. 373, no. 39, pp. 3550–3553, 2009.

[34] M. A. Uman, *The Lightning Discharge*, Academic, Orlando, Fla, USA, 1987.

[35] D. D. Sentman and E. M. Wescott, "Observations of upper atmospheric optical flashes recorded from an aircraft," *Geophysical Research Letters*, vol. 20, no. 24, pp. 2857–2860, 1993.

[36] C. Haldoupis, R. J. Steiner, A. Mika et al., ""Early/slow" events: a new category of VLF perturbations observed in relation with sprites," *Journal of Geophysical Research A*, vol. 111, no. 11, Article ID A11321, 2006.

[37] L. Liszka, "On the possible infrasound generation by sprites," *Journal of Low Frequency Noise Vibration and Active Control*, vol. 23, no. 2, pp. 85–93, 2004.

[38] L. Liszka and Y. Hobara, "Sprite-attributed infrasonic chirps-their detection, occurrence and properties between 1994 and 2004," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 68, no. 11, pp. 1179–1188, 2006.

[39] T. Farges, E. Blanc, A. Le Pichon, T. Neubert, and T. H. Allin, "Identification of infrasound produced by sprites during the Sprite2003 campaign," *Geophysical Research Letters*, vol. 32, no. 1, pp. 1–4, 2005.

[40] T. Neubert, M. Rycroft, T. Farges et al., "Recent results from studies of electric discharges in the mesosphere," *Surveys in Geophysics*, vol. 29, no. 2, pp. 71–137, 2008.

[41] S. B. Mende, H. U. Frey, R. R. Hsu et al., "D region ionization by lightning-induced electromagnetic pulses," *Journal of Geophysical Research A*, vol. 110, no. 11, Article ID A11312, 2005.

[42] Z. Cheng, S. A. Cummer, H. T. Su, and R. R. Hsu, "Broadband very low frequency measurement of D region ionospheric perturbations caused by lightning electromagnetic pulses," *Journal of Geophysical Research A*, vol. 112, no. 6, Article ID A06318, 2007.

[43] C. Haldoupis, N. Amvrosiadi, B. R. T. Cotts, O. A. van der Velde, O. Chanrion, and T. Neubert, "More evidence for

a one-to-one correlation between Sprites and Early VLF perturbations," *Journal of Geophysical Research*, vol. 115, p. A07304, 2010.

[44] C. Haldoupis, A. Mika, and S. Shalimov, "Modeling the relaxation of early VLF perturbations associated with transient luminous events," *Journal of Geophysical Research A*, vol. 114, no. 10, Article ID A00E04, 2009.

[45] A. Mika, *Very low frequency EM wave studies of transient luminous events in the lower ionosphere*, Ph.D. thesis, University of Crete, 2007.

[46] Y. N. Taranenko, U. S. Inan, and T. F. Bell, "Interaction with the lower ionosphere of electromagnetic pulses from lightning: heating, attachment, and ionization," *Geophysical Research Letter*, vol. 20, pp. 1539–1542, 1993.

[47] H. L. Rowland, "Theories and simulations of elves, sprites and blue jets," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 60, no. 7–9, pp. 831–844, 1998.

[48] Y. Hiraki, T. Lizhu, H. Fukunishi, K. Nambu, and H. Fujiwa, "Development of a new numerical model for investigation the energetic of sprites," *Eos Transactions of American Geophysical Union*, vol. 83, no. 47, 2002.

[49] J. M. Galloway, F. J. Dentener, D. G. Capone et al., "Nitrogen cycles: past, present and future," *Biogeochemistry*, vol. 70, pp. 153–226, 2004.

[50] D. Siingh, R. P. Singh, A. K. Kamra et al., "Review of electromagnetic coupling between the Earth's atmosphere and the space environment," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 67, no. 6, pp. 637–658, 2005.

[51] U. Schumann and H. Huntrieser, "The global lightning-induced nitrogen oxides source," *Atmospheric Chemistry and Physics*, vol. 7, no. 14, pp. 3823–3907, 2007.

[52] P. J. Crutzen, "Photochemical reactions initiated by and influencing ozone in unpolluted tropospheric air," *Tellus*, vol. 26, pp. 47–57, 1974.

[53] W. C. Burgess and U. S. Inan, "The role of ducted whistlers in the precipitation loss and equilibrium flux of radiation belt electrons," *Journal of Geophysical Research*, vol. 98, pp. 15643–15650, 1993.

[54] T. J. Rosenberg, R. A. Helliwell, and J. P. Katsufrakis, "Electron precipitation associated with discrete very low frequency emissions," *Journal of Geophysical Research*, vol. 76, p. 8445, 1971.

[55] M. J. Rycroft, "Enhanced energetic electron intensities at 100 km altitude and a whistler propagating through the plasmasphere," *Planetary and Space Science*, vol. 21, no. 2, pp. 239–251, 1973.

[56] R. J. Goldberg, S. A. Curtis, and J. R. Barcus, "Detailed spectral structure of magnetospheric electron bursts precipitated by lightning," *Journal of Geophysical Research*, vol. 92, pp. 2505–2512, 1987.

[57] H. D. Voss, W. L. Imhof, M. Walt et al., "Lightning-induced electron precipitation," *Nature*, vol. 312, no. 5996, pp. 740–742, 1984.

[58] H. D. Voss, "Satellite observations of lightning-induced electron precipitation," *Journal of Geophysical Research A*, vol. 103, no. 6, Article ID 97JA02878, pp. 11725–11744, 1998.

[59] M. J. Rycroft, S. Israelsson, and C. Price, "The global atmospheric electric circuit, solar activity and climate change," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 62, no. 17-18, pp. 1563–1576, 2000.

[60] E. R. Williams and S. J. Heckman, "The local diurnal variation of cloud electrification and the global diurnal variation of negative charge on the Earth," *Journal of Geophysical Research*, vol. 98, no. 3, pp. 5221–5234, 1993.

[61] M. J. Rycroft, A. Odzimek, N. F. Arnold, M. Füllekrug, A. Kułak, and T. Neubert, "New model simulations of the global atmospheric electric circuit driven by thunderstorms and electrified shower clouds: the roles of lightning and sprites," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 69, no. 17-18, pp. 2485–2509, 2007.

[62] C. L. Ziegler and D. R. Macgorman, "Observed lightning morphology relative to modeled space charge and electric field distributions in a tornadic storm," *Journal of the Atmospheric Sciences*, vol. 51, no. 6, pp. 833–851, 1994.

[63] V. P. Pasko, "Electric jets," *Nature*, vol. 423, no. 6943, pp. 927–929, 2003.

[64] H. T. Su, R. R. Hsu, A. B. Chen et al., "Gigantic jets between a thundercloud and the ionosphere," *Nature*, vol. 423, no. 6943, pp. 974–976, 2003.

[65] Y. P. Raizer, G. M. Milikh, and M. N. Shneider, "On the mechanism of blue jet formation and propagation," *Geophysical Research Letters*, vol. 33, no. 23, Article ID L23801, 2006.

[66] Y. P. Raizer, G. M. Milikh, and M. N. Shneider, "Leader-streamers nature of blue jets," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 69, no. 8, pp. 925–938, 2007.

[67] M. Johnston, "Electromagnetic fields generated by earthquakes," in *International Handbook of Earthquake and Engineering Seismology*, W. Lee, H. Kanamori, P. Jennings, and C. Kisslinger, Eds., part A, pp. 621–634, Academic Press, 2002.

[68] M. Parrot, "Statistical study of ELF/VLF emissions recorded by a low-altitude satellite during seismic events," *Journal of Geophysical Research*, vol. 99, pp. 23339–23347, 1994.

[69] V. A. Liperovsky, O. A. Pokhotelov, C. V. Meister, and E. V. Liperovskaya, "Physical models of coupling in the lithosphere-atmosphere-ionosphere system before earthquakes," *Geomagnetism and Aeronomy*, vol. 48, no. 6, pp. 795–806, 2008.

[70] N. F. Arnold and T. R. Robinson, "Solar cycle changes to planetary wave propagation and their influence on the middle atmosphere circulation," *Annales Geophysicae*, vol. 16, no. 1, pp. 69–76, 1998.

[71] S. A. Bowhill, "Interaction between the stratosphere and ionosphere," *Annals of the International Year of the Quiet Sun*, vol. 5, pp. 83–95, 1969.

[72] P. R. Krehbiel, J. A. Riousset, V. P. Pasko et al., "Upward electrical discharges from thunderstorms," *Nature Geoscience*, vol. 1, no. 4, pp. 233–237, 2008.

[73] T. C. Marshall, M. Stolzenburg, P. R. Krehbiel, N. R. Lund, and C. R. Maggio, "Electrical evolution during the decay stage of New Mexico thunderstorms," *Journal of Geophysical Research D*, vol. 114, no. 2, Article ID D02209, 2009.

[74] V. A. Rakov and M. A. Uman, *Lightning: Physics and Effects*, Cambridge University Press, Cambridge, UK, 2003.

[75] M. Stolzenburg, "Observations of high ground flash densities of positive lightning in summertime thunderstorms," *Monthly Weather Review*, vol. 122, no. 8, pp. 1740–1750, 1994.

[76] M. Stolzenburg, W. D. Rust, B. F. Smull, and T. C. Marshall, "Electrical structure in thunderstorm convective regions 1. Mesoscale convective systems," *Journal of Geophysical Research D*, vol. 103, no. 12, pp. 14059–14078, 1998.

[77] W. D. Rust, D. R. MacGorman, E. C. Bruning et al., "Inverted-polarity electrical structures in thunderstorms in the Severe Thunderstorm Electrification and Precipitation Study (STEPS)," *Atmospheric Research*, vol. 76, no. 1-4, pp. 247–271, 2005.

[78] D. Siingh, V. Gopalakrishnan, R. P. Singh et al., “The atmospheric global electric circuit: an overview,” *Atmospheric Research*, vol. 84, no. 2, pp. 91–110, 2007.

[79] R. G. Roble and L. Tzur, “The global atmospheric-electrical circuit,” in *The Earth's Electrical Environment*, pp. 206–231, National Academy Press, Washington, DC, USA, 1986.

[80] S. S. Davydenko, E. A. Mareev, T. C. Marshall, and M. Stolzenburg, “On the calculation of electric fields and currents of mesoscale convective systems,” *Journal of Geophysical Research D*, vol. 109, no. 11, pp. D11103–10, 2004.

[81] R. G. Harrison and K. S. Carslaw, “Ion-aerosol-cloud processes in the lower atmosphere,” *Reviews of Geophysics*, vol. 41, no. 3, pp. 2–1, 2003.

[82] W. O. Schumann, “U‘ ber die strahlunglosen eigenschwingungen einer leitenden Kugel, die von einer Luftschicht und einer Ionospharenh ulle umgeben ist,” *Zeitschrift Naturforschung*, vol. 7, pp. 6627–6628, 1952.

[83] E. Huang, E. Williams, R. Boldi et al., “Criteria for sprites and elves based on Schumann resonance observations,” *Journal of Geophysical Research D*, vol. 104, no. 14, pp. 16943–16964, 1999.

[84] E. Greenberg, C. Price, Y. Yair, M. Ganot, J. Bór, and G. Sátori, “ELF transients associated with sprites and elves in eastern Mediterranean winter thunderstorms,” *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 69, no. 13, pp. 1569–1586, 2007.

[85] E. Greenberg, C. Price, Y. Yair, C. Haldoupis, O. Chanrion, and T. Neubert, “ELF/VLF signatures of sprite-producing lightning discharges observed during the 2005 EuroSprite campaign,” *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 71, no. 12, pp. 1254–1266, 2009.

[86] R. A. Helliwell, “40 years of whistlers,” in *Modern Radio Science*, pp. 189–212, Oxford University Press, New York, NY, USA, 1993.

[87] C. J. Rodger, “Subionospheric VLF perturbations associated with lightning discharges,” *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 65, no. 5, pp. 591–606, 2003.

[88] U. S. Inan, S. A. Cummer, and R. A. Marshall, “A survey of ELF and VLF research on lightning-ionosphere interactions and causative discharges,” *Journal of Geophysical Research A*, vol. 115, no. 6, Article ID A00E36, 2010.

[89] M. Hayakawa, “Association of whistlers with lightning discharges on the Earth and on Jupiter,” *Journal of Atmospheric and Terrestrial Physics*, vol. 57, no. 5, pp. 525–535, 1995.

[90] M. J. Rycroft, “Interactions between whistler-mode waves and energetic electrons in the coupled system formed by the magnetosphere, ionosphere and atmosphere,” *Journal of Atmospheric and Terrestrial Physics*, vol. 53, no. 9, pp. 849–858, 1991.

[91] A. Mika, C. Haldoupis, R. A. Marshall, T. Neubert, and U. S. Inan, “Subionospheric VLF signatures and their association with sprites observed during EuroSprite-2003,” *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 67, no. 16, pp. 1580–1597, 2005.

[92] A. K. Singh, “Ultra low frequency waves,” in *Solar Terrestrial Environment: Space Weather*, R. P. Singh, R. Singh, and A. K. Singh, Eds., pp. 429–440, Allied, New Delhi, India, 2003.

[93] P. P. Belyaev, S. V. Polyakov, V. O. Rapoport, and V. Y. Trakhtengerts, “The ionospheric Alfven resonator,” *Journal of Atmospheric and Terrestrial Physics*, vol. 52, no. 9, pp. 781–788, 1990.

[94] A. I. Sukhorukov and P. Stubbe, “Problems of blue jet theories,” *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 60, no. 7–9, pp. 725–732, 1998.

[95] P. P. Belyaev, S. V. Polyakov, E. N. Ermakova, and S. V. Isaev, “Solar cycle variations in the ionospheric Alfven resonator 1985-1995,” *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 62, no. 4, pp. 239–248, 2000.

[96] G. A. Mikhailova, Y. M. Mikhailov, and O. V. Kapustina, “ULF-VLF electric fields in the external ionosphere over powerful typhoons in Pacific Ocean,” *International Journal of Geomagnetism and Aeronomy*, vol. 2, no. 2, 2000.

[97] V. P. Pasko, U. S. Inan, Y. N. Taranenko, and T. F. Bell, “Heating, ionization and upward discharges in the mesosphere due to intense quasi-electrostatic thunderstorm fields,” *Geophysical Research Letter*, vol. 22, pp. 365–370, 1995.

[98] C. J. Rodger, M. Cho, M. A. Clilverd, and M. J. Rycroft, “Lower ionospheric modification by lightning-EMP: simulation of the night ionosphere over the United States,” *Geophysical Research Letters*, vol. 28, no. 2, pp. 199–202, 2001.

[99] U. S. Inan, “VLF heating of the lower ionosphere,” *Geophysical Research Letters*, vol. 17, pp. 729–732, 1990.

[100] P. Stubbe, “Review of ionospheric modification experiments at Tromsø,” *Journal of Atmospheric and Terrestrial Physics*, vol. 58, no. 1-4, pp. 349–368, 1996.

[101] V. P. Pasko, U. S. Inan, and T. F. Bell, “Sprites as luminous columns of ionization produced by quasi-electrostatic thundercloud fields,” *Geophysical Research Letters*, vol. 23, no. 6, pp. 649–652, 1996.

[102] M. Cho and M. J. Rycroft, “Computer simulation of the electric field structure and optical emission from cloud-top to the ionosphere,” *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 60, no. 7–9, pp. 871–888, 1998.

[103] L. P. Babich, A. Y. Kudryavtsev, M. L. Kudryavtseva, and I. M. Kutsyk, “Atmospheric gamma-ray and neutron flashes,” *Journal of Experimental and Theoretical Physics*, vol. 106, no. 1, pp. 65–76, 2008.

[104] M. Füllekrug, R. Roussel-Dupré, E. M. D. Symbalisty et al., “Relativistic runaway breakdown in low-frequency radio,” *Journal of Geophysical Research A*, vol. 115, no. 1, Article ID A00E09, 2010.

[105] O. Chanrion and T. Neubert, “Production of runaway electrons by negative streamer discharges,” *Journal of Geophysical Research A*, vol. 115, no. 6, Article ID A00E32, 2010.

[106] E. Gerken and U. Inan, “Streamers and diffuse glow observed in upper atmospheric electrical discharges,” *IEEE Transactions on Plasma Science*, vol. 33, no. 2 I, pp. 282–283, 2005.

[107] T. Adachi, H. Fukunishi, Y. Takahashi, and M. Sato, “Roles of the EMP and QE field in the generation of columniform sprites,” *Geophysical Research Letters*, vol. 31, no. 4, pp. L04107–4, 2004.

[108] O. A. van der Velde, A. Mika, S. Soula, C. Haldoupis, T. Neubert, and U. S. Inan, “Observations of the relationship between sprite morphology and in-cloud lightning processes,” *Journal of Geophysical Research D*, vol. 111, no. 15, Article ID D15203, 2006.

[109] N. Liu and V. P. Pasko, “Effects of photoionization on propagation and branching of positive and negative streamers in sprites,” *Journal of Geophysical Research A*, vol. 109, Article ID A04301, 2004.

[110] J. Qin, S. Celestin, and V. P. Pasko, “On the inception of streamers from sprite halo events produced by lightning discharges with positive and negative polarity,” *Journal of Geophysical Research A*, vol. 116, Article ID A06305, 2011.

[111] V. P. Pasko and J. J. George, “Three-dimensional modeling of blue jets and blue starters,” *Journal of Geophysical Research A*, vol. 107, Article ID 1458, 2002.

[112] V. P. Pasko, M. A. Stanley, J. D. Mathews, U. S. Inan, and T. G. Wood, “Electrical discharge from a thundercloud top to the lower ionosphere,” *Nature*, vol. 416, no. 6877, pp. 152–154, 2002.

[113] C. L. Kuo, A. B. Chen, Y. J. Lee et al., “Modeling elves observed by FORMOSAT-2 satellite,” *Journal of Geophysical Research A*, vol. 112, no. 11, Article ID A11312, 2007.

[114] C. L. Kuo, J. K. Chou, L. Y. Tsai et al., “Discharge processes, electric field, and electron energy in ISUAL recorded gigantic jets,” *Journal of Geophysical Research A*, vol. 114, no. 4, Article ID A04314, 2009.

[115] S. A. Cummer, J. Li, F. Han et al., “Quantification of the troposphere-to-ionosphere charge transfer in a gigantic jet,” *Nature Geoscience*, vol. 2, no. 9, pp. 617–620, 2009.

[116] V. I. Ermakov and Y. I. Stozhkov, *Thunderstorm Cloud Physics*, Lebedev Physical Institute, Russian Academy of Sciences, 2004.

[117] M. P. McCarthy and G. K. Parks, “On the modulation of X ray fluxes in thunderstorms,” *Journal of Geophysical Research*, vol. 97, no. 5, pp. 5857–5864, 1992.

[118] G. J. Fishman, P. N. Bhat, R. Mallozzi et al., “Discovery of intense gamma-ray flashes of atmospheric origin,” *Science*, vol. 264, no. 5163, pp. 1313–1316, 1994.

[119] A. C. Aikin and N. C. Maynard, “A Van de Graaf source mechanism for middle atmospheric vertical electric fields,” *Journal of Atmospheric and Terrestrial Physics*, vol. 52, no. 9, pp. 695–705, 1990.

[120] S. V. Polyakov, V. O. Rapoport, and V. Y. Trakhtengerts, “the generation of electric fields in the upper atmosphere,” *Geomagnetism Aeronomy*, vol. 30, p. 869, 1990.

[121] R. A. Goldberg, “Middle atmospheric electrodynamics during map,” *Advances in Space Research*, vol. 10, no. 10, pp. 209–217, 1990.

[122] M. C. Kelley, C. L. Siefring, and R. F. Pfaff Jr., “Large middle atmospheric electric fields, fact or fiction?” *Geophysical Research Letter*, vol. 10, p. 733, 1983.

[123] A. M. Zadorozhny and A. A. Tyutin, “Effects of geomagnetic activity on the mesospheric electric fields,” *Annales Geophysicae*, vol. 16, no. 12, pp. 1544–1551, 1998.

[124] A. M. Gokov and S. I. Martynenko, “Changes in the electron collision frequency and electric field in the lower ionosphere,” *Geomagnetism Aeronomy* , vol. 37, p. 76, 1997.

[125] S. I. Martynenko, V. T. Rozumenko, A. M. Tsymbal, O. F. Tyrnov, and A. M. Gokov, “Mesospheric electric field measurements with a partial reflection radar,” *Journal Atmospheric Electricity*, vol. 19, p. 81, 1999.

[126] S. I. Martynenko, V. T. Rozumenko, and O. F. Tyrnov, “New possibilities for mesospheric electricity diagnostics,” *Advances in Space Research*, vol. 27, no. 6-7, pp. 1127–1132, 2001.

[127] C. E. Meek, A. H. Manson, S. I. Martynenko, V. T. Rozumenko, and O. F. Tyrnov, “Remote sensing of mesospheric electric fields using MF radars,” *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 66, no. 10, pp. 881–890, 2004.

[128] I. M. Fuks, R. S. Shubova, and S. I. Martynenko, “Lower ionosphere response to conductivity variations of the near-earth atmosphere,” *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 59, no. 9, pp. 961–965, 1997.

[129] S. I. Martynenko and S. F. Clifford, “On the electrical coupling between the troposphere and the mesosphere,” *International Journal of Geomagnetism and Aeronomy*, vol. 6, Article ID GI2007, 2006.

[130] D. Siingh, R. P. Singh, A. K. Singh, M. N. Kulkarni, A. S. Gautam, and A. K. Singh, “Solar activity, lightning and climate,” *Surveys in Geophysics*, vol. 32, no. 6, pp. 659–703, 2011.

[131] T. Ogawa, Y. Tanaka, A. Huzita, and M. Yasuhara, “Horizontal electric fields in the middle latitude,” *Planetary and Space Science*, vol. 23, no. 5, pp. 825–830, 1975.

View publication stats

(No Model.)

M. W. DEWEY.

METHOD OF UTILIZING NATURAL ELECTRIC ENERGY.

No. 414,943. Patented Nov. 12, 1889.

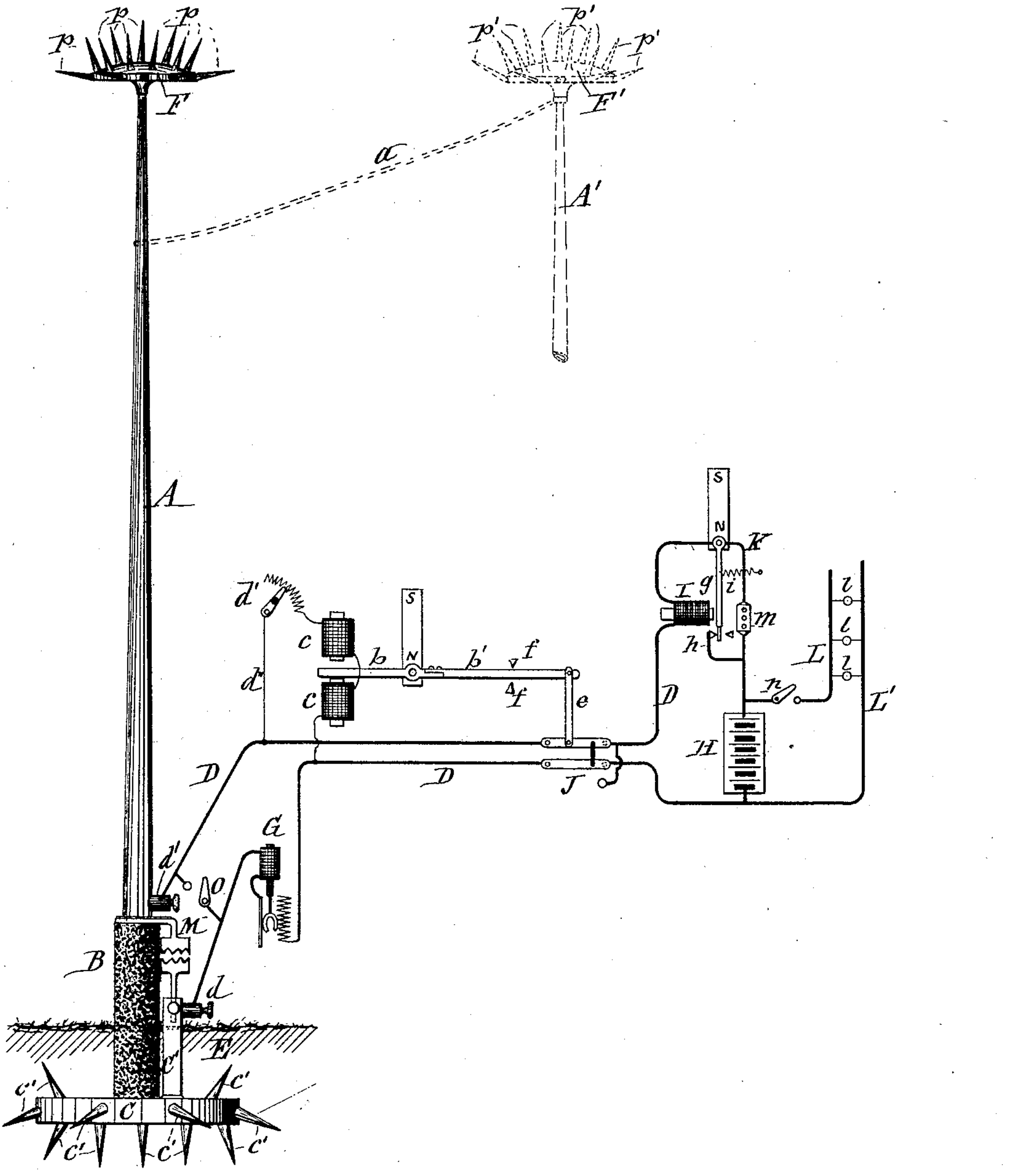
p
p
p
F
p'
p'
p'
F'
a
A'
A
S
N
K
I
g
i
m
h
L
l
l
l
d'
c
b
N
S
b'
f
f
c
e
d
D
D
H
L'
n
J
D
G
d'
o
M
B
d
E
c'
C
c'
c'
c'
c'
c'
c'

WITNESSES:

C. L. Bendixon

A. F. Walz.

Mark W. Dewey INVENTOR

BY

Duell, Laass & Duell

ATTORNEYS

N. PETERS, Photo-Lithographer, Washington, D. C.

UNITED STATES PATENT OFFICE.

MARK W. DEWEY, OF SYRACUSE, NEW YORK, ASSIGNOR TO THE DEWEY CORPORATION, OF SAME PLACE.

METHOD OF UTILIZING NATURAL ELECTRIC ENERGY.

SPECIFICATION forming part of Letters Patent No. 414,943, dated November 12, 1889.

Application filed May 1, 1889. Serial No. 309,171. (No model.)

To all whom it may concern:

Be it known that I, MARK W. DEWEY, of Syracuse, in the county of Onondaga, in the State of New York, have invented new and useful Improvements in the Method of Utilizing Natural Electric Energy, of which the following, taken in connection with the accompanying drawings, is a full, clear, and exact description.

The object of this invention is to provide a method whereby natural electrical energy—such as the so-called "atmospheric electrical energy," or electrical energy that may be derived from the difference of potential existing between two points, one being the earth and the other the atmosphere at an elevation above the earth—may be collected or utilized for the various uses to which electricity is applied.

It has been found that the presence of electricity in the upper regions of the atmosphere is not confined to thunder-clouds, but can be detected at all times and often in great quantities in different conditions of the atmosphere. In fine weather this electricity is mostly positive; but in showery or stormy weather negative electricity is as frequently met with as positive, and it is in such weather that the indications of electricity, whether positive or negative, are usually the strongest. It has also been found that as we proceed farther from the earth's surface, whether upward from a level plane thereof or horizontally from an elevation, the potential of points in the air becomes more and more different from that of the earth, the difference being, in a broad sense, simply proportional to the distance; hence we can infer that there is electricity residing on the surface of the earth, the density of which at any moment in the locality of observation is measured by the difference of potential found to exist between the earth and a given point in the air near it. The results of observations show that the variations of the electricity residing in the atmosphere is the main cause of the variations of the electricity on the surface of the earth. A charged cloud or body of air induces electricity of the opposite kind to its own on the parts of the earth's surface over which it passes and produces such variations. The difference of potential in increasing the distance from the earth is due to electricity induced on the surface of the earth by opposite electricity in the air overhead, and the air being a nonconductor the electricities are unable to combine. As electrical density is greater on projecting parts of a surface than on those which are plane or concave, stronger indications are obtained on hills than in valleys, if the collecting apparatus be at the same distance from the ground in both cases. The average difference of potential is greater in the winter than in the summer. Little or no effects can be obtained within inclosures or under trees, as they tend to screen the apparatus.

Inasmuch as electricity travels in preference through the best conductors, it follows that if a path of low resistance is formed (such as the erection of a metal pole) to a sufficient elevation above the earth the electricities in the atmosphere and that on the surface of the earth will tend to combine and travel through the said path in a current or currents, and if this pole terminates in a metal point or a number of such points the earth and clouds exchange their opposite electricities without a disruptive discharge—as the lightning—but in a slow and gradual way through convection. Besides supplying the top of the pole with points, a metal plate to which the said points are fixed attracts the opposite electricity in the atmosphere. In order to obtain a greater and increased effect, a large metal plate is buried at the foot of the pole and electrically connected therewith, and is provided with points or branches extending in different directions in the ground. The plate and branches may be surrounded by metallic refuse, coke, or other good conducting substance. The metallic points on the top of the pole should be sharp, and preferably of copper, and may be platinized, gilded, or galvanized to prevent corrosion. It having been ascertained by practical experiments that either a flame or dropping water at an elevation above the surface of the earth produces convection of electricity, it is obvious that such means may be employed in place of the points hereinbefore referred to. As the electricity in the atmosphere is sometimes positive and other times negative, the direction of the currents is not always the same—that is, the atmospheric electrical energy is composed of a current of an alternating character, flowing in one direction on an average about as much of the time as in the opposite direction, but the length of the current in a certain direction, or the lengths of time between reversals or changes in the direction of its flow, is greatly varied. This and also the varied strength of the currents have prevented the utilization of atmospheric electrical energy in commercial quantities for the various purposes for which electricity is generally employed. In order to utilize such a current or currents, they should be transformed into a continuous direct current of uniform strength. The apparatus for accomplishing the transformation of atmospheric electricity into a direct current of uniform strength is susceptible of being greatly modified. The preferred form of apparatus, however, for carrying the invention practically into effect I will now proceed to describe to show that the method is capable of actual performance. Said apparatus is illustrated in the diagram accompanying this specification.

Referring specifically to said diagram, A represents a metallic pole, which is shown bare, but may be enveloped in suitable insulating material, if desirable. B is the base of the aforesaid pole, which base is of insulating material set in the ground E. C is a large metal plate beneath the said base, and has points or branches *c'* extending therefrom in different directions in the ground. C' is a metallic post extending from the said plate above the surface of the ground and having a terminal *d* of the circuit D. On the top of said pole is mounted a metallic cap F, consisting, preferably, of a convex disk provided with sharp iron or copper points *p*, which project in all directions from the same. As before mentioned, the said points may be plated with a suitable metal that forms a good conductor and prevents corrosion. Similar caps F' may be placed on other insulated poles, as A', in the vicinity, and connected with the main pole A by an electric conductor *a*, for increasing the effect. The pole A' may be of wood, and the pole A may also be of the same material if provided with a metal conductor within or on the outside, extending from the cap to the other terminal *d'* of the circuit D. The said circuit D leads from the terminal *d* through an automatic variable resistance G, thence to one of the poles of a secondary or storage battery H, and from the other pole of said battery through an automatic current-regulator to the terminal *d'*. An automatic current-reverser or pole-changer J is located in the said circuit for reversing the current whenever there is a change in its direction, so that it may be rectified or straightened during transit and caused to travel at all times whether its direction is toward or from the earth in one and the same direction through a portion of the circuit containing the secondary battery. The reversals are accomplished automatically by means of a pivoted polarized armature *b*, located between two electro-magnets *c c*, having their coils included in a shunt-circuit between the leads of the circuit D. The magnets are wound so that a north pole will be presented to the armature on one side and a south pole on the other. When the current is flowing in a certain direction, the said polarized armature will be repelled by one magnet and attracted by the other, and thereby moved to one side. When the current changes its direction through the magnets, the poles of the said magnets are reversed and the armature is both repelled and attracted to the other side. An arm *b'*, of diamagnetic material, is fixed to and extends from the armature and is moved by the same. Between a movable end of the arm and the said current-reverser is a pivoted connection or link *e*, by which the motion of the arm is conveyed to the reverser. Stops *f* are provided for limiting the movement of the said arm, and an adjustable resistance *d'* is included in the shunt *d''* to regulate the current through the same.

The current-regulator hereinbefore referred to prevents short-circuiting or the rapid discharge of the secondary battery into the air and ground when the strength of said battery-current becomes greater than that passing to the battery. Said regulator is composed of an electro-magnet I, having its coil in the circuit D. A pivoted polarized armature *g* is connected at its pivot to one terminal of the coil of said magnet, and when the current is flowing to the battery said armature is attracted by the magnet I and held in contact with the stop *h*, to which the terminal of the secondary battery is connected. In the aforesaid condition a free or low-resistance path for the current is provided to the battery; but when the battery-current exceeds the charging-current the magnet-poles are reversed and the armature is repelled by the magnet, and the free path of the circuit is broken between the armature *g* and stop *h*. In order to maintain the armature in the latter position until the charging-current has been increased in strength above that of the battery-current or discharging-current, and so that the said regulator will automatically operate, a shunt path K of high resistance is provided around the armature *g* and stop *h*. The high resistance of the said shunt is obtained by including a rheostat *m*. This shunt path or circuit K, with the resistance, permits a small but sufficient amount of current to flow through the magnet I to hold the armature away from the stop *h* until the current is reversed, and then move the armature back to stop *h* to close the free or low-resistance path. A spring *i* is provided to assist the movement of the armature from the magnet when it is repelled by the same.

The electric current may be directly conducted to translating devices—such as lamps or electromotors; but, as hereinbefore stated, said current is preferably employed to charge one or more cells of the secondary battery H, and this battery stores or accumulates the electrical energy and supplies the said translating devices. To illustrate the latter feature, leads or wires L and L′ are extended from the poles or electrodes of said battery, and translating devices, in the shape of incandescent lamps *l*, are connected with the said wires in multiple arc. In the wire L is a common circuit maker and breaker *n* to close and open the circuit to the lamps. The said battery may be charged in series or parallel.

The automatic variable resistance G maintains the current flowing through the circuit D approximately uniform by increasing the resistance therein upon an increase of strength. Said resistance is not absolutely necessary and may be dispensed with.

The lightning-arrester M is to short-circuit a very heavy current to prevent the same from passing through the other parts of the apparatus and injuring it. A low-resistance shunt O, with a circuit maker and breaker therein, is connected between the leads of the circuit D, near the terminals *d* and *d′*, to completely short-circuit the apparatus when desired.

Having described my invention, what I claim as new, and desire to secure by Letters Patent, is—

1. The method of utilizing atmospheric electrical energy, consisting in conducting the electric current or currents between the earth and a point in the atmosphere at an elevation above the earth, rectifying or straightening the currents during transit, and storing or accumulating the electrical energy of said rectified current or currents.

2. As preliminary steps in the method of utilizing atmospheric electrical energy, conducting the electric current or currents through a path of low resistance between the earth and the atmosphere at an elevation above the earth, maintaining an approximately uniform strength of current and rectifying or straightening the same during transit.

3. As preliminary steps in the method of utilizing atmospheric electrical energy, conducting the electric current or currents through a path of low resistance between the earth and the atmosphere at an elevation above the earth, and rectifying or straightening said currents during transit.

4. The method of utilizing atmospheric electrical energy or deriving energy from the difference of electrical potential existing between the earth and a point or points in the atmosphere at an elevation above the earth, consisting in conducting the current or currents between the two points through a path of low resistance, rectifying or straightening the said currents in a portion of the path during transit, and storing or accumulating the electrical energy thereof.

5. The method of utilizing atmospheric electrical energy or deriving energy from the difference of electrical potential existing between the earth and a point or points in the atmosphere at an elevation above the earth, consisting in conducting the current or currents between the two points through a path of low resistance, maintaining an approximately uniform strength of current, rectifying or straightening the said currents in a portion of the path during transit, and storing or accumulating the electrical energy.

6. The method of utilizing atmospheric electrical energy composed of a current of a varied alternating character, consisting in conducting the electric current or currents through a path of low resistance between the earth and the atmosphere at an elevation above the earth, and then accumulating the electrical energy of said current or currents while flowing in both directions in one or more cells of a secondary or storage battery, as described.

In testimony whereof I have hereunto signed my name this 29th day of April, 1889.

MARK W. DEWEY. [L. S.]

Witnesses:
C. H. DUELL,
C. L. BENDIXON.

home Forum Users Enter or register
Search messages Last messages
home > Forum > Free energy - discussions > The energy of nature

Friends, the "fever" around topics about alternative energy has made scammers excited too! Be vigilant and don't fall for cheap scams. Remember that if you are offered to buy a working BTG at a price less than a million dollars, then by 99% this is a scam for money. If you are offered to buy BTG for more than a million dollars, then this is a 100% scam. Alas, there are no miracles and exceptions yet, although you can always check it yourself... :-)

atmospheric electricity

one 2 Next >

NNN
Anti-Tenturian

hmm... here I came across such a scheme:

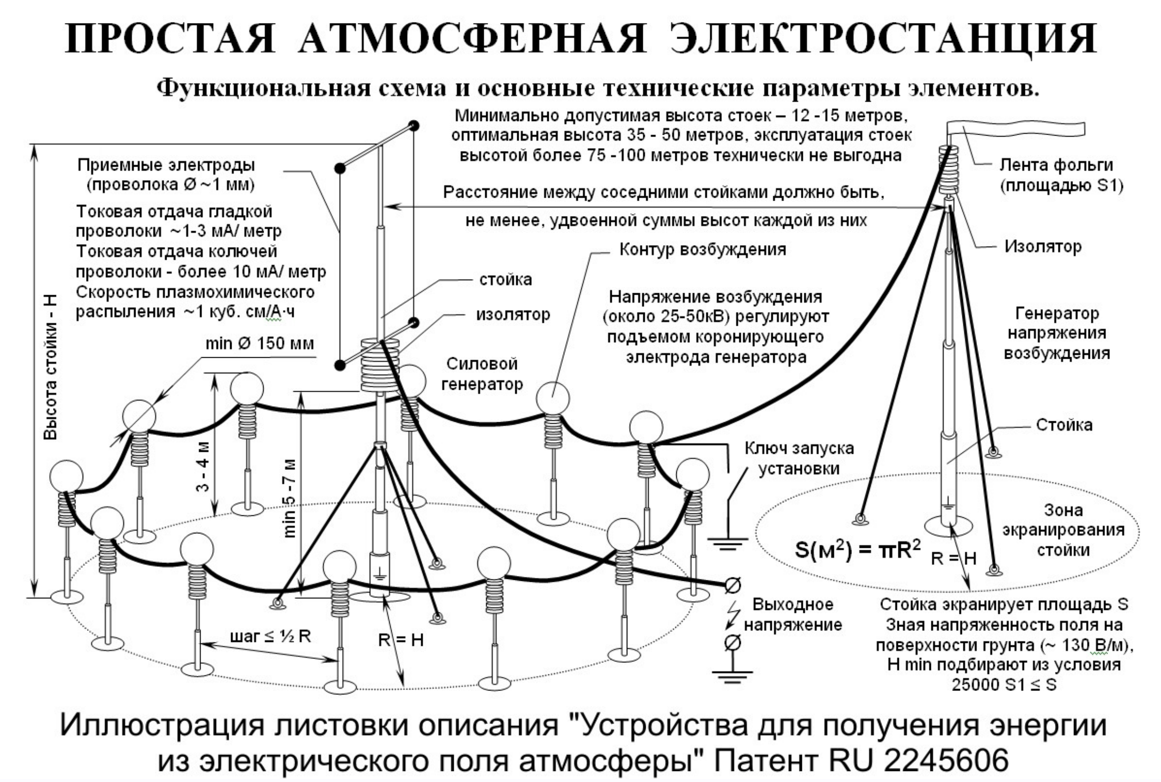

ПРОСТАЯ АТМОСФЕРНАЯ ЭЛЕКТРОСТАНЦИЯ

Функциональная схема и основные технические параметры элементов.

Минимально допустимая высота стоек – 12 -15 метров, оптимальная высота 35 - 50 метров, эксплуатация стоек высотой более 75 -100 метров технически не выгодна
Приемные электроды (проволока Ø ~1 мм)
Токовая отдача гладкой проволоки ~1-3 мА/ метр
Токовая отдача колючей проволоки - более 10 мА/ метр
Скорость плазмохимического распыления ~1 куб. см/А·ч
Лента фольги (площадью S1)
Расстояние между соседними стойками должно быть, не менее, удвоенной суммы высот каждой из них
Изолятор
Контур возбуждения
стойка
изолятор
Напряжение возбуждения (около 25-50кВ) регулируют подъемом коронирующего электрода генератора
Генератор напряжения возбуждения
Высота стойки - H
min Ø 150 мм
Силовой генератор
Стойка
Ключ запуска установки
3 - 4 м
min 5 -7 м
Зона экранирования стойки
S(м²) = πR²
R = H
Выходное напряжение
шаг ≤ ½ R
R = H
Стойка экранирует площадь S Зная напряженность поля на поверхности грунта (~ 130 В/м), H min подбирают из условия 25000 S1 ≤ S

Иллюстрация листовки описания "Устройства для получения энергии из электрического поля атмосферы" Патент RU 2245606

maybe it will be useful for someone for experiments...

---

here is a small excerpt from the "theoretical" description:

===

Does the Earth have its own electric field? Who remembers its characteristics?

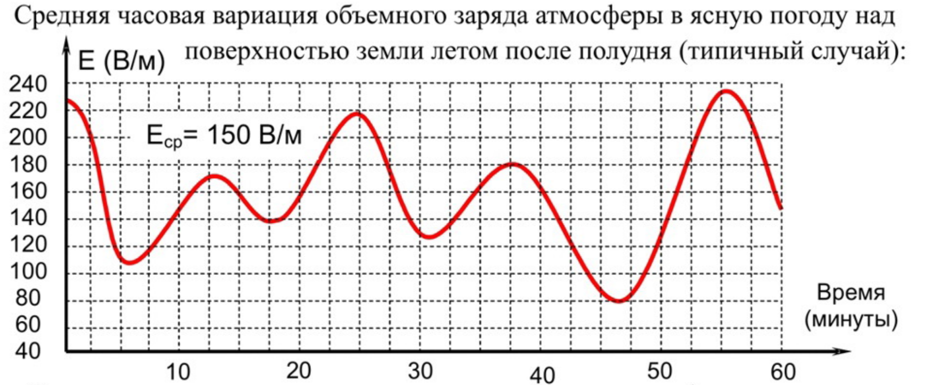
Средняя часовая вариация объемного заряда атмосферы в ясную погоду над поверхностью земли летом после полудня (типичный случай):

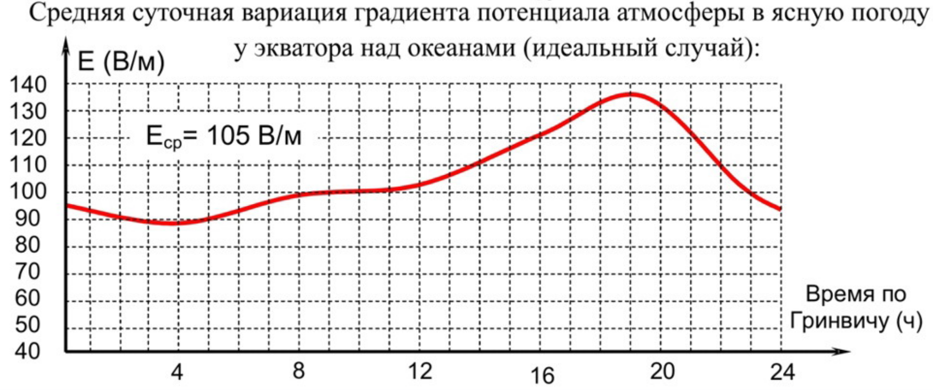
Средняя суточная вариация градиента потенциала атмосферы в ясную погоду у экватора над океанами (идеальный случай):

For absolutely clear weather, near the equator, the vertical electric field strength at the sea surface is ~ 130 volts per meter. That is, by raising your hands above your head, you will get between your fingers and boots, a voltage difference equal to the amplitude voltage in a conventional outlet, about 310 volts. Why doesn't anyone notice this? Firstly, the body is a conducting object and its potential almost coincides with the ground potential. Secondly, air is an excellent insulator and electric current flows through it at much higher intensities. About 30 kilovolts / cm, yeah ... And even in this case, the current through the air is indecently small, milliamps and fractions of a milliamp. Anyone who has been in the mountains and seen how the “fires of St. Elmo” appear on the fingers and points, in bad weather, is familiar with the extreme version of atmospheric electricity. And personally, such lights were shown to me at school, with the help of an electric car. You are standing on a dielectric stool, in a darkened classroom, someone is turning the handle of the unit, and you have your hair sticking up in all directions and blue corona discharges are burning on your outstretched arms. M-dya ... And if, on the same "electric stool" to put a girl with long hair - a spectacle outrageous. The field spreads her hair in a "sheaf" in all directions, each hair separately ... Screeching, screaming, sparks ... to put a girl with long hair is a sight to behold. The field spreads her hair in a "sheaf" in all directions, each hair separately ... Screeching, screaming, sparks ... to put a girl with long hair is a sight to behold. The field spreads her hair in a "sheaf" in all directions, each hair separately ... Screeching, screaming, sparks ...

The same effect can be obtained with very simple, almost home remedies. It is enough to have a high lightning rod at hand or launch a balloon or a kite into the sky, on a metal wire. The simplest arithmetic. If with each meter the voltage increases by 100-200 volts, then already a few tens of meters above the ground conditions arise for the breakdown of air by an electric discharge.

It can be calculated specifically. Do you know the lightning rod formula? Come on! Okay

... I understand that you would never need it in your life, therefore you didn't fill your head with nonsense ... Who knew that fortune would turn its back? Well, I remind you - "The lightning rod protects the site from atmospheric discharge (lightning), in the form of a circle circumscribed around its base, with a radius equal to its height." Why? Because the lightning rod is a voltage concentrator, and all the force lines of the atmospheric electric field, which were previously evenly distributed over this site, are now "pulled" to its top. There, on occasion, lightning will strike from a thundercloud. Towards the "leader" developing from the top of the lightning rod. I mean, first, a corona discharge will appear above its elevated part, and from there it will pull a chain of air ions, sufficient to break through lightning of the "linear" type. God, guys, why do you have such gloomy faces?

The question is, is it possible to obtain a similar effect (a gas discharge on a highly raised electrode) in clear, cloudless weather? And if so, what does it give us in a practical sense? After all, outside the discharge, the air will remain a dielectric. Then why? There is a reason and meaning, however... Where does the electric charge of the atmosphere come from? What feeds and sustains it, um... billions of years?

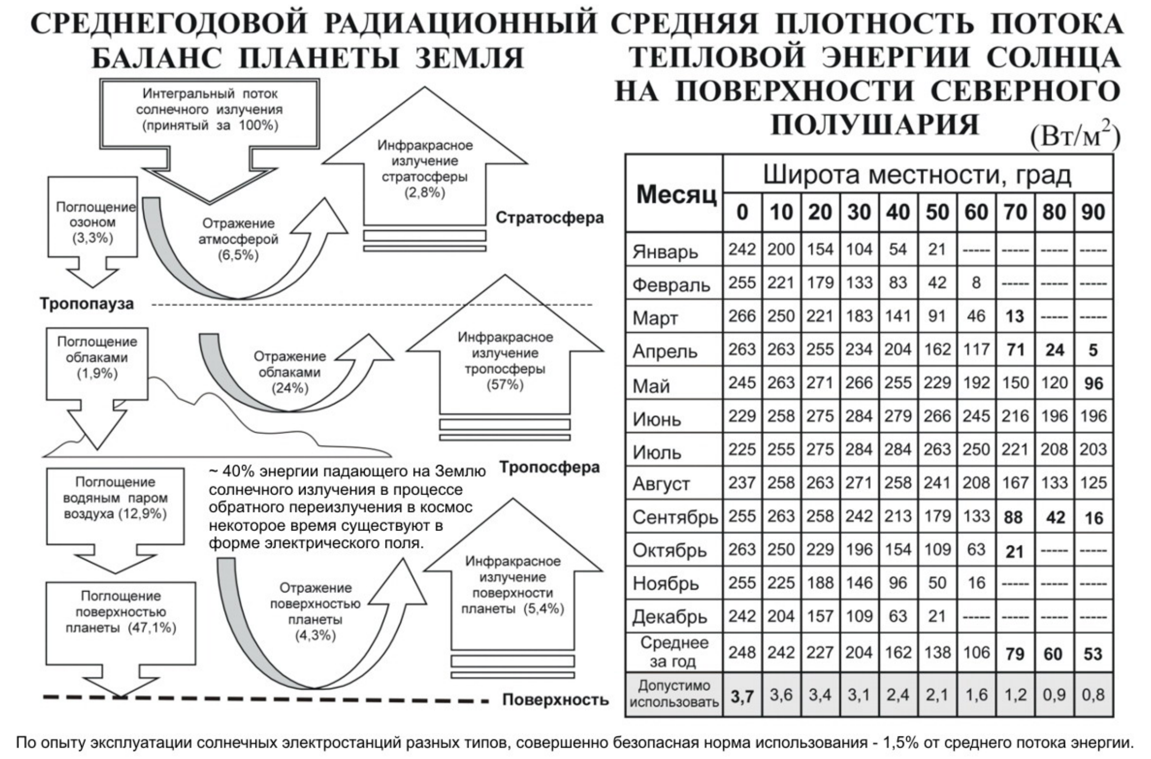
СРЕДНЕГОДОВОЙ РАДИАЦИОННЫЙ БАЛАНС ПЛАНЕТЫ ЗЕМЛЯ

Интегральный поток солнечного излучения (принятый за 100%)
Инфракрасное излучение стратосферы (2,8%)
Поглощение озоном (3,3%)
Отражение атмосферой (6,5%)
Стратосфера
Тропопауза
Поглощение облаками (1,9%)
Отражение облаками (24%)
Инфракрасное излучение тропосферы (57%)
Тропосфера
Поглощение водяным паром воздуха (12,9%)
~ 40% энергии падающего на Землю солнечного излучения в процессе обратного переизлучения в космос некоторое время существуют в форме электрического поля.
Инфракрасное излучение поверхности планеты (5,4%)
Поглощение поверхностью планеты (47,1%)
Отражение поверхностью планеты (4,3%)
Поверхность

СРЕДНЯЯ ПЛОТНОСТЬ ПОТОКА ТЕПЛОВОЙ ЭНЕРГИИ СОЛНЦА НА ПОВЕРХНОСТИ СЕВЕРНОГО ПОЛУШАРИЯ (Вт/м^2)

Месяц	Широта местности, град									
	0	10	20	30	40	50	60	70	80	90
Январь	242	200	154	104	54	21	-----	-----	-----	-----
Февраль	255	221	179	133	83	42	8	-----	-----	-----
Март	266	250	221	183	141	91	46	**13**	-----	-----
Апрель	263	263	255	234	204	162	117	**71**	**24**	**5**
Май	245	263	271	266	255	229	192	150	120	**96**
Июнь	229	258	275	284	279	266	245	216	196	196
Июль	225	255	275	284	284	263	250	221	208	203
Август	237	258	263	271	258	241	208	167	133	125
Сентябрь	255	263	258	242	213	179	133	**88**	**42**	**16**
Октябрь	263	250	229	196	154	109	63	**21**	-----	-----
Ноябрь	255	225	188	146	96	50	16	-----	-----	-----
Декабрь	242	204	157	109	63	21	-----	-----	-----	-----
Среднее за год	248	242	227	204	162	138	106	**79**	**60**	**53**
Допустимо использовать	**3,7**	3,6	3,4	3,1	2,4	2,1	1,6	1,2	0,9	0,8

По опыту эксплуатации солнечных электростанций разных типов, совершенно безопасная норма использования - 1,5% от среднего потока энергии.

Approximately 40% of the energy of the sunlight falling on the Earth, in the course of the return radiation into space, for some time (the section of convective heat exchange) is converted into the energy of a potential electric field. The charge of the tropopause and the stratosphere is produced by flows of heated air with an admixture of water vapor ascending from the surface of the planet. The natural leakage of this charge is carried out due to the ionization of air by cosmic rays, thunderstorms, rain and snow falling on mountain peaks. Have you ever wondered why high peaks are always covered with snow caps? There, the same electric field, day and night, pulls the charged ice dust hanging in the upper layers of the atmosphere.

However, let's start from afar, from space. The temperature of any celestial body is rigidly set by the integral power of external radiation absorbed by its surface and internal heat release from the depths. For the Earth, the average annual flux of sunlight ranges from 250-300 W/sq. meter at the equator up to 50-120 W/sq. meter in the polar regions. Cooling in a vacuum, of course, is purely radiative.

The sum of the energy fluxes reaching the surface is exactly equal to the energy flux radiated back into space. But, many planets of the solar system (in particular the Earth) have an atmosphere. They have only the stratosphere (from where thermal radiation goes into the world space) is close to the norm of the thermal balance (for the Earth, about -25 degrees C). On the surface of a celestial body surrounded by a gaseous shell, it is always much warmer. This thing is called the "greenhouse effect".

Any atmosphere is a spectral filter with multiple transparency windows. It can transmit light of a strictly defined range. Since the maximum of the return (thermal) radiation from the surface is always shifted, relative to the absorbed, to the long-wave region, then radiative cooling (like naked asteroids) is also impossible. Radiation from the surface does not diverge, heating the surrounding air. To remove thermal energy into space, a very productive mechanism with a material coolant is needed. In the dense part of the gas envelope of the planets (up to and including the stratosphere), heat transfer is almost entirely convective. The Earth's troposphere (this is below 12-17 km) contains 90% of the mass of air and 99% of atmospheric moisture. The heat flow "surface-space" there is due to the mechanical movement of air masses. This is an extremely rough model of the process. Now, its important detail.

The balance of radiative energy inflow and convective heat outflow requires a rapid vertical movement of the coolant. There is a sharp drop in temperature and pressure along the height. The thermal pressure between the surface of the planet and the upper boundary of the troposphere (altitude 12-17 km) is quite large (on Earth in the equator region from +45 degrees C at sea level to -70 degrees C in the tropopause). But, by itself, convective heat transfer still cannot cope with the load. Energy stupidly "gets stuck" in the lower layers of the atmosphere. Air is an extremely poor heat transfer medium. As a result, the main burden of energy transfer through dense layers of surface air usually lies with the much more productive effect of phase transitions, which acts in parallel with convection. The atmosphere is clearly stratified in height into layers that differ in composition, density and temperature. Moving from heat to cold and back, some of the components of air flows cyclically change their state of aggregation (evaporate, absorbing heat, and condense, giving it away). At the moment of condensation, in the upper layers of the atmosphere, each molecule emits a quantum of radiation leaving for the world space. On Earth, the "working body" of the described heat pump during phase transitions is water, on Venus - sulfuric acid, on Jupiter - ammonia.

In different regions of the planet, due to the evaporation of water from the surface and its re-evaporation in the clouds (water aerosol strongly absorbs infrared radiation), up to 10-55% of the integral solar energy flux is transferred from the troposphere to the stratosphere. The average content of water vapor in the Earth's atmosphere does not exceed 0.3-0.4%, but the energy consumption of its evaporation-condensation is huge and the total contribution of evaporation and condensation in the work of the "atmospheric heat pump" prevails. This ensures more than 95% of the heat exchange between the surface and the stratosphere. The process goes on around the clock. It is most intense in the tropical zone over the oceans, but is noticeable even over the eternal ice during the polar night. In middle latitudes, during the winter, up to 25-30% of the snowfall evaporates.

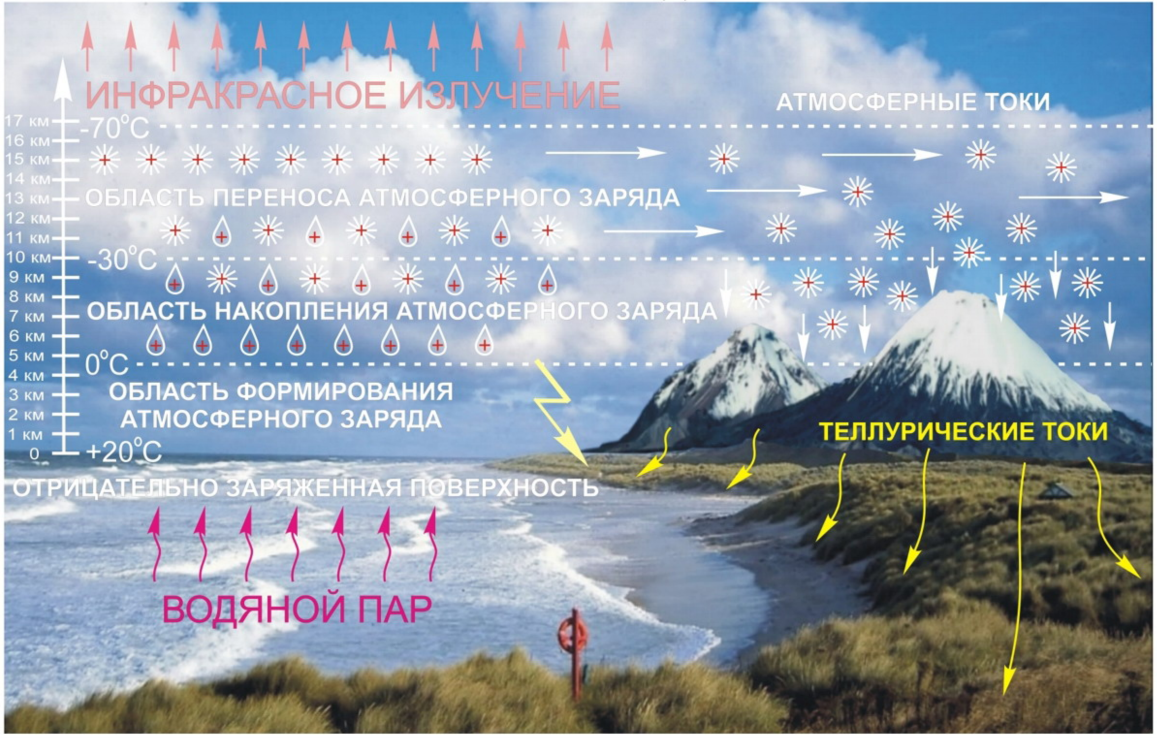
КРУГОВОРОТ ЭЛЕКТРИЧЕСКИХ ЗАРЯДОВ В АТМОСФЕРЕ ЗЕМЛИ

ИНФРАКРАСНОЕ ИЗЛУЧЕНИЕ
АТМОСФЕРНЫЕ ТОКИ
17 км
16 км
15 км
14 км
13 км
12 км
11 км
10 км
9 км
8 км
7 км
6 км
5 км
4 км
3 км
2 км
1 км
0
-70°C
ОБЛАСТЬ ПЕРЕНОСА АТМОСФЕРНОГО ЗАРЯДА
-30°C
ОБЛАСТЬ НАКОПЛЕНИЯ АТМОСФЕРНОГО ЗАРЯДА
0°C
ОБЛАСТЬ ФОРМИРОВАНИЯ АТМОСФЕРНОГО ЗАРЯДА
+20°C
ТЕЛЛУРИЧЕСКИЕ ТОКИ
ОТРИЦАТЕЛЬНО ЗАРЯЖЕННАЯ ПОВЕРХНОСТЬ
ВОДЯНОЙ ПАР

The work of the described atmospheric mechanism is accompanied by peculiar side effects. A change in the state of aggregation of a substance dramatically changes its dielectric constant. Spontaneous electrification always occurs at the phase boundary. Part of the thermal energy of the air components is converted into electricity... Condensation of water vapor in the cold upper layers of the gas envelope is accompanied by the accumulation of a large number of positively charged particles there. This is how space charges arise, which form a continuous layer in the upper layers of the Earth's atmosphere.

The uppermost layers of the atmosphere are highly ionized by cosmic radiation and represent a region of high conductivity. Result? Noticeable changes in the field strength over any point on the planet's surface are accompanied by a rapid redistribution of charges in the stratosphere and ionosphere. Therefore, the average intensity of the atmospheric field throughout the planet is almost stable.

The electric field of the lowest layers of the atmosphere, however, is very variable. It is associated with the smallest droplets of water and ice crystals suspended in the air. Free charge carriers (ions and electrons) are almost absent in the troposphere. This gives insulating properties to dense layers of air and prevents self-discharge of the aerosol mass. Due to the low mobility of condensate particles, volumetric electric charges (clouds, jets of fog, etc.) are strongly associated with carrier turbulent flows and move with them for a long time (often for weeks) until they evaporate, discharge onto mountain peaks or fall to the surface with precipitation. So "jet" electric currents in the stratosphere and mirror-like "telluric" currents in the upper layers of the earth's crust. Result?

A positive space charge of about 0.57 million coulombs constantly hangs in the atmosphere. It creates an electric field with the already mentioned average intensity of 130 V/m. The field pulsates in time with the rotation of the planet (its intensity is maximum at 17:00 GMT, when the Pacific Ocean floats on the sunny side). The average potential difference between the surface and the stratosphere is about 400 kV. This is in any good physics textbook.

The total energy resource of the charged atmosphere (~ 40% of the integral power of solar radiation in the earth's orbit) is estimated at about 2.5-5 per ten to the seventh power of gigawatts. It covers the Earth from pole to pole and is like a global DC distribution network connected to an eternal, free and environmentally friendly source of energy - the Sun.

We return to where we started. There is a stick. On a stick - an insulated wire to the very top. At the end of the wire is an air ionizer. Yes, it doesn’t matter what ... It doesn’t matter at all ... A corona discharge, a heated wire, an open flame, a piece of a radioactive isotope ... The main thing is that around this elevated conductor, let’s call it a “receiver”, an excess of ions. Free charge carriers. Yeah, I see it's starting to get...

Air is a moving substance. If the charge left the conductor and escaped into the atmosphere, then its fate is indifferent to us. The electric field, of course, will drag him to the right address. And since there is a flow of charges, then an electric current flows in the circuit. Well, yes, through the air ... The strength of this current depends only on the performance of the ionizer. How many charge carriers he throws out of the metal every second, so many pendants of electricity will flow through this conductor. What is air resistance? And it doesn't exist, practically. can be considered equal to zero. It doesn't matter how we connected directly to the stratosphere, all 400 kilovolts of celestial voltage hanging over our heads seem to be sitting at the end of a lightning rod...

Unclear? Oh, guys and girls, what are you all ... Okay, I'll explain about air resistance. On the contrary... Do you know what grounding is? Well, yes, an iron pin driven into the ground to drain current from the protected equipment. If anything gets there, everything will “close” to the ground and disappear safely in the ground. Now a question for backfilling. What is the earth resistance value? Well, yes, if we take a couple of identical grounds, measure the resistance between them and divide them in two, we will just get the desired value. So, how much? But no! A good ground should have resistance in the order of an ohm. Yes, sir. In any soil, even in dry sand or rock. And don't care about the resistance of the material. In-in ... The main thing is to ensure good contact with the volume, albeit of poorly conductive material. The earth is big. Consequently, the earth's resistance tends to zero, and all resistance losses - only at the point of contact with the earth. Where the charge carriers move densely enough. And all problems are solved. The cross-sectional area of the worst "conductor", when the current spreads in the volume, with the distance from the ground electrode, increases as the square of the distance. Exponentially... Very fast. And just as quickly, its electrical resistance drops. Out of habit - it seems like a paradox ... And just as quickly, its electrical resistance drops. Out of habit

- it seems like a paradox ... And just as quickly, its electrical resistance drops. Out of habit - it seems like a paradox ...

When an ionizer operates at the upper end of our "atmospheric generator", its resistance, relative to the atmosphere, also tends to zero. More precisely, it is negligible. Volumetric effect, as in the case of grounding in the ground. The strength of the current in the circuit depends only on the performance of the ionizer .. Convenient ...

No, we will not pile the burner on the upper end of the pantograph. And firecrackers - we will not. And we don't have a piece of a radioactive element. It's for the best, by the way... During discharge, in a gaseous environment, the electrodes lose weight. For every ampere-hour of charge flowing through the gas, about a cubic centimeter of electrode material is sprayed by the plasma. Its composition does not matter. There are no materials resistant to plasma chemistry ... At least put tungsten, even silicon, even graphite. Everything burns and oxidizes. Guessed? If you put a radioactive source on top, then atmospheric corrosion will quickly corrode it to holes and all the muck will be sprayed into the environment, settling around in a thin uniform layer. Do we need it? Live easier.

The expendable knob for the receiver of atmospheric current, on all lightning rods, put a simple, environmentally friendly iron. From it, except for rust, no harm. We, too, will not break away from the team ... How to make a banal piece of iron emit plasma ions into the air? There is such a way!

We consider the area of the soil, which is shielded from the atmospheric electric field by a pin, for example, twenty meters high. Yeah, this is the same pine tree that we appointed under the receiver stand. Pi multiplied by "er square", right? Here "er" is the radius of the platform and the height of the receiver. The result is 1256 square meters.

Now we find the area of the receiving electrode, under given conditions, sufficient to ignite a corona discharge on it in clear, cloudless weather. It's simple too. The field strength in the corona is about 30 kV/cm or 3 MV/m. The field strength over flat ground is 130 V/m. We consider the proportion ... Approximately 500 square centimeters of bare metal, a ball with a diameter of 12-13 centimeters, at the top, will already be surrounded by a weak cloud of ionized air molecules ... If you take a thinner rod, then the corona discharge around it will be proportionally stronger. Just?

Not just... The speed of movement of free ions in the air, at atmospheric pressure, is measured in centimeters per second. The top of the current collector will be enveloped by a sedentary cloud of lazily drifting ions, and the discharge will stabilize at the threshold of ignition. As it actually happens with any real lightning rod. Something, slowly, always leaks through it into the atmosphere, but this current is almost imperceptible. Its consequences can only be detected indirectly, by the unnatural smoothness of the metal surface. All the irregularities there seem to be polished. Really. The protrusions and burrs are the first to be "eaten up" by a quiet discharge. For a similar reason, grass is so evenly "trimmed" in all high-mountain "alpine meadows". The field of atmospheric electricity in the mountains is strong. On each blade of grass, sticking out a little higher than the rest, maybe, in bad weather, catch fire "

Conclusion? If you need a strong discharge, then you need a "point". Not simple, motionless, but dangling in the wind, blown by air pressure. Steel foil (ideally) or metallized paper, varnished from moisture. A long strip of such foil is now fluttering in the wind. Melting, melting, melting... Consumable! And the current adjustment is very simple - the position of the support. Tilt lower - the discharge field weakened. Raised higher - intensified. Need for inventions is cunning ...

Since the earth is negatively charged relative to the atmosphere, and the installation case is grounded, the voltage on the central electrode of the cement pipe is positive. Equal to the potential on the "receiver". When the process stops, this high-voltage part is closed to the housing with a grounded steel pin. Several tens of kilovolts are still present there. The simplest example of using an atmospheric generator "directly" as a ready-made power source for an "electrostatic precipitator". No matching devices for you, no converters. On one side of the chain, a corona discharge burns around an "electricity receiver" splashing in the wind, on the other, a corona discharge burns in a pipe through which dusty gas rushes from the reactor. From the central electrode (barbed wire) the ion flow pushes its way to the surrounding grounded walls. On the way, it charges cement particles and they drift there, after the discharge they settle on the metal. The pipe trembles as the cement slides down the spiral path into the hopper. And nothing else is needed. The "corona" sweeps the entire condensed fraction out of the gas like an invisible broom. It's very simple and clean, yeah... Adjustment of voltage and current - by tilting (lifting height) of the "receiver" bar. How much exactly the voltage in the circuit - no one knows. There were no people willing to measure. And why? You can see from the exhaust port of the pipe ... If the outgoing gas is clean of dust, everything works fine. As soon as the clubs appeared - lift the rack higher. The discharge does not cope ... And sparks of breakdowns began to click loudly in the iron pipe - lower it lower. On the way, it charges cement particles and they drift there, after the discharge they settle on the metal. The pipe trembles as the cement slides down the spiral path into the hopper. And nothing else is needed. The "corona" sweeps the entire condensed fraction out of the gas like an invisible broom. It's very simple and clean, yeah... Adjustment of voltage and current - by tilting (lifting height) of the "receiver" bar. How much exactly the voltage in the circuit - no one knows. There were no people willing to measure. And why? You can see from the exhaust port of the pipe ... If the outgoing gas is clean of dust, everything works fine. As soon as the clubs appeared - lift the rack higher. The discharge does not cope ... And sparks of breakdowns began to click loudly in the iron pipe - lower it lower. On the way, it charges cement particles and they drift there, after the discharge they settle on the metal. The pipe trembles as the cement slides down the spiral path into the hopper. And nothing else is needed. The "corona" sweeps the entire condensed fraction out of the gas like an invisible broom. It's very simple and clean, yeah... Adjustment of voltage and current - by tilting (lifting height) of the "receiver" bar. How much exactly the voltage in the circuit - no one knows. There were no people willing to measure. And why? You can see from the exhaust port of the pipe ... If the outgoing gas is clean of dust, everything works fine. As soon as the clubs appeared - lift the rack higher. The discharge does not cope ... And sparks of breakdowns began to click loudly in the iron pipe - lower it lower. the cement slides along a spiral path into the accumulator. And nothing else is needed. The "corona" sweeps the entire condensed fraction out of the gas like an invisible broom. It's very simple and clean, yeah... Adjustment of voltage and current - by tilting (lifting height) of the "receiver" bar. How much exactly the voltage in the circuit - no one knows. There were no people willing to measure. And why? You can see from the exhaust port of the pipe ... If the outgoing gas is clean of dust, everything works fine. As soon as the clubs appeared - lift the rack higher. The discharge does not cope ... And sparks of breakdowns began to click loudly in the iron pipe - lower it lower. the cement slides along a spiral path into the accumulator. And nothing else is needed. The "corona" sweeps the entire condensed fraction out of the gas like an invisible broom. It's very simple and clean, yeah... Adjustment of voltage and current - by tilting (lifting height) of the "receiver" bar. How much exactly the voltage in the circuit - no one knows. There were no people willing to measure. And why? You can see from the exhaust port of the pipe ... If the outgoing gas is clean of dust, everything works fine. As soon as the clubs appeared - lift the rack higher. The discharge does not cope ... And sparks of breakdowns began to click loudly in the iron pipe - lower it lower. How much exactly the voltage in the circuit - no one knows. There were no people willing to measure. And why? You can see from the exhaust port of the pipe ... If the outgoing gas is clean of dust, everything works fine. As soon as the clubs appeared - lift the rack higher. The discharge does not cope ... And sparks of breakdowns began to click loudly in the iron pipe - lower it lower. How much exactly the voltage in the circuit - no one knows. There were no people willing to measure. And why? You can see from the exhaust port of the pipe ... If the outgoing gas is clean of dust, everything works fine. As soon as the clubs appeared - lift the rack higher. The discharge does not cope ... And sparks of breakdowns began to click loudly in the iron pipe - lower it lower.

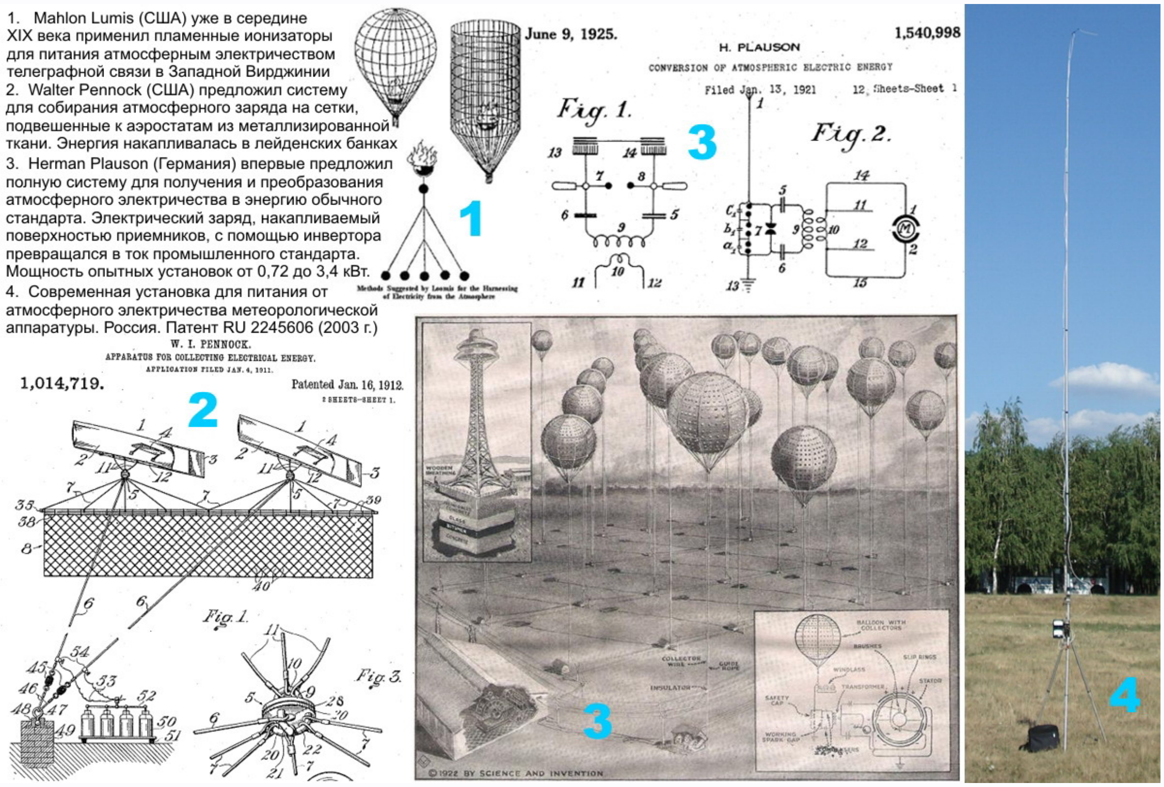

1. Mahlon Lumis (США) уже в середине XIX века применил пламенные ионизаторы для питания атмосферным электричеством телеграфной связи в Западной Вирджинии
2. Walter Pennock (США) предложил систему для собирания атмосферного заряда на сетки, подвешенные к аэростатам из металлизированной ткани. Энергия накапливалась в лейденских банках
3. Herman Plauson (Германия) впервые предложил полную систему для получения и преобразования атмосферного электричества в энергию обычного стандарта. Электрический заряд, накапливаемый поверхностью приемников, с помощью инвертора превращался в ток промышленного стандарта. Мощность опытных установок от 0,72 до 3,4 кВт.
4. Современная установка для питания от атмосферного электричества метеорологической аппаратуры. Россия. Патент RU 2245606 (2003 г.)

W. I. PENNOCK.
APPARATUS FOR COLLECTING ELECTRICAL ENERGY.
APPLICATION FILED JAN. 4, 1911.
1,014,719.
Patented Jan. 16, 1912.
2 SHEETS—SHEET 1.
Fig. 1.
Fig. 3.
June 9, 1925.
1,540,998
H. PLAUSON
CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY
Filed Jan. 13, 1921
12 Sheets-Sheet 1
Fig. 1.
Fig. 2.
Methods Suggested by Loomis for the Harnessing of Electricity from the Atmosphere
1
2
3
3
4
© 1922 BY SCIENCE AND INVENTION

NNN , 5 January 2013 #one

Klon likes this.

Contact
Patsak

I came up with another idea for generating electricity from ionized air.
Man has long been able to receive electricity directly from the air. When we breathe, we do it. If we run fast, we consume more energy, and therefore we breathe more intensively.
How to collect ions in the air and get electricity?
It is possible to drive a stream of air through a strong magnetic field. Positive and negative ions will scatter in different directions, separated by a magnetic field. They should fall on the plates to which the current collector will be connected. For example, a capacitor. Only I need a person on the forum who will help to calculate this installation. To know at what speed it is necessary to drive air through the pipe through a magnetic field, what strength the magnetic field should be, the area of the plates on which the ions fall. How much energy is contained in a cubic meter of ionized air?<br /><br />-- Apr 20, 2013, 23:32 --<br /><br />I tried to conduct the experiment I mentioned above - the effect is zero. Those. using a magnet, it was not possible to separate air ions.
I bought two plate magnets, fixed them with adhesive tape between two metal plates. I tried two plates of copper, aluminum and experience when one plate is made of copper and the second of zinc (as in galvanic cells). I connected a multimeter in voltmeter mode to the plates. He showed a potential difference of zero point, horseradish hundredths of milliamps. I tried to blow on the plates so that air passes between them and blow on them also with the help of a fan - the effect is the same. It turns out that ions in a magnetic field will simply rotate in a circle, as in this model . To force the ions to be attracted to metal plates, you need to apply an electric field of several kilovolts to them, but then no perpetuity will naturally work, because. we will spend more energy than we get.

The only thing that successfully came to my mind about the implemented projects of "vechnyaks" was to use horsepower. After all, horses also breathe air, eat free grass and water.

Who has any ideas how to make ions from the air separate into positive and negative, so that positive ones fly to one plate, and negative ones to another, so that a potential difference is created between the plates?

Attachments:

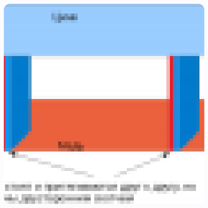

shema-opita.png
File size: 16.7 KB
Views: 439

Kontakter , March 31, 2013 #2

Good afternoon.
I decided to conduct some experiments for myself regarding atmospheric electricity.
First experiment:
- galvanized sheet s=0.3mm, 2 sq.m.
- height 25m,

Max Rud
Anti-Tenturian

- grounding - galvanized sheet 0.3mm, 0.5 sq.m.
- between them a wire through a voltmeter.
Result: 0.5V maximum.
Thinking that a brick building with iron elements distorts the electric field, we decided to inflate the balloon and go out into the open field.

Here is a video of experiment #2.
[media]https://www.youtube.com/watch?v=iFk3zOwr1wQ[/media]
nothing works yet....
Now we think that we need to introduce a capacitor into the circuit. Perhaps there is an equalization of potentials, which we cannot even notice.
While like this

Max Rud , 3 April 2014 #3

---

moderator
Administrator
Forum Team

Thanks for the material.
Shot perfectly, the work is done, but there are a couple of errors.
1. The resistance of a household voltmeter is very low, and the current obtained in this way is small, which is why you did not manage to measure the voltage!

Well, here's an example for you, charge a capacitor with a small capacity (up to ten microfarads) and measure its voltage with your voltmeter. You will see that the capacitor itself practically does not discharge, but as soon as you touch it with a voltmeter, the voltage across it immediately drops ...

Therefore, you also need to put a capacitor in the circuit and measure the voltage on it, but not constantly, but only occasionally touching with a voltmeter.

2. Take a steel broom instead of a frame and aluminum. that is, a bunch of wires sticking out like a panicle on one side and soldered or wrapped with wire on the other. So the efficiency will be many times higher.

moderator , 4 April 2014 #four

Rostislav
Administrator

> Max Rud said:
>
> *Thinking that a brick building with iron elements distorts the electric field, we decided to inflate the balloon and go out into the open field.*

What was used as a rope for the ball? The wire? In isolation?

Rostislav , 4 April 2014 #5

Thank you very much for your comment and attention.
I think now an active lightning rod should be used as an antenna.

Max Rud
Anti-Tenturian

It ionizes the air around it.
We will reach the voltage, but then there will be a question with the current strength ...
But we have assembled the circuit. But not tested due to extremely low voltage.
Let's move on.

> Rostislav said:
>
> *What was used as a rope for the ball? The wire? In isolation?*

As a cable, there was a stranded wire 2mm.kv.

Max Rud , 4 April 2014 #6

---

loginza22096
Anti-Tenturian

The topic is interesting. I will share my observations.
During a thunderstorm, I observed several times how a discharge of a couple of centimeters jumps between a disconnected antenna cable and a nail in the floor. This was before the army. We had strong dry thunderstorms then in Altai. There is lightning, but no precipitation.
Somehow, after the repair, I took the plasma home for testing. Installed and hooked up all right. In the evening it began to rain and a thunderstorm approached. I decided to be safe, the device is a client, you need to pull out the antenna. When trying to pull out the antenna, I received a millimeter discharge between the antenna and the antenna connector. (Plasma turned out to be alive🙂) I bring my finger to the antenna cable and get a discharge. Reddish in color, about a millimeter long. Played for about five minutes.
Later I decided to conduct an experiment to connect a voltmeter between the antenna and ground and take measurements. Waited until the storm approached again. All day before the thunderstorm it was roasting for 30. I went to the service. Time of measurements is 22-23 hours. I quickly bungled an antenna of 11 m of an antenna cord with a section of 3 mm, a decrease of 5 m of a 75 ohm cable (the cable braid was not connected either from above or from the bottom), grounding was 1 ohm. The height of the suspension top point is 6 meters (attached to the ridge of the building), the bottom point is 1.5 meters. In general, an impromptu inclined beam turned out.
The device used a switch, if memory serves ts4341. The selected mode is DC measurement, the measurement limit is 2.5 volts. Installed the device. I turned the head of the device to the exact "0". I'm sitting waiting. Readings "0" Thunderstorm is approaching readings "0", even the arrow does not twitch at the time of the lightning strike. As soon as it started to rain, the device came to life. The readings are very small, so I connected the dt890 in parallel. The measurement limit is 2 volts. The voltage between the antenna and ground is constant 0.02 volts. I didn't measure the current. The alternating voltage is not fixed by the devices. (Maybe not, or maybe the devices are not like that). The switch circuit during lightning discharges reacted but not more than 0.3 volts. Perhaps the inertia of the device is superimposed. The digital camera did not record flashes at all. The storm front proceeded from the place of measurements in 8-10 km. I counted the time after the lightning flash and multiplied by 300. Later this experiment was carried out in the daytime at about 14-15 hours. The storm passed in 3-5 km. The voltmeter reading is about twice as high. I did not write down the exact data (a working

H. C. VION.

Electric Apparatus.

No. 28,793.

Patented June 19, 1860.

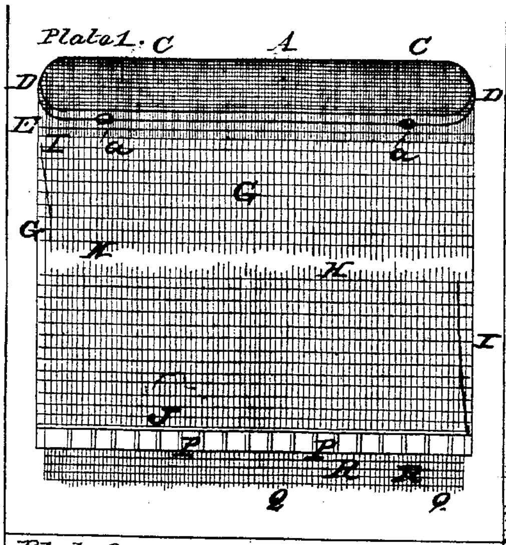
Plate 1.
C
A
C
D
D
E
I
a
a
G
G
K
K
I
J
P
P
R
R
Q
Q

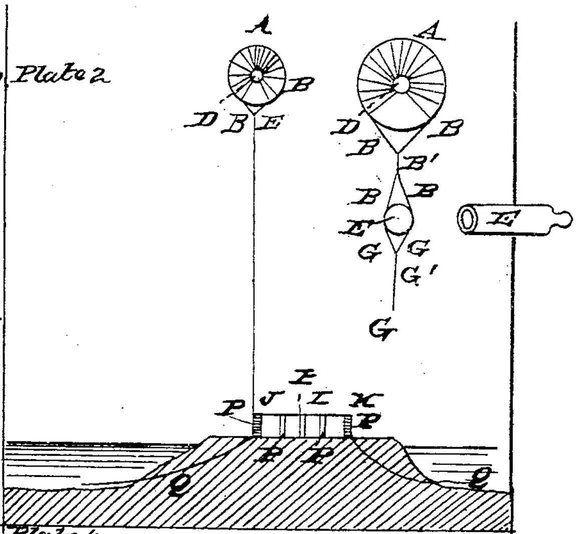
Plate 2.
A
B
D
B
E
A
D
B
B
B
B'
B
B
E
G
G
G'
G
E
I
P
J
I
K
P
P
P
Q
Q

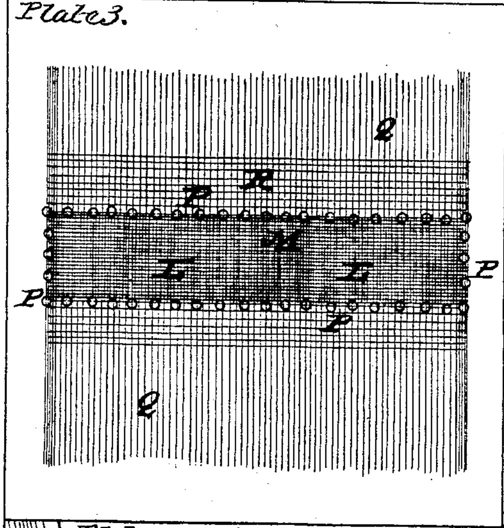
Plate 3.
Q
R
P
M
L
L
P
P
P
Q

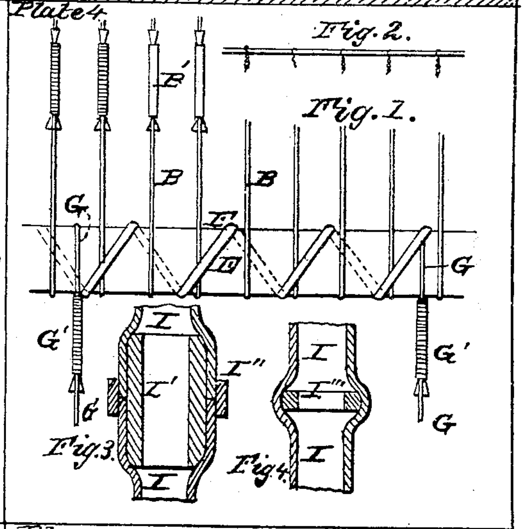
Plate 4.
Fig. 2.
E'
Fig. 1.
B
B
G
E
D
G
G'
I
I'
I''
I
I'''
G'
G
G
Fig. 3.
I
Fig. 4.
I

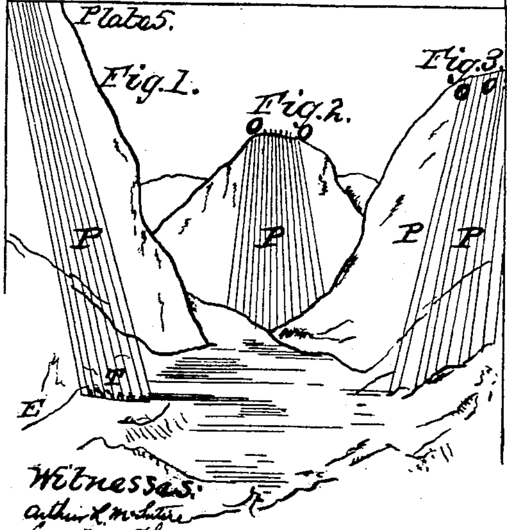
Plate 5.
Fig. 1.
Fig. 2.
Fig. 3.
O
O
O
O
P
P
P
P
P
E

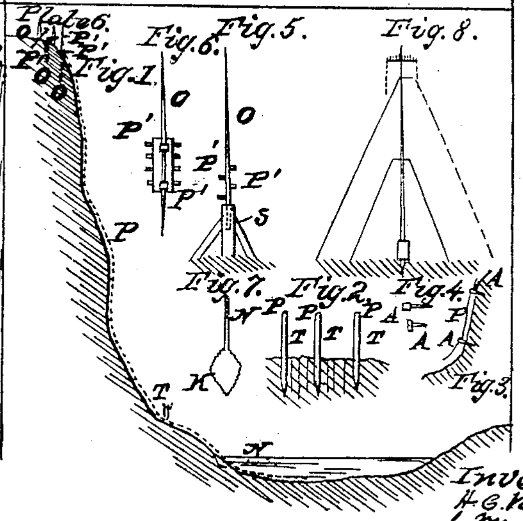
Plate 6.
Fig. 1.
Fig. 6.
Fig. 5.
Fig. 8.
O
O
P'
P'
P'
P
S
Fig. 7.
Fig. 2.
Fig. 4.
N
P
P
T
T
T
A
A
A
P
A
K
T
Fig. 3.
N

Witnesses:
Arthur L. McIntire
Lewis Bosworth

N. PETERS, Photo-Lithographer, Washington, D. C.

Inventor:
H. C. Vion
by Munn & Co.
Attorneys.

UNITED STATES PATENT OFFICE.

H. CHARLES VION, OF PARIS, FRANCE.

IMPROVED METHOD OF UTILIZING ATMOSPHERIC ELECTRICITY.

Specification forming part of Letters Patent No. **28,793**, dated June 19, 1860.

To all whom it may concern:

Be it known that I, HIPPOLYTE CHARLES VION, of Paris, in the Empire of France, engineer, have invented a new Mode of Obtaining Atmospheric Electricity and Terrestrial Electricity and its Industrial Applications; and I do hereby declare that the following is a full, clear, and exact description of the same, reference being had to the accompanying drawings, making part of this specification, in which—

Plate 1 represents a front elevation; Plate 2, a side elevation of an aerostat, in combination with certain other devices for obtaining atmospheric and terrestrial electricity; Plate 3, a plan of the device for obtaining terrestrial electricity; Plate 4, detached portions of the apparatus. Plate 5 shows a front view, and Plate 6 a vertical section and various details of the apparatus when applied in mountainous regions, the use of an aerostat being dispensed with.

The object of my invention is to form an electric pile of great power by using the positive electricity contained in the atmosphere, and the negative electricity contained in the earth, so as to make the electricity therein contained available for industrial purposes.

To enable others skilled in the art to make and use my invention, I will proceed to describe its construction and operation.

A is aerostat of a tubular form closed at both ends. It is made of suitable material, so as to be impervious to air. Its dimensions are such as to give it great ascensional power when filled with gas.

a a are valves in the surface of the aerostat, to be opened when the inflation of the aerostat should become too great.

A long india-rubber tube, I, communicates with the interior of the aerostat, being attached to the latter, near one end of its ends D, as seen in Plate 1, while the lower end of the tube I is attached to a gasometer, (not shown in the drawings.) The gasometer is to be fed with hydrogen gas, produced by the action of the pile itself, the negative wires of which (afterward to be described) enter a body of water at or near the base of the apparatus and decompose the water so as to produce the hydrogen gas. As the aerostat is supposed to be at a considerable height the tube I must be of corresponding length, and is constructed of a number of tubes, short wooden tubes I′ being inserted where the tubes I are joined, and a fastening-ring, I″, being slipped over each of the joints, as seen in Fig. 3 of Plate 4. At certain distances the tube I is fastened to the net-work of the positive wires (afterward to be described) in order to secure the tube against the action of the wind, and at each of these fastening-places a washer, I‴, is inserted in the tube in order that the tube shall not be compressed by the wire or other means employed to fasten it to the net-work of positive wires. (See Fig. 4, Plate 4.)

The aerostat is surrounded with a net-work of wires, one layer of the wires, C, being parallel with the axis of the aerostat and fastened to rings D at both ends of the aerostat, and the other layer of wires, B, extending partially around the aerostat at right angles to the wires C. One end of each of the wires B extends around an iron tube, E, some distance below the aerostat and meets the other end of it between the tube E and the aerostat. The two ends are fastened together by a ligature, B′. (See Fig. 1, Plate 4.) Each end of the tube E terminates into a ball, *e*. The wires B are fastened to the surface of the tube E by means of a helical wire, F, wound around the tube and across the wires B, as seen in Fig. 1, Plate 4. The upper ends of long vertical wires G are also wound around cylinder E, each wire G between two of the wires B, and the ends secured by a ligature, G′, as seen in Fig. 1, Plate 4. The helical wire F is also wound across the wires G, so as to keep them in their places on tube E.

The vertical wires G, which are to be the conductors of the positive electricity of the atmosphere, must be of a length proportionate to the desired efficacy of the electric pile, and the size and ascensional power of the aerostat must, of course, be adequate to sustain the weight of and keep suspended the wires G, (a weight still further increased by the horizontal cross-wires H, with which the vertical wires G are interlaced, in order to form a net-work not liable to be deranged by the action of the winds or similar influences.) The two outside wires, G, are stronger than the rest of them, and their lower ends are fastened to dynamometers of any suitable construction. These dynamometers are attached to the ends or a massive iron cylinder, J, and they serve to indicate the tension in the outside wires, G, and the corresponding ascensional power of the aerostat. According to the reading of these dynamometers the aerostat has to be supplied (through tube I) with more or less gas. The lower end of each of the wires G is wound around the cylinder J, and secured by ligatures similar to those above described. The wires G are all insulated (by a coating of gutta-percha or similar substance,) except where they are in contact with tube E and with cylinder J, and a similar insulating coating is laid on cylinder J, after the lower ends of the wires G have been fastened to it.

Another cylinder, K, similar to J, is placed at some distance from and parallel to cylinder J. It is connected with cylinder J by wires L, wound around both cylinders and interlaced with cross-wires M. The wire-work L M and cylinders are insulated (in a manner already described) against outside influences, so that the only electric communication between the two cylinders will be through the wires L. The two cylinders are placed upon insulated columns P. The cylinder K may be used as a substitute for cylinder J, and vice versa, whenever repairs become necessary. Insulated branch wire or wires are attached to the cylinders J K and wires L, so as to conduct the positive electricity obtained from the atmosphere by means of the above-described apparatus to wherever it is desired for industrial purposes. The insulated wires Q (interlaced with cross-wires R) are placed on the ground underneath and parallel to the positive wires L. Both ends of each of the wires Q are sunk into the earth or submerged in water, and fastened to a metallic plate coated with a metal not subject to oxidation. These wires Q are the conductors for the negative electricity of the earth, and a branch wire or wires attached to the wires Q serve to transmit the negative terrestrial electricity to wherever it is wanted for industrial or other purposes.

By uniting to the ends of the positive and the negative branch wire or wires a powerful electric current will be obtained, one pole of which is the atmosphere and the other the earth, and may be applied to any suitable useful purpose.

I will now proceed to describe the modification of the above-described apparatus when to be applied in mountainous countries.

P represents the positive electric copper or other metal wires coated over with an insulating substance. The upper ends of each of the positive wires is soldered to a prompter, O, at P′, Figs. 1, 5, and 6, Plate 6. The lower portion of each of the positive-wires is secured to an insulator, T, Fig. 2, Sheet 6. The positive wires are held above the ground by joints A, Figs. 3 and 4, Sheet 6, projecting from the soil at suitable distances from each other. The wires P are intended to follow the inequalities of the ground on which they are laid.

The prompters O, Figs. 5 and 6, Sheet 6, are iron rods sharpened to a point and silvered or coppered at their upper ends. The lower part of the prompter is fastened into a pole, S, covered with tar, which isolates the prompter and holds it in a firm position. A large metallic plate may be soldered to each prompter, as shown in Fig. 6. The positive wires may be soldered to the rod of each prompter or to the plate which is fastened thereto.

One or more branch lines, E, are soldered up to the positive wires to transmit the positive atmospheric electricity for which the wires P are the conductors to any desirable point.

N are negative iron or other metal conductors coated with an insulating substance. The upper ends of these wires rest on the ground near the positive insulators. The lower ends of these wires are soldered to a metallic plate or plates, V, Fig. 7, Plate 6, coated with a metal not subject to oxidation. The negative wires are sunk into the ground at very great depth or into wells, rivers, or into the sea. The negative electric branch wires are attached to the negative conductors N in the same manner as the positive branch wires are to the positive conductors. The branch wires and the soldering are coated over with an insulating substance. They are intended to carry the negative terrestial electricity to any desired point.

What I claim as my invention, and desire to secure by Letters Patent, is—

1. The peculiar arrangement of means herein specified, whereby I am enabled to use the positive electricity contained in the atmosphere and the negative electricity contained in the earth, and thus form an electric pile of considerable power and make the electricity therein contained available for industrial purposes, as set forth.

2. The combination of an aerostat and vertical wire-work with a tube, I, for admitting gas into the aerostat, in the manner and for the purposes above set forth.

CHARLES VION.

Witnesses:
HARRY W. SPENCER,
A. GUION, Jr.

US 20100220424A1

(19) **United States**
(12) **Patent Application Publication**
IBOK
(10) **Pub. No.: US 2010/0220424 A1**
(43) **Pub. Date: Sep. 2, 2010**

(54) **METHOD OF ATMOSPHERIC DISCHARGE ENERGY CONVERSION, STORAGE AND DISTRIBUTION**

(75) Inventor: **EFFIONG ETUKUDO IBOK**, SUNNYVALE, CA (US)

Correspondence Address:
THE TRAVIS BUSINESS GROUP, INC.
1005 E. HOMESTEAD RD.
SUNNYVALE, CA 94087 (US)

(73) Assignees: **EFFIONG ETUKUDO IBOK**, SUNNYVALE, CA (US); **THE TRAVIS BUSINESS GROUP, INC**, SUNNYVALE, CA (US)

(21) Appl. No.: **12/534,115**

(22) Filed: **Aug. 1, 2009**

Publication Classification

(51) **Int. Cl.**
H02H 3/22 (2006.01)
H02J 7/00 (2006.01)

(52) **U.S. Cl.** **361/117**; 320/166

(57) **ABSTRACT**

A method of converting atmospheric electrical discharge to a useable form of energy by arresting, storage and retransmission of lightning induced electrical discharge is disclosed. The invention discloses methods of deploying this technology even in isolated locations where no electricity infrastructure exists. Additionally, the potential for achieving in excess of 1 GWe of electrical power supply at costs orders of magnitude lower than fossil fuel or solar is also disclosed. Isolated collection units are disclosed. A method of deployment of these devices on individual cars and recharge stations is disclosed which in effect unplugs these automotive and recharge stations from the grid. This capability significantly decreases their dependence on the electricity grid infrastructure, enabling a much more enhanced rollout capability for the industry. The concept of Energy Dams is disclosed. These are facilities with substantial electrical energy storage capacities with the capability of receiving energy feeds from various generation sources.

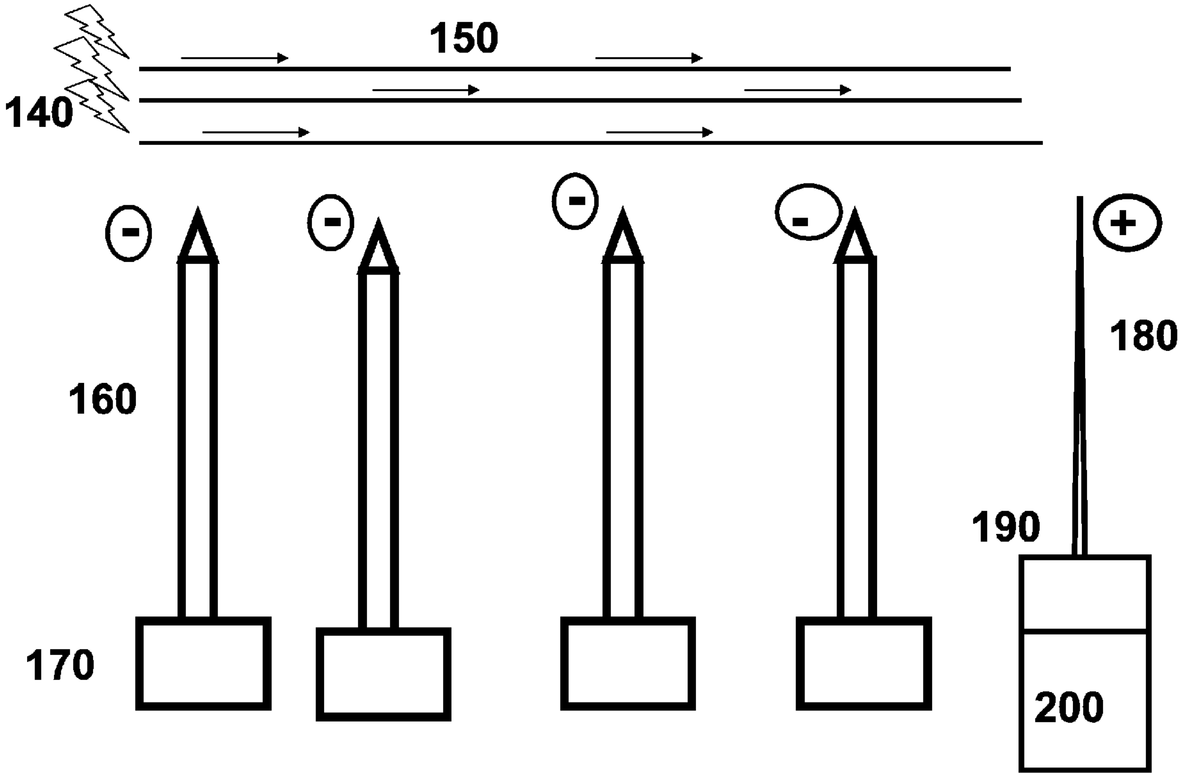

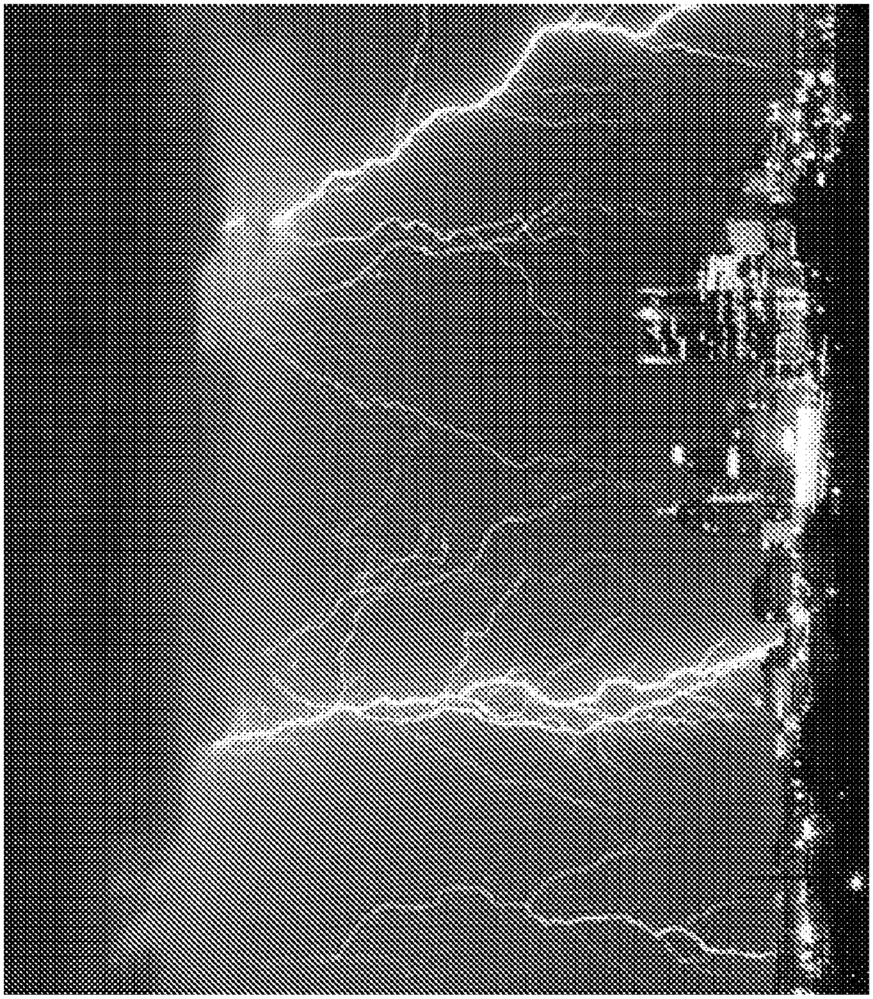

Fig. 1a

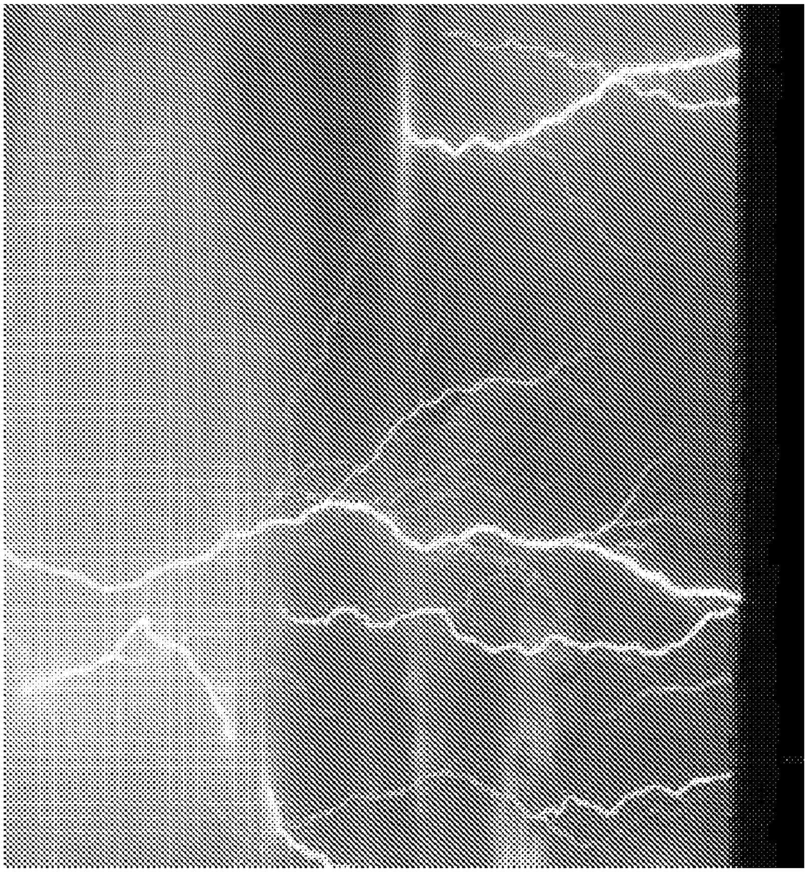

Fig. 1b

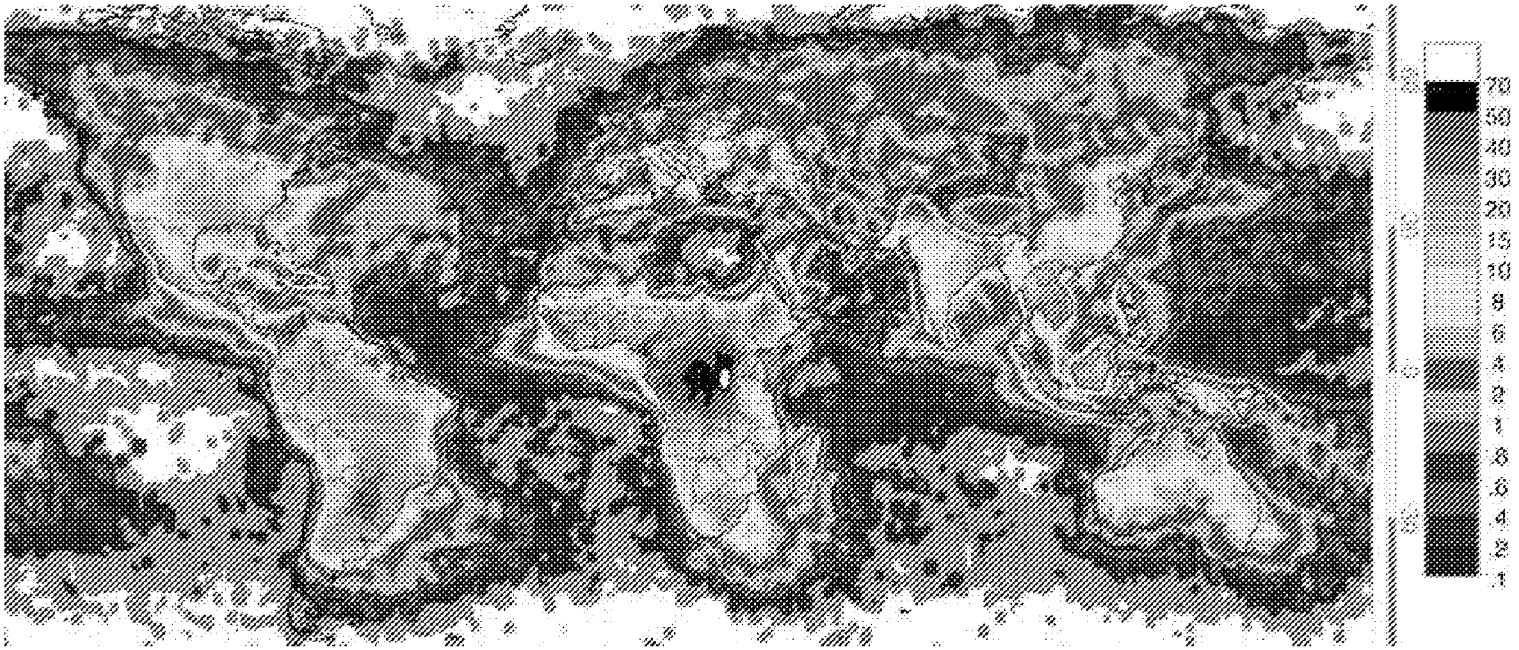
70
50
40
30
20
15
10
8
6
4
2
1
.8
.6
.4
.2
.1

Global Distribution of Lightning Strikes: April 1995 – February, 2003. (Source: NASA)

Fig. 1c

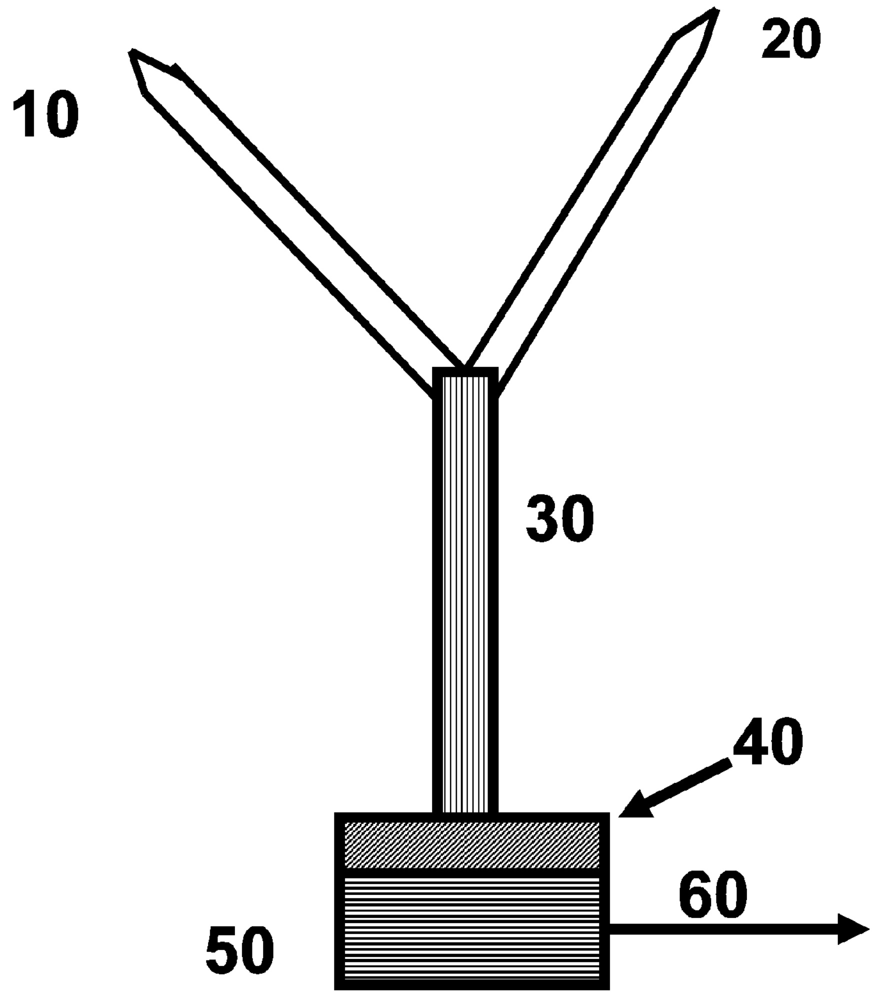
10
20
30
40
60
50

Fig. 2

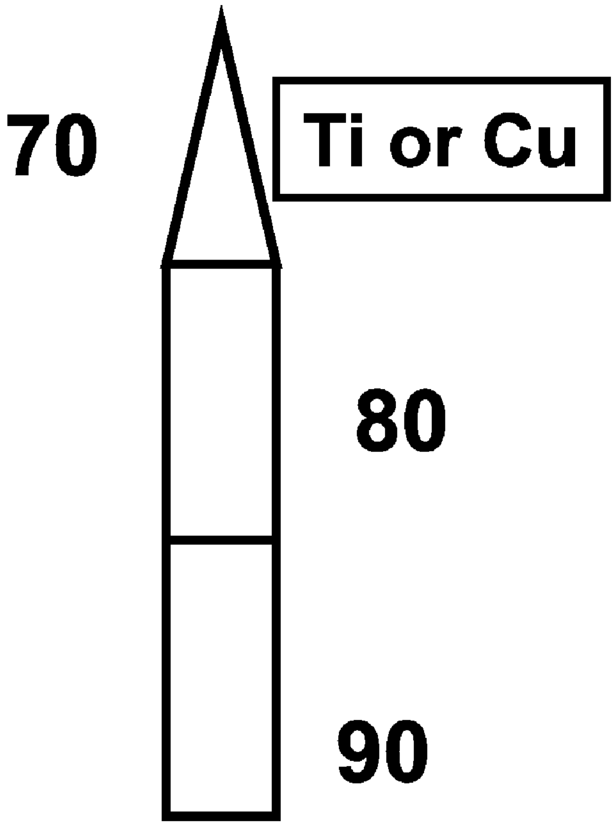
70
Ti or Cu
80
90

Fig. 3

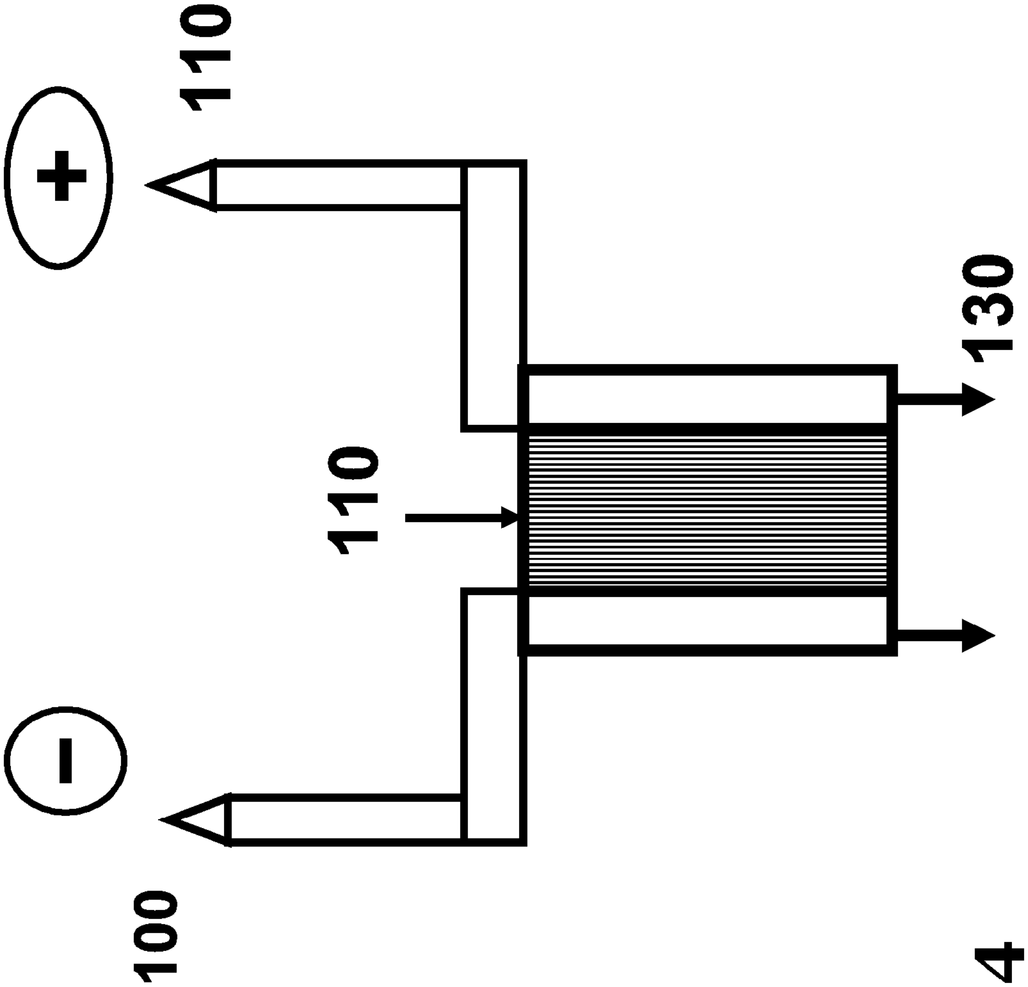
110
+
110
130
−
100

Fig. 4

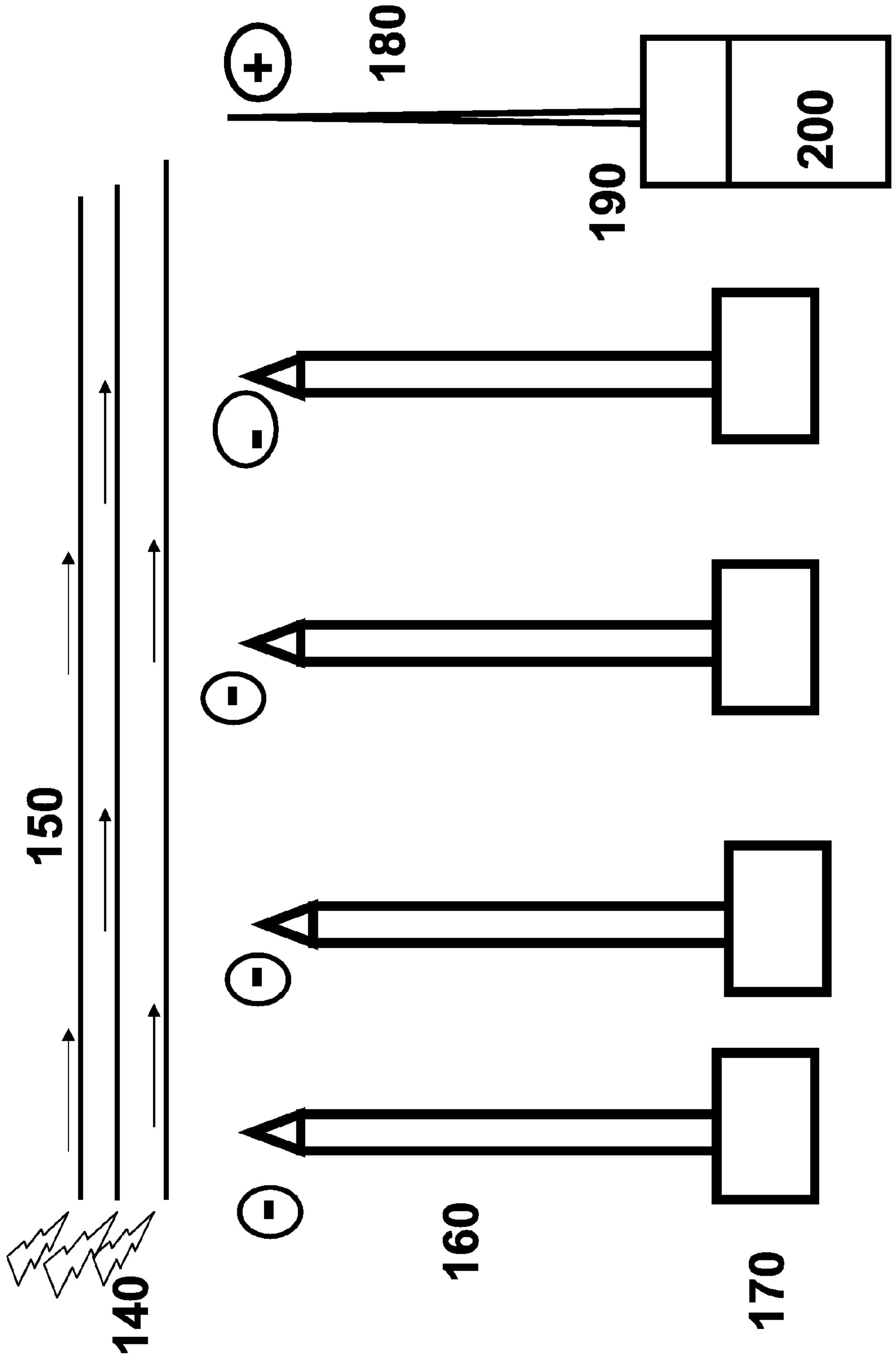
140
150
160
170
180
190
200
+
-
-
-
-

Fig. 5

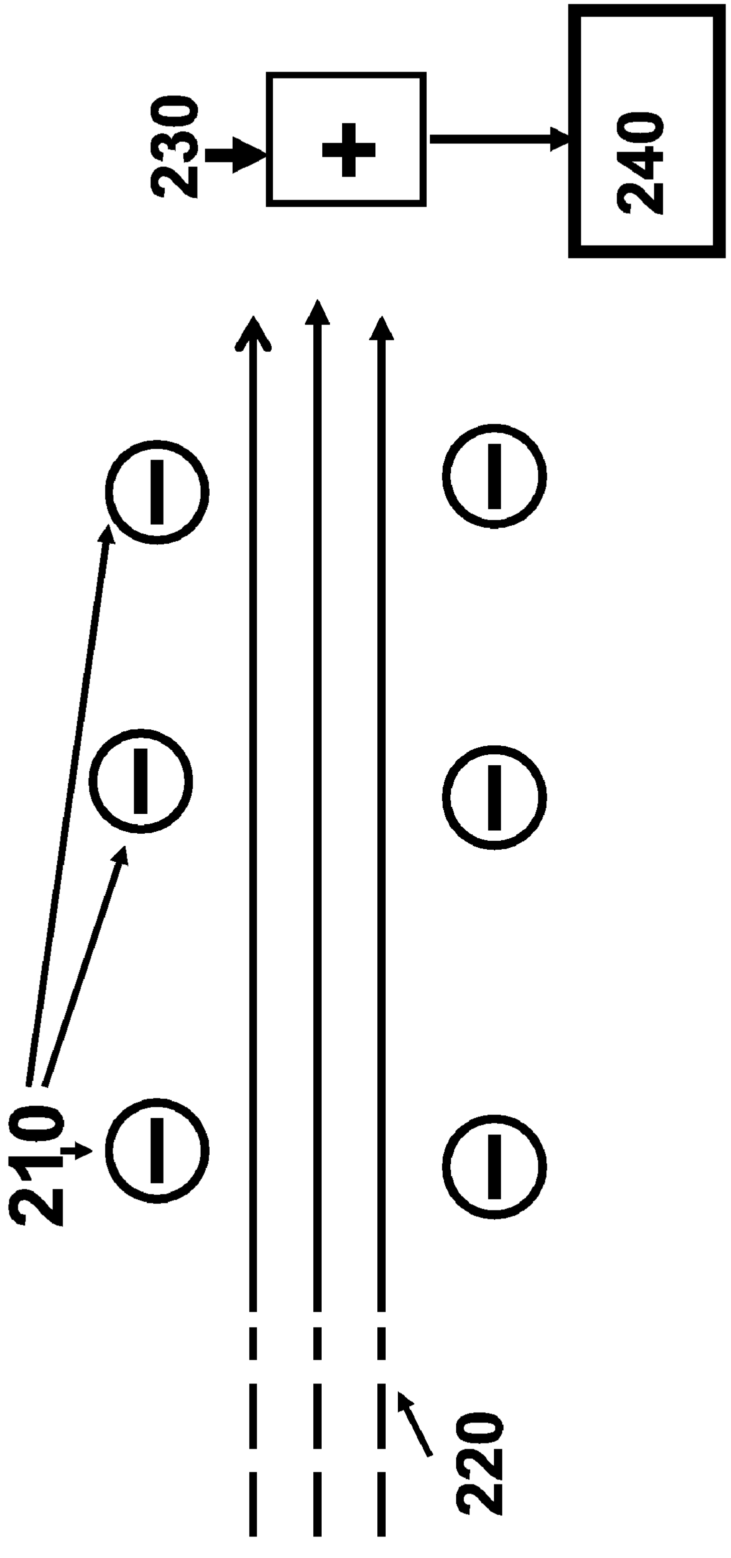
230
+
240
210
220

Fig. 6

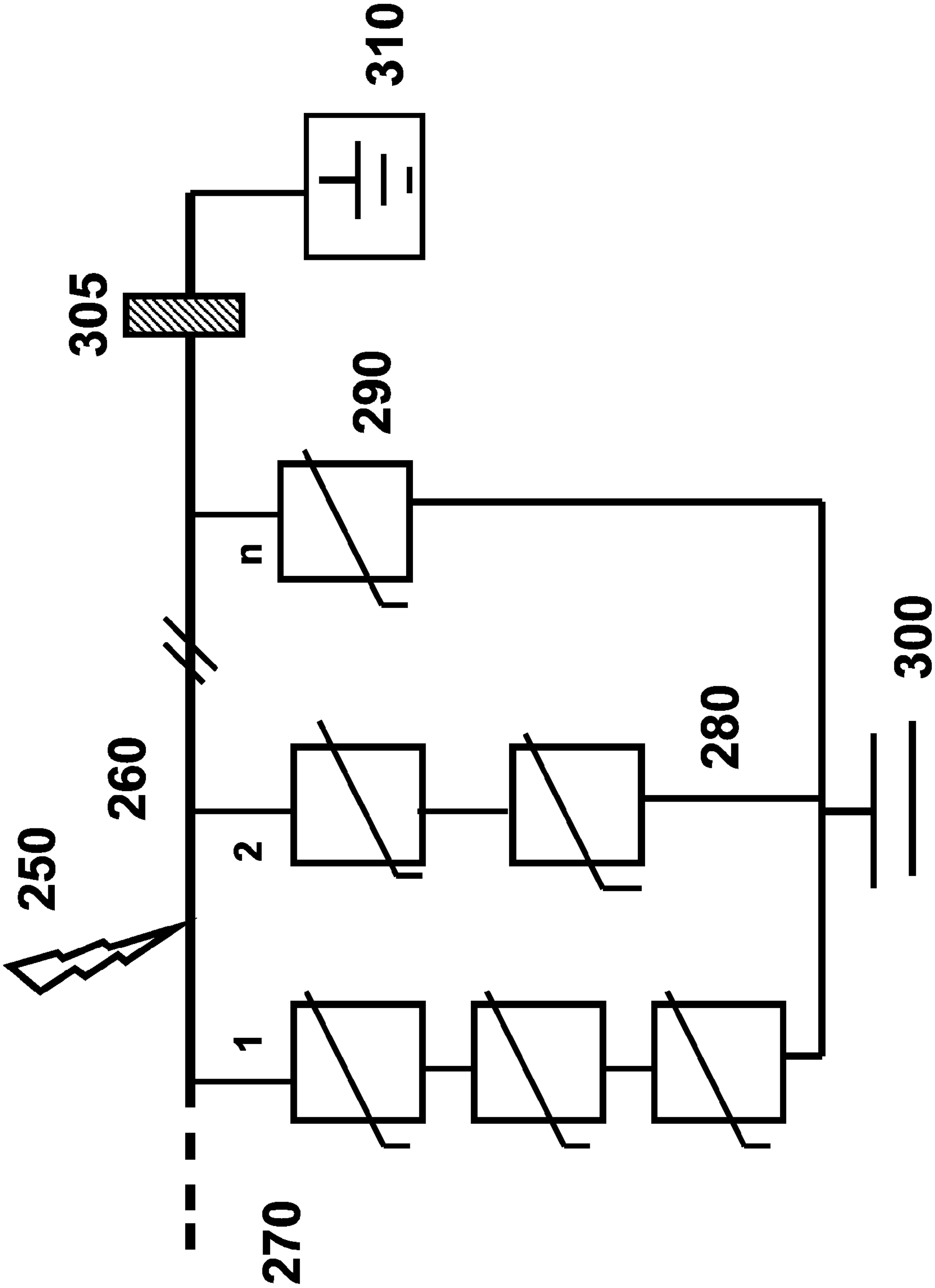
250
260
270
280
290
300
305
310
1
2
n

Fig. 7a

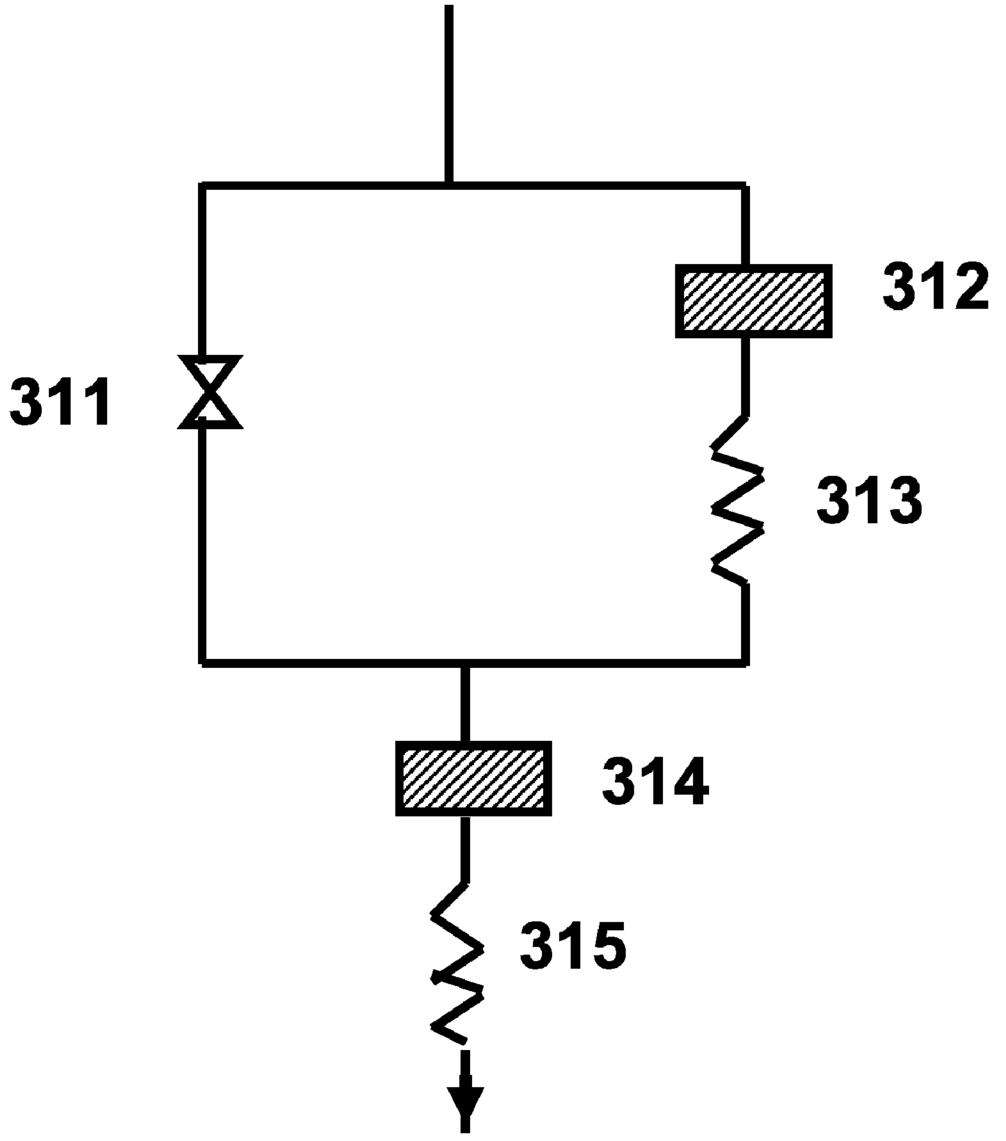
311
312
313
314
315

Fig. 7b

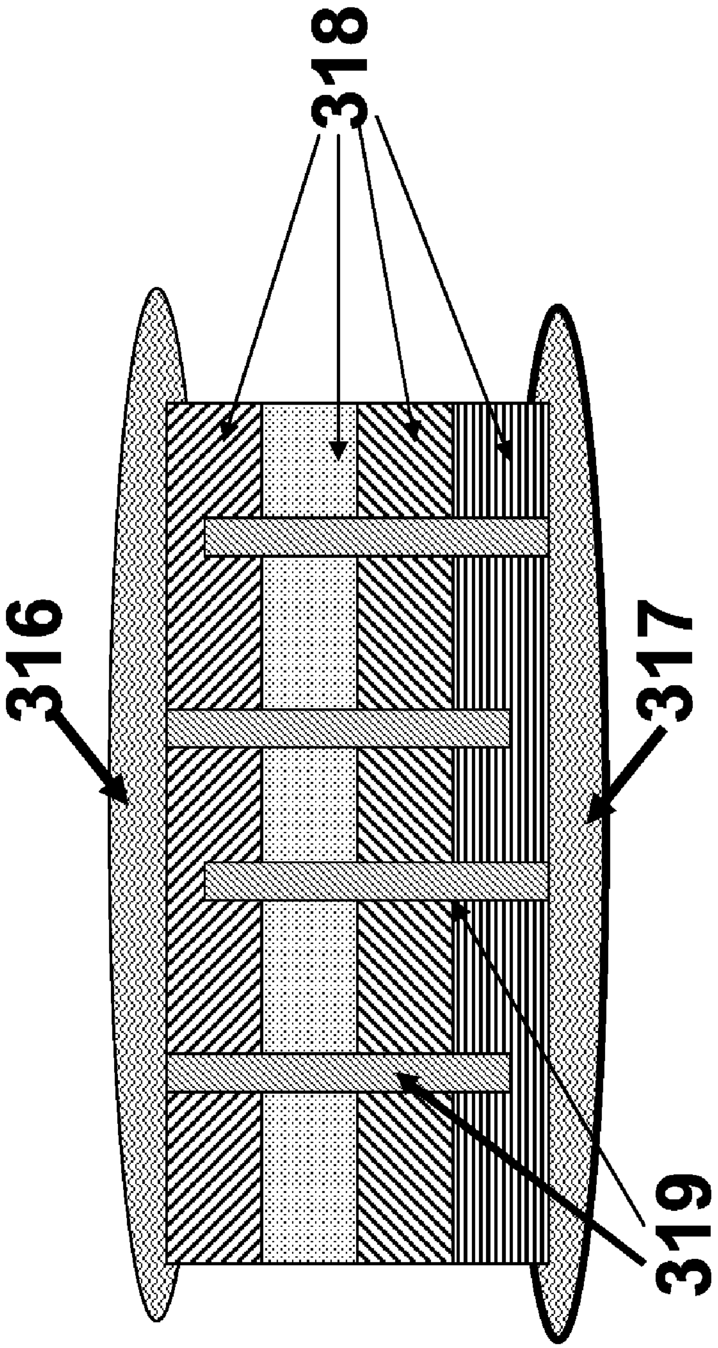
318
316
317
319

Fig 7c

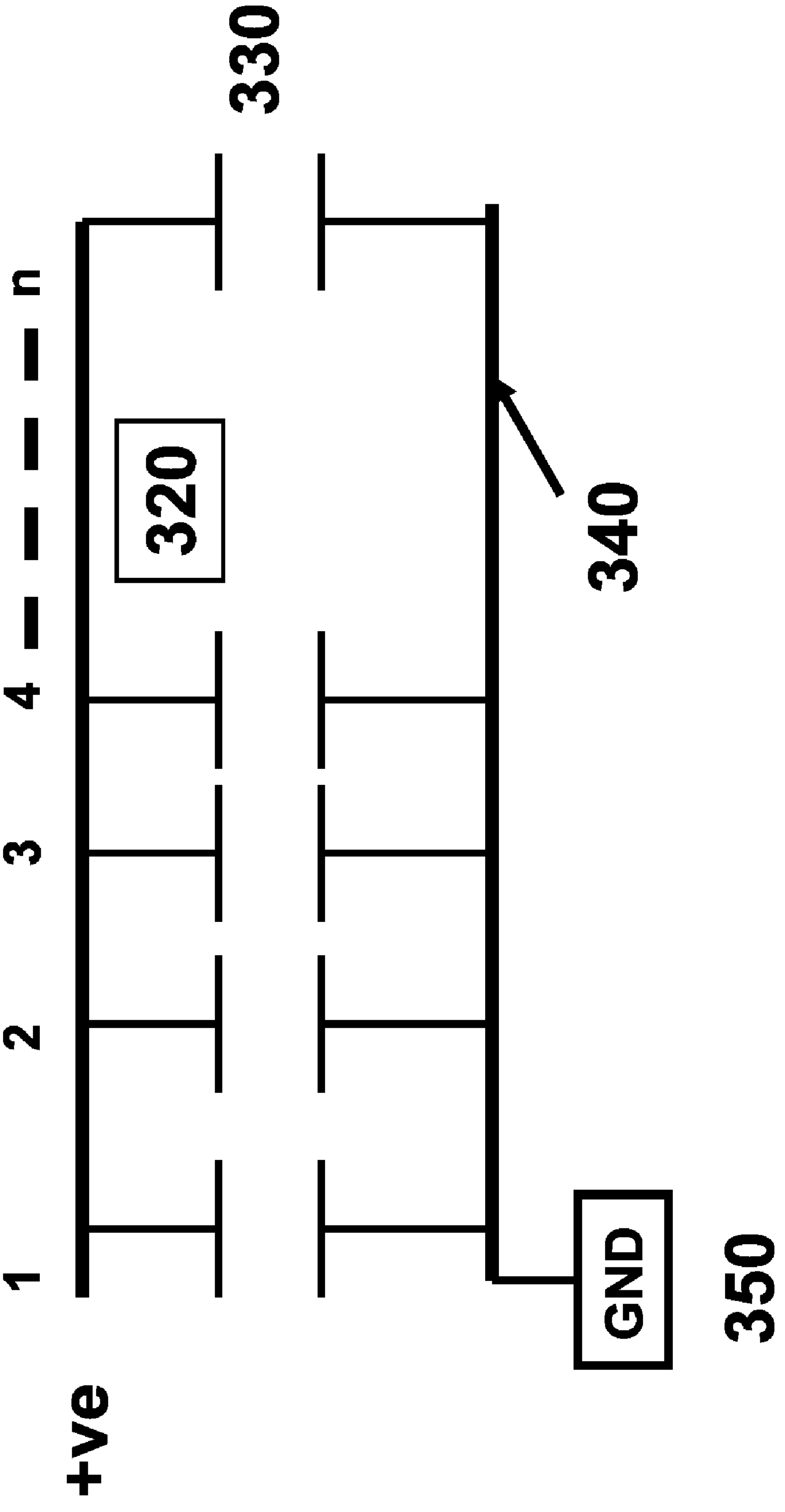
+ve
1
2
3
4
n
320
330
340
GND
350

Fig. 8a

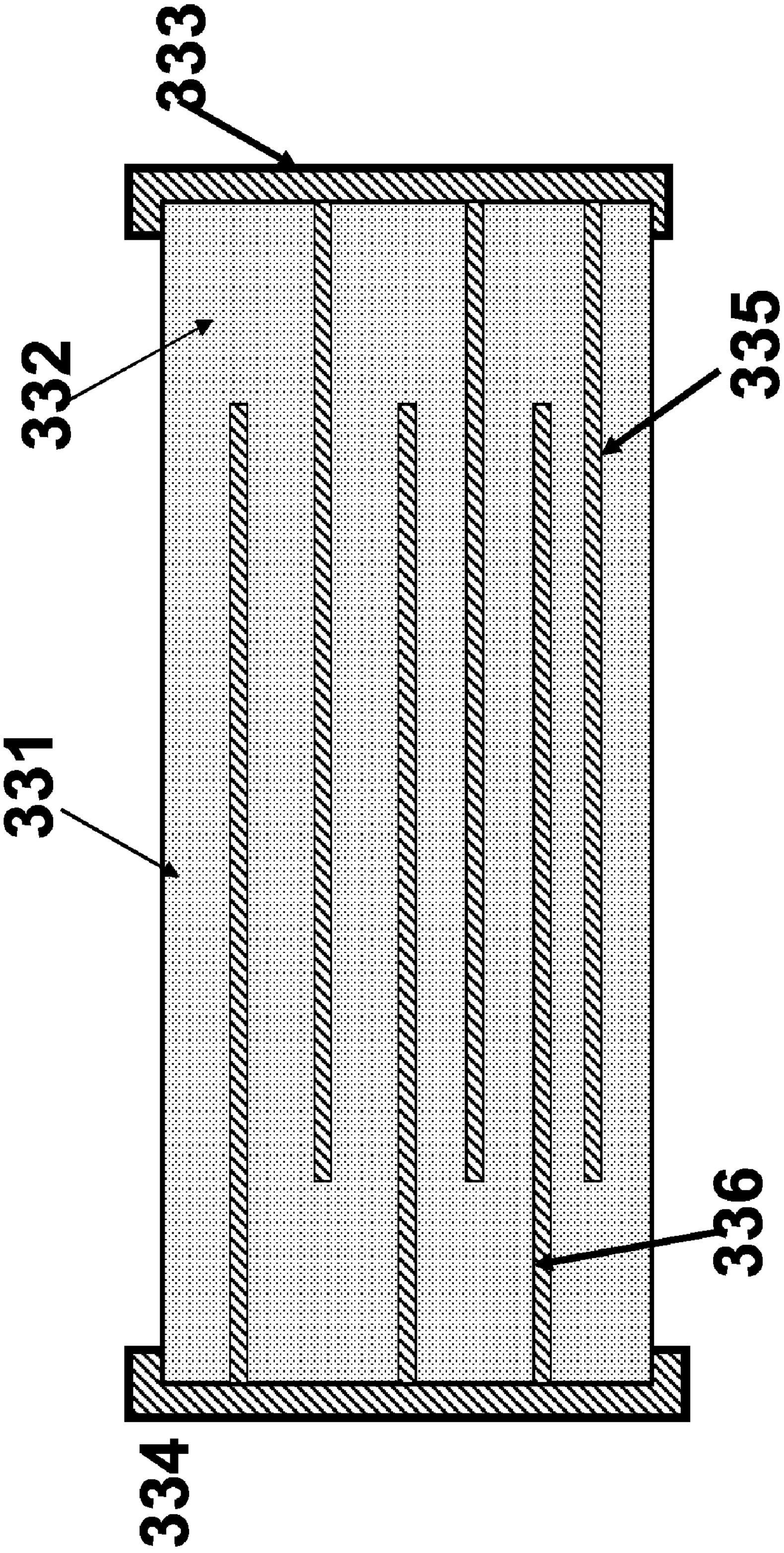
333
332
335
331
336
334

Fig. 8b

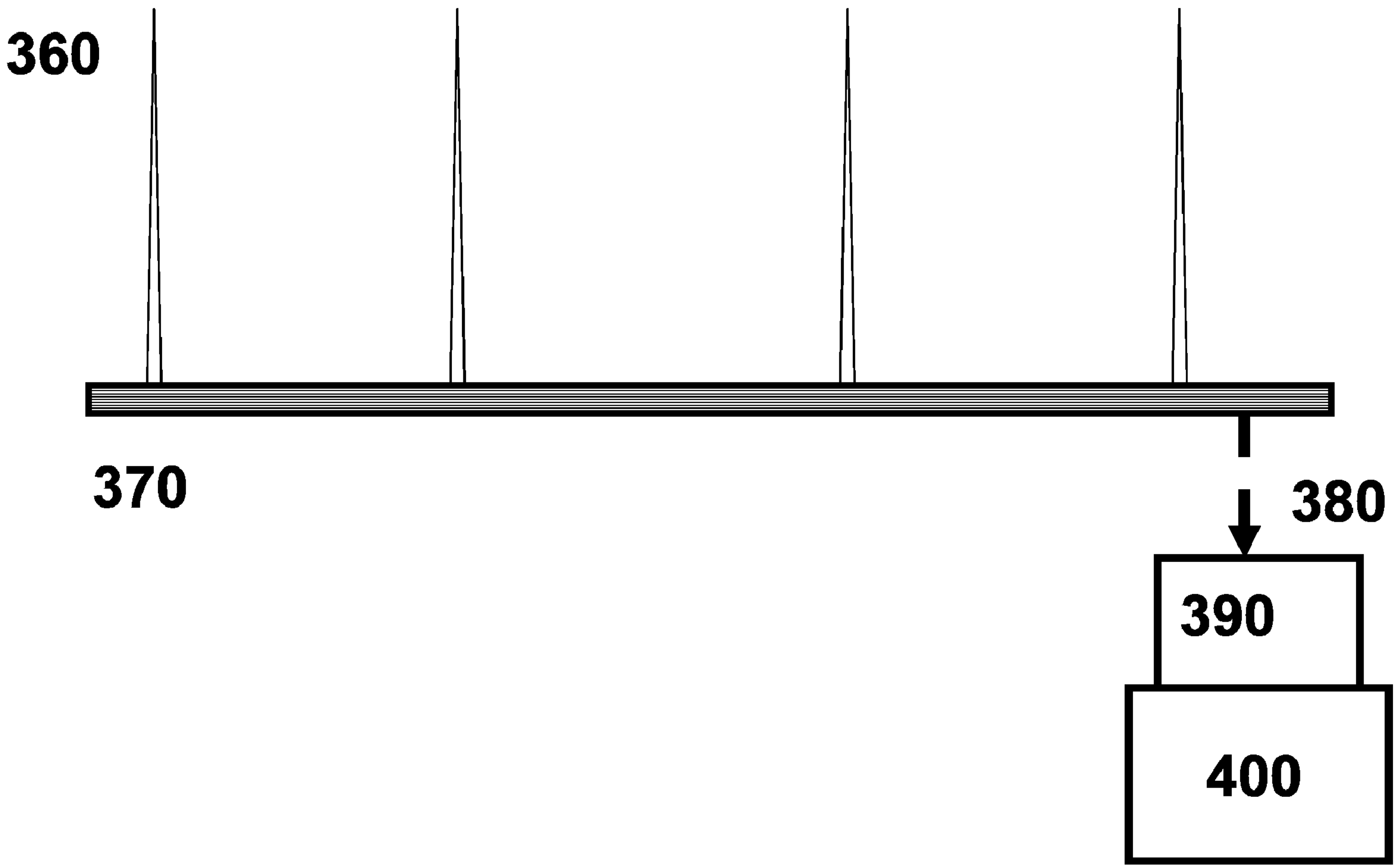
360
370
380
390
400

Fig. 9

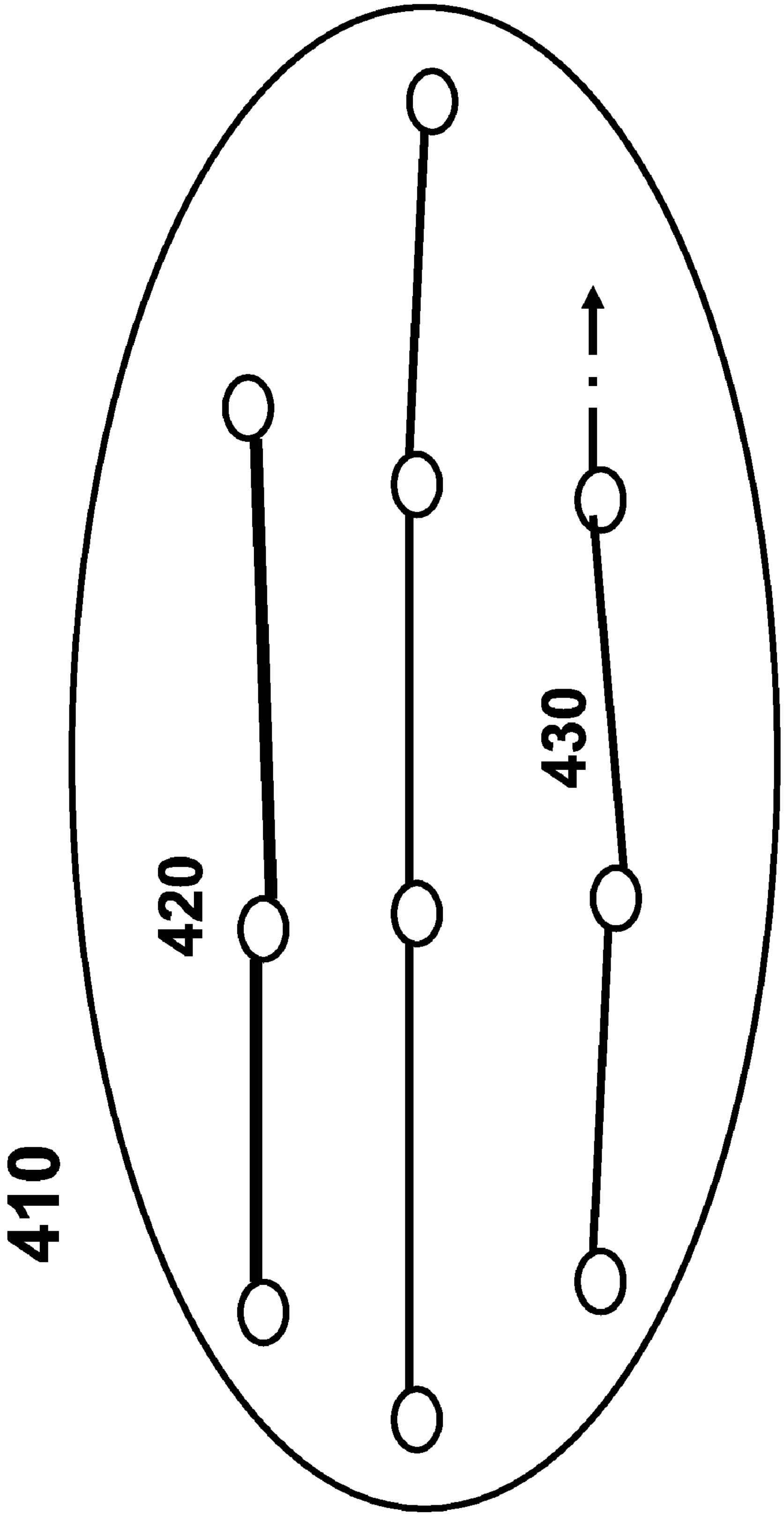

Fig. 10

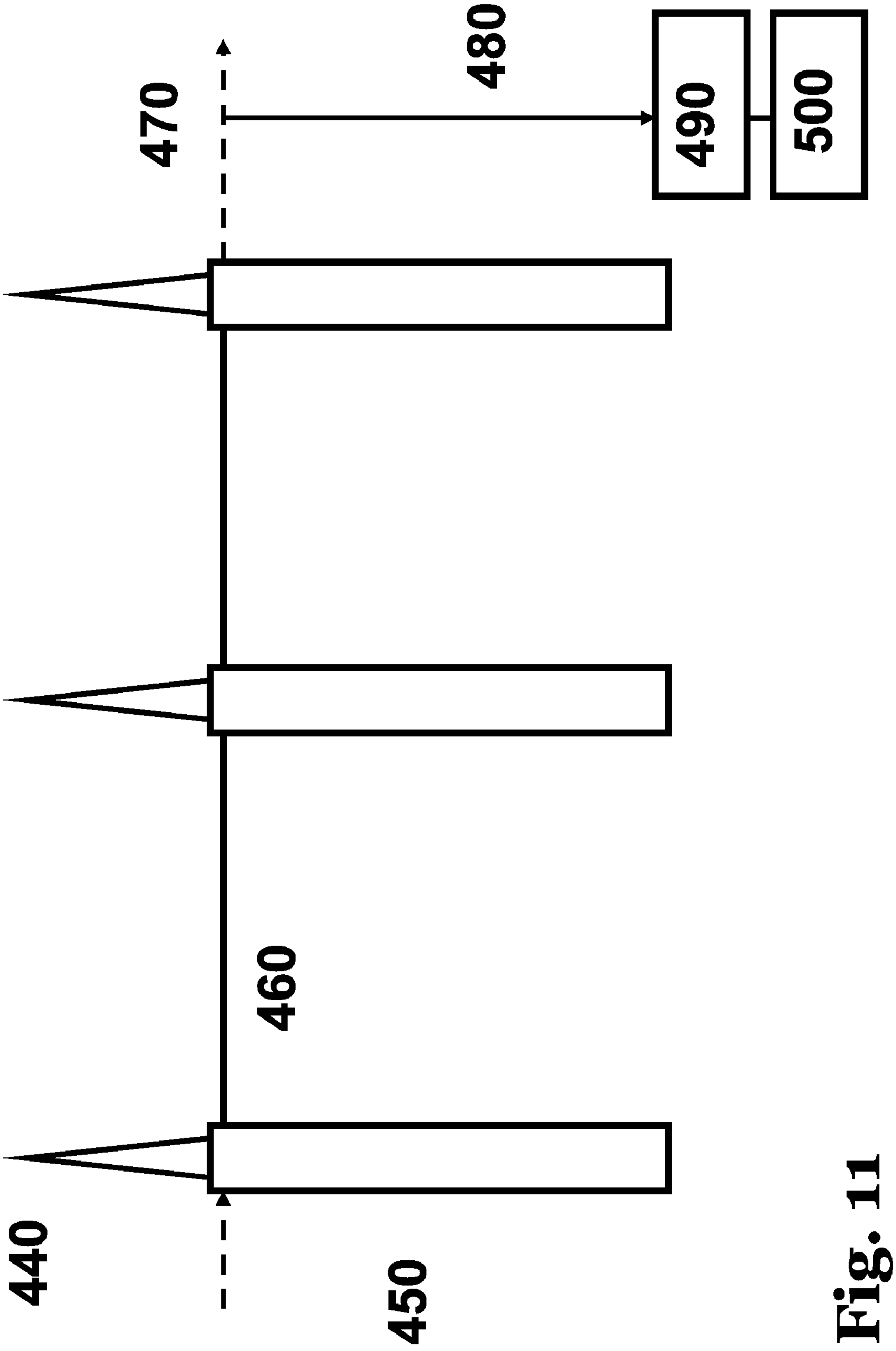
440
450
460
470
480
490
500

Fig. 11

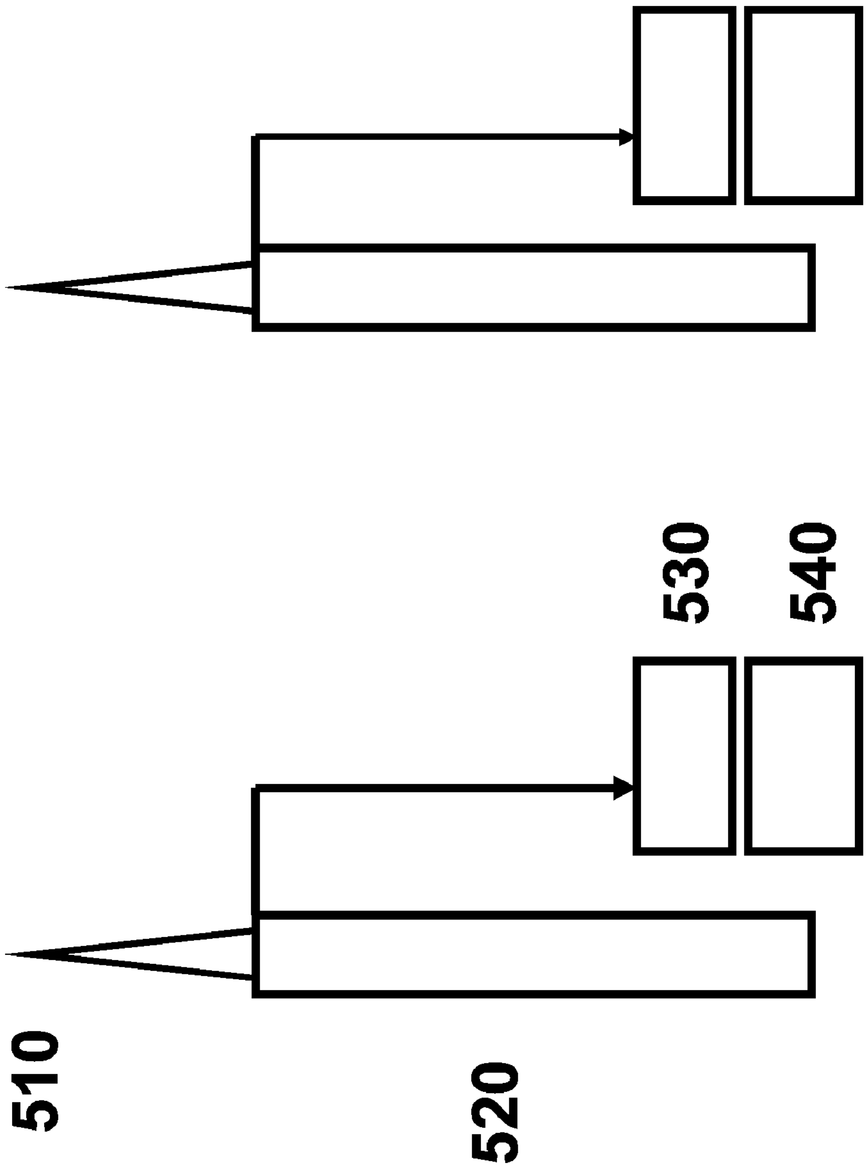
510
520
530
540

Fig. 12

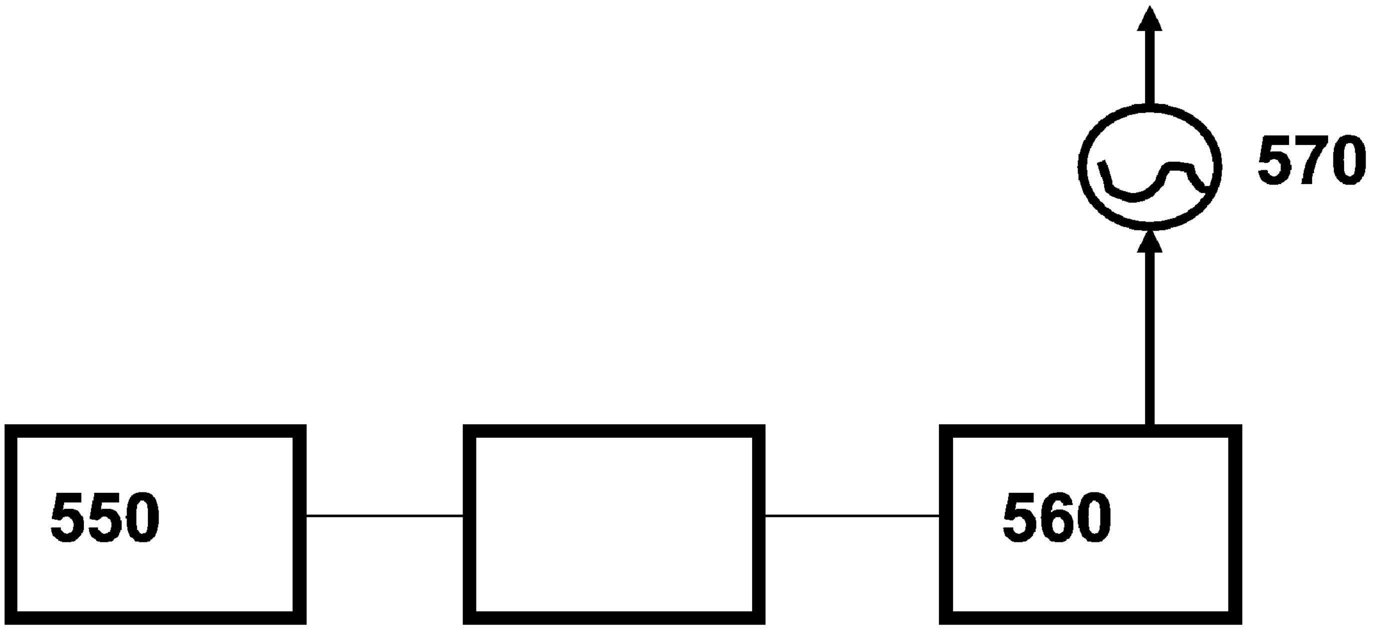

Fig. 13

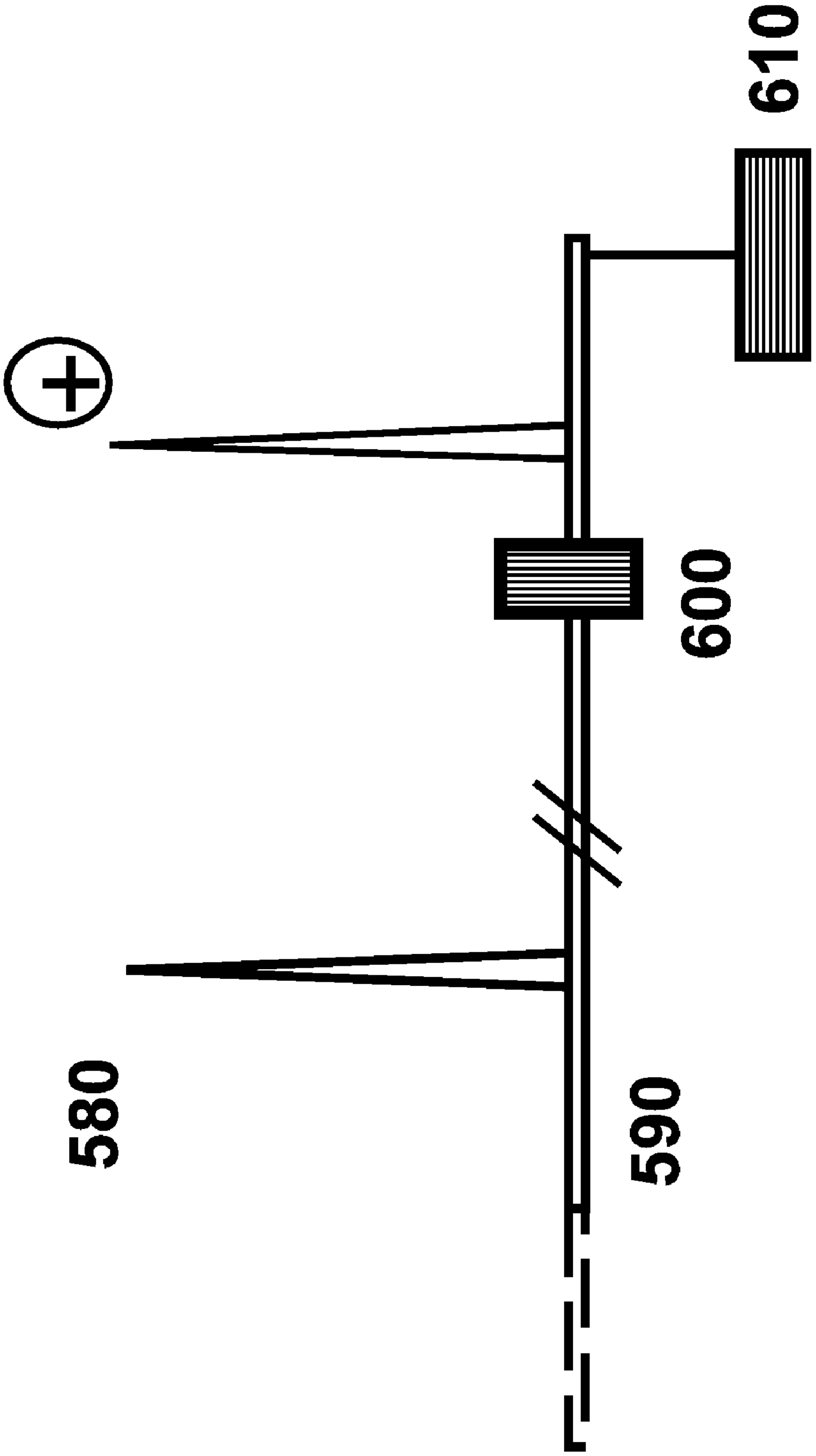
610
600
580
590

Fig. 14

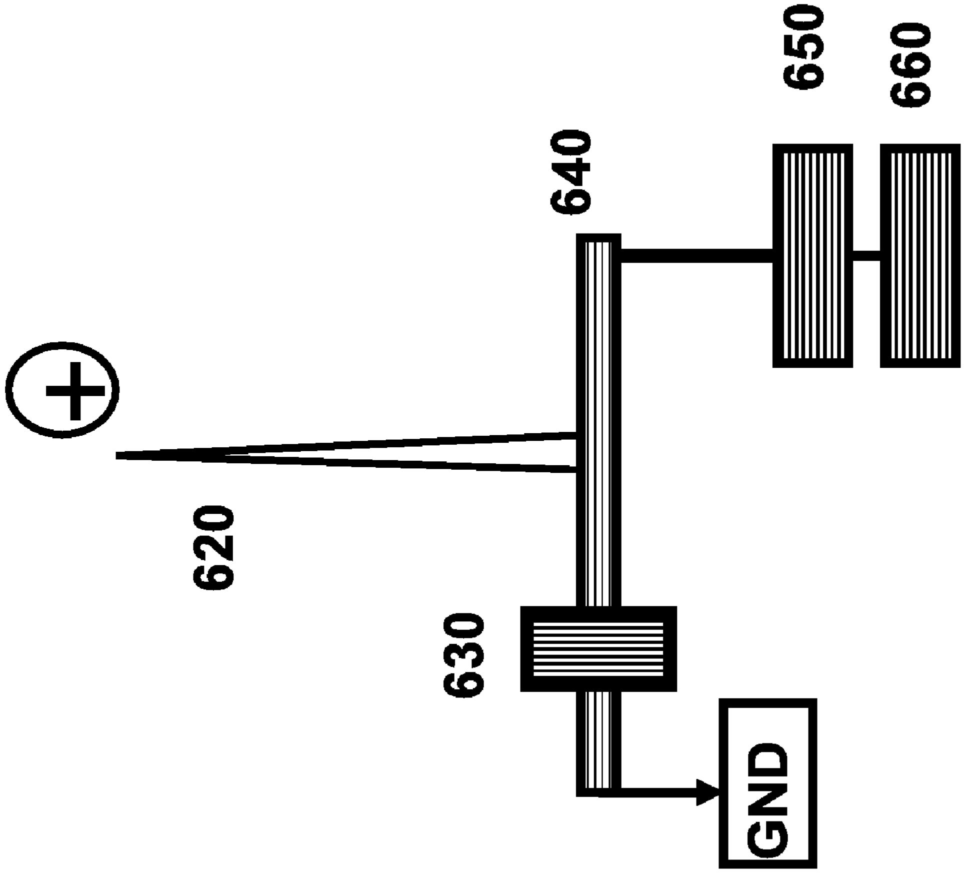
620
630
640
650
660
GND
+

Fig. 15

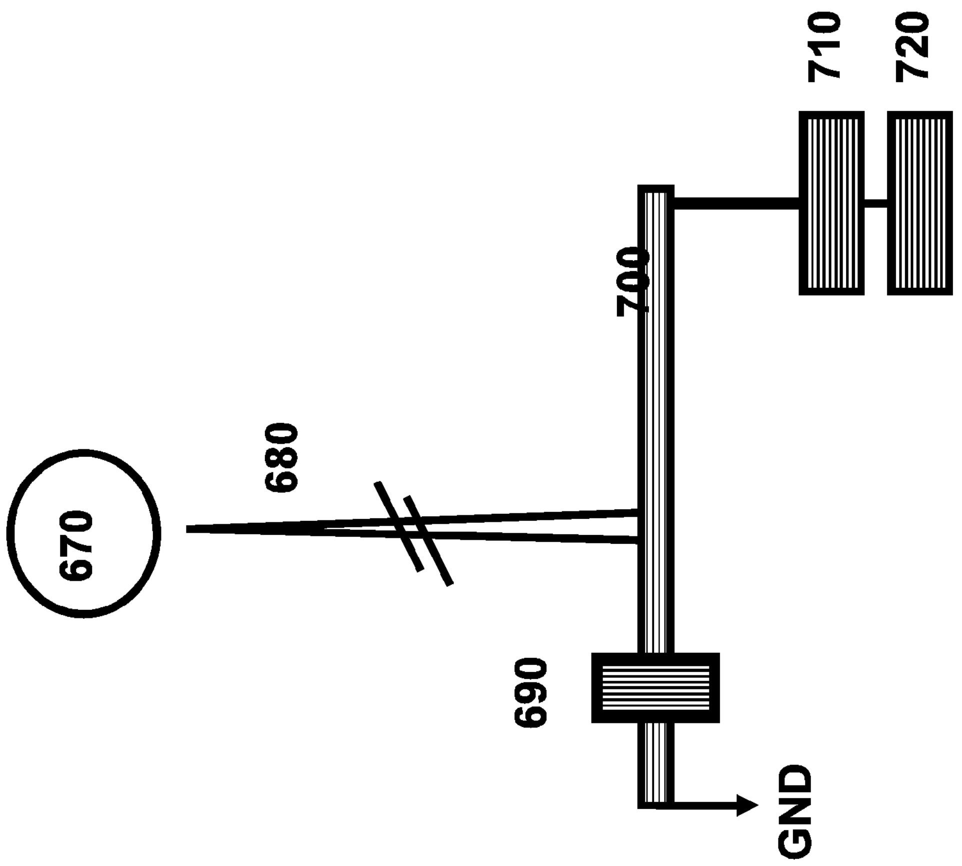
670
680
690
700
710
720
GND

Fig. 16

METHOD OF ATMOSPHERIC DISCHARGE ENERGY CONVERSION, STORAGE AND DISTRIBUTION

TECHNICAL FIELD

[0001] This invention relates to the field of electrical energy generation and storage in general and the conversion of atmospheric electrical discharge to a useable form of energy by arresting, storage and retransmission of lightning induced electrical discharge in particular.

BACKGROUND OF INVENTION

[0002] Lightning has always been a destructive force of nature. Causing death by electrocution and starting fires both in residential areas and in forests. In either case the loss of life and property is substantial. Lightning strikes have affected communication systems, and by uncontrolled arcing, electricity grids leading to loss of power with substantial economic loss

[0003] Lightning is the number one cause of forest and farm fires just in the United States and causes over 80 percent of all livestock losses due to accidents. Property losses run into billions of dollars annually and thousands of deaths and injury are reported yearly. Business losses from power and communications failure run into billions of dollars annually. Up to now, lightning strike damage has generally been controlled primarily by grounding.

[0004] A lightning discharge can have voltage of 10 MV to 100 MV and current of 1 kA to 300 kA. It packs so much voltage that it can leap a mile through the air and strike another object. Lightning can strike a building directly or leaping to it after striking a tree or other nearby object or by following a power line. It can strike any object that provides an easier path to the ground for it than the air. Because of the large network of lines, lightning can strike a utility pole at one point and be transmitted to any location in the network. Because of this, utility companies use multiple grounding and other protective devices per mile to ameliorate the lightning effects.

[0005] There has been a lot of activity in the alleviation of lightning damage but none in utilizing its energy potential.

[0006] U.S. Pat. No. 7,495,168 discloses a dipole lightning conductor which can gather electrical charges from the atmosphere during a thunderstorm, opposite to the polarity of the adjacent cloud on one side and induce a different polarity on the other end contacting the ground. When a sufficient potential difference is reached between the dipole and the earth, a dielectric breakdown occurs sending a large amount of earth charge to the thundercloud, inducing a thunderbolt.

[0007] In U.S. Pat. No. 6,012,330, Palmer discloses a method of inducing lightning strikes by shooting a stream of ionized water into a thundercloud and triggering electrical conduction through the ionized water column to the ground. In this invention, while ionized water is used, advantage is taken also of capillary wetting of conducting mesh cable at the tip of which is tethered a balloon of conductor coated Mylar. It is easier to float a balloon than to pump a column of water to any significant height.

[0008] In U.S. Pat. No. 6,320,119 Gumley discloses a capacitively coupled composite air terminal with sharp and rounded end components for voltage manipulation to minimize electric field ahead of streamers to minimize corona discharge.

[0009] Kato, in U.S. Pat. No. 5,280,335, discloses a compound lightning low voltage arrester with high energy tolerance for power and telephone lines comprising a spark gap is connected to a serial combination of Zinc Oxide arrestor and a non-inductive resistor. By cooperative execution between the spark gap and the Zinc Oxide arrester, the Zinc Oxide is chosen to be of a rating such that thermal breakdown does not occur before the spark gap ignition voltage is reached. The arrangement ensures reliable and predictable discharge and arrester survival even under multiple lightning strikes of close proximity.

[0010] Subbarao in U.S. Pat. No. 4,338,648, discloses a discharge counter and arrester current meter to determine arrester steady state current and the number of discharge events occurring during a thunder storm lightning arrester event.

[0011] Greenwald and Moses in U.S. Pat. No. 5,610,813 disclose a thunderstorm cell detection and mapping system for acquiring localized lightning strike information and identifying and locating active thunderstorm cells.

[0012] In U.S. Pat. No. 4,272,720, a method of differentiating between cloud-cloud and cloud-ground discharges is disclosed. Indicating discharge events with no accompanying lightning ground strike. This indicates that even in the absence of streamers, there is significant electrical discharge that can be converted and stored.

[0013] Weir and Nelson disclose in U.S. Pat. No. 7,033,406, a high energy density ceramic capacitor with energy capacity of 52 kW.h in 2005 cu.in, the equivalent of 1600 kW.h/m^3. Hansen in U.S. Pat. No. 6,078,494 discloses multilayer ceramic capacitors comprising doped barium-calcium-zirconium-titanate dielectric, the materials basis for U.S. Pat. No. 7,033,406. It is claimed to be characterized by high dielectric constant, high stability of capacitance value, long service life, low loss factor, high insulation resistance capacitance, low voltage dependence, and wide temperature range stability. Electrodes of base metal alloys from the group of Ni, Fe, Co or their alloys are claimed to be perhaps just as effective as noble metals containing gold, silver, platinum and palladium and may also contain Cr, Ti, Zr, V, Al, Zn, Cu, Sn, Pb, Mn, Mo and W.

[0014] Although the energy density capacity is not disclosed, it may be in the same range as U.S. Pat. No. 7,033,406 showing a path for storing the energy that can be extracted from lightning discharge.

[0015] U.S. Pat. No. 5,361,187, claims dielectric constant (k) of up to 19000 using similar group of dielectrics and processes. In U.S. Pat. No. 5,604,167 dielectric constant values of between 11,000 and 14,000 are disclosed by the same inventor with equivalent materials and processes.

[0016] Since these capacitors use standard powder processing technique, it can reasonably be expected that equivalent characteristics may be observed with naturally occurring doped silicates such as phyllosilicates for example, montmorillonite: $(Na,Ca)_{0.33}(Al,Mg)_2(Si_4O_{10})(OH)_2.nH_2O$, or cyclosilicates such as Benitoite —$BaTi(Si_3O_9)$ as base materials appropriately calcined and doped.

[0017] In U.S. Pat. No. 6,078,494 multilayer ceramic capacitors comprising doped barium-calcium-zirconium-titanate dielectric is disclosed. It is claimed to be characterized by high dielectric constant, high stability of capacitance value, long service life, low loss factor, high insulation resistance and low voltage dependence, and wide temperature range stability. Electrodes of base metal alloys from the group of Ni, Fe, Co or their alloys are claimed to be perhaps just as effective as noble metals containing gold, silver, platinum and palladium and may also contain Cr, Ti, Zr, V, Al, Zn, Cu, Sn, Pb, Mn, Mo and W. In U.S. Pat. No. 5,361,187, the inventors claim dielectric constants of up to 19000 using similar group of dielectrics and processes. In U.S. Pat. No. 5,604,167 dielectric constant values of between 11,000 and 14,000 is disclosed by the same inventors with equivalent materials and processes. Since these capacitors use standard powder processing technique, it can reasonably be expected that equivalent characteristics may be observed with naturally occurring doped silicates such as phyllosilicates for example, various grades of montmorillonite: $(Na,Ca)_{0.33}(Al,Mg)_2(Si_4O_{10})(OH)_2.nH_2O$, or cyclosilicates such as Benitoite—$BaTi(Si_3O_9)$ as base materials appropriately calcined and doped.

[0018] Although the technology to arrest lightning strikes has been around for at least a hundred years, this invention to harness and store that power has only become viable due to recent advances in super-capacitor technology which has made possible the fabrication of ultrahigh capacitance capacitors with very long lifetimes, high power, high energy densities and very fast charge and discharge rates.

[0019] Just as flood water, another destructive force of nature in times past was controlled by building dams, and from the dams eventually came hydroelectric power; and wind including thunderstorms has now been harnessed to generate electricity, this invention describes the technology to harness the power of atmospheric discharges including lightening storms to generate useable electrical energy.

[0020] This disclosure also introduces the concept of an Energy Dam. Facilities of substantial energy storage capacity capable of receiving energy feeds from various energy sources such as Solar, Wind, Thermal Power plants and Lightning. For example, during a thunderstorm, electricity generated by wind and lightning is collected and stored.

[0021] Lightning strikes every part of the globe but not uniformly. The regions with the highest historical concentration of lightning strikes are shown in FIG. **1***c*. These include Florida and the Gulf Coast in the Americas, the Equatorial Highlands of DRC, Rwanda and Burundi in Central Africa and the Monsoon Belt in Asia.

[0022] Except for the Americas, typically, these regions have very little electricity infrastructure. With the capability disclosed here, substantial reserves of electricity can be generated, stored and possibly traded.

[0023] In regions of epileptic electricity supply, disruptions in power do not necessarily translate into disruption in service. Since when power is generated, it is fed into these energy dams or local storage units for eventual use. Just as in water supply, the customer is not necessarily conscious of when the local water company is pumping water into storage tanks or local dams, in energy dams, the customer need not be conscious of whether or when electricity is being generated by the utility company because the power to his home is drawn not necessarily live from the grid but from the local energy storage units or energy dams.

BRIEF SUMMARY OF THE INVENTION

[0024] A method of harnessing atmospheric discharge and storing said energy in a useful form for subsequent use is presented.

[0025] Since the voltage of a lighting bolt can be as high as 100 MV, a methodology for safely channeling this discharge for storage is disclosed.

[0026] It is the also the object of this invention to present a storage medium in which an energy capacity of 1 MW.h in a volume of under 1 m^3 can be achieved.

[0027] It is also a further objective of this invention to show how lightening strikes on tall buildings can be harnessed to provide energy for the building either directly or fed into the grid.

[0028] It is also an objective of this invention to present a methodology for harnessing lightening, heretofore a source of power disruptions for electricity grids and converting it into a useable form of energy and feeding it back into the grid.

[0029] Since each electricity grid has a lightning suppression unit, a methodology for harnessing the impulse from these units from lightening strikes and feeding it back into the grid is disclosed.

[0030] a method of deploying a network of said storage units along a utility grid is also disclosed.

[0031] It is also the objective of this invention to disclose a method of modifying the storage units for single building deployment.

[0032] A method of deploying said storage units as vehicular charging centers even in locations not readily served by a utility grid is also disclosed.

[0033] A method of manipulating the direction and movement of the lightening bolt in open air locations and redirecting them to predetermined collection points is disclosed.

[0034] Lightning storms are some of the major causes of wild fires in the world. A method of redirecting the electrical discharges in any given space via a network of polarized and non-polarized collectors is presented. The discharges can be stored, fed into the grid or grounded.

[0035] A method for non-spontaneous discharge generation and storage is disclosed. And a method for adapting the methodology for electric automotive deployment is also disclosed.

BRIEF DESCRIPTION OF DRAWINGS

[0036] For understanding of the present invention reference is made of the accompanying drawings in the following DETAILED DESCRIPTION OF THE INVENTION. In the drawings:

[0037] FIG. **1***a*. Picture of lightning strike over a city.

[0038] FIG. **1***b*. Picture of lightning strike over open plain.

[0039] FIG. **1***c*: Global Distribution of Lightning Strikes: April 1995-February, 2003. (Source: NASA).

[0040] FIG. **2** shows a schematic of discharge collection and storage system.

[0041] FIG. **3** shows a schematic of collection unit showing metallic composition variation.

[0042] FIG. **4** Shows a Schematic of composite collector showing polarized ends.

[0043] FIG. **5** Schematic of collector network configuration with negatively polarized ends to manipulate and redirect lightning electrical discharge away from the earth for eventual collection at a predetermined location.

[0044] FIG. **6** Plan view of polarized collection network as shown in FIG. **5** acting to concentrate and modulate the electrical discharge and redirect such for collection at predetermined location.

[0045] FIG. **7***a*. Schematic showing a representation of a network of varistors arranged in rows and columns with the net effect of applying resistive arrestors in parallel and in series to effect voltage mitigation of atmospheric discharges prior to collection and storage.

[0046] FIG. **7***b:* Schematic of varistor of power mitigation unit showing cascade of ZnO arrestor in series with non-inductive resistor in parallel with spark gap.

[0047] FIG. **7***c:* Multilayer varistor with ZnO, ZnS and WO_3/WSi_2

[0048] FIG. **8***a*. Schematic showing cross section of a collection unit.

[0049] FIG. **8***b*. Schematic showing cross-section of multilayer ceramic capacitor of **8***a*.

[0050] FIG. **9** Schematic of collection unit network mesh in forested area with air frames deployed on each tree as collection of trees linked in a mesh network.

[0051] FIG. **10** Plan-view schematic of collection unit network mesh of FIG. **9**

[0052] FIG. **11** Schematic showing collection unit deployment on electricity transmission poles linked in a network and fed into the electricity grid.

[0053] FIG. **12** Schematic showing collection unit deployment on light poles or wireless towers.

[0054] FIG. **13** Schematic showing network of linked collection units and storage units on transmission lines, light poles or wireless towers fed into the grid.

[0055] FIG. **14**. Schematic of a system for non-spontaneous atmospheric discharge generation and storage.

[0056] FIG. **15** Schematic of atmospheric discharge collection and storage unit deployable on automotive units.

[0057] FIG. **16** Schematic of balloon-assisted discharge collection and storage unit.

DETAILED DESCRIPTION OF THE INVENTION

[0058] This invention describes a method of harnessing the electrical energy in atmospheric discharges, including, but not limited to lightning, storing said electrical energy and transmitting said energy for use. To this end, as an illustration, the preferred embodiments are represented in the drawings shown in FIGS. **2** to **15**.

[0059] Since lightning discharge voltages range from 10 MV to 100 MV with current of 10 kA to 100 kA, the technology challenge will be in mitigating the ultrahigh power which range from 100 GW to in excess of 10,000 GW. Although individual strikes last less than 100 μsec, thunderstorms usually last longer than 30 min. in any cloud formation and location. Indicating that the energy that can be mined from any location in a thunderstorm can run in excess of 50 GW.H. Enough energy for 500,000 homes consuming 1000 kW.h per month. At $100 per 1000 kWh, the revenue equivalent is $5 million per thunderstorm per location.

[0060] In solar energy generation, the cost of a 1 MWe rating PV panel is about $1 million. To achieve 1 GWe, the cost implications will be about $1 billion. In fossil fuel plants, the construction cost is approximately $1.3 million per MWe, putting the cost for a 1 GWe at ~$1.3 billion excluding fuel, carbon footprint, and other operating cost. Though thunderstorms are not continuous, with storage capability, the cost basis for an atmospheric discharge energy source will be several orders of magnitude lower than fossil fuel or solar of any architecture.

[0061] One embodiment of this invention is depicted in FIG. **2**. Lightning is collected at the air terminals **10**, **20** and transmitted through the conductor **30** to an arrestor **40** and on to the capacitor **50** where the electricity is stored. From which it can be extracted **60** at point of use or transmitted into the grid.

[0062] The degree of atmospheric discharge charge collection can be influenced by the metallurgical composition of the collectors. FIG. **3** illustrates the fabrication of the collector with different metals of varying conductivities and atmospheric and electrical stability. In this embodiment, metallurgical compositions for the various segments comprise Ti, Cu, Al. Other metals or their alloys can be used depending on the required properties and the locations of deployment.

[0063] In yet another embodiment of the invention as shown in FIG. **4**, a composite collector with opposite polarity air terminals **100**, **110** is used. The air terminals are isolated from each other by an insulator **120**. The negative polarity terminal serves to deflect and manipulate the charge to the positive terminal from which it is conducted to the storage unit **130**.

[0064] Another embodiment of the invention is illustrated in FIGS. **5** and **6**. In this representation, negative polarity collectors, **160** deflect the atmospheric discharge so that they do not reach the ground but are deflected for collection at a predetermined terminal **180** of positive polarity and thereafter storage **200**. A plan view representation of this embodiment is shown in FIG. **6**.

[0065] It is essential to mitigate the power of the lightning before storage. In one embodiment of this invention, a collection of varistors of various power capacities, linked in series **270**,**280** and **290** are connected in parallel in lines **1** to n, to achieve a desired capability as shown in FIG. **7***a*. The resultant power is stored in the capacitor **300**. An insulator or fuse **305** is placed at the end of the varistor chain to protect the capacitor **300** against over-current. In the event of overstressing, the dielectric or varistor is deformed and the excess current is grounded.

[0066] Each component of the power mitigation unit is any of the following configurations as shown in FIG. **7***b:* A ZnO varistor **312** connected in series with a non-inductive resistor **313** and in parallel with a spark gap **311**. The resultant current is further modulated through another ZnO unit **314** and resistor **315**. A cascade of this configuration could be adopted to achieve the desired properties.

[0067] Zinc Oxide has a relatively large band gap of ~3.3 eV at room temperature. The consequence of this large band gap includes, higher breakdown voltages, ability to sustain large electric field and high power and temperature operation. Because of this, ZnO is the primary component in varistors and in voltage suppression devices. Zinc Sulfide has a band gap of 3.5 eV for the cubic form and as high as 3.91 eV for the hexagonal form at 300K. While ZnO band gap can be tuned from between 3-4 eV by alloying with magnesium oxide or cadmium oxide, the co-processing of ZnO with ZnS of the hexagonal wurtzite form should provide electrical field and other operational advantages over ZnO alone.

[0068] Multilayers of ZnO and ZnS or as sandwich layers with each other with ZnO achieving 50-99.9 mol. % of each layer can be adopted to improve on the suppression properties of the single layer ZnO varistor as is shown in FIG. **7***c*. Said varistor comprising a ceramic base body **318** of multiple layers, terminal electrodes **316**, **317** and internal electrodes **319**.

[0069] Alloying of ZnO with metal oxides and nitrides of transition elements including W, Zr, Mo, Mg, Cr, and Ti may improve its band gap characteristics and improve its voltage response non-linearity. Tungsten, in the form of tungsten silicide has very unique crystallization characteristics, particularly in the presence of dopants such as phosphorus or arsenic. Upon post deposition anneal, the Si in the non-stoichiometric tungsten silicide (WSix) segregates to the grain boundaries and the film interfaces. The dopant atoms diffuse preferentially along the grain boundaries. When oxidized, it is the segregated silicon that is oxidized selectively. If the W is exposed to the oxidizing ambient of O_2/Ar or O_2/N_2 at temperatures in excess of 700° C., tungsten oxide WOx (x=3, or 1.5) is formed. It has no corresponding crystallographic orientation to the underlying tungsten silicide or metal, being in the form of an amorphous powder. The fabrication of multilayer varistors with ZnO and WSi_2/WO_3, SiO_2 should provide the flexibility of modulation of the effective grain sizes of the host material to within 0.5-10μ as the case may be. This precise grain size control provides corresponding modulation of the electrical properties including overvoltage suppression.

[0070] A schematic of the storage unit **320** is shown in FIG. **8***a*. A collection of capacitors **330** are connected in parallel to achieve a desired capacitance. A schematic of the cross-section of each capacitor unit is shown in FIG. **8***b*. The unit **331** with dielectric layers **332** is a multilayer capacitor comprising alternately stacked internal electrodes **335**, **336** connected to external electrodes **333**, **334**. The dielectric layers may be of uniform composition, or layers of different compositions optimized for flexibility in the modulation of desired electrical, physical and reliability properties.

[0071] Capacitance in excess of 1 kF/m^3 with energy storage capacity of 1.5 MW.h is reported as being possible by existing art. With compositional and manufacturing improvements disclosed here these benchmarks are expected to be surpassed making possible energy storage capacities of up to 1 GW.h. Sufficient energy for 1000 households consuming 1000 kW.h of electricity per month for one month.

[0072] Lightning storms are a major cause of wildfire. To address this, in yet another embodiment of this invention as represented in FIG. **9**, air terminals **360** are set up on a selection of trees and connected by a conductor **370** in a mesh array as shown in FIG. **10**. The impinging charge is transmitted through a conductor **380**, through a suppressor unit **390** and on to the storage unit **400**.

[0073] Utility transmission lines usually have lightning collectors along the length of the network. In this invention as demonstrated in FIG. **11** air terminals **440** deployed transmission poles **450** are linked by a polarized conductor **460**. At predetermined locations the impinging charge is conducted through a suppression unit **490** and thereafter into the storage unit **500**. The stored energy is periodically released into the grid.

[0074] In yet another embodiment of the invention, air terminals **510** are deployed on light poles **520**. The collected charged are modified and stored in a storage unit **540**.

[0075] In another embodiment, the air terminal/collection units are deployed on radio transmission towers. The charge could be stored individually at each tower location or linked in a network as shown in FIG. **13**. In either case, the stored energy be used at each individual location or fed into the grid **570**.

[0076] In the absence of a thundercloud, lightening flash can be generated by the modification of the natural bias between the cloud and ground. This is demonstrated in FIG. **14**, where a positively biased conductor **580** of height greater than nearest ground elevation or nearest structure. The increased elevation and bias facilitates cloud to ground discharge even when there is no active lightning storm. The incoming charge is isolated by an insulator **600** and stored in the storage unit **610**.

[0077] In yet another embodiment of this invention, a balloon of conductive material is tethered to a terminal via a mesh line of conducting material. The balloon captures cloud to cloud discharges which are transmitted down to storage. In yet another embodiment this tether line may be wetted by a conductive solution such as but not limited to salt solution to facilitate the discharge conduction.

[0078] Another embodiment of this invention is shown in FIG. **15**. In an isolated collection system as shown here, the air terminal **620** which may be biased or unbiased is connected to a collection/storage unit **650**, **660** respectively by conductor **640** and isolated by a dielectric or insulator **670**. Should the discharge exceed the capacity of the isolation, the isolation is bridged and the charge grounded. This isolated collection system is particularly useful where there is no utility grid, particularly in very remote and isolated locations.

[0079] This isolated collection capability enables deployment on individual houses and isolated farms. In electric automotive recharge stations, this capability enables the location of these facilities in any location even those not served by regular electricity utilities lines.

[0080] In isolated rural areas, the deployment of these isolated collection units enables the provision of electricity where electricity grid infrastructure is not available or practical. The storage units can be deployed in banks as the estimated anticipated usage dictates.

[0081] In another embodiment of this invention, the isolated collection unit is modified to be portable for deployment on automobiles. The charge in the storage unit is linked with the automobile's electrical power or storage unit. Alternatively, it can replace the power storage unit of the automobile for pure EV or hybrid cars. Additionally, it can unplug these cars from the grid. With the ambitious industry plans for electric or hybrid cars the infrastructure for charging these cars either in garages or charging stations is somewhat limited. The capability to both 'charge as you go' and unplug from the grid would expand the infrastructure significantly.

[0082] In yet another embodiment of this invention, the isolated collection unit is deployed on trains. This could be in commuter trains that directly fed from the grid, or long-haul trains to augment or replace the fossil fuel used.

[0083] It is understood that the presentation of these steps in this disclosure is not exhaustive. Only the preferred embodiments of the invention and but a few of the examples of its versatility are shown and described in the present disclosure. It should be readily apparent to those of ordinary skill in the art that the invention is capable of use in various other combinations, environments and applications and is capable of changes or modifications within the scope of the inventive concept as expressed herein. These changes and modifications may be made without departing from the spirit and scope of the invention as set forth in the appended claims.

REFERENCES

[0084] U.S. Pat. No. 7,495,168,

[0085] U.S. Pat. No. 6,012,330

[0086] U.S. Pat. No. 6,078,494

[0087] U.S. Pat. No. 4,338,648

[0088] U.S. Pat. No. 6,407,900

[0089] U.S. Pat. No. 6,924,598

[0090] U.S. Pat. No. 5,569,972

[0091] U.S. Pat. No. 5,289,335
[0092] U.S. Pat. No. 4,272,720
[0093] U.S. Pat. No. 4,338,648
[0094] U.S. Pat. No. 5,610,813
[0095] U.S. Pat. No. 6,320,119
[0096] U.S. Pat. No. 5,604,167

1. A method of converting an atmospheric electrical discharge into a useable form of electrical energy comprising the following steps:
   - (a) Arresting the discharge via an air terminal,
   - (b) Said air terminal connected via a conducting unit to a power mitigating (step down) unit which can reduce the discharge voltage from approximately higher than 100 MV and 100 KA to below 20 kV, 10 kA.
   - (c) The step down unit connected to an energy storage unit which may comprise modules of high energy density capacitors
   - (d) Said storage units arranged such that the stored electrical energy can be discharged into the grid or to an appliance at point of use as desired.

2. A method of claim **1** wherein the arrestor and mitigating unit comprise an array of self-healing varistors in series and parallel chosen to ensure total conversion of the incident lightning discharge to a voltage/current level suitable for storage in said capacitors.

3. A method of claim **2**, where the power mitigation unit is any of the following configurations:
   - (a) Single ZnO varistor or
   - (b) Zinc Oxide arrester connected in series with a non-inductive resistor
   - (c) Zinc Oxide arrester in series with said non-inductive resistor and in parallel with a spark gap.
   - (d) A cascade of configuration as described in (c).

4. A method of claim **1** wherein the storage unit comprises modular array of high energy density doped ceramic capacitors with storage capacity of 10 kWe.h to 1 GWe.h and volume 0.1 to 1 m^3 and up to 1000 units as the desired capacity dictates.

5. A method of broad spatial lightning protection by the manipulation of the path of an electrical discharge via the manipulation of the air terminal bias magnitude and polarity Said arrestors being arranged in an array to achieve the redirection of the discharge path. The discharge path being directed to a collection unit at some location significantly removed from the initial discharge impact position.

6. A method of claim **5** wherein the discharge redirection is via a biased conductor to the said collection and storage unit.

7. A method of claim **5** wherein the area of protection is a forest whereby:
   - (a) A network array of air terminals is deployed on trees, in close proximity to said trees with tip approximating the treetops, or girding the trees to redirect the discharge along a biased conductor to a collection unit for the purpose of mitigating lightning induced fires.
   - (b) Said discharge electrical energy being stored or fed into the electricity grid.

8. A method of claim **1** wherein the discharge collection is on electrical power lines and the stored energy is fed back into the grid.

9. A method of claim **1** wherein the discharge collection and storage are at telecom towers, light poles or other free standing structures. Said storage being at individual collection sites or some location serving an array of collectors and said stored energy used for individual structure's operational energy need or fed back into the utility grid.

10. A method of non-spontaneous generation of atmospheric discharge by the following:
    - (a) Modulation of the magnitude and bias in a dipole air terminal relative to the ground
    - (b) Conductive material balloon tethered by a conductive mesh cable.
    - (c) Said cable being soaked by an ionic liquid, including, a salt solution to facilitate said atmospheric discharge conduction.

The discharge thus formed is stored for feeding into the greed or for operation as a stand-alone energy dam.

11. A method of claim **1** in which the discharge collection unit is configured as a mobile unit. Said unit being deployed on an automobile or any mobile platform. The said deployed unit extracting electrical charge from the atmosphere and storing such energy in a deployed storage unit on the said automobile. The energy thus stored, feeds the car battery which obviates the need for recharging at charging centers or from the grid.

12. A method of claim **1** wherein the electrical discharge energy collection unit is a stand-alone unit comprising modular capacitors configured as an electric vehicle charging station.

13. A method of claim **1** wherein the air terminals are deployed on buildings and the discharge energy collected thereby are:
    - (a) Stored and fed back into the grid.
    - (b) Stored for use by each building
    - (c) Stored in a central energy storage facility for use by the community in general.

14. A method of claim **1** wherein the said storage units are underground.

15. A method of claim **1** wherein the storage unit is of capacity and configuration to accept energy feeds from multiple electricity generation sources: Thermal, Hydro, Solar, Wind and Lightning.

16. A method of claim **4** wherein the storage unit is integrated into a Utility's Switch Farm.

17. A method of claim **4** wherein the multilayer ceramic capacitor unit comprises dielectric layers of varying compositions individually or as sandwich layers with each other containing Barium-Titanate silicates as host material doped with Zr, WSi_2/WO_3 to modulate grain growth to between 0.1 to 10 μm as the case may be; said modulation providing corresponding control of electrical and reliability properties.

18. A method of claim **1** wherein the varistor comprise multilayers of ZnO and ZnS individually layered or as sandwich layers with each other with ZnO achieving 50-99.9% of each layer; said layers being alloyed metal oxides of transition elements including W, Zr, Mo, Mg, Cr, Ti to improve voltage response non-linearity.

* * * * *

WILLIAM H. WARD.

Improvement in Collecting Electricity for Telegraphing, &c.

No. 126,356. Patented April 30, 1872.

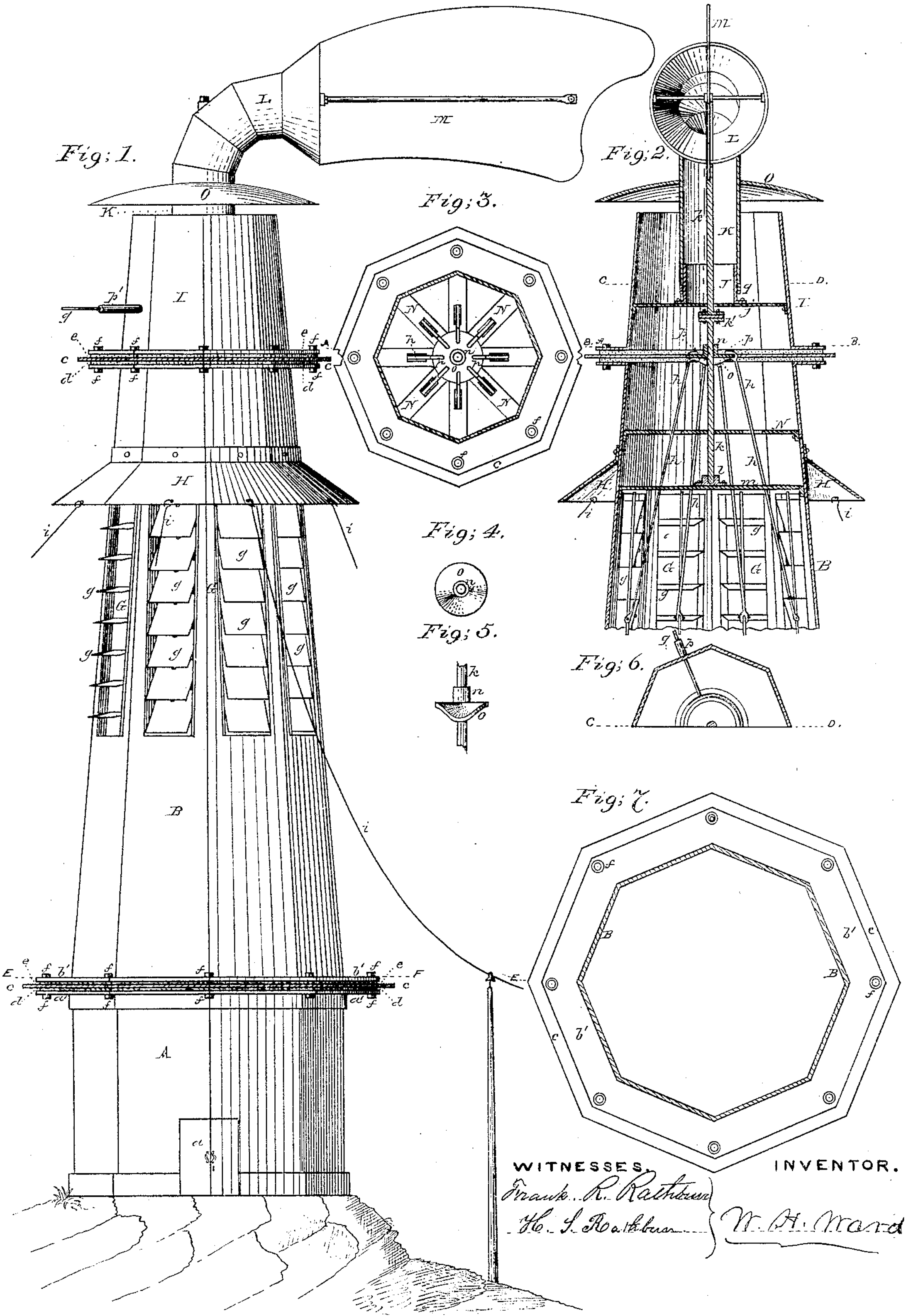
Fig; 1.
Fig; 2.
Fig; 3.
Fig; 4.
Fig; 5.
Fig; 6.
Fig; 7.
M
L
O
K
I
p'
g
e
f
c
d
H
i
G
B
E
F
b'
a'
A
a
J
k
k'
n
p
o
N
m
l
h
C
D

WITNESSES.

Frank R. Rathbun

H. S. Rathbun

INVENTOR.

W. H. Ward

AM. PHOTO-LITHOGRAPHIC CO. N. Y. (OSBORNE'S PROCESS.)

126,356

UNITED STATES PATENT OFFICE.

WILLIAM HENRY WARD, OF AUBURN, NEW YORK.

IMPROVEMENT IN COLLECTING ELECTRICITY FOR TELEGRAPHING, &c.

Specification forming part of Letters Patent No. 126,356, dated April 30, 1872.

I, WILLIAM HENRY WARD, of Auburn, in the county of Cayuga and State of New York, have invented an Electrical Tower for Accumulating Natural Electricity for Telegraphic Purposes, of which the following is a specification:

My invention consists of a tower for the purpose of receiving and imparting natural electricity, so as to be in constant contact with that upper stratum of electricity which surrounds the earth, by tapping which a never-failing supply is formed when brought into contact with the earth, as will be more fully explained hereinafter.

In the accompanying drawing, Figure 1 represents a side elevation of my improved electrical tower. Fig. 2 is a vertical central section of the upper part of the same. Figs. 3, 4, 5, 6, and 7 are detached views of several parts of the same.

The tower is constructed in three separate sections, a lower, a middle, and an upper one, and is placed on elevated mountain tops or peaks. The lower one, A, is a mere shell, having a door, *a*, and is constructed of any material suitable for the purpose. It is insulated from the middle portion B by means of a glass diaphragm, *c*, held between similar diaphragms of rubber, *d*, and of gutta-percha, *e*, by means of flanges *a'* and *b'*, formed on the upper end of the lower portion and the lower end of the middle portion of the tower respectively. These flanges are secured to each other by insulated bolts *f*. The middle portion B is provided with suitable openings or windows, G, having shutters or slats *g* pivoted in them, so that, by means of raising or lowering rods *h*, suitably connected to said shutters, the openings G may be shut or opened. A projecting roof, H, is formed on or secured to the middle portion B just over the openings G, and which serves the double purpose of protecting the said openings from the effects of the weather, such as rain or snow, and also for receiving the aerial electricity, which may be drawn from it by wires *i*, for land-line purposes. Above this roof H the middle portion of the tower is again insulated by diaphragms of glass, rubber, and gutta-percha from the upper portion I of the tower in the same manner as it is insulated from its lower portion. A circular plate, *j*, secured on the inside of this upper portion, bears a short tube, J, which is surrounded by the tube K of the ventilator L, from which latter the vane M extends; or the tube J may be held by rods extending from the side of the tower centrally. This ventilator is supported by a rod or shaft, *k*, firmly attached to the tube K, and having its lower bearing in a step, *l*, on a brace, *m*, crossing the middle portion B of the tower just above the openings G. This rod or shaft *k* is formed in two parts, insulated from each other, as shown at *k'*. On the lower portion of this shaft *k* is keyed or otherwise secured a sleeve, *n*, from which a horizontal serpentine cam-plate, *o*, shown in detail in Figs. 4 and 5, extends, over the rim of which the forked ends *p* of rods *h* seize, and which is so arranged relatively to the vane M and the shutters or slats *g* that the revolution of the vane by the wind will open the windward and close the leeward shutters or slats *g* and openings G by means of the rods *h*. Guide-plates or their equivalents N, provided with suitable openings, through which the rods *h* pass, keep the upper forked ends *h* of the latter in constant contact with the cam-plate *o*, which raises or lowers the rods, and consequently opens or closes the openings G as it revolves. A circular roof, O, extends from the tube K of the ventilator so as to cover and protect the open upper end of the tower. The upper portion I of the tower is provided with an insulating-tube, *p'*, through which a copper wire, *q*, enters the portion I of the tower, and is coiled around the tube J just under the tube K, which wire may connect that upper portion of the tower with any land line of wires. The upper portion I of the tower, as well as the tubes J K, may be constructed of zinc, while the vane may be made of zinc, copper, or any equivalent material, which, with the zinc and the moisture of the atmosphere, would form an electrical current. As the vane revolves it opens the openings to windward and closes those to leeward, thus helping to drive an aerial current of electricity into the insulated middle portion of the tower, which current passes upwardly through the upper portion of the tower and out through the ventilator or the top, which is swung around by the wind or aerial electrical current, thus forming a draught in addition to the closing of the shutters or openings to leeward, by means of which draught the electrical current is forced upwardly and out at the vane. The top portion of the tower is of course completely insulated from all below it, including the revolving central ventilator-shaft. As the middle portion of the tower allows the electrical current of atmosphere, wind, &c., to pass within, up, and out at the top it forms a continuous current, whereby the tower is receiving continually fresh and new supplies of electricity, which can be drawn from the projecting roof H by the wires *i* for the use of land lines of telegraphs or for other purposes, such as light, heat, &c. By the use of aerial electricity I entirely dispense with artificial batteries, forming my circuit merely by connecting the aerial current with the earth current. For instance, to bring Buenos Ayres, in South America, into direct connection with New York, the following plan would be pursued: One electrical tower is erected on Pike's Peak or any other suitable high mountain in North America, and another similar tower on some suitable peak of the Andes in South America. The former would, by means of land-lines, be connected directly with Denver, which place is again connected with all the prominent cities of the States. In a similar manner the southern tower is connected by land-lines with the prominent cities via Quito. New York telegraphs to the tower on Pike's Peak, and the operator having connected the land-line with the aerial current, the signals are transmitted through the aerial current to the tower in the Andes in South America, and from there—the land-lines being suitably connected with the aerial current—to Quito and Buenos Ayres. In this manner a message would be sent entirely by natural electricity in place of artificial. In the same manner a message may be sent across the ocean by having a high tower on each continent, each of which towers would have to be, of course, through land-lines connected with the earth to enable the ground current with the aerial current to form a circuit. Different towers may be erected on the different continents, and if they are all what is technically called hooked on—that is to say, connected to the earth—a signal given at one tower will be repeated at all the towers, they being connected with each other by the aerial current. If the earth-connection is severed, or the insulation with the tower destroyed, there is no power; but by insulating the tower and concentrating its force to a point, bringing the same corresponding effect from the earth current in connection, an exceedingly powerful electrical force is created.

Having described my invention, I claim—

1. A tower constructed so as to collect, hold, distribute, and utilize aerial currents of natural electricity for telegraphic and other purposes, essentially as described.

2. A tower for collecting aerial currents of natural electricity, constructed of three sections, insulated from each other and the earth, or their equivalents, substantially in the manner described.

3. In an electrical tower, the combination of the shutters or slats *g* with the ventilator-vane in such a manner that the vane, through suitable mechanism, always opens the shutters to windward and closes them to leeward, substantially as described.

4. In an electrical tower, the collecting, distributing, and protecting roof H, substantially as and for the purpose described.

5. The combination of the tube J and ventilator L with its vane M and the coiled wire or small cable *q*, substantially as and for the purpose set forth.

6. The combination of the insulated shaft *k* and rods *h* having forked upper ends *k'* with the shutters or blinds *g*, serpentine cam-plate *o*, and ventilator L, substantially as and for the purpose described.

The above specification of my improvement, being a tower for accumulating natural electricity for telegraphic purposes, signed this 29th day of June, A. D. 1871.

W. H. WARD.

Witnesses:
A. L. PALMER,
J. H. STARIN.

Dr Prem
Life Improving Guide

Today's Mortgage Rate
3.75% APR
15-YEAR FIXED
Select Loan A

Guide Home Parenting Guide Family Guide Green Guide Style Guide Design Guide Home Imp

Home › Green Guide › Capturing free atmospheric electricity: Feasibility and challenges

GREEN GUIDE

Capturing free atmospheric electricity: Feasibility and challenges

Earth's atmosphere is in itself an infinite storeroom of energy. The regular diurnal variations of the Earth's electromagnetic network produce strong electric currents on a gigantic scale. The Earth has its own negative electr while the atmosphere is positively charged. The electric potential increases with the altitude, hence the electric for proportional to the atmospheric altitude. In the present times, where alternative energy holds the key to the energ crisis, the abundance of this source of energy has attracted the researchers' attention. Evermore exciting is its free availability in air and clouds. There have been efforts to capture it since long but a considerable success is still har come by.

Atmospheric Electricity, nature lends it for nothing:

All atmospheric electricity is a result of the potential difference between any point on the surface of the negatively charged earth and the other in the positively charged air above it. Earth's electromagnetic network consists of thunderstorms, the electrification of the air in a combined or uncombined state and the polar Aurora effects. Equa experiences more of thunderstorms than anywhere on the entire planet. These thunderstorms express themselve the form of lightning and atmospheric pressure, wind levels, solar winds and the solar particles. Estimates reflect t lightning strikes somewhere on the surface of the Earth about 100 times every second.

Secondly, the Polar Aurora provides a great natural spectacle and is an outcome of the atmospheric electricity. Ear magnetosphere abounds in profuse solar-wind-transferred-energy, which is visible from the polar regions of the p

Moreover, this simple, robust and low cost electricity is much cheaper than wind turbines. You don't have to count the weather conditions, and it's available day and night. No doubt, it's in abundance at the poles but that doesn't d its presence elsewhere.

Feasible ways to capture it:

Capturing Atmospheric Electricity means to exploit the natural electrostatic potential gradient of the Earth. The electrostatic potential gradient between the Earth and the ionosphere is about 100Vm-1 near the surface in summ and 300Vm-1 in winter. Electrically charging a bank of capacitors or operating an electrostatic motor/parametric conversion machine are some of the possible ways to convert this static atmospheric electricity into conventional Alternating Current.

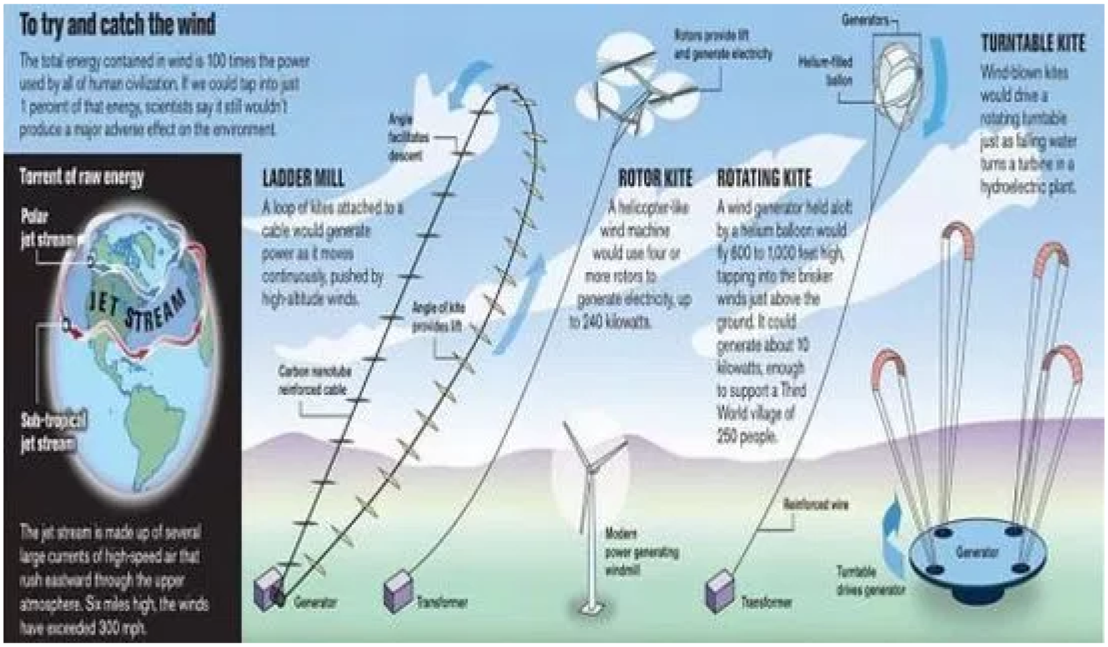
To try and catch the wind
Torrent of raw energy
Polar jet stream
JET STREAM
Sub-tropical jet stream
LADDER MILL
Angle facilitates descent
Angle of kite provides lift
Carbon nanotube reinforced cable
Generator
Transformer
Rotors provide lift and generate electricity
ROTOR KITE
ROTATING KITE
Modern power generating windmill
Generators
Helium-filled ballon
Reinforced wire
Transformer
TURNTABLE KITE
Generator
Turntable drives generator

The total energy contained in wind is 100 times the power used by all of human civilization. If we could tap into just 1 percent of that energy, scientists say it still wouldn't produce a major adverse effect on the environment.

The jet stream is made up of several large currents of high-speed air that rush eastward through the upper atmosphere. Six miles high, the winds have exceeded 300 mph.

A loop of kites attached to a cable would generate power as it moves continuously, pushed by high-altitude winds.

A helicopter-like wind machine would use four or more rotors to generate electricity, up to 240 kilowatts.

A wind generator held aloft by a helium balloon would fly 600 to 1,000 feet high, tapping into the brisker winds just above the ground. It could generate about 10 kilowatts, enough to support a Third World village of 250 people.

Wind-blown kites would drive a rotating turntable just as falling water turns a turbine in a hydroelectric plant.

Image Source: SFGate

1) Tapping the Jet Streams:

The jet streams generally blow at speeds up to 310 mph over the northern hemisphere and are a possible source atmospheric electricity. If the researchers could succeed in tapping these ferocious winds, the world's entire electr needs can be met. Scientists believe that a kite-like wind-power generator, lobbed into the upper atmosphere to capture the power of the jet stream, can produce energy, which is transmittable to Earth via aluminum or copper or invisible microwave beams. Next step includes its storage and transmission to target consumption points. The energy generation via the jet streams involved the inclusion of the designs enlisted below:

a. Ladder Mill: Rowing against the high-altitude winds, kites attached to a cable could generate a sizeable amount power.

b. Rotor Kite: A unique machine design includes more than four rotors to generate electricity, up to 240 kilowatts

c. Rotating Kite: The concept lies deep-rooted in conception of a wind generator at 600 to 1,000 feet. A helium ba keeps the kite in the air. The proposed electrical generation is about 10 kilowatts.

d. Turntable Kite: Though no data on the projected capacity is available, a rotating turntable still ensures better p generation.

2) Through Aerostats, as suggested by Dr. Plauson in 1920:

Dr. Hermann Plauson, the director of the Fischer-Tropsch "Otto Traun Research Laboratories" in Hamburg, Germa during the Weimar Republic of the 1920s, drew on the aerostats made from magnesium-aluminum alloy. Some electrolytically deposited needles, further doped with radium, perforated the aerostat all through it. Patched with amalgam, the aerostat went on to produce 0.72kW and 3.4kW 300m above ground level. The power could further increased up to 81-1/2 kilowatts in 24 hours by using two balloons with a special condenser battery.

3) The Pyramidal Electric Transducer:

Inspired by the Great Pyramid of Giza (GPG), a Californian researcher, Peter Grandics says that pyramidal horn antennas suitably detect the short-pulse waveforms. Fed by the hypothesis, a pyramidal antenna, modeled on the can transfer the electrostatic discharge into a resonant circuit. While coupling into the atmosphere, it effectively converts the random impulses into an alternating current. Resultantly, it has the potential to utilize the renewable electric power.

Remarkable technology has some obstructions:

Moreover, this simple, robust and low cost electricity is much cheaper than the wind turbines. You don't have to c on the weather conditions and it's available day and night. Frequent use of this technology will definitely lead to th discharge of storm clouds and the risk of lightning strikes will get minimized. No doubt, it's in abundance at poles that doesn't deny its presence elsewhere.

Despite it's being a feasible energy source, there are certain hitches scientists need to deal with:

I. As this technology makes use of tethers, it's hazardous to aircraft. On the contrary, you can't compromise even tether's thickness since density means more output. As a direct consequence, it necessitates bigger balloon size a more maintenance.

II. Like the airplanes, the balloons are also defenseless when it comes to bad weather. So, as is the case with wind solar power, adverse weather conditions play a role here also.

III. At 200 meters or above altitudes, the amassed charge goes far beyond the wires' potential. It requires electrica insulation around the wire to avoid the energy loss.

IV. After you've exploited the device for quite sometime, the electrical potential goes down drastically. It further hu the ionization capability of the device. Therefore, it requires extending of the surface electrical field via the use of thin wires.

OLEG D. JEFIMENKO
West Virginia University

Electrostatic Motors

Their History, Types, and
Principles of Operation —

With many illustrations, of which 57 are
by
David K. Walker
of
Waynesburg College

Electret Scientific Company
Star City

UNIVERSITY OF COLORADO
DENVER CAMPUS LIBRARY

Copyright © 1973 by
Electret Scientific Company
P.O. Box No. 4132
Star City, West Virginia 26505

All rights reserved

Printed in the United States of America

This book is dedicated to an accomplished
farmer, craftsman, musician,
educator, and lawyer
MY FATHER

PREFACE

The fascinating science of electrostatics was developed mainly in the 18th century at a time when the technology and industry were too primitive to put this science to practical uses. As a result, serious research in electrostatics soon lost its momentum and, except for a few isolated efforts, was practically nonexistent during the entire 19th century. Only very recently practical aspects of electrostatics began to make their impact on the industry and economy, and the once glamorous but long forgotten science has again appeared at the focal point of serious scientific investigations.

A peculiar obstacle stands, however, in the way of many such investigations: because of the one and one half century of neglect of electrostatic explorations there is a singular lack of easily available quantitative and qualitative information on even the most basic electrostatic phenomena, techniques, and devices. Whereas experimental and theoretical data pertaining to most modern sciences are well documented and are easily retrievable from numerous reference sources, many data in the field of electrostatics must be extracted through thorough and laborious firsthand studies of old books and magazines, the very existence of which is not generally known, and which are not readily available in any of the present-day libraries or repositories.

It is precisely this kind of obstacle that the author encountered, when several years ago he started his research on electrostatic motors. Luckily, however, the obstacle soon transformed itself into a highly rewarding experience of studying various "ancient" books and periodicals and searching through electrostatic inventories of various European and American science museums. In the course of these studies he found numerous forgotten publications on electrostatic motors and found that several old electrostatic motors can actually be seen in some of the museums. (It is interesting to note that the latter motors are usually not on public display, apparently because of a lack of adequate information about their purpose and mode of operation available to museum personnel. It is also interesting to note that certain types of electrostatic motors were frequently employed in various animated toys during the second half of the 18th century, and that, although some of these toys are shown in museums, they are usually shown without an explanation of what made them move or how they were supposed to function.)

The purpose of this book is to describe the various types of electrostatic motors reported in the scientific literature between 1700 and the present, and to discuss in general terms their various design features and principles of operation. The book is written for a wide circle of readers, and care has been taken to avoid technical details that would be useful only to a small group of specialists. The readers interested in additional scientific and technical data on the various motors are referred to appropriate original publications cited in this book.

The material is presented in several chapters, each chapter describing a particular class of motors. The sequence of the chapters corresponds to the chronological order in which the earliest motor belonging to each particular chapter was invented. The sequence of material in each chapter is normally arranged also in a chronological order. The last chapter describes some of the most recent research results on electrostatic motors and extrapolates these results into the immediate future.

The author is grateful to his wife Valentina for typing the manuscript and for otherwise helping him to make the book ready for publication. He is also grateful to Dr. David K. Walker for reading the manuscript and for drawing the illustrations for the book.

Oleg D. Jefimenko

CONTENTS

Preface vii

1. What are Electrostatic Motors? 1

   Electromagnetic and Electrostatic Motors. Some Unique Properties of Electrostatic Motors. Classification of Electrostatic Motors.

2. Electric Pendulum Motors; Contact Motors. 6

   Gordon Electric Bells. Electric Swing. Electric Seesaw. Electric Perpetuum Mobile. Dailey Reciprocating Motor. Electric Racing Ball.

3. Electric Wind Motors. 22

   Gordon Electric Fly. Hamilton Electrometer. Langenbucher Electrometer. Electric Orrery. Kinnersley Electric Horse-Race. Bichat Electric Tourniquet. Corona-Kraft Levitation. Electric Wind Wheels and Turbines. King Electric Rabbit Hunt.

4. Franklin's Spark Motors and Their Descendants. 39

Franklin First Motor. Franklin Second Motor. Moore Motors. Bell Jar Motors. Another Electric Perpetuum Mobile. Dressel Nuclear Battery Motor.

5. Corona Motors and Motors Similar to Them. 56

Holtz Rotation Phenomenon. Poggendorff Motor. Christiansen Motors. Gruel Electric Tourbillion. Ruhmkorff Motor. Johnson Motors. McVay Motors.

6. Capacitor Motors. 73

Zipernowsky Motor. Van Huffel Motor. Synchronous and Asynchronous Motors. Bollée Capacitor Motors. Moore Interdigital Motor.

7. Induction Motors. 84

Arno Motor. Weiler Induction Motors. Bollée Induction Motors. Choi and Dunn Motor.

8. Liquid-Immersed Motors. 91

Weiler Liquid-Immersed Motor. Secker Motors. Quincke Motor.

9. Electret Motors. 97

Gubkin Motor. Slot-Effect Motors. Synchronous Electret Motor.

10. What to Expect from Electrostatic Motors. 106

Motors for Miniaturized Systems. New Corona Motors. Sources of Power for Electrostatic Motors. Operation of Electrostatic Motors from Atmospheric Electricity.

11. Index 120

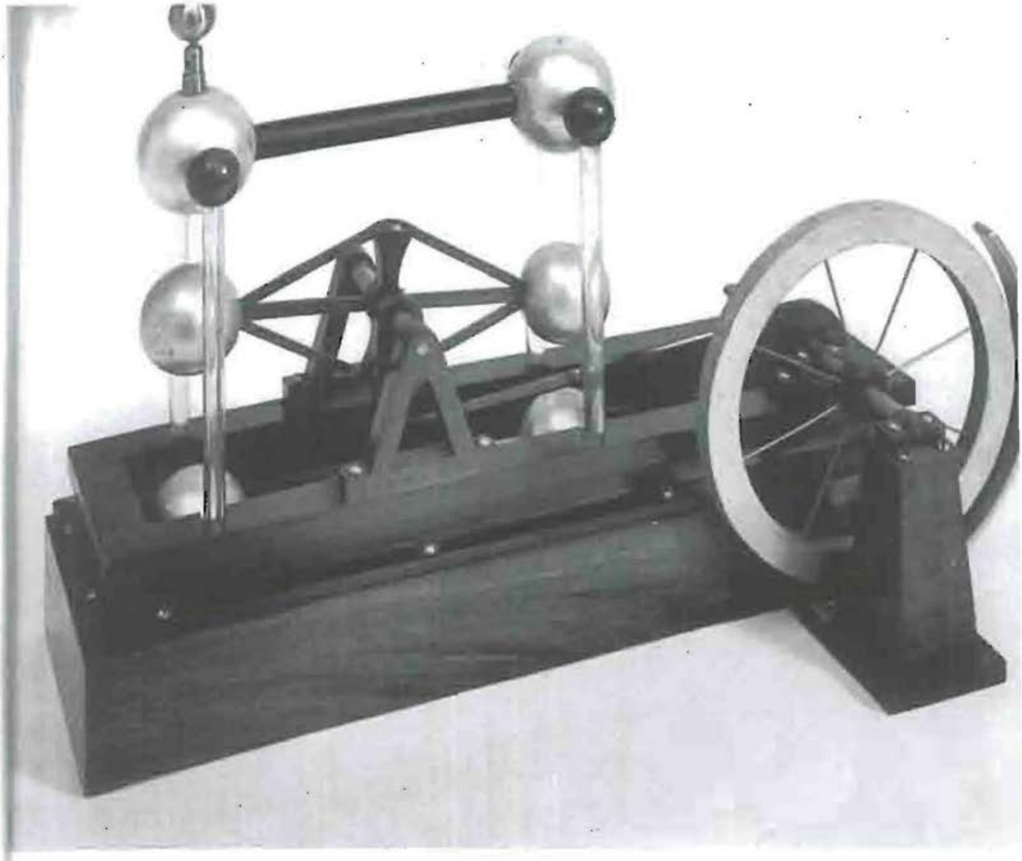

PLATE 1

Here is what H. B. Dailey wrote about this electrostatic motor which he and his father built in 1880: "Whether or not there is about the instrument and its history that which would ever give it a claim to any measure or serious interest, it is at least an electric motor most unique. A motor without magnetism, wiring, or any iron in its make up. A motor that runs by the action of the direct push and pull of the pure unconverted electricity itself".

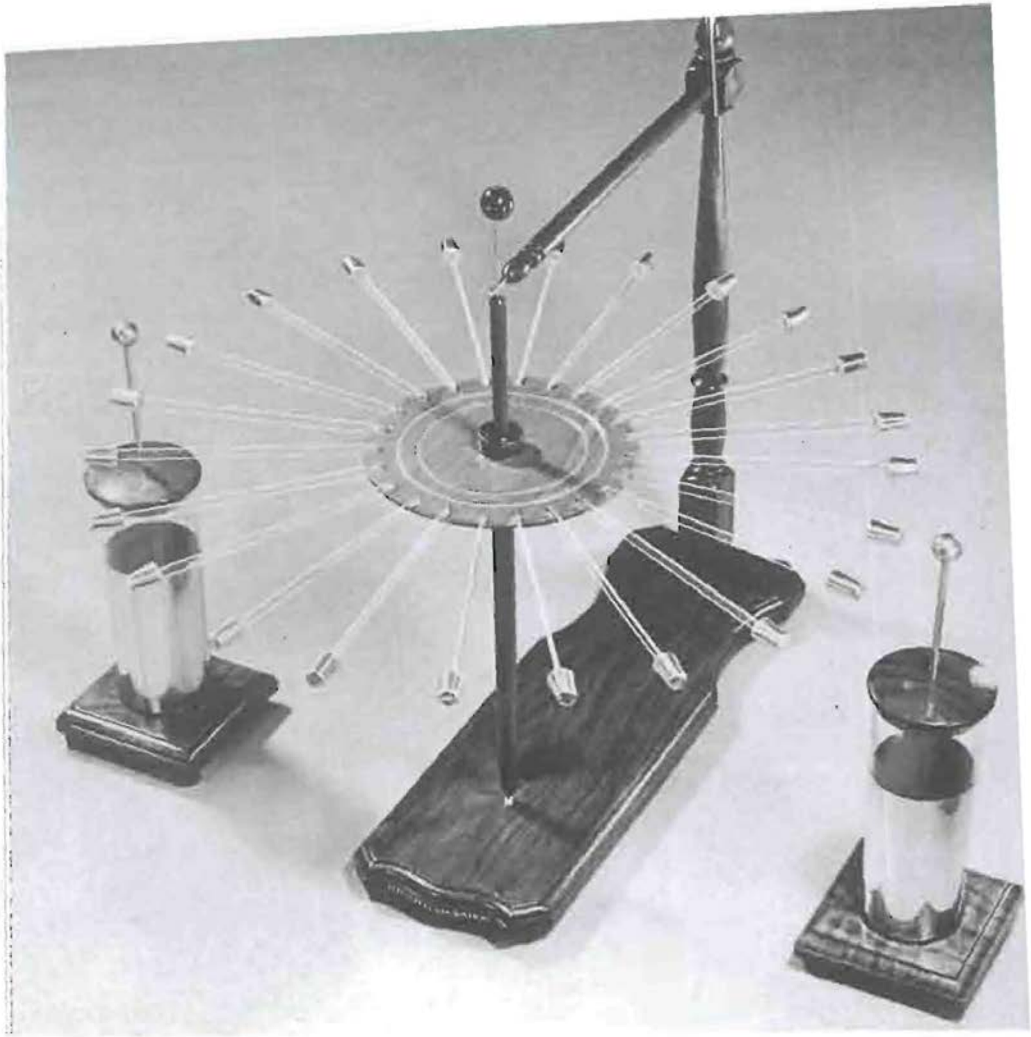

PLATE 2

It was Benjamin Franklin who in 1748 constructed the first electrostatic motors. No original drawings or models of his motors are known to exist, but his first motor must have looked very similar to this replica designed by the author for the Electret Scientific Company.

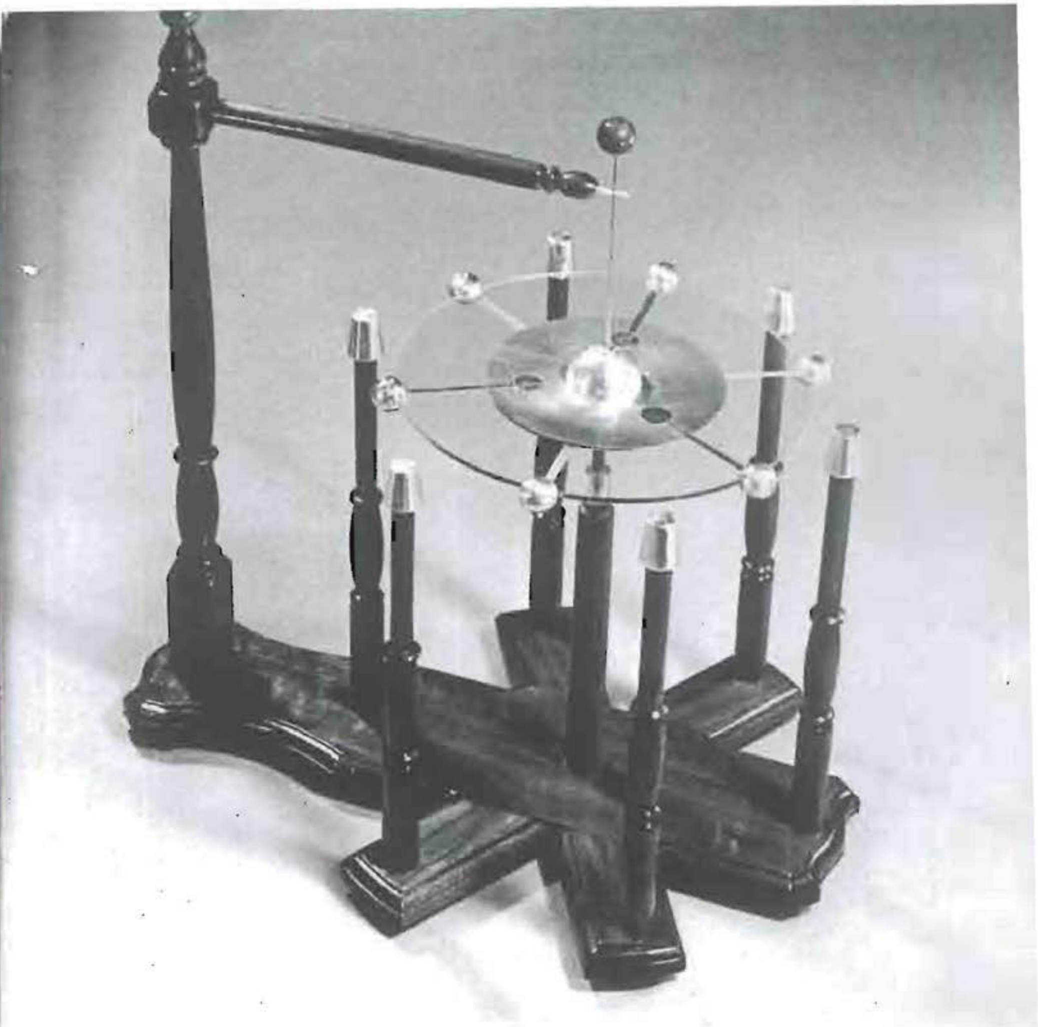

PLATE 3

Franklin's second motor probably looked very much like this replica also designed by the author for the Electret Scientific Company. Whereas the first motor operated from the electricity stored in Leyden jars, this motor operated from the electricity stored in the motor itself.

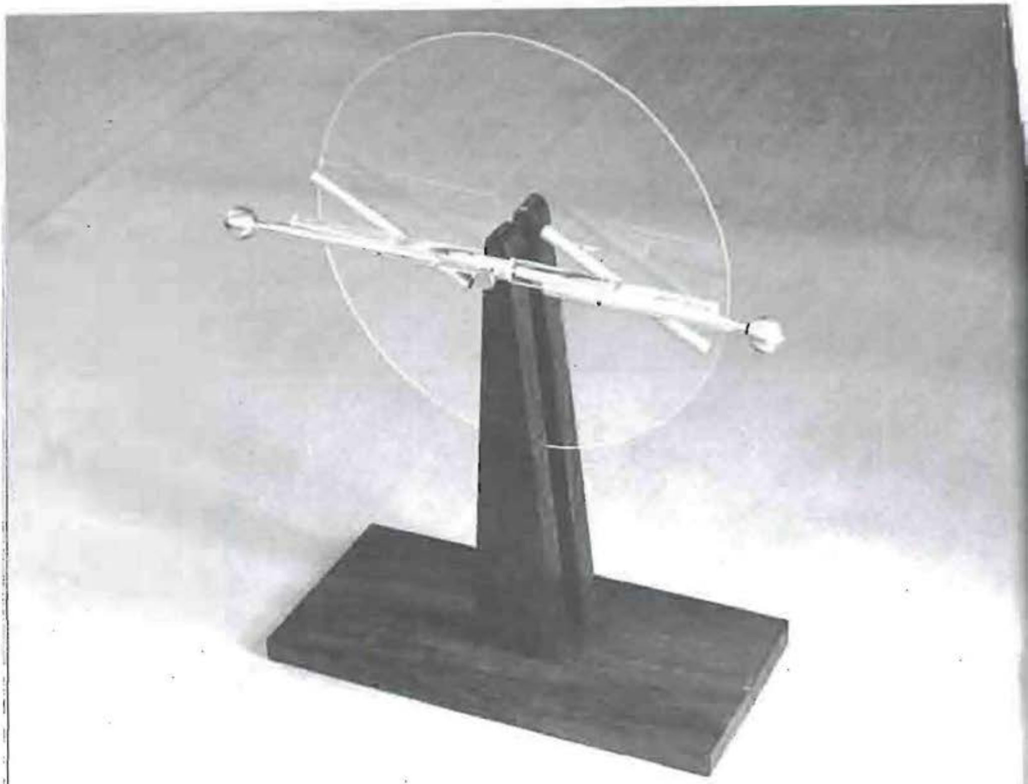

PLATE 4

The first corona motor was designed around 1869 by the German physicist Poggendorff. Poggendorff made a thorough study of the motor, but failed to appreciate its possibilities in the mistaken belief that no sources of electricity could supply enough power to any electric motor to make it do useful work. This simplified version of Poggendorff's motor was built by the author.

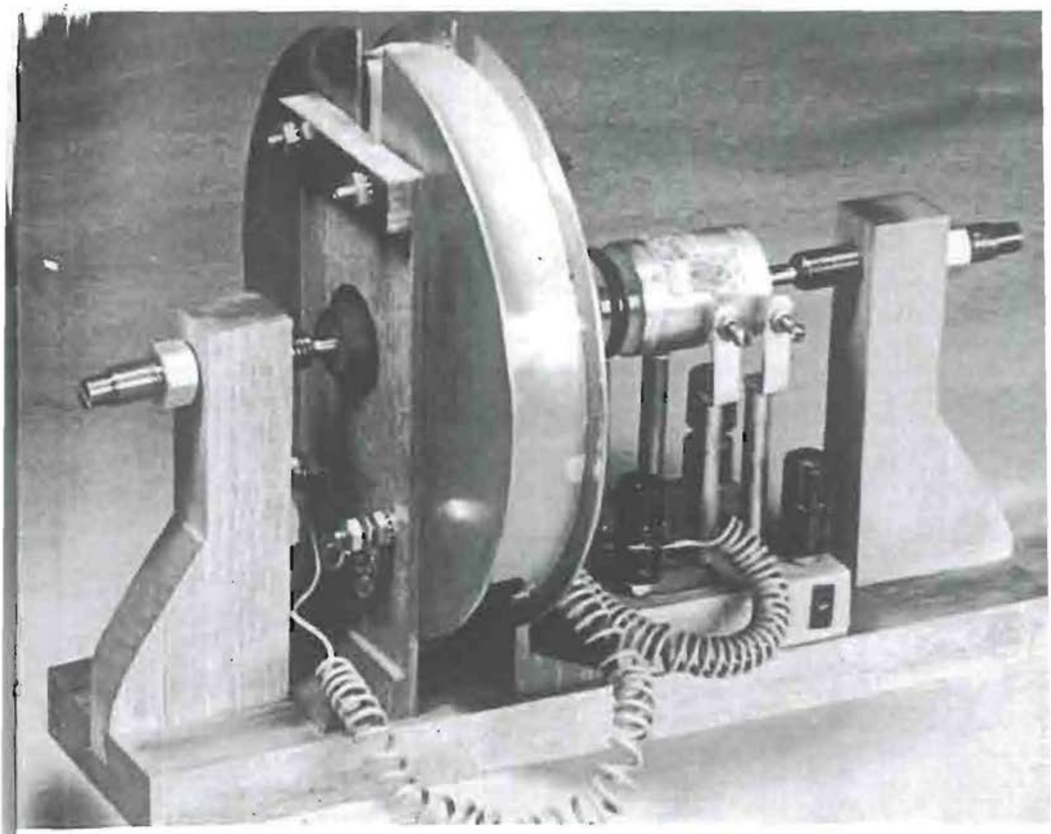

PLATE 5

The electret—a permanently electrized dielectric—is essentially a product of the 20th century. When an electret is placed between slotted electrodes to which a voltage is applied, the electret experiences a force. This motor designed by the author utilizes the slot effect. The rotor of the motor is a carnauba wax electret made of two oppositely polarized half-disks. The motor was built in 1966.

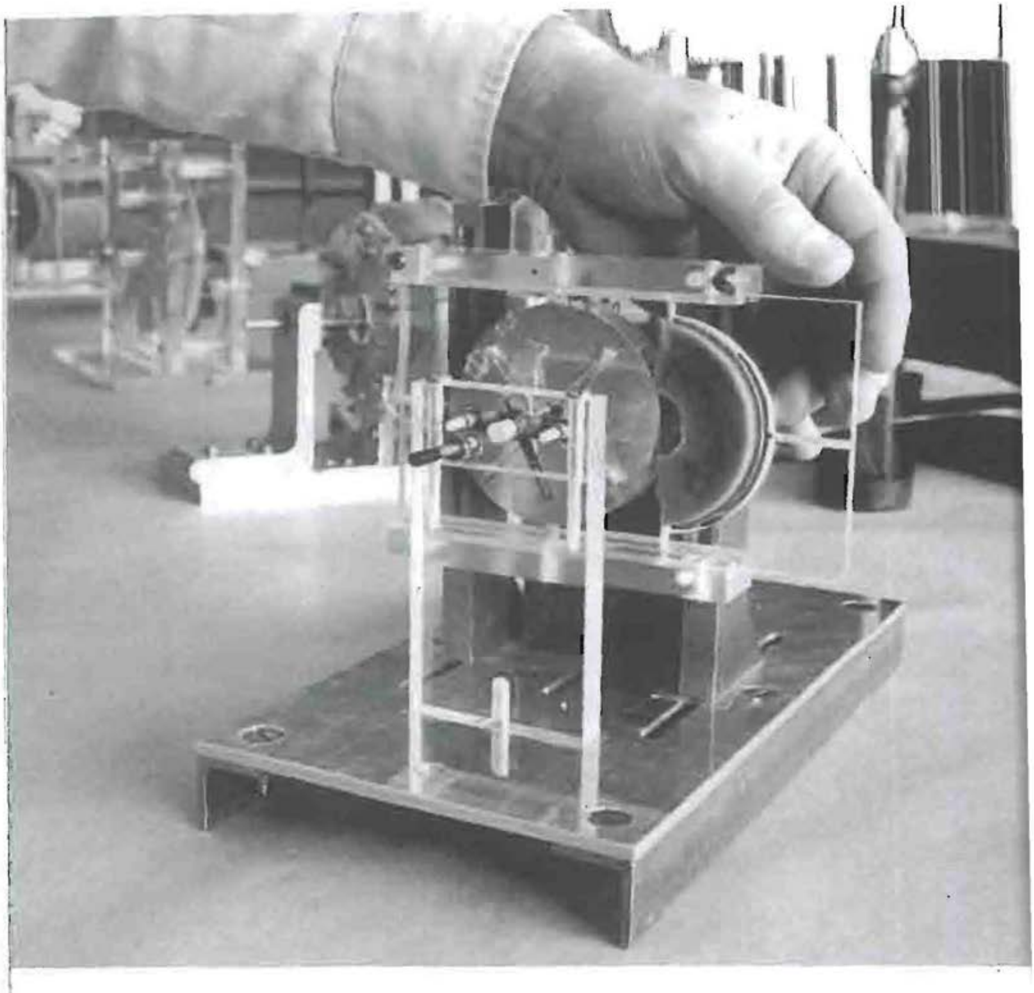

PLATE 6

A more practical design of an electret motor incorporates stationary electrets and rotating electrodes. The electrets can be easily withdrawn from the motor for servicing or replacement. The motor operates from approximately 60 volts dc. This particular motor was the first ever operated from atmospheric electricity.

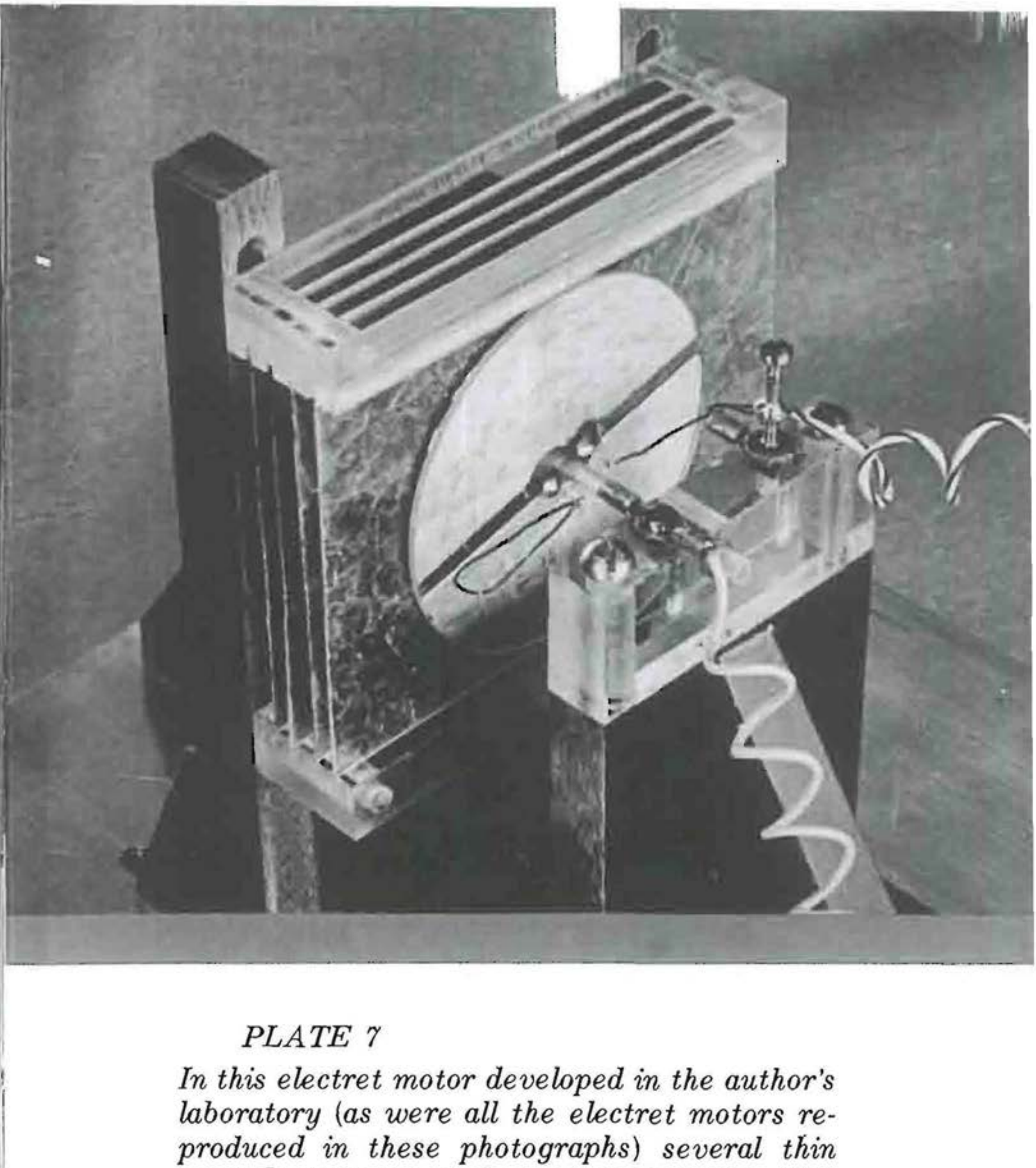

PLATE 7

In this electret motor developed in the author's laboratory (as were all the electret motors reproduced in these photographs) several thin mica electrets are used as active elements. This arrangement of electrets makes good use of the available space and allows one to construct relatively powerful electret motors.

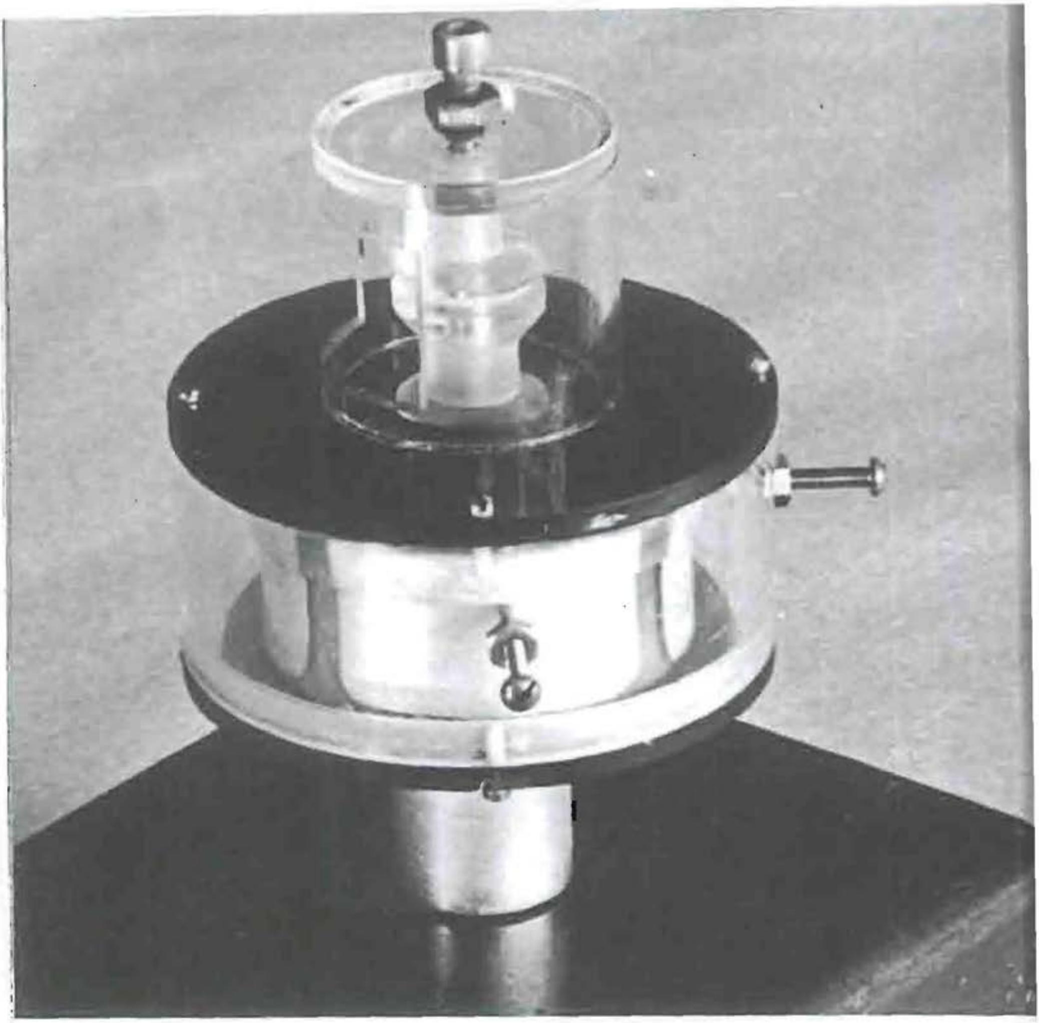

PLATE 8

This "hoop" electret motor built in 1967 by D. K. Walker has a stationary cylindrical electret. The rotor is made of four pairs of bent aluminum plates enclosing the electret. The motor requires no commutator. The motor has been operated for several years without any servicing.

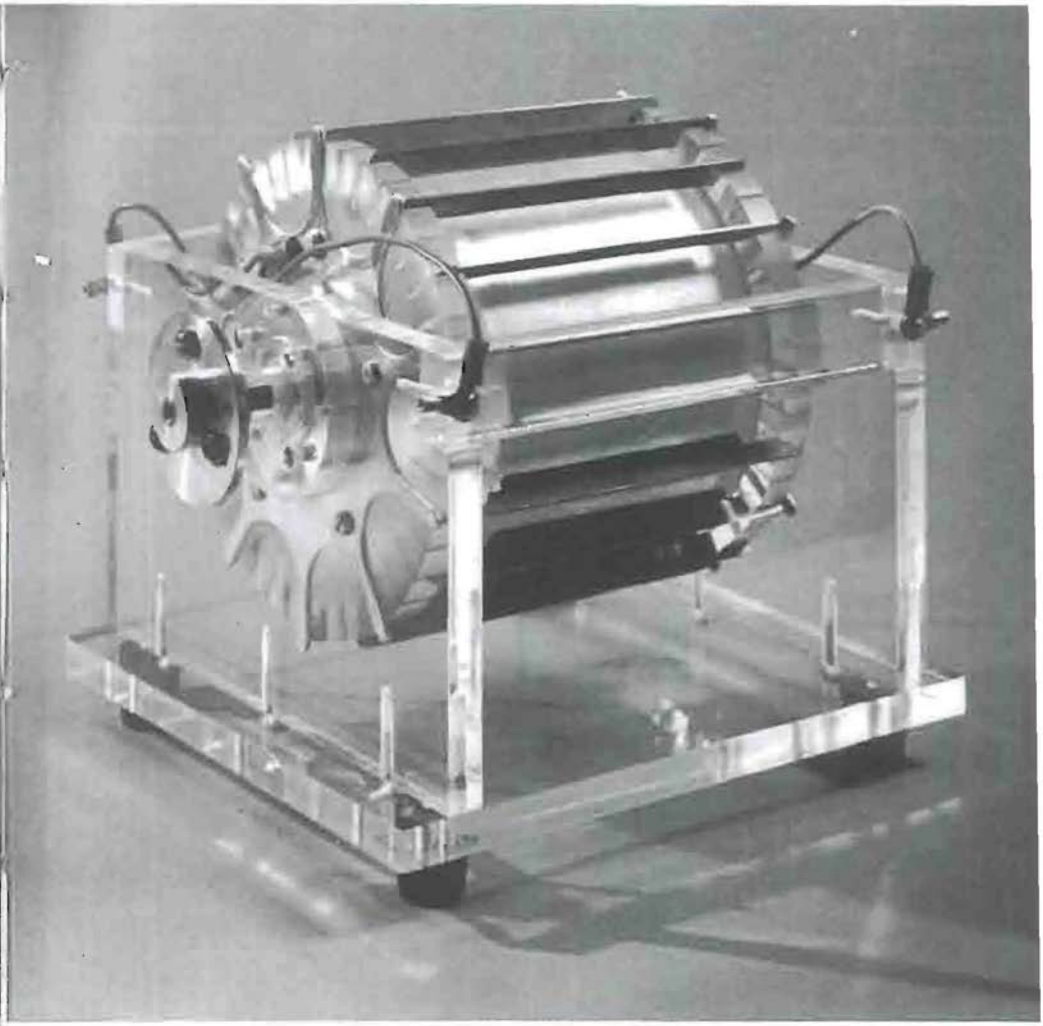

PLATE 9

Corona motors are probably the most promising electrostatic motors. This 0.1-hp modern version of the Poggendorff motor has a cylindrical rotor instead of a disk. The motor operates from a 6000-volt power supply as well as from an earth field antenna.

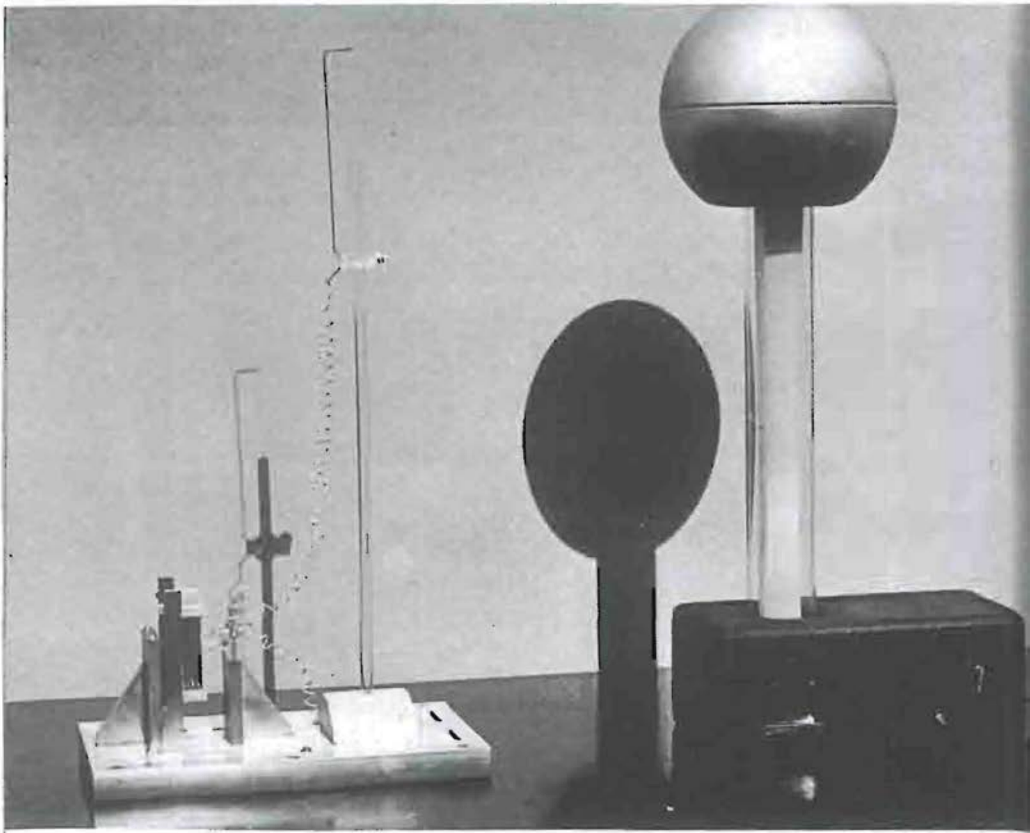

PLATE 10

The operation of electrostatic motors from antennas can be demonstrated with this lecture-room apparatus. The Van de Graaff generator produces electric charges in the air. The sharp-point antenna collects the charges from the air and delivers them to one terminal of an electrostatic motor whose other terminal is grounded.

PLATE 11

The author and D. K. Walker attempt to operate an electret motor from a 20-foot pole antenna in front of the Physics Building of the West Virginia University. The tall building screened the atmospheric field, and the motor did not run, although it operated very well from the same antenna in the nearby unobstructed parking lot.

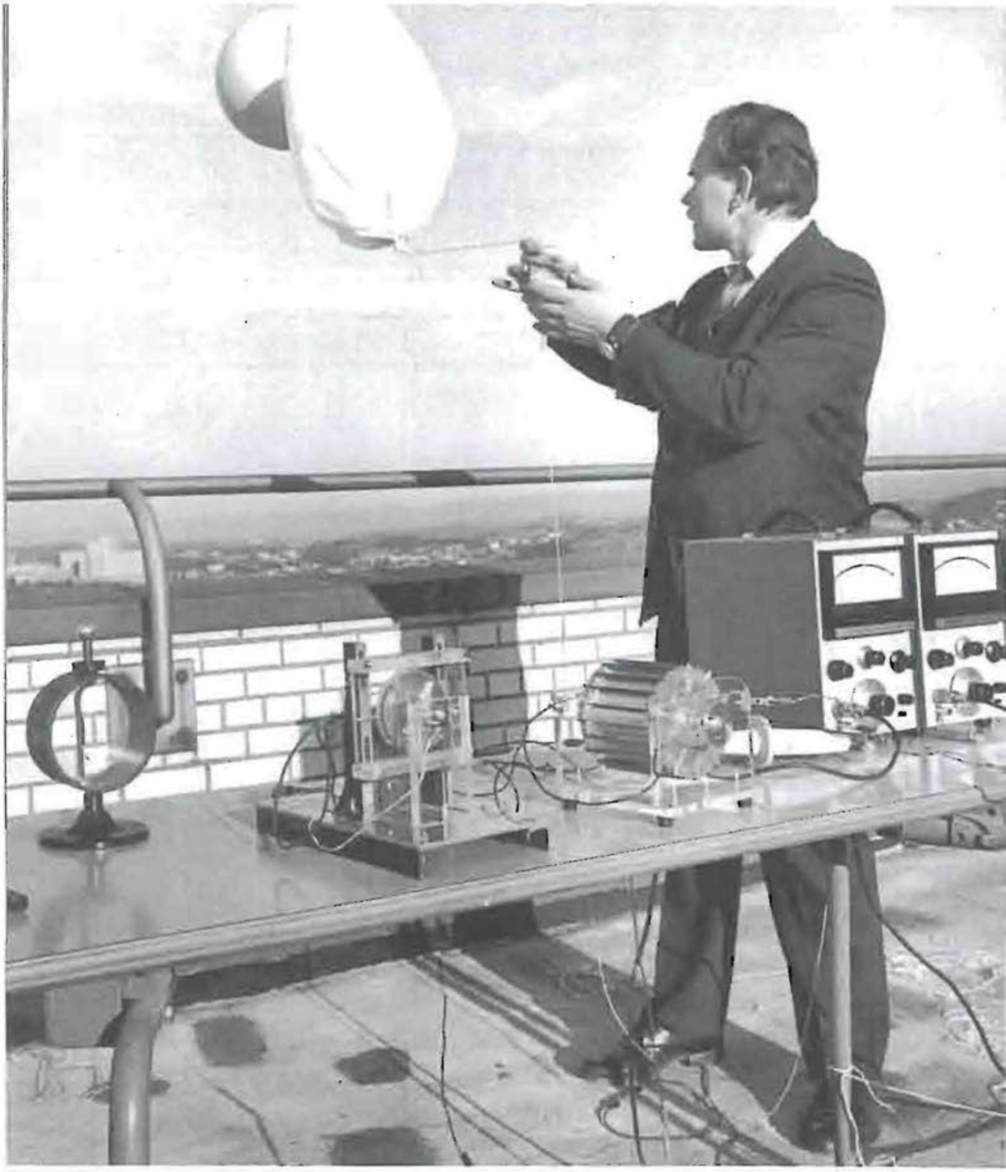

PLATE 12

The author operates electrostatic motors by means of an airborne antenna.

1

WHAT ARE ELECTROSTATIC MOTORS?

Conventional electric motors create mechanical motion as a result of magnetic forces acting upon electric currents. These motors are properly called *electromagnetic motors*. There is, however, another type of electric motor, in which the motion is created as a result of electric, or "electrostatic", forces acting between electric charges. Motors of this type are called *electrostatic motors*.

It is interesting to note that in nature the electrostatic forces are much stronger than the magnetic ones. There are many ways in which this can be demonstrated. For example, although a considerable effort may be needed to separate a magnet from an object attracted by it, a much greater effort is needed to break the magnet; this is because the magnet and the object are held together by magnetic forces, while the molecules of matter in the magnet (as well as in any other body) are held together by electrostatic forces.

Why, then, do we not use the electrostatic forces in our electric devices, and in motors in particular, at least on as wide a scale as we use the "inferior" magnetic forces? There are two main reasons for that. First, it is difficult to establish appreciable concentrations of electric charges without causing an electric breakdown in the medium surrounding or supporting the charges (although with modern insulating materials and techniques this difficulty becomes progressively less serious). Second, powerful electrostatic devices require voltages of many kilovolts for their operation, and until recently such voltages could not be produced conveniently and economically.

Motors of great power, however, are not the only ones needed. Equally important are low-power motors capable of performing various special tasks. In this respect electrostatic motors may compete successfully with their electromagnetic cousins even now.

Although electrostatic motors are not yet widely known or used, they already hold at least five very impressive records as compared with the electromagnetic motors:

1. The first electric motor ever invented was an electrostatic one. It was built about 100 years before the first electromagnetic motor was conceived.

2. The electric motor that operated without interruption longer than any other electric motor was an electrostatic one. This was a pendulum-type motor, known as the "electric perpetuum mobile", installed at the University of Insbruck, Austria, in 1823. Since then it operated continuously at least until 1909, powered by a Zamboni pile (an early high-voltage battery).
3. Electrostatic motors have been operated from voltages in excess of 10^5 volts, which is much higher than the voltages suitable for operating electromagnetic motors.
4. Electrostatic motors have been operated by using currents smaller than 10^{-9}amp, which is much less than the currents needed to operate electromagnetic motors.
5. The first electric motor that operated directly from the atmospheric electricity was an electrostatic motor. None of the presently available electromagnetic motors can operate directly from this source.

Even this short list reflecting some of the more obvious peculiarities of electrostatic motors shows quite clearly that electrostatic motors possess a number of unique properties. These properties undoubtedly will make electrostatic motors increasingly more important for the science, engineering, and technology of the future.

Many different types and designs of electrostatic motors are possible. It is customary to classify electrostatic motors in accordance with some prominent feature of their mode of operation or some prominent feature of their design. Thus, in reference to the techniques used for delivering electric charges to the active part of a motor one speaks of *contact motors*, *spark motors*, *corona motors*, *induction motors*, and *electret motors*. In reference to the medium in which the active part of a motor is located one speaks of *liquid-* or *gas-immersed motors*. In reference to the material and design of the active part of a motor one speaks of *dielectric motors* (the active part is made mainly of dielectric material) and *conducting-plate*, or *capacitor*, *motors* (the active part is made mainly of metal and resembles a variable capacitor). Finally, in reference to the rate of rotation of a motor relative to the period of the applied voltage (for ac operated motors) one speaks of *synchronous motors* and *asynchronous motors*.

This classification of electrostatic motors does not make it possible, however, to specify uniquely each individual motor. Also, two or more different operation modes are usually possible for most electrostatic motors (for example, certain motors can be operated at will as contact, spark, or corona motors). Therefore an assignment of a particular motor to one or another of the above types or categories is frequently more or less arbitrary. This applies, of course, also to the present book, where certain motors described in the chapters that follow could be equally well discussed under different chapter headings, and where the inclusion of certain motors in a particular chapter was occasionally dictated by the fact that such an inclusion resulted in a more coherent development of the subject matter.

2

ELECTRIC PENDULUM MOTORS; CONTACT MOTORS

If "electric motor" is understood to be a device converting electric energy into a continuous mechanical motion, then the first two electric motors were invented in the early 1740's by Andrew Gordon, a Scottish Benedictine monk and professor of philosophy at Erfurt, Germany.[1] Gordon's first motor was a device known as the "electric bells" (his second motor will be described in the next chapter). The device and its operation were as follows. A metallic clapper (pendulum) was suspended by a silk thread between two oppositely charged bells (Fig. 1). From an initial contact with one of the bells the clapper acquired a charge of the same polarity as that of the bell. Due to repulsion of like charges and attraction of opposite charges the clapper was then repelled by this bell and attracted by the second bell. As the clapper struck the second bell, it gave off its initial charge and acquired a charge of the same polarity as that of the second bell. Then the clapper was repelled by the latter and attracted by the first bell, which it struck again, and so on.

Numerous variations of Gordon's bells have been described by later authors, mostly as devices for lecture-room demonstrations of electrostatic forces. Such devices are widely used for this purpose even now.

An ingenious application of Gordon's bells was made by Benjamin Franklin in 1752. He connected the bells to an insulated lightning rod as a warning device "to give notice when the rod should be electrified".[2] (It appears that at that time neither Franklin nor any other scientist suspected how dangerous this "warning" device could be. The extreme danger of experiments with insulated lightning rods became clear a year later, when in 1753 the Russian physicist George Wilhelm Richmann was killed by lightning that entered his laboratory through such a rod, as he approached it in order to measure its electrification with a specially constructed electrometer.)

In a later modification of Gordon's invention one of the two bells was mounted on a Leyden jar (Fig. 2); the bells would then ring for as long as there was enough electric energy stored in the jar to move the clapper. (Franklin described the operation of an electric pendulum powered by a Leyden jar in 1747; see Ref. 2, p. 189.)

Sometimes Gordon's bells were made as a set

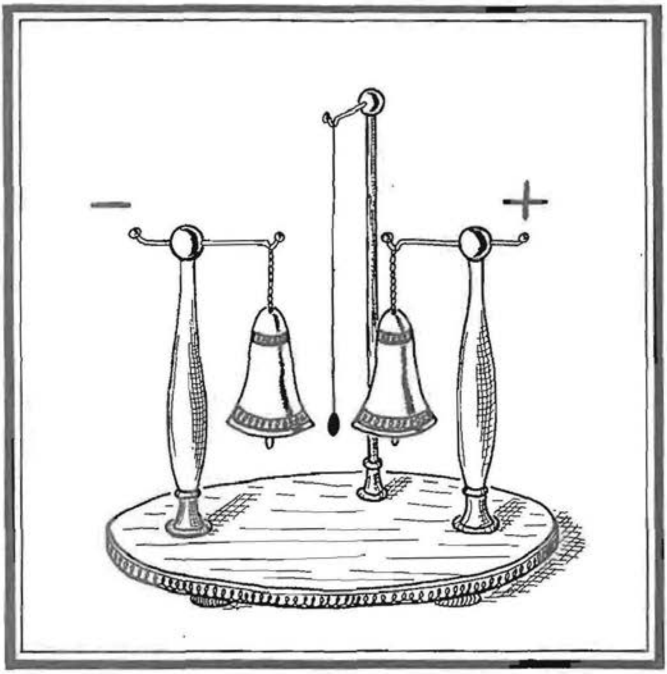

FIGURE 1

The so-called "electric bells" constituted the first device that converted electrical energy into a continuous mechanical motion. They were invented in about 1742 by Andrew Gordon, a professor of philosophy at Erfurt, Germany.

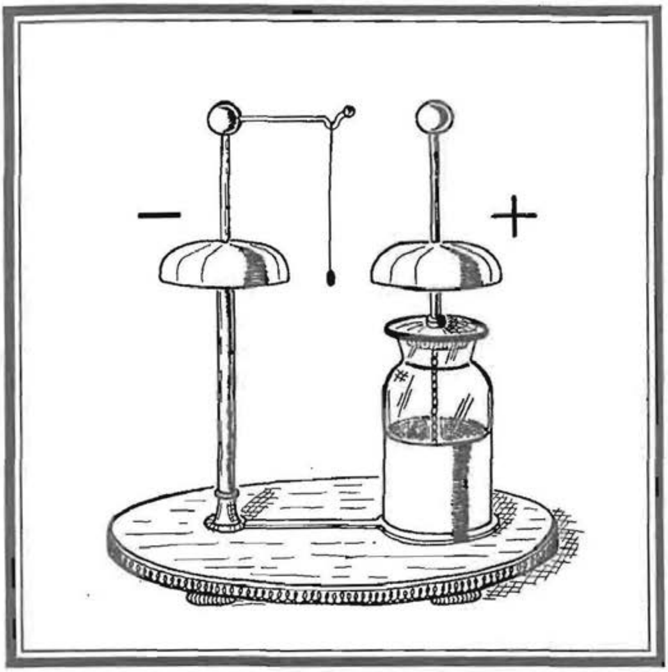

FIGURE 2

In a later modification of Gordon's invention one of the two bells was mounted on a Leyden jar. The bells would then ring for as long as there was enough electrical energy stored in the jar to move the clapper.

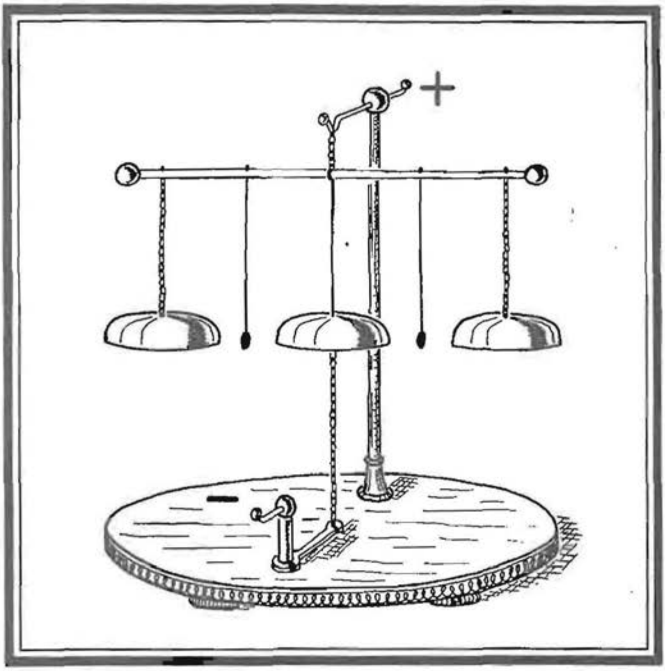

FIGURE 3

Sometimes Gordon's bells were made as a set of three bells and two clappers, all suspended from a horizontal bar, the central bell being insulated from the bar and grounded by means of a light chain.

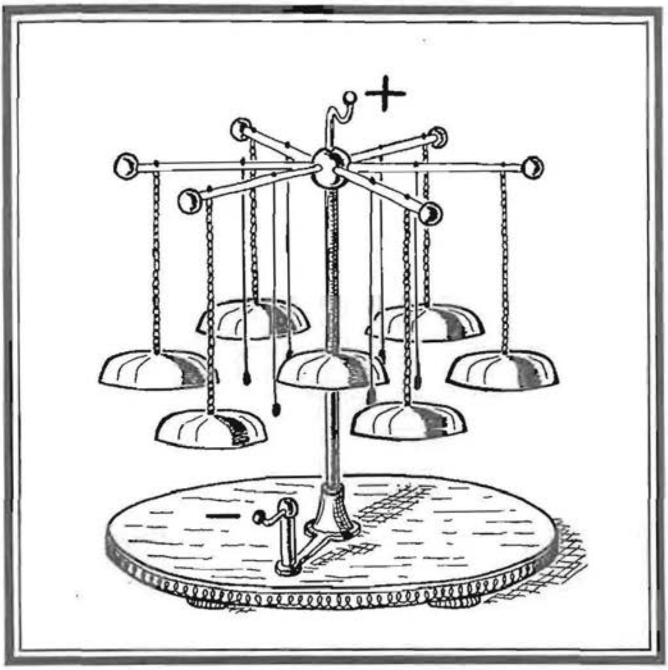

FIGURE 4

Another modification of Gordon's bells was a set of bell chimes arranged on a wooden base in a circular formation around the central bell. A separate clapper was then present for each of the outside bells.

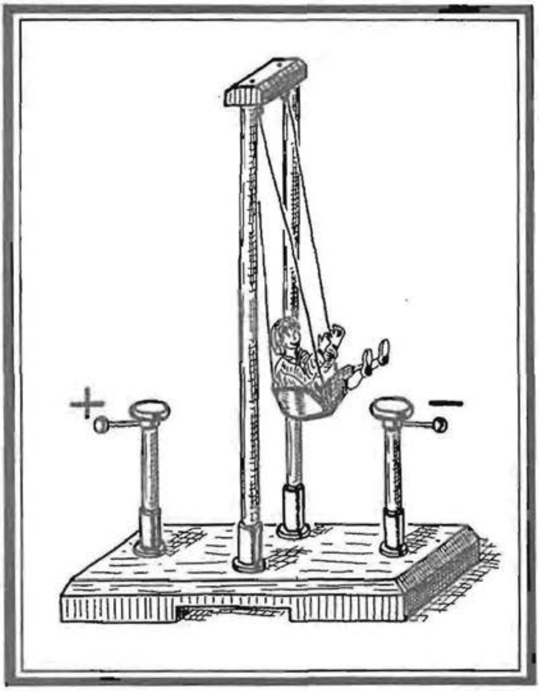

FIGURE 5

The electrical and mechanical principle of Gordon's invention could be used not only for ringing bells but for creating a mechanical motion in general. In this "electric swing" this principle was used to operate an animated toy.

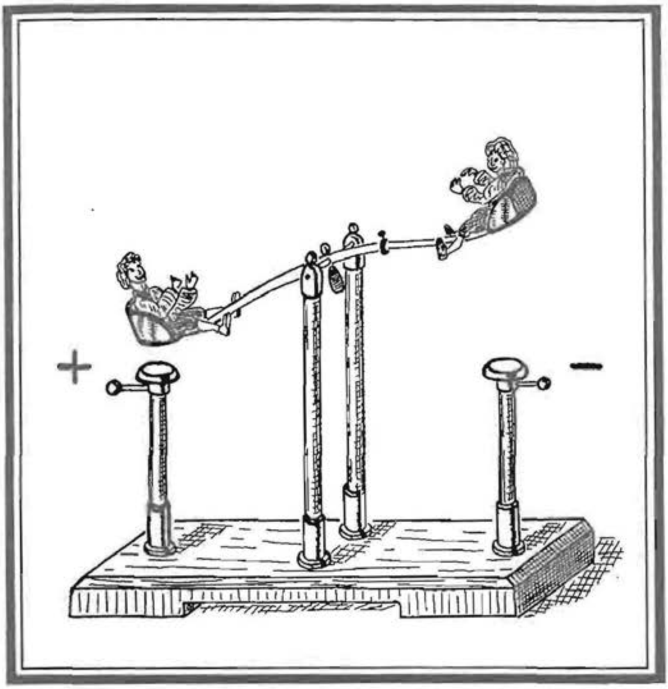

FIGURE 6

Various electrostatically operated toys were invented in the 18th and 19th centuries. This "electric seesaw" was basically a conducting bar oscillating about a horizontal axis on the same principle as the clapper of Gordon's bells.

of three bells and two clappers, all suspended from a horizontal bar, the central bell being insulated from the bar and grounded by means of a light chain (Fig. 3). Another modification of Gordon's bells was a set of bell chimes arranged in a circular formation around the central bell (Fig. 4); a separate clapper was then present for each of the outside bells, each clapper moving between one of the latter and the central bell (the central bell was usually grounded, the outside bells were connected to an electrostatic generator).

The electrical and mechanical principle of Gordon's invention could be used, of course, not only for ringing bells but also for a variety of other purposes. In particular, this principle was later used in two electrically operated toys: the "electric swing" and the "electric seesaw", which utilized, respectively, an insulated pendulum made as a swing (Fig. 5) and an insulated conducting swing bar capable of oscillating about a horizontal axis (Fig. 6).[3,4]

Except for Franklin's bells, all of the above devices were designed for operation from an electrostatic generator or from Leyden jars charged by such a generator. However, in 1806 a high-voltage chemical battery ("dry pile", later known as the "Zamboni pile") was invented by Georg Behrens, and in 1810 this pile was adapted by Giuseppe Zamboni to operate primitive electrostatic motors.[5] Inasmuch as a Zamboni pile could remain active for many years, the little motors that operated from it were known as the "electric perpetuum mobile". An example of such a perpetuum mobile is shown in Fig. 7 (see Chapter 4 for another design). The apparatus consisted of a light rigid insulated pendulum, pivoted just above the center of gravity, with a light conducting ring at the top. The ring was located between the two knobs of a Zamboni pile. When the ring contacted one of the knobs, it acquired a charge from the knob and was repelled from this knob and attracted to the second knob, and so on, just like the clapper in Gordon's bells. One such perpetuum mobile is reported to have operated without interruption for at least 86 years.[6]

The most sophisticated device derivable from Gordon's electric bells was probably the reciprocating electrostatic motor (Fig. 8 and Plate 1) built in 1880 by Howard B. Dailey and Elijah M. Dailey.[7] The motor, which is now at the Museum of History and Technology of the Smithsonian Institution, was described by one of its builders as follows:

"This machine operates by the direct action of static electric attractions and repulsions. It is constructed entirely of fine wood, glass and hard rubber, there being no magnetic materials used. The flywheel is of laminated, soft wood and runs in journal bearings of very small diameter. The moving balls, mounted on the walking beam of vulcanite, are made of wood, hollowed out so that the walls are about 2 millimeters thick.

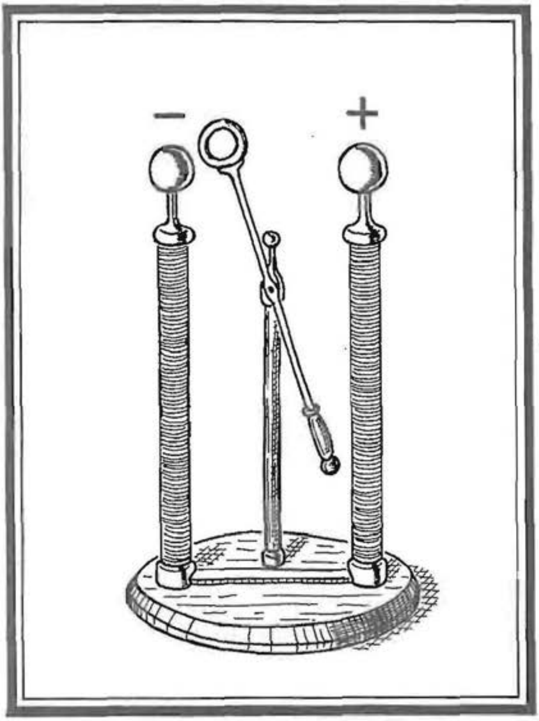

FIGURE 7

This electric pendulum operated from a Zamboni pile, an early high-voltage battery. Since the pile remained active for years, the pendulum was known as an "electric perpetuum mobile".

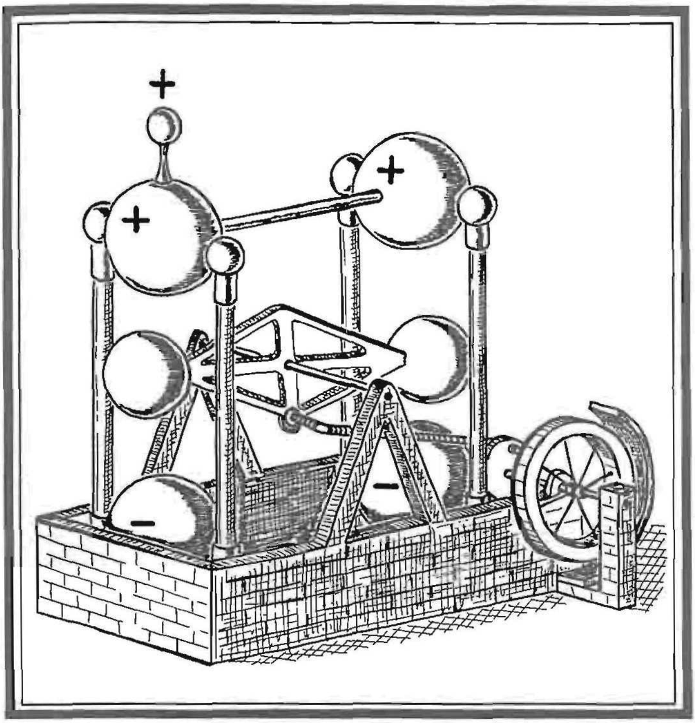

FIGURE 8

The most sophisticated device derivable from Gordon's bells was probably this reciprocating electrostatic motor built in 1880 by H. B. Dailey and E. M. Dailey.

They are covered with aluminum foil for static conductivity. The stationary balls are of solid wood.

"To operate the engine the stationary balls are charged with electricity from a static electric generator, such as a Holtz machine, the upper balls being connected through the brass ball to one pole of the machine while the lower stationary balls are connected through the binding post on the bed frame to the opposite pole of the machine. Under proper conditions, when charged, the engine will make about 375 revolutions per minute.

"The walnut base upon which the engine is mounted is 14" long, 4" wide and 1¾" thick. The movable balls are about 1½" in diameter; the upper stationary balls are 1¾" in diameter; and the lower stationary balls, 1½". The four glass rods, mounted vertically, are about 6" high and spaced 6" apart along the bed. The diameter of the flywheel is 5¾". It is gilded and has small wire spokes. The connecting rod is 7" in length."*

The electric principle of Gordon's bells could be used also for producing a continuous unidirectional motion. A fascinating device of this type was the "electric racing ball", or the "electric planetarium", an excellent description of

---

*The author is grateful to Mr. Elliot N. Sivowitch, Smithsonian Institution Museum Specialist, for kindly providing this description and the photograph of Dailey's motor.

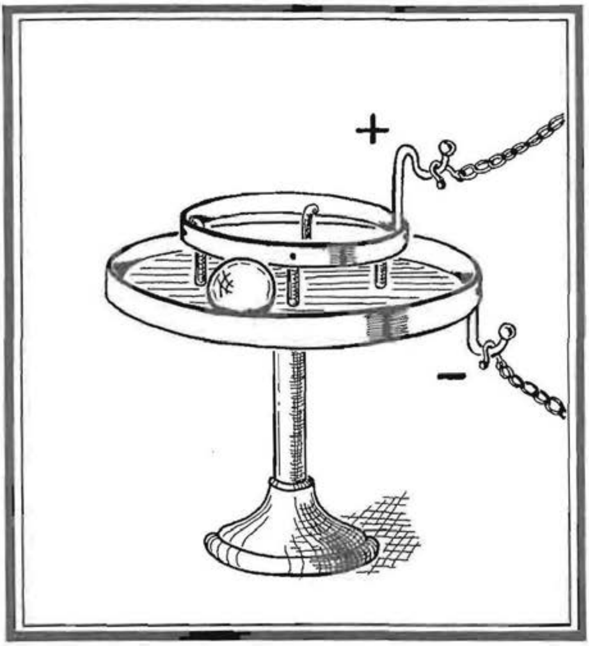

FIGURE 9

The electric principle of Gordon's bells could be used also for producing a continuous unidirectional motion. This "electric racing ball" was propelled along a conducting hoop by the same mechanism that moved the clapper in Gordon's bells.

which has been provided by M. Guyot.[8] The apparatus consisted of an insulated horizontal conducting hoop positioned above a conducting disk having a vertical rim concentric with the hoop (Fig. 9). The arrangement constituted a track for a light sphere of glass placed on the disk between the hoop and the rim. When the hoop and the disk were connected to an electrostatic generator, the sphere ran along this track due to attraction and repulsion exerted by the hoop and the disk upon the points of the sphere that acquired charges from contact with the disk and the hoop (the rolling of the sphere usually had to be initiated with a light push by hand, but once the sphere had started to run, it continued running for as long as the electrostatic generator was in operation).

Since the moving elements in the above described motors were charged by contact, the motors may be classified as "contact motors". This method of charging, however, is neither the only possible nor the most expedient one. Therefore contact motors will probably always be subordinate to most of the motors described in the following chapters. In this connection it may be useful to note that, upon a closer examination of the operation of Gordon's bells and other similar devices, one recognizes that the charging of the moving elements in these devices usually occurs by means of a spark that jumps from a stationary electrode (bell) to a moving element (clapper) as soon as the latter comes close enough to the former. Thus the actual physical contact between the moving and stationary components in these devices is not at all necessary for their operation. In fact, such a contact is harmful insofar as it results in energy losses from impact and friction.

REFERENCES

1. P. Benjamin, *A History of Electricity*, John Wiley and Sons, New York, 1898, pp. 506, 507.
2. J. Sparks, Ed., *The Works of Benjamin Franklin*, Whittemore, Niles, and Hall, Boston, 1856, Vol. 5, p. 301.
3. F. M. Avery, *Modern Electricity and Magnetism*, Sheldon and Company, New York, 1885, p. 241.
4. A. Neuburger, *Ergötzliches Experimentierbuch*, Ullstein & Co., Berlin, 1920, pp. 434-436.
5. Gilbert, *Annalen der Physik*, Ser. 1, Vol. 49, pp. 35-46, (1815).
6. Müller-Pouillet, *Lehrbuch der Physik und Meteorologie*, Friedrich Vieweg und Sohn, Braunschweig, 1909, 10th edn., Vol. 4, p. 338.
7. H. B. Dailey, *Modern Electrics*, Vol. 5, pp. 916-917 (1912).
8. M. Guyot, *Nouvelles Recreations Physiques et Mathematiques*, Gueffier, Paris, 1786, Vol. 1, pp. 272-274 and Plate 27.

3

ELECTRIC WIND MOTORS

As already mentioned in the preceding chapter, Andrew Gordon invented two electric motors. His second motor was the "electric fly",[1,2] also known as the "electric whirl", "electric pinwheel", or "electric reaction wheel" (Fig. 10). The electric fly consisted of one or more light metal arms with sharp-point ends bent at right angles to each arm and in the same circumferential direction. The fly was pivoted at its center on an insulated needle. When the needle was connected to an electrostatic generator, a corona discharge* took place from the sharp points of the fly. The air near these points became charged with charges of the same polarity as that of the fly and was then repelled from the points due to repulsion of like charges (the resulting motion of the air is known as the "electric wind"). Similarly, the points themselves were repelled from the charges in the air, and the fly rotated therefore in the opposite sense to that in which the points were directed.

---

*Corona discharge is a spontaneous electric conduction in a gas (air) originating from sharp conducting bodies whose electrostatic potential relative to the ground is 3000 volts or higher.

Also the electric fly was used for ringing bells or chimes. This was accomplished by suspending from the fly a clapper, which, as the fly rotated, struck in turn several bells positioned in a circular formation underneath the fly (Fig. 11).[3]

In 1760, Hamilton, professor of philosophy at Dublin, suggested to use a similar device as an electrometer.[4] In this electrometer small weights were suspended from the fly. As the fly turned, the weights were deflected from the vertical direction by centrifugal forces. This deflection was a measure of the fly's speed and, hence, of the strength of the source to which the fly was connected. A more sophisticated electrometer utilizing the electric fly was described by Jakob Langenbucher,[5] a silversmith at Augsburg, Germany, who spent a large part of his time and fortune on improving electrostatic generators and studying electrostatic phenomena. In this electrometer the fly was mounted on a thin vertical cylinder. A fine string was fastened to the cylinder. The string passed over a small pulley and had a light pan attached to its free end. A small weight was placed on the pan. As the fly turned, the string

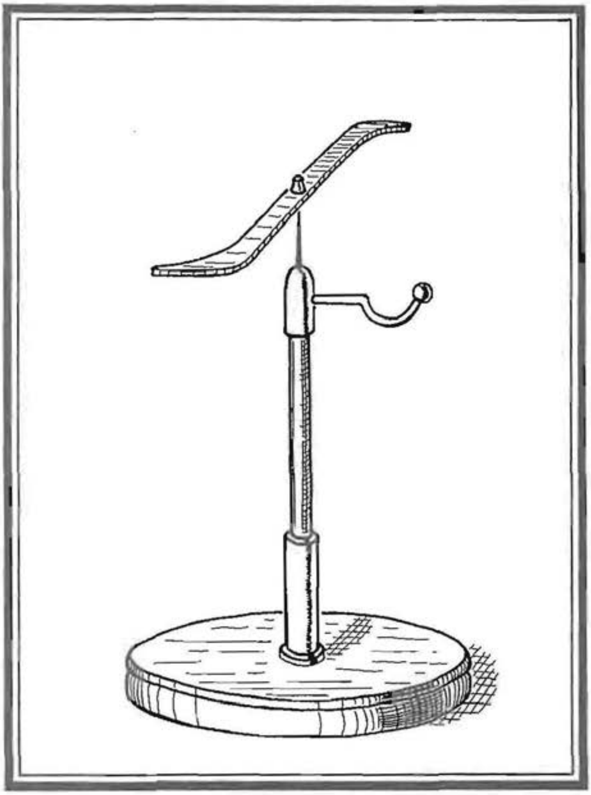

FIGURE 10

Andrew Gordon invented two electric motors. His second motor was the so-called "electric fly".

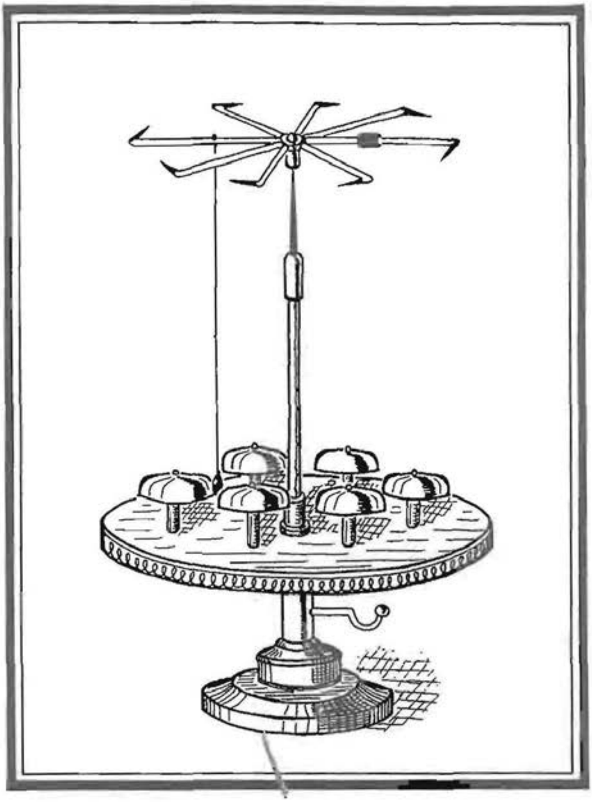

FIGURE 11

Also the electric fly was used for ringing bells with the aid of a clapper suspended from it.

wound itself on the cylinder lifting the pan and the weight. The maximum weight that could be lifted by this electrometer was a measure of the strength of the source from which the fly was operated.

The principle of the electric fly was later used to operate an orrery depicting the orbital motions of the Sun, the Moon, and the Earth (Fig. 12).[6] In such an orrery, also known as the "electrical tellurium", the longer bent arm carried a large sphere at one end, while at the other end it carried a similar but smaller bent arm, each end of which also carried a small sphere. The three spheres represented the Sun, the Moon, and the Earth. The arms were pivoted on sharp points, and each arm had a perpendicular sharp point in the horizontal plane. When the instrument was connected to an electrostatic generator, the two systems rotated about their points of support, the smaller system, being lighter, rotated more rapidly than the larger one. The relative masses of the spheres were adjusted to make the Earth-Moon system rotate twelve times for every single rotation of the Sun.

An interesting variation of the electric fly was constructed in 1761 by Ebenezer Kinnersley, a Philadelphia school teacher and close friend of Benjamin Franklin. Kinnersley called his device the "electrical horse-race".[7] This was a cross formed by two pieces of wood of equal length (probably about 18 inches long) supported horizontally on a central pin. A light figure of horse

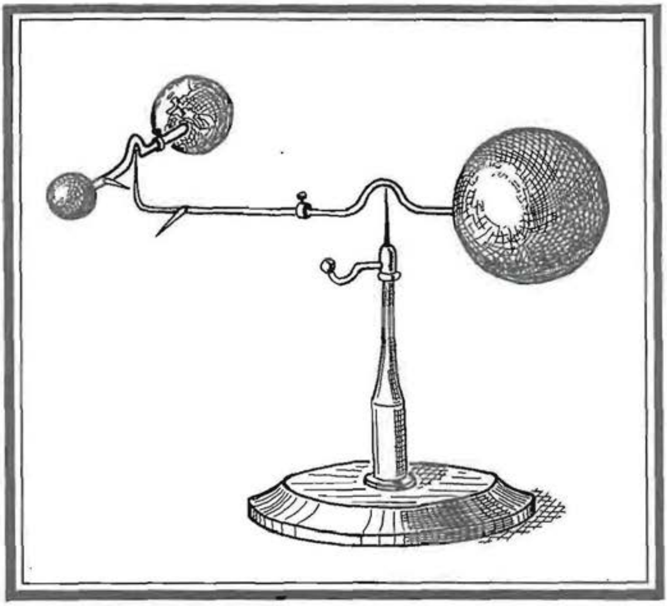

FIGURE 12

The principle of the electric fly was later used to operate an orrery depicting the orbital motions of the Sun, the Moon, and the Earth. The masses of the spheres were adjusted to make the Earth-Moon system rotate twelve times for every single rotation of the Sun.

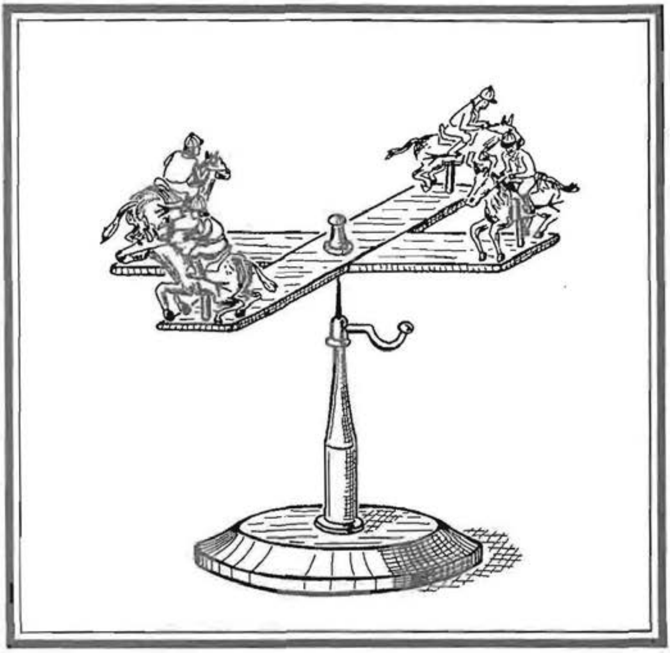

FIGURE 13

An interesting variation of the electric fly was constructed in 1761 by Ebenezer Kinnersley, a friend of Benjamin Franklin. Kinnersley called his device the "electrical horse-race". The electrical horse-race was a popular toy in the latter part of the 18th century.

with his rider was placed upon each extremity of the cross. The motion was produced by corona discharge from the spurs of the riders (Fig. 13). The electrical horse-race was a popular toy in the latter part of the 18th century. Interesting descriptions of this toy have been provided by Langenbucher (Ref. 5, p. 206 and Plate 8), and by Guyot.[8]

To show more convincingly that the electric fly converted electrical energy into mechanical energy, electric flies were placed on horizontal axes and were arranged to lift small weights (Fig. 14) or to climb inclined rails (Fig. 15).[9]

Operation of the electric fly in liquids was investigated by Aimé.[10] A historical account of electric fly studies was presented by Tomlinson.[11] The mechanism of the fly was discussed at some length by Mascart.[12] A quantitative study of the forces acting upon the electric fly was performed by Kämpfer.[13]

A more efficient modification of the electric fly was proposed in 1887 by Bichat.[14] Bichat's electric "tourniquet" consisted of a light vertical frame capable of turning about a vertical axis and supporting two thin vertical wires producing a corona discharge along their entire length. The vertical sides of the frame acted as "spoilers" preventing a corona discharge from the "back side" of the wires. A simple motor based on Bichat's device is shown in Fig. 16.

In recent years the principle of the electric fly (reaction force experienced by a sharp point as a

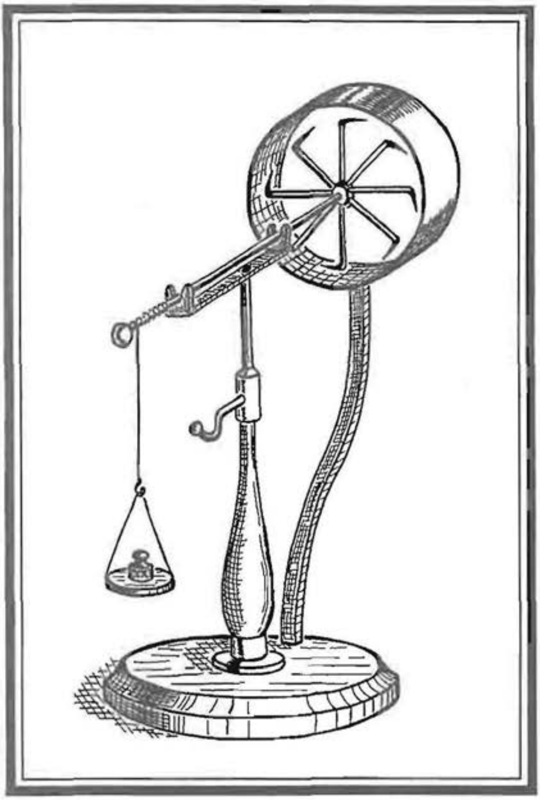

FIGURE 14

Electric fly converted electrical energy into mechanical energy and could lift small weights.

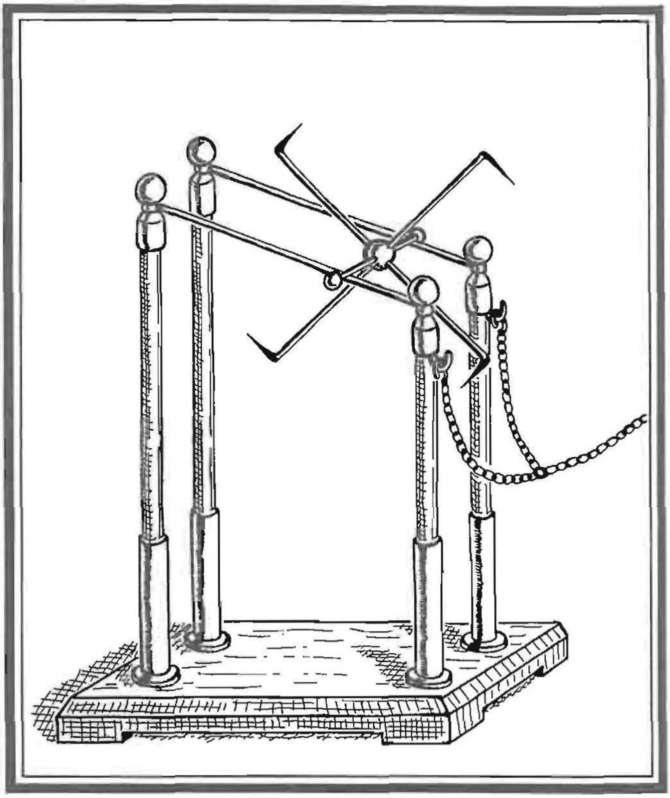

FIGURE 15

Electric fly developed enough mechanical energy for climbing inclined rails.

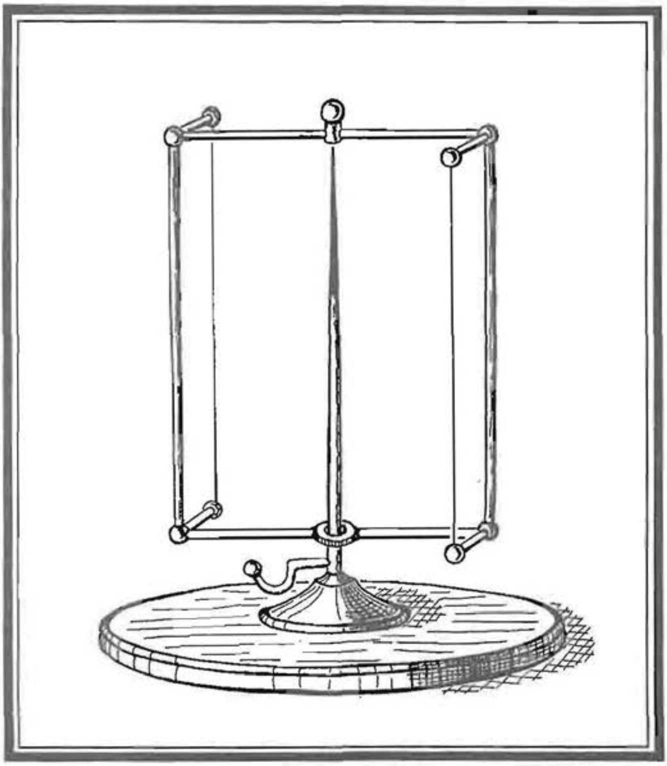

FIGURE 16

A more efficient modification of the electric fly was proposed in 1887 by Bichat, who called his device the "electric tourniquet".

result of a corona discharge from it) has been suggested for a flying machine ("corona-kraft levitation").[15] The lift realizable from such a machine does not, however, appear to be of practical significance.[16]

Closely related to the electric fly were the "electric wind wheels" and the "electric wind turbines". Like the electric fly, they were invented in the 1740's and utilized a corona discharge from sharp points for their operation. Benjamin Franklin studied these devices in 1747 (he first learned of them from his friend Philip Syng). According to Franklin (Ref. 7, p. 184) they were "light windmill-wheels made of stiff paper vanes; also. . .little wheels, of the same matter, but formed like water-wheels". Their probable appearance was as shown in Figs. 17 and 18. Later versions of these devices have been described by Guyot (Ref. 8, pp. 274, 275 and Plate 27), by Langenbucher (Ref. 5, pp. 75, 76 and Plate 3), by Gale,[17] and by Neuburger.[18]

A delightful electric toy using the electric wind for its propulsion was made by W. R. King, an associate of James Wimshurst (the designer of the electrostatic machine carrying his name). This toy, which may be called an "electrical rabbit hunt", now at the Science Museum in London, is shown in a simplified form in Fig. 19. It consists of a small electric wind turbine mounted on a vertical shaft carrying five light metal arms on its upper end; the arms support

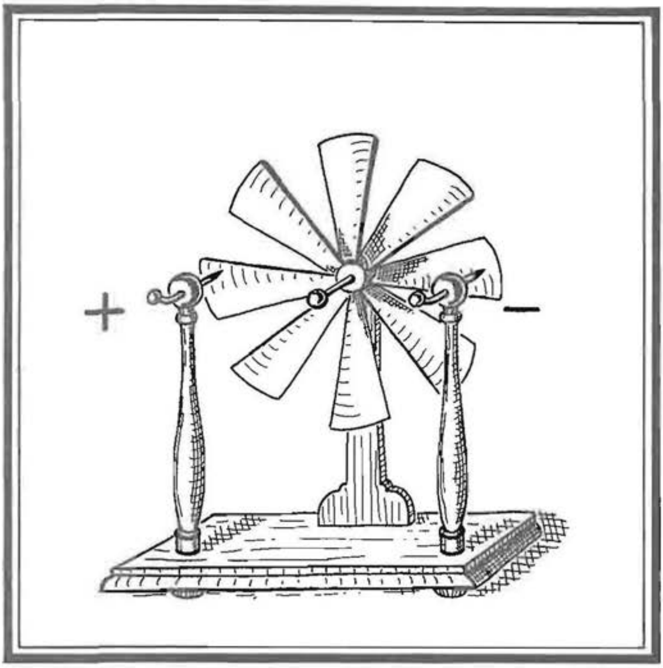

FIGURE 17

Closely related to Gordon's electric fly was the "electric wind wheel". Like the electric fly, it was invented in the 1740's and utilized a corona discharge from two sharp points for its operation.

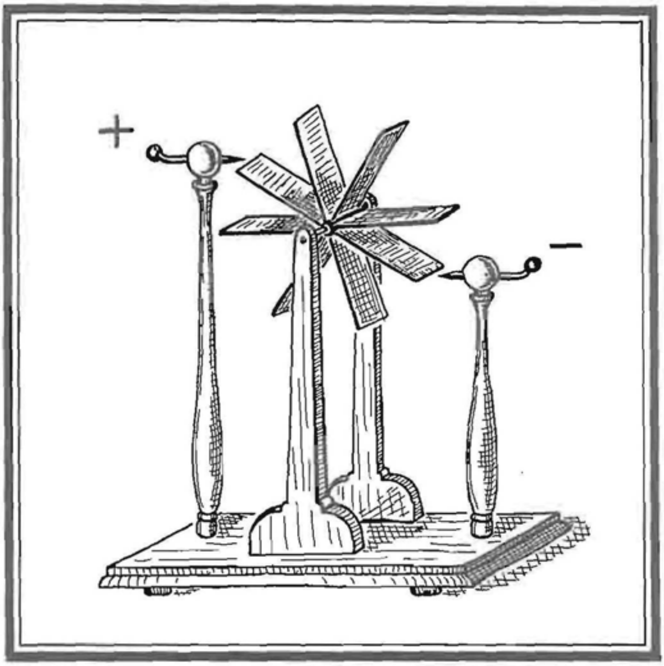

FIGURE 18

Benjamin Franklin studied this "electric wind turbine" in 1747. He concluded that its action depended not as much on the electric wind as on the forces between the charges present on its moving and stationary parts.

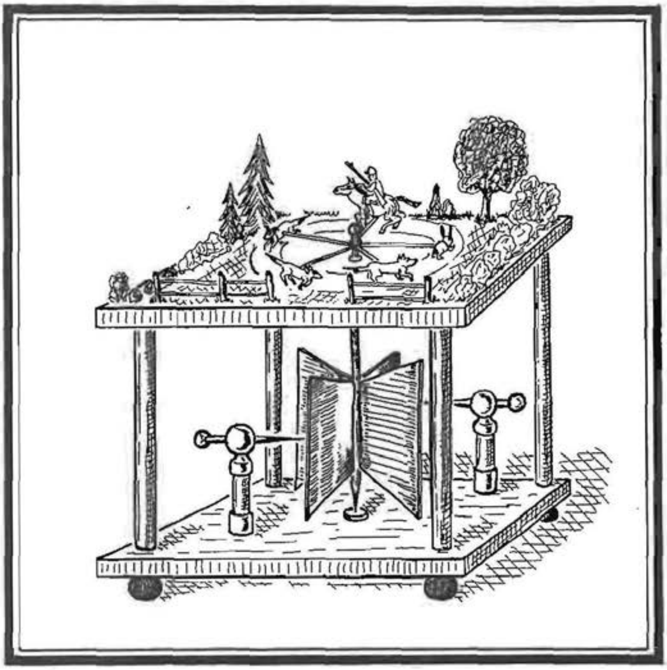

FIGURE 19

A delightful electric toy utilizing the electric wind for its propulsion was made by E. R. King, an associate of James Wimshurst. This toy, which may be called an "electrical rabbit hunt", is now at the Science Museum in London.

figurines of a rabbit, three dogs, and a horseman, who chase each other in a beautiful little garden, as the arms turn powered by the turbine below.

In studying the electric wind motors, Benjamin Franklin concluded that their action depended not as much on the electric wind as such, but on the repulsion and attraction of electric charges present on the moving and stationary parts of the motors (Ref. 7, p. 184). It is very probable that this conclusion led him to the invention of his big electric motors, the famous "electrical wheels", which are described in the next chapter.

REFERENCES

1. P. Benjamin, *A History of Electricity*, John Wiley and Sons, New York, 1898, pp. 506, 507.
2. J. C. Poggendorff, *Geschichte der Physik*, Johann A. Barth, Leipzig, 1879, p. 846.
3. V. K. Chew, *Physics for Princes*, Her Majesty's Stationery Office, London, 1968, p. 13.
4. B. Wilson, *Philosophical Transactions*, Vol. 51, p. 896 (1760).
5. J. Langenbucher, *Beschreibung einer beträchtlich verbesserten Elektrisiermaschine*, Mathäus Reigers, Augsburg, 1780, pp. 54-56 and Plate 3.
6. G. Adams, *An Essay on Electricity*, London, 1799, 5th edn., pp. 580-581 and Plate 4.

7. J. Sparks, Ed., *The Works of Benjamin Franklin*, Whittemore, Niles, and Hall, Boston, 1856, Vol. 5, p. 371.
8. M. Guyot, *Nouvelles Recreations Physiques et Mathematiques*, Gueffier, Paris, 1786, Vol. 1, pp. 249, 250 and Plate 25.
9. H. J. Oosting, *Zeitschrift für den Physikalischen und Chemischen Unterricht*, Vol. 9, pp. 84-85 (1896).
10. R. Aimé, *Annales de Chimie et de Physique*, Ser. 2, Vol. 62, p. 421 (1833).
11. C. Tomlinson, *Philosophical Magazine*, Ser. 4, Vol. 27, p. 202 (1864).
12. M. E. Mascart, *Traité d'Electricité Statique*, G. Masson, Paris, 1876, Vol. 1, pp. 175-179.
13. D. Kämpfer, *Annalen der Physik*, Ser. 3, Vol. 20, pp. 601-614 (1887)
14. E. Bichat, *Annales de Chimie et de Physique*, Vol. 12, pp. 64-79 (1887).
15. A. Christenson and S. Moller, *AIIA Journal*, Vol. 5, pp. 1768-1733, (1967).
16. L. B. Loeb, *Electrical Coronas*, University of California Press, Berkley, 1965, pp. 402-406.
17. L. D. Gale, *Elements of Natural Philosophy*, Mark H. Newman, New York, 1846, p. 164.
18. A. Neuburger, *Ergötzliches Experimentierbuch*, Ullstein & Co., Berlin, 1920, pp. 432, 433.

4

FRANKLIN'S SPARK MOTORS AND THEIR DESCENDANTS

If one defines the "electric motor" as a machine capable of converting substantial amounts of electrical energy into mechanical energy, then it was Benjamin Franklin who invented the electric motor, for he was the first person to design and to construct a machine capable of accomplishing such a conversion.

Franklin built two electric motors, which he called "electrical wheels". He described them in a letter to Peter Collinson, Fellow of the Royal Society of London, written in Philadelphia in 1748.[1]

His first motor was a big machine, about 40 inches in diameter. The main part of this motor was a wooden disk mounted horizontally on a vertical wooden axle and carrying 30 glass spokes with brass thimbles on their ends (Fig. 20 and Plate 2). The axle turned on a sharp point of

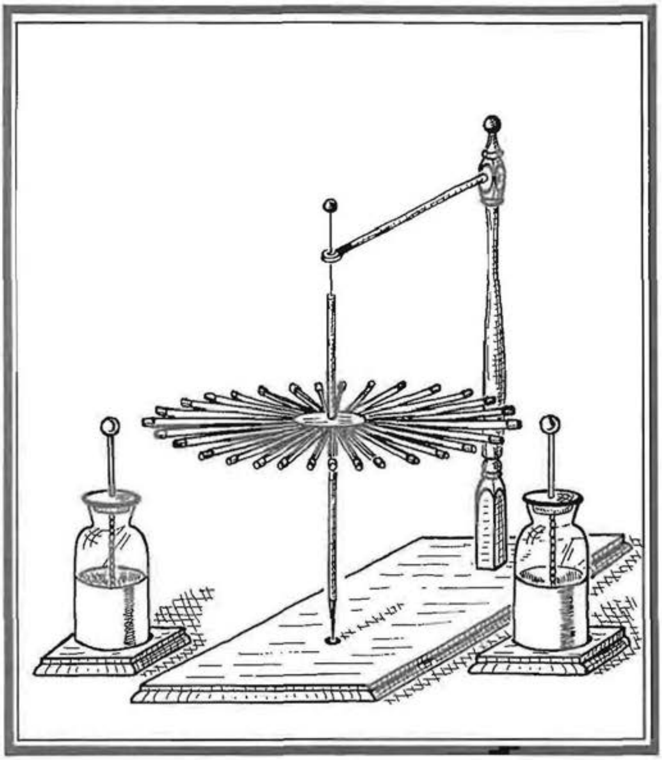

FIGURE 20

Franklin's first "electrical wheel" was the earliest device converting substantial amounts of electrical energy into mechanical energy.

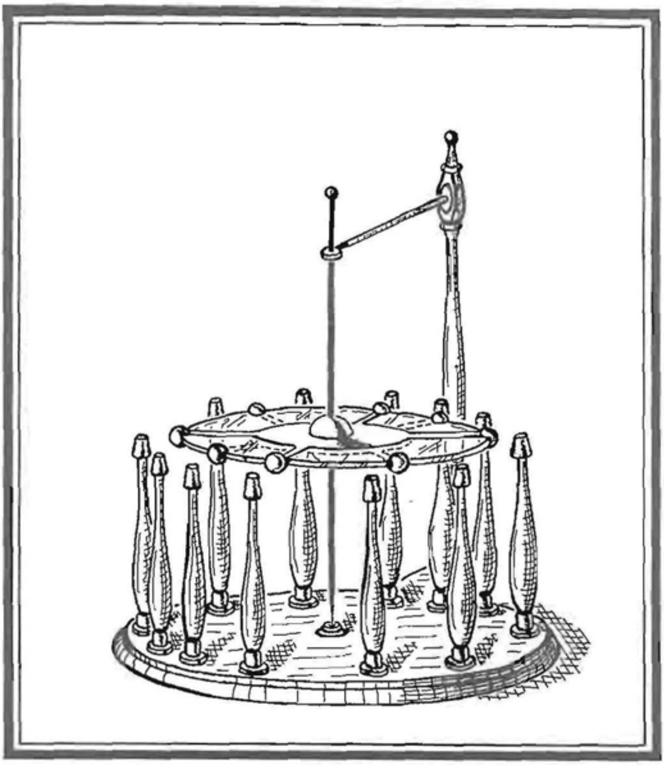

FIGURE 21

Franklin's second "electrical wheel" was more efficient than the first. It was powered by the electric charge stored in its capacitor-rotor.

iron fixed in its lower end and was kept in the vertical position by a strong wire mounted in its upper end and passing through a small hole in a thin stationary brass plate. Two oppositely charged Leyden jars (bottles) were placed in close proximity to the thimbles on the spokes, so that when the wheel turned, the thimbles barely missed the knobs of the jars.

As the thimbles moved past the Leyden jars, a spark jumped between a jar knob and the passing thimble, charging the latter with a charge of the same polarity as that of the knob. Therefore each knob attracted the oncoming thimbles ("opposite charges attract") and repelled the departing thimbles ("like charges repel"), causing the motor to turn. The motor, being perfectly symmetric, could turn in either direction, and usually required a starting push by hand to initiate the charging of the thimbles.

Here is how Franklin himself described the operation of his motor:[1]

"If now the wire of a bottle, electrified in the common way, be brought near the circumference of this wheel, it will attract the nearest thimble, and so put the wheel in motion; that thimble, in passing by, receives a spark, and thereby being electrified, is repelled, and so driven forwards; while a second being attracted, approaches the wire, receives a spark, and is driven after the first, and so on till the wheel has gone once round, when the thimbles before electrified approaching the wire, instead of being attracted as they were at first, are repelled, and the motion presently ceases. But if another bottle, which had been charged through the coating (e.i. charged oppositely to the first, O. J.), be placed near the same wheel, its wire will attract the thimble repelled by the first, and thereby double the force that carries the wheel round; and, not only taking out the fire (electric charge, O. J.) that had been communicated to the thimbles by the first bottle, but even robbing them of their natural quantity, instead of being repelled when they come again towards the first bottle, they are more strongly attracted, so that the wheel mends its pace, till it goes with a great rapidity, twelve or fifteen rounds in a minute, and with such strength, as that the weight of one hundred Spanish dollars, with which we once loaded it, did not seem in the least to retard its motion".

Franklin was not quite pleased with this motor, however, because it required "a foreign force, to wit, that of the bottles" for its operation. He constructed therefore a second, "self-moving", motor, which itself contained the electrical energy to be converted into the mechanical motion.

The second motor was made as follows (Fig. 21 and Plate 3). A disk of window glass, 17 inches in diameter, was gilded on each side, except for narrow areas next to the edge. Several lead spheres (bullets) were fixed to the edge of the disk and were connected in alternation to the two gold layers. The disk was mounted on a vertical axle consisting of two pieces of strong wire, each about 9 inches long; the upper wire being in contact with the top layer of gold, the lower wire being in contact with the bottom layer. The lower wire terminated in a sharp point resting on a piece of brass cemented within a glass salt-cellar. The upper wire passed through a hole in a thin brass plate cemented on a long strong piece of glass. This wire carried a small ball on its upper end for preventing a corona discharge from it. Several glass pillars with thimbles on their tops were positioned in a circle around the disk, so that the spheres on the edge of the disk barely missed the thimbles when the disk was in motion. The motor was powered by electricity stored in the disk itself, which, due to the two layers of gold, constituted a capacitor capable of storing an appreciable quantity of charge (to charge the rotor, the two layers of gold were temporarily connected to an electrostatic friction generator).

The motor operated on the same principle as Franklin's first motor, except that now the thimbles were stationary and received positive or negative charges from the passing spheres, rather than the other way around. As in the first motor, the rotation persisted as long as there was enough charge stored in the capacitor (rotor) to produce sparks between the spheres and the thimbles.

Franklin reported that his second motor

(which was more powerful than the first one) could make 50 turns in a minute and could run for up to 30 minutes on a single charging of the rotor.[1] Experiments with replicas of Franklin's motors performed by the author indicate that the original motors developed a power of the order of 0.1 watt,[2,3] which was a very impressive achievement for that very early period in the history of electric science.

Several authors have described simplified versions of Franklin's motors.[4,5] Miniature Franklin-type motors have been constructed and described by A. D. Moore[6] (Fig. 22). One such motor consists of a Plexiglas disk about 4 inches in diameter with small conducting balls embedded in the rim. The disk turns on a horizontal axis supported by a base behind the disk. The motor runs when placed between two conductors connected to an electrostatic generator. To make the motor self-starting and one-way running, the conductors may be provided with short auxiliary rods tangential to the disk; since the rods dispatch sparks predominantly in the direction of their free ends, the disk also turns in this direction. Another little motor of the same type designed by Moore is shown in Fig. 23.

An obvious improvement of Franklin's motors would be to use a cylindrical rotor with electrodes in the shape of long strips. Such electrodes would accept more charge than thimbles or spheres, and the power of the motor would be increased considerably. Motors of this type

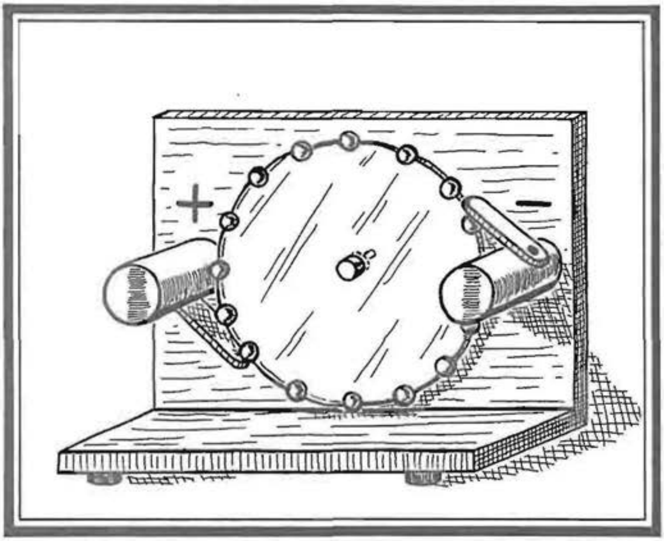

FIGURE 22

Several authors have described simplified versions of Franklin's motors. One such motor consists of a Plexiglas disk about 4 inches in diameter with small conducting balls embedded in the rim. The disk turns on a horizontal axis supported by a base behind the disk.

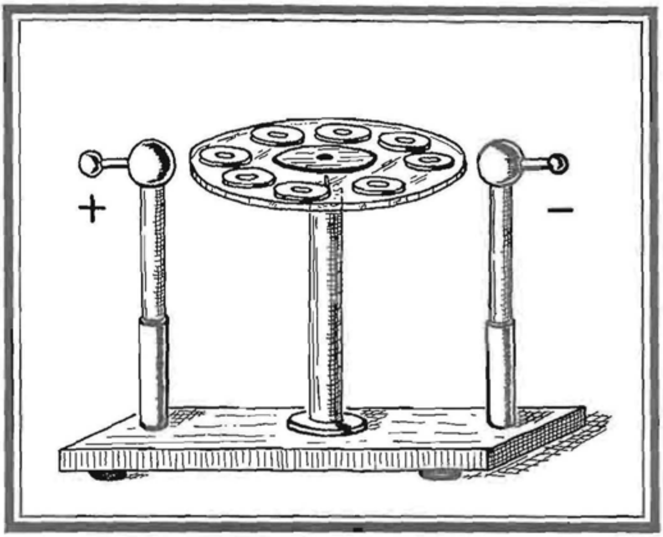

FIGURE 23

This inexpensive miniature Franklin-type motor of recent design can be used for lecture-room demonstrations of electrostatic forces. The rotor disk can be made from cardboard, wood, or plastic. For the rotor electrodes one can use thumbtacks or small aluminum foil disks.

have indeed been constructed. Here is a description of a "bell jar" motor taken from a book published in 1847 (Fig. 24):

"On a pivot, *a*, suspend a bell jar having four pieces of tinfoil pasted on its sides, *b*, *c*, *d*; connect the jar, by means of the insulated wire, *y*, with the prime conductor (high-voltage terminal of an electrostatic friction generator, O. J.) so that the pieces of tinfoil may receive sparks. On the opposite side arrange a conductor, *x*, in connection with the ground by a chain. On putting the machine into activity, the jar will commence rotating on its pivot."[7]

Small motors closely related to both Franklin's first motor and to pendulum motors were used in the rotary-type "perpetuum mobile" (see Chapter 2) operated from a Zamboni pile.[8] The main part of these motors (Fig. 25) was a light insulated arm pivoted on a sharp point and carrying a piece of tinsel on each end. The tinsels could just touch the knobs of the pile when the arm was moving on its pivot. The device functioned in the same manner as Franklin's first motor, the two tinsels corresponding to the thimbles of the latter.

An analogous design has recently been described as a "millifleapower motor" operating from a nuclear battery.[9] In this motor a light aluminum vane with pointed ends (Fig. 26) bent so as to form an inverted V is mounted at its midpoint on the tip of a sharp needle protruding from the positive terminal of the battery. As the

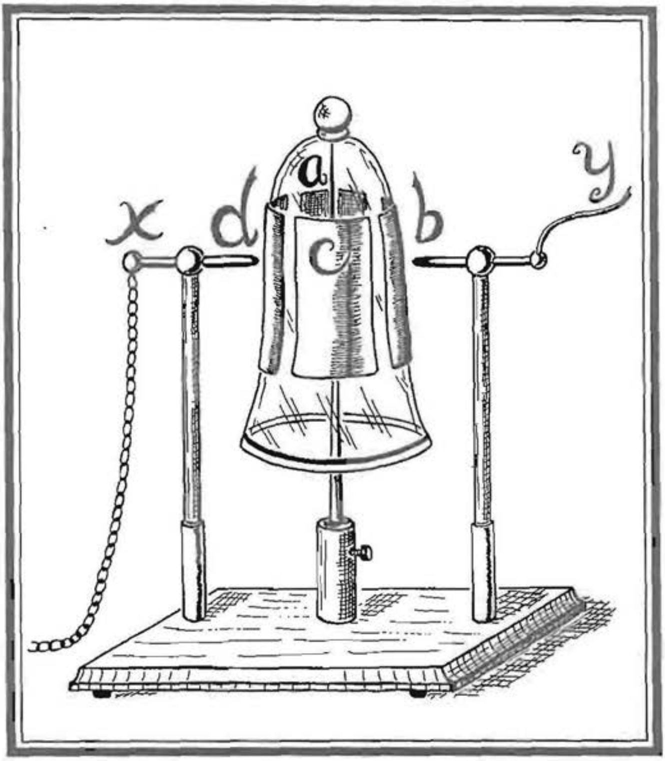

FIGURE 24

This "bell jar" Franklin-type motor had a cylindrical rotor with tinfoil electrodes in the shape of long strips.

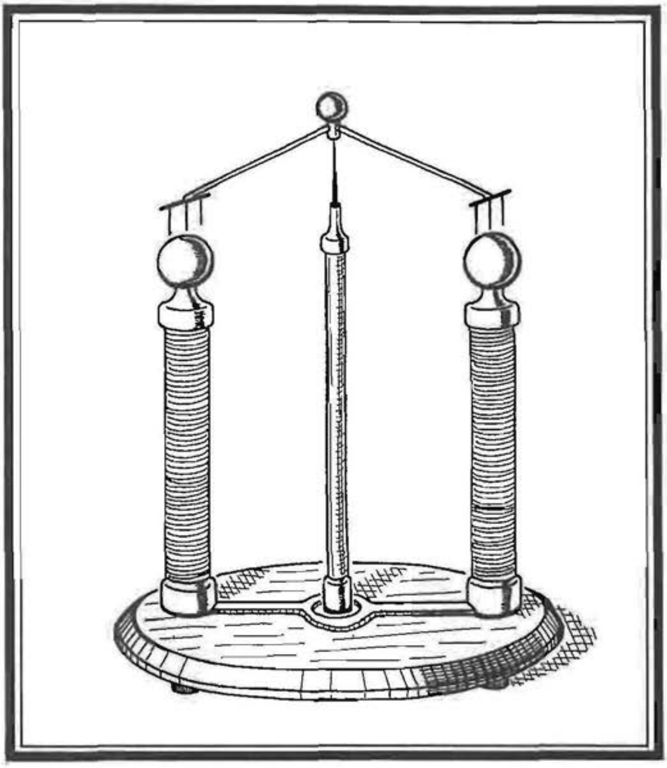

FIGURE 25

This small motor was used as a rotary-type "perpetuum mobile" operating from a Zamboni pile. Such motors turned for years without stopping.

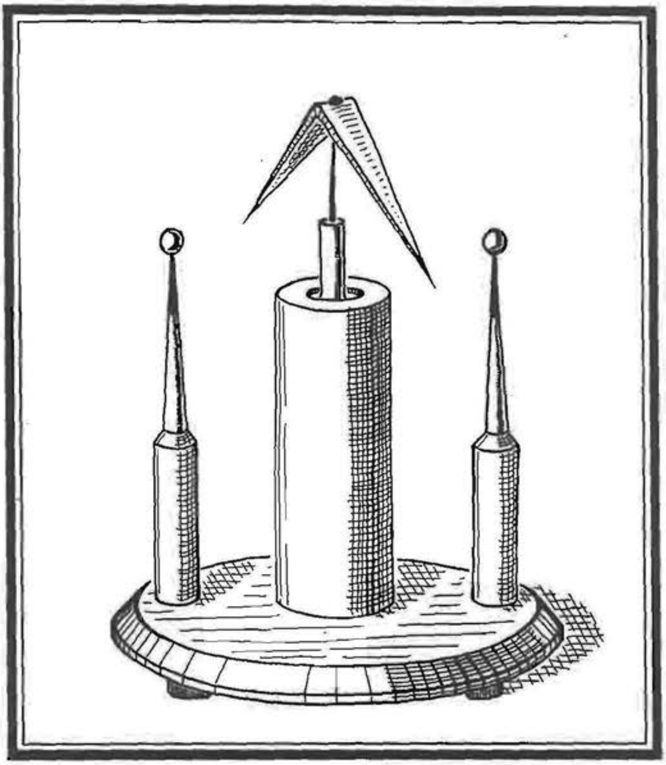

FIGURE 26

An analogous design has recently been described as a "millifleapower motor" operating from a nuclear battery.

vane turns, its ends come close to two spheres connected with the ground through a high resistance and become attracted by them. A discharge from the ends of the vane charges the two spheres, and the spheres then repel the ends of the vane. During the time it takes the vane to complete one half of its revolution, the charges of the spheres leak off to the ground, so that the spheres can again attract the ends of the vane, accept new charges from them, and then again repel them. The motor is expected to run for many years.

An interesting motor derivable from Franklin's first motor and from pendulum motors was suggested for operation from an electrophorus.[10] The main part of this motor (Fig. 27) was a light conducting ring supported in a vertical plane on an insulated needle. In front of the ring there were two insulated spheres, which were so positioned that the ring just missed them as it turned on its needle. The spheres were either charged oppositely by the electrophorus, or only one sphere was charged, in which case the second sphere was grounded. The device operated on the same principle as Franklin's first motor.

Of all the motors discussed above, Franklin's original motors and the bell jar motor were the most powerful ones. One can make such motors even more powerful by increasing the number of moving and stationary electrodes. But this can be done only up to a limit, because if the electrodes are too close to one another, sparks will

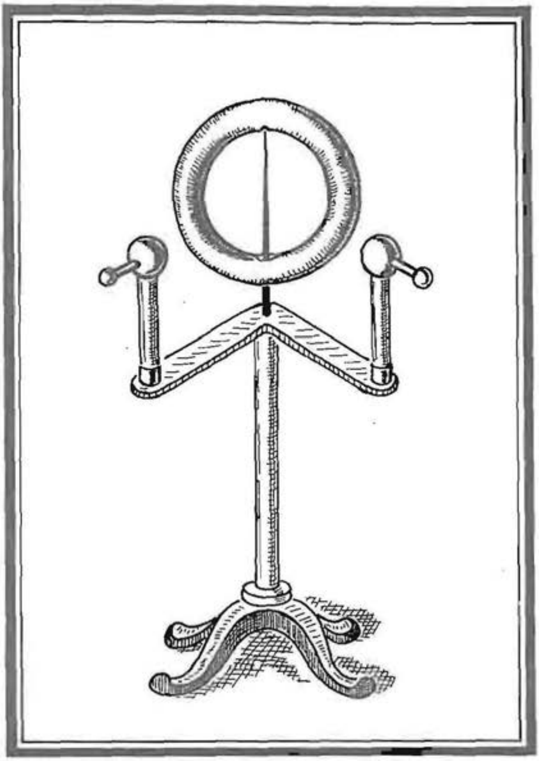

FIGURE 27

This motor derivable from Franklin's first motor and from pendulum motors was suggested for operation from an electrophorus.

jump all around from one electrode to the next, thus short-circuiting the motor. The motor will then not work at all, unless the operating voltage is lowered. But by lowering the voltage one also lowers the power, so that the gain obtained by increasing the number of electrodes becomes offset by the decrease in the operating voltage. Franklin type motors have therefore an inherent limitation in their power characteristics, and the fact that they depend for their operation on the presence of many metallic electrodes is one of their disadvantages.

Since in Franklin's motors and in other motors described in this chapter the rotor is charged by means of electric sparks, these motors may be classified as "spark motors". Toward the end of the 19th century a way was found to construct electrostatic motors based on the same mechanism of attraction and repulsion between electric charges as in Franklin's motors, but with rotors (or stators) charged more efficiently than it is possible to do by means of the sparks. These motors are described in the next chapter.

REFERENCES

1. J. Sparks, Ed., *The Works of Benjamin Franklin*, Whittemore, Niles, and Hall, Boston, 1856, Vol. 5, pp. 204-207.
2. O. Jefimenko and D. K. Walker, *The Physics Teacher*, Vol. 9, pp. 121-129 (1971).
3. O. Jefimenko, *American Journal of Physics*, Vol. 39, pp. 1139-1140 (1971).

4. M. Guyot, *Nouvelles Recreations Physiques et Mathematiques*, Gueffier, Paris, 1786, Vol. 1, pp. 272-274 and Plate 27.
5. Dayton C. Miller, *Sparks, Lightning, Cosmic Rays*, MacMillan, New York, 1939, pp. 63, 64.
6. A. D. Moore, *Electrostatics*, Doubleday and Company, Inc., New York, 1968, pp. 104-108.
7. J. W. Draper, *Text-Book on Chemistry*, Harper and Brothers, New York, 1847, p. 105.
8. M. J. Jamin, *Cours de Physique*, Gauthier-Villars, Paris, 1869, 2nd edn., Vol. 3, p. 39.
9. R. H. Dressel, *High School Science Bulletin of the New Mexico State University*, Vol. 6, pp. 3-4 (April 1, 1964).
10. H. Wommelsdorf, *Annalen der Physik*, Ser. 4, Vol. 70, pp. 135-138 (1923).

5

CORONA MOTORS AND MOTORS SIMILAR TO THEM

In 1867, W. Holtz, the inventor of the electrostatic "influence" machine (generator) carrying his name, discovered that this machine could function as a motor when powered by another similar machine.[1] In order to study this effect in some detail, the famous German physicist J. C. Poggendorff ordered the construction of a special device, analogous to Holtz's machine, but designed to operate only as a motor and provided with numerous adjustable components so as to operate under a wide range of conditions.[2]

The basic part of this motor was a vertical glass-disk rotor placed between two ebonite crosses carrying metallic sharp-needle "combs" (Fig. 28 and Plate 4). The sharp points of the combs almost touched the surface of the disk, so that the disk could be charged by means of a corona discharge from the combs when they were connected in alternation to an electrostatic generator. Each comb could be oriented so as to make an arbitrary angle with respect to a radial direction.

Just like the sparks in Franklin's motors, the corona discharge from a comb deposited on the rotor a charge of the same polarity as that of the comb. Each comb repelled therefore the segment of the rotor carrying charges which it sprayed onto the rotor and attracted the segment carrying charges sprayed onto the rotor by the preceding comb. Because of a continuous discharge between the combs and the rotor, almost the entire surface of the rotor was sprayed with electric charges. The repulsion and attraction between the combs and the rotor, were therefore not only much stronger than in Franklin's motors, (where only the electrodes were charged) but were also much steadier, since the distances between the combs and the charges on the rotor were always the same. As a result, the torque and power obtainable with this motor were far greater than those obtainable with Franklin's motors of similar dimensions.

Poggendorff made a thorough study of his motor, varying all its important parameters (his article comprises 32 pages). He used several rotors of different thickness and surface characteristics, whose weight ranged from 2.5 to 4.5 pounds (all rotors were 15 inches in diameter). He varied the number, the orientation, and the polarity of the combs.

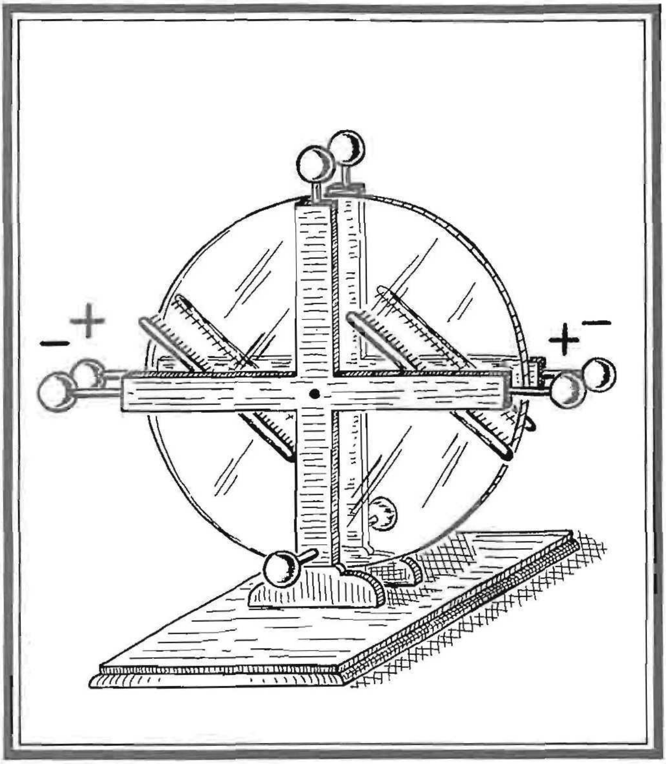

FIGURE 28

Poggendorff's corona motor had a glass-disk rotor placed between two ebonite crosses carrying metallic sharp-needle "combs".

He found that a motor with combs arranged along the radii of the rotor could rotate equally well in either direction but was not self-starting, since such combs sprayed charges onto the rotor in a perfectly symmetric fashion. However, if the combs were slanted relative to the radii, the motor became self-starting and unidirectional, because such combs sprayed charges predominantly in one direction. He also fould that the performance of the motor could be improved by placing pieces of glass, cardboard, or metal close to the surface of the rotor (this is the principle of the modern "backing plate" or "lining"; see Chapter 10). When properly adjusted, the motor could run at better than 300 rpm.

Poggendorff had fully appreciated the superiority of his motor in comparison fo Franklin's motors, and had stated so at the end of his article. In fact, he had indicated that his motor used from 1200 to 1800 times as much current as Franklin's motors, in which case, assuming that the operating voltage of his motor was the same as that of Franklin's motors (approximately 100 kilovolts), the power of his motor must have been close to 100 watts.

At the time when Poggendorff's article was written, however, it was considered unscientific to expect that electricity (*any* electricity) could ever become a significant source of motive power. This may be the reason why Poggendorff, speaking of the rotation phenomenon that he investigated, declared at the end of his article: "...it would be a sanguine hope if one wanted to believe that any useful mechanical effect could be achieved with it. That this is not possible, follows already from the consideration of how small is the quantity of electricity that is here put into play in comparison to that developed by a voltaic battery, with which one nevertheless, even with the help of the magnetismus produced by it, has so far achieved nothing substantial ("Erkleckliches")". The argument was obviously weak, but taking into account Poggendorff's reputation as an eminent scientist, it was probably sufficient to discourage further serious investigations of the corona-induced rotation for many years to come.

Since Poggendorff's motor depended on a corona discharge for its operation, the motor may be classified as a "corona motor".

Holtz's rotation phenomenon was also studied by the Danish physicist C. Christiansen (whose interest in the phenomenon had been aroused by a preliminary report of Poggendorff's investigations published in 1867).[3] Christiansen described two motors based on Holtz's discovery. His first motor consisted of a horizontal disk rotor and two stationary corona-producing combs. His second motor was similar to the first except that in it the glass disk was stationary and the combs constituted the rotor. The rotor combs in this second motor were located below the glass disk and were charged by means of two foil strips hanging from them and sliding along

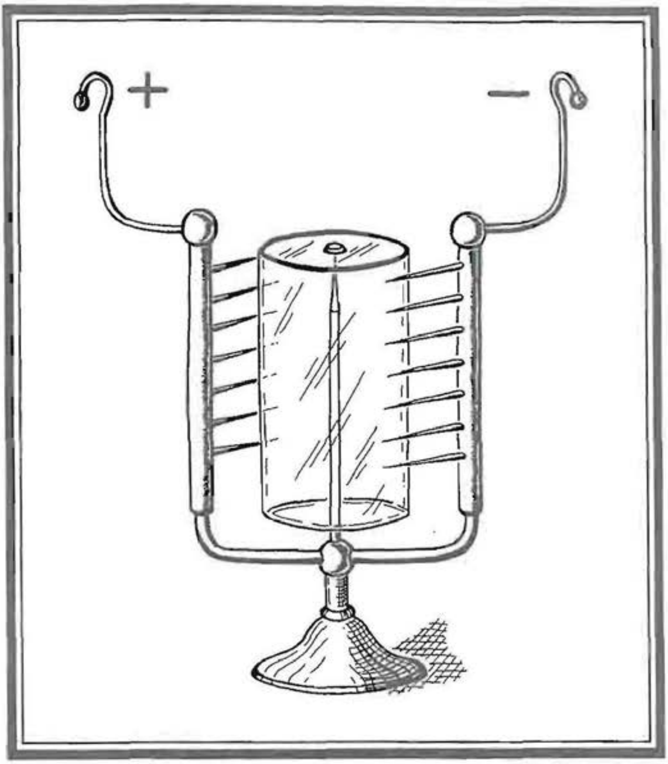

FIGURE 29

This self-starting corona motor with a cylindrical rotor was described in 1871 under the name of "electric tourbillion".

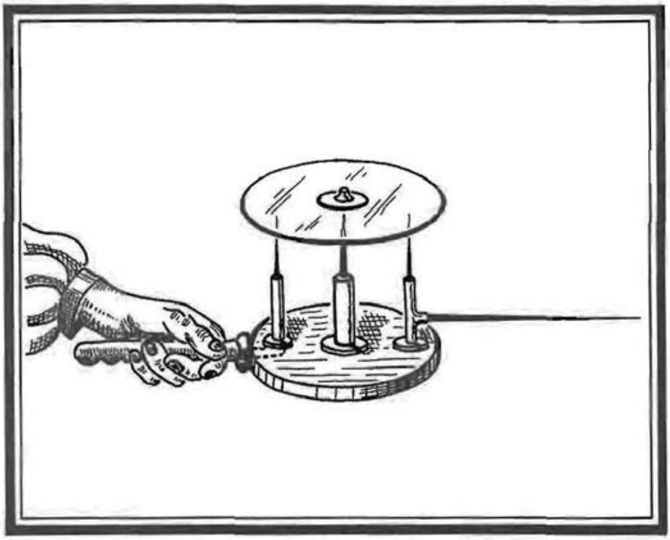

FIGURE 30

In this hand-held corona motor (attributed to Ruhmkorff) the horizontal wire, by a corona discharge, extracted charges from the power source, and the vertical needle connected to this wire, also by a corona discharge, sprayed these charges onto the rotor.

two horizontal concentric contact rings mounted on the base of the motor.

Cylindrical and spherical motors operating on the same principle as Poggendorff's motor were described in 1871 by W. Gruel (Fig. 29), who called them the "electric tourbillions".[4] Gruel pointed out that these motors could be made self-starting and unidirectional by inclining the combs in the same rotational direction with respect to the surface of the rotor. He also suggested that by providing the rotors with a series of holes, the motors could be used as sirens.

A disk motor not noticeably different from Poggendorff's motor was patented in the U.S.A. in 1891 by J. W. Davis and J. B. Farrington.[5]

An interesting hand-held corona motor (attributed to Ruhmkorff) was described in 1876 by M. E. Mascart.[6] The motor had a horizontal mica disk rotor (Fig. 30) supported on a vertical needle by a jewel bearing, such as are used for supporting compass needles. Below the disk there were two vertical corona-producing needles mounted on a hard rubber base. One of the latter needles was connected to a metal handle, the other was connected to a long, stiff, horizontal wire terminating in a sharp point. To set the motor in motion, the operator held the motor so that the sharp point of the horizontal wire was in the proximity of a charged conductor. By a corona discharge, the horizontal wire then extracted charges from the conductor, and the ver-

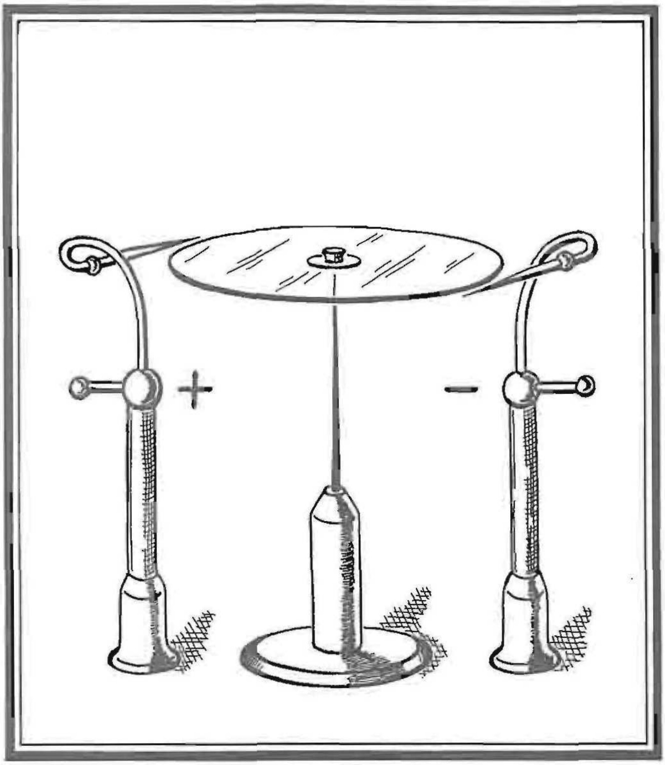

FIGURE 31

This motor had two insulated corona-producing points tangential to the rotor disk and oriented in the same rotational direction.

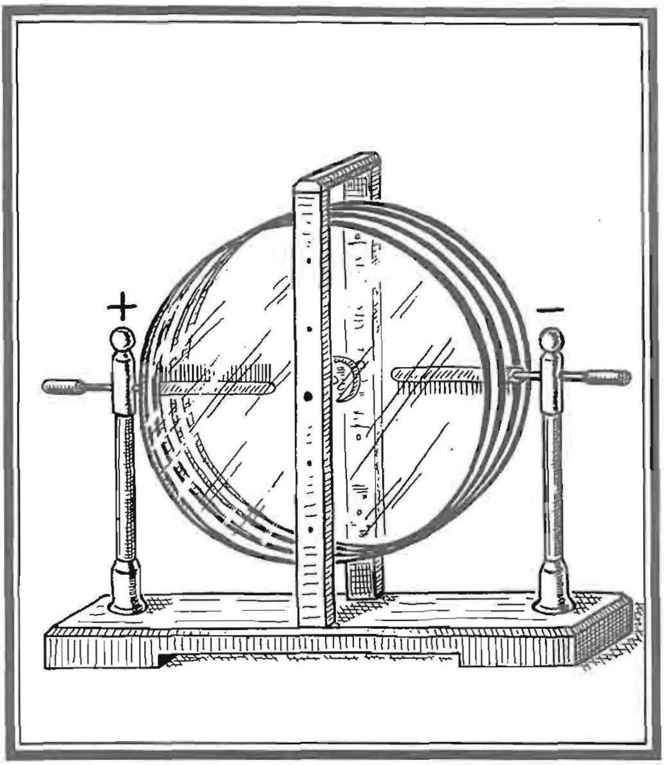

FIGURE 32

This corona motor with a compound glass rotor could run at about 2000 rpm and had a power of about 90 watts.

tical needle connected to this wire, also by a corona discharge, sprayed these charges onto the rotor. The second vertical needle discharged the rotor and conducted its charges to the ground through the hand and body of the operator.

A number of experiments with Poggendorff type corona motors were reported in 1921 by V. E. Johnson.[7] One of his motors was similar to Ruhmkorff's motor (Fig. 31). This was a horizontal mica disk, about 6 inches in diameter, supported on a vertical needle point. At the edge of the disk there were two insulated corona-producing points tangential to the disk and oriented in the same rotational direction. Another of Johnson's motors was a big machine consisting of two stationary glass disks and a three-piece glass rotor that could rotate on ball bearings between the stationary disks. The rotor was made of two large glass disks separated by a smaller disk so as to have a deep slot along its edge. Two flat sharp-point combs were inserted diametrically opposite to one another into the slot of the rotor (Fig. 32). Johnson estimated that his motor could run at about 2000 rpm and had a power of about 90 watts.

Just as Poggendorff's motor was derived from Holtz's electrostatic machine, a series of electrostatic motors were similarly derived from Wimshurst's electrostatic machine.

In 1891 five such motors were constructed by William McVay of New York City.[8] McVay's first motor (Fig. 33) consisted of two horizontal glass disks, about 12 inches in diameter, one stationary and the other rotating on the vertical axis just above the first. The lower disk had two quadrants of tinfoil, and the upper disk had 16 tinfoil sectors, as shown in the figure. The power (from a Wimshurst machine) was delivered to the motor by means of two insulated arms, each of which terminated in two brushes, one touching continually one of the lower quadrants, the other charging a sector on the upper disk just clear of the edge of the quadrant. Charges of the same polarity were thus deposited on the quadrant and on the sector, causing them to repel each other. An "equalizer" reduced the charge of the sectors before they passed over the further edge of an oppositely charged quadrant, thus reducing the back torque on the rotating disk. An important new feature of this motor was the simultaneous charging of the stationary quadrants and of the moving sectors, which assured a relatively strong starting torque and, together with the neutralizing system, assured a reliable unidirectional operation. McVay also constructed motors of cylindrical geometry, one of which is shown in Fig. 34. In this motor the quadrants were located on the inner cylinder, while the charging and neutralizing brushes were on the neck of the outside cylinder.

McVay's first motor was later modified for charging by combs, rather than by brushes (Fig. 35). The motor then operated essentially as

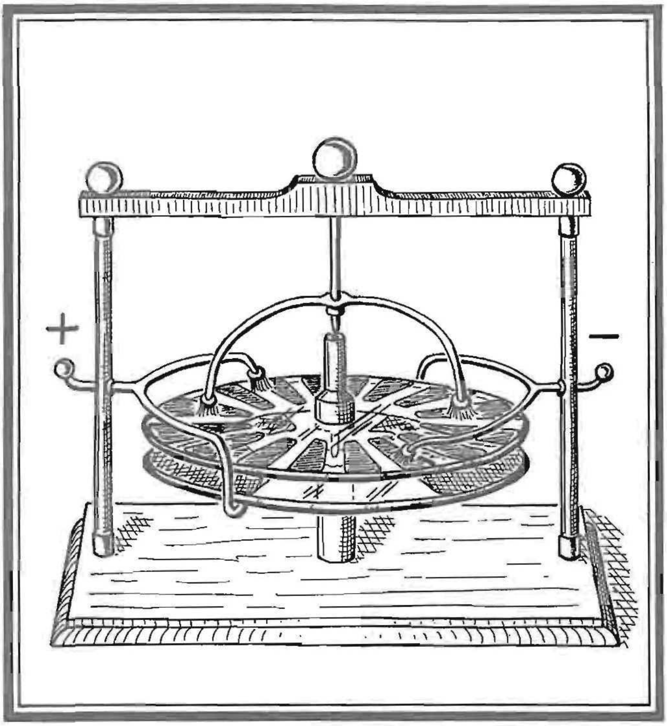

FIGURE 33

McVay's first motor consisted of two horizontal glass disks, about 12 inches in diameter, one stationary and the other rotating on the vertical axis just above the first.

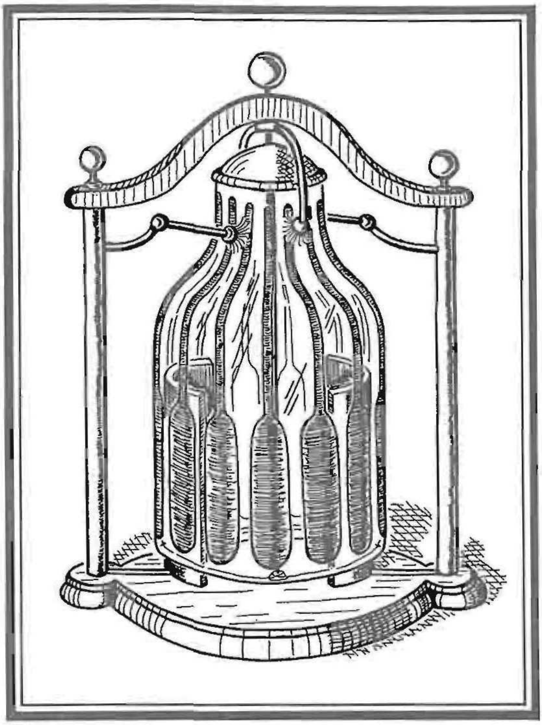

FIGURE 34
In this McVay's motor the quadrants were located on the inner cylinder (shown as two segments), while the charging and neutralizing brushes were on the neck of the outside cylinder.

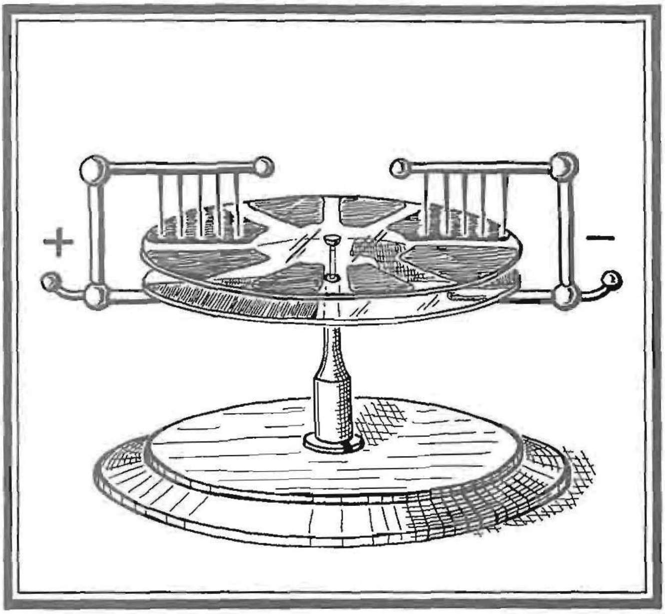

FIGURE 35

McVay's first motor was later modified for charging by combs, rather than by brushes. The motor then operated essentially as Poggendorff's corona motor, retaining, however, its self-starting and unidirectional qualities.

Poggendorff's corona motor, retaining, however, its self-starting and unidirectional qualities.

Instructions for building a Poggendorff motor may be found in the May 1971 issue of Popular Science.[9] Instructions for building a simple McVay motor may be found in the January 1914 issue of Electrical Experimenter[10] (the author is grateful to Mr. Thorn L. Mayes for this information).

Since a corona discharge in air at atmospheric pressure requires a minimum voltage of about 3000 volts, ordinary corona motors can operate only from sources capable of supplying a voltage of this magnitude. Such voltages are, of course, easily attainable at the present time, and since the corona motors are very simple and efficient devices, they have been further developed and studied in recent years. We shall return to them in the last chapter of this book.

REFERENCES

1. W. Holtz, *Annalen der Physik*, Ser. 2, Vol. 130, pp. 168-171 (1867).
2. J. C. Poggendorff, *Annalen der Physik*, Ser. 2, Vol. 139, pp. 513-546 (1870).
3. C. Christiansen, *Annalen der Physik*, Ser. 2 Vol. 137, p. 490 (1869).
4. "Elektrischer Tourbillion", *Annalen der Physik*, Ser. 2 Vol. 144, p. 644 (1871).
5. U. S. Patent No. 459678.

6. M. E. Mascart, *Traité d'Electricité Statique*, G. Masson, Paris, 1876, Vol. 1 pp. 179, 180.
7. V. E. Johnson, *Modern High Speed Influence Machines*, E. and F. N. Spon, London, 1921, pp. 175-205.
8. "A Group of Static Motors", *Electrical World*, Vol. 18, p. 418 (1891).
9. C. P. Gilmore and W. J. Hawkins, *Popular Science Monthly*, Vol. 198, pp. 95-97 and 114 (May 1971).
10. "A Static Electric Motor", *Electrical Experimenter*, January 1914, p. 137.

6

CAPACITOR MOTORS

In 1889, Karl Zipernowsky, a Hungarian engineer (co-inventor of practical electrical transformers), constructed a new type of electrostatic motor,[1] which was derived from Thomson's quadrant electrometer.[2] The rotor of this motor (Fig. 36) consisted of two pairs of aluminum sectors insulated from each other and from the rest of the apparatus. The stator consisted of four double (hollow) sectors of brass enclosing the rotor. The rotor was fitted with a commutator in four parts, by means of which the sectors of the rotor were charged oppositely to those sectors of the stator into which they were entering and identically to those sectors of the stator which they were leaving. An interesting property of this motor was that it could operate from high-voltage dc as well as from high-voltage ac sources.

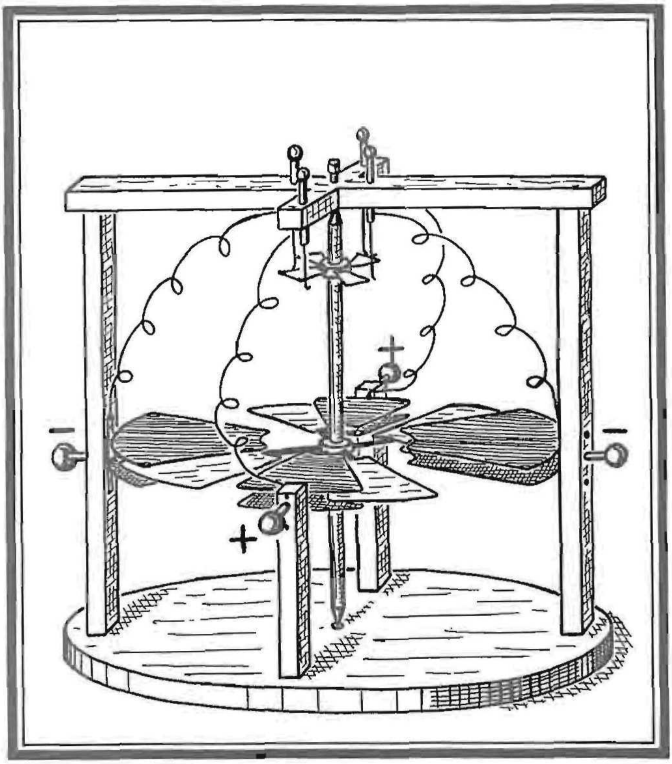

FIGURE 36

This motor was built in 1889 by Karl Zipernowsky, a Hungarian engineer. It operated from dc as well as from ac sources.

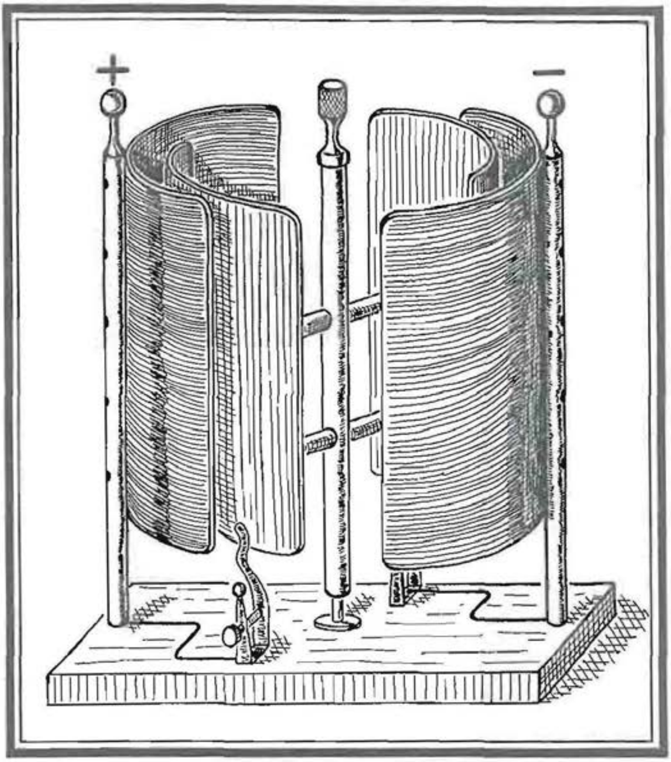

FIGURE 37

This capacitor motor was described in 1904 by van Huffel. Its rotor could be charged by contact or by sparks.

Inasmuch as Zipernowsky's motor operated as a result of the electric forces exerted by one charged conducting plate upon a second charged conducting plate (which are the same forces that act upon the two plates of a capacitor) it constituted what is now called an electrostatic "capacitor motor".

A simpler version of a capacitor motor was described in 1904 by van Huffel.[3] This motor was based on the so-called Thomson's replenisher.[4] The motor had an essentially cylindrical geometry (Fig. 37) and consisted of a rotor with two insulated bent brass plates mounted on a vertical axle and located between two similar, but slightly larger, plates of the stator. The rotor plates (which were not quite coaxial) were charged by means of two metallic tongues connected to the stator plates; the tongues touched the rotor plates just as the latter began to clear the stator plates. The charging of the rotor plates could occur by a direct contact as well as through a spark, depending on the adjustment of the tongues.

Since capacitor motors do not require sparks or a corona discharge for their operation, they can operate, at least in principle, from as low a voltage as one desires to use. This is one of their important advantages and is one of the reasons that such motors have been given considerable attention in recent years. Furthermore, as already indicated, capacitor motors can operate not only from dc sources, but also from ac sources. Finally, when powered by an ac source, they can operate both as synchronous and asynchronous motors (Zipernowsky's original motor operated from ac as an asynchronous motor).

A synchronous capacitor-type electrostatic motor is merely a multi-electrode capacitor motor without a commutator, the proper charging of the rotor being accomplished by continuously supplying an ac voltage of proper frequency between the stator and the rotor (Fig. 38). It is easy to see that if the rotor moves by one electrode in one period of the supply voltage, then the ac voltage accomplishes the same effect as that accomplished by a dc voltage with a commutator. The synchronous velocity is therefore $2\pi f/N$, where f is the frequency of the supply voltage and N is the number of the electrodes.

Several precision-made synchronous motors were described in 1969 by B. Bollée.[5] One of these motors ran on sapphire bearings and had a cylindrical aluminum rotor 1 centimeter long and 0.45 centimeters in diameter; the rotor and the stator had 60 electrodes each, with a 0.001 centimeter gap between stator and rotor electrodes. The maximum power of this motor was about 100 microwatts at 220 volts and 50 hertz. Another motor (for an electric clock) had a dielectric disk rotor, with four rows of 100 electrodes on each side, located between two bearing covers, each with five rows of 100 elec-

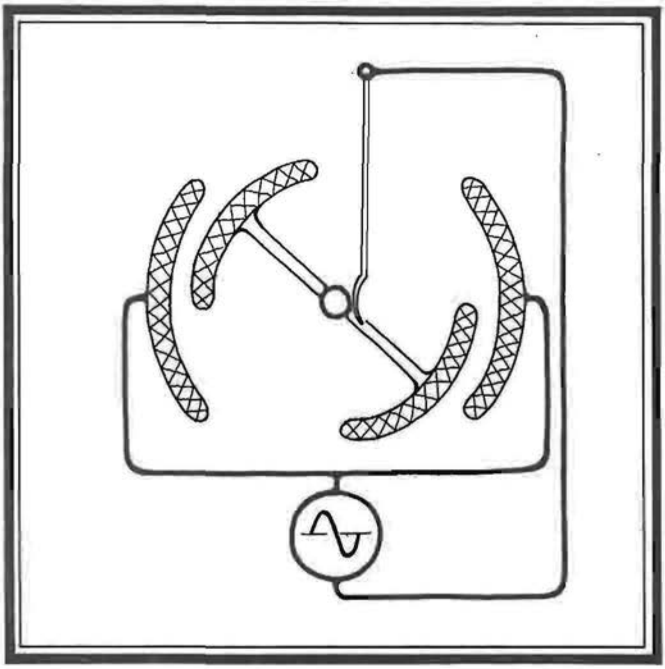

FIGURE 38

In a synchronous capacitor-type electrostatic motor the proper charging of the rotor is accomplished by continuously supplying an ac voltage of proper frequency between the stator and the rotor.

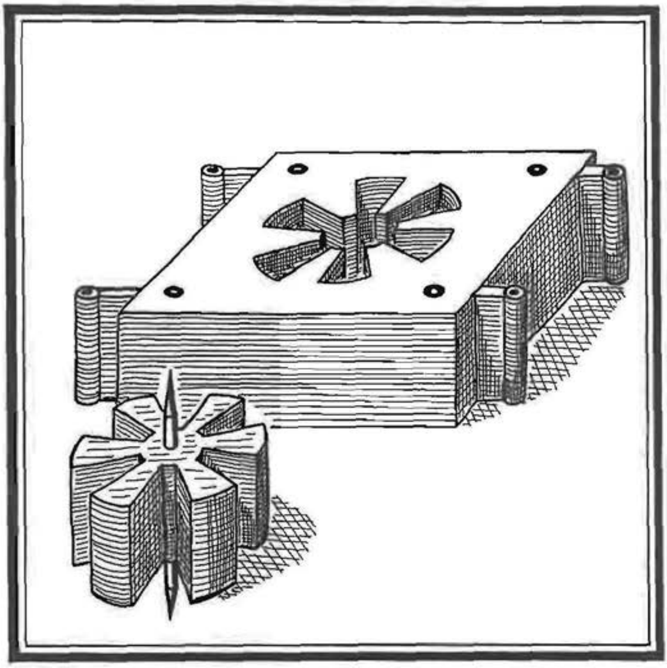

FIGURE 39

This synchronous electrostatic motor is essentially a variable capacitor with a rotatable shaft. The design of this motor is similar to the design of an electrostatic variable-capacitance generator.

trodes; the gap width in the axial direction was 0.01 centimeter. At 200 volts this motor had a maximum output of 600 microwatts. An interesting feature of this motor was that it had no slip ring on the rotor shaft, and that the rotor was therefore not connected to the power source. Instead, the stator was divided into two mutually insulated halves. This arrangement reduced the friction of the rotor and improved the performance of the motor. Still another motor was essentially a variable capacitor with a rotatable shaft. It had 15 plates in the stator and 14 in the rotor, each 0.03 centimeters thick; the gaps between stator and rotor plates were 0.015 centimeters (Fig. 39). The design of this motor is similar to the design of the electrostatic variable-capacitance generator described by N. J. Felici.[6]

According to Bollée, the maximum average torque on the rotor in a capacitor-type electrostatic motor is $T_{\max} = kNV^2$, where k is a geometrical constant, and V is the peak voltage applied. A motor with many electrodes is therefore slow but produces greater torque. The power of the motor is proportional to the square of the applied voltage and to the frequency but does not depend on the number of electrodes.

An important parameter in capacitor-type electrostatic motors is the capacitance difference $C_{\max} - C_{\min}$, where $C_{\max}$ is the greatest capacitance between the rotor and the stator and $C_{\min}$ is the smallest capacitance as they oc-

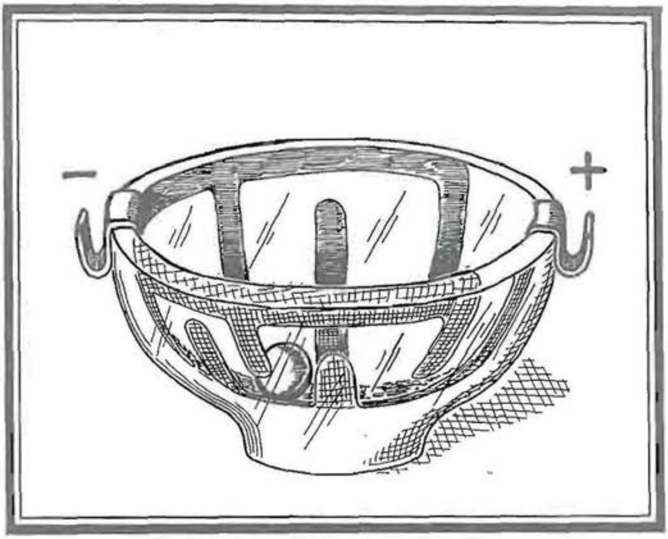

FIGURE 40

This interesting motor has been described by Moore, who calls it the "Interdigital Motor". The stator consists of a glass bowl with strips of aluminum foil glued to it. The rotor is a single conducting ball inside the bowl.

cur when the rotor turns. The larger this parameter is, the better are both the torque and the power. The main objective in the design of such motors is therefore to obtain the greatest possible variation of the capacitance.

An interesting motor which can be classified either as a modified Franklin's second motor or as a modified capacitor-type motor has been described by Moore,[7] who calls it the Interdigital Motor. The stator consists of a glass bowl with strips of aluminum foil glued to it (Fig. 40). The rotor is a single conducting ball inside the bowl; the ball runs along the sloping side when the strips are connected to an electrostatic generator (positive and negative strips are connected in alternation). Several balls can be placed into the bowl at the same time, and all of them will then run.

REFERENCES

1. "Zipernowsky Electrostatic Motor", *Electrical World*, Vol. 14, p. 260 (1889).
2. W. Thomson, *British Association Report*, 1855 (2), p. 22.
3. N. G. van Huffel, *Zeitschrift für den Physikalischen und Chemischen Unterricht*, Vol. 17, pp. 316-317 (1904).
4. W. Thomson, *British Association Report*, 1867, pp. 489-513.
5. B. Bollée, *Philips Technical Review*, Vol. 30, pp. 178-194 (1969).

6. N. J. Felici, *Electronics and Power*, Vol. 11, pp. 169-171 (1965).
7. A. D. Moore, *Electrostatics*, Doubleday and Company, Inc., New York, 1968, p. 109.

7

INDUCTION MOTORS

In 1892-1893, electrostatic motors of a fundametally new kind were described by Riccardo Arno[1,2] and by W. Weiler.[3] Their motors operated on the principle that the polarization of a dielectric in a variable electric field lags behind the field inducing the polarization. Therefore if a dielectric body is placed into a rotating electric field, the polarization charges* induced on the body will experience a force causing the body to follow the rotation of the field.

Arno's motor (Fig. 41) had a cylindrical stator formed by four insulated copper segments enclosing the space in which the rotating field was to be produced. The rotor was a hollow closed ebonite cylinder supported on two steel points turning in holes in glass; the cylinder weighed 40.33 grams, was 18 centimeters long and 8 centimeters in diameter. The rotating field was produced by means of a high-voltage transformer (3800 volts output), an RC-circuit, and a mercury commutator. The rotor attained a speed of about 250 rpm and developed a torque of 176 $\text{centimeter}^2 \times \text{gram} \times \text{second}^{-2}$.

---

*See, for example, O. Jefimenko "Electricity and Magnetism", Appleton-Century-Crofts, Inc. (1966), pp. 245-249.

Weiler used a hand-operated double commutator for his motors. The commutator delivered high voltage from an electrostatic generator in sequence to the four segments of a stator, thus producing the needed rotating electric field. Weiler described four different motors utilizing rotating fields. One of his motors was similar to Arno's motor. The operation of this motor (and that of Arno's motor) can be explained with the aid of Fig. 42, representing the top view of the motor. When segment A of the stator is positive, and segment B is negative, the electric field of the stator induces in the dielectric rotor (assumed to be at rest) polarization charges as shown in the figure. After the commutator has completed ¼ revolution, segment C is positive and segment D is negative. Thus the field has rotated by 90°, and now it polarizes the rotor in a new direction (at 90° with respect to the first). However, since it takes a certain time for the polarization charges to relax to zero, some of the initial polarization charges are still on the rotor, so that the distribution of the polarization charges on the rotor is not symmetrical relative

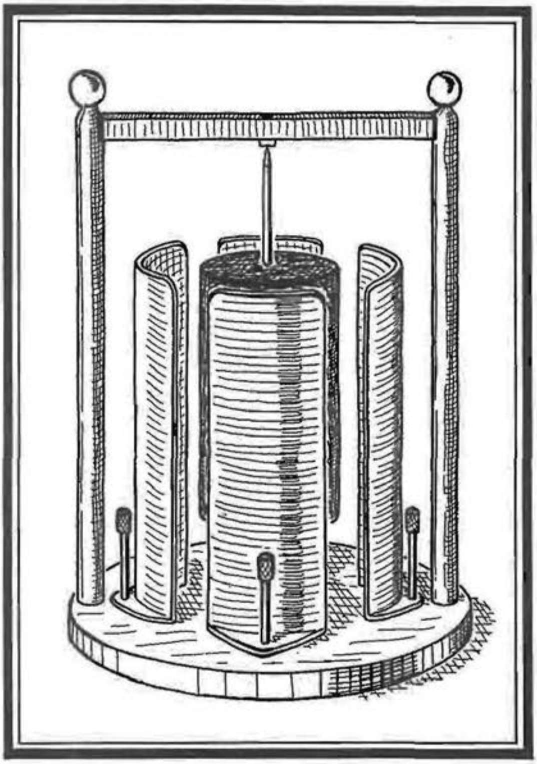

FIGURE 41

Arno's induction motor operated on the principle that the polarization of a dielectric in a variable electric field lags behind the field.

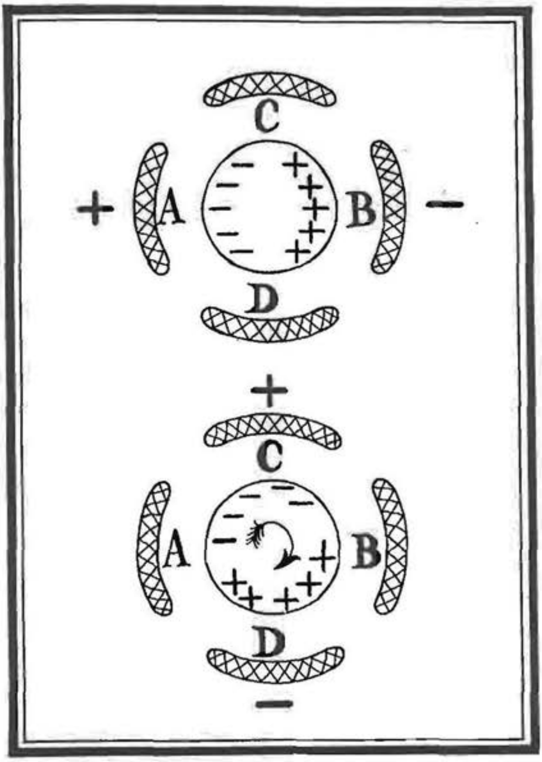

FIGURE 42

The operation of Arno's and Weiler's induction motors can be explained with the aid of this diagram.

to the charges on the stator. The rotor experiences therefore a torque due to attraction and repulsion between the rotor charges and stator charges, and the rotor turns.

Weiler also experimented with motors having noncylindrical rotors. One such motor is shown in Fig. 43. Rotating field electrostatic motors with disk-shaped rotors were built and studied between 1894 and 1901 in Japan by Hiderato Ho.[4]

Since Arno's and Weilers motors depend for their operation on induced polarization charges, they may be called "induction motors".

Induction motors, too, have a number of appealing features which attracted considerable interest to such motors in recent years: the motors can operate from both dc and ac; the motors can operate from low-voltage sources; since the motors require no brushes or slip rings, friction losses in these motors are very small.

Several carefully designed induction motors (both cylindrical and flat) were described in 1969 by Bollée.[5] One such motor of cylindrical geometry operated from a three-phase power supply at 220 volts and 50 hertz. The stator had 12 electrodes, which surrounded a rotor 15 centimeters in diameter. A disk-type induction motor operating from a three-phase power supply had a rotor consisting of 10 glass disks, each 6 centimeters in diameter, coated with a thin layer of slightly conductive material and mounted on a common axle. The rotor turned inside a ring

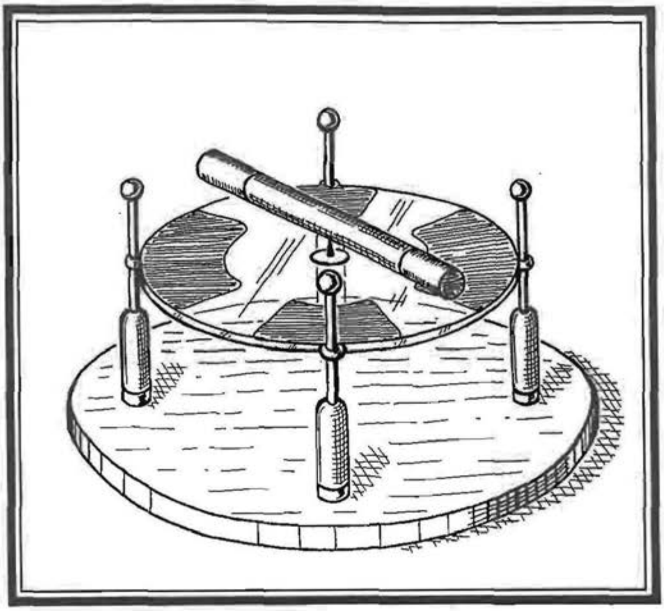

FIGURE 43

In 1892-1893, Weiler also experimented with induction motors having noncylindrical rotors, such as the motor shown in this drawing. Induction motors have a number of appealing features which attracted considerable interest to such motors in recent years.

formed by 60 comb-shaped stator electrodes. The maximum torque was 2000 centimeter^2 × gram × second $^{-2}$ at 220 volts and 50 hertz.

Another interesting motor described by Bollée was a linear induction motor, which "may be regarded as a segment of rotating-field motor bent straight".[5]

An extensive mathematical analysis of the induction motors was published in 1970 by Soon Dal Choi and D. A. Dunn.[6] They tested their theoretical results on a cylindrical motor with 72 electrodes. The motor was 14 inches wide, 14 inches high, and 19¼ inches long. It had a Plexiglas rotor that turned at a maximum rate of about 1000 rpm when powered by a 10-kilovolt source.

REFERENCES

1. R. Arno, *Rediconti della Atti Reale Accademia dei Lincei*, Vol. 1 (2), p. 284 (1892).
2. R. Arno, *The Electrician*, Vol. 29, pp. 516-518 (1892).
3. W. Weiler, *Zeitschrift für den Physikalischen und Chemischen Unterricht*, Vol. 7, pp. 1-4 (1893).
4. "A Rotary Field Electrostatic Induction Motor", *Electrical World*, Vol. 37, pp. 1012-1014 (1901).
5. B. Bollée, *Philips Technical Review*, Vol. 30, pp. 178-194 (1969).
6. S. D. Choi and D. A. Dunn, *Proceedings of the IDEE*, Vol. 59, pp. 737-738 (1971).

8

LIQUID-IMMERSED MOTORS

In 1893, W. Weiler discovered that a glass cylinder placed in a poorly conducting liquid between two spherical electrodes began to rotate when the electrodes were connected to an electrostatic generator.[1] He then constructed a small motor based on this discovery (Fig. 44). The principle of the operation of this motor was essentially the same as that of a cylindrical corona motor, except that the charges were now deposited on the rotor not by a corona discharge, but by the conduction current in the liquid.

In 1896, G. Quincke reported the same phenomenon and published a very comprehensive experimental study of it.[2] His device is shown in a simplified form in Fig. 45. It is interesting to note that Quincke, rather than Weiler, has been credited by subsequent investigators with the discovery of the rotation of dielectric bodies in

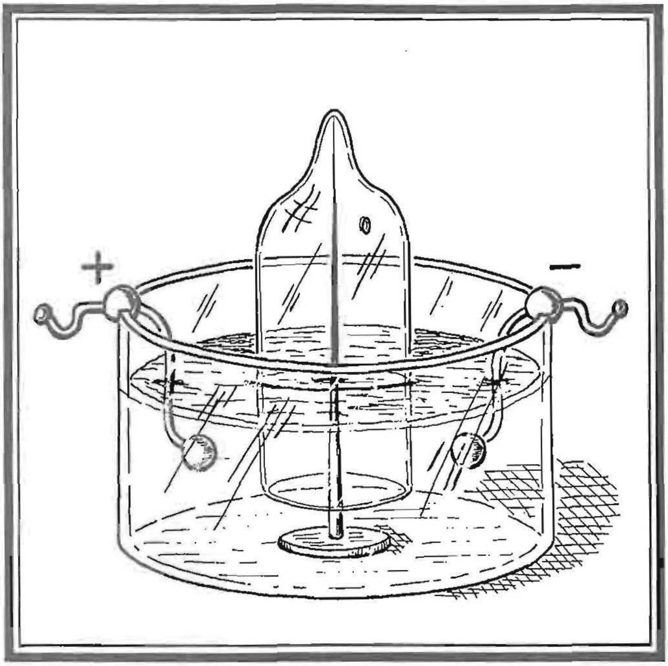

FIGURE 44

In 1893, W. Weiler discovered that a glass cylinder placed in poorly conducting liquid between two spherical electrodes began to rotate when the electrodes were connected to an electrostatic generator.

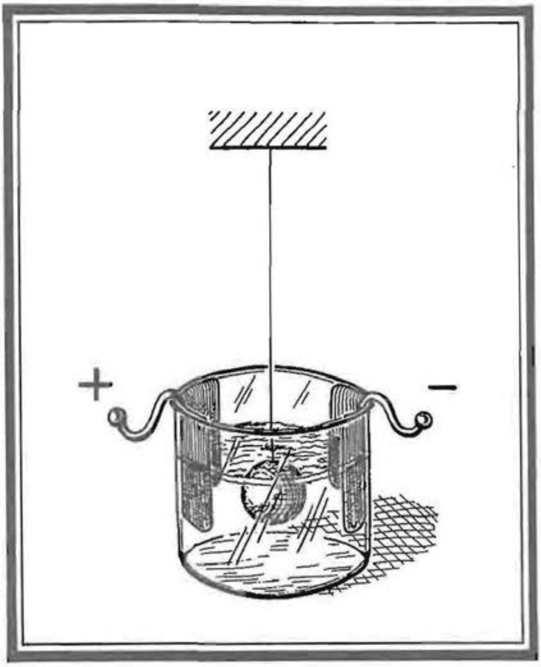

FIGURE 45

In 1896, G. Quincke reported the same phenomenon of rotation and published a very comprehensive experimental study of it. His liquid-immersed motor is shown in a simplified form in the above drawing.

poorly conducting liquids.

Weiler-type liquid-immersed dielectric motors were recently studied in considerable detail by P. E. Secker and his co-workers.[3,4] Their motors operated in various semi-insulating liquids such as hexane, hexane doped with amyl alcohol or ethyl alcohol, and isoamyl alcohol.

In one of their motors the electrodes were stainless steel squares 1.8 centimeters on a side. Their separation was varied from 1 to 2.5 centimeters. The rotor, of ebonite or Perspex, was 0.95 centimeters in diameter and 2.2 centimeters long. The operating voltage was from 3 to 30 kilovolts. The speed of the motor was found to be a linear function of the applied voltage, with the maximum recorded speed of about 2500 rpm.

In another motor, a stator with six electrodes was used (Fig. 46). The rotor was made of Perspex and was 7/8 inches in diameter and 2 inches long; it had a layer of high-permittivity material on its surface (titanium ceramic in polystyrene). Powered by a 30-kilovolt source, the motor turned at 1700 rpm, and the total power input amounted to 5.4 watts, of which 2.7 watts appeared at the rotor shaft.

A discussion of electrohydrodynamic effects that may take place during an operation of liquid-immersed motors has been presented by J. R. Melcher and G. I. Taylor.[5]

Because of considerable hydrodynamical losses, the motors of this type will apparently always have a relatively low output torque.

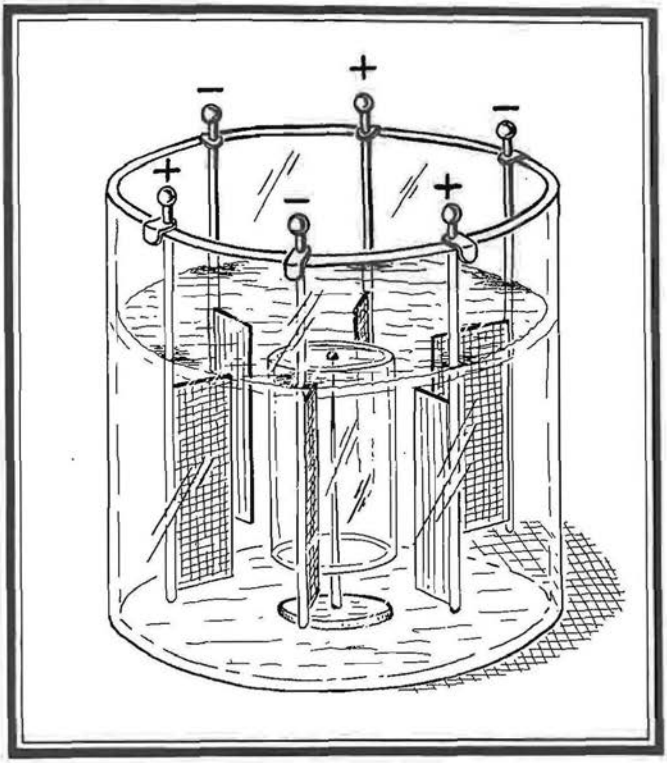

FIGURE 46

This liquid-immersed dielectric motor was recently studied in considerable detail by P. E. Secker and his co-workers.

REFERENCES

1. W. Weiler, *Zeitschrift für den Physikalischen und Chemischen Unterricht*, Vol. 6, pp. 194-195 (1893).
2. G. Quincke, *Annalen der Physik*, Ser. 3, Vol. 59, pp. 417-486 (1896).
3. P. E. Secker and I. N. Scialom, *Journal of Applied Physics*, Vol. 39, pp. 2957-2961 (1968).
4. P. E. Secker and M. R. Belmont, *Journal of Physics D: Applied Physics*, Vol. 3, pp. 216-220 (1970).
5. J. R. Melcher and G. I. Taylor, *Annual Review of Fluid Mechanics*, Vol. 1, pp. 111-146 (1969).

9

ELECTRET MOTORS

In general, an electret is a permanently charged dielectric body. In a more restricted sense of the word, an electret is a permanently polarized dielectric body and may be considered to constitute the electrical counterpart of a permanent magnet. It is this latter type of electret that has been used for constructing electrostatic motors.

In 1961, the Russian physicist A. N. Gubkin described an electret motor[1] shown schematically in Fig. 47. The motor consisted of a stator formed by two horizontal, axially symmetric, parallel-plate capacitors and a rotor formed by two flat, axially symmetric, oppositely polarized electrets mounted on a vertical axle and capable of passing between the plates of the capacitors. The motor operated as follows. When a voltage was applied to the two capacitors, so that

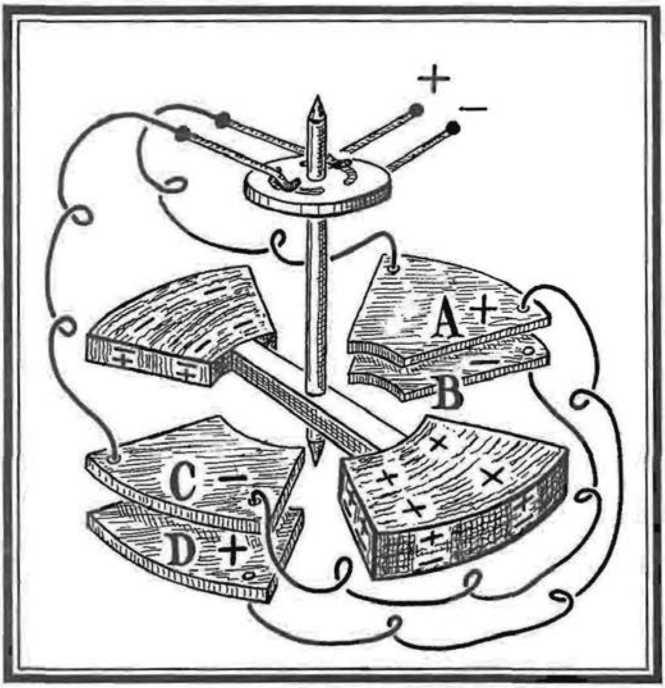

FIGURE 47

In 1961, the Russian physicist A. N. Gubkin described this motor with two electrets in the rotor and two parallel-plate capacitors in the stator. A commutator was used to change the polarity of the capacitors.

plates A and D were positive and plates B and C were negative, the capacitors attracted the electrets as soon as the latter came close to the plates, and the electrets were pulled into the capacitors. A commutator changed the polarity of the plates just as the electrets were coming out from the capacitors, and the electrets were then repelled from them, and so on. Thus the rotor was set in a continuous rotation.

A well-known property of electrets is that they lose their polarization in the absence of adequate shielding. Electret motors with almost perfect shielding were described by this author and by D. K. Walker in 1970.[2] These motors were based on the so-called "electret slot effect".[3] The slot effect works as follows (Fig. 48). An electret is placed between two pairs of electrodes, each pair forming a slot. If a voltage is applied to one or both electrode pairs, the electret (or the electrodes) experiences a force in a direction perpendicular to the slot and parallel to the plates. This arrangement insures both a nearly perfect shielding and a relatively large force.

A simple electret motor utilizing the slot effect is shown in Fig. 49 and Plate 5.* It uses a disk-shaped carnauba wax electret rotor consisting of two oppositely polarized half-disks. The stator has two pairs of electrodes connected to a

---

*The motor was designed by the author and built by one of his students, Charles Lynn Walls, in 1966.

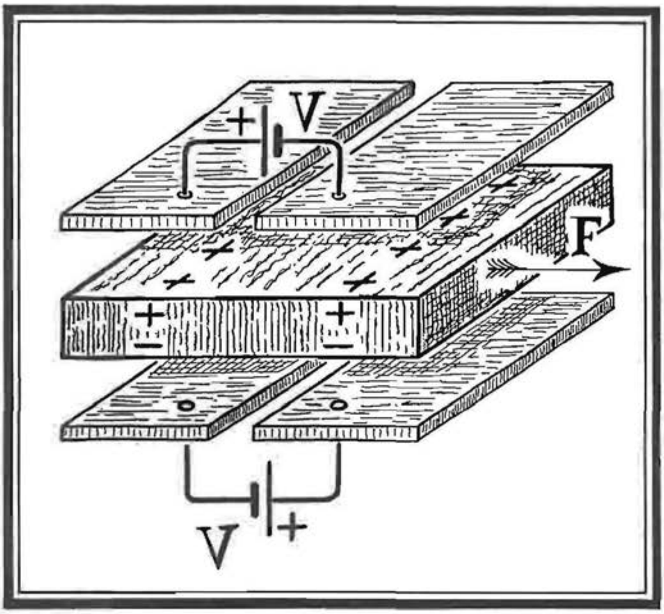

FIGURE 48

The slot effect works as follows. An electret is placed between two pairs of electrodes, each pair forming a slot. If a voltage is applied to one or both electrode pairs, the electret experiences a force in a direction perpendicular to the slot and parallel to the plates.

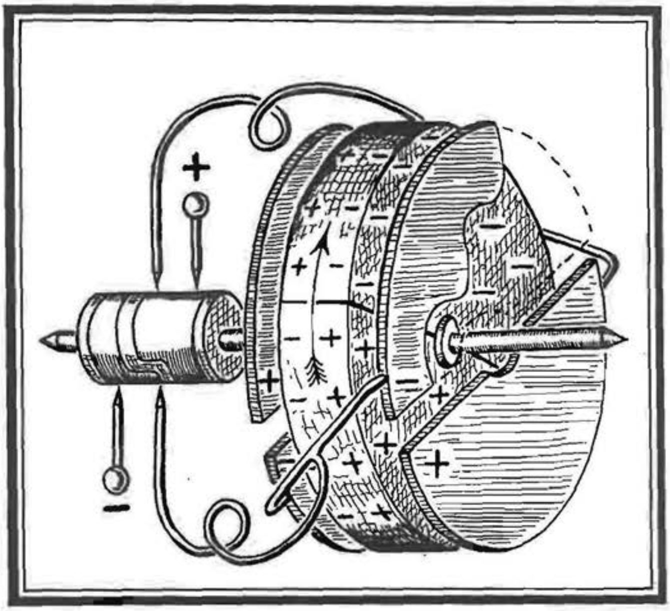

FIGURE 49

This simple electret motor utilizes the slot effect. The rotor is made of a disk-shaped electret consisting of two oppositely polarized half-disks. The stator has two pairs of slotted electrodes connected to a cylindrical commutator, which is charged by contact or by sparks.

cylindrical commutator. The thickness of the rotor is ½ inch, the diameter is 5 inches. The motor operates from an 8-kilovolt power supply and rotates at 1500 rpm.

Similar motors with stationary electrets and rotating electrodes are shown in Plates 6 and 7.

A more sophisticated slot-effect motor[2] uses a cylindrical electret (Fig. 50 and Plate 8). The electret is stationary; it is shaped as a hollow cylinder and has four sections of opposite radial polarization. The rotor is made of four internal and four external aluminum electrodes forming two cylinders with four slots in each. The inner electrodes are cross-connected with the outer ones, and all electrodes are supported by a Plexiglas disk mounted on a steel axle. This motor uses no commutator. The power is delivered to two adjacent external electrodes by means of two sharp points which charge the rotor through a corona discharge. The overall diameter of this motor is 3 inches, the operating voltage is 6 kilovolts, the speed is up to 5000 rpm, the power is about 20 milliwatts.

Synchronous electret motors for electric clocks have been recently announced by the General Time Corporation. The motors are about 1¼ inches in diameter and ¼ inch thick. In these motors (Fig. 51) a thin plastic electret disk with 15 active sectors and 15 equally large cut-outs is the rotor. The rotor is placed between two stator plates fabricated as printed circuit boards. Each stator plate has 30 elec-

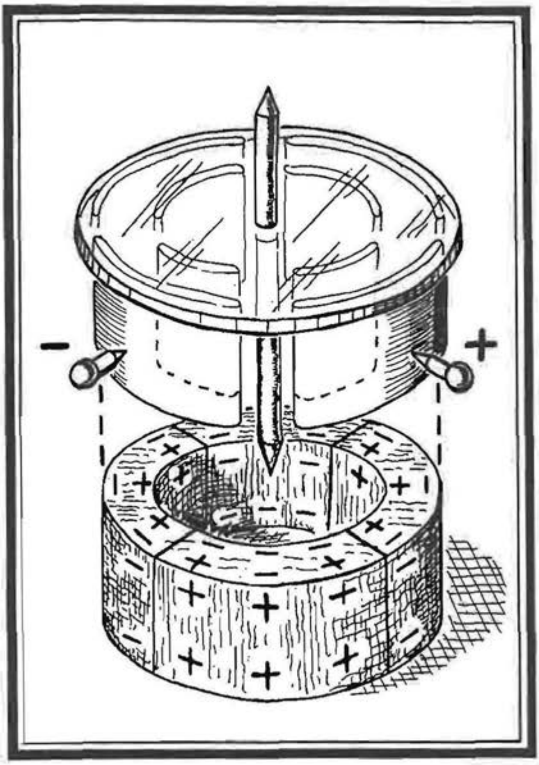

FIGURE 50

This more advanced slot-effect motor uses a stationary cylindrical electret and works without a commutator. The rotor has 8 slots.

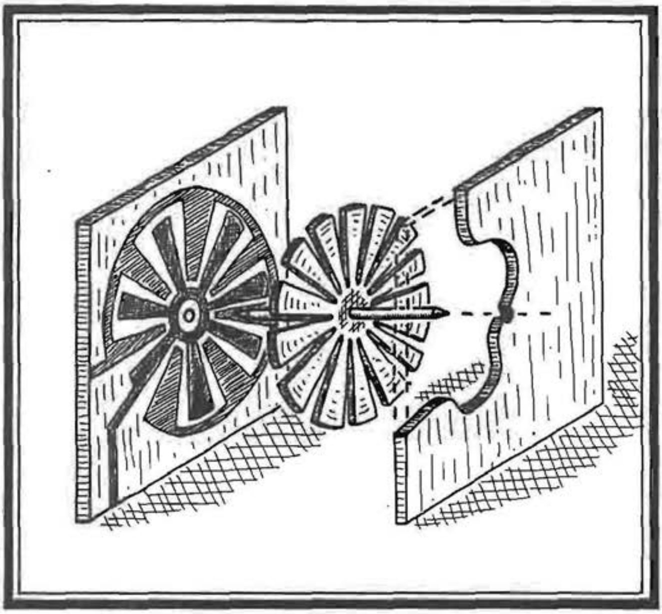

FIGURE 51

In this simplified version of a synchronous electret motor for electric clocks a thin plastic electret disk with several active sectors and several cut-outs is used as the rotor.

trodes connected in alternation to the two input terminals. The operation of this motor is similar to that of the first slot-effect motor described above (Fig. 49) except that the reversal of polarity of the electrodes is accomplished directly by the applied ac voltage rather than by a commutator.

Of all presently known types of electrostatic motors, electret motors are the newest and (together with corona motors) the most highly promising ones. We shall return to them once again in the next chapter.

REFERENCES

1. A. N. Gubkin, *Electrets*, Academy of Sciences, Moscow, 1961, pp. 130-133 (in Russian).
2. O. Jefimenko and D. K. Walker, *Conference on Dielectric Materials, Measurements and Applications*, The Institution of Electrical Engineers, London, 1970, pp. 146-149.
3. O. Jefimenko, *Proceedings of the West Virginia Academy of Sciences*, Vol. 40, pp. 345-348 (1968).

10

WHAT TO EXPECT FROM ELECTROSTATIC MOTORS

Having described the various types of electrostatic motors we shall now present a brief discussion of the current aims in the electrostatic motor research and a brief discussion of the most probable future uses and applications of these motors.

It has been pointed out by Bollée[1] that electromagnetic motors rapidly lose their efficiency in scaled-down versions (which is due to the relative increase in energy dissipation in the magnet coils) and that very small capacitor-type and induction-type electrostatic motors may be a better choice for miniaturized systems. Therefore one may expect that miniature electrostatic motors of these types will find applications in various sensor and control devices where only very small torques and power are needed.

Experiments with electret motors conducted in the author's laboratory[2] indicate that these motors may be very useful in systems where powers of up to 1 watt are needed.

The most promising electrostatic motors appear to be, however, the corona motors. These motors possess a number of highly desirable features. Here are some of them: the motors are extremely simple in design; they require no expensive materials; their maintenance is very simple; having only few metal parts they possess a very good power-to-weight ratio; they are fully capable of developing appreciable amounts of power; and they can attain very high speeds.

Our present-day awareness of the many attractive properties of corona motors is to a great extent due to the work of the Russian engineers Yu. Karpov, V. Krasnoperov, and Yu. Okunev published in 1958 and 1960.[3,4] They described a 6-watt cylindrical corona motor of improved design (Fig. 52) operating from a 7-kilovolt power supply and turning at a rate of 6000 rpm. This motor had a hollow Plexiglas rotor 10.5 centimeters in diameter and 17 centimeters long with a conducting lining on its inner surface. The stator supported 16 knife-like electrodes inclined relative to the surface of the rotor in the direction of the desired rotation. The lining of the rotor served to increase the electric field in the gap between the electrodes and the surface of the rotor and thus to enhance the corona discharge from the electrodes.

A high-speed corona motor with a disk rotor

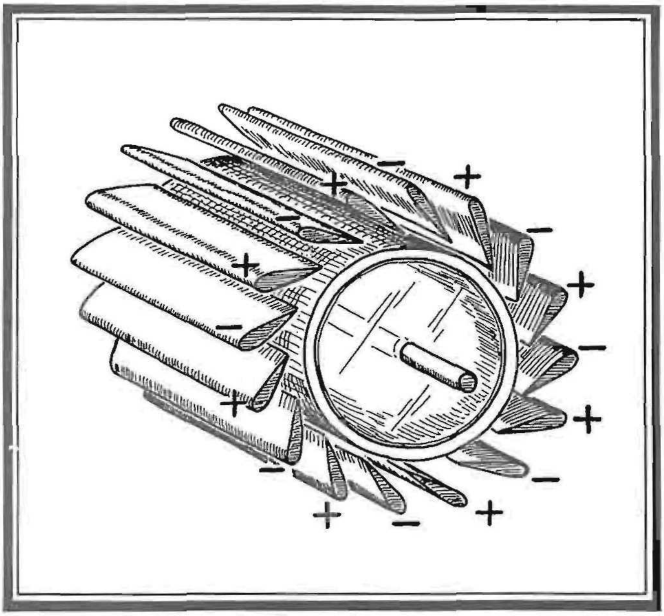

FIGURE 52

In 1958, the Russian engineers Yu. Karpov, V. Krasnoperov, and Yu. Okunev described a 6-watt cylindrical corona motor of an improved design operating from a 7-kilovolt power supply and turning at a rate of 6000 rpm. This motor had a Plexiglas rotor.

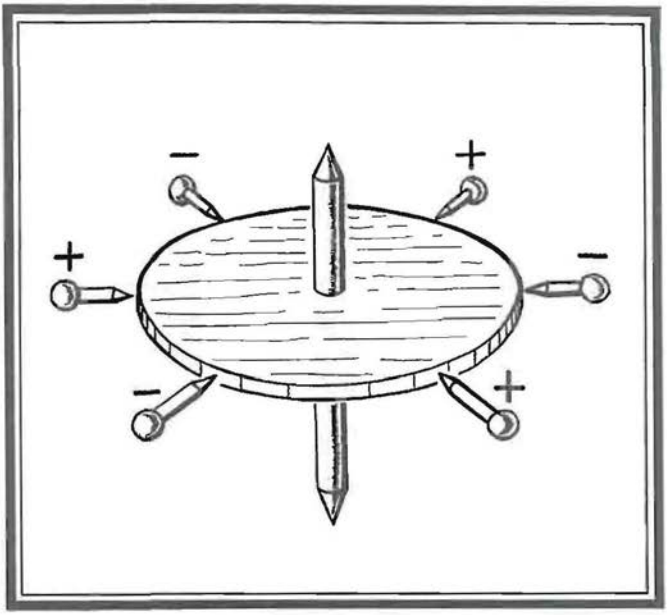

FIGURE 53

This corona motor with a disk rotor and circumferential electrodes was described by J. D. N. Van Wyck and G. J. Kühn of South Africa in 1961. It had a rotor 1.5 inches in diameter turning in jewel bearings. Operating from 13 kilovolts, the motor ran at 12000 rpm.

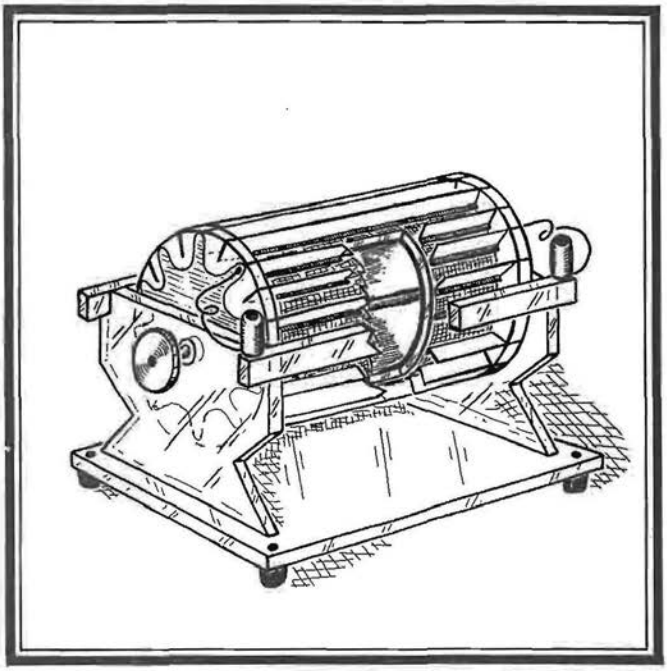

FIGURE 54

The performance of a corona motor depends on the shape and arrangement of electrodes and the structure of the rotor. The curved electrodes of this motor produce an especially large torque. The lining of the rotor increases the power.

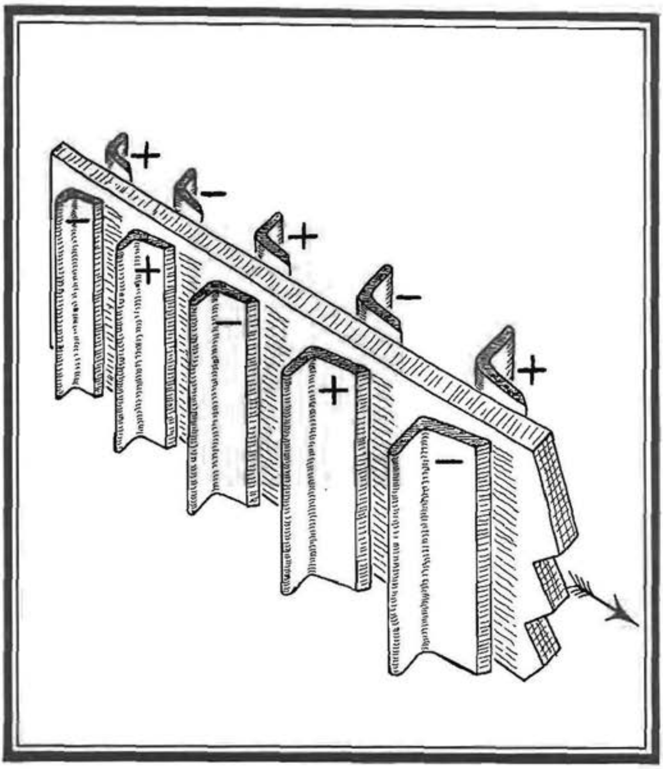

FIGURE 55

The mechanism of electric interaction producing rotational motion in corona motors can be used to achieve a translational motion as well. This drawing shows the principle of a linear corona motor with multiple electrodes.

and circumferential electrodes was described by J. D. N. Van Wyck and G. J. Kühn of South Africa in 1961.[5] The motor (Fig. 53) had a rotor 1.5 inches in diameter turning in jewel bearings. The stator supported 6 sharp-point electrodes. Operating from 8-13 kilovolts, the motor developed speeds of up to 12000 rpm.

A similar motor was studied in Poland by B. Sujak and W. Heffner in 1963.[6]

A number of advanced corona motors were studied in the author's laboratory;[7] an example of these motors is given in Fig. 54 and Plate 9. A diagram of a linear corona motor designed by the author is shown in Fig. 55. Instructions for building the corona motors of the author's design may be found in the May 1971 issue of Popular Science.[8]

It appears that corona motors with input power of 100 to 1000 watts and efficiency of at least 60% can be constructed without any difficulty. There seems to be no reason why even more powerful corona motors could not be built. It is likely that the motors can be further improved by using rotors immersed in a gas other than the air at atmospheric pressure.

Performance characteristics of corona motors strongly depend on the design and geometrical configuration of the electrodes spraying charges onto the rotor. Several typical electrode types and electrode configurations for corona motors are shown in Fig. 56. Especially interesting is the electrode arrangement shown in Fig. 56c.

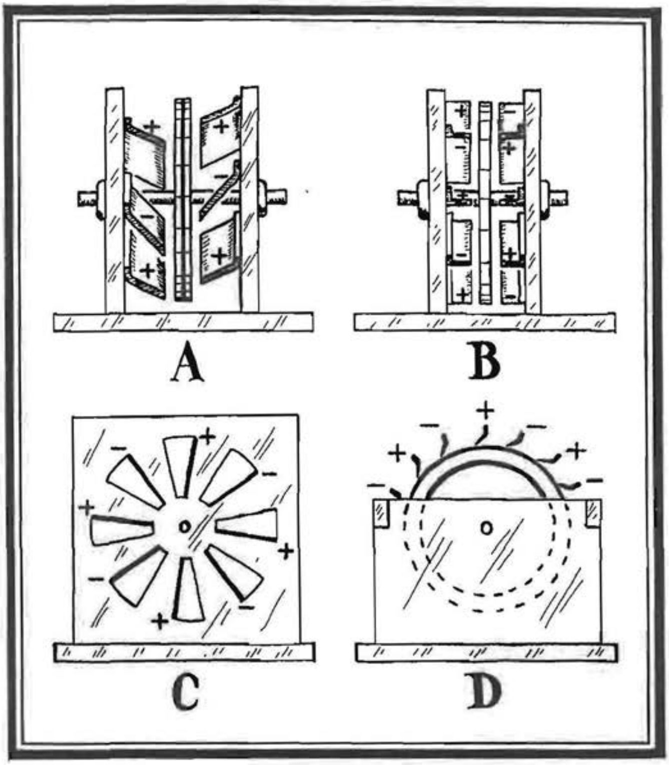

FIGURE 56

(a) Slanted electrodes and disk rotor with a conducting lining in the middle; (b) Straight electrodes; (c) Foil electrodes shielded on one side; (d) Curved electrodes and cylindrical rotor with a conducting lining.

This arrangement is used in conjunction with thin disk-shaped rotors "sandwiched" between two insulating stator plates. An even number of windows is cut in each stator plate, and metal foil electrodes are glued to the walls of the holes, as shown by heavy lines in Fig. 56c. In this arrangement there is a dielectric medium on one side of each electrode, and there is air on the other side. Therefore the electrodes spray charges predominantly in the direction of the side facing the air. Thus a motor with electrodes of this type is unidirectional and self-starting. Furthermore, the entire "sandwich" comprising the rotor and the two stator plates may be very thin. Therefore many such "sandwiches" can be placed on a single axle thus forming a compound motor in which good use is made of the entire volume occupied by the motor. There are indications that compound motors of this type can develop up to 1000 horsepower for each cubic meter of their volume.

An important property of electrostatic motors is that they can operate from a much greater variety of sources than the electromagnetic motors. It is, of course, obvious that any sources used for operating conventional electromagnetic motors can be used to operate the electrostatic motors. However, the electrostatic motors can operate also from sources from which no other motors can operate. The reason for this is that the electrostatic motors are extra-high impedence devices and thus require extremely small currents for their operation.

One very interesting source of electricity for electrostatic motors is the ordinary capacitor. It is true that capacitors, as we know them now, do not store appreciable amounts of electric energy. However, if high-permittivity low-loss dielectric materials were developed, then the capacitor would become a very useful device for storage of electric energy, and possibly could be used in place of the chemical batteries for operating electrostatic motors.

Another possible source of power for electrostatic motors are high-impedance high-voltage batteries of the type of Zamboni pile.

A potentially very important source of electricity for powering electrostatic motors are electrostatic generators. Considerable advances in the development of such generators have been made in recent years[9,10] and it is conceivable that electrostatic motor-generators will be used to convert the high-voltage dc produced by such generators into the conventional low-voltage dc or ac.

Finally, a very interesting source for powering electrostatic motors is the atmospheric electric field. In fact, it appears possible to extract the energy contained in this field by means of electrostatic motors. Experiments on such an energy extraction have been conducted by the author.[11,12] In these experiments an electret motor and a corona motor were powered by simple earth-field antennas (Fig. 57; Plates 10 to

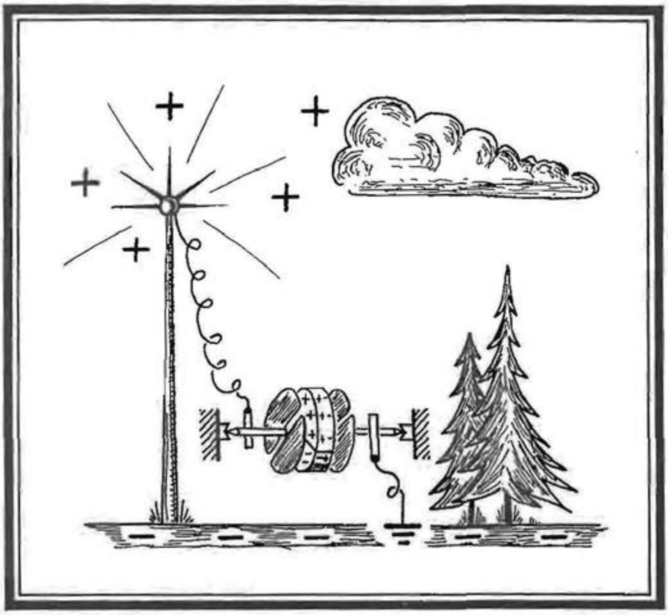

FIGURE 57

Small electrostatic motors can be powered from atmospheric electricity by means of simple earth-field antennas. Whether or not it will be possible to operate large motors in this manner will depend on how successful we are in designing antennas capable of extracting appreciable power from the earth electric field.

12). The corona motor was the one shown in Plate 9. The electret motor was similar to that shown in Fig. 49, but with the electret stationary and the electrodes rotating (Plate 6). These experiments indicate that it is entirely possible to operate small electrostatic motors from atmospheric electricity. Whether or not it will be possible to operate large motors in this manner will depend on how successful we will be in designing and building earth-field antennas capable of extracting appreciable power from the earth electric field.

In conclusion it may be useful to mention that a quantitative electrostatic motor research is now only in its very rudimentary stage. Almost all papers on electrostatic motors published thus far deal with the qualitative aspects of the performance of the various types of motors. Only very few papers deal with the optimization of design and with the quantitative aspects of the operation of the motors, likewise only very few papers present a theoretical analysis of the electrostatic motors. It is clear therefore that the electrostatic motors still constitute an essentially unexplored area of physics and engineering, and that the electrostatic motor research is presently one of the potentially most rewarding research fields in modern electrostatics.

REFERENCES

1. B. Bollée, *Philips Technical Review*, Vol. 30, pp. 178-194 (1969).

2. O. Jefimenko and D. K. Walker, *Conference on Dielectric Materials, Measurements and Applications*, the Institution of Electrical Engineers, London, 1970, pp. 146-149.
3. Yu. Karpov, V. Krasnoperov, and Yu. Okunev, *Tekhnika Molodezhi*, Vol. 26, pp. 36-37 (Sept. 1958) (in Russian).
4. Yu. Karpov, V. Krasnoperov, Yu. T. Okunev, and V. V. Pasynkov, "On the Motion of Dielectrics in an Electric Field", in *Dielectric Physics*, Academy of Sciences, Moscow, 1960, pp. 124-131 (in Russian).
5. J. D. N. Van Wyck and G. J. Kühn, *Nature*, Vol. 192, pp. 640-650 (1961).
6. B. Sujak and W. Heffner, *Acta Physica Polonica*, Vol. 23, pp. 715-726 (1963).
7. O. Jefimenko and H. Fischbach-Nazario, *Proceedings of the West Virginia Academy of Sciences*, Vol. 42, pp. 216-221 (1971).
8. C. P. Gilmore and W. J. Hawkins, *Popular Science Monthly*, Vol. 198, pp. 95-97, 114 (May 1971).
9. N. J. Felici, *Elektrostatische Hochspannungs Generatoren*, G. Braun, Karlsruhe, 1957.
10. J. Hughes and P. Secker, *New Scientist and Science Journal*, Vol. 49, pp. 468-470 (1971).
11. O. Jefimenko, *American Journal of Physics*, Vol. 39, pp. 776-778 (1971).
12. W. Aston, *West Virginia University Magazine*, Vol. 3, Number 4, Spring 1971, pp. 6-11.

INDEX

INDEX

Adams, G., 37
Aimé, R., 29,38
Arno, Q., 84,90
Arno Motor, 84,85,87
Aston, W., 118
Asynchronous motors, 4, 77
Atmospheric electricity, 3, 115-117
Avery, F. M., 21

Backing plate, 59, 107
Behrens, G., 14
Bell jar motor, 48, 49, 52
Belmont, M. R., 96
Benjamin, P., 21
Bichat, E., 29, 32, 38
Bollée, B., 77, 80, 82, 90, 106, 117
Bollée capacitor motors, 77, 79, 80
Bollée induction motors, 88, 90

Capacitor motors, 4, 73-77, 80, 82, 106
Chew, V. K., 37
Choi, S. D., 90
Christenson, A., 38
Christiansen, C., 60, 71
Christiansen motor, 60
Classification of electrostatic motors, 4
Collinson, P., 39
Contact motors, 4, 6, 20
Corona discharge, 22, 56, 71
Corona-kraft levitation, 33
Corona motors, 4, 56, 58, 60-66, 68-71, 107-112, 115, 116

Dailey, H. B. and E. M., 15, 21
Dailey motor, 15-18
Davis, J. W., 63
Draper, J. W., 55
Dressel, R. H., 51, 55
Dunn, D. A., 90

Earth-field antennas, 115, 116, 117
Electret, 97, 99
Electret motors, 4, 97-104, 106, 114, 115
Electret slot effect, 99, 100
Electric bells, 6, 8-11, 23, 25
Electric chimes, 11, 14, 23, 25
Electric fly, 22-26, 29-31
Electric horse-race, 26, 28, 29

Electric motor, 1, 6, 39
Electric orrery, 26, 27
Electric pendulum, 6
Electric pendulum motors, 6, 15, 16
Electric perpetuum mobile, 3, 15, 16, 48, 50
Electric pinwheel, 22
Electric planetarium, 18, 19
Electric rabbit hunt, 33, 36
Electric racing ball, 18, 19
Electric reaction wheel, 22
Electric seesaw, 13, 14
Electric swing, 12, 14
Electric tellurium, 26, 27
Electric tourbillion, 61, 63, 71
Electric tourniquet, 29, 32
Electric toys, 12, 13, 14, 28, 29, 33, 36
Electric wheels, 37, 39
Electric whirl, 22
Electric wind, 23
Electric wind motors, 22 ff, 33-35
Electric wind turbines, 33, 35
Electric wind wheels, 33, 34
Electromagnetic motors, 1-3, 106, 114
Electrophorus, 52
Electrostatic forces, 1
Electrostatic generators, 80, 115
Electrostatic motors, 1-4, 106, 107, 114-117

Farrington, J. B., 63
Felici, N. J., 80, 83, 118
Franklin, Benjamin, 7, 33, 37, 39, 42, 43
Franklin bells, 7, 14
Franklin first motor, 39-45, 48, 52
Franklin motors, 39-45, 54, 57, 59
Franklin second motor, 41, 43-45, 82

Gale, L. D., 33, 38
Gilmore, C. P., 72, 118
Gordon, A., 6, 22
Gordon bells, 6, 8, 15, 18, 20
Gruel, W., 63
Gruel motor, 61, 63
Gubkin, A. N., 97, 105
Gubkin motor, 97-99
Guyot, M., 20, 21, 29, 33, 38, 55

Hamilton, 23
Hamilton electrometer, 23
Hawkins, W. J., 72, 118
Heffner, W., 112, 118
Ho, H., 88
Holtz, W., 56, 71
Holtz rotation phenomenon, 56, 59, 60
Van Huffel, N. G., 76, 82
Van Huffel motor, 75, 76
Hughes, J., 118

Induction motors, 4, 84-90, 106
Interdigital motor, 81, 82

Jamin, M. J., 55
Jefimenko, O., 54, 84, 105, 117, 118
Johnson, R., 66
Johnson motors, 64, 65, 66

Kämpfer, D., 29, 38
Karpov, Yu., 107, 108, 118
King, W. R., 33
Kinnersley, E., 26
Krasnoperov, V., 107, 108, 118
Kühn, G. J. 110, 118

Langenbucher, J., 23, 29, 33, 37
Langenbucher electrometer, 23, 24
Levitation, electrostatic, 33
Leyden jar, 6, 42, 43
Linear corona motor, 111, 112
Linear induction motor, 90
Liquid-immersed motors, 4, 91-95
Loeb, L. B., 38

McVay, W., 66
McVay motors, 66-71
Magnetic forces, 1
Mascart, M. E., 29, 38, 63, 72
Mayes, T. L., 71
Melcher, J. R., 94, 96
Miller, D. C., 55
Millifleapower motor, 48, 51
Moller, S., 38
Moore, A. D., 45-47, 55, 81-83
Müller-Pouillet, 21

Neuburger, A., 21, 33, 38

Okunev, Yu., 107, 108, 118
Oosting, H. J., 38

Pendulum motors, 3, 6, 15, 16
Perpetuum mobile, electric, 3, 15, 16, 48, 50
Poggendorff, J. C., 37, 56, 59, 60, 71
Poggendorff motor, 56-60, 66, 71

Quincke, G., 91, 93, 96

Reciprocating motor, 15, 17, 18
Richmann, G. W., 7
Rotating field motors, 84-90
Ruhmkorff, 63
Ruhmkorff motor, 62, 63

Scialom, I. N., 96
Secker, P. E., 94, 96, 118
Secker motor, 94, 95
Sivowitch, E. N., 18
Slot effect, 99, 100
Slot-effect motors, 99-103
Spark motors, 4, 39-41, 46, 47, 49, 50, 53, 54
Sparks, J., 21, 38, 54
Sujak, B., 110, 118
Synchronous motors, 4, 77, 102
Syng, P., 33

Taylor, G. I., 94, 96
Thomson, W., 82
Thomson electrometer, 73
Thomson replenisher, 76
Tomlinson, C., 29, 38

Walker, D. K., 54, 99, 105
Weiler, W., 84, 90, 91, 96
Weiler induction motors, 87-89
Weiler liquid-immersed motor, 91, 92
Wilson, B., 37
Wimshurst, J., 33
Wommelsdorf, H., 55
Van Wyck, J. D. N., 110, 118

Zamboni, G., 14
Zamboni pile, 3, 14, 15, 48, 115
Zipernowsky, K., 73, 77
Zipernowsky motor, 73, 74, 82

This book was set in Century Textbook.
The format was designed by the author.
The drawings were done by
Dr. D. K. Walker from the author's sketches.
The printer and binder was
Morgantown Printing & Binding Company.

Picture Credits

Plate 1: courtesy Smithsonian Institution, Washington, D.C.
Plates 2 and 3: courtesy Electret Scientific Company, Star City, W. Va.
Plates 6, 9, 11, and 12: courtesy West Virginia University, Morgantown, W. Va.
Plates 4, 5, 7, 8, and 10: author's photographs.

Visite 109 117
Pagine 244 336

MAREA SISTEMI ©

HOME

YOU ALSO PARTICIPATE IN THE SOCIAL NETWORK OF MAREASISTEMI

GOOGLE + FACEBOOK CONTACT FACEBOOK PAGE YOUTUBE

facebook
mareasistemi.com

Electricity from the earth's atmosphere. An old forgotten idea

Source: http://scienzaearte.blogspot.it/2011/03/energia-elettrica-dallatmosfera.html

Solar, wind, hydroelectric, geothermal, biogas. These are the main renewable and non-polluting energy sources. But sure not to leave something out?
It is known that the Earth's atmosphere has a potential difference that increases with height. For non-professionals it the atmosphere was a charged battery. This is evident during thunderstorms, when this charge finds its outlet in th frightening and fascinating phenomena that are lightning. Many have imagined and studied methods for storing ligh energy, but this presents numerous problems:

1) it is not known in advance where a lightning will strike or in any case where a thunderstorm will take place; 2) ev lightning strike could be intercepted, for example by attracting it with a kind of lightning rod, it would be extrem difficult to store all its energy, since the discharge is very short; 3) moreover, most of the lightning energy dissipat light, heat and sound. In short, a "power plant" based on the interception of lightning would not be very profitable. **W less known, however, is that the atmosphere has this charge even when there is not even a cloud in the sk**

. The idea of deriving electricity directly from the earth's atmosphere dates back more than a century ago. To m knowledge - I urge readers to post comments if they are aware of other studies - the latest and most promising study field was carried out by the Estonian engineer and inventor **Hermann Plauson** . His studies took place in the 1920 almost a century ago. Is it possible that since then there has been no progress in this field? Why did the idea of extra electricity from the Earth's atmosphere go into oblivion, even though it seemed promising at the time of Plauson's st

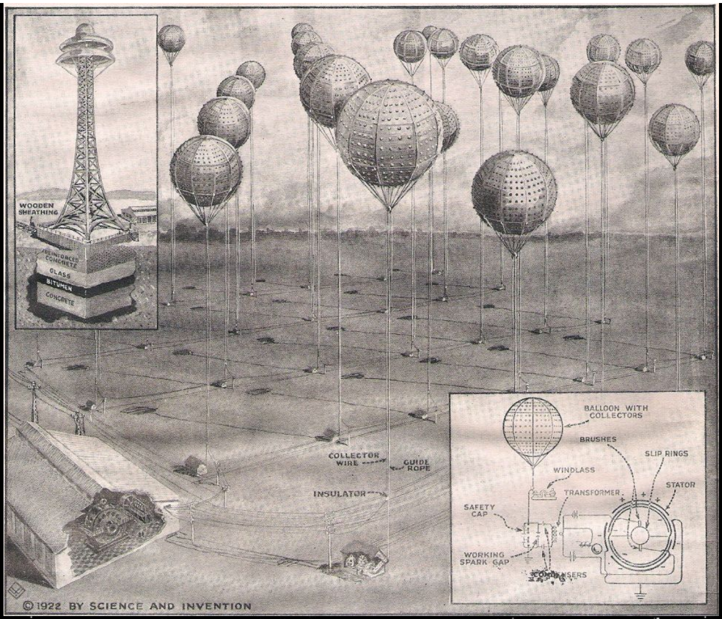

Image taken from the magazine "Science and Invention" (1922), where we describe the invention of Hermann Plau
Let's see before giving a brief description of the method used by Plauson. An electricity collector, consisting for exam
a balloon filled with helium, was raised to an altitude of several hundred meters, connected to the ground by a condu
cable. The balloon itself was made of conductive material and was covered with innumerable little needles, capabl
collecting atmospheric electricity through a phenomenon called field effect emission . High-voltage electricity w
conveyed to the ground by the conductor cable, and was converted into low-voltage electricity to be used by comm
electric users.

In the description of his invention , Plauson reports that he managed to get about 3.4 kW (kilowatt) in a "pilot" exper
using two of these balloons.
Returning to the first question: why such a promising technology was abandoned?
Several hypotheses can be made. The Plauson method provided - even if it was not bound - the use of radioactive ma
(radio and / or polonium), whose ionizing properties were used as a "catalyst" to increase the current inflow. It is ob
that today any alternative energy proposal based on the use of radioactive material would not be frowned upon,
considering the possible environmental impact that could result from accidental contamination. As a note
it is interesting *to mention* that until the eighties a type of lightning conductor covered with a thin layer of americiu
radioactive material), similar in concept to the Plauson collectors, was used in Italy: it is useless to say that these de
are currently become outlaws.
Another possible reason why the idea was abandoned may be related to the fact that the energy needs in the 20s wer
lower than today's ones: a typical "central" "at Plauson", composed of 100 separate balls the one from the other of a
100 meters (so 1 km of side), would have generated some hundreds of kilowatt: a power of all respect for those time
that pales a bit 'when compared to the energy produced by a single wind generator - about 1 Megawatt.
This reason, however, would justify only part of the abandonment of research in this field, since a system that allows
obtain even a few kilowatts using just a couple of balloons could be competitive - compared to wind energy - in area
windy and not served by the electricity network. **Let's** add at this point **an advantage that this system would hav
wind and solar: uninterrupted production 24 hours a day and 365 days a year** .
Apparently, therefore, the only major obstacle would seem to be represented by the use of radioactive material. And
we are meeting the recent technological development. In recent years, in fact, techniques have been developed to o
nanomaterials that are very efficient from the point of view of field effect emission: in practice they are super-pins
could function as effective collectors of atmospheric electricity without resorting to use of radioactive material.
Some might argue, and rightly so, "But are these particular nanomaterials safe from the health point of view?"
Unfortunately, since these are new technologies, it is still difficult to estimate the impact in this sense, but some studi
under way. If, as is hoped, these materials will be considered harmless to health, then it can not be excluded that t
research on electric energy obtained directly from the earth's atmosphere will find a new vigor and this source not
exploited can go alongside the best known renewable and non-polluting sources.
As a final note it should be added that even without using radioactive material and the nanomaterials mentioned abo

system similar to that of Plauson could provide modest but useful quantities of electricity, finding application especi
the areas of developing countries far from large centers inhabited.

Hermann PLAUSON

Conversion of Atmospheric Electricity

Source: Source: rexresearch.com

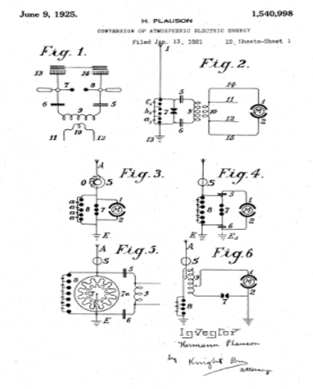

In 1920 the German engineer Hermann Plauson published a text "Gew und Verwertung der Atmospharischen Elektrizitat" (Acquisition and atmospheric electricity).

Where he concluded that atmospheric electricity could be captur exploited by men.

In this regard he wrote:

"Humanity will be free from worry about the threat of freezing death, knowing that in a pred time it will be possible to exploit the natural resources to the fullest ... Humanity will no longer suffer from the cold, because mother nature will give her abundant gift power."

Plauson designed a series of helium-filled balloons, able to collect atmospheric electricity in t through the use of thermionic rectifiers, Leida bottle condensers and induction coils, his idea supply energy to whole Germany.

Biography (Wikipedia)

Meridian International Research: Atmospheric Electricity Research

Science & Invention (Feb. 1922) : "Power from the Air" (I)

Science & Invention (March 1922) : "Power from the Air" (II)

Plauson's Patents (List)

H. Plauson: USP # 1,540,998 - Conversion of Atmospheric Electricity

H. Plauson: British Patent # 157,262 - Improvements in Electric Motors

H. Plauson: British Patent # 157,263 - Process & Apparatus for Converting Static Atmospheric Electrical Energy into Dynamic Electrical Energ

Science & Invention (June 1928) - "Harnessing Nature's Electricity"

H. Plauson: British Patent # 299735 - Rapidly Moving Electron Process for Producing [PDF]

H. Plauson: *Gewinnung und Verwertung der Atmosphatischen Elecktrizitat* (1922) [PDF]

---

http://en.wikipedia.org/wiki/Hermann_Plauson

Biography

Hermann Plauson was an Estonian engineer and inventor. Plauson investigated the production of energy and power via atmospheric electricity.

Plauson was the director of the Fischer-Tropsch "Otto Traun Research Laboratories" in Hamburg, Germany during the Weimar Republic of the 1920s. Nikola Tesla's idea for connecting machinery to the "wheelwork of nature". Plauson's US Patent # 1,540,998 methods of converting alternating radian electricity into rectified continuous current pulses. He developed the Plauson's converter, an electrostatic generator. In 1920, Plauson published a b *Production and Utilization of the Atmospheric Electricity* " (*Gr., Gewinnung und Verwertung der Atmospharischen Elektrizitat*). A copy of this boo British Library.

It is believed to be Gertrud Plauson (the exact relationship is unknown;

"Power from the Air". *Science and Invention* , Feb. 1922, no. 10. Vol IX, Whole No. 106. New York. (nuenergy.org)
"Power from the Air". *Science and Invention* , March 1922.

***Science and Invention,* Vol. IX (106) # 10 (February 1922)**

Power from the Air (I)

by

Hugo Gernsback

In the war there was developed in Germany in the new art --- or science --- that bids fair to revolutionize our present means of obtaining power.

This art, which is as new now as wireless, is 25 years ago, will be attainments during the next 25 years that may appear fantastic today. Hermann Plauso of the new science, has devoted years of labor to his researches and he has now been using small power plants, generating electricity direct from the a night, without interruption at practically no cost, once the plant is constructed.
We have occasion, to describe the system, roughly, from cabled dispatches, but complete information is available now. The amount of electrical power in our atmosphere is astounding. Herr Plauson found in his experiments a single balloon sent to a height of 300 yards gave a constant current at 400 amperes, or in 24 hours over 17-1 / 4 kilowatts! By using two balloons in connection with a special condenser battery, the power obtained was 81-1 / 2 24 hours. The actual current delivered was 6.8 amperes at 500 volts.

The best balloons used by the inventor are made of thin aluminum leaf. No fabric was used. A simple internal system of ribs, stays and wires, gives the b certain amount of elasticity. The balloon, when made airtight, is filled with hydrogen or better, with helium. It will then stay aloft for weeks at a time surface is made of extremely sharp pins, made sharp electrolytically. Ordinary pins did not test good current collectors, as they lacked extreme sharpnes themselves were made from amalgamated zinc, containing a radium preparation, in order to ionize the air. It was also found that by the outer surface of with zinc-amalgam. Even better results were obtained with polonium amalgam.

One hundred of these captive balloons, separated one hundred yards from each other, will give a steady yield of 200 horsepower. This is the minimum horsepower, due to the higher electrification of the atmosphere.

We need to go into the technic of how the problem is now solved by Herr Plauson. By using batteries of condensers, high tension transformers, etc., the be changed to any form desires. Such as lighting, running motors, charging storage batteries, etc.

Plauson also invented a sort of electrostatic rotary transformer which gives alternating current without the use of condensers and transformers. Indeed, very great, as it is actually sucks the current from the collector balloons. There is no doubt that this invention will become a universal use all over the wor see the land with captive balloons, especially in the country. Indeed, the time is not all that is coming from the atmosphere. It may be the least form of po as the cheapest form of power known today. Not only that, but not as devastating thunder storms occur

---

***Science & Invention* (March 1922), page 1006, 1007**

Power from the Air (II)

by

Hugo Gernsback

[For many years electrical engineers have been able to use the free electrical energy ever present in the atmosphere, but they were not the elevated co endanger the lives of the experimenters, or else destroy the apparatus connected with it. A German engineer has, however, devised the somewhat elabor here shown in the brief, and he has succeeded, at least so his report states, in safely extracting several kilowatts of electrical power from the atmos metallic surfaced balloons, elevated to a height of only 1000 feet.]

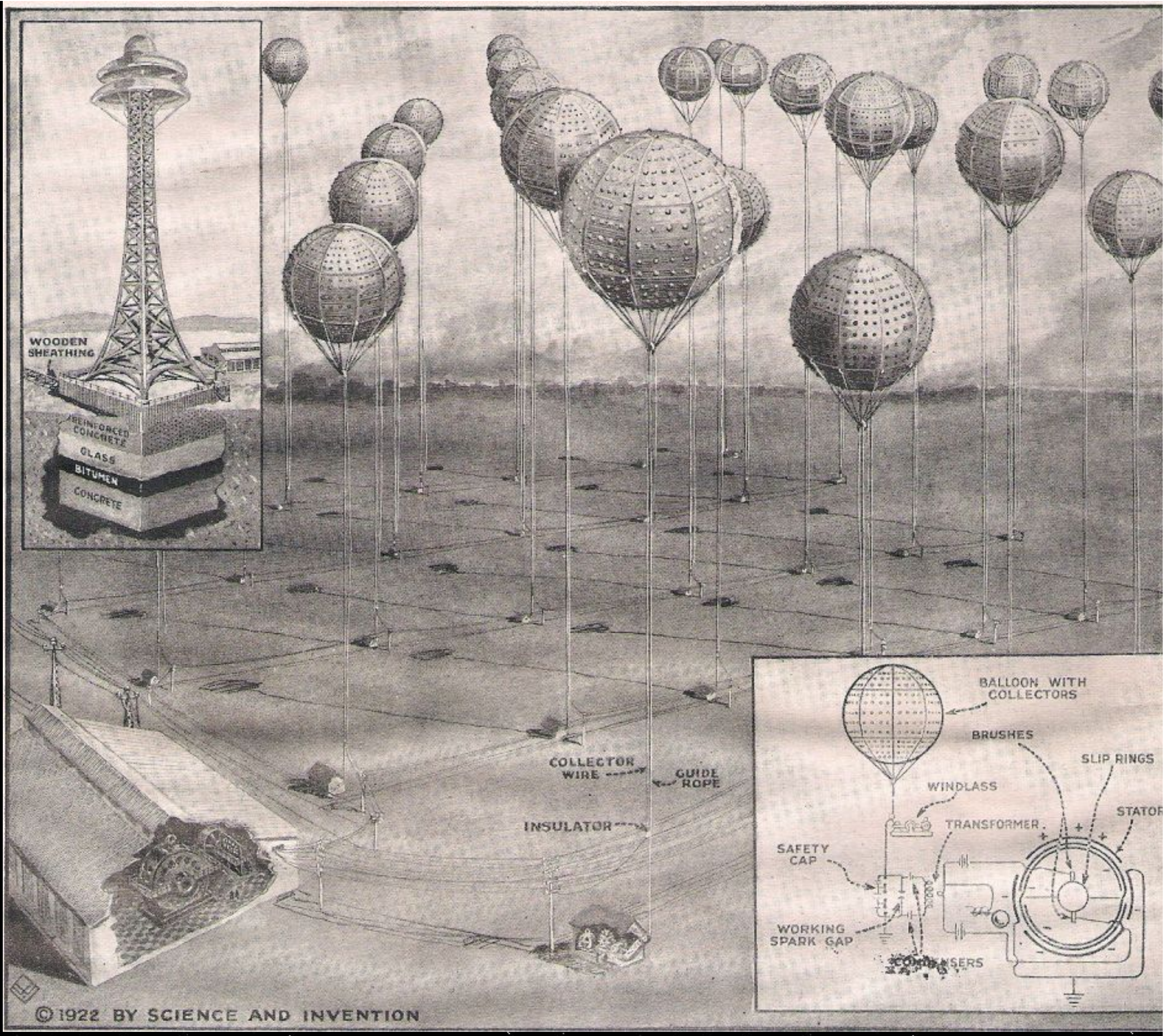

© 1922 BY SCIENCE AND INVENTION

We have been treated with the extraction of electrical energy from the atmosphere. The difference of the electric potential in different parts of the atmo
the difference between the upper air and the earth. The power would take the form of high potential. It has long appeared rather than to conservative e
such a source of power should really be available. It is not the same, but when the lightning flash is flashing, it is likely to be very small. seriously, as th
now being seriously investigated. A German scientist, Hermann Plauson,

We will first speak of the methods used for collecting electricity from the upper air. The author cites several German patents. One of them shows the u
balloon. The balloon is shown floating in the air, kite fashion, and from a great net or aerial for the collection of electricity. The conductor from the ae
the ground station; quite an elaborate description of the net-work which the patentee proposes to have covered with needle points. A windlass takes in
cable for the balloon, and the patentee claims that will send you to 225,000 volts to draw upon. He then speaks or a battery of 20,000 cells in series, wh
up to 40,000 to 50,000 volts in the charging. This certainly provides for a large large fall of potential.

But our author discards this idea and first suggest something more permanent. He proposes the erection of towers to the height of 1,000 feet. At the su
his collecting aerial. The appliance consists of a number of copper tubes; to the top of the tubes, with the aim of collecting net-work covering the tops o
"One of his apprehensions" should be his "trouble trouble", he would like to propose himself to "Siamese pagoda". He also compares the form of the pro
great petticoat insulator. His is insulated from the earth. He, therefore describes a complicated foundation for his structure. A foundation of simple concr
subject, it is a layer of cast glass, three to ten feet thick, and then comes to a reinforced concrete foundation. Sides of the ground, this foundation must
least seven feet The author's idea is based on the number of these towers connected by a horizontal cable. three to ten feet thick, and then to a reinforc
foundation, to which the metallic foot of the tower is to be anchored. Sides of the ground, this foundation must be raised at least seven feet The auth
based on the number of these towers connected by a horizontal cable. three to ten feet thick, and then to a reinforced concrete foundation, to which the m
of the tower is to be anchored. Sides of the ground, this foundation must be raised at least seven feet The author's idea is based on the number of th
connected by a horizontal cable.

The author strongly advocates balloons as collectors of the electric power of the air. These he depicts covered with spots. These spots indicate areas to be
coated and prepared to collect potential from the atmosphere.

In the first place he described the balloon as "made of thin metallic leaf" supported by internal ribs. Steel wires silver-plated, copper-plated, or alumin
run from the balloon to the pendant or junction ring. To this ring the cable is attached to the surface of the earth. 300 feet to three miles.

The coating of the spots is to be of the thinnest amalgam, of mercury and gold, or zinc, or even polonium, perhaps only 1/2500 inch thick. Numberless m
To prepare the needle-like wires, they are collected into bundles and are treated electrolytically in a bath, so as to be dissolved into part. This gives a shar
a roughened surface. The points may be of copper, steel, or some hard metallic alloy. After this corrosion. As it may be termed, the wires are plated w
other of the so-called noble metals. It is advised that polonium or radium salts be added to the plating bath.

Dr Plauson devotes This is a rotary motor including a stator and rotor and its peculiarity is that it contains no coils, develops no electromagnetic fie
speaking, but works by static excitation. One typical arrangement is shown in our illustration. The stator plates and rotor plates are concentric with
representing segments of cylinders. The alternation of negative and positive charged plates produces the rotation. In the connection there is included a s
gap to take care of dangerous potentials. Inductances and capacities are also used and indicated. The plates heated, owing to the Foucalt currents, are
subdividing the stator and rotor plates, are described by the author.

The whole subject is quite captivating, and it really seems as if the use of the air may be in sight. It would seem to be possible to carry out experim
direction by means of the Eiffel Tower.

And now our author gives us some practical details. He says that he is doing it with a radium preparation as an ionizer. The surface of the balloon wa
over with zinc amalgam. It was sent to a height of 300 meters, early 1,000 feet, and was held by a copper-plated steel wire. A constant current of 1.8 am
average of 400 volts potential difference was obtained. This gave nearly three-quarters of a kilowatt, or close to one horsepower. The collector of the ball
to be a tension of 42,000 volts. A passing a second balloon with an antenna. Antenna, and the antenna connection with the 500 volts mean tension. The k
the use of these two balloons.

Science & Invention (June 1928)

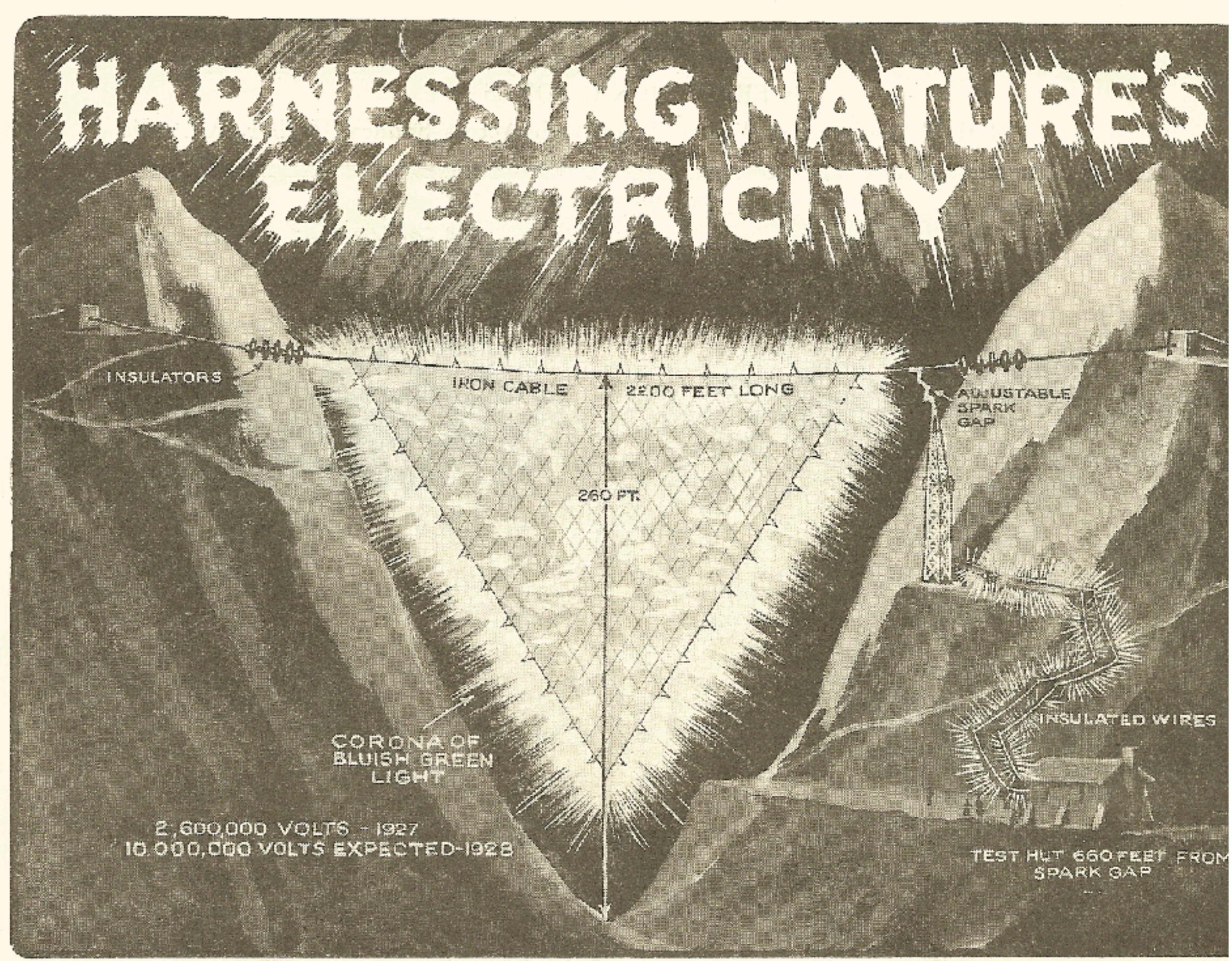
HARNESSING NATURE'S
ELECTRICITY
INSULATORS
IRON CABLE
2200 FEET LONG
ADJUSTABLE SPARK GAP
260 FT.
INSULATED WIRES
CORONA OF BLUISH GREEN LIGHT
2,600,000 VOLTS - 1927
10,000,000 VOLTS EXPECTED-1928
TEST HUT 660 FEET FROM SPARK GAP

Remarkable European Experiments with Atmospheric Electrical Discharges with Potentials as High as 3,000,000 Volts

By HENRY TOWNSEND

elevation of 350 ft., and these students of natural electrical phenomena have found a very desirable location in the Alps, where they can suspend between one mountain and an adjacent one, a strong iron cable having a length of about 2,000 ft. This cable is about 250 feet above the intervening valley, and from it these daring engineers have suspended a coarsely woven wire net, which serves as an electrical capacity to gather the electricity from the atmosphere. As shown in the pictures, the wire net is supplied with numerous sharp points to aid in collecting the current from the air.

As the accompanying photographs of the actual apparatus and wire cable used last year clearly show, an adjustable spark gap of considerable length is provided. By adjusting this spark gap to various lengths, it is possible to judge the voltage of the discharge which leaps the gap at any moment. Mr. F. W. Peek, Jr., the well-known American worker in the realm of high voltage measurements, together with other engineers, have provided tabulated data and curves for various lengths of both needle and sphere type spark gaps. As one of the accompanying diagrams shows, it is a simple matter to calculate the voltage when a certain length of gap is used. The engineer first checks the length of the gap on the chart; he then follows a line horizontally from the gap length, to where it intersects with the angular line on the chart; and from the point of intersection he looks in a visual line downward to a place where the voltage is given. For needle spark gap measurements, the characteristic curve on the chart is practically a straight line, while for sphere gaps the characteristic curve on the voltage versus gap length, is a curved line. Those interested in high voltage measurements by means of the spark gap method can find the voltage-gap tables and charts in the *Standardization Rules* of the American Institute of Electrical Engineers. According to Mr. Peek's researches, the voltage per foot of atmospheric electrical discharges is about 100,000, while in laboratory measurements with A.C., transformer high potential discharges, the average voltage per foot of spark was found to be about 150,000 volts. The voltage of a lightning flash may

(*Continued on page* 156)

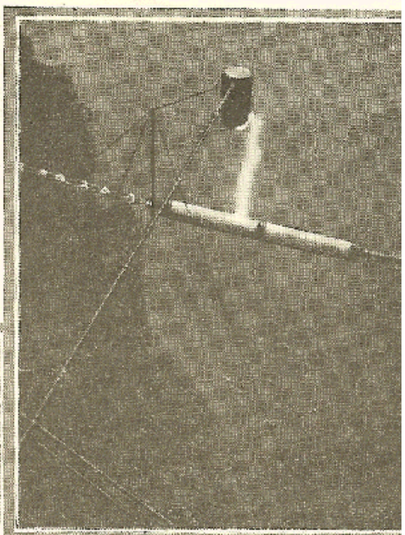

↑ Actual photo at
shows 13 ft., he
spark obtained i
the collecting net
the Alps by the
man scientists.
voltage is about 2,
000. The spark occu
once per second fo
minutes.

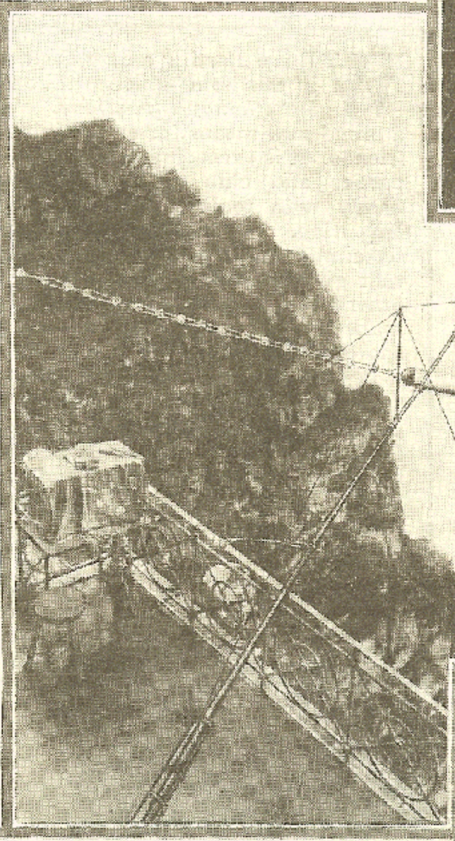

Photo, left, shows
adjustable spark
used in the Alps.
tice the heavy e
trode on the end of
adjustable arm to wh
the spark jumps

←

Actual photograph of the experimental "kite" used by the German experimenters in the Alps Mountains, for the purpose of accumulating high potential electrical discharges from the atmosphere. Note the size of the insulators.

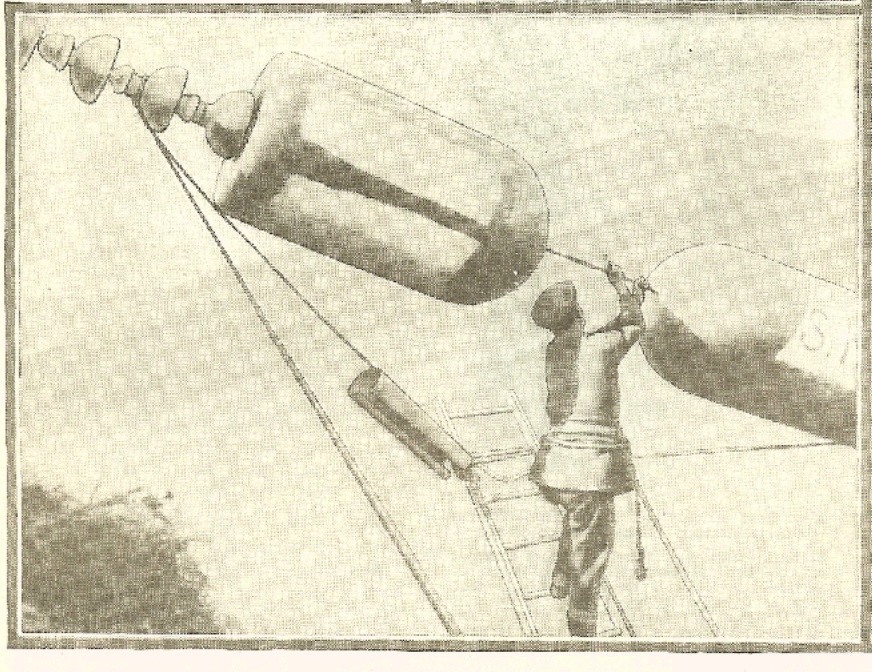

Below we see 3,000
volt artificial light
stroke produced in
E. Laboratory at P
field, Mass. Note n ↓

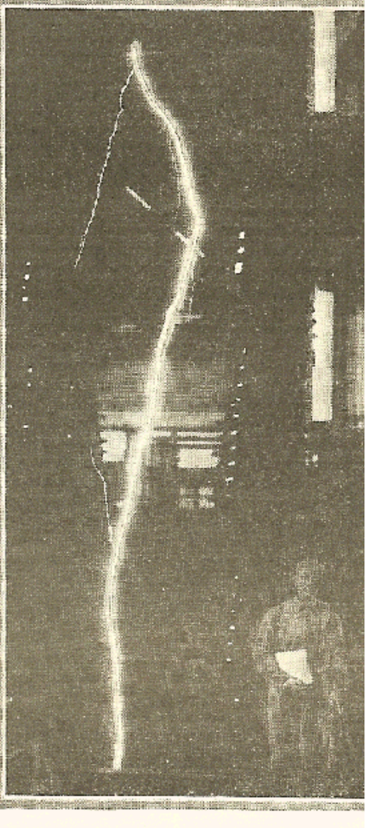

if-
you pledge
yourself to secrecy we will teach you the secrets of
Real Professional
Magic

Dr. Harlan Tarbell, internationally known magician and teacher of some of the world's greatest magicians.

Learn to be a Magician

For the first time in the history of Magic the age-old, sacredly guarded secrets of the Mystic Art are being revealed. Now at last you can learn to be a Real Professional Magician. Yes, and you can learn this wonderful art *easily and quickly* AT HOME. You are taught the mysterious *Principles* by which the Professional Magician works. You are taught everything from wonderful impromptu effects to massive stage illusions. But before you can study, you must sign the Magician's Solemn Pledge of Secrecy. This means that you are getting the *guarded secrets* of the Magic Profession. Think of that!

Earn $250 to $1000 a Month
There's a big demand for Magic entertainment. Magicians get big money. Dr. Tarbell gets as high as $250 for a half hour's work. Unlimited opportunities for you!

Low Prices Easy Terms

Write Today Send today for free Magic book, low prices and easy terms. Don't delay. Mail the coupon NOW!

Tarbell System, Inc., Studio A-142
1920 Sunnyside Avenue, Chicago, Ill.

Gentlemen: Without any obligation send me your free literature and information all about your wonderful "Tarbell" Course in Magic." Also tell me about your Low Price Offer and Easy Payment Plan.

Name..

Address..

....*Age*........

HARNESSING NATURE'S ELECTRICITY

By HENRY TOWNSEND

(*Continued from page* 111)

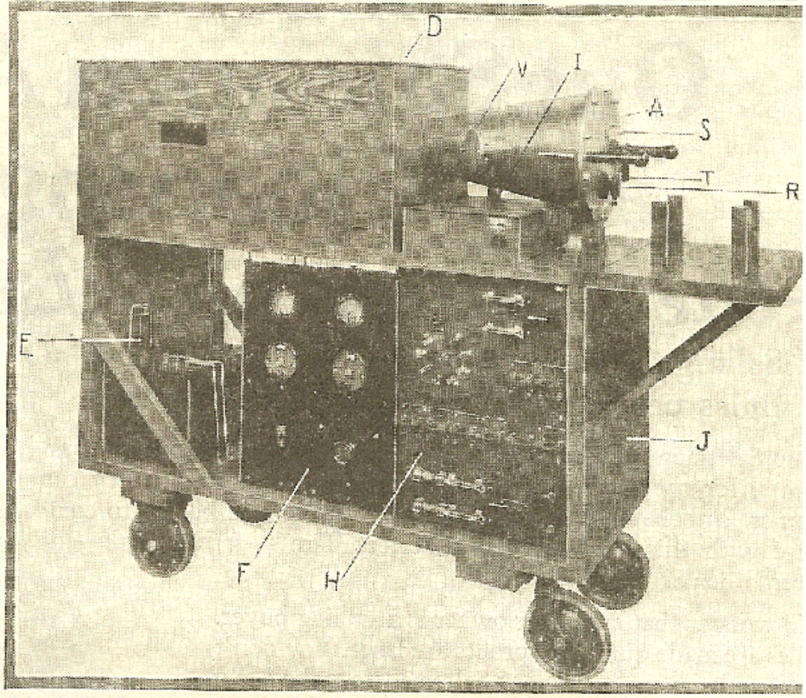

Remarkable cathode ray oscillograph used by Mr. F. W. Peck, Jr., in causing lightning t write its autograph with a pencil of electrons on a photographic film.

easily be 100,000,000 volts, as Mr. Peek has pointed out in one of his scientific papers.

WHY THE EXPERIMENTS ARE BEING MADE

ONE of the main reasons why these dangerous experiments are being carried on by the three young German scientists, whose names we have already learned, is because science believes that with a sufficiently high voltage, it will be possible for man to disintegrate the atom, and in this way make available a tremendous source of power as yet untapped. These experts have calculated that they will be able to obtain electrical energy in sufficient quantity from one of these powerful atmospheric discharges, to equal the Alpha rays obtained from 220 pounds of radium. As we have mentioned before, these experiments are of course fraught with great danger, and for that reason the experimenters seek refuge in a special lightning-proof hut, which is located about 600 ft., from the spark gap. When electrical storms are in the vicinity, it is especially important that the scientists keep within their protected fortress, for otherwise they would very probably be killed by a stray electrical discharge.

One of the peculiar things about this whole line of experiment is that the average layman does not realize perhaps, that there is a high electrical stress in the atmosphere on clear days, as well as when thunder storms are overhead. This fact has been known for a hundred years and more, and many years ago measurements of the various electrical potentials at increasing altitudes, were observed and measured by scientific investigators. There are a number of different ways in which these high electrical potentials found in the atmosphere can be measured; one of these methods involves the use of a calibrated spark gap. In this case the gap is set to a predetermined length, and when a discharge jumps this gap, the engineers know of course from previous experience and measurements, just what voltage is present. Another method of measuring
extra high potentials, such as here
countered, requires the use of a static v
meter, which involves the use of a statio
and of a movable or rotary set of m
plates, forming a condenser, to which
indicating needle is attached. For volt
above 2,000, static voltmeters have been
in a great many American central stati
and they have many desirable and us
characteristics. Of course as the vol
to be measured increases, the space betw
the quadrant shaped stationary and mov
plates is increased and vice versa.
electrostatic field from voltages below 2
is not sufficient to warrant the use o
static voltmeter. Another method of m
uring high potentials involves the use
the so-called vacuum tube voltmeter.

The general characteristics of the
mospheric electrical discharges, inclu
lightning, have been measured and reco
by one of the newest scientific instrum
known as the cathode ray oscillograph.
means of this quite remarkable, high
age scientific apparatus. Mr. Peck, one
the well-known General Electric Comp
staff of research engineers, has made s
very interesting and remarkable discove
concerning the nature of natural elect
discharges, particularly of lightning
charges. Many people will probably wo
why Mr. Peek and some of his collea
in the engineering profession, including
three daring German students, Mes
Brasch, Lange and Urban, play with
dangerous electrical discharges, and
they are at all interested in them. We
plained previously why the German sav
are intent on finding out all they can a
these tremendous voltages obtained
the atmosphere, while Mr. Peck, we
say, also has a very practical reason
carrying on experiments with these de
dealing bolts of Thor. Mr. Peek has
for many years intent on finding out
causes the huge insulators on long dista
high potential transmission lines to b
down when electrical storms break loose
these regions.

Plauson's Electrical Patents

USP # 1,540,998
Conversion of Atmospheric Electric Energy
6-09-1925

GB157262
Improvements in Electric Motors
1922-07-10

GB157263
Process and Apparatus for Converting Static Atmospheric Electrical Energy into Dynamic Electrical Energy of any Suitable High Periodicity
7-10-1922

British Patent # 299,735
Apparatus for Producing Rapidly Moving Electrons
7-15-1930

FI21227
Elektrisk uppvärmningsanordning
4-25-1946

Varmelegeme med elektriske varmemodstande
DK67691C
9-27-1948

FR877362
Dispositif de chauffage électrique
12-04-1942

DE734794
Elektrisches Heizsystem
4-24-1943

CH222509
Elektrischer Heizkörper zur Erwärmung von Flüssigkeiten
7-31-1942

DE738107
Elektrolyt fuer unmittelbare elektrische Warmwasser-Radiatorenheizung mit Elektroden
8-03-1943

DE433476
Verfahren zur Herstellung von Elektroden und Schleifkontakten fuer Dynamomaschinen
8-31-1926

CH94021
Elektrode und Verfahren zu deren Herstellung
4-01-1922

CA226423
Electrode for Electrolytic Apparatuses
11-21-1922

---

http://www.meridian-int-res.com/Energy/Atmospheric.htm

Atmospheric Electricity Research
[Excerpts]

In the nineteenth and early twentieth centuries, a large number of investigators

Dr Hermann Plauson who in the 1920s succeeded in generating significant quantities of electrical power with modern solar photovoltaic systems of a s
...

Dr Hermann Plauson. Dr Plauson was an Estonian citizen who lived in Hamburg and Switzerland. He carried out experiments in Finland with ma magnesium-aluminum alloys, covered with electrolytically deposited needles. The needles were further doped with a radium compound to increase the i the air. (This was the era in which the hands were painted with radium to make them luminous in the dark). Zinc amalgam patches were also paint aerostats. Plauson obtained with power output of 0.72kW and 3.4kW from one and two aerostats 300m above ground level. Dr Plauson filed patents i Great Britain and Germany in the 1920s. His book "Gewinnung und Verwertung der Atmosphärischen Elektrizität"

Other atmospheric electricity researchers contemporary to Dr Plauson included Walter Pennock and MW Dewey in the USA, Andor Palencsar in Hung Heinrich Rudolph in Germany. Hippolyte Charles Vion in Paris predated them all, putting forward proposals in the 1850s and 1860s.

Heinrich Rudolph made an interesting contribution to the design of the aerostat collectors. In 1898 he designed an elliptical aerostat made up of faceted minimize the effect of wind. The design bears a strong resemblance to Northrop's 2003 UCARS unamnned helicopter UAV project. The design uses Effect to help keep the aerostat on station and minimise wind effects.

Dr Oleg Jefimenko: In recent times, Dr Oleg Jefimenko has been active in this field. Dr Jefimenko carried out experiments on electrostatic motors from the 1970s.

MIR's Research Program

Since 1997 we have been carrying out theoretical research into conversion of atmospheric electricity into useable electrical power.

From a low level (5m high) simple zinc antenna we are able to obtain sufficient charge. Further experimental investigations with metallic aerostat col cavity resonant slow-wave antennae concepts are ongoing ...

Advantages of Atmospheric Electricity

Simple and robust technology
Low Cost Technology - much cheaper than photovoltaics or wind turbines
Available day and night in all weather conditions - in fact, blackberries power is produced at night than during the day
Available at any point on the Earth's surface

1. *Gewinnung und Verwertung der Atmosphärischen Elektrizität* , Dr Hermann Plauson, Hamburg, (1920)
2. Conversion of Atmospheric Electric Energy, USP 1,540,998; Dr Hermann Plauson, (1925)
3. Assembly for the Induction of Lightning into a Superconducting Magnetic Energy Storage System, USP 5,367,245 Goven Mims, (1994)
4. Electrostatic Motors are Powered by Electric Field of the Earth; CL Stong, *Scientific American* , (October 1974)
5. Operation of Electric Motors from the Atmospheric Electric Field; Dr. Oleg Jefimenko, *American Journal of Physics* , vol. 39, July 1971.
6. *Electrostatic Motors: Their Principles, Types and Theory of Operation*; Dr Oleg Jefimenko, Electret Scientific, (1972).
7. Parametric Electric Machine, USP 4,622,510, Ferdinand Cap, (1986).

US Patent # 1,540,998

Conversion of Atmospheric Electric Energy

(9 June 1925)

Hermann PLAUSON

Be it known that I, Hermann Plauson, Estonian subject, residing in Hamburg, Germany, have been invented in the Conversion of Atmospheric Electric Ene following is a specification,

A method of obtaining atmospheric electricity by means of metallic nettings set with spikes which are held by means of ordinary or anchored kite balloons m filled with hydrogen, are in theory already known. Direct current for the charging of accumulators. This knowledge is only presented as theoretical as the conv has hitherto been a failure. Apparatus of destruction by lightning. The balloons used for collecting the charge must also be made of very large size.

Non-conducting materials of non-conducting materials that are liable to be used as non-conducting materials

(a) The metallic cases are impenetrable to helium and hydrogen; they also represent large metallic weather-proof collecting surfaces.

(b) Radio active means the like may be easily applied internally or externally; whereby the ionization is considerable increased and there is also the quanti electricity capable of being collected.

(c) These balloons have been produced by light weight.

(d) The entire system offers a small surface for the action of storm and wind.

(e) Each balloon can be raised and lowered by means of a winch so that all repairs can be carried out without danger during the operation.

Collaterals are interconnected by electrical conductors.

This article was previously published underneath this article. strength is converted into electro-dynamic energy into the form of high frequency vibrations. Ma make the purchase and all disadvantages avoided.

This article is based on the current law. It is possible to obtain electromagnetic waves of various amplitude and to increase the degree of resonance of suc resonance can be used to provide the possibility to be chosen again by the starting and stopping of the machine by simply transforming the resonance betwee machine and the transformer forming the resonance. Further, such currents have the property of being directly available for various uses, even without em driving motors, of which there may be

Further, with such currents a series of apparatus may be fed without direct current supply through conductors and also the electro-magnetic high frequency converted by means of special motors adapted for electro-magnetic oscillations into mechanical energy, or finally converted by special machines into alternatin

The invention is more particularly with reference to the accompanying diagrams in which:

Figure 1 is and explanatory figures.

Figure 2 is a diagrammatic view of the simplest form.

Figure 3 shows a method of converting

Figure 4 is a diagram showing the use of protective means.

Figure 5 is a diagram of an arrangement for converting large current strengths.

Figure 6 is a diagram of an arrangement including controlling means.

Figure 7 shows means whereby the spark gap can be adjusted.

Figure 8 shows a unipolar connection for the motor.

Figure 9 shows a weak coupled system suitable for use with small power motors.

Figures 10, 11, and 12 show modified arrangements.

Figure 13 shows a form of inductive coupling for the motor circuit.

Figure 14 is a modified form of Figure 13 with inductive coupling.

Figure 15 is an arrangement with a non-inductive motor.

Figure 16 is an arrangement with coupling by condenser.

Figures 17, 18, and 19 are diagrams of further modifications.

Figure 20 shows a simple form in which the serial network is combined with special collectors.

Figure 21 shows diagrammatically an arrangement suitable for collecting large quantities of energy.

Figure 22 is a modified arrangement with two rings of collectors.

Figure 23 shows the connection for three rings of collectors.

Figure 24 shows a collecting balloon and a diagram of its connection of condenser batteries.

Figures 25 and 26 show modified collector balloon arrangements.

Figure 27 shows a second method of connecting conductor for the balloon aerials.

Figure 28 shows an auto-transformer method of connection.

Figure 29 shows the simplest form of construction with incandescent cathode.

Figure 30 shows a form with a cigar shaped balloon.

Figure 31 is a modified arrangement.

Figure 32 shows a form with cathode and electrode enclosed in a vacuum chamber.

Figure 33 is a modified form of Figures 32.

Figure 34 shows an arc light collector.

Figure 35 shows such an arrangement for alternating current.

Figure 36 shows an incandescent collector with Nernst lamp.

Figure 37 shows a form with a gas flame.

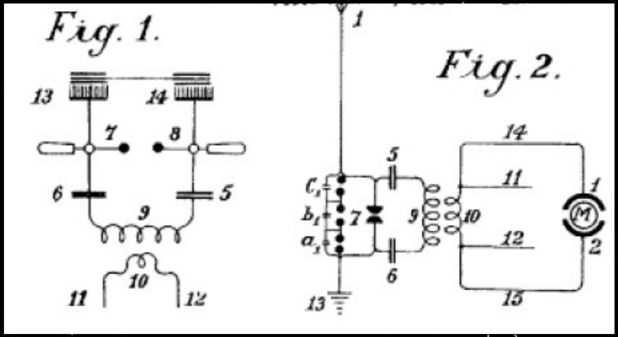

Figure 1 illustrates a simple diagram for converting static electricity into a dynamic power of a high number of oscillations. An illustration of an influence ma of clarity. 13 and 14 are combs for collecting the static electricity. 7 and 8 are spark discharging electrodes. 10 and 8 inductive coil, 10 secondary coils, 1 conductors of the secondary coil 10. When the disc of the static influence is rotated by mechanical means, the combs collect the electric charges one the and and 6 until the potential is formed across the spark gap 7-8, that the spark gap is jumped. As the spark gap 7-8 forms a closed circuit with condensers 5 and 6,

The high frequency of the oscillations produced in the primary circuit. Thus in the primary circuit electromagnetic oscillations are formed by the passage of the

The ratio between the number of the coils in the primary and the secondary circuits with the correct use of the co-efficients of the resonance (especiall resistance) voltage and high current strength.

The liquid gap until the accumulated charge again breaks the spark gap. This is how it is produced by the static machine employing mechanical energy.

Figure 2 in which two spark gaps in parallel are shown in Figure 2, while the second one is used as a safety device for excess voltage and consists of a larger in series and are bridged by very small capacities as shown in a, b, c, Figure 2

In Figure 2, it is an aerial antenna for collecting charges of atmospheric electricity. 13 is the earth connection of the second part of the spark gap, 5 and 6 are primary coil. Now when you go through the aerial The spark gaps. The resistance of the spark gap is shown in the drawings, lower than that of the other safe consists of three spark gaps connected in series, and consequently to three times.

So long, therefore, as the resistance of the spark gap, I know that the other spark gaps have an equal resistance with the discharges take place it could be dange the condensers 5 and 6 for the coil insulation 9 and 10 in consequence of the break down, by a correct regulation of this spark gap .

Without this second spark gap, it is impossible to collect and make available large quantities of electrical energy.

; Capacitor 5 and 6, primary coil 9, and also secondary coil 10 is the same as the one described in Figure 1 with the arrangement of the static induction mach difference that here the second spark gap is provided. The electromagnetic high frequency alternating current can be tapped off from the conductors 11 and 1 heating purposes. Special kinds of motors with special features such as electrical systems may be connected to 14 and 15 which can work with static e frequency oscillations.

In addition to the use of spark gaps in parallel a second measure of security is also necessary for taking off the current. This is a precautionary measure based in the introduction of a method of connecting certain protective electromagnets or choking coils in the aerial circuit as shown by S in Figure 3.

A single electromagnet only having a core of the thinnest possible separate laminations is connected with the aerial.

In the case of high frequencies in the aerial network or at places where there are frequent thunder storms, several such magnets may be connected in series.

In the case of parallel or parallel series.

The windings of these electromagnets may be connected in series with the aerials. In this case the winding preferably consists of several thin parallel wires together, the necessary section.

The winding may be made of primary and secondary windings in the form of a transformer. The primary windings will be connected in series with the aerial secondary winding is more or less short-circuited over a regulating resistance or an induction coil. It is possible to regulate to a certain extent the effect of the c aerial electromagnet choke coil is indicated by a simple ring

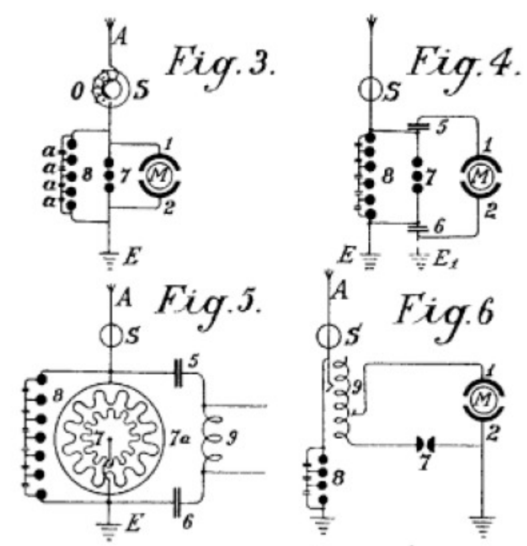

Figure 3 shows the simplest way of converting atmospheric electricity into electromagnetic waves. Recent improvements in motorsport with motorsports, work States

Diagrammatically indicated by the two semicircles 1 and 2 and the rotor of the motor by a ring M (Figure 3). A is a vertical aerial or aerial network. The electr electromagnet and the electromagnet with the spark gap, the circuit 8 with the spark gap , and then a circuit including the stator terminal 1, the rotor and st which the connection is made to the earth wire. The two spark gaps are also connected metallically with the earth wire. These methods are as follows:

The positive atmospheric electric charge collected with the earth. It goes through the electromagnet. Further, its progress is arrested by two sparks gaps place the stator condenser surfaces. The stator condenser surfaces are charged until the charge is higher than the resistance of the spark gap 7, whereupon a spark spark gap 7, and an oscillatory charge is obtained from the motor M, stator surfaces 1 and 2, and spark gap 7, a closed oscillation circuit is obtained electromagnetic oscillations. The motor here forms the capacity and the necessary inductance and resistance, which, as is well known,

By the electromagnet or choke. A light source, a periodic difference in the electromagnet, a periodic difference between the two, and is. This is the case of the which is sufficiently high, and is supplied to the flow of atmospheric electricity.

With the wave of the motor, it does not end with the natural frequency of the motor, it does not endanger the motor through the spark gap.

In the diagram illustrated in Figure 4 the spark gap is shunted across condensers 5 and 6 from the motor.

In Figure 5 a diagram is illustrated for large current strengths which may be direct direct without motors, for example, for lighting or heating purposes. The r that the spark gap consists of 7 When separated from one another, discharges take place, thus forming an oscillation circuit over condensers 5 and 6, and oscillatory discharges. It is evident that a motor may be connected to the ends of the spiral 9.

The construction of the diagram is shown in Figure 6 of the oscillating circuit of the motor being connected with an induction coil. Motorcycles can be more or coiled 9 (coupled inductively to the aerial).

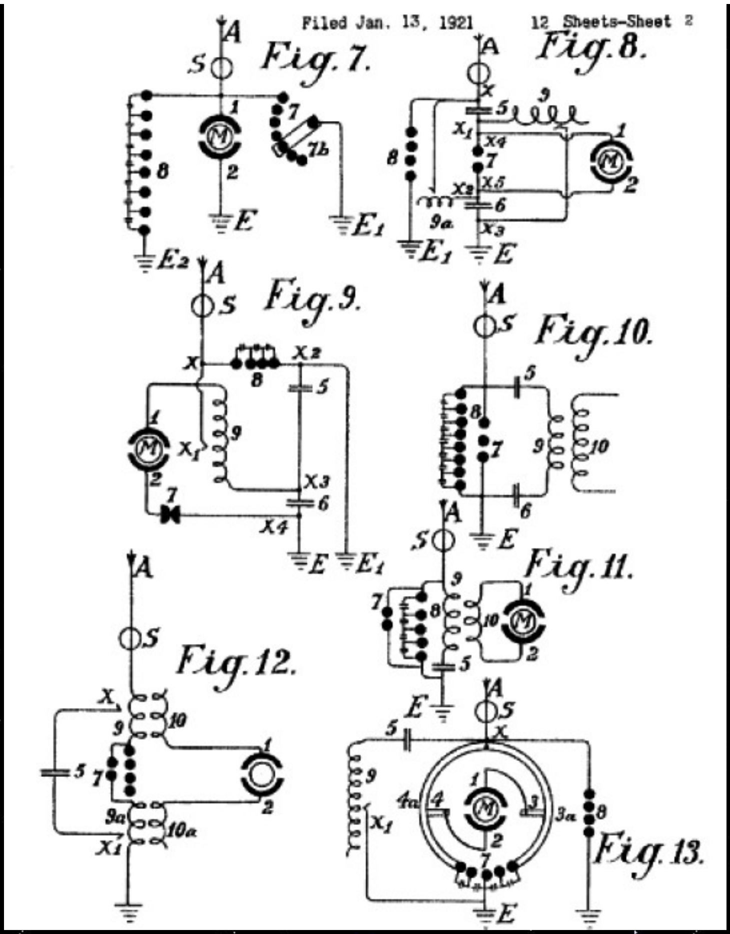

In Figure 7 the oscillation circuit is closed through the earth (E and E1). Spark gap 7 may be prolonged or shortened by more or fewer spark gaps being succe by means of a contact arm 7b.

Diagram 8 shows a unipolar connection of the motor with the aerial network. Here two oscillation circuits are closed through the same motor. Electroma inductance 9a to the earth condenser 6 and further, over spark gap 7 to the aerial condenser 5 and back to *x* . The second oscillation circuit from the aerial c point x^{1} over the inductance 9 to the earth condenser 6 at the point x^{3} and through the condenser 6 over the spark gap 7 back to x^{1}. The motor itself is ins two points of the spark gap 7. From this arrangement, slightly damped oscillation wave currents are produced.

In the diagram illustrated in Figure 9, a loosely coupled system of connections is shown as being for small motors for measuring purposes. A the aerial electromagnet in the aerial conductor, 9 the inductance, 7 the spark gap, 5 and 6 condensers, E the earth, M the motor, and 1 and 2 stator connections of the mo directly metallically connected with the oscillation circuit.

In Figure 10 a purely inductive coupling is employed for the motor circuit. The motor is connected with the secondary wire 10 as may be seen in Figure modified diagram connection. Figure 12.

The diagrams are hitherto described as preferable of motors of small and medium strength to be operated. For large quantities of energy is difficult; swit greater.

A means of overcoming such difficulties is shown in Figure 13. The oscillation circuit starts from the point x over condenser 5, variable inductance 9, spark g segments (3rd and 4th) forming arms of a Wheatstone bridge, electromagnetic oscillations of equal and electromagnetic oscillations are shown in the c approximate. If the the brushes 3 and 4 are moved in common with the conducting wires 1 and 2 which connect the brushes with the stator poles to ce displacement of the polarity is obtained and the motor commences to revolve.

The maximum action will result if one brush 3 comes on the central sparking contact 7 and the other brush 4 on the part *x* . Gaps with the motor oscillation circ

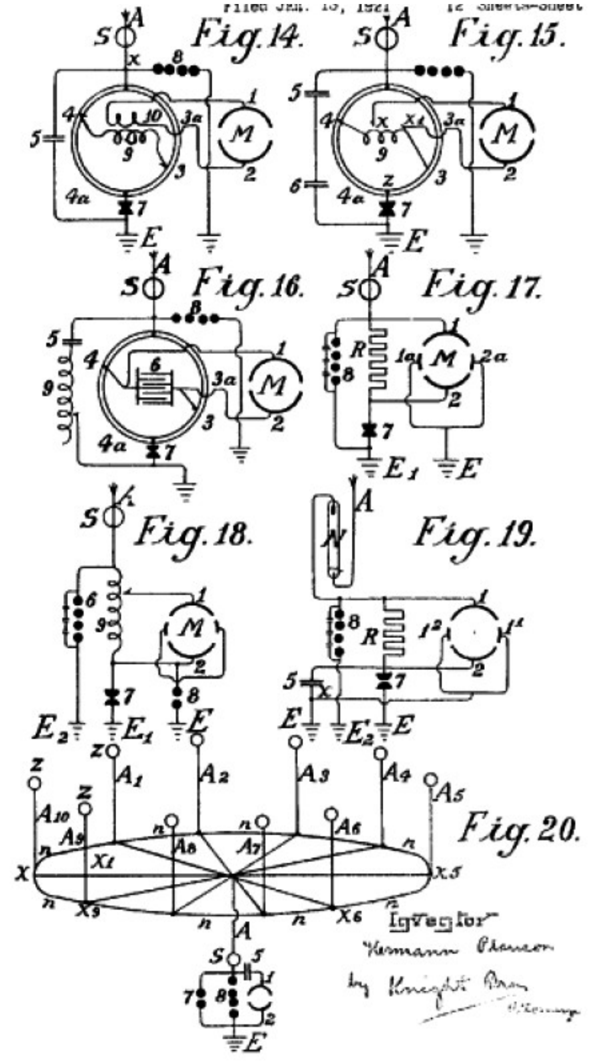

14. The diagram 14 differs from the foregoing only by the motor not being directly related to the motor commutator, but only a primary coil 9 which induces i
10, current which feeds the motor M and takes the place of the rotor. A loose coupling and an oscillation circuit without a spark gap.

In Figure 15 the motor is not purely inductive as in 14, but directly metallically branched off from the primary coil (at x and x^1) after the principle of the auto-

In Figure 16 instead of an inductance condenser 6 is similarly, and for the same object inserted between the segments 3a and 4a. This has the advantage that t
need to be made of solid metal but may consist of spiral coils whereby a more exact regulation is possible and further motors of high inductance may be employ

The arrangements of Figures 17, 18 and 19 may be employed for use with resonance and particularly with induction condenser motors; between the larg
condenser surfaces, small reversing pole condenser surfaces, mall reversing pole condensers are connected, which can be seen from Figures 17, 18 and 19 a
earth. Such reversing poles

Figure 19 shows a method of electromagnetic oscillations of the high number of alternations. It is based on the well-known principle of a mercury lamp, one e
is formed of mercury, the other of solid metal. . The mercury electrode of the vacuum tube. From this it can be only from the aerial through the vacuum tube
circuit, but not vice versa. Oscillations which are formed on being transformed into an oscillating circuit.

In practice these vacuum tubes must be connected behind the danger of lightning.

As regards the use of spark gaps, can be used for wireless telegraphy may be used. Of course the spark gaps in large machines must have sufficiently large
cooled in liquid carbonic acid or better still in liquid nitrogen or hydrogen; in most cases the cooling of the metal series by the means of hydrocarbons the
which lies between -90 ° C and -40 ° C. The spark gap casing must also be insulated and may be strength to be able to resist any pressure which may arise.
excess super-pressure which may be formed must be automatically let off. Mercury electrodes which were frozen in liquid carbonic acid,

Figure 20 is one of the simplest forms of construction of an aerial network in combination with collectors, transformers and the like illustrated diagrammatic
earth wire, 8 the spark spark gap, 7 the working spark gap, 1 and 2 the stator surfaces of the motor, 5 to the condenser battery, the protective magnet which is
coil in aerial conductor, A^1 to ^{10} aerial antennae with collecting balloons, horizontal collecting or connecting wire from which, to the center to number of conn

The actual collectors consist of metal sheaths, preferably made of aluminum magnesium alloy, and are filled with hydrogen or helium and are attached to c
wires. Is the actual weight of the balloon and the weight of the conducting wire. On top of the balloon aluminum spikes, made and gilded in a special m
described, are arranged in order to produce a conductor action. Small quantities of radium preparations, more particularly polonium-ionium or meso-thor
considerably increase the ionization, and therewith the action of these collectors.

In addition to metal balloons, fabric balloons which are superficially metal coated according to Schoop's metal spraying process, may also be employed. Th
may also be produced by lacquering with metallic bronze powders in two electrical series of widely different metals.

Instead of the ordinary round balloons, they can be used. Non-conducting substances, which produces electricity by the wind, also produces electricity collecting effect is substantially increased, the wind will impart a portion of its energy in the form of frictional electricity.

In practice however, very high towers may be used as an antenna. In these towers, copper tubes freely rise above the top of the tower. Lamp to form a col conveyed through the interior of the tube up to the summit. The copper tube must be protected from the effects of the tower and the rain must be prevented by walls of the tower. This is done by a series of enlargements of the Siamese pagodas.

Special attention must be devoted to the foundations of such towers. They can be well insulated from the ground, which may be obtained by first embedding a in a box. . There is an iron-concrete layer in which the metal foot of the tube is secured. This concrete block must be at least 2 meters from the ground. In the tower a wood or glass house for the large condenser batteries or for the motors may be constructed. In order to lead the earth to a well-insulated pit constructed must be provided. These towers are erected at equal distances. To be used on high voltage conductors. Can be taken at any suitable places.

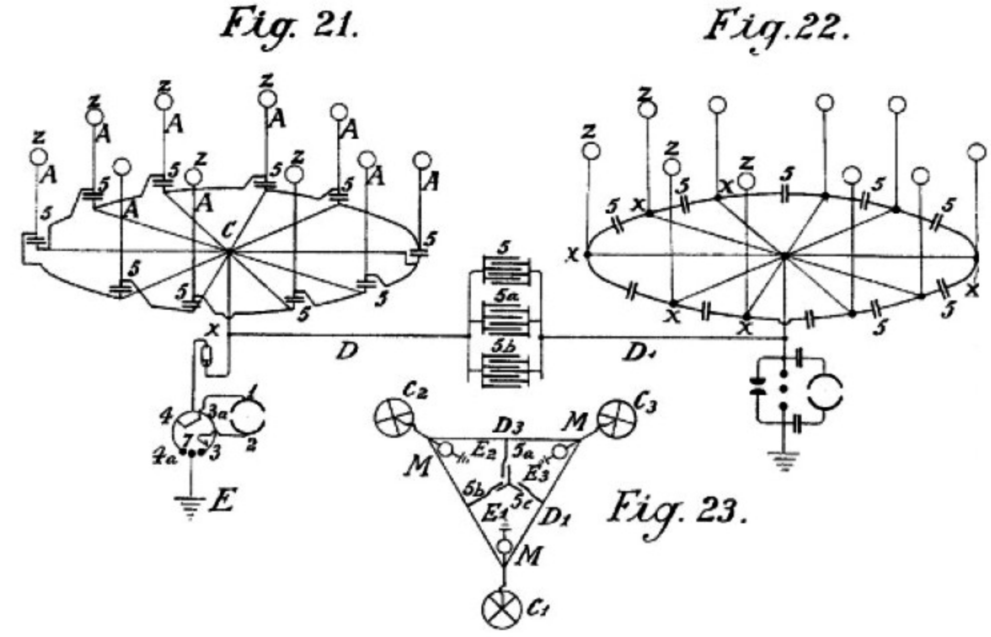
Fig. 21.

Fig. 22.

Fig. 23.

In order to collect a large number of units of electricity, please refer to Figures 21 and 22. In Figure 21 the batteries of condensers Z by the aerial conductor, t the other conductor.

Figure 22 shows a similar arrangement, or two different series of antennae rings for sufficiently large capacitor batteries (5, 5a, 5b) by means of Maji star cond In Figure 23 a connection of three such rings with a central condenser battery is illustrated.

The condenser batteries of such large installations must be embedded in liquid gases or liquids freezing at very low temperatures. In these cases a portion o energy must be used for liquefying these gases. It is also preferable to employ pressure. This is the reason why this is not the case. For smaller installations the condensers in well-insulated oil or the like, suffices. Solid substances on the other hand can not be used as insulators.

The arrangement in the diagrams was shown to be such that the condenser batteries were connected with both poles directly to the aerial condensers. An imp the connections for the supply of electricity to the condenser batteries has been found to be very advantageous. This arrangement consists of one pole (unipolar network. Such a method of arrangement is very important, as a means of being a constant current and an increase of the normal working pressure or voltage i have a collecting balloon, which is allowed to rise to a height of 300 meters, shows 40,000 volts above earth voltage, in practice it is only about 400 volts. Bal of the condenser, the capacity of the condenser surfaces up to and beyond 500 volts. This can be ascribed to the favorable action of the connecting method. t aircraft in the air with the exception of 500 volts. . This can be ascribed to the favorable action of the connecting method. the capacity of the aircraft in the air of 500 volts. . This can be ascribed to the favorable action of the connecting method.

In addition to this substantial improvement, it has also been found to be a double inductance with electromagnets. It can also be used as an induction coil or as or as a condenser itself as an induction condenser. This is a condenser that can be applied to the accumulated force. Condenser pole, and if the spark gap is accumulated energy is again given back to the free condenser pole. The same as the collector network. Increased energy. Inserted in the same space of time.

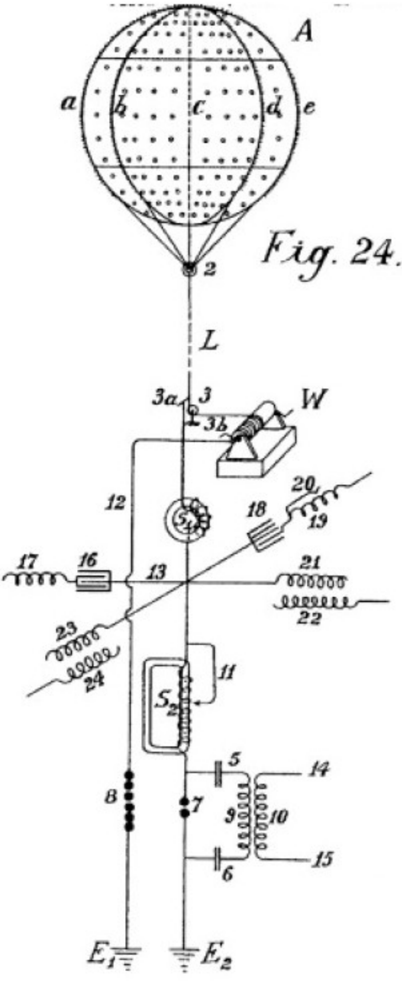

Fig. 24.

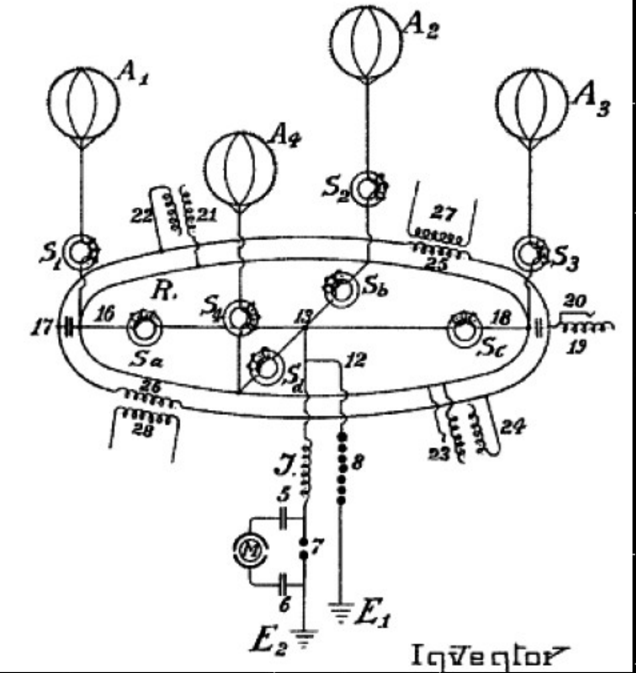

Fig. 25.

In Figures 24 and 25 two different diagrams of connections are shown, Figure 24 shows a collecting balloon and the diagram of the connections to earth. Figu collecting balloons and the parallel connection of the condenser batteries belonging thereto.

A is the collecting balloon made of an aluminum magnesium alloy (electron metal, magnalium) of a specific gravity of 1.8 and a thickness of plate 0.1 to 0.2 are eight strong vertical ribs of T-shaped section about 10 to 20 mm in height and about 3 mm in thickness with the projecting part directed inwards (indicated so forth); They are riveted together to form a skeleton and are stiffened in a horizontal direction by two cross ribs. Internally and transversely by means o whereby the balloon obtains great power of resistance and elasticity. Rolled plates of 0.1 to 0. 2 mm in thickness made of magnalium alloy are then either sold this skeleton coppered steel hawser L twisted out of separate thin wires (shown in dotted lines in Figure 24) and which should be long enough to allow the bal desired height, to metal roller or pulley 3 and from thence to a winch W, well insulated from the earth. The means of this winch, the balloon, which is filled helium, and brought to the ground for recharging or repairs. the coppered steel hawser The twisted out of separate thin wires (shown in dotted lines in Figu should be long enough to allow the balloon to rise in the desired height, to metal roller or pulley 3 and from thence to a winch W, well insulated from the ea this winch, the balloon, which is filled with hydrogen, or helium, and brought to the ground for recharging or repairs. the coppered steel hawser The twisted o wires (shown in dotted lines in Figure 24) and which should be long enough to allow the balloon to rise in the desired height, to metal roller or pulley 3 and winch W, well insulated from the earth. The means of this winch, the balloon, which is filled with hydrogen, or helium, and brought to the ground for rechargin

The actual current is taken directly through a friction contact from the metal 3 or from the wire, or even from the winch or from the three by means of brush

Beyond the brushes the conductor is divided, the paths being: --- firstly over 12 to the safety spark gap 8, from the earth to the conductor E^1, and secondly o S^1, point 13, to a second loose electromagnet having an adjustable coil S^2, then to the spark gap 7 and to the second earth conductor E^2. The actual working through the spark gap 7, condensers 5 and 6, and through the primary coil 9; here the static electricity formed by oscillatory discharges is accumulated and co

frequency electromagnetic oscillations. Between the electromagnets S^1 and S^2 at the crossing point 13, four condenser batteries are introduced diagrammatically in the drawings each by one condenser. Two of these batteries (16 and 18) are induction condensers and prolonged by regulating induction and 19 while the two others (21 and 23) are induction condensers. The condenser batteries 16, 18, 21 and 23 are connected by only one pole to the aerial conductor. The second poles 17, 19, 22 and 24 are open. In the case of plate condensers having no inductive resistance and an induction coil is inserted. Th spiral or coil is the displacement of the current phase by 1/4 periods, while that of the current charge of the condenser poles which lie free in the air, works bac aerial. The consequence of this is that in the aerial collectibles. It has also been found in The length of the coil in the size of the induction coil. It has also b length of the coil in the size of the induction coil. It has also been found in The length of the coil in the size of the induction coil.

S^1 and S^2 may be provided with such regulating devices in the case of S^2 (illustrated by 11). 12 and spark gap 8 or through any other suitable apparatus, si would be dangerous for the other apparatus.

The small circles on the collector balloon indicate places at which zinc amalgam or gold amalgam or other photoelectric acting metals in the form of small pat thin layers (0.01 to 0.05 mm in thickness) are applied to the entire balloon as well as in greater thickness to the conducting network. The capacity of the c increase strengthened at the surface. Polonium amalgams and the like. On the surface of the collector balloon metal points or spikes are also fixed along the serve for collecting the collector charge. For this purpose it is very important to employ as sharp spikes as possible. Spikes made of bars or rollers with a large spikes made of bars or rollers with rough surfaces. Here you are experimented with the collector balloons hereinbefore mentions. The best results were give were made in the following way. End points made of steel, copper, nickel, or copper and nickel alloys, were fastened together in bundles and then placed points in a suitable electrolyte (preferably in hydrochloric acid or iron) at 2 to 3 volts pressure. Spikes or pins the points become extremely sharp and the bo have a rough surface. The bundle can be washed off with water. Platinum, iridium, palladium or wolfram salts or their compounds and coated with a cathode w precious metal, which must be sufficiently firm to protect them from atmospheric oxidation.

Such spikes act at a 20-fold lower voltage almost as well as the best and finest points made by mechanical means. Polonium or radium salts are added to the g forming and protective layer or coating. These pins have a low resistance at their points.

In Figure 24 the three unconnected poles are not connected with one another in parallel. That is quite possible in practice without altering the principle of the f preferable to interconnect in parallel to a common collector network, a series of collecting aerials.

Figure 25 shows a diagram for such an installation. A^1, A^2, A^3, A^4 are four metal collector balloons with gold or platinum coated spikes which are electr the presence of polonium emanations or radium salts, which spikes or needles are connected over four electromagnets S^1, S^2, S^3, S^4, through an annular c this annular conductor four wires run over four further electromagnets Sa, Sb, Sc, Sd, to the connecting point 13. the safety spark gap 8 to the earth at E^1, inductive resistance J and the working spark gap 7 to the earth at E^2. The working circuit, consisting of the condenser 5 and 6 and a resonance motor herebefore described, is connected in proximity with the sparking gap section 7.

Of the primary circuit for high frequency oscillatory current may also be connected.

The condenser batteries are connected by one and the inductionless (16 and 18) or made as induction condensers as shown by 21 and 23. The free poles condensers are indicated by 17 and 19, those of the induction condensers by 22 and 24. May be interconnected in parallel through a second annular conductor that the principle of the free pole connection will be injured. In addition to the advantages already set up the parallel connection also provides an equalizati pressure. Suitable constructed and calculated induction coils 25 and 26 may also be included in the annular conductor of the free poles,

According to what has been hereinbefore stated separate collector balloons may be connected to equidistant stations distributed over the entire country, or c with one other metallically or by means of intermediate suitably connected capacitor batteries through high voltage conductors insulated from earth. This i following criteria: The wires leading from the collector balloons have hitherto been connected through an annular conductor without this endless connection, b any action on the whole conductor system.

A condenser, or short circuited in the form of coils over a condenser battery or spark gap or through thermionic tubes or valves or audions, then the total c exhibits quite new properties. The collection of atmospheric electricity is not only available in the collector network. Further, the atmospheric electrica themselves in the higher regions. In Figures 26 and 28 a form of construction is shown on the basis of which the further foundations of the method will be parti

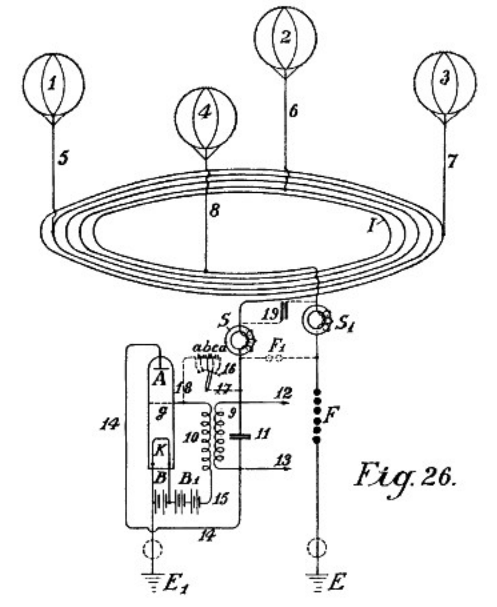

Fig. 26.

In Figure 26, 1, 2, 3, 4 are metal collector balloons, 5, 6, 7, 8 their metallic aerial conductors and the actual collector network. High voltage masts in high v coil has a diameter of 1 or 100 km or more. S and S^1 are two protective electromagnets, F the second safety section against excess voltage, and its earth con earth conductor of the working section. Anion of the electromagnet S, primary induction coil 9, conductor 14, anode tube, incandescent cathode K, as electromagnet and the safety spark gap F offers considerably greater resistance. An electromagnetic alternating field is produced in the interior of the colle whereby the whole free electrons are directed into the interior of the coil. An increased ionization of the atmosphere. In addition to this, they will be show balloon show a reduced reduced resistance and therefore increased static charges between the points on the balloon and the surrounding atmosphere are pro this is a significantly increased collector effect.

A second effect, which can not be obtained from the electromagnetic alternating field, which runs parallel to the earth surface, whereby in the case of fluctuati a return induction current of reversed signal is always produced in the collector coil by earth magnetism. Now if the current periodicity is also produced network coil. As the same alternating field is further transmitted to the aerial balloon, while the collector action is considerably increased. A further advantage i of potential of the collector area. The field of collectable surfaces, when the collateral surfaces occur, when the collectable surface is produced by the law induction of reversed sign on the collector surface and so on charge). In addition to the advantages hereinbefore set, the construction of connecting conductor when sufficiently large, also in the simplest way. Sturmo's fire or northern lights. These energy quantities have not been available to be used up to now. By these kinds of energy, as they are of an electromagnetic nature and the direction of the axis of the collector coils stands at the right angles to the earth's surfac less absorbed in the same way as a receiver in a wireless telegraphy absorbs waves coming from to distance. It is possible to connect large surface and to take u of energy. can be more or less absorbed in the same way as a receiver in a wireless telegraphy absorbs waves coming from a far distance. It is possible to con and to take up large quantities of energy. can be more or less absorbed in the same way as a receiver in a wireless telegraphy absorbs waves coming from a possible to connect large surface and to take up large quantities of energy.

It is well known that it is large diameter. If it is constantly taking place in the atmosphere. Particularly in the tropics or still better in the polar regions where t are present, large quantities of energy may be obtained in this way. A coil with several windings should act the best. In similar manner, any alteration of the should act inductively on such a coil.

This is an induction in that collector coils of sufficient size. In the same way, this collector will react on the earth. By combining the previous kind of current form of electricity are considerably increased.

In order to produce the improved collector coil, a uniform current oscillations of an undamped nature may be called a "high vacuum or thermionic tubes of sui

instead of the previously known spark gaps (Figures 26, Nos. 9-18). Electromagnet The main aerial current flows through electromagnet ($^{1)}$) and may be c primary coils in the induction winding through wire. Parallel with the induction resistance 9 a regulating capacity of a suitable size, such as condenser 11 lower part of the tube, the tube is arranged for the incandescent filament or the cathode K which is fed through a battery B. From the battery B two branches ru

conductor E^1 and the other through battery B^1 and secondary coil 10 to the grid above *g* in the vacuum tube. By the method of connections shown in dotte

voltage at the grid electrode *g*may also be produced through the wire 17 and some small capacitors (*a, b, c, d*) connected in series, and conductor 18, witho being required.

The action of the entire system is somewhat as follows: -

The condenser pole 11 is charged and slightly damped oscillations are formed in the short circuited existing oscillation circuit of the condenser 11 and self ir fluctuations of the voltage in the anode circuit, with the same frequency, which changes the strength of the electrode . A permanent supply of energy to the osc and 10 consequently takes place, until a condition of balance is set up, in which the consumed oscillation is equal to that absorbed.

For a period of 180 ° so that if the grid is negatively charged, then the anode is positively charged and vice versa . This is the possibility of separating the osci the connection between the anodes and the grid circuit and so forth.

A second important factor in this way of converting the use of the grid circuit, or as shown by dotted lines 16, 17, 18 by means of a larger or smaller numbe

suitable size connected in series ; in this case the battery B^1 may be omitted. With a suitable selection of the grid potential in glow discharge takes place betw the anode A and a dark space is formed. The size of this cathode is influenced by the ions which are emitted in the lower space. On the other hand the numb through the grid is dependent on the grid between the grid and the cathode. Thus is the grid voltage undergoes periodic fluctuations (in the present case) t cathode drop at the grid fluctuates and consequently the internal resistance of the tube correspondingly fluctuates, so that takes place, the necessary means ar production of undamped oscillations and of taking current,

The frequency of the undamped oscillations produced with the coupling of the oscillation circuits 9 and 10. By a suitable selection of the self induction of th electromagnetic oscillation of only a few meters wavelength down to the lowest High vacuum transmission tubes of 0.5 to 2 kw in size can be connected in para

Dynamics circuit is also known to be used as such tubes for producing undamped oscillations, and also on the contrary their use at lower voltages is considerab Gaps must be considered as an ideal solution for the use of high-voltage electricity.

By the application of safety spark gaps, with interpolation of electromagnets, not only is short-circuating but also the taking up of current is regulated. Osc inserted in the above way form a constantly acting electromagnetic alternating field in the collector coil, whereby accumulated effect takes place. electromagnetic alternating field, so long as the direction of its axis running parallel with that of the main current.

In oscillation circuits 9 and 11, electromagnets S and S^1 must be inserted if the high frequency oscillations are not to be penetrated the collector oil, betwe producers and the collector coil. In all other cases they are connected shortly before the earthing (as in Figures 27 and 28).

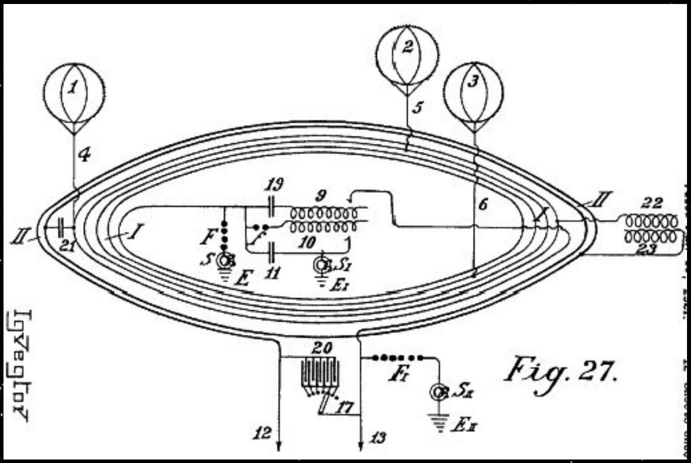

Fig. 27.

In Figure 27 a second method of construction of the connecting conductor is shown in the form of a coil. The other annular conductor II is inserted parallel to t the air but both in the form of a coil. The connection of the balloon aerials is indicated as a primary conductor and also as a current producing network; the ot and is not connected to the current production network.

In Figure 27, the current producing network is shown with three balloon collectors 1, 2, 3 and aerial conductors 4, 5, 6; it is short-circuited through condenser

9. The oscillation forming circuit consists of this diagram of spark gap f, inductance 10, and condenser 11; the earth wire E 1 is connected to earth over elect the spark spark gap which is also connected to earth through a second electromagnet S at E. On connecting the condenser circuit. This discharging cur inductance 10 on the inductively coupled secondary 9, whereby in the producing network a modification of the potential of the condenser 19 is produced. producer network. These oscillations induces a current in the secondary circuit, which corresponds to the proportion of the number of windings and of the oh greater.

In order to be able to convert the current wavelengths, a sufficiently large regulatable capacity of 20 into inserted between the ends 12 and 13 of the secondary and to be connected to the earth over an electromagnet S2.

The over-induction condenser 21 or over an induction condenser 22, 23. In this case the secondary conductor is unipolarly connected with the energy conducto

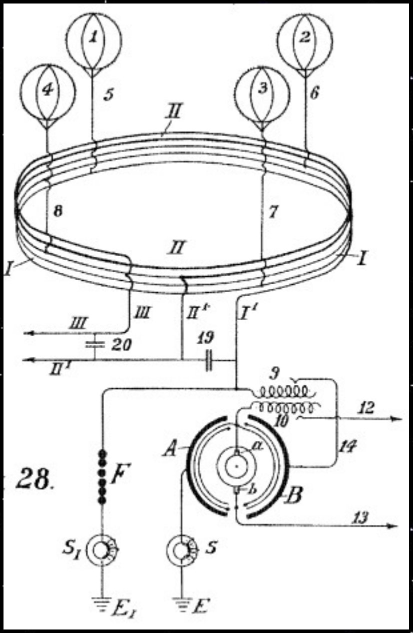

28.

In Figure 28, the connecting conductor between the separate accumulator balloons is carried out according to the autotransformer principle. The collecting c aerial balloons 1, 2, 3, 4, the windings of which are not made side by side but one above the other. In Figure 28 the collector coil I is shown with a thin lin connected prolongation coils II with a thick line. Between the ends I^{1} and II^{1} of the energy networks in the regulating capacity 19 is inserted. The wire ^{1} is c wire output and the spark gap F.

As a transformer of the electricity supply, it is made of rotary pairs of condensers in which the one is connected to the earth pole. Collector rings and brushes, alternating current, the frequency of which is dependent on the number of balloons and the revolutions of the rotor. In this invention, in this improved method this invention, through coils 1 on the inductance 9, an increase or decrease of the current feed in I can be obtained according to the direction of the current by ba

These rhythms produce short-circuited through the regulatable condenser 19 these rhythms produce short-circuited undamped oscillations in the energ periodicity and wave wavelength and therewith also to a given frequency. These currents may Also be employed in this form directly as working current throu II^{1} and III. By inserting the condenser 20 a connection between these conductors may also be made, whereby harmonic oscillations of desired wavelengths ar means quite new effects as regards current distribution: Coil with a cable or a cable to a point in the interior of the producing network is firmly or movably mo conductor. In this case it is induced in the producing network, the size of which is dependent on the total capacity and resistance and also on the periodi potential or afford afforded in the future, network by wireless means. This is also the case for the acquisition of energy for the future.

Instead of spark gaps grid vacuum tubes may be used as producers for undamped oscillations. The separated coils of the producer may be connected with on separate conductors all in parallel or in series or in groups in series. By collulating the number of oscillations and also the extent of the voltage more or less T be spirally over the entire section. Angular form or also in triangular, quadrangular, hexagonal or octagonal form.

The current waves as guides. This is necessary when the currents have been taken over mountains and valley and so forth. In all these cases, the current must b current of suitable periodicity.

As already hereinbefore mentioned separate collecting balloons may be directly metallically interconnected at equidistant stations distributed over the entire c connected by interpolation of suitable capacitors. A range of oscillations, and could be in such forms, a wit to a suitable arrangement of the connections, measures of protection, regulation.

Collecting effect of the balloon in the aerial collector of the balloon, radiating collectors are employed. These consist either of incandescent metal or oxide elect of vacuum tubes, or electric arcs (mercury and the like electrodes), Nernst lamps, or finally flames of various kinds may be connected with the respective condu

An incandescent body with an incandescent body and an electric tube. Hitherto however, a cathode was always first directly placed as an anode, and secondly t consisted of a closed circuit.

Light and flame arcs in which a cathode only as a source unipolar discharge (which represent group and point discharges in electrostatic machines si discharges), it may be ascertained that incandescent cathodes and less all incandescent radiators, flames and the like radiate into the open space in the form o as transmitters.

The object of this invention is as described below, if such incandescent oxide electrodes or other incandescent radiators or flames are not supported in spa metallically with the earth so that they can be charged with negative terrestrial electricity, these radiators possess the property of absorbing the free positive e They can serve as collectors and have, in comparison to the action of the spikes, or points, a very large radius of action R; the effective capacity of these c greater than the geometrical capacity (R$_{or-}$) calculated in an electrostatic sense.

Now as our earth is surrounded by an electrostatic field and the difference of potential

$$\frac{\delta V}{\delta h}$$

of the earth field according to the latest investigations, is in the summer about 60 to 100 volts and in the winter 300 to 500 volts per meter of difference in heig calculation gives the result that when a radiation collector or flame collector is arranged for example on the ground, with a distance of 2000 meters and both a conducting cable, there is a difference in potential in the summer of about 2,000,000 volts and in winter even of 6,000,000 volts and more.

According to Stefan Boltzmann's law of radiation, the quantity of energy, which is an incandescent surface (temperature T) of 1 sq cm, radiates into a unit of t air (temperature T_o) is expressed by the following formula:

$S = \delta(T^4 - T_o^4)$ watt./sq cm.

Equal to 6.30 x 10^{-12} watt / sq. Cm. And the universal radiation δ is based on the latest researches of Ferry (*Annales de Chimie et de Physique* , 17: 267 [1909

Now if an in incandescent surface of 1 sq cm shows, as Compared with the surrounding space a periodic fall of potential δV it radiates (independent of the That is to say of the sign) secondo with the above formula, for example at a temperature of 3725 ° C an energy of 1.6 kw / sq cm / second. As for the radiatio can be calculated for the collection of energy, but reversed. 60 cm to 65 cm per cm per cm, no carbon will result in this direction in employing radia accumulators.

If the earth be regarded as a cosmically insulated condenser in the sense of geometrical electrostatics *x* there results from the geometric (appears Edwald Rasch *Bogenlicht* [The Electric Arc Light], page 169) capacity of the earth according to Functional Chwolson:

For negative charging 1.3 x 10^6 Coulomb

For negative potential V = 10 x 10^8 volts.

From this there results however, EJT = 24.7 x 10^{24} watts / sec. This is an example of an electrician, who works for about 79,500 x 10^{10} kilowatt years. thermodynamically, electromagnetically and also kinematically coupled with the sun and a system of cosmic radiation and gravitation. The energies Which collectors would withdraw from the earth field can only houses by the withdrawal of motor work in lowering of the earth temperatures (temperatures T_E = 30

to the world of space (T = O) by using the entire energy. This is not a cosmically entirely insulated system. From the sun an energy of 18,500 x 10^{10} kw. lowering of the earth temperature (T_E) without a simultaneous lowering of the sun's temperature (T_s) would contradict Stefan Boltzmann's law of radiation.

$S = \delta(s^4 - T^4)$.

From this it must be Concluded That if the earth temperature (T_E) sinks the total radiation S absorbed by the earth Increases, and Further Also That the secula of the earth is directly dependent on That of the sun and the other radiators cosmically coupled with the sun and is connected with most closely.

The incandescent radiation collectors may be used for negative energy and if they are (1) are charged with the negative earth electricity and (2) if large surfaces) charged with electricity are mounted opposite them as positive poles in the air. It is not possible to collect with an incandescent collector, sufficiently the electrical charges contained in the atmosphere as the technology requires; very small, especially if it be considered that the very small surface density (e about = 2 x 7. 10^9 St. E. per sq cm) does not allow large quantities of charge being absorbed from the atmosphere.

x) Calculated according to Poisson's calculation:

$\Delta V = -4\pi\delta$; as here the alteration of the potential or potential is taken place in the direction or the normal, this calculation assumes the simple form

$$\delta = \frac{1}{4\pi} \times \frac{\delta^2 V}{\delta n^2}$$

It has been already proposed to be used as collaterals for collectibles. It is however, not known that the quantities of current which could be hitherto be s purposes. According to my experiments the reason for this is too small capacities of the collector conductor poles. If such flame or radiating collectors have positive surfaces, their radius of action is too small. If the incandescent collectors are being held in practice, they are not yet being carried out in practice.

By this invention the collector effect is considerably increased by a positive potential and the possibility of being held floating in the air at a desired height. A of sheet metal or of metalized balloon made to mount to 300 meters to 3000 meters in the air and as a positive pole it is brought forward to a radiating collect conductor to the earth, quite different results are obtained.

The metallic balloon shell (with a large surface) This is the greater the collecting balloon is above the incandescent collector. It is being attracted through th ionization, proceeding from the incandescent cathode. Collection of the incandescent cathode collector is appreciatively enhanced and also the collection balloon surface. More than a large piece of water, if it is not enough .

In the air and being in the air in the air conductively connected to earth.

Condenser and an inductive resistance being switched on in parallel, a process that can be carried out by two such contacts.

Such as radiating collectors in series. An incandescent cathode may be seen from the open ground and an incandescent cathode which is heated by specia currents. Of course for this the special vacuum Liebig tubes wit or without grids may also be employed. An ordinary arc lamp with oxide electrodes may be ground and the positive pole is not directly connected to the collecting balloon, but through the upper inc. The method of connecting the incandescent cathode may be seen in Figures 29-33.

B is the air balloon, K a Cardan ring (connection with the hawser), C the balloon, L a good connecting cable, P a positive pole, N negative incandescent cat conductor.

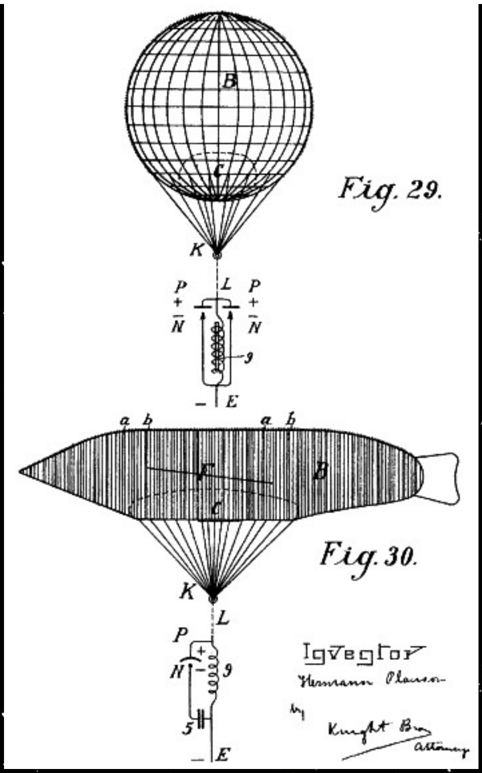
B
C
Fig. 29.
K
P
+
N
L
P
+
N
9
E
a b
a b
F
B
C
Fig. 30.
K
L
P
N
9
5
E
Inventor
Hermann Plauson
by Knight Bros
attorneys

Figure 29 represents the simplest form of construction. The electric oscillations are produced below the ground by means of a carbon arc lamp or other suitabl the electrical resistance. is formed, and as, over N and P only an inductionless ohmic resistance is present, a spark will spring over as long as the separate induc and the like are correctly calculated. The oxide electrode (carbon or the like) is rendered incandescent and then shows as an incandescent cathode an increased The positive poles must be larger than the negative. A large balloon that has a large capacity and is charged at high voltage, an incandescent body that is hel positive pole that can collect large capacities. The incandescent cathode is the first to become incandescent by means of separate energy produced on the earth.

Figure 3 only shows the difference between the incandescent cathode and the earth conductor PN 5 and 9 is obtained. This is the small quantity of electrici cathode to become incandescent.

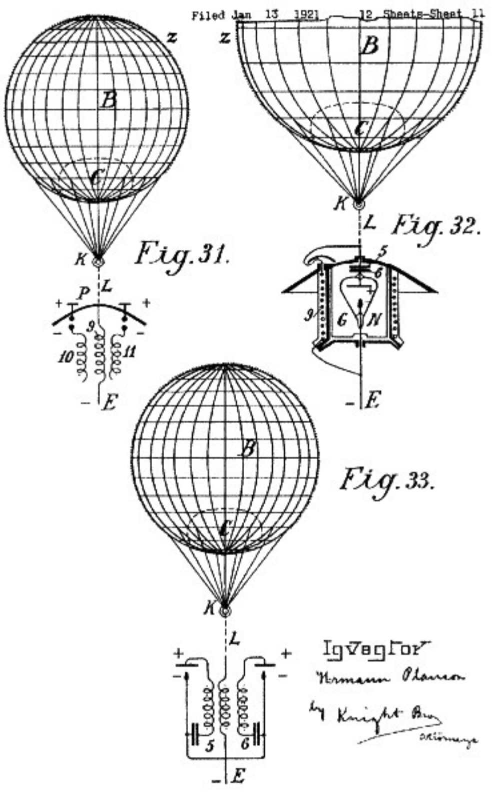
Filed Jan 13 1921 12 Sheets-Sheet 11
Z
B
C
K
Fig. 31.
P
L
9
10
11
E
Z
B
C
K
L
Fig. 32.
5
6
9
6
N
E
B
Fig. 33.
C
K
L
5
6
E
Inventor
Hermann Plauson
by Knight Bros
attorneys

In this form of construction, both the incandescent cathode and the positive electrode may be enclosed in Figure 32. A cable is insulated in a condenser dis arched in order to keep off the rain. The vessel is completely made of magnetic metal and well insulated inside and outside. Opposite the disc 5 another disc 6 a positive pile of the vacuum tube with the incandescent cathode (oxide electrode) N is arranged. The negative electrode is being connected to the earth. Figur of an open incandescent cathode one enclosed in vacuo is employed. As in such collectors only small bodies can be brought to incandescence in large installatio such vacuum tubes must be inserted in proximity to one another. According to the previous constructions Figures 31 and 33 are quite self evident without furthe

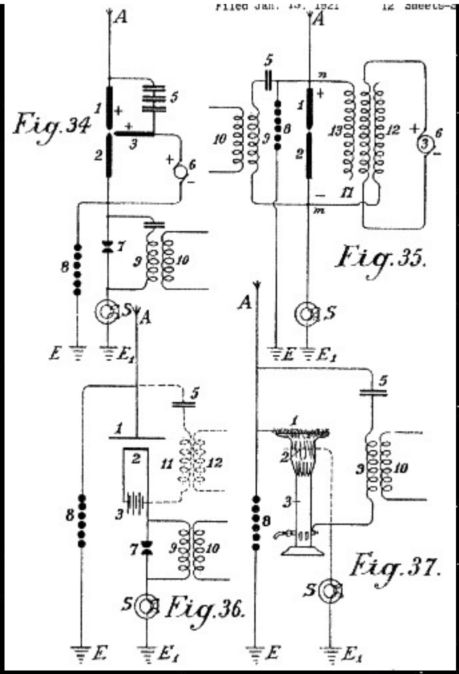

Figures 34-37 represent further diagrams of connections over radiating and flame collectors, and in fact, they are to be arranged on the ground.

Figure 34 shows an arc light collector with oxide electrodes for direct current and its connection; Figures 35 a similar one for a current flame, Figure 36 a simi flame.

The positive pole 1 of the radiating collectors is always directly connected to the aerial collecting conductor A. In Figure 34 this is connected Further over the 5 with a second positive electrode 3. The direct current dynamo *b* Produces Which current flows over between the electrodes 3 and 2 as an arc light. The neg electron 2 absorbs electricity from the positive poles The spark gap 7, inductive resistance 9 and induction coil 10 are like the ones previously describe electromagnet S the guard against the circuitry, the safety spark gap 8 from excess voltage or overcharging.

In Figure 35 the connection is so far that the alternating current dynamo feeds the exciting coil 11 of the induction condenser. 12 is its negative and 13 its po coil 3 on the magnet of the dynamo is correctly calculated and the periodicity of the alternating current is sufficiently high and can be formed between the two the cathode 2 is connected with the negatively charged earth, and therefore always acts as a negative pole, a form of rectification of the alternating current dynamo 3 is obtained, the second half of the period is always suppressed. The working circuit may be carried out in the same way as in Figure 34; the wo however be dispensed with,

Figures 36 represent a form of construction similar to the figure of an arc lamp in Nernst incandescent body is employed. The Nernst lamp is fed through t working section is connected with the negative pole. It is also available for 12 over the oscillation circuit 5, 11 (shown in dotted lines).

Flame collectors (Figure 37) may also be employed according to this invention. The wire network 1 is connected to the aerial collector conductor A and th earth. At the upper end of the dairy, long points are provided which project into the flame. The positive electrode is connected with the negative over a co induction coil 9 with the earth.

The novelty in this invention is firstly, the use of the incandescent cathodes opposite the positive poles which are connected with the large metallic capac collecting surfaces, (2) the connection of the incandescent cathodes with the earth whereby, in addition to the electricity conveyed to them from the battery c causes the incandescing, and (3) the connection of the positive and negative poles of the radiating collectors over a condenser circuit alone or with the introdu inductive resistance, whereby an oscillatory oscillation circuit may be obtained. The collecting effect is from these methods quite considerably increased.

I declare that what I claim is: --- [Claims not included here]

British Patent # 157,262

(10 July 1922)

Improvements in Electric Motors

Hermann Plauson // Otto Traun's Forchungs-Laboratorium GmbH

This invention relates to that type of motor in which rotation is produced by means of the attraction and repulsion of surfaces.

According to this invention a stator and rotor are formed of condensor surfaces and charges of electricity.

The invention is more particularly with reference to the accompanying drawing in which:

Figure 1 shows a simple form of motor and feed.

Fig.1

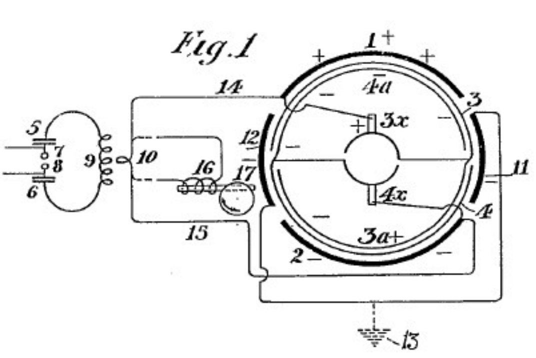
1+
14
4a
3
3x
12
11
5
7
8
9
10
16
17
6
4x
4
15
3a+
2
13

Figure 2 is a modification of Figure 1.

Fig.2.

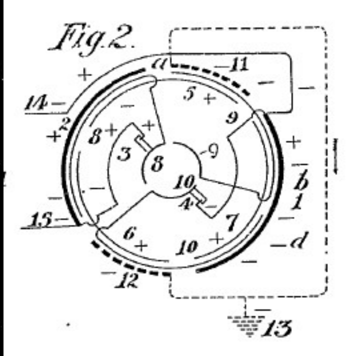
a
11
14
5
9
8
3
b
1
10
4
7
15
6
10
d
12
13

Figure 3 shows one form of a spiral condenser surface.

Figure 4 shows a wire wound condenser surface.

Figure 5 is a diagram of one type of rotor,

Fig.3

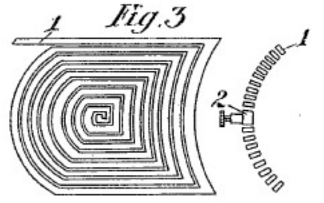

Fig.4

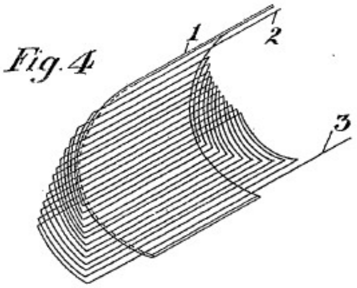

Fig.5.

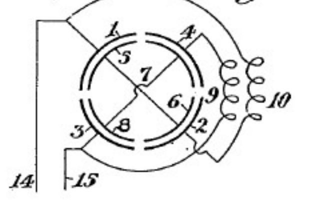
1
4
5
7
9
10
6
3
8
2
14
15

The inner plates of the condenser 7, 8 connected to a source of energy of sufficiently high pressure (alternating or direct current), until the potential has ris springs over.

The spark gap 7, 8 forms with the condenser 5 and self-inductance 9 and condenser 6 a closed oscillatory circuit and alternating currents of high frequency w this circuit. Currents of the same periodicity.

The improved type of motor is fed by the discharges produced by the induction in the secondary circuit.

Hitherto only Tesla's motor system (shown diagrammatically in Figures 1, 16 and 17) was known for this purpose. The above-mentioned diagram is only sho the fundamental principle. It has no practical interest for carrying out large machines by reason of the impossibility of the regulation and the low efficiency.

All of these processes will be overcome by the construction of a machine which is applicable for high frequencies. The difference between the principle of co motors as compared to those of the principle of magnetic induction only (as they have been all motors hitherto and also Tesla's motors).

It will be fed directly with a static electricity but if it is connected to a source of high frequency current it will operate.

The applicants call this new type of motors 'condenser motors' to differentiate them from the hitherto existing types.

The simplest form of construction of such condenser motors is shown in Figure 1, and this motor can be fed with high frequency alternating currents.

To a moment of positive energy is charged by means of the lead 14 to the stator surface 1 and to the brush 3x (Figure 1). The 3x brush is connected with the that both the surface stator and the rotor surface 3 is charged with positive electricity. The stator surface 1 and the rotor surface 3 and the rotor surfaces 3 an with negative electricity, such motors can be a version of the alternating current in a manner similar to the starting of the synchronous motors of known constru

Although this motor is easy to start it can only be employed for small experimental and measuring purposes because the stator and rotor surfaces are made of s heated by Foucalt (eddy) currents. In spite of its simplicity and its unsuitability for use in practice it must however be regarded as a basic type for technical calc

The condenser motor shown in Figure 2 differs from Figure 1 by the rotor surfaces consisting of six condenser surfaces connected one behind the other in s connected with three collector surfaces, so that at any one moment only two adjacent collector surfaces come under the two brushes (3 and 4). In its other acti to Figure 1. The leads 14 and 15 may be connected either to the ends of the secondary coil 10 or directly with the source of energy. The outer thicker line i surfaces 1 and 2 (that is to say the unmoving part of the motor), 11 and 12 shown by thick dotted lines means earthed additional poles of the stator, 8, 9, an parts of the rotor condenser surfaces which in turn are connected with the collector surfaces 8, 9, and 10. 5, 6, and 7 are the inner parts of the condenser surfac 3 and 4 are brushes.

Hitherto stator and rotor surfaces. These conditions become highly heated with eddy currents and hardly yield 10-15% of useful effect. It is found in certain the stator and condenser surfaces. It was not possible to be cut in the metal surface of the stator and rotor in the form of a spiral, not only was it to be improved.

Experiments have shown that this is possible to build a very useful motor for high frequency alternating currents more particularly those of an undamped natur

Shown in Figure 1, but with the model shown below in Figure 5, but with the form of construction of the condensers of the stator as well as of the rotor accor condenser motor is obtained which works well in either direction for high frequency alternating current. It was also observed that the motors in such forms of found to be sensitive to resonance effects. Such a motor can be used in the same way if the winder is in the stator and also in the rotor.

A motor construction according to the foregoing kind is already fully technically applicable. But even these motors have a series of faults, more particularl construction. For example, the attachment of the spiral condenser both in the stator and of the rotor sown in Figure 3 are in practice difficult to carry out. In shown in Figure 4. These stator and rotor surfaces may be without further difficulty, be considered as electromagnetic poles, although they are not made of electromagnets.

In the case of the coil as shown in Figure 4, the coil can be embedded in insulating material or as the motorist as it has been done in the case of ordinary sin motors. To the same time the possibility is afforded by increasing the number of changes to the individual induction co-efficients.

In Figure 5 is shown a modified construction of a capacitor surfaces 1, 2, 3, 4, of which 1 and 2 are connected though an iductance 9 coupled with the c surfaces 5 , 6, 7, 8 are provided of which 5 and 6 are connected directly 7 and 8 similarly connected.

The pair of like poles are connected by wires 14 and 15 to the source of energy. By a suitable selection of the values of the reactance and capacity in these circu

We have been able to do this, and we will be able to declare what we claim is: --- [Claims not included here]

British Patent # 157,263

Process and Apparatus for Converting Static

Hermann Plauson // HO Traun's Forschungs-Laboratorium GmbH

In the form of direct current in the form of a direct current in the form of a wave in the form of a direct current in the form or by means of a special kin 'condenser motors' --- ready to be used for technical purposes as mechanical energy.

For small installations, this system may be very well employed; about 100 horsepower may be stated as practical limit. In constructions of larger aggregate dif the spark gaps however increase considerably. It can be used as an alternative to a number of alternatives.

British Patent # 157,262 It was observed that the rotor, if one pole of the stator was connected to the air, collecting aerial electricity and the other pole with could a motor, but if vice versa the rotor connection with the stator be interrupted and the rotor caused to rotate by means of another motor, that when t alternating current the periodicity of which is dependent on the number of poles and the revolutions of the rotor. Such an apparatus may be considered as static into dynamic electrical energy.

The invention is more particularly with reference to the accompanying diagrams in which:

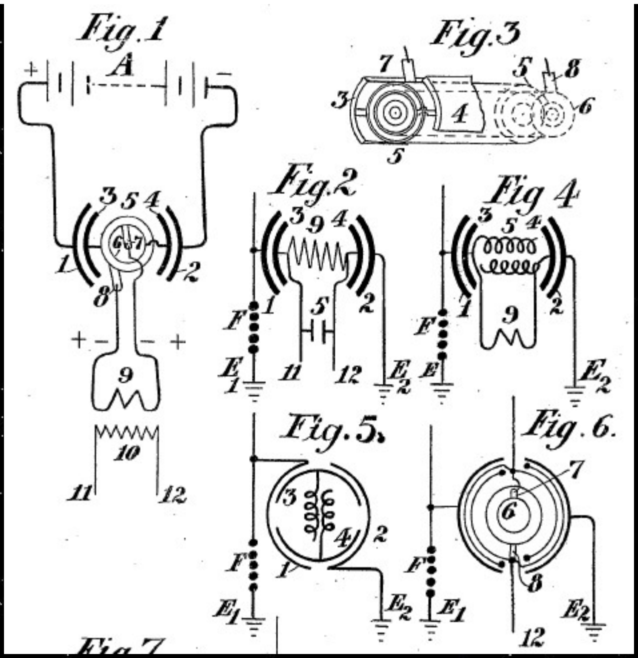

In Figure 1A is a strong accumulator battery, 1 and 2 are the outer poles of the transformer, consisting of simple metallic plates as shown in Figures 8-11, m without an electromagnet being present. These armors are similar to cylindrically curved plates 3 and 4. These are metallically connected with two collectors which two brushes 7 and 8 freely run which again are short 11- 12: If the accumulator battery is used with the free ends 11 and 12. 3 which is connected by th over the primary coil 9 with the second rotor surface 4. This is the case that you are using, if you are in the current circuit, or if you are in the same circu mechanical power, the rotor, the surface conditions are altered. After a revolution the plates are between the stator plates and therefore no condenser surface This is the case for the entire system. 9. Now if the rotor is turned further through 90 degrees by mechanical energy the rotor plate 3 plate 2 and the rotor pla stator plate 1, so that the rotor pates in a field of reverse sign. The first condition is again produced. The result of such a revolution is an alternating current which is equal to the number of revolution. In practice of course, not two poles but many poles as possible would be to increase the number of alternati alternating current thus obtained from the winding of the coil. Figure 7 shows a multipolar machine. The first condition is again produced. The result of such alternating current the periodicity of which is equal to the number of revolution. In practice of course, not two poles but many poles as possible would be to inc of alternations. The primary alternating current thus obtained from the winding of the coil. Figure 7 shows a multipolar machine. The first condition is aga result of such a revolution is an alternating current the periodicity of which is equal to the number of revolution. In practice of course, not two poles but as would have been to increase the number of alternations. The primary alternating current thus obtained from the winding of the coil. Figure 7 shows a multipo result of such a revolution is an alternating current the periodicity of which is equal to the number of revolution. In practice of course, not two poles but many would be to increase the number of alternations. The primary alternating current thus obtained from the winding of the coil. Figure 7 shows a multipolar mach such a revolution is an alternating current the periodicity of which is equal to the number of revolution. In practice of course, not two poles but many poles as p to increase the number of alternations. The primary alternating current thus obtained from the winding of the coil. Figure 7 shows a multipolar machine. The pr current thus obtained from the winding of the coil. Figure 7 shows a multipolar machine. The primary alternating current thus obtained from the winding of shows a multipolar machine.

If the stator surface 2, but the rotor wich is otherwise constructed as hereinbefore, be rotated by a separate motor a much stronger alternating current result ascribed to the pole of the stator by reason of the higher pressure of the static electricity than where accumulators are employed. By this means the transformer scope of supply to it.

Figure 2 shows a mode of connections. The stator surface 1 is connected to the aerial antennae which is connected through the safety spark gap F to earth surface 2 is directly earthed at E2. The inner revoluble rotor surfaces 3 and 4 are interconnected by means of an induction coil which is constructed directly current is assumed as in Figure 1, to be shown for the sake of clearness, and further conveyed through the 11 and 12 conductors. There is a short circuit oscill spark gaps, which circuit consists of the induction coil 9 and condenser 5 and is fed by the periodic charging current impulse. This is the possibility of ob current which is characterized by longer periods and is undamped and oscillatory. Of course a simple alternating current may be obtained by cutting the conden

Instead of the induction coil the condenser may also be b constructed in the rotor. It can be carried out in such a way that it serves directly as collector rings through the brushes. In Figure 3 such a motor is sketched in perspective, 3 and 4 are the rotor surfaces, 5 and 6 are the condenser surfaces constructed to form consisting of two co-axial cylinders fitting free on the brushes 7 on one end of the condenser cylinder 6.

Is shown in Figure 4. A further type of transformer is shown in Figure 5. The difference consists in the stator and rotor surfaces. of the circuit but almost the h condenser is better utilized. Full face of the stator surfaces.

In addition to condition is obtained in which the stator surfaces are inductively connected by the rotor surfaces. The connection is as before.

Figure 6 shows the alternation of the rotor surfaces; two cylindrical condenser plates arranged concentrically, each divided into two halves and connected so t cylinder is connected to half the outer. Transformer action.

Fig.7.

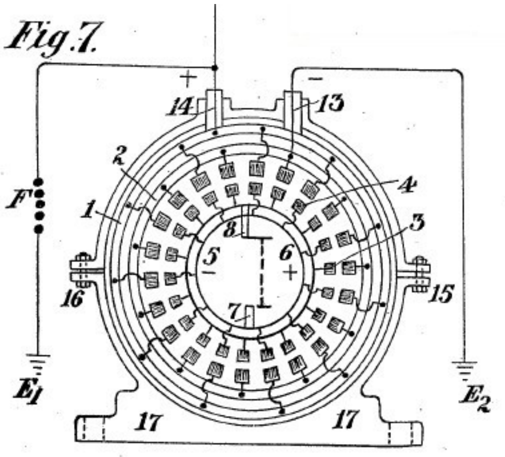

Figure 7 shows a four polar transformer. It consists of a metal casing, the lower half of which is fastened with the foundation plates. 17 to the support or foun
half, the cover, is connected by bolts 15 and 16 This upper sleeve or case is insulated from the under part. Two rings 1 and 2 are cylindrically constructed in th
1 is metallically connected with the collector and the ring 2 with the earth. On both rings and an equal number of stator surfaces are mounted side by side
from one another and thus form an electrostatic field similar to the electromagnetic in many alternating current machines. The rotor consists of the similarity o
6 on which an equal number of rotor surfaces are fixed. By the brushes 7 and 8 The charge is conveyed by the conductor 14 to and by 13 away. Alternating cu
the rotor, the periods of time are rotated by revolutions per second.

This article was previously published under Q140870 and was used as an alternative to a conversion to the friction. To the effect of the considerable energy
caused by a conductor being moved through strongly electrostatic fields since the electrostatic lines of force must be cut to the right angles and that to
apparatus should not only be regarded as a transformer, but also as an energy producer, with the difference that the excitation is here, instead of by means of ele
entire system may, to some extent, be applied with a dynamo in which the excitation takes place. It was further ascertained that this way of using the atmo
produced as a result of suction on the collector network.

The effects which in this instrument became evident are extremely interesting and open to the future. Merely that these transformers become already the extre
apparatus. If you are a future manufacturer, you will be able to produce a certain quantity of energy which may be constructed in the future. according to
system.

The results of the examinations made for this may be construed as follows.

(1) If solid electrodes (condenser surfaces, rotor and stator surfaces) are used they become hot. Ribbed form electromagnetism Figure 8, but not entirely re
allows the surface of the condenser plate to be enlarged or increased; the electrodes may be fastened in a simple manner on the under frame by perforation 1, 2,

Fig.8

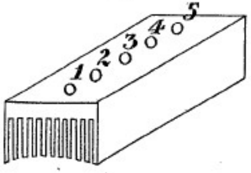

Fig. 9.

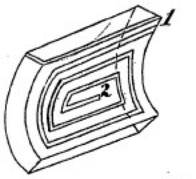

(2) If nicks or notches in spiral form as shown in Figure 9 seen from the side in Figure 10 in section are employed, not only is the transformer effect greater b
their movement than a simple commutator action would require.

Fig. 10

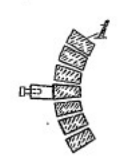

Fig. 11

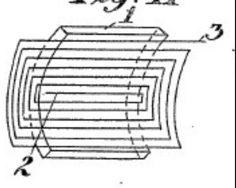

(3) The greatest effect is obtained if the rotor and stator surfaces are wound into a flat spiral form of suitably thick wire, and in that a way that the inductive ef
with the capacity suitable periodicity. In practice this is done by spiral wire (see Figure 11) so that a smooth pole is similar to that in phase motors.

Fig.12

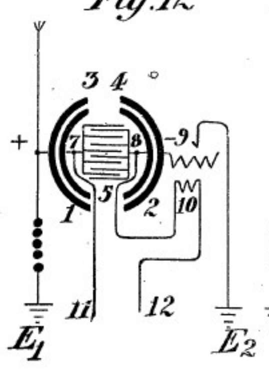

Fig. 13.

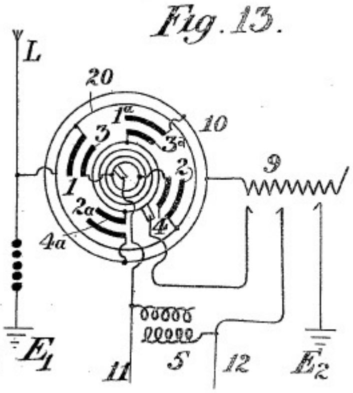

The aerial wire is shown in Figure 13. The aerial wire is shown in Figure 13. The aerial wire is connected to the ring. The inductive earth pole is also connec
ring 10 from which again two, poles 1a and 2a are branched off. Of course in the same manner any suitable number of poles may be branched of. In similar ma
two poles fastened to one another (3 and 4 and 3a and 4a) connected with separate collector rings. From these two rings the current is collected by means of
induced alternating current is however directly metallically connected with an inductive earth stator conductor over an induction coil 9. In parallel with the co
sparkless oscillatory circuit that can be applied to the exciting current in the stator. This produces however, a periodic alteration of the charge according to the
rotor currents in consequence of which the stator charge also commences with resonance oscillations and if the stator and rotor surfaces are calculated to ar
oscillations of waves of similar length the form of which is dependent on the amplitude of the main alternating current and is caused by the number of the pole
per second. Thus an alternating current of, for example, 100 periods of time is formed, the separated periods of which are formed by undamped oscillations o
of alternations. In Figures 14-16 four other diagrams of converters are shown, the object of which is not to produce the usual alternating current, but os
frequency.

Fig.14

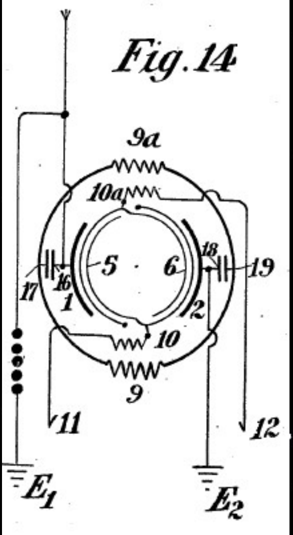

The main difference of these systems from the one that is already present in the stator pole 1 (Figure 14) and one pole 16 of the condenser 17 and the earth con
second stator pole 2 and the pole 18 of the condenser 19. The other poles of these condensers 17 and 19 are short circuited through a ring over two inductive p
9a with one another. The secondary coils form the rotor conductors 10 and 10a. The rotor itself is constructed in the manner shown in Figure 6. In similar man
stator surfaces may also be formed. The collector rings of the rotor with the two brushes for the collecting current are shown in order to simplify the drawings.
maximum action by the connection of the two alternating current and the alternating current produced in the rotor on the stator circuit, with a correct calculati
and the self-induction may be obtained. The kind of current production will be similar to that described for Figure 12.

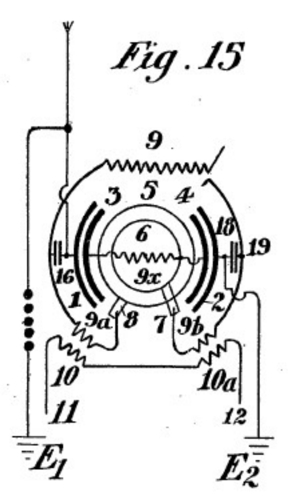

The novelty of the converter illustrated in Figure 15 consists mainly of the current in the rotor is not directly employed, but only serves the excitation of the pri 9b. The working current is produced in the secondary coils 10 and 10a and further conducted through the 11 and 12 conductors. The stator current may b regulatable inductive resistance to the same resonance as the rotor current.

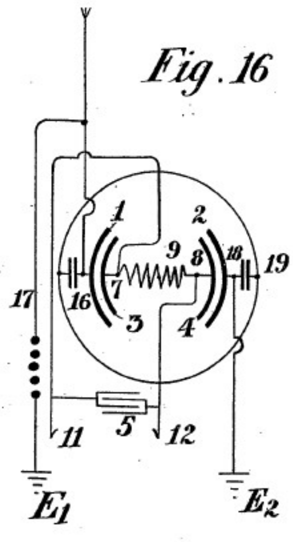

In Figure 16 a very similar system is shown in Figure 14. The condenser 5 is connected in parallel with the converter; and in the rotor a short circuited os formed which gives extraordinarily good results and is simple in construction.

The inductive resistance may be considered as primary coil used outside the rotor and short circuit (see Figure 17).

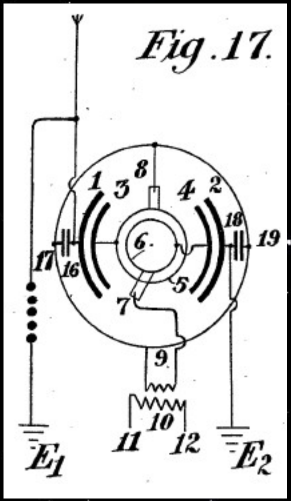

The last six types serves only for producing oscillations of a high number of alternations. If it be desired to be ordinary alternating current there complicat arrangements are not required as the types illustrated in Figures 1 to 11 suffice. It is self-evident that these arrangements may be altered in various ways.
We have been able to describe this as well as the fact that we are doing it, we declare that what we claim is: --- [Claims not included here]

US008253639B2

(12) **United States Patent**
Cohen

(10) **Patent No.: US 8,253,639 B2**
(45) **Date of Patent: Aug. 28, 2012**

(54) **WIDEBAND ELECTROMAGNETIC CLOAKING SYSTEMS**

(76) Inventor: **Nathan Cohen**, Belmont, MA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 498 days.

(21) Appl. No.: **12/547,104**

(22) Filed: **Aug. 25, 2009**

(65) **Prior Publication Data**

US 2010/0156556 A1 Jun. 24, 2010

Related U.S. Application Data

(60) Provisional application No. 61/189,966, filed on Aug. 25, 2008.

(51) **Int. Cl.**
H01Q 15/02 (2006.01)
H01Q 19/06 (2006.01)

(52) **U.S. Cl.** **343/753**; 343/909

(58) **Field of Classification Search** 343/753, 343/909, 910; 333/135
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,471,224 A *	11/1995	Barkeshli	....	343/909
5,714,961 A *	2/1998	Kot et al.	....	343/769
6,473,048 B1 *	10/2002	Diaz	....	343/753

* cited by examiner

Primary Examiner — Michael C Wimer

(74) *Attorney, Agent, or Firm* — McDermott Will & Emery LLP

(57) **ABSTRACT**

Arrangement of resonators in an aperiodic configurations are described, which can be used for electromagnetic cloaking of objects. The overall assembly of resonators, as structures, do not all repeat periodically and at least some of the resonators are spaced such that their phase centers are separated by more than a wavelength. The arrangements can include resonators of several different sizes and/or geometries arranged so that each size or geometry corresponds to a moderate or high "Q" response that resonates within a specific frequency range, and that arrangement within that specific grouping of akin elements is periodic in the overall structure. The relative spacing and arrangement of groupings can be defined by self similarity and origin symmetry.

14 Claims, 4 Drawing Sheets

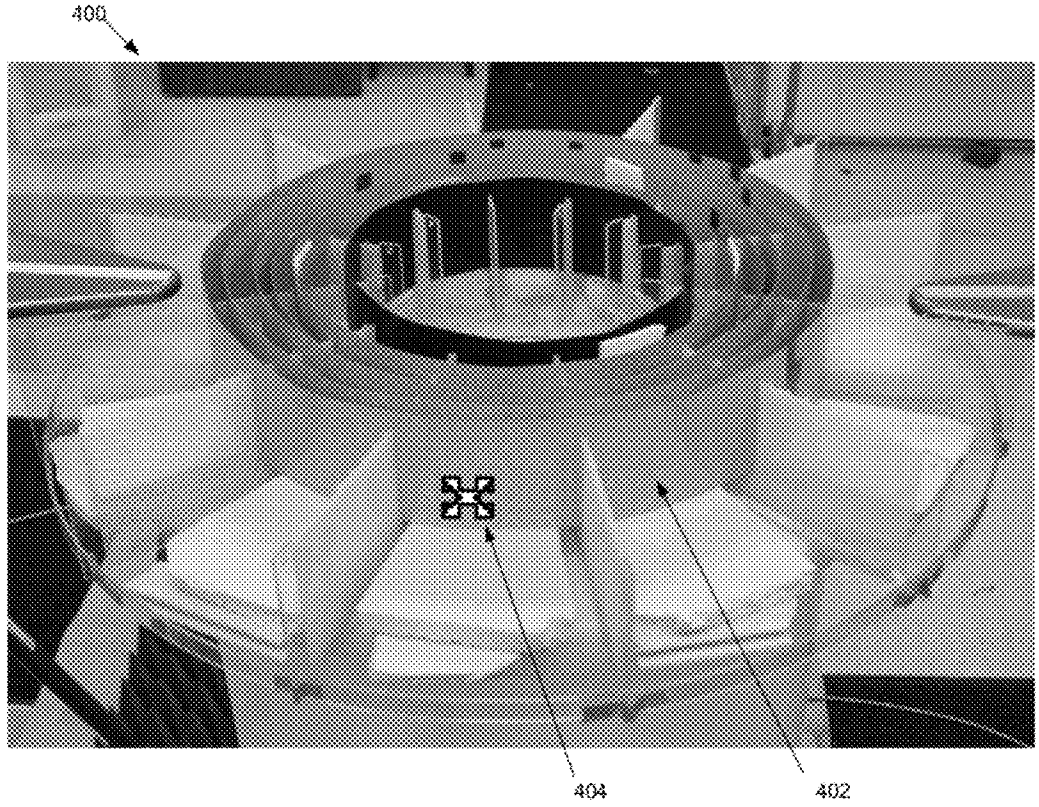

FIG. 1

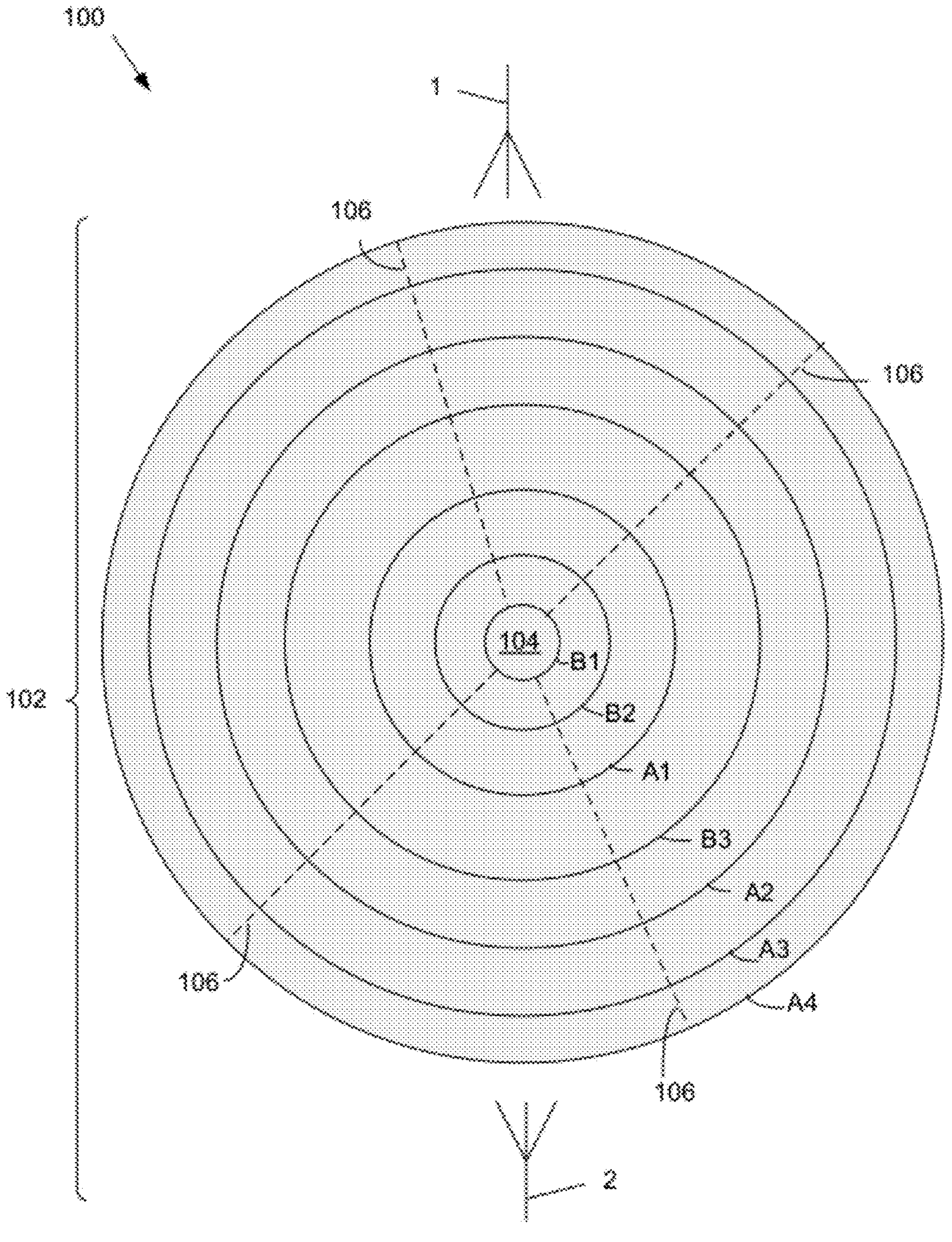
100
1
106
106
104
B1
B2
A1
B3
A2
A3
A4
102
106
106
2

FIG. 2

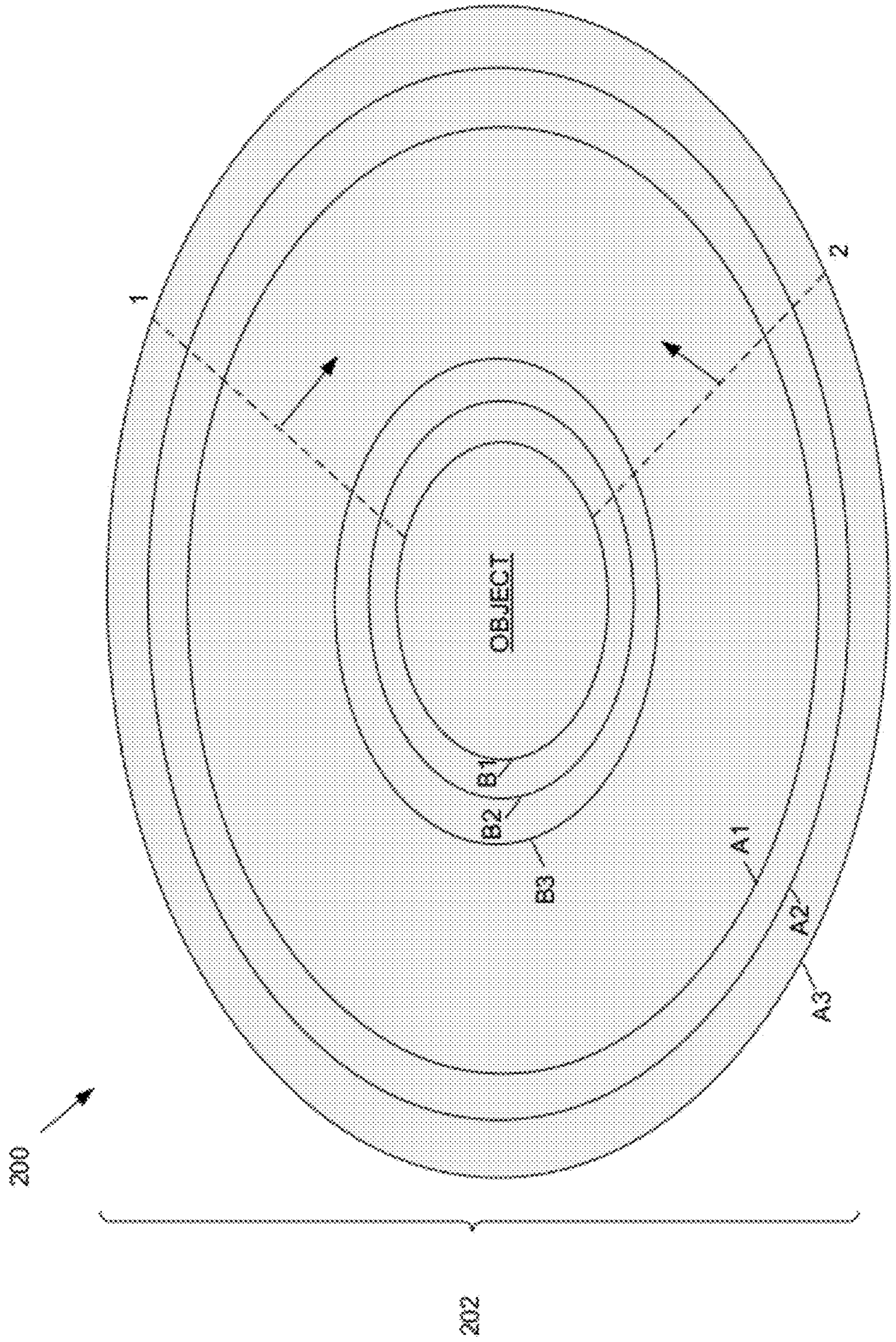
200
202
1
2
OBJECT
B1
B2
B3
A1
A2
A3

FIG. 3

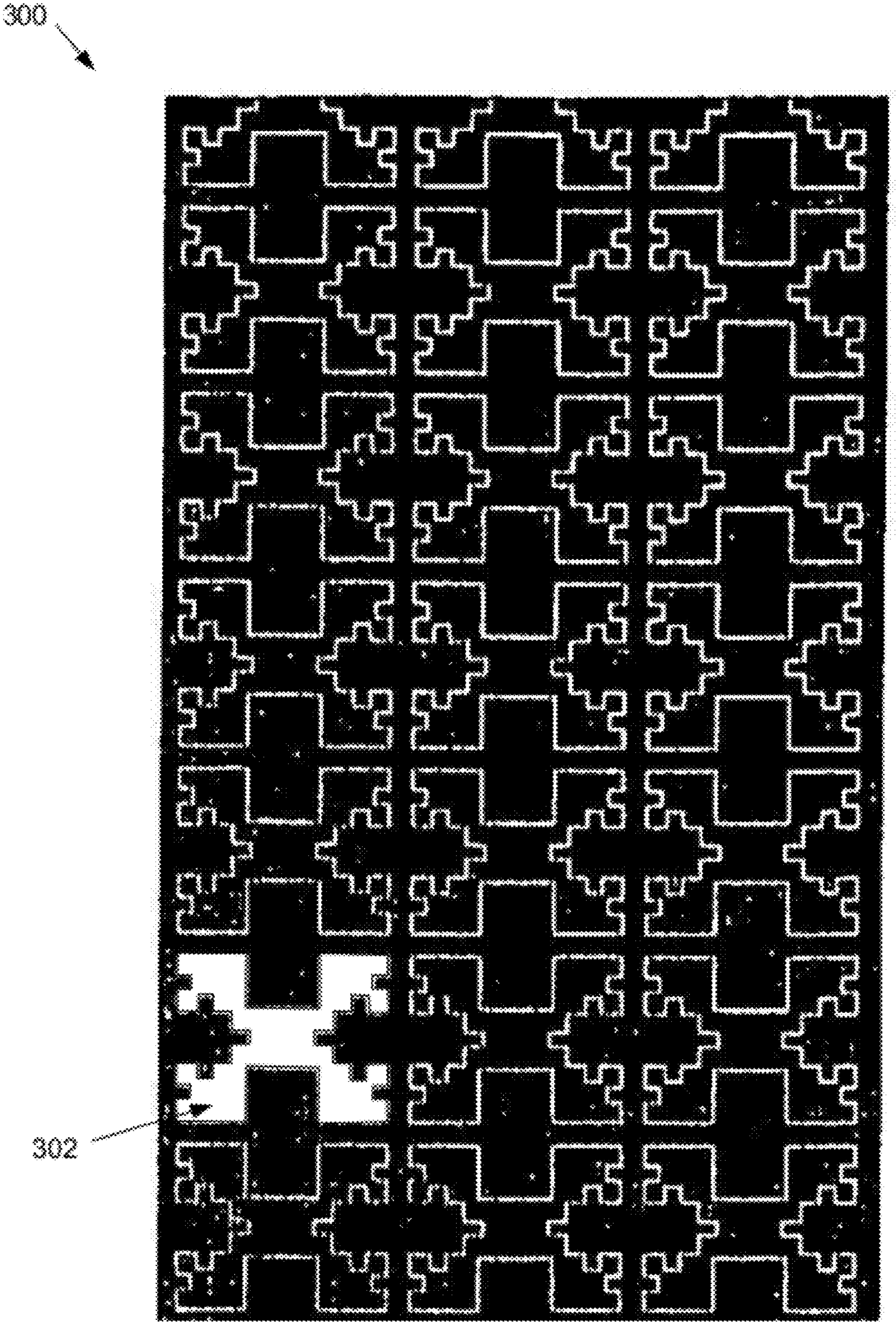

FIG. 4

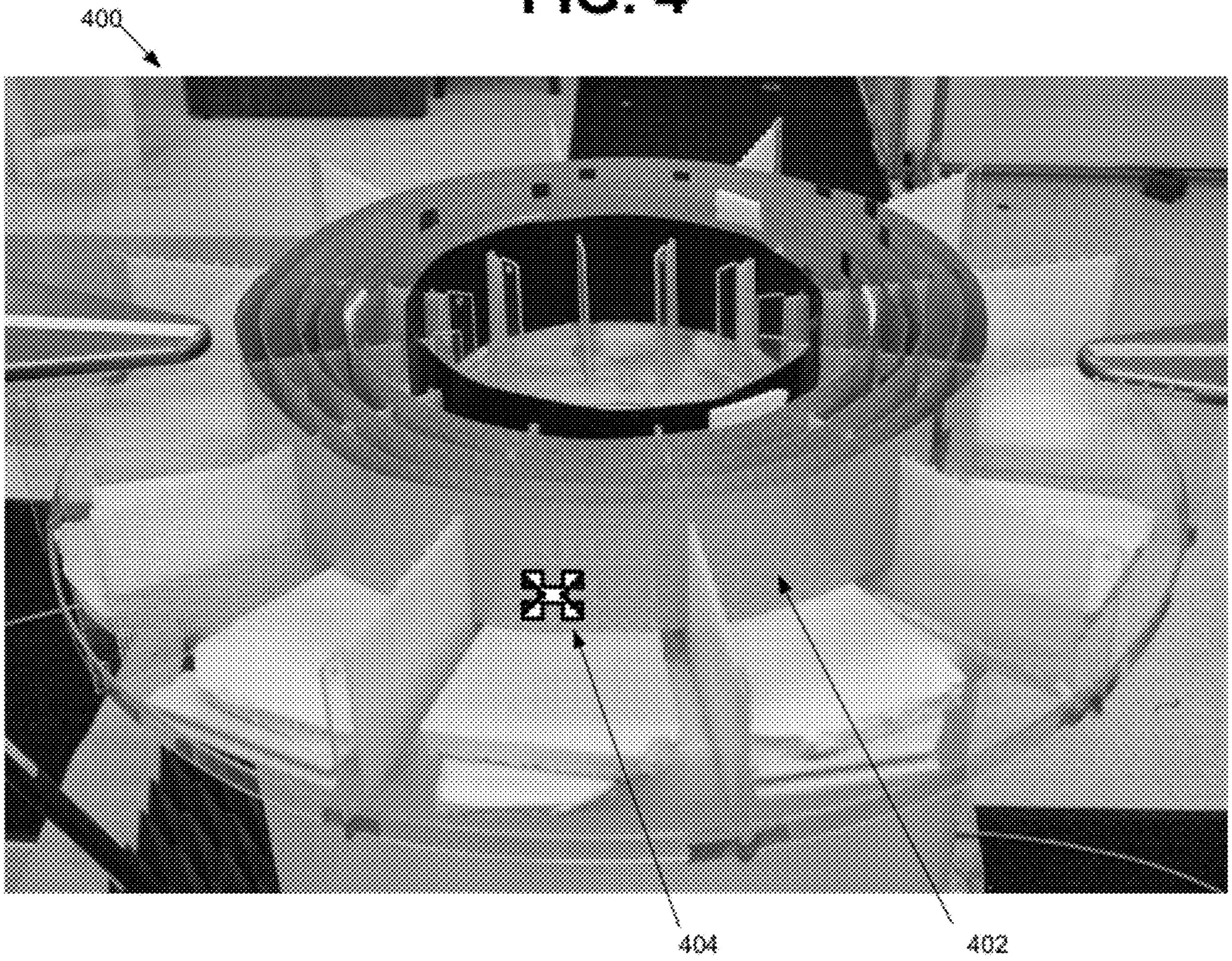

WIDEBAND ELECTROMAGNETIC CLOAKING SYSTEMS

RELATED APPLICATION

This application claims priority to U.S. Provisional Patent Application No. 61/189,966, filed 25 Aug. 2008 and entitled "Method and Apparatus for Wideband Electromagnetic Cloaking, Negative Refractive Index Lensing and Metamaterial Applications," the entire contents of which are incorporated herein by reference.

BACKGROUND

Much time and effort has been devoted to the quest for so-called invisibility machines. Beyond science fiction, however, there has been little if any real progress toward this goal.

Materials with negative permittivity and permeability leading to negative index of refraction were theorized by Russian noted physicist Victor Veselago in his seminal paper in *Soviet Physics* USPEKHI, 10, 509 (1968). Since that time, metamaterials have been developed that produce negative index of refraction, subject to various constraints. Such materials are artificially engineered micro/nanostructures that, at given frequencies, show negative permeability and permittivity. Metamaterials have been shown to produce narrow band, e.g., typically less than 5%, response such as bent-back lensing. Such metamaterials produce such a negative-index effect by utilizing a closely-spaced periodic lattice of resonators, such as split-ring resonators, that all resonate. Previous metamaterials provide a negative index of refraction when a sub-wavelength spacing is used for the resonators.

In the microwave regime, certain techniques have been developed to utilize radiation-absorbing materials or coatings to reduce the radar cross section of airborne missiles and vehicles. While such absorbing materials can provide an effective reduction in radar cross section, these results are largely limited to small ranges of electromagnetic radiation.

SUMMARY

Embodiments of the present disclosure can provide techniques, including systems and/or methods, for cloaking objects at certain wavelengths/frequencies or over certain wavelength/frequency ranges (bands). Such techniques can provide an effective electromagnetic lens and/or lensing effect for certain wavelengths/frequencies or over certain wavelength/frequency ranges (bands).

The effects produced by such techniques can include cloaking or so-called invisibility of the object(s) at the noted wavelengths or bands. Representative frequencies of operation can include, but are not limited to, those over a range of 500 MHz to 1.3 GHz, though others may of course be realized. Operation at other frequencies, including for example those of visible light, infrared, ultraviolet, and as well as microwave EM radiation, e.g., K, Ka, X-bands, etc. may be realized, e.g., by appropriate scaling of dimensions and selection of shape of the resonator elements.

Exemplary embodiments of the present disclosure can include a novel arrangement of resonators in an aperiodic configuration or lattice. The overall assembly of resonators, as structures, do not all repeat periodically and at least some of the resonators are spaced such that their phase centers are separated by more than a wavelength. The arrangements can include resonators of several different sizes and/or geometries arranged so that each size or geometry ("grouping") corresponds to a moderate or high "Q" (that is moderate or low bandwidth) response that resonates within a specific frequency range, and that arrangement within that specific grouping of akin elements is periodic in the overall structure—even though the structure as a whole is not an entirely periodic arrangement of resonators. The relative spacing and arrangement of groupings (at least one for each specific frequency range) can be defined by self similarity and origin symmetry, where the "origin" arises at the center of a structure (or part of the structure) individually designed to have the wideband metamaterial property.

For exemplary embodiments, fractal resonators can be used for the resonators in such structures because of their control of passbands, and smaller sizes compared to non-fractal based resonators. Their benefit arises from a size standpoint because they can be used to shrink the resonator (s), while control of passbands can reduce or eliminates issues of harmonic passbands that would resonate at frequencies not desired.

It should be understood that other embodiments of wideband electromagnetic resonator or cloaking systems and methods according to the present disclosure will become readily apparent to those skilled in the art from the following detailed description, wherein exemplary embodiments are shown and described by way of illustration. The systems and methods of the present disclosure are capable of other and different embodiments, and details of such are capable of modification in various other respects. Accordingly, the drawings and detailed description are to be regarded as illustrative in nature and not as restrictive.

BRIEF DESCRIPTION OF THE DRAWINGS

Aspects of the disclosure may be more fully understood from the following description when read together with the accompanying drawings, which are to be regarded as illustrative in nature, and not as limiting. The drawings are not necessarily to scale, emphasis instead being placed on the principles of the disclosure. In the drawings:

FIG. **1** depicts a diagrammatic plan view of a resonator cloaking system utilizing a number of cylindrical shells, in accordance with exemplary embodiments of the present disclosure;

FIG. **2** depicts a diagrammatic plan view of a resonator cloaking system utilizing a number of shells having an elliptical cross-section, in accordance with an alternate embodiment of the present disclosure;

FIG. **3** depicts an exemplary embodiment of a portion of shell that includes repeated conductive traces that are configured in a fractal-like shape; and

FIG. **4** depicts a perspective view (photograph) of an exemplary embodiment of the present disclosure.

While certain embodiments depicted in the drawings, one skilled in the art will appreciate that the embodiments depicted are illustrative and that variations of those shown, as well as other embodiments described herein, may be envisioned and practiced within the scope of the present disclosure.

DETAILED DESCRIPTION

The present disclosure is directed to novel arrangements of resonators useful for obscuring or hiding objects at given bands of electromagnetic radiation. Embodiments of the present disclosure can provide techniques, including systems and/or methods, for hiding or obscuring objects at certain wavelengths/frequencies or over certain wavelength/frequency ranges or bands. Such techniques can provide an

effective electromagnetic lens and/or lensing effect for certain wavelengths/frequencies or over certain wavelength/frequency ranges or bands. The effects produced by such techniques can include cloaking or so-called invisibility of the object(s) at the noted wavelengths or bands.

Representative frequencies of operation can include, but are not limited to, those over a range of about 500 MHz to about 1.3 GHz, though others may of course be realized. Operation at other frequencies, including for example those of visible light, infrared, ultraviolet, and as well as microwave EM radiation, e.g., K, Ka, X-bands, etc. may be realized, e.g., by appropriate scaling of dimensions and selection of shape of the resonator elements.

Embodiments of the present disclosure include arrangement of resonators or resonant structures in aperiodic configurations or lattices. The overall assembly of resonator structures can include nested or concentric shells, that each include repeated patterns of resonant structures. The resonant structures can be configured as a close-packed arrangement of electrically conductive material. The resonant structures can be located on the surface of a circuit board.

The overall assemblies, as structures, do not all repeat periodically and at least some of the resonators are spaced such that their phase centers are separated by more than a wavelength. The arrangements can include resonators of several different sizes and/or geometries arranged so that each size or geometry ("grouping") corresponds to a moderate or high quality-factor "Q" response (that is, one allowing for a moderate or low bandwidth) that resonates within a specific frequency range, and that arrangement within that specific grouping of like elements is periodic in the overall structure—even though the structure as a whole is not an entirely periodic arrangement of resonators. The relative spacing and arrangement of groupings (at least one for each specific frequency range) can be defined by self similarity and origin symmetry, where the "origin" arises at the center of a structure (or part of the structure) individually designed to have the wideband metamaterial property.

For exemplary embodiments, fractal resonators can be used for the resonators because of their control of passbands, and smaller sizes. A main benefit of such resonators arises from a size standpoint because they can be used to shrink the resonator(s), while control of passbands can reduce/mitigate or eliminate issues of harmonic passbands that would resonate at frequencies not desired.

Exemplary embodiments of a resonator system for use at microwave (or nearby) frequencies can be built from belts of circuit boards festooned with resonators. These belts can function to slip the microwaves around an object located within the belts, so the object is effectively invisible and "see thru" at the microwave frequencies. Belts, or shells, having similar closed-packed arrangements for operation at a first passband can be positioned within a wavelength of one another, e.g., $\frac{1}{10}\lambda$, $\frac{1}{8}\lambda$, $\frac{1}{4}\lambda$, $\frac{1}{2}\lambda$, etc.

An observer can observe an original image or signal, without it being blocked by the cloaked object. Using no power, the fractal cloak can replicates the original signal (that is, the signal before blocking) with great fidelity. Exemplary embodiments can function over a bandwidth from about 500 MHz to approximately 1500 MHz (1.5 GHz), providing 3:1 bandwidth; operation within or near such can frequencies can provide other bandwidths as well, such as 1:1 up to 2:1 and up to about 3:1.

FIG. **1** depicts a diagrammatic plan view of a cloaking system **100** and RF testing set up in accordance with exemplary embodiments of the present disclosure. As shown in FIG. **1**, a number of concentric shells (or bands) **102** are placed on a platform (parallel to the plane of the drawing). The shells include a flexible substrate (e.g., polyimide with or without composite reinforcement) with conductive traces (e.g., copper, silver, etc.) in fractal shapes or outlines. The shells **102** surround an object to be cloaked (shown as **104** in FIG. **1**). A transmitting antenna **1** and a receiving antenna **2** are configured at different sides of the system **100**, for verifying efficacy of the cloaking system **100** and recording results. The shells **102** can be held in place by radial supports **106** (while only four are shown, **12** were used in the exemplary embodiment indicated).

The shells indicated in FIG. **1** are of two types, one set (A**1**-A**4**) configured for optimal operation over a first wavelength/frequency range, and another set (B**1**-B**3**) configured for optimal operation over a second wavelength/frequency range. (The numbering of the shells is of course arbitrary and can be reordered, e.g., reversed.)

For an exemplary embodiment of system **100**, the outer set of shells (A**1**-A**4**, with A**1** being the innermost and A**4** the outmost) had a height of about 3 to 4 inches (e.g., 3.5 inches) and the inner set of shells had a height of about 1 inch less (e.g., about 2.5 to 3 inches). The spacing between the shells with a larger fractal shape (A**1**-A**4**) was about 2.4 cm while the spacing between shells of smaller fractal generator shapes (B**1**-B**3**) was about 2.15 cm (along a radial direction). In a preferred embodiment, shell A**4** was placed between shell B**2** and B**3** as shown. The resonators formed on each shell by the fractal shapes can be configured so as to be closely coupled (e.g., by capacitive coupling) and can serve to propagate a plasmonic wave.

It will be appreciated that while, two types of shells and a given number of shells per set are indicated in FIG. **1**, the number of shell types and number of shells for each set can be selected as desired, and may be optimized for different applications, e.g., wavelength/frequency bands.

FIG. **2** depicts a diagrammatic plan view of a cloaking system (or electrical resonator system) according to an alternate embodiment in which the individual shells have an elliptical cross section. As shown in FIG. **2**, a system **200** for cloaking can include a number of concentric shells (or bands) **202**. These shells can be held in place with respect to one another by suitable fixing means, e.g., they can be placed on a platform (parallel to the plane of the drawing) and/or held with a frame. The shells **202** can include a flexible substrate (e.g., polyimide with or without composite reinforcement) with a close-packed arrangement of electrically conductive material formed on the first surface. As stated previously for FIG. **1**, the closed-packed arrangement can include a number of self-similar electrical resonator shapes. The resonator shapes can be made from conductive traces (e.g., copper, silver, gold, silver-based ink, etc.) having a desired shape, e.g., fractal shape, split-ring shape, and the like. The shells **202** can surround an object to be cloaked, as indicated in FIG. **2**.

As indicated in FIG. **2** (by dashed lines **1** and **2** and arrows), the various shells themselves do not have to form closed surfaces. Rather, one or more shells can form open surfaces. This can allow for preferential cloaking of the object in one direction or over a given angle (solid angle). Moreover, while dashed lines **1** and **2** are shown intersecting shells B**1**-B**3** and A**1**-A**3** of system **200**, one or more shells of each group of shells (B**1**-B**3** and A**1**-A**3**) can be closed while others are open.

With further regard to FIGS. **1**-**2**, it should be appreciated that the cross-sections shown for each shell can represent closed geometric shapes, e.g., spherical and ellipsoidal shells.

As indicated previously, each shell of a cloaking system can include multiple resonators. The resonators can be repeated patterns of conductive traces. These conductive traces can be closed geometric shapes, e.g., rings, loops, closed fractals, etc. The resonator(s) can being self similar to at least second iteration. The resonators can include split-ring shapes, for some embodiments. The resonant structures are not required to be closed shapes, however, and open shapes can be used for such.

In exemplary embodiments, the closed loops can be configured as a fractals or fractal-based shapes, e.g., as depicted by **302** in FIG. **3** for an exemplary embodiment of a shell **300**, or **402** in FIG. **4**. The dimensions and type of fractal shape can be the same for each shell type but can vary between shell types. This variation (e.g., scaling of the same fractal shape) can afford increased bandwidth for the cloaking characteristics of the system (e.g., system **100** of FIG. **1**) This can lead to periodicity of the fractal shapes of common shell types but aperiodicity between the fractal shapes of different shell types.

Examples of suitable fractal shapes (for use for shells and/or a scatting object) can include, but are not limited to, fractal shapes described in one or more of the following patents, owned by the assignee of the present disclosure, the entire contents of all of which are incorporated herein by reference: U.S. Pat. Nos. 6,452,553; 6,104,349; 6,140,975; 7,145,513; 7,256,751; 6,127,977; 6,476,766; 7,019,695; 7,215,290; 6,445,352; 7,126,537; 7,190,318; 6,985,122; 7,345,642; and, U.S. Pat. No. 7,456,799.

Other suitable fractal shape for the resonant structures can include any of the following: a Koch fractal, a Minkowski fractal, a Cantor fractal, a torn square fractal, a Mandelbrot, a Caley tree fractal, a monkey's swing fractal, a Sierpinski gasket, and a Julia fractal, a contour set fractal, a Sierpinski triangle fractal, a Menger sponge fractal, a dragon curve fractal, a space-filling curve fractal, a Koch curve fractal, an Iypanov fractal, and a Kleinian group fractal.

FIG. **3** depicts an exemplary embodiment of a shell **300** (only a portion is shown) that includes repeated conductive traces that are configured in a fractal shape **302** (the individual closed traces). For the exemplary embodiment shown, each resonator shape **302** is about 1 cm on a side. Such resonator could, e.g., be used for the fractal shapes of shells B**1**-B**3** of FIG. **1**, in which case similar fractal shapes of larger size (e.g., about 1.5 cm on a side) could be used for shells A**1**-A**4**. The conductive trace is preferably made of copper. While exemplary fractal shapes are shown in FIG. **3**, the present disclosure is not limited to such and any other suitable fractal shapes (including generator motifs) may be used in accordance with the present disclosure.

It will be appreciated that the resonant structures of the shells may be formed or made by any suitable techniques and with any suitable materials. For example, semiconductors with desired doping levels and dopants may be used as conductive materials. Suitable metals or metal containing compounds may be used. Suitable techniques may be used to place conductors on/in a shell, including, but no limited to, printing techniques, photolithography techniques, etching techniques, and the like.

It will also be appreciated that the shells may be made of any suitable material(s). Printed circuit board materials may be used. Flexible circuit board materials are preferred. Other material may, however, be used for the shells and the shells themselves can be made of non-continuous elements, e.g., a frame or framework. For example, various plastics may be used.

FIG. **4** depicts a perspective view (photograph) of an exemplary embodiment of a cloak system **400** according to the present disclosure. As shown, the system includes a number of resonator shells **402** each having a close-packed arrangement of electrically conductive material (self-similar resonators **404**) formed on one surface. Two different shell configurations are shown, four larger shells and two smaller shells. The smaller shells included close-packed arrangements of resonator structures in which each resonator shape (as shown by **302** in FIG. **3**) was about 1 cm on a side. Similar fractal shapes of larger size (e.g., about 1.5 cm on a side) were used for the larger shells.

In FIG. **4**, a transmitting (source) antenna and a receiving antenna are shown as triangular shapes on the left and right, respectively (though functionally of each could of course be interchanged for the other). Twelve radially arrayed spacers are shown in FIG. **4**. The system **400** is shown supported on a Nalgene tank and Delrin platform and Delrin supports (radial supports) RF absorbers were placed in the immediate vicinity of the set up; further RF tripods (e.g., available from ETS) were used; all such materials were substantially transparent at the RF frequencies investigated/used. The cloak system **400** consists of six belts of fractal metamaterial (i.e., fractal-resonant structures shown in FIG. **3**) on flexible Taconic EF35 (low loss) circuit board. The belts are shown surround a scattering ring (object). The arrangement is supported by RF transparent plastics in a comb support. The entire system **400** was shown to be easily built up and broken down within a minute or two. The scale in FIG. **4** is about 0.7 meters across. The height of each shell can of course be selected as desired depending on the situation/application.

While embodiments are shown and described herein as having shells in the shape of concentric rings (circular cylinders), shells can take other shapes in other embodiments. For example, one or more shells could have a generally spherical shape (with minor deviations for structural support). In an exemplary embodiment, the shells could form a nested arrangement of such spherical shapes, around an object to be shielded (at the targeted/selected frequencies/wavelengths). Shell cross-sections of angular shapes, e.g., triangular, hexagonal, while not preferred, may be used.

One skilled in the art will appreciate that embodiments and/or portions of embodiments of the present disclosure can be implemented in/with computer-readable storage media (e.g., hardware, software, firmware, or any combinations of such), and can be distributed and/or practiced over one or more networks. Steps or operations (or portions of such) as described herein, including processing functions to derive, learn, or calculate formula and/or mathematical models utilized and/or produced by the embodiments of the present disclosure, can be processed by one or more suitable processors, e.g., central processing units ("CPUs") implementing suitable code/instructions in any suitable language (machine dependent on machine independent).

While certain embodiments and/or aspects have been described herein, it will be understood by one skilled in the art that the methods, systems, and apparatus of the present disclosure may be embodied in other specific forms without departing from the spirit thereof.

For example, while certain wavelengths/frequencies of operation have been described, these are merely representative and other wavelength/frequencies may be utilized or achieved within the scope of the present disclosure.

Furthermore, while certain preferred fractal generator shapes have been described others may be used within the scope of the present disclosure. Accordingly, the embodi-

ments described herein are to be considered in all respects as illustrative of the present disclosure and not restrictive.

What is claimed is:

1. An electrical resonator system, comprising:

a plurality of concentric electrical resonator shells, each shell including a substrate having first and second surfaces and a close-packed arrangement of electrically conductive material formed on the first surface, wherein the closed-packed arrangement comprises a plurality of self-similar electrical resonator shapes and is configured to operate at a desired passband of electromagnetic radiation;

wherein the close-packed arrangements of at least two of the electrical resonator shells are different in size and/or shape; and

wherein a resonator in the close-packed arrangement comprises a second order or higher fractal.

2. The system of claim **1**, wherein said passband is about 2:1.

3. The system of claim **2**, wherein said passband is about 3:1.

4. The system of claim **1**, wherein the electrical system is configured and arranged so that radiation incident on the system from a given direction has an intensity on a point-by-point basis such at each respective antipodal point, relative to an object placed at the center of the system, the radiation has the same or similar intensity.

5. The system of claim **1**, wherein the system is configured and arranged so that radiation incident on the system from a direction in cylindrical coordinates has the same or similar intensity at the antipodal point after having traversed the system.

6. The system of claim **1**, wherein the plurality of shells comprises a first pair of shells having similar closed-packed arrangements for operation at a first passband, wherein the two shells are positioned within ⅛λ of one another.

7. The system of claim **6**, wherein the plurality of shells comprises a second pair of shells having similar closed-packed arrangements for operation at a second frequency band, wherein the two shells are positioned within ⅛λ of one another.

8. The system of claim **1**, wherein the plurality of shells are hemispherical.

9. The system of claim **1**, wherein the plurality of shells are cylindrical.

10. The system of claim **1**, wherein the plurality of shells are spherical.

11. The system of claim **1**, wherein said fractal is selected from the group consisting of a Koch fractal, a Minkowski fractal, a Cantor fractal, a torn square fractal, a Mandelbrot, a Caley tree fractal, a monkey's swing fractal, a Sierpinski gasket, and a Julia fractal.

12. The system of claim **1**, wherein the fractal is selected from the group consisting of a contour set fractal, a Sierpinski triangle fractal, a Menger sponge fractal, a dragon curve fractal, a space-filling curve fractal, a Koch curve fractal, a Lypanov fractal, and a Kleinian group fractal.

13. The system of claim **1**, wherein the plurality of concentric electrical resonator shells are configured and arranged for operation at K band, Ka band, or X-band.

14. The system of claim **1**, wherein the system is operational over a bandwidth from about 500 MHz to about 1500 MHz.

* * * * *

US005590031A

United States Patent [19]

Mead, Jr. et al.

[11] **Patent Number: 5,590,031**

[45] **Date of Patent: Dec. 31, 1996**

[54] **SYSTEM FOR CONVERTING ELECTROMAGNETIC RADIATION ENERGY TO ELECTRICAL ENERGY**

[76] Inventors: **Franklin B. Mead, Jr.**, 44536 Avenida Del Sol, Lancaster, Calif. 93535; **Jack Nachamkin**, 12314 Teri Dr., Poway, Calif. 92064

[21] Appl. No.: **281,271**

[22] Filed: **Jul. 27, 1994**

[51] **Int. Cl.^{6}** **H02M 1/00**

[52] **U.S. Cl.** **363/8**; 363/178; 342/6

[58] **Field of Search** 363/8, 13, 178; 342/6, 61, 73, 173, 175

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,882,503	5/1975	Gamara	343/100 R
4,725,847	2/1988	Poirier	343/840
5,008,677	4/1991	Trigon et al.	342/17

Primary Examiner—Peter S. Wong
Assistant Examiner—Adolf Berhane
Attorney, Agent, or Firm—Chris Papageorge

[57] **ABSTRACT**

A system is disclosed for converting high frequency zero point electromagnetic radiation energy to electrical energy. The system includes a pair of dielectric structures which are positioned proximal to each other and which receive incident zero point electromagnetic radiation. The volumetric sizes of the structures are selected so that they resonate at a frequency of the incident radiation. The volumetric sizes of the structures are also slightly different so that the secondary radiation emitted therefrom at resonance interfere with each other producing a beat frequency radiation which is at a much lower frequency than that of the incident radiation and which is amenable to conversion to electrical energy. An antenna receives the beat frequency radiation. The beat frequency radiation from the antenna is transmitted to a converter via a conductor or waveguide and converted to electrical energy having a desired voltage and waveform.

14 Claims, 8 Drawing Sheets

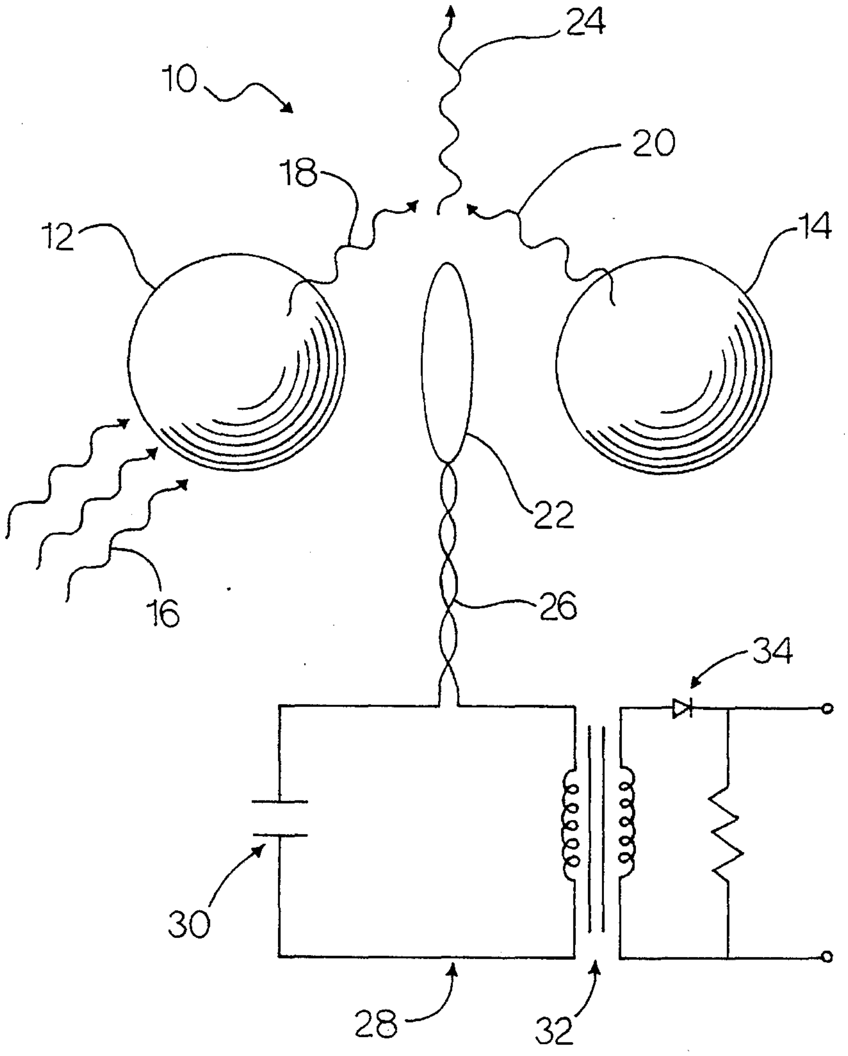

FIG. 1

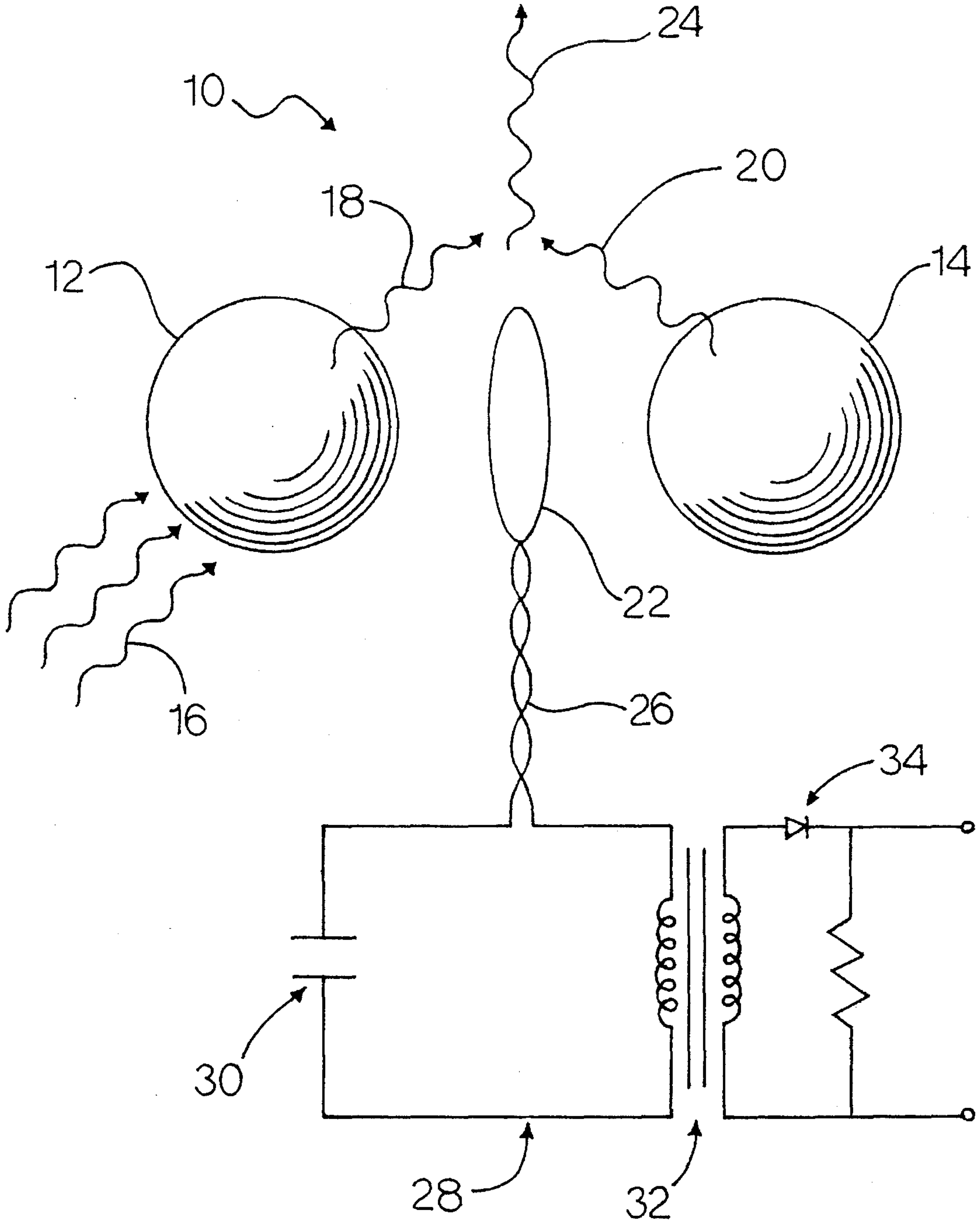
24
10
20
18
12
14
22
16
26
34
30
28
32

FIG. 2

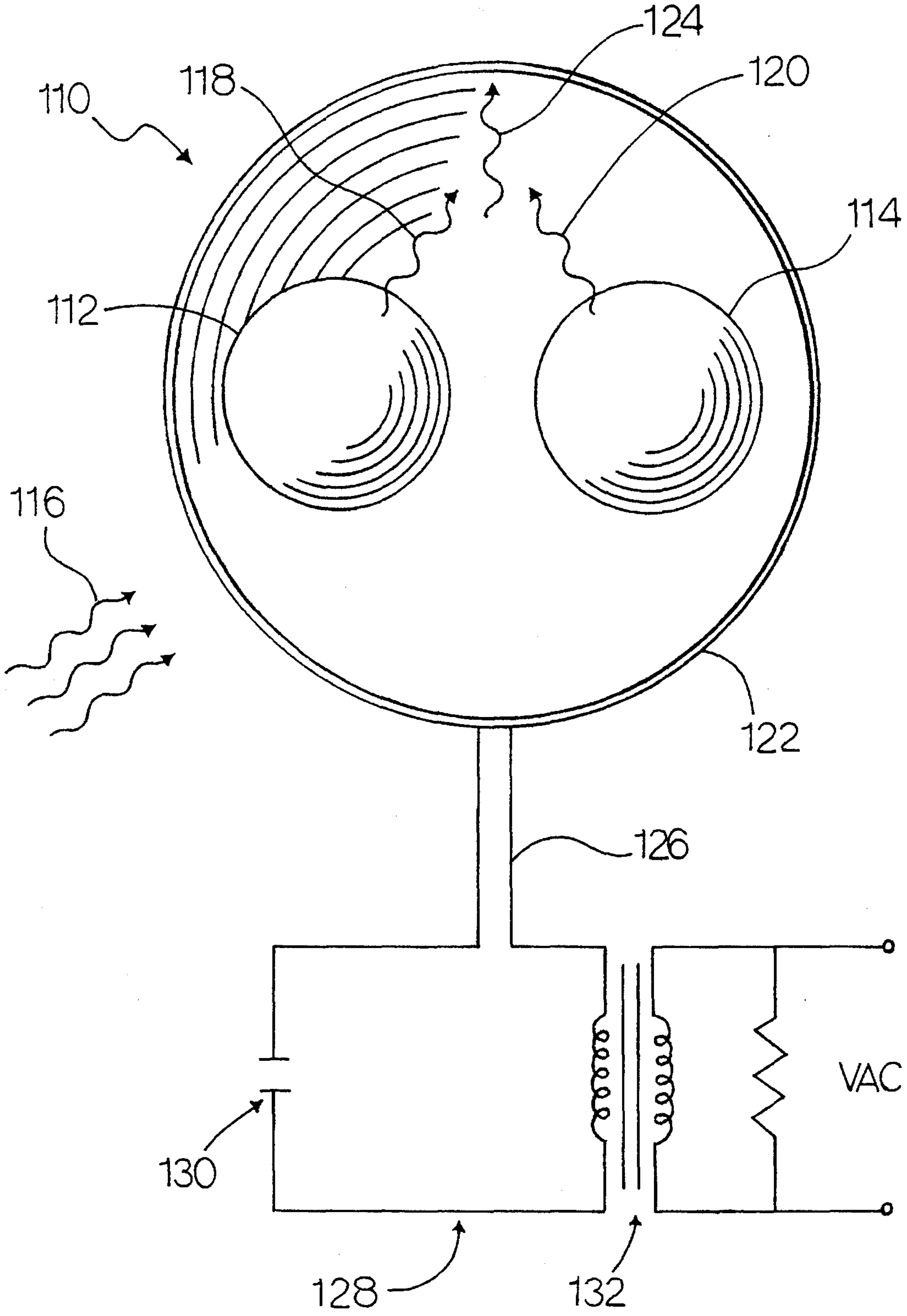
124
118
110
120
114
112
116
122
126
VAC
130
128
132

FIG. 3

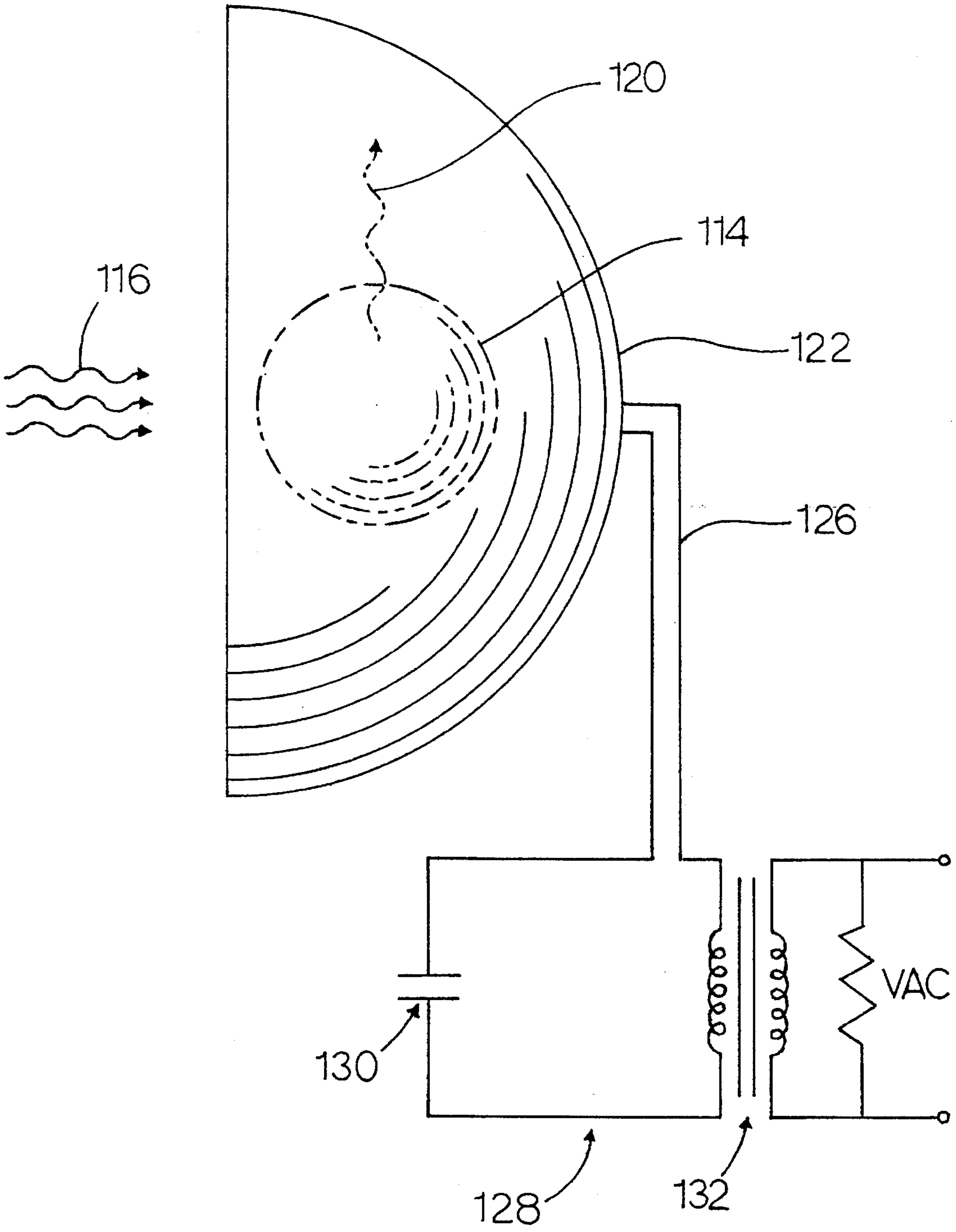
120
114
116
122
126
VAC
130
128
132

FIG. 4

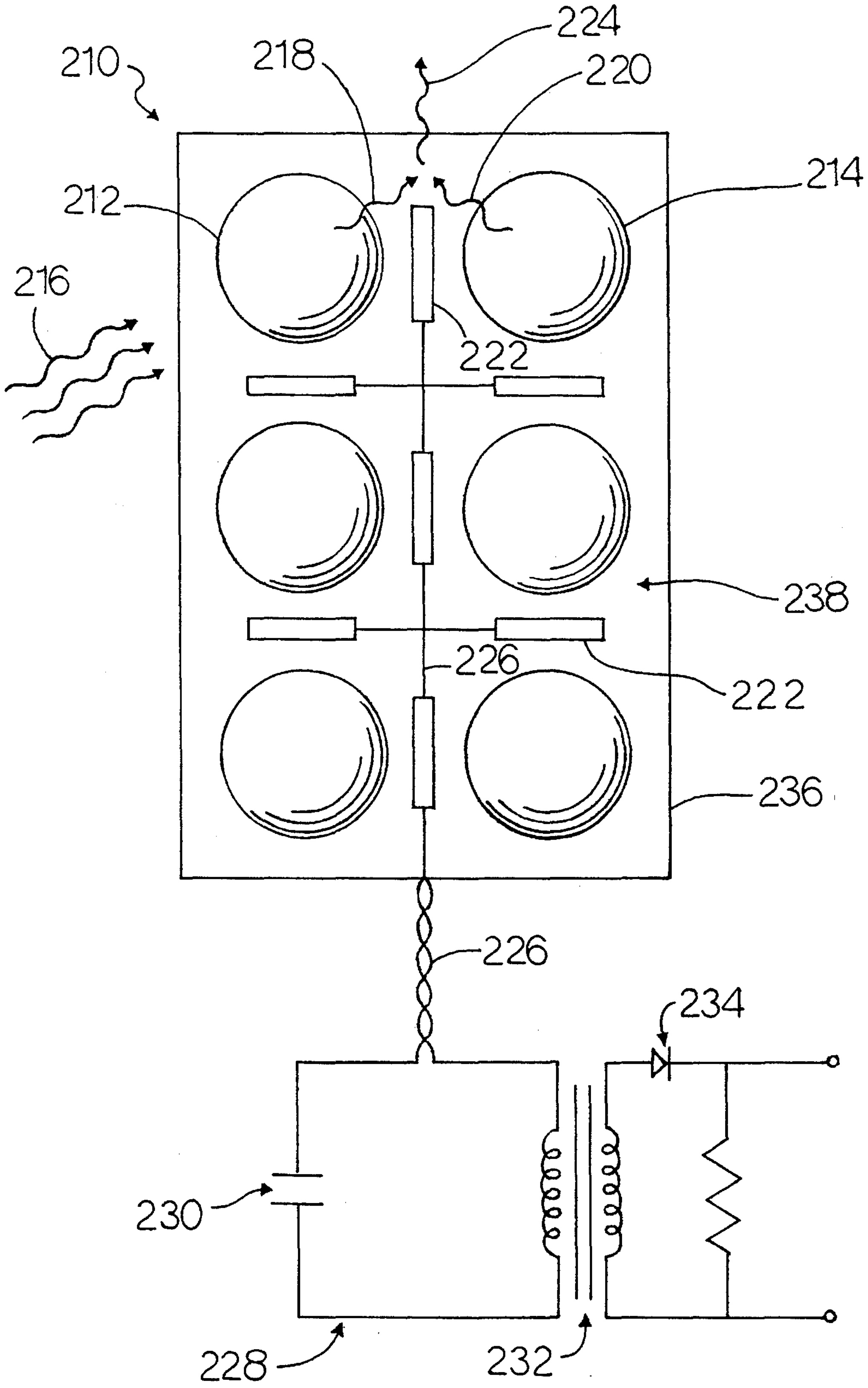

FIG. 5

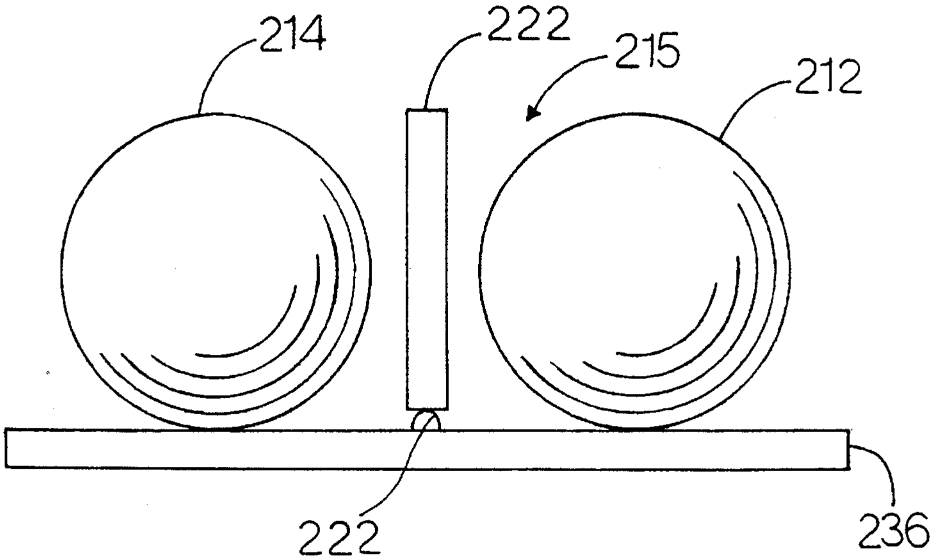
214
222
215
212
222
236

FIG. 6

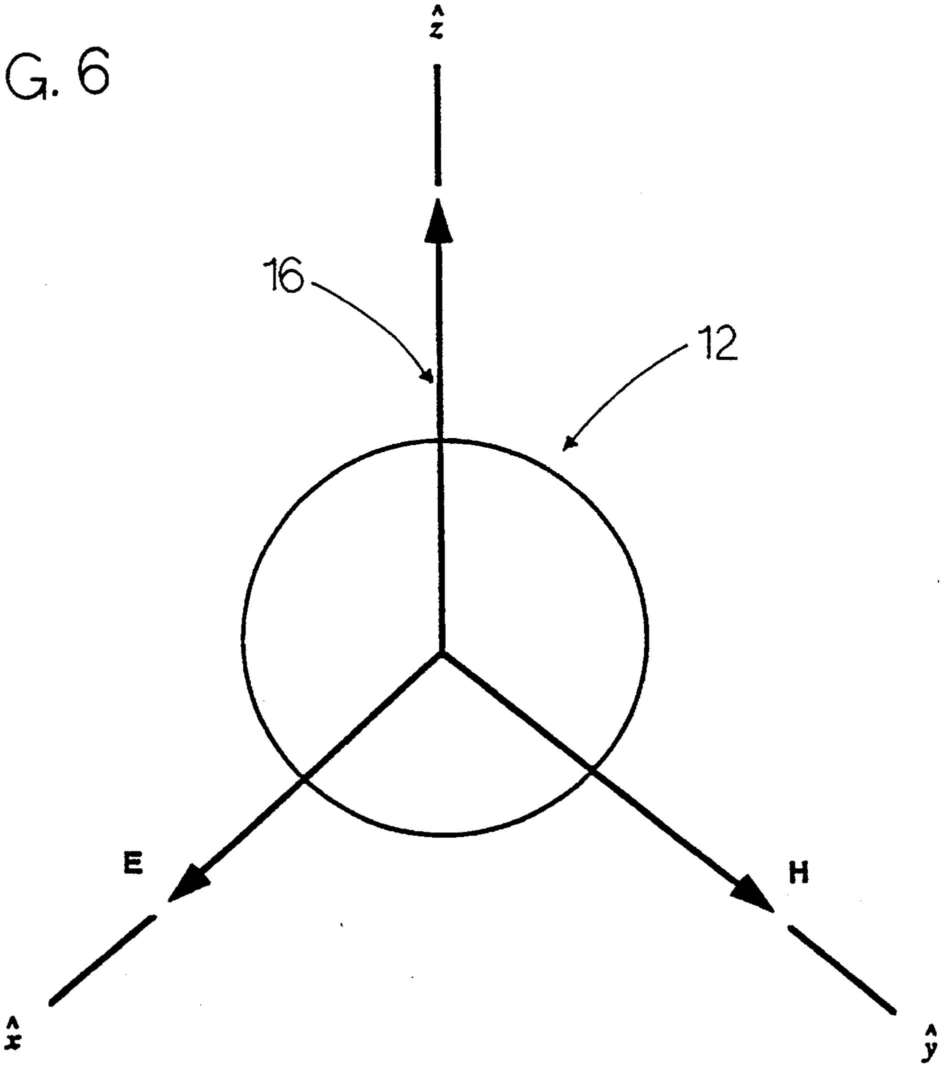
$\hat{z}$
16
12
E
H
$\hat{x}$
$\hat{y}$

FIG. 7

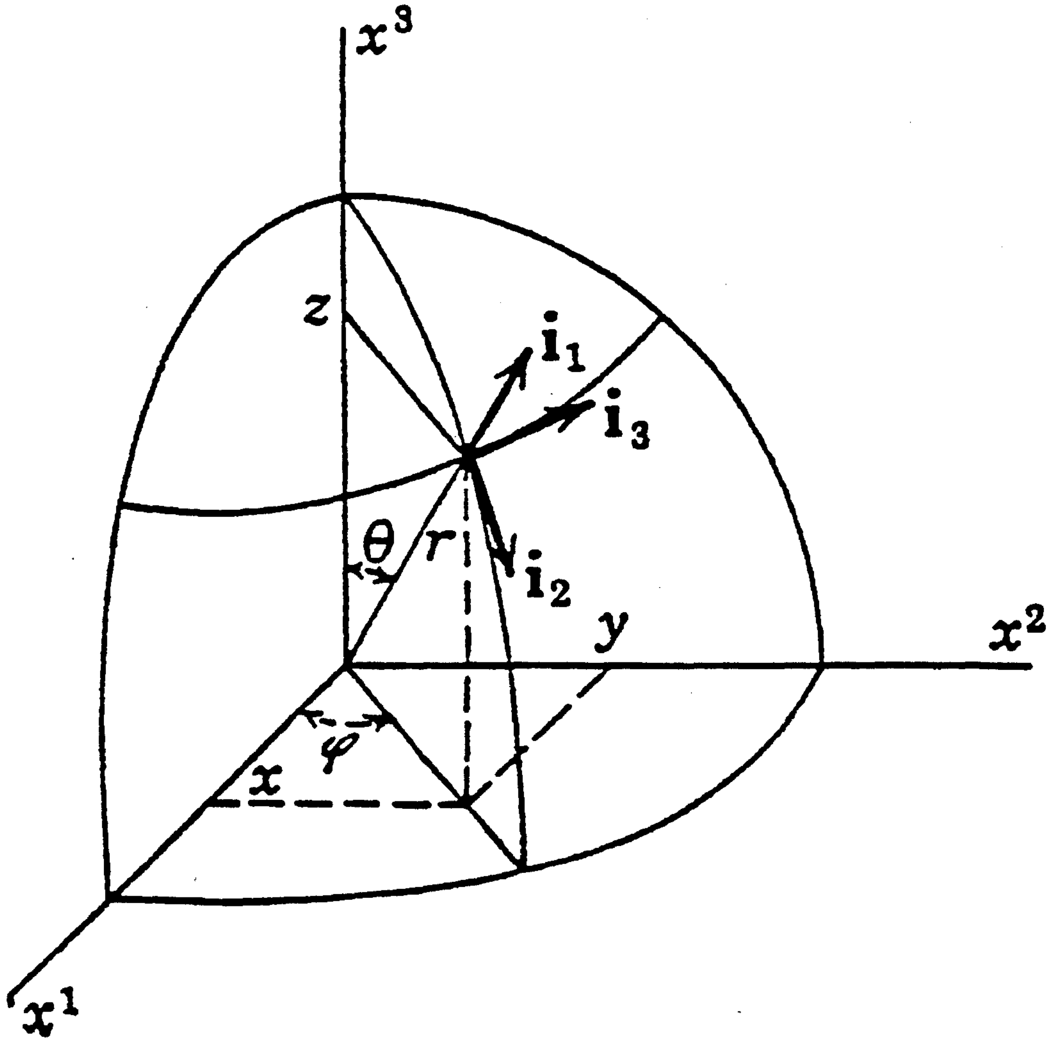
x^3
z
$\mathbf{i}_1$
$\mathbf{i}_3$
θ
r
$\mathbf{i}_2$
y
x^2
φ
x
x^1

FIG. 8

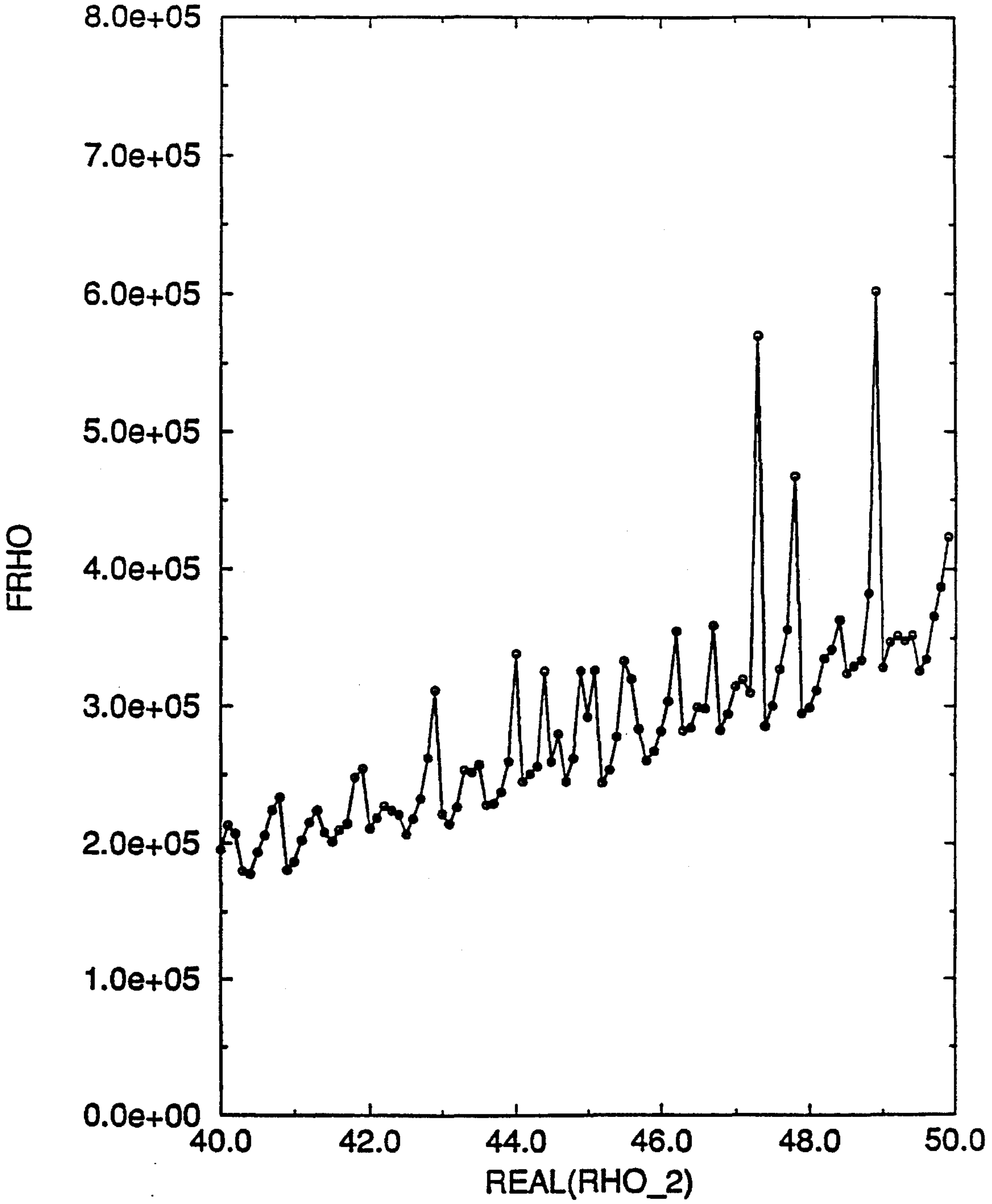
8.0e+05
7.0e+05
6.0e+05
5.0e+05
4.0e+05
3.0e+05
2.0e+05
1.0e+05
0.0e+00
FRHO
40.0
42.0
44.0
46.0
48.0
50.0
REAL(RHO_2)

FIG.9

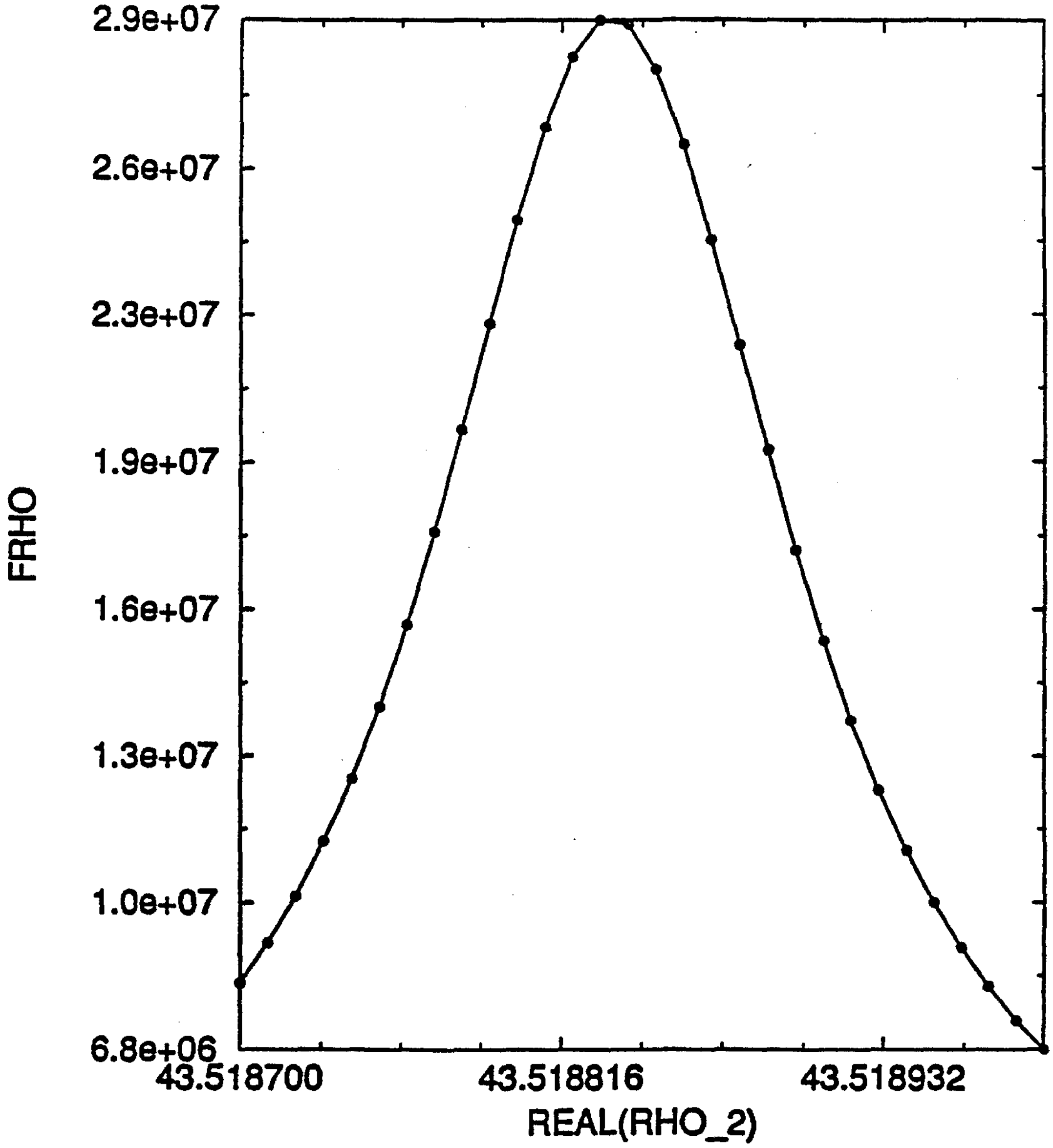
2.9e+07
2.6e+07
2.3e+07
1.9e+07
1.6e+07
1.3e+07
1.0e+07
6.8e+06
FRHO
43.518700
43.518816
43.518932
REAL(RHO_2)

SYSTEM FOR CONVERTING ELECTROMAGNETIC RADIATION ENERGY TO ELECTRICAL ENERGY

BACKGROUND OF THE INVENTION

The invention relates generally to conversion of electromagnetic radiation energy to electrical energy, and, more particularly, to conversion of high frequency bandwidths of the spectrum of a type of radiation known as zero point electromagnetic radiation to electrical energy.

The existence of zero point electromagnetic radiation was discovered in 1958 by the Dutch physicist M. J. Sparnaay. Mr. Sparnaay continued the experiments carried out by Hendrik B. G. Casimir in 1948 which showed the existence of a force between two uncharged parallel plates which arose from electromagnetic radiation surrounding the plates in a vacuum. Mr. Sparnaay discovered that the forces acting on the plates arose from not only thermal radiation but also from another type of radiation now known as classical electromagnetic zero point radiation. Mr. Sparnaay determined that not only did the zero point electromagnetic radiation exist in a vacuum but also that it persisted even at a temperature of absolute zero. Because it exists in a vacuum, zero point radiation is homogeneous and isotropic as well as ubiquitous. In addition, since zero point radiation is also invariant with respect to Lorentz transformation, the zero point radiation spectrum has the characteristic that the intensity of the radiation at any frequency is proportional to the cube of that frequency. Consequently, the intensity of the radiation increases without limit as the frequency increases resulting in an infinite energy density for the radiation spectrum. With the introduction of the zero point radiation into the classical electron theory, a vacuum at a temperature of absolute zero is no longer considered empty of all electromagnetic fields. Instead, the vacuum is now considered as filled with randomly fluctuating fields having the zero point radiation spectrum. The special characteristics of the zero point radiation which are that it has a virtually infinite energy density and that it is ubiquitous (even present in outer space) make it very desirable as an energy source. However, because high energy densities exist at very high radiation frequencies and because conventional methods are only able to convert or extract energy effectively or efficiently only at lower frequencies at which zero point radiation has relatively low energy densities, effectively tapping this energy source has been believed to be unavailable using conventional techniques for converting electromagnetic energy to electrical or other forms of easily useable energy. Consequently, zero point electromagnetic radiation energy which may potentially be used to power interplanetary craft as well as provide for society's other needs has remained unharnessed.

There are many types of prior art systems which use a plurality of antennas to receive electromagnetic radiation and provide an electrical output therefrom. An example of such a prior art system is disclosed in U.S. Pat. No. 3,882,503 to Gamara. The Gamara system has two antenna structures which work in tandem and which oscillate by means of a motor operatively attached thereto in order to modulate the radiation reflected from the antenna surfaces. The reflecting surfaces of the antennas are also separated by a distance equal to a quarter wavelength of the incident radiation. However, the Gamara system does not convert the incident radiation to electrical current for the purpose of converting the incident electromagnetic radiation to another form of readily useable energy. In addition, the relatively large size of the Gamara system components make it unable to resonate at and modulate very high frequency radiation.

What is therefore needed is a system which is capable of converting high frequency electromagnetic radiation energy into another form of energy which can be more readily used to provide power for transportation, heating, cooling as well as various other needs of society. What is also needed is such a system which may be used to provide energy from any location on earth or in space.

SUMMARY OF THE INVENTION

It is a principle object of the present invention to provide a system for converting electromagnetic radiation energy to electrical energy.

It is another object of the present invention to provide a system for converting electromagnetic radiation energy having a high frequency to electrical energy.

It is another object of the present invention to provide a system for converting zero point electromagnetic radiation energy to electrical energy.

It is another object of the present invention to provide a system for converting electromagnetic radiation energy to electrical energy which may used to provide such energy from any desired location on earth or in space.

It is another object of the present invention to provide a system for converting electromagnetic radiation energy to electrical energy having a desired waveform and voltage.

It is an object of the present invention to provide a miniaturized system for converting electromagnetic radiation energy to electrical energy in order to enhance effective utilization of high energy densities of the electromagnetic radiation.

It is an object of the present invention to provide a system for converting electromagnetic radiation energy to electrical energy which is simple in construction for cost effectiveness and reliability of operation.

Essentially, the system of the present invention utilizes a pair of structures for receiving incident electromagnetic radiation which may be propagating through a vacuum or any other medium in which the receiving structures may be suitably located. The system of the present invention is specifically designed to convert the energy of zero point electromagnetic radiation; however, it may also be used to convert the energy of other types of electromagnetic radiation. The receiving structures are preferably composed of dielectric material in order to diffract and scatter the incident electromagnetic radiation. In addition, the receiving structures are of a volumetric size selected to enable the structures to resonate at a high frequency of the incident electromagnetic radiation based on the parameters of frequency of the incident radiation and propagation characteristics of the medium and of the receiving structures. Since zero point radiation has the characteristic that its energy density increases as its frequency increases, greater amounts of electromagnetic energy are available at higher frequencies. Consequently, the size of the structures are preferably miniaturized in order to produce greater amounts of energy from a system located within a space or area of a given size. In this regard, the smaller the size of the receiving structures, the greater the amount of energy that can be produced by the system of the present invention.

At resonance, electromagnetically induced material deformations of the receiving structures produce secondary fields of electromagnetic energy therefrom which may have eva-

nescent energy densities several times that of the incident radiation. The structures are of different sizes so that the secondary fields arising therefrom are of different frequencies. The difference in volumetric size is very small so that interference between the two emitted radiation fields, and the receiving structures at the two different frequencies produces a beat frequency radiation which has a much lower frequency than the incident radiation. The beat frequency radiation preferably is at a frequency which is sufficiently low that it may be relatively easily converted to useable electrical energy. In contrast, the incident zero point radiation has its desirable high energy densities at frequencies which are so high that conventional systems for converting the radiation to electrical energy either cannot effectively or efficiently so convert the radiation energy or simply cannot be used to convert the radiation energy for other reasons.

The system of the present invention also includes an antenna which receives the beat frequency radiation. The antenna may be a conventional metallic antenna such as a loop or dipole type of antenna or a rf cavity structure which partially encloses the receiving structures. The antenna feeds the radiation energy to an electrical conductor (in the case of a conventional dipole or comparable type of antenna) or to a waveguide (in the case of a rf cavity structure). The conductor or waveguide feeds the electrical current (in the case of the electrical conductor) or the electromagnetic radiation (in the case of the waveguide) to a converter which converts the received energy to useful electrical energy. The converter preferably includes a tuning circuit or comparable device so that it can effectively receive the beat frequency radiation. The converter may include a transformer to convert the energy to electrical current having a desired voltage. In addition, the converter may also include a rectifier to convert the energy to electrical current having a desired waveform.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a plan view of the receiving structures and antenna of a first embodiment of the system of the present invention with a schematic view of the conductor and converter thereof and also showing the incident primary and emitted secondary electromagnetic radiation.

FIG. **2** is a front view of the receiving structures, antenna and waveguide of a second embodiment of the system of the present invention with a schematic view of the converter thereof and also showing the incident primary and emitted secondary electromagnetic radiation.

FIG. **3** is a perspective view of the receiving structures, antenna and waveguide of the second embodiment shown in FIG. **2** with a schematic view of the converter thereof and also showing the incident primary and emitted secondary electromagnetic radiation.

FIG. **4** is a front view of the substrate and a plurality of pairs of the receiving structures and a plurality of antennas of a third embodiment of the system of the present invention with a schematic view of the conductor and converter thereof and also showing the incident primary and emitted secondary electromagnetic radiation.

FIG. **5** is a top view of some of the components of the third embodiment of the system of the present invention showing two of the plurality of pairs of receiving structures and two of the plurality of antennas mounted on the substrate.

FIG. **6** is a diagram of a receiving structure of the system of the present invention showing an incident electromagnetic plane wave impinging on the receiving structure and illustrating the directions of the electric and magnetic field vectors thereof.

FIG. **7** is a diagram of a spherical coordinate system as used in the formulas utilized in the system of the present invention.

FIG. **8** is a graph showing an imaginary ρ parameter plotted against a real ρ parameter illustrating the values thereof at resonance as well as values thereof at other than resonance.

FIG. **9** is a graph showing a portion of the graphical representation shown in FIG. **8** illustrating the real and imaginary ρ values at or near a single resonance.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring to the drawings, a first embodiment of the present invention is generally designated by the numeral **10**. The system **10** includes a first and second means for receiving **12** and **14** incident electromagnetic radiation **16**. The means for receiving **12** and **14** are preferably a pair of spherical structures **12** and **14** which are preferably composed of a dielectric material. Alternatively, the spheres **12** and **14** may be cubical structures or any other suitable shape. The spheres **12** and **14** may be mounted on a suitable foundation by any suitable mounting means (not shown), or spheres **12** and **14** may be suspended from a suitable foundation by any suitable suspension means (not shown). The spheres **12** and **14** are preferably composed of a dielectric material. The dielectric spheres **12** and **14** scatter and concentrate electromagnetic waves. At very sharpely defined frequencies, the spheres **12** and **14** will have resonances wherein the internal energy densities can be five orders of mangitude larger than the energy density of the incident electromagnetic field driving the spheres **12** and **14**. At resonance, the electromagnetic stresses, equivalent to pressures proportional to the energy density, can cause material deformation of the spheres **12** and **14** which produce a secondary electromagnetic field. The spheres **12** and **14** are preferably positioned proximal to each other, as shown in FIG. **1**. Although the proximity of the spheres to each other will adversely affect the resonances, the very high "Q"s of the isolated-sphere resonances results in such adverse affect being relatively small. However, the proximity of the spheres **12** and **14** allows the spheres to interact electromechanically which increases the magnitude of the secondary radiation emitted therefrom.

The electromagnetic radiation incident upon the spheres **12** and **14** which drives the spheres to resonance is preferably zero point radiation **16**. However, other types of electromagnetic radiation may also be used to drive the spheres **12** and **14**, if desired.

The effect of a dielectric sphere such as **12** or **14** on an incident electromagnetic radiation such as a plane wave thereof is shown in FIG. **6**. The plane wave propagates in the z axis direction and is diffracted by the sphere **12** resulting in scattering thereof. This scattering is commonly known as Mie scattering. The incident radiation wave has an electric vector component which is linearly polarized in the x axis direction and a magnetic vector component which is linearly polarized in the y axis direction.

An electromagnetic wave incident upon a structure produces a forced oscillation of free and bound charges in synch with the primary electromagnetic field of the incident electromagnetic wave. The movements of the charges produce a

secondary electromagnetic field both inside and outside the structure. The secondary electromagnetic radiation comprising this secondary electromagnetic field is shown in FIG. **1** and designated by the numerals **18** and **20**. An antenna which is shown simply as a loop antenna but may also be a dipole or any other suitable type of antenna is also shown in FIG. **1** and designated by the numeral **22**. The nonlinear mutual interactions of the spheres produces interference between the secondary electromagnetic radiation **18** and **20** produces a beat frequency radiation **24** which is preferably at a much lower frequency than the primary radiation **16**. It is this beat frequency radiation **24** which is desired for conversion into electrical energy because it preferably is within the frequency range of rf radiation which may be converted into electrical energy by generally conventional systems. Thus, the radiation **24** received by the antenna **22** is fed via an electrical conductor **26** to a means for converting the beat frequency radiation **24** to electrical energy. This means for converting is designated by the numeral **28** and preferably includes a tuning capacitor **30** and a transformer **32** and a rectifier (preferably a diode) **34**. Instead of including the capacitor **30**, transformer **32** and rectifier **34**, the converter **28** may alternatively include an rf receiver of any suitable type.

The resultant field at any point is the vector sum of the primary and secondary fields. For the equations that follow, the structure receiving the incident plane wave is a sphere of radius a having a propagation constant k_1 positioned in an infinite, homogeneous medium having a propagation constant k_2. The incident plane wave propagates in the z axis direction and is as shown in FIG. **6**. The spherical coordinate system used for the vector spherical wave functions is shown in FIG. **7**. Expansion of the incident field provides:

$$E_i = E_o e^{-i\omega t} \sum_{n=1}^{\infty} i^n \frac{2n+1}{n(n+1)} \left(m^{(1)}_{01n} - i n^{(1)}_{eln}\right)$$

$$H_i = -\frac{k_2}{\omega\mu_2} E_o e^{-i\omega t} \sum_{n=1}^{\infty} i^n \frac{2n+1}{n(n+1)} \left(m^{(1)}_{eln} + i n^{(1)}_{01n}\right)$$

where E is the electric field and H is the magnetic field; and

$$m^{(1)}_{\substack{01n\\e1n}} = \pm \frac{1}{\sin\theta} j_n(k_2R) P_n^{1}(\cos\theta) \substack{\cos\\\sin} \phi i_2 - j_n(k_2R) \frac{\partial p_n^{1}}{\partial\theta} \substack{\sin\\\cos} \phi i_3$$

$$n^{(1)}_{\substack{0ln\\e1n}} = \frac{n(n+1)}{k_2R} j_n(k_2R) P_n^{1}(\cos\theta) \substack{\sin\\\cos} \phi i_1 + \frac{1}{k_2R} [k_2R j_n(k_2R)]' \times$$

$$\frac{\partial P_n^{1}}{\partial\theta} \substack{\sin\\\cos} \phi i_2 \pm \frac{1}{k_2R\sin\theta} [k_2R j_n(k_2R)]' P_n^{1}(\cos\theta) \substack{\cos\\\sin} \phi i_3.$$

The electric and magnetic fields of the incident wave transmitted into the sphere i.e., R<a, can be similarly expanded:

$$E_t = E_0 e^{-i\omega t} \sum_{n=1}^{\infty} i^n \frac{2n+1}{n(n+1)} \left(a_n^{t} m^{(1)}_{01n} - i b_n^{t} n^{(1)}_{e1n} \right)$$

$$H_t = \frac{k_2}{\phi\mu_1} E_0 e^{-i\omega t} \sum_{n=1}^{\infty} i^m \frac{2n+1}{n(n+1)} \left(b_n^{t} m^{(1)}_{e1n} - i a_n^{t} n^{(1)}_{01n} \right)$$

If $j_n(k_2R)$ is replaced by $h_n^{(1)}(k_2R)$ in the previous equations, the functions $m^{(1)}$ and $n^{(1)}$ become $m^{(3)}$ and $n^{(3)}$. The outgoing fields i.e., R>a, are represented by:

$$E_R = E_0 e^{-i\omega t} \sum_{n=1}^{\infty} i^n \frac{2n+1}{n(n+1)} \left(a_n^{r} m^{(3)}_{01n} - i b_n^{r} n^{(3)}_{e1n} \right)$$

$$H_r = \frac{k_2}{\phi\mu_1} E_0 e^{-i\omega t} \sum_{n=1}^{\infty} i^m \frac{2n+1}{n(n+1)} \left(b_n^{r} m^{(3)}_{e1n} - i a_n^{r} n^{(3)}_{01n} \right)$$

where H_r represents the resultant wave in the medium surrounding the sphere. At resonance, the values of ρ at resonance require that the a_n^t and b_n^t coefficients be infinite. In order to determine these values of a_n^t and b_n^t, the boundary conditions at the sphere radius are needed. Since there must be continuity of the E and H values at the surface, the following equations are used:

$$i_1 \times (E_i + E_r) = i_1 \times E_t \text{ and}$$

$$i_1 \times (H_i + H_r) = i_1 \times H_t$$

which lead to two pairs of inhomogeneous equations:

$$a_n^t j_n(N\rho) - a_n^r h_n^{(1)}(\rho) = j_n(\rho)$$

$$\mu_2 a_n^t [N\rho j_n(N\rho)]' - \mu_1 a_n^r [\rho h_n^{(1)}(\rho)]' = \mu_1 [\rho j_n(\rho)]' \text{ and}$$

$$\mu_2 N b_n^t j_n(N\rho) - \mu_1 b_n^r h_n^{(1)}(\rho) = \mu_1 j_n(\rho)$$

$$b_n^t [N\rho j_n(N\rho)]' - N b_n^r [\rho h_n^{(1)}(\rho)]' = N[\rho j_n(\rho)]'$$

where $k_1=Nk_2$, $\rho=k_2a$, $k_1a=N\rho$. Spherical Bessel functions of the first kind are denoted by j_n, while those of the third kind are denoted by $h_n^{(1)}$. The resulting equations are:

$$a_n^t = \frac{\mu_1 j_n(\rho)[\rho h_n^{(1)}(\rho)]' - \mu_1 h_n^{(1)}\rho)[\rho j_n(\rho)]'}{\mu_1 j_n(N\rho)[\rho h_n^{(1)}(\rho)]' - \mu_2 h_n^{(1)}(\rho)[N\rho j_n(N\rho)]'}$$

and

$$b_n^t = \frac{\mu_1 N j_n(\rho)[\rho h_n^{(1)}(\rho)]' - \mu_1 N h_n^{(1)}\rho)[\rho j_n(\rho)]'}{\mu_2 N^2 j_n(N\rho)[\rho h_n^{(1)}(\rho)]' - \mu_1 h_n^{(1)}(\rho)[N\rho j_n(N\rho)]'}$$

At a resonance, the denominator of either a_n^t or b_n^t will be zero. Thus, ρ values are found using the above equations that correspond to a resonant combination of angular frequency (ω) and radius (a) for a given sphere material and given surrounding medium. In determining such values of ρ, the following equations are also specifically used:

$$\rho = ak_2 = a\omega\sqrt{\epsilon_2\mu_2} \text{ and}$$

$$\rho_1 = (k_1/k_2)\rho$$

where ρ_1 corresponds to the sphere material. An iterative method is preferably used to find the desired values of ρ at resonance. In calculating ρ utilizing the above equations for purposes of example, it was assumed that $\mu_1=\mu_2=\mu_0=4\pi\times 10^{-7}$ and $\epsilon_2=\epsilon_0=8.85419\times10^{-12}$.

One major root of ρ which was found has a value of:

Real (ρ)=+66.39752607619131

Imaginary (ρ)=−0.6347867071968998.

These particular values are not shown in FIG. **8**. However, other values of ρ found using the equations set forth herein are shown in FIG. **8**. The peaks in FIG. **8** are the resonances. One of these resonances shown in FIG. **8** is shown in detail in FIG. **9**. These resonance values are shown for purposes of example. Other resonances also exist which have not been determined; thus, not all possible resonance values are shown in FIGS. **8** and **9**.

Calculation of these values also allows the determination of a possible am combination which would have these root values. For ρ, ϵ (epsilon)=ϵ_0 and $\mu=\mu_0$, and

$$\rho = a\omega\sqrt{\epsilon_0\mu_0} = a\omega/c.$$

Expressed in SI units, the speed of light $c=2.99792458\times10^{14}$ m/s. If an a value of 10^{-6} m is assumed for the examples shown herein, then:

$$\omega=\rho c/a\approx1.9919\times10^{16}-i1.9044\times10^{14}\text{ radians/s.}$$

This is an example of the angular frequency required within the impingent EM radiation in order to create a resonant situation. Examples of other resonances were indicated, and these are shown in FIG. **8**. No complex-frequency plane waves exist. Therefore, the calculations were made by considering only the real portion of the above root and setting the imaginary portion equal to zero. However, upon doing this, the iterative calculation procedure becomes insensitive to any root in the vicinity of the root's real portion. In the iterative calculation procedure, initially a range of ρ values is input into the equations. These ρ values are in the neighborhood of the prospective root. A range of ρ values is subsequently studied to find any imaginary ρ i.e., fρ (a function of ρ), peaks in that range. Next, once a peak has been chosen, the function order n giving the dominant fρ is determined. This also gives a clue as to whether the peak is due to a magnetic resonance (a_n approaches infinity) or an electrical resonance (b_n approaches infinity). A large number of Newton-Raphson iterations is preferably performed in order to converge upon a root ρ value.

FIGS. **2** and **3** show a second embodiment of the present invention generally designated by the numeral **110**. Embodiment **110** is essentially the same as embodiment **10** except that the antenna is a rf cavity structure **122** which feeds the received beat frequency radiation **124** to a waveguide **126**. Embodiment **110** also preferably includes two spheres **112** and **114** which receive the primary incident electromagnetic radiation **116** and emit the secondary electromagnetic radiation **118** and **120**. As with the spheres **18** and **20** of embodiment **10**, spheres **118** and **120** are preferably composed of a dielectric material. Embodiment **110** also includes converter **128**, capacitor **130**, transformer **132** and rectifier **134** which are essentially identical to the correspondingly numbered elements of embodiment **10**. Therefore, a description of these components of embodiment **110** will not be repeated in order to promote brevity. In addition, the same equations and method of calculation set forth above with regard to embodiment **10** also apply to embodiment. Therefore, their description will not be repeated in order to promote brevity.

FIGS. **4** and **5** show a third embodiment of the present invention generally designated by numeral **210**. Embodiment **210** is essentially identical to the first embodiment **10** except that the embodiment **210** includes a plurality of pairs **215** of receiving means (spheres) **212** and **214** mounted on a substrate **236**. The spheres **212** and **214** are thus in the form of an array **238**. The pairs **215** of the array **238** are preferably positioned proximal to each other in order to maximize the amount of energy extracted from a particular area or space of a given size. Since, as set forth hereinabove, the energy density of the zero point radiation increases as the frequency of the radiation increases, it is desirable that the spheres resonate at as high a bandwidth of frequencies as possible. Because the spheres **212** and **214** must be small in direct proportion to the wavelength of the high frequencies of the incident electromagnetic radiation **216** at which resonance is desirably obtained, the spheres **212** and **214** are preferably microscopic in size. Current lithographic techniques are capable of manufacturing such microscopically small spheres mounted on a suitable substrate thereby providing a suitably miniaturized system **210**. A miniaturized system enhances the energy output capability of the system by enabling it to resonate at higher frequencies at which there are correspondingly higher energy densities. Consequently, utilization of array **238** in the system **210** enhances the maximum amount of electrical energy provided by the system **210**.

Lithographic techniques may be more amenable to manufacturing microscopically small receiving structures **212** and **214** which may be disc shaped, semispherical or have another shape other than as shown in FIGS. **4** and **5**. Consequently, the receiving means **212** and **214** may accordingly have such alternative shapes rather than the spherical shape shown in FIGS. **4** and **5**. In addition, a large number of small spheres may be manufactured by bulk chemical reactions. Packing a volume with such spheres in close proximity could enhance the output of energy.

Embodiment **210** also includes a plurality of antennas **222** positioned preferably between the spheres **212** and **214** which receive the beat frequency radiation **224** produced by the interference between the secondary radiation **218** and **220**. The antennas **222** are shown as loop antennas **222** but may be any other suitable type of antennas as well.

Embodiment **210** has a plurality of electrical conductors **226** which preferably include traces mounted on the substrate **236** which occupies a finite volume. The electrical conductors **226** feed the electrical output from the antennas **222** to a suitable converter **228** which preferably includes tuning capacitor **230**, transformer **232** and rectifier **234**, as with embodiments **10** and **110**. Except as set forth above, the components of embodiment **210** are identical to embodiment **10** so the detailed description of these components will not be repeated in order to promote brevity. In addition, the same equations and method of calculation set forth above for embodiment **10** also apply to embodiment **210**. Therefore, the description of these equations and method of calculation will not be repeated in order to promote brevity.

Accordingly, there has been provided, in accordance with the invention, a system which converts high frequency zero point electromagnetic radiation into electrical energy effectively and efficiently and thus fully satisfies the objectives set forth above. It is to be understood that all terms used herein are descriptive rather than limiting. Although the invention has been specifically described with regard to the specific embodiments set forth herein, many alternative embodiments, modifications and variations will be apparent to those skilled in the art in light of the disclosure set forth herein. Accordingly, it is intended to include all such alternatives, embodiments, modifications and variations that fall within the spirit and scope of the invention as set forth in the claims hereinbelow.

What is claimed is:

1. A system for converting incident electromagnetic radiation energy to electrical energy, comprising:

a first means for receiving incident primary electromagnetic radiation, said means for receiving producing emitted secondary electromagnetic radiation at a first frequency, said first means for receiving having a first volumetric size selected to resonate at a frequency within the frequency spectrum of the incident primary electromagnetic radiation in order to produce the secondary electromagnetic radiation at the first frequency at an enhanced energy density;

a second means for receiving the incident primary electromagnetic radiation, said means for receiving producing emitted secondary electromagnetic radiation at a second frequency, the secondary radiation at the first frequency and the secondary radiation at the second frequency interfering to produce secondary radiation at a lower frequency than that of the incident primary radiation, said second means for receiving having a second volumetric size selected to resonate at a frequency within the frequency spectrum of the incident primary electromagnetic radiation in order to produce the emitted secondary electromagnetic radiation at the second frequency at an enhanced energy density;

an antenna for receiving the emitted secondary electromagnetic radiation at the lower frequency, said antenna providing an electrical output responsive to the secondary electromagnetic radiation received;

a converter electrically connected to said antenna for receiving electrical current output from said antenna and converting the electrical current output to electrical current having a desired voltage and waveform.

2. The system of claim **1** wherein:

said first means for receiving is composed of a dielectric material; and

said second means for receiving is composed of a dielectric material.

3. The system of claim **1** wherein:

said first means for receiving is spherical; and

said second means for receiving is spherical.

4. A system for for converting incident zero point electromagnetic radiation energy to electrical energy, comprising:

a first means for receiving incident primary zero point electromagnetic radiation, said means for receiving producing emitted secondary electromagnetic radiation at a first frequency;

a second means for receiving the incident primary zero point electromagnetic radiation, said means for receiving producing emitted secondary electromagnetic radiation at a second frequency, the secondary radiation at the first frequency and the secondary radiation at the second frequency interfering to produce secondary radiation at a beat frequency which is lower than that of the incident primary radiation;

an antenna for receiving the emitted secondary electromagnetic radiation at the lower frequency, said antenna providing an electrical output responsive to the secondary electromagnetic radiation received;

means for transmitting the emitted secondary electromagnetic radiation at the beat frequency from said antenna, said means for transmitting connected to said antenna;

a converter connected to said means for transmitting for receiving the emitted secondary electromagnetic radiation at the beat frequency from said antenna and converting the same to electrical current having a desired voltage and waveform.

5. The system of claim **4** wherein:

said first means for receiving has a first volumetric spherical size selected to resonate in response to the incident primary electromagnetic radiation in order to produce the secondary electromagnetic radiation at the first frequency at an enhanced energy density; and

said second means for receiving has a second volumetric spherical size selected to resonate in response to the incident primary electromagnetic radiation in order to produce the emitted secondary electromagnetic radiation at the second frequency at an enhanced energy density, said first and second volumetric sizes selected based on parameters of propagation constant of said first and second means for receiving, propagation constant of medium in which said first and second means for receiving are located and frequency of the incident primary electromagnetic radiation.

6. The system of claim **5** wherein the first and second volumetric sizes are selected by utilizing the formulas:

$$a_n{}^{t}=\frac{\mu_1 j_n(\rho)[\rho h_n^{(1)}(\rho)]'-\mu_1 h_n^{(1)}\rho)[\rho j_n(\rho)]'}{\mu_1 j_n(N\rho)[\rho h_n^{(1)}(\rho)]'-\mu_2 h_n^{(1)}(\rho)[N\rho j_n(N\rho)]'}$$

$$b_n{}^{t}=\frac{\mu_1 N j_n(\rho)[\rho h_n^{(1)}(\rho)]'-\mu_1 N h_n^{(1)}\rho)[\rho j_n(\rho)]'}{\mu_2 N^2 j_n(N\rho)[\rho h_n^{(1)}(\rho)]'-\mu_1 h_n^{(1)}(\rho)[N\rho j_n(N\rho)]'}$$

$$\rho=a\omega\sqrt{\epsilon_2\mu_2}$$

wherein at a resonance, the denominator of either equation for $a_n{}^{t}$ or $b_n{}^{t}$ will be approximately zero and wherein k_1=propagation constant of the means for receiving, k_2=propagation constant of medium through which the incident electromagnetic radiation propagates, a is the radius of either means for receiving, $N=k_1/k_2$, $\rho=k_2a$, $k_1a=N\rho$, $a_n{}^{t}$=magnitude of oscillations of the electric field of the nth order, $b_n{}^{t}$=magnitude of oscillations of the magnetic field of the nth order, ω=angular frequency of the incident electromagnetic radiation, ϵ is the permittivity of the material or medium and μ is the permeability of the material or medium.

7. The system of claim **6** wherein the radius of the first means for receiving is different from the radius of the second means for receiving, difference between the radius of said first means for receiving and the radius of said second means for receiving selected so that the beat frequency resulting from the difference is a frequency which facilitates conversion of the beat frequency electromagnetic radiation to electrical energy.

8. The system of claim **4** wherein:

said first means for receiving is composed of a dielectric material; and

said second means for receiving is composed of a dielectric material.

9. The system of claim **4** wherein:

said first means for receiving is spherical; and

said second means for receiving is spherical.

10. The system of claim **4** wherein said antenna is positioned generally between said first and second means for receiving.

11. The system of claim **4** wherein said antenna is a loop antenna.

12. The system of claim **4** wherein said antenna is a generally concave shell partially enclosing said first and second means for receiving.

13. The system of claim **4** wherein said means for transmitting is a waveguide.

14. A system for for converting incident zero point electromagnetic radiation energy to electrical energy, comprising:

a substrate;

a plurality of pairs of first means for receiving incident primary zero point electromagnetic radiation and second means for receiving incident primary zero point electromagnetic radiation, said plurality of pairs of means for receiving mounted on said substrate, said first means for receiving producing emitted secondary

electromagnetic radiation at a first frequency, said second means for receiving the incident primary zero point electromagnetic radiation producing emitted secondary electromagnetic radiation at a second frequency, the secondary radiation at the first frequency and the secondary radiation at the second frequency interfering to produce secondary radiation at a beat frequency which is lower than that of the incident primary radiation, said first means for receiving having a first volumetric size selected to resonate in response to the incident primary electromagnetic radiation in order to produce the secondary electromagnetic radiation at the first frequency at an enhanced energy density, and said second means for receiving having a second volumetric size selected to resonate in response to the incident primary electromagnetic radiation in order to produce the emitted secondary electromagnetic radiation at the second frequency at an enhanced energy density, said first and second volumetric sizes selected based on parameters of propagation constant of said first and second means for receiving, propagation constant of medium in which said first and second means for receiving are located and frequency of the incident primary electromagnetic radiation, said first and second volumetric sizes being different from each other;

a plurality of antennas for receiving the emitted secondary electromagnetic radiation at the lower frequency, said antenna providing an output responsive to the secondary electromagnetic radiation received, said plurality of antennas mounted on said substrate, each of said plurality of antennas receiving the emitted secondary electromagnetic radiation of one of said pairs of first and second means for receiving;

means for transmitting the emitted secondary electromagnetic radiation at the beat frequency from said antenna, said means for transmitting connected to said plurality of antennas;

a converter connected to said means for transmitting for receiving the emitted secondary electromagnetic radiation at the beat frequency from said antenna and converting the same to electrical current having a desired voltage and waveform.

* * * * *

advertisement

FREE ENERGY MICROWAVE

Technology (/technology/) > Electronics | by GEORGE CHANIOTAKIS (/member/GEORGE+CHANIOTAKIS/)

Follow

4,604 ♥ 56 5

Posted Dec. 22, 2015 | (cc) BY-NC-SA

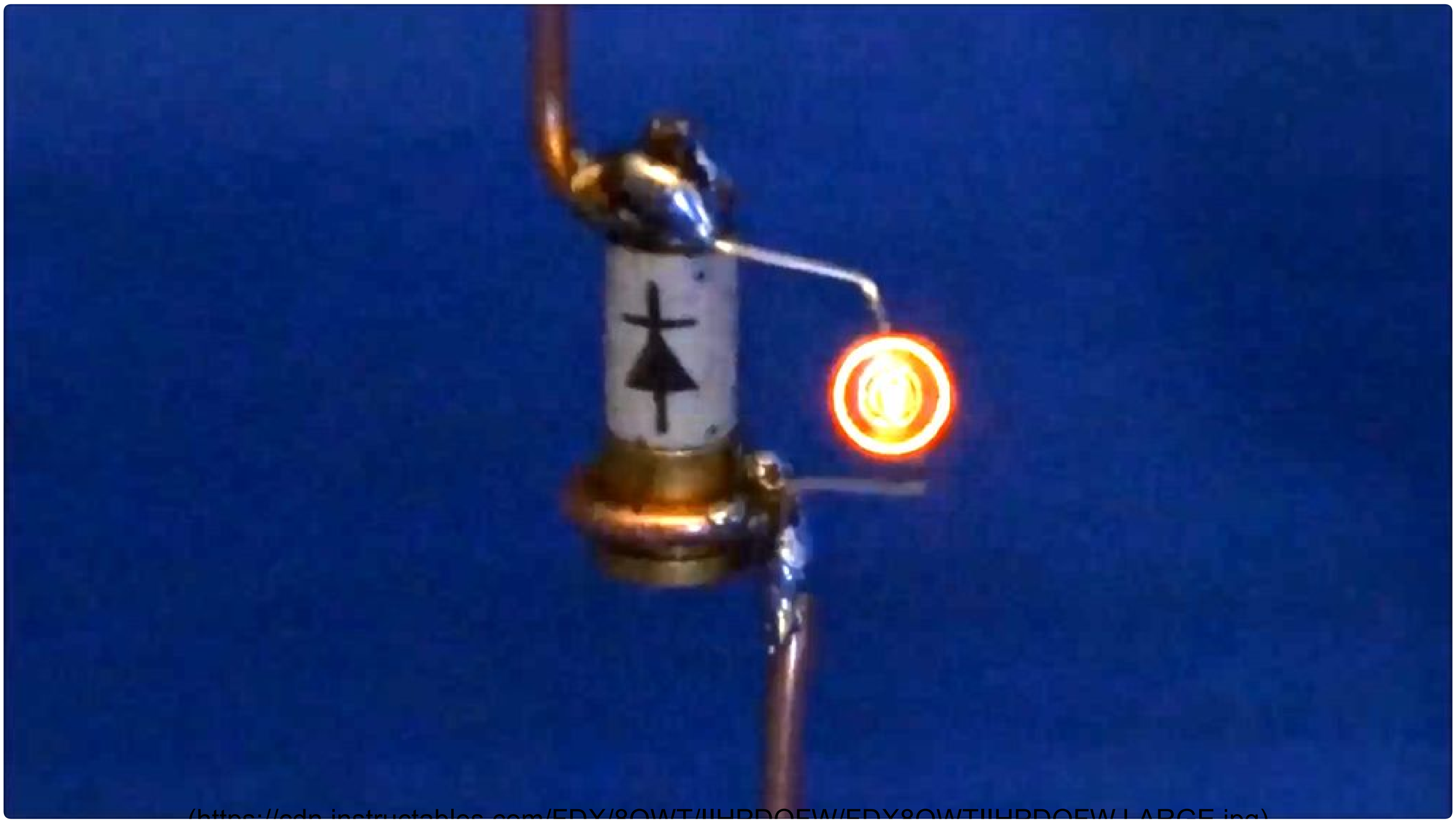

(https://cdn.instructables.com/FDY/8QWT/IIHPDOFW/FDY8QWTIIHPDOFW.LARGE.jpg)

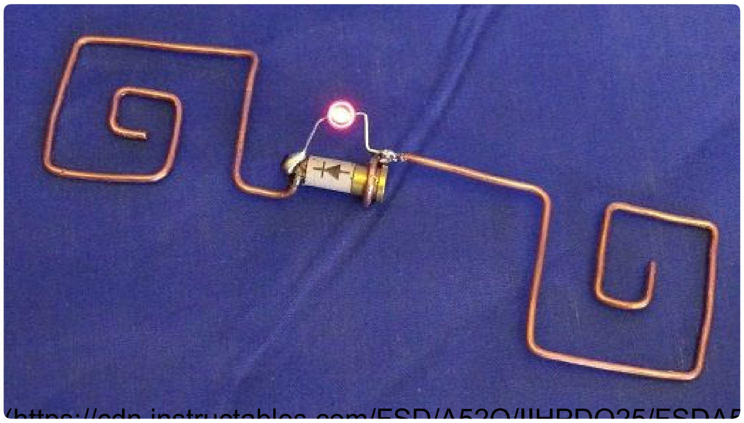

(https://cdn.instructables.com/FSD/A52Q/IIHPDQ35/FSDA5

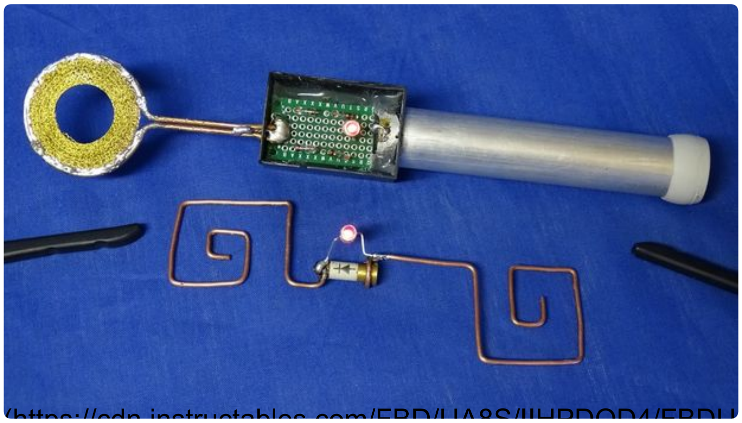

(https://cdn.instructables.com/FBD/UA8S/IIHPDQD4/FBDU

In electronics, a diode is a two-terminal electronic component that conducts primarily in one direction; it has low resistance to the flow of current in one direction, and high resistance in the other. A semiconductor diode, the most common type today, is a crystalline piece of semiconductor material with a p–n junction connected to two electrical terminals. A vacuum tube diode has two electrodes, a plate and a heated cathode. Semiconductor diodes were the first semiconductor electronic devices. The discovery of crystals' rectifying abilities was made by German physicist Ferdinand Braun in 1874. The first semiconductor diodes, called cat's whisker diodes, developed around 1906, were made of mineral crystals such as galena. Today, most diodes are made of silicon, but other semiconductors such as selenium or germanium are sometimes used.

This diode is able to detect RF signals and converts it to DC electricity to power a Led the amazing thing about this is that its completely passive and doesn't require any batteries... Gold Plated Slug 1N23B is Vintage U.S. ARMY Military for General Purpose UHF-MW Silicon Mixer. Primarily used in the Allies’ radar systems during World War II.

Add Tip Ask Question

Step 1:

ATMOSPHERIC ELECTRICITY ON EARTH AND PLANETS

J.J. BERTHELIER, LATMOS/IPSL

9th International Planetary Probe Workshop, Toulouse, June 16-17, 2012.

OUTLINE (1)

TERRESTRIAL ATMOSPHERIC ELECTRICITY

1- THE GLOBAL ATMOSPHERIC CIRCUIT

- The Conductive Atmosphere and Boundaries
- Generators
- Coupling with outer space

2- PHYSICAL PROCESSES

- Charging/Discharging Mechanisms in Clouds and Thunderstorms
- High Altitude Phenomena: Transient Luminous Events
- Acceleration Processes: Transient Gamma Ray Flashes
- EM emissions and Schumann resonances

PLANETARY ATMOSPHERIC ELECTRICITY

1- PLANETARY GLOBAL ELECTRICAL CIRCUITS

- Terrestrial Planets: Venus, Mars
- Giant Planets: Jupiter, Saturn and their moons

2- OBSERVATIONS

OUTLINE (2)

INSTRUMENTATION

1- MEASUREMENT CONDITIONS
  - Ground Based Observations
  - Observations from Space
  - Balloons

2- MEASUREMENTS TECHNIQUES
  - Conductivity
  - Electric Fields and Waves

3- SOME EXAMPLES

THE EARTH'S GLOBAL ATMOSPHERIC ELECTRIC CIRCUIT

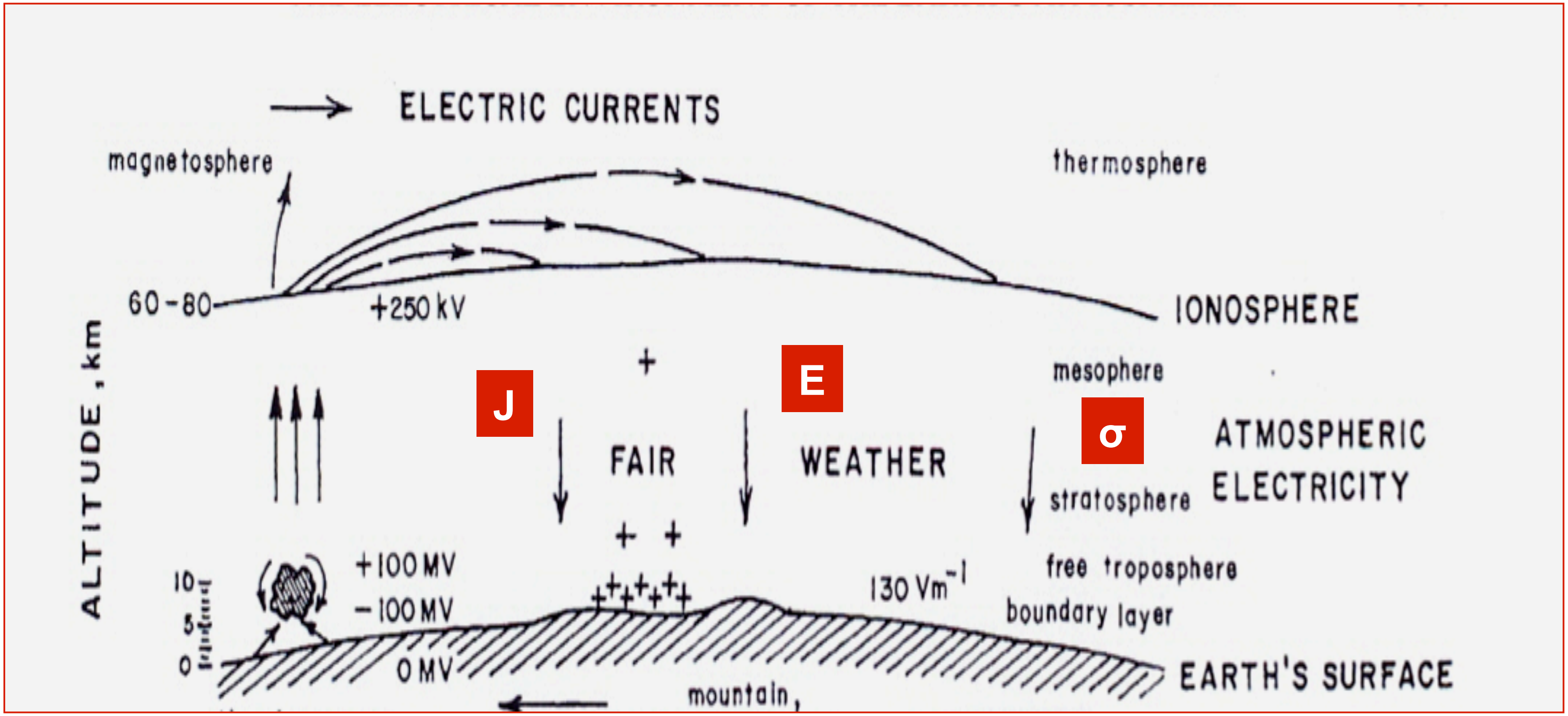
ELECTRIC CURRENTS
magnetosphere
thermosphere
60-80
+250 kV
IONOSPHERE
ALTITUDE, km
+
J
E
mesophere
σ
FAIR
WEATHER
ATMOSPHERIC
ELECTRICITY
stratosphere
+ +
+100 MV
−100 MV
10
5
0
130 Vm⁻¹
free troposphere
boundary layer
0 MV
mountain,
EARTH'S SURFACE

From M.J. Rycroft et al., JASTP, 2000

ATMOSPHERIC CONDUCTIVITY

Production Mechanisms

- Soil radioactivity at low altitude
- Cosmic rays
- Solar sources: UV, X-rays,
- auroral and polar regions: magnetospheric electrons, solar protons

Charged Particles

-Positive and Negative cluster ions

Conductivity profile

- 10^{-14} S/m at ground, scale height H_{σ}~ atmospheric scale height ~ 6-7 km
- Isotropic till ~ 70 km, anisotropic in the ionosphere $\sigma_{//} >> \sigma$ Hall, σPedersen
- day/night and latitude variations

ATMOSPHERIC CONDUCTIVITY

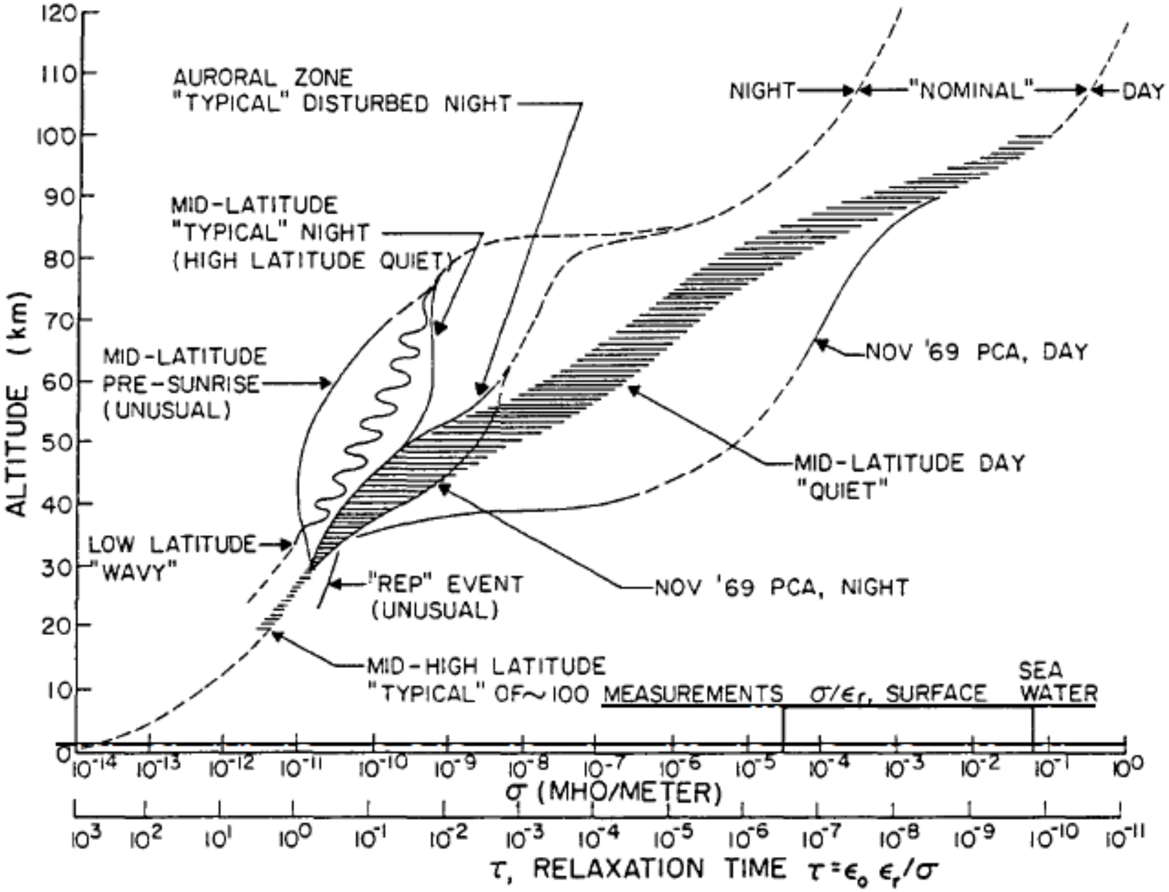

Fig. 3. Total conductivity and corresponding relaxation time under a variety of conditions.

JASR 4:4-M

From Hale, ASR, 1984

GENERATORS, THUNDERSTORMS

CHARGING MECHANISMS

Convective air motion in thunderstorm clouds, temperature profile

Collisional Charging between graupels and rain drops

Electrical structure of Thunderstorm clouds

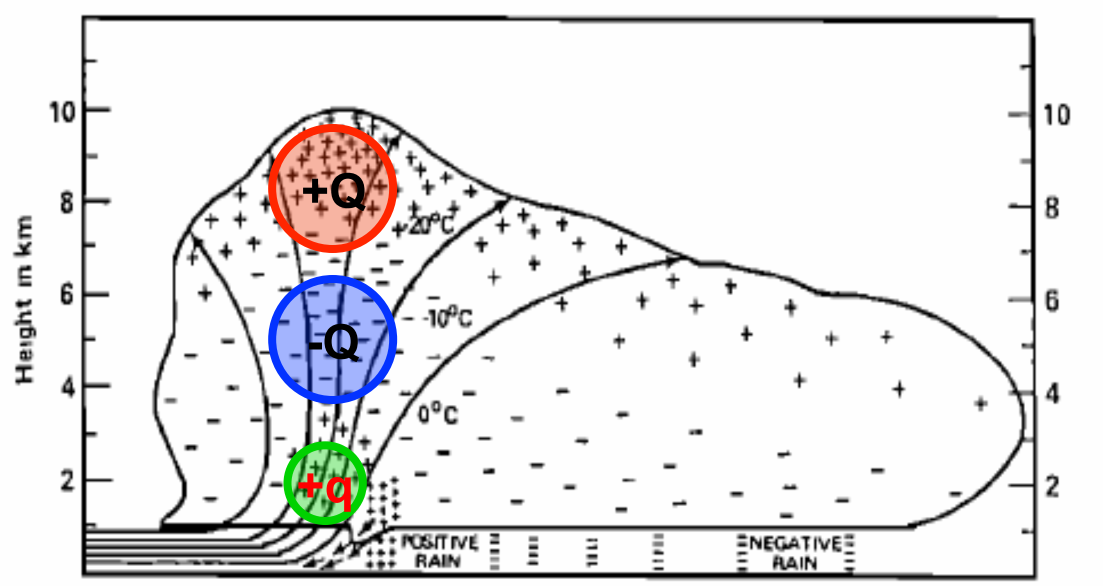
Height in km
10
8
6
4
2
10
8
6
4
2
+Q
-Q
+q
-20°C
-10°C
0°C
POSITIVE RAIN
NEGATIVE RAIN

Fig. 1. Illustration of the tripole structure of thunderclouds based on in situ measurements by *Simpson and Scrase* [1937].

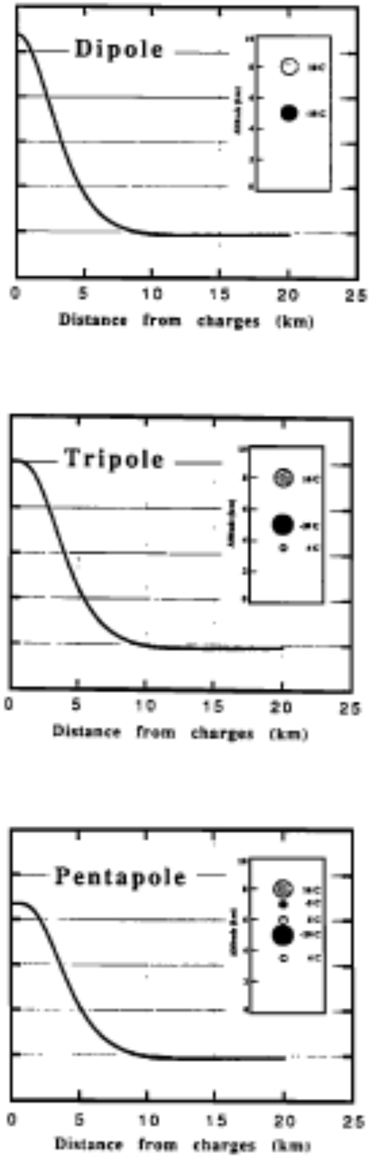

he electric field at the ground beneath three differ
utions. The distributions are shown in the inset box
tion is net neutral. For each distribution the elect
ound changes polarity at a horizontal distance of ab

From Marshall & Rust, JGR, 1991

GENERATORS, LIGHTNING

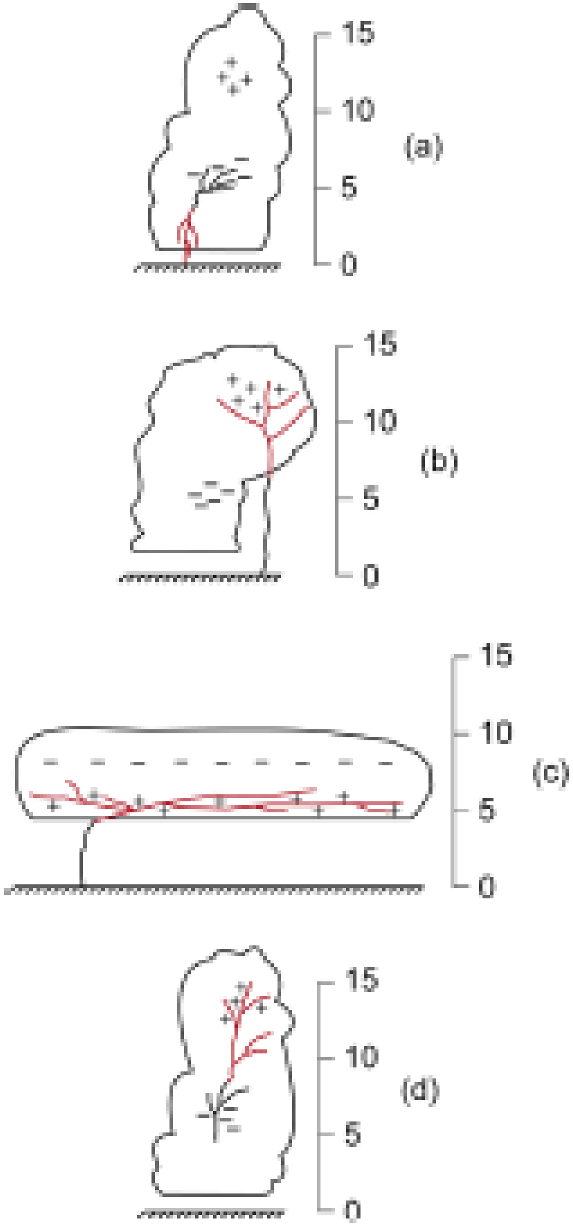

Negative Cloud to Ground, CG-

Positive Cloud to Ground, CG+

Intracloud IC

Cloud to Cloud CC

GENERATORS, LIGHTNING

Geographic Average Distribution

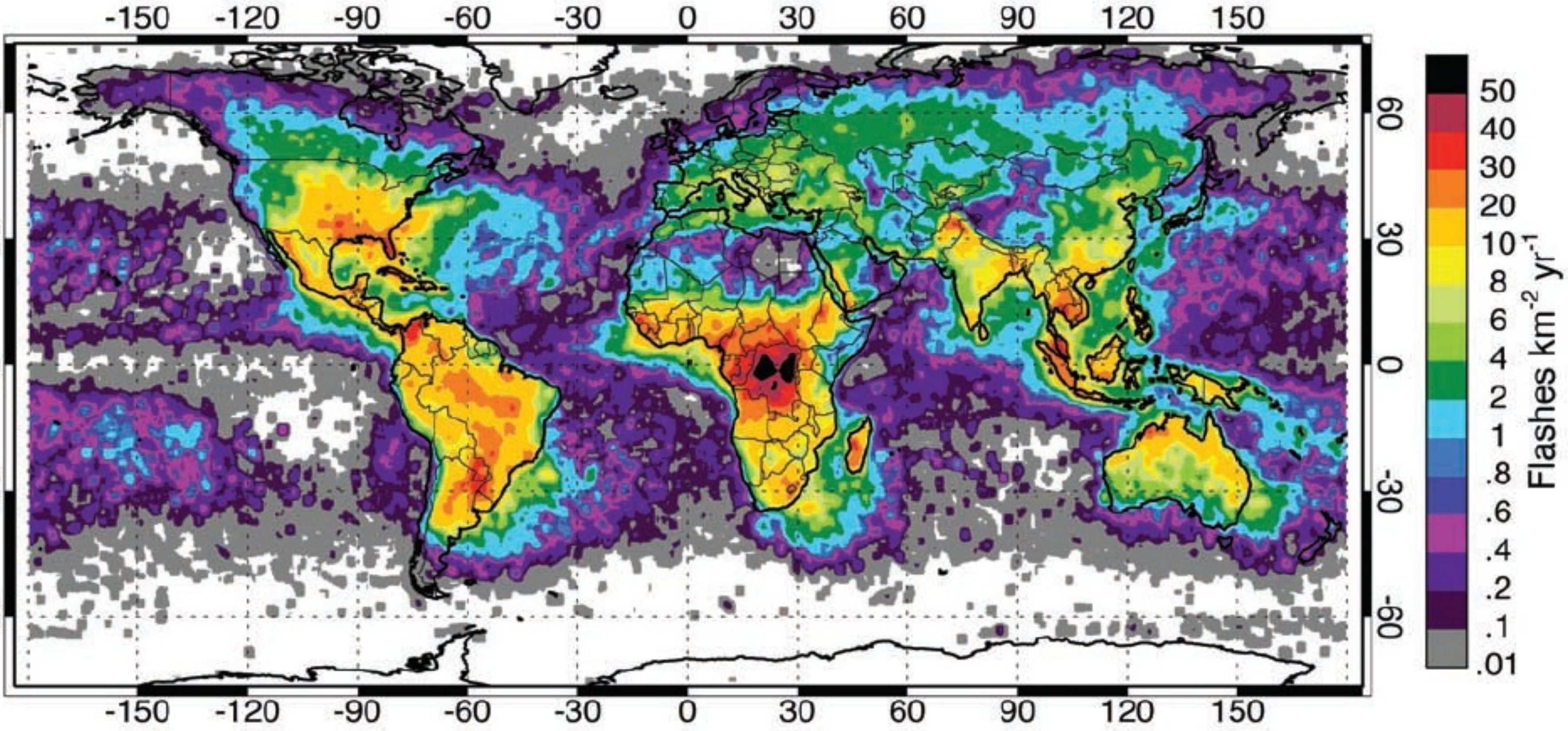
-150
-120
-90
-60
-30
0
30
60
90
120
150
60
30
0
-30
-60
50
40
30
20
10
8
6
4
2
1
.8
.6
.4
.2
.1
.01
Flashes km^{-2} yr^{-1}

Figure 4. The annualized distribution of total lightning activity (in units of fl km^{-2} yr^{-1}).

From Christian et al, JGR 2003

GENERATORS, LIGHTNING

SEASONAL VARIATIONS

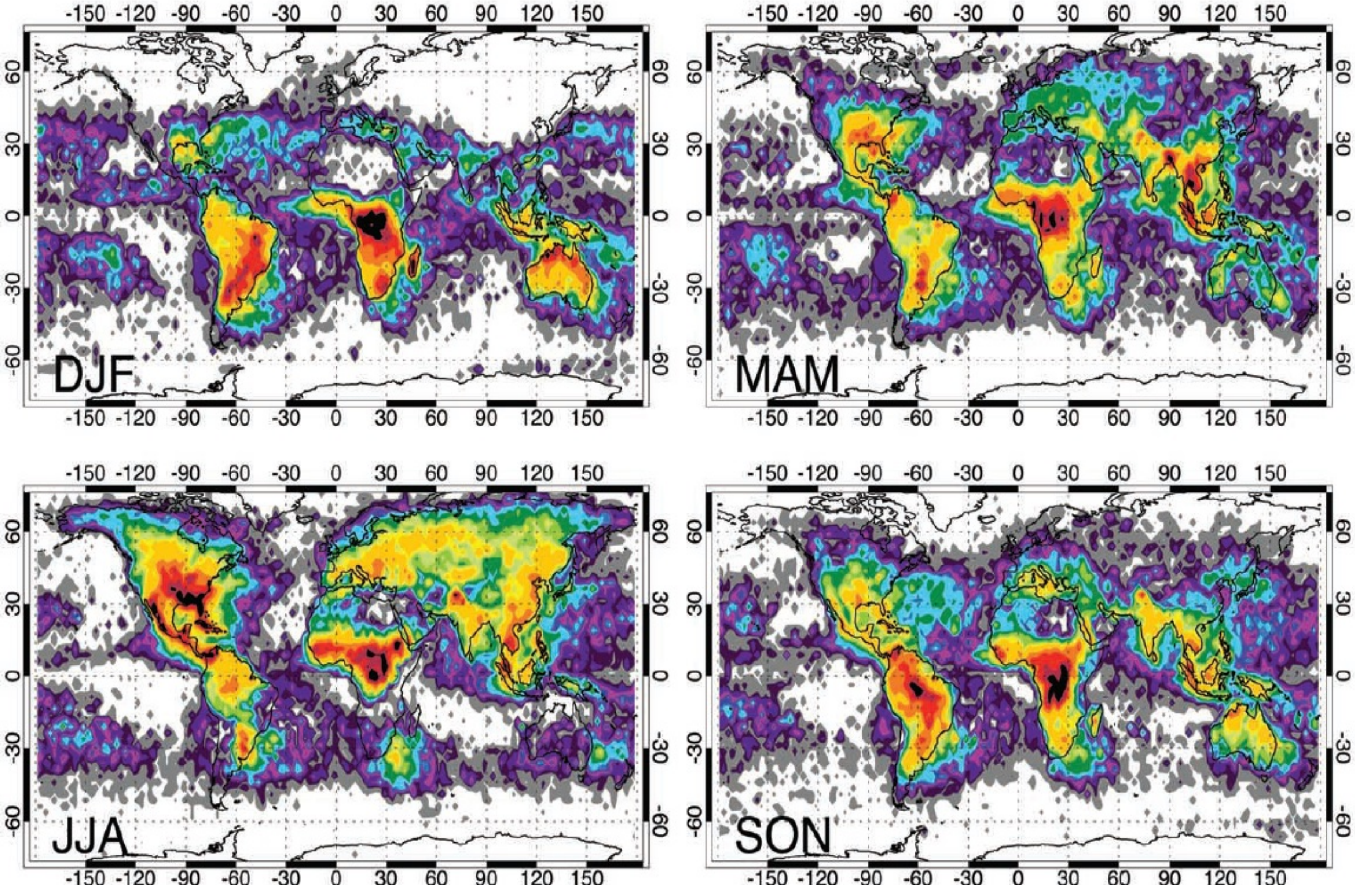
DJF
MAM
JJA
SON
-150 -120 -90 -60 -30 0 30 60 90 120 150
60 30 0 -30 -60

GENERATORS, HIGH ALTITUDE DISCHARGES, TLE’s

PHYSICAL PROCESSES

TLE’s (Sprites, Blue Jets, Elves, Gigantic jets) are electrical discharges in the stratosphere and mesosphere above active thunderstorm clouds

Sprites

Large Electric Field between cloud and ionosphere following a +CG

Break-down threshold reached at ~ 70 km

Elves

Initiated from the EMP following a CG

Ionization and luminous halo produced in the mesosphere at ~ 90 km

Blue Jets and Gigantic Jets

Streamer initiated at cloud top (~ 15-18 km) by localized intensification of the electric field

Propagate up to ~40 km for BJ, ~ 70 km or GJ

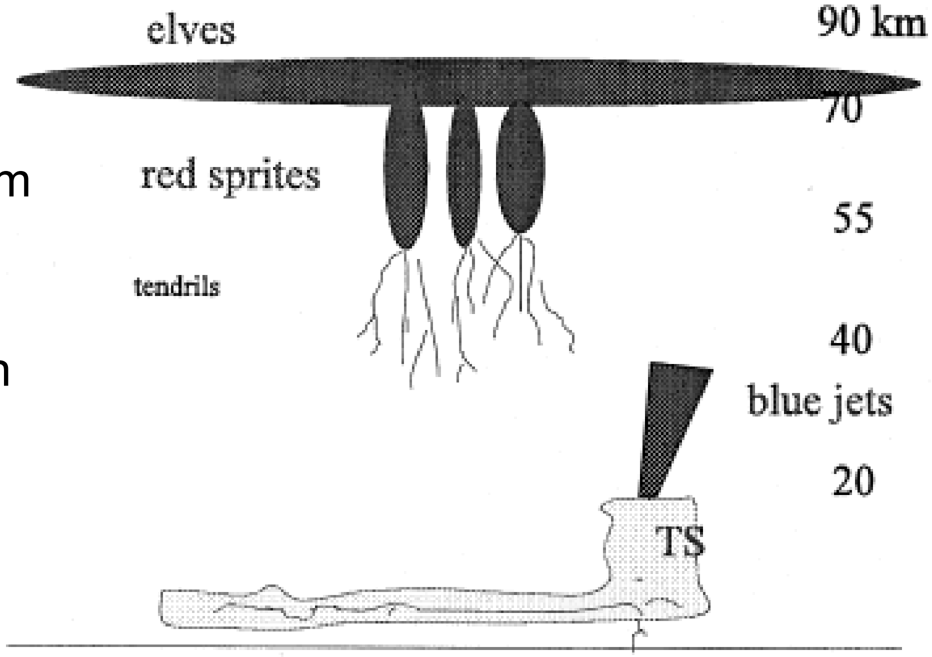

GENERATORS, HIGH ALTITUDE DISCHARGES, TLE’s

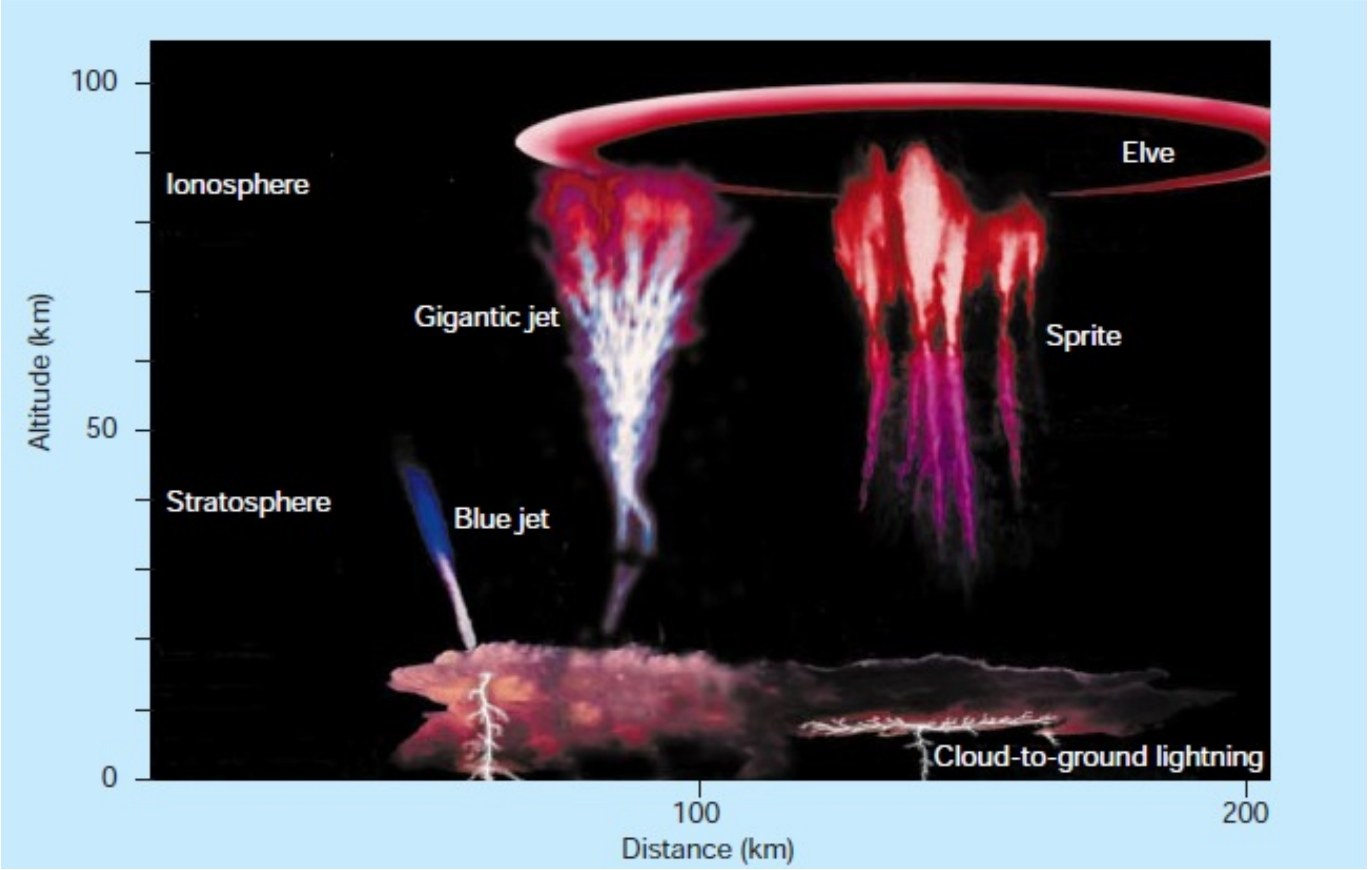
100
50
0
Altitude (km)
Ionosphere
Stratosphere
Elve
Gigantic jet
Sprite
Blue jet
Cloud-to-ground lightning
100
200
Distance (km)

GENERATORS, HIGH ALTITUDE DISCHARGES, TLE’s

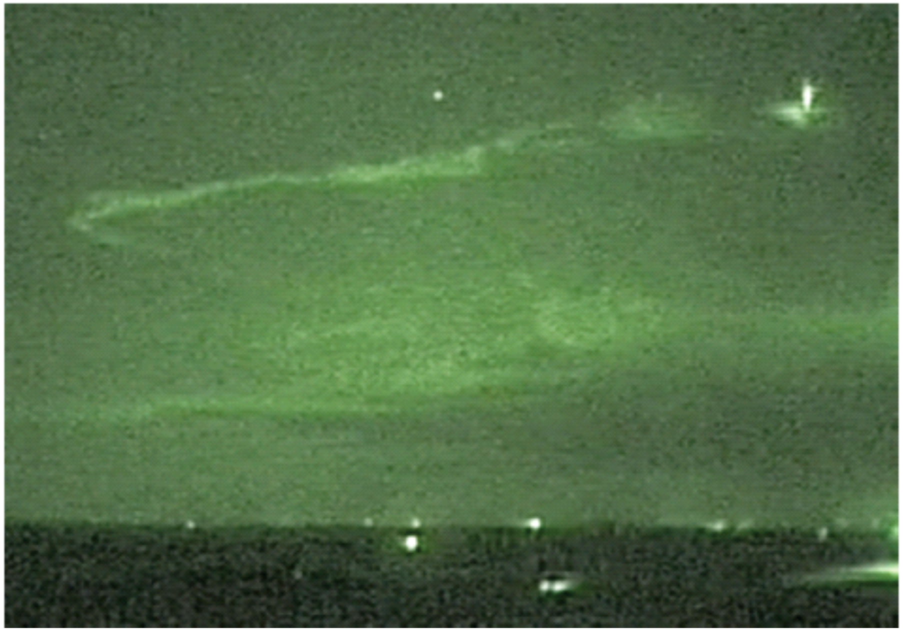

ACCELERATION PROCESSES, TGF

PHYSICAL PROCESSES

Acceleration of Electrons up to ~ 100 MeV

- **Relativistic Runaway Electron Avalanche from cosmic ray generated electrons initially proposed. But**
  (i) cosmic ray showers not a sufficient electron source
  (ii) RREA cannot account for the intensity of TGF fluxes.

-**Two mechanisms recently proposed (Dwyer2007, 2008):**
  - **Relativistic Feedback mechanism from backward propagating positrons and scattered X,γ rays**
- **Runaway Electron production in large E-fields reproduces the duration and intensity of TGF**

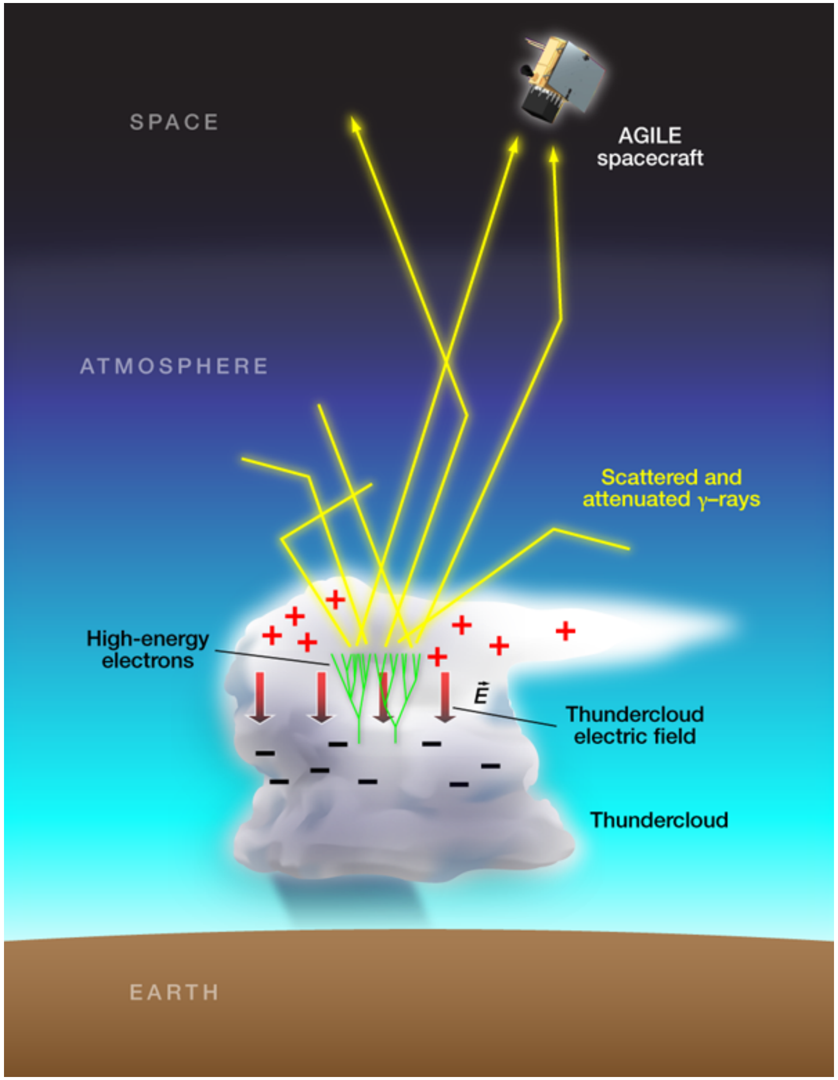
SPACE
AGILE spacecraft
ATMOSPHERE
Scattered and attenuated γ–rays
High-energy electrons
$\vec{E}$
Thundercloud electric field
Thundercloud
EARTH

From Dwyer, 2011

ACCELERATION PROCESSES, TGF

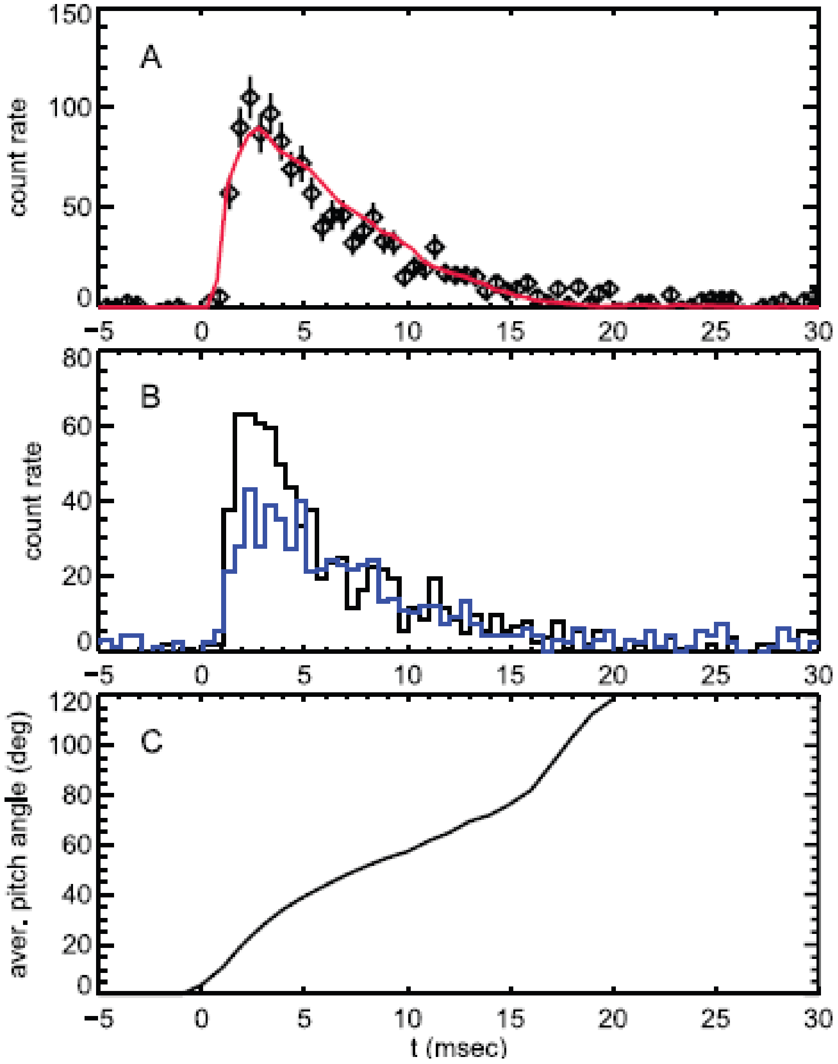

Simulation/Observations comparisons

Electron fluxes:

RHESSI data (black dots)

Simulated fluxes (red curve)

(Dwyer, JGR, 2008)

ELECTRO-MAGNETIC EMISSIONS

Sferics

- frequency spectrum peaked at ~ 10 kHz, extends up to a few MHz

Whistlers

- Ionospheric propagation along Earth's magnetic field and ducts at ELF/VLF

Transverse Resonance

- Between the surface and the lower ionosphere, ~ 1.5 to 3 kHz

Schumann Resonances

-Resonant modes of the Earth-Ionosphere wave guide $\omega \sim [n(n+1)]^{1/2}$ (c/R)

- frequencies 7.8, 14.3, 20.8, 27.3, 33.8, …

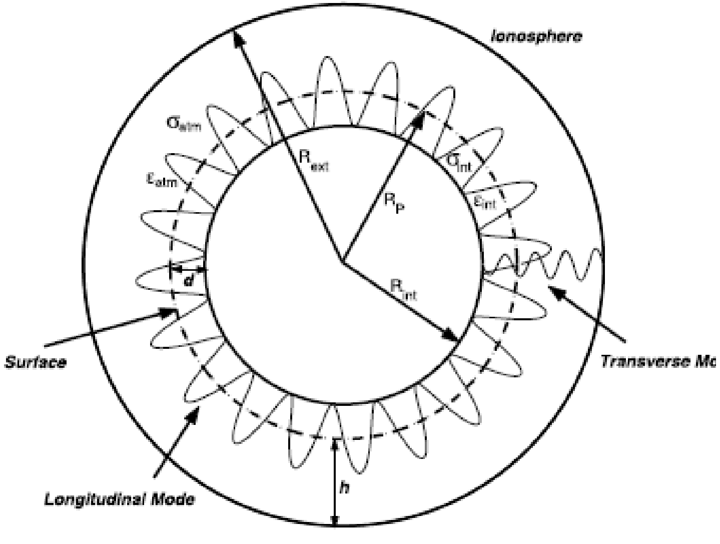
Ionosphere
σ_{atm}
ε_{atm}
R_{ext}
σ_{int}
ε_{int}
R_P
d
R_{int}
Surface
Transverse Mo
h
Longitudinal Mode

COUPLING WITH OUTER SPACE

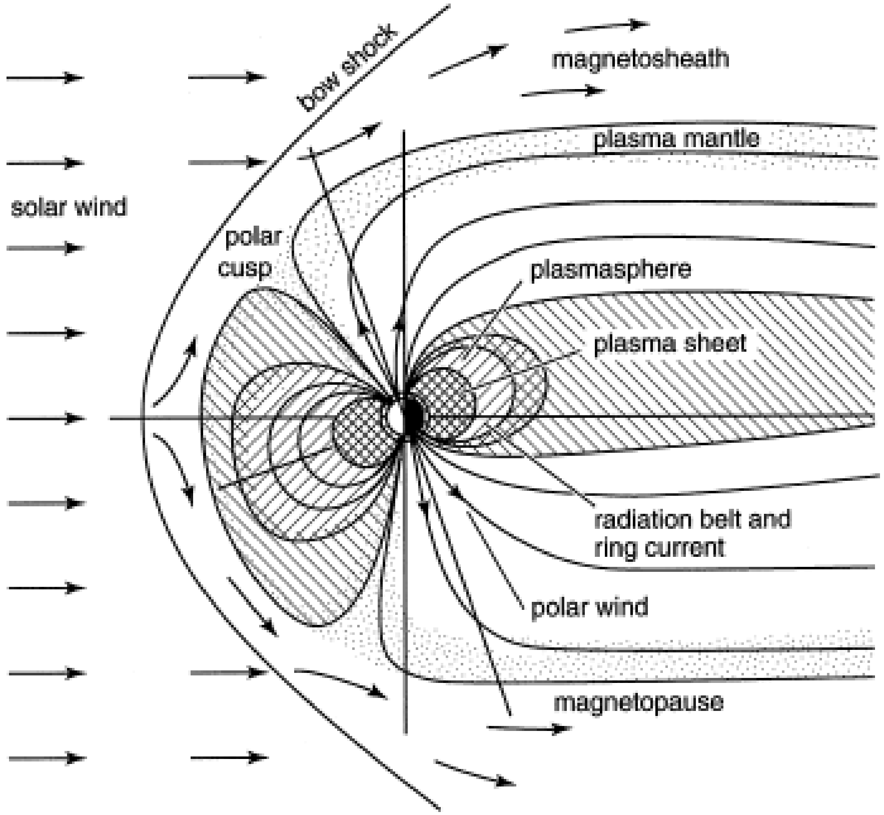

bow shock
magnetosheath
plasma mantle
solar wind
polar cusp
plasmasphere
plasma sheet
radiation belt and ring current
polar wind
magnetopause

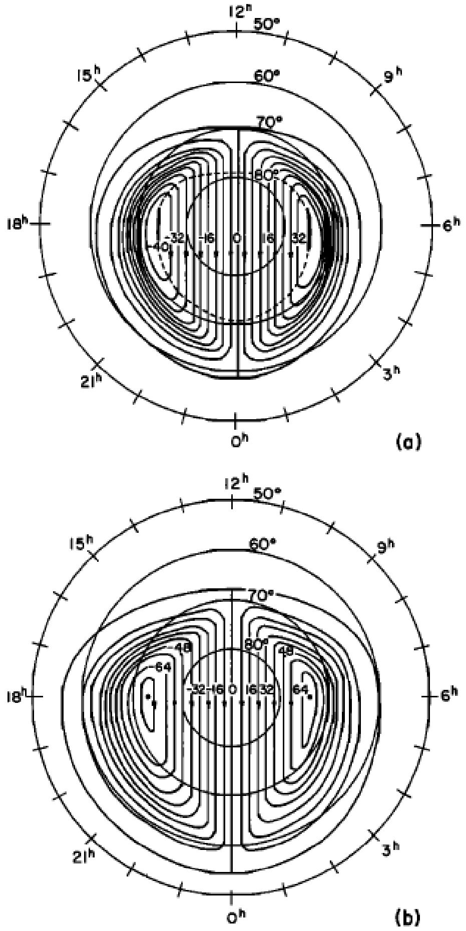

12h
50°
60°
70°
80°
15h
9h
18h
6h
21h
3h
0h
-40
-32
-16
0
16
32
(a)
12h
50°
60°
70°
80°
15h
9h
18h
6h
21h
3h
0h
-64
48
-32
-16
0
16
32
64
48
(b)

COUPLING WITH OUTER SPACE

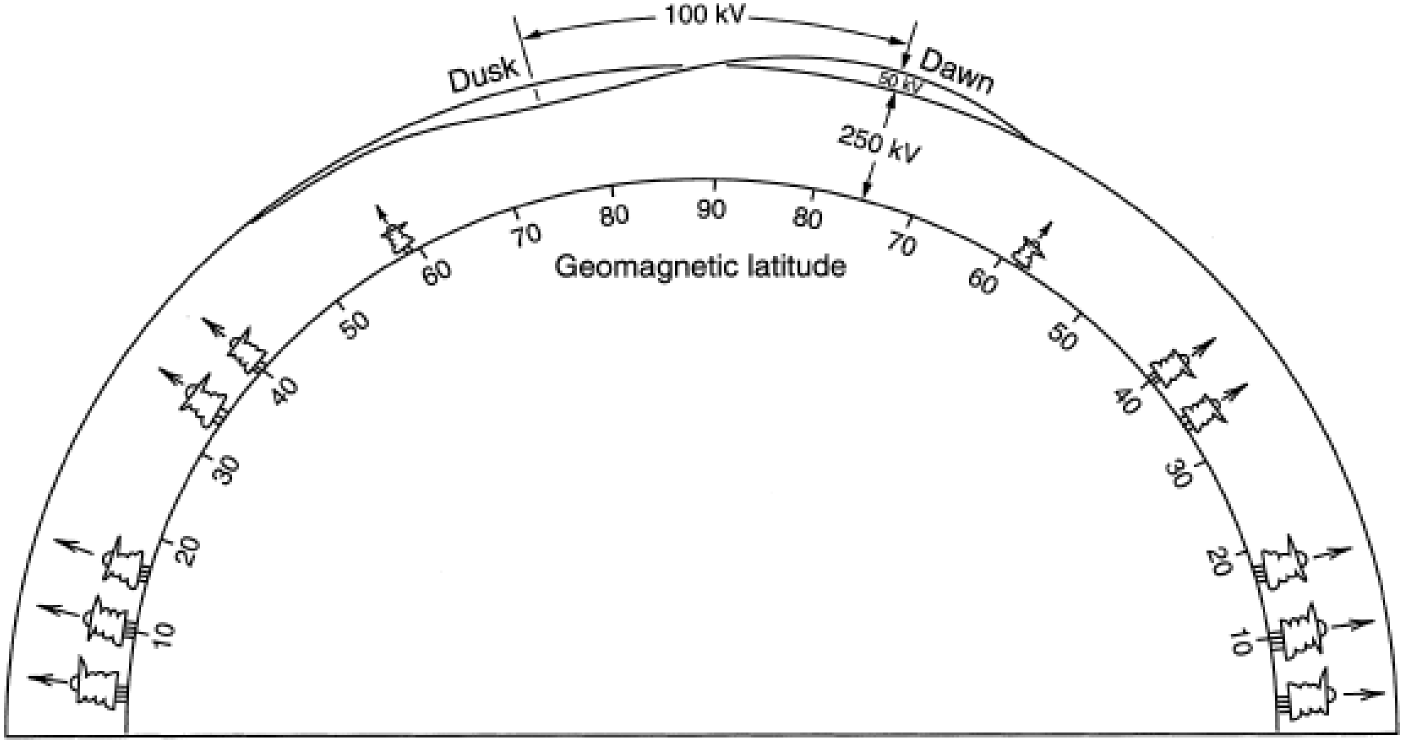
100 kV
Dusk
Dawn
50 kV
250 kV
90
80
70
60
50
40
30
20
10
80
70
60
50
40
30
20
10
Geomagnetic latitude

From Rycroft, 2011

PLANETARY ATMOSPHERIC ELECTRICITY

CONDITIONS FOR A GLOBAL ELECTRICAL CIRCUIT ON OTHER PLANETS

	Ion mobility in lower atmosphere	Upper Conductive Boundary	Lower Conductive Boundary	Clouds	Electrification Mechanisms
Venus	yes	Ionosphere	Yes σ_g small	yes	Charge separation Lightning
Mars	yes	Ionosphere	σ_g? Water reservoirs?	Faint, high altitude	Dust impact
Jupiter	Probably not in deep atmosphere	Ionosphere	? Deep layers?	yes	Charge separation lightning
Saturn	Probably not in deep atmosphere	Ionosphere	? Deep layers?	yes	Charge separation Lightning
Titan	Yes, weak at low altitude	Ionosphere	Buried ocean	yes	? Lightning not observed

VENUS

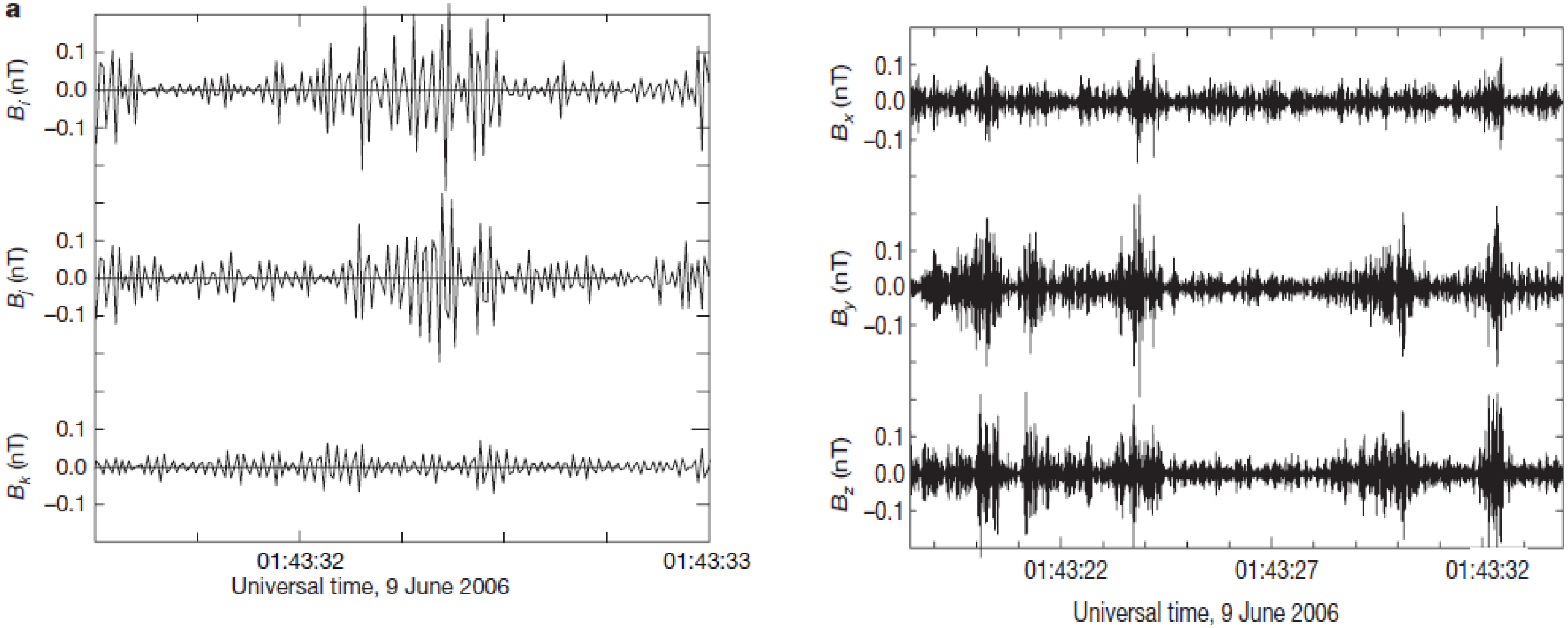

Whistler mode ELF signals from the Venus Express magnetometer
Indicating the existence of lightning

Russel et al., ASR 2008

MARS

Conductivity

- profile similar to Earth
- σ at surface ~ 10^3 σ on Earth

Electrification Mechanism

- Dust Impact
- Charge depends on size, material
- Breakdown at ~ 10 kV/m

Generators

- Dust storms
- Dust Devils

Observations

-SR of extremely high amplitude (?)

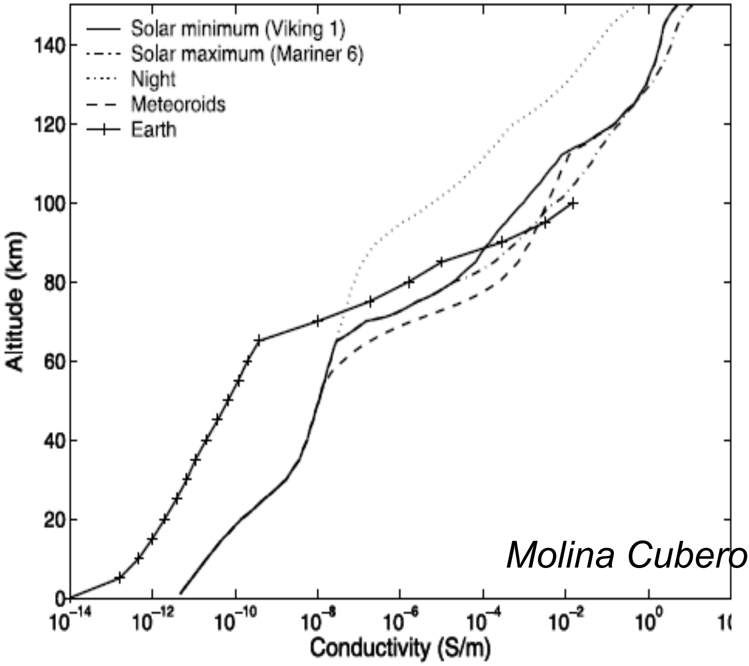

Molina Cuberos, 2006

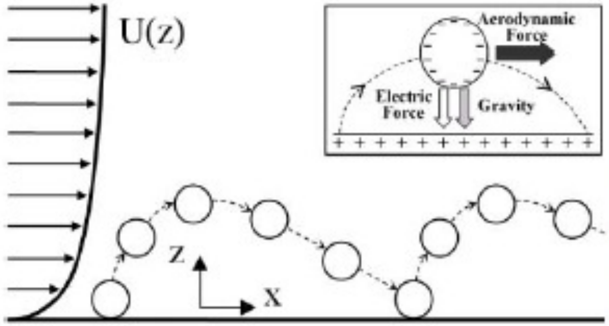
U(z)
Aerodynamic Force
Electric Force
Gravity
z
x

(Kok, Renno, 2007)

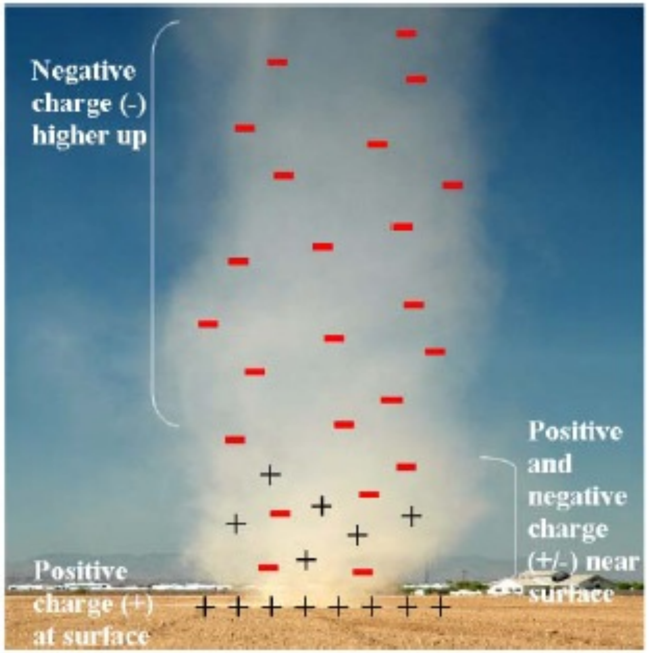
Negative charge (-) higher up
Positive and negative charge (+/-) near surface
Positive charge (+) at surface

SATURN

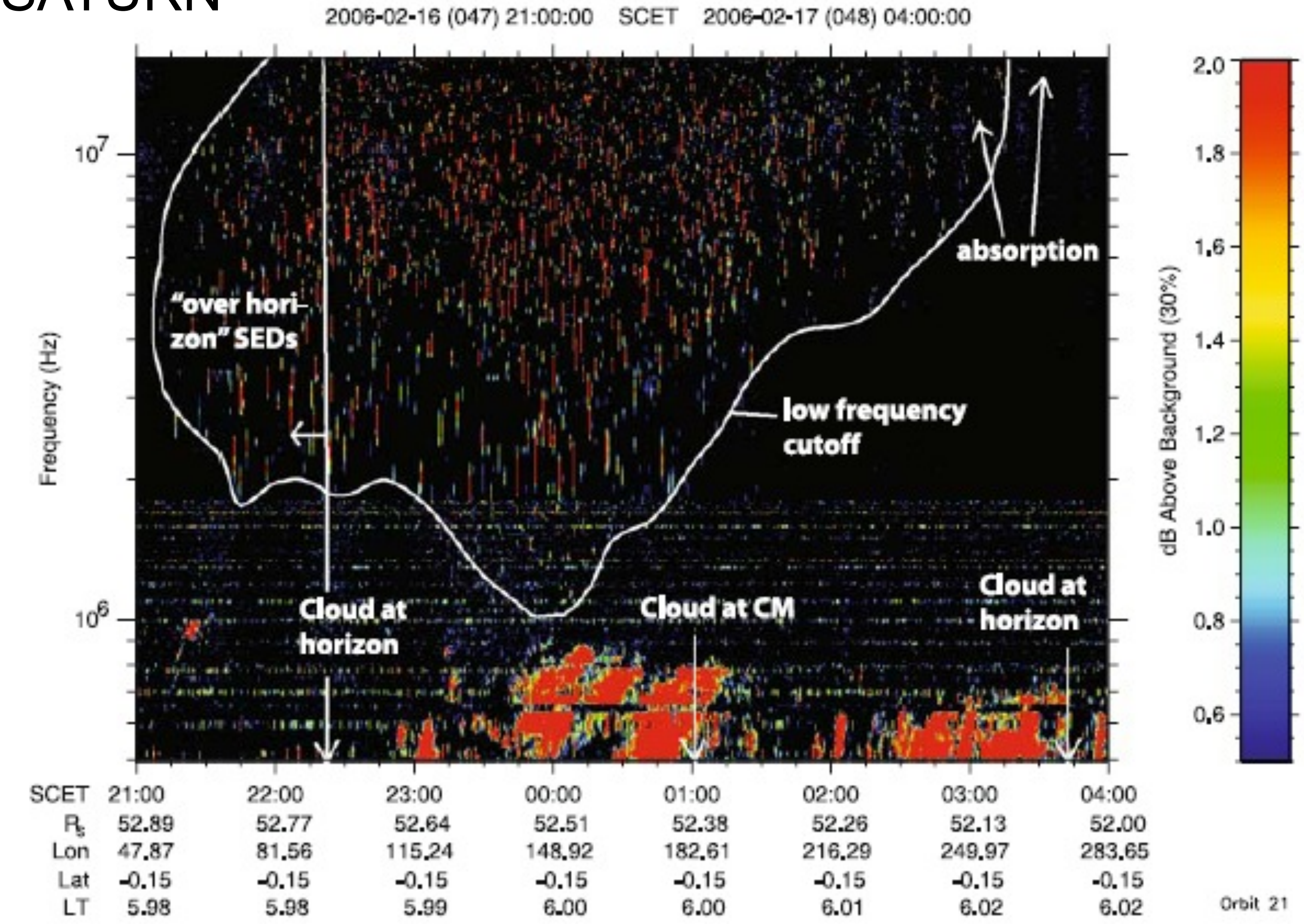

« Saturn Electronic Discharges » (SED) observed by RPWS on Cassini
EM emissions from lightning

SATURN and JUPITER

Comparison between Saturn, Jupiter and Earth lightning

Lightning sources :

- Saturn: Giant convective storm systems 3000 km in diameter, equator or 35°S
- Jupiter: numerous storms in ~ 5° latitude bands at ~ 50° N and S
- Updrafting water clouds at levels of ~ 10 bar (Saturn), ~ 5 bar (Jupiter)

Characteristics

- Frequency spectrum: up to 20 MHz / a few kHz / a few MHz
- Spectral power at MHz frequencies 100 W/Hz / ~10 / 0.01

Physical process

- Saturn: Elves more likely than Sprites
- Jupiter: lightning a few ms long

Desch et al., 2002; Fisher et al., 2008

TITAN

36 Hz peak in ELF AC electric field spectrum
interpreted as the second harmonic of Schumann resonnance
of the planet-ionosphere wave-guide due to a buried ocean at ~ 60-80 km depth

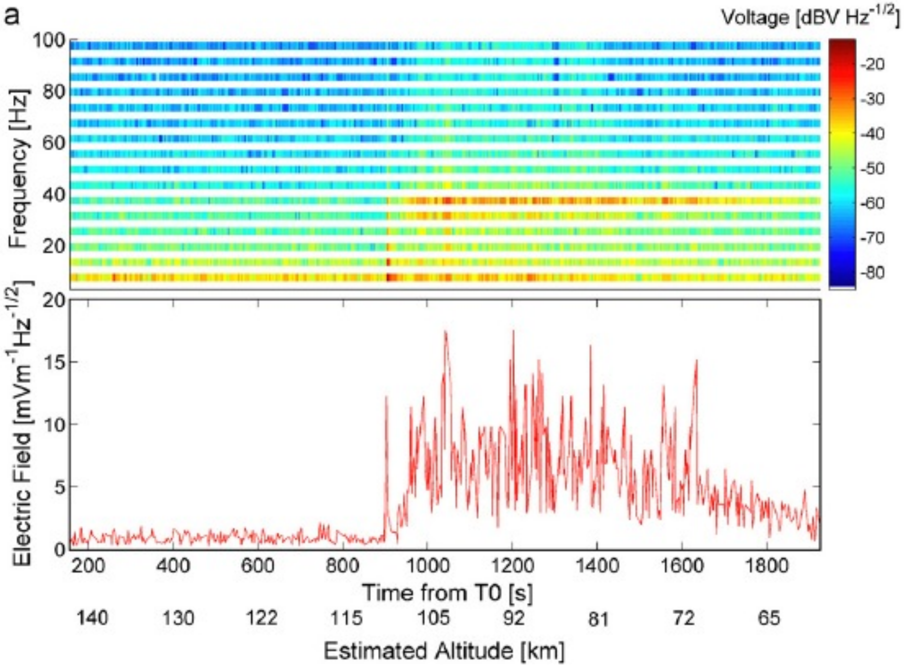

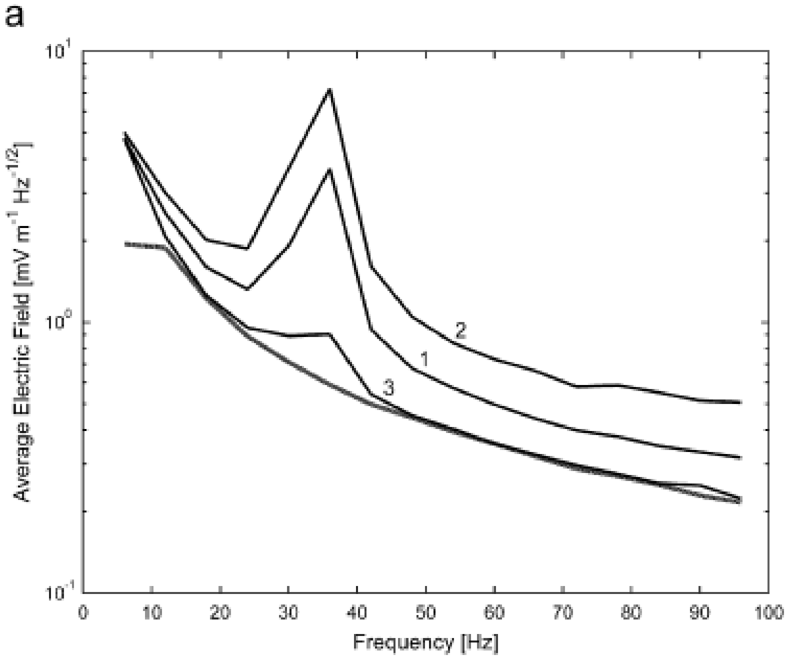

Simoes et al., Icarus, 2007
Beghin et al, 2012

INSTRUMENTATION AND OBSERVATIONS

1- CONDUCTIVITY

2- ELECTRIC FIELDS AND WAVES

3- CURRENTS

4- EXAMPLES OF OBSERVATIONS
    - Stratospheric balloon measurements
    - Huygens probe measurements in the atmosphere of Titan

ATMOSPHERIC ELECTRICITY PARAMETERS AND THEIR MEASUREMENTS

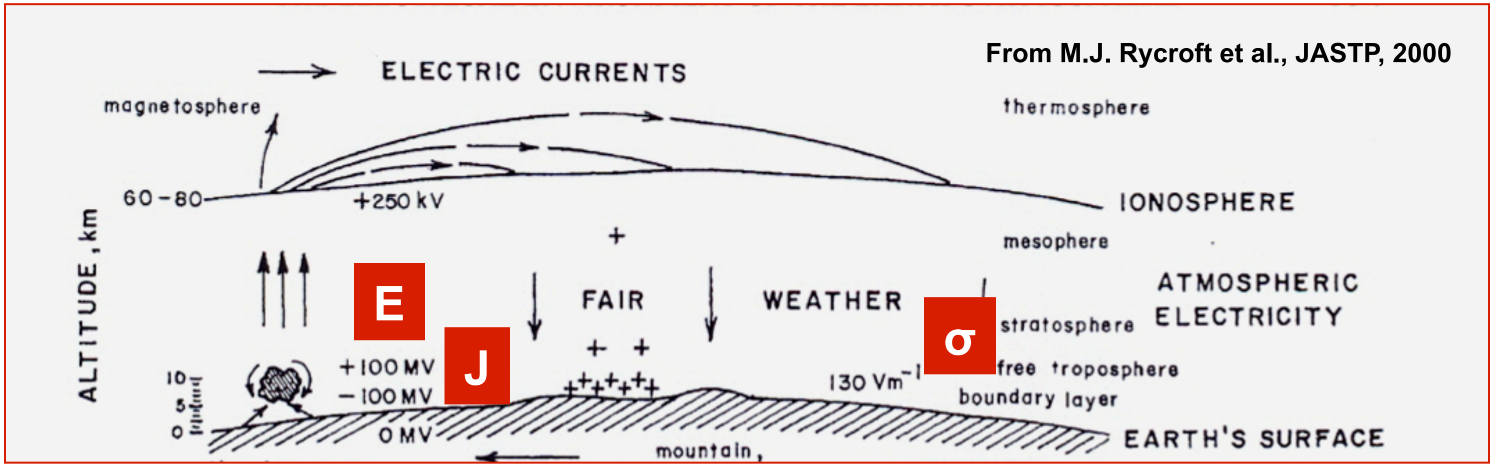
ELECTRIC CURRENTS
magnetosphere
thermosphere
60 - 80
+250 kV
IONOSPHERE
+
mesophere
ALTITUDE, km
E
FAIR
WEATHER
σ
ATMOSPHERIC
ELECTRICITY
stratosphere
J
+ +
+100 MV
-100 MV
130 Vm^{-1}
free troposphere
boundary layer
10
5
0
0 MV
mountain,
EARTH'S SURFACE

From M.J. Rycroft et al., JASTP, 2000

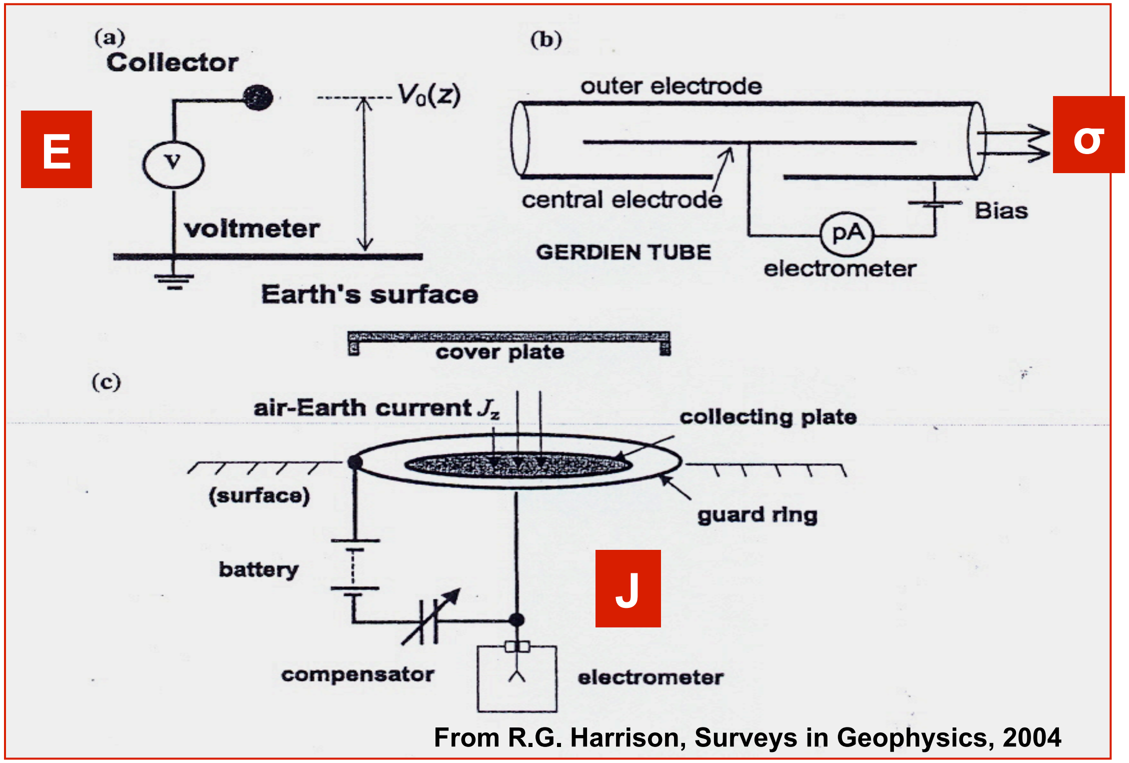
(a)
Collector
$V_0(z)$
E
V
voltmeter
Earth's surface
(b)
outer electrode
σ
central electrode
Bias
pA
GERDIEN TUBE
electrometer
cover plate
(c)
air-Earth current J_z
collecting plate
(surface)
guard ring
battery
J
compensator
electrometer

From R.G. Harrison, Surveys in Geophysics, 2004

ATMOSPHERE ELECTRIC CONDUCTIVITY (1)

Gerdien Condenser

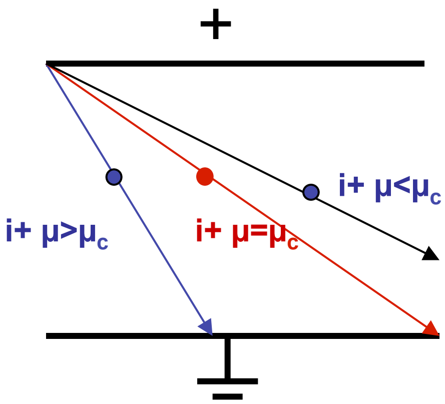
+
i+ μ<μc
i+ μ>μc
i+ μ=μc

$\sigma^+ = \Sigma\ e.n_i.\mu_i$

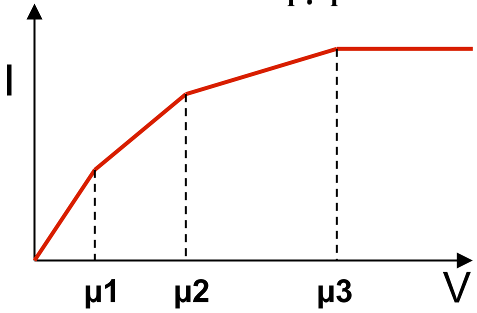

Relaxation method

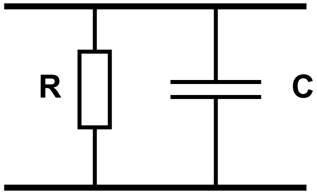

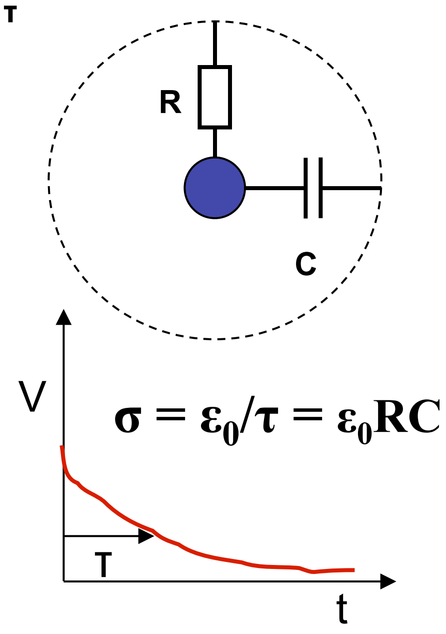

$\sigma = \varepsilon_0/\tau = \varepsilon_0 RC$

Mutual Impedance

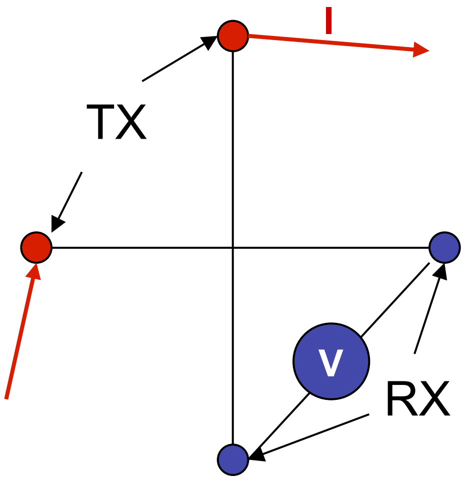

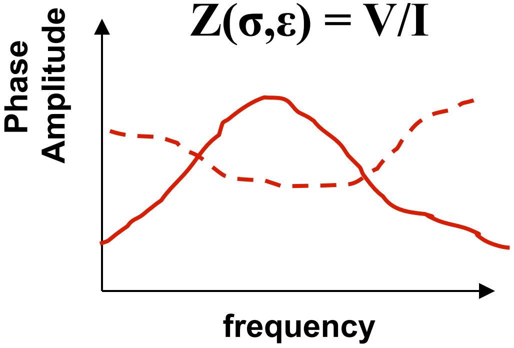

$Z(\sigma,\varepsilon) = V/I$

ATMOSPHERE ELECTRIC CONDUCTIVITY (2)

Gerdien condenser

- gives access to the ion mobility spectrum
- relaxation method can be included
- requires accurately controlled air flow
- complexity (fan, electronics…)

Relaxation method (with electrostatic probes)

- simple, can be easily implemented on double probe
- accurate technique for large enough σ and stable electric field
- analysis difficult with a variable background electric field
- requires specific modeling in case of space charge

Mutual impedance

- accurate in atmospheres with both ions and electrons
- able to provide soil measurements if landed on planetary surface
- complexity (booms, electronics…)

TERRESTRIAL CONDUCTIVITY PROFILES

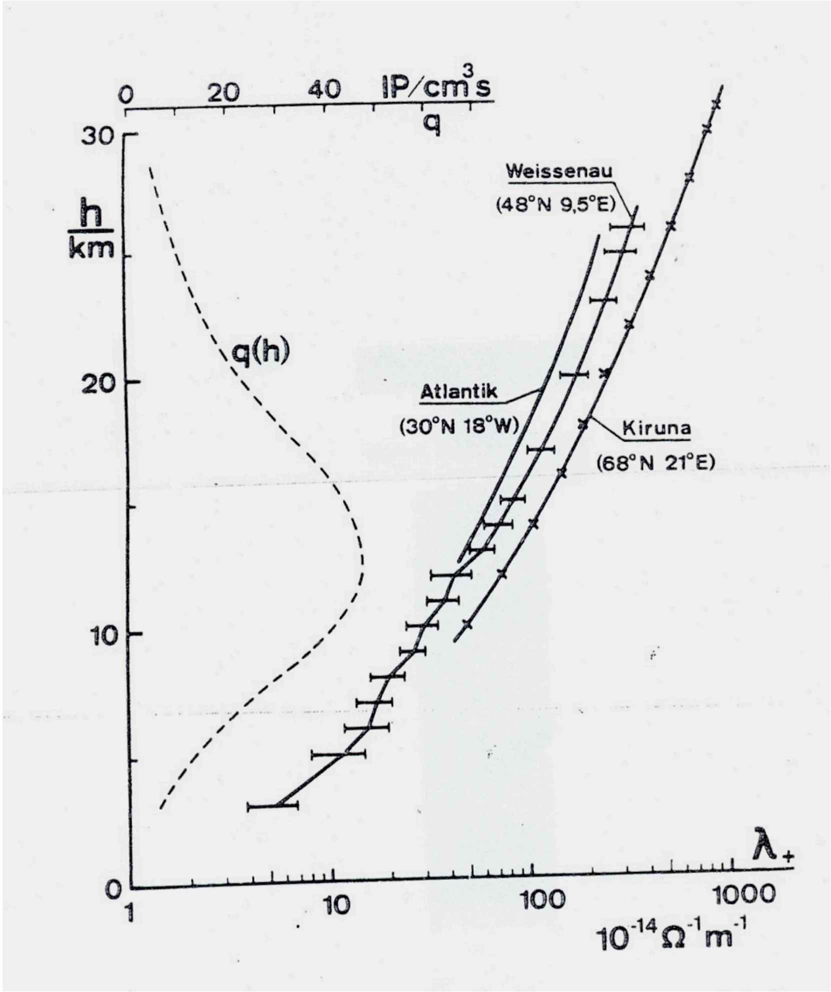

From W. Gringel, Prometheus, 1977

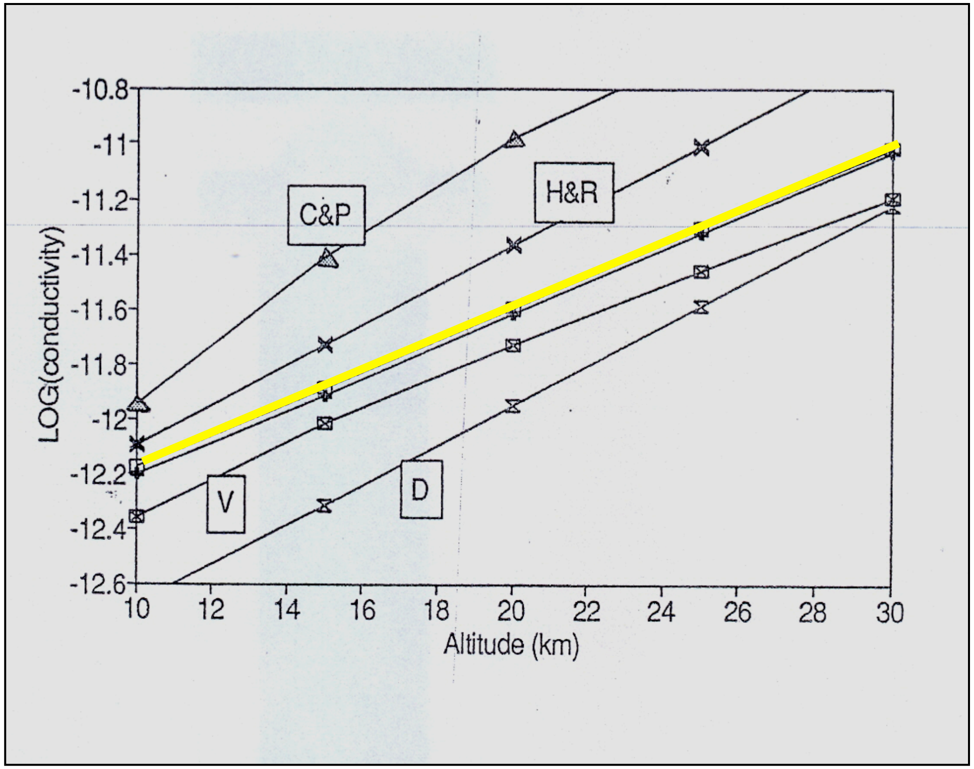

From R.H. Holzworth, JGR, 1991

ELECTRIC FIELD MEASUREMENTS

INDUCTIVE COUPLING

THE « FIELD MACHINES »

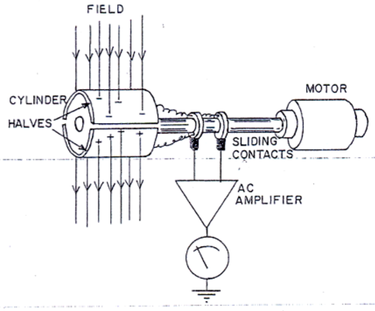
FIELD
CYLINDER
HALVES
MOTOR
SLIDING CONTACTS
AC AMPLIFIER

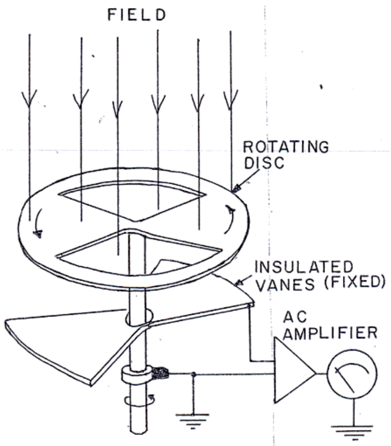
FIELD
ROTATING DISC
INSULATED VANES (FIXED)
AC AMPLIFIER

RESISTIVE COUPLING

THE DOUBLE PROBE INSTRUMENT

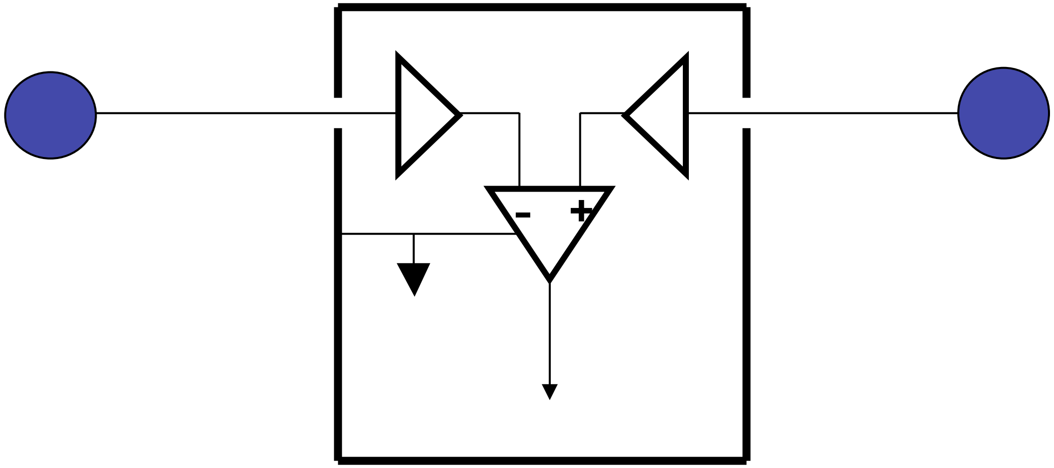

ELECTRIC FIELD MEASUREMENTS

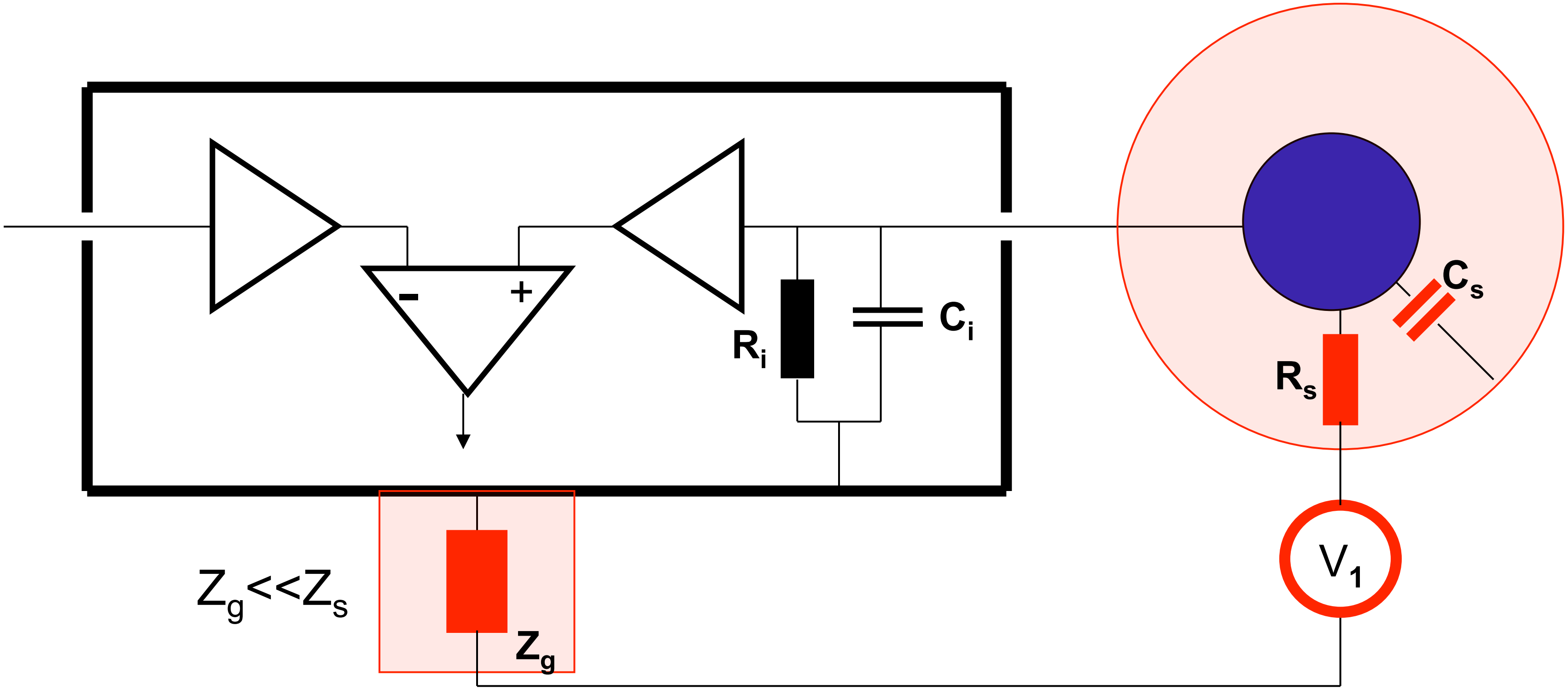
-
+
R_i
C_i
C_s
R_s
V_1
$Z_g<<Z_s$
Z_g

$$V_{out}=\left[(V_1-V_2)+(WF_2-WF_1)\right]\cdot\left[\frac{Z_i}{Z_i+Z_s}\right]$$

ELECTRIC FIELD MEASUREMENTS

Transition frequency $$\omega_s = \frac{1}{R_s \cdot C_s}$$

Low Frequency limit $$V_{out} = (V_1 - V_2) \cdot \frac{R_i}{R_i + R_s} \cong (V_1 - V_2)$$ if $R_i >> R_s$

High Frequency Limit $$Z_s \approx {}^{1}\!/_{\omega C_s} \quad Z_i \approx {}^{1}\!/_{\omega C_i} \qquad V_{out} = (V_1 - V_2) \cdot \frac{C_s}{C_s + C_i}$$

DC-ELF : spherical sensors (with polarization current)

HF : cylindrical long antenna

ELECTRIC FIELD MEASUREMENTS

INDUCTIVE COUPLING, THE « FIELD MACHINES »

- high impedance, can operate in very low σ (Earth's surface)
- can measure large DC electric fields
- Some directional capability, 2D measurement with 1 sensor
- innovative designs in progress, miniaturized devices, vibrating system, ASIC

- low sensitivity
- electro-mechanical device (harsh environments, dust, EMI…), signal processing
- low temporal resolution and low frequency capability (no waves)

RESISTIVE COUPLING, THE DOUBLE PROBE INSTRUMENT

- simple device and electronics
- high sensitivity (in AC better than $1\mu V/mHz^{1/2}$)
- high temporal resolution (lightning), high frequency capability (MHz)
- direct measurement, no signal processing
- high amplitude DC and AC electric field measurements in dedicated modes

- booms (deployment, mass,…)
- limited to $\sigma \geq 10^{-13}$ S/m (on Earth above ~ 8 kilometers)

ATMOSPHERIC CURRENT MEASUREMENTS

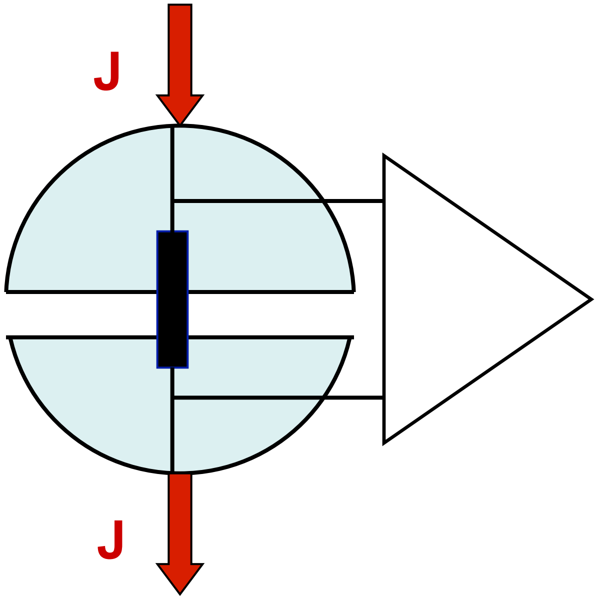

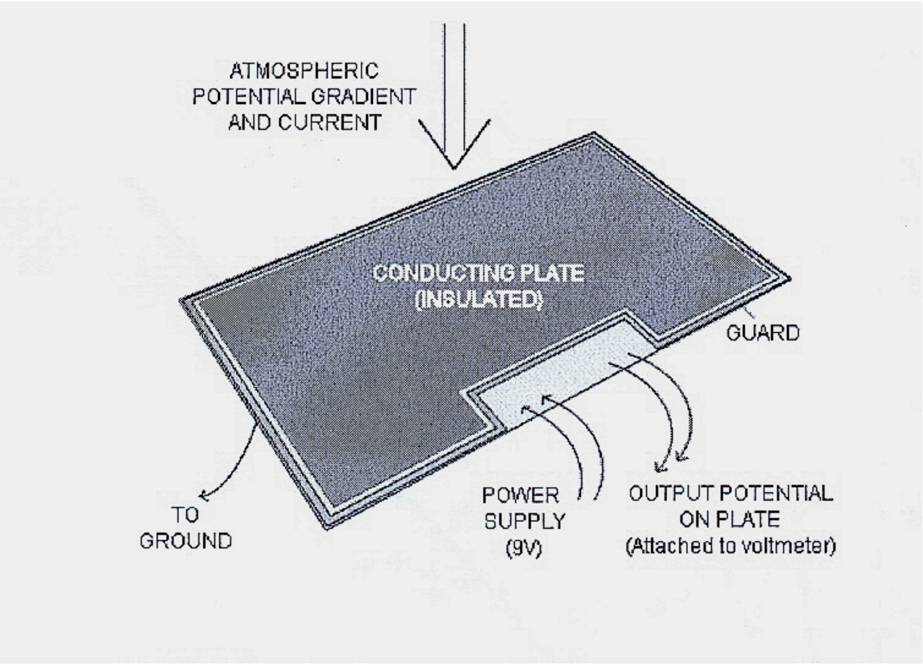
ATMOSPHERIC
POTENTIAL GRADIENT
AND CURRENT
CONDUCTING PLATE
(INSULATED)
GUARD
TO
GROUND
POWER
SUPPLY
(9V)
OUTPUT POTENTIAL
ON PLATE
(Attached to voltmeter)

From A.J. Bennet and R.G. Harrison, Sub. Adv in Geosciences

ELECTRIC FIELDS AND CONDUCTIVITY MEASUREMENTS ON STRATOSPHERIC BALLOON FLIGHTS

HVAIRS Gondola
AMMA Campaign in Niger

Electric Field Intrument

- Vertical component of Electric Field
  - DC to 3 kHz
  - Large signal « DC channel »
    ~ ± 50 mV/m to ± 200 V/m
    (up to ± 10 kV/m in special mode)
  - Small signal « AC channel »
    sensitivity ~ 30 µV/m. Hz$^{1/2}$
- Conductivity measurements
  - relaxation method

Optical sensors

- photodiode lightning detectors

On-Board Data Storage

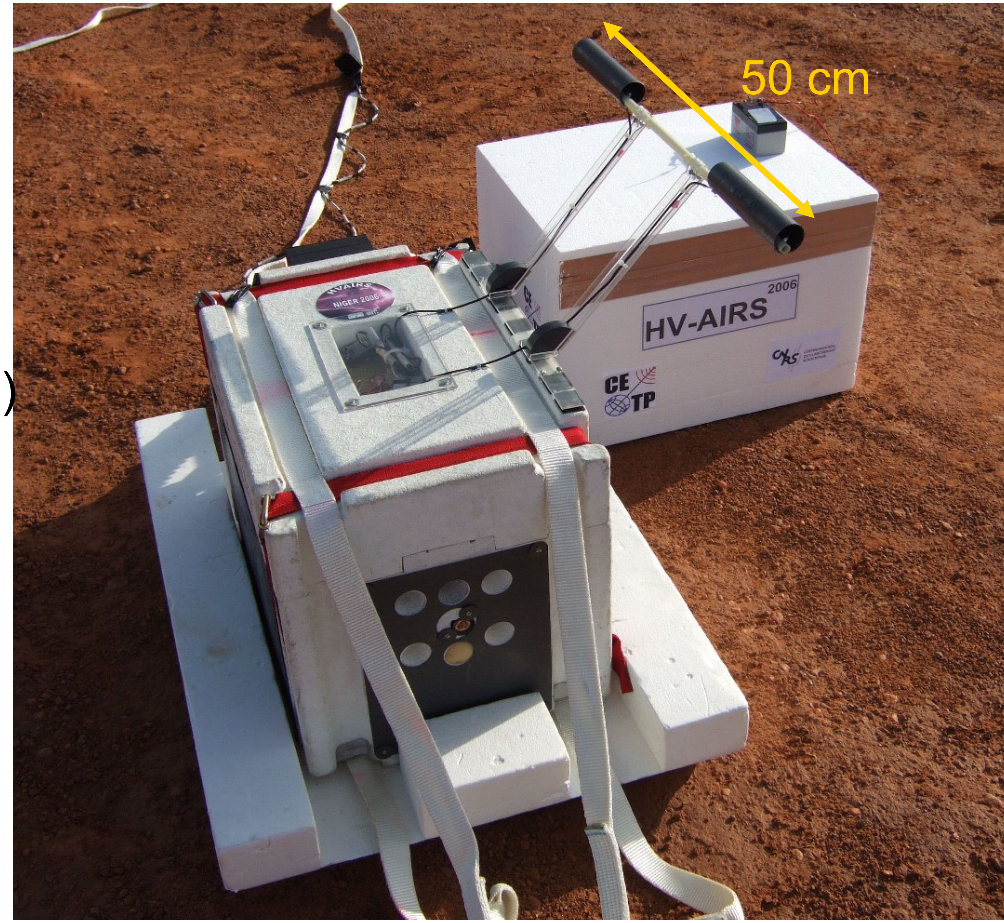

HVAIRS_AMMA, Meteorological Conditions (1)

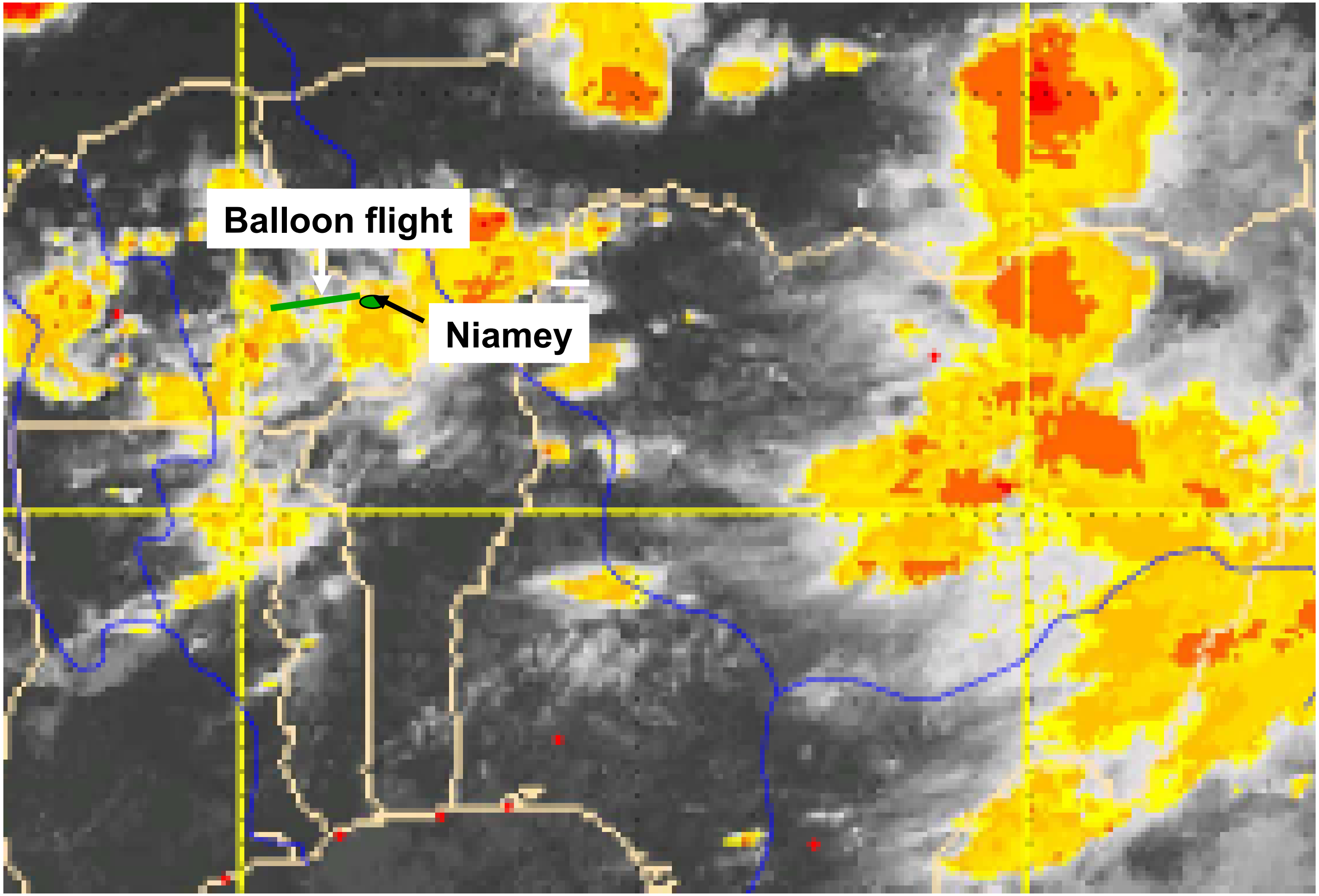
Balloon flight
Niamey

HVAIRS_AMMA, Meteorological conditions (2)

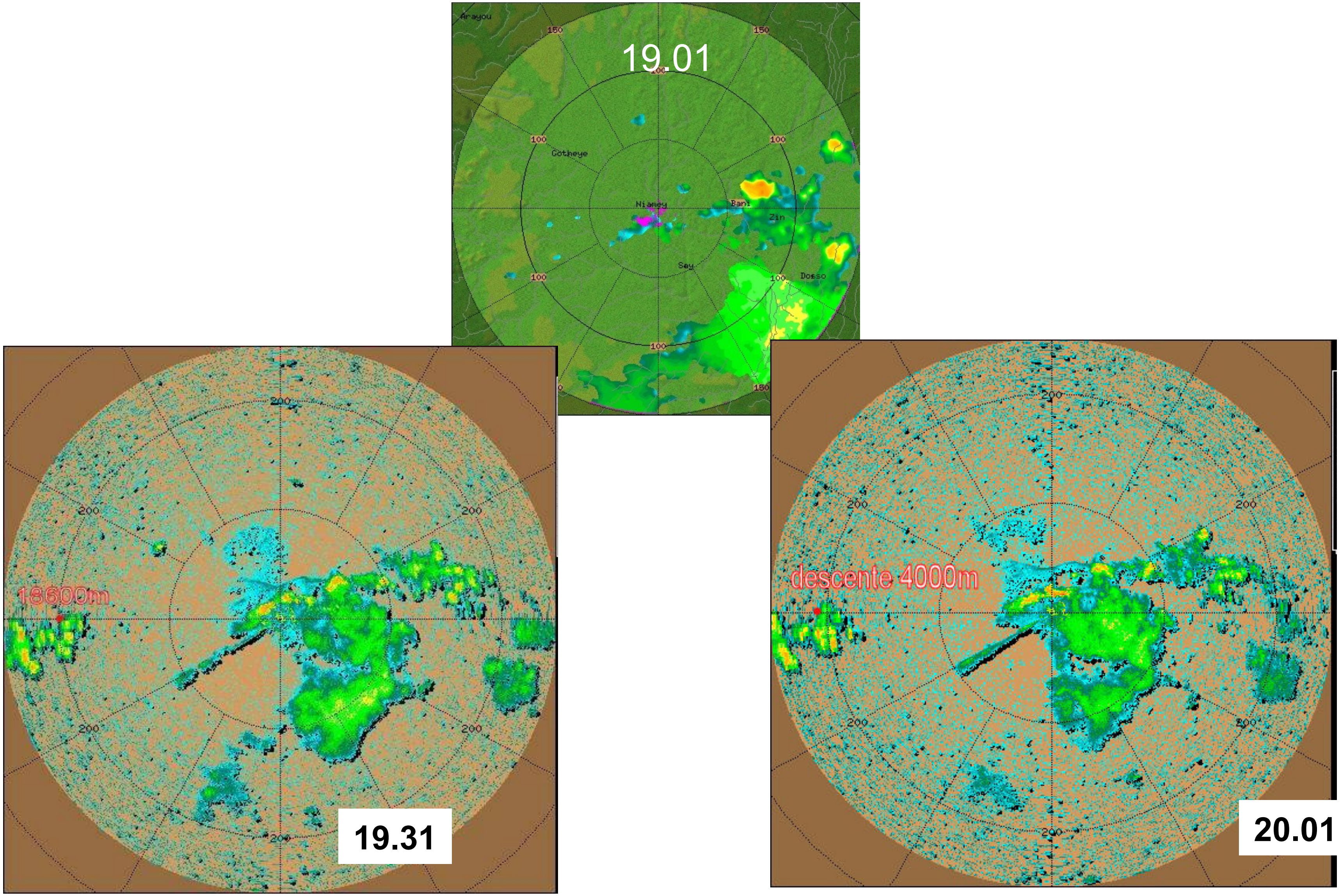
Arayou
150
150
19.01
100
100
100
Gotheye
Niamey
Bani
Zin
Say
100
100
Dosso
100
150
200
200
200
18600m
200
200
200
19.31
200
200
200
descente 4000m
200
200
200
20.01

HVAIRS_AMMA, Conductivity Measurements

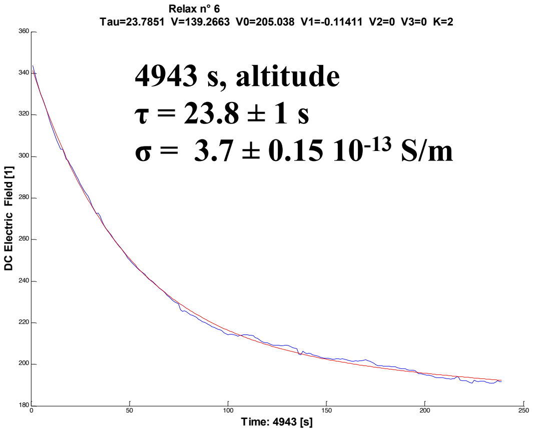

4943 s, altitude
$\tau = 23.8 \pm 1$ s
$\sigma = 3.7 \pm 0.15\ 10^{-13}$ S/m

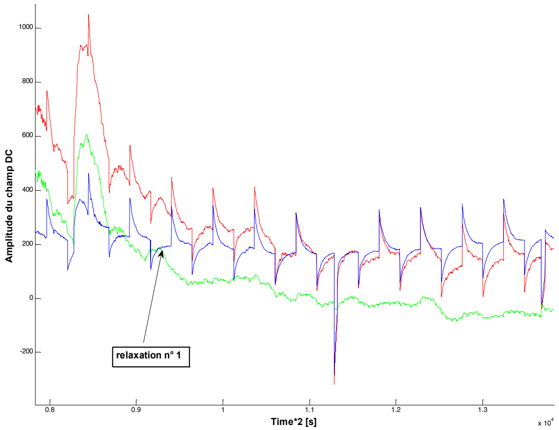

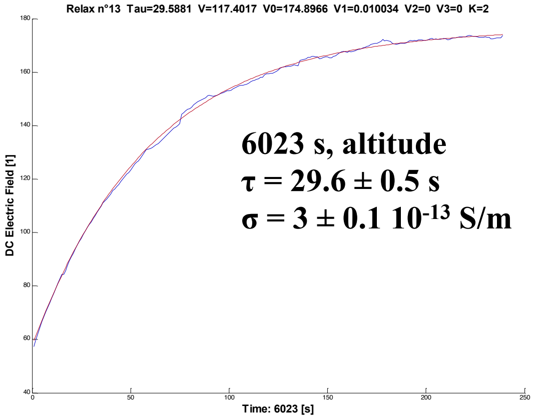

6023 s, altitude
$\tau = 29.6 \pm 0.5$ s
$\sigma = 3 \pm 0.1\ 10^{-13}$ S/m

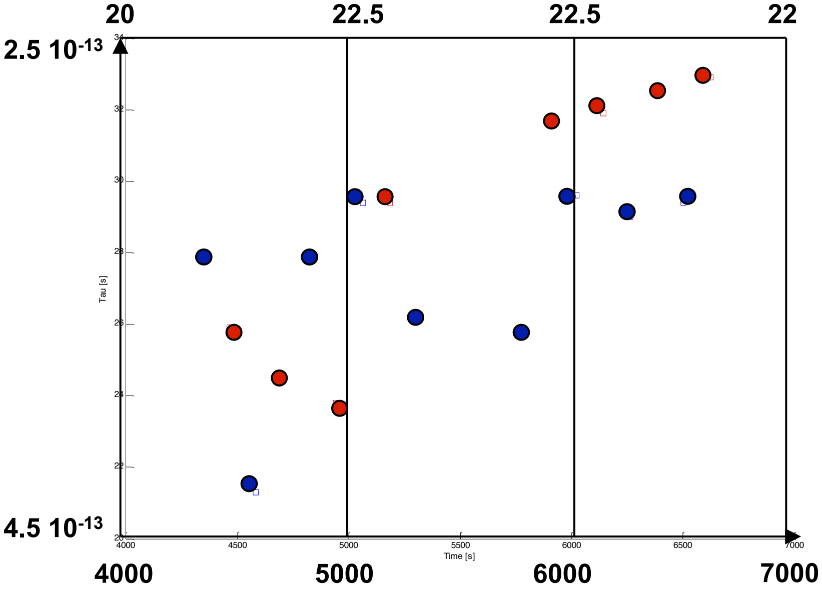

● σ- ~ 3.1 10^{-13} S/m

● σ+ ~ 3.3 10^{-13} S/m

σ ~ 6.4 10^{-13} S/m

σ ~ 2 10^{-12} S/m
R.H. Holzworth JGR, 1991

σ ~ 1.4 10^{-12} S/m
From W. Gringel, Prometheus, 1977

HVAIRS_AMMA, AC ELECTRIC FIELDS
Background noise during ascent and Schumann resonnances

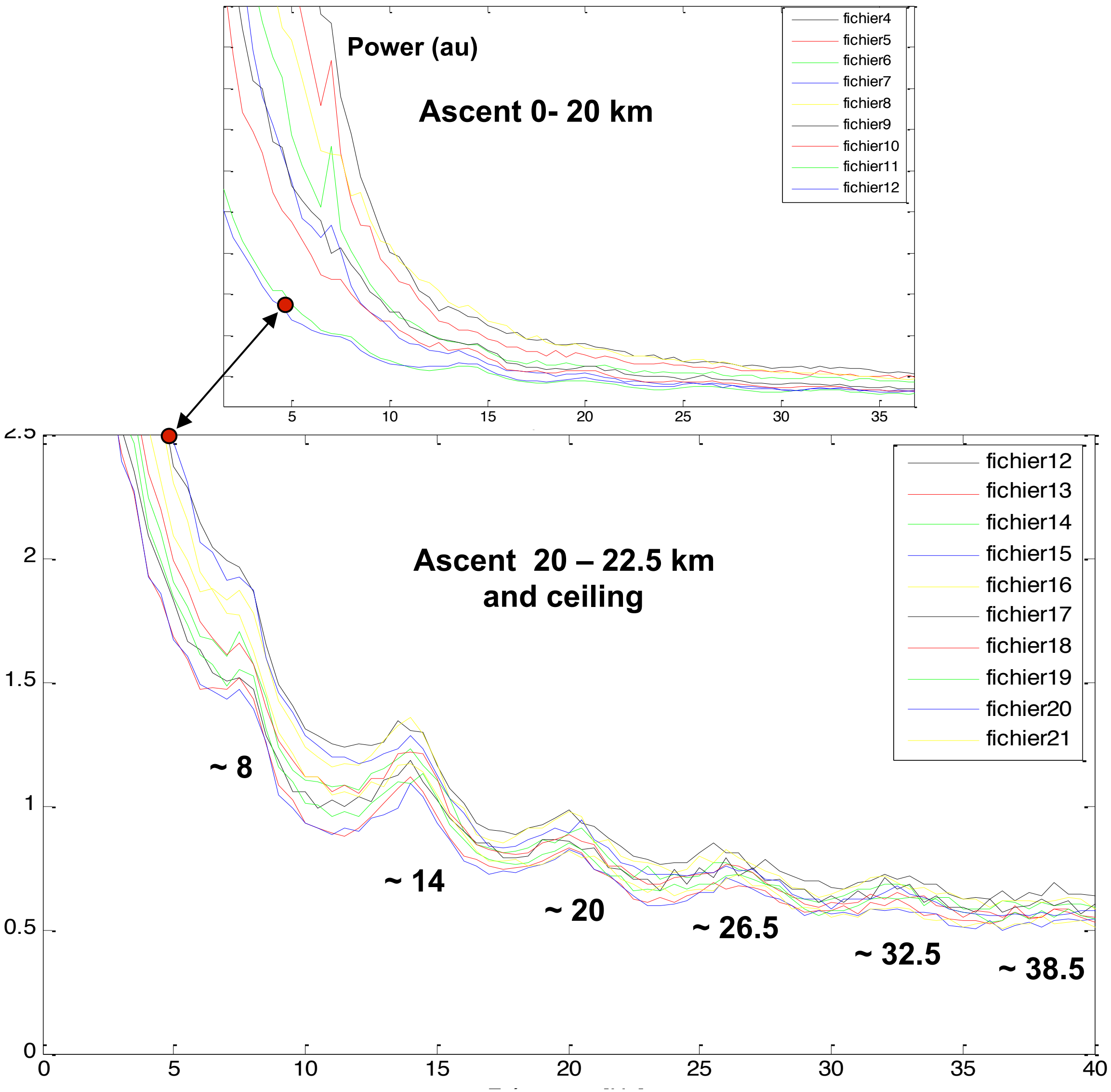

HVAIRS_AMMA, LIGHTNING and DC ELECTRIC FIELDS

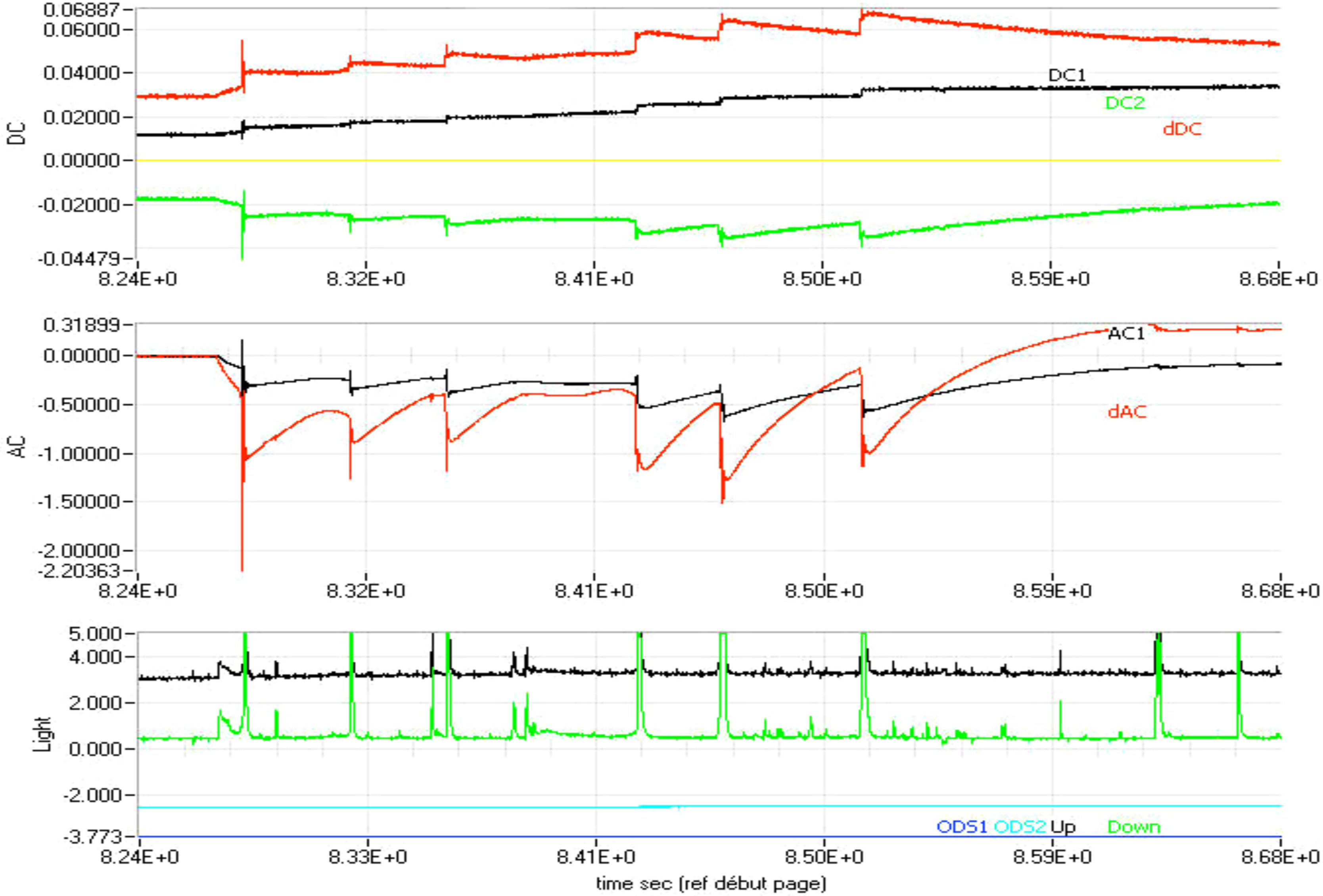

DC Electric Fields variations induced by lightning
Intra-Cloud Charge neutralization

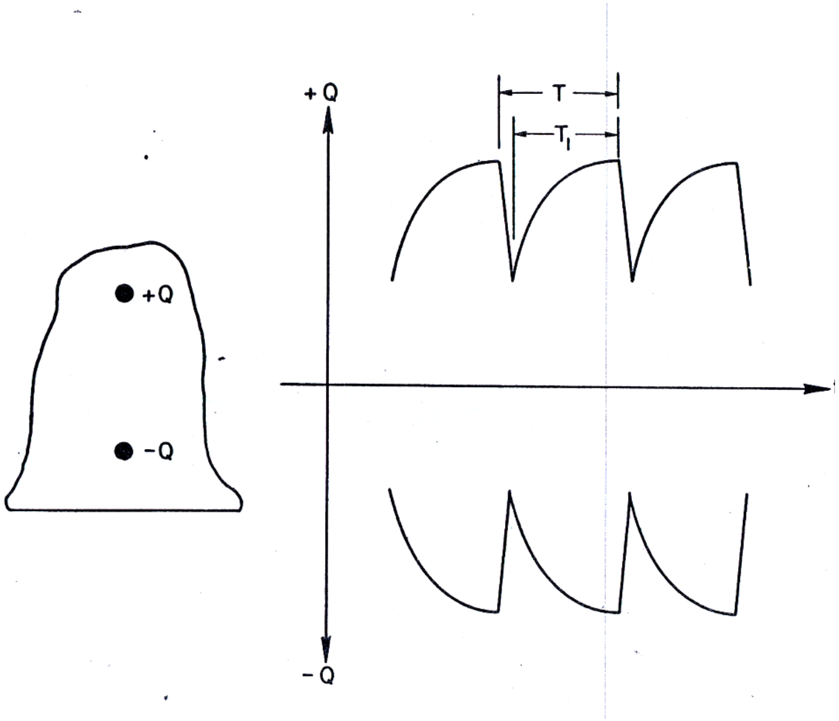

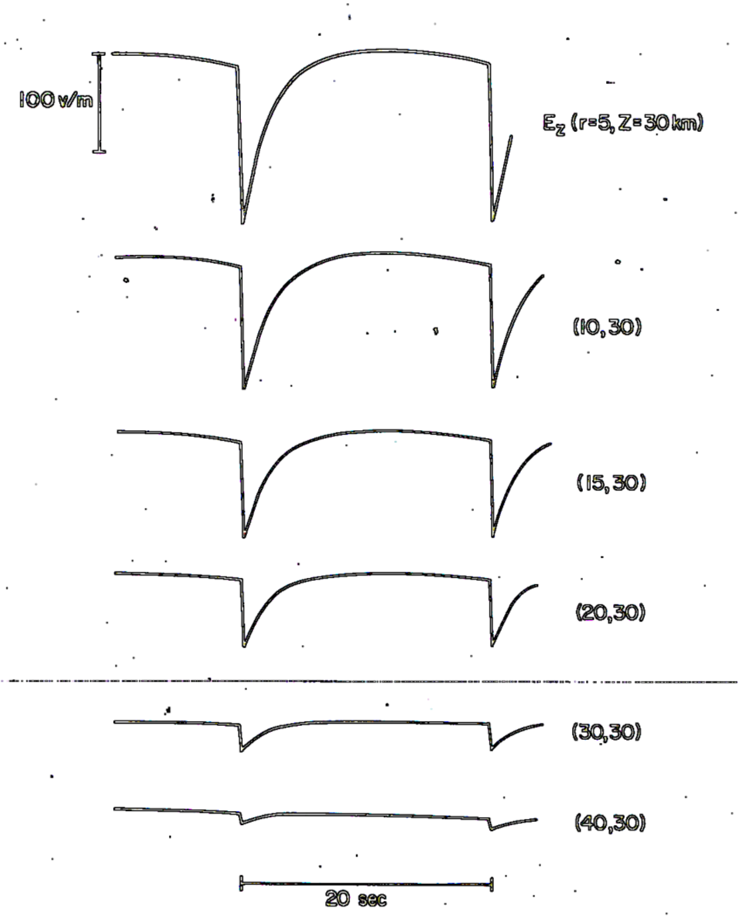

FIGURE 5-26. PLOTS OF E_z AT z = 30 km AND SELECTED VALUES OF r.

115

Dejnakarintra, 1973

HV-AIRS LIGHTNING and E-FIELD
Precursors and Continuing Currents

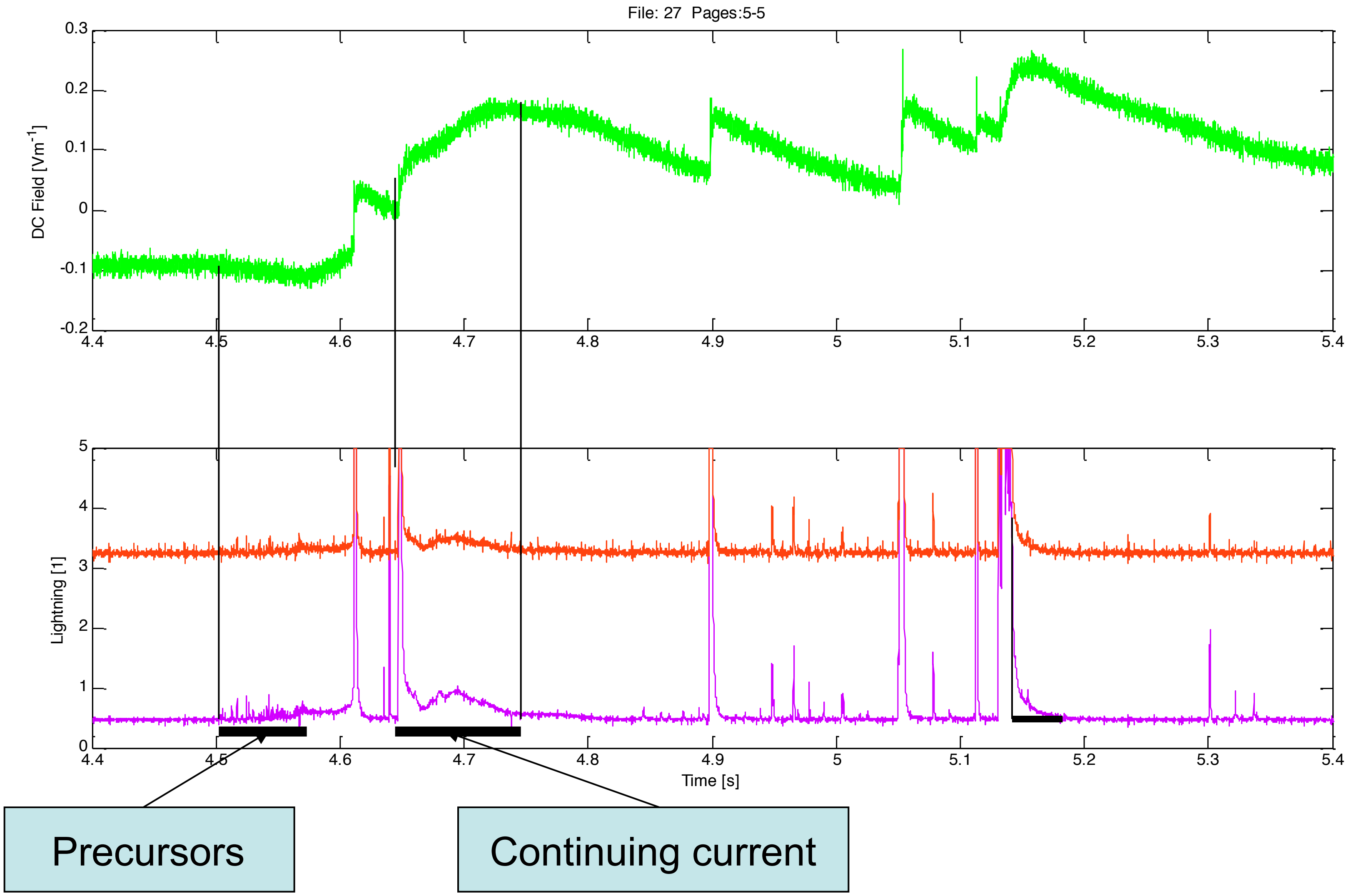

Precursors

Continuing current

HVAIRS_AMMA, DC Electric Fields
ULF signatures of stratospheric charged clouds

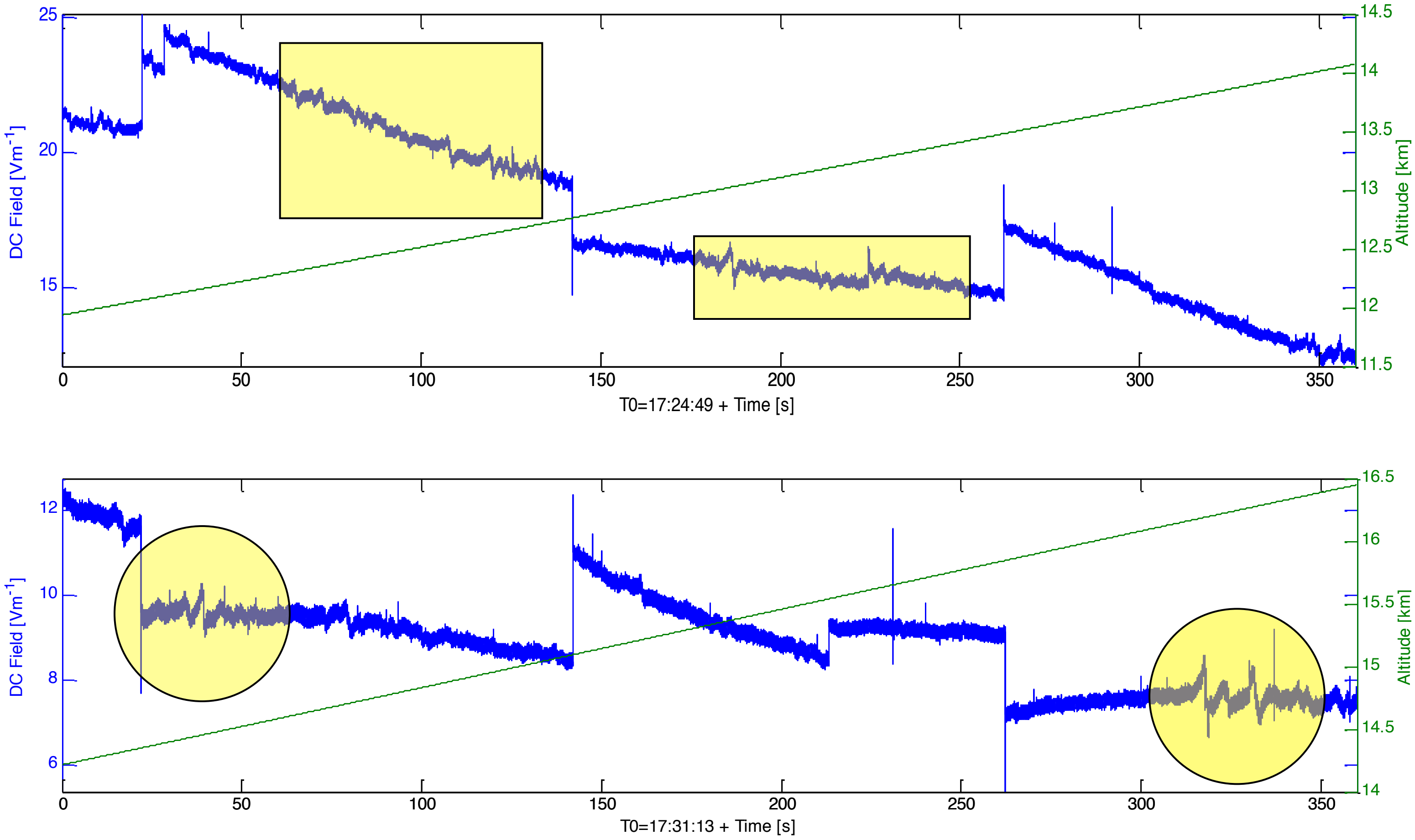

HVAIRS_AMMA, DC Electric Fields
ULF signatures of stratospheric Charged Clouds

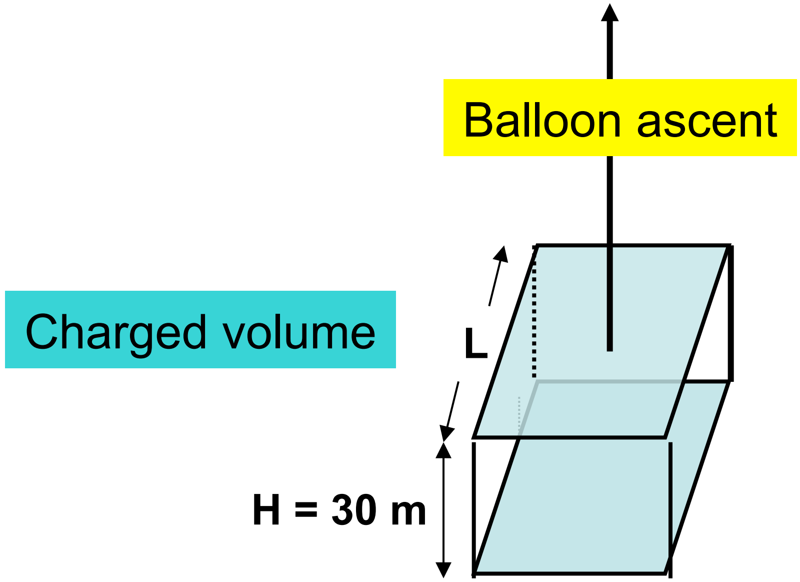

Balloon ascent

Charged volume

H = 30 m

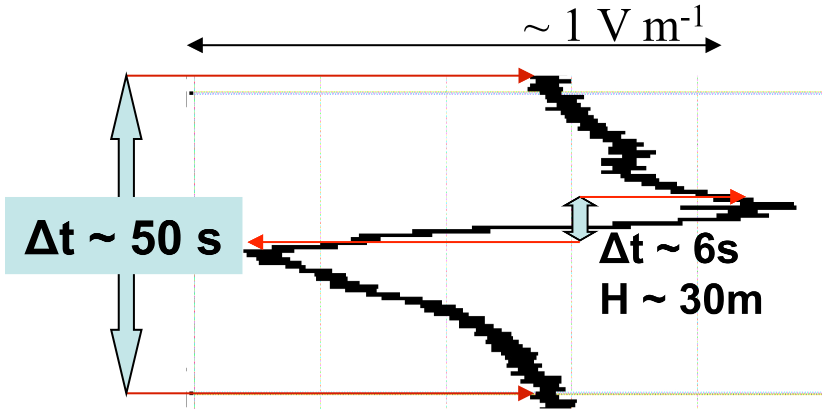

~ 1 V m^{-1}

Δt ~ 50 s

Δt ~ 6s
H ~ 30m

H ~ 30m
L ~ 50-100m
Q~ 10-30 µC

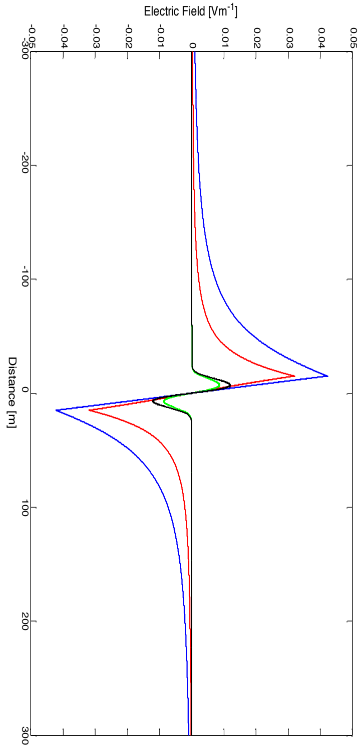

HV-AIRS AMMA, LIGHTNING, EM Pulse and Transverse resonance

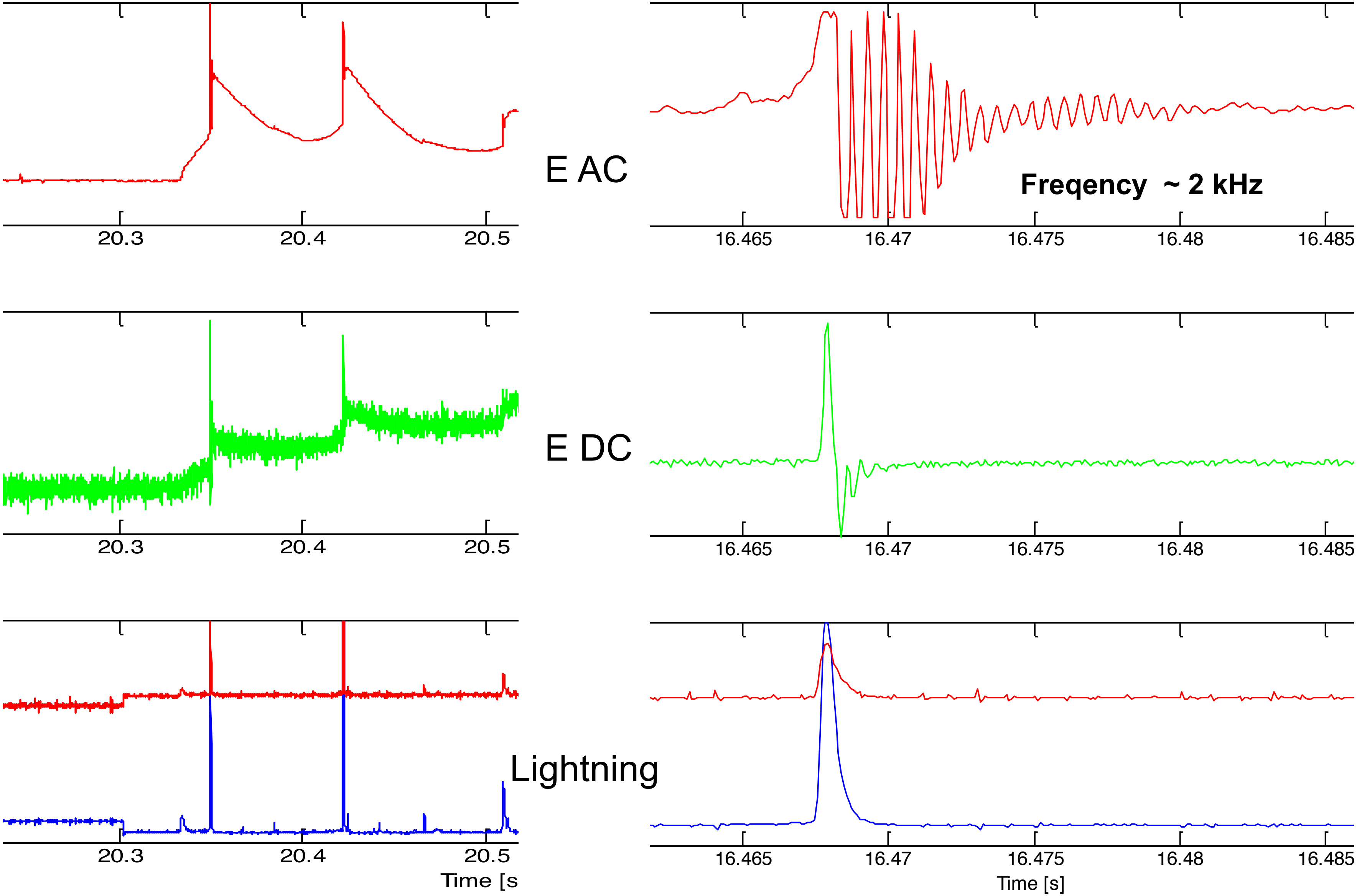

LIGHTNING DETECTION FROM ORBIT
DEMETER Observations at 650 km altitude

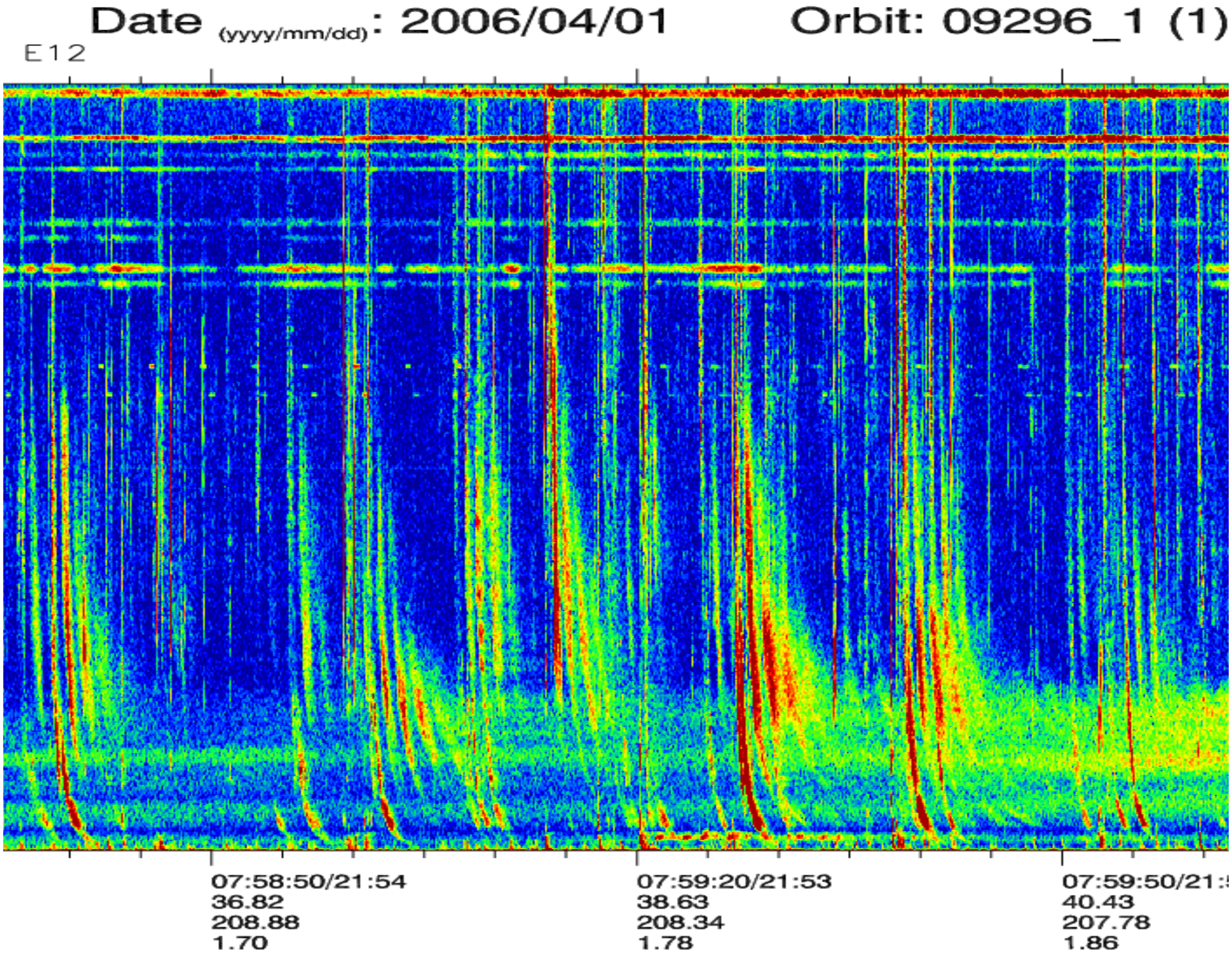

LIGHTNING DETECTION FROM ORBIT

DEMETER Observations at 650 km altitude

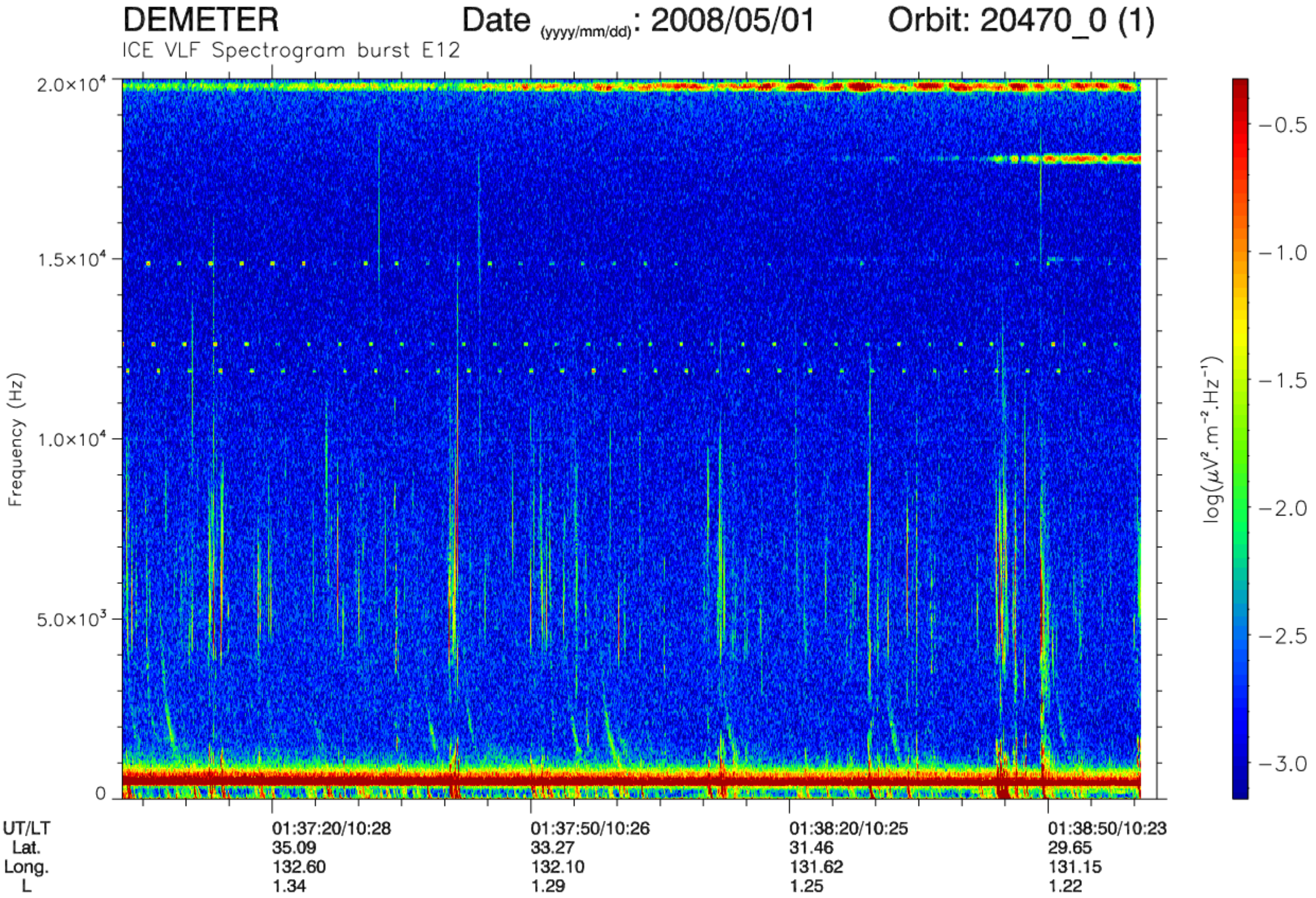

THE HASI/PWA EXPERIMENT ON HUYGENS

PWA aimed at contributing to answering the following questions:

- What are the ion and electron conductivity profiles?
- What is the role of aerosols in atmospheric chemistry?
- Is there lightning on Titan?
- Do standing waves form in the surface-ionosphere cavity?
- What are the dielectric properties of the surface?
- Does a global electric circuit exist on Titan?

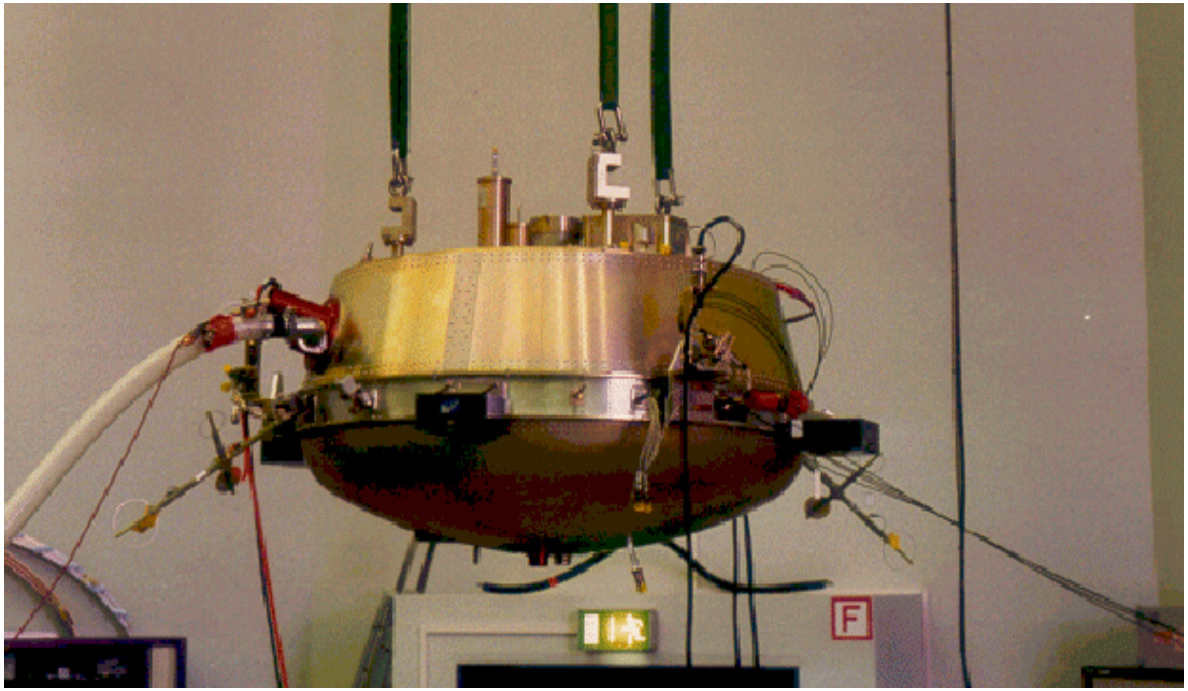

PWA ELECTRODES

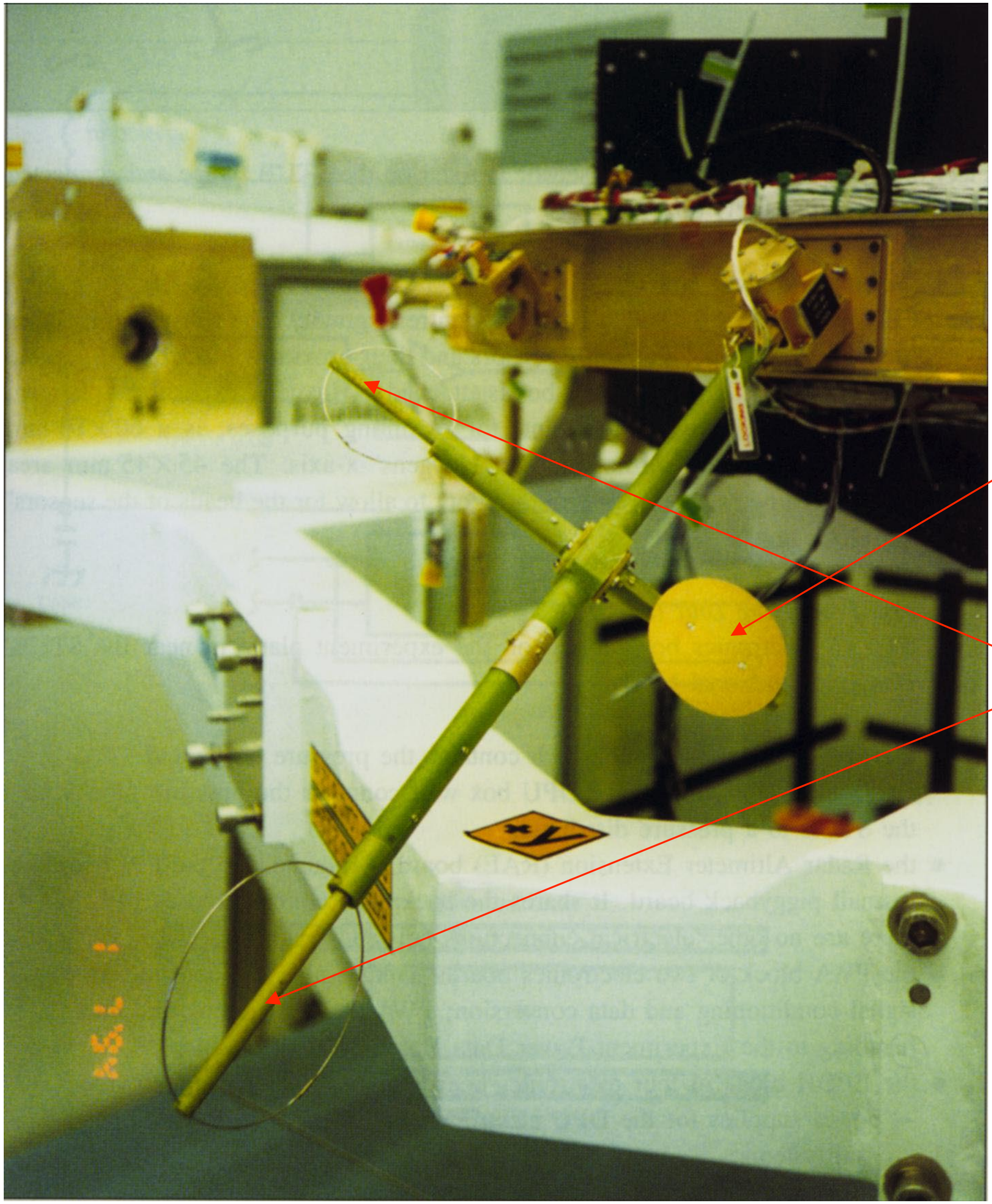

Relaxation Probe RP

Mutual Impedance Probe MIP

PLANETARY CONDUCTIVITY PROFILES: TITAN

HASI_PWA experiment on HUYGENS

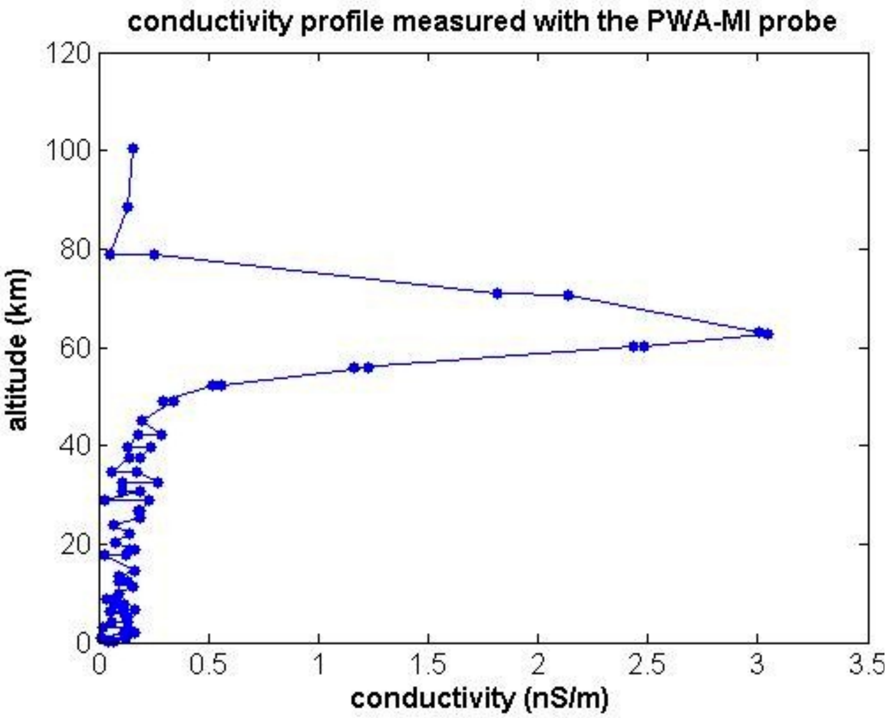

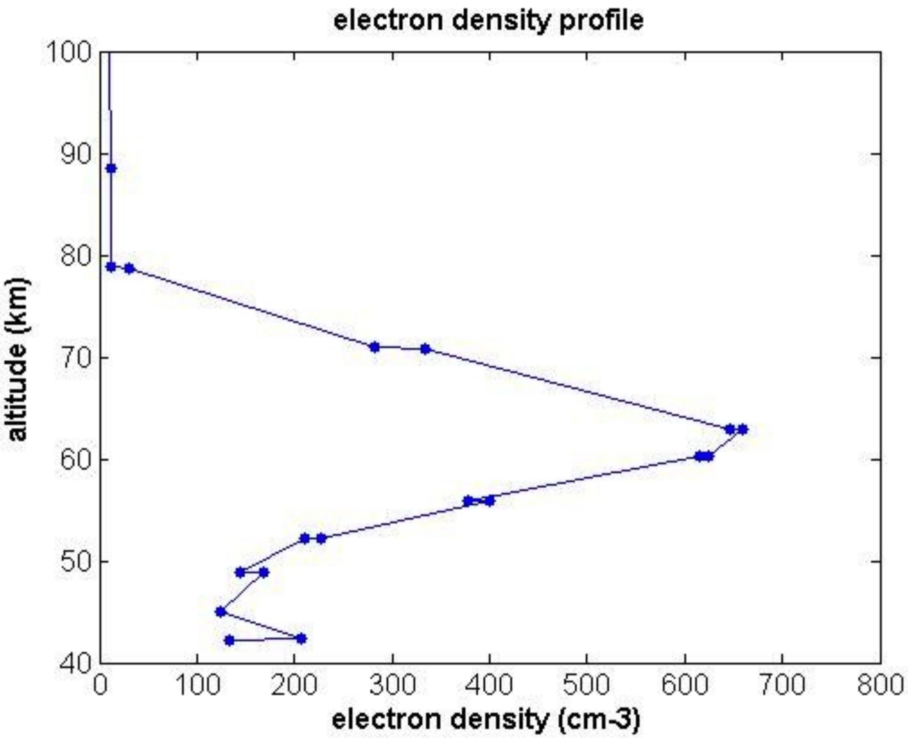

From M. Hamelin et al., PSS, 2007.

HUYGENS: Relaxation Curve with plateaus

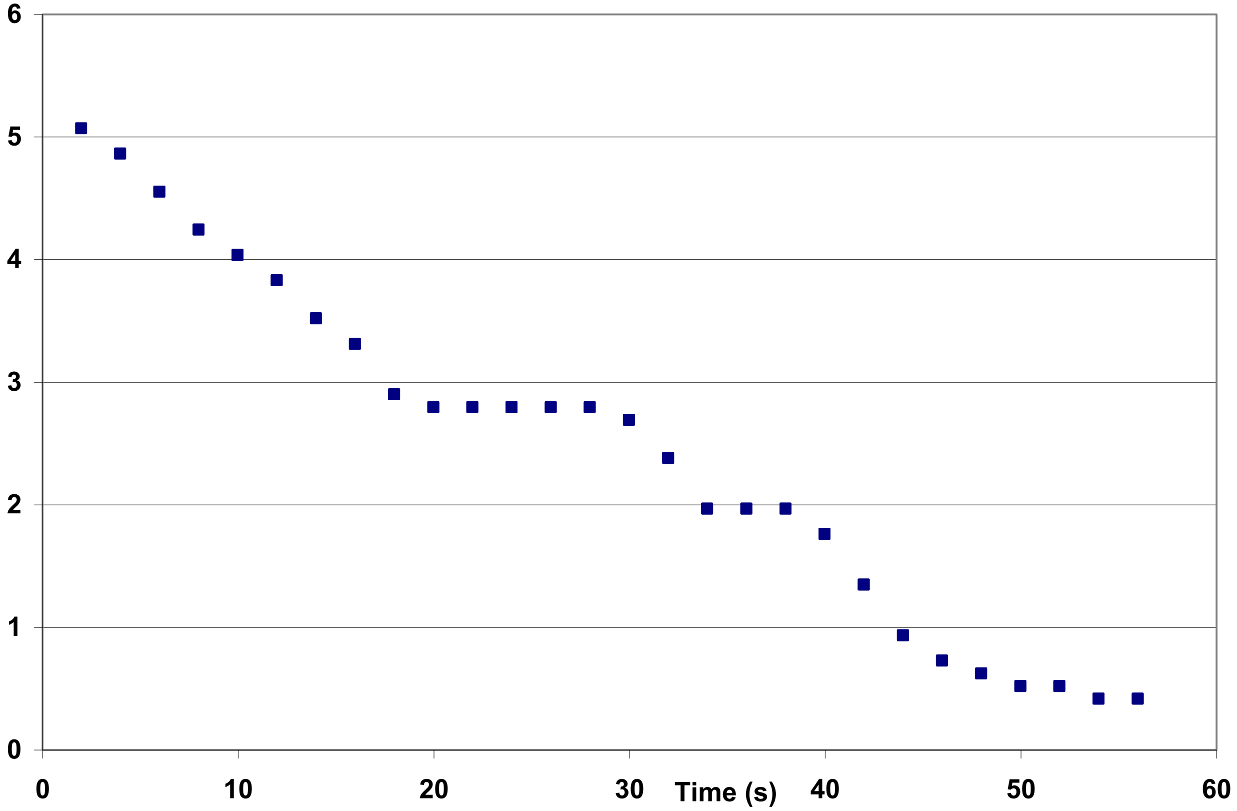

RELAXATION PROBE WAVEFORMS
DETECTION OF AEROSOL LAYER IN TITAN ATMOSPHERE

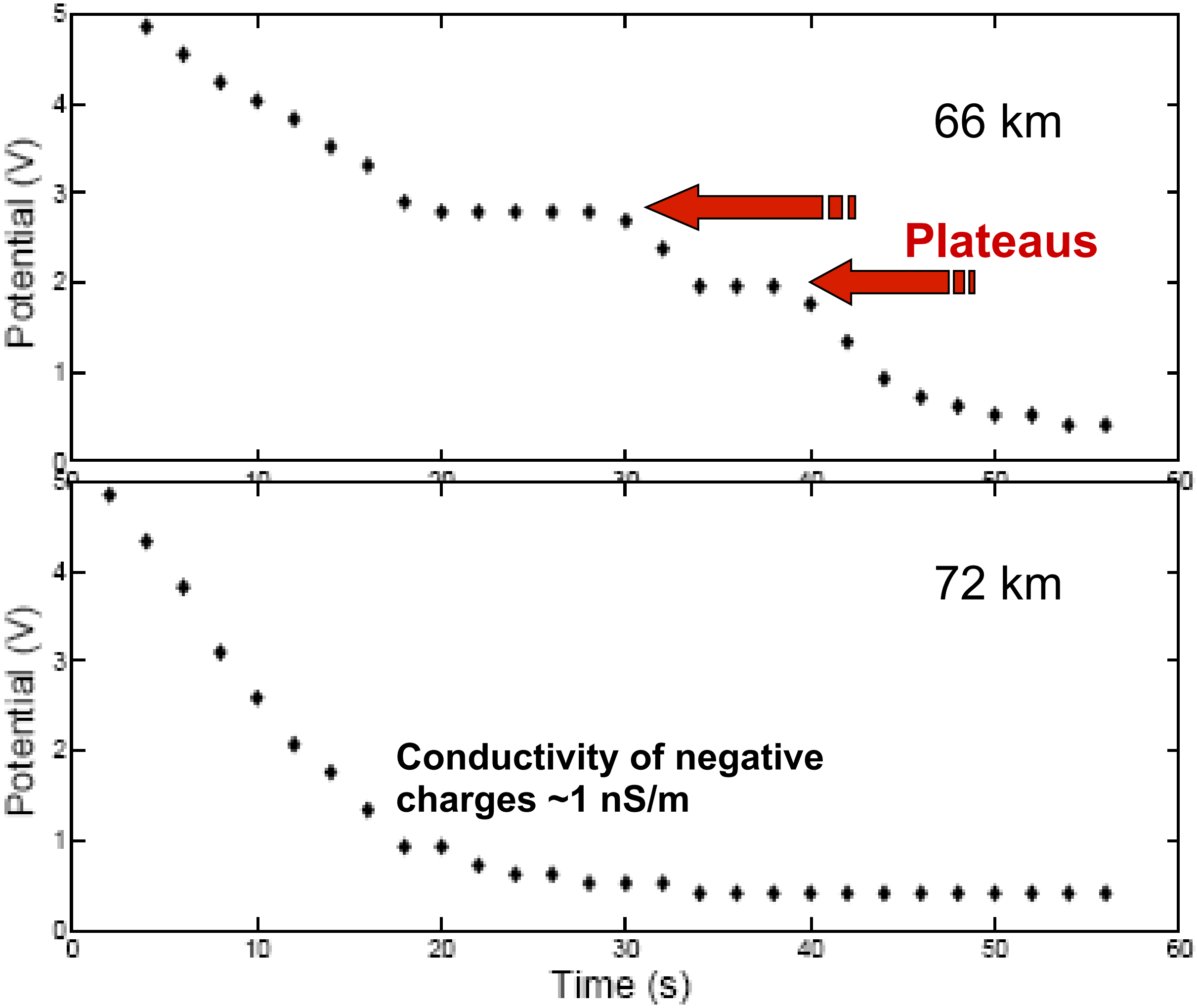
66 km

Plateaus

72 km

Conductivity of negative charges ~1 nS/m

RP measured positive ions and negative ions+electrons

No artefact has been found to explain this behaviour.

Plateaus likely correspond to absence of electrons - aerosol layers or bubbles (*ongoing work*).

Aerosol structure

Altitude [km]	**Thickness [km]**
94.5	3.3
70 69.7	0.3 0.1
62.2	0.1
57.1 54.3	0.2 0.2
56. 6 50.9	0.1 0.1

López-Moreno et al., 2008

United States Patent [19]
Eastlund

[11] **Patent Number:** **4,686,605**
[45] **Date of Patent:** **Aug. 11, 1987**

[54] **METHOD AND APPARATUS FOR ALTERING A REGION IN THE EARTH'S ATMOSPHERE, IONOSPHERE, AND/OR MAGNETOSPHERE**

[75] Inventor: **Bernard J. Eastlund**, Spring, Tex.

[73] Assignee: **APTI, Inc.**, Los Angeles, Calif.

[21] Appl. No.: **690,333**

[22] Filed: **Jan. 10, 1985**

[51] **Int. Cl.**^{4} **H05B 6/64**; H05C 3/00; H05H 1/46

[52] **U.S. Cl.** .... **361/231**; 89/1.11; 380/59; 244/158 R

[58] **Field of Search** .... 361/230, 231; 244/158 R; 376/100; 89/1.11; 380/59

[56] **References Cited**

PUBLICATIONS

Liberty Magazine, (2/35) p. 7 N. Tesla.
New York Times (9/22/40) Section 2, p. 7 W. L. Laurence.
New York Times (12/8/15) p. 8 Col. 3.

Primary Examiner—Salvatore Cangialosi
Attorney, Agent, or Firm—Roderick W. MacDonald

[57] **ABSTRACT**

A method and apparatus for altering at least one selected region which normally exists above the earth's surface. The region is excited by electron cyclotron resonance heating to thereby increase its charged particle density. In one embodiment, circularly polarized electromagnetic radiation is transmitted upward in a direction substantially parallel to and along a field line which extends through the region of plasma to be altered. The radiation is transmitted at a frequency which excites electron cyclotron resonance to heat and accelerate the charged particles. This increase in energy can cause ionization of neutral particles which are then absorbed as part of the region thereby increasing the charged particle density of the region.

15 Claims, 5 Drawing Figures

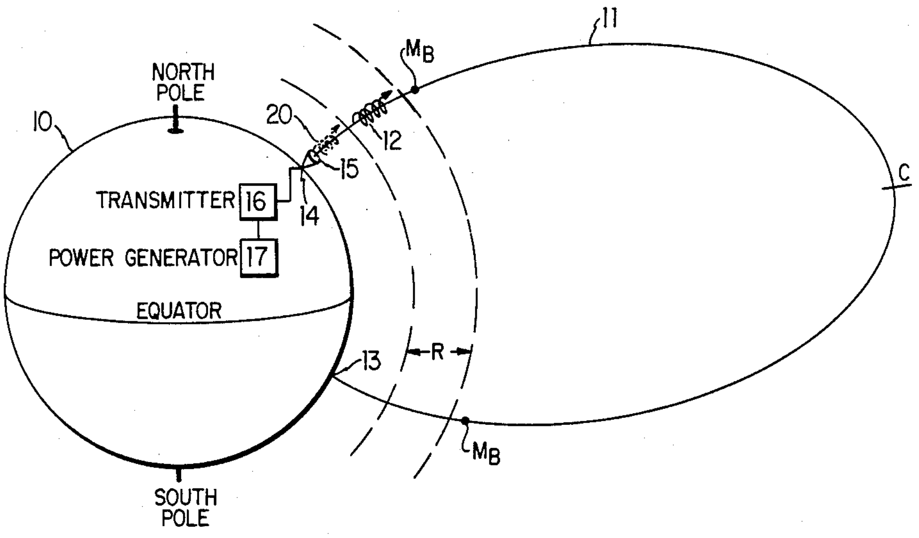

FIG. 1

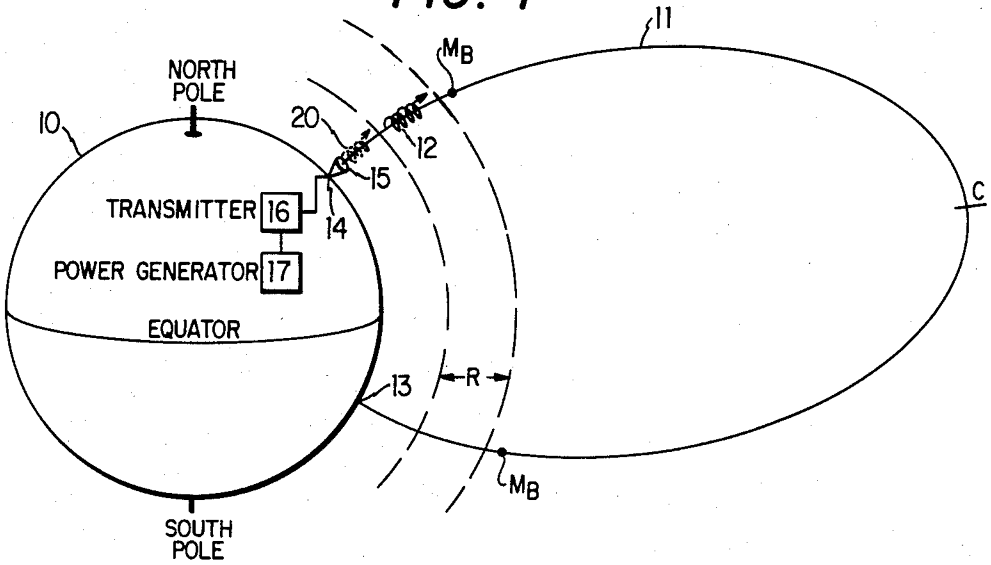
11
MB
NORTH POLE
10
20
12
15
TRANSMITTER 16
14
POWER GENERATOR 17
C
EQUATOR
13
R
MB
SOUTH POLE

FIG. 2

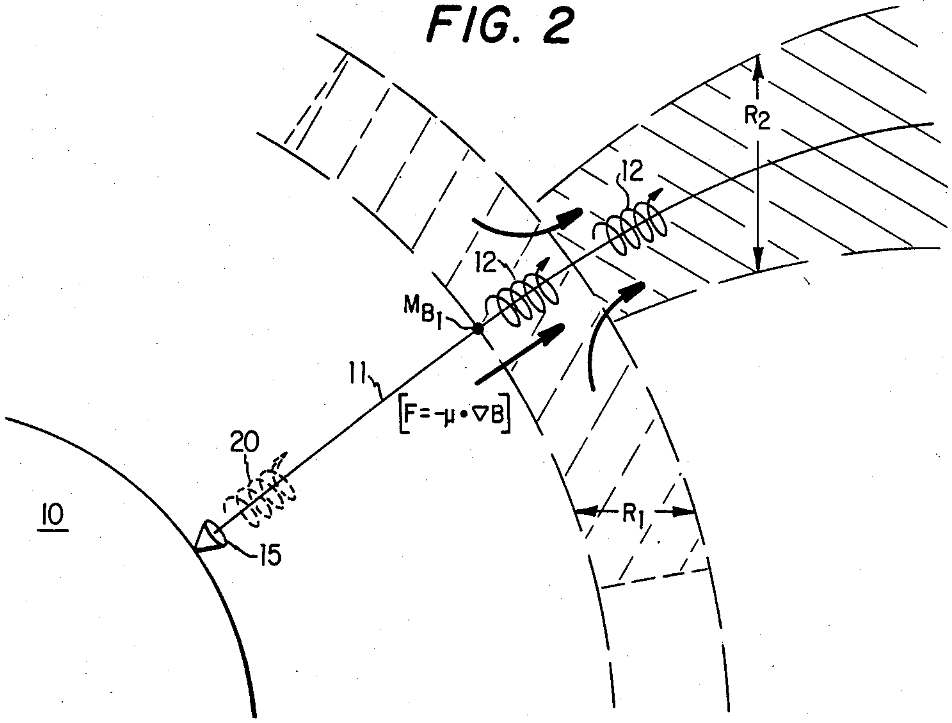
R2
MB1
[F=-μ•∇B]
R1

FIG. 3

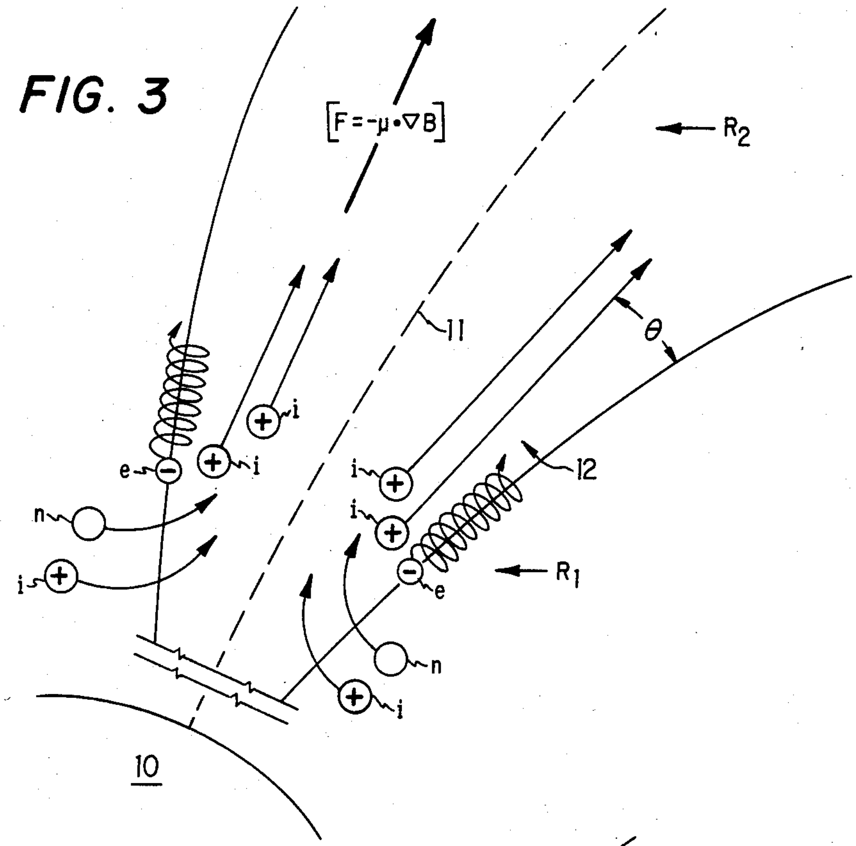
[F=-μ•∇B]
R2
11
θ
i
e
i
12
n
i
i
R1
i
e
n
i
10

FIG. 4

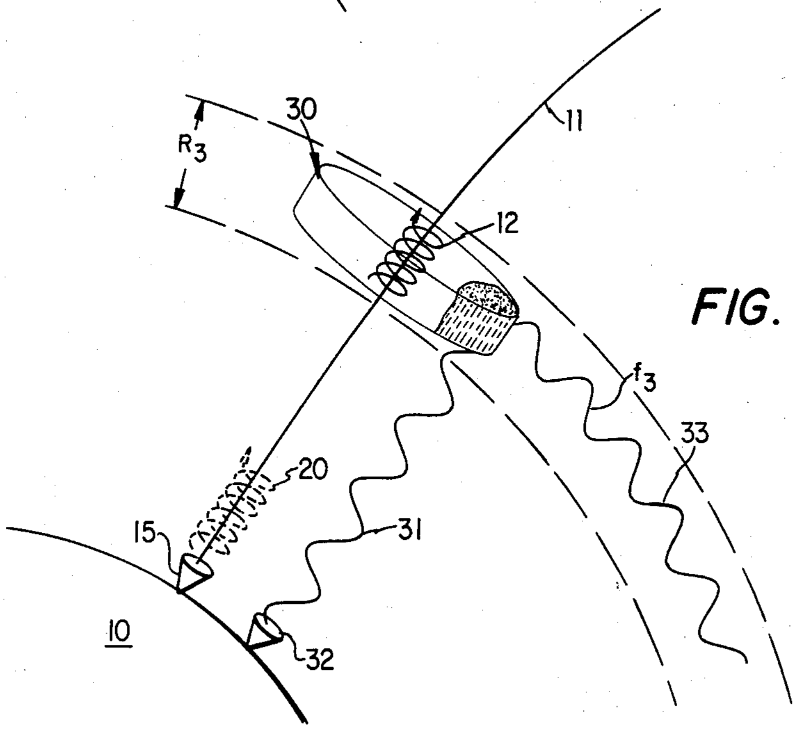
30
R3
11
12
f3
33
20
15
31
10
32

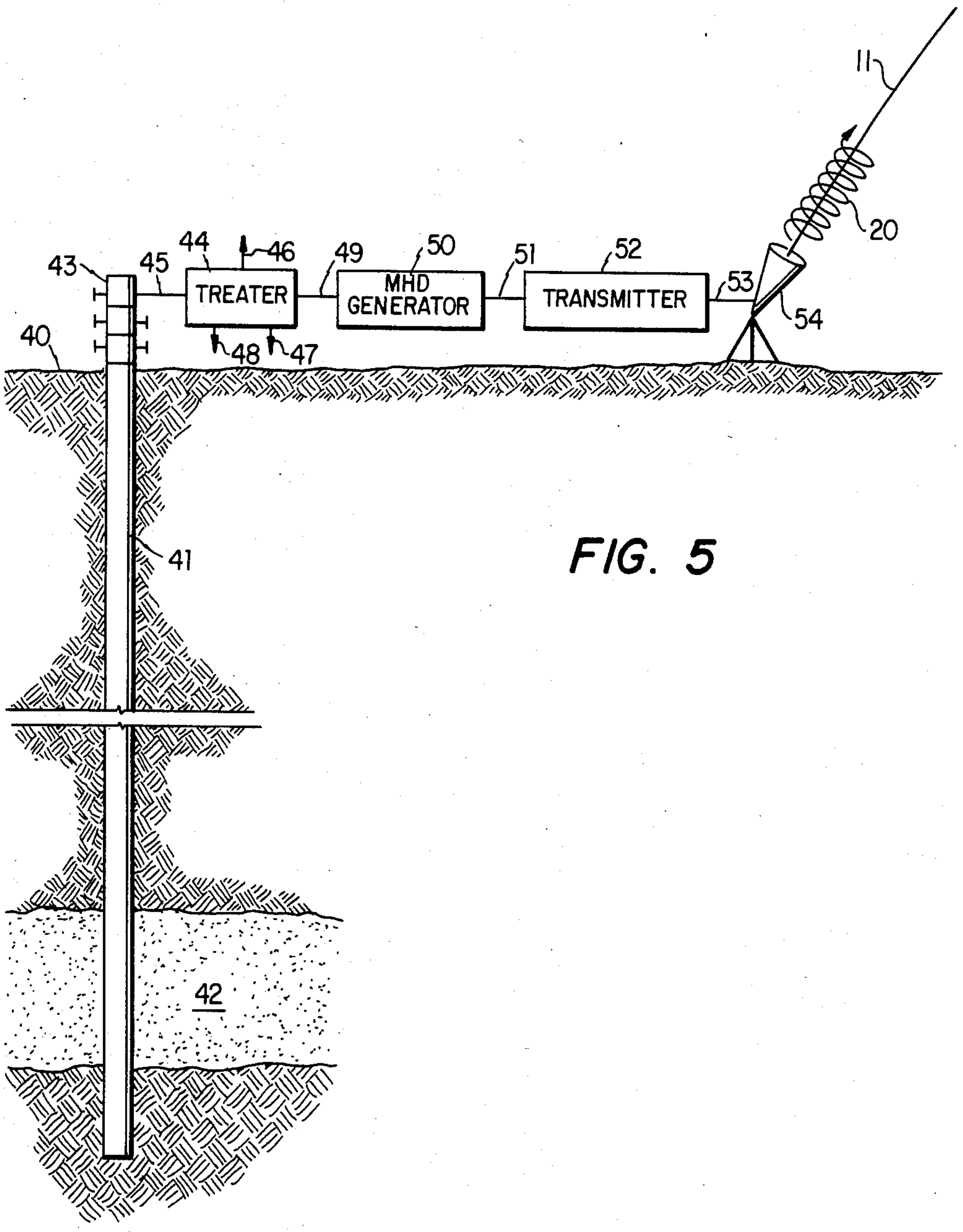
TREATER
MHD GENERATOR
TRANSMITTER

FIG. 5

METHOD AND APPARATUS FOR ALTERING A REGION IN THE EARTH'S ATMOSPHERE, IONOSPHERE, AND/OR MAGNETOSPHERE

DESCRIPTION

1. Technical Field

This invention relates to a method and apparatus for altering at least one selected region normally existing above the earth's surface and more particularly relates to a method and apparatus for altering said at least one region by initially transmitting electromagnetic radiation from the earth's surface essentially parallel to and along naturally-occurring, divergent magnetic field lines which extend from the earth's surface through the region or regions to be altered.

2. Background Art

In the late 1950's, it was discovered that naturally-occuring belts exist at high altitudes above the earth's surface, and it is now established that these belts result from charged electrons and ions becoming trapped along the magnetic lines of force (field lines) of the earth's essentially dipole magnetic field. The trapped electrons and ions are confined along the field lines between two magnetic mirrors which exist at spaced apart points along those field lines. The trapped electrons and ions move in helical paths around their particular field lines and "bounce" back and forth between the magnetic mirrors. These trapped electrons and ions can oscillate along the field lines for long periods of time.

In the past several years, substantial effort has been made to understand and explain the phenomena involved in belts of trapped electrons and ions, and to explore possible ways to control and use these phenomena for beneficial purposes. For example, in the late 1950's and early 1960's both the United States and U.S.S.R. detonated a series of nuclear devices of various yields to generate large numbers of charged particles at various altitudes, e.g., 200 kilometers (km) or greater. This was done in order to establish and study artifical belts of trapped electrons and ions. These experiments established that at least some of the extraneous electrons and ions from the detonated devices did become trapped along field lines in the earth's magnetosphere to form artificial belts which were stable for prolonged periods of time. For a discussion of these experiments see "The Radiation Belt and Magnetosphere", W. N. Hess, Blaisdell Publishing Co., 1968, pps. 155 et sec.

Other proposals which have been advanced for altering existing belts of trapped electrons and ions and/or establishing similar artificial belts include injecting charged particles from a satellite carrying a payload of radioactive beta-decay material or alpha emitters; and injecting charged particles from a satellite-borne electron accelerator. Still another approach is described in U.S. Pat. No. 4,042,196 wherein a low energy ionized gas, e.g., hydrogen, is released from a synchronous orbiting satellite near the apex of a radiation belt which is naturally-occurring in the earth's magnetosphere to produce a substantial increase in energetic particle precipitation and, under certain conditions, produce a limit in the number of particles that can be stably trapped. This precipitation effect arises from an enhancement of the whistler-mode and ion-cyclotron mode interactions that result from the ionized gas or "cold plasma" injection.

It has also been proposed to release large clouds of barium in the magnetosphere so that photoionization will increase the cold plasma density, thereby producing electron precipitation through enhanced whistler-mode interactions.

However, in all of the above-mentioned approaches, the mechanisms involved in triggering the change in the trapped particle phenomena must be actually positioned within the affected zone, e.g., the magnetosphere, before they can be actuated to effect the desired change.

The earth's ionosphere is not considered to be a "trapped" belt since there are few trapped particles therein. The term "trapped" herein refers to situations where the force of gravity on the trapped particles is balanced by magnetic forces rather than hydrostatic or collisional forces. The charged electrons and ions in the ionosphere also follow helical paths around magnetic field lines within the ionosphere but are not trapped between mirrors, as in the case of the trapped belts in the magnetosphere, since the gravitational force on the particles is balanced by collisional or hydrostatic forces.

In recent years, a number of experiments have actually been carried out to modify the ionosphere in some controlled manner to investigate the possibility of a beneficial result. For detailed discussions of these operations see the following papers: (1) Ionospheric Modification Theory; G. Meltz and F. W. Perkins; (2) The Platteville High Power Facility; Carrol et al.; (3) Arecibo Heating Experiments; W. E. Gordon and H. C. Carlson, Jr.; and (4) Ionospheric Heating by Powerful Radio Waves; Meltz et al., all published in Radio Science, Vol. 9, No. 11, November, 1974, at pages 885–888; 889–894; 1041–1047; and 1049–1063, respectively, all of which are incorporated herein by reference. In such experiments, certain regions of the ionosphere are heated to change the electron density and temperature within these regions. This is accomplished by transmitting from earth-based antennae high frequency electromagnetic radiation at a substantial angle to, not parallel to, the ionosphere's magnetic field to heat the ionospheric particles primarily by ohmic heating. The electron temperature of the ionosphere has been raised by hundreds of degrees in these experiments, and electrons with several electron volts of energy have been produced in numbers sufficient to enhance airglow. Electron concentrations have been reduced by a few percent, due to expansion of the plasma as a result of increased temperature.

In the Elmo Bumpy Torus (EBT), a controlled fusion device at the Oak Ridge National Laboratory, all heating is provided by microwaves at the electron cyclotron resonance interaction. A ring of hot electrons is formed at the earth's surface in the magnetic mirror by a combination of electron cyclotron resonance and stochastic heating. In the EBT, the ring electrons are produced with an average "temperature" of 250 kilo electron volts or kev (2.5×10^9K) and a plasma beta between 0.1 and 0.4; see, "A Theoretical Study of Electron—Cyclotron Absorption in Elmo Bumpy Torus", Batchelor and Goldfinger, Nuclear Fusion, Vol. 20, No. 4 (1980) pps. 403–418.

Electron cyclotron resonance heating has been used in experiments on the earth's surface to produce and accelerate plasmas in a diverging magnetic field. Kosmahl et al. showed that power was transferred from the electromagnetic waves and that a fully ionized plasma

was accelerated with a divergence angle of roughly 13 degrees. Optimum neutral gas density was 1.7×10^{14}per cubic centimeter; see, "Plasma Acceleration with Microwaves Near Cyclotron Resonance", Kosmahl et al., Journal of Applied Physics, Vol. 38, No. 12, Nov., 1967, pps. 4576–4582.

DISCLOSURE OF THE INVENTION

The present invention provides a method and apparatus for altering at least one selected region which normally exists above the earth's surface. The region is excited by electron cyclotron resonance heating of electrons which are already present and/or artifically created in the region to thereby increase the charged particle energy and ultimately the density of the region.

In one embodiment this is done by transmitting circularly polarized electromagnetic radiation from the earth's surface at or near the location where a naturally-occurring dipole magnetic field (force) line intersects the earth's surface. Right hand circular polarization is used in the northern hemisphere and left hand circular polarization is used in the southern hemisphere. The radiation is deliberately transmitted at the outset in a direction substantially parallel to and along a field line which extends upwardly through the region to be altered. The radiation is transmitted at a frequency which is based on the gyrofrequency of the charged particles and which, when applied to the at least one region, excites electron cyclotron resonance within the region or regions to heat and accelerate the charged particles in their respective helical paths around and along the field line. Sufficient energy is employed to cause ionization of neutral particles (molecules of oxygen, nitrogen and the like, particulates, etc.) which then become a part of the region thereby increasing the charged particle density of the region. This effect can further be enhanced by providing artificial particles, e.g., electrons, ions, etc., directly into the region to be affected from a rocket, satellite, or the like to supplement the particles in the naturally-occurring plasma. These artificial particles are also ionized by the transmitted electromagnetic radiation thereby increasing charged particle density of the resulting plasma in the region.

In another embodiment of the invention, electron cyclotron resonance heating is carried out in the selected region or regions at sufficient power levels to allow a plasma present in the region to generate a mirror force which forces the charged electrons of the altered plasma upward along the force line to an altitude which is higher than the original altitude. In this case the relevant mirror points are at the base of the altered region or regions. The charged electrons drag ions with them as well as other particles that may be present. Sufficient power, e.g., 10^{15}joules, can be applied so that the altered plasma can be trapped on the field line between mirror points and will oscillate in space for prolonged periods of time. By this embodiment, a plume of altered plasma can be established at selected locations for communication modification or other purposes.

In another embodiment, this invention is used to alter at least one selected region of plasma in the ionosphere to establish a defined layer of plasma having an increased charged particle density. Once this layer is established, and while maintaining the transmission of the main beam of circularly polarized electromagnetic radiation, the main beam is modulated and/or at least one second different, modulated electromagnetic radiation beam is transmitted from at least one separate source at a different frequency which will be absorbed in the plasma layer. The amplitude of the frequency of the main beam and/or the second beam or beams is modulated in resonance with at least one known oscillation mode in the selected region or regions to excite the known oscillation mode to propagate a known frequency wave or waves throughout the ionosphere.

BRIEF DESCRIPTION OF THE DRAWINGS

The actual construction, operation, and apparent advantages of this invention will be better understood by referring to the drawings in which like numerals identify like parts and in which:

FIG. **1** is a simplified schematical view of the earth (not to scale) with a magnetic field (force) line along which the present invention is carried out;

FIG. **2** is one embodiment within the present invention in which a selected region of plasma is raised to a higher altitude;

FIG. **3** is a simplified, idealized representation of a physical phenomenon involved in the present invention; and

FIG. **4** is a schematic view of another embodiment within the present invention.

FIG. **5** is a schematic view of an apparatus embodiment within this invention.

BEST MODES FOR CARRYING OUT THE INVENTION

The earth's magnetic field is somewhat analogous to a dipole bar magnet. As such, the earth's magnetic field contains numerous divergent field or force lines, each line intersecting the earth's surface at points on opposite sides of the Equator. The field lines which intersect the earth's surface near the poles have apexes which lie at the furthest points in the earth's magnetosphere while those closest to the Equator have apexes which reach only the lower portion of the magnetosphere.

At various altitudes above the earth's surface, e.g., in both the ionosphere and the magnetosphere, plasma is naturally present along these field lines. This plasma consists of equal numbers of positively and negatively charged particles (i.e., electrons and ions) which are guided by the field line. It is well established that a charged particle in a magnetic field gyrates about field lines, the center of gyration at any instance being called the "guiding center" of the particle. As the gyrating particle moves along a field line in a uniform field, it will follow a helical path about its guiding center, hence linear motion, and will remain on the field line. Electrons and ions both follow helical paths around a field line but rotate in opposite directions. The frequencies at which the electrons and ions rotate about the field line are called gyromagnetic frequencies or cyclotron frequencies because they are identical with the expression for the angular frequencies of gyration of particles in a cyclotron. The cyclotron frequency of ions in a given magnetic field is less than that of electrons, in inverse proportion to their masses.

If the particles which form the plasma along the earth's field lines continued to move with a constant pitch angle, often designated "alpha", they would soon impact on the earth's surface. Pitch angle alpha is defined as the angle between the direction of the earth's magnetic field and the velocity (V) of the particle. However, in converging force fields, the pitch angle does change in such a way as to allow the particle to

turn around and avoid impact. Consider a particle moving along a field line down toward the earth. It moves into a region of increasing magnetic field strength and therefore sine alpha increases. But sine alpha can only increase to 1.0, at which point, the particle turns around and starts moving up along the field line, and alpha decreases. The point at which the particle turns around is called the mirror point, and there alpha equals ninety degrees. This process is repeated at the other end of the field line where the same magnetic field strength value B, namely Bm, exists. The particle again turns around and this is called the "conjugate point" of the original mirror point. The particle is therefore trapped and bounces between the two magnetic mirrors. The particle can continue oscillating in space in this manner for long periods of time. The actual place where a particle will mirror can be calculated from the following:

$$\sin^2 \text{alpha}_o = B_o/B_m \qquad (1)$$

wherein:

alpha_o = equatorial pitch angle of particle

B_o = equatorial field strength on a particular field line

B_m = field strength at the mirror point

Recent discoveries have established that there are substantial regions of naturally trapped particles in space which are commonly called "trapped radiation belts". These belts occur at altitudes greater than about 500 km and accordingly lie in the magnetosphere and mostly above the ionosphere.

The ionosphere, while it may overlap some of the trapped-particle belts, is a region in which hydrostatic forces govern its particle distribution in the gravitational field. Particle motion within the ionosphere is governed by both hydrodynamic and electrodynamic forces. While there are few trapped particles in the ionosphere, nevertheless, plasma is present along field lines in the ionosphere. The charged particles which form this plasma move between collisions with other particles along similar helical paths around the field lines and although a particular particle may diffuse downward into the earth's lower atmosphere or lose energy and diverge from its original field line due to collisions with other particles, these charged particles are normally replaced by other available charged particles or by particles that are ionized by collision with said particle. The electron density (N_e) of the plasma will vary with the actual conditions and locations involved. Also, neutral particles, ions, and electrons are present in proximity to the field lines.

The production of enhanced ionization will also alter the distribution of atomic and molecular constituents of the atmosphere, most notably through increased atomic nitrogen concentration. The upper atmosphere is normally rich in atomic oxygen (the dominant atmospheric constituent above 200 km altitude), but atomic nitrogen is normally relatively rare. This can be expected to manifest itself in increased airglow, among other effects.

As known in plasma physics, the characteristics of a plasma can be altered by adding energy to the charged particles or by ionizing or exciting additional particles to increase the density of the plasma. One way to do this is by heating the plasma which can be accomplished in different ways, e.g., ohmic, magnetic compression, shock waves, magnetic pumping, electron cyclotron resonance, and the like.

Since electron cyclotron resonance heating is involved in the present invention, a brief discussion of

same is in order. Increasing the energy of electrons in a plasma by invoking electron cyclotron resonance heating, is based on a principle similar to that utilized to accelerate charged particles in a cyclotron. If a plasma is confined by a static axial magnetic field of strength B, the charged particles will gyrate about the lines of force with a frequency given, in hertz, as $f_g = 1.54 \times 10^3 B/A$, where: B = magnetic field strength in gauss, and A = mass number of the ion.

Suppose a time-varying field of this frequency is superimposed on the static field B confining the plasma, by passage of a radiofrequency current through a coil which is concentric with that producing the axial field, then in each half-cycle of their rotation about the field lines, the charged particles acquire energy from the oscillating electric field associated with the radio frequency. For example, if B is 10,000 gauss, the frequency of the field which is in resonance with protons in a plasma is 15.4 megahertz.

As applied to electrons, electron cyclotron resonance heating requires an oscillating field having a definite frequency determined by the strength of the confining field. The radio-frequency radiation produces time-varying fields (electric and magnetic), and the electric field accelerates the charged particle. The energized electrons share their energy with ions and neutrals by undergoing collisions with these particles, thereby effectively raising the temperature of the electrons, ions, and neutrals. The apportionment of energy among these species is determined by collision frequencies. For a more detailed understanding of the physics involved, see "Controlled Thermonuclear Reactions", Glasstone and Lovberg, D. Van Nostrand Company, Inc., Princeton, N.J., 1960 and "The Radiation Belt and Magnetosphere", Hess, Blaisdell Publishing Company, 1968, both of which are incorporated herein by reference.

Referring now to the drawings, the present invention provides a method and apparatus for altering at least one region of plasma which lies along a field line, particularly when it passes through the ionosphere and/or magnetosphere. FIG. **1** is a simplified illustration of the earth **10** and one of its dipole magnetic force or field lines **11**. As will be understood, line **11** may be any one of the numerous naturally existing field lines and the actual geographical locations **13** and **14** of line **11** will be chosen based on a particular operation to be carried out. The actual locations at which field lines intersect the earth's surface is documented and is readily ascertainable by those skilled in the art.

Line **11** passes through region R which lies at an altitude above the earth's surface. A wide range of altitudes are useful given the power that can be employed by the practice of this invention. The electron cyclotron resonance heating effect can be made to act on electrons anywhere above the surface of the earth. These electrons may be already present in the atmosphere, ionosphere, and/or magnetosphere of the earth, or can be artificially generated by a variety of means such as x-ray beams, charged particle beams, lasers, the plasma sheath surrounding an object such as a missile or meteor, and the like. Further, artificial particles, e.g., electrons, ions, etc., can be injected directly into region R from an earth-launched rocket or orbiting satellite carrying, for example, a payload of radioactive beta-decay material; alpha emitters; an electron accelerator; and/or ionized gases such as hydrogen; see U.S. Pat. No. 4,042,196. The altitude can be greater than about 50 km if desired,

e.g., can be from about 50 km to about 800 km, and, accordingly may lie in either the ionosphere or the magnetosphere or both. As explained above, plasma will be present along line **11** within region R and is represented by the helical line **12**. Plasma **12** is comprised of charged particles (i.e., electrons and ions) which rotate about opposing helical paths along line **11**.

Antenna **15** is positioned as close as is practical to the location **14** where line **11** intersects the earth's surface. Antenna **15** may be of any known construction for high directionality, for example, a phased array, beam spread angle (θ) type. See "The MST Radar at Poker Flat, Alaska", Radio Science, Vol. 15, No. 2, Mar.-Apr. 1980, pps. 213–223, which is incorporated herein by reference. Antenna **15** is coupled to transmitter **16** which generates a beam of high frequency electromagnetic radiation at a wide range of discrete frequencies, e.g., from about 20 to about 1800 kilohertz (kHz).

Transmitter **16** is powered by power generator means **17** which is preferably comprised of one or more large, commercial electrical generators. Some embodiments of the present invention require large amounts of power, e.g., up to 10^9 to 10^{11} watts, in continuous wave or pulsed power. Generation of the needed power is within the state of the art. Although the electrical generators necessary for the practice of the invention can be powered in any known manner, for example, by nuclear reactors, hydroelectric facilities, hydrocarbon fuels, and the like, this invention, because of its very large power requirement in certain applications, is particularly adapted for use with certain types of fuel sources which naturally occur at strategic geographical locations around the earth. For example, large reserves of hydrocarbons (oil and natural gas) exist in Alaska and Canada. In northern Alaska, particularly the North Slope region, large reserves are currently readily available. Alaska and northern Canada also are ideally located geographically as to magnetic latitudes. Alaska provides easy access to magnetic field lines that are especially suited to the practice of this invention, since many field lines which extend to desirable altitudes for this invention intersect the earth in Alaska. Thus, in Alaska, there is a unique combination of large, accessible fuel sources at desirable field line intersections. Further, a particularly desirable fuel source for the generation of very large amounts of electricity is present in Alaska in abundance, this source being natural gas. The presence of very large amounts of clean-burning natural gas in Alaskan latitudes, particularly on the North Slope, and the availability of magnetohydrodynamic (MHD), gas turbine, fuel cell, electrogasdynamic (EGD) electric generators which operate very efficiently with natural gas provide an ideal power source for the unprecedented power requirements of certain of the applications of this invention. For a more detailed discussion of the various means for generating electricity from hydrocarbon fuels, see "Electrical Aspects of Combustion", Lawton and Weinberg, Clarendon Press, 1969. For example, it is possible to generate the electricity directly at the high frequency needed to drive the antenna system. To do this, typically the velocity of flow of the combustion gases (v), past magnetic field perturbation of dimension d (in the case of MHD), follow the rule:

$$v=df$$

where f is the frequency at which electricity is generated. Thus, if $v=1.78\times10^6$ cm/sec and d=1 cm then electricity would be generated at a frequency of 1.78 mHz.

Put another way, in Alaska, the right type of fuel (natural gas) is naturally present in large amounts and at just the right magnetic latitudes for the most efficient practice of this invention, a truly unique combination of circumstances. Desirable magnetic latitudes for the practice of this invention interest the earth's surface both northerly and southerly of the equator, particularly desirable latitudes being those, both northerly and southerly, which correspond in magnitude with the magnetic latitudes that encompass Alaska.

Referring now to FIG. 2 a first ambodiment is illustrated where a selected region R_1 of plasma **12** is altered by electron cyclotron resonance heating to accelerate the electrons of plasma **12**, which are following helical paths along field line **11**.

To accomplish this result, electromagnetic radiation is transmitted at the outset, essentially parallel to line **11** via antenna **15** as right hand circularly polarized radiation wave **20**. Wave **20** has a frequency which will excite electron cyclotron resonance with plasma **12** at its initial or original altitude. This frequency will vary depending on the electron cyclotron resonance of region R_1 which, in turn, can be determined from available data based on the altitudes of region R_1, the particular field line **11** being used, the strength of the earth's magnetic field, etc. Frequencies of from about 20 to about 7200 kHz, preferably from about 20 to about 1800 kHz can be employed. Also, for any given application, there will be a threshhold (minimum power level) which is needed to produce the desired result. The minimum power level is a function of the level of plasma production and movement required, taking into consideration any loss processes that may be dominant in a particular plasma or propagation path.

As electron cyclotron resonance is established in plasma **12**, energy is transferred from the electromagnetic radiation **20** into plasma **12** to heat and accelerate the electrons therein and, subsequently, ions and neutral particles. As this process continues, neutral particles which are present within R_1 are ionized and absorbed into plasma **12** and this increases the electron and ion densities of plasma **12**. As the electron energy is raised to values of about 1 kilo electron volt (kev), the generated mirror force (explained below) will direct the excited plasma **12** upward along line **11** to form a plume R_2 at an altitude higher than that of R_1.

Plasma acceleration results from the force on an electron produced by a nonuniform static magnetic field $(\bar{B})$. The force, called the mirror force, is given by

$$F=-\mu\nabla B \qquad (2)$$

where μ is the electron magnetic moment and $\nabla \bar{B}$ is the gradient of the magnetic field, μ being further defined as:

$$W_\perp/B=mV_\perp^2/2B$$

where $W_\perp$ is the kinetic energy in the direction perpendicular to that of the magnetic field lines and B is the magnetic field strength at the line of force on which the guiding center of the particle is located. The force as represented by equation (2) is the force which is responsible for a particle obeying equation (1).

Since the magnetic field is divergent in region R_1, it can be shown that the plasma will move upwardly from the heating region as shown in FIG. **1** and further it can be shown that

$$\tfrac{1}{2}M_eV_{e\perp}{}^2(x) \approx \tfrac{1}{2}M_eV_{e\perp}{}^2(Y) + \tfrac{1}{2}M_iV_{i\,||}{}^2(Y) \qquad (3)$$

where the left hand side is the initial electron transverse kinetic energy; the first term on the right is the transverse electron kinetic energy at some point (Y) in the expanded field region, while the final term is the ion kinetic energy parallel to B at point (Y). This last term is what constitutes the desired ion flow. It is produced by an electrostatic field set up by electrons which are accelerated according to Equation (2) in the divergent field region and pulls ions along with them. Equation (3) ignores electron kinetic energy parallel to B because $V_{e\,||} \approx V_{i\,||}$, so the bulk of parallel kinetic energy resides in the ions because of their greater masses. For example, if an electromagnetic energy flux of from about 1 to about 10 watts per square centimeter is applied to region R, whose altitude is 115 km, a plasma having a density (N_e) of 10^{12} per cubic centimeter will be generated and moved upward to region R_2 which has an altitude of about 1000 km. The movement of electrons in the plasma is due to the mirror force while the ions are moved by ambipolar diffusion (which results from the electrostatic field). This effectively "lifts" a layer of plasma **12** from the ionosphere and/or magnetosphere to a higher elevation R_2. The total energy required to create a plasma with a base area of 3 square kilometers and a height of 1000 km is about 3×10^{13} joules.

FIG. **3** is an idealized representation of movement of plasma **12** upon excitation by electron cyclotron resonance within the earth's divergent force field. Electrons (e) are accelerated to velocities required to generate the necessary mirror force to cause their upward movement. At the same time neutral particles (n) which are present along line **11** in region $R_1$are ionized and become part of plasma **12**. As electrons (e) move upward along line **11**, they drag ions (i) and neutrals (n) with them but at an angle θ of about 13 degrees to field line **11**. Also, any particulates that may be present in region R_1, will be swept upwardly with the plasma. As the charged particles of plasma **12** move upward, other particles such as neutrals within or below R_1, move in to replace the upwardly moving particles. These neutrals, under some conditions, can drag with them charged particles.

For example, as a plasma moves upward, other particles at the same altitude as the plasma move horizontally into the region to replace the rising plasma and to form new plasma. The kinetic energy developed by said other particles as they move horizontally is, for example, on the same order of magnitude as the total zonal kinetic energy of stratospheric winds known to exist.

Referring again to FIG. **2**, plasma **12** in region R_1 is moved upward along field line **11**. The plasma **12** will then form a plume (cross-hatched area in FIG. **2**) which will be relatively stable for prolonged periods of time. The exact period of time will vary widely and be determined by gravitational forces and a combination of radiative and diffusive loss terms. In the previous detailed example, the calculations were based on forming a plume by producing 0+energies of 2 ev/particle. About 10 ev per particle would be required to expand plasma **12** to apex point C (FIG. **1**). There at least some of the particles of plasma **12** will be trapped and will oscillate between mirror points along field line **11**. This oscillation will then allow additional heating of the trapped plasma **12** by stochastic heating which is associated with trapped and oscillating particles. See "A New Mechanism for Accelerating Electrons in the Outer Ionosphere" by R. A. Helliwell and T. F. Bell, Journal of Geophysical Research, Vol. 65, No. 6, June, 1960. This is preferably carried out at an altitude of at least 500 km.

The plasma of the typical example might be employed to modify or disrupt microwave transmissions of satellites. If less than total black-out of transmission is desired (e.g., scrambling by phase shifting digital signals), the density of the plasma (N_e) need only be at least about 10^6 per cubic centimeter for a plasma orginating at an altitude of from about 250 to about 400 km and accordingly less energy (i.e., electromagnetic radiation), e.g., 10^8 joules need be provided. Likewise, if the density $N_e$is on the order of 10^8, a properly positioned plume will provide a reflecting surface for VHF waves and can be used to enhance, interfere with, or otherwise modify communication transmissions. It can be seen from the foregoing that by appropriate application of various aspects of this invention at strategic locations and with adequate power sources, a means and method is provided to cause interference with or even total disruption of communications over a very large portion of the earth. This invention could be employed to disrupt not only land based communications, both civilian and military, but also airborne communications and sea communications (both surface and subsurface). This would have significant military implications, particularly as a barrier to or confusing factor for hostile missiles or airplanes. The belt or belts of enhanced ionization produced by the method and apparatus of this invention, particularly if set up over Northern Alaska and Canada, could be employed as an early warning device, as well as a communications disruption medium. Further, the simple ability to produce such a situation in a practical time period can by itself be a deterring force to hostile action. The ideal combination of suitable field lines intersecting the earth's surface at the point where substantial fuel sources are available for generation of very large quantitities of electromagnetic power, such as the North Slope of Alaska, provides the wherewithal to accomplish the foregoing in a practical time period, e.g., strategic requirements could necessitate achieving the desired altered regions in time periods of two minutes or less and this is achievable with this invention, especially when the combination of natural gas and magnetohydrodynamic, gas turbine, fuel cell and/or EGD electric generators are employed at the point where the useful field lines intersect the earth's surface. One feature of this invention which satisfies a basic requirement of a weapon system, i.e., continuous checking of operability, is that small amounts of power can be generated for operability checking purposes. Further, in the exploitation of this invention, since the main electromagnetic beam which generates the enhanced ionized belt of this invention can be modulated itself and/or one or more additional electromagnetic radiation waves can be impinged on the ionized region formed by this invention as will be described in greater detail herein after with respect to FIG. **4**, a substantial amount of randomly modulated signals of very large power magnitude can be generated in a highly nonlinear mode. This can cause confusion of or interference with or even complete disruption of guidance systems employed by

even the most sophisticated of airplanes and missiles. The ability to employ and transmit over very wide areas of the earth a plurality of electromagnetic waves of varying frequencies and to change same at will in a random manner, provides a unique ability to interfere with all modes of communications, land, sea, and/or air, at the same time. Because of the unique juxtaposition of usable fuel source at the point where desirable field lines intersect the earth's surface, such wide ranging and complete communication interference can be achieved in a resonably short period of time. Because of the mirroring phenomenon discussed hereinabove, it can also be prolonged for substantial time periods so that it would not be a mere transient effect that could simply be waited out by an opposing force. Thus, this invention provides the ability to put unprecedented amounts of power in the earth's atmosphere at strategic locations and to maintain the power injection level, particularly if random pulsing is employed, in a manner far more precise and better controlled than heretofore accomplished by the prior art, particularly by the detonation of nuclear devices of various yeilds at various altitudes. Where the prior art approaches yielded merely transitory effects, the unique combination of fuel and desirable field lines at the point where the fuel occurs allows the establishment of, compared to prior art approaches, precisely controlled and long-lasting effects which cannot, practically speaking, simply be waited out. Further, by knowing the frequencies of the various electromagnetic beams employed in the practice of this invention, it is possible not only to interfere with third party communications but to take advantage of one or more such beams to carry out a communications network even though the rest of the world's communications are disrupted. Put another way, what is used to disrupt another's communications can be employed by one knowledgeable of this invention as a communications network at the same time. In addition, once one's own communication network is established, the far-reaching extent of the effects of this invention could be employed to pick up communication signals of other for intelligence purposes. Thus, it can be seen that the disrupting effects achievable by this invention can be employed to benefit by the party who is practicing this invention since knowledge of the various electromagnetic waves being employed and how they will vary in frequency and magnitude can be used to an advantage for positive communication and eavesdropping purposes at the same time. However, this invention is not limited to locations where the fuel source naturally exists or where desirable field lines naturally intersect the earth's surface. For example, fuel, particularly hydrocarbon fuel, can be transported by pipeline and the like to the location where the invention is to be practiced.

FIG. 4 illustrates another embodiment wherein a selected region of plasma R_3which lies within the earth's ionosphere is altered to increase the density thereof whereby a relatively stable layer **30** of relatively dense plasma is maintained within region R_3. Electromagnetic radiation is transmitted at the outset essentially parallel to field line **11** via antenna **15** as a right hand circularly polarized wave and at a frequency (e.g., 1.78 megahertz when the magnetic field at the desired altitude is 0.66 gauss) capable of exciting electron cyclotron resonance in plasma **12** at the particular altitude of plasma **12**. This causes heating of the particles (electrons, ions, neutrals, and particulates) and ionization of the uncharged particles adjacent line **11**, all of which are absorbed into plasma **12** to increase the density thereof. The power transmitted, e.g., 2×10^6 watts for up to 2 minutes heating time, is less than that required to generate the mirror force F required to move plasma **12** upward as in the previous embodiment.

While continuing to transmit electromagnetic radiation **20** from antenna **15**, a second electromagnetic radiation beam **31**, which is at a defined frequency different from the radiation from antenna **15**, is transmitted from one or more second sources via antenna **32** into layer **30** and is absorbed into a portion of layer **30** (cross-hatched area in FIG. 4). The electromagnetic radiation wave from antenna **32** is amplitude modulated to match a known mode of oscillation f_3 in layer **30**. This creates a resonance in layer **30** which excites a new plasma wave **33** which also has a frequency of f_3 and which then propogates through the ionosphere. Wave **33** can be used to improve or disrupt communications or both depending on what is desired in a particular application. Of course, more than one new wave **33** can be generated and the various new waves can be modulated at will and in a highly nonlinear fashion.

FIG. **5** shows apparatus useful in this invention, particularly when those applications of this invention are employed which require extremely large amounts of power. In FIG. **5** there is shown the earth's surface **40** with a well **41** extending downwardly thereinto until it penetrates hydrocarbon producing reservoir **42**. Hydrocarbon reservoir **42** produces natural gas alone or in combination with crude oil. Hydrocarbons are produced from reservoir **42** through well **41** and wellhead **43** to a treating system **44** by way of pipe **45**. In treater **44**, desirable liquids such as crude oil and gas condensates are separated and recovered by way of pipe **46** while undesirable gases and liquids such as water, H_2S, and the like are separated by way of pipe **47**. Desirable gases such as carbon dioxide are separated by way of pipe **48**, and the remaining natural gas stream is removed from treater **44** by way of pipe **49** for storage in conventional tankage means (not shown) for future use and/or use in an electrical generator such as a magnetohydrodynamic, gas turbine, fuel cell or EGD generator **50**. Any desired number and combination of different types of electric generators can be employed in the practice of this invention. The natural gas is burned in generator **50** to produce substantial quantities of electricity which is then stored and/or passed by way of wire **51** to a transmitter **52** which generates the electromagnetic radiation to be used in the method of this invention. The electromagnetic radiation is then passed by way of wire **53** to antenna **54** which is located at or near the end of field line **11**. Antenna **54** sends circularly polarized radiation wave **20** upwards along field line **11** to carry out the various methods of this invention as described hereinabove.

Of course, the fuel source need not be used in its naturally-occurring state but could first be converted to another second energy source form such as hydrogen, hydrazine and the like, and electricity then generated from said second energy source form.

It can be seen from the foregoing that when desirable field line **11** intersects earth's surface **40** at or near a large naturally-occurring hydrocarbon source **42**, exceedingly large amounts of power can be very efficiently produced and transmitted in the direction of field lines. This is particularly so when the fuel source is natural gas and magnetohydrodynamic generators are employed. Further, this can all be accomplished in a

relatively small physical area when there is the unique coincidence of fuel source **42** and desirable field line **11**. Of course, only one set of equipment is shown in FIG. **5** for sake of simplicity. For a large hydrocarbon reservoir **42**, a plurality of wells **41** can be employed to feed one or more storage means and/or treaters and as large a number of generators **55** as needed to power one or more transmitters **52** and one or more antennas **54**. Since all of the apparatus **44** through **54** can be employed and used essentially at the sight where naturally-occurring fuel source **42** is located, all the necessary electromagnetic radiation **20** is generated essentially at the same location as fuel source **42**. This provides for a maximum amount of usable electromagnetic radiation **20** since there are no significant storage or transportation losses to be incurred. In other words, the apparatus is brought to the sight of the fuel source where desirable field line **11** intersects the earth's surface **40** on or near the geographical location of fuel source **42**, fuel source **42** being at a desirable magnetic latitude for the practice of this invention, for example, Alaska.

The generation of electricity by motion of a conducting fluid through a magnetic field, i.e., magnetohydrodynamics (MHD), provides a method of electric power generation without moving mechanical parts and when the conducting fluid is a plasma formed by combustion of a fuel such as natural gas, an idealized combination of apparatus is realized since the very clean-burning natural gas forms the conducting plasma in an efficient manner and the thus formed plasma, when passed through a magnetic field, generates electricity in a very efficient manner. Thus, the use of fuel source **42** to generate a plasma by combustion thereof for the generation of electricity essentially at the site of occurrence of the fuel source is unique and ideal when high power levels are required and desirable field lines **11** intersect the earth's surface **40** at or near the site of fuel source **42**. A particular advantage for MHD generators is that they can be made to generate large amounts of power with a small volume, light weight device. For example, a 1000 megawatt MHD generator can be construed using superconducting magnets to weigh roughly 42,000 pounds and can be readily air lifted.

This invention has a phenomenal variety of possible ramifications and potential future developments. As alluded to earlier, missile or aircraft destruction, deflection, or confusion could result, particularly when relativistic particles are employed. Also, large regions of the atmosphere could be lifted to an unexpectedly high altitude so that missiles encounter unexpected and unplanned drag forces with resultant destruction or deflection of same. Weather modification is possible by, for example, altering upper atmosphere wind patterns or altering solar absorption patterns by constructing one or more plumes of atmospheric particles which will act as a lens or focusing device. Also as alluded to earlier, molecular modifications of the atmosphere can take place so that positive environmental effects can be achieved. Besides actually changing the molecular composition of an atmospheric region, a particular molecule or molecules can be chosen for increased presence. For example, ozone, nitrogen, etc. concentrations in the atmosphere could be artificially increased. Similarly, environmental enhancement could be achieved by causing the breakup of various chemical entities such as carbon dioxide, carbon monoxide, nitrous oxides, and the like. Transportation of entities can also be realized when advantage is taken of the drag effects caused by regions of the atmosphere moving up along diverging field lines. Small micron sized particles can be then transported, and, under certain circumstances and with the availability of sufficient energy, larger particles or objects could be similarly affected. Particles with desired characteristics such as tackiness, reflectivity, absorptivity, etc., can be transported for specific purposes or effects. For example, a plume of tacky particles could be established to increase the drag on a missile or satellite passing therethrough. Even plumes of plasma having substantially less charged particle density than described above will produce drag effects on missiles which will affect a lightweight (dummy) missile in a manner substantially different than a heavy (live) missile and this affect can be used to distinguish between the two types of missiles. A moving plume could also serve as a means for supplying a space station or for focusing vast amount of sunlight on selected portions of the earth. Surveys of global scope could also be realized because the earth's natural magnetic field could be significantly altered in a controlled manner by plasma beta effects resulting in, for example, improved magnetotelluric surveys. Electromagnetic pulse defenses are also possible. The earth's magnetic field could be decreased or disrupted at appropriate altitudes to modify or eliminate the magnetic field in high Compton electron generation (e.g., from high altitude nuclear bursts) regions. High intensity, well controlled electrical fields can be provided in selected locations for various purposes. For example, the plasma sheath surrounding a missile or satellite could be used as a trigger for activating such a high intensity field to destroy the missile or satellite. Further, irregularities can be created in the ionosphere which will interfere with the normal operation of various types of radar, e.g., synthetic aperture radar. The present invention can also be used to create artificial belts of trapped particles which in turn can be studied to determine the stability of such parties. Still further, plumes in accordance with the present invention can be formed to simulate and/or perform the same functions as performed by the detonation of a "heave" type nuclear device without actually having to detonate such a device. Thus it can be seen that the ramifications are numerous, far-reaching, and exceedingly varied in usefulness.

I claim:

1. A method for altering at least one region normally existing above the earth's surface with electromagnetic radiation using naturally-occurring and diverging magnetic field lines of the earth comprising transmitting first electromagnetic radiation at a frequency between 20 and 7200 kHz from the earth's surface, said transmitting being conducted essentially at the outset of transmission substantially parallel to and along at least one of said field lines, adjusting the frequency of said first radiation to a value which will excite electron cyclotron resonance at an initial elevation at least 50 km above the earth's surface, whereby in the region in which said electron cyclotron resonance takes place heating, further ionization, and movement of both charged and neutral particles is effected, said cyclotron resonance excitation of said region is continued until the electron concentration of said region reaches a value of at least 10^6 per cubic centimeter and has an ion energy of at least 2 ev.

2. The method of claim **1** including the step of providing artificial particles in said at least one region which are excited by said electron cyclotron resonance.

3. The method of claim **2** wherein said artificial particles are provided by injecting same into said at least one region from an orbiting satellite.

4. The method of claim **1** wherein said threshold excitation of electron cyclotron resonance is about 1 watt per cubic centimeter and is sufficient to cause movement of a plasma region along said diverging magnetic field lines to an altitude higher than the altitude at which said excitation was initiated.

5. The method of claim **4** wherein said rising plasma region pulls with it a substantial portion of neutral particles of the atmosphere which exist in or near said plasma region.

6. The method of claim **1** wherein there is provided at least one separate source of second electromagnetic radiation, said second radiation having at least one frequency different from said first radiation, impinging said at least one second radiation on said region while said region is undergoing electron cyclotron resonance excitation caused by said first radiation.

7. The method of claim **6** wherein said second radiation has a frequency which is absorbed by said region.

8. The method of claim **6** wherein said region is plasma in the ionosphere and said second radiation excites plasma waves within said ionosphere.

9. The method of claim **8** wherein said electron concentration reaches a value of at least 10^{12} per cubic centimeter.

10. The method of claim **8** wherein said excitation of electron cyclotron resonance is initially carried out within the ionosphere and is continued for a time sufficient to allow said region to rise above said ionosphere.

11. The method of claim **1** wherein said excitation of electron cyclotron resonance is carried out above about 500 kilometers and for a time of from 0.1 to 1200 seconds such that multiple heating of said plasma region is achieved by means of stochastic heating in the magnetosphere.

12. The method of claim **1** wherein said first electromagnetic radiation is right hand circularly polarized in the northern hemisphere and left hand circularly polarized in the southern hemisphere.

13. The method of claim **1** wherein said electromagnetic radiation is generated at the site of a naturally-occurring hydrocarbon fuel source, said fuel source being located in at least one of northerly or southerly magnetic latitudes.

14. The method of claim **13** wherein said fuel source is natural gas and electricity for generating said electromagnetic radiation is obtained by burning said natural gas in at least one of magnetohydrodynamic, gas turbine, fuel cell, and EGD electric generators located at the site where said natural gas naturally occurs in the earth.

15. The method of claim **14** wherein said site of natural gas is within the magnetic latitudes that encompass Alaska.

* * * * *

B1—Battery, 9 VDC, type 2U6
or equiv.
M1—Meter, 0-1-mA DC

56 Simple AM Mod. Monitor

☐ This simple modulation monitor for AM ham or CB transmitters requires no connection to the transmitter. Just position the loop near the final tank or antenna matching coil until the signal is heard in the headphones.

PARTS LIST FOR
SIMPLE AM MOD. MONITOR

C1—100-pF disc capacitor
D1—1N914 diode
E1—Magnetic headphone, 2000
ohms or better
L1—Coil, 3 turns on 1½-in. dia.
form, use any thin gauge wire

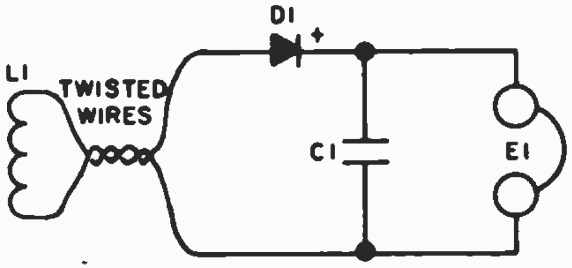

Prehistoric Radio!

If you've been following the "Below 500 kHz" column, you know that natural phenomena right here on earth have been producing, transmitting, and receiving radio signals since time began. Lightning bolts, and some other mighty reactions in nature, produce great amounts of electromagnetic radiation in the light, heat, and the radio portions of the electromagnetic-wave spectrum. Lightning discharges from tropical storms in particular are continually producing wide-band radio waves that propagate worldwide. You receive these signals as bothersome static noise on the LF, MF, and on HF bands.

FIGURE 1:

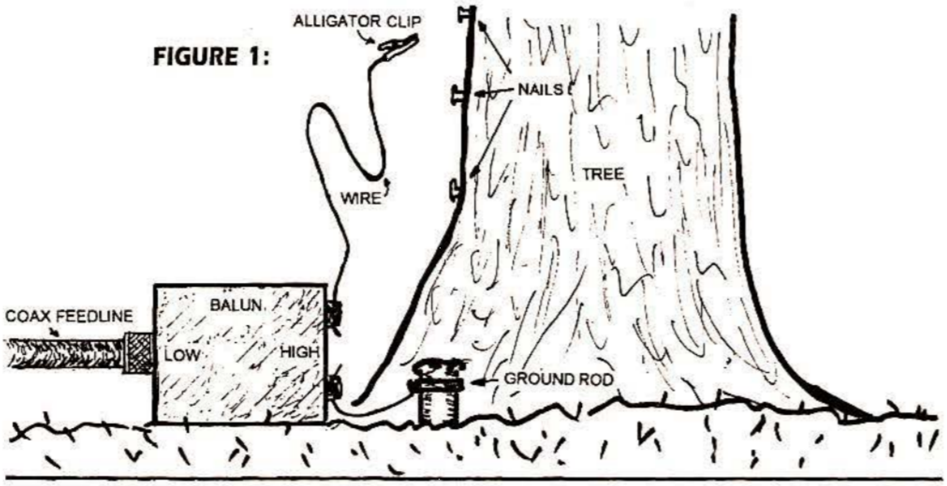

Connecting a feedline to a tree antenna. Nail small, thort, thin nails into the tree about every two or three feet, up to 10 or 15 feet high. Connect the clip to the various nails until best results are achieved. The balun is a 4:1 ratio or higher, high side to the tree.

Weak-signal operators who do moon-bounce work are only too well acquainted with the radio signals which our own sun transmits. These noisy signals can totally mask over moon-bounce signals when the sun and moon are both in the same direction from moon-bouncer's antenna. Radio waves known as "cosmic noise" are continually bombarding our antennas with radio signals generated by natural reactions occurring in galactic space. As a matter of fact, the discovery that the source of this particular radio noise lay outside our solar system marked the beginning of radio astronomy. In radio astronomy, naturally occurring radio signals from space are studied in an attempt to learn more about the cosmos.

OK, so there were radio signals before we were around to know about it, but what about the reception of those waves? Who received those signals? No person received them, but they were received by every naturally-occurring conductor to which they propagated. And, once trees evolved, they provided natural, grounded-vertical Marconi antennas to receive these waves.

Well, but where was the receiver for those antennas? The resistance of the tree's "body" was the "receiver." This resistance dissipated much of the received energy as heat rather than as sound, like the loudspeaker does in our HF receivers.

Of course, natural transmission and reception still goes on today. Have you asked the trees near your home if they have received any RF signals lately? I have. And the trees which I asked told me that they receive lots of signals—and I haven't gone off my rocker! They receive lightning-generated signals just as they have since time immemorial, and they also receive the signals which man puts onto the airwaves today.

How did I ask the trees? It's simple. I just connected a feedline to a tree, and connected the line to my receiver. I received a multitude of signals across the LF, MF, and HF bands. If I had tried other bands I would have, no doubt, received signals on those bands, too. As a matter of fact, a tall, live tree makes a decent LF-MF-HF antenna.

■ Tree-Antennas Throughout Radio's History

Old time radio expert Squires reported using trees as receiving aerials for distances of 50 kilometers. His method was to hammer a nail into the tree a few yards above the ground, and connect the receiver's antenna input connectors to the nail and to a ground.

Texanna Loomis, daughter of wireless communication pioneer Mahlon Loomis, says in her *Radio Theory and Operating* that "It has been found that a tree can be used for a receiving antenna, preferably an oak, by attaching a lead-in wire to the trunk of the tree."

Another old-time radio expert named Morecroft felt that reception came from the lead-in wire, rather than the tree itself. In my own tests I can't agree with Morecroft, because my lead-in was shielded, grounded coax, the attachment connectors to the tree were short, and the antenna performed much more like a long antenna than a short stub.

In more recent times there have also been reports of tree antennas. Cohen, in the April 1996 issue of *73 Magazine*, reports using a tree antenna on VHF and UHF with "mediocre-to-good results," but no acceptable results for lower bands. Perhaps his use of trees only 20 feet tall limited their performance on lower bands.

Also in *73* (May 1990), JA6HW and JA6AUI report using a 12 foot high tree for both transmitting and receiving on the 10 meter band with good results. Their article also says that hams used live-tree antennas in the 1930's, as did the U. S. Forces in Vietnam. In *Monitoring Times*, April 1989, Dr. Kosta reports that trees "frequently work better than conventional antennas" for television reception. In the September 1990 issue of *Monitoring Times* yours truly reports good results with a tall maple tree on LF through HF.

Tapping into Natural Radio

Figure 1 shows suggestions on utilizing a tree as an antenna. For HF, I suggest using the tallest tree available. Perhaps shorter trees would be better for VHF-UHF as in the 1990 *73* article discussed above—or perhaps not. As with any outside antenna it is wise to utilize some form of lightning protection.

RADIO RIDDLES

Last month:

I said, "Let's say that we could stain some radio waves, frozen at one moment in time, with some kind of dye so that we could actually see them. Of course we can't do this, except in our imagination. But if we did, how would the waves appear to us?"

For the area occupied by the waves at any one instant in time we would be trying to visualize the variations in the signal's electrical-field strength, and the reversals in field orientation (showing when the RF current changes direction of flow twice each cycle). This wouldn't look like a wavy line, or a sine wave graph as is represented in most texts on radio.

One way to imagine seeing the frozen waves is to envision that the space which the radio waves fill will show a darker shade of gray where their field strength is more intense, and a lighter shade where their field strength is weaker. For simplicity, let's consider only direct-wave propagation with no reflections or other impediments to the wave's travel between the transmitting antenna and receiving antenna.

In the space between the two antennas we would see bands of differing shades of gray oriented at right angles to the wave's direction of travel. There would be darker bands, less dark bands just next, even less dark bands next till a relatively light shade of gray was reached. Then somewhat darker bands, then even darker bands, and on to produce a sort of zebra-skin look. The darkest bands would be a half wavelength apart as would the lightest ones.

If we added a green tinge to our gray to indicate one orientation of the electrical field, and a red tinge for the field's other orientation then, starting at the middle of a darker band there would be greenish gray for a half wavelength along the direction of wave travel, a reddish gray for the next half wavelength, and so on.

Of course this picture is an over-simplification. For instance, we haven't talked about the wave's magnetic field, the polarization of the wave, or the spherical shape of the wavefront. But this simplified picture does cover the basic idea of radio waves in space.

This Month:

Let's say that, instead of freezing the waves, we had the advanced eyes and brains of some specialized android so that we could actually watch radio waves zip along on their 186,000 miles per hour flight from the transmitting antenna to our receiving antenna? How would those rascals look then?

You'll find an answer for this month's riddle, and much more, in next month's issue of *Monitoring Times*. 'Til then Peace, DX, and 73.

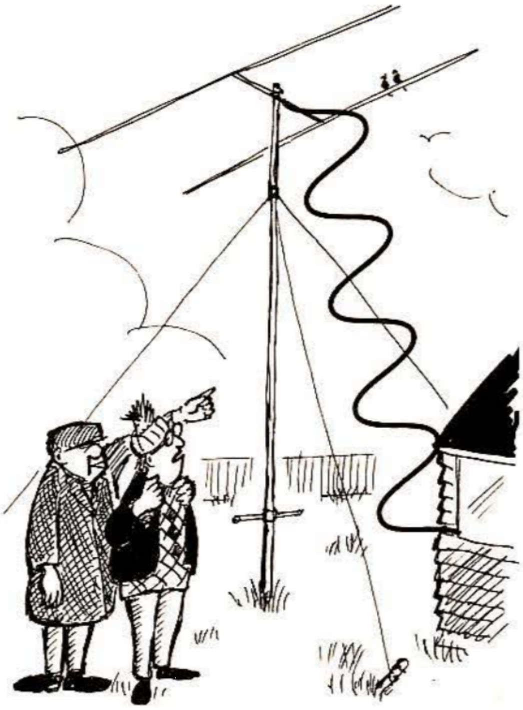

"You know, I thought those were classic standing radio waves until I read Clem Small's column this month. Now, I don't know what the heck they are!"

Motron Electronics
310 Garfield St Suite 4
PO Box 2748
Eugene, Oregon 97402
http://www.motron.com

$59.00 Plus S/H

DTMF DECODER

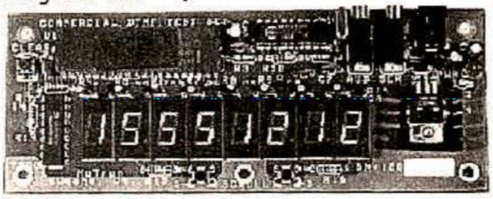

The TDD-8X features a large 8-digit LED display and decodes all 16 DTMF digits. The 104 character memory is viewed, without loss of data, by scrolling either left or right. Although a computer is not required, the serial ASCII output allows for automatic logging of decoded number groups. ToneLog ™, our IBM/Compatible logging software, is included with each TDD-8X purchase.

TDD-8X Assembled & Tested PCB $ 59.00 **PMK-1** Plastic Mounting Kit $15.00
CAB-1 Audio & Computer Cables 20.00 **PS-12** 12VDC AC Power Adaptor 10.00

Visa, MasterCard, American Express, Discover and Government Purchase Orders accepted.
COD on Cash or Money Order basis only.
Minimum Government Purchase Order: $99.
S/H: $8 USA; $11 Canada; $16 Foreign COD: $5.
Premium Shipping available at additional charge

Info: (541) 687-2118 **Orders: (800) 338-9058** Fax: (541) 687-2492

17

Solar TV Reception Booster

Here is a versatile rooftop TV reception booster that will add a full 18 dB of gain to the incoming TV signals at your receiver. This means no more ghosts, clearer pictures, and perhaps even reception of a distant station you could not get before. Best of all, this gadget is powered by an inexpensive solar cell that will virtually run forever without requiring replacement. Additionally, the circuit is designed with a charger circuit and reserve pack.

You can build it as is, or only include as much as you would like. For example, if you break the circuit where the X's are indicated, your booster will thrive entirely off its solar battery. If you break it at the two Y's, the solar cell charges a 4½-volt battery pack during the day so the booster will work just as well at night as in direct sunlight. If you include all the circuit (taking into account the indoor AC supply shown), you can occasionally recharge the 4½-volt pack just in case you have had a protracted period of rain or overcast weather which prevented the solar cell from sending its energy down to the reserve battery.

Construction is not complicated, but since you are dealing with both vhf and uhf signals, you will have to keep all leads extremely short and direct. In fact, you would do best to cram the parts close together so as to cut down on the length of interconnecting wires. Looking at the schematic, you will note that Q1 must be grounded; this is no problem; however, since the HEP-3 comes with four leads instead of the regular three. Turning the transistor upside down, you will notice an arrangement of leads as shown and identified in Fig. 17.

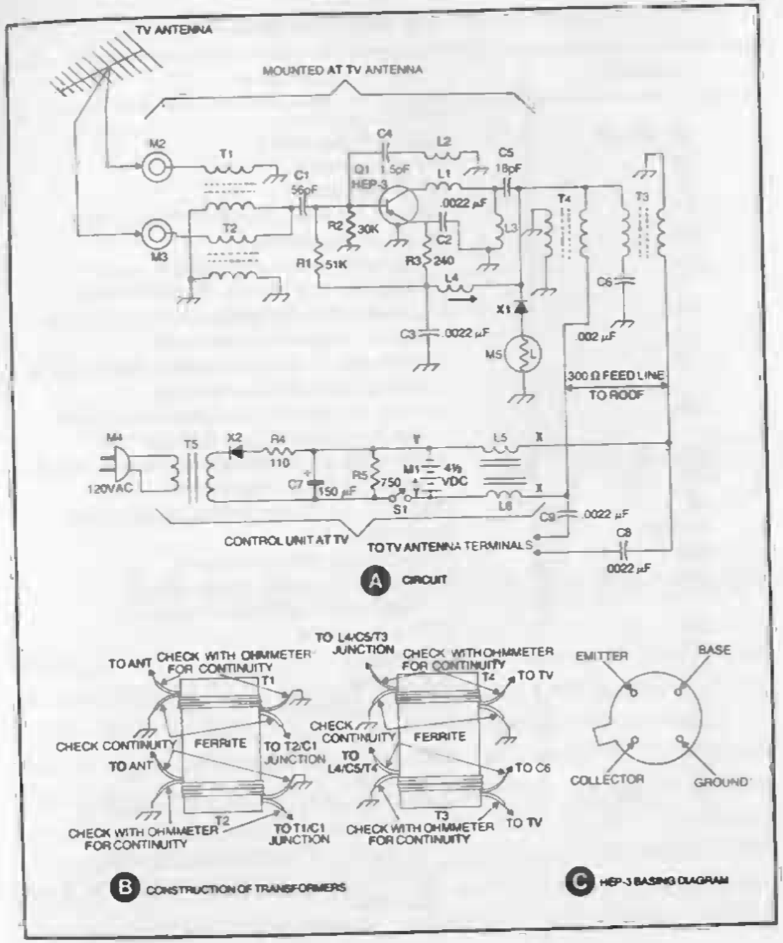
TV ANTENNA
MOUNTED AT TV ANTENNA
M2
T1
C1
56pF
T2
M3
C4
1.5pF
L2
Q1
HEP-3
L1
C5
18pF
.0022 μF
R2
30K
R1
51K
C2
R3
240
L3
L4
X1
C3
.0022 μF
M5
L
T4
T3
C6
.002 μF
300 Ω FEED LINE
TO ROOF
M4
120VAC
T5
X2
R4
110
C7
150 μF
R5
750
M1
4½
VDC
S1
Y
L5
L6
X
C9
.0022 μF
C8
0022 μF
CONTROL UNIT AT TV
TO TV ANTENNA TERMINALS
A CIRCUIT
TO ANT
CHECK WITH OHMMETER FOR CONTINUITY
T1
CHECK CONTINUITY
FERRITE
TO T2/C1 JUNCTION
TO ANT
T2
CHECK WITH OHMMETER FOR CONTINUITY
TO T1/C1 JUNCTION
TO L4/C5/T3 JUNCTION
CHECK WITH OHMMETER FOR CONTINUITY
T4
TO TV
CHECK CONTINUITY
FERRITE
TO L4/C5/T4
TO C6
T3
CHECK WITH OHMMETER FOR CONTINUITY
TO TV
EMITTER
BASE
COLLECTOR
GROUND
B CONSTRUCTION OF TRANSFORMERS
C HEP-3 BASING DIAGRAM

Fig. 17. Solar TV reception booster.

Take care in constructing your coils that you duplicate exactly those turns and techniques recommended in Table 17.

Once complete, you can test out your booster by hooking it to the TV set and placing a lit 60-watt lamp bulb in rather close proximity to M5. As the lamp is turned on, you should get a much-improved picture. Adjust the best reception on all channels by tuning L4 and the coupling between L1 and L2. Now mount at the TV antenna, positioning M5 for best sunlight hits.

Hint: Leave the AC plug plugged in at all times. It will not draw power until you throw switch S1 onto "charge."

Table 17. Parts List for Solar TV Reception Booster.

Item No.	Description
C1	56-pF capacitor.
C2, C3, C6	.0022-μF capacitors.
C4	1.5-pF capacitor.
C5	18-pF capacitor.
C7	150-μF, 15-w VDC electrolytic capacitor.
C8, C9	.0022-μF capacitors.
L1	7½ turns No. 24 enameled wire, evenly wound on a 3/16-inch-diameter form.
L2	3 turns No. 24 enameled wire evenly wound on a 3/16-inch-diameter form.
L3	17 turns No. 24 enameled wire evenly wound on a 3/16-inch-diameter form.
L4	11½ turns of No. 24 enameled wire evenly wound on Speer Type E ferrite form.
L5, L6	10 turns No. 24 enameled wire evenly wound on Speer Type E ferrite form.
M1	4½ VDC, with three penlites of NiCd cells.
M2, M3	Binding posts.
M4	AC wall plug with cable clamp.
M5	4½-VDC Solar pack. (International Rectifier No. SP5G26C or equiv.)
Q1	HEP-3 transistor.
R1	51K resistor.
R2	30K resistor.
R3	240-ohm resistor.
R4	110-ohm resistor.
R5	750-ohm resistor.
S1	Spst switch.
T1, T2, T3, T4	Using Ferroxcube K5050-06 ferrite cores, insert 2 turns of special 300-ohm miniature twinlead in each form hole. Pull tight, and connect leads where shown in diagram.
T5.	6.3-VAC filament trnasformer. (Triad F-14X or equiv.)

18

Free-Power AM Radio Receiver

Strange as it may seem, this transistor broadcast-band radio receiver "steals" power from one station to give to another! The principle is basic: By tuning the battery-section antenna coil (L2) to the strongest broadcast station on the band, diode X1 can rectify the rf and convert it into DC current. Naturally, the closer you are to a strong station, the more current the "radio battery" section of your radio receiver will be able to supply. Once you have found this spot, the DC current is passed on to power the transistor circuit which acts as a genuine receiver, with the full tuning it affords. See Fig. 18 and Table 18.

The basic consideration is a good antenna and ground, the latter preferably being made to a water pipe or solid external ground composed of a pipe driven at least 4 feet into moist earth. This procedure not only ensures maximum signal pickup for the radio-battery portion of the circuit, but also provides best results for the GE-2 receiver circuit.

Once completed, just tune the radio battery as explained in the first paragraph and calibrate your receiver by adjusting L1 so that the bottom of the band occurs when C1 is fully meshed. If you have a local broadcast station operating near 540 kHz, this simplifies things tremendously. Once the calibration procedure is complete, forget entirely about adjusting L1 and do all your listening by tuning C1.

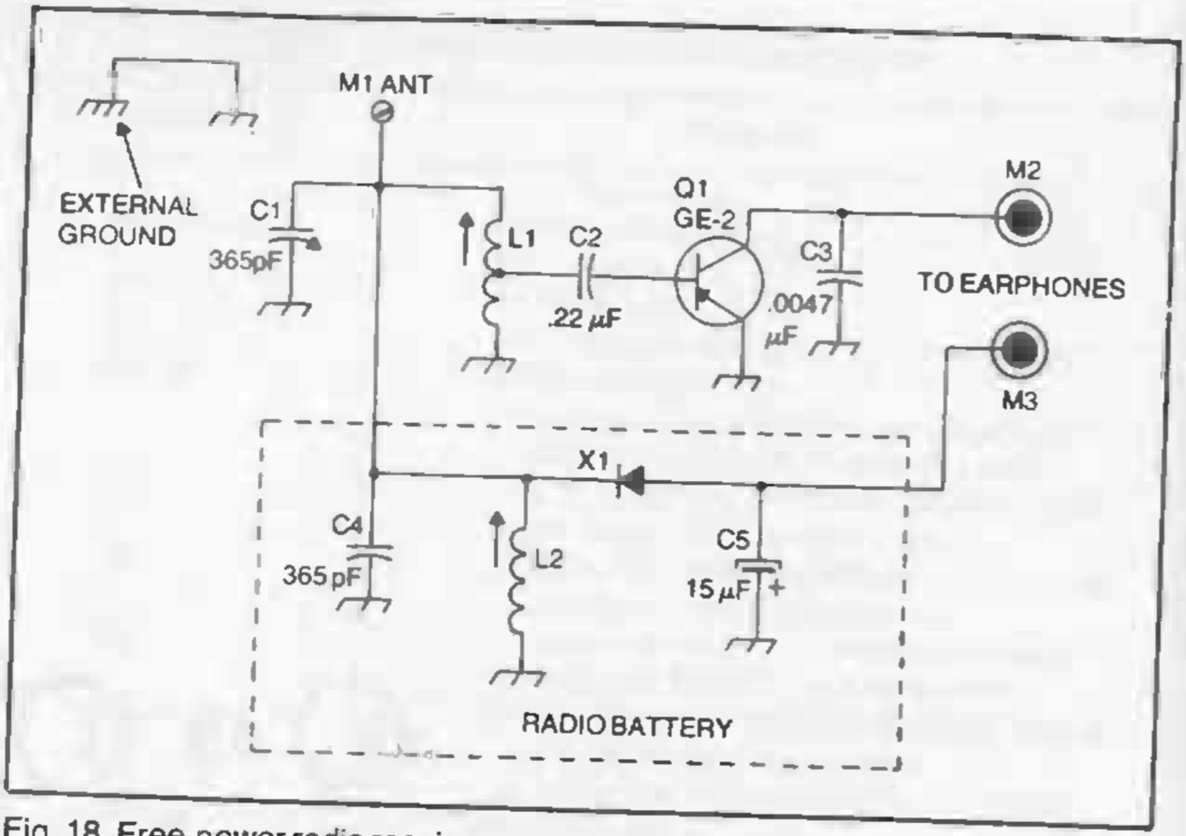
EXTERNAL GROUND
M1 ANT
C1
365pF
L1
C2
.22 µF
Q1
GE-2
C3
.0047 µF
M2
TO EARPHONES
M3
X1
C4
365 pF
L2
C5
15 µF
+
RADIO BATTERY

Fig. 18. Free-power radio receiver.

Table 18. Parts List for Free-Power AM Radio Receiver.

Item No.	Description
C1, C4	365-pF variable capacitors.
C2	.22-µF capacitor.
C3	.0047-µF capacitor.
C5	15-µF, 6-w VDC electrolytic capacitor.
L1, L2	Loopsticks. (Superex VLT-240 or equiv.)
M1	Binding post.
M2, M3	Test jacks.
Q1	GE-2 transistor.
X1	1N38B diode.

RADIO SCIENCE Journal of Research NBS/USNC–URSI
Vol. 69D, No. 8, August 1965

Schumann Resonances[1]

Janis Galejs

Applied Research Laboratory, Sylvania Electronic Systems, a Division of Sylvania Electric Products Inc.,
40 Sylvan Road, Waltham, Mass. 02154

(Received February 3, 1965)

This paper reviews the current work in the field of Schumann resonances and discusses the topics of waveforms and frequency estimates, resonance frequencies and Q factors, source distributions and noise spectra, diurnal and seasonal changes of power spectra, and variations of resonance frequencies. Observations of the above phenomena are correlated with theoretical considerations.

1. Introduction

The resonances in the spherical shell between the earth and ionosphere that are due to azimuthal waves have been considered in two recent survey papers [Galejs, 1964a; Wait, 1964a]. This work contains experimental data as well as the required theory. The basic theory of the resonances has been reviewed by Wait [1964b]. In the present summary the emphasis will be on correlation between measurements and available theory. There will be no extensive analytical developments, but frequent references will be made to literature.

Based on theoretical considerations, Schumann [1952a, b, 1957] postulated resonances of the earth-to-ionosphere cavity. Koenig [1958, 1959, 1961] obtained the first experimental indication of Schumann resonances by observing noise waveforms in the output of a narrow band amplifier. Detailed frequency spectra of this noise were first obtained by Balser and Wagner [1960]. Other measurements have been reported by Fournier [1960]; Benoit and Houri [1961, 1962]; Lokken et al. [1961, 1962]; Polk and Fitchen [1962]; Gendrin and Stefant [1962a]; Balser and Wagner [1962a, b, 1963]; Rycroft [1963]; Chapman and Jones [1964].

The fundamental theory of Schumann resonances [Schumann 1952a, b, 1957] is discussed in a book by Wait [1962]. Raemer [1961a, b] considers the observable noise spectra as the response of the earth-to-ionosphere cavity due to lightning flashes all over the world, but the homogeneous sharply bounded ionosphere model of Raemer introduces high losses, and he does not succeed in reproducing the spectral measurements of Balser and Wagner [1960]. Galejs [1961a, b] uses an isotropic ionosphere model of exponentially increasing conductivity which is based on measured or calculated characteristics of the lower ionosphere. This model permits a close reproduction of the noise spectra measurements in the resonance region, and it also provides an agreement with measured ELF attenuation rates [Chapman and Macario, 1956; Jean et al., 1961].

Models of an anisotropic ionosphere are difficult to apply to the problem of earth-to-ionosphere cavity resonances because of the variations of the magnetic field vector along the surface of the earth. The propagation parameters can be estimated in the presence of a horizontal magnetic field component from the work of Galejs and Row [1964] and of Galejs [1964b]. The earth-to-ionosphere cavity resonances have been analyzed for a nonhomogeneous ionosphere in the presence of a radial (vertical) magnetic field by Thompson [1963] and Galejs [1965]. A thin-shell approximation of the ionosphere with a superimposed radial magnetic field has been considered by Wait [1964c]. Some other work of Wait [1963a, b], although intended for a different frequency range, can also be applied to the earth-to-ionosphere resonance problem in the presence of a radial magnetic field.

Additional references are available in the comprehensive bibliography by Brock-Nannestad [1962].

2. Waveforms and Frequency Estimates

In wide band recordings of atmospheric signals an ELF component or slow tail follows the initial VLF component of the transient [Hepburn and Pierce, 1953; Liebermann, 1956a, b; Tepley, 1959; Pierce, 1960]. An oscillatory structure of the signal becomes noticeable by passing it through a band-pass filter. An example of a noise recording in the output of a 2 to 30 c/s band-pass filter [Lokken, 1964, private communication] is shown in figure 1. The two horizontal magnetic field components have been observed simultaneously in Canada and in Antarctica. There is a correlation between the high intensity ELF noise bursts observed at the various sites. A recording of one of the waveforms is shown in figure 2. The darkened areas indicate resonance frequencies around 8, 14, and 20 c/s. Noise waveforms in the 5 to 20 c/s band have been reported by Polk and Fitchen [1962] and Polk [1962]. Figure 3 shows sample recordings of the outputs of two coils the axes of which are oriented in the north to south (N–S) and east to west (E–W) directions. Frequency of these waveforms has been estimated by counting the cycles between the second markers shown on the records. The average frequency computed in this procedure should be higher than the resonance frequencies of a spectral analysis [Galejs, 1962a, 1964a]. Noise recordings made simultaneously in Germany and in Rhode Island have been reported by Keefe, Polk, and Koenig [1964].

---

[1] Paper presented at the ULF Symposium, Boulder, Colo., Aug. 17–20, 1964.

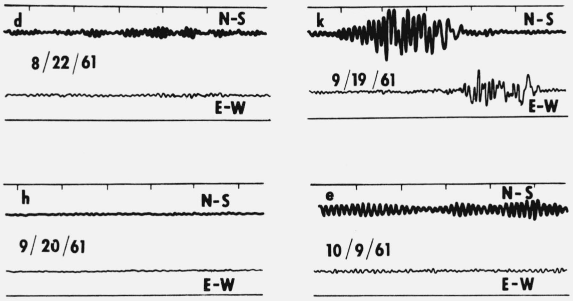

FIGURE 3. *Noise recording in the 5 to 20 c/s band.*

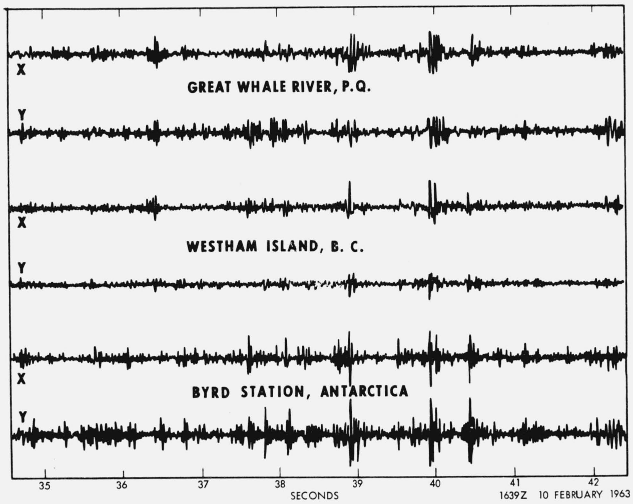

FIGURE 1. *Noise recording in the 2 to 30 c/s band.*

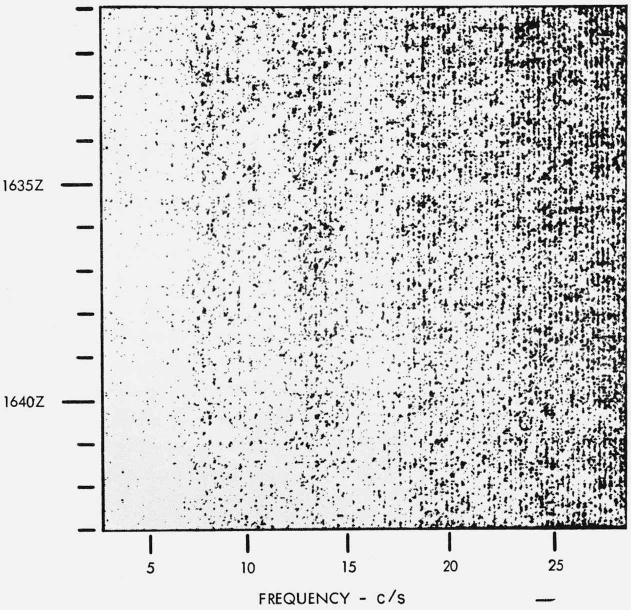

FIGURE 2. *Recording of the noise waveform X of Westham Island, B.C.*

3. Resonance Frequencies and Q Factors

3.1. Definitions

A resonance frequency and Q factor may be determined from spectral measurements. Analytically it is also possible to compute the power spectrum of the noise waveform. However, it is more convenient to relate the resonance frequency and the Q factor to the analytically determined propagation parameters. The radial electrical field due to the vertical dipole excitation can be represented by

$$E_r = -\frac{iIds}{4\omega\epsilon h a^2}\,\frac{\nu(\nu+1)}{\sin\nu\pi}\,P_\nu(-\cos\theta) \tag{1}$$

where

$$\nu(\nu+1) = (k_0 a S)^2. \tag{2}$$

For a wave that propagates in the spherical shell between the earth and the ionosphere S is defined as the ratio of the complex wave number k to free space wave number $k_0(S = k/k_0)$. The real part of S can be seen to be inversely proportional to the phase velocity, and the imaginary part of S is proportional to the attenuation constant. This parameter S is determined from solving the appropriate boundary value problem of the earth, the air space and the chosen ionosphere model [Wait, 1962; Galejs, 1964a]. The resonance frequencies of the nth mode are determined from the minima of $\sin\nu\pi$ in (1) [Galejs, 1962a, 1964a] which gives

$$f_n \approx \frac{7.5\sqrt{n(n+1)}}{\mathrm{Re}S}. \tag{3}$$

Typical values of ReS are 1.4 to 1.2 for frequencies in the resonance region.

The cavity Q may be determined as a ratio between the stored energy and energy loss per cycle or simply by the width of the resonance curves. As shown in the appendix, in the present problem the electrically stored energy W_e differs from the magnetically stored energy W_h, and the Q definition should consider the sum of the average stored energies W_e and W_h. This expression for Q is

$$Q=\frac{(\mathrm{Re}S)^2+1}{4\mathrm{Re}S\,\mathrm{Im}S}. \tag{4}$$

When considering only the magnetically stored energy Q is given by

$$Q=\frac{1}{2\,\mathrm{Re}S\,\mathrm{Im}S}. \tag{5}$$

For a given set of propagation parameters this will give a lower Q figure than (4). This expression has been used in the past work of Galejs [1961a, 1962a, b, 1964a]. Considering only the electrically stored energy gives

$$Q=\frac{\mathrm{Re}S}{2\,\mathrm{Im}S}. \tag{6}$$

The last expression can also be derived from the concept of complex resonance frequencies [Wait, 1964a] or by considering the half-power band width of the resonance curve. All three Q definitions yield the same results as ReS approaches unity, but they will differ near the lower resonance frequency of the earth ionosphere cavity. Equation (4) will be used in the present calculation.

Experimentally determined Q values that are estimated from the half-power bandwidths of the resonance curves neglect the effects of adjacent resonances and of near-field noise that will tend to add to the background level. Hence, the half-power level is estimated too low, the apparent half-power bandwidth is larger and the estimated Q factor may be too low.

3.2. Characteristics of the Boundary

In the theoretical models that are advanced for explaining the resonance frequencies and the Q-factor measurements [Wait, 1962; Galejs, 1964a; Wait, 1964a] it is necessary to consider the boundary properties of the spherical shell.

In the frequency range of the resonances the displacement currents of ground are negligible and the ground conductivity $\sigma_g=10^{-3}$ to 1 mho/m is much higher than the effective ionospheric conductivity σ_i that is estimated to be of the order of 10^{-7} to 10^{-6} mho/m. The ground conductivity σ_g may be assumed as infinite in the first approximation and only the ionospheric properties require a detailed analysis.

The ionosphere acts as an anisotropic conductor and for a z-directed magnetic field its conductivity tensor is of the form

$$[\sigma]=\begin{bmatrix}\sigma_1 & -\sigma_2 & 0\\ \sigma_2 & \sigma_1 & 0\\ 0 & 0 & \sigma_0\end{bmatrix}. \tag{7}$$

Equations for computing the conductivity components are available in literature [Galejs, 1964a and b] and typical conductivity profiles are shown in figure 4 but some uncertainties still apply to such models [VanZandt, 1964]. The above computations consider the ion effects, but assume the operating frequency to be zero. These profiles can be further approximated by straight lines through the lower ionosphere in the semilogarithmic plots of figure 4, which corresponds to an expoential height variation of the conductivity components. In a further simplification the anisotropy may be ignored ($\sigma_1=\sigma_0$, $\sigma_2=0$). Even a simple one or two layer model can be assumed, such as shown in figure 5. These particular models have been used in past work by Raemer [1961a and b], Galejs [1962a and b], and by Chapman and Jones [1964]. In the isotropic exponential model the conductivity is given by

$$\sigma(z)=\sigma(z_0)\,\exp\,[\beta(z-z_0)] \tag{8}$$

where $z_0=60$ km and the day and night models are characterized by $\sigma(z_0)=4.63\times10^{-8}$ mho/m and $\beta=0.308$ km^{-1} and by $\sigma(z_0)=6.5\times10^{-10}$ mho/m and $\beta=0.44$ km^{-1}, respectively.

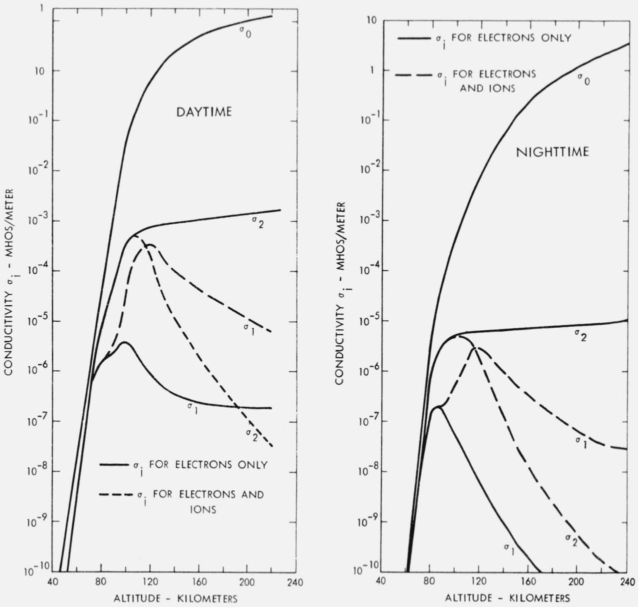

FIGURE 4. *Components of the tensor conductivity* F = 0.

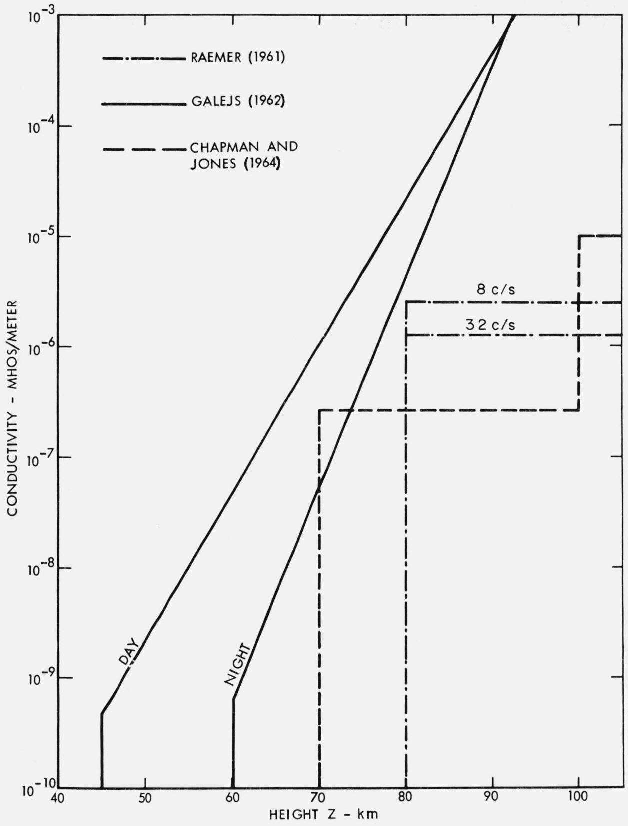

FIGURE 5. *Models of ionospheric conductivity.*

3.3. Single Layer Ionosphere

The single layer ionosphere model of Raemer [1961a and b] assumes the boundary height of the ionosphere to be constant, but the conductivity to change with frequency so as to obtain the correct resonant frequencies. However, the resulting *Q* factors are approximately two times lower than in measurements. The same difficulties are experienced with an ionosphere model of constant conductivity, but frequency dependent boundary height. Similar results are obtained by Madden [1961] with a conductivity of $\sigma = 1.8 \times 10^{-6}$ mho/m and the ionospheric boundary at $h = 80$ km. The deficiencies of a single layer ionosphere model are less pronounced if it is attempted to achieve only a partial match of the resonant frequencies, as seen from Wait [1964a]. However, the *Q* factors will be too low in particular at the higher resonance frequencies ($f \approx 30$ c/s).

Compressibility effects of the ionosphere have been considered by Seshadri [1965b] using the model of a perfectly conducting flat earth and a sharply bounded homogeneous isotropic and lossless ionosphere. The compressibility of the ionosphere does not affect the wave propagation below it for frequencies in the range of Schumann resonances and in the ELF band. However, the compressibility should be considered for waves that propagate inside the ionosphere [Seshadri, 1965a].

3.4. Two-Layer Ionosphere Models

Two-layer models give more accurate values of the resonance frequencies and of cavity *Q* or of the complex propagation parameter *S*. Two-layer and multilayer models have been discussed by Wait [1958, 1962]; Madden [1961]; Jean et al. [1961], and a particularly successful two-layer model has been constructed by Chapman and Jones [1964a, b]. This model, which is indicated in figure 5, has been applied in the frequency range of earth-to-ionosphere cavity resonances and also for frequencies up to 1 kc/s.

Such models are usually derived by trial and error procedures and they are difficult to correlate with continuous ionosphere profiles, such as shown in figure 4.

3.5. Models of Continuously Varying Ionosphere

The isotropic exponential model gives a reasonable approximation to the ionospheric conductivity versus height profile through the lower ionosphere layers, and the propagation parameters of the model can be determined from closed form expressions of the surface impedance [Galejs, 1961a, 1962a, b, 1964a]. Calculated resonance frequencies and *Q* factors that are based on (3), (4), and available values of *S* [Galejs, 1962a] are listed in tables I and II. Calculations are based on daytime or day and nighttime ionosphere models. The *Q* factors considered also the effects of cosmic ray ionization [Bourdeau et al., 1959] that give a nearly exponential conductivity versus height variation as indicated by Galejs [1962a]. Cosmic ray ionization decreases the effective *Q* figures, in particular for the lower frequencies. The calculated *Q* values are still somewhat higher than those estimated from experimental data.

TABLE I

RESONANCE FREQUENCIES OF THE SPHERICAL SHELL BETWEEN THE EARTH AND IONOSPHERE

	n	1	2	3	4
RESONANCE FREQUENCY f_n IN C/S	LOSSLESS CAVITY	10.6	18.3	25.9	33.5
	DAYTIME WITH h = 45 KM	7.6	13.5	19.4	25.4
	AVERAGE DATA OF DAY WITH h = 45 KM AND OF NIGHT WITH h = 60 KM	8.05	14.25	20.5	26.8
	MEASUREMENTS	8	14.1	20.3	26.4

TABLE II

Q FACTORS OF EARTH-IONOSPHERE CAVITY RESONANCES

	f IN C/S	10	30	100
Q - FACTORS	DAY WITH h = 45 KM, NIGHT WITH h = 60 KM	8.5	7	6.7
	DAYTIME ONLY WITH h = 45 KM	6.5	5.4	5.2
	DAY WITH h = 45 KM, NIGHT WITH h = 60 KM, COSMIC RAY IONIZATION	5.5	6	6.4
	MEASUREMENTS	4	6	

3.6. Magnetic Field Effects

The magnetic field effects for a homogeneous sharply bounded ionosphere model have been discussed by Wait [1962] in a quasi-longitudinal approximation for a superimposed radial magnetic field and also for a transverse magnetic field (east-west or west-east propagation along the magnetic equator).

The propagation parameters have been determined by Galejs and Row [1964] and Galejs [1964a, 1964b] in the presence of a horizontal magnetic field. These data provide an indication of propagation characteristics of the waves near the magnetic equator, but they do not apply directly to the earth-ionosphere resonance problem, where the **H** vector is nearly horizontal only over a small surface portion of the earth.

The earth-ionosphere cavity resonance frequencies have been considered in the thesis of Thompson [1963] for a radial magnetic field. He uses a multilayer ionosphere model that approximates the continuous altitude variation of the conductivity tensor. The azimuthal field variation is expressed in terms of the Legendre polynomials $P_n(\cos\theta)$ for the nth mode. The wave equation has four different solutions for each of the ionosphere layers, but it has only two solutions for the space between the ionosphere and the perfectly conducting ground surface. The solution of the problem and the resonance condition for the mode n are obtained after multiplying a sequence of 4×4 matrices. In these computations the displacement currents of the ionosphere are neglected. The computed resonance frequencies are in agreement with measurements, but the Q values are in the range of 15 to 20, with the exception of nighttime anisotropic solutions, where $Q=7$ to 10. The lowering of the nighttime Q figures by the magnetic field has been attributed to wave penetration through the ionosphere and to energy thus escaping from the earth-ionosphere cavity.

Another solution to the problem of earth's ionosphere cavity resonance in the presence of a radial magnetic field is given by Galejs [1965]. The azimuthal field variation is expressed in terms of the Legendre functions $P_\nu(-\cos\theta)$, and the solution of the wave equation is set up so that the four solutions of the ionospheric layers where $\sigma > \omega\epsilon_0$ degenerate into two solutions for the lower ionospheric layers where $\sigma << \omega\epsilon_0$ and for free space. After computing the surface impedance at the lower boundary of the ionosphere, the propagation parameters are determined from the solutions of the modal equation for TM modes, althought the coupling between TE and TM modes is considered in the processes of computing the surface impedance. Only the vertically polarized TM modes are assumed to propagate below the ionosphere. Such assumptions have been justified for a single layer ionosphere model at higher frequencies [Wait 1962, p. 269] and for a thin ionospheric shell also for frequencies in the resonance range [Wait, 1964c].

In the numerical calculations the anisotropic ionosphere models are those of figure 4 and further calculations are made for ionosphere models which consider electrons and ions at an operating frequency of 20 c/s, as indicated in figures 3 and 4 of Galejs [1964b]. In the isotropic ionosphere models σ_0 remains unaltered, $\sigma_1 = \sigma_0$, and $\sigma_2 \rightarrow 0$. The calculated resonance frequencies f_n are listed in table III. The isotropic day or night models exhibit too high resonance frequencies, but the average for the anisotropic day or night models gives nearly correct results, although the first resonance occurs near 7.7 c/s. At night the various anisotropic models give different results, while their results are nearly the same for daytime or for the isotropic models. The calculated Q factors are listed in table IV. At daytime the Q factors are lowered slightly due to anisotropy. The losses are principally due to absorption in the lower D region where the collision frequencies are high and the magnetic field has only small effects. At nighttime the anisotropy reduces the Q values drastically, and the ions have a considerable effect. For the isotropic models the energy remains below the ionosphere, the losses in the lower ionosphere are small and the Q values are high. For the anisotropic models some energy penetrates the ionosphere and escapes from the earth to ionosphere cavity. Hence the Q figure is low.

TABLE III

RESONANCE FREQUENCIES OF THE SPHERICAL SHELL BETWEEN THE EARTH AND IONOSPHERE IN THE PRESENCE OF A RADIAL MAGNETIC FIELD

	IONOSPHERE MODEL	n	1	2	3	4
RESONANCE FREQUENCY f_n IN C/S	DAY, ANISOTROPIC		7.8	13.8	19.8	26
	DAY, ISOTROPIC		8.1	14.2	20.4	26.6
	NIGHT, ANISOTROPIC	EL.+ IONS F = 20	7.6	15.2	23.2	29.5
		EL.+ IONS F = 0	7.7	13.8	20.1	26.9
		EL. ONLY F = 0	7.8	14.1	20.7	27.8
	NIGHT, ISOTROPIC		8.6	15.3	21.3	27.4
	MEASUREMENTS		8	14.1	20.3	26.4

TABLE IV

Q-FACTORS OF EARTH-IONOSPHERE CAVITY RESONANCES IN THE PRESENCE OF A RADIAL MAGNETIC FIELD

	IONOSPHERE MODEL	F	10	30
Q-FACTORS	DAY, ANISOTROPIC		6.8	5.6
	DAY, ISOTROPIC		8.3	7
	NIGHT, ANISOTROPIC	EL. + IONS F = 20	1.7	6.6
		EL. + IONS F = 0	2.0	2.7
		EL. ONLY F = 0	2.8	4.5
	NIGHT, ISOTROPIC		8.8	9.3
	MEASUREMENTS		4	6

The previous calculations have assumed a constant radial magnetic field (S pole) all over the globe. A reversal of the static magnetic field (N pole) causes negligible differences in the propagation parameters or in the values of f_n and Q. Hence a model using the average of the radial magnetic field may provide a first order estimate of the anisotropy effects. The present calculations assume the radial magnetic of the polar regions to be extended over the whole surface of the globe and the anisotropy effects should be too pronounced. The analytical results are highly dependent on the detailed structure of the nighttime anisotropic ionosphere, and further studies of it may be in order.

4. Source Distributions and Noise Spectra

Most of the ELF energy is of terrestrial origin and is caused by thunderstorm activity, although there are some high latitude events of an apparent extraterrestrial origin. Such emissions of approximately 800 c/s have been detected near Kiruna, Sweden [Egeland, 1964] and of 14 to 17 c/s on Kerguelen Island in the Southern part of the Indian Ocean [Gendrin and Stephant, 1964, private communication]. The Spectran recording of the signals received at Kerguelen is shown in figure 6. The upper trace shows a normal signal with resonances near 8, 14, 20, and 25 c/s. There is an additional signal of 14 to 17 c/s on the two lower traces. When neglecting the presence of such extraneous signals the terrestrial noise spectra can be calculated after estimating the dipole moments of the sources and using propagation parameters that are appropriate to the selected ionosphere model.

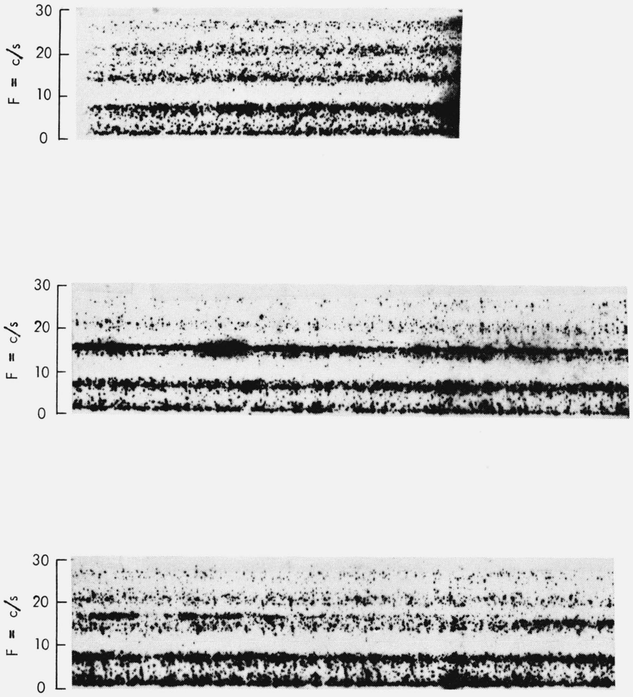

FIGURE 6. *Spectran recordings of noise waveforms at Kerguelen Island.*

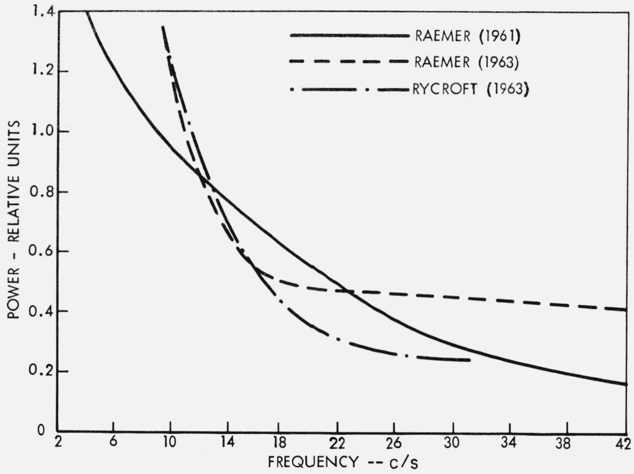

FIGURE 7. *Estimated lighting flash spectrum.*

Raemer [1961a, b] has estimated the power spectrum of the dipole moment in a median lightning flash Ids (or the first derivative of the electrostatic moment dM/dt) based on statistics of lightning strokes compiled by Williams [1959]. Later work of Raemer [1963 private communication] has lead to a revised estimate. These spectra are shown in figure 7. Pierce [1963] has computed the magnitude of the frequency spectra of the second derivative d^2M/dt^2, and Rycroft [1963] has calculated the frequency spectra of dM/dt using a similar representation of the lightning moments. The square of the frequency spectrum computed by Rycroft [1963] is also indicated in figure 7.

An equivalent power spectrum for the dipole moment of terrestrial noise sources has been deduced by Harris and Tanner [1962] from the measurements of Balser and Wagner [1960] by estimating the complex propagation constant within this frequency range in a "trial and error" approach. Harris and Tanner [1962] do not require an ionosphere model or any knowledge of lightning waveforms and their statistics for obtaining this equivalent power spectrum.

After determining the source spectra and the propagation parameters the power spectrum of the received noise may be computed from

$$G(i\omega)=\frac{g(i\omega)\,|\nu(\nu+1)|^2}{8(ak_0h)^2\,|\nu+\frac{1}{2}|\,[\cosh(2\pi\,\mathrm{Im}\,\nu)-\cos(2\pi\,\mathrm{Re}\,\nu)]}\cdot\left\{\frac{\sinh[2\,\mathrm{Im}\,\nu\,(\pi-\theta)]}{(-\mathrm{Im}\,\nu)}-\frac{\sin[(2\,\mathrm{Re}\,\nu+1)(\pi-\theta)-(\pi/2)]}{\mathrm{Re}\,\nu+\frac{1}{2}}\right\}_{\theta_1}^{\theta_2}, \quad (9)$$

where the ν is given by (2) and where the sources are assumed to be uniformly distributed in the θ interval of $\theta_1 < \theta < \theta_2$. The derivation of (9) starts from (1) and Legendre functions $P_\nu(-\cos\,\theta)$ are replaced by an asymptotic expansion. There are also equivalent spectral representations in terms of Legendre polynomials [Raemer, 1961a, b].

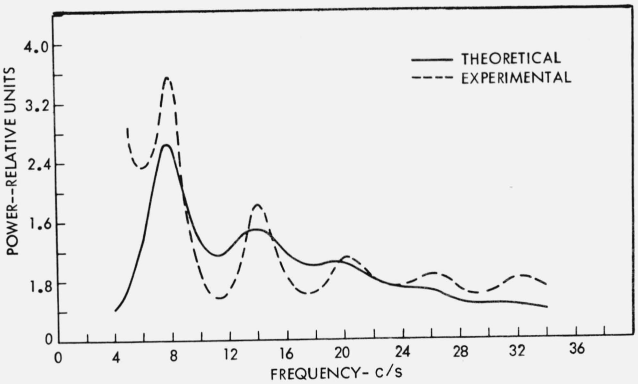

FIGURE 8. *ELF Noise spectrum—effective ionospheric height assumed constant.*

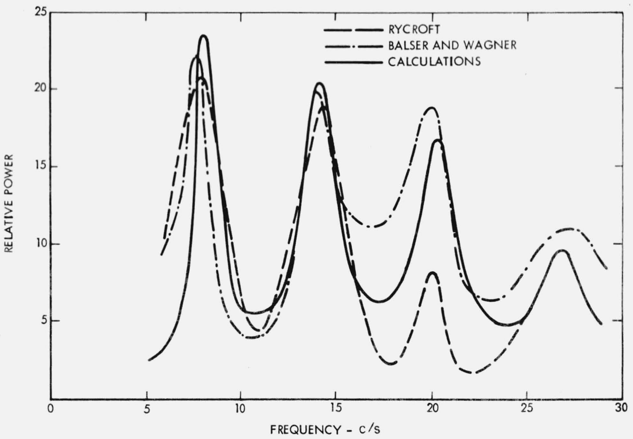

FIGURE 10. *Noise spectrum for 47° ≤ θ ≤ 128°.*

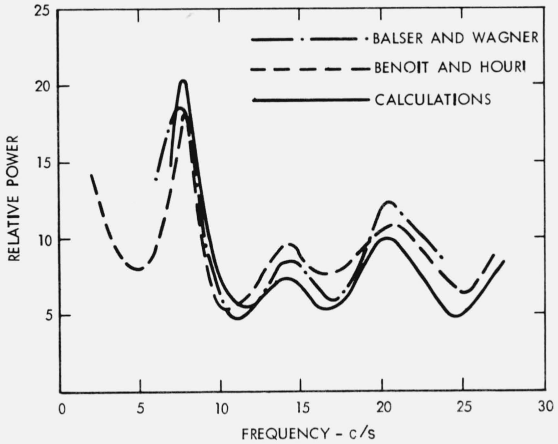

FIGURE 9. *Noise spectrum for 38° ≤ θ ≤ 72°.*

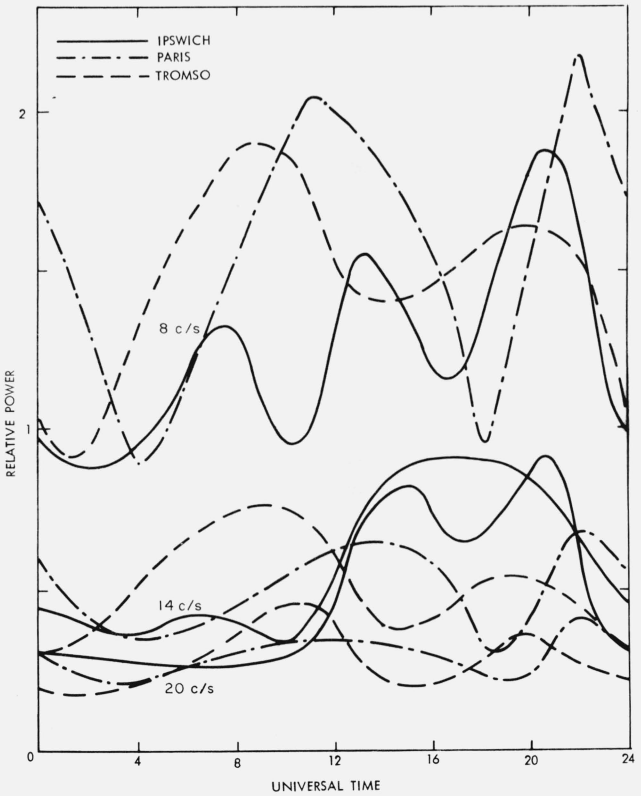

FIGURE 11. *Diurnal variation of the resonance modes at Ipswich, Mass., February 1961 (vertical electric field), Chambon-la-Foret (near Paris) July 1962 and at Tromso (Norway) April 1962 (horizontal magnetic field).*

The noise spectrum of Raemer [1961b] has been used for computing the observed noise spectra in conjunction with a sharply bounded ionosphere model. This spectrum is shown in figure 8 and is seen to exhibit excessively damped higher resonance peaks.

The exponential ionosphere model [Galejs, 1961b, 1962a, 1964a] produces a better agreement with measurements as seen in figures 9 and 10. In this work propagation is assumed to be uniform around the surface of the earth and the propagation parameters represent the average of day and night conditions. The sources are assumed to be uniformly distributed over the θ interval indicated in the figures (θ is defined with respect to the observer).

5. Diurnal and Seasonal Changes of Power Spectra

Balser and Wagner [1962a] have recorded diurnal variations of measured noise spectra. Their measurements of the diurnal variations of the power level at the first three resonance frequencies are shown in figure 11 together with similar measurements by Gendrin and Stefant [1962a]. Diurnal variations of the fields near the two lower resonance frequencies have been measured at Kingston, R.I., by Polk, Huck, and Yu [1963, private communication]. The maximum signal intensity occurs between 1500 and 1800 UT. In an amplitude versus time plot the signal peak is broader during summer than winter. Seasonal and diurnal variations of the noise power in the ELF frequency band below 35 c/s have been described by Wright [1963] without singling out the individual resonance modes. The measurements are made at Byrd Station, Antarctica (80°S, 120°W) and show peak intensities near 1000 UT during spring and summer and near 2300 UT during most of the year.

The gross features of the spectrum, such as the number of peaks and their approximate time of occurance, may be explained qualitatively by noting areas of globe with particularly high thunderstorm activity and by measuring the polar angle θ from these source regions to the observation point. The vertical electric field E_r and the horizontal magnetic field H_ϕ of the mode n exhibit the θ dependence of

$$E_r \sim P_n(\cos\theta) \tag{10}$$

$$H_\phi \sim \frac{d}{d\theta} P_n(\cos\theta), \tag{11}$$

where $P_n(\cos\theta)$ is a Legendre polynomial of the order n. A particular geographical region will contribute to the measured values of E_r or H_ϕ in the mode n if the right hand sides of (10) or (11) are sizeable for those values of θ. Such considerations have been reported by Balser and Wagner [1962a] and Gendrin and Stefant [1962a].

More detailed calculations have been carried out by Abraham [1964, private communication; Galejs 1964a] who used available estimates of geographical distributions of the noise sources and who calculated the day and night propagation effects after expressing the fields as a sum of recirculating azimuthal waves.

In the expression for the electric field component E_r of (1), the Legendre function is replaced by the leading term of its asymptotic expansion. The function $\sin \nu\pi$ is expanded into

$$(\sin \nu\pi)^{-1} = -2i \sum_{n=0}^{\infty} e^{i(2n+1)\nu\pi} \tag{12}$$

which is valid for Im $\nu > 0$. This results in [Abraham, 1964, private communication]

$$E_r = \frac{Ids}{2\omega\epsilon h a^2} \frac{\nu(\nu+1)}{\sqrt{2\pi(\nu+\frac{1}{2})\sin\theta}} e^{i\frac{3\pi}{4}}$$
$$\sum_{n=0}^{\infty} \left\{ e^{i(\nu+\frac{1}{2})(\theta+2n\pi) - i2n\frac{\pi}{2}} + e^{i(\nu+\frac{1}{2})[2\pi(n+1)-\theta] - i(2n+1)\frac{\pi}{2}} \right\}. \tag{13}$$

The first term of the summation represents a wave which has traveled the direct distance from the source to the observation point (θ) plus n full circles around the sphere ($2n\pi$), while experiencing a phase shift of ($-\pi/2$) radians at each traversal of the source point or its antipode. The other term of the summation represents a wave which has traveled toward the observation point across the antipode of the source ($2\pi-\theta$) and has made furthermore n circles around the sphere ($2n\pi$) and has also experienced a phase shift of ($-\pi/2$) radians at each traversal of the source point or its antipode. This physical interpretation of (13), which is equivalent to (1) for $\nu=$const., permits a heuristic consideration of the ν differences between the day and night hemispheres. The multiplicative ν factors in front of the n summation are assumed to be constant and equal to $\overline{\nu}$ which is intermediate between the day and night ν values. The θ factors of the exponentials are replaced by integrals in order to account for the ν variation seen by the circular wave front as it propagates in the θ-direction with respect to the source. The amplitude changes at the day and night boundary can be also considered for each traveling wave by assuming that the energy of the wave remains constant after passing the day and night boundary. This results in decreased amplitudes for the larger ionospheric height of the night hemisphere. In the numerical work the surface of the earth was subdivided into 10° cells. Each cell had a constant surface density of noise sources but a variation was allowed from cell to cell. The plots of thunderstorm days over the world were used to set relative values of lightning activity for each of the cells, by assuming that the distribution of squared dipole moment per unit area is directly proportional to the thunderstorm day plots according to the Handbook of Geophysics [1960]. The diurnal variations were accounted for by a multiplicative factor in terms of local time at the source [Williams, 1959]. The final equation used in the spectrum calculations is similar to (9), although it is algebraically more involved [Abraham, 1964, private communication].

The diurnal variations of the noise power at the first three Schumann resonances have been computed using the exponential ionosphere model of figure 5 and are shown in figure 12. The calculations have disregarded the noise sources near the antipode where the usual asymptotic approximation of the Legendre functions is inaccurate. Figure 12 shows a qualitative agreement with the measurements of Balser and Wagner [1962a], that are shown in figure 11.

An example of seasonal power variations is shown in figure 13. It indicates a decreased noise power near the higher frequencies and/or during winter months.

Based on measurements at one location of the earth it is also possible to predict the noise levels at other locations using this procedure of noise spectrum calculations. The noise that is due to a worldwide distribution of thunderstorms depends on the frequency variation and also on the absolute value of the power spectrum of the mean dipole moment of the source $g(i\omega, \theta, \phi)$. The θ and ϕ dependence may be estimated from the thunderstorm frequency maps as was indicated before, which leaves the absolute value and the frequency dependence of $g(i\omega)$ as an unknown. Assuming that $g(i\omega)$ is the same for all noise sources, it may be determined by making the calculations to agree with measurements at one particular location. This gives a uniquely determined noise distribution over the surface of the earth and the accuracy of this model can be checked by com-

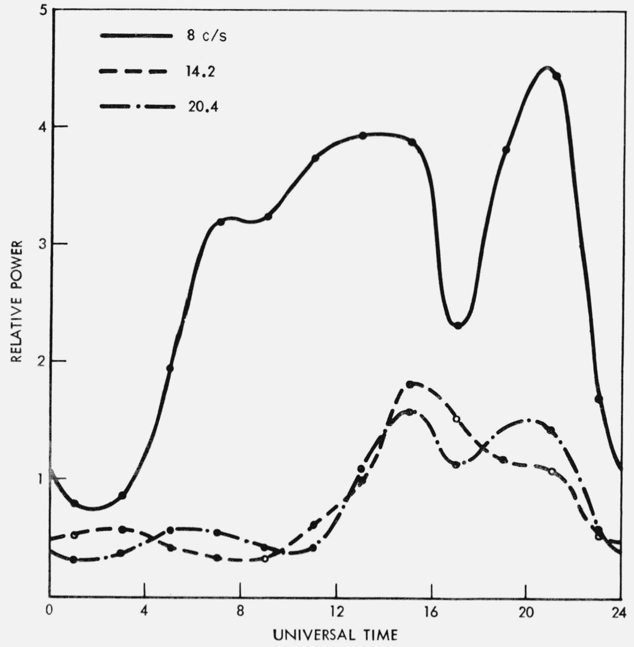

FIGURE 12. *Calculated diurnal variation of noise power at the peak of the resonance modes—wintertime—Boston, Mass.*

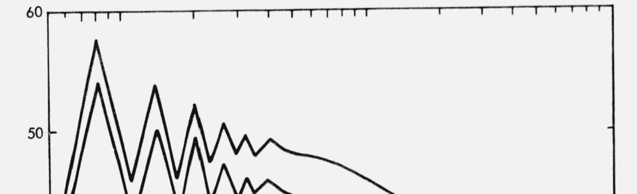

FIGURE 13. *Calculated noise spectrum for local noon in northern hemisphere.*

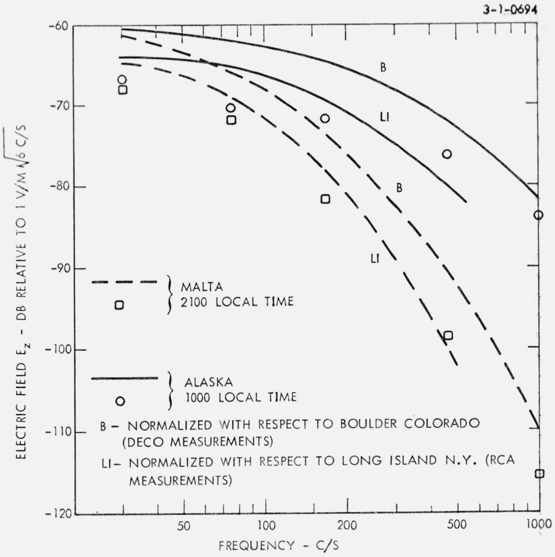

FIGURE 14. *Comparison of calculated field strength's data with July 1963 measurements by DECO and RCA.*

paring the calculations with measurements at other locations. In an example, measurements by DECO at Boulder, Colo., [Maxwell, 1963, private communication] and by RCA at Long Island, N.Y. [Powers, 1964, private communication], were used to determine $g(i\omega)$. The calculations by Abraham [1964, private communication] based on these values of $g(i\omega)$ have been compared with simultaneous measurements made in Malta and Alaska [Maxwell, 1963, private communication]. As seen from figure 14 the analytic curves agree in shape with the measurements, but the normalization with respect to Boulder gives somewhat higher predictions of the noise level. This may possibly be caused by summertime thunderstorm activity in the mountains near Boulder, Colo., that would give significant near-fields and a high estimate for the function $g(i\omega)$. The calculations are extended above the region of the earth-ionosphere cavity resonances to show that more significant noise level changes occur at higher frequencies.

In the calculations leading to figures 12 through 14 the day and night effects are considered most accurately if the day and night boundary is located symmetrically with respect to source (i.e., source at local noon or midnight). For nonsymmetrical day and night boundaries it is difficult to estimate the effects of wave front distortion. When neglecting such distortions (or assuming propagation strictly in the θ-direction), this method can account for $\nu(\theta)$ variations, which may be caused by ionosphere parameter changes or by changes of the static magnetic field along the surface of the earth. The assymptotic expansions of $P_\nu(-\cos\theta)$ used in (13) or (9) can be modified to increase their accuracy for observations near the source or its antipode [Abraham, 1965].

6. Variations of Resonance Frequencies

Diurnal frequency variations of the earth-ionosphere cavity modes have been reported by Balser and Wagner

[1962b], Gendrin and Stefant [1963], Rycroft [1963] and by Chapman and Jones [1964b]. The resonance frequencies vary by about ±0.2 to 0.3 c/s, but the changes are not simultaneous for the different modes; a low resonance frequency of one mode occurs simultaneously with a high frequency of a different mode.

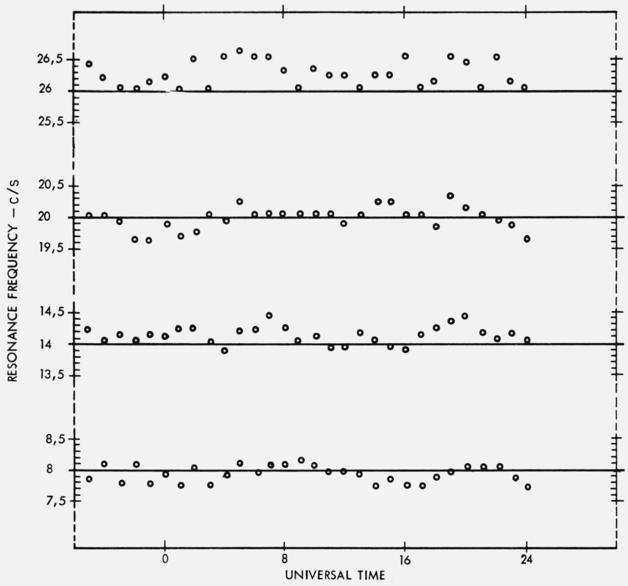

FIGURE 15. *Diurnal frequency variations measured by Gendrin and Stefant.*

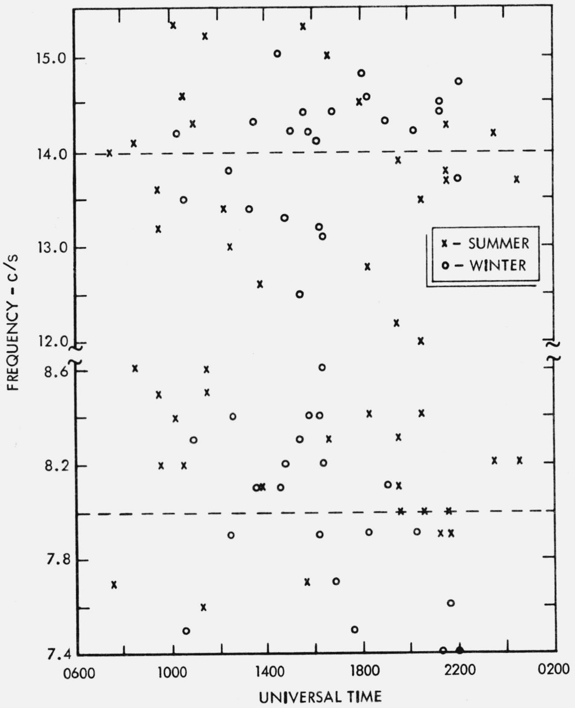

FIGURE 16. *Diurnal frequency variations measured by Rycroft.*

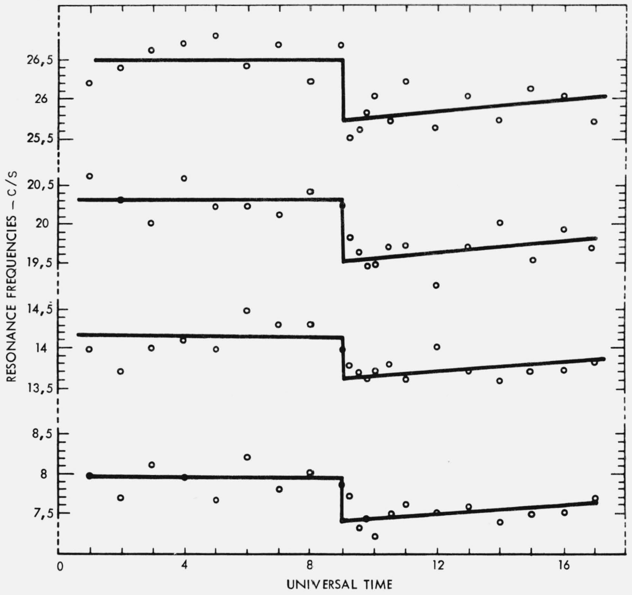

FIGURE 17. *Nuclear effects on earth-ionosphere cavity resonances.* Measurements by Gendrin and Stefant.

A high altitude nuclear explosion has been shown to affect the earth-to-ionosphere cavity resonances by simultaneously lowering all the resonance frequencies [Gendrin and Stefant, 1962a and b; Balser and Wagner, 1963]. Tepley et al., [1963] noted also an increase of signal strength for observations near the conjugate point of the explosion. The lowering of resonance frequency shown in figure 17 has been attributed to worldwide lowering of the effective ionosphere in a model of a single layer isotropic ionosphere [Gendrin and Stefant, 1962b]. The depression of the resonance frequency can be also explained by assuming an exponential daytime ionosphere to be effective around the globe [Row, 1964, private communication]. As an illustration the daytime resonance frequencies of table 1 compare with the measured resonance frequencies of figure 17. There are VLF observations [Zmuda et al, 1963] that can be explained by a lowering (or increasing electron density) of the *D*-region boundary. Brady et al., [1964] interpreted VLF diurnal phase shifts to show that some enhanced *D*-region ionization persists for several days following the event.

The lowering of the resonance frequency following a nuclear explosion has been attributed by Chapman and Jones [1964a] in their two-layer ionosphere model solely to a decreasing conductivity of the *E*-region.

After measuring a change of the resonance frequencies following a nuclear explosion, it is possible to advance simple ionosphere models for explaining a particular change. However, no attempt has been reported to predict resonance shifts for different nuclear yields and detonation altitudes. This rather complex problem requires a quantitative evaluation of the various nuclear mechanisms that may lead to ionosphere perturbations. If such information is available, the computation of the resonance shifts is relatively simple. It would be also of interest to investigate the effects of sudden ionosphere disturbances (SID) and to discriminate between SID's and nuclear effects.

The resonant frequency changes can be considered more accurately with nonuniform models of the propagation geometry. A perturbation method [Wait, 1964b] or numerical techniques [Madden and Thompson, 1964] have been used for solving such problems.

Madden and Thompson [1964] compute the local surface impedance and propagation constant for a uniform cylindrical shell between the anisotropic ionosphere and the earth using matrix multiplication techniques. There is a smooth variation of the parameters between polar and equatorial regions and also between day and night hemispheres. These local surface impedances and propagation constants are assumed to characterize a spherical transmission line of nonuniform parameters. The resonance frequencies and Q factors are computed numerically using a lumped parameter approximation of the two-dimensional transmission line. This model has been applied to explain the diurnal frequency variations and also the frequency changes due to a nuclear explosion.

7. Areas of Future Work

There are numerous observations of earth-ionosphere cavity resonances and existing ionosphere and source models account for most observed characteristics like the resonance frequency, Q factors, shape of the noise spectra, and diurnal variations of the received power.

Future work appears to be required in a systematic study of diurnal and seasonal amplitude and frequency variations of resonances. These changes should also be correlated with ionospheric disturbances. The parameters of the nighttime ionosphere require also a future clarification. The variations of the ionosphere parameters and the variations of the magnetic field vector along the surface of the earth should be incorporated in future calculations [Wait, 1964b; Madden and Thompson, 1964]. Much work remains to be done for assessing nuclear effects on the earth ionosphere resonance and for discriminating between lower level nuclear and sudden ionospheric disturbances.

The work of Sylvania reported in this paper has been supported principally by the Office of Naval Research under Contract Nonr 3185(00).

8. Appendix. Q Factors

The basic definition of the Q factor of a cavity with lossy walls is

$$\frac{1}{Q_i}=\frac{\int\int \mathbf{E}\times\mathbf{H}^*\cdot\mathbf{ds}}{\frac{\omega\epsilon_0}{2}\int\int\int \mathbf{E}\cdot\mathbf{E}^*dv+\frac{\omega\mu_0}{2}\int\int\int \mathbf{H}\cdot\mathbf{H}^*dv} \tag{14}$$

where the surface integral is extended over the cavity walls and the volume integral covers the inside of the cavity. The ground surface is assumed to be a perfect conductor and E_θ is related to H_ϕ at the ionospheric boundary by the surface impedance Z_s as $E_\theta=Z_sH_\phi$. However E_θ is assumed to be small relative to E_r for vertically polarized waves. Equation (14) is rewritten as

$$\frac{1}{Q_i}=\frac{2\int\int \mathrm{Re}Z_s|H_\phi|^2(a+h)^2\sin\theta d\theta d\phi}{\omega\int\int\int[\epsilon_0|E_r|^2+\mu_0|H_\phi|^2]r^2dr\sin\theta d\theta d\phi} \tag{15}$$

$$\approx\frac{2\mathrm{Re}Z_s}{\omega h}\frac{\int|H_\phi|^2\sin\theta d\theta}{\int[\epsilon_0|E_r|^2+\mu_0|H_\phi|^2]\sin\theta d\theta} \tag{16}$$

where a is the radius of earth, h is the ionospheric height and where the cavity is assumed to be uniform. For excitation by a radial electric dipole of moment Idr the field component E_r and H_ϕ follow from (21), (61), and (62) of Galejs [1964a] as

$$E_r=\frac{iIdr}{4\pi h\omega\epsilon a^2}\frac{n(n+1)(2n+1)}{D_n}P_n(\cos\theta) \tag{17}$$

$$H_\phi=-\frac{Idr}{4\pi ha}\frac{(2n+1)}{D_n}\frac{\partial}{\partial\theta}P_n(\cos\theta) \tag{18}$$

with $D_n=n(n+1)-\nu(\nu+1)$. Substituting (17) and (18) in (16) and evaluating the integrals gives

$$\frac{1}{Q_i}=\frac{2\mathrm{Re}Z_s}{\omega\mu_o h\left[\frac{n(n+1)}{(k_0a)^2}+1\right]} \tag{19}$$

where $k_0=\omega\sqrt{\mu_0\epsilon_0}$. From the approximate modal equation

$$S=\sqrt{1+\frac{iZ_s}{\omega\mu_0h}} \tag{20}$$

it follows that

$$2\ \mathrm{Re}S\ \mathrm{Im}S=\frac{\mathrm{Re}Z_s}{\omega\mu_0h}\cdot \tag{21}$$

Applying (21) and the resonance condition

$$n(n+1)\approx(k_0a\ \mathrm{Re}S)^2 \tag{22}$$

(19) simplifies to

$$\frac{1}{Q_i}=\frac{4\ \mathrm{Re}S\ \mathrm{Im}S}{(\mathrm{Re}S)^2+1} \tag{23}$$

which was listed as (4). When considering the differences of E_r, H_ϕ, Z_s, and h between the day and night hemispheres it follows that

$$\frac{1}{Q_i}\approx\frac{2[\mathrm{Re}S\ \mathrm{Im}S|_{\mathrm{day}}+\mathrm{Re}S\ \mathrm{Im}S|_{\mathrm{night}}]}{(\mathrm{Re}S|_{\mathrm{avg}})^2+1}\,. \tag{24}$$

The form (5) or (6) is obtained by neglecting the E integral or H integral, and by doubling the value of the H integral or E integral in (14).

9. References

Abraham, L. G. Jr., Aspects of the terrestrial ELF noise spectrum when near the source or its antipode, Radio Sci., J. Res. NBS **69D** (July 1965).

Balser, M., and C. A. Wagner (Nov. 1960), Observations of earth-ionosphere cavity resonances, Nature **188,** 638–641.

Balser, M., and C. A. Wagner (Feb. 1962a), Diurnal power variations of the earth-ionosphere cavity modes and their relationship to worldwide thunderstorm activity, J. Geophys. Res. **67,** No. 2, 619–625.

Balser, M., and C. A. Wagner (Sept. 1962b), On frequency variations of the earth-ionosphere cavity modes, J. Geophys. Res. **67,** No. 10, 4081–4083.

Balser, M., and C. A. Wagner (July 1, 1963), Effect of a high-altitude nuclear detonation on the earth-ionosphere cavity, J. Geophys. Res. **68,** No. 13, 4115–4118.

Benoit, R., and A. Houri (1961), Propagation of very-low-frequencies in the earth-ionosphere system, Ann. Geophys. (France) **17,** No. 4, 370–373.

Benoit, R., and A. Houri (Nov. 5, 1962), Power spectrum measurements of geophysical noise. Application to earth-to-ionosphere cavity, C. R. Acad. Sci. (France) **255,** 2496–2498.

Bourdeau, R. E., E. C. Whipple, and J. F. Clark (Oct. 1959), Analytic and Experimental Conductivity between the stratosphere and the ionosphere, J. Geophys. Res. **64,** 1363–1370.

Brady, A. H., D. D. Crombie, A. G. Jean, A. C. Murphy, and F. K. Stede (May 1, 1964), Long-lived effects in the D region after the high altitude nuclear explosion of 9 July 1962, J. Geophys. Res. **69,** 1921.

Brock-Nannestad, L. (Sept. 1962), Bibliography electromagnetic phenomena with special reference to ELF (1–3000 c/s) TR 10 Saclant, ASW Research Center, La Spezia, Italia.

Chapman, F. W., and D. L. Jones (May 16, 1964a), Earth-ionosphere cavity resonances and the propagation of extremely low frequency radio waves, Nature **202,** 654–657.

Chapman, F. W. and D. L. Jones, Observations of earth-ionosphere cavity resonances and their interpretation in terms of a two-layer ionosphere model, Radio Sci., J. Res. NBS **68D,** 1177–1185 (Nov. 1964b).

Chapman, F. W., and R. C. Macario (May 1956), Propagation of audio frequency radio waves to great distances, Nature **177,** 930–933.

Egeland, A. (1964a) ELF (500–1000 c/s) emissions at high latitude, Proc. NATO Advanced Study Institute on Natural Electromagnetic Phenomena Below 30 kc/s, Bad Homburg, West Germany, July–Aug. 1963, pp. 157–166 (Plenum Press, New York, N.Y.).

Fournier, H. (Aug. 17, 1960), Some aspects of the first high-frequency geomagnetic recordings obtained at Garchy, C. R. Acad. Sci. (France) **251,** No. 7, 962–964.

Galejs, J. (1961a), Terrestrial extremely-low-frequency noise spectrum in the presence of exponential ionospheric conductivity profiles, J. Geophys. Res. **66,** No. 9, 2787–2792.

Galejs, J. (Nov. 1961b), ELF waves in presence of exponential ionospheric conductivity profiles, IRE Trans. Ant. Prop. **AP-9,** 554–562.

Galejs, J. (July 1962a), A further note on terrestrial extremely-low frequency propagation in the presence of an isotropic ionosphere with an exponential conductivity-height profile, J. Geophys. Res. **67,** No. 7, 2715–2728.

Galejs, J. (1962b), Terrestrial extremely-low frequency propagation in the presence of an isotropic ionosphere with an exponential conductivity height profile, Proc. International Conference on the Ionosphere, London, pp. 467–474 (Chapman and Hall).

Galejs, J. (1964a), Terrestrial extremely-low frequency propagation, Proc. NATO Advanced Study Institute on Natural Electromagnetic Phenomena Below 30 kc/s, Bad Homburg, West Germany, July–Aug. 1963, pp. 205–258 (Plenum Press, New York, N.Y.).

Galejs, J. (June 1964b), ELF and VLF waves below an inhomogeneous anisotropic ionosphere, Radio Sci. J. Res. NBS **68D,** No. 6, 693–707.

Galejs, J., and R. V. Row (Jan. 1964), Propagation of ELF waves below an inhomogeneous anisotropic ionosphere, IEEE Trans. Ant. Prop. **AP-12,** No. 1, 74–83.

Galejs, J., On the terrestrial propagation of ELF and VLF waves in the presence of a radial magnetic field, Radio Sci., J. Res. NBS **69D,** No. 5 (May 1965).

Gendrin, R., and R. Stefant (Sept. 1962a), Magnetic records between 0.2 and 30 c/s, AGARD Conference on Propagation of Radio Frequencies Below 300 kc/s, Munich, Germany.

Gendrin, R., and R. Stefant (1962b), Effects of the very high altitude nuclear explosion of the July 9, 1962 on the earth to ionosphere cavity resonances, experimental results and interpretation, C. R. Acad. Sci. **255,** 2273–2275 and 2493–2495.

Handbook of Geophysics (1960), USAF Air Research and Development Command, Airforce Cambridge Research Center (Macmillan Co., New York, N.Y.).

Harris, F. B., Jr., and R. L. Tanner (1962), A method for determination of lower ionosphere properties by means of field measurements on sferics, J. Res. NBS **66D** (Radio Prop.), No. 4, 463–478.

Hepburn, F., and E. T. Pierce (May 9, 1953), Atmospherics with very-low-frequency components, Nature (GB) **171,** 837–838.

Jean, A. G., Jr., and A. C. Murphys, J. R. Wait, and D. F. Wasmundt (1961), Propagation attenuation rates at ELF, J. Res NBS **65D** (Radio Prop.), No. 5, 475–479.

Keefe, T. J., C. Polk, and H. L. Koenig (1964), Results of simultaneous ELF measurements at Brannenburg (Germany) and Kingston, R.I., ULF Symposium, Boulder, Colo. Aug. 17–20, 1964.

Koenig, H. (Dec. 1958), Atmospherics geringster frequenzen, Dissertation an der Technischen Hochschule München.

Koenig, H. (July 1959), Atmospherics geringster frequenzen, Z. Angew. Phys. **11,** No. 7, 264–274.

Koenig, H., E. Haine, and C. H. Antoniadis (Aug. 1961), Messung von 'atmospherics' geringster frequenzen in Bonn, Z. Angew. Phys. **13,** No. 8, 364–367.

Liebermann, L. (Dec. 1956a), Extremely-low-frequency electromagnetic waves: I. Reception from lightning, J. Appl. Phys. **27,** No. 12, 1473–1476.

Liebermann, L. (Dec. 1956b), Extremely-low-frequency electromagnetic waves: II. Propagation properties, J. Appl. Phys. **27,** No. 12, 1477–1483.

Lokken, J. E., J. A. Schand, and C. S. Wright (1962), A note on the classification of geomagnetic signals below 30 c/s, Can. J. Phys. **40,** 1000–1009.

Madden, T. R. (1961), An analysis of the earth-ionosphere electromagnetic resonances, unpublished class notes for MIT course 12.88, April 24.

Madden, T. Rand, W. Thompson, Low-frequency electromagnetic oscillations of the earth-ionosphere cavity, Report on Project NRL 371–401, Geophysics Laboratory, MIT, Cambridge, Mass. (Sept. 22, 1964). Also Rev. Geophys. **3,** No. 2 (May 1965).

Pierce, E. T. (July–Aug. 1960), Some ELF phenomena, J. Res. NBS **64D** (Radio Prop.), No. 4, 383–386.

Pierce, E. T. (July 1, 1963), Excitation of earth-ionosphere cavity resonances by lightning flashes, J. Geophys. Res. **68,** No. 13, 4125–4127.

Polk, C., and F. Fitchen (May–June 1962), Schumann resonances of the earth-ionosphere cavity—extremely low frequency reception at Kingston, R.I., J. Res. NBS **66D** (Radio Prop.), No. 3, 313–318.

Polk, C. (1962), ELF propagation and the earth-ionosphere resonant cavity, paper presented at the meeting of the Ionospheric Research Committee of AGARD in Munich, West Germany, Sept. 17–21, 1962.

Raemer, H. R. (1961a), On the extremely-low-frequency spectrum of earth-ionosphere cavity response to electric storms, J. Geophys. Res. **66,** No. 5, 1580–1583.

Raemer, H. R. (Nov.–Dec. 1961b), On the spectrum of terrestrial radio noise at extremely low frequencies, J. Res. NBS **65D** (Radio Prop.), No. 6, 581–594.

Rycroft, M. J. (Dec. 1963), Low frequency disturbances of natural origin of the electric and magnetic fields of the earth, Ph.D. Thesis, University of Cambridge.

Schumann, W. O. (1952a), Über die strahlungslosen Eigenschwingungen einer leitenden Kugel, die von einer Luftschicht und einer Ionosphärenhülle umgeben ist, Z. Naturforsch, **72,** 149–154.

Schumann, W. O. (1952b), Über die Dämpfung der elektromagnetischen Eigenschwingungen des Systems Erde-Luft-Ionosphäre, Z. Naturforsch **72,** 250–252.

Schumann, W. O. (Aug. 1957), Elektrische Eigenschwingungen des Hohlraumes Erde-Luft-Ionosphäre, Z. Angew. Phys. **9,** 373–378.

Seshadri, S. R., Radiation from electromagnetic sources in a plasma, IEEE Transactions on Antennas and Prop. **AP–13,** No. 1, 79–88 (Jan. 1965a).

Seshadri, S. R. (March 1965b), VLF propagation in a sharply bounded, compressible isotropic ionosphere, IEEE Trans. Ant. Prop. **AP–13,** No. 2, 311–317.

Tepley, L. R. (Dec. 1959), A comparison of sferics as observed in the very-low-frequency and extremely-low-frequency bands, J. Geophys. Res. **64,** No. 12, 2315–2329.

Tepley, L., R. C. Wentworth, K. G. Lambert, H. V. Prentiss, K. D. Amundsen and D. R. Hillendahl (26 Dec. 1963), Sub ELF geomagnetic fluctuations, III final report Contract AF 19(628)–462, Lockheed Missile and Space Co., Palo Alto, Calif.

Thompson, W. B. (Feb. 1963), A layered model approach to the earth-ionosphere cavity resonance problem, Ph.D. Thesis, Department of Geology and Geophysics, MIT.

VanZandt, T. E. (1964), Ionosphere, ULF Symposium, Boulder, Colo., Aug. 17–20.

Wait, J. R. (March 1958), Extension to mode theory of VLF propagation, J. Geophys. Res. **63,** 125–135.

Wait, J. R. (1962), Electromagnetic waves in stratified media (Macmillan Co., New York, N.Y.).

Wait, J. R. (Feb. 1963a), The mode theory of VLF radio propagation for a spherical earth and a concentric anisotropic ionosphere, Can. J. Phys. **41,** 299–315.

Wait, J. R. (May–June 1963b), Concerning solutions of the VLF mode problem for an anisotropic curved ionosphere, J. Res. NBS **67D** (Radio Prop.), No. 3, 297–302.

Wait, J. R. (April 1964a), On the theory of Schumann resonances in the earth-ionosphere cavity, Can. J. Phys. **42,** 575–582.

Wait, J. R. (1964b), Earth-ionosphere cavity resonances and the propagation of ELF radio waves, ULF Symposium, Boulder, Colo., Aug. 17–20. Also Radio Sci. J. NBS **69D,** No. 8.

Wait, J. R. (1964c), Cavity resonances for a spherical earth with a concentric anisotropic shell (private communication).

Wait, J. R. (1964d), A note on the Schumann problem for nonuniform boundaries (private communication).

Williams, J. C. (May 1959), Thunderstorms and VLF radio noise, Ph.D. Thesis, Harvard University, Division of Engineering and Applied Physics.

C. S. Wright (July–Aug. 1963), Conjugate relations of micropulsations and other ionospheric phenomena, NATO Advanced Study Institute on Natural Electromagnetic Phenomena Below 30 kc/s, Bad Homburg, Germany.

Zmuda, A. J., B. W. Shaw, and C. R. Haave (Feb. 1, 1963), VLF disturbances and the high-altitude nuclear explosion of July 9, 1962, J. Geophys. Res. **68,** No. 3, 745–758.

(Paper 69D8–538)

US 20110075316A1

(19) **United States**
(12) **Patent Application Publication** (10) **Pub. No.: US 2011/0075316 A1**
Rodriguez Montes et al. (43) **Pub. Date: Mar. 31, 2011**

(54) **SYSTEM FOR RECOVERING AND USING THE ELECTROSTATIC CHARGE GENERATED BY LIGHTNING**

(76) Inventors: **Angel Rodriguez Montes**, Ordino (ES); **Antonio Claveria Canal**, La Massana (ES)

(21) Appl. No.: **12/891,388**

(22) Filed: **Sep. 27, 2010**

(30) **Foreign Application Priority Data**

Sep. 29, 2009 (ES) 200901924

Publication Classification

(51) **Int. Cl.**
H05F 7/00 (2006.01)

(52) **U.S. Cl.** **361/225**

(57) **ABSTRACT**

A system for the recovery and use of the electrostatic charge present in the atmosphere during storms generated by lightning, transforming natural energy into electrical energy useful for consumption, and providing the necessary protection for lives and property against electro-atmospheric phenomena, has an ecological lightning conductor, an electrical conductor connected by one end to the low-frequency electrode of the lightning conductor, and an enhanced intelligent earth wire, whereto the conductor is joined by its other end. The conductor includes a series of probes and magnetic field detectors which will send information to the microprocessor and from there to the mainframe computer. The control device allows regulation of the earth resistance, thereby generating more or less charge, diverting the energy to the accumulators or to earth through dissipator electrodes.

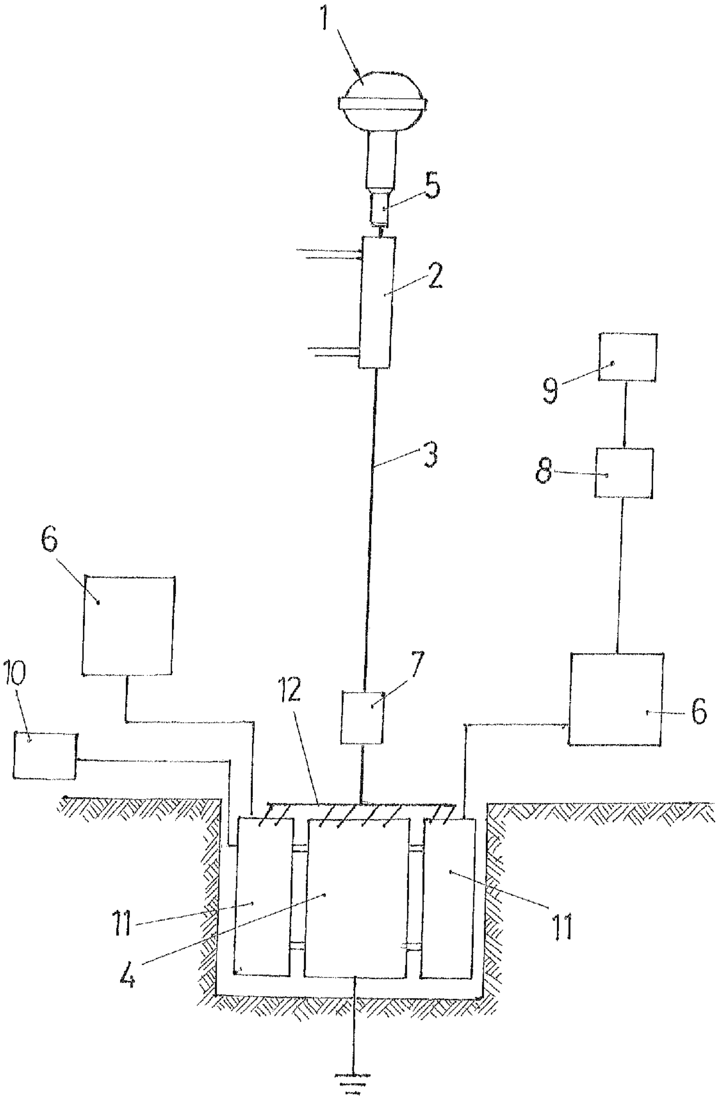

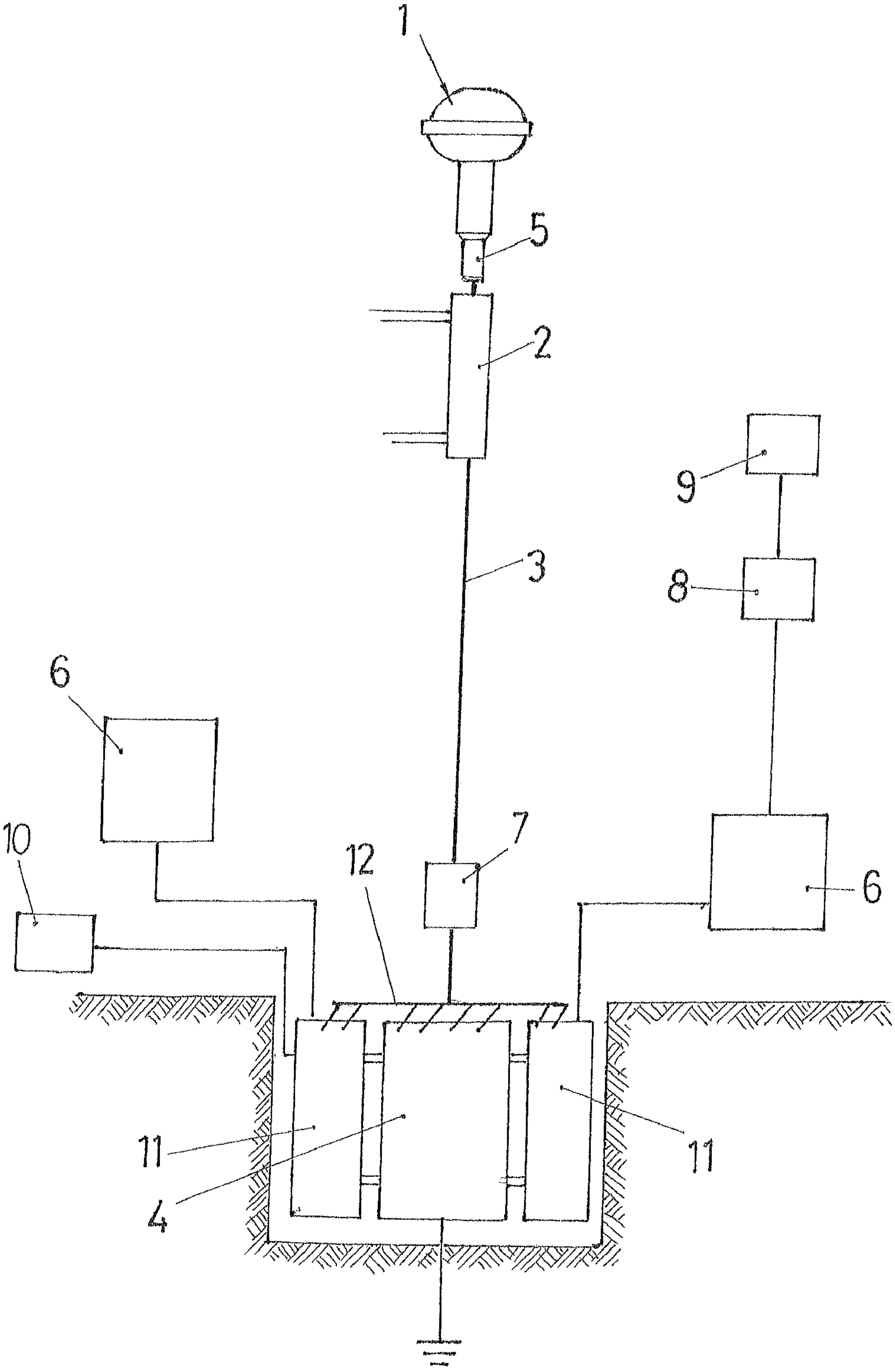
1
5
2
9
3
8
6
10
7
12
6
11
4
11

FIG. 1

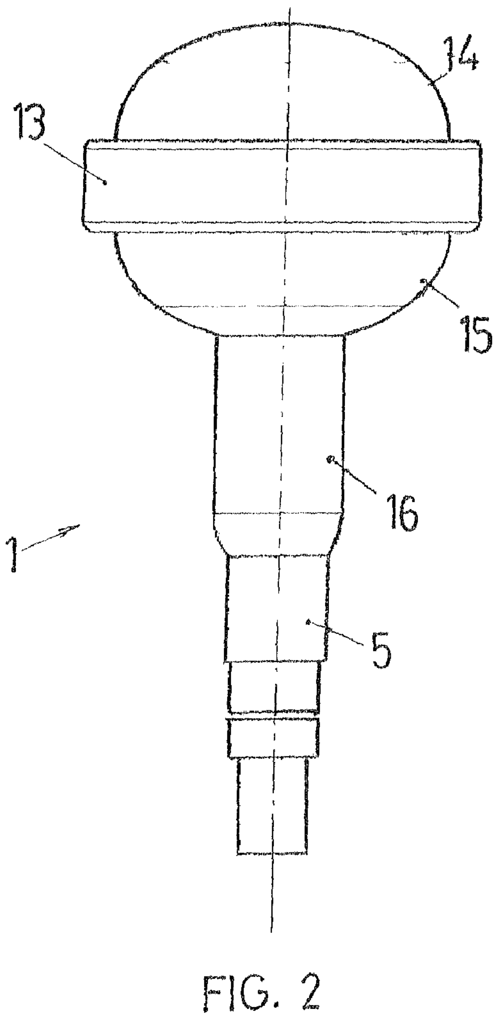
13
14
15
16
1
5

FIG. 2

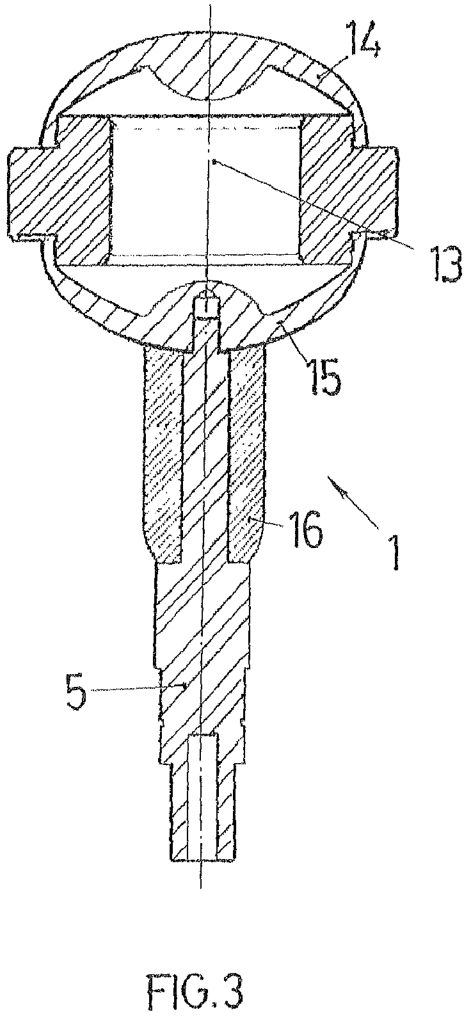
14
13
15
16
1
5

FIG.3

SYSTEM FOR RECOVERING AND USING THE ELECTROSTATIC CHARGE GENERATED BY LIGHTNING

CROSS-REFERENCE TO RELATED APPLICATION

[0001] The invention described and claimed hereinbelow is also described in Spanish Patent Application ES 2009 01924 filed on Sep. 29, 2009. This Spanish Patent Application, whose subject matter is incorporated here by reference, provides the basis for a claim of priority of invention under 35 U.S.C. 119(a)-(d)

BACKGROUND OF THE INVENTION

[0002] The present invention relates to a system which allows the recovery and use of the electrostatic charge present in the atmosphere during storms which generates lightning, thereby achieving the transformation of natural energy into electrical energy useful for domestic and industrial consumption, while providing essential and adequate protection to persons, animals and property against the aforementioned atmospheric electrical phenomena.

[0003] Different technical and scientific arguments relative to the effective application of the system that is the object of this invention will be explained in detail in this description, in addition to describing the elements that compose the aforementioned system.

[0004] During an atmospheric storm formation process, variable electrical fields appear which are responsible for the generation and appearance of lightning, which are discharges between the ground and the atmosphere. As a result of the generation of electrical charges within the cloud, the atmosphere increases its electrical potential, inducing a variable ground voltage.

[0005] It has been scientifically proven that if this ground voltage is controlled, the current flowing through the ground can also be controlled, thereby adequately leveraging the full potential of this energy.

[0006] This will primarily avoid the electrical saturation of that area, which could call, excite or create a lightning bolt therein and, secondly, this energy may be accumulated in the most adequate manner.

[0007] This process for recovering energy from the atmosphere is carried out before and during the first phase of the generation of lightning charges. The effect of attenuating the electrical field of the atmosphere causes the disappearance of the lightning discharge to ground, on not having sufficient potential for formation thereof, thereby allowing prevention and protecting persons, animals and property.

[0008] Widely known are the experiments and achievements of Benjamin Franklin in the 18th century, inventor of the passive lightning conductor equipped with a priming device or those developed subsequently using the "multipoint" system, the action of which is based on ionisation and excitation, constant or by impulses, of the electro-atmospheric field in order to capture the lightning discharge.

[0009] Mention must be made of radioactive lightning conductors, currently in disuse for evident safety reasons due to radiation hazard; likewise, other devices based on the filtration of high frequencies and harmonic components are also known to exist, equipped with two or more external electrical connections and having different components such as coils, resistors and condensers, duly connected and having the necessary siliceous sand for absorbing the thermal inertia produced when the system functions as a whole.

[0010] Reference is made herein to patent of invention P 200202884/2, by the same inventors, which relates to a deionising lightning conductor of electrostatic charge, based on three insulators equipped with a variable noble gas dielectric in its interior.

[0011] According to the data provided by the world meteorological detection system, more than four thousand storms form in the Earth's atmosphere every day, which generate around eight million lightning bolts.

[0012] The phenomenon called "lightning" is the electrical reaction caused by the saturation of electrostatic charges generated by the progressive accumulation of an electric field between the ground and the cloud, during the activation of a typical storm.

[0013] This phenomenon is randomly present as of an electro-atmospheric potential of 10/45 kV. It is generated between two points of attraction of different polarity and same potential for the purpose of compensating the saturation of the electrostatic charge. Lightning charge density is proportional to the saturation of the electrostatic charge of the area: the higher the charge density, the higher the risk of generating a tracker or "leader" and, subsequently, a lightning discharge.

[0014] Lightning charge intensity is variable, depending on the critical moment when the resistance in the air between the two transfer points break down; it will also be influenced by serially exposed materials such as soils, rocks, wood, iron or other metals, lightning conductors, earth wires, etc. A lightning bolt can transport an average charge of 20 GW of power per second, equivalent to one hundred million ordinary incandescent light bulbs.

[0015] This electrical phenomenon is represented in the form of an electronic shadow, which determines the elements that will be affected by charge exchange.

[0016] Studies carried out on the electro-atmospheric field at ground level reveal that charge distribution during the generation of the high-voltage field is not static but rather dynamic, on being randomly formed and generating the moving tracker or leader and different geographical points at the same time. The intensity and location of this electronic shadow can change radically and affect the low or lateral areas of high structures or buildings.

[0017] The behavioural study of a conventional pointed lightning conductor has demonstrated that they generate, to a greater or lesser extent, electromagnetic effects in the installations themselves, as well as in neighbouring installations, upon capturing the lightning bolt. In the case of ionising lightning conductors, accidents frequently occur for this reason.

[0018] Recently, enhancements made in electrostatic charge deionising lightning conductors through the application and incorporation of new technologies, such as a special cable adaptable to different forms of energy transport, eliminates causes of accidents and improves the reliability of the facilities, in addition to acting within an ecological electrical environment free of electromagnetic interferences.

SUMMARY OF THE PRESENT INVENTION

[0019] Accordingly, it is an object of the present invention to provide a new system which allows recovering and using the electrostatic charge generated by lightening and provides for highly advantageous results.

[0020] In keeping with these objects this system allows to cover all levels of electrical risk, diversity in the type and polarity of lightning bolts for any type of structure. It is an individual and collective protection system which bases its operating principle on the orderly transformation of ions flowing in a current, and is useful for working at low voltages and previously be derived to an accumulation area.

[0021] The system is composed of an external electrode acting as an ecological lightning conductor, an electric conductor, designed and calculated for a wide range of frequencies, and having anti-corrosive protection and an intelligent earth wire, complemented by energy accumulators and dissipator electrodes, capable of deriving all current leakage to earth, all of said components being controlled by a microprocessor which manages the energy as it appears, controlling its intensity, as well as the charge of the aforementioned accumulators for subsequent consumption.

[0022] The system also incorporates an alarm-based alert device, both private and public, for occupational risk prevention management against lightning risk.

[0023] At meteorological level, the system allows data to be obtained relative to the behaviour of the electric field, temperatures, atmospheric pressure, humidity and quantity of discharges.

[0024] The invention being described incorporates magnetic field detectors and detectors for controlling earth leakage current intensity, for subsequent analysis thereof and the consequent alarms. These detectors send the information to the microprocessor, upon recording the abrupt magnetic field variations.

[0025] The deionisation system of the lightning conductor itself facilitates the appearance of small currents in the earth wire, where charges are intelligently accumulated for subsequent use thereof. Depending on the intensity of the natural earth electric field, the ecological lightning conductor system generates the appearance of an earth leakage current which passes through the electrical conductor from the lightning conductor, strictly speaking, to the earth wire. The automatic control system regulates earth resistance in order to generate more or less charge and divert energy to the accumulators or to the enhanced intelligent earth wire.

[0026] The intelligent earth wire incorporates an electronic system which automatically regulates its own electrical resistance with reference to the physical earth. This control system allows controlled energy transfer, maintaining a resistance of less than 5 ohms in the earth installation during the four seasons of the year and varying, as required, the electrical conductivity of the physical earth and its humidity in order to facilitate or stop leakage currents.

[0027] The intelligent earth wire consists of special electrodes, according to the type of soil and/or terrain, aimed at facilitating current leakage and avoiding corrosion of the metals. These electrodes are installed conveniently protected against the aggressiveness of any chemical or electrochemical reaction.

[0028] The novel features which are considered as characteristic for the present invention are set forth in particular in the appended claims. The invention itself, however, both as to its construction and its method of operation, together with additional objects and advantages thereof, will be best understood from the following description of specific embodiments when read in connection with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0029] FIG. **1** is a longitudinal diagram of the system in accordance with the present invention, illustrating the layout of the different components thereof;

[0030] FIG. **2** shows an outer view of the ecological lightning conductor of the inventive system; and

[0031] FIG. **3** is a complementary view of the preceding figure, showing a fully sectional view of the lightning conductor of the inventive system.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0032] A system for recovering and using the electrostatic charge generated by the lightning of the present invention includes an ecological lightning conductor **1**, disposed on a mast **2**, which is installed at the highest point of the existing structure by means of adequate bearings or on a newly built tower, an electrical conductor **3** and an intelligent earth wire **4**.

[0033] The electrical conductor **3** is connected by one end to the mast or low electrical resistance electrode **5** of the lightning conductor **1**, and connected to the intelligent earth wire **4** by the other end, said conductor incorporating a series of probes **6** and magnetic field detectors **7** that will send information to the microprocessor **8**.

[0034] These data, recorded by the microprocessor **8**, are sent to a mainframe computer **9** where they are processed, being immediately displayed on a screen or sent to an "Internet"-type data transport network.

[0035] The intelligent earth wire **4** is equipped with a control device **10** which will analyse, in real time, the earth current leakages, sending the data to the aforementioned mainframe computer **9**. Depending on the intensity of the leakage, the system will manage the information, generating alarms or modifying the processes for sending current to the energy accumulators **11**, disposed next to the intelligent earth wire **4**. This earth wire **4** also has dissipator electrodes **12** at ground level.

[0036] The ecological lightning conductor **1** has a toroidal-shaped body **13**, closed by its upper part by an upper frame **14** and completed by a lower frame **15** whereto, in the central appendix **16** of its lower part, is fixed to the mast or low electrical resistance electrode **5**.

[0037] The dielectric product, which is essential for the object to act as a lightning conductor in itself, is disposed in the interior of the toroidal body **13** of the ecological lightning conductor. The dielectric product can be of noble gas.

[0038] The previously described characteristics of the ecological lightning conductor confer the quality of deionising lightning conductor to this system component, which allows the appearance, as mentioned earlier, of small currents in the intelligent earth wire **4**, where the charges are accumulated for their subsequent consumption in the energy accumulators **11**.

[0039] During a storm process, a high-voltage field is generated at ground level which is proportional to the charge of the cloud and the distance that separates it from the ground.

[0040] As of a certain magnitude of the natural electric field on the physical earth, the equipotential installation of earth wires of the system facilitates transfer of the charges through the electrical conductor **3**, charges which, regardless of their polarity, are concentrated on the mast or lower electrode **5** of the lightning conductor **1**, disposed at the highest point of the installation that configures the system, which is connected to the earth wire by means of the electrical conductor **3**. The current that circulates through the lightning conductor-conductor-earth wire system is proportional to the charge within the cloud.

[0041] The low resistance of this mast or lower electrode **5** of the lightning conductor **1**, facilitates the capture of opposing charges in the electrode or upper body **13**. During this energy transfer process, a small current flow takes place between the cathode and anode in the interior of the lightning conductor, in the direction of the current flow.

[0042] The resulting effect causes the appearance of a weak leakage current that derives to the electrical earth wiring of the installation, which is controlled by a microprocessor **8** which, in turn, causes current shunting to an accumulation system **11** or to dissipator electrodes **12** at ground level.

[0043] The evolution of the electro-atmospheric phenomenon is recorded in a database by means of electric probes **6** installed in the electrical conductor **3**, disposed between the lightning conductor **1** and intelligent earth wire **4**, and magnetic field detectors **7**. The data are stored, in real time, in a microprocessor **8** thanks to the transfer of data from the probes **6** to the mainframe computer **9**, via radio or conductor. These data may be viewed on a PC screen and statistics calculated, both real time and historical, for the purpose of carrying out electromagnetic and electrical prevention studies.

[0044] The novel technology of the system that is the object of the invention, and which is being described, allows optimisation of industrial process management and enhancement of personal electrical safety.

[0045] Depending on the intensity of the current that flows through the conductor **3** towards the earth wire **4**, high electric field alarms may be generated, consequently alerting of the risk of electrical storm in the area, with lightning activity.

[0046] Having sufficiently described the characteristics of the system that is the object of the present invention, we must point out that any variation in dimensions, shapes, types of materials and layout of the different components shall not alter the essentiality of the invention and may be necessary according to the protection requirements of the desired area, based on factors such as the geographical area, environmental context and dimensions.

[0047] It will be understood that each of the elements described above, or two or more together, may also find a useful application in other types of constructions differing from the types described above.

[0048] While the invention has been illustrated and described as embodied in a system for recovering and using the electrostática charge generated by ligntening, it is not intended to be limited to the details shown, since various modifications and structural changes may be made without departing in any way from the spirit of the present invention.

[0049] Without further analysis, the foregoing will so fully reveal the gist of the present invention that others can, by applying current knowledge, readily adapt it for various applications without omitting features that, from the standpoint of prior art, fairly constitute essential characteristics of the generic or specific aspects of this invention.

What is claimed as new and desired to be protected by Letters Patent is set forth in the appended claims.

1. A system for recovering and using an electrostatic charge generated by lightning and being present in an atmosphere during storms for transforming a natural energy into electrical energy and providing a necessary protection to persons, animals and property, the system comprising components including an element selected from the group consisting of an external electrode and an ecological lightning conductor, a mast fixing said electrode, an intelligent earth wire, an electrical conductor joining said element to said intelligent earth wire, energy accumulators, dissipation electrodes, electric probes, magnetic field detectors, and a microprocessor controlling all the components of the system and transferring data to a mainframe computer.

2. A system for recovering and using the electrostatic charge as defined in claim **1**, wherein said ecological lightning conductor has a toroidal-shaped body housing containing a dielectric product, an upper frame closing said body, a lower frame completing said body, and a lower central appendix in which a member selected from a group consisting of said mast and a low electrical resistance electrode is fixed.

3. A system for recovering and using the electrostatic charge as defined in claim **2**, wherein said toroidal-shaped boyd of said ecological lightening conductor contains a noble gas.

4. A system for recovering and using the electrostatic charge as defined in claim **1**, wherein said electrical conductor has one end connected to a member selected from the group consisting of a mast and an electrode and is designed for a wide range of frequencies and has adequate anti-corrosion protection, said electrical conductor also having another end joined to said intelligent earth wire, said electrical conductor incorporating said electric probes and said magnetic field detectors.

5. A system for recovering and using the electrostatic charge as defined in claim **1**, wherein said intelligent earth wire is provided with disipator electrodes at a ground level according to a type of a component selected from the group consisting of soil, terrain and both.

6. A system for recovering and using the electrostatic charge as defined in claim **1**, further comprising a control device analyzing earth current leakages in real time and sending data to the mainframe computer, generating according to a leakage intensity, alarms or modifying a process for sending current to said accumulators.

* * * * *

Aug. 2, 1932. **H. L. BUMBAUGH** **1,869,661**

LIGHTNING PROTECTION SYSTEM AND METHOD

Filed July 26, 1927 2 Sheets-Sheet 1

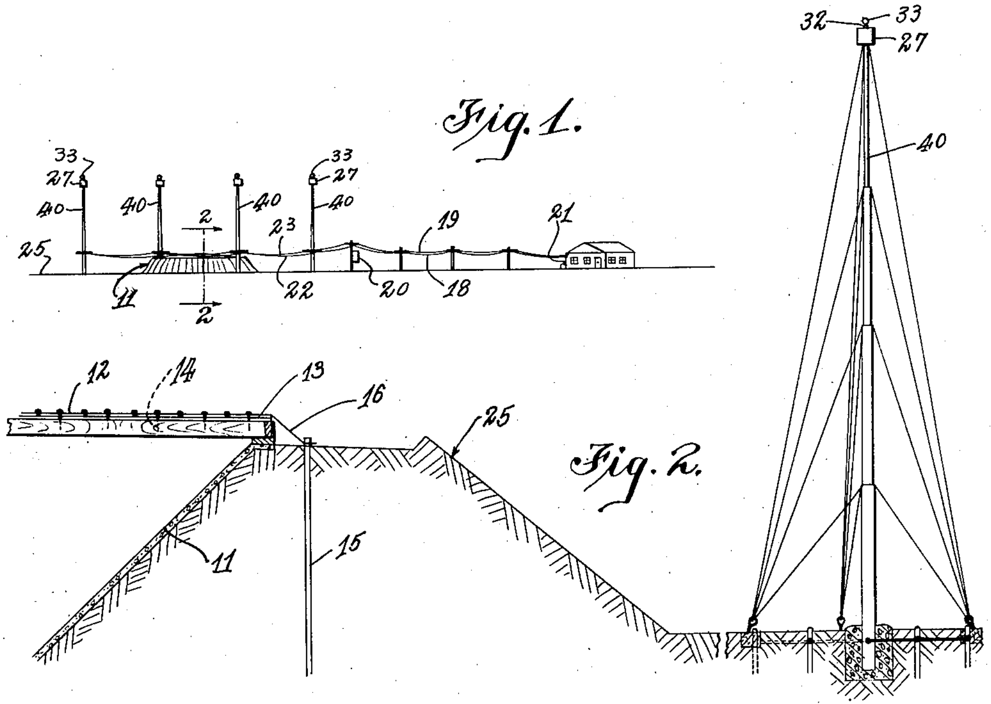
Fig. 1.
33
27
40
40
40
40
33
27
2
2
25
11
23
22
20
19
18
21
Fig. 2.
32
33
27
40
12
14
13
16
25
11
15

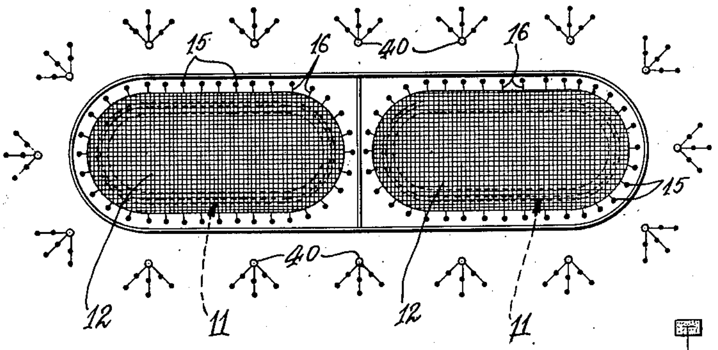
15
16
40
16
15
40
12
11
12
11

Fig. 3.

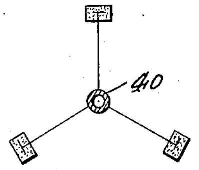

Fig. 5.

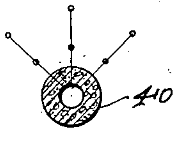

Fig. 4.

Inventor

Harold L. Bumbaugh

By Lyon & Lyon

Attorneys

Aug. 2, 1932. **H. L. BUMBAUGH** **1,869,661**

LIGHTNING PROTECTION SYSTEM AND METHOD

Filed July 26, 1927 2 Sheets-Sheet 2

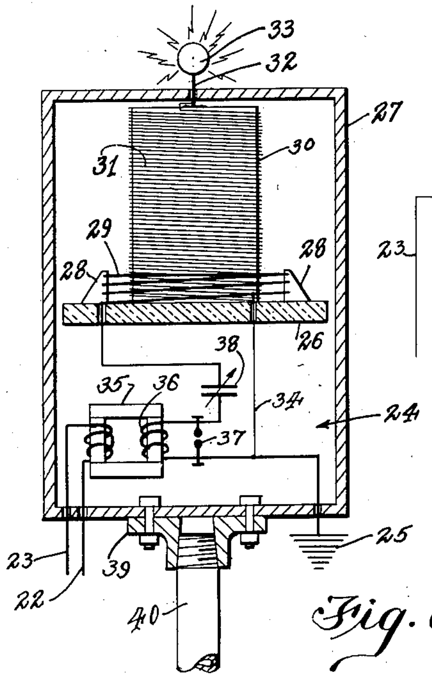
33
32
27
31
30
29
28
28
26
38
35
36
34
24
37
25
23
39
22
40

Fig. 6.

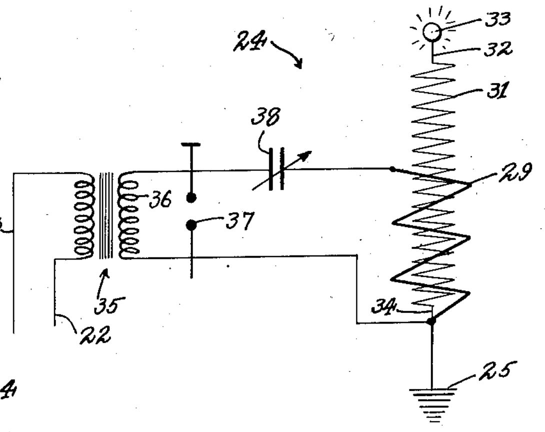
24
33
32
31
38
29
36
37
23
35
22
34
25

Fig. 7.

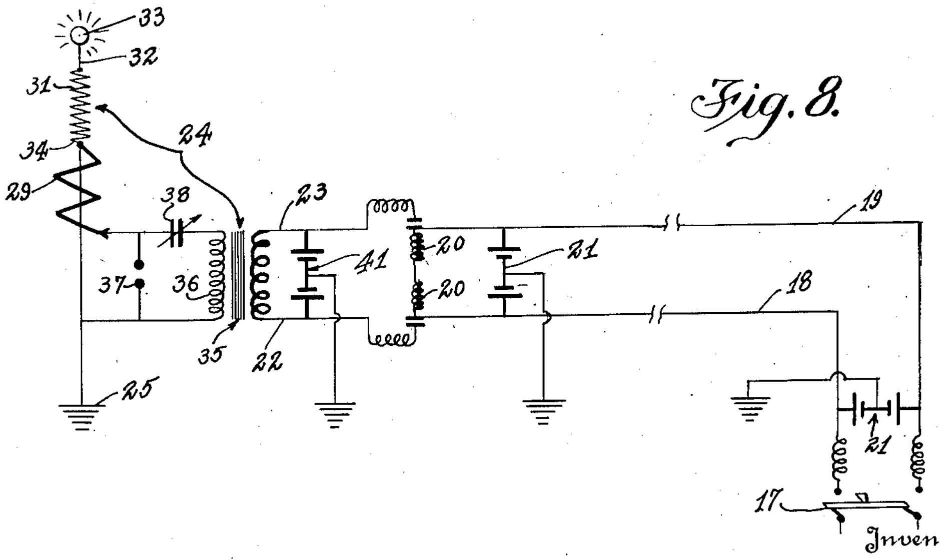
33
32
31
34
24
29
38
23
41
20
20
21
19
18
37
36
35
22
25
21
17

Fig. 8.

Inventor

Harold L. Bumbaugh

By Lyon & Lyon

Attorneys

Patented Aug. 2, 1932 **1,869,661**

UNITED STATES PATENT OFFICE

HAROLD L. BUMBAUGH, OF SAN FRANCISCO, CALIFORNIA, ASSIGNOR TO STANDARD OIL COMPANY OF CALIFORNIA, OF SAN FRANCISCO, CALIFORNIA, A CORPORATION OF DELAWARE

LIGHTNING PROTECTION SYSTEM AND METHOD

Application filed July 26, 1927. Serial No. 208,536.

This invention relates to lightning protection systems and, also, to lightning protection methods.

In general, an object of the invention is to protect particular areas or structures against being struck by lightning.

More particularly, an object of the invention is to provide a system and method whereby the point or points for receiving the lightning strokes are predetermined.

Another object is to make provision for taking care of the consequential secondary discharges.

Heretofore, it has been a practice for protecting certain areas and structures from lightning strokes, to provide towers, masts and network, singly or enclosing said areas or structures, thus providing a multiplicity of points in the expectation that one or more of said points would receive the lightning stroke in event of the stroke occurring in the vicinity of the area or structure thus protected, thus diverting the lightning stroke from said area or structure. Protection against lightning might possibly be afforded in this manner if it were practicable to erect towers in sufficient number and of sufficient height or to provide metallic enclosures of a type that would insure immunity for the enclosed area or structure against lightning strokes. This, however, is not, generally, feasible economically. Accordingly, another object of the present invention is to afford adequate protection in an economical and practicable manner.

In order to more clearly understand the invention, let it be assumed that there is a cloud sufficiently charged with electricity at a considerable height above the earth, imposing by its influence, an equal charge of electricity of opposite sign on the earth. The result will be a potential difference between the cloud and the earth, the presence of which is manifested in the form of an electrostatic atmospheric potential gradient. As a further result, there will be tendency toward equalization of this potential gradient through the medium of whatever conduction may exist across or through the dielectric or dielectrics separating the charged bodies which, in this instance, are the cloud and the earth.

As the cloud approaches the earth, or as the charge on the cloud increases, the potential gradient becomes steeper and the air intervening between the cloud and the earth becomes subjected to a greater electrostatic stress. The greater this stress the greater will be the tendency of the dielectric to break down and become to a greater or less degree conductive of electric current. In this case, conductivity is primarily due to ionization, probably the ionization of collision, when electrons are pulled from the atoms of the dielectric by the more or less intense strain exerted by the potential gradient. The theoretical explanation of what occurs implies a breaking up of the atoms of the dielectric into electrons and protons which, under the influence of the potential gradient, are drawn, respectively, toward the positively charged body and the negatively charged body.

Other factors being equal, the conduction across a stratum of air will be in direct ratio to the degree of ionization which, in turn, is largely dependent on the potential gradient. Therefore, when two electrically charged bodies of opposite sign exist and there is a tendency toward equalization of potential between them, under such conditions a reduction in the potential difference may take the form of a visible phenomena, when the so-termed ionization potential of the dielectric has been reached. This will be visibly indicated by the formation of corona and under slightly higher potentials, by the formation of streamers, these manifestations generally preceding a more or less complete breakdown of the dielectric, particularly under conditions of increasing or extremely high potential gradient. Under these conditions it is probable that the ionization of collision is supplemented by ionization due to increase in temperature, since the increased molecular activity causes more electrons to give up their energy in the form of heat which results from collision. The effects are undoubtedly cumulative when the potential gradient is sufficiently high.

The rate at which equalization of potential occurs between an electrically charged cloud and the earth depends on the steepness of the potential gradient and the character of the dielectric and, under a sufficiently steep potential gradient, every projecting point on the surface of the earth, such as blades of grass and leaves on the trees, in the vicinity, will have an electric discharge ranging from an infinitesimal amount to an amount that results in visible corona and streamers. If the cloud attains sufficient electric charge and has sufficient capacity, to make it a potential source of lightning strokes, some path of ionization or breakdown of the dielectric may occur between the cloud and the earth and along this highly conducting path, the current flow which constitutes a lightning stroke will take place. The earth end of this path of ionization will be at some point which, under the existing potential gradient, has a greater discharge, or ionizing influence, than any other nearby point. This is because the power of ionization is the result of the introduction into the space under stress of a number of free electrons (created through ionization) with a consequent tendency to render that particular portion of the space relatively conducting by moving under the influence of the potential gradient and tending to leave a more or less ionized path, and, therefore, one that is more or less conducting of electric current. It follows from this that the ground point first furnishing ionization or electrons in sufficient degree, will be the point from which the path of ionization will start and, if the path of ionization be then completed to the charged cloud, said point will be the one where lightning will strike the earth, all other conditions being the same.

In nature, there are so many variables which influence the behavior and occurrence of lightning strokes that it is manifestly impossible to predetermine and locate every danger point with a view toward avoiding the hazard at said point and in the area adjacent to said point. It is an object of the present invention to artificially create, at will, and at any desired point, an ionized area corresponding to the ionized area at the earth end or point of the lightning stroke so that, by locating this point removed from a particular structure or area, the same will be protected or rendered immune against lightning discharges, any discharge in that particular area being taken by the specially prepared point.

The potential of metallic objects or bodies on the earth is that of the earth and, under the influence of electrical disturbances, a considerable electric charge may be carried to a point considerably above the surface of the earth. As long as the potential gradient remains unchanged, this elevated charge will be in equilibrium and, accordingly, no hazard is present from such conditions; but, in the event of a sudden change in the potential gradient following a lightning discharge, there is a tremendous surge of current from the metallic objects or bodies to the earth. This surge is composed of the elevated charges, the potential of which, when the gradient is reduced, is above that of the earth. If an unbroken path, such as is formed by a metallic conductor of ample capacity, exists from the elevated charges to the earth, the charges, suddenly released, are carried harmlessly to the earth along said path and the electrical equilibrium is again restored. However, any break in the conducting path or lack of capacity of the metallic conductor to properly take care of the current flow will result in a bridging of the gap, resulting from the break in the conductor, by the electric current and the consequent formation of an arc across said gap, thus producing a serious fire hazard.

These electric charges, which take up their position under the stress of a potential gradient are known as bound charges and, when released, following lightning strokes, constitute secondary electric surges which are particularly hazardous in the presence of inflammable liquids and gases. The roof of a reservoir or tank used for the storage of petroleum is a typical example of hazard due to secondary electric discharges or surges. Accordingly, it is an object of this invention to provide a path of electric current conductive material of a capacity to handle the maximum surges to the earth and to make this path a part of the structure to be protected.

In the system I at present propose to employ, there is an electric device for producing corona, brush, streamer, or other high frequency discharges from the top of suitable towers, masts or other elevated structures grounded to the earth, to provide a path along which lightning discharges will flow, and included in the system is a grounded metallic network or grid having a capacity sufficient to relieve the area or structure being protected from any bound electrical charge released by a sudden decrease in the atmospheric potential gradient, such as may follow a stroke of lightning in the vicinity of said area or structure. This grounded metallic network is in all instances placed directly on the surface of and in contact with any roof, tank roof, reservoir roof or other structure which may under a high gradient be the seat of bound charges which upon release may form a secondary surge to ground.

The accompanying drawings illustrate a system constructed in accordance with the provisions of this invention and capable of operating to perform the method in accordance with the invention, the drawings being more or less diagrammatic.

Figure 1 is a view of a lightning protection system embodying the invention.

Figure 2 is an enlarged view, mainly in

section, from the line indicated by 2—2, Figure 1.

Figure 3 is an enlarged plan view of the left portion of Figure 1 showing the reservoir and the elements adjacent thereto.

Figure 4 is a plan showing in detail the base of one of the masts, together with the guy wires and grounding wires.

Figure 5 is a plan view of one of the masts, together with its guy wires and the anchors for said wires, the mast being shown in section.

Figure 6 is an enlarged view of the ionizing apparatus, the housing being shown in section.

Figure 7 is a wiring diagram of the wiring of the ionizing apparatus.

Figure 8 is a diagram of electrical connections.

Referring to the drawings, the area or structure to be protected against lightning strokes may be, for example, a tank farm of the character used for the storage of petroleum and its products. In Figure 3, two tanks or reservoirs of such tank farm are indicated at 11 and on each of them is a metallic network or grid 12 for conducting to ground the bound charges that may exist on the reservoir roof. As indicated in Figure 2 the grids 12 may be metallically bonded to the reservoir roofs 13 as indicated at 14 so that, under the influence of an unbalanced potential, any secondary electrical discharges will be carried to the earth, without arcing, by grounds 15 from the grids. Said grounds may be pipes. The grids are connected to the pipes by wires 16.

Now referring, more particularly, to Figure 8, there is provided a manually controlled switch 17 connecting with a source of commercial or other electric current, preferably at a location that is remote from the area and structure that are to be protected. The manually operated switch 17 is connected by wires 18, 19 to remote control switches 20. 21 indicates a line protector for grounding transients and lightning surges that may occur in the wires 18, 19. The remote control switches 20 are connected by wires 22, 23 to a high frequency, high voltage generator or resonance coil 24 which, when in operation, produces the electrical discharge for ionization. The generator 24 is grounded at 25 to the earth so that lightning discharges will be conducted to the earth.

In Figure 8 but one resonance coil or generator 24 is indicated, but, preferably, the upper end of each mast 40 will be provided with such coil or generator and this comprises a portion of the ionization apparatus which is illustrated, more or less diagrammatically in toto in Figure 6. Referring more particularly to said Figure 6 there is provided a base 26 mounted in a housing 27. The base 26 is of suitable electrical insulating material such as hard rubber, bakelite or the like. Mounted on the base 26 is a support 28 for primary winding 29. Also mounted on the base 26 is a cylinder 30 of suitable insulating material, such as hard rubber, bakelite or the like, and on said cylinder is a secondary winding 31. Connected to one end of the secondary winding 31 and projecting through the top of the housing 27 into the atmosphere is a vertical metal rod 32. The rod 32 is surmounted by a metal ball 33. The other end of the secondary winding 31 is connected to a point on the primary winding 29 and the secondary winding 31 is connected by a wire 34 to the ground 25 hereinbefore mentioned. Mounted in the housing 27 is a line-voltage transformer 35. The secondary winding 36 of the transmitter 35 connects with a spark gap 37 and also connects with the ground 25 and with a variable condenser 38 which, in turn, connects with the primary winding 29.

The housing 27 is constructed of suitable weather proof material and covers all elements of the ionizing apparatus excepting the ball 33 and a portion of the rod 32 and the housing may be connected to the mast, with which it is associated, by a pipe flange 39 which is secured to the bottom of the housing and which is threaded on to the upper end of the mast.

The above described system operates to perform the various operations of the method, as follows: Assuming that there are indications of an approaching storm or that a storm may eventually occur in the area or near the structure to be protected, the manually operated switch 17 will be closed, thus energizing the lines 18, 19 so as to operate the remote control switches 20. This connects the ionization apparatus with the source of power and renders the system operative for producing the desired effect on lightning discharges. The high frequency, high voltage generator operates to produce sufficient free electrons adjacent the ball 33 to insure that, if a path of ionization is even completed from the earth to a cloud in that area, the earth end of that path will be in the ionized zone surrounding the ball 33.

It is to be understood that the remote control switches may be dispensed with and the system put into and out of operative condition by manually operated switch or switches at the base of the mast or other supporting structure.

It is to be understood that the system above described may be modified in any suitable manner to suit local conditions.

It will appear from the foregoing that the methhod of protecting an area or structure against lightning involves the operation of producing a definite zone of relatively high ionization at a point or points removed from the area that is to be protected. The method also includes the operation of conducting to the earth from any elevated objects that may be in the protected area, the secondary electrical surge resulting from a lightning discharge in the vicinity.

It is to be particularly noted that in this system the metallic grid or network lies directly on the structure that is to be protected with no dielectric separating the grid and structure. Consequently, when the grid drains the charges to ground, following a lightning discharge in the vicinity, it drains from the structure to ground charges on the structure and, therefore, no fire hazard can exist there.

I claim:

1. The combination with a reservoir for inflammable substances, of a metallic grid in contact with and extending over said reservoir and electrically connected with the earth, an elevated structure at one side of the reservoir, and a means mounted on the elevated structure operable to artificially produce a higher ionized zone above the level of the grid than is producible in said zone by the natural charged condition of the earth, the last mentioned means being electrically connected with the earth.

2. The combination with a reservoir for inflammable substances, of a metallic grid resting on the roof of said reservoir in contact with and bonded thereto, means for electrically connecting the metallic grid with the earth, and means for artificially producing a measurably higher ionized zone at a point removed from the grid than is producible in said zone by the natural charged condition of the earth, the last mentioned means being electrically connected with the earth.

Signed at Ventura, Calif., this 5th day of July, 1927.

HAROLD L. BUMBAUGH.

rexresearch.com

Oleg JEFIMENKO

Electrostatic Motors

---

1 | **Tesla Generator - Build Your Own Generator**
Eliminate Your Power Bill Easy. Do It Yourself. Great Discount Now! Just $49. tesla-energy.org/Generator

2 | **Start Download - View PDF**
Convert From Doc to PDF, PDF to Doc Simply With The Free Online App! download.fromdoctopdf.com

1 | **Tesla Generator - Build Your Own Generator**
Eliminate Your Power Bill Easy. Do It Yourself. Great Discount Now! Just $49. tesla-energy.org/Generator

2 | **Start Download - View PDF**
Convert From Doc to PDF, PDF to Doc Simply With The Free Online App! download.fromdoctopdf.com

VIEW PDF
DOCtoPDF
Start Download
Begin Download - Free PDF Converter. Download & Start Now!

» JEFIMENKO, Oleg : Electrostatic Motors ~ PDF copy of the out-of-print classic. Run motors with atmospheric electricity.

Popular Science **(April 1971)**

The Amazing Motor That Draws Power from the Air

by

C.P. Gilmore & William J. Hawkins

Would you believe an electric motor made almost entirely of plastic? That can run on power transmitted through open air? And sneak free electricity right out of the earth's electrical field?

At the University of West Virgina we saw a laboratory full of such exotic devices spinning, humming, and buzzing away like a swarm of bees. They are electrostatic motors, run by charges similar to those that make your hair stand on end when you comb it on a cold winter's day.

Today, we use electromagnetic motors almost exclusively. but electrostatics have a lot of overlooked advantages. They're far lighter per horsepower than electromagnetics, can run at extremely high speeds, and are incredibly simple and foolproof in construction.

"And, in principle," maintains Dr Oleg Jefimenko, "they can do anything electromagnetic motors can do, and some things they can do better."

Jewel-like Plastic Motors.

Jefimenko puts on an impressive demonstration. He showed us motors that run on the voltage developed when you hold them in your hands and scuff across a carpet, and other heavier, more powerful ones that could do real work. Up on the roof of the University's physics building in a blowing snowstorm, he connected an electrostatic motor to a specially designed earth-field antenna. It twirled merrily from electric power drawn out of thin air.

These remarkable machines are almost unknown today. Yet the world's electric motor was an electrostatic. It was invented in 1748 by Benjamin Franklin.

Franklin's motor took advantage of the fact that like charges repel, unlike ones attract. He rigged a wagon-wheel-sized, horizontally mounted device with 30 glass spokes. On the end of each spike was a brass thimble. Two oppositely charged leyden jars -- high voltage capacitors -- were so placed that the thimbles on the rotating spokes barely missed the knobs on the jars (see photo).

As a thimble passed close to a jar, a spark leaped from knob to thimble. That deposited a like charge on the thimble, so they repelled each other. then, as the thimble approached the oppositely charged jar, it was attracted. As it passed this second jar, a spark jumped again, depositing a new charge, and the whole repulsion-attraction cycle began again.

In 1870, the German physicist J.C. Poggendorff built a motor so simple it's hard to see what makes it work. The entire motor, as pictured here, is a plastic disk (Poggendorff used glass) and two electrodes. The electrodes set up what physicists call a corona discharge; their sharp edges ionize air molecules that come in contact with them. These charged particles floating through the air charge the surface of the palstic disk nearby. Then the attraction-repulsion routine that Franklin used takes place.

A few papers on electrostatic motors have trickled out of the laboratories in recent years. But nobody really showed much interest until Dr Jefimenko came on the scene.

The Russian-born physicist was attending a class at the University of Gottingen one day shortly after World War II when the lecturer, a Prof. R.W. Pohl, displayed two yard-square metal plates mounted on the end of a pole. He stuck the device outside and flipped it 180 degrees. A galvanometer hooked to the plates jumped sharply.

"I could never forget that demonstration," said Jefimenko. "And I wondered why, if there is electricity in the air, you couldn't use it light a bulb or something."

Electricity Everywhere

The earth's electrical field has been known for centruries. Lightning and St Elmo's fire are the most dramatic manifestations of atmospheric electricity. But the field doesn't exist just in the vicinity of these events; it's everywhere.

The earth is an electrical conductor. So is the ionosphere, the layer of ionized gas about 70 kilometers over our heads. The air between is a rather poor insulator. Some mechanisms not yet explained constantly pumps large quantitites of charged particles into the air. The charged particles cause the electrical field that Jefimenko saw demonstrated. Although it varies widely, strength of the field averages 120 volts per meter.

You can measure this voltage with an earth-field antenna -- a wire with a sharp point at the top to start a corona, or with a bit of radioactive materials that ionizes the air in its immediate vicinity. near the earth, voltage is proportional to altitude; on an average day you might measure 1200 volts with a 10-meter antennas.

Over that past few years, aided by graduate-student Henry Fischbach-Nazario, Jefimenko designed advanced corona motors. With David K. Walker, he experimented with electret motors. An electret is an insulator with a permanent electrostatic charge. It produces a permanent electrostatic charge in the surrounding space, just as a magnet produces a permanent magnetic field. And like a magnet, it can be used to build a motor.

Jefimenko chose the electrostatic motor for his project because the earth-field antennas develop extremely high-voltage low-current power -- and unlike the electromagnetic motor -- that's exactly what it needs.

The Climactic Experiment

On the night of Sept. 29, 1970, Jefimenko and Walker strolled into an empty parking lot, and hiked a 24-foot pole painted day-glow orange into the sky. On the pole's end was a bit of radioactive material in a capsule connected to a wire. The experimenters hooked an electret motor to the antenna, and, as Jefimenko describes it, "the energy of the earth's electrical field was converted into continuous mechanical motion."

Two months later, they successfully operated operated a corona motor from electricity in the air.

Any Future In It?

Whether the earth's electrical field will ever be an important source of power is open to question. There are millions -- perhaps billions -- of kilowatts of electrical energy flowing into the earth constantly. Jefimenko thinks that earth-field antennas could be built to extract viable amounts of it.

But whether or not we tap this energy source, the electrostatic motor could become important on its own.

* In space or aviation, it's extreme light weight could be crucial. Jefimenko estimates that corona motors could deliver one horsepower for each 3 pounds of weight.

* They'd be valuable in laboratories where even the weakest magnetic field could upset an experiment.

* Suspended on air bearings, they'd make good gyroscopes.

In a particularly spectacular experiment, Jefimenko turned on a Van de Graaff generator -- a device that creates a very-high-voltage field. About a yard away he placed a sharp-pointed corona antenna and connected it to an electrostatic motor. The rotor began to spin. The current was flowing from the generator through the air to where it was being picked up by the antenna.

The stunt had a serious purpose: The earth's field is greatest on mountaintops. Jefimenko would like to set up a large antenna in such a spot, then aim an ultraviolet laser beam at a receiving site miles away at ground level. The laser beam would ionize the air, creating an invisible conductor through apparently empty space.

To be sure, many difficulties exist; and no one knows for sure whether we'll ever get useful amounts of power out of the air. But with thinking like that, Jefimenko's a hard man to ignore.

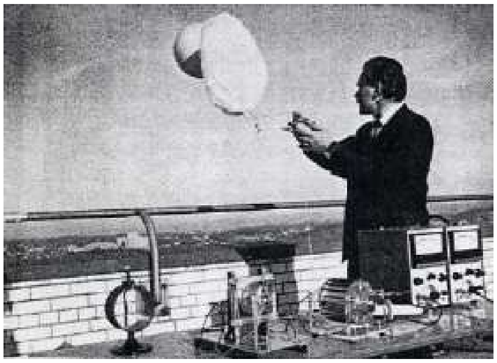

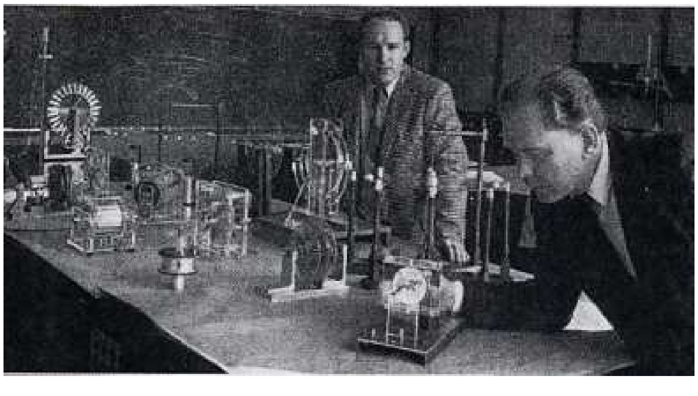

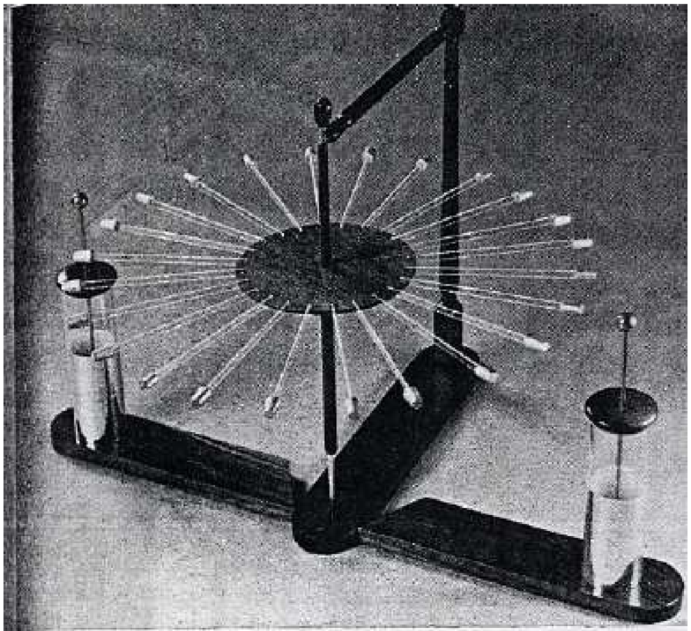

World's first electrostatic motor looked like this, according to Benjamin Franklin's description. Franklin said he'd use it to turn a roasting turkey, but there's no evidence that he did.

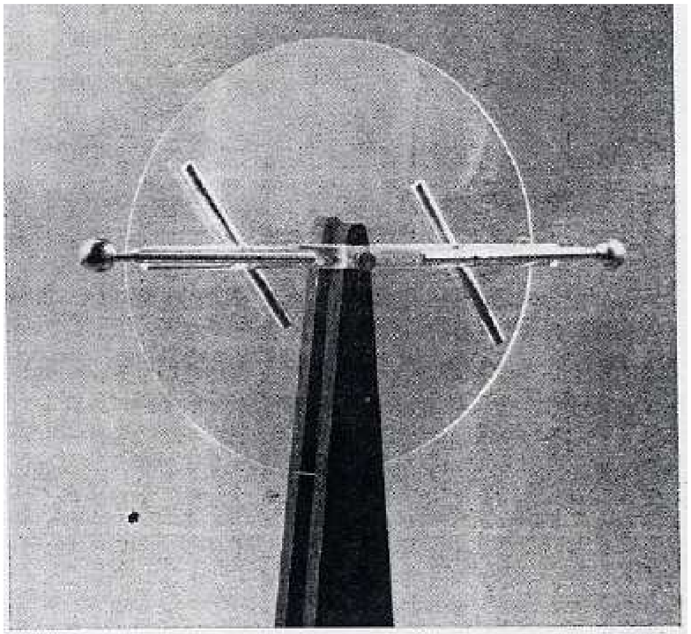

Simplest electrostatic motor ever designed was built by physicist J. C. Poggendorff. Two slanted electrodes spray charges onto disk, alternately repel and attract it to make it revolve.

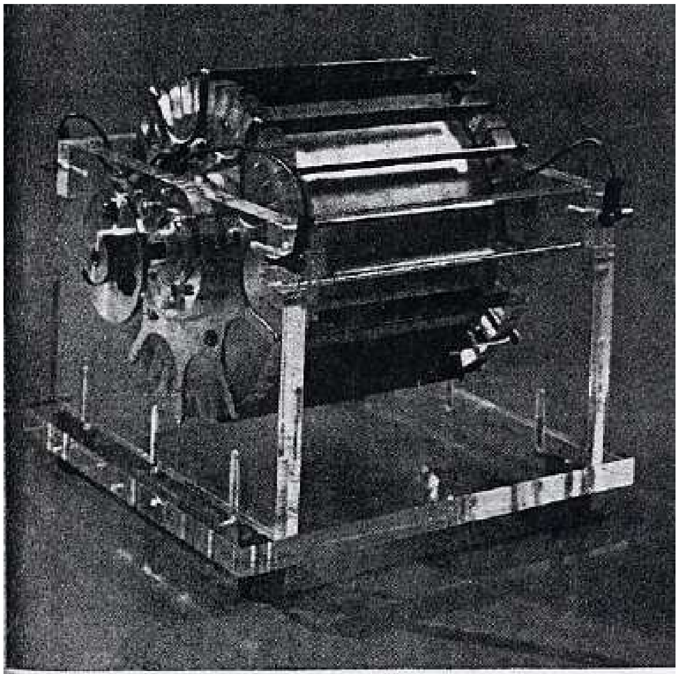

Most efficient, most powerful electrostatic shown to date was designed by Dr. Oleg Jefimenko. It has multiple electrodes, cylindrical rotor, and develops a maximum of 70 watts of power.

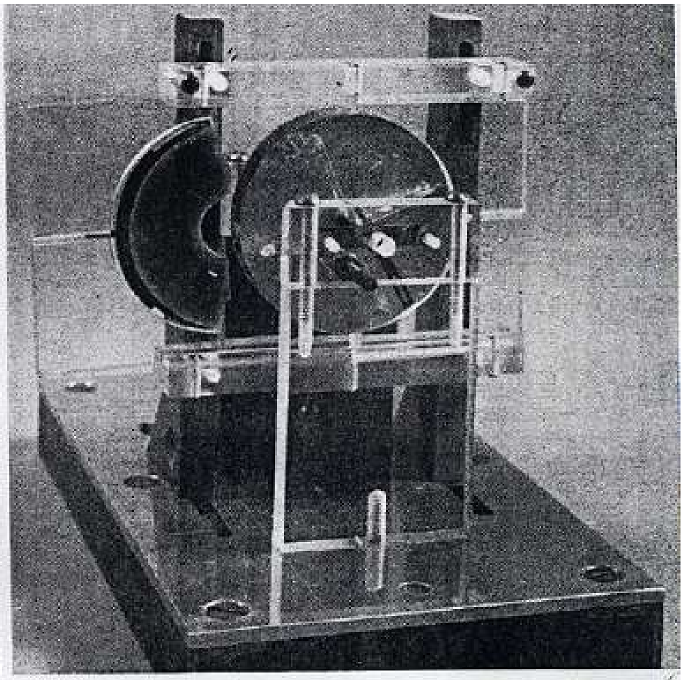

Electret motor above turns on as little as 60 volts. It was the first to run on power from the earth's electrical field. Such motors can achieve extremely high speeds, and use little power

---

***Popular Science* (May 1971)**

Electrostatic Motors You Can Build

by

C.P. Gilmore & William J. Hawkins

When we crank up the electrostatic motor at the top of this page, people always want to know what makes it run. It is mysterious -- there's nothing but a plastic disk and two strange electrodes. Yet there it is, spinning merrily.

In "The Amazing Motor That Draws Power From the Air", last month, we told about our visit to the laboratory of Dr Oleg Jefimenko at the University of West Virginia, who has designed and built a variety of these ingenious machines. now, as promised, we bring you details on how you can build your own electrostatic motor from simple materials.

The devices that you see here are corona-discharge motors. The sharp-pointed or knife-edge electrodes create a corona, which ionizes or charges the air particles floating by. These charged particles transfer their charge to the closest part of the plastic rotor and charge it up, just as you can charge your body by walking across a wool rug on a dry winter's day.

Once a spot on the rotor assumes a charge, it is repelled from the chargin electrode by electrostatic forces, and at the same time is attracted to the other electrode, which has an opposite charge. When the charged section of the rotor reaches the opposite electrode, another corona discharge reverses the polarity and starts the whole thing over again.

The Concept is Simple

And so are the motors. But that doesn't mean thery're easy to build. These motors run on millionths of a watt; they've got no power to waste turning stiff bearings or slightly misaligned rotors. So they must be built with watch-making precision.

They're made of acrylic sheet, rod, and tube stock -- Plexiglas and Lucite are two of the better-known brands. Acrylic cuts and works beautifully. Cut edges can be sanded so they have a white, frosted appearance that, in contrast with clear surfaces, gives your finished motor a sparkling, jewel-like appearance. If you like clear edges, you can buff them on a wheel and the whole thing becomes transparent.

Drill and tap the acrylic and assemble parts with machine screws. This allows for fine adjustment and alignment. Later, you can make the whole thing permanent by putting a little solvent along the joints. The solvent flows into the joint and fuses it permanently.

Details of framework, support and so on aren't important; change them if you like. but work with care if you want to avoid headaches. The Poggendorff motor looked simple; we slapped it together in a couple of hours, hooked up the power source -- and nothing happened. We gave it a few helpful spins by hand, but it wouldn't keep running.

The cure took about 3 hours. First, we noticed that the outer edge of the disk wobbled from side to side about 1/16 of an inch as the wheel revolved. So the rotor-electrode distance was constantly changing. There was a little play in the 1/4" hole we had drilled for the electrodes -- so they weren't lined up absolutely square with the disk. Then we noticed that the disk always stopped with one side down. The imbalance was only a fraction of an ounce -- but it was too much.

We drilled out the old hub and cemented in a new one -- this time, carefully. We lined up the electrodes -- precisely. Then, once more spinning the disk by hand, we added bit of masking tape until it was perfectly balanced. We connected the power -- and slowly... slowly... the disk began to turn. After about a minute, we clocked some turns with a watch and found it was spinning at 200 rpm. A moment later, we lost count. It was a great feeling.

Where Tolerances Are Brutal

We had even more trouble with the octagonal-window machine. When it wouldn't run and we turned the shaft by hand, we could feel the rotor dragging. We took it apart, felt all the surfaces on the rotor and the framework's insides and found a few bits of hardened cement, which we removed. We filed down all edges on the rotor adn the windows to make sure there were no beads or chips dragging.

The rotor and corner separators are made from the same sheet of 1/2" plastic, so rotor clearance is achieved by putting shims at the corners to hold the side plates slightly more than 1/2" apart. With the 1/16" shims we were using, we could see that the sides were slightly misaligned so the shaft was not being held at a true 90 degrees. We drilled slightly oversized holes in the corners of one side piece and carefully adjusted until the rotor was turning true in the slot. To give the motor more torque, we put a bead of cement along the outer edge of each aluminum-foil electrode to stop corona leakage. The motor ran.

Take A Giant Step

Once you've built these machines, why not design your own? Start with the Jefimenko 1/10 hp model (pictured) as a challenge. Then plan one from scratch. You can power your motors with a laboratory high-voltage supply, a Van de Graff generator, or a Wimhurst machine or any other high-voltage source. We've been running ours on the home-built Wimhurst machine shown in the photos. (If you don't want to build one, Wimhurst machiens are available from scientific supply houses such as Edmund Scientific).

The discharge globes are traditional for high-voltage machines. They aren't necessary, but they give a quick check on machine operation and a satifying arc when you move them within 1/2" of each other. Incidentally, that funny smell is ozone. But its concentration is too low to be harmful. The generator is safe, too. You can hold both electrodes in your hands and all you'll feel is a tingle. This particular generator, we estimate, puts out about 30,000 volts.

To make wiring simple, we used standard connectors on the Wimhurst collectors, and meter leads with regular banana plugs and alligator clips to hook up the motors.

Last month, we mentioned seeing Dr Jefimenko run his electrostatic motors on electricity tapped from the earth's field. We haven't had a chance to try this yet with ours, but it should work. If you want to try, you'll need a needle-pointed piece of music wire a few inches long to start a corona, plus several hundred feet of fine copper wire.

Connect the pointed wire to the fine conductor, get the sharp point up into the air at least 200-300 feet with a kite or balloon, and hook the wire to one side of the motor. Hook the other side of the motor to ground. The earth field antenna should at times be able to develop up to 20,000 volts from the earth's electrical field. If nothing happens, check your equipment, or try another day. The field changes constantly.

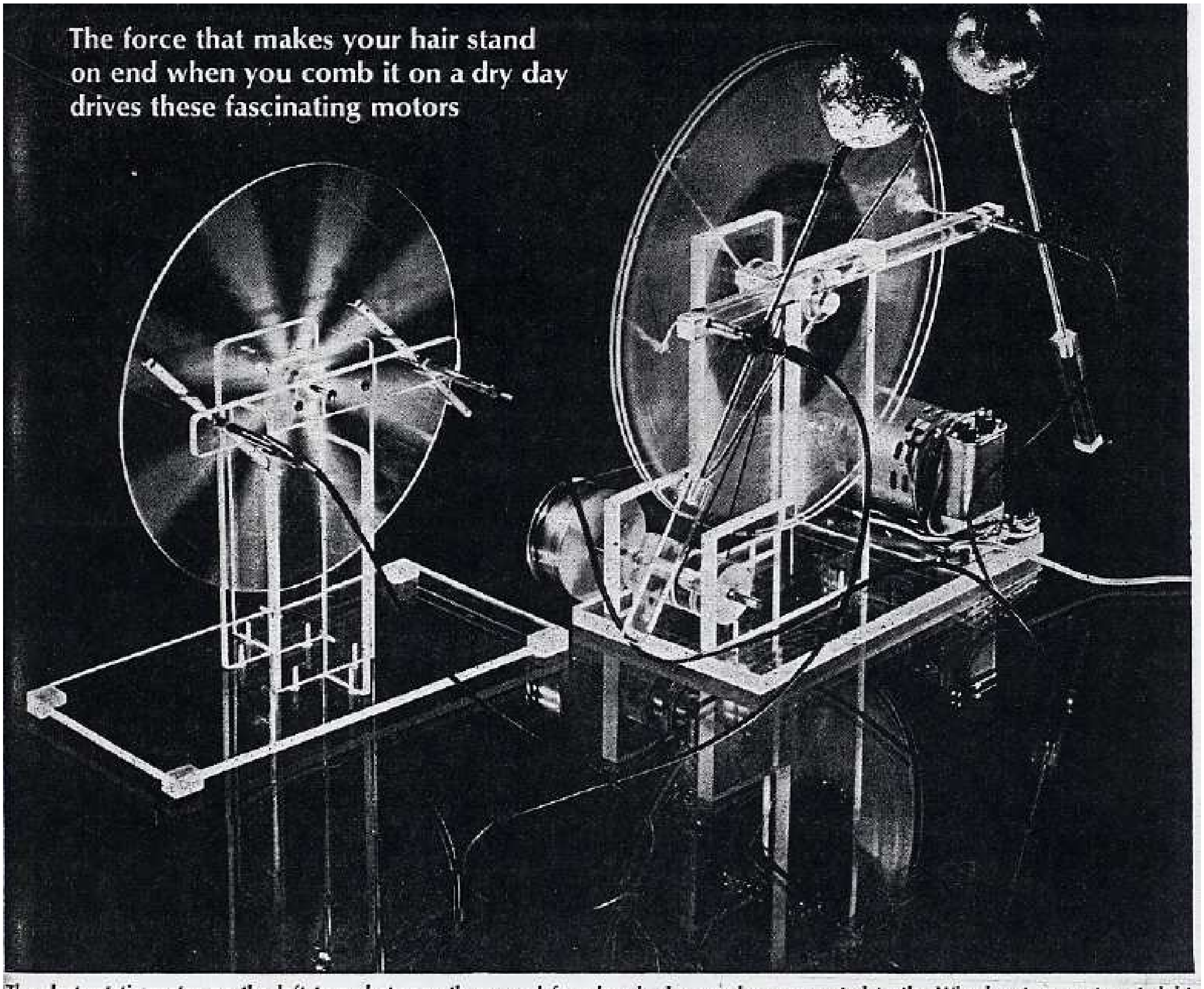
The force that makes your hair stand
on end when you comb it on a dry day
drives these fascinating motors

The electrostatic motor on the left turns between three- and four-hundred rpm when connected to the Wimshurst generator at right.

This octagonal-window design makes motor handsome, self-starting

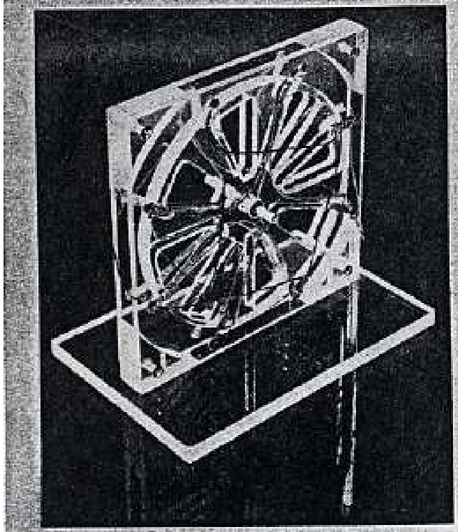

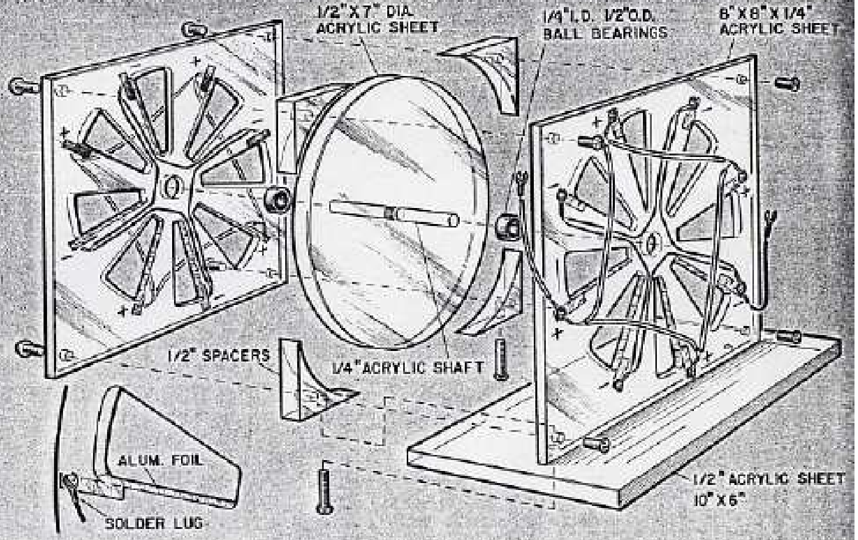
1/2"X7" DIA. ACRYLIC SHEET
1/4"I.D. 1/2"O.D. BALL BEARINGS
8"X8"X1/4" ACRYLIC SHEET
1/2" SPACERS
1/4"ACRYLIC SHAFT
ALUM. FOIL
SOLDER LUG
1/2" ACRYLIC SHEET 10"X6"

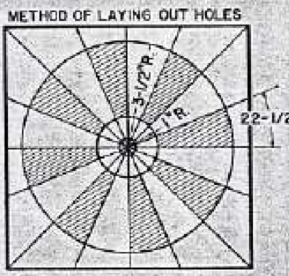
METHOD OF LAYING OUT HOLES
3-1/2"R.
1"R.
22-1/2°

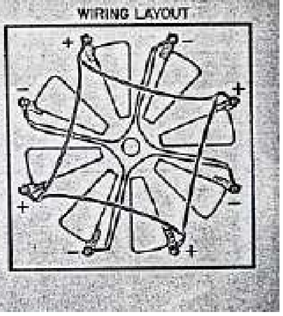
WIRING LAYOUT
+
−
+
−
+
−
+

Rotor of the Jefimenko-designed motor shown here is a ½″ plastic disk with a ¼″ shaft cemented carefully in place. Clamp the two outer sheets together before you drill the bearing holes and cut the windows. Aluminum foil electrodes can be cut to approximate shape, glued in place, and trimmed with razor blade. Keep + and − wiring separated by ½″ to avoid arcing. Note that electrodes oppose each other directly; opposing electrodes are of opposite polarity. Cement corner separators to one side for stability, attach other side with screws. To assemble, shim sides apart. Shim thickness of 1/16″ or less is best if rotor can be made to turn freely within this narrow gap. Avoid sharp corners, points, or edges in electrical circuits. They develop corona and leak power into air.

Wimshurst machine generates 30,000 volts

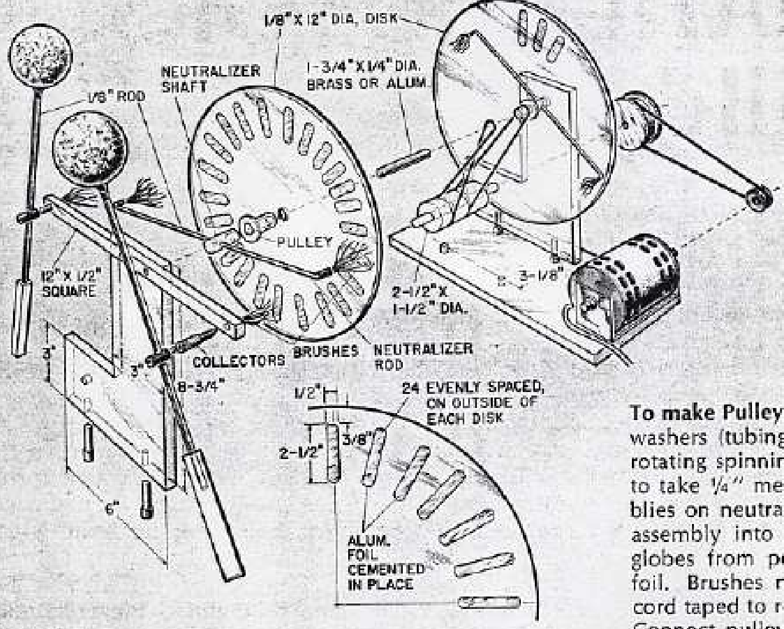
1/8"X12" DIA. DISK
NEUTRALIZER SHAFT
1-3/4"X1/4"DIA. BRASS OR ALUM.
1/8" ROD
PULLEY
12"X1/2" SQUARE
2-1/2"X 1-1/2" DIA.
3-1/8"
3"
3"
8-3/4"
COLLECTORS
BRUSHES
NEUTRALIZER ROD
6"
1/2"
2-1/2"
3/8"
24 EVENLY SPACED, ON OUTSIDE OF EACH DISK
ALUM. FOIL CEMENTED IN PLACE

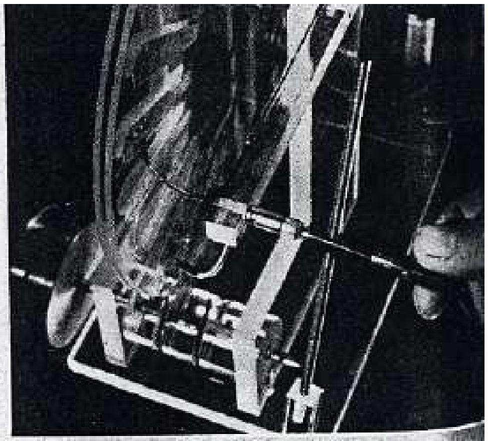

To make Pulleys, drill rotors, cement in ½″ i.d. tubing plus plastic washers (tubing protrudes 1/16″ on inner surface to hold counter-rotating spinning disks ⅛″ apart). Drill ½″ o.d. neutralizer shafts to take ¼″ metal shaft; assemble by slipping rotor-pulley assemblies on neutralizer shafts, neutralizer shafts on metal shaft. Slip assembly into ¼″-deep holes drilled in supports. Make spark globes from polystyrene Christmas balls wrapped in aluminum foil. Brushes must be flexible; use aluminum-braided gift-wrap cord taped to rods. Use any drive that turns disks about 1,200 rpm. Connect pulleys with tape-recorder drive belts.

Best first project: motor designed by physicist J. C. Poggendorff in 1870

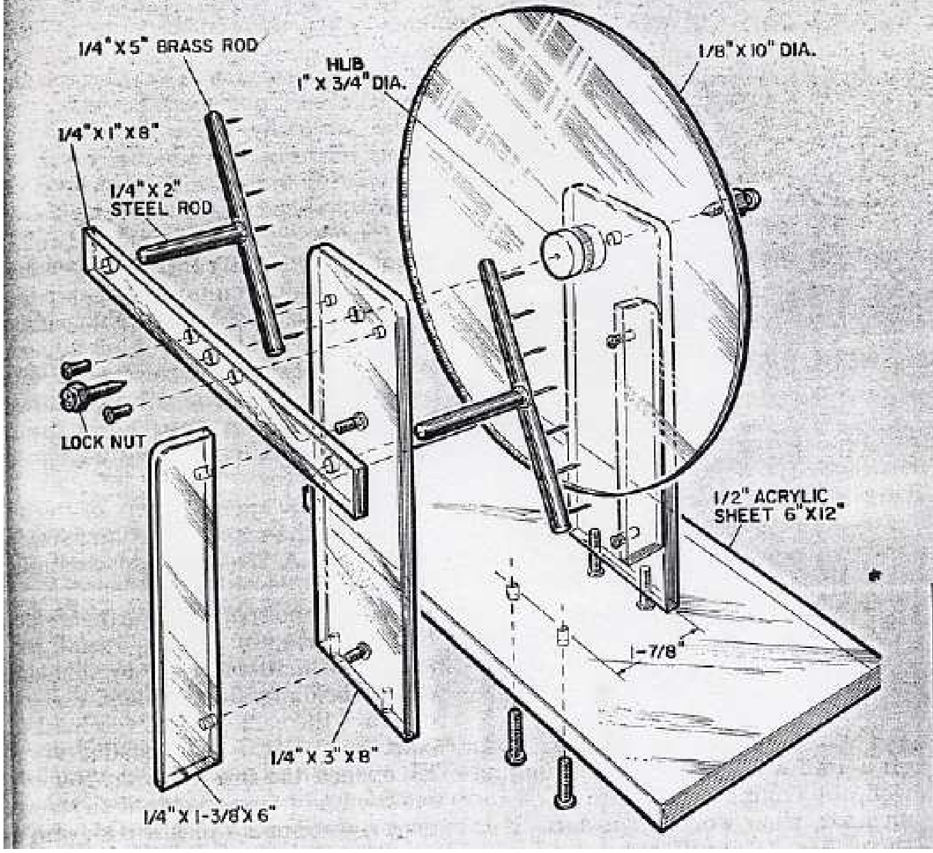

To build Poggendorff motor, drill center of the rotor disk accurately, cement ¾" hub in place, use centerpunch to make bearing indentations in the ends of the hub. Center gap between needle bearings must be adjustable; easiest way is to make bearings from machine screws sharpened on grindstone. Use locknuts to hold fine adjustment. For electrodes, drill seven equidistant holes in brass, steel, or aluminum rods; sharpen brads to a needle point, press-fit them into the holes, grind off heads flush. Brads should protrude about ½". File flat on the back of electrodes, solder or drill, tap, and thread support at right angles. Get extra torque if desired by adding an extra set of electrodes on the opposite side of rotor.

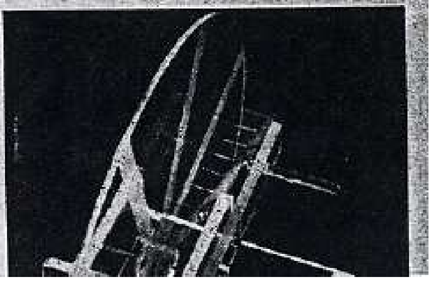

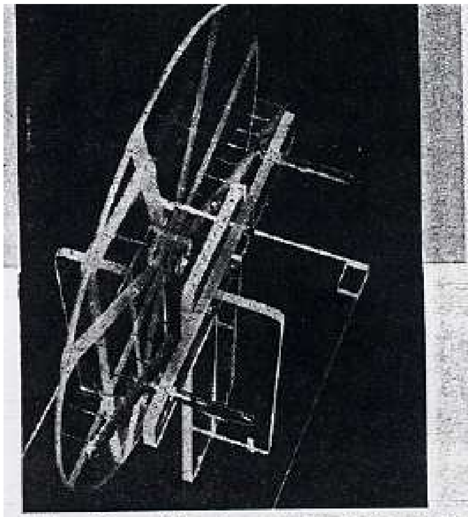

Mounting, with rotor removed, shows the relative positions of supports, needle bearings, and electrodes. Electrode points should be close to disk, but not touching.

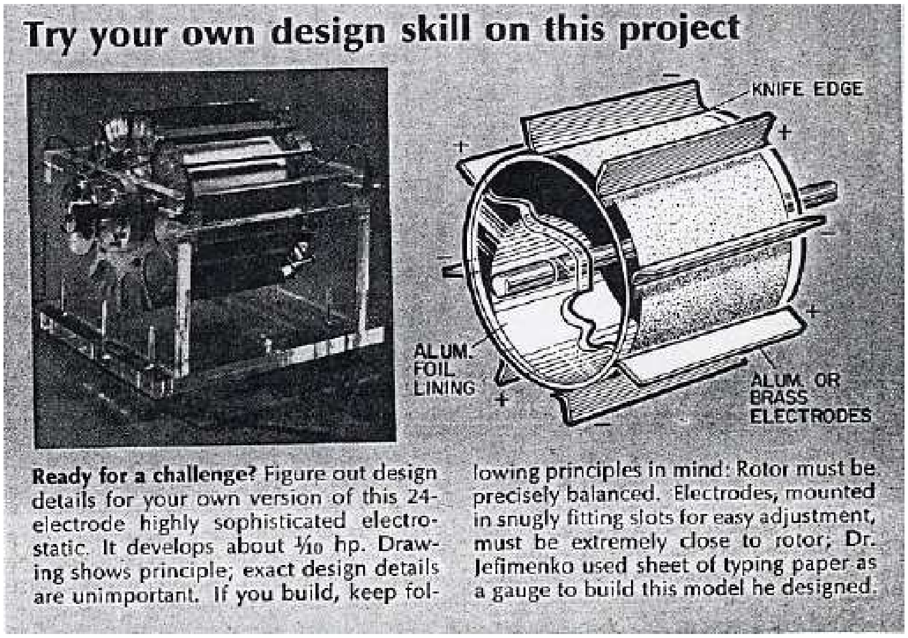

Try your own design skill on this project

KNIFE EDGE
ALUM. FOIL LINING
ALUM. OR BRASS ELECTRODES
+
−

Ready for a challenge? Figure out design details for your own version of this 24-electrode highly sophisticated electrostatic. It develops about 1/10 hp. Drawing shows principle; exact design details are unimportant. If you build, keep following principles in mind: Rotor must be precisely balanced. Electrodes, mounted in snugly fitting slots for easy adjustment, must be extremely close to rotor; Dr. Jefimenko used sheet of typing paper as a gauge to build this model he designed.

---

http://en.wikipedia.org/wiki/Oleg_D._Jefimenko

Oleg D. Jefimenko

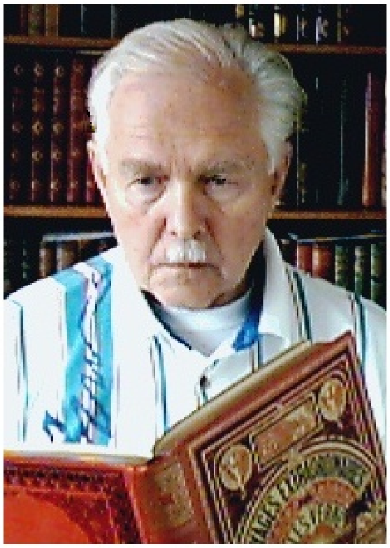

(October 14, 1922, Kharkiv, Ukraine - May 14, 2009, Morgantown, West Virginia, USA) - physicist and Professor Emeritus at West Virginia University.

Biography

Jefimenko received his B.A. at Lewis and Clark College (1952). He received his M. A. at the University of Oregon (1954). He received his Ph.D. at the University of Oregon (1956). Jefimenko has worked for the development of the theory of electromagnetic retardation and relativity. In 1956, he was awarded the Sigma Xi Prize. In 1971 and 1973, he won awards in the AAPT Apparatus Competition. Jefimenko has constructed and operated electrostatic generators run by atmospheric electricity.

Jefimenko has worked on the generalization of Newton's gravitational theory to time-dependent systems. In his opinion, there is no objective reason for abandoning Newton's force-field gravitational theory (in favor of a metric gravitational theory). He is actively trying to develop and expand Newton's theory, making it compatible with the principle of causality and making it applicable to time-dependent gravitational interactions.

Jefimenko's expansion, or generalization, is based on the existence of the second gravitational force field, the "cogravitational, or Heaviside's, field". This is might also be called a gravimagnetic field. It represents a physical approach profoundly different from the time-space geometry approach of the Einstein general theory of relativity. Oliver Heaviside first predicted this field in the article "A Gravitational and Electromagnetic Analogy" (1893).

Selected publications

Books

* "Electrostatic motors; their history, types, and principles of operation". Star City [W. Va.], Electret Scientific Co. [1973]. LCCN 73180890
* "Electromagnetic Retardation and Theory of Relativity: New Chapters in the Classical Theory of Fields", 2nd ed., Electret Scientific, Star City, 2004.
* "Causality, Electromagnetic Induction, and Gravitation: A Different Approach to the Theory of Electromagnetic and Gravitational Fields", 2nd ed., Electret Scientific, Star City, 2000.
* "Electricity and Magnetism: An Introduction to the Theory of Electric and Magnetic Fields", 2nd ed., Electret Scientific, Star City, 1989.
* "Scientific Graphics with Lotus 1-2-3: Curve Plotting, 3D Graphics, and Pictorial Compositions". Electret Scientific, Star City, 1987.

Book chapters

* "What is the Physical Nature of Electric and Magnetic Forces?" in Has the Last Word Been Said on Classical Electrodynamics? -- New Horizons, A. E. Chubykalo, Ed., (Rinton Press, Paramus, 2004).
* "Does special relativity prohibit superluminal velocities?" in Instantaneous Action at a Distance in Modern Physics: "Pro" and "Contra", A. E. Chubykalo, Ed., (Nova Science, New York, 1999).

Papers

* "A neglected topic in relativistic electrodynamics: transformation of electromagnetic integrals". arxiv.org, 2005.
* "Presenting electromagnetic theory in accordance with the principle of causality", Eur. J. Phys. 25 287-296, 2004. doi:10.1088/0143-0807/25/2/015
* "Causality, the Coulomb field, and Newton's law of gravitation" (Comment), American Journal of Physics, Volume 70, Issue 9, p. 964, September 2002.
* "The Trouton-Noble paradox," J. Phys. A. 32, 3755–3762, 1999.
* "On Maxwell's displacement current," Eur. J. Phys. 19, 469-470, 1998.
* "Correct use of Lorentz-Einstein transformation equations for electromagnetic fields", European Journal of Physics 18, 444-447, 1997.
* "Retardation and relativity: Derivation of Lorentz-Einstein transformations from retarded integrals for electric and magnetic fields", American Journal of Physics 63 (3), 267-72.
* "Retardation and relativity: The ease of a moving line charge", American Journal of Physics, 63 (5), 454-9.
* "Direct calculation of the electric and magnetic fields of an electric point charge movingwith constant velocity," Am.J.Phys. 62, 79-84, 1994.
* "Solutions of Maxwell's equations for electric and magnetic fields in arbitrary media," Am. J. Phys. 60, 899-902 1992.
* "Electrets," (with D. K. Walker) Phys. Teach. 18, 651-659, 1980.
* "How can An Electroscope be Charged This Way?", TPT 56, 1979.
* "Water Stream 'Loop-the-Loop'", AJP 42, 103-105, 1974.
* "Franklin electric motor," Am. J. Phys. 39, 1139-1141, 1971.
* "Operation of electric motors from atmospheric electric field," Am. J. Phys. 39, 776-779, 1971.
* "Demonstration of the electric fields of current-carrying conductors," Am. J. Phys. 30, 19-21, 1962.
* "Effect of the earth's magnetic field on the motion of an artificial satellite," Am. J. Phys. 27, 344-348, 1959.

Encyclopedia Article

* "'Maxwell's Equations'", Macmillan Encyclopedia of Physics, Macmillan, New York, 1996.

---

http://www.integrityresearchinstitute.org/FutureEnergy/FutureEnergyTech.html

The next energy breakthrough is Dr. Oleg Jefimenko's electrostatic motors. Discovered by Ben Franklin in the 18th century, electrostatic motors are an all-American invention. They are based on the physics of the fair-weather atmosphere that has an abundance of positive electric charges up to an altitude of 20 km. However, the greatest concentration is near the ground and diminishes with altitude rapidly. Dr. Jefimenko discovered that when sharp-pointed antennas are designed for a sufficient length to obtain at least 6000 volts of threshold energy, the fair-weather current density available is about a picoampere per square meter. Such antennas produce about a microampere of current. However, small radioactive source antennas may be used instead that have no threshold voltage and therefore no height requirements. Similar to a nuclear battery design of Dr. Brown, these antennas have larger current potentials depending upon the radioactive source used (alpha or beta source) and ionize the air in the vicinity of the antenna. Electrostatic motors are lighter than electromagnetic motors for the same output power since the motor occupies the entire volume. For example, it is expected that a motor one meter on a side will provide a power of one megawatt and weigh 500 kg or less. Electrostatic motors also require very little metal in their construction and can use mostly plastic for example. They can also operate from a variety of sources and range of voltages. As Dr. Jefimenko points out, "It is clear that electrostatic motor research still constitutes an essentially unexplored area of physics and engineering, and that electrostatic motor research must be considered a potentially highly rewarding area among the many energy-related research endeavors."[5] The atmospheric potential of the planet is not less than 200,000 megawatts. He has succeeded in constructing demonstration motors that run continuously off atmospheric electricity. Jefimenko's largest output motor was an electret design that had a 0.1 Hp rating. [6] Certainly the potential for improvement and power upgrade exists with this free energy machine.

---

http://forum.allaboutcircuits.com/newsgroups/viewtopic.php?t=67792

Book Report

So I go out on my step and what to my wondering eyes doth appear but the box from Amazon.com containing the Jefimenko books that Bill Miller shamed me into finally buying. So I take a quick look inside and decide to give an initial report here. (Too much vector math in there for a thorough review.)

These books have some amazing advanced thinking in the understanding of Maxwell and EM. One first thought is the consideration of causality. This is typically totally ignored in the EM community. Evidence of that is the fact that EM waves are widely held to be propagated by the E field creating the H field and the H field creating the E field as it goes along. Too bad it's just not true! Jefimenko points out that causality demands that that an event must be PRECEDED by it's cause! Simultaneous events CANNOT be "causal" of each other. Hence E and H fields of waves are created by the WAVE SOURCE not each other! Same things goes for the E field created by Faraday induction. It simply cannot be "caused" by the time-varying Magnetic Field. "Magnetic induction" is therefore a misnomer. Such induction is caused by the source CURRENT and NOT the magnetic field!

This leads Jefimenko on to note that contrary to the "one E field" theory that has been believed for so many years, the inductive E field is clearly NOT the same field as an electrostatic E field. Jefimenko terms this inductive E field the "Electrokinetic Field" to show that it is a different field from the electrostatic E field. Very good. However, old habits die hard and even Jefimenko persists in writing an expression for a "total E field" following Maxwell as consisting of a sum of the electrostatic and electrokinetic parts as if they were both the "same" kind of E field.

Jefimenko then proceeds to illustrate the electrokinetic fields with a series of calculations and examples using his formulas as an approach. He presents it as basically a "new" way of doing this and in one sense it is compared to the commonly used and non-causal bogus "flux linkage" methods. However, he fails to note that the causal Neumann formula is in essence identical to his formulation and has been a standard formulation for years for the calculation of the "electrokinetic field" or what is usually termed "mutual induction". Nevertheless his example calculations are important basic references to the topic of Faraday induction. And the consideration of causality clearly shows that the Neumann approach has the edge over the "flux linkage" ideas with at times fail to give correct results.

But the subject doesn't end with induction, he pulls gravity into the mix. Of particular interest is that he shows that once you introduce causality into Newton's theory of gravitation, interesting things start to happen! Relativity suddenly begins to show up and even more interesting "action-reaction" is soon discovered to actually be a law that does NOT hold in all cases! The electromagnetic nature of gravity quickly becomes strongly hinted at and without action-reaction laws, those dreamed of devices such as anti-gravity ships and the "force-glove" that you wear to push over a building become theoretical possibilities! These are truly books full of thought-provoking new ways of looking at tired old physics!

I'm not going to be going through the large quantity of field theory math in these books in a hurry, but that's OK because clearly taking the time to go through in detail WILL be worth the effort! What can I say? Listen to Bill Miller and get that order off to Amazon.com now. At roughly $25 each these two books are a huge bargain to the usual EM text books costing hundreds of dollars and then being full of bogus ideas and misunderstandings of the established theories. Just do it! I did and I'm not sorry I did!

Benj

---

http://stupac2.blogspot.com/2007/04/bizarre-and-intriguing-story-of-oleg.html

The Bizarre and Intriguing Story of Oleg Jefimenko and the Solutions to Maxwell's Equations

I recently heard the story of Oleg Jefimenko during a lecture on Electrodynamics, specifically the general solution to Maxwell's Equations.

Jefimenko's tiny bit of fame comes from Jefimenko's Equations, which are the general solution to Maxwell's equations expressed solely in terms of sources, that is charge and current distributions. The equations are messy and difficult to work with, and aren't used much in practice. But they do reveal certain bits of physics (such as the applicability of the quasistatic approximation (the link goes to a thermodynamics page, but the idea is the same) and that fields must be created by sources), and it's always nice to have the general solution to a problem available.

These equations weren't written down until 1966, about a century after Maxwell's Equations were known. Some people will claim (as the Wikipedia article cited does) that Jefimenko's Equations were written down earlier, but those earlier versions are always slightly different and not quite complete. What's really funny is that Jefimenko wrote them down in an attempt to formulate an alternative to Maxwell's equations.

When my current Professor, David Griffiths, was in the process of writing a paper on the subject, he independently derived Jefimenko's equations, and tried to figure out if anyone had done it before. Other than some slightly tricky and annoying math, they're not hard to derive, so someone must have done it. He found that Jefimenko had written them in a book that was published by a company that had only published one other work, also by Jefimenko (apparently regular publishers wouldn't take his books, so he went to a prestige press). He contacted Jefimenko, and Jefimenko didn't believe that he had solved Maxwell's equations, but that he had created an electromagnetic theory separate from (and doubtless better than) Maxwell's. Of course he had done no such thing, his formulation is exactly equivalent to Maxwell's, but he wasn't buying it.

According to Griffiths, Jefimenko currently submits one or two papers a week to American journals, gets denied, then publishes them in Europe (where review is apparently not as stringent). I don't know what they're about, the Wikipedia article says he focuses on overthrowing Einstein's General Relativity and Maxwell.

I found this story behind some esoteric equations to be pretty amusing, and thought others might agree. I hope you've enjoyed the convoluted and intriguing story behind Jefimenko's equations.

[Most of my information comes from a lecture with Griffiths, and as such could not be found online. Anything that is available online has been referenced.]

---

http://electretscientific.com/

Electret Scientific Co
P.O. Box 4132
Star City, WV 26504 USA

Books by Professor Oleg Jefimenko

(also available from www.Amazon.com)

Gravitation and Cogravitation -- Developing Newton's Theory of Gravitation to its Physical and Mathematical Conclusion,
Paperback - List Price US$ 22.00
Hardback - List Price US$ 32.00

Electrostatic Experiments -- An Encyclopedia of Early Electrostatic Experiments, Demonstrations, Devices, and Apparatus,
by G.W. Francis (author), Oleg Jefimenko (editor)
Paperback - List Price US$ 24.00
Hardback - List Price US$ 48.00

Electromagnetic Retardation and Theory of Relativity -- New Chapters in the Classical Theory of Fields, 2nd edition,
Paperback - List Price US$ 24.00
Hardback - List Price US$ 44.00

Causality, Electromagnetic Induction, and Gravitation -- A Different approach to the Theory of Electromagnetic and Gravitational Fields, 2nd edition,
Paperback - List Price US$ 22.75
Hardback - List Price US$ 32.50

An Introduction to the Theory of Electric and Magnetic Fields, 2nd edition,
Hardback - List Price US$ 72.00

Electrostatic Motors -- Their History, Types, and Principles of Operation,

Out of Print -- free e-book download available soon

---

http://www.amazon.com/Electrostatic-Experiments-Encyclopedia-Demonstrations-Apparatus/dp/0917406133

Electrostatic Experiments: An Encyclopedia of Early Electrostatic Experiments, Demonstrations, Devices, and Apparatus (Paperback)

---

http://www.amazon.com/Electrostatic-Motors-Oleg-D-Jefimenko/dp/0917406028

Electrostatic Motors (Paperback)
by Oleg D. Jefimenko

***Scientific American* (October, 1974)**

Electrostatic Motors Are Powered by Electric Field of the Earth

by

C. L. Stong

Although no one can make a perpetual motion machine, anyone can tap the earth's electric field to run a homemade motor perpetually. The field exists in the atmosphere between the earth's surface and the ionosphere as an electric potential of about 360,000 volts. Estimates of the stored energy range from a million kilowatts to a billion kilowatts.

Energy in this form cannot be drawn on directly for driving ordinary electric motors. Such motors develop mechanical force through the interaction of magnetic fields that are generated with high electric current at low voltage, as Michael Faraday demonstrated in 1821. The earth's field provides relatively low direct current at high voltage, which is ideal for operating electrostatic motors similar in principle to the machine invented by Benjamin Franklin in 1748.

Motors of this type are based on the force of mutual attraction between unlike electric charges and the mutual repulsion of like charges. The energy of the field can be tapped with a simple antenna in the form of a vertical wire that carries one sharp point or more at its upper end. During fair weather the antenna will pick up potential at the rate of about 100 volts for each meter of height between the points and the earth's surface up to a few hundred feet. At higher altitudes the rate decreases. During local thunderstorms the pickup can amount to thousands of volts per foot. A meteorological hypothesis is that the field is maintained largely by thunderstorms, which pump electrons out of the air and inject them into the earth through bolts of lightning that continuously strike the surface at an average rate of 200 strokes per second.

Why not tap the field to supplement conventional energy resources? Several limitations must first be overcome. For example, a single sharp point can draw electric current from the surrounding air at a rate of only about a millionth of an ampere. An antenna consisting of a single point at the top of a 60-foot wire could be expected to deliver about a microampere at 2,000 volts; the rate is equivalent to .002 watt. A point-studded balloon tethered by a wire at an altitude of 75 meters might be expected to deliver .075 watt. A serious limitation appears as the altitude of the antenna exceeds about 200 meters. The correspondingly higher voltages become difficult to confine.

At an altitude of 200 meters the antenna should pick up some 20,000 volts. Air conducts reasonably well at that potential. Although nature provides effective magnetic materials in substances such as iron, nickel and cobalt, which explains why the electric-power industry developed around Faraday's magnetic dynamo, no comparably effective insulating substances exist for isolating the high voltages that would be required for electrostatic machines of comparable power. Even so, electrostatic motors, which are far simpler to build than electromagnetic ones, may find applications in special environments such as those from which magnetism must be excluded or in providing low power to apparatus at remote, unmanned stations by tapping the earth's field.

Apart from possible applications electrostatic motors make fascinating playthings. They have been studied extensively in recent years by Oleg D. Jefimenko and his graduate students at West Virginia University. The group has reconstructed models of Franklin's motors and developed advanced electrostatic machines of other types.

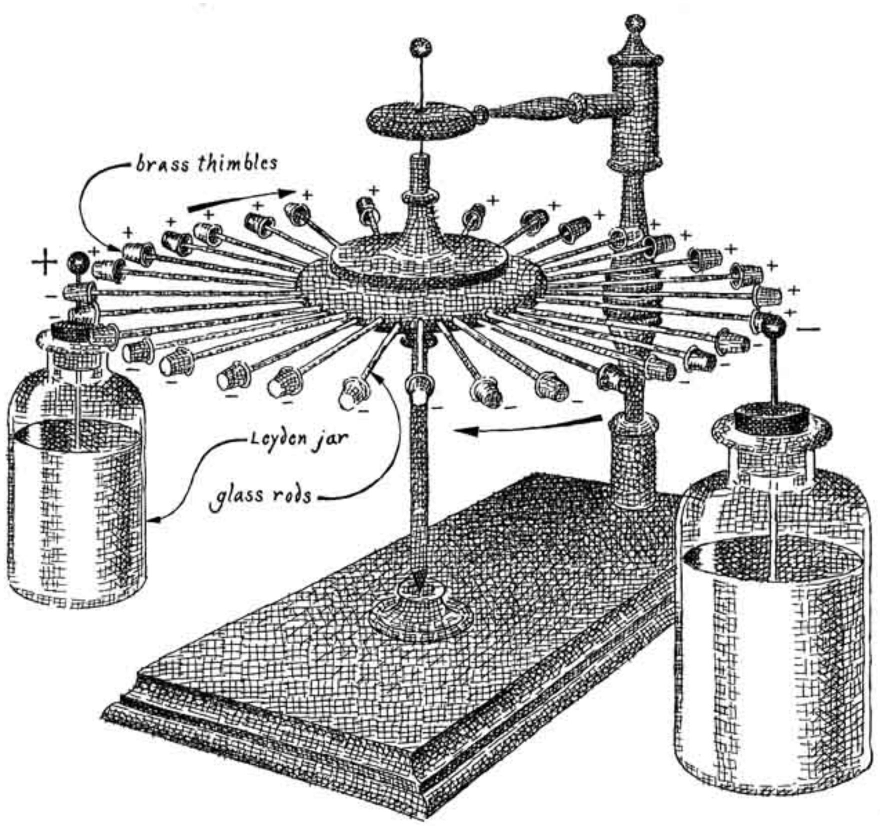

Although Franklin left no drawing of his motor, his description of it in a letter to Peter Collinson, a Fellow of the Royal Society, enabled Jefimenko to reconstruct a working model [see illustration at right]. Essentially the machine consists of a rimless wheel that turns in the horizontal plane on low-friction bearings. Each spoke of the "electric wheel," as Franklin called the machine, consists of a glass rod with a brass thimble at its tip. An electrostatic charge for driving the motor was stored in Leyden jars. A Leyden jar is a primitive form of the modern high-voltage capacitor. Franklin charged his jars with an electrostatic generator.

The high-voltage terminals of two or more Leyden jars that carried charges of opposite polarity were positioned to graze the thimbles on opposite sides of the rotating wheel. The motor was started by hand. Thereafter a spark would jump from the high-voltage terminal to each passing thimble and impart to it a charge of the same polarity as that of the terminal. The force of repulsion between the like charges imparted momentum to the wheel.

Conversely, the thimbles were attracted by the oppositely charged electrode of the Leyden jar Franklin placed on the opposite side of the wheel. As the thimbles grazed that jar, a spark would again transfer charge, which was of opposite polarity. Thus the thimbles were simultaneously pushed and pulled by the high-voltage terminals exactly as was needed to accelerate the wheel.

Franklin was not altogether happy with his motor. The reason was that running it required, in his words, "a foreign force, to wit, that of the bottles." He made a second version of the machine without Leyden jars.

In this design the rotor consisted principally of a 17-inch disk of glass mounted to rotate in the horizontal plane on low-friction bearings. Both surfaces of the disk were coated with a film of gold, except for a boundary around the edge. The rotor was thus constructed much like a modern flat-plate capacitor.

Twelve evenly spaced metal spheres, cemented to the edge of the disk, were connected alternately to the top and bottom gold films. Twelve stationary thimbles supported by insulating columns were spaced around the disk to graze the rotating metal spheres. When Franklin placed opposite charges on the top and bottom films and gave the rotor a push, the machine ran just as well as his first design, and for the same reason. According to Franklin, this machine would make up to 50 turns a minute and would run for 30 minutes on a single charge.

Jefimenko gives both motors an initial charge from a 20,000-volt generator. They consume current at the rate of about a millionth of an ampere when they are running at full speed. The rate is equivalent to .02 watt, which is the power required to lift a 20-gram weight 10 centimeters (or an ounce 2.9 inches) in one second.

Jefimenko wondered if Franklin's motor could be made more powerful. As Jefimenko explains, the force can be increased by adding both moving and fixed electrodes. This stratagem is limited by the available space. If the electrodes are spaced too close, sparks tend to jump from electrode to electrode around the rotor, thereby in effect short-circuiting the machine. Alternatively the rotor could be made cylindrical to carry electrodes in the form of long strips or plates. This scheme could perhaps increase the output power by a factor of 1,000.

Reviewing the history of electrostatic machines, Jefimenko came across a paper published in 1870 by Johann Christoff Poggendorff, a German physicist. It described an electrostatic motor fitted with a rotor that carried no electrodes. The machine consisted of an uncoated disk of glass that rotated in the vertical plane on low-friction bearings between opposing crosses of ebonite. Each insulating arm of the crosses supported a comblike row of sharp needle points that grazed the glass.

When opposing combs on opposite sides of the glass were charged in opposite polarity to potentials in excess of 2,000 volts, air in the vicinity of the points on both sides of the glass was ionized. A bluish glow surrounded the points, which emitted a faint hissing sound. The effect, which is variously known as St. Elmo's fire and corona discharge, deposited static charges on both sides of the rotor.

Almost the entire surface of the glass acquired a coating of either positive or negative fixed charges, depending on the polarity of the combs. The forces of repulsion and attraction between glass so charged and the combs were substantially larger than they were in Franklin's charged thimbles. The forces were also steadier, because in effect the distances between the combs and the charged areas remained constant. It should be noted that adjacent combs on the same side of the glass carried charges of opposite polarity, so that the resulting forces of attraction and repulsion acted in unison to impart momentum to the disk, as they did in Franklin's motor.

By continued experimentation Poggendorff learned that he should slant the teeth of the combs to spray charge on the glass at an angle. The resulting asymmetrical force made the motor self-starting and unidirectional. When the teeth were perpendicular to the glass surface, the forces were symmetrical, as they were in Franklin's motor. When the machine was started by hand, it ran equally well in either direction.

Poggendorff was immensely pleased by the rate at which his machine converted charge into mechanical motion. He concluded his paper with a faintly odious reference to Franklin's device. "That such a quantity of electricity must produce a far greater force than that in the [Franklin] electric roasting spit," he wrote, "is perfectly obvious and nowadays would not be denied by Franklin himself. With one grain of gunpowder one cannot achieve so much as with one hundred pounds."

Electrostatic motors are now classified in general by the method by which charge is either stored in the machine or transferred to the rotor. Poggendorff's machine belongs to the corona type, which has attracted the most attention in recent years. Although its measured efficiency is better than 50 percent, Poggendorff regarded it merely as an apparatus for investigating electrical phenomena. He wrote that "it would be a sanguine hope if one wanted to believe that any useful mechanical effect could be achieved with it."

Poggendorff's negative attitude toward the usefulness of his design may well have retarded its subsequent development. A modern version of the machine constructed in Jefimenko's laboratory has an output of approximately .1 horsepower. It operates at speeds of up to 12,000 revolutions per minute at an efficiency of substantially more than 50 percent. In one form the modern corona motor consists of a plastic cylinder that turns on an axial shaft inside a concentric set of knife-edge electrodes that spray charge on the surface of the cylinder [see illustration at left]. Forces that act between the sprayed charges and the knife-edge electrodes impart momentum to the cylinder.

Machines of this kind can be made of almost any inexpensive dielectric materials, including plastics, wood and even cardboard. The only essential metal parts are the electrodes and their interconnecting leads. Even they can be contrived of metallic foil backed by any stiff dielectric. The shaft can be made of plastic that turns in air bearings. By resorting to such stratagems experimenters can devise motors that are extremely light in proportion to their power output. Corona motors require no brushes or commutators. A potential of at least 2,000 volts, however, is essential for initiating corona discharge at the knife-edges.

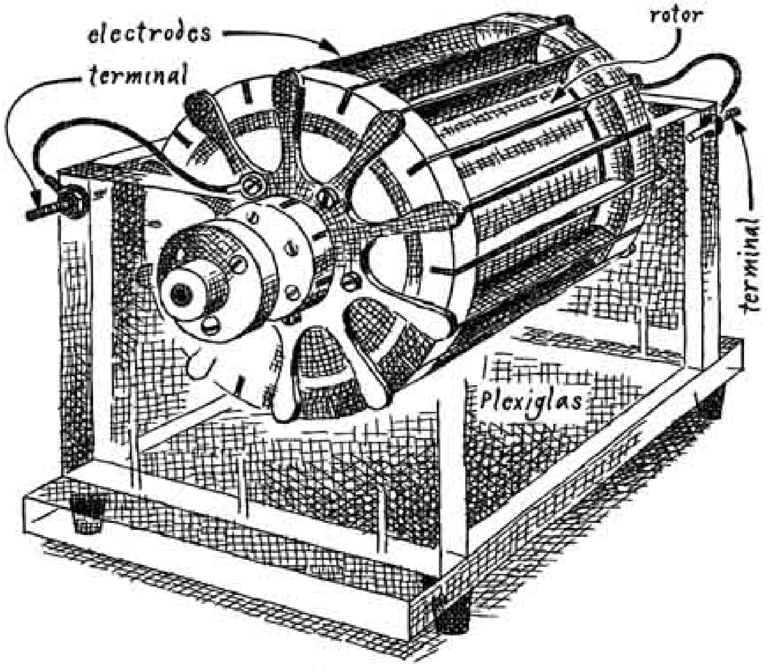

A smaller and simpler version of the machine was demonstrated in 1961 by J. D. N. Van Wyck and G. J. Kühn in South Africa. This motor consisted of a plastic disk about three millimeters thick and 40 millimeters in diameter supported in the horizontal plane by a slender shaft that turned in jeweled bearings. Six radially directed needle points grazed the rim of the disk at equal intervals. When the machine operated from a source of from 8,000 to 13,000 volts, rotational speeds of up to 12,000 revolutions per minute were measured.

I made a corona motor with Plexiglas tubing two inches in diameter and one and a half inches long. It employed stiffbacked single-edge razor blades as electrodes. The bore of the tube was lined with a strip of aluminum foil, a stratagem devised in Jefimenko's laboratory to increase the voltage gradient in the vicinity of the electrodes and thus to increase the amount of charge that can be deposited on the surface of the cylinder. I coated all surfaces of the razor blades except the cutting edges and all interconnecting wiring with "anticorona dope," a cementlike liquid that dries to form a dielectric substance that reduces the loss of energy through corona discharges in nonproductive portions of the circuit.

The axial shaft that supports the cylinder on pivot bearings was cut out of a steel knitting needle. The ends of the shaft were ground and polished to 30degree points. To form the points I chucked the shaft in an electric hand drill, ground the metal against an oilstone and polished the resulting pivots against a wood lap coated with tripoli.

The bearings that supported the pivots were salvaged from the escapement mechanism of a discarded alarm clock. A pair of indented setscrews could be substituted for the clock bearings. The supporting frame was made of quarter-inch Lucite. The motor can be made self-starting and unidirectional by slanting the knife-edges. Those who build the machine may discover, as I did, that the most difficult part of the project, balancing the rotor, is encountered after assembly. The rotor must be balanced both statically and dynamically.

Static balance was achieved by experimentally adding small bits of adhesive tape to the inner surface of the aluminum foil that lines the cylinder until the rotor remained stationary at all positions to which it was set by hand. When the rotor was balanced and power was applied, the motor immediately came up to speed, but it shook violently. I had corrected the imbalance caused by a lump of cement at one end of the rotor by adding a counterweight on the opposite side at the opposite end of the cylinder. Centrifugal forces at the ends were 180 degrees out of phase, thus constituting a couple.

The dynamic balancing, which is achieved largely by cut-and-try methods, took about as much time as the remainder of the construction. To check for dynamic balance suspend the motor freely with a string, run it at low speed and judge by the wiggle where a counterweight must be added. Adhesive tape makes a convenient counterweight material because it can be both applied and shifted easily.

I made the motor as light and frictionless as possible with the objective of operating it with energy from the earth's field. The field was tapped with an antenna consisting of 300 feet of No. 28 gauge stranded wire insulated with plastic. It is the kind of wire normally employed for interconnecting electronic components and is available from dealers in radio supplies.

The upper end of the wire was connected to a 20-foot length of metallic tinsel of the kind that serves for decorating a Christmas tree. The tinsel functioned as multiple needle points. Strips cut from window screening would doubtless work equally well.

The upper end of the tinsel was hoisted aloft by a cluster of three weather balloons. Such balloons, each three feet in diameter, and the helium to inflate them are available from the Edmund Scientific Co. (300 Edscorp Building, Barrington, N.J. 08007). The weight in pounds that a helium-filled balloon of spherical shape can lift is roughly equal to a quarter of the cube of its radius in feet. To my delight the motor began to run slowly when the tinsel reached an altitude of about 100 feet. At 300 feet the rotor made between 500 and 700 revolutions per minute.

A note of warning is appropriate at this point. Although a 300-foot vertical antenna can be handled safely in fair weather, it can pick up a lethal charge during thunderstorms. Franklin was incredibly lucky to have survived his celebrated kite experiment. A European investigator who tried to duplicate Franklin's observations was killed by a bolt of lightning. The 300-foot antenna wire can hold enough charge to give a substantial jolt, even during fair weather. Always ground the lower end of the wire when it is not supplying a load, such as the motor.

To run the motor connect the antenna to one set of electrodes and ground the other set. Do not connect the antenna to an insulated object of substantial size, such as an automobile. A hazardous charge can accumulate. Never fly the balloon in a city or in any other location where the antenna can drift into contact with a high-voltage power line. Never fly it below clouds or leave it aloft unattended.

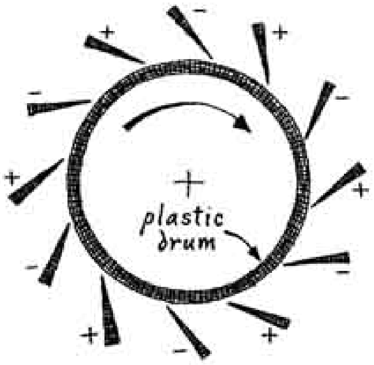

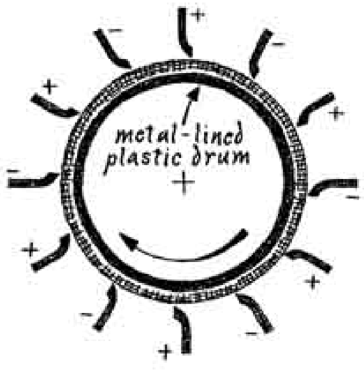

A variety of corona motors have been constructed in Jefimenko's laboratory. He has learned that their performance can be vastly improved by properly shaping the corona-producing electrodes [see illustration at right]. The working surface of the rotors should be made of a fairly thin plastic, such as Plexiglas or Mylar. Moreover, as I have mentioned, the inner surface of the cylinder should be backed by conducting foil to enhance the corona. Effective cylinders can be formed inexpensively out of plastic sewer pipe. Corona rotors can of course also be made in the form of disks.

One model consists of a series of disks mounted on a common shaft. Double-edged electrodes placed radially between adjacent disks function much like Poggendorff's combs. This design needs no foil lining or backing because a potential gradient exists between electrodes on opposite sides of the disks. It is even possible to build a linear corona motor, a design that serves to achieve translational motion. A strip of plastic is placed between sets of knife-edge electrodes slanted to initiate motion in the desired direction.

Notwithstanding the problem of handling potentials on the order of a million volts without effective insulation materials, Jefimenko foresees the possibility of at least limited application of corona power machines. In The Physics Teacher (March, 1971) he and David K. Walker wrote: "These motors could be very useful for direct operation from high-voltage d.c. transmission lines as, for example, the 800 kV Pacific Northwest-Southwest Intertie, which is now being constructed between the Columbia River basin and California. It is conceivable that such motors could replace the complex installations now needed for converting the high-voltage d.c. to low-voltage a.c. All that would be required if corona motors were used for this purpose would be to operate local low-voltage a.c. generators from corona motors powered directly from the high-voltage d.c. line."

As Jefimenko points out, a limiting factor of the corona motor is its required minimum potential of 2,000 volts. This limitation is circumvented by a novel electrostatic motor invented in 1961 by a Russian physicist, A. N. Gubkin. The motor is based on an electret made in 1922 by Mototaro Eguchi, professor of physics at the Higher Naval College in Tokyo.

An electret is a sheet or slab of waxy dielectric material that supports an electric field, much as a permanent magnet carries a magnetic field. Strongly charged carnauba-wax electrets are available commercially, along with other electrostatic devices, from the Electret Scientific Company (P.O. Box4132, Star City, W.Va. 26505). A recipe for an effective electret material is 45 percent carnauba wax, 45 percent water-white rosin and 10 percent white beeswax. Some experimenters substitute Halowax for the rosin.

The ingredients are melted and left to cool to the solid phase in a direct-current electric field of several thousand volts. The wax continues to support the field even though the external source of potential is turned off [see "The Amateur Scientist, SCIENTIFIC AMERICAN, November, 1960, and July, 1968]. The electret reacts to neighboring charges exactly as though it were a charged electrode, that is, it is physically attracted or repelled depending on the polarity of the neighboring electrode.

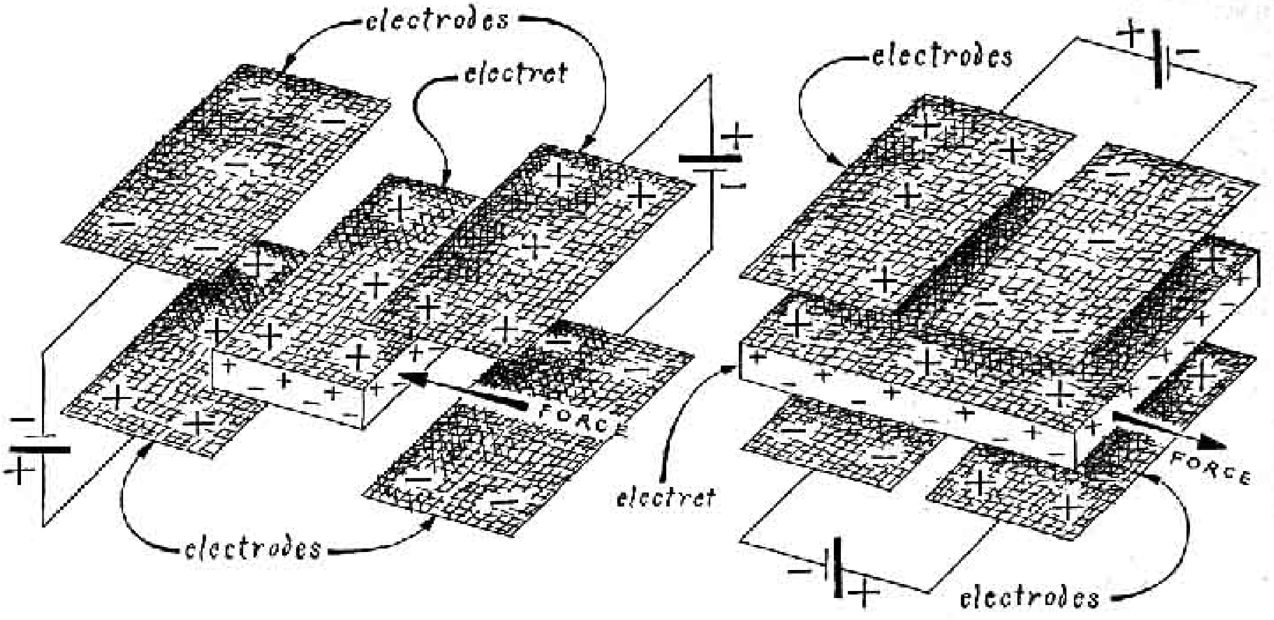

Gubkin harnessed this effect to make a motor. The rotor consisted of a pair of electrets in the shape of sectors supported at opposite ends of a shaft. The center of the shaft was supported transversely by an axle. When the rotor turned, the electrets were swept between adjacent pairs of charged metallic plates, which were also in the form of sectors.

The plates were electrified by an external source of power through the polarity-reversing switch known as a commutator. The commutator applied to the electrodes a charge of polarity opposite to the charge of the attracted electret. As the electret moved between the attracting plates, however, the commutator switched the plates to matching polarity. The alternate push and pull imparted momentum to the rotor in exact analogy to Franklin's motor.

Gubkin's motor was deficient in two major respects. The distances between the electrodes and the electrets were needlessly large, so that the forces of attraction and repulsion were needlessly weak. Moreover, during the electret's transit between electrodes its surfaces were unshielded. Unshielded electrets attract neutralizing ions from the air and lose their charge within hours or days.

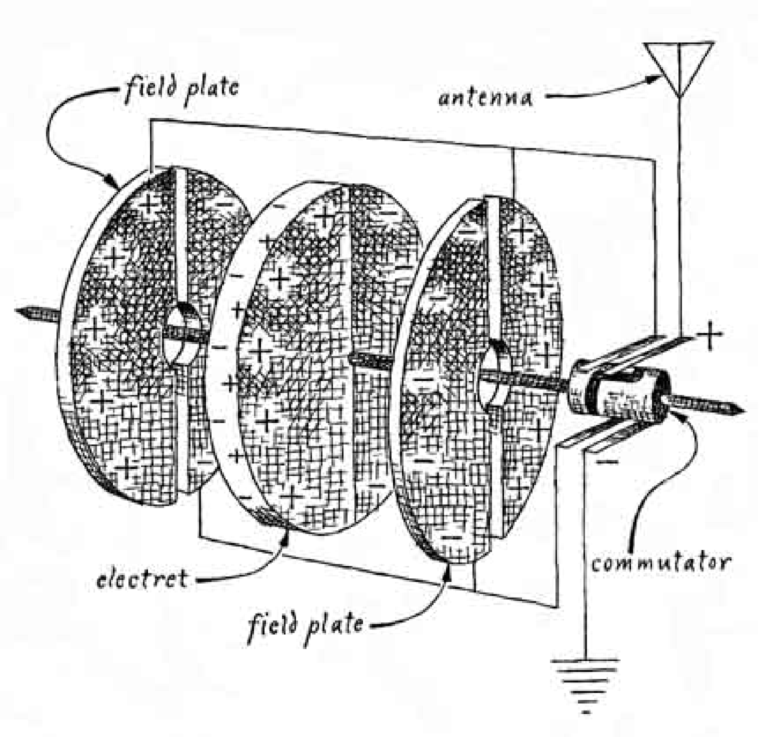

Both inherent deficiencies of Gubkin's motor have been corrected in Jefimenko's laboratory by taking advantage of what is termed the slot effect. Instead of sandwiching the electret alternately between pairs of metal plates, Jefimenko employs opposing pairs of adjacent plates [see illustration at rightt]. The adjacent plates are separated by a narrow slot. When adjacent plates carry charges of opposite polarity, the electret experiences a force at right angles to the slot and in the plane of the electret. The strength of the force is at a maximum because the electret is close to the electrodes. Simultaneously the electrodes function as shields to prevent the neutralization of the electret by free ions.

Motors based on the slot effect can be designed in a number of forms. One design consists of an electret in the shape of a wafer-thin sheet of Mylar supported by a flat disk of balsa wood 100 millimeters in diameter and three millimeters thick. (A long-lasting charge is imparted to the Mylar by immersing it in a field of a few thousand volts from an electrostatic generator after the motor is assembled.) This rotor is sandwiched between four semicircular sectors that are cross-connected [see illustration].

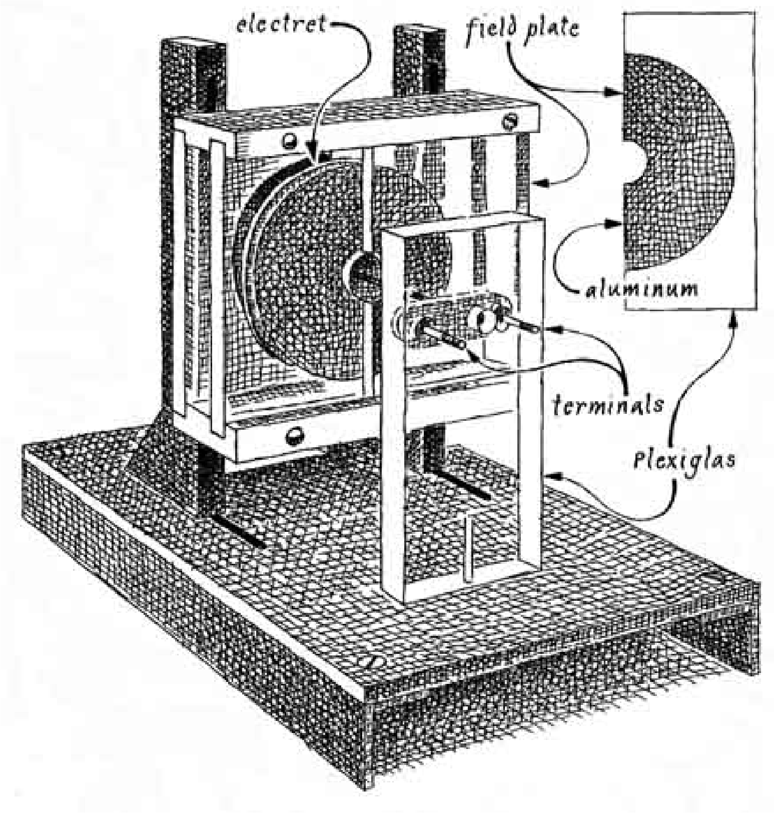

The electret is mounted on a four-millimeter shaft of plastic that turns in jeweled bearings. The conducting surfaces of the commutator consist of dried India ink. The brushes are one-millimeter strips of kitchen aluminum foil. The motor operates on a few microwatts of power.

Jefimenko has demonstrated a similar motor that was designed to turn at a rate of about 60 revolutions per minute and develop a millionth of a horsepower on a 24-foot antenna having a small polonium probe at its upper end. (By emitting positive charges probes of this type tap the earth's field somewhat more efficiently than needle points do.) The performance of the motor easily met the design specifications. The charm of these motors lies in the fact that, although they do not accomplish very much, they can run forever.

Bibliography

ATMOSPHERIC ELECTRICITY. J. Alan Chalmers. Pergamon Press, 1968.

ELECTROSTATIC MOTORS: THEIR HISTORY, TYPES AND PRINCIPLES OF OPERATION. Oleg D. Jefimenko. Electret Scientific Company, 1973.

ELECTROSTATICS AND ITS APPLICATIONS. Edited by A. D. Moore. John Wiley & Sons, 1973.

***Annales de la Fondation Louie de Broglie* 32 (1) : 117 (2007)**

Oleg Jefimenko : Gravitation and Cogravitation -- Note de Lecture (PDF)

Your Support Maintains this Service --

BUY

US 20140152016A1

(19) **United States**

(12) **Patent Application Publication**

JENNINGS

(10) **Pub. No.: US 2014/0152016 A1**

(43) **Pub. Date: Jun. 5, 2014**

(54) **ATMOSPHERIC TRANSDUCTION SYSTEM**

(71) Applicant: **JAMES EDWARD JENNINGS**, SUPERIOR, CO (US)

(72) Inventor: **JAMES EDWARD JENNINGS**, SUPERIOR, CO (US)

(21) Appl. No.: **13/692,121**

(22) Filed: **Dec. 3, 2012**

Publication Classification

(51) **Int. Cl.**
H02N 11/00 (2006.01)
F03D 9/00 (2006.01)

(52) **U.S. Cl.**
CPC ***H02N 11/002*** (2013.01); ***F03D 9/003*** (2013.01)
USPC **290/55**; 310/308

(57) **ABSTRACT**

Atmospheric Transduction System including a Power Frequency broadcast station, a receiver, and a network. The Power Frequency broadcast station includes a transmitter and a computer server. The receiver is in communication with the Power Frequency broadcast transmitter and also includes a user interface for receiving user input commands comprising a request for information from the Power Frequency broadcast station. The receiver is configured to establish a two-way communication path between the receiver and the Power Frequency broadcast transmitter. The network is in communication with the transducer, controller and the receiver for exchanging information therebetween. In response to oscillation translation and/or rotation of the electronic transducer, portions of forces induced by the mass are transferred to the piezoelectric elements. Electrical energy output by these piezoelectric elements is received in a power controller and can be applied to the battery as self charging. The piezoelectric transducer includes a conductive rotor and bearings, at least one of them incorporating a vibrator of mechanical oscillation, having a piezoelectric transducer converting mechanical vibrations into electric power.

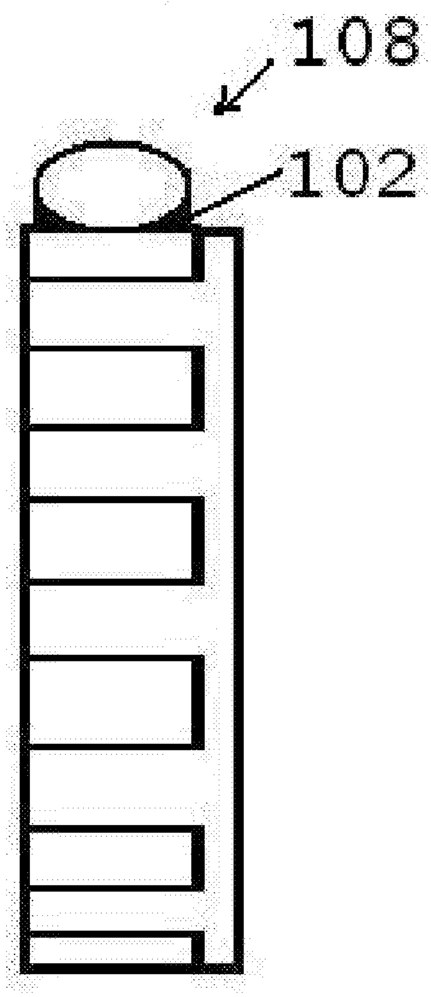

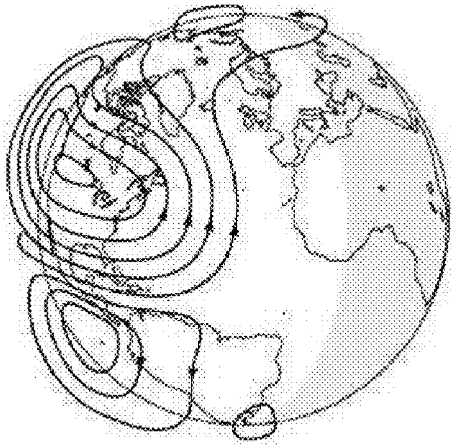

FIG.1

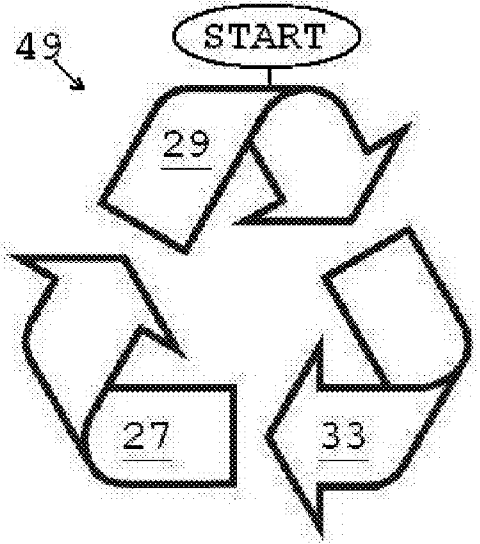
49
START
29
27
33

FIG.2

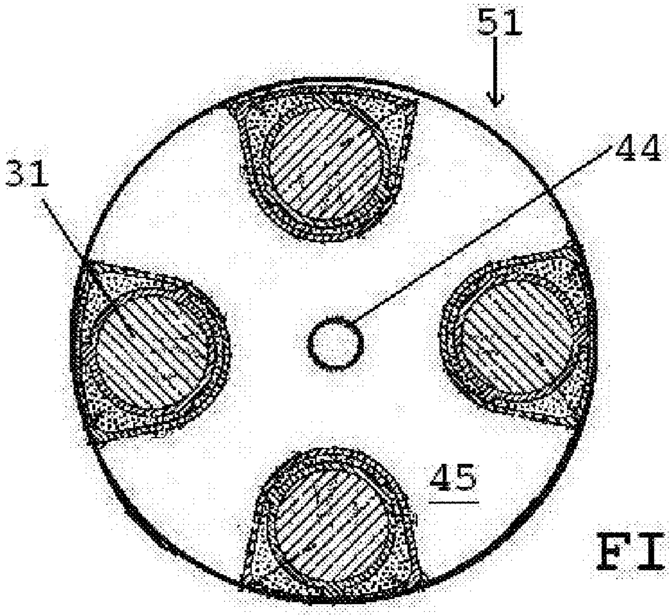

FIG.3

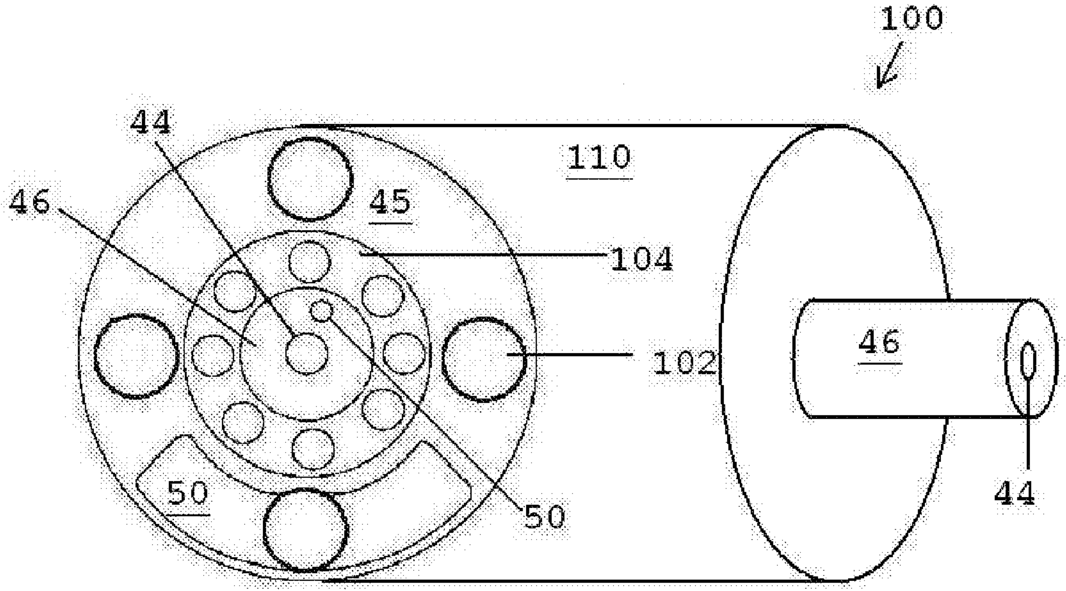
100
44
110
46
45
104
102
46
50
50
44

FIG. 4

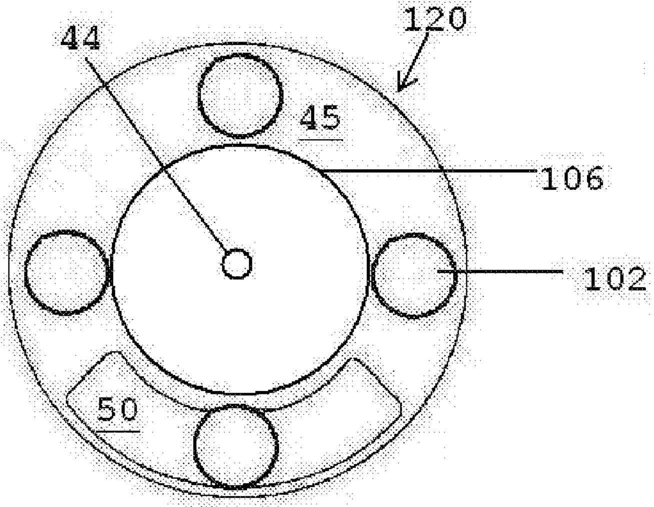
120
44
45
106
102
50

FIG. 5

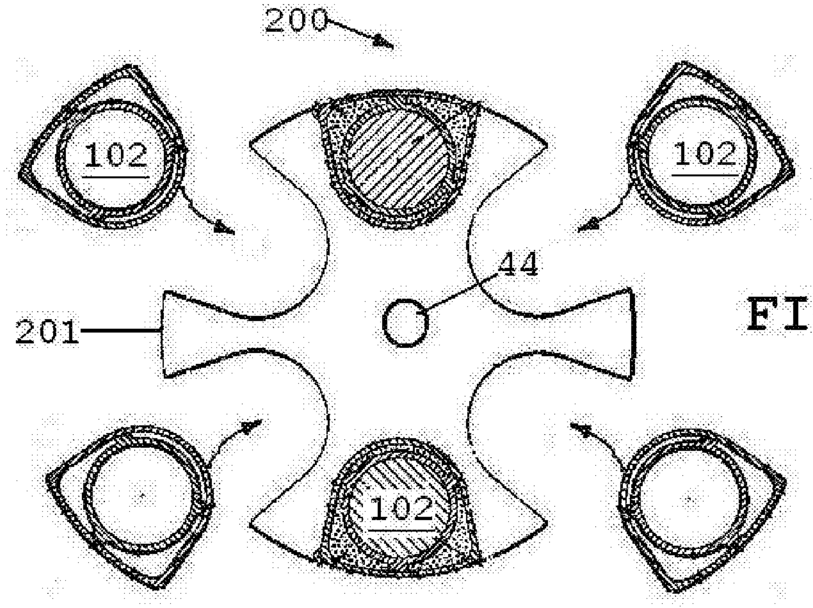
200
102
102
44
201
102

FIG. 6

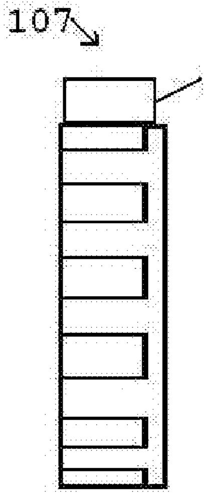

FIG. 7

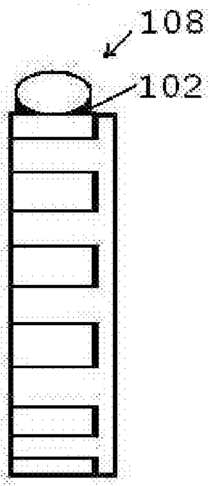

FIG. 8

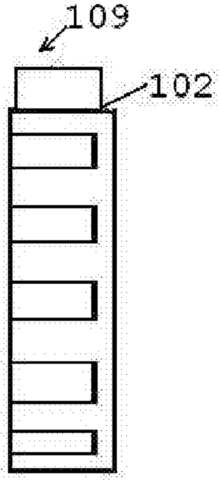

FIG. 9

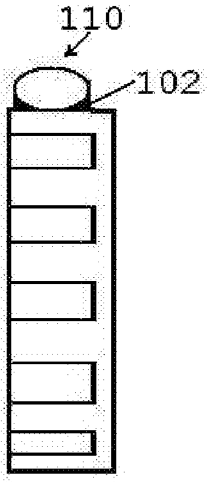

FIG. 10

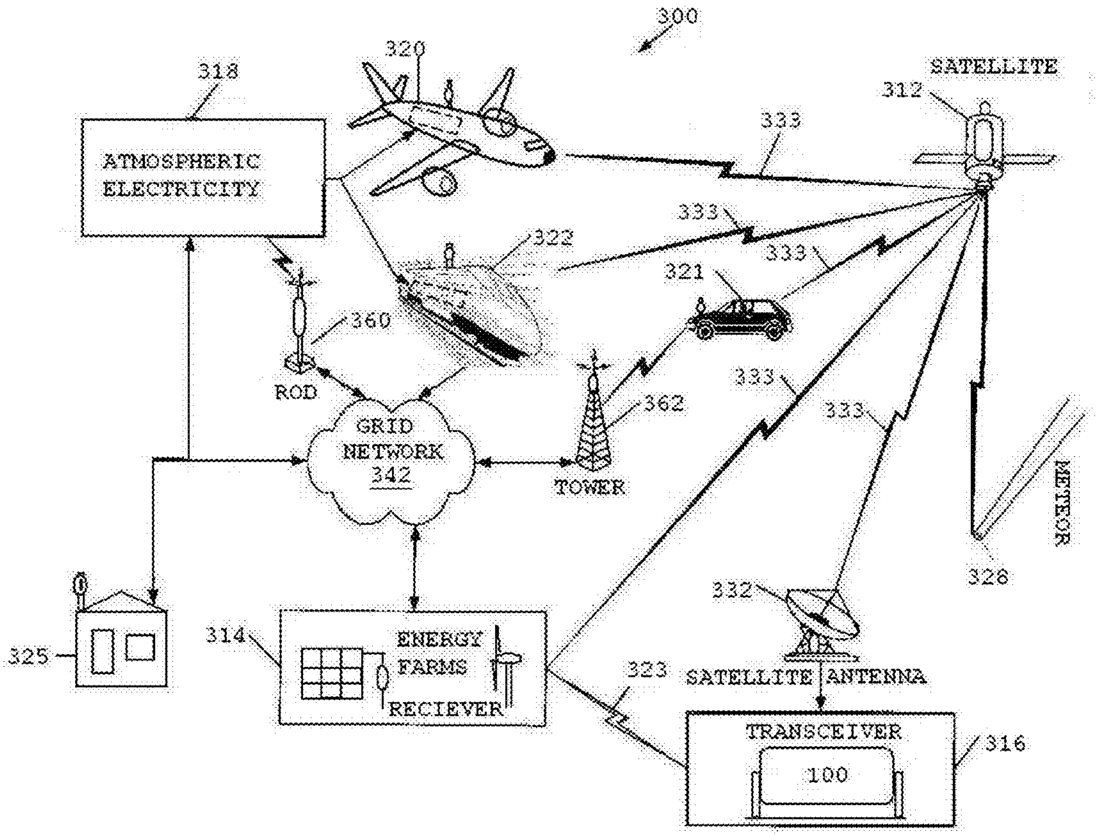
ATMOSPHERIC ELECTRICITY
SATELLITE
ROD
GRID NETWORK 342
TOWER
METEOR
ENERGY FARMS
RECIEVER
SATELLITE ANTENNA
TRANSCEIVER

FIG.11

FIG. 12

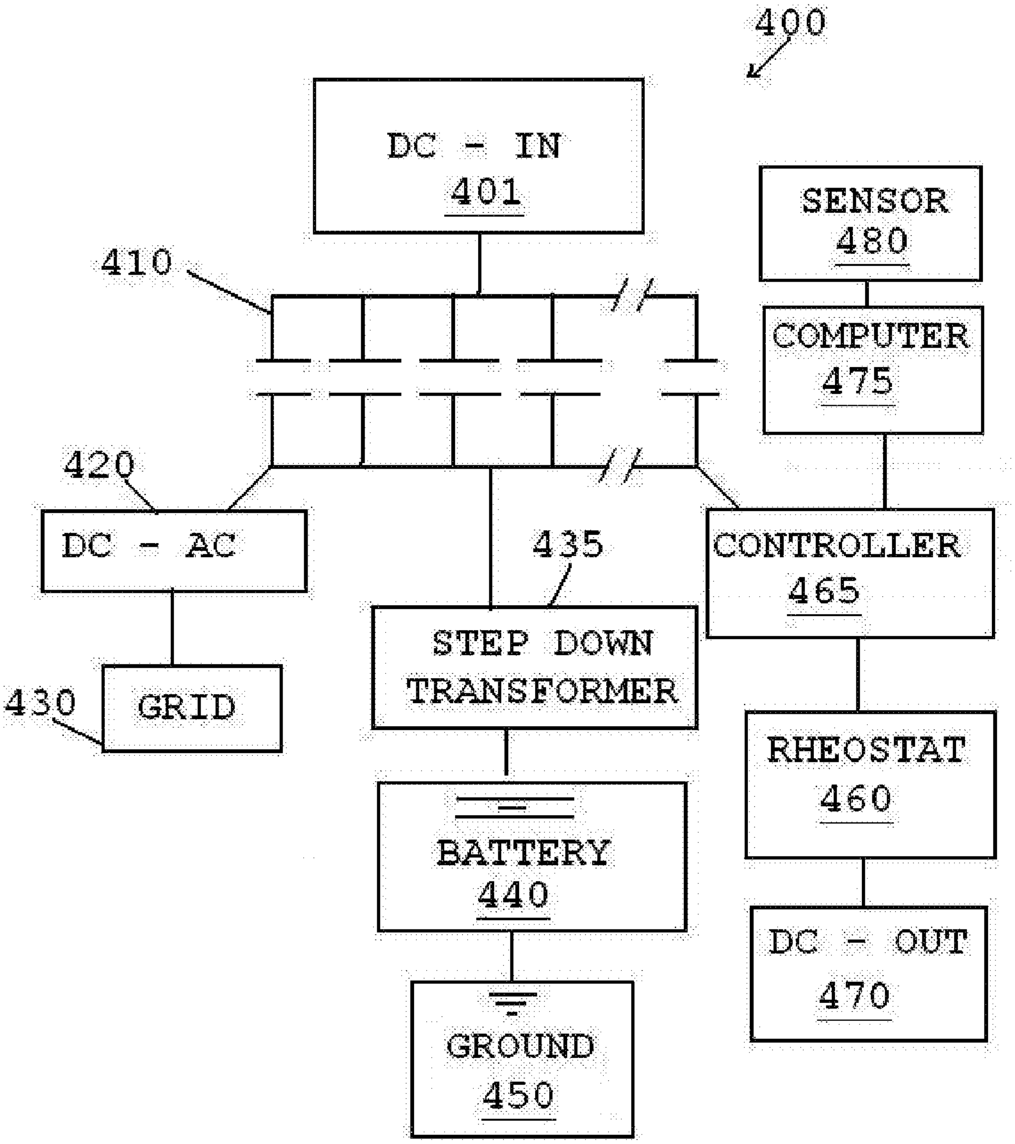
400
DC - IN
401
410
SENSOR
480
COMPUTER
475
420
DC - AC
435
CONTROLLER
465
STEP DOWN
TRANSFORMER
430
GRID
RHEOSTAT
460
BATTERY
440
DC - OUT
470
GROUND
450

FIG.13

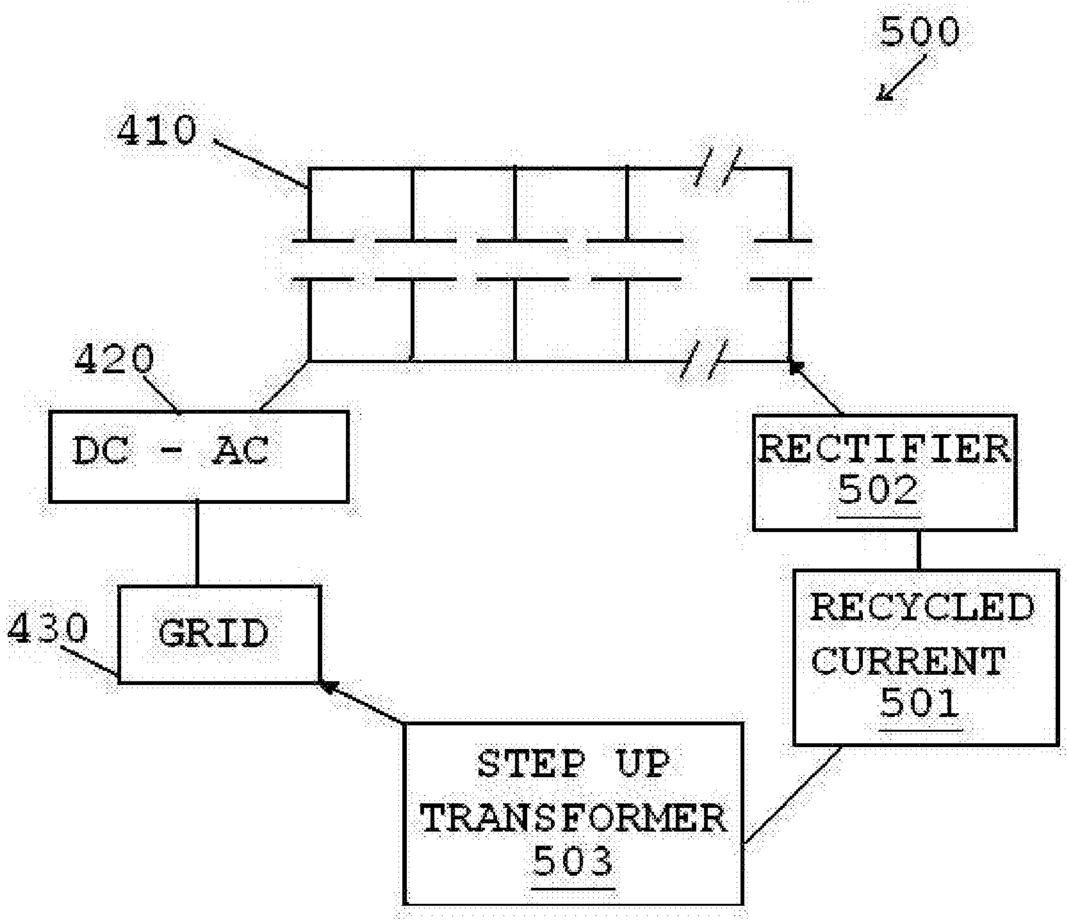
500
410
420
DC - AC
RECTIFIER
502
430
GRID
RECYCLED
CURRENT
501
STEP UP
TRANSFORMER
503

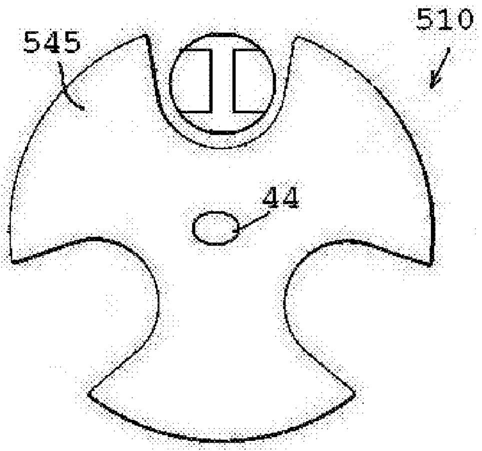

FIG. 14

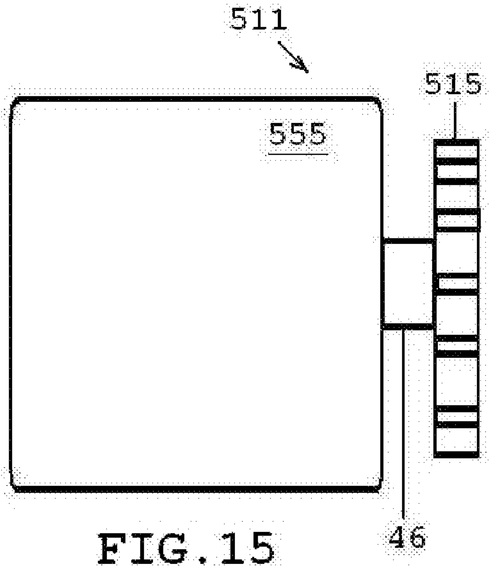
511
515
555
46

FIG. 15

FIG. 16

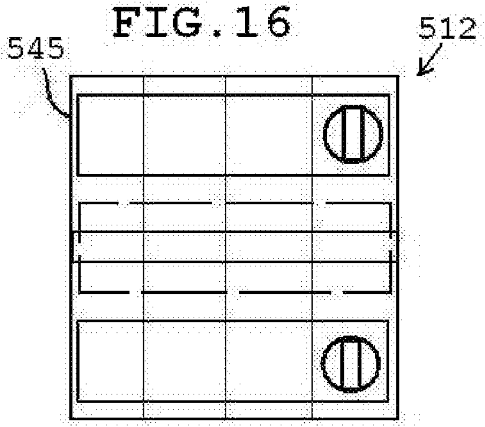

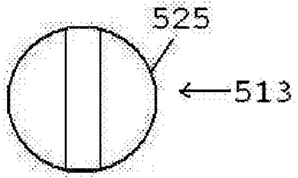

FIG. 17

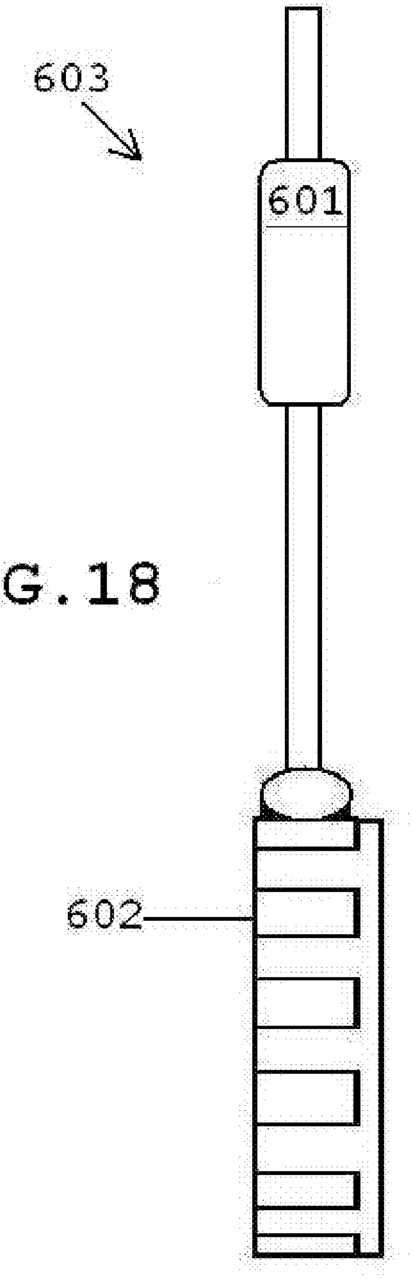

FIG.18

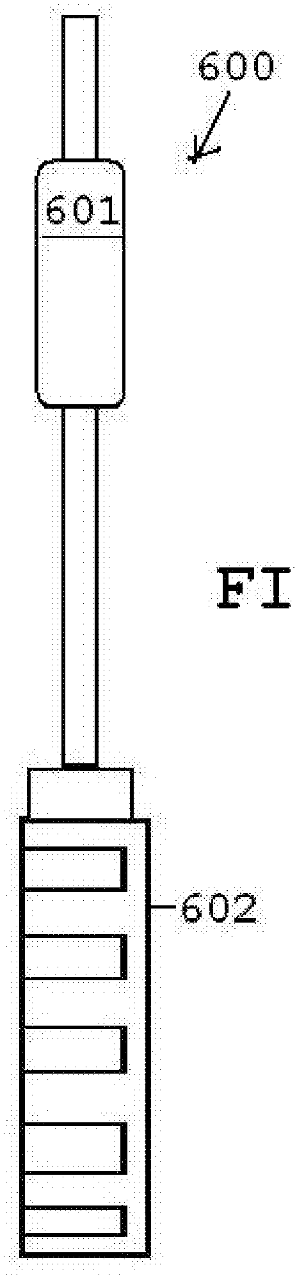

FIG.19

FIG.20

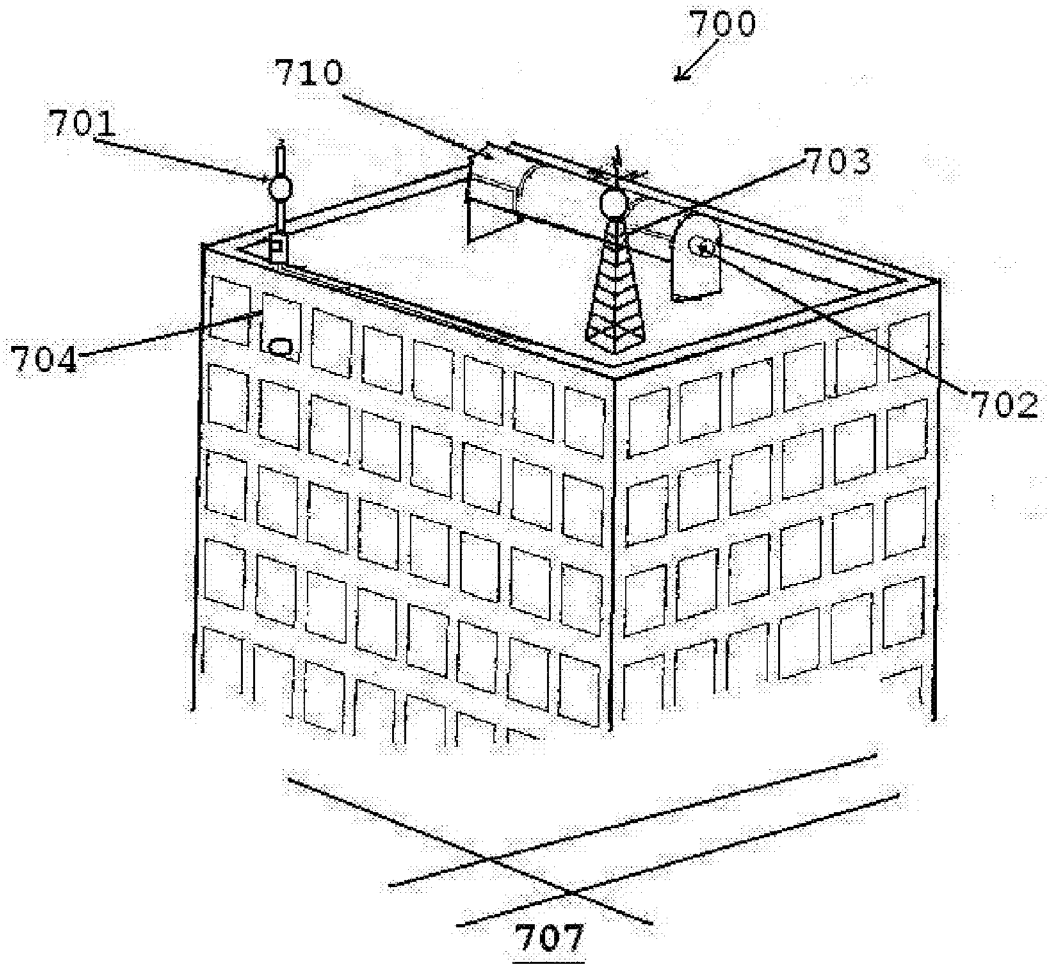
700
710
701
703
704
702
707

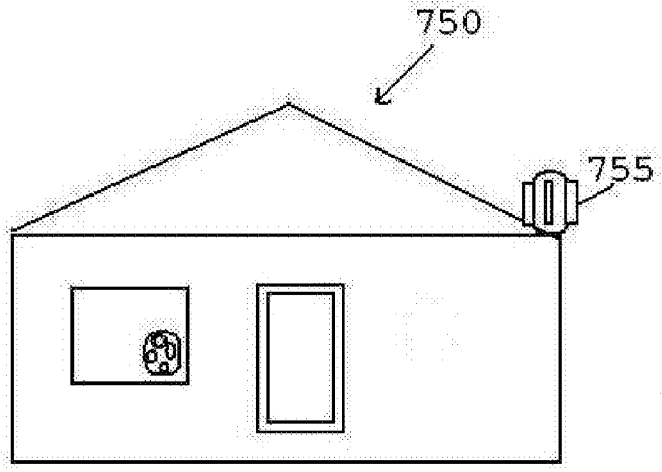

FIG. 21

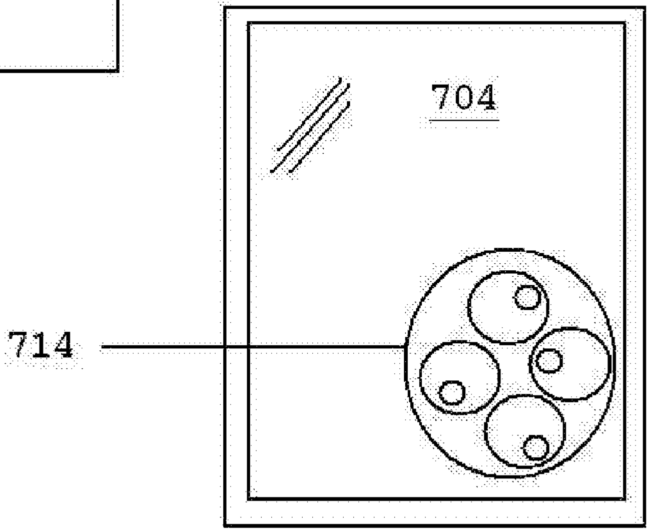

FIG. 22

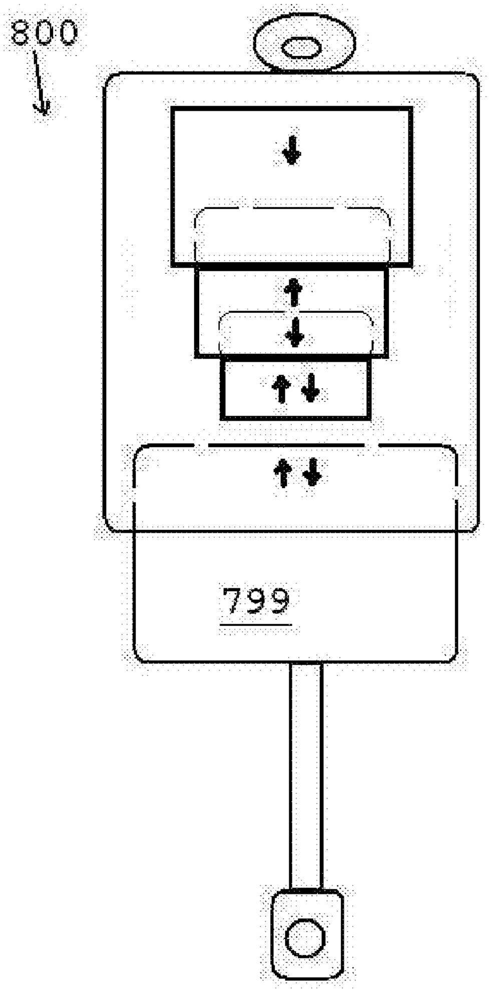

FIG.23

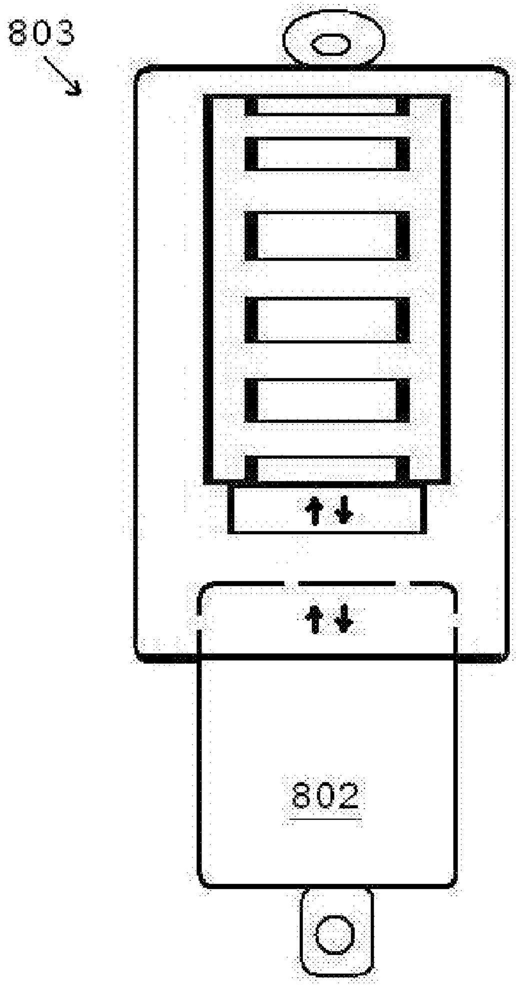

FIG.24

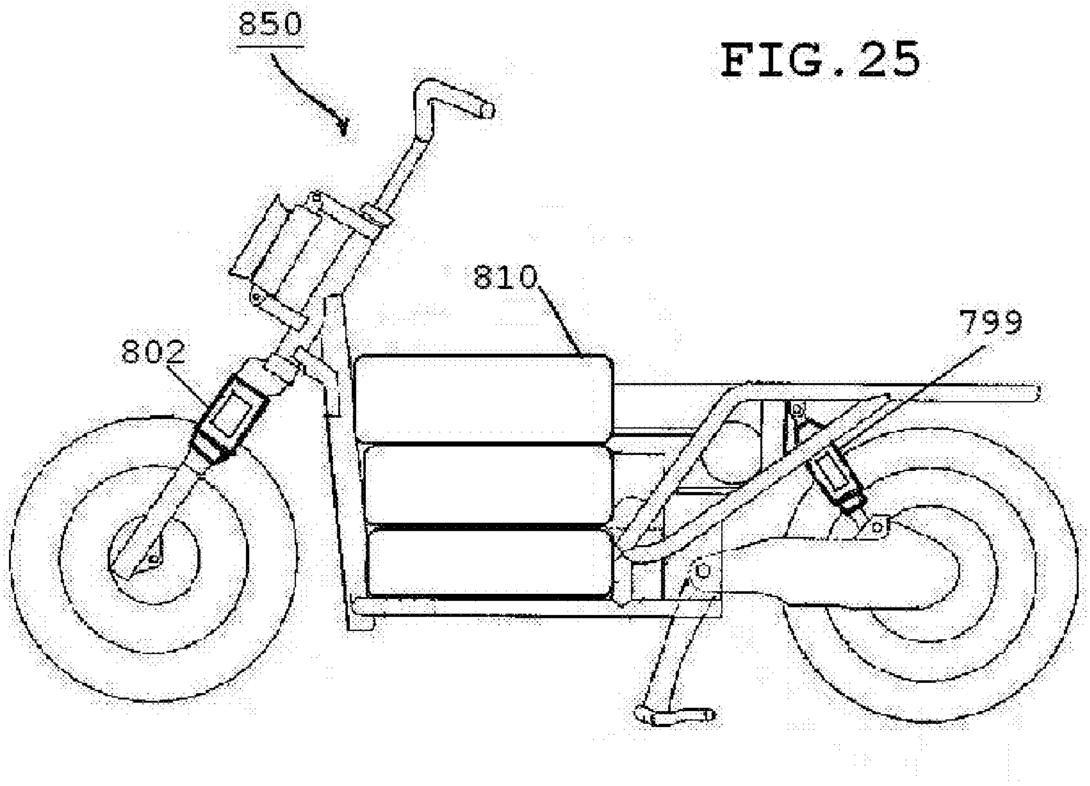
850
810
799
802

FIG.25

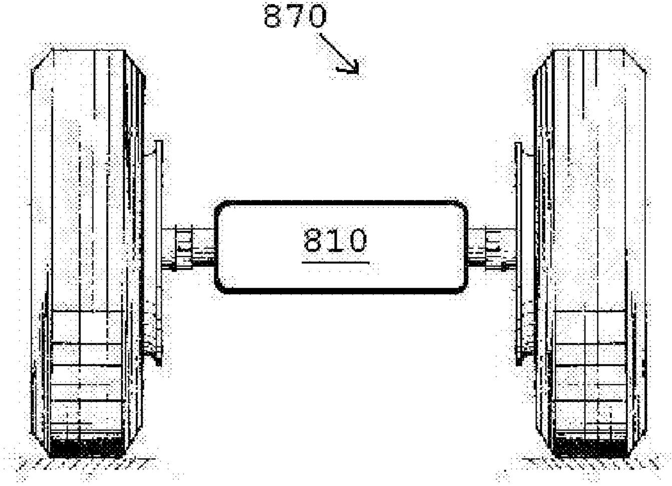

FIG.26

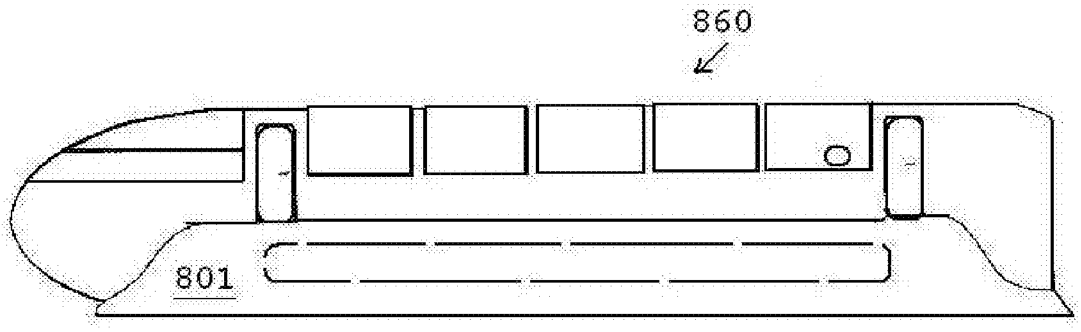

FIG.27

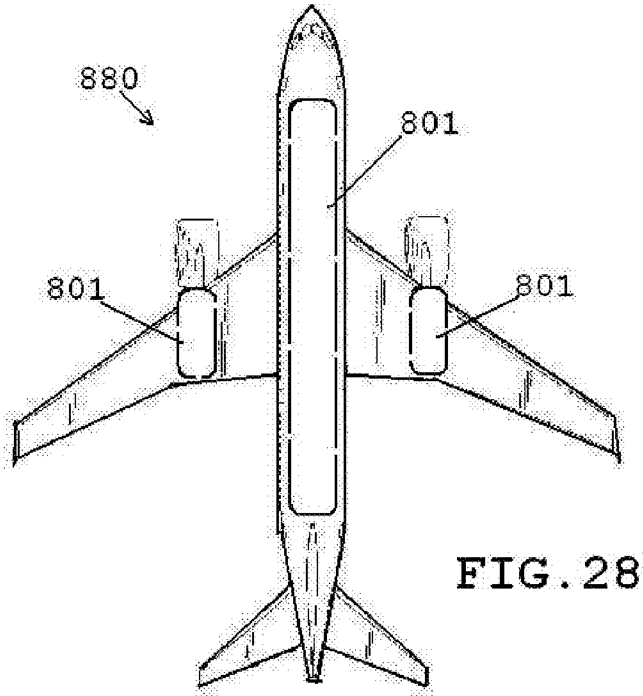
880
801
801
801

FIG.28

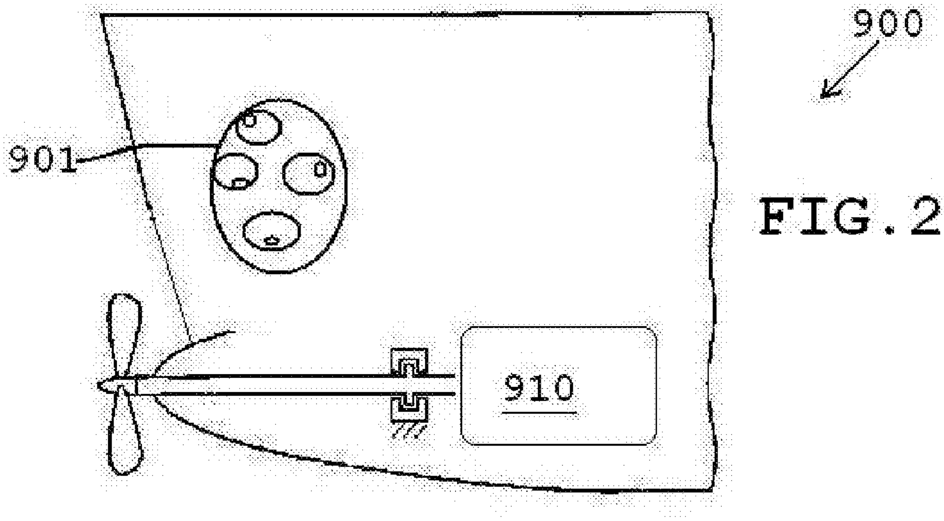

FIG.29

FIG.30

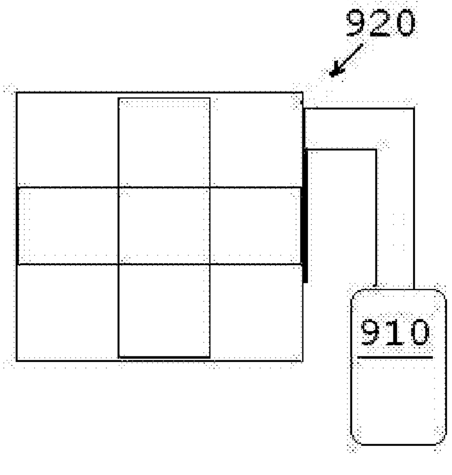

FIG.31

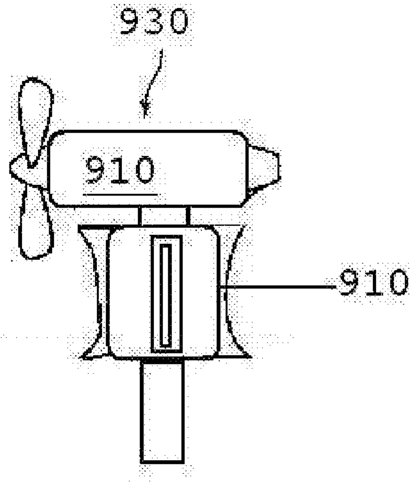

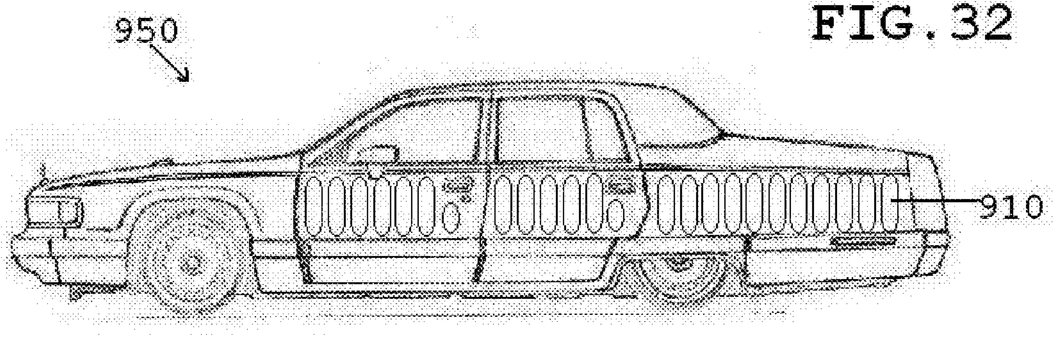

FIG.32

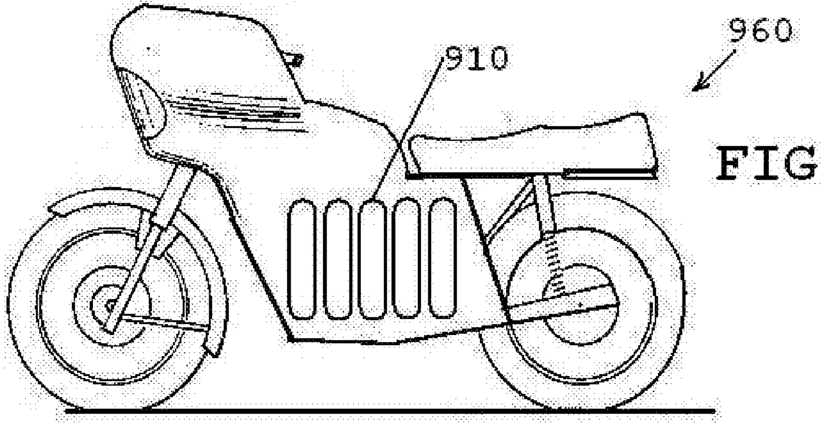

FIG.33

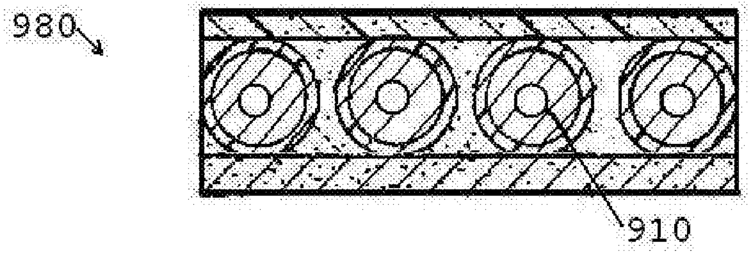

FIG.34

ATMOSPHERIC TRANSDUCTION SYSTEM

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] NONE

BACKGROUND OF THE INVENTION

Field of the Invention

[0002] Self-propelled travel is a type of recreational adventure travel using human powered transport. This includes non-motorized machines such as a bicycle or skateboard. It is in contrast to traveling in a powered vehicle (such as an automobile) as in that case it is the vehicle which powers itself. Self-propelled travel is used to travel short distances or even for much longer distances such as bicycle touring. Self propelled describes something that moves, progresses or acts on its own power without needing outside help. Leonardo da Vinci's 1478 Self-Propelled Car: It was more than 500 years ago, however—sometime around the year 1478 to be more or less specific—when Leonardo drew out his plans for the world's first self-propelled vehicle. Experts originally believed two leaf springs, the simplest form of the spring typically used for automotive suspensions, somehow powered the vehicle. Closer inspection eventually revealed the power came from bigger, coiled springs located in tambours, cylindrical drum-like casings, inside the car's frame. The machine works like a robot or a wind-up toy simply by rotating the wheels opposite of their intended direction, which winds up the springs inside and gives it power. Self propulsion (of a vehicle) provided with its own source of tractive power rather than requiring an external means of propulsion.

[0003] In the middle of the 18th century, Benjamin Franklin's experiments showed that electrical phenomena of the atmosphere were not fundamentally different from those produced in the laboratory. By 1749, Franklin observed lightning to possess almost all the properties observable in electrical machines.

[0004] In July 1750, Franklin hypothesized that electricity could be taken from clouds via a tall metal aerial with a sharp point. Before Franklin could carry out his experiment, in 1752 Thomas-Francois Dalibard erected a 40-foot (12 m) iron rod at Marly-la-Ville, near Paris, drawing sparks from a passing cloud. With ground-insulated aerials, an experimenter could bring a grounded lead with an insulated wax handle close to the aerial, and observe a spark discharge from the aerial to the grounding wire. In May 1752, Dalibard affirmed that Franklin's theory was correct.

Piezoelectric Motor

[0005] A piezoelectric motor or piezo motor is a type of electric motor based upon the change in shape of a piezoelectric material when an electric field is applied. Piezoelectric motors make use of the converse piezoelectric effect whereby the material produces acoustic or ultrasonic vibrations in order to produce a linear or rotary motion. In one mechanism, the elongation in a single plane is used to make a series stretches and position holds, similar to the way a caterpillar moves. A transducer is a device that converts one form of energy to another. Energy types include (but are not limited to) electrical, mechanical, electromagnetic (including light), chemical, acoustic or thermal energy. While the term transducer commonly implies the use of a sensor/detector, any device which converts energy can be considered a transducer. Transducers are widely used in measuring instruments. Piezoelectric materials can also be used to harvest low levels of mechanical energy into electrical energy suitable for powering wireless sensors, low power microprocessors or charging batteries. Rotary Uses include rotating machines such as fans, turbines, drills, the wheels on electric cars, locomotives and conveyor belts. Also, in many vibrating or oscillating machines, an electric motor spins an unbalanced mass, causing the motor (and its mounting structure) to vibrate.

Atmospheric Electricity

[0006] There is always free electricity in the air and in the clouds, which acts by induction on the earth and electromagnetic devices. Experiments have shown that there is always free electricity in the atmosphere, which is sometimes negative and sometimes positive, but most generally positive, and the intensity of this free electricity is greater in the middle of the day than at morning or night and is greater in winter than in summer. In fine weather, the potential increases with altitude at about 30 volts per foot (100 V/m).

Atmospheric Layers

[0007] The electrical conductivity of the atmosphere increases exponentially with altitude. The amplitudes of the electric and magnetic components depend on season, latitude, and height above the sea level. The greater the altitude the more atmospheric electricity abounds. The exosphere is the uppermost layer of the atmosphere and is estimated to be 500 km to 1000 km above the Earth's surface, and its upper boundary at about 10,000 km. The thermosphere (upper atmosphere) is the layer of the Earth's atmosphere directly above the mesosphere and directly below the exosphere. Within this layer, ultraviolet radiation causes ionization. Theories that have been proposed to explain the phenomenon of the polar aurora, but it has been demonstrated by experiments that it is due to currents of positive electricity passing from the higher regions of the atmosphere to the earth.

[0008] The mesosphere (middle atmosphere) is the layer of the Earth's atmosphere that is directly above the stratosphere and directly below the thermosphere. The mesosphere is located about 50-80/85 km above Earth's surface. The stratosphere (middle atmosphere) is a layer of Earth's atmosphere that is stratified in temperature and is situated between about 10 km and 50 km altitude above the surface at moderate latitudes, while at the poles it starts at about 8 km altitude. The stratosphere sits directly above the troposphere and directly below the mesosphere. The troposphere (lower atmosphere) is the densest layer of the atmosphere.

[0009] The planetary boundary layer (PBL), also known as the atmospheric boundary layer (ABL), is the lowest part of the atmosphere and its behavior is directly influenced by its contact with the planetary surface. It is also known as the "exchange layer". (see also: p-n junction.)

[0010] There is a potential gradient at ground level ("Atmosphere ground layer") and this vertical field corresponds to the negative charge in and near the Earth's surface. The negative potential gradient falls rapidly as altitude increases from the ground. Most of this potential gradient is in the first few kilometers. The positive potential gradient rises rapidly as altitude increases from the ground. Volta, over two centuries before the 21st century, discovered with some degree of exactitude that the proportions of the ordinates of the curve or gradient of electric potential increased as the distance from the earth increases, and, more recently, Engel has provided data to calculate the increase (Image to the right).

Drum-Type Generator

[0011] A drum-type homopolar generator has a magnetic field (B) that radiates radially from the center of the drum and induces voltage (V) down the length of the drum. A conducting drum spun from above in the field of a "loudspeaker" type of magnet that has one pole in the center of the drum and the other pole surrounding the drum could use conducting ball bearings at the top and bottom of the drum to pick up the generated current.

Astrophysical Unipolar Inductors

[0012] Unipolar inductors occur in astrophysics where a conductor rotates through a magnetic field, for example, the movement of the highly conductive plasma in a cosmic body's ionosphere through its magnetic field. In their book, Cosmical Electrodynamics, Hannes Alfvén and Carl-Gunne Fälthammar write:

[0013] "Since cosmical clouds of ionized gas are generally magnetized, their motion produces induced electric fields [. . .] For example the motion of the magnetized interplanetary plasma produces electric fields that are essential for the production of aurora and magnetic storms" [. . .]

[0014] " . . . the rotation of a conductor in a magnetic field produces an electric field in the system at rest.

[0015] This phenomenon is well known from laboratory experiments and is usually called 'homopolar' or 'unipolar' induction.

The Faraday Disc

[0016] The homopolar generator was developed first by Michael Faraday during his experiments in 1831. It is frequently called the Faraday disc in his honor. It was the beginning of modern dynamos—that is, electrical generators which operate using a magnetic field. It was very inefficient and was not used as a practical power source, but it showed the possibility of generating electric power using magnetism, and led the way for commutated direct current dynamos and then alternating current alternators.

Boeing 737-800

[0017] The Boeing Fuel Cell Demonstrator Airplane has a Proton Exchange Membrane (PEM) fuel cell/lithium-ion battery hybrid system to power an electric motor, which is coupled to a conventional propeller. The fuel cell provides all power for the cruise phase of flight. During takeoff and climb, the flight segment that requires the most power, the system draws on lightweight lithium-ion batteries.

[0018] The demonstrator aircraft is a Dimona motor glider, built by Diamond Aircraft Industries of Austria, which also carried out structural modifications to the aircraft. With a wing span of 16.3 meters (53.5 feet), the airplane will be able to cruise at approximately 100 kilometers per hour (62 miles per hour) on power from the fuel cell.

[0019] Nikola Tesla explored the wireless transmission of energy through his work with radio and microwaves and his creation of the Tesla coil and the magnifying transmitter. In 1898, Tesla demonstrated his radio-controlled boat, which he was able to control remotely. In the 1930s, Tesla claimed to have invented a particle beam weapon, or, as some called it, a "peace ray." The device was, in theory, capable of generating an intense, targeted beam of energy and sending it across great distances to demolish warplanes, foreign armies, or anything else you'd rather didn't exist.

[0020] "Roy J. Meyers, British Patent Number 1098"

[0021] This invention relates to improvements in apparatus for the production of electrical currents, and the primary object in view is the production of a commercially serviceable electrical current without the employment of mechanical or chemical action. To this end the invention comprises means for producing what I believe to be dynamic electricity from the earth and its ambient elements.

[0022] Edward Leedskalnin: Magnetic Current—

[0023] The Perpetual Motion Holder is primarily a teaching device but it has many functions including an electromagnet, this is easy enough to see; it is a generator—spin a magnet between the coils it will generate electricity; it functions as a transformer; it demonstrates how permanent magnets are made, and is a holder of perpetual motion.

Strategic Defense Initiative

[0024] The Strategic Defense Initiative (SDI) was proposed to use ground and space-based systems to protect the United States from attack by strategic nuclear ballistic missiles.

Description of the Related Art

[0025] The present invention relates to a ball bearing assembly structure, an electromagnetic clutch having the ball bearing assembly structure, and a gas compressor equipped with the electromagnetic clutch.

[0026] When operating the gas compressor, the electromagnet of the electromagnetic clutch is energized to attract or adsorb the follower armature plate to an end surface of the prime-mover pulley and join the prime-mover pulley and the rotor shaft, thereby rotating the rotor shaft.

[0027] The ball bearing of the electromagnetic clutch conventionally has used one having an even number of balls per row. Generally, the ball bearing causes vibration and noise due to rotation. In the case of the ball bearing rotating while undergoing a radial load due to a tension of the belt, vibration and noise considerably occur. Particularly when other vibration and noise levels are lowered during engine idling, the vibration and noise of the ball bearing transmitted to the vehicular compartment is not negligible.

[0028] The inventor has conducted various experiments and discovered that the one factor of high vibration and noise level is an even number of balls of the ball bearing. In the ball bearing having an even number of balls per one row, the balls are in a facing relation to have linear-symmetry arrangement between the inner race and the outer race. The deformation and vibration at a regular particular frequency is caused in the inner and outer races. It is to be considered that the vibration as a source also increases noise.

[0029] The present invention relates generally and in various embodiments to piezoelectric mechanical systems. More specifically, the present invention relates generally and in various embodiments to atmospheric oscillation transducer apparatuses, systems, and methods.

[0030] Although various implementations of the present invention, among many, may be described herein with reference to the specific illustrative embodiments related to particular applications, those skilled in the art will understand that the invention is not in any way intended to be limited to such embodiments and/or applications. Those having ordinary skill in the art and reference to the description of the embodiments herein will recognize additional modifications, applications, and other embodiments falling within the scope of the claimed invention and additional fields in which the present invention may be practiced.

[0031] Digital Radio (also known as Satellite Radio or Satellite Digital Audio Radio Service (SDARS)) is a subscriber-based digital radio service that is broadcast via satellites. Digital radio service provides compact-disc (CD) quality programming that may be digitally transmitted via one or more satellites and/or space stations to one or more Earth-based (terrestrial) digital radio stations, receivers, and/or repeaters. In satellite-based direct-broadcast radio services, digitally-encoded audio program material may be broadcast to terrestrial fixed or mobile digital radio receivers. Fixed receivers may include, for example, stand alone digital radio receivers or digital radio receivers connected via computer networks, including for example, the Internet. Mobile receivers may include, for example, digital radio receivers located in automobiles, aircrafts, watercrafts, and the like.

[0032] Satellite-based digital audio radio services such as SDARS, for example, may be broadcast to one or more digital radio receivers either directly from an orbiting satellite, or indirectly from one or more repeater stations. Such repeater stations may be useful where the digital radio receiver is located in a shielded location or where there is no direct line of sight between the radio and the satellite. In other digital audio radio services systems, the audio programs also may be transmitted in digital form by one or more space stations directly to fixed, mobile, and/or portable radio stations. Such systems may comprise, for example, orbiting satellites, complementary repeating terrestrial transmitters, telemetry, tracking, and control facilities.

[0033] Combinations of mechanical devices U.S. Pat. Nos. 4,019,073, 6,615,968 and atmospheric system interaction are disclosed in U.S. Pat. Nos. 1,119,732, 787,412, 6,902,513 to Nikola Tesla; 28,793 to Charles Vion; and U.S. Pat. No. 1,540,998 to Herman Plauson. Lastly, U.S. Pat. No. 8,102,078 and US2008/0009240. Agnoff discloses an Oscillating watch winder in U.S. Pat. No. 6,543,929, Jennings further discloses an oscillating smart device in application No. 13,572,679.

[0034] As illustrated by a large body of prior art, including the above-noted patents, and a large number of commercial devices, efforts are continuously being made in an attempt to improve helmets, headsets and their methods of fabrication. Nothing in the prior art, however, suggests the present inventive combination of materials and method steps as herein described and claimed. The present invention achieves its purposes, objects and advantages over the prior art through a new, useful and unobvious combination of components and method steps which improve safety, comfort and noise abatement performance.

[0035] Therefore, it is an object of this invention to provide Effectively, the provision of energy such that it meets the needs of the present without compromising the ability of future generations to meet their own needs . . . . Sustainable Energy has two key components: renewable energy and energy efficiency.

[0036] It is still a further objection of this invention to promote Dynamic harmony between equitable availability of energy-intensive goods and services to all people and the preservation of the earth for future generations." And, "the solution will lie in finding sustainable energy sources and more efficient means of converting and utilizing energy.

[0037] It is a further object of the present invention to produce Green Power Energy: is energy that can be extracted, generated, and/or consumed without any significant negative impact to the environment, green power; as electricity produced from solar, wind, geothermal, biogas, biomass, and low-impact small hydroelectric sources.

[0038] Thus, there is a need for a clean energy system that uses atmospheric electricity.

Prior Art

[0039] Quartz crystals have been in regular use for many years to give an accurate frequency for all radio transmitters, radio receivers and computers. Their accuracy comes from an amazing set of coincidences: Quartz—which is silicon dioxide like most sand—is unaffected by most solvents and remains crystalline to hundreds of degrees Fahrenheit. The property that makes it an electronic miracle is the fact that, when compressed or bent, it generates a charge or voltage on its surface. This is a fairly common phenomenon called the Piezoelectric effect. In the same way, if a voltage is applied, quartz will bend or change its shape very slightly.

[0040] If a bell were shaped by grinding a single crystal of quartz, it would ring for minutes after being tapped. Almost no energy is lost in the material. A quartz bell—if shaped in the right direction to the crystalline axis—will have an oscillating voltage on its surface, and the rate of oscillation is unaffected by temperature. If the surface voltage on the crystal is picked off with plated electrodes and amplified by a transistor or integrated circuit, it can be re-applied to the bell to keep it ringing.

[0041] The electronics of the watch initially amplifies noise at the crystal frequency. This builds or regenerates into oscillation—it starts the crystal ringing. The output of the watch crystal oscillator is then converted to pulses suitable for the digital circuits.

Polymers

[0042] Polyvinylidene fluoride (PVDF): PVDF exhibits piezoelectricity several times greater than quartz. Unlike ceramics, where the crystal structure of the material creates the piezoelectric effect, in polymers the intertwined long-chain molecules attract and repel each other when an electric field is applied.

Near Space

[0043] Solar particles become trapped within the Earth's magnetic field and form radiation belts. The Van Allen radiation belt is a torus of energetic charged particles (i.e. a plasma) around Earth, trapped by Earth's magnetic field.

[0044] At elevations above the clouds, atmospheric electricity forms a continuous and distinct element (called the electrosphere) in which the Earth is surrounded. The electrosphere layer (from tens of kilometers above the surface of the earth to the ionosphere) has a high electrical conductivity and is essentially at a constant electric potential. The ionosphere is the inner edge of the magnetosphere and is the part of the atmosphere that is ionized by solar radiation. (Photoionisation is a physical process in which a photon is incident on an atom, ion or molecule, resulting in the ejection of one or more electrons.)

Advantages/Disadvantages

[0045] Energy in electronic elements: Electric potential energy, or electrostatic potential energy, is a potential energy (measured in joules) that results from conservative Coulomb forces and is associated with the configuration of a particular set of point charges within a defined system. The term "electric potential energy" is used to describe the potential energy in systems with time-variant electric fields, while the term "electrostatic potential energy" is used to describe the potential energy in systems with time-invariant electric fields.

[0046] Capacitance is the ability of a body to store an electrical charge. Any body or structure that is capable of being charged, either with static electricity or by an electric current, exhibits capacitance. A common form of energy storage device is a parallel-plate capacitor. In a parallel plate capacitor, capacitance is directly proportional to the surface area of the conductor plates and inversely proportional to the separation distance between the plates. If the charges on the plates are +q and −q, and V gives the voltage between the plates, then the capacitance C is given by

$$C=q/V.$$

[0047] The capacitance is a function only of the physical dimensions (geometry) of the conductors and the permittivity of the dielectric. It is independent of the potential difference between the conductors and the total charge on them.

[0048] Piezoelectricity is the combined effect of the electrical behavior of the material:

$$D=\in E$$

where D is the electric charge density displacement (electric displacement), $\in$ is permittivity and E is electric field strength, and

Hooke's Law: $S=sT$

where S is strain, s is compliance and T is stress.

[0049] Polyvinylidene fluoride, or polyvinylidene difluoride (PVDF) is a highly non-reactive and pure thermoplastic fluoropolymer produced by the polymerization of vinylidene difluoride. PVDF is a specialty plastic material in the fluoropolymer family; it is used generally in applications requiring the highest purity, strength, and resistance to solvents, acids, bases and heat and low smoke generation during a fire event. Compared to other fluoropolymers, it has an easier melt process because of its relatively low melting point of around 177° C.

[0050] It has a low density (1.78) and low cost compared to the other fluoropolymers. It is available as piping products, sheet, tubing, films, plate and an insulator for premium wire. It can be injected, molded or welded and is commonly used in the chemical, semiconductor, medical and defense industries, as well as in lithium ion batteries. It is also available as a crosslinked closed cell foam, used increasingly in aviation and aerospace. PVDF has a glass transition temperature (Tg) of about −35° C. and is typically 50-60% crystalline. To give the material its piezoelectric properties, it is mechanically stretched to orient the molecular chains and then poled under tension. PVDF exists in several forms: alpha (TGTG'), beta (TTTT), and gamma (TTTGTTTG') phases, depending on the chain conformations as trans (T) or gauche (G) linkages. When poled, PVDF is a ferroelectric polymer, exhibiting efficient piezoelectric and pyroelectric properties. These characteristics make it useful in sensor and battery applications. Thin films of PVDF are used in some newer thermal camera sensors.

[0051] Copolymers: Copolymers of PVDF are also used in piezoelectric and electrostrictive applications. One of the most commonly-used copolymers is P(VDF-trifluoroethylene), usually available in ratios of about 50:50 wt % and 65:35 wt % (equivalent to about 56:44 mol % and 70:30 mol %). Another one is P(VDF-tetrafluoroethylene). They improve the piezoelectric response by improving the crystallinity of the material.

[0052] While the copolymers' unit structures are less polar than that of pure PVDF, the copolymers typically have a much higher crystallinity. This results in a larger piezoelectric response: d33 values for P(VDF-TrFE) have been recorded to be as high as −38 pC/N versus −33 pC/N in pure PVDF.

[0053] Applications:

[0054] The piezoelectric properties of PVDF are used to advantage to manufacture tactile sensor arrays, inexpensive strain gauges and lightweight audio transducers. Piezoelectric panels made of PVDF are used on the Venetia Burney Student Dust Counter, a scientific instrument of the New Horizons space probe that measures dust density in the outer solar system. PVDF is the standard binder material used in the production of composite electrodes for lithium ion batteries. 1-2% weight solution of PVDF dissolved in N-Methyl-2-pyrrolidone (NMP) is mixed with an active lithium storage material such as graphite, silicon, tin, $LiCoO_2$, $LiMn_2O_4$, or $LiFePO_4$ and a conductive additive such as carbon black or carbon nanofibers. This slurry is cast onto a metallic current collector and the NMP is evaporated to form a composite or paste electrode. PVDF is used because it is chemically inert over the potential range used and does not react with the electrolyte or lithium. Piezoelectric elements can be used in laser mirror alignment, where their ability to move a large mass (the mirror mount) over microscopic distances is exploited to electronically align some laser mirrors. By precisely controlling the distance between mirrors, the laser electronics can accurately maintain optical conditions inside the laser cavity to optimize the beam output. Piezoelectric sensors especially are used with high frequency sound in ultrasonic transducers for medical imaging and also industrial nondestructive testing (NDT).

[0055] For many sensing techniques, the sensor can act as both a sensor and an actuator—often the term transducer is preferred when the device acts in this dual capacity, but most piezo devices have this property of reversibility whether it is used or not. Ultrasonic transducers, for example, can inject ultrasound waves into the body, receive the returned wave, and convert it to an electrical signal (a voltage). Most medical ultrasound transducers are piezoelectric.

[0056] Advantageously,

[0057] Sustainable energy is the sustainable provision of energy that meets the needs of the present without compromising the ability of future generations to meet their needs. Technologies that promote sustainable energy include renewable energy sources, such as hydroelectricity, solar energy, wind energy, wave power, geothermal energy, and tidal power, and also technologies designed to improve energy efficiency.

[0058] This sequence of oscillations causes the rotor within the watch to spin rapidly thereby winding the watch in a manner closely simulating the spinning of the rotor that occurs during normal winding of the watch when the watch is worn by a user. Due to the forces that are exerted, the rotor spins around the watch shaft during the oscillations, as opposed to the partial rotation observed in prior art mechanisms. Therefore, the time required to wind the watch, and the energy required, is substantially reduced. Moreover, since the rotor is spinning about the shaft, as opposed to merely being held in a downward position while the watch is rotated, winding more closely approximating the design mechanism is achieved, thereby putting less wear on the watch.

[0059] This invention relates to satellite communications systems using multiple spot beams from a geosynchronous earth orbit satellite to provide selective coverage of the continental United States and, more particularly, relates to a system having a satellite receiving hub in every spot beam that allows for asynchronous communications between each hub and the satellite for maximizing frequency re-use.

[0060] These purposes, objects and advantages should be construed as merely illustrative of some of the more prominent features and applications of the present invention. Many other beneficial results can be obtained by applying the disclosed invention in a different manner or by modifying the invention within the scope of the disclosure. Accordingly, other purposes, objects and advantages as well as a fuller understanding of the invention may be had by referring to the summary herein mentioned and detailed description describing the preferred embodiments of the invention, in addition to the scope of the invention, as defined by the claims, taken in conjunction with the accompanying drawings.

SUMMARY OF THE INVENTION

[0061] In one general respect, an embodiment of the present invention is directed to a system. The system includes a Power Frequency broadcast station, a receiver, and a network. The Power Frequency broadcast station includes a transmitter and a server. The receiver is in communication with the Power Frequency broadcast transmitter and also includes a user interface for receiving user input commands comprising a request for information from the Power Frequency broadcast station.

[0062] The receiver is configured to establish a two-way communication path between the receiver and the Power Frequency broadcast transmitter. The network is in communication with the server and the receiver for exchanging information therebetween. The request for information is provided to the server via the network and the server is configured to receive the request and transmit a response message to the receiver in accordance with the request.

Continuously Outboard Recharged Electric Vehicle (COREY)

[0063] Given suitable infrastructure, permissions and vehicles, BEVs (battery electric vehicles) can be recharged while the user drives. The BEV establishes contact with an electrified rail, plate or overhead wires on the highway via an attached conducting wheel or other similar mechanism (see Conduit current collection). The BEV's batteries are recharged by this process—on the highway—and can then be used normally on other roads until the battery is discharged. Some of battery-electric locomotives used for maintenance trains on the London Underground are capable of this mode of operation. Power is picked up from the electrified rails where possible, switching to battery power where the electricity supply is disconnected.

[0064] The present invention is directed to overcome the disadvantages of the prior art. The invention is a mechanism which taps into the naturally occurring static electricity in the atmosphere. Whereas heretofore, the attempt to garner electricity from the atmosphere has focused exclusively on capturing lightning, the present invention syphons off the static electricity which is generated from any agitated air and voids lightning.

[0065] Lightning is only the final discharge of the static electricity, whether that lightning is intra-cloud lightning, cloud-to-ground lightning, or inter-cloud lightning. Other types of final discharges are known as heat lightning, summer lightning, sheet lightning, ribbon lightning, silent lightning, ball lightning, bead lightning, elves, jets, and sprites. Well before these discharges are observed, as the atmosphere becomes agitated by wind or thermal, static electricity is being generated. The present invention recognizes that this static electricity is being formed and creates a mechanism to capture it.

[0066] In the preferred embodiment, a sensor array is used to monitor the activities both at the base unit (such as electrical flow within the conductor) and in the surrounding locale. A sensor monitoring the electrical flow (i.e. voltage and/or current) within the conductor is used to monitor the electrical activity within the conductor.

[0067] In the preferred embodiment, a lightning sensor monitors for lightning activity within the locale. As noted earlier, the electrical characteristic of lightning is so extreme that ideally this discharge is avoided as it might damage the mechanism of this invention. The sensor array is utilized by a controller, such as microprocessor, programmed to operate the mechanism as outlined herein.

[0068] The controller operates the winch motor to extend or withdraw the conductive line and by extension the altitude of the balloon. The controller is programmed to operate the winch by monitoring the electrical characteristics of the conductor and adjusting the balloon's altitude to maintain these characteristics within the conductor within a preset range.

[0069] This preset range is established either in the base programming of the controller or is established by an operator of the system. As example, by controlling the amount of current being withdrawn from the atmosphere, the mechanism operates within a safe range and also provides a relatively stable current flow from which a variety of activities can take place (such as DC-AC conversion).

[0070] The controller also utilizes the lightning sensor to protect the mechanism from a lightning strike. Should lightning be detected within a pre-determined range (as established by the software or defined by an operator), then the balloon is pulled down to minimize the risk of damage from a lightning strike.

[0071] Another aspect of the invention relates to the electrical system which accepts the fluctuating atmospheric charge and changes it into an acceptable configuration for either the desired load or for the existing power grid.

[0072] Power grids in the United States operate with a frequency of 60 hertz in an alternating current arrangement. While this basic configuration seems to be universally accepted, the voltage within the grid varies dramatically, such as 15 kv, 34 kv, 69 kv, and even 112 kv.

[0073] Each atmospheric generator is placed proximate to or within easy access to a specific grid; this establishes the required electrical output configuration (i.e. that which is accepted by the power grid). As example, one of the atmospheric electrical collector units as described above collects the atmospheric electrical power as direct current and then supplies the appropriate power grid a specific flow (as example, AC, 60 hertz, at 69 kv).

[0074] The difficulty lies the fact that the DC current being garnered from the atmosphere varies depending on the actual agitation being generated in the atmosphere. This means that the source of DC current is fluctuating.

[0075] The present invention uses a monitoring system which checks the input DC voltage. Depending on the actual voltage being received, the appropriate converter is connected to the input DC voltage so that the desired output is obtained.

[0076] As example, suppose the DC input voltage is 1500 volts, the monitoring system, sensing this input, closes the switch connecting the DC voltage to a converter which accepts DC voltage in the range of 1000-2000 volts which then delivers an AC, 60 hertz 69 kv signal to the power grid. If the DC input voltage increases to 2100 volts, then the monitoring system opens the switch to the first converter (1000-2000 volts) and closes the switch to a second converter (such as 2000-4000 volts) to deliver the desired output of AC (60 hertz, 69 kv) for the power grid.

[0077] In this manner, regardless of the fluctuating input DC voltage, the electrical grid is supplied with a fully configured electrical input conforming to the needs of that specific electrical grid.

[0078] Another aspect of the present invention is the use of a tower or permanent structure instead of an aircraft. In this embodiment, the building or tower is electrically isolated from the ground and a rod (similar to a lightning rod) is extended into the atmosphere. The rod collects the atmospheric charge which is conveyed via an electrical conduit (ideally insulated) where the collected DC charge is reconfigured to meets the need of the locale.

[0079] In this context, for one embodiment of the invention, a tower is placed onto the top of a building. The tower is electrically isolated from the building using such mechanisms well known to those of ordinary skill in the art such as rubber mats. A rod ideally extends from the top of the tower to facilitate the collection of the DC electrical energy.

[0080] Piezoelectricity is a material property that is manifested when voltage is produced by applying mechanical forces, and vice versa, the effect has been described as direct and converse. Piezoelectricity has been described as coupling between a quasi-static electric field and dynamic mechanical motion.

[0081] A piezoelectricity converter mechanism such as described above, is connected to the tower to flow the DC electricity to a converter which modifies the DC current for the specific application. In one application, the DC current is converted to the electrical needs of the building, thereby providing at least some of the electrical requirements of the building itself.

[0082] As noted earlier, the dynamic converter system of the present invention allows a power generator to address a variable voltage in an efficient manner. This makes the dynamic converter system ideal for a variety of alternative energy sources such as the above described atmospheric electrical generator and other alternative energy sources such as wind and wave powered systems. In these systems, the energy being generated must be converted to a proper electrical configuration for a identified load. This may be a particular motor or connection to the power grid which act as a load to the power generating mechanism.

[0083] For these energy generating systems, the converter assembly of this invention utilizes multiple converters. Each converter is configured to accept a unique range of voltages and from these voltages, create the desired electrical configuration. By using multiple converters, a full range is available, from a minimum voltage input to a maximum voltage input.

[0084] The present invention can include systems and methods for integrating sensors for tracking atmospheric transducer system performance metrics into media devices and accessories therefor, thereby reducing or eliminating the need for additional independent monitoring devices. In one embodiment of the present invention.

[0085] It also is known to provide such transducers with connectors to allow their rechargeable batteries to be charged. In some cases, the connector is a Universal Serial Bus (USB) connector, allowing the transducer to be charged by plugging it into the USB port of a computer, grid circuit or other device.

[0086] These and other objects and advantages of the invention will appear more clearly from the following description in which the preferred embodiment of the invention has been set forth in conjunction with the drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0087] FIG. **1** Prior Art depicts perspective view of Electric currents created in sunward ionosphere;

[0088] FIG. **2** is a box flow chart of the propulsion cycle systems present invention;

[0089] FIG. **3** is an top view of the piezoelectric disk cylinder orb of the present invention;

[0090] FIG. **4** is a longitudinal sectional view showing an embodiment of this invention;

[0091] FIG. **5** is a top view of the (ATS) slip rotor piezoelectric chamber of the present invention;

[0092] FIG. **6** is an alternate top view of the piezoelectric cylinder orb of the present invention;

[0093] FIG. **7-10** elevation views of spine piezo stacks embodiments of the present invention;

[0094] FIG. **11** illustrates one embodiment of a power service (ATS) system architecture;

[0095] FIG. **12** is a block diagram of the ATS charge schematics systems of the present invention;

[0096] FIG. **13** is a block diagram of an ATS charge and recycle schematics of the present invention;

[0097] FIG. **14** is a view of the piezoelectric ball race cylinder of the present invention;

[0098] FIG. **15** is a view of the piezoelectric housing and gear of the present invention;

[0099] FIG. **16** is a view of a multiple piezoelectric ball race cylinder of the present invention;

[0100] FIG. **17** is another embodiment of the piezoelectric stack ball of the present invention;

[0101] FIGS. **18 & 19** are antenna rod transmit device embodiments of the (ATS) present invention;

[0102] FIG. **20** illustrates substation embodiment of the power service (ATS) system architecture;

[0103] FIG. **21** illustrates a home embodiment of the power service (ATS) system architecture;

[0104] FIG. **22** illustrates a window embodiment of the power service (ATS) system architecture;

[0105] FIG. **23** is (ATS) shock device in accordance with an embodiment of the present invention;

[0106] FIG. **24** is (ATS) shock device in accordance with an embodiment of the present invention;

[0107] FIG. **25** is (ATS) motorcycle device embodiment of the present invention;

[0108] FIG. **26** is (ATS) axle device in accordance with an embodiment of the present invention;

[0109] FIG. **27** is (ATS) train device in accordance with an embodiment of the present invention;

[0110] FIG. **28** is (ATS) plane device in accordance with an embodiment of the present invention;

[0111] FIG. **29** is (ATS) boat device in accordance with an embodiment of the present invention;

[0112] FIG. **30** is (ATS) solar device in accordance with an embodiment of the present invention;

[0113] FIG. **31** a (ATS) turbine device in accordance with an embodiment of the present invention;

[0114] FIG. **32** is (ATS) auto body panel device embodiment of the present invention;

[0115] FIG. **33** is (ATS) motorcycle body fairing panel device in of the present invention; and

[0116] FIG. **34** is (ATS) cross section view of body embed panel device of the present invention.

DETAILED DESCRIPTION

[0117] This sequence of oscillations causes the conductive rotor within piezoelectric molded housing device to spin rapidly thereby winding the mechanism in a manner closely simulating the spinning of the conductive rotor that occurs during normal electric activity when the device is activated. Due to the forces that are exerted, the conductive rotor spins around the piezoelectric cylinder device shaft during the oscillations, as opposed to the partial rotation observed in prior art mechanisms. Therefore, the time required to charge the Atmospheric Transduction System (ATS) device, and the energy required, is substantially reduced. Moreover, since the rotor is spinning about the shaft, as opposed to merely being held in a downward position while the ATS device is rotated, recharging more closely approximating the design mechanism is achieved, thereby putting less wear on the ATS chamber device Innovative Piezoelectric housing and ball bearing, coupler and book spine stacks. There relies the notion of negative ground electricity and positive aerial electricity which is in abundance. This invention substantiates land vehicles recycle recharge by reverse oscillation. Aerial vehicles recycle recharge by forward oscillation in accumulation of environmental positive and negative electricity. These aforementioned activities are integral or synchronous with frequency.

[0118] Prior Art FIG. **1**

[0119] Electric currents created in sunward ionosphere. FIG. **2** is a recycle box flow chart **49** of a self propulsion unit consisting of a battery **27**, engine **29** and piezoelectric transducer **33** unit. FIG. **3** is a multiple disk **31** load part **51** about a 360 degree cylinder **45** and shaft **44**. FIG. **4** is a piezoelectric molded device **100** housing **110** containing a ball bearing race **104**, disk stacks capacitor **102** and conductive rotor **46**. FIG. **5** is a piezoelectric chamber **120** configuration including a counterweights **50**, shaft **44**, conductive slip rotor **106**. FIG. **6** is an alternate part **200** chamber cylinder **201** embodiment containing piezoelectric spine disk stack capacitors **102**. FIG. **7-10** are variations of spine **107**,**108**,**109**,**110** case piezoelectric stack capacitor **102** with plate and ball heads. FIG. **11** Various embodiments of the present invention, among others, will now be described with reference to the accompanying drawings. Accordingly, FIG. **11** illustrates an embodiment of a Atmospheric Transduction (ATS) System **300** architecture. The system **300** may include, for example, a satellite broadcast station **318** that transmits signals **333** containing frequency content from a geostationary satellite **312** by way of airplane antenna **326**. In turn, the satellite **312** transmits line-of-sight (LOS) signals **333** to one or more ATS terrestrial frequency power receiver farms **314**. The system **300** also may include one or more terrestrial repeaters **316** which receive and retransmit the satellite signals **333** as repeater signals **323** to facilitate reliable reception in geographic areas where LOS reception from the satellite **312** is obstructed by tall buildings, hills, tunnels, and other similar impediments to the signals **333**. The ATS receivers **314** maybe designed to receive one or more signals **333** from the satellite **312** and/or from the terrestrial repeater transceiver **316**. In operation, such ATS receivers **314** may receive both the satellite signals **333** and the repeater signals **323**. The receivers **314** also may be located in mobile environments **320**, **321**, **322** which include, but are not limited to, land vehicles **321**, **322**, aircraft **320**, watercraft **900**, and handheld devices, among others. The receivers **314** also may be fixed in stationary units for residential use (e.g., home **325**,**750** entertainment, etc.) or commercial **314**, **328**,**360** use (e.g., business **314**, office **700**, security **328**, etc.). The power frequency broadcast station **318** also may be in communication with a grid network **342**. Two-way communication between the ATS receivers **314** and the power frequency broadcast station **318** may occur via the network **342**. Furthermore, information feedback from the power frequency broadcast station **318** may be transmitted to the ATS receiver **314** both by way of the network **342** as well as via the satellite **312**. Information also may be transmitted to the power broadcast station **318** wirelessly via a wireless network **342**,**707** by way of transducer tower **360**.

[0120] Further disclosed in FIG. **12** and FIG. **13** (in block diagrams **400**, **500**) are electrical schematics for handling the static charge from the atmosphere. By maintaining the voltage being collected in a prescribed range, an electrical conversion system is easily designed. While FIGS. **12**, and **13** illustrate some electrical configurations, those of ordinary skill in the art readily recognize a variety of other configurations which will serve the same function.

[0121] Referencing FIG. **12**, Direct Current In (DC IN) **401** is buffered by a gang of capacitors **410** before being communicated to a DC/AC converter **420**. The DC/AC converter **420** converts the direct current into a an alternating current suitable for placement over an existing electrical grid **430** such as normally found from a power-plant. FIG. **12** Also illustrates an electrical arrangement suitable for use in charging a battery **440**. DC IN **401** is buffered by capacitor **410** bank before entering into a step down transformer **435**. Step down transformer **435** reduces the voltage so that the voltage can safely be introduced into battery **440** which is connected to ground **450** at the battery's other pole. In FIG. **12**, DC IN **401** is fed into an adjustable rheostat **460** which is controlled by the controller **465** so that the DC OUT **470** falls within a computer **475** monitored and sensor **480** specified range. FIG. **13** Hypothetically, unused energy may be recycled current **501** and/or recaptured by reversing the oscillated spin rotation of devices **51**, **100**, **120** with the use of a rectifier **502** and Step up transformer **503**, returned to grid **430** capacitor **410**. This theory lends itself to the concept of positive and negative frequency. FIG. **14-17** Self charging propulsion embodiment of the invention where Da Vinci's ball race **510**,**512** is combined with a disk cylinder **45** and **513** stack balls **525**, conductive rotor **46**, **511** tooth gear **515** and piezoelectric molded housing **555**. FIGS. **18** & **19** Improvement structures Franklin's lightening rod **600**,**603** and a molded vibration transducer **601** quasi replicating Vion's tubes and piezoelectric spine stacks capacitor **602** improving Tesla's Atmospheric transmit device. FIG. **20** is an atmospheric receiving building sub station **700** where energy is consumed and excess rendered to the grid **707** by conductive rotor **702** transducer **710**, tower transducer **703**, antenna rod **701** and transducer windows **704**.

[0122] FIG. **12** and FIG. **2** flowchart illustration also includes battery **440**. Battery **440** may provide electrical power to components of ATS devices within FIG. **11**. Charging circuitry may also be provided to charge battery **440** when an external power supply is connected to an ATS device **100**. FIG. **14-17** eliminates one or more steps by presenting a self charge retaining transducer **510-513** may be configured with an accelerator sensor **480** controller **465** and gears **515**, provide reciprocal power incorporated within piezoelectric molded and ceramic housing **555** along with stack balls **525** and cylinder **545**. This assembly more resembles a motor by characteristics given power with application.

[0123] In operation, as illustrated in FIGS. **2-20** and FIGS. **21-34**, is periodically energized by movement to rotate Orb in either a clockwise or counterclockwise direction. The length of time or activity is energized, and the length of time between the period when the capacitor **31**, **102** battery is energized, will depend on the particular ATS device design. As the Orb rotates, the outer end of the disk moves along a 360° circular pathway to push against with forward and rearwardly spinning. Upon engagement of the Orb, ATS device disk is rotated until carried to the apex or top of the circular pathway. Upon reaching the apex, the gravitational and vibrational force or counterweight **50** promotes additional oscillation. ATS device movement rapidly rotate on Orb at a rotational speed greater than the speed of rotation of Orb. Counterweight **50** is then carried beyond the bottom or lowest point of the pathway by its momentum to a point near the apex on the opposite side of the pathway. The cycle is repeated through multiple increasing oscillations of the ATS device until counterweight **50** stops at the bottom position, or until once again engages to again move counterweight to the top of its circular pathway.

[0124] FIG. **2** is a flow chart showing generation of energy using a rotor according to one or more of the above-described embodiments. First, battery **27** starts the engine **29** and/or mobile transducer **33** is oscillated. In response to this acceleration, forces are imposed on one or more rotation piezoelectric devices. In response to those forces, the piezoelectric devices output electrical energy, which energy is extracted at a power controller **465**. The power controller **465** sensor **480** then makes this energy available to recharge a capacitor **410** battery **440** and/or to electronic components of the mobile terminal. Although FIG. **2** shows a serial flow of events, it is to be appreciated that the events of blocks **33**, **27** and **29** occur substantially instantaneously upon acceleration of the mobile terminal.

Preferred Alternate Embodiments

[0125] The present invention (ATS) device in accordance with an embodiment of the present invention overcomes the foregoing problem in the conventional art and provides an electro energy vibration and alternative to gas, oil or fossil fuel consumption in FIG. **21** homes **750** transducer **755**, FIG. **28** airplane **880** transducer **801**, FIG. **27** train **860** transducer **801**, FIG. **32** auto **950** transducer **910**, FIG. **25** and FIG. **33** motorcycles **960**,**850** transducers **799**, **802**, **810**, **910** and FIG. **29** boat **900** transducers **901**,**910**.

[0126] In order to solve the foregoing problems in the conventional art, the present invention provides an electro transducer having a ball bearing assembly which is compatible with FIG. **26** axle shaft **44** transducer **810** wheel **870** assembly. In order to solve the foregoing problems in the conventional art, the present invention provides an electro transducer having an FIG. **34** housing panel **980** transducer **910** assembly which is compatible with an exterior body assembly. Thereby, an Atmospheric transducer device may be shock **800** integrated as within the FIG. **23** coupler **802**, and FIG. **24** ladder **799**, (vertical friction) piezoelectric absorbers assemblies. In order to solve the foregoing problems in the conventional art, the present invention provides an electro transducer having a capable flat assembly which is compatible with FIG. **22** window **704** disk **31** plate transducer **714**, FIG. **30** solar panel **920** transducer **910**, FIG. **31** wind turbine **930** transducer **910** assembly. This invention provided FIG. **3** piezoelectric plate disks **31**, FIG. **7-10** spine piezoelectric disks stacks **107-110** capacitor **102**, FIG. **17** piezoelectric stack balls **525** capacitor, the arrangement of balls **525** will not be in a facing relation. The deformation in the inner and outer races during rotation of the ball bearing while undergoing a radial load is made irregular and complicated. The spine eliminates the deformation and vibration level increased by a combination of ceramic and molded piezoelectric materials at a regular predetermined frequency thereby multiplying the level of vibration and noise reduction.

[0127] Certain modifications and improvements will occur to those skilled in the art upon a reading of the foregoing description. It should be understood that all such modifications and improvements have been deleted herein for the sake of conciseness and readability but are properly within the scope of the following claims.

[0128] Thus it is seen that an atmospheric transducer device may be integrated and/or provided. It will be understood that the foregoing is only illustrative of the principles of the invention, and that various modifications can be made by those skilled in the art without departing from the scope and spirit of the invention, and the present invention is limited only by the claims that follow.

1-3. (canceled)

4. An atmospheric transduction system comprising:

means for recycle recharge by oscillation and frequency in accumulation of environmental positive and negative electricity, maintaining the voltage being collected in a prescribed range, providing an electrical conversion broadcast network; and

means for collecting the atmospheric electrical power as direct current and then supplies the appropriate power grid, transceiver and capacitates a charge; and

means for self charging propulsion provides motor characteristics, and frequency engine.

5. The atmospheric transduction system in accordance with claim **1**, wherein said means for recycle recharge by oscillation and frequency in accumulation of environmental positive and negative electricity, maintaining the voltage being collected in a prescribed range, an electrical conversion broadcast network comprises piezoelectric transducer molded device(s), rotor, stack ball bearing, coupler, book spine stacks and antenna rod stacks;

wherein said means for collecting the atmospheric electrical power as direct current and then supplies the appropriate power grid and capacitates a charge comprises a piezoelectric network, piezoelectric grid, piezoelectric spine stack antenna, piezoelectric transmitters, piezoelectric receivers, piezoelectric devices, piezoelectric cylinders and orbs, power frequency broadcast;

wherein said means for charging providing motor characteristics, said power frequency engine comprises a sensor accelerator, rotor/gear, battery/capacitor, antenna, controller, for propulsion.

6. An atmospheric transduction system comprising:

a piezoelectric transducer molded device(s), rotor, spine ball bearing, coupler and book spine stack transducers, for recycle recharge by oscillation and frequency in accumulation of environmental positive and negative electricity, maintaining the voltage being collected in a prescribed range, providing an electrical conversion broadcast server network; and

a piezoelectric network, piezoelectric grid, piezoelectric spine stack antenna, piezoelectric transmitters, piezoelectric receivers, piezoelectric devices, piezoelectric cylinders and disc orbs, stack rods, power frequency broadcast, for collecting the atmospheric electrical power as direct current and then supplies the appropriate power grid, transceiver and capacitates a charge; and

a sensor accelerator, rotor/gear, battery/capacitor, controller, antenna, means for self charge propulsion providing motor characteristics, thereby an engine of power frequency.

* * * * *

WIKIPEDIA

Atmospheric physics

Atmospheric physics is the application of physics to the study of the atmosphere. Atmospheric physicists attempt to model Earth's atmosphere and the atmospheres of the other planets using fluid flow equations, chemical models, radiation budget, and energy transfer processes in the atmosphere (as well as how these tie into other systems such as the oceans). In order to model weather systems, atmospheric physicists employ elements of scattering theory, wave propagation models, cloud physics, statistical mechanics and spatial statistics which are highly mathematical and related to physics. It has close links to meteorology and climatology and also covers the design and construction of instruments for studying the atmosphere and the interpretation of the data they provide, including remote sensing instruments. At the dawn of the space age and the introduction of sounding rockets, aeronomy became a subdiscipline concerning the upper layers of the atmosphere, where dissociation and ionization are important.

Contents

Remote sensing
Radiation
Cloud physics
Atmospheric electricity
Atmospheric tide
Aeronomy
Centers of research
See also
References
Bibliography

Remote sensing

Remote sensing is the small or large-scale acquisition of information of an object or phenomenon, by the use of either recording or real-time sensing device(s) that is not in physical or intimate contact with the object (such as by way of aircraft, spacecraft, satellite, buoy, or ship). In practice, remote sensing is the stand-off collection through the use of a variety of devices for gathering information on a given object or area which gives more information than sensors at individual sites might convey.[1] Thus, Earth observation or weather satellite collection platforms, ocean and atmospheric observing weather buoy platforms, monitoring of a pregnancy via ultrasound, Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and space probes are all examples of remote sensing. In modern usage, the term generally refers to the use of imaging sensor technologies including but not limited to the use of instruments aboard aircraft and spacecraft, and is distinct from other imaging-related fields such as medical imaging.

There are two kinds of remote sensing. Passive sensors detect natural radiation that is emitted or reflected by the object or surrounding area being observed. Reflected sunlight is the most common source of radiation measured by passive sensors. Examples of passive remote sensors include film photography, infra-red, charge-coupled devices, and radiometers. Active collection, on the other hand, emits energy in order to scan objects and areas whereupon a sensor then detects and

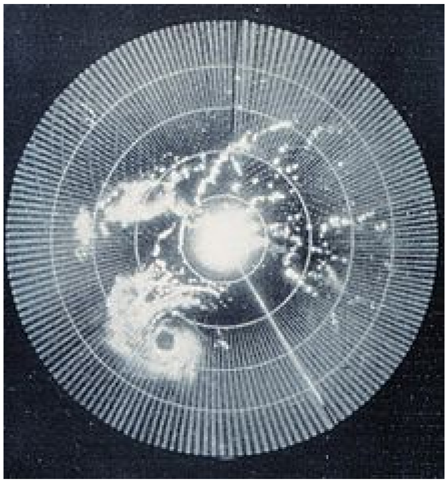

Brightness can indicate reflectivity as in this 1960 weather radar image (of Hurricane Abby). The radar's frequency, pulse form, and antenna largely determine what it can observe.

measures the radiation that is reflected or backscattered from the target. radar, lidar, and SODAR are examples of active remote sensing techniques used in atmospheric physics where the time delay between emission and return is measured, establishing the location, height, speed and direction of an object.[2]

Remote sensing makes it possible to collect data on dangerous or inaccessible areas. Remote sensing applications include monitoring deforestation in areas such as the Amazon Basin, the effects of climate change on glaciers and Arctic and Antarctic regions, and depth sounding of coastal and ocean depths. Military collection during the cold war made use of stand-off collection of data about dangerous border areas. Remote sensing also replaces costly and slow data collection on the ground, ensuring in the process that areas or objects are not disturbed.

Orbital platforms collect and transmit data from different parts of the electromagnetic spectrum, which in conjunction with larger scale aerial or ground-based sensing and analysis, provides researchers with enough information to monitor trends such as El Niño and other natural long and short term phenomena. Other uses include different areas of the earth sciences such as natural resource management, agricultural fields such as land usage and conservation, and national security and overhead, ground-based and stand-off collection on border areas.[3]

Radiation

Atmospheric physicists typically divide radiation into solar radiation (emitted by the sun) and terrestrial radiation (emitted by Earth's surface and atmosphere).

Solar radiation contains variety of wavelengths. Visible light has wavelengths between 0.4 and 0.7 micrometers.[4] Shorter wavelengths are known as the ultraviolet (UV) part of the spectrum, while longer wavelengths are grouped into the infrared portion of the spectrum.[5] Ozone is most effective in absorbing radiation around 0.25 micrometers,[6] where UV-c rays lie in the spectrum. This increases the temperature of the nearby stratosphere. Snow reflects 88% of UV rays,[6] while sand reflects 12%, and water reflects only 4% of incoming UV radiation.[6] The more glancing the angle is between the atmosphere and the sun's rays, the more likely that energy will be reflected or absorbed by the atmosphere.[7]

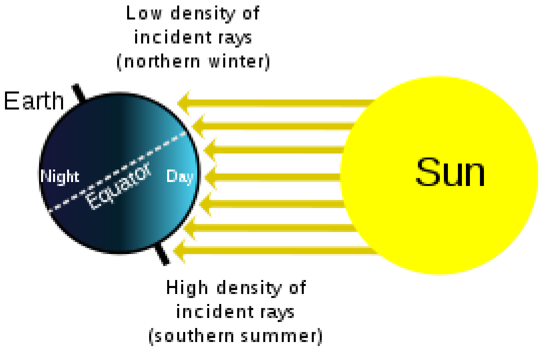

This is a diagram of the seasons. In addition to the density of incident light, the dissipation of light in the atmosphere is greater when it falls at a shallow angle.

Terrestrial radiation is emitted at much longer wavelengths than solar radiation. This is because Earth is much colder than the sun. Radiation is emitted by Earth across a range of wavelengths, as formalized in Planck's law. The wavelength of maximum energy is around 10 micrometers.

Cloud physics

Cloud physics is the study of the physical processes that lead to the formation, growth and precipitation of clouds. Clouds are composed of microscopic droplets of water (warm clouds), tiny crystals of ice, or both (mixed phase clouds). Under suitable conditions, the droplets combine to form precipitation, where they may fall to the earth.[8] The precise mechanics of how a cloud forms and grows is not completely understood, but scientists have developed theories explaining the structure of clouds by studying the microphysics of individual droplets. Advances in radar and satellite technology have also allowed the precise study of clouds on a large scale.

Atmospheric electricity

Atmospheric electricity is the term given to the electrostatics and electrodynamics of the atmosphere (or, more broadly, the atmosphere of any planet). The Earth's surface, the ionosphere, and the atmosphere is known as the **global atmospheric electrical circuit**.[9] Lightning discharges 30,000 amperes, at up to 100 million volts, and emits light, radio waves, x-rays and even gamma rays.[10] Plasma temperatures in lightning can approach 28,000 kelvins and electron densities may exceed $10^{24}/\text{m}^3$.[11]

Cloud to ground Lightning in the global atmospheric electrical circuit.

Atmospheric tide

The largest-amplitude atmospheric tides are mostly generated in the troposphere and stratosphere when the atmosphere is periodically heated as water vapour and ozone absorb solar radiation during the day. The tides generated are then able to propagate away from these source regions and ascend into the mesosphere and thermosphere. Atmospheric tides can be measured as regular fluctuations in wind, temperature, density and pressure. Although atmospheric tides share much in common with ocean tides they have two key distinguishing features:

i) Atmospheric tides are primarily excited by the Sun's heating of the atmosphere whereas ocean tides are primarily excited by the Moon's gravitational field. This means that most atmospheric tides have periods of oscillation related to the 24-hour length of the solar day whereas ocean tides have longer periods of oscillation related to the lunar day (time between successive lunar transits) of about 24 hours 51 minutes.[12]

ii) Atmospheric tides propagate in an atmosphere where density varies significantly with height. A consequence of this is that their amplitudes naturally increase exponentially as the tide ascends into progressively more rarefied regions of the atmosphere (for an explanation of this phenomenon, see below). In contrast, the density of the oceans varies only slightly with depth and so there the tides do not necessarily vary in amplitude with depth.

Note that although solar heating is responsible for the largest-amplitude atmospheric tides, the gravitational fields of the Sun and Moon also raise tides in the atmosphere, with the lunar gravitational atmospheric tidal effect being significantly greater than its solar counterpart.[13]

At ground level, atmospheric tides can be detected as regular but small oscillations in surface pressure with periods of 24 and 12 hours. Daily pressure maxima occur at 10 a.m. and 10 p.m. local time, while minima occur at 4 a.m. and 4 p.m. local time. The absolute maximum occurs at 10 a.m. while the absolute minimum occurs at 4 p.m.[14] However, at greater heights the amplitudes of the tides can become very large. In the mesosphere (heights of ~ 50 – 100 km) atmospheric tides can reach amplitudes of more than 50 m/s and are often the most significant part of the motion of the atmosphere.

Aeronomy

Aeronomy is the science of the upper region of the atmosphere, where dissociation and ionization are important. The term aeronomy was introduced by Sydney Chapman in 1960.[15] Today, the term also includes the science of the corresponding regions of the atmospheres of other planets. Research in aeronomy requires access to balloons, satellites, and sounding rockets which provide valuable data about this region of the atmosphere. Atmospheric tides play an important role in interacting with both the lower and upper atmosphere. Amongst the phenomena studied are upper-atmospheric lightning discharges, such as luminous events called red sprites, sprite halos, blue jets, and elves.

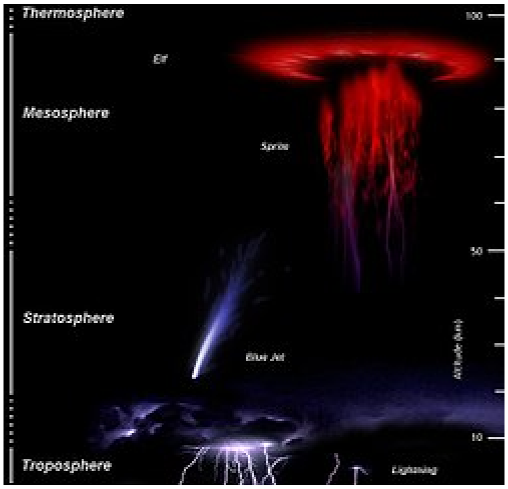

Representation of upper-atmospheric lightning and electrical-discharge phenomena

Centers of research

In the UK, atmospheric studies are underpinned by the Met Office, the Natural Environment Research Council and the Science and Technology Facilities Council. Divisions of the U.S. National Oceanic and Atmospheric Administration (NOAA) oversee research projects and weather modeling involving atmospheric physics. The US National Astronomy and Ionosphere Center also carries out studies of the high atmosphere. In Belgium, the Belgian Institute for Space Aeronomy studies the atmosphere and outer space.

See also

- Adiabatic lapse rate
- Atmospheric thermodynamics
- Baroclinic instability
- Barotropic vorticity equation
- Convective instability
- Coriolis effect
- Euler equations
- FluxNet
- Geostrophic wind
- Gravity wave
- Hydrostatic balance
- Kelvin–Helmholtz instability
- Madden–Julian oscillation
- Navier–Stokes equations
- Potential vorticity
- Pressure-gradient force
- Primitive equations
- Rossby number
- Rossby radius of deformation
- Space weather
- Space physics
- Thermal wind
- Vorticity equation

References

1. COMET program (1999). Remote Sensing. (http://www.comet.ucar.edu/nsflab/web/remote/121.htm) University Corporation for Atmospheric Research. Retrieved on 2009-04-23.
2. Glossary of Meteorology (2009). Radar. (http://amsglossary.allenpress.com/glossary/search?id=radar1) American Meteorological Society. Retrieved on 2009-24-23.
3. NASA (2009). Earth. (http://hurricanes.nasa.gov/earth-sun/technology/remote_sensing.html) Archived (https://web.archive.org/web/20060929081013/http://hurricanes.nasa.gov/earth-sun/technology/remote_sensing.html) 2006-09-29 at the Wayback Machine. Retrieved on 2009-02-18.

4. Atmospheric Science Data Center. What Wavelength Goes With a Color? (http://science-edu.larc.nasa.gov/EDDOCS/Wavelengths_for_Colors.html) Archived (https://web.archive.org/web/20110720105431/http://science-edu.larc.nasa.gov/EDDOCS/Wavelengths_for_Colors.html) 2011-07-20 at the Wayback Machine. Retrieved on 2008-04-15.
5. Windows to the Universe. Solar Energy in Earth's Atmosphere. (http://www.windows.ucar.edu/tour/link=/earth/Atmosphere/earth_atmosph_radiation_budget.html) Retrieved on 2008-04-15.
6. University of Delaware. Geog 474: Energy Interactions with the Atmosphere and at the Surface. (http://www.udel.edu/Geography/DeLiberty/Geog474/geog474_energy_interact.html) Retrieved on 2008-04-15.
7. Wheeling Jesuit University. Exploring the Environment: UV Menace. (http://www.cet.edu/ete/modules/ozone/ozatmo.html) Archived (https://web.archive.org/web/20070830094306/http://www.cet.edu/ete/modules/ozone/ozatmo.html) August 30, 2007, at the Wayback Machine. Retrieved on 2007-06-01.
8. Oklahoma Weather Modification Demonstration Program. CLOUD PHYSICS. (http://www.evac.ou.edu/okwmdp/physics.html) Archived (https://web.archive.org/web/20080723154301/http://www.evac.ou.edu/okwmdp/physics.html) 2008-07-23 at the Wayback Machine. Retrieved on 2008-04-15.
9. Dr. Hugh J. Christian and Melanie A. McCook. Lightning Detection From Space: A Lightning Primer. (http://thunder.msfc.nasa.gov/primer/primer3.html) Archived (https://web.archive.org/web/20080430142456/http://thunder.msfc.nasa.gov/primer/primer3.html) April 30, 2008, at the Wayback Machine. Retrieved on 2008-04-17.
10. NASA. Flashes in the Sky: Earth's Gamma-Ray Bursts Triggered by Lightning. (http://www.nasa.gov/vision/universe/solarsystem/rhessi_tgf.html) Retrieved on 2007-06-01.
11. Fusion Energy Education.Lightning! Sound and Fury. (http://fusedweb.llnl.gov/CPEP/Chart_Pages/5.Plasmas/Lightning.html) Retrieved on 2008-04-17.
12. Glossary of Meteorology. Atmospheric Tide. (http://amsglossary.allenpress.com/glossary/search?id=atmospheric-tide1) Retrieved on 2008-04-15.
13. Scientific American. Does the Moon have a tidal effect on the atmosphere as well as the oceans?. (http://www.sciam.com/article.cfm?id=does-the-moon-have-a-tida) Retrieved on 2008-07-08.
14. Dr James B. Calvert. Tidal Observations. (http://mysite.du.edu/~jcalvert/geol/tides.htm) Retrieved on 2008-04-15.
15. Andrew F. Nagy, p. 1-2 in *Comparative Aeronomy*, ed. by Andrew F. Nagy *et al.* (Springer 2008, ISBN 978-0-387-87824-9)

Bibliography

- J. V. Iribarne, H. R. Cho, *Atmospheric Physics*, D. Reidel Publishing Company, 1980

Retrieved from "https://en.wikipedia.org/w/index.php?title=Atmospheric_physics&oldid=824104541"

This page was last edited on 5 February 2018, at 10:06.

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.

Don’t Let The Big Energy Corporations Control You – Free Energy Is HERE

(**+6** *rating*, **6** *votes*, ***rated***)

You need to be a registered member to rate this post.

May 26, 2012 *in* *Activism*, *Preparedness*, *Resistance*, *Science*, *Technology*

by Kellie

The below added text I have been looking into has been an easily made design. Our government goes to no end to stop men and women who help human kind with their awesome God given inventions. I am not afraid to share this as Alex does with his knowledge and research. It is to open the minds of those that are dumbed down by corporate America and the global elite to do as they bid…No…stand up and fight against their system.

Free energy is NOT a hoax as they want you to believe.

If you have any engineer in you or technical expertise – Read…It works !

Jes Ascanius’ Version of Nikola Tesla’s Aerial System

This sort of information may seem confusing and maybe a little too technical for you, so let me tell you about the practical and useful applications used by Jes Ascanius, a Danish developer, to whom thanks is due for sharing his design. Initially, he set up a system to charge his mobile phone battery overnight from an aerial. Then he went on to produce a full-size Tesla Aerial System as described at the start of this chapter. Let’s start with the very simple system and progress from that to the more powerful arrangements.

The initial circuit uses one strand of solid wire which rises vertically to a 700 mm diameter drum where there are some twenty turns. The arrangement is like this:

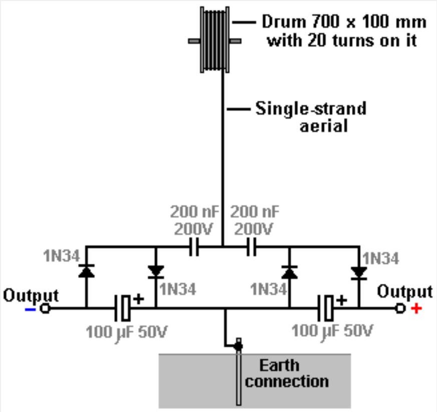

The aerial wire is several metres long, and in the prototype, was supported by (and insulated from) the eaves of a house. The aerial should be vertical or near vertical and a proper earth connection provided by driving a metal rod into the ground or connecting a wire to a metal plate and burying the plate in the ground as a good electrical connection is needed here. The earth connection used here is a 12 mm copper pipe 3 metres long, driven into the ground and the ground around it saturated with water:

The wire used to connect with the earthing rod is very important and should not be less than 8 swg copper wire, that is, 4 mm diameter and 13 sq. mm. cross-sectional area. As with all free-energy devices, the exact constructional details are vital.

The diodes used are germanium 1N34 or 1N34a as germanium diodes drop far less voltage than do silicon diodes and the 1N34 types are low-loss diodes, particularly suited to this application. Ceramic disc types are recommended for the 200 nF capacitors. The prototype build looked like this:

REFEED 24/7
LISTEN NOW

UPCOMING MISSIONS

- Complete Your 'Base, Loc
Interests' Sections Of You
Complete by:Thursday Dec 31, 2
- Flood Alex With Demands
Does And Does So Very T
Complete by:Friday Jan 1, 2016 /
- #InfowarsMusic
Complete by:Saturday Apr 30, 20
- Prevent Tyranny In All Nat
Complete by:Sunday Jan 1, 2017
- Collaborators Remembere
Complete by:Sunday Dec 31, 201
- Mission #4: Earth Psychic
Complete by:Thursday Dec 31, 2
- N.D.A.A (National Defense
Complete by:Sunday Jul 31, 2022
- Be The Best You Can Be
Complete by:Wednesday Nov 30
- Agenda 21
Complete by:Saturday Dec 31, 2
- U.N. Gun Arms Trade Trea
Complete by:Saturday Dec 31, 2

INSPIRATION SITES

Infowars.com
PrisonPlanet.tv
Infowars Shop
PrisonPlanet.com

Asked by Infowars

How do you fee
players kneelin
anthem?

I stand with

I stand with

(https://www.squareoffs

INFOWARS.COM NE

UPCOMING EVENTS

There are no upcoming group e

SEARCH:

Search

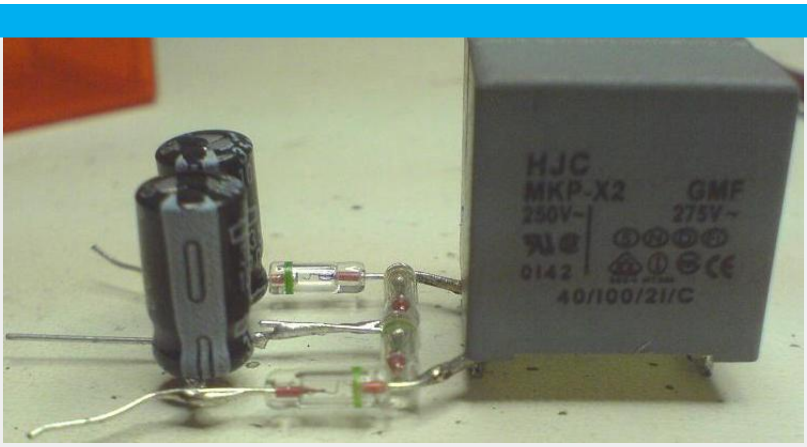

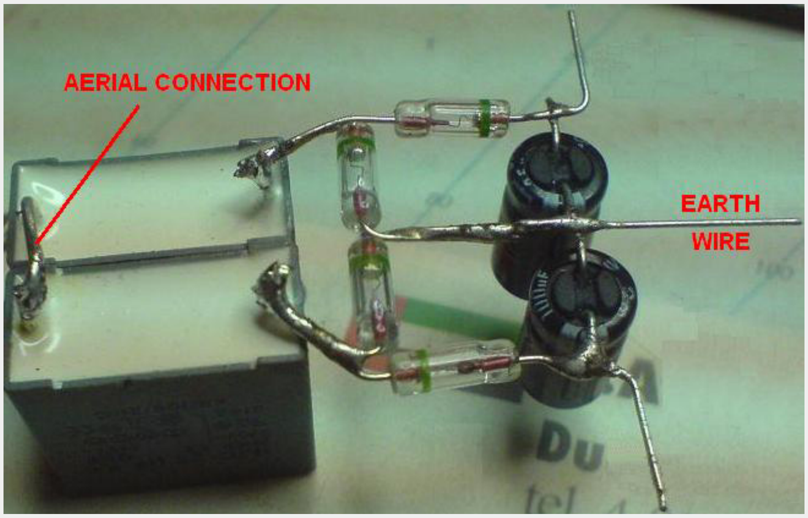

Now, consider this circuit as described, to be one modular building block which can lead to unlimited power from an aerial. I will represent the circuit shown above as a rectangle, showing the above circuit as:

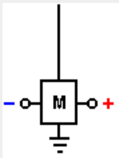

While it is possible to use more than one module with the aerial to get more power, the Danish developer then switched to the full-blown Tesla arrangement by attaching a 800 x 600 x 2 mm aluminium plate inside the sloping roof of his house:

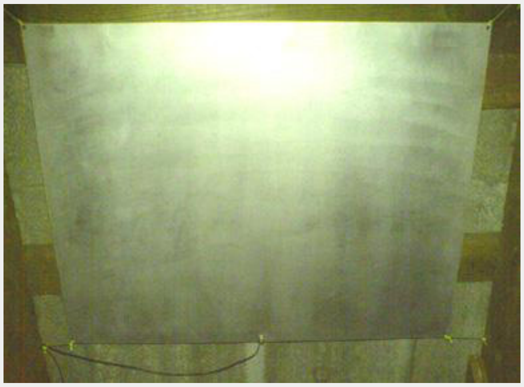

The plate being suspended using nylon cord to prevent it touching the roof or anything else:

The plate is positioned between 3 and 3.5 metres (10 to 12 feet) above the ground and the attachment to the plate is also heavy-duty 8 swg cable:

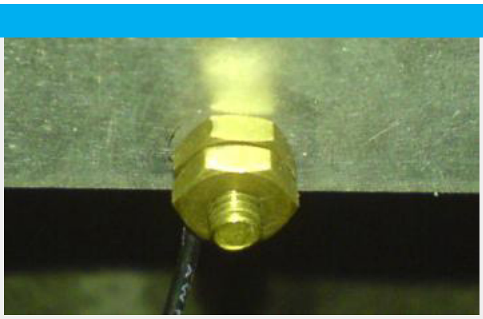

The cable is connected to the aluminium plate using a brass bolt and nuts which the builder thinks may be significant, quite apart from avoiding any galvanitic connection to the circuit. The cable is then run vertically downwards to the circuit. For this arrangement a second earthing point is also used. This is a galvanised iron pipe 3 metres long, driven vertically into the ground which is saturated with water. The second earth is 2 metres away from the first earth.

This arrangement provides serious power, enough to cause injury to, or kill a careless human. With two modules, it will light an LED very brightly, driving it to 2.6 volts. If the LED is removed, then the voltage climbs to about twenty volts and is easily sufficient to charge a 12V battery or battery bank although that takes time. With twenty modules as 12V battery can be charged over night. It is estimated that with two hundred modules, the power would be sufficient to power a household although that has not yet been done. It should be borne in mind that each module is easy and cheap to make, so arranging for a stack of them where additional modules can be added at a later date for more power, is an ideal arrangement. The circuit is like this:

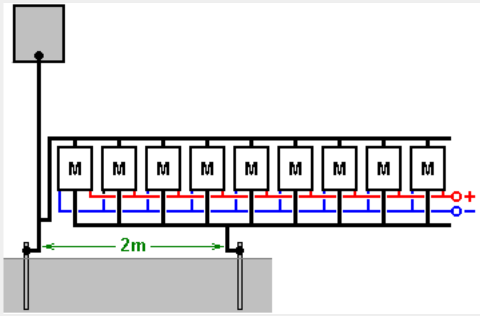

This circuit looks completely mad as the aerial input to the circuit appears to be directly short-circuited by the second earth connection. In spite of this, the circuit works very well when connected this way. Additional modules can be added without any known limit. Increased power can be had by either raising the aluminium plate higher above the ground, to say, 10 metres (33 feet), or by adding one or more additional aerial plates. As you have a good aerial connected through to a very good earth, there has to be the possibility of the equipment being hit by lightning, and so it is suggested that a protective spark-gap is installed between the aerial and the earth, close to the circuit, so that if high-voltage is suddenly applied to the aerial, the spark gap will fire and shunt the excess power through to the earth. Alternatively, possibly a better solution is to install a standard lightning rod system a few metres away from the aerial and a metre or two higher up, so that it forms a more attractive point for a lightning strike.

Further experimentation has shown that altering the connection point for the aerial has a significant effect on the results. If the connection is made at the mid point between the aerial plate and the earth connection, it produces a greater output:

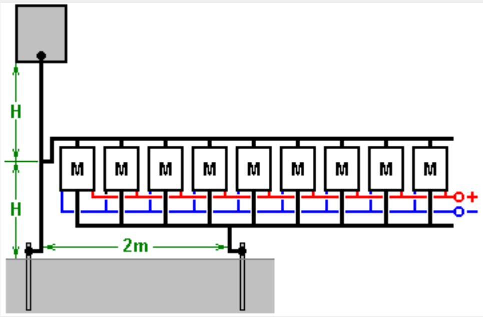

With this arrangement a single module produces around 30 volts while the original method of connecting near the earth was giving about 26 volts with two modules.

Dragan Kljajic has been experimenting with this circuit and has started by building many of these modules on a printed circuit board like this:

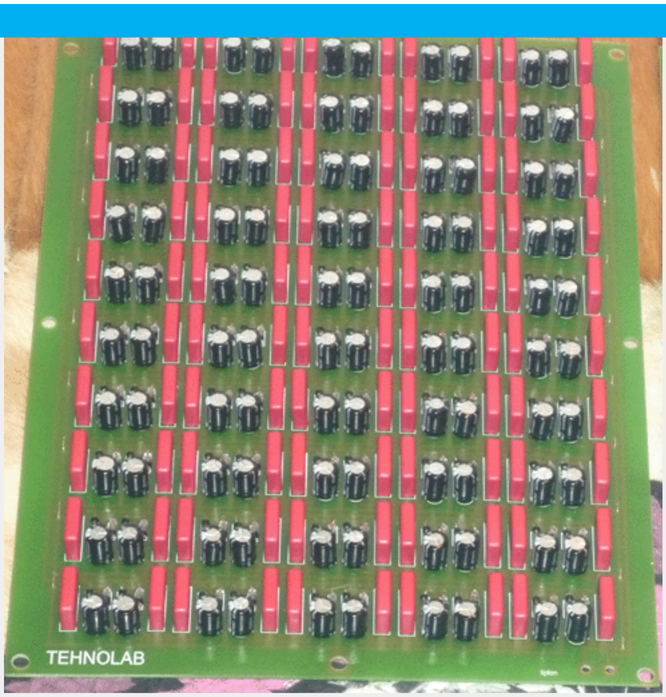

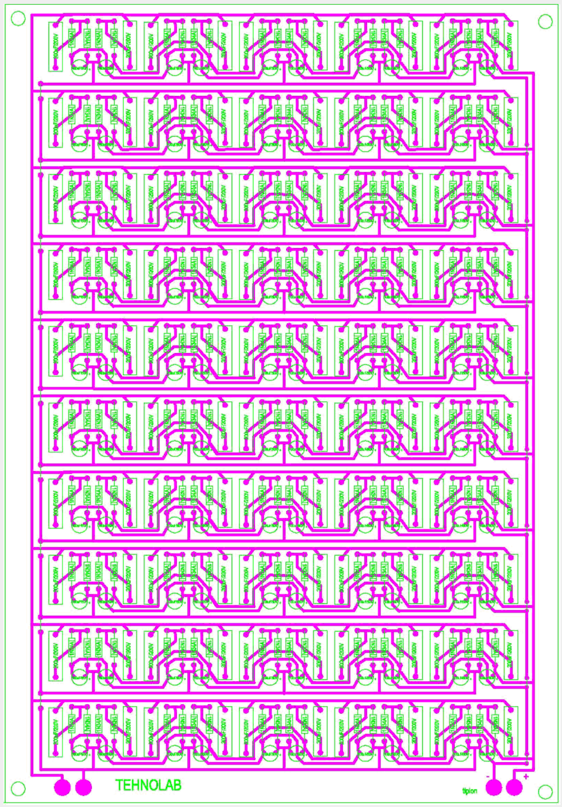

Using two of these boards, Dragan is pulling 96 watts continuously from his aerial plate. He intends to extend this arrangement much further, but is being hindered at present by a local civil war.

There is a forum where some builders of this system share comments. One comment is that there is an increased risk of a lightning strike where you have an earthed aerial, and so it is advisable not to place the aerial plate inside a house, but perhaps suspended between two trees. Also, using a car spark plug connected across the module set can protect against lightning strikes damaging the circuitry.

← Truth is a commodity… Sad →

INFOWARS.COM VIDEOS:

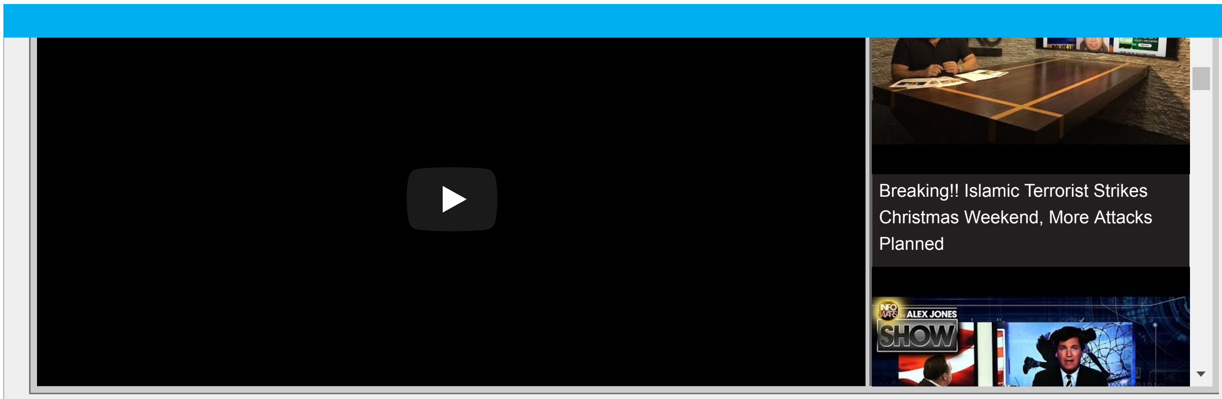
Breaking!! Islamic Terrorist Strikes Christmas Weekend, More Attacks Planned
ALEX JONES
SHOW

COMMENT ON THIS ARTICLE:

Comment rules

9 responses to *Don’t let the big energy corporations control you – Free energy is HERE*

Hendra said on April 26, 2016

Hello, I am new to this website and I am trying to build this things right now with the pcb. Has anyone successfully built this things with the PCB that Dragan make ?

THank you

Hendra

Log in to Reply

ifree said on June 1, 2012

cool. I remember reading an article about a farmer who was able to light a bunch of bulbs from a really long cable that somehow just inducted energy… this was after that idiotic show “Mythbusters” said it was impossible.. so, yeah. I don’t recommend that show.

Log in to Reply

Shaun said on June 1, 2012

I have made two of these boards but in A4 size with 32 modules on each board im in the process of drilling holes.
The alexkor circuit in Chap 7 is also worth looking at I will let you know when Im done with the Tesla /Jes Ascanius
circuit and what results I get I know its important to polish the plate and to have a very good earthing plate.

Log in to Reply

United said on May 28, 2012

Great post! thanks

Planet Infowars Log In Visit

Kakashi said on May 28, 2012

Is this transduction?

Log in to Reply

Kellie said on May 28, 2012

The whole article is here http://www.free-energy-info.com/Chapt7.html

Log in to Reply

WizarDave said on May 26, 2012

Gotta love Tesla!

Kellie said on July 6, 2012

Yes, and look how our great government treated him later in life up unto his death, and still to this day…Corrupticians !

Log in to Reply

ZerMaster said on May 26, 2012

Don’t forget about Tesla’s Turbine at 1PSI Steam hooked up to a generator (1PSI! think about it …) cheap to build and use. http://books.google.ro/books?id=BOMDAAAAMBAJ&pg=PA188&dq=popular+mechanics+Tesla&hl=en&sa=X&ei=BeC_T4G1FdDn-gbS-rCcCg&sqi=2&ved=0CCwQ6AEwAA#v=onepage&q=popular%20mechanics%20Tesla&f=false

http://www.youtube.com/watch?v=Oqyh-JyCmQc

Log in to Reply

Leave a reply

You must be logged in to post a comment.

BE ACTIVE

- LOGIN
- GROUPS
- MEMBERS
- F.A.Q.
- TUTORIALS

YOUR ACCOUNT

- LOGIN

MORE

- CONTACT
- WRITING TIPS
- TERMS OF SERVICE
- DIGITAL MILLENNIUM COPYRIGHT ACT

©2012 Planet.Infowars.com is a Planet Infowars, LLC company. All rights reserved - Digital Millennium Copyright Act Notice

MDS975 www.mds975.co.uk

CRYSTAL SETS 5
Experimental
Crystal Sets

MDS975 www.mds975.co.uk

Home | Contact | Site Map | Radio Stations & Memorabilia | Amateur Radio

Crystal Sets Introduction | Resistor & Capacitor Conversion Tables

Crystal Sets (Part1) | Build Your Own Crystal Set (Part 2)

Spider's Web Crystal Set (Part 3) | Crystal Set By Kenneth Rankin (Part 4) | Crystal Radio Links

CRYSTAL SETS 5: EXPERIMENTAL CRYSTAL SETS

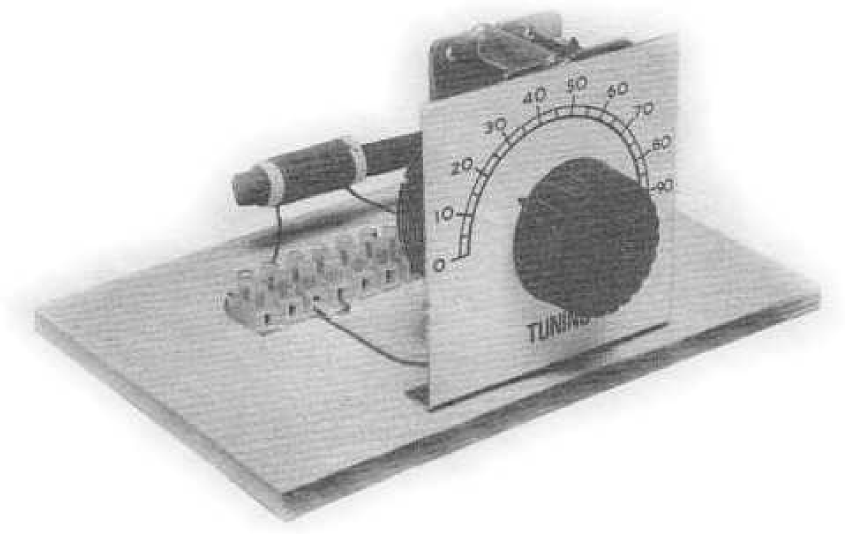

Picture 1 - The Complete Experimental Crystal Set

THE POPULARITY of the crystal radio arises from its simplicity, and the fact that it needs no power supply. The circuit here allows for easy experiments with tuning, aerial and diode coupling, and frequency coverage. Wrong connections can cause no damage to any components.

A Crystal Set is more often than not used for the reception of medium and long wave radio, but short wave reception is also quite feasible. It will normally be possible to receive some of the stronger international radio stations.

This is adapted from an article that appeared in the 1970's in Everyday Electronics, and gave me almost endless hours of fun!

BASIC CIRCUIT

The basic circuit is shown in Picture 2 below. The coil L1 can be air cored, or have a ferrite rod placed in its winding. The variable capacitor C1, in conjunction with aerial-earth capacitance, tunes the circuit to resonate with the wanted radio station frequency. The diode D1 "detects" or demodulates the radio signal so that the programme is heard in the earpiece.

This basic circuit can be modified in various ways to obtain better performance.

EARPHONE

As most constructors will be using a Crystal Earpice to listen to the crystal set it is essential that a 47k Ohm resistor is connected across the earphone terminals (TB1/1 and TB1/2 in the diagram), i.e. in parallel with the earphone, otherwise results will be very quiet.

A High Impedance headset of 20k Ohms (20,000 Ohms) may give even better results, but these are very difficult to obtain , so unless you happen to already own such a headset the Crystal Earphone with 47k resistor will be the only option. An ordinary magnetic earpiece or Walkman headphones will not work with a crystal set.

ASSEMBLY

Construction is of a 'breadboard' type using a wooden board of about 165 x 130 mm. A 12-way block connector, TB1, is used to connected together the components and this is screwed onto the wooden board. The use of a block connector provides an easy method of connecting the components together and then subsequently rearranging them as the experiments progress.

Tuning capacitor C1 is screwed to a bracket made of some scrap metal which is then also screwed firmly down to the baseboard, see Picture 1 above. Thin plywood screwed to the front edge of the baseboard would also provide a suitable method of fixing the tuning capacitor to the base. A knob with pointer is fitted to C1, and a scale is drawn and fitted behind this.

Except for C1, all connections are made by the terminals of the 12-way terminal block as shown in Picture 4. Loosen the screws with a small screwdriver, insert the bared ends of the wires, and tighten the screws. The various locations on the terminal block, TB1, are also shown in the circuit diagram, Picture 2.

AERIAL AND EARTH

Crystal receivers need a long wire aerial preferably strung outside and about 25m long, or as long as is possible to install. If this is outside it should be high and clear of earthed objects as this will improve performance.

An earth is absolutely essential for a crystal set to work properly. The earth lead can be run to an earth rod or spike that is buried to a depth of about 1 meter into damp soil. Or it may be soldered to a bare metal can which is buried in damp soil.

It is feasible, though not recommended, that the earth lead can be connected to the earthing terminal of a hi-fi system or even to the bare metal case of a personal computer that is plugged into an earthed mains outlet, but is switched OFF.

Stranded, insulated wire, or purpose made aerial wire can be used for the aerial and earth leads.

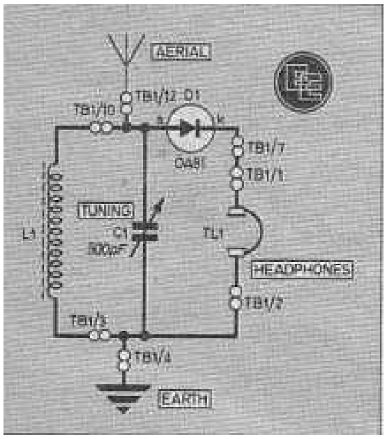

Picture 2 - The Basic Circuit

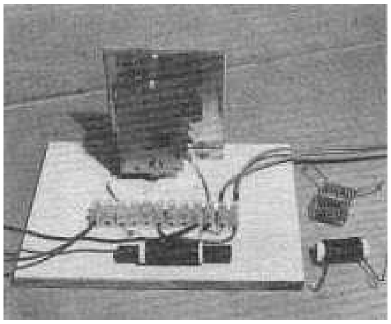

Picture 3 - Photo Of The General Layout

INDUCTORS (The Tuning Coils)

The following four coils are suggested for initial use as L1 :

Coil 1: Make a thin card tube to slide on a 10mm diameter ferrite rod, and on this tube wind about 105 turns of 32 s.w.g. enamelled copper wire, side by side. Secure ends with sticky tape.

Coil 2: Make a similar coil to to coil 1 having about 15 turns of 24 s.w.g. enamelled wire on the card tube. Loops of cotton will help hold the ends in place.

Coil 3: Wind 9 turns of 20 s.w.g. bare tinned copper wire on an object about 20mm in diameter. Remove and stretch to separate the turns, to obtain a coil about 25mm long.

Coil 4: Make a similar coil to coil 3, but with 5 turns.

The Ferrite Rod

It will be necessary to have a ferrite rod of about 60mm to 75mm long available. Coils 1 and 2 will provide reception of medium wave and the longer short wave bands. Coil 3 should cover about 3 - 10MHz shortwave with the ferrite placed in it, or about 6 - 18MHz with the ferrite rod removed. Coil 4 should cover about 6 -13MHz with the rod in, and about 9 - 20MHz without the' rod.

It will be noted that as the ferrite rod is inserted, any particular signal has to be re-tuned by opening Cl. This arises because the ferrite increases the inductance of the winding, so less parallel capacitance is needed for the same resonant frequency.

EFFICIENCY CHECKS

Tune in a m.w. transmission using coil 1 which gives good headphone volume. Place a microammeter or multi-range meter on a sensitive range in series with the headphones. A reading of 50-100uA or more may be obtained, depending on aerial, earth, earphone resistance and resistor value, coil and detector efficiency and strength of signals at your locality.

Placing the ferrite rod in the coil and re-tuning should boost the meter reading to some extent. Surplus or other detector diodes can be tried by substituting them in turn and noting the meter reading. Improvements to the aerial (or earth) will also show up as a rise in meter reading.

If experimenting with a crystal earpiece, which gives no direct current circuit, the meter may be clipped across the phone leads, i.e. D1 cathode to earth.

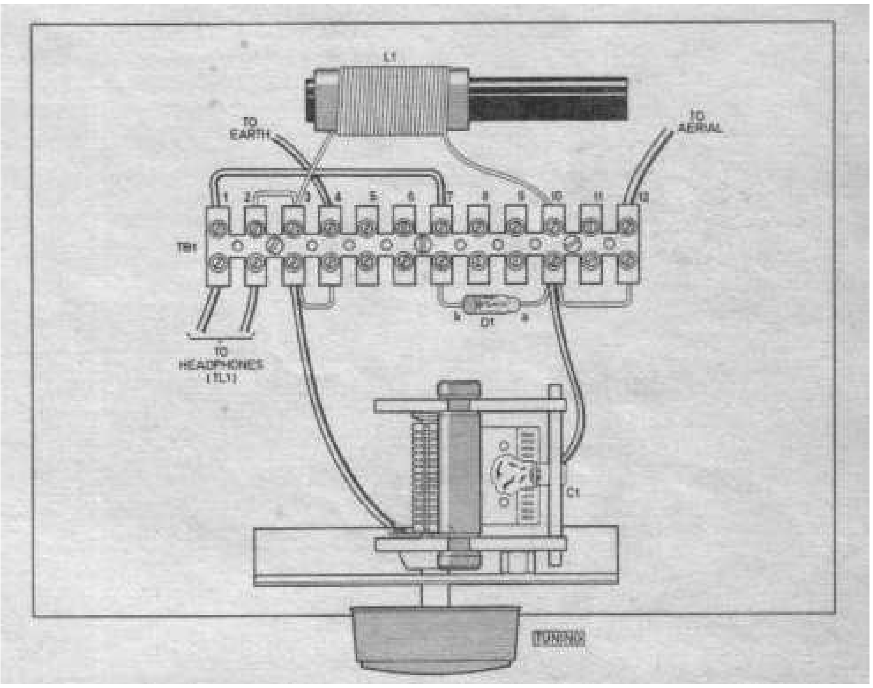

Picture 4 - Baseboard Layout Of The Crystal Set

AERIAL COUPLING

The aerial loads the tuned circuit heavily when connected directly to the top of the tuned circuit, as in Picture 2. This damps the tuning action and it can be found that stations spread out all over the dial, which is unsatisfactory.

The series capacitor, C2 connected in Picture 5(a) reduces the loading and thus improves the sharpness of the tuning. A variable or pre-set capacitor of about 250pF maximum is most suitable. for this role, though it is possible to experiment with a variety of fixed value capacitors in this range also.

Connecting the aerial to a tapping on the coil, as in Picture 5 (b) also sharpens tuning. It may also increase volume. Try about 2 turns from earth for coil 4, or 4 turns from earth for coil 3.

Another method is to have a coupling primary, as in Picture 5 (c). This consists of a second coil, with about one third the turns of the original wound on top of the existing coil.

You can even combine these methods to find what arrangement best suits the aerial in use.

The diode can be disconnected from the end of L1 and taken to a spare position on TB1 for example location TB1/9. You can then run a flying-lead fitted with a crocodile clip from this position, connecting it to various tappings on the coil as required as in Picture 5 (d). This method also reduces loading on the tuned circuit.

Coils with spaced turns of bare wire are readily tapped. For other coils, small loops can be made every ten turns or so, and crocodile clips can be attached to these when selecting tapping points.

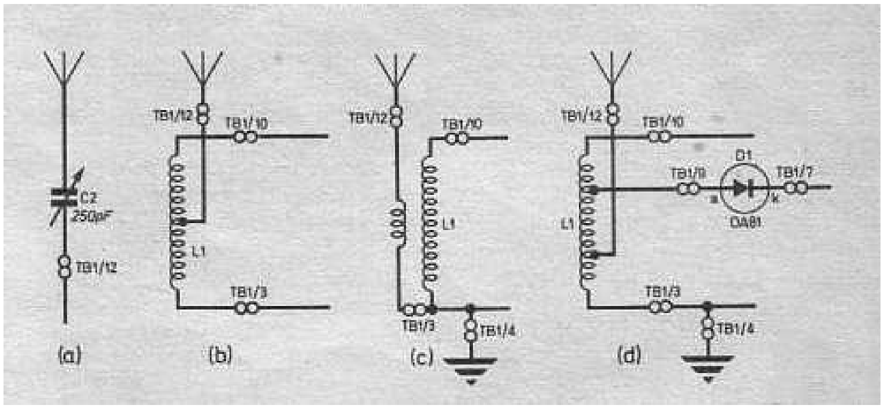

Picture 5 - Alternative Methods Of Aerial Coupling

SHORT WAVES

For shortwave reception, a good efficient outdoor aerial is certainly recommended. Evening listening in the region around 5 - 9MHz in often proves to be the most fruitful.

Since there is no amplification, as with a valve or transistor receiver, certain frequencies will seem to be completely dead at particular times of day. So if the crystal receiver works satisfactorily on medium wave and longwave, but no shortwave signals are heard, check again in the evening, or after dark, when conditions are different.

PARTS REQUIRED

C1	**365pF or 500pF Air Spaced Tuning Capacitor**
D1*	**OA47, IN34, OA81, OA90, OA91, IN94 or similar point contact small signal Germanium Diode** ***** The OA47 will be of particular interest since it has the lowest forward bias voltage of any of these diodes which will make the crystal set somewhat more sensitive and therefore louder. The US equivalent of the British OA47 is the IN34.
TL1	**High Impedance Headphones (20,000 Ohms)** **or Crystal Earphone**
TB1	**12-Way Plastic Screw Block Terminal**
Also Required:	**47 k Ohm Resistor for Crystal Earphone: Enamelled Copper Wire: 32 and 24 s.w.g. for L1: 20 s.w.g. tinned wire for L1: Ferrite Rod 10mm diameter x 75 mm long: 25m of wire for aerial: Wire and rod or spike etc for earth: Wood for base e.g. 10mm x165mm x 130mm: Scrap of metal of thin plywood for C1 bracket/front panel: Knob: Crocodile clip(s)**

Adapted from an article in Everyday Electronics magazine, November 1981, By F.G. Rayer.

HERE ARE A COUPLE OF VERY INTERESTING CRYSTAL SET DESIGNS SENT IN BY KRYSATEC - "THE RAT" - FROM THE CZECH REPUBLIC

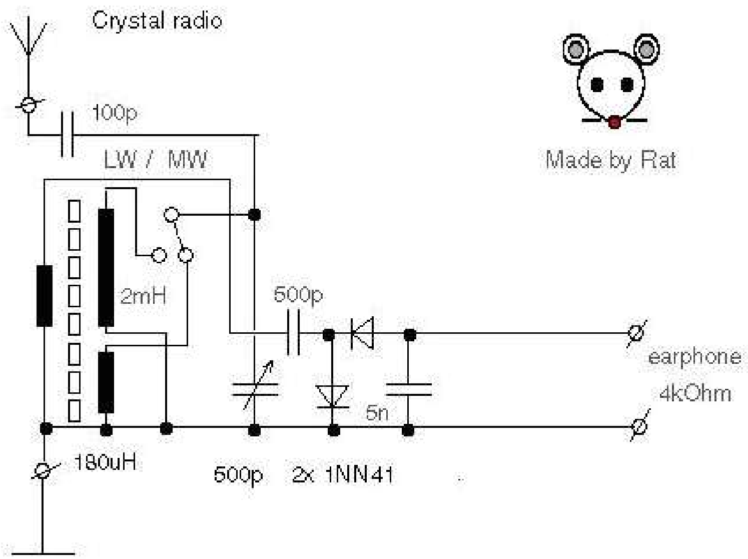
Crystal radio
100p
LW / MW
Made by Rat
2mH
500p
earphone
4kOhm
5n
180uH
500p
2x 1NN41

1/ Using old coils from old bulb radio for MW and LW band. Though it would be straightforward to wind the coils - one for Long Wave, one for Medium Wave and a coupling coil. Variable capacitor is 2 x 500pF only one half is used: 500pF. For the crystal earphone a resistor of about 82k ohm in parallel is required. This set also uses two Ge diodes as a multiplier in the quest for for higher audio signal output.

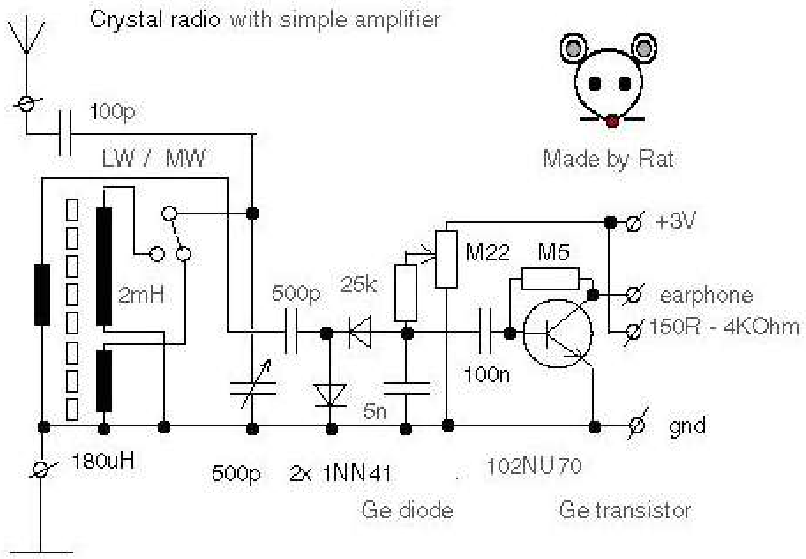
Crystal radio with simple amplifier
100p
LW / MW
Made by Rat
+3V
M22
M5
2mH
500p
25k
earphone
150R - 4KOhm
100n
5n
gnd
180uH
500p
2x 1NN41
102NU70
Ge diode
Ge transistor

2. If signals are not strong signal in your location, then the above circuit design can be considered. A simple transistor amplifier is used. A variable resistor M22 is used for better sensitivity which can be adjusted for poor signals. This crystal radio is aversion from cca 1960 - 1970 y.

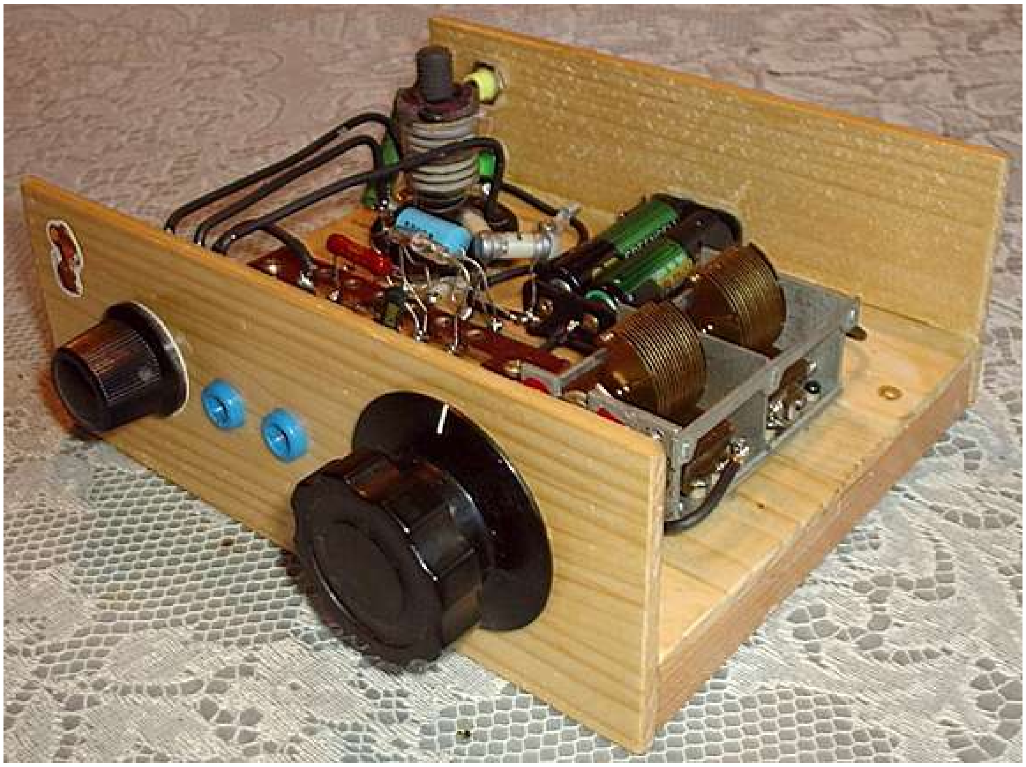

Rat's finished Crystal Set with additional amplification - very neat!

'Minilabs' Crystal Radio

BELOW: Ian Tomlinson kindly sent in a photograph of the box that contained the kit for his John Adams Toys 'Minilabs' Crystal Radio.

It is a very simple circuit consisting of the coil (inductor) with a sliding contact that provides variable tapping points, a diode and crystal earphone. All that is added is the aerial and earth. There is no variable tuning capacitor for simplicity and to keep costs down.

The coil provides the inductance required for tuning into a certain frequency (wavelength). These days a variable "tuning" capacitor is normally wired in parallel across the inductance (coil) in order to vary the resonance of the tuned circuit and therefore enable to easily tune into various transmitters on different frequencies. This crystal is tuned varying the number of turns on the coil (ie varying the inductance) by tapping off at different points using the sliding contact ("ball").

The crystal earpiece, or high Z headphone, is connected between the output of the detector diode (the other end from the coil) and earth. The volume from a crystal earpiece may be considerably improved by connecting a resistor of - somewhere between - 4.7 k and 47k ohms in parallel with the earpiece. A crystal earpiece cannot directly allow current to flow through it and the parallel resistor therefore allows current to better flow through the circuit.

John Adams Toys
Crystal
Radio Kit
Make a radio complete with tuning coil, antenna and earphone.
Pack contains all you need including full instructions.
No battery required.
Ages 8–14

'Minilabs' crystal set by John Adams Toys

A discussion on configurations for Crystal Sets by Felix Scerri VK4FUQ

This discussion, by Felix Scerri VK4FUQ, was posted at this address which no longer appears on the web www.tarc.org.au/techinfo2.htm (error 404) so here it is reproduced:

Crystal Set design is one of my passions closely allied with my obsession for audio and high fidelity.

My main interest in crystal sets, apart from the wonder of a radio receiver that does not require a power source, is the potential excellence of the recovered audio quality from normal AM broadcast stations.

Personally, it is one of my great laments that most people have never heard how good wideband AM can sound. A high performance crystal set or similar TRF approach is, in my opinion,the only way to do it. There are a few people around who have heard the audible results of my efforts,and can only agree.

I have often wondered,given the ultimate simplicity of the crystal set, being essentially a tuned circuit,a diode detector and some form of output device, what it takes to achieve optimum performance. What follows are my thoughts on the matter.

Crystal Set optimisation, is in my opinion, all about reduction of circuit losses. Essentially this means high "Q" tuned circuits and high quality detectors. Efficient output devices also help too. But as we will see, there are some trade-offs required as well. A high "Q" tuned circuit is always benefical, as a high "Q" tuned circuit has lowest RF losses,highest potential selectivity,and highest voltage at resonance, which is very useful for the diode being fed from the tuned circuit. Variable capacitors, even the "modern" miniature variable capacitors (although the older air dielectric units, as used in old valve receivers are more desirable) for various reasons,are generally quite efficient, and a higher "Q" coil will produce the most worthwhile improvements.The best (highest "Q") coils are wound with "Litz" wire, which is a multistranded woven wire with all strands insulated from each other. The performance of Litz wire wound coils is spectacular, unfortunately, although I know Litz wire is still being made, from personal experience, it is VERY rare in Australia.

Efficient coil design can be quite complex and all my coils are wound on ferrite rods. There seems to be,at least for ordinary single wire windings (close wound), an optimum wire thickness for optimum coil "Q". I have determined .315 mm winding wire to be about optimum for simple (single wire) coils on ferrite rods. Thicker wire is NOT better, believe it or not.

Lacking Litz wire, an interesting winding approach I have developed is to use two slightly thinner wires wound as a bifilar winding connected together at the beginning and end of the coil, yields considerably higher "Q" compared to a simple single wire winding. I have found 0.25 mm winding wire optimum in this application.

Whilst high "Q" coils are beneficial from the RF point of view, there is a possible downside. If one is interested in maximum selectivity and sensitivity, there is no problem, but remember highest "Q" results in a narrowed audio band-width as a simple consequence of band-width. For high fidelity applications this could be a disadvantage under some circumstances, although there are clever ways around this.

Regardless of ultimate coil "Q", selectivity is a major issue with crystal sets generally. Here another trade-off is evident. For the maximum voltage into the diode, connecting the diode to the high impedance end of the coil (i.e. the top) yields the greatest voltage but the selectivity is usually terrible, because of severe "loading" by the diode circuit. For this reason, tapping well own the coil improves selectivity at the expense of signal volume (reduced voltage). Once again there are ways around this. As described in my "Double Tuned Crystal Set Tuner" article in "Amateur Radio" magazine, March 2002, the use of two separately tuned coupled resonant circuits allows top connection into the diode without compromising overall selectivity, thanks to the use of a second tuned circuit which is fed from the external antenna. The whole network forms a double tuned input bandpass filter and in practice this approach works very well. For single coil crystal sets I recommend the use of an un-tuned "antenna" winding adjacent to the "hot" end of the main coil, preferably adjustable (old paper reels from sewing cotton threads are ideal). This allows the degree of coupling to be optimised under actual listening conditions. The double tuned set up is best, yielding superb selectivity, but the un-tuned antenna coil arrangement also works quite well, especially if the diode is tapped well down the main coil.Tapping halfway works well.

The other method of performance improvement involves the use of the most effective detector system possible. Here things get very interesting. In fact the temptation is to use more complex circuitry, but that gets away from the charming simplicity of the crystal set. As an example, my own crystal set tuner has at times mutated into a TRF tuner complete with FET RF preamplifiers, active(powered) detectors and other enhancements. These modifications do work well, but loses the simplicity of a basic crystal set. In actuality, a simple diode detector can work extremely well, subject to some qualification. Diodes like to work with a reasonable level of RF input voltage. Audio distortion can result under conditions of low signal level, due to diode transfer curve non linearity and other factors, such as the widespread use of broadcast station "processing". The actual type of diode makes a difference. The 1N34A germanium diode is very popular for crystal set use, although in my experience just about ANY germanium diode will work, although it is worth trying different specimens. Some are definitely better than others. Even from a pack of twenty 1N34A's from the same source, some were definitely better than others. Measuring the average value of rectified output voltage across the diode load resistor will show which diodes are best. By the way, I regard a diode load resistor as being mandatory. I find a value of about 47K about right, especially if a crystal earpiece is being used or the crystal set is being used as a tuner feeding an audio pre-amplifier and following amplifier. If using high impedance magnetic type headphones, the headphones provide the diode DC load.

Another type of diode that is very interesting, is the hot carrier diode. There seem to be a lot of different hot carrier diodes around these days. There are even hot carrier diodes now being sold as "germanium diode equivalents". I have tried them and they do work acceptably well, but they are not quite as good as genuine germanium diodes such as the 1N34A. Typical UHF mixer hot carrier diodes, such as the 1N5711 will not work well in crystal set service simply because their "turn on voltage"is too high, similar to silicon diodes such as the 1N4148/914 series, which require a lot of RF input to function adequately as RF detectors, however a simple technique can be used to turn hot carrier diodes such as the 1N5711 into superlative detectors.

I guess we are cheating a little, because the technique is to use a little voltage bias supplied via a 1.5v battery, through a simple potentiometer voltage divider arrangement, with capacitor (for DC isolation) fed into the diode from the tuned circuit. With applied adjustable bias, I find the 1N5711 diodes absolutely superlative detectors under ANY signal strength conditions. I find the detection quality also superlative, with a clarity and low noise profile unmatched by any other diode arrangement. In my opinion, hot carrier diodes, running bias,are the best detectors overall.

Regarding other detector arrangements, the diode "voltage doubler" is often recommended, however my own experiments with the doubler arrangement have been inconclusive and slightly disappointing overall. I have found no real advantage in their use over a simple (one) diode detector, believe it or not.

Yes, they do work, but they're nothing special, at least in my opinion.

Any comments on this general subject of crystal set optimisation would be welcome.

73's Felix Scerri VK4FUQ.
22nd July 2002

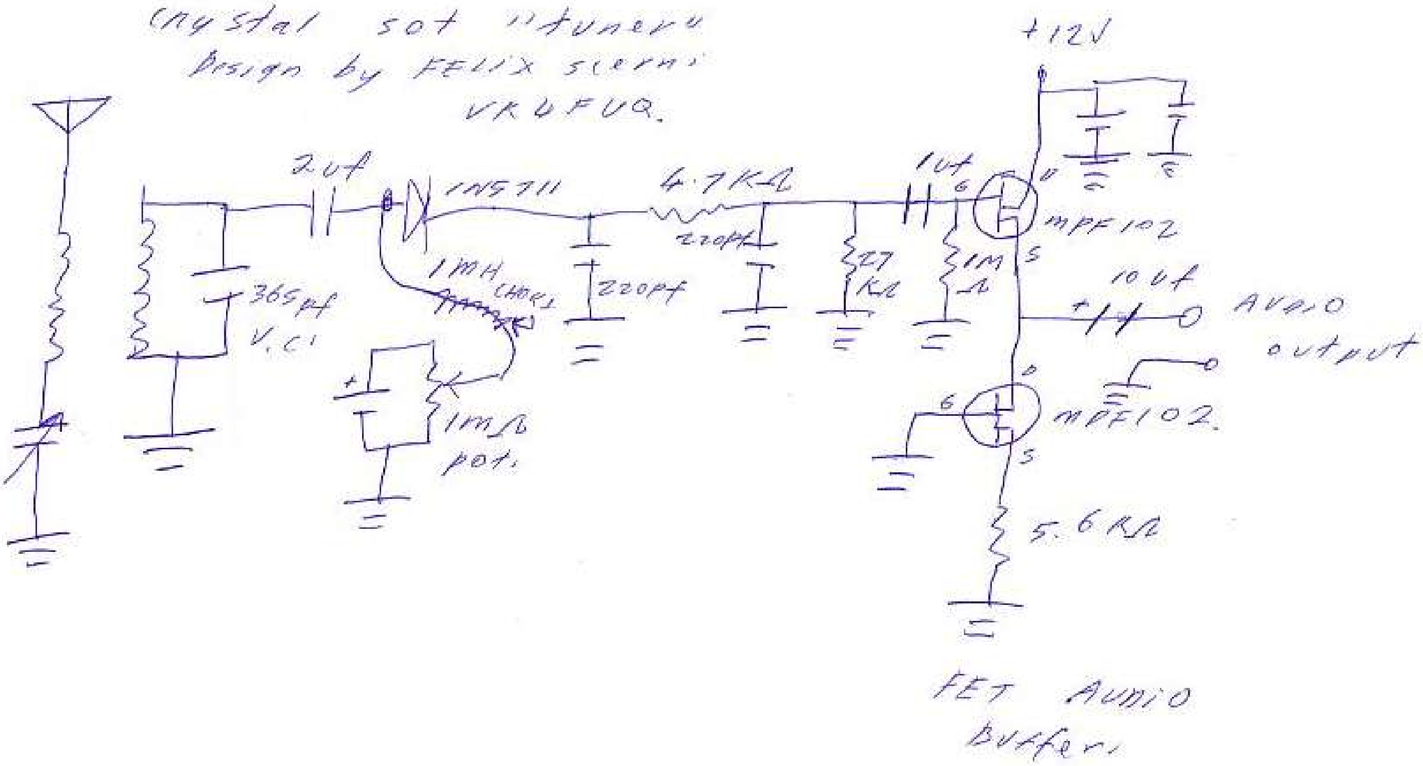

Above: CRYSTAL SET BASED CIRCUIT PROVIDING A HIGH QUALITY PROGRAMME SOURCE

IMPROVED VERSION OF THE ABOVE CONCEPT !! New update from Felix Scerri February 2010:

New 'two FET infinite impedance AM detector'

I've developed a new version of my old favourite FET 'infinite impedance' AM detector that I think sounds very nice. I include a short audio of one of our local AM stations. I picked this station as it is my reference 'torture test AM station' as they run very heavy 'processing' which normally sounds yuck with all my other (diode and non diode) detectors! However it's quite clean with this detector. What do you reckon? I'll do up a circuit if you'd like to feature it in your TRF radio section. A general draft article follows.

'A favourite non diode based AM detector that I've built and used many times over the years is the FET based infinite impedance detector, offering very good general AM detector performance, especially under weak RF signal conditions where diode based detectors do not perform well, especially in terms of audio distortion.

However one of the slightly strange things I've noticed about the simple FET based infinite impedance detector is the variable audio quality noted, even when using the same type of FET. Some I've built have sounded good and others slightly fuzzy when used with an audio preamp and fed into a high quality audio system. I've been giving this a considerable bit of thought of late and I've wondered if the audio distortion might be a result not necessarily of the detection process itself, but the FET stage in its guise as a 'source follower' audio stage which essentially, it is.

I have long been aware that as a simple audio buffer stage, the FET based 'source follower' can exhibit a considerable amount of audio distortion, and a technique I've long used to greatly reduce this audio distortion is to use a second FET in the source lead of the first FET as a 'constant current source' which serves to 'linearise' and greatly reduce audio distortion in the buffer stage overall. So, to test the theory I built a simple one FET infinite impedance AM detector which worked well, but with just a hint of audio 'fuzziness' on received AM stations. So I added a second FET in the source lead of the first FET wired as a constant current source, taking the output from the source of the first RF detector FET and the source resistor and RF bypass capacitor off the source lead of the second FET 'constant current source'. The result, totally clean audio! The theory seems proved! I call this modified detector the 'Two FET infinite impedance detector'

))) Here's what is sounds like - click to play the audio file (((

Here is the circuit diagram:

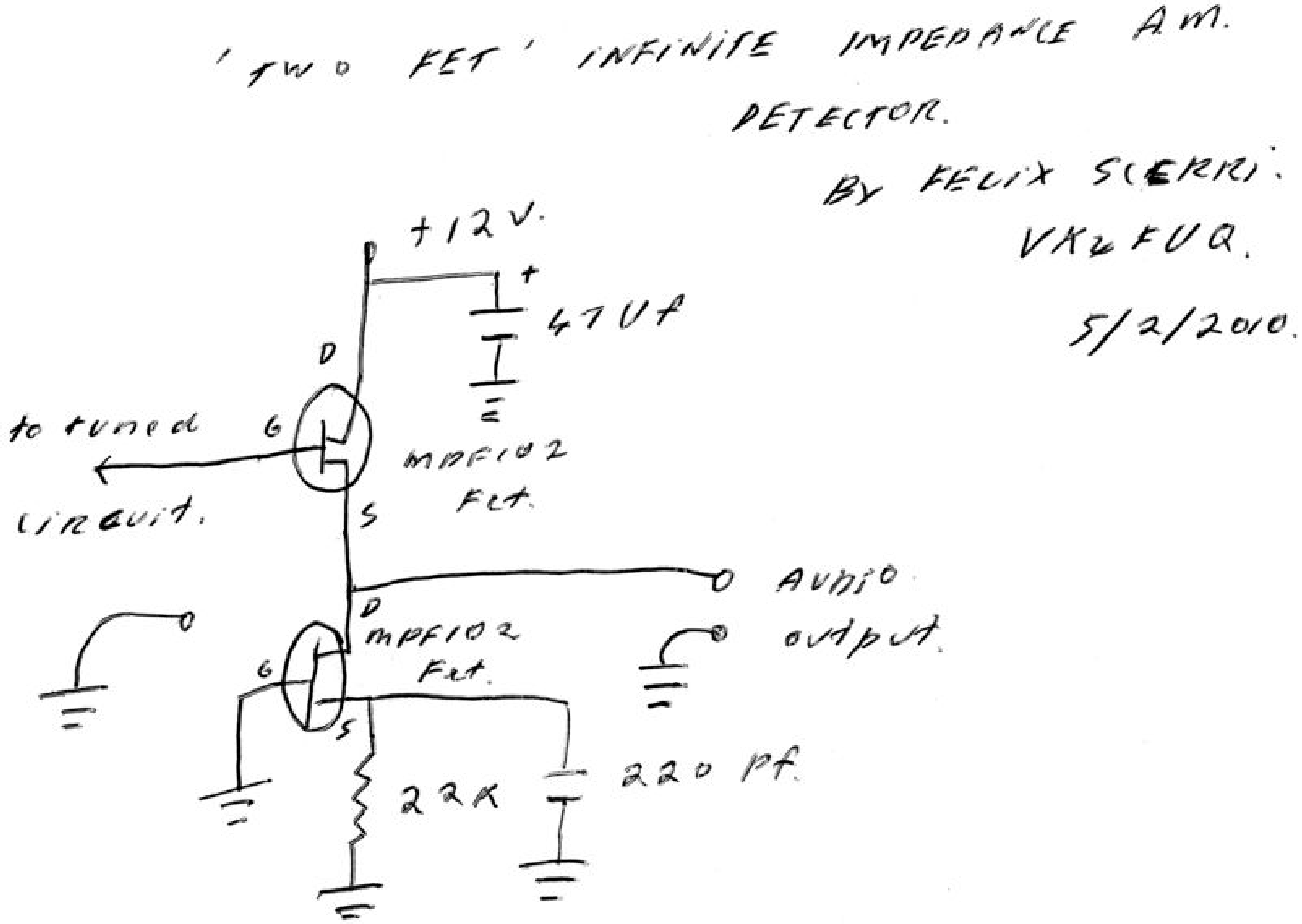

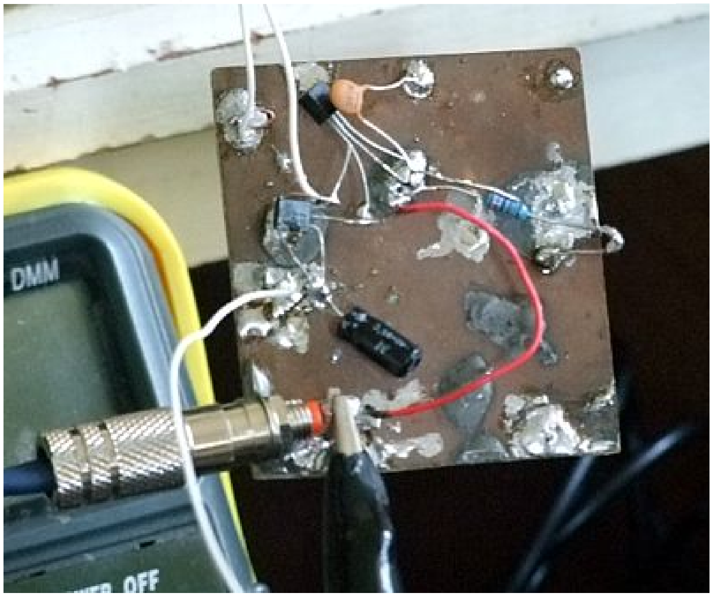

This detector has been a real eye opener for me in terms of its excellent performance, especially considering its circuit simplicity. Indeed in the past I have designed other more complex FET based infinite impedance circuits that do not quite work as well in practical terms as this latest circuit, at least according to my well calibrated ears!

I do not have access to any precise test equipment but my well calibrated ears tell me this 'two FET infinite impedance detector' is a beauty, surpassing practically every other AM detector I've built even at low RF input, and that's rather a impressive claim and the audio quality when used as an AM tuner feeding a high quality audio system is quite remarkable. Possibly the best thing about this detector is its excellent performance under weak signal conditions. Diode based detectors also work beautifully, but the use of an RF stage to ensure detection over a linear portion of the diode's curve is mandatory! This compound infinite impedance detector works beautifully on the sniff of a useable RF signal.

Just add a high Q tuned circuit and that's it!

Felix Scerri
VK4FUQ

A better FET for the 'basic' Infinite Impedance Detector:

Quite recently by accident, I've realised the MPF102 FET that I've long used in my FET based infinite impedance detectors is possibly not the best FET to use. This was the reason why I developed the 'two FET' infinite impedance detector some time ago which works very well.

However I've found the choice of a more suitable FET works beautifully in the basic FET based infinite impedance detector circuit, which has appeared for many years in many editions of the ARRL Handbook.

I use the 2N5457 and others of the same 'family' may be equally suitable, but I haven't tried them! However with a 2N5457 in place of an MPF102, the basic infinite impedance detector has became my AM detector of choice. It works beautifully even at low signal input with lovely and clean low distortion audio along with a very high input impedance for good tuning selectivity. It's a beauty! The basic generic circuit is attached, courtesy of Rod Elliott's ESP website.

73 Felix VK4FUQ
10 / 02 / 2012.

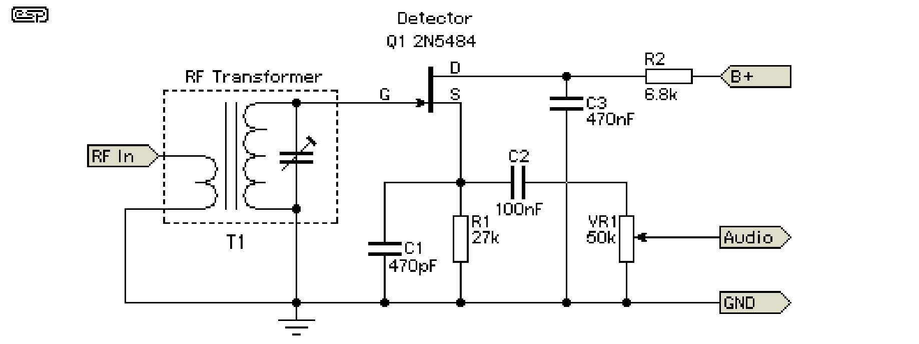

The basic generic circuit is attached, courtesy of Rod Elliott's ESP website
Felix Scerri VK4FUQ

As often happens with me, my renewed interest in FET based 'infinite impedance detectors' of late has led to some interesting new research and I may have considerably improved the 'two FET infinite impedance detector' as a result.

My research suggests that although the use of a CCS (constant current source) reduces audio distortion in an audio stage, the value of the 'source resistor' in the CCS stage is somewhat critical for best results.

By using a potentiometer in lieu of a fixed resistor I have found that a resistance value of around 470 kohms cleaned up all overall audio distortion. I used an MPF102 as the CCS in this circuit. An interesting and worthwhile little circuit refinement.

73 Felix VK4FUQ
21 / 02 / 2012.

A Minimum Component Count High Quality AM Detector

I was generally messing around with various circuit ideas and I came up with this AM detector circuit, a simple diode detector along with a FET stage. It was an attempt to provide good performance along with minimum number of components. Actually I've been pleasantly surprised at the excellent level of general performance and the best of all, it sounds great!

The circuit is quite conventional being a BAT 46 diode detector feeding an MPF102 FET buffer/ common source amplifier stage. I would ordinarily use a FET source buffer stage in this application, but opted to use a simple low gain FET 'common source' amplifier stage instead, with excellent results. I also used a BAT 46 Schottky diode instead of an ordinary germanium signal diode.

This was done for several reasons. Firstly, germanium diodes are now very hard to find but in any case these 'germanium diode equivalent' Schottky diodes are actually a superior diode, having very low noise, almost zero back leakage and an essentially complete absence of carrier storage effects and very good weak signal sensitivity. I call these diodes high fidelity diodes as they sound wonderful as RF detectors.

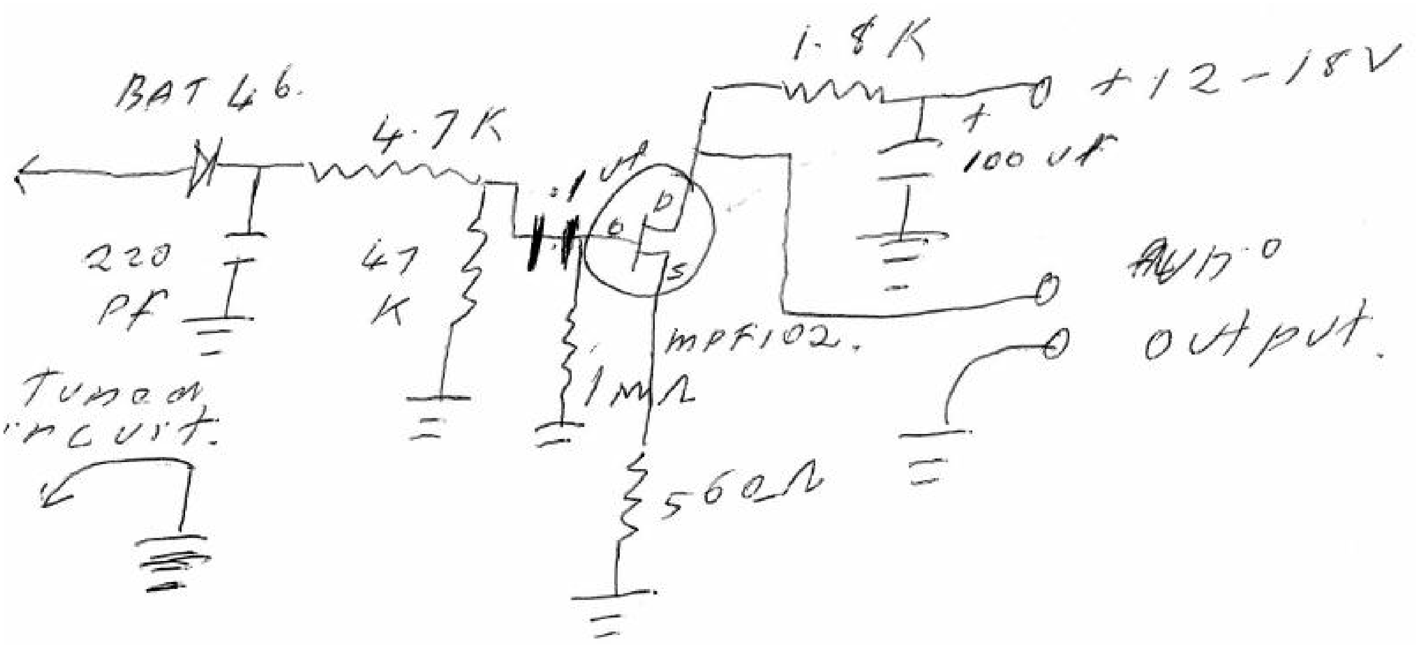

Circuit Diagram of the Minimum Component Count AM Detector by Felix Scerri

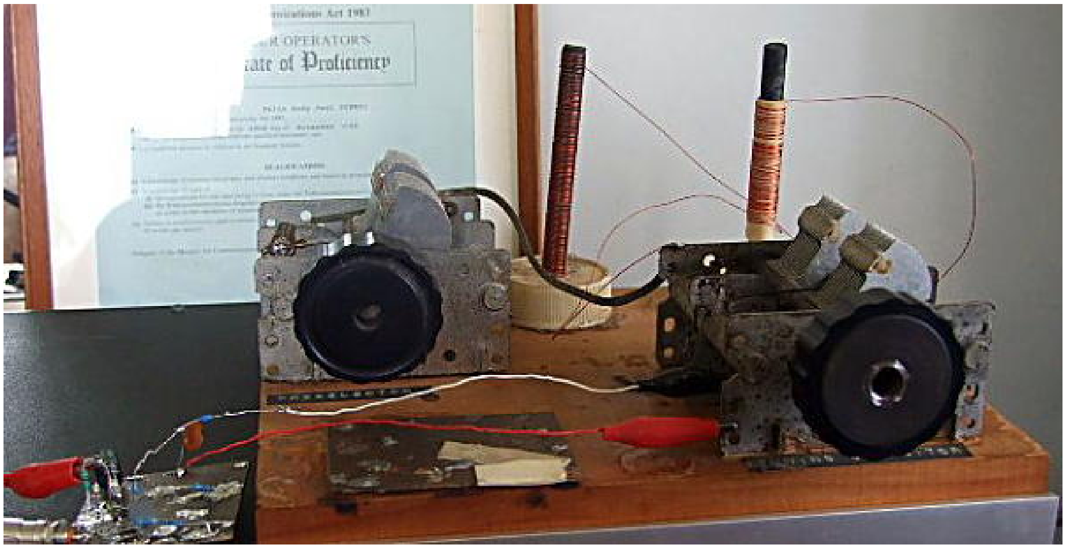

Minimum Component Count AM Detector by Felix Scerri

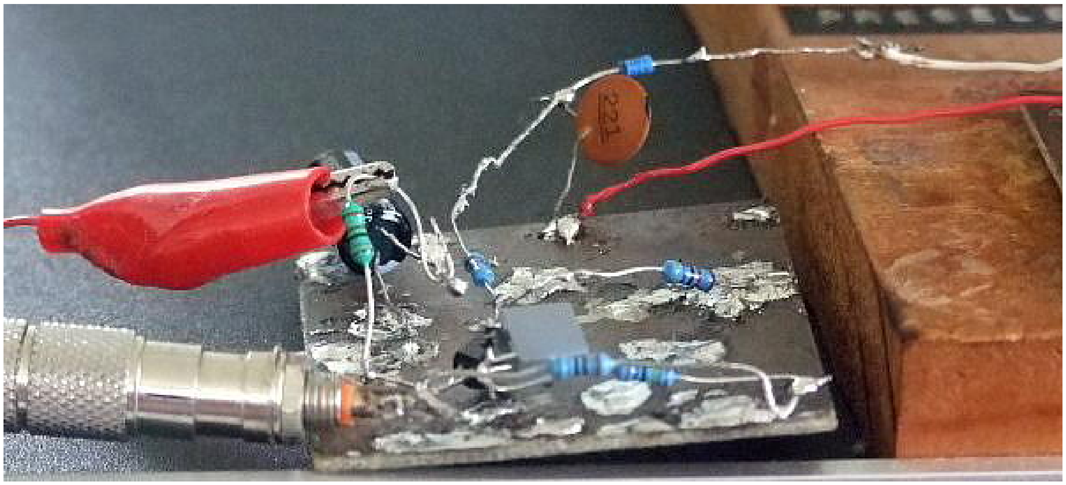

A Closer View

The high impedance of the FET's gate circuit is perfect for optimal buffering of the diode detector, something very important for good low distortion diode detector performance. Apart from providing slight voltage gain, the use of the common source FET amplifier is a new idea, as this prevents the possibility of incidental RF rectification occurring in a FET source follower stage, which can happen. A 1 uf plastic film capacitor may be added in series with the audio 'hot' output lead to block the DC offset out of the FET drain, if required.

Despite no additional RF stage ahead of the diode, audio quality on even relatively weak RF strength stations is actually very good, and of course the audio quality will be even better with increasing RF signal strength, something which will also increase the audio output level. Just on this, for a long time I was somewhat negative regarding diode detectors, as one AM station locally (the strongest one) was always distorted when using a diode detector.

I strongly suspected a transmitter fault, but my complaints were ignored, until one day some time ago when all audio distortion suddenly disappeared! Nothing was ever 'said', but I realised that my suspicions of a long standing transmitter fault were correct, after all!

73 Felix
VK4FUQ
02 March 2012

Voltage Doubler Detector

This is an AM detector circuit that I've long known about and which worked ok, but never seemed to work as well as it should have. However I spent some time late last night trying to optimise the circuit, with some success.

It is a curious circuit being essentially a 'voltage doubler' originally developed for power supply applications, and its use as an AM detector is hard to analyse! It seems that the component values in the circuit are somewhat critical for good performance and if not, the performance is rather 'ordinary'. The circuit that I eventually came up with uses a 150 picofarad 'input capacitor' with a 150 kohm 'load' resistor and loaded into a FET common source voltage amplifier stage (as previously described) through a coupling capacitor with a 1 Mohm input resistance.

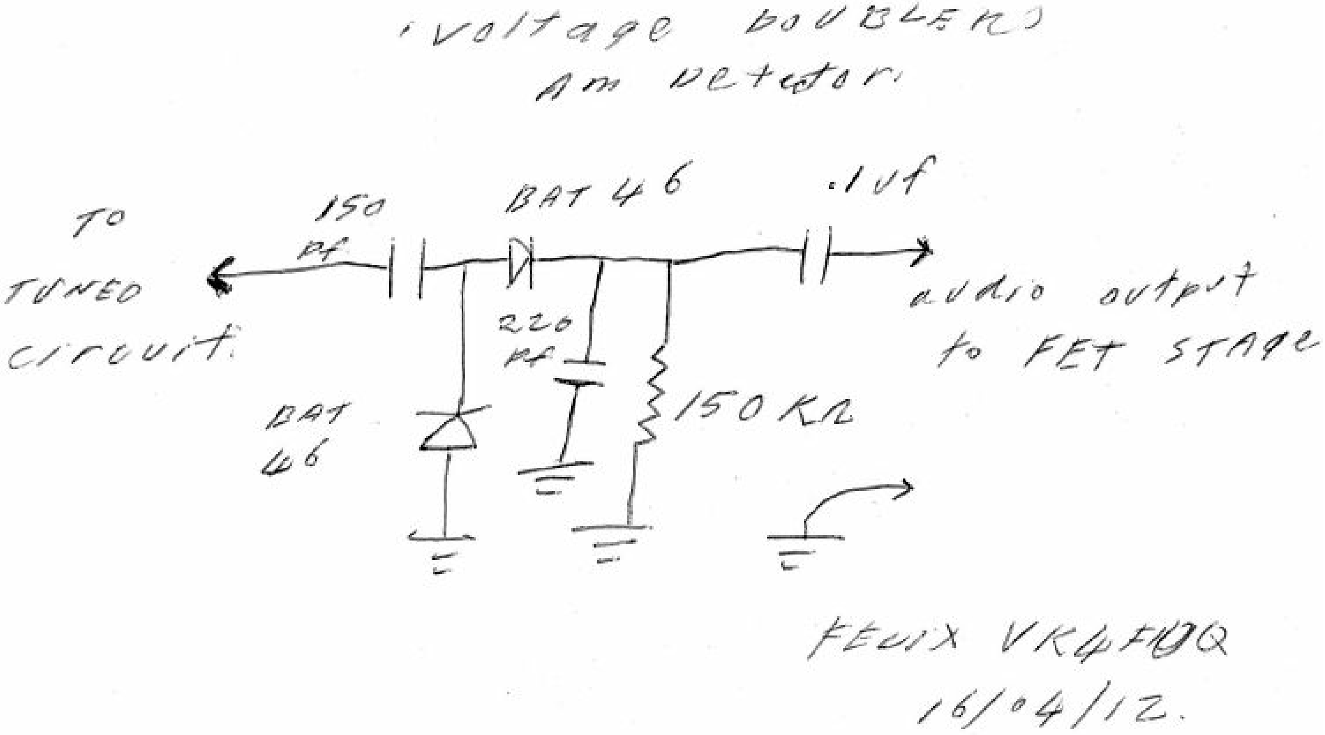

With these circuit values, it all works 'quite well'. Give it a go! It's an interesting AM detector with quite good 'sensitivity' and clean audio quality, and it seems to work well at low signal levels.

73 Felix
vk4fuq.

16th April 2012.

Simple AM detectors: What works best? A practical experimenters viewpoint.

I have written a lot about simple AM detectors for use as tuners for feeding into an audio amplifier, and it has been a long time interest. These days I use either diode based or 'infinite impedance' types of AM detectors. In this location our 'local' AM stations are quite distant and are therefore quite weak in terms of signal strength.

As such I find infinite impedance detectors based on field effect transistors give consistently better results for tuner applications due to their lower apparent overall detector distortion. Diode based detectors are quite 'fussy' as they require both optimal output buffering (AC/DC ratio) and an 'adequate' (beyond the diode knee) level of RF signal injection. http://www.tonnesoftware.com/appnotes/demodulator/diodedemod.html

Diode based detectors will happily 'detect' at very low signal levels, however the (inevitable) audio distortion that results, can be extremely irritating to the ear! Under these conditions I find infinite impedance detectors (even without additional RF preamplification and subject to individual FET characteristics), generally sound 'cleaner' and more pleasant to the ear.

FET's of course require a power source for operation whereas diodes are passive (un-powered) detectors (most of the time), however this is of no real advantage in a tuner application as an 'active' audio amplifier stage will generally be required anyway for audio level boosting, buffering etc.

In the end it will come down to a consideration of prevailing RF signal levels and other related circuit considerations at one's location. If local RF levels are strong, a well designed diode detector will give excellent results. If not, an 'infinite impedance' type of detector is most likely the better option unless one goes towards the option of additional RF preamplification prior to the diode detector.

73 Felix
vk4fuq.

17th April 2012.

ANOTHER DETECTOR by Felix Scerri !

I've tried diode based 'voltage doubler' (or more correctly 'diode integrator') AM detectors before with indifferent results, however the other day, just trying a few ideas I came up with this version that works rather well, with low audio distortion, high audio output and really 'nice' audio quality and the best of all, it seems to work very at very low RF input level.

The two diode 'voltage doubler' detector using two BAT 46 silicon schottky diodes feed directly into a MPF102 source follower stage set at 1 Megohm input resistance. The 'input' capacitor feeding the diodes from a tuned circuit is 68 picofarads.

I have the simple RF filter right on the output of the FET stage. In that respect this circuit is vaguely similar to the old 'Selstead-Smith' valve AM detector of the past. An interesting one! I am very happy with its general performance. Regards, Felix vk4fuq 11/04/2013

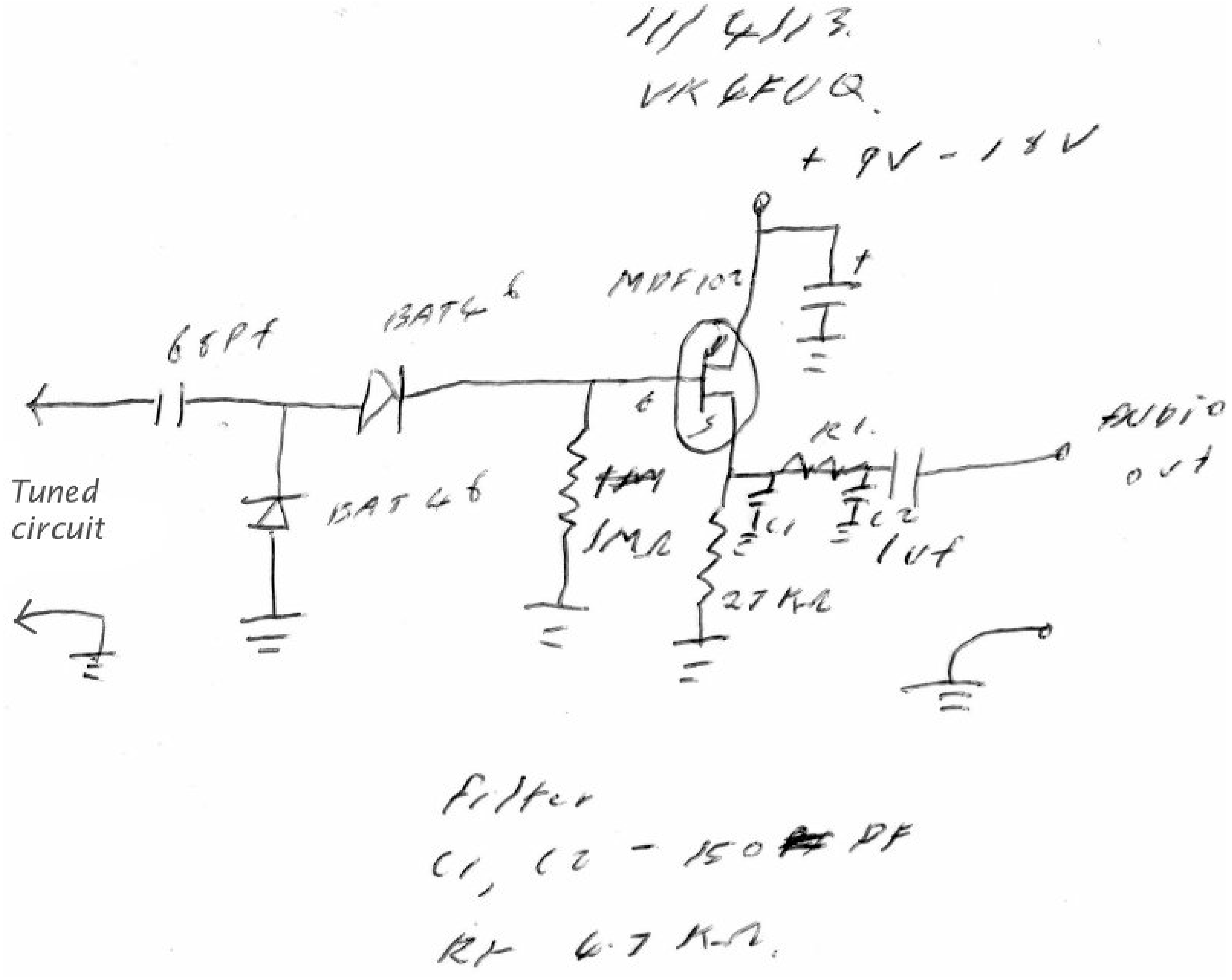
11/4/13.
VK4FUQ.
+ 9V - 18V
MPF102
BAT46
68Pf
Tuned circuit
BAT46
G
S
~~1M~~
1MΩ
R1.
C1
C2
1uf
27KΩ
AUDIO out
Filter
C1, C2 - 150 PF
R1 4.7 KΩ.

Felix Scerri VK4FUQ

UPDATE - JUNE 2013

G’day all, readers may recall the two FET infinite impedance detector I developed some time ago. That circuit worked well, but some samples of the MPF102 regretfully produced distorted output. However a recent discovery has resulted in an improved version that has truly exemplary performance.

The MPF102 is in all honesty a device essentially designed for VHF applications, but is often used in simple audio applications albeit with occasionally indifferent results, due to general device parameter ‘spread’. More or less by empirical trial and error I have found that a simple change in source resistor value to 100 k (from a much lower value), pretty well fixes everything.

In a FET source follower circuit this is interesting as the output impedance is actually a lot lower than 100 k due to the action of the FET’s transconductance. It is similar to the action of a bipolar transistor emitter follower circuit. In fact a simple infinite impedance detector with a 100 k source resistor actually works very well and will be good enough for many applications, however the addition of a second FET as a constant current source does markedly reduce overall audio distortion in the stage overall, and produces very clean and low distortion audio quality and also reduces the output impedance considerably (good for tuner applications), so take your pick!

Regards, Felix vk4fuq 02/06/2013.

UPDATE - JANUARY 2014

A simple modification to the basic FET 'infinite impedance' AM detector that dramatically improves performance.

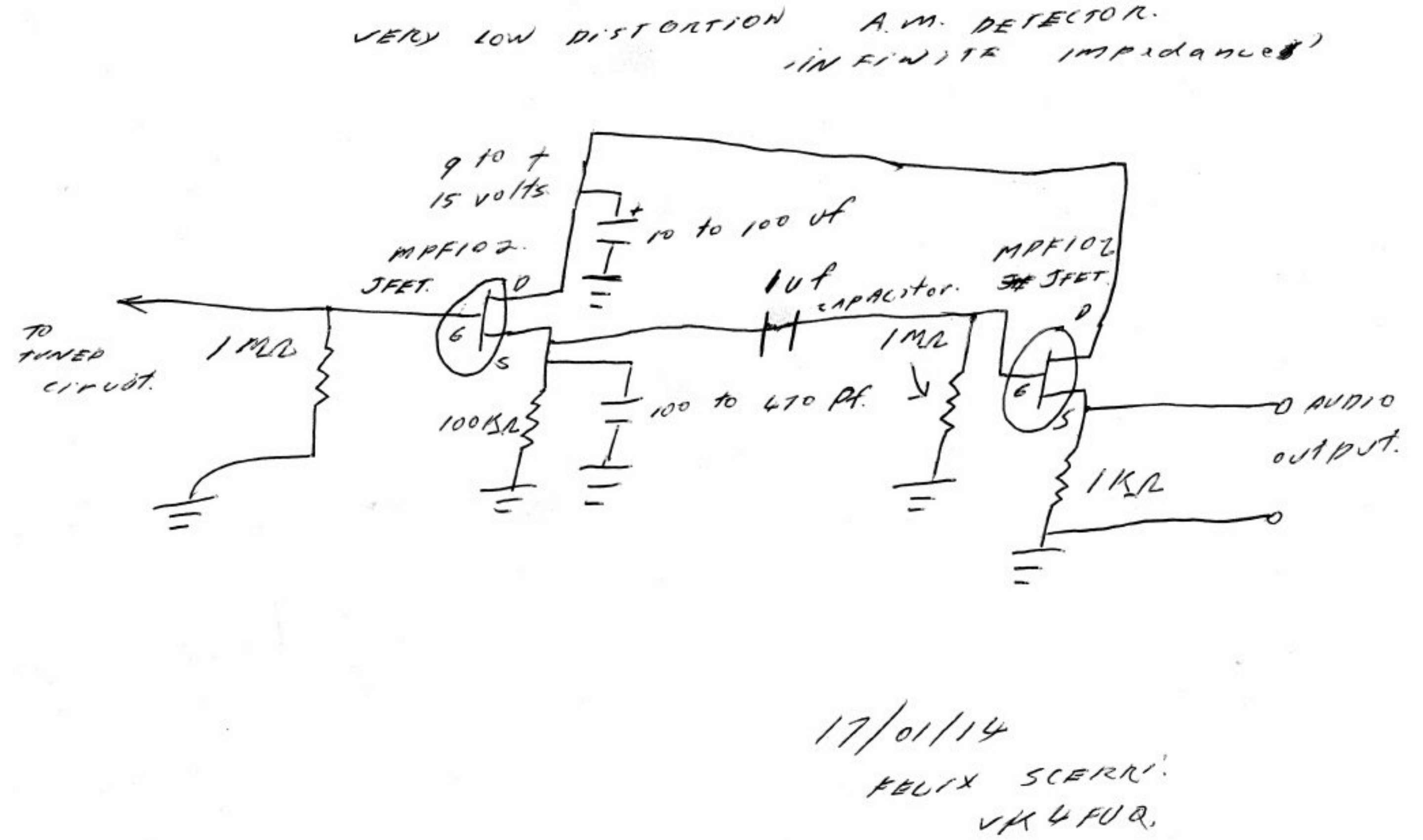

G’day all, over the Christmas break just messing around I worked out (mostly by accident), a simple modification for the simple FET based ‘infinite impedance’ detector circuit that dramatically improves weak signal performance and also greatly reduces audio distortion.

Essentially by the addition of another FET ‘buffer’ stage, another source follower, capacitively coupled from the first detector stage. The circuit is actually a simplified version of the circuit that I described in this link, http://sound.westhost.com/articles/am-radio.htm (figure 6) and testing the two head to head, they both sound superb and the simpler version is actually somewhat easier to build. I cannot get over how low distortion and ‘nice’ the recovered audio sounds. It is a joy to listen to!

Regards, Felix vk4fuq 12/01/2014.

Diodes for ‘weak signal’ crystal set applications.

As I am primarily interested in using crystal sets as AM tuners for feeding into a preamplifier/amplifier and loudspeakers, the actual type of diode can be relatively critical. In a strong signal area, not so much, but in a weak signal area such as where I reside, definitely. A diode with a good ‘square law’ performance (the area below the ‘knee’ in the diode curve), generally results in much cleaner and lower distortion than other good performing diodes, and believe me diode distortion under weak conditions, even with optimised diode ‘buffering’ used is very nasty sounding to the ear!

Testing many, many diodes in actual working crystal sets and listening critically to the audio output, it seems to me that the best diodes to use under weak signal conditions are the so called ‘gold bonded’ germanium diodes. I have sampled many different gold bonded germanium diodes and they all work well in this specific application, although sometimes the rectified output voltage may not be as high as other germanium or silicon schottky diodes, however the audio quality is much cleaner and shows much less apparent distortion!

I have tried OA5, OA47, IN141 and several ‘CG’ gold bonded germanium diodes with consistently excellent results. Other ‘ordinary’ germanium diodes may also work well under weak signal conditions, but they will need to be tested individually to check actual performance in a working circuit. One thing that I have noticed about germanium diodes is that due the ‘point contact’ nature of their construction, even diodes of the same specific type can exhibit rather different levels of performance!

The otherwise very good BAT46 silicon schottky diode works extremely well at good RF input level but not so well at weak RF input, especially when heavy broadcast ‘processing’ (commonplace these days), is used. New ‘gold bonded’ germanium diodes are probably no longer made although I am aware that they can be purchased through vendors over the internet. Apart from that, they may be found in old gear. As stated earlier, ordinary germanium diodes may be quite good but will need to be checked individually. Diodes are complex things!

Felix (vk4fuq). 29/01/2014.

* * * * * * * * * * * * * *

Just when I thought that I’ve tried everything I realised that I’ve not tried a voltage doubler (diode integrator) detector with an RF stage in front. It is an unfortunate fact that all diode detectors ‘need’ a good RF injection level (and buffering) or audio distortion becomes ‘bad’ (an understatement!)

Consider yourself very lucky if you live in a strong signal area! (unfortunately I don’t). Anyway I quickly built a couple of prototypes which worked ok but not as well as hoped. Initially I tried using a source follower buffer (no voltage gain), but converting it to a common source amplifier (with appreciable voltage gain), seemed no better, which was strange.

After staring at my prototype for what seemed like forever with my magnifiers on, I (finally) realised my mistake. A common source stage takes its output off the ‘drain’ terminal of the FET, not the ‘source’ terminal as it does with the FET source follower buffer stage. I shifted one wire, and all worked as expected and it sounds fantastic! All diode detectors (one diode envelope detectors, doubler detectors etc), need good RF input and a simple untuned FET RF gain stage works very well. As to the sound, it sounds great.

73 vk4fuq 10/03/2014.

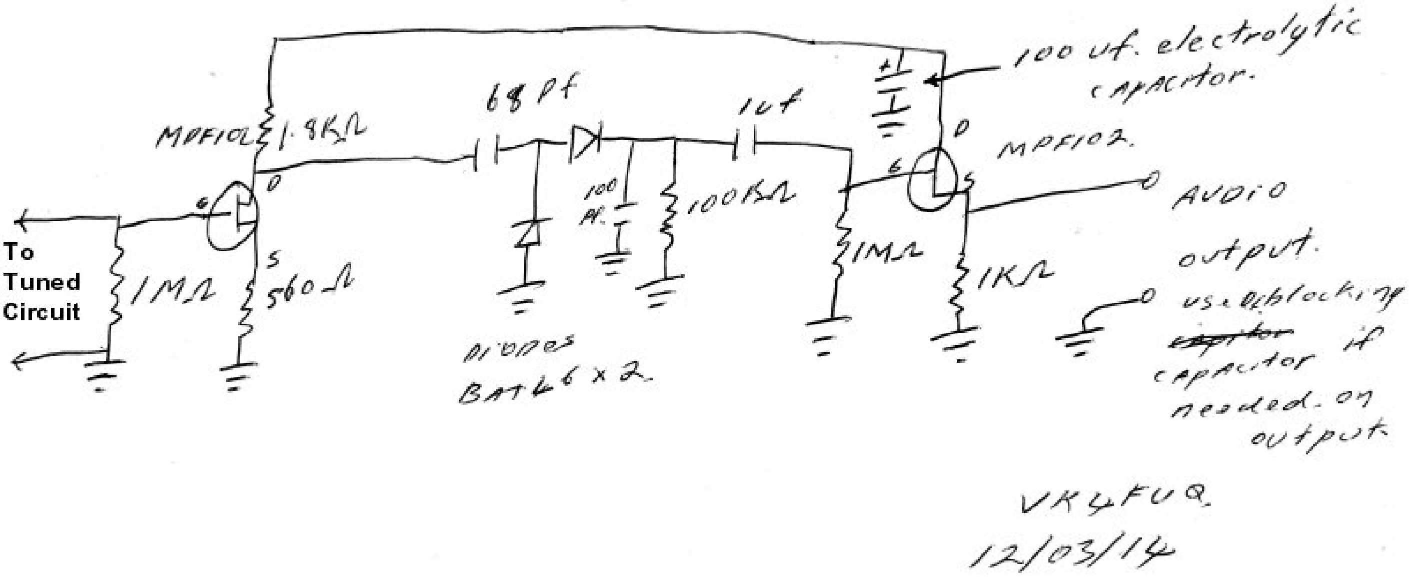

Felix Scerri's FET Crystal Set with Voltage Doubler

A 'nice sounding' AM detector - update October 2015

Looking through these pages the other day I was slightly shocked to realize that I’ve completely forgotten about some of the circuits that I’ve contributed in earlier days! They all work pretty well, however in my advancing older age and a little like my tastes in high fidelity generally, I’m starting to show a particular preference for ‘nice sounding’ bits of audio gear.

I guess that this also means low noise/low distortion too, although with AM broadcasting, at least in this country and probably elsewhere in the world too, the very common use of broadcasting ‘processing’ tends to make it hard for AM detectors generally, often resulting in a ‘hard/compressed’ sort of sound although still low in distortion, is not ‘nice sounding’, if that makes any sense!

Well of all the AM detectors that I’ve tried and/or developed, only one sounds ‘nice’ when confronted by heavy broadcast processing and that is an AM detector ‘based’ on a voltage doubler/diode integrator circuit (similar to the above circuit, actually the 68 pf capacitor should be changed to .1 uf, for slightly greater output).

It has taken me a very long time (years) to ‘optimise’ this circuit, but as it presently stands this is my favourite ‘nice sounding’ AM detector, and gets most use for general high quality listening on the AM broadcast band. It sounds really good! It somewhat reminds me of an old OA47 gold bonded diode detector from years ago before such abhorrent ‘processing’ became commonplace!

Ah yes the OA47….now that is a lovely sounding detector diode!

Regards, Felix vk4fuq. 17 / 10 / 2015.

* * * * * * * * * * * * * * *

That's it for crystal sets. I hope you try building one, it's easy and great fun!

See some useful links below....

73's
Mike

[Crystal Sets (Part1)] | [Build Your Own Crystal Set (Part 2)]

[Spider's Web Crystal Set (Part 3)] | [Crystal Set By Kenneth Rankin (Part 4)] | [Crystal Radio Links]

Amateur Radio

MØMTJ Visit my amateur radio pages MØMTJ

Visit my Amateur Radio Pages >

No AM radio stations or transmitters in your locality or country?

Spitfire
AM Transmitter

Has your local medium wave broadcast station closed or been moved to VHF/FM or Digital? Don't worry. You can still build and experiment with crystal sets and TRF radios by also buying or even building a simple low power AM transmitter. So, not only can you use your crystal sets but you can also run your own radio station that can be heard in and around your home - playing the music or programmes that you want to hear!

SSTRAN AMT3000 Superb high fidelity medium wave AM transmitter kits from SSTRAN. Versions available for 10kHz spacing in the Americas (AMT3000 or AMT3000-SM) and 9kHz spacing in Europe and other areas (AMT3000-9 and AMT3000-9SM). Superb audio quality and a great and well designed little kit to build: http://www.sstran.com/pages/products.html

POWER
GAIN
MODULATION
COMPRESSION
SSTRAN
AMT3000

http://www.sstran.com/

Other AM transmitters available:

Spitfire & Metzo Complete, high quality ready built medium wave AM Transmitters from Vintage Components: http://www.vcomp.co.uk/index.htm Vintage Components offer a choice of the high quality Spitfire and Metzo transmitters:

SPITFIRE AM Medium Wave Transmitter with 100 milliwatt RF output power:

http://www.vcomp.co.uk/spitfire/spitfire.htm

Spitfire
AM Transmitter

METZO AM Medium Wave Transmitter with built in compressor:

Gain
Comp
Mod
Metzo

http://www.vcomp.co.uk/metzo/metzo.htm

AM88 LP A basic AM transmitter kit from North County Radio.
http://www.northcountryradio.com/Kitpages/am88.htm

LINKS: Fine links to more Crystal Radio websites here

Component Suppliers: Links to electronic component suppliers here

Amateur Radio
M0MTJ Visit my amateur radio pages M0MTJ

Visit my Amateur Radio Pages >

Crystal Sets (Part1) | Build Your Own Crystal Set (Part 2)

Spider's Web Crystal Set (Part 3) | Crystal Set By Kenneth Rankin (Part 4) | Crystal Radio Links

^Top Of Page

Home | Contact | Site Map | Radio Stations & Memorabilia | Amateur Radio

Crystal Sets Introduction | Resistor & Capacitor Conversion Tables

www.mds975.co.uk © 2004 - 2015

May 9 2018, 4:29:05 pm

MY WEATHER STATION, BY DIMITRIS

Location: Peristeri of Athens, Greece
Position: 38° 00.58' NORTH / 23° 42.04' EAST
Elevation: 63 meters
contact me at meteo@otenet.gr

Click to enable Adobe Flash Player

Home Page | Lightnings | External Data | NEW My Projects | Knowledge | Friendly Sites | Photo Album | Guestbook | About

SWITCH TO CONTENTS

RF field strength meter version 1.

I made a RF simple field strength meter for test the remote control from doors ,alarm e.t.c at frequency VHF-UHF, i used one μA meter ,2 diodes ,2 ceramic capacitors,a potentiometer,a box ,and one telescopic antenna.When press the button at remote control,output there is small current RF,where with this simple circuit ,easy i can to know if remote controls it work or no.

Here it is the electronic design

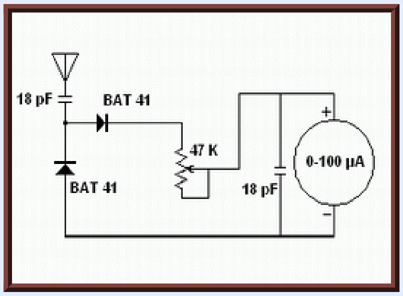

Here it is the electronic material,without yet the potentiometer 47 Kohm, 2 capacitor 18 pF,2 diodes BAT 41,and a μA meter

Here it is the plastic box,i have open holes for μA meter ,and i have touch the telescopic antenna

Here with soldering iron the components it is ready

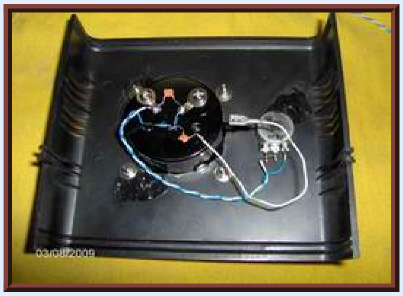

Here it is the sensitivity test

Here it the test with remote temperature /humidity sensor

It is ready

Ready with full open telescopic antenna

Click over to photo for to see a video about it

Click here for to see video
03/08/2009
Click here for to see video

RF field strength meter version 2.

I made a second version.The electronic design i saw from Here.

Here it is a electronic design.I left the old potentiometer 47 kohm and the μA meter 100 μA.

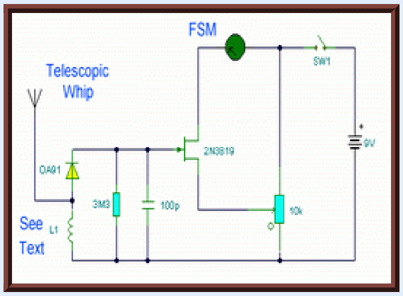
FSM
SW1
Telescopic Whip
OA91
2N3819
9V
3M3
100p
10k
See Text
L1

Here i added this electronic material.One transistor JFET 2N3819,one diode germanium AA119,one cap 100 pF,one resistor 3,3 Mohm,one clip 9 volt battery,one battery 9 volt,one switch on-off,and one inductor over ferrite 1,82 μH.

I use a coil over to ferrite about 1,82 μH (i removed little coil for it is about 1,82 μH)

Here i added a switch on-off,and i put the aerial antenna to other side.

From the other side.

I finnish with the soldering iron material.

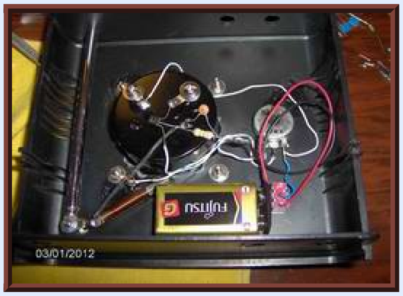

RF field strength meter,It is ready

Here i make a test with out remote sensor from weather station at frequency 433 Mhz.

Here i make a test with remote control 433 Mhz

Here i make a test with mobile phone 1,8 Ghz.

Click here for you see a video

04/01/2012

Click here for you see video
04/01/2012
Click here for you see video

PHGN 300
Crystal Radio
(in 45 minutes)

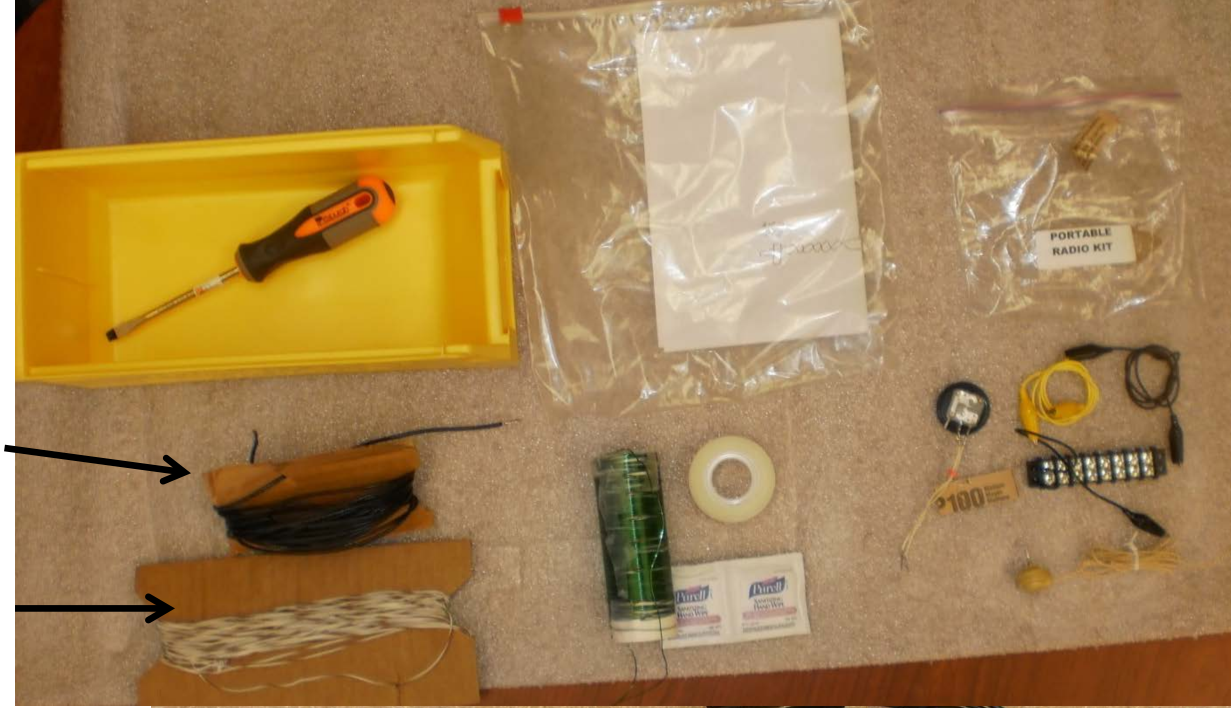
PORTABLE RADIO KIT
100

Ground Wire

Antenna
Need double
length

PORTABLE RADIO KIT

Please put away like
this when done!

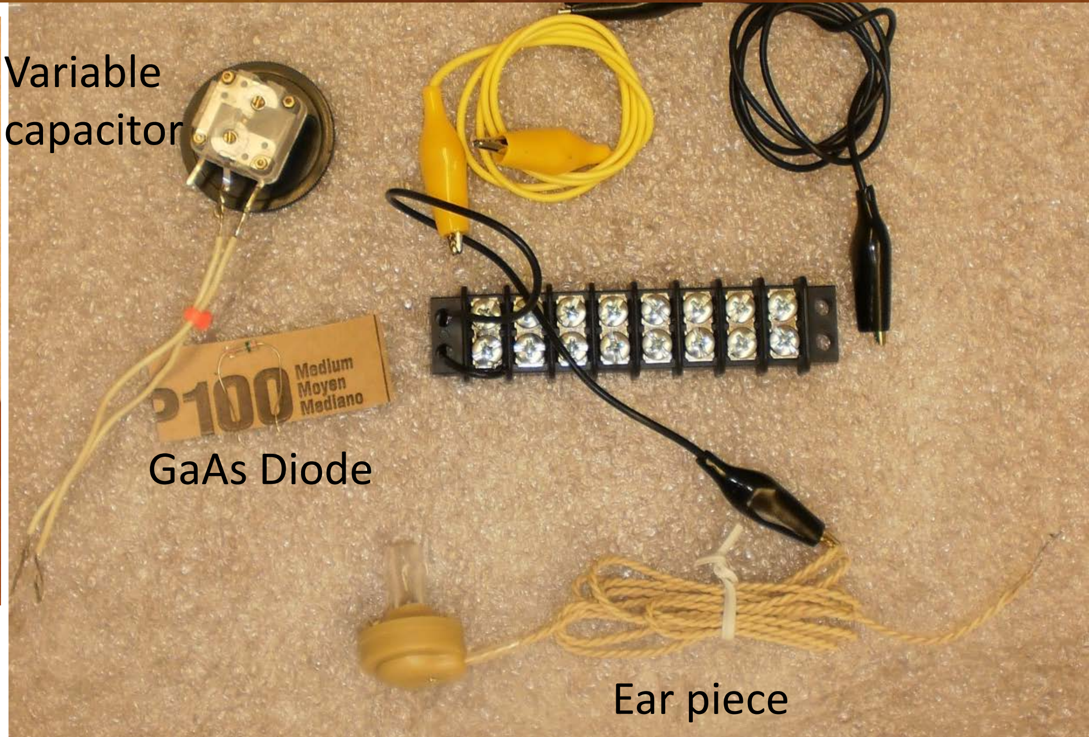
100 Medium Moyen Mediano

Variable
capacitor

GaAs Diode

Ear piece

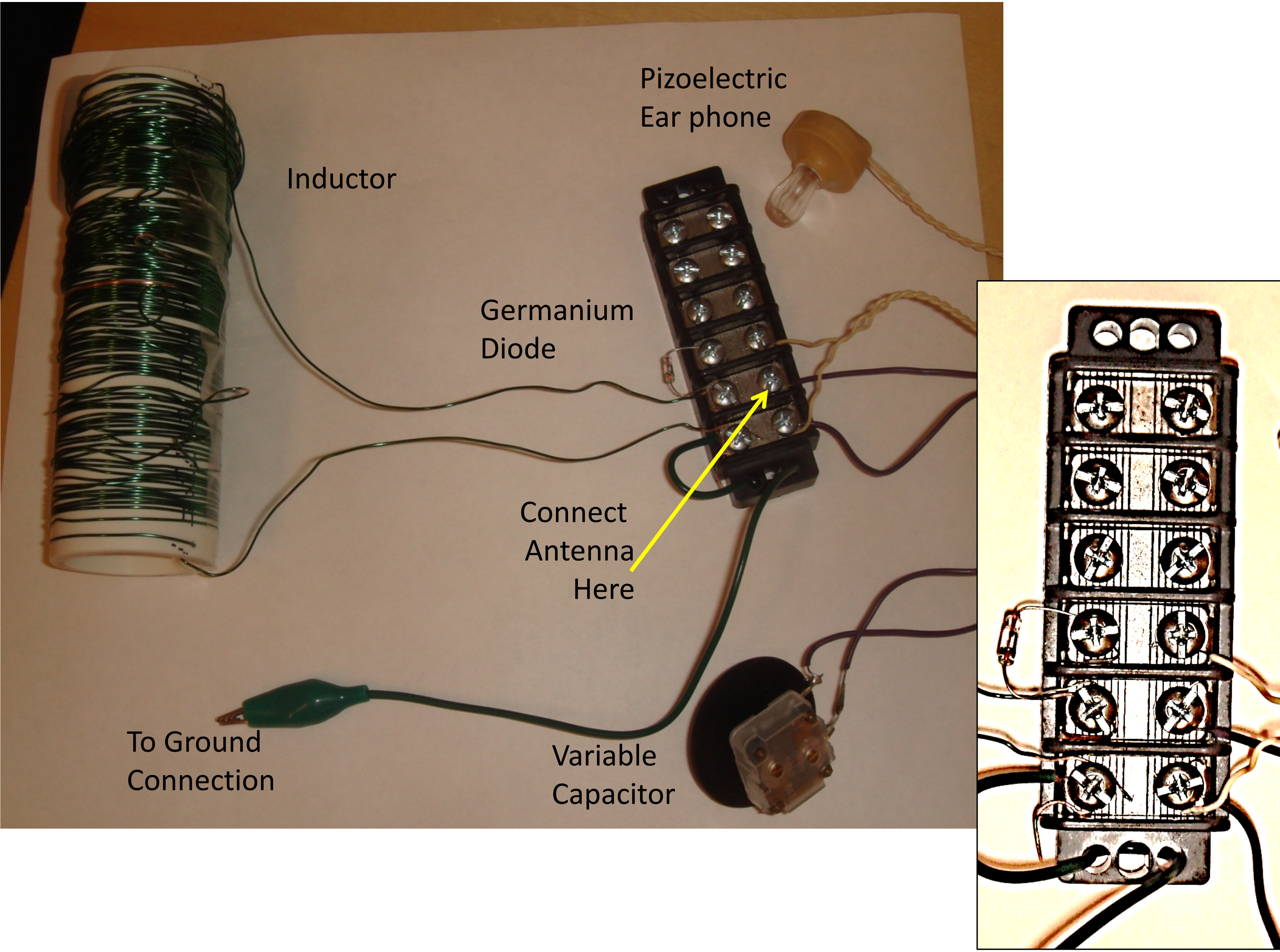
Pizoelectric
Ear phone
Inductor
Germanium
Diode
Connect
Antenna
Here
To Ground
Connection
Variable
Capacitor

Crystal Radio in 45 minutes.

Work in groups of 4.

2 people should assemble the radio and connect the ground wire.

2 people should string the antenna

If you get your radio to work using the GeAs diode,
Try the Gallinium + Cat Whisker

Return everything the way you found it before leaving

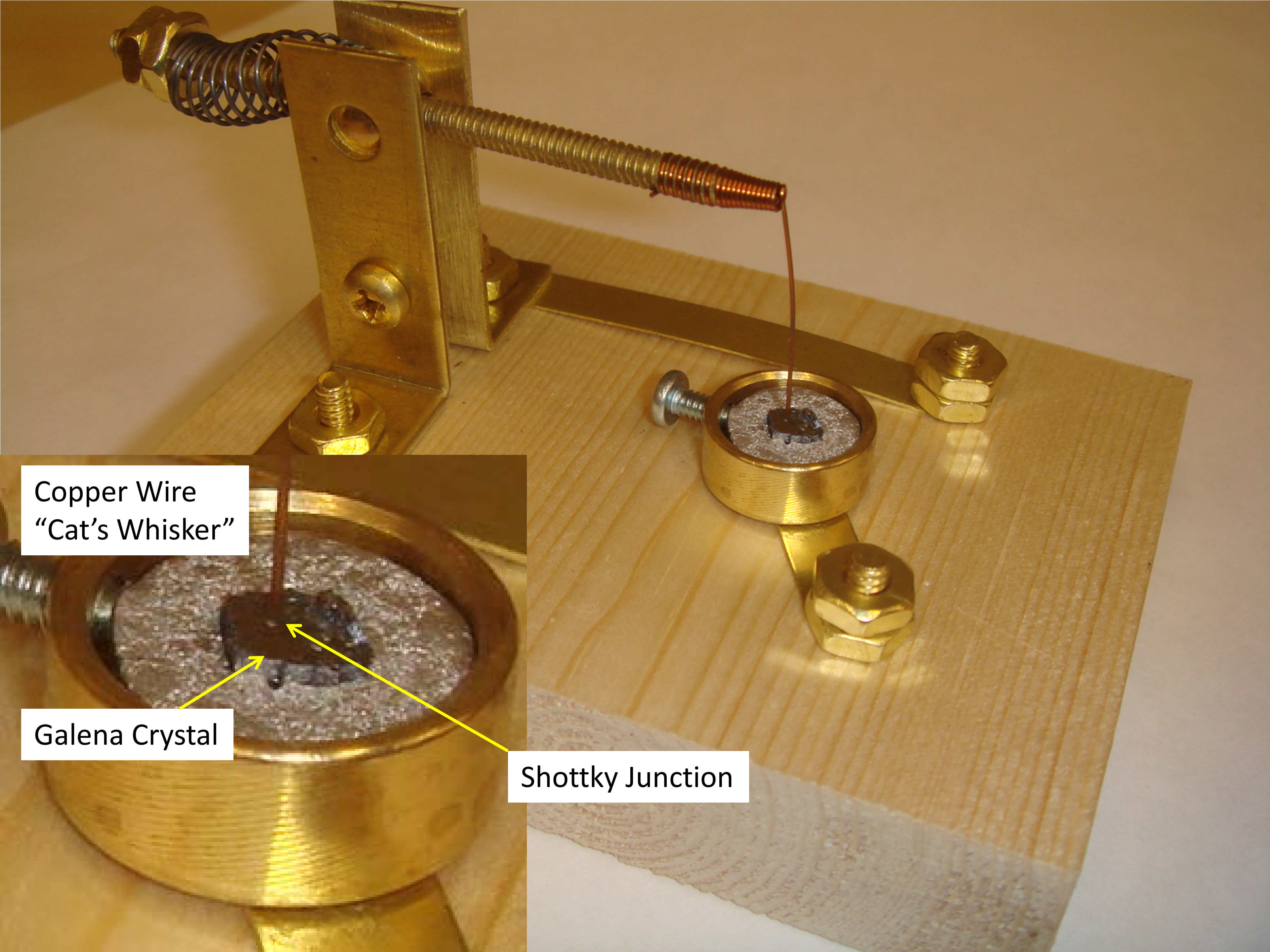
Copper Wire
"Cat's Whisker"
Galena Crystal
Shottky Junction

The radios **must** have a good ground connection and a long antenna. Long means 100’-200’ of antenna wire outside. 200’ is better

Two possible routes are shown here. Use one.

Attach to tree branches you can reach.

When crossing a sidewalk, make it high so pedestrians don’t have to duck (they won’t).

You will take antenna down when class ends.

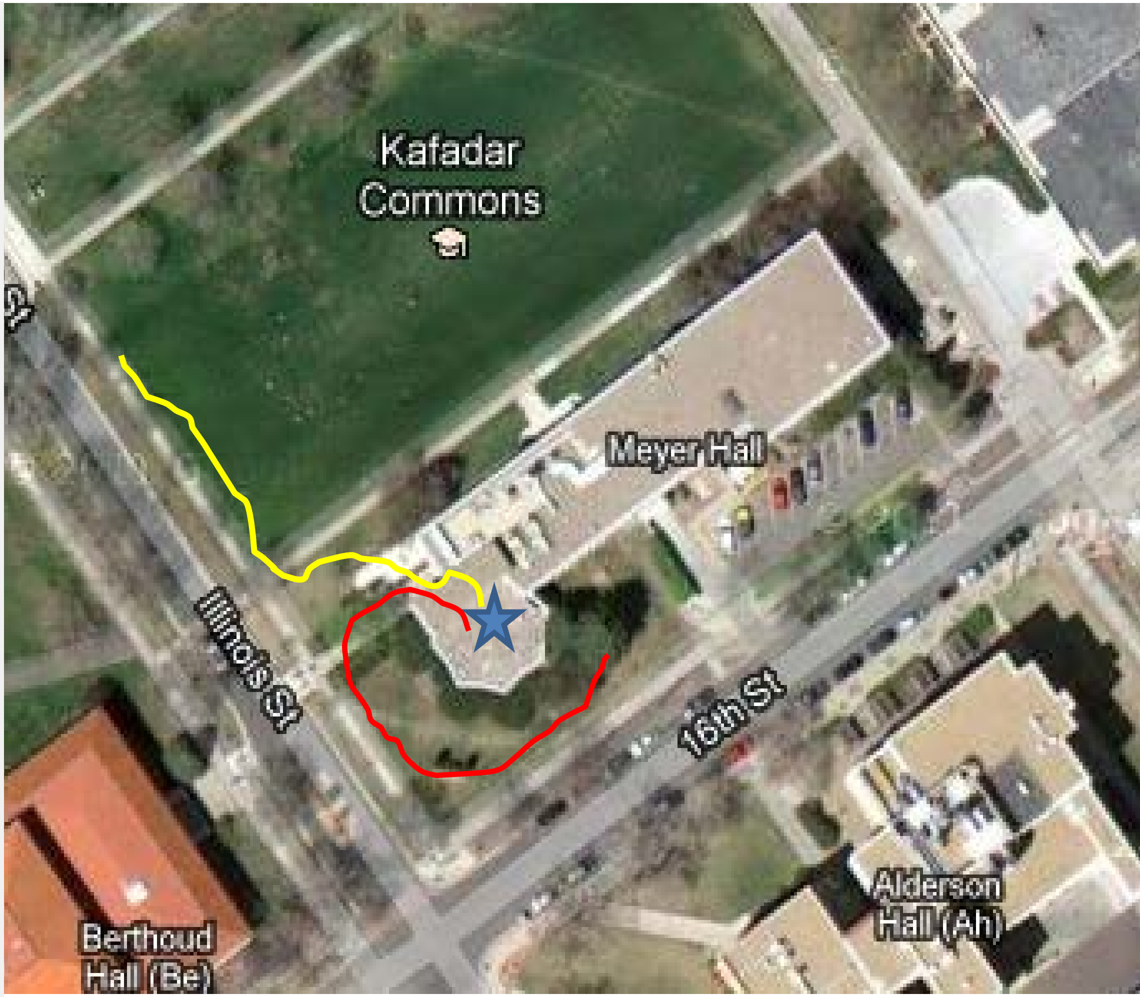
Kafadar Commons
Meyer Hall
Illinois St
16th St
Alderson Hall (Ah)
Berthoud Hall (Be)

rexresearch.com

Roy J. MEYERS

Absorber

How To Patent Your Idea
3 Easy Steps To Patent & Make Money Get A Free Patent & Invention Kit
www.Patent.InventionHome.com

Superior X-Ray Solutions
Performance, Reliability & Speed from the world leader in SCD
www.Bruker-AXS.com

Do You Have an Idea?
We Can Help Turn Your Idea Into A Real Product. Get Free Information!
NewProductHelp.com

Buckyballs Magnetic Toy
Get the Gadget on Everyone's List. Best prices on this year's hot toy!
www.getbuckyballs.com

Car Door Magnets: 50% Off
Magnetic Car Signs $4.99. No Mins. Full-Color. Upload Photos & Logos!
www.Vistaprint.com

Power Your Own Home
Reduce Utility Bills by 50% Lower C02 Emissions by 1/3
www.ClearEdgePower.com

Ads by Google

Renewable Clean Energy
Compact Electricity & Heat Unit For Homes & Small Businesses
www.ClearEdgePower.com

Do You Have an Idea?
We Can Help Turn Your Idea Into A Real Product. Get Free Information!
NewProductHelp.com

Patent Attorney/Engineer
Over 36 years experience. Staff available 24/7 by phone.
www.invention.net

C. Edholm: "Picks Power from Air" (*Technology World Magazine*, 1912 ?)
Dr L. Hirshberg: "Electricity from Air New Great Discovery (*Modern Electrics*, 1914 ?)
Roy J. Meyers: British Patent # 1098 (1913) -- Improvements in and Relating to Apparatus for Producing Electricity (PDF Format, requires Adobe Reader)
R. J. Meyers: British Patent # 1098 (1913) -- Transcription w/ enlarged figures
Notes & Comments
R. J. Meyers: British Patent # 1098 (1913) -- JPG version

***Technology World Magazine*, p. 279-281 (Year unknown, apparently circa 1912; another article about Meyers appeared in the November 1912 issue of *Electronic World*)**

"Picks Power from the Air"

by
Charlton Lawrence Edholm

A remarkable scene took place in the legislature of Arizona this spring when the lawmakers enthusiastically voted for the parole of a certain convict in the State penitentiary, granting him a leave of absence for 30 days and by means of private contributions raising a fund to defray his expenses to Washington DC and return.

The prisoner, Roy J. Meyers, is serving a 3-1/2 year sentence, but in spite of the fact that he bears the stigma of a convicted lawbreaker, he has demonstrated that a convict can be a useful member of society. During his imprisonment he perfected an electrical device of such original character as to arouse feelings of wonder and skepticism until experts had seen it in actual operation. It is a device to draw electricity from the atmosphere for light and power, and the 30-day parole was granted in order that the inventor might protect his rights through the patent office at Washington.

With the acquiescence of the legislature, Governor Hunt granted the parole and the prisoner was allowed to go free without any guard or any assurance but his word of honor that he would return. Two days before the period had elapsed, Meyer again presented himself before the governor, having accomplished his mission, and then returned to the penitentiary at Florence, where he continues to serve his sentence.

This, in brief, is the picturesque story which has called attention of the civilized world to a newly discovered electrical genius, and to another feature of the case which is of equal importance and human interest; namely, the enlightened policy pursued by our youngest State in its treatment of convicts...

Before entering the prison, Meyer had already applied for various patents, among them one for an improved trolley wheel head which prevents the trolley wheel from jumping the wire. Meyers had a conference with Superintendent Sims and Parole Clerk Sanders, and it was to these gentlemen that the inventor first explained the principles of his new device for securing electrical energy from the air. The officials were willing to give the man the opportunity to develop his plan and a little wooden building outside the walls was turned over to Meyers and was fitted up as a workshop and a laboratory. The first demonstration of the new apparatus was made shortly thereafter, the electricity drawn from the atmosphere being used to spark the gas engines of the pump house, and although the device was crude yet it did the work, and removed the doubts of his friends. Furtter development of the "absorber" followed, and his second model was constructed, and developed 8 volts. The machine came to the attention of the remarkable woman who brought his name before the legislature.

This was Miss Kate Barnard, State Commissioner of Charities and Corrections of Oklahoma, who was a guest of Mr. Sims, while studying prison conditions. She saw the machine at work, became familiar with the facts of Meyers' case, and was impressed by his rather blunt and unaffected personality, for Meyers has nothing of the polish or glibness of the poseur. He is a simple, earnest student of mechanical problems and not the sort of man to make a sentimental appeal for sympathy because of any grace of person or manner. Therefore it was the value of Meyers' invention, together with his essential integrity (in spite of his lapse) which so strongly impressed Miss Barnard that when she appeared before the Arizona legislature not long afterwards, addressing that body on the need of enlightened legislature along the line of her own work, she told the story of Roy Meyers and his epoch-making invention.

So, early in May, Meyers set out for Washington, unaccompanied.

In his own words: "When I arrived in Washington and laid my plans before the patent office experts, they merely smiled and told me that I would have to build a model and demonstrate my claims --- that it seemed strange that I, unknown as I am in the electrical world, should have accomplished the things for which Edison, Tesla and other experts have been striving for years.

"They could grasp the meaning of my drawings nor the explanation I tried to make to them. There was little time to spare, as I had only 20 days left of my leave, but I set to work in a few days was able to take a crude model around to the patent office to make a demonstration.

"Arriving at the patent office I telephoned to a friend who had been so kind as to introduce me and aid me in reaching the proper officials. The absorber was hoisted on two short poles and made to work. While they were as yet unable to understand the principles involved and hardly willing to believe their eyes, they were forced to admit that I had something new and different, and they told me that there would be no further objection; that I might file my application without further delay.

"I hope to construct my first large machine right here in Phoenix. I feel grateful; to Governor Hunt and others for what they have done for me and to the help they have given in securing protection I might not otherwise have had, and I am desirous of demonstrating this gratitude. I am going back to Florence today to resume the serving of my sentence, which will expire in 10 months. Then, here in Phoenix, I will begin the work of making my machines."

While there are some details of the device which the inventor refuses to make public, yet there are many general features that may be explained. It is planned that the machine, to be set up in Phoenix, will generate sufficient power to light the city, and will consist of a 200 foot tower upon which is placed the "absorber". The latter consists of a series of magnetized steel plates set in a circle (the manner of preparing them is kept secret) and this mechanism attracts the electricity from the atmosphere. This is carried by wires to a transformer in the engine house below and thence is applied to produce either power or light after the usual manner.

In an authorized statement Meyers says: "The flow of electricity is constant. When it emerges into the transformer it is in the form of a direct current. It will absorb the electricity day and night and will work whenever the wireless will work. I can put up a plant to supply such a building as the Adams Hotel for about $1500, and one of the principal items of the expense is the cost of the towers, the wires, the magnetizing of one set of plates, which is part of the secret of the treatment which makes it respond to the accumulations of the atmosphere.

"For use in the case of an electrical storm I have made what I call a modified form of circuit breaker, such as is commonly used as a lightning arrester on telegraph lines. In case of a storm the accumulator would suddenly become overcharged, possibly, and as the electricity would not of itself flow back into the air, the result might be disastrous. So I send it down into the ground, whenever the voltage rises above a certain amount."

Roy J. Meyers & the "Absorber"

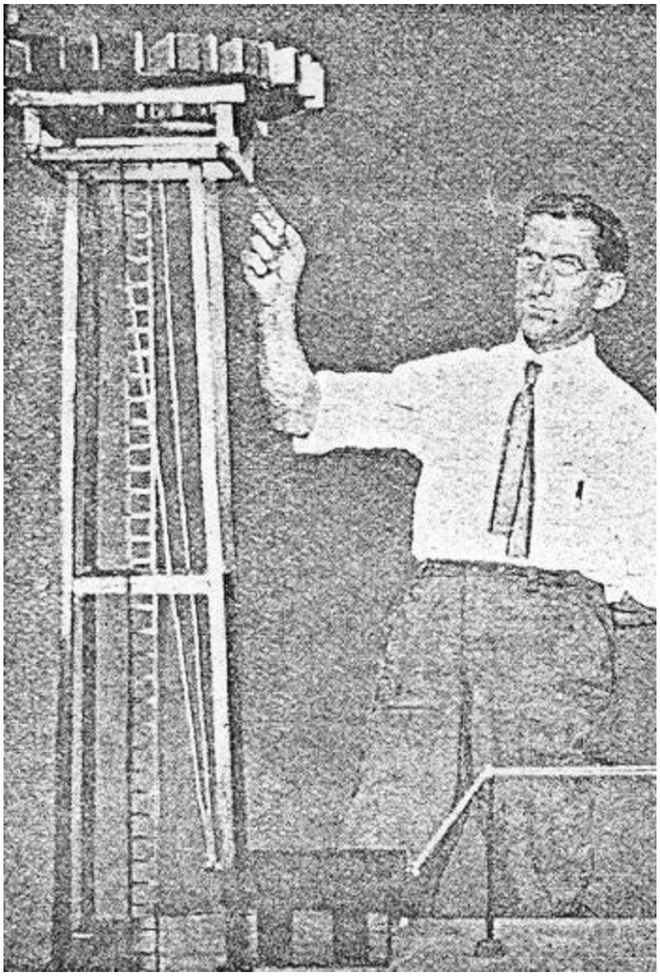

Modern Electronics **(1914 ?)**

Electricity From Air New Great Discovery

by

Dr Leonard Keene Hirshberg

Working quietly in the heart of Baltimore for weeks on an invention which some critics say will revolutionize the method of converting electricity to practical use has been Roy J. Meyers, who like Benjamin Franklin, extracts the electric current from the air.

Mr Meyers invention was made last summer while he was confined in the penitentiary at Florence, Arizona. His first finished apparatus was made in Baltimore.

A practical, unlettered electrician, Mr Meyers, while in Arizona, was arrested on a comparatively minor charge and sent to the penitentiary. There he was placed in charge of the prison electrical plant, and there he says he made his discovery that the current which the civilized world is beginning to use most extensively for light and power could be transformed from the atmosphere without the aid of moving machines or batteries.

Miss Kate Barnard, Commissioner of Charities and Corrections, of Oklahoma, hearing of Meyers' invention and of his desire to have it patented, appeared before the Arizona Legislature to make an appeal in behalf of the young convict. As a result a special bill was passed which granted Meyers a month's leave of absence on parole. He went unaccompanied to Washington, filed his patent applications and returned to the penitentiary. Since then he has been indefinitely paroled.

He came to Baltimore as the place where he could easily obtain the mechanical parts needed to make a more nearly perfect machine than the crude model he has fashioned in the penitentiary workshop, and is making his headquarters here while working on his invention. With him is W.E. Chenot, who has been his assistant in assembling and testing the machine and who says that he has bought Meyers' patent rights for Germany.

They have proved beyond doubt that the invention is practical and that when finally brought to a state of perfection it will introduce a new epoch in the industrial use of electricity. By Westinghouse meters they tested the strength of the current gathered from the air, and with the use of only two of the four rectifying transformers the voltmeter recorded four and one-half volts, and the ammeter, which had the capacity of recording 75 amperes, was broken by the force of the current.

The machine itself is simple. It is in reality a transformer, which is familiar to anyone knowing anything at all about electricity in its practical uses. On a high tripod, which resembles somewhat the framework of a windmill tower, is the transformer, which Mr Meyers calls his 'absorber'. It is made up of an iron core, wrapped with copper wire. The secret of the invention is the manner in which the disks composing this 'absorber' are magnetized, and this secret Meyers says he found by accident while at work in prison.

What the machine, when finally perfected, will do is yet to be seen. Its inventor claims that it will greatly reduce the cost of making electricity. No batteries of any kind are needed, he says, and not a part of the machine turns upon the other. It is as durable, apparently, as an electric light pole. One of these machines, says Meyers, when perfected may be placed on a vehicle and transform enough electricity to give motive power, be that vehicle a locomotive or an automobile. He declared it can be placed on a building to furnish electric lights or power, and that the only wear will be upon the machinery which its current runs.

Meyers is 34 years old and he gained his knowledge of electricity by working in shops along the Pacific Coast. The depths of the mysteries of electricity he has not explored, but he is certain that he has found the means of absorbing it from the air and of converting it to the use of mankind.

British (GB) Patent # 191301098

Improvements in and Relating to Apparatus for Producing Electricity.

1-14-1914

Roy Jerome Meyers

Classification: - international: H05F7/00; H05F7/00; - european: H05F7/00
Application number: GBD191301098 19130114
Priority number(s): GBT191301098 19130114

Abstract ~ Vapour apparatus, arrangements of. - A rectifier for use with apparatus for producing electricity from the earth consists of mercury- vapour lamps constructed and arranged as shown in Fig. 4. Each lamp comprises two wires 6<1>, 7<1> wound around a steel tube 15 surrounding a

mercury tube 11 preferably of copper. The coil 6<1> is connected between the electrode 14 and the terminal 18, and the coil 7<1> between the terminals 19, 5. The coils 6<1>, 7<1> are preferably composed of soft iron. Reference has been directed by the Comp- troller to Specifications 16,709/87, 14,033/99, and 5457/11, [all in Class 53, Galvanic batteries], and 15,412/06.

British Patent # 1098

(January 14, 1913)

Improvements in and Relating to Apparatus for Producing Electricity

Roy Jerome Meyers

This invention relates to improvements in apparatus for the production of electrical currents, and the primary object in view is the production of a commercially serviceable electrical current without the employment of mechanical or chemical action. To this end the invention comprises means for producing what I believe to be dynamic electricity from the earth and its ambient elements.

I am, of course aware that it has been proposed to obtain static charges from upper strata of the atmosphere, but such charges are recognized as of widely variant potential and have thus far proved of no practical commercial value, and the present invention is distinguished from all such apparatus as has heretofore been employed for attracting static charges by the fact that this improved apparatus is not designed or employed to produce or generate irregular, fluctuating or other electrical charges which lack constancy, but on the other hand I have by actual test been able to produce from a very small apparatus at comparatively low elevation, say about 50 or 60 feet above the earth's surface, a substantially constant current at a commercially usable voltage and amperage. This current I ascertained by repeated tests is capable of being readily increased by additions of the unit elements in the apparatus hereinafter set forth, and I am convinced from the constancy of the current obtained and its comparatively low potential that the current is dynamic and not static, although, of course, it is not impossible that certain static discharges occur and, in fact, I have found occasion to provide against the damage which might result from such discharge by the provision of lightning arresters and cut-out apparatus which assist in rendering the obtained current stable by eliminating sudden fluctuations which sometimes occur during conditions of high humidity from what I consider static discharges. The nature of my invention is obviously such that I have been unable to establish authoritatively all of the principles involved, and some of the theories herein expressed may possibly prove erroneous, but I do know and am able to demonstrate that the apparatus which I have discovered does produce, generate, or otherwise acquire a difference of potential representing a current amperage above stated, or varied therefrom at the will of the operator according to the uses which the current is to be subjected.

The invention comprises generically means for producing electrical currents of serviceable potential substantially without the employment of mechanical or chemical action, and in this connection I have been able to observe no chemical action whatever on the parts utilized although deterioration may possibly occur in some of the parts, but so far as I am able to determine such deterioration does not add to the current supply but is merely incidental to the effect of climatic action.

The invention more specifically comprises the employment of a magnet or magnets and a co-operating element, such as zinc disposed adjacent to the magnet or magnets and connected in such manner and arranged relative to the earth so as to produce current, my observation being that current is produced only when such magnets have their poles facing substantially to the north and south and the zincs are disposed substantially along the magnets.

The invention also comprehends other details of construction, combinations and arrangements of parts as will hereinafter be fully set forth and claimed.

In the accompanying drawings:

Figure 1 is a top plan view of an apparatus embodying the features of the present invention, the arrow accompanying the figure indicating substantially the geographical north, parts of the figure being diagrammatic for condensing the showing.

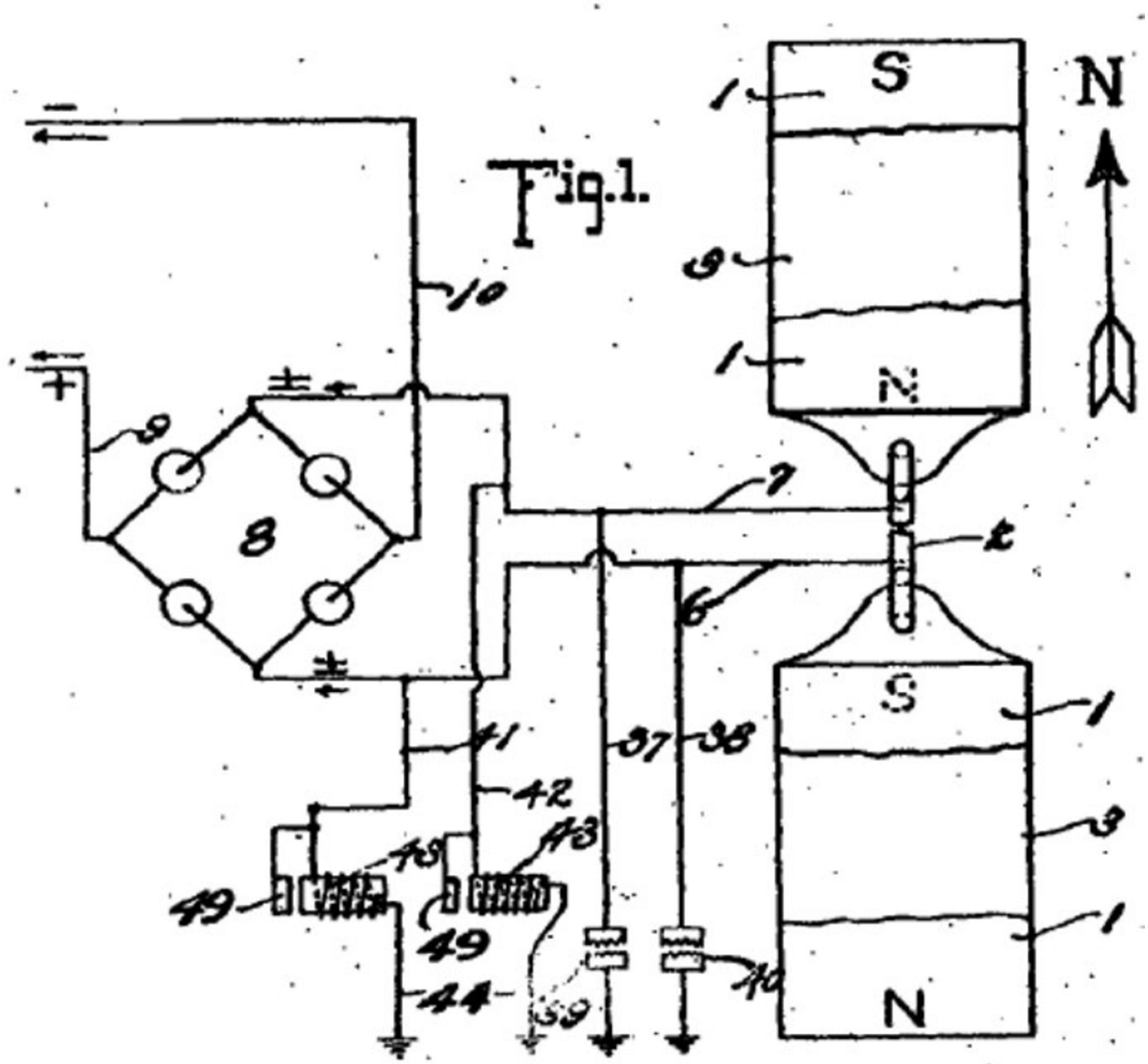
Fig.1.

Figure 2 is a view is side elevation of the parts seen in plan in Figure 1.

Fig.2.

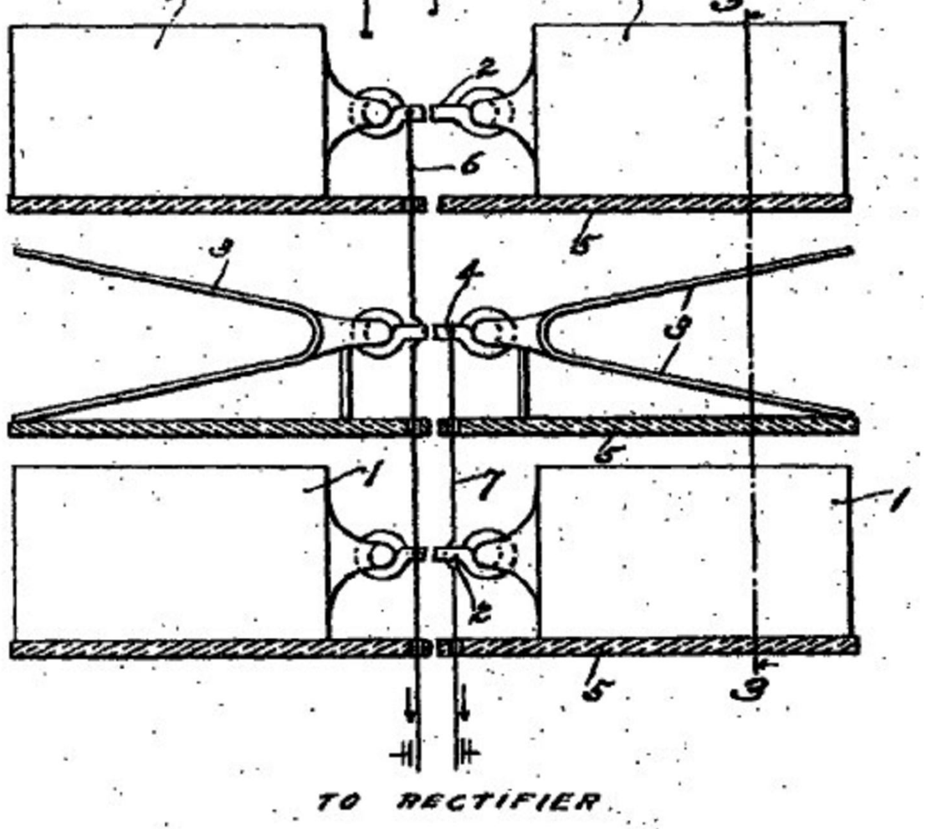
1
1
3
2
6
5
9
4
3
6
1
7
1
2
5
3
TO RECTIFIER

Figure 3 is a vertical section taken on the plane indicated by the line 3-3 of Figure 2 and looking in the direction indicated by the arrow.

Fig.5.

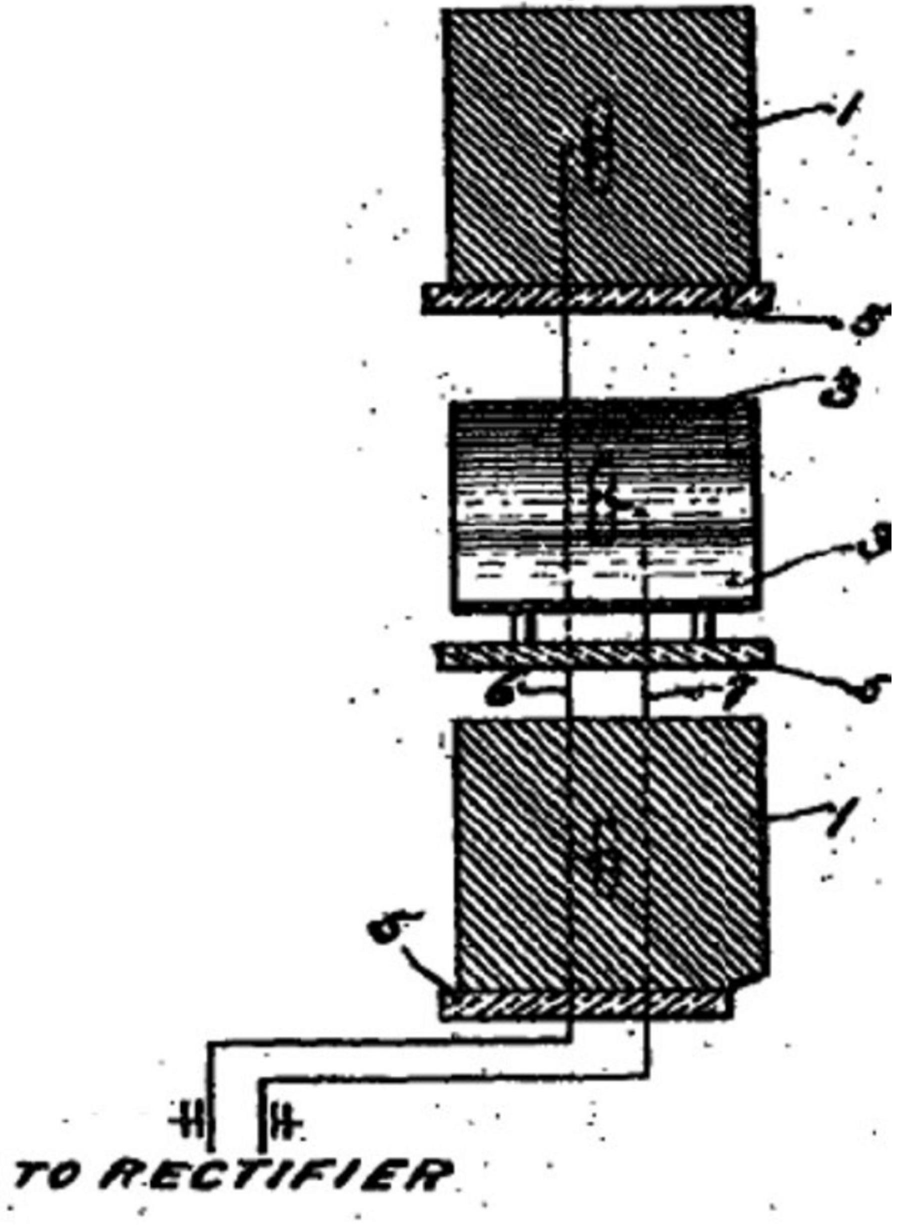
1
4
8
3
6
3
5
6
7
1
4
5
TO RECTIFIER

Figure 4 is a detail view partly in elevation and partly in section showing the detail connections of the converter and intensifier.

Fig4

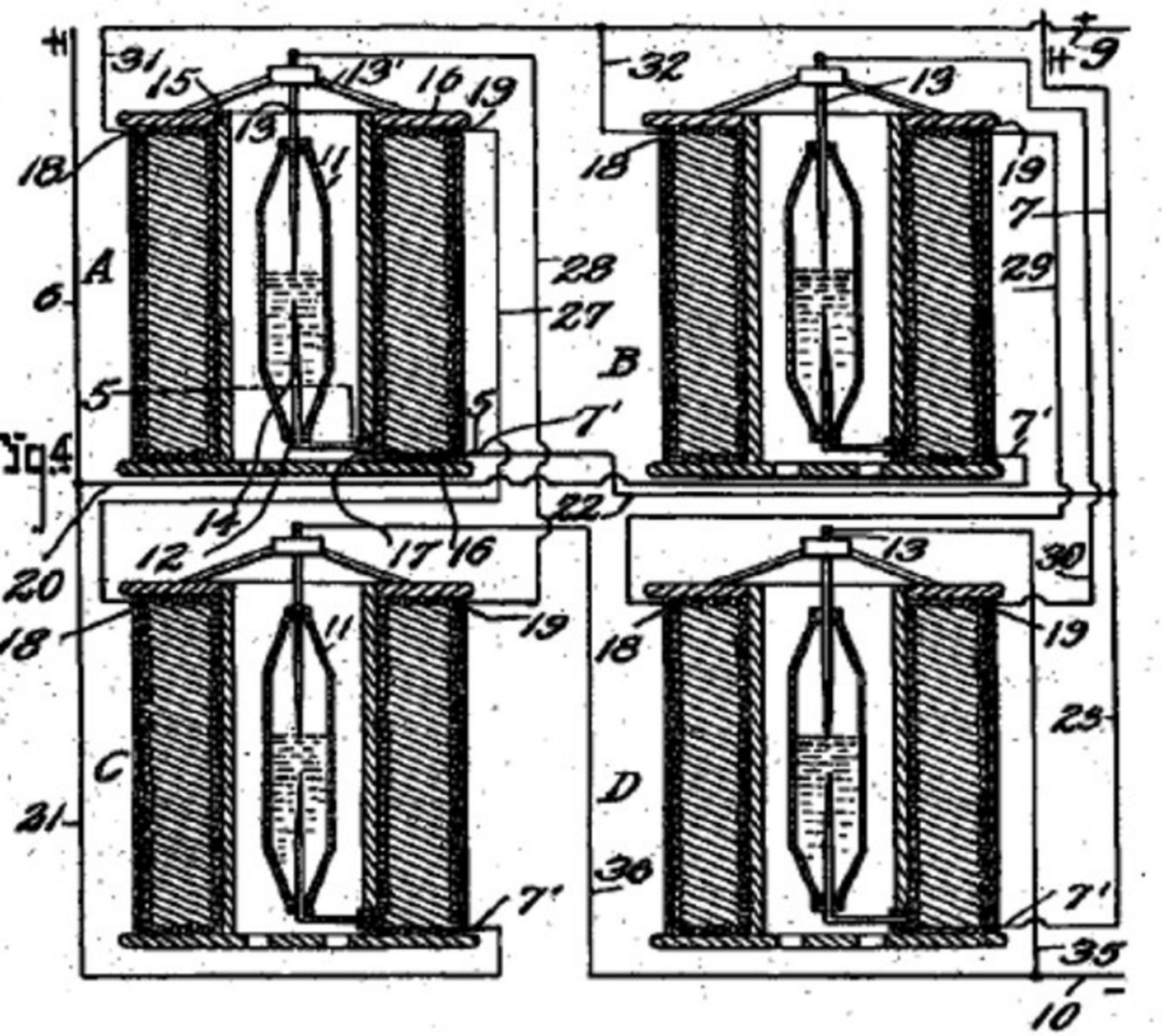
+
31
15
13'
16
19
13
11
18
6
A
5
28
27
B
5
7'
14
20
22
17
16
12
32
13
9
19
7
29
30
23
C
21
D
36
35
10
-

Figure 5 is a transverse section taken on the planes indicated by line 5-5 of Figure 4 and looking downwardly.

Fig5.

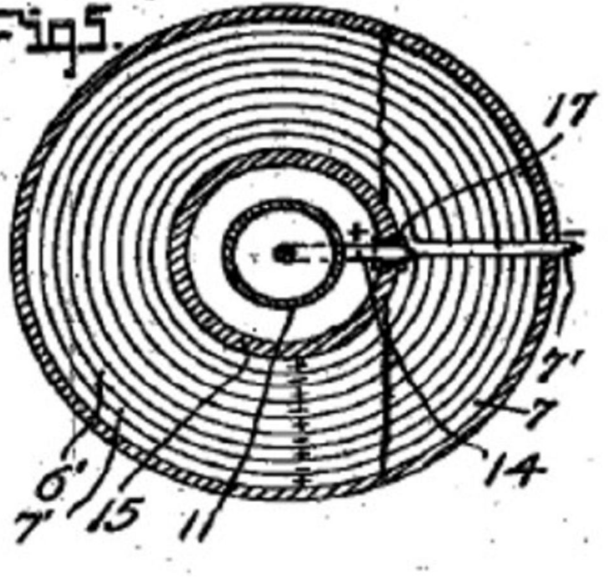
17
+
-
7'
7
14
6'
7'
15
11

Figure 6 is an enlarged detail fragmentary section illustrating the parts at the juncture of the conductors and one of the intensifiers.

Fig. 6

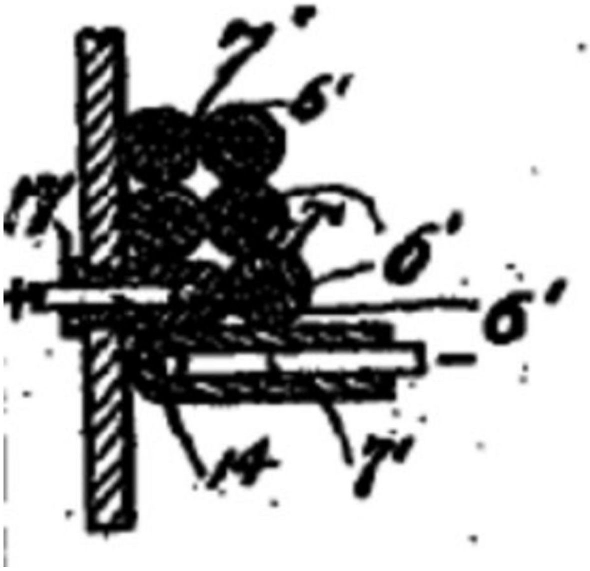

Figure 7 is an enlarged detail view partly in elevation and partly in section of one of the automatic cut-outs and

Fig. 7.

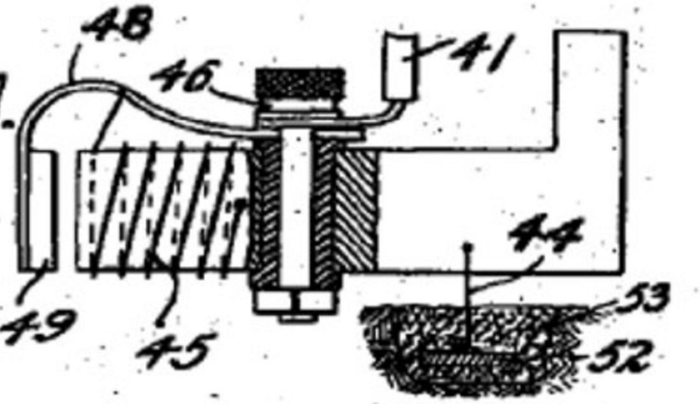

Figure 8 is a diagrammatic view of one of the simplest forms of embodiment of the invention.

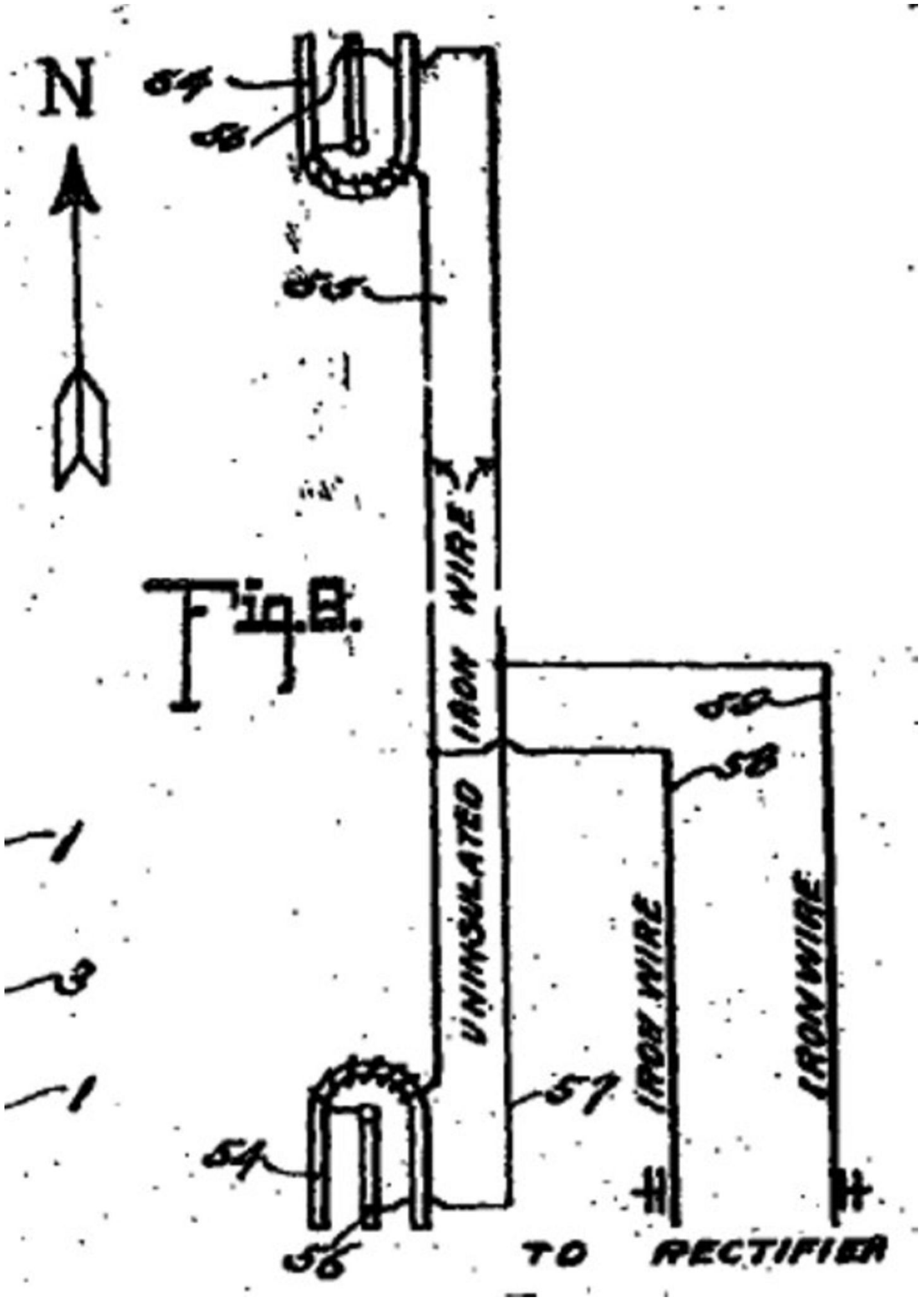

Fig. 8.

Referring to the drawing by numerals, 1,1 indicates magnets connected by a magnetic substance 2, preferably an iron wire. The magnets 1 are arranged in pairs, one pair being spaced beneath the other, and interposed between the magnets are zinc plates 3,3 connected by an iron wire conductor 4. Suitable insulating supports 5 are arranged for sustaining the respective magnets 1 and plates 3,3. Each plate 3 is preferably bent substantially into V form, as clearly seen in Figure 1, and the V1 of one of the plates opens or faces toward the north and the V of the other plate to the South. I have determined by experimentation that it is essential that the plates 3 be disposed substantially north and south with their flat faces approximately parallel to the adjacent faces of the co-operating magnets, although by experience I have not discovered any material difference in the current obtained when the plates are disposed slightly to one side of north and south, as for instance when the plates are disposed slightly to one side of north and south, as for instance when disposed in the line of the magnetic polarity of the earth. The same is true with respect to the magnets 1, the said magnets being disposed substantially north and south for operative purposes, although I find that it is immaterial whether the north pole of one of the magnets is disposed to the north and the south pole to the south, or vice versa, and it is my conviction from experience that it is essential to have the magnets of each pair connected by magnetic material so that the magnets substantially become one with a pole exposed to the north and a pole exposed to the north. In Figure 1, I have indicated in full lines by the letters 8 and N the respective polarities of the magnets 1, and have indicated in dotted lines the other pole of those magnets when the connection 2 is severed. I have found that the magnets and zinc plates operate to produce, whether by collection or generation I am not certain, electrical currents when disposed substantially north and south, but when disposed substantially east and west no such currents are produced. I also find that the question of elevation is by no means vital, but it is true that more efficient results are obtained by placing the zincs and magnets on elevated supports. I furthermore find from tests that it is possible to obtain currents from the apparatus with the zincs and magnets disposed in a building or otherwise enclosed, although more efficient results are obtained by having the said elements arranged in the open.

While in Figures 1, 2, and 3, I have shown the magnets and the zinc plates as superimposed, it will be apparent, as hereinafter fully set forth, that these elements may be juxtaposed in horizontal planes, and substantially the same results will be secured. Furthermore, the magnets 1 with the interposed zincs 3, as shown in Figures 1, 2 and 3 merely represent a unit which may be repeated either horizontally or vertically for increasing the current supply, and when the unit is repeated the zinc plates are arranged alternating with the magnets throughout the entire series as hereinafter indicated.

A conductor 6 is connected in multiple with the conductors 2 and a conductor 7 is connected with conductor 4, the conductor 6 extending to one terminal of a rectifier which I have indicated by the general reference character 8, and the conductor 7 extending to the other terminal of said rectifier. The rectifier as seen in diagram in Figure 1 may assume any of several well known embodiments of the electrical valve type and may consist of four asymmetric cells or Cooper-Hewitt mercury vapor lamps connected as indicated in Figure 1 for permitting communication of the positive impulses from the conductor 6 only to the line conductor 9 and the negative impulses from conductor 6 on only to the line conductor 10. The current from this rectifier may be delivered through the conductors 9 and 10 to any suitable source for consumption.

While the said rectifier 8 may consist of any of the known types, as above outlined, it preferably consists of a specially constructed rectifier which also has the capacity of intensifying the current and comprises specifically the elements shown in detail in Figures 4, 5, and 6 wherein I have disclosed the detail wiring of the rectifier when composed of four of the rectifying and intensify in elements instead of asymmetric cells or simple mercury vapor valves. As each of these structures is an exact embodiment of all the others, one only will be described, and the description will apply to all. The rectifying element of each construction consists of a mercury tube 11 which is preferably formed of glass or other suitable material, and comprises a cylinder having its end portions tapered and each terminating in an insulating plug or stopper 12. Through the upper stopper 12 is extended the electrode 13 which extends well into the tube and preferably substantially one-half the length thereof to a point adjacent the inner end of an opposing electrode 14 which latter electrode extends thence downwardly through the insulation 12 at the lower end of the tube. The tube 11 is supplied with mercury and is adapted to operate on the principle of the mercury vapor lamp, serving to rectify current by checking back impulses of one sign and permitting passage of impulses of the other. To avoid the necessity for utilizing a starter, as is common with the lamp type of electrical valve, the supply of mercury within the tube may be sufficient to contact with the lower end of the electrode 13 when current is not being supplied, so that as soon as current is passed from one electrode to the other sufficiently for volatilizing that portion of the mercury immediately adjacent the lower end of electrode 13, the structure begins its operation as a rectifier. The tube 11 is surrounded by a tube 15 which is preferably spaced from tube 11 sufficiently for allowing atmospheric or other cooling circulation to pass the tube 11. In some instances, it may be desirable to cool the tube 11 by a surrounding body of liquid, as hereinafter indicated. The tube 15 may be of insulating material but I find efficient results attained by the employment of a steel tube, and fixed to the ends of the of the tube are insulating disks 16, 16 forming a spool on which are wound twin wires 6^1 and 7^1, the wire 6^1 being connected at the inner helix of the coil with the outer end of the electrode 14, the lower portion of said electrode being extended to one side of the tube 11 and passed through an insulating sleeve 17 extending through the tube 15, and at its outer end merging into the adjacent end of the wire 6^1. The wire 7^1 extends directly from the outer portion of the spool through the several helices to a point adjacent the juncture of the electrode 14 with wire 6^1 and thence extends in mechanical parallelism with the wire throughout the coil, the wire 6^1 ending in a terminal 18 and the wire 7^1 ending in a terminal 19. For the sake of convenience of description and of tracing the circuits, each of the apparatus just above described and herein known as an intensifier and rectifier will be mentioned as A, B, C and D, respectively. Conductor 6 is formed with branches 20 and 21 and conductor 7 is formed with similar branches 22 and 23. Branch 20 from conductor 6 connects with conductor 7^1 of intensifier B and branch 21 of conductor 6 connects with the conductor 7^1 of intensifier C, while branch 22 of conductor 7 of intensifier C, while branch 22 of conductor 7 connects with conductor 7^1 of intensifier D. A conductor 27 is connected with terminal 19 of intensifier A and extends to and is connected with the terminal 18 of intensifier C, and a conductor 7 connects with conductor 7^1 of intensifier D. A conductor 27 is connected with terminal 19 of intensifier A. and extends to and is connected with terminal 18 of intensifier C, and a conductor 28 is connected with the terminal 19 of intensifier C and extends from the terminal 19 of intensifier B to the terminal 18 of intensifier D to electrode 13 of intensifier B. Each electrode 13 is supported on a spider 13^1 resting on the upper disk 16 of the respective intensifier. Conductors 31 and 32 are connected with the terminals 18 of intensifiers A and B and are united to form the positive line wire 9 which co-operates with the negative line wire 10 and extends to any suitable point of consumption. The line wire 10 is provided with branches 35 and 36 extending to the electrodes 13 of intensifiers C and D for completing the negative side of the circuit.

Thus it will be seen that alternating currents produced in the wires 6 and 7 will be rectified and delivered in the form of a direct current through the line wires 9 and 10, and I find by experiment that the wires 6 and 7 should be of iron, preferably soft, and may of course be insulated, the other wiring not specified as iron being of copper or other suitable material.

In carrying out the operation as stated, the circuits may be traced as follows: A positive impulse starting at the zincs 3 is directed along conductor 7 to branch 23 to conductor 7^1 and the winding of the rectifier of intensifier B through said rectifier to the conductor 6^1, through the winding thereof to the contact 18, conductor 32 and to the line wire 9. The next or negative impulse directed along conductor 7 cannot find its way along branch 23 and the circuit just above traced because it cannot pass across the rectifier of intensifier B but instead the negative impulse passes along conductor 22 to conductor 7 of intensifier A and the winding thereof to the contact 19 and to conductor 27 to contact 18 of intensifier C, to the winding of the wire 6^1 thereof to the electrode 14 through the rectifier to the of the electrode 13 and conductor of intensifier A, electrode 14 thereof and conductor 6^1 to contact 18 and wire 31 to line wire 9. Obviously the positive impulse cannot pass along the wire 20 because of its inverse approach to the rectifier of intensifier B. The next impulse or negative impulse delivered to conductor 6 cannot pass along conductor 21 because of its connection with electrode 13 of the rectifier of intensifier A, but instead passes along conductor 20 to the wire 7^1 and its winding forming part of intensifier B to the contact 19 and conductor 29 to contact 18 and the winding of wire 6^1 of intensifier D to the electrode 14 and through the rectifier to the electrode 13 and conductor 35 to line wire 10. Thus the current is rectified and all positive impulses directed along one line and all negative impulses along the other lie s that the potential difference between the two lines will be maximum for the given current of the alternating circuit. It is, of course, apparent that a less number of intensifiers with their accompanying rectifier elements may be employed with a sacrifice of the impulses which are checked back from a lack of ability to pass the respective rectifier elements, and in fact I have secured efficient results by the use of a single intensifier with its rectifier elements, as hereinafter set forth.

Grounding conductors 37 and 38 are connected respectively with the conductors 6 and 7 and are provided with the ordinary lightning arresters 39 and 40 respectively for protecting the circuit against high tension static charges.

Conductors 41 and 42 are connected respectively with the conductors 6 and 7 and each connects with an automatic cutout 43 which is grounded as at 4. Each of said automatic cutouts is exactly like the other and one of the same is shown in detail in Figure 7 and comprises the inductive resistance 45 provided with an insulated binding post 46 wit which the respective conductor 6 or 7 is connected, said post also supporting a spring 48 which sustains an armature 49 adjacent to the core of the resistance 45. The helix of resistance 45 is connected preferably through the spring to the binding post at one end and at the other end is grounded on the core of the resistance, the said core being grounded by ground conductor 44 which extends to the metallic plate 52 embedded in moist carbon or other inductive material buried in the earth. Each of the conductors 41, 42 and 44 is of iron, and in this connection I wish it understood that where I state the specific substance I am able to verify the accuracy of the statement by the results of tests which I have made, but of course I wish to include along with such substances al equivalents, as for instance, where iron is mentioned its byproducts, such as steel, and its equivalents such as nickel and other magnetic substances are intended to be comprehended. The cutout apparatus seen in detail in Figure 7 is employed particularly for insuring against high tension currents, it being obvious from the structure shown that when potential rises beyond the limit established by the tension of the spring sustaining the armature 40, the armature will be moved to a position contacting with the core of the cutout device and thereby directly close the ground connection for line wire 41 with conductor 44, eliminating the resistance of winding 45 and allowing the high tension current to be discharged to the ground. Immediately upon such discharge the winding 45 losing its current will allow the core to become demagnetized and release the armature 49 whereby the ground connection is substantially broken leaving only the connection through the winding 45 the resistance of which is sufficient for insuring against loss of low tension current.

In Figure 8 I have illustrated an apparatus which though apparently primitive in construction and arrangement comprehends the first successful embodiment which I produced in the course of discovery of the present invention, and it will be observed that the essential features of the invention are therein disclosed. The structure delineated in said figure consists of horseshoe magnets 54, 55, one facing north and the other south, that is, each opening in the respective directions indicated and the two being connected by an iron wire 55 which is uninsulated and wrapped about the respective magnets each end portion of the wire 55 being extended from the respective magnets to and connected with, as by being soldered to, a zinc plate 56, there being a plate 56 for each magnet and each plate being arranged longitudinally substantially parallel with the legs of the magnet and with the faces of the plate exposed toward the respective legs of the magnet, the plate being thus arranged endwise toward the north and south. An iron wire 57 connects the plates 56, the ends of the wire being preferably connected adjacent the outer ends of the plates but from experiment I find that the wire may be connected at practically any point to the plate. Lead wires 58 and 59 are connected respectively with the wires 55 and 57 and supply an alternating current at a comparatively low tension, and to control such current the wires 58 and 59 may be extended to a rectifier or combined rectifier and intensifier, as above set forth.

The tests which I have found successful with the apparatus seen in Figure 8 were carried out by the employment first of horseshoe magnets approximately 4 inches in length, the bar comprising the horseshoe being about one inch square, the zincs being dimensioned proportionately and from this apparatus with the employment of a single intensifier and rectifier, as above stated, I was able to obtain a constant current of 8 volts.

It should be obvious that the magnets forming one of the electrodes of this apparatus may be permanent or may be electromagnets, or a combination of the two.

While the magnets mentioned throughout the above may be formed of any magnetic substance, I find the best results obtained by the employment of the nickel chrome steel.

While the successful operation of the various devices which I have constructed embodying the present invention have not enabled me to arrive definitely and positively at fixed conclusion relative to the principles and theories of operation and the source from which current is supplied, I wish it to be understood that I consider myself as the first inventor of the general type hereinbefore described capable of producing commercially serviceable electricity, for which reason my claims hereinafter appended contemplate that I may utilize a wide range of equivalents so far as concerns details of construction suggested as preferably employed.

The current which I am able to obtain is dynamic in the sense that it is not static and its production is accomplished without chemical or mechanical action either incident to the actual chemical or mechanical motion or incident to changing caloric conditions so that the elimination of necessity for the use of chemical or mechanical action is to be considered as including the elimination of the necessity for the use of heat or varying degrees thereof.

Having now particularly described and ascertained the nature of my said invention, and in what manner the same is to be performed, I declare that what I claim is: --- [Claims not included here]

NOTES & COMMENTS

From the Article in *Tech. World Mag*.:

1. First demo model was powerful enough to spark a gas engine.
2. Second model developed 8 volts.
3. Demo model at Patent Office was elevated on short poles.
4. The model planned to power Phoenix AZ would be elevated 200 feet.
5. The Absorber "consists of a series of magnetized steel plates set in a circle (the manner of preparing them is kept secret)".
6. "[T]he magnetizing of one set of plates... is part of the secret of the treatment which makes it respond to the accumulations of the atmosphere".

From British Patent # 1098 (1913):

"I have been able... to produce from a very small apparatus at comparatively low elevation, say about 50 or 60 feet above the earth's surface, a substantially constant current at a commercially useable voltage and amperage".

"This current... is capable of being readily increased by additions of the unit elements in the apparatus".

Fig. 1 and Fig. 2 show the magnet poles are connected N-S by a thick iron rod (thick compared to the lines used for wires in the drawings).

No angle is specified for the V-shaped zinc plates. The article (but not the patent) states that the plates are magnetized (obviously not zinc). Zinc-galvanized steel? Will a thin film of Zn work? Or, powdered Zn in a binder (more surface area)? Or, zinc-galvanized iron wire in a coil?

The Palmer Craig device (www.rexresearch.com/craig/craig.htm) is powered by the terrestrial magnetic field, and employs a thin film of bismuth to capture the energy as diamagnetism. Perhaps this can be integrated with Meyers' device.

Figure 8 (the demonstration of principle) show uninsulated iron wire being used to connect the plates and magnets. The wire is wound around the Bloch wall area of the horseshoe magnets. Perhaps Coler-type windings around the poles could be used here (See: www.rexresearch.com/coler/coler.htm). Coler used copper plates as "condensers" in his device. Could copper plates be used for the Meyers device? Perhaps flat (Tesla non-inductive) coils could be integrated here.

"It is essential that the plates 3 be disposed substantially N and S with their flat faces approximately parallel to the adjacent faces of the co-operating magnets....

"I find that it is immaterial whether the N pole of one of the magnets is disposed to the N and the S pole to the S, or vice versa".

"[T]he magnets and zinc plates... produce electrical currents when disposed... N and S, but when disposed... E and W no such currents are produced".

"[E]levation is by no means vital, but... more efficient results are obtained by placing the zincs and magnets on elevated supports".

"The elements may be disposed in horizontal planes [or vertically]...".

The "zinc plate 56... [is] arranged longitudinally substantially parallel with the legs of the magnet and with the faces of the plate exposed toward the respective legs of the magnet, the plate being thus arranged endwise toward the north and south".

The first model used "horseshoe magnets approximately 4 inches in length, the bar comprising the horseshoe being about one inch square, the zincs being dimensioned proportionately and from this apparatus with the employment of a single intensifier and rectifier, as above stated, I was able to obtain a constant current of 8 volts... [T]he magnets... may be permanent or may be electromagnets, or a combination of the two... I find the best results obtained by... nickel chrome steel".

Comments & Questions:

The rectifier is described as a preferred embodiment, but other designs also work. The Ed Gray capacitor design comes to mind (www.rexresearch.com/evgray/1gray.htm). The Tate Ambient Power Module also might apply (www.rexresearch/tate/tate.htm).

Would non-ferrous magnets work? Is there a frequency involved (oscilloscope tests)? Coler found that ferromagnetism has a resonant frequency about 180 KHz. Can the components be made adjustable for RLC-resonance?

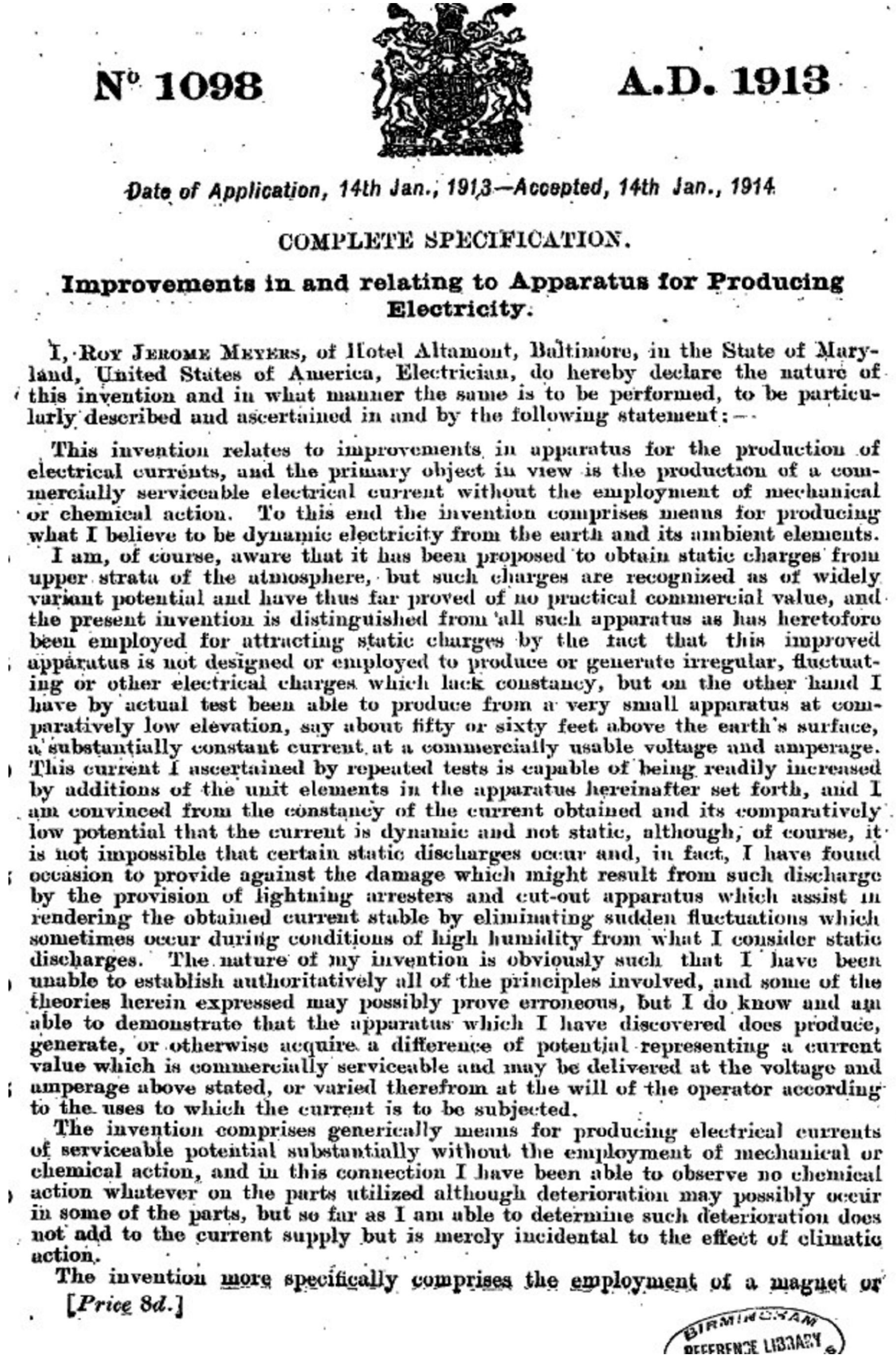

Nº 1098 A.D. 1913

Date of Application, 14th Jan., 1913—Accepted, 14th Jan., 1914.

COMPLETE SPECIFICATION.

Improvements in and relating to Apparatus for Producing Electricity.

I, ROY JEROME MEYERS, of Hotel Altamont, Baltimore, in the State of Maryland, United States of America, Electrician, do hereby declare the nature of this invention and in what manner the same is to be performed, to be particularly described and ascertained in and by the following statement:—

This invention relates to improvements in apparatus for the production of electrical currents, and the primary object in view is the production of a commercially serviceable electrical current without the employment of mechanical or chemical action. To this end the invention comprises means for producing what I believe to be dynamic electricity from the earth and its ambient elements. I am, of course, aware that it has been proposed to obtain static charges from upper strata of the atmosphere, but such charges are recognized as of widely variant potential and have thus far proved of no practical commercial value, and the present invention is distinguished from all such apparatus as has heretofore been employed for attracting static charges by the fact that this improved apparatus is not designed or employed to produce or generate irregular, fluctuating or other electrical charges which lack constancy, but on the other hand I have by actual test been able to produce from a very small apparatus at comparatively low elevation, say about fifty or sixty feet above the earth's surface, a substantially constant current at a commercially usable voltage and amperage. This current I ascertained by repeated tests is capable of being readily increased by additions of the unit elements in the apparatus hereinafter set forth, and I am convinced from the constancy of the current obtained and its comparatively low potential that the current is dynamic and not static, although, of course, it is not impossible that certain static discharges occur and, in fact, I have found occasion to provide against the damage which might result from such discharge by the provision of lightning arresters and cut-out apparatus which assist in rendering the obtained current stable by eliminating sudden fluctuations which sometimes occur during conditions of high humidity from what I consider static discharges. The nature of my invention is obviously such that I have been unable to establish authoritatively all of the principles involved, and some of the theories herein expressed may possibly prove erroneous, but I do know and am able to demonstrate that the apparatus which I have discovered does produce, generate, or otherwise acquire a difference of potential representing a current value which is commercially serviceable and may be delivered at the voltage and amperage above stated, or varied therefrom at the will of the operator according to the uses to which the current is to be subjected.

The invention comprises generically means for producing electrical currents of serviceable potential substantially without the employment of mechanical or chemical action, and in this connection I have been able to observe no chemical action whatever on the parts utilized although deterioration may possibly occur in some of the parts, but so far as I am able to determine such deterioration does not add to the current supply but is merely incidental to the effect of climatic action.

The invention more specifically comprises the employment of a magnet or

[*Price 8d.*]

BIRMINGHAM REFERENCE LIBRARY

magnets and a co-operating element, such as zinc disposed adjacent to the magnet or magnets and connected in such manner and arranged relative to the earth so as to produce current, my observation being that current is produced only when such magnets have their poles facing substantially to the north and south and the zincs are disposed substantially along the magnets.

The invention also comprehends other details of construction, combinations and arrangements of parts as will hereinafter be fully set forth and claimed.

In the accompanying drawings:

Figure 1 is a top plan view of an apparatus embodying the features of the present invention, the arrow accompanying the figure indicating substantially the geographical north, parts of the figure being diagrammatic for condensing the showing.

Figure 2 is a view in side elevation of the parts seen in plan in Figure 1.

Figure 3 is a vertical section taken on the plane indicated by the line 3—3 of Figure 2 and looking in the direction indicated by the arrow.

Figure 4 is a detail view partly in elevation and partly in section showing the detail connections of the converter and intensifier.

Figure 5 is a transverse section taken on the planes indicated by line 5—5 of Figure 4 and looking downwardly.

Figure 6 is an enlarged detail fragmentary section illustrating the parts at the juncture of the conductors and one of the intensifiers.

Figure 7 is an enlarged detail view partly in elevation and partly in section of one of the automatic cut-outs and

Figure 8 is a diagrammatic view of one of the simplest forms of embodiments of the invention.

Referring to the drawing by numerals, 1, 1 indicates magnets connected by a magnetic substance 2, preferably an iron wire. The magnets 1 are arranged in pairs, one pair being spaced beneath the other, and interposed between the magnets are zinc plates 3, 3 connected by an iron wire conductor 4. Suitable insulating supports 5 are arranged for sustaining the respective magnets 1 and plates 3, 3. Each plate 3 is preferably bent substantially into **V** form, as clearly seen in Figure 1, and the **V** of one of the plates opens or faces toward the north and the **V** of the other plate to the south. I have determined by experimentation that it is essential that the plates 3 be disposed substantially north and south with their flat faces approximately parallel to the adjacent faces of the co-operating magnets, although by experience I have not discovered any material difference in the current obtained when the plates are disposed slightly to one side of north and south, as for instance when disposed in the line of the magnetic polarity of the earth. The same is true with respect to the magnets 1, the said magnets being disposed substantially north and south for operative purposes, although I find that it is immaterial whether the north pole of one of the magnets is disposed to the north and the south pole to the south, or *vice versa*, and it is my conviction from experience that it is essential to have the magnets of each pair connected by magnetic material so that the magnets substantially become one with a pole exposed to the north and a pole exposed to the south. In Figure 1, I have indicated in full lines by the letters S and N the respective polarities of the magnets 1, and have indicated in dotted lines the other poles of those magnets when the connection 2 is severed. I have found that the magnets and zinc plates operate to produce, whether by collection or generation I am not certain, electrical currents when disposed substantially north and south, but when disposed substantially east and west no such currents are produced. I also find that the question of elevation is by no means vital, but it is true that more efficient results are obtained by placing the zincs and magnets on elevated supports. I furthermore find from tests that it is possible to obtain currents from the apparatus with the zincs and magnets disposed in a building or otherwise enclosed, although more efficient results are obtained by having the said elements arranged in the open.

While in Figures 1, 2 and 3, I have shown the magnets and the zinc plates as superimposed, it will be apparent, as hereinafter fully set forth, that these elements may be juxtaposed in horizontal planes, and substantially the same results will be secured. Furthermore, the magnets 1 with the interposed zincs 3, as shown in Figures 1, 2 and 3 merely represent a unit which may be repeated either horizontally or vertically for increasing the current supply, and when the unit is repeated the zinc plates are arranged alternating with the magnets throughout the entire series as hereinafter indicated.

A conductor 6 is connected in multiple with the conductors 2 and a conductor 7 is connected with conductor 4, the conductor 6 extending to one terminal of a rectifier which I have indicated by the general reference character 8, and the conductor 7 extending to the other terminal of said rectifier. The rectifier as seen in diagram in Figure 1 may assume any of several well known embodiments of the electrical valve type and may consist of four assymmetric cells or Cooper-Hewitt mercury vapor lamps connected as indicated in Figure 1 for permitting communication of the positive impulses from conductor 7 only to the line conductor 9 and the negative impulses from the conductor 7 only to the line conductor 10, and permitting the positive impulses from the conductor 6 only to the line conductor 9 and the negative impulses from conductor 6 only to the line conductor 10. The current from this rectifier may be delivered through the conductors 9 and 10 to any suitable source for consumption.

While the said rectifier 8 may consist of any of the known types, as above outlined, it preferably consists of a specially constructed rectifier which also has the capacity of intensifying the current and comprises specifically the elements shown in detail in Figures 4, 5 and 6 wherein I have disclosed the detail wiring of the rectifier when composed of four of the rectifying and intensifying elements instead of assymmetric cells or simple mercury vapor valves. As each of these structures is an exact embodiment of all the others, one only will be described, and the description will apply to all. The rectifying element of each construction consists of a mercury tube 11 which is preferably formed of copper but may be formed of glass or other suitable material, and comprises a cylinder having its end portions tapered and each terminating in an insulating plug or stopper 12. Through the upper stopper 12 is extended the electrode 13 which extends well into the tube and preferably substantially one-half the length thereof to a point adjacent the inner end of an opposing electrode 14 which latter electrode extends thence downwardly through the insulation 12 at the lower end of the tube. The tube 11 is supplied with mercury and is adapted to operate on the principle of the mercury vapor lamp, serving to rectify current by checking back impulses of one sign and permitting passage of impulses of the other. To avoid the necessity for utilizing a starter, as is common with the lamp type of electrical valve, the supply of mercury within the tube may be sufficient to contact with the lower end of the electrode 13 when current is not being supplied, so that as soon as current is passed from one electrode to the other sufficiently for volatilizing that portion of the mercury immediately adjacent the lower end of electrode 13, the structure begins its operation as a rectifier. The tube 11 is surrounded by a tube 15 which is preferably spaced from tube 11 sufficiently for allowing atmospheric or other cooling circulation to pass the tube 11. In some instances, it may be desirable to cool the tube 11 by a surrounding body of liquid, as hereinafter indicated. The tube 15 may be of insulating material but I find efficient results attained by the employment of a steel tube, and fixed to the ends of the tube are insulating disks 16, 16 forming a spool on which are wound twin wires 6^1 and 7^1, the wire 6^1 being connected at the inner helix of the coil with the outer end of the electrode 14, the lower portion of said electrode being extended to one side of the tube 11 and passed through an insulating sleeve 17 extending through the tube 15, and at its outer end merging into the adjacent end of the wire 6^1. The wire 7^1 extends directly from the outer portion of the spool through the several helices to a point adjacent the juncture of the electrode 14 with wire 6^1 and thence extends in mechanical parallelism with the wire throughout the coil, the wire 6^1 ending in a terminal 18 and the wire 7^1 ending in a terminal 19. For the sake of convenience of description and of tracing the circuits, each of the apparatus just above described and herein known as an intensifier and rectifier will be mentioned as A, B, C and D, respectively. Conductor 6 is formed with branches 20 and 21 and conductor 7 is formed with similar branches 22 and 23. Branch 20 from conductor 6 connects with conductor 7^1 of intensifier B and branch 21 of conductor 6 connects with the conductor 7^1 of intensifier C, while branch 22 of conductor 7 connects with conductor 7^1 of intensifier A and branch 23 of conductor 7 connects with conductor 7^1 of intensifier D. A conductor 27 is connected with terminal 19 of intensifier A and extends to and is connected with terminal 18 of intensifier C, and a conductor 28 is connected with the terminal 19 of intensifier C and extends to and is connected with electrode 13 of intensifier A. A conductor 29 extends from the terminal 19 of intensifier B to the terminal 18 of intensifier D and a conductor 30 extends from the terminal 19 of intensifier D to electrode 13 of intensifier B. Each electrode 13 is supported on a spider 13^1 resting on the upper disk 16 of the respective intensifier. Conductors 31 and 32 are connected with the terminals 18 of intensifiers A and B and are united to form the positive line wire 9 which co-operates with the negative line wire 10 and extends to any suitable point of consumption. The line wire 10 is provided with branches 35 and 36 extending to the electrodes 13 of intensifiers C and D for completing the negative side of the circuit.

Thus it will be seen that alternating currents produced in the wires 6 and 7 will be rectified and delivered in the form of a direct current through the line wires 9 and 10, and I find by experiment that the wires 6 and 7 should be of iron, preferably soft, and may of course be insulated, the other wiring not specified as iron being of copper or other suitable material.

In carrying out the operation as stated, the circuits may be traced as follows: A positive impulse starting at the zincs 3 is directed along conductor 7 to branch 23 to conductor 7^1 and the winding of intensifier D to terminal 19 through conductor 30 to electrode 13 of the rectifier of intensifier B through said rectifier to the conductor 6^1, through the winding thereof to the contact 18, conductor 32 and to the line wire 9. The next or negative impulse directed along conductor 7 cannot find its way along branch 23 and the circuit just above traced because it cannot pass across the rectifier of intensifier B but instead the negative impulse passes along conductor 22 to conductor 7 of intensifier A and the winding thereof to the contact 19 to conductor 27 to contact 18 of intensifier C, to the winding of the wire 6^1 thereof to the electrode 14 through the rectifier to the electrode 13 and conductor 36 to the line wire 10. A positive impulse delivered to wire 6 passes along the said wire to the branch 21 to the conductor 7^1 of intensifier C and the winding thereof to the contact 19, conductor 28, electrode 13 of the rectifier of intensifier A, electrode 14 thereof and conductor 6^1 to contact 18 and wire 31 to line wire 9. Obviously the positive impulse cannot pass along the wire 20 because of its inverse approach to the rectifier of intensifier B. The next impulse or negative impulse delivered to conductor 6 cannot pass along conductor 21 because of its connection with electrode 13 of the rectifier of intensifier A, but instead passes along conductor 20 to the wire 7^1 and its winding forming part of intensifier B to the contact 19 and conductor 29 to contact 18 and the winding of wire 6^1 of intensifier D to the electrode 14 and through the rectifier to the electrode 13 and conductor 35 to line wire 10. Thus the current is rectified and all positive impulses directed along one line and all negative impulses along the other line so that the potential difference between the two lines will be maximum for the given current of the alternating circuit. It is, of course, apparent that a less number of intensifiers with their accompanying rectifier elements may be employed with a sacrifice of the impulses which are checked back from a lack of ability to pass the respective rectifier elements, and in fact I have secured efficient results by the use of a single intensifier with its rectifier elements, as hereinafter set forth.

Grounding conductors 37 and 38 are connected respectively with the conductors 6 and 7 and are provided with the ordinary lightning arresters 39 and 40 respectively for protecting the circuit against high tension static charges.

Conductors 41 and 42 are connected respectively with the conductors 6 and 7 and each connects with an automatic cut-out 43 which is grounded as at 44. Each of said automatic cut-outs is exactly like the other and one of the same is shown in detail in Figure 7 and comprises the inductive resistance 45 provided with an insulated binding post 46 with which the respective conductor 6 or 7 is connected, said post also supporting a spring 48 which sustains an armature 49 adjacent the core of the resistance 45. The helix of resistance 45 is connected preferably through the spring to the binding post at one end and at the other end is grounded on the core of the resistance, the said core being grounded by ground conductor 44 which extends to the metallic plate 52 imbedded in moist carbon or other inductive material 53 buried in the earth. Each of the conductors 41, 42 and 44 is of iron, and in this connection I wish it understood that where I state the specific substance I am able to verify the accuracy of the statement by the results of tests which I have made, but of course I wish to include along with such substance all equivalents, as for instance, where iron is mentioned its by-products, such as steel, and its equivalents such as nickel and other magnetic substances are intended to be comprehended. The cut-out apparatus seen in detail in Figure 7 is employed particularly for insuring against high tension currents it being obvious from the structure shown that when potential rises beyond the limit established by the tension of the spring sustaining the armature 49, the armature will be moved to a position contacting with the core of the cut-out device and thereby directly close the ground connection for line wire 41 with conductor 44, eliminating the resistance of winding 45 and allowing the high tension current to be discharged to the ground. Immediately upon such discharge the winding 45 losing its current will allow the core to become demagnetized and release the armature 49 whereby the ground connection is substantially broken leaving only the connection through the winding 45 the resistance of which is sufficient for insuring against loss of low tension current.

In Figure 8 I have illustrated an apparatus which though apparently primitive in construction and arrangement comprehends the first successful embodiment which I produced in the course of the discovery of the present invention, and it will be observed that the essential features of the invention are therein disclosed. The structure delineated in said figure consists of horse shoe magnets 54, 54, one facing north and the other south, that is, each opening in the respective directions indicated and the two being connected by an iron wire 55 which is uninsulated and wrapped about the respective magnets at or adjacent the neutral zone thereof, and the wire 55 is preferably soldered to the respective magnets each end portion of the wire 55 being extended from the respective magnet to and connected with, as by being soldered to, a zinc plate 56, there being a plate 56 for each magnet and each plate being arranged longitudinally substantially parallel with the legs of the magnet and with the faces of the plate exposed toward the respective legs of the magnet, the plate being thus arranged endwise toward the north and south. An iron wire 57 connects the plates 56, the ends of the wire being preferably connected adjacent the outer ends of the plates but from experiment I find that the wire may be connected at practically any point to the plate. Lead wires 58 and 59 are connected respectively with the wires 55 and 57 and supply an alternating current ordinarily at a comparatively low tension, and to control such current the wires 58 and 59 may be extended to a rectifier or combined rectifier and intensifier, as above set forth.

The tests which I have found successful with the apparatus seen in Figure 8 were carried out by the employment first of horse shoe magnets approximately four inches in length, the bar comprising the horse shoe being about one inch square, the zincs being dimensioned proportionately and from this apparatus with the employment of a single intensifier and rectifier, as above stated, I was able to obtain a constant current of 8 volts.

It should be obvious that the magnets forming one of the electrodes of this apparatus may be permanent or may be electro-magnets, or a combination of the two.

While the magnets mentioned throughout the above may be formed of any magnetic substance, I find the best results obtained by the employment of the nickel chrome steel.

While the successful operation of the various devices which I have constructed embodying the present invention have not enabled me to arrive definitely and positively at fixed conclusions relative to the principles and theories of operation and the source from which current is supplied, I wish it to be understood that I consider myself as the first inventor of an apparatus of the general type hereinbefore described capable of producing commercially serviceable electricity, for which reason my claims hereinafter appended contemplate that I may utilize a wide range of equivalents so far as concerns details of construction suggested as preferably employed.

The current which I am able to obtain is dynamic in the sense that it is not static and its production is accomplished without chemical or mechanical action either incident to the actual chemical or mechanical motion or incident to changing calorific conditions so that the elimination of necessity for the use of chemical or mechanical action is to be considered as including the elimination of the necessity for the use of heat or varying degrees thereof.

Having now particularly described and ascertained the nature of my said invention, and in what manner the same is to be performed, I declare that what I claim is:—

1. Means for producing dynamic electricity without mechanical or chemical action.

2. Electricity producing means as claimed in Claim 1, characterized in that said means is capable of producing low tension current.

3. Electricity producing means as claimed in Claim 1, characterized in that said means comprises a magnet, and means co-operating with the latter.

4. Means as set forth in Claim 3, characterized in that the parts are stationary.

5. Means as claimed in Claim 3, characterized in that the magnet is disposed substantially north and south.

6. Means as claimed in Claims 1, 3 and 5, wherein the means coacting with the magnet is a metal and said parts are disposed substantially north and south.

7. Means as claimed in Claims 1, 3 and 6, in which the metal is zinc, and combined with conductors connected with the co-operating parts to deliver current therefrom.

8. Means as set forth in Claim 1, characterized in that said means comprises spaced magnets with an adjacent co-operating zinc, and an iron wire connecting the magnets, and means to deliver electrical current from said parts.

9. Means as set forth in Claim 8, characterized in that the current delivery means comprises an intensifier and rectifier.

10. Means as set forth in Claim 8, characterized in that the current delivery means comprises wires connected with the magnets and zinc, an electrical valve connected with one of said wires, and independent helices surrounding said valve, one of said helices being connected with the terminal of the valve at the opposite side from the contact of said wire and the other of said helices being connected with the other wire, the electrical valve comprising a mercury vapor rectifier.

11. The process of producing electricity comprising exposing a magnet and

co-operating means disposed substantially north and south and taking off current therefrom.

12. The process of producing electricity comprising disposing stationary elements in co-operative relation in respect to each other and with respect to the earth for giving off relatively low tension electrical current, substantially without chemical action.

13. Electricity producing means constructed and operating substantially as described with reference to the accompanying drawings.

Dated this 13th day of January, 1913.

W. P. THOMPSON & Co.,
6, Lord Street, Liverpool, and at
Bradford & London,
Agents for the Applicant.

Reference has been directed, in pursuance of Section 7, Sub-section 4, of the Patents and Designs Act, 1907, to Specifications No. 16,709 of 1887, No. 14,033 of 1899, No. 15,412 of 1906, and No. 5457 of 1911.

Redhill: Printed for His Majesty's Stationery Office, by Love & Malcomson, Ltd.—1914.

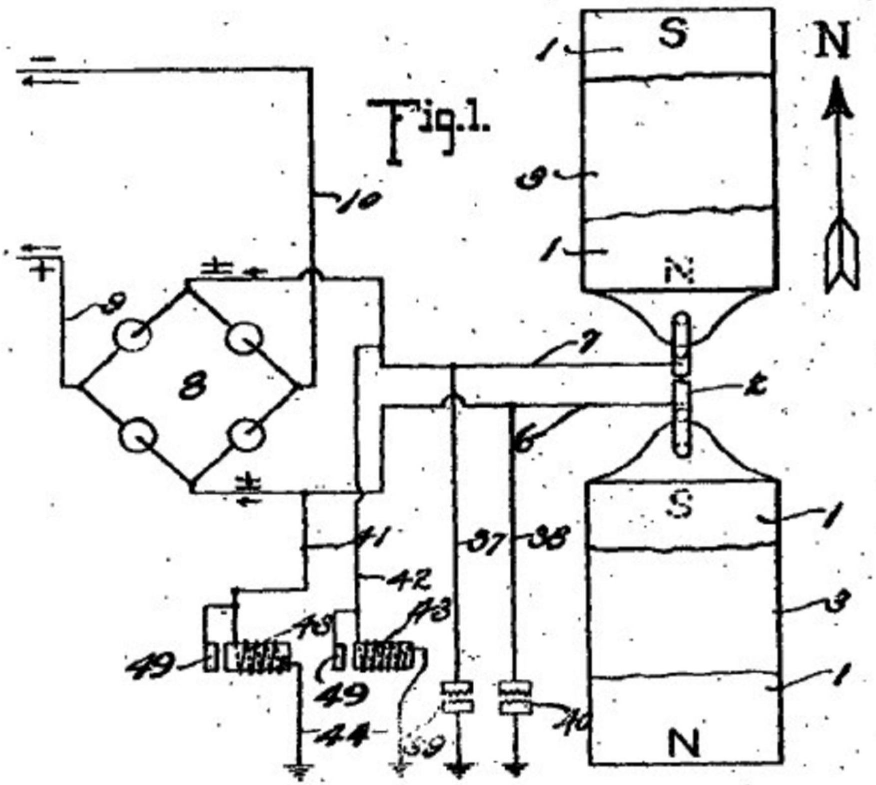
Fig.1.
S
N
S
N
N
1
9
1
10
8
9
7
6
t
41
42
37
38
49
48
43
49
44
39
40
1
3
1
+
−

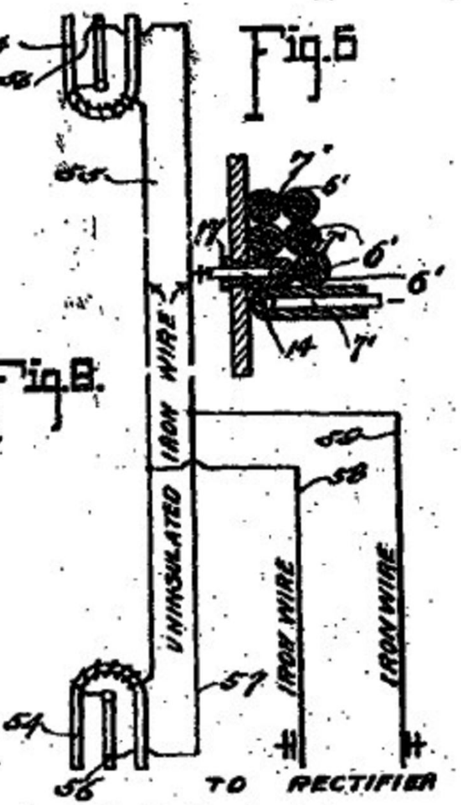
Fig.6
Fig.8.
54
55
56
57
58
59
6'
7'
14
17
IRON WIRE
UNINSULATED
IRON WIRE
IRON WIRE
TO RECTIFIER

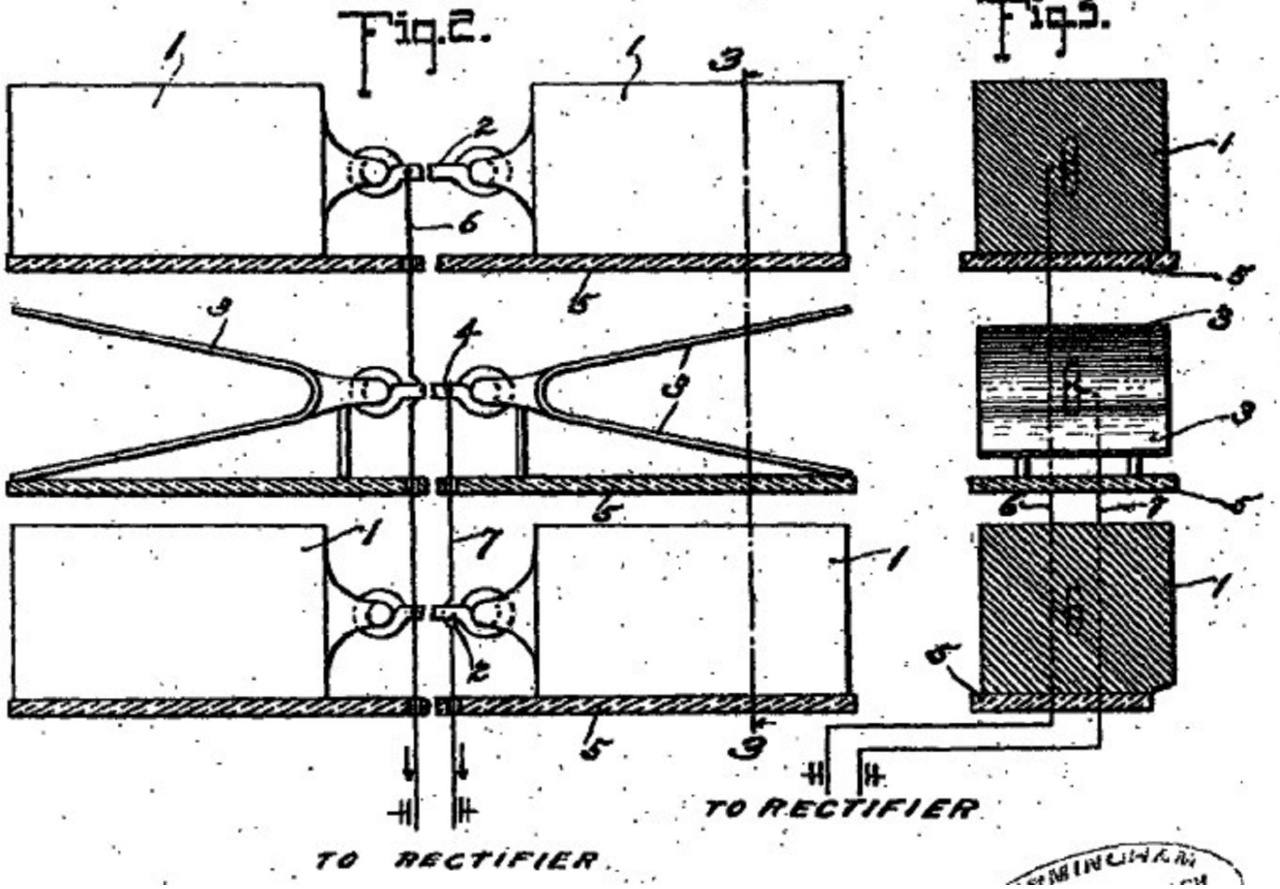
Fig.2.
Fig.3.
1
1
3
2
6
5
3
4
3
5
1
7
1
t
5
3
TO RECTIFIER
TO RECTIFIER
1
5
3
3
6
7
5
5
1

Fig. 7.

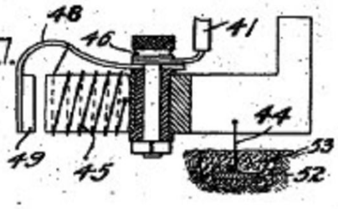
48
46
41
44
53
52
49
45

Fig. 4.

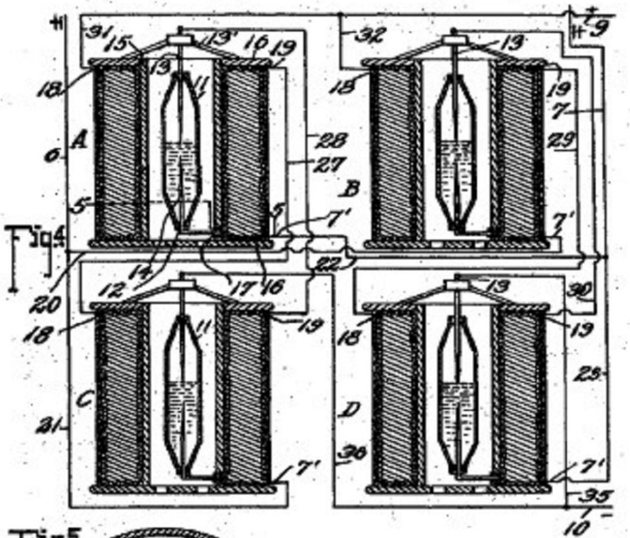
11
31
15
13'
16
19
13
18
A
6
5
28
27
B
7'
32
19
7
29
9
10
12
14
17
16
22
20
18
C
21
D
36
30
13
25
35

Fig. 5.

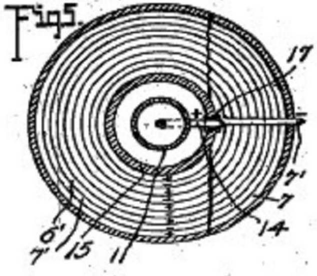
17
7'
7
14
6'
7'
15
11

Your Support Maintains this Service. Please Invest in the mind-boggling Rex Research Website CD -- Always The Best Investment ... Everything on this Website & More (the Bonus Files CD) -- Only $13, Postpaid Anywhere ! Secure Transaction via PayPal: All Major Credit Cards Accepted:

Buy Now

No Credit ? Don't like PayPal ? Send a Check/Money Order to:

Rex Research, PO Box 19250, Jean, NV 89019 USA

rexresearch.com

rexresearch.com

Roy J. MEYERS

Absorber

How To Patent Your Idea
3 Easy Steps To Patent & Make Money Get A Free Patent & Invention Kit
www.Patent.InventionHome.com

Superior X-Ray Solutions
Performance, Reliability & Speed from the world leader in SCD
www.Bruker-AXS.com

Do You Have an Idea?
We Can Help Turn Your Idea Into A Real Product. Get Free Information!
NewProductHelp.com

Buckyballs Magnetic Toy
Get the Gadget on Everyone's List. Best prices on this year's hot toy!
www.getbuckyballs.com

Car Door Magnets: 50% Off
Magnetic Car Signs $4.99. No Mins. Full-Color. Upload Photos & Logos!
www.Vistaprint.com

Power Your Own Home
Reduce Utility Bills by 50% Lower C02 Emissions by 1/3
www.ClearEdgePower.com

Ads by Google

Renewable Clean Energy
Compact Electricity & Heat Unit For Homes & Small Businesses
www.ClearEdgePower.com

Do You Have an Idea?
We Can Help Turn Your Idea Into A Real Product. Get Free Information!
NewProductHelp.com

Patent Attorney/Engineer
Over 36 years experience. Staff available 24/7 by phone.
www.invention.net

C. Edholm: "Picks Power from Air" (*Technology World Magazine*, 1912 ?)
Dr L. Hirshberg: "Electricity from Air New Great Discovery (*Modern Electrics*, 1914 ?)
Roy J. Meyers: British Patent # 1098 (1913) -- **Improvements in and Relating to Apparatus for Producing Electricity (PDF Format, requires Adobe Reader)**
R. J. Meyers: British Patent # 1098 (1913) -- **Transcription w/ enlarged figures**
Notes & Comments
R. J. Meyers: British Patent # 1098 (1913) -- **JPG version**

***Technology World Magazine*, p. 279-281 (Year unknown, apparently circa 1912; another article about Meyers appeared in the November 1912 issue of *Electronic World*)**

"Picks Power from the Air"

by
Charlton Lawrence Edholm

A remarkable scene took place in the legislature of Arizona this spring when the lawmakers enthusiastically voted for the parole of a certain convict in the State penitentiary, granting him a leave of absence for 30 days and by means of private contributions raising a fund to defray his expenses to Washington DC and return.

The prisoner, Roy J. Meyers, is serving a 3-1/2 year sentence, but in spite of the fact that he bears the stigma of a convicted lawbreaker, he has demonstrated that a convict can be a useful member of society. During his imprisonment he perfected an electrical device of such original character as to arouse feelings of wonder and skepticism until experts had seen it in actual operation. It is a device to draw electricity from the atmosphere for light and power, and the 30-day parole was granted in order that the inventor might protect his rights through the patent office at Washington.

With the acquiescence of the legislature, Governor Hunt granted the parole and the prisoner was allowed to go free without any guard or any assurance but his word of honor that he would return. Two days before the period had elapsed, Meyer again presented himself before the governor, having accomplished his mission, and then returned to the penitentiary at Florence, where he continues to serve his sentence.

This, in brief, is the picturesque story which has called attention of the civilized world to a newly discovered electrical genius, and to another feature of the case which is of equal importance and human interest; namely, the enlightened policy pursued by our youngest State in its treatment of convicts...

Before entering the prison, Meyer had already applied for various patents, among them one for an improved trolley wheel head which prevents the trolley wheel from jumping the wire. Meyers had a conference with Superintendent Sims and Parole Clerk Sanders, and it was to these gentlemen that the inventor first explained the principles of his new device for securing electrical energy from the air. The officials were willing to give the man the opportunity to develop his plan and a little wooden building outside the walls was turned over to Meyers and was fitted up as a workshop and a laboratory. The first demonstration of the new apparatus was made shortly thereafter, the electricity drawn from the atmosphere being used to spark the gas engines of the pump house, and although the device was crude yet it did the work, and removed the doubts of his friends. Furtter development of the "absorber" followed, and his second model was constructed, and developed 8 volts. The machine came to the attention of the remarkable woman who brought his name before the legislature.

This was Miss Kate Barnard, State Commissioner of Charities and Corrections of Oklahoma, who was a guest of Mr. Sims, while studying prison conditions. She saw the machine at work, became familiar with the facts of Meyers' case, and was impressed by his rather blunt and unaffected personality, for Meyers has nothing of the polish or glibness of the poseur. He is a simple, earnest student of mechanical problems and not the sort of man to make a sentimental appeal for sympathy because of any grace of person or manner. Therefore it was the value of Meyers' invention, together with his essential integrity (in spite of his lapse) which so strongly impressed Miss Barnard that when she appeared before the Arizona legislature not long afterwards, addressing that body on the need of enlightened legislature along the line of her own work, she told the story of Roy Meyers and his epoch-making invention.

So, early in May, Meyers set out for Washington, unaccompanied.

In his own words: "When I arrived in Washington and laid my plans before the patent office experts, they merely smiled and told me that I would have to build a model and demonstrate my claims --- that it seemed strange that I, unknown as I am in the electrical world, should have accomplished the things for which Edison, Tesla and other experts have been striving for years.

"They could grasp the meaning of my drawings nor the explanation I tried to make to them. There was little time to spare, as I had only 20 days left of my leave, but I set to work in a few days was able to take a crude model around to the patent office to make a demonstration.

"Arriving at the patent office I telephoned to a friend who had been so kind as to introduce me and aid me in reaching the proper officials. The absorber was hoisted on two short poles and made to work. While they were as yet unable to understand the principles involved and hardly willing to believe their eyes, they were forced to admit that I had something new and different, and they told me that there would be no further objection; that I might file my application without further delay.

"I hope to construct my first large machine right here in Phoenix. I feel grateful; to Governor Hunt and others for what they have done for me and to the help they have given in securing protection I might not otherwise have had, and I am desirous of demonstrating this gratitude. I am going back to Florence today to resume the serving of my sentence, which will expire in 10 months. Then, here in Phoenix, I will begin the work of making my machines."

While there are some details of the device which the inventor refuses to make public, yet there are many general features that may be explained. It is planned that the machine, to be set up in Phoenix, will generate sufficient power to light the city, and will consist of a 200 foot tower upon which is placed the "absorber". The latter consists of a series of magnetized steel plates set in a circle (the manner of preparing them is kept secret) and this mechanism attracts the electricity from the atmosphere. This is carried by wires to a transformer in the engine house below and thence is applied to produce either power or light after the usual manner.

In an authorized statement Meyers says: "The flow of electricity is constant. When it emerges into the transformer it is in the form of a direct current. It will absorb the electricity day and night and will work whenever the wireless will work. I can put up a plant to supply such a building as the Adams Hotel for about $1500, and one of the principal items of the expense is the cost of the towers, the wires, the magnetizing of one set of plates, which is part of the secret of the treatment which makes it respond to the accumulations of the atmosphere.

"For use in the case of an electrical storm I have made what I call a modified form of circuit breaker, such as is commonly used as a lightning arrester on telegraph lines. In case of a storm the accumulator would suddenly become overcharged, possibly, and as the electricity would not of itself flow back into the air, the result might be disastrous. So I send it down into the ground, whenever the voltage rises above a certain amount."

Roy J. Meyers & the "Absorber"

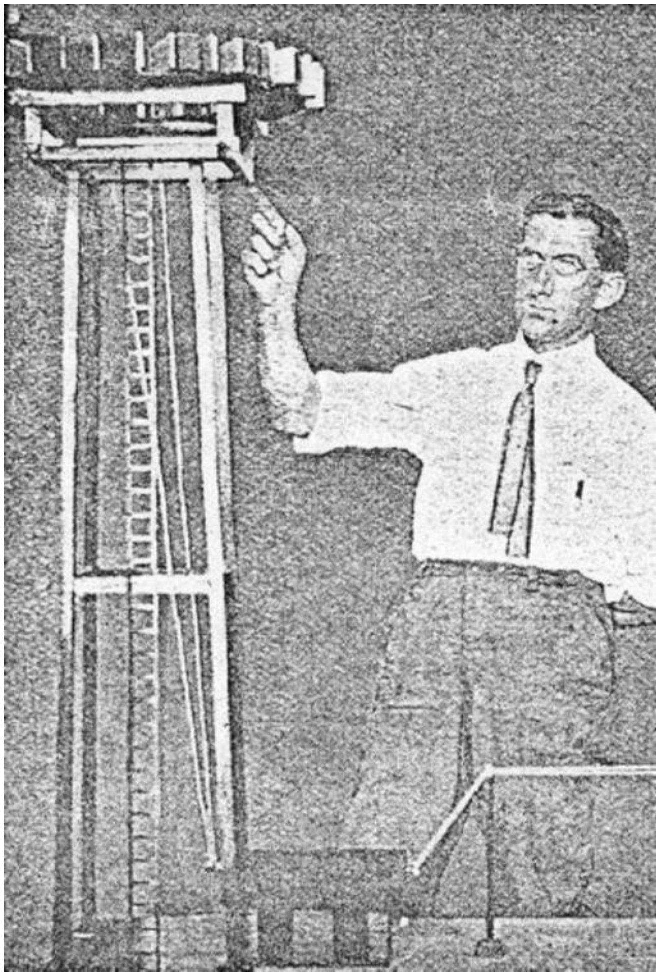

Modern Electronics **(1914 ?)**

Electricity From Air New Great Discovery

by

Dr Leonard Keene Hirshberg

Working quietly in the heart of Baltimore for weeks on an invention which some critics say will revolutionize the method of converting electricity to practical use has been Roy J. Meyers, who like Benjamin Franklin, extracts the electric current from the air.

Mr Meyers invention was made last summer while he was confined in the penitentiary at Florence, Arizona. His first finished apparatus was made in Baltimore.

A practical, unlettered electrician, Mr Meyers, while in Arizona, was arrested on a comparatively minor charge and sent to the penitentiary. There he was placed in charge of the prison electrical plant, and there he says he made his discovery that the current which the civilized world is beginning to use most extensively for light and power could be transformed from the atmosphere without the aid of moving machines or batteries.

Miss Kate Barnard, Commissioner of Charities and Corrections, of Oklahoma, hearing of Meyers' invention and of his desire to have it patented, appeared before the Arizona Legislature to make an appeal in behalf of the young convict. As a result a special bill was passed which granted Meyers a month's leave of absence on parole. He went unaccompanied to Washington, filed his patent applications and returned to the penitentiary. Since then he has been indefinitely paroled.

He came to Baltimore as the place where he could easily obtain the mechanical parts needed to make a more nearly perfect machine than the crude model he has fashioned in the penitentiary workshop, and is making his headquarters here while working on his invention. With him is W.E. Chenot, who has been his assistant in assembling and testing the machine and who says that he has bought Meyers' patent rights for Germany.

They have proved beyond doubt that the invention is practical and that when finally brought to a state of perfection it will introduce a new epoch in the industrial use of electricity. By Westinghouse meters they tested the strength of the current gathered from the air, and with the use of only two of the four rectifying transformers the voltmeter recorded four and one-half volts, and the ammeter, which had the capacity of recording 75 amperes, was broken by the force of the current.

The machine itself is simple. It is in reality a transformer, which is familiar to anyone knowing anything at all about electricity in its practical uses. On a high tripod, which resembles somewhat the framework of a windmill tower, is the transformer, which Mr Meyers calls his 'absorber'. It is made up of an iron core, wrapped with copper wire. The secret of the invention is the manner in which the disks composing this 'absorber' are magnetized, and this secret Meyers says he found by accident while at work in prison.

What the machine, when finally perfected, will do is yet to be seen. Its inventor claims that it will greatly reduce the cost of making electricity. No batteries of any kind are needed, he says, and not a part of the machine turns upon the other. It is as durable, apparently, as an electric light pole. One of these machines, says Meyers, when perfected may be placed on a vehicle and transform enough electricity to give motive power, be that vehicle a locomotive or an automobile. He declared it can be placed on a building to furnish electric lights or power, and that the only wear will be upon the machinery which its current runs.

Meyers is 34 years old and he gained his knowledge of electricity by working in shops along the Pacific Coast. The depths of the mysteries of electricity he has not explored, but he is certain that he has found the means of absorbing it from the air and of converting it to the use of mankind.

British (GB) Patent # 191301098

Improvements in and Relating to Apparatus for Producing Electricity.

1-14-1914

Roy Jerome Meyers

Classification: - international: H05F7/00; H05F7/00; - european: H05F7/00
Application number: GBD191301098 19130114
Priority number(s): GBT191301098 19130114

Abstract ~ Vapour apparatus, arrangements of. - A rectifier for use with apparatus for producing electricity from the earth consists of mercury- vapour lamps constructed and arranged as shown in Fig. 4. Each lamp comprises two wires 6<1>, 7<1> wound around a steel tube 15 surrounding a

mercury tube 11 preferably of copper. The coil 6<1> is connected between the electrode 14 and the terminal 18, and the coil 7<1> between the terminals 19, 5. The coils 6<1>, 7<1> are preferably composed of soft iron. Reference has been directed by the Comp- troller to Specifications 16,709/87, 14,033/99, and 5457/11, [all in Class 53, Galvanic batteries], and 15,412/06.

British Patent # 1098

(January 14, 1913)

Improvements in and Relating to Apparatus for Producing Electricity

Roy Jerome Meyers

This invention relates to improvements in apparatus for the production of electrical currents, and the primary object in view is the production of a commercially serviceable electrical current without the employment of mechanical or chemical action. To this end the invention comprises means for producing what I believe to be dynamic electricity from the earth and its ambient elements.

I am, of course aware that it has been proposed to obtain static charges from upper strata of the atmosphere, but such charges are recognized as of widely variant potential and have thus far proved of no practical commercial value, and the present invention is distinguished from all such apparatus as has heretofore been employed for attracting static charges by the fact that this improved apparatus is not designed or employed to produce or generate irregular, fluctuating or other electrical charges which lack constancy, but on the other hand I have by actual test been able to produce from a very small apparatus at comparatively low elevation, say about 50 or 60 feet above the earth's surface, a substantially constant current at a commercially usable voltage and amperage. This current I ascertained by repeated tests is capable of being readily increased by additions of the unit elements in the apparatus hereinafter set forth, and I am convinced from the constancy of the current obtained and its comparatively low potential that the current is dynamic and not static, although, of course, it is not impossible that certain static discharges occur and, in fact, I have found occasion to provide against the damage which might result from such discharge by the provision of lightning arresters and cut-out apparatus which assist in rendering the obtained current stable by eliminating sudden fluctuations which sometimes occur during conditions of high humidity from what I consider static discharges. The nature of my invention is obviously such that I have been unable to establish authoritatively all of the principles involved, and some of the theories herein expressed may possibly prove erroneous, but I do know and am able to demonstrate that the apparatus which I have discovered does produce, generate, or otherwise acquire a difference of potential representing a current amperage above stated, or varied therefrom at the will of the operator according to the uses which the current is to be subjected.

The invention comprises generically means for producing electrical currents of serviceable potential substantially without the employment of mechanical or chemical action, and in this connection I have been able to observe no chemical action whatever on the parts utilized although deterioration may possibly occur in some of the parts, but so far as I am able to determine such deterioration does not add to the current supply but is merely incidental to the effect of climatic action.

The invention more specifically comprises the employment of a magnet or magnets and a co-operating element, such as zinc disposed adjacent to the magnet or magnets and connected in such manner and arranged relative to the earth so as to produce current, my observation being that current is produced only when such magnets have their poles facing substantially to the north and south and the zincs are disposed substantially along the magnets.

The invention also comprehends other details of construction, combinations and arrangements of parts as will hereinafter be fully set forth and claimed.

In the accompanying drawings:

Figure 1 is a top plan view of an apparatus embodying the features of the present invention, the arrow accompanying the figure indicating substantially the geographical north, parts of the figure being diagrammatic for condensing the showing.

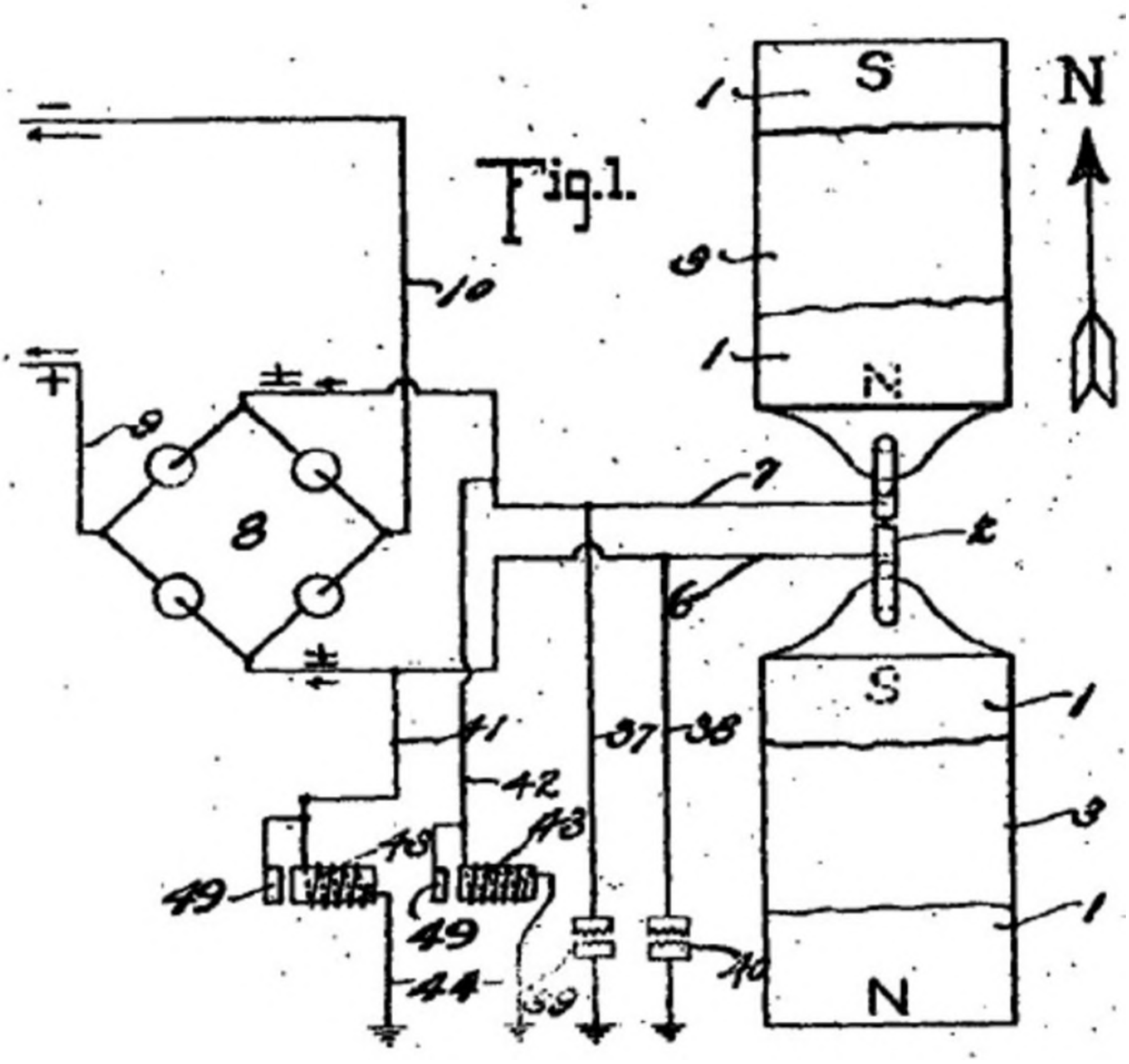

Figure 2 is a view is side elevation of the parts seen in plan in Figure 1.

Fig.2.

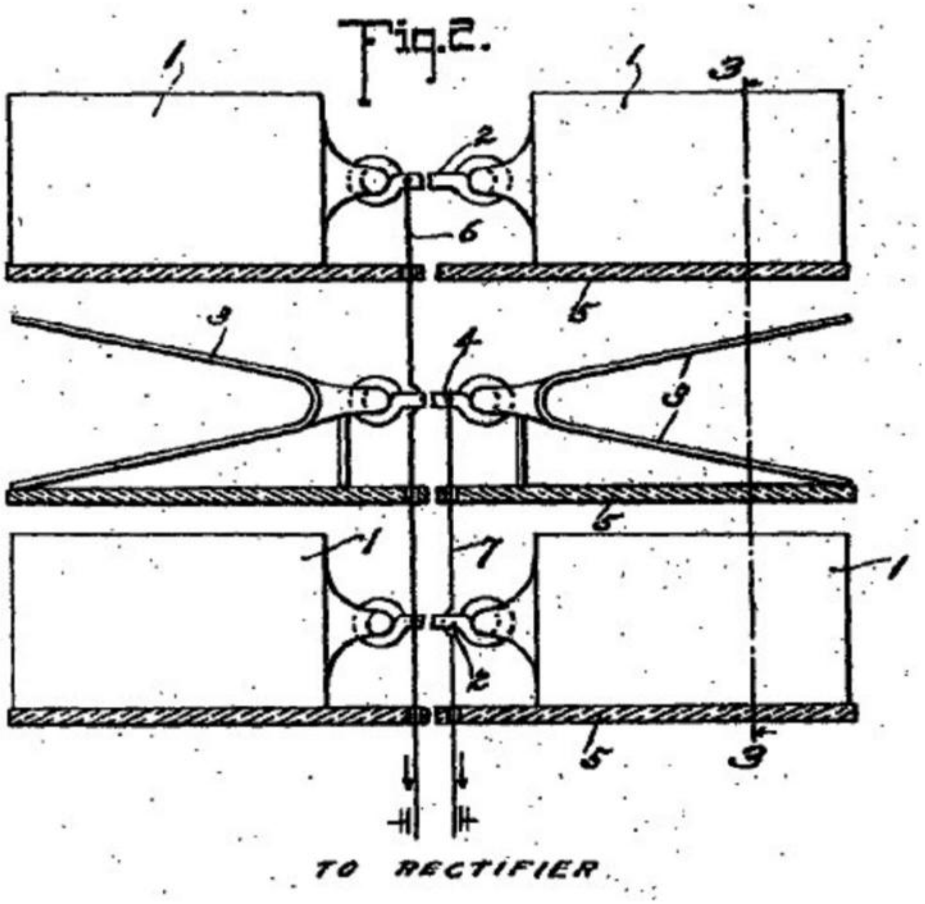
1
1
3
2
6
5
9
4
9
6
1
7
1
2
5
3
TO RECTIFIER

Figure 3 is a vertical section taken on the plane indicated by the line 3-3 of Figure 2 and looking in the direction indicated by the arrow.

Fig.5.

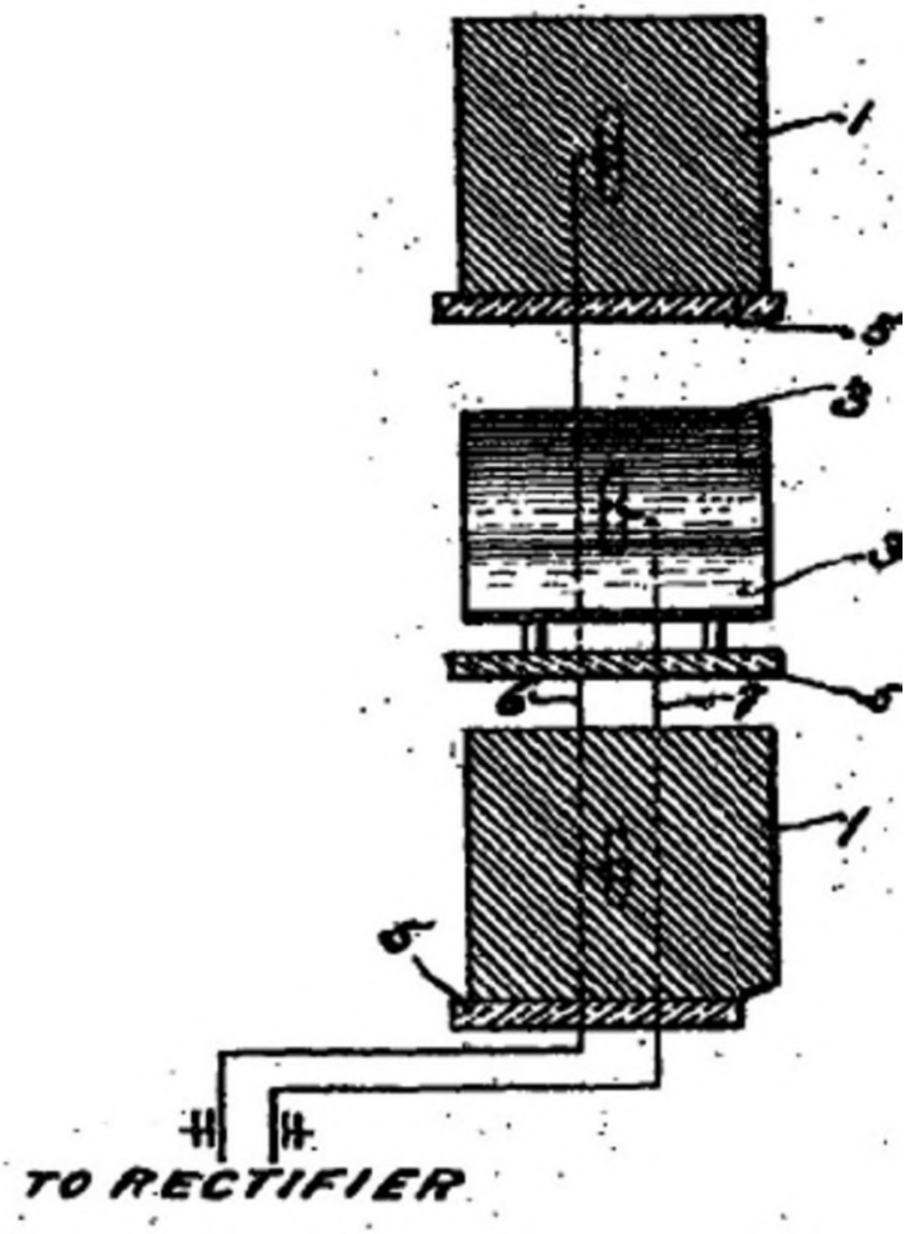

Figure 4 is a detail view partly in elevation and partly in section showing the detail connections of the converter and intensifier.

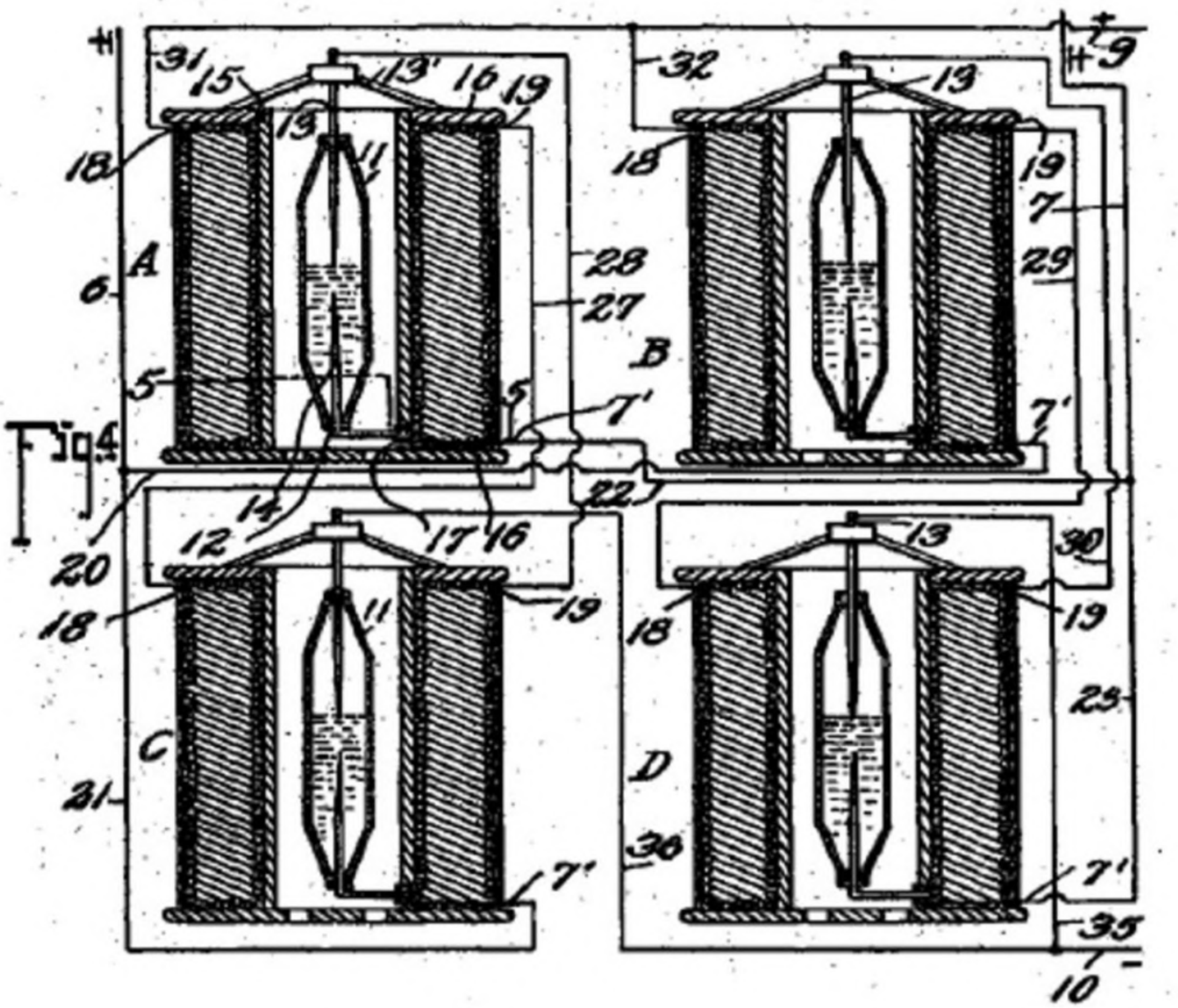
Fig. 4

31
15
13'
16
19
32
13
9
18
13
11
A
6
28
27
B
5
7'
20
14
12
17
16
22
30
18
19
29
7
C
21
D
36
35
10
23

Figure 5 is a transverse section taken on the planes indicated by line 5-5 of Figure 4 and looking downwardly.

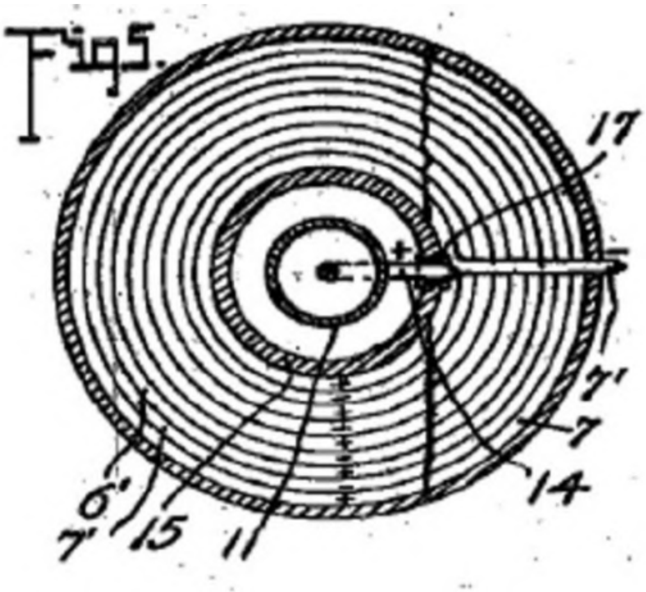
Fig. 5

17
7'
7
14
6'
7'
15
11

Figure 6 is an enlarged detail fragmentary section illustrating the parts at the juncture of the conductors and one of the intensifiers.

Fig. 6

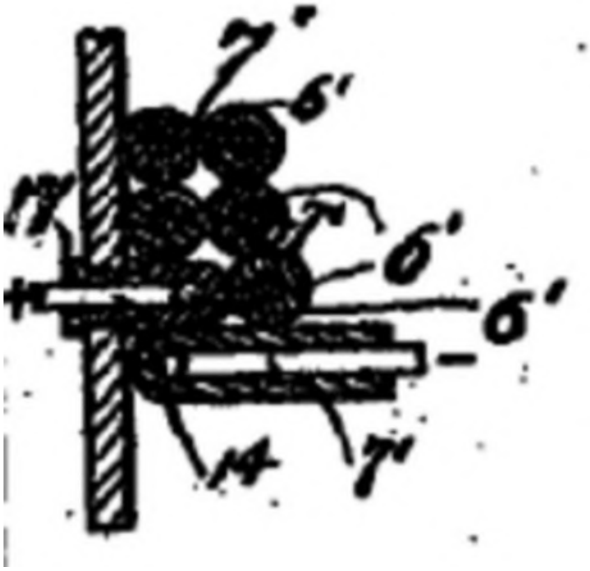

Figure 7 is an enlarged detail view partly in elevation and partly in section of one of the automatic cut-outs and

Fig. 7.

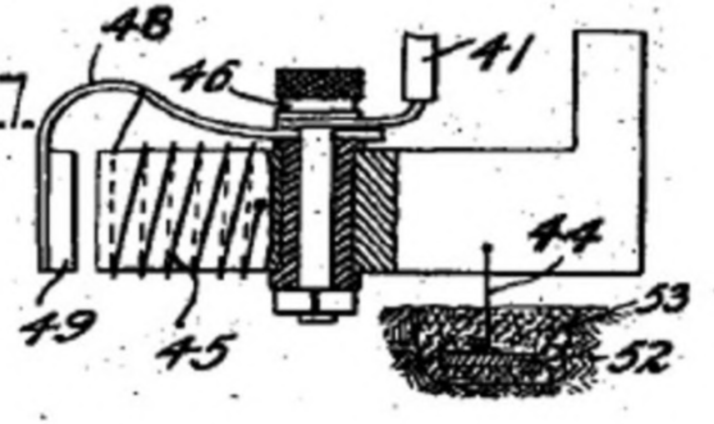

Figure 8 is a diagrammatic view of one of the simplest forms of embodiment of the invention.

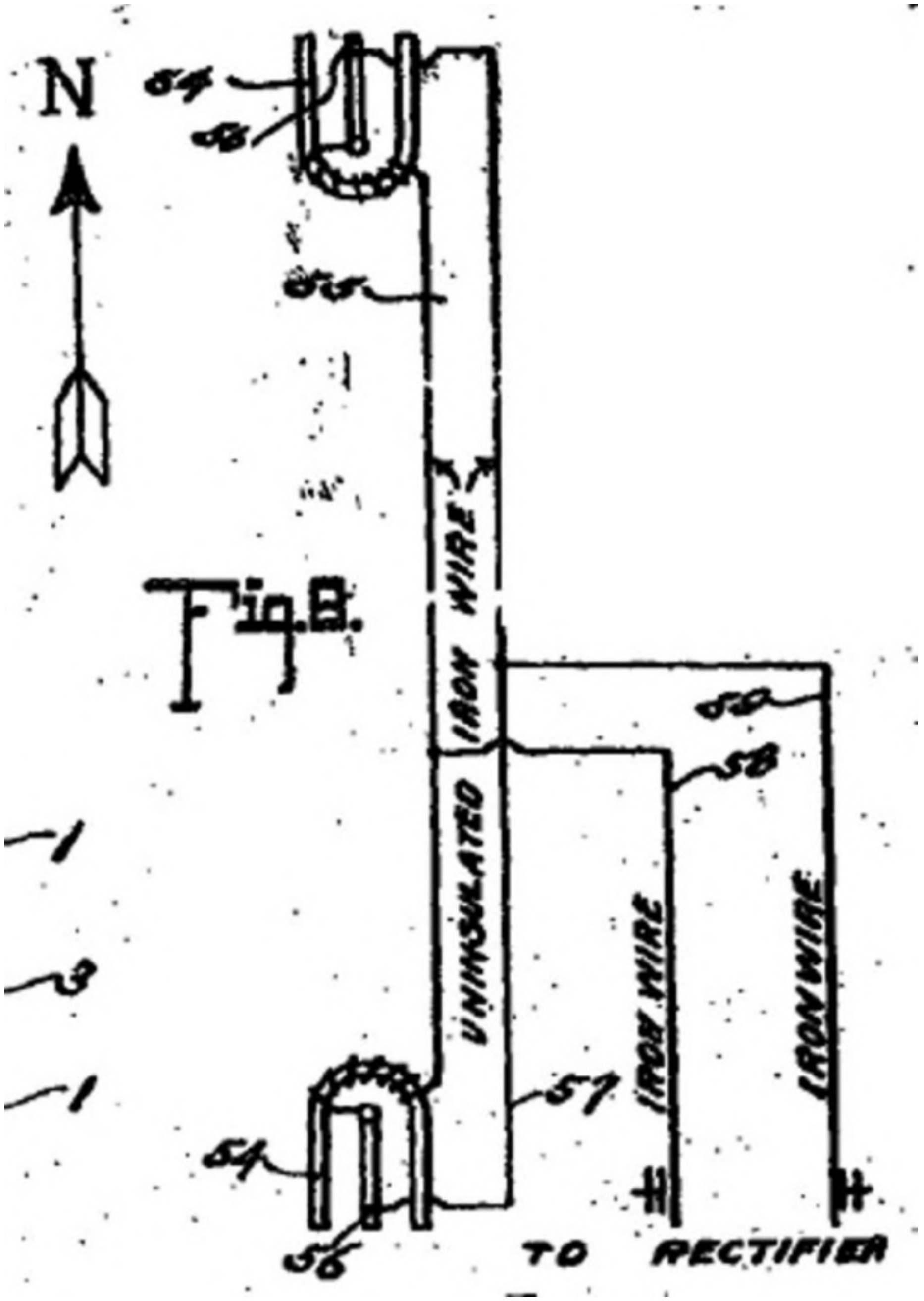

Fig. 8.

Referring to the drawing by numerals, 1,1 indicates magnets connected by a magnetic substance 2, preferably an iron wire. The magnets 1 are arranged in pairs, one pair being spaced beneath the other, and interposed between the magnets are zinc plates 3,3 connected by an iron wire conductor 4. Suitable insulating supports 5 are arranged for sustaining the respective magnets 1 and plates 3,3. Each plate 3 is preferably bent substantially into V form, as clearly seen in Figure 1, and the V1 of one of the plates opens or faces toward the north and the V of the other plate to the South. I have determined by experimentation that it is essential that the plates 3 be disposed substantially north and south with their flat faces approximately parallel to the adjacent faces of the co-operating magnets, although by experience I have not discovered any material difference in the current obtained when the plates are disposed slightly to one side of north and south, as for instance when the plates are disposed slightly to one side of north and south, as for instance when disposed in the line of the magnetic polarity of the earth. The same is true with respect to the magnets 1, the said magnets being disposed substantially north and south for operative purposes, although I find that it is immaterial whether the north pole of one of the magnets is disposed to the north and the south pole to the south, or vice versa, and it is my conviction from experience that it is essential to have the magnets of each pair connected by magnetic material so that the magnets substantially become one with a pole exposed to the north and a pole exposed to the north. In Figure 1, I have indicated in full lines by the letters 8 and N the respective polarities of the magnets 1, and have indicated in dotted lines the other pole of those magnets when the connection 2 is severed. I have found that the magnets and zinc plates operate to produce, whether by collection or generation I am not certain, electrical currents when disposed substantially north and south, but when disposed substantially east and west no such currents are produced. I also find that the question of elevation is by no means vital, but it is true that more efficient results are obtained by placing the zincs and magnets on elevated supports. I furthermore find from tests that it is possible to obtain currents from the apparatus with the zincs and magnets disposed in a building or otherwise enclosed, although more efficient results are obtained by having the said elements arranged in the open.

While in Figures 1, 2, and 3, I have shown the magnets and the zinc plates as superimposed, it will be apparent, as hereinafter fully set forth, that these elements may be juxtaposed in horizontal planes, and substantially the same results will be secured. Furthermore, the magnets 1 with the interposed zincs 3, as shown in Figures 1, 2 and 3 merely represent a unit which may be repeated either horizontally or vertically for increasing the current supply, and when the unit is repeated the zinc plates are arranged alternating with the magnets throughout the entire series as hereinafter indicated.

A conductor 6 is connected in multiple with the conductors 2 and a conductor 7 is connected with conductor 4, the conductor 6 extending to one terminal of a rectifier which I have indicated by the general reference character 8, and the conductor 7 extending to the other terminal of said rectifier. The rectifier as seen in diagram in Figure 1 may assume any of several well known embodiments of the electrical valve type and may consist of four asymmetric cells or Cooper-Hewitt mercury vapor lamps connected as indicated in Figure 1 for permitting communication of the positive impulses from the conductor 6 only to the line conductor 9 and the negative impulses from conductor 6 on only to the line conductor 10. The current from this rectifier may be delivered through the conductors 9 and 10 to any suitable source for consumption.

While the said rectifier 8 may consist of any of the known types, as above outlined, it preferably consists of a specially constructed rectifier which also has the capacity of intensifying the current and comprises specifically the elements shown in detail in Figures 4, 5, and 6 wherein I have disclosed the detail wiring of the rectifier when composed of four of the rectifying and intensify in elements instead of asymmetric cells or simple mercury vapor valves. As each of these structures is an exact embodiment of all the others, one only will be described, and the description will apply to all. The rectifying element of each construction consists of a mercury tube 11 which is preferably formed of glass or other suitable material, and comprises a cylinder having its end portions tapered and each terminating in an insulating plug or stopper 12. Through the upper stopper 12 is extended the electrode 13 which extends well into the tube and preferably substantially one-half the length thereof to a point adjacent the inner end of an opposing electrode 14 which latter electrode extends thence downwardly through the insulation 12 at the lower end of the tube. The tube 11 is supplied with mercury and is adapted to operate on the principle of the mercury vapor lamp, serving to rectify current by checking back impulses of one sign and permitting passage of impulses of the other. To avoid the necessity for utilizing a starter, as is common with the lamp type of electrical valve, the supply of mercury within the tube may be sufficient to contact with the lower end of the electrode 13 when current is not being supplied, so that as soon as current is passed from one electrode to the other sufficiently for volatilizing that portion of the mercury immediately adjacent the lower end of electrode 13, the structure begins its operation as a rectifier. The tube 11 is surrounded by a tube 15 which is preferably spaced from tube 11 sufficiently for allowing atmospheric or other cooling circulation to pass the tube 11. In some instances, it may be desirable to cool the tube 11 by a surrounding body of liquid, as hereinafter indicated. The tube 15 may be of insulating material but I find efficient results attained by the employment of a steel tube, and fixed to the ends of the of the tube are insulating disks 16, 16 forming a spool on which are wound twin wires 6^1 and 7^1, the wire 6^1 being connected at the inner helix of the coil with the outer end of the electrode 14, the lower portion of said electrode being extended to one side of the tube 11 and passed through an insulating sleeve 17 extending through the tube 15, and at its outer end merging into the adjacent end of the wire 6^1. The wire 7^1 extends directly from the outer portion of the spool through the several helices to a point adjacent the juncture of the electrode 14 with wire 6^1 and thence extends in mechanical parallelism with the wire throughout the coil, the wire 6^1 ending in a terminal 18 and the wire 7^1 ending in a terminal 19. For the sake of convenience of description and of tracing the circuits, each of the apparatus just above described and herein known as an intensifier and rectifier will be mentioned as A, B, C and D, respectively. Conductor 6 is formed with branches 20 and 21 and conductor 7 is formed with similar branches 22 and 23. Branch 20 from conductor 6 connects with conductor 7^1 of intensifier B and branch 21 of conductor 6 connects with the conductor 7^1 of intensifier C, while branch 22 of conductor 7 of intensifier C, while branch 22 of conductor 7 connects with conductor 7^1 of intensifier D. A conductor 27 is connected with terminal 19 of intensifier A and extends to and is connected with the terminal 18 of intensifier C, and a conductor 7 connects with conductor 7^1 of intensifier D. A conductor 27 is connected with terminal 19 of intensifier A. and extends to and is connected with terminal 18 of intensifier C, and a conductor 28 is connected with the terminal 19 of intensifier C and extends from the terminal 19 of intensifier B to the terminal 18 of intensifier D to electrode 13 of intensifier B. Each electrode 13 is supported on a spider 13^1 resting on the upper disk 16 of the respective intensifier. Conductors 31 and 32 are connected with the terminals 18 of intensifiers A and B and are united to form the positive line wire 9 which co-operates with the negative line wire 10 and extends to any suitable point of consumption. The line wire 10 is provided with branches 35 and 36 extending to the electrodes 13 of intensifiers C and D for completing the negative side of the circuit.

Thus it will be seen that alternating currents produced in the wires 6 and 7 will be rectified and delivered in the form of a direct current through the line wires 9 and 10, and I find by experiment that the wires 6 and 7 should be of iron, preferably soft, and may of course be insulated, the other wiring not specified as iron being of copper or other suitable material.

In carrying out the operation as stated, the circuits may be traced as follows: A positive impulse starting at the zincs 3 is directed along conductor 7 to branch 23 to conductor 7^1 and the winding of the rectifier of intensifier B through said rectifier to the conductor 6^1, through the winding thereof to the contact 18, conductor 32 and to the line wire 9. The next or negative impulse directed along conductor 7 cannot find its way along branch 23 and the circuit just above traced because it cannot pass across the rectifier of intensifier B but instead the negative impulse passes along conductor 22 to conductor 7 of intensifier A and the winding thereof to the contact 19 and to conductor 27 to contact 18 of intensifier C, to the winding of the wire 6^1 thereof to the electrode 14 through the rectifier to the of the electrode 13 and conductor of intensifier A, electrode 14 thereof and conductor 6^1 to contact 18 and wire 31 to line wire 9. Obviously the positive impulse cannot pass along the wire 20 because of its inverse approach to the rectifier of intensifier B. The next impulse or negative impulse delivered to conductor 6 cannot pass along conductor 21 because of its connection with electrode 13 of the rectifier of intensifier A, but instead passes along conductor 20 to the wire 7^1 and its winding forming part of intensifier B to the contact 19 and conductor 29 to contact 18 and the winding of wire 6^1 of intensifier D to the electrode 14 and through the rectifier to the electrode 13 and conductor 35 to line wire 10. Thus the current is rectified and all positive impulses directed along one line and all negative impulses along the other lie s that the potential difference between the two lines will be maximum for the given current of the alternating circuit. It is, of course, apparent that a less number of intensifiers with their accompanying rectifier elements may be employed with a sacrifice of the impulses which are checked back from a lack of ability to pass the respective rectifier elements, and in fact I have secured efficient results by the use of a single intensifier with its rectifier elements, as hereinafter set forth.

Grounding conductors 37 and 38 are connected respectively with the conductors 6 and 7 and are provided with the ordinary lightning arresters 39 and 40 respectively for protecting the circuit against high tension static charges.

Conductors 41 and 42 are connected respectively with the conductors 6 and 7 and each connects with an automatic cutout 43 which is grounded as at 4. Each of said automatic cutouts is exactly like the other and one of the same is shown in detail in Figure 7 and comprises the inductive resistance 45 provided with an insulated binding post 46 wit which the respective conductor 6 or 7 is connected, said post also supporting a spring 48 which sustains an armature 49 adjacent to the core of the resistance 45. The helix of resistance 45 is connected preferably through the spring to the binding post at one end and at the other end is grounded on the core of the resistance, the said core being grounded by ground conductor 44 which extends to the metallic plate 52 embedded in moist carbon or other inductive material buried in the earth. Each of the conductors 41, 42 and 44 is of iron, and in this connection I wish it understood that where I state the specific substance I am able to verify the accuracy of the statement by the results of tests which I have made, but of course I wish to include along with such substances al equivalents, as for instance, where iron is mentioned its byproducts, such as steel, and its equivalents such as nickel and other magnetic substances are intended to be comprehended. The cutout apparatus seen in detail in Figure 7 is employed particularly for insuring against high tension currents, it being obvious from the structure shown that when potential rises beyond the limit established by the tension of the spring sustaining the armature 40, the armature will be moved to a position contacting with the core of the cutout device and thereby directly close the ground connection for line wire 41 with conductor 44, eliminating the resistance of winding 45 and allowing the high tension current to be discharged to the ground. Immediately upon such discharge the winding 45 losing its current will allow the core to become demagnetized and release the armature 49 whereby the ground connection is substantially broken leaving only the connection through the winding 45 the resistance of which is sufficient for insuring against loss of low tension current.

In Figure 8 I have illustrated an apparatus which though apparently primitive in construction and arrangement comprehends the first successful embodiment which I produced in the course of discovery of the present invention, and it will be observed that the essential features of the invention are therein disclosed. The structure delineated in said figure consists of horseshoe magnets 54, 55, one facing north and the other south, that is, each opening in the respective directions indicated and the two being connected by an iron wire 55 which is uninsulated and wrapped about the respective magnets each end portion of the wire 55 being extended from the respective magnets to and connected with, as by being soldered to, a zinc plate 56, there being a plate 56 for each magnet and each plate being arranged longitudinally substantially parallel with the legs of the magnet and with the faces of the plate exposed toward the respective legs of the magnet, the plate being thus arranged endwise toward the north and south. An iron wire 57 connects the plates 56, the ends of the wire being preferably connected adjacent the outer ends of the plates but from experiment I find that the wire may be connected at practically any point to the plate. Lead wires 58 and 59 are connected respectively with the wires 55 and 57 and supply an alternating current at a comparatively low tension, and to control such current the wires 58 and 59 may be extended to a rectifier or combined rectifier and intensifier, as above set forth.

The tests which I have found successful with the apparatus seen in Figure 8 were carried out by the employment first of horseshoe magnets approximately 4 inches in length, the bar comprising the horseshoe being about one inch square, the zincs being dimensioned proportionately and from this apparatus with the employment of a single intensifier and rectifier, as above stated, I was able to obtain a constant current of 8 volts.

It should be obvious that the magnets forming one of the electrodes of this apparatus may be permanent or may be electromagnets, or a combination of the two.

While the magnets mentioned throughout the above may be formed of any magnetic substance, I find the best results obtained by the employment of the nickel chrome steel.

While the successful operation of the various devices which I have constructed embodying the present invention have not enabled me to arrive definitely and positively at fixed conclusion relative to the principles and theories of operation and the source from which current is supplied, I wish it to be understood that I consider myself as the first inventor of the general type hereinbefore described capable of producing commercially serviceable electricity, for which reason my claims hereinafter appended contemplate that I may utilize a wide range of equivalents so far as concerns details of construction suggested as preferably employed.

The current which I am able to obtain is dynamic in the sense that it is not static and its production is accomplished without chemical or mechanical action either incident to the actual chemical or mechanical motion or incident to changing caloric conditions so that the elimination of necessity for the use of chemical or mechanical action is to be considered as including the elimination of the necessity for the use of heat or varying degrees thereof.

Having now particularly described and ascertained the nature of my said invention, and in what manner the same is to be performed, I declare that what I claim is: --- [Claims not included here]

NOTES & COMMENTS

From the Article in *Tech. World Mag*.:

1. First demo model was powerful enough to spark a gas engine.
2. Second model developed 8 volts.
3. Demo model at Patent Office was elevated on short poles.
4. The model planned to power Phoenix AZ would be elevated 200 feet.
5. The Absorber "consists of a series of magnetized steel plates set in a circle (the manner of preparing them is kept secret)".
6. "[T]he magnetizing of one set of plates... is part of the secret of the treatment which makes it respond to the accumulations of the atmosphere".

From British Patent # 1098 (1913):

"I have been able... to produce from a very small apparatus at comparatively low elevation, say about 50 or 60 feet above the earth's surface, a substantially constant current at a commercially useable voltage and amperage".

"This current... is capable of being readily increased by additions of the unit elements in the apparatus".

Fig. 1 and Fig. 2 show the magnet poles are connected N-S by a thick iron rod (thick compared to the lines used for wires in the drawings).

No angle is specified for the V-shaped zinc plates. The article (but not the patent) states that the plates are magnetized (obviously not zinc). Zinc-galvanized steel? Will a thin film of Zn work? Or, powdered Zn in a binder (more surface area)? Or, zinc-galvanized iron wire in a coil?

The Palmer Craig device (www.rexresearch.com/craig/craig.htm) is powered by the terrestrial magnetic field, and employs a thin film of bismuth to capture the energy as diamagnetism. Perhaps this can be integrated with Meyers' device.

Figure 8 (the demonstration of principle) show uninsulated iron wire being used to connect the plates and magnets. The wire is wound around the Bloch wall area of the horseshoe magnets. Perhaps Coler-type windings around the poles could be used here (See: www.rexresearch.com/coler/coler.htm). Coler used copper plates as "condensers" in his device. Could copper plates be used for the Meyers device? Perhaps flat (Tesla non-inductive) coils could be integrated here.

"It is essential that the plates 3 be disposed substantially N and S with their flat faces approximately parallel to the adjacent faces of the co-operating magnets....

"I find that it is immaterial whether the N pole of one of the magnets is disposed to the N and the S pole to the S, or vice versa".

"[T]he magnets and zinc plates... produce electrical currents when disposed... N and S, but when disposed... E and W no such currents are produced".

"[E]levation is by no means vital, but... more efficient results are obtained by placing the zincs and magnets on elevated supports".

"The elements may be disposed in horizontal planes [or vertically]...".

The "zinc plate 56... [is] arranged longitudinally substantially parallel with the legs of the magnet and with the faces of the plate exposed toward the respective legs of the magnet, the plate being thus arranged endwise toward the north and south".

The first model used "horseshoe magnets approximately 4 inches in length, the bar comprising the horseshoe being about one inch square, the zincs being dimensioned proportionately and from this apparatus with the employment of a single intensifier and rectifier, as above stated, I was able to obtain a constant current of 8 volts... [T]he magnets... may be permanent or may be electromagnets, or a combination of the two... I find the best results obtained by... nickel chrome steel".

Comments & Questions:

The rectifier is described as a preferred embodiment, but other designs also work. The Ed Gray capacitor design comes to mind (www.rexresearch.com/evgray/1gray.htm). The Tate Ambient Power Module also might apply (www.rexresearch/tate/tate.htm).

Would non-ferrous magnets work? Is there a frequency involved (oscilloscope tests)? Coler found that ferromagnetism has a resonant frequency about 180 KHz. Can the components be made adjustable for RLC-resonance?

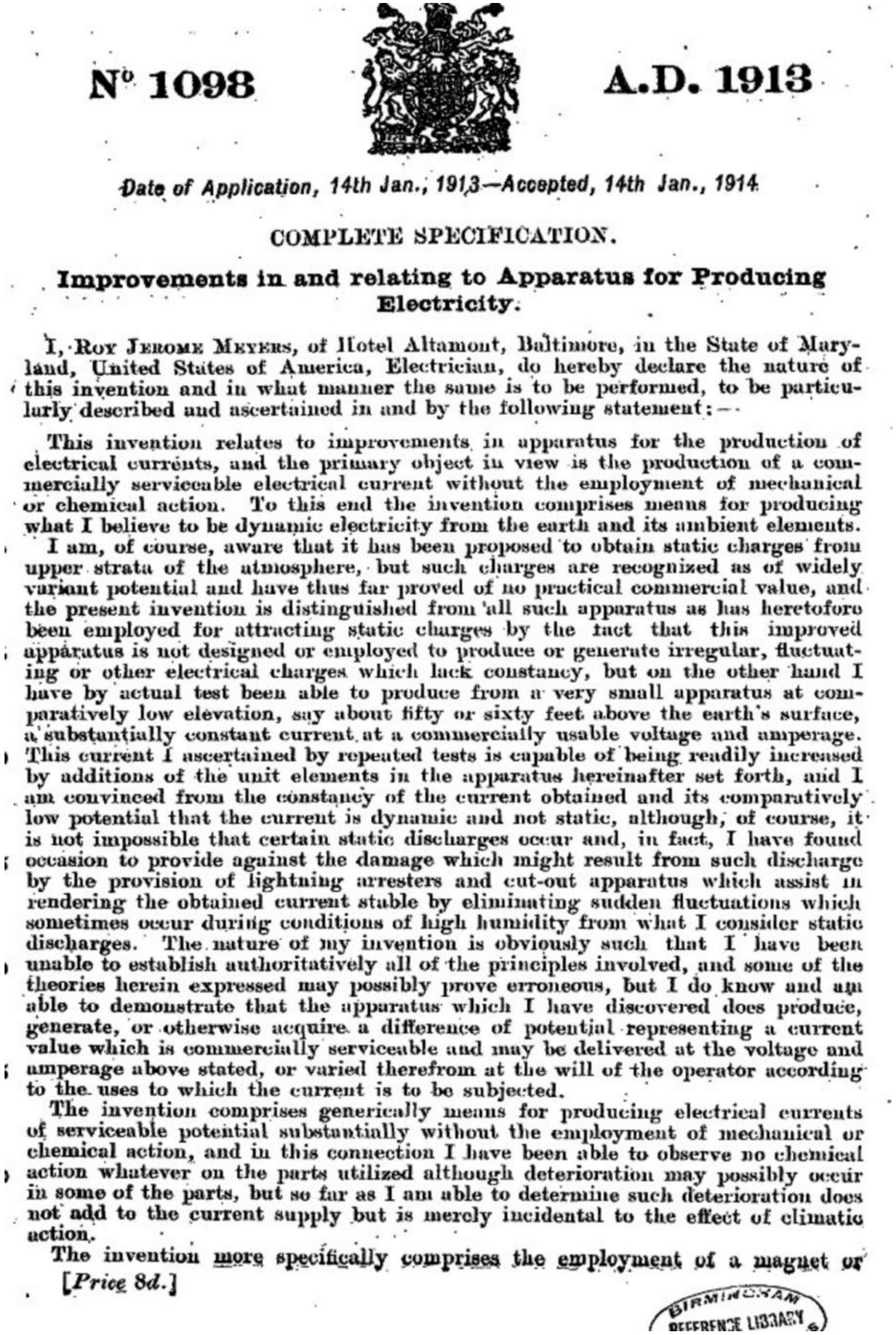

Nº 1098 A.D. 1913

Date of Application, 14th Jan., 1913—Accepted, 14th Jan., 1914

COMPLETE SPECIFICATION.

Improvements in and relating to Apparatus for Producing Electricity.

I, ROY JEROME MEYERS, of Hotel Altamont, Baltimore, in the State of Maryland, United States of America, Electrician, do hereby declare the nature of this invention and in what manner the same is to be performed, to be particularly described and ascertained in and by the following statement:—

This invention relates to improvements in apparatus for the production of electrical currents, and the primary object in view is the production of a commercially serviceable electrical current without the employment of mechanical or chemical action. To this end the invention comprises means for producing what I believe to be dynamic electricity from the earth and its ambient elements. I am, of course, aware that it has been proposed to obtain static charges from upper strata of the atmosphere, but such charges are recognized as of widely variant potential and have thus far proved of no practical commercial value, and the present invention is distinguished from all such apparatus as has heretofore been employed for attracting static charges by the fact that this improved apparatus is not designed or employed to produce or generate irregular, fluctuating or other electrical charges which lack constancy, but on the other hand I have by actual test been able to produce from a very small apparatus at comparatively low elevation, say about fifty or sixty feet above the earth's surface, a substantially constant current at a commercially usable voltage and amperage. This current I ascertained by repeated tests is capable of being readily increased by additions of the unit elements in the apparatus hereinafter set forth, and I am convinced from the constancy of the current obtained and its comparatively low potential that the current is dynamic and not static, although, of course, it is not impossible that certain static discharges occur and, in fact, I have found occasion to provide against the damage which might result from such discharge by the provision of lightning arresters and cut-out apparatus which assist in rendering the obtained current stable by eliminating sudden fluctuations which sometimes occur during conditions of high humidity from what I consider static discharges. The nature of my invention is obviously such that I have been unable to establish authoritatively all of the principles involved, and some of the theories herein expressed may possibly prove erroneous, but I do know and am able to demonstrate that the apparatus which I have discovered does produce, generate, or otherwise acquire a difference of potential representing a current value which is commercially serviceable and may be delivered at the voltage and amperage above stated, or varied therefrom at the will of the operator according to the uses to which the current is to be subjected.

The invention comprises generically means for producing electrical currents of serviceable potential substantially without the employment of mechanical or chemical action, and in this connection I have been able to observe no chemical action whatever on the parts utilized although deterioration may possibly occur in some of the parts, but so far as I am able to determine such deterioration does not add to the current supply but is merely incidental to the effect of climatic action.

The invention more specifically comprises the employment of a magnet or

[*Price 8d.*]

BIRMINGHAM REFERENCE LIBRARY

magnets and a co-operating element, such as zinc disposed adjacent to the magnet or magnets and connected in such manner and arranged relative to the earth so as to produce current, my observation being that current is produced only when such magnets have their poles facing substantially to the north and south and the zincs are disposed substantially along the magnets.

The invention also comprehends other details of construction, combinations and arrangements of parts as will hereinafter be fully set forth and claimed.

In the accompanying drawings:

Figure 1 is a top plan view of an apparatus embodying the features of the present invention, the arrow accompanying the figure indicating substantially the geographical north, parts of the figure being diagrammatic for condensing the showing.

Figure 2 is a view in side elevation of the parts seen in plan in Figure 1.

Figure 3 is a vertical section taken on the plane indicated by the line 3—3 of Figure 2 and looking in the direction indicated by the arrow.

Figure 4 is a detail view partly in elevation and partly in section showing the detail connections of the converter and intensifier.

Figure 5 is a transverse section taken on the planes indicated by line 5—5 of Figure 4 and looking downwardly.

Figure 6 is an enlarged detail fragmentary section illustrating the parts at the juncture of the conductors and one of the intensifiers.

Figure 7 is an enlarged detail view partly in elevation and partly in section of one of the automatic cut-outs and

Figure 8 is a diagrammatic view of one of the simplest forms of embodiments of the invention.

Referring to the drawing by numerals, 1, 1 indicates magnets connected by a magnetic substance 2, preferably an iron wire. The magnets 1 are arranged in pairs, one pair being spaced beneath the other, and interposed between the magnets are zinc plates 3, 3 connected by an iron wire conductor 4. Suitable insulating supports 5 are arranged for sustaining the respective magnets 1 and plates 3, 3. Each plate 3 is preferably bent substantially into **V** form, as clearly seen in Figure 1, and the **V** of one of the plates opens or faces toward the north and the **V** of the other plate to the south. I have determined by experimentation that it is essential that the plates 3 be disposed substantially north and south with their flat faces approximately parallel to the adjacent faces of the co-operating magnets, although by experience I have not discovered any material difference in the current obtained when the plates are disposed slightly to one side of north and south, as for instance when disposed in the line of the magnetic polarity of the earth. The same is true with respect to the magnets 1, the said magnets being disposed substantially north and south for operative purposes, although I find that it is immaterial whether the north pole of one of the magnets is disposed to the north and the south pole to the south, or *vice versa*, and it is my conviction from experience that it is essential to have the magnets of each pair connected by magnetic material so that the magnets substantially become one with a pole exposed to the north and a pole exposed to the south. In Figure 1, I have indicated in full lines by the letters S and N the respective polarities of the magnets 1, and have indicated in dotted lines the other poles of those magnets when the connection 2 is severed. I have found that the magnets and zinc plates operate to produce, whether by collection or generation I am not certain, electrical currents when disposed substantially north and south, but when disposed substantially east and west no such currents are produced. I also find that the question of elevation is by no means vital, but it is true that more efficient results are obtained by placing the zincs and magnets on elevated supports. I furthermore find from tests that it is possible to obtain currents from the apparatus with the zincs and magnets disposed in a building or otherwise enclosed, although more efficient results are obtained by having the said elements arranged in the open.

While in Figures 1, 2 and 3, I have shown the magnets and the zinc plates as superimposed, it will be apparent, as hereinafter fully set forth, that these elements may be juxtaposed in horizontal planes, and substantially the same results will be secured. Furthermore, the magnets 1 with the interposed zincs 3, as shown in Figures 1, 2 and 3 merely represent a unit which may be repeated either horizontally or vertically for increasing the current supply, and when the unit is repeated the zinc plates are arranged alternating with the magnets throughout the entire series as hereinafter indicated.

A conductor 6 is connected in multiple with the conductors 2 and a conductor 7 is connected with conductor 4, the conductor 6 extending to one terminal of a rectifier which I have indicated by the general reference character 8, and the conductor 7 extending to the other terminal of said rectifier. The rectifier as seen in diagram in Figure 1 may assume any of several well known embodiments of the electrical valve type and may consist of four assymmetric cells or Cooper-Hewitt mercury vapor lamps connected as indicated in Figure 1 for permitting communication of the positive impulses from conductor 7 only to the line conductor 9 and the negative impulses from the conductor 7 only to the line conductor 10, and permitting the positive impulses from the conductor 6 only to the line conductor 9 and the negative impulses from conductor 6 only to the line conductor 10. The current from this rectifier may be delivered through the conductors 9 and 10 to any suitable source for consumption.

While the said rectifier 8 may consist of any of the known types, as above outlined, it preferably consists of a specially constructed rectifier which also has the capacity of intensifying the current and comprises specifically the elements shown in detail in Figures 4, 5 and 6 wherein I have disclosed the detail wiring of the rectifier when composed of four of the rectifying and intensifying elements instead of assymmetric cells or simple mercury vapor valves. As each of these structures is an exact embodiment of all the others, one only will be described, and the description will apply to all. The rectifying element of each construction consists of a mercury tube 11 which is preferably formed of copper but may be formed of glass or other suitable material, and comprises a cylinder having its end portions tapered and each terminating in an insulating plug or stopper 12. Through the upper stopper 12 is extended the electrode 13 which extends well into the tube and preferably substantially one-half the length thereof to a point adjacent the inner end of an opposing electrode 14 which latter electrode extends thence downwardly through the insulation 12 at the lower end of the tube. The tube 11 is supplied with mercury and is adapted to operate on the principle of the mercury vapor lamp, serving to rectify current by checking back impulses of one sign and permitting passage of impulses of the other. To avoid the necessity for utilizing a starter, as is common with the lamp type of electrical valve, the supply of mercury within the tube may be sufficient to contact with the lower end of the electrode 13 when current is not being supplied, so that as soon as current is passed from one electrode to the other sufficiently for volatilizing that portion of the mercury immediately adjacent the lower end of electrode 13, the structure begins its operation as a rectifier. The tube 11 is surrounded by a tube 15 which is preferably spaced from tube 11 sufficiently for allowing atmospheric or other cooling circulation to pass the tube 11. In some instances, it may be desirable to cool the tube 11 by a surrounding body of liquid, as hereinafter indicated. The tube 15 may be of insulating material but I find efficient results attained by the employment of a steel tube, and fixed to the ends of the tube are insulating disks 16, 16 forming a spool on which are wound twin wires 6^1 and 7^1, the wire 6^1 being connected at the inner helix of the coil with the outer end of the electrode 14, the lower portion of said electrode being extended to one side of the tube 11 and passed through an insulating sleeve 17 extending through the tube 15, and at its outer end merging into the adjacent end of the wire 6^1. The wire 7^1 extends directly from the outer portion of the spool through the several helices to a point adjacent the juncture of the electrode 14 with wire 6^1 and thence extends in mechanical parallelism with the wire throughout the coil; the wire 6^1 ending in a terminal 18 and the wire 7^1 ending in a terminal 19. For the sake of convenience of description and of tracing the circuits, each of the apparatus just above described and herein known as an intensifier and rectifier will be mentioned as A, B, C and D, respectively. Conductor 6 is formed with branches 20 and 21 and conductor 7 is formed with similar branches 22 and 23. Branch 20 from conductor 6 connects with conductor 7^1 of intensifier B and branch 21 of conductor 6 connects with the conductor 7^1 of intensifier C, while branch 22 of conductor 7 connects with conductor 7^1 of intensifier A and branch 23 of conductor 7 connects with conductor 7^1 of intensifier D. A conductor 27 is connected with terminal 19 of intensifier A and extends to and is connected with terminal 18 of intensifier C, and a conductor 28 is connected with the terminal 19 of intensifier C and extends to and is connected with electrode 13 of intensifier A. A conductor 29 extends from the terminal 19 of intensifier B to the terminal 18 of intensifier D and a conductor 30 extends from the terminal 19 of intensifier D to electrode 13 of intensifier B. Each electrode 13 is supported on a spider 13^1 resting on the upper disk 16 of the respective intensifier. Conductors 31 and 32 are connected with the terminals 18 of intensifiers A and B and are united to form the positive line wire 9 which co-operates with the negative line wire 10 and extends to any suitable point of consumption. The line wire 10 is provided with branches 35 and 36 extending to the electrodes 13 of intensifiers C and D for completing the negative side of the circuit.

Thus it will be seen that alternating currents produced in the wires 6 and 7 will be rectified and delivered in the form of a direct current through the line wires 9 and 10, and I find by experiment that the wires 6 and 7 should be of iron, preferably soft, and may of course be insulated, the other wiring not specified as iron being of copper or other suitable material.

In carrying out the operation as stated, the circuits may be traced as follows: A positive impulse starting at the zincs 3 is directed along conductor 7 to branch 23 to conductor 7^1 and the winding of intensifier D to terminal 19 through conductor 30 to electrode 13 of the rectifier of intensifier B through said rectifier to the conductor 6^1, through the winding thereof to the contact 18, conductor 32 and to the line wire 9. The next or negative impulse directed along conductor 7 cannot find its way along branch 23 and the circuit just above traced because it cannot pass across the rectifier of intensifier B but instead the negative impulse passes along conductor 22 to conductor 7 of intensifier A and the winding thereof to the contact 19 to conductor 27 to contact 18 of intensifier C, to the winding of the wire 6^1 thereof to the electrode 14 through the rectifier to the electrode 13 and conductor 36 to the line wire 10. A positive impulse delivered to wire 6 passes along the said wire to the branch 21 to the conductor 7^1 of intensifier C and the winding thereof to the contact 19, conductor 28, electrode 13 of the rectifier of intensifier A, electrode 14 thereof and conductor 6^1 to contact 18 and wire 31 to line wire 9. Obviously the positive impulse cannot pass along the wire 20 because of its inverse approach to the rectifier of intensifier B. The next impulse or negative impulse delivered to conductor 6 cannot pass along conductor 21 because of its connection with electrode 13 of the rectifier of intensifier A, but instead passes along conductor 20 to the wire 7^1 and its winding forming part of intensifier B to the contact 19 and conductor 29 to contact 18 and the winding of wire 6^1 of intensifier D to the electrode 14 and through the rectifier to the electrode 13 and conductor 35 to line wire 10. Thus the current is rectified and all positive impulses directed along one line and all negative impulses along the other line so that the potential difference between the two lines will be maximum for the given current of the alternating circuit. It is, of course, apparent that a less number of intensifiers with their accompanying rectifier elements may be employed with a sacrifice of the impulses which are checked back from a lack of ability to pass the respective rectifier elements, and in fact I have secured efficient results by the use of a single intensifier with its rectifier elements, as hereinafter set forth.

Grounding conductors 37 and 38 are connected respectively with the conductors 6 and 7 and are provided with the ordinary lightning arresters 39 and 40 respectively for protecting the circuit against high tension static charges.

Conductors 41 and 42 are connected respectively with the conductors 6 and 7 and each connects with an automatic cut-out 43 which is grounded as at 44. Each of said automatic cut-outs is exactly like the other and one of the same is shown in detail in Figure 7 and comprises the inductive resistance 45 provided with an insulated binding post 46 with which the respective conductor 6 or 7 is connected, said post also supporting a spring 48 which sustains an armature 49 adjacent the core of the resistance 45. The helix of resistance 45 is connected preferably through the spring to the binding post at one end and at the other end is grounded on the core of the resistance, the said core being grounded by ground conductor 44 which extends to the metallic plate 52 imbedded in moist carbon or other inductive material 53 buried in the earth. Each of the conductors 41, 42 and 44 is of iron, and in this connection I wish it understood that where I state the specific substance I am able to verify the accuracy of the statement by the results of tests which I have made, but of course I wish to include along with such substance all equivalents, as for instance, where iron is mentioned its by-products, such as steel, and its equivalents such as nickel and other magnetic substances are intended to be comprehended. The cut-out apparatus seen in detail in Figure 7 is employed particularly for insuring against high tension currents, it being obvious from the structure shown that when potential rises beyond the limit established by the tension of the spring sustaining the armature 49, the armature will be moved to a position contacting with the core of the cut-out device and thereby directly close the ground connection for line wire 41 with conductor 44, eliminating the resistance of winding 45 and allowing the high tension current to be discharged to the ground. Immediately upon such discharge the winding 45 losing its current will allow the core to become demagnetized and release the armature 49 whereby the ground connection is substantially broken leaving only the connection through the winding 45 the resistance of which is sufficient for insuring against loss of low tension current.

In Figure 8 I have illustrated an apparatus which though apparently primitive in construction and arrangement comprehends the first successful embodiment which I produced in the course of the discovery of the present invention, and it will be observed that the essential features of the invention are therein disclosed. The structure delineated in said figure consists of horse shoe magnets 54, 54, one facing north and the other south, that is, each opening in the respective directions indicated and the two being connected by an iron wire 55 which is uninsulated and wrapped about the respective magnets at or adjacent the neutral zone thereof, and the wire 55 is preferably soldered to the respective magnets each end portion of the wire 55 being extended from the respective magnet to and connected with, as by being soldered to, a zinc plate 56, there being a plate 56 for each magnet and each plate being arranged longitudinally substantially parallel with the legs of the magnet and with the faces of the plate exposed toward the respective legs of the magnet, the plate being thus arranged endwise toward the north and south. An iron wire 57 connects the plates 56, the ends of the wire being preferably connected adjacent the outer ends of the plates but from experiment I find that the wire may be connected at practically any point to the plate. Lead wires 58 and 59 are connected respectively with the wires 55 and 57 and supply an alternating current ordinarily at a comparatively low tension, and to control such current the wires 58 and 59 may be extended to a rectifier or combined rectifier and intensifier, as above set forth.

The tests which I have found successful with the apparatus seen in Figure 8 were carried out by the employment first of horse shoe magnets approximately four inches in length, the bar comprising the horse shoe being about one inch square, the zincs being dimensioned proportionately and from this apparatus with the employment of a single intensifier and rectifier, as above stated, I was able to obtain a constant current of 8 volts.

It should be obvious that the magnets forming one of the electrodes of this apparatus may be permanent or may be electro-magnets, or a combination of the two.

While the magnets mentioned throughout the above may be formed of any magnetic substance, I find the best results obtained by the employment of the nickel chromo steel.

While the successful operation of the various devices which I have constructed embodying the present invention have not enabled me to arrive definitely and positively at fixed conclusions relative to the principles and theories of operation and the source from which current is supplied, I wish it to be understood that I consider myself as the first inventor of an apparatus of the general type hereinbefore described capable of producing commercially serviceable electricity, for which reason my claims hereinafter appended contemplate that I may utilize a wide range of equivalents so far as concerns details of construction suggested as preferably employed.

The current which I am able to obtain is dynamic in the sense that it is not static and its production is accomplished without chemical or mechanical action either incident to the actual chemical or mechanical motion or incident to changing calorific conditions so that the elimination of necessity for the use of chemical or mechanical action is to be considered as including the elimination of the necessity for the use of heat or varying degrees thereof.

Having now particularly described and ascertained the nature of my said invention, and in what manner the same is to be performed, I declare that what I claim is:—

1. Means for producing dynamic electricity without mechanical or chemical action.
2. Electricity producing means as claimed in Claim 1, characterized in that said means is capable of producing low tension current.
3. Electricity producing means as claimed in Claim 1, characterized in that said means comprises a magnet, and means co-operating with the latter.
4. Means as set forth in Claim 3, characterized in that the parts are stationary.
5. Means as claimed in Claim 3, characterized in that the magnet is disposed substantially north and south.
6. Means as claimed in Claims 1, 3 and 5, wherein the means coacting with the magnet is a metal and said parts are disposed substantially north and south.
7. Means as claimed in Claims 1, 3 and 6, in which the metal is zinc, and combined with conductors connected with the co-operating parts to deliver current therefrom.
8. Means as set forth in Claim 1, characterized in that said means comprises spaced magnets with an adjacent co-operating zinc, and an iron wire connecting the magnets, and means to deliver electrical current from said parts.
9. Means as set forth in Claim 8, characterized in that the current delivery means comprises an intensifier and rectifier.
10. Means as set forth in Claim 8, characterized in that the current delivery means comprises wires connected with the magnets and zinc, an electrical valve connected with one of said wires, and independent helices surrounding said valve, one of said helices being connected with the terminal of the valve at the opposite side from the contact of said wire and the other of said helices being connected with the other wire, the electrical valve comprising a mercury vapor rectifier.
11. The process of producing electricity comprising exposing a magnet and

co-operating means disposed substantially north and south and taking off current therefrom.

12. The process of producing electricity comprising disposing stationary elements in co-operative relation in respect to each other and with respect to the earth for giving off relatively low tension electrical current, substantially without chemical action.

13. Electricity producing means constructed and operating substantially as described with reference to the accompanying drawings.

Dated this 13th day of January, 1913.

W. P. THOMPSON & Co.,
6, Lord Street, Liverpool, and at
Bradford & London,
Agents for the Applicant.

Reference has been directed, in pursuance of Section 7, Sub-section 4, of the Patents and Designs Act, 1907, to Specifications No. 16,709 of 1887, No. 14,033 of 1899, No. 15,412 of 1906, and No. 5457 of 1911.

Redhill: Printed for His Majesty's Stationery Office, by Love & Malcomson, Ltd.—1914.

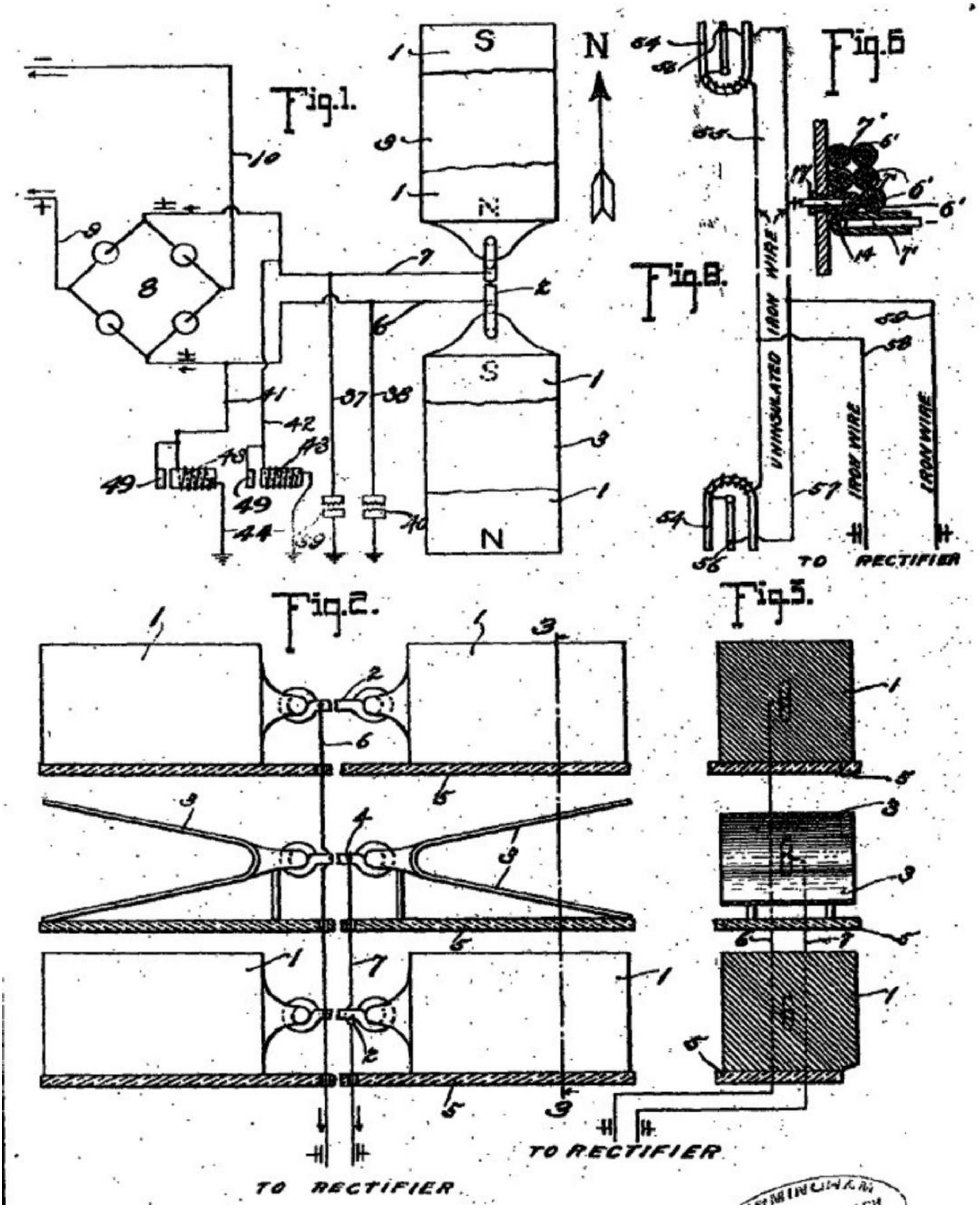
Fig.1.
Fig.6
Fig.8.
Fig.2.
Fig.5.
S
N
N
S
N
t
UNINSULATED
IRON WIRE
IRON WIRE
IRON WIRE
TO RECTIFIER
t
TO RECTIFIER
TO RECTIFIER

Fig7.

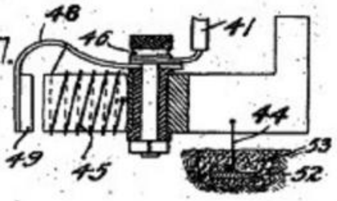
48
46
41
44
53
52
49
45

Fig4

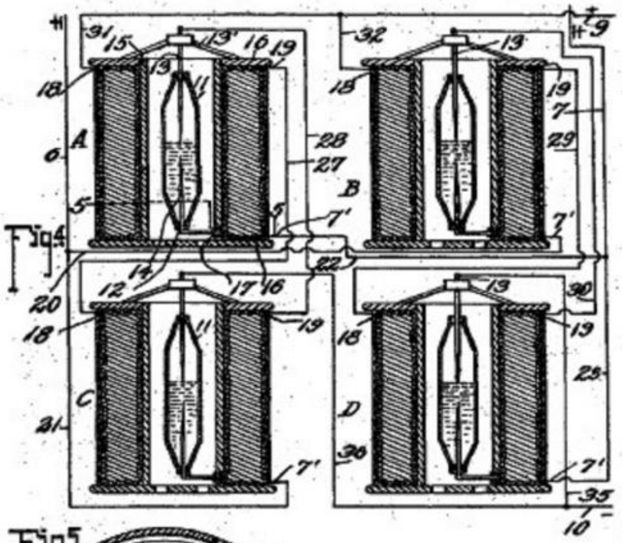
+
31
15
13'
16
19
32
13
+
9
18
11
13
18
19
7
A
28
27
29
6
B
5
5
7'
7'
22
12
14
17
16
13
30
20
18
11
19
18
19
25
C
21
D
36
7'
7'
35
10

Fig5.

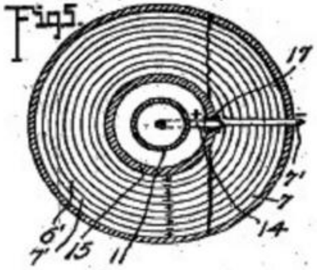
17
7'
7
14
6'
7
15
11

Your Support Maintains this Service. Please Invest in the mind-boggling Rex Research Website CD -- Always The Best Investment ... Everything on this Website & More (the Bonus Files CD) -- Only $13, Postpaid Anywhere ! Secure Transaction via PayPal: All Major Credit Cards Accepted:

Buy Now

No Credit ? Don't like PayPal ? Send a Check/Money Order to:

Rex Research, PO Box 19250, Jean, NV 89019 USA

---

rexresearch.com

Roy J. Meyers Absorber, GB Patent 191301098.

This device was brought to my attention by Grumage. Here is my take on why it worked.

In its simplest form the device consists of two horseshoe magnets, some zinc plates and some uninsulated iron wire. The iron wire is wound around the curved ends of the horseshoe magnet and connected to the zinc plates as shown in figure 1. The system has to be aligned with the earth's magnetic field.

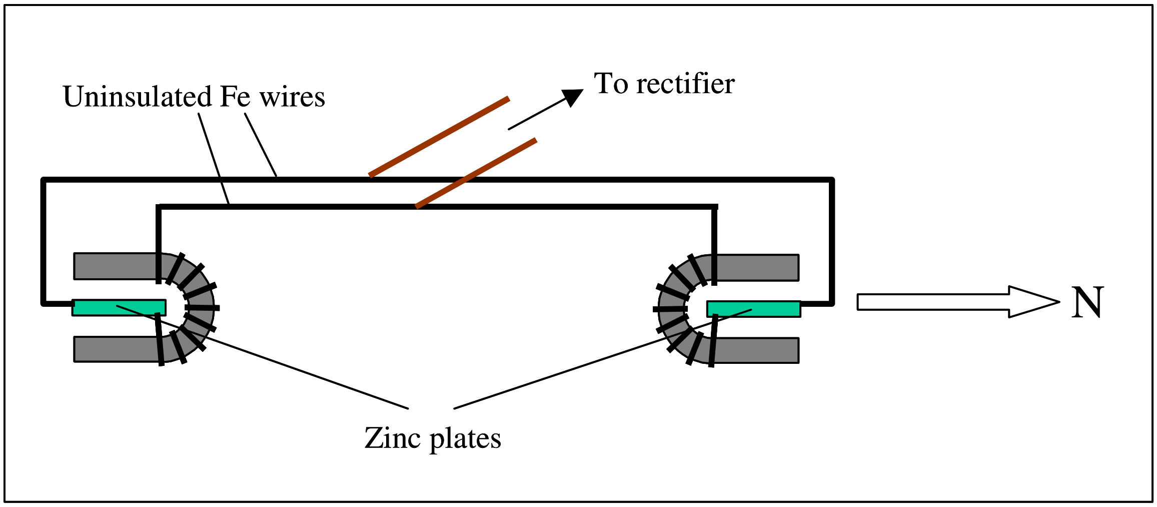

Figure 1. Simple Scheme reported to supply alternating current at low voltage

At first sight this seems incapable of providing alternating voltage but here is a possible explanation. If you follow the magnetization axis of each horseshoe it is seen that the earth's field supplies a magnetic gradient around the curved section, hence spin polarized conduction electrons therein will be dragged along thus creating a tiny potential difference between the two legs of each magnet. The iron wires wound over that curved end do nothing except create contact points at the ends of each curvature. We can discount any galvanic potential between iron wire and magnet since the magnets used were of magnetized steel. However we cannot ignore the galvanic potential between the zinc and the iron connection, or indeed the solder used for the connection. Although the two connections to the zinc would normally result in a zero overall potential, there exists a difference in the magnetic field at each end of the zinc, which could result in a significant (but low) voltage present. This magneto-Seebeck effect is a fairly new phenomenon and would not have been known in 1913 when the patent application was made. Also unknown at that time was the presence of spin-polarized conduction electrons.

This all predicts the presence of DC voltage, so how could this become AC? Well the magneto-Seebeck effect can result in a positive or a negative potential, and it also changes with temperature. In fact there can be a crossover temperature where it changes from positive to negative. Hence the DC current flowing around the closed loop of iron wire could heat the thin zinc plates to take it through that crossover temperature, whereupon the current decreases, the zinc cools down then the whole process repeats itself. This would result in a very low frequency AC voltage being observed.

US 20140152016A1

(19) **United States**
(12) **Patent Application Publication**
JENNINGS
(10) **Pub. No.: US 2014/0152016 A1**
(43) **Pub. Date: Jun. 5, 2014**

(54) **ATMOSPHERIC TRANSDUCTION SYSTEM**

(71) Applicant: **JAMES EDWARD JENNINGS**, SUPERIOR, CO (US)

(72) Inventor: **JAMES EDWARD JENNINGS**, SUPERIOR, CO (US)

(21) Appl. No.: **13/692,121**

(22) Filed: **Dec. 3, 2012**

Publication Classification

(51) **Int. Cl.**
H02N 11/00 (2006.01)
F03D 9/00 (2006.01)
(52) **U.S. Cl.**
CPC ***H02N 11/002*** (2013.01); ***F03D 9/003*** (2013.01)
USPC **290/55**; 310/308

(57) **ABSTRACT**

Atmospheric Transduction System including a Power Frequency broadcast station, a receiver, and a network. The Power Frequency broadcast station includes a transmitter and a computer server. The receiver is in communication with the Power Frequency broadcast transmitter and also includes a user interface for receiving user input commands comprising a request for information from the Power Frequency broadcast station. The receiver is configured to establish a two-way communication path between the receiver and the Power Frequency broadcast transmitter. The network is in communication with the transducer, controller and the receiver for exchanging information therebetween. In response to oscillation translation and/or rotation of the electronic transducer, portions of forces induced by the mass are transferred to the piezoelectric elements. Electrical energy output by these piezoelectric elements is received in a power controller and can be applied to the battery as self charging. The piezoelectric transducer includes a conductive rotor and bearings, at least one of them incorporating a vibrator of mechanical oscillation, having a piezoelectric transducer converting mechanical vibrations into electric power.

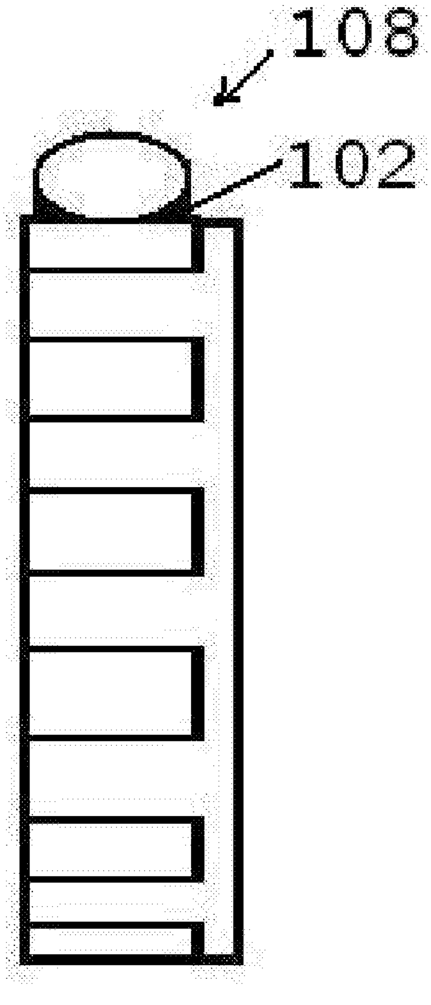

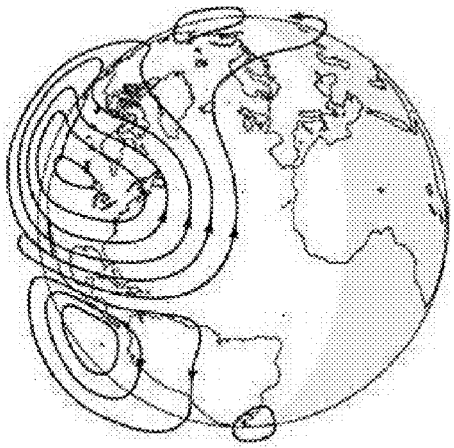

FIG. 1

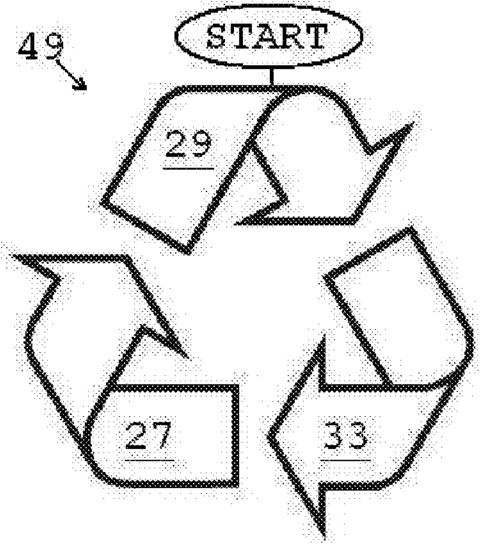
49
START
29
27
33

FIG. 2

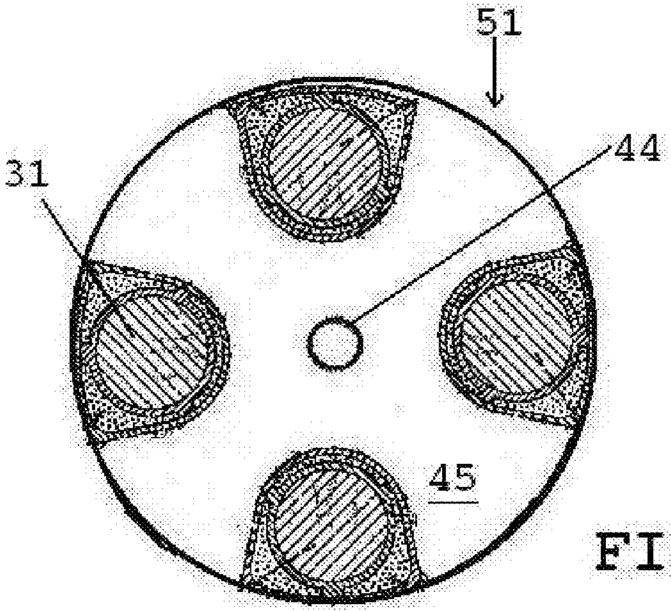

FIG. 3

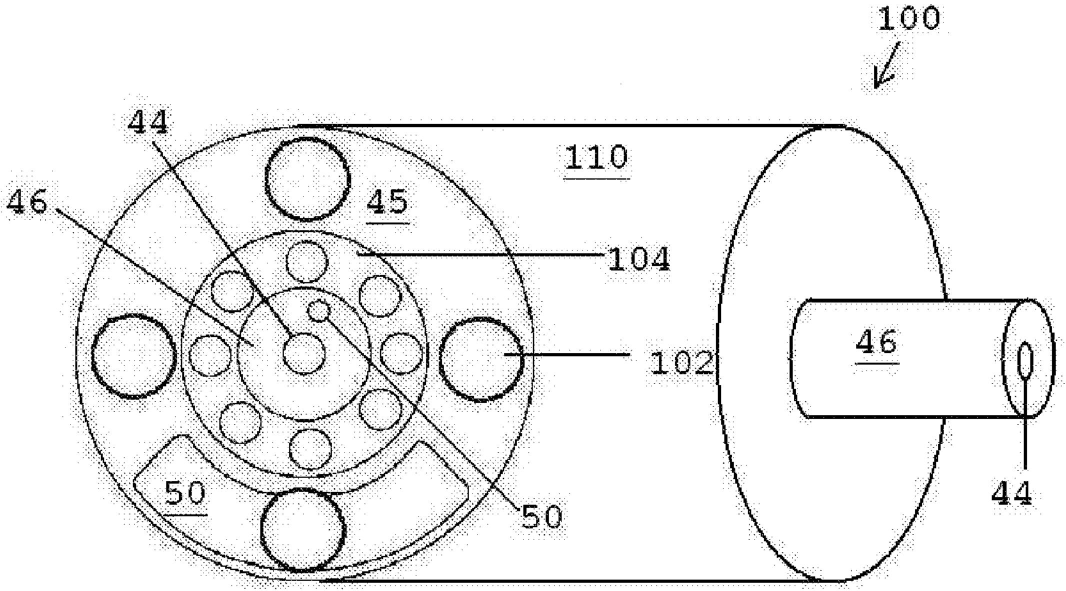
100
44
110
46
45
104
46
102
50
50
44

FIG.4

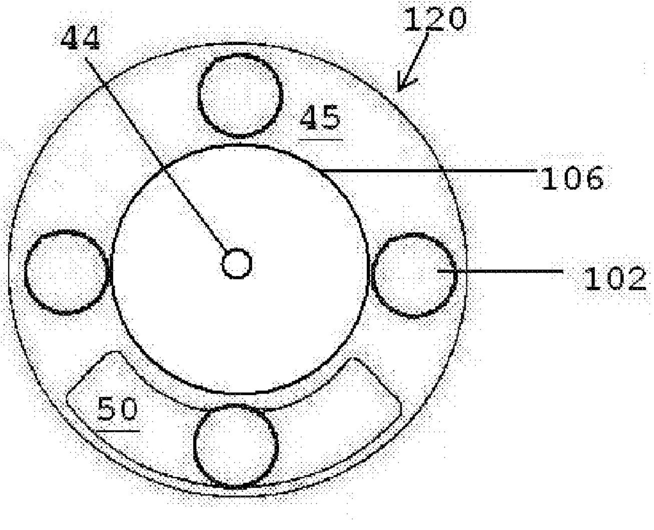
120
44
45
106
102
50

FIG.5

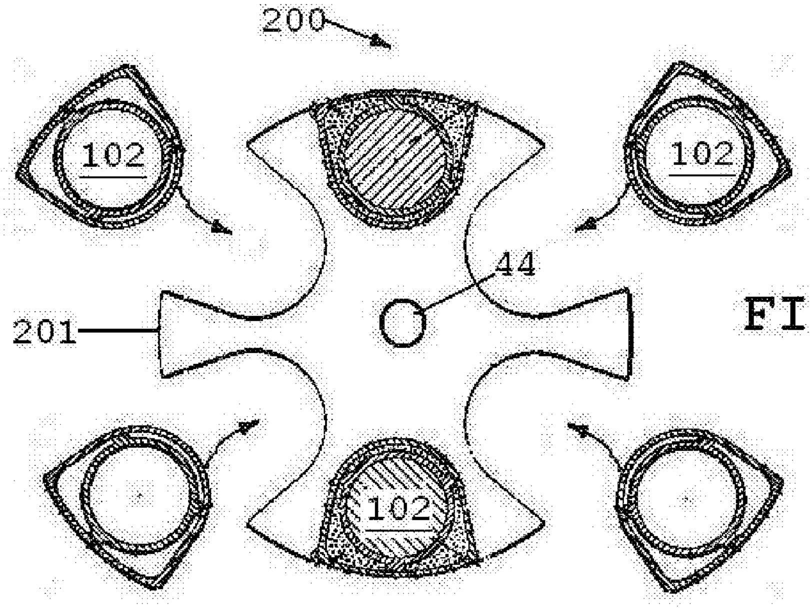
200
102
102
44
201
102

FIG. 6

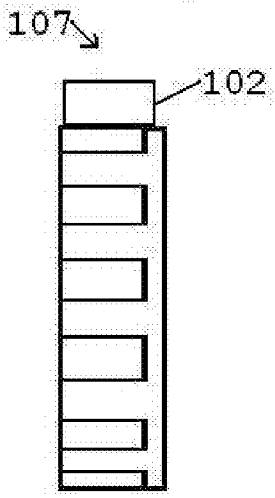

FIG. 7

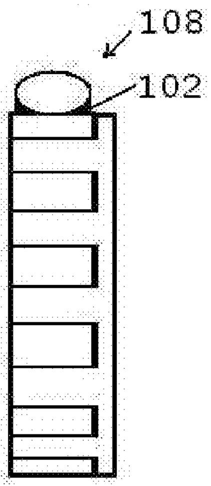

FIG. 8

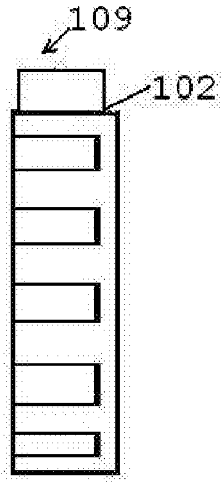

FIG. 9

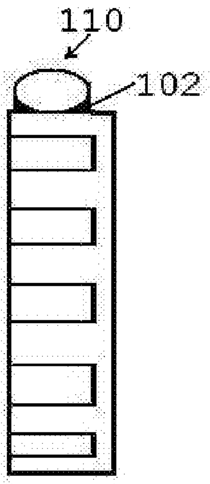

FIG. 10

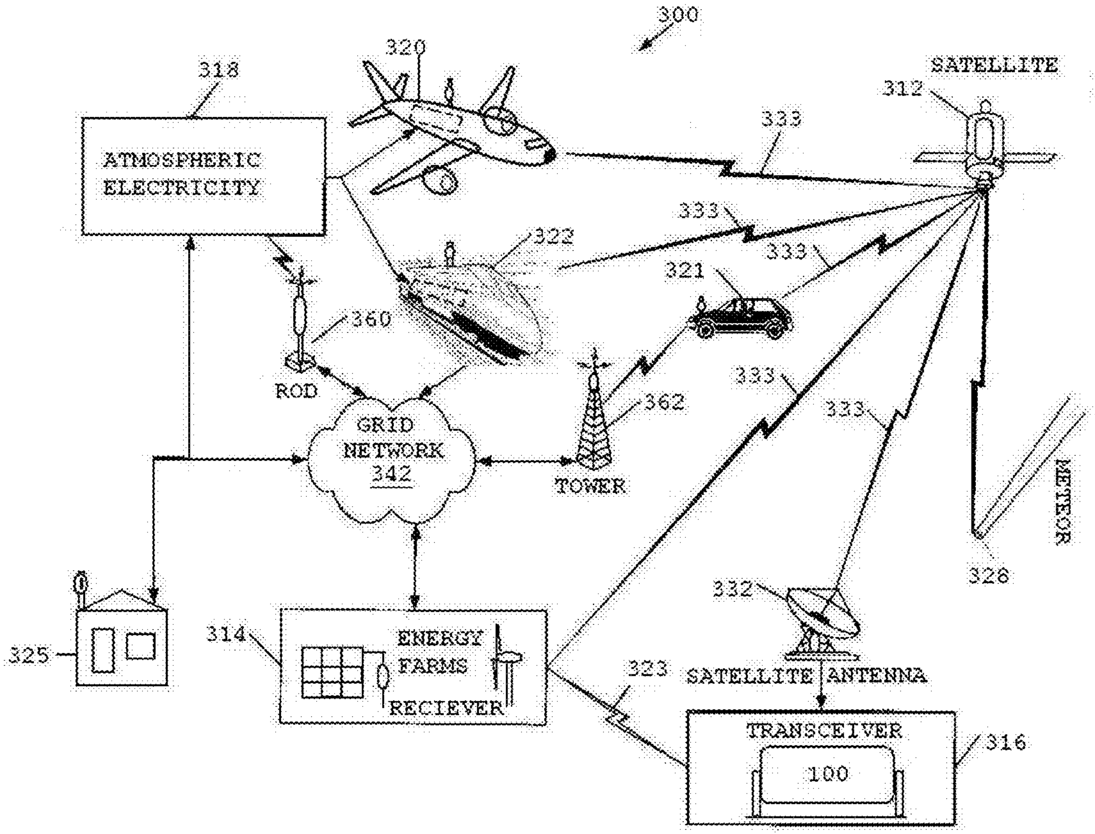
ATMOSPHERIC ELECTRICITY
SATELLITE
ROD
GRID NETWORK 342
TOWER
METEOR
ENERGY FARMS
RECIEVER
SATELLITE ANTENNA
TRANSCEIVER

FIG. 11

FIG.12

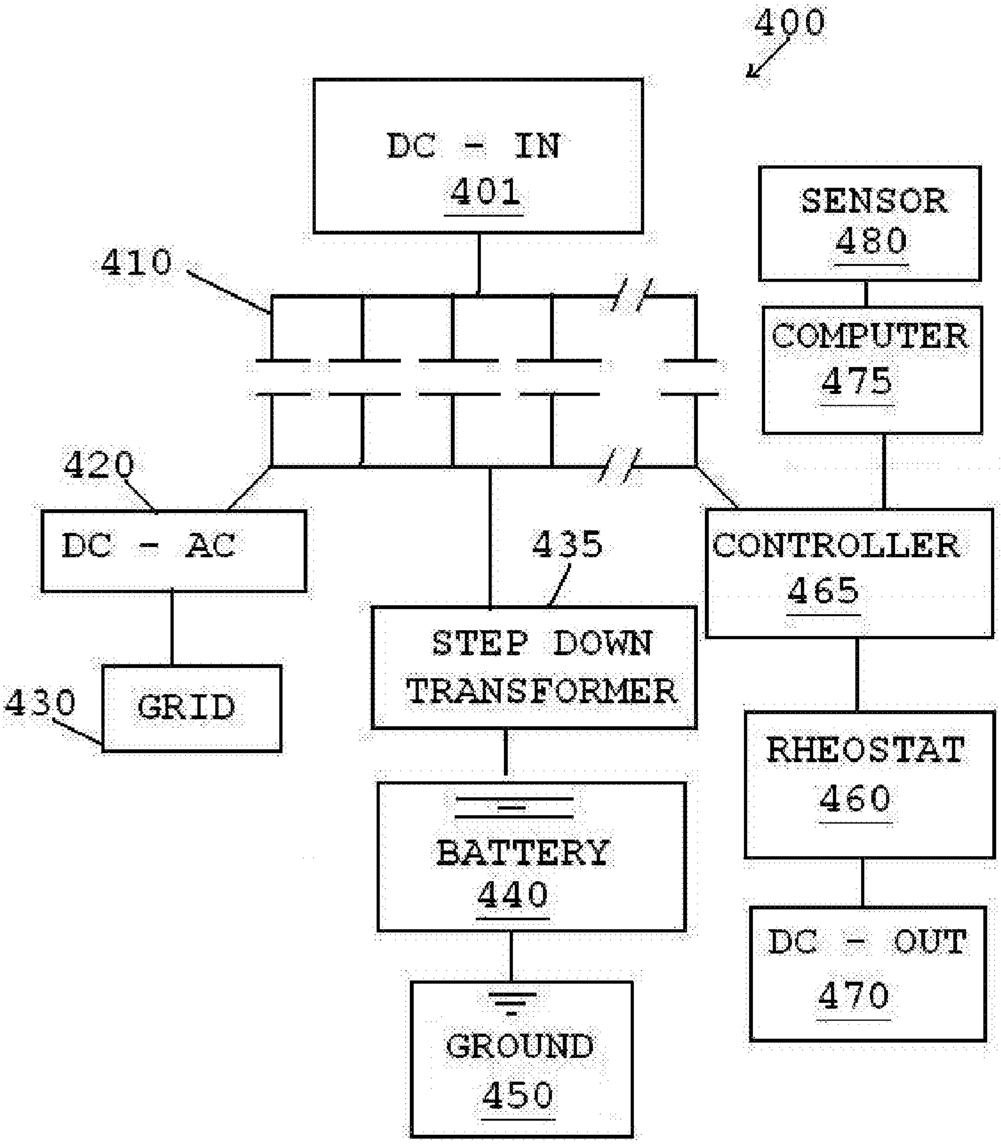
400
DC - IN
401
410
SENSOR
480
COMPUTER
475
420
DC - AC
435
CONTROLLER
465
STEP DOWN
TRANSFORMER
430
GRID
RHEOSTAT
460
BATTERY
440
DC - OUT
470
GROUND
450

FIG.13

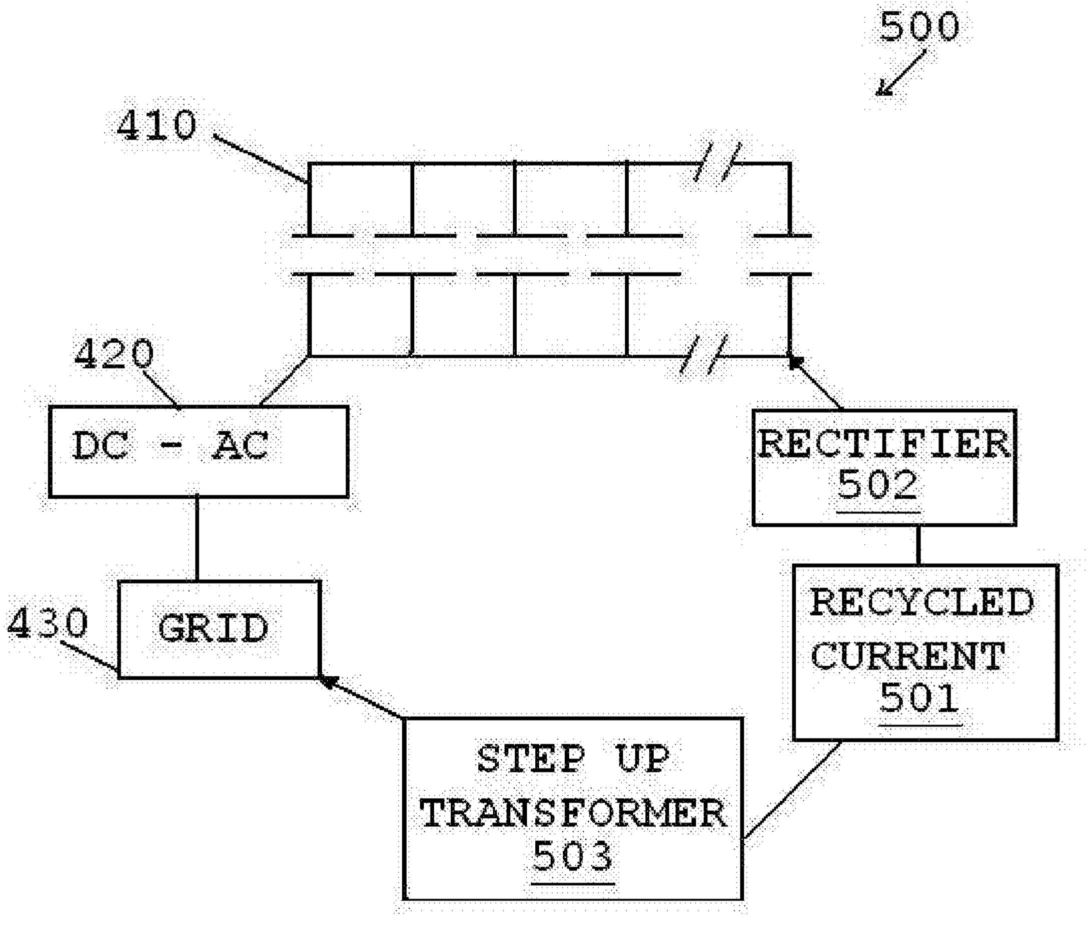
500
410
420
DC - AC
RECTIFIER
502
430
GRID
RECYCLED
CURRENT
501
STEP UP
TRANSFORMER
503

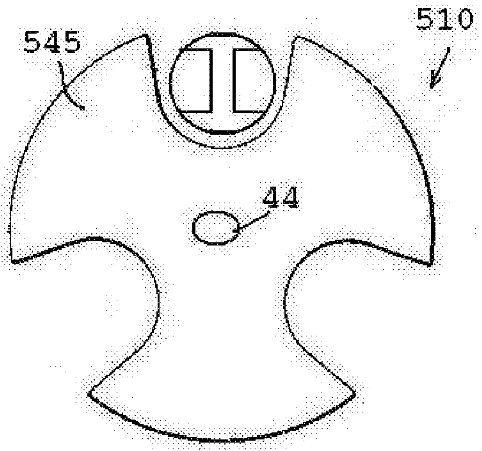

FIG.14

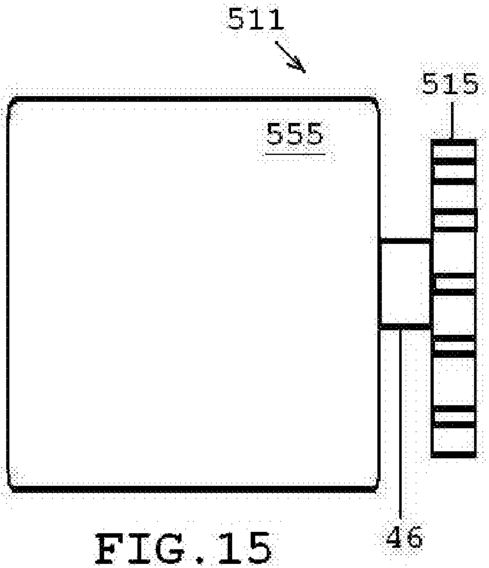
511
515
555
46

FIG.15

FIG.16

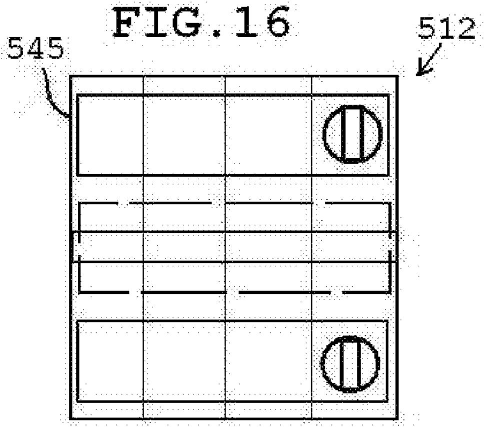

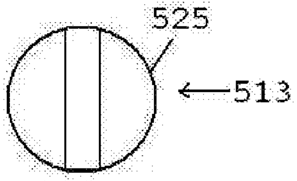

FIG.17

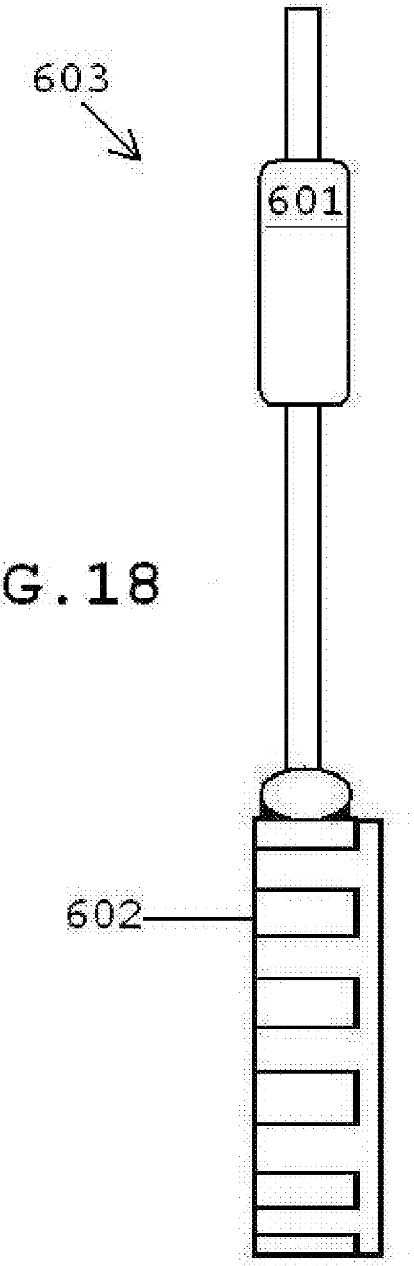

FIG.18

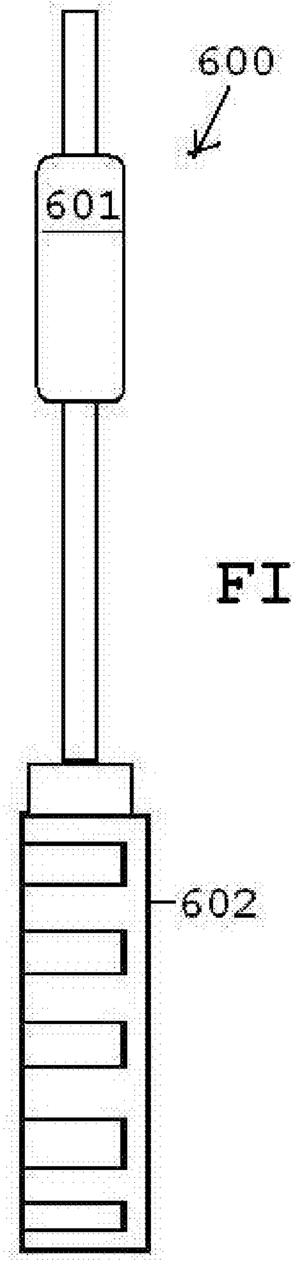

FIG.19

FIG.20

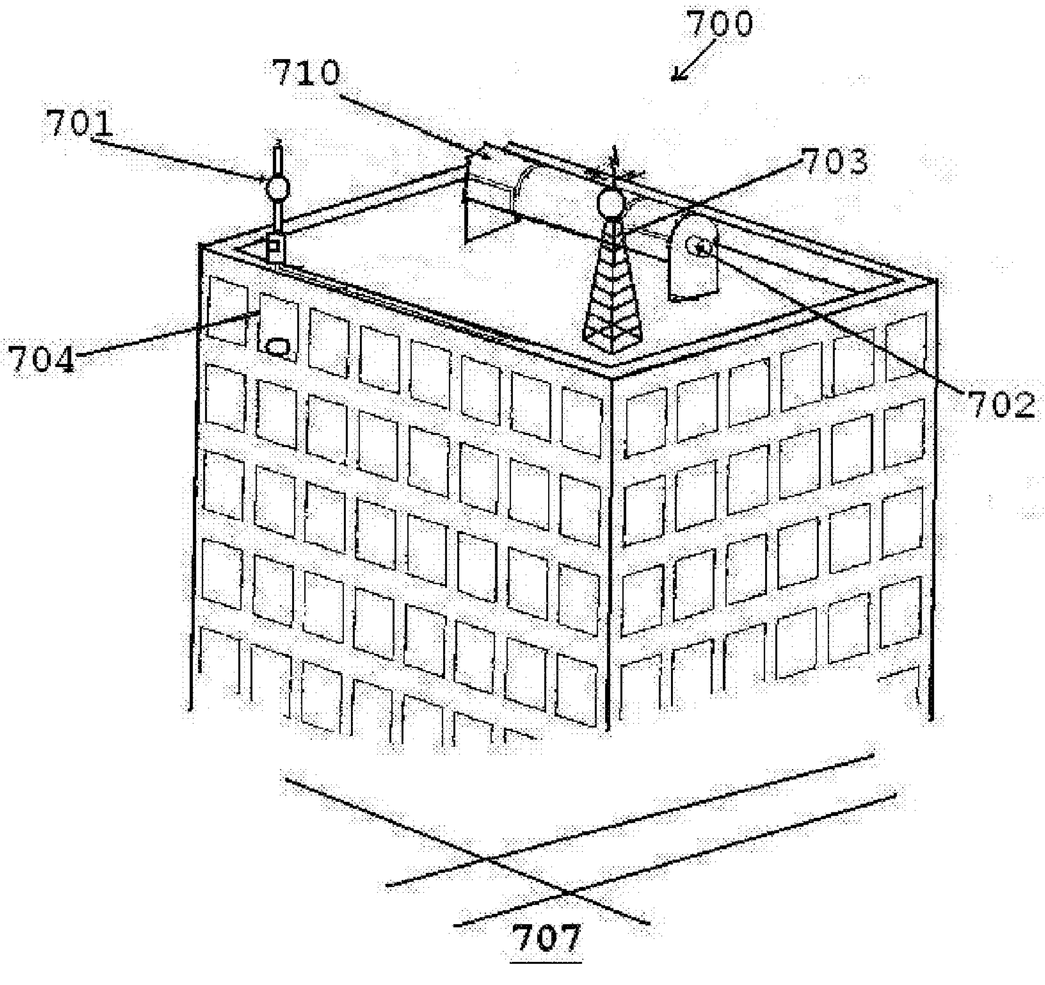
700
710
701
703
704
702
707

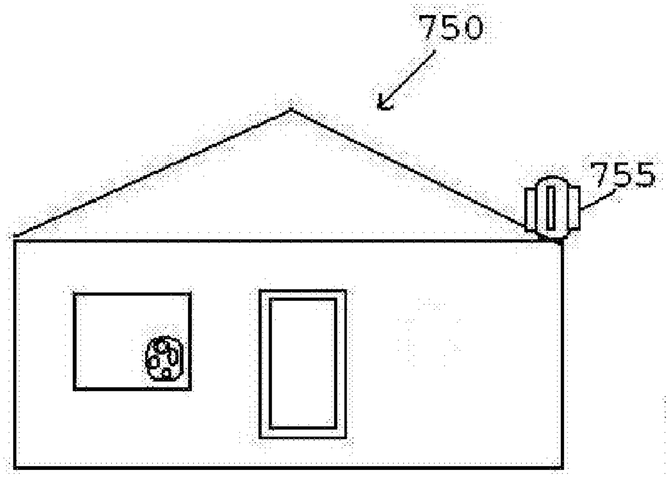

FIG.21

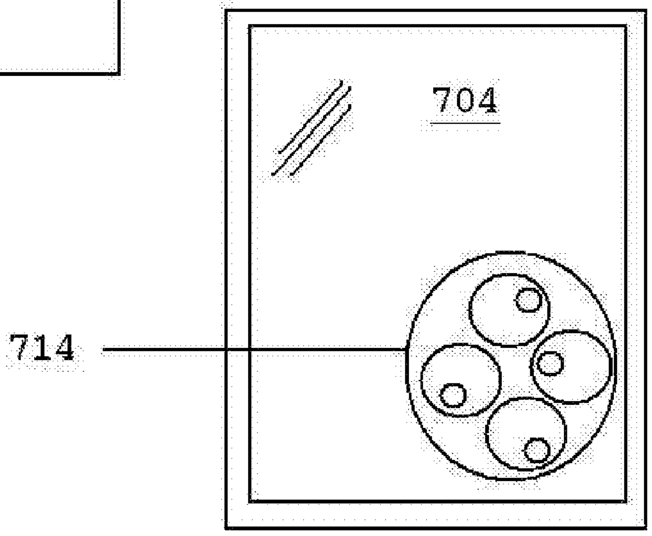

FIG.22

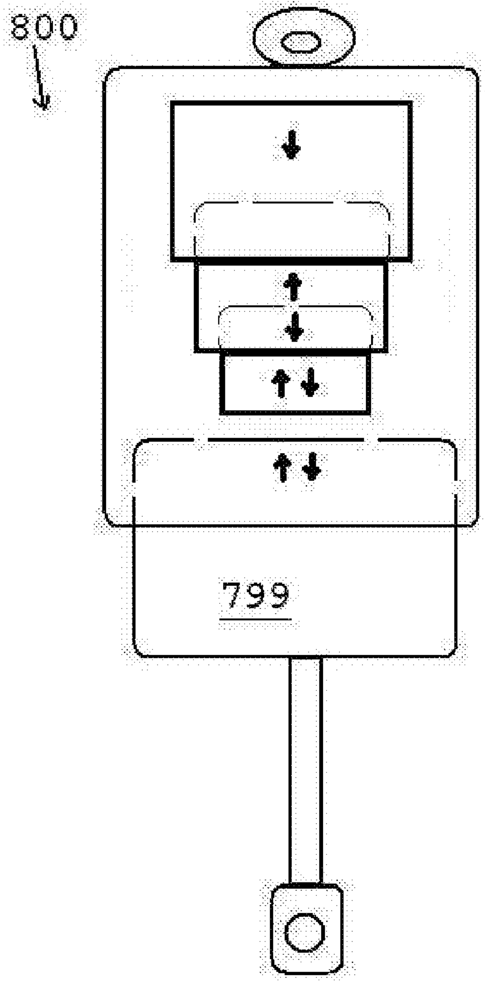

FIG.23

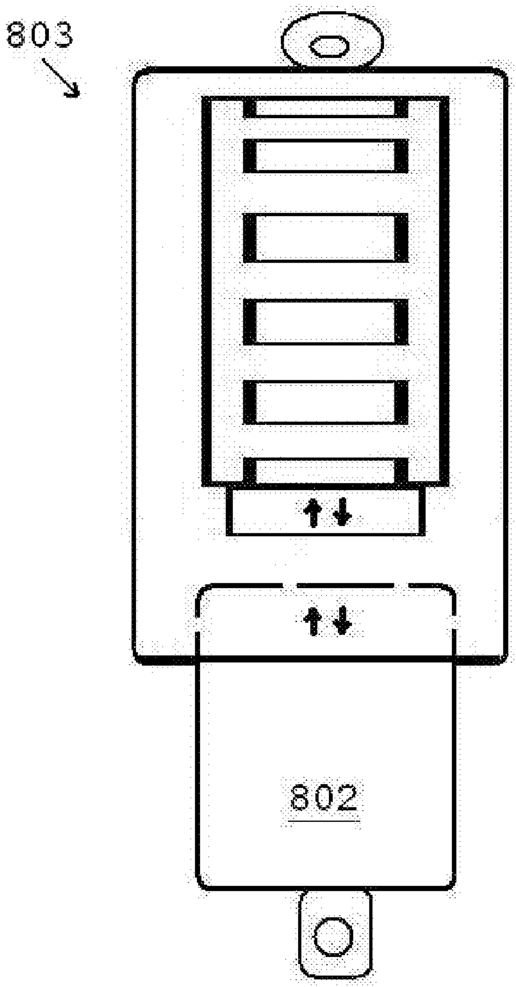

FIG.24

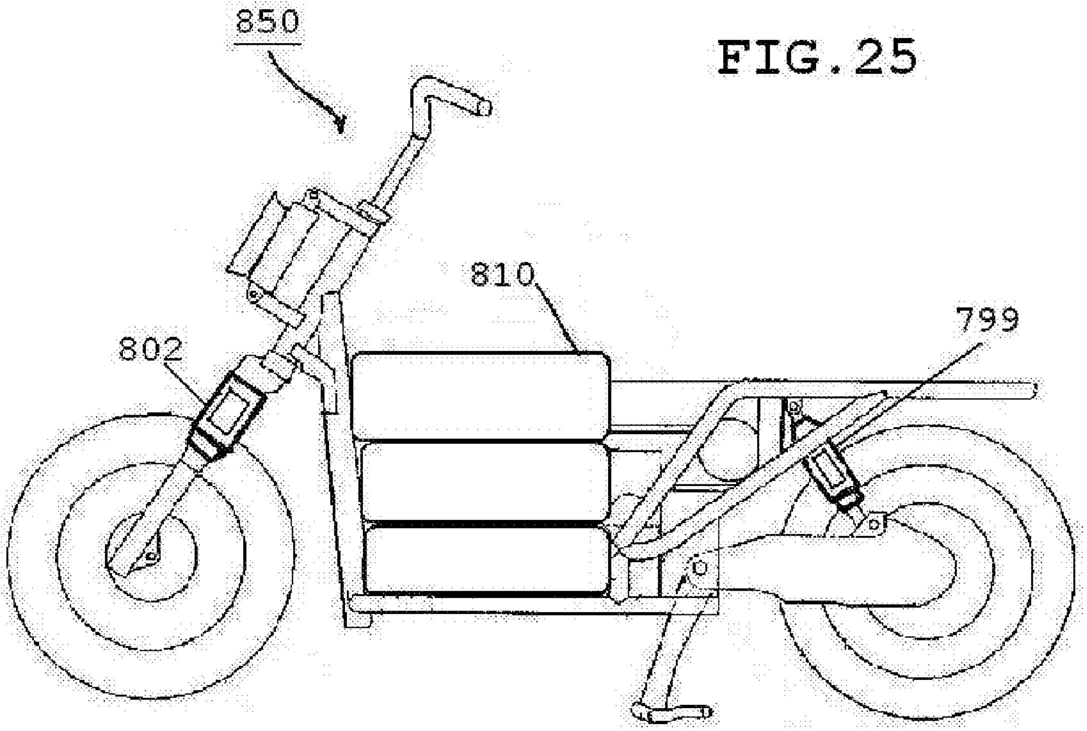
850
810
799
802

FIG. 25

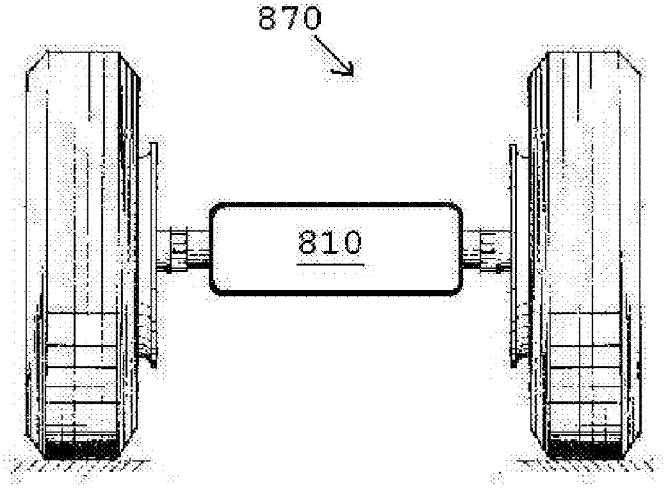

FIG. 26

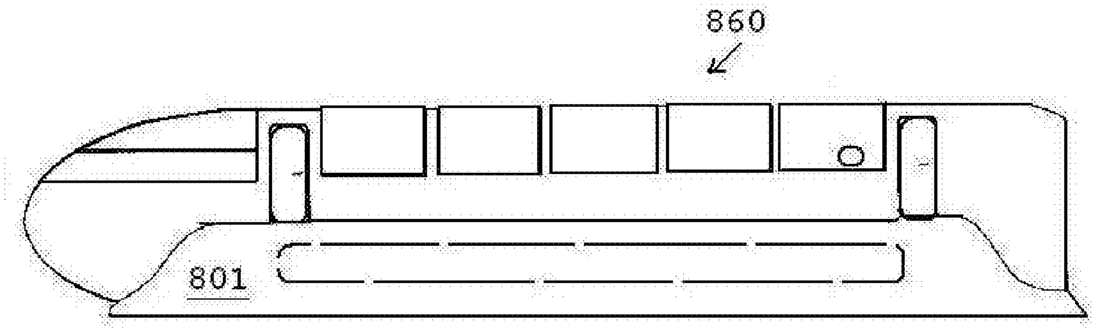

FIG. 27

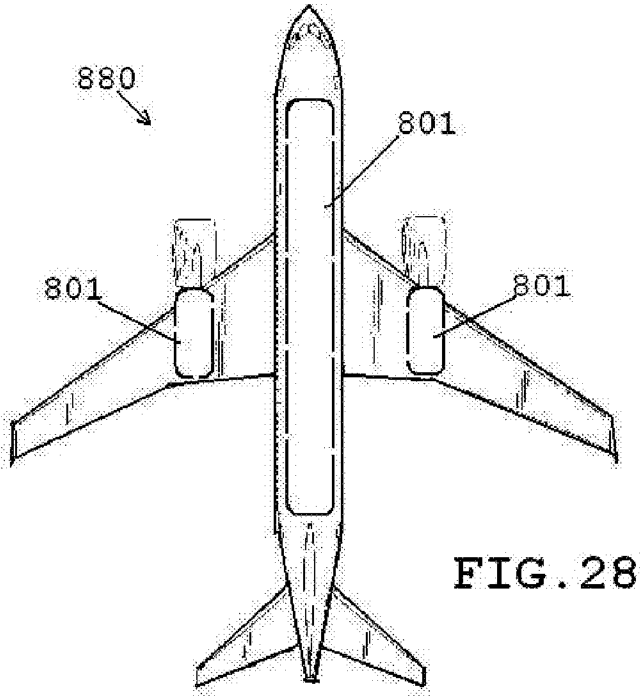
880
801
801
801

FIG. 28

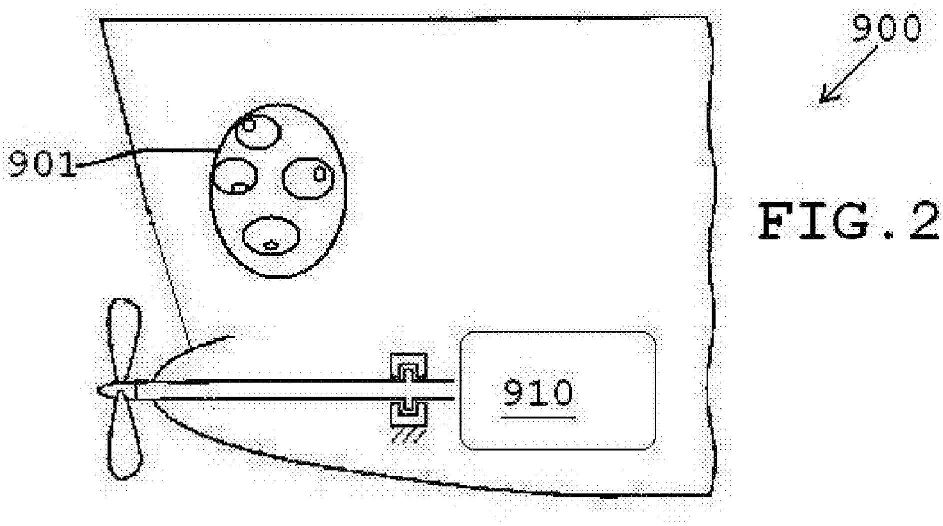

FIG.29

FIG.30

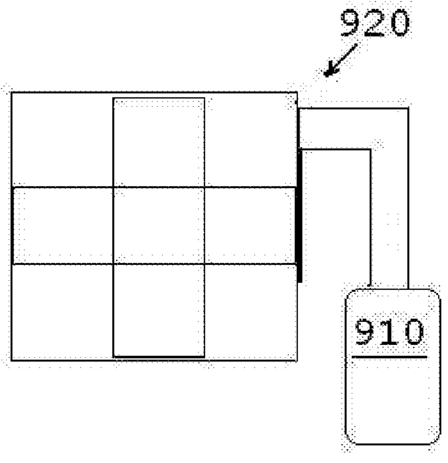

FIG.31

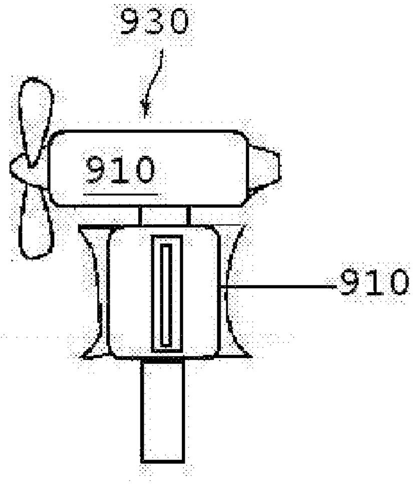

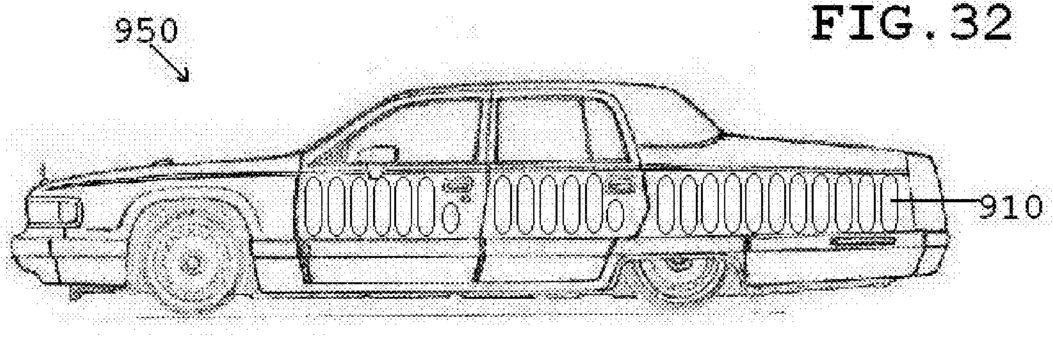

FIG. 32

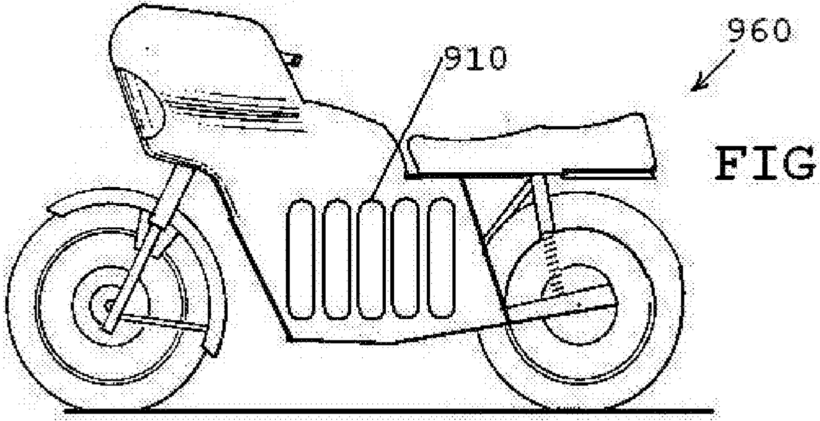

FIG. 33

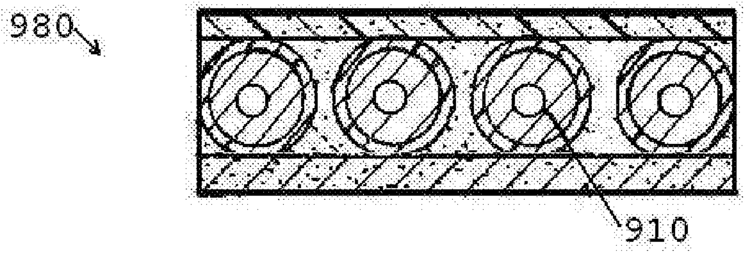

FIG. 34

ATMOSPHERIC TRANSDUCTION SYSTEM

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] NONE

BACKGROUND OF THE INVENTION

Field of the Invention

[0002] Self-propelled travel is a type of recreational adventure travel using human powered transport. This includes non-motorized machines such as a bicycle or skateboard. It is in contrast to traveling in a powered vehicle (such as an automobile) as in that case it is the vehicle which powers itself. Self-propelled travel is used to travel short distances or even for much longer distances such as bicycle touring. Self propelled describes something that moves, progresses or acts on its own power without needing outside help. Leonardo da Vinci's 1478 Self-Propelled Car: It was more than 500 years ago, however—sometime around the year 1478 to be more or less specific—when Leonardo drew out his plans for the world's first self-propelled vehicle. Experts originally believed two leaf springs, the simplest form of the spring typically used for automotive suspensions, somehow powered the vehicle. Closer inspection eventually revealed the power came from bigger, coiled springs located in tambours, cylindrical drum-like casings, inside the car's frame. The machine works like a robot or a wind-up toy simply by rotating the wheels opposite of their intended direction, which winds up the springs inside and gives it power. Self propulsion (of a vehicle) provided with its own source of tractive power rather than requiring an external means of propulsion.

[0003] In the middle of the 18th century, Benjamin Franklin's experiments showed that electrical phenomena of the atmosphere were not fundamentally different from those produced in the laboratory. By 1749, Franklin observed lightning to possess almost all the properties observable in electrical machines.

[0004] In July 1750, Franklin hypothesized that electricity could be taken from clouds via a tall metal aerial with a sharp point. Before Franklin could carry out his experiment, in 1752 Thomas-Francois Dalibard erected a 40-foot (12 m) iron rod at Marly-la-Ville, near Paris, drawing sparks from a passing cloud. With ground-insulated aerials, an experimenter could bring a grounded lead with an insulated wax handle close to the aerial, and observe a spark discharge from the aerial to the grounding wire. In May 1752, Dalibard affirmed that Franklin's theory was correct.

Piezoelectric Motor

[0005] A piezoelectric motor or piezo motor is a type of electric motor based upon the change in shape of a piezoelectric material when an electric field is applied. Piezoelectric motors make use of the converse piezoelectric effect whereby the material produces acoustic or ultrasonic vibrations in order to produce a linear or rotary motion. In one mechanism, the elongation in a single plane is used to make a series stretches and position holds, similar to the way a caterpillar moves. A transducer is a device that converts one form of energy to another. Energy types include (but are not limited to) electrical, mechanical, electromagnetic (including light), chemical, acoustic or thermal energy. While the term transducer commonly implies the use of a sensor/detector, any device which converts energy can be considered a transducer. Transducers are widely used in measuring instruments. Piezoelectric materials can also be used to harvest low levels of mechanical energy into electrical energy suitable for powering wireless sensors, low power microprocessors or charging batteries. Rotary Uses include rotating machines such as fans, turbines, drills, the wheels on electric cars, locomotives and conveyor belts. Also, in many vibrating or oscillating machines, an electric motor spins an unbalanced mass, causing the motor (and its mounting structure) to vibrate.

Atmospheric Electricity

[0006] There is always free electricity in the air and in the clouds, which acts by induction on the earth and electromagnetic devices. Experiments have shown that there is always free electricity in the atmosphere, which is sometimes negative and sometimes positive, but most generally positive, and the intensity of this free electricity is greater in the middle of the day than at morning or night and is greater in winter than in summer. In fine weather, the potential increases with altitude at about 30 volts per foot (100 V/m).

Atmospheric Layers

[0007] The electrical conductivity of the atmosphere increases exponentially with altitude. The amplitudes of the electric and magnetic components depend on season, latitude, and height above the sea level. The greater the altitude the more atmospheric electricity abounds. The exosphere is the uppermost layer of the atmosphere and is estimated to be 500 km to 1000 km above the Earth's surface, and its upper boundary at about 10,000 km. The thermosphere (upper atmosphere) is the layer of the Earth's atmosphere directly above the mesosphere and directly below the exosphere. Within this layer, ultraviolet radiation causes ionization. Theories that have been proposed to explain the phenomenon of the polar aurora, but it has been demonstrated by experiments that it is due to currents of positive electricity passing from the higher regions of the atmosphere to the earth.

[0008] The mesosphere (middle atmosphere) is the layer of the Earth's atmosphere that is directly above the stratosphere and directly below the thermosphere. The mesosphere is located about 50-80/85 km above Earth's surface. The stratosphere (middle atmosphere) is a layer of Earth's atmosphere that is stratified in temperature and is situated between about 10 km and 50 km altitude above the surface at moderate latitudes, while at the poles it starts at about 8 km altitude. The stratosphere sits directly above the troposphere and directly below the mesosphere. The troposphere (lower atmosphere) is the densest layer of the atmosphere.

[0009] The planetary boundary layer (PBL), also known as the atmospheric boundary layer (ABL), is the lowest part of the atmosphere and its behavior is directly influenced by its contact with the planetary surface. It is also known as the "exchange layer". (see also: p-n junction.)

[0010] There is a potential gradient at ground level ("Atmosphere ground layer") and this vertical field corresponds to the negative charge in and near the Earth's surface. The negative potential gradient falls rapidly as altitude increases from the ground. Most of this potential gradient is in the first few kilometers. The positive potential gradient rises rapidly as altitude increases from the ground. Volta, over two centuries before the 21st century, discovered with some degree of exactitude that the proportions of the ordinates of the curve or gradient of electric potential increased as the distance from the earth increases, and, more recently, Engel has provided data to calculate the increase (Image to the right).

Drum-Type Generator

[0011] A drum-type homopolar generator has a magnetic field (B) that radiates radially from the center of the drum and induces voltage (V) down the length of the drum. A conducting drum spun from above in the field of a "loudspeaker" type of magnet that has one pole in the center of the drum and the other pole surrounding the drum could use conducting ball bearings at the top and bottom of the drum to pick up the generated current.

Astrophysical Unipolar Inductors

[0012] Unipolar inductors occur in astrophysics where a conductor rotates through a magnetic field, for example, the movement of the highly conductive plasma in a cosmic body's ionosphere through its magnetic field. In their book, Cosmical Electrodynamics, Hannes Alfvén and Carl-Gunne Fälthammar write:

[0013] "Since cosmical clouds of ionized gas are generally magnetized, their motion produces induced electric fields [. . .] For example the motion of the magnetized interplanetary plasma produces electric fields that are essential for the production of aurora and magnetic storms" [. . .]

[0014] " . . . the rotation of a conductor in a magnetic field produces an electric field in the system at rest.

[0015] This phenomenon is well known from laboratory experiments and is usually called 'homopolar' or 'unipolar' induction.

The Faraday Disc

[0016] The homopolar generator was developed first by Michael Faraday during his experiments in 1831. It is frequently called the Faraday disc in his honor. It was the beginning of modern dynamos—that is, electrical generators which operate using a magnetic field. It was very inefficient and was not used as a practical power source, but it showed the possibility of generating electric power using magnetism, and led the way for commutated direct current dynamos and then alternating current alternators.

Boeing 737-800

[0017] The Boeing Fuel Cell Demonstrator Airplane has a Proton Exchange Membrane (PEM) fuel cell/lithium-ion battery hybrid system to power an electric motor, which is coupled to a conventional propeller. The fuel cell provides all power for the cruise phase of flight. During takeoff and climb, the flight segment that requires the most power, the system draws on lightweight lithium-ion batteries.

[0018] The demonstrator aircraft is a Dimona motor glider, built by Diamond Aircraft Industries of Austria, which also carried out structural modifications to the aircraft. With a wing span of 16.3 meters (53.5 feet), the airplane will be able to cruise at approximately 100 kilometers per hour (62 miles per hour) on power from the fuel cell.

[0019] Nikola Tesla explored the wireless transmission of energy through his work with radio and microwaves and his creation of the Tesla coil and the magnifying transmitter. In 1898, Tesla demonstrated his radio-controlled boat, which he was able to control remotely. In the 1930s, Tesla claimed to have invented a particle beam weapon, or, as some called it, a "peace ray." The device was, in theory, capable of generating an intense, targeted beam of energy and sending it across great distances to demolish warplanes, foreign armies, or anything else you'd rather didn't exist.

[0020] "Roy J. Meyers, British Patent Number 1098"

[0021] This invention relates to improvements in apparatus for the production of electrical currents, and the primary object in view is the production of a commercially serviceable electrical current without the employment of mechanical or chemical action. To this end the invention comprises means for producing what I believe to be dynamic electricity from the earth and its ambient elements.

[0022] Edward Leedskalnin: Magnetic Current—

[0023] The Perpetual Motion Holder is primarily a teaching device but it has many functions including an electromagnet, this is easy enough to see; it is a generator—spin a magnet between the coils it will generate electricity; it functions as a transformer; it demonstrates how permanent magnets are made, and is a holder of perpetual motion.

Strategic Defense Initiative

[0024] The Strategic Defense Initiative (SDI) was proposed to use ground and space-based systems to protect the United States from attack by strategic nuclear ballistic missiles.

Description of the Related Art

[0025] The present invention relates to a ball bearing assembly structure, an electromagnetic clutch having the ball bearing assembly structure, and a gas compressor equipped with the electromagnetic clutch.

[0026] When operating the gas compressor, the electromagnet of the electromagnetic clutch is energized to attract or adsorb the follower armature plate to an end surface of the prime-mover pulley and join the prime-mover pulley and the rotor shaft, thereby rotating the rotor shaft.

[0027] The ball bearing of the electromagnetic clutch conventionally has used one having an even number of balls per row. Generally, the ball bearing causes vibration and noise due to rotation. In the case of the ball bearing rotating while undergoing a radial load due to a tension of the belt, vibration and noise considerably occur. Particularly when other vibration and noise levels are lowered during engine idling, the vibration and noise of the ball bearing transmitted to the vehicular compartment is not negligible.

[0028] The inventor has conducted various experiments and discovered that the one factor of high vibration and noise level is an even number of balls of the ball bearing. In the ball bearing having an even number of balls per one row, the balls are in a facing relation to have linear-symmetry arrangement between the inner race and the outer race. The deformation and vibration at a regular particular frequency is caused in the inner and outer races. It is to be considered that the vibration as a source also increases noise.

[0029] The present invention relates generally and in various embodiments to piezoelectric mechanical systems. More specifically, the present invention relates generally and in various embodiments to atmospheric oscillation transducer apparatuses, systems, and methods.

[0030] Although various implementations of the present invention, among many, may be described herein with reference to the specific illustrative embodiments related to particular applications, those skilled in the art will understand that the invention is not in any way intended to be limited to such embodiments and/or applications. Those having ordinary skill in the art and reference to the description of the embodiments herein will recognize additional modifications, applications, and other embodiments falling within the scope of the claimed invention and additional fields in which the present invention may be practiced.

[0031] Digital Radio (also known as Satellite Radio or Satellite Digital Audio Radio Service (SDARS)) is a subscriber-based digital radio service that is broadcast via satellites. Digital radio service provides compact-disc (CD) quality programming that may be digitally transmitted via one or more satellites and/or space stations to one or more Earth-based (terrestrial) digital radio stations, receivers, and/or repeaters. In satellite-based direct-broadcast radio services, digitally-encoded audio program material may be broadcast to terrestrial fixed or mobile digital radio receivers. Fixed receivers may include, for example, stand alone digital radio receivers or digital radio receivers connected via computer networks, including for example, the Internet. Mobile receivers may include, for example, digital radio receivers located in automobiles, aircrafts, watercrafts, and the like.

[0032] Satellite-based digital audio radio services such as SDARS, for example, may be broadcast to one or more digital radio receivers either directly from an orbiting satellite, or indirectly from one or more repeater stations. Such repeater stations may be useful where the digital radio receiver is located in a shielded location or where there is no direct line of sight between the radio and the satellite. In other digital audio radio services systems, the audio programs also may be transmitted in digital form by one or more space stations directly to fixed, mobile, and/or portable radio stations. Such systems may comprise, for example, orbiting satellites, complementary repeating terrestrial transmitters, telemetry, tracking, and control facilities.

[0033] Combinations of mechanical devices U.S. Pat. Nos. 4,019,073, 6,615,968 and atmospheric system interaction are disclosed in U.S. Pat. Nos. 1,119,732, 787,412, 6,902,513 to Nikola Tesla; 28,793 to Charles Vion; and U.S. Pat. No. 1,540,998 to Herman Plauson. Lastly, U.S. Pat. No. 8,102,078 and US2008/0009240. Agnoff discloses an Oscillating watch winder in U.S. Pat. No. 6,543,929, Jennings further discloses an oscillating smart device in application No. 13,572,679.

[0034] As illustrated by a large body of prior art, including the above-noted patents, and a large number of commercial devices, efforts are continuously being made in an attempt to improve helmets, headsets and their methods of fabrication. Nothing in the prior art, however, suggests the present inventive combination of materials and method steps as herein described and claimed. The present invention achieves its purposes, objects and advantages over the prior art through a new, useful and unobvious combination of components and method steps which improve safety, comfort and noise abatement performance.

[0035] Therefore, it is an object of this invention to provide Effectively, the provision of energy such that it meets the needs of the present without compromising the ability of future generations to meet their own needs . . . . Sustainable Energy has two key components: renewable energy and energy efficiency.

[0036] It is still a further objection of this invention to promote Dynamic harmony between equitable availability of energy-intensive goods and services to all people and the preservation of the earth for future generations." And, "the solution will lie in finding sustainable energy sources and more efficient means of converting and utilizing energy.

[0037] It is a further object of the present invention to produce Green Power Energy: is energy that can be extracted, generated, and/or consumed without any significant negative impact to the environment, green power; as electricity produced from solar, wind, geothermal, biogas, biomass, and low-impact small hydroelectric sources.

[0038] Thus, there is a need for a clean energy system that uses atmospheric electricity.

Prior Art

[0039] Quartz crystals have been in regular use for many years to give an accurate frequency for all radio transmitters, radio receivers and computers. Their accuracy comes from an amazing set of coincidences: Quartz—which is silicon dioxide like most sand—is unaffected by most solvents and remains crystalline to hundreds of degrees Fahrenheit. The property that makes it an electronic miracle is the fact that, when compressed or bent, it generates a charge or voltage on its surface. This is a fairly common phenomenon called the Piezoelectric effect. In the same way, if a voltage is applied, quartz will bend or change its shape very slightly.

[0040] If a bell were shaped by grinding a single crystal of quartz, it would ring for minutes after being tapped. Almost no energy is lost in the material. A quartz bell—if shaped in the right direction to the crystalline axis—will have an oscillating voltage on its surface, and the rate of oscillation is unaffected by temperature. If the surface voltage on the crystal is picked off with plated electrodes and amplified by a transistor or integrated circuit, it can be re-applied to the bell to keep it ringing.

[0041] The electronics of the watch initially amplifies noise at the crystal frequency. This builds or regenerates into oscillation—it starts the crystal ringing. The output of the watch crystal oscillator is then converted to pulses suitable for the digital circuits.

Polymers

[0042] Polyvinylidene fluoride (PVDF): PVDF exhibits piezoelectricity several times greater than quartz. Unlike ceramics, where the crystal structure of the material creates the piezoelectric effect, in polymers the intertwined long-chain molecules attract and repel each other when an electric field is applied.

Near Space

[0043] Solar particles become trapped within the Earth's magnetic field and form radiation belts. The Van Allen radiation belt is a torus of energetic charged particles (i.e. a plasma) around Earth, trapped by Earth's magnetic field.

[0044] At elevations above the clouds, atmospheric electricity forms a continuous and distinct element (called the electrosphere) in which the Earth is surrounded. The electrosphere layer (from tens of kilometers above the surface of the earth to the ionosphere) has a high electrical conductivity and is essentially at a constant electric potential. The ionosphere is the inner edge of the magnetosphere and is the part of the atmosphere that is ionized by solar radiation. (Photoionisation is a physical process in which a photon is incident on an atom, ion or molecule, resulting in the ejection of one or more electrons.)

Advantages/Disadvantages

[0045] Energy in electronic elements: Electric potential energy, or electrostatic potential energy, is a potential energy (measured in joules) that results from conservative Coulomb forces and is associated with the configuration of a particular set of point charges within a defined system. The term "electric potential energy" is used to describe the potential energy in systems with time-variant electric fields, while the term "electrostatic potential energy" is used to describe the potential energy in systems with time-invariant electric fields.

[0046] Capacitance is the ability of a body to store an electrical charge. Any body or structure that is capable of being charged, either with static electricity or by an electric current, exhibits capacitance. A common form of energy storage device is a parallel-plate capacitor. In a parallel plate capacitor, capacitance is directly proportional to the surface area of the conductor plates and inversely proportional to the separation distance between the plates. If the charges on the plates are +q and −q, and V gives the voltage between the plates, then the capacitance C is given by

$$C=q/V.$$

[0047] The capacitance is a function only of the physical dimensions (geometry) of the conductors and the permittivity of the dielectric. It is independent of the potential difference between the conductors and the total charge on them.

[0048] Piezoelectricity is the combined effect of the electrical behavior of the material:

$$D=\in E$$

where D is the electric charge density displacement (electric displacement), $\in$ is permittivity and E is electric field strength, and

$$\text{Hooke's Law: } S=sT$$

where S is strain, s is compliance and T is stress.

[0049] Polyvinylidene fluoride, or polyvinylidene difluoride (PVDF) is a highly non-reactive and pure thermoplastic fluoropolymer produced by the polymerization of vinylidene difluoride. PVDF is a specialty plastic material in the fluoropolymer family; it is used generally in applications requiring the highest purity, strength, and resistance to solvents, acids, bases and heat and low smoke generation during a fire event. Compared to other fluoropolymers, it has an easier melt process because of its relatively low melting point of around 177° C.

[0050] It has a low density (1.78) and low cost compared to the other fluoropolymers. It is available as piping products, sheet, tubing, films, plate and an insulator for premium wire. It can be injected, molded or welded and is commonly used in the chemical, semiconductor, medical and defense industries, as well as in lithium ion batteries. It is also available as a crosslinked closed cell foam, used increasingly in aviation and aerospace. PVDF has a glass transition temperature (Tg) of about −35° C. and is typically 50-60% crystalline. To give the material its piezoelectric properties, it is mechanically stretched to orient the molecular chains and then poled under tension. PVDF exists in several forms: alpha (TGTG'), beta (TTTT), and gamma (TTTGTTTG') phases, depending on the chain conformations as trans (T) or gauche (G) linkages. When poled, PVDF is a ferroelectric polymer, exhibiting efficient piezoelectric and pyroelectric properties. These characteristics make it useful in sensor and battery applications. Thin films of PVDF are used in some newer thermal camera sensors.

[0051] Copolymers: Copolymers of PVDF are also used in piezoelectric and electrostrictive applications. One of the most commonly-used copolymers is P(VDF-trifluoroethylene), usually available in ratios of about 50:50 wt % and 65:35 wt % (equivalent to about 56:44 mol % and 70:30 mol %). Another one is P(VDF-tetrafluoroethylene). They improve the piezoelectric response by improving the crystallinity of the material.

[0052] While the copolymers' unit structures are less polar than that of pure PVDF, the copolymers typically have a much higher crystallinity. This results in a larger piezoelectric response: d33 values for P(VDF-TrFE) have been recorded to be as high as −38 pC/N versus −33 pC/N in pure PVDF.

[0053] Applications:

[0054] The piezoelectric properties of PVDF are used to advantage to manufacture tactile sensor arrays, inexpensive strain gauges and lightweight audio transducers. Piezoelectric panels made of PVDF are used on the Venetia Burney Student Dust Counter, a scientific instrument of the New Horizons space probe that measures dust density in the outer solar system. PVDF is the standard binder material used in the production of composite electrodes for lithium ion batteries. 1-2% weight solution of PVDF dissolved in N-Methyl-2-pyrrolidone (NMP) is mixed with an active lithium storage material such as graphite, silicon, tin, $LiCoO2$, $LiMn2O4$, or $LiFePO4$ and a conductive additive such as carbon black or carbon nanofibers. This slurry is cast onto a metallic current collector and the NMP is evaporated to form a composite or paste electrode. PVDF is used because it is chemically inert over the potential range used and does not react with the electrolyte or lithium. Piezoelectric elements can be used in laser mirror alignment, where their ability to move a large mass (the mirror mount) over microscopic distances is exploited to electronically align some laser mirrors. By precisely controlling the distance between mirrors, the laser electronics can accurately maintain optical conditions inside the laser cavity to optimize the beam output. Piezoelectric sensors especially are used with high frequency sound in ultrasonic transducers for medical imaging and also industrial nondestructive testing (NDT).

[0055] For many sensing techniques, the sensor can act as both a sensor and an actuator—often the term transducer is preferred when the device acts in this dual capacity, but most piezo devices have this property of reversibility whether it is used or not. Ultrasonic transducers, for example, can inject ultrasound waves into the body, receive the returned wave, and convert it to an electrical signal (a voltage). Most medical ultrasound transducers are piezoelectric.

[0056] Advantageously,

[0057] Sustainable energy is the sustainable provision of energy that meets the needs of the present without compromising the ability of future generations to meet their needs. Technologies that promote sustainable energy include renewable energy sources, such as hydroelectricity, solar energy, wind energy, wave power, geothermal energy, and tidal power, and also technologies designed to improve energy efficiency.

[0058] This sequence of oscillations causes the rotor within the watch to spin rapidly thereby winding the watch in a manner closely simulating the spinning of the rotor that occurs during normal winding of the watch when the watch is worn by a user. Due to the forces that are exerted, the rotor spins around the watch shaft during the oscillations, as opposed to the partial rotation observed in prior art mechanisms. Therefore, the time required to wind the watch, and the energy required, is substantially reduced. Moreover, since the rotor is spinning about the shaft, as opposed to merely being held in a downward position while the watch is rotated, winding more closely approximating the design mechanism is achieved, thereby putting less wear on the watch.

[0059] This invention relates to satellite communications systems using multiple spot beams from a geosynchronous earth orbit satellite to provide selective coverage of the continental United States and, more particularly, relates to a system having a satellite receiving hub in every spot beam that allows for asynchronous communications between each hub and the satellite for maximizing frequency re-use.

[0060] These purposes, objects and advantages should be construed as merely illustrative of some of the more prominent features and applications of the present invention. Many other beneficial results can be obtained by applying the disclosed invention in a different manner or by modifying the invention within the scope of the disclosure. Accordingly, other purposes, objects and advantages as well as a fuller understanding of the invention may be had by referring to the summary herein mentioned and detailed description describing the preferred embodiments of the invention, in addition to the scope of the invention, as defined by the claims, taken in conjunction with the accompanying drawings.

SUMMARY OF THE INVENTION

[0061] In one general respect, an embodiment of the present invention is directed to a system. The system includes a Power Frequency broadcast station, a receiver, and a network. The Power Frequency broadcast station includes a transmitter and a server. The receiver is in communication with the Power Frequency broadcast transmitter and also includes a user interface for receiving user input commands comprising a request for information from the Power Frequency broadcast station.

[0062] The receiver is configured to establish a two-way communication path between the receiver and the Power Frequency broadcast transmitter. The network is in communication with the server and the receiver for exchanging information therebetween. The request for information is provided to the server via the network and the server is configured to receive the request and transmit a response message to the receiver in accordance with the request.

Continuously Outboard Recharged Electric Vehicle (COREY)

[0063] Given suitable infrastructure, permissions and vehicles, BEVs (battery electric vehicles) can be recharged while the user drives. The BEV establishes contact with an electrified rail, plate or overhead wires on the highway via an attached conducting wheel or other similar mechanism (see Conduit current collection). The BEV's batteries are recharged by this process—on the highway—and can then be used normally on other roads until the battery is discharged. Some of battery-electric locomotives used for maintenance trains on the London Underground are capable of this mode of operation. Power is picked up from the electrified rails where possible, switching to battery power where the electricity supply is disconnected.

[0064] The present invention is directed to overcome the disadvantages of the prior art. The invention is a mechanism which taps into the naturally occurring static electricity in the atmosphere. Whereas heretofore, the attempt to garner electricity from the atmosphere has focused exclusively on capturing lightning, the present invention syphons off the static electricity which is generated from any agitated air and voids lightning.

[0065] Lightning is only the final discharge of the static electricity, whether that lightning is intra-cloud lightning, cloud-to-ground lightning, or inter-cloud lightning. Other types of final discharges are known as heat lightning, summer lightning, sheet lightning, ribbon lightning, silent lightning, ball lightning, bead lightning, elves, jets, and sprites. Well before these discharges are observed, as the atmosphere becomes agitated by wind or thermal, static electricity is being generated. The present invention recognizes that this static electricity is being formed and creates a mechanism to capture it.

[0066] In the preferred embodiment, a sensor array is used to monitor the activities both at the base unit (such as electrical flow within the conductor) and in the surrounding locale. A sensor monitoring the electrical flow (i.e. voltage and/or current) within the conductor is used to monitor the electrical activity within the conductor.

[0067] In the preferred embodiment, a lightning sensor monitors for lightning activity within the locale. As noted earlier, the electrical characteristic of lightning is so extreme that ideally this discharge is avoided as it might damage the mechanism of this invention. The sensor array is utilized by a controller, such as microprocessor, programmed to operate the mechanism as outlined herein.

[0068] The controller operates the winch motor to extend or withdraw the conductive line and by extension the altitude of the balloon. The controller is programmed to operate the winch by monitoring the electrical characteristics of the conductor and adjusting the balloon's altitude to maintain these characteristics within the conductor within a preset range.

[0069] This preset range is established either in the base programming of the controller or is established by an operator of the system. As example, by controlling the amount of current being withdrawn from the atmosphere, the mechanism operates within a safe range and also provides a relatively stable current flow from which a variety of activities can take place (such as DC-AC conversion).

[0070] The controller also utilizes the lightning sensor to protect the mechanism from a lightning strike. Should lightning be detected within a pre-determined range (as established by the software or defined by an operator), then the balloon is pulled down to minimize the risk of damage from a lightning strike.

[0071] Another aspect of the invention relates to the electrical system which accepts the fluctuating atmospheric charge and changes it into an acceptable configuration for either the desired load or for the existing power grid.

[0072] Power grids in the United States operate with a frequency of 60 hertz in an alternating current arrangement. While this basic configuration seems to be universally accepted, the voltage within the grid varies dramatically, such as 15 kv, 34 kv, 69 kv, and even 112 kv.

[0073] Each atmospheric generator is placed proximate to or within easy access to a specific grid; this establishes the required electrical output configuration (i.e. that which is accepted by the power grid). As example, one of the atmospheric electrical collector units as described above collects the atmospheric electrical power as direct current and then supplies the appropriate power grid a specific flow (as example, AC, 60 hertz, at 69 kv).

[0074] The difficulty lies the fact that the DC current being garnered from the atmosphere varies depending on the actual agitation being generated in the atmosphere. This means that the source of DC current is fluctuating.

[0075] The present invention uses a monitoring system which checks the input DC voltage. Depending on the actual voltage being received, the appropriate converter is connected to the input DC voltage so that the desired output is obtained.

[0076] As example, suppose the DC input voltage is 1500 volts, the monitoring system, sensing this input, closes the switch connecting the DC voltage to a converter which accepts DC voltage in the range of 1000-2000 volts which then delivers an AC, 60 hertz 69 kv signal to the power grid. If the DC input voltage increases to 2100 volts, then the monitoring system opens the switch to the first converter (1000-2000 volts) and closes the switch to a second converter (such as 2000-4000 volts) to deliver the desired output of AC (60 hertz, 69 kv) for the power grid.

[0077] In this manner, regardless of the fluctuating input DC voltage, the electrical grid is supplied with a fully configured electrical input conforming to the needs of that specific electrical grid.

[0078] Another aspect of the present invention is the use of a tower or permanent structure instead of an aircraft. In this embodiment, the building or tower is electrically isolated from the ground and a rod (similar to a lightning rod) is extended into the atmosphere. The rod collects the atmospheric charge which is conveyed via an electrical conduit (ideally insulated) where the collected DC charge is reconfigured to meets the need of the locale.

[0079] In this context, for one embodiment of the invention, a tower is placed onto the top of a building. The tower is electrically isolated from the building using such mechanisms well known to those of ordinary skill in the art such as rubber mats. A rod ideally extends from the top of the tower to facilitate the collection of the DC electrical energy.

[0080] Piezoelectricity is a material property that is manifested when voltage is produced by applying mechanical forces, and vice versa, the effect has been described as direct and converse. Piezoelectricity has been described as coupling between a quasi-static electric field and dynamic mechanical motion.

[0081] A piezoelectricity converter mechanism such as described above, is connected to the tower to flow the DC electricity to a converter which modifies the DC current for the specific application. In one application, the DC current is converted to the electrical needs of the building, thereby providing at least some of the electrical requirements of the building itself.

[0082] As noted earlier, the dynamic converter system of the present invention allows a power generator to address a variable voltage in an efficient manner. This makes the dynamic converter system ideal for a variety of alternative energy sources such as the above described atmospheric electrical generator and other alternative energy sources such as wind and wave powered systems. In these systems, the energy being generated must be converted to a proper electrical configuration for a identified load. This may be a particular motor or connection to the power grid which act as a load to the power generating mechanism.

[0083] For these energy generating systems, the converter assembly of this invention utilizes multiple converters. Each converter is configured to accept a unique range of voltages and from these voltages, create the desired electrical configuration. By using multiple converters, a full range is available, from a minimum voltage input to a maximum voltage input.

[0084] The present invention can include systems and methods for integrating sensors for tracking atmospheric transducer system performance metrics into media devices and accessories therefor, thereby reducing or eliminating the need for additional independent monitoring devices. In one embodiment of the present invention.

[0085] It also is known to provide such transducers with connectors to allow their rechargeable batteries to be charged. In some cases, the connector is a Universal Serial Bus (USB) connector, allowing the transducer to be charged by plugging it into the USB port of a computer, grid circuit or other device.

[0086] These and other objects and advantages of the invention will appear more clearly from the following description in which the preferred embodiment of the invention has been set forth in conjunction with the drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0087] FIG. **1** Prior Art depicts perspective view of Electric currents created in sunward ionosphere;

[0088] FIG. **2** is a box flow chart of the propulsion cycle systems present invention;

[0089] FIG. **3** is an top view of the piezoelectric disk cylinder orb of the present invention;

[0090] FIG. **4** is a longitudinal sectional view showing an embodiment of this invention;

[0091] FIG. **5** is a top view of the (ATS) slip rotor piezoelectric chamber of the present invention;

[0092] FIG. **6** is an alternate top view of the piezoelectric cylinder orb of the present invention;

[0093] FIG. **7-10** elevation views of spine piezo stacks embodiments of the present invention;

[0094] FIG. **11** illustrates one embodiment of a power service (ATS) system architecture;

[0095] FIG. **12** is a block diagram of the ATS charge schematics systems of the present invention;

[0096] FIG. **13** is a block diagram of an ATS charge and recycle schematics of the present invention;

[0097] FIG. **14** is a view of the piezoelectric ball race cylinder of the present invention;

[0098] FIG. **15** is a view of the piezoelectric housing and gear of the present invention;

[0099] FIG. **16** is a view of a multiple piezoelectric ball race cylinder of the present invention;

[0100] FIG. **17** is another embodiment of the piezoelectric stack ball of the present invention;

[0101] FIGS. **18** & **19** are antenna rod transmit device embodiments of the (ATS) present invention;

[0102] FIG. **20** illustrates substation embodiment of the power service (ATS) system architecture;

[0103] FIG. **21** illustrates a home embodiment of the power service (ATS) system architecture;

[0104] FIG. **22** illustrates a window embodiment of the power service (ATS) system architecture;

[0105] FIG. **23** is (ATS) shock device in accordance with an embodiment of the present invention;

[0106] FIG. **24** is (ATS) shock device in accordance with an embodiment of the present invention;

[0107] FIG. **25** is (ATS) motorcycle device embodiment of the present invention;

[0108] FIG. **26** is (ATS) axle device in accordance with an embodiment of the present invention;

[0109] FIG. **27** is (ATS) train device in accordance with an embodiment of the present invention;

[0110] FIG. **28** is (ATS) plane device in accordance with an embodiment of the present invention;

[0111] FIG. **29** is (ATS) boat device in accordance with an embodiment of the present invention;

[0112] FIG. **30** is (ATS) solar device in accordance with an embodiment of the present invention;

[0113] FIG. **31** a (ATS) turbine device in accordance with an embodiment of the present invention;

[0114] FIG. **32** is (ATS) auto body panel device embodiment of the present invention;

[0115] FIG. **33** is (ATS) motorcycle body fairing panel device in of the present invention; and

[0116] FIG. **34** is (ATS) cross section view of body embed panel device of the present invention.

DETAILED DESCRIPTION

[0117] This sequence of oscillations causes the conductive rotor within piezoelectric molded housing device to spin rapidly thereby winding the mechanism in a manner closely simulating the spinning of the conductive rotor that occurs during normal electric activity when the device is activated. Due to the forces that are exerted, the conductive rotor spins around the piezoelectric cylinder device shaft during the oscillations, as opposed to the partial rotation observed in prior art mechanisms. Therefore, the time required to charge the Atmospheric Transduction System (ATS) device, and the energy required, is substantially reduced. Moreover, since the rotor is spinning about the shaft, as opposed to merely being held in a downward position while the ATS device is rotated, recharging more closely approximating the design mechanism is achieved, thereby putting less wear on the ATS chamber device Innovative Piezoelectric housing and ball bearing, coupler and book spine stacks. There relies the notion of negative ground electricity and positive aerial electricity which is in abundance. This invention substantiates land vehicles recycle recharge by reverse oscillation. Aerial vehicles recycle recharge by forward oscillation in accumulation of environmental positive and negative electricity. These aforementioned activities are integral or synchronous with frequency.

[0118] Prior Art FIG. **1**

[0119] Electric currents created in sunward ionosphere. FIG. **2** is a recycle box flow chart **49** of a self propulsion unit consisting of a battery **27**, engine **29** and piezoelectric transducer **33** unit. FIG. **3** is a multiple disk **31** load part **51** about a 360 degree cylinder **45** and shaft **44**. FIG. **4** is a piezoelectric molded device **100** housing **110** containing a ball bearing race **104**, disk stacks capacitor **102** and conductive rotor **46**. FIG. **5** is a piezoelectric chamber **120** configuration including a counterweights **50**, shaft **44**, conductive slip rotor **106**. FIG. **6** is an alternate part **200** chamber cylinder **201** embodiment containing piezoelectric spine disk stack capacitors **102**. FIG. **7-10** are variations of spine **107**,**108**,**109**,**110** case piezoelectric stack capacitor **102** with plate and ball heads. FIG. **11** Various embodiments of the present invention, among others, will now be described with reference to the accompanying drawings. Accordingly, FIG. **11** illustrates an embodiment of a Atmospheric Transduction (ATS) System **300** architecture. The system **300** may include, for example, a satellite broadcast station **318** that transmits signals **333** containing frequency content from a geostationary satellite **312** by way of airplane antenna **326**. In turn, the satellite **312** transmits line-of-sight (LOS) signals **333** to one or more ATS terrestrial frequency power receiver farms **314**. The system **300** also may include one or more terrestrial repeaters **316** which receive and retransmit the satellite signals **333** as repeater signals **323** to facilitate reliable reception in geographic areas where LOS reception from the satellite **312** is obstructed by tall buildings, hills, tunnels, and other similar impediments to the signals **333**. The ATS receivers **314** maybe designed to receive one or more signals **333** from the satellite **312** and/or from the terrestrial repeater transceiver **316**. In operation, such ATS receivers **314** may receive both the satellite signals **333** and the repeater signals **323**. The receivers **314** also may be located in mobile environments **320**, **321**, **322** which include, but are not limited to, land vehicles **321**, **322**, aircraft **320**, watercraft **900**, and handheld devices, among others. The receivers **314** also may be fixed in stationary units for residential use (e.g., home **325**,**750** entertainment, etc.) or commercial **314**, **328**,**360** use (e.g., business **314**, office **700**, security **328**, etc.). The power frequency broadcast station **318** also may be in communication with a grid network **342**. Two-way communication between the ATS receivers **314** and the power frequency broadcast station **318** may occur via the network **342**. Furthermore, information feedback from the power frequency broadcast station **318** may be transmitted to the ATS receiver **314** both by way of the network **342** as well as via the satellite **312**. Information also may be transmitted to the power broadcast station **318** wirelessly via a wireless network **342**,**707** by way of transducer tower **360**.

[0120] Further disclosed in FIG. **12** and FIG. **13** (in block diagrams **400**, **500**) are electrical schematics for handling the static charge from the atmosphere. By maintaining the voltage being collected in a prescribed range, an electrical conversion system is easily designed. While FIGS. **12**, and **13** illustrate some electrical configurations, those of ordinary skill in the art readily recognize a variety of other configurations which will serve the same function.

[0121] Referencing FIG. **12**, Direct Current In (DC IN) **401** is buffered by a gang of capacitors **410** before being communicated to a DC/AC converter **420**. The DC/AC converter **420** converts the direct current into a an alternating current suitable for placement over an existing electrical grid **430** such as normally found from a power-plant. FIG. **12** Also illustrates an electrical arrangement suitable for use in charging a battery **440**. DC IN **401** is buffered by capacitor **410** bank before entering into a step down transformer **435**. Step down transformer **435** reduces the voltage so that the voltage can safely be introduced into battery **440** which is connected to ground **450** at the battery's other pole. In FIG. **12**, DC IN **401** is fed into an adjustable rheostat **460** which is controlled by the controller **465** so that the DC OUT **470** falls within a computer **475** monitored and sensor **480** specified range. FIG. **13** Hypothetically, unused energy may be recycled current **501** and/or recaptured by reversing the oscillated spin rotation of devices **51**, **100**, **120** with the use of a rectifier **502** and Step up transformer **503**, returned to grid **430** capacitor **410**. This theory lends itself to the concept of positive and negative frequency. FIG. **14-17** Self charging propulsion embodiment of the invention where Da Vinci's ball race **510**,**512** is combined with a disk cylinder **45** and **513** stack balls **525**, conductive rotor **46**, **511** tooth gear **515** and piezoelectric molded housing **555**. FIGS. **18** & **19** Improvement structures Franklin's lightening rod **600**,**603** and a molded vibration transducer **601** quasi replicating Vion's tubes and piezoelectric spine stacks capacitor **602** improving Tesla's Atmospheric transmit device. FIG. **20** is an atmospheric receiving building sub station **700** where energy is consumed and excess rendered to the grid **707** by conductive rotor **702** transducer **710**, tower transducer **703**, antenna rod **701** and transducer windows **704**.

[0122] FIG. **12** and FIG. **2** flowchart illustration also includes battery **440**. Battery **440** may provide electrical power to components of ATS devices within FIG. **11**. Charging circuitry may also be provided to charge battery **440** when an external power supply is connected to an ATS device **100**. FIG. **14-17** eliminates one or more steps by presenting a self charge retaining transducer **510-513** may be configured with an accelerator sensor **480** controller **465** and gears **515**, provide reciprocal power incorporated within piezoelectric molded and ceramic housing **555** along with stack balls **525** and cylinder **545**. This assembly more resembles a motor by characteristics given power with application.

[0123] In operation, as illustrated in FIGS. **2-20** and FIGS. **21-34**, is periodically energized by movement to rotate Orb in either a clockwise or counterclockwise direction. The length of time or activity is energized, and the length of time between the period when the capacitor **31**, **102** battery is energized, will depend on the particular ATS device design. As the Orb rotates, the outer end of the disk moves along a 360° circular pathway to push against with forward and rearwardly spinning. Upon engagement of the Orb, ATS device disk is rotated until carried to the apex or top of the circular pathway. Upon reaching the apex, the gravitational and vibrational force or counterweight **50** promotes additional oscillation. ATS device movement rapidly rotate on Orb at a rotational speed greater than the speed of rotation of Orb. Counterweight **50** is then carried beyond the bottom or lowest point of the pathway by its momentum to a point near the apex on the opposite side of the pathway. The cycle is repeated through multiple increasing oscillations of the ATS device until counterweight **50** stops at the bottom position, or until once again engages to again move counterweight to the top of its circular pathway.

[0124] FIG. **2** is a flow chart showing generation of energy using a rotor according to one or more of the above-described embodiments. First, battery **27** starts the engine **29** and/or mobile transducer **33** is oscillated. In response to this acceleration, forces are imposed on one or more rotation piezoelectric devices. In response to those forces, the piezoelectric devices output electrical energy, which energy is extracted at a power controller **465**. The power controller **465** sensor **480** then makes this energy available to recharge a capacitor **410** battery **440** and/or to electronic components of the mobile terminal. Although FIG. **2** shows a serial flow of events, it is to be appreciated that the events of blocks **33**, **27** and **29** occur substantially instantaneously upon acceleration of the mobile terminal.

Preferred Alternate Embodiments

[0125] The present invention (ATS) device in accordance with an embodiment of the present invention overcomes the foregoing problem in the conventional art and provides an electro energy vibration and alternative to gas, oil or fossil fuel consumption in FIG. **21** homes **750** transducer **755**, FIG. **28** airplane **880** transducer **801**, FIG. **27** train **860** transducer **801**, FIG. **32** auto **950** transducer **910**, FIG. **25** and FIG. **33** motorcycles **960**,**850** transducers **799**, **802**, **810**, **910** and FIG. **29** boat **900** transducers **901**,**910**.

[0126] In order to solve the foregoing problems in the conventional art, the present invention provides an electro transducer having a ball bearing assembly which is compatible with FIG. **26** axle shaft **44** transducer **810** wheel **870** assembly. In order to solve the foregoing problems in the conventional art, the present invention provides an electro transducer having an FIG. **34** housing panel **980** transducer **910** assembly which is compatible with an exterior body assembly. Thereby, an Atmospheric transducer device may be shock **800** integrated as within the FIG. **23** coupler **802**, and FIG. **24** ladder **799**, (vertical friction) piezoelectric absorbers assemblies. In order to solve the foregoing problems in the conventional art, the present invention provides an electro transducer having a capable flat assembly which is compatible with FIG. **22** window **704** disk **31** plate transducer **714**, FIG. **30** solar panel **920** transducer **910**, FIG. **31** wind turbine **930** transducer **910** assembly. This invention provided FIG. **3** piezoelectric plate disks **31**, FIG. **7-10** spine piezoelectric disks stacks **107-110** capacitor **102**, FIG. **17** piezoelectric stack balls **525** capacitor, the arrangement of balls **525** will not be in a facing relation. The deformation in the inner and outer races during rotation of the ball bearing while undergoing a radial load is made irregular and complicated. The spine eliminates the deformation and vibration level increased by a combination of ceramic and molded piezoelectric materials at a regular predetermined frequency thereby multiplying the level of vibration and noise reduction.

[0127] Certain modifications and improvements will occur to those skilled in the art upon a reading of the foregoing description. It should be understood that all such modifications and improvements have been deleted herein for the sake of conciseness and readability but are properly within the scope of the following claims.

[0128] Thus it is seen that an atmospheric transducer device may be integrated and/or provided. It will be understood that the foregoing is only illustrative of the principles of the invention, and that various modifications can be made by those skilled in the art without departing from the scope and spirit of the invention, and the present invention is limited only by the claims that follow.

1-3. (canceled)

4. An atmospheric transduction system comprising:

means for recycle recharge by oscillation and frequency in accumulation of environmental positive and negative electricity, maintaining the voltage being collected in a prescribed range, providing an electrical conversion broadcast network; and

means for collecting the atmospheric electrical power as direct current and then supplies the appropriate power grid, transceiver and capacitates a charge; and

means for self charging propulsion provides motor characteristics, and frequency engine.

5. The atmospheric transduction system in accordance with claim **1**, wherein said means for recycle recharge by oscillation and frequency in accumulation of environmental positive and negative electricity, maintaining the voltage being collected in a prescribed range, an electrical conversion broadcast network comprises piezoelectric transducer molded device(s), rotor, stack ball bearing, coupler, book spine stacks and antenna rod stacks;

wherein said means for collecting the atmospheric electrical power as direct current and then supplies the appropriate power grid and capacitates a charge comprises a piezoelectric network, piezoelectric grid, piezoelectric spine stack antenna, piezoelectric transmitters, piezoelectric receivers, piezoelectric devices, piezoelectric cylinders and orbs, power frequency broadcast;

wherein said means for charging providing motor characteristics, said power frequency engine comprises a sensor accelerator, rotor/gear, battery/capacitor, antenna, controller, for propulsion.

6. An atmospheric transduction system comprising:

a piezoelectric transducer molded device(s), rotor, spine ball bearing, coupler and book spine stack transducers, for recycle recharge by oscillation and frequency in accumulation of environmental positive and negative electricity, maintaining the voltage being collected in a prescribed range, providing an electrical conversion broadcast server network; and

a piezoelectric network, piezoelectric grid, piezoelectric spine stack antenna, piezoelectric transmitters, piezoelectric receivers, piezoelectric devices, piezoelectric cylinders and disc orbs, stack rods, power frequency broadcast, for collecting the atmospheric electrical power as direct current and then supplies the appropriate power grid, transceiver and capacitates a charge; and

a sensor accelerator, rotor/gear, battery/capacitor, controller, antenna, means for self charge propulsion providing motor characteristics, thereby an engine of power frequency.

* * * * *

June 6, 1933. C. F. STREET 1,913,010

GRATE ASSEMBLY

Filed Feb. 21, 1929 2 Sheets-Sheet 1

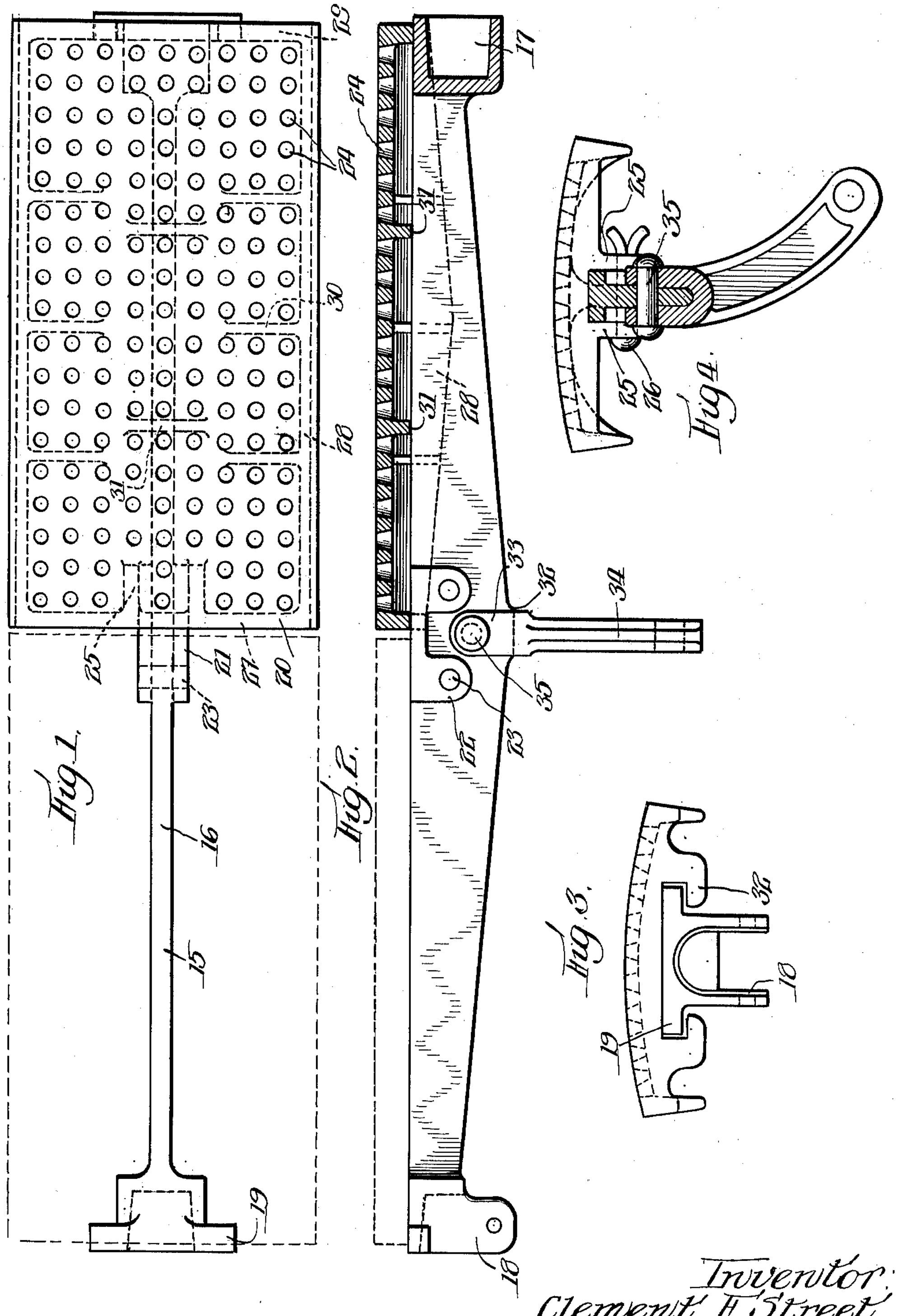
29
24
24
30
28
31
25
21
27
20
23
16
15
19
17
24
24
31
31
28
33
36
34
35
22
23
18
25
35
25
26
32
18
19

Fig. 1.

Fig. 2.

Fig. 3.

Fig. 4.

Inventor:
Clement F. Street,
By Wilkinson, Huxley, Byron & Knight
Attys.

June 6, 1933. C. F. STREET 1,913,010

GRATE ASSEMBLY

Filed Feb. 21, 1929 2 Sheets-Sheet 2

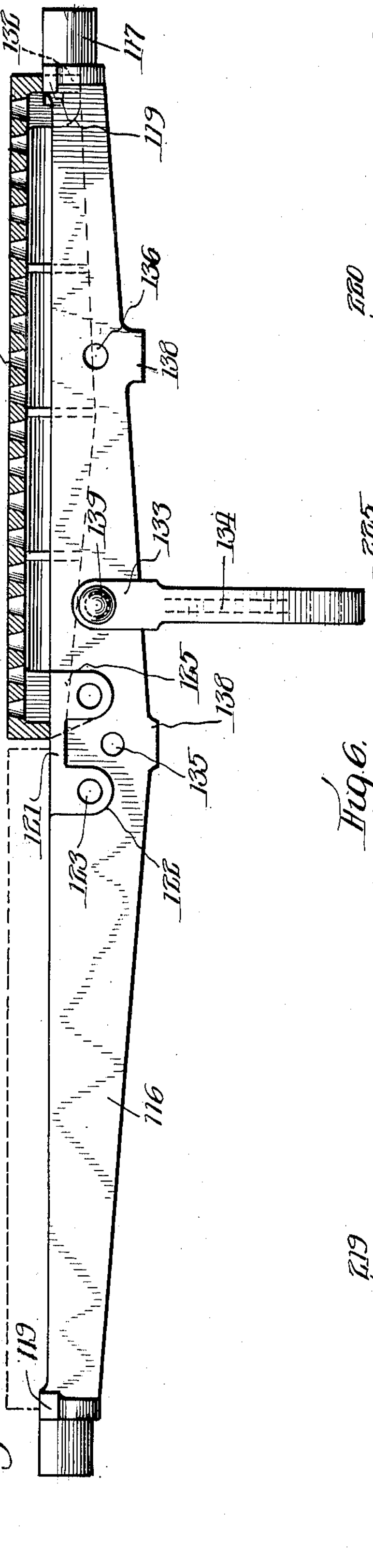
136
117
119
136
120
130
139
133
134
125
138
121
135
123
122
116
119

Fig. 5.

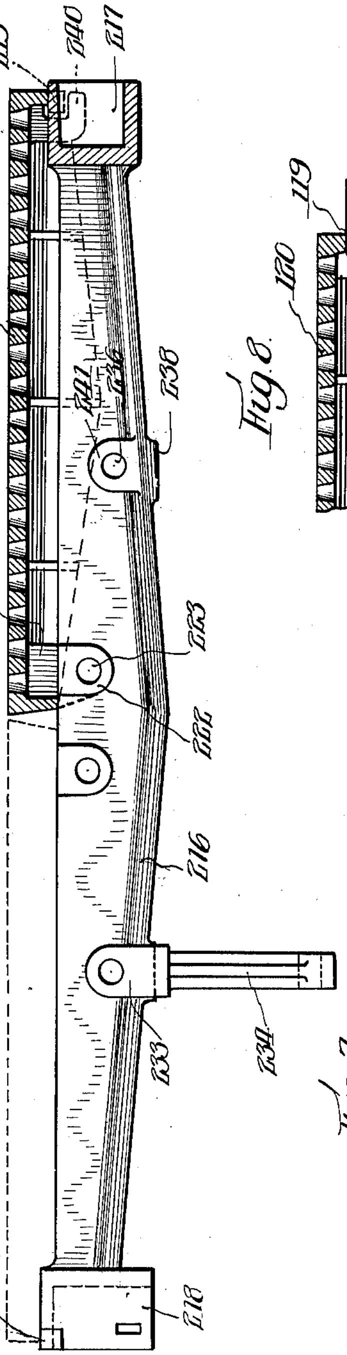
219
240
217
220
241
236
236
225
223
222
216
234
233
219
218

Fig. 6.

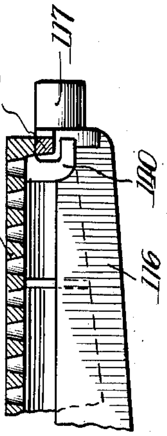
117
119
120
140
116

Fig. 8.

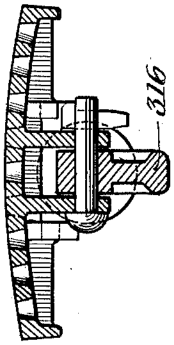

Fig. 7.

Inventor:
Clement F. Street,
By Wilkinson, Huxley, Byron & Knight
Attys.

UNITED STATES PATENT OFFICE

CLEMENT F. STREET, OF GREENWICH, CONNECTICUT, ASSIGNOR TO AMERICAN STEEL FOUNDRIES, OF CHICAGO, ILLINOIS, A CORPORATION OF NEW JERSEY

GRATE ASSEMBLY

Application filed February 21, 1929. Serial No. 341,683.

This invention pertains to grates, and more particularly to the multiple grate assemblies used on locomotives.

Locomotive grates are usually of the rocking type and it is common practice to shake them at frequent intervals while the locomotive is running. This is required for the reason that the coal commonly used leaves a heavy ash and clinkers between the burning coal and the grates, and if this is not removed, it will retard the flow to the fire of the air required for combustion. But there is a wide variation in the character of the coal used on locomotives, and while some of it when burning leaves a heavy ash and clinkers, other leaves a very light ash and few clinkers, if any.

It is common practice to use a drop grate in combination with a set of rocking grates, and this drop grate is operated by means of a set of levers which are handled from the ground. They can therefore be used only when the locomotive is standing.

This drop grate is pivoted near one side and is very different in construction from either the non-rocking type or from the rocking type shown in the construction herewith, and should not be confused therewith.

Grates of this type when used in this service are subject to a high degree of heat on their upper surface while the lower parts are immersed in a continuous flow of cold air. The different parts are therefore subject to different amounts of expansion and contraction, and an object of this invention is to provide a construction in which this will not result in warping and distortion of the parts of the assembly.

Another object of this invention is to provide a grate assembly which is inexpensive to build and maintain and one which fulfills all requirements for economic operation and maintenance.

Still another object is to provide a grate bar assembly which permits rapid cleaning of the fire and changes of the grate bars.

Yet another object is to provide a grate bar assembly which resists distortion, is readily removable and is capable of rapid conversion to provide different drafts to accommodate different fuels.

A further object is to provide a grate assembly wherein the grate members are secured to their supports by a minimum of parts, yet one which is so constructed that the grate members will be effectively and securely kept in place.

Still a further object is to provide a grate assembly which offers substantially no resistance to expansion and contraction yet at the same time is fixedly supported and effectively resists warping and twisting.

Yet a further object is to provide a grate assembly wherein the securing members between parts are reduced to a minimum, and wherein the clearances are predetermined and arranged to permit ease of operation and assemblage.

With these and various other objects in view, the invention may consist of certain novel features of construction and operation, as will be more fully described and particularly pointed out in the specification, drawings and claims appended hereto.

In the drawings, which illustrate an embodiment of the device and wherein like reference characters are used to designate like parts—

Figure 1 is a top plan view, showing the supporting relation of the grate frame to one of the grate plate members;

Figure 2 is a side elevation, partly in section, of the assemblage illustrated in Figure 1, the position occupied by the grate member completing the assembly being indicated in dotted lines;

Figure 3 is an end elevation of the grate assembly, the same being taken from the end of the grate frame provided with the trunnion;

Figure 4 is a cross section of the grate frame assembly illustrated in Figure 2, showing the relation of the shaker arm to the remainder of the assembly;

Figure 5 is a side elevation, partly in section, of a modified form of grate assembly, the position occupied by the grate member completing the assembly being indicated in dotted lines;

Figure 6 is a side elevation, partly in section, of still another modified form of grate assembly;

Figure 7 is a cross sectional view of the grate assembly illustrated in Figure 6; and

Figure 8 is a fragmentary side elevation, partly in section, of a modified form of securing means provided on the plate grate of the modified assembly illustrated in Figure 5.

Referring first of all to the modification shown in Figures 1 to 4 inclusive, the supporting frame 15 of the grate assembly consists essentially of a single longitudinal rib 16 provided at one end thereof with a trunnion socket 17 and at the other end thereof with a horseshoe or U-shaped trunnion socket 18, each of said sockets being provided with outstanding lug members 19 provided for locking purposes in a manner to be later described. It will of course be understood that the frame may be of any length to accommodate any size or any number of plate grates, the embodiment illustrated being for an assembly including two of such plate grates 20. The supporting frame is provided with an enlarged portion 21 adjacent the center thereof, said portion having downwardly extending bosses 22, the said bosses being provided with apertures 23. The grate plates 20 may be either of the arcuate or flat type provided with suitably disposed apertures 24, said apertures being of a size and spacing for proper draft purposes.

The plates 20 adjacent their inner or inboard end are provided with downwardly extending lugs 25 adapted to straddle the rib 16 and the bosses 22, said lugs being provided with aligned apertures adapted to register with the aperture 23 for the reception of a locking pin 26 for securing said plate grate to the support. The lugs 25 merge into and are integral with the plate and also with a reinforcing rib 27 formed on the inner end of the plate, said rib merging into and being integral with longitudinal ribs 28 which in turn merge into and are integral with a rib 29 formed on the outer end of said plate. Ribs 28 are preferably made deeper adjacent their centers than at their ends for the purpose of preventing the plate from sagging as a result of service, and to prevent said ribs from bowing outwardly, a series of local cross-ribs 30 is provided which ribs are integral with said ribs and merge into the plate.

In service, it has been found that under heat the ends of the plate will go up or the center will go down, so in order to prevent either one or both of these movements, local ribs 31 are provided suitably spaced and preferably in between or in staggered position with respect to the ribs 30, said ribs contacting with the longitudinal rib 16 of the frame. The inner end of the plate of course is restrained from going up by the pin 26, and the outer end is restrained from going up by providing the depending flange 29 with inwardly extending hooks 32 adapted to have locking engagement with the lugs 19. The frame may be suitably apertured at any place along its length, shown adjacent the center thereof, and may be provided with a guiding boss 38 adapted to cooperate with a portion of the jaws 33 of the shaker arm 34 secured by means of a rivet or lock connection 35 to said frame.

The structure is assembled with sufficient lost motion between the parts to allow them to go together easily and of course the first effect of warping will be to take up this lost motion and bring all the parts firmly in contact. After this has occurred, any further movement will be resisted by the vertical strength of the frame. In assembling the parts of the device, it is only necessary to slip the hooks 32 into engagement with the lugs 19, dropping the plate into place whereby the lugs 25 straddle the frame and the apertures in lugs 25 register with the aperture 23 provided in the bosses 22 whereupon the locking member 26 may be put in place. The adjacent grate plate may likewise be put in place, there being sufficient clearance for the parts to permit their ready assemblage.

Referring now more particularly to the form of the device illustrated in Figure 5, the longitudinal frame 116 is provided with trunnions 117 at each end thereof and outwardly extending lugs 119 are formed on said frame adjacent said trunnions. The plate grates 120 may be of the same construction as has been already described with respect to grate plate 20, the transverse hook members 132 being provided for cooperation with the lugs 119 for locking one end of the plate grate in operative position. As before, the longitudinal member is provided with an enlarged portion 121 provided with bosses 122 having apertures 123 therein for the reception of locking members cooperating with the depending lugs 125 provided on the grate plates. The longitudinal frame member may be provided at convenient places along its length with apertures 135 and 136 having bosses 137 and 138 disposed adjacent thereto for cooperation with parts of the shaker arm 134 to maintain the same in substantially fixed position with respect to said frame members, said shaker arm being provided with jaws 133 having apertures therein adapted to receive the locking member 139.

In the modification illustrated in Figure 8, the frame member 116 is provided with a trunnion 117 and outwardly extending lugs 119. The grate plate 120 is suitably supported on said frame but is provided with longitudinally extending hooks 140 adapted to cooperate with the lugs 119 to maintain the outer end of the grate plate in its operative position with respect to said frame.

Referring now more particularly to the modification illustrated in Figures 6 and 7, the longitudinal frame member 216 is provided at one end with a trunnion socket 217 and at the other end with the U-shaped trunnion 218. The frame member 216 is deeper adjacent the center thereof as is the case in the modifications already described, and this member may also be of substantially truss formation, as illustrated in Figure 7, that is, it is provided with an enlarged rib or bulbous portion 316 provided along the lower edge thereof. Outwardly extending lugs 219 are provided adjacent the trunnions and are adapted to provide cooperating locking relation with the longitudinal hook members 240 provided on the plate grate 220. The grate plate is provided with reinforcing flanges, as before described, the inner flange thereof being provided with depending lugs 225 provided with aligned apertures adapted to cooperate with apertures 223 provided in the lugs 222 formed on the longitudinal supporting member 216. Bosses 241 may also be provided on the supporting member adjacent suitably spaced apertures 236, said apertures being for the reception of securing means passing through the upstanding jaws 233 of the shaker arm 234, the longitudinal frame member being further provided with bosses 238 cooperating with the shaker arms to maintain the same in fixed position with respect to said frame.

The operation and assemblage of the modified forms are substantially the same as has already been described with respect to the first modification, and it is to be understood that I do not wish to be limited by the exact embodiment of the device shown, which is merely by way of illustration and not limitation, as various and other forms of the device will of course be apparent to those skilled in the art without departing from the spirit of the invention or the scope of the claims.

I claim:

1. In a grate assembly, the combination of a support, a grate member secured thereto, said member having staggered reinforcements thereon, certain of said reinforcements having contacting relation to said support.

2. In a grate assembly, the combination of a support, a grate member secured thereto, said member having reinforcements disposed at the sides and center thereof, the reinforcements at the center being in staggered relation to the reinforcements at the sides, said center reinforcements having contacting relation to said support.

3. In a grate assembly, the combination of a support having securing means thereon, a grate member carried by said support and having securing means at each end thereof adapted to cooperate with said first named securing means for fastening said grate member to said support, and reinforcing ribs disposed on said grate member and contacting with said support to prevent distortion of said grate member.

4. In a grate assembly, the combination of a support having securing means thereon, a grate member carried by said support and having a marginal reinforcement thereon, securing means at each end of said member adapted to cooperate with said first named securing means for fastening said grate member to said support, reinforcing ribs disposed on said grate member and integral with said marginal reinforcement to prevent distortion of the latter, and center reinforcing ribs disposed on said grate member and contacting said support to prevent distortion of said grate member.

5. In a grate assembly, the combination of a support having securing means thereon, a grate member carried by said support and having a marginal reinforcement thereon, securing means at each end of said member, certain of said means being integral with said marginal reinforcement, said last named securing means being adapted to cooperate with securing means on the support for fastening said grate member to said support, and reinforcing ribs disposed on said grate member and integral with said marginal reinforcement to prevent distortion of the latter.

6. In a grate assembly, the combination of a support having securing means thereon, a grate member carried by said support and having securing means at each end thereof adapted to cooperate with the securing means on the support for fastening said grate member to said support, one of said securing means including members disposed transversely of said grate member, and reinforcing ribs disposed on said grate member and contacting said support to prevent distortion of said grate member.

7. In a grate assembly, the combination of a support having securing means thereon, a grate member carried by said support and having securing means at each end thereof adapted to cooperate with the securing means on the support for fastening said grate member to said support, reinforcing ribs disposed on said grate member and contacting said support to prevent distortion of said grate member, and bosses and fastening means on said support for cooperation with shaking means for moving said assembly.

8. As an article of manufacture, a plate grate having marginal flanges, reinforcing means on said grate and integral with certain of said flanges to prevent distortion of said flanges, and fastening means on said grate integral with other of said flanges, certain of said fastening means including facing hook members.

9. As an article of manufacture, a plate grate having marginal flanges, reinforcing means on said grate and integral with certain of said flanges to prevent distortion of said flanges, reinforcing means adjacent the longitudinal center line of said grate and in staggered relation to said first named reinforcing means, and fastening means on said grate integral with other of said flanges, certain of said fastening means including facing hook members.

Signed at Chicago, Illinois, this 14th day of February, 1929.

CLEMENT F. STREET.

Patents

Device for obtaining energy from the electric field of the atmosphere

Abstract

The invention relates to electrical engineering and is intended for uninterrupted power supply of autonomous electrical equipment, for example, automatic weather stations or space probes. The technical result of the invention is the uninterrupted generation of energy from the electric field of the atmosphere in the required quantity. The device is easy, lack of moving parts, the simplicity of structural elements and the convenience of their transportation. The device contains electrodes, a supporting structure and ionizers of atmospheric gas. The electrodes are spaced along the lines of force of the electric field. Surrounded by the atmosphere, the electrodes are located on a supporting structure. The outer surface of these electrodes is structurally combined with ionizers of atmospheric gas. The lower electrode is a grounding electrode. Between the electrodes, Distances along the lines of force of the atmospheric electric field include the load. In the absence of contact of the device with the planet, the outer surface of all atmosphere-surrounded electrodes is structurally combined with ionizers of atmospheric gas. Current-carrying parts of the device, which are under voltage, are electrically isolated from the supporting structure. 2 yl.

Images (1)

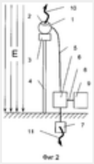

RU2245606C2

RU Grant

Find Prior Art Similar

Other languages: English

Inventor: P.A. The coachman (RU), P.A. Coachman, IN AND. Kolomiets (RU), IN AND. Kolomiets

Original Assignee: Kucher Pavel Alekseevich, Kolomiyets Valentin Ivanovich

Priority date : 2003-03-11

Family: RU (1)

Date	App/Pub Number	Status
2003-03-11	RU2003106714A	
2004-09-27	RU2003106714A	Application
2005-01-27	RU2245606C2	Grant

Info: Non-patent citations (2), Similar documents, Priority and Related Applications

External links: Espacenet, Global Dossier, Discuss

Claims (1)

1. A device for obtaining energy from the electric field of the atmosphere, including electrodes spaced along field lines, the lower of which is an earth electrode, a supporting structure with atmosphere-surrounded electrodes, ionizers of atmospheric gas structurally aligned with the outer surface of these electrodes, characterized in that between spaced apart field lines with electrodes, the load is switched on, in the absence of contact of the device with the planet, the outer surface of all atmosphere-surrounded electrodes about aligned with the ionization of atmospheric gas, the conductive portion of the device under voltage, electrically insulated from the supporting structure.

Description

The invention relates to electrical engineering and is intended for uninterrupted power supply of autonomous electrical equipment, for example automatic weather station or space probes on the Earth and other planets having an atmospheric electric field.

Known difficulties associated with the uninterrupted power supply of autonomous electrical equipment. Internal sources of energy (batteries, batteries, fuel cells) require periodic replacement, refueling or recharging, that is, manual, manual maintenance. External sources of energy (running water, wind, sunlight) require massive large structures, converters with moving parts and are not always available.

To solve the problem, we need a low-power, but reliable uninterrupted source of energy at any time and in any place. Taking into account the operating conditions of the autonomous electrical equipment, it must be lightweight, structurally simple and convenient for transportation to any distances.

There are known devices for the indirect generation of energy from the atmosphere by means of water wheels, turbines and other hydropower plants, as well as for direct energy production from the atmosphere by means of windmills, turbines and other windmills. In these devices, the energy of the atmosphere is converted into heat, work or electricity.

The continuity of the supply of energy generated with their help can not be ensured. It depends on the state of the atmosphere, climatic conditions and season. (Chaly G. Energy yesterday, today, tomorrow - Chisinau: Kartia Moldovenna, 1977. - 202 pp., Ill., Pp. 44-64,).

For the prototype a lightning rod is adopted, which is designed to neutralize the energy of the atmospheric electric field. The lightning rod contains a vertically oriented bearing structure that rises above the terrain, and electrodes spaced along field lines, connected by a current collector. At the top of the structure, there is an electrode surrounded by an atmosphere - a lightning receiver in the form of a massive metal rod. It can be combined with an ionizer of atmospheric gas. The lower electrode is a grounding electrode and provides electrical contact of the device with the planet's surface (VVBazutkin et al., High Voltage Engineering .- Moscow: Energoatomizdat, textbook for high schools, 1986. - 464 p., Ill., Pp. 219-220). The construction of the lightning rod does not include the payload.

The features of the prototype, which coincide with the essential features of the claimed invention, are as follows. The lightning conductor contains electrodes, a supporting structure and ionizers of atmospheric gas. The electrodes are spaced along the lines of force of the electric field. Surrounded by the atmosphere, the electrodes are located on a supporting structure. The outer surface of these electrodes is structurally combined with ionizers of atmospheric gas. The lower electrode is a grounding electrode.

The reasons for preventing the required technical result from the prototype are as follows. Lightning conductor ensures the flow through the atmosphere of electric current only in the form of a periodic spark discharge. The transit time of this current is unpredictable, and the magnitude is random and can not be adjusted. The lightning rod does not contain a payload. It is designed to work in an abnormally strong electric field and for the most part of time is uselessly idle. Lightning conductor does not work in places where there are no thunderclouds. The lightning rod does not work in the absence of contact with the planet. The live parts of the lightning rod do not have electrical insulation from the supporting structure.

The technical result is the uninterrupted reception of energy from the electric field of the atmosphere in the required quantity and the creation for this purpose of the claimed device.

The device is advantageous for its lightness, lack of moving parts, simplicity of necessary structural elements and convenience of their transportation. It is more reliable than all devices of autonomous power supply, including wind generators and solar panels, since the electric field of the atmosphere depends little on the time of the year, exists around the clock and is accessible anywhere in the world.

The technical result is achieved as follows. Between the electrodes spaced along the lines of force of the atmospheric electric field, a load is applied. In the absence of contact of the device with the planet, the outer surface of all atmosphere-surrounded electrodes is structurally combined with ionizers of atmospheric gas. Current-carrying parts of the device, which are under voltage, are electrically isolated from the supporting structure.

The essential features of the claimed invention are as follows. An apparatus for recovering energy from the atmospheric electric field comprises electrodes spaced along the electric field lines, the lower electrode is an earth electrode. The electrodes surrounded by the atmosphere are placed on a supporting structure. Ionizers of atmospheric gas are structurally combined with the outer surface of the atmosphere surrounded by an electrode.

Unlike the prototype between spaced along the electric field lines included electrodes load. In the absence of the device in contact with the outer surface of the planet electrodes structurally combined with the atmospheric gas ionizers. The current-carrying live parts of the device are electrically isolated from the supporting structure.

The farther apart the electrodes are separated, the higher the electric field strength near their surface and the greater the spreading rate of free charge carriers in the surrounding gas. Surrounded by the atmosphere, the electrodes are advantageously placed in the vertically spaced points of the supporting structure. Spreading of charge carriers in the atmosphere is facilitated by free movement of gas at the electrodes and the absence of other voltage concentrators around the device.

If there is a contact between the device and the planet, the lower electrode is an earth electrode. This provides a low electrical resistance to the passage of atmospheric electric current through the ground. If there is no contact of the device with the planet, the atmospheric electric current created by it can pass only through the gas discharge channel. In this case, all electrodes surrounded by atmosphere (both upper and lower) are structurally combined with ionizers of atmospheric gas. Thus, the uninterrupted conversion of the energy of the atmospheric electric field into the energy of the electric current flowing through the load is ensured. Further useful conversion of this energy into heat, work or electricity is accomplished by choosing the type of load (heater, electric motor or other electrical equipment).

The effect of the essential features of the claimed invention on the resulting technical effect is as follows. The supporting structure ensures the correct orientation of the device in the atmospheric electric field, which coincides with the direction of its lines of force. It keeps the electrodes spaced along the field lines at the required distance, provides mechanical strength of the device and integrates its individual parts into a single whole. Electrodes should be moved apart from each other for the maximum technically possible distance. Proportional to the distance increases the strength of the electric field on their outer surface, which increases the mobility of charge carriers in the surrounding atmosphere and facilitates the flow of electric current through it.

If there is a contact between the device and the planet, the lower electrode is an earth electrode. This is the simplest, cheapest and reliable way to provide low resistance to the current flowing through it into the ground. Surrounded by the atmosphere, the electrodes are mounted on a supporting structure. This ensures the stability of their spatial position in an external electric field. Ionizers of atmospheric gas are structurally combined with the outer surface of the atmosphere surrounded by an electrode. This ensures a constant presence of a sufficient number of free charge carriers in the space adjacent to the electrodes and the free movement of these charges along the lines of force of the electric field. Thus, an electrical contact is provided with a low resistance between the solid electrodes and the atmospheric gas. The strength of a current passing through such a contact within a wide range does not depend on the strength of the external electric field and is determined only by the capacity of the ionizers. This guarantees the uninterrupted availability of energy, regardless of external factors, minimizes the power consumption for maintaining the channel of a non-self-sustaining electrical discharge through a neutral atmosphere and prevents the occurrence of inrush currents through the device in conditions of thunderstorm activity.

If there is no contact of the device with the planet, the ionizers of the atmospheric gas are structurally combined with the outer surface of all the electrodes surrounded by the atmosphere, both upper and lower. This is the only available way to ensure the flow of electric current through the atmosphere on both sides of the device deployed along field lines of force.

The load is included between the electrodes spaced along the lines of force of the electric field. This ensures its serial connection (together with the electrodes) to the circuit of the current passing through the atmosphere of a non-self-sustaining electric discharge. The product of the voltage drop across the load and the amount of current passing through it determines the useful power received by the device from the atmospheric electric field. The coefficient of efficiency of the device is determined by the balance between the load resistance and the transient resistance of the electrodes with the surrounding space (ionizer capacity). The lower the transient resistance of the electrodes, the greater the proportion of energy released in the load.

The insulation of the current-carrying parts of the device from the supporting structure ensures an optimal arrangement of the electric field lines in the surrounding space and helps prevent the flow of electric current through the conductive parts of the device that are not related to its electrical circuit.

The uninterrupted supply of energy is provided by the fact that the electric field of the atmosphere does not depend on the time of the year, exists around the clock in all weathers, is accessible anywhere in the troposphere. This energy resource of the atmosphere is constantly fueled by the full power of the planetary mechanism for separating electric charges.

Reliability of energy supply is ensured by the simplicity of the device, the lack of moving parts in its most important elements. Obtaining energy from the electric field of the atmosphere does not require large structures (dams, towers of great height) and complex technological methods. The maintenance of such a device is much simpler than in analogues.

The simplicity of the design of the elements of the device for obtaining energy from the electric field of the atmosphere follows from the fact that it contains only standard electrotechnical units without movable parts that do not require adjustment, adjustment and regular maintenance. Parts of the device and parts of the supporting structure do not require careful manufacturing.

The ease of transporting the device is achieved by the fact that it is manufactured with a minimum margin of safety, since it does not experience dynamic loads and, therefore, it does not require massive, large-sized units and parts. The device is made foldable or collapsible.

1 shows a general view of an apparatus for generating energy from the electric field of the atmosphere in the absence of contact with the planet.

The device comprises a vertically oriented support structure, for example a balloon, 1, electrodes 2, 3, air ionizers 4 and 5, insulators 6 and 7, cables 8 and 9, voltage converter 10 and a load, for example a radiosonde 11. In the surrounding balloon space, lines of the atmospheric electric field E. Electrodes, structurally combined with air ionizers, are attached to insulators. The balloon keeps the electrodes at a sufficient distance from each other, and the cables connect them to the voltage converter. The voltage converter is connected to the payload electrically and is connected by a common housing.

The device works as follows. After switching on the device, the ionizers 4, 5 saturate the air surrounding the electrodes 2, 3 with free charge carriers. They start drifting through neutral air, moving along the lines of force of the atmospheric electric field E, additionally amplified by the spaced position of the electrodes on the balloon 1. The leakage of charge carriers from the electrodes is compensated by the constant work of the ionizers. A constant electric current flows through the channels of the non-self-sustained gas discharge 12 and 13 from the electrodes through the atmosphere. A difference of potentials appears between the spaced electrodes. They keep it due to insulators 6 and 7. After the appearance of the working difference of potentials (≈5 kV) on the electrodes, the voltage converter 10 switches on. Atmospheric electric current is closed through it via cables 8 and 9.

Determine the electrical power required for autonomous uninterruptible power supply of the payload. For example, to operate the equipment of a modern radiosonde, suspended to an aerostat, a constant electric power of 10 W is sufficient. Determine the limiting electrical power necessary to supply auxiliary devices and air ionizers. For example, in an amount not exceeding 150% of the useful value, that is, 15 watts. Taking into account operating conditions of current collectors, the limiting potential difference between open live parts of the installation is determined. For example, no more than 10 kV is recommended, but actually 5 kV is chosen. The maximum current in the discharge channel is calculated. In this case, no more than (10 W + 15 W) / 5 kV = 5 mA. This value sets the productivity of ionizers, structurally combined with the electrodes. The density of the vertical conduction current and the field strength in the troposphere at the calculated flight altitude are determined. For example, the conduction current is not more than 1.5 μA / m^2, the electric field strength is 2.2-3.5 V / m. Calculate the distance between the electrodes, their shape and design, ensuring efficient spreading of the charge carriers created by ionizers in the atmosphere. For example, the distance is 50 m, the shape of the receiving electrodes is a ball, the area of each is not less than 0.5 m^2. They fix the electrodes with insulators on the balloon and the suspension system, lay cables, mount a power converter in the container with the equipment.

In Fig. 2, a general view of the device for obtaining energy from the electric field of the atmosphere when the device is in contact with the planet is presented.

The device comprises an electrode 1, an air ionizer 2, an insulator 3, a supporting structure 4, a cable 5, a voltage converter 6, a ground connection 7, a connecting cable 8 and a payload (weather station) 9. In the surrounding space, electric field E. The electrode, structurally combined with the ionizer of air, is attached to the insulator. The stand keeps the electrode at a sufficient height, and the cable connects it to the voltage converter. The voltage converter is electrically connected to the ground through the ground and with the payload through the connecting cable.

The device works as follows:

After switching on the device, the ionizer 2 saturates the air surrounding the electrode 1 with free charge carriers. They begin to drift through neutral air, moving along the lines of the atmospheric electric field E, additionally amplified by the elevated position of the electrode on the column 4. The leakage of the charge carriers from the receiving electrode is compensated by the constant work of the ionizer. A constant electric current flows through the channel of the non-self-sustained gas discharge 10 from the electrode through the atmosphere. The electrode acquires an electrical potential with respect to the ground surface and retains it due to the insulator 3. After the working potential (≈25 kV) appears on the electrode, a voltage converter 6 switches on. The atmospheric electric current closes through it via cable 5 to ground 7.

Determine the electrical power required for autonomous uninterruptible power supply of the payload. For example, for the operation of a modern automatic meteorological station, a constant electric power of 100 W at a nominal voltage of 27 V is sufficient. The limiting electrical power necessary for supplying auxiliary devices and ionizers of air is determined. For example, in an amount not more than 50% of the useful, that is 50 watts. From the climatic conditions and design features of the current collector determine the limiting potential difference between the open live parts of the installation. For example, no more than 30 kV is recommended, but actually 25 kV are chosen. The maximum current in the discharge channel is calculated. In this case, no more than (100 W + 50 W) / 25 kV = 6 mA. This value specifies the productivity of the ionizer, which is structurally combined with the electrode. The electrical activity of the atmosphere (the density of the vertical conduction current and the field strength) in the given area is measured or found from tables. For example, the conduction current is not more than 0.1 μA / m^2, the electric field strength is 110-250 V / m. Calculate the height of the rise of the electrode over the terrain, its shape and working surface, sufficient for efficient spreading of charge carriers created by the ionizer in the near-earth atmosphere. For example, the height is not less than 10 m, the shape is a ball, the area is not less than 1 m^2. The calculations take into account the wind speed, the presence of other voltage concentrators, the geological structure of the ground and other significant factors. Establish the required height of the rack with an electrode on the insulator, arrange grounding, mount auxiliary electrical equipment and power converter.

Thus, uninterrupted energy supply of autonomous electrical equipment is ensured due to the free force of the atmospheric electric field. The device is advantageously distinguished by the lack of moving parts, the simplicity of the necessary structural elements and the ease of their transportation. It is more reliable than all the devices of autonomous power supply, including wind generators and solar panels, since the electric field of the atmosphere depends little on the time of the year, exists around the clock and is accessible anywhere in the world.

Non-Patent Citations (2)

Title
A.P.MERKULOV. Magnetic fields are workers. - M .: Mechanical Engineering, 1978, p. 142-144. *
VVBAZUTKIN and others. The technique of high stresses. - Moscow: Energoatomizdat. Textbook for high schools. 1986, p. 219-220. *

* Cited by examiner, † Cited by third party

Similar Documents

Publication	Publication Date	Title
US7469541B1	2008-12-30	Portable power system
US7105940B2	2006-09-12	Mobile renewable energy generator
US4257823A	1981-03-24	Direct conversion solar electric generator
US20110062717A1	2011-03-17	Urility grid vertical axis wind turbine system
US6974110B2	2005-12-13	Method and apparatus for converting electrostatic potential energy
US7478712B2	2009-01-20	Energy collection
US20110100433A1	2011-05-05	System and method for grounding photovoltaic modules
US5942806A	1999-08-24	Method and device for generating electricity
Kauranen et al.	1994	Development of a self-sufficient solar-hydrogen energy system
Fritts	1967	Special applications of thermoelectric generators
GB2076036A	1981-11-25	Elevated housing for equipment
DE4205521A1	1993-04-01	Recording storage system for lightning storage system
US8006933B2	2011-08-30	Airborne power station
GB2449102A	2008-11-12	Urban power generation platform for use on street lampposts
Djafour et al.	2011	Photovoltaic-assisted alkaline water electrolysis: Basic principles
Al-Ismaily et al.	1998	Photovoltaic electricity prospects in Oman
CN201562024U	2010-08-25	Test device for gas insulated metal-enclosed transmission line with UHV AC/DC
Baer et al.	1976	420 kV SF 6-insulated tubular bus for the wehr pumped-storage plant electric tests
Gomaa et al.	1995	Design and analysis of photovoltaic and wind energy hybrid systems in Alexandria, Egypt
WO2013030897A1	2013-03-07 SYSTEM	VOLTAGE STABILIZING APPARATUS
EP2107617A1	2009-10-07	Architectural covering with photovoltaic modules
WO2011098906A2	2011-08-18	Electric generator pole device
US20120118353A1	2012-05-17	Electrical grid solar energy harvesting system
US8045314B2	2011-10-25	Method of atmospheric discharge energy conversion, storage and distribution
US20120260967A1	2012-10-18	Photovoltaic array utilizing phyllotaxic architecture

Priority And Related Applications

Priority Applications (1)

Application	Priority date	Filing date	Title
RU2003106714A	2003-03-11	2003-03-11	Device for obtaining energy from the electric field of the atmosphere

Applications Claiming Priority (1)

Application	Filing date	Title
RU2003106714A	2003-03-11	Device for obtaining energy from the electric field of the atmosphere

About Send Feedback Public Datasets Terms Privacy Policy

United States Patent [19]

Yukl

[11] **Patent Number: 4,494,009**

[45] **Date of Patent: Jan. 15, 1985**

[54] **METHOD AND APPARATUS FOR CAPTURING AN ELECTRICAL POTENTIAL GENERATED BY A MOVING AIR MASS**

[76] Inventor: **Tex Yukl**, Pleasant Valley Rte., Box 21, Baker, Oreg. 97814

[21] Appl. No.: **533,237**

[22] Filed: **Sep. 19, 1983**

[51] **Int. Cl.**³ **F03D 5/04;** H02P 9/04

[52] **U.S. Cl.** **290/55;** 290/43; 290/44; 290/54; 310/308

[58] **Field of Search** .... 290/42, 43, 44, 45, 290/52, 53, 54, 55; 416/132 B; 310/10, 11, 308, 309, 310; 60/202; 322/2 R, 2 A

[56] **References Cited**

U.S. PATENT DOCUMENTS

28,793	6/1860	Uion	310/308 X
414,943	11/1889	Dewey	310/308 X
911,260	2/1909	Pennock	310/308 X
1,014,719	1/1912	Pennock	310/308 X
2,239,693	4/1941	Benett	310/308 X
2,279,586	2/1939	Benett	310/308 X
4,146,800	3/1979	Gregory et al.	290/44
4,206,396	6/1980	Marks	290/44 X
4,302,684	11/1981	Gogins	290/44 X
4,433,248	2/1984	Marks	290/55 X

Primary Examiner—J. V. Truhe
Assistant Examiner—Paul Shik Luen Ip
Attorney, Agent, or Firm—Kolisch, Hartwell, Dickinson & Anderson

[57] **ABSTRACT**

A method and apparatus for making available an electrical current, which comprises placing a conductive airfoil in a moving air mass, sinking a conductive element subsurface, and insulating the airfoil from the conductive element and producing an electrical potential between the airfoil and the element in a moving air mass.

4 Claims, 5 Drawing Figures

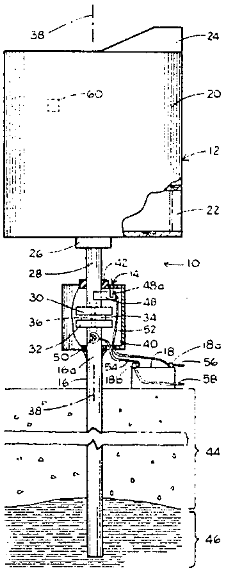

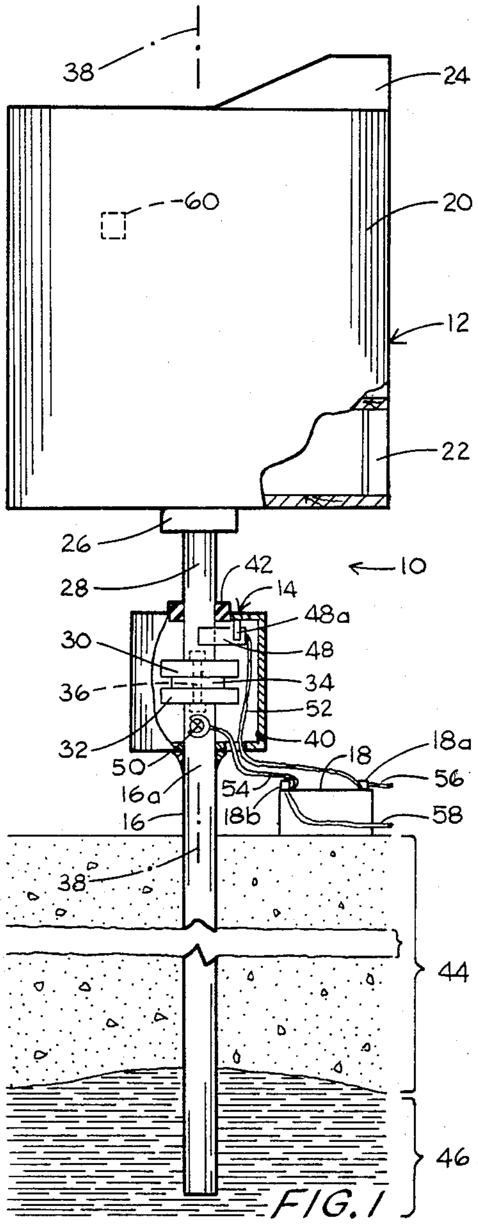
38
24
60
20
12
22
26
28
42
10
14
48a
48
30
34
36
52
32
40
18
18a
50
56
16a
54
18b
58
16
38
44
46

FIG. 1

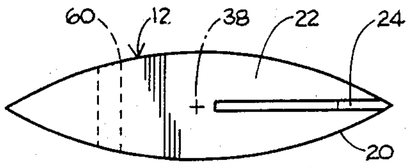
60
12
38
22
24
20

FIG. 2

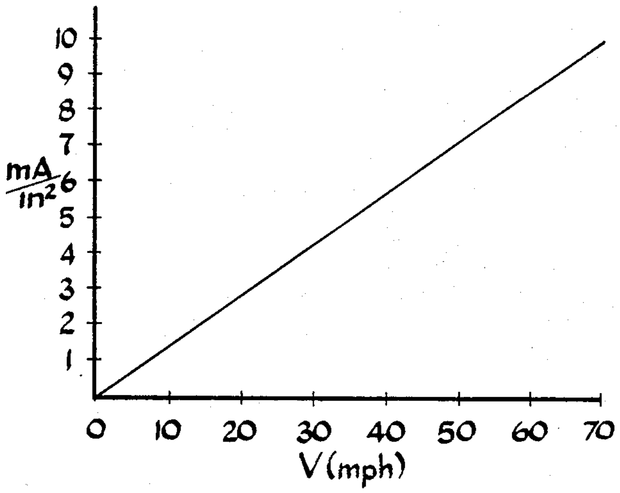

FIG. 5

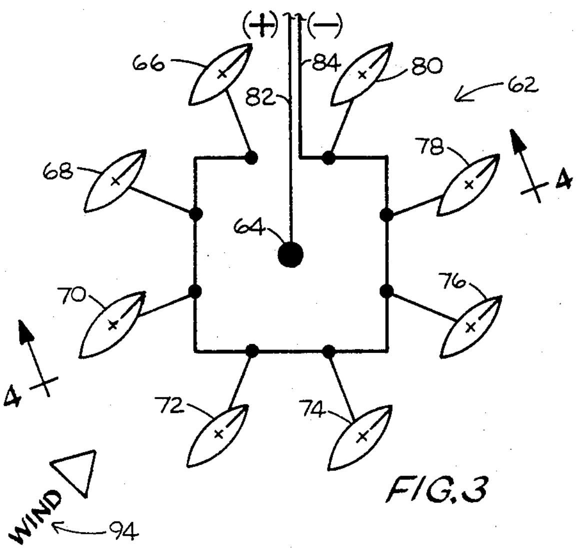
(+)
(−)
66
84
80
82
62
78
4
68
64
70
76
4
72
74
WIND
94

FIG. 3

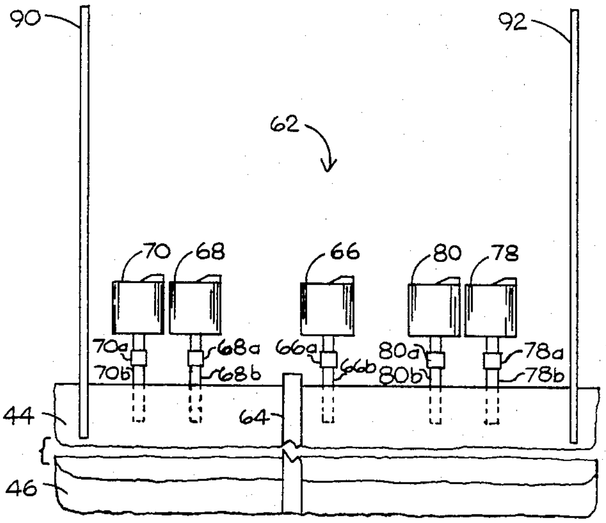
90
92
62
70
68
66
80
78
70a
70b
68a
68b
66a
66b
80a
80b
78a
78b
44
64
46

FIG. 4

METHOD AND APPARATUS FOR CAPTURING AN ELECTRICAL POTENTIAL GENERATED BY A MOVING AIR MASS

BACKGROUND AND SUMMARY OF THE INVENTION

The instant invention relates to a method and apparatus for capturing an electrical potential generated by a moving air mass. Specifically, the instant invention describes a method and apparatus for capturing an earth-air static electrical potential which is produced by an air mass moving over a stationary airfoil.

Known apparatus and methods of producing electricity from a moving air mass utilize windmills and windmill-driven turbines. Such devices are generally mounted on fairly high towers and require complex propeller systems to effectively drive electrical turbines. The systems also tend to have a rather large affect on the air mass as it passes over the blades of a particular windmill. As with the placement of any artificial barriers in an air mass, the air mass is disturbed, and turbulence is induced into the air mass. As a particular particle of the air mass reaches the next windmill, the induced turbulence results in less driving force on the blades of the secondarily driven windmill.

The present invention does not have a propeller system, nor does it induce significant turbulence into a moving air mass. It is well documented that an object, particularly an airfoil, which is moving relative to an air mass, develops a substantial static electrical charge on its surface relative to the air mass as the result of friction between the airfoil and the air mass. This naturally occurring static charge is particularly strong when the airfoil is a conductive surface, such as a metal-covered aircraft wing. For this reason, virtually all metal-covered aircraft employ static discharge strips along the trailing edges of airfoil surfaces in order to dispel any accumulated static electrical charge.

An object of the instant invention is to collect electrical energy from movement of an air mass over an airfoil.

Another object of the instant invention is to collect electrical energy with a non-complex collecting element.

A further object of the instant invention is to collect electrical energy from movement of an air mass over an airfoil with a substantially non-turbulence producing system.

A further object of the instant invention is to describe a method of collecting electrical energy from the frictional effects of an air mass moving over a conductive airfoil.

The electrical system of the instant invention includes a conductive airfoil, which is rotatably mounted on an insulated bearing, the airfoil being free to align itself with a moving air mass. A second component of the system is a ground rod which extends down to a water table level. Electrical leads are connected to a conductive portion of the airfoil and to the ground rod, and are then connected to a storage battery.

The method of collecting electrical energy from the system involves connecting the negative side of the storage battery to the airfoil and connecting the positive side of the battery to the ground rod. As an air mass moves over the surface of the airfoil, an electrical charge is produced and transmitted to the storage battery. Electrical energy may be drawn from the storage battery as needed.

These and other objects and advantages of the instant invention will become more fully apparent as the detailed description of the invention which now follows is read in conjunction with the drawings.

DESCRIPTION OF THE DRAWINGS

FIG. **1** is a front elevation of an earth-air electrical current friction generator constructed according to the instant invention, with portions broken away to show interior detail.

FIG. **2** is a top plan view of the generator of FIG. **1**.

FIG. **3** is a schematic plan view of a generator field constructed according to the instant invention.

FIG. **4** is a view of the generator field of FIG. **3**, taken generally along line **4**—**4** in that figure.

FIG. **5** is a graph depicting the relationship between current production and wind speed associated with the generator of FIG. **1**.

DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT

Turning now to the drawings, and in particular FIG. **1**, an earth-air electrical current friction generator made according to my invention is depicted generally at **10**. The generator includes a metal airfoil, or conductive airfoil means, shown generally at **12**, which is supported by a rotatable airfoil mount **14** and an electrically conductive electrode **16**. A battery **18** provides storage capability for the collecting apparatus.

Airfoil **12** is vertically mounted above ground level. The airfoil is a zero-lift structure in the preferred embodiment, since both active surfaces are symmetrical. Airfoil **12** is formed of a conductive metal skin **20**, formed of lightweight aluminum sheeting in the preferred embodiment, over a frame **22** which is shown behind the cut-away portion of skin **20**. Frame **22** may be formed of any suitable material, such as wood, plastic or metal. In the preferred embodiment, frame **22** is a substantially square structure.

A wind vane **24** is mounted on the top and at one side of airfoil **12**. Wind vane **24** is also referred to as an air mass alignment means, and provides that airfoil **12** will be windfollowing.

An airfoil support bracket **26** is attached to the base of airfoil **12**. One end of a support post **28** is received within bracket **26**. The other end of post **28** terminates in a non-conductive airfoil bearing race **30** and is received within mount **14**. Bracket **26** and post **28** are electrically conductive and are electrically connected to skin **20**.

Rotatable airfoil mount **14**, or rotatable airfoil mounting means, electrically insulates airfoil **12** from electrode **16**, and is referred to herein as isolation means or isolator means.

Airfoil mount **14** is, itself, supported by an above-ground portion **16*a*** of electrode **16**. Portion **16*a*** terminates in a non-conductive electrode bearing race **32**. A bearing **34** is interposed race **30** and race **32**. The respective airfoil and electrode components are retained together by means of a fastener **36**, which is a nut and bolt combination in the preferred embodiment. Fastener **36** aligns with an airfoil axis of rotation **38**.

Mount **14** is enclosed by a cylindrical enclosure **40** which is secured to portion **16*a*** and which allows post **28** to rotate within a weathertight seal **42**. Enclosure **40** and seal **42** are, of course, electrically non-conductive.

Electrode **16** extends through whatever ground layers exist beneath the surface, which are depicted generally at **44**, and then extends into a permanent water table layer **46**. Extending electrode **16** into a water table provides an electrical surface which is as large as the water table. Although the generator would function if electrode **16** did not extend into a water table, the operation of the generator is greatly enhanced by extending electrode **16** to a water table, for reasons which will be explained later herein.

An electrical commutator **48**, and a grounding screw **50**, make electrical contact with post **28** and portion **16***a*, respectively. Commutator **48** is insulated from and fastened to enclosure **40** by bracket **48***a*. Commutator **48** is connected to a lead line **52** which is attached to battery negative terminal **18***a*, and screw **50** is attached to lead line **54** which is attached to battery positive terminal **18***b*. The commutator, screw and lead line comprise what is referred to herein as means for connecting, or connecting means.

Battery **18**, although represented in FIG. **1** as a solitary unit, may obviously include a plurality of units, connected in series or parallel, depending on the output of generator **10**. A pair of leads **56**, **58**, are shown coming off of battery **18**, and would presumably be attached to an electrical distribution system, possibly through an A.C. inverter. The battery and leads **56**, **58** compise an electrical system.

As an air mass moves over skin **20** of airfoil **12**, there occurs an exchange of ions between the air mass and the surface of the airfoil. This exchange takes place as an air mass moves over any stationary object. In most instances, however, the object is sitting on the surface of the earth, in electrical communication therewith, and a charge does not accumulate on the surface of the object. Since the airfoil of the instant invention is insulated from the earth, a static charge builds up on the surface of the airfoil as an air mass flows over the airfoil.

As the air mass moves over airfoil **12**, a collection of negatively charged ions builds up on the airfoil surface. Were there no electrical connection whatsoever between airfoil **12** and ground surface **44**, and were it not for the fact that eventually a charge built up on airfoil **12** would dissipate into the air mass, a charge would eventually build up on airfoil **12** of sufficient strength to arc, such as thc arc produced between a thunderstorm and the ground in the case of lightning, resulting in the complete discharge of the negative ions accumulated on airfoil **12**.

The electrical charge which accumulates on airfoil **12** is drawn off of the airfoil by means of commutator **48**, which is attached to battery **18**. A corresponding positive charge is drawn off of electrode **16** by means of screw **50**.

The surface area of airfoil **12** is defined by what is referred to herein as a cross-section square inch. This unit is depicted at **60** in FIGS. **1** and **2**. The unit is measured on both sides of the airfoil. The distance between the sides of the airfoil is not a factor. So long as the surface area in electrical contact with electrode **16** equals or exceeds the surface area of airfoil **12**, the generator will produce a current flow of 1-milliamp-per-square-inch of airfoil surface cross section with a wind speed of 7-miles-per-hour.

Because the airfoil collects negative ions, it is sometimes referred to as a negative collector element. Likewise, since electrode **16** acts as a source of a positive charge, it is sometimes referred to as a positive collector element, or conductive earth means.

Turning now to FIGS. **3** and **4**, a friction generator farm **62** is depicted schematically in FIG. **3**, and in a pictorial cross section side view in FIG. **4**. A generator farm, according to the instant invention, would include at least one conductive electrode **64**, placed in the ground down to the level of a permanent water table, which would act as a positive collector element. A plurality of airfoils is depicted by even numbers **66** through **80**. Electrode **64** is connected to a positive side of an electrical system by a conductive line **82**. Airfoils **66** through **80** are connected to a common negative lead **84** which connects to the negative side of an electrical system.

Turning now to FIG. **4**, a side sectional view of generator farm **62** is shown. Electrode **64** is placed in ground, represented by **44** as are the support posts for the airfoils. It should be noted that in this type of arrangement, a single commutator would be required within airfoil mounts **70***a*, **68***a*, **66***a*, **80***a* and **78***a*. An airfoil structure support post, represented at **70***b*, **68***b*, **66***b*, **80***b* and **78***b* may be formed from an electrically insulative material when a single electrode, like that shown at **64**, is utilized, thereby further simplifying the internal portions of the airfoil mount.

In addition to earth-air friction generators, a pair of lightning rod devices **90**, **92**, are shown in FIG. **4**. Rods **90** and **92** also are extended subsurface to provide protection from lightning strikes in the vicinity of generator farm **62**. A generator farm of the type described would be particularly susceptible to damaging lightning strikes were such lightning rods not provided.

Returning for a moment to FIG. **3**, a wind arrow **94** is depicted to show a prevailing wind and direction of air mass movement. It will be noted that airfoils **66** through **80** have aligned themselves with the wind indicated by arrow **94**. It should also be noted that with a generator farm in the layout proposed in FIG. **3**, there would be very little disturbance of wind flow by the airfoils of the generator farm. This is in extreme contrast to the turbulence and generally delaminating effect of conventional windmill generating farms.

It should also be obvious to those skilled in the art that the construction of the generator of the instant invention is many orders of magnitude less complex than the construction required for a conventional windmill-type turbine generator.

Turning now to FIG. **5**, a graph of current output per crosssection square inch mA/in^2 versus wind speed V (mph) is depicted. The above-described generator will produce 1-milliamp of electrical current per cross-section square inch with a 7-mile-per-hour wind. Generator output increases linearly, such that the generator will produce 10-milliamps per cross-section square inch in a 70-mile-per-hour wind. It should also be noted that very few known windmill turbine generators function in a 7- or 70-mile-per-hour wind. The generator of the instant application does not have such a limitation. The generator will produce current in any amount of wind, light or strong. As long as the structure is sufficiently durable to withstand a high wind, the system will continue functioning and producing an electrical current. Most systems are restricted to operations over a set wind speed range, generally in the 10- to 45-mile-per-hour range. Systems which produce electricity in extremely light wind conditions are very costly.

Assume for a moment, that an individual airfoil is constructed with sides which are 10-feet in length, thereby producing a cross section dimension on the order of 100-square feet, and further assume that ten airfoils are arranged similar to generator farm **62**, thereby resulting in a total cross section area of 1,000-square feet. The generator farm so described, when connected to a 12-volt D.C. system, and when excited by a wind of 7-miles-per-hour, will produce 1,728-watts of electrical energy.

Thus, a new method of capturing a usable electric current, and an apparatus for practicing the method described have been disclosed. While a preferred embodiment of the invention has been described, it is appreciated that variations and modifications may be made without departing from the spirit of the invention.

It is claimed and desired to secure by Letters Patent:

1. An earth-air electrical current friction generator which comprises

conductive earth means extending subsurface into a water table,

conductive airfoil means operatively associated with said earth means extending above surface into an air mass responsive to airflow to produce an electrical potential relative to said earth means,

isolation means electrically insulating said airfoil means from said earth means, said isolation means including a rotatable airfoil mounting means which allows said airfoil means to rotate in such an air mass and which supports said airfoil means above surface,

an electrical system, and

means for connecting said airfoil means and said earth means to said electrical system for conduction thereto of an electrical current.

2. An electrical charge collector which comprises

a positive collector element extending to a water table,

a negative collector element operatively associated with said positive collector element, said negative collector element being wind-following and including wind alignment means for aligning said negative collector element with a prevailing wind,

isolator means for insulating said negative collector element electrically from said positive collector element, said isolator means including a rotatable mounting which rotatably supports said negative collector element, and

connecting means for providing an electrical connection from each of said elements to an electrical system.

3. The collector of claim **2**, wherein said positive collector element includes at least one conductive electrode.

4. The collector of claim **2**, wherein said negative collector element includes at least one conductive airfoil.

* * * * *

US006116544A

United States Patent [19]

Forward et al.

[11] **Patent Number: 6,116,544**

[45] **Date of Patent: Sep. 12, 2000**

[54] **ELECTRODYNAMIC TETHER AND METHOD OF USE**

[75] Inventors: **Robert L. Forward**, Houston, Tex.; **Robert P. Hoyt**, Seattle, Wash.

[73] Assignee: **Tethers Unlimited, Inc.**, Seattle, Wash.

[21] Appl. No.: **08/929,271**

[22] Filed: **Sep. 12, 1997**

[51] **Int. Cl.**^{7} **B64G 1/00**

[52] **U.S. Cl.** .... **244/158 R**; 244/166; 244/172

[58] **Field of Search** .... 244/158 R, 164, 244/166, 172

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,097,010	6/1978	Colombo et al.	244/158 R
4,824,051	4/1989	Engelking	244/158 R
4,923,151	5/1990	Roberts	244/158 R

OTHER PUBLICATIONS

"Tethers Open New Space Options" by Ivan Bekey, *Astronautics & Aeronautics*, 1983.

Primary Examiner—J. Woodrow Eldred
Attorney, Agent, or Firm—Arthur M. Dula

[57] **ABSTRACT**

The present invention comprises an electrodynamic tether structure and a method of use. The structure of the tether taught by the present invention is a short, wide, interconnected multiwire (compared to the long, narrow single wires of the prior art) conductive tether whose area maximizes electrodynamic drag while simultaneously minimizing the Area-Time-Product swept by the tether during its operating life. The preferred tether length is two kilometers to five kilometers. The preferred tether mass is one percent to five percent of the spacecraft mass. The method of operation comprises orienting the tether structure at an angle to the local vertical to maximize electrodynamic drag on the host spacecraft and minimize tether instability. The angle of 35.26 degrees is preferred.

26 Claims, 12 Drawing Sheets

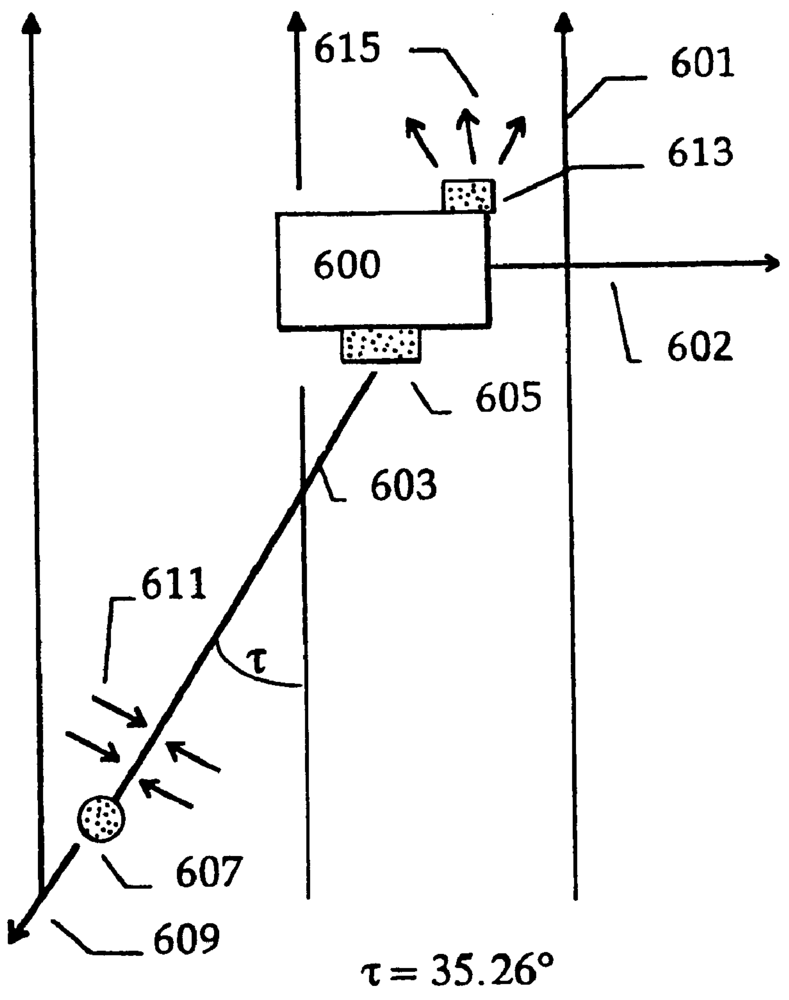

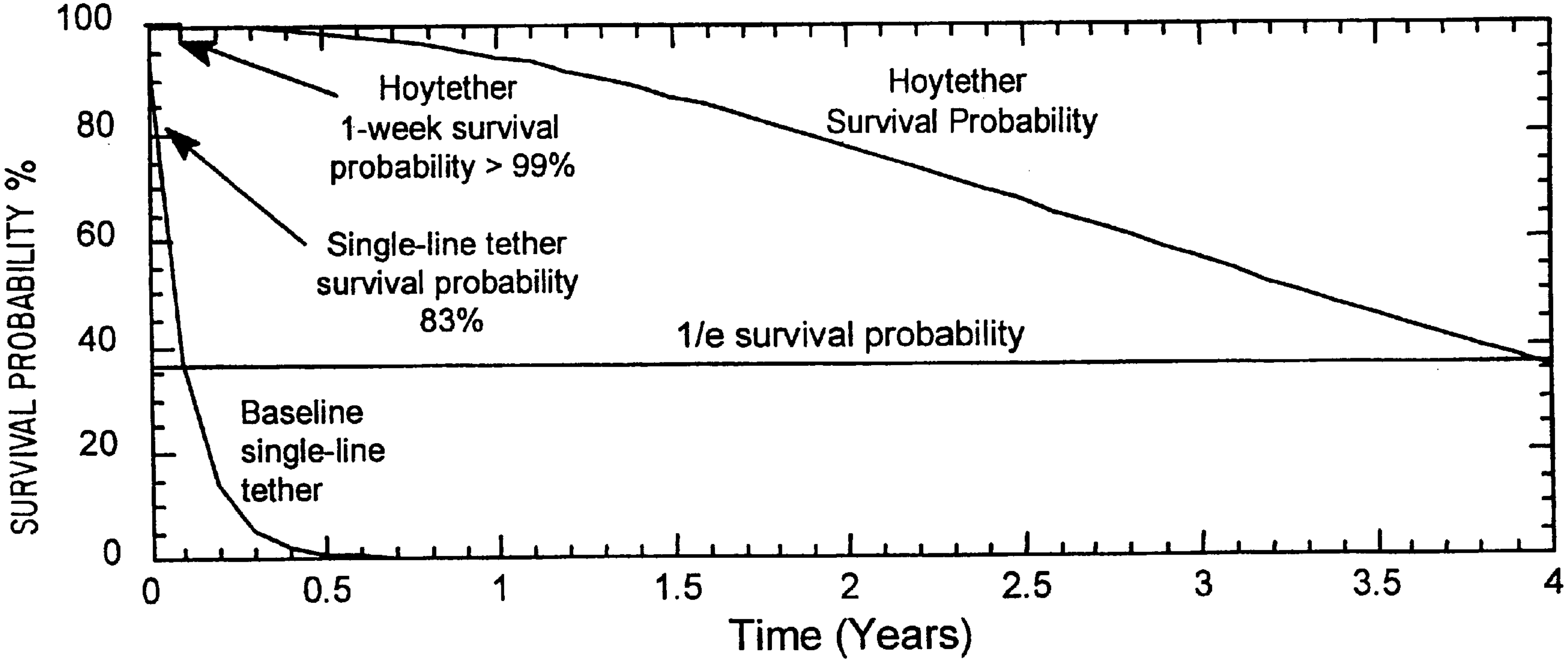
100
80
60
40
20
0
SURVIVAL PROBABILITY %
0
0.5
1
1.5
2
2.5
3
3.5
4
Time (Years)
Hoytether
1-week survival
probability > 99%
Single-line tether
survival probability
83%
Hoytether
Survival Probability
1/e survival probability
Baseline
single-line
tether

FIG. 1a

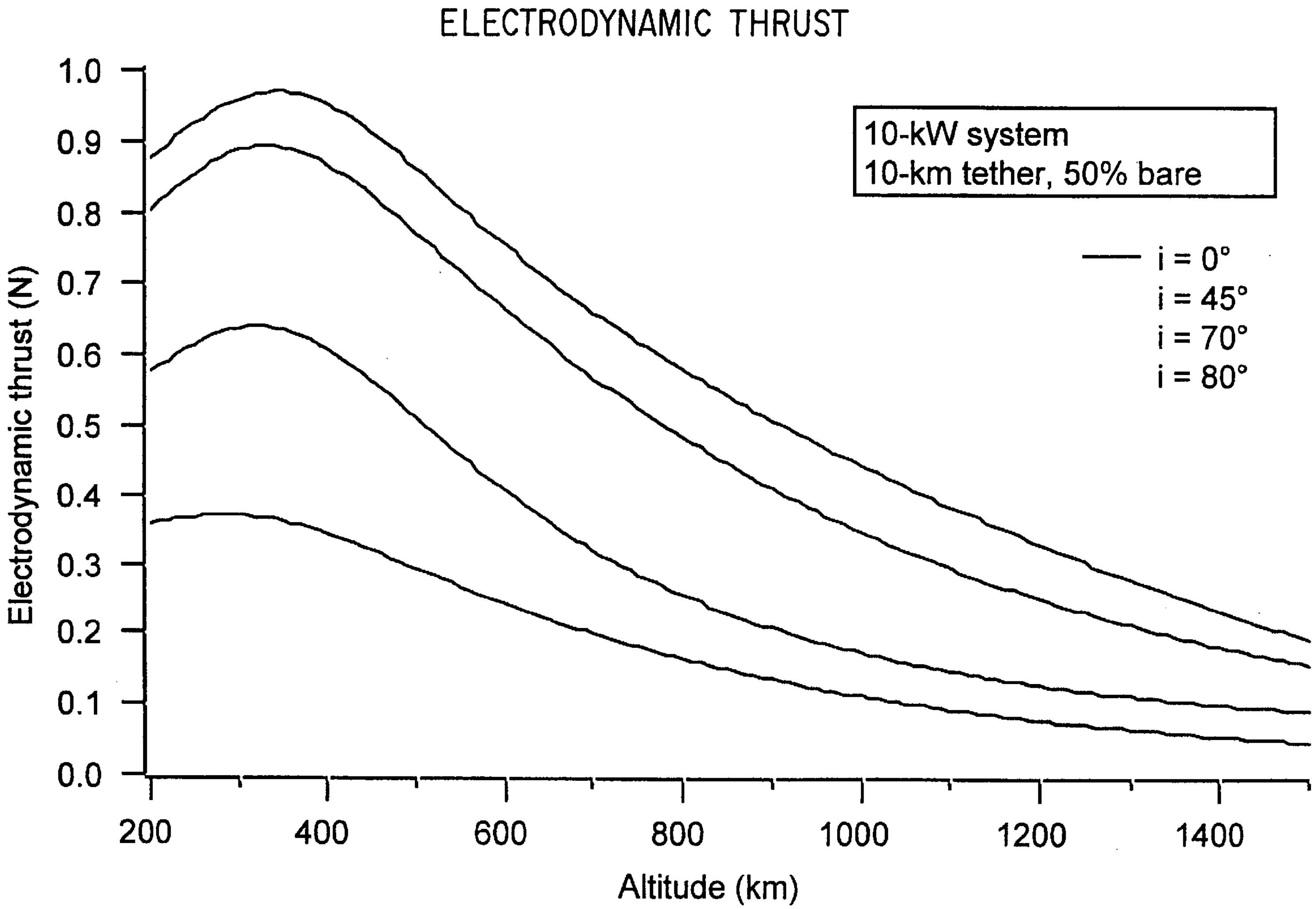
ELECTRODYNAMIC THRUST
10-kW system
10-km tether, 50% bare
i = 0°
i = 45°
i = 70°
i = 80°
Electrodynamic thrust (N)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
200
400
600
800
1000
1200
1400
Altitude (km)

FIG. 1b

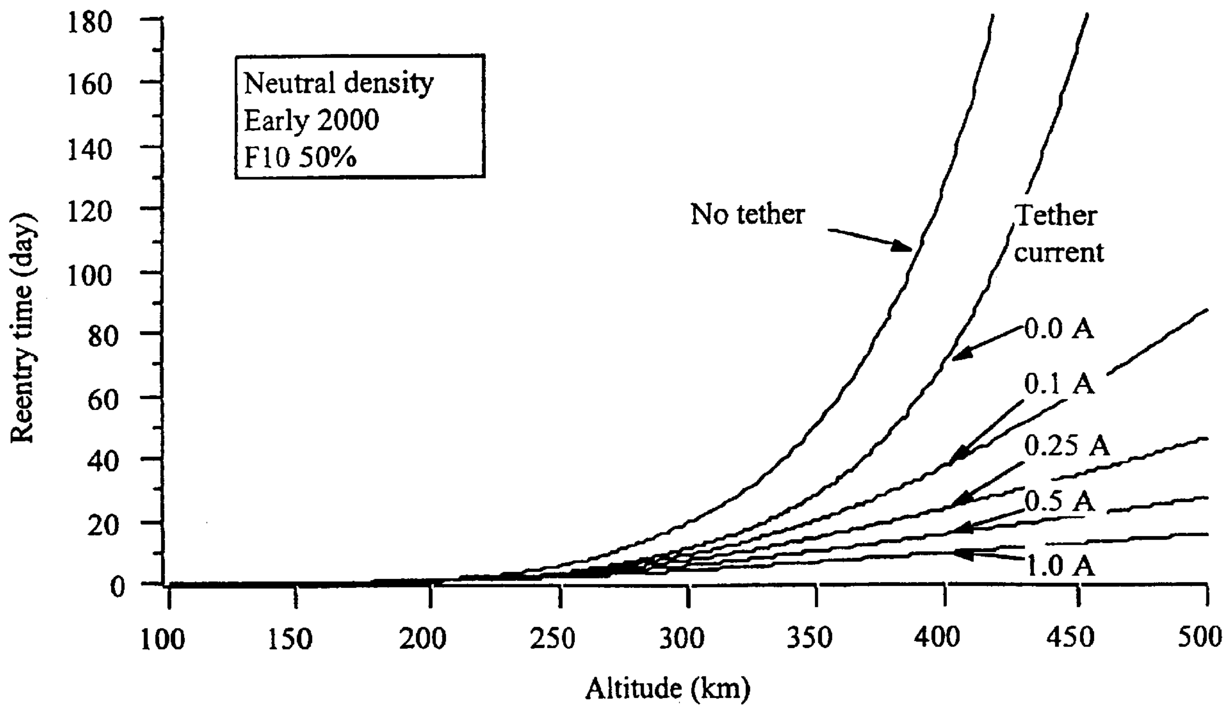

Reentry Time Sensitivity
Unpowered 5 km ProSEDS Tether

Enrico Lorenzini, SAO, NASA ProSEDS Mission (July 1997), NASA/MSFC Grant NAG8-1303

FIG. 1c

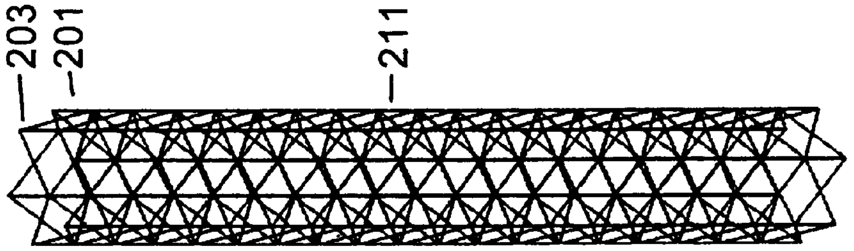

FIG. 2a

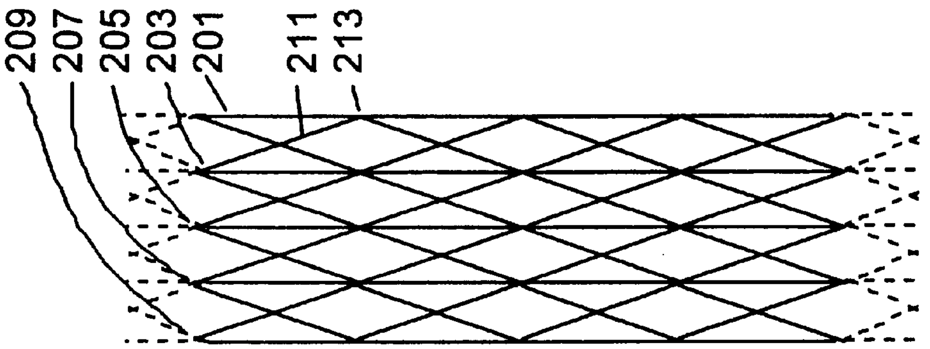
209
207
205
203
201
211
213

FIG. 2b

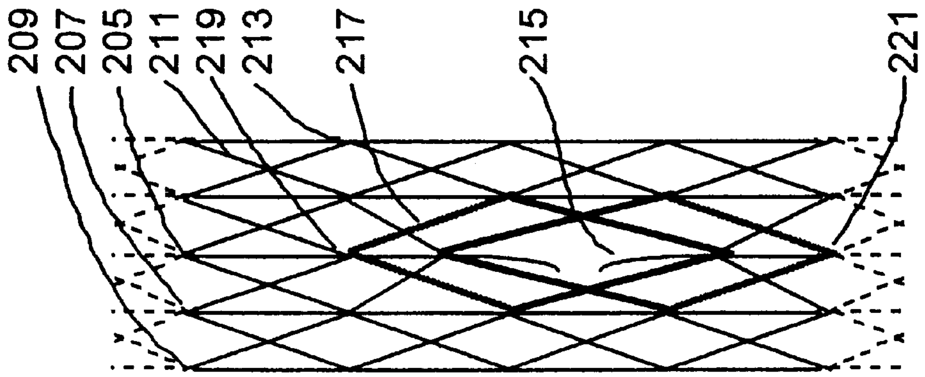
209
207
205
211
219
213
217
215
221

FIG. 2c

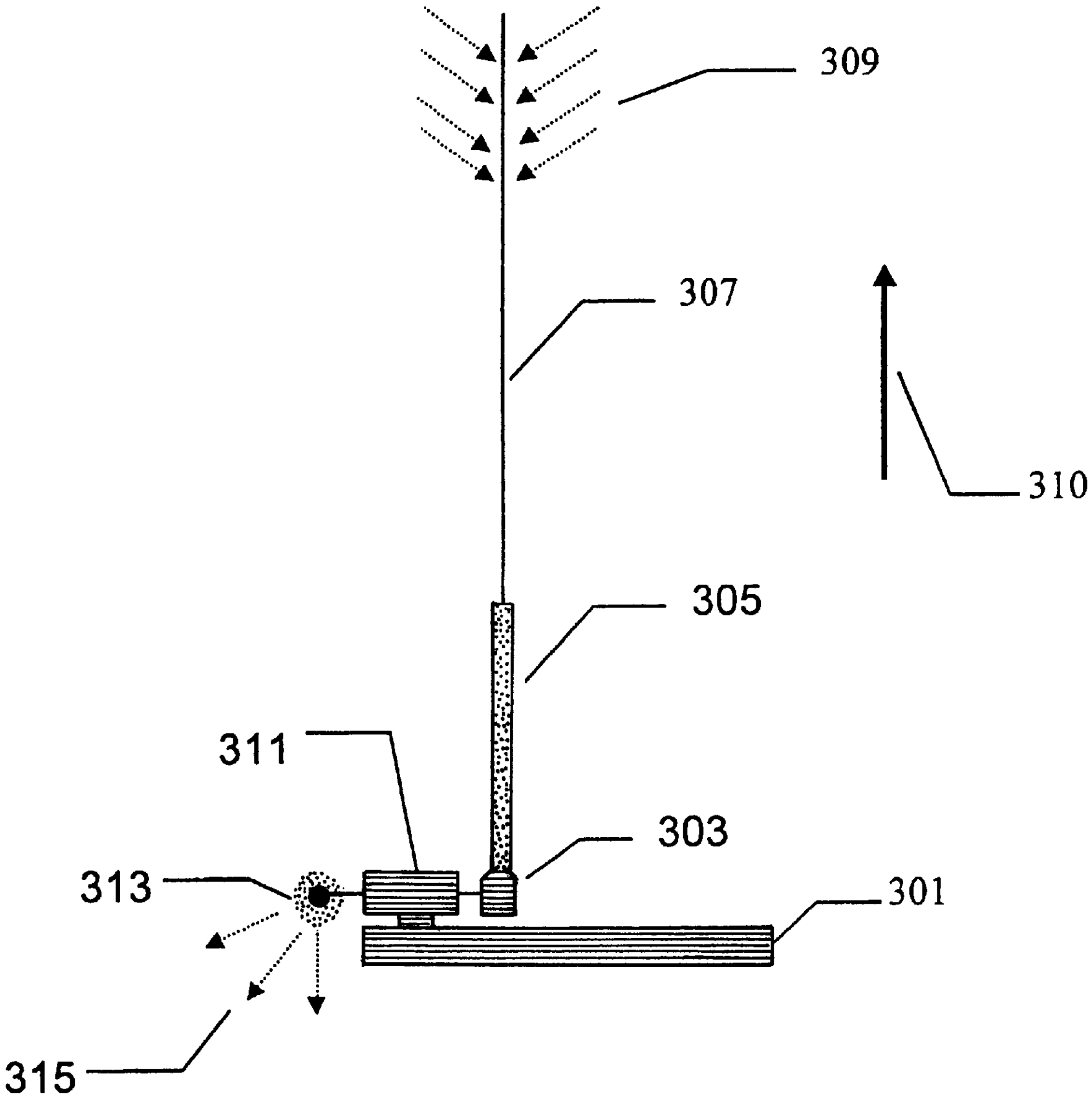
309
307
310
305
311
303
313
301
315

FIG. 3

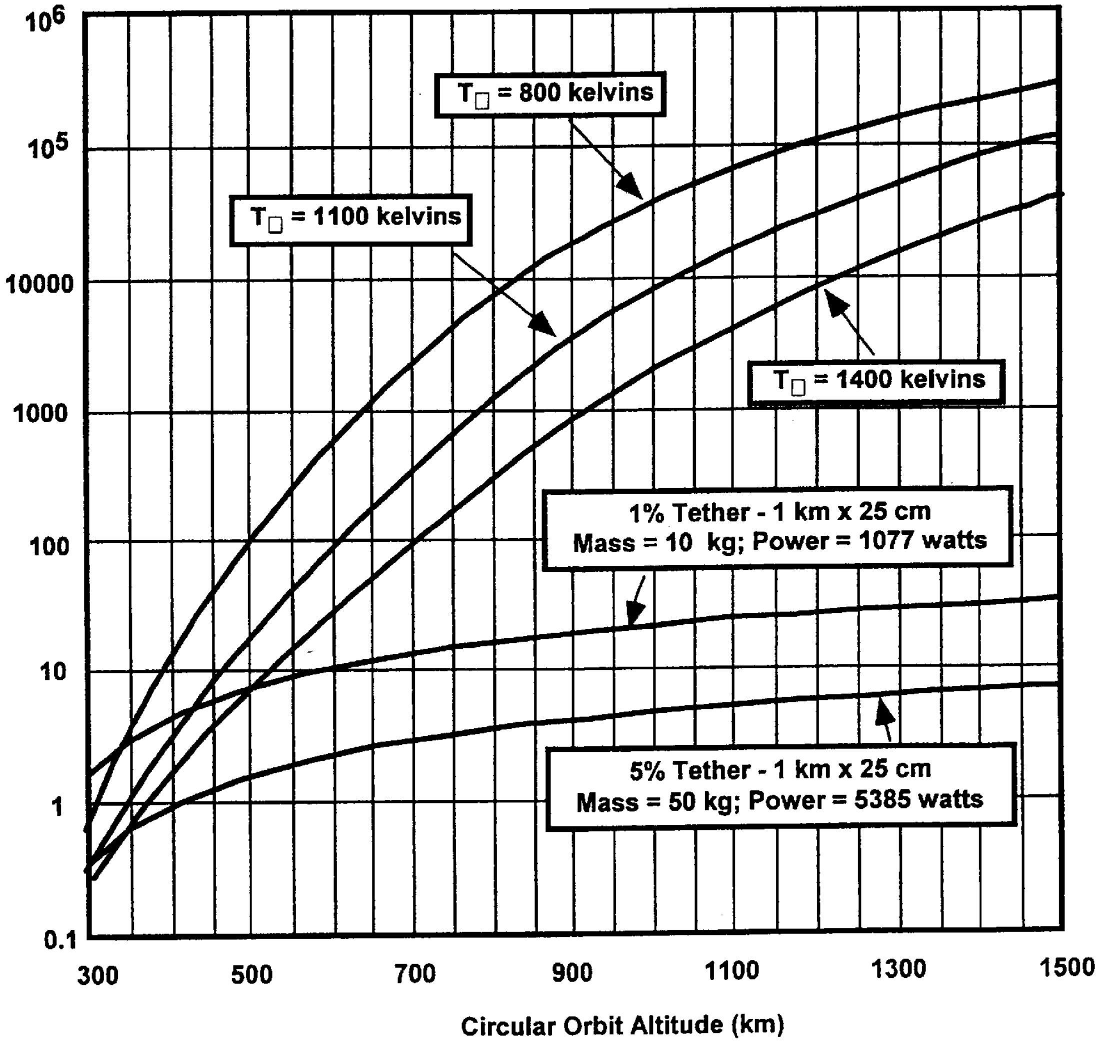

Area-Time Product for Mean and Extremes of Exospheric Temperature

(m = 1000 kg, Cd = 2.0 ; 1% Tether, 1 km x 25 cm; power = 1077 watts)

(5% Tether, 1 km x 25 cm; power =5385 watts)

FIG. 4

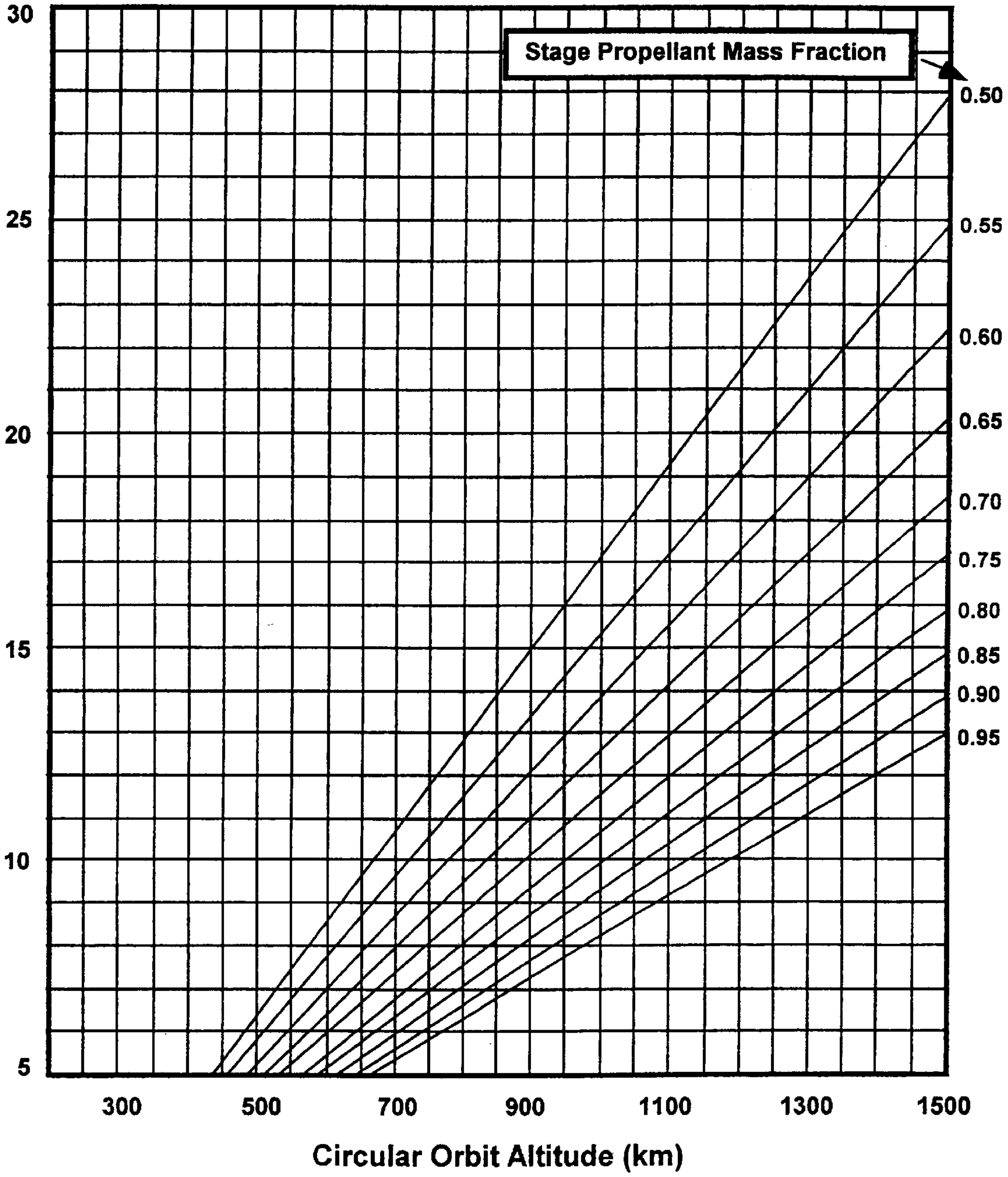

Conventional Solid Rocket Motor System

Percent Increase in Orbital Mass vs. Altitude and I' (Isp = 288 sec)

FIG. 5

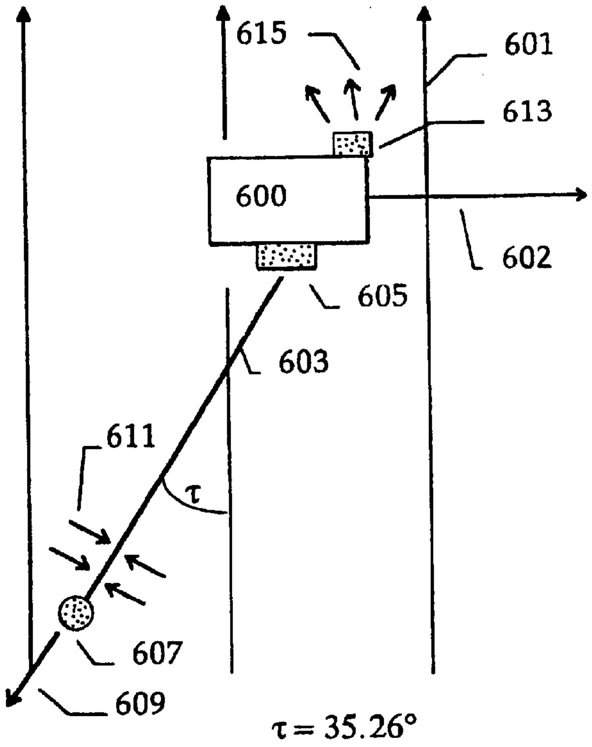
τ
τ = 35.26°

FIG. 6

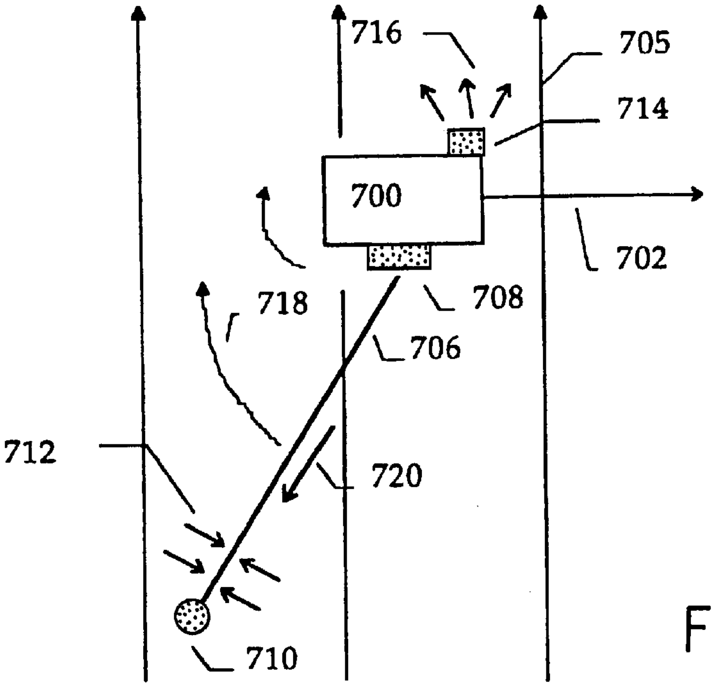

FIG. 7

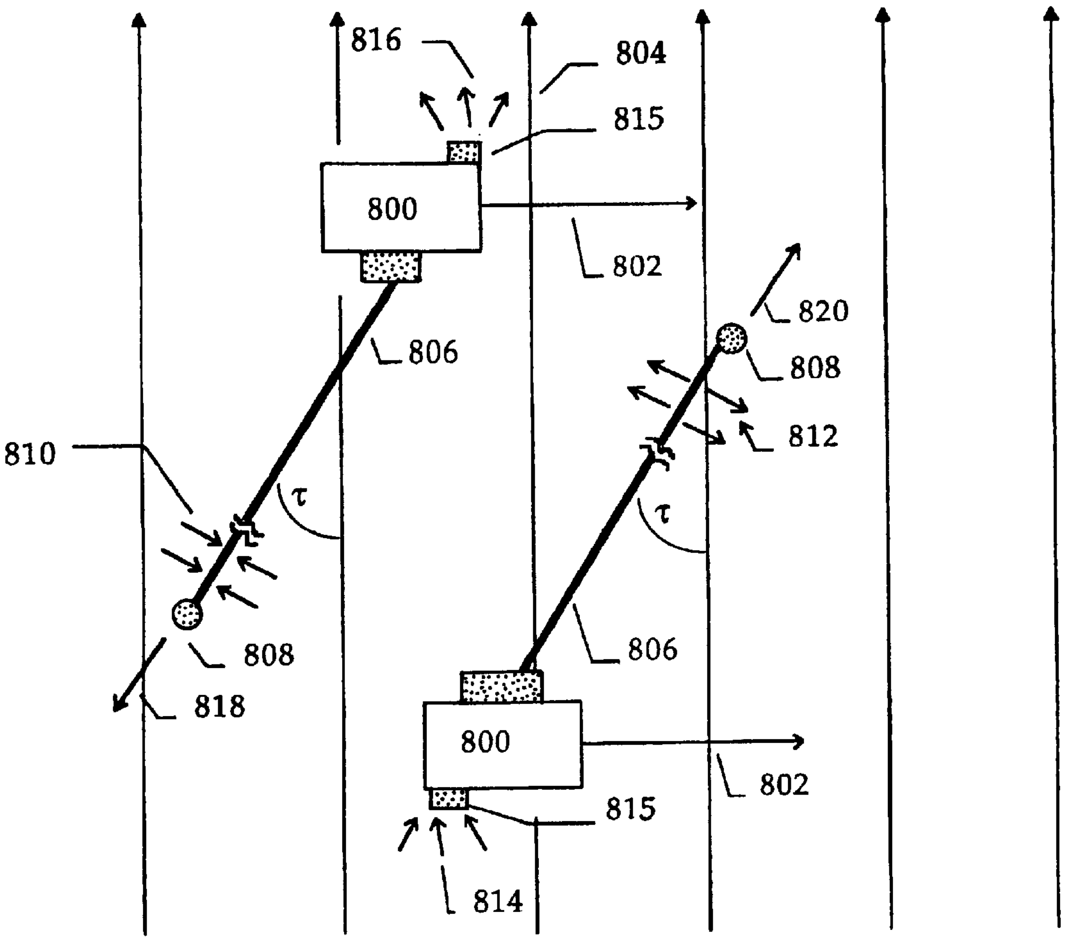
τ
τ

FIG. 8

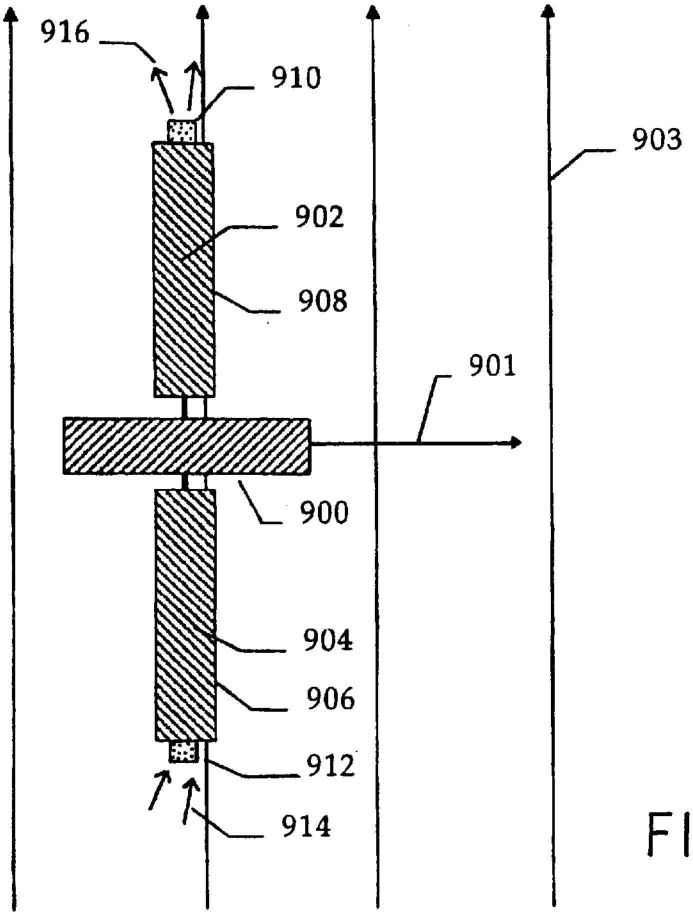

FIG. 9

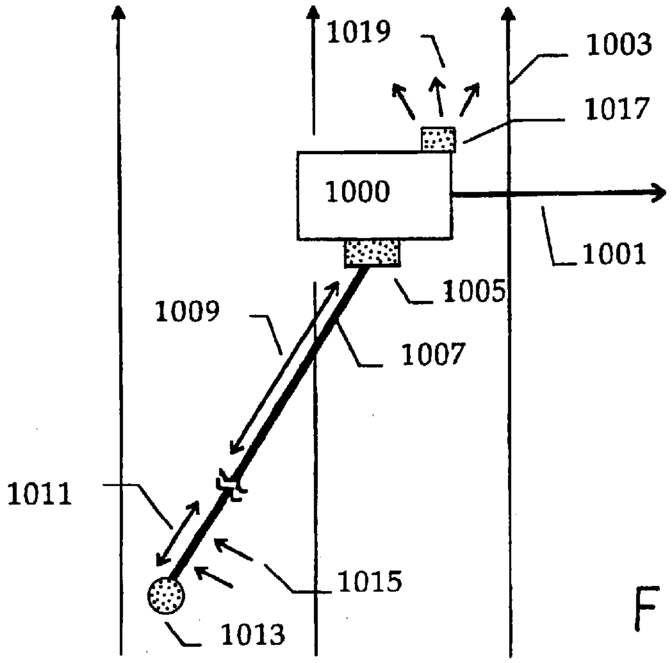

FIG. 10

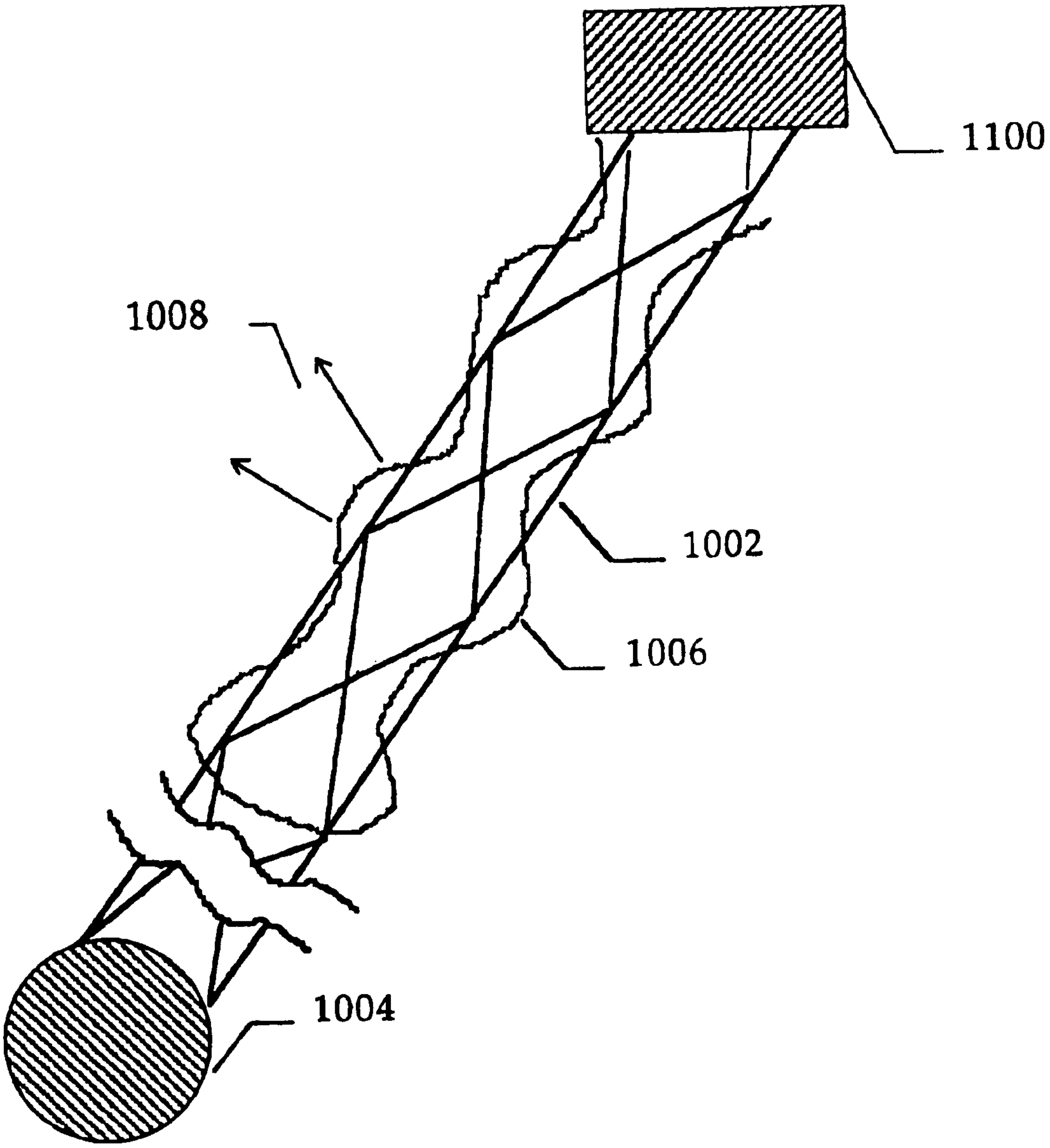
1100
1008
1002
1006
1004

FIG. 11

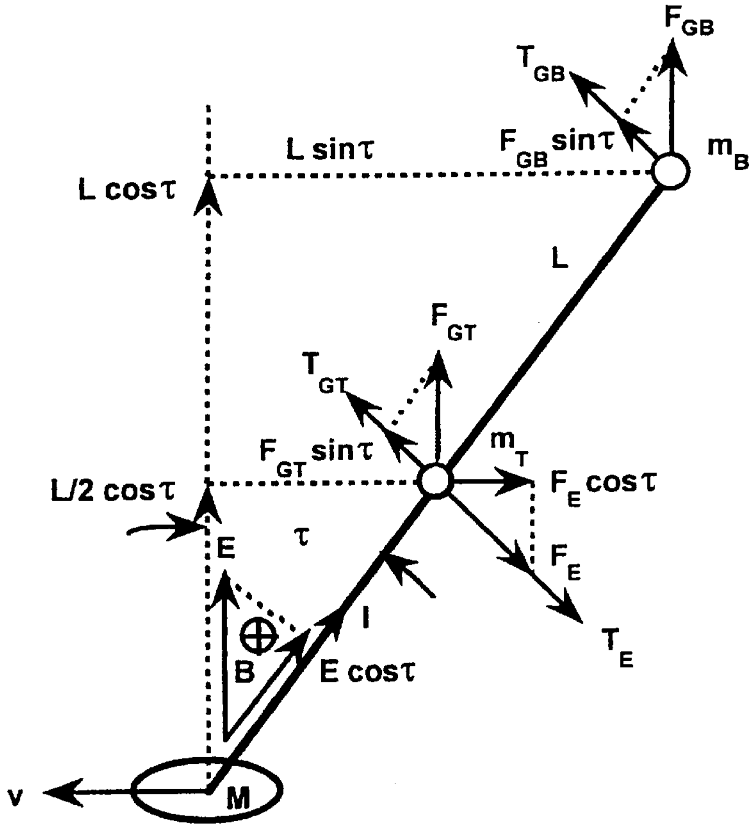

F_{GB}
T_{GB}
$F_{GB} \sin\tau$
m_B
$L \sin\tau$
$L \cos\tau$
L
F_{GT}
T_{GT}
$F_{GT} \sin\tau$
m_T
$F_E \cos\tau$
$L/2 \cos\tau$
E
τ
F_E
I
T_E
$\oplus$
B
$E \cos\tau$
v
M

FORCES AND TORQUES ON AN UPWARD DEPLOYED CONDUCTING TETHER DUE TO THE MOTION OF THE HOST SPACECRAFT THROUGH THE MAGNETIC FIELD OF THE EARTH.

FIG. 12

ELECTRODYNAMIC TETHER AND METHOD OF USE

TECHNICAL FIELD

This invention relates generally to apparatus and methods useful for changing the state vector of a space object when the space object is moving relative to a magnetic field. More specifically, the present invention relates to an apparatus and method of using a conducting tether to produce an electrodynamic force to deorbit a satellite from its orbit around a celestial body, such as the Earth, which has an associated magnetic field.

RESERVATION OF RIGHTS

This application is subject to certain rights of the U.S. government as a result of contracts between the U.S. government and the inventors.

BACKGROUND ART

The present invention has its principal utility in outer space, primarily for deorbiting satellites at the end of their useful life to mitigate the harm and reduce the liability created by the proliferation of space debris. In order to obtain a better understanding of the present invention it is helpful to understand the prior art of space tethers, especially tether dynamics and tether electrodynamics. The present invention may be more readily understood through a review of the experimental prior art and a mathematical analysis of electrodynamic space tethers.

Prior Art Tethers:

A tether was originally a rope or chain used to fasten an animal so that it grazed only within certain limits. Tethers have been used for decades in space to attach astronauts to their spacecraft.

In 1974 Professor Guiseppe Colombo, holder of the Galileo chair of astronomy at the University of Padua in Italy, proposed using a long tether to support a satellite from an orbiting platform. U.S. Pat. No. 4,097,010, which issued to Professor Colombo and Mario Grossi on Jun. 27, 1978, teaches a satellite connected by means of a long tether to a powered spacecraft. Colombo actively pursued the design of a tethered satellite system.

Several NASA experiments, such as the two Small Expendable Deployer System (SEDS 1 & 2) experiments and the Plasma Motor Generator (PMG) experiment used tethers in space. SEDS used a nonconducting tether. The PMG used a 500-meter conducting tether. The Tethered Satellite System flights in 1992 and 1996 (TSS-1 & 1R) used a 20,000-meter conducting tether.

On the TSS-1 mission the tether deployed only 260 meters (853 feet) before the deployer failed. On the TSS-1R the tether was deployed 19,500 meters. In the SEDS-2 flight, a 0.8-mm diameter, 20,000-meter long braided single-line tether was deployed to study tether dynamics and lifetime. Orbital debris or a meteoroid severed this tether in less than four days.

In the TSS-1R flight, the conducting single-line tether was severed after five hours of deployment. This failure was caused by an electric arc produced by the 3,500 volts of electric potential generated by the conductive tether's movement through the Earth's magnetic field.

The Tether Physics and Survivability (TiPS) satellite consists of two end masses connected by a 4,000-meter long non-conducting tether. This satellite was deployed on Jun. 20, 1996 at an altitude of 1,022 kilometers (552 nautical miles). Its tether is an outer layer of Spectra™ 1000 braid over a core of acrylic yarn. The yarn will "puff" its outer braid to two millimeters to "give it a larger cross section to improve its resistance to debris and small micrometeoroids", according to the National Reconnaissance Office (NRO), which is a sponsor of the TiPS mission. As of Jun. 21, 1997 the TiPS tether had survived one year.

References:

1. Joseph A. Carroll, "SEDS Deployer Design and Flight Performance", paper WSEDSA-1 at the 4^{th} International Conference on Tethers in Space, Washington, D.C., April 1995.
2. James E. McCoy, et. al. "Plasma Motor-Generator (PMG) Flight Experiment Results", pp.57–84, Proceedings of the 4^{th} International conference on Tethers in Space, Washington, D.C., April 1995.
3. W. John Raitt, et. al. "The NASA.ASI-TSS-1 Mission, Summary of Results and Reflight Plans, pp. 107–118, Proceedings of the 4^{th} International conference on Tethers in Space, Washington, D.C., April 1995.
4. Joseph C. Anselmo, "NRO Orbiting Spacecraft Studies Tether Survivability", Aviation Week, page 24, Jul. 1, 1996.

These experiments all used single line tethers.

The following reference is illustrative of the current state of the art in space tethers: Paul A. Penzo and Paul W. Ammann. Tethers in Space Handbook—Second Edition. NASA Office of Space Flight, NASA Headquarters, Washington, D.C. 20546. See also the hundreds of references in the 33 page bibliography at the end of the handbook.

The "Hoytether"™, an Improved, High-Reliability Tether:

In 1991, one of the present inventors, Robert Hoyt, invented a lightweight net-like structure that provides many redundant load-bearing paths. A number of primary load bearing lines running the length of the structure are connected periodically by diagonal secondary lines. The disclosed embodiment of this invention has the secondary lines firmly connected by knots to the primary lines. The secondary lines are connected only to the primary lines. At either end of the disclosed structure, a support ring enforces the cylindrical spacing between the primary lines. The secondary lines are designed with a small amount of slack. These secondary lines are only put under load if a primary line fails. This specific tether structure was disclosed to the public in 1992 (Forward, R. L., "Failsafe Multistrand Tether Structures for Space Propulsion", AIAA paper 92-3214, 28^{th} Joint Propulsion Conference, Nashville, Tenn., 1992 (hereinafter "1992 Hoytether structure"). This structure was named a "Hoytether". The term "Hoytether" is used throughout the remainder of this specification for this type of tether structure.

The present invention uses an improved Hoytether, which was invented by the same inventors as the present invention. This improved Hoytether is the subject of a copending PCT application. The Hoytether is discussed briefly in this specification to aid understanding of the present invention.

The 1992 Hoytether design teaches that the normally slack secondary lines have half the cross-section (0.707 the diameter) of the primary lines. There are twice as many secondary lines as primary lines, thus the mass of the secondary lines is equal to the mass of the primary lines. In an undamaged Hoytether, the primary lines carry the entire load, while none of the secondary lines are under load.

While the survival probability of a single-line tether decreases exponentially with time, the Hoytether can maintain a high, i.e. greater than 99 percent, survival probability for periods of months or years (forward and Hoyt, "Failsafe Multiline Hoythther Lifetimes", Paper AIAA 95-2890, 31st Joint Propulsion Conference, July 1995).

References:

1. Robert L. Forward, Failsafe Multistrand Tethers for Space Propulsion, Forward Unlimited, P.O. Box 2783, Malibu, Calif. 90265, July 1992, Final Report on NASA Contract NAS8-39318 SBIR 91-1 Phase I.
2. Robert L. Forward and Robert P. Hoyt, Failsafe Multistrand Tether SEDS Technology Demonstration, Final Report on NAS8-40545 with NASA/MSFC (Jun. 14, 1995).
3. Robert L. Forward and Robert P. Hoyt, "High Strength-to-Weight Tapered Hoytether for LEO to GEO Payload Transfer" Final Report on contract number NAS8-40690 with NASA/MSFC (Jul. 10, 1996).

The Hoytether is essentially a tri-axial net structure, with 'primary' lines running along the length of the tether and two sets of 'secondary' lines connecting these primaries diagonally. They can be made by hand and connected with knots as is taught by the 1992 Hoytether structure. Because knotted connections severely limit the strength of a structure, it is desirable to use a knotless fabrication technique to achieve interconnections that have strengths approaching the limits of the constituent material. As these tethers may be many kilometers long; fast and inexpensive mechanical methods are required for their practical fabrication.

Hoytethers may be made by mechanical braiding, i.e. three-dimensional braiding, such as 3-D rotation braiding using braiding machines such as those developed by the Herzog Company in Germany (August Herzog Maschinenfabrik GmbH & Co., Postfach 2260.26012, Oldenburg, Germany. The specialized loom developed by the Nichimo Company of Japan (Nichimo Company Ltd., 2-6-2 Ohtemachi, Chiyoda-Ku, Tokyo, Japan) is used to produce "Ultracross" knotless fishing nets in which the individual strands are braided as a 4-braid line, and the strands are interbraided where they cross. This produces netting that has slipless interconnections that are very strong, approaching the maximum capability of the fiber. Such a loom could, with some modifications, produce the present invention's structure. Only two such machines exist, one in Japan, the other in Washington State. Unfortunately neither can work with the small line diameters needed to practice the preferred embodiment of the present invention. See generally, Ko, F. K., "Braiding", in *Engineered Materials Handbook, Vol. 1., Composites*. ASM International, Metals Park, Ohio, 1957. Pp. 519–528.

The most common 3-dimensional braiding machines are 4-step braiders based upon the designs of Maistre (German Patent P230-16986, issued 1973) and Forentine (U.S. Pat. No. 4,312,261, issued 1982). Braiding is accomplished by using pneumatics or solenoids to push the parts of the braiding machine to the proper positions. This is a slow process and making a Hoytether kilometers long with these machines would be very time consuming and expensive. The composites division of Albany International (Albany International Research Company, 777 West Street, Mansfield, Mass.) also produces a 3-D braiding machine. This machine uses modular braiding components that arc assembled breadboard fashion on a large wall.

Although braiding is the preferred technique, alternate fabrication methods such as Raschel knitting and crocheting can be used successfully. Multikilometer long Hoyththers are presently being produced for the inventors by the vendors Culzean Fabrics and Flemings Textiles using an electronically controlled crochet machine produced by Comez in Italy.

Space Tether Systems:

The prior art teaches the use of tethers in space applications. U.S. Pat. No. 5,163,641, issued on Apr. 9, 1990 to Yasaka, teaches the use of a powered spacecraft connected by a tether to a satellite. This tether is disconnected to change the state vector of the satellite. The state of the art of energy and momentum transfer using space tethers is discussed in Ivan Beckey's article "Tethering, a new Technique for Payload Deployment", Aerospace America, March 1997, at pages 36–40. Beckey concludes, "Tethers can perform the same functions as propulsive upper stages of direct payload injection, but at lower weight and cost per pound." U.S. Pat. No. 4,923,151, issued Mar. 1, 1988 to Roberts, Wilknson and Webster, teaches a tether power generator for earth orbiting satellites. U.S. Pat. No. 4,580,747, issued Mar. 15, 1983 to Pearson, teaches use of a long tether extending downward into the atmosphere from a satellite. The state vector of the satellite is changed by forces acting on a lifting body connected to the end of the tether. U.S. Pat. No. 4,824,051, issued Jan. 12, 1987 to Engelking, teaches passing an electric current through a conductive tether attached to a satellite to provide propulsive force to alter the orbit of the satellite. U.S. Pat. No. 5,082,211, issued Jan. 21, 1992 to Werka, teaches use of a tether to deorbit space debris. U.S. Pat. No. 4,727,373, issued Mar. 31, 1986 to Hoover, teaches an orbiting stereo imaging radar system having two spacecraft in synchronous parallel orbits connected by a tether.

Tether Dynamics:

In order to understand the forces that cause a tethered satellite to move upward and away from an orbiting satellite, for example, it is first necessary to explain briefly how a satellite remains in orbit. An orbiting satellite is acted on by the force of gravity which pulls it toward Earth, and by a centrifugal force, which pushes it away from Earth. The centrifugal force" (actually inertia) results from the motion of the satellite around its circular orbit. This is the same force that one can experience by swinging a ball around on the end of a string. A satellite is maintained in its orbit when it travels at the natural speed for its altitude and, as a result, the centrifugal force is equal to the gravitational force.

At the typical orbital altitude of 250 kilometers for a low-Earth orbit satellite, for example, a speed of approximately 7.6-km per second is required to create sufficient centrifugal force to balance gravitational attraction on the satellite. If the altitude is changed, the two opposing forces will no longer be in balance unless the satellite also changes its speed. A higher orbital altitude requires a slightly lower speed so the satellite will take longer to complete an orbit. Because of this, if two free-flying satellites are in orbits at different altitudes, the lower satellite will circle the Earth in less time than the satellite in the higher orbit.

If two satellites, at different altitudes, are connected to each other by a tether, they are forced to travel around their orbits together—in the same period of time, which is longer than the natural period of the lower satellite but shorter than that of the upper satellite. The lower satellite will, therefore, slow down below the natural speed for its orbit and will tend to fall to a lower orbit because the centrifugal force will now be less than the gravitational attraction of the Earth. An upward force in the tether that makes up the difference between centrifugal and gravitational forces holds it in place, however.

Correspondingly, the upper satellite will be accelerated above its natural orbiting speed (increasing its centrifugal force above the gravitational attraction) and will tend to move to a higher orbit. It, too, is held in place by an additional force (downward) in the tether. In other words, the net force downward on the lower satellite is balanced, through the tether, by the net force upward on the upper satellite. The effect of unbalanced forces on the two satellites is, therefore, to create tension in the tether. During the TSS-1 & 1R experiments, the inertia of the tethered satellite causes the satellite to rise above the orbiter as the tether is reeled out. Very close to the orbiter, there is little difference in the two orbits, and the tension force is insufficient to overcome friction in the deployer mechanism; therefore, until the satellite reaches a separation of approximately 1000-meters, the tension is augmented by small tether-aligned thrusters on the satellite. Beyond this point, the tension in the tether is the only force required.

By experimenting with a ball hung on a piece of elastic cord (a paddleball, for example) it is possible to simulate all the different types of oscillations that are possible on a space-based tether system. The elastic cord, representing the tether, may compress and stretch, causing the ball to bounce up and down (longitudinal oscillation). It also may move in a circular (skip-rope) motion or may develop wave-like motions (transverse oscillations). Even if the string itself remains straight, it is possible to get the ball swinging back and forth about its attachment point on the paddle like a child on a swing rope (pendulous motion).

Each type of motion occurs with a particular frequency, which depends on the length and tension of the tether. When the frequencies are different, the motions do not interact; however, at some tether lengths, the frequencies of two or more types of oscillation can become very close. At this point, energy can be transferred from one type of motion to another, a phenomenon known as resonance. For instance, the transverse oscillations in the tether may cause the satellite to swing back and forth in pendulous motion.

Many different factors may cause oscillations; the movements of the satellite or Shuttle are but two of these. For an electrodynamic tether, the skip-rope and pendulous oscillations are of particular interest. If a current is passed through a tether, the current will interact with Earth's magnetic field, resulting in a force that may produce skip-rope and pendulous oscillations. Because it is necessary to maintain control of the satellite, much study has gone into identifying the different types of possible motions and the methods used to control them.

One way to control the magnitude of those motions that cause a change in tension or transverse motion at the end of the tether is to have an end mass connected to the Hoytether that maintains a controlled tension on the tether, working much like a spring-loaded 'dog leash'. This may be as simple as a coiled spring, or as complex as an active control system that measures the tension and transverse forces on the tether and adjusts the applied tension according to a local or remotely operating algorithm.

Electrodynamic Effects of Conductive Tethers:

Electric potential is generated across a conductive tether by its motion through the Earth's magnetic field. Electromagnetic forces acting on a conductive tether in orbit can make the tether system behave like an electric motor or generator, thereby exerting a useful force to alter the state vector of the tether and any mass attached to it.

Electrodynamic tether propulsion is unlike most other types of space propulsion in use or being developed for space application today—there is no hot gas expelled to provide thrust. Instead, the environment of near-Earth space is being utilized to propel a spacecraft or upper stage via electrodynamic interactions.

A charged particle moving in a magnetic field experiences a force that is perpendicular to its direction of motion and the direction of the field. When a current flows through a long, conducting tether the electrons flowing through the tether experience this force due to the fact that they are moving along the wire in the presence of Earth's magnetic field. This force is transferred to the tether and to whatever the tether is attached (like a spacecraft, satellite, space station or upper stage). It can be an orbit-raising thrust force or orbit-lowering drag force, depending upon the direction of current flow. Operation in one mode allows boost from LEO to higher orbit while reversing the current flow provides negative thrust for deboost. The principle is much the same for an electric motor; reverse its operation and it acts as a generator. The current passing through the tether is returned through the ionosphere to complete the current loop with collection and emission occurring on opposite ends of the tether.

The PMG experiment demonstrated that a conducting tether can be used as both a motor and a generator. The TSS experiments, especially TSS-1R showed that very large voltages (about 3500 volts) can be generated by a sufficiently long tether.

Uses of an electrodynamic tether as an orbit raising and lower propulsion system has many advantages over competing systems:

a. It is nearly propellantless. Most other systems expel hot gases and require extensive resupply. To emit current, the electrodynamic tether propulsion system may use plasma contactors developed as a part of the International Space Station Program. These contactors consume less than 20 kg of xenon gas per year with a 50% duty cycle. The electrodynamic tether propulsion system can also use field emitter arrays, which emit electrons without the use of any gas.

b. It can change both altitude and inclination. The Earth's magnetic field is non-uniform and can therefore provide both in- and out-of-plane forces for inclination changes as well as altitude changes. This is of particular interest to payloads requiring polar orbits in that they can be launched on a small launch vehicle into a lower inclination orbit and have it raised in space by the proper phasing of current through the tether.

A demonstration of the propulsive capabilities of electrodynamic tethers was recently approved for a flight test in 1999. The Propulsive SEDS or ProSEDS mission, will fly as a secondary payload on a Delta II launch vehicle and deploy a 5-km conducting tether using the existing SEDS deployer concept. The ProSEDS experiment will be followed by the Electrodynamic Tether Upper Stage (EDTUS) experiment that will demonstrated the use of electrodynamic forces to change both the altitude and inclination of the experimental spacecraft. FIGS. **1B** and **1C** show the calculated electrodynamic thrust at several inclinations and the reentry time sensitivity of the ProSEDS tether, respectively.

One application for long-life conducting electrodynamic tethers is as a "Terminator Tether™" for removing from orbit unwanted Earth orbiting spacecraft at the end of their useful lives. When the mission of the satellite is completed, the Terminator Tether™, weighing a small fraction of the mass of the satellite, would be deployed. The electrodynamic interaction of the conducting tether with the Earth's magnetic field will induce current flow in the tether conductor. The resulting energy loss from the heat generated by the current flowing through the ohmic resistance in the conducting tether will remove energy from the spacecraft, eventually causing it to deorbit, thus reducing the amount of orbital space debris that must be coped with in the future.

In the following analysis, it is shown that the amount of energy loss generated by an electrodynamic tether is essentially independent of its length or area, and instead is primarily proportional to the tether mass and the physical properties of the conductor metal chosen. In the typical example calculated, a 1000-kg spacecraft can be deorbited from a 1000-km high Earth orbit by a 10-kg mass tether in a month, while a 1-kg tether can deorbit a 1000-kg spacecraft in less than a year.

To the knowledge of the inventors, Joseph P. Loftus of NASA/JSC first proposed the general concept of using an electrodynamic tether to deorbit spent satellites. (Joseph P. Loftus <JLoftus@ems.jsc.nasa.gov>, personal communication via email to Robert Forward, Monday Jun. 10, 1996 15:50:10.) In order to show that the Loftus deorbit concept was not obvious to those skilled in the art of electrodynamic tethers, Forward contacted the leading expert, Joseph Carroll, of Chula Vista, Calif., who built and participated in the flight test of the PMG. After being told of the Loftus concept in a telephone conversation, his reply in an Email message dated Aug. 5, 1996, was "such a system would be feasible . . . by it is still not obvious to me that it would be useful . . . "

Loftus was considering the use of electrodynamic drag from a conducting tether to achieve this goal of bringing the unwanted spacecraft down from its high orbit (where atmospheric drag is negligible) to a 200-km orbit, where atmospheric drag would rapidly finish off the task of removing the unwanted spacecraft from orbit. The tether Loftus was considering was a single-line, conducting tether, typically 1-mm in diameter, 1-km long, and, if made of aluminum, 2-kg in mass. He would include means at the ends of the tether to contact the ambient space plasma around the Earth to complete the current loop.

Unfortunately it is probable that space impactors would sever the 1-mm diameter, 1-km long single-line tether proposed by loftus within a 1/e lifetime of four months. This would produce orbital debris rather than removing it. The motivation for this work is the NASA Safety Standard NSS 1740.14 "(Guidelines and Assessment Procedures for Limiting Orbital Debris." The relevant portion of the Standard starts on page 6-3: General Policy Objective-Postmission Disposal of Space Structures. Item 6-1: "Disposal for final mission orbits passing through LEO: A spacecraft or upper stage with perigee altitude below 2000 km in its final orbit will be disposed of by one of three methods." The method of interest is the atmospheric reentry option, Option a: "Leave the structure in an orbit in which, using conservative projections for solar activity, atmospheric drag will limit the lifetime to no longer than 25 years after completion of mission. If drag enhancement devices are to be used to reduce the orbit lifetime, it should be demonstrated that such devices will significantly reduce the area-time product of the system or will not cause the spacecraft or large debris to fragment if a collision occurs while the system is decaying from orbit."

The NASA standard applies only to NASA spacecraft and even then only to completely new spacecraft designs. New versions of existing designs are to make a "best effort" to meet the standard, but will not be required to change their design to do so. The Department of Defense has adopted the NASA standard with the same provisos. An Interagency Group report has recommended that the NASA standard be taken as a starting point for a national standard. It is NASA's recommendation to the Interagency Group that the safety requirement be phased in only as spacefaring nations reach consensus internationally, which is being done through the International Debris Coordination Working Group whose members are Russia, China, Japan, ESA, UK, India, France, Italy, and the US.

Thus, although the NASA Safety Standard in its present form is not the "Law", the existence of the standard means that some time in the future a similar requirement may be imposed on all spacecraft. This could result in major growth in future space tether business, with a sale to every non-geostationary spacecraft being "mandated" by government safety regulations, somewhat as the sale of seat belts and airbags for every car are mandated.

In fact, three of the companies planning to set up "constellations" of low to medium orbit communications: Teledesic, Iridium and Odessey have committed their companies to abide by the spirit of NASA Safety Standard 1740.14 by using one means or another to deorbit their spacecraft before they reach end of life.

Problems with Prior Art Tethers:

All electrodynamic tether designs proposed by the prior art teach that the tether should be operated at a right angle to the magnetic field through which the tether is moving. This is a problem because the electrodynamic force acting on the tether causes the tether to align itself with the magnetic field force lines. To overcome this problem the prior art teaches the use of a large ballast mass attached to the end of the tether and/or use of a very long (tens to hundreds of kilometers) tether. The large ballast mass is expensive to take to orbit because it replaces useful payload. The long tether sweeps a larger Area-Time-Product during its useful life and thus is more likely to impact other space objects, either debris or another spacecraft.

Another problem common to all proposed prior art tethers is tether instability. If the tether produces a large electrodynamic drag force, which is desirable because a large drag force will cause the satellite to deorbit quickly, then the tether will be dynamically unstable. This instability can cause the tether to lose its effectiveness, act uncontrollably and even wrap around the satellite or otherwise malfunction. Experts skilled in the art of tether design have opined that this dynamic instability is inherently unavoidable in any electrodynamic tether system. The prior art solution, such as that presently being used in the ProSEDS experiment, has been to use a large ballast mass to increase the stabilizing gravity-gradient force and/or to limit the electrodynamic drag of the tether to less that the maximum that could be produced. In the ProSEDS experiment, the conducting electrodynamic tether is five kilometers long. To insure stability, it will be augmented by a 20–35 kilometer long non-conducting tether, which to further have stability will have a 40 kilogram ballast end mass.

Yet another problem of all proposed prior art electrodynamic tether systems is how to radiate away the energy produced by the tether's operation. A satellite moving at an orbital velocity of 18,000 miles per hour has a kinetic energy of over 10,000 calories per gram. To put this amount of energy in an understandable perspective, it may be noted that when nitroglycerine explodes it produces about 1,500 calories per gram. Prior art designs of electrodynamic drag tethers teach the use of the electrical energy generated by the tether to charge batteries or operate electronics, with the excess energy being converted into heat by a resistive load. This excess heat must be radiated to the space environment or it will melt the resistive load. Thus the resistive load, and/or its associated radiator structures, must be massive and replace useful payload.

DISCLOSURE OF THE INVENTION

The present invention comprises an electrodynamic tether structure and a method of use. The principal industrial utility of the present invention is to deorbit satellites in Earth orbit at the end of their useful life. This embodiment of the present invention is sometimes referred to in this specification as a "Terminator Tether™" because it terminates the orbital lifetime of the host spacecraft. The structure of the tether taught by the present invention is a short, wide, interconnected-multiwire (compared to the long single wires of the prior art) conductive Hoytether whose area maximizes electrodynamic drag while simultaneously minimizing the Area-Time-Product swept by the tether during its operating life. The preferred tether length is two to five kilometers. The preferred tether mass is one to five percent (1%–5%) of the spacecraft mass. The method of operation comprises orienting the tether structure at a 35.26-degree trailing angle to the local vertical to maximize electrodynamic force on the tether while avoiding tether instability and allowing use of a small tether end mass.

The present invention also teaches that the satellite-tether system may be rotated around its common center of mass to centrifugally produce tension force in the tether structure to oppose forces causing tether instability. The angle of the conductive tether structure of the present invention with respect to the velocity vector of the host spacecraft may be controlled by the method of the present invention so it interacts with the encountered magnetic field to induce a maximum current flow in the tether. This produces maximum electrodynamic drag. All or a portion of this electric power may be stored and then controllably applied to the conductive tether to produce an induced electrodynamic force. This induced electrodynamic force may by used to enhance the drag force, to rotate the tether-satellite system and/or to provide satellite propulsion, i.e. to change the state vector of the satellite for any useful purpose, e.g. to avoid collision or to change the host spacecraft's orbit to an orbit more favorable for more rapid deorbiting.

The present invention also teaches a tether structure that also functions as a thermal radiator and/or plasma contactor. An embodiment of the present invention using conducting elements of the satellite, e.g. the solar arrays, as electrodynamic tether structures is also disclosed.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention may be better understood with reference to the following drawings:

FIG. **1***a* is a graph comparing the survival probability of a single line tether as taught by the prior art to the Hoytether used by the present invention.

FIG. **1***b* is a graph showing the level of electrodynamic thrust for a 10 kilometer, 10 kilowatt tether at various altitudes and inclinations.

FIG. **1***c* is a graph showing reentry time sensitivity of a five kilometer ProSEDS tether for several conditions of current flow in the electrodynamic tether.

FIGS. **2***a*, **2***b* and **2***c* show the Hoytether structure used by the preferred embodiment of the present invention.

FIG. **2***d* shows the Hoytape structure used by the preferred embodiment of the present invention.

FIG. **3** is a block electrical diagram of an electrodynamic tether system.

FIG. **4** is a graph showing a graph of the Area-Time product for three separate levels of exospheric temperature (for neutral drag calculations) and for two separate assumptions regarding tether mass and power drawn from the ambient plasma (for the Terminator Tether™ calculations).

FIG. **5** is the percent additional mass required in orbit to drop the perigee of a circular orbit at altitude, at, to a value of 200 kilometers using standard chemical rocket as taught by the prior art for deorbiting unwanted spacecraft.

FIG. **6** is a deorbit tether system shown operating with its electrodynamic tether at a trailing 35.26-degree angle to the encountered magnetic field, as is taught by the preferred embodiment of the present invention.

FIG. **7** shows an embodiment of the present invention in which the electrodynamic tether—satellite system is rotated about its center of mass to avoid tether instability.

FIG. **8** shows an embodiment of the present invention wherein the tether is powered to provide induced electrodynamic force.

FIG. **9** shows an embodiment of the present invention wherein the solar power system structure of the satellite is used as an electrodynamic tether.

FIG. **10** shows the use of a portion of the tether structure as a thermal radiator and plasma contactor.

FIG. **11** shows the braiding of an ohmic resistive load into the tether structure.

FIG. **12** is a force diagram showing the forces and torques on an upward deployed conducting tether due to the motion of the host spacecraft through the magnetic field of the Earth.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

The principal industrial utility of the present invention is the deorbiting of satellites in Earth orbit, which necessarily can happen only in outer space. Until the present invention is reduced to actual practice by use with satellites in outer space and a body of practical experience is thereby obtained, the present invention can only be constructively reduced to practice, such as by this written specification and its associated drawings, diagrams and graphs and by reference to electric measurements made on conducting tethers in space by the PMG and TSS experiments. The inventors provide herein a detailed discussion of the theory of the present invention to help those skilled in the art of aerospace and tether engineering to understand the present invention and to make and use the best embodiment of the present invention known to the inventors at the time this specification was prepared, without undue experimentation.

Although the present invention is discussed in this specification in its preferred embodiment as a means of deorbiting satellites, it must be emphasized that discussion of this specific use in no way limits the broad scope of the present invention. The present invention can be used to change the direction and speed, i.e. the state vector, of any space object providing that object is in motion relative to any magnetic field. This magnetic field may be that of the Earth or of any other celestial body, for example Jupiter or the sun.

In this specification the material forming the structure of the electrodynamic tether is specified as being a conductor. For example, a metal, such as copper or aluminum wire could be used. Likewise, nonmetallic conductor, such as carbon nanotubes, or a conductive polymer could form the conductive structure of the tether.

FIG. **1***a* shows the survival curve for a Hoytether as used by the present invention vs. the survival curve of a prior art single line tether. The survival probability curve of the Hoytether as a function of time does not conform to the well known "1/e decay" shape of a single-line tether. The Hoytether can maintain a high level of survival probability, about 99 percent, until it nears its 'lifetime'. Its survival probability then drops rapidly to zero. A detailed mathematical analysis of the difference between Hoytether and single tether survival probabilities is given in Appendix E "Small Impactor Survival Probabilities of Hoytethers" and Appendix F "Large Orbital Debris Survival Probabilities of Hoytethers", both contained in the Final Report of NASA Contract NAS8-40545. The resulting cut probability with time for the Hoytether has a "bingo curve" shapes. In a bingo game, at least five numbers must be called before anyone can win, and usually many numbers have to be called before one of the bingo cards gets five in a row. In the Hoytether at least four cuts must happen at the same level before any failure occurs, and many cuts have to be made before any one of the levels has all four lines cut. The bingo curve has the property that the probability of survival stays very high for periods short compared to the lifetime. The probability of survival is greater than 99% for periods shorter than 10% of the lifetime. This is much better performance than the 1/e curve of a single line tether, where the probability of survival is only 90% at 10% of the 1/e lifetime.

FIG. **1***b* shows the calculated level of electrodynamic thrust, in Newtons, produced by a ten kilowatt, ten kilometer single conductor tether at altitudes between 200 and 1400 kilometers and at orbital inclinations between zero degrees and eighty degrees.

FIG. **1***c* is a graph showing the reentry time sensitivity calculated for a five kilometer ProSEDS single conductor tether at altitudes between 200 and 500 kilometers with current flow in the tether between on flow and one ampere.

FIGS. **1***b* and **1***c* are the work of Enrico Lorenzini of the Smithsonian Astrophysical Observatory from his study "Performance Evaluation of the Electrodynamic Tether Tug", NASA/MSFC Grant NAG8-1303 (Nov. 12, 1966). They show that experts in the field believe that the present invention will have industrial utility, as the current flow of one ampere in FIG. **1***c* decreases the deorbit time for the calculated case from more than 180 days to less than 20 days.

FIG. **2***a* is an isometric drawing showing the generally cylindrical structure of the Hoytether. In FIG. **2***a*, primary lines **201** and **203** are shown connected via secondary lines **211**. In FIG. **2***b*, which is a schematic illustration of the structure shown in FIG. **2***a*, primary lines **201**, **203**, **205**, **207** and **209** are connected, each to their respective adjacent primary lines, by a plurality of secondary lines **221**. These connections, for example as shown at interconnection **213**, are made by knotless, slipless interconnections, such as Soutache braiding of twisted connections.

FIG. **2***c* shows the Hoytether structure of FIG. **2***b* but with a severed primary line **205** at breakpoint **215**. The second level of secondary lines **217**, shown as bold lines in FIG. **2***c*, redistribute the load from the severed primary line **215** at point **217** back to line **215** at points **219** and **221**, above and below the break, respectively.

For the electrodynamic tether, a preferred embodiment would be to unroll the tubular Hoytetherr™ into a flat Hoytape™ as shown in FIG. **2***d*.

FIG. **3** generally shows the arrangement of functional elements of an electrodynamic tether system. Of course such a system must have a means of attachment to the spacecraft, a deployer, the tether itself, some form of space plasma cathode contactor, some form of space plasma anode contactor, a power dissipation system, and a communication and control system. Each of these individual elements are known in some form to those skilled in the art of aerospace engineering. Thus the present inventors do not feel it is necessary to describe them in detail in this specification.

In FIG. **3**, a space object **301**, i.e. a satellite in Earth orbit, or any other space object either natural or man made, is physically connected to the tether system. The tether system comprises a deployer **303** from which a conductive Hoytether **305** having a bare segment **307** extends upward from space object **301**. The positively biased anode end **309** of tether **305** collects electrons from the ionosphere as space object **301** moves in direction **310** across the Earth's magnetic field. These electrons flow through the conductive structure of the Hoytether to the power system interface **311**, where it supplies power to an associated load, not shown. The electrons then flow to the negatively biased cathode **313** where electrons are ejected into the space plasma **315**, thus completing the electric circuit.

In order to allow those skilled in the art to better appreciate the broad scope of the present invention, the inventors will now provide an example using specific tether and system values to compare the Terminator Tether™ to prior art means, ie a rocket, for deorbiting a satellite.

Terminator Tethers™:

When an Earth orbiting conducting space tether moves through the magnetic field of the Earth, an electric field is generated in the tether that is proportional to the velocity of the conductor, the magnetic field strength of the Earth, and the angle between the conducting tether and the magnetic field lines. From data obtained during the various electrodynamic experiments that have been conducted in space to date, such as the PMG, TSS-1, and TSS-1R experiments, a typical value of the generated emf per unit length of tether of E=100–200 Volts/kilometer can be assumed. The electric potential V developed at the ends of a tether of length L is then V=EL. For a tether of length L=10 km, the electric potential developed is V=1000 V. For calibration, the 20-km long TSS-1R tether, at the moment of failure, was developing a potential of 3500 volts, which is 175 volts/km.

The mass of a conducting tether of length L, cross-sectional area A, and densin, d is given by m=dLA. Typical values for the density are d=2700 kg per cubic meter for aluminum and 8900-kg per cubic meter for copper. For a typical aluminum tether of mass m=10 kg and length L=10 km, the cross-sectional area will be A=0.37 mm squared. If this were a solid-wire single-line tether, the diameter of the wire would be D=0.69 mm (21.5 gauge). If this were an 18 line tubular Hoytether, the diameter of the lines in the Hoytether would be D=0.16 mm (34 gauge).

The resistance of a conducting tether of length L and cross-sectional area A is given by R=rL/A, where r is the resistivity of the conductor in nano-ohm-meters (nΩ-m). Typical values for the resistivity are r=27.4nΩ-m for aluminum and 17.0 -nΩ-m for copper. For the m=10 kg aluminum tether of length L=10 km and cross-sectional area A=0.37 mm squared at 20 degrees C, the end-to-end resistance is 750 Ω. This value of resistance is the essentially the same whether the conductor area is concentrated into a single-line tether or divided up into many lines as in a multiline Hoytether.

The current I generated in the conducting tether by the electric potential V between the ends of the tether applied across the tether resistance R is I=V/R. For the 10 kg mass aluminum tether of length L=10 km, electric potential V=EL=1000 volts, and resistance R=750 Ω the current is I=V/R=1.33 Amps. Currents near these values were measured in the TSS-1R experiment at the time of failure. At the time of failure of the TSS-1R tether, none of the plasma contactors on the Space Shuttle or the Italian Satellite were operating. Thus the current was being collected by the conductive surface area of the Space Shuttle and the Italian

Satellite at the two ends of the tether. The Space Shuttle area is quite large, so it was not the limiting factor in current collection. The diameter of the Italian Satellite was 1.6 m, which would give it an effective plasma contact area of about 8 square meters. This shows that, if the plasma contact area of the ends of a conductive space tether can be make large enough, then ampere level currents can be extracted from the ambient space plasma without the use of space plasma contactors.

The power dissipated as ohmic heating in the tether is given by P=IE. For an aluminum tether of mass m=10 kg, resistivity r=27 nΩ-m, and density d=2700 kg/cubic meter, subjected to an electric field of E=100 V/m, the power dissipated in the ohmic losses of the conductor is P=IV=1330 Watts. For a 1 kg mass tether, it would be still a considerable 133 Watts of dissipation.

There will no doubt be additional dissipation of energy in plasma ohmic losses, plasma wave generation, and plasma ion acceleration, but the ohmic losses in the conducting tether alone are sufficient for the task of deorbiting an unwanted spacecraft massing 100 to 1000 times more than the tether.

The decay time of a metric ton spacecraft moving from a 1000 km altitude orbit to a 200 km altitude orbit with an energy difference of dU=3.3GJ, when its energy is being dissipated at a power of P=1330 W by an aluminum tether massing just 10 kg, or 1% the mass of the spacecraft, is about one month. This is a remarkably short time, and indicates that the concept of using a conductive tether to deorbit a spacecraft is indeed feasible. If the aluminum tether massed only 1 kg, or 1/1000th the mass of the spacecraft it was deorbiting, then the decay time would rise to 10 months, still a reasonable value.

In reality, of course, the actual decay time will be longer than this. If the electrodynamic drag force is very large, and becomes larger than the gravity gradient forces pulling on the ends of the tether (which force is proportional to the mass of the tether), then the tether will tend to align itself along the magnetic field lines instead of across them, and the drag force will decrease because of the small angle between the conductor length and the magnetic field lines. The tether will then settle into an angle determined by the balance between these two forces.

Is the Terminator Tether™ theoretically a better means than atmospheric drag or a rocket engine for deorbiting satellites at the end of their useful life?

In the formula for da/dt (change in altitude per change in time), using the assumptions of near-circular spiral trajectories, the Area-Time-Product (Z), the criteria by which NASA judges compliance with Safety Standard 1740.14, is given by:

$$Z = A\int dt = -\frac{m}{C_D}\int \frac{da}{\rho(a)\sqrt{\mu a}},$$

where ρ(a) is simply the density as a function of semi-major axis. Thus, for a static atmosphere, the problem of area-time product is reduced to quadrature. Note that the area-time product depends only on the density profile and the ratio of spacecraft mass to drag-coefficient and is linearly related to that ratio.

The basic altitude-dependence of the atmospheric density, represented here by ρ(a) can be expressed in a low-order power series of ln(density)=f(ln(altitude)), so that it will be possible to develop a representation of the log—log relationship between altitude and density. The three static levels of atmospheric density are modeled as 5th order polynomials in the natural log of the altitude and are representative of well known atmosphere models, for example, Jacchia, L. G., "Thermospheric Temperature, Density, and Composition: New Model," SAO Special Report 375, March 1977. 2. Anon., Marshall Space Flight Center, "Long-Range Statistical Solar Activity Estimation," Atmospheric Sciences Division, 1989.) The values of exospheric temperature are reasonable (about 2-sigma) high and low solar activity values of 1400 and 800 kelvins respectively. The mean value is taken to be 1100 kevins.

FIG. 4 shows a graph of the Area-Time product for three separate levels of exospheric temperature (for neutral drag calculations) and for two separate assumptions regarding tether mass and power drawn from the ambient plasma (for the Terminator Tether™ calculations). The density profiles for the three values of exospheric temperature represent the extremes and mean values to be expected during the next several decades. The exospheric temperature goes through a cycle of about an 11 year period with maximum about 1400 kelvins and minimum about 800 kelvins.

The tether calculations include an inherent assumption that the Terminator Tether™ transfers all the energy extracted from the ambient space plasma into drag that decreases the orbital energy of the spacecraft at a rate given by the power drawn by the tether. Thus, the change in energy for a spacecraft of mass m at an initial altitude h is given by the expression:

$$\Delta E = -m\left\{\frac{\mu}{2(R_e + h)} - \frac{\mu}{2(R_e + 250)}\right\},$$

and the time required to effect this change in energy of the mass, m, is just

$$\Delta t = \Delta E / P,$$

where P is the power drawn by the tether from the ambient plasma charge.

The curves for 1% and 5% tethers do not include the effects of atmospheric drag because the electrodynamic drag is orders of magnitude greater than the neutral atmospheric drag. At altitudes greater than about 700 km, the electrodynamic drag is 200 to 3000 times greater than the neutral drag forces.

It is clear, from FIG. 4 that the Terminator Tether™ concept is far superior to neutral drag in removing spacecraft from orbit, no matter how much additional area is added to the passive spacecraft to increase the atmospheric drag. The analyst should note that FIG. 4 is proportional to mass; that is, the Area-Time-Product values should be multiplied by the mass of the spacecraft and divided by 1000 kg. This is true for the neutral drag and Terminator Tether™ curves alike. The power levels assumed for the 1% and 5% tethers are only 80% of their theoretical values. This is done to provide a 20% margin on the power available from the ambient plasma and electrodynamic gradient.

Conventional rocket mechanisms can remove spacecraft from orbit, but this mechanism is apparently not viable from mass considerations when these factors are compared with the capabilities of the Terminator Tether™. A satellite owner may decide to satisfy the NASA safety requirements by adding a small solid motor and the associated hardware, software, sensors and structure to make the package independent, to be used to deorbit the spacecraft in case the main spacecraft power, attitude, or propulsion system fails.

The requirements of such a rocket motor system are more stringent than those attributed to ordinary spacecraft. The rocket motor deorbit system must operate when some or all other systems of the spacecraft have failed. These more stringent requirements are balanced by lesser requirements of performance. The backup system must simply deorbit the spacecraft; it does not have to perform all the other duties of the spacecraft. But the backup system must know when to fire under all kinds of anomalous situations, including tumbling, offset of center of mass (because of loss of parts due to collisions), and lack of knowledge of the orbital position.

FIG. **5** shows the percent additional mass required by the rocket motor system to drop the perigee of a circular orbit at some initial perigee altitude to a value of 200 km. The atmospheric drag at this 200 km perigee altitude will remove any spacecraft (in the range considered) from orbit in a few revolutions. The contours of constant stage propellant mass fraction, λ, range from low values of 0.5 through reasonable values of 0.65 to 0.7, up to the values associated with the best solid motors ($\lambda \nabla 0.93$) without adding any extra hardware to the emergency stage. An effective, independent stage to provide a retro deltaV of from 50 to 325 m/s will almost certainly have a λ of the order of 0.6 to 0.75. If the emergency stage is required to perform its own attitude determination, the stage propellant mass fraction may be as low as 0.55 or 0.50. Note that the additional mass, shown in FIG. **5**, must also be lofted to orbit in the first place, to provide the mass on orbit for the originally intended service. The stage propellant mass fraction, λ, is a key stage performance parameter that describes the ratio of the mass of propellant to the mass of the stage. That is $\lambda=m_p/(m_p+m_I)$, where m_p is the mass of the propellant and m_I is the mass of everything else in the stage (not including the payload or any stages above or below the stage being considered). Typical independent small rocket stages have values of λ from 0.60 to 0.75. The curves of stage propellant mass fraction are shown as straight lines (a minor approximation) and are included only from 0.5 to 0.95, the extremes of reasonable design practice.

These calculations that show, beyond any reasonable doubt, that the Terminator Tether™ concept is far superior to conventional mechanisms such as drag enhancement devices or small rocket deorbit propulsion systems. The superiority is measured in terms of Area-Time-Product, NASA's measure of the likelihood of collision with other spacecraft in the path of the descending spent member of a constellation. Tether calculations were made using conservative assumptions that the power extractable from the ambient plasma and electrodynamic gradient is only 80% of the theoretical power available to a perfect tether crossing the magnetic field lines at a right angle, i.e. normally.

The following analysis is presented by the inventors to help those skilled in the art to better understand the present invention when it is used as a Terminator Tether™ for removing from orbit unwanted non-geostationary Earth-orbiting spacecraft at the end their useful lives. The primary result of that analysis is that the electrical power P in the tether that is converted into heat by the resistance of the tether and radiated away into space will remove energy from the spacecraft, causing it to rapidly deorbit, thus reducing the amount of orbital space debris that must be coped with in outer space. Tether electric power P is given by:

$$P=(vB)^2m/2rd$$

where m is the mass of the conducting tether, r and d are the resistivity and density of the conducting material, and v is the velocity of the spacecraft's motion through the Earth's magnetic field B. For a m=10 kg tether of aluminum with resistivity of r=27.4 nΩ-m and density d=2700 kg/m^3, moving at a velocity v=7037 m/s relative to the Earth's horizontal magnetic field B=26.5 μT, the power dissipated is P=2350 W! This energy loss in the form of heat must necessarily come out of the kinetic energy of the host spacecraft. For a typical example, a 1000 kg spacecraft in a 1000 km high orbit subjected to an energy loss of 2350 J/s from a 10 kg tether (1% the mass of the host spacecraft) will be deorbited in a few weeks.

Power levels of the magnitude estimated in the previous paragraph have been measured in a real orbital space experiment, the TSS-1R mission carried out on the Shuttle Orbiter in 1995. In that experiment, a large Italian spacecraft, 1.6 m in diameter, was deployed upward from the Shuttle Orbiter at the end of a conducting copper wire tether covered with electrical insulation. As the tether was slowly deployed upwards, a series of measurements were made of the open circuit voltage induced in the tether by its motion through the Earth's magnetic field. The voltage between the end of the tether and the Orbiter ground varied from zero volts at the start to 3500 V when the amount of tether deployed approached its maximum length of 20 km. Periodically, the end of the tether was connected either to one of two different electron guns, which supplied contact to the surrounding space plasma, or to the Orbiter ground. The bare surfaces of the Shuttle Orbiter proved to be a surprisingly good plasma contactor via a combination of ton collection and secondary electron emission. The current flow through the tether was deliberately limited by control circuits and the current capacity of the electron guns to about 0.5 amperes, but power levels of 1800 Watts were reached.

The tether was intended to have a fully deployed length of 20 km, but at a deployed length of 19.5 km, when about 3500 V was being induced at the end of the tether inside the Orbiter reel mechanism, a flaw in the insulation allowed an electrical spark to jump in an uncontrolled manner from the tether to the Orbiter ground. With no control circuits to keep the current level down to 0.5 amperes, the current flow jumped to 1.1 amperes, and the total power generated was P=3850 Watts Most of this energy went into the electrical arc, which burned through the tether, causing it to break and halting the experiment. This experiment showed that large areas of bare conducting material, such as that provided by the 8 square meter area of the Italian spacecraft at one end of the tether and the very large surface area of the Shuttle Orbiter spacecraft at the other end of the tether, can collect and emit amperes of current, while thousands of volts of potential can be generated by sufficiently long tethers moving at orbital speeds.

Thus, both theory and experimental data collected in the space environment indicate that significant amounts of electrodynamic drag force can be obtained from a low mass conducting tether attached to a host spacecraft, provided the ends of the conductor can exchange sufficient numbers of electrons with the surrounding space plasma.

Experimental data from the TSS-1R data also produced the amazing result that the efficiency of a bare metal surface in "contacting" the space plasma is many times better than the standard theory would predict. The 8 square meters of bare surface area of the Italian spacecraft were sufficient to collect the 1.1 A of electron current. This amount of area is easily replicated by a few hundred meters of bare wire, considering that the effective collection diameter around the wire is the Debye length, which is a few centimeters at the typical values for space plasma density and temperature.

Because of this result, that a bare wire can easily collect electrons, Les Johnson, Nobie Stone, Chris Rupp, and others at NASA Marshall Space Flight Center have formed a team, which includes the present inventors, which is embarked on a new flight experiment. The experiment is scheduled for a piggy-back flight on a Delta II launch of an AF Global Positioning Satellite in early 2000. The goal of the experiment is to demonstrate that electrodynamic drag from a wire moving at orbital speeds through the Earth's magnetic field will create a large enough electrodynamic drag force to deorbit the Delta II second stage, whose mass is greater than 1000 kilograms, in a few weeks. This is essentially a demonstration of the Loftus electrodynamic drag deorbit concept and the first step in the development of a Terminator Tether™.

The ProSEDS (Propulsion Small Expendable-tether Deployer System) mission will use a 5 km long copper wire conductive tether massing 18 kg connected to a 20–35 km long nonconducting tether, which is in turn connected to a 25–40 kg ballast mass. The total of 25–40 km of tether length and the 25–40 kg ballast mass on the end will provide enough gravity gradient force to keep the tether aligned near the zenith, so that the direction of the current in the tether is at right angles to both the direction of the spacecraft motion in the nominal EW direction and the Earth's near-equatorial magnetic field in the nominal NS direction.

An important feature of the ProSEDS experiment is that it is designed to be completely self-powered. It uses a battery to initiate deployment and to power up the plasma contactor, but once current is flowing through the tether, some of the power is tapped off and used to recharge the battery. The battery, in turn, powers the current control electronics, the telemetry, system, and the plasma contactor. The ProSEDS mission will not be designed to allow ground control changes in operation, primarily because of the increase in complexity and cost associated with that option.

The present invention is the use of a small, low-mass deployer/controller package containing a large collecting area, short length, multiline space tether, such as a Hoytape mesh made of aluminum wire, as a "Terminator Tether™" for a constellation spacecraft. The Terminator Tether™ would be deployed when the host spacecraft is no longer working or no longer wanted. The electrodynamic drag from the Terminator Tether™ would rapidly remove the unwanted spacecraft from the constellation and a few weeks later complete the deorbit of the host spacecraft from space by burnup in the upper atmosphere of the Earth. For a Terminator Tether™ to be of maximum usefulness for constellation spacecraft, it would be desirable to minmize the mass and the length of the tether. A lower added mass means more mass for revenue producing transponders, while a shorter tether length means a lower collision cross-section Area-Time Product during deorbit.

FIG. **6** shows a spacecraft **600** having a state vector **602** that causes the spacecraft to move across magnetic field lines **601**. Spacecraft **600** has a conductive tether **603** attached to it by tether deployer and control system **605**. The outer end of tether **603** is attached to tether end mass **607**. Spacecraft **600** also has an attached electron emitter means **613**.

In FIG. **6**, spacecraft **600** has a velocity and direction defined by state vector **602**. As spacecraft **600** moves it causes conductive tether **603** to cut magnetic field lines **601**. As this happens the distal end of the conductive tether collects electrons, shown as arrows **611**, from the space plasma. Electrons **611** move through the conductive tether **603**, are passed through a resistive load, not shown, in control system **605**, and are emitted back into the space plasma as electrons **615** from electron emitter **613**. This produces an electrodynamic drag in the direction shown by arrow **609**.

As is discussed in detail in the electrodynamic drag analysis below, especially in that sections of the analysis titled "Optimization of Tether Angle", the maximum electrodynamic drag of the tether is achieved when the angle, shown as the Greek letter τ in FIG. **6**, is 35.26 degrees, trailing, to the direction of the magnetic field lines. This angle may be maintained by a feedback control circuit that maximizes the electrodynamic drag of the tether system, as is discussed in detail below. The tether structure **603** is a Hoytether having a length of about 2 to 5 kilometers. The use of the 35.26 degree angle to the magnetic field reduces the tether instability, allowing the use of a short tether which will have a small Area-Time-Product. This will minmize the possibility that the tether will impact another space object during its use.

In FIG. **7**, spacecraft **700** has a state vector **702** across magnetic field lines **704**. Conductive tether **706** is attached to spacecraft **700** by tether deployer and control system **708**. The distal end of tether **706** is attached to a tether end mass **710**. Spacecraft **700** has an electron emitter means **714**.

Functionally, electrons **712** flow from the space plasma into conductive tether **706**, are passed through a resistive load in control unit **708** and are emitted into the space plasma by electron emitter **714** as free electrons **716**. As is discussed in detail in the analysis below, this produces an electrodynamic drag on satellite **700**, which causes it to deorbit.

In FIG. **7**, spacecraft **700** is rotating with an angular velocity indicated by arrow **718**. This rotation causes a centrifugal force **720** to place tension on conductive tether **706**. This force **720** places tension on the tether **706** in a direction that counters tether instability. The Tether **706** may be at any angle to the magnetic field while the rotation of the tether-satellite system is producing this useful tension. FIG. **7** shows the angle to be 35.26 degrees, which is optimum, but any angle will work.

FIGS. **8***a* and **8***b* show a powered Terminator Tether™. In FIG. **8***a* the spacecraft **800** has a state vector **802** that causes it and its associated conductive tether **806** to move across magnetic field **804**. Tether **806** has an end mass **808** that is proximate a plasma contactor **810**. Just as was discussed in the embodiment of the present invention shown in FIG. **6** and FIG. **7**, an electric current moves through the circuit formed by the space plasma, the plasma contactor **810**, the conductive tether **806**, and the electron emitter **815**, thence back into the space plasma as electrons **816**. This creates an electrodynamic drag on the tether in the direction shown by arrow **818**.

In FIG. **8***b*, similar numbers indicate similar structures. In FIG. **8***b*, however, an electric power source, not shown, in spacecraft **800** provides electrons to conductive tether **806**. The electrons are emitted from the end **812** of the tether **806**. Electrons are collected to from the space plasma **814** by contactor **815**, thus completing the electric circuit. The result is an electrodynamic force on tether **806**, and therefore on spacecraft **800**, in the direction show by arrow **820**.

Spacecraft **800** in FIG. **8***b* may be rotated as is taught in the embodiment of the invention shown in FIG. **7**, above, to allow the control system to time the application of electric current to the tether **806** to cause force **820** to be exert force in any desired direction along the circle made by the rotating tether. This allows the state vector of the spacecraft to be modified to raise or lower the spacecraft's orbital altitude, or to otherwise usefully change its orbital elements.

FIG. **9** show an embodiment of the present invention wherein the conductive structure of a satellite is used as the tether. In FIG. **9**, a spacecraft **900** which has a state vector

901 and is moving across magnetic field **903**, has two attached solar panels, **902** and **904**. Solar panels **902** and **904** have conductive aluminum frames **908** and **906**, respectively. Frame **906** is electrically and mechanically connected to plasma contactor **912**. Frame **908** is electrically and mechanically connected to electron emitter **910**. As spacecraft **900** moves across magnetic field **903**, electrons from the space plasma **914** flow through contactor **912** and frame **906** to the body of spacecraft **900**, where the current flows through a resistive load, not shown, and then into frame **908** and electron emitter **910** to the space plasma **916**, thus completing the circuit. This creates an electrodynamic drag on spacecraft **900**. If the power output from solar panels **902** and **904** was directed to flow into the frames, **906** and **908**, then the electrodynamic interaction with the space plasma would cause a useful force to be impressed on spacecraft **900**, as was discussed in connection with FIG. **8** above. The spacecraft may be rotated and/or oriented to a desired angle with the magnetic field lines, as was also discussed above.

FIG. **10** shows an embodiment of the present invention wherein the plasma contactor and the resistive load are structurally part of the conductive tether. In FIG. **10** spacecraft **1000** has a state vector **1001** that causes it to pass through magnetic field **1003**. Spacecraft **1000** has a tether control and deployment system **1005** that is connected to a conductive tether **1007** having an end mass **1013**. A contactor portion **1015** of the conductive tether **1007** near end mass **1013** is adapted to contact the space plasma so as to receive or emit electrons. This may be the bare wire strands of a wide Hoytether, or it may be any other plasma contactor means that can be conveniently made electrically part of the Hoytether structure, such as points or metal fuzz. A second resistive portion **1009** of tether structure **1007** is adapted to be a resistive load. This may be done by making a portion of the tether structure from a material, such as nichrome steel wire, that has suitable electrical resistance The large surface area of the Hoytether provides an excellent radiator structure into the three degree Kelvin radiation sink of outer space.

FIG. **11** shows an alternative embodiment of the load resistance/radiator structure of the present invention. In FIG. **11**, a biline Hoytether **1002** is attached to its control and deployer system **1100** on a spacecraft, not shown, and to a tether end mass **1004**. A resistive load, **1006** is woven into the Hoytether structure, whereby this load **1006** and the section of the Hoytether **1002** into which it is woven act together as a thermal radiator allowing photons **1008** to radiate into the thermal sink of outer space.

The following detailed analysis of the present invention will help those skilled in the art to better understand and use the present invention.

Electrodynamic Drag Analysis:

The following is an optimization analysis of the electrodynamic drag produced by a conducting tether deployed from a host spacecraft that it is to deorbit. The force and torque diagram used in the analysis is shown in FIG. **12**, which shows the forces and torques on an upward deployed conducting tether due to the motion of the host spacecraft through the magnetic field of the Earth.

Physical Constants and Assumptions:

The analysis that follows will use the following physical constants and assumptions:

Newton's gravitational constant $G=6.67\times10^{-11}$ $m^3/kg\text{-}s^2$.

Mass of Earth $M_e=5.976\times10^{24}$ kg.

Radius of Earth near equator $R_e=6378$ km.

Assumed host spacecraft altitude h=622 km.

Assumed host spacecraft orbital radius a=7000 km.

Host spacecraft orbital velocity $w=\omega a=[GM_e/a]^{1/2}=7546$ m/s.

Vertical gravity gradient at spacecraft $2\Gamma=2GM_e/a^3=2.32\times10^{-6}\ s^{-2}$.

Centrifugal gradient at spacecraft $\omega^2=GM_e/a^3=1.16\times10^{-6}\ s^{-2}$.

Combined gradient at spacecraft $3\Gamma=3GM_e/a^3=3\omega^2=3.49\times10^{-6}\ s^{-2}$.

Magnetic field of Earth (tilted dipole approximation):

Magnitude $B_o=35\ \mu T=0.35$ gauss.

Angle between magnetic pole and spin pole 11.5°.

Field components at spacecraft with altitude a and angle β between radius vector of spacecraft and magnetic equatorial plane.

Horizontal $B_H=B_o(R_e/a)^3\cos\beta=26.5\ \mu T\cos\beta$.

Vertical $B_V=2B_o(R_e/a)^3\sin\beta=52.9\ \mu T\sin\beta$.

Electrodynamic Tether Constants and Assumptions:

The electrodynamic tether is assumed to be made of a conducting metal, and have a length L, density d, resistivity r, and cross-sectional area A that is constant along the length of the tether. If the tether is a single round wire of diameter D, then the cross-sectional area is $A=\pi D^2/4$. Because of the micrometeorite and space debris hazard, however, it is likely the tether will be made up of redundantly interconnected multiple lines whose individual cross-sectional areas add up to A. Given these assumptions, the tether mass is then $m_T=dLA$, while the end-to-end tether resistance is $R_T=rL/A=rdL^2/m_T$.

Specific Conductivity Parameter:

The choice of the metal conductor to be used in a space tether is determined by a combination of low resistivity (high conductivity) and low density, with cost, strength, and melting point as secondary considerations for certain applications. Copper has a resistivity r=17.0 nΩ-m, a density d=8933 kg/m, and a "specific conductivity" of 1/rd=6,585 $m^2/\Omega\text{-kg}$. Aluminum has a resistivity r=27.4 nΩ-m, which is significantly greater than that of copper, but it has a much lower density of $d=2700\ kg/m^3$. As a result, aluminum's "specific conductivity" of $1/rd=13{,}500\ m^2/\Omega\text{-kg}$ is twice the conductivity per unit mass of copper. Silver, because of its higher density and higher cost, is not competitive as an electrodynamic space tether even though its resistivity of 16.1 nΩ-m is slightly better than that of copper. An alternate candidate material would be beryllium, with a resistivity r=32.5 nΩ-m, density $d=1850\ kg/m^3$, and a "specific conductivity" of $1/rd=16{,}630\ m^2/\Omega\text{-kg}$, slightly better than that of the much cheaper aluminum. Beryllium also has a higher melting point at 1551 K than aluminum at 933 K, so some of its alloys may be a preferred material for some electrodynamic applications despite its higher materials cost. Unfortunately, despite decades of metallurgical research by the nuclear power industry, highly ductile alloys of beryllium have not been found, so it is difficult to pull beryllium into wire. As a result, because of its high specific conductivity, low cost, and ready availability in ductile wire form, it will be assumed for this analysis that the electrodynamic tether will be made of aluminum wire.

Typical Resistance Values:

To be economically competitive, the mass of the tether needs to be a small fraction of the mass of the host spacecraft it is required to deorbit. Since a typical constellation spacecraft has a mass of about 1000 kg, a typical Terminator Tether™ with a mass that is 2% of the host spacecraft mass would consist of a deployer/controller package with a mass $m_D=10$ kg, containing an aluminum tether with a mass $m_T=10$ kg with a volume of $LA=m_T/d=3.70\times10^{-3}\ m^3$. If this 10 kg of aluminum were formed into a tether with a length of $L=2$ km and a cross-sectional area of $A=1.85\ \text{mm}^2$, then the end-to-end resistance of the tether would be $R=rL/A=rdL^2/m_T=29.6\ \Omega$ A longer tether would have a proportionately smaller cross-sectional area and a higher resistance; for example, a 5 km long tether with the same mass would have a resistance of 185 Ω.

Orbit Inclination Assumption:

In order to make the remainder of the analysis mathematically tractable, it will be assumed that the orbit of the host spacecraft is circular and above the magnetic equator, so that the angle between the radius vector to the spacecraft and the magnetic equatorial plane is $\beta=0$ degrees. In this orbit, the velocity v of the spacecraft with respect to the magnetic field is the orbital velocity w less the rotational velocity of the magnetic field at the orbital radius a due to the once per day rotation of the Earth, or:

$$v=W-2\pi a/1\text{day}=(7546-509)\text{m/s}=7037\text{m/s}$$

This equation also shows why electrodynamic drag will not be useful for removing geostationary spacecraft from orbit. At the geostationary orbital radius of $a=42{,}200$ km, the relative velocity of the spacecraft and the rotating magnetic field of the Earth is zero.

With this assumption of an orbit above the magnetic equator, the vertical component of the Earth's magnetic field is zero and the horizontal component of the Earth's magnetic field is at right angles to both the local vertical and the direction of motion of the spacecraft. The total magnetic field seen by the host spacecraft and its Terminator Tether™ is then orthogonally horizontal and has the magnitude:

$$B=B_H=B_o[R_e/a]^3 \cos \beta=26.5\mu T.$$

Where for simplicity the subscript H will be dropped for the remainder of this analysis.

This orbit, with its inclination of 11.5° with respect to the spin equator, will not stay in the plane of the magnetic equator, but will vary ±11.5° above and below it as the Earth rotates, causing a variation in magnetic field strength of $\pm 0.5\ \mu T$, as well as a slight variation in angle. This 2% variation is negligible in terms of the other uncertainties in this analysis. It is not until the orbital inclination of the host spacecraft orbit reaches 60°, where $\cos 60°=0.5$, that there is a significant drop in the expected magnitude of the electrodynamic drag forces calculated in this analysis.

Electromagnetic Drag Effects in Polar Orbit:

In many medium Earth orbit communication satellite constellations, there are a significant number of spacecraft at high inclinations and in nearly polar orbits. The high inclination spacecraft, with inclinations between 60° and 78.5°, will all have orbits that stay between the magnetic poles. Although the amount of electrodynamic drag will be significantly less than that experienced by spacecraft with orbits at lower inclinations, the direction of the induced electric fields in the tether will always be in the proper direction. If the tether is vertically upward, the outer tip of the tether will be positively charged and the bare wire in the tether will pull electrons out of the surrounding space plasma, while the electron emitter at the host spacecraft end will eject the electrons back into the space plasma to complete the circuit.

For spacecraft in near polar orbits with inclinations between 78.5° and 90°, however, there will be much more drastic variations. First of all, for a spacecraft in a 90° orbit that happens to pass directly over the magnetic poles, there is no horizontal component of the magnetic field when it is passing over one of the magnetic poles, so no voltage is generated in the tether during that part of the orbit, while the horizontal component of the magnetic field near the magnetic equator, while strong in magnitude, is along the spacecraft velocity vector, so no voltage is generated in the tether in that part of the orbit either. As a result, there will be negligible electrodynamic drag experienced by the host spacecraft during that particular orbit. The Earth is rotating, however, and the magnetic pole is rotating with it. A few orbits later, the host spacecraft will be passing over the Earth's spin pole at a point where the horizontal component of the magnetic field is exactly at right angles to the direction of motion of the spacecraft, so the full voltage is generated in the tether. When passing over the spin pole, the spacecraft radius vector is at an angle of 11.5° from the magnetic pole and at an angle $\beta=78.5°$ away from the magnetic equatorial plane. With these assumptions, the magnitude of the horizontal component of the magnetic field at the Earth's spin pole is a respectable:

$$B=B_o[R_e/a]^3 \cos 78.5°=5.28\mu T$$

or 20% of the maximum value experiences by spacecraft orbiting above the magnetic equator. This value will drop slightly as the orbit continues, then build back up as the spacecraft passes over the opposite pole. As the Earth continues to rotate, bringing the magnetic pole again under the orbit, and the interaction of the tether with the horizontal component of the magnetic field again drops to zero. The average coupling of a tether to the Earth's magnetic field over all polar trajectories has been calculated to be 12.3 percent.

A problem experienced only by spacecraft with orbit inclinations greater than 78.5°:

The spacecraft will no longer be traveling from west to east with respect to the magnetic field axis, but will have a retrograde motion as it moves through the magnetic field. As a result, the voltage generated by the motion of the tether through the Earth's magnetic field will switch direction. The outer tip will be negatively charged and will attempt to collect ions, which is a much less efficient process than collecting electrons.

There are a number of solutions to this problem. The first is to increase the mass and length of the tether supplied to a spacecraft assigned to a polar orbit, so that higher voltages, currents, and drag are generated during the limited times the spacecraft is passing over the poles in the right direction. The second is to supply a tether with plasma contactors at both ends that can emit electrons from either end, allowing the current to flow either way, depending upon which direction the spacecraft is passing around the magnetic pole. The third is to utilize the first number of passes to torque the orbit of the spacecraft until the orbit inclination has been shifted below 60°, then turn on the electrodynamic drag full time to deorbit the spacecraft from this more favorable orbit inclination. This orbit torqueing maneuver is accomplished by activating the electrodynamic drag mechanism only when the magnetic field orientation is such that a strong out-of-plane component of force is created. This orbit torqueing maneuver can be augmented by switching to a propulsion mode, where power saved in the batteries during the drag force mode is pumped back into the tether when the magnetic field is in the opposite direction, applying electrodynamic propulsion to torque the orbit even further. With the tether at a large and stable trailing orientation, the coupling to the magnetic field can be significant to a vertical tether.

In summary, spacecraft in near-polar orbits might take longer to bring down, and might have to utilize specially designed Terminator Tethers™ that might cost and mass more than the simpler Terminator Tethers™ usable in lower inclination orbits, but a Terminator Tether™ can still remove a spacecraft from a polar orbit when desired.

The inventors now present a detailed analysis and optimization of a typical Terminator Tether™ attached to a typical host spacecraft in a typical low inclination orbit.

Electromotive Generation of Voltage and Current in the Tether:

When an object is moved at a velocity v through a magnetic field B, an electric field is generated in the frame of reference of the moving object given by:

$$E=v\times B=vB$$

where the magnetic field B of the Earth, being mostly tangent to the Earth's surface in the north-south direction, is at right angles to the velocity vector v of the spacecraft, assumed to be orbiting in a generally west-east direction. The direction of the electric field E will be at right angles to both v and B, or along the local vertical. It should be noted that this electric field exists in the moving frame of reference because a moving magnetic field creates an electric field. No object actually has to be there, but if it is, then the relative motion of the magnetic field of the Earth will not only apply magnetic forces to whatever material the object is made out of, but electric forces too.

Note also that the velocity used in this equation is the relative velocity between the object and the magnetic field. Because the Earth's magnetic field rotates with the Earth, the motion of the magnetic field must be subtracted from the orbital velocity of the object to obtain the relative velocity.

Voltage generation in a Conducting Tether:

If the moving object is a long conducting wire of length L, the electric field E, generated in the wire produces a voltage V between the opposite ends of the wire given by:

$$V=E\bullet L=EL\cos\tau=vBL\cos\tau$$

where τ is the angle between the length vector L of the tether and the electric field vector E, assumed to be in the vertical direction at right angles to the velocity vector v in the plane of FIG. **12** and the magnetic field vector B out of the plane of FIG. **12**. A typical value for the voltage level built up in a vertically oriented tether with length L=5 km and $\tau=0$, moving at a velocity of v=7037 m/s through the Earth's horizontal magnetic field of strength 26.5 μT, is 932 V, or 0.186 V/m. Spacecraft in higher inclination orbits would experience somewhat smaller electric fields. For calibration, the voltage measured between the ends of the 19.5 km long TSS-1R tether was 3500 V, or 0.175 V/m.

Contacting the Space Plasma:

Although a voltage will build up between the ends of the conducting tether, no current will flow unless the circuit is completed. The circuit cannot be completed with another wire, for it too will have a similar voltage generated in it by the moving magnetic field. Fortunately, empty space is not empty, and in near-Earth regions not too distant from the Earth's atmosphere there exists highly electrically conductive space plasma, kept partially ionized by radiation from the Sun. The electron and ion density varies from the dark to light side, with altitude, with season, with sunspot cycle, with contamination level, etc. but it typically varies from 10^{11} to 10^{13} electrons/cc.

Fortunately, it has been found that if a bare conductive surface such as the spherical Italian spacecraft in the TSS-1R experiment, or the long bare wire to be used in the ProSEDS experiment, or a multiline conducting wire Hoytape mesh, is charged to a few hundred volts, the bare conductor will readily pull electrons out of the space plasma. So all that is needed to complete a connection to the space plasma at the positively charged (upper) end of the tether is a sufficient large area of uninsulated conductor.

It is more difficult to eject electrons from a wire or to collect positive ions from the space plasma. Although it is conceivable that a very large area at the other end of the tether could collect enough ions to complete the circuit as was demonstrated during the TSS-1R mission when the Shuttle orbiter was found to be an adequate plasma contactor for over an amp of current, the present method chosen is to use an electron emitter of some sort, either a hot cathode, a plasma cathode or contactor, a field-emission device, or something similar. Once provisions have been made at both ends of the tether to allow the flow of electrons out of one end of the tether and into the other end, and the altitude of the host spacecraft is not too high, then there will be sufficient conductivity in the space plasma surrounding the host spacecraft to allow current to flow through the tether.

Current Flow In The Tether:

The amount of current I flowing through the tether depends upon the total resistance R in the circuit. This resistance will consist of three components, the effective resistance of the plasma, the resistance of the tether, calculated earlier as $R_T=rL/A=rdL^2/m_T$, and a control resistor R_C, which will be varied as needed to optimize the Terminator Tether™ performance. There will also need to be a parasitic load on the current in the form of a charging device to charge some batteries. The batteries in turn will be used to power the control and communication circuits, and drive the electron emitting devices at the negative end of the tether. A well-designed Terminator Tether™ will thus be completely self-powered, except for an initial charge in the batteries to provide electrical power for the deployment and startup procedure. For simplicity of analysis it will be assumed that this battery charging load, which absorbs power like a resistor, but which stores it and uses it later instead of dissipating it immediately as heat, is included in the control resistor R_c. Normally, this load would act as a "base resistance" below which the control resistor could not be lowered, but since the charging circuit can be turned off when desired, and operations continued without interruption using the power stored in the batteries, it will be assumed that the control resistor can be taken to near zero value in those circumstances where the space plasma conductivity is low, or the magnetic field is in the wrong orientation and the voltage being generated in the tether is not large. Under these circumstances, lowering the control resistor to near zero allows a much higher current to flow for a given generated voltage, thus increasing the power being dissipated in the tether and maintaining a high level of electrodynamic drag on the host spacecraft.

A properly designed Terminator Tether™ will have plenty of bare metal area for electron collection at the positive end, while the electron emitters at the other end are efficient in terms of emitting large electron currents at low voltages and therefore low powers, while at the same time the mass, length, and area of the conducting tether have been made such that the resistance of the tether is moderately high. Under these conditions, the effective resistance of the space plasma will be much less than the design resistance of the tether. To make the mathematics more tractable it will be assumed that the plasma resistance is can be neglected and that the voltage available to drive the tether is the full voltage generated by the moving magnetic field. Although voltage will be needed to power the electron emitter, whatever its form, it will be assumed that the required voltage and power will be supplied by the batteries. The batteries in turn will be powered by the battery charging circuit, whose effective resistance is included in the control resistor (except for those short periods when the plasma or magnetic field interactions are weak).

Given all these assumptions concerning the total resistance in the circuit, the current I flowing through the tether is then given by:

$$I=V/R=vBL\cos\tau/(R_C+rdL^2/m_T)$$

A typical current level can be estimated by assuming that the control resistance R_C should be about the same size as the tether resistance R_T, which for an aluminum tether with resistivity r=27.4 nΩ-m, density d=2700 kg/m^3, length L=5 km, and mass m_T=10 kg, is R_T=185 Ω, giving a total resistance, including the control resistance R_C, of R=370 Ω. Thus the current flowing though a tether of length L=5 km, generating a voltage of V=932 V, would be I=V/R=2.52 A.

Power Dissipation Analysis:

The induced current I flowing through the tether resistance R_T will generate heat in the wire, which will be radiated into space and lost. That radiated energy must come from somewhere. There is also power taken out of the current flow by the battery charging circuit and ultimately dissipated. In order to conform to energy conservation laws, this power and energy must come from a decrease in the total kinetic and potential energy of the host spacecraft, causing it to deorbit. Set forth below is a detailed, optimized force analysis which will calculate the drag force on the host spacecraft. In this section, by using some simplifying assumptions, it is possible to produce a general argument based on energy conservation laws, which will illustrate the broad scope of the present invention.

For this energy conservation analysis it will be assumed that the plasma resistance is small compared with the tether resistance, and that the control resistance is equal to the tether resistance:

$$R_C=R_T=rL/A=rdL^2/m_T$$

Assume that the ballast mass at the end of the tether is a large piece of the defunct host spacecraft, such as a solar panel, antenna, or battery pack, so that the gradient force is large and the tether is always oriented along the vertical so that the angle τ=0°. Also assume the spacecraft is in an orbit above the magnetic equator and thus is moving at right angles to the Earth's magnetic field, so that the angle between the orbital velocity vector and the magnetic field vector is 90°, and the electric field vector is in the vertical direction, aligned perfectly with the tether length vector. Under these ideal conditions, the voltage between the ends of the tether is given by:

$$V=E\bullet L=(v\times B)\bullet L=vBL$$

The current in the tether is then just:

$$I=V/(R_C+R_T)=vBm_T/2rdL$$

The power dissipated as ohmic heating in the tether is then given by:

$$P=IV=(vB)^2m_T/2rd$$

This equation shows the interesting result that the power dissipated does not depend upon the length or the area of the tether (within reason), but only the mass of the tether m_T, the resistivity r and density d of the tether material, and the velocity v of the conducting tether through the Earth's magnetic field B.

For a typical aluminum tether of mass m_T=10 kg, the power dissipated is an impressive P=2350 W. Even if only a small fraction of this power is dissipated in a real system, the kinetic energy of the host spacecraft would suffer a significant energy loss with time.

Orbital Energy Extraction Analysis:

The total energy U of a spacecraft of mass M in orbit around the Earth consists of two components, its positive orbital kinetic energy and its negative gravitational potential energy. If a circular orbit of altitude h and radius a=R_e+h is assumed, then the total energy of the spacecraft moving at a velocity w=$(GM_e/a)^{1/2}$ is:

$$U=+Mw^2/2-GMM_e/a=-GMM_e/2a$$

The system will be dissipating energy in the tether to decrease the energy U of the unwanted spacecraft from its relatively low negative value in a high Earth orbit to a greater negative value in a lower Earth orbit. To give a specific example, an unwanted spacecraft with a mass of M=1000 kg (a metric ton) in an orbit with an altitude of h=1000 km and a high orbital radius of a(H)=R_e+h=7371 km, then its (negative) total energy is:

$$U(H)=-GMM_e/2a(H)=-27.1\text{ GJ or }-27.1\text{ MJ/kg}$$

It is desired to lower the spacecraft to an orbit with an altitude just outside the atmosphere of h=200 km or a lower orbital radius of a(L)=R_e+h=6571 km, where atmospheric drag will finish the job. The total (negative) energy of a metric ton spacecraft at 200 km altitude is:

$$U(L)=-GMM_e/2a(L)=-30.4\text{ GJ}$$

Thus, the amount of energy needed to be dissipated in order to move the 1000 kg spacecraft from a 1000 km orbit with a total energy U(H)=−27.1 GJ down to a 200 km orbit with energy U(L)=−30.4 CJ is:

$$dU=U(H)-U(L)=3.3\text{ GJ.}$$

The decay time T of a metric ton spacecraft moving from a 1000 km altitude orbit to a 200 km altitude orbit with an energy difference of dU=3.3 GJ, when its energy is being dissipated at a power of P=2350 W by a typical aluminum tether massing just 10 kg or 1% the mass of the spacecraft, is found to be T=dU/P=16 days.

The actual decay time will be longer than this. If the electrodynamic drag force is very large, and becomes larger than the gravity gradient forces pulling on the ends of the tether, then the tether will tend to align itself along the magnetic field lines instead of across them, and the drag force will decrease because of the small angle between the conductor length and the magnetic field lines. The tether current will need to be controlled until the angle of the tether settles into an angle determined by the balance between these two forces. These electrodynamic torque vs. gradient torque dynamic interaction effects are discussed in the next section.

Force and Torque Balance Analysis:

It is now possible to calculate the forces and torques on the tether. It is also possible to use the fact that the electrodynamic and gravity forces and torques on the tether must balance each other out to calculate and teach some optimum values for some of the Terminator Tether™ parameters.

Electrodynamic Force and Torque:

As discussed above, both theory and experiment show that: provided the conducting tether is moved rapidly through the Earth's magnetic field in order to generate a voltage across it, and provided good contact is made with the space plasma, the conducting tether will have a current flowing through it. When a wire (moving or not) carrying a current I is embedded in a magnetic field B, there will be an electrodynatnic force F_E generated on each element of the wire. The electrodynamic force will be at right angles to the magnetic field vector and the length vector of the wire element, with a magnitude given by:

$$F_E = J \times B = ILB$$

where B is the horizontal component of the magnetic field, which is perpendicular to the assumed magnetic equator orbit plane, while the tether length vector L is assumed to lie in the orbit plane. If, as will be the case most of the time, the electron current is leaving the space plasma and entering the tether along the length of the tether, then IL needs to be replaced with an integral of the current along the length of the tether.

Note that the electrodynamic force amplitude or direction is not directly dependent on the motion of the tether through the Earth's magnetic field. The electrodynamic force would be the same if the tether were not moving and the current was being supplied by a battery. Since the current I is a function of the orbital velocity, however, there is a secondary dependence of the electrodynamic force magnitude on the orbital velocity, but not its direction. The electrodynamic force is always at right angles to the conductor, and stays at right angles to the conductor as the angle τ varies, as shown in FIG. **12**.

Assuming that the electrodynamic drag force is applied uniformly along the length of the tether, it is possible to make the simplifying assumption that the integrated force is effectively applied at right angles to the center of mass of the tether at the point L/2 as shown in FIG. **12**. The electrodynamic torque on the tether is then:

$$T_E = F_E L/2 = IBL^2/2 = vB^2L^3 \cos \tau/2(R_C + rdL^2/m_T) = vB^2L^3 \cos \tau/2R$$

Gravity Gradient Forces and Torques:

When a tether and its ballast end mass are deployed from a host spacecraft, the gravity gradient force field of the Earth, combined with the orbital centrifugal gradient force field, will cause the tether to deploy either up or down from the host spacecraft. The direction desired depends on which end of the tether is connected to the electron emitter. Normally, the electron emitter will be on the end attached to the host spacecraft, in which case the desired direction of deployment will be upward so that the induced voltage in the tether will produce an excess of electrons at the electron emitter end of the tether. The desired upward direction is chosen by having the deployer eject the ballast mass in the upward direction. Once the ballast mass has been started in that direction, the centrifugal force due to the orbital motion of the ballast mass will cause the ballast mass to continue to accelerate in the upward direction until it is brought to a halt by the full deployment of the tether.

If there were no electrodynamic or atmospheric drag, the equilibrium direction of the tether would be exactly along the vertical, since the combined gradient field is a maximum in that direction. Because a significant amount of electromagnetic drag is expected, the actual angle of the tether with respect to the local vertical will be at some angle τ, lagging behind the spacecraft motion in the plane of the orbit, as shown in FIG. **12**. In the following analysis shows that there is an optimum angle for τ that produces the largest electrodynamic drag force on the host spacecraft without producing tether instability.

The combined vertical gravity gradient and centrifugal gradient field 3Γ acting on the ballast mass m_B at the end of the tether of length L will produce a gradient force F_{GB} given by:

$$F_{GB} = -3_{\Gamma} m_B L \cos \tau$$

The strength of the force depends not only on the ballast mass m_B and the strength of the gradient field 3Γ, but also the component along the radial direction of the distance of the ballast mass from the center of mass, which is L cos τ. This force acts in the vertical direction along the radius vector leading from the ballast mass away from the center of the Earth. The amount of gradient force is not large. For a ballast mass of m_B=10 kg and a tether length of L=5 km, the gradient force is about 0.175 N.

As can be seen in FIG. **12**, the component of this gradient force that is at right angles to the tether, given by F_{GB} sin τ, will produce a torque T_{GB} on the tether that tends to restore the tether toward the vertical, lessening the angle τ.

$$T_{GB} = LF_{GB} \sin \tau = -3_{\Gamma} m_B L^2 \sin \tau \cos \tau$$

The tether mass m_T also contributes to the gradient force and torque. If it is assumed that the tether has a uniform cross section, then it is possible to replace the distributed mass of the tether with an equivalent point mass m_T placed at the center of mass of the tether, which is the point L/2 along the tether, and a distance L/2 cos τ in the radial direction. The gradient force due to the tether mass is then:

$$F_{GT} = -3/2_{\Gamma} m_T L \cos \tau$$

While the gradient torque is:

$$T_{GT} = -L/2F_{GT} \sin \tau = -3/4_{\Gamma} m_B L^2 \sin \tau \cos \tau$$

The total gradient torque attempting to restore the tether to its vertical orientation is then:

$$T_G = T_{GB} + T_{GT} = 3_{\Gamma}(m_B + m_T/4)L^2 \sin \tau \cos \tau$$

It is important to notice the variation of the total gravity gradient torque as the tether angle τ is varied. Since the gradient force is always in the radial or vertical direction, there is no torque on the tether when the tether is vertical, as is the case when there are no aerodynamic or electromagnetic drag forces. Once the drag forces become important and start to apply torque to the tether, increasing the tether angle τ, those drag torques causing an increase in tether angle τ will be opposed by a rising gradient torque which will attempt to decrease the tether angle. For small tether angles, the gradient torque increases first linearly with τ, then as sin τ, since cos τ is near unity for small τ, then at τ=45° the gradient torque reaches its maximum, where sin τ=cos τ=0.707 and sin τ cos τ=0.50. When this angle is reached, the tether is a point of catastrophic instability, for if there is a further increase in the electrodynamic drag force, causing the angle τ to become greater than 45°, the gradient torque, instead of growing stronger to counteract the increased drag torque, will become weaker and the tether angle will go rapidly to 90°.

To restore control to the tether angle if the instability occurs, it will be necessary to turn off the electrodynamic drag forces by shutting off the current flow through the tether. The τ=90° position for the tether and ballast mass is a gravitationally unstable orientation. After a time, slight fluctuations in the gravity field will allow the gradient force to slowly take over and restore the tether to the vertical orientation, which, unless it can be controlled in some way, is equally likely to be up or down. It would therefore be desirable to maintain control of the tether angle so as to avoid the tether angle getting into the region of instability.

Torque Balance on the Tether:

The angle τ of the tether is determined by a balance between the electrodynamic torque T_E attempting to increase the angle τ and the gradient torque T_G attempting to decrease the angle τ. Balance is achieved when the two torques are equal:

$$T_E=T_G=T_{GB}+T_{GT}$$

or:

$$vB^2L^3 \cos \tau/2R=3_{\Gamma}(m_B+m_T/4)L^2 \sin \tau \cos \tau$$

Simplifing and rearranging yields an equation giving us the angle τ at which torque balance occurs:

$$\tau=\arcsin[vB^2L/6_{\Gamma}(m_B+m_T/4)R]$$

The physical interpretation of this equation is that the maximum electrodynamic force that can be sustained on the tether is limited by the gradient force on the tether. It is possible to increase the electrodynamic force by decreasing the total resistance R of the tether (for example, by using a tether with a larger mass and a lower tether resistance), but if the tether resistance is too low, then the quantity in brackets becomes greater than unity and this equation has no solution, indicating that the tether has reached an angle where instability sets in.

To maintain control of the tether angle, it will be necessary to vary the control resistance of the tether to compensate for variations in magnetic field strength and direction, plasma density (which affects the plasma resistance), and other factors, and thereby maintain the tether at an intermediate angle where both the electrodynamic and gradient forces are at an appreciable level and balance each other.

Optimization of Tether Angle:

At first glance, it might seem that the optimum angle for the tether would be 45°, since at that angle the gradient torque is largest and therefore can counteract a larger electrodynamic drag force. However, since the 45° angle is the point where instability sets in, it is desirable to set the tether angle at some value below 45°. The optimum angle is that which maximizes the horizontal or drag component of the electrodynamic force. This optimum angle τ is derived from the equation for the horizontal component of the electrodynamic force, or the electrodynamic drag force, since it opposes the host spacecraft motion:

$$F_{ED}=F_E \cos \tau=6[m_B+m_T/4]\Gamma L \sin \tau \cos^2\tau$$

By setting the partial derivative$\square F_{ED}/\square\tau=0$ and solving, it is possible to calculate that the optimum angle for the tether that gives the maximum electrodynamic drag force F_{ED}, while still keeping the tether torques balanced and under control, is τ=arctan(0.707)=35.26°. With this angle selected and maintained by varying the control resistor R_C to compensate for variations in plasma contact resistance and variatins in the strength and direction of the Earth's magnetic field B seen at the spacecraft, the tether experience the maximum stable value for the electromagnetic drag force of:

$$F_{ED}(\max,\tau=35.26°)=2.31[m_B+m_T/4]\Gamma L$$

There are many ways to provide the sensing information needed to provide the feedback signals to the control resistor, but the simplest is to merely measure the drag acceleration on the host spacecraft with a set of accelerometers, and maximiize the force in the direction opposite to the host spacecraft motion. Another method would be to measure the current in the tether, and knowing the tether resistance and the amount of control resistance, calculate the power being extracted and maximize that value. Alternate methods would be to use an optical position sensor or GPS receivers at both ends of the tether to measure the angle of the tether or the position of the ballast mass with respect to the host spacecraft.

To make an estimate of the magnitude of drag force attainable, assume that a typical aluminum tether with a length of L=5 km, a tether mass of m_T=10 kg and a ballast mass of m_B=10 kg is being used. Then the maximum gradient force limited electrodynamic drag force is F_{ED}=0.167 N. The power being dissipated in the canoical tether at the angle τ, with $R_C=R_T=rdL^2/m_T=185\ \Omega$ is given by:

$$P=IR=[vBL \cos \tau]^2(R_C+rdL^2m_T)$$

which for an angle τ=35.26° results in a power dissipation of 1570 W. Thus, the thrust level per unit power obtained by the typical Terminator Tether™ would be about 0.106 N/kW. This number is comparable to the value of 0.148 N/kW estimated for the much heavier and longer TSS-1R tether.

Reconciliation of Energy and Force Analysis:

By a force and torque balance analysis an optimum angle for the tether has been discovered at which it is possible to obtain a maximum in the drag component of the electrodynamic force. There is also an additional component of the electrodynamic force, the component in the vertical direction, $F_{EV=FE}$ sin(τ), which is downward for an upwardly deployed tether. This component of force combines with the gravity force of the Earth to effectively allow the host spacecraft to orbit a little faster than normal for that orbital altitude. It does not contribute to the deorbiting of the host spacecraft. But since this vertical component of force is created by current running through the tether, and that current is creating heat and dissipating energy as it passes through the tether, there might be some concern that the force and torque balance analysis above does not conform to the law of conservation of energy.

The inventors will now show, in a very general manner, that no matter what the tether angle, the electrical power being dissipated in the tether is exactly equal to the power being lost by the slowing of the host spacecraft.

The "deorbit power" P_D that must be removed from a spacecraft moving at a velocity v when that motion is opposed by a drag force F_{ED} is:

$$P_D=v\bullet F_{ED}=vF_E \cos \tau=vILB \cos \tau=I[vBL \cos \tau]$$

But, since the voltage V induced across the tether of length L and tilt angle τ moving at a velocity v through a horizontal magnetic field B is

$$V=E\bullet L=EL \cos \tau=vBL \cos \tau$$

Therefore:

$$P_D=IV=P_E$$

And the deorbit power P_D extracted from the slowing of the spacecraft by the drag component of the electrodynamic force is always exactly equal to the electrical power P_E being dissipated as heat in the tether circuits, independent of the tether angle.

Optimization of Tether Mass Distribution:

If it is not possible to use a piece of the host spacecraft as ballast mass, then the mass of the ballast must be included in the Terminator Tether™ mass. It would be desirable to minimize the total Terminator Tether™ mass, since every kilogram saved means that another kilogram's worth of revenue-producing transponders can be added to the working payload of the host spacecraft. Given a total mass for the Terminator Tether™ and the mass of the deployer/controller unit, it is possible to optimize the mass distribution between the ballast mass and the tether mass to obtain a minimum total Terminator Tether™ mass. A well-designed Terminator Tether™ will also have most, if not all, of the deployer mass incorporated into the ballast mass.

Assume that the total Terminator Tether™ mass consists of the deployer/controller mass m_D, the tether mass m_T, and the ballast mass m_B, with $m_B > m_T > m_D$. Of the three mass components in the Terminator Tether™, two of them affect the electrodynamic drag performance. If it is assumed that the ballast mass is a factor X larger than the tether mass, or $m_B = Xm_T$, the maxinum drag force that can be obtained is now:

$$F_{ED}(\max)=2.31m_T(X+0.25)\Gamma L$$

If it is assumed that the control resistor has been adjusted so that this maximum value for the electrodynamic drag force is maintained as the motion of the spacecraft along its orbit moves the spacecraft into regions with different magnetic field strengths and plasma densities, then for the angle $\tau=35.26°$, $\cos(35.26°)=0.817$, and the maximum electrodynamic drag force in terms of the electrodynamic parameters will be:

$$0.817vB^2L^2/(R_C+rdL^2/m_T)=F_{ED}(\max)=2.31m_T(X+0.25)\Gamma L$$

or canceling out terms and rearranging:

$$(X+0.25)=0.353vB^2L/\Gamma m_T(R_C+rdL^2/m_T)$$

It is now possible to make the further simplifying assumption that to maintain control of the tether, the control resistor needs to be roughly the same size as the tether resistance or $R_C \approx rdL^2/m_T$. Using this assumption produces an equation for the optimum value for the ratio X of the ballast mass to the tether mass:

$$\begin{aligned} X = vB^2/\Gamma rdL - 0.25 &= -0.25 + \frac{vB^2}{12\sin\gamma\cos\gamma\Gamma rdL} \\ &= -0.25\frac{\Lambda}{L} \end{aligned}$$

Where Λ is the "effective electrodynamic length".

If the tether is short, so that L is small, this equation indicates that the ballast mass must be increased to increase the gradient force, since it is the gradient force which determines the upper limit to the amount of electrodynamic drag force that can be generated without losing control of the tether.

Assuming the usual values for the Earth's magnetic and gradient fields, the velocity of the host spacecraft through the magnetic field, and the resistivity and density of aluminum, this equation becomes:

$$X=10.9/L-0.25$$

where L is in kilometers.

For a typical tether of length L=5 km, X=1.93. If the total mass for the Terminator Tether™ is 20 kg, or 2% of the host spacecraft mass, and the mass of the deployer is 10 kg, then the remaining 10 kg should be distributed so that 6.59 kg is in the ballast mass while 3.41 kg is in tether mass.

Optimization of Tether Length:

If the ballast mass consists of a piece of the host spacecraft, then the mass of the ballast does not have to be considered in the optimization of the Tether system. In this case, the optimum distribution of the Tether mass is to put as much mass as possible into the tether, with a minimum in the deployer/controller package. Under this assumption, the optimum length of the tether is determined by the desire to keep the Area-Time Product of the host spacecraft plus Tether to a minimum.

In a long tether of length L, the effective "collision" cross-sectional area is not the area of the tether, but the larger area produced by multiplying the length of the tether by sum of the width of the tether plus the width of the "target" spacecraft that would be damaged by the tether hitting it. Thus, to decrease the "area" portion of the Area-Time Product, it would be desirable to shorten the tether, making it thicker, and perhaps slightly wider, at the same time, thus keeping the same tether mass and electron collection area, and thereby maintaining the drag force and keeping the decay time at the same level. The result will be a shortening of the Area-Time Product.

If the tether is shortened too much, however, the smaller voltage generated across the shorter tether will leave less voltage margin for the operation of the plasma contactors and the battery recharge system. Since the maximum voltage that can be generated is about 200 V/km, and typical plasma contactors and battery, chargers take 25–100 V to operate, a minimum length for a Tether would be roughly 2 km. Host spacecraft operating in polar orbits, where the conditions for voltage generation are not as good, may require a tether length of 5 km or more.

Since many watts of heat power will be dissipated in the control resistor, means must usually be provided to radiate the heat away into space. One low-mass method of accomplishing this is to make the control resistor (or resistors) in the form of a long, electrically insulated, high resistivity wire similar to those used in electric blankets, but designed to operate at a higher temperature, and during the fabrication of the electrodynamic tether, incorporate the high resistivity wire (or wires) into the end of the tether closest to the control circuit by weaving or braiding the insulated high resistivity wire in with the uninsulated aluminum wire of the electrodynamic tether. The surface area of a long wire is very large, so in this extended wire configuration, exposed to the space all around it, the hot control resistor wire can self-radiate its power into cold space without incurring the mass penalty of a separate radiating surface.

Implementation:

The basic optimum structure for a Terminator Tether™ would be one of the many types of Hoytethers. A multiline (6–12 primary line) Hoytape™ will provide the largest contact area with the plasma, since both sides of the tape would be able to pass current to the plasma. If the spacing between the primary lines is chosen to be larger than twice the Debye length of the plasma, then the effective current collection area per unit length of the Hoytape is proportional to the width of the Hoytape mesh, not the diameter of the wires in the mesh. Thus, a Hoytape not only provides an assured longer life for the Terminator Tether™, but very short lengths will also provide a very large current collection areas.

The deployer for the tether can deploy the Terminator Tether™ either down or up or both. The deployer can stay attached to the spacecraft as was done in the SEDS missions, or perhaps a better alternative would have the deployer ejected from the spacecraft, with one end of the tether still attached to the spacecraft, reeling out tether as it leaves. The empty deployer would then act as a ballast mass at the end of the Terminator Tether™, which would improve the performance.

The "Remora Remover"

In addition to attaching Terminator Tethers to spacecraft before launch, it is possible to consider a missile-like application. This "Remora Remover" missile would use a Terminator Tether™ carried by a seeker missile similar to the small "hit-to-kill" missiles developed by the Space Defense Initiative Office that deployed a net loaded with oriented metal rods. The Remora Remover missile would hunt down a spacecraft that needs to be removed from space, but instead of hitting the spacecraft, the missile would be programmed to rendezvous with the spacecraft and attach itself to the host spacecraft using a hooked net, harpoon, or adhesive "sucker". The Remora Remover missile would then deploy the Terminator Tether™, which would bring down both the spacecraft and the missile.

The inventors have presented two analyses, backed up with data obtained from space flight experiments. One analysis was based on generalized energy conservation laws while the other analysis use force and torque balance arguments. Both analyses show that optimized aluminum wire tethers 2 to 5 km in length and massing just 1% to 5% of the mass of the host spacecraft can deorbit the host spacecraft in the order of a few weeks, thus mitigating the long-lived orbital debris hazard created by a constellation spacecraft after their end-of-life.

Power Augmented Operation of the Tether:

It is well known in the aerospace literature that if a spacecraft has a power supply and a conducting tether hanging from it, and the spacecraft is in an orbit or trajectory that takes it near a body, such as the Earth or Jupiter, which has both a magnetic field and a region of moderate density electron-ion plasma, and electrical current is pumped from the power supply through the tether and back through the plasma, that the current flowing through the tether will interact with the magnetic field of the body, producing forces on the tether and thence on the spacecraft, changing its orbit or trajectory. These forces can be used to increase or decrease the spacecraft altitude and/or inclination. The amount of altitude or inclination change is proportional to the ratio of the power available from the power supply divided by the mass of the spacecraft and varies with altitude. The unpowered Terminator Tether™ is a simple example of such a system, in which only altitude decrease is possible, although small amounts of both inclination increase or decrease are possible at the same time. Some typical references to this prior art are: Les Johnson, "Propulsive Small Expendable Deployer System Mission (ProSEDS)", Proceedings of the OAST 8^{th} Advanced Propulsion Workshop, JPL, Pasadena, Calif., 20–22 May 1997; and Les Johnson, Joe Carroll, Robert D. Estes, Enrico Lorenzini, Brian Gilchrist, Manuel Martinez-Sanchez, Juan Sanmartin, and Irwin Vas, "Electrodynamic Tethers for Reboost of the International Space Station and Spacecraft Propulsion, AIAA Paper 1996.

Since the Terminator Tether™ of the invention is normally associated with a satellite that has a power supply in the form of a solar panel combined with a storage battery, the Terminator Tether™ can be operated in the "powered propulsion" mode if desired. Such a mode would be useful when attempting to avoid a collision between the Terminator Tether™ and another spacecraft. In such cases, the power from the solar panel, augmented by the power stored in the battery, can be used to temporarily operate the Terminator Tether™ in the propulsion mode instead of the normal drag mode. This would enable to Terminator Tether™ to raise its altitude and/or increase or decrease its inclination to avoid a collision.

In addition, the Terminator Tether™ could be operated in the "power-augmented drag" mode. In this mode, there will be a current induced in the tether by the motion of the spacecraft through the magnetic field, which will cause a drag force on the spacecraft. This induced current would then be augmented by additional current in the same direction generated by the solar panel and the storage batteries in the power supply. The increased current will cause an increased drag force. If there is a significant vertical component of the magnetic field there will also be an increased force normal to the orbital plane, which will cause an increased rate of change of inclination.

When the Terminator Tether™ is operated in the "powered propulsion" or "power-augmented drag" mode, the conditions for stable optimum operation will be essentially the same as in the "unpowered drag" mode. The reason the conditions stay the same is that the maximum tether force than can be effectively utilized is limited by the strength of the gradient force, which has not changed, since the length and mass of the tether and the ballast mass has not changed for these different modes of operation. As a result the optimum operational angle for the tether is still 35.26 degrees, and the optimum ratio of ballast mass to tether mass for different tether lengths and tether material is still determined by the analysis of the invention as is set forth above.

In the analysis given above, where the optimum tether angle was found, the tether was assumed to be in the plane of the orbit and lagging behind the spacecraft motion, since in the "unpowered drag" mode, that would be its normal position, although if the component of the magnetic field along the orbit were high, there would be some tilt of the tether out of the plane of the orbit, producing a force tending to change the inclination of the orbit. In either of the "powered" modes of operation of the Terminator Tether™, however, the tilt of the tether could be forward toward the motion of the spacecraft or strongly tilted to one side or the other of the orbital plane. In all these cases, if the electrodynamic force is allowed to become too large, and the tether angle exceeds 45 degrees, the restoring force of the gradient field will drop off and the tether angle will go unstable. As a result, the optimum angle for the tether that will give the maximum stable force, whether it is a maximum drag force, and maximum propulsion force, or a maximum inclination change force, will be 35.26 degrees, and the optimum ratio for the mass of the tether versus the mass of the ballast will be the same as for the unpowered drag case.

The analysis that gave the optimum angle as 35.26 degrees assumed that the tether would be oriented like a rigid rotor. In reality, variations in forces along the tether will probably cause the tether to hang in a slightly curved shape, where the optimum angle may not be exactly 35.26 degrees from the local vertical. The optimum angle of the tether as it leaves the host spacecraft in such a case is where the drag force is largest.

The present invention is discussed in this disclosure in terms of its space applications as a Terminator Tether™ useful for the removal of unwanted satellites. It should be understood, however, that the present invention is useful in any application where a space object can use a conductive tether to produce electrodynamic force through interaction with a magnetic field. The invention, therefore, should be limited not by this description, but only by the appended claims and their equivalents.

What is claimed is:

1. A method of operating a conductive tether attached to a space object, said conductive tether being oriented at an angle to magnetic field lines of an external magnetic field and said conductive tether moving across the magnetic field lines of said external magnetic field at said angle to producing an electric current in said conducting tether and a resulting electrodynamic force acting on said tether and said attached space object comprising,

Controlling the amount of electric current flowing in said conducting tether to vary the angle between the conductive tether the field lines of the external magnetic field and the electrodynamic force acting on said tether and attached space object.

2. A method as in claim **1** wherein said angle is less than 45 degrees.

3. A method as in claim **1** wherein said angle is 35.26 degrees.

4. A method as in claim **1** wherein said angle is controlled to average 35.26 degrees over time.

5. A method as in claim **1** wherein said angle is controlled to average 35.26 degrees over the length of the tether.

6. A method as in claim **1** wherein the space object and the tether connected to it are rotated about their center of mass while the conductive tether interacts with the external magnetic field.

7. A method as in claim **1** wherein the angle of the conductive tether to the external magnetic field is controlled by sensing a measurable parameter of the space object/conductive tether system.

8. A method as in claim **7** wherein the parameter is current flow in the conductive tether.

9. A method as in claim **7** wherein the parameter is the electrical power dissipated in the conductive tether and the current control system.

10. A method as in claim **7** wherein the parameter is the relative position of the space object and the position of the far end tip of the tether.

11. A method as in claim **7** wherein the parameter is the acceleration of the space object.

12. A method as in claim **6** wherein the parameter is the state vector of the space object as measured by the Global Positioning System.

13. A method as in claim **1** including the step of applying electric power to the conductive tether to change the state vector of the space object.

14. A method of operating a conductive tether attached to a space object, said conductive tether being oriented at an angle to the local vertical between said space object and a celestial body having a magnetic field having magnetic field lines external to said conductive tether; said conductive tether moving across said magnetic field lines of said external magnetic field at an angle to the local vertical to produce an electric current in said conducting tether and a resulting electrodynamic force acting on said tether and said attached space object comprising,

Controlling the amount of electric current flowing in said conducting tether to vary the angle between the conductive tether the local vertical between said space object and said celestial body to vary the electrodynamic force acting on said tether and attached space object.

15. A method as in claim **14** wherein said angle is less than 45 degrees.

16. A method as in claim **14** wherein said angle is controlled to average 35.26 degrees over time.

17. A method as in claim **14** wherein the space object and the tether connected to it are rotated about their center of mass while the conductive tether interacts with the external magnetic field.

18. A method as in claim **14** wherein the angle of the conductive tether to the external magnetic field is controlled by sensing a measurable parameter of the space object/conductive tether system.

19. A method as in claim **18** wherein the parameter is current flow in the conductive tether.

20. A method as in claim **18** wherein the parameter is the acceleration of the space object caused by the electrodynamic force acting on the space object.

21. A method as in claim **18** wherein the parameter is the change in the state vector of the space object as measured by the Global Positioning System.

22. A method as in claim **14** including the step of applying electric power to the conductive tether to change the state vector of the space object.

23. A method as in claim **14** wherein the celestial body is the Earth and the magnetic field is the Earth's magnetic field.

24. A method as in claim **14** wherein the celestial body is the sun and the magnetic field is the Sun's magnetic field.

25. A method as in claim **14** wherein the magnetic field is the vector sum of the magnetic fields from all celestial bodies in the solar system.

26. A method as in claim **14** wherein the conductive tether is a Hoytether.

* * * * *

WILLIAM H. WARD.

Improvement in Collecting Electricity for Telegraphing, &c.

No. 126,356. Patented April 30, 1872.

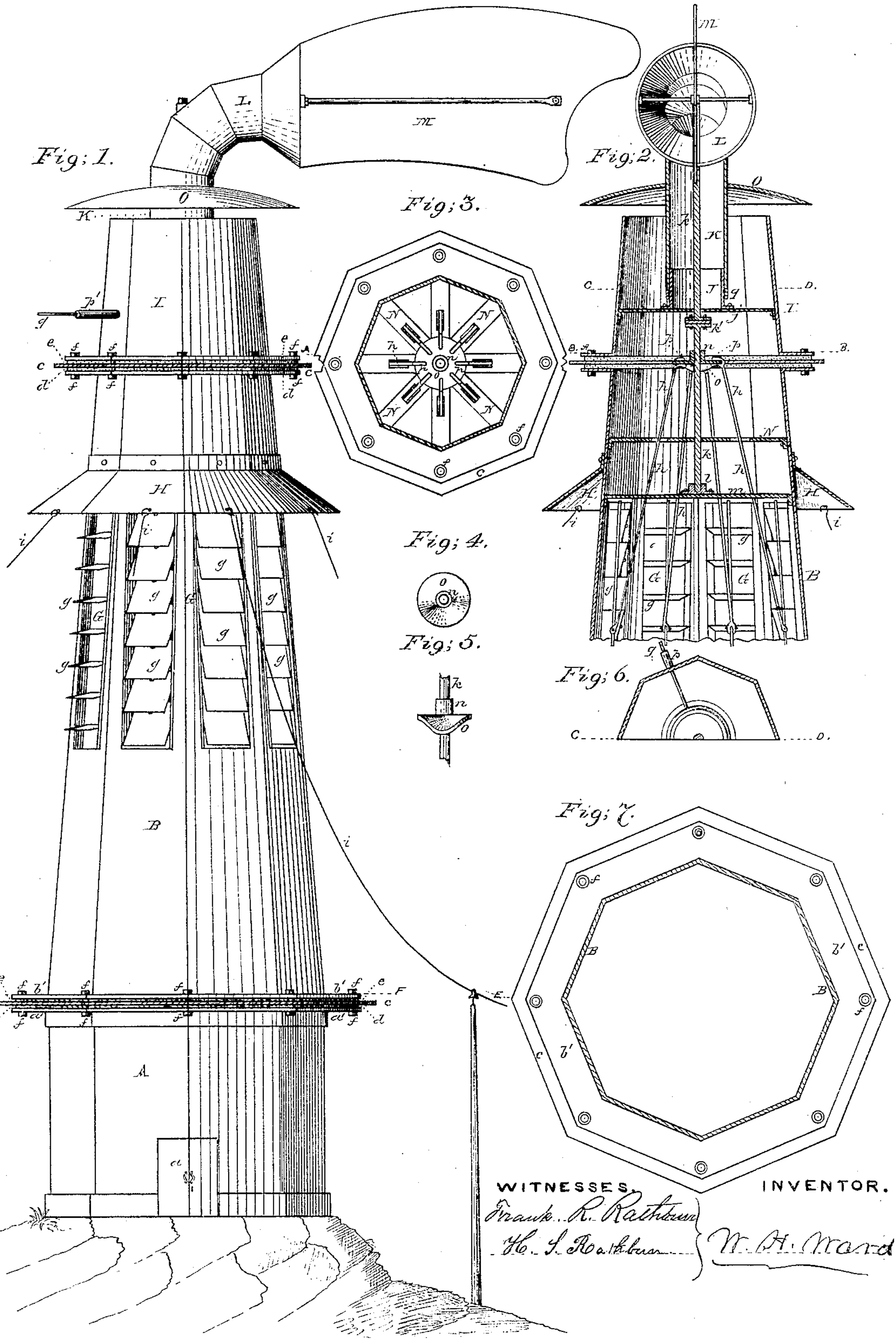
Fig; 1.
Fig; 2.
Fig; 3.
Fig; 4.
Fig; 5.
Fig; 6.
Fig; 7.
A
B
G
H
I
K
L
M
N
O
a
b'
c
d
e
f
g
h
i
k
l
m
n
o
p
p'
C
D
E
F

WITNESSES.

Frank R. Rathbun

H. S. Rathbun

INVENTOR.

W. H. Ward

AM. PHOTO-LITHOGRAPHIC CO. N. Y. (OSBORNE'S PROCESS.)

126,356

UNITED STATES PATENT OFFICE.

WILLIAM HENRY WARD, OF AUBURN, NEW YORK.

IMPROVEMENT IN COLLECTING ELECTRICITY FOR TELEGRAPHING, &c.

Specification forming part of Letters Patent No. 126,356, dated April 30, 1872.

I, WILLIAM HENRY WARD, of Auburn, in the county of Cayuga and State of New York, have invented an Electrical Tower for Accumulating Natural Electricity for Telegraphic Purposes, of which the following is a specification:

My invention consists of a tower for the purpose of receiving and imparting natural electricity, so as to be in constant contact with that upper stratum of electricity which surrounds the earth, by tapping which a never-failing supply is formed when brought into contact with the earth, as will be more fully explained hereinafter.

In the accompanying drawing, Figure 1 represents a side elevation of my improved electrical tower. Fig. 2 is a vertical central section of the upper part of the same. Figs. 3, 4, 5, 6, and 7 are detached views of several parts of the same.

The tower is constructed in three separate sections, a lower, a middle, and an upper one, and is placed on elevated mountain tops or peaks. The lower one, A, is a mere shell, having a door, *a*, and is constructed of any material suitable for the purpose. It is insulated from the middle portion B by means of a glass diaphragm, *c*, held between similar diaphragms of rubber, *d*, and of gutta-percha, *e*, by means of flanges *a'* and *b'*, formed on the upper end of the lower portion and the lower end of the middle portion of the tower respectively. These flanges are secured to each other by insulated bolts *f*. The middle portion B is provided with suitable openings or windows, G, having shutters or slats *g* pivoted in them, so that, by means of raising or lowering rods *h*, suitably connected to said shutters, the openings G may be shut or opened. A projecting roof, H, is formed on or secured to the middle portion B just over the openings G, and which serves the double purpose of protecting the said openings from the effects of the weather, such as rain or snow, and also for receiving the aerial electricity, which may be drawn from it by wires *i*, for land-line purposes. Above this roof H the middle portion of the tower is again insulated by diaphragms of glass, rubber, and gutta-percha from the upper portion I of the tower in the same manner as it is insulated from its lower portion. A circular plate, *j*, secured on the inside of this upper portion, bears a short tube, J, which is surrounded by the tube K of the ventilator L, from which latter the vane M extends; or the tube J may be held by rods extending from the side of the tower centrally. This ventilator is supported by a rod or shaft, *k*, firmly attached to the tube K, and having its lower bearing in a step, *l*, on a brace, *m*, crossing the middle portion B of the tower just above the openings G. This rod or shaft *k* is formed in two parts, insulated from each other, as shown at *k'*. On the lower portion of this shaft *k* is keyed or otherwise secured a sleeve, *n*, from which a horizontal serpentine cam-plate, *o*, shown in detail in Figs. 4 and 5, extends, over the rim of which the forked ends *p* of rods *h* seize, and which is so arranged relatively to the vane M and the shutters or slats *g* that the revolution of the vane by the wind will open the windward and close the leeward shutters or slats *g* and openings G by means of the rods *h*. Guide-plates or their equivalents N, provided with suitable openings, through which the rods *h* pass, keep the upper forked ends *h* of the latter in constant contact with the cam-plate *o*, which raises or lowers the rods, and consequently opens or closes the openings G as it revolves. A circular roof, O, extends from the tube K of the ventilator so as to cover and protect the open upper end of the tower. The upper portion I of the tower is provided with an insulating-tube, *p'*, through which a copper wire, *q*, enters the portion I of the tower, and is coiled around the tube J just under the tube K, which wire may connect that upper portion of the tower with any land line of wires. The upper portion I of the tower, as well as the tubes J K, may be constructed of zinc, while the vane may be made of zinc, copper, or any equivalent material, which, with the zinc and the moisture of the atmosphere, would form an electrical current. As the vane revolves it opens the openings to windward and closes those to leeward, thus helping to drive an aerial current of electricity into the insulated middle portion of the tower, which current passes upwardly through the upper portion of the tower and out through the ventilator or the top, which is swung around by the wind or aerial electrical current, thus forming a draught in addition to the closing of the shutters or openings to leeward, by means of which draught the electrical current is forced upwardly and out at the vane. The top portion of the tower is of course completely insulated from all below it, including the revolving central ventilator-shaft. As the middle portion of the tower allows the electrical current of atmosphere, wind, &c., to pass within, up, and out at the top it forms a continuous current, whereby the tower is receiving continually fresh and new supplies of electricity, which can be drawn from the projecting roof H by the wires *i* for the use of land lines of telegraphs or for other purposes, such as light, heat, &c. By the use of aerial electricity I entirely dispense with artificial batteries, forming my circuit merely by connecting the aerial current with the earth current. For instance, to bring Buenos Ayres, in South America, into direct connection with New York, the following plan would be pursued: One electrical tower is erected on Pike's Peak or any other suitable high mountain in North America, and another similar tower on some suitable peak of the Andes in South America. The former would, by means of land-lines, be connected directly with Denver, which place is again connected with all the prominent cities of the States. In a similar manner the southern tower is connected by land-lines with the prominent cities via Quito. New York telegraphs to the tower on Pike's Peak, and the operator having connected the land-line with the aerial current, the signals are transmitted through the aerial current to the tower in the Andes in South America; and from there—the land-lines being suitably connected with the aerial current—to Quito and Buenos Ayres. In this manner a message would be sent entirely by natural electricity in place of artificial. In the same manner a message may be sent across the ocean by having a high tower on each continent, each of which towers would have to be, of course, through land-lines connected with the earth to enable the ground current with the aerial current to form a circuit. Different towers may be erected on the different continents, and if they are all what is technically called hooked on—that is to say, connected to the earth—a signal given at one tower will be repeated at all the towers, they being connected with each other by the aerial current. If the earth-connection is severed, or the insulation with the tower destroyed, there is no power; but by insulating the tower and concentrating its force to a point, bringing the same corresponding effect from the earth current in connection, an exceedingly powerful electrical force is created.

Having described my invention, I claim—

1. A tower constructed so as to collect, hold, distribute, and utilize aerial currents of natural electricity for telegraphic and other purposes, essentially as described.

2. A tower for collecting aerial currents of natural electricity, constructed of three sections, insulated from each other and the earth, or their equivalents, substantially in the manner described.

3. In an electrical tower, the combination of the shutters or slats *g* with the ventilator-vane in such a manner that the vane, through suitable mechanism, always opens the shutters to windward and closes them to leeward, substantially as described.

4. In an electrical tower, the collecting, distributing, and protecting roof H, substantially as and for the purpose described.

5. The combination of the tube J and ventilator L with its vane M and the coiled wire or small cable *q*, substantially as and for the purpose set forth.

6. The combination of the insulated shaft *k* and rods *h* having forked upper ends *k'* with the shutters or blinds *g*, serpentine cam-plate *o*, and ventilator L, substantially as and for the purpose described.

The above specification of my improvement, being a tower for accumulating natural electricity for telegraphic purposes, signed this 29th day of June, A. D. 1871.

W. H. WARD.

Witnesses:

A. L. PALMER,
J. H. STARIN.

No. 649,621. Patented May 15, 1900.

N. TESLA.

APPARATUS FOR TRANSMISSION OF ELECTRICAL ENERGY.

(Application filed Feb. 19, 1900.)

(No Model.)

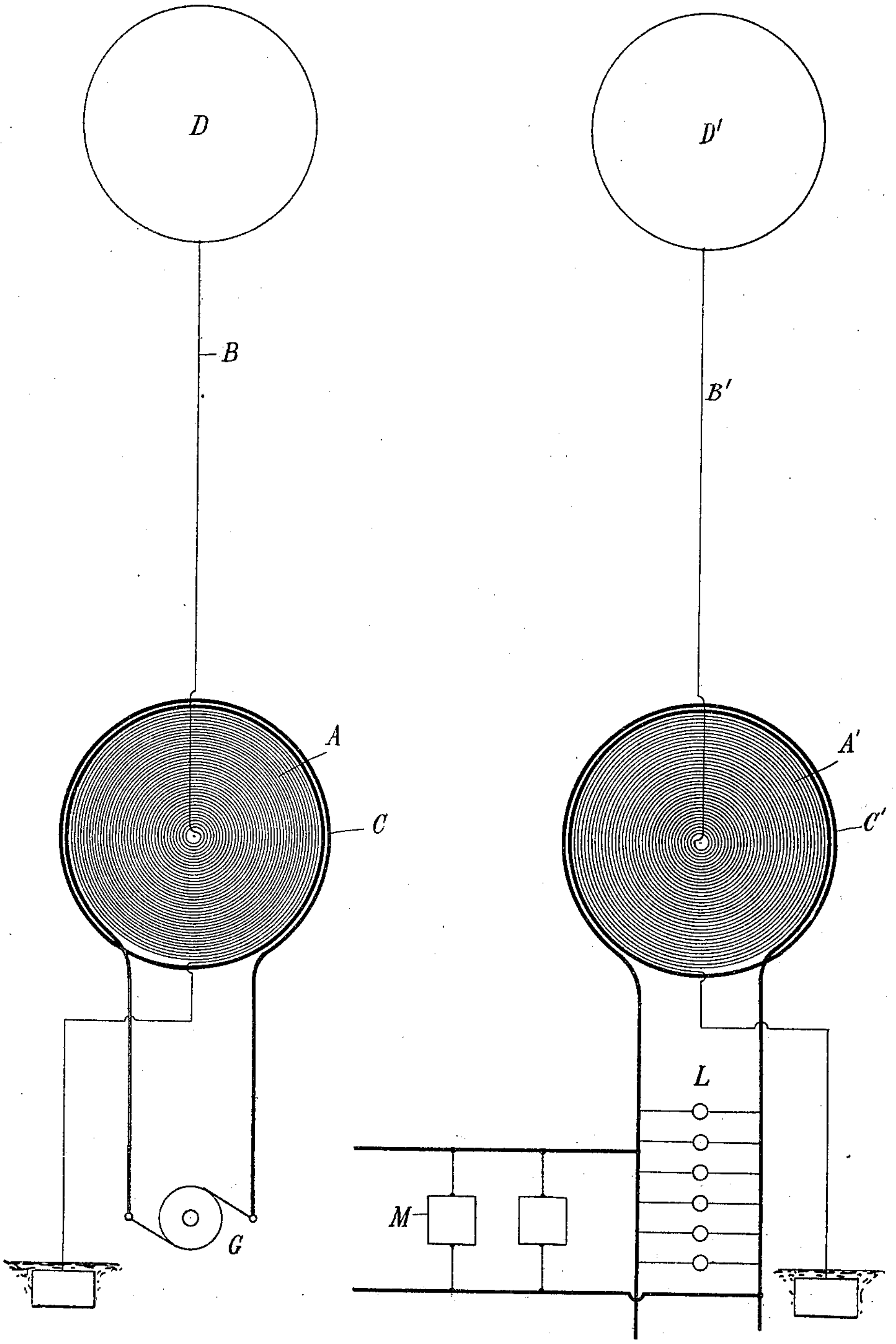

Witnesses:
Benjamin Miller.
G. W. Martling.

Nikola Tesla, Inventor
by Kerr, Page & Cooper
Att'ys

THE NORRIS PETERS CO., PHOTO-LITHO., WASHINGTON, D. C.

UNITED STATES PATENT OFFICE.

NIKOLA TESLA, OF NEW YORK, N. Y.

APPARATUS FOR TRANSMISSION OF ELECTRICAL ENERGY.

SPECIFICATION forming part of Letters Patent No. 649,621, dated May 15, 1900.

Original application filed September 2, 1897, Serial No. 650,343. Divided and this application filed February 19, 1900. Serial No. 5,780. (No model.)

To all whom it may concern:

Be it known that I, NIKOLA TESLA, a citizen of the United States, residing at the borough of Manhattan, in the city of New York, county and State of New York, have invented certain new and useful Improvements in Apparatus for the Transmission of Electrical Energy, of which the following is a specification, reference being had to the drawing accompanying and forming a part of the same.

This application is a division of an application filed by me on September 2, 1897, Serial No. 650,343, entitled "Systems of transmissions of electrical energy," and is based upon new and useful features and combinations of apparatus shown and described in said application for carrying out the method therein disclosed and claimed.

The invention which forms the subject of my present application comprises a transmitting coil or conductor in which electrical currents or oscillations are produced and which is arranged to cause such currents or oscillations to be propagated by conduction through the natural medium from one point to another remote therefrom and a receiving coil or conductor at such distant point adapted to be excited by the oscillations or currents propagated from the transmitter.

This apparatus is shown in the accompanying drawing, which is a diagrammatic illustration of the same.

A is a coil, generally of many turns and of a very large diameter, wound in spiral form either about a magnetic core or not, as may be desired. C is a second coil formed by a conductor of much larger size and smaller length wound around and in proximity to the coil A.

The apparatus at one point is used as a transmitter, the coil A in this case constituting a high-tension, secondary, and the coil C the primary, of much lower tension, of a transformer. In the circuit of the primary C is included a suitable source of current G. One terminal of the secondary A is at the center of the spiral coil, and from this terminal the current is led by a conductor B to a terminal D, preferably of large surface, formed or maintained by such means as a balloon at an elevation suitable for the purposes of transmission. The other terminal of the secondary A is connected to earth, and, if desired, to the primary also, in order that the latter may be at substantially the same potential as the adjacent portions of the secondary, thus insuring safety. At the receiving-station a transformer of similar construction is employed; but in this case the longer coil A′ constitutes the primary, and the shorter coil C′ the secondary, of the transformer. In the circuit of the latter are connected lamps L, motors M, or other devices for utilizing the current. The elevated terminal D′ connects with the center of the coil A′, and the other terminal of said coil is connected to earth and preferably, also, to the coil C′ for the reasons above stated.

The length of the thin wire coil in each transformer should be approximately one-quarter of the wave length of the electric disturbance in the circuit, this estimate being based on the velocity of propagation of the disturbance through the coil itself and the circuit with which it is designed to be used. By way of illustration, if the rate at which the current traverses the circuit including the coil be one hundred and eighty-five thousand miles per second then a frequency of nine hundred and twenty-five per second would maintain nine hundred and twenty-five stationary moves in a circuit one hundred and eighty-five thousand miles long and each wave would be two hundred miles in length.

For such a low frequency, which would be resorted to only when it is indispensable for the operation of motors of the ordinary kind under the conditions above assumed, I would use a secondary of fifty miles in length. By such an adjustment or proportioning of the length of wire in the secondary coil or coils the points of highest potential are made to coincide with the elevated terminals D D′, and it should be understood that whatever length be given to the wires this requirement should be complied with in order to obtain the best results.

It will be readily understood that when the above-prescribed relations exist the best conditions for resonance between the transmitting and receiving circuits are attained, and owing to the fact that the points of highest potential in the coils or conductors A A′ are coincident with the elevated terminals the maximum flow of current will take place in the two coils, and this, further, necessarily implies that the capacity and inductance in each of the circuits have such values as to secure the most perfect condition of synchronism with the impressed oscillations.

When the source of current G is in operation and produces rapidly pulsating or oscillating currents in the circuit of coil C, corresponding induced currents of very much higher potential are generated in the secondary coil A, and since the potential in the same gradually increases with the number of turns toward the center and the difference of potential between the adjacent turns is comparatively small a very high potential impracticable with ordinary coils may be successively obtained.

As the main object for which the apparatus is designed is to produce a current of excessively-high potential, this object is facilitated by using a primary current of very considerable frequency; but the frequency of the currents is in a large measure arbitrary, for if the potential be sufficiently high and the terminals of the coils be maintained at the proper elevation where the atmosphere is rarefied the stratum of air will serve as a conducting medium for the current produced and the latter will be transmitted through the air, with, it may be, even less resistance than through an ordinary conductor.

As to the elevation of the terminals D D′, it is obvious that this is a matter which will be determined by a number of things, as by the amount and quality of the work to be performed, by the condition of the atmosphere, and also by the character of the surrounding country. Thus if there be high mountains in the vicinity the terminals should be at a greater height, and generally they should always be at an altitude much greater than that of the highest objects near them. Since by the means described practically any potential that is desired may be produced, the currents through the air strata may be very small, thus reducing the loss in the air.

The apparatus at the receiving-station responds to the currents propagated from the transmitter in a manner which will be well understood from the foregoing description. The primary circuit of the receiver—that is, the thin wire coil A′—is excited by the currents propagated by conduction through the intervening natural medium from the transmitter, and these currents induce in the secondary coil C′ other currents which are utilized for operating the devices included in the circuit thereof.

Obviously the receiving-coils, transformers, or other apparatus may be movable—as, for instance, when they are carried by a vessel floating in the air or by a ship at sea. In the former case the connection of one terminal of the receiving apparatus to the ground might not be permanent, but might be intermittently or inductively established without departing from the spirit of my invention.

It is to be noted that the phenomenon here involved in the transmission of electrical energy is one of true conduction and is not to be confounded with the phenomena of electrical radiation which have heretofore been observed and which from the very nature and mode of propagation would render practically impossible the transmission of any appreciable amount of energy to such distances as are of practical importance.

What I now claim as my invention is—

1. The combination with a transmitting coil or conductor connected to ground and to an elevated terminal respectively, and means for producing therein electrical currents or oscillations, of a receiving coil or conductor similarly connected to ground and to an elevated terminal, at a distance from the transmitting-coil and adapted to be excited by currents caused to be propagated from the same by conduction through the intervening natural medium, a secondary conductor in inductive relation to the receiving-conductor and devices for utilizing the current in the circuit of said secondary conductor, as set forth.

2. The combination with a transmitting coil or conductor having its ends connected to ground and to an elevated terminal respectively, a primary coil in inductive relation thereto and a source of electrical oscillations in said primary circuit, of a receiving conductor or coil having its ends connected to ground and to an elevated terminal respectively and adapted to be excited by currents caused to be propagated from the transmitter through the natural medium and a secondary circuit in inductive relation to the receiving-circuit and receiving devices connected therewith, as set forth.

3. The combination with a transmitting instrument comprising a transformer having its secondary connected to ground and to an elevated terminal respectively, and means for impressing electrical oscillations upon its primary, of a receiving instrument comprising a transformer having its primary similarly connected to ground and to an elevated terminal, and a translating device connected with its secondary, the capacity and inductance of the two transformers having such values as to secure synchronism with the impressed oscillations, as set forth.

4. The combination with a transmitting instrument comprising an electrical transformer having its secondary connected to ground and to an elevated terminal respectively, and means for impressing electrical oscillations upon its primary, of a receiving instrument comprising a transformer having its primary similarly connected to ground and to an elevated terminal, and a translat-

ing device connected with its secondary, the capacity and inductance of the secondary of the transmitting and primary of the receiving instruments having such values as to secure synchronism with the impressed oscillations, as set forth.

5. The combination with a transmitting coil or conductor connected to ground and an elevated terminal respectively, and means for producing electrical currents or oscillations in the same, of a receiving coil or conductor similarly connected to ground and to an elevated terminal and synchronized with the transmitting coil or conductor, as set forth.

6. The combination with a transmitting instrument comprising an electrical transformer, having its secondary connected to ground and to an elevated terminal respectively, of a receiving instrument comprising a transformer, having its primary similarly connected to ground and to an elevated terminal, the receiving-coil being synchronized with that of the transmitter, as set forth.

7. The combination with a transmitting coil or conductor connected to ground and to an elevated terminal respectively, and means for producing electrical currents or oscillations in the same, of a receiving coil or conductor similarly connected to ground and to an elevated terminal, the said coil or coils having a length equal to one-quarter of the wave length of the disturbance propagated, as set forth.

8. The combination with a transmitting coil or conductor connected to ground and to an elevated terminal respectively, and adapted to cause the propagation of currents or oscillations by conduction through the natural medium, of a receiving-circuit similarly connected to ground and to an elevated terminal, and of a capacity and inductance such that its period of vibration is the same as that of the transmitter, as set forth.

9. The transmitting or receiving circuit herein described, connected to ground and an elevated terminal respectively, and arranged in such manner that the elevated terminal is charged to the maximum potential developed in the circuit, as set forth.

10. The combination with a transmitting coil or conductor connected to ground and to an elevated terminal respectively of a receiving-circuit having a period of vibration corresponding to that of the transmitting-circuit and similarly connected to ground and to an elevated terminal and so arranged that the elevated terminal is charged to the highest potential developed in the circuit, as set forth.

NIKOLA TESLA.

Witnesses:
PARKER W. PAGE,
MARCELLUS BAILEY.

No. 674,427. Patented May 21, 1901.

A. PALENCSÁR.

APPARATUS FOR COLLECTING ATMOSPHERIC ELECTRICITY.

(Application filed July 10, 1900.)

(No Model.) 3 Sheets—Sheet 1.

Fig: 1.

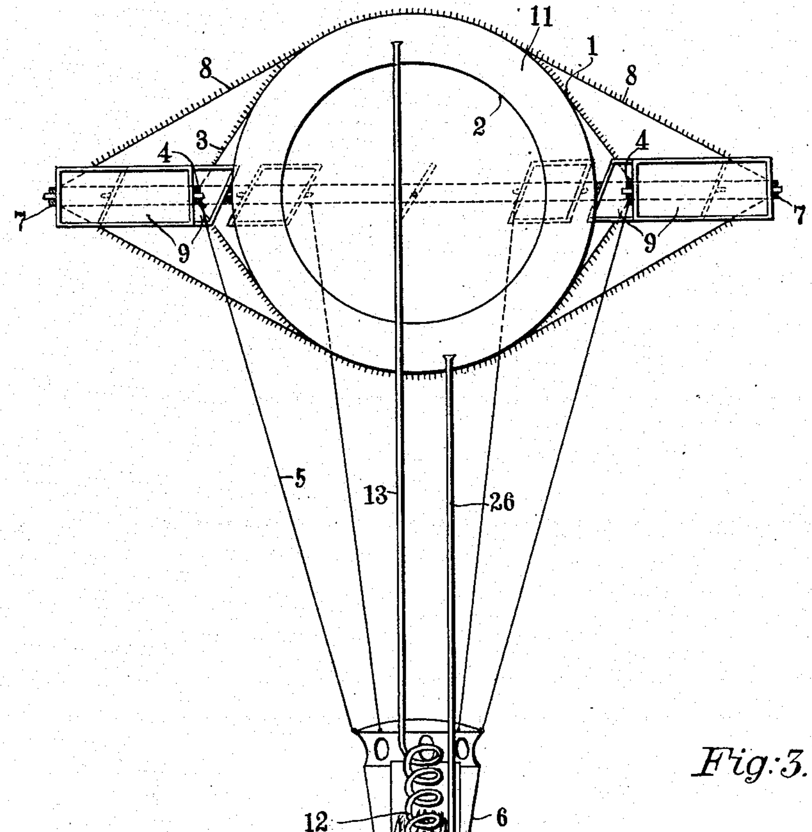
11
1
8
8
2
3
4
4
7
7
9
9
9
5
13
26
12
6
14
15

Fig: 3.

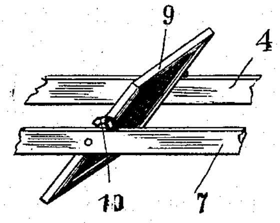

WITNESSES:
Ella L. Giles
Otto Munk

INVENTOR
Andor Palencsár
BY

ATTORNEYS

THE NORRIS PETERS CO., PHOTO-LITHO., WASHINGTON, D. C.

No. 674,427. Patented May 21, 1901.

A. PALENCSÁR.

APPARATUS FOR COLLECTING ATMOSPHERIC ELECTRICITY.

(Application filed July 10, 1900.)

(No Model.) 3 Sheets—Sheet 2.

Fig: 2.

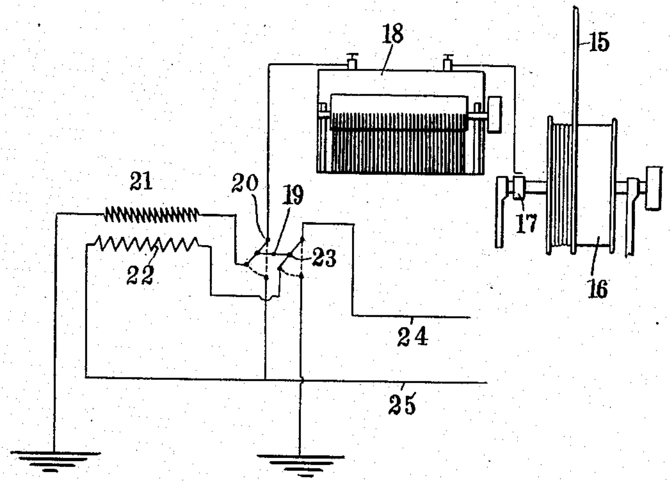

WITNESSES:
Ella L. Giles
Otto Munk

INVENTOR
Andor Palencsár
BY
Richards & Co
ATTORNEYS

THE NORRIS PETERS CO., PHOTO-LITHO., WASHINGTON, D. C.

No. 674,427. Patented May 21, 1901.

A. PALENCSÁR.

APPARATUS FOR COLLECTING ATMOSPHERIC ELECTRICITY.

(Application filed July 10, 1900.)

(No Model.) 3 Sheets—Sheet 3.

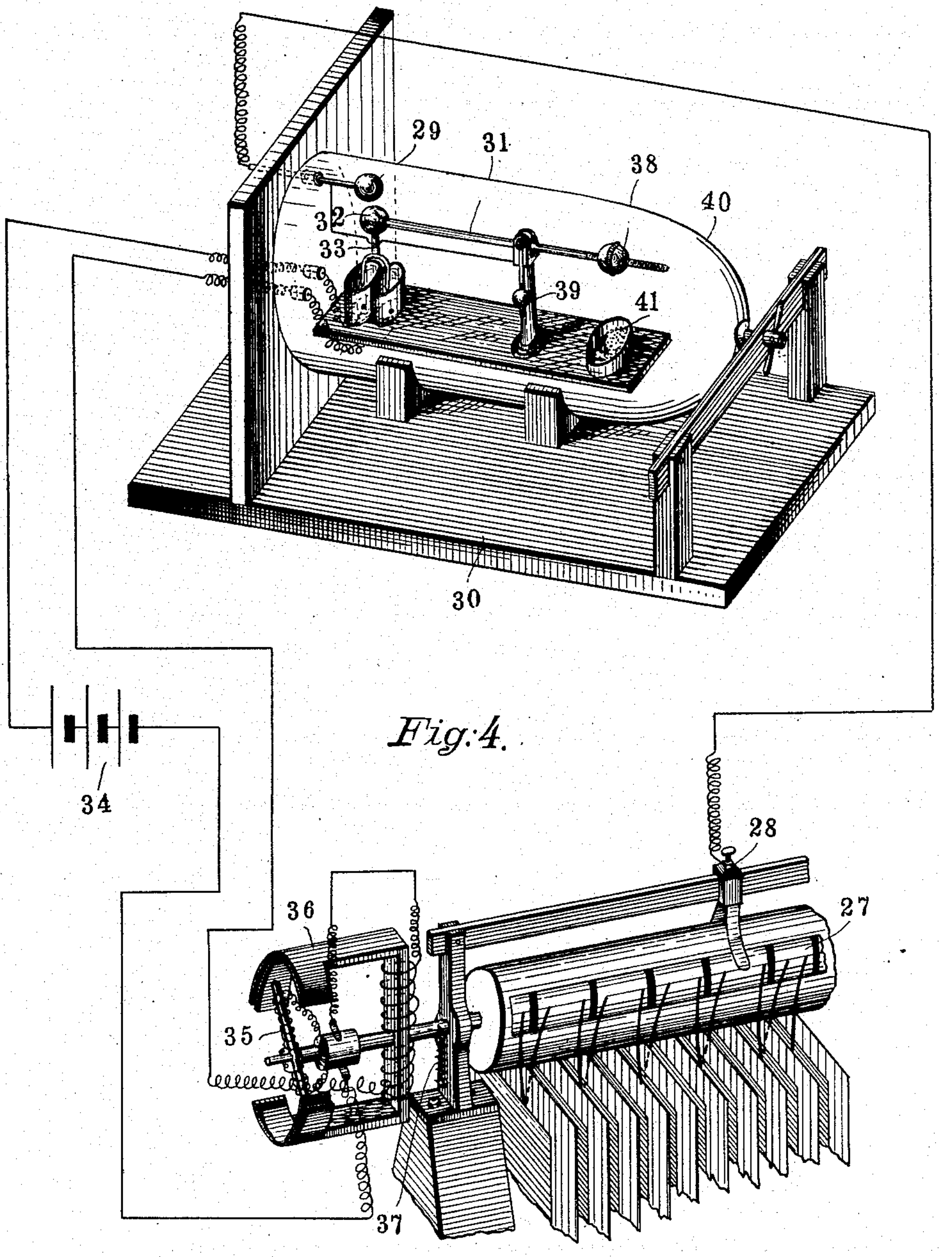

29
31
38
40
32
33
39
41
30
34
28
27
36
35
37

Fig. 4.

WITNESSES:
Ella L. Giles
Otto Munk

INVENTOR
Andor Palencsár
BY
Richards & Co
ATTORNEYS

THE NORRIS PETERS CO., PHOTO-LITHO., WASHINGTON, D. C.

UNITED STATES PATENT OFFICE.

ANDOR PALENCSÁR, OF BUDA-PESTH, AUSTRIA-HUNGARY.

APPARATUS FOR COLLECTING ATMOSPHERIC ELECTRICITY.

SPECIFICATION forming part of Letters Patent No. 674,427, dated May 21, 1901.

Application filed July 10, 1900. Serial No. 23,102. (No model.)

To all whom it may concern:

Be it known that I, ANDOR PALENCSÁR, a subject of the King of Hungary, residing at Buda-Pesth, Austria-Hungary, have invented a certain new and useful Apparatus for Collecting Atmospherical Electricity, of which the following is a full, clear, and exact specification.

The present invention concerns an apparatus for collecting and driving or conducting atmospherical electricity which renders it practicable to obtain material quantities of the same in serviceable form.

The experiments made hitherto have been limited to deviating or conducting the electricity by simple lightning-rods without making allowance for the progress of the theory of atmospherical electricity. It is obvious that only extremely small quantities of electricity can be collected in this manner, and, moreover, it was obtained in a form which entirely precluded its utilization. It is evident that a source of electricity of irregular yield can only be utilized by means of accumulators and for charging accumulators only a current of constant potential can be employed. The potential of electricity derived or deviated by means of a lightning-rod varies within wide limits, and, moreover, it is so high that it cannot be used at all for charging accumulators. All these drawbacks are remedied by the present invention, by which the atmospheric electricity is obtained in larger quantities with as low and constant a potential as may be desired.

The idea underlying the invention is based on the modern theory of atmospheric electricity, according to which it is produced by the condensation of steam or aqueous vapors, and that the increase of potential is effected by the concentration of the small drops of water into larger ones, as the proportion of the surface of the drops to the volume of same is materially reduced thereby. According to this theory the water-droplets floating in the layers of air are considered as vehicles or carriers of the electricity, and a rational system of the deviation of the atmospheric electricity must derive it from the water-drops. This is attained by the present invention in the following manner: A collecting-body of as large a surface as possible is provided with sharp needles. This is moved in the higher layers of air, while being continually heated by a suitable heating device. Owing to the heat the water-drops immediately ambient to the collecting-body will be evaporated, their capacity is gradually reduced, while the potential of the charge grows until it reaches an infinite height with infinite smallness of the drops. It is readily apparent that the whole charge of the drops as soon as they have been evaporated will have passed to the collector or collecting-body, from which it can be conducted. For the purpose of replacing the evaporated drops, which have been deprived of their charge by new-charged drops, the collector is moved in relation to the ambient air.

One way of carrying out the invention is represented in diagram in the accompanying drawings, in which—

Figure 1 is a longitudinal section of the collecting-balloon. Fig. 2 shows the deviating and conversion device, and Fig. 3 a detail. Fig. 4 is a diagrammatic perspective view of a form of automatic regulator for the rheostatic machine.

The apparatus consists of a balloon having two walls and covered with a light wire net, preferably of aluminium wire, said net being studded with needles. Besides, the balloon carries the net 3, on which the ring 4, made of a solid but light material, (wood, cane, &c.,) is fixed. This ring carries the basket 6 by means of the cords or ropes 5. On a level with the ring 4 is the ring 7, which is kept spread by the blades or wings 9, which are journaled in a manner to rotate easily. The ring 7 is kept fixedly in position by the cords or ropes 8. The blades or wings consist of a frame covered with light material, and their rotation in either direction is limited by the stops or ledges 10. The blades or wings form advantageously an angle of sixty to seventy degrees with the vertical line.

All wire nets, ropes, rings, and blades or wings may be covered with small metal needles which are electrically connected with each other.

From the space 11, formed between the double walls of the balloon, the pipe 26 leads from the lowest point of the same to the serpentine 12, arranged in the basket 6, which in turn is

connected with the pipe 13, ending in the upper part of the space 11 between the two walls 12. The serpentine is heated by means of a suitable source of heat, whereby a warm current of gas or air circulates continually between the double walls of the balloon.

Under the basket the ball-bearing 14 is arranged in electrical connection with the wire nets, and its stud is electrically connected with the carefully-insulated light though sufficiently strong cable 15.

On the earth's surface is a winch 16, Fig. 2, by means of which the balloon can be made to ascend or descend as soon as the interior space of the balloon is filled with illuminating gas or hydrogen.

The end of the cable-core is soldered to a collector arranged upon and insulated from the axle of the winch, and the electricity is conducted from this collector by means of a sliding contact. The collecting of the electricity takes place by moving the balloon continuously up and down by means of the winch. In this movement the balloon is turned by means of the wings or blades 9, which are adapted to turn the balloon always in the same direction whether ascending or descending, as in the change from ascending into descending of the balloon, or vice versa, the blades are turned over by the aerial resistance, and thus impart the rotating motion to the balloon in the same direction. In order to avoid torsion of the cable, the ball-bearing 14 is provided. This up-and-down motion and rotating of the balloon accomplishes the purpose of bringing it into contact with as many water particles floating in the air as possible.

As the electricity conducted from the collector-wing 17 possesses a much too high and varying potential for making its direct application practicable, and as it is usual with an irregular source of electricity to first charge accumulators and to further utilize the easily-regulated current of the same only, it becomes necessary to seek to maintain the electricity conducted from the collector 17 for the charging of the accumulators at a constant potential and convert the potential to a much lower one; but as we deal in this case with a direct current ordinary converters cannot be used for this purpose. Moreover, the electricity possesses in this case a much too high potential, so that with the employment of ordinary converters the greatest part of the collected electricity would be lost again. The only practical method for this purpose is the converting by means of the Planté rheostatic machine, by which this high-voltage electricity can be transformed almost without any loss whatever. Thus the question of conversion would be solved, and only the question of maintenance of constant potential remains. This is obtained in the following manner: The rheostatic machine 18 or only a part of the plates is connected with an electrometer of any construction, the movable part of which closes a contact which eventually actuates an electromagnet which effects the switching or reversion of the rheostatic machine. After the switching of the rheostatic machine it is discharged, the potential falls to zero, (*nil*,) and the electrometer resumes its initial position, whereby the current of the electromagnet which effects the reversion is interrupted and the plates of the rheostatic machine are reconnected to potential. The machine is then ready for renewed charge and is again discharged when the determined potential is reached. This action is continually repeated as long as the apparatus is in operation.

A form of the automatic regulator for the rheostatic machine is represented in Fig. 4. 27 is the contact-cylinder of the rheostatic machine, on which, for clearness sake, only the contacts for the charging position of the condenser-plates are shown, while the contacts for the discharging position, which come into action after the contact-cylinder has been turned, are omitted. The coating of the condenser-plates is connected in electrical circuit with the stationary ball 29 and the movable ball 32 of the charge-meter 30. When the charge of the rheostatic machine rises, the ball 32, arranged on the one extremity of the lever 31, is repelled and at a certain stroke actuated by means of the fork 33, which is fastened on the ball 32, and by dipping into a mercury-cup it closes the circuit of the source of current 34. This current passes through the windings of the anchor or armature 35, fixed on the shaft of the contact-cylinder, and it also passes through the electromagnet 36. Thereby the contact-cylinder is turned by a certain angle and the rheostatic machine is thus reversed. Then if the tension decreases by discharge and the repulsion of the balls 29 32 declines so far that the contact is interrupted at 33 the spring 37 turns the contact-cylinder into its normal position and the rheostatic machine is again switched to tension. The regulation of the electrometer is effected by the adjustable weights 38 39. When larger quantities of electricity shall be derived or deviated, two rheostatic machines may operate alternately so, that while the one is being discharged the other can be charged. It is readily apparent that if the capacity of the rheostatic machine is not changed the switching or reversion by the electrometer will always take place at the same potential of the rheostatic machine, and as the number of plates, and hence the proportion of conversion, remains the same the current impulses derived or deviated from the rheostatic machine will also have the same potential.

The irregularities of the source of electricity change the interval of time in which the charges follow each other; but as long as the potential remains constant this has no injurious effect on the charge of the accumulators. The current derived. or deviated from the

rheostatic machine can be further transformed by an ordinary converter 21 22, and whenever it is sufficiently constant it can be utilized without the intervention of the accumulators. The converters can be connected or disconnected by means of the double switch 19 20 23.

24 and 25 are the conducting-wires, which run either directly to the place of consumption or to an accumulator-battery.

Having now described my invention, what I claim as new, and desire to secure by Letters Patent, is—

1. An apparatus for collecting atmospheric electricity comprising a collecting-body adapted to be kept in motion, heating means for said body, and a rheostatic machine and a converter connected with the said body electrically, substantially as described.

2. An apparatus for collecting atmospheric electricity for storage comprising a collecting-body adapted to be kept in motion in the ambient air, heating means for said body, a conductor leading from said body, a rheostatic machine connected with the said conductor, an electrometer connected with said rheostatic machine electrically, a contact controlled by the electrometer, and an electromagnet controlled by said contact, said electromagnet controlling the reversing of the rheostatic machine, substantially as described.

3. In combination, the balloon-like collecting-body, having the collecting-points, means for heating the interior space of said body, a conductor leading from the balloon-like body, and electrical devices for receiving the current therefrom, substantially as described.

4. In combination, the balloon-like collecting-body, means for turning the same constantly in one direction in both ascending and descending, conducting means leading from the balloon and electrical devices for receiving the current therefrom, substantially as described.

In witness whereof I have hereunto set my hand in presence of two witnesses.

ANDOR PALENCSÁR.

Witnesses:
EUGÈNE HARRÄNJO,
PAUL BÖLESKEY.

No. 685,957. Patented Nov. 5, 1901.

N. TESLA.

APPARATUS FOR THE UTILIZATION OF RADIANT ENERGY.

(Application filed Mar. 21, 1901.)

(No Model.)

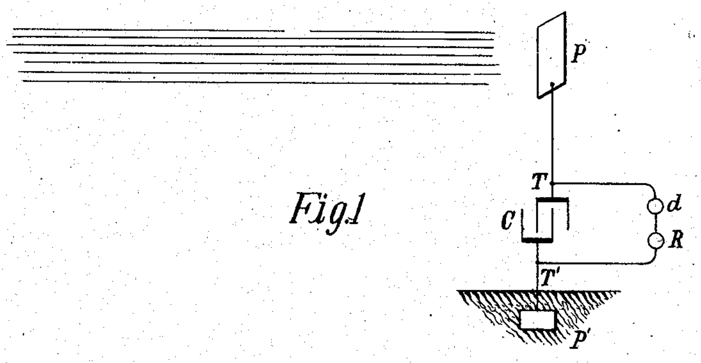

Fig.1

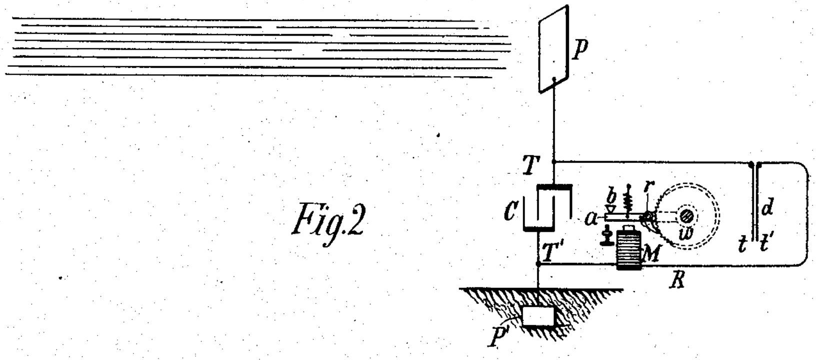
P
T
C
b
r
a
w
d
t
t'
T'
M
R
P'

Fig.2

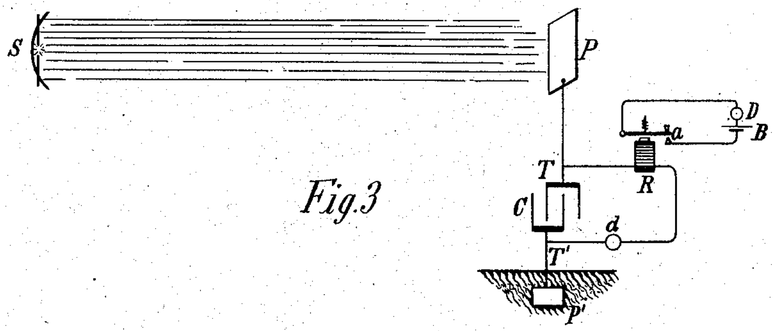
S
P
D
B
a
R
T
C
d
T'
P'

Fig.3

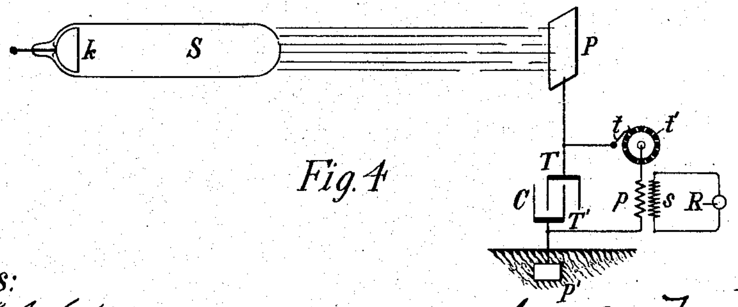
k
S
P
t
t'
T
C
p
s
R
T'
P'

Fig.4

Witnesses:
Raphaël Netter
M. Lawson Dyer.

Inventor
Nikola Tesla
by Kerr. Page & Cooper Attys.

THE NORRIS PETERS CO., PHOTO-LITHO., WASHINGTON, D. C.

UNITED STATES PATENT OFFICE.

NIKOLA TESLA, OF NEW YORK, N. Y.

APPARATUS FOR THE UTILIZATION OF RADIANT ENERGY.

SPECIFICATION forming part of Letters Patent No. 685,957, dated November 5, 1901.

Application filed March 21, 1901. Serial No. 52,153. (No model.)

To all whom it may concern:

Be it known that I, NIKOLA TESLA, a citizen of the United States, residing at the borough of Manhattan, in the city, county, and State of New York, have invented certain new and useful Improvements in Apparatus for the Utilization of Radiant Energy, of which the following is a specification, reference being had to the drawings accompanying and forming a part of the same.

It is well known that certain radiations—such as those of ultra-violet light, cathodic, Roentgen rays, or the like—possess the property of charging and discharging conductors of electricity, the discharge being particularly noticeable when the conductor upon which the rays impinge is negatively electrified. These radiations are generally considered to be ether vibrations of extremely small wave lengths, and in explanation of the phenomena noted it has been assumed by some authorities that they ionize or render conducting the atmosphere through which they are propagated. My own experiments and observations, however, lead me to conclusions more in accord with the theory heretofore advanced by me that sources of such radiant energy throw off with great velocity minute particles of matter which are strongly electrified, and therefore capable of charging an electrical conductor, or, even if not so, may at any rate discharge an electrified conductor either by carrying off bodily its charge or otherwise.

My present application is based upon a discovery which I have made that when rays or radiations of the above kind are permitted to fall upon an insulated conducting-body connected to one of the terminals of a condenser while the other terminal of the same is made by independent means to receive or to carry away electricity a current flows into the condenser so long as the insulated body is exposed to the rays, and under the conditions hereinafter specified an indefinite accumulation of electrical energy in the condenser takes place. This energy after a suitable time interval, during which the rays are allowed to act, may manifest itself in a powerful discharge, which may be utilized for the operation or control of mechanical or electrical devices or rendered useful in many other ways.

In applying my discovery I provide a condenser, preferably of considerable electrostatic capacity, and connect one of its terminals to an insulated metal plate or other conducting-body exposed to the rays or streams of radiant matter. It is very important, particularly in view of the fact that electrical energy is generally supplied at a very slow rate to the condenser, to construct the same with the greatest care. I use, by preference, the best quality of mica as dielectric, taking every possible precaution in insulating the armatures, so that the instrument may withstand great electrical pressures without leaking and may leave no perceptible electrification when discharging instantaneously. In practice I have found that the best results are obtained with condensers treated in the manner described in a patent granted to me February 23, 1897, No. 577,671. Obviously the above precautions should be the more rigorously observed the slower the rate of charging and the smaller the time interval during which the energy is allowed to accumulate in the condenser. The insulated plate or conducting-body should present as large a surface as practicable to the rays or streams of matter, I having ascertained that the amount of energy conveyed to it per unit of time is under otherwise identical conditions proportionate to the area exposed, or nearly so. Furthermore, the surface should be clean and preferably highly polished or amalgamated. The second terminal or armature of the condenser may be connected to one of the poles of a battery or other source of electricity or to any conducting body or object whatever of such properties or so conditioned that by its means electricity of the required sign will be supplied to the terminal. A simple way of supplying positive or negative electricity to the terminal is to connect the same either to an insulated conductor supported at some height in the atmosphere or to a grounded conductor, the former, as is well known, furnishing positive and the latter negative electricity. As the rays or supposed streams of matter generally convey a positive charge to the first condenser-terminal, which is connected to the plate or conductor above mentioned, I usually connect the second terminal of the condenser to the ground, this being the most convenient way of obtaining negative electricity, dispensing with the necessity of providing an artificial source. In order to utilize for any useful purpose the energy accumulated in the condenser, I furthermore connect to the terminals of the same a circuit including an instrument or apparatus which it is desired to operate and another instrument or device for alternately closing and opening the circuit. This latter may be any form of circuit-controller, with fixed or movable parts or electrodes, which may be actuated either by the stored energy or by independent means.

My discovery will be more fully understood from the following description and annexed drawings, to which reference is now made, and in which—

Figure 1 is a diagram showing the general arrangement of apparatus as usually employed. Fig. 2 is a similar diagram illustrating more in detail typical forms of the devices or elements used in practice, and Figs. 3 and 4 are diagrammatical representations of modified arrangements suitable for special purposes.

As illustrative of the manner in which the several parts or elements of the apparatus in one of its simplest forms are to be arranged and connected for useful operation, reference is made to Fig. 1, in which C is the condenser, P the insulated plate or conducting-body which is exposed to the rays, and P′ another plate or conductor which is grounded, all being joined in series, as shown. The terminals T T′ of the condenser are also connected to a circuit which includes a device R to be operated and a circuit-controlling device *d* of the character above referred to.

The apparatus being arranged as shown, it will be found that when the radiations of the sun or of any other source capable of producing the effects before described fall upon the plate P an accumulation of electrical energy in the condenser C will result. This phenomenon, I believe, is best explained as follows: The sun, as well as other sources of radiant energy, throws off minute particles of matter positively electrified, which, impinging upon the plate P, communicate continuously an electrical charge to the same. The opposite terminal of the condenser being connected to the ground, which may be considered as a vast reservoir of negative electricity, a feeble current flows continuously into the condenser, and inasmuch as these supposed particles are of an inconceivably small radius or curvature, and consequently charged to a relatively very high potential, this charging of the condenser may continue, as I have actually observed, almost indefinitely, even to the point of rupturing the dielectric. If the device *d* be of such character that it will operate to close the circuit in which it is included when the potential in the condenser has reached a certain magnitude, the accumulated charge will pass through the circuit, which also includes the receiver R, and operate the latter.

In illustration of a particular form of apparatus which may be used in carrying out my discovery I now refer to Fig. 2. In this figure, which in the general arrangement of the elements is identical to Fig. 1, the device *d* is shown as composed of two very thin conducting-plates *t t′*, placed in close proximity and very mobile, either by reason of extreme flexibility or owing to the character of their support. To improve their action, they should be inclosed in a receptacle, from which the air may be exhausted. The plates *t t′* are connected in series with a working circuit, including a suitable receiver, which in this case is shown as consisting of an electromagnet M, a movable armature *a*, a retractile spring *b*, and a ratchet-wheel *w*, provided with a spring-pawl *r*, which is pivoted to armature *a*, as illustrated. When the radiations of the sun or other radiant source fall upon plate P, a current flows into the condenser, as above explained, until the potential therein rises sufficiently to attract and bring into contact the two plates *t t′*, and thereby close the circuit connected to the two condenser-terminals. This permits a flow of current which energizes the magnet M, causing it to draw down the armature *a* and impart a partial rotation to the ratchet-wheel *w*. As the current ceases the armature is retracted by the spring *b*, without, however, moving the wheel *w*. With the stoppage of the current the plates *t t′* cease to be attracted and separate, thus restoring the circuit to its original condition.

Fig. 3 shows a modified form of apparatus used in connection with an artificial source of radiant energy, which in this instance may be an arc emitting copiously ultra-violet rays. A suitable reflector may be provided for concentrating and directing the radiations. A magnet R and circuit-controller *d* are arranged as in the previous figures; but in the present case the former instead of performing itself the whole work only serves the purpose of alternately opening and closing a local circuit, containing a source of current B and a receiving or translating device D. The controller *d*, if desired, may consist of two fixed electrodes separated by a minute air-gap or weak dielectric film, which breaks down more or less suddenly when a definite difference of potential is reached at the terminals of the condenser and returns to its original state upon the passage of the discharge.

Still another modification is shown in Fig. 4, in which the source S of radiant energy is a special form of Roentgen tube devised by me, having but one terminal *k*, generally of aluminium, in the form of half a sphere, with a plain polished surface on the front side, from which the streams are thrown off. It may be excited by attaching it to one of the terminals of any generator of sufficiently high electromotive force; but whatever apparatus be used it is important that the tube be exhausted to a high degree, as otherwise it might prove entirely ineffective. The working or discharge circuit connected to the terminals T T′ of the condenser includes in this case the primary *p* of a transformer and a circuit-controller comprising a fixed terminal or brush *t* and a movable terminal *t′* in the shape of a wheel, with conducting and insulating segments, which may be rotated at an arbitrary speed by any suitable means. In inductive relation to the primary wire or coil *p* is a secondary *s*, usually of a much greater number of turns, to the ends of which is connected a receiver R. The terminals of the condenser being connected, as indicated, one to an insulated plate P and the other to a grounded plate P′, when the tube S is excited rays or streams of matter are emitted from the same, which convey a positive charge to the plate P and condenser-terminal T, while terminal T′ is continuously receiving negative electricity from the plate P′. This, as before explained, results in an accumulation of electrical energy in the condenser, which goes on as long as the circuit including the primary *p* is interrupted. Whenever the circuit is closed owing to the rotation of the terminal *t′*, the stored energy is discharged through the primary *p*, this giving rise in the secondary *s* to induced currents, which operate the receiver R.

It is clear from what has been stated above that if the terminal T′ is connected to a plate supplying positive instead of negative electricity the rays should convey negative electricity to plate P. The source S may be any form of Roentgen or Lenard tube; but it is obvious from the theory of action that in order to be very effective the electrical impulses exciting it should be wholly or at least preponderatingly of one sign. If ordinary symmetrical alternating currents are employed, provision should be made for allowing the rays to fall upon the plate P only during those periods when they are productive of the desired result. Evidently if the radiations of the source be stopped or intercepted or their intensity varied in any manner, as by periodically interrupting or rythmically varying the current exciting the source, there will be corresponding changes in the action upon the receiver R, and thus signals may be transmitted and many other useful effects produced. Furthermore, it will be understood that any form of circuit-closer which will respond to or be set in operation when a predetermined amount of energy is stored in the condenser may be used in lieu of the device specifically described with reference to Fig. 2 and also that the special details of construction and arrangement of the several parts of the apparatus may be very greatly varied without departure from the invention.

Having described my invention, what I claim is—

1. An apparatus for utilizing radiant energy, comprising in combination a condenser, one armature of which is subjected to the action of rays or radiations, independent means for charging the other armature, a circuit and apparatus therein adapted to be operated or controlled by the discharge of the condenser, as set forth.

2. An apparatus for utilizing radiant energy, comprising in combination, a condenser, one armature of which is subjected to the action of rays or radiations, independent means for charging the other armature, a local circuit connected with the condenser-terminals, a circuit-controller therein and means adapted to be operated or controlled by the discharge of the condenser when the local circuit is closed, as set forth.

3. An apparatus for utilizing radiant energy, comprising in combination, a condenser, one terminal of which is subjected to the action of rays or radiations, independent means for charging the other armature, a local circuit connected with the condenser-terminals, a circuit-controller therein dependent for operation on a given rise of potential in the condenser, and devices operated by the discharge of the condenser when the local circuit is closed, as set forth.

4. An apparatus for utilizing radiant energy, comprising in combination, a condenser, one terminal of which is subjected to the action of rays or radiations, and the other of which is connected with the ground, a circuit and apparatus therein adapted to be operated by the discharge of the accumulated energy in the condenser, as set forth.

5. An apparatus for utilizing radiant energy, comprising in combination, a condenser, one terminal of which is subjected to the action of rays or radiations and the other of which is connected with the ground, a local circuit connected with the condenser-terminals, a circuit-controller therein and means adapted to be operated by the discharge of the condenser when the local circuit is closed, as set forth.

6. An apparatus for utilizing radiant energy, comprising in combination, a condenser, one terminal of which is subjected to the action of rays or radiations and the other of which is connected with the ground, a local circuit connected with the condenser-terminals, a circuit-controller therein adapted to be operated by a given rise of potential in the condenser, and devices operated by the discharge of the condenser when the local circuit is closed, as set forth.

7. An apparatus for utilizing radiant en-

ergy, comprising a condenser, having one terminal connected to earth and the other to an elevated conducting-plate, which is adapted to receive the rays from a distant source of radiant energy, a local circuit connected with the condenser-terminals, a receiver therein, and a circuit-controller therefor which is adapted to be operated by a given rise of potential in the condenser, as set forth.

NIKOLA TESLA.

Witnesses:
M. LAWSON DYER,
RICHARD DONOVAN.

N. TESLA.

APPARATUS FOR TRANSMITTING ELECTRICAL ENERGY.

APPLICATION FILED JAN. 18, 1902. RENEWED MAY 4, 1907.

1,119,732. Patented Dec. 1, 1914.

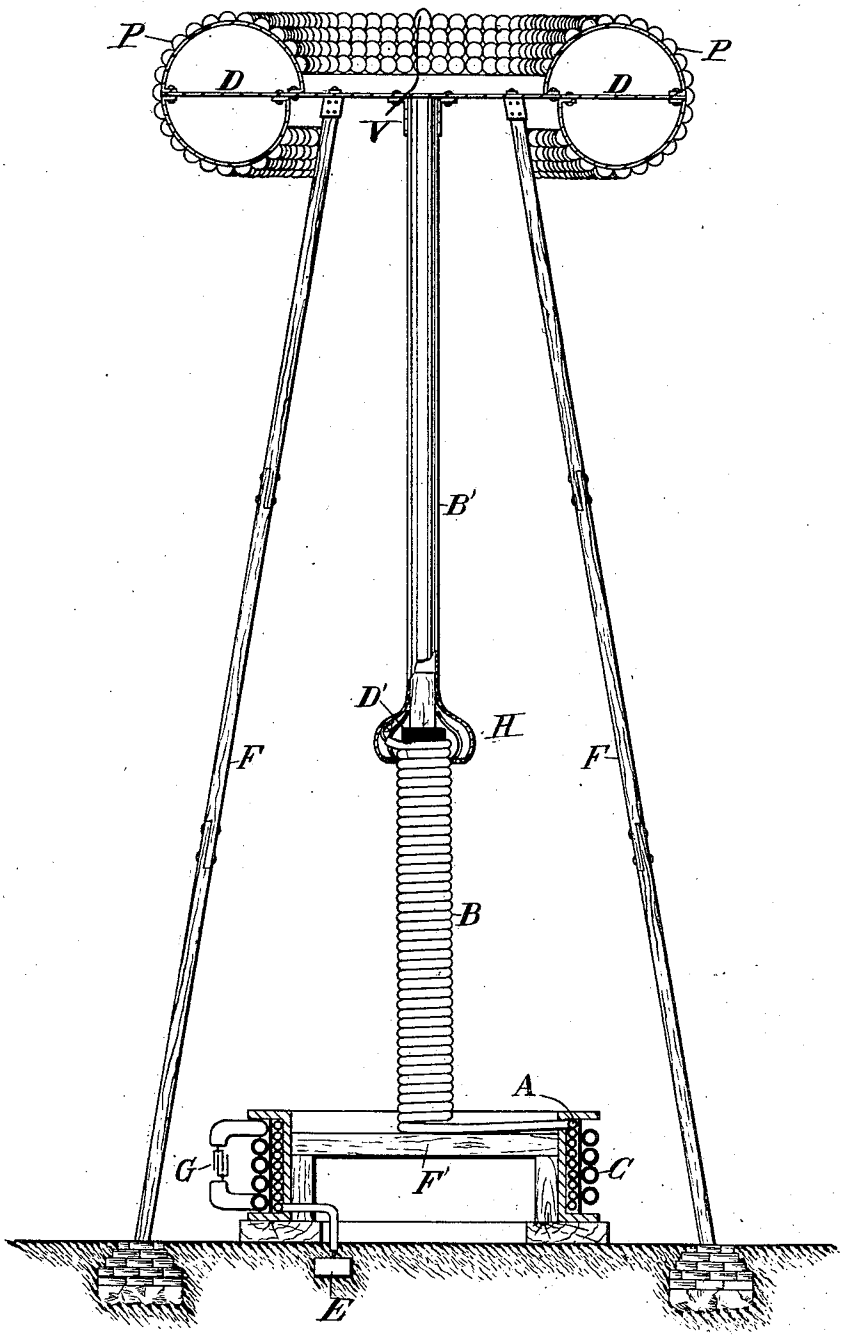

WITNESSES:

M. Lawson Dyer

Benjamin Miller.

Nikola Tesla, INVENTOR,

BY Kerr, Page & Cooper,

his ATTORNEYS.

UNITED STATES PATENT OFFICE.

NIKOLA TESLA, OF NEW YORK, N. Y.

APPARATUS FOR TRANSMITTING ELECTRICAL ENERGY.

1,119,732. Specification of Letters Patent. **Patented Dec. 1, 1914.**

Application filed January 18, 1902, Serial No. 90,245. Renewed May 4, 1907. Serial No. 371,817.

To all whom it may concern:

Be it known that I, Nikola Tesla, a citizen of the United States, residing in the borough of Manhattan, in the city, county, and State of New York, have invented certain new and useful Improvements in Apparatus for Transmitting Electrical Energy, of which the following is a specification, reference being had to the drawing accompanying and forming a part of the same.

In endeavoring to adapt currents or discharges of very high tension to various valuable uses, as the distribution of energy through wires from central plants to distant places of consumption, or the transmission of powerful disturbances to great distances, through the natural or non-artificial media, I have encountered difficulties in confining considerable amounts of electricity to the conductors and preventing its leakage over their supports, or its escape into the ambient air, which always takes place when the electric surface density reaches a certain value.

The intensity of the effect of a transmitting circuit with a free or elevated terminal is proportionate to the quantity of electricity displaced, which is determined by the product of the capacity of the circuit, the pressure, and the frequency of the currents employed. To produce an electrical movement of the required magnitude it is desirable to charge the terminal as highly as possible, for while a great quantity of electricity may also be displaced by a large capacity charged to low pressure, there are disadvantages met with in many cases when the former is made too large. The chief of these are due to the fact that an increase of the capacity entails a lowering of the frequency of the impulses or discharges and a diminution of the energy of vibration. This will be understood when it is borne in mind, that a circuit with a large capacity behaves as a slackspring, whereas one with a small capacity acts like a stiff spring, vibrating more vigorously. Therefore, in order to attain the highest possible frequency, which for certain purposes is advantageous and, apart from that, to develop the greatest energy in such a transmitting circuit, I employ a terminal of relatively small capacity, which I charge to as high a pressure as practicable. To accomplish this result I have found it imperative to so construct the elevated conductor, that its outer surface, on which the electrical charge chiefly accumulates, has itself a large radius of curvature, or is composed of separate elements which, irrespective of their own radius of curvature, are arranged in close proximity to each other and so, that the outside ideal surface enveloping them is of a large radius. Evidently, the smaller the radius of curvature the greater, for a given electric displacement, will be the surface-density and, consequently, the lower the limiting pressure to which the terminal may be charged without electricity escaping into the air. Such a terminal I secure to an insulating support entering more or less into its interior, and I likewise connect the circuit to it inside or, generally, at points where the electric density is small. This plan of constructing and supporting a highly charged conductor I have found to be of great practical importance, and it may be usefully applied in many ways.

Referring to the accompanying drawing, the figure is a view in elevation and part section of an improved free terminal and circuit of large surface with supporting structure and generating apparatus.

The terminal D consists of a suitably shaped metallic frame, in this case a ring of nearly circular cross section, which is covered with half spherical metal plates P P, thus constituting a very large conducting surface, smooth on all places where the electric charge principally accumulates. The frame is carried by a strong platform expressly provided for safety appliances, instruments of observation, etc., which in turn rests on insulating supports F F. These should penetrate far into the hollow space formed by the terminal, and if the electric density at the points where they are bolted to the frame is still considerable, they may be specially protected by conducting hoods as H.

A part of the improvements which form the subject of this specification, the transmitting circuit, in its general features, is identical with that described and claimed in my original Patents Nos. 645,576 and 649,621. The circuit comprises a coil A which is in close inductive relation with a primary C, and one end of which is connected to a ground-plate E, while its other end is led through a separate self-induction coil B and a metallic cylinder B′ to the terminal D.

The connection to the latter should always be made at, or near the center, in order to secure a symmetrical distribution of the current, as otherwise, when the frequency is very high and the flow of large volume, the performance of the apparatus might be impaired. The primary C may be excited in any desired manner, from a suitable source of currents G, which may be an alternator or condenser, the important requirement being that the resonant condition is established, that is to say, that the terminal D is charged to the maximum pressure developed in the circuit, as I have specified in my original patents before referred to. The adjustments should be made with particular care when the transmitter is one of great power, not only on account of economy, but also in order to avoid danger. I have shown that it is practicable to produce in a resonating circuit as E A B B′ D immense electrical activities, measured by tens and even hundreds of thousands of horse-power, and in such a case, if the points of maximum pressure should be shifted below the terminal D, along coil B, a ball of fire might break out and destroy the support F or anything else in the way. For the better appreciation of the nature of this danger it should be stated, that the destructive action may take place with inconceivable violence. This will cease to be surprising when it is borne in mind, that the entire energy accumulated in the excited circuit, instead of requiring, as under normal working conditions, one quarter of the period or more for its transformation from static to kinetic form, may spend itself in an incomparably smaller interval of time, at a rate of many millions of horse power. The accident is apt to occur when, the transmitting circuit being strongly excited, the impressed oscillations upon it are caused, in any manner more or less sudden, to be more rapid than the free oscillations. It is therefore advisable to begin the adjustments with feeble and somewhat slower impressed oscillations, strengthening and quickening them gradually, until the apparatus has been brought under perfect control. To increase the safety, I provide on a convenient place, preferably on terminal D, one or more elements or plates either of somewhat smaller radius of curvature or protruding more or less beyond the others (in which case they may be of larger radius of curvature) so that, should the pressure rise to a value, beyond which it is not desired to go, the powerful discharge may dart out there and lose itself harmlessly in the air. Such a plate, performing a function similar to that of a safety valve on a high pressure reservoir, is indicated at V.

Still further extending the principles underlying my invention, special reference is made to coil B and conductor B′. The latter is in the form of a cylinder with smooth or polished surface of a radius much larger than that of the half spherical elements P P, and widens out at the bottom into a hood H, which should be slotted to avoid loss by eddy currents and the purpose of which will be clear from the foregoing. The coil B is wound on a frame or drum D^1 of insulating material, with its turns close together. I have discovered that when so wound the effect of the small radius of curvature of the wire itself is overcome and the coil behaves as a conductor of large radius of curvature, corresponding to that of the drum. This feature is of considerable practical importance and is applicable not only in this special instance, but generally. For example, such plates at P P of terminal D, though preferably of large radius of curvature, need not be necessarily so, for provided only that the individual plates or elements of a high potential conductor or terminal are arranged in proximity to each other and with their outer boundaries along an ideal symmetrical enveloping surface of a large radius of curvature, the advantages of the invention will be more or less fully realized. The lower end of the coil B—which, if desired, may be extended up to the terminal D—should be somewhat below the uppermost turn of coil A. This, I find, lessens the tendency of the charge to break out from the wire connecting both and to pass along the support F′.

Having described my invention, I claim:

1. As a means for producing great electrical activities a resonant circuit having its outer conducting boundaries, which are charged to a high potential, arranged in surfaces of large radii of curvature so as to prevent leakage of the oscillating charge, substantially as set forth.

2. In apparatus for the transmission of electrical energy a circuit connected to ground and to an elevated terminal and having its outer conducting boundaries, which are subject to high tension, arranged in surfaces of large radii of curvature substantially as, and for the purpose described.

3. In a plant for the transmission of electrical energy without wires, in combination with a primary or exciting circuit a secondary connected to ground and to an elevated terminal and having its outer conducting boundaries, which are charged to a high potential, arranged in surfaces of large radii of curvature for the purpose of preventing leakage and loss of energy, substantially as set forth.

4. As a means for transmitting electrical energy to a distance through the natural media a grounded resonant circuit, comprising a part upon which oscillations are impressed and another for raising the ten-

sion, having its outer conducting boundaries on which a high tension charge accumulates arranged in surfaces of large radii of curvature, substantially as described.

5. The means for producing excessive electric potentials consisting of a primary exciting circuit and a resonant secondary having its outer conducting elements which are subject to high tension arranged in proximity to each other and in surfaces of large radii of curvature so as to prevent leakage of the charge and attendant lowering of potential, substantially as described.

6. A circuit comprising a part upon which oscillations are impressed and another part for raising the tension by resonance, the latter part being supported on places of low electric density and having its outermost conducting boundaries arranged in surfaces of large radii of curvature, as set forth.

7. In apparatus for the transmission of electrical energy without wires a grounded circuit the outer conducting elements of which have a great aggregate area and are arranged in surfaces of large radii of curvature so as to permit the storing of a high charge at a small electric density and prevent loss through leakage, substantially as described.

8. A wireless transmitter comprising in combination a source of oscillations as a condenser, a primary exciting circuit and a secondary grounded and elevated conductor the outer conducting boundaries of which are in proximity to each other and arranged in surfaces of large radii of curvature, substantially as described.

9. In apparatus for the transmission of electrical energy without wires an elevated conductor or antenna having its outer high potential conducting or capacity elements arranged in proximity to each other and in surfaces of large radii of curvature so as to overcome the effect of the small radius of curvature of the individual elements and leakage of the charge, as set forth.

10. A grounded resonant transmitting circuit having its outer conducting boundaries arranged in surfaces of large radii of curvature in combination with an elevated terminal of great surface supported at points of low electric density, substantially as described.

NIKOLA TESLA.

Witnesses:
M. LAMSON DYER,
RICHARD DONOVAN.

Dec. 17, 1963 W. C. BROWN **3,114,517**

MICROWAVE OPERATED SPACE VEHICLES

Filed May 12, 1959 2 Sheets-Sheet 1

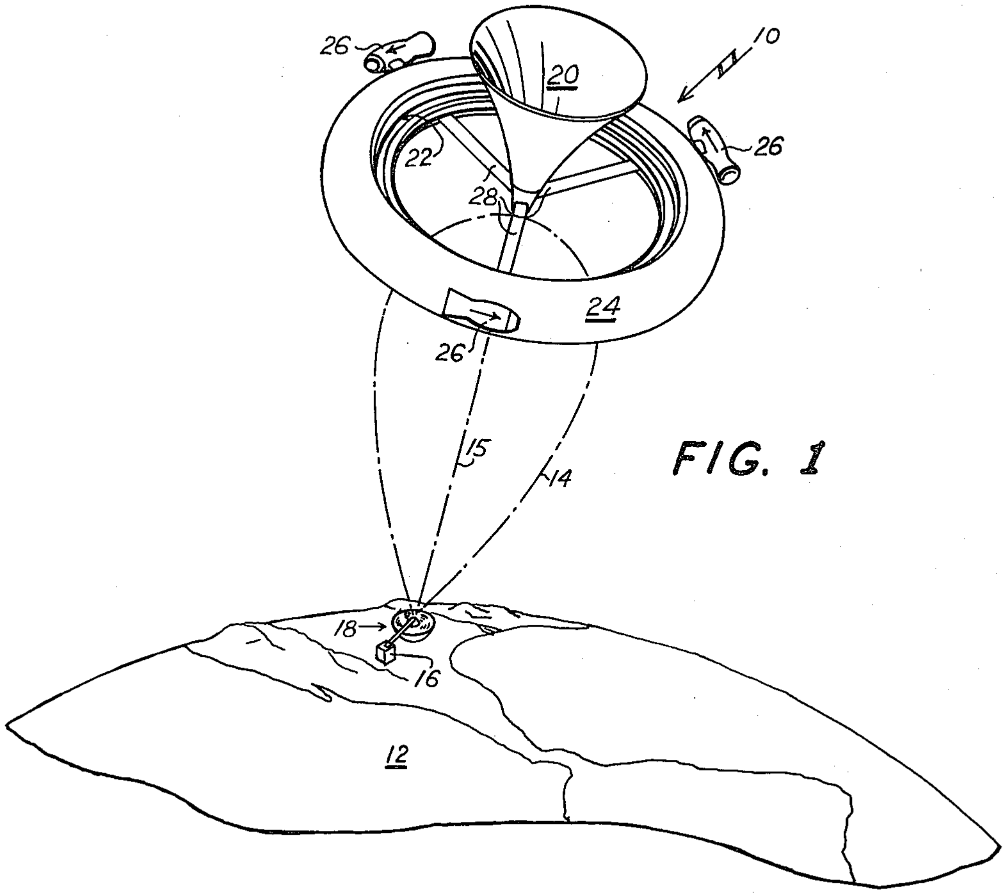

26
10
20
26
22
28
24
26
15
14
18
16
12

FIG. 1

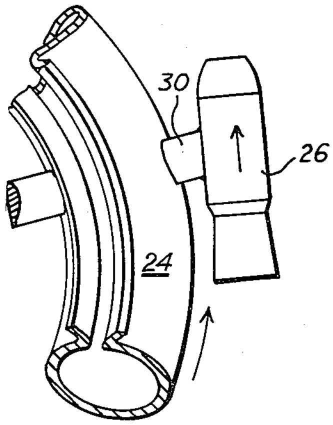

FIG. 3

INVENTOR
WILLIAM C. BROWN
BY H. Vincent Harsha
ATTORNEY

Dec. 17, 1963 W. C. BROWN **3,114,517**

MICROWAVE OPERATED SPACE VEHICLES

Filed May 12, 1959 2 Sheets-Sheet 2

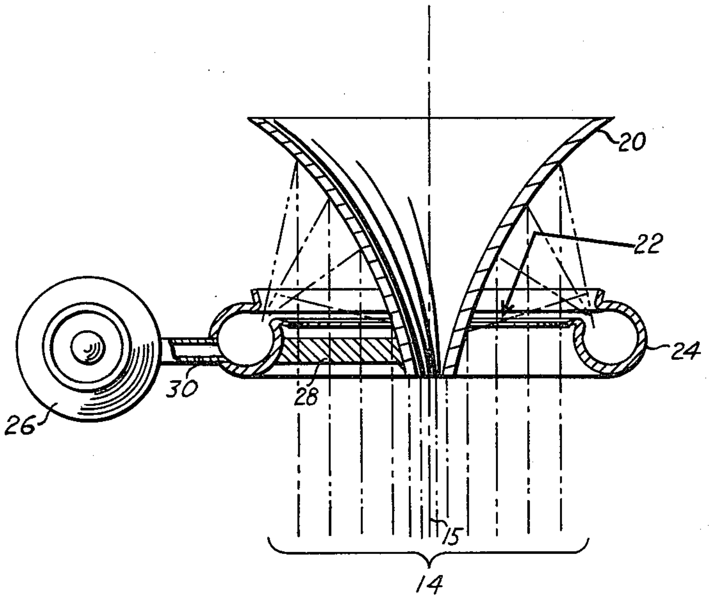
20
22
24
26
30
28
15
14

FIG. 2

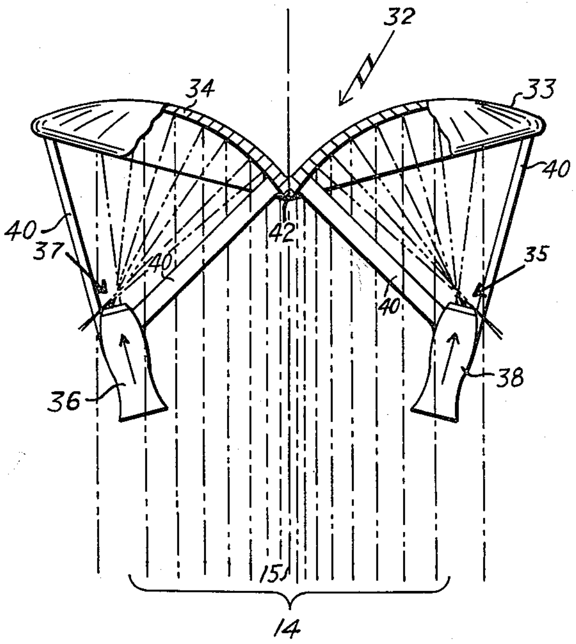
32
34
33
40
40
37
40
42
35
40
36
38
15
14

FIG. 4

INVENTOR
WILLIAM C. BROWN
BY H. Vincent Harsha
ATTORNEY

3,114,517
MICROWAVE OPERATED SPACE VEHICLES

William C. Brown, Weston, Mass., assignor to Raytheon Company, Lexington, Mass., a corporation of Delaware

Filed May 12, 1959, Ser. No. 812,710
11 Claims. (Cl. 244—1)

This invention pertains generally to space vehicles and the like, and more particularly to a novel form of space vehicle and a system for maintaining such vehicle in flight or at a desired location in spaced-apart relationship with respect to a mother planet or the like.

The prior art affords many examples of apparatus adapted to be propelled in spaced-apart relationship with the earth, as well as systems for maintaining this relationship in opposition to the gravitational field thereof. Exemplary devices of this type are conventional aircraft (both jet and propeller driven) and the more modern un-manned or pilotless vehicles such as rocket-propelled missiles. However, the operation of each of these vehicles of the prior art necessitates the expenditure of a fuel carried by the vehicle. Further, while an orbital Earth satellite requires no applied energy in executing a given orbit, enormous quantities of fuel are consumed in launching the satellite into orbit. Also, where the satellite vehicle is designed for the subsequent correction of or other change in the orbital path, a fuel supply must be carried by the satellite in order to provide the desired corrective thrusts.

It is thus apparent that in each of these prior art vehicles a fuel supply must be carried thereby for local consumption in order to initiate, maintain or alter a given flight pattern, whether such pattern involves motion of the vehicle relative to the Earth or a substantially motionless hovering with respect thereto. It therefore follows that the maximum payload capabilities of such vehicles are limited by the fact that the vehicle must support the weight of the fuel carried thereby. Also, such vehicles are not naturally adapted to indefinitely long periods of continuous operation, since the fuel supply therein will eventually be exhausted, necessitating a return of the vehicle to a refueling base, or alternatively, an in-flight refueling process as is accomplished with some conventional aircraft.

The utility of a space vehicle or the like which is capable of sustained flight over an indefinitely long period of time without employing a local fuel supply is evident, and the applications therefor are numerous. For example, a vehicle capable of perpetual operation (except for mechanical failure) and stationed in space above the surface of the Earth, either within or without the atmosphere thereof, may perform alarm and surveillance tasks by radar techniques, along with guidance and protection of defense vehicles, and may simultaneously provide long-haul broadband communications. Systems of such vehicles or platforms may be established to provide radar and communication networks.

These desired operational characteristics are provided by the vehicle of the present invention, which vehicle is energized by means of transmitted microwave electromagnetic energy that is beamed toward the vehicle. The microwave energy incident upon the vehicle is converted into appropriate mechanical forces which produce the desired flight operation of the vehicle. The advantages attendant the utilization of microwave energy, in contrast to electromagnetic energy of other wavelengths, are readily apparent. Microwaves have been generally defined as radio waves whose wavelength is less than 30 centimeters, with a lower wavelength limit on the order of one millimeter or one centimeter sometimes being applied to what is called "the microwave region." The superiority of microwaves is due in part to the fact that

it is generally necessary to focus the transmitted energy so as to achieve a desirably high power density at a remote point or area with respect to a given generator or transmitter power level. In accordance with the laws of optics, the sharpness of the beam produced by an antenna varies as the ratio of antenna dimensions to the wavelength of the transmitted energy. Thus, for a given or desired power density or beam sharpness, a decrease in the wavelength of the transmitted energy permits a corresponding decrease in the dimensions of the antenna. From the standpoint of mechanical convenience, it is generally desirable to employ small antennas and other components, and it is therefore advantageous to employ energy of very short wavelength. In addition, the difficulties encountered in relatively long wave transmission as a result of natural and man-made interference or noise (which noise would interfere with the transmission of radar or communications intelligence via the microwave power beam) do not occur with any appreciable significance at microwave frequencies. Also, where the space vehicle is to be operated at an altitude above the ionosphere, long wave transmission will generally be reflected by the reflecting layers thereof, while microwaves pass through such layers substantially unaffected.

In view of the several advantages stated above, it might appear desirable to employ the shortest wavelength possible commensurate with the power-generating capabilities of the transmitter at the wavelength chosen. A lower limit is fixed, however, for the wavelength that may be used in a practical radiating system in view of the increasing losses due to absorption occurring in the atmosphere at wavelengths below five to ten centimeters. In the one to two centimeter region, there is a peak in absorption by water vapor, and even for dry air the absorption of electromagnetic energy increases very rapidly below one or two centimeters. It may thus be seen that microwaves in a region having the approximate bounds of two and thirty centimeters are readily adaptable to convenient radiation of energy to a remote point with small transmission loss, with the preferred wavelengths being of the order of five or ten centimeters in order to accomplish efficient focusing with a transmitting lens system of reasonable size without inflicting an intolerable power loss by absorption.

The key to the practical utilization of high-power electromagnetic beams for remote energization of space vehicles and the like is a device which will generate large amounts of power within this wavelength region. In addition, many applications of such high-power beams require such refinements as broad electronic bandwidth and low phase distortion, placing the additional requirement of sophisticated performance on the high-power generator. A device which currently satisfies the duel requirements of high power output and refined performance is the Amplitron tube, a relatively new type of crossed-field vacuum tube which may be used as a compact, highly efficient, broadband amplifier capable of handling high peak and average powers, and which generally comprises a circular but non-reentrant, dispersive network matched at both ends over the frequency region of interest, and a reentrant electron beam originating from a continuously-coated (or nearly so) cathode coaxial with the network, with a D.-C. potential being applied between the cathode and anode, and a magnetic field applied parallel to the axis of the cathode and transverse to the electric field between the anode and the cathode. For a more complete description of the Amplitron, reference may be had to my application Serial No. 706,812, filed January 2, 1958, for "Low Level Duplexer System," now United States Letters Patent No. 2,933,723, which issued April 19, 1960. Amplitrons currently available are capable of producing 15 or 20 kilowatts of average radio-frequency power in the neighborhood of ten centimeters in wavelength, and future models are expected to yield 500 kilowatts or more average power, with 50 megawatt peak power.

It is accordingly a primary object of this invention to provide a space vehicle or the like designed for an indefinitely long period of operation either in moving flight or at a selectable location with respect to the Earth or other mother planet, with energy necessary for such operation being supplied to such device in the form of microwave electormagnetic energy.

A further object of the invention is to provide a system for maintaining a space vehicle either in moving flight or at a selectable location with respect to a mother planet or the like in response to microwave electromagnetic energy transmitted from such mother planet to the vehicle.

In accordance with an exemplary embodiment of the present invention, this system for maintaining flight of the space vehicle includes means on the Earth or some other mother planet for radiating microwave energy in a beam directed toward and focused on the vehicle for substantially continuously supplying such vehicle with the power necessary for the operation thereof. The space vehicle of the invention has means associated therewith for receiving and converting incident microwave energy into appropriate forces applied to the vehicle to sustain flight. The received microwave energy may be directly converted into a flight-producing force as by an ion engine or the like, or the microwave energy may be subjected to a mesne conversion to heat. In the latter case, means are provided for applying the heat converted to a suitable fluid medium which is expanded thereby to perform the work necessary to propel or sustain the vehicle.

In one preferred form of the apparatus of the invention, a plurality of thrust-providing engines are mounted for rotary movement about a substantially vertical common axis along which the beam of energy is normally transmitted, with suitable airfoil means being mounted for movement with such engines to provide the lifting force necessary to maintain flight of such vehicle. Means are provided for apportioning the received energy between the several engines in such manner as to selectively energize the engines to maintain the vehicle centered on the transmitted beam.

In an alternative form of the apparatus of this invention, a plurality of thrust-providing engines are symmetrically positioned with respect to the vertical central axis of the vehicle along which the beam of energy is normally transmitted. Means are provided for distributing the received energy between the engines so as to differentially energize the engines to apply corrective forces to the vehicle upon the occurrence of angular or translational displacement of the axis thereof from the center line of the radiated beam.

With the above considerations and objects in mind, the invention itself will now be described in connection with a preferred embodiment thereof given by way of example and not of limitation, and with reference to the accompanying drawings, in which:

FIG. 1 is a perspective view of a preferred form of the apparatus of the present invention, schematically indicating the control system therefor;

FIG. 2 is a partial sectional view in elevation of the vehicle shown in FIG. 1, the section being taken substantially centrally of the vehicle;

FIG. 3 is a plan view of a portion of the apparatus of FIG. 1; and

FIG. 4 is a partial sectional view in elevation of an alternative form of the apparatus of this invention.

Referring now to FIG. 1, one preferred form of the vehicle of the present invention is indicated generally at **10**. As shown, the vehicle **10** is adapted to be maintained in flight or in hovering position with respect to the Earth or other mother planet **12** by means of energy transmitted to the vehicle as radiated microwave electromagnetic energy in beam form, as indicated by lobe **14** having center line **15**. A suitable transmitter **16** is provided for this purpose, along with a beam-forming transmitting antenna **18**. As will be appreciated by those skilled in the art, the representations of the transmitting equipment and microwave beam in FIG. 1 are highly diagrammatic in form, and the present invention is not limited to any particular transmitter or antenna assembly. As a practical matter, however, the reflector of the antenna assembly **18** is considerably larger than most of the reflectors of the prior art in order to focus a large amount of power at a high-altitude (e.g. 65,000 feet) space vehicle. In an exemplary embodiment, the antenna reflector is at least partially supported in a large bowl or other dug-out area on the Earth's surface so as to provide a convenient means of support therefor. Also, the antenna may actually comprise a large number of small horn and reflector assemblies. In addition, the transmitter **16** is capable of much greater power output than most of the prior art transmitters, and, as stated in the earlier paragraphs of this specification, the power-generating element of the transmitter is preferably an Amplitron tube in order to achieve sufficient power output.

The microwave energy that illuminates the vehicle **10** is intercepted by the under or concave side of the reflecting member **20**, the curvature of which is such that the energy received along a substantially straight line parallel to the central axis of the radiated beam of energy is reflected into an annular opening **22** in the frame member **24**. The latter frame member may preferably comprise a hollow waveguide for the received radio-frequency energy, and the energy applied thereto through the opening **22** is, in turn, applied to the several engines **26** connected thereto.

The means for maintaining flight or hovering conditions in the vehicle **10** include the several engines **26**, which effect a rotary motion of the vehicle frame in the direction shown by the arrows about a substantially vertical central axis normally coinciding with the center line of the transmitted beam **14** of microwave energy. Upon such rotary motion of the vehicle frame, a sustaining lift is provided by means of a plurality of airfoil members **28**, which may generally be in the form of the rotary wings of a helicopter or the like. It will be understood that only the basic elements of the vehicle **10** are shown in the drawing, and that other frame members and the like would normally be provided to support a payload of radar, communication or other similar equipment. In this connection, the beam **14** also provides a signal-information link between the ground station and the payload equipment on the vehicle.

FIG. 2 shows in somewhat greater detail the construction of the apparatus of FIG. 1, with the showing of FIG. 2 being taken along a vertical plane through one of the supporting airfoil members **28**. The construction and operation of the several thrust engines will not be discussed in detail herein, since a complete description of engines suitable for this purpose is provided in my application Serial Number 812,697, filed May 12, 1959, now United States Letters Patent No. 3,083,528, which issued April 2, 1963, and assigned to the assignee of the present application. As disclosed in such application, an exemplary embodiment of an engine suitable for use in the vehicle of the present application includes a heat-exchanging container means adapted to be energized by microwave energy to heat a gaseous medium therein and cause expansion of same to produce a jet stream for providing a propulsive thrust. In one preferred form of the invention, the container means comprises an electromagnetic waveguide of lossy characteristics which produces heat upon energization by suitable electromagnetic energy to raise the temperature of a gaseous medium passing therethrough. A member of stacked and closely-spaced electrically lossy plates is preferably mounted within the waveguide to serve as the main means for absorbing microwave energy and converting same into heat, as described in detail in the aforementioned copending application. The plurality of closely spaced plates form therebetween a series of narrow passages through which passes the gaseous medium which is to be heated thereby. The microwave energy employed is of sufficiently high frequency to produce a pronounced skin effect in the waveguide and associated heat-exchanger, so that the energy expended in creating heat in the steady state condition is effectively and quickly transferred to the gaseous medium in which the particular conductor exhibiting the skin effect is immersed. This latter feature is particularly advantageous in transient-state conditions with respect to the ability of the engine of this invention to provide almost instantaneous response to a change in the level of the applied microwave energy, since little or no time lag is involved in converting electromagnetic energy to heat and transferring it to the cooling gas. As is evident, the heated gas may be formed into a jet stream in order to provide propulsion, or a turbine may be driven thereby to provide a rotary shaft output for driving a propeller or the like.

As may be seen in FIG. 2, the engines 26 are supported from the rotary frame member 24 by means of suitable supporting members 30 which also serve as the means for transferring the microwave energy in the interior of the frame member 24 to the interior of the engines 26. Thus, each of the supporting members 30 must be adapted to act as a transmission line of short length for the radio-frequency energy employed. The particular curvature of the reflecting member 20 is chosen so as to effect the reflection of radiated energy indicated in FIG. 2, viz., the energy transmitted toward the vehicle reaches the reflector member 20 along substantially straight parallel paths as shown by the vertical dash-dot lines, and upon reflection all of the received energy is focused on the opening 22 in the frame member 24 so that it may be transferred into the several engines 26.

FIG. 3 shows that the direction of thrust produced by each of the engines 26 is not quite tangential to the circle of rotation thereof. As shown, the engines 26 are mounted at a small angle with respect to the periphery of the frame member 24, so that while the major portion of the thrust produced by each of the engines is tangential to such periphery, a small component thereof is directed away from the center of the vehicle. This construction is employed in order to provide a vehicle which is capable of a self-centering or beam riding operation, as will be more fully described below.

In the operation of the device thus far described, transmitter 16 generates a suitable microwave signal for transmission to the vehicle 10 by means of the radiating antenna 18. In the normal operation of the device, the reflecting member 20 will be centrally positioned in the radiated beam 14, and the energy of the beam will therefore be reflected in equal amounts toward all portions of the annular opening 22 in frame member 24. As a result, the three engines 26 will receive equal amounts of the radio-frequency energy, and normal operation is thereby effected, resulting in a rotary movement of the engines 26 and the frame member 24 about the central vertical axis of the vehicle 10. As mentioned above, such rotational movement of the frame member 24 produces a corresponding rotation of the airfoil members 28 to provide a vertical thrust for maintaining the vehicle 10 in flight or in a hovering position with respect to the mother planet 12.

In order to maintain the vehicle 10 in a desired location with respect to the center-line of the transmitted radio-frequency beam 14, viz., at a centered position with the vertical central axis of the vehicle coinciding with the center line of the transmitted beam, the shape of the concave reflecting surface of member 20 is chosen as described above to reflect energy of the beam 14 equally to the several engines 26 when the vehicle 10 is properly centered in the beam 14. Upon a lateral displacement of the vehicle 10 from the beam 14, one side of the reflecting surface 20 will be illuminated by more radiated energy than the opposite side, and the several engines 26 are thereupon differentially energized to provide corrective lateral forces to achieve recentering of the vehicle in the beam. It will be understood that the desired tolerance as to vehicle position and the lateral dimensions of the beam and the vehicle are so interrelated that a displacement beyond such tolerance will result in a significant degree of such differential energization. For example, if the vehicle 10 were to be displaced toward the right in FIG. 2, the left-hand portion of the reflecting member 20 would lie nearer the center of the beam 14, and the engine shown in FIG. 2 would thereupon receive a greater amount of energy than the other engines. Since the engine 26 of FIG. 2 (as well as each of the remaining engines of the vehicle) provides a component of thrust normal to the circle of rotation thereof and directed away from the center of the vehicle, such additional energization of the engine shown in FIG. 2 will provide a resultant lateral thrust toward the left in this figure. It is evident that the particular engine 26 shown in FIG. 2 will be in motion, and will therefore not remain at the position shown to continuously receive the major portion of the energy intercepted by reflector member 20. As each of the engines 26 passes through the azimuthal position of rotation opposite to the direction of lateral displacement of the vehicle 10 from the center of the beam 14, it then receives a greater energization than the remaining engines, and such differential energization of the several engines 26 continues with the rotational movement of such engines until the vehicle is once again centered in the beam 14. As will be appreciated by those skilled in the art, there must be a minimum lag in the transient response of each of the engines 26 in order to achieve the self-centering action. In this connection, the efficacy of the engine of my aforementioned copending application is evident. As stated above, the thermal time lag of this engine is minimal as a result of the fact that there is a significant skin effect in the microwave heat exchanger, and the transient-state heat changes are produced in intimate proximity to the fluid medium to be heated. As a result of the propinquity of the heat source (the shallow skin effect penetration of the heat-exchanging wall) and the sink (the fluid medium to be heated), such engines are able to respond almost instantaneously to a change in the power input, thus making the self-centering of vehicle 10 feasible.

An alternative form of the apparatus of this invention is shown in FIG. 4, with the vehicle therein being indicated generally at 32. As shown, the vehicle 32 includes reflector numbers 33 and 34 having concave lower surfaces for reflecting incident microwave energy of beam 14 to respective focal points indicated generally at 35 and 37. A plurality of supporting arms 40 are provided for mounting the respective engines 36 and 38 on the reflecting members 34 and 33. The diagrammatic showings of engines 36 and 38 in FIG. 4 are intended to represent jet engines of the type disclosed in my aforementioned copending application, and it is to be understood that the microwave energy necessary for operation of each engine 36 and 38 is received at the upper or intake end thereof. The focus "points" 35 and 37 are in reality areas rather than points, and the engines 36 and 38 are intentionally mounted slightly inboard of a centered position with respect to the associated area 35 or 37, so that normally not all of the energy at each area is directed into the respective engine. This condition is indicated in FIG. 4 by the several reflected dot-dash lines which pass by rather than into the mouths of engines 36 and 38. The particular proportion of energy that is thus normally wasted, as compared to that portion which is collected by the engines, is a matter of choice in design in a given instance, but it is generally desirable to utilize a major portion (e.g., 85 or 90 percent) of the available energy for normal energization of the engines. The magnitude of the portion that is normally wasted will generally be determined in accordance with the peak wind velocities occurring at the altitude at which the vehicle is to operate, as will be better understood in connection with the description of the operation of vehicle 32.

As in the case of the vehicle described in connection with the earlier figures, there will normally be other additional members employed in actual practice in order to support such operational or "pay load" equipment as radar, communications transceivers and the like. The configuration of the reflecting surface of members 33 and 34 is such as to equally divide energy of beam 14 between the engines mounted on the vehicle 32 when such vehicle is centered (both translationally and angularly) with respect to the beam 14, as shown in FIG. 4. As may be seen, the positions in which the engines 36 and 38 are mounted is such as to provide respective thrusts directed upwardly and away from the center of the vehicle 32. Thus, the thrust produced by each of the engines 36 and 38 has a major component substantially parallel to the direction of transmission of the radio-frequency energy, along with a minor component normal thereto and directed away therefrom. By means of this construction, the vehicle 32 may be maintained in flight or hovering condition and may also be controlled as to attitude in the plane of the drawing sheet. As will become apparent in connection with the description of the operation of the apparatus of FIG. 4, it will be necessary to provide three such engines to achieve attitude control in all directions.

In the operation of the vehicle of FIG. 4, when the vehicle is operating in a centered position with respect to the beam of microwave energy 14, with the central vertical axis of the vehicle in registry with the center line of the radiated beam, a propulsive thrust is provided for either propelling the vehicle upwardly or for maintaining a given altitude in opposition to gravitational forces, and the thrust components normal to the direction of the radiated beam 14 are equal in magnitude and so directed as to cancel each other.

When the vehicle 32 is moved from this desired position, as by wind forces or the like, the vehicle is re-centered on the beam by virtue of the structure described. For example, if the vehicle were to suffer translation toward the right as viewed in FIG. 4, the reflecting member 34 would lie nearer the center of the beam 14 and would therefore intercept a greater amount of radio-frequency energy that would member 33. Accordingly, engine 36 would be illuminated with more energy than falls on engine 38, and the oppositely-directed horizontal or lateral components of force exerted by the respective engines would result in a net lateral force applied to the vehicle 32 and directed toward the left in FIG. 4. This lateral correcting force will be so applied until the vehicle is once again entered on the beam. A corresponding operation obtains where the vehicle 32 is displaced laterally to the left in FIG. 4.

In addition, the structure of FIG. 4 provides automatic attitude correction to restore the vehicle to a desired position upon an unwanted angular displacement thereof with respect to the center line of the beam 14 of radiated energy. For example, if the vehicle 32 were to be angularly displaced in a clockwise direction about an axis perpendicular to the drawing sheet and passing through apex 42, the reflecting member 33 would, as a result, be moved toward a position more nearly normal to the direction of the incoming radiant energy. As a result, the angles of incidence and reflection are correspondingly increased, and the focus point 35 is moved toward the engine 38 to supply a greater level of energization thereto. A corresponding decrease in the energization of engine 36 obtains, and a net resultant force is applied to the vehicle 32 about the aforementioned axis at apex 42 to correct the attitude of the vehicle. As is evident, a similar but oppositely-directed operation occurs upon angular displacement of the vehicle 32 in a counter-clockwise direction.

The invention has been described above in considerable detail, and particularly with reference to its application to a space vehicle adapted to be maintained at a desired altitude and lateral position in spaced-apart relationship with respect to the Earth or other mother planet. However, it will be apparent to those skilled in the art that the invention is equally applicable to vehicles adapted for movement relative to such mother planet, whether involving a change in altitude, latitude or longitude, or some combination thereof, so long as a suitable beam of radiant energy is provided. The vehicle of FIG. 4 is shown and described as having a separate dished reflector for each engine. However, it is evident that a single reflecting surface of revolution may equally well be employed, along with a suitable annular collecting and distributing means, as indicated in FIGS. 1 and 2. Conversely, the separate reflectors of FIG. 4 may equally well be employed in the embodiment of FIGS. 1 and 2. Further, the embodiment of FIG. 4 is not necessarily strictly alternative to that of the preceding figures, since the two embodiments may be combined in a single vehicle to perform the combined results of each embodiment.

In addition, the invention is not limited to use of any particular form of thrust-providing engine (except as to the requirement that it be energized by received radiant energy), and the particular means for creating the propulsive thrust may be either a jet stream, a propeller or the like. In this connection, it should be noted that while the vehicle of the present invention is described as being energized solely by microwave electromagnetic energy, the vehicle may also include a conventional engine and fuel supply system for supplementing the action of the microwave engines until such fuel supply is exhausted. Also, the invention is applicable to tethered air vehicles, and where the operating altitude is low, the tethering means may serve as a transmission line for supplying energy to the vehicle without prohibitive losses. Hence the invention is not to be considered as being limited to the particular details given, nor to the specific application to which reference has been made during the description of the apparatus, except insofar as may be required by the scope of the appended claims.

What is claimed is:

1. A thrust-propelled space vehicle adapted to be maintained in a centered position with respect to a beam of radiated energy, comprising a plurality of thrust engines mounted for rotary movement about a common axis, a major portion of the thrust applied to such vehicle by each of said engines being substantially tangential to the circle of rotation thereof with a minor portion of such thrust being directed outwardly from such circle, airfoil means mounted for movement with said engines to apply a lifting force to such vehicle, means for receiving energy radiated toward such vehicle in a beam normally lying along said axis and for applying such energy to said plurality of engines to energize the latter thereby, the energy thus being applied to said engines being apportioned in accordance with the relative positions of each of said engines and the center of such beam upon a displacement of such beam away from said axis, with those engines disposed more nearly in the direction of displacement of such beam center from said axis at any given instant receiving relatively greater energization than others, whereby corrective forces are applied to such vehicle in directions substantially normal to said axis to maintain such vehicle centered on such beam.

2. A thrust-propelled space vehicle comprising a plurality of thrust engines mounted for rotation about a common axis at equal distances therefrom, a major portion of the thrust applied to such vehicle by each of said engines being substantially tangential to the circle of rotation thereof with a minor portion of such thrust being

directed outwardly from such circle, airfoil means mounted for movement with said engines to provide a lift force for sustaining air flight of such vehicle, means for receiving energy radiated toward such vehicle in a beam normally lying along said axis and for applying such energy to said plurality of engines in amounts apportioned in direct relationship with the respective distances of each of said engines from the center line of such beam at any given instant, whereby said engines are energized by such received energy to provide movement of said airfoil means and to provide corrective forces to such vehicle in directions substantially normal to said axis to maintain such vehicle centered on such beam.

3. A thrust-propelled space vehicle in accordance with claim 2, with such radiated energy being microwave electromagnetic energy, said engines including means for converting incident electromagnetic energy into propulsive thrust, and said energy receiving and applying means comprising a concave surface of revolution positioned on said axis to reflect the electromagnetic energy of such beam to said engines.

4. A thrust-propelled space vehicle adapted to be maintained in a centered position with respect to a beam of radiated energy, comprising a pair of thrust engines adapted to be energized by such radiated energy and mounted in symmetrical relationship on opposite sides of a central axis of such vehicle, said pair of engines being mounted in a substantially common plane including such central axis and at angles with respect to such axis to apply to such vehicle respective thrusts having components directed away from such axis, and means symmetrically disposed about such axis for directing energy from such beam to said engines in respective amounts apportioned in accordance with the relative distances between each of said engines and the center line of such beam, whereby said engines are differentially energized to exert corrective thrusts upon displacement of such axis from the center line of such beam.

5. A thrust-propelled space vehicle in accordance with claim 4, with such radiated energy being microwave electromagnetic energy, said engines including means for converting electromagnetic energy into propulsive thrust, and said directing means comprising a pair of concave reflecting members each reflecting a portion of the energy of such beam to a respective one of said converting means.

6. A thrust-propelled space vehicle adapted to be maintained in a centered position with respect to a beam of radiated energy, comprising a plurality of thrust engines adapted to be energized by such radiated energy and mounted in symmetrical relationship with respect to a central axis of such vehicle, the thrust produced by each of said engines being substantially confined to a respective plane including such central axis and having components both parallel with and directed away from such central axis, and means symmetrically disposed about such axis for directing energy from such beam to said engines in respective amounts apportioned in accordance with the relative distances between each of said engines and the center line of such beam, whereby said engines are differentially energized to exert corrective thrusts upon displacement of such axis from the center line of such beam.

7. A system for maintaining a thrust-propelled space vehicle in spaced-apart relationship with a mother planet at a selectable location with respect thereto, comprising a control station on such mother planet for transmitting microwave electromagnetic energy in a substantially vertically-directed beam to such vehicle, a plurality of thrust engines on such vehicle mounted for rotary movement about a substantially vertical common axis, a major portion of the thrust applied to such vehicle by each of said engines being substantially tangential to the circle of rotation thereof with a minor portion of such thrust being directed outwardly from such circle, airfoil means mounted for movement with said engines to provide a

lifting force for sustaining air flight of such vehicle, means for receiving energy transmitted toward such vehicle along said axis from said control station and for applying such energy to said plurality of engines to energize the latter thereby, the energy thus being applied to said engines being apportioned in accordance with the relative positions of each of said engines and the center of such beam upon a displacement of such beam away from said axis, with those engines disposed more nearly in the direction of displacement of such beam center from said axis at any given instant receiving relatively greater energization than others, whereby corrective forces are applied to such vehicle in directions substantially normal to said axis to maintain such vehicle centered on such beam.

8. A system for maintaining a thrust-propelled space vehicle in spaced-apart relationship with a mother planet at a selectable location with respect thereto, comprising a control station on such mother planet for transmitting microwave electromagnetic energy in a substantially vertically-directed beam toward such vehicle, a pair of thrust engines on such vehicle adapted to be energized by such radiated energy and mounted in symmetrical relationship on opposite sides of a central axis of such vehicle normally coinciding with the center line of such beam, said pair of engines being mounted in a substantially common plane including such central axis and at angles with respect to such axis to provide respective thrusts to such vehicle having components both normal to and parallel with such axis, and means symmetrically disposed about such axis for directing energy from such beam to said engines in respective amounts apportioned in accordance with the relative distances between each of said engines and the center line of such energy beam, whereby said engines are differentially energized to exert corrective thrusts upon displacement of such axis from the center line of such beam.

9. A system for maintaining a thrust-propelled space vehicle in spaced-apart relationship with a mother planet at a selectable location with respect thereto, comprising a control station on such mother planet for transmitting microwave electromagnetic energy in a substantially vertically-directed beam toward such vehicle, a plurality of thrust engines on such vehicle adapted to be energized by such transmitted energy and mounted in symmetrical relationship with respect to a central axis of such vehicle normally coinciding with the center line of such beam, the thrust produced by each of said engines being substantially confined to a respective plane including such central axis and having components both parallel with and directed away from such central axis, and means symmetrically disposed about such axis for directing energy from such beam to said engines in respective amounts apportioned in accordance with the relative distances between each of said engines and the center line of such energy beam, whereby said engines are differentially energized to exert corrective thrusts upon displacement of such axis from the center line of such beam.

10. A space vehicle, comprising body means, antenna means on said body means for intercepting incident microwave energy, means on said body means for converting said intercepted microwave energy directly into a flight-producing force applied to said body means, and means for directing microwave energy intercepted by said antenna means to said converting means.

11. A system for remotely energizing a space vehicle, comprising means at a control station on a mother planet for generating microwave energy, means at such control station to radiate such microwave energy to such space vehicle, antenna means on said vehicle for intercepting such radiated microwave energy, means on such vehicle for converting said intercepted microwave energy directly into a flight-producing force applied to such vehicle, and

means for directing microwave energy intercepted by said antenna means to said converting means.

References Cited in the file of this patent

UNITED STATES PATENTS

889,790	Kitsee	June 2, 1908
930,508	Vreeland	Aug. 10, 1909
2,471,744	Hershberger	May 31, 1949
2,769,601	Hagopian	Nov. 6, 1956
2,813,242	Crump	Nov. 12, 1959
2,941,764	Lee et al.	June 21, 1960
2,949,550	Brown	Aug. 16, 1960

OTHER REFERENCES

"Vistas in Astronautics," published by Pergamon Press in 1958; page 192.

Advanced Propulsion Systems, published 1959 by Pergamon Press, pp. 81–88.

United States Patent [19]
Winn et al.

[11] **4,025,913**
[45] **May 24, 1977**

[54] ELECTRICAL FIELD SENSING AND TRANSMITTING APPARATUS

[75] Inventors: **William Paul Winn; Leon Geddis Byerley, III,** both of Socorro, N. Mex.

[73] Assignee: **The United States of America as represented by the Secretary of the Navy,** Washington, D.C.

[22] Filed: **Dec. 22, 1975**

[21] Appl. No.: **642,926**

[52] **U.S. Cl.** .... **340/189 M;** 244/32; 324/72; 325/113; 343/706

[51] **Int. Cl.**^{2} **G08C 17/00;** H04B 1/034

[58] **Field of Search** .... 340/184 M, 177 R; 325/111, 113, 115; 343/706; 244/31, 32; 324/72; 317/262 E, 262 AE

[56] **References Cited**

UNITED STATES PATENTS

674,427	5/1901	Palencsar	343/706
911,260	2/1909	Pennock	343/706
2,409,155	10/1946	Schellens et al.	325/113
3,273,066	9/1966	Ruhnke	340/189 M
3,458,805	7/1969	Kasemir	324/72
3,614,031	10/1971	Demboski	244/32
3,925,726	12/1975	Few	324/72

OTHER PUBLICATIONS

Runnalls, "Bluebird Racer's Telemetry System," *Electronics*, Oct. 28, 1960, pp. 70–72.

Primary Examiner—John W. Caldwell
Assistant Examiner—James J. Groody
Attorney, Agent, or Firm—Richard S. Sciascia; Paul N. Critchlow

[57] ABSTRACT

A pair of conductive spherical shells are carried by a balloon or other means into an electrical field which induces equal and opposite charges on the shells. Rotation of the shells causes the charges to alternate. The alternating sinusoidal charge is applied to an amplifier and the amplified output modulated and applied to a transmitter. The spherical shells become a telemetry antenna for transmitting the sensed electrical field information to a remote receiver.

2 Claims, 4 Drawing Figures

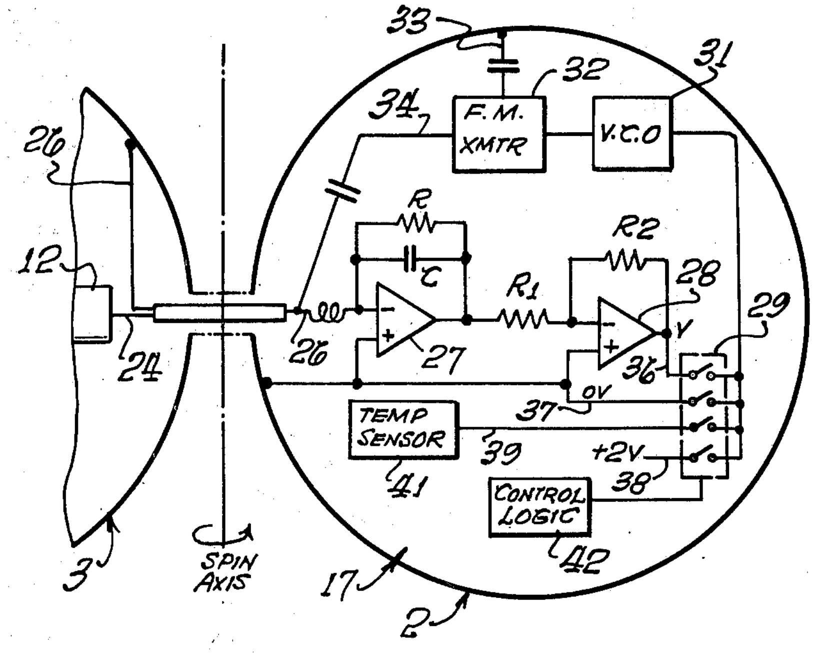

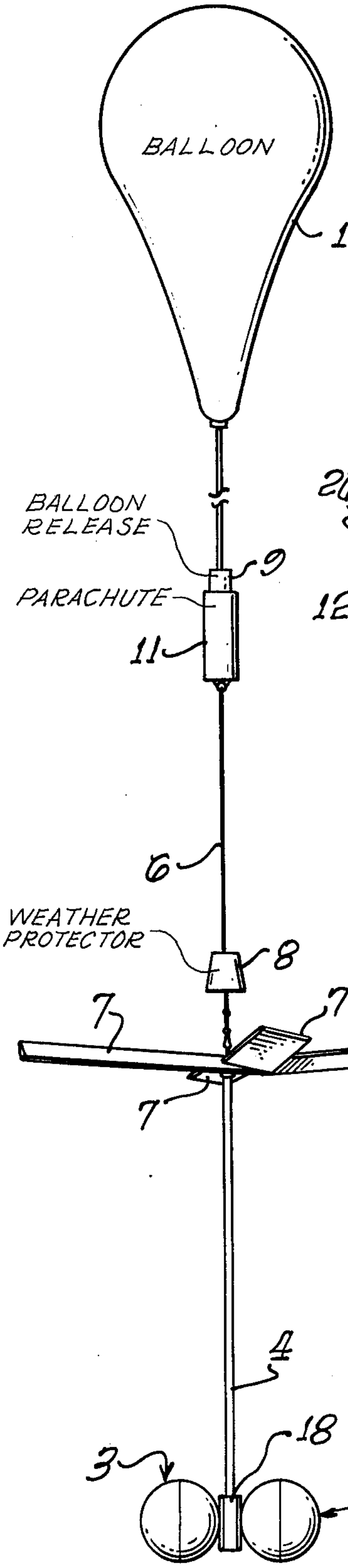
BALLOON
1
BALLOON RELEASE
9
PARACHUTE
11
6
WEATHER PROTECTOR
8
7
7
7
7
4
18
3
2

Fig. 1.

Fig. 3.

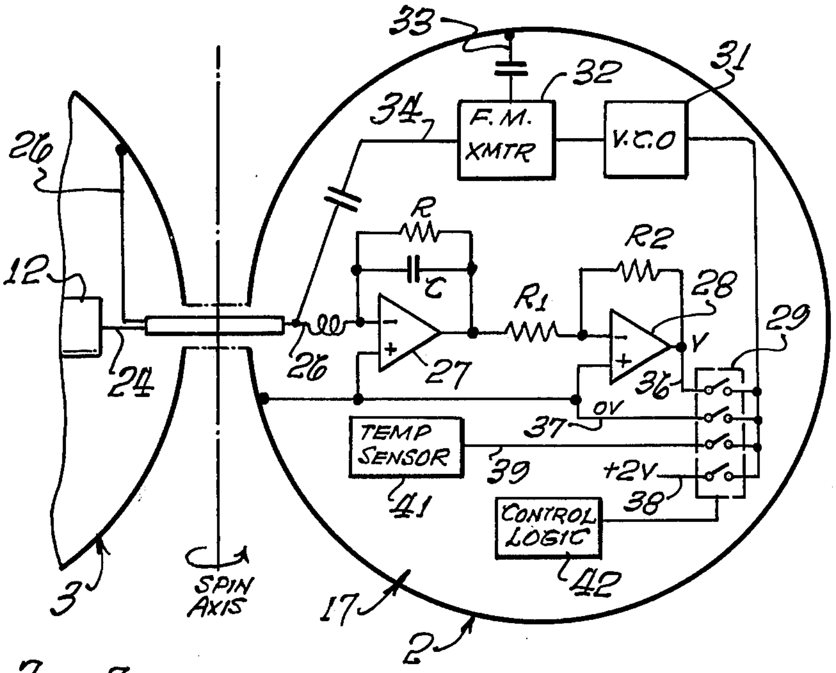
33
32
31
34
F.M. XMTR
V.C.O
26
R
C
R2
28
12
R1
29
V
26
24
27
36
0V
37
TEMP SENSOR
39
+2V
38
41
CONTROL LOGIC
42
3
SPIN AXIS
17
2

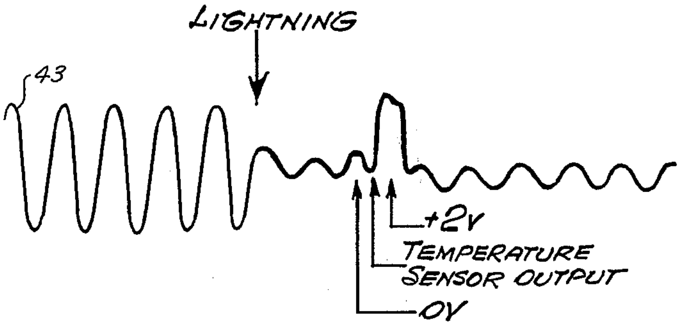

Fig. 4.

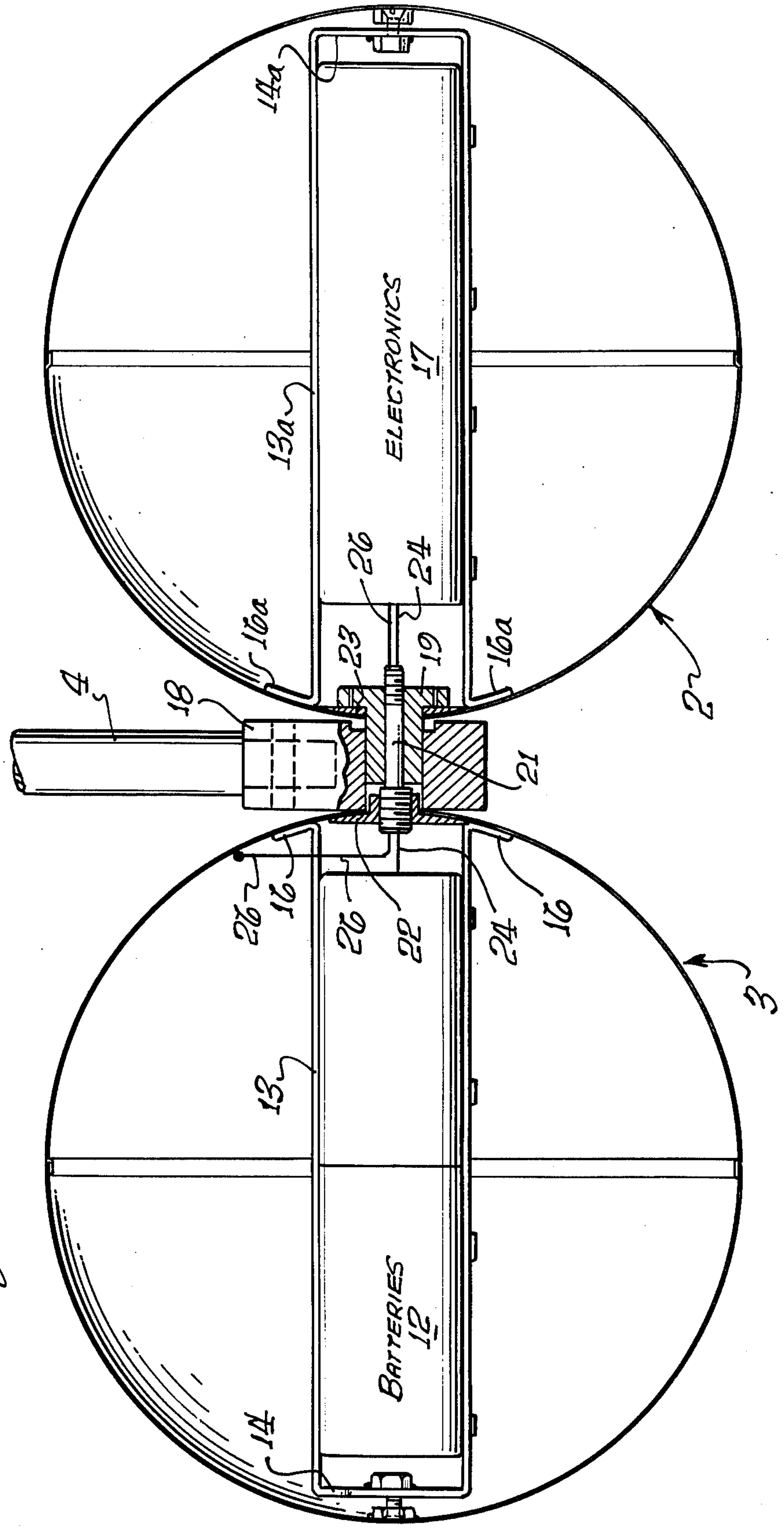
ELECTRONICS
BATTERIES

FIG. 2.

ELECTRICAL FIELD SENSING AND TRANSMITTING APPARATUS

BACKGROUND OF THE INVENTION

The invention relates to apparatus for sensing and transmitting environmental information.

As will become apparent, the present invention, which will be referred to as an electric field meter, has been developed primarily for the purpose of measuring the growth of electrical fields in thunderclouds. For this reason, the ensuing discussion will be particularly with regard to the thundercloud application. Other uses, however, are considered to be within the present inventive scope.

There have been a number of previous efforts directed at obtaining reliable data regarding the growth and behavior of electric fields in thunderclouds. In particular, various investigators and authors have predicted that the behavior of the electric field in a cloud should be different than that at ground. For one reason, the electric field at the ground is affected by point discharge from the ground. Also, it is a fact that the air outside of clouds, especially at high altitudes, has known finite conductivity. Consequently, estimates based upon measurements made outside of a cloud, such as at ground level, are not reliable and their unreliability leaves unresolved several important questions. For example, it is important to determine how the behavior of the electric field in the cloud compares with that at the ground. Also, reliable data is needed regarding the total quantity of the charge in regions affected by lightning as well as such matters as the behavior of the electric field as a function of time inside a cloud at a point fixed relative to the cloud. Obviously, data capable of providing reliable estimates must be obtained from positions directly within the cloud. Also, it is important, particularly in obtaining the so-called 'recovery curves' (electrical field strength vs. time between lightning flashes) to assure to the maximum possible extent that the measurements are obtained from a fixed disposition relative to the cloud. To obtain such data, some previous investigators have utilized slowly moving vehicles, such as balloons or parachutes to carry the sensing instrumentation into cloud formations. For example, the following authors have described results obtained by releasing free balloons with instruments that measured the corona current through a long vertical wire:

Simpson, G. C. and Scrase, F. J., 1937, 'The distribution of electricity in thunderclouds,' *Proc. R. Soc. London A*, 161, pp. 308–352.

Simpson, G. C. and Robinson, G. D., 1940, 'The distribution of electricity in thunderclouds, II,' *Proc. R. Soc. London A*, 177, pp. 281–329

Chapman, S, 1953, 'Thundercloud electrication in relation to rain and snow particles,' *Thunderstorm Electricity*, Chicago Press, pp. 207–230

Yüan, C., Chiang, P., Jan, L., Chou, S., and Chen, T., 1965, 'A stong electrical field radiosonde and analysis of results obtained in thunderclouds,'*Ch'i-hsiang Hsueh-pao* (*Acta Meterologica Sinica*), 35, pp. 440–448.

However, although these authors were able to deduce the sign of the vertical component of the field, their estimates of the magnitude of vertical component were very rough because magnitude is related to the corona current by a non-linear function that varies with altitude. Other recorded data has been obtained by W. H. Evans, 1969, 'Electric fields and conductivity in thunderclouds,' *J. Geophys. Res.*, 74, pp. 939–948, and W. D. Rust and C. B. Moore, 1974, 'Electrical conditions near the bases of thunderclouds over New Mexico,'-*Quart. J. R. Met. Soc.*, 100, pp. 450–468. The data reported by Evans was obtained by employing parachuteborne instruments while the Rust and Moore records were obtained with the use of a captive balloon. J. A. Chalmers in a review appearing in 1967, Atmospheric Electricity, Pergamon Press, also describes a so-called "electronic flux meter" which has proven to be a useful device that in some respects is related to the present meter.

The present invention is particularly concerned with providing an improved electric field meter which is slowly moved vertically up and down through a cloud formation preferably by a balloon and parachute. In particular, it provides a combination field sensor and telemetry antenna by means of which the sensed information can be transmitted to a data-collection center for comparison with other data obtained at ground level.

The objects of the invention are concerned generally with obtaining reliable data relative to the matters which have been discussed. Also, as will become apparent, the invention seeks to obtain this data by the use of expendable apparatus or, in other words, apparatus that is so simple and inexpensive that its recovery costs normally would exceed its fabrication cost.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention is illustrated in the accompanying drawings of which:

FIG. 1 is a somewhat schematic illustration of the present meter suspended from a balloon, and

FIG. 2 is a sectional view of the two-sphere meter showing its electronics and power supply in block form;

FIG. 3 is a view similar to FIG. 2 showing primarily a circuit diagram for achieving the present sensing and transmitting functions, and

FIG. 4 is a short segment of actual recorded data obtained by using the present system.

DETAILED DESCRIPTION OF THE INVENTION

FIG. 1 illustrates one intended use of the present apparatus in which it is suspended in a cloud formation by means of a balloon **1**. The electric field meter portion of the apparatus consists generally of a pair of hollow spherical casings or shells **2** and **3** in which are mounted the sensing and amplifying components. A support rod **4** and flexible line **6** secure the spheres to the balloon and, of particular interest, rotator blades or propellers **7** are secured to and project radially from rod **4**. The attachment fittings for coupling the rod and line should permit free rotation of the rod relative to the balloon. A bell-shaped weather protector **8** aids in keeping moisture, etcetera away from the fittings.

As will be apparent, the balloon is used to carry sensor casings **2** and **3** aloft into a cloud formation to measure its electric field strength. The function of propellers **7** is to rotate the rod and the sensor casings during ascent. The spherical casings, being conductive, receive an induced charge from the electric field and, in a particular manner to be described, this charge is transmitted to a ground station or the like.

At a predetermined or selected level, the balloon is freed by actuating a release device such as an explosive squib **9**. At the same time a parachute **11** is released to permit the meter apparatus to descend at a controlled rate. Field measurements are obtained during descent as well as ascent. The explosive squib and parachute release can be implemented in any desired manner so that it functions either responsively to a ground control or automatically at a timed interval or a predetermined altitude.

Considering the identified components in greater detail, balloon **1**, in practice, is made of 0.02 mm thick polyethylene inflated to a diameter of about 2.5 m so that the total lifting force is between 1.5–2 times the weight of the load. Thus the balloon rises even after accumulating some water and ice. Ascent velocities before entering clouds typically are 3 to 5 m/s. Neoprene meterological balloons initially were tried but they often burst before reaching the cloud base apparently as a result of hail impacts.

Contained parachute **11** is placed about 5 m above spherical casings **2** and **3**. It simply is a 1.2 m square of polyethylene with shroud lines from each corner. Descent velocity is about twice the ascent velocity. As has been indicated, no attempt is made to recover the electric field meters since they are inexpensive compared to the cost of finding and recovering them in the usually rugged terrain in which they are used.

Spherical casings **2** and **3** formed of hollow copper spheres of about 15 cm diameter and, as shown best in FIG. **3**, they are held in a manner to be described, 1.9 cm apart at the bottom end of rod **4**. The spherical shape is an important feature since it minimizes corona discharge which undesirably affects the reliability of the electric field measurements. Corona discharge from rotator blades **7** also is a factor to be considered since, during ascent, the vertical winds tend to carry such space charges past the spheres so as to contribute to the electric field component that is being measured. To minimize this undesired effect, the blades, which are made of balsa wood, are coated with a urethane finish. Other precautions also can be taken although, generally, it has been found that the blade rotation which is high at the blade tips prevents water from accumulating.

The sensing and transmitting functions of the meter are provided by spheres **2** and **3** and by the components contained in their hollow interiors. FIG. **2** shows their structure and support while FIG. **3** illustrates the system electronics. As shown in FIG. **2**, each sphere is formed of lapped hemispheres which facilitate fabrication. Sphere **3** in the implementation contains the power supply for the system which is a battery pack **12** mounted on a bracket support **13** having an outer end **14** securely bolted to the sphere and an open-flanged inner end **16** which is soft-soldered to the sphere. Sphere **2**, correspondingly contains the system electronics of block **17**. The mounting for the electronics is achieved by a bracket **13***a* formed and secured in the same manner as bracket **13**.

For reasons which will become apparent, spheres **2** and **3** should be mutually insulated as well as spaced one from the other a fixed distance. To achieve these purposes, the bottom end of rod **4** is received and pinned in a spacer block **18** formed of a dielectric material, such as teflon. Block **18**, in turn, is formed with a transverse bore receiving a nut **19** also formed of an insulating material. A bolt **21** is threaded into nut **19** to secure the two spheres in their spaced disposition. Inserts **22** and **23** are fitted flush against the inner arcuate walls of each sphere to accommodate the nut and bolt arrangement. Obviously, other arrangements can be substituted. It is, however, important to note that the present bolt **21** is provided with a central bore to receive a conductor **24** which supplies the battery power to the system electronics. It also is to be noted that another conductor **26** conducts the charge induced on sphere **3** to the system electronics for amplification and transmission. Again, the particular paths for the leads is purely a matter of choice although the fact that the charge on the casings are interconnected through the electronics is a significant feature.

The circuitry illustrated in FIG. **3** amplifies and transmits the so-called charge received through lead **26** from spherical casing **3**. Conventional electronic circuitry is used and, of course, it can be modified in any desired manner providing it achieves the desired functions. Included are a pair of amplifiers **27** and **28**, a switch arrangement **29** and a voltage controlled oscillator (VCO) **31** to modulate the switched outputs for application to an FM transmitter **32** which, in turn is electrically coupled to both spheres by conductors **33** and **34**. It also should be noted that the amplifiers are grounded to the copper surface of spherical casing **2**.

A better understanding of the operation of the meter can be obtained by considering the nature of the charges induced on the spherical casings when they are carried into an electrical field by balloon **1**. First, however, it is to be noted that the illustrated arrangement is intended primarily to measure the strength of the horizontal component of the electrical field into which it is carried. Modifications, which will be discussed, can be made if, for example, the vertical, another component, such as is involved.

As to the charges induced on spheres **2** and **3**, there are two important factor to be noted. First, due to the arrangement, these charges will be equal and opposite. Thus, because the spheres are coupled through conductor **26**, there will be a charge transfer between the spheres when they are disposed in the electrical field and the transfer will result in a positive charge on one sphere, such as sphere **2**, and a negative charge on the other. The second important factor is that, as the apparatus is borne aloft by balloon **1**, the vertical wind causes the rotator blades to rotate rod **4** so that the spheres rotate in a horizontal direction on the longitudinal axis of the rod. Although the spin rate can be a matter of choice, the present apparatus achieved a rate of 3 rps.

Rotation of the spheres alternates their relative positons to cause the charge to alternate. This alternation results in a sinusoidally varying input through line **26** to the amplifiers so that this input can be amplified, modulated and transmitted. Further, the fact that the spheres are disposed to receive horizontal field component charges coupled with the fact that the rotational axis of the spheres is normal to this horizontal component, assures output data representative of the amplitude variations of this component. The mathematical symbol used to designate the horizontal component is $\underset{\sim}{E}\perp$ and, as stated, the desired output of illustrated apparatus is $\underset{\sim}{E}\perp$. If there is a pendulum motion during the balloon or parachute flight, the vertical component will be introduced to cause a modulation of the sinusoidal signal. If this component is to be determined, the amount of modulation and the instruments maximum angle of pendulum excursion can be employed. Also, as will be discussed, other geometries can be used and other rotational drives employed to permit investigations of any desired components.

Although the electronics of FIG. 3 is straightforward, there are several factors to be especially noted. As has been described, it is important that the charge on sphere 3 be applied through the amplifiers, et cetera to the other sphere and that the output of transmitter 32 be applied to the spheres. Insulation of the spheres is used to assure that all of the charge is applied to the amplifiers.

Switching arrangement **29** is employed because it is important to provide reference voltages to permit calibration of the entire telemetry and recording components of the system. As would be expected, the transmitted data is recorded at a ground station. FIG. 4 shows a short segment of one of the records. In addition to the reference voltages, it is important in thundercloud investigations to obtain temperature data related to the event being recorded. For these purposes, the system employs a switching arrangement having four inputs. Thus, as shown, switch **29** has an input **36** which is the amplifier output for the signal of interest and inputs **37, 38** and **39.** Inputs **37** and **38** are reference voltages used for calibrations while input **39** is directly from a temperature sensor **41.** Each input has its own switch arm and the switch operation is controlled by a control logic **42** to connect and disconnect the inputs in a desired time sequence. In practice, as shown in FIG. 4, the normal output of the instrument is a sine wave **43.** A lightning event which is marked on FIG. 4 interrupts the sine wave and is recorded. Thereafter, the switches disconnect the electric field signal every 5 seconds and connect, in sequence, the reference voltage OV, the temperature sensor output and the +2V reference. It will be appreciated that other sensors and other references can be added as desired.

In practice, the value of $\underset{\sim}{E}\perp$ can be derived from V_0 by solving the electrostatic problem of two spheres in a uniform external field for the relation between $\underset{\sim}{E}\perp$ and the charges Q and —Q induced on the spheres and by analyzing the circuit for the relation between Q and V_o. These derivations have been made and the result is that:

$$\underset{\sim}{E}\perp = -(8.8\times 10 \text{ inches})(1+(\omega RC)^{-2})\,½\ A^{-1}CV_o$$

(in S1 units)

The constant 8.8×10^{11} volts meter $^{-1}$ Coul^{-1} is the ratio of electric field to induced charge for 15 cm diameter spheres held 1.9 cm apart; ω is the angular spin rate (radians/s) of the spheres about the vertical rod, R and C are the feedback resistance and capacitance of the charge amplifier 27 and A is the gain of the second amplifier. A value for V_0 and thus $\underset{\sim}{E}\perp$ is obtained for each half revolution which corresponds to a time of less than 0.2 sec. and a vertical ascent (or descent) distance relative to the surrounding air of less than 1 meter.

The operation of the apparatus should be clear from the foregoing description. In use, significant data has been recorded in which corona effect and the effect of finite conductivity of the air outside a cloud seem to be negligible. Measurements of the electric field within a cloud just before and after a lightning flash, when combined with measurements at the ground, can be used to obtain a value for the total quantity of charge in a region of the cloud just before it is partially discharged by lightning. Also, an occasion, the shape of the so-called recovery curves have been obtained for comparison with theoretical models. The recovery data is achievable in situations where there are frequent lightning flashes and the electric field appears to be a slowly varying function of position in the vicinity of the balloon.

There are a number of factors which should provide some confidence in the use of the illustrated arrangement. Thus, although the copper spheres and rotating blades both may accumulate a net electrical charge by capturing or releasing charged particles or by going into corona discharge, this accumulated charge does not move relative to the instrument and thus it does not contribute to the sinusoidal output voltage. In general, the collection and breaking away of charged particles, such as precipitation particles, produces such a small step voltage that it is too small to resolve on the records. Noise, such as might be caused by collisions, is not typically present in the records and the few examples where it has appeared probably were due to the somewhat microsphonic nature of the FM transmitter. As another factor, it has been found that the spin rate of the spheres are usually reasonably constant. Variations, of course, can affect the calibrations and they can occur due to accumulations of ice which adds mass to the instrument. However, such variations as have been noted lasted only for a short period of time and were distinquishable.

It is again to be noted that the apparatus shown in the drawings is intended only as an example of a successful arrangement and that other arrangements and applications are well within the purview of the present invention. For example, the illustrated arrangement utilizes two spheres 2 and 3. The same purpose can be served using a single sphere divided into two insulated hemispheres. The two spheres are preferred because they serve as two halves of a biconical antenna and thus have a characteristic impedance that is mostly resistive at the chosen telemetry frequency of 400 MHz. Two insulated hemispheres would have a large capacitative part which would impose difficulties. Even so, the hemisphere arrangement may be better adapted for some purposes and its use is contemplated provided, as indicated, the two halves combine to provide a spherically-shaped pair of mutually-insulated casings or shells. Further, to avoid loss of charge Q, one casing should be electrically coupled through the amplifiers circuitry to the other casing.

It also is well within present contemplation to use the arrangement for purposes other than thundercloud studies. For example, it is important to monitor the electric field charge near storage tanks or other areas containing explosives. In these situations and others, the balloon is not needed. Instead, the spheres can be mounted in a fixed position and rotated positively at any desire spin rate. Further, the disposition of the spheres is not necessarily relative to the horizontal component of the electric field. Any desired disposition can be used and rotation either can be about a horizontal axis or precessed in a controlled manner to obtain information on various components. The basic idea common to all such uses is that of employing the spherical pair to induce the essential potential difference which then is converted into a sinusoidal input for amplification and transmission. It is a significant consideration that this basic arrangement can be implemented in a relatively simple and inexpensive manner to reliably sense the electrical field and simultaneously transmit it using the same spherical arrangement used in the sensing.

It has been indicated that measurements obtained with the present instruments are utilized in conjunction with field changes obtained at the ground. The ground measurements, of course, are not a part of the present invention. However, one manner in which they may be used may be helpful in utilizing the present system. In particular, information or estimates about the net quantity of negative charge present inside thunderclouds can be obtained if both the quantity of charge neutralized and the ratio of charge neutralized to the total amount present is known. Measurements of abrupt field changes at ground accompanying lightning and of the location of the charge that was neutralized together give values of the quantity of charge neutralized. Brook et al., 1962, "Quantitative Study of Strokes and Continuing Currents in Lightning Discharges to Ground", J. Geophys. Res, 67, pp 649–659 discuss this aspect of the measurements. Measurements obtained with the present instruments provide the values necessary to obtain an estimate for the ratio. Such estimates appear to be true on an average although actual fractional variations between the compared values may be more or less than the assumed constant.

Obviously many modifications and variations of the present invention are possible in the light of the above teachings. It is therefore to be understood that within the scope of the appended claims the invention may be practiced otherwise than as specifically described.

We claim:

1. Apparatus for measuring the magnitude of the horizontal component of an electrical field in a thundercloud environment comprising:

electrical field charge collecting and transmitting means including:

a pair of substantially identical spherically shaped hollow electrically-conductive casings disposed in a closely-spaced side-by-side arrangement,

spacer block means for maintaining said side-by-side arrangement, said block means being formed of a dielectric material for structurally insulating said casings one from the other,

a charge amplifier mounted within one of said casings and provided with an input and output,

charge conduction means coupling one of said casings to said input, and

a second charge conducting means electrically coupling said amplifier output to said casings for utilizing the pair as a transmitter antenna; and

means for carrying said collecting and transmitting pair of casings upwardly into and downwardly out of said thundercloud environment, said means including:

a balloon for producing said upward movement,

rigid rod-like suspension means having its upper end carried by said balloon and its lower end fixedly secured to said spacer block means

controllable release means carried by said suspension means for releasing said balloon,

parachute means coupled to said rod-like suspension means below said release means, said means being operable upon balloon release for controlling the downward movement of said casing pair, and

propeller means fixedly secured to said suspension means below said parachute means for rotating said rod-like means and said casing pair during said upward and downward movement,

said pair of casings having a common horizontal axis normally aligned during said upward and downward movement with said horizontal component of the thundercloud field whereby the propeller-induced rotation of said casing pair produces a transmittable sinusoidal amplifier output corresponding in magnitude to said horizontal component of the electrical field, said sinusoidal output being applied to said casing pair for transmission.

2. The apparatus of claim 1 wherein said second charge-conducting means includes means for modulating said transmittable sinusoidal amplifier output.

* * * * *

United States Patent [19]

Yukl

[11] **Patent Number: 4,494,009**

[45] **Date of Patent: Jan. 15, 1985**

[54] METHOD AND APPARATUS FOR CAPTURING AN ELECTRICAL POTENTIAL GENERATED BY A MOVING AIR MASS

[76] Inventor: **Tex Yukl**, Pleasant Valley Rte., Box 21, Baker, Oreg. 97814

[21] Appl. No.: **533,237**

[22] Filed: **Sep. 19, 1983**

[51] **Int. Cl.**³ **F03D 5/04**; H02P 9/04

[52] **U.S. Cl.** **290/55**; 290/43; 290/44; 290/54; 310/308

[58] **Field of Search** .... 290/42, 43, 44, 45, 290/52, 53, 54, 55; 416/132 B; 310/10, 11, 308, 309, 310; 60/202; 322/2 R, 2 A

[56] **References Cited**

U.S. PATENT DOCUMENTS

28,793	6/1860	Uion	310/308 X
414,943	11/1889	Dewey	310/308 X
911,260	2/1909	Pennock	310/308 X
1,014,719	1/1912	Pennock	310/308 X
2,239,693	4/1941	Benett	310/308 X
2,279,586	2/1939	Benett	310/308 X
4,146,800	3/1979	Gregory et al.	290/44
4,206,396	6/1980	Marks	290/44 X
4,302,684	11/1981	Gogins	290/44 X
4,433,248	2/1984	Marks	290/55 X

Primary Examiner—J. V. Truhe
Assistant Examiner—Paul Shik Luen Ip
Attorney, Agent, or Firm—Kolisch, Hartwell, Dickinson & Anderson

[57] **ABSTRACT**

A method and apparatus for making available an electrical current, which comprises placing a conductive airfoil in a moving air mass, sinking a conductive element subsurface, and insulating the airfoil from the conductive element and producing an electrical potential between the airfoil and the element in a moving air mass.

4 Claims, 5 Drawing Figures

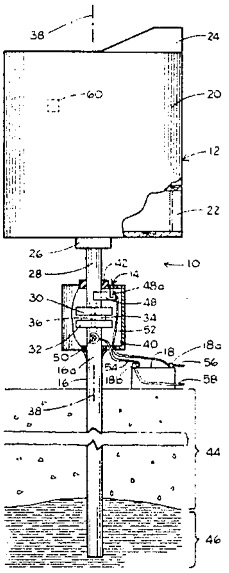

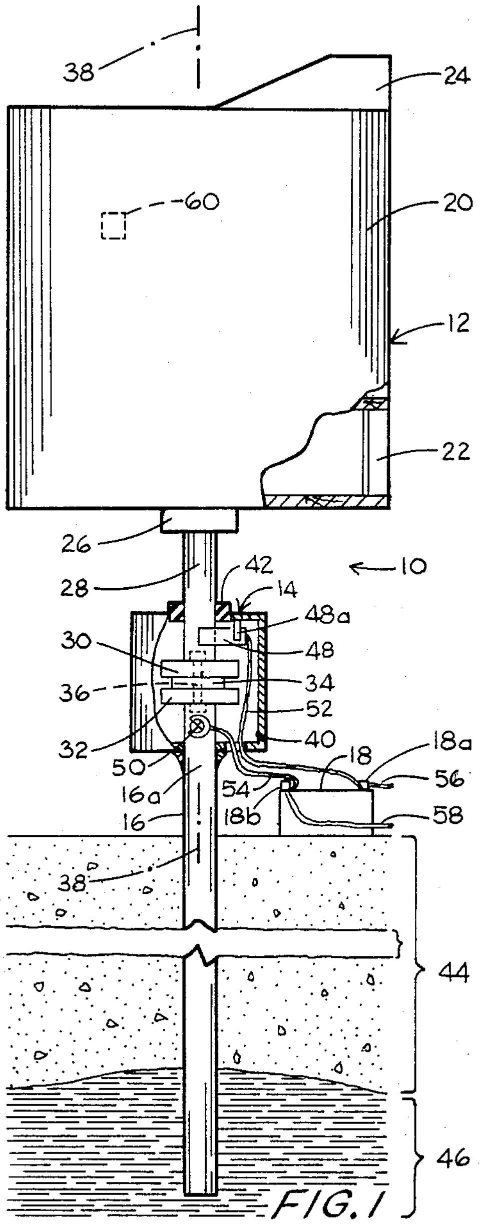
38
24
60
20
12
22
26
28
42
10
14
48a
48
30
34
36
52
32
40
18a
18
50
56
16a
54
18b
58
16
38
44
46

FIG. 1

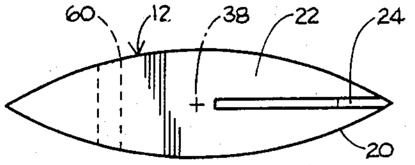
60
12
38
22
24
20

FIG. 2

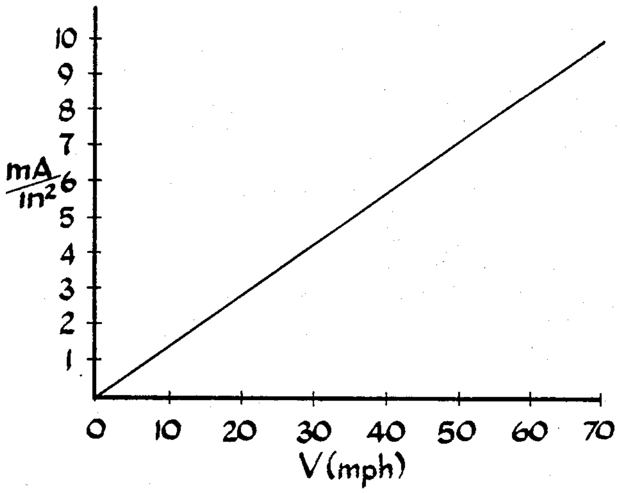

FIG. 5

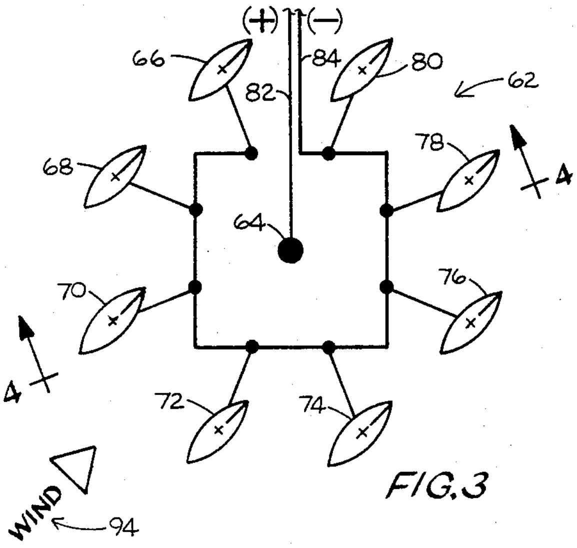
(+)
(−)
66
84
80
82
62
68
78
4
64
70
76
4
72
74
WIND
94

FIG. 3

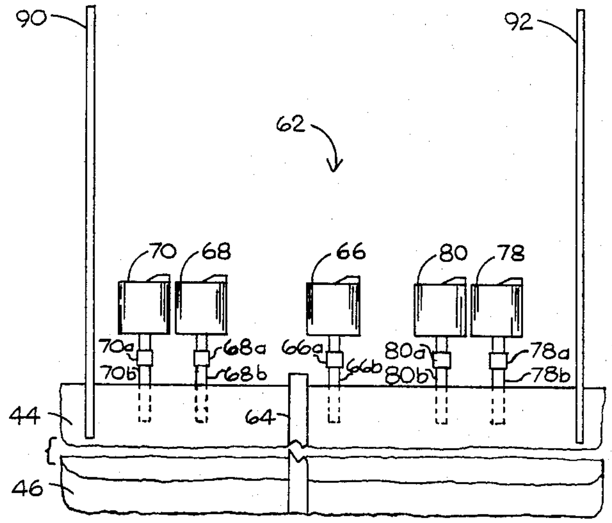
90
92
62
70
68
66
80
78
70a
70b
68a
68b
66a
66b
80a
80b
78a
78b
44
64
46

FIG. 4

METHOD AND APPARATUS FOR CAPTURING AN ELECTRICAL POTENTIAL GENERATED BY A MOVING AIR MASS

BACKGROUND AND SUMMARY OF THE INVENTION

The instant invention relates to a method and apparatus for capturing an electrical potential generated by a moving air mass. Specifically, the instant invention describes a method and apparatus for capturing an earth-air static electrical potential which is produced by an air mass moving over a stationary airfoil.

Known apparatus and methods of producing electricity from a moving air mass utilize windmills and windmill-driven turbines. Such devices are generally mounted on fairly high towers and require complex propeller systems to effectively drive electrical turbines. The systems also tend to have a rather large affect on the air mass as it passes over the blades of a particular windmill. As with the placement of any artificial barriers in an air mass, the air mass is disturbed, and turbulence is induced into the air mass. As a particular particle of the air mass reaches the next windmill, the induced turbulence results in less driving force on the blades of the secondarily driven windmill.

The present invention does not have a propeller system, nor does it induce significant turbulence into a moving air mass. It is well documented that an object, particularly an airfoil, which is moving relative to an air mass, develops a substantial static electrical charge on its surface relative to the air mass as the result of friction between the airfoil and the air mass. This naturally occurring static charge is particularly strong when the airfoil is a conductive surface, such as a metal-covered aircraft wing. For this reason, virtually all metal-covered aircraft employ static discharge strips along the trailing edges of airfoil surfaces in order to dispel any accumulated static electrical charge.

An object of the instant invention is to collect electrical energy from movement of an air mass over an airfoil.

Another object of the instant invention is to collect electrical energy with a non-complex collecting element.

A further object of the instant invention is to collect electrical energy from movement of an air mass over an airfoil with a substantially non-turbulence producing system.

A further object of the instant invention is to describe a method of collecting electrical energy from the frictional effects of an air mass moving over a conductive airfoil.

The electrical system of the instant invention includes a conductive airfoil, which is rotatably mounted on an insulated bearing, the airfoil being free to align itself with a moving air mass. A second component of the system is a ground rod which extends down to a water table level. Electrical leads are connected to a conductive portion of the airfoil and to the ground rod, and are then connected to a storage battery.

The method of collecting electrical energy from the system involves connecting the negative side of the storage battery to the airfoil and connecting the positive side of the battery to the ground rod. As an air mass moves over the surface of the airfoil, an electrical charge is produced and transmitted to the storage battery. Electrical energy may be drawn from the storage battery as needed.

These and other objects and advantages of the instant invention will become more fully apparent as the detailed description of the invention which now follows is read in conjunction with the drawings.

DESCRIPTION OF THE DRAWINGS

FIG. **1** is a front elevation of an earth-air electrical current friction generator constructed according to the instant invention, with portions broken away to show interior detail.

FIG. **2** is a top plan view of the generator of FIG. **1**.

FIG. **3** is a schematic plan view of a generator field constructed according to the instant invention.

FIG. **4** is a view of the generator field of FIG. **3**, taken generally along line **4**—**4** in that figure.

FIG. **5** is a graph depicting the relationship between current production and wind speed associated with the generator of FIG. **1**.

DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT

Turning now to the drawings, and in particular FIG. **1**, an earth-air electrical current friction generator made according to my invention is depicted generally at **10**. The generator includes a metal airfoil, or conductive airfoil means, shown generally at **12**, which is supported by a rotatable airfoil mount **14** and an electrically conductive electrode **16**. A battery **18** provides storage capability for the collecting apparatus.

Airfoil **12** is vertically mounted above ground level. The airfoil is a zero-lift structure in the preferred embodiment, since both active surfaces are symmetrical. Airfoil **12** is formed of a conductive metal skin **20**, formed of lightweight aluminum sheeting in the preferred embodiment, over a frame **22** which is shown behind the cut-away portion of skin **20**. Frame **22** may be formed of any suitable material, such as wood, plastic or metal. In the preferred embodiment, frame **22** is a substantially square structure.

A wind vane **24** is mounted on the top and at one side of airfoil **12**. Wind vane **24** is also referred to as an air mass alignment means, and provides that airfoil **12** will be windfollowing.

An airfoil support bracket **26** is attached to the base of airfoil **12**. One end of a support post **28** is received within bracket **26**. The other end of post **28** terminates in a non-conductive airfoil bearing race **30** and is received within mount **14**. Bracket **26** and post **28** are electrically conductive and are electrically connected to skin **20**.

Rotatable airfoil mount **14**, or rotatable airfoil mounting means, electrically insulates airfoil **12** from electrode **16**, and is referred to herein as isolation means or isolator means.

Airfoil mount **14** is, itself, supported by an above-ground portion **16*a*** of electrode **16**. Portion **16*a*** terminates in a non-conductive electrode bearing race **32**. A bearing **34** is interposed race **30** and race **32**. The respective airfoil and electrode components are retained together by means of a fastener **36**, which is a nut and bolt combination in the preferred embodiment. Fastener **36** aligns with an airfoil axis of rotation **38**.

Mount **14** is enclosed by a cylindrical enclosure **40** which is secured to portion **16*a*** and which allows post **28** to rotate within a weathertight seal **42**. Enclosure **40** and seal **42** are, of course, electrically non-conductive.

Electrode **16** extends through whatever ground layers exist beneath the surface, which are depicted generally at **44**, and then extends into a permanent water table layer **46**. Extending electrode **16** into a water table provides an electrical surface which is as large as the water table. Although the generator would function if electrode **16** did not extend into a water table, the operation of the generator is greatly enhanced by extending electrode **16** to a water table, for reasons which will be explained later herein.

An electrical commutator **48**, and a grounding screw **50**, make electrical contact with post **28** and portion **16***a*, respectively. Commutator **48** is insulated from and fastened to enclosure **40** by bracket **48***a*. Commutator **48** is connected to a lead line **52** which is attached to battery negative terminal **18***a*, and screw **50** is attached to lead line **54** which is attached to battery positive terminal **18***b*. The commutator, screw and lead line comprise what is referred to herein as means for connecting, or connecting means.

Battery **18**, although represented in FIG. **1** as a solitary unit, may obviously include a plurality of units, connected in series or parallel, depending on the output of generator **10**. A pair of leads **56**, **58**, are shown coming off of battery **18**, and would presumably be attached to an electrical distribution system, possibly through an A.C. inverter. The battery and leads **56**, **58** compise an electrical system.

As an air mass moves over skin **20** of airfoil **12**, there occurs an exchange of ions between the air mass and the surface of the airfoil. This exchange takes place as an air mass moves over any stationary object. In most instances, however, the object is sitting on the surface of the earth, in electrical communication therewith, and a charge does not accumulate on the surface of the object. Since the airfoil of the instant invention is insulated from the earth, a static charge builds up on the surface of the airfoil as an air mass flows over the airfoil.

As the air mass moves over airfoil **12**, a collection of negatively charged ions builds up on the airfoil surface. Were there no electrical connection whatsoever between airfoil **12** and ground surface **44**, and were it not for the fact that eventually a charge built up on airfoil **12** would dissipate into the air mass, a charge would eventually build up on airfoil **12** of sufficient strength to arc, such as thc arc produced between a thunderstorm and the ground in the case of lightning, resulting in the complete discharge of the negative ions accumulated on airfoil **12**.

The electrical charge which accumulates on airfoil **12** is drawn off of the airfoil by means of commutator **48**, which is attached to battery **18**. A corresponding positive charge is drawn off of electrode **16** by means of screw **50**.

The surface area of airfoil **12** is defined by what is referred to herein as a cross-section square inch. This unit is depicted at **60** in FIGS. **1** and **2**. The unit is measured on both sides of the airfoil. The distance between the sides of the airfoil is not a factor. So long as the surface area in electrical contact with electrode **16** equals or exceeds the surface area of airfoil **12**, the generator will produce a current flow of 1-milliamp-per-square-inch of airfoil surface cross section with a wind speed of 7-miles-per-hour.

Because the airfoil collects negative ions, it is sometimes referred to as a negative collector element. Likewise, since electrode **16** acts as a source of a positive charge, it is sometimes referred to as a positive collector element, or conductive earth means.

Turning now to FIGS. **3** and **4**, a friction generator farm **62** is depicted schematically in FIG. **3**, and in a pictorial cross section side view in FIG. **4**. A generator farm, according to the instant invention, would include at least one conductive electrode **64**, placed in the ground down to the level of a permanent water table, which would act as a positive collector element. A plurality of airfoils is depicted by even numbers **66** through **80**. Electrode **64** is connected to a positive side of an electrical system by a conductive line **82**. Airfoils **66** through **80** are connected to a common negative lead **84** which connects to the negative side of an electrical system.

Turning now to FIG. **4**, a side sectional view of generator farm **62** is shown. Electrode **64** is placed in ground, represented by **44** as are the support posts for the airfoils. It should be noted that in this type of arrangement, a single commutator would be required within airfoil mounts **70***a*, **68***a*, **66***a*, **80***a* and **78***a*. An airfoil structure support post, represented at **70***b*, **68***b*, **66***b*, **80***b* and **78***b* may be formed from an electrically insulative material when a single electrode, like that shown at **64**, is utilized, thereby further simplifying the internal portions of the airfoil mount.

In addition to earth-air friction generators, a pair of lightning rod devices **90**, **92**, are shown in FIG. **4**. Rods **90** and **92** also are extended subsurface to provide protection from lightning strikes in the vicinity of generator farm **62**. A generator farm of the type described would be particularly susceptible to damaging lightning strikes were such lightning rods not provided.

Returning for a moment to FIG. **3**, a wind arrow **94** is depicted to show a prevailing wind and direction of air mass movement. It will be noted that airfoils **66** through **80** have aligned themselves with the wind indicated by arrow **94**. It should also be noted that with a generator farm in the layout proposed in FIG. **3**, there would be very little disturbance of wind flow by the airfoils of the generator farm. This is in extreme contrast to the turbulence and generally delaminating effect of conventional windmill generating farms.

It should also be obvious to those skilled in the art that the construction of the generator of the instant invention is many orders of magnitude less complex than the construction required for a conventional windmill-type turbine generator.

Turning now to FIG. **5**, a graph of current output per crosssection square inch mA/in^2 versus wind speed V (mph) is depicted. The above-described generator will produce 1-milliamp of electrical current per cross-section square inch with a 7-mile-per-hour wind. Generator output increases linearly, such that the generator will produce 10-milliamps per cross-section square inch in a 70-mile-per-hour wind. It should also be noted that very few known windmill turbine generators function in a 7- or 70-mile-per-hour wind. The generator of the instant application does not have such a limitation. The generator will produce current in any amount of wind, light or strong. As long as the structure is sufficiently durable to withstand a high wind, the system will continue functioning and producing an electrical current. Most systems are restricted to operations over a set wind speed range, generally in the 10- to 45-mile-per-hour range. Systems which produce electricity in extremely light wind conditions are very costly.

Assume for a moment, that an individual airfoil is constructed with sides which are 10-feet in length, thereby producing a cross section dimension on the order of 100-square feet, and further assume that ten airfoils are arranged similar to generator farm **62**, thereby resulting in a total cross section area of 1,000-square feet. The generator farm so described, when connected to a 12-volt D.C. system, and when excited by a wind of 7-miles-per-hour, will produce 1,728-watts of electrical energy.

Thus, a new method of capturing a usable electric current, and an apparatus for practicing the method described have been disclosed. While a preferred embodiment of the invention has been described, it is appreciated that variations and modifications may be made without departing from the spirit of the invention.

It is claimed and desired to secure by Letters Patent:

1. An earth-air electrical current friction generator which comprises

conductive earth means extending subsurface into a water table,

conductive airfoil means operatively associated with said earth means extending above surface into an air mass responsive to airflow to produce an electrical potential relative to said earth means,

isolation means electrically insulating said airfoil means from said earth means, said isolation means including a rotatable airfoil mounting means which allows said airfoil means to rotate in such an air

mass and which supports said airfoil means above surface,

an electrical system, and

means for connecting said airfoil means and said earth means to said electrical system for conduction thereto of an electrical current.

2. An electrical charge collector which comprises

a positive collector element extending to a water table,

a negative collector element operatively associated with said positive collector element, said negative collector element being wind-following and including wind alignment means for aligning said negative collector element with a prevailing wind,

isolator means for insulating said negative collector element electrically from said positive collector element, said isolator means including a rotatable mounting which rotatably supports said negative collector element, and

connecting means for providing an electrical connection from each of said elements to an electrical system.

3. The collector of claim **2**, wherein said positive collector element includes at least one conductive electrode.

4. The collector of claim **2**, wherein said negative collector element includes at least one conductive airfoil.

* * * * *

United States Patent [19]
Phillips, Sr.

[11] **Patent Number: 4,685,047**
[45] **Date of Patent: Aug. 4, 1987**

[54] APPARATUS FOR CONVERTING RADIO FREQUENCY ENERGY TO DIRECT CURRENT

[76] Inventor: **Raymond P. Phillips, Sr.,** 1200 S. Frances, Terrell, Tex. 75160

[21] Appl. No.: **886,138**

[22] Filed: **Jul. 16, 1986**

[51] **Int. Cl.**^{4} **H02M 7/06**
[52] **U.S. Cl.** **363/126**; 307/151
[58] **Field of Search** 363/125, 126; 307/151

[56] References Cited

U.S. PATENT DOCUMENTS

649,621	11/1901	Tesla	455/39
685,957	5/1902	Tesla	455/619
3,434,678	3/1969	Brown et al.	307/151
3,445,767	5/1969	Beard	324/78
3,678,365	8/1972	Fremiot	307/151
3,723,891	3/1973	Whitely	307/246
3,760,257	9/1973	Fletcher et al.	307/151
3,784,845	1/1974	Haas	307/519
3,792,290	2/1974	Brocker	307/522
3,944,935	3/1976	Plant	307/519
4,079,268	3/1978	Fletcher et al.	307/151
4,079,325	3/1978	Mawhinney et al.	307/261
4,085,377	4/1978	Turski et al.	307/519
4,195,237	3/1980	Meslener	307/261
4,297,557	10/1981	Tyler et al.	219/10.55 B
4,360,741	11/1982	Fitzsimmons et al.	307/151
4,360,865	11/1982	Yasumura et al.	363/126
4,393,441	8/1983	Enge	363/126

Primary Examiner—Patrick R. Salce
Assistant Examiner—Jeffrey Sterrett
Attorney, Agent, or Firm—Millen & White

[57] ABSTRACT

Apparatus for and methods of converting radio frequency energy into direct current for generating electric power includes a dipolar antenna for receiving radio frequency energy and a circuit connected thereto for converting the radio frequency energy to direct current. The circuit has a positive output line connected to one pole of the antenna and a negative output line connected to the other pole of the antenna. A positive transmitting diode is in the positive output line and a negative transmitting diode is in the negative output line. First and second bus lines and a pair of tuned circuits of opposite polarity couple the positive output line and negative line to the bus line with one of the bus lines being connected to ground. Each tuned circuit includes a first bridging line connecting the positive output line to the first and second ground lines and a second bridging line connecting the negative output line to the first and second ground lines. Each bridging line has a diode therein oriented at a polarity which is reverse with respect to the input diode. The bridging lines of each tuned circuit are connected to one another by an inductance and have capacitors disposed between the diode and the bus lines. A direct current device is connected to the positive line of the circuit.

11 Claims, 1 Drawing Figure

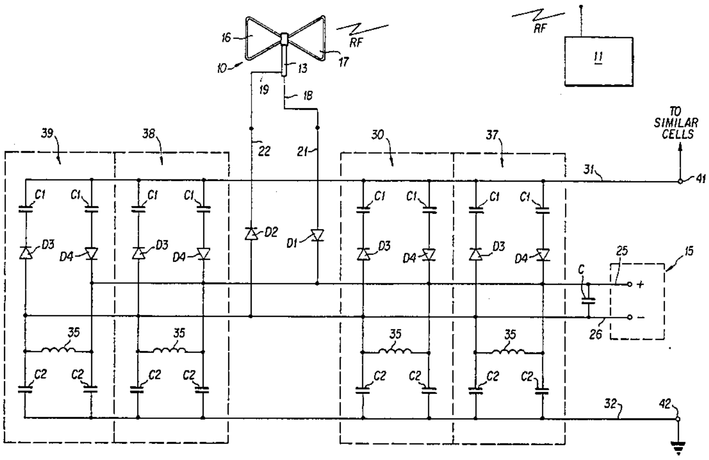

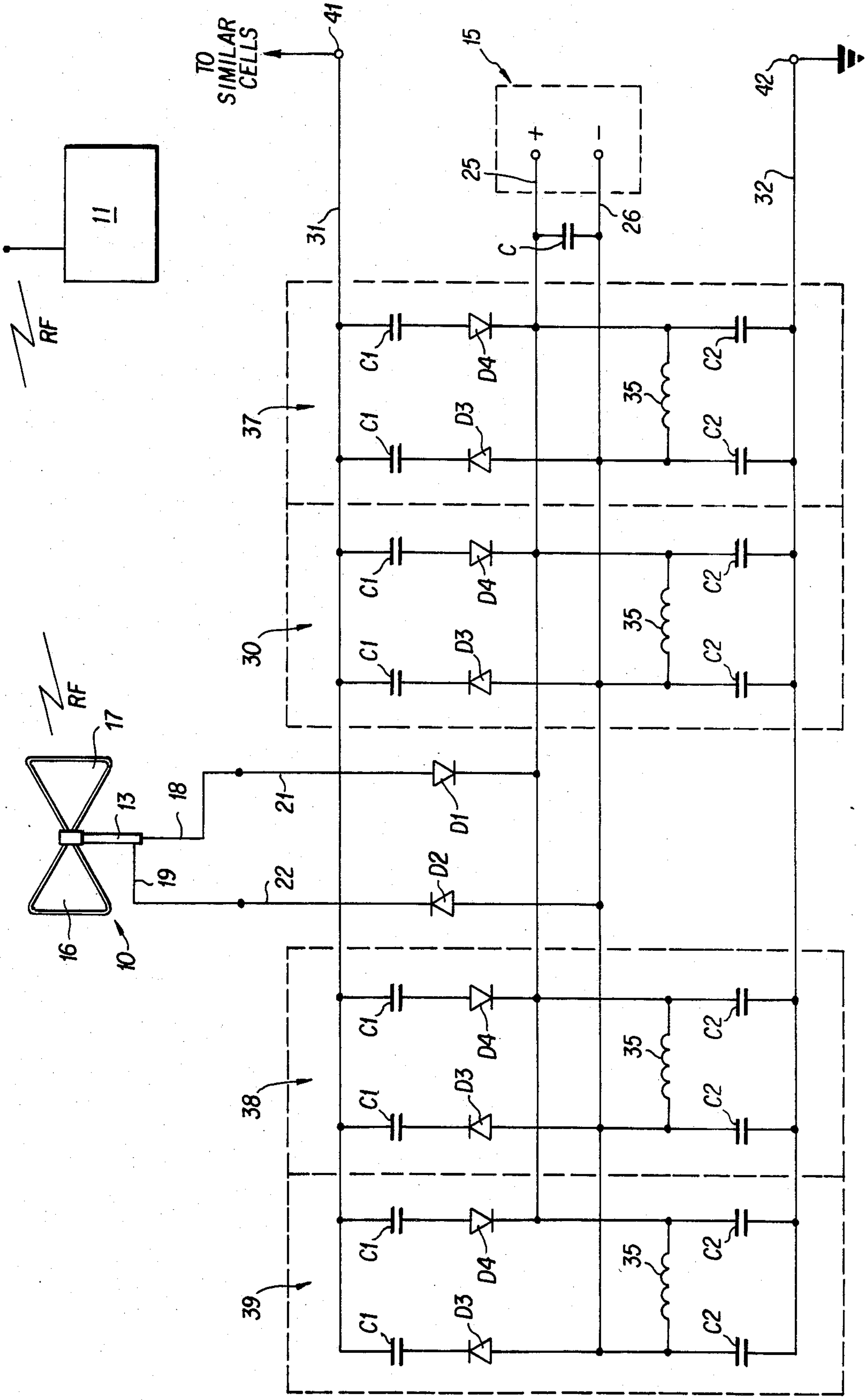
TO SIMILAR CELLS
+
−
C
RF
RF
C1
C1
D4
D3
C2
C2
D1
D2

APPARATUS FOR CONVERTING RADIO FREQUENCY ENERGY TO DIRECT CURRENT

BACKGROUND OF THE INVENTION

1. Field of the Invention

The instant invention relates to methods of an apparatus for converting radio frequency energy to direct current; more particularly, the instant invention relates to methods and apparatus for converting radio frequency energy to direct current wherein the direct current is of sufficient magnitude to power devices such as battery charges and electric motors without the use of amplification.

2. Technical Considerations in Prior Art

There has long been interest in technology directed to transmitting electrical energy over a distance without utilizing conductors, such as wire conductors. Development of such a technology has enormous potential. This was first recognized by Nikola Tesla who in 1899 constructed a 200 foot Tesla coil rated at 300 kilowatts and 150 kilocycles. Tesla hoped to set up standing waves of electrical energy around the whole surface of the earth, so that receiving antennas set at optimum points could tap the power when needed. Tesla was able to light hundreds of lamps at a distance of about 40 kilometers with his device without utilizing electrical conductors. The scheme has generally remained a scientific curiosity but has provided the initial groundwork for current developments wherein attempts are being made to transmit power by microwaves. However, power transmitted by microwaves is envisioned in the form of a beam of very high intensity which is focused from a microwave generator to a receiving antenna. This technology is envisioned as being used for many types of purposes such as transmitting microwave energy collected from gigantic solar power satellites and "star wars"-type weapons systems. However, the focused microwave beam is not suitable for many applications in that the beam must be directed toward a receiving antenna and cannot be transmitted through most objects, including living objects, without destroying the objects.

The instant invention relies on converting energy from standing waves which are emitted from radio frequency antennas in the RF range rather than the microwave range. Of particular interest are very low frequencies which are not used in communications and are available for transmitting power. Also of interest with respect to the instant invention are the low frequency waves emitted by the earth due to pulsation thereof caused by its magnetic field. These low frequency standing "earth" waves can be picked up by receivers tuned thereto.

SUMMARY OF THE INVENTION

It is an object of the invention to provide new and improved methods of an apparatus for converting radio frequency currents to direct current for practical uses other than communications, wherein the direct current energy converted from the radio frequency input energy does not require amplification.

Upon further study of the specification and appended claims, further objects and advantages of this invention will become apparent to those skilled in the art.

The instant invention contemplates an RF antenna for receiving radio waves. The RF antenna connected to a circuit configured in accordance with the principles of the instant invention to convert the RF signals to direct current. The radio frequency signals received by the antenna are transmitted to first and second leads, with one lead being rectified to transmit positive voltage and the other lead being rectified to transmit negative voltage. The positive voltage lead being connected directly to a positive output line and the negative voltage lead being connected directly to a negative output line. The positive output line is connected to a pair of bus lines through a first pair of capacitors, while the negative output line is connected to the pair of bus lines by a second pair of capacitors. Disposed between the first bus line and the positive output line is a reverse diode of negative polarity, while disposed between the negative output line and first bus line is a reverse diode of positive polarity. The positive and negative output lines are connected to one another through an inductance which is in parallel with the capacitors of the first and second pair connected between the second bus line and the positive and negative output lines.

In accordance with one embodiment of the invention the afore-described circuit is duplicated for each positive and negative output line. In accordance with another embodiment of the invention, the afore-described circuitry is coupled to additional circuits identically configured in order to increase the direct current output of the arrangement.

In accordance with a further configuration of the invention, the antenna utilized is a dipolar antenna of aluminum wire arranged in a "butterfly" configuration.

The instant invention further contemplates the method of utilizing the afore-described elements so as to generate direct current having sufficient power to perform tasks such as charging batteries, lighting lamps and powering direct current electric motors without the use of amplifiers.

BRIEF DESCRIPTION OF THE DRAWING

Various other objects, features and attendant advantages of the present invention will be more fully appreciated as the same becomes better understood when considered in connection with the accompanying drawing, in which like reference characters designate the same or similar parts throughout the several views, and wherein:

There is shown a diagram of a circuit in accordance with the instant invention in combination with a driven device and a dipolar antenna which receives radio frequency waves which are converted to DC current for powering the driven device.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring now to the drawing, there is shown a dipolar antenna, designated generally by the numeral **10**, which receives radio frequency waves from an RF transmitter **11**. The radio frequency waves are transmitted to a radio frequency to direct current converting circuit, designated generally by the numeral **12**, through a coaxial cable **13** and DC current from the circuit **12** is used to power an output device, designated generally by the numeral **15**, which may for example by a battery charger DC motor or lighting device. The circuit **12** has no other power inputs other than the radio frequency energy transmitted thereto by the antenna **10** and therefore includes no amplifiers for amplifying the RF energy.

The source of radio frequencies convertible to direct current by the circuit shown may include sources of high frequency, low frequency (LF), very low frequency (VLF) and extremely low frequency (ELF) radio waves as well as seismic vibration of the earth's magnetic fields.

Preferably, the dipolar antenna **10** is formed of two triangular loops of aluminum wire **16** and **17**, one of which is connected to the annular conductor **18** of the coaxial cable **13** and the other of which is connected to the center conductor **19** of the coaxial cable. The size of the bipolar antenna **10** is dependent on the particular application to which it is put. In accordance with one embodiment of the invention, the antenna **10** is approximately 12 inches in width and 18 inches in length. Such an antenna is used to receive five watt energy, such as that generated by a walkie-talkie or citizen-band radio.

The annular conductor **18** of the coaxial cable **13** is connected to a positive lead **21** of the circuit **12**, while the center conductor **19** of the coaxial cable is connected to a negative lead **22** of the circuit. A positive transmitting diode D1 is disposed between the lead **21** and the remainder of the circuit **12** while a negative transmitting diode D2 is disposed between the lead **21** to a positive output line **25** while the negative diode D2 is connected to a negative output line **26**. Accordingly, the positive voltages with respect to ground are produced on output line **25** and negative voltages with respect to ground are produced on output line **26**.

In order to provide a DC output of sufficient power, a plurality of inductance-capacitance, RF, tuned circuits **30**, each forming a positive cell, or a negative cell, are utilized for connecting the positive output line **25** and negative output line **26** to first and second bus lines **31** and **32**, respectively. Bus line **32** is connected to ground while bus line **32** can be connected to circuits similar to circuit **12**. The positive output line **25** is connected by a first bridging line **33** to the first and second bus lines **31** and **32** while the negative output line **26** is connected by a second bridging line **34** to the first and second bus lines. The bridging line **33** has capacitors C1 and C2 disposed between the positive output line **25** and the first and second bus lines **31** and **32**, while the bridging line **34** also has capacitors C1 and C2 dispensed between the negative output line **26** and the first and second bus lines **31** and **32**. Connected between the bridging lines **33** and **34**, is an inductor **35** which serves as an RF choke, while disposed between the positive output line **25** and the capacitor C1 there is a negative polarity diode D4 referred to herein as a bridging diode and disposed between the negative output line **26** and capacitor C1 in line **34** there is a positive polarity diode D3 referred to herein as a bringing diode. As is seen of the drawing, the RF tuned circuit cell **30** is repeated a plurality of times. In the specific example shown, the circuit **12** has separate cells **30**, **37**, **38** and **39**. The cells **30** and **38** are of opposite polarity and balance one another while the cells **37** and **39** of opposite polarity and also balance one another. In order for the system to function, a pair of opposite polarized cells must be utilized. The particular number of cells **30** and the value of the components thereof are determined by the configuration of the dipolar antenna **10** and the power and frequency of the RF transmitter **11**.

The radio frequency to direct current conversion circuit **12** may itself be connected to a duplicate circuit via pins **41** so as to provide additional direct current output on lines similar to positive output line **25** and negative output line **26** the output lines may be connected together in order to boost the total output of the system.

An operative embodiment of the invention utilizes the following elements:

Diodes D1, D2, D3 and D4—Germanium Diodes, Archer **1** N34A,Catalog #1123.
Inductor 35-47 Milli henry R. F. Choke
Capacitors C1 and C2-0.47 Pico Farads at 200 volts
Coaxial Cable 13—50 ohms
Dipolar Antenna 10—aluminum wire triangular loops approximately 12 inches by 18 inches.

From the foregoing description, one skilled in the art can easily ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

What is claimed is:

1. Apparatus for converting radio frequency energy into direct current for generating electric power, the apparatus comprising:

a dipolar input for receiving radio frequency energy;

a positive output line connected to one pole of the dipolar input and a negative output line connected to the other pole of the dipolar input;

a positive transmitting input diode means in the positive output line and a negative transmitting input diode means in the negative output line;

first and second bus lines and a pair of tuned circuits of opposite polarity coupling the positive output line and negative output line to the bus lines, one of the bus lines being connected to ground;

each tuned circuit including a first bridging line connecting the positive output line to the first and second bus lines and a second bridging line connecting the negative output line to the first and second bus lines, each bridging line having a bridging diode means therein oriented at a polarity which is reverse with respect to the transmitting input diode means of the respective output line; the bridging lines of each tuned circuit being connected to one another by an inductance and having capacitors disposed between the bridging diode means thereon and the bus lines, and

a direct current device connected to the bridging lines of the tuned circuit.

2. The apparatus of claim **1**, wherein there are a plurality of similarly configured tuned circuits connected between the output lines and the bus lines.

3. The apparatus of claim **2**, wherein there are a plurality of radio frequency-to-direct current conversion circuits connected to one another to provide a direct current power array tuned to a specific radio frequency such as a high frequency source, low frequency source, very low frequency source, extremely low frequency source, or source created by seismic vibrations of the earth's magnetic field.

4. The apparatus of claim **3** wherein the dipolar input is a dipolar antenna.

5. The apparatus of claim **1**, wherein the device connected to the output lines is a direct current motor.

6. The apparatus of claim **1**, wherein the device connected to the output line is an illuminating device.

7. The apparatus of claim **1**, wherein the device connected to the output lines is a battery charger.

8. The apparatus of claim **1**, wherein the device connected to the output lines is a DC-to-AC inverter.

9. The apparatus of claim **1**, wherein there are a plurality of radio frequency to direct current converter circuits connected to one another to form an array of circuits tuned to a particular radio frequency.

10. The apparatus of claim **1** wherein the dipolar input is a dipolar antenna.

11. The apparatus of claim **10**, wherein the dipolar antenna utilizes aluminum wire arranged in pair of triangular loops.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 4,685,047

DATED : August 4, 1987

INVENTOR(S) : Raymond P. Phillips, Sr.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

On the Title Page "11 Claims" should read -- 12 Claims --.

Column 4, Claim 1, Line 21:

Reads: "into direct current for generating electric power, the"

Should Read: --into direct current, the--

Column 4, Claim 5, Line 62:

Reads: "nected to the output lines is a direct current motor."

Should Read: --nected to the output lines is a direct-current motor.--

Column 4, Claim 8, Line 68:

Reads: "nected to the output lines is a DC-to-AC inverter. "

Should Read: --nected to the output lines is a DC-to-AC converter.--

Please add the following new claim:

--12. The apparatus of claim 1 wherein the dipolar input is a coazial cable.--

Signed and Sealed this

Third Day of November, 1987

Attest:

DONALD J. QUIGG

Attesting Officer

Commissioner of Patents and Trademarks

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 4,685,047

Page 1 of 4

DATED : August 4, 1987

INVENTOR(S) : Raymond P. Phillips, Sr.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

On the Title Page, Item [57],
IN THE ABSTRACT:

Line 1: Reads: "Apparatus for and methods of converting radio fre-"
Should Read: --Apparatus for converting radio fre- --

Line 9: "transmitting diode is in the positive output line and a"
--transmitting diode is located in the positive output line and a"

Line 10: "negative transmitting diode is in the negative output"
--negative transmitting diode is located in the negative output--

Line 24: "to the positive line of the circuit"
--to the output lines of the circuit--

COL. 1
Line 9: "the instant invention relates to methods of an appara-"
--the instant invention relates to appara- --

COL. 1
Line 8: "to methods and apparatus for converting radio fre-"
--to apparatus for converting radio fre- --

COL. 1
Line 14: "as battery charges and electric motors without the use"
--as battery chargers and electric motors without the use--

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 4,685,047 Page 2 of 4
DATED : August 4, 1987
INVENTOR(S) : Raymond P. Phillips, Sr.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

COL. 1
Line 29: "tific curiosity but has provided the initial groundwork"
--tific curiosity, but has provided the initial groundwork--

COL. 1
Line 57: "improved methods of an apparatus for converting radio"
--improved apparatus for converting radio--

COL. 2
Line 6: "tive voltage. The positive voltage lead being connected"
--tive voltage. The positive voltage lead is connected--

COL. 2
Line 8: "age lead being connected directly to a negative output"
--age lead is connected directly to a negative output--

COL. 2
Line 58: "ted to a radio frequency to direct current converting"
--ted to a radio frequency-to-direct current converting--

COL. 2
Line 62: "by the numeral 15, which may for example by a battery"
--by the numeral 15, which may for example be a battery--

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 4,685,047

DATED : August 4, 1987

INVENTOR(S) : Raymond P. Phillips, Sr.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

COL. 3
Line 23: "and the remainde of the circuit 12 while a negative"
--and the remainder of the circuit 12, while a negative--

COL. 3
Line 25: "to a positive output line 25 while the negative diode D2"
--a positive output line 25, while the negative transmitting diode D2--

COL. 3
Line 31: "a plurality of inductance-capacitance, RF, tuned cir-"
--a plurality of inductance-capacitance, RF-tuned cir- --

COL. 3
Line 32: "cuits 30, each forming a positive cell, or a negative cell,"
--cuits 30, each forming either a positive cell or a negative cell,--

COL. 3
Line 50: "polarity diode D4 referred to herein as a bridging diode"
--polarity diode D4, referred to herein as a bridging diode--

COL. 3
Line 53: "D3 referred to herein as a bringing diode. As is seen of"
--D3, referred to herein as a bridging diode. As is seen of--

COL. 3
Line 58: "other while the cells 37 and 39 of opposite polarity and"
--other, while the cells 37 and 39 of opposite polarity--

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 4,685,047

Page 4 of 4

DATED : August 4, 1987

INVENTOR(S) : Raymond P. Phillips, Sr.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

COL. 3
Line 67: "via pins 41 so as to provide additional direct current"
--via pin 41 so as to provide additional direct current--

COL. 4
Line 1: "negative output line 26 the output lines may be con-"
--negative output line 26. The output lines may be con- --

Claim 12, line 2: Change "coazial" to --coaxial--.

Signed and Sealed this

Eleventh Day of July, 1989

Attest:

DONALD J. QUIGG

Attesting Officer

Commissioner of Patents and Trademarks

US005942806A

United States Patent [19]

Veliadis

[11] **Patent Number: 5,942,806**

[45] **Date of Patent: Aug. 24, 1999**

[54] **METHOD AND DEVICE FOR GENERATING ELECTRICITY**

[76] Inventor: **Konstantinos D. Veliadis**, 520 Salem St., Lafayette, Ind. 47905

[21] Appl. No.: **08/695,135**

[22] Filed: **Aug. 8, 1996**

[51] **Int. Cl.**^{6} **H02P 9/04**

[52] **U.S. Cl.** .... **290/1 R**; 290/2; 290/46; 290/1 A

[58] **Field of Search** .... 290/1 R, 44, 55, 290/1 P

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,206,396	6/1980	Marks	290/44
4,314,192	2/1982	Caro	322/2 R
4,433,248	2/1984	Marks	290/44
4,489,269	12/1984	Edling et al.	322/2

FOREIGN PATENT DOCUMENTS

0571786 10/1977 U.S.S.R. .

Primary Examiner—Elvin G. Enad
Attorney, Agent, or Firm—Woodard, Emhardt, Naughton Moriarty & McNett Patent and Trademark Attorneys

[57] **ABSTRACT**

A method and device for generating an electric current and power from the earth's electric field. The electric current is generated by providing a thermoemissive material at some distance above the earth's surface, connecting the thermoemissive material to the earth's surface with a conductor (e.g., wire) such that a conductive path is established between the thermoemissive material and the earth's surface through the conductor, and heating the thermoemissive material to cause it to emit electrons into the surrounding air, thereby drawing replacement electrons from the earth's surface through the conductor to the thermoemissive material. One way to produce power in the above setting is by installing a D.C. electric motor or other electrical device along the said conductive pathway.

5 Claims, 1 Drawing Sheet

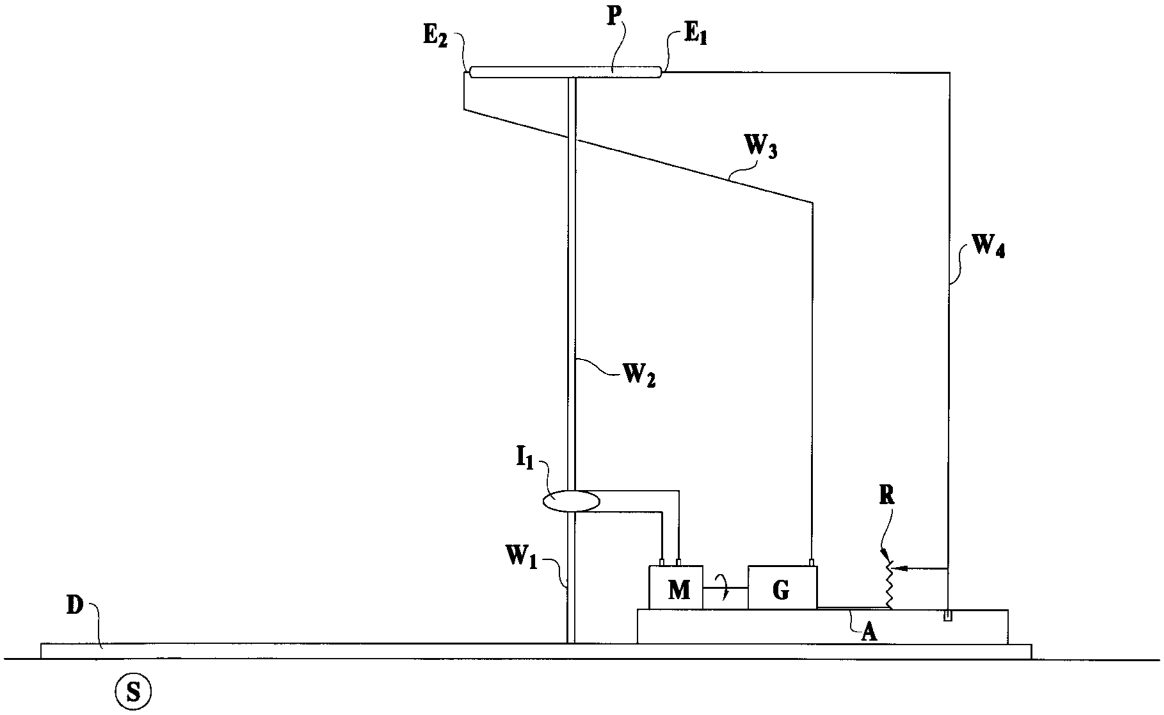

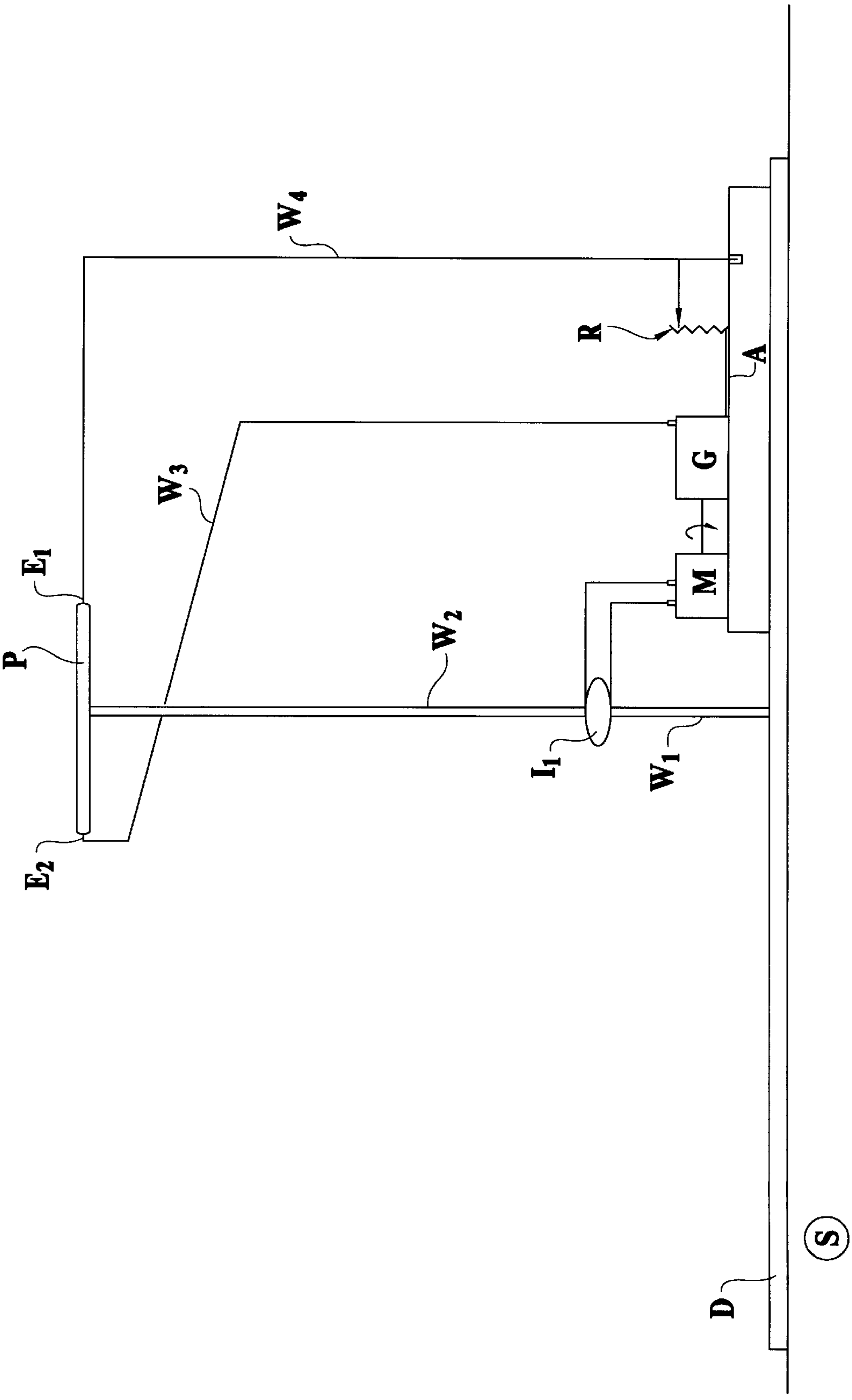
W_4
R
A
G
M
W_3
E_1
P
W_2
E_2
I_1
W_1
D
S

Fig. 1

METHOD AND DEVICE FOR GENERATING ELECTRICITY

FIELD OF THE INVENTION

The present application relates generally to a method and device for producing electricity, and more particularly to a method and device for generating an electric current and power using the earth's electric field.

BACKGROUND OF THE INVENTION

It is known that the earth is under a relatively strong electric field, observed between the surface of the earth and its surrounding air. This field arises from the fact that the air has a net excess of positive ions (i.e. is positively charged), whereas the earth's surface has a net excess of electrons (i e. is negatively charged). By the laws of electrostatics, each of these net charges tends to be uniformly distributed everywhere in the air and on the ground surface, respectively.

In spite of these quasi-uniform distributions, the induced electric field is still relatively strong, approximately 130 V/m measured on the surface of the sea. Although this voltage looks attractive for generating electricity, the problem is that it is accompanied by a relatively small electric-current density: the air's net positive-ion density n_i is very low, so that the resulting net ion-current density $J=n_i \times v_i$ toward the earth's surface is almost zero. Thus, for example, if one connects one bare end of an insulated wire with the earth's surface (negatively charged) and the other bare end with a horizontal metallic surface in the (positively charged) air, one would measure very little current flowing through the wire, proportional to the exposed area of the metallic surface. Accordingly, to date there has been no successful way developed to use the earth's electric field for significant electricity production.

A need therefore exists for a method of generating electricity by taking advantage of the earth's electric field. The present invention addresses that need.

SUMMARY OF THE INVENTION

Briefly describing one aspect of the present invention, there is provided a method for generating an electric current and power using the earth's electric field. The electric current is generated by providing a thermoemissive material at a distance above the earth's surface, connecting the thermoemissive material to the earth's surface with a conductor so that a conductive path is established between the thermoemissive material and the earth's surface through the conductor, and heating the thermoemissive material to cause it to emit electrons into the surrounding air, thereby drawing replacement electrons from the earth's surface through the conductor to the thermoemissive material. Electric power is produced by installing a direct-current electric motor or other electrical device somewhere along the conductive path. One objective of the present invention is to provide a method for generating an electric current and power. Further objectives and advantages of the present invention will be made obvious by the following description of the preferred embodiments.

BRIEF DESCRIPTION OF THE DRAWING

FIG. **1** shows the apparatus of the present invention, according to one preferred embodiment, where:

W—wire or other electrical conductor;
D—conductive disc, in contact with one end of W_1;

M—direct-current (D.C.) electric motor;
G—alternating-current (A.C.) generator, powered by M;
P—thermoemissive material, could be a plaque or plate;
E—electrode, in contact along one whole side of P;
R—roostat, or other suitable voltage-regulating device;
A—conductor, connecting a pole of G with an end of R; and
I_1—insulating material, which takes on the weight of W_2, P, E_1, E_2 and part of W_3 and W_4. That weight is passed on to W_1, and from there onto D. Other weight-supporting schemes are possible.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

For the purpose of promoting an understanding of the principles of the invention, reference will now be made to preferred embodiments and specific language will be used to describe the same. It will nevertheless be understood that no limitation of the scope of the invention is thereby intended, such alterations and further modifications in the illustrated device, and such further applications of the principles of the invention as illustrated therein being contemplated as would normally occur to one skilled in the art to which the invention relates.

As previously indicated, the present invention relates generally to a method for generating an electric current and power by taking advantage of the earth's electric field. The method generally comprises providing a thermoemissive material at some distance above the earth's surface, connecting the thermoemissive material to the earth's surface with a conductor (e.g., wire) and heating the thermoemissive material to cause it to emit electrons into the surrounding air, thereby drawing replacement electrons from the earth's surface through the conductor to the thermoemissive material. A sample apparatus for practicing the present invention is shown in FIG. **1**.

One of the most critical elements of the inventive apparatus is the thermoemissive material. For the purposes of this invention, a thermoemissive material is any material that emits electrons when appropriately heated or irradiated. In some preferred embodiments the thermoemissive material ("thermat") is a solid sheet or layer, while in other embodiments the thermat is in a molten or fluid state.

Preferred thermats include pure metals (such as Mo), and their alloys, and unsintered or sintered ceramics (especially metallic ceramics, "cermets", such as Mo-sintered BaO/SrO). Ni/BaS and Pt/BaS sintered cermets are preferably used in some embodiments.

As shown in FIG. **1**, the thermat may be plated onto a plaque which is suspended (e.g., horizontally) at some distance above the earth's surface. Preferably, the thermat is suspended at a distance of at least a two meters above the earth's surface. As is known to the art, the magnitude of the potential difference increases with increasing height above the earth's surface (except for random effects of significantly reduced local net-ion density in the surrounding air), so the higher the thermat is placed above the earth's surface (up to a naturally-prescribed limit), the greater the total voltage available for producing electricity.

The thermat is generally connected by a conductor to the earth's surface. In one preferred embodiment the conductor is copper wire, although other materials, especially tungsten, may be used. The conductor provides a conductive path for electrons to pass from the earth's surface, through a D.C. electric motor, to the thermat, where they are emitted into the surrounding air. The D.C. electric motor powers a generator, which in turn produces the A.C. electricity. In one embodiment the conductor is made of many thin wires, as further described below.

With the present invention the thermat must first be heated to put it in a state of emitting electrons. Once electrons are being emitted from the thermat, they will be driven away (skywards) by the earth's electric field and simultaneously be scattered by collisions with air molecules. That emission will cause replacement electrons to be drawn through the conductive path to the thermat, forming the electric current that powers the D.C. electric motor. When the thermat has been heated to the point that the desired level of constant flow of electrons is fully established, a steady state operation of the device has been achieved.

Because it is necessary to keep the thermat hot, the conductors attached to it will also heat-up and may thus need adequate cooling to keep their temperature down. To this end, the conductors W can be left partially or totally uninsulated, with the many and thin individual wires of each pulled apart from one another and their bare surface roughened, for maximum ventilation effect. Also, many thin cooling ffins can be vertically welded all over the free upper surface of the disc, made of the same material as the disc and with their outer surface roughened. Lastly, the free upper surface of the disc can be carefully sprayed periodically with regular tap water, to induce on it refrigeration via evaporation of liquid H_2O. Additional cooling measures, such as artificial refrigeration of the conductors W and/or the disc, etc., can also be adopted if necessary.

The basic procedure for practicing the present invention, according to one preferred embodiment, is described below. First, a small piece of land with a generally flat area is selected. The elevation of the land is preferably close to sea level. All bushes and/or trees are removed from inside a 10-meter radius of the geometric center of the flat area of the land. All other trees or bushes in the flat area of the land are trimmed, so that none exceeds in height the elevation of the thermat to be installed.

Moist soft soil is provided over the 10-meter radius of the flat-ground area described above, to a thickness of about 70 centimeters. A copper disc D, approximately 20 meters in diameter and 10 cm in thickness, is laid horizontally on the soft soil. This disc is the base of the device. For best operation of the device, the disc D is connected with one or more thick (uninsulated) copper wires to seawater (if the sea is nearby). In other situations, lakewater, riverwater or even the main pipe of potable tap-water can be used to establish electrical connection with the disc.

One end of a copper wire W_1 is welded to the center of the (upper) surface of the copper disc D. Wire W_1 is positioned perpendicular with respect to the copper disc D, and its other end is steadfasted (fixed) to the bottom of an insulating material I_1.

A second copper wire W_2 has one end steadfasted (fixed) to the top of I_1 and extends upward from I_1. The upper end of W_1 and the lower end of W_2 are respectively connected (via externally insulated conductors) to the two poles of a D.C. electric motor M. The weight of the motor is preferably supported with a pedestal base fastened on the copper disc D and made of a strong, hydrophobic, insulating plastic; the motor's weight is not carried at all then by W_1.

An A.C. generator G is mounted on a similar pedestal base; that base is also fastened on the copper disc D and its height and position are chosen such that the rotating axis of the generator is along the same straight line as that of the motor M. The motor's rotor is then mechanically joined with the generator's rotor. The proper thermal-expansion/contraction margin for the two rotors is allowed in the design of their mechanical joint.

The generator and the motor must not be in contact with any materials except for the externally insulated conductors whose ends are electrically attached to their poles, the insulating pedestals they sit on, the mechanical joint between their rotors, and the surrounding air.

To protect from the rain, the whole system (except for the thermat)may be placed under a waterproof tent, preferably made of thin plastic. The bottom part of the walls of the tent is cut-off and removed (e.g. the bottom 1 meter), to secure adequate ventilation and cooling of the device.

The composite thermoemissive plaque simply consists of a tungsten sheet whose one free surface is plated with a very thin (e.g. 0.1 micron) layer of the thermoemissive material of choice (e.g. 50% BaO+50% SrO). Such plaques have already been made and tested for their thermoemissivity routinely, as mentioned in the pertinent literature. An example of such a fabrication is given in the next two paragraphs. First, a very thin (e.g. 0.1 micron) ceramic sheet is supported horizontally via four thin rods, that are made of insulating material and are placed at the four corners of the mid-rectangle of one side (referred to henceforth as "the bottom") of the sheet. The melting point of the insulating material of the rods must be higher than the melting point of tungsten. For example, cement or a ceramic material may be used.

A tungsten sheet of the same perimeter as the ceramic sheet is then placed on the top-surface of the latter, so that their four edges align exactly. The bottom of the ceramic sheet is then heated until its top (where the tungsten sheet rests) reaches the melting point of tungsten; this temperature is maintained until the tungsten sheet thermally bonds with the top of the ceramic sheet by infusion of liquid tungsten into the microscopic pits, scratches and cracks of the top-surface of the ceramic sheet.

When the above process is over, the thermoemissive plaque P has been formed and is allowed to cool down to room temperature, via natural convection.

The upper (free) end of the wire W_2 is made of tungsten and is welded to the center of the tungsten surface of the plaque, so that the W_1-I_1-W_2 structure supports the plaque and, via the electric motor M, also provides the link for the plaque's electric contact with the ground.

In one embodiment, an electrode is attached along each of two opposite sides of the tungsten's periphery on the thermoemissive plaque P. One such electrode, E_2, is connected by the conductor W_3 to one pole of the generator G, and the other such electrode, E_1, is connected via the conductor W_4 to a very strong electric heat resistance, which is part of a very large roostat R, whose steady arm (conductor) A is in turn connected to the other pole of the generator. The purpose of this auxiliary circuit between the generator's two poles, is to electrically heat the thermat, to cause the emission of electrons from the thermat into the surrounding air. Of course, other means for heating the thermat may be used, such as very hot gas(es), electromagnetic or other radiation, etc. When electrical-resistance heat is used as shown in the preferred embodiment, all electric connections are preferably done with insulated copper wires, and, in the vicinity of the plaque, tungsten wires.

As one example of how the device may be started, an external electric power source that can supply 5000V and 2000 A (10 MW) D.C. to the poles of the electric motor may be used. Such a power source could be one or more diesel engines coupled with the proper generator(s); or a transformer with A.C. input from the city network and 5000V/2000 A D.C. output. A roostat can regulate the percentage of the above external electric power given to the motor.

When the above 100 MW electricity is supplied to the motor, the generators roostat R FIG. **1** is set so that all the electric output of the generator G goes through the whole regulating resistance of FIG. **1**. Gradually, more of the generator's power is slowly provided for heating the thermat, causing the device to start delivering its own electricity on the motor (at 5000V D.C.). Simultaneously, we start lowering the externally-supplied D.C. current to the motor, so that the total D.C. current supplied to the motor is at all times kept constant at 2000 A. This procedure is continued until all 2000 A D.C. delivered to the motor eventually come from electrons flowing from the copper disc D to the thermoemissive plaque P. Steady-state operation of the device has then been achieved.

Other start-up procedures for heating-up the thermat and bringing the device to a desired steady state of operation exist. For example, the thermat may be bombarded with electromagnetic radiation and/or hot gas(es), etc. In one preferred embodiment, hot gases are used to heat the thermat and bring the device to its steady state of operation.

It is to be appreciated that some secondary components of the device have not been illustrated in the Figure, because they would make the schematic too dense in detail and hard to understand. For example, the external power supply used to start-up the operation of the device has not been shown, although persons of skill in the art can easily appreciate its role.

While the invention has been illustrated and described in detail in the drawing and the foregoing description, the same is to be considered as illustrative and not restrictive in character, it being understood that only the preferred embodiment has been shown and described and that all changes and modifications that come within the spirit of the invention are desired to be protected.

What is claimed is:

1. A method for generating an electric current and power, comprising:

(a) providing a thermoemissive material positioned above the earth's surface;

(b) connecting the thermoemissive material to the earth's surface with a conductor, so that a conductive path is established between the thermoemissive material and the earth's surface through the conductor;

(c) providing a D.C. electric motor or other electrical device to harness the electric energy produced; and

(d) heating the thermoemissive material to cause it to emit electrons into the surrounding air, thereby drawing replacement electrons from the earth's surface through the conductor to the thermoemissive material.

2. An apparatus for generating an electric current and power, comprising:

(a) a thermoemissive material positioned above the surface of the earth;

(b) a conductive path between the earth's surface and the thermoemissive material;

(c) a direct-current electric motor or other electrical device, operatively positioned to harness the electric energy produced;

(d) an arrangement for heating the said thermoemissive material, so that electrons are emitted from the latter; and, if an electric motor is chosen in (c) above,

(e) a generator (optional) for generating electricity from the power provided by said electric motor.

3. The apparatus of claim **2** wherein said thermoemissive material is provided as a plaque.

4. The apparatus of claim **2** wherein said thermoemissive material is a sintered or unsintered ceramic material, or any metal or metallic alloy, or any combination thereof, operating at any state of interest.

5. The apparatus of claim **4** wherein said thermoemissive material is selected from the group consisting of BaO, SrO, Mo-sintered BaO/SrO, Ni/BaS, Pt/BaS, Ni, Pt, Sr, Mo, K, Na, Ba, S, Cs, Ca, Li and Fr, or combinations thereof.

* * * * *

US007643312B2

(12) **United States Patent**
Vanderelli et al.

(10) **Patent No.: US 7,643,312 B2**
(45) **Date of Patent: *Jan. 5, 2010**

---

(54) **METHOD AND APPARATUS FOR A WIRELESS POWER SUPPLY**

(75) Inventors: **Timm A. Vanderelli**, Ligonier, PA (US); **John G. Shearer**, Ligonier, PA (US); **John R. Shearer**, Pittsburgh, PA (US)

(73) Assignee: **Powercast Corporation**, Ligonier, PA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **11/357,578**

(22) Filed: **Feb. 17, 2006**

(65) **Prior Publication Data**

US 2006/0164866 A1 Jul. 27, 2006

Related U.S. Application Data

(63) Continuation of application No. 10/966,880, filed on Oct. 15, 2004, now Pat. No. 7,027,311.

(60) Provisional application No. 60/511,860, filed on Oct. 17, 2003.

(51) **Int. Cl.**
H02M 5/00 (2006.01)

(52) **U.S. Cl.** **363/8**

(58) **Field of Classification Search** 363/8, 363/59, 60, 157, 184, 165; 455/270, 272, 455/273, 275, 276.1, 277.1; 307/151
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,471,344	A		9/1984	Williams	
4,685,047	A	*	8/1987	Phillips, Sr.	363/126
4,700,179	A		10/1987	Fancher	
4,703,190	A	*	10/1987	Tamura et al.	307/2
4,857,893	A		8/1989	Carroll	
5,296,866	A		3/1994	Sutton	
5,612,652	A	*	3/1997	Crosby	333/24 R
5,731,762	A		3/1998	Gila et al.	
5,737,177	A	*	4/1998	Mett et al.	361/234
5,889,383	A		3/1999	Teich	
5,898,911	A	*	4/1999	White	455/232.1

(Continued)

FOREIGN PATENT DOCUMENTS

EP	0 168 947 A1	1/1986

(Continued)

OTHER PUBLICATIONS

Soohoo, "Magnetic Thin Film Inductors for Integrated Circuit Applications," IEEE Transactions in Magnetic, vol. MAG-15, No. 6, pp. 1803-1805 (Nov. 1979).

(Continued)

Primary Examiner—Matthew V Nguyen

(57) **ABSTRACT**

An apparatus for a wireless power supply including a mechanism for receiving a range of RF radiation across a collection of frequencies. The apparatus includes a mechanism for converting the RF radiation across the collection of frequencies, preferably at a same time into DC. A method for a wireless power supply including the steps of receiving a range of RF radiation across a collection of frequencies. There is the step of converting the RF radiation across the collection of frequencies, preferably at a same time into DC.

20 Claims, 2 Drawing Sheets

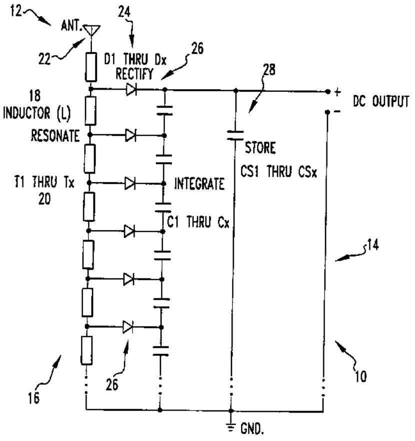

U.S. PATENT DOCUMENTS

5,952,814	A		9/1999	Van Lerberghe	
6,054,925	A		4/2000	Proctor et al.	
6,094,161	A	*	7/2000	Cassen et al.	342/175
6,100,804	A		8/2000	Brady et al.	
6,127,799	A		10/2000	Krishnan	
6,130,603	A		10/2000	Briechle	
6,177,872	B1		1/2001	Kodukula et al.	
6,184,651	B1		2/2001	Fernandez et al.	
6,184,789	B1		2/2001	Richley et al.	
6,284,651	B1		9/2001	Sharan et al.	
6,289,237	B1		9/2001	Mickle et al.	
6,480,699	B1		11/2002	Lovoi	
6,549,064	B2		4/2003	Bandy et al.	
6,633,750	B1	*	10/2003	Dacus et al.	455/126
6,841,981	B2	*	1/2005	Smith et al.	323/312
6,999,327	B2	*	2/2006	Smith et al.	363/60
7,027,311	B2	*	4/2006	Vanderelli et al.	363/8
7,274,921	B2	*	9/2007	Smith et al.	455/313

FOREIGN PATENT DOCUMENTS

JP	3034862	3/1997
JP	10-257697	9/1998
JP	10-295043	11/1998
JP	2000-278887	10/2000

OTHER PUBLICATIONS

Saleh and Qureshi, "Permalloy ThinFilm Inductors," Electronic Letters, vol. 6, No, 26, pp. 850-852 (Dec. 31, 1970).

Reinhold Rydenberg, "The Receipt of Electric Waves in the Wireless Telegraphy," Annalen de Physik IV, 25, 1908, p. 466.

PCT/US2004/034194 International Preliminary Report on Patentability dated Apr. 18, 2006.

PCT/US2004/034194 International Search Report dated Jun. 17, 2005.

* cited by examiner

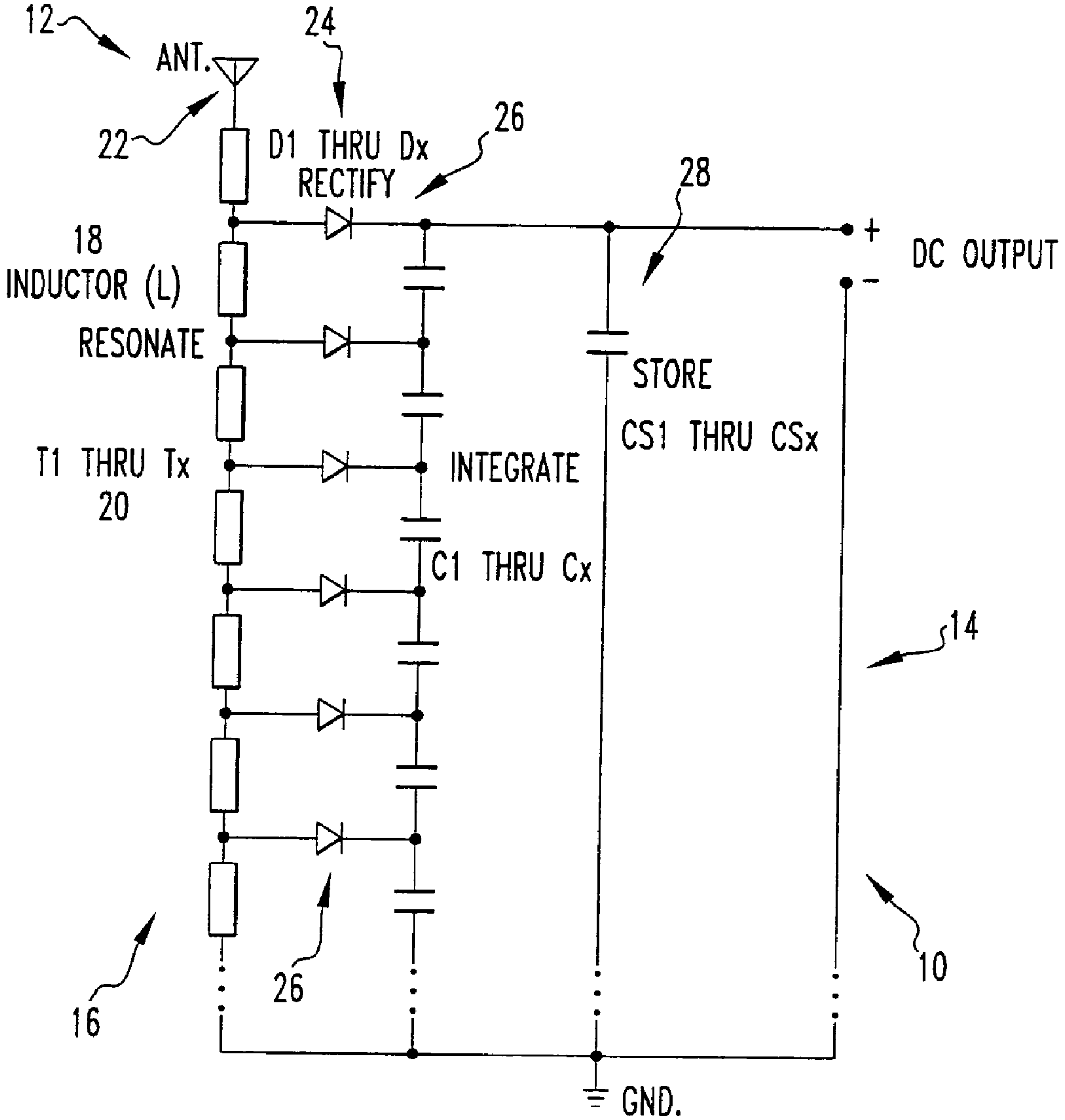
12
ANT.
22
24
D1 THRU Dx
RECTIFY
26
28
+
DC OUTPUT
-
18
INDUCTOR (L)
RESONATE
STORE
CS1 THRU CSx
T1 THRU Tx
20
INTEGRATE
C1 THRU Cx
14
26
16
10
GND.

FIG.1

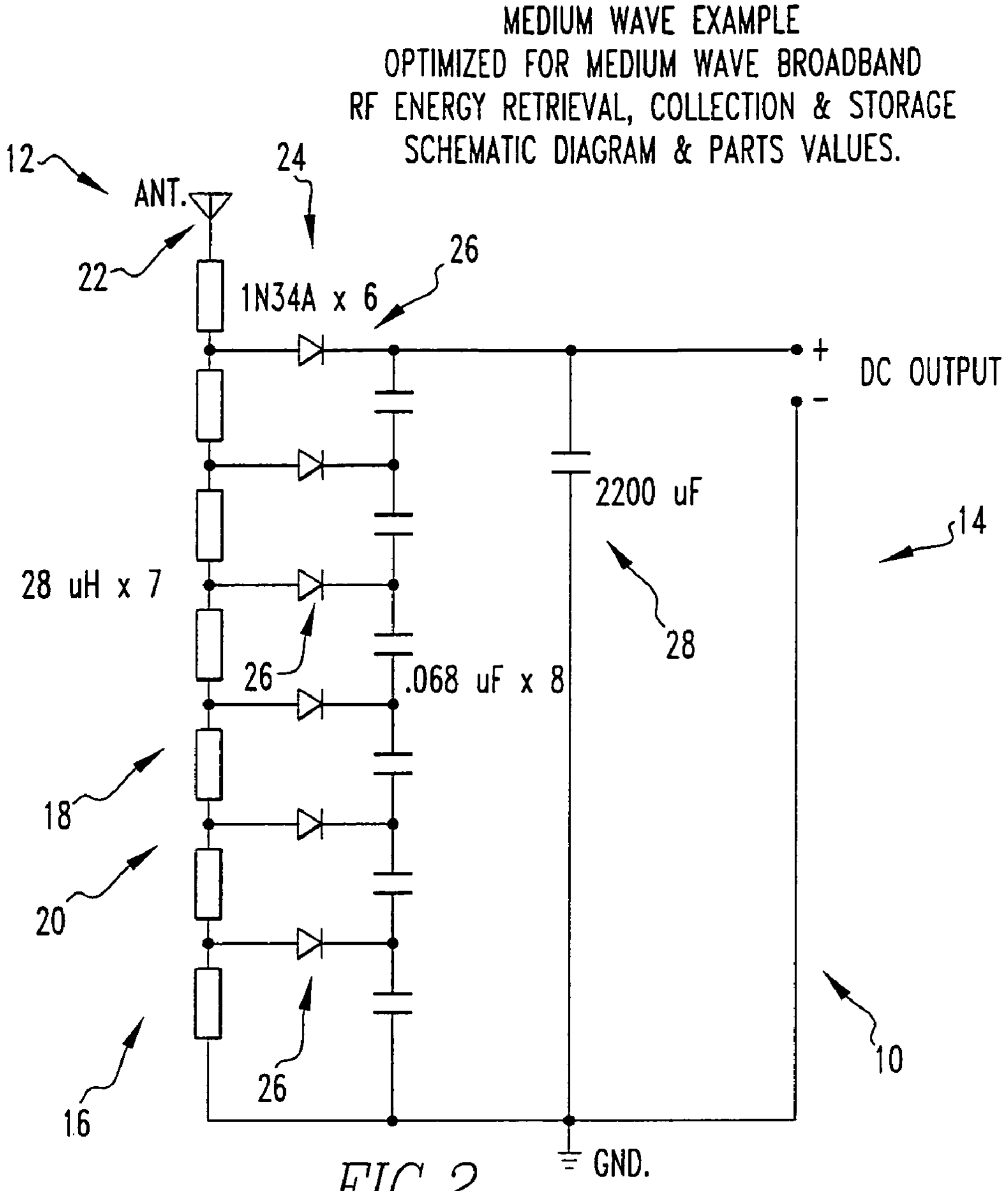
MEDIUM WAVE EXAMPLE
OPTIMIZED FOR MEDIUM WAVE BROADBAND
RF ENERGY RETRIEVAL, COLLECTION & STORAGE
SCHEMATIC DIAGRAM & PARTS VALUES.
12
ANT.
24
22
1N34A x 6
26
+
DC OUTPUT
-
2200 uF
14
28 uH x 7
28
26
.068 uF x 8
18
20
10
16
26
GND.

FIG. 2

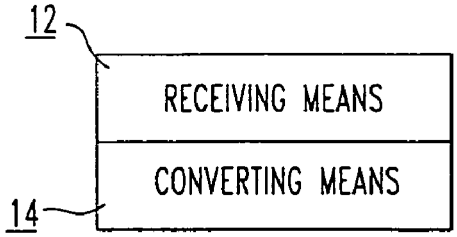
12
RECEIVING MEANS
CONVERTING MEANS
14

FIG. 3

METHOD AND APPARATUS FOR A WIRELESS POWER SUPPLY

This application is a continuation application of application Ser. No. 10/966,880 filed on Oct. 15, 2004, now U.S. Pat. No. 7,027,311.

The nonprovisional application designated above, namely application Ser. No. 10/966,880, filed Oct. 15, 2004, claims the benefit of U.S. Provisional Application No.:

APPLICATION NO. 60/511,860
FILING DATE Oct. 17,2003

FIELD OF THE INVENTION

The present invention is related to the retrieval of radiated electrical energy. More specifically, the present invention is related to the retrieval of radiated electrical energy that is optimized for any given portion of the RF spectrum using a plurality of taps.

BACKGROUND OF THE INVENTION

In the operation of the invention, ambient RF and generated RF signals provide a source of potential energy that can be gathered, stored and supplied to a multitude of devices requiring electrical energy or that can restore energy lost by a discharged source.

Traditional RF receiving devices utilize an antenna to capture a narrow band of frequencies within the RF spectrum, whereby the collection of RF frequencies is then filtered, or tuned, to a specific frequency(s) for the purposes of maximizing the signal being transmitted within the chosen frequency (s). The potential energy contained in the signal is then used for its intended purpose, such as audio, video or data processing. These RF receiving devices have focused on maximizing selectivity of the frequency in order to isolate and to be coherent without interference from other sources.

SUMMARY OF THE INVENTION

The present invention pertains to an apparatus for a wireless power supply. The apparatus comprises means for receiving a range of RF radiation across a collection of frequencies, preferably at a same time. The apparatus comprises means for converting the RF radiation across the collection of frequencies, preferably at a same time into DC.

The present invention pertains to a method for a wireless power supply. The method comprises the steps of receiving a range of RF radiation across a collection of frequencies, preferably at a same time. There is the step of converting the RF radiation across the collection of frequencies, preferably at a same time into DC.

BRIEF DESCRIPTION OF THE DRAWINGS

In the accompanying drawings, the preferred embodiment of the invention and preferred methods of practicing the invention are illustrated in which:

FIG. **1** is a schematic representation of a preferred embodiment of an apparatus of the present invention.

FIG. **2** is a schematic representation of a preferred embodiment of an apparatus of the present invention optimized for medium wave bandwidth RF energy retrieval, collection and storage.

FIG. **3** is a block diagram of the apparatus of the present invention.

DETAILED DESCRIPTION

Referring now to the drawings wherein like reference numerals refer to similar or identical parts throughout the several views, and more specifically to FIG. **1** thereof, there is shown an apparatus **10** for a wireless power supply. The apparatus **10** comprises means **12** for receiving a range of RF radiation across a collection of frequencies, preferably at a same time. The apparatus **10** comprises means **14** for converting the RF radiation across the collection of frequencies, preferably at a same time into DC.

Preferably, the converting means **14** includes an absorbing mechanism **16** which is resonant for a desired band of RF spectrum. The absorbing mechanism **16** preferably includes an inductor **18** which is resonant for the desired band of RF spectrum. Preferably, the converting means **14** includes a plurality of taps **20** placed at points along the inductor **18** to access the RF energy.

The tap points preferably are calculated by matching the inductor's **18** impedance to the desired band of RF spectrum. Preferably, the receiving means **12** includes an antenna **22**. The converting means **14** preferably includes a rectifying mechanism **24** which rectifies the RF energy and converts it into DC voltage. Preferably, the rectifying mechanism **24** includes a plurality of diodes **26** at each tap point which rectifies the RF energy and converts it into DC voltage.

The apparatus **10** preferably includes a storage device **28** for storing the DC voltage. Preferably, the antenna **22** impedance is matched 1:1 with the inductor **18** impedance. The RF spectrum preferably is between 60 Hz to 28 gigahertz.

The present invention pertains to a method for a wireless power supply. The method comprises the steps of receiving a range of RF radiation across a collection of frequencies. There is the step of converting the RF radiation across the collection of frequencies, preferably at a same time into DC.

Preferably, the converting step includes the step of absorbing the energy. The absorbing step preferably includes the step of absorbing the energy with an inductor **18**. Preferably, the converting step includes the step of accessing the absorbing energy with a plurality of taps **20** on the inductor **18**. There is preferably the step of matching the inductor's impedance to a desired RF range.

Preferably, the converting step includes the step of rectifying energy available at each tap and converting it into DC voltages. The rectifying step preferably includes the step of rectifying the energy available at each tap and converting it into DC voltages with diodes **26**. Preferably, the converting step includes the step of summing the DC voltages. The summing step preferably includes the step of adding the DC voltages among a series capacitor integrator. Preferably, there is the step of storing the summed DC voltages. There is preferably the step of using the stored DC voltages.

A method and apparatus **10** for retrieval of radiated electrical energy is described herein. The radiated energy to be captured is being transmitted in the portion of the electromagnetic spectrum sometimes referred to as RF, or Radio Frequency. The primary purpose of the method and apparatus **10** described herein, is to receive RF energy and convert the energy into a usable form of power. The method and apparatus **10** does not discern or interpret individual signals or frequencies. It is designed to absorb and convert signal, carrier and any associated interference for a chosen band or range of frequencies into reusable power.

In contrast, to traditional RF receiving devices, this methodology and apparatus **10** avoids selectivity. It has the unique characteristic of accepting broad ranges of the RF spectrum as a collection of frequencies. Each collected range of frequencies is then rectified, or converted, as a whole into a single voltage. Preferably, at the same time of RF absorption, the resultant voltage is generated. The apparatus **10** makes no attempt to tune for any specific frequency or signal. Each voltage, which is gathered from a given range of frequencies, is then added together and made available to power a device directly, to be stored, or to supply energy to a recharging apparatus. As is implicit from the aforementioned description, preferably, the apparatus **10** receives the RF radiation across the collection of frequencies at the same time.

The radiated electrical energy, to be utilized by the circuit, can be in the form of a wide range of the RF spectrum. Some examples of ambient RF sources can include, but are not limited to: Very Low Frequency—VLF (Maritime/Aeronautical Mobile), Medium Frequency—MF (AM Radio Broadcast), High Frequency—HF (Shortwave Radio Broadcast), Very High Frequency—VHF (TV and FM Radio Broadcast), Ultra High Frequency—UHF (TV, HDTV, PCS, WiFi) and certain Microwave transmissions. In addition, the apparatus **10** allows for the reception of dedicated RF transmission that are generated and broadcast for the specific purpose of transmitting power to the apparatus **10** for absorption, collection and utilization. In this case, it is not necessary for the dedicated RF transmission to contain a specific signal or data that needs to be interpreted for ancillary purposes such as audio/video or data reception and interpretation.

Using the technique described herein, one can design and create an apparatus **10** that is optimized for any given portion of the RF Spectrum. The necessary electrical and magnetic characteristics of the apparatus **10** components will vary depending on the chosen portion of the spectrum. Because of this, it is impractical to create one single apparatus **10** to cover the entire RF spectrum. However, it is possible to create individual apparatus **10**, each designed for a given RF band, and combine both the apparatus **10**, their outputs for maximum power efficiency.

A portion of a selected RF frequency band is intercepted by an antenna **22** placed in the field of emitted energy. The antenna **22** receives energy, in accordance with its design efficiency, and directs it into a system where it is absorbed, rectified, summed and delivered for use or storage.

RF Energy→Antenna→
[Absorbed→Rectified→Integrated→Delivered]
→Used

RF signals striking an antenna **22** are fed into an inductor (L), which is resonant for the desired band of RF spectrum. Note: In areas with a high concentration of RF energy, there is no need to attach an antenna **22**. The absorbed RF energy, consisting of fundamental, harmonic, inter-harmonic and standing waves is accessed via taps **20** (T**1**-Tx) on the inductor **18** which are placed at points along the inductor **18**. A key characteristic of this device is that a capacitor-less front-end allows for the inductors' wide bandwidth and maximum admittance of the incoming RF energy. The tap points are calculated by matching the inductor **18** section's impedance to the desired RF range.

The resultant RF energy, available at each tap point, is rectified by a device, such as diodes **26** (D**1**-Dx), and converted into DC Voltages. The individual rectified voltages are spread among a series capacitor integrator consisting of capacitors (C**1**-Cx). This broadband approach allows maximum energy to be spread among the series capacitor stack.

The sum of the voltages available from C**1**-Cx is stored in any storage device **28** such as a capacitor or group of capacitors Cs (s**1**-sx) and made available for immediate use, or to supply electronic device(s) requiring intermittent power. The electrical characteristics of the storage devices or capacitors, the configuration and actual number of storage devices is dependent on the voltage and power requirements of the device the apparatus **10** is delivering power to. (See Figure One)

Although not considered part of the apparatus **10**, the antenna **22** is an integral component of any practical device utilizing the method and apparatus **10** described. The key characteristics of the antenna **22** would be that it is capable of wide band reception, optimized for the chosen bandwidth, and takes into consideration the necessary effective area to support the power requirements of the target device.

Ideally, the antenna **22** impedance is matched 1:1 with the inductor **18** impedance of the apparatus **10**.

Note: In areas with a high concentration of RF energy, there is no need to attach an antenna **22** to the apparatus **10**.

Inductor **18**:

The characteristics of the inductor **18** is dependent on the chosen bandwidth of frequencies to be collected and utilized. The ideal inductor **18** should be constructed so that the mid point of total inductance would be resonant at the center frequency of the chosen RF segment or spectrum.

Multiple taps **20** provide fundamental and inter-harmonic output voltages from the selected band segments of radio frequency energy.

For example, a medium wave circuit (FIG. **2**), utilizing an antenna **22** impedance of 375 ohms, into an inductive circuit with 375 ohms of reactance, with a center frequency of 1.2 MHz would require an inductance of 100 uH. The effective bandwidth would be approximately 2 MHz wide. (−3 db down at each end of the band).

The inductor **18** can be calculated using the following standard resonance formula (Formula 1):

L=(*d* squared times *n* squared) divided by (18 times *d* plus 40 times *j*)

Where

L=inductance in micro-henrys.
d=conductor diameter in inches.
j=conductor length in inches.
n=number of conductor iterations.

Using similar formulae, the required inductance can be re-calculated for henrys, milli-henrys, pico-henrys and nano-henrys. ie. VLF, LF, MW, HF, VHF, UHF and Microwave frequency band segments.

Utilizing a capacitor-less front-end insures the inductors' wide bandwidth, and maximum admittance to the incoming RF energy.

Taps **20**:

Taps **20** are to be constructed and placed at points along the inductor **18**. Each tap provides an individual output voltage into the rectifying portion of the apparatus **10**.

The number of taps **20** from the inductor **18** can be calculated by the following formula (Formula 2):

Tn=Bw times pi

Where

Tn=Total number of taps.
Bw=Effective Bandwidth of inductor (in Megahertz).
Pi=3.1416

The default position of each tap on the inductor **18** is equidistant along the inductor **18**. Tap positions can also be calculated for optimum output voltage. When calculating the taps **20**, one must take into consideration known frequencies within the chosen band segment that contain higher RF energies, And using a standard resonance inductance formula (1) each individual tap can be calculated for the required frequency and optimum voltage output.

Rectifiers:

The RF energy available at each tap is converted to DC voltage via a rectifying device. The type of rectifying device to be used is dependent on the chosen frequency band, and includes crystal, germanium, silicon and any other types.

Integrator:

A voltage integrator is composed of capacitors C**1**-Cx. The values of these capacitors are dependent on the chosen frequency band, the unique characteristics of the rectifiers and the load imposed by the Storage stage. The reactance of this circuit varies greatly, even during normal operation. However, one can use a standard formula for capacitive reactance as a starting point for preliminary calculations:

$$Xc=1/(2*pi*F*C)$$

Where
Xc=Capacitive reactance in ohms
C=Capacitance in Microfarads.
F=Frequency in Hertz.
Pi=3.1416

Storage:

Storage component(s) are determined by the power requirements of the attached device(s), and the available RF energy absorbed by the inductor (L).

Using a Medium Wave example, a 2,200 micro-farad electrolytic capacitor is used as storage.

Sample Apparatus 10: Medium Wave (AM) Wireless Power Supply

A device has been constructed, using the method stated above, which uses the ambient (existing) AM Broadcast band of the RF spectrum as its source of energy. The device's primary purpose is to optimize the energy absorbed, collected and converted to reusable power.

The size and characteristics of the antenna **22** required for the circuit to operate are not considered a design requirement for the apparatus **10**. The antenna **22** needed to obtain sufficient energy to charge a storage device **28** in a typical urban area with several AM radio stations, would be similar to one used for a standard AM radio. In areas where there is a higher concentration of RF energy, the apparatus **10** itself, without an antenna **22**, is sufficient to develop stored power.

The inductor **18** is in the form of an air coil comprised of enameled #28 gauge wire wound onto a 2" form. The coil is a continuous tightly wound wire with taps **20** placed every twenty turns with a total of six taps **20** available (T**1**-T**6**). The top of the coil is where the antenna **22** is connected. The bottom of the coil is connected to ground.

Germanium diodes (IN34A) (D**1**-D**6**) are connected to each tap on the coil. The series capacitor integrator (C**1**-C**6**) is constructed as illustrated with the C**6** attached to ground. C**1**-C**6** are poly capacitors with a 0.068 uF rating. The power storage device **28** utilized in this sample apparatus **10**, C7, is a 2200 uF electrolytic capacitor.

Very wide band operation can be utilized by coupling multiple instances of the Broadband Wireless Power Supply together.

For Example:

A BWPS circuit designed and constructed (see design considerations) for a Very Low Frequency wave segment (60 Hz center frequency), can be coupled into another BWPS circuit designed and constructed (see design considerations) for an Ultra High Frequency wave segment (5 GHz center frequency). The outputs of each individual circuit connect (via another integrator circuit) into a common storage device **28** (i.e., capacitor) to "pool" collected and converted RF energy together. This technique can be repeated for any or all segments of the energy spectrum.

Although the invention has been described in detail in the foregoing embodiments for the purpose of illustration, it is to be understood that such detail is solely for that purpose and that variations can be made therein by those skilled in the art without departing from the spirit and scope of the invention except as it may be described by the following claims.

What is claimed is:

1. An apparatus, comprising:

means for receiving RF radiation within a plurality of frequency bands; and

means for converting separately the received RF radiation associated with each of the plurality of frequency bands into DC, the converting means includes an inductor and at least three taps, each tap from the at least three taps being placed at a unique point along the inductor to access an energy associated with the RF radiation.

2. An apparatus as described in claim **1** wherein the inductor includes one or more portions, each portion of the inductor is uniquely associated with one frequency band from the plurality of frequency bands, each portion of the inductor has a resonant frequency within that frequency band.

3. An apparatus as described in claim **2** wherein the tap points are calculated based on the resonant frequency for each portion of the inductor.

4. An apparatus as described, in claim **1** wherein the receiving means includes an antenna.

5. An apparatus as described in claim **1** wherein the converting means includes a rectifying mechanism which rectifies the energy associated with the received RE radiation to produce rectified energy and converts the rectified energy into a DC voltage.

6. An apparatus as described in claim **5** wherein the rectifying mechanism includes a diode at each tap point to rectify the energy associated with the received RF radiation into the DC voltage.

7. An apparatus as described in claim **5** further comprising a storage device for storing the DC voltage.

8. A method, comprising:

receiving RF radiation within a plurality of frequency bands at a single antenna;

separating the received RE radiation into a plurality of portions of the received RE radiation independent of a capacitor, each portion from the plurality of portions of the received RE radiation uniquely associated with a frequency band from the plurality of frequency bands; and

converting separately the plurality of portions of the received RE radiation into DC.

9. A method as the described in claim **8** wherein the converting includes absorbing an energy associated with the received RF radiation.

10. A method as described in claim **9** wherein the absorbing includes absorbing the energy with an inductor.

11. A method as described in claim **10** wherein the converting includes accessing the absorbed energy with a plurality of taps on the inductor.

12. A method as described in claim **11** wherein the converting includes rectifying the energy associated with each tap to produce rectified energy and converting the rectified energy associated with each tap into a DC voltage.

13. A method as described in claim **12** wherein the rectifying includes rectifying the energy associated with each tap to produce rectified energy and converting the rectified energy associated with each tap into the DC voltage with a diode.

14. A method as described in claim **13** wherein the converting includes adding the DC voltage associated with each tap from the plurality of taps into a single DC voltage.

15. A method as described in claim **14** wherein the adding includes integrating the DC voltage associated with each tap from the plurality of taps with a series capacitor integrator.

16. A method as described in claim **14** further comprising storing the single DC voltage.

17. An apparatus, comprising:

means for receiving a range of RF radiation across a collection of frequencies; and

means for converting RF radiation across the collection of frequencies into DC, including multiple taps to access the PT radiation across the collection of frequencies.

18. An apparatus, comprising:

means for receiving a range of RF radiation across a collection of frequencies with a single antenna; and

means for converting RF radiation across the collection of frequencies from the single antenna into DC, the converting means includes an inductor that is resonant for a desired band of RF spectrum, the inductor having a plurality of taps placed at points along the inductor to access the RE energy.

19. A method, comprising:

receiving a range of RF radiation across a collection of frequencies at a single antenna;

converting the RF radiation received at the single antenna across the collection of frequencies into DC by absorbing the energy with an inductor; and

accessing the absorbed energy with a plurality of taps on the inductor.

20. A method, comprising:

receiving radiation within an RF spectrum band via a single antenna, a first portion of the received radiation being associated with a first portion of the RF spectrum band, a second portion of the received radiation being associated with a second portion of the RF spectrum band different than the first portion;

converting separately the received radiation from each of the first portion of the RF spectrum band and the second portion of the RF spectrum band into a DC voltage; and

combining the DC voltage associated with the first portion of the RF spectrum band and the DC voltage associated with the second portion of the RE spectrum band.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 7,643,312 B2
APPLICATION NO. : 11/357578
DATED : January 5, 2010
INVENTOR(S) : Timm A. Vanderelli et al.

Page 1 of 1

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 6, line 36 (claim 5, line 3), replace “RE radiation” with -- RF radiation --.

Column 6, line 48 (claim 8, line 4), replace “RE radiation” with -- RF radiation --.

Column 6, line 49 (claim 8, line 5), replace “RE radiation” with -- RF radiation --.

Column 6, line 51 (claim 8, line 7), replace “RE radiation” with -- RF radiation --.

Column 6, line 55 (claim 8, line 11), replace “RE radiation” with -- RF radiation --.

Column 7, line 19 (claim 17, line 6), replace “PT radiation” with -- RF radiation --.

Column 8, line 3 (claim 18, line 9), replace “RE energy” with -- RF energy --.

Column 8, line 24 (claim 20, line 13), replace “RE spectrum band” with -- RF spectrum band --.

Signed and Sealed this
Eighteenth Day of January, 2011

David J. Kappos

David J. Kappos
Director of the United States Patent and Trademark Office

US007855476B2

(12) United States Patent

Ogram

(10) **Patent No.: US 7,855,476 B2**
(45) **Date of Patent: Dec. 21, 2010**

(54) **ATMOSPHERIC ELECTRICAL GENERATOR**

(76) Inventor: **Mark Ellery Ogram**, 780 S. Freeman Rd., Tucson, AZ (US) 85748

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 254 days.

(21) Appl. No.: **12/218,297**

(22) Filed: **Jul. 14, 2008**

(65) **Prior Publication Data**

US 2010/0007218 A1 Jan. 14, 2010

(51) **Int. Cl.**
H02G 11/00 (2006.01)

(52) **U.S. Cl.** **307/145**; 307/149; 361/218; 361/230

(58) **Field of Classification Search** 307/149, 307/145; 244/30, 31; 174/2; 361/212, 215–218, 361/230, 231
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427	A *	5/1901	Palencsar	307/149
911,260	A *	2/1909	Pennock	307/149
1,014,719	A *	1/1912	Pennock	307/149
1,540,998	A *	6/1925	Plauson	310/309
3,121,196	A *	2/1964	Kasemir	324/464
4,234,167	A *	11/1980	Lane	254/291
4,486,669	A *	12/1984	Pugh	290/44
4,842,219	A *	6/1989	Jakubowski et al.	244/31
5,203,542	A	4/1993	Coley et al.	
6,254,034	B1 *	7/2001	Carpenter	244/153 R
6,829,911	B2	12/2004	Davis et al.	
6,961,662	B2	11/2005	Murphy	
7,016,785	B2	3/2006	Makela et al.	
7,200,418	B2	4/2007	Kaikuranta et al.	
7,330,366	B2	2/2008	Lee et al.	
7,375,492	B2	5/2008	Calhoon et al.	
7,378,181	B2	5/2008	Skinlo	
7,388,350	B1	6/2008	Wright	
7,394,671	B2	7/2008	Fukumoto et al.	
7,397,220	B2	7/2008	Uchida et al.	

* cited by examiner

Primary Examiner—Fritz M Fleming
(74) *Attorney, Agent, or Firm*—Mark Ogram

(57) **ABSTRACT**

A mechanism to tap an electrical source which uses an aircraft (preferably a lighter than air balloon) tethered by a conductive line. The conductive line is extended/withdrawn by a winch motor to adjust the altitude of the aircraft. The conductive line is isolated from the ground and an electrical conductor is connected to the conductive line and to an electrical load. In this manner, static electricity generated in the atmosphere is gathered for use.

15 Claims, 4 Drawing Sheets

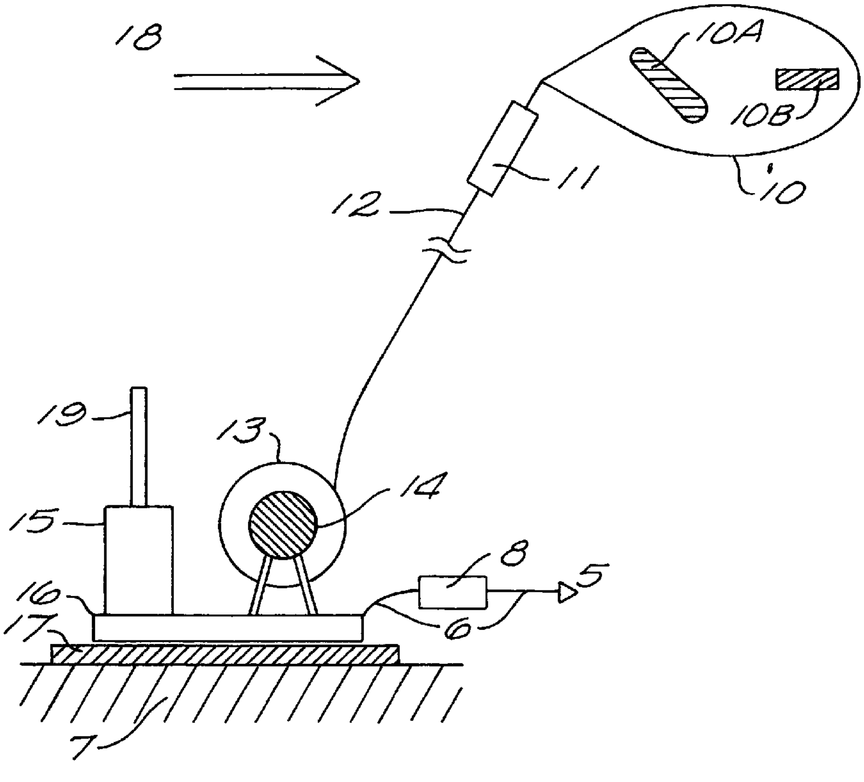

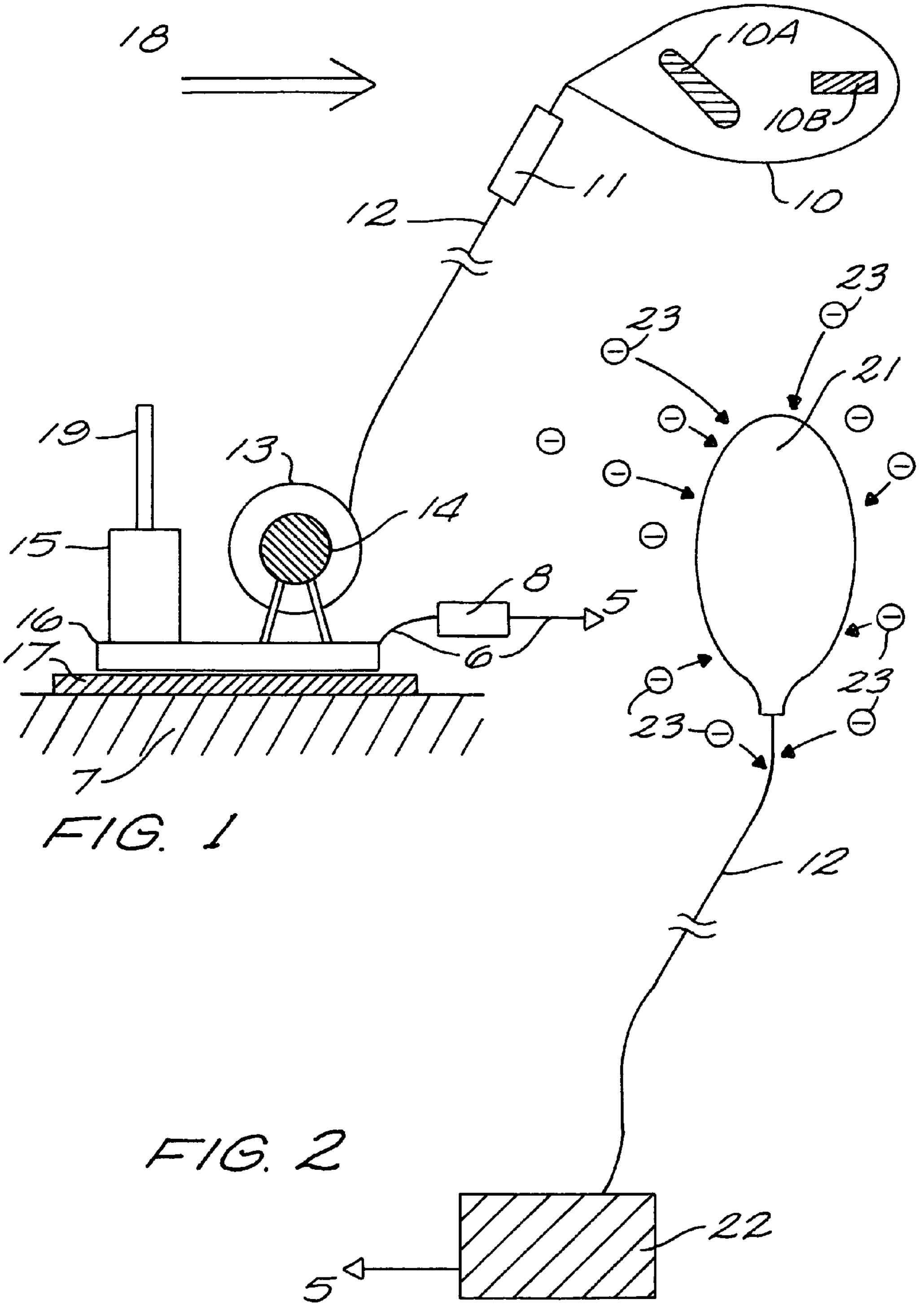
18
10A
10B
10
11
12
19
13
14
15
8
5
16
17
6
7
23
23
21
23
23
12
22
5

FIG. 1

FIG. 2

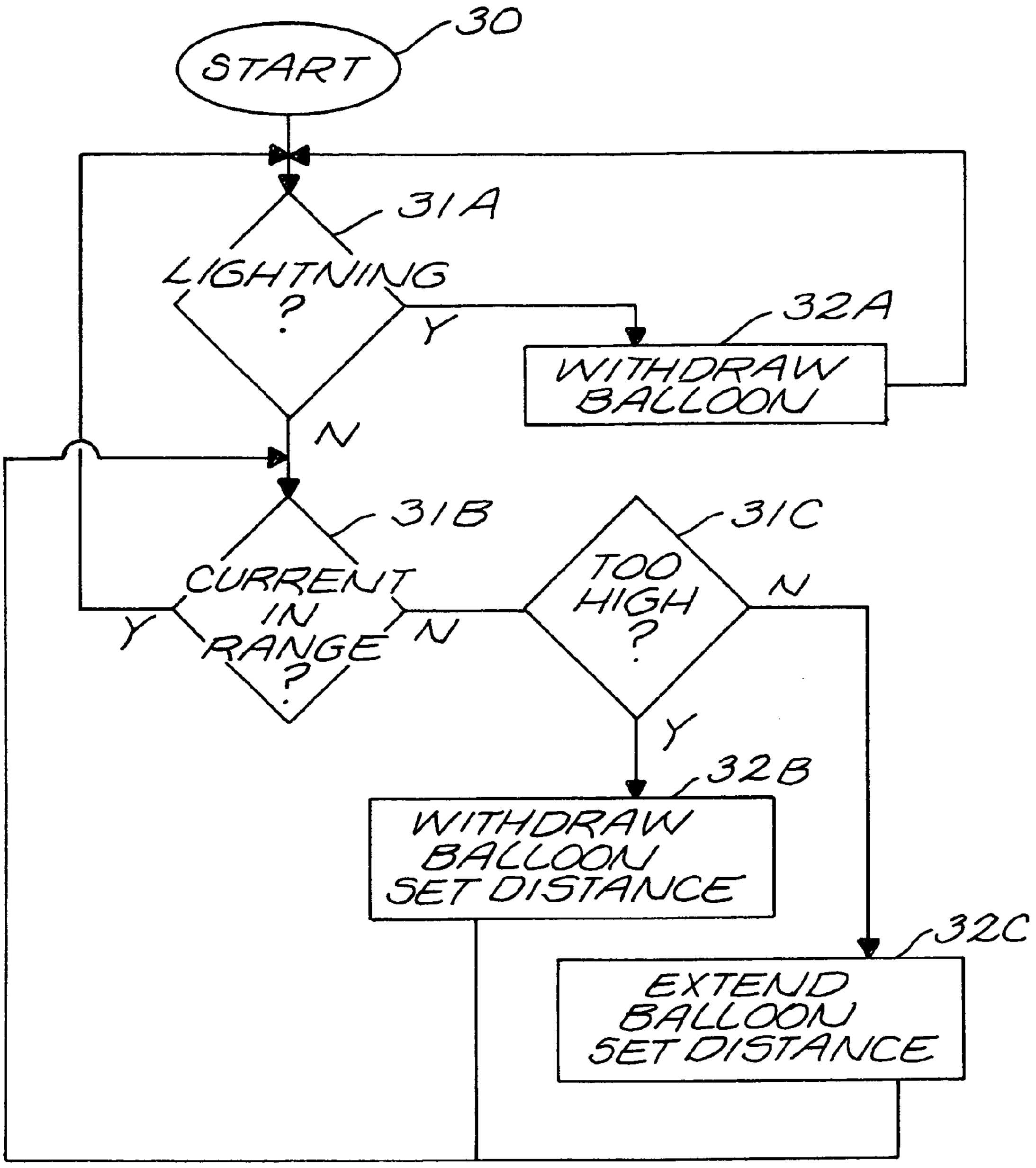
30
START
31A
LIGHTNING ?
Y
32A
WITHDRAW BALLOON
N
31B
CURRENT IN RANGE ?
Y
N
31C
TOO HIGH ?
N
Y
32B
WITHDRAW BALLOON SET DISTANCE
32C
EXTEND BALLOON SET DISTANCE

FIG. 3

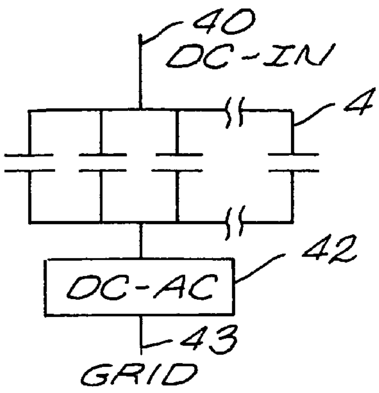
40
DC-IN
41
42
DC-AC
43
GRID

FIG. 4A

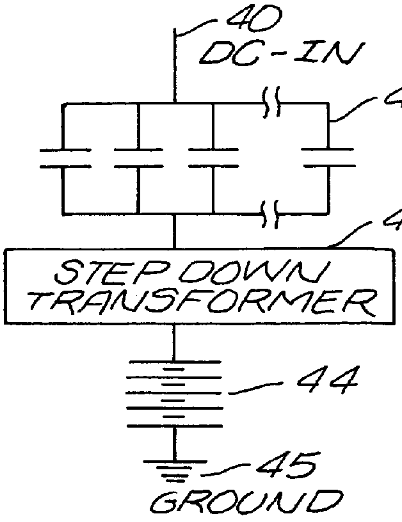
40
DC-IN
41
43
STEP DOWN TRANSFORMER
44
45
GROUND

FIG. 4B

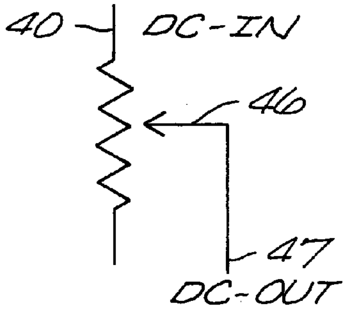
40
DC-IN
46
47
DC-OUT

FIG. 4C

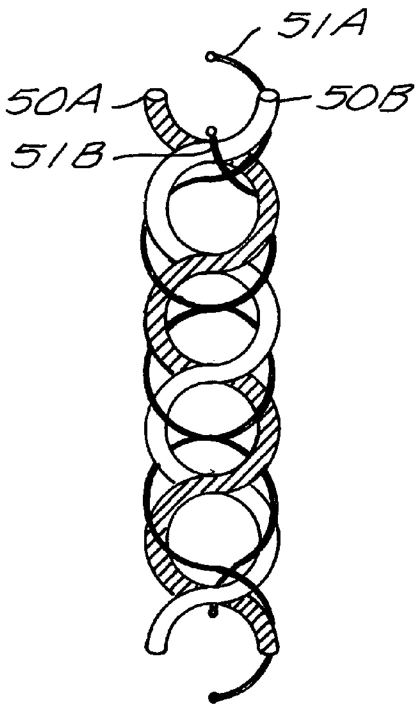
51A
50A
50B
51B

FIG. 5

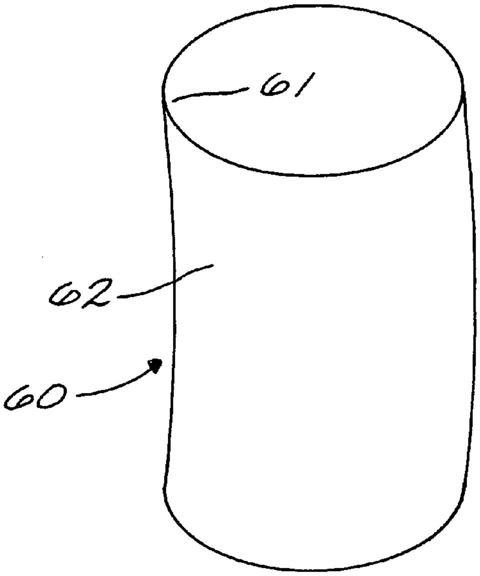

FIG. 6A

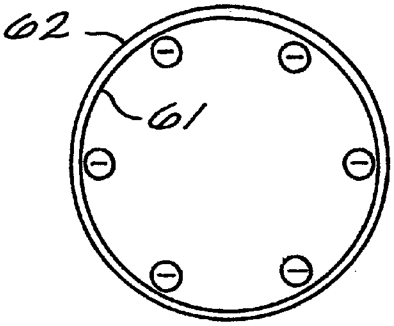

FIG. 6B
REPLACEMENT

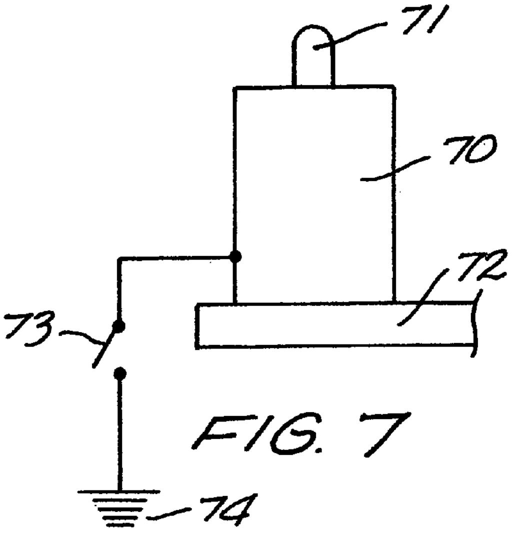
71
70
72
73
74

FIG. 7

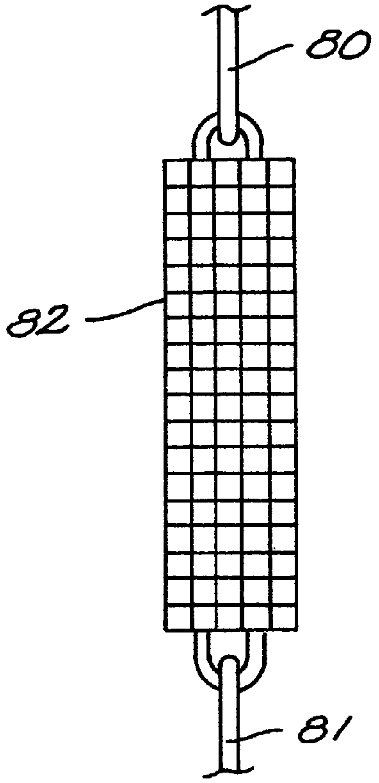

FIG. 8A

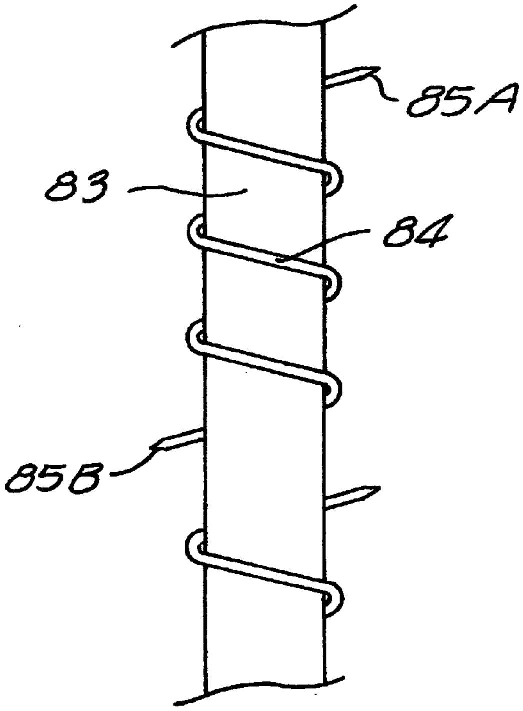
85A
83
84
85B

FIG. 8B

ATMOSPHERIC ELECTRICAL GENERATOR

BACKGROUND OF THE INVENTION

This invention relates to the production of electrical energy and more particularly to the production of electrical energy from the atmosphere.

Everyone is familiar with Benjamin Franklin's kite experiment of 1752. Using a kite whose string had become wet, negative charges from the passing clouds flowed into the string, down to the suspended key, and then into a Leyden jar via a thin metal wire. Franklin was protected by a dry silk string; but, when Franklin's knuckle came too close to the key, he received a strong shock. Fortunately, Benjamin Franklin was not killed, others who tried this same experiment were not so lucky.

Since then, the formation of lightning has remained something of a mystery. Lightning bolts are triggered when a negatively charged cloud base induces a positive charge from the ground, thereby forming a "pathway" for the discharge of the collected electrical energy.

Lightning travels up to 60,000 miles per hour with a flash that is brighter than ten million 100-watt lightbulbs. This wattage is as much power as is produced by all of the electricity plants in the United States and with a voltage of up to 300 million volts.

It is this very fact, the power within lightning is immense, that has prevented any successful collection of the electrical energy from lightning. The electricity in lightning is far too extreme for current technology to harness.

While lightning has attracted a energy starved industrial world, no one has developed any technique to harness this naturally occurring electrical source.

It is clear there is a continuing need for an electrical source other than carbon-based fuels and that the naturally occurring electricity in the atmosphere is being ignored.

SUMMARY OF THE INVENTION

The invention is a mechanism which taps into the naturally occurring static electricity in the atmosphere. Whereas heretofore, the attempt to garner electricity from the atmosphere has focused exclusively on capturing lightning, the present invention syphons off the static electricity which is generated from any agitated air and avoids lightning.

Lightning is only the final discharge of the static electricity, whether that lightning is intra-cloud lightning, cloud-to-ground lightning, or inter-cloud lightning. Other types of final discharges are known as heat lightning, summer lightning, sheet lightning, ribbon lightning, silent lightning, ball lightning, bead lightning, elves, jets, and sprites. Well before these discharges are observed, as the atmosphere becomes agitated by wind or thermal activity, static electricity is being generated.

The present invention recognizes that this static electricity is being formed and creates a mechanism to capture it.

The mechanism of this invention utilizes an aircraft such as a lighter than air balloon. While the preferred embodiment uses a foil balloon, a variety of other aircraft are obvious to those of ordinary skill in the art, including, but not limited to: gliders, rubber balloons (such as weather balloons), biaxially-oriented polyethylene terephthalate polyester film balloons, and latex balloons.

Within this discussion, the balloon is referenced, but, the invention is not intended to be limited solely to balloons.

The balloon is sent aloft and is tethered by a conductive line. In this context, the conductive line may be any obvious to those of ordinary skill in the art. For the preferred embodiment, the conductive line is a generically referred to as a "poly-rope" and is commercially available through a variety of sources. A suitable conductive line is described in U.S. Pat. No. 5,203,542, entitled "Apparatus for Improved Electric Fence Wire Construction for use with Intensive Grazing" issued Apr. 20, 1993, to Coley, et al. and incorporated hereinto by reference.

The conductive line is played out of a winch to control the altitude of the balloon. The motor controlling the winch is able to reverse direction to both extend and withdraw the conductive line which is wrapped around a spool on the winch. The winch/spool combination are part of a base unit.

In some embodiments of the invention, the spool is constructed of rubber so as to insulate the conductive line from the winch assembly. In this embodiment, only the conductive line is charged by the atmospheric static electricity while the winch remains neutral.

In yet another embodiment, the winch/spool are part of a base unit which is itself isolated from the ground by an insulator. In this embodiment, the entire base unit is charged by the atmospheric static electricity.

A conductor, such as an insulated wire, is electrically connected to the conductive line. In one embodiment, where the conductive line is electrically isolated from the spool and winch motor, the conductor is connected to the conductive line. In the embodiment where the conductive line is electrically connected to the base unit, then the conductor is connected anywhere on a metallic base unit.

The other end of the conductor is connected to a load. The load in this case can be any of a variety of electrical loads well known to those of ordinary skill in art, including, but not limited to a motor, a battery system, or the electrical grid for the system.

In the preferred embodiment, a sensor array is used to monitor the activities both at the base unit (such as electrical flow within the conductor) and in the surrounding locale.

A sensor monitoring the electrical flow (i.e. voltage and/or current) within the conductor is used to monitor the electrical activity within the conductor.

In the preferred embodiment, a lightning sensor monitors for lightning activity within the locale. As noted earlier, the electrical characteristic of lightning is so extreme that ideally this discharge is avoided as it might damage the mechanism of this invention.

The sensor array is utilized by a controller, such as microprocessor, programmed to operate the mechanism as outlined herein.

The controller operates the winch motor to extend or withdraw the conductive line and by extension the altitude of the balloon. The controller is programmed to operate the winch by monitoring the electrical characteristics of the conductor and adjusting the balloon's altitude to maintain these characteristics within the conductor within a preset range.

This preset range is established either in the base programming of the controller or is established by an operator of the system.

As example, by controlling the amount of current being withdrawn from the atmosphere, the mechanism operates within a safe range and also provides a relatively stable current flow from which a variety of activities can take place (such as DC-AC conversion).

The controller also utilizes the lightning sensor to protect the mechanism from a lightning strike. Should lightning be detected within a pre-determined range (as established by the software or defined by an operator), then the balloon is pulled down to minimize the risk of damage from a lightning strike.

The invention, together with various embodiments thereof will be more fully explained by the following description of the accompanying drawings.

DRAWINGS IN BRIEF

FIG. **1** diagrams the preferred embodiment of the invention.

FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

DRAWINGS IN DETAIL

FIG. **1** diagrams the preferred embodiment of the invention.

Balloon **10** is an aircraft which, in this illustration, is a lighter than air balloon. Wings **10**A, extending from the body of balloon **10**, provide additional lift in air flow **18**. Tail **10**B helps to stabilize balloon **10**.

Balloon **10** is tethered to the ground via conductive line **12**. As noted earlier, a variety of configurations and materials are available to serve as conductive line **12**. In this illustration, a poly-wire is used Poly-wire is commercially available through a variety of vendors, including, but not limited to: Jeffers Livestock and Zareba Systems, Inc, of Ellendale, Minn.

In this embodiment, located proximate to balloon **10**, is an electrical collection enhancement lead **11** which assists in the collection of the static electrical charge in the atmosphere. Electrical collection enhancement lead **11** is configured to attract the static charge and conduct the charge into the conductive line **12**.

The electricity flows down the conductive line into spool **13**, where the conductive line **12** is collected and either withdrawn or dispensed through operation of winch motor **14**.

Winch motor **14** and spool **13** are mounted onto base unit **16** which is electrically isolated from ground **7** using insulator **17**. Note, in this embodiment of the invention, when electricity is being collected from the atmosphere, the entire base unit **16** becomes charged. In another embodiment of the invention, spool **13** is constructed of rubber, thereby preventing base unit **16** from becoming charged, thereby restricting the charging from the atmosphere to only conductive line **12**.

In this embodiment, conductor **6** is connected to base unit **16** (since the entire base unit **16** is charged and the base unit is metallic) to communicate the electrical current to load **5**. Conductor **6** is ideally an insulated wire.

The electrical current through conductor **6** is measured using sensor **8**.

In the alternative embodiment discussed above, where only the conductive line **12** is charged, then conductor **6** is connected to conductive line **12**.

Controller **15**, located in this embodiment on base unit **16**, operates winch motor **14** in response to signals from sensor **8** (measuring the current being discharged to load **5**) to maintain the current flow within a pre-defined range. As the current flow diminishes, then the conductive line **12** extended from spool **13** to increase the altitude of balloon **10** to that more static charge from the atmosphere is gathered; as the current flow falls exceeds a preset level, conductive line **12** is withdrawn onto spool **13** to decrease the static charge being collected from the atmosphere.

The range of current flow through conductor **6** is ideally set by the program, although some embodiments of the invention permit an operator to establish this range of operation.

In an alternative embodiment, the sensor monitoring conductor **6** monitors the voltage therein.

In the preferred embodiment of the invention, controller **15** is also equipped with a lightning sensor **19**. In this embodiment, when lightning is sensed within a preset range, then substantially all of conductive line **12** is wound onto spool **13** to pull balloon **10** near the ground and protect the entire mechanism from being damaged from a lightning discharge.

In the preferred embodiment, the "safe" distance form lightning is set in the programming of controller **15** and is ideally two miles; other embodiments permit the operator to "safe" distance.

There are a variety of lightning sensors well known to those of ordinary skill in the art, including, but not limited to those described in: U.S. Pat. No. 7,016,785, entitled "Lightning Detection" issued to Makela, et al. on Mar. 21, 2006; U.S. Pat. No. 6,829,911, entitled "Lightning Detection and Prediction Alarm Device" issued to Jones, et al. on Dec. 7, 2004; U.S. Pat. No. 7,200,418, entitled "Detection of Lightning" issued to Karikuranta, et al. on Apr. 3, 2007; and U.S. Pat. No. 6,961,662, entitled "Systems and Methods for Spectral Corrected Lightning Detection" issued to Murphy on Nov. 1, 2005; all of which are incorporated hereinto by reference.

In another embodiment of the invention, controller **15** is not located on base unit **16**, rather it is remote and communicates its control signals to winch motor **14** using radio waves.

FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

Static charges **23** are generated in the atmosphere by agitated air. These static charges are often collected at the bottom of clouds, but exist in other environments as well.

Balloon **21** is extended into this strata of static charges **23** which are then attracted to conductive line **12** to flow to base unit **22** and then onto load **5**.

By increasing or decreasing the altitude of balloon **21** (defined by the length of the extended conductive line **12**), conductive line **12** is selectively exposed to varying densities and levels of the static charge strata, and by extension, the current flow or voltage is increased or decreased.

FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

Once the program starts **30**, the lightning sensor is checked to determine if lightning has occurred within the unsafe range **31**A, if it has, then the balloon is lowered **32**A, and the program continues monitoring the status of lightning until no lightning is detected.

When the lightning status is acceptable, then the current within the conductor is checked to see if the current is within the prescribed range **31**B. If the current is acceptable (within range) the program returns to check the lightning status **31**A; otherwise a determination is made to see if the current is above the prescribed range **31**C.

If the current is above the prescribed range, then the altitude of the balloon is withdrawn a set amount **32**B (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

If the current is below the prescribed range, then the altitude of the balloon is extended a set amount **32**C (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

In this manner of feed-back and minor adjustments in the altitude of the balloon, the current is maintained within a prescribed range which can be handled by the downstream electrical system.

As noted earlier, some embodiments of the invention monitor the voltage instead of the current.

FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

By maintaining the voltage being collected in a prescribed range, an electrical conversion system is easily designed. While FIGS. **4**A, **4**B, and **4**C illustrate some electrical configurations, those of ordinary skill in the art readily recognize a variety of other configurations which will serve the same function.

Referencing FIG. **4**A, Direct Current In (DC IN) **40** is buffered by a gang of capacitors **41** before being communicated to a DC/AC converter **42**. The DC/AC converter converts the direct current into a an alternating current suitable for placement over an existing electrical grid **43** such as normally found from a power-plant.

Those of ordinary skill in the art readily recognize a variety of DC/AC converters, including, but not limited to: U.S. Pat. No. 7,394,671, entitled "Controller IC, DC-AC Conversion Apparatus, and parallel running system of DC-AC Conversion Apparatuses" issued to Fukumoto, et al. on Jul. 1, 2008; and, U.S. Pat. No. 7,330,366, entitled "DC-AC Converter" issued to Lee, et al. on Feb. 12, 2008; all of which are incorporated hereinto by reference.

FIG. **4**B illustrates an electrical arrangement suitable for use in charging a battery. DC IN **40** is buffered by capacitor bank **41** before entering into a step down transformer **43**. Step down transformer **43** reduces the voltage so that the voltage can safely be introduced into battery **44** which is connected to ground **45** at the battery's other pole.

Those of ordinary skill in the art readily recognize a variety of batteries which will work in this capacity, including, but not limited to those described in: U.S. Pat. No. 7,378,181, entitled "Electric Storage Battery Construction and Manufacture" issued to Skinlo on May 27, 2008; U.S. Pat. No. 7,388,350, entitled "Battery with Electronic Compartment" issued to Wright on Jun. 17, 2008; U.S. Pat. No. 7,397,220, entitled "Connection Member and Battery Pack" issued to Uchida, et al. on Jul. 8, 2008; and, U.S. Pat. No. 7,375,492, entitled "Inductively Charged Battery Pack" issued to Calhoon, et al. on May 20, 2008; all of which are incorporated hereinto by reference.

In FIG. **4**C, DC IN **40** is fed into an adjustable rheostat **46** which is controlled by the controller so that the DC OUT **47** falls within a specified range.

FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

This type of conductive line is commonly called poly-wire and consists of multiple interwoven strands of plastic **50**A and **50**B woven into a cord or rope arrangement having intertwined therein exposed metal wires **51**A and **51**B. While this illustration shows two plastic strands and two metal wires, any number of possible combinations is possible.

The exposed metal wires **51**A and **51**B attract the atmospheric static charge and transmit the charge down to the base unit (not shown).

FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

This conductive line utilizes a tube **60** having an outer layer **62** of PET Film (Biaxially-oriented polyethylene terephtalate polyester film) which provides exceptionally high tensile strength and is chemically and dimensionally stable. The tube has an ideal diameter of between two and three inches.

An interior metal coating **61** provides an initial conduit for the flow of static charge. The static charge through the metal forces the tube to expand due to the repulsion experienced by like charges. Further, the flow of electricity causes the interior of the tube **60** to become ionized to provide an additional pathway for the atmospheric static charges to the base unit (not shown).

Because outer layer **62** provides a gas barrier, the resulting ionization is not dissipated by air currents, thereby providing a highly stable pathway.

FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

In this embodiment, controller box **70**, resting on insulating pad **72**, is in communication with the sensors as described above. Using the input from these sensors, when there is flow of electricity through the base unit, warning flashing light **71** is illuminated. To electrically neutralize the mechanism, switch **73** is activated to pass any existing current into the ground **74**.

FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

Referencing FIG. **8**A, enhanced electrical collection lead **82** is a wire mesh which is in electrical communication with conductive line **81** and balloon **80**. Because of the significant amount of metal exposed by enhanced electrical collection lead **82**, more static electricity from the atmosphere is drawn to the collection lead **82**, and then down conductive line **81** to the base unit (not shown).

Conductive lead **82** is positioned proximate to balloon **80**.

In FIG. **8**B, poly-wire **83** has enhanced electrical collection leads **84** wrapped therearound. Collection leads **84** have pointed ends **85**A and **85**B which have a propensity to attract more electricity than rounded ends do.

It is clear from the foregoing that the present invention captures an entirely new source of electrical energy.

What is claimed is:

1. A mechanism to tap an electrical source comprising:

a) a lighter than air balloon suspended in the atmosphere;

b) a base unit having a spool of conductive line on a winch motor, one end of said conductive line secured to said lighter than air balloon, a portion of said conductive line collecting electricity in the atmosphere, said winch motor capable of selectively extending or withdrawing said conductive line from said spool;

c) an insulator electrically isolating said conductive line from ground;

d) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load being powered by collected electricity from said conductive line;

e) an electrical flow sensor monitoring electrical flow through said conductor and generating an electrical flow indicia indicative of said electrical flow in said conductor; and,

f) a controller receiving said electrical flow indicia and selectively operating said winch motor such that said electrical flow indicia remains within a selected operating range.

2. The mechanism to tap an electrical source according to claim **1**, wherein said selected operating range is established by an operator.

3. The mechanism to tap an electrical source according to claim **2**, further including a warning light activated when said electrical flow indicia is non-zero.

4. The mechanism to tap an electrical source according to claim **1**,

a) further including a lightning sensor generating a lightning presence indicia indicative of lightning within a prescribed range, said presence indicia being communicated to said controller; and,

b) wherein said control mechanism, in response to said lightning presence indicia, operates said winch motor to withdraw substantially all of said conductive line onto said spool.

5. The mechanism to tap an electrical source according to claim **4**, wherein said prescribed range is established by an operator.

6. The mechanism to tap an electrical source according to claim **1**, further including an electrical collection enhancement lead in electrical contact with a first end of said conductive material, said electrical collection enhancement lead configured to attract static electricity.

7. The mechanism to tap an electrical source according to claim **6**, wherein said electrical collection enhancement lead is positioned proximate to said lighter than air balloon.

8. The mechanism to tap an electrical source according to claim **7**, wherein said electrical collection enhancement lead includes at least two pointed electrical conductors.

9. The mechanism to tap an electrical source according to claim **1**, wherein said insulator electrically isolates said conductive line from said winch motor.

10. The mechanism to tap an electrical source according to claim **1**, wherein said insulator electrically isolates said base unit from ground.

11. A mechanism comprising:

a) an airborne aircraft having a conductive line secured to a winch capable of extending or withdrawing said conductive line from a spool, when said airborne aircraft is aloft, said conductive line collecting electricity from the atmosphere;

b) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load such that collected electricity from said conductive line powers said load;

c) an electrical flow sensor monitoring electrical flow through said conductor and generating an electrical flow indicia indicative of said electrical flow in said conductor; and,

d) a controller receiving said electrical flow indicia adjusting an altitude of said airborne aircraft via said winch such that said electrical flow indicia remains within a selected operating range.

12. The mechanism to tap an electrical source according to claim **11**,

a) further including a lightning sensor communicating with said controller; and,

b) wherein said controller, in response to selected signals from said lightning sensor withdraws substantially all of said conductive line.

13. A mechanism to tap an electrical source comprising:

a) a lighter than air balloon suspended in the atmosphere;

b) a base unit having a spool of conductive line on a winch motor, one end of said conductive line suspended by said balloon, a portion of said conductive line collecting electricity from the atmosphere, said winch motor capable of adjusting an altitude of said lighter than air balloon by selectively extending or withdrawing said conductive line from said spool;

c) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load being powered by said collected electricity from said conductive line;

d) a sensor array having,

1) an electrical flow sensor monitoring electrical flow through said conductor and generating an electrical flow indicia indicative of said electrical flow in said conductor, and,

2) a lightning sensor monitoring existence of proximate lightning; and,

e) a controller responsive to said electrical flow indicia from said sensor array to selectively operate said winch motor.

14. The mechanism to tap an electrical source according to claim **13**,

a) further including a warning light; and,

b) wherein said controller activates said warning light when said electrical flow in said conductor is non-zero.

15. The mechanism to tap an electrical source according to claim **14**, further including an electrical collection enhancement lead in electrical contact with said conductive material, said electrical collection enhancement lead configured to attract static electricity and positioned proximate to said lighter than air balloon.

* * * * *

US008102078B2

(12) United States Patent
Ogram

(10) **Patent No.: US 8,102,078 B2**
(45) **Date of Patent: *Jan. 24, 2012**

(54) **DYNAMIC ELECTRICAL CONVERTER SYSTEM**

(75) Inventor: **Mark Ellery Ogram**, Tucson, AZ (US)

(73) Assignee: **Sefe, Inc.**, Scottsdale, AZ (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 561 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **12/231,663**

(22) Filed: **Sep. 4, 2008**

(65) **Prior Publication Data**

US 2010/0007321 A1 Jan. 14, 2010

Related U.S. Application Data

(63) Continuation-in-part of application No. 12/218,297, filed on Jul. 14, 2008, now Pat. No. 7,855,476.

(51) **Int. Cl.**
H02G 11/00 (2006.01)

(52) **U.S. Cl.** **307/82**; 307/145; 307/149; 361/218; 361/230

(58) **Field of Classification Search** 307/80, 307/82–84, 130, 145, 149; 361/212, 215, 361/216, 217, 218, 230, 231; 174/2; 244/30, 244/31

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,730,122 A *	3/1988	Dreibelbis et al.	....	307/75
5,521,603 A *	5/1996	Young	....	342/198
5,923,516 A *	7/1999	Young	....	361/111
2009/0212758 A1*	8/2009	Asinovski et al.	....	323/355

* cited by examiner

Primary Examiner — Fritz M Fleming

(74) *Attorney, Agent, or Firm* — Mintz Levin Cohn Ferris Glovsky and Popeo, P.C.

(57) **ABSTRACT**

A dynamic electrical converter system configured to receive a variable voltage input which is then converted to a proper electrical configuration for a load. The converter assembly utilized multiple converters, each of which is configured to accept a unique range of voltages and from these voltages creating the desired electrical output configuration. A monitor checks the incoming voltage and by controlling switches going to each of the converters, selectively closes the circuit to the appropriate converter; thereby, routing the incoming voltage to the converter capable of addressing the incoming voltage.

20 Claims, 5 Drawing Sheets

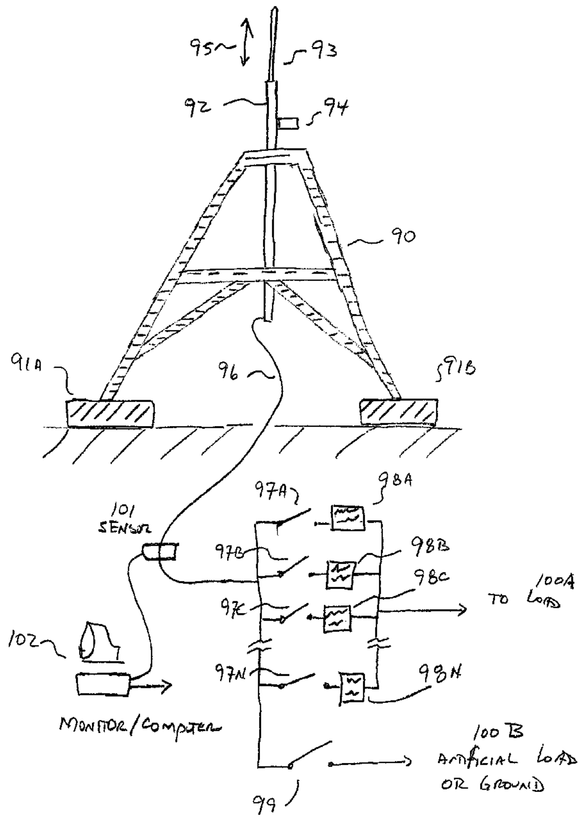

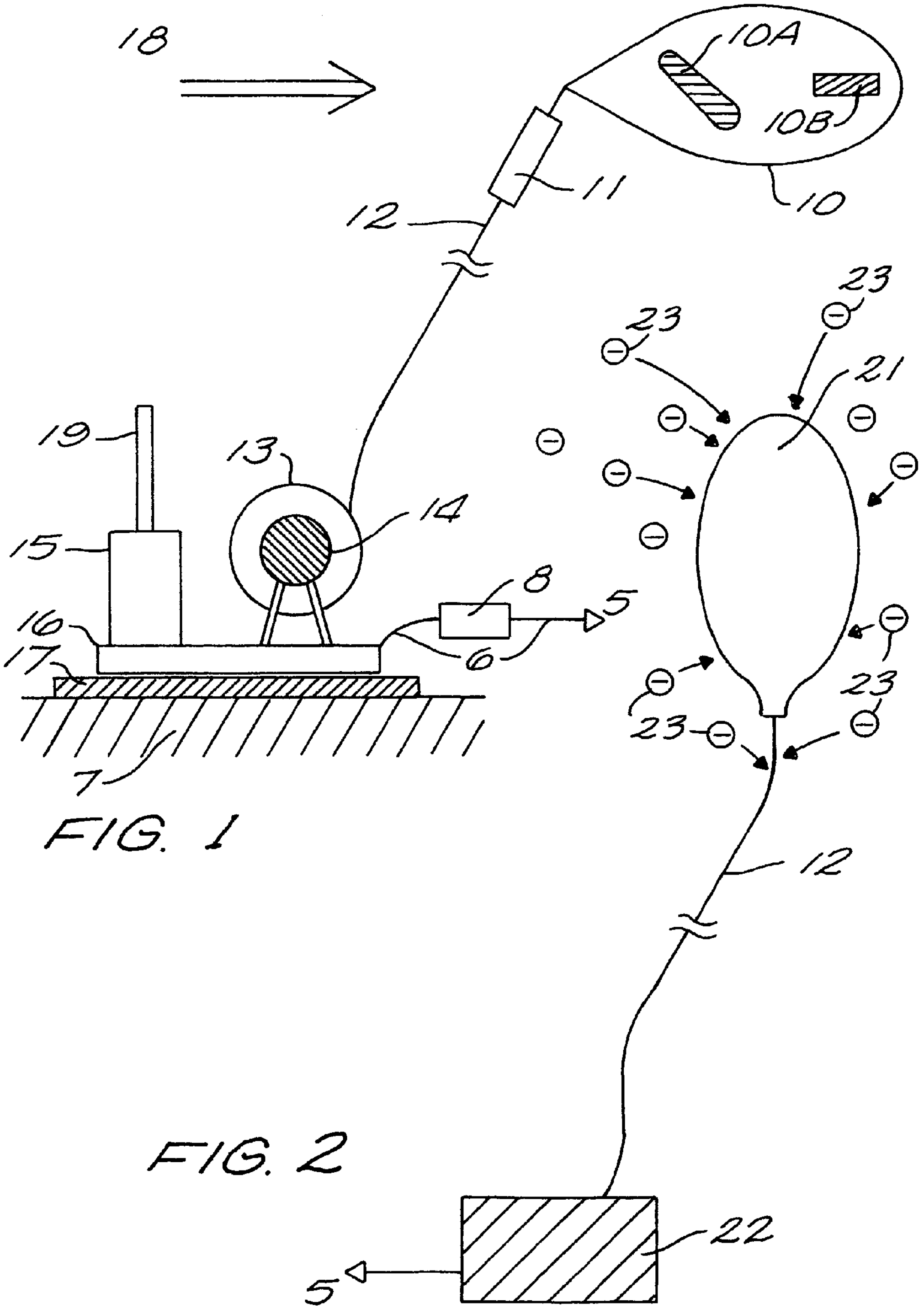
18
10A
10B
10
11
12
19
13
14
15
8
5
16
17
6
7
23
23
21
23
23
23
12
22
5

FIG. 1

FIG. 2

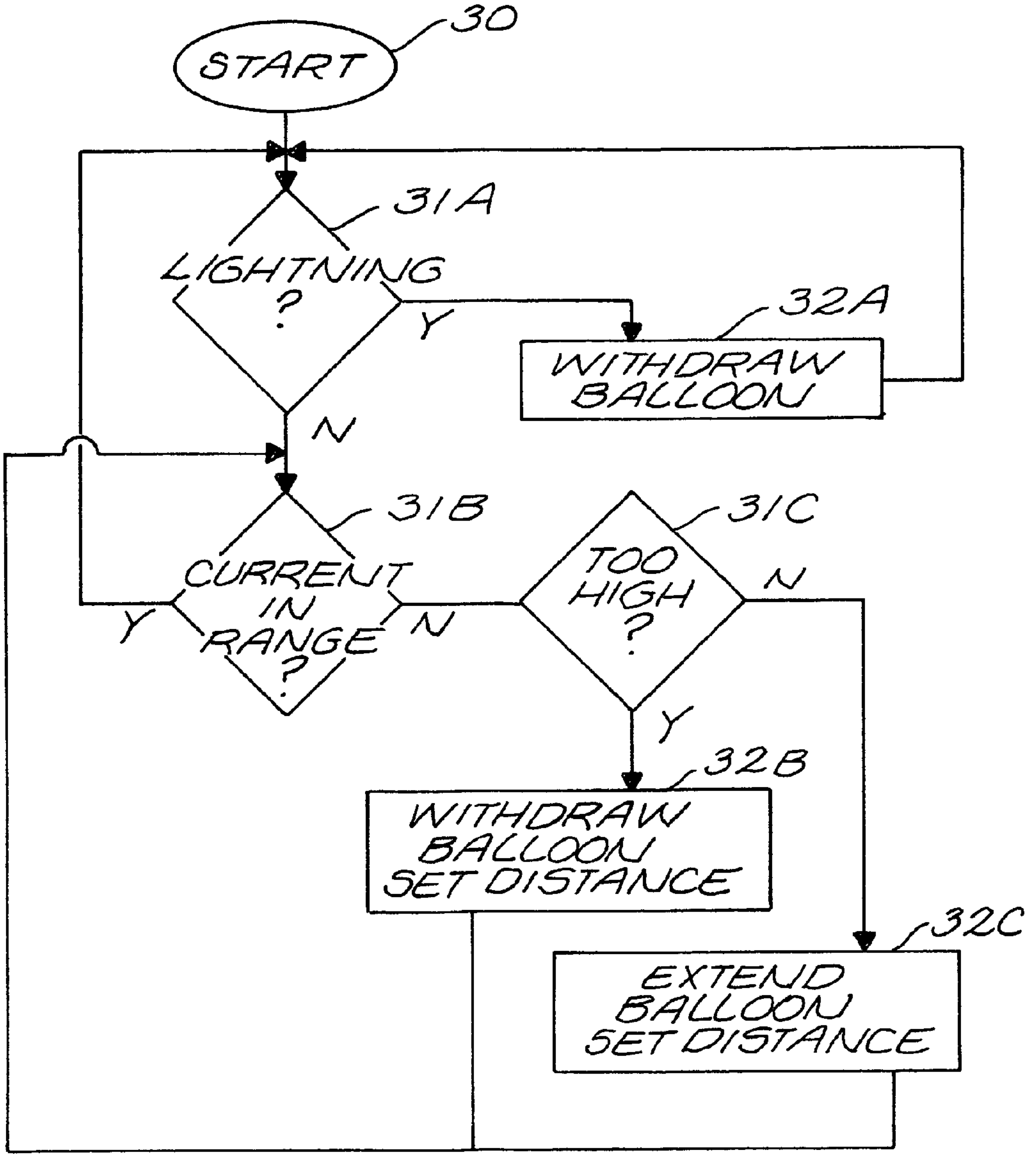
30
START
31A
LIGHTNING ?
Y
32A
WITHDRAW BALLOON
N
31B
CURRENT IN RANGE ?
Y
N
31C
TOO HIGH ?
N
Y
32B
WITHDRAW BALLOON SET DISTANCE
32C
EXTEND BALLOON SET DISTANCE

FIG. 3

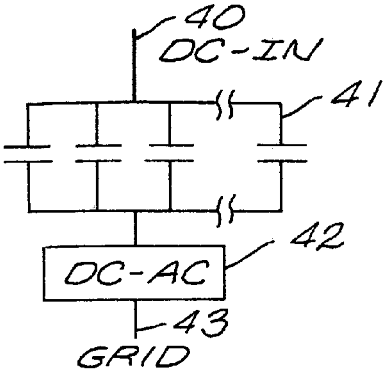
40
DC-IN
41
DC-AC
42
43
GRID

FIG. 4A

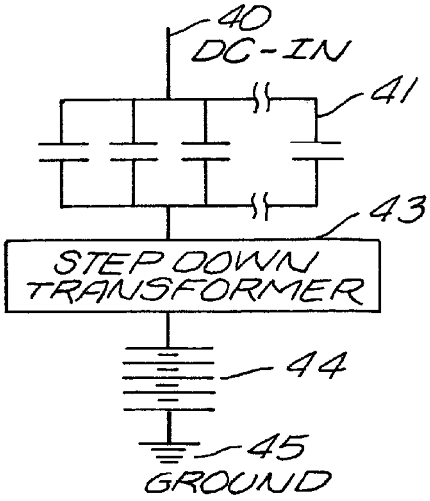
40
DC-IN
41
43
STEP DOWN TRANSFORMER
44
45
GROUND

FIG. 4B

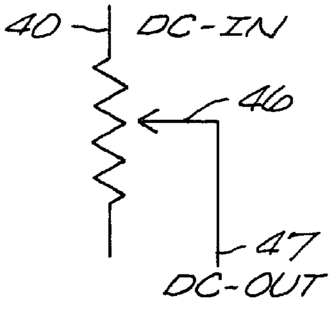
40
DC-IN
46
47
DC-OUT

FIG. 4C

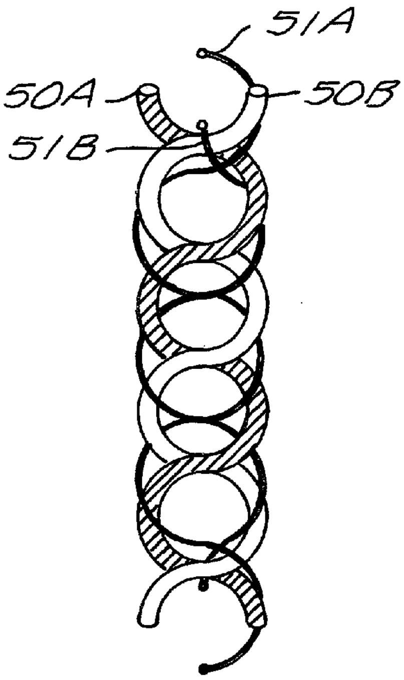
51A
50A
50B
51B

FIG. 5

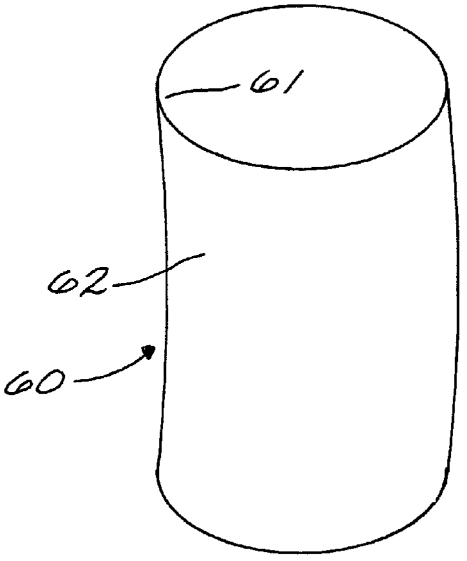

FIG. 6A

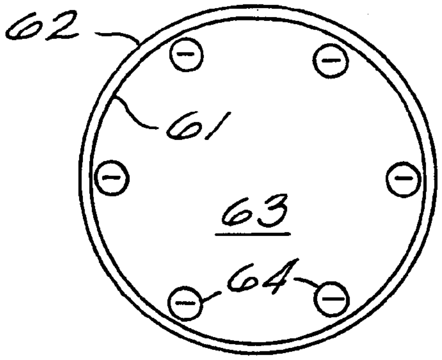

FIG. 6B

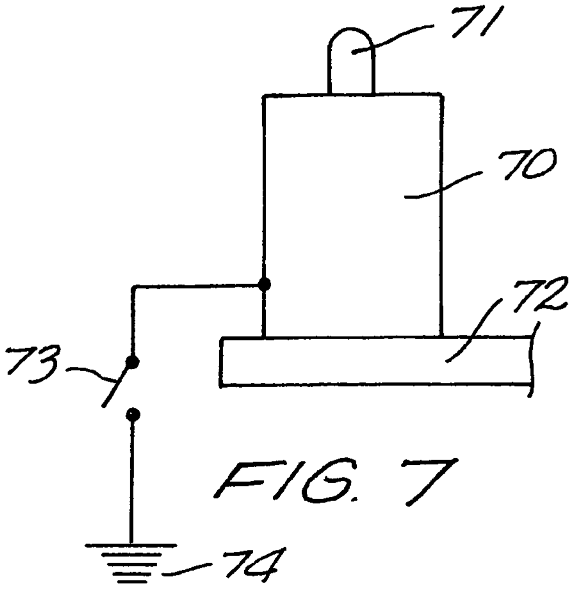
71
70
72
73
74

FIG. 7

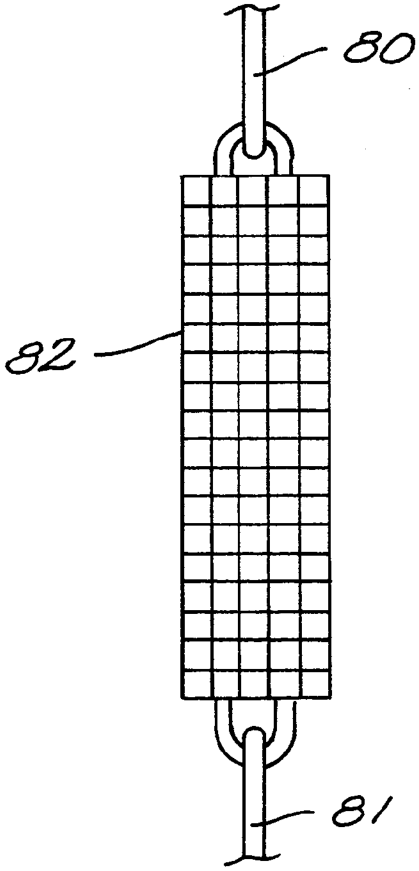

FIG. 8A

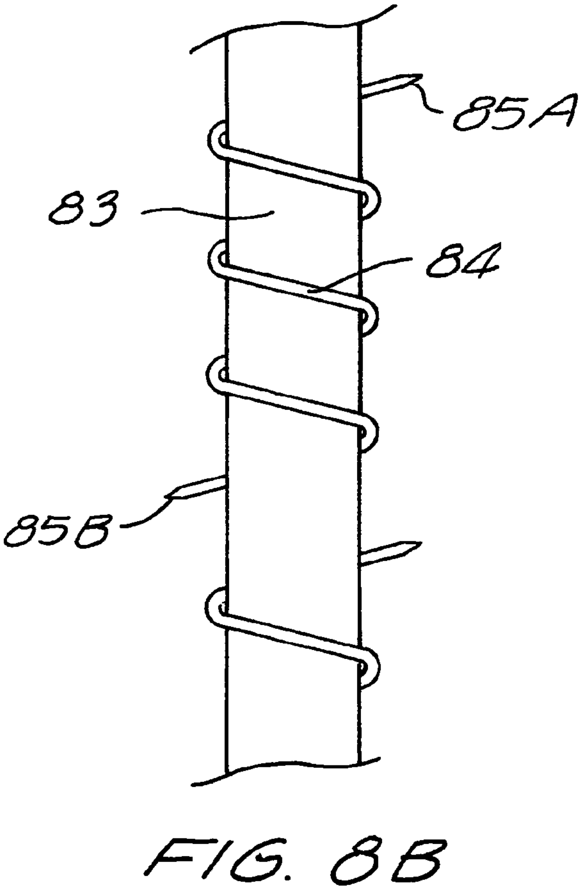
85A
83
84
85B

FIG. 8B

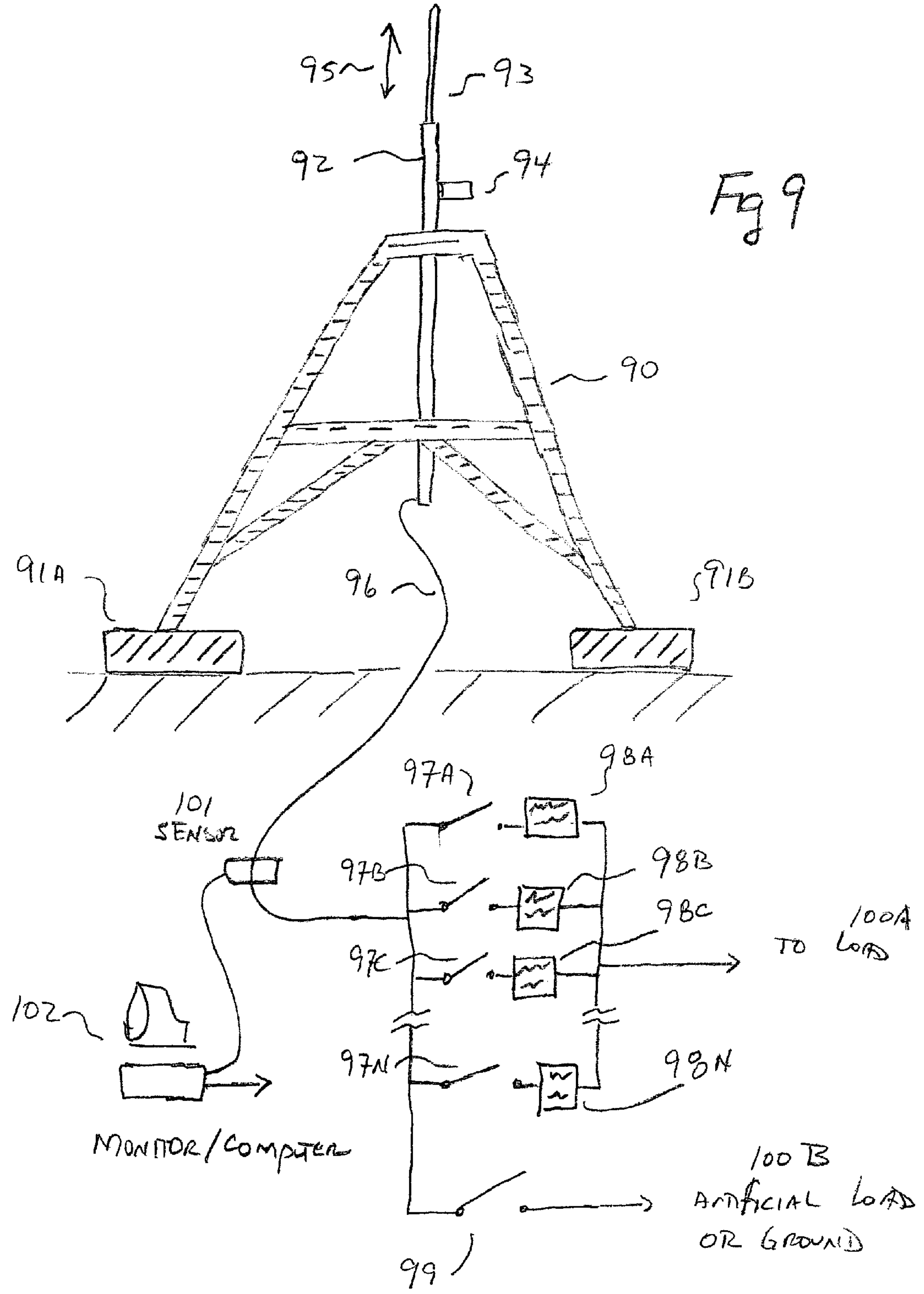
Fig 9
SENSOR
TO LOAD
97N
98N
MONITOR/COMPUTER
100 B
ARTIFICIAL LOAD
OR GROUND

DYNAMIC ELECTRICAL CONVERTER SYSTEM

This is a continuation in part of U.S. patent application Ser. No. 12/218,297, filed on Jul. 14, 2008 now U.S. Pat. No. 7,855,476, and entitled, "Atmospheric Electrical Generator".

BACKGROUND OF THE INVENTION

This invention relates to the production of electrical energy and more particularly to the a conversion system for variable electrical voltages.

Everyone is familiar with Benjamin Franklin's kite experiment of 1752. Using a kite whose string had become wet, negative charges from the passing clouds flowed into the string, down to the suspended key, and then into a Leyden jar via a thin metal wire. Franklin was protected by a dry silk string; but, when Franklin's knuckle came too close to the key, he received a strong shock. Fortunately, Benjamin Franklin was not killed, others who tried this same experiment were not so lucky.

Since then, the formation of lightning has remained something of a mystery. Lightning bolts are triggered when a negatively charged cloud base induces a positive charge from the ground, thereby forming a "pathway" for the discharge of the collected electrical energy.

Lightning travels up to 60,000 miles per hour with a flash that is brighter than ten million 100-watt lightbulbs. This wattage is as much power as is produced by all of the electricity plants in the United States and with a voltage of up to 300 million volts.

It is this very fact, the power within lightning is immense, that has prevented any successful collection of the electrical energy from lightning. The electricity in lightning is far too extreme for current technology to harness.

While lightning has attracted a energy starved industrial world, no one has developed any technique to harness this naturally occurring electrical source.

It is clear there is a continuing need for an electrical source other than carbon-based fuels and that the naturally occurring electricity in the atmosphere is being ignored.

SUMMARY OF THE INVENTION

The invention is a mechanism which taps into the naturally occurring static electricity in the atmosphere. Whereas heretofore, the attempt to garner electricity from the atmosphere has focused exclusively on capturing lightning, the present invention syphons off the static electricity which is generated from any agitated air and voids lightning.

Lightning is only the final discharge of the static electricity, whether that lightning is intra-cloud lightning, cloud-to-ground lightning, or inter-cloud lightning. Other types of final discharges are known as heat lightning, summer lightning, sheet lightning, ribbon lightning, silent lightning, ball lightning, bead lightning, elves, jets, and sprites. Well before these discharges are observed, as the atmosphere becomes agitated by wind or thermal, static electricity is being generated.

The present invention recognizes that this static electricity is being formed and creates a mechanism to capture it.

The mechanism of this invention utilizes an aircraft such as a lighter than air balloon. While the preferred embodiment uses a foil balloon, a variety of other aircraft are obvious to those of ordinary skill in the art, including, but not limited to: gliders, rubber balloons (such as weather balloons), biaxially-oriented polyethylene terephthalate polyester film balloons, and latex balloons.

Within this discussion, the balloon is referenced, but, the invention is not intended to be limited solely to balloons.

The balloon is sent aloft and is tethered by a conductive line. In this context, the conductive line may be any obvious to those of ordinary skill in the art. For the preferred embodiment, the conductive line is a generically referred to as a "poly-rope" and is commercially available through a variety of sources. A suitable conductive line is described in U.S. Pat. No. 5,203,542, entitled "Apparatus for Improved Electric Fence Wire Construction for use with Intensive Grazing" issued Apr. 20, 1993, to Coley, et al. and incorporated hereinto by reference.

The conductive line is played out of a winch to control the altitude of the balloon. The motor controlling the winch is able to reverse direction to both extend and withdraw the conductive line which is wrapped around a spool on the winch. The winch/spool combination are part of a base unit.

In some embodiments of the invention, the spool is constructed of rubber so as to insulate the conductive line from the winch assembly. In this embodiment, only the conductive line is charged by the atmospheric static electricity while the winch remains neutral.

In yet another embodiment, the winch/spool are part of a base unit which is itself isolated from the ground by an insulator. In this embodiment, the entire base unit is charged by the atmospheric static electricity.

A conductor, such as an insulated wire, is electrically connected to the conductive line. In one embodiment, where the conductive line is electrically isolated from the spool and winch motor, the conductor is connected to the conductive line. In the embodiment where the conductive line is electrically connected to the base unit, then the conductor is connected anywhere on a metalic base unit.

The other end of the conductor is connected to a load. The load in this case can be any of a variety of electrical loads well known to those of ordinary skill in art, including, but not limited to a motor, a battery system, or the electrical grid for the system.

In the preferred embodiment, a sensor array is used to monitor the activities both at the base unit (such as electrical flow within the conductor) and in the surrounding locale.

A sensor monitoring the electrical flow (i.e. voltage and/or current) within the conductor is used to monitor the electrical activity within the conductor.

In the preferred embodiment, a lightning sensor monitors for lightning activity within the locale. As noted earlier, the electrical characteristic of lightning is so extreme that ideally this discharge is avoided as it might damage the mechanism of this invention.

The sensor array is utilized by a controller, such as microprocessor, programmed to operate the mechanism as outlined herein.

The controller operates the winch motor to extend or withdraw the conductive line and by extension the altitude of the balloon. The controller is programmed to operate the winch by monitoring the electrical characteristics of the conductor and adjusting the balloon's altitude to maintain these characteristics within the conductor within a preset range.

This preset range is established either in the base programming of the controller or is established by an operator of the system.

As example, by controlling the amount of current being withdrawn from the atmosphere, the mechanism operates within a safe range and also provides a relatively stable current flow from which a variety of activities can take place (such as DC-AC conversion).

The controller also utilizes the lightning sensor to protect the mechanism from a lightning strike. Should lightning be detected within a pre-determined range (as established by the software or defined by an operator), then the balloon is pulled down to minimize the risk of damage from a lightning strike.

Another aspect of the invention relates to the electrical system which accepts the fluctuating atmospheric charge and changes it into an acceptable configuration for either the desired load or for the existing power grid.

Power grids in the United States operate with a frequency of 60 hertz in an alternating current arrangement. While this basic configuration seems to be universally accepted, the voltage within the grid varies dramatically, such as 15 kv, 34 kv, 69 kv, and even 112 kv.

Each atmospheric generator is placed proximate to or within easy access to a specific grid; this establishes the required electrical output configuration (i.e. that which is accepted by the power grid). As example, one of the atmospheric electrical collector units as described above collects the atmospheric electrical power as direct current and then supplies the appropriate power grid a specific flow (as example, AC, 60 hertz, at 69 kv).

The difficulty lies the fact that the DC current being garnered from the atmosphere varies depending on the actual agitation being generated in the atmosphere. This means that the source of DC current is fluctuating.

The present invention uses a monitoring system which checks the input DC voltage. Depending on the actual voltage being received, the appropriate converter is connected to the input DC voltage so that the desired output is obtained.

As example, suppose the DC input voltage is 1500 volts, the monitoring system, sensing this input, closes the switch connecting the DC voltage to a converter which accepts DC voltage in the range of 1000-2000 volts which then delivers an AC, 60 hertz 69 kv signal to the power grid. If the DC input voltage increases to 2100 volts, then the monitoring system opens the switch to the first converter (1000-2000 volts) and closes the switch to a second converter (such as 2000-4000 volts) to deliver the desired output of AC (60 hertz, 69 kv) for the power grid.

In this manner, regardless of the fluctuating input DC voltage, the electrical grid is supplied with a fully configured electrical input conforming to the needs of that specific electrical grid.

Another aspect of the present invention is the use of a tower or permanent structure instead of an aircraft. In this embodiment, the building or tower is electrically isolated from the ground and a rod (similar to a lightning rod) is extended into the atmosphere. The rod collects the atmospheric charge which is conveyed via an electrical conduit (ideally insulated) where the collected DC charge is reconfigured to meets the need of the locale.

In this context, for one embodiment of the invention, a tower is placed onto the top of a building. The tower is electrically isolated from the building using such mechanisms well known to those of ordinary skill in the art such as rubber mats. A rod ideally extends from the top of the tower to facilitate the collection of the DC electrical energy.

A converter mechanism such as described above, is connected to the tower to flow the DC electricity to a converter which modifies the DC current for the specific application. In one application, the DC current is converted to the electrical needs of the building, thereby providing at least some of the electrical requirements of the building itself.

As noted earlier, the dynamic converter system of the present invention allows a power generator to address a variable voltage in an efficient manner. This makes the dynamic converter system ideal for a variety of alternative energy sources such as the above described atmospheric electrical generator and other alternative energy sources such as wind and wave powered systems. In these systems, the energy being generated must be converted to a proper electrical configuration for a identified load. This may be a particular motor or connection to the power grid which act as a load to the power generating mechanism.

For these energy generating systems, the converter assembly of this invention utilizes multiple converters. Each converter is configured to accept a unique range of voltages and from these voltages, create the desired electrical configuration. By using multiple converters, a full range is available, from a minimum voltage input to a maximum voltage input.

As example, suppose the output voltage needs to be 55 kv AC, but the input voltage usually ranges from 20 kv DC to 100 kv DC. By using four different converts, this 80 kv DC range is broken into:

1) converter A accepting 20 kv DC to 40 kv DC and converting it to 55 kv AC;

2) converter B accepting 40 kv DC to 60 kv DC and converting it to 55 kv AC;

3) converter C accepting 60 kv DC to 80 kv DC and converting it to 55 kv AC; and,

4) converter A accepting 80 kv DC to 100 kv DC and converting it to 55 kv AC.

Those of ordinary skill in the art readily recognize a variety of other arrangement.

A monitor checks the incoming voltage recognizes this voltage as being addressed by a particular converter, and, by controlling switches going to each of the converters directs the input voltage appropriately. Carrying on with the above example, assume the monitor determines that the incoming voltage is 56 kv DC, then converter B is activated by selectively closing the switch associated with converter B and keeping the switches for the other converters open.

In this manner, only the appropriate converter is used for whatever the incoming voltage is.

For the situation of wind power, this permits the wind power generator to operate even in "light" winds and "heavy" winds when the amount of electricity being generated varies significantly from the desired levels. With the increased operating capability, the wind power generator is able to provide a much larger overall energy flow.

The invention, together with various embodiments thereof will be more fully explained by the following description of the accompanying drawings.

DRAWINGS IN BRIEF

FIG. **1** diagrams the preferred embodiment of the invention.

FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

FIG. **9** diagrams the electrical converter system being employed on a stationary structure and the dynamic converter system.

DRAWINGS IN DETAIL

FIG. **1** diagrams the preferred embodiment of the invention.

Balloon **10** is an aircraft which, in this illustration, is a lighter than air balloon. Wings **10**A, extending from the body of balloon **10**, provide additional lift in air flow **18**. Tail **10**B helps to stabilize balloon **10**.

Balloon **10** is tethered to the ground via conductive line **12**. As noted earlier, a variety of configurations and materials are available to serve as conductive line **12**. In this illustration, a poly-wire is used. Poly-wire is commercially available through a variety of vendors, including, but not limited to: Jeffers Livestock and Sareba Systems, Inc. of Ellendale, Minn.

In this embodiment, located proximate to balloon **10**, is an electrical collection enhancement lead **11** which assists in the collection of the static electrical charge in the atmosphere. Electrical collection enhancement lead **11** is configured to attract the static charge and conduct the charge into the conductive line **12**.

The electricity flows down the conductive line into spool **13**, where the conductive line **12** is collected and either withdrawn or dispensed through operation of winch motor **14**.

Winch motor **14** and spool **13** are mounted onto base unit **16** which is electrically isolated from ground **7** using insulator **17**. Note, in this embodiment of the invention, when electricity is being collected from the atmosphere, the entire base unit **16** becomes charged. In another embodiment of the invention, spool **13** is constructed of rubber, thereby preventing base unit **16** from becoming charged, thereby restricting the charging from the atmosphere to only conductive line **12**.

In this embodiment, conductor **6** is connected to base unit **16** (since the entire base unit **16** is charged and the base unit is metalic) to communicate the electrical current to load **5**. Conductor **6** is ideally an insulated wire.

The electrical current through conductor **6** is measured using sensor **8**.

In the alternative embodiment discussed above, where only the conductive line **12** is charged, then conductor **6** is connected to conductive line **12**.

Controller **15**, located in this embodiment on base unit **16**, operates winch motor **14** in response to signals from sensor **8** (measuring the current being discharged to load **5**) to maintain the current flow within a pre-defined range. As the current flow diminishes, then the conductive line **12** extended from spool **13** to increase the altitude of balloon **10** to that more static charge from the atmosphere is gathered; as the current flow falls exceeds a preset level, conductive line **12** is withdrawn onto spool **13** to decrease the static charge being collected from the atmosphere.

The range of current flow through conductor **6** is ideally set by the program, although some embodiments of the invention permit an operator to establish this range of operation.

In an alternative embodiment, the sensor monitoring conductor **6** monitors the voltage therein.

In the preferred embodiment of the invention, controller **15** is also equipped with a lightning sensor **19**. In this embodiment, when lightning is sensed within a preset range, then substantially all of conductive line **12** is wound onto spool **13** to pull balloon **10** near the ground and protect the entire mechanism from being damaged from a lightning discharge.

In the preferred embodiment, the "safe" distance form lightning is set in the programming of controller **15** and is ideally two miles; other embodiments permit the operator to "safe" distance.

There are a variety of lightning sensors well known to those of ordinary skill in the art, including, but not limited to those described in: U.S. Pat. No. 7,016,785, entitled "Lightning Detection" issued to Makela, et al. on Mar. 21, 2006; U.S. Pat. No. 6,829,911, entitled "Lightning Detection and Prediction Alarm Device" issued to Jones, et al. on Dec. 7, 2004; U.S. Pat. No. 7,200,418, entitled "Detection of Lightning" issued to Karikuranta, et al. on Apr. 3, 2007; and U.S. Pat. No. 6,961,662, entitled "Systems and Methods for Spectral Corrected Lightning Detection" issued to Murphy on Nov. 1, 2005; all of which are incorporated hereinto by reference.

In another embodiment of the invention, controller **15** is not located on base unit **16**, rather it is remote and communicates its control signals to winch motor **14** using radio waves.

FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

Static charges **23** are generated in the atmosphere by agitated air. These static charges are often collected at the bottom of clouds, but exist in other environments as well.

Balloon **21** is extended into this strata of static charges **23** which are then attracted to conductive line **12** to flow to base unit **22** and then onto load **5**.

By increasing or decreasing the altitude of balloon **21** (defined by the length of the extended conductive line **12**), conductive line **12** is selectively exposed to varying densities and levels of the static charge strata, and by extension, the current flow or voltage is increased or decreased.

FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

Once the program starts **30**, the lightning sensor is checked to determine if lightning has occurred within the unsafe range **31**A, if it has, then the balloon is lowered **32**A, and the program continues monitoring the status of lightning until no lightning is detected.

When the lightning status is acceptable, then the current within the conductor is checked to see if the current is within the prescribed range **31**B. If the current is acceptable (within range) the program returns to check the lightning status **31**A; otherwise a determination is made to see if the current is above the prescribed range **31**C.

If the current is above the prescribed range, then the altitude of the balloon is withdrawn a set amount **32**B (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

If the current is below the prescribed range, then the altitude of the balloon is extended a set amount **32**C (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

In this manner of feed-back and minor adjustments in the altitude of the balloon, the current is maintained within a prescribed range which can be handled by the downstream electrical system.

As noted earlier, some embodiments of the invention monitor the voltage instead of the current.

FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

By maintaining the voltage being collected in a prescribed range, an electrical conversion system is easily designed. While FIGS. **4**A, **4**B, and **4**C illustrate some electrical configurations, those of ordinary skill in the art readily recognize a variety of other configurations which will serve the same function.

Referencing FIG. **4**A, Direct Current In (DC IN) **40** is buffered by a gang of capacitors **41** before being communicated to a DC/AC converter **42**. The DC/AC converter converts the direct current into a an alternating current suitable for placement over an existing electrical grid **43** such as normally found from a power-plant.

Those of ordinary skill in the art readily recognize a variety of DC/AC converters, including, but not limited to: U.S. Pat. No. 7,394,671, entitled "Controller IC, DC-AC Conversion Apparatus, and parallel running system of DC-AC Conversion Apparatuses" issued to Fukumoto, et al. on Jul. 1, 2008; and, U.S. Pat. No. 7,330,366, entitled "DC-AC Converter" issued to Lee, et al. on Feb. 12, 2008; all of which are incorporated hereinto by reference.

FIG. **4**B illustrates an electrical arrangement suitable for use in charging a battery. DC IN **40** is buffered by capacitor bank **41** before entering into a step down transformer **43**. Step down transformer **43** reduces the voltage so that the voltage can safely be introduced into battery **44** which is connected to ground **45** at the battery's other pole.

Those of ordinary skill in the art readily recognize a variety of batteries which will work in this capacity, including, but not limited to those described in: U.S. Pat. No. 7,378,181, entitled "Electric Storage Battery Construction and Manufacture" issued to Skinlo on May 27, 2008; U.S. Pat. No. 7,388,350, entitled "Battery with Electronic Compartment" issued to Wright on Jun. 17, 2008; U.S. Pat. No. 7,397,220, entitled "Connection Member and Battery Pack" issued to Uchida, et al. on Jul. 8, 2008; and, U.S. Pat. No. 7,375,492, entitled "Inductively Charged Battery Pack" issued to Calhoon, et al. on May 20, 2008; all of which are incorporated hereinto by reference.

In FIG. **4**C, DC IN **40** is fed into an adjustable rheostat **46** which is controlled by the controller so that the DC OUT **47** falls within a specified range.

FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

This type of conductive line is commonly called poly-wire and consists of multiple interwoven strands of plastic **50**A and **50**B woven into a cord or rope arrangement having intertwined therein exposed metal wires **51**A and **51**B. While this illustration shows two plastic strands and two metal wires, any number of possible combinations is possible.

The exposed metal wires **51**A and **51**B attract the atmospheric static charge and transmit the charge down to the base unit (not shown).

FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

This conductive line utilizes a tube **60** having an outer layer **62** of PET Film (Biaxially-oriented polyethylene terephtalate polyester film) which provides exceptionally high tensile strength and is chemically and dimensionally stable. The tube has an ideal diameter of between two and three inches.

An interior metal coating **61** provides an initial conduit for the flow of static charge. The static charge through the metal forces the tube to expand due to the repulsion experienced by like charges. Further, the flow of electricity causes the interior of the tube **60** to become ionized to provide an additional pathway for the atmospheric static charges to the base unit (not shown).

Because outer layer **62** provides a gas barrier, the resulting ionization is not dissipated by air currents, thereby providing a highly stable pathway.

FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

In this embodiment, controller box **70**, resting on insulating pad **72**, is in communication with the sensors as described above. Using the input from these sensors, when there is flow of electricity through the base unit, warning flashing light **71** is illuminated. To electrically neutralize the mechanism, switch **73** is activated to pass any existing current into the ground **74**.

FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

Referencing FIG. **8**A, enhanced electrical collection lead **82** is a wire mesh which is in electrical communication with conductive line **81** and balloon **80**. Because of the significant amount of metal exposed by enhanced electrical collection lead **82**, more static electricity from the atmosphere is drawn to the collection lead **82**, and then down conductive line **81** to the base unit (not shown).

Conductive lead **82** is positioned proximate to balloon **80**.

In FIG. **4**B, poly-wire **83** has enhanced electrical collection leads **84** wrapped therearound. Collection leads **84** have pointed ends **85**A and **85**B which have a propensity to attract more electricity than rounded ends do.

FIG. **9** diagrams the electrical converter system being employed on a stationary structure and the dynamic converter system.

This illustration uses a structure for the collection of the atmospheric energy. The circuit described in this illustration is not to be construed only to this energy collection mechanism, but, is adapted work with any number of energy collection/generation mechanisms, including, but not limited to: solar collectors; wind power generators; wave motion power generators; and, the atmospheric energy collection mechanisms described herein.

The structure of this embodiment is tower **90** which is electrically isolated from the ground via insulators **91**A and **91**B. In this manner, any atmospheric energy collected is maintained within tower **90**.

Conduit **90** contains antenna **93** which is moved up/down as illustrated by arrow **95** by motor **94**. Motor **94** is controlled by monitor/computer **102** based upon the electrical energy being collected by antenna **93** and communicated via insulated conductor **96**, monitored by sensor **101**.

The collected energy is communicated to a bank of switches (**97**A, **97**B, **97**C, through **97**N, N being any number of switches). In this manner, each switch receives the electrical energy being gathered by antenna **93**. As noted earlier, this electrical energy varies depending on the ambient conditions.

Each switch addresses a single converter. As illustrated, switch **97**A, when closed, communicates the electrical energy to converter **98**A; switch **97**B to converter **98**B; switch **97**C to converter **98**C; and so on until switch **97**N to converter **98**N.

Each converter is designed to accept a range of electrical input and generate a uniform electrical configuration to load **100**A.

Monitor/computer **102** monitors the varying electrical input from antenna **93** via sensor **101**. Based upon the electrical input, monitor/computer **102** closes a selected one of the switches (and opens all of the rest) so that the voltage being received is directed to the converter designed to handle that voltage.

In some situations, the voltage is so large that (or during emergency) that emergency switch **99** must be closed to direct the voltage to an artificial load or ground either during the switching of switches **97**A-**97**N or to "shut down" the system for repairs/maintenance or if lightning is sensed within a "dangerous" range of tower **90**.

In this context, an artificial load is anything that expends the incoming energy, such as lifting a weight to let it fall once switch **99** is opened, or pumping water out of and then into the same reservoir.

It is clear from the foregoing that the present invention captures an entirely new source of electrical energy.

What is claimed is:

1. An electrical converter system comprising:

a static electricity collector having a conductor that collects static electricity from the atmosphere and outputs a collected energy, the conductor having a height from a stationary structure, the height being adjustable by a motor coupled with the stationary structure;

a plurality of converters in electrical communication with the conductor, each converter configured to accept a unique range of voltages and producing a uniform electrical output;

a plurality of switches, each switch receiving an input voltage based on the collected energy from the conductor, and connected to corresponding one of said plurality of converters; and

a monitor having a sensor that measures the collected energy from the conductor, the monitor operating the motor to adjust the height of the conductor relative to the stationary structure such that the collected energy remains within a selected operating range, the sensor further identifying a level of the input voltage of each of switch to selectively close one of the plurality of switches in response to the level of the input voltage and based on the unique range of voltages.

2. The electrical converter system according to claim **1**, further comprising:

a load; and

a load switch connecting said input voltage to said load,

wherein said monitor closes, when the level of input voltage exceeds a predetermined value, said load switch prior to closing one of said plurality of switches.

3. The electrical converter system according to claim **2**, wherein said load switch routes said input voltage to ground.

4. The electrical converter system according to claim **2**, wherein said monitor closes said load switch if said input voltage exceeds the unique ranges of voltages for all of said plurality of converters.

5. The electrical converter system according to claim **2**, further comprising:

a lightning sensor generating a signal indicative of a distance to a lightning activity, said monitor closing said load switch when said signal indicative of a distance to a lightning activity is within a prescribed range.

6. The electrical converter system according to claim **2**, wherein said monitor opens said load switch after closing one of said plurality of switches.

7. The electrical converter system according to claim **1**, wherein said input voltage is direct current.

8. The electrical converter system according to claim **1**, wherein said input voltage is alternating current.

9. The electrical converter system according to claim **1**, wherein each converter of the plurality of converters converts direct current to alternating current.

10. An electrical system comprising:

a static electricity collector having a conductor that collects static electricity from the atmosphere and outputs a collected energy, the conductor having a height from a stationary structure, the height being adjustable by a motor coupled with the stationary structure;

a power generating mechanism in electrical contact with the conductor and generating a variable voltage from the static electricity;

a power grid electrically connected to consumers of electricity, said power grid operating at a uniform electrical configuration; and,

a conversion system comprising:

- a plurality of converters, each converter configured to accept a unique range of voltages and producing said uniform electrical configuration to said power grid;
- a plurality of switches, each switch receiving the variable voltage and connected to corresponding one of said plurality of converters; and

a monitor comprising a sensor that measures the collected energy from the conductor, the monitor operating the motor to adjust the height of the conductor relative to the stationary structure such that the collected energy remains within a selected operating range, the sensor further identifying a level of the variable voltage, and, in response to said level of the variable voltage, selectively closing one of said plurality of switches.

11. The electrical system according to claim **10**, wherein said variable voltage is direct current, the direct current being converted to alternating current by the corresponding one of said plurality of converters.

12. The electrical system according to claim **10**, wherein said power generating mechanism is reactive to atmospheric voltage.

13. The electrical system according to claim **12**, further comprising:

a lighting sensor generating a signal indicative of a distance to a lightning activity said monitor closing an artificial load switch when said signal indicative of a distance to a lightning activity is within a prescribed range.

14. The electrical converter system according to claim **10**, wherein said variable voltage is alternating current.

15. The electrical system according to claim **14**, wherein said power generating mechanism is reactive to wind power.

16. An electrical converter system receiving a variable voltage input and connected to a load requiring a set electrical configuration, said system comprising:

a static electricity collector having a conductor that collects static electricity from the atmosphere and outputs a collected energy, the conductor having a height from a stationary structure, the height being adjustable by a motor coupled with the stationary structure;

a converter assembly in electrical communication with the conductor comprising:

a plurality of converters, each converter configured to accept a unique range of voltages from the collected energy and producing the set electrical configuration, and,

a plurality of switches, each switch receiving the variable voltage input and connected to corresponding one of said plurality of converters;

a sensor that measures the collected energy from the conductor, the sensor further identifying a level of the variable voltage input; and

a monitor that operates the motor to adjust the height of the conductor relative to the stationary structure such that the collected energy remains within a selected operating range, the monitor further selectively closing, in response to said level of the input voltage, one of said plurality of switches.

17. The system according to claim **16**, further comprising:

a load; and

a load switch connecting said variable voltage input to said load, said monitor closing said load switch prior to closing one of said plurality of switches.

18. The system according to claim **17**, wherein said load switch routes said variable voltage input to ground.

19. The system according to claim **17**, further comprising:

a lighting sensor generating a signal indicative of a distance to a lightning activity said monitor closing said load switch when said signal indicative of a distance to a lightning activity is within a prescribed range.

20. The system according to claim **16**, wherein said monitor closes said load switch if said variable voltage input exceeds the unique range of voltages for all of said plurality of converters.

* * * * *

US008283811B2

(12) **United States Patent**
Ogram

(10) **Patent No.:** **US 8,283,811 B2**
(45) **Date of Patent:** ***Oct. 9, 2012**

(54) **ATMOSPHERIC STATIC ELECTRICITY COLLECTOR**

(75) Inventor: **Mark Ellery Ogram**, Tucson, AZ (US)

(73) Assignee: **Sefe, Inc.**, Scottsdale, AZ (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **13/356,631**

(22) Filed: **Jan. 23, 2012**

(65) **Prior Publication Data**

US 2012/0119590 A1 May 17, 2012

Related U.S. Application Data

(63) Continuation of application No. 12/321,306, filed on Jan. 16, 2009, now Pat. No. 8,102,082, which is a continuation-in-part of application No. 12/218,297, filed on Jul. 14, 2008, now Pat. No. 7,855,476.

(51) **Int. Cl.**
H02G 11/00 (2006.01)

(52) **U.S. Cl.** **307/145**; 307/149; 307/151; 361/218; 361/230

(58) **Field of Classification Search** 307/145, 307/149, 151; 244/30, 31; 174/2; 361/212, 361/215, 216, 217, 218, 230, 231
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,855,476 B2 * 12/2010 Ogram 307/145
8,102,082 B2 * 1/2012 Ogram 307/145

* cited by examiner

Primary Examiner — Fritz M Fleming

(74) *Attorney, Agent, or Firm* — Mintz Levin Cohn Ferris Glovsky and Popeo, P.C.

(57) **ABSTRACT**

An antenna for the collection of atmospheric static electricity in which an electrically conductive hub is suspended from a balloon or blimp via a tether. The hub is either solid or uses a spoke/arm arrangement. A number of rods extend from the hub enhance the collection of atmospheric static electricity. The collected atmospheric electricity is conducted from the rods to an electrical connection where the electricity is conducted to earth via a conductive line.

16 Claims, 7 Drawing Sheets

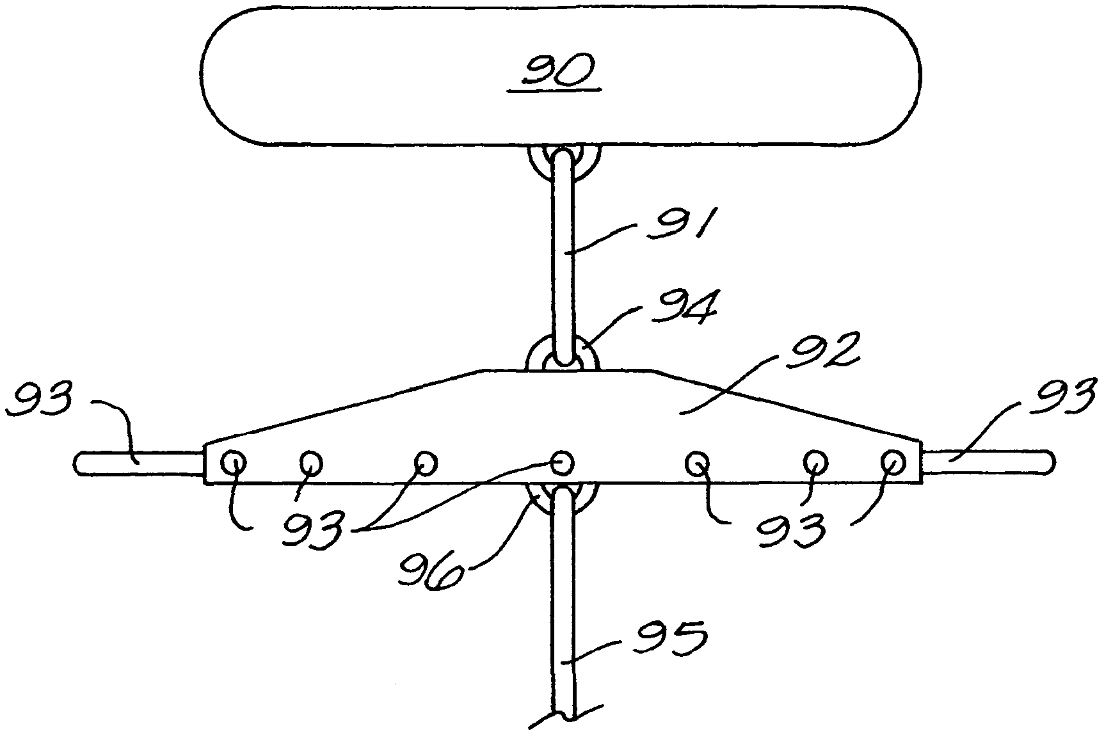

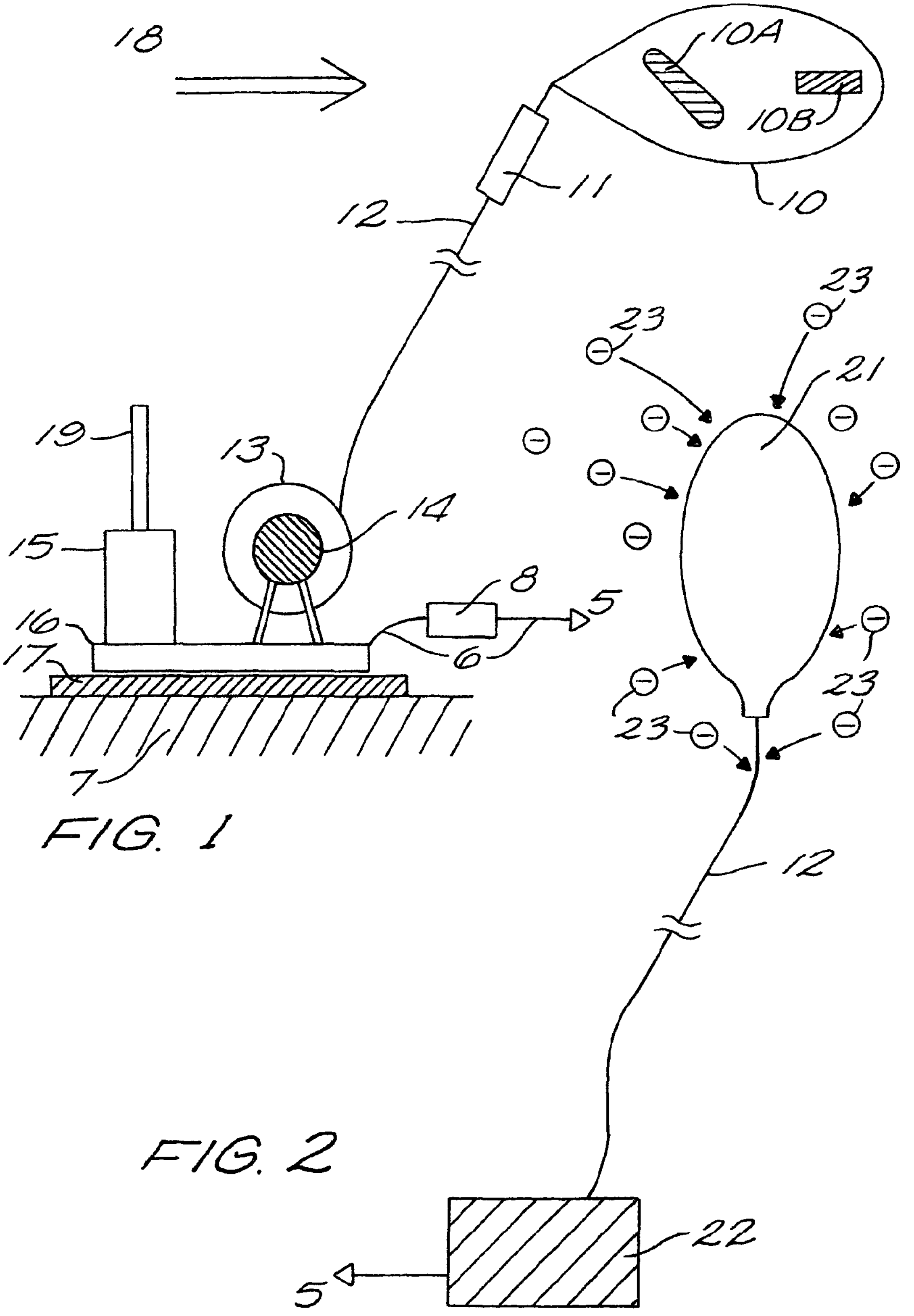
18
10A
10B
10
11
12
19
13
14
15
8
5
16
6
17
7
23
23
21
23
23
23
12
22
5

FIG. 1

FIG. 2

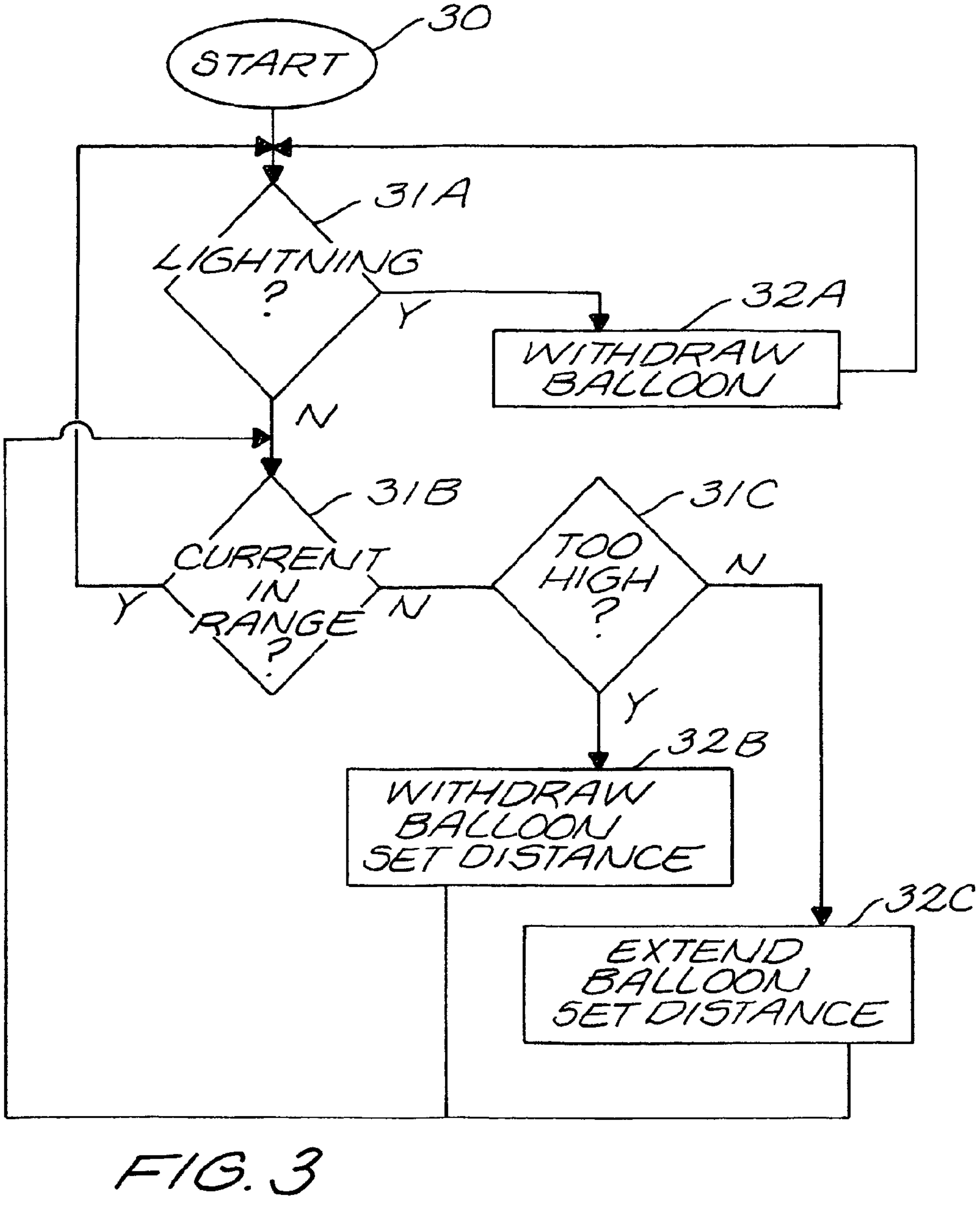
30
START
31A
LIGHTNING ?
Y
32A
WITHDRAW BALLOON
N
31B
CURRENT IN RANGE ?
Y
N
31C
TOO HIGH ?
N
Y
32B
WITHDRAW BALLOON SET DISTANCE
32C
EXTEND BALLOON SET DISTANCE

FIG. 3

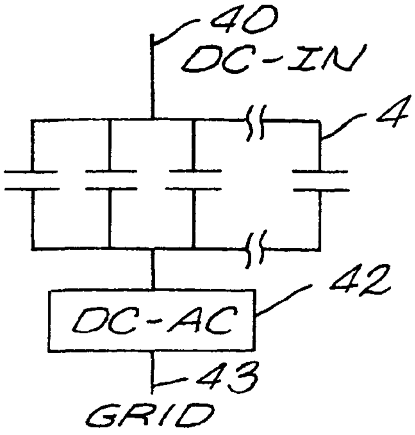
40
DC-IN
41
42
DC-AC
43
GRID

FIG. 4A

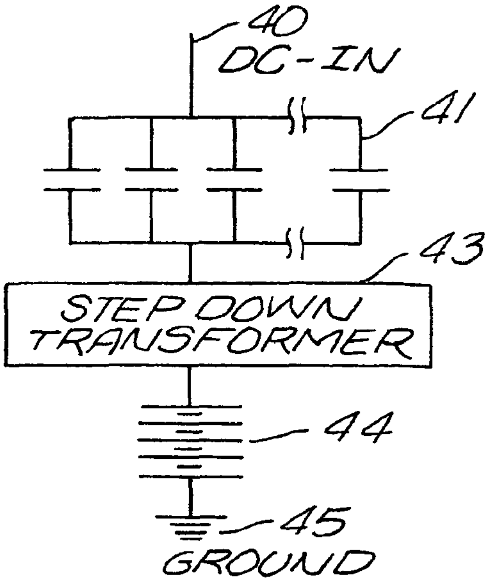
40
DC-IN
41
43
STEP DOWN TRANSFORMER
44
45
GROUND

FIG. 4B

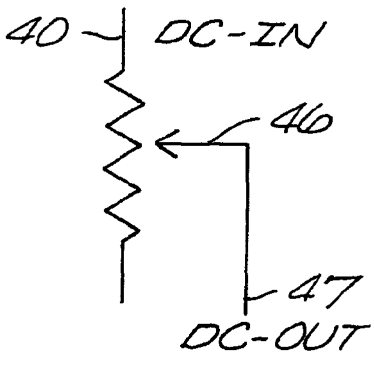
40
DC-IN
46
47
DC-OUT

FIG. 4C

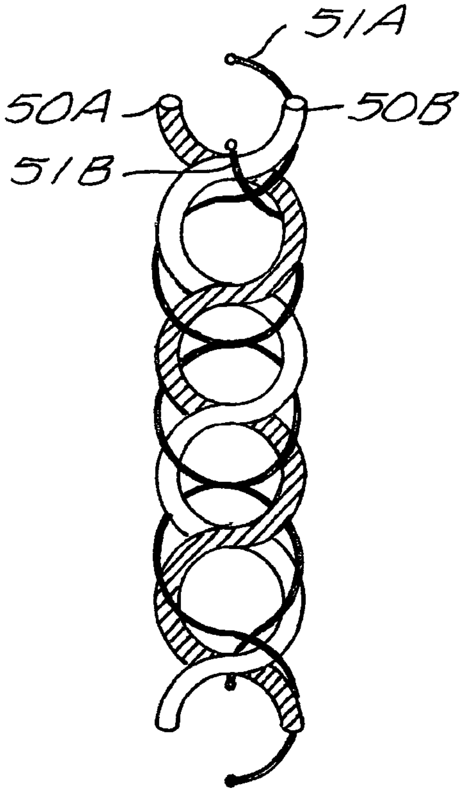
51A
50A
50B
51B

FIG. 5

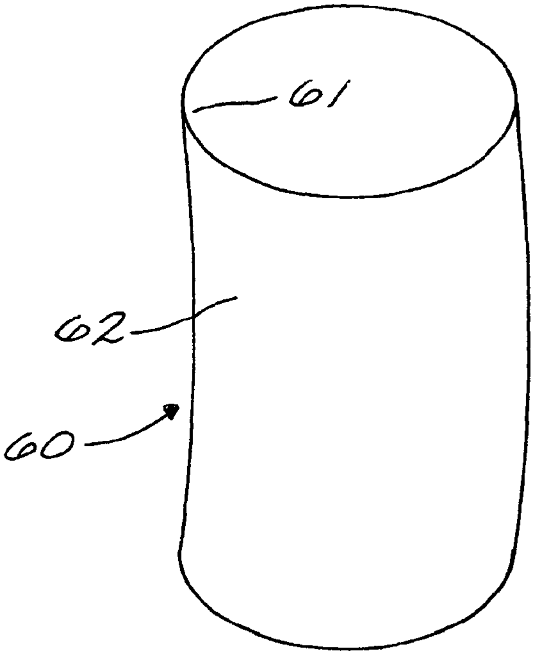

FIG. 6A

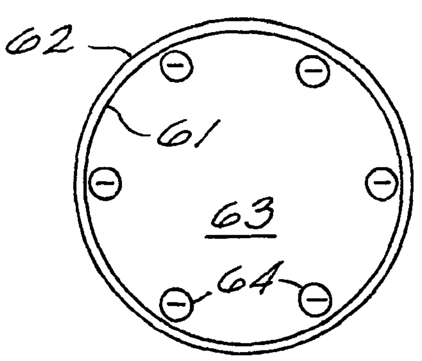

FIG. 6B

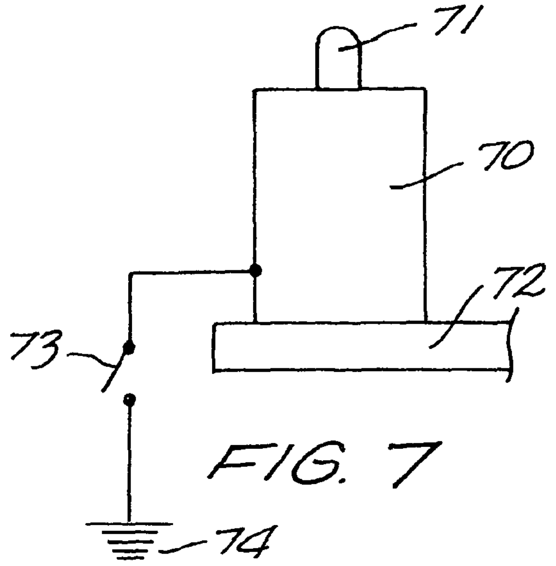
71
70
72
73
74

FIG. 7

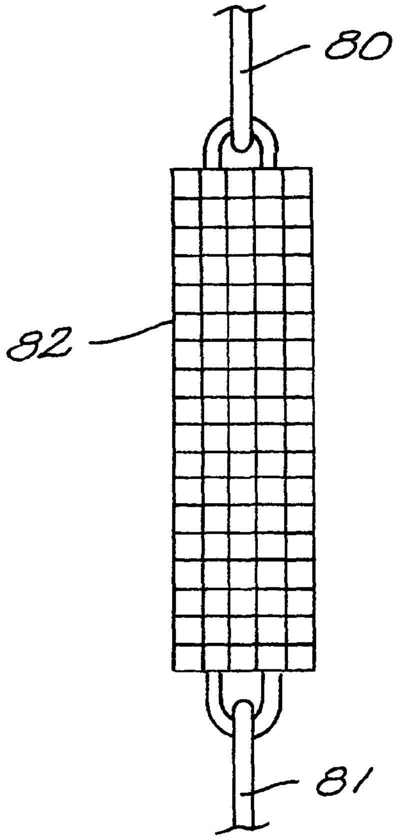

FIG. 8A

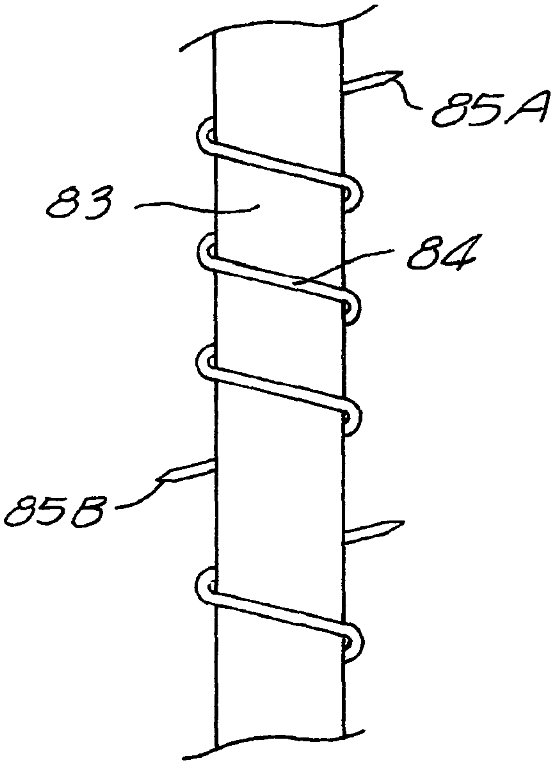
85A
83
84
85B

FIG. 8B

FIG. 9A

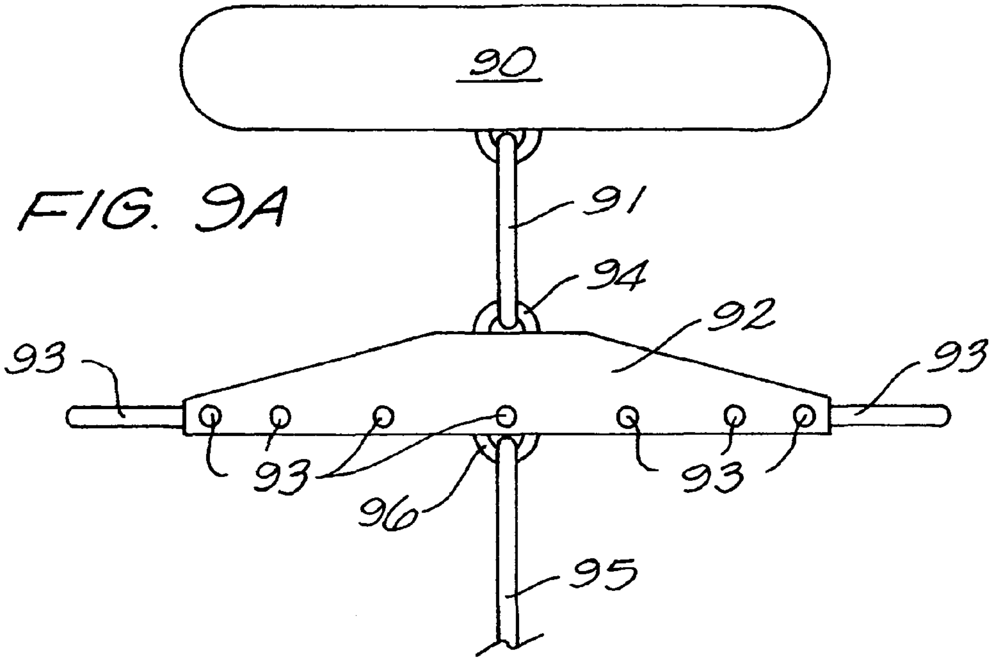
90
91
94
92
93
93
93
93
96
95

FIG. 9B

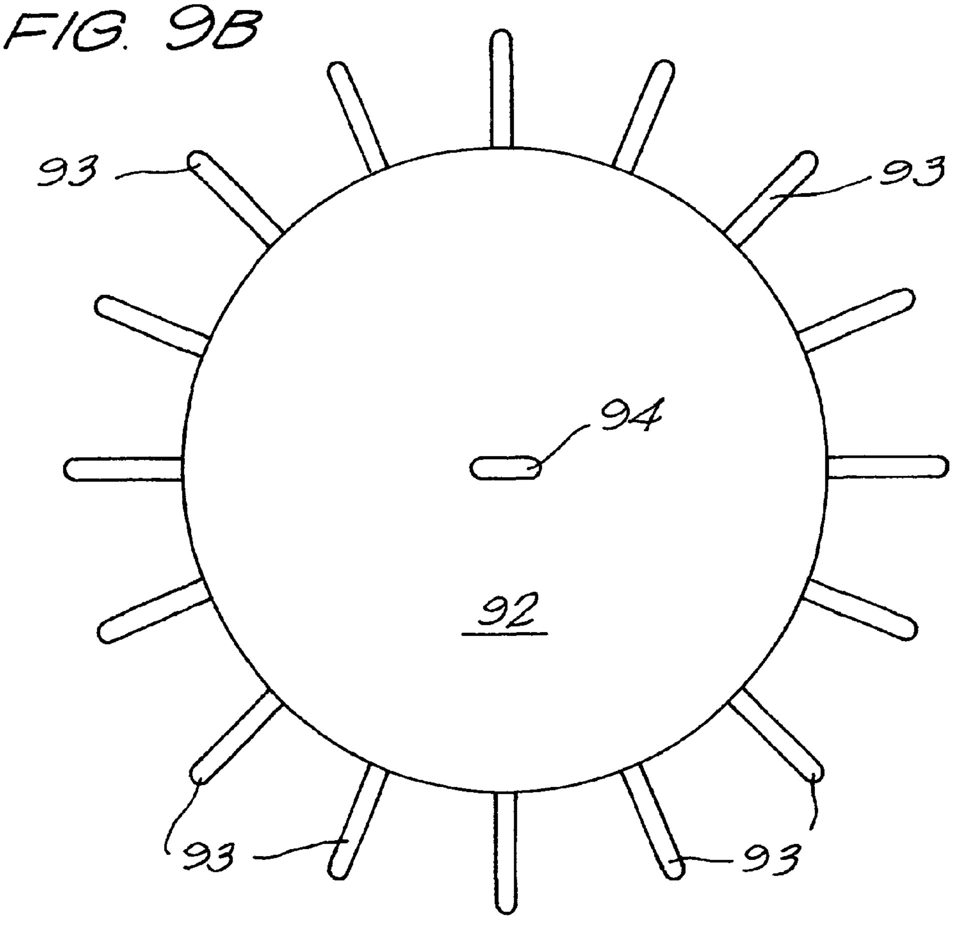
93
93
94
92
93
93
93
93

FIG. 10

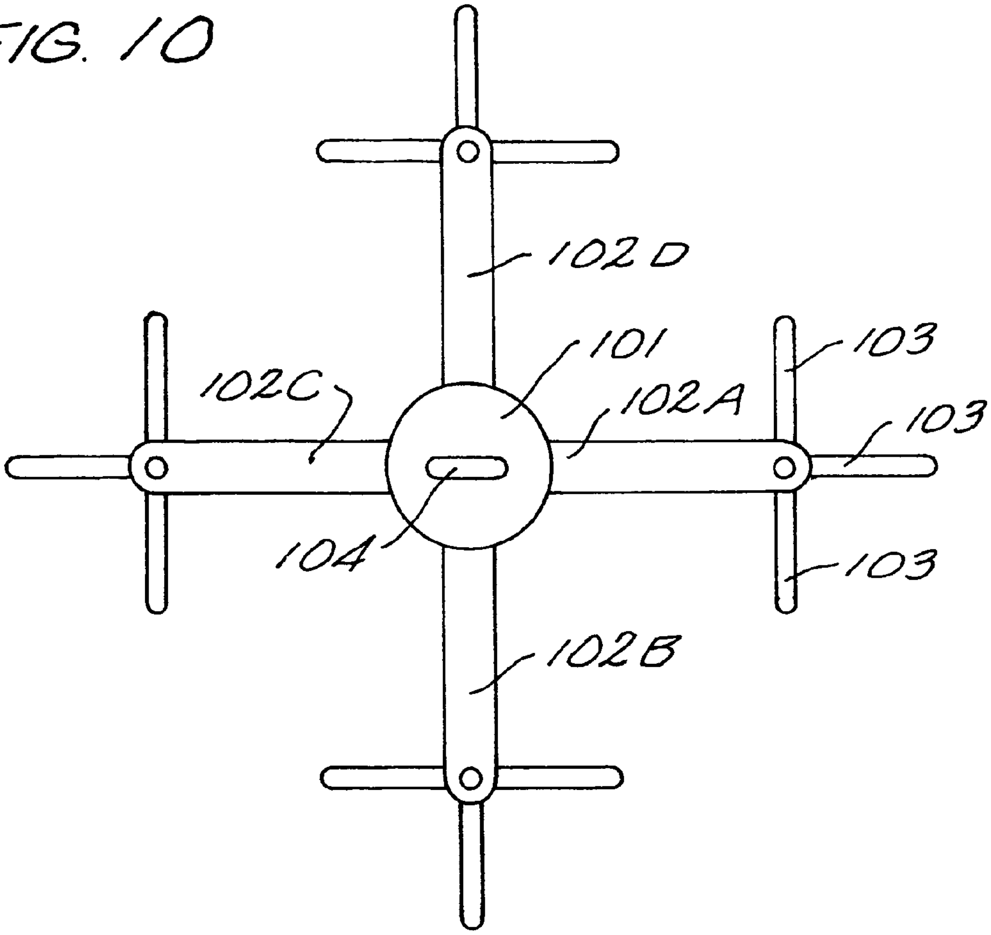
102D
101
103
102C
102A
103
104
103
102B

FIG. 11

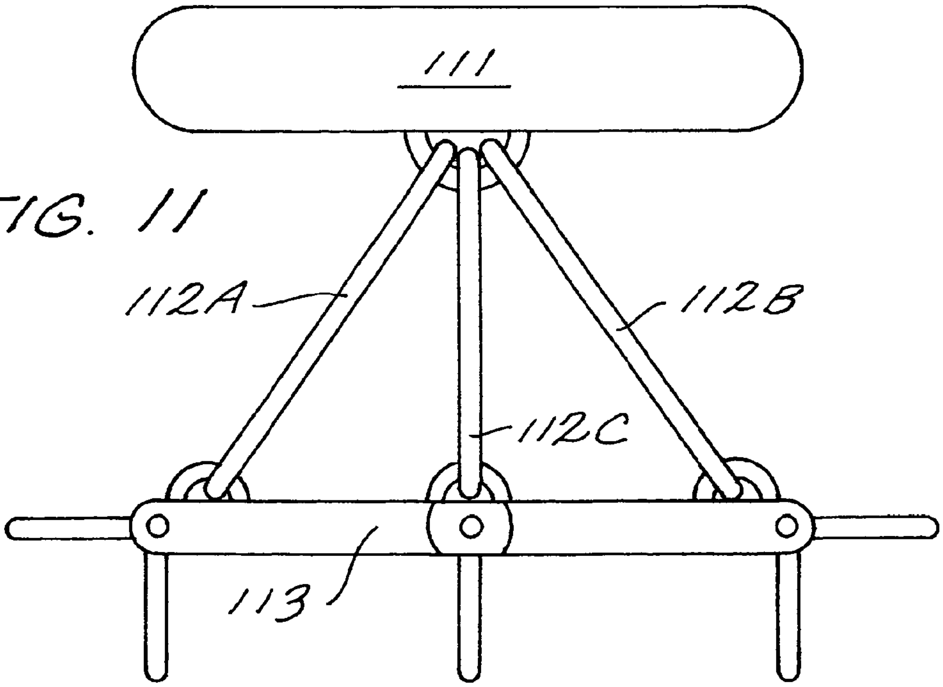
111
112A
112B
112C
113

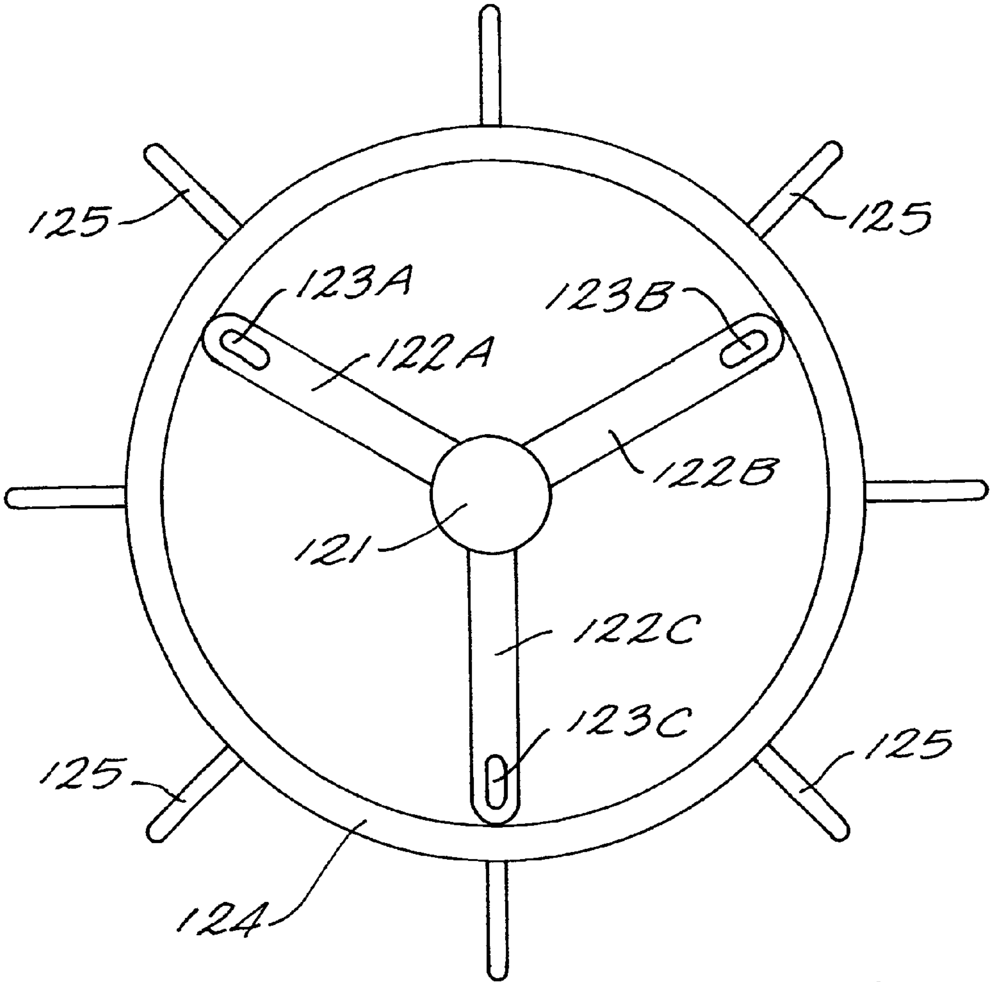
125
125
123A
123B
122A
122B
121
122C
123C
125
125
124

FIG. 12

ATMOSPHERIC STATIC ELECTRICITY COLLECTOR

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation and claims the benefit of priority under 35 U.S.C. §120 of U.S. patent application Ser. No. 12/321,306, filed Jan. 16, 2009, entitled "Atmospheric Static Electricity Collector", now U.S. Pat. No. 8,102,082, which is a continuation-in-part of U.S. patent application Ser. No. 12/218,297, filed Jul. 14, 2008, entitled "Atmospheric Electrical Generator", now U.S. Pat. No. 7,855,476, which the disclosures of the priority applications are incorporated by reference herein.

BACKGROUND OF THE INVENTION

This invention relates to the production of electrical energy and more particularly to the collection of static electricity from the atmosphere.

Everyone is familiar with Benjamin Franklin's kite experiment of 1752. Using a kite whose string had become wet, negative charges from the passing clouds flowed into the string, down to the suspended key, and then into a Leyden jar via a thin metal wire. Franklin was protected by a dry silk string; but, when Franklin's knuckle came too close to the key, he received a strong shock. Fortunately, Benjamin Franklin was not killed, others who tried this same experiment were not so lucky.

Since then, the formation of lightning has remained something of a mystery. Lightning bolts are triggered when a negatively charged cloud base induces a positive charge from the ground, thereby forming a "pathway" for the discharge of the collected electrical energy.

Lightning travels up to 60,000 miles per hour with a flash that is brighter than ten million 100-watt lightbulbs. This wattage is as much power as is produced by all of the electricity plants in the United States and with a voltage of up to 300 million volts.

It is this very fact, the power within lightning is immense, that has prevented any successful collection of the electrical energy from lightning. The electricity in lightning is far too extreme for current technology to harness.

While lightning has attracted a energy starved industrial world, no one has developed any technique to harness this naturally occurring electrical source.

It is clear there is a continuing need for an electrical source other than carbon-based fuels and that the naturally occurring electricity in the atmosphere is being ignored.

SUMMARY OF THE INVENTION

The invention is a mechanism which taps into the naturally occurring static electricity in the atmosphere. Whereas heretofore, the attempt to garner electricity from the atmosphere has focused exclusively on capturing lightning, the present invention syphons off the static electricity which is generated from any agitated air and voids lightning.

Lightning is only the final discharge of the static electricity, whether that lightning is intra-cloud lightning, cloud-to-ground lightning, or inter-cloud lightning. Other types of final discharges are known as heat lightning, summer lightning, sheet lightning, ribbon lightning, silent lightning, ball lightning, bead lightning, elves, jets, and sprites. Well before these discharges are observed, as the atmosphere becomes agitated by wind or thermal, static electricity is being generated.

The present invention recognizes that this static electricity is being formed and creates a mechanism to capture it.

The mechanism of this invention utilizes an aircraft such as a lighter than air balloon. While the preferred embodiment uses a foil balloon, a variety of other aircraft are obvious to those of ordinary skill in the art, including, but not limited to: gliders, rubber balloons (such as weather balloons), biaxially-oriented polyethylene terephthalate polyester film balloons, and latex balloons.

Within this discussion, the balloon is referenced, but, the invention is not intended to be limited solely to balloons.

The balloon is sent aloft and is tethered by a conductive line. In this context, the conductive line may be any obvious to those of ordinary skill in the art. For the preferred embodiment, the conductive line is a generically referred to as a "poly-rope" and is commercially available through a variety of sources. A suitable conductive line is described in U.S. Pat. No. 5,203,542, entitled "Apparatus for Improved Electric Fence Wire Construction for use with Intensive Grazing" issued Apr. 20, 1993, to Coley, et al. and incorporated hereinto by reference.

The conductive line is played out of a winch to control the altitude of the balloon. The motor controlling the winch is able to reverse direction to both extend and withdraw the conductive line which is wrapped around a spool on the winch. The winch/spool combination are part of a base unit.

In some embodiments of the invention, the spool is constructed of rubber so as to insulate the conductive line from the winch assembly. In this embodiment, only the conductive line is charged by the atmospheric static electricity while the winch remains neutral.

In yet another embodiment, the winch/spool are part of a base unit which is itself isolated from the ground by an insulator. In this embodiment, the entire base unit is charged by the atmospheric static electricity.

A conductor, such as an insulated wire, is electrically connected to the conductive line. In one embodiment, where the conductive line is electrically isolated from the spool and winch motor, the conductor is connected to the conductive line. In the embodiment where the conductive line is electrically connected to the base unit, then the conductor is connected anywhere on a metallic base unit.

The other end of the conductor is connected to a load. The load in this case can be any of a variety of electrical loads well known to those of ordinary skill in art, including, but not limited to a motor, a battery system, or the electrical grid for the system.

In the preferred embodiment, a sensor array is used to monitor the activities both at the base unit (such as electrical flow within the conductor) and in the surrounding locale.

A sensor monitoring the electrical flow (i.e. voltage and/or current) within the conductor is used to monitor the electrical activity within the conductor.

In the preferred embodiment, a lightning sensor monitors for lightning activity within the locale. As noted earlier, the electrical characteristic of lightning is so extreme that ideally this discharge is avoided as it might damage the mechanism of this invention.

The sensor array is utilized by a controller, such as microprocessor, programmed to operate the mechanism as outlined herein.

The controller operates the winch motor to extend or withdraw the conductive line and by extension the altitude of the balloon. The controller is programmed to operate the winch by monitoring the electrical characteristics of the conductor and adjusting the balloon's altitude to maintain these characteristics within the conductor within a preset range.

This preset range is established either in the base programming of the controller or is established by an operator of the system.

As example, by controlling the amount of current being withdrawn from the atmosphere, the mechanism operates within a safe range and also provides a relatively stable current flow from which a variety of activities can take place (such as DC-AC conversion).

The controller also utilizes the lightning sensor to protect the mechanism from a lightning strike. Should lightning be detected within a pre-determined range (as established by the software or defined by an operator), then the balloon is pulled down to minimize the risk of damage from a lightning strike.

An aspect of the present invention is the use of an antenna which are used to collect the atmospheric static electricity. The antenna is shaped as a hub which is suspended from the blimp/balloon. The hub is ideally spoked shaped although an alternative embodiment uses a solid hub.

A number of rods extend from the hub so as collect atmospheric static electricity. These rods are ideally rounded at the ends to enhance the attraction of the atmospheric static electricity.

The collected atmospheric electricity is conducted from the rods to an electrical connection on the hub where the electricity is conducted to a power plant on earth such as described above.

The invention, together with various embodiments thereof will be more fully explained by the following description of the accompanying drawings.

DRAWINGS IN BRIEF

FIG. **1** diagrams the preferred embodiment of the invention.

FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

FIGS. **9**A and **9**B are side views and top views of an embodiment of the antenna used to collected atmospheric electricity.

FIG. **10** is a top view of an alternative embodiment of the antenna of this invention.

FIG. **11** is a side view of yet another alternative embodiment of the antenna used to collect atmospheric electricity.

FIG. **12** is the preferred embodiment of the antenna of this invention.

DRAWINGS IN DETAIL

FIG. **1** diagrams the preferred embodiment of the invention.

Balloon **10** is an aircraft which, in this illustration, is a lighter than air balloon. Wings **10**A, extending from the body of balloon **10**, provide additional lift in air flow **18**. Tail **10**B helps to stabilize balloon **10**.

Balloon **10** is tethered to the ground via conductive line **12**. As noted earlier, a variety of configurations and materials are available to serve as conductive line **12**. In this illustration, a poly-wire is used. Poly-wire is commercially available through a variety of vendors, including, but not limited to: Jeffers Livestock and Sareba Systems, Inc. of Ellendale, Minn.

In this embodiment, located proximate to balloon **10**, is an electrical collection enhancement lead **11** which assists in the collection of the static electrical charge in the atmosphere. Electrical collection enhancement lead **11** is configured to attract the static charge and conduct the charge into the conductive line **12**.

The electricity flows down the conductive line into spool **13**, where the conductive line **12** is collected and either withdrawn or dispensed through operation of winch motor **14**.

Winch motor **14** and spool **13** are mounted onto base unit **16** which is electrically isolated from ground **7** using insulator **17**. Note, in this embodiment of the invention, when electricity is being collected from the atmosphere, the entire base unit **16** becomes charged. In another embodiment of the invention, spool **13** is constructed of rubber, thereby preventing base unit **16** from becoming charged, thereby restricting the charging from the atmosphere to only conductive line **12**.

In this embodiment, conductor **6** is connected to base unit **16** (since the entire base unit **16** is charged and the base unit is metallic) to communicate the electrical current to load **5**. Conductor **6** is ideally an insulated wire.

The electrical current through conductor **6** is measured using sensor **8**.

In the alternative embodiment discussed above, where only the conductive line **12** is charged, then conductor **6** is connected to conductive line **12**.

Controller **15**, located in this embodiment on base unit **16**, operates winch motor **14** in response to signals from sensor **8** (measuring the current being discharged to load **5**) to maintain the current flow within a pre-defined range. As the current flow diminishes, then the conductive line **12** extended from spool **13** to increase the altitude of balloon **10** to that more static charge from the atmosphere is gathered; as the current flow falls exceeds a preset level, conductive line **12** is withdrawn onto spool **13** to decrease the static charge being collected from the atmosphere.

The range of current flow through conductor **6** is ideally set by the program, although some embodiments of the invention permit an operator to establish this range of operation.

In an alternative embodiment, the sensor monitoring conductor **6** monitors the voltage therein.

In the preferred embodiment of the invention, controller **15** is also equipped with a lightning sensor **19**. In this embodiment, when lightning is sensed within a preset range, then substantially all of conductive line **12** is wound onto spool **13** to pull balloon **10** near the ground and protect the entire mechanism from being damaged from a lightning discharge.

In the preferred embodiment, the “safe” distance form lightning is set in the programming of controller **15** and is ideally two miles; other embodiments permit the operator to “safe” distance.

There are a variety of lightning sensors well known to those of ordinary skill in the art, including, but not limited to those described in: U.S. Pat. No. 7,016,785, entitled “Lightning Detection” issued to Makela, et al. on Mar. 21, 2006; U.S. Pat. No. 6,829,911, entitled “Lightning Detection and Prediction Alarm Device” issued to Jones, et al. on Dec. 7, 2004; U.S. Pat. No. 7,200,418, entitled “Detection of Lightning” issued to Karikuranta, et al. on Apr. 3, 2007; and U.S. Pat. No. 6,961,662, entitled “Systems and Methods for Spectral Corrected Lightning Detection" issued to Murphy on Nov. 1, 2005; all of which are incorporated hereinto by reference.

In another embodiment of the invention, controller **15** is not located on base unit **16**, rather it is remote and communicates its control signals to winch motor **14** using radio waves.

FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

Static charges **23** are generated in the atmosphere by agitated air. These static charges are often collected at the bottom of clouds, but exist in other environments as well.

Balloon **21** is extended into this strata of static charges **23** which are then attracted to conductive line **12** to flow to base unit **22** and then onto load **5**.

By increasing or decreasing the altitude of balloon **21** (defined by the length of the extended conductive line **12**), conductive line **12** is selectively exposed to varying densities and levels of the static charge strata, and by extension, the current flow or voltage is increased or decreased.

FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

Once the program starts **30**, the lightning sensor is checked to determine if lightning has occurred within the unsafe range **31**A, if it has, then the balloon is lowered **32**A, and the program continues monitoring the status of lightning until no lightning is detected.

When the lightning status is acceptable, then the current within the conductor is checked to see if the current is within the prescribed range **31**B. If the current is acceptable (within range) the program returns to check the lightning status **31**A; otherwise a determination is made to see if the current is above the prescribed range **31**C.

If the current is above the prescribed range, then the altitude of the balloon is withdrawn a set amount **32**B (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

If the current is below the prescribed range, then the altitude of the balloon is extended a set amount **32**C (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

In this manner of feed-back and minor adjustments in the altitude of the balloon, the current is maintained within a prescribed range which can be handled by the downstream electrical system.

As noted earlier, some embodiments of the invention monitor the voltage instead of the current.

FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

By maintaining the voltage being collected in a prescribed range, an electrical conversion system is easily designed. While FIGS. **4**A, **4**B, and **4**C illustrate some electrical configurations, those of ordinary skill in the art readily recognize a variety of other configurations which will serve the same function.

Referencing FIG. **4**A, Direct Current In (DC IN) **40** is buffered by a gang of capacitors **41** before being communicated to a DC/AC converter **42**. The DC/AC converter converts the direct current into a an alternating current suitable for placement over an existing electrical grid **43** such as normally found from a power-plant.

Those of ordinary skill in the art readily recognize a variety of DC/AC converters, including, but not limited to: U.S. Pat. No. 7,394,671, entitled "Controller IC, DC-AC Conversion Apparatus, and parallel running system of DC-AC Conversion Apparatuses" issued to Fukumoto, et al. on Jul. 1, 2008; and, U.S. Pat. No. 7,330,366, entitled "DC-AC Converter" issued to Lee, et al. on Feb. 12, 2008; all of which are incorporated hereinto by reference.

FIG. **4**B illustrates an electrical arrangement suitable for use in charging a battery. DC IN **40** is buffered by capacitor bank **41** before entering into a step down transformer **43**. Step down transformer **43** reduces the voltage so that the voltage can safely be introduced into battery **44** which is connected to ground **45** at the battery's other pole.

Those of ordinary skill in the art readily recognize a variety of batteries which will work in this capacity, including, but not limited to those described in: U.S. Pat. No. 7,378,181, entitled "Electric Storage Battery Construction and Manufacture" issued to Skinlo on May 27, 2008; U.S. Pat. No. 7,388,350, entitled "Battery with Electronic Compartment" issued to Wright on Jun. 17, 2008; U.S. Pat. No. 7,397,220, entitled "Connection Member and Battery Pack" issued to Uchida, et al. on Jul. 8, 2008; and, U.S. Pat. No. 7,375,492, entitled "Inductively Charged Battery Pack" issued to Calhoon, et al. on May 20, 2008; all of which are incorporated hereinto by reference.

In FIG. **4**C, DC IN **40** is fed into an adjustable rheostat **46** which is controlled by the controller so that the DC OUT **47** falls within a specified range.

FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

This type of conductive line is commonly called poly-wire and consists of multiple interwoven strands of plastic **50**A and **50**B woven into a cord or rope arrangement having intertwined therein exposed metal wires **51**A and **51**B. While this illustration shows two plastic strands and two metal wires, any number of possible combinations is possible.

The exposed metal wires **51**A and **51**B attract the atmospheric static charge and transmit the charge down to the base unit (not shown).

FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

This conductive line utilizes a tube **60** having an outer layer **62** of PET Film (Biaxially-oriented polyethylene terephtalate polyester film) which provides exceptionally high tensile strength and is chemically and dimensionally stable. The tube has an ideal diameter of between two and three inches.

An interior metal coating **61** provides an initial conduit for the flow of static charge. The static charge through the metal forces the tube to expand due to the repulsion experienced by like charges. Further, the flow of electricity causes the interior of the tube **60** to become ionized to provide an additional pathway for the atmospheric static charges to the base unit (not shown).

Because outer layer **62** provides a gas barrier, the resulting ionization is not dissipated by air currents, thereby providing a highly stable pathway.

FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

In this embodiment, controller box **70**, resting on insulating pad **72**, is in communication with the sensors as described above. Using the input from these sensors, when there is flow of electricity through the base unit, warning flashing light **71** is illuminated. To electrically neutralize the mechanism, switch **73** is activated to pass any existing current into the ground **74**.

FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

Referencing FIG. **8**A, enhanced electrical collection lead **82** is a wire mesh which is in electrical communication with conductive line **81** and balloon **80**. Because of the significant amount of metal exposed by enhanced electrical collection lead **82**, more static electricity from the atmosphere is drawn to the collection lead **82**, and then down conductive line **81** to the base unit (not shown).

Conductive lead **82** is positioned proximate to balloon **80**.

In FIG. **8**B, poly-wire **83** has enhanced electrical collection leads **84** wrapped therearound. Collection leads **84** have pointed ends **85**A and **85**B which have a propensity to attract more electricity than rounded ends do.

FIGS. **9**A and **9**B are side views and top views of an embodiment of the antenna used to collected atmospheric electricity.

Blimp **90** supports antenna **92** via tether **91** which is attached to antenna **92** by connection **94**. In this embodiment, antenna **92** is made of electrically conductive material and is ideally light in weight to lessen the payload requirements on blimp **90**.

Encircling antenna **92** and extending therefrom are rods **93** which are rounded at their distal ends so as to enhance attraction of the atmospheric static electricity. The static electricity is communicated to connection **96** and then to electrical line **95** which communicates the electricity to the ground based station (not shown) as described above.

FIG. **10** is a top view of an alternative embodiment of the antenna of this invention,

In this embodiment of the antenna a central hub **101** has the blimp connection **104** (capable of receiving the tether to the blimp) secured thereto. In this illustration, four arms **102**A, **102**B, **102**C, and **102**D extend from hub **101**. Hub **101**, and arms **102**A, **102**B, **102**C, and **102**D are all electrically conductive in this embodiment.

At the end of each arm are rods **103** which are used to enhance the collection of the atmospheric static electricity.

Although this embodiment illustrates four arms, the invention is not intended to be limited to four arms, rather, any number of arms may be used and the number of rods extending from the distal ends of the arms also varies.

FIG. **11** is a side view of yet another alternative embodiment of the antenna used to collect atmospheric electricity.

In this embodiment, several tethers **112**A, **112**B, and **112**C, are used to secure the antenna **113** to the blimp **111**. This arrangement of several tethers provides heightened stability of the antenna by reducing the affects wind will have on the antenna.

FIG. **12** is the preferred embodiment of the antenna of this invention.

In this embodiment of the antenna, arms **122**A, **122**B, and **122**C extend from a central hum and are electrically connected to rim **124**. Tether connectors **123**A, **123**B, and **123**C, are used to secure the antenna to the blimp or balloon.

Rods **125** extend from rim **124** to increase the collection of the static charges in the atmosphere.

It is clear from the foregoing that the present invention captures an entirely new source of electrical energy.

What is claimed is:

1. A system for collection of atmospheric static electricity comprising:

a balloon;

a hub that is electrically conductive and connected to the balloon by a tether, the hub having:

two or more arms that are electrically conductive, a first end of each of the two or more arms electrically connected to the hub; and

a plurality of rods, a first end of each of the rods being electrically connected to a second end of at least one of the two or more arms;

a base unit having a spool of conductive line on a winch motor, one end of the conductive line secured to the hub, a portion of the conductive line and the hub collecting electricity in the atmosphere, the winch motor selectively extending or withdrawing the conductive line from the spool;

a conductor having a first end electrically connected to the conductive line and a second end electrically connected to a load being powered by collected electricity from the conductive line;

an electrical flow sensor monitoring electrical flow through the conductor and generating an electrical flow indicia indicative of the electrical flow in the conductor; and

a controller to operate the winch motor to maintain the electrical flow indicia within a selected operating range.

2. The system according to claim **1**, wherein each of the plurality of rods has a rounded second end.

3. The system according to claim **1**, wherein at least two rods are attached to the second end of each of the two or more arms.

4. The system according to claim **1**, further comprising an electrically conductive rim connected to a second end of the two or more arms.

5. The system according to claim **4**, wherein at least a portion of the two or more electrically conductive arms are connected substantially at right angles to an exterior of the electrically conductive rim.

6. A system for collection of atmospheric static electricity comprising:

two or more arms that are electrically conductive;

a plurality of rods, each of the plurality of rods being electrically connected to at least one of the two or more arms; and

a balloon tether connected between a balloon and the two or more arms;

a base unit having a spool of conductive line on a winch motor, one end of the conductive line secured to the two or more arms, a portion of the conductive line and the two or more arms collecting electricity in the atmosphere, the winch motor selectively extending or withdrawing the conductive line from the spool;

a conductor having a first end electrically connected to the conductive line and a second end electrically connected to a load being powered by collected electricity from the conductive line;

an electrical flow sensor monitoring electrical flow through the conductor and generating an electrical flow indicia indicative of the electrical flow in the conductor; and

a controller to operate the winch motor to maintain the electrical flow indicia within a selected operating range.

7. The system according to claim **6**, wherein each of the plurality of rods has a rounded end distal from the electrically conductive arm.

8. The system according to claim **7**, further including a balloon tether connection secured to the first end of the two or more arms.

9. The system according to claim **8**, further including an electrical connection secured to the first end of each of the two or more arms.

10. The system according to claim **9**, wherein the at least two rods attached to the two or more arms includes three or more rods.

11. The system according to claim **9**, further including an electrically conductive rim connected to a second end of the two or more arms.

12. The system according to claim **11**, wherein the plurality of arms are electrically connected to the electrically conductive rim.

13. The system according to claim **12**, wherein at least a portion of the plurality of rods are connected to an exterior of the electrically conductive rim substantially at right angles thereto.

14. A system for the collection of atmospheric static electricity comprising:

an electrically conductive body;

a balloon tether connected to a first side of the electrically conductive body and to a balloon; and

an electrical connection connected to a second side of the electrically conductive body;

a base unit having a spool of conductive line on a winch motor, one end of the conductive line secured to the electrical connection, a portion of the conductive line and the electrically conductive body collecting electricity in the atmosphere, the winch motor selectively extending or withdrawing the conductive line from the spool;

a conductor having a first end electrically connected to the conductive line and a second end electrically connected to a load being powered by collected electricity from the conductive line;

an electrical flow sensor monitoring electrical flow through the conductor and generating an electrical flow indicia indicative of the electrical flow in the conductor; and

a controller receiving the electrical flow indicia and selectively operating the winch motor such that the electrical flow indicia remains within a selected operating range.

15. The system according to claim **14**, wherein the electrically conductive body comprises:

an electrically conductive hub;

at least three electrically conductive arms, each of the arms connected at a first end to the hub; and

an electrically conductive rim connected to the second end of each of the at least three arms.

16. The system according to claim **15**, further comprising a plurality of electrically conductive rods, a first end of each of the rods connected to the rim and extending therefrom substantially at rights angles thereto.

* * * * *

US008686575B2

(12) United States Patent
McCowen

(10) **Patent No.: US 8,686,575 B2**
(45) **Date of Patent: Apr. 1, 2014**

(54) **ENERGY COLLECTION**

(75) Inventor: **Clint McCowen**, Navarre, FL (US)

(73) Assignee: **Ion Power Group, LLC**, Navarre, FL (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **12/255,130**

(22) Filed: **Oct. 21, 2008**

(65) **Prior Publication Data**

US 2009/0040680 A1 Feb. 12, 2009

Related U.S. Application Data

(63) Continuation of application No. 11/358,264, filed on Feb. 21, 2006, now Pat. No. 7,439,712.

(51) **Int. Cl.**

F02B 63/04	(2006.01)
F03G 7/08	(2006.01)
H02K 7/18	(2006.01)
H02N 1/00	(2006.01)

(52) **U.S. Cl.**
USPC .... **290/1 R**; 322/2 A

(58) **Field of Classification Search**
USPC .... 290/1 R; 322/2 A; 126/660; 60/641.8, 60/641.11, 641.12, 659, 698; 244/158.2
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427	A	5/1901	Palencsar	307/149
911,260	A	2/1909	Pennock	307/149
1,014,719	A	1/1912	Pennock	307/149
2,473,819	A *	6/1949	Pittman	315/36
3,532,959	A *	10/1970	Erickson	320/166
3,780,722	A *	12/1973	Swet	126/680
3,946,227	A	3/1976	Bingham	250/281
4,079,225	A *	3/1978	Warner	219/110
4,092,250	A	5/1978	Sano et al.	210/500.29
4,104,696	A	8/1978	Cochran, Jr.	361/229
4,146,800	A *	3/1979	Gregory et al.	290/44
4,180,056	A *	12/1979	Schnabel et al.	126/660
4,201,197	A *	5/1980	Dismer	126/600
4,206,396	A *	6/1980	Marks	322/2 A
4,224,496	A *	9/1980	Riordan et al.	219/110
4,307,936	A *	12/1981	Ochiai	385/120
4,314,192	A *	2/1982	Caro	322/2 R
4,346,478	A *	8/1982	Sichling	398/107
4,425,905	A *	1/1984	Mori	126/578
4,433,248	A *	2/1984	Marks	290/44
RE31,678	E *	9/1984	Ochiai	385/115

(Continued)

FOREIGN PATENT DOCUMENTS

WO	2007098341	8/2007

OTHER PUBLICATIONS

"Emissivity Values for Common Materials & Non-Metals", Raytek—A Fluke Company, http://www.raytek.com/Raytek/en-r0/IREducation/EmissivityNonMetals.htm, 1999.*

Primary Examiner — Pedro J Cuevas
(74) *Attorney, Agent, or Firm* — Benjamin A. Balser; Balser & Grell IP Law

(57) **ABSTRACT**

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support wire system supported by poles. The support wire system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but carbon and graphite are preferred. Diodes may be used to restrict the backflow or loss of energy.

25 Claims, 10 Drawing Sheets

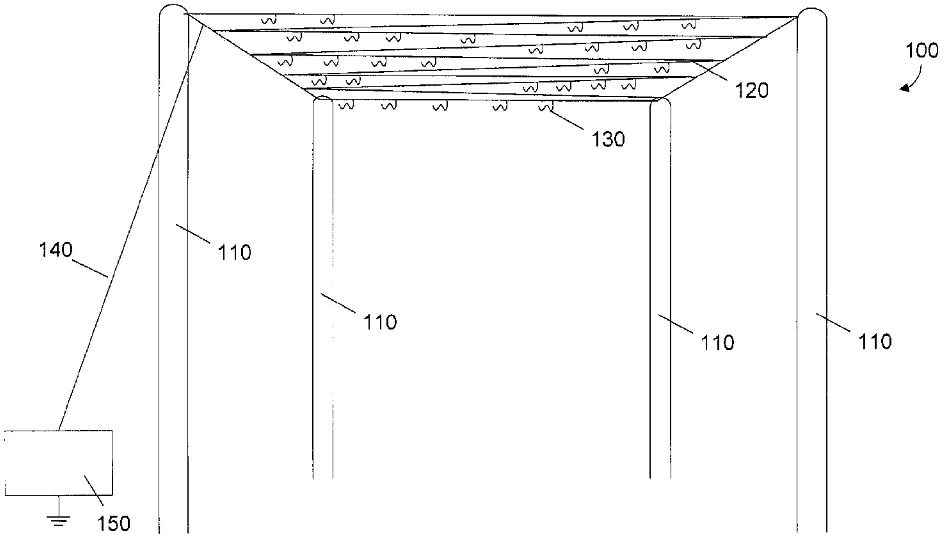

(56) References Cited

U.S. PATENT DOCUMENTS

4,483,311	A *	11/1984	Whitaker	126/602
4,489,269	A *	12/1984	Edling et al.	322/2 R
4,512,335	A *	4/1985	Mori	126/678
4,653,223	A *	3/1987	Mori	47/17
4,653,472	A *	3/1987	Mori	126/683
4,676,226	A *	6/1987	Mori	126/680
4,809,675	A *	3/1989	Mori	126/578
4,852,454	A *	8/1989	Batchelder	89/1.11
4,943,125	A *	7/1990	Laundre' et al.	385/35
4,988,159	A *	1/1991	Turner et al.	385/33
5,047,892	A	9/1991	Sakata et al.	361/231
5,114,101	A *	5/1992	Stern et al.	244/172.8
5,145,257	A *	9/1992	Bryant et al.	374/131
5,379,103	A *	1/1995	Zigler	356/73
5,851,309	A *	12/1998	Kousa	136/248
5,942,806	A *	8/1999	Veliadis	290/1 R
6,025,591	A	2/2000	Taylor et al.	250/292
6,038,363	A *	3/2000	Slater et al.	385/147
6,116,544	A *	9/2000	Forward et al.	244/158.2
6,173,922	B1 *	1/2001	Hoyt et al.	244/158.2
6,226,440	B1 *	5/2001	Lyons	385/119
6,419,191	B1 *	7/2002	Hoyt et al.	244/158.2
6,425,391	B1 *	7/2002	Davoren et al.	126/683
6,735,830	B1	5/2004	Merciel	250/324
6,765,212	B2 *	7/2004	Goetz et al.	250/339.11
6,846,447	B2 *	1/2005	Mbachu et al.	264/406
6,853,447	B2 *	2/2005	Goetz	356/237.1
6,894,772	B2 *	5/2005	Goetz et al.	356/237.1
6,925,852	B2 *	8/2005	Susko	73/23.2
6,974,110	B2	12/2005	Grandics	244/171.5
7,109,597	B1 *	9/2006	Bose	290/43
7,128,867	B2 *	10/2006	Mbachu et al.	264/406
7,231,809	B2 *	6/2007	Susko	73/31.05
7,259,475	B2 *	8/2007	Hong et al.	307/64
7,439,712	B2 *	10/2008	McCowen	322/2 A
7,449,668	B2 *	11/2008	Schutten et al.	250/214 SW
7,478,712	B2 *	1/2009	McCowen	191/45 R
7,532,819	B1 *	5/2009	Triebes et al.	398/121
2002/0026933	A1 *	3/2002	Gottlieb	126/628
2002/0108892	A1 *	8/2002	Goetz et al.	209/576
2002/0109094	A1 *	8/2002	Goetz et al.	250/339.11
2002/0109835	A1 *	8/2002	Goetz	356/237.1
2003/0107869	A1	6/2003	Pan et al.	361/306.2
2003/0125630	A1 *	7/2003	Furnish	600/461
2003/0193319	A1 *	10/2003	Wood et al.	322/2 A
2004/0011925	A1 *	1/2004	Grandics	244/62
2004/0083793	A1 *	5/2004	Susko	73/31.05
2004/0094853	A1 *	5/2004	Mbachu et al.	264/40.1
2004/0160711	A1	8/2004	Stumberger	361/20
2005/0101024	A1 *	5/2005	Mbachu et al.	436/85
2005/0136311	A1 *	6/2005	Ueda et al.	429/30
2005/0270525	A1 *	12/2005	Susko	356/326
2007/0107765	A1 *	5/2007	Schutten et al.	136/243
2007/0195481	A1 *	8/2007	McCowen	361/233
2007/0218323	A1 *	9/2007	Sudo et al.	429/13
2007/0223171	A1	9/2007	Guy et al.	361/126
2007/0273206	A1 *	11/2007	McCowen	307/9.1
2007/0297892	A1 *	12/2007	Kildegaard	415/1
2008/0145714	A1 *	6/2008	Kagami	429/13
2009/0107842	A1	4/2009	Park et al.	204/554
2009/0114495	A1 *	5/2009	McCowen	191/45 R
2009/0216292	A1 *	8/2009	Pless et al.	607/33
2010/0090562	A1 *	4/2010	McCowen	310/309
2010/0090563	A1 *	4/2010	McCowen	310/309
2012/0164555	A1 *	6/2012	Yamazaki et al.	429/482
2012/0299559	A1 *	11/2012	McCowen	322/2 A
2013/0276776	A1 *	10/2013	Te Kamp et al.	126/660

* cited by examiner

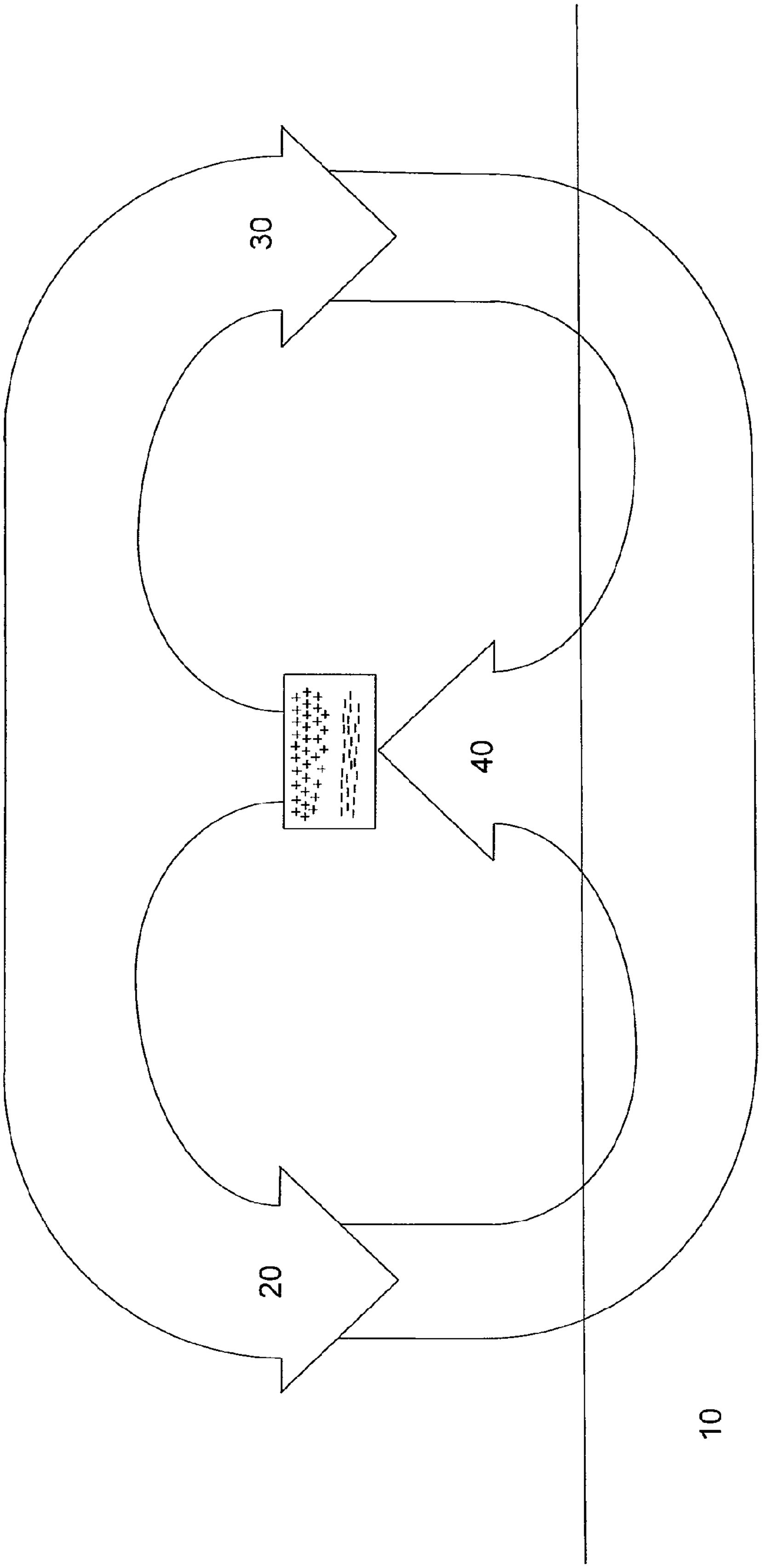

FIGURE 1

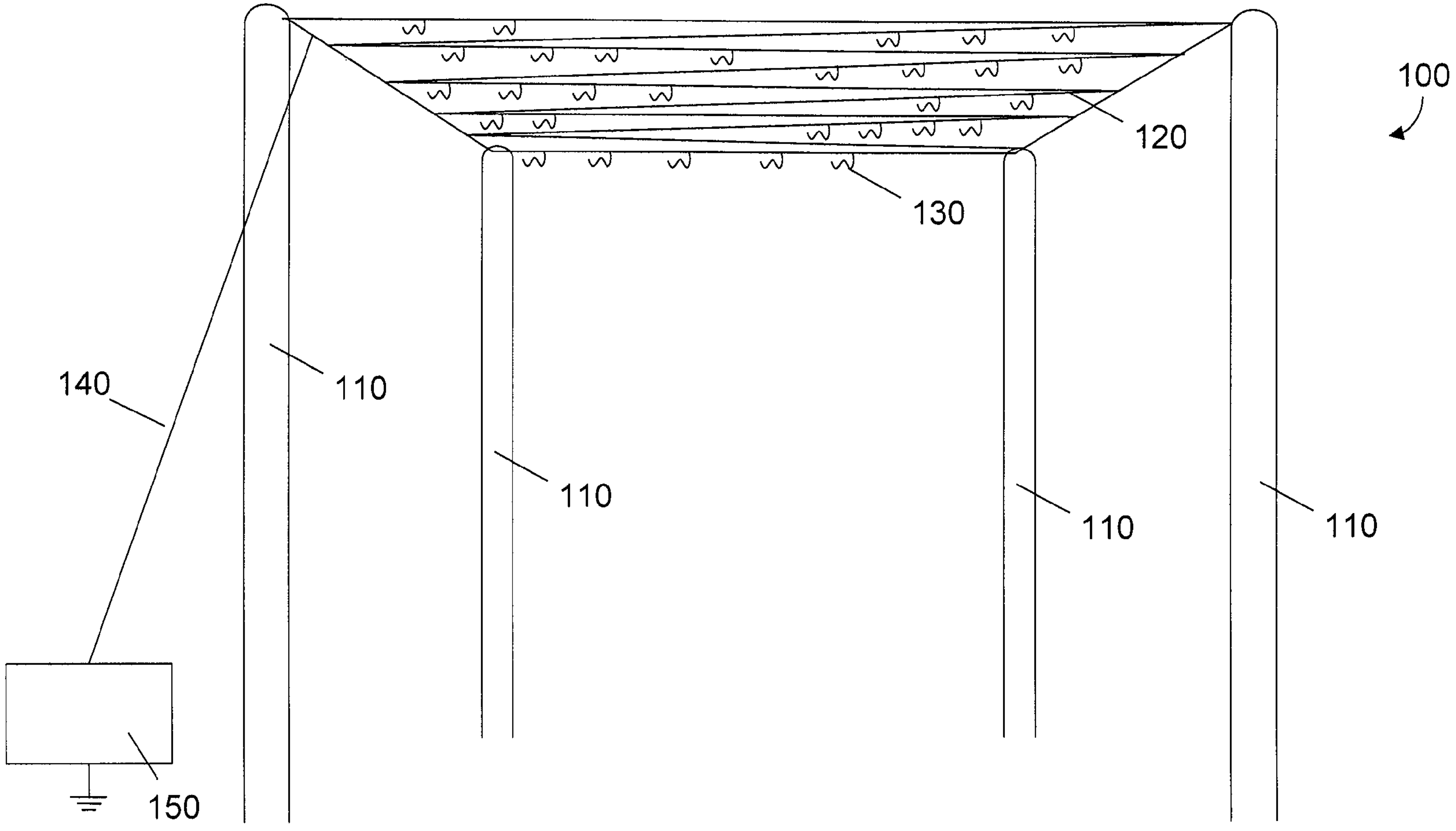
100
120
130
140
110
110
110
110
150

FIGURE 2

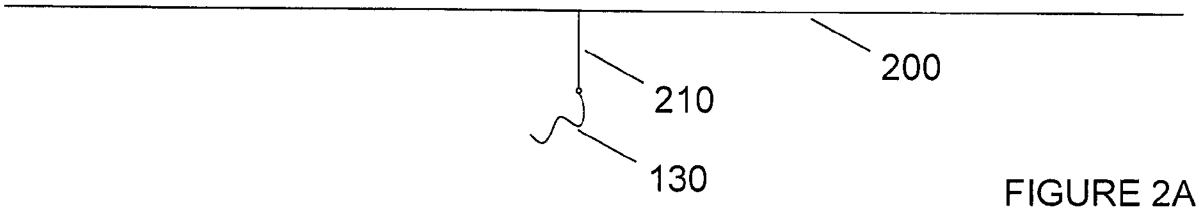

FIGURE 2A

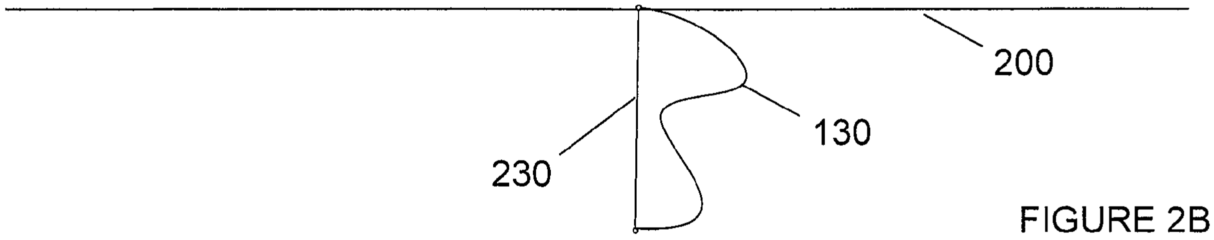

FIGURE 2B

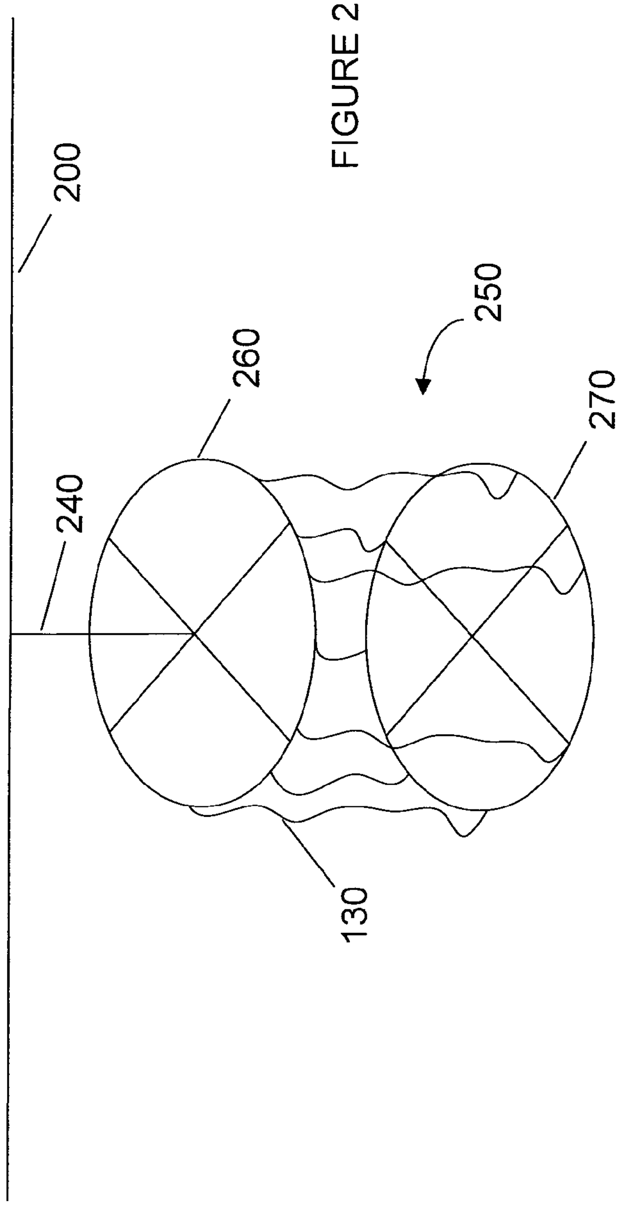
200
240
260
250
270
130

FIGURE 2C

FIGURE 2D

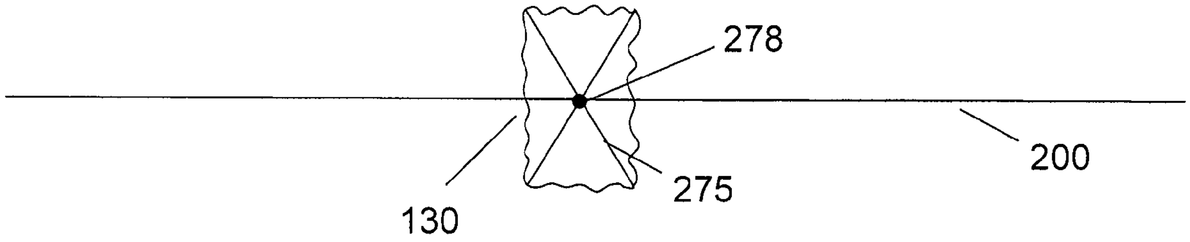
278
200
275
130

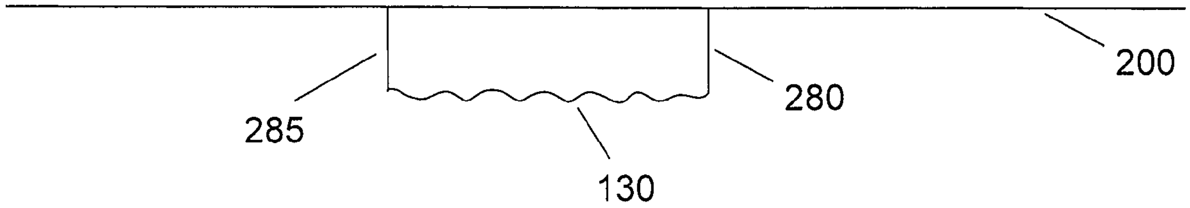
200
280
285
130

FIGURE 2E

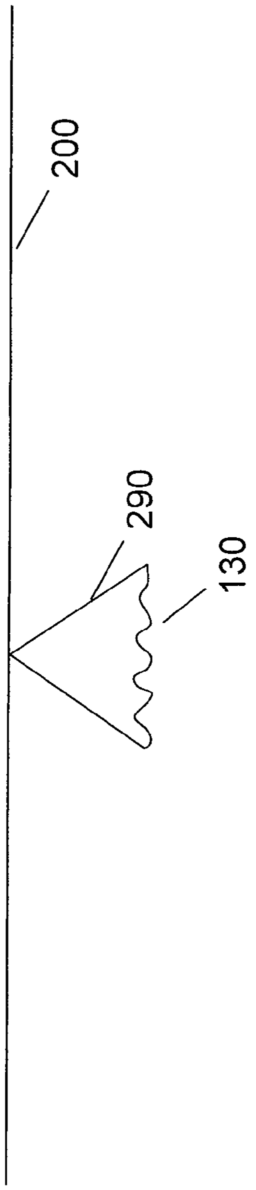

FIGURE 2F

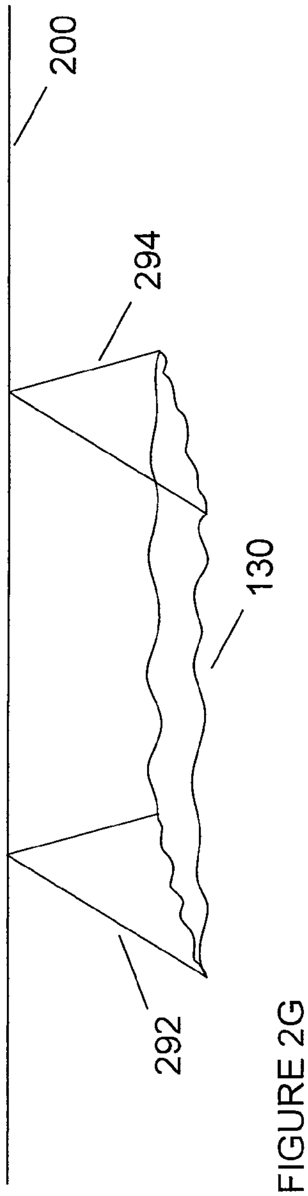
200
294
130
292

FIGURE 2G

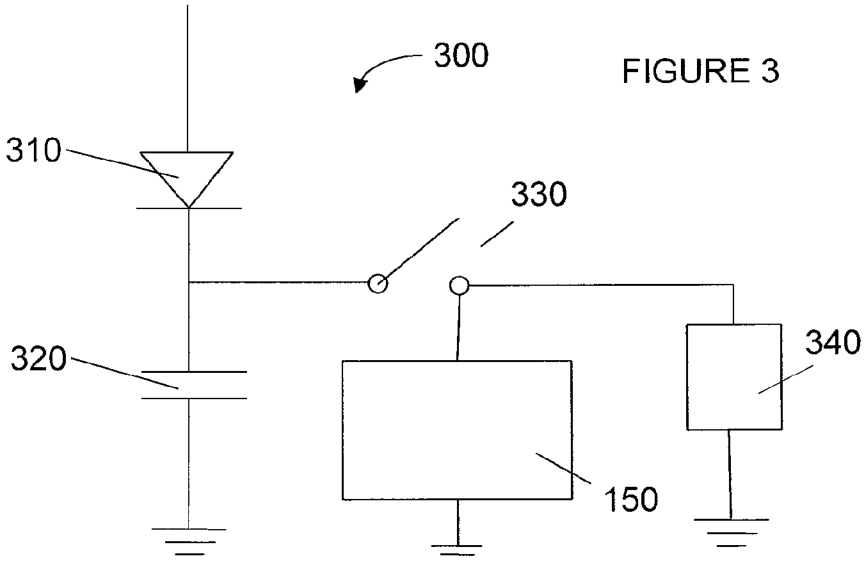
300
310
330
320
340
150

FIGURE 3

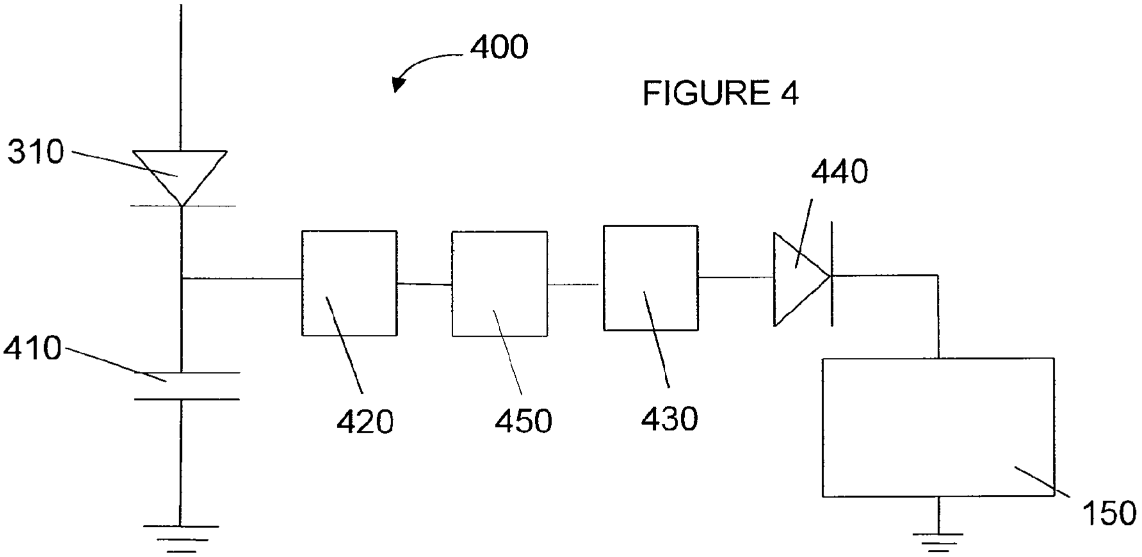
400
310
440
410
420
450
430
150

FIGURE 4

FIGURE 5

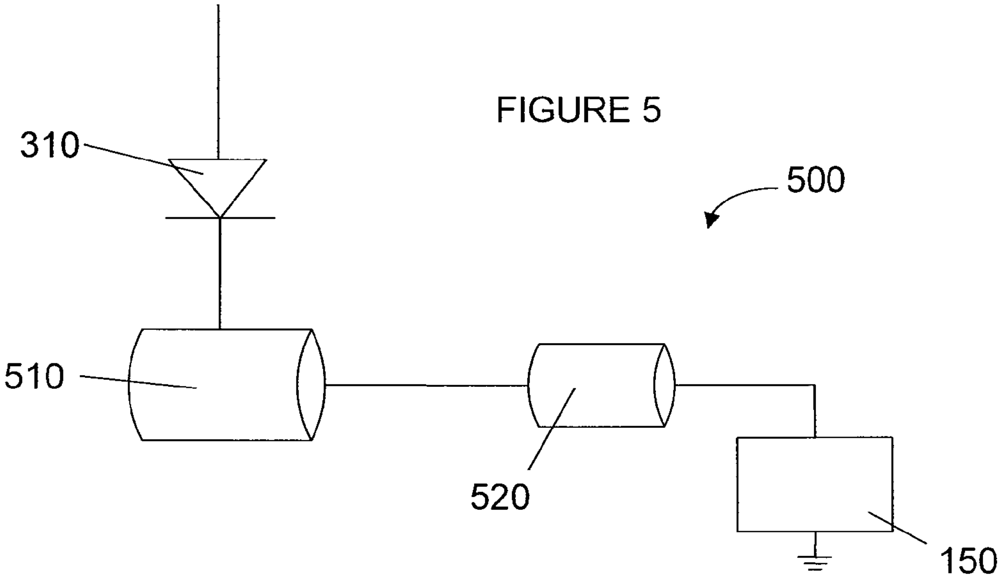
310
500
510
520
150

FIGURE 6

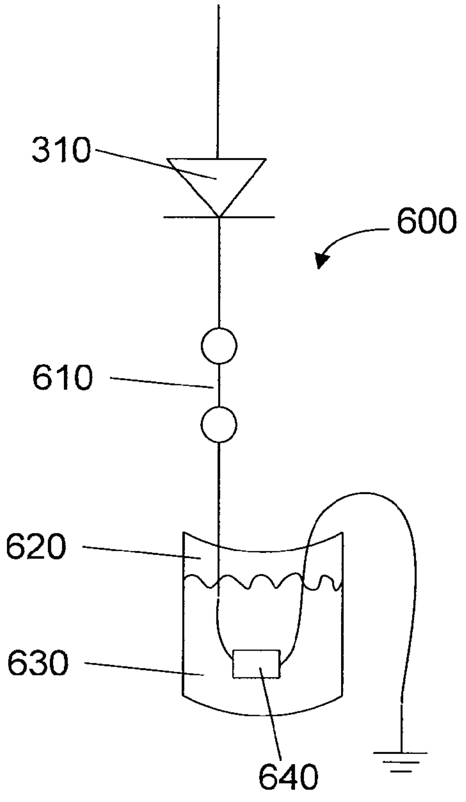
310
600
610
620
630
640

FIGURE 7

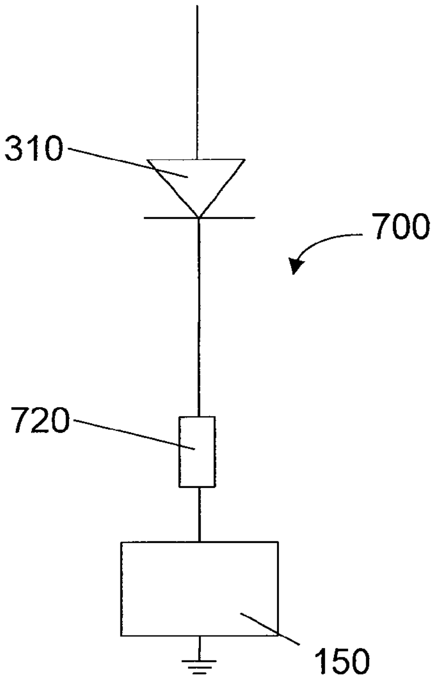
310
700
720
150

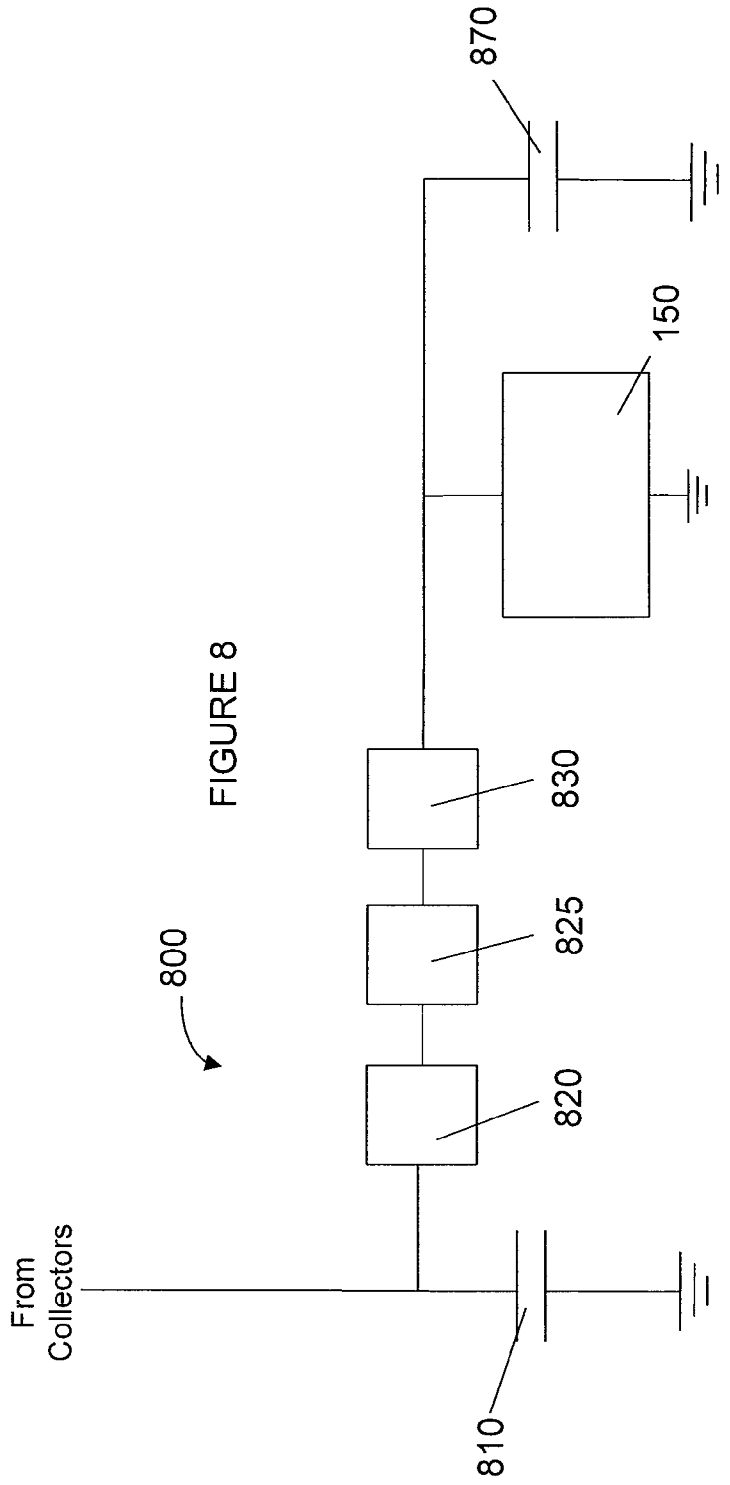
From Collectors
810
820
825
830
800
150
870

FIGURE 8

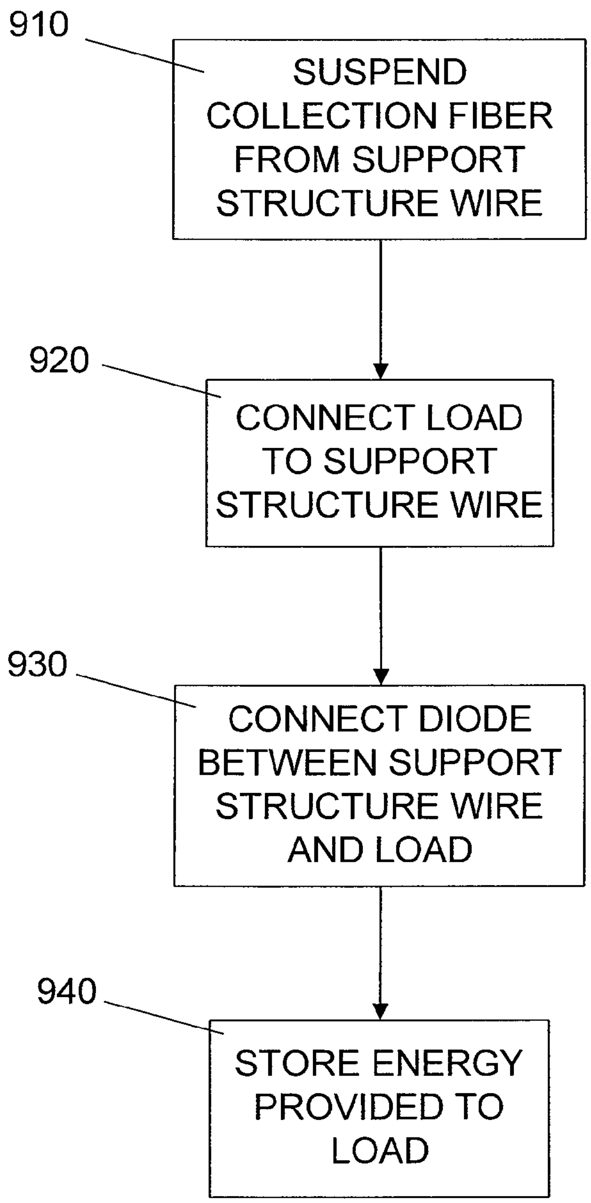
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

ENERGY COLLECTION

CROSS REFERENCE TO RELATED APPLICATION

This application is a continuation application of U.S. patent application Ser. No. 11/358,264, filed on Feb. 21, 2006, which is incorporated by reference herein.

TECHNICAL FIELD

The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure wire elevated above a ground level, at least one collection fiber electrically connected to the support structure wire; a load electrically connected to the support structure wire; and a diode electrically connected between the load and at least one collection fiber.

Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection fiber from a support structure wire elevated above ground level, the fiber electrically connected to the support structure wire; providing a load with an electrical connection to the support structure wire to draw current; and providing a diode electrically connected between the collection fiber and the load.

Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. **1** is a circuit diagram of a weather energy circuit.

FIG. **2** is a perspective view of an exemplary embodiment of many energy collectors elevated above ground by a structure.

FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

FIG. **2**B is a side view of an exemplary embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

FIG. **2**D is a side view of an exemplary embodiment of a support structure for multiple energy collection fibers.

FIG. **2**E is a side view of a support structure for an energy collection fiber.

FIG. **2**F is a side view of an exemplary embodiment of a support structure for an energy collection fiber.

FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

FIG. **3** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **4** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **5** is a circuit diagram of an exemplary embodiment of an energy collection circuit for powering a generator and motor.

FIG. **6** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

FIG. **7** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

FIG. **8** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy.

FIG. **9** is a flow diagram of an exemplary embodiment of collecting energy with a collection fiber.

DETAILED DESCRIPTION

Electric charges on conductors reside entirely on the external surface of the conductors, and tend to concentrate more

around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An exemplary embodiment of this disclosure takes advantage of this property, among others, to collect and use the energy generated by an electric field in the atmosphere. Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An exemplary embodiment utilizes carbon or graphite fibers for static electricity collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including aluminum or steel, but most notably, copper. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

An exemplary embodiment of the collection fibers as collection device **130** includes graphite or carbon fibers. Graphite and carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, or graphite fibers are examples that demonstrate exemplary collection ability.

In at least one exemplary embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than 100 volts per meter under some conditions. When support wire **120** is suspended in the air at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An exemplary embodiment of the collection device comprises the diode with no collection fiber. A preferred embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restricts energy that may be collected through one fiber from escaping through another fiber.

Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2A-2G** using a collection fiber embodiment. FIG. **2A** presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**.

Likewise, FIG. **2B** shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

FIG. **2C** presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

FIG. **2D** presents another exemplary embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2E** provides another exemplary embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

FIG. **2F** presents another exemplary embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2G** provides two supports as provided in FIG. **2F** such that at least two support members **292**, **294** may be connected

to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and between multiple support structures. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**.

FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load. Capacitor **320** may be connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

In an exemplary embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel. Non-limiting examples of spark gap **420** include mercury-reed switches and mercury-wetted reed switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground. The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes may be placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches and mercury-wetted reed switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel.

FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit. Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

FIG. **8** presents exemplary circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches and mercury-wetted reed switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

It should be emphasized that the above-described embodiments of the present disclosure, particularly, any "preferred" embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A method of collecting energy comprising:

suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to the support structure;

providing a load with an electrical connection to the at least one collection device to passively draw current; and

powering a fuel cell between the support structure and the load, the powering performed with energy collected from the at least one collection device.

2. The method of claim **1**, wherein the collection device comprises a diode.

3. The method of claim **1**, wherein the collection device comprises a collection fiber.

4. The method of claim **1**, wherein the collection device comprises a diode and a collection fiber and the diode is electrically connected between the collection fiber and the load.

5. The method of claim **1**, further comprising storing energy provided to the load.

6. The method of claim **5**, wherein storing energy provided to the load comprises storing energy in a capacitor or an inductor.

7. The method of claim **3**, wherein the collection fiber comprises carbon fiber or graphite fiber.

8. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the collection device electrically connected to and suspended from the support structure, the at least one collection device configured to passively draw current;

a load electrically connected to the at least one collection device; and

a diode electrically connected between the at least one collection device and the support structure.

9. The system of claim **8**, wherein the collection device comprises a diode.

10. The system of claim **8**, wherein the collection device comprises a collection fiber.

11. The system of claim **8**, wherein the collection device comprises a collection fiber and a diode electrically connected between the load and the collection fiber.

12. The system of claim **9**, wherein the diode is elevated relative to the ground level.

13. The system of claim **10**, wherein the collection fiber comprises a carbon fiber or a graphite fiber.

14. The system of claim **8**, further comprising:

a switch connected in series between the at least one collection device and the load; and

a capacitor connected in parallel with the switch and the load.

15. The system of claim **14**, wherein the switch comprises an interrupter connected between the at least one collection device and the load, and wherein the interrupter comprises at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

16. The system of claim **15**, further comprising a transformer connected between the interrupter and the load.

17. The system of claim **8**, further comprising:

a motor for providing power, the motor connected between the at least one collection device and the load; and

a generator powered by the motor.

18. The system of claim **8**, further comprising a fuel cell between the support structure and the load.

19. The system of claim **18**, wherein the fuel cell produces hydrogen and oxygen.

20. A system of collecting energy comprising:

means for suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to the means for suspending;

means for passively inducing current flow, the means for inducing current flow electrically connected to the means for suspending; and

means for restricting the backflow of charge carriers, the means for restricting the backflow of charge carriers electrically connected between the at least one collection device and the means for inducing current flow.

21. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the collection device electrically connected to and suspended from the support structure, the at least one collection device configured to passively draw current;

a load electrically connected to the at least one collection device; and

a switch connected in series between the at least one collection device and the load, the switch comprising an interrupter connected between the at least one collection device and the load, and wherein the interrupter comprises at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

22. The system of claim **21**, further comprising a transformer connected between the interrupter and the load.

23. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to and suspended from the support structure, the collection device configured to passively draw current;

a load electrically connected to the at least one collection device;

a motor for providing power, the motor connected between the at least one collection device and the load; and

a generator powered by the motor.

24. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to and suspended from the support structure the at least one collection device configured to passively draw current;

a load electrically connected to the at least one collection device; and

a fuel cell between the support structure and the load.

25. The system of claim **24**, wherein the fuel cell produces hydrogen and oxygen.

* * * * *

US 20070195481A1

(19) **United States**
(12) **Patent Application Publication** (10) **Pub. No.: US 2007/0195481 A1**
McCowen (43) **Pub. Date: Aug. 23, 2007**

(54) **ENERGY COLLECTION**

(76) Inventor: **Clint McCowen**, Navarre, FL (US)

Correspondence Address:
THOMAS, KAYDEN, HORSTEMEYER &
RISLEY, LLP
100 GALLERIA PARKWAY, NW
STE 1750
ATLANTA, GA 30339-5948 (US)

(21) Appl. No.: **11/358,264**

(22) Filed: **Feb. 21, 2006**

Publication Classification

(51) **Int. Cl.**
H01G 7/02 (2006.01)
G01R 31/02 (2006.01)
(52) **U.S. Cl.** **361/233**; 324/72

(57) **ABSTRACT**

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support wire system supported by poles. The support wire system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but carbon and graphite are preferred. Diodes may be used to restrict the backflow or loss of energy.

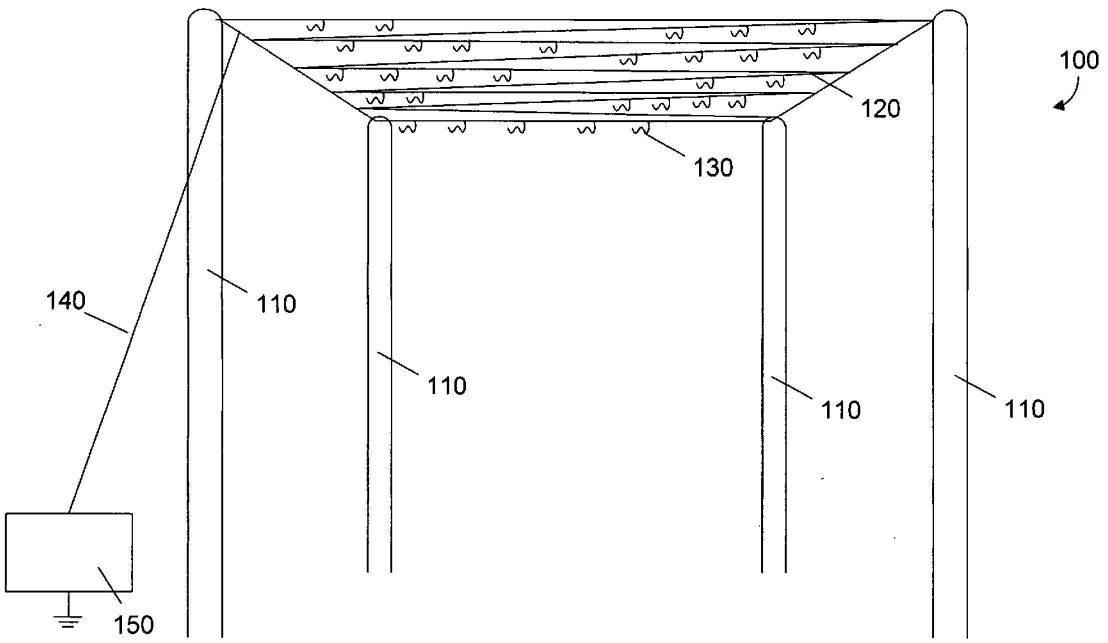

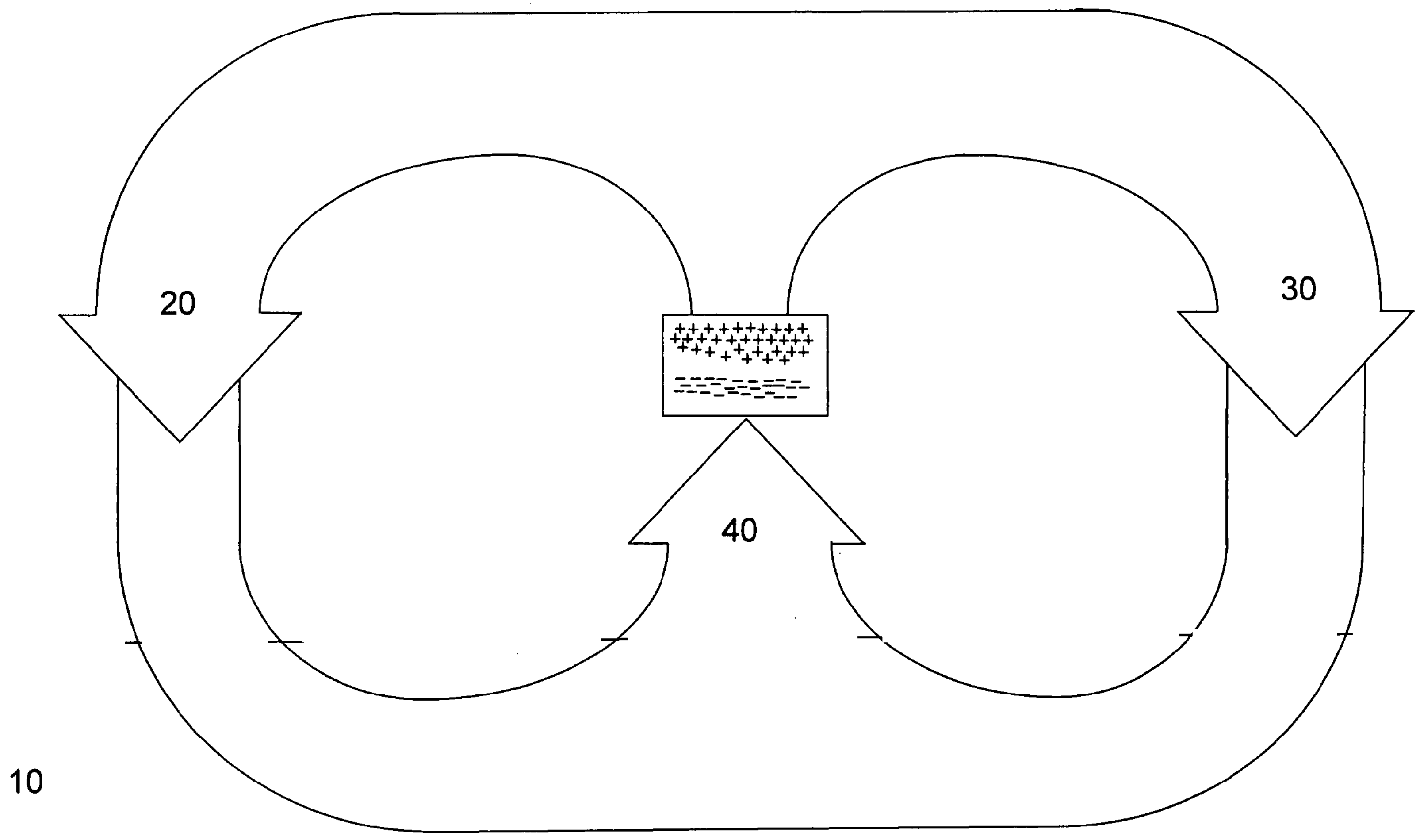

FIGURE 1

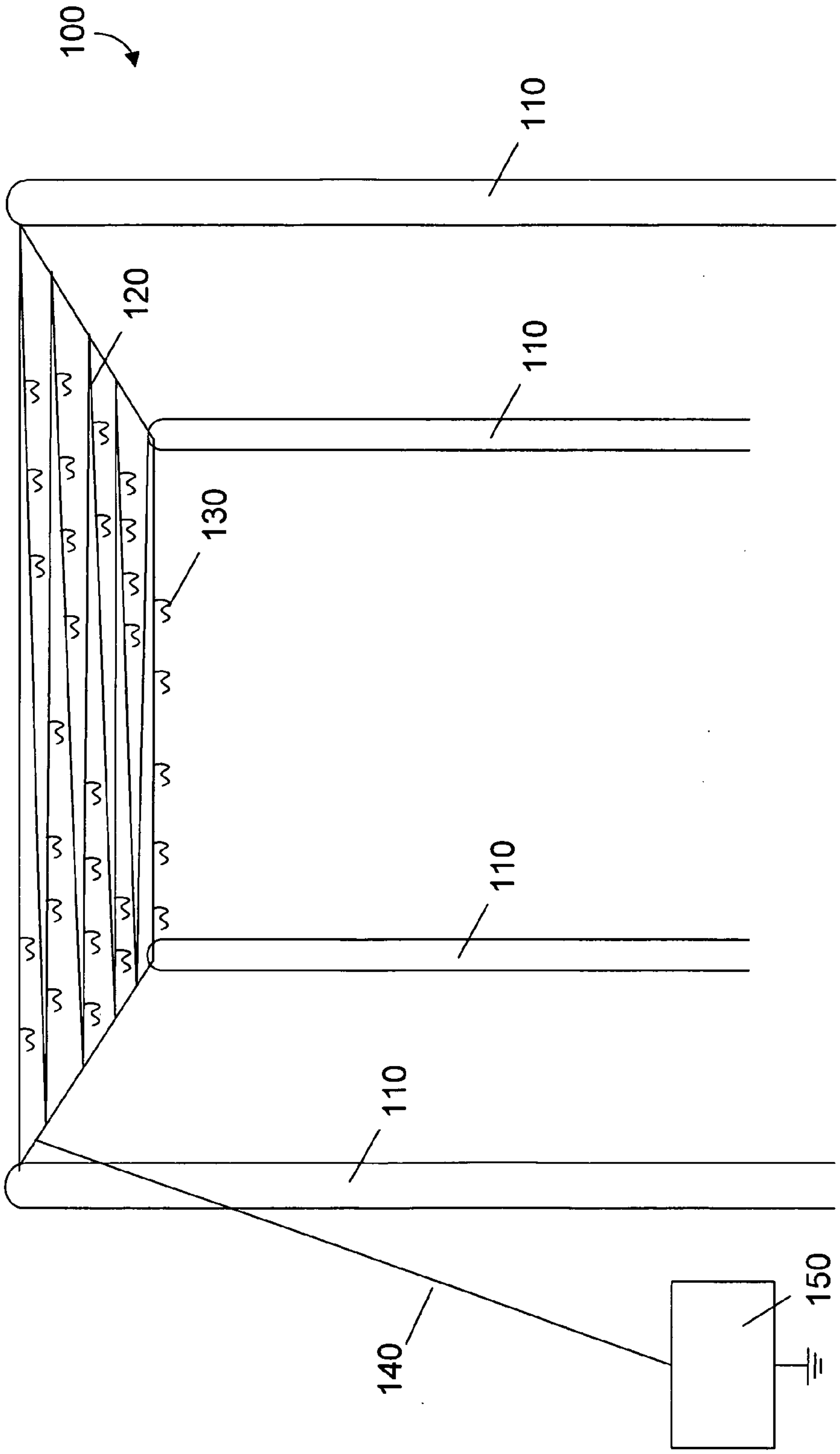
100
110
120
110
130
110
110
140
150

FIGURE 2

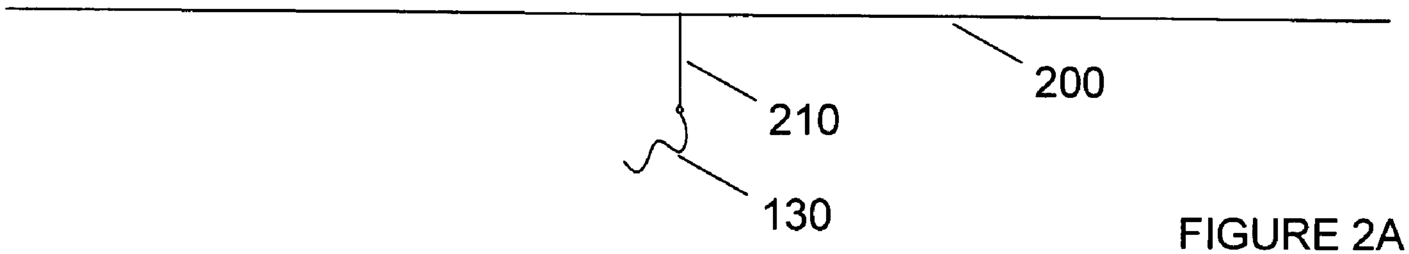

FIGURE 2A

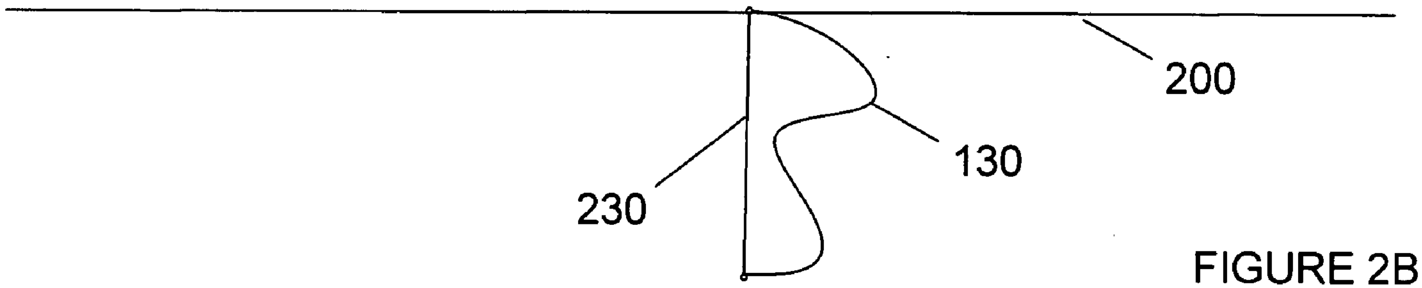

FIGURE 2B

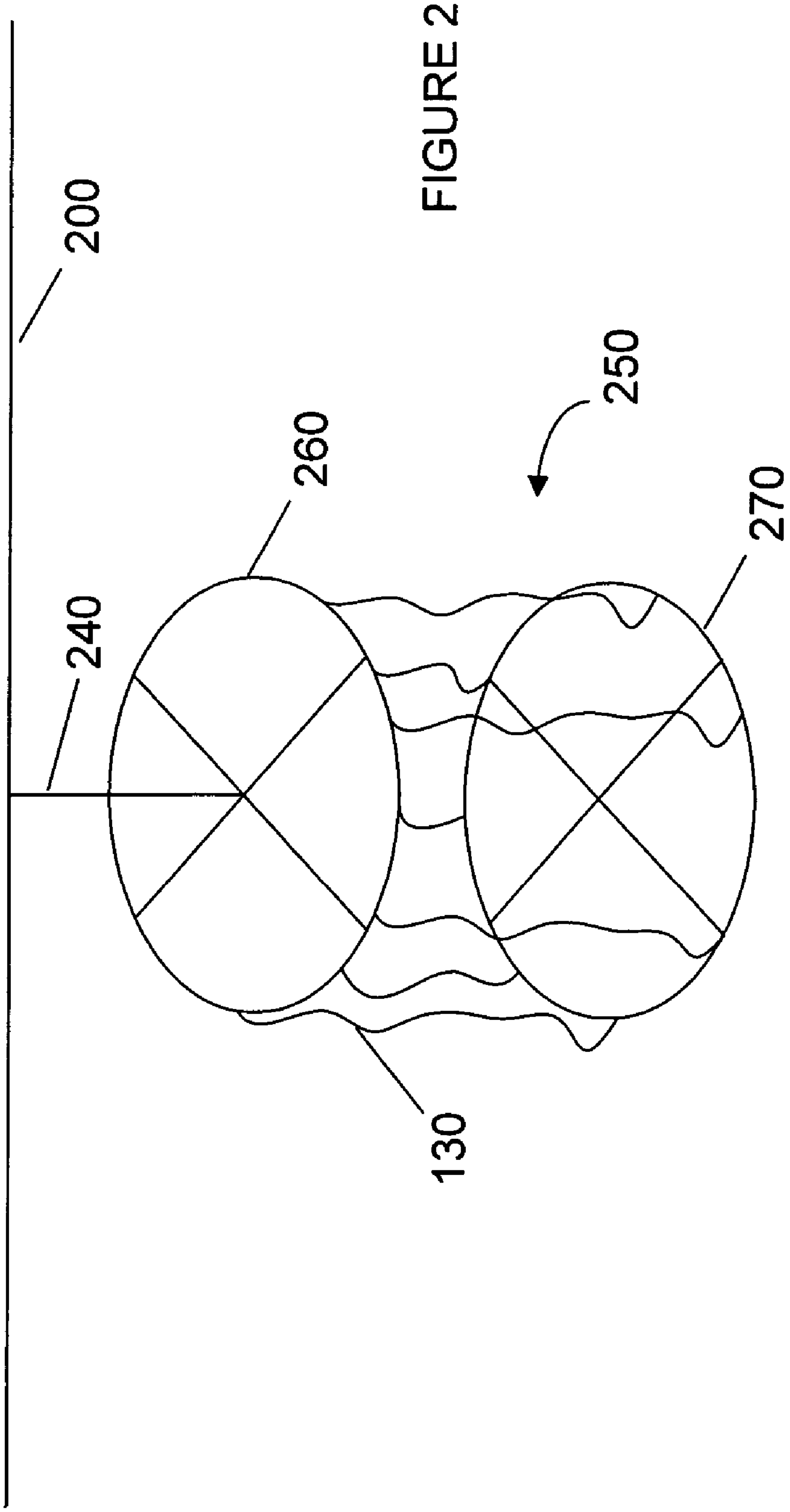
200
240
260
250
270
130

FIGURE 2C

FIGURE 2D

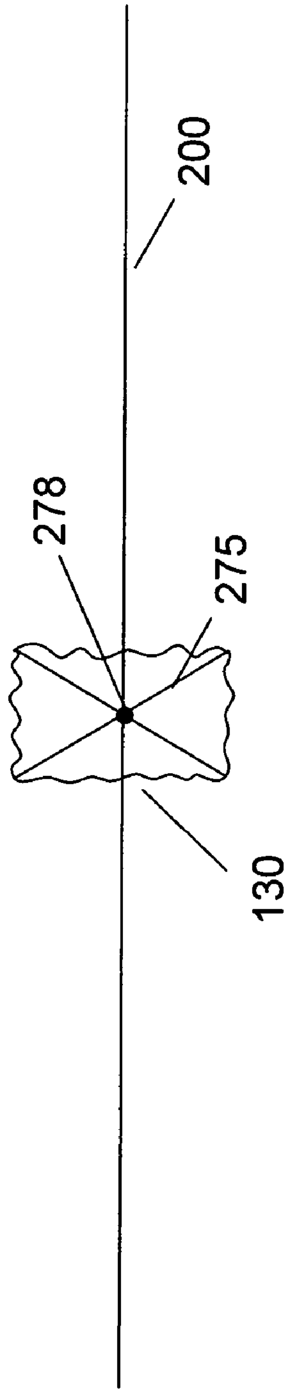
200
278
275
130

FIGURE 2E

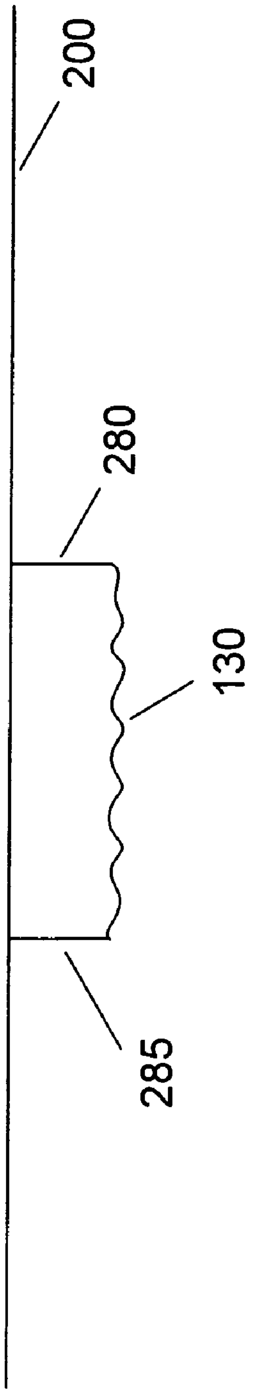
200
280
130
285

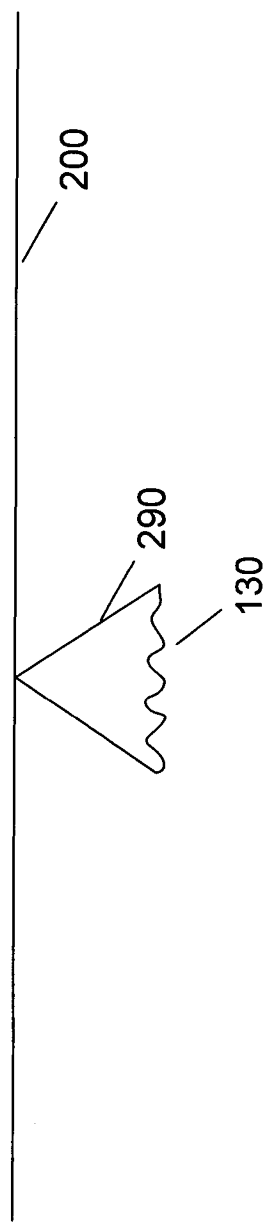

FIGURE 2F

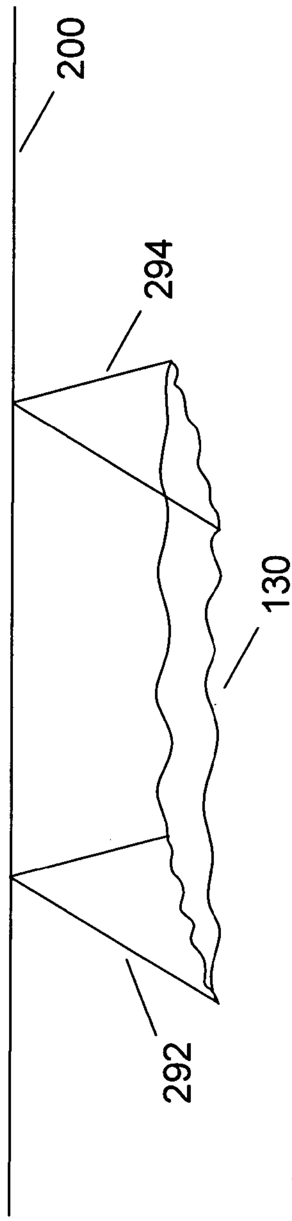
200
294
130
292

FIGURE 2G

FIGURE 3

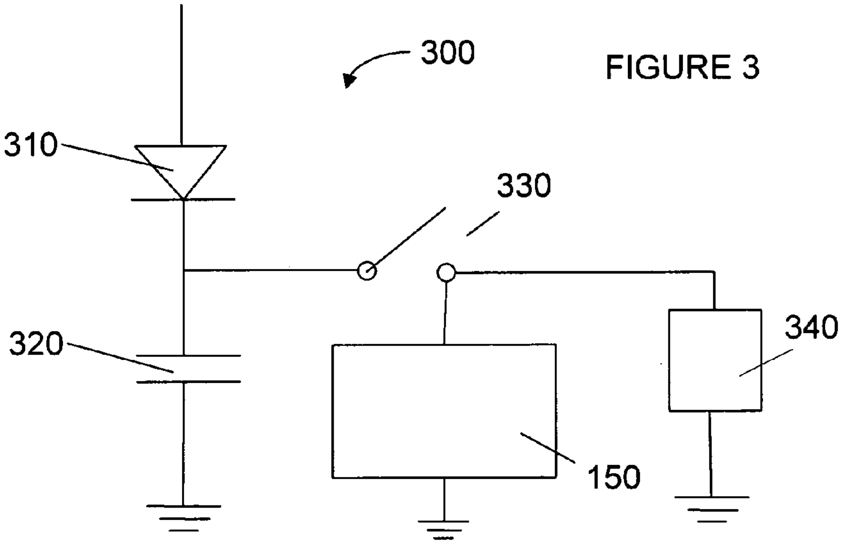
300
310
330
320
340
150

FIGURE 4

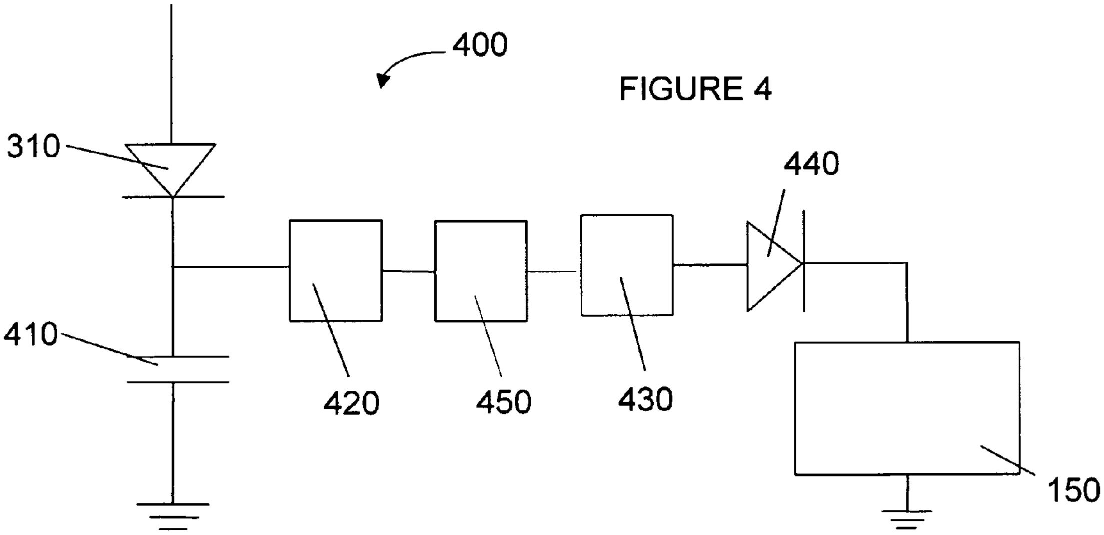
400
310
440
410
420
450
430
150

FIGURE 5

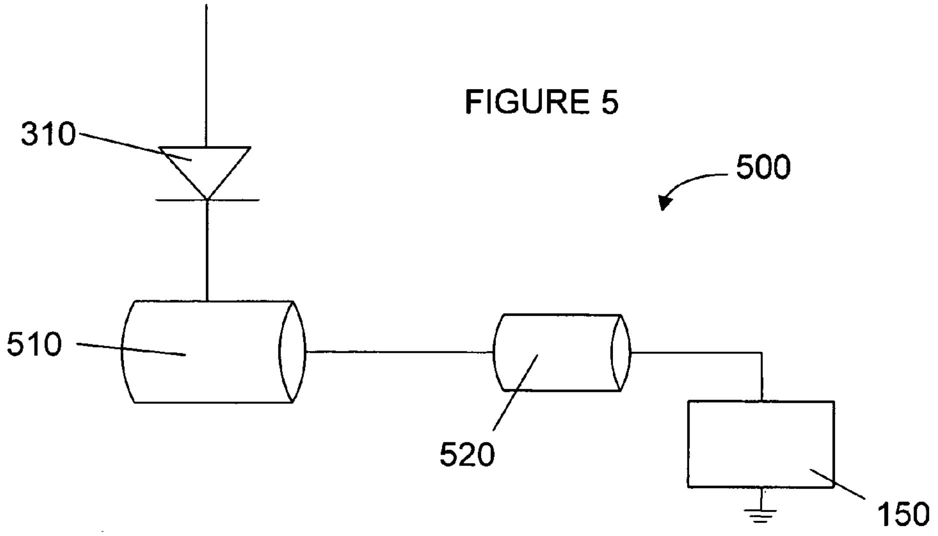
310
500
510
520
150

FIGURE 6

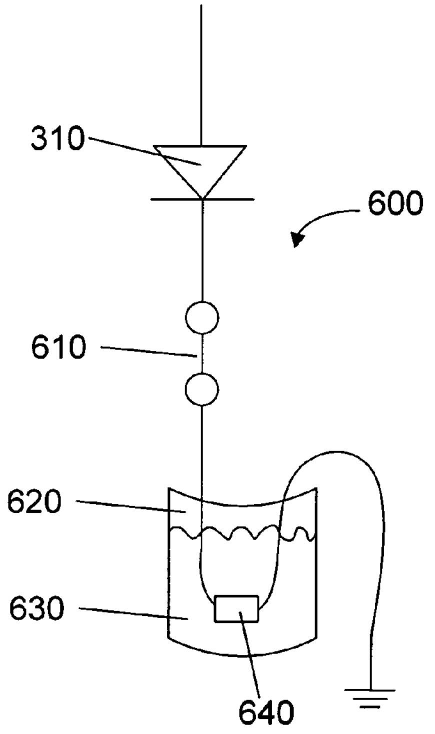
310
600
610
620
630
640

FIGURE 7

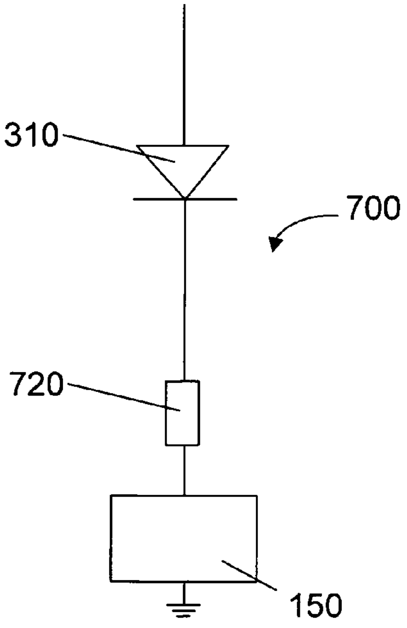
310
700
720
150

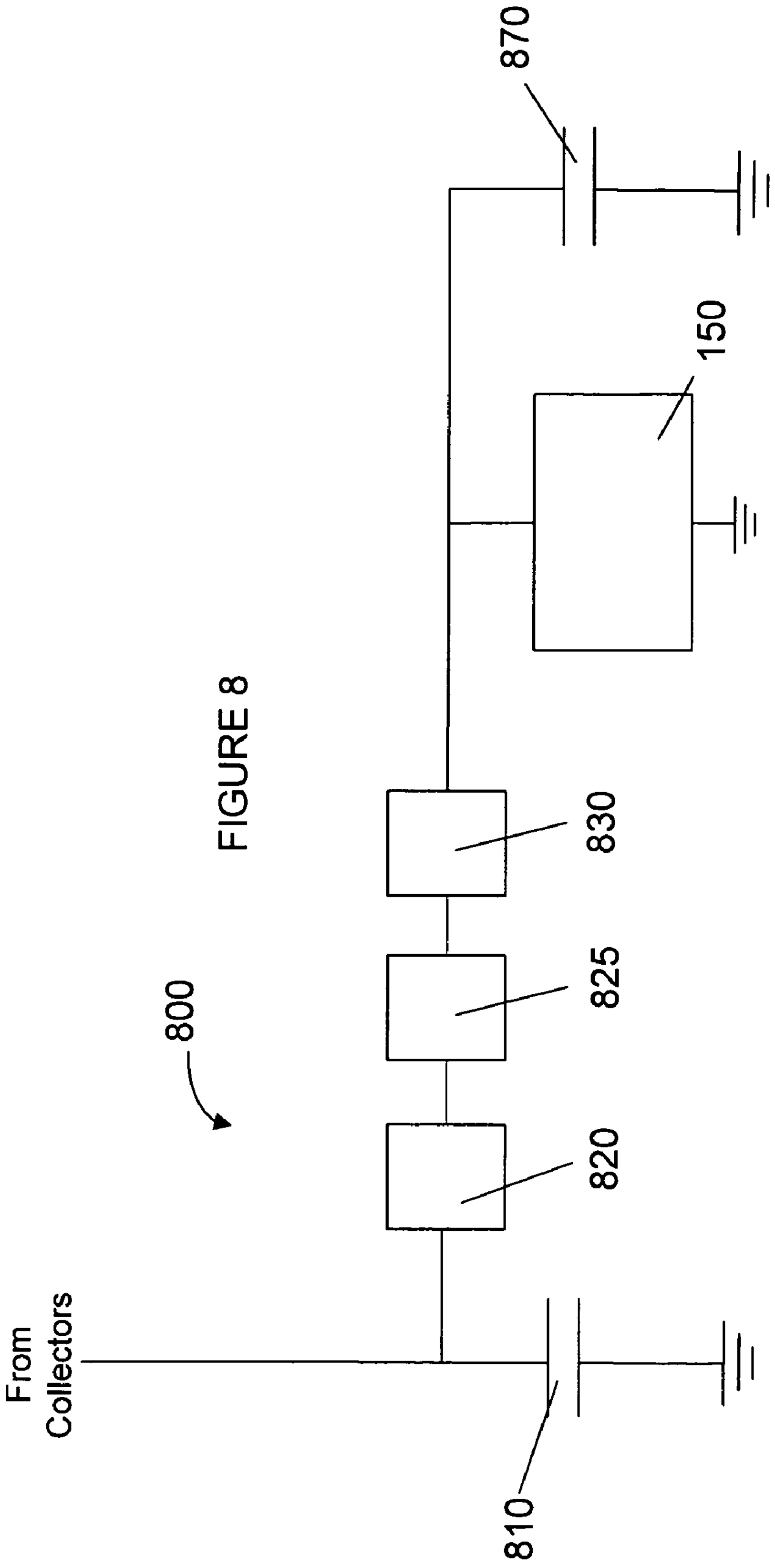
From Collectors
810
800
820
825
830
150
870

FIGURE 8

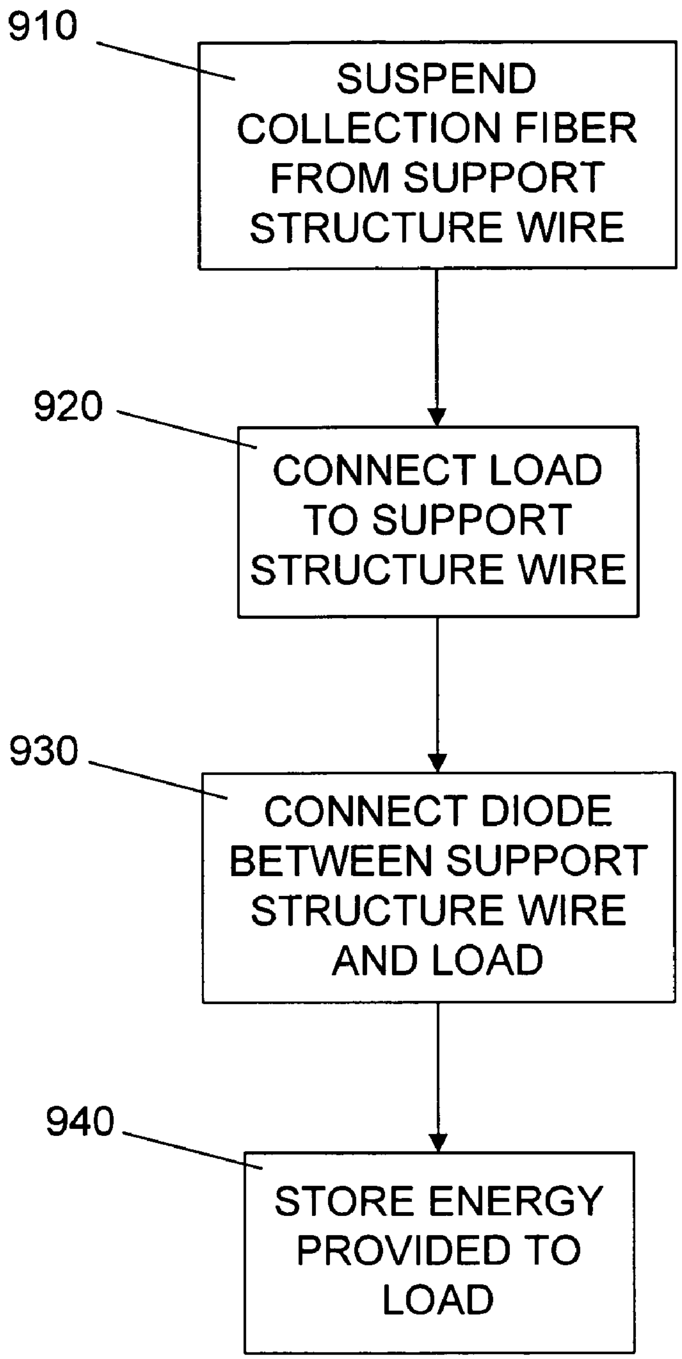
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

ENERGY COLLECTION

TECHNICAL FIELD

[0001] The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

[0002] The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

[0003] In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

[0004] Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

[0005] Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

[0006] While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

[0007] Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure wire elevated above a ground level, at least one collection fiber electrically connected to the support structure wire; a load electrically connected to the support structure wire; and a diode electrically connected between the load and at least one collection fiber.

[0008] Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection fiber from a support structure wire elevated above ground level, the fiber electrically connected to the support structure wire; providing a load with an electrical connection to the support structure wire to draw current; and providing a diode electrically connected between the collection fiber and the load.

[0009] Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

[0011] FIG. **1** is a circuit diagram of a weather energy circuit.

[0012] FIG. **2** is a perspective view of an exemplary embodiment of many energy collectors elevated above ground by a structure.

[0013] FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

[0014] FIG. **2**B is a side view of an exemplary embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

[0015] FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

[0016] FIG. **2**D is a side view of an exemplary embodiment of a support structure for multiple energy collection fibers.

[0017] FIG. **2**E is a side view of a support structure for an energy collection fiber.

[0018] FIG. **2**F is a side view of an exemplary embodiment of a support structure for an energy collection fiber.

[0019] FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

[0020] FIG. **3** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

[0021] FIG. **4** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

[**0022**] FIG. **5** is a circuit diagram of an exemplary embodiment of an energy collection circuit for powering a generator and motor.

[**0023**] FIG. **6** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

[**0024**] FIG. **7** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

[**0025**] FIG. **8** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy.

[**0026**] FIG. **9** is a flow diagram of an exemplary embodiment of collecting energy with a collection fiber.

DETAILED DESCRIPTION

[**0027**] Electric charges on conductors reside entirely on the external surface of the conductors, and tend to concentrate more around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An exemplary embodiment of this disclosure takes advantage of this property, among others, to collect and use the energy generated by an electric field in the atmosphere. Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An exemplary embodiment utilizes carbon or graphite fibers for static electricity collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including aluminum or steel, but most notably, copper. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

[**0028**] An exemplary embodiment of the collection fibers as collection device **130** includes graphite or carbon fibers. Graphite and carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, or graphite fibers are examples that demonstrate exemplary collection ability.

[**0029**] In at least one exemplary embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than **100** volts per meter under some conditions. When support wire **120** is suspended in the air at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

[**0030**] A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An exemplary embodiment of the collection device comprises the diode with no collection fiber. A preferred embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restricts energy that may be collected through one fiber from escaping through another fiber.

[**0031**] Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. 2A-2G using a collection fiber embodiment. FIG. **2A** presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**.

[**0032**] Likewise, FIG. **2B** shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

[**0033**] FIG. **2C** presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

[**0034**] FIG. **2**D presents another exemplary embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

[**0035**] FIG. **2**E provides another exemplary embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

[**0036**] FIG. **2**F presents another exemplary embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

[**0037**] FIG. **2**G provides two supports as provided in FIG. **2**F such that at least two support members **292**, **294** may be connected to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and between multiple support structures. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**.

[**0038**] FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load.

[**0039**] Capacitor **320** may be connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

[**0040**] In an exemplary embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel. Non-limiting examples of spark gap **420** include mercury-reed switches and mercury-wetted reed switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground.

[**0041**] The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

[**0042**] FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

[**0043**] FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes may be placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches and mercury-wetted reed switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel.

[**0044**] FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit. Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

[**0045**] FIG. **8** presents exemplary circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches and mercury-wetted reed switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

[**0046**] FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

[**0047**] Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

[**0048**] It should be emphasized that the above-described embodiments of the present disclosure, particularly, any "preferred" embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A method of collecting energy comprising:

suspending at least one collection device from a plurality of support structure wires elevated above a ground level, the collection device electrically connected to the support structure wire; and

providing a load with an electrical connection to the plurality of support structure wires to draw current.

2. The method of claim 1, wherein the collection device comprises a diode.

3. The method of claim 1, wherein the collection device comprises a collection fiber.

4. The method of claim 1, wherein the collection device comprises a diode and a collection fiber and the diode is electrically connected between the collection fiber and the load.

5. The method of claim 1, further comprising storing energy provided to the load.

6. The method of claim 5, wherein storing energy provided to the load comprises storing energy in a capacitor or an inductor.

7. The method of claim 3, wherein the collection fiber comprises carbon fiber or graphite fiber.

8. The method of claim 2, wherein the diode is elevated relative to the ground level.

9. A system of energy collection comprising:

a plurality of support structure wires elevated above ground level;

at least one collection device electrically connected to the plurality of support structure wires; and

a load electrically connected to the plurality of support structure wires.

10. The system of claim 9, wherein the collection device comprises a diode.

11. The system of claim 9, wherein the collection device comprises a collection fiber.

12. The system of claim 9, wherein the collection device comprises a collection fiber and a diode electrically connected between the load and the collection fiber.

13. The system of claim 10, wherein the diode is elevated relative to the ground level.

14. The system of claim 11, wherein the collection fiber comprises a carbon fiber or a graphite fiber.

15. The system of claim 9, further comprising a diode electrically connected between the at least one collection device and the plurality of support structure wires.

16. The system of claim 11, wherein the collection fiber comprises a first end and a second opposing end, and wherein the system further comprises a support structure connected to both ends of the collection fiber.

17. The system of claim 11, further comprising:

a plurality of collection fibers;

a support frame having a top side and a bottom side; and

a conducting connecting wire between the plurality of support structure wires and the support frame,

wherein one end of each of a plurality of collection fibers is connected to the top side of the support frame, and the opposing end of each of a plurality of collection fibers is connected to the bottom side of the support frame.

18. The system of claim 11, further comprising a rigid structure, the rigid structure comprising multiple supports extending outward from a single point of the support structure, wherein each end of the collection fiber is connected to an end of the multiple supports.

19. The system of claim 11, further comprising a rigid structure, the rigid structure comprising multiple supports extending outward from multiple points of the support structure, wherein each end of the collection fiber is connected to an end of the multiple supports.

20. The system of claim 18, further comprising a plurality of rigid structures wherein collection fibers are connected between a plurality of the rigid structures.

21. The system of claim 9, further comprising:

a switch connected in series between the plurality of support structure wires and the load; and

a capacitor connected in parallel with the switch and the load.

22. The system of claim 21, wherein the switch comprises an interrupter connected between the plurality of support structure wires and the load.

23. The system of claim 22, wherein the interrupter comprises one of a fluorescent tube, a neon bulb, an AC light, or a spark gap.

24. The system of claim 22, further comprising a transformer connected between the interrupter and the load.

25. The system of claim 9, further comprising:

a motor for providing power the motor connected between the plurality of support structure wires and the load; and

a generator powered by the motor.

26. The system of claim 9, wherein the load comprises a spark gap in a container of fluid, and the load is used to produce a chemical reaction.

27. The system of claim 26, wherein the fluid comprises water and the chemical reaction comprises the production of hydrogen and oxygen.

28. The system of claim 9, further comprising a fuel cell between the plurality of support structure wires and the load.

29. The system of claim 9, wherein the load comprises a fuel cell.

30. The system of claim 28, wherein the fuel cell produces hydrogen and oxygen.

31. The system of claim 28, further comprising a diode connected between the plurality of support structure wires and the fuel cell.

32. A system of collecting energy comprising:

means for suspending collection fibers, the means elevated above a ground level, the collection fibers electrically connected to the means for suspending collection fibers;

means for inducing current flow, the means for inducing current flow electrically connected to the means for suspending collection fibers; and

means for restricting the backflow of charge carriers, the means for restricting the backflow of charge carriers electrically connected between the collection fibers and the means for inducing current flow.

* * * * *

US 20100007239A1

(19) **United States**

(12) **Patent Application Publication**

Ogram

(10) **Pub. No.: US 2010/0007239 A1**

(43) **Pub. Date: Jan. 14, 2010**

(54) ATMOSPHERIC ELECTRICAL GENERATOR WITH CHANGE OF STATE

(76) Inventor: **Mark Ellery Ogram**, Tucson, AZ (US)

Correspondence Address:
Mark E. Ogram
Ste. 110-186, 6990 E. 22nd Street
Tucson, AZ 85710-5192 (US)

(21) Appl. No.: **12/460,555**

(22) Filed: **Jul. 21, 2009**

Related U.S. Application Data

(63) Continuation-in-part of application No. 12/218,297, filed on Jul. 14, 2008.

Publication Classification

(51) **Int. Cl.**
H02N 3/00 (2006.01)

(52) **U.S. Cl.** **310/309**

(57) **ABSTRACT**

A mechanism to tap an electrical source which uses an aircraft (**10**) (preferably a lighter than air balloon) tethered by a conductive line (**12**). The conductive line (**12**) is extended/withdrawn by a winch motor (**13**, **14**) to adjust the altitude of the aircraft (**10**). The conductive line (**12**) is isolated from the ground (**17**) and an electrical conductor (**6**) is connected to the conductive line and to an electrical load (**5**). In this manner, static electricity generated in the atmosphere is gathered for use.

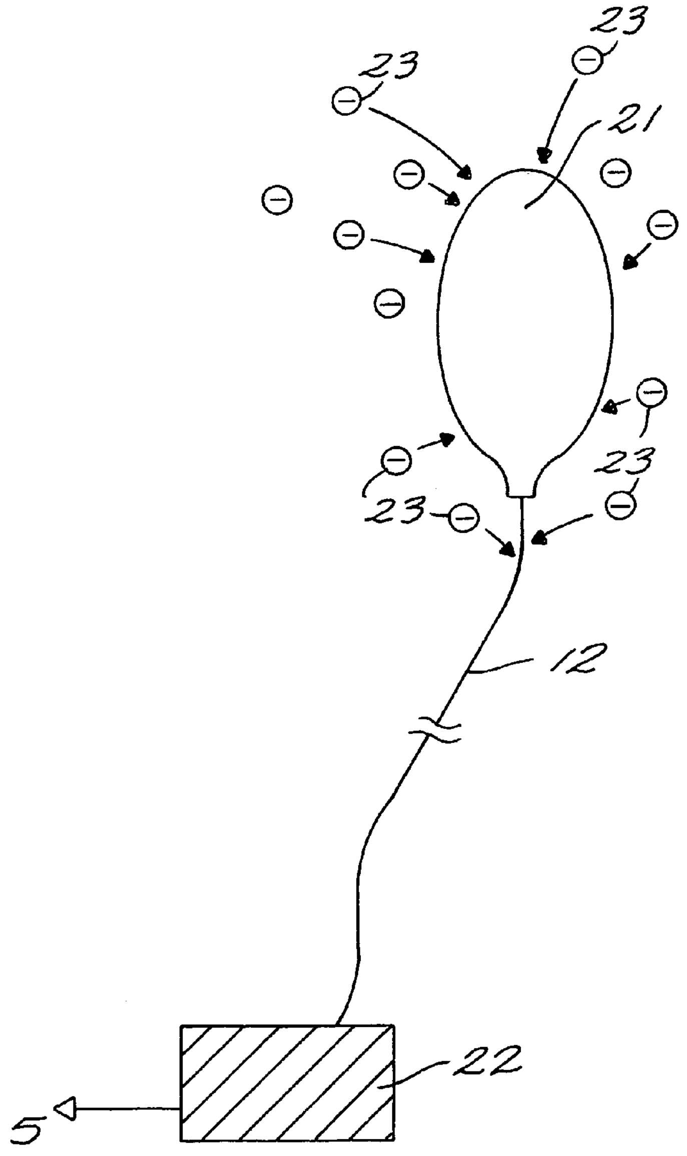

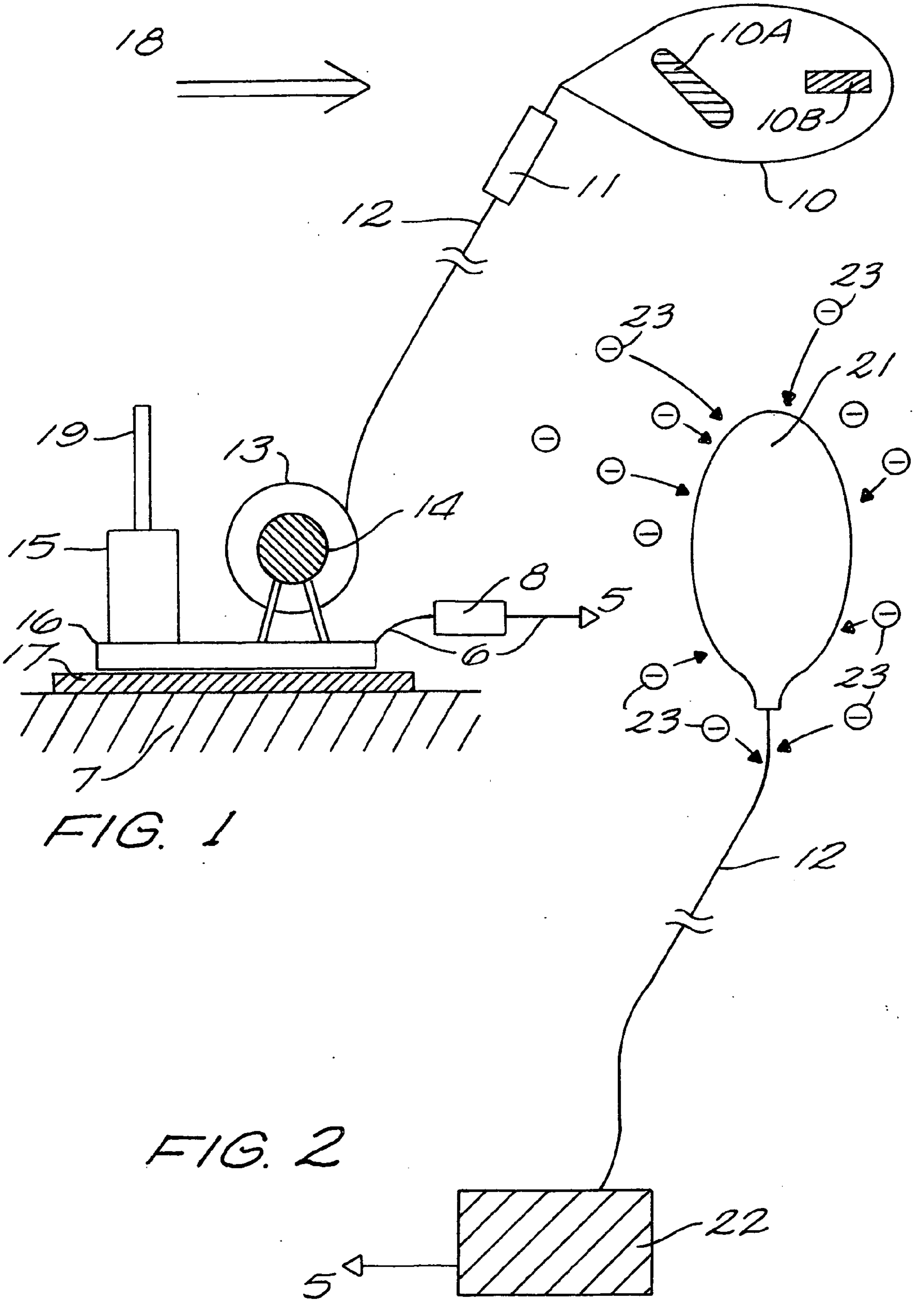
18
10A
10B
10
11
12
23
23
21
19
13
14
15
8
5
16
6
17
7
23
23
23
12
22
5

FIG. 1

FIG. 2

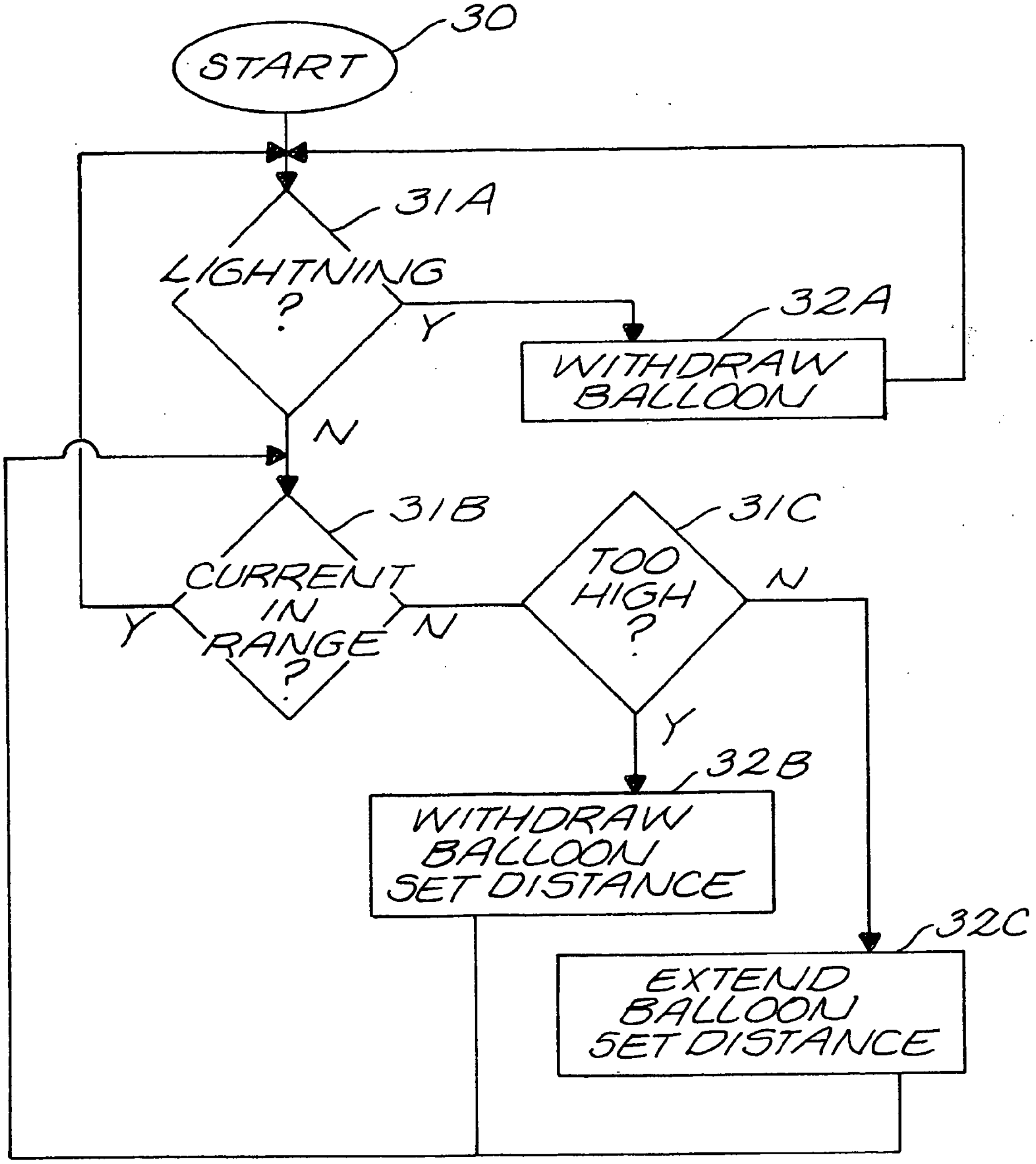
30
START
31A
LIGHTNING ?
Y
32A
WITHDRAW BALLOON
N
31B
CURRENT IN RANGE ?
Y
N
31C
TOO HIGH ?
N
Y
32B
WITHDRAW BALLOON SET DISTANCE
32C
EXTEND BALLOON SET DISTANCE

FIG. 3

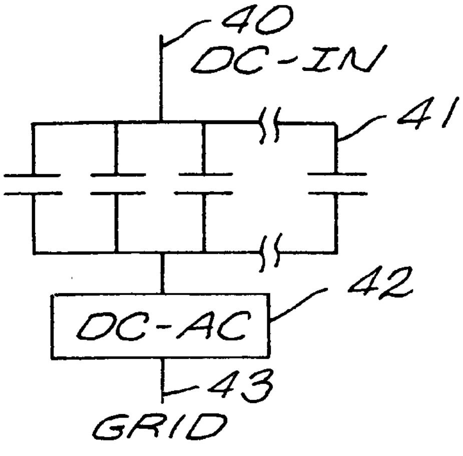
40
DC-IN
41
42
DC-AC
43
GRID

FIG. 4A

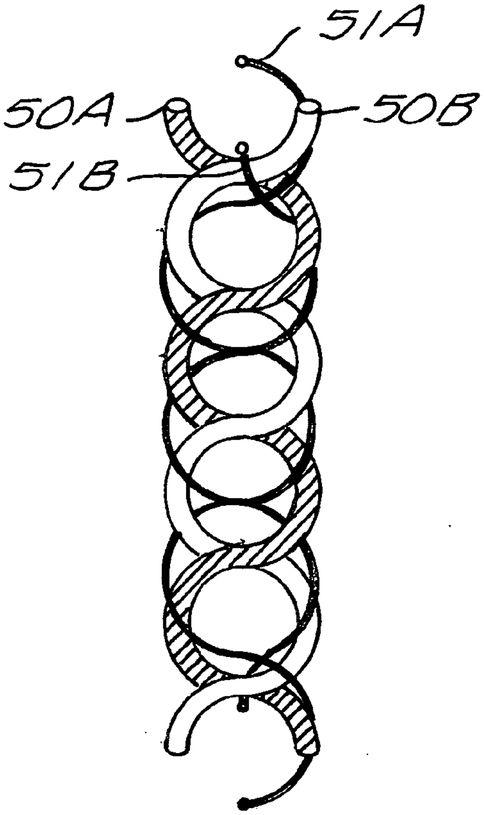
51A
50A
50B
51B

FIG. 5

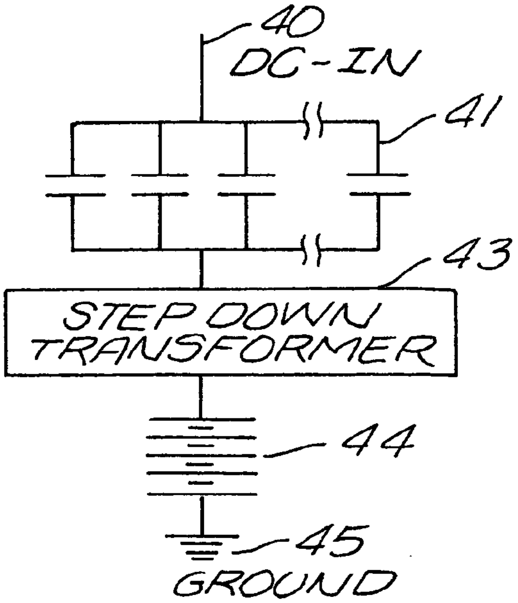
40
DC-IN
41
43
STEP DOWN TRANSFORMER
44
45
GROUND

FIG. 4B

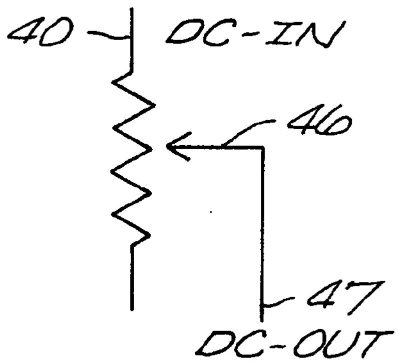
40
DC-IN
46
47
DC-OUT

FIG. 4C

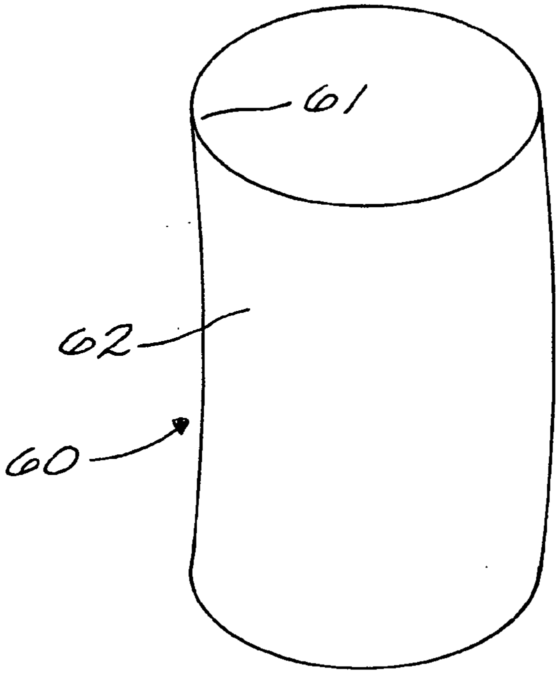

FIG. 6A

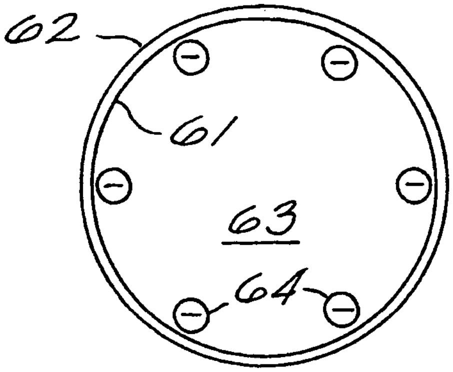

FIG. 6B

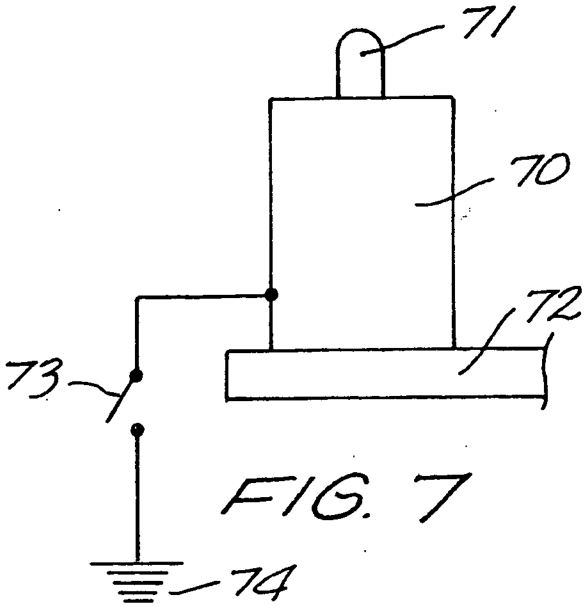
71
70
72
73
74

FIG. 7

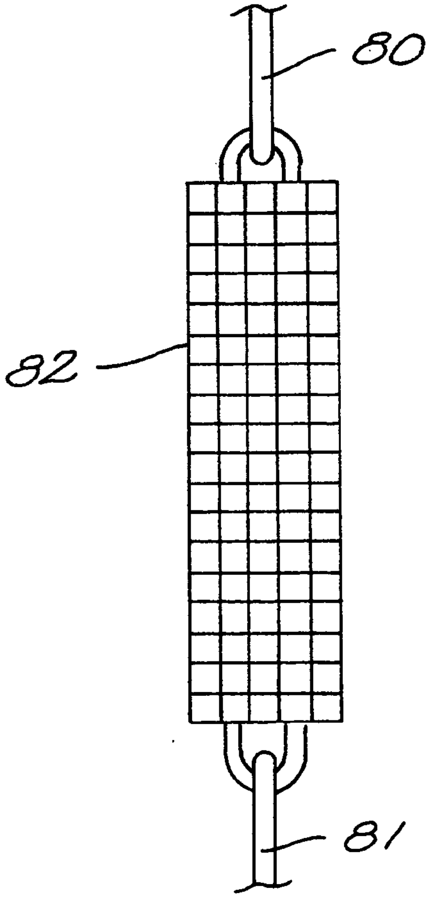

FIG. 8A

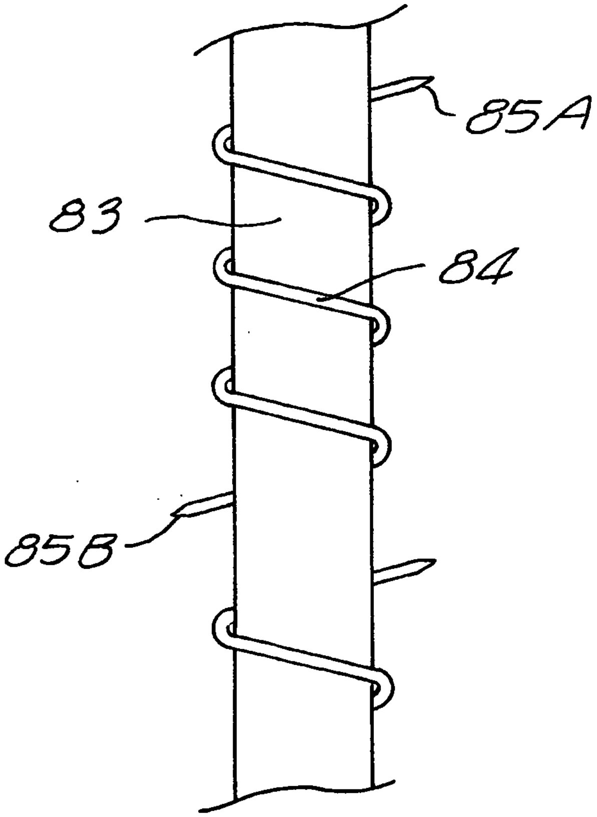
85A
83
84
85B

FIG. 8B

ATMOSPHERIC ELECTRICAL GENERATOR WITH CHANGE OF STATE

[0001] This is a continuation-in-part of U.S. patent application Ser. No. 12/218,297, entitled "Atmospheric Electrical Generator", filed on Jul. 14, 2008, by Ogram.

BACKGROUND OF THE INVENTION

[0002] This invention relates to the production of electrical energy and more particularly to the production of electrical energy from the atmosphere.

[0003] Everyone is familiar with Benjamin Franklin's kite experiment of 1752. Using a kite whose string had become wet, negative charges from the passing clouds flowed into the string, down to the suspended key, and then into a Leyden jar via a thin metal wire. Franklin was protected by a dry silk string; but, when Franklin's knuckle came too close to the key, he received a strong shock. Fortunately, Benjamin Franklin was not killed, others who tried this same experiment were not so lucky.

[0004] Since then, the formation of lightning has remained something of a mystery. Lightning bolts are triggered when a negatively charged cloud base induces a positive charge from the ground, thereby forming a "pathway" for the discharge of the collected electrical energy.

[0005] Lightning travels up to 60,000 miles per hour with a flash that is brighter than ten million 100-watt lightbulbs. This wattage is as much power as is produced by all of the electricity plants in the United States and with a voltage of up to 300 million volts.

[0006] It is this very fact, the power within lightning is immense, that has prevented any successful collection of the electrical energy from lightning. The electricity in lightning is far too extreme for current technology to harness.

[0007] While lightning has attracted a energy starved industrial world, no one has developed any technique to harness this naturally occurring electrical source.

[0008] It is clear there is a continuing need for an electrical source other than carbon-based fuels and that the naturally occurring electricity in the atmosphere is being ignored.

SUMMARY OF THE INVENTION

[0009] The invention is a mechanism which taps into the naturally occurring static electricity in the atmosphere. Whereas heretofore, the attempt to garner electricity from the atmosphere has focused exclusively on capturing lightning, the present invention syphons off the static electricity which is generated from any agitated air and voids lightning.

[0010] Lightning is only the final discharge of the static electricity, whether that lightning is intra-cloud lightning, cloud-to-ground lightning, or inter-cloud lightning. Other types of final discharges are known as heat lightning, summer lightning, sheet lightning, ribbon lightning, silent lightning, ball lightning, bead lightning, elves, jets, and sprites. Well before these discharges are observed, as the atmosphere becomes agitated by wind or thermal, static electricity is being generated.

[0011] The present invention recognizes that this static electricity is being formed and creates a mechanism to capture it.

[0012] The mechanism of this invention utilizes an aircraft such as a lighter than air balloon. While the preferred embodiment uses a foil balloon, a variety of other aircraft are obvious to those of ordinary skill in the art, including, but not limited to: gliders, rubber balloons (such as weather balloons), biaxially-oriented polyethylene terephthalate polyester film balloons, and latex balloons.

[0013] Within this discussion, the balloon is referenced, but, the invention is not intended to be limited solely to balloons.

[0014] The balloon is sent aloft and is tethered by a conductive line. In this context, the conductive line may be any obvious to those of ordinary skill in the art. For the preferred embodiment, the conductive line is a generically referred to as a "poly-rope" and is commercially available through a variety of sources. A suitable conductive line is described in U.S. Pat. No. 5,203,542, entitled "Apparatus for Improved Electric Fence Wire Construction for use with Intensive Grazing" issued Apr. 20, 1993, to Coley, et al. and incorporated hereinto by reference.

[0015] The conductive line is played out of a winch to control the altitude of the balloon. The motor controlling the winch is able to reverse direction to both extend and withdraw the conductive line which is wrapped around a spool on the winch. The winch/spool combination are part of a base unit.

[0016] In some embodiments of the invention, the spool is constructed of rubber so as to insulate the conductive line from the winch assembly. In this embodiment, only the conductive line is charged by the atmospheric static electricity while the winch remains neutral.

[0017] In yet another embodiment, the winch/spool are part of a base unit which is itself isolated from the ground by an insulator. In this embodiment, the entire base unit is charged by the atmospheric static electricity.

[0018] A conductor, such as an insulated wire, is electrically connected to the conductive line. In one embodiment, where the conductive line is electrically isolated from the spool and winch motor, the conductor is connected to the conductive line. In the embodiment where the conductive line is electrically connected to the base unit, then the conductor is connected anywhere on a metalic base unit.

[0019] The other end of the conductor is connected to a load. The load in this case can be any of a variety of electrical loads well known to those of ordinary skill in art, including, but not limited to a motor, a battery system, or the electrical grid for the system.

[0020] In the preferred embodiment, a sensor array is used to monitor the activities both at the base unit (such as electrical flow within the conductor) and in the surrounding locale.

[0021] A sensor monitoring the electrical flow (i.e. voltage and/or current) within the conductor is used to monitor the electrical activity within the conductor.

[0022] In the preferred embodiment, a lightning sensor monitors for lightning activity within the locale. As noted earlier, the electrical characteristic of lightning is so extreme that ideally this discharge is avoided as it might damage the mechanism of this invention.

[0023] The sensor array is utilized by a controller, such as microprocessor, programmed to operate the mechanism as outlined herein.

[0024] The controller operates the winch motor to extend or withdraw the conductive line and by extension the altitude of the balloon. The controller is programmed to operate the winch by monitoring the electrical characteristics of the conductor and adjusting the balloon's altitude to maintain these characteristics within the conductor within a preset range.

[0025] This preset range is established either in the base programming of the controller or is established by an operator of the system.

[0026] As example, by controlling the amount of current being withdrawn from the atmosphere, the mechanism operates within a safe range and also provides a relatively stable current flow from which a variety of activities can take place (such as DC-AC conversion).

[0027] The controller also utilizes the lightning sensor to protect the mechanism from a lightning strike. Should lightning be detected within a pre-determined range (as established by the software or defined by an operator), then the balloon is pulled down to minimize the risk of damage from a lightning strike.

[0028] The invention, together with various embodiments thereof will be more fully explained by the following description of the accompanying drawings.

DRAWING IN BRIEF

[0029] FIG. **1** diagrams the preferred embodiment of the invention.

[0030] FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

[0031] FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

[0032] FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

[0033] FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

[0034] FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

[0035] FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

[0036] FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

DRAWINGS IN DETAIL

[0037] FIG. **1** diagrams the preferred embodiment of the invention.

[0038] Balloon **10** is an aircraft which, in this illustration, is a lighter than air balloon. Wings **10**A, extending from the body of balloon **10**, provide additional lift in air flow **18**. Tail **10**B helps to stabilize balloon **10**.

[0039] Balloon **10** is tethered to the ground via conductive line **12**. As noted earlier, a variety of configurations and materials are available to serve as conductive line **12**. In this illustration, a poly-wire is used. Poly-wire is commercially available through a variety of vendors, including, but not limited to: Jeffers Livestock and Sareba Systems, Inc. of Ellendale, Minn.

[0040] In this embodiment, located proximate to balloon **10**, is an electrical apparatus **11**. The electrical apparatus **11**, is either a collection enhancement mechanism which enhances the collection of static electricity or a transformer.

[0041] In the collection enhancement lead embodiment, an electrical collection enhancement lead is configured to attract the static charge and conduct the charge into the conductive line **12**.

[0042] In another embodiment, the electrical apparatus **11** is a transformer which converts the atmospheric static electricity from a direct current configuration to an alternating current configuration. An insulated conductor extends from the ground to the transformer to assist in this operation.

[0043] Those of ordinary skill in the art readily recognize a variety of DC-AC transformers such as, but not limited to: U.S. Pat. No. 7,471,528, entitled "Parallel Operating System of DC-AC Converters and Controller IC Therefor" issued to Fukumoto on Dec. 30, 2008; U.S. Pat. No. 7,499,290, entitled "Power Conversion" issued to Mazzola, et al. on Mar. 3, 2009; U.S. Pat. No. 7,531,993, entitled "Half Bridge Circuit and Method of Operating a Half Bridge Circuit" issued to Udrea et al. on May 12, 2009; U.S. Pat. No. 5,538,536, entitled "DC-AC Converter and Controller IC Therefor" issued to Fukumoto on May 26, 2009; and, U.S. Pat. No. 7,554,823, entitled "Controller IC, DC-AC Conversion Appratus, and Parallel Running System of DC-AC Conversion Apparatuses" issued to Fukumoto, et al. on Jun. 30, 2009; all of which are incorporated hereinto by reference.

[0044] The electricity flows, whether direct current or alternating current, using the conductive line into spool **13**, where the conductive line **12** is collected and either withdrawn or dispensed through operation of winch motor **14**.

[0045] Winch motor **14** and spool **13** are mounted onto base unit **16** which is electrically isolated from ground **7** using insulator **17**. Note, in this embodiment of the invention, when electricity is being collected from the atmosphere, the entire base unit **16** becomes charged. In another embodiment of the invention, spool **13** is constructed of rubber, thereby preventing base unit **16** from becoming charged, thereby restricting the charging from the atmosphere to only conductive line **12**.

[0046] In this embodiment, conductor **6** is connected to base unit **16** (since the entire base unit **16** is charged and the base unit is metalic) to communicate the electrical current to load **5**. Conductor **6** is ideally an insulated wire.

[0047] The electrical current through conductor **6** is measured using sensor **8**.

[0048] In the alternative embodiment discussed above, where only the conductive line **12** is charged, then conductor **6** is connected to conductive line **12**.

[0049] Controller **15**, located in this embodiment on base unit **16**, operates winch motor **14** in response to signals from sensor **8** (measuring the current being discharged to load **5**) to maintain the current flow within a pre-defined range. As the current flow diminishes, then the conductive line **12** extended from spool **13** to increase the altitude of balloon **10** to that more static charge from the atmosphere is gathered; as the current flow falls exceeds a preset level, conductive line **12** is withdrawn onto spool **13** to decrease the static charge being collected from the atmosphere.

[0050] The range of current flow through conductor **6** is ideally set by the program, although some embodiments of the invention permit an operator to establish this range of operation.

[0051] In an alternative embodiment, the sensor monitoring conductor **6** monitors the voltage therein.

[0052] In the preferred embodiment of the invention, controller **15** is also equipped with a lightning sensor **19**. In this embodiment, when lightning is sensed within a preset range, then substantially all of conductive line **12** is wound onto spool **13** to pull balloon **10** near the ground and protect the entire mechanism from being damaged from a lightning discharge.

[0053] In the preferred embodiment, the "safe" distance form lightning is set in the programming of controller **15** and is ideally two miles; other embodiments permit the operator to "safe" distance.

[0054] There are a variety of lightning sensors well known to those of ordinary skill in the art, including, but not limited to those described in: U.S. Pat. No. 7,016,785, entitled "Lightning Detection" issued to Makela, et al. on Mar. 21, 2006; U.S. Pat. No. 6,829,911, entitled "Lightning Detection and Prediction Alarm Device" issued to Jones, et al. on Dec. 7, 2004; U.S. Pat. No. 7,200,418, entitled "Detection of Lightning" issued to Karikuranta, et al. on Apr. 3, 2007; and U.S. Pat. No. 6,961,662, entitled "Systems and Methods for Spectral Corrected Lightning Detection" issued to Murphy on Nov. 1, 2005; all of which are incorporated hereinto by reference.

[0055] In another embodiment of the invention, controller **15** is not located on base unit **16**, rather it is remote and communicates its control signals to winch motor **14** using radio waves.

[0056] FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

[0057] Static charges **23** are generated in the atmosphere by agitated air. These static charges are often collected at the bottom of clouds, but exist in other environments as well.

[0058] Balloon **21** is extended into this strata of static charges **23** which are then attracted to conductive line **12** to flow to base unit **22** and then onto load **5**.

[0059] By increasing or decreasing the altitude of balloon **21** (defined by the length of the extended conductive line **12**), conductive line **12** is selectively exposed to varying densities and levels of the static charge strata, and by extension, the current flow or voltage is increased or decreased.

[0060] FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

[0061] Once the program starts **30**, the lightning sensor is checked to determine if lightning has occurred within the unsafe range **31**A, if it has, then the balloon is lowered **32**A, and the program continues monitoring the status of lightning until no lightning is detected.

[0062] When the lightning status is acceptable, then the current within the conductor is checked to see if the current is within the prescribed range **31**B. If the current is acceptable (within range) the program returns to check the lightning status **31**A; otherwise a determination is made to see if the current is above the prescribed range **31**C.

[0063] If the current is above the prescribed range, then the altitude of the balloon is withdrawn a set amount **32**B (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

[0064] If the current is below the prescribed range, then the altitude of the balloon is extended a set amount **32**C (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

[0065] In this manner of feed-back and minor adjustments in the altitude of the balloon, the current is maintained within a prescribed range which can be handled by the downstream electrical system.

[0066] As noted earlier, some embodiments of the invention monitor the voltage instead of the current.

[0067] FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

[0068] By maintaining the voltage being collected in a prescribed range, an electrical conversion system is easily designed. While FIGS. **4**A, **4**B, and **4**C illustrate some electrical configurations, those of ordinary skill in the art readily recognize a variety of other configurations which will serve the same function.

[0069] Referencing FIG. **4**A, Direct Current In (DC IN) **40** is buffered by a gang of capacitors **41** before being communicated to a DC/AC converter **42**. The DC/AC converter converts the direct current into a an alternating current suitable for placement over an existing electrical grid **43** such as normally found from a power-plant.

[0070] Those of ordinary skill in the art readily recognize a variety of DC/AC converters, including, but not limited to: U.S. Pat. No. 7,394,671, entitled "Controller IC, DC-AC Conversion Apparatus, and parallel running system of DC-AC Conversion Apparatuses" issued to Fukumoto, et al. on Jul. 1, 2008; and, U.S. Pat. No. 7,330,366, entitled "DC-AC Converter" issued to Lee, et al. on Feb. 12, 2008; all of which are incorporated hereinto by reference.

[0071] FIG. **4**B illustrates an electrical arrangement suitable for use in charging a battery. DC IN **40** is buffered by capacitor bank **41** before entering into a step down transformer **43**. Step down transformer **43** reduces the voltage so that the voltage can safely be introduced into battery **44** which is connected to ground **45** at the battery's other pole.

[0072] Those of ordinary skill in the art readily recognize a variety of batteries which will work in this capacity, including, but not limited to those described in: U.S. Pat. No. 7,378,181, entitled "Electric Storage Battery Construction and Manufacture" issued to Skinlo on May 27, 2008; U.S. Pat. No. 7,388,350, entitled "Battery with Electronic Compartment" issued to Wright on Jun. 17, 2008; U.S. Pat. No. 7,397,220, entitled "Connection Member and Battery Pack" issued to Uchida, et al. on Jul. 8, 2008; and, U.S. Pat. No. 7,375,492, entitled "Inductively Charged Battery Pack" issued to Calhoon, et al. on May 20, 2008; all of which are incorporated hereinto by reference.

[0073] In FIG. **4**C, DC IN **40** is fed into an adjustable rheostat **46** which is controlled by the controller so that the DC OUT **47** falls within a specified range.

[0074] FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

[0075] This type of conductive line is commonly called poly-wire and consists of multiple interwoven strands of plastic **50**A and **50**B woven into a cord or rope arrangement having intertwined therein exposed metal wires **51**A and **51**B. While this illustration shows two plastic strands and two metal wires, any number of possible combinations is possible.

[0076] The exposed metal wires **51**A and **51**B attract the atmospheric static charge and transmit the charge down to the base unit (not shown).

[0077] FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

[0078] This conductive line utilizes a tube **60** having an outer layer **62** of PET Film (Biaxially-oriented polyethylene terephtalate polyester film) which provides exceptionally high tensile strength and is chemically and dimensionally stable. The tube has an ideal diameter of between two and three inches.

[0079] An interior metal coating **61** provides an initial conduit for the flow of static charge. The static charge through the metal forces the tube to expand due to the repulsion experienced by like charges. Further, the flow of electricity causes the interior of the tube **60** to become ionized to provide an additional pathway for the atmospheric static charges to the base unit (not shown).

[0080] Because outer layer **62** provides a gas barrier, the resulting ionization is not dissipated by air currents, thereby providing a highly stable pathway.

[0081] FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

[0082] In this embodiment, controller box **70**, resting on insulating pad **72**, is in communication with the sensors as described above. Using the input from these sensors, when there is flow of electricity through the base unit, warning flashing light **71** is illuminated. To electrically neutralize the mechanism, switch **73** is activated to pass any existing current into the ground **74**.

[0083] FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

[0084] Referencing FIG. **8**A, enhanced electrical collection lead **82** is a wire mesh which is in electrical communication with conductive line **81** and balloon **80**. Because of the significant amount of metal exposed by enhanced electrical collection lead **82**, more static electricity from the atmosphere is drawn to the collection lead **82**, and then down conductive line **81** to the base unit (not shown).

[0085] Conductive lead **82** is positioned proximate to balloon **80**.

[0086] In FIG. **4**B, poly-wire **83** has enhanced electrical collection leads **84** wrapped therearound. Collection leads **84** have pointed ends **85**A and **85**B which have a propensity to attract more electricity than rounded ends do.

[0087] It is clear from the foregoing that the present invention captures an entirely new source of electrical energy.

What is claimed is:

1. A mechanism to tap an electrical source comprising:

a) a lighter than air balloon;

b) a direct current to alternating current transformer electrically connected via an electrically exposed tether to said lighter than air balloon and said base unit such that alternating current is communicated to said base unit; and,

c) a base unit having a spool of insulated conductive wire on a winch motor, one end of said insulated conductive wire electrically connected to an output of said direct current to alternating current transformer, said winch motor capable of selectively extending or withdrawing said insulated conductive wire from said spool, a second end of said insulated conductive wire electrically connected to a load.

2. The mechanism to tap an electrical source according to claim **1**, further including a controller, said controller selectively operating said winch motor based upon a voltage level in said insulated conductor wire.

3. The mechanism to tap an electrical source according to claim **2**,

a) further including an electrical flow sensor monitoring electrical flow through said conductor and generating an electrical flow indicia indicative of said electrical flow in said conductor, said electrical flow indicia communicated to said controller; and,

b) wherein said controller operates said winch motor such that said electrical flow indicia remains within a selected operating range.

4. The mechanism to tap an electrical source according to claim **3**, wherein said conductor further includes an insulated ground wire communicating between said direct current to alternating current transform and ground.

5. The mechanism to tap an electrical source according to claim **4**,

a) further including a lightning sensor generating a lightning presence indicia indicative of lightning within a prescribed range, said presence indicia being communicated to said controller; and,

b) wherein said controller, in response to said lightning presence indicia, operates said winch motor to withdraw substantially all of said insulated conductive wire onto said spool.

6. A mechanism comprising:

a) an airborne aircraft tethered by a conductive line secured to a winch capable of extending or withdrawing said conductive line from a spool; and,

b) a direct current to alternating current transformer connected via an electrically conductive tether to said lighter than air balloon.

7. The mechanism according to claim **6**, further including a base unit selectively extending and withdrawing an insulated conductive wire to an output of said direct current to alternating current transformer.

8. The mechanism according to claim **7**, wherein a second end of said insulated conductive wire is electrically connected to an earthbound load.

9. The mechanism according to claim **8**,

a) further including a lightning sensor communicating with said base unit; and,

b) wherein said base unit, in response to selected signals from said lightning sensor withdraws substantially all of said insulated conductive wire.

10. A mechanism to tap an electrical source comprising:

a) a lighter than air balloon;

b) a direct current to alternating current transformer, said transformer having an electrically exposed tether connected to said lighter than air balloon such that direct current from the atmosphere is communicated to said transformer;

b) a base unit having a spool of conductive line on a winch motor, one end of said conductive material secured to said transformer so that alternating current is provided to said conductive line, said winch motor capable of adjusting an altitude of said lighter than air balloon by selectively extending or withdrawing said conductive line from said spool;

c) an insulator electrically isolating said conductive line from ground; and,

d) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load;

e) a sensor array having,

1) an electrical flow sensor monitoring electrical flow through said conductor, and,

2) a lightning sensor monitoring existence of proximate lightning; and,

f) a controller responsive to signals from said sensor array to selectively operate said winch motor.

11. The mechanism to tap an electrical source according to claim **10**,

a) further including a warning light; and,

b) wherein said controller activates said warning light when said electrical flow in said conductor is non-zero.

12. The mechanism to tap an electrical source according to claim **10**, further including an electrical collection enhancement lead in electrical contact with said conductive material, said electrical collection enhancement lead configured to attract static electricity and positioned proximate to said lighter than air balloon.

* * * * *

US 20100008011A1

(19) **United States**

(12) **Patent Application Publication**

Ogram

(10) **Pub. No.: US 2010/0008011 A1**

(43) **Pub. Date: Jan. 14, 2010**

(54) **ATMOSPHERIC STATIC ELECTRICITY COLLECTOR**

(76) Inventor: **Mark Ellery Ogram**, Tucson, AZ (US)

Correspondence Address:
Mark E. Ogram (30343)
Suite 110-186, 6990 E. 22nd Street
Tucson, AZ 85710-5192 (US)

(21) Appl. No.: **12/321,306**

(22) Filed: **Jan. 16, 2009**

Related U.S. Application Data

(63) Continuation-in-part of application No. 12/218,297, filed on Jul. 14, 2008.

Publication Classification

(51) **Int. Cl.**
H05F 7/00 (2006.01)

(52) **U.S. Cl.** **361/222**

(57) **ABSTRACT**

An antenna for the collection of atmospheric static electricity in which an electrically conductive hub is suspended from a balloon or blimp via a tether. The hub is either solid or uses a spoke/arm arrangement. A number of rods extend from the hub enhance the collection of atmospheric static electricity. The collected atmospheric electricity is conducted from the rods to an electrical connection where the electricity is conducted to earth via a conductive line.

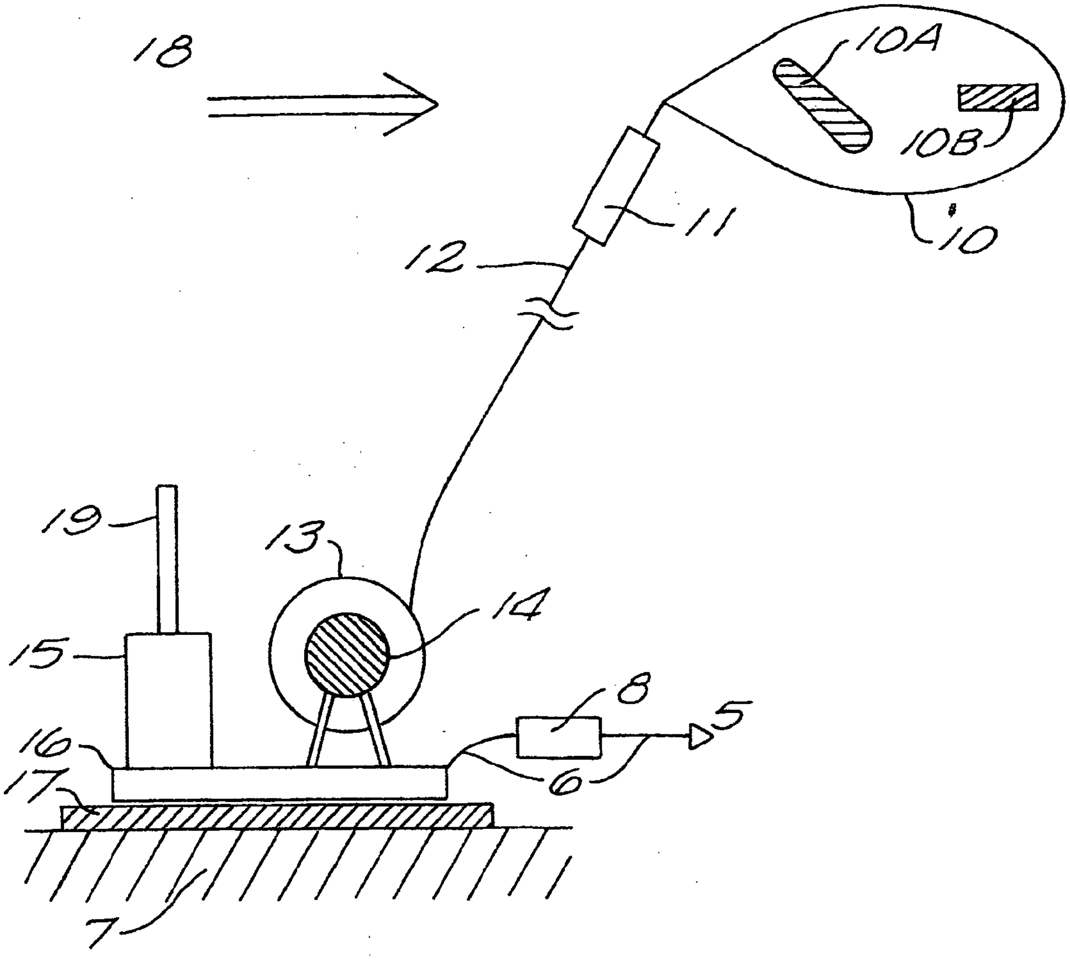

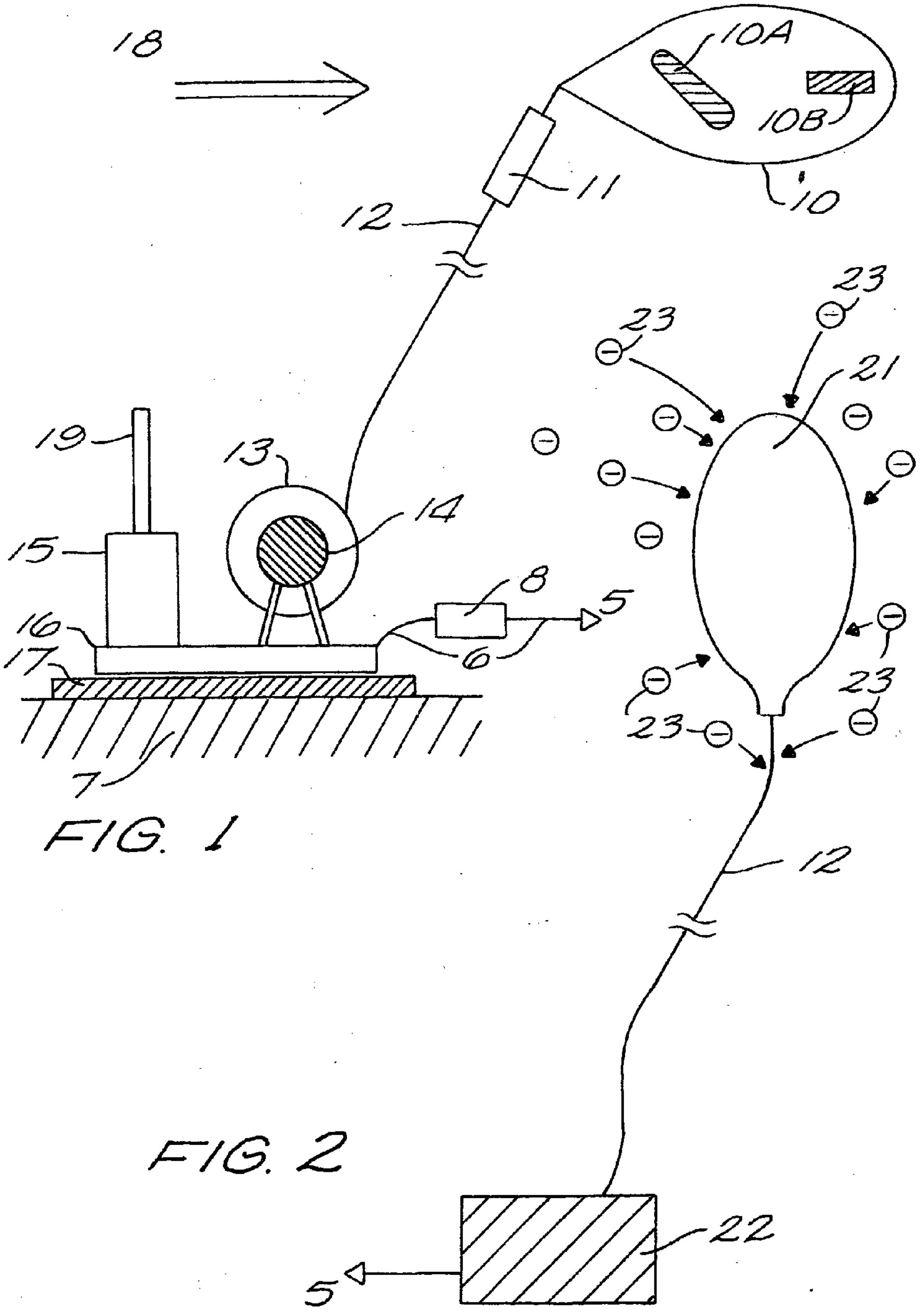
18
10A
10B
10
11
12
19
13
14
15
8
5
16
6
17
7
23
21
12
22
5

FIG. 1

FIG. 2

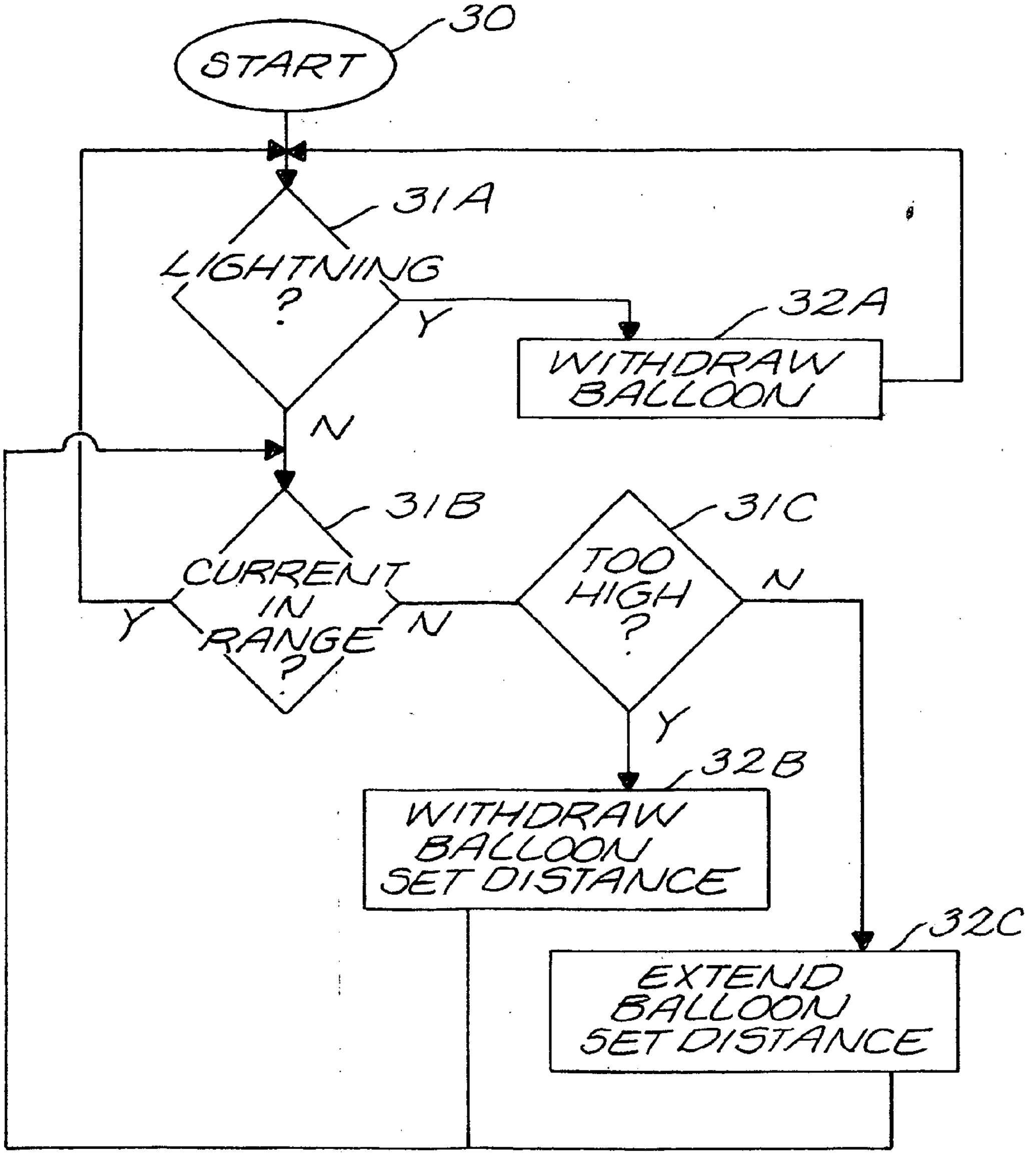
30
START
31A
LIGHTNING ?
Y
32A
WITHDRAW BALLOON
N
31B
CURRENT IN RANGE ?
Y
N
31C
TOO HIGH ?
N
Y
32B
WITHDRAW BALLOON SET DISTANCE
32C
EXTEND BALLOON SET DISTANCE

FIG. 3

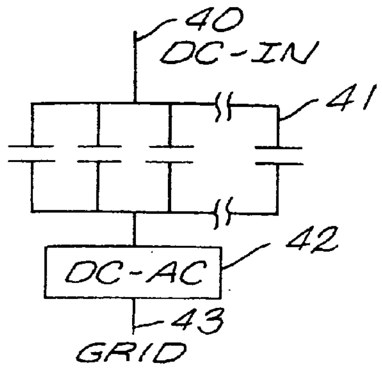
40
DC-IN
41
DC-AC
42
43
GRID

FIG. 4A

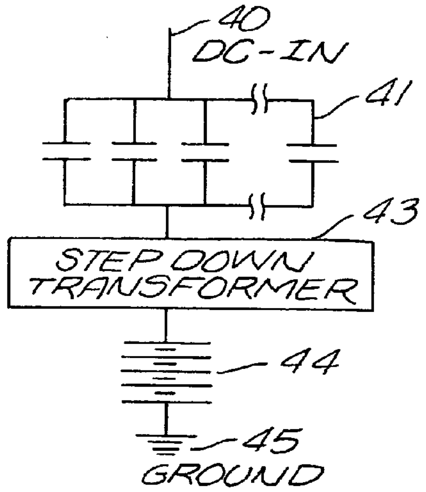
40
DC-IN
41
43
STEP DOWN TRANSFORMER
44
45
GROUND

FIG. 4B

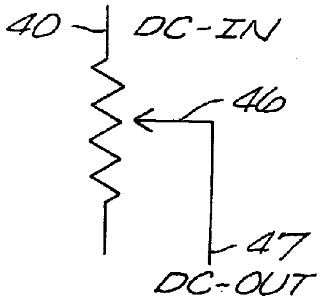
40
DC-IN
46
47
DC-OUT

FIG. 4C

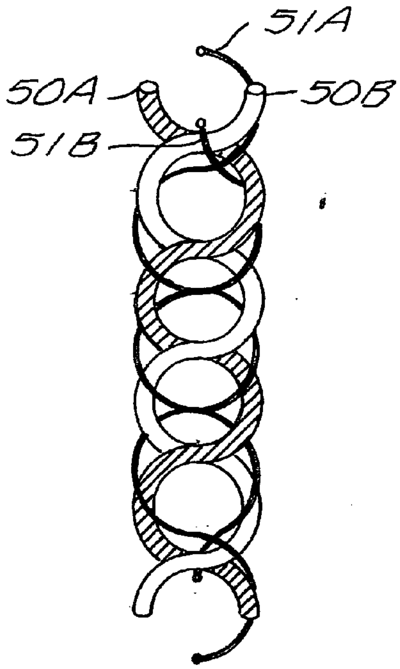
51A
50A
50B
51B

FIG. 5

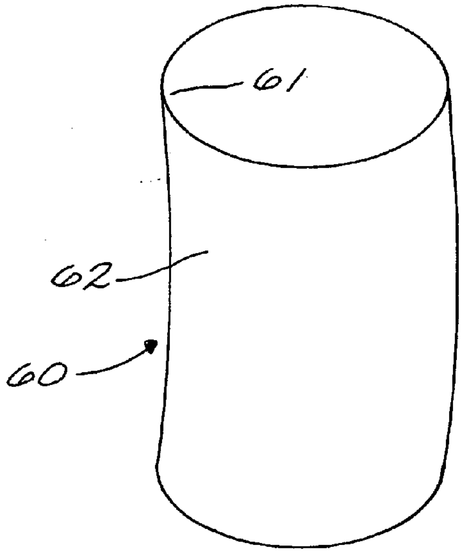

FIG. 6A

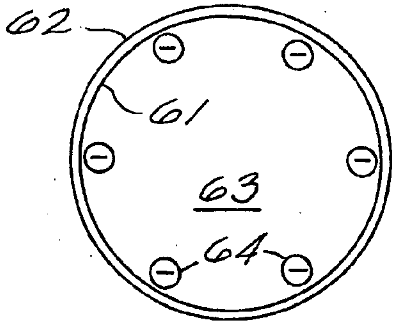

FIG. 6B

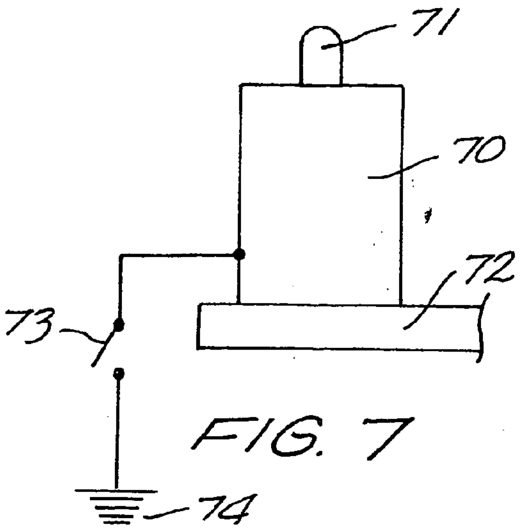
71
70
72
73
74

FIG. 7

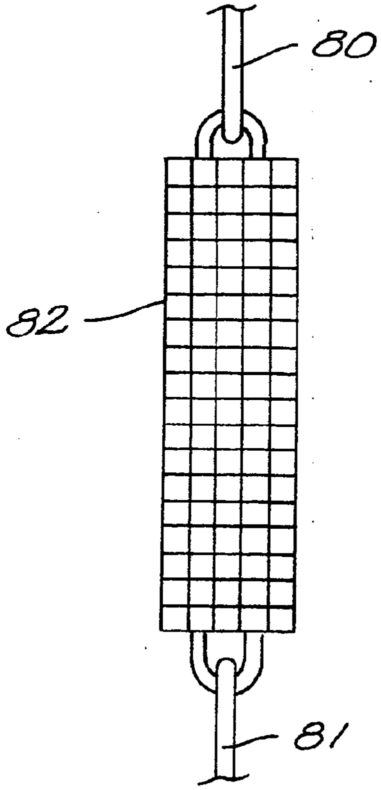

FIG. 8A

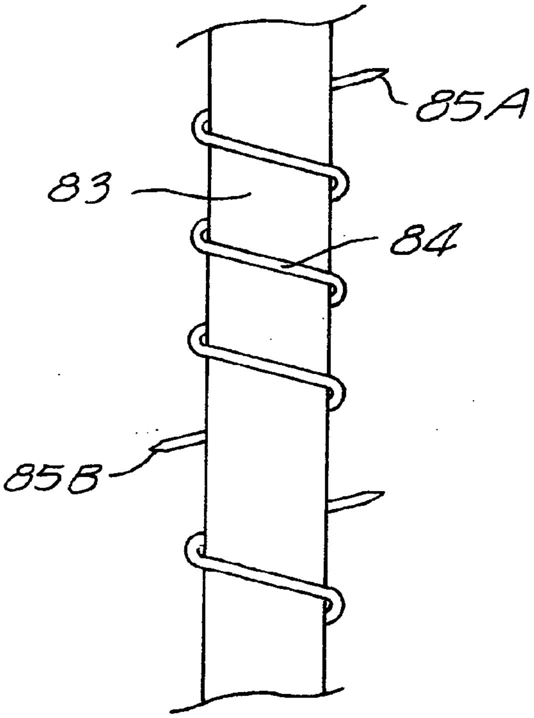
85A
83
84
85B

FIG. 8B

FIG. 9A

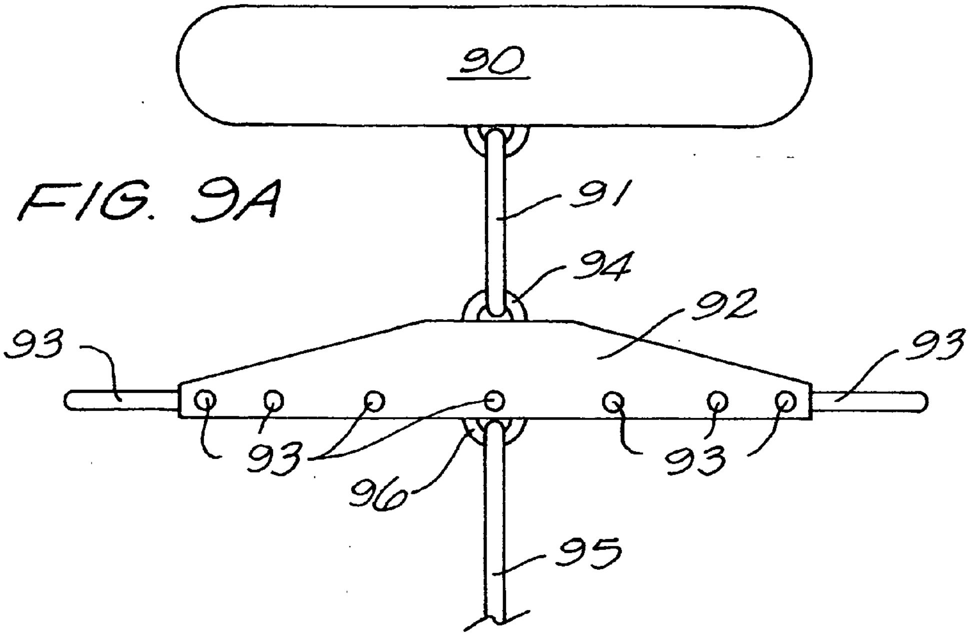
90
91
94
92
93
93
93
93
93
96
95

FIG. 9B

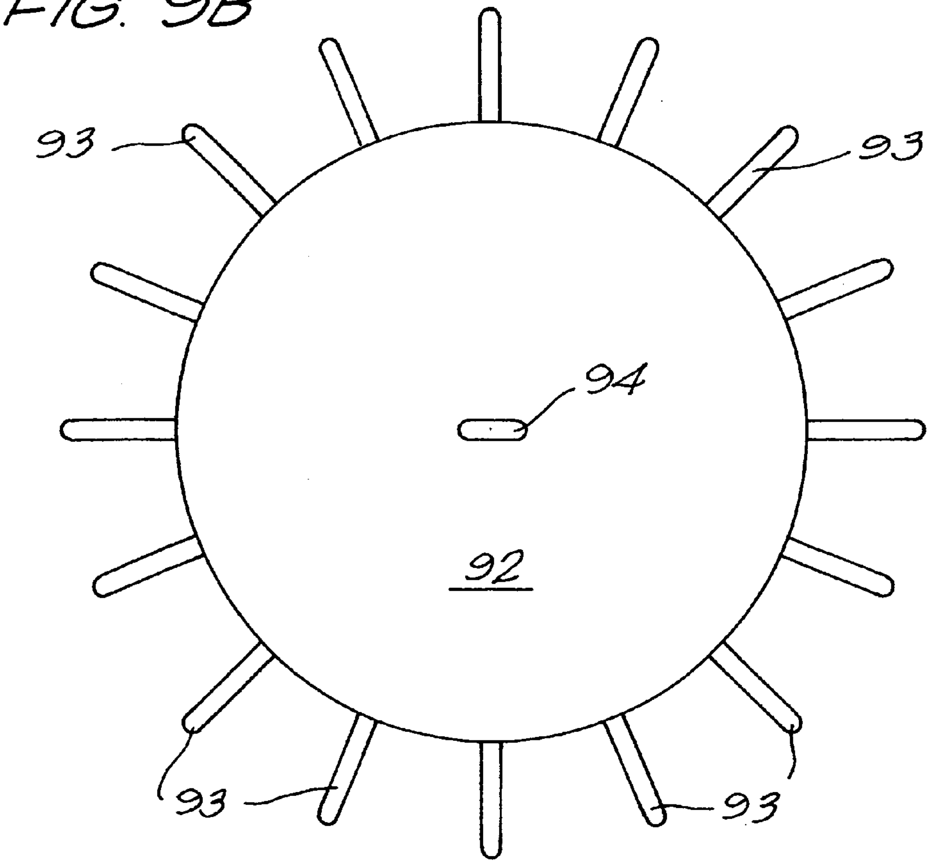
93
93
94
92
93
93

FIG. 10

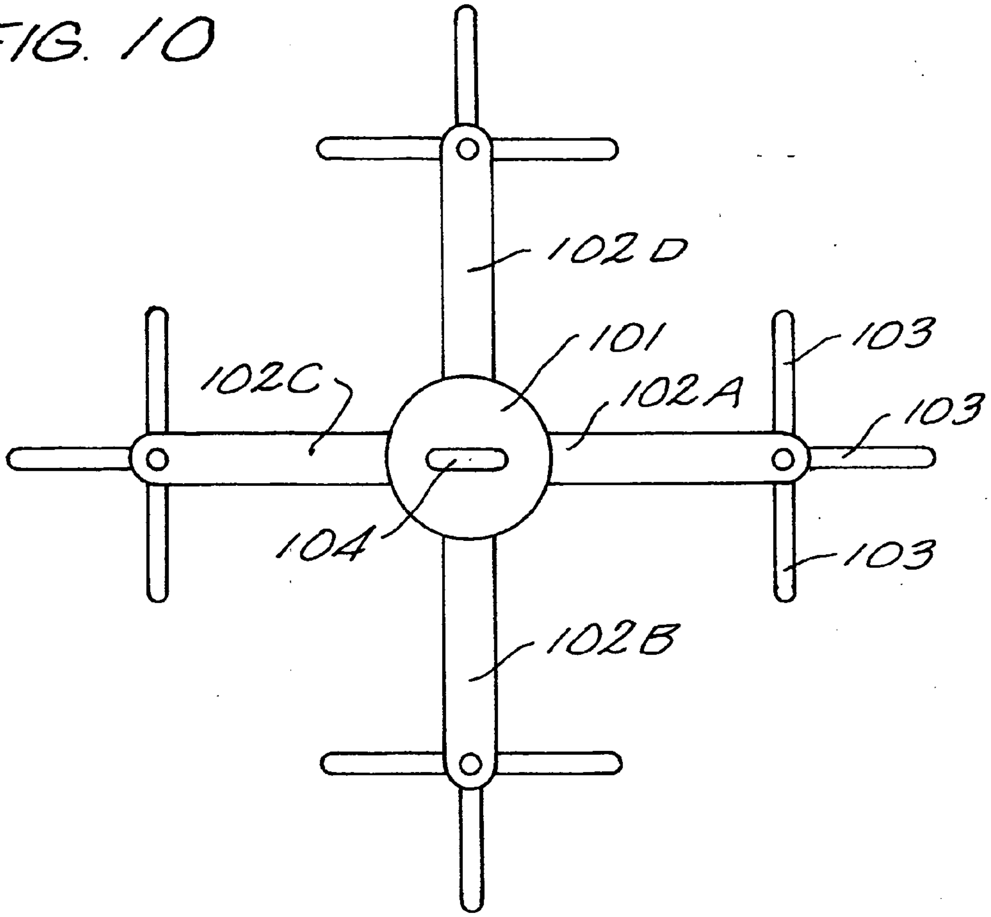
102D
101
103
102C
102A
103
104
103
102B

FIG. 11

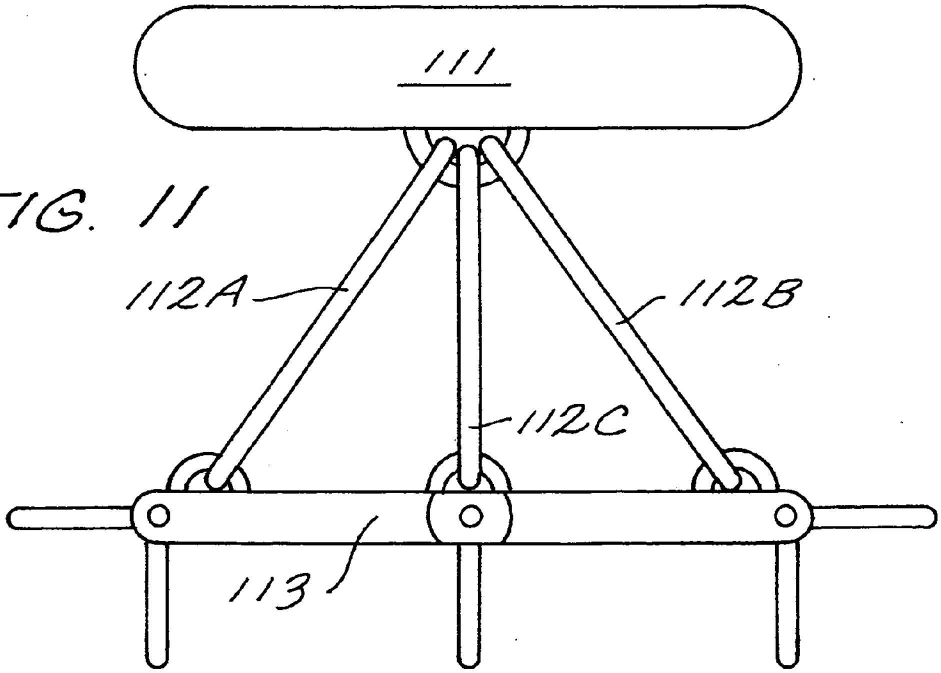
111
112A
112B
112C
113

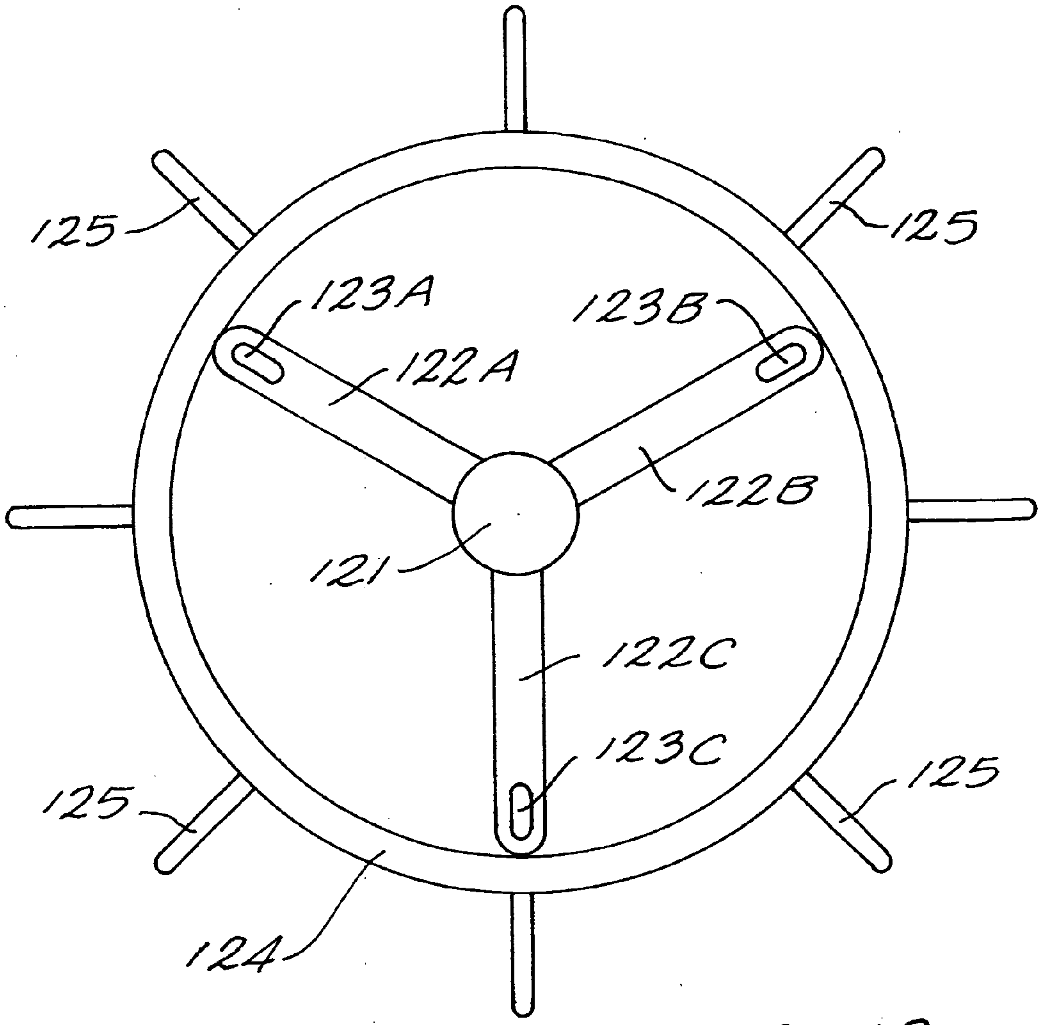
125
125
123A
123B
122A
122B
121
122C
123C
125
125
124

FIG. 12

ATMOSPHERIC STATIC ELECTRICITY COLLECTOR

BACKGROUND OF THE INVENTION

[0001] This invention relates to the production of electrical energy and more particularly to the collection of static electricity from the atmosphere.

[0002] This is a continuation in part of U.S. patent application Ser. No. 12/218,297, filed on Jul. 14, 2008, and entitled, "Atmospheric Electrical Generator".

[0003] Everyone is familiar with Benjamin Franklin's kite experiment of 1752. Using a kite whose string had become wet, negative charges from the passing clouds flowed into the string, down to the suspended key, and then into a Leyden jar via a thin metal wire. Franklin was protected by a dry silk string; but, when Franklin's knuckle came too close to the key, he received a strong shock. Fortunately, Benjamin Franklin was not killed, others who tried this same experiment were not so lucky.

[0004] Since then, the formation of lightning has remained something of a mystery. Lightning bolts are triggered when a negatively charged cloud base induces a positive charge from the ground, thereby forming a "pathway" for the discharge of the collected electrical energy.

[0005] Lightning travels up to 60,000 miles per hour with a flash that is brighter than ten million 100-watt lightbulbs. This wattage is as much power as is produced by all of the electricity plants in the United States and with a voltage of up to 300 million volts.

[0006] It is this very fact, the power within lightning is immense, that has prevented any successful collection of the electrical energy from lightning. The electricity in lightning is far too extreme for current technology to harness.

[0007] While lightning has attracted a energy starved industrial world, no one has developed any technique to harness this naturally occurring electrical source.

[0008] It is clear there is a continuing need for an electrical source other than carbon-based fuels and that the naturally occurring electricity in the atmosphere is being ignored.

SUMMARY OF THE INVENTION

[0009] The invention is a mechanism which taps into the naturally occurring static electricity in the atmosphere. Whereas heretofore, the attempt to garner electricity from the atmosphere has focused exclusively on capturing lightning, the present invention syphons off the static electricity which is generated from any agitated air and voids lightning.

[0010] Lightning is only the final discharge of the static electricity, whether that lightning is intra-cloud lightning, cloud-to-ground lightning, or inter-cloud lightning. Other types of final discharges are known as heat lightning, summer lightning, sheet lightning, ribbon lightning, silent lightning, ball lightning, bead lightning, elves, jets, and sprites. Well before these discharges are observed, as the atmosphere becomes agitated by wind or thermal, static electricity is being generated.

[0011] The present invention recognizes that this static electricity is being formed and creates a mechanism to capture it.

[0012] The mechanism of this invention utilizes an aircraft such as a lighter than air balloon. While the preferred embodiment uses a foil balloon, a variety of other aircraft are obvious to those of ordinary skill in the art, including, but not limited to: gliders, rubber balloons (such as weather balloons), biaxially-oriented polyethylene terephthalate polyester film balloons, and latex balloons.

[0013] Within this discussion, the balloon is referenced, but, the invention is not intended to be limited solely to balloons.

[0014] The balloon is sent aloft and is tethered by a conductive line. In this context, the conductive line may be any obvious to those of ordinary skill in the art. For the preferred embodiment, the conductive line is a generically referred to as a "poly-rope" and is commercially available through a variety of sources. A suitable conductive line is described in U.S. Pat. No. 5,203,542, entitled "Apparatus for Improved Electric Fence Wire Construction for use with Intensive Grazing" issued Apr. 20, 1993, to Coley, et al. and incorporated hereinto by reference.

[0015] The conductive line is played out of a winch to control the altitude of the balloon. The motor controlling the winch is able to reverse direction to both extend and withdraw the conductive line which is wrapped around a spool on the winch. The winch/spool combination are part of a base unit.

[0016] In some embodiments of the invention, the spool is constructed of rubber so as to insulate the conductive line from the winch assembly. In this embodiment, only the conductive line is charged by the atmospheric static electricity while the winch remains neutral.

[0017] In yet another embodiment, the winch/spool are part of a base unit which is itself isolated from the ground by an insulator. In this embodiment, the entire base unit is charged by the atmospheric static electricity.

[0018] A conductor, such as an insulated wire, is electrically connected to the conductive line. In one embodiment, where the conductive line is electrically isolated from the spool and winch motor, the conductor is connected to the conductive line. In the embodiment where the conductive line is electrically connected to the base unit, then the conductor is connected anywhere on a metalic base unit.

[0019] The other end of the conductor is connected to a load. The load in this case can be any of a variety of electrical loads well known to those of ordinary skill in art, including, but not limited to a motor, a battery system, or the electrical grid for the system.

[0020] In the preferred embodiment, a sensor array is used to monitor the activities both at the base unit (such as electrical flow within the conductor) and in the surrounding locale.

[0021] A sensor monitoring the electrical flow (i.e. voltage and/or current) within the conductor is used to monitor the electrical activity within the conductor.

[0022] In the preferred embodiment, a lightning sensor monitors for lightning activity within the locale. As noted earlier, the electrical characteristic of lightning is so extreme that ideally this discharge is avoided as it might damage the mechanism of this invention.

[0023] The sensor array is utilized by a controller, such as microprocessor, programmed to operate the mechanism as outlined herein.

[0024] The controller operates the winch motor to extend or withdraw the conductive line and by extension the altitude of the balloon. The controller is programmed to operate the winch by monitoring the electrical characteristics of the conductor and adjusting the balloon's altitude to maintain these characteristics within the conductor within a preset range.

[0025] This preset range is established either in the base programming of the controller or is established by an operator of the system.

[0026] As example, by controlling the amount of current being withdrawn from the atmosphere, the mechanism operates within a safe range and also provides a relatively stable current flow from which a variety of activities can take place (such as DC-AC conversion).

[0027] The controller also utilizes the lightning sensor to protect the mechanism from a lightning strike. Should lightning be detected within a pre-determined range (as established by the software or defined by an operator), then the balloon is pulled down to minimize the risk of damage from a lightning strike.

[0028] An aspect of the present invention is the use of an antenna which are used to collect the atmospheric static electricity. The antenna is shaped as a hub which is suspended from the blimp/balloon. The hub is ideally spoked shaped although an alternative embodiment uses a solid hub.

[0029] A number of rods extend from the hub so as collect atmospheric static electricity These rods are ideally rounded at the ends to enhance the attraction of the atmospheric static electricity.

[0030] The collected atmospheric electricity is conducted from the rods to an electrical connection on the hub where the electricity is conducted to a power plant on earth such as described above.

[0031] The invention, together with various embodiments thereof will be more fully explained by the following description of the accompanying drawings.

DRAWINGS IN BRIEF

[0032] FIG. **1** diagrams the preferred embodiment of the invention.

[0033] FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

[0034] FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

[0035] FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

[0036] FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

[0037] FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

[0038] FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

[0039] FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

[0040] FIGS. **9**A and **9**B are side views and top views of an embodiment of the antenna used to collected atmospheric electricity.

[0041] FIG. **10** is a top view of an alternative embodiment of the antenna of this invention.

[0042] FIG. **11** is a side view of yet another alternative embodiment of the antenna used to collect atmospheric electricity.

[0043] FIG. **12** is the preferred embodiment of the antenna of this invention.

DRAWINGS IN DETAIL

[0044] FIG. **1** diagrams the preferred embodiment of the invention.

[0045] Balloon **10** is an aircraft which, in this illustration, is a lighter than air balloon. Wings **10**A, extending from the body of balloon **10**, provide additional lift in air flow **18**. Tail **10**B helps to stabilize balloon **10**.

[0046] Balloon **10** is tethered to the ground via conductive line **12**. As noted earlier, a variety of configurations and materials are available to serve as conductive line **12**. In this illustration, a poly-wire is used. Poly-wire is commercially available through a variety of vendors, including, but not limited to: Jeffers Livestock and Sareba Systems, Inc. of Ellendale, Minn.

[0047] In this embodiment, located proximate to balloon **10**, is an electrical collection enhancement lead **11** which assists in the collection of the static electrical charge in the atmosphere. Electrical collection enhancement lead **1 1** is configured to attract the static charge and conduct the charge into the conductive line **12**.

[0048] The electricity flows down the conductive line into spool **13**, where the conductive line **12** is collected and either withdrawn or dispensed through operation of winch motor **14**.

[0049] Winch motor **14** and spool **13** are mounted onto base unit **16** which is electrically isolated from ground **7** using insulator **17**. Note, in this embodiment of the invention, when electricity is being collected from the atmosphere, the entire base unit **16** becomes charged. In another embodiment of the invention, spool **13** is constructed of rubber, thereby preventing base unit **16** from becoming charged, thereby restricting the charging from the atmosphere to only conductive line **12**.

[0050] In this embodiment, conductor **6** is connected to base unit **16** (since the entire base unit **16** is charged and the base unit is metalic) to communicate the electrical current to load **5**. Conductor **6** is ideally an insulated wire.

[0051] The electrical current through conductor **6** is measured using sensor **8**.

[0052] In the alternative embodiment discussed above, where only the conductive line **12** is charged, then conductor **6** is connected to conductive line **12**.

[0053] Controller **15**, located in this embodiment on base unit **16**, operates winch motor **14** in response to signals from sensor **8** (measuring the current being discharged to load **5**) to maintain the current flow within a pre-defined range. As the current flow diminishes, then the conductive line **12** extended from spool **13** to increase the altitude of balloon **10** to that more static charge from the atmosphere is gathered; as the current flow falls exceeds a preset level, conductive line **12** is withdrawn onto spool **13** to decrease the static charge being collected from the atmosphere.

[0054] The range of current flow through conductor **6** is ideally set by the program, although some embodiments of the invention permit an operator to establish this range of operation.

[0055] In an alternative embodiment, the sensor monitoring conductor **6** monitors the voltage therein.

[0056] In the preferred embodiment of the invention, controller **15** is also equipped with a lightning sensor **19**. In this embodiment, when lightning is sensed within a preset range, then substantially all of conductive line **12** is wound onto spool **13** to pull balloon **10** near the ground and protect the entire mechanism from being damaged from a lightning discharge.

[0057] In the preferred embodiment, the "safe" distance form lightning is set in the programming of controller **15** and is ideally two miles; other embodiments permit the operator to "safe" distance.

[0058] There are a variety of lightning sensors well known to those of ordinary skill in the art, including, but not limited to those described in: U.S. Pat. No. 7,016,785, entitled "Lightning Detection" issued to Makela, et al. on Mac. 21, 2006; U.S. Pat. No. 6,829,911, entitled "Lightning Detection and Prediction Alarm Device" issued to Jones, et al. on Dec. 7, 2004; U.S. Pat. No. 7,200,418, entitled "Detection of Lightning" issued to Karikuranta, et al. on Apr. 3, 2007; and U.S. Pat. No. 6,961,662, entitled "Systems and Methods for Spectral Corrected Lightning Detection" issued to Murphy on Nov. 1, 2005; all of which are incorporated hereinto by reference.

[0059] In another embodiment of the invention, controller **15** is not located on base unit **16**, rather it is remote and communicates its control signals to winch motor **14** using radio waves.

[0060] FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

[0061] Static charges **23** are generated in the atmosphere by agitated air. These static charges are often collected at the bottom of clouds, but exist in other environments as well.

[0062] Balloon **21** is extended into this strata of static charges **23** which are then attracted to conductive line **12** to flow to base unit **22** and then onto load **5**.

[0063] By increasing or decreasing the altitude of balloon **21** (defined by the length of the extended conductive line **12**), conductive line **12** is selectively exposed to varying densities and levels of the static charge strata, and by extension, the current flow or voltage is increased or decreased.

[0064] FIG. 3 is a flowchart of the operation of the controller for the preferred embodiment of the invention.

[0065] Once the program starts **30**, the lightning sensor is checked to determine if lightning has occurred within the unsafe range **31**A, if it has, then the balloon is lowered **32**A, and the program continues monitoring the status of lightning until no lightning is detected.

[0066] When the lightning status is acceptable, then the current within the conductor is checked to see if the current is within the prescribed range **31**B. If the current is acceptable (within range) the program returns to check the lightning status **31**A; otherwise a determination is made to see if the current is above the prescribed range **31**C.

[0067] If the current is above the prescribed range, then the altitude of the balloon is withdrawn a set amount **32**B (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

[0068] If the current is below the prescribed range, then the altitude of the balloon is extended a set amount **32**C (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

[0069] In this manner of feed-back and minor adjustments in the altitude of the balloon, the current is maintained within a prescribed range which can be handled by the downstream electrical system.

[0070] As noted earlier, some embodiments of the invention monitor the voltage instead of the current.

[0071] FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

[0072] By maintaining the voltage being collected in a prescribed range, an electrical conversion system is easily designed. While FIGS. **4**A, **4**B, and **4**C illustrate some electrical configurations, those of ordinary skill in the art readily recognize a variety of other configurations which will serve the same function.

[0073] Referencing FIG. **4**A, Direct Current In (DC IN) **40** is buffered by a gang of capacitors **41** before being communicated to a DC/AC converter **42**. The DC/AC converter converts the direct current into a an alternating current suitable for placement over an existing electrical grid **43** such as normally found from a power-plant.

[0074] Those of ordinary skill in the art readily recognize a variety of DC/AC converters, including, but not limited to: U.S. Pat. No. 7,394,671, entitled "Controller IC, DC-AC Conversion Apparatus, and parallel running system of DC-AC Conversion Apparatuses" issued to Fukumoto, et al. on Jul. 1, 2008; and, U.S. Pat. No. 7,330,366, entitled "DC-AC Converter" issued to Lee, et al. on Feb. 12, 2008; all of which are incorporated hereinto by reference.

[0075] FIG. **4**B illustrates an electrical arrangement suitable for use in charging a battery. DC IN **40** is buffered by capacitor bank **41** before entering into a step down transformer **43**. Step down transformer **43** reduces the voltage so that the voltage can safely be introduced into battery **44** which is connected to ground **45** at the battery's other pole.

[0076] Those of ordinary skill in the art readily recognize a variety of batteries which will work in this capacity, including, but not limited to those described in: U.S. Pat. No. 7,378,181, entitled "Electric Storage Battery Construction and Manufacture" issued to Skinlo on May 27, 2008; U.S. Pat. No. 7,388,350, entitled "Battery with Electronic Compartment" issued to Wright on Jun. 17, 2008; U.S. Pat. No. 7,397,220, entitled "Connection Member and Battery Pack" issued to Uchida, et al. on Jul. 8, 2008; and, U.S. Pat. No. 7,375,492, entitled "Inductively Charged Battery Pack" issued to Calhoon, et al. on May 20, 2008; all of which are incorporated hereinto by reference.

[0077] In FIG. **4**C, DC IN **40** is fed into an adjustable rheostat **46** which is controlled by the controller so that the DC OUT **47** falls within a specified range.

[0078] FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

[0079] This type of conductive line is commonly called poly-wire and consists of multiple interwoven strands of plastic **50**A and **50**B woven into a cord or rope arrangement having intertwined therein exposed metal wires **51**A and **51**B. While this illustration shows two plastic strands and two metal wires, any number of possible combinations is possible.

[0080] The exposed metal wires **51**A and **51**B attract the atmospheric static charge and transmit the charge down to the base unit (not shown).

[0081] FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

[0082] This conductive line utilizes a tube **60** having an outer layer **62** of PET Film (Biaxially-oriented polyethylene terephtalate polyester film) which provides exceptionally high tensile strength and is chemically and dimensionally stable. The tube has an ideal diameter of between two and three inches.

[0083] An interior metal coating **61** provides an initial conduit for the flow of static charge. The static charge through the metal forces the tube to expand due to the repulsion experienced by like charges. Further, the flow of electricity causes the interior of the tube **60** to become ionized to provide an additional pathway for the atmospheric static charges to the base unit (not shown).

[0084] Because outer layer **62** provides a gas barrier, the resulting ionization is not dissipated by air currents, thereby providing a highly stable pathway.

[0085] FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

[0086] In this embodiment, controller box **70**, resting on insulating pad **72**, is in communication with the sensors as described above. Using the input from these sensors, when there is flow of electricity through the base unit, warning flashing light **71** is illuminated. To electrically neutralize the mechanism, switch **73** is activated to pass any existing current into the ground **74**.

[0087] FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

[0088] Referencing FIG. **8**A, enhanced electrical collection lead **82** is a wire mesh which is in electrical communication with conductive line **81** and balloon **80**. Because of the significant amount of metal exposed by enhanced electrical collection lead **82**, more static electricity from the atmosphere is drawn to the collection lead **82**, and then down conductive line **81** to the base unit (not shown).

[0089] Conductive lead **82** is positioned proximate to balloon **80**.

[0090] In FIG. **8**B, poly-wire **83** has enhanced electrical collection leads **84** wrapped therearound. Collection leads **84** have pointed ends **85**A and **85**B which have a propensity to attract more electricity than rounded ends do.

[0091] FIGS. **9**A and **9**B are side views and top views of an embodiment of the antenna used to collected atmospheric electricity.

[0092] Blimp **90** supports antenna **92** via tether **91** which is attached to antenna **92** by connection **94**. In this embodiment, antenna **92** is made of electrically conductive material and is ideally light in weight to lessen the payload requirements on blimp **90**.

[0093] Encircling antenna **92** and extending therefrom are rods **93** which are rounded at their distal ends so as to enhance attraction of the atmospheric static electricity. The static electricity is communicated to connection **96** and then to electrical line **95** which communicates the electricity to the ground based station (not shown) as described above.

[0094] FIG. **10** is a top view of an alternative embodiment of the antenna of this invention.

[0095] In this embodiment of the antenna a central hub **101** has the blimp connection **104** (capable of receiving the tether to the blimp) secured thereto. In this illustration, four arms **102**A, **102**B, **102**C, and **102**D extend from hub **101**. Hub **101**, and arms **102**A, **102**B, **102**C, and **102**D are all electrically conductive in this embodiment.

[0096] At the end of each arm are rods **103** which are used to enhance the collection of the atmospheric static electricity.

[0097] Although this embodiment illustrates four arms, the invention is not intended to be limited to four arms, rather, any number of arms may be used and the number of rods extending from the distal ends of the arms also varies.

[0098] FIG. **11** is a side view of yet another alternative embodiment of the antenna used to collect atmospheric electricity.

[0099] In this embodiment, several tethers **112**A, **112**B, and **112**C, are used to secure the antenna **113** to the blimp **111**. This arrangement of several tethers provides heightened stability of the antenna by reducing the affects wind will have on the antenna.

[0100] FIG. **12** is the preferred embodiment of the antenna of this invention.

[0101] In this embodiment of the antenna, arms **122**A, **122**B, and **122**C extend from a central hum and are electrically connected to rim **124**. Tether connectors **123**A, **123**B, and **123**C, are used to secure the antenna to the blimp or balloon.

[0102] Rods **125** extend from rim **124** to increase the collection of the static charges in the atmosphere.

[0103] It is clear from the foregoing that the present invention captures an entirely new source of electrical energy.

What is claimed is:

1. An antenna for the collection of atmospheric static electricity comprising:
- a) a hub being electrically conductive and having,
  - 1) a balloon tether, and,
  - 2) an electrical connection;
- b) at least three electrically conductive arms, a first end of each of said conductive arms electrically connected to said hub; and,
- b) a plurality of rods, a first end of each of said rods being electrically connected to a second end of att least one of said at least three arms.

2. The antenna according to claim **1**, wherein each of said plurality of rods has a rounded second end.

3. The antenna according to claim **2**, wherein at least two rods are attached to the second end of each of said at least three arms.

4. The antenna according to claim **1**, further including an electrically conductive rim connected to a second end of said electrically conductive arms.

5. The antenna according to claim **4**, wherein at least a portion of said plurality of electrically conductive arms are connected substantially at right angles to an exterior of said electrically conductive rim.

6. An antenna for the collection of atmospheric static electricity comprising:
- a) a radial arrangement of at least three electrically conductive arms; and,
- b) a plurality of rods, each of said rods being electrically connected to at least one of said at least three arms.

7. The antenna according to claim **6**, wherein each of said plurality of rods has a rounded end distal from said electrically conductive arm.

8. The antenna according to claim **7**, further including a balloon tether connection secured to the first end of said at least three arms.

9. The antenna according to claim **8**, further including an electrical connection secured to the first end of said at least three arms.

10. The antenna according to claim **9**, wherein said at least two rods attached to said at least three arms is five rods.

11. The antenna according to claim **9**, further including an electrically conductive rim connected to a second end of said electrically conductive arms.

12. The antenna according to claim **11**, wherein said plurality of arms are electrically connected to said electrically conductive rim.

13. The antenna according to claim **12**, wherein at least a portion of said plurality of rods are connected to an exterior of said electrically conductive rim substantially at right angles thereto.

14. An antenna for the collection of atmospheric static electricity comprising:

a) a generally circular electrically conductive body;

b) a a balloon tether connected to a first side of said generally circular electrically conductive body; and,

c) an electrical connection connected to a second side of said generally circular electrically conductive body.

15. The antenna according to claim **14**, wherein said generally circular electrically conductive body includes:

a) an electrically conductive hub;

b) at least three electrically conductive arms, each of said arms connected at a first end to said hub; and,

c) an electrically conductive rim connected to the second end of each of said at least three arms.

16. The antenna according to claim **15**, further including a plurality of electrically conductive rods, a first end of each of said rods connected to said rim and extending therefrom substantially at rights angles thereto.

* * * * *

US 20120274144A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2012/0274144 A1**

Gasparini et al. (43) **Pub. Date: Nov. 1, 2012**

(54) SYSTEM AND METHOD FOR EFFICIENTLY HARVESTING ENVIRONMENTAL ENERGY

(75) Inventors: **Alessandro Gasparini**, Cusano Milanino (IT); **Stefano Ramorini**, Arluno (IT); **Giorgio Massimiliano Membretti**, Milano (IT)

(73) Assignee: **STMICROELECTRONICS S.R.L.**, Agrate Brianza (MB) (IT)

(21) Appl. No.: **13/456,814**

(22) Filed: **Apr. 26, 2012**

(30) **Foreign Application Priority Data**

Apr. 29, 2011 (IT) TO2011A000379

Publication Classification

(51) **Int. Cl.**
H02J 1/00 (2006.01)

(52) **U.S. Cl.** **307/80**

(57) ABSTRACT

A transducer converts energy coming from an energy source into an electrical signal for storage as electrical energy on a first storage element. A switch is selectively actuated to pass electrical energy from the first storage element to a second storage element. The selective actuation of the switch is driven by sensing electrical energy stored in the second storage element. The switch is closed when electrical energy in the second storage element is sensed to fall below a first threshold. The switch is opened when electrical energy in the second storage element is sensed to rise above a second threshold.

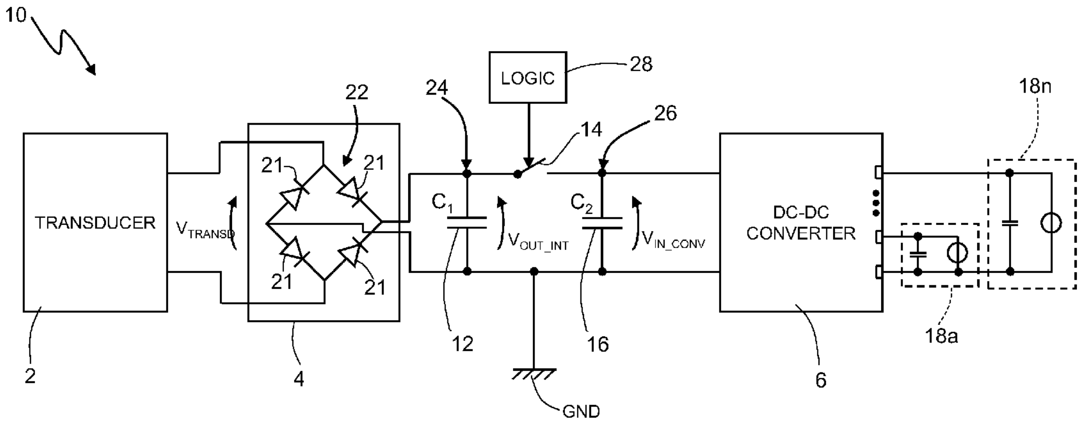

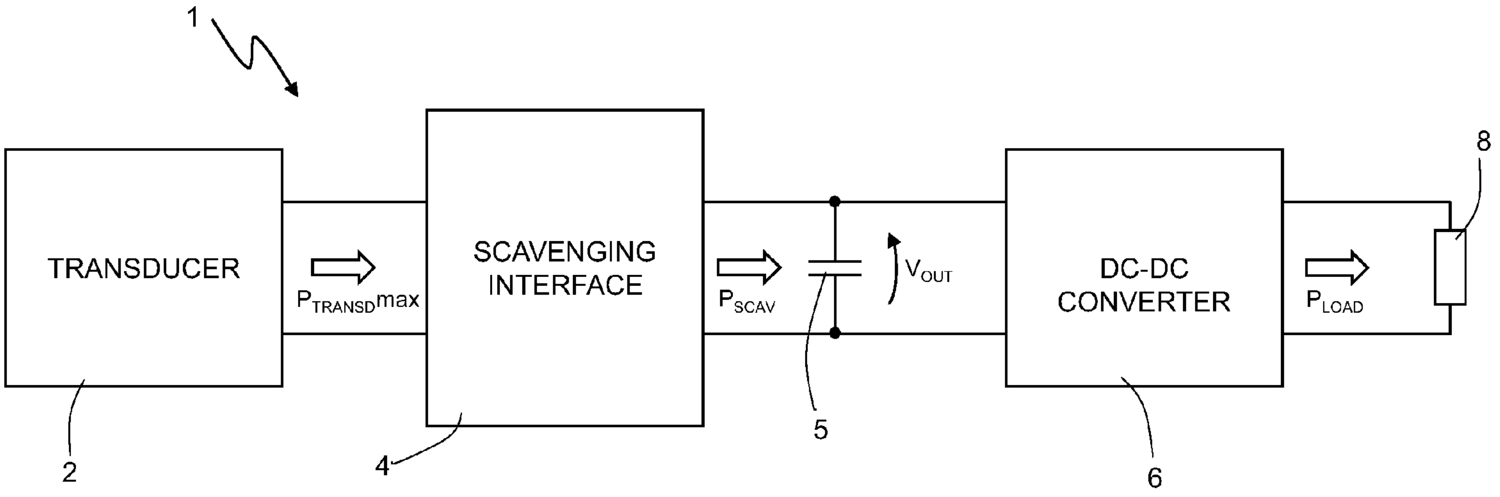
1
TRANSDUCER
$P_{TRANSD}max$
SCAVENGING INTERFACE
P_{SCAV}
V_{OUT}
DC-DC CONVERTER
P_{LOAD}
8
2
4
5
6

Fig.1

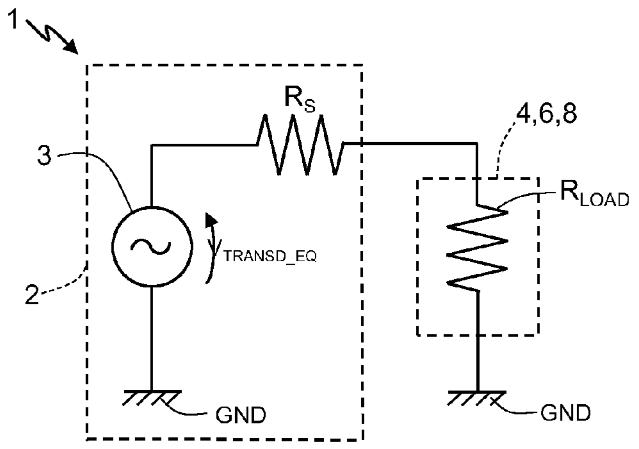
1
R_S
4,6,8
3
R_{LOAD}
V_{TRANSD_EQ}
2
GND
GND

Fig.2

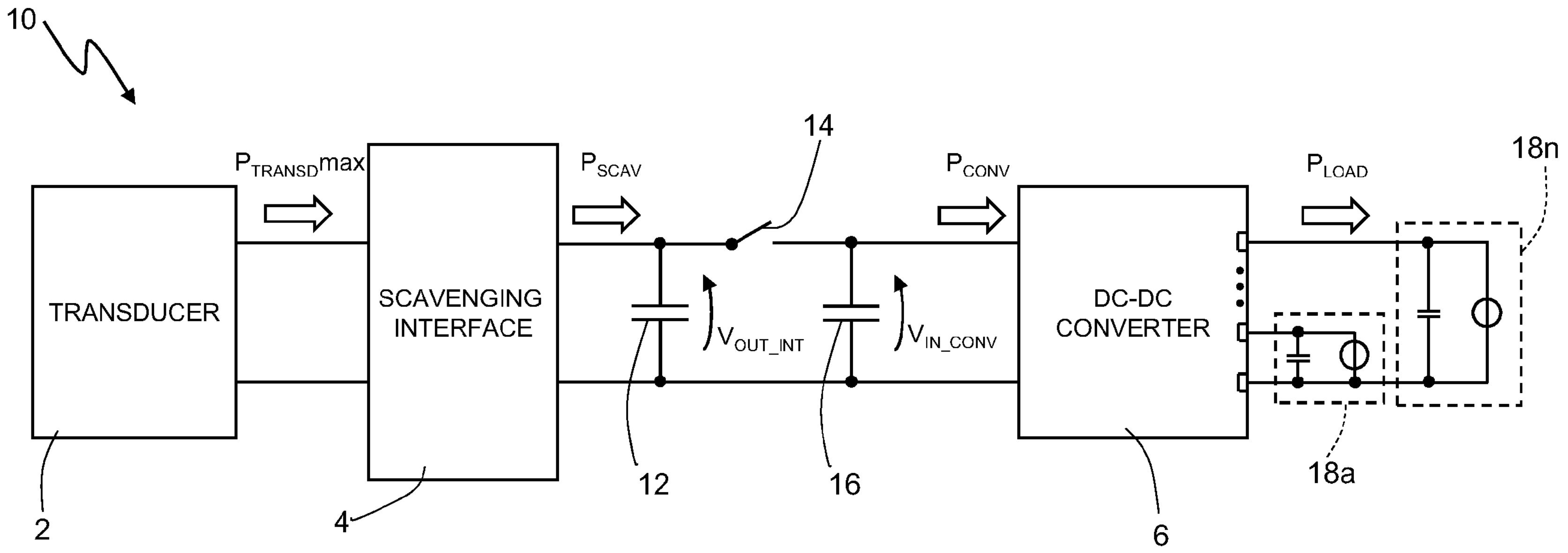
10
TRANSDUCER
2
$P_{TRANSD}max$
SCAVENGING INTERFACE
4
P_{SCAV}
12
V_{OUT_INT}
14
16
V_{IN_CONV}
P_{CONV}
DC-DC CONVERTER
6
P_{LOAD}
18a
18n

Fig.3

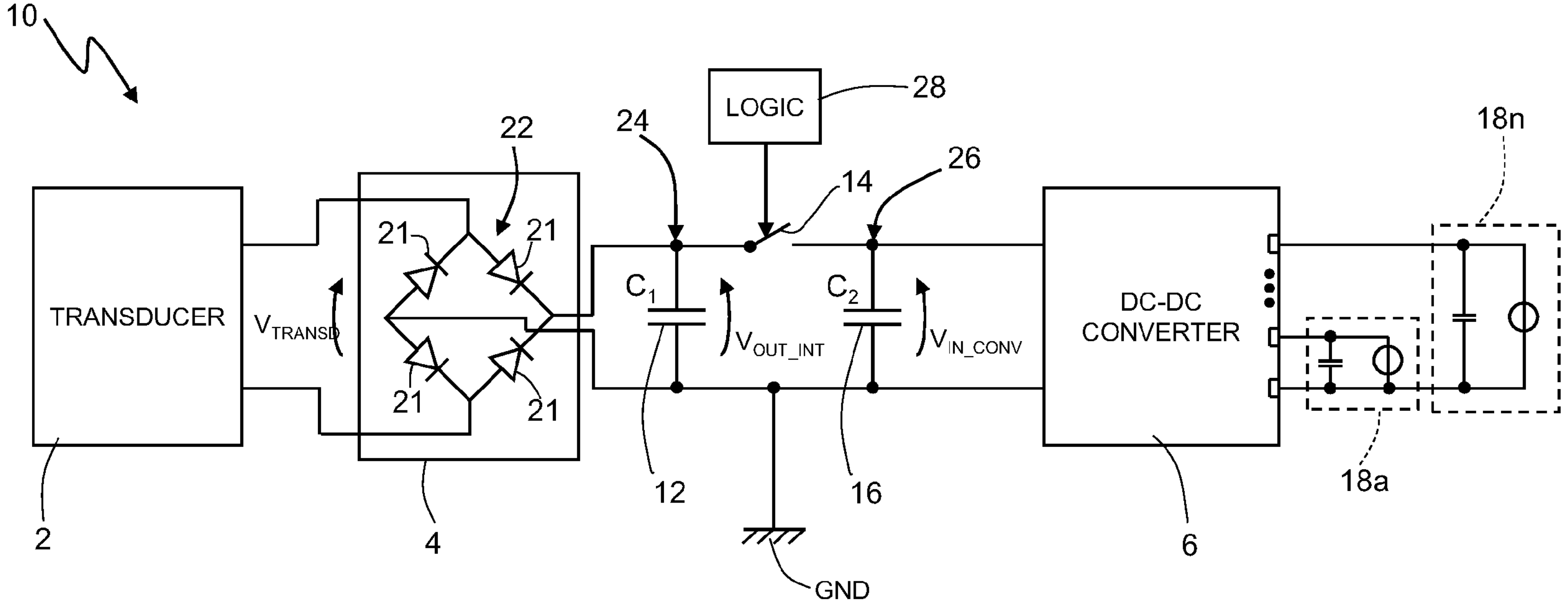
10
TRANSDUCER
2
V_{TRANSD}
22
21
21
21
21
4
24
LOGIC
28
14
26
C_1
V_{OUT_INT}
C_2
V_{IN_CONV}
12
16
GND
DC-DC
CONVERTER
6
18a
18n

Fig.4

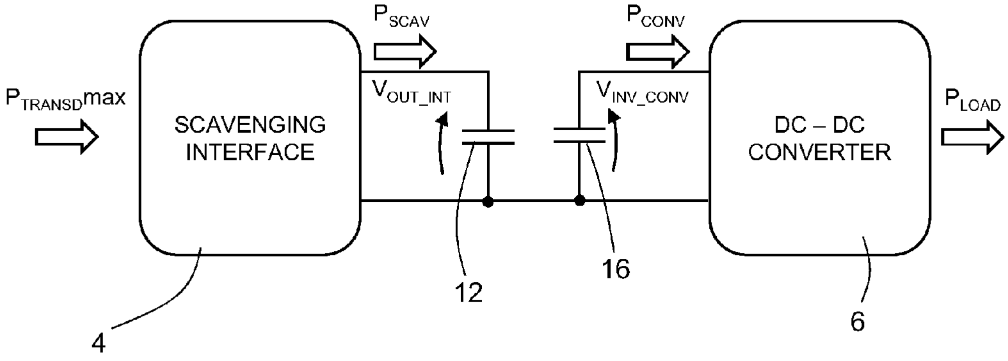
$P_{TRANSD}max$
SCAVENGING INTERFACE
P_{SCAV}
V_{OUT_INT}
P_{CONV}
V_{INV_CONV}
DC – DC CONVERTER
P_{LOAD}
4
12
16
6

Fig.5

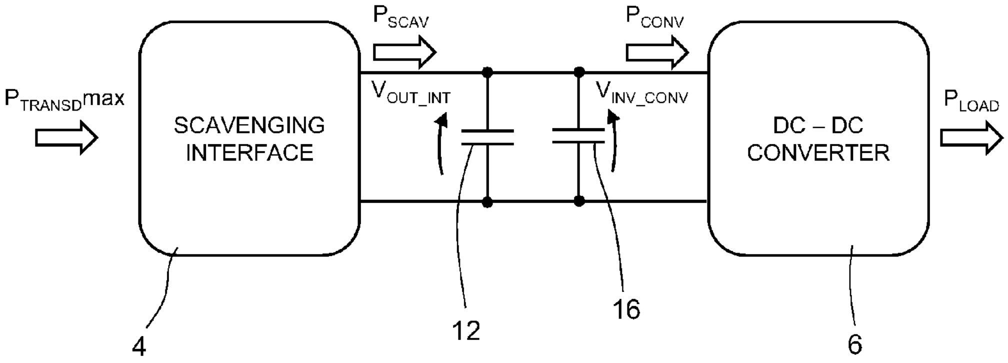
$P_{TRANSD}max$
SCAVENGING INTERFACE
P_{SCAV}
V_{OUT_INT}
P_{CONV}
V_{INV_CONV}
DC – DC CONVERTER
P_{LOAD}
4
12
16
6

Fig.6

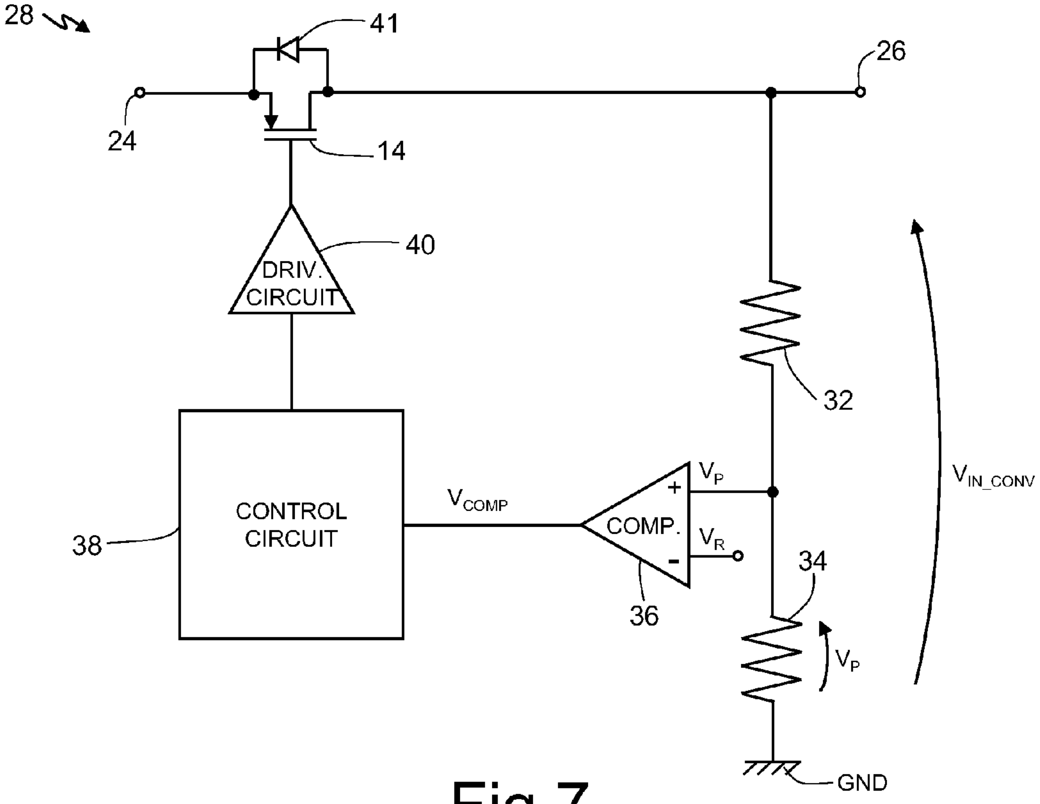
28
41
26
24
14
40
DRIV. CIRCUIT
32
V_{IN_CONV}
V_{COMP}
V_P
+
COMP.
V_R
-
CONTROL CIRCUIT
38
34
36
V_P
GND

Fig.7

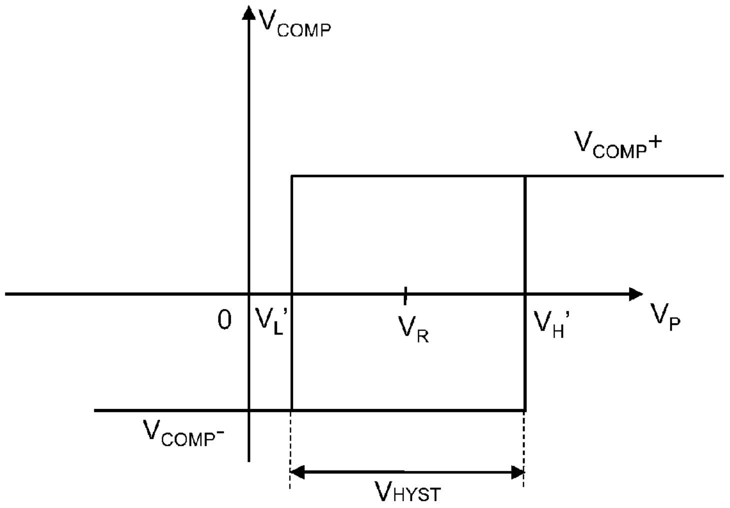
V_{COMP}
$V_{COMP}+$
0
V_L'
V_R
V_H'
V_P
$V_{COMP}-$
V_{HYST}

Fig.8

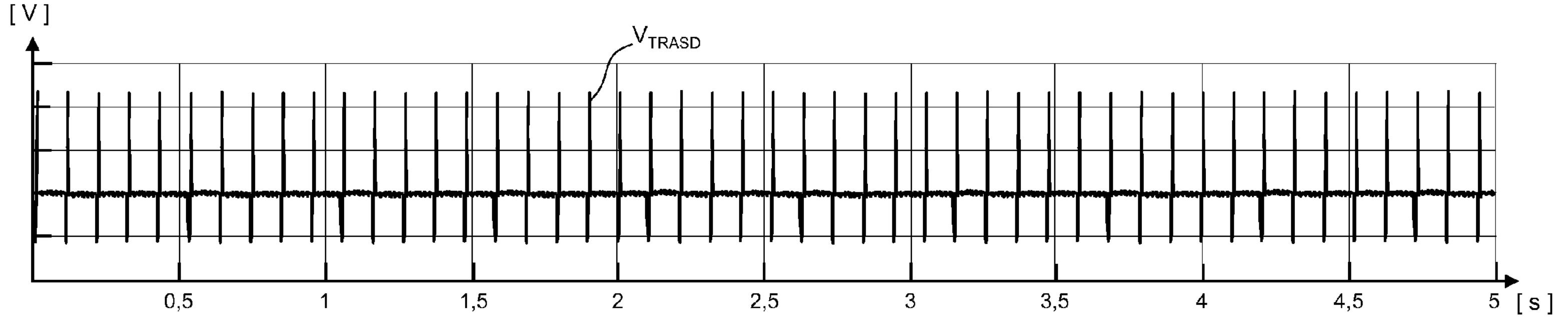
[V]
V_{TRASD}
0,5
1
1,5
2
2,5
3
3,5
4
4,5
5
[s]

Fig.9

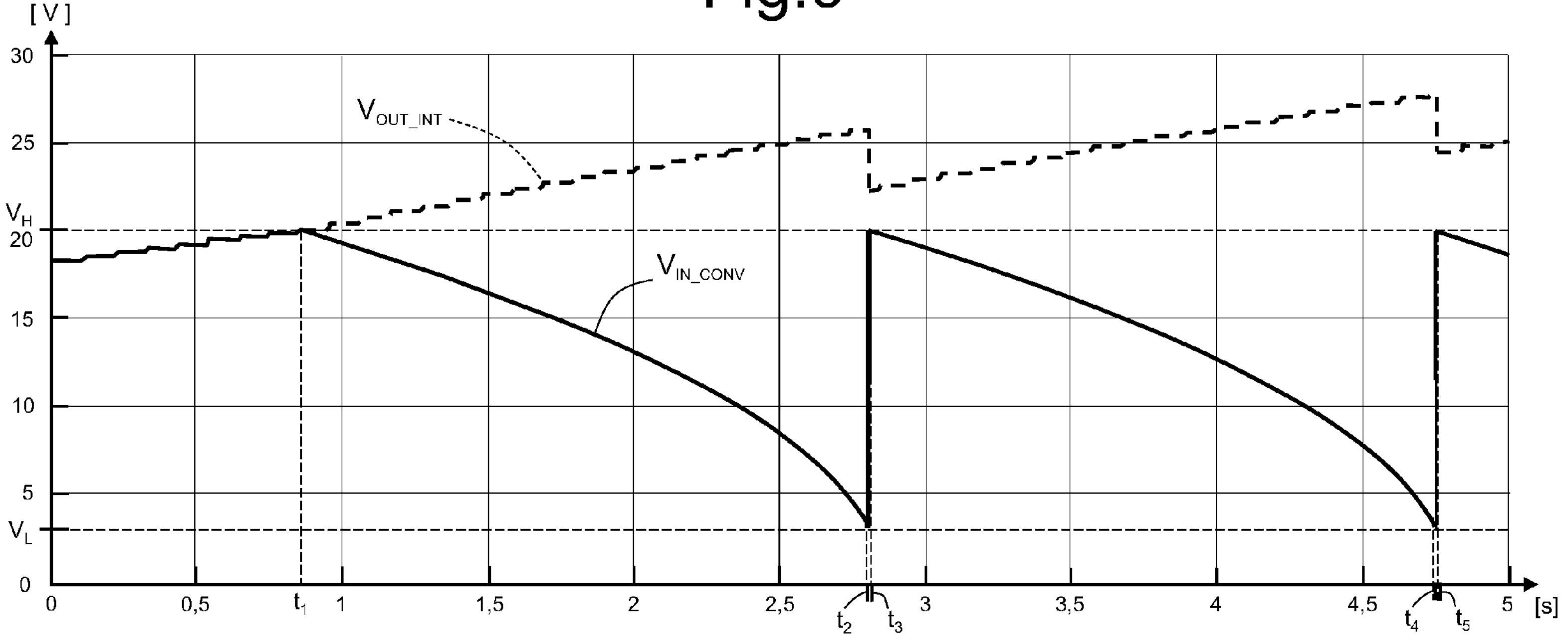
[V]
30
25
V_H
20
15
10
5
V_L
0
V_{OUT_INT}
V_{IN_CONV}
0
0,5
t_1
1
1,5
2
2,5
t_2
t_3
3
3,5
4
4,5
t_4
t_5
5
[s]

Fig.10

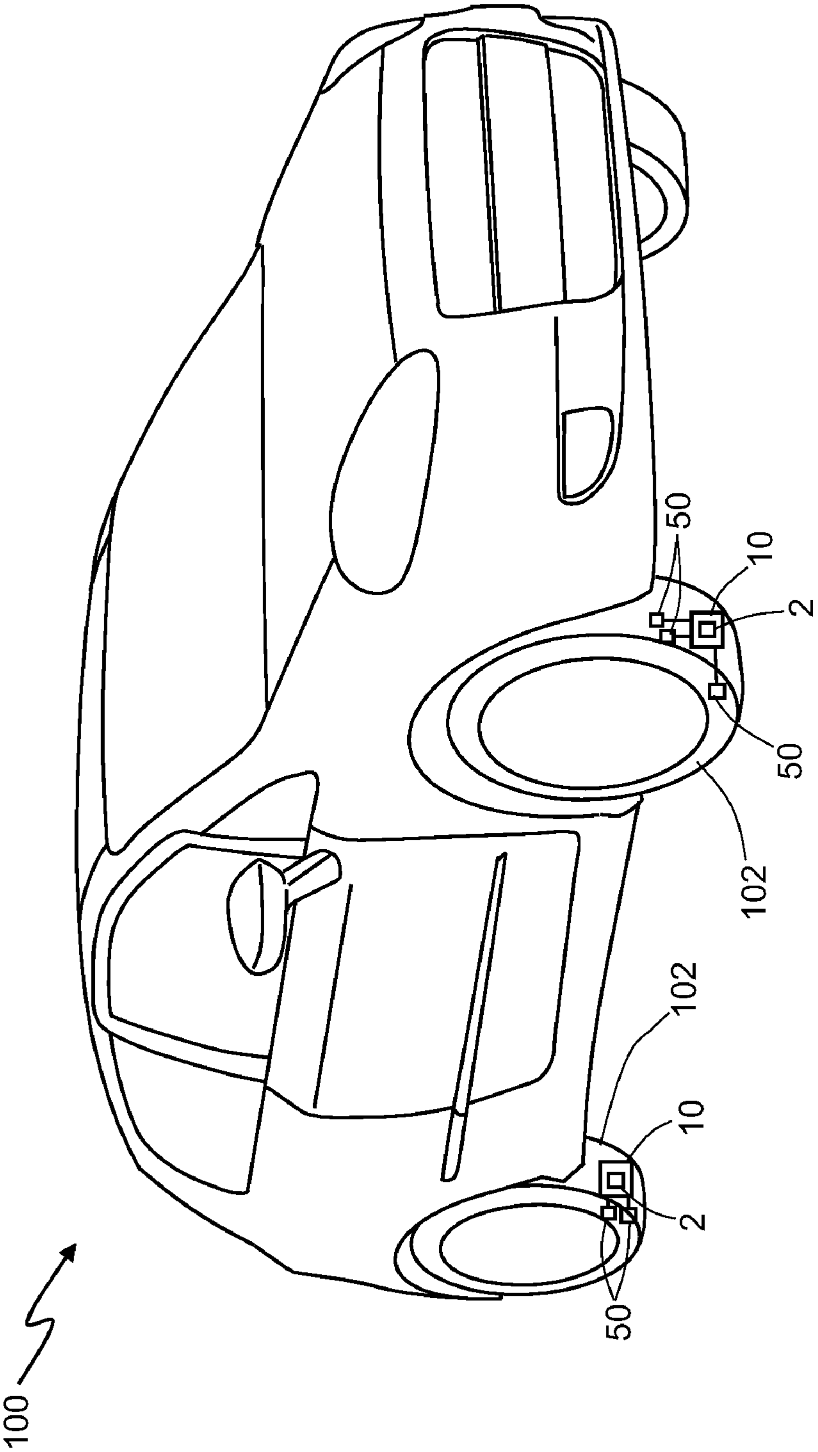
100
50
10
2
50
102
102
10
2
50

Fig.11

SYSTEM AND METHOD FOR EFFICIENTLY HARVESTING ENVIRONMENTAL ENERGY

PRIORITY CLAIM

[0001] This application claims priority from Italian Application for Patent No. TO2011A000379 filed Apr. 29, 2011, the disclosure of which is hereby incorporated by reference.

TECHNICAL FIELD

[0002] The present invention relates to a system and to a method for efficiently harvesting environmental energy.

BACKGROUND

[0003] As is known, systems for harvesting energy (also known as "energy harvesting" or "energy scavenging" systems) from intermittent environmental energy sources, have aroused and continue to arouse considerable interest in a wide range of technological fields. Typically, energy harvesting systems are designed to harvest, store, and transfer energy generated by mechanical sources to a generic load of an electrical type.

[0004] One of the main energy sources that can be used for harvesting mechanical energy and converting it into electrical energy is constituted by piezoelectric or electromagnetic devices. Low-frequency vibrations, such as for example mechanical vibrations of disturbance in systems with moving parts, can be a valid energy source.

[0005] The main needs that are felt in the field of systems for harvesting energy from environmental sources include minimum consumption of energy of the components of the systems themselves, maximum efficiency for harvesting, conversion, and storage of energy, and the need to supply the energy stored to a plurality of devices different from one another that use it for their operation.

[0006] FIG. **1** is a schematic illustration, by means of functional blocks, of an energy harvesting system of a known type.

[0007] The energy harvesting system **1** of FIG. **1** comprises: a transducer **2**, for example of an electromagnetic or piezoelectric type, subjected in use to environmental mechanical vibrations and configured for converting mechanical energy into electrical energy; a scavenging interface **4**, for example comprising a rectifier circuit, configured for receiving at input an AC signal generated by the transducer **2** and supplying at output a DC signal for charging a capacitor **5** connected to the output of the scavenging interface **4**; and a DC-DC converter **6**, connected to the capacitor **5** for receiving at input the electrical energy stored by the capacitor **5** and supplying it to an electrical load **8**.

[0008] The global efficiency η_{TOT} of the energy harvesting system **1** is given by Eq. (1) below:

$$\eta_{TOT}=\eta_{TRANSD}\cdot\eta_{SCAV}\cdot\eta_{DCDC} \tag{1}$$

[0009] where: η_{TRANSD} is the efficiency of the transducer **2**, indicating the amount of energy available in the environment that has been effectively converted, by the transducer **2**, into electrical energy; η_{SCAV} is the efficiency of the scavenging interface **4**, indicating the energy consumed by the scavenging interface **4** and the factor of impedance decoupling between the transducer and the interface; and η_{DCDC} is the efficiency of the DC-DC converter **6**.

[0010] As is known, in order to supply to the load the maximum power available, the impedance of the load should be equal to that of the source. As illustrated in FIG. **2**, the transducer **2** can be represented schematically, in this context, as a voltage generator **3** provided with an internal resistance R_S of its own. The maximum power P_{TRANSD}^{MAX} that the transducer **2** can supply at output may be defined as:

$$P_{TRANSD}^{MAX}=V_{TRANSD\_EQ}^{2}/4R_S;\ \text{if } R_{LOAD}=R_S \tag{2}$$

[0011] Where: V_{TRANSD_EQ} is the voltage produced by the equivalent voltage generator; and R_{LOAD} is the equivalent electrical resistance on the output of the transducer **2** (or, likewise, seen at input to the scavenging interface **4**), which takes into due consideration the equivalent resistance of the scavenging interface **4**, of the DC-DC converter **6**, and of the load **8**.

[0012] On account of the impedance decoupling ($R_{LOAD}\neq R_S$), the power at input to the scavenging interface **4** is lower than the maximum power available P_{TRANSD}^{MAX}.

[0013] The power P_{SCAV} stored by the capacitor **5** is a fraction of the power recovered by the interface, and is given by Eq. (3):

$$P_{SCAV}=\eta_{TRANSD}\cdot\eta\text{SCAV}\cdot P_{TRANSD}^{MAX} \tag{3}$$

[0014] whilst the power P_{EL_LOAD} supplied at output by the DC-DC converter to the electrical load **8** is given by the following Eq. (4):

$$P_{EL\_LOAD}=P_{DCDC}\cdot\eta_{DCDC} \tag{4}$$

[0015] where P_{DCDC} is the power received at input by the DC-DC converter **8**, in this case coinciding with P_{SCAV}.

[0016] The main disadvantage of the configuration according to FIG. **1** regards the fact that the maximum voltage supplied at output by the scavenging interface **4** is limited by the input dynamics of the DC-DC converter **8**.

[0017] The voltage V_{OUT} across the capacitor **5** (supplied at output by the scavenging interface **4** and at input to the DC-DC converter **8**) is in fact determined on the basis of the power balancing according to the following Eq. (5):

$$P_{STORE}=P_{SCAV}-P_{DCDC} \tag{5}$$

[0018] where P_{STORE} is the excess power with respect to the power required by the load, recovered by the interface and stored in the capacitor.

[0019] In applications where the transducer **2** converts mechanical energy into electrical energy in a discontinuous way (i.e., the power P_{TRANSD}^{MAX} varies significantly in time) and/or the power P_{EL_LOAD} required by the electrical load **8** varies significantly in time, also the voltage V_{OUT} consequently presents a plot that is variable in time.

[0020] This causes, for example, a variation of the efficiency factor η_{DCDC}, which assumes low values at high values of V_{OUT}. The maximum value of V_{OUT} is moreover limited by the range of input voltages allowed by the DC-DC converter. Maximization of the window of values allowed at input by the DC-DC converter **6** requires a specific design of the DC-DC converter; however, also in the latter case, an upper limit of the range of allowable values for V_{OUT} is imposed.

[0021] There is a need in the art to provide system and a method for efficiently harvesting environmental energy that will enable the aforesaid problems and disadvantages to be overcome.

SUMMARY

[0022] In accordance with an embodiment, a system and method enable maximization of the storage of electrical charge without requiring constraints of design of other components of the system itself, maximizing the global efficiency.

[0023] According to the present invention a system and a method for efficiently harvesting environmental energy are provided as defined in the annexed claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0024] For a better understanding of the present invention, preferred embodiments thereof are now described, purely by way of non-limiting example and with reference to the attached plates of drawings, wherein:

[0025] FIG. **1** shows an energy harvesting system according to a known embodiment;

[0026] FIG. **2** shows an equivalent circuit representation of the system of FIG. **1**;

[0027] FIG. **3** shows an energy harvesting system according to one embodiment of the present invention;

[0028] FIG. **4** shows at a greater circuit detail the system of FIG. **3**;

[0029] FIG. **5** shows the system of FIG. **3** during a first operating step;

[0030] FIG. **6** shows the system of FIG. **3** during a second operating step;

[0031] FIG. **7** is a schematic illustration of control logic for the system of FIG. **3**;

[0032] FIG. **8** shows a hysteresis curve of a hysteretic comparator of the control logic of FIG. **7**;

[0033] FIGS. **9** and **10** show the time plots of electrical signals internal to the system of FIG. **3**; and

[0034] FIG. **11** shows a vehicle comprising the system of FIG. **3**.

DETAILED DESCRIPTION OF THE DRAWINGS

[0035] FIG. **3** shows an energy harvesting system **10** according to one embodiment.

[0036] In a way similar to what is illustrated in FIG. **1** and described with reference to said figure, the energy harvesting system **10** comprises the transducer **2** and the scavenging interface **4**, electrically coupled to one another in a way in itself known.

[0037] The energy harvesting system **10** further comprises: a first storage element **12**, for example one or more capacitors; a second storage element **16**, for example one or more capacitors; a switch **14**, connected between the first storage element **12** and the second storage element **16**, which can be operated for connecting the first and second storage elements **12**, **16** to one another; and the DC-DC converter **6**, electrically coupled to the second storage element **16** and configured for supplying one or more loads **18***a* . . . **18***n* connected to respective outputs **6***a* . . . **6***n* of the DC-DC converter **6**.

[0038] In greater detail, as illustrated in FIG. **4**, the scavenging interface **4** comprises, according to one embodiment, a rectifier **20**, for example comprising diodes **21** connected to one another to form a Graetz-bridge structure **22**. The rectifier **20** has the function of receiving at input an AC signal (for example, a voltage signal) generated by the transducer **2** and supplying at output a corresponding DC signal (once again, a voltage signal). It is evident that the rectifier **20** is not necessary in the case where the transducer **2** is configured for generating at output a DC signal (in particular, a voltage signal) V_{TRANSD}.

[0039] The input terminals of the bridge structure **22** are connected to respective output terminals of the transducer **2** so as to be biased at the input voltage V_{TRANSD} generated at output by the transducer **2**. The output terminals of the bridge structure **22** are, instead, connected between a ground reference terminal GND, at reference voltage V_{REF}, and a biasing terminal **24**, at voltage V_{OUT_INT}. The first storage element **12** is connected between the biasing terminal **24** and the ground reference terminal GND, and, in use, is charged to a voltage $V_{OUT\_NT}-V_{REF}$. For example, the voltage V_{REF} is 0 V. The voltage V_{REF} can, however, assume values other than 0 V, for example be in a neighborhood of 0 V, or assume other values still, higher or lower than of 0 V, indifferently.

[0040] The switch **14** is connected between the biasing terminal **24** and an intermediate terminal **26**. The second storage element **16** is connected between the ground reference terminal GND and the intermediate terminal **26**. In this way, in use, the switch **14** can be operated for coupling the second storage element **16** with the biasing terminal **24**, providing ideally a connection in parallel between the first and second storage elements **12**, **16** and, alternately, uncoupling the second storage element **16** from the biasing terminal **24** insulating electrically the first storage element **12** from the second storage element **16**. The switch **14** is controlled in opening and closing by an appropriate control logic **28**, which will be described more fully in what follows.

[0041] In use, the voltage V_{OUT_INT} generated at output by the scavenging interface **4** is stored in the first storage element **12**. In the case where the switch **14** is open, there is no transfer of charge from the first storage element **12** to the second storage element **16** and the DC-DC converter **6**. The switch **14** can be driven in closing for the time necessary to charge the second storage element **16**, by transfer of charge from the first storage element **12**. Once a desired state of charge of the second storage element **16** has been reached, the switch **14** is opened, insulating the first storage element **12** from the second storage element **16**.

[0042] The first storage element **12** is configured for storing a high electrical charge. In particular, the storage element **12** is an element of a capacitive type having a capacitance C_1 of between 10 μF and 100 μF, for example equal to 50 μF. The second storage element **16** is also of a capacitive type, having a capacitance C_2 of between 1 μF and 20 μF, for example equal to 10 μF. The second storage element **16** is moreover configured in such a way that the voltage that is set up between its conduction terminals does not exceed the range of voltages allowed at input by the DC-DC converter **6** used. In this way, it is possible to use DC-DC converters **6** of a known type, designed without particular constraints of input dynamics deriving from the specific technical application.

[0043] If V_L is the lower limit of the range of voltages allowed at input by the DC-DC converter, and V_H is the upper limit of the range of voltages allowed at input by the DC-DC converter, the condition that brings about opening of the switch **14** is given by $V_{IN\_CONV}>V_H$, whereas the condition that brings about closing of the switch **14** is given by $V_{IN\_CONV}<V_L$. Hence, in use, we find that $V_L<V_{IN\_CONV}<V_H$. Said condition is ensured, in use, by the logic **28**, which implements, for example, a hysteretic algorithm such as to satisfy the aforesaid condition for the voltage V_{IN_CONV}.

[0044] FIG. **5** shows a portion of the energy harvesting system **10** during an operating step in which the switch **14** is open, in particular during a step of charging of the first storage element **12** via the energy supplied by the transducer **2** and of simultaneous supply of the DC-DC converter **6** by means of the second storage element **16**. The DC-DC converter **6** is supplied at input by the charge stored in the second storage element **16**, which consequently, during the step according to FIG. **5**, decreases. The first storage element **12** is electrically uncoupled from the load **18**a . . . **18**n and receives at input the voltage V_{TRANSD} appropriately rectified by the scavenging interface **4** (and reduced by a factor depending upon the efficiency η_{TRANSD} of the scavenging interface **4**). The first storage element **12** is hence charged.

[0045] With reference to what has been described in regard to FIG. **5**, it is evident that the supply of the DC-DC converter **6** is possible only if the second storage element **16** has been previously pre-charged. In an initial step of start-up of the energy harvesting system **10**, the operating step of FIG. **5** comprises the step of charging both the first and the second storage elements **12**, **16**, in a way similar to what is described in what follows with reference to FIG. **6**.

[0046] FIG. **6** shows a portion of the energy harvesting system **10** during an operating step in which the switch **14** is closed. Said operating condition is achieved when the voltage V_{IN_CONV} across the second storage element **16** drops below the limit value V_L and in the start-up step when both the first and the second storage elements **12**, **16** are discharged. In this case, the logic **28** controls the switch **14** in closing. The first and second storage elements **12**, **16** are hence electrically coupled to one another via the connection provided by the switch **14** closed (ideally, considering a switch **14** with zero internal resistance, they are connected in parallel to one another), and there is a transfer of charge between the first and second storage elements **12**, **16**. In particular, the second storage element is charged by means of the charge stored in the first storage element **12**. Consequently, in the absence of supply by the transducer **2** (via the scavenging interface **4**), the voltage V_{OUT_INT} drops and the voltage V_{IN_CONV} increases up to the value V_H (provided that, of course, the voltage V_{OUT_INT} is sufficiently high as to enable the voltage V_{IN_CONV} to reach the value V_H). When the value $V_{IN\_CONV}=V_H$ is reached, the logic **28** controls the switch **14** in opening, and flow returns to the condition of FIG. **5**.

[0047] FIG. **7** is a schematic illustration of the control logic **28**, configured for controlling the switch **14** in opening and closing, according to one embodiment of the present invention.

[0048] The control logic **28** comprises: a first resistor **32** and a second resistor **34**, connected between the intermediate terminal **26** and the ground reference terminal GND; a comparator **36**, including a first input terminal configured for receiving a reference signal V_R at input, a second input terminal connected between the first and second resistors **32**, **34** for receiving a partition signal V_p proportional to the signal V_{IN_CONV}, and an output terminal for supplying at output a comparison signal V_{COMP} indicating a result of the comparison between the reference signal V_R and the partition signal V_P; a control circuit **38**; and a driving circuit **40**, connected to the control circuit **38**.

[0049] The comparator **36** is a hysteretic comparator, of a known type. The comparator hence receives at input the signal to be monitored (partition signal V_P) and the reference signal V_R, around which it generates, in a known way, the hysteresis. The triggering thresholds are hence $V_H'=V_R+V_{HYST}/2$ and $V_L'=V_R-V_{HYST}/2$ (see also FIG. **8**). Alternatively, the comparator **36** can be without hysteresis. In this case, the comparison signal V_{COMP} varies when the partition signal V_P rises/drops with respect to the threshold value given by the reference signal V_R. A hysteretic comparator is advantageous as compared to a non-hysteretic comparator in so far as it is less subject to undesirable variations of the output signal (comparison signal V_{COMP}) due to electrical noise.

[0050] The control circuit **38** is configured for receiving at input the comparison signal V_{COMP} and, on the basis of the value assumed by the comparison signal V_{COMP}, controlling, via the driving circuit **40**, the switch **14** in opening and closing. The control circuit **38** is, for example, a microprocessor circuit configured for carrying out the aforementioned step of checking on the comparison signal V_{COMP}. Furthermore, the control circuit **38** and the driving circuit **40** can be built in integrated form as a single control and driving circuit.

[0051] The switch **14** is, for example, a MOSFET of a P type, having a first conduction terminal (source terminal), connected to the biasing terminal **24**, a second conduction terminal (drain terminal), connected to the intermediate terminal **26**, and a control terminal (gate terminal). In this case, the switch **14** is of a type designed to sustain high voltages between its source and drain terminals (e.g., drift MOS, DMOS, etc.). The diode **41** must be connected as illustrated in FIG. **7** in order to prevent conduction when the switch **14** is open.

[0052] In detail, the diode **41** is connected between the source terminal S and the drain terminal D of the switch **14**, in antiparallel configuration (with respect to the normal direction of flow of the current through the switch **14**). As is known, a characteristic of a MOSFET is that of displaying, under certain operating conditions, the electrical properties of a diode (parasitic diode). Said diode is electrically set (integrated) between the source and drain terminals of the MOSFET. In other words, the switch **14** can present the electrical behavior of a diode, in which the cathode of the diode corresponds to the source terminal and the anode to the drain terminal. The diode **41** is hence the diode integrated in the MOSFET that forms the switch **14**.

[0053] If technologies different from MOSFET technology are used for the switch **14**, the diode **41** may not be present.

[0054] The driving circuit **40** is connected to the gate terminal of the switch **14** and is configured for biasing appropriately the gate terminal of the switch **14** in order to drive the transistor in conduction or inhibition. Said circuit is in effect a level shifter capable of converting the signal (typically at low voltage) produced by the control logic **38** into a CMOS signal referenced to the terminal **24** (source of the switch **14**). By so doing it is possible to generate a correct driving signal of the switch **14** without jeopardizing operation thereof and without damaging the gate oxide thereof.

[0055] The value of resistance of the resistors **32** and **34** is chosen as high as possible according to the area on silicon available so as to guarantee the lowest current absorption at the node **26**, and so as to maximize the overall efficiency of the system **10**. The ratio between the resistors **32** and **34** is chosen so as to be able to compare the voltage at the node **26** with the low-voltage reference V_R normally generated by a bandgap circuit.

[0056] The comparator **36** is a hysteretic comparator, of a type in itself known. As illustrated schematically in FIG. **8**, the comparator **36** is provided with two different triggering thresholds: a lower threshold (positioned at a value V_L' proportional to V_L) for falling-edge values of the signal V_P applied at input to the comparator **36**, and an upper threshold (positioned at a value V_H' proportional to V_H) for rising-edge values of the signal V_P applied at input to the comparator **36**. For example, $V_L'=3$ V and $V_H'=20$ V.

[0057] The output V_{COMP} of the comparator **36** assumes two values, indicated in FIG. **8** as V_{COMP}^{+} and V_{COMP}^{-}. FIG. **8** exemplifies a comparator **36** with hysteresis around the value of the reference V_R.

[0058] In use, during charging of the second storage element **16**, the output of the comparator **36** is at the value V_{COMP}^{-}; the switch **14** is driven into a closed state by the control circuit **38** and by the driving circuit **40**; and the voltage V_{IN_CONV} across the second storage element **16** increases until it reaches the value V_H. When the voltage V_{IN_CONV} reaches a value equal to V_H (i.e., the voltage V_P reaches a value equal to V_H'), the output V_{COMP} of the comparator **36** switches from V_{COMP}^{-} to V_{COMP}^{+}. The control logic **38**, via the driving circuit **40**, opens the switch **14**, and the DC-DC converter **6** is supplied at input exclusively by the second storage element **16**. Then, the voltage V_{IN_CONV} (i.e., the signal V_P) decreases until it reaches the lower limit value V_L (i.e., V_L'); at this point, the output V_{COMP} of the comparator **36** switches to the value V_{COMP}^{-} and the control logic **38**, via the driving circuit **40**, closes the switch **14**, and the second storage element **16** is again charged by means of the charge stored in the first storage element **12**.

[0059] What has been described above is also illustrated in FIGS. **9** and **10**, which show the time plots of the signals V_{TRANSD} (FIG. **9**), and V_{OUT_INT} and V_{IN_CONV} (FIG. **10**). It should here be noted that the signal V_{TRANSD}, supplied at output by the transducer **2**, has an ideal time plot of an impulsive type, with pulses having all the same amplitude and the same frequency (they are uniform in time with respect to one another). The signal V_{OUT_INT} is defined on the basis of the ideal signal V_{TRANSD}. Each pulse of the signal V_{TRANSD} of the transducer **2** generates an increase in the signal V_{OUT_INT}, which assumes a steplike local profile. A real signal source V_{TRANSD} is typically of a discontinuous type, with pulses having amplitudes different from one another and random onset frequency. A source of this type will cause a corresponding variation of the profile of increase of the signal V_{OUT_INT}.

[0060] In an initial step of start-up of the energy harvesting system **10** (not illustrated in FIG. **10**), the voltage present across the first storage element **12** is lower than the value V_L. The switch **14** is closed and the transducer **2** supplies energy to the first and second storage elements **12**, **16**, charging them (time interval preceding time T_1). Then, when the voltage across the first and second storage elements **12**, **16** reaches the limit value of V_H (time t_1), the switch **14** is opened. The voltage V_{IN_CONV} across the second storage element **16** starts to decrease, whilst the voltage V_{OUT_INT} across the first storage element **12** continues to rise (i.e., the process for charging the first storage element **12** continues).

[0061] Of course, what has been illustrated in FIGS. **9** and **10** indicates an optimal operating condition, in which the environmental energy converted by the transducer **2** is sufficiently high and such as to be higher than the energy used for charging the second storage element **16** and used by the electrical load **18***a*, . . . , **18***n*.

[0062] When the voltage V_{IN_CONV} reaches the lower-limit value $V_{IN\_CONV}=V_L$ (time t_2), the switch **14** is closed and there is a transfer of charge from the first storage element **12** to the second storage element **16**. The voltage V_{IN_CONV} increases until the value $V_{IN\_CONV}=V_H$ is reached (time t_3).

[0063] The switch **14** is then opened again, and the step of discharge of the second storage element **16** is resumed.

[0064] As is known, the operation of charge sharing between the first storage element **12** and the second storage element **16** following upon closing of the switch **14** is the cause of a power loss due to the presence of the switch **14** itself. This causes a consequent reduction in the global efficiency of the energy harvesting system **10**. The power P_{CHARGE} useful for charging the second storage element **16** is given by the following Eq. (6):

$$P_{CHARGE}=P_{SCAV}-P_{CONV}-P_{SH} \tag{6}$$

[0065] where P_{SH} is the mean power lost during the operation of charge sharing.

[0066] The mean power P_{SH} is defined as the ratio between the energy lost ΔE and the time interval T_{SH} elapsing between a given charge-sharing event and a subsequent charge-sharing event (for example, with reference to FIG. **10**, the interval t_2-t_4).

[0067] The energy ΔE and the interval T_{SH} are defined by Eqs. (7) and (8):

$$T_{SH}=\frac{0.5\cdot C_2\cdot(V_H^2-V_L^2)}{\eta_{DCDC}\cdot P_{LOAD}} \tag{7}$$

$$\Delta E=0.5\cdot C_2\cdot(V_H-V_L)\cdot\left[2\cdot V_{OUT}+(V_H-V_L)\cdot\left(\frac{C_1-C_2}{C_1}\right)\right] \tag{8}$$

[0068] Consequently, P_{SH} is given by Eq. (9):

$$P_{SH}=\frac{\Delta E}{T_{SH}}=\left[\frac{2\cdot V_{OUT}-\frac{C_2}{C_1}\cdot(V_H-V_L)}{(V_H-V_L)}-1\right]\cdot\eta_{DCDC}\cdot P_{LOAD} \tag{9}$$

[0069] If the charge balancing is positive, the voltage V_{OUT_INT} across the first storage element **12** increases during the cycles T_{SH}; otherwise, it decreases. In order to maximize the value of energy stored in the first storage element **12** (and hence maximize the interval T_{SH}), it is advisable to maximize the threshold value V_H, which in any case must be kept within the input dynamics of the DC-DC converter **6**. It is hence advisable to use DC-DC converters **6** with high input dynamics (in particular, with an upper limit of input dynamics of a value that is as high as possible) so as to enable high threshold values V_H.

[0070] When the interval T_{SH} is maximized, the power dissipated on account of operations of charge sharing is reduced in so far as also the frequency of occurrence of the latter is reduced.

[0071] FIG. **11** shows a vehicle **100** comprising the energy harvesting system **10**, according to one embodiment of the present invention. The vehicle **100** is, in particular, a motor vehicle. It is evident, however, that the energy harvesting system **10** can be used in any vehicle **100** or in systems or apparatuses other than a vehicle. In particular, the energy harvesting system **10** can find application in generic systems in which it is desirable to harvest, store, and use environmental energy, in particular by means of conversion of mechanical energy into electrical energy.

[0072] With reference to FIG. **11**, the vehicle **100** comprises one or more transducers **2** coupled in a known way to a portion of the vehicle **100** that is subject to mechanical stresses and/or vibrations, for converting said mechanical stresses and/or vibrations into electric current.

[0073] The energy harvesting system **10** is connected to one or more electrical loads **18***a* . . . **18***n*. In particular, according to one application of the present invention, the electrical loads **18***a* . . . **18***n* comprise TPM (tire-parameter monitoring) sensors **50** for monitoring parameters of tires **102**. In this case, the TPM sensors **50** are coupled to an internal portion of the tires **102** of the vehicle **100**. Likewise, also the transducers **2** (for example, of an electromagnetic or piezoelectric type) are coupled to an internal portion of the tire **102**. The stress on of the transducers **2** when the vehicle **100** is travelling causes the production of a current/voltage electrical signal at output from the transducer **2** by means of conversion of the mechanical energy into electrical energy. The electrical energy thus produced is stored, as described previously, in the first storage element **12**, and supplied, when necessary, to the second storage element **16**, which in turn makes it available for the DC-DC converter **6** to which two or more TPM sensors **50** are connected. Thanks to the fact that the first storage element **12** has a capacitance greater than that of the second storage element **16**, which is not constrained by parameters of design of the DC-DC converter **6**, when the vehicle **100** stops it is in any case possible to supply the TPM sensors **50** using the electrical energy stored in the first storage element **12**. By appropriately sizing the first storage element **12** it is possible to increase or reduce as desired the range of autonomy of supply of the TPM sensors **50** when the vehicle is stationary.

[0074] According to one embodiment of the present invention, the energy harvesting system **10**, comprising one or more transducers, and the TPM sensors, are glued inside one or more tires **102**.

[0075] The impact of the tire **102** on the ground during motion of the vehicle **100** enables production of electrical energy.

[0076] As an alternative to what is illustrated in FIG. **11**, the energy harvesting system **10** can be set in any other portion of the vehicle **100**, and/or used for supplying an electrical load other than or additional to the TPM sensors **50**.

[0077] Another possible application of the energy harvesting system **10** is the generation of electrical energy by exploiting the mechanical energy produced by an individual when he is walking or running. In this case, the energy harvesting system **10** is located inside the shoes of said individual (for example, inside the sole). In systems aimed at fitness, where it is particularly interesting to count the steps, it is useful to recover energy from the vibrations induced by walking/running to be able to supply without the use of a battery acceleration sensors and/or RFID transmitters capable of communicating with cellphones, devices for playing music, or with any other apparatus involved in information on the steps performed.

[0078] From an examination of the characteristics of the invention obtained according to the present disclosure the advantages that it affords are evident.

[0079] In particular, the system according to the present invention has a high efficiency and autonomy irrespective of external conditions, such as possible energy peaks acquired by the transducer or a temporary absence of energy. In fact, the first storage element **12**, which has a high capacity for storing energy, has the function of reserving energy in situations of temporary absence of energy; the switch **14** is configured for uncoupling, when necessary, the first storage element **12** from the DC-DC converter **6** and from the load of the latter, rendering operativeness of the energy harvesting system **10** independent of any particular temporary conditions external to the system itself.

[0080] Furthermore, the architecture proposed for the energy harvesting system **10** enables use of standard electronic components, i.e., ones not developed purposely for this application (for example, the DC-DC converter **6** does not need to be designed to satisfy particular conditions of input dynamics), which means a considerable economic saving and saving in terms of overall efficiency.

[0081] Finally, the hysteretic control of the switch **14** enables storage of a high electrical charge in the capacitor **16** without jeopardizing the functionality of the DC-DC converter **6** and without requiring any modification thereof.

[0082] Finally, it is clear that modifications and variations may be made to what has been described and illustrated herein, without thereby departing from the sphere of protection of the present invention, as defined in the annexed claims.

[0083] For example, there may be present a plurality of transducers **2**, all of the same type or of a type different from one another, indifferently. For example, the transducer/transducers can be chosen in the group comprising: electrochemical transducers (designed to convert chemical energy into an electrical signal), electromechanical transducers (designed to convert mechanical energy into an electrical signal), electroacoustic transducers (designed to convert variations of pressure into an electrical signal), electromagnetic transducers (designed to convert a magnetic field into an electrical signal), photoelectric transducers (designed to convert light energy into an electrical signal), electrostatic transducers, or thermoelectrical transducers.

What is claimed is:

1. A system, comprising:

a transducer configured to convert energy coming from an external energy source into an electrical signal;

a first storage element coupled to an output of the transducer and configured to receive said electrical signal and store electrical energy,

a second storage element configured to store electrical energy; and

a switch coupled between the first and the second electrical-energy storage elements and operable to electrically couple the first and second storage elements to one another and, alternatively, uncouple the first and second storage elements from one another.

2. The system according to claim **1**, wherein the first storage element comprises a first conduction terminal and a second conduction terminal, wherein the second storage element comprises a first conduction terminal and a second conduction terminal, wherein the second conduction terminals of the first and second storage elements are connected to a common reference terminal, wherein the first conduction terminals of the first and second storage elements are connected to respective terminals of the switch in such a way that, when the switch is operated to electrically couple the first and second storage elements, the respective first conduction terminals of the first and second storage elements are coupled to one another.

3. The system according to claim **1**, wherein said first storage element has an electrical energy storage capacity greater than that of said second storage element.

4. The system according to claim **1**, further comprising control logic having an output coupled to control said switch, said control logic configured to close the switch when the electrical energy stored in the second storage element is lower than a first threshold value, and further configured to open the switch when the electrical energy stored in the second storage element is higher than a second threshold value.

5. The system according to claim **4**, wherein the control logic comprises:

a comparator with hysteresis configured to receive on a first input a first comparison signal indicative of electrical energy stored in the second storage element, and receive on a second input a second predefined comparison signal, and supply at the output a third comparison signal indicative of a result of a comparison between the first and second comparison signals;

a control and driving circuit configured to receive the third comparison signal and, on the basis of the third comparison signal, control opening or closing of the switch.

6. The system according to claim **1**, wherein the first storage element includes a capacitor and wherein the electrical signal generated by the transducer is an AC voltage signal, the system further comprising a scavenging interface having an input connected to the transducer and an output connected to the first storage element, the scavenging interface configured to receive at an input the electrical signal and generate at an output a first intermediate DC signal to charge the first storage element.

7. The system according to claim **1**, further comprising a DC-DC converter coupled between the second storage element and an electrical load, said DC-DC converter configured to receive a second intermediate signal depending upon the electrical energy stored in the second storage element and supply said electrical load via said second intermediate signal.

8. The system according to claim **7**, wherein the second storage element includes a capacitor having biasing terminals, the DC-DC converter further configured to accept a voltage signal having a value lower than a maximum value, and the second storage element being configured in such a way that the difference of potential that, in use, is set up across its terminals has a value lower than said maximum value.

9. The system according to claim **1**, comprising an energy-harvesting system for use in harvesting energy from movement of a support.

10. The system according to claim **9**, wherein said support is a means of transport.

11. The system according to claim **9**, wherein said transducer is chosen in the group comprising: electrochemical transducer, electromechanical transducer, electroacoustic transducer, electromagnetic transducer, photoelectric transducer, electrostatic transducer, thermoelectrical transducer.

12. An energy-harvesting method, comprising:

converting, by a transducer, energy from an external energy source into electrical energy to generate an electrical signal;

storing said electrical energy in a first storage element;

coupling the first storage element to a second storage element so as to store, in the second storage element, at least a portion of the electrical energy stored in the first storage element;

alternatively, uncoupling the first storage element from the second storage element so as to insulate the first storage element from the second storage element; and

during uncoupling, supplying an electrical load with electrical energy stored in the second storage element.

13. The method according to claim **12**, further comprising monitoring the electrical energy stored in the second storage element, wherein a coupling of the first storage element to the second storage element is carried out when the electrical energy stored in the second storage element falls lower than a first threshold value, and wherein uncoupling of the first storage element from the second storage element is carried out when the electrical energy stored in the second storage element rises higher than a second threshold value.

14. The method according to claim **12**, wherein monitoring:

supplying, on a first input of a comparator with hysteresis, a first comparison signal indicative of the electrical energy stored in the second storage element;

supplying, on a second input of the comparator with hysteresis, a second predefined comparison signal; and

generating, at an output of the comparator with hysteresis, a third comparison signal indicative of a result of a comparison between the first and second comparison signals,

wherein said coupling and uncoupling actions are performed on the basis of a value assumed by the third comparison signal.

15. The method according to claim **12**, wherein the first storage element includes a capacitor and wherein the electrical signal generated by the transducer is an AC voltage signal, the method further comprising:

rectifying the electrical signal generated by the transducer to generate a first intermediate DC signal; and

charging the first storage element with the first intermediate signal.

16. A circuit, comprising:

a first capacitor having a first terminal coupled to a reference voltage node and a second terminal coupled to an input node and configured to store first electrical energy;

a second capacitor having a third terminal coupled to said reference voltage node and a fourth terminal and configured to store second electrical energy;

a switching circuit having an input terminal coupled to the first terminal and an output terminal coupled to the fourth terminal;

a hysteresis comparator coupled to sense the second electrical energy at the fourth terminal and configured to close the switching circuit in response to sensed second electrical energy falling below a first threshold and open the switching circuit in response to sensed electrical energy rising above a second threshold.

17. The circuit of claim **16**, further comprising a rectifying circuit configured to rectify an AC signal to generate an DC signal at said input node.

18. The circuit of claim **17**, further comprising a transducer circuit configured to convert non-electrical energy into said AC signal.

19. The circuit of claim **18**, further comprising a DC-DC converter circuit having an input coupled to said fourth terminal and configured to supply power to a load from the second electrical energy.

20. The circuit of claim **16**, wherein said first capacitor has an electrical energy storage capacity greater than that of said second capacitor.

* * * * *

US 20120286623A1

(19) **United States**

(12) **Patent Application Publication**

Ogram

(10) **Pub. No.: US 2012/0286623 A1**

(43) **Pub. Date: Nov. 15, 2012**

(54) ATMOSPHERIC ENERGY COLLECTION

(75) Inventor: **Mark E. Ogram**, Tucson, AZ (US)

(73) Assignee: **Sefe, Inc.**, Tempe, AZ (US)

(21) Appl. No.: **13/103,963**

(22) Filed: **May 9, 2011**

Publication Classification

(51) **Int. Cl.**
H02N 3/00 (2006.01)

(52) **U.S. Cl.** **310/309**

(57) **ABSTRACT**

The subject matter described herein is an atmospheric energy collector. The atmospheric energy collector includes of a windsock arrangement that has a large up-wind opening on one side and that tapers from the larger up-wind opening on the one side to a small down-wind opening on the other side. The up-wind side is secured to a tether such that an electrically conducting material (e.g. metal) included in construction of the atmospheric energy collector is connected to the tether. The windsock arrangement is extended outwards by wind and the like atmospheric conditions such that the electrically conducting material collects the atmospheric energy and transfers the collected energy to the tether.

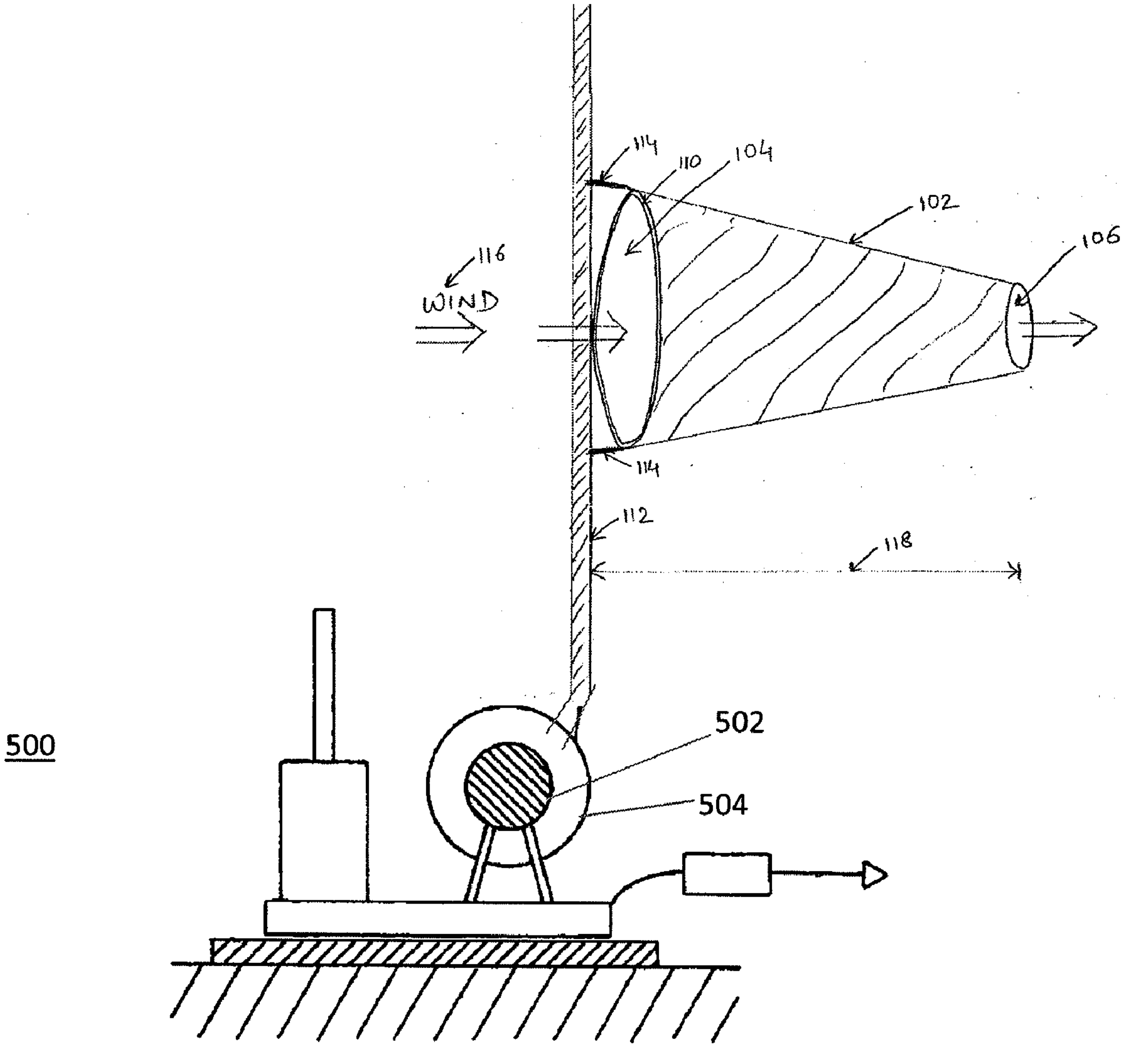

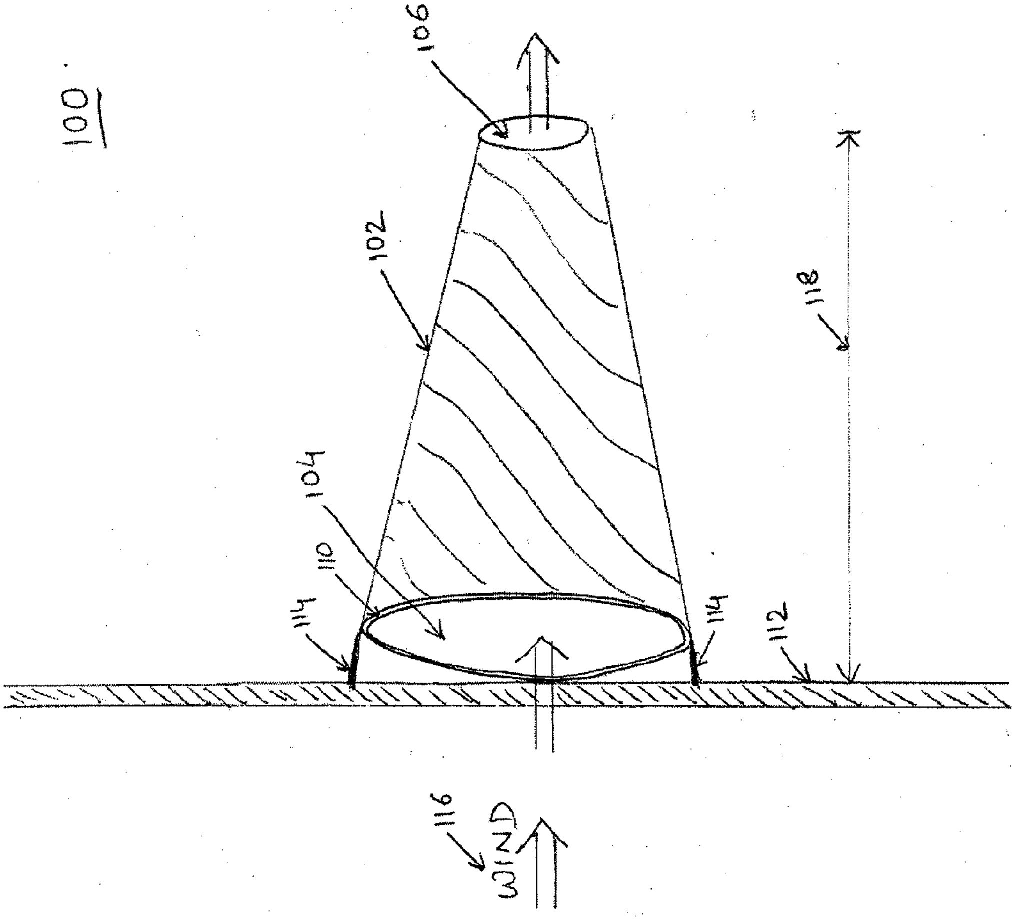
100
106
102
118
104
110
114
114
112
116
WIND

FIG. 1A

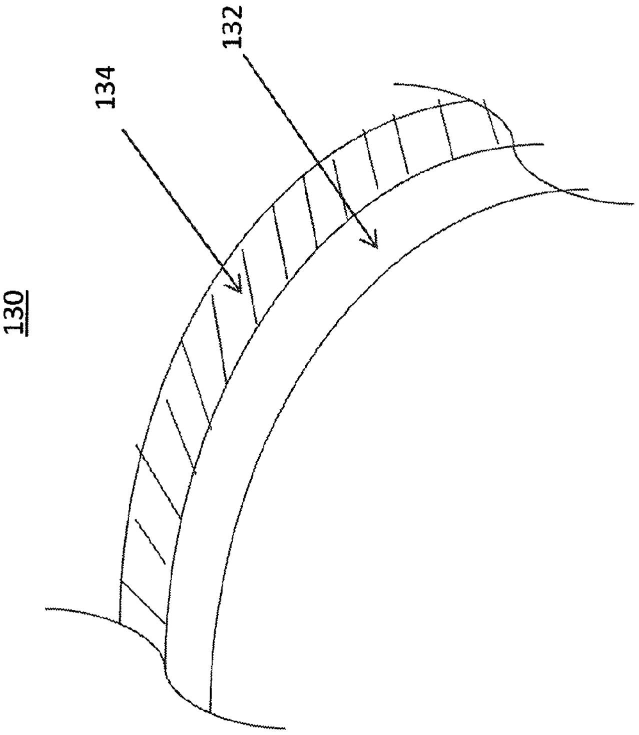

FIG. 1B

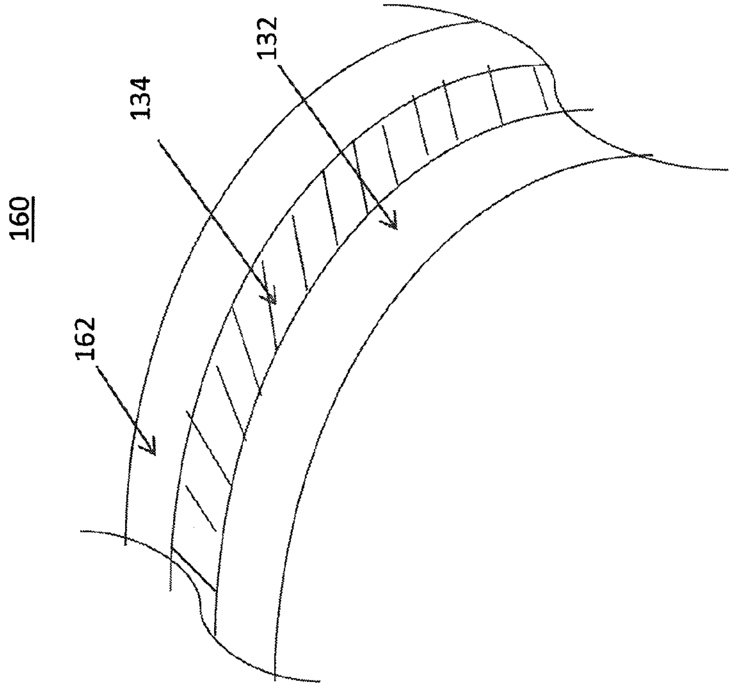
160
162
134
132

FIG. 1C

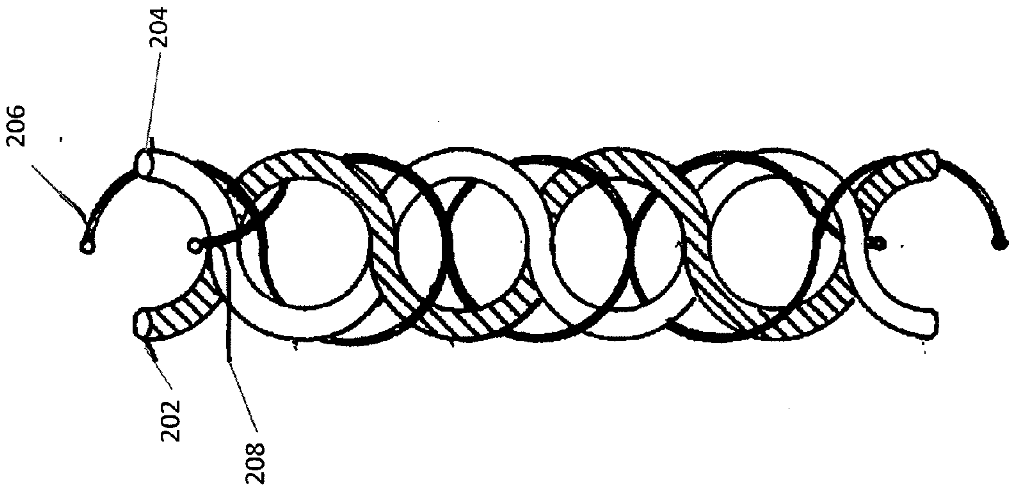
204
206
202
208

FIG. 2

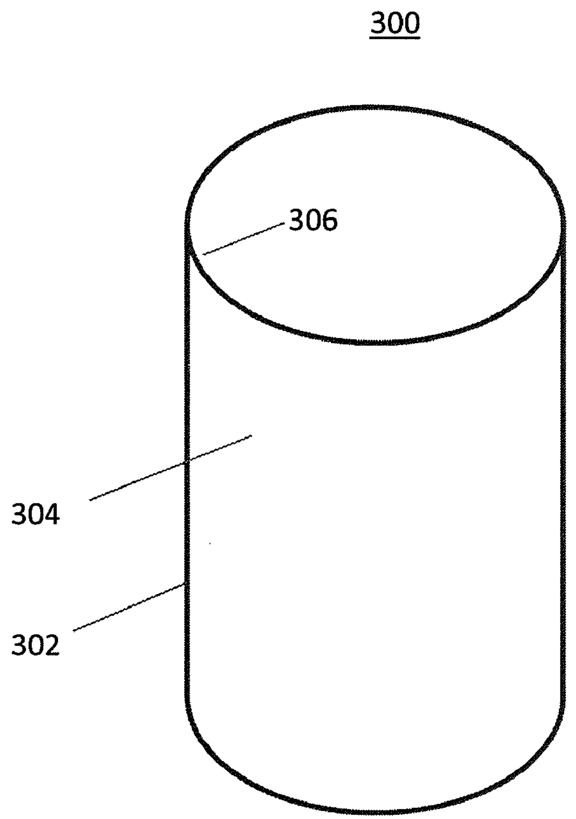
300
306
304
302

FIG. 3A

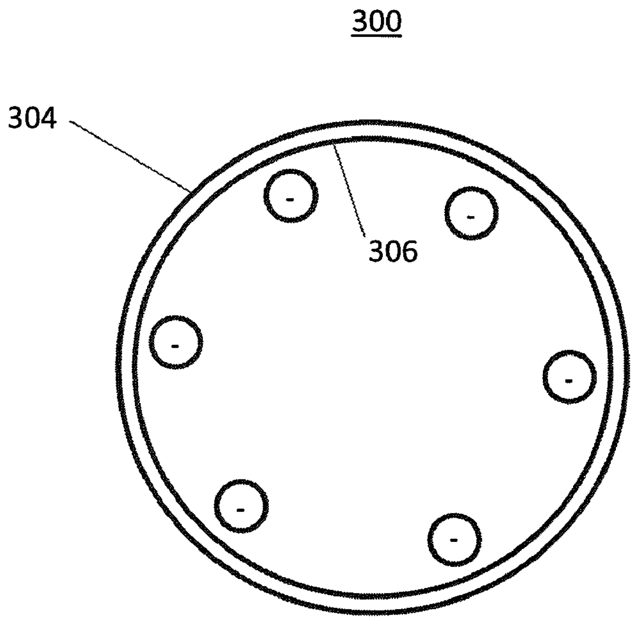

FIG. 3B

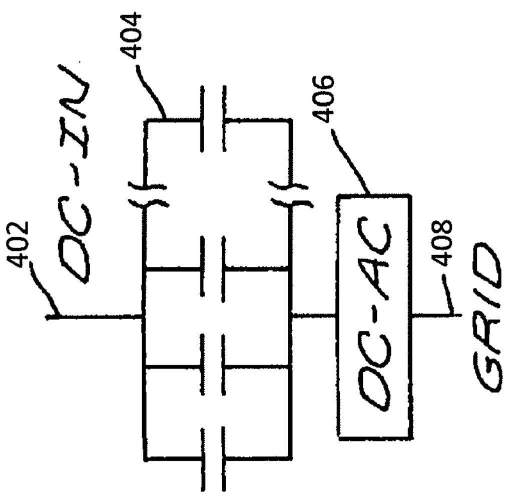
402
DC-IN
404
406
DC-AC
408
GRID

FIG. 4A

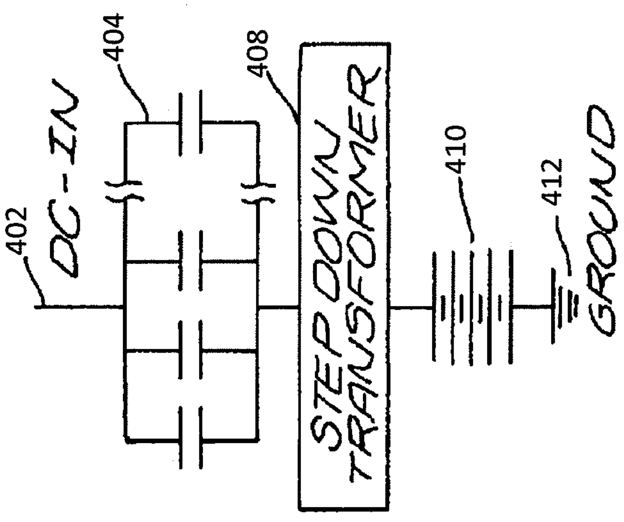
402
DC-IN
404
408
STEP DOWN TRANSFORMER
410
412
GROUND

FIG. 4B

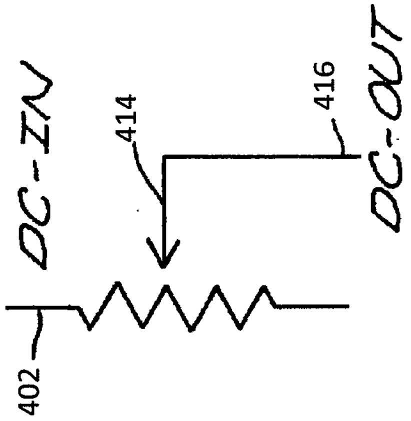
402
DC-IN
414
416
DC-OUT

FIG. 4C

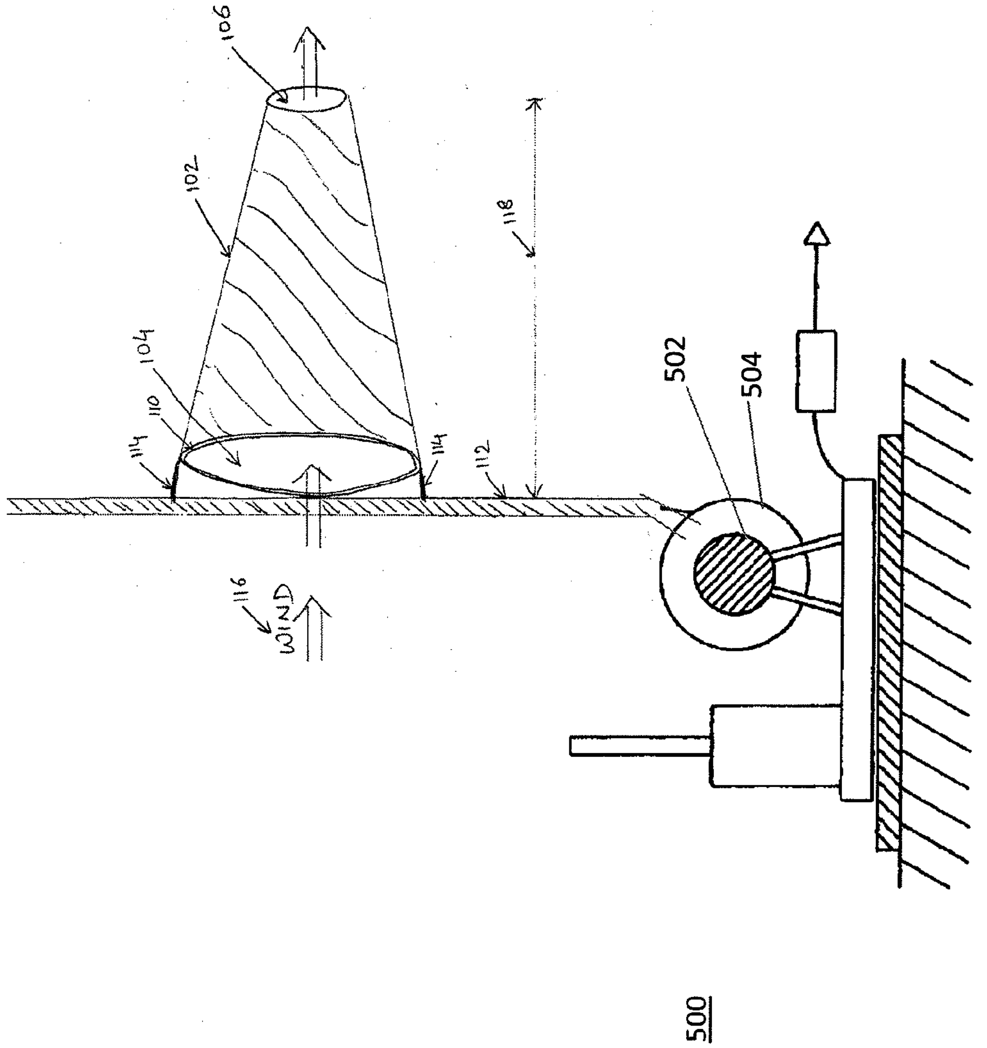
106
102
118
104
110
114
114
112
116
WIND
502
504
500

FIG. 5

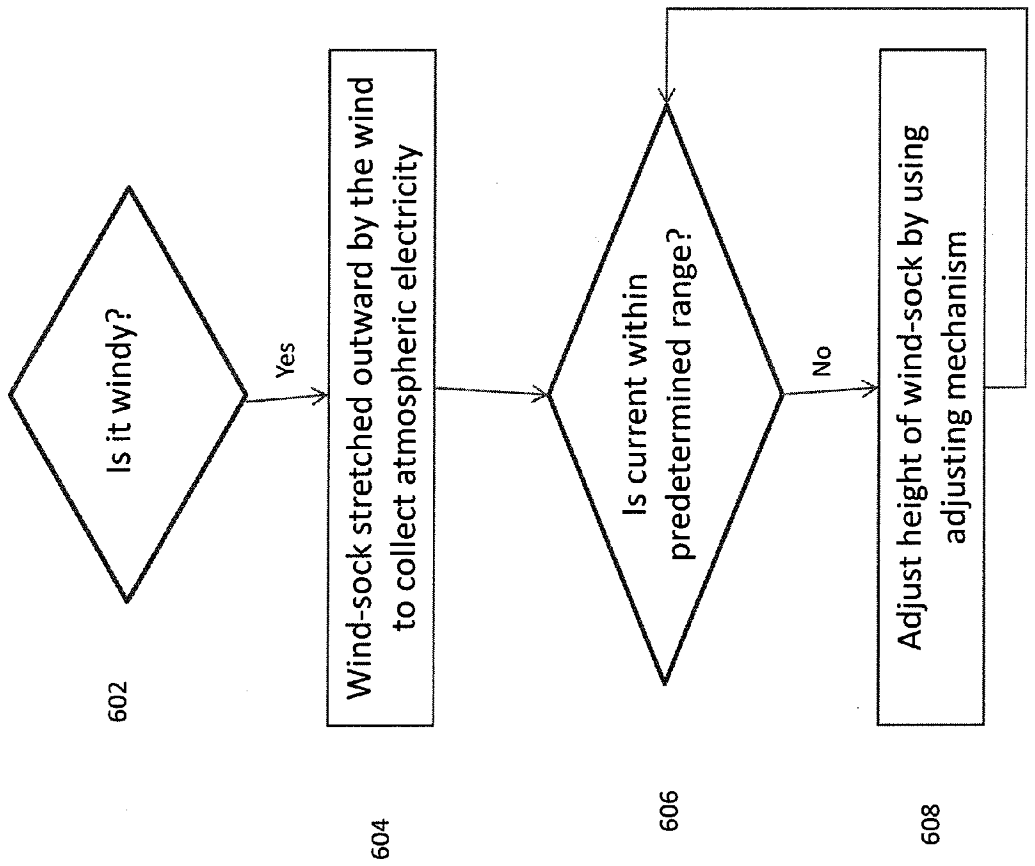
602
Is it windy?
Yes
604
Wind-sock stretched outward by the wind to collect atmospheric electricity
606
Is current within predetermined range?
No
608
Adjust height of wind-sock by using adjusting mechanism

FIG. 6

ATMOSPHERIC ENERGY COLLECTION

TECHNICAL FIELD

[0001] The subject matter described herein relates to a light-weight, extendible electricity collecting windsock arrangement comprising an enhanced collection surface for atmospheric electrical energy collection.

BACKGROUND

[0002] The atmosphere above the earth is known to include electrical charge. The earth's surface is negatively charged, while the air above it is positively charged. All atmospheric effects are a result of an interplay between these two huge areas of opposite electrical energy. The potential difference between the positively charged atmosphere and the negatively charged earth surface causes atmospheric electrical energy to be developed. Some of this atmospheric electrical energy may be experienced through thunderstorms.

[0003] Even though there is a huge amount of atmospheric electrical energy in the atmosphere, collection of the atmospheric electrical energy still remains a problem. Due to high voltage and low current conditions, there is a need for large collection surfaces for maximizing the collection of atmospheric electrical energy. Note that the terms atmospheric electricity, atmospheric energy, and atmospheric electrical energy have been used interchangeably in this specification.

[0004] The problem of low current is compounded by a heavy weight of the collectors of atmospheric electrical energy. The heavy weight of the collectors increases payload requirements of a lift mechanism, which may be used to lift the collectors to an appropriate height or altitude so that the collection of atmospheric electrical energy is maximized. Thus, a need exists for large yet light-weight collectors such that payload requirements for the lift mechanism remain reasonable.

SUMMARY

[0005] The subject matter described herein relates to an atmospheric energy collector. The atmospheric energy collector includes an electricity collecting windsock arrangement that has a large up-wind opening on one side and that tapers from the larger up-wind opening on the one side to a small down-wind opening on the other side. The up-wind side is secured to a tether such that an electrically conducting material (e.g. metal) included in construction of the atmospheric energy collector is connected to the tether. The electricity collecting windsock arrangement is extended outwards by wind and the like atmospheric conditions such that the electrically conducting material collects the atmospheric energy and transfers the collected energy to the tether. Further, the electricity collecting windsock includes light-weight collectors such that payload requirements for the lift mechanism remain reasonable (i.e. remain within a predetermined value).

[0006] In one aspect, an electricity collection apparatus includes a tether and a windsock. The windsock is formed of an insulation material with a metal deposited on at least a portion of the insulation material. The metal of the windsock is electrically connected to the tether. The windsock extends in a direction of ambient wind to provide a surface area based on the ambient wind, and the surface area provides the metal to collect electrical energy from the ambient wind, the collected electrical energy being transferred to the tether.

[0007] In another aspect, an electricity collection apparatus includes a windsock formed of an insulation material the windsock to extend in a direction of ambient wind to provide a surface area based on the ambient wind. The apparatus further includes one or more electrical conductors provided on at least a portion of the insulation material to collect electrical energy from the ambient wind. The apparatus further includes an electrically conductive tether connected with windsock to anchor the windsock in the direction of the ambient wind, the electrically conductive tether being electrically connected to the one or more electrical conductors provided on at least a portion of the insulation material to transfer the collected electrical energy from the one or more electrical conductors to an electrical storage.

[0008] The subject matter described herein provides many advantages. For example, large collection surfaces allow a maximized collection of atmospheric energy. Moreover, the light-weight collectors allow payload requirements for the lift mechanism to remain reasonable (i.e. remain within a predetermined value).

[0009] The details of one or more variations of the subject matter described herein are set forth in the accompanying drawings and the description below. Other features and advantages of the subject matter described herein will be apparent from the description and drawings, and from the claims.

DESCRIPTION OF DRAWINGS

[0010] The accompanying drawings, which are incorporated in and constitute a part of this specification, show certain aspects of the subject matter disclosed herein and, together with the description, help explain some of the principles associated with the disclosed implementations. In the drawings,

[0011] FIG. **1**A illustrates extendible electricity collecting windsock arrangement that is used for collection of atmospheric electrical energy in accordance with implementations of the current subject matter;

[0012] FIG. **1**B illustrates a cross-section of one implementation of the electrically conducting windsock in accordance with implementations of the current subject matter;

[0013] FIG. **1**C illustrates a cross-section of another implementation of the electricity collecting windsock in accordance with implementations of the current subject matter;

[0014] FIG. **2** illustrates an exemplary conductive line in accordance with implementations of the current subject matter;

[0015] FIG. **3**A illustrates a front view of a conductive line in accordance with implementations of the current subject matter;

[0016] FIG. **3**B illustrates a top view of the conductive line shown in FIG. **3** in accordance with implementations of the current subject matter;

[0017] FIGS. **4**A, **4**B, and **4**C illustrate electrical schematics for handling the static charge from the atmosphere in accordance with implementations of the current subject matter;

[0018] FIG. **5** illustrates an adjusting apparatus (winch motor and spool) that enables adjustment of location/height of the electricity collecting windsock arrangement in accordance with implementations of the current subject matter; and

[0019] FIG. **6** is a process flow diagram illustrating aspects of a method in accordance with implementations of the current subject matter.

[0020] When practical, similar reference numbers denote similar structures, features, or elements.

DETAILED DESCRIPTION

[0021] To address these and potentially other issues with currently available solutions, one or more implementations of the current subject matter provide methods, systems, articles or manufacture, and the like that can, among other possible advantages, provide an energy collector formed as a windsock and having an enhanced collection surface for atmospheric electrical energy collection. In preferred implementations, an electricity collecting windsock is formed as a flexible cylinder, a truncated cone, or a cone with a metal surface, and is mounted to a mast or other tether so as to capture ambient wind which extends the windsock outward from the mast, in turn forming the largest possible surface area for the wind conditions to enable the metal surface to collect atmospheric amperage. The collected amperage is transferred from the metal surface of the electricity collecting windsock to the mast or tether, and eventually to an electricity storage device. Although implementations in which windsock arrangements having one windsock each are discussed below, those or ordinary skill in the art understand that a windsock arrangement may include two or more windsocks for enhanced atmospheric electricity collection.

[0022] FIG. **1**A illustrates an implementation of electricity collecting windsock arrangement **100** that is used to collect atmospheric electrical energy. The electricity collecting windsock arrangement **100** includes an extendible electricity collecting windsock **102**. The extendible electricity collecting windsock **102** is made of a material that has a high tensile strength and provides electrical insulation. This electrically insulating material **134** (discussed below with respect to FIG. **1**B) may be a plastic or a flexible glass, such as polyester like polyethylene terephthalate (PET) or biaxially-oriented polyethylene terephthalate (BoPET). BoPET is known in the industry by different trade names, some of which are Mylar, Melinex and Hostaphan.

[0023] As is described later with respect to FIGS. **1**B and **1**C, which illustrate cross-sections of implementations of electricity collecting windsock **102**, a thin coat of an electrically conducting material **132**, **162** is deposited on the electrically insulating material **134**. The deposited electrically conducting material **132**, **162** is a good conductor of electricity, and helps electricity collecting windsock **102** to tap atmospheric electrical energy. The electrically conducting material may be any one of a metal or a suitable alloy, such as one of or a combination of one or more of gold, silver, copper, aluminum, and the like.

[0024] The electricity collecting windsock **102** has a large up-wind opening up-wind opening **104** on one side. The electricity collecting windsock **102** tapers from the large up-wind opening **104** on the one side to a small down-wind opening **106** on the other side of the electricity collecting windsock **102**. The up-wind opening **104** has a metal loop **110** on the circumference of the up-wind opening **104**. The metal loop **110** is attached to a tether **112** using an attachment mechanism **114**. The attachment mechanism **114** may include a soldering mechanism. In some implementations, other attachment mechanisms **114** are also known to be used, such as nut and bolt mechanism, threading mechanism, gluing mechanism, and the like. The electricity collecting windsock **102** is extended outwards by wind **116** and/or like atmospheric conditions such that the electrically conducting material (e.g. metal) collects the atmospheric electrical energy and transfers the collected energy to the tether **112**.

[0025] The tether **112** includes a conductive line that is isolated from the ground. The conductive line is used to transfer the collected atmospheric electrical energy from the electricity collecting windsock **102** to an electricity storage device (not shown) where the collected atmospheric charge may be gathered for later or simultaneous use.

[0026] When wind **116** strikes against the electricity collecting windsock arrangement **100** in the direction illustrated, the electricity collecting windsock **102** extends outward, thus allowing the electrically conducting material **132** (discussed below with respect to FIG. **1**B) to collect maximum possible atmospheric electricity. In one implementation, the outward extension **118** is fifty feet or more. When there is no wind, the electricity collecting windsock **102** may be in a relatively compressed state such that the length of the electricity collecting windsock **102** is less than the outward extension **118** caused by windy conditions or like atmospheric conditions. The collected atmospheric electrical energy is communicated to the tether **112**.

[0027] FIG. **1**B illustrates a cross-section **130** of one implementation of the electrically conducting windsock **102**. The cross-section **130** illustrates the electrically conducting material **132** deposited on the inner surface of the electrically insulating material **134**. The electricity collecting windsock **102** includes the electrically conducting material **132** and the electrically insulating material **134** such that the electrically conducting material **132** forms the inner surface of the electricity collecting windsock **102**. When the wind **116** strikes the electricity collecting windsock arrangement, the electricity collecting windsock **102** extends outward, thus allowing the electrically conducting material **132** to collect maximum possible atmospheric electricity. The electrically conducting material **132** may include multiple collectors that are spaced apart such that the collection surface is greatly enhanced. The multiple collectors may be spaced apart in patterns, such as spaced apart lines, squares, rectangles, circles, and the like, or any other pattern that may maximize the collection of the atmospheric electrical energy. As noted above, the electrically insulating material **134** may be a plastic or a flexible glass, such as polyester like polyethylene terephthalate (PET) or biaxially-oriented polyethylene terephthalate (BoPET). The electrically conducting material **132** may be a metal or alloy, such as one of or a combination of one or more of gold, silver, copper, aluminum, and the like.

[0028] FIG. **1**C illustrates a cross-section **160** of another implementation of the electricity collecting windsock **102**. The electrically conducting material (e.g. metal deposit) **132**, **162** may be placed on both sides of the electrically insulating material **134** such that the collection ability is increased even further. Multiple collectors are implemented and are spaced apart on electrically conducting material **132**, **162** such that the collection surface is greatly enhanced. The multiple collectors may form one or more patterns, as noted above with respect to FIG. **1**B. Furthermore, the electrically insulating material **134** (e.g. mylar) used in the electricity collecting windsock **102** is light, thereby minimizing the overall weight of the electricity collecting windsock **102** and reducing the weight for payload calculations. The electrically conducting materials **132**, **162** may be an electrically conducting metal or alloy, such as one of or a combination of at least one of gold, copper, aluminum, silver, and the like. The electrically conducting materials **132**, **162** may be made of the same metal/alloy or a different metal/alloy.

[0029] The electrically conducting materials **132**, **162** may be attached to the electrically insulating material **134** by a gluing mechanism. In some implementations, other attachment mechanisms are known to be implemented, such as paint coating mechanism, a threading mechanism, a nut and bolt mechanism, soldering mechanism, and the like. The attachment mechanism between the electrically conducting material **132** and the electrically insulating material **134** may or may not be the same as the attachment mechanism between the electrically conducting material **162** and the electrically insulating material **134**.

[0030] FIG. **2** illustrates an exemplary conductive line in accordance with one implementation. This type of conductive line is commonly referred to as poly-wire. The conductive line consists of multiple interwoven strands of plastic **202** and **204** woven into a cord or rope arrangement having intertwined therein exposed metal wires **206** and **208**. Although FIG. **2** illustrates two plastic strands and two metal wires, any number of possible combinations of plastic strands and metal wires is possible. The exposed metal wires **206** and **208** attract the atmospheric static charge and transmit the charge down to the electricity storage device (not shown).

[0031] FIG. **3**A illustrates a front view of a conductive line **300** in accordance with another implementation. FIG. **3**B illustrates a top view of the conductive line **300** illustrated in FIG. **3**. The conductive line **300** creates an ionized pathway for the flow of the static charges from the atmosphere to the electricity storage device via the electricity collecting windsock arrangement **100**. This conductive line utilizes a tube **302** having an outer layer **304** of PET Film (Biaxially-oriented polyethylene terephtalate polyester film). The tube **302** provides exceptionally high tensile strength and is chemically and dimensionally stable. In one implementation, the tube **302** may have an ideal diameter of between two and three inches. An interior metal coating **306** provides an initial conduit for the flow of static charge. The static charge through the metal may force the tube **302** to expand due to the repulsion experienced by like charges. Further, the flow of electricity causes the interior of the tube **302** to become ionized to provide an additional pathway for the atmospheric static charges to the electricity storage device (not shown).

[0032] FIGS. **4**A, **4**B, and **4**C illustrate electrical schematics for handling the static charge from the atmosphere. By maintaining the voltage being collected in a prescribed range, an electrical conversion system is easily designed. While FIGS. **4**A, **4**B, and **4**C illustrate some electrical configurations, those of ordinary skill in the art readily recognize a variety of other configurations that may serve the same function.

[0033] Referencing FIG. **4**A, Direct Current In (DC IN) **402** is buffered by a gang of capacitors **404** before being communicated to a DC/AC converter **406**. The DC/AC converter converts the direct current into alternating current suitable for placement over an existing electrical grid **408** such as normally found from a power-plant. Those of ordinary skill in the art readily recognize a variety of DC/AC converters that may work in this capacity.

[0034] FIG. **4**B illustrates an electrical arrangement suitable for use in charging a battery. DC IN **402** is buffered by capacitor bank **404** before entering into a step down transformer **408**. Step down transformer **408** reduces the voltage so that the voltage can safely be introduced into battery **410**, which is connected to ground **412** at the battery's other pole. Those of ordinary skill in the art readily recognize a variety of batteries that may work in this capacity.

[0035] In FIG. **4**C, DC IN **402** is fed into an adjustable rheostat **414**, which is controlled by the controller so that the DC OUT **416** falls within a specified range.

[0036] FIG. **5** illustrates an adjusting apparatus **500** that enables adjustment of height of the electricity collecting windsock arrangement **100** in accordance with one implementation. As per one implementation, the height of the electricity collecting windsock **102** can be adjusted using a winch motor **502** and a spool **504**. The winch motor **502** and the spool **504** can release or withdraw the tether **112** to adjust the height of the electricity collecting windsock **102** that is connected to the tether **112**. In one implementation, this release and the withdrawal may be performed manually. In another implementation, this release and the withdrawal may be performed automatically based on information obtained from a sensor system that measures atmospheric electrical energy being collected per unit time. If the sensor indicates that a current flow is diminishing, then the tether **112** is released/extended from the spool **504** to increase the altitude of the electricity collecting windsock **102** such that more static charge from the atmosphere is gathered. If the sensor indicates that collected atmospheric electrical energy exceeds a preset level per unit time, the tether **112** is withdrawn onto the spool **504** to decrease the static charge being collected from the atmosphere.

[0037] FIG. **6** is a process flow diagram illustrating aspects of a method consistent with implementations of the current subject matter. At step **602**, the functioning of the electricity collecting windsock arrangement **100** depends on whether atmospheric conditions are windy. If the atmospheric conditions are windy, the wind **116** stretches/extends outward the electricity collecting windsock **102**—Step **604**. Next, it is determined whether the atmospheric electrical energy (atmospheric current) that is collected is in a predetermined range—Step **606**. If the collected atmospheric electrical energy is in the predetermined range, the flow goes back to step **604**. If the current is not in the predetermined range, the height of the electricity collecting windsock **102** may be adjusted using adjusting apparatus **500**, such as the winch motor **502** and spool **504**. One skilled in the art understands that other implementations may include any other adjusting apparatus **500**, such as a pulley, a wheel mechanism, or the like.

[0038] The implementations set forth in the foregoing description do not represent all implementations consistent with the subject matter described herein. Instead, they are merely some examples consistent with aspects related to the described subject matter. Although a few variations have been described in detail herein, other modifications or additions are possible. In particular, further features and/or variations can be provided in addition to those set forth herein. For example, the implementations described above can be directed to various combinations and sub-combinations of the disclosed features and/or combinations and sub-combinations of one or more features further to those disclosed herein. In addition, the logic flows depicted in the accompanying figures and/or described herein do not necessarily require the particular order shown, or sequential order, to achieve desirable results. The scope of the following claims may include other implementations or embodiments.

What is claimed is:

1. An electricity collection apparatus comprising:

a tether; and

a windsock formed of an insulation material with a metal deposited on at least a portion of the insulation material, the metal of the windsock being electrically connected to the tether, the windsock to extend in a direction of ambient wind to provide a surface area based on the ambient wind, the surface area providing the metal to collect electrical energy from the ambient wind, the collected electrical energy being transferred to the tether.

2. The apparatus in accordance with claim **1**, wherein the metal is deposited on at least one of opposite sides of the insulation material.

3. The apparatus in accordance with claim **2**, wherein the metal is a metal selected from a group of metals comprising gold, silver, copper and aluminum.

4. The apparatus in accordance with claim **1**, wherein the insulation material is a polyester film.

5. The apparatus in accordance with claim **1**, wherein the transferred electrical energy at the tether is further transferred from the tether to an electricity storage device.

6. The apparatus in accordance with claim **1**, wherein the metal comprises conductors that are light-weight thereby allowing payload requirements for adjusting the windsock to an optimum location to be within a predetermined value, the optimum location comprising a height of the windsock from ground level.

7. The apparatus in accordance with claim **6**, wherein the deposited metal conductors form a pattern that maximizes the collection of the electrical energy at the optimum location.

8. An electricity collection apparatus comprising:

a windsock formed of an insulation material having a metal deposited on at least a portion of the insulation material, the windsock to extend in a direction of ambient wind to provide a surface area based on the ambient wind, the surface area providing the metal to collect electrical energy from the ambient wind; and

an electrically conductive tether connected with windsock to anchor the windsock in the direction of the ambient wind, the electrically conductive tether being electrically connected to the metal of the insulation material, the collected electrical energy being transferred to the tether.

9. The apparatus in accordance with claim **8**, wherein the metal is deposited on at least one of opposite sides of the insulation material.

10. The apparatus in accordance with claim **9**, wherein the metal is a metal selected from a group of metals comprising gold, silver, copper and aluminum.

11. The apparatus in accordance with claim **8**, wherein the insulation material is a polyester film.

12. The apparatus in accordance with claim **8**, wherein the transferred electrical energy at the tether is further transferred from the tether to an electricity storage device.

13. The apparatus in accordance with claim **8**, wherein the metal comprises conductors that are light-weight thereby allowing payload requirements for adjusting the windsock to an optimum location to be within a predetermined value, the optimum location comprising a height of the windsock from ground level.

14. The apparatus in accordance with claim **13**, wherein the metal conductors form a pattern that maximizes the collection of the electrical energy at the optimum location.

15. An electricity collection apparatus comprising:

a windsock formed of an insulation material the windsock to extend in a direction of ambient wind to provide a surface area based on the ambient wind;

one or more electrical conductors provided on at least a portion of the insulation material to collect electrical energy from the ambient wind; and

an electrically conductive tether connected with windsock to anchor the windsock in the direction of the ambient wind, the electrically conductive tether being electrically connected to the one or more electrical conductors provided on at least a portion of the insulation material to transfer the collected electrical energy from the one or more electrical conductors to an electrical storage.

* * * * *

US 20130033120A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2013/0033120 A1**

Ogram (43) **Pub. Date: Feb. 7, 2013**

(54) **ATMOSPHERIC STATIC ELECTRICITY COLLECTOR**

(71) Applicant: **Mark Ellery Ogram**, Tucson, AZ (US)

(72) Inventor: **Mark Ellery Ogram**, Tucson, AZ (US)

(21) Appl. No.: **13/648,246**

(22) Filed: **Oct. 9, 2012**

Related U.S. Application Data

(63) Continuation of application No. 13/356,631, filed on Jan. 23, 2012, now Pat. No. 8,283,811, which is a continuation of application No. 12/321,306, filed on Jan. 16, 2009, now Pat. No. 8,102,082, which is a continuation-in-part of application No. 12/218,297, filed on Jul. 14, 2008, now Pat. No. 7,855,476.

Publication Classification

(51) **Int. Cl.**
H02G 11/00 (2006.01)

(52) **U.S. Cl.** **307/145**

(57) **ABSTRACT**

An antenna for the collection of atmospheric static electricity in which an electrically conductive hub is suspended from a balloon or blimp via a tether. The hub is either solid or uses a spoke/arm arrangement. A number of rods extend from the hub enhance the collection of atmospheric static electricity. The collected atmospheric electricity is conducted from the rods to an electrical connection where the electricity is conducted to earth via a conductive line.

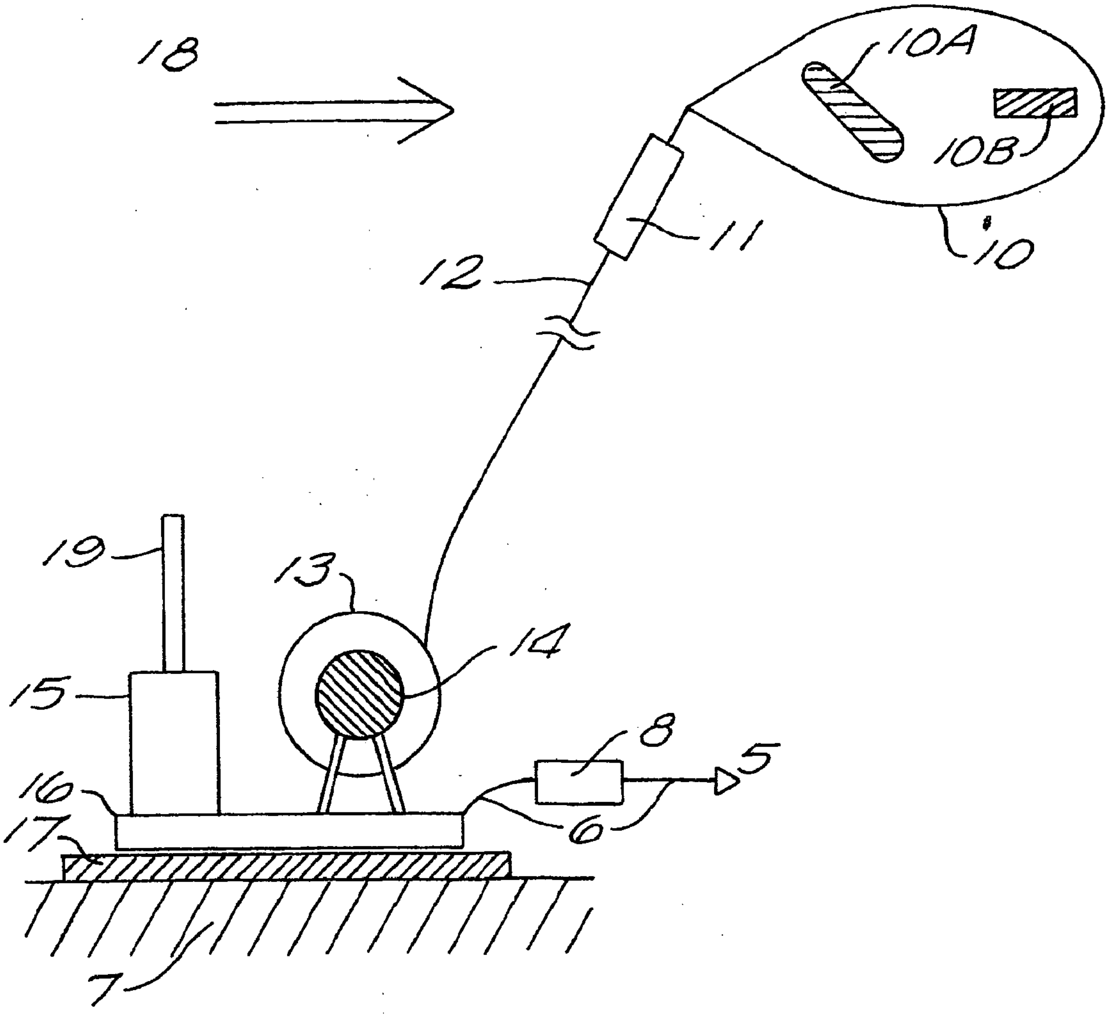

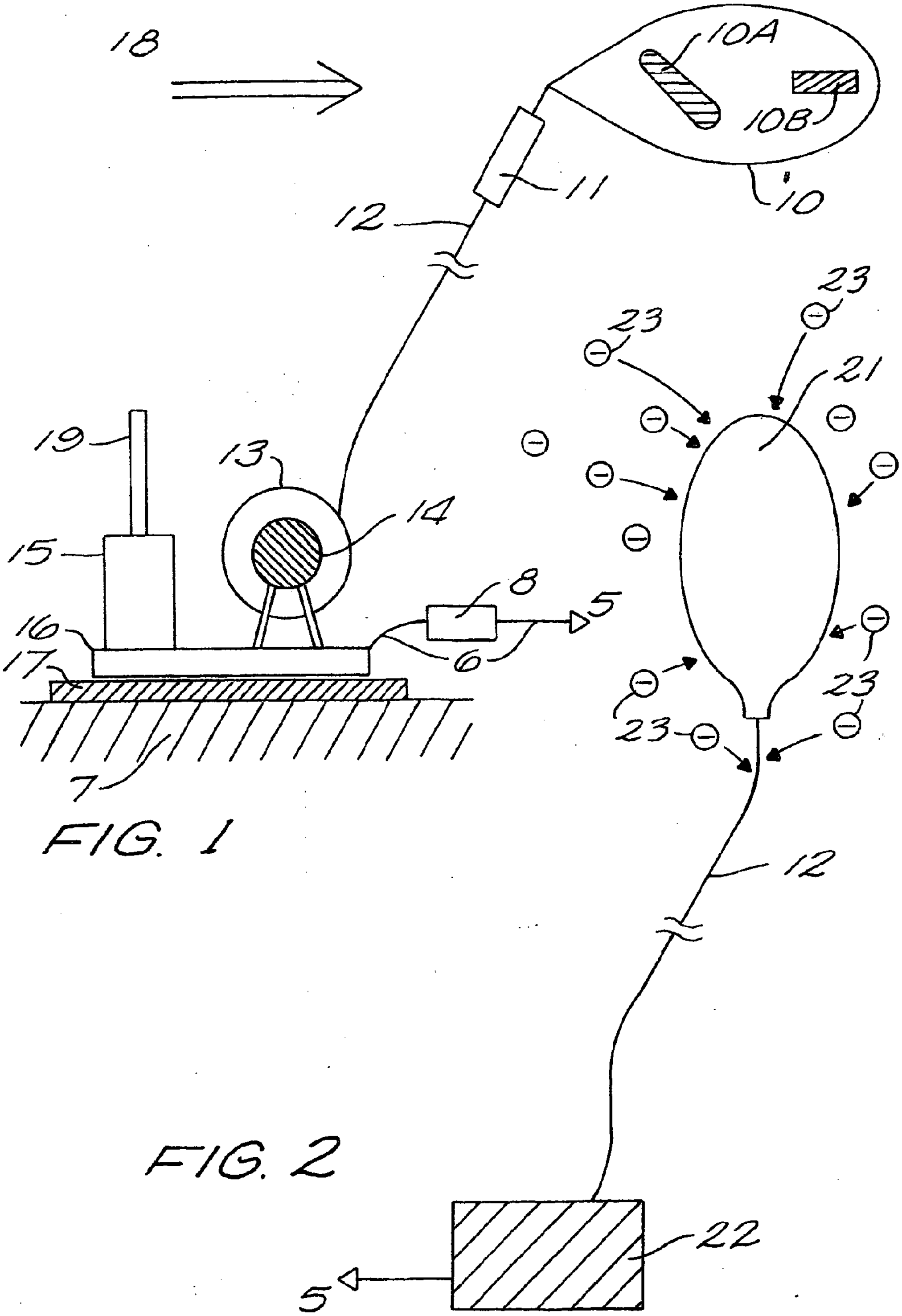
18
10A
10B
10
11
12
19
13
14
15
16
17
8
5
6
7
23
23
21
23
23
23
12
22
5

FIG. 1

FIG. 2

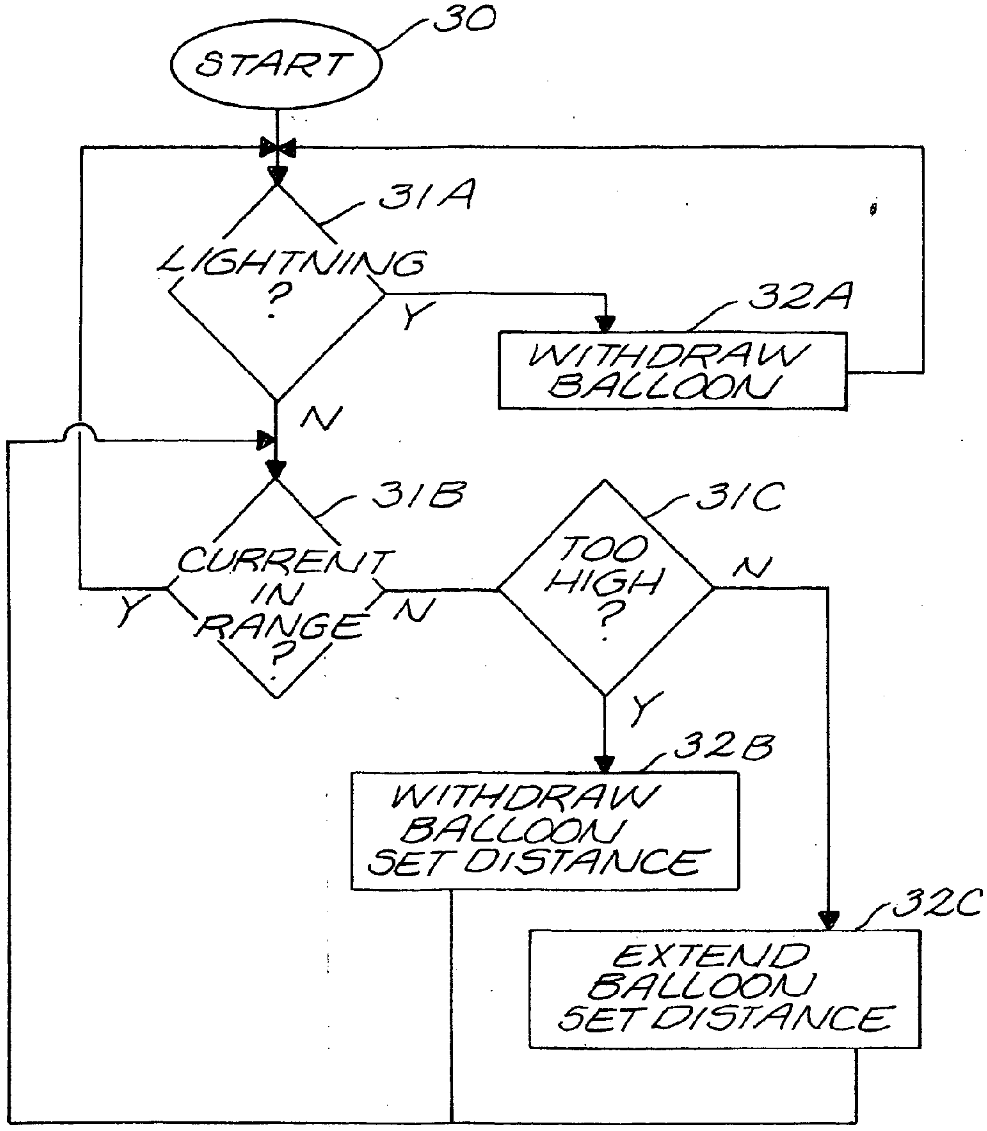
30
START
31A
LIGHTNING ?
Y
N
32A
WITHDRAW BALLOON
31B
CURRENT IN RANGE ?
Y
N
31C
TOO HIGH ?
N
Y
32B
WITHDRAW BALLOON SET DISTANCE
32C
EXTEND BALLOON SET DISTANCE

FIG. 3

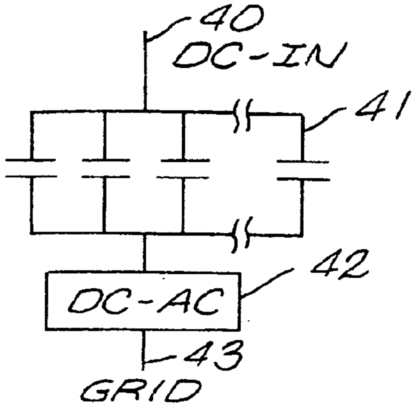
40
DC-IN
41
DC-AC
42
43
GRID

FIG. 4A

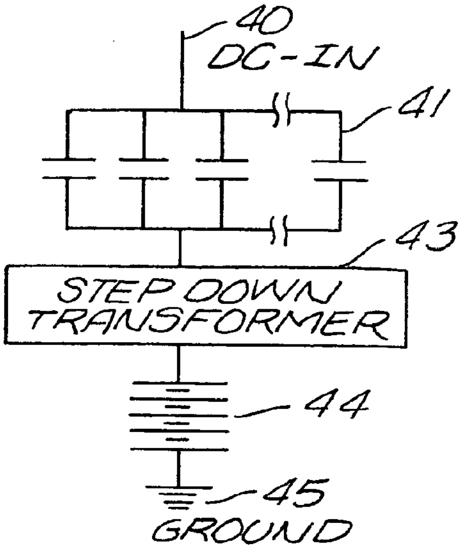
40
DC-IN
41
43
STEP DOWN TRANSFORMER
44
45
GROUND

FIG. 4B

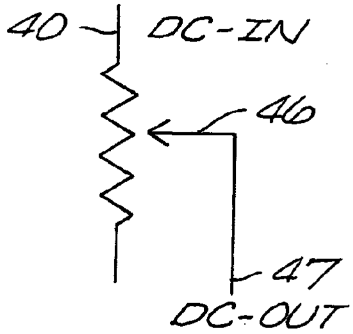
40
DC-IN
46
47
DC-OUT

FIG. 4C

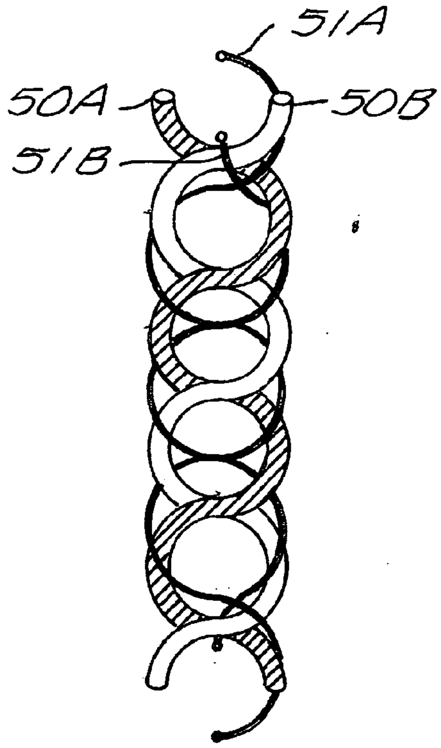
51A
50A
50B
51B

FIG. 5

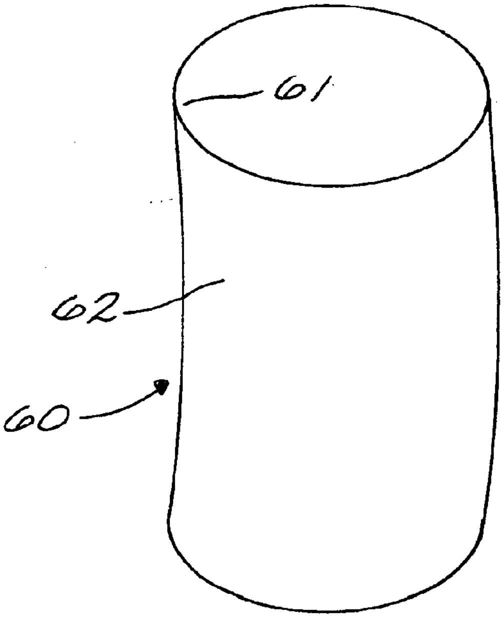

FIG. 6A

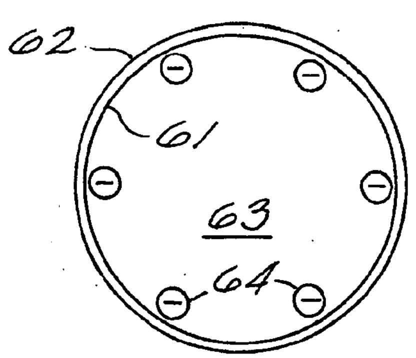

FIG. 6B

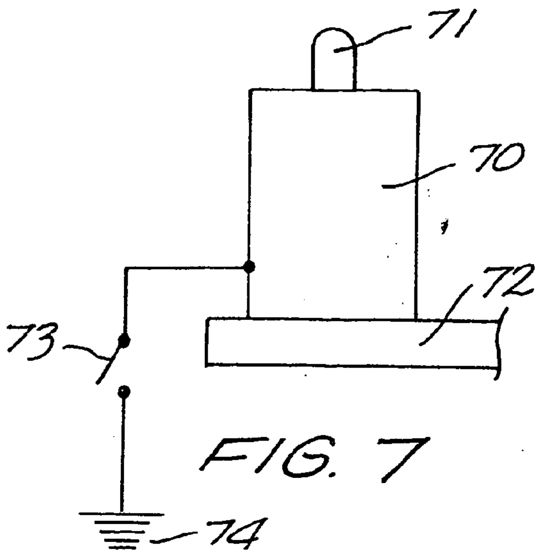
71
70
72
73
74

FIG. 7

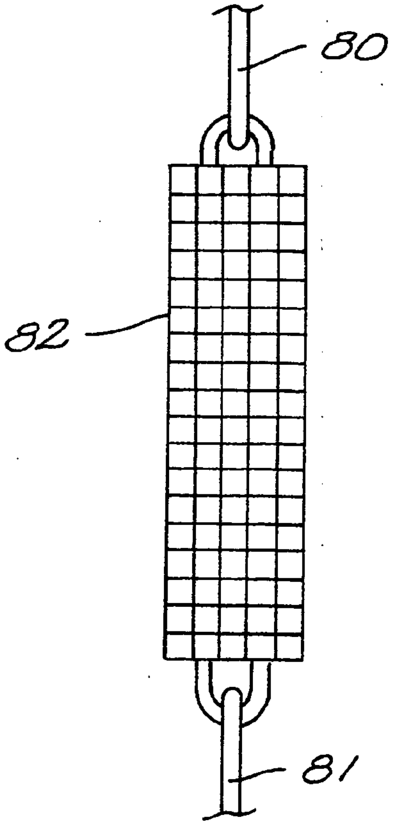

FIG. 8A

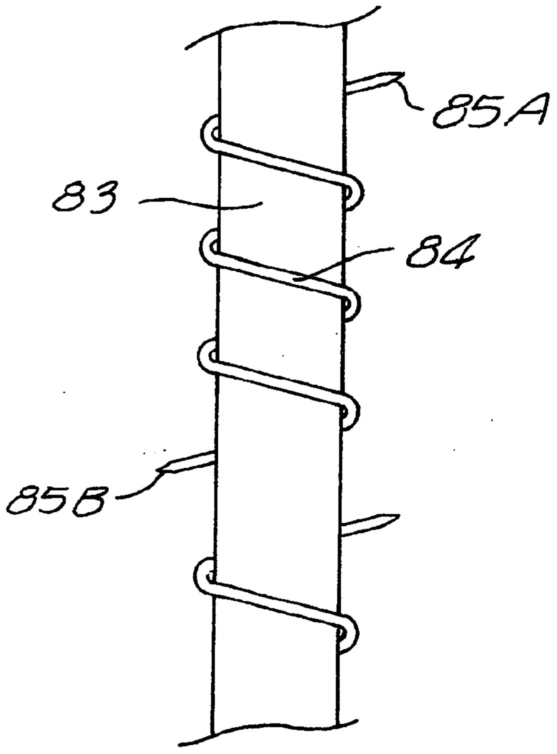
85A
83
84
85B

FIG. 8B

FIG. 9A

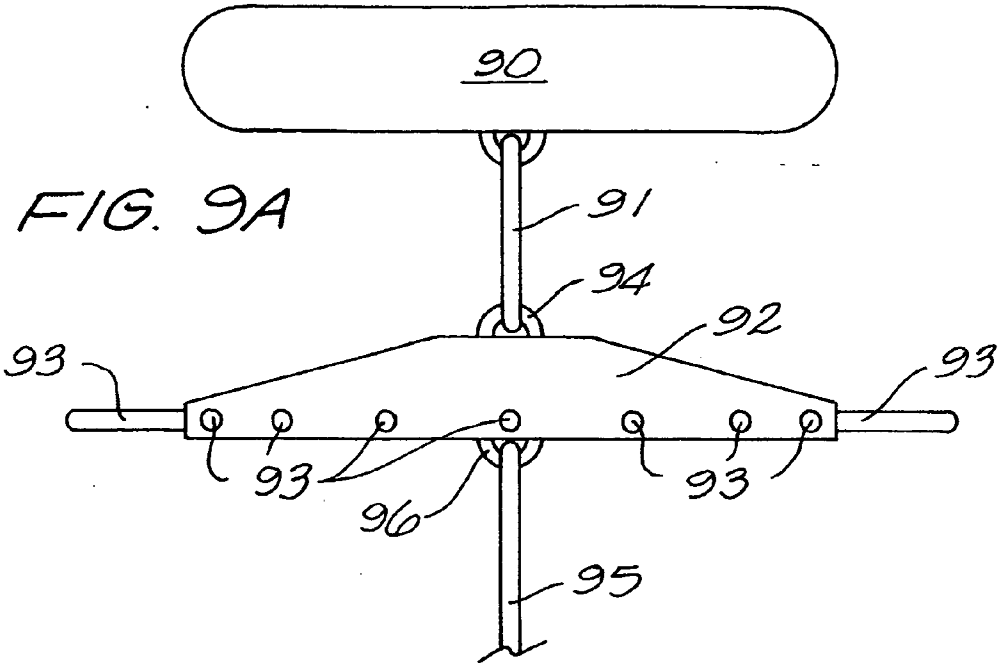
90
91
94
92
93
93
93
93
93
96
95

FIG. 9B

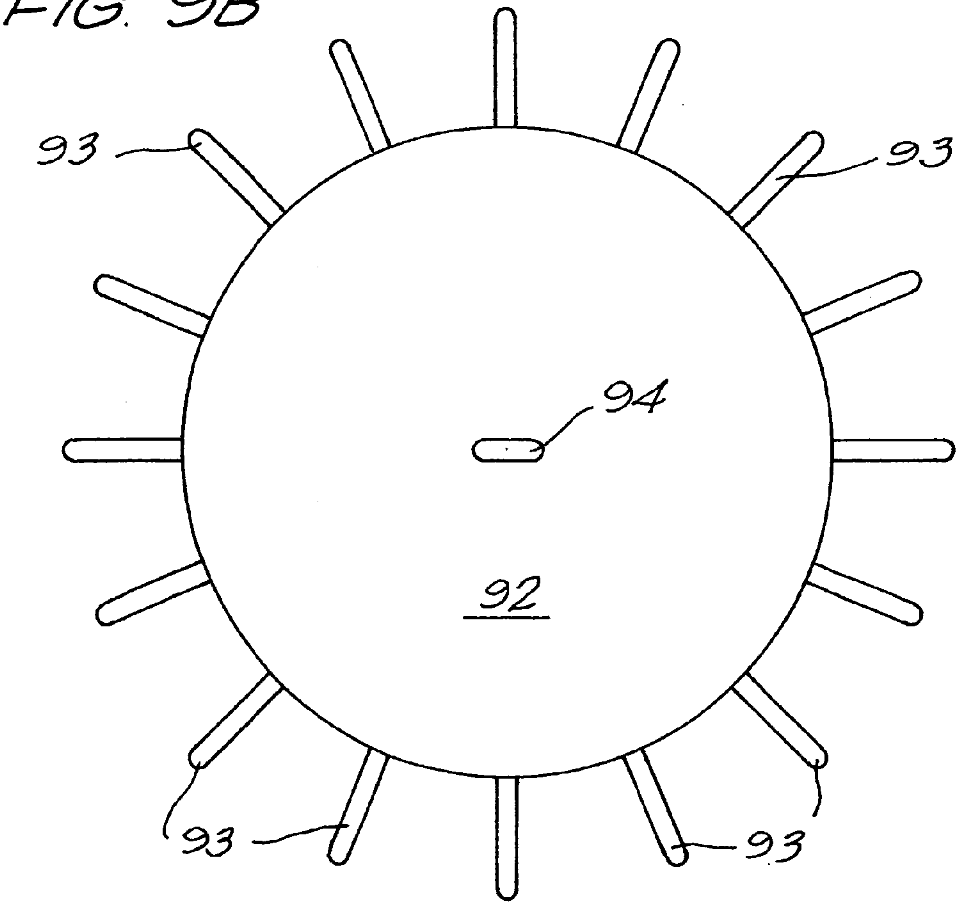
93
93
94
92
93
93
93
93

FIG. 10

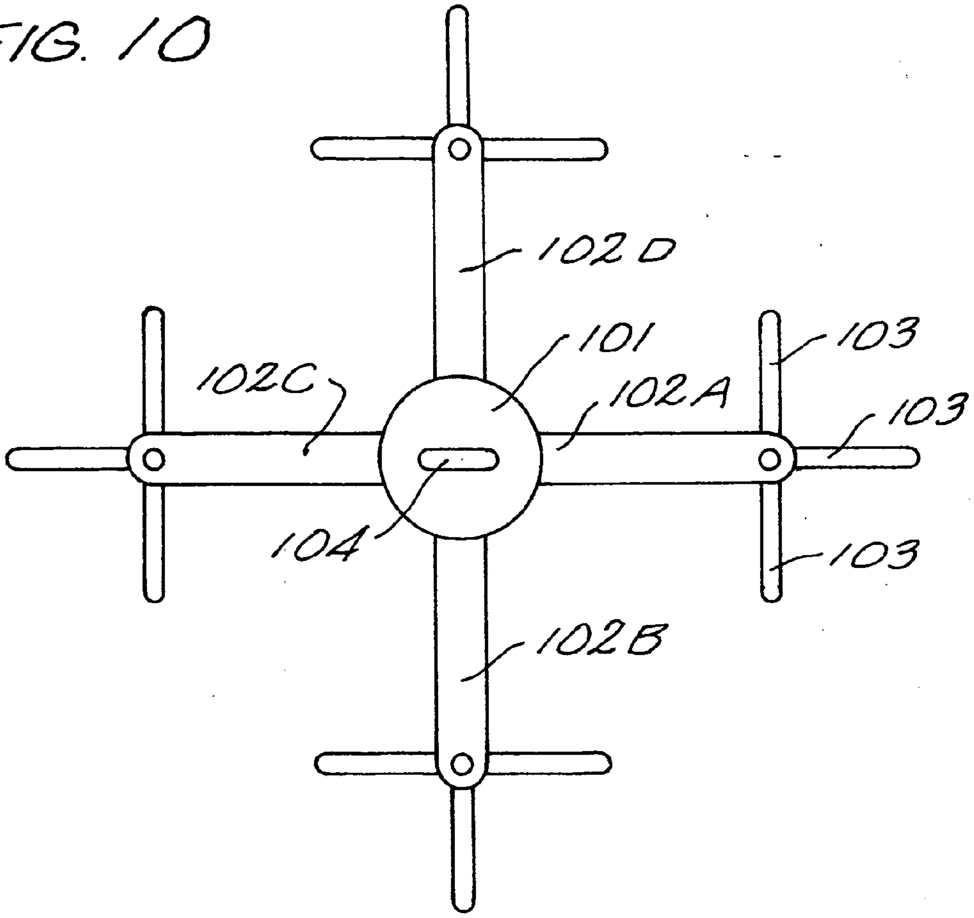
102D
101
103
102C
102A
103
104
103
102B

FIG. 11

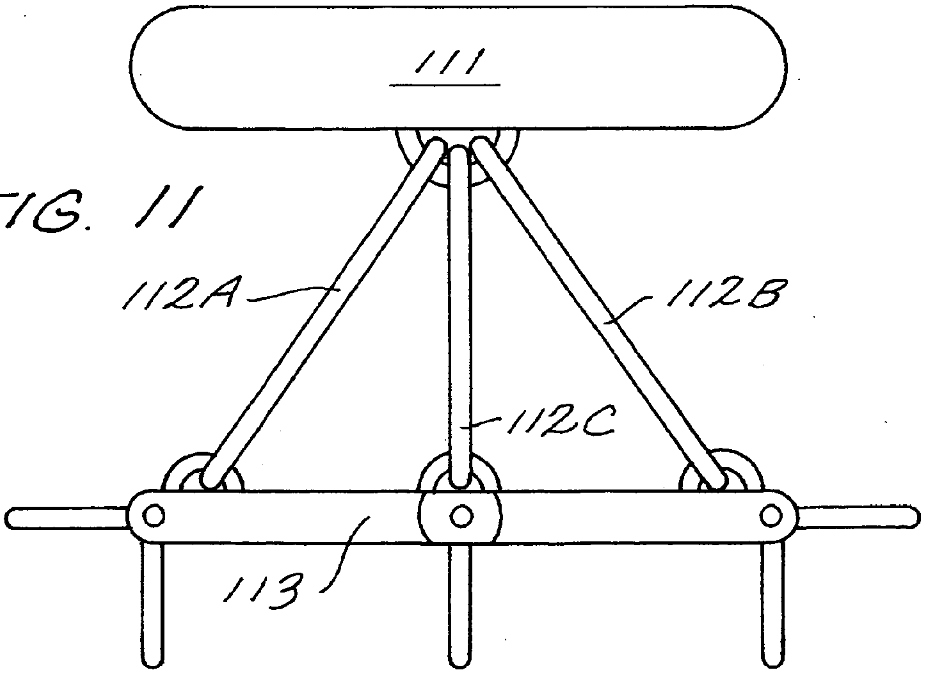
111
112A
112B
112C
113

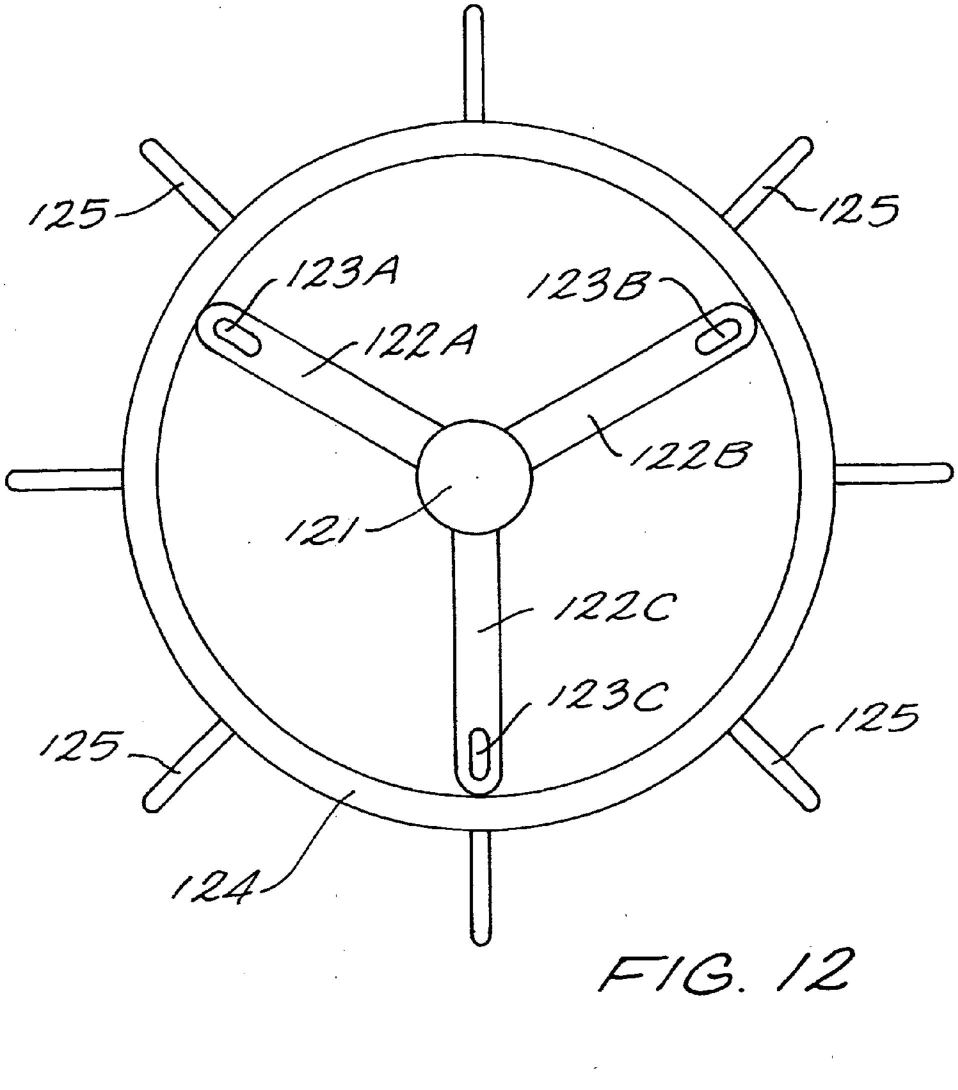
125
125
123A
123B
122A
122B
121
122C
123C
125
125
124

FIG. 12

ATMOSPHERIC STATIC ELECTRICITY COLLECTOR

BACKGROUND OF THE INVENTION

[0001] This invention relates to the production of electrical energy and more particularly to The collection of static electricity from the atmosphere.

[0002] This is a continuation in part of U.S. patent application Ser. No. 12/218,297, filed on Jul. 14, 2008, and entitled, "Atmospheric Electrical Generator".

[0003] Everyone is familiar with Benjamin Franklin's kite experiment of 1752. Using a kite whose string had become wet, negative charges from the passing clouds flowed into the string, down to the suspended key, and then into a Leyden jar via a thin metal wire. Franklin was protected by a dry silk string; but, when Franklin's knuckle came too close to the key, he received a strong shock. Fortunately, Benjamin Franklin was not killed, others who tried this same experiment were not so lucky.

[0004] Since then, the formation of lightning has remained something of a mystery. Lightning bolts are triggered when a negatively charged cloud base induces a positive charge from the ground, thereby forming a "pathway" for the discharge of the collected electrical energy.

[0005] Lightning travels up to 60,000 miles per hour with a flash that is brighter than ten million 100-watt lightbulbs. This wattage is as much power as is produced by all of the electricity plants in the United States and with a voltage of up to 300 million volts.

[0006] It is this very fact, the power within lightning is immense, that has prevented any successful collection of the electrical energy from lightning. The electricity in lightning is far too extreme for current technology to harness.

[0007] While lightning has, attracted a energy starved industrial world, no one has developed any technique to harness this naturally occurring electrical source.

[0008] It is clear there is a continuing need for an electrical source other than carbon-based fuels and that the naturally occurring electricity in the atmosphere is being ignored.

SUMMARY OF THE INVENTION

[0009] The invention is a mechanism which taps into the naturally occurring static electricity in the atmosphere. Whereas heretofore, the attempt to gamer electricity from the atmosphere has focused exclusively on capturing lightning, the present invention syphons off the static electricity which is generated from any agitated air and voids lightning.

[0010] Lightning is only the final discharge of the static electricity, whether that lightning is intra-cloud lightning, cloud-to-ground lightning, or inter-cloud lightning. Other types of final discharges are known as heat lightning, summer lightning, sheet lightning, ribbon lightning, silent lightning, ball lightning, bead lightning, elves, jets, and sprites. Well before these discharges are observed, as the atmosphere becomes agitated by wind or thermal, static electricity is being generated.

[0011] The present invention recognizes that this static electricity is being formed and creates a mechanism to capture it.

[0012] The mechanism of this invention utilizes an aircraft such as a lighter than air balloon. While the preferred embodiment uses a foil balloon, a variety of other aircraft are obvious to those of ordinary skill in the art, including, but not limited to: gliders, rubber balloons (such as weather balloons), biaxially-oriented polyethylene terephthalate polyester film balloons, and latex balloons.

[0013] Within this discussion, the balloon is referenced, but, the invention is not intended to be limited solely to balloons.

[0014] The balloon is sent aloft and is tethered by a conductive line. In this context, the conductive line may be any obvious to those of ordinary skill in the art. For the preferred embodiment, the conductive line is a generically referred to as a "poly-rope" and is commercially available through a variety of sources. A suitable conductive line is described in U.S. Pat. No. 5,203,542, entitled "Apparatus for Improved Electric Fence Wire Construction for use with Intensive Grazing" issued Apr. 20, 1993, to Coley, et al. and incorporated hereinto by reference.

[0015] The conductive line is played out of a winch to control the altitude of the balloon. The motor controlling the winch is able to reverse direction to both extend and withdraw the conductive line which is wrapped around a spool on the winch. The winch/spool combination are part of a base unit.

[0016] In some embodiments of the invention, the spool is constructed of rubber so as to insulate the conductive line from the winch assembly. In this embodiment, only the conductive line is charged by the atmospheric static electricity while the winch remains neutral.

[0017] In yet another embodiment, the winch/spool are part of a base unit which is itself isolated from the ground by an insulator. In this embodiment, the entire base unit is charged by the atmospheric static electricity,

[0018] A conductor, such as an insulated wire, is electrically connected to the conductive line. In one embodiment, where the conductive line is electrically isolated from the spool and winch motor, the conductor is connected to the conductive line. In the embodiment where the conductive line is electrically connected to the base unit, then the conductor is connected anywhere on a metalic base unit.

[0019] The other end of the conductor is connected to a load. The load in this case can be any of a variety of electrical loads well known to those of ordinary skill in art, including, but not limited to a motor, a battery system, or the electrical grid for the system.

[0020] In the preferred embodiment, a sensor array is used to monitor the activities both at the base unit (such as electrical flow within the conductor) and in the surrounding locale.

[0021] A sensor monitoring the electrical flow (i.e. voltage and/or current) within the conductor is used to monitor the electrical activity within the conductor,

[0022] In the preferred embodiment, a lightning sensor monitors for lightning activity within the locale. As noted earlier, the electrical characteristic of lightning is so extreme that ideally this discharge is avoided as it might damage the mechanism of this invention.

[0023] The sensor array is utilized by a controller, such as microprocessor, programmed to operate the mechanism as outlined herein.

[0024] The controller operates the winch motor to extend or withdraw the conductive line and by extension the altitude of the balloon. The controller is programmed to operate the winch by monitoring the electrical characteristics of the conductor and adjusting the balloon's altitude to maintain these characteristics within the conductor within a preset range.

[0025] This preset range is established either in the base programming of the controller or is established by an operator of the system.

[0026] As example, by controlling the amount of current being withdrawn from the atmosphere, the mechanism operates within a safe range and also provides a relatively stable current flow from which a variety of activities can take place (such as DC-AC conversion).

[0027] The controller also utilizes the lightning sensor to protect the mechanism from a lightning strike. Should lightning be detected within a pre-determined range (as established by the software or defined by an operator), then the balloon is pulled down to minimize the risk of damage from a lightning strike.

[0028] An aspect of the present invention is the use of an antenna which are used to collect the atmospheric static electricity. The antenna is shaped as a hub which is suspended from the blimp/balloon. The hub is ideally spoked shaped although an alternative embodiment uses a solid hub.

[0029] A number of rods extend from the hub so as collect atmospheric static electricity. These rods are ideally rounded at the ends to enhance the attraction of the atmospheric static electricity.

[0030] The collected atmospheric electricity is conducted from the rods to an electrical connection on the hub where the electricity is conducted to a power plant on earth such as described above.

[0031] The invention, together with various embodiments thereof will be more fully explained by the following description of the accompanying drawings.

DRAWINGS IN BRIEF

[0032] FIG. **1** diagrams the preferred embodiment of the invention.

[0033] FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

[0034] FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

[0035] FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

[0036] FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

[0037] FIGS. **6**A and **613** illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

[0038] FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

[0039] FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

[0040] FIGS. **9**A and **9**B are side views and top views of an embodiment of the antenna used to collected atmospheric electricity.

[0041] FIG. **10** is a top view of an alternative embodiment of the antenna of this invention.

[0042] FIG. **11** is a side view of yet another alternative embodiment of the antenna used to collect atmospheric electricity.

[0043] FIG. **12** is the preferred embodiment of the antenna of this invention.

DRAWINGS IN DETAIL

[0044] FIG. **1** diagrams the preferred embodiment of the invention.

[0045] Balloon **10** is an aircraft which, in this illustration, is a lighter than air balloon. Wings **10**A , extending from the body of balloon **10**, provide additional lift in air flow **18**. Tail **10**B helps to stabilize balloon **10**.

[0046] Balloon **10** is tethered to the ground via conductive line **12**. As noted earlier, a variety of configurations and materials are available to serve as conductive line **12**. In this illustration, a poly-wire is used. Poly-wire is commercially available through a variety of vendors, including, but not limited to: Jeffers Livestock and Sareba Systems, Inc. of Ellendale, Minn.

[0047] In this embodiment, located proximate to balloon **10**, is an electrical collection enhancement lead **11** which assists in the collection of the static electrical charge in the atmosphere. Electrical collection enhancement lead **11** is configured to attract the static charge and conduct the charge into the conductive line **12**.

[0048] The electricity flows down the conductive line into spool **13**, where the conductive line **12** is collected and either withdrawn or dispensed through operation of winch motor **14**.

[0049] Winch motor **14** and spool **13** are mounted onto base unit **16** which is electrically isolated from ground **7** using insulator **17**. Note, in this embodiment of the invention, when electricity is being collected from the atmosphere, the entire base unit **16** becomes charged. In another embodiment of the invention, spool **13** is constructed of rubber, thereby preventing base unit **16** from becoming charged, thereby restricting the charging from the atmosphere to only conductive line **12**.

[0050] In this embodiment, conductor **6** is connected to base unit **16** (since the entire base unit **16** is charged and the base unit is metalic) to communicate the electrical current to load **5**. Conductor **6** is ideally an insulated wire.

[0051] The electrical current through conductor **6** is measured using sensor **8**.

[0052] In the alternative embodiment discussed above, where only the conductive line **12** is charged, then conductor **6** is connected to conductive line **12**.

[0053] Controller **15**, located in this embodiment on base unit **16**, operates winch motor **14** in response to signals from sensor **8** (measuring the current being discharged to load **5**) to maintain the current flow within a pre-defined range. As the current flow diminishes, then the conductive line **12** extended from spool **13** to increase the altitude of balloon **10** to that more static charge from the atmosphere is gathered; as the current flow falls exceeds a preset level, conductive line **12** is withdrawn onto spool **13** to decrease the static charge being collected from the atmosphere.

[0054] The range of current flow through conductor **6** is ideally set by the program, although some embodiments of the invention permit an operator to establish this range of operation.

[0055] In an alternative embodiment, the sensor monitoring conductor **6** monitors the voltage therein.

[0056] In the preferred embodiment of the invention, controller **15** is also equipped with a lightning sensor **19**. In this embodiment, when lightning is sensed within a preset range, then substantially all of conductive line **12** is wound onto spool **13** to pull balloon **10** near the ground and protect the entire mechanism from being damaged from a lightning discharge.

[0057] In the preferred embodiment, the "safe" distance form lightning is set in the programming of controller **15** and is ideally two miles; other embodiments permit the operator to "safe" distance.

[0058] There are a variety of lightning sensors well known to those of ordinary skill in the art, including, but not limited to those described in: U.S. Pat. No. 7,016,785, entitled "Lightning Detection" issued to Makela, et al. on Mar. 21, 2006; U.S. Pat. No. 6,829,911, entitled "Lightning Detection and Prediction Alarm Device" issued to Jones, et al. on Dec. 7, 2004; U.S. Pat. No. 7,200,418, entitled "Detection of Lightning" issued to Karikuranta, et al. on Apr. 3, 2007; and U.S. Pat. No. 6,961,662, entitled "Systems and Methods for Spectral Corrected Lightning Detection" issued to Murphy on Nov. 1, 2005; all of which are incorporated hereinto by reference.

[0059] In another embodiment of the invention, controller **15** is not located on base unit **16**, rather it is remote and communicates its control signals to winch motor **14** using radio waves.

[0060] FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

[0061] Static charges **23** are generated in the atmosphere by agitated air. These static charges are often collected at the bottom of clouds, but exist in other environments as well.

[0062] Balloon **21** is extended into this strata of static charges **23** which are then attracted to conductive line **12** to flow to base unit **22** and then onto load **5**.

[0063] By increasing or decreasing the altitude of balloon **21** (defined by the length of the extended conductive line **12**), conductive line **12** is selectively exposed to varying densities and levels of the static charge strata, and by extension, the current flow or voltage is increased or decreased.

[0064] FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

[0065] Once the program starts **30**, the lightning sensor is checked to determine if lightning has occurred within the unsafe range **31**A, if it has, then the balloon is lowered **32**A, and the program continues monitoring the status of lightning until no lightning is detected.

[0066] When the lightning status is acceptable, then the current within the conductor is checked to see if the current is within the prescribed range **31**B. lithe current is acceptable (within range) the program returns to check the lightning status **31**A; otherwise a determination is made to see if the current is above the prescribed range **31**C.

[0067] If the current is above the prescribed range, then the altitude of the balloon is withdrawn a set amount **32**B (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

[0068] If the current is below the prescribed range, then the altitude of the balloon is extended a set amount **32**C (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

[0069] In this manner of feed-back and minor adjustments in the altitude of the balloon, the current is maintained within a prescribed range which can be handled by the downstream electrical system.

[0070] As noted earlier, some embodiments of the invention monitor the voltage instead of the current

[0071] FIGS. **4**A, **413**, and **4**C are electrical schematics for handling the static charge from the atmosphere.

[0072] By maintaining the voltage being collected in a prescribed range, an electrical conversion system is easily designed. While FIGS. **4**A, **4**B, and **4**C illustrate some electrical configurations, those of ordinary skill in the art readily recognize a variety of other configurations which will serve the same function.

[0073] Referencing FIG. **4**A, Direct Current In (DC INT) **40** is buffered by a gang of capacitors **41** before being communicated to a DC/AC converter **42**. The DC/AC converter converts the direct current into a an alternating current suitable for placement over an existing electrical grid **43** such as normally found from a power-plant.

[0074] Those of ordinary skill in the art readily recognize a variety of DC/AC converters, including, but not limited to: U.S. Pat. No. 7,394,671, entitled "Controller IC, DC-AC Conversion Apparatus, and parallel running system of DC-AC Conversion Apparatuses" issued to Fukumoto, et al. on Jul. 1, 2008; and, U.S. Pat. No. 7,330,366, entitled "DC-AC Converter" issued to Lee, et al. on Feb. 12, 2008; all of which are incorporated hereinto by reference.

[0075] FIG. **4**B illustrates an electrical arrangement suitable for use in charging a battery. DC IN **40** is buffered by capacitor bank **41** before entering into a step down transformer **43**. Step down transformer **43** reduces the voltage so that the voltage can safely be introduced into battery **44** which is connected to ground **45** at the battery's other pole.

[0076] Those of ordinary skill in the art readily recognize a variety of batteries which will work in this capacity, including, but not limited to those described in: U.S. Pat. No. 7,378,181, entitled "Electric Storage Battery Construction and Manufacture" issued to Skinlo on May 27; 2008; U.S. Pat. No. 7,388,350, entitled "Battery with Electronic Compartment" issued to Wright on Jun. 17, 2008; U.S. Pat. No. 7,397,220, entitled "Connection Member and Battery Pack" issued to Uchida, et al. on Jul. 8, 2008; and, U.S. Pat. No. 7,375,492, entitled "Inductively Charged Battery Pack" issued to Calhoon, et al. on May 20, 2008; all of which are incorporated hereinto by reference.

[0077] In FIG. **4**C, DC IN **40** is fed into an adjustable rheostat **46** which is controlled by the controller so that the DC OUT **47** falls within a specified range.

[0078] FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

[0079] This type of conductive line is commonly called poly-wire and consists of multiple interwoven strands of plastic **50**A and **50**B woven into a cord or rope arrangement having intertwined therein exposed metal wires **51**A and **51**B. While this illustration shows two plastic strands and two metal wires, any number of possible combinations is possible.

[0080] The exposed metal wires **51**A and **51**B attract the atmospheric static charge and transmit the charge down to the base unit (not shown).

[0081] FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

[0082] This conductive line utilizes a tube **60** having an outer layer **62** of PET Film (Biaxially-oriented polyethylene terephtalate polyester film) which provides exceptionally high tensile strength and is chemically and dimensionally stable. The tube has an ideal diameter of between two and three inches.

[0083] An interior metal coating **61** provides an initial conduit for the flow of static charge. The static charge through the metal forces the tube to expand due to the repulsion experienced by like charges. Further, the flow of electricity causes the interior of the tube **60** to become ionized to provide an additional pathway for the atmospheric static charges to the base unit (not shown).

[0084] Because outer layer **62** provides a gas barrier, the resulting ionization is not dissipated by air currents, thereby providing a highly stable pathway.

[0085] FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

[0086] In this embodiment, controller box **70**, resting on insulating pad **72**, is in communication with the sensors as described above. Using the input from these sensors, when there is flow of electricity through the base unit, warning flashing light **71** is illuminated. To electrically neutralize the mechanism, switch **73** is activated to pass any existing current into the ground **74**.

[0087] FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

[0088] Referencing FIG. **8**A, enhanced electrical collection lead **82** is a wire mesh which is in electrical communication with conductive line **81** and balloon **80**. Because of the significant amount of metal exposed by enhanced electrical collection lead **82**, more static electricity from the atmosphere is drawn to the collection lead **82**, and then down conductive line **81** to the base unit (not shown).

[0089] Conductive lead **82** is positioned proximate to balloon **80**.

[0090] In FIG. **8**B, poly-wire **83** has enhanced electrical collection leads **84** wrapped therearound. Collection leads **84** have pointed ends **85**A and **85**B which have a propensity to attract more electricity than rounded ends do.

[0091] FIGS. **9**A and **9**B are side views and top views of an embodiment of the antenna used to collected atmospheric electricity.

[0092] Blimp **90** supports antenna **92** via tether **91** which is attached to antenna **92** by connection **94**. In this embodiment, antenna **92** is made of electrically conductive material and is ideally light in weight to lessen the payload requirements on blimp **90**.

[0093] Encircling antenna **92** and extending therefrom are rods **93** which are rounded at their distal ends so as to enhance attraction of the atmospheric static electricity. The static electricity is communicated to connection **96** and then to electrical line **95** which communicates the electricity to the ground based station (not shown) as described above.

[0094] FIG. **10** is a top view of an alternative embodiment of the antenna of this invention.

[0095] In this embodiment of the antenna a central hub **101** has the blimp connection **104** (capable of receiving the tether to the blimp) secured thereto. In this illustration, four arms **102**A, **102**B, **102**C, and **102**D extend from hub **101**. Hub **101**, and arms **102**A, **10213**, **102**C, and **1020** are all electrically conductive in this embodiment.

[0096] At the end of each arm are rods **103** which are used to enhance the collection of the atmospheric static electricity.

[0097] Although this embodiment illustrates four arms, the invention is not intended to be limited to four arms, rather, any number of arms may be used and the number of rods extending from the distal ends of the arms also varies.

[0098] FIG. **11** is a side view of yet another alternative embodiment of the antenna used to collect atmospheric electricity.

[0099] In this embodiment, several tethers **112**A, **11213**, and **1120**, are used to secure the antenna **113** to the blimp **111**. This arrangement of several tethers provides heightened stability of the antenna by reducing the affects wind will have on the antenna.

[0100] FIG. **12** is the preferred embodiment of the antenna of this invention.

[0101] In this embodiment of the antenna, arms **122**A, **122**B, and **122**C extend from a central hum and are electrically connected to rim **124**. Tether connectors **123**A, **123**B, and **123**C, are used to secure the antenna to the blimp or balloon.

[0102] Rods **125** extend from rim **124** to increase the collection of the static charges in the atmosphere.

[0103] It is clear from the foregoing that the present invention captures an entirely new source of electrical energy.

1-16. (canceled)

17. A system for collection of atmospheric static electricity comprising:

a balloon;

a hub that is electrically conductive and connected to the balloon by a tether, the hub having two or more arms that are electrically conductive, a first end of each of the two or more arms electrically connected to the hub, the hub further having a plurality of rods, a first end of each of the rods being electrically connected to a second end of at least one of the two or more arms;

a winch motor having a spool of conductive line, one end of the conductive line secured to the hub, a portion of the conductive line and the hub collecting electricity in the atmosphere, the winch motor selectively extending or withdrawing the conductive line from the spool;

a conductor having a first end electrically connected to the conductive line and a second end electrically connected to a load being powered by collected electricity from the conductive line;

an electrical flow sensor monitoring electrical flow through the conductor and generating an electrical flow indicia indicative of the electrical flow in the conductor; and

a controller to operate the winch motor to maintain the electrical flow indicia within a selected operating range.

18. The system according to claim **17**, wherein each of the plurality of rods has a rounded second end.

19. The system according to claim **18**, wherein at least two rods are attached to the second end of each of the two or more arms.

20. The system according to claim **17**, further comprising an electrically conductive rim connected to a second end of the two or more arms.

21. The system according to claim **20**, wherein at least a portion of the two or more electrically conductive arms are connected substantially at right angles to an exterior of the electrically conductive rim.

22. A system for collection of atmospheric static electricity comprising:

two or more arms that are electrically conductive;

a plurality of rods, each of the plurality of rods being electrically connected to at least one of the two or more arms; and

a balloon tether connected between a balloon and the two or more arms;

a winch motor that extends or withdraws a spool of conductive line, one end of the conductive line being secured to the two or more arms, an extended portion of the conductive line and the two or more arms being adapted to collect electricity in the atmosphere;

a conductor having a first end electrically connected to the conductive line and a second end electrically connected to a load being powered by collected electricity from the conductive line;

an electrical flow sensor monitoring electrical flow through the conductor and generating an electrical flow indicia indicative of the electrical flow in the conductor; and

a controller to operate the winch motor to maintain the electrical flow indicia within a selected operating range.

23. The system according to claim **22**, wherein each of the plurality of rods has a rounded end distal from the electrically conductive arm.

24. The system according to claim **23**, further including a balloon tether connection secured to the first end of the two or more arms.

25. The system according to claim **24**, further including an electrical connection secured to the first end of each of the two or more arms.

26. The system according to claim **25**, wherein the at least two rods attached to the two or more arms includes three or more rods.

27. The system according to claim **25**, further including an electrically conductive rim connected to a second end of the two or more arms.

28. The system according to claim **27**, wherein the plurality of arms are electrically connected to the electrically conductive rim.

29. The system according to claim **28**, wherein at least a portion of the plurality of rods are connected to an exterior of the electrically conductive rim substantially at right angles thereto.

30. A system for the collection of atmospheric static electricity comprising:

an electrically conductive body;

a balloon tether connected to a first side of the electrically conductive body and to a balloon; and

an electrical connection connected to a second side of the electrically conductive body;

a winch motor that selectively extends or withdraws a spool of conductive line, one end of the conductive line secured to the electrical connection, a portion of the conductive line and the electrically conductive body collecting electricity in the atmosphere, the winch motor selectively extending or withdrawing the conductive line from the spool;

a conductor having a first end electrically connected to the conductive line and a second end electrically connected to a load being powered by collected electricity from the conductive line;

an electrical flow sensor monitoring electrical flow through the conductor and generating an electrical flow indicia indicative of the electrical flow in the conductor; and

a controller receiving the electrical flow indicia and selectively operating the winch motor such that the electrical flow indicia remains within a selected operating range.

31. The system according to claim **30**, wherein the electrically conductive body comprises:

an electrically conductive hub;

at least three electrically conductive arms, each of the arms connected at a first end to the hub; and

an electrically conductive rim connected to the second end of each of the at least three arms.

32. The system according to claim **31**, further comprising a plurality of electrically conductive rods, a first end of each of the rods connected to the rim and extending therefrom substantially at rights angles thereto.

* * * * *

US 20150102676A1

(19) **United States**

(12) **Patent Application Publication**

Dinwiddie et al.

(10) **Pub. No.: US 2015/0102676 A1**

(43) **Pub. Date: Apr. 16, 2015**

(54) METHOD AND APPARATUS FOR EXTRACTING AND CONVEYING ELECTRICAL ENERGY FROM THE EARTH'S IONOSPHERE CAVITY

(71) Applicant: **Earth Energies, Inc.**, Johns Creek, GA (US)

(72) Inventors: **John Dinwiddie**, Cary, NC (US); **Terry L. Wright**, Suwanee, GA (US)

(21) Appl. No.: **14/509,846**

(22) Filed: **Oct. 8, 2014**

Related U.S. Application Data

(60) Provisional application No. 61/889,894, filed on Oct. 11, 2013.

Publication Classification

(51) **Int. Cl.**
H01F 38/14 (2006.01)

(52) **U.S. Cl.**
CPC *H01F 38/14* (2013.01); *H01F 2038/146* (2013.01)

(57) ABSTRACT

The system and apparatus of one or more embodiments of the present invention extracts, conditions, and conveys electric power from the earth ionosphere cavity through the integrated and collaborative operation of the system and apparatus consisting of a capacitively-coupled insulated elevated terminal (coupled capacitor upper plate), an evacuated spark gap, an integrated step-down transformer and resonant capacitor, and a ground terminal (coupled capacitor lower plate). The architecture governing exemplary embodiments of the system and apparatus of the present invention emulates the natural architecture governing the interaction of living trees with the electrical energy resident in the earth ionosphere cavity. The implementation of such exemplary embodiments of the present invention utilizes standard and customized components appropriate for their function within the system and apparatus.

The extracted electric power from the earth ionosphere cavity manifests in the form of broadband electromagnetic waves in the 0 to 200 Hz frequency range. Electric energy from these waves produce magnetic fields inside the integrated step-down transformer which are added together by the primary coil of the integrated step-down transformer, and delivered as 60 Hz 120/240 VAC to an electric load via the secondary coil of the integrated step-down transformer resonating at 60 Hz due to the influence of the integrated resonant capacitor.

The system and apparatus of exemplary embodiments of the present invention extracts, conditions, and conveys ionosphere cavity-resident electric power in alternating current and direct current realms.

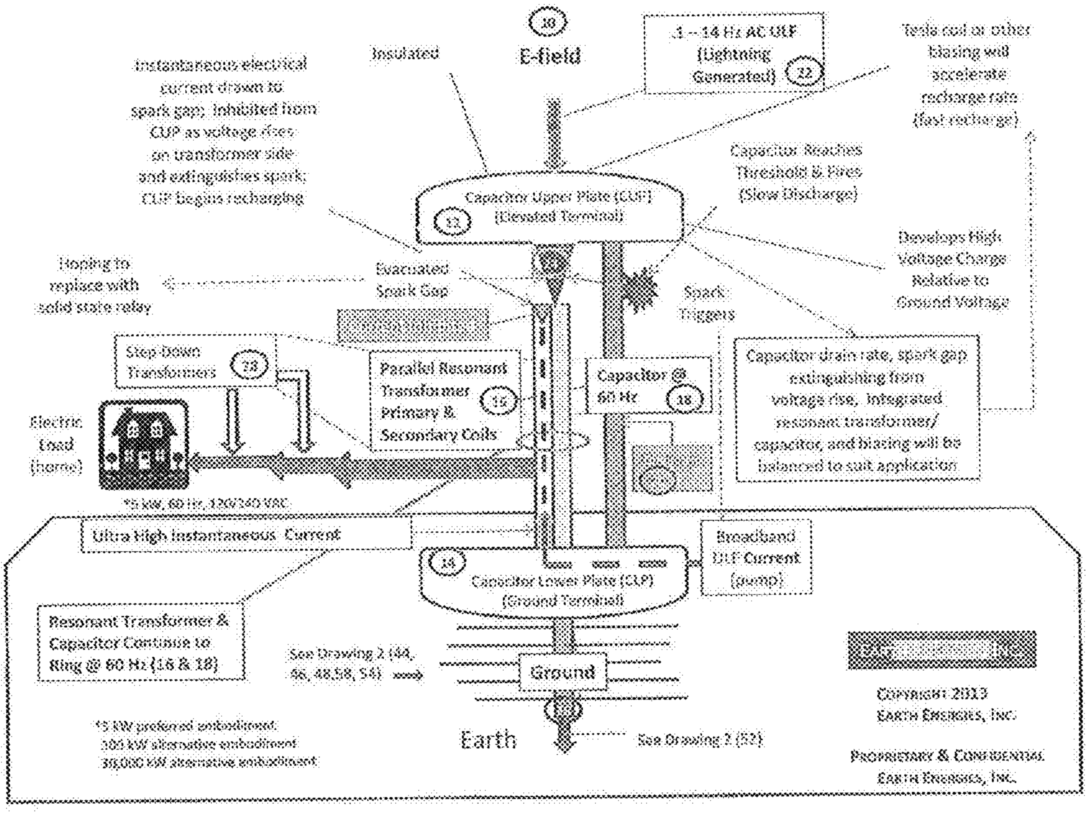
10
E-field
Insulated
Instantaneous electrical current drawn to spark gap; Inhibited from CUP as voltage rises on transformer side and extinguishes spark; CUP begins recharging
.1 – 14 Hz AC ULF (Lightning Generated) 22
Tesla coil or other biasing will accelerate recharge rate (fast recharge)
Capacitor Reaches Threshold & Fires (Slow Discharge)
Capacitor Upper Plate (CUP) (Elevated Terminal)
12
Develops High Voltage Charge Relative to Ground Voltage
Hoping to replace with solid state relay
Evacuated Spark Gap
Spark Triggers
Step-Down Transformers 28
Parallel Resonant Transformer Primary & Secondary Coils 16
Capacitor @ 60 Hz 18
Capacitor drain rate, spark gap extinguishing from voltage rise, integrated resonant transformer/ capacitor, and biasing will be balanced to suit application
Electric Load (home)
*5 kW, 60 Hz, 120/240 VAC
Ultra High Instantaneous Current
14
Capacitor Lower Plate (CLP) (Ground Terminal)
Broadband ULF Current (pump)
Resonant Transformer & Capacitor Continue to Ring @ 60 Hz (16 & 18)
See Drawing 2 (44, 46, 48,58, 54)
Ground
Earth
See Drawing 2 (52)
*5 kW preferred embodiment
100 kW alternative embodiment
35,000 kW alternative embodiment
Copyright 2013
Earth Energies, Inc.
Proprietary & Confidential
Earth Energies, Inc.

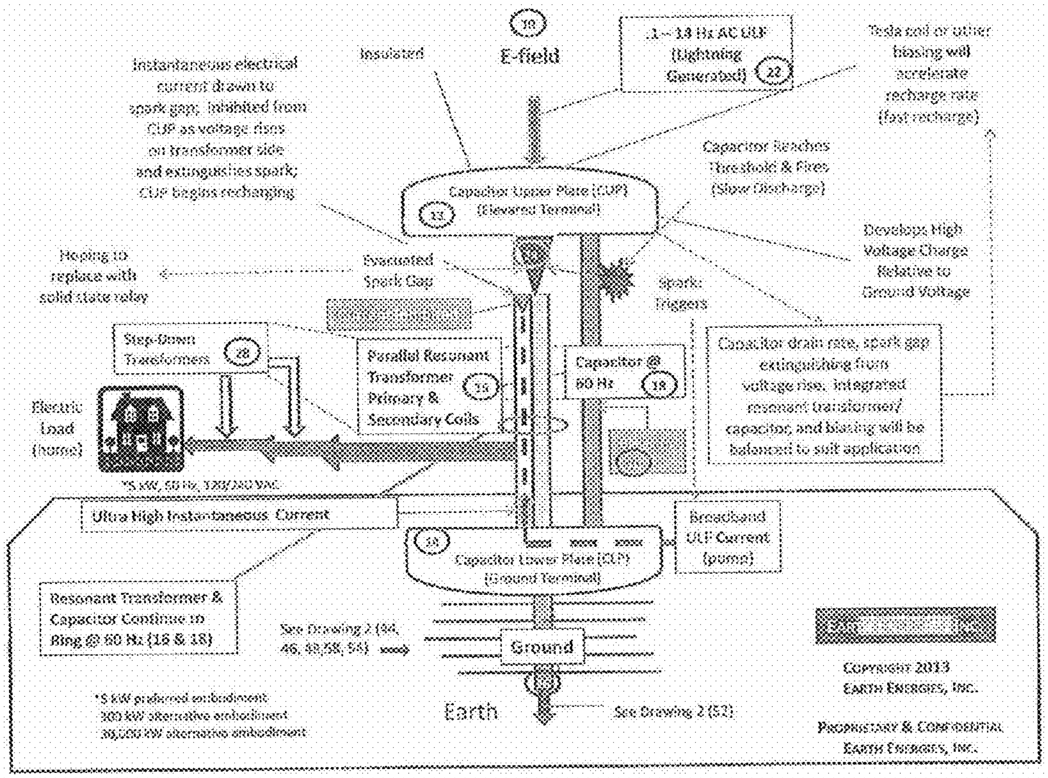
10
E-field
Insulated
.1 – 14 Hz AC ULF (Lightning Generated) 22
Tesla coil or other biasing will accelerate recharge rate (fast recharge)
Instantaneous electrical current drawn to spark gap; inhibited from CUP as voltage rises on transformer side and extinguishes spark; CUP begins recharging
Capacitor Upper Plate (CUP) (Elevated Terminal) 12
Capacitor Reaches Threshold & Fires (Slow Discharge)
Develops High Voltage Charge Relative to Ground Voltage
Hoping to replace with solid state relay
Evacuated Spark Gap
Spark Triggers
Step-Down Transformers 28
Parallel Resonant Transformer Primary & Secondary Coils 15
Capacitor @ 60 Hz 18
Capacitor drain rate, spark gap extinguishing from voltage rise, integrated resonant transformer/ capacitor, and biasing will be balanced to suit application
Electric Load (home)
*5 kW, 60 Hz, 120/240 VAC
Ultra High Instantaneous Current
Broadband ULF Current (pump)
18
Capacitor Lower Plate (CLP) (Ground Terminal)
Resonant Transformer & Capacitor Continue to Ring @ 60 Hz (16 & 18)
See Drawing 2 (44, 46, 48,58, 54)
Ground
Earth
See Drawing 2 (52)
*5 kW preferred embodiment
300 kW alternative embodiment
30,000 kW alternative embodiment
Copyright 2013
Earth Energies, Inc.
Proprietary & Confidential
Earth Energies, Inc.

FIG. 1

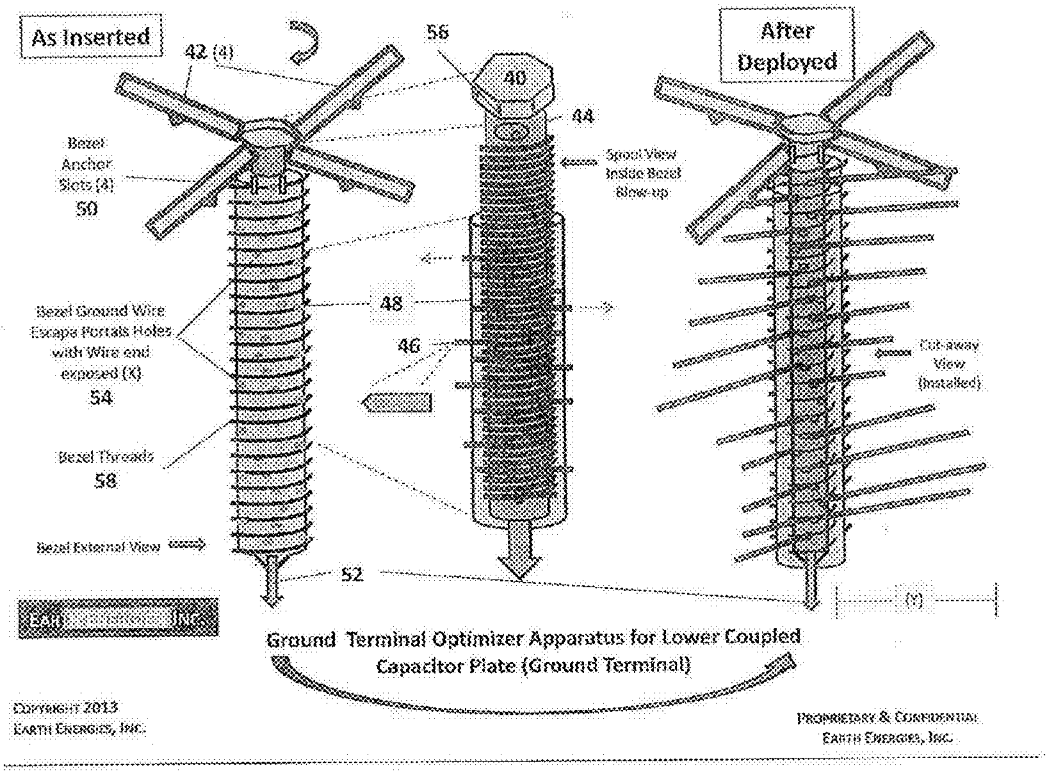
As Inserted
42 (4)
56
40
44
After Deployed
Bezel Anchor Slots (4)
50
Spool View Inside Bezel Blow-up
Bezel Ground Wire Escape Portals Holes with Wire end exposed (X)
54
48
46
Cut-away View (Installed)
Bezel Threads
58
Bezel External View
52
(Y)
Ground Terminal Optimizer Apparatus for Lower Coupled Capacitor Plate (Ground Terminal)
COPYRIGHT 2013
EARTH ENERGIES, INC.
PROPRIETARY & CONFIDENTIAL
EARTH ENERGIES, INC.

FIG. 2

METHOD AND APPARATUS FOR EXTRACTING AND CONVEYING ELECTRICAL ENERGY FROM THE EARTH'S IONOSPHERE CAVITY

[0001] This application claims benefit of U.S. Provisional Application No. 61/889894, filed 11 Oct. 2013, titled, "Method and Apparatus for Extracting and Conveying Electrical Energy From the Earth's Ionosphere Cavity," the disclosure of which is incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION

[0002] A portion of the disclosure of this patent document contains material which is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by any-one of the patent document or the patent disclosure, as it appears in the Patent and Trademark Office patent file or records, but otherwise reserves all copyright whatsoever.

[0003] The invention is an apparatus and methodology for extracting electric power manifesting as a broadband collection of electric field oscillations in the 0 to 200 Hz frequency range within the Earth's ionosphere cavity, through the use of an elevated biased hemispherical or other curved surface capacitor coupling the ionosphere cavity and Earth, and then conveying this electrical energy at the proper frequency, amperage, and voltage to existing homes and buildings where it is delivered for consumption over existing home or building wiring.

DESCRIPTION OF THE RELATED ART

[0004] The current state of the art in this field is focused on the bending of radio frequency waves in the Earth's ionosphere cavity for the purpose of distorting (through reflection) the resident cavity's magnetic field to physically direct the stimulating electrical energy to targeted locations within the Earth's ionosphere cavity and/or coordinates on the Earth's surface at specific energy levels appropriate to specific applications. The regenerative feedback was done by Sutton's active antenna.

[0005] Nikola Tesla attempted to transmit wireless power in the "natural medium"; however, Tesla bad no description of a device to receive natural oscillations as described in the present disclosure.

[0006] There is no known prior art in the field of electric energy extraction from the ionosphere cavity and conveyance to other terminals within the cavity for delivery of usable electric power.

SUMMARY OF THE INVENTION

[0007] In one exemplary embodiment of the present invention, a properly-scaled electric power extraction/conveyance apparatus is installed at a typical consumer residential home to deliver 60 Hz 120/240 VAC electric power, sufficient to convey the equivalent of 5 kW of continuous electricity, to the home's existing electric service entry component to power electrical appliances in the home over existing home electric wiring.

[0008] In another exemplary embodiment of the present invention, a properly-scaled electric power extraction/conveyance apparatus is installed at an appropriate location within an industrial park or neighborhood to deliver 60 Hz 120/240 V AC electric power, sufficient to convey the equivalent of 300 kW of continuous electricity, to the service entry components of industrial buildings or neighborhood homes to power electrical devices and appliances over existing building or home electric wiring.

[0009] Alternative exemplary embodiments of the present invention will extract direct current electric power from the Earth ionosphere cavity, generated by the earth's rotating magnetosphere, and convey in the form of conditioned direct current to direct current load sources, or convert to alternating current and convey in the form of conditioned alternating current to alternating current load sources.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] A disclosure embodiments of the present invention, is set forth more particularly in the remainder of the specification, which makes reference to the accompanying figures, in which:

[0011] FIG. **1** provides an example of an electric power extraction and conveyance apparatus for one residential-scale embodiment of the present invention. Certain additional example embodiments of the present invention utilize similar fundamental technologies, but differ in the scale (size) of the components and characteristics of the electricity conveyed for consumption; and

[0012] FIG. **2** is a schematic illustration of ground terminals for use with embodiments of the present invention.

DETAILED DESCRIPTION

[0013] Applicant has discovered that trees extract and use electric power from the Earth ionosphere cavity. The present invention emulates the architecture of electric power extraction and usage methodology inherent in the natural operation of living trees, but uses standard and customized components, common in the field, for its implementation. The natural electric power extraction and use of electric power by trees leverages the shape and physical composition of trees as instances of large electrical capacitors coupled to the ionosphere cavity. The canopy of trees (upper limbs and leaf structures) emulates the upper plates of a coupled capacitor, the trunks of trees emulate a combined dielectric and resonant transformer/capacitor, and the root systems of trees emulate ground terminals (lower coupled capacitor plates). The extraction of electric power from the ionosphere cavity by trees is in the form of ultra-low frequency broadband waves (oscillations) in the frequency range of 0.1 to 14 Hertz (ULF). Inherent tree physical composition and control structure (i.e., DNA/RNA) govern the use of the extracted electric power to accomplish tree growth, seasonal metabolic adjustments, and reproductive processes. Any excess extracted electric power by trees is returned to the ionosphere cavity. The present invention is an exemplary instance of this same natural architecture using standard and customized components that facilitate extraction and conveyance of ionosphere cavity-resident broadband electric power at useful scales sufficient to provide continuous electrical power for electrical devices used by people.

[0014] One preferred exemplary embodiment of the present invention is in the form shown in Drawing **1** above and will deliver continuous 60 Hz 120/240 V AC electric power to consumer residences (and other structures with similar electricity conditioning and consumption requirements) equivalent to electric power capacity of 5 kilowatts.

[0015] A secondary exemplary embodiment of the present invention is also in the form shown in FIG. **1**, appropriately scaled for larger electrical power demands and related conveyance capacity, and will deliver continuous 60 Hz 120/240 V AC electric power to larger buildings or electrical distribution facilities for commercial office parks with electric power capacity of 300 kilowatts.

[0016] A third exemplary embodiment of the present invention is also in the form shown in FIG. **1**, appropriately scaled for larger electrical power capacity, and will deliver continuous 60 Hz 120/240 VAC electric power to utility substations with electric power capacity of 30,000 kilowatts (or higher).

[0017] The following system, power sources, and component descriptions provide additional details of the present invention in various embodiments.

[0018] Earth's ionosphere cavity (**10**) is a reservoir of continuously-generated electrical energy (E-Field) in both direct current (DC) and alternating current (AC) realms. The DC electric energy is created by the constant rotation of the Earth's magnetosphere exposed to the solar wind. Earth's perpetually moving magnetic fields create the DC electric power in the ionosphere cavity E-Field per the equations in James Clerk Maxwell's fundamental theory of electromagnetism. AC electrical energy in the ionosphere cavity is generated from the persistent E-Field disturbances (i. e., oscillations) caused by lightning discharges. The present invention includes the direct extraction of AC electric power from the ionosphere cavity, and the direct extraction of DC electric power from the ionosphere cavity which is converted to AC electric power before conveyance. Optional operational configurations of the present invention will also enable conveyance of extracted DC electric power for potential scenarios where the conveyance of conditioned DC electricity is required.

[0019] Electric-field oscillations (**22**) are induced by lightning and encapsulated in the earth-ionosphere cavity, creating a continuous source of renewable electrical energy. The frequency of lightning occurrences in the ionosphere cavity averages thirty (30) strikes per second, with each strike representing approximately one (1) terawatt (TW) of instantaneous energy. Assuming an average energy dissipation rate for each lightning strike of fifteen (15) microseconds (μsec), i. e., the time involved for one (1) terawatt of instantaneous energy to dissipate to zero (0) energy, these lightning-induced electric field oscillations are sufficient to provide 4,500 times the entire electric energy consumption on planet Earth at the projected consumption rate for the year 2100 (55.3 TWh annually). Lightning-induced E-Field oscillations vary between 0.1 and 14 Hertz with wavelengths exceeding the practicality of an electromagnetic antenna for collecting the energy. The present invention utilizes capacitive coupling to overcome this obstacle.

[0020] An insulated upper capacitor plate (terminal) (12) elevated above the influence of the ground floor collects E-Field oscillations from the ionosphere cavity. This insulated elevated terminal, capacitively-coupled to the E-Field bounded by the earth-ionosphere cavity, is arranged in a surface of large radii of curvature, supporting maximum surface connection to the E-Field without leaking the high voltage charge applied to the elevated terminal. The capacitance and resistance of the elevated terminal are responsive to the reception of broadband electric-field frequencies in the 0 to 200 hertz range. The elevated terminal is subject to high voltage alternating current biasing used to collect electric-field oscillations generated by lightning impulses in the earth-ionosphere cavity.

[0021] An evacuated spark gap (**24**) connected to the elevated terminal (**12**) prevents electrical discharge occurring between the elevated terminal and the ionosphere cavity. As the high voltage alternating current approaches peak voltage the spark gap energizes and a large voltage is applied to the step-down transformer primary coil (**16**). The high voltage pumps high current from the ground terminal (**14**, **26**) converting the electric-field to a magnetic field within the transformer (**16**). The step-down transformer is connected to a 60 hertz resonant capacitor (**18**), the secondary side of the transformer resonating at 60 hertz will supply a 60 hertz 120/240 VAC power to the load. Filter circuits will condition the power. The evacuated spark gap function may be accomplished by solid state circuitry in the implementation of the present invention.

[0022] The resonant transformer (**16**) is integrated with a parallel resonant capacitor across the transformer where the integrated combination resonates at 60 hertz supplying power to the electricity consuming structure (home or business). Various step-down transformers (**28**) are utilized to condition the conveyed electricity to accommodate the requirements of the receiving station for various embodiments of the present invention.

[0023] Ground (**26**) is the source of electric current for the present invention. A ground terminal, connected to the soil, is arranged in a surface of large radii of curvature to permit instantaneous sourcing of high currents from earth ground without temperature and voltage rise inhibiting the ground terminal collection of current. The capacitance and resistance of the ground terminal to earth is minimized to promote on demand current sourcing. Drawing **2** illustrates an exemplary apparatus of the present invention that optimizes ground connection quality by maximizing the penetration extent and resulting surface area of the ground terminal's extension into the soil.

[0024] Step-down transformers (**28**) are scaled to transform the electrical current and voltage into structured and conditioned electricity that is compatible with the electric power needs of the load sources for various embodiments of the present invention.

[0025] A mechanical nut (**40**) solidly attached to shaft **44** to facilitate extension, when turned, of large gauge copper wire extrusions (**46**).

[0026] A downward pointed anchor flat shaft (**42**) attached to the bezel **50** via a T-slot in bezel.

[0027] A copper (or other highly-conductive metal) spool base (**44**) sourcing "X" independently attached large gauge pointed copper wires (**46**) of length "Y", where "X" is equal to the number of escape portals in bezel (**50**), and "Y" is the length determined to yield optimum surface area of wires of type **46** for high quality connectivity to soil. The spool can freely rotate inside the bezel (**48**) and emulates the primary root structure of a living tree.

[0028] Sharpened large gauge (reinforced) copper wires (**46**), quantity "X", of length "Y" emulating the root structure of living trees.

[0029] Threaded bezel (spool-housing sleeve) (**48**) with drilled escape portal holes with end of copper wires (**46**) protruding through holes ½ inch.

[0030] Bezel anchor slots (**50**) of twice the height of anchor flat shafts (**42**) and T-slotted to accommodate anchor flat shafts.

[0031] Extended pointed end (**52**) of inner spool base shaft.

[0032] Escape portals (**54**) in bezel to allow for large gauge copper wire (**46**) extension into soil.

[0033] Connector port (**56**) for connection to ground terminal (**14**) in Drawing **1**.

[0034] Bezel threads (**58**).

[0035] Various embodiments of the present invention will include one or more of the following characteristics in various combinations. These descriptions are provided for purposes of example. Not all of the following characteristics will be required for any given exemplary embodiment.

[0036] Characteristic 1, The form (shape), physical composition, size, insulation, location altitude, discharge capacitive threshold and discharge rate, and the strength of AC or DC biasing applied to the (CUP) upper coupled capacitor plate (**12**: elevated terminal) all form an integrated whole, the purpose of which is to extract and temporarily store DC or AC electric power from the Earth's ionosphere cavity (**10**) at high voltage, and to trigger the energizing of the evacuated spark gap (**24**).

[0037] Characteristic 2, The elevated terminal (**12**) is capacitively coupled to the E-Field (**10**) bounded by the earth ionosphere cavity, and arranged in a surface of large radii of curvature, supporting maximum surface connection to the E-Field without leaking the high voltage charge applied to the elevated terminal (**12**).

[0038] Characteristic 3, The capacitance and resistance of the elevated terminal (**12**) are responsive to the reception of broadband electric-field frequencies in the 0 to 200 hertz range.

[0039] Characteristic 4, The elevated terminal (**12**) is subject to high voltage alternating current biasing used to collect E-Field oscillations generated by lightning impulses in the earth-ionosphere cavity.

[0040] Characteristic 5, An alternative instantiation of elevated terminal (**12**) is subject to high voltage direct current biasing used to collect E-Field direct current generated by the Earth's rotating magnetosphere.

[0041] Characteristic 6, The evacuated spark gap (**24**) may be in the form of a cohesive physical spark ignition or an operational emulation of this function through solid state circuitry.

[0042] Characteristic 7, The evacuated spark gap (**24**) connected to the elevated terminal (**12**) prevents electrical discharge occurring between the elevated terminal and the ionosphere cavity.

[0043] Characteristic 8, The high voltage alternating current from the E-Field, induced into the elevated terminal (**12**), approaches peak voltage and causes the evacuated spark gap (**24**) to energize and a large voltage is applied to the step-down transformer primary coil (**16**).

[0044] Characteristic 9, The high voltage alternating current impulse from the elevated terminal (**12**), which causes the evacuated spark gap (**24**) to energize, manifests as alternating current on broadband frequencies between 0.1 and 200 Hz.

[0045] Characteristic 10, In an alternative instantiation of the elevated terminal (**12**) the stored high voltage direct current, which causes the evacuated spark gap (**24**) to energize, manifests as direct current on broadband frequencies between 0 and 200 Hz.

[0046] Characteristic 11, The high voltage applied to the step-down transformer primary coil (**16**) pumps high current from the ground terminal (**14**), converting the electric-field to a magnetic field within the transformer (**16**).

[0047] Characteristic 12, The high current pumped from the ground terminal (**14**), caused by high voltage applied to the step-down transformer primary coil (**16**), will attempt to normalize the field and extinguish the evacuated spark gap (**24**).

[0048] Characteristic 13, The transformer's (**16**) magnetic field expands in response to the high current electric-field inrush from the ground terminal (**14**).

[0049] Characteristic 14, The primary winding inside the step-down transformer (**16**) will add the energy from each broadband frequency together into a single magnetic field, where it will become a 60 Hz frequency at the secondary winding of the transformer (**16**).

[0050] Characteristic 15, The step-down transformer (**16**) is connected to a 60 Hertz resonant capacitor (**18**).

[0051] Characteristic 16, The secondary (coil) side of the step-down transformer (**16**), resonating at 60 hertz, will supply a 60 hertz 120/240 VAC power to the electric load.

[0052] Characteristic 17, Filter circuits within the step-down transformer will condition the power according to the electric power needs of the electric load.

[0053] Characteristic 18, Electric-field lines exist (within the Earth's ionosphere cavity) that are horizontal to the surface of the Earth and extend to heights above 30,000 feet, with voltage gradients typically 100 Volts per vertical meter.

[0054] Characteristic 19, Lightning impulses create global broadband oscillations in the Earth ionosphere cavity electric-field between 0.1 and 14 Hertz.

[0055] Characteristic 20, Rotation of the Earth's magnetosphere creates global electric field oscillations in the Earth ionosphere cavity that manifest as electromagnetic waves in the 0 to 200 Hertz frequency range.

[0056] Characteristic 21, Global broadband oscillations in the Earth ionosphere cavity electric-field can be used for wireless energy transmission without harm to people, plants, or trees.

[0057] Characteristic 22, A maximum quality ground connection is achieved by expanding the surface area of the ground connection (**48** and **44**) through extension of high-conductor large gauge copper wires (**46**) horizontally through bezel (**48**) escape portals (**54**) along depth of source spool (**44**) and bezel (**48**).

[0058] Characteristic 23, The technique described in Characteristic 22 effectively expands the diameter of the ground source spool (**44**) with regard to quality of ground connection.

[0059] Characteristic 24, Bezel threads (**58**) expand surface area of bezel (**48**) touching earth, enhancing ground connection quality.

[0060] Characteristic 25, Bezel threads (**58**) enhance ground connection quality of bezel (**48**), source spool (**44**), and high conductor large gauge copper wires (**46**) through rotational insertion and resulting friction hold v. other (direct drive) insertion methods which can result in corruption of ground hole integrity caused by imprecise angle and/or vibration during insertion.

[0061] While one or more embodiments of the present invention have been described above, it should be understood that any and all equivalent realizations of the present invention are included within the scope and spirit thereof. Thus, the embodiments presented herein are by way of example only and are not intended as limitations of the present invention. Therefore, it is contemplated that any and all such embodiments are included in the present invention.

What is claimed is:

1. A power receiver for extracting electrical energy from the earth's electric field, said power receiver comprising:

a resonant transformer connected between a ground terminal disposed below the surface of the earth and an elevated terminal;

impulse generator for generating and applying a high voltage electrical impulse to a primary winding of the resonant transformer to induce current flow from the ground terminal through the primary winding of the transformer; and

a power conversion circuit connected to a secondary winding of the resonant transformer to convert electrical current flowing through the secondary winding to a desired form.

2. The power receiver of claim **1** wherein a resonant frequency of the resonant transformer is below 200 Hz.

3. The power receiver of claim **1** wherein the elevated terminal comprises an upper capacitive plate coupled to the earth's ionosphere cavity.

4. The power receiver of claim **3** wherein the impulse generator comprises:

the upper capacitive plate;

a pair of electrodes separated by a spark gap, said electrodes connected between the upper capacitive plate and the primary winding of the resonant transformer and configured to generate a spark when a voltage difference between the electrodes reaches a predetermined level.

5. The power receiver of claim **1** further comprising a lower capacitive plate coupled to the ground terminal.

6. The power receiver of claim **1** wherein the resonant transformer comprises a capacitor connected in parallel with a primary winding of the resonant transformer.

* * * * *

455-619 AU 233 EX
FIP8106 OR 685,958

55

No. 685,958. **Patented Nov. 5, 1901.**

N. TESLA.

METHOD OF UTILIZING RADIANT ENERGY.

(Application filed Mar. 21, 1901.)

(No Model.)

Electric steping motor
energized by corpuscular
energy from sun

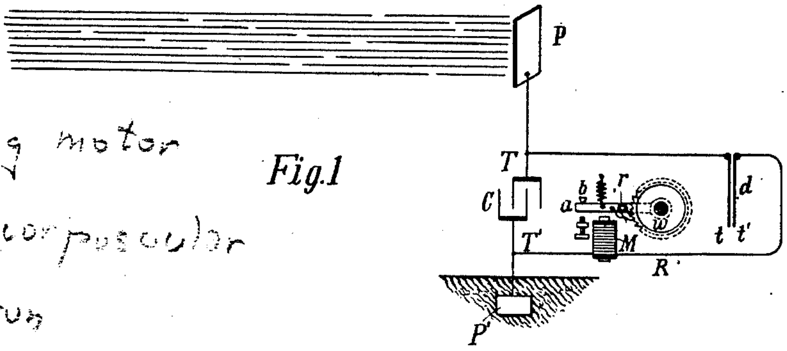

Fig.1

No refs.

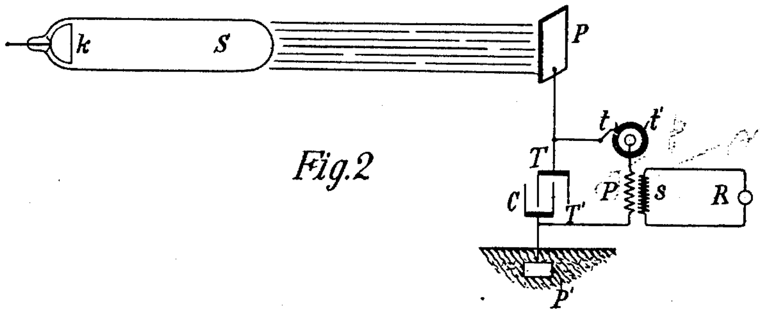

Fig.2

photo pickup

Witnesses:
Raphaël Netter
M. Lawson Dyer

Nikola Tesla, Inventor
by Kerr, Page & Cooper Attys

THE NORRIS PETERS CO., PHOTO-LITHO., WASHINGTON, D. C.

UNITED STATES PATENT OFFICE.

NIKOLA TESLA, OF NEW YORK, N. Y.

METHOD OF UTILIZING RADIANT ENERGY.

SPECIFICATION forming part of Letters Patent No. 685,958, dated November 5, 1901.

Application filed March 21, 1901. Serial No. 52,154. (No model.)

To all whom it may concern:

Be it known that I, NIKOLA TESLA, a citizen of the United States, residing at the borough of Manhattan, in the city, county, and State of New York, have invented certain new and useful Improvements in Methods of Utilizing Radiant Energy, of which the following is a specification, reference being had to the drawings accompanying and forming a part of the same.

It is well known that certain radiations—such as those of ultra-violet light, cathodic, Roentgen rays, or the like—possess the property of charging and discharging conductors of electricity, the discharge being particularly noticeable when the conductor upon which the rays impinge is negatively electrified. These radiations are generally considered to be ether vibrations of extremely small wave lengths, and in explanation of the phenomena noted it has been assumed by some authorities that they ionize or render conducting the atmosphere through which they are propagated. My own experiments and observations, however, lead me to conclusions more in accord with the theory heretofore advanced by me that sources of such radiant energy throw off with great velocity minute particles of matter which are strongly electrified, and therefore capable of charging an electrical conductor, or even if not so may at any rate discharge an electrified conductor either by carrying off bodily its charge or otherwise.

My present application is based upon a discovery which I have made that when rays or radiations of the above kind are permitted to fall upon an insulated conducting body connected to one of the terminals of a condenser, while the other terminal of the same is made by independent means to receive or to carry away electricity, a current flows into the condenser so long as the insulated body is exposed to the rays, and under the conditions hereinafter specified an indefinite accumulation of electrical energy in the condenser takes place. This energy after a suitable time interval, during which the rays are allowed to act, may manifest itself in a powerful discharge, which may be utilized for the operation or control of mechanical or electrical devices or rendered useful in many other ways.

In applying my discovery I provide a condenser, preferably of considerable electrostatic capacity, and connect one of its terminals to an insulated metal plate or other conducting body exposed to the rays or streams of radiant matter. It is very important, particularly in view of the fact that electrical energy is generally supplied at a very slow rate to the condenser, to construct the same with the greatest care. I use by preference the best quality of mica as dielectric, taking every possible precaution in insulating the armatures, so that the instrument may withstand great electrical pressures without leaking and may leave no perceptible electrification when discharging instantaneously. In practice I have found that the best results are obtained with condensers treated in the manner described in a patent granted to me February 23, 1897, No. 577,671. Obviously the above precautions should be the more rigorously observed the slower the rate of charging and the smaller the time interval during which the energy is allowed to accumulate in the condenser. The insulated plate or conducting body should present as large a surface as practicable to the rays or streams of matter, I having ascertained that the amount of energy conveyed to it per unit of time is under otherwise identical conditions proportionate to the area exposed, or nearly so. Furthermore, the surface should be clean and preferably highly polished or amalgamated. The second terminal or armature of the condenser may be connected to one of the poles of a battery or other source of electricity or to any conducting body or object whatever of such properties or so conditioned that by its means electricity of the required sign will be supplied to the terminal. A simple way of supplying positive or negative electricity to the terminal is to connect the same either to an insulated conductor, supported at some height in the atmosphere, or to a grounded conductor, the former, as is well known, furnishing positive and the latter negative electricity. As the rays or supposed streams of matter generally convey a positive charge to the first condenser-terminal, which is connect-

RADIANT ENERGY.

EXAMINER

ed to the plate or conductor above mentioned, I usually connect the second terminal of the condenser to the ground, this being the most convenient way of obtaining negative electricity, dispensing with the necessity of providing an artificial source. In order to utilize for any useful purpose the energy accumulated in the condenser, I furthermore connect to the terminals of the same a circuit including an instrument or apparatus which it is desired to operate and another instrument or device for alternately closing and opening the circuit. This latter may be any form of circuit-controller, with fixed or movable parts or electrodes, which may be actuated either by the stored energy or by independent means.

The rays or radiations which are to be utilized for the operation of the apparatus above described in general terms may be derived from a natural source, as the sun, or may be artificially produced by such means, for example, as an arc-lamp, a Roentgen tube, and the like, and they may be employed for a great variety of useful purposes.

My discovery will be more fully understood from the following detailed description and annexed drawings, to which reference is now made, and in which—

Figure 1 is a diagram showing typical forms of the devices or elements as arranged and connected in applying the method for the operation of a mechanical contrivance or instrument solely by the energy stored; and Fig. 2 is a diagrammatical representation of a modified arrangement suitable for special purposes, with a circuit-controller actuated by independent means.

Referring to Fig. 1, C is the condenser, P the insulated plate or conducting body, which is exposed to the rays, and P′ another plate or conductor, all being joined in series, as shown. The terminals T T′ of the condenser are also connected to a circuit including a receiver R, which is to be operated, and a circuit-controlling device *d*, which in this case is composed of two very thin conducting-plates *t t′*, placed in close proximity and very mobile, either by reason of extreme flexibility or owing to the charater of their support. To improve their action, they should be inclosed in a receptacle from which the air may be exhausted. The receiver R is shown as consisting of an electromagnet M, a movable armature *a*, a retractile spring *b*, and a ratchet-wheel *w*, provided with a spring-pawl *r*, which is pivoted to armature *a*, as illustrated. The apparatus being arranged as shown, it will be found that when the radiations of the sun or of any other source capable of producing the effects before described fall upon the plate P an accumulation of electrical energy in the condenser C will result. This phenomenon, I believe, is best explained as follows: The sun as well as other sources of radiant energy throw off minute particles of matter positively electrified, which, impinging upon the plate P, communicate an electrical charge to the same. The opposite terminal of the condenser being connected to the ground, which may be considered as a vast reservoir of negative electricity, a feeble current flows continuously into the condenser, and inasmuch as these supposed particles are of an inconceivably small radius or curvature, and consequently charged to a relatively very high potential, this charging of the condenser may continue, as I have found in practice, almost indefinitely, even to the point of rupturing the dielectric. Obviously whatever circuit-controller be employed it should operate to close the circuit in which it is included when the potential in the condenser has reached the desired magnitude. Thus in Fig. 2 when the electrical pressure at the terminals T T′ rises to a certain predetermined value the plates *t t′*, attracting each other, close the circuit connected to the terminals. This permits a flow of current which energizes the magnet M, causing it to draw down the armature *a* and impart a partial rotation to the ratchet-wheel *w*. As the current ceases the armature is retracted by the spring *b* without, however, moving the wheel *w*. With the stoppage of the current the plates *t t′* cease to be attracted and separate, thus restoring the circuit to its original condition.

Many useful applications of this method of utilizing the radiations emanating from the sun or other source and many ways of carrying out the same will at once suggest themselves from the above description. By way of illustration a modified arrangement is shown in Fig. 2, in which the source S of radiant energy is a special form of Roentgen tube devised by me having but one terminal *k*, generally of aluminium, in the form of half a sphere with a plain polished surface on the front side, from which the streams are thrown off. It may be excited by attaching it to one of the terminals of any generator of sufficiently-high electromotive force; but whatever apparatus be used it is important that the tube be exhausted to a high degree, as otherwise it might prove entirely ineffective. The working or discharge circuit connected to the terminals T T′ of the condenser includes in this case the primary *p* of a transformer and a circuit-controller comprising a fixed terminal or brush *t* and a movable terminal *t′* in the shape of a wheel with conducting and insulating segments which may be rotated at an arbitrary speed by any suitable means. In inductive relation to the primary wire or coil *p* is a secondary *s*, usually of a much greater number of turns, to the ends of which is connected a receiver R. The terminals of the condenser being connected as indicated, one to an insulated plate P and the other to a grounded plate P′, when the tube S is excited rays or streams of matter are emitted from the same, which convey a positive charge to the plate P and condenser-terminal T, while terminal T′ is continuously receiving negative electricity from the plate

P'. This, as before explained, results in an accumulation of electrical energy in the condenser, which goes on as long as the circuit including the primary p is interrupted. Whenever the circuit is closed, owing to the rotation of the terminal t', the stored energy is discharged through the primary p, this giving rise in the secondary s to induced currents which operate the receiver R.

It is clear from what has been stated above that if the terminal T' is connected to a plate supplying positive instead of negative electricity the rays should convey negative electricity to plate P. The source S may be any form of Roentgen or Lenard tube; but it is obvious from the theory of action that in order to be very effective the electrical impulses exciting it should be wholly or at least preponderatingly of one sign. If ordinary symmetrical alternating currents are employed, provision should be made for allowing the rays to fall upon the plate P only during those periods when they are productive of the desired result. Evidently if the radiations of the source be stopped or intercepted or their intensity varied in any manner, as by periodically interrupting or rythmically varying the current exciting the source, there will be corresponding changes in the action upon the receiver R, and thus signals may be transmitted and many other useful effects produced. Furthermore, it will be understood that any form of circuit-closer which will respond to or be set in operation when a predetermined amount of energy is stored in the condenser may be used in lieu of the device specifically described with reference to Fig. 1, and also that the special details of construction and arrangement of the several parts of the apparatus may be very greatly varied without departure from the invention.

Having described my invention, what I claim is—

1. The method of utilizing radiant energy, which consists in charging one of the armatures of a condenser by rays or radiations, and the other armature by independent means, and discharging the condenser through a suitable receiver, as set forth.

2. The method of utilizing radiant energy, which consists in simultaneously charging a condenser by means of rays or radiations and an independent source of electrical energy, and discharging the condenser through a suitable receiver, as set forth.

3. The method of utilizing radiant energy, which consists in charging one of the armatures of a condenser by rays or radiations, and the other by independent means, controlling the action or effect of said rays or radiations and discharging the condenser through a suitable receiver, as set forth.

4. The method of utilizing radiant energy, which consists in charging one of the armatures of a condenser by rays or radiations and the other by independent means, varying the intensity of the said rays or radiations and periodically discharging the condenser through a suitable receiver, as set forth.

5. The method of utilizing radiant energy, which consists in directing upon an elevated conductor, connected to one of the armatures of a condenser, rays or radiations capable of positively electrifying the same, carrying off electricity from the other armature by connecting the same with the ground, and discharging the accumulated energy through a suitable receiver, as set forth.

6. The method of utilizing radiant energy, which consists in charging one of the armatures of a condenser by rays or radiations, and the other by independent means, and effecting by the automatic discharge of the accumulated energy the operation or control of a suitable receiver, as set forth.

NIKOLA TESLA.

Witnesses:
M. LAWSON DYER,
RICHARD DONOVAN.

Russian Military Outposts

"The Russian Woodpecker"
Duga Radar Array, Chenobyl, Ukraine

51°18'20.17"N, 30°04'02.60"E

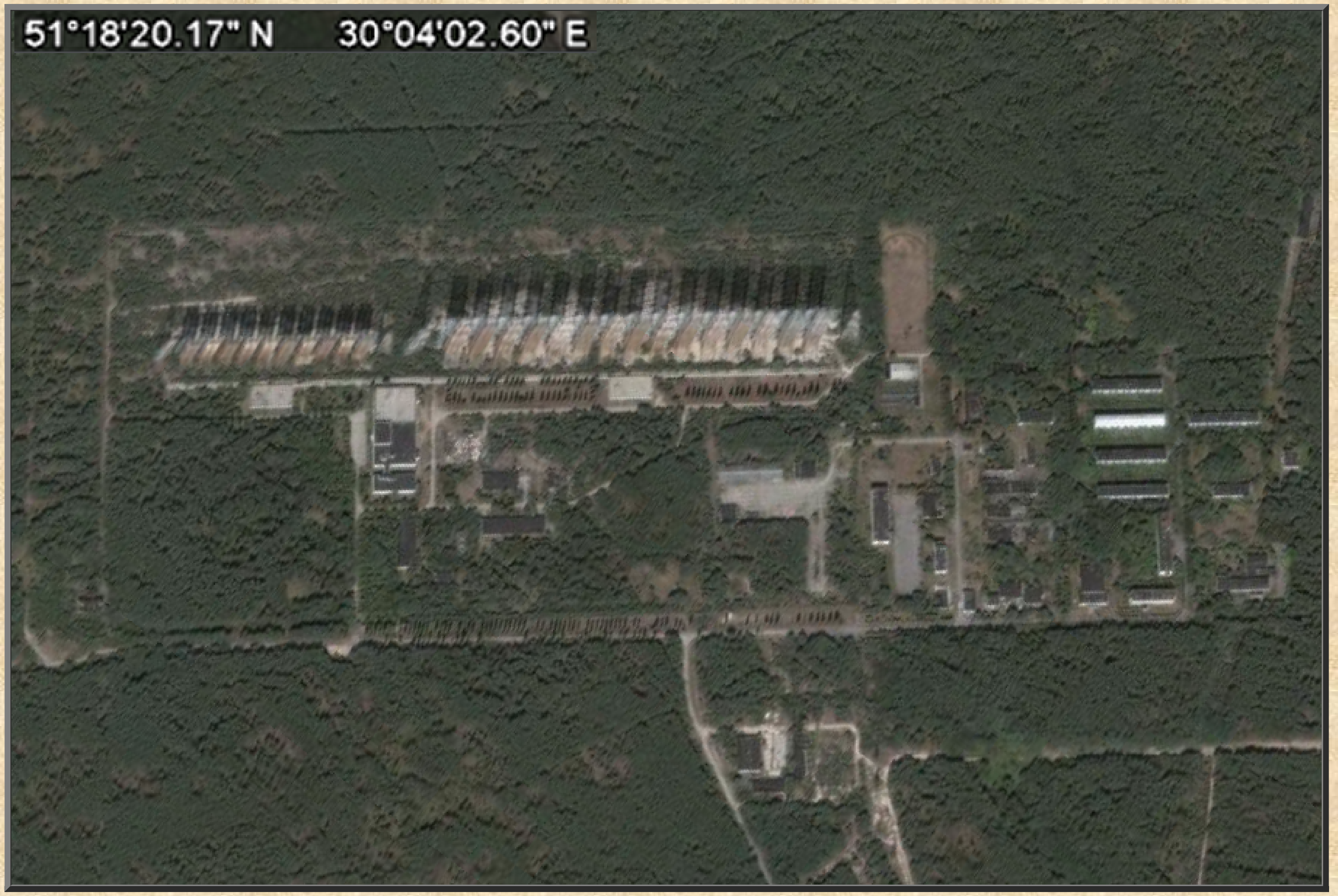

"Woodpecker" Duga radar array, Chenobyl, Ukraine

The Russian Woodpecker

Duga-3 array outside Chernobyl. The array of pairs of cylindrical/conical cages on the right are the driven elements, fed at the facing points with a form of ladder line suspended from stand-off platforms at top right. A back plane reflector of small wires can just be seen left of center, most clearly at the bottom of the image.

The Russian Woodpecker was a notorious Soviet signal that could be heard on the shortwave radio bands worldwide between July 1976 and December 1989. It sounded like a sharp, repetitive tapping noise, at 10 Hz, giving rise to the "Woodpecker" name. The random frequency hops disrupted legitimate broadcast, amateur radio, and utility transmissions and resulted in thousands of complaints by many countries worldwide.

The signal was long believed to be that of an over-the-horizon radar (OTH) system. This theory was publicly confirmed after the fall of the Soviet Union, and is now known to be the Duga-3[1] system, part of the Soviet ABM early-warning network. This was something that NATO military intelligence was well aware of all along, having photographed it and giving it the NATO reporting name Steel Yard.

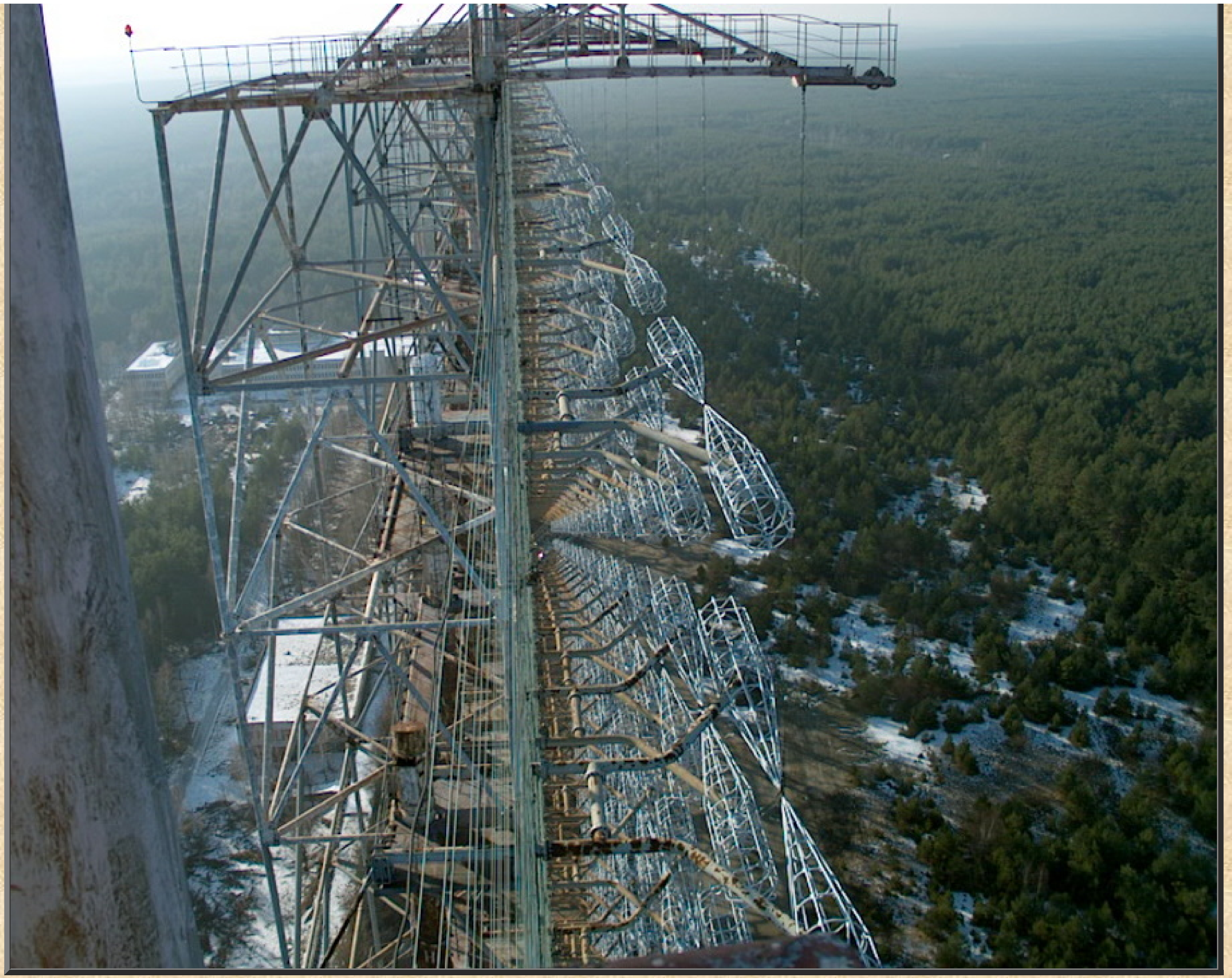

"Woodpecker" Duga radar array, Chenobyl, Ukraine by Necator

Woodpecker - Woodpecker on shortwave radio interfering with WWVH, November 2, 1984.

History

The Soviets had been working on early warning radars for their anti-ballistic missile systems through the 1960s, but most of these had been line-of-sight systems that were useful for raid analysis and interception only. None of these systems had the capability to provide early-warning of a launch, which would give the defenses time to study the attack and plan a response. At the time the Soviet early-warning satellite network was not well developed, so work started on over-the-horizon radar systems for this associated role in the late 1960s.

The first experimental system, Duga-1, was built outside Mykolaiv in the Ukraine, successfully detecting rocket launches from Baikonur Cosmodrome at 2,500 kilometers. This was followed by the prototype Duga-2, built on the same site, which was able to track launches from the far east and submarines in the Pacific Ocean as the missiles flew towards Novaya Zemlya. Both of these radars were aimed east and were fairly low power, but with the concept proven work began on an operational system. The new Duga-3 systems used a transmitter and receiver separated by about 60 km.[2]

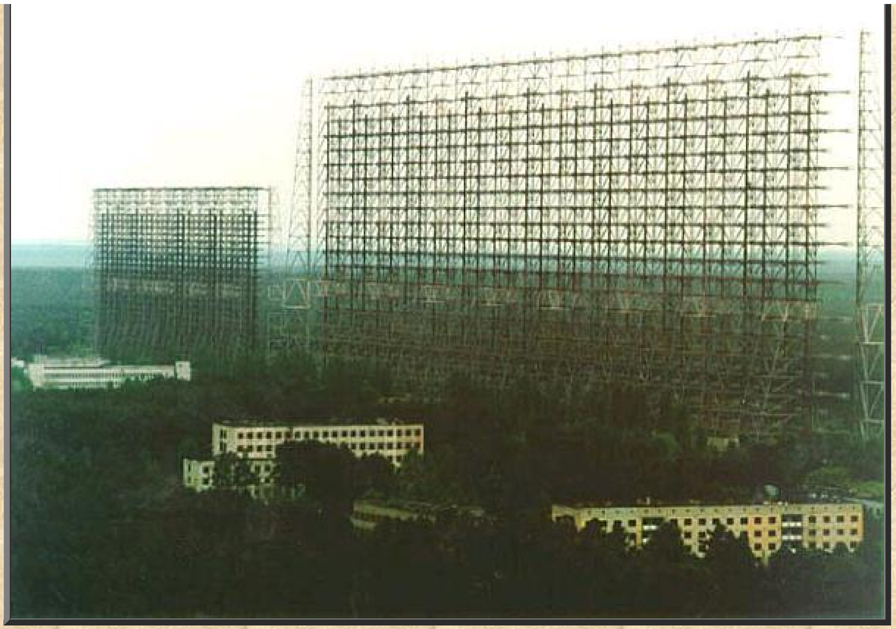

РЛС Дуга-1 Чернобыль-2 (Radar arch-1 Chernobyl-2) by DM.HANTER

Appearance

Starting in 1976 a new and powerful radio signal was detected worldwide, and quickly dubbed the Woodpecker by amateur radio operators. Transmission power on some woodpecker transmitters was estimated to be as high as 10 MW EIRP. As well as disrupting shortwave amateur radio and broadcasting it could sometimes be heard over telephone circuits due to the strength of the signals. This led to a thriving industry of "Woodpecker filters" and noise blankers.

One idea amateur radio operators used to combat this interference was to attempt to "jam" the signal by transmitting synchronized unmodulated continuous wave signals, at the same pulse rate as the offending signal. This idea was considered, but abandoned as impractical. Simple CW pulses didn't appear to have any effect. However, playing back recordings of the woodpecker transmissions sometimes caused the woodpecker transmissions to shift frequency leading to speculation that the receiving stations were able to differentiate between the "signature" waveform of the woodpecker transmissions and a simple pulsed carrier.[4]

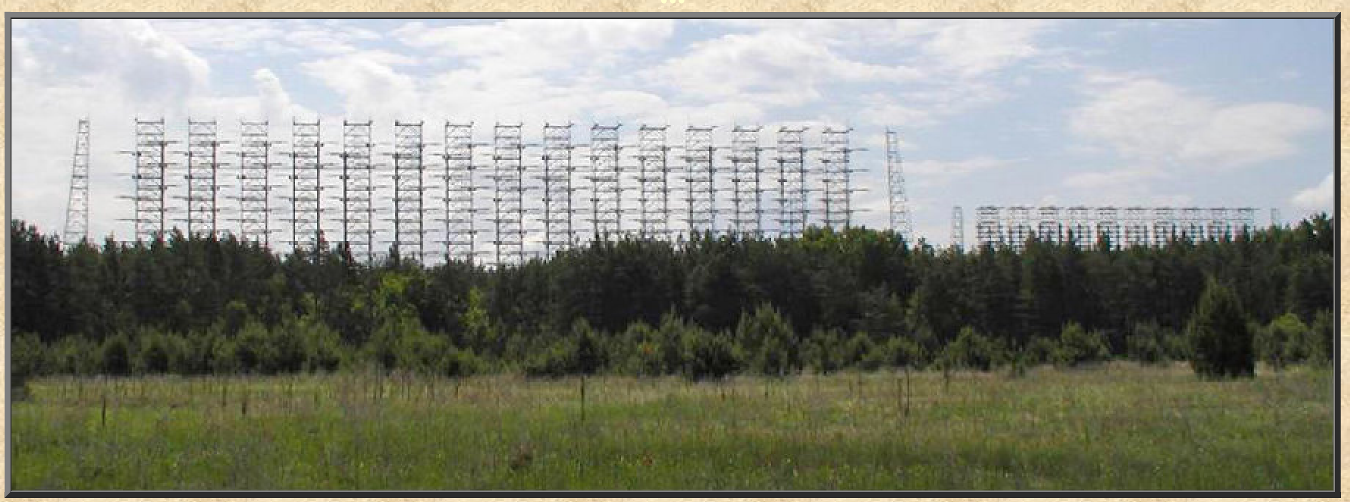

Чорнобиль-2. Приймальні антени. (Chernobyl-2. RX antennas.) by djcrok

Identification

Triangulation quickly revealed the signals came from Ukraine. Confusion due to small differences in the reports being made from various military sources led to the site being alternatively located near Kiev, Minsk, Chernobyl, Gomel or Chernihiv. All of these reports were describing the same deployment, with the transmitter only a few kilometers southwest of Chernobyl (south of Minsk, northwest of Kiev) and the receiver about 50 km northwest of Chernobyl (just west of Chernihiv, south of Gomel). Unknown to most observers, NATO was well aware of the new radar installation, which they referred to as Steel Yard.

Even from the earliest reports it was suspected they were tests of an over-the-horizon radar., and this remained the most popular theory during the cold war. Several other theories were floated as well, including everything from jamming western broadcasts to submarine communications. The broadcast jamming theory was debunked early on when a monitoring survey showed that Radio Moscow and other pro-Soviet stations were just as badly affected by woodpecker interference as Western stations. More speculative explanations were also offered, claiming it was a system for weather control or even an attempt at mass subconscious mind control.

As more information about the signal became available, its purpose as a radar signal became increasingly obvious. In particular, its signal contained a clearly recognizable structure in each pulse, which was eventually identified as a 31-bit pseudo-random binary sequence, with a bit-width of 100 µs resulting in a 3.1 ms pulse. This sequence is usable for a 100 µs chirped pulse amplification system, giving a resolution of 15 km (10 mi) (the distance light travels in 50 µs). When a second Woodpecker appeared, this one located in eastern Russia but also pointed toward the US and covering blank spots in the first system's pattern, this conclusion became inescapable.

In 1988, the Federal Communications Commission conducted a study on the Woodpecker signal. Data analysis showed an inter-pulse period of about 90 ms, a frequency range of 7 to 19 MHz, a bandwidth of 0.02 to 0.8 MHz, and typical transmission time of 7 minutes.

- The signal was observed using three repetition rates: 10 Hz, 16 Hz and 20 Hz.
- The most common rate was 10 Hz, while the 16 Hz and 20 Hz modes were rather rare.
- The pulses transmitted by the woodpecker had a wide bandwidth, typically 40 kHz.

Пропускной пункт на территорию РЛС (Checkpoint to the territory of the radar) - by Alex Taranenko

Disappearance

Starting in the late 1980s, even as the FCC was publishing studies of the signal, the signals became less frequent, and in 1989 disappeared altogether. Although the reasons for the eventual shutdown of the Duga-3 systems have not been made public, the changing strategic balance with the fall of the cold war in the late 1980s likely had a major part to play. Another factor was the success of the US-KS early-warning satellites, which entered preliminary service in the early 1980s, and by this time had grown into a complete network. The satellite system provides immediate, direct and highly secure warnings, whereas any radar-based system is subject to jamming, and the effectiveness of OTH systems is also subject to atmospheric conditions.

According to some reports, the Komsomolsk-na-Amure installation in Siberia was taken off combat alert duty in November 1989, and some of its equipment was subsequently scrapped. Google Maps photography of the area shows that the antenna has been removed. The original Duga-3 site lies within the 30 kilometer Zone of Alienation around the Chernobyl power plant. It appears to have been permanently deactivated, since their continued maintenance did not figure in the negotiations between Russia and Ukraine over the active early warning radars at Mukachevo and Sevastopol. The antenna still stands, however, and has been used by amateurs as a transmission tower (using their own antennas) and has been extensively photographed.

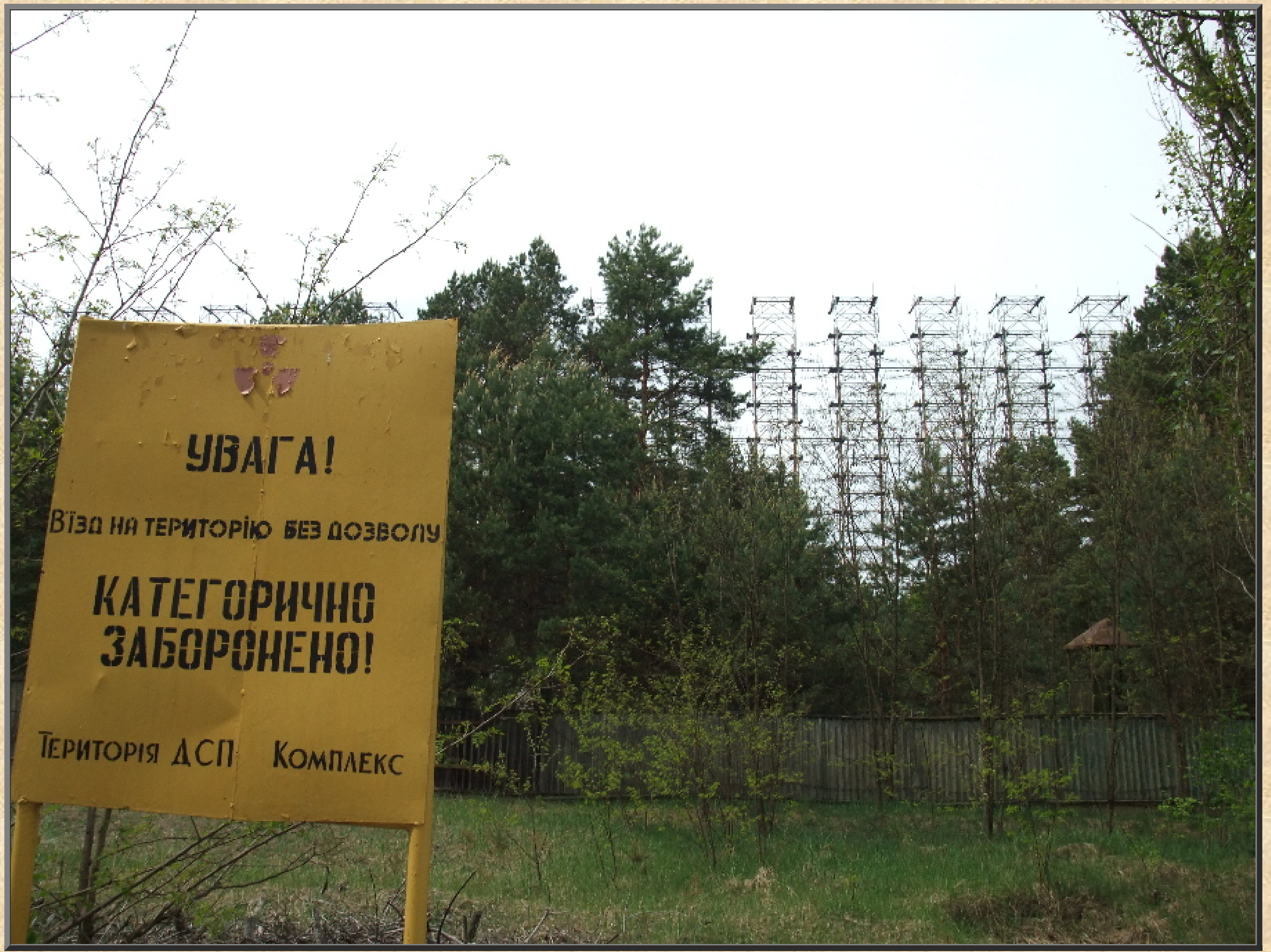

Warning Sign and Antennas by markus_online_world

Positions

Duga-1 and -2, Nikolayev, 47°02′28.33″N, 32°11′57.29″E
Duga-3, Gomel/Minsk, (Chernobyl-2) 51°18′19.06″N, 30°03′57.35″E and 51°38′15.98″N, 30°42′10.41″E
Duga-3, Komsomolsk-na-Amure, 50°53′34.66″N, 136°50′12.38″E and 50°23′07.98″N, 137°19′41.87″E

РЛС / Radar station by Alex Taranenko

Appearances in media

The Ukrainian-developed computer game S.T.A.L.K.E.R has a plot focused on the Chernobyl Nuclear Power Plant and the nuclear accident there. The game heavily features actual locations in the area, including the Duga-3 array. It is presented as the "Brain Scorcher", a wide-area mind control device which must be deactivated by the player.

References:

1. **The Russian word "duga" means "arc" in English.**
2. **Sean O'Connor. Russian Strategic Defense - Part 2, The ABM Network**
3. **Dave Finley (7 July 1982). "Radio hams do battle with 'Russian Woodpecker'". *The Miami Herald***
4. **David L. Wilson (Summer 1985). "The Russian Woodpecker... A Closer Look". *Monitoring Times***
5. **"Mystery Soviet over-the-horizon tests". *Wireless World*: 53. February 1977**
6. **J.P. Martinez (April 1982). "Letter from J. P. Martinez". *Wireless World*: 89**

Further reading

- **Headrick, James M., Looking over the horizon , IEEE Spectrum, pp. 36 - 39 (July 1990).**
- **Headrick, James M. and Skolnik, Merrill I., Over-the-horizon radar in the HF band, Proceedings of the IEEE, vol. 62, no. 6 pp. 664 - 673 (June 1974).**
- **Headrick, James M., Ch. 24: "HF over-the-horizon radar," in: Radar Handbook, 2nd ed., Merrill I. Skolnik, ed. [New York: McGraw-Hill, 1990].**
- **Kosolov, A. A., ed. Fundamentals of Over-the-Horizon Radar (translated by W. F. Barton) [Norton, Mass.: Artech House, 1987].**

External links

- **Chernobyl-2. Sercet Military Facility in the territory of exclusion zone. Text and photos 2008**
- **OTH-Radar "Chornobyl - 2" and Center of space-communication**
- **"Circle" is an auxiliary system for OTH-Radar "Chornobyl - 2"**

- **The Russian Woodpecker, *Miami Herald*, July 1982.**
- **Steel Yard OTH, globalsecurity.org**
- **Sound clip of Russian Woodpecker interfering with WWVH time station, November 1984 (Wikimedia Commons)**
- **Some pictures of Chernobyl-2**
- **Images of the Gomel installation.**

Retrieved from http://en.wikipedia.org/wiki/Russian_Woodpecker

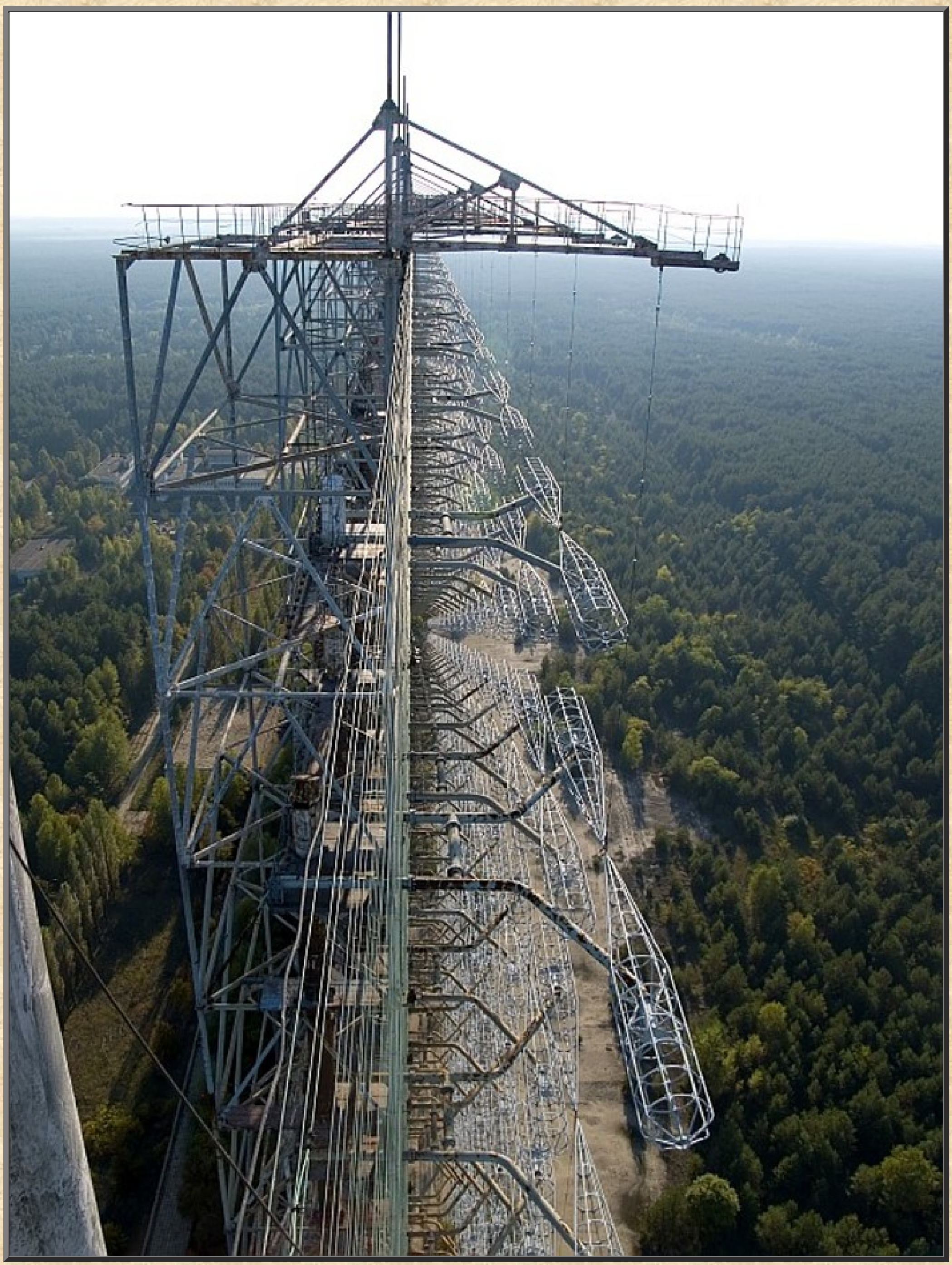

Чорнобиль-2. Приймальні антени. (Chernobyl-2. RX antennas.) by djcrok

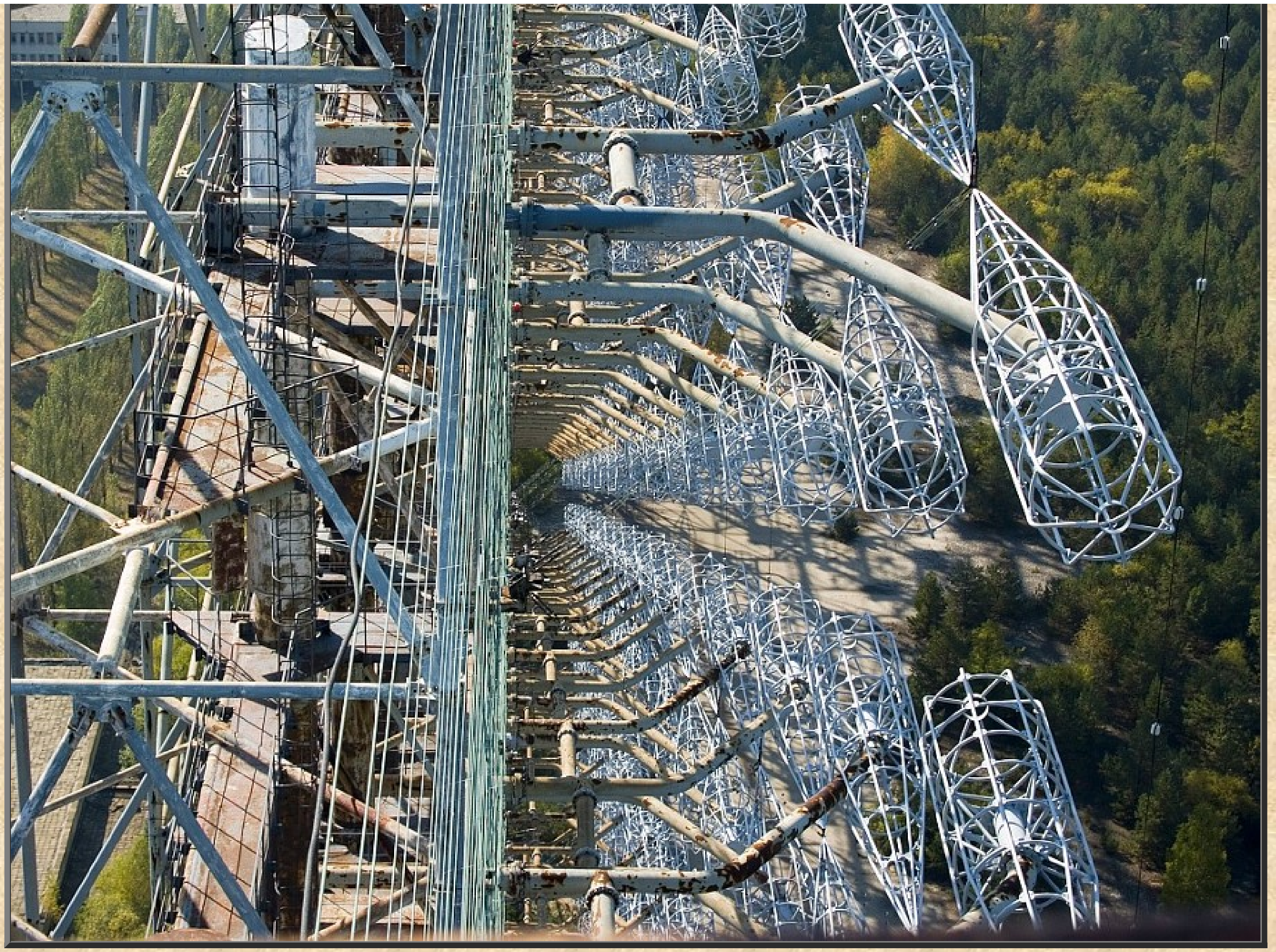

Чорнобиль-2. Приймальні антени (Chernobyl-2. Receiving Antenna) by djcrok

Чорнобиль-2. (Chernobyl-2.) by djcrok

www.pripyat.com

Чорнобиль-2. Приймальні антени. (Chernobyl-2. RX antennas.) by djcrok

Чорнобиль-2. Приймальні антени. (Chernobyl-2. RX antennas.) by djcrok

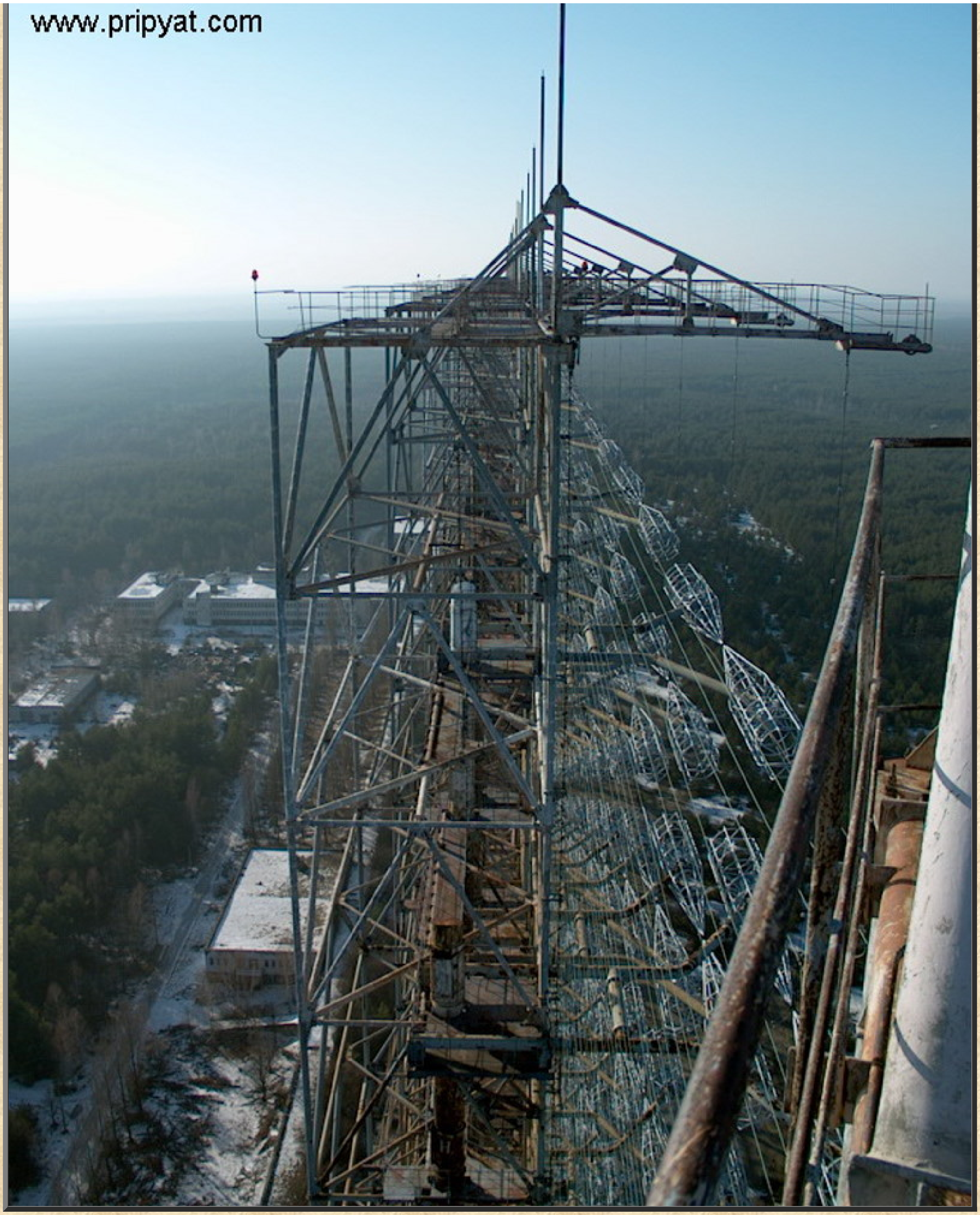

Чорнобиль-2. Приймальні антени. (Chernobyl-2. RX antennas.) by djcrok

Чорнобиль-2. Приймальні антени. (Chernobyl-2. RX antennas.) by djcrok

Дуга (Arc) by A.Miched

WWW.KRANZ.COM.UA

Чорнобиль-2. Залишки соціалізму. (Remains of socialism.) by djcrok

…

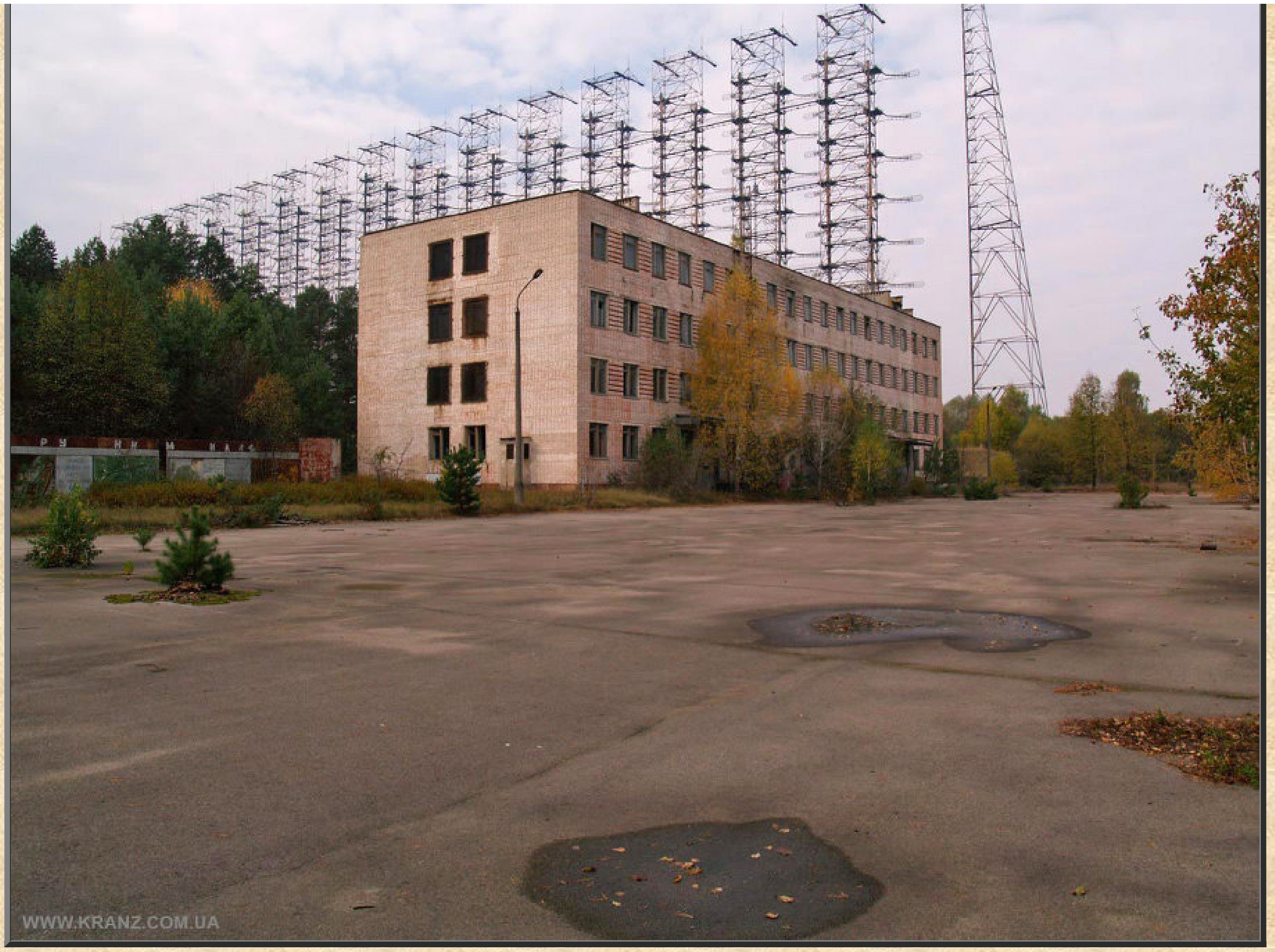

Чорнобиль-2. Залишки соціалізму. (Remains of socialism.) by djcrok

www.kranz.com.ua

http://www.kranz.com.ua/oth_radar_chornobyl2_R.htm

http://www.kranz.com.ua/report/oth_radar_chornobyl2_52.htm

FAIR USE NOTICE: This page contains copyrighted material the use of which has not been specifically authorized by the copyright owner. Pegasus Research Consortium distributes this material without profit to those who have expressed a prior interest in receiving the included information for research and educational purposes. We believe this constitutes a fair use of any such copyrighted material as provided for in 17 U.S.C § 107. If you wish to use copyrighted material from this site for purposes of your own that go beyond fair use, you must obtain permission from the copyright owner.

~ MENU ~

Webpages © 2001-2016
Blue Knight Productions

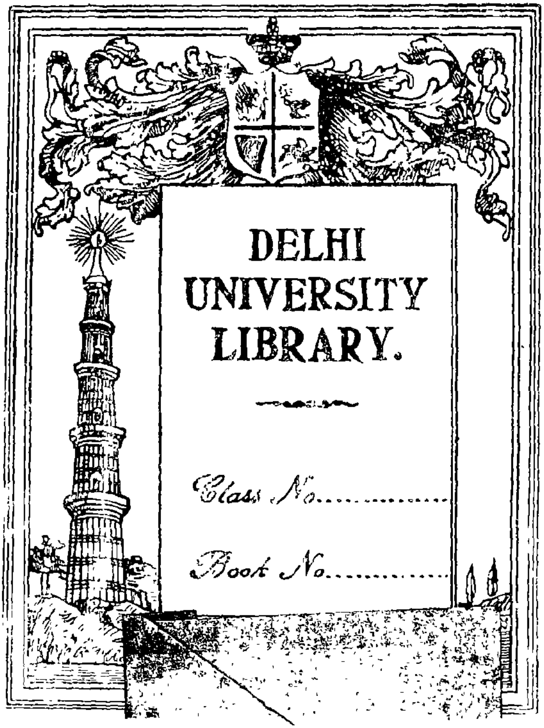

DELHI
UNIVERSITY
LIBRARY.

Class No..............

Book No.............

DELHI UNIVERSITY LIBRARY

Cl. No. U28626 G2

Ac. No. 36281 Date of release for loan

This book should be returned on or before the date last stamped below. A fine of **one anna** will be charged for each day the book is kept overtime.

22 JAN 1954

4 OCT 1957

19 JUN 1958

3 AUG 1959

3 AUG 1959

[illegible] AUG 1959

Methuen's Monographs on Physical Subjects
General Editor: B. L. Worsnop, B.Sc., Ph.D.

ATMOSPHERIC ELECTRICITY

METHUEN'S MONOGRAPHS ON PHYSICAL SUBJECTS

F'cap 8vo, 2s. 6d. net each

General Editor : B. L. Worsnop, B.Sc., Ph.D.
Senior Lecturer in Physics
King's College, London

SPECTRA. By R. C. Johnson, D.Sc. *Second Edition.*

WAVE MECHANICS. By H. T. Flint, D.Sc. *Second Edition.* 3s. 6*d. net.*

THE PHYSICAL PRINCIPLES OF WIRELESS. By J. A. Ratcliffe, M.A. *Second Edition.*

THE CONDUCTION OF ELECTRICITY THROUGH GASES. By K. G. Emeléus, M.A., Ph.D.

MAGNETISM. By E. C. Stoner, Ph.D.

X-RAYS. By B. L. Worsnop, B.Sc., Ph.D.

X-RAY CRYSTALLOGRAPHY. By R. W. James, M.A., B.Sc.

THE COMMUTATOR MOTOR. By Professor F. J. Teago.

APPLICATIONS OF INTERFEROMETRY. By W. E. Williams, M.Sc.

PHOTOCHEMISTRY. By D. W. G. Style, B.Sc., Ph.D.

THERMODYNAMICS. By Professor A. W. Porter, F.R.S.

THERMIONIC VACUUM TUBES. By Professor E. V. Appleton, M.A., D.Sc., F.R.S. 3s. *net.*

WIRELESS RECEIVERS. By C. W. Oatley, M.A., M.Sc.

ATMOSPHERIC ELECTRICITY. By B. F. J. Schonland, M.A., Ph.D.

THE METHOD OF DIMENSIONS. By Professor A. W. Porter, D.Sc., F.R.S.

In Preparation

SURFACE TENSION. By A. H. Ferguson, D.Sc.

THE NUCLEUS. By Professor J. A. Crowther.

TABLES OF PHYSICAL CONSTANTS. By W. H. J. Childs, B.Sc., Ph.D.

ATMOSPHERIC ELECTRICITY

BY

B. F. J. SCHONLAND, O.B.E., M.A., PH.D.
SENIOR LECTURER AND FELLOW IN PHYSICS AT THE
UNIVERSITY OF CAPE TOWN

WITH A PREFACE BY

O. W. RICHARDSON, F.R.S.
YARROW RESEARCH PROFESSOR OF THE ROYAL SOCIETY
NOBEL LAUREATE IN PHYSICS, 1928

METHUEN & CO. LTD.
36 ESSEX STREET W.C.
LONDON

First Published in 1932

PRINTED IN GREAT BRITAIN

GENERAL PREFACE

THIS series of small monographs is one which should commend itself to a wide field of readers.

The reader will find in these volumes an up-to-date résumé of the developments in the subject considered. The references to the standard works and to recent papers will enable him to pursue further those subjects which he finds of especial interest. The monographs should therefore be of great service to physics students who have examinations to consider, to those who are engaged in research in other branches of physics and allied sciences, and to the large number of science masters and others interested in the development of physical science who are no longer in close contact with recent work.

From a consideration of the list of authors it is clear that the reader need have no doubt of the accuracy of the general accounts found in these volumes.

O. W. RICHARDSON

KING'S COLLEGE
LONDON

CONTENTS

PAGE

GENERAL PREFACE v

CHAP.

I. THE IONISATION OF THE ATMOSPHERE . . . 1

II. THE PENETRATING RADIATION 21

III. ELECTRIC FIELDS AND ELECTRIC CURRENTS IN THE ATMOSPHERE 43

IV. THE ELECTRIFICATION OF THUNDERSTORMS . . 72

GENERAL REFERENCES 96

INDEX 97

ATMOSPHERIC ELECTRICITY

CHAPTER I

THE IONISATION OF THE ATMOSPHERE

THE electrical conductivity of the air, first established by Linss in the year 1887, plays an important part in all the phenomena included under the title Atmospheric Electricity. Since 1901, when Elster and Geitel and C. T. R. Wilson independently discovered that this conductivity is due to the presence of small carriers of positive and negative charge called ions, the nature and the origin of these ions have been the subject of many investigations. Two different paths have been followed, the one dealing with the number and the nature of the atmospheric ions, the other with the radiations producing them. *The two paths converge on the problem of the* ionisation-balance of the atmosphere, the manner in which an equilibrium is maintained between the rate of creation and the rate of disappearance of the ions. It is only of recent years that an answer to this problem has emerged.

The investigation of the ionisation of the air has been rendered exceptionally difficult by the number of separate factors involved and by the variation of each with local geological and meteorological conditions. Much remains to be done before the conditions governing the conductivity of even the lowest layers of the atmosphere are completely understood; observations at greater heights in the regions accessible to balloons and airships are just emerging from the difficult pioneer stage, and the very important regions of the upper air which lie beyond fifteen kilometres above the ground can at present be examined only indirectly by deductions based upon the mode of propagation of upwardly directed wireless waves and upon the variations of the earth's magnetic field.

§ 1. The Atmospheric Ions, Small and Large

A gas molecule is ionised when some external agency supplies the energy necessary to remove an electron from it, whereupon the positively charged residue of the molecule and the ejected electron, both of which under ordinary pressures very soon attach themselves to one or more uncharged molecules, form what is called an ion-pair. Such ions are known as "normal," "small," or "fast," and when found in the atmosphere have mobilities of the same order as those determined for pure dry air in laboratory experiments, about 1·5 cm./sec. in a field of gradient 1 volt per cm.

The presence in the air of minute particles of dust, of the products of combustion, and of extremely small drops of water even when the air is unsaturated, leads very often to the capture of a normal ion by one of these condensation or Aitken nuclei, as they are termed, and so to the formation of another type of ion of low mobility, ranging from ·0005 to ·0003 cm./sec., known as "large," "slow," or "Langevin" ions. Roughly speaking, one-third of the condensation nuclei present in the air is electrically neutral, the other two-thirds carry, nearly equally, positive and negative charges captured from small ions. The interplay between the large and the small ions and the uncharged nuclei is of considerable importance.

Besides these two main types of atmospheric ion, a group of "intermediate" ions whose mobilities lie between ·1 and ·01 cm./sec. has been found to be present under special meteorological conditions such as low relative humidity. While the origin of this type of ion is at present uncertain, its existence emphasises the fact *that the customary division into the two main classes,* large and small, requires refinement, and future work will no doubt proceed in the direction of dividing the ions according to their mobilities into a larger number of groups. J. Nolan and de Sachy have recently found evidence of such groups in the small and the large ion "families."[1]

It has been found that the great majority of the ions, large and small, carry a single elementary charge, either positive or negative.[2] Over land areas the large ions considerably outnumber the small ones, often in the ratio of ten to one. Thus while the numbers, n_+ and n_-, of small ions of each sign per c.c. of the air over land usually range from 300 to 1000, those of the large ions, N_+ and N_-, amount to from 1,000 to 80,000 per c.c. Since practically all the large ions are formed from the capture of small ones by nuclei, an increase in the number of these nuclei, such as is due to the fires of an industrial area, will increase the proportion of the large ions in the air at the expense of the small ions.

Over the oceans this state of affairs is reversed ; the air away from land areas is but poorly supplied with nuclei and the large ions are in the minority. The mean values of n_+ and n_- at sea lie between 500 and 700 per c.c., while N_+ and N_- are about 200 per c.c. In view of their comparative freedom from capture by nuclei, it is at first sight rather surprising that the number of small ions per c.c. in mid-ocean differs but little from the same quantity over the land, where the average values of n_+ and n_- are also about 600 per c.c. This is, however, to be explained (§ 12) by the fact that the decreased rate of disappearance of the small ions over the sea is balanced by a decrease in the rate at which they are produced.

§ 2. THE DETERMINATION OF THE NUMBER OF IONS PER C.C.

To determine the numbers of the small ions, positive and negative, per c.c. (n_+ and n_-) an arrangement known as an Ebert ion-counter is usually employed. This consists in essentials of a long earthed metal tube A (Fig. 1) on the axis of which is mounted a charged insulated rod B, connected to a quartz fibre electroscope. Air is sucked through the tube for about five minutes by means of a clock-work driven fan F and the speed of the air-stream is arranged to be slow enough for all the small ions entering A with charges opposite in sign to that on the central rod to be drawn to B before reaching the bottom of the tube. The speed of flow must at the same time be great enough to make it impossible for any appreciable number of large ions to have time to reach the rod. If v is the volume of the air sucked through in the course of the measurement (indicated by an anemometer C registering its revolutions on a dial), the quantity of electricity gathered in by the central electroscope system is ven_-, where e is the elementary charge on a single ion and the rod is supposed to have been positively charged. If then C is the capacity of the whole central system, rod and electroscope, and V and V′ are the potentials indicated by the electroscope before and after the experiment, $C(V - V') = ven_-$, an equation from which n_- can be

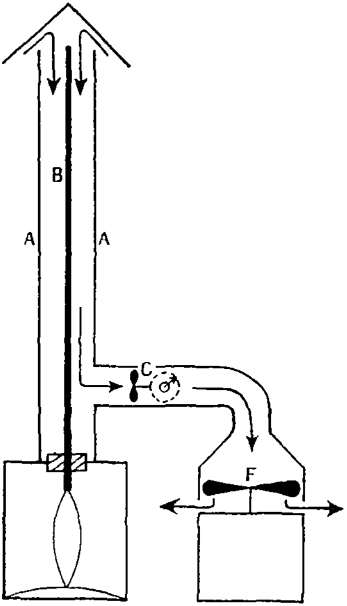

FIG. 1.—Ebert ion-counter.

found. To determine n_+, the measurement must be repeated with the rod B charged negatively.

The same method, using a longer tube, a stronger field in its interior, and a slower draught of air, and thus giving time for the capture of all the slow ions as well, can be applied to the determination of $N + n$. A combination of the two determinations then yields N_+ and N_-, the numbers of the large ions per c.c. of each sign.

The values found for n_+ and n_- are not usually the same; the ratio n_+/n_- over both land and sea has a mean value of 1·22. Although, as explained in § 30, the earth's electric field may sometimes affect the apparatus in such a way as to give a spuriously high ratio, Swann, who has made a special study of the point, concludes that when all precautions are taken a real excess of positive ions must be admitted. Probably the main factor causing this excess is the superior mobility of the small negative ion; Nolan has found that k_-/k_+ has the value 1·16. A lower mobility implies a lower velocity of thermal agitation for the positive than the negative ion, and consequently less likelihood of capture by condensation nuclei. Another contributing cause may be the Ebert effect discussed in § 38. This general excess of positive small ions does not necessarily involve a nett positive charge in the air; it is easily compensated for by a comparatively small excess in the much larger number of large negative ions.

An instrument known as the Aitken dust-counter is used to find the total number of nuclei, charged and uncharged, per c.c. It consists of a small expansion chamber filled with a sample of the air and saturated with water-vapour. The sudden movement of a piston gives an adiabatic expansion which causes drops of water to condense upon the nuclei; the resulting cloud falls upon a glass plate divided into squares and the number of droplets is counted under a microscope. Though called a dust-counter it does not record the coarser dust in the air.[3]

§ 3. The Polar Conductivities and the Ionic Mobilities

A simple relation holds between the conductivity of the air and the numbers per c.c. and the mobilities of the ions responsible for conduction. Consider a point in the atmosphere at which the electric field intensity is F and the specific conductivity, the reciprocal of the specific resistance, is λ. The conduction current, i, through unit area drawn perpendicular to the field is then equal to Fλ, for F is numerically equal to the potential difference, and $1/\lambda$ is the resistance, between the ends of a unit cube of the air. In this equation we assume that conditions are such that Ohm's law can be applied to the air. Since the current is carried by positive and negative streams of ions moving in opposite directions we may write

$$i = \mathrm{F}(\lambda_+ + \lambda_-), \quad . \quad . \quad . \quad (1)$$

where λ_+ and λ_- are called the polar conductivities.

Actually the existence of the two main types of ion, large and small, leads to the flow of four different ion-streams, two up and two down if the field is vertical, in each of which the velocity is different, being given by F × k, where k is the corresponding mobility. The total current is thus

$$i = \mathrm{F}e(n_+k_+ + n_-k_- + \mathrm{N}_+\mathrm{K}_+ + \mathrm{N}_-\mathrm{K}_-), \quad . \quad (2)$$

where k_+ and k_-, K_+ and K_-, are the mobilities of the positive and the negative ions, large and small respectively, and e is the elementary charge. From (1) and (2), separating out the oppositely moving streams,

$$\lambda_+ = (k_+n_+ + \mathrm{K}_+\mathrm{N}_+) \text{ and } \lambda_- = (k_-n_- + \mathrm{K}_-\mathrm{N}_-). \quad (3)$$

In clear country air over land, n_+ and n_- number about 600 per c.c., while N_+ and N_- are about 2000 per c.c. ; k_+ and k_- are 1·5 approximately and K_+ and K_-, ·0004. Here, therefore, the ratio of the first to the second term in each bracket above is 1100 : 1 and most of the conductivity of the air is due to the small ions. Near large towns, however, n is usually of the order of 100 per c.c. and N, on account of the heavy pollution of the atmosphere may be 50,000. Under these conditions the second term would be one-seventh of the first and the large ions would play a part in the transfer of electricity through the air.*

It is not usual to find the polar conductivities from equations (3) above, but directly, from measurements of the rate which the charge on an insulated body exposed to the free air is dissipated by the flow of ions to it. Consider a negatively charged body exposed to moving air and let σ be the density of the electrification on any portion of area dS.[1] Positive ions will move to the body, and since the field F just outside the surface is $4\pi\sigma$, the current per unit area tending to dissipate the charge will be $F\lambda_+ = 4\pi\sigma\lambda_+$. The total dissipation current, obtained by integrating over the whole exposed surface, is $4\pi\lambda_+\int\int\sigma dS$ or $4\pi\lambda_+Q$, where Q is the charge exposed to the air. Since this current represents the rate at which the body loses charge, we have $-dQ/dt = 4\pi\lambda_+Q$, a relation which provides a means of determining λ_+. A similar equation, with λ_- substituted for λ_+, holds for the case of a positively charged body.

An arrangement used for this purpose is due to Gerdien (Fig. 2); it resembles the Ebert apparatus already described, but the air-stream is now made so rapid and the field in the interior of the tube so weak, that only a very small fraction of the total number of ions passing through A has time to reach the central rod. In this case the numbers of ions per c.c. are not appreciably altered by the experiment and Ohm's Law applies.

If the potential V of the negatively charged central electroscope system is observed to rise at a rate dV/dt and if C is the total capacity of the central system, rod and electroscope, we have $-dQ/dt = -CdV/dt$. Also, if C′ is the capacity of that portion of the central system exposed to the stream of air, the rod B and its support, Q = C′V, consequently

---

* The methods used to determine ionic mobilities are described in text-books on the conduction of electricity through gases

$$-CdV/dt = 4\pi\lambda_+ C'V, \quad . \quad . \quad . \quad (4)$$

a relation from which λ_+ can be determined. A similar experiment with the rod positively charged yields λ_-. The use made of the values found for the polar conductivities and an adaptation of the Gerdien method due to Schering are described in Chapter III, § 30.

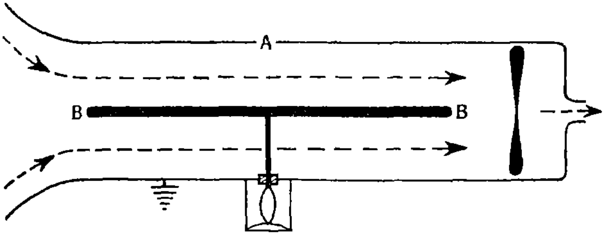

Fig. 2.—Gerdien conductivity apparatus.

§ 4. The Ionising Agencies

Three principal agencies are responsible for the ionisation of the lower atmosphere: radiations from radioactive substances in the surface of the earth, radiations from radioactive matter present in the air itself, and the penetrating radiation. For a discussion of their relative importance it is convenient to use the symbol q to represent the number of "pairs of ions produced per c.c. per second in air at N.T.P.," a quantity which will be referred to as the ionising power of the radiation. This phrase in inverted commas is usually represented by the symbol I. Thus the statement $q = 4$ I means that, at the point in question, the radiations create four pairs of ions per c.c. per second, if the medium there is air at N.T.P. The contributions of the three different agencies just mentioned to the total ionising power, q, are distinguished by the symbols q_e, q_a, and q_h.

§ 5. The Effect of Radioactive Matter in the Earth

Uranium, thorium, and their products are widely distributed through the earth's crust, whose surface layers therefore emit α, β, and γ rays into the air. Since the α rays can only emerge from a very thin layer and can only ionise the first few cms. of the air, their effect is inconsiderable. The β rays can come from greater depths and penetrate much further into the atmosphere. Calculations based upon average values for the measured concentration of radioactive matter in the earth indicate that an effect ranging from 1 I at the surface to 0·1 I at about 10 metres above it is due to these rays. The γ rays are a more important factor, for they can come from still greater depths and consequently from a still larger quantity of radioactive material. It is estimated that they produce on the average 3 I at the surface, 1·5 I at 150 metres above it, and 0·3 I at a height of 1 kilometre. The recent discovery of the γ ray activity of potassium introduces another source of ionisation, for the potassium content of the earth's crust is considerable. No very detailed survey of this question has yet been made, but measurements indicate that at a height of 10 metres the contributions of radium, thorium, and potassium to the γ ray ionisation of the air are in the ratio of 1 : 1 : 0·7, in accord with sampling of the surface content of the earth.[5] The average value for the ionising power of the earth-radiations near the ground has been found directly to lie between 2 and 10 I, which is in accord with the measurements of the radioactive content of the earth's crust. Local surface and geological conditions naturally have a marked effect upon the value of q_e at any given spot.

The radioactive content of sea-water is found to be negligible in comparison with that of soil and rocks, and over the oceans the earth-radiation effect is very small indeed.

§ 6. THE EFFECT OF RADIOACTIVE MATTER IN THE AIR

The atmosphere contains a considerable quantity of the radioactive emanations, radon and thoron, and their successive products. A wire charged negatively to a potential of a few hundred volts and exposed to the air collects an easily detectable quantity of radium A and thorium A by recoil from the disintegration of these emanations. The gases themselves arise from the decay of radium and thorium within the earth, from which they escape by diffusion, by thermal convection, and as a result of decreases in the external atmospheric pressure. The greater portion of the ionisation thus produced is due to the α rays from the emanations themselves and from their A and C products, for here, unlike the case of the radiations from the earth, there is no absorbing layer between the disintegrating atom and the air. The radioactive gases and their products are fairly evenly distributed by atmospheric turbulence through the lower air. This "air-radiation" effect can be calculated from radioactive data if the emanation content of the air is determined. The emanations may be removed from a measured volume of air by drawing it through tubes cooled with liquid air or filled with coconut charcoal. The quantity of emanation condensed or absorbed is then determined by the effect produced in an ionisation electroscope, corrected for the decay of the material during the time of the experiment. A more direct method is to fill a closed ionisation chamber first with ordinary air and then with air which has been freed by the above method from the emanations and from their products, and to determine the saturation currents in the two cases. If these are i_1 and i_2 respectively, v the volume of the vessel and e the elementary charge, the value of the air-radiation ionising power, q_a, is given by $i_1 - i_2 = q_a ve$.

The results of such measurements agree with estimates based upon observations of the average emanation content of the air in placing the air-radiation effect at about 5 I in the neighbourhood of the earth's surface. Of this about 60 per cent. is due to radon and its α ray products and the remainder to the thoron series. It varies somewhat with locality and with those meteorological conditions which determine the rate of escape of the emanations Apart from the effect of strong upward winds in elevating the richer surface layers, the emanation content of the air should decrease fairly rapidly with height. Direct information on this point is difficult to obtain, but what exists indicates that the air-radiation effect falls to less than 2 per cent. of its surface value at a height of 5 kilometres.

The effect, like that of the earth-radiation, is extremely small over the oceans, where the emanation content of the air is only about 1 per cent. of its value over land areas.

§ 7. The Effect of the Penetrating Radiation

The third factor in the ionisation of the atmosphere is a radiation of peculiar interest, whose origin and nature are still the subject of much investigation. It has a penetrating power which is considerably greater than that of the most penetrating γ rays from radioactive bodies, and it travels to the earth's surface in a downward direction, undergoing a certain amount of absorption on the way. Its ionising power at sea-level over both land and ocean areas lies between 1·5 and 2·5 I. Though these figures correctly suggest that the absolute value of the effect is not accurately known, relative observations can be made with great certainty and show that the ionising power of the radiation at a given height above sea-level is practically independent of geographical position and alters only slightly with meteorological conditions such as pressure. The ionisation due to this radiation increases rapidly with height above sea-level, rising ten-fold in the first 5 kilometres. It is discussed in greater detail in the next chapter.

§ 8. The Ionisation Inside a Closed Vessel

It is a matter of considerable difficulty to determine separately the ionising powers of the earth-radiation and of the penetrating rays. Calculations of the former must be based upon measurements of the radioactive content of the earth's crust, and these are necessarily tedious and rough. To proceed further, one must examine the ionisation produced in the air inside a sealed metal chamber under conditions which are modified from those existing in the free air by the presence of the metal walls. This involves correction to free-air conditions. A further disadvantage is that a new ionising agency is introduced, the radioactive matter in and on the inside walls. The ionising power of this "wall-effect," or the "residual ionisation" of the vessel as it is called, is a constant depending on the material of which the walls are made and the cleanliness of the inner surface. It is usually represented by the symbol q_0.

The total ionising power of all the radiations operating upon the gas in a closed vessel may be called q' and the modified effects of the three main ionising agencies q'_e, q'_a, and q'_h : we then have

$$q' = q'_e + q'_a + q'_h + q_0. \quad . \quad . \quad (5)$$

We may suppose the vessel to take the form of a sealed ionisation chamber filled with air at N.T.P. which has been freed from radioactive matter. The saturation current carried by the ions to the central collecting electrode is then given by $i = q've$, where v is the volume of the vessel and e the elementary charge on an ion.

Various methods have been employed to separate out the four terms on the right-hand side of equation (5) from one another. The most satisfactory of these begins by determining q_0. The instrument, whose form is described in the following chapter, is taken to a large snow-fed lake whose radioactive content, like that of the sea, is negligible. When it is sunk to the depth of a few feet in such a lake, the effects of earth and air-radiations are non-existent and equation (5) becomes $q' = q'_h + q_0$; on lowering the vessel still further, q'_h diminishes as the penetrating radiation becomes absorbed in the increasing thickness of water, and it is possible to reach depths at which the saturation current is practically constant and unaffected by further sinking of the instrument. At this stage $q' = q_0$, and the residual ionisation is determined. By combining this observation with that made close to the surface of the lake, q'_h may be found.

Since the effects of the penetrating radiation and of the walls of the vessel are both independent of local conditions other than height above sea-level, they remain the same when the instrument is transported to a site on land at the same level. A measurement of the saturation current in the chamber under land conditions gives q' as stated by equation (5), and hence the sum of the earth and air effects, q'_e and q'_a can be found In actual practice the air-radiation effect within a closed vessel is extremely small: the walls exclude α and β rays coming from outside and reduce q'_a to the ionising power of the γ rays emitted by the B and C products of the radioactive emanations. This amounts to but 4 per cent. of the effect produced by the air-radiation in the free atmosphere, in the case of a vessel with steel walls a few millimetres thick. It is therefore sufficient to determine the air-radiation effect, q'_a, by calculation based upon a knowledge of the emanation content of the air and of the nature of the walls of the vessel. The fourth factor, the effect of the earth-radiation, q'_e, is then obtained by substitution in equation (5).

To pass from these closed vessel determinations of the effect of the radioactive matter in the earth to the effect to be expected in the free air, experiments may be made with screens of different thicknesses placed around the ionisation chamber and the values thus obtained for the earth-radiation effect extrapolated to zero thickness. In doing so account has to be taken of the effect of the screens upon q'_h.

Until more is known of the properties of the penetrating radiation, the relation of its ionising powers inside and outside a closed vessel cannot be stated with any certainty, but it is probable that unless the walls are very thick they will differ little from one another.

The greater part of the residual ionisation arises from α rays and takes place close to the walls ; q_0 is therefore the effect of the residual ionisation near the walls divided by the volume of the vessel.

§ 9. SUMMARY OF IONISING POWERS OF VARIOUS AGENCIES

The table on opposite page gives a rough idea of the effects of the three main factors in the ionisation of the air at sea-level, both in closed vessels of steel, a few millimetres thick, and in the free atmosphere. The values given refer to measurements at about a metre from the surface of the earth ; they are averages and, except in the case of the penetrating radiation, are subject to considerable fluctuations with local conditions.

It is noteworthy that the rate of production of ions over land areas is about six times as great as over the oceans, where the ionisation is practically all due to the action of the penetrating rays. Over land close to the surface of the earth, these rays are much less important, but at greater heights the effects of the earth and the air-radiations fall off rapidly as we have seen, so that here also, from a height of one of two kilometres upwards, the penetrating radiation is the chief agent in making the air a conductor.

§ 10. THE RATE OF DISAPPEARANCE OF THE SMALL IONS

The number of ions produced by the agencies we have discussed will not increase indefinitely as time goes on. Even in pure filtered air, free from nuclei, the small ions disappear by diffusion to the walls and by recombination with one another. Experiments have shown

TABLE I

IONISING POWERS AT THE EARTH'S SURFACE, SEA-LEVEL

Ionising Agent.	Ionising Power in Ion-pairs per c.c. per sec. in Air at N.T.P.			
	Unscreened Closed Vessel.		Free Air.	
	Over Land.	Over Sea.	Over Land.	Over Sea.
Radioactive matter in the earth . .	4·0	0	4·0	0
Radioactive matter in the air . . .	0·2	0	5·0	0
Penetrating Radiation	2·0	2·0	2·0	2·0
	Total ionising power for the free air .		11·0	2·0

that when the diffusion loss can be neglected, the equation $dn/dt = q - \alpha n^2$ applies to such filtered air, n being as before the number of positive or negative small ions per c.c., q the number of pairs of ions produced per c.c. per second, and α the so-called recombination constant of small ions. In the course of time a state of equilibrium will be reached, in which the processes of creation and disappearance of ions balance one another ; then $dn/dt = 0$ and the equilibrium equation $q = \alpha n^2$ holds. Applying this equation to the free air, taking $q = 11$ I from the preceding section and $\alpha = 1{\cdot}6 \times 10^{-6}$ cm.^{3}/sec. from laboratory experiments, we find $n = 2622$. This is very much larger than the number of small ions actually observed, which is round about 600 per c.c. The discrepancy is due to neglect of the important part played by the Aitken or condensation nuclei in capturing the small ions.

An equation applicable to air containing nuclei was first given by von Schweidler [6] in the form

$$dn/dt = q - \alpha n^2 - \gamma n Z,$$

where γ is a constant and Z the total number of nuclei, charged and uncharged, per c.c. of the air. For equilibrium, when dn/dt is zero, it becomes

$$q = \alpha n^2 + \gamma n Z = n(\alpha n + \gamma Z) = n(\alpha n + \beta) \qquad (6)$$

or

$$q = \beta' n \quad . \quad . \quad . \quad . \quad (7)$$

The quantity β' is known as the disappearance constant of the small ions. Experiment shows that in ordinary air, where nuclei are plentiful, the term αn is small compared with β ; even over the oceans it is probably only one-seventh of β. Thus β', while not a true constant, for it depends upon the concentration Z of the nuclei in the air, is to a large extent independent of the number of small ions per c.c. Equation (7) thus states a linear law of recombination as opposed to the quadratic law which holds for filtered air.

If the action of the ionising agents were removed, the small ions would disappear at a rate $dn/dt = \beta' n$, giving $n = n_0 e^{-\beta' t}$, for the number present at any subsequent time. It is assumed in the integration that there are sufficient nuclei present to make β' independent of n. Von Schweidler has pointed out that the quantity $\theta = 1/\beta'$ can, by analogy with the equation for radioactive decay, be termed the average life of a small ion. Equation (7) then becomes

$$q\theta = n \quad . \quad . \quad . \quad . \quad (8)$$

in which θ should vary inversely as the number of nuclei per c.c. of the air.

In a development of this theory put forward by J. J. Nolan and his co-workers, [2, 1] separate consideration is given to the capture of small ions by charged nuclei of opposite sign, leading to neutralisation, and by uncharged nuclei, leading to the formation of large ions. For the small positive ions, they write

$$dn_+/dt = q - \alpha n_+ n_- - \eta_1 n_+ N_0 - \eta_2 n_+ N_- = 0, \text{ for equilibrium} \quad . \quad (9)$$

Here N_0 and N_- are the numbers of uncharged and negatively charged nuclei per c.c. respectively, while η_1 and η_2 are the recombination coefficients for the two types of capture. By a theoretical discussion, too long to be reproduced, this equation is converted into the simpler form $q = [\eta_2 Z/2{\cdot}1]n$, which corresponds to Schweidler's equation (6) and again indicates that the average life of an ion should vary as $1/Z$.

This average life θ is found over land areas to lie between 10 and 60 seconds according to the purity of the air, the higher values being associated with observations made in the country. Over the oceans it may be as much as 230 seconds, according to the measurements of Hess [7] and of Mathias [8] on the island of Heligoland. The table below, taken from Hess, illustrates the great difference in the equilibrium between small and large ions brought about when the wind direction changes ; the first set of observations was made when the wind blew from the nuclei-laden European mainland, the second when it blew from the polar regions :—

TABLE II

OBSERVATIONS OF THE STATE OF ATMOSPHERIC IONISATION (HESS, HELIGOLAND)

Wind.	No of Small Ions per c.c (n).		No. of Large Ions per c.c.(N).		Total No. of Nuclei.	Average Life of Small Ion, secs.	
	n_+	n_-	N_+	N_-	$Z = N_0 + N_+ + N_-$	θ_+	θ_-
S.W. (Land)	220	267	1480	1510	6300	22	22
N.N.W. (Arctic)	794	843	200	200	1100	115	204

These authors find that for values of θ lying between 36 and 100 seconds, Schweidler's linear equation (8) is valid; in purer air, where θ is greater than 100 seconds, the simple equation no longer holds, for the quadratic recombination term αn^2 then becomes significant.

Experiments by P. and J. Nolan[9] and their co-workers have, however, suggested that the form of the accepted equilibrium equation is incorrect and should be replaced by a new empirical relation $q = n(\alpha n + \zeta\sqrt{Z})$, where ζ is a constant. For air with a sufficient concentration of nuclei this reduces to $q = [\zeta\sqrt{Z}]n$, which would imply that the mean life of an ion varies as $1/\sqrt{Z}$. It is therefore possible that the processes governing the disappearance of the small ions in the atmosphere are more complex than has hitherto been supposed. Account may have to be taken of variations in the actual nature of the nuclei, of complexities in the small and the large ion families, such as are indicated by the existence of "intermediate" ions, of the possible occurrence of multiply-charged large ions when the number of nuclei is small,[10] and of the effect of actual dust-nuclei.

§ 11. METHODS OF DETERMINING THE AVERAGE LIFE OF A SMALL ION

The average life, θ, of the small ions can be determined for a sample of air enclosed in an ionisation chamber by two methods suggested by Schweidler.[6] The first involves the separate measurement of the ionising power, q, of the ionising radiations, and of the number of small ions per c.c., n, and substitution in the equation $q\theta = n$. q is obtained from the saturation current, i, using the relation $i = qev$, where v is the volume of the vessel. The corrections to be applied have already been discussed (§ 8). The quantity n is found by allowing the air to stand undisturbed for a sufficient time for equilibrium between small ions and nuclei to be attained and then applying a strong field for a few seconds to sweep the small ions on to the collecing electrode. The second method involves the measurement of the currents flowing in an ionisation vessel for various values of the applied potential difference.[11]

§ 12. THE IONISATION-BALANCE IN THE LOWER ATMOSPHERE

It is of considerable importance to decide to what extent the various factors in the creation and destruction of ions account for the state of ionisation actually observed in the air over land and sea. Any failure in this respect would suggest that new processes must be sought.

The evidence available indicates, however, that the chief factors in the ionisation-balance have been discovered. Over land areas where nuclei are usually plentiful, the Schweidler equilibrium equation, $q\theta = n$, applies. An average value of 60 seconds for θ, the mean life of a small ion, and of 650 per c.c. for the small ion concentration, requires that the rate of creation, q, should be 650/60 or 10·8 ions/c.c./sec. in satisfactory agreement with direct measurements of q, shown in Table I. No direct measurements of θ have yet been made over the oceans. The concentration of nuclei in mid-ocean has been estimated by Hess, from measurements on air from the Arctic, to be such as to give an average life of 230 seconds to a small ion. Since n was found on the *Carnegie* to amount to about 550 small ions/c.c., the ionising power at sea should amount to 550/230 or 2·4 I. The agreement of this figure with the value 2·0 I shown in Table I. is reasonable, in view of the roughness of the estimate of θ used, and of the present uncertainty in the value of q itself.

Hess's estimate of the concentration of Aitken nuclei over the oceans, about 800 per c.c., is borne out by direct measurements on the *Carnegie*, which give means of 870 and 774 per c.c. in the Atlantic and Pacif measure-
respectively, and by a recent count eaks, or in balloons,
the Atlantic. oust and portable, as

The problem of the ionisation

cannot be said to be completely solved until agreement is reached on the form of the interaction-relation between small ions and nuclei. The general nature of the factors in the ionisation-balance has, however, been established, and an explanation can be given of the paradox that the conductivity of the air undergoes no large change as one passes from mid-ocean to land, despite a six-fold increase in the rate at which ions are formed. The increased birth-rate of the small ions is counterbalanced by an increased death-rate, owing to the greater concentration of predatory condensation-nuclei over the land areas.

REFERENCES TO PAPERS

[1] J. J. Nolan and de Sachy, *Proc. Roy. Irish Acad.*, A, 37, p. 71 (1927).
[2] J. J. Nolan, Boylan, and de Sachy, *ibid.*, A, 37, p. 1 (1925).
[3] Boylan, *ibid* , A, 37, p. 58 (1926).
[4] C. T. R. Wilson, *Dict. of Appd. Physics*, Art. *Atmospheric Electricity*, vol. 3, p. 84.
[5] Suckstorff, *Naturwissenschaften*, 19, p. 87 (1931).
[6] von Schweidler, *Wien Sitz. Ber.*, 2*a*, 127 (1918), 128 (1919), 133 (1924).
[7] Hess, *Gerland's Beit. z. Geophysik*, Bd. 22, p. 256 (1929).
[8] O. Mathias, *ibid.*, Bd. 27, p. 360 (1930).
[9] J. J. Nolan and P. J. Nolan, *Proc. Roy. Irish Acad.*, A, 40, p. 11 (1931).
[10] Boylan, *ibid.*, A, 40, p. 76 (1931).
[11] O'Brolchain, *Gerland's Beit. z. Geophysik*, 29, p. 1 (1931).

CHAPTER II

THE PENETRATING RADIATION

UNTIL the year 1911 it was generally supposed that the ionisation inside a closed vessel was entirely due to radiations from the walls, and from radioactive matter in the earth and the outside air. It was therefore to be expected that this ionisation would decrease when the vessel was raised above the surface of the earth. This was tested in balloon ascents by Hess in 1911, and Kolhörster in 1913, who found that the diminution with height was only appreciable during the first kilometre, after which an increase occurred. These observations led Hess to postulate an ionising radiation coming from above and suffering partial absorption in the atmosphere.

Something has already been said of this radiation in § 7 ; the accurate determination of its properties has proved a difficult matter. Its great penetrating power implies that the ionisation it creates within a closed vessel is so small as to require specially refined technique to measure accurately in the presence of the other ionising agencies discussed in Chapter I. Further, the fact that it comes from above and suffers some absorption necessitates investigations at high altitudes so that any softer and more easily absorbed components which do not reach sea-level may be examined. Such measurements must be made on mountain peaks, or in balloons, under conditions which require robust and portable, as well as sensitive, instruments.

§ 13. Experimental Methods of Measuring the Intensity of the Radiation

The most extensive measurements of the intensity of the radiation have been made by the ionisation method. The first point to be attended to is to eliminate as far as possible the disturbing effects of the earth and air-radiations. This can best be done by sinking the ionisation chamber beneath the waters of a lake, or by surrounding it with a thick shield of non-radioactive iron or lead and correcting for the small fraction of the γ rays which are still able to enter the chamber.

The ionising power of the penetrating rays at sea-level is about 2·0 I (ions per c.c. per sec.), so that their contribution to the saturation current in a vessel of volume 1 litre, filled with air at N.T.P., would be $3{\cdot}2 \times 10^{-16}$ amperes. In order to increase this effect, it has become the general practice to fill the vessel with carbon dioxide at a pressure of 30 atmospheres or more.* This carries with it an important advantage, in that it considerably reduces the effect of the residual ionisation, which, we have already seen, is caused by α rays from the walls. The range of these rays in CO_2 at high pressure is small, and they produce all their effect within a few millimetres of the walls themselves. Here the electric field driving ions to the central collecting electrode is weakest and cannot separate out more than a small fraction of those produced before they recombine. A further decrease in the residual effect can be obtained by using a wire cage as an electrostatic screen to form the walls of an inner ionisation chamber. This cuts down the amount of wall material productive of undesirable α rays (Hoffmann, Fig. 4).

The quantity actually measured in these experiments is the sum of the penetrating ray and the residual ionisations, q_h and q_0, and fluctuations in the latter limit the accuracy with which q_h can be found. These arise from the fact that in a given time-interval the number of α particles emitted from the walls undergoes variations about a mean value which are subject to well-known statistical laws. To smooth them out it is generally necessary to make observations over a period of several hours. For accurate measurements in a shorter time instruments have been devised by Hoffmann[1] and by the author[2] which are sensitive enough to show the sudden increases in the ionisation due to single α particles and so to determine the residual ionisation directly.

---

* Increasing the pressure does not give a proportional increase in ionisation. For 30 atmospheres the effect is only 14 times that at N.T.P. and the actual factor is somewhat uncertain. Hence the difficulty of finding the absolute value of q_h.

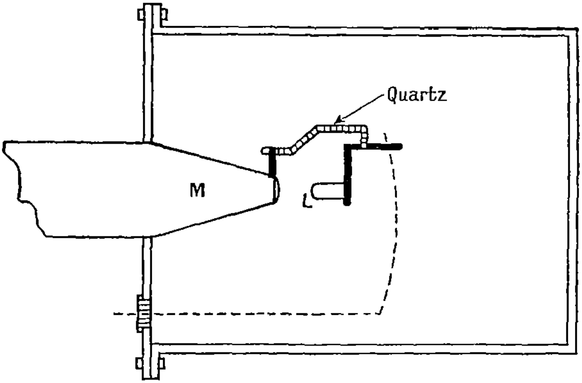

FIG. 3.—Kolhörster ionisation chamber and electroscope.

The portable instrument developed by Kolhörster is shown diagrammatically in Fig. 3. The ionisation chamber is of steel, coated inside with zinc, and is filled with purified air at atmospheric pressure. Two small loops L of platinised quartz fibre, mounted upon a rod held in a quartz insulator, form the central electrode and can be charged by a device operated from outside by an electromagnet (shown dotted). Their potential is determined from their distance apart as read in the eye-piece of a microscope M. In Millikan's instrument quartz fibres are also used as the central electrode, but the gas-filling is CO_2 at a pressure of 30 atmospheres. In both cases the quartz and invar-steel framework supporting the fibres is designed to reduce the effects of temperature variations. The chief disadvantage of this type of chamber is the small volume permissible in view of the pressure and the use of a high-power microscope. The penetrating radiation ionises the gas through the agency of fast-moving particles (§ 17) which fluctuate in number like the α particles from the walls. To smooth them out in the case of small ionisation vessels necessitates observations extending over many hours.

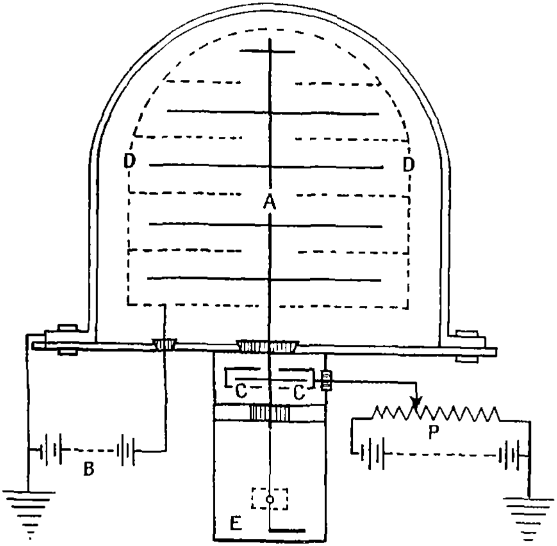

FIG. 4.—Hoffmann's null method for the measurement of the ionisation due to the penetrating radiation.

Another type of apparatus, due to Hoffmann,[3] employs an electrometer which is outside the ionisation vessel (Fig. 4). The chamber is filled with CO_2 at a pressure of about 30 atmospheres, the residual radiation is reduced by the use of a wire cage for the walls of the ionisation vessel proper, and saturation conditions are obtained by adopting the branched form of central electrode A and cage D. The electrometer E is of a specially sensitive type developed from the quadrant electrometer. It serves as a null instrument, for arrangements are provided to maintain the whole central system at zero potential. The cage is kept permanently charged by means of the battery B, and the charge carried to the central electrode by the ions is compensated for by an equal and opposite charge induced upon the inner plates of the small condenser C by means of the potentiometer P. The amount of charge collected during a given time can be found from the capacity of C and the potential applied to the outer plates. Both C and the electrometer are kept evacuated. Though more complicated and less portable than the first type, this instrument has the advantages of freedom from all insulation losses on the collecting system and of greater volume.

The more recent portable instruments show a reduction in the ionising power of the residual wall radiation; in Millikan's 1926 form the ratio of this quantity to the ionising power of the penetrating radiation at sea-level was 4·5/1; in the form used by him in 1930 this had been reduced to 1/30. The ratio for Hoffmann's instrument is less than 1/5.

§ 14. The Ionisation-depth Curve in Air and Water

In studying the variation of the ionisation intensity of the radiation above and below sea-level, it has been found convenient to measure distance downwards from the top of the atmosphere. This distance is usually expressed in equivalent metres of water, i.e. in terms of that column of water which would possess the same mass per unit area of cross-section as the column of air above the point of observation. Thus sea-level is 10·33 equivalent metres of water below the top of the atmosphere.

Observations of the ionisation at various depths below the top of the atmosphere are most conveniently made by sinking an ionisation vessel in a lake situated at a very great altitude. The results when plotted against the total depth, equivalent and real, of water below the top of the atmosphere constitute an ionisation-depth curve. An extensive series of investigations has been carried out by Millikan and Cameron to determine the shape of this curve, using lakes on mountain peaks in both North and South America.[4] Thus they were able to examine the absorption of the rays from near the surface of Gem Lake, California, where the depth below the top of the atmosphere was 7·4 equivalent metres of water (altitude 3000 metres), down to 81 actual metres below the surface of the lake, 86 equivalent metres from the top of the atmosphere. Owing to the need for some water-screening to cut down the air-radiation effect, q'_a, trustworthy observations began at a total depth of some 8 metres.

They have repeated these observations in other lakes at lower altitudes and have been able to show that at the same total depth below the top of the atmosphere the ionisation is always the same. This indicates that within the reasonably high accuracy of their measurements, the absorbing power of the additional thickness of air involved in the observations at lower altitudes is in fact that of the equivalent layer of water. As a consequence of this result measurements made at still greater altitudes on mountain peaks and in balloons can be plotted on the ionisation-depth curve for water with some degree of confidence that the absorption in the air can be reduced to that in water in this way. Observations made in air alone, however, require the instrument to be shielded with lead from the external earth and air-radiations and have not only to be corrected for the small fraction of these which can still penetrate the shields but also for the absorption of the penetrating radiation in the lead. For these reasons they are not so reliable for plotting on the curve as the measurements made in water.

The ionisation-depth curves obtained by Millikan and Cameron are shown in Fig. 5, the portion below 30 metres being inset on a larger scale. The observations made in Lake Arrowhead (altitude 1700 metres) and Gem Lake (altitude 3000 metres) constitute the main "water" curve. The curve marked "land" shows the land observations made at various heights from sea-level to the top of Pike's Peak, California (altitude 4700 metres, 6·2 equivalent metres of water below the top of the atmosphere), inside a thick lead shield. By comparing the abscissæ of the land and water curves for the same ordinate, the water-equivalent of the shield at the higher altitudes was found to be 1·22 metres. The dotted portion of the water curve is obtained by adding this amount to the equivalent depth of the land observations.

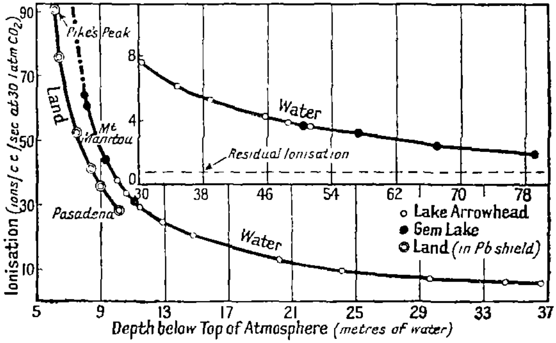

FIG. 5.—Ionisation-depth curve in air and water (Millikan and Cameron).

Even at a total depth of 80 metres the water curve is not parallel to the horizontal axis and the value of the residual ionisation (1·2 ions/c.c./sec. in CO_2 at 30 atmospheres) has to be obtained by extrapolation.

These measurements show clearly the very great penetrating power of the radiation, for the most penetrating known gamma rays (from thorium C″) would be reduced to one-fiftieth of their original intensity in their passage through a single metre of water. The lower part of the ionisation-depth curve has been extended still further by Regener [5] in experiments on Lake Constance. His ionisation vessel had a volume of 34 litres and was filled with CO_2 at a pressure of 30 atmospheres. This was surrounded by a protecting tank filled with water from the surface of the lake which formed a screen

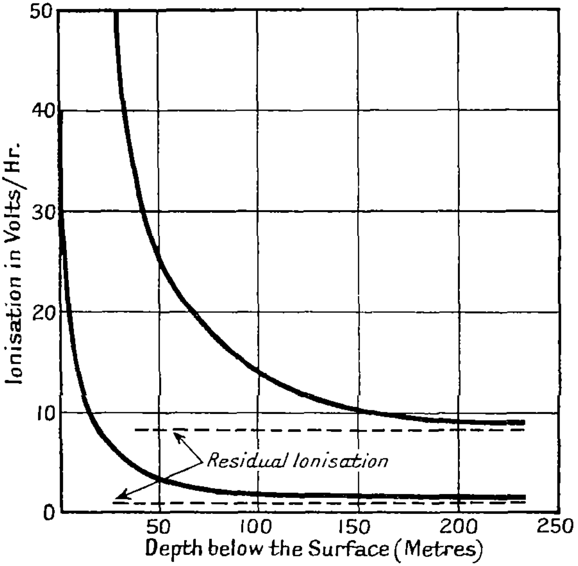

Fig. 6.—Ionisation-depth curve (Regener).

1 metre in thickness and prevented possible variations with depth in the small radioactive content of the surrounding water from affecting the measurements. With this arrangement Regener has followed the radiation down to a depth of 230 metres of water below sea-level, the large volume of the chamber enabling accurate measurements to be made even at great depths. At 230 metres the curve was found to be still falling and could still be determined with an error of the order of 2 per cent. The movement of the fibres of the electrometer was recorded automatically by photographing their position at intervals of one hour. Regener's curve is shown in Fig. 6, with the ordinates of the portion below 30 metres repeated on a tenfold scale.

§ 15. ANALYSIS OF THE IONISATION-DEPTH CURVE

The form of the ionisation-depth curve suggests that it may be made up of three or four separate curves corresponding to component radiations of different penetrating powers. On this view the initial steep portion would correspond in the main to an easily absorbed component which does not reach sea-level in any quantity compared to the others, while only the most penetrating component can reach to the great depths examined by Millikan and Regener. In endeavouring to analyse the curve in this way it has been assumed that the process of absorption of a parallel beam of the s-th component is represented by the well-known equation $-dI = \mu_s I dH$, where $-dI$ is the decrease in the ionisation I due to passage through a thickness dH of the absorbing medium and μ_s is a constant, characteristic of the radiation and of the medium, which is called the absorption coefficient of the s-th component. Integration of this equation gives $I = I_0 e^{-\mu_s H}$. The significance to be attached to μ_s will be discussed later.

There is, however, no reason to believe that the radiation strikes the top of the atmosphere as a parallel beam, and we must assume incidence with equal intensity in all directions and take account of the increased thickness of air or water to be traversed as the angle of incidence is increased. When this is done, it is found that the relation between the ionisation I at a depth H and that which would be found at the top of the atmosphere, I_0, takes the form $I = I_0\int_1^\infty [e^{-\mu_s H x} \,.\, dx/x^2]$.* Tables of this integral have been given by Gold.

The analysis of the ionisation-depth curve must be carried out by a trial and error method, for not only the number of components but the absorption coefficient μ_s and the initial intensity I_0 of each has to be chosen to fit the curve. Analyses of this kind have been made by Millikan and Cameron, and by Regener, whose results are shown in the following table :—

TABLE III

ANALYSES OF THE IONISATION-DEPTH CURVE INTO COMPONENTS

Compt.	Millikan.		Regener.	
	Absorption Coefficient, μ_s.	Intensity at Top of Atmosphere, I_0.	Absorption Coefficient, μ_s.	Intensity at Top of Atmosphere, I_0.
1	0·80 *metres*$^{-1}$	141,000	?	?
2	0·20 ,,	130	0·21	130
3	0·10 ,,	80	0·073	51
4	0·03 ,,	33	0·02	6·5

The values of μ have been expressed in reciprocal metres of water ; in the same units the mean absorption coefficient of the gamma rays from radium B + C after filtration through 10 mm. of lead would be 5·4. Since Regener's experiments did not extend above sea-level he has not been able to determine the constants of the most easily absorbed component.

---

* See Rutherford, *Radioactive Substances and their Radiations*, p. 262, where x = sec θ.

The agreement obtained by these independent observers for the other three components is striking; slight discrepancies in μ may be explained by the uncertainty in the exact value of the residual ionisation, for the curves are still falling at the lowest depths reached. The difference in relative intensities may be due to some selective action arising from the difference in the size of the chambers used.

§ 16. IONISATION-DEPTH CURVES IN HEAVY METALS: THE TRANSITION ZONE

The absorption of the radiation in heavy metals has been studied by placing metal screens round the ionisation vessel. Since this corresponds to sinking it in a lake of the metal, the results obtained should agree with the ionisation-depth curve taken in water when the thickness of each screen is expressed in equivalent metres of water. It is, however, found that the absorbing power of a metal such as lead is initially very high. The curve falls abnormally rapidly at first (for 10 cm. in the case of Pb); afterwards it runs parallel to, but lower than, the water curve, additional screens producing the same reduction as their water-equivalents.* It has been shown that this effect is due to the transition from air to lead, and something similar is exhibited in all transitions from one medium to another. Fig. 7, after Steinke,[6] shows the ionisation-depth curves of aluminium and lead as thick lines. The effects of adding an inside layer of Pb after 28 cm. of Al have been used, and of adding inside layers of Al after 4 and 10 cm. of Pb, are shown by the dotted lines. The increase in the ionisation in two of these cases is very remarkable. It is suggested that the transition zone effect is due to an alteration in the equilibrium between the primary rays and their secondary electrons (§ 20), but no quantitative explanation can yet be given. The phenomenon is of great importance in connection with the problem of the nature of the radiation.

---

* In this work the water-equivalent is that thickness of water with the same number of extra-nuclear electrons in a column one sq. cm. in cross-section. Millikan's water-equivalents (§ 14) are mass-equivalents.

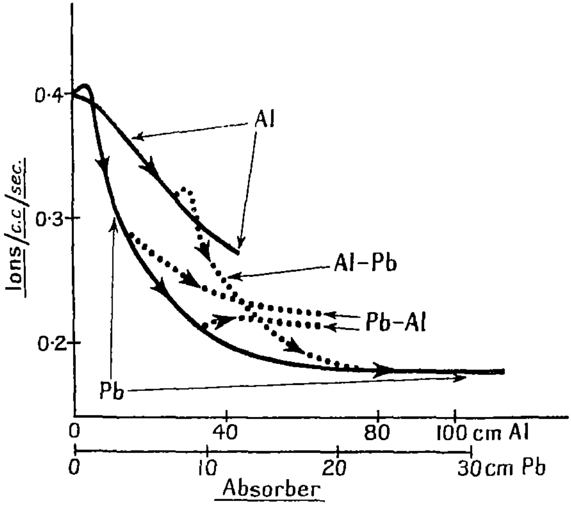

Fig. 7.—Transition-zone effects in Pb and Al (Steinke).

A striking instance of the initial drop in the absorption curve in lead is afforded by a screen used by Millikan in mountain observations. Its calculated water-equivalent was 85 cm., but a comparison with the ionisation-depth curve in water showed that it was effectively 170 cm. at sea-level and 122 cm. at a height of 5000 metres.

§ 17. PHOTOGRAPHIC REGISTRATION AND COUNTING OF THE IONISING PARTICLES

The ions produced by the rays can be studied by means of the Wilson cloud chamber, in which a water-drop is condensed on each. Skobelzyn [7] has shown in this way that much of the ionisation is due to the action of fast-moving particles which travel in a downward direction at the rate of about one particle per square centimetre per minute at sea-level. The photographs of their tracks show that they create some 40 ion-pairs per centimetre of their path in ordinary air.

It is not yet certain whether all the ionisation arises from particles of this kind. A number of the tracks show a paired relation which may be significant, two or more particles entering the chamber during one expansion in directions making a small angle with one another.

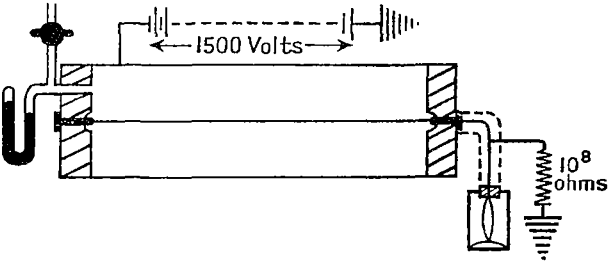

FIG. 8.—Geiger-Müller tube-counter.

The passage of these ionising particles can be continuously recorded with the tube-counter of Geiger and Müller. In this (Fig. 8) the ionisation vessel is a tube of zinc and the collecting electrode a thin wire along the axis, coated with a semi-insulating layer of oxide or varnish. Conditions of pressure and potential difference are adjusted below the sparking value so that the entry of each ionising particle gives rise to a sudden pulse of current of very short duration. The advantage of this device is that the effect of an α particle is equal to that of a β particle, instead of 200 times greater as in the ordinary ionisation chamber. The same cleaning and screening precautions as have been described for the latter must be taken. This counter has been used by Regener [5] to confirm the readings on his ionisation-depth curve; the submerged counter was joined to an arrangement which recorded each pulse automatically.

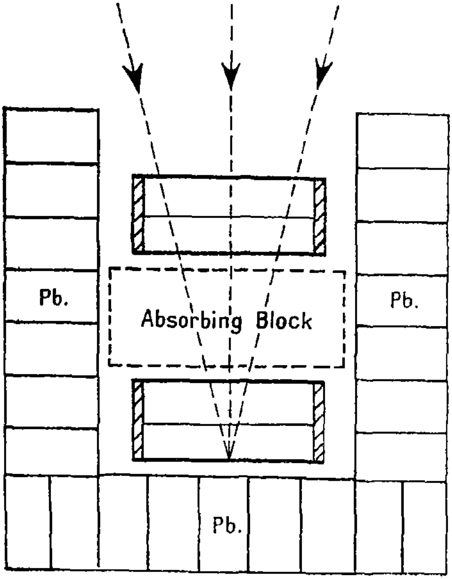

Fig. 9.—The Bothe-Kolhörster experiment.

An apparatus which counts the penetrating ray particles only has been developed by Kolhörster and Bothe (Fig. 9). Two counters are mounted vertically above one another, and account is taken only of those pulses which coincide in time with one another and are considered to be due to the passage of particles which are able to penetrate both counters. The coincidences were originally determined by recording the sudden deflections of both electroscopes on the same moving photographic film. In a development due to Bothe,[8] the two counters are joined to the two grids of a screened-grid valve which is so arranged that a pulse of plate current occurs only when both grids are practically simultaneously changed in potential.

§ 18. The Bothe-Kolhörster Experiment

Bothe and Kolhörster[9] have used this arrangement to determine the absorption coefficient of the ionising particles themselves. A block of gold 4·1 cm. thick was placed between the counters, and the reduction in the number of coincidences was observed. The value found for the absorption coefficient was so close to that given by the usual ionisation chamber method as to lead them to conclude provisionally that the ionising particles are the primary radiation and not of secondary origin. In an extension of this experiment carried out by Rossi,[10] however, evidence was found of the actual production of fresh ionising particles when the radiation was absorbed. These presumably arise from a primary radiation of γ ray type.

§ 19. The Energy of the Ionising Particles

If the ionising particles are electrically charged, the deformation of their paths in a transverse magnetic field should give information as to their nature and energy. Several unsuccessful experiments of this kind have been interpreted as indicating that their momentum is too great for them to be bent in the fields used. Mott-Smith[11] in this way has placed a lower limit of 2×10^9 *e*-volts * to their energy if they are electrons, and 10^9 *e*-volts if they are protons.

The Bothe-Kolhörster-Rossi experiment suggests that the effective range of the particles is about 10 equivalent metres of water or $7{\cdot}7 \times 10^5$ cm. of air at N.T.P. They produce some 40 ion-pairs per centimetre of air and each ion-pair costs the particle about 30 *e*-volts. Hence the total initial energy of a particle must be of the order of $40 \times 30 \times 7{\cdot}7 \times 10^5$ or $9{\cdot}2 \times 10^8$ *e*-volts.

In experiments carried out under thunderstorms,[2] the author has found that the strong electric fields within and outside the cloud are able to stop a large proportion of these particles. He concludes that their energy must be less than 5×10^9 *e*-volts, if they are electrons or protons.

Such energies are more than a hundred times greater than those of the fastest particles emitted by radioactive bodies. If the particles are electrons they must be moving with a velocity only about 40 metres per second less than that of light and their inertial mass is 2000 times their rest-mass, slightly greater than the mass of a slow proton. At such energies electrons and protons may be expected to give approximately the same ionisation per unit path and the same absorption coefficient.

It has recently proved possible to obtain magnetic bending of the tracks in a Wilson cloud-chamber.[18] The paired particles are found to be of two types, positively and negatively charged, presumably protons and electrons. Their energies range from 10^7 to 10^9 *e*-volts.

§ 20. The Nature of the Penetrating Radiation

In view of the difficulties in the way of determining the nature of the fast ionising particles and of some uncertainty as to whether there are not other ionising rays of much shorter range, any discussion of the nature of the penetrating radiation itself is purely tentative. In this section two of the many important points which have been raised in this connection will be outlined.

---

* Electron-volts, the energy expressed in terms of the difference of potential in volts through which an electron must travel to attain this energy.

The first concerns the effects to be expected if the primary radiation is itself constituted of fast-moving charged particles. In such a case the earth's magnetic field should produce a deflection of their paths leading to a concentration of the particles in the higher magnetic latitudes. Brucke has shown that for a fast electron to reach the earth's surface below magnetic latitude 50° its energy must exceed 5×10^9 volts, and for it to attain the equator it must have twice this energy, even if loss of energy in passage through the air be neglected. This question is being investigated at the present time; Millikan[12] has shown that the sea-level intensities in latitudes 59° and 34° are the same within the limits of accuracy of his measurements (1 per cent.). These and other observations speak against a corpuscular nature for the primary rays.

The second point concerns the view that the ionising particles are secondary electrons arising from the passage of primary ultra-gamma ray quanta through the atmosphere, and the utilisation of absorption coefficients derived from the ionisation-depth curve to find the frequencies of these quanta.[13] On this view an ionising particle is produced when a quantum of energy $h\nu$ makes a Compton collision with an electron in an atom. The electron moves forward with energy $h(\nu - \nu_1)$ and the quantum recoils with a lower "degraded" frequency ν_1. Consideration of the equations governing the Compton scattering process shows that for high-frequency quanta the collision in most cases leaves quantum and electron moving forward along paths only slightly inclined to one another, so that the radiation at any point within the atmosphere should consist of a mixture of primary and degraded quanta of various orders (the results of one, two, and more successive collisions) together with the fast electrons arising from collisions of primary and degraded quanta.

Absorption measurements are made upon this mixed beam and yield an absorption coefficient $\bar{\mu}$; what relation this bears to the absorption coefficient of the pure primary radiation, μ, can be determined only by involved calculation. Since the degraded quanta have frequencies ν_1, ν_2, etc., which are less than the primary frequency ν, they are more easily absorbed than the primary quanta. Thus, as the radiation travels through the air, there will come a stage when as many degraded quanta of each kind disappear in each centimetre of path as are produced by Compton processes in the same centimetre. The parent beam is then in equilibrium with its *quantum* products in the same way as a long-lived parent radioactive body such as radium reaches equilibrium with its descendants. From this stage onwards, say after a thickness equivalent to D cm. of water, the constitution of the quantum part of the beam will remain unaltered, and this part will disappear at a rate set by the rate of disappearance of the primary quanta. Until this stage is reached $\bar{\mu}$ cannot equal μ.

Actual absorption measurements, however, are concerned with the ionising electrons and not with the quanta which give rise to them. Consider the first scattering process only, which we may assume for simplicity to give rise to electrons all of the same range, R cm. of water. Then the ionisation at a depth H cm. of water below the top of the atmosphere will be due to electrons from points lying between H and H − R cm. from the top. Thus the actual measurements reflect the state of the beam some distance above the point of observation and can give only an apparent absorption coefficient $\bar{\mu}$ which equals μ, provided that $H - R > D$. If this condition is not fulfilled, the number of electrons brought to rest in 1 cm. is less than the number generated in the same distance and $\bar{\mu}$ will be less than μ.

The ionisation-depth curve has been analysed into four component curves, distinguished by different values of $\bar{\mu}$. The significance of these apparent absorption coefficients obviously depends upon how far the criteria for equilibrium are satisfied in each case. Since according to Millikan, only the least penetrating component reaches equilibrium with its electrons and degraded quanta, the values of $\bar{\mu}$ found for the others will be less than μ. The deduction of their true absorption coefficients is at present impossible, for it involves a wide extrapolation of the laws governing the passage of γ and β rays through matter. Recent experiments have shown that these processes are not completely accounted for by existing theories. In particular, high-frequency quanta interact with atomic nuclei in a manner which is likely to be very important in the case of the penetrating radiation. The paired proton and electron tracks referred to in §§ 17 and 20 show that the rays frequently disintegrate atomic nuclei. This process is not taken into account in the γ ray absorption formulæ. If the radiation does consist of a mixed beam of ultra-gamma quanta and fast electrons, the observed absorption coefficients do not in general approximate to the true coefficients, and it is not possible to deduce the quantum frequencies except perhaps in the case of the softest component.

§ 21. The Origin of the Radiation

The intensity of the radiation is remarkably constant, and no certain variation with solar or sidereal time has been established. Corlin[14] and others have put forward evidence of a sidereal time effect which decreases with depth below the top of the atmosphere and amounts at sea-level to about 2 per cent. of the mean intensity. The effect only shows in averages of a long series of hourly observations, for alterations in the effective thickness of the atmosphere (not always faithfully represented by changes in the barometer) and other causes give rise to larger and irregular fluctuations.

Hoffmann, with the sensitive instrument described in § 13, can find no evidence for a sidereal time variation. A suggested solar effect also awaits confirmation; if it exists at all it is very small ($\frac{1}{3}$ per cent).[16]

It is most natural to look for the source of quanta of such high energy in the interiors of the hot stars, where high temperatures may lead to the conversion of matter into radiation. This possibility has, however, been rejected owing to the absence of appreciable solar and sidereal time effects.

Another argument against it is provided by the estimated total energy reaching the outer atmosphere in the form of penetrating radiation. This is more than one-tenth, probably one-fifth, of the total energy arriving from all the stars as light and heat. If the radiation originated in stellar interiors, very little of it could escape unaltered and most of it would be turned into heat. Unless, therefore, the radiation is generated only in the outermost layers of the stars, it is not possible for it to convey anything like such a large fraction of the total stellar radiation as one-fifth.

Millikan [4] has considered the possibility of the radiation arising from the formation of more complex elements out of hydrogen in interstellar space. The mass of a heavy atom being less, by Δm, than the separate masses of its constituent protons and electrons, energy of amount $\Delta m \times c^2$ ergs would be radiated out as a quantum whose frequency is given by $h\nu = \Delta m \times c^2$, where c is the velocity of light. He considers that the four main components of the radiation (§ 15) represent the formation of the four most abundant groups of elements, apart from hydrogen, viz. helium, "oxygen" (C, N, and O), "silicon" (Mg, Al, and Si), and "iron" (Fe, Ni, and Co). These elements make up 99 per cent. of all known matter, excluding hydrogen. The atomic weight determinations of Aston with the positive ray mass spectrograph give Δm in each case, and thus the quantum frequencies, ν, can be calculated. The corresponding absorption coefficients, μ, are calculated by extrapolating the relation which holds approximately in the gamma-ray region. The theoretical coefficients for the four groups are ·796, ·241, ·142, and ·075 m^{-1} H_2O, respectively, and are to be compared with the four values of the apparent absorption coefficients, $\bar{\mu}$, from the ionisation-depth curve (Table III), ·80, ·20, ·10, and ·03. Millikan attributes the disagreement in the last three cases to a failure on the part of the more penetrating components to reach equilibrium with their secondary electrons. This, as we have seen (§ 20), would cause $\bar{\mu}$ to be less than μ, but the difficulty of making a quantitative check on this point and the impracticability of any extrapolation of the $\nu - \mu$ relation have delayed acceptance of the theory.

C. T. R. Wilson[15] has suggested that part at least of the radiation may be of terrestrial origin and come from thunderclouds. A β particle of radioactive origin emitted upwards in the strong field within the electrified cloud would, he has shown, rapidly increase in energy, for the gain of energy from the electric field would exceed the loss of energy by ionisation along its path. Since he and others consider the electric field within the cloud to be in general directed downwards, a thundercloud should spray upwards a stream of " runaway " electrons of energy up to 5×10^9 *e*-volts (see § 42). These would either be bent down in the earth's magnetic field to strike the earth at distant points or would be converted by direct hits with atomic nuclei into ultra-gamma quanta. Attempts to find direct evidence of this effect are at present being made.

A suggestion that the dissolution of matter into radiation may be responsible for the more penetrating components of the rays has recently been supported by the argument that the smallest absorption coefficient found by Regener (·0205 m^{-1} H_2O) corresponds on certain reasonable assumptions (Jeans[17]) to the value calculated for the radiation arising from the complete dissolution of a particle within the nucleus of a heavy atom (·020 m^{-1})

Some workers consider that apparent contradictions in the results of experiments on the radiation arise from the fact that it consists of fast-moving "neutrons," proton-electron combinations endowed with no resultant charge but having a magnetic moment.

REFERENCES TO PAPERS

[1] Hoffmann, *Ann. d. Physik*, 80, p. 779 (1926).
[2] Schonland, *Proc. Roy. Soc.*, A, 130, p. 37 (1930).
[3] Hoffmann, *Gerland's Beit. z. Geophysik*, 20, 12 (1928).
[4] Millikan and Cameron, *Phys. Rev.*, 37, p. 235 (1931).
[5] Regener, *Naturwiss.*, 11, p. 183 (1929): *Zeit. f. Physik*, 74, p. 433 (1932).
[6] Steinke, *Phys. Zeit.*, 22, p. 1019 (1930).
[7] Skobelzyn, *Zeit. f. Physik*, 54, p. 686 (1929).
[8] Bothe, *ibid.*, 59, p. 1 (1929).
[9] Bothe and Kolhörster, *ibid.*, 56, p. 751 (1929).
[10] Rossi, *ibid.*, 68, p. 64 (1931).
[11] Mott-Smith, *Phys. Rev.*, 39, p. 397 (1932).
[12] Millikan, *ibid.*, 36, p. 1595 (1930).
[13] Gray, *Proc. Roy. Soc.*, A, 122, p. 647 (1929); Kulenkampff, *Phys. Zeit.*, 30, p. 561 (1929).
[14] Corlin, *Naturwiss.*, Jan. 9 (1931); Steinke, *Zeit. f. Physik*, 64, p. 48 (1930); Hoffmann, *ibid.*, 69, p. 703 (1931).
[15] C. T. R. Wilson, *Proc. Camb. Phil. Soc.*, vol. 22, p. 534 (1925).
[16] Millikan, *Phys. Rev.*, 39, p. 391 (1932).
[17] Jeans, *Nature* (April 18, 1931).
[18] Millikan and Anderson, *Phys. Rev.*, 40, p. 325 (1932).

CHAPTER III

ELECTRIC FIELDS AND ELECTRIC CURRENTS IN THE ATMOSPHERE

§ 22. THE FINE-WEATHER FIELD

IT was first found in the year 1752 by Lemonnier that the air above the earth is the seat of a persistent electric field during fine weather. The direction of this field shows that the earth carries a negative charge and the upper layers of the air a positive one. The state of affairs close to a flat portion of the surface of the earth can be expressed in three ways: (*a*) the earth carries a charge of surface density σ per unit area; (*b*) there is a vertical field of strength $F = 4\pi\sigma$, just above the ground; or (*c*) between two horizontal planes close to the ground there is a difference of potential

$$dV = V_{h+dh} - V_h = - Fdh = - 4\pi\sigma dh,$$

where $h + dh$ and h are the heights of the planes.

The quantity dV/dh is called the potential gradient and is positive, since σ is found to be negative. For the same reason, the field F is directed downwards, and this downward direction is by convention adopted as the positive direction for electric fields in the atmosphere. In these definitions it is assumed that the portion of the ground considered is flat and far removed from projections, such as trees and buildings, which would disturb the distribution of charge and concentrate the lines of force at certain points.

The average value of the fine-weather potential gradient is about 100 volts per metre; the corresponding value for the average charge density is $2{\cdot}7 \times 10^{-4}$ electrostatic units per square centimetre or ·0009 coulomb per square kilometre. The total fine-weather charge on the earth is of the order of 500,000 coulombs. Measurements of the fine-weather field are made by determining either σ or dV/dh; they are always expressed in terms of the latter, in volts per metre.

§ 23. DIRECT MEASUREMENT OF THE SURFACE DENSITY OF THE EARTH'S CHARGE

The charge on such a limited portion of the ground as can be isolated for the determination of σ is so small that a sensitive electrometer must be used to measure it. Fig. 10 illustrates the "Universal portable electrometer" devised by C. T. R. Wilson for this purpose. A flat circular plate P, called the test-plate, is surrounded by a guard-ring and mounted flush with the surface of the earth but insulated from it. This plate is joined to a gold-leaf electroscope G, whose leaf moves inside a positively charged case and is observed through a telemicroscope. C is a variable cylindrical condenser, called a compensator, with its inner plate joined to the gold-leaf system, and its outer plate maintained at a constant potential, $-$ V.

To make a measurement, the instrument is initially shielded from the earth's field by placing over it the earthed metal cover Q, indicated by the dotted lines. The reading of the leaf corresponding to earth or zero potential is found by momentarily earthing the plate P and the gold-leaf system with the key K. This earthing key is then withdrawn, the cover removed and the plate exposed to the field, with the compensator capacity zero. A negative charge thus appears by induction upon P, an equal positive charge is set free upon the central system, and the leaf moves inwards. It can be restored to its zero position if a negative charge is induced on the central system by increasing the capacity of the compensator to a value C′. The charge on the test-plate is then − C′V and the surface density, σ, is − C′V/A, where A is the area of the test-plate.

In its original form the apparatus is small and portable,

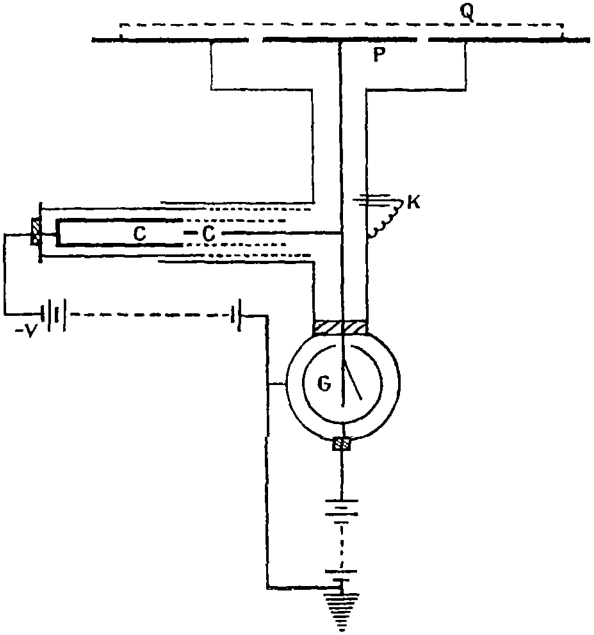

Fig. 10.—Universal portable electrometer (Wilson).

and the two batteries shown in the figure are replaced by small Leyden jars of silvered quartz, which keep their charge for a considerable time. For absolute measurements, with the plate mounted at ground level, a pit must be provided to house the observer ; very often relative measurements are made with the instrument mounted on a tripod, and the factor necessary to reduce the observations to standard conditions determined from simultaneous observations in a pit or with a potential gradient method. Since the lines of force are concentrated upon any projection above the earth, the factor is greater than unity; in the case of a tripod 1 metre high it amounts to 3 or 4. A useful method of calibrating the instrument is to place it in artificial electric fields of known strength. Two large flat metal sheets, one flush with, and the other above, the test-plate, are charged to a measured difference of potential by means of a battery of small cells.[1]

§ 24. MEASUREMENT OF THE POTENTIAL GRADIENT

The potential gradient dV/dh may be found by measuring the difference of potential between two insulated conductors at different heights above the ground. Various forms of electrometer are employed, and one conductor may be the earth itself, while the other is a stretched horizontal wire about a metre above ground-level. Sometimes two wires at different heights are used. It is of course essential that the supports of the wires and the presence of the observer and his instruments should not materially alter the field to be measured. In the arrangement in Fig. 11 (*a*), the wire

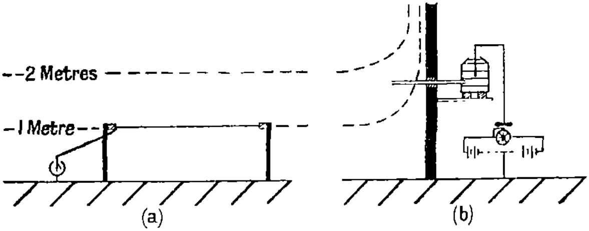

FIG. 11.—Measurement of the potential gradient.
(*a*) Absolute determination in open field.
(*b*) Continuous recording with the Kelvin water-dropper.

is hung between amber or sulphur insulators at the ends of two vertical rods; its length should exceed twice its height above ground.

The wire when first set up is at a different potential to that of the air in its neighbourhood, and a current will flow between them which tends to remove this potential difference. The ordinary conductivity of the air is, however, too small for this process of equalisation to take place very rapidly and the electrometer reading will always lag behind a fluctuating field. It is necessary for this reason to place at the centre of the wire some active ionising agent which speeds up the process by increasing the local conductivity of the air. This may take the form of a glowing fuse of filter-paper, impregnated with a 5 per cent. solution of lead nitrate, or of a disc or spiral of metal coated with a deposit of ionium or radio-thorium. Such "collectors," as they are called, have the advantage of causing the wire and the electrometer to acquire the potential of the air at the centre of the stretched length, where the electric field is least disturbed by supports and observer. The time taken for equalisation of potential depends upon the type and activity of the collector and varies from 1 to 60 seconds. There is always some uncertainty as to the exact position at which the ionising agent exerts its effect, especially if a wind is blowing, but this can be avoided by making observations at different heights.

§ 25. CONTINUOUS RECORDING OF THE POTENTIAL GRADIENT

An adaptation of the arrangement just described is employed to obtain continuous records of the potential gradient, the electrometer being housed in a building and the collector carried on an insulated rod projecting from the wall. The earth's field is very much distorted by the presence of the building, and a reduction factor for the installation must occasionally be determined by making simultaneous measurements in an open field. Fig. 11 (*b*) shows the nature of the distortion of the equipotential surfaces due to a building.

The Kelvin water-dropper is still largely used at recording stations as a potential equaliser. A jet of water from an insulated cistern within the building escapes from the end of a pipe passing through and insulated from the wall, and breaks up into fine drops at the point where the potential is required (Fig. 11). On starting the jet, the potential of the insulated system will in general be somewhat lower than that of the air at the end of the pipe ; this end will carry an induced negative charge and the cistern a positive one. Each drop carries a negative charge away and leaves the system slightly higher in potential than before. Ultimately the negative charge at the end of the pipe disappears and the system is at the potential of the air at the point at which the drops break away. In practice this takes a time of the order of 30 seconds.

Radioactive collectors are also employed, but whatever the type of collector, the insulated system is usually joined to the needle of a quadrant electrometer of low sensitivity, and opposite pairs of quadrants connected to a battery whose centre is earthed. The record can be made intermittently, by means of a clockwork arrangement which depresses an inked pointer attached to the needle (Benndorf), or continuously, by receiving the spot of light from a small mirror on a moving strip of bromide paper.

§ 26. VARIATION OF THE POTENTIAL GRADIENT WITH HEIGHT ABOVE THE GROUND : SPACE-CHARGE

A good many observations of the potential gradient have been made with balloons fitted with collectors of the radioactive or the glowing fuse pattern ; they all indicate a rapid decrease in the field with increasing height above the ground (Table V, § 33). Though no great accuracy can be claimed for these measurements —for they are subject to errors arising from the distortion of the field by free and induced charges on the balloon and are but momentary observations of a quantity subject to considerable variation—it is well established that at a height of 10 kilometres the field has fallen to less than 1/50 of its value at ground level and is diminishing still further. Some recent experiments of Idrac,[2] in which a thermionic valve was employed as a voltmeter, suggest that at very great heights the gradient may sometimes be reversed in direction, but the significance of these results is not yet clear.

This rapid diminution in the potential gradient with height indicates that a free positive charge, practically equal to the negative charge on the surface of the earth, resides within the lower 10 or 15 kilometres of the atmosphere. Integration of the values of dV/dh obtained in balloon ascents gives a total potential difference between the earth and the air at a height of 15 kilometres of about 3×10^5 volts; at this height, and above it, the conductivity of the air is so great that this figure may be taken to represent the potential of the highly conducting "Kennelly-Heaviside layer" some 80 kilometres above the ground.

The density, ρ, of the positive "space-charge" referred to can be determined from Poisson's equation, which, if the lines of force are vertical, takes the form $\frac{d}{dh}(dV/dh) = -4\pi\rho$. Several investigators have endeavoured to examine the nature of the space-charge quite close to the ground, using collectors at different heights. The results obtained are conflicting as to the magnitude, and even the sign, of ρ in this region; evidently local conditions are of main importance.

§ 27. VARIATIONS OF THE POTENTIAL GRADIENT WITH LOCALITY, TIME AND SEASON

Before passing on to the more important variations of the potential gradient, mention should be made of the fact that incessant fluctuations are observed over short periods of time; these are due to local changes in the space-charge and the conductivity of the air in the vicinity of the collector.

For many years continuous records of the potential gradient have been taken at certain stations in the Northern Hemisphere. Less extensive information is available for the Southern Hemisphere, but observations have been made over considerable periods in the Arctic and Antarctic regions. Some very important measurements over the oceans have been carried out by the survey ship *Carnegie* of the Carnegie Institute of Washington. It appears that the sign of the potential gradient in fine weather is positive all over the earth's surface, but that over land areas its value varies considerably with local conditions. The mean of dV/dh at Kew is 317 volts/metre, while that at Davos in Switzerland is but 64 v./m. Over the oceans a value of 126 v./m., which varies little if at all with geographical position, has been found. The average value for the whole earth is not far from 120 v./m.

It would seem that the potential gradient observed on land is not a quantity of much fundamental significance, its value depending primarily upon local conditions of atmospheric conductivity. This is evident from the fact that a simple periodic variation according to *local* time is the basis of all fine-weather land records. The minimum values are obtained in the early morning at about 4 a.m. and the maxima in the evening between 6 and 8 p m. In many places a second maximum at 8 a.m. and a second minimun at midday are found. The amplitude of these daily variations at land stations sometimes reaches 50 per cent. of the mean value for the day.

The local time variation has been shown by Whipple[3] to be closely correlated with the amount and distribution of the smoke pollution of the atmosphere near large cities, where these continuous records have generally been made. Fig. 12 shows the variation at Kew of the potential gradient during the summer months for the two periods 1898-1915 and 1916-28, together with the variation in the amount of pollution (determined by aspirating the air through filter-paper) for the period 1921-28. The figure shows not only the close relation which exists between potential gradient and pollution but also the striking effect of the introduction of " summer time " into Great Britain in the year 1916, after which the morning minimum and maximum moved back by approximately one real hour. The atmospheric pollution curve has maxima and minima which have been explained by Simpson as due to the combined effect of variations in the amount of smoke produced in the city and variations in the stability of the atmosphere, that is in the mixing action of surface winds and general turbulence of the lower air.

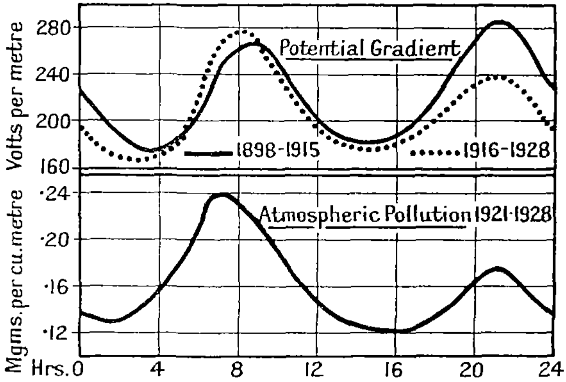

FIG. 12.—Relation between potential gradient and atmospheric pollution (Whipple).

As regards the annual variation of the potential gradient, land stations in both hemispheres show a maximum in the local winter and a minimum in the local summer. The only significant exception is the Antarctic region, where a reversal of phase occurs, giving rise to a maximum in the local summer and a minimum in the winter.

In view of the local origin of the variations of the gradient found at land stations near towns, observations over the oceans, where no pronounced local effects are to be expected, are of very great importance. Mauchly[4] has analysed the measurements of the *Carnegie* cruises and found that there is a well-marked diurnal variation at sea, with maxima and minima occurring at the same moment in all parts of the globe. Thus the potential gradient in all oceans is found to be about 15½ per cent. below the mean at 5 hours, Greenwich Mean Time, and about 20 per cent. above it at 19 hours, G.M.T. It is remarkable that the same 24-hour wave, progressing according to universal solar time, has been observed by expeditions to the Arctic and Antarctic regions and to Lapland. Fig. 13 (*a*) (after Whipple) illustrates these results, which have been confirmed by the last cruise of the *Carnegie*. In each case shown in the figure, the mean potential gradient has been reduced to 100 v./m. No appreciable annual variation has so far been observed over the sea.

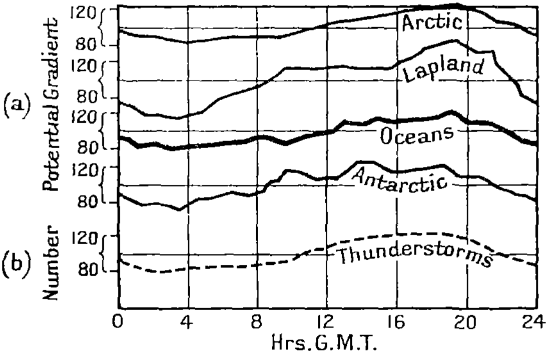

FIG. 13.

(*a*) Variation of potential gradient with universal time.
(*b*) Variation of thunderstorm activity over the earth's surface (Whipple).

Mauchly's discovery of a variation of potential gradient according to universal time is generally accepted as being of fundamental significance. The variation must clearly be ascribed to an alteration in the difference of potential between the upper conducting layers of the atmosphere and the earth ; it affords an important means of testing any theory of the origin and maintenance of this potential difference.

§ 28. THE FINE-WEATHER CURRENT : CONDUCTION AND CONVECTION CURRENTS

In the ordinary fine-weather field positive ions are driven towards the earth and negative ions away from it; the motion of these ions constitutes a downwardly directed conduction current. We have seen (§ 3) that the conduction current passing through 1 sq. cm. of a horizontal plane is given by $I_0 = F(\lambda_+ + \lambda_-)$, where F is the strength of the electric field and λ_+ and λ_- are the polar conductivities of the air at the point in question.

There is, however, a fine-weather current of a different nature which plays a part in the transfer of electricity by the atmosphere. If the air should contain at any point an excess of ions of one sign (a space-charge), movement due to wind or ordinary turbulence will give rise to a mechanical transference of electric charge. Thus if v is the upward vertical component of the velocity of the air and ρ the space-charge per c.c., the upward convection current due to this cause will be $v\rho$ per square centimetre. The real current-density in fine weather is thus the resultant of the downward conduction and the upward convection currents, and is given by $I = I_0 - v\rho$.

The mean value of the total fine-weather current is not far from 2×10^{-16} amperes per square centimetre or 2 microamperes per square kilometre, so that the total current flowing in this way between the upper atmosphere and the whole earth is about 1000 amperes. It would seem, from measurements which have been made in various parts of the world, that this current varies much less with changes in geographical position, in time of day and season of the year, than does the potential gradient. Although in the past it has not received a great deal of attention, it would appear to be a more fundamental quantity than the potential gradient and less affected by purely local conditions.

§ 29. Direct Determination of the Conduction Current at the Surface of the Earth

Direct measurements of the magnitude of the conduction current which flows into the earth's surface have been made with the Wilson universal electrometer, with the test-plate mounted flush with the surface of the ground, as described in § 23. The curve of Fig. 14 illustrates the procedure followed in the measurements. At time t_1 the test-plate is exposed to the earth's field by removing the cover, and the capacity of the compensating condenser is adjusted to the value C necessary to bring the gold leaf back to the zero position. The

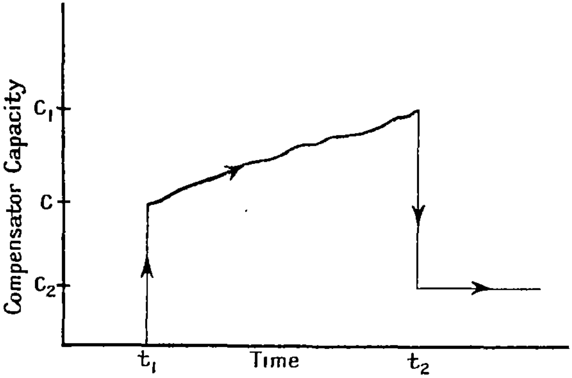

Fig. 14.—Measurement of conduction current with the Wilson universal electrometer.

charge on the plate is then $-$ CV and the field strength $F = 4\pi\frac{CV}{A}$, where, as before, A is the area of the plate and $-$ V the potential of the outside of the compensator. The test-plate is now left exposed until time t_2 and kept continuously at earth potential by increasing the compensator capacity to balance the charge received from the positive ions driven from the air on to the plate. Owing to variations in the strength of the field, and so in the induced charge on the plate, this increase is not a regular one. At time t_2 the plate is again shielded with the cover and the compensator capacity is reduced from C_1 to C_2 to maintain the shielded system at earth potential. At this moment the field strength is $4\pi(C_1 - C_2)V/A$. Since at the beginning of the measurement the compensator capacity was zero, it is evident that during the interval, $t_2 - t_1$, the plate received a positive charge C_2V and the average value of the conduction current per square centimetre was $C_2V/A(t_2 - t_1)$. For routine observations the instrument is mounted on a tripod above the ground, and the readings are reduced to standard conditions by occasional determinations of the reduction factor, from observations in a pit.

§ 30. Indirect Determination of the Conduction Current at a Point in the Air

The conduction current at a point in the air can also be determined by separate measurements of the field strength, F, and the polar conductivities, λ_+ and λ_-. The methods used to determine F, or the potential gradient, have already been described in this chapter; the conductivities are usually found by the Gerdien method sketched in § 2. In another form of the Gerdien method, due to Schering, the cylindrical condenser is replaced by a large earthed wire cage, along the axis of which is a charged insulated wire connected to an electrometer. The rate at which the charge is dissipated by ions of opposite sign is determined in exactly the same manner as in the Gerdien method, and the same equations apply if certain conditions are fulfilled. The cage must be large enough, and the charge on the wire small enough, to ensure that no considerable alteration in the number of ions is caused by the flow of current. The arrangement must be set up in the open air to ensure that ordinary atmospheric circulation prevents the development of the electrode space-charges to be described in the next section. And finally the whole cage must be shielded from the earth's field by a rough roof or by trees, otherwise it will carry an induced negative charge on the outside and rob the air entering it of negative ions.

§ 31. THE COMPARISON OF THE DIRECT AND INDIRECT METHODS: ELECTRODE SPACE-CHARGE

The direct method due to Wilson measures the conduction current entering the earth; the indirect method measures the current at a certain height above the earth. In the latter case the current is carried by two oppositely moving streams of ions of opposite sign; in the former it consists of a single stream, for while positive ions enter the earth, negative ions do not pass from it into the atmosphere. The two methods, therefore, do not measure the same thing. One determines the true conduction current at a height, and the other its positive component at ground-level.

Consider a vertical cylinder of perfectly still air whose height is h and cross-section 1 sq. cm., with its base on the ground (Fig. 15). Let F' and F be the electric field strengths at top and bottom of this cylinder, and λ'_+, λ'_- and λ_+, λ_- the polar conductivities. The conduction current through the top is $F'(\lambda'_+ + \lambda'_-)$; a positive charge $F'\lambda'_+$ enters, and a negative charge $F'\lambda'_-$ leaves, in each second. Through the bottom there is only a downward flow of positive ions, carrying away a positive charge $F\lambda_+$ per second. The cylinder as a whole thus loses a positive charge $F\lambda_+ - F'\lambda'_+$ per second, and a negative charge $F'\lambda'_-$ per second. This may be expressed as a development of a positive space-charge at a rate $F'\lambda'_- - (F\lambda_+ - F'\lambda'_+) = F'(\lambda'_+ + \lambda'_-) - F\lambda_+$ per second. The growth of this "electrode" space-charge will automatically decrease the field F′ and will cease only when F′ has such a value that as many negative as positive ions leave the cylinder in each second, i.e. when

$$F'(\lambda'_+ + \lambda'_-) = F\lambda_+ \quad . \quad (10)$$

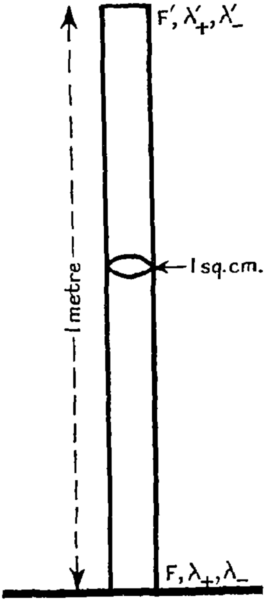

FIG. 15.—Electrode space-charge.

The two sides of this last equation are respectively the conduction currents at height *h* and at ground level. Thus, for perfectly still air, the two methods we have described should give the same numerical values, though they measure different quantities. It appears probable that this conclusion is not realised in practice and that the Wilson method yields much smaller values than the other. Some average figures from the more extensive series of measurements are shown in Table IV.

Still another consequence of the argument we have considered fails to appear in practice. Watson [1] has shown that during fine weather at Kew there is no sign of any electrode space-charge within the first metre above the ground, and that F′ and F are the same.

As there can be no doubt that a space-charge would develop if the air were perfectly still, it has been

TABLE IV

MEASUREMENTS OF THE AIR-EARTH CONDUCTION CURRENT

Indirect Method.		Direct Method.	
Station.	Current (amps./sq. cm.).	Station.	Current (amps /sq. cm.).
Potsdam	$2{\cdot}2\ 10^{-16}$	Kew	$1{\cdot}1\ 10^{-16}$
Davos	2·0 ,,	Munich (Lutz)	1·0 ,,
Gottingen	2·7 ,,		

suggested that ordinary atmospheric turbulence, by continually mixing the lower layers of the air, carries the excess positive charge upwards as rapidly as it is formed. If $F' = F$ and $\lambda'_+ = \lambda_+$, the positive charge developed in the cylinder per second is $F'\lambda'_-$. This must be removed by the air at such a rate that an upward connection current $v\rho = F'\lambda'_-$ per sq. cm. crosses the top of the cylinder, and the total resultant air-earth current at the height h is

$$I = I_0 - v\rho = F'(\lambda'_+ + \lambda'_-) - F'\lambda'_- = F'\lambda'_+ \simeq F\lambda_+.$$

If this view is correct, the quantity $F\lambda_+$, which is the portion of the conduction current measured by the Wilson method, is actually equal to the total current, conductive and convective, passing between the atmosphere and the earth. While this result may not hold in practice with great exactitude, the arguments outlined above indicate that the test-plate method gives a closer approximation to the total fine-weather current than the indirect method which determines the conduction current only. The test-plate result is the more valuable because it appears that any direct estimate of the convection current by itself is very difficult.

§ 32. VARIATIONS IN THE FINE-WEATHER CURRENT

Both direct and indirect methods of studying the air-earth current indicate that in fairly fine weather it is less subject to variations than is the potential gradient. Simultaneous measurements of the field and of the total conductivity of the air, $\lambda = \lambda_+ + \lambda_-$, show that these quantities vary to some extent in inverse ratio, with the result that their product, the conduction current, is less variable than either. Factors such as an increase in the number of condensation nuclei or the presence of fog particles, which decrease λ by raising the proportion of the slow ions present, cause a corresponding increase in the potential gradient. The simple form of the daily variation in the earth's field has its counterpart in an inverse variation of the conductivity, with a maximum in the early morning between 3 and 4 a.m. The extreme effect of atmospheric pollution near large cities, which increases the potential gradient and diminishes the conductivity, has already been discussed (§ 27). In general the reciprocal relation between λ and F is not exact enough to keep the conduction current invariable. At land stations in Europe it is found to have an annual variation which coincides in phase with that of the potential gradient but has a smaller amplitude. At most stations its daily variation follows the changes in the conductivity.

As shown by Table IV the mean values obtained at European stations for the conduction current, using the indirect method, all lie close to $2{\cdot}0 \times 10^{-16}$ amps./sq. cm. The observations over the sea and near the poles give about double this value. Since, as we have seen, to make allowance for the upward convection current near the ground these figures must probably be reduced by half, the total air-earth current in fine weather may be taken to have an average value of 2×10^{-16} amps./sq. cm. over the whole globe.

§ 33. THE CONDUCTION CURRENT AT GREAT HEIGHTS

It has been found from simultaneous determinations of potential gradient and atmospheric conductivity during balloon ascents that the rapid decrease in F with height above the ground is accompanied by an increase in λ, of such magnitude as to keep the conduction current approximately constant. Table V shows the results of an ascent by Wigand,[5] in which it will be seen that when the field had fallen to 1/14 of its value at the ground, the conductivity had risen eleven-fold.

TABLE V

CONDUCTION CURRENT DURING BALLOON ASCENT (WIGAND)

Height (metres).	F (volts/metre).	λ (e.s.u.).	I (amps/sq. cm.).
0	136	1·1 10^{-4}	1·7 10^{-16}
2500	27	4·8 ,,	1·4 ,,
4400	18	8·2 ,,	1·6 ,,
6500	8·8	12·6 ,,	1·2 ,,

§ 34. THE SIGNIFICANCE OF THE AIR-EARTH CURRENT

The general constancy of the air-earth current with height above the ground and in different parts of the world suggests that, at any rate to a first approximation, it has its origin in a constant potential difference between the conducting layers of the upper atmosphere and the earth, and that its actual value depends upon the resistance of the air between the two regions. Now the conductivity of the upper portion of this air-path is entirely due to the penetrating radiation * and is unlikely to alter at all with time of day, locality or season (§§ 20, 21). But below a height of about 2 kilometres over land

---

* This refers to the regions from 15 to 30 kilometres above the ground. The Heaviside layer 80 kilometres up has a solar origin.

areas, an additional ionising influence is exercised by radioactive matter in the air and the earth, which, as we have seen (§ 5), varies very much. Moreover, the conductivity of the air in this region is dependent upon the relative concentrations of the large and the small ions, and these alter with the number of nuclei present. The current-bearing column of air above the point of observation thus consists of an upper constant resistance and a lower variable one in series with each other. The extent to which the current carried by the air-column alters with local conditions will depend on the ratio which the lower variable resistance bears to the total resistance.

The conductivity above 14 kilometres is so high that we may, for the sake of illustration, suppose that this is the actual height of the upper positively charged layer and take the mean specific resistance of the air between it and the earth to be $7{\cdot}1 \times 10^{14}$ ohms-cm., the value found by Wigand at a height of 6·5 kilometres. Then the total resistance of an air-column of cross-section 1 sq. cm. would be 10^{21} ohms. If the variable part of the column be the first 100 metres above the earth, where the specific resistance is usually 8×10^{15} ohm-cm., this portion will contribute 8×10^{19} ohms to the total. We see that even if the conductivity of the first 100 metres were halved, the effect upon the air-earth current would amount to but 4 per cent. The potential gradient near the ground, on the other hand, would be doubled, for from this point of view it represents the fall of potential per unit length of the lowest part of the current-bearing column. It is evident that potential gradient measurements over land, and in particular near cities, cannot be expected to yield information which bears directly upon the question of the total potential difference between upper atmosphere and earth. In the case of observations at sea and in polar regions the position is different; practically all the ionisation throughout the whole column is then due to the penetrating radiation, and the only local effect which can influence the fine-weather gradient is a change in the number of nuclei available for the capture of small ions.

§ 35. The Electric Field during Disturbed Weather

The normal downwardly directed field, due to a positively charged upper atmosphere and a negatively charged earth, is frequently disturbed when the weather ceases to be fine. During fog, as explained in § 32, it is very much increased, and may reach ten times its normal value ; during dust-storms, of the type common in semi-arid regions and deserts, powerful reversed or negative fields are usual and reach 10,000 volts per metre. The effects found during cloudy weather and rain are variable, and range from gradients of the order of a few hundred volts per metre in a fine drizzle to as much as 50,000 volts per metre under thunderclouds.

Generally speaking, negative fields predominate when light or steady rain is falling, though occasional positive excursions are observed. In the case of heavy rainstorms and thunderclouds, the sign of the field depends upon the portion of the cloud passing over the point of observation, but here, too, there is evidence that negative gradients are the more frequent.

The field during thunderstorms fluctuates a great deal if the cloud is active in producing lightning. Its value at the ground does not as a rule exceed 10,000 to 20,000 volts per metre except for very brief intervals occupying fractions of a second. Equally high fields are observed during heavy rainstorms unaccompanied by lightning. During a snowstorm the field is usually positive and in a heavy storm may attain 10,000 v./m. The methods employed for the study of the large and rapidly varying fields of thunderclouds are described in Chapter IV.

§ 36. The Charge on Rain and Snow ; the Precipitation Current

The apparatus required to examine the charge carried to the earth by precipitation—rain, hail, snow, and sleet—is comparatively simple, and consists essentially of an insulated metal vessel to catch the rain, etc., and an electrometer to measure the charge received. Great precautions must, however, be taken to eliminate the disturbing effect of the negative charge given to the air as the drops splash on the container (Lenard effect, § 43). If some of this is lost a spurious positive charge will be indicated. For a similar reason, the entry of spray due to the splashing of rain falling outside the receiving vessel must be prevented. The receiver must be screened electrostatically from the electric field outside. The rate of rainfall, and sometimes the average size of the drops, is also observed ; with sensitive forms of the apparatus the charges carried by individual drops may be measured. As the charges brought down vary considerably, it is necessary in all such work to carry out a long series of observations.

Since the year 1908, when the modern series of observations began, all observers agree in finding a preponderance of positive charge on rain in general. The average ratio of the quantities of positive and negative electricity brought down by all kinds of precipitation varies according to the situation of the observer. Some of the values found for this ratio are : Potsdam (Schindelhauer), 1·4 : 1 ; Puy-en-Velay (Baldit), 2·4 : 1 ; Dublin (McClelland and Nolan), 4·8 : 1 ; Simla (Simpson), 2·4 : 1 ; Otago, New Zealand (Marwick), 3·2 : 1. As exceptions to the general rule, it may be noted that snow usually carries a considerable negative charge, and that the fine drops of a drizzle are also negative. Simpson showed in 1908 that thunderstorm rain—the most heavily charged of all forms—is positively charged when it falls from the front of the cloud and negatively charged elsewhere. Ordinary steady rain, unaccompanied by sudden changes in the barometer, shows the most regular as well as the highest excess of positive charge. If such rainfall alone is considered, Baldit's average ratio of 2·4 : 1 for all types of precipitation rises to 4·3 : 1.

The charge per c.c. in the case of heavy thunderstorm rains reaches average values of 30 to 40 e.s.u., while single drops may carry 100 to 200 e.s.u. (Gschwend) and attain potentials as great as 300 volts. The mean values reported by different observers for the nett charge per c.c. on all forms of precipitation during a long series of observations range from + ·029 e.s.u. at Potsdam (Schindelhauer) to + ·176 e.s.u. at Simla (Simpson).

The transfer of electricity from atmosphere to earth in this manner constitutes a "precipitation" current which appears on the whole to be in the same direction as the fine-weather conduction current, since it carries more positive than negative charge to the earth. In the case of thunderstorms this current sometimes reaches values as high as 2×10^{-11} amps./sq. cm., and the total convection current carried by the rain in such a storm may amount to more than 0·1 amp. Before it is possible to frame a trustworthy estimate of the amount of the excess of positive over negative electricity conveyed to the whole earth in this way, much more information is required from tropical and equatorial regions.

§ 37. CURRENTS DUE TO POINT-DISCHARGE

It is well known that an electric glow or brush discharge takes place from a pointed conductor when placed in a sufficiently strong field. It consists of a stream of ions of the same sign as the charge upon the point. Discharges of this kind must, therefore, be expected from the ends of exposed conductors on the surface of the earth whenever the potential gradient and the height and sharpness of the ends are sufficiently great. Indeed the earliest investigators, Franklin with his kite, Dalibard and Lemonnier with their high metal rods, detected electric fields, the first two in thundery, and the last in fine weather, by means of this current. Under favourable conditions, afforded by mountain peaks or the masts of ships exposed to the intense fields of a thundercloud, the glow discharge may become very conspicuous (St. Elmo's fire).

It was first pointed out by C. T. R. Wilson [6] that point-discharge currents of much smaller intensity must play an important part in the interchange of electricity between the earth and the atmosphere. The fields prevailing under rainstorms and thunderclouds very frequently reach values sufficiently great for considerable currents to be discharged from exposed conductors, such as trees, bushes, housetops and even fields of grass. A conductor need not end in a sharp point or project to a great height in order that it should begin to act as a discharger. For example, an earth-connected sphere of radius 1 cm. need only be raised to a height of 3 metres in the not unusual field of 10,000 volts per metre for the electric intensity at its surface to reach the critical value of 30,000 volts per cm. at which brush discharge begins. A practical instance of the correctness of this suggestion is afforded by some experiments made by the author, in which a small tree, 4 metres high, was cut off at its base, mounted on insulators and connected to earth through a galvanometer. Exposure to the fields due to nearby thunderstorms yielded the average point-discharge currents shown below:—[7]

TABLE VI

POINT-DISCHARGE CURRENTS FROM A SMALL TREE (SCHONLAND)

Field (Volts/metre).	Current (Microamperes).
– 3,500	0·07
– 5,500	0·20
– 11,000	1·00
– 16,000	4·00

The contribution of a single tree such as this must be multiplied by a very large factor to obtain the total current flowing between the earth and a thundercloud; in the case quoted, the tree was typical of the exposed natural conductors for many miles around, and a rough surface integration over the area affected by the cloud indicated that the total point-discharge effect was of the order of 2 amperes in an upward direction.

Since the direction and magnitude of the large fields associated with disturbed weather vary considerably, it is extremely important to know the relative amounts of positive and negative electricity discharged from an exposed point over a long period of time. In a series of observations by Wormell,[8] a sharp point carried on insulators at the top of a pole was connected by cable to one electrode of a voltameter filled with dilute sulphuric acid, the other electrode being earthed. The voltameter (Fig. 16) was constructed of fine capillary tubing (diameter 0·8 mm.) and the gases evolved were separately collected, their volumes being determined from the lengths of the bubbles formed. If v_1 and v_2 are the volumes of the mixed gases collected at the earthed and the point-connected electrodes respectively, the quantities of positive and negative electricity, q_1 and q_2, which have been discharged from the point can be separately determined in the following way: If one unit

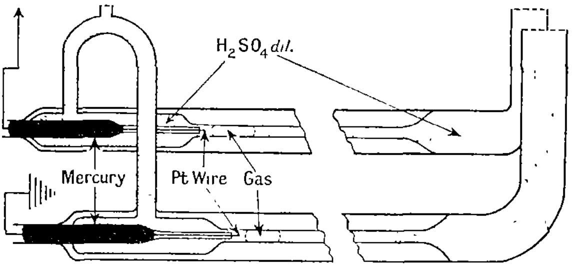

FIG. 16.—Water micro-voltameter (Wormell).

of quantity liberates in electrolysis c c.c. of oxygen and $2c$ c.c. of hydrogen, we have

$$q_1c + 2q_2c = v_1 \quad \text{and} \quad 2q_1c + q_2c = v_2,$$

from which we may determine q_1 and q_2.

The quantities discharged in this way have been examined by Wormell at Cambridge over a period of three years. Some of his results are shown in the table below :—

TABLE VII

INTEGRATED EFFECTS OF POINT-DISCHARGE (WORMELL)

Year	1927	1928.	1929.
Positive quantity discharged, q_1 (coulombs).	·28	·24	—
Negative quantity discharged, q_2 (coulombs).	·14	·11	—
Nett positive discharge, $q_1 - q_2$ (coulombs) .	·14	·13	·11

Each year shows a considerable nett loss of positive electricity by the discharger.

In a more detailed discussion it is noted that of 147 separate periods of disturbed weather accompanied by precipitation, 103 showed an excess of upward current (+ ve discharge), 34 an excess of downward current (— ve discharge), and in 10 the nett discharge was zero. This is what would be expected from the predominance of negative potential gradients in disturbed weather. It is to be hoped that similar measurements will be made in other parts of the world.

§ 38. The Interchange of Electricity Between the Earth and the Atmosphere; the Maintenance of the Earth's Charge

The negatively charged earth and the positively charged upper layers of the atmosphere form two plates of a spherical condenser with the lower air as the dielectric. Although the conductivity of the air between the plates is small, the applied potential difference is great enough to make the leakage of charge through it very considerable. The average value of the charge per unit area of the earth's surface is 9×10^{-14} coulombs/sq. cm., and that of the fine-weather conduction current is 2×10^{-16} amps. per sq. cm. Left to itself, a condenser of this kind would be discharged by internal leakage in a time of the order of ten minutes. But this is not all: to the dissipation of the earth's charge by the fine-weather current we must add the effect of the charge conveyed by rain and snow, for it has been shown that on the whole the nett charge conveyed to the earth by precipitation is positive in sign. Thus to the conduction current of 1000 amperes over the whole surface of the earth must be added a precipitation current in the same direction, estimated by Wigand[9]—though any such estimate can only be provisional—at 400 amperes, making a total of 1400 amperes tending to dissipate the 500,000 coulombs with which the earth is charged.

In spite of the continuous operation of these two factors, the earth's charge remains practically constant, so it is clear that there must exist some reverse or compensatory process which neutralises their effect. Some agency must be continuously at work replenishing the charges on the earth and in the upper air. The discovery of the mechanism responsible for this replenishment has been one of the chief problems of atmospheric electricity, and many suggestions have been made to this end. Two of these, though they have not proved acceptable, will be briefly mentioned.

The suggestion of Ebert is based upon the experimentally verified view that ionised air escaping through the narrow capillaries of the earth's crust is richer in positive than in negative ions, for the higher mobility of the negative ion causes it to diffuse more rapidly to the walls of a narrow passage. The air escaping from the earth and ionised by radioactive emanations, etc., thus gives a positive charge to the lower atmosphere, which Ebert suggested was conveyed to considerable altitudes by convection currents. It appears, however, that the Ebert effect in the lower atmosphere is actually extremely small and that the upward air-currents are quite insufficient to produce the necessary convection.

Another suggestion, made by Simpson and developed by Swann, is that negative charge is conveyed to the earth by fast-moving negatively charged corpuscles or β particles. This view meets with the difficulty that the ionisation produced in the air by the number of β particles needed to maintain the earth's charge would be very large, many times greater than that observed and known to be due to other causes. The objection can be avoided only by ascribing a very small ionising power to an extremely fast β ray (Swann). This is not in accord with what is known of the fastest β rays from radioactive substances, whose ionising powers vary very slowly with their energies.

It was first suggested by C. T. R. Wilson that it is to regions of disturbed weather that we must look for the mechanism of replenishment, that under rain and thunderclouds, where the potential gradient is more often reversed than in the normal direction, there are two processes at work which convey considerable quantities of negative charge to the earth. The first of these is the action of point-discharge from conductors projecting from the ground. This has already been discussed, and it has been shown that experimental tests indicate that it is of great importance. The second process is the charge conveyed to the ground by lightning flashes from thunderclouds. Every second, as will be seen in Chapter IV, some 100 lightning discharges occur over the whole surface of the earth, each involving the passage of a charge of the order of 20 coulombs. Only a fraction of these, perhaps one in four on the average, strike the ground. The effectiveness of this process of transference of charge will depend upon whether more charges of one sign than another are conveyed to the earth. If, for example, practically all the charges transferred were negative, the process would be equivalent to a continuous compensation current of about 500 amperes. Though there is evidence that this is actually the case, the question is still under investigation.

An estimate of the annual electrical " balance-sheet " of a square kilometre of ground at Cambridge has been made by Wormell,[8] on the basis of his point-discharge measurements and observations on lightning discharges. The air-earth conduction current loss and the effect of rain and snow were estimated from average values of these quantities. The results were :—

	Coulombs/sq. km /annum.
By natural point-discharge, gained .	−100
By lightning, gained	− 20
By atmospheric conduction, lost . .	− 60
By precipitation, lost	− 20
Nett gain of negative charge . . .	− 40

Though such an estimate is only a very rough one, it appears to be quite possible that in this locality the four processes approximately balance one another, or even that the earth gains a negative charge.

This suggestion of Wilson involves more than the interchange of electricity between the bases of cumulonimbus clouds, showers and thunderstorms, and the ground; he regards them as equally active in supplying positive electricity to the upper air. On this view the cloud acts as an electrical generator which removes positive electricity from the earth and supplies it to the conducting layers of the atmosphere above, by which it is rapidly distributed in such a way as to maintain them at a constant potential of about 3×10^5 volts, in spite of losses.

If this view is correct, we should expect the potential difference, and so the potential gradient, to show a maximum when the thundery regions of the earth are at their maximum activity.[10] The diurnal variation of the gradient discovered by Mauchly (§ 27) should agree in phase with the diurnal variation of thunderstorm activity over the earth. It has recently been shown by Whipple [11] that this appears to be the case. Whipple's examination of the diurnal variation of the world's thunderstorms is represented by the bottom curve of Fig. 13 (*b*), where it will be seen that the parallelism with the variation of the potential gradient over the oceans and the polar caps is indeed very close.

REFERENCES TO PAPERS

[1] Watson, *Met. Office Geophys. Memoir*, No. 45 (1929).
[2] Idrac, *Compt. Rend.*, 182, p. 1634 (1926).
[3] Whipple, *Quart. Journ. Roy. Met. Soc.*, 55, p. 351 (1929).
[4] Mauchly, *Terr. Mag.*, 28, p. 61 (1923). See also Hoffmann, *Beitr. Physik. Atmos.*, Leipzig, 11, p. 1 (1923).
[5] Wigand, *Phys. Zeit.*, 26, p. 81 (1925).
[6] C. T. R. Wilson, *Proc. Phys. Soc.*, 37, p. 32*d* (1925).
[7] Schonland, *Proc. Roy. Soc.*, A, 118, p. 252 (1928).
[8] Wormell, *ibid.*, A, 115, p. 443 (1927); 117, p. 567 (1929).
[9] Wigand, *Ann. d. Physik*, 66, p. 81 (1921).
[10] C. T. R. Wilson, *The Observatory*, 44 (1922).
[11] Whipple, *Quart. Journ. Roy. Met. Soc.*, 55, p. 1 (1929).

CHAPTER IV

THE ELECTRIFICATION OF THUNDERSTORMS

§ 39. The Number of Thunderstorms over the Globe

WHILE a thunderstorm is a comparatively rare event at any one station in temperate latitudes, the number of thunderstorms per day experienced by the earth as a whole is very large. Even in such a limited area as France there are comparatively few days in the year where thunder is nowhere reported. The total amount of thunderstorm activity which takes place at various seasons of the year has been determined by Brookes[1] from information taken from all parts of the world. According to his estimate, the earth experiences 16,000,000 thunderstorms per annum, or 44,000 per day. Taking one hour as the average duration of each—a modest estimate—there will on the average be 1800 thunderstorms in progress in different parts of the world at any one moment. In terms of lightning flashes this would mean that something like 100 flashes occur over the earth in every second, or 360,000 per hour.

It appears from recent measurements (§ 42) that the average thunderstorm develops and expends two or three million kilowatts of electrical energy continuously during its hour or two of activity. Over the whole earth, therefore, thunderstorms are continuously expending electrical energy at a rate of some 4×10^9 kilowatts. This figure is so large that by the employment of even a small fraction of its power in an apparently insignificant "side-line," the thunderstorm could be of importance in a number of geophysical problems. Amongst these may be mentioned the generation of Hertzian waves (atmospherics), the maintenance of the earth's negative charge and of the positive charge on the upper air (§§ 37 and 38), the production of high-speed electrons (§ 21), and of ionisation in the upper atmosphere.

§ 40. The Measurement of the Electric Fields and Field-changes due to Thunderstorms

Practically all the quantitative information on the electrification of thunderstorms which has so far been obtained has been derived from measurements of the electric fields and field-changes produced by them at the surface of the earth. The development of these measurements and the interpretation of the results obtained is mainly due to Professor C. T. R. Wilson.[2]

The general principle is similar to that of his test-plate or induced charge method of studying the fine-weather field (§ 23), in which an exposed conductor is kept at zero potential by alteration in the capacity of a compensating condenser, when the charge induced upon the conductor by the field is equal to that given by the compensator to the system. In the measurement of rapidly changing thunderstorm fields, however, compensation is produced automatically and practically instantaneously by means of an ingenious form of capillary electrometer. As shown in Fig 17, this consists of a small bubble of dilute sulphuric acid enclosed between mercury threads in a narrow capillary tube. The threads lead to end-cups of mercury, and one of these, A, is connected to a test-plate T insulated from, but flush with, the surrounding ground, while the other is connected to the earth. Between the mercury and the glass there is everywhere a thin film of dilute acid, so that an electrical "double layer" surrounds the mercury and makes each half of the electrometer a condenser charged to the potential difference of the double layer, about ·90 volt, the mercury being positively charged with respect to the acid. If a positive charge is given to the left-hand side, it will increase the charge on the left-hand condenser and the left-hand thread will move forward a certain distance so as to absorb this charge, by increasing the area of the acid-mercury surface and the capacity of the left-hand condenser. A corresponding movement of the right-hand thread causes a reduction in area on this side and the passage to earth of a quantity of electricity exactly equal to that originally supplied to the electrometer. The distance moved, y, is easily seen to be related to the charge passed through, q, by the equation

$$q = 2\pi r C V'(1 - r/R)y,$$

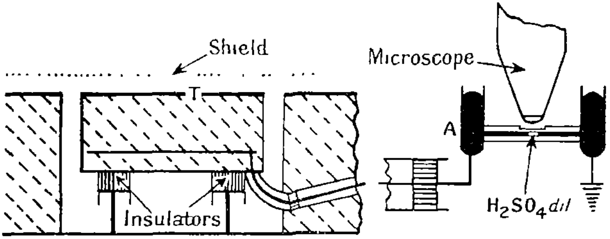

FIG. 17.—Measurement of the electric fields and field-changes due to thunderstorms (C. T. R. Wilson).

where r is the radius of the capillary and R that of the end-tube, V′ the potential difference, and C the capacity per unit area, of the double layer. Thus $q = ky$, where k is an instrumental constant which can be determined by discharging a known quantity of electricity through the instrument. The arrangement thus gives a linear relation between the movement of the acid bubble and the quantity discharged. In spite of the large value of C, about 30 microfarads per square centimetre, it is extremely sensitive. By using a microscope focussed upon one end of the bubble quantities of the order of 10^{-9} coulombs may be measured. A permanent record is best obtained by placing a slit and a moving photographic plate in the focal plane of the objective.

When connected to a test-plate the electrometer automatically maintains it at earth potential and records by its movements the quantities of electricity passing to and from the earth as a result of changes in the external field and of the induced charges on the plate. The type of record obtained from this arrangement during a thunderstorm is illustrated by the thick line of Fig. 18. An earthed cover is swung over the test-plate for a few seconds at the beginning of each minute, causing the movements A′AA″, B′BB″, etc., as the induced charge flows away to earth and then returns. The points A, B, C, E would lie on a horizontal line if there were no actual transfer of charge between the plate and the atmosphere taking place, that is to say, no conduction current, point-discharge, or transport of charge by rain. The first two minutes in the figure show the slight slope due to the conduction current only, the last two show the effect of a fall of positively charged rain. The charge, for example, conveyed to the

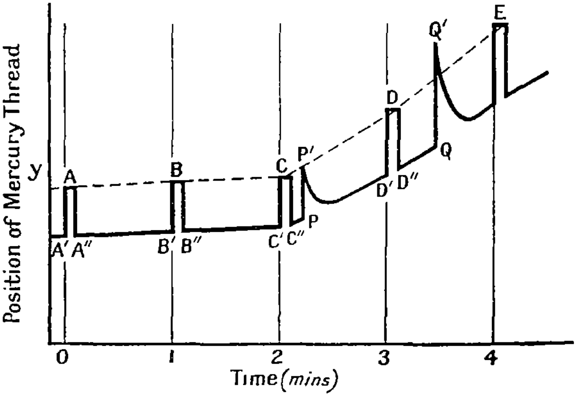

Fig. 18.—Capillary electrometer record during a thunderstorm.

plate by rain and conduction during the third minute is $k(y_D - y_C)$ and the mean value of the current per square centimetre $k(y_D - y_C)/60A$, where A is the area of the plate and k the constant of the electrometer. The dotted line joining the points A, B, C, D, E gives the reading of the instrument when the test-plate is shielded at any time during the record, and this line serves as the zero from which the field and field-changes are measured. The field at the end of the first minute, for instance, is given by $F = 4\pi k(y_B - y_{B'})/A$.

Two sudden interruptions in the record, such as are caused by lightning flashes, are shown at P and Q. At P the record suddenly touches the dotted line, indicating that for a moment the field was completely destroyed and the original induced charge ran to earth through the electrometer. At Q the movement extends beyond the dotted line, showing that the field was actually reversed in sign by the discharge and reached a reverse value equal to half its original strength. The shape of the record immediately after the passage of these lightning flashes, the recovery curve as it is called, is of considerable interest; it will be seen in these two examples that the thundercloud fields and charges recovered their original values after the lapse of something like twenty seconds. The rate of recovery was approximately exponential, being most rapid immediately after the flash and becoming very slow as the cloud approached its original electrified condition. We may find the magnitude of these sudden field-changes by the same method used for the field itself; if Δy be the movement PP′, or QQ′, the sudden field-change which caused it is given by: $\Delta F = 4\pi k \,.\, \Delta y/A$.

The test-plate arrangement, using a plate 50 cm. in diameter, is suitable for the measurement of fields of strength exceeding 1000 volts/metre, such as are caused by thunderstorms when they come within a distance of about 8 kilometres. For more distant storms Wilson employs a different form of exposed conductor, a copper ball 1 foot in diameter mounted on insulators at the end of an iron pipe 5 metres in length (Fig. 19). When the arrangement is in use the pipe is held vertical; by turning it about a horizontal axis through its base, the pipe is lowered and the ball enters an earthed metal case which shields it from the field. Connection to the capillary electrometer is made by an insulated wire passing down the centre of the pipe.

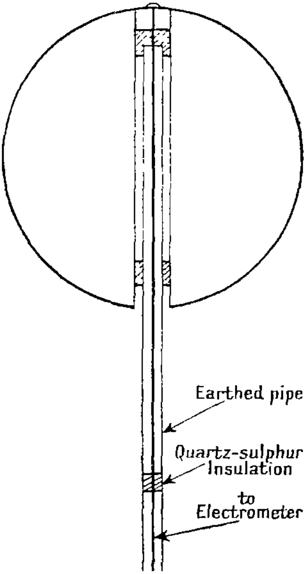

FIG. 19.—Elevated sphere for measuring fields of distant thunderstorms.

The field producing a given movement of the electrometer is calculated as follows: The charge q induced upon the ball when raised to a height h above the ground must be such as to maintain it at earth potential, for the ball is earth-connected through the electrometer. Thus if V is the undisturbed potential of the air at the point occupied by the centre of the sphere of radius r, we have $V + q/r - q/2h = 0$, the third term allowing for the effect of the electrical image of the charged sphere in the earth. If the movement of the electrometer caused by raising the ball from its case to the height h is y, we have $q = -ky$ and $V = k(1/r - 1/2h)y$. This observation therefore gives for the mean potential gradient between the ground and the height h,

$$F = V/h = (k/h)(1/r - 1/2h)y.$$

The same equation relates the sudden changes of field ΔF to the corresponding sudden movements Δy of the electrometer. Since these relations, like those derived for the test-plate, are derived on the assumption of a perfectly plane earth, small corrections must be applied to allow for the effect of the concentration of the lines of force upon the hut used to house the recording instrument.

The effects obtained from the elevated sphere become too small for accurate measurement when the storm is more than 20 kilometres away ; for still greater distances a wireless aerial has been used by some workers. Thunderstorm fields are of the same order as the fine-weather field, about 100 v./m., when the cloud is 20 kilometres off. At distances of 6 to 8 kilometres they often reach 5000 v./m., and when the storm passes overhead values of 10,000 to 20,000 v./m. are commonly observed, while the sudden changes may exceed these figures very considerably for a small fraction of a second. In an active thunderstorm the fields are constantly altering as a result of the neutralisation of part or all of the charge by lightning discharges, of the subsequent rebuilding of the destroyed charges, and of movements of the cloud as a whole.

§ 41. THE CHANGES OF FIELD ACCOMPANYING LIGHTNING DISCHARGES

The methods described have been used for the measurement of the field-changes due to lightning in England and in South Africa,[2, 3] and have led to valuable information as to the magnitude of the electrical quantities involved in the thunderstorm and the distribution of the charges in the cloud. The principles to be followed in the interpretation of the observations have been very completely discussed by Wilson.[2] Their application is not always an easy matter, for though the structure of most thunderclouds may be the same in principle, meteorological conditions lead to frequent departures from any simple model which can be framed. Anything like a full discussion of this question would occupy too much space, and so we will limit ourselves to the detailed consideration of a type of lightning discharge which takes place within the thundercloud without reaching the ground and which is the most frequent kind of flash in a tropical storm. In the Cape Province of South Africa nine-tenths of the discharges are of this nature, and simply involve the downward movement of a positive charge from a height of 4 or 5 kilometres to one of 2 or 3 kilometres.

Consider a simple cloud-model in which a positive charge Q is distributed through a spherical region with its centre at a height H_2 above the ground. To calculate the resulting downwardly directed field at the earth's surface we may regard the charge induced on the conducting earth as replaced by the image-charge — Q at a depth H_2 below the ground and we may disregard the curvature of the earth. The field at any point such as P (Fig. 20), at a horizontal distance L, is then that due to an electric doublet of moment $2QH_2$, and its strength is given by $F_2 = 2QH_2(H_2^2 + L^2)^{-3/2}$, vertically downwards. If a lightning discharge causes the positive cloud-charge to pass to a lower level H_1, the new field $F_1 = 2QH_1/(H_1^2 + L^2)^{-3/2}$ and the sudden field-change due to the discharge is

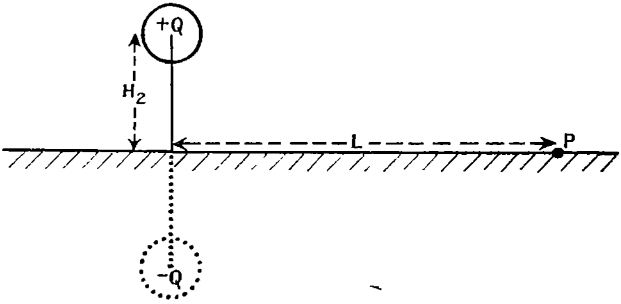

FIG. 20.—Thundercloud charge above the earth.

$$\Delta F = F_1 - F_2 = 2Q[H_1(H_1^2+L^2)^{-3/2} - H_2(H_2^2+L^2)^{-3/2}]. \quad (11)$$

It may be noted that this equation applies equally well if the act of lowering the charge Q involves the neutralisation of an equal and opposite charge $-Q$ at the height H_1, for the effect of adding the field due to this charge to both F_2 and F_1 disappears in computing the field-change ΔF.

Equation (11) shows that the change of field should be zero at a distance L_0 such that

$$H_1/(H_1^2 + L_0^2)^{3/2} = H_2/(H_2^2 + L_0^2)^{3/2},$$

that it is negative for $L > L_0$ and positive for $L < L_0$. This reversal distance L_0 is found to be 6·7 kilometres, if H_2 and H_1 are 6 and 3 kilometres respectively.

South African thunderstorms show this reversal in the sign of the majority of the field-changes very clearly. The following Table[3] gives the results of observations upon some fifty storms, giving 2800 observed field-changes which were grouped as "distant" when more than 15 kilometres from the station and as "near" when within 7 kilometres; in many cases the same storm was observed in both groups.

TABLE VIII

SUDDEN FIELD-CHANGES DUE TO THUNDERSTORMS (SCHONLAND)

Distant Storms.		Near Storms.	
Positive ΔF.	Negative ΔF.	Positive ΔF.	Negative ΔF
250	2375	188	9

The phenomenon of zero values for the field-changes, due to discharges at the reversal distance, can only be observed in isolated instances. In practice, owing to the discharge not taking a vertical but an inclined path, and for other reasons, the simple concept of a reversal point must be replaced by that of a reversal zone, in which small field-changes of either sign are equally probable, and which, in the storms of Table VIII, may extend from 7 to 15 kilometres from the point of observation.

When the distance L is great compared with the heights H_1 and H_2 of the cloud-charges, equation (11) takes an important form; it becomes

$$\Delta F = - 2Q(H_2 - H_1)/L^3 \text{ or } 2Q(H_2 - H_1) = - \Delta F \times L^3.$$

The quantity on the left is now the difference between the electric moments of the thundercloud charges before and after the discharge; it is referred to as the electric moment of the discharge. All that is needed to determine it is a measurement of the field-change due to a distant lightning flash and of the distance L, which can be found from the interval between the flash and the thunder it causes.

Measurements of the electric moments of discharges have been made both in England and South Africa, with similar results. The average value found is about 2×10^{16} e.s.u. × centimetres, or 60 coulomb-kilometres, in Europe, and 90 coulomb-kilometres in the more elevated tropical storms. Direct observation in the latter case gives for the vertical lengths of cloud-flashes, $H_2 - H_1$, values lying between 1·5 and 3 kilometres, so that the average quantity of electricity conveyed by a lightning discharge is about 20 coulombs. The usual range of variation for individual discharges is not large, from 2 to 100 coulombs if the length of the discharge is supposed constant, and probably much less if this quantity were actually observed in every case.

§ 42. MAGNITUDES OF THE ELECTRICAL QUANTITIES INVOLVED IN THUNDERSTORMS

The determination of the average quantity of electricity destroyed by a flash of lightning provides the starting-point for the evaluation of a number of important electrical quantities involved. A single flash, of the kind we have discussed, generally completely discharges

6

the whole cloud and momentarily reduces the field from it to zero. A fairly active cloud produces one flash every twenty seconds ; to feed the flashes such a storm would have to generate electricity at an average rate of 1 coulomb per second, i.e. the mechanism would have to provide an average current of 1 ampere. Reference has already been made to the rapid rate at which the initial regeneration of the cloud-charges takes place. It is usually such as to be able to restore the whole electrification in about five seconds if it continued for that time. The exponential form of the recovery curves suggests that this initial rate may remain constant, but that, as the cloud charges grow, effects occur which tend to dissipate them and which are proportional to the quantity of charge present at any time. On this view the rate of regeneration involves a current of 20/5 coulombs/sec. or 4 amperes. The electrical energy generated by the storm is thus only partially employed in the feeding of lightning flashes ; at the moment just before a discharge most of the power of the machine is being expended in overcoming various leakage effects. It may happen in this way that the field in the cloud never reaches the sparking value : this appears to be the case in many heavy rainstorms.

Before proceeding further it is necessary to know the critical value of the electric field strength at which the spark which we call lightning takes place. In air at ordinary pressure this is 30,000 volts per centimetre, but in a thundercloud the pressure is lower and the presence of small drops of water introduces a new and important factor. Macky[4] has recently investigated the behaviour of water-drops in strong electric fields, and finds that when exposed to increasing fields a drop of radius r cm. becomes elongated, until at a definite field strength, given in volts per cm. by $F\sqrt{r} = 3875$, it becomes unstable and filaments are drawn out from the ends. When instability occurs, a discharge passes through the drop in exactly the same manner as it would through a conductor pointed at both ends, with characteristic luminous glows. The fields producing discharge through the drops are unaltered by reduction of the air-pressure over a wide range ; for example, the field required to cause discharge through a drop 0·227 cm. in radius remains at 8400 volts/cm. as the pressure drops from 76 to 25 cm. of mercury. At still lower pressures the discharge is unaffected by the drop and the sparking potential falls in the usual way. The removal of water by the formation of filaments sets a limit to the size of the drops which can exist inside a thundercloud. Macky concludes from his experiments that no drops larger than 0·15 cm. in radius can be present in the fully charged cloud, and that the critical value of the field at which breakdown occurs is of the order of 10,000 volts/cm. or 33 e.s.u.

With this information the potential and the volume of the charged portion of the cloud can be estimated. Although the results might be expected to depend very much upon the manner in which the charge is distributed, it appears that whatever assumptions are made on this point the values obtained are of the same order of magnitude. Consider a simple spherical distribution of a charge q on a sphere of radius R . the electric field at the surface of the sphere is given by $F = q/R^2$, and at the moment at which a discharge takes place q is 20 coulombs or 6×10^{10} e.s.u. while F has the value 33 e s.u. Hence $R = 427$ metres and the dimensions of the charged region are of the order of 1 kilometre. The potential at the surface is given by $V = q/R = FR = 33 \times 4{\cdot}3 \times 10^4$ e s.u. or $4{\cdot}3 \times 10^8$ volts. If the radial electric force within the sphere everywhere reaches the sparking limit, the potential at the centre is 2FR or 9×10^8 volts. Two such spheres carrying opposite charges would therefore be at a difference of potential of the order of a thousand million volts.

A similar result follows if we make the very different assumption that the two poles of the cloud are distributed in horizontal layers with a vertical thickness L of neutral water-drops between them. The order of magnitude of

L is known from the lengths of lightning discharges to be 2 kilometres, and the critical value of the field, as we have seen, is 10,000 volts/cm. Thus the difference of potential between the poles is $10.000 \times 2 \times 10^5$ or 2×10^9 volts.

When a cloud-charge is conveyed to earth, or neutralised by another charge of opposite sign, by a lightning discharge, the energy dissipated is of the order of 2×10^{17} ergs or 5×10^9 calories; most of this is converted into heat along the path of the lightning flash. A cloud giving one flash every twenty seconds is dissipating electrical energy in the form of lightning at an average rate of 10^6 kilowatts. We have seen that other effects may account for an expenditure of energy at a rate considerably exceeding this.

§ 43. The Distribution and Origin of the Electric Charges

It is generally agreed that the seat of the generation of the electricity of a thundercloud lies within the cloud itself. Equal quantities of positive and negative electricity are produced and then separated by some agency into different regions, where they constitute the "poles" of the electrical machine. During the period in which the charges are built up the dissipation losses in these two regions may well be quite different and the quantities ultimately collected upon the two poles are not necessarily the same. The distribution of the charges has been the subject of a great deal of investigation, both practical and theoretical. At the present time two quite different views as to the distribution and the mechanism which gives rise to it, suggested by G. C. Simpson and by C. T. R. Wilson, are being examined.

According to Wilson the polarity of a thundercloud is what is called positive, the positive pole being higher up in the cloud than the negative one. The cloud is to be considered as made up of large negatively charged drops and small positively charged droplets.[5] When a flash of lightning has just removed the free charges at the poles, the situation is momentarily that illustrated by Fig. 21 (*a*), in which the cloud as a whole is electrically neutral. Immediately afterwards, however, the faster rate of fall under gravity of the large negative drops leads to the development of free positive and negative charges at the top and bottom of the cloud, separated by a neutral region, as represented in Fig. 21 (*b*). Since the free poles attract each other, the rate of separation will become slower as the charges grow and will lead to a rate of fall of the larger drops which is just sufficient to maintain the charges at their existing values, in the face of the dissipating action of point discharge currents within and outside the cloud. Ultimately the field between the poles may become strong enough for a discharge to pass. If the charges are equal, this leaves the cloud in its initial state and the large drops in the neutral region once more separate under the action of

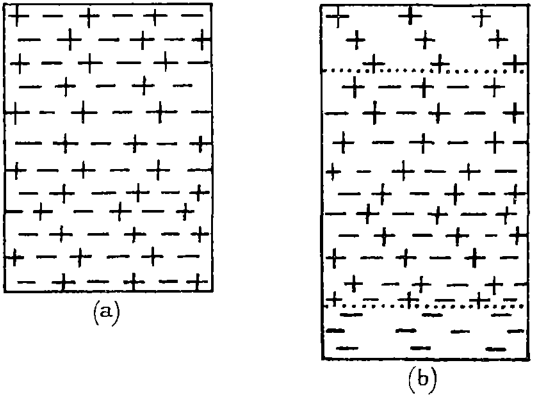

FIG. 21.—Distribution of thundercloud charges (C. T. R. Wilson).
(*a*) Just after a lightning flash.
(*b*) Just before a lightning flash.

gravity. The relative rate of separation may be taken from Macky's results, already discussed, to be the terminal velocity of a drop 0·15 cm. in radius, which is about 6 metres per second. Relative to such a drop the much slower small droplets may be considered as stationary. The average rate at which the initial regeneration of the moment of the cloud takes place is 60/5 coulomb-kilometres per second, using the data given in § 42. From these figures Wilson[5] finds that the average total charge on the positive or negative carriers in the neutral part of the cloud is about 500 coulombs, occupying a volume of some 4 cubic kilometres. The charge per cubic centimetre of water on the larger negatively charged drops he finds to be less than 30 e.s.u., a result which we have seen (§ 36) to be in accord with direct measurements of the charge on rain from thunderstorms.

The mechanism suggested by Wilson as the cause of the opposite charges on large and small drops depends upon the presence in the cloud of numbers of slow ions, attached to Aitken nuclei and small water-drops. Since their mobility (§ 1) is about ·0003 cm./sec., the velocity with which these ions will move, even when the field attains its maximum value of 10,000 volts per metre, is only 3 cm./sec., which is less than that of a drop of water a few thousandths of a centimetre in radius falling under gravity. In the earliest stages of the development of the charged cloud, these may arise from natural sources, but later on they will be supplied in enormous numbers by brush-discharge from water-drops drawn out by the field into pointed forms.

Consider a large drop, say 1 mm. in radius, inside a cloud in which a strong downwardly directed field already exists. Owing to this field, the drop carries a negative polarisation charge upon its upper surface and a positive charge below (Fig. 22). As it falls downwards, it will interact with positive ions moving more slowly down and with negative ions moving up; encounters with other charged drops will be relatively rare. The polarisation charge on the drop and the fact that its velocity, 590 cm./sec., is large compared with that of the slowly moving ions, will cause it to behave differently in its encounters with the two streams. At its under-surface a selective capture of negative ions will occur, for the negative ions which it meets will be attracted and the positive ions repelled by the positive polarisation charge there. A reverse process, a selective capture of positive ions, does not, however, take place at the upper surface, for here the drop is receding from the ions and the positive ions have been repelled away from the drop at the earlier encounter. Innumerable episodes of this kind will lead to the acquisition of a considerable resultant negative charge by the drop from the ion-streams, without seriously affecting the possibility of further captures. For example, a drop 1 mm. in radius falling in a field of 10,000 volts/cm. carries at each end a polarisation charge of ·25 e.s.u. and if it had collected a nett charge equivalent to the very reasonable value of 12 e.s.u. per cubic centimetre of water (§ 36), the negative ions captured would only increase the upper negative and decrease the lower positive charges by 10 per cent.

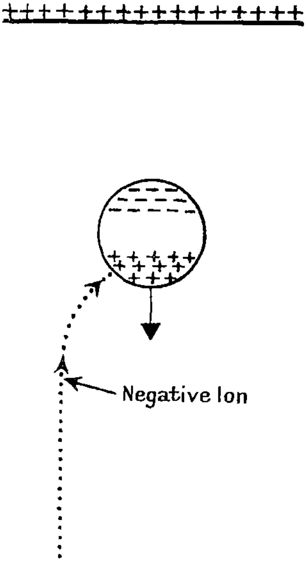

FIG. 22.—Capture of negative ions by large falling drops.

The tiny droplets in the cloud, which are falling with a velocity of the same order as that of the ion-streams, would be unable to exercise the selective action described; they will therefore acquire a nett positive charge, the complement of the negative charge collected by the larger drops.

A complete account of this theory has not yet been given; it would involve consideration of the early stages in the building up of a charged cloud, the action of the normal potential gradient of fine weather in determining the direction of the field to be generated, and the initial building up with the aid of natural ions.

A very different view of the origin and distribution of thundercloud charges has been developed by Simpson,[6] who considers the generation of the electricity to be a consequence of the disruption of rain-drops. Lenard showed in 1892 that when pure water splashes against a solid obstacle the water becomes positively charged, while a complementary negative charge is given to the air. Simpson in 1908 extended these experiments to the breaking of drops in an ascending current of air and obtained the same result, and quantitative values for the charge developed in the process. In extending this observation to conditions within a thundercloud, account has to be taken of the fact that a water-drop of radius greater than ·25 cm. becomes flattened out and unstable when it falls through the air, with the result that it breaks up into a number of smaller drops. Since the terminal velocity of fall for a drop ·25 cm. in radius is 8 metres per second, it follows that no drops of water can ever fall downwards through an ascending current of air whose vertical velocity component exceeds 8 metres/sec. All drops of water under such circumstances will be carried upwards.

The meteorological conditions inside a thunderstorm of the heat-type are represented by Simpson as in Fig. 23; the continuous lines represent stream-lines of the air, and their distance apart is made inversely proportional to the wind-velocity at each point. The air enters the cloud from the right and passes upwards into the front of the cloud. The vertical component of the velocity of the wind is taken to reach a value exceeding 8 metres per second within the oval unshaded region. No water, as we have seen, can pass vertically downwards through this region. The paths of the larger drops, as they fall through the cloud, are represented by the dotted lines. On the right-hand half of the figure these drops are deflected to the left by the air-stream. Above the oval region, therefore, there will be a considerable accumulation of water, and it is here that the larger drops will form and break up on account of instability, giving rise to electric charge. The largest drops, ·25 cm. in radius,* will be falling fast enough to penetrate to just above the forbidden oval, where they will be broken into pieces by the upward-moving air. The fragments, now positively charged, re-form drops and fall back again, and the negative charge given to the rising air will be absorbed by small droplets and carried to the main body of the cloud. The region in which the process of generation of electricity occurs is indicated by the heavily shaded area above the oval.

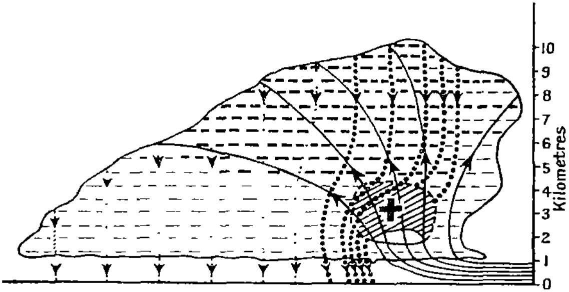

FIG. 23.—Thundercloud electrification (Simpson).

Simpson shows that a reasonable estimate for the average mass of water in the shaded region is 3×10^{11} grams, the dimensions of the region being of the order of a kilometre. If at any one time half this water is in the form of drops large enough to be broken up, the total number of drops available for breaking is $2{\cdot}2 \times 10^{12}$. Since the laboratory experiments showed that a drop of radius ·25 cm , when broken, gives rise to $5{\cdot}10^{-3}$ e.s.u. of charge, the quantity developed if every available drop broke once would be 3·6 coulombs. The drops would have to break six times to produce the 20 coulombs on the poles of the cloud.

---

* The theory in this form was put forward before Macky's results (§ 42) were published.

To decide between these two views on the charge distribution, various experimental tests have been applied to thunderclouds, the description of which is beyond the scope of this book. For the discussions on this question, the reader may be referred to the references at the end of the chapter[2, 3, 6, 7, 8, 9]. It is not impossible that both the Simpson and the Wilson mechanism may be at work in a thundercloud at one and the same time.

§ 44. LIGHTNING

Studies of the lightning discharge are of considerable interest in relation to the problems just discussed; they also furnish valuable material for the study of the spark discharge in air at ordinary pressures, over the mechanism of which general agreement has not yet been reached.

When photographs of the lightning channel taken on a moving camera are examined, they show that many apparently continuous discharges are made up of a number of separate strokes, separated by short periods in which faintly luminous channels sometimes continue. These discontinuities are often apparent to the eye as flicker of the flash of lightning. Though the whole flash often occupies more than a second, the component strokes are generally over in a few thousandths of a second. After each such stroke, the ionisation in places along its track persists sufficiently long to provide the succeeding stroke with a ready-made conducting path. Fig. 24 (after Walter)[10] illustrates how the apparent branching of a flash observed by the eye (24 (*a*)) is often the result of the superposition of a number of separate strokes, whose tracks coincide in part. Fig. 24 (*b*) is from a photograph of the same flash as (*a*), taken on a rotating camera.

These discontinuous strokes must arise from discontinuities in the rate at which the thundercloud can supply electric charge to the end of the channel. Since the cloud itself, though charged, is only a conductor by virtue of the ions produced by point-discharge from drawn-out water-drops, the conditions within the cloud, during the passage of a discharge, must be very complex.[11]

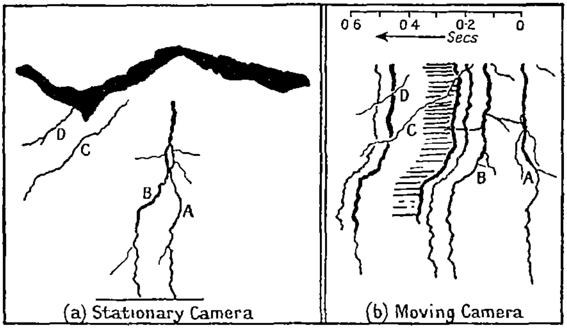

(a) Stationary Camera (b) Moving Camera

FIG. 24.—Component lightning strokes (Walter).

The field-change apparatus of Wilson, in the form described, does not usually yield information as to the quantities of electricity conveyed by these component strokes, the recording arrangement being too slow to record them separately. The results given in § 41 refer to the total integrated field-change and so to the total charge transferred by the flash.

§ 45. OSCILLATIONS IN THE LIGHTNING STROKE

Much attention has been paid to the question as to whether a lightning discharge is oscillatory or aperiodic, and several writers have examined the applicability of the well-known theoretical condition for aperiodicity, $R^2 > 4L/C$, where R is the total resistance of the circuit, C the capacity and L the inductance. The thundercloud itself is considered by some as a condenser plate whose capacity to earth can be estimated from its dimensions, by others as a non-conductor. Simpson,[11] who takes the latter view, has pointed out that the conducting channel of the flash, with its ramifications and tributaries within the cloud, is a conducting system possessing distributed capacity and self-inductance. He has shown that a straight lightning channel 2 km. high and 5 cm. in diameter will be capable of oscillation if its total resistance is less than 2000 ohms, or 1 ohm/metre. A potential difference of 10^9 volts applied to a channel of this resistance would give rise to a current of 500,000 amperes; there is evidence that momentary currents as large as this do actually flow.

On this view the oscillations may be expected to take the form of ripples superimposed upon the main current, without reversing its direction of flow. Such ripples have been observed by several workers (Appleton, Watt and Herd[7]; Norinder[14]; Mathias[20]) who have used the cathode-ray oscillograph to record the field-changes due to distant lightning flashes. The main effect is a unidirectional pulse which rises to a maximum and then falls to zero. On this are superimposed ripples which have a frequency corresponding to a wave-length in air of about 30 km. If the lightning channel be regarded as a gigantic wireless aerial 3 km. high, connected to earth at its base, its natural wave-length would be 12 km. Since the actual channel has extensive branches within the cloud, the agreement with the observed wave-length as regards order of magnitude can be considered satisfactory.

§ 46. THE ELECTRICAL EFFECTS OF DIRECT STROKES

We have seen (§ 42) that the potential of a thundercloud pole is about 10^9 volts. An elevated and insulated conducting system when struck by lightning may be expected to attain a considerable fraction of this potential. Experiments of this nature have been made by Brasch and Lange,[12] who used a wire stretched across an Alpine valley and insulated at each end. They have reported potentials of about 10,000,000 volts, giving sparks to earth 18 metres long. Pittman and Torok,[13] using a cathode-ray oscillograph connected to an overhead power line, have recorded a potential of 5,000,000 volts caused by a travelling wave from a lightning stroke which hit the line four miles away.

The time occupied by a lightning stroke is an important quantity in connection with the current in the discharge. Measurements by Norinder,[14] Mathias [20] and Watson-Watt [7] agree in giving values of the order of ·001 second for the duration of single strokes. Since we know from the work of Wilson that the average quantity of electricity carried by the stroke is about 20 coulombs, the average value of the current must be of the order of 20,000 amps. One would, however, expect that for very short periods of time the discharge current is much greater than this average. Lightning strokes to the supporting towers of power transmission lines have been found to carry currents as high as 250,000 amps.

The question of the diameter of the discharge channel has been the subject of laboratory experiments by Petersen,[15] who has found that powerful spark discharges carrying currents of the order of 60,000 amps. have a constant current-density of 10 amps. per square millimetre. If this result may be extrapolated to currents of 200,000 amps. it would yield 16 cm. for the diameter of the channel of a powerful lightning flash. The fulgurites formed by the fusion of sand struck by lightning have a diameter of some 5 cm., and Petersen reports that in his heavy current discharges the diameter of the channel at the electrodes contracted to one-third of its value elsewhere.

A simple and valuable instrument for the study of direct lightning strokes has been developed in America under the name of the Klydonograph. This is an adaptation of the well-known Lichtenberg figure experiment and operates as a recording voltmeter for high potential differences. The positive electrode (Fig. 25) takes the form of a rod A, touching the emulsion of a photographic plate or film. Below this is a thin layer of varnished paper insulation resting upon a flat metal plate connected to earth. When A is raised to a high positive potential a discharge passes along the photographic emulsion, which on development shows as a circular patch or figure. It has been found that the relation between the diameter of this figure and the applied voltage is a linear one up to potentials of 25,000 volts. Above this the instrument must be used with a potential divider, tapped off, for example, from a string of insulators connected between the elevated conductor and the earth. To measure negative potentials the connections must be reversed, the rod A earthed and the plate connected to the high potential side. For lightning recording, two oppositely connected klydonographs have been joined to the transmission lines under examination.

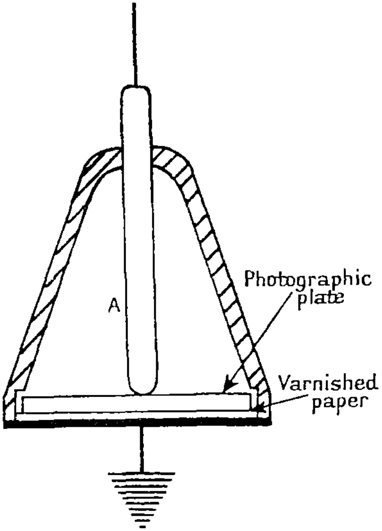

Fig. 25. — Klydonograph arrangement for recording potentials in lightning strokes.

Lewis and Foust [16] have recently published the results of some years' investigations with 1500 klydonographs connected to the supporting towers of transmission lines. One hundred direct strokes were recorded and all these conveyed negative charge from cloud to ground. This result is of interest in connection with the problems of the maintenance of the earth's charge (§ 38), and of the distribution of electricity in the thundercloud.

§ 47. The Mechanism of the Discharge

Two discussions of the actual mechanism of the lightning discharge have been given. Dorsey[17] considers that the principal factor is the movement of a stream of fast electrons from the negative pole, the ions necessary for the subsequent passage of the main discharge arising from the Townsend collision process. According to Simpson,[18] however, the discharge proceeds away from the positive pole; the slower-moving positive ions, on this view, constitute the tip or spear-head of the discharge as it develops, and the negative ions produced at the tip flow into the positive pole along the channel.

Light may be thrown on this important question, which is directly connected with the vexed problem of the mechanism of the electric spark between metal electrodes, by photographs taken with a camera moving sufficiently rapidly to show the manner in which the first stroke in a lightning flash proceeds. C. V. Boys[19] has developed a camera of this type and has published one such photograph. It shows a flash beginning at a point some 500 metres above the earth and developing its track simultaneously above and below.

REFERENCES TO PAPERS

[1] Brookes, *Met. Office Geophys. Memoir*, No. 24 (1925).

[2] C. T. R. Wilson, *Proc. Roy. Soc.*, A, 92, p. 555 (1916); *Phil. Trans.*, A, 221, p. 73 (1921).

[3] Schonland, *Proc. Roy. Soc.*, A, 118, p. 233 (1928); A, 114, p. 229 (1927) (with J. Craib).

[4] Macky, *ibid.*, A, 133, p. 565 (1931); *see also* J. Nolan and O'Keefe, *Proc. Roy. Irish Acad.*, A, 40, p. 86 (1932).

[5] C. T. R. Wilson, *Journ. Franklin Inst.*, 208, p. 1 (1929).
[6] Simpson, *Proc. Roy. Soc.*, A, 114, p. 376 (1927).
[7] Appleton, Watt and Herd, *ibid.*, A, 111, p. 615 (1926).
[8] Banerji, *Journ. Roy. Met. Soc.*, vol. 56, p. 305 (1930).
[9] Wormell, *Proc Roy Soc.*, A, 117, p. 567 (1929).
[10] Walter, *Phys. Zeit.*, 19, p. 273 (1928).
[11] Simpson, *Journ. I.E E.*, 67, p. 1269 (1929).
[12] Brasch and Lange, *Zeit. f. Physik*, 70, p. 1 (1931).
[13] Pittman and Torok, *Trans. Am. I.E.E.* (Sept., 1931).
[14] Norinder, *Journ. Franklin Inst.*, 205, p. 760 (1928).
[15] Petersen, *A.E.G. Progress* (Oct., 1930).
[16] Lewis and Foust, *G.E. Rev.*, 34, p. 452 (1931).
[17] Dorsey, *Journ. Franklin Inst.*, 20, p. 485 (1926).
[18] Simpson, *Proc. Roy. Soc.*, A, 111, p. 56 (1926).
[19] Boys, *Nature*, 118, p. 749 (1926); 122, p. 310 (1928).
[20] A. Mathias, *Mitteilungen der V.D.E W.*, No. 424 (1927).

GENERAL REFERENCES

Victor F. Hess, *The Electrical Conductivity of the Atmosphere and its Causes*, English translation by Codd, gives a detailed account of the matter treated in Chapter I, up to the year 1927.

The Dictionary of Applied Physics, vol. 3, contains a valuable article on "Atmospheric Electricity," by C. T. R. Wilson.

Encyclopædia Britannica, 1910, art. "Atmospheric Electricity," by C. Chree.

K. Kahler, *Einfuhrung in die Atmospharische Elektrizität*, gives details of instruments and the technique of measurement.

E. von Schweidler, *Einfuhrung in die Geophysik*, II, art. "Luftelektrizitat".

K. Kahler, *Die Elektrizität der Gewitter.*

A. Gockel, *Die Luftelektrizitat.*

B. Chauveau, *Électricité Atmosphérique.*

E. Mathias and collaborators, *Traite d'Électricité Atmosphérique et Téllurique.*

A discussion in the *Proceedings of the Royal Society*, A, volume 132, p. 331, and the chapter on the absorption of the gamma rays in Rutherford, Chadwick and Ellis, *Radiations from Radioactive Substances*, will be found valuable in connection with the penetrating radiation. Critical summaries of recent work, in addition to original papers, are published in the journal *Terrestrial Magnetism and Atmospheric Electricity.* For information on the effect of lightning upon overhead transmission lines, etc., reference should be made to the *Transactions of the American Institute of Electrical Engineers.* As a source-book for recent work there is *Science Abstracts.*

INDEX

Air-earth current, *see* Fine-weather current
Air-radiation, 10
Aitken counter, 5
Aitken nuclei, *see* Nuclei
Appleton, 92

Baldit, 63
Benndorf, 48
Bothe, 34, 35
Bothe-Kolhorster expt., 35
Boys, 95
Brookes, 72

Cameron, 26
Capillary electrometer, 73
Capture of small ions, 2, 14
Carnegie, S.S., 19, 50, 52
Charge on the earth, replenishment of, 68
Closed vessel, ionisation in, 12, 14
Collectors, 47
Condensation nuclei, *see* Nuclei
Conduction current, *see* Fine-weather current
Conductivity, atmospheric, measurement of, 7, 55
  effect of pollution on, 52, 59
  influence of large ions on, 7
  polar, 6
  variation with height, 60
  variation with time and season, 59
Convection current in atmosphere, 53, 58
Corlin, 39
Counter, Geiger-Müller, 33

Dalibard, 64
De Sachy, 3
Diameter of lightning channel, 93
Disappearance constant of ions, 16
Dissipation of charge in atmosphere, 7
Distribution of thundercloud charges, 84
Dorsey, 95
Drops, electrification by breaking, 88
Drops deformation in electric fields, 82
Dust-counter, 5
Dust-storms, field in, 62

Earth-radiation, 9
Earth's charge, 43, 68
Ebert effect, 5
Ebert ion-counter, 4
Electrode space-charge, 56
Electrometer, Benndorf, 48
  capillary, 73
  universal portable, 44, 54
Elster, 1
Emanation content of air, 10
Equilibrium of atmospheric ionisation, 14, 19
Equivalent depth, 25

Fast ions, *see* Small ions
Field, earth's in disturbed weather, 62

Field, fine-weather, 43 *et seq.*
  continuous recording of, 47
  measurement of, 43
  origin of, 68
  variation with height, 48; with time and season, 49
Field-changes, due to lightning, measurement of, 73
  sign of, 79
Fine-weather current, measurement of, 53, 54
  significance of, 60
  variation with height, 60; with time and season, 59
Fog, field during, 62
FOUST, 94
FRANKLIN, 64
Fulgurites, 93

Gamma radiation, absorption formulæ, 39
  from air, 13
  from earth, 9
  from potassium, 9
GEIGER-MULLER counter, 33
GEITEL, 1
GERDIEN, conductivity apparatus, 7
GOLD, integral, 30
GSCHWEND, 64

HEAVISIDE, 49
HERD, 92
HESS, 17, 19, 21
HOFFMANN, 23, 24, 40

IDRAC, 49
Intermediate ions, 2, 18
Ionisation at sea, 15, 19
Ionisation balance, 19
Ionisation-depth curve, air and water, 25
  analysis of, 29
  in heavy metals, 31
Ionising agents, 8
Ionising particles, penetrating radiation, 33
Ionising power of radiations, 8, 14
Ions, atmospheric, 2 *et seq.*, *see* Large ions *and* Small ions
  charge on, 3, 18
  equilibrium of, 14 *et seq.*
  life of, 16
  mobilities of, 2
  number of, 2, 3

JEANS, 41

KELVIN, water-dropper, 47
KENNELLY-HEAVISIDE layer, 49
Klydonograph, 94
KOLHORSTER, 21, 23, 35

LANGE, 93
LANGEVIN ions, *see* Large ions
Large ions, 3
  charge on, 3
  complexities of, 2, 18
  effect on conductivity, 6, 59
  formation of, 2, 14
  mobility of, 2
  number of, 3
LEMONNIER, 43, 64
LENARD, 63, 88
LEWIS, 94
LICHTENBERG, 94
Life of small ions, 16, 17, 18, 19
Lightning channel, diameter of, 93
Lightning, charge carried by, 81
  current in, 93
  discontinuous nature of, 90
  duration of, 93
  effect on earth's charge, 69
  field-changes due to, 73
  frequency of flashes, 72
  mechanism of, 95
  oscillations in, 91
  photography of, 90, 95
  polarity of strokes, 95
LINSS, 1
LUTZ, 58

M'CLELLAND, 63
MACKY, 82
Maintenance of earth's charge, 68
MARWICK, 63
MATHIAS, 17, 92, 93
MAUCHLY, 52
MILLIKAN, 24, 26, 29, 40
Mobilities of ions, 2
Moments of flashes, 81
MOTT-SMITH, 35
MULLER, 33
Multiply-charged ions, 18

Neutron, 42
NOLAN, J. J., 3, 16, 18
NOLAN, P., 18
NORINDER, 92
Normal ions, *see* Small ions
Nuclei, AITKEN or condensation, 2, 5
  counting of, 5
  effect on life of small ions, 14
  effect on potential gradient, 59, 60

Penetrating radiation, 11, 21 *et seq.*
  absorption in air and water, 25; in heavy metals, 25
  fluctuations, statistical, 21
  intensity of, 11, 22
  ionising particles from, 33
  ionising power of, 15
  nature of, 35
  origin of, 39
  variation with time, 39
PETERSEN, 93
PITTMAN, 93
Point-discharge, 64
  effect on earth's charge, 69
  effect on thundercloud charges, 82
Polar conductivity, *see* Conductivity
Polarity of thunderclouds, 84; of lightning strokes, 95
Pollution, atmospheric, 50; *see* Nuclei
Potassium, γ rays from, 9
Potential gradient, *see* Field
Precipitation, charge on, 62
  effect on earth's charge, 68

Radioactive matter in air, 10
  in earth, 9
Rain, charge on, 62
Recovery curves, thundercloud, 76, 82
REGENER, 28, 41
Residual ionisation, 12, 22, 25
  fluctuations in, 23
Reversal distance, lightning field-changes, 80
ROSSI, 35
Runaway electrons, 41
RUTHERFORD, 30

St. Elmo's fire, 64
SCHERING, 55
SCHINDELHAUER, 63
SCHONLAND, 23, 36, 64, 80
SCHWEIDLER, 16, 18
SIMPSON, 51, 69, 88, 95
SKOBELZYN, 33
Slow ions, *see* Large ions
Small ions, 2; capture of, 2, 14; charge on, 2; complexities of, 2, 18
  effect on conductivity, 6
  life of, 16 *et seq.*
  mobility of, 2
  number of, 2
Smoke pollution, effect on field, 50
Snow, charge on, 62
  field during, 62
Space-charge, at earth's surface, 56
  in lower atmosphere, 49
Sparking field, 82
STEINKE, 31
SWANN, 5, 69

Thunderclouds, charges on, 81, 86
  distribution of charge on, 84
  effect on upper atmosphere, 70
  electrons from, 41
  fields of, 62, 73
  moments of, 80
  number of, 72
  origin of charge on, 84
  point-discharge effects from, 65, 66
  polarity of, 84
  potentials in, 83
  rate of working, 84
Torok, 93
Transition zone, 31
Universal electrometer, 44, 54
Upper atmosphere: potential of, 49; effect of thunderclouds on, 70

Voltameter, for point-discharge currents, 65

Wall effect, *see* Residual ionisation
Walter, 90
Water-dropper, 47
Watson, 46, 58
Watson-Watt, 92, 93
Whipple, 50, 71
Wigand, 68
Wilson, C. T. R., 1, 41, 44, 54, 65, 69 *et seq.*
Wormell, 66, 70

PRINTED IN GREAT BRITAIN AT THE UNIVERSITY PRESS, ABERDEEN

No. 841,386. PATENTED JAN. 15, 1907.

L. DE FOREST.

WIRELESS TELEGRAPHY.

APPLICATION FILED AUG. 27, 1906.

2 SHEETS—SHEET 1.

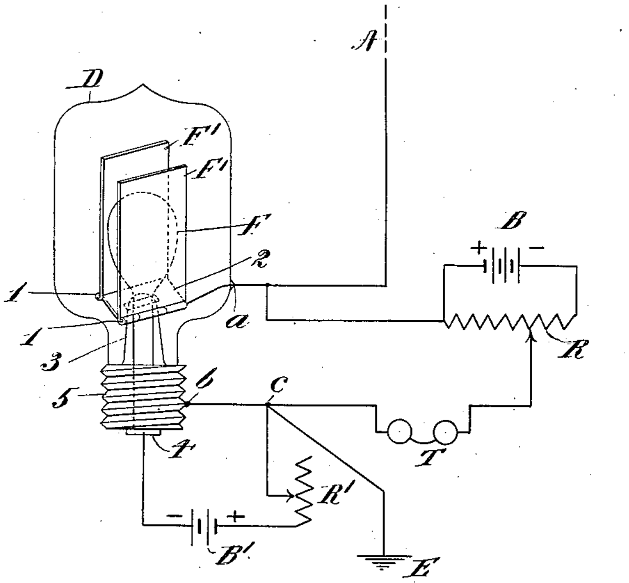
A
D
F′
F′
F
B
+
−
2
1
1
a
3
R
5
6
C
T
4
R′
−
+
B′
E

Fig. 1.

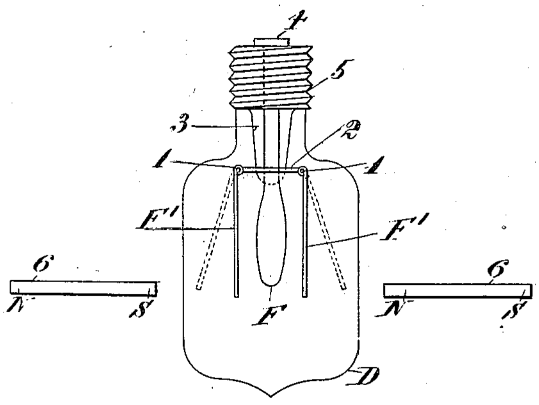
4
5
3
2
1
1
F′
F′
6
6
N
S
N
S
F
D

Fig. 2.

WITNESSES=

Frank G. Parker

John Buckler,

INVENTOR=

Lee de Forest

by Geo. K. Woodworth

Atty.

No. 841,386. PATENTED JAN. 15, 1907.

L. DE FOREST.

WIRELESS TELEGRAPHY.

APPLICATION FILED AUG. 27, 1906.

2 SHEETS—SHEET 2.

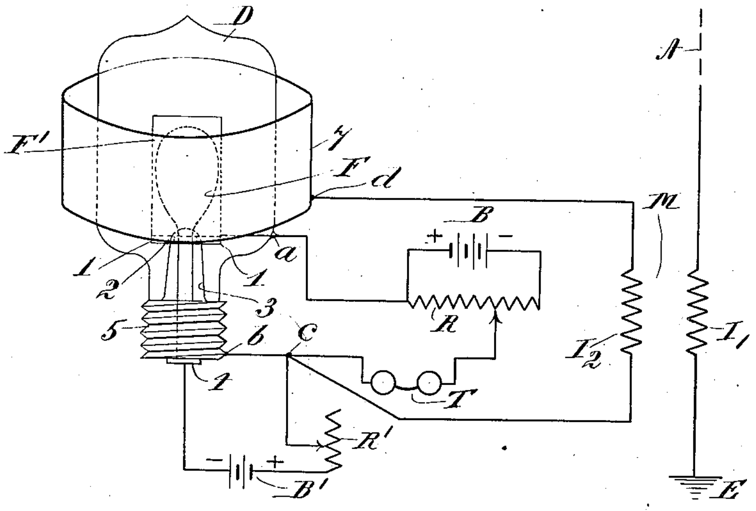
D
F'
4
F'
d
A
M
B
+
-
1
2
a
1
3
R
5
6
C
I2
I1
4
T
R'
-
+
B'
E

Fig. 3.

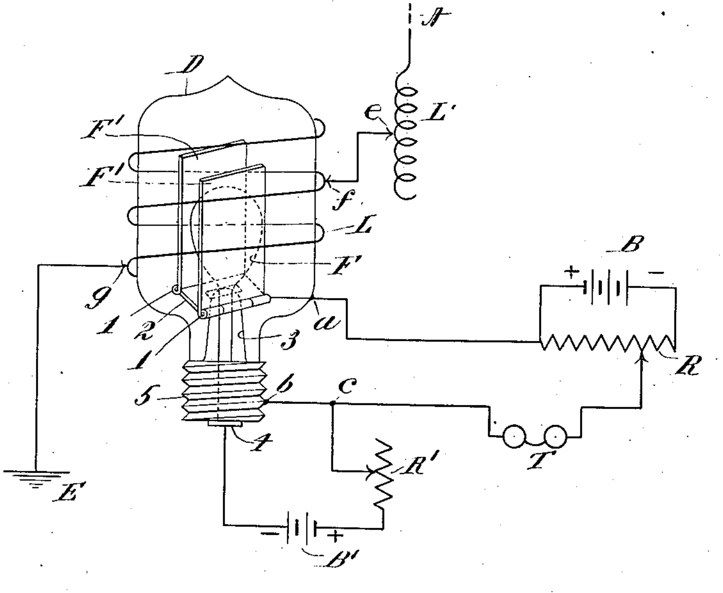
A
D
e
L'
F'
F'
f
L
F'
B
+
-
9
1
2
a
1
3
R
5
6
C
4
R'
T
E'
-
+
B'

Fig. 4.

WITNESSES:

Frank G. Parker

John Buckler,

INVENTOR:

Lee de Forest

by Geo. K. Woodworth

Atty.

UNITED STATES PATENT OFFICE.

LEE DE FOREST, OF NEW YORK, N. Y.

WIRELESS TELEGRAPHY.

No. 841,386. **Specification of Letters Patent.** **Patented Jan. 15, 1907.**

Application filed August 27, 1906. Serial No. 332,213.

To all whom it may concern:

Be it known that I, Lee de Forest, a citizen of the United States, and a resident of New York, in the county of New York and State of New York, have invented a new and useful Improvement in Wireless Telegraphy, of which the following is a specification.

My invention relates to wireless telegraphy and more particularly to an oscillation-detector which, in addition to performing the usual functions of such detectors, may be used as a tuning device and as a static valve, and which has a variety of other uses.

I have discovered that if the gaseous medium intervening between separated electrodes be put into a condition of molecular activity, as by heating the same or otherwise, said medium becomes highly sensitive to electrical oscillations, as more fully set forth in my application, Serial No. 243,913.

I have shown in my application, Serial No. 300,975, that the sensitiveness of an oscillation-detector comprising electrodes separated by such sensitive gaseous medium may be varied by the action of a magnetic field, so that the response of said detector *per se* is a function of a characteristic, such as the frequency, of the electrical oscillations operating upon the same. Thus, for certain frequencies a magnetic field of given strength will greatly decrease the sensitiveness of the detector and render it practically non-responsive, while for other frequencies the same magnetic field will greatly increase the sensitiveness of the detector. In short, said detector *per se* may be made selectively responsive without having recourse to tuned or resonant circuits, so that the use of such circuits may be dispensed with and the detector itself employed to perform the functions heretofore effected by said circuits.

The subject-matter of the present invention is an oscillation-detector which, like that above referred to, may be made selectively responsive to electrical oscillations having a given frequency or electrical rate of vibration but which does not require a proximate magnetic field for this purpose; and which also may be made selectively responsive to electrical oscillations having a given spark, or wave-train, frequency—that is to say, may be so adjusted that its response to electrical oscillations is a function of the number of groups or trains of waves per second, irrespective of the electrical frequency of such waves.

By virtue of the selective properties of my oscillation-detector as above set forth, I am enabled to employ said detector as a tuning device and also as a static valve or leak; and on account of its indestructibility I am enabled to employ the detector in those systems where large amounts of energy are created in the immediate proximity of said detector, such, for example, as the duplex systems described in my United States Letters Patent No. 772,879, dated October 18, 1904, and in the systems which permit "breaking" between two operators as set forth in my United States Letters Patent Nos. 827,523 and 827,524, dated July 31, 1906.

My invention may best be understood by having reference to the drawings which accompany and form a part of this specification and which diagrammatically illustrate four embodiments thereof which have successfully been employed; but I do not wish to limit myself to the details shown therein inasmuch as many other forms of construction and arrangement of circuits embodying the principles of my invention, as particularly pointed out in the appended claims, will readily occur to those skilled in the art.

In the drawings, Figure 1 represents a simple form of my oscillation-detector conductively connected with a receiving-antenna. Fig. 2 represents a detail of construction used in connection with the detectors shown in Figs. 1, 3 and 4. Fig. 3 represents a modification of the detector shown in Fig. 1 in which the detector is energized by an electric field developed by the oscillations to be detected. Fig. 4 represents a further modification in which the detector is energized by a magnetic field created by the oscillations to be detected.

In the figures, A is a receiving-antenna.

E is an earth connection.

M is a transformer, of which I_1 and I_2 are the primary and secondary windings.

B B′ are batteries.

R R′ are resistances.

L L′ are inductances.

T is a telephone or other signal-indicating device.

D represents an evacuated vessel of glass or other suitable material having two separated electrodes F and F′ between which intervenes the gaseous medium which, when sufficiently heated or otherwise made highly conducting, forms the sensitive element of my oscillation-detector. While various means may be employed for heating said gaseous medium I find it convenient to employ electrical means and, therefore, I prefer to employ as the electrode F a filament of carbon or metal and to connect the same in series with a battery B′ and a rheostat R′. I have ascertained that a filament of metal, especially one of tantalum, renders the oscillation-detector much more sensitive than one of carbon.

The electrode F′ may consist of a non-heated metallic body such, as shown in the present instance, as the wings or plates F′ F′ secured to the metallic base 2, which may be supported by the upwardly-extending portion 3 of the vessel D, although any suitable means may be used to support the electrodes F′ within said vessel.

While I have shown and described the electrode F′ as consisting of two wings or flat plates, it will be understood that said electrode may have any other form suitable for the purpose, although I have found that better results are obtained by using two flat plates or wings than by using a flattened cylinder and that a flattened cylinder gives better results than a cylinder of circular section. The reason for this is that the sensitiveness of the device, other things being equal, depends in a measure upon the distance between the electrodes F F′, and the minimum separation can be obtained by using flat plates for the electrode F′. As will be obvious, the electrode F′ may consist of one flat plate.

The evacuated vessel D may be provided with a screw-socket 5 after the manner of ordinary incandescent lamps, and one end of the filament F may be conductively connected thereto while the other end may be insulated therefrom and conductively connected to the contact-piece 4 in the base of the socket. In such case the terminals of the heating battery B′ are connected respectively to the socket at *b* and to the contact-piece 4.

The terminals of the local circuit, which includes the battery B and signal-indicating device, such as the telephone T, are connected respectively to the electrodes F and F′ at *a* and *b*. Means may be included in or associated with the local circuit for varying the potential developed by the local circuit B between the points *a* and *b*. For this purpose the number of cells which make up the battery B may be varied, or the adjustable resistance R, which, with the battery, constitutes a potentiometer, may be used.

The oscillation-detector, as above described, may be associated with the receiving system in a variety of ways now well known in the art by connecting the electrode F to one of the terminals of the source of oscillations and the electrode F′ to the other terminal thereof. In the present instance the antenna A is connected to the electrode F′ by a leading-in wire which is connected to the base 2 and which is sealed in the vessel D at *a*, and the earth-connector is connected to the electrode F through the socket 5 and the conductor *b c*.

It is preferable that the positive pole of the battery B should be connected to the electrode F′, and better results are obtained if the negative pole of the battery B be connected to that end of the filament F to which the positive pole of the battery B′ is connected, (shown in Figs. 1, 3, and 4 as that end of said filament which is connected to the screw-socket.)

By regulating and adjusting the amount of heating-current flowing in the circuit of the battery B′, the sensitiveness of my detector may be varied, so that the response of the detector *per se* will be a function of the frequency of the electrical oscillations operating upon the same, and said detector itself, as distinguished from the circuit in which it is included, will be selectively responsive to oscillations having a given electrical frequency. For regulating and adjusting the heating-current flowing through the filament F, I may employ the rheostat R′ or any other means suitable for the purpose.

For tuning the oscillation-detector to oscillations of given electrical frequency, the amplitude of the heating-current must be determined empirically, as by adjusting the rheostat R′, because a number of factors such, for example, the separation of the electrodes F F′ and the difference of potential developed by battery B across the points *a b*, are involved.

By regulating and adjusting the potential impressed by the battery B across the electrodes F F′, the sensitiveness of the detector may be varied so that the response of the detector *per se*, other things being equal, will be a function not only of the electrical frequency or wave length of the oscillations operating upon the detector, but also of the spark frequency which, *caeteris paribus*, is a measure of the total energy received, and the detector itself, as distinguished from a circuit tuned electrically to electrical or to spark frequency and from a mechanism tuned mechanically or acoustically to spark frequency, will be selectively responsive to oscillations having a given electrical frequency, and also to oscillations having a given spark frequency irrespective of their electrical frequency or wave length. Thus, with a given separation of the electrodes F F′ and a given amplitude of heating-current flowing through the filament F, I have been able by merely varying the potential of the battery B to attune the oscillation-detector to the oscillations developed in the receiving system by one of several transmitting systems and to receive the energy thereof to the exclusion of the energies of the oscillations developed therein by the other of said transmitting systems, even when the several sets of oscillations had the same electrical frequency and wave length, but differed in spark frequency or in average energy conveyed per unit of time. Here also the adjustment of the potential impressed by the battery B upon the electrodes F F′ must be determined empirically.

By virtue of the selective properties of my detector, as above set forth, I am enabled to effect by the detector itself a method of double-tuning which affords the greatest security from waves the energy of which is not intended to affect the detector as well as from atmospheric or "static" electricity and other disturbing electrical forces.

As above mentioned, the sensitiveness of my oscillation detector, which in common with others of my invention employing a conducting gaseous medium as the sensitive element I call the "audion," is a function of the separation of the electrodes or of the length of the sensitive conducting gaseous medium; but I find that here again the frequency is a determining factor, so that by varying the separation of the electrodes I can make the response of the audion a function of the frequency. In other words, by suitably varying the length of the interelectrode medium, other adjustments, such as those of the batteries B B′, remaining unchanged, I am enabled to make the audion *per se* selectively responsive to oscillations of a given frequency independently of the usual tuned or resonant circuit heretofore employed in wireless telegraphy as a current or oscillation selector.

A convenient means of varying the distance between the electrodes F F′ is to hinge the wings which constitute the electrode F′ to the base 2 as by the hinges 1 1. By approaching the magnets 6 6 to the inverted tube, as shown in Fig. 2, the wings, which may be made wholly or partly of iron, will open outwardly away from the filament and thereby increase the path of the conducting ions in the interelectrode medium.

When the audion is to be attuned to oscillations of a given frequency by varying the separation of the electrodes, the adjustment produced by the magnets 6 6 must be empirically determined, just as the adjustments of the batteries B B′, above referred to are purely empirical.

It will be understood that each of the three methods above set forth for varying the response of the audion in accordance with the electrical frequency and the spark frequency of the oscillations to be detected may be separately employed and also that any two or all three of said methods may be used conjointly. For example, by appropriately varying the separation of the electrodes F F′ and the amplitude of the heating-current from the battery B′, the audion may be made selectively responsive to waves of a given electrical frequency, and then by the regulation and adjustment of the potential impressed upon the audion by the battery B, the audion may be made to select a particular one of a number of sets of waves all having the said electrical frequency and wave length, but each set having a spark frequency different from that of any other set.

The oscillation-detectors shown in Figs. 3 and 4 are identical in construction with those above described in connection with Figs. 1 and 2, and differ therefrom merely in the manner in which they are associated with the source of oscillations to be detected.

In Fig. 3, 7 represents a metallic member brought into proximity with the vessel D and shown in the present instance as a cylinder surrounding said vessel. In this case one terminal of the oscillation source which in the present instance is shown as a circuit including the secondary of the transformer M, is conductively connected to the cylinder 7 at the point *d*, while the other terminal may be connected to either one of the electrodes F or F′ and in the present instance is shown as connected to the electrode F through the socket 5 and conductor *b c*. In this case the oscillating electric field developed by the electrical oscillations in the secondary circuit, which includes the secondary winding I_2, operates to alter the conducting properties of the sensitive conducting gaseous medium in the vessel D and thereby to vary the current flowing in the local circuit which includes the telephone T, just as in the case of the system shown in Fig. 1 in which the terminals of the oscillation source are connected directly to the interelectrode medium.

The arrangement shown in Fig. 3 affords a fourth method of rendering the audion selectively responsive, for in addition to the three methods above set forth in connection with Figs. 1 and 2 I find that, other things being equal, the response of the audion to oscillations of a given frequency is a function of the area of the surface of the number 7 and of the separation thereof from the elements within the vessel D.

If the form shown in Fig. 3 be employed, the cylinder 7 may be moved with respect to the axis of the vessel D until, *cœteris paribus*, the response of the audion is a maximum for oscillations of a given frequency. It will be understood of course that the separation of the electrodes F, F′, and the values of the resistances R, R′, of the audion shown in Fig. 3 may all be varied in the manner above explained in connection with Figs. 1 and 2.

When the connections shown in Fig. 3 are used I find that the quality of the sound produced in the telephone T is quite different from that produced therein when the connections shown in Fig. 1 are used, although the sounds may be, and under the same conditions are, of equal intensity. When the connections shown in Fig. 1 are employed, the sounds produced by the telephone are sharp and crackling, as is the case where various other forms of oscillation-detector are employed; but when the connections shown in Fig. 3 are used, the sounds produced by the telephone, under the same conditions as before, are muffled or drummy. This latter sound is distinctive and is not produced by the use of any other form of oscillation-detector so far as I am aware. The advantage of obtaining the latter sound in the telephone is that it may readily be distinguished from the sounds produced therein by atmospheric electricity, whereas the sharp crackling sounds produced in the telephone by the signal-waves when other types of oscillation-detector are used very nearly imitate the sounds produced therein by static disturbances or atmospheric electricity.

In the arrangement shown in Fig. 4 the audion is energized by the oscillating magnetic field developed by the passage of the oscillations to be detected through the coil L which surrounds the sensitive conducting gaseous medium and which in the present instance is wound around the outside of the tube. This coil preferably is made adjustable, as by the contacts *f* and *g*, and the antenna A may include an inductance L′ having an adjusting-contact *e*. It will be understood that the adjustments of the separation of the electrodes F, F′, of the amplitude of the heating-current and of the value of the potential impressed between the points *a* and *b*, as set forth above in connection with Fig. 1, may also be employed to render the audion shown in Fig. 4 selectively responsive. I have ascertained also that the variation of the magnetic field created by the passage of oscillations through the coil L varies the response of the audion, other things being the same, and furthermore that this variation also is a function of frequency. In other words, having adjusted the audion to be most responsive to oscillations of given frequency in the manner above described in connection with Fig. 1, there may be found empirically a certain position for the contacts *f* or *g*, at which for oscillations of a said frequency the response of the detector is a maximum.

The syntonizing-inductance L′, or other electrical tuning device, may be employed in connection with any of the several embodiments of my present invention so as to combine the selectivity of the antenna or tuned receiving-circuit with the selective properties of the oscillation-detector itself, and in such cases I have found the selectivity of a receiving system so constructed to be very high indeed.

It will be understood that in all cases, if desired, the vessel D may include the vapor of a halogen salt or other substance the vapor of which is conducting or a gas whose conductivity is greater than that of air under an attenuated pressure.

I claim—

1. An electrical-oscillation-selecting device comprising a vessel, a filament sealed in said vessel, a source of electric current connected in series with said filament, means for varying the amplitude of said current, whereby said device is rendered selective of electrical oscillations having a given electrical frequency, a member of conducting material inclosed within said vessel and a circuit connecting said member with said filament.

2. An electrical-oscillation-selecting device comprising a vessel, a filament sealed in said vessel, a source of electric current connected in series with said filament, a member of conducting material inclosed within said vessel, a circuit connecting said member with said filament, a source of electromotive force associated with said circuit and means for varying said electromotive force, whereby said device is rendered selective of electrical oscillations having a given spark frequency.

3. An electrical-oscillation-selecting device comprising a vessel, a filament sealed in said vessel, a source of electric current connected in series with said filament, means for varying the amplitude of said current, whereby said device is rendered selective of electrical oscillations having a given electrical frequency, a member of conducting material inclosed within said vessel, a circuit connecting said member and said filament, a source of electromotive force associated with said circuit and means for varying said electromotive force, whereby said device is rendered selective of electrical oscillations having a given spark frequency.

4. An electrical-oscillation-selecting device comprising a vessel inclosing a gaseous medium maintained in a condition of molecular activity, two electrodes inclosed within said vessel, and means for varying the separation of said electrodes.

5. An electrical-oscillation-selecting device comprising a vessel inclosing a gaseous medium maintained in a condition of molecular activity, two electrodes inclosed within said vessel, and means outside said vessel for varying the separation of said electrodes.

6. An electrical-oscillation-selecting device comprising a vessel, a filament sealed in said vessel, a source of electric current connected in series with said filament, a member of conducting material inclosed within said vessel, a circuit connecting said filament and said member and means for varying the separation of said filament and said member.

7 An electrical-oscillation-selecting device comprising a vessel, a filament sealed in

said vessel, a source of electric current connected in series with said filament, and means for varying the amplitude of said current, whereby said device is rendered selective of electrical oscillations having a given electrical frequency.

8. An electrical - oscillation - selecting device comprising a vessel, a filament sealed in said vessel, a source of electric current connected in series with said filament, a member of conducting material inclosed within said vessel, a circuit connecting said filament with said member, a source of electromotive force associated with said circuit, means for varying said electromotive force and means for varying the separation of said filament and said member.

9. An electrical - oscillation - selecting device comprising a vessel, a filament sealed in said vessel, a source of electric current connected in series with said filament, means for varying the amplitude of said current, whereby said device is rendered selective of electrical oscillations having a given electrical frequency, a conductor inclosed within said vessel, a circuit connecting said conductor and filament, a source of electromotive force associated with said circuit and means for varying said electromotive force, whereby said device is rendered selective of electrical oscillations having a given spark frequency.

10. An oscillation-detector and means whereby said detector *per se* is rendered selective of electrical oscillations having a given spark frequency.

11. An oscillation-detector, means whereby said detector *per se* is rendered selective of electrical oscillations having a given electrical frequency and means whereby said detector *per se* is rendered selective to electrical oscillations having a given spark frequency.

12. An electrical-oscillation-selecting device comprising a vessel containing a sensitive conducting gaseous medium, two electrodes within said vessel, a local circuit connecting said electrodes, a receiving-circuit having one terminal connected with one of said electrodes, a conducting member outside said vessel and movable with respect thereto, and an electrical connection from the other terminal of said receiving-circuit to said conducting member.

13. An electrical-oscillation-selecting device comprising a vessel inclosing a gaseous medium maintained in a condition of molecular activity, two electrodes inclosed within said vessel, means for impressing an electrical potential upon said electrodes, and means for varying said potential.

14. An electrical-oscillation-selecting device comprising a vessel inclosing a gaseous medium, an electrode within said vessel, means for heating said electrode, means for regulating the temperature thereof, a second electrode inclosed within said vessel and a circuit connecting said electrodes.

15. An electrical-oscillation-selecting device comprising a vessel inclosing a gaseous medium, an electrode within said vessel, means for heating said electrode, means for regulating the temperature thereof, a second electrode inclosed within said vessel, means for impressing an electrical potential upon said electrodes and means for varying said potential.

16. An electrical-oscillation-selecting device comprising a vessel inclosing a gaseous medium maintained in a condition of molecular activity, two electrodes inclosed within said vessel, a local circuit connecting said electrodes and means for varying the separation of said electrodes.

17. An electrical-oscillation-selecting device comprising a vessel inclosing a gaseous medium, an electrode within said vessel, means for heating said electrode, means for varying the temperature thereof, a second electrode inclosed within said vessel and means for varying the separation of said electrodes.

18. An electrical-oscillation-selecting device comprising a vessel inclosing a gaseous medium, an electrode within said vessel, means for heating said electrode, means for varying the temperature thereof, a second electrode inclosed within said vessel, a circuit connecting said electrodes, an adjustable source of electromotive force associated with said electrodes, and means for varying the separation of said electrodes.

19. An electrical-oscillation detector comprising a vessel inclosing a gaseous medium, means for maintaining said gaseous medium in a condition of molecular activity, a member in proximity to said vessel, and an oscillation-circuit electrically connected to said member and to said gaseous medium.

20. An electrical-oscillation-selecting device comprising a vessel inclosing a gaseous medium, means for heating said medium, and means for varying the temperature thereof, whereby said device is rendered selective of electrical oscillations having a given electrical frequency.

21. An oscillation-detector comprising two conductively-connected flat plates of conducting material inclosed within an evacuated vessel, an electrode sealed within said vessel and located between said plates, and means for heating said electrode.

22. An oscillation-detector comprising two conductively-connected flat plates of conducting material inclosed within an evacuated vessel, a metallic filament sealed within said vessel and located between said plates, and means for heating said filament.

23. An oscillation-detector comprising two conductively-connected flat plates of con-

ducting material inclosed within an evacuated vessel, a filament of tantalum sealed within said vessel and located between said plates, and means for heating said filament.

24. An oscillation-detector comprising two flat plates of conducting material inclosed within an evacuated vessel, an electrode sealed within said vessel and located between said plates, and means for heating said electrode.

25. An oscillation-detector comprising two flat plates of conducting material inclosed within an evacuated vessel, a filament sealed within said vessel and located between said plates, and means for heating said filament.

In testimony whereof I have hereunto subscribed my name this 24th day of August, 1906.

LEE DE FOREST.

Witnesses:
Geo. E. Terry,
Geo. K. Woodworth.

No. 841,387.

PATENTED JAN. 15, 1907.

L. DE FOREST.

DEVICE FOR AMPLIFYING FEEBLE ELECTRICAL CURRENTS.

APPLICATION FILED OCT. 25, 1906.

2 SHEETS—SHEET 1.

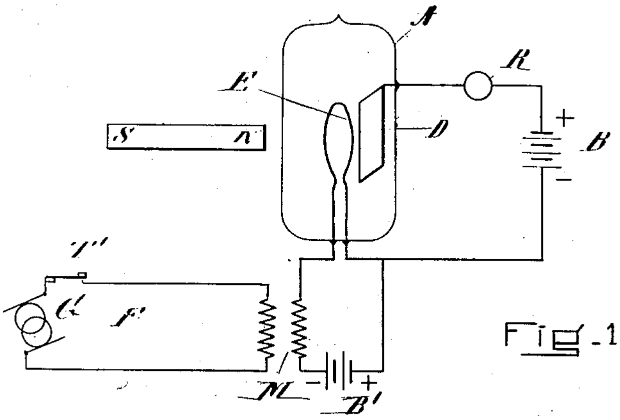
A
E
R
S
n
D
B
T''
G
F
M
B'

Fig-1.

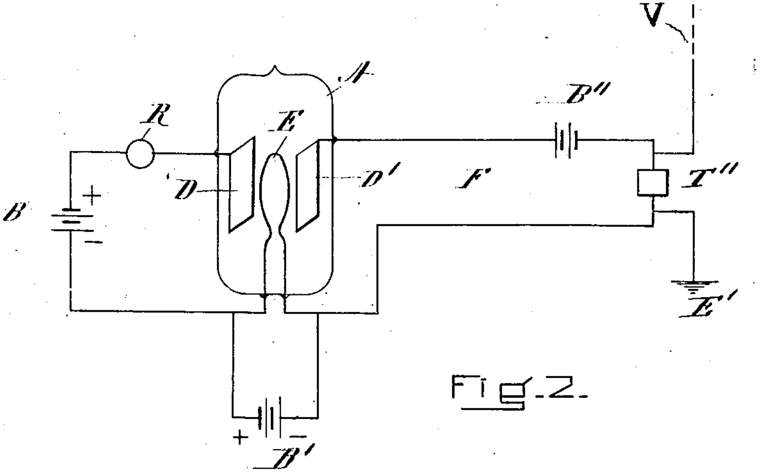
V
A
B''
R
E
D
D'
F
T''
B
E'
B'

Fig-2.

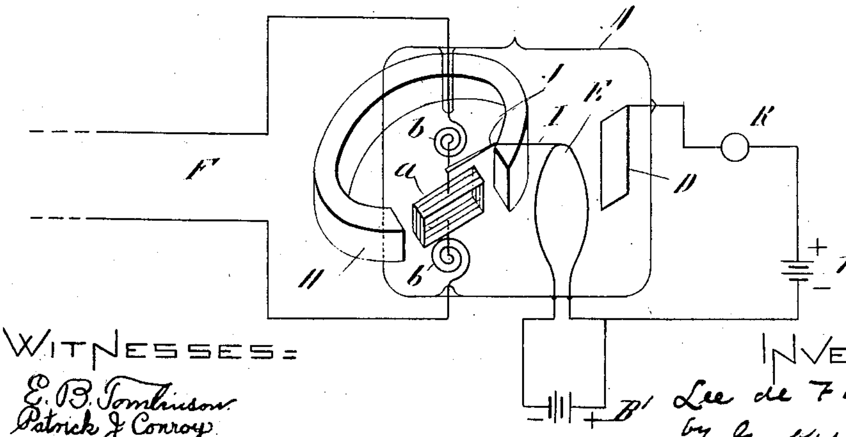
A
J
E
I
b
a
b
R
D
F
H
B
B'

Fig-3.

WITNESSES:

E. B. Tomlinson

Patrick J. Conroy

INVENTOR:

Lee de Forest

by Geo. K. Woodworth

Attorney.

No. 841,387. PATENTED JAN. 15, 1907.

L. DE FOREST.

DEVICE FOR AMPLIFYING FEEBLE ELECTRICAL CURRENTS.

APPLICATION FILED OCT. 25, 1906.

2 SHEETS—SHEET 2.

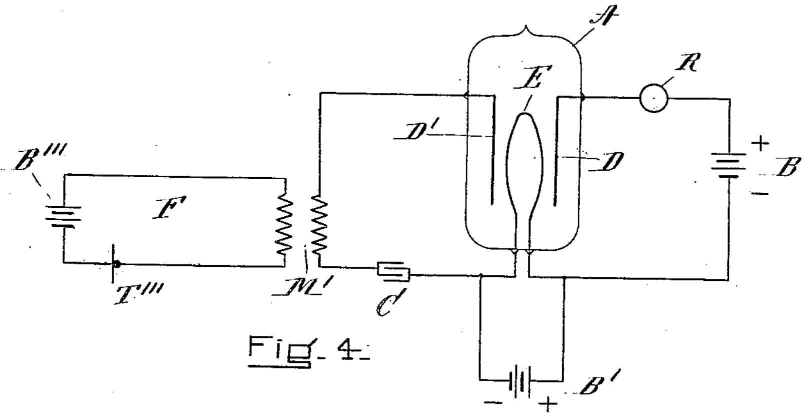
A
E
R
D'
D
B'''
F
+
B
−
T'''
M'
C
B'
−
+

Fig. 4.

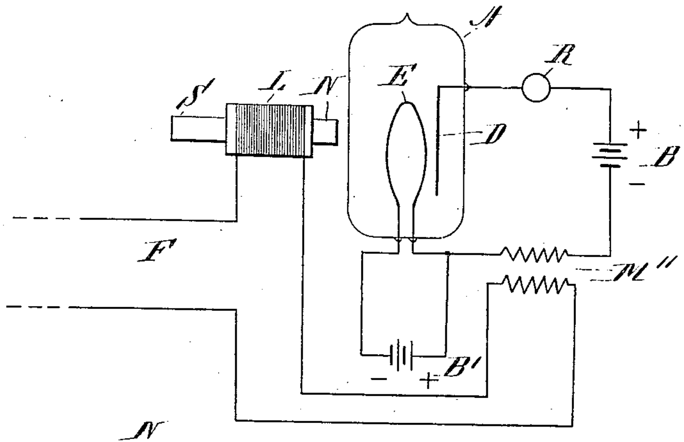
A
R
E
S
L
N
D
+
B
−
F
M''
B'
−
+

Fig. 5.

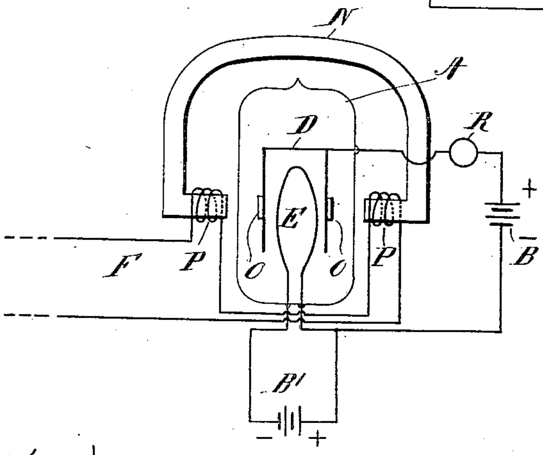
N
A
D
R
+
E
B
F
P
O
O
P
B'
−
+

Fig. 6.

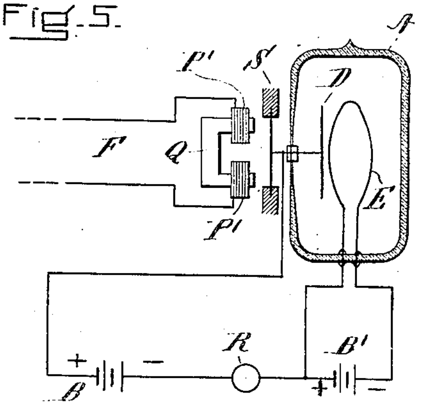
A
P'
S
D
F
Q
E
P'
R
B'
+
−
B
+
−

Fig. 7.

WITNESSES:
E. B. Tomlinson
Patrick J. Conroy

INVENTOR:
Lee de Forest
by Geo. K. Woodworth
Attorney.

UNITED STATES PATENT OFFICE.

LEE DE FOREST, OF NEW YORK, N. Y.

DEVICE FOR AMPLIFYING FEEBLE ELECTRICAL CURRENTS.

No. 841,387. **Specification of Letters Patent.** Patented Jan. 15, 1907.

Application filed October 25, 1906. Serial No. 340,487.

To all whom it may concern:

Be it known that I, Lee de Forest, a citizen of the United States, and a resident of New York, in the county of New York and State of New York, have invented a new and useful Improvement in Devices for Amplifying Feeble Electrical Currents, of which the following is a specification.

My invention relates to devices for amplifying feeble electrical currents—such, for example, as telephone-currents; and its object is to produce an amplifying device of greater efficiency and simplicity than those heretofore employed.

My invention will be described with reference to the drawings accompanying and forming a part of this specification, and in which—

Figures 1, 2, 3, 4, 5, 6, and 7 represent conventionally or diagrammatically various arrangements of apparatus and circuits whereby my invention may be carried into effect.

In the figures, A represents an evacuated vessel inclosing a sensitive conducting gaseous medium maintained in a condition of molecular activity.

R is a signal-indicating device.

B B are batteries or other sources of electrical energy.

D E D', are electrodes sealed within the receptacle A.

The circuit B R D E is a local receiving-circuit. The circuit F is a line-circuit conveying the currents to be amplified to the amplifying device. The electrode E, which may be of platinum, tantalum, carbon, or other suitable material, is heated and preferably maintained incandescent by the battery B'. The electrodes D and D , which may be plates of platinum or other suitable material, are placed in close proximity to the electrode E, and when the electrode D' is employed its separation from the electrode E preferably is less than that of the electrode D therefrom.

In Fig. 1, N S represent a magnet placed adjacent to the vessel A. The currents to be amplified may be impressed upon the circuit which includes the heated electrode or filament E—as, for example, by means of the transformer M—and the magnetic field set up by these currents reacts upon the field set up by the magnet N S, thereby causing a slight variation in the separation between the electrodes D E. I have found that the slightest variation in the separation of the hot and cold electrodes produces a large and disproportionately greater variation in the flux between said electrodes, especially if the latter are close together, and such variation in flux may be made manifest by the signal-indicating device R.

In Fig. 2 the current to be amplified may be impressed upon the medium intervening between the electrodes D and E, and thereby alter, by electrostatic attraction, the separation between the electrodes. In this case D' may be a strip of platinum-foil, and the slightest approach thereof toward the filament will act to slightly cool the gaseous medium, and thereby alter the current in the local circuit, or, if D' is rigid, the increase in electrostatic attraction between D' and E will cause E to recede from D, and thereby alter the current in the local circuit.

In Fig. 3 the filament E is connected, by means of a minute platinum wire I, to the arm J which is secured to the coil *a*, placed between the poles of the magnet H and secured to the walls of the vessel A through the spiral springs *b b*. The line-current to be amplified in this case is passed through the coil *a* through the springs *b b*, and the resulting rotation of the coil varies the separation between the electrodes D E, thereby altering the current in the local circuit.

In Fig. 4 the currents to be amplified may be impressed upon the gaseous medium intervening between D' and E by means of the transformer M'. A condenser C may be included in series with the secondary of said transformer and the electrodes D' E. In this case there may or may not be a variation between the separation of the electrodes, and the currents to be amplified may vary the motions of the ions around the filament, thereby controlling to a greater degree the flux between said filament and the electrode D.

In Fig. 5 the currents to be amplified may be passed through the solenoid surrounding the magnet N S, and thereby vary the field, which by coacting with the magnetic field surrounding the electrode E determines the normal separation of the electrodes D and E. Even without creating actual movements between the electrodes D and E the varying magnetic flux produced by the line-current passing through the solenoid L will affect the motion of the ions in the gaseous medium, and thereby alter the current in the local circuit.

In Fig. 6 the electrode D may be constituted of iron or may consist of platinum plates provided with small iron armatures O O. In either case the currents to be amplified by passing through the solenoids P P, which surround the poles of the magnet N, effect the desired variation in the separation of the electrodes D and E.

In Fig. 7 the electrode D and diaphragm S may both be rigidly secured to the inside and outside, respectively, of the glass wall of the vessel A at a point where said wall has been flattened and made very thin, like the crystal of a watch. The currents to be amplified in this case by operating upon the coils P′ P′, surrounding the magnet Q, effect slight movements of the diaphragm S, and these movements are mechanically transmitted through the glass wall of the vessel A to the electrode D, thereby varying the current in the local receiving-circuit.

It will be obvious that the amplifying device, which constitutes the subject-matter of the present invention, is not limited in its use to any particular kind of electrical circuit or apparatus, but that it is capable of general application wherever an amplifying device is required. By way of example of its application to a wire telegraph or cable system I have shown the line F in Fig. 1 as including a telegraph transmitting-key T′ and source of vibratory current G. In Fig. 2 I have shown the line F as constituting the local circuit of a wireless-telegraph receiving system including the battery B″ and oscillation-detector T″, the latter being connected in series with an antenna V and the earth E′. In Fig. 4 I have shown the line F as constituting a telephone-circuit including the microphone-transmitter T‴ and battery B‴. In all instances it will be understood by those skilled in the art and without going into further detail that the signal-indicating device R, which is included in the local receiving-circuit, may be any device suitable for the purpose of reproducing the signal initiated in the line F

I do not limit myself to any of the specific embodiments of my invention herein described, inasmuch as many modifications will readily occur to those skilled in the art without departing from the principle of my invention.

I claim—

1. In a device for amplifying electrical currents, an evacuated vessel inclosing a sensitive conducting gaseous medium maintained in a condition of molecular activity, two electrodes sealed within said vessel, a local receiving-circuit associated with said electrodes, and means whereby the separation of said electrodes may be varied by the currents to be amplified.

2. In a device for amplifying electrical currents, an evacuated vessel, two electrodes sealed within said vessel, means for heating one of said electrodes, a local receiving-circuit associated with said electrodes, and means whereby the separation of said electrodes may be varied by the currents to be amplified.

3. In a device for amplifying electrical currents, an evacuated vessel, two electrodes sealed within said vessel, a circuit including a source of electric energy connected in series with one of said electrodes, a local receiving-circuit associated with said electrodes, and means whereby the separation of said electrodes may be varied by the currents to be amplified.

4. In a device for amplifying electrical currents, an evacuated vessel, three electrodes sealed within said vessel, means for heating one of said electrodes, a local receiving-circuit including two of said electrodes, and means for passing the current to be amplified between one of the electrodes which is included in the receiving-circuit and the third electrode.

5. In a device for amplifying electrical currents, an evacuated vessel inclosing a gaseous medium, means other than the received energy for maintaining said gaseous medium in a condition of molecular activity, means for impressing the currents to be amplified upon said gaseous medium, and a local receiving-circuit having its electrodes sealed within said vessel.

6. In a device for amplifying electrical currents, an evacuated vessel, a heated electrode and two non-heated electrodes sealed within said vessel, the non-heated electrodes being unequally spaced with respect to said heated electrode, a local receiving-circuit including said heated electrode and that one of the non-heated electrodes which has the greater separation from the heated electrode, and means for passing the current to be amplified between the heated electrode and the other non-heated electrode.

In testimony whereof I have hereunto subscribed my name this 17th day of October, 1906.

LEE DE FOREST.

Witnesses:
Ralph Polk Buell,
Sidney Williams

M0DAD RADIO WEBSITE

Basic Crystal Radio Set

I have been very interested in crystal radios for many years, I can remember building my first set back in my schooldays in the early 1980s. It was the first and still is the most basic of all radio receivers , If you are new to radio and have never seen one of these sets you will be very amazed at how easy it is to build a working radio needing so few components. What also may surprise a newcomer is that it does not need any battery or power to work, The only power it uses is taken from the transmitter of the station you are listening to, In other words it pulls the power from the radio wave.

The biggest craze of building and listening to crystal sets was at its peak in the 1920s. The old style sets used a real mineral crystal for the detector, The most popular being the galena and a fine bit of wire called a cats whisker which you would have to touch the crystal in the correct place to find the best reception where the radio station would come in. During world war II the underground people of Europe used crystal sets to listen to prohibited broadcast stations, Because of there simplicity to build. It is now getting harder to obtain all the original parts like the high impedance headphones, But I will show you how to build the modern day circuit using readily available parts.

My basic crystal radio project is aimed at the absolute beginner or younger generation, to try and interest them in the radio hobby as many amateur radio operators began there interest in radio with building the crystal set. The circuit can be soldered or built using chock block connectors for safety if a young person is doing this project.

CIRCUIT

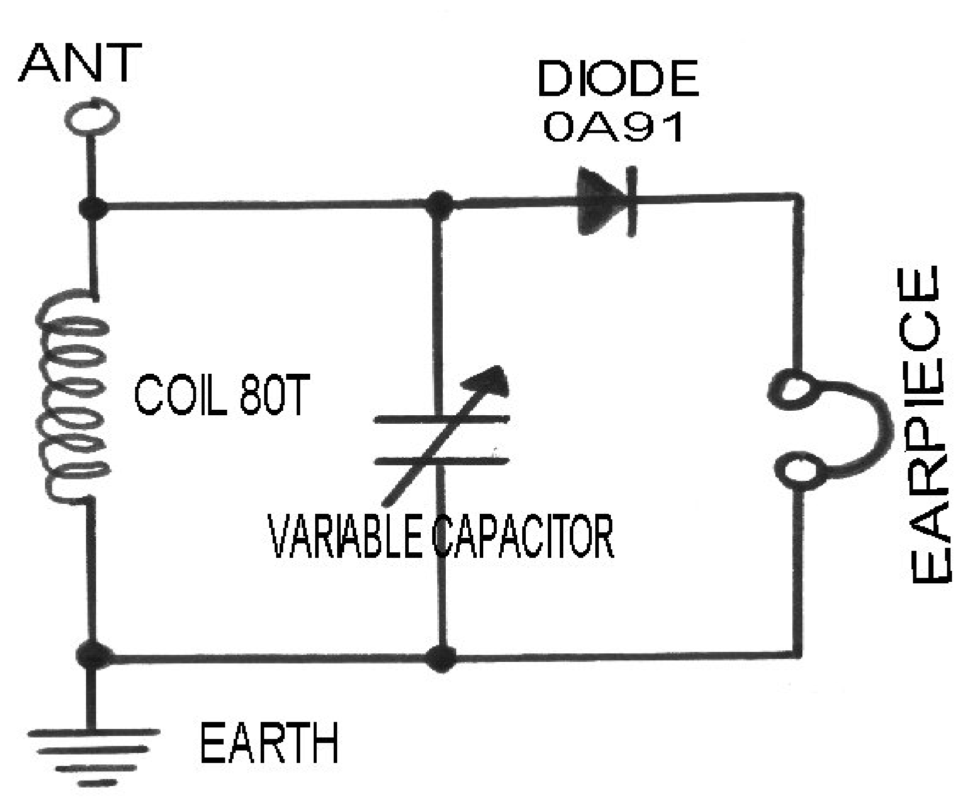
ANT
DIODE
0A91
COIL 80T
VARIABLE CAPACITOR
EARPIECE
EARTH

HOW THE CIRCUIT WORKS

First of all there are many different signals from stations which will be traveling down the long wire antenna, The coil and capacitor together form a tuned circuit, And because the capacitor is a variable type you can change the resonant frequency of the tuned circuit by rotating the tuning knob on the capacitor. Only signals from stations on or near the resonant frequency can pass through the tuned circuit, The other remaining frequencies are bypassed to earth. The passed signals then go through the germanium diode (detector) which rectifies the positive half cycles changing it to DC current and then sending it to the crystal earpiece or high impedance headphones.

BOARD LAYOUT

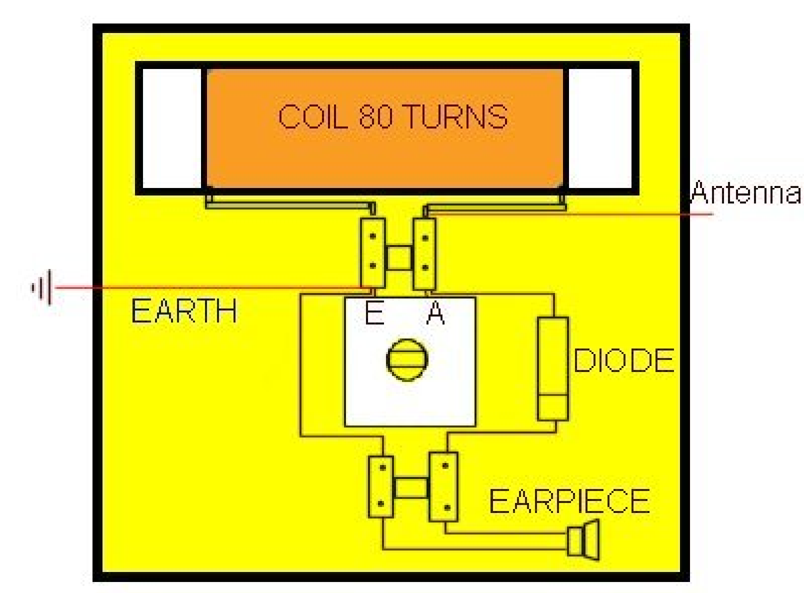
COIL 80 TURNS
Antenna
EARTH
E
A
DIODE
EARPIECE

PARTS

I personally find it a lot more fun to salvage all the parts from an old radio as it will cost you nothing, But just for this project I have listed all the parts that you will need and where to buy them from.

Variable capacitor, available from Maplin Electronics (UK) Code FT78K

Diode: Maplin Electronics 0A91, Code QH72ZP, 0A90 Code QH71N or any germanium point contact diode such as an OA70,OA71, or 1N34.

Chock block connectors Maplin Electronics (UK).

Crystal earplug, avail able from Maplin Electronics (UK) Code LB25C

26 s.w.g. Enameled copper wire.

Coil former: An empty card board toilet roll or an off cut piece of plastic waste pipe.

Base: Any piece of soft scrap wood about 4"x 5"

PVC coated wire for links.

A roll of any kind of wire PVC coated or copper wire for antenna and an earth connection.

CONSTRUCTION

Start the construction with the coil by securing the former (toilet roll, or PVC pipe) and then winding 80 turns of 26 s.w.g enameled copper wire around. When finished secure the windings with sticky tape or glue, Then scrape the wire enamel off both ends of the coil with a knife or sand paper until you can see a shiny tip. The coil can be secured to the wood base with two drawing pins.

Now the most difficult part is over , you can secure a couple of chock block connectors in front of the coil and the two behind the capacitor then wire the circuit as in the board layout, It is straight forward, but make sure you get the capacitor the correct way, the earth and antenna pins are marked e and a on the bottom, There is no connection to the pin marked o. The diode has a glass body with a band on to show the flow of the conducting current. Strip the plastic coating off the end of the crystal earpiece leads and insert them into the two rear chock blocks.

You are now ready to connect the antenna, Just string up the longest length of wire you can, run it from the set through a window and secure it to an insulated pole, tree or fence as high and as long as you possibly can to achieve the best results. If it is impossible for you to have an outdoor wire, String a wire several times around a room ceiling or an attic, It will still receive some stations but at reduced volume.

You will need an earth, run a wire from the set to a cold water pipe, connect it to the pipe or to a water tap with a crocodile clip, Or if you want a more permanent arrangement you can run a wire outside and connect it to an earth rod driven into the ground, This can be a length of copper pipe or brass rod as long and deep into the ground as you can. It is said that you must have a good earth for a crystal set to work, In practice I found it works slightly better but have also achieved good results without any earth.

OPERATING THE SET

This is the best part, When you are finished building the circuit connect the earth and antenna wires, pick up the earpiece and have a listen, You should be able to hear noise, turn the screw on the variable capacitor until you tune into a station, You should be able to hear several stations on the MW band, Depending on your location, length and how high the wire antenna is, Because this is a basic circuit you will soon notice that the set has poor selectivity, This means that sometimes you can hear two or more stations coming in together and may find it difficult to separate the one you want to listen to, That is if you can at all. If you can only receive one or two stations and the screw on the variable capacitor is fully turned, try unwinding about 4 or 5 turns off the coil. One of the most difficult requirements is having the patience, Some stations will come in loud and then may fade out altogether after a wile, If you leave it on the same frequency they can sometimes return as the propagation changes.

FAULT FINDING

If it does not work, Check your wiring with that of the board layout, And make sure all leads are not slack.

You must make sure you have a germanium point contact diode, other diodes will not work.

Make sure you have a proper crystal earpiece or high impedance headphones, Do not be confused with the standard white coloured magnetic earpiece, Unlike the standard 8 or 16 ohm earphones of today crystal earpieces are into the mega ohms, So if you used the wrong type it would be like trying to run your washing machine on a PP3 9 volt battery.

EXPERIMENTING

When I built this crystal set for the web site I hooked a wire to a radiator for an earth, strung a 60 FT length of PVC coated wire through a window down to the rear garden fence, The first station I heard was quite a surprise, Radio Korea, Other stations where Virgin Radio - 1215 KHz, Talk Sport 1071 KHz, Magic 1152 KHz, BBC Radio 5 Live 693 KHz, BBC Radio 4 - 603 KHz, So it had a good range through the medium wave broadcast band.

Do not be afraid to experiment, Due to the simplicity of the set you wont break it, You can try changing the diode around, Try other coils with different number of windings, The more turns should take you lower in frequency, and the smaller number higher, try a ferrite rod as a former or salvage a ready made coil from an old broken radio. Experiment with different lengths of antenna wire. I also find these sets work better with a higher quality air spaced variable capacitor which you can buy from a radio rally or you could try J. Birkett suppliers of electronic components who always stocks them at reasonable prices, TEL NO. +44 (0)1522 520767. The company is UK based and may not export, but I am sure that you will know of local electronics suppliers for such parts.

I hope you have enjoyed reading about the crystal set and if you have built the one from this page e mail me and let me know how it went. If this page has interested you and you would like to build more sets check out my other web pages or web links page.

BACK TO M0DAD HOMEPAGE

28th June 2013

Battery Free Loop FM Crystal Radio

Battery Free Loop FM Radio using an 8-mm diameter copper tube.

This FM radio runs without battery and you can listen to FM radio by just placing this close to your window.
If reception is not good, you can attach an external FM antenna (or with antenna amplifier), to the Loop at the point 1/3 of the loop close to the ground.

Parts you will need:
A Diode that works in high frequency (200Mhz or above, e.g. AA112, 1N34A, 1N48, Iss86 or Iss106.
A 0-12 pf AIR Variable capacitor - you can get a 0-22pf one, then remove all the copper plates and leave just 4.
50pf and 180pf fixed ceramic capacitor
100K 1/4W resistor
33uH Choke coil.
3.5mm mini headphone jack
A high impedance 4K Ohm earphone or a crystal earphone.
Some wires.

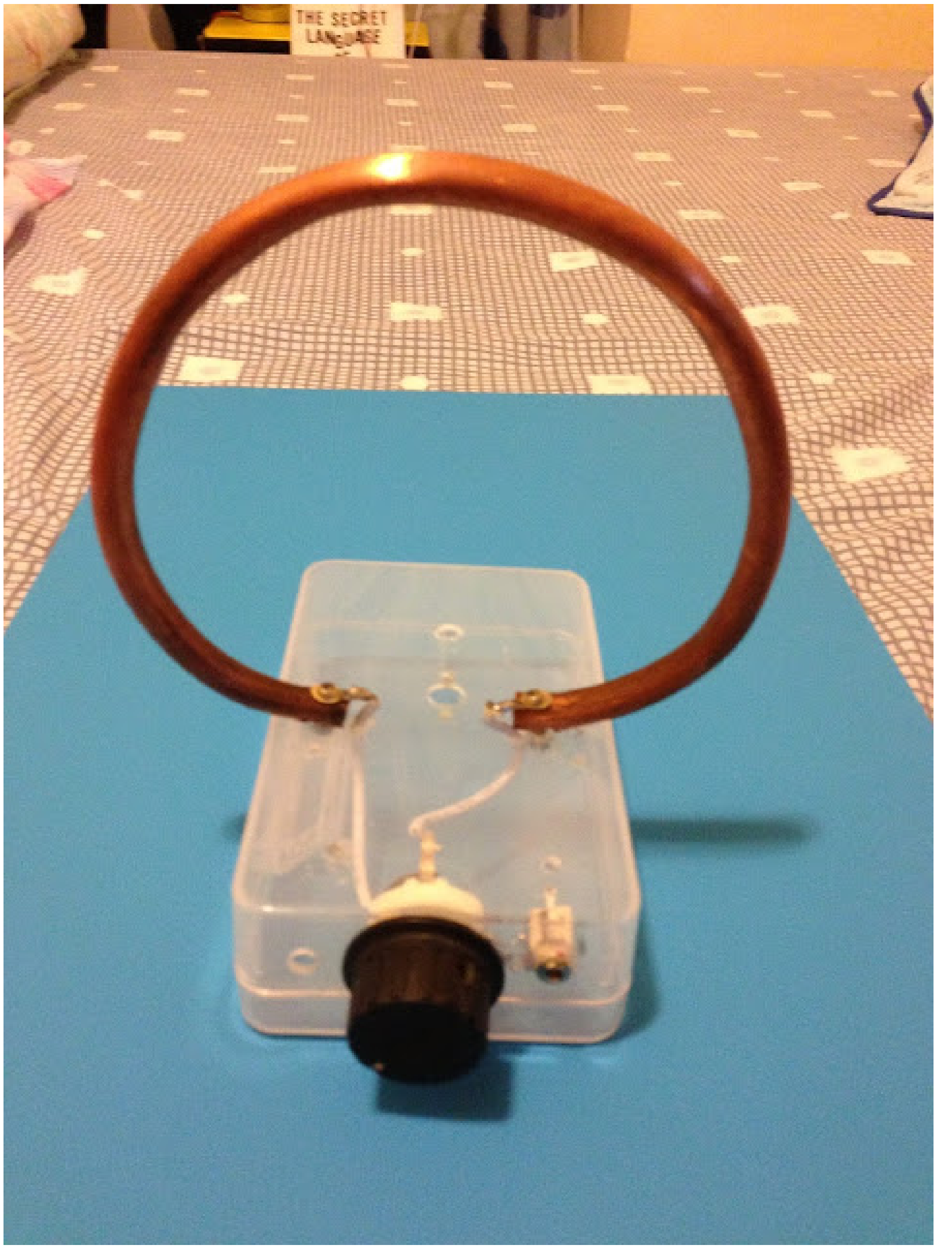

[http://2.bp.blogspot.com/-1QLVTR6aixI/Uc2EVLvYurI/AAAAAAAABLU/FHVgksSmYt0/s1600/IMG_2527.JPG]

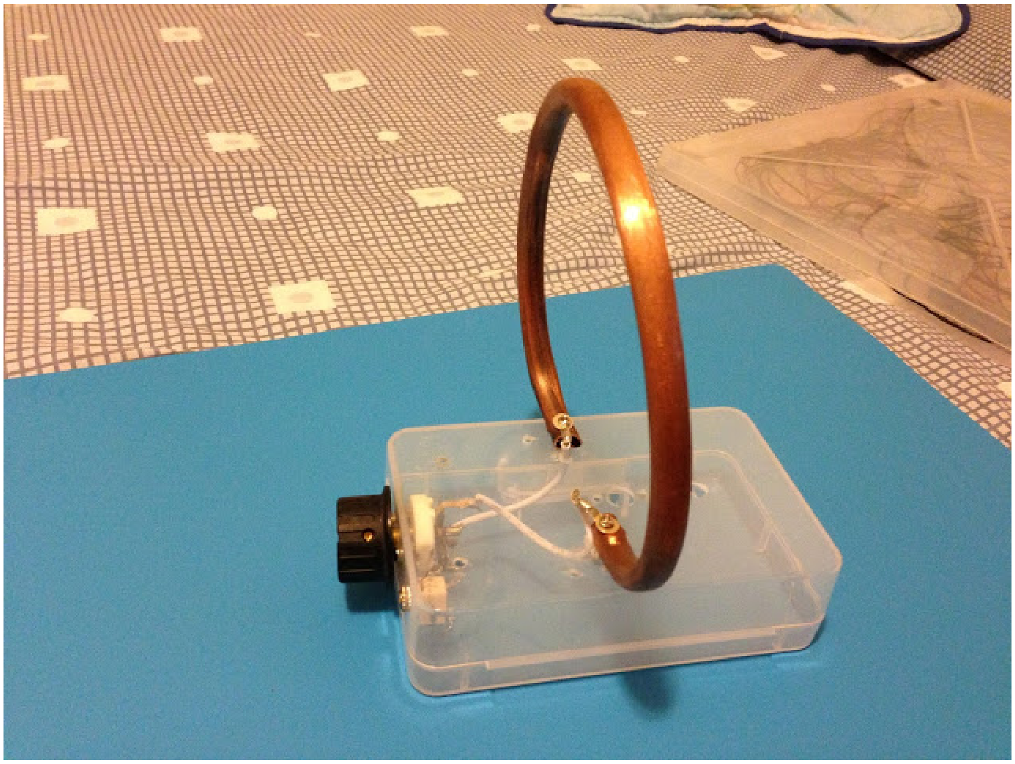

[http://1.bp.blogspot.com/-1DHVXlxHjMY/Uc2EVCFGUkI/AAAAAAAABLk/K03mBQ-9OXM/s1600/IMG_2528.JPG]

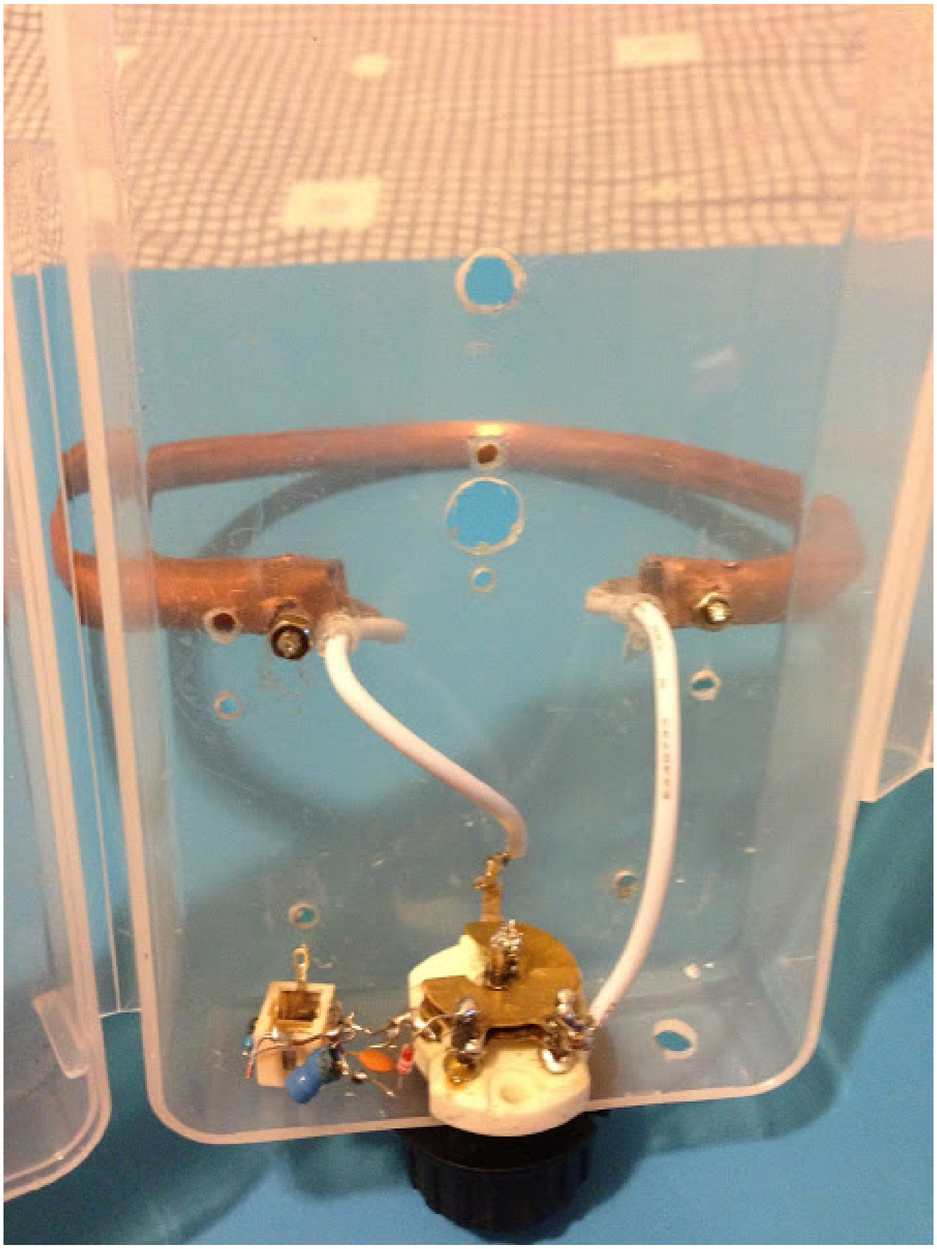

[http://1.bp.blogspot.com/-U8f34Fh_4EM/Uc2EVA6kIQI/AAAAAAAABLY/aI6OZ2c5Edc/s1600/IMG_2529.JPG]

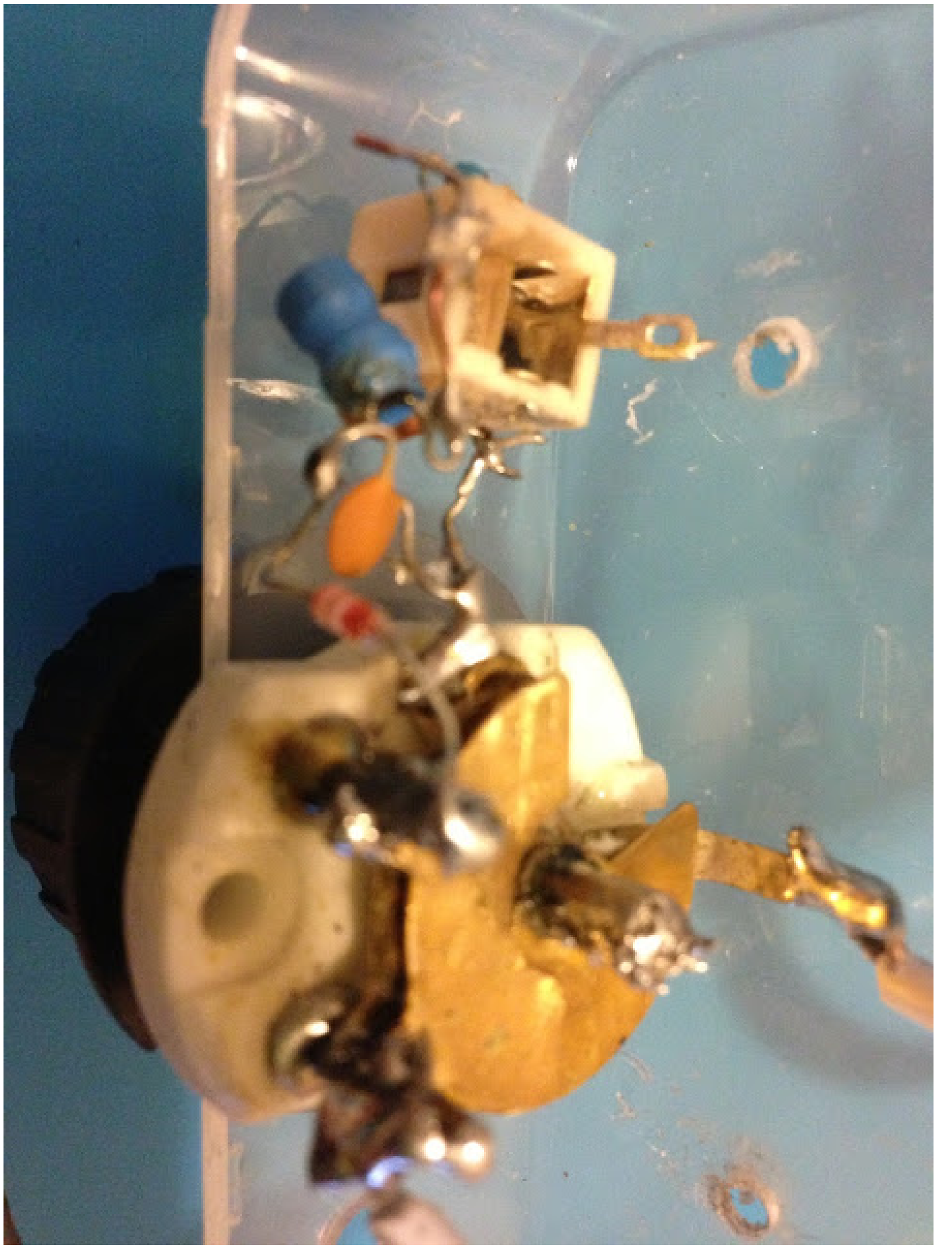

[http://1.bp.blogspot.com/-ONsnEMeXv-g/Uc2EWEFWNnI/AAAAAAAABLo/ap0dwwE1PKY/s1600/IMG_2531.JPG]

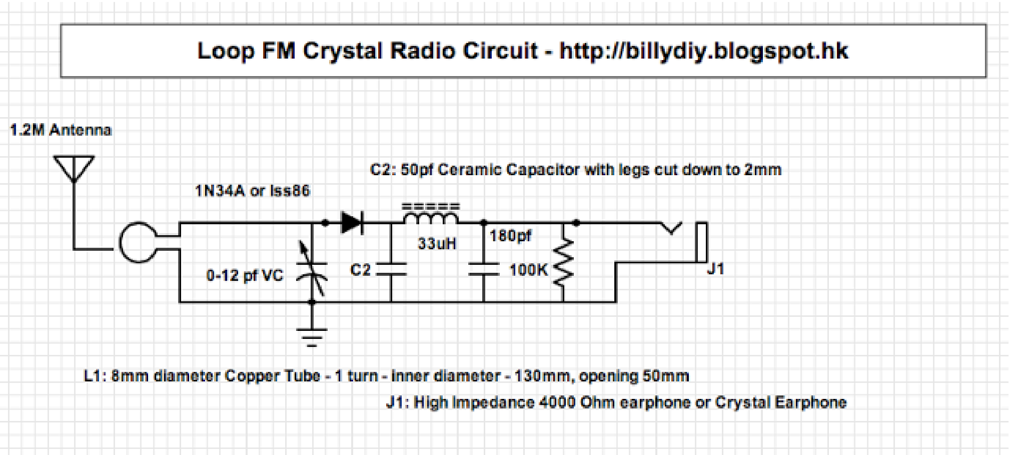
Loop FM Crystal Radio Circuit - http://billydiy.blogspot.hk
1.2M Antenna
1N34A or Iss86
C2: 50pf Ceramic Capacitor with legs cut down to 2mm
33uH
180pf
100K
0-12 pf VC
C2
J1
L1: 8mm diameter Copper Tube - 1 turn - inner diameter - 130mm, opening 50mm
J1: High Impedance 4000 Ohm earphone or Crystal Earphone

[http://1.bp.blogspot.com/-84ZQcpL-iDM/Uc2DeEbsVzI/AAAAAAAABLA/R_rFIJDIWn8/s763/Screen+Shot+2013-06-28+at+8.36.46+PM.png]

Posted 28th June 2013 by Billy

Free Printable Coupons
Free Coupons and Online Discounts. Print Your Free Coupons Now! get.couponunity.com/

Free Printable Forms
PDF, Word, PPT, Excel + More. Download YourTemplateFinder™ Now! yourtemplatefinder.com

0 Add a comment

Enter your comment...

Comment as: Karl Smirn ▼ Sign out

Publish Preview ☐ Notify me

US007490538B2

(12) United States Patent
Lowell et al.

(10) **Patent No.:** **US 7,490,538 B2**
(45) **Date of Patent:** **Feb. 17, 2009**

(54) **WEAPON HAVING LETHAL AND NON-LETHAL DIRECTED-ENERGY PORTIONS**

(75) Inventors: **Reid F. Lowell**, Ontario, CA (US); **Kenneth W. Brown**, Yucaipa, CA (US); **A-Lan V. Reynolds**, Tucson, AZ (US); **Alan A. Rattray**, Alta Loma, CA (US)

(73) Assignee: **Raytheon Company**, Waltham, MA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 428 days.

(21) Appl. No.: **11/207,049**

(22) Filed: **Aug. 18, 2005**

(65) **Prior Publication Data**

US 2007/0040725 A1 Feb. 22, 2007

(51) **Int. Cl.**
F41H 13/00 (2006.01)
F41C 27/00 (2006.01)

(52) **U.S. Cl.** **89/1.11**; 42/1.08; 42/84

(58) **Field of Classification Search** 89/1.11; 42/1.08, 84
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,325,305 A *	4/1982	Pastine	....	102/201
6,487,950 B2 *	12/2002	Samland	....	89/1.13
6,799,499 B2 *	10/2004	Seregelyi et al.	....	89/1.13
6,843,178 B2 *	1/2005	Wood et al.	....	102/275.5
2005/0156743 A1	7/2005	Gallivan et al.		

FOREIGN PATENT DOCUMENTS

EP	1396699 A2	3/2004	
WO	WO-2007022339 A2	2/2007	

OTHER PUBLICATIONS

GlobalSecurity.org, Vehicle-Mounted Active Denial Technology (V-MADS), Internet Archived as shown in box at top og page, Jan. 7, 2003, 6 pages.*

"International Search Report and Written Opinion- Application No. PCT/US2006/032117", (Feb. 2007).

Hackett, K. E., et al., "Active Denial Technology", http://www.afrlhorizons.com/Briefs/Sept01/De0101.html>, Internet Article,(Retrieved from the Internet Jan. 22, 2007).

* cited by examiner

Primary Examiner—Bret Hayes
(74) *Attorney, Agent, or Firm*—Schwegman, Lundberg & Woessner, P.A.; Gregory J. Gorrie

(57) **ABSTRACT**

A portable weapon comprises a non-lethal portion and a lethal portion. The lethal portion may comprise a rifle, and the non-lethal portion may comprise a millimeter-wave directed energy weapon. The non-lethal portion may comprise a kit to add non-lethal capability to a lethal weapon. The non-lethal portion may comprise an output antenna to generate a high-power millimeter-wave initial wavefront, a main reflector, and a sub-reflector to reflect the initial wavefront to the main reflector. The main reflector may direct the reflected wavefront in a bore-sighted direction toward a target. The wavefront directed by the main reflector may have a power density selected to deliver a non-lethal deterring effect on the target. In some embodiments, the non-lethal portion may include a replaceable energy-storage module.

23 Claims, 8 Drawing Sheets

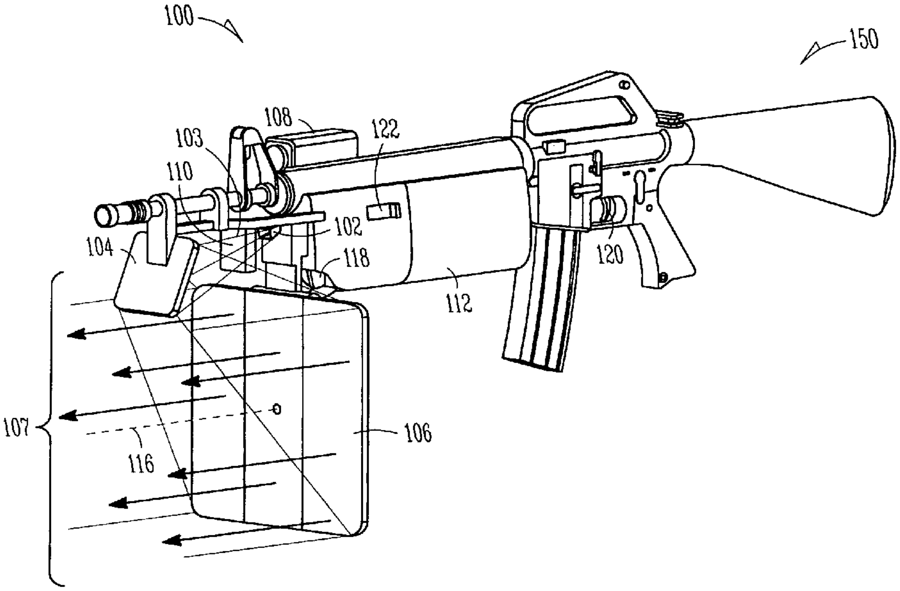

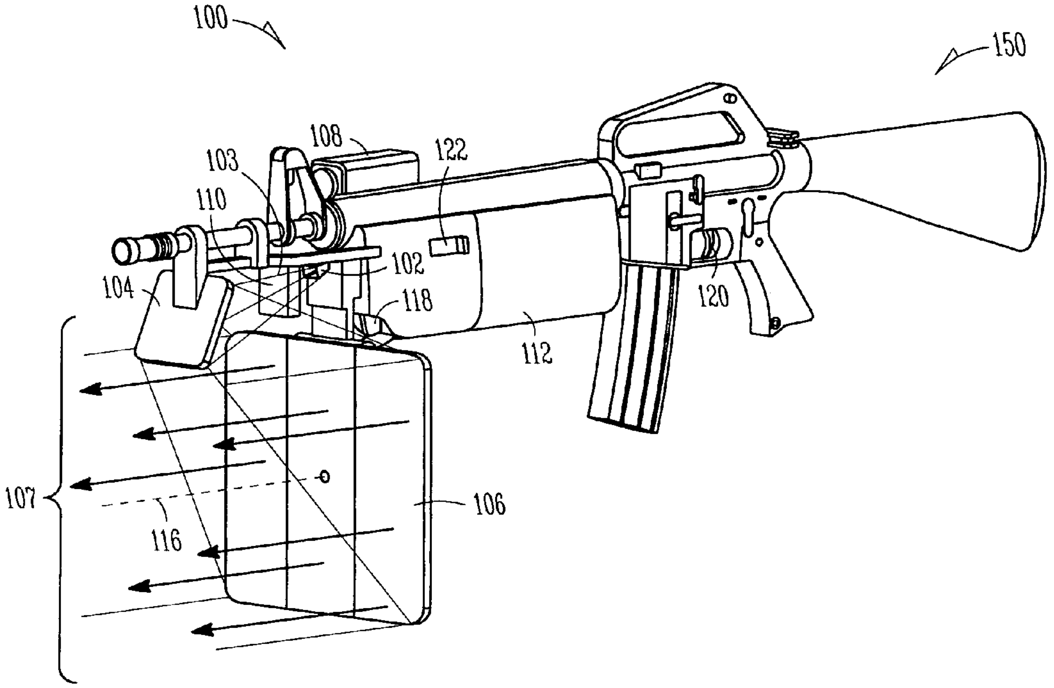
100
150
108
122
103
110
102
104
118
120
112
107
116
106

FIG. 1

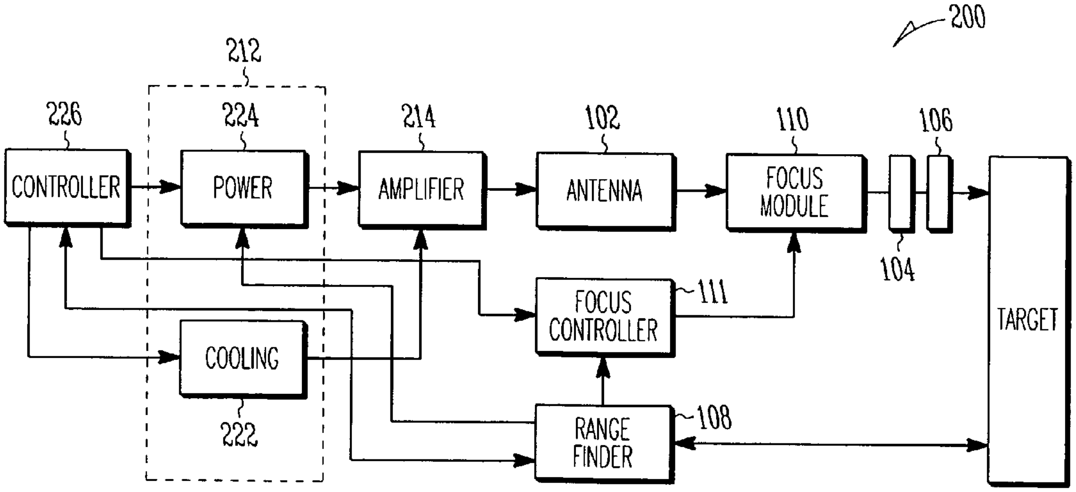
200
212
226
224
214
102
110
106
CONTROLLER
POWER
AMPLIFIER
ANTENNA
FOCUS MODULE
104
TARGET
FOCUS CONTROLLER
111
COOLING
222
RANGE FINDER
108

NON-LETHAL PORTION

FIG. 2

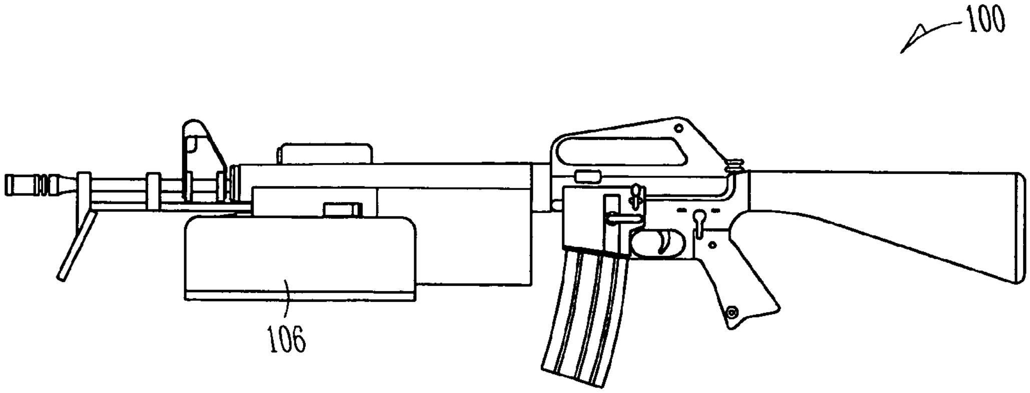

FIG. 3A

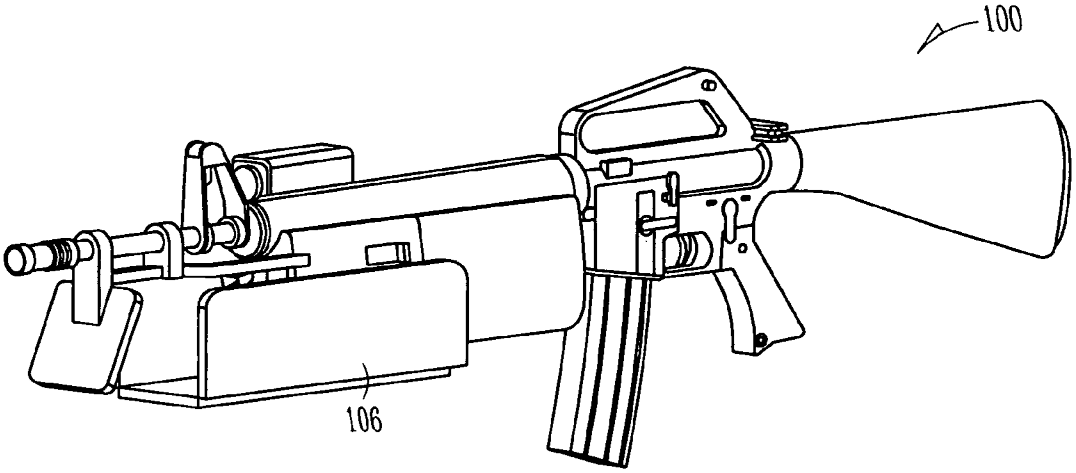

FIG. 3B

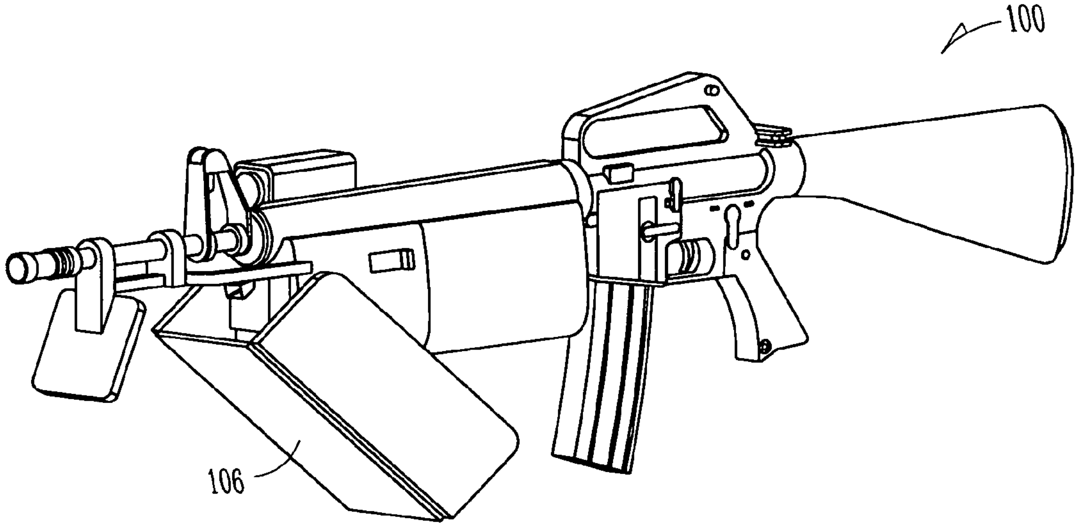

FIG. 3C

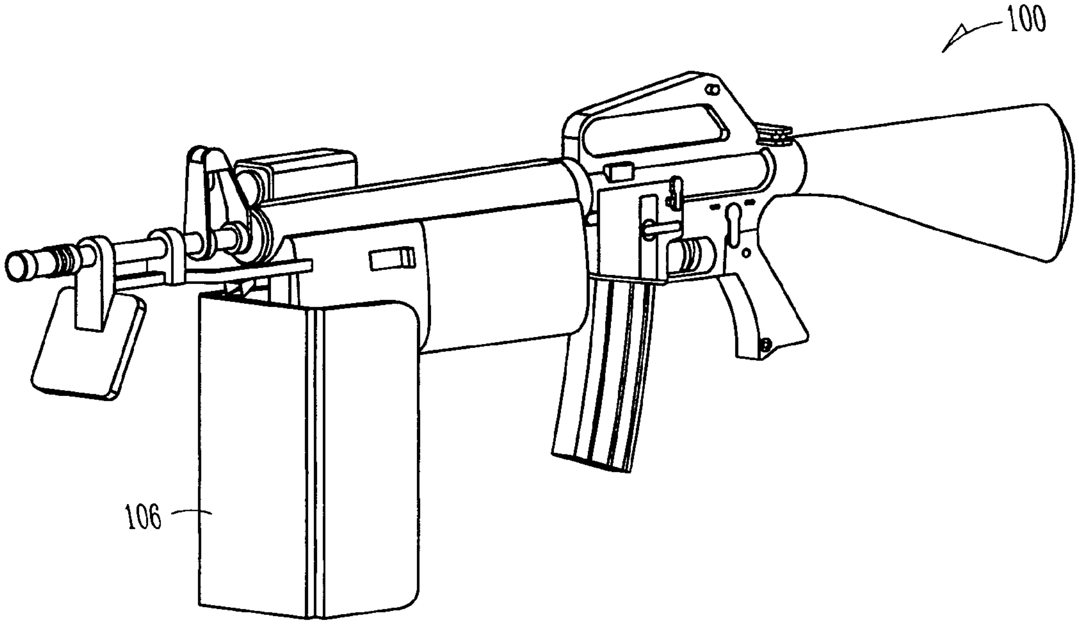

FIG. 3D

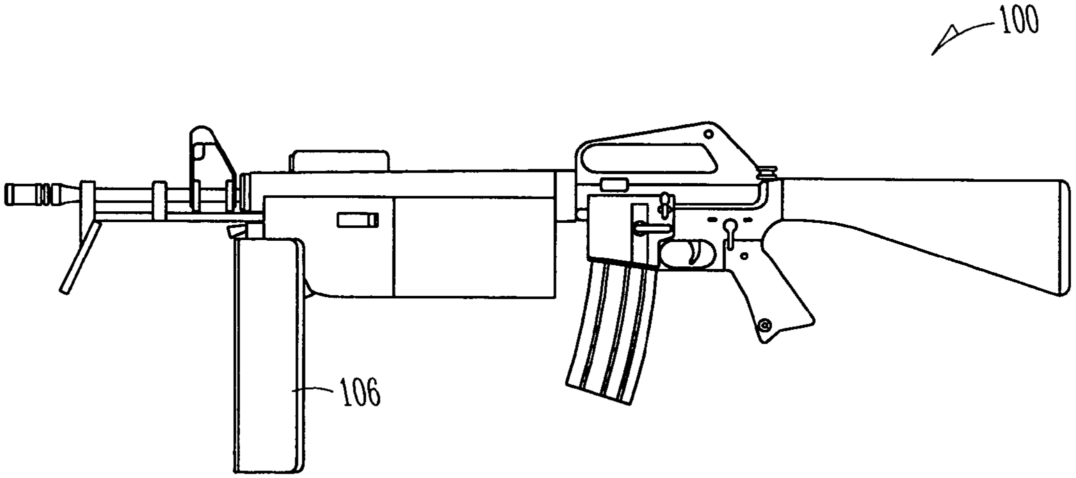

FIG. 3E

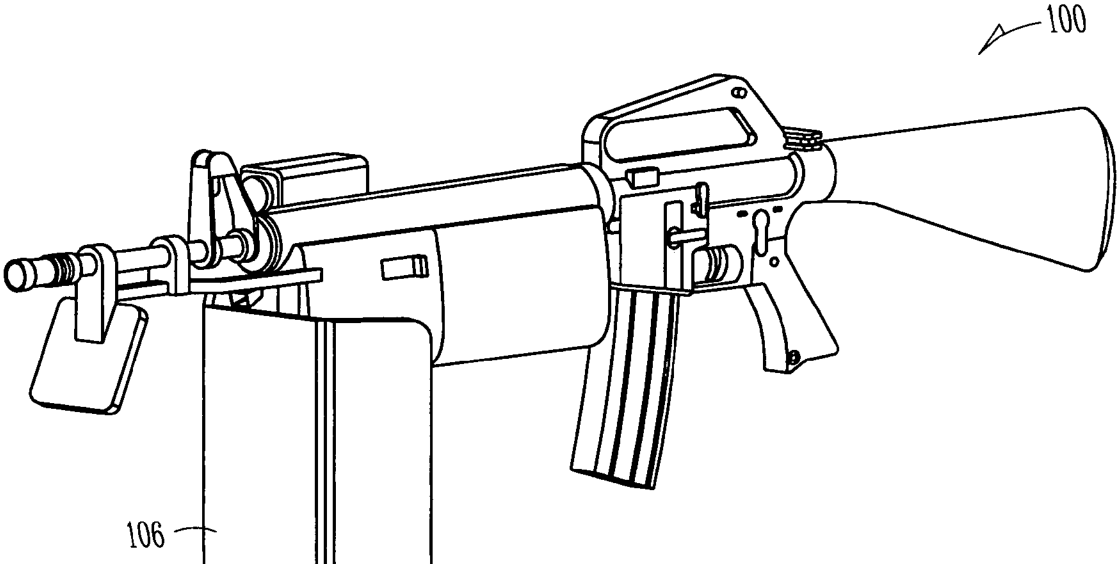

FIG. 3F

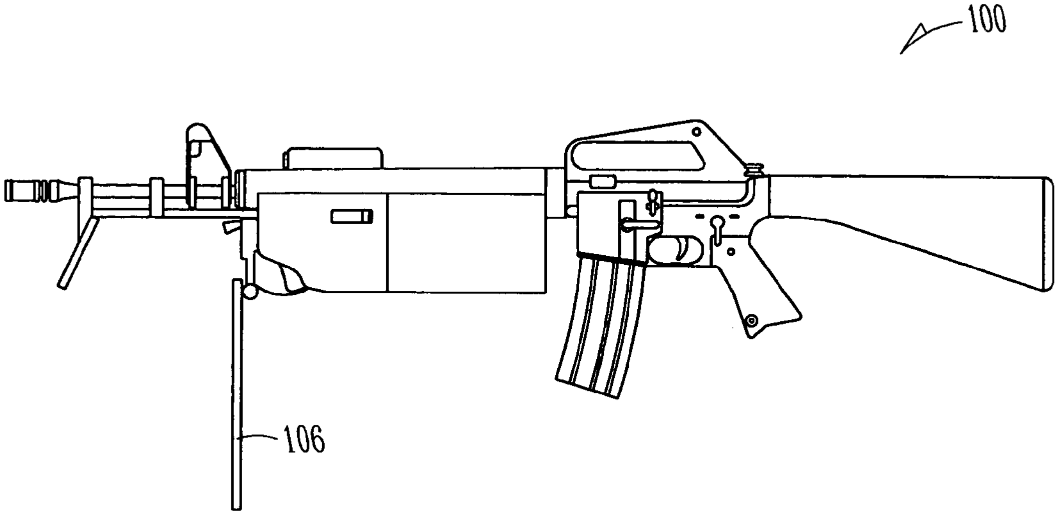

FIG. 4A

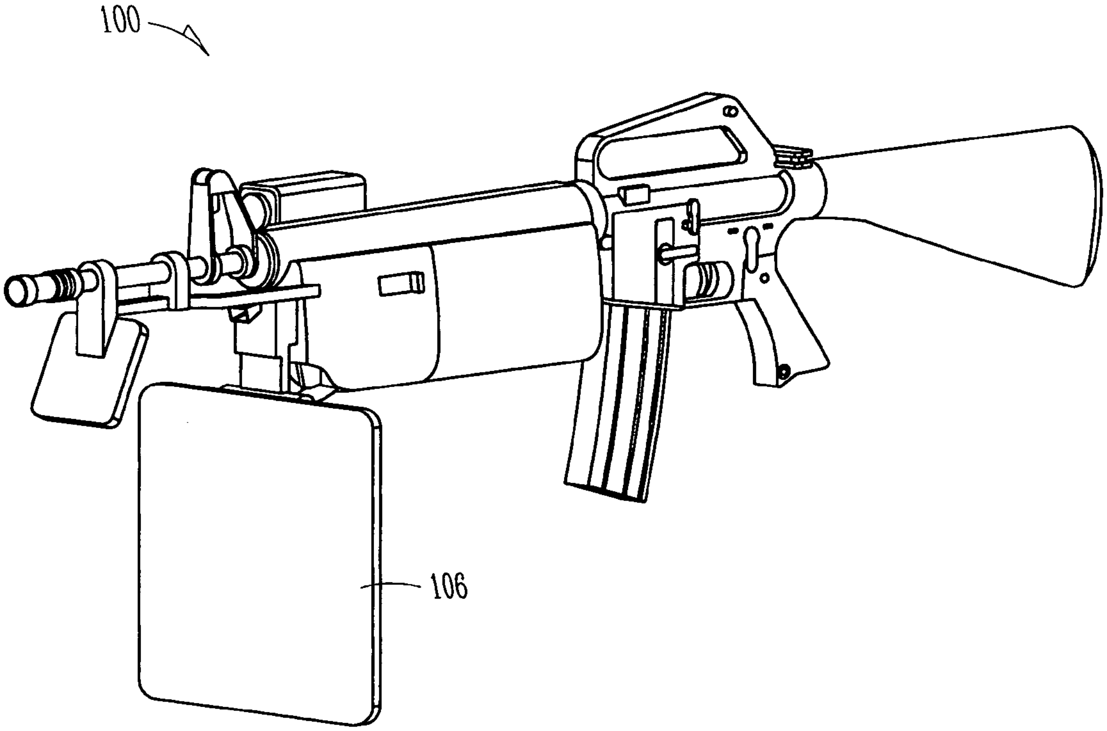

FIG. 4B

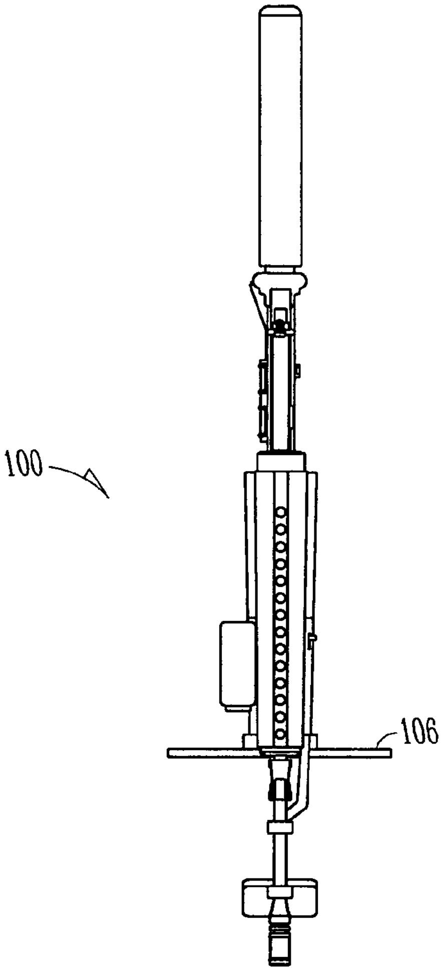

FIG. 4C

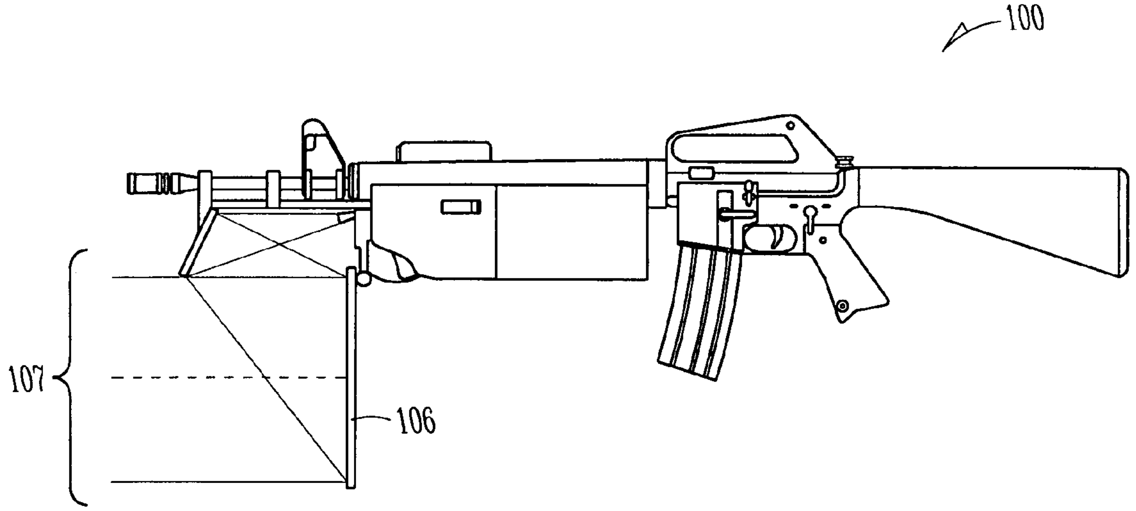

FIG. 5A

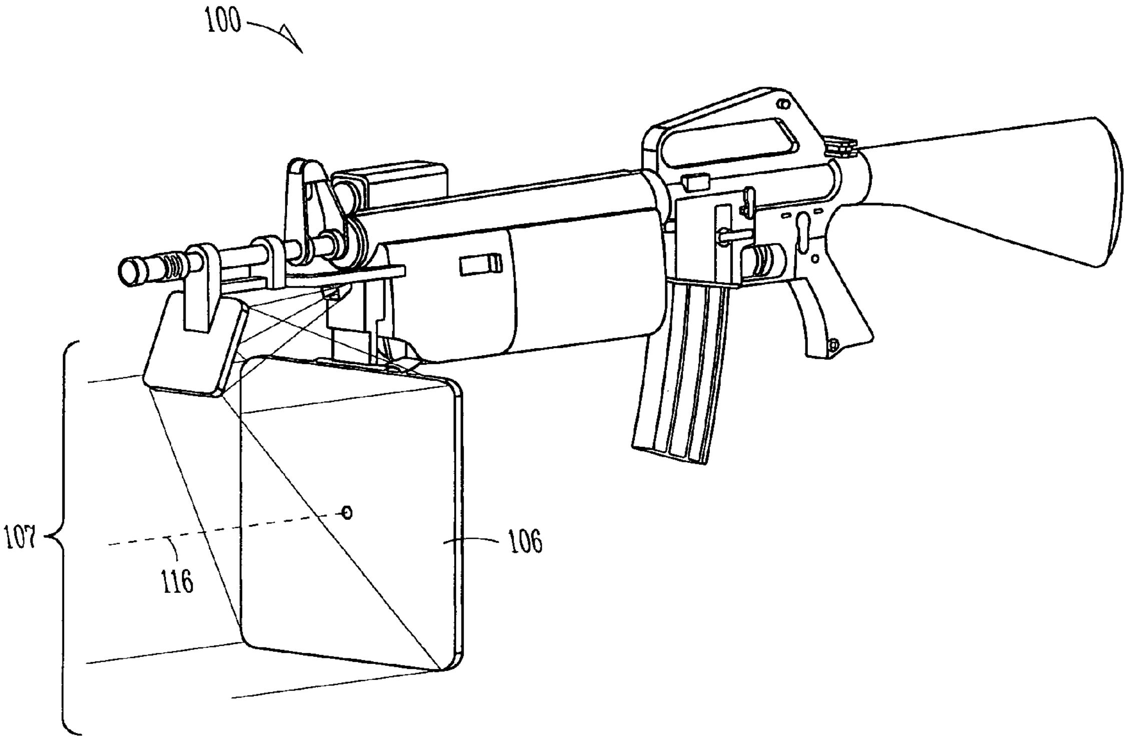
100
107
116
106

FIG. 5B

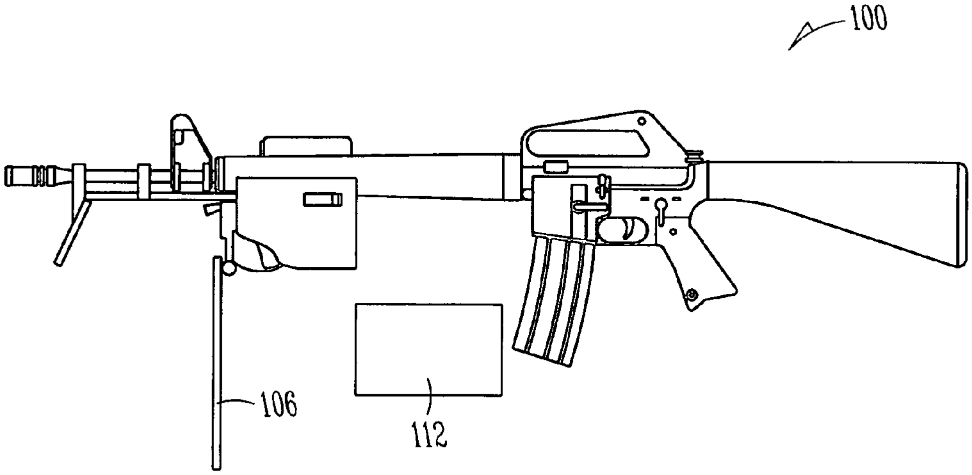

FIG. 6A

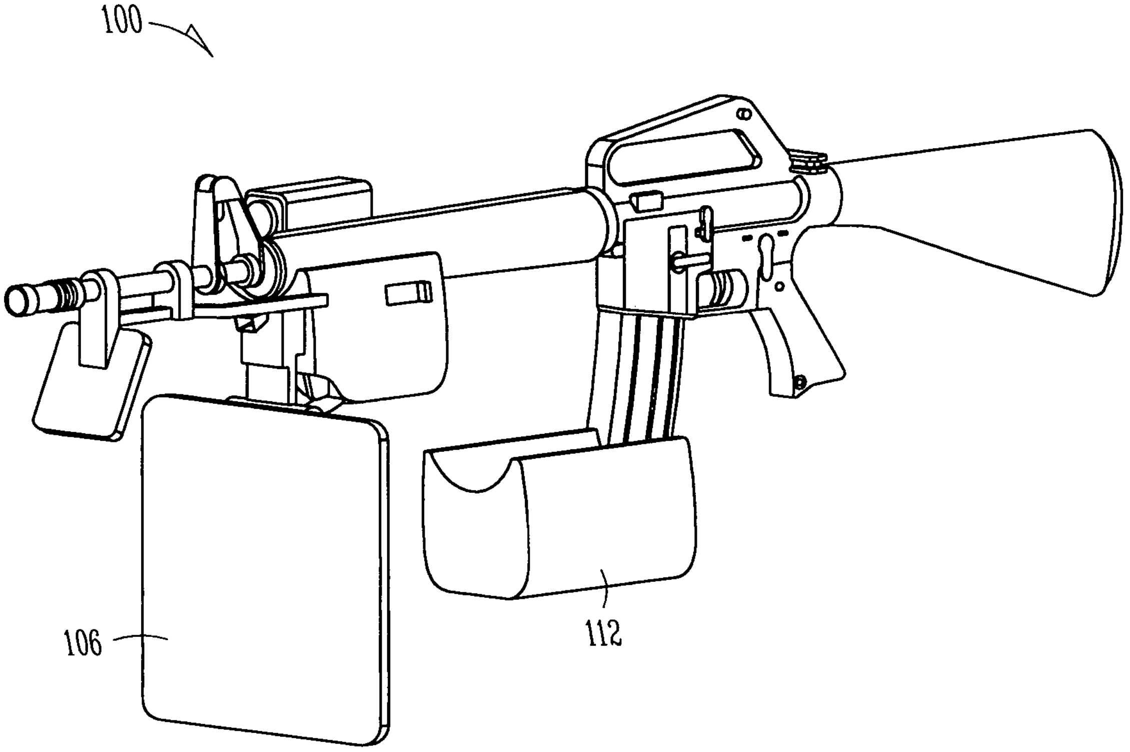

FIG. 6B

WEAPON HAVING LETHAL AND NON-LETHAL DIRECTED-ENERGY PORTIONS

TECHNICAL FIELD

Embodiments of the present invention pertain to portable weapons. Some embodiments pertain to directed-energy weapons and some embodiments pertain to kits for adding non-lethal capability to lethal weapons.

BACKGROUND

There are many difficult quickly changing situations in modern urban conflicts that security personnel must deal with. In modern urban conflicts, security personnel must execute dynamically changing missions that could shift rapidly between direct action, security patrols and civil stability. Among the mix of unarmed civilians, non-lethal combatants (e.g., rock throwing) and lethal combatants, it is often not immediately clear who is an innocent bystander and who poses an immediate threat to security personnel. Options for security personnel many times progress quickly from shouting to shooting. Modern urban conflicts many times require a delicate balance between the use of non-lethal force and the use of lethal force. Non-lethal weapons, when available, are generally carried separate from lethal weapons resulting in a potentially life-threatening delay for security personnel when switching between the types of weapons. Urban riot situations, for example, can easily escalate in a moment's notice and require security personnel to switch between a non-lethal response and a lethal response.

One problem with many non-lethal weapons is that they are largely ineffective over the range that lethal weapons are effective. For example, a non-lethal kinetic weapon that sends projectiles (e.g., rubber bullets) must have a reasonable range to maintain its nonlethality, however, the weapon becomes potentially lethal at close range when powerful enough to be used for longer ranges due to the initial velocity required to project the projectile over these longer ranges.

Thus, there are general needs for a non-lethal weapon that can easily be deployed along with a lethal weapon. There are also needs for a combined lethal/non-lethal weapon that has an effective non-lethal range comparable to its lethal range. There are also needs for a combined lethal/non-lethal weapon that allows security personnel to easily and quickly switch between non-lethal and lethal capabilities.

SUMMARY

A weapon having non-lethal and lethal portions is provided. The non-lethal portion directs a high-power millimeter-wave wavefront toward a target. The non-lethal portion comprises an output antenna to direct a high-power millimeter-wave initial wavefront at a sub-reflector, and a main reflector to reflect the wavefront to the target. The high-power wavefront may produce a non-lethal deterring effect on the target. The main reflector may be bore-sighted with the lethal weapon portion of the weapon to easily allow switching between non-lethal and lethal capabilities.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a perspective view of a weapon in accordance with some embodiments of the present invention;

FIG. **2** is a functional block diagram of the non-lethal portion of the weapon of FIG. **1** in accordance with some embodiments of the present invention;

FIG. **3**A is a side view illustrating the main reflector in a fully folded-up position in accordance with some embodiments of the present invention;

FIG. **3**B is a perspective view illustrating the main reflector in a fully folded-up position in accordance with some embodiments of the present invention;

FIG. **3**C is a perspective view illustrating the main reflector in a partially folded-up position in accordance with some embodiments of the present invention;

FIG. **3**D is a perspective view illustrating wings of the main reflector folded-up in accordance with some embodiments of the present invention;

FIG. **3**E is a side view illustrating wings of the main reflector partially folded-up in accordance with some embodiments of the present invention;

FIG. **3**F is a perspective view illustrating wings of the main reflector partially folded-up in accordance with some embodiments of the present invention;

FIG. **4**A is a side view illustrating the main reflector in a fully-deployed position in accordance with some embodiments of the present invention;

FIG. **4**B is a perspective view illustrating the main reflector in a fully-deployed position in accordance with some embodiments of the present invention;

FIG. **4**C is a top view of the weapon illustrated in FIGS. **4**A and **4**B;

FIG. **5**A is a side view illustrating the operation of the non-lethal portion of the weapon in accordance with embodiments of the present invention;

FIG. **5**B is a perspective view illustrating the operation of the non-lethal portion of the weapon in accordance with embodiments of the present invention;

FIG. **6**A is a side view illustrating the removable energy-storage module in accordance with embodiments of the present invention; and

FIG. **6**B is a perspective view illustrating the removable energy-storage module in accordance with embodiments of the present invention.

DETAILED DESCRIPTION

The following description and the drawings illustrate specific embodiments of the invention sufficiently to enable those skilled in the art to practice them. Other embodiments may incorporate structural, logical, electrical, process, and other changes. Examples merely typify possible variations. Individual components and functions are optional unless explicitly required, and the sequence of operations may vary. Portions and features of some embodiments may be included in or substituted for those of others. Embodiments of the invention set forth in the claims encompass all available equivalents of those claims. Embodiments of the invention may be referred to, individually or collectively, herein by the term "invention" merely for convenience and without intending to limit the scope of this application to any single invention or inventive concept if more than one is in fact disclosed.

A weapon in accordance with some embodiments of the present invention combines lethal capability with non-lethal capability allowing a user to easily switch between lethal and non-lethal force in a moment's notice. In many urban conflict situations, this ability may help save the lives of security personnel as well as the lives of innocent non-combatants. The non-lethal portion uses directed energy which, unlike many other non-lethal weapons (e.g., rubber bullets, taser, water cannons), generally causes no residual damage to a person fired upon. Because energy is the ammunition, the logistical burdens associated with conventional non-lethal weapons are significantly reduced.

FIG. **1** is a perspective view of a weapon in accordance with some embodiments of the present invention. Weapon **100** comprises non-lethal portion and lethal portion **150**. Lethal portion **150** may be any lethal weapon including a rifle or machine gun. The non-lethal portion may comprise a directed energy weapon and may be bore-sighted or aligned with the lethal portion. In some embodiments, the non-lethal portion may be a kit allowing non-lethal capability to be added to a lethal weapon, although the scope of the invention is not limited in this respect.

Non-lethal portion of weapon **100** may comprise output antenna **102** to transmit a high-power millimeter-wave initial wavefront **103**, main reflector **106**, and sub-reflector **104** to reflect initial wavefront **103** to main reflector **106**. Main reflector **106** may direct wavefront **107** in a bore-sighted direction toward a target.

In some embodiments, non-lethal portion of weapon **100** may also comprise rangefinder **108** to determine a range to the target, focus module **110** to focus wavefront **107**, replaceable energy-storage module **112** to provide energy to the non-lethal portion, and laser designator **116** for designating the target. In some embodiments, non-lethal portion may also include trigger **118** to cause the generation of wavefront **107** and on-off switch **122**. These elements are discussed in more detail below.

In some embodiments, wavefront **107** may comprise a millimeter-wave frequency, such as a W-band frequency between 94 and 96 GHz, although the scope of the invention is not limited in this respect. In some embodiments, wavefront **107** directed by main reflector **106** may have a power density selected to deliver a non-lethal deterring effect on the target. In some embodiments, wavefront **107** comprises a frequency selected to penetrate a shallow skin-depth (i.e., of less than five millimeters). In some embodiments, a power density of wavefront **107** at the target may be calculated and selected to cause a deterring effect by inducing pain on human skin. In some embodiments, wavefront **107** may comprise W-band millimeter-wave or higher frequency radiation selected to penetrate only a shallow skin-depth allowing the energy to heat the region of the skin's pain sensors, although the scope of the invention is not limited in this respect.

In some embodiments, main reflector **106** may be a collimating reflector to generate a collimated wavefront toward the target. The collimated wavefront may be substantially uniform in amplitude and/or substantially coherent in phase in a planar cross-section of a column of energy emanating from main reflector **106**, although the scope the invention is not limited in this respect.

In some other embodiments, main reflector **106** may generate a converting wavefront which may converge at or near an intended target. In these embodiments, a convergence distance may be selected to provide a predetermined power density at or near a surface of the target. In some embodiments, main reflector **106** may generate a slightly diverging wavefront. In some embodiments, the focus of wavefront **107** (i.e., whether wavefront **107** is converging, collimated or diverging) may be at least partially controlled by focus module **110**.

In some embodiments, main reflector **106** may be aligned with sights of lethal portion **150** of weapon **100**. In some embodiments, the non-lethal portion may be a bore-sighted attachment kit to add non-lethal capability to a lethal weapon. The kit may be a "B-kit" add-on to a rifle, such as an M-16 rifle, although the scope of the invention is not limited in this respect.

FIG. **2** is a functional block diagram of the non-lethal portion of the weapon of FIG. **1** in accordance with some embodiments of the present invention. Elements illustrated in FIG. **2** having the same reference number as elements in FIG. **1** may refer to the same element. As illustrated in FIG. **2**, non-lethal portion **200** may include system controller **226** to control the operation of the various elements of non-lethal portion **200**. Rangefinder **108** may determine a distance to the target and system controller **226** may determine the proper power density of wavefront **107** at or near the target based on the distance. System controller **226** may control the RF power output of amplifier **214** accordingly. On-off switch **122** may allow power to be turned off to the active elements of the non-lethal portion.

Referring to FIGS. **1** and **2** together, in some embodiments, laser designator **116** may be used to visually designate the target. Laser designator **116** may be bore-sighted with both the lethal portion and non-lethal portion. In some embodiments, rangefinder **108** may be aligned with laser designator **116**. In some embodiments, rangefinder **108** comprises a laser-rangefinder, although the scope of the invention is not limited in this respect.

In some embodiments, laser designator **116** may generate a laser-beam in parallel to wavefront **107** and may comprise a laser-diode mounted on main reflector **106**. In some embodiments, laser designator **116** may shine through a small hole in main reflector **106**. In some alternate embodiments, a laser diode may be provided at or near output antenna **102** and a laser-beam may be reflected by optically reflective portions on reflectors **104** and **106** and may be provided parallel to wavefront **107**.

Focus module **110** may change a focus of initial wavefront **103** generated by output antenna **102** based on a distance to the target. This may allow the power-density of wavefront **107** to be adjusted based on the distance to the target. In some embodiments, focus module **110** may be moveable by a user allowing the user to select a position for a focusing element based on the distance to the target. In some embodiments, the focusing element may be manually slidable by a user.

In some embodiments, focus module **110** comprises a millimeter-wave radio-frequency (RF) lens that may be positioned by focus controller **111** based on a distance to a target. In some embodiments, focus module **110** may include one or more RF lenses that may be switched in and out of the RF path by focus controller **111** to focus wavefront **107**. In some embodiments, focus controller **111** may change the relative position of sub-reflector **104** to focus wavefront **107**. In some other embodiments, system controller **226** may change the phasing of electronic phase shifters within main reflector **106** to change the focusing and phase distribution of wavefront **107**.

In yet some other embodiments, focus module **110** may be an active-lens array in which a plurality of active array elements receive the wavefront, amplify the wavefront and retransmit the wavefront. In some embodiments, focus module **110** may provide a continually variable focusing distance, while in other embodiments; focus module **110** may provide selectable discrete focusing steps. In some embodiments, focus controller **111** and/or system controller **226** may configure main reflector **104**, sub-reflector **104** and/or focus module **110** to generate a collimated wavefront, while in other embodiments; focus controller **111** and/or system controller **226** may configure main reflector **106**, sub-reflector **104** and/or focus module **110** to generate a converting wavefront. In some other embodiments, focus controller **111** and/or system controller **226** may configure main reflector **104**, sub-reflector **104** and/or focus module **110** to generate a slightly diverging wavefront.

In some embodiments, rangefinder **108** and/or focus module **110** are optional. In these embodiments, the focus of the non-lethal portion may be set at a predetermined distance or at infinity. In these embodiments, the power output of amplifier **214** may be varied, although the power output may also be set to a predetermined level.

In some embodiments, focus controller **111** may change a focus of focus module **110** in response to changes in a distance to the target provided by rangefinder **108**. In some of these embodiments, a convergence point of wavefront **107** may be selected by system controller **226** to generate a predetermined power density at or near a target.

In some embodiments, amplifier **214** may be high-power millimeter-wave amplifier coupled to output antenna **102** to generate a high-power RF signal. In some embodiments, amplifier **214** may comprise a solid-state millimeter-wave amplifier comprising a plurality of either Silicon (Si) or Gallium-Arsenide (GaAs) semiconductor amplifier elements, although the scope of the invention is not limited in this respect. In some other embodiments, amplifier **214** comprises vacuum tube amplifier elements, although the scope of the invention is not limited in this respect. In some embodiments, additional amplifiers may not be needed when a powerful enough tube source is used for amplifier **214**.

In some embodiments, output antenna **102** comprises a horn antenna and initial wavefront **103** may be a substantially spherical wavefront. In some embodiments, initial wavefront **103** may be generated from a pulsed W-band millimeter-wave signal generated by amplifier **214**, although the scope of the invention is not limited in this respect.

Replaceable and removable energy-storage module **212** may provide electrical energy for the millimeter-wave amplifier **214** and/or other elements of the non-lethal portion. In some embodiments, energy-storage module **212** comprises power element **224** which may include, for example, either batteries or a fuel cell. In some embodiments, energy-storage module **212** may comprise a disposable battery or power pack, although the scope of the invention is not limited in this respect.

In some embodiments, weapon **100** may further comprise cooling element **222** to cool amplifier **214**. In some embodiments, cooling element **222** may be part of replaceable energy-storage module **212**, although the scope of the invention is not limited in this respect.

In some embodiments, cooling element **222** may circulate a phase-change fluid to cool amplifier **214**. In some embodiments, the phase-change fluid may comprise a refrigerant, although the scope of the invention is not limited in this respect. In some embodiments, cooling element **222** may circulate a coolant and may include a reservoir to store the fluid. In some other embodiments, cooling element **222** may comprise a semiconductor-based thermo-electric cooling (TEC) element to remove heat from amplifier **214** using electric current, although the scope of the invention is not limited in this respect.

In some embodiments, cooling element **222** may use an expanding gas to cool the amplifier **214**. In these embodiments, cooling element **222** may include a pressurized chamber to store the gas. In these embodiments, the chamber, including the gas therein, may be replaceable and may be replaced as part of energy-storage module **212**. In these embodiments, the gas may be contained in a pressurized chamber that may be replaced when energy-storage module **212** is replaced. In some embodiments, the gas may comprise carbon-dioxide (CO2), although the scope of the invention is not limited in this respect.

Although FIG. **2** illustrates replaceable module **212** as including power element **224** and cooling element **222**, the scope of the invention is not limited in this respect. In some embodiments, replaceable module may also include other elements of non-lethal portion **200**.

In some embodiments, weapon **100** may further comprise lethal-weapon trigger **120** to fire the lethal portion of the weapon and non-lethal-weapon trigger **118** to fire the non-lethal portion by generating wavefront **107**.

In accordance with some embodiments, non-lethal portion may operate as follows. Switch **122** may be turned on providing power to amplifier **214** and other elements of non-lethal portion and allowing range-finder **108** to determine a distance to a target. Laser-designator **116** may also be activated to designate the target to the user, however in some embodiments; laser-designator **116** may part of lethal portion **150** and may operate independent of the non-lethal portion. Focus module **110** may adjust the power output of amplifier **214** and/or may focus the various elements based on the target's distance. In some embodiments, the power output and focus may be adjusted based on a distance to the target to provide a predetermined power density (i.e., spot size) at the target. When trigger **118** is pulled, wavefront **107** is generated to deter the target. If use of non-lethal force is not successful, the user may easily switch to lethal force.

In some embodiments, lethal portion **150** comprises a machine gun; however, lethal portion **150** of weapon **100** may comprise almost any type of gun including hand-held guns. In some embodiments, the lethal portion may comprise a rifle or a machine gun, such as an M-16 rifle, although the scope of the invention is not limited in this respect.

In some embodiments, sub-reflector **104** may have a substantially flat millimeter-wave reflective surface. In some other embodiments, sub-reflector **104** has a millimeter-wave reflective surface comprising at least a portion of a substantially hyperboloidal, ellipsoidal or paraboloidal surface. Other specifically tailored reflective surfaces or lenses may also be used.

In some embodiments, main reflector **106** may comprise a geometrically-flat electrically-parabolic surface reflector antenna having a plurality of antenna elements to receive and retransmit an incident wavefront, although the scope of the invention is not limited in this respect. In these embodiments, the antenna elements may have circumferentially varying sizes and may be arranged around a center of the main reflector. In some embodiments, the antenna elements may have their electrical shapes optimized to generate either a collimating or converging wavefront of desired power densities. In some embodiments, the antenna elements may comprise a plurality of dual-polarized dipoles that circumferentially vary in size, although the scope of the invention is not limited in this respect. In some embodiments, the antenna elements may each provide approximately a 180 degree phase shift, although the scope of the invention is not limited in this respect. In some embodiments, the individual antenna elements may have varying sizes and shapes to receive the wavefront reflected by sub-reflector **104** and generate output wavefront **107** as either a collimated wavefront or a converting wavefront. An example of a reflector suitable for use as main reflector **106** may include the geometrically-flat electrically-parabolic surface reflector antenna disclosed in U.S. Pat. No. 4,905,014, although other reflective elements may also be suitable.

In some embodiments, main reflector **106** comprises an active reflect-array antenna comprising a plurality of active elements. Each element may have a receive antenna to receive portions of the reflected wavefront, an amplifier to amplify signals from the receive antenna, and a transmit antenna to transmit the amplified signals. In these embodiments, the plurality of active elements may generate an amplified wavefront in the direction toward the target. In these embodiments, the receive and transmit antennas may be orthogonally polarized.

In some embodiments, main reflector **106** may be coupled by a hinge to the weapon **100** to allow main reflector **106** to fold back when the non-lethal portion is not being used. In some embodiments, main reflector **106** is foldable and may fold into two or more flat sections. One of the sections may be coupled by a hinge to weapon **100**, and the two or more flat sections may fold up at least partially around the weapon.

In some other embodiments, main reflector **106** comprises a single flat panel and is detachable from weapon **100**. In these embodiments, main reflector **106** may be stored in a user's backpack, for example, although the scope of the invention is not limited in this respect. In some embodiments, main reflector **106** may be able to be snapped-on to weapon **100**.

FIGS. **3**A-**3**F illustrate embodiments of the present invention having a foldable main reflector in which main reflector **106** folds up and wraps around the body of weapon **100** when the non-lethal portion is not in use. FIG. **3**A is a side view illustrating main reflector **106** in a fully folded-up position in accordance with some embodiments of the present invention. FIG. **3**B is a perspective view illustrating main reflector **106** in a fully folded-up position in accordance with some embodiments of the present invention. FIG. **3**C is a perspective view illustrating main reflector **106** in a partially folded-up position in accordance with some embodiments of the present invention. FIG. **3**D is a perspective view illustrating wings of main reflector **106** folded-up in accordance with some embodiments of the present invention. FIG. **3**E is a side view illustrating wings of main reflector **106** partially folded-up in accordance with some embodiments of the present invention. FIG. **3**F is a perspective view illustrating wings of main reflector **106** partially folded-up in accordance with some embodiments of the present invention.

FIG. **4**A is a side view illustrating main reflector **106** in a fully-deployed position in accordance with some embodiments of the present invention. FIG. **4**B is a perspective view illustrating main reflector **106** in a fully-deployed position in accordance with some embodiments of the present invention. FIG. **4**C is a top view of the weapon illustrated in FIGS. **4**A and **4**B. In these embodiments, main reflector **106** may be coupled by a hinge to weapon **100** to allow main reflector **106** to fold back when the non-lethal portion is not being used. In these embodiments, main reflector **106** comprises a single flat panel and may be detachable from weapon **100**. In these embodiments, main reflector **106** may be stored in a user's backpack, for example, although the scope of the invention is not limited in this respect. In some embodiments, main reflector **106** may be able to be snapped-on to weapon **100**.

FIG. **5**A is a side view illustrating the operation of the non-lethal portion of the weapon in accordance with embodiments of the present invention. FIG. **5**B is a perspective view illustrating the operation of the non-lethal portion of the weapon in accordance with embodiments of the present invention. The non-lethal portion is illustrated generating wavefront **107** toward a target.

FIG. **6**A is a side view illustrating the removable energy-storage module **112** in accordance with embodiments of the present invention. FIG. **6**B is a perspective view illustrating the removable energy-storage module **112** in accordance with embodiments of the present invention.

Although non-lethal portion **200** (FIG. **2**) is illustrated as having several separate functional elements, one or more of the functional elements may be combined and may be implemented by combinations of software-configured elements, such as processing elements including digital signal processors (DSPs), and/or other hardware elements. For example, some elements, such as system controller **226** (FIG. **2**) and or focus controller **111** (FIG. **2**) may comprise one or more microprocessors, DSPs, application specific integrated circuits (ASICs), and combinations of various hardware and logic circuitry for performing at least the functions described herein. In some embodiments, the functional elements of non-lethal portion **200** (FIG. **2**) may refer to one or more processes operating on one or more processing elements.

The Abstract is provided to comply with 37 C.F.R. Section 1.72(b) requiring an abstract that will allow the reader to ascertain the nature and gist of the technical disclosure. It is submitted with the understanding that it will not be used to limit or interpret the scope or meaning of the claims.

In the foregoing detailed description, various features may be occasionally grouped together in a single embodiment for the purpose of streamlining the disclosure. This method of disclosure is not to be interpreted as reflecting an intention that the claimed embodiments of the subject matter require more features than are expressly recited in each claim. Rather, as the following claims reflect, invention may lie in less than all features of a single disclosed embodiment. Thus the following claims are hereby incorporated into the detailed description, with each claim standing on its own as a separate preferred embodiment.

What is claimed is:

1. A weapon having a lethal and a non-lethal portion, the non-lethal portion comprising:

an output antenna to transmit a high-power millimeter-wave initial wavefront;

a main reflector;

a sub-reflector to reflect the initial wavefront to the main reflector; and

a high-power millimeter-wave amplifier coupled to the output antenna to generate a high-power RF signal,

wherein the main reflector is to direct the reflected wavefront in a bore-sighted direction of the lethal portion toward a target, and

wherein the weapon is portable and both the lethal and non-lethal portions are coupled together to be co-aligned in the bore-sighted direction.

2. A weapon having a lethal and a non-lethal portion comprising:

an output antenna to transmit a high-power millimeter-wave initial wavefront;

a main reflector; and

a sub-reflector to reflect the initial wavefront to the main reflector,

wherein the main reflector is to direct the reflected wavefront in a bore-sighted direction toward a target,

wherein the non-lethal portion further comprises a rangefinder to determine a distance to the target, and

wherein a predetermined power density of the directed wavefront at or near the target is generated based on the distance.

3. The weapon of claim **2** further comprising a laser designator to visually designate the target, and

wherein the laser designator is bore-sighted with both the lethal portion and non-lethal portion.

4. A weapon having a lethal and a non-lethal portion comprising:

an output antenna to transmit a high-power millimeter-wave initial wavefront;

a main reflector;

a sub-reflector to reflect the initial wavefront to the main reflector; and

further comprising a focusing element to change a focus the initial wavefront generated by the output antenna based on a distance to the target,

wherein the main reflector is to direct the reflected wavefront in a bore-sighted direction toward a target.

5. The weapon of claim **4** wherein the focusing element is moveable by a user allowing the user to select a focusing position for the focusing element based on the distance to the target.

6. The weapon of claim **4** wherein the focusing element comprises a millimeter-wave radio frequency lens.

7. The weapon of claim **4** further comprising a controller to change a focus of the focusing element in response to changes in a distance to the target,

wherein the focusing element is to change a convergence point of the directed wavefront to generate a predetermined power density at or near the target.

8. A weapon having a lethal and a non-lethal portion comprising:

an output antenna to transmit a high-power millimeter-wave initial wavefront;

a main reflector; and

a sub-reflector to reflect the initial wavefront to the main reflector,

wherein the main reflector is to direct the reflected wavefront in a bore-sighted direction toward a target, and

wherein the non-lethal portion comprises a high-power millimeter-wave amplifier coupled to the output antenna to generate a high-power RF signal.

9. The weapon of claim **8** wherein a wavefront directed by the main reflector has a power density selected to deliver a non-lethal deterring effect on the target.

10. The weapon of claim **8** wherein the wavefront comprises W-band millimeter-wave frequency radiation selected to penetrate a shallow skin-depth of human skin allowing the radiation to heat a region of the skin that includes pain sensors, and

wherein a power density at the target is selected to cause a deterring effect by inducing pain on the human skin.

11. The weapon of claim **8** wherein the main reflector is a collimating reflector to generate a collimated wavefront toward the target.

12. The weapon of claim **8** wherein the main reflector is to generate a converting wavefront to converge at or near the target, wherein a convergence distance is determined to provide a predetermined power density at or near a surface of the target.

13. The weapon of claim **8** wherein the main reflector is aligned with sights of the lethal portion of the weapon, and

wherein the non-lethal portion is a bore-sighted attachment kit to add non-lethal capability to a lethal weapon.

14. The weapon of claim **8** wherein the non-lethal portion further comprises a replaceable energy-storage module to provide electrical energy for the millimeter-wave amplifier.

15. The weapon of claim **8** further comprising a cooling element to cool the amplifier, wherein the cooling element circulates a phase-change fluid to cool the amplifier.

16. The weapon of claim **8** further comprising a cooling element to cool the amplifier, wherein the cooling element uses an expanding gas to cool the amplifier, the cooling element including a pressurized chamber to store the gas,

wherein the chamber including the gas therein is replaceable and is replaces as part of a replaceable energy-storage module.

17. The weapon of claim **8** further comprising:

a lethal-weapon trigger to fire the lethal portion; and

a non-lethal-weapon trigger to activate the non-lethal portion to generate the initial wavefront,

wherein the lethal portion comprises a machine gun.

18. The weapon of claim **8** wherein the main reflector comprises a single flat panel and is detachable from the weapon.

19. The weapon of claim **8** wherein the sub-reflector has a substantially flat millimeter-wave reflective surface.

20. The weapon of claim **8** wherein the sub-reflector has a millimeter-wave reflective surface comprising at least a portion of one of a substantially hyperboloid, ellipsoid or paraboloid surface.

21. A weapon having a lethal and a non-lethal portion comprising:

an output antenna to transmit a high-power millimeter-wave initial wavefront;

a main reflector; and

a sub-reflector to reflect the initial wavefront to the main reflector,

wherein the main reflector is to direct the reflected wavefront in a bore-sighted direction toward a target,

wherein the main reflector is a geometrically-flat electrically-parabolic surface reflector antenna having a plurality of antenna elements to receive and retransmit an incident wavefront, the antenna elements having circumferentially varying sizes arranged circumferentially around a center of the main reflector to generate either a collimating or converging wavefront.

22. A weapon having a lethal and a non-lethal portion comprising:

an output antenna to transmit a high-power millimeter-wave initial wavefront;

a main reflector; and

a sub-reflector to reflect the initial wavefront to the main reflector,

wherein the main reflector is to direct the reflected wavefront in a bore-sighted direction toward a target,

wherein the main reflector comprises an active reflect-array antenna comprising a plurality of active elements, each element having a receive antenna to receive portions of the reflected wavefront, an amplifier to amplify signals from the receive antenna, and a transmit antenna to transmit the amplified signals, and

wherein the plurality of active elements generate an amplified wavefront for direction toward the target.

23. A weapon having a lethal and a non-lethal portion comprising:

an output antenna to transmit a high-power millimeter-wave initial wavefront;

a main reflector; and

a sub-reflector to reflect the initial wavefront to the main reflector,

wherein the main reflector is to direct the reflected wavefront in a bore-sighted direction toward a target,

wherein the main reflector is foldable into two or more flat sections,

wherein one of the sections is hinged, and

wherein the two or more flat sections fold up at least partially around the weapon.

* * * * *

The Tom Bearden Website

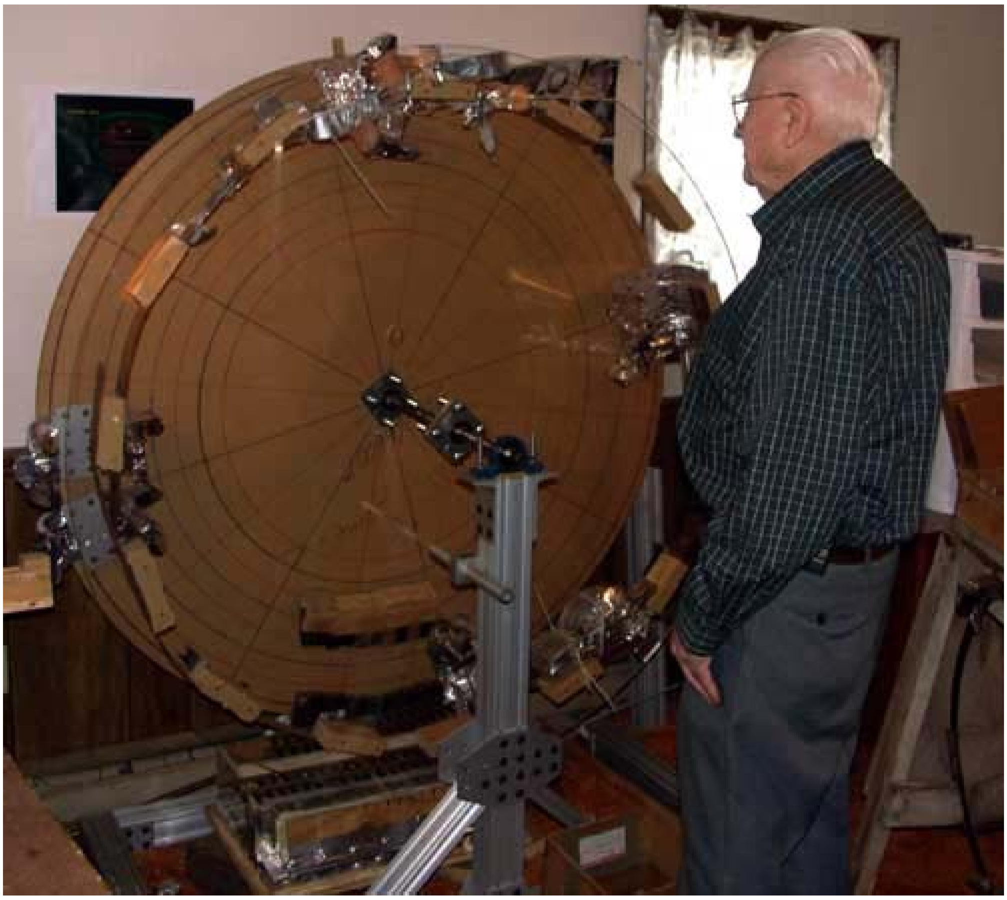

Howard Johnson constructing one of his magnetic motors

A candid shot taken during the filming of the upcoming William Gazecki Documentary
ENERGY FROM THE VACUUM

Johnson's Motors

Johnson [[xxvii]] has built many novel linear and rotary motors and at least one self-powering magnetic rotary device — later stolen in a mysterious break-in at his laboratory — personally tested by the present author. Johnson uses a bidirectional "two particle" theory of magnetic flux lines which can be justified by Whittaker's earlier work showing the internal bidirectional energy flows in all potentials and fields. He also utilizes controlled spin-waves and self-initiated precise exchange forces, which are known to momentarily produce bursts of very strong forcefields [[xxviii]]. His approach is to use highly nonlinear assemblies of magnets which initiate the foregoing phenomena at very precise points in the rotation cycle. In short, he seeks to produce precisely located and directed sudden magnetic forces, using self-initiated nonlinear magnetic phenomena. This is analogous to what the Wankel engine did using the Lenz law effect by sharply interrupting a weak current in a external coil. We point out that the Lenz law effect and other very abrupt field changes momentarily produce not only an amplified *Poynting* energy flow component, but also an amplified *Heaviside* energy flow component as well.

[xxvii]. Howard R. Johnson, "Permanent Magnet Motor." U.S. Patent No. 4,151,431, Apr. 24, 1979. See also Johnson's U.S. Patents 4,877,983, Oct. 31, 1989 and 5,402,021, Mar. 28, 1995.

[xxviii]. For an exposition of exchange forces and exchange energy, see B. D. Cullity, Introduction to Magnetic Materials, Addison-Wesley, Reading, MA, 1972; A.G. Gurevich and G.A. Melkov, Magnetization Oscillations and Waves, CRC Press, 1996; Victor S. L'vov, Wave Turbulence Under Parametric Excitation: Applications to Magnets, Springer-Verlag, Berlin, 1994. See also V.S. L'vov and L.A. Prozorova, "Spin Waves Above the Threshold of Parametric Excitation," in A.S. Borovik-Romanov and S.K. Sinha, Eds., Spin Waves and Magnetic Excitations, North-Holland, Amsterdam, 1988.

Excerpted from "On Extracting Electromagnetic Energy from the Vacuum," IC-2000, by Tom Bearden.

The photos below are excerpted from Howard Johnson's book "Discovering Magnetism", shortly to be
re-published by Cheniere Press

The computer plots were taken of one of Howard Johnson's magnetic gates, and clearly show the spin forces
created by the magnets, and also the "firing" of a magnet, with a burst of energy that is often several hundred
times larger than the baseline magnetic strength. Howard Johnson was working in the field that is now known
as "spintronics" long before it was "officially" recognized. There are over 200 magnetic effects, and only
around half of them are understood.

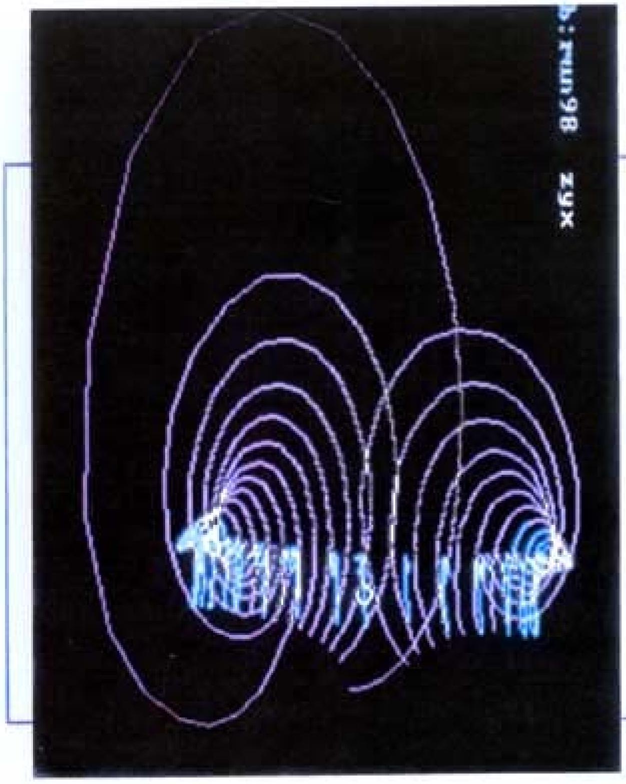

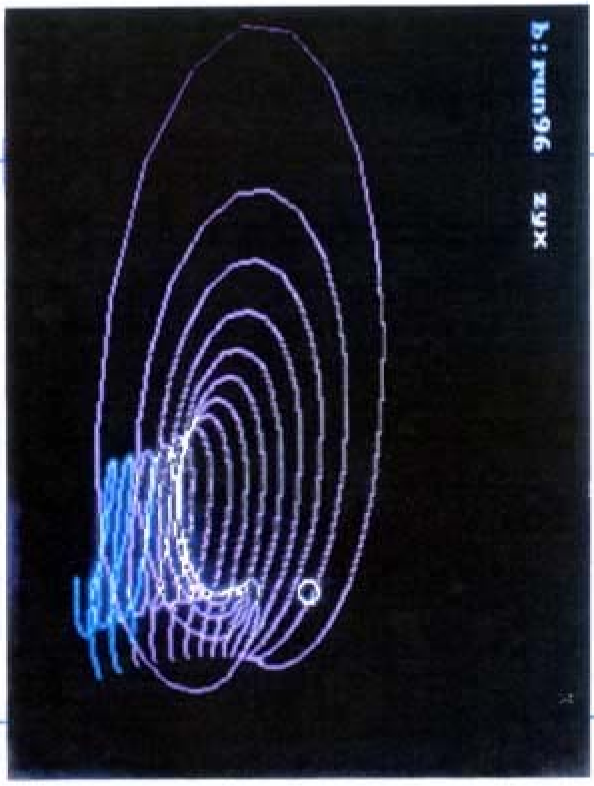

Th
be
pic
in
the
pr
cu
ma

This picture is the interaction
of the curved metallic magnet
as it enters into the gate:

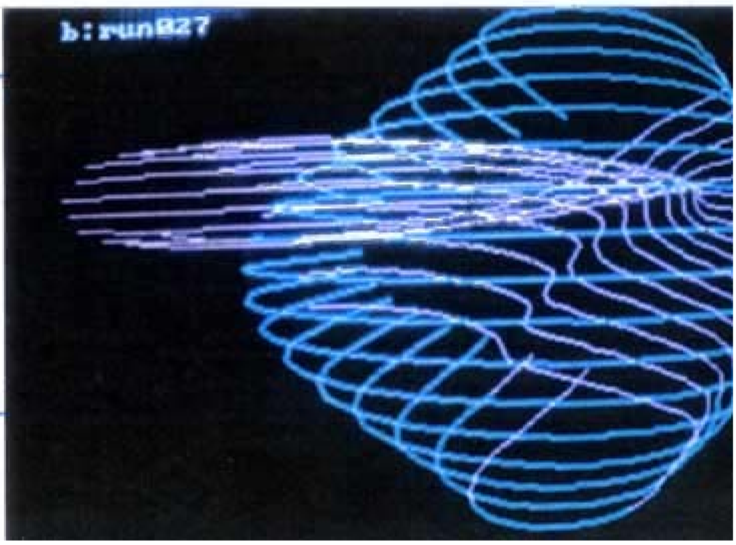

CTEC Inc. 1996 report on testing one of Howard Johnson's magnetic gates

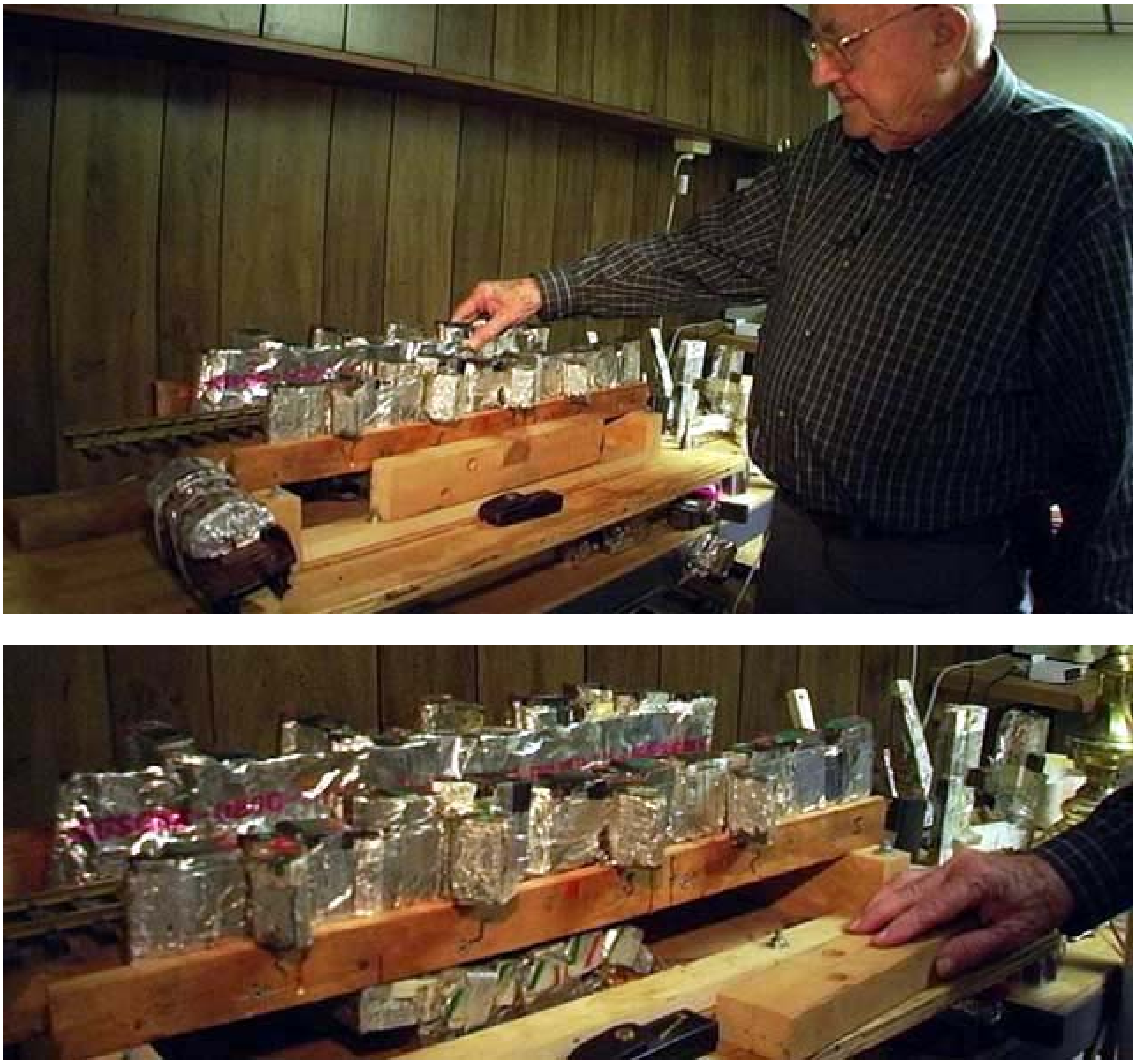

Howard demonstrates one of his magnetic rail cars

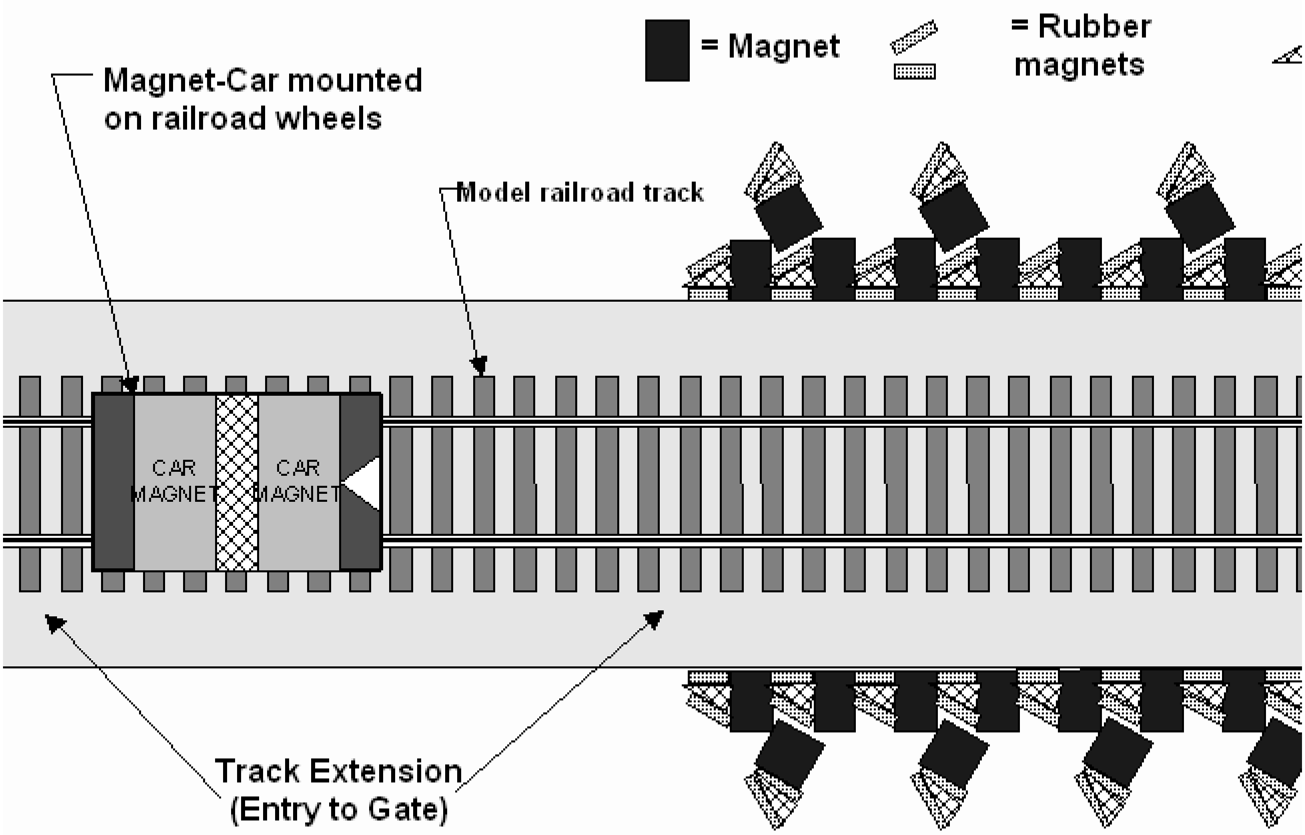
= Magnet
= Rubber magnets
Magnet-Car mounted on railroad wheels
Model railroad track
CAR MAGNET
CAR MAGNET
Track Extension (Entry to Gate)

Figure 2. Howard Johnson's Magnetic Gate with Tra Extension and Magnet-Car

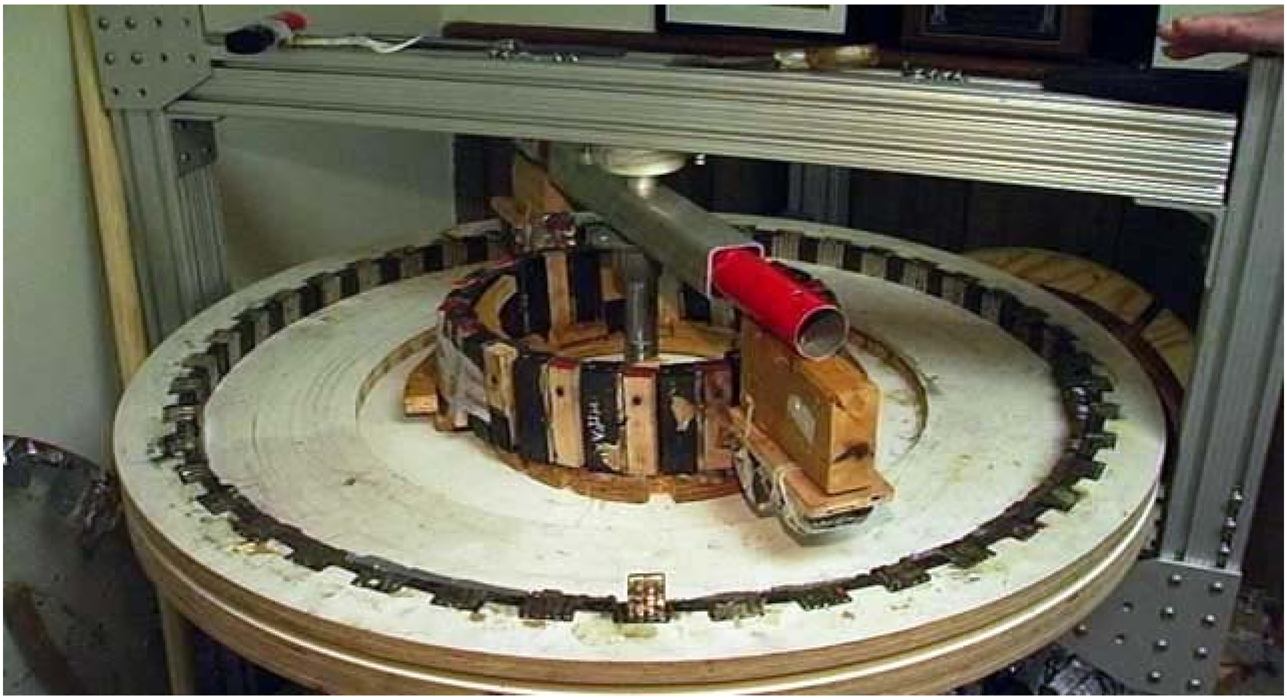

Another Johnson self-powering all-magnetic motor

Howard Johnson - unsung pioneer

US006392370B1

(12) **United States Patent**
Bedini

(10) **Patent No.: US 6,392,370 B1**
(45) **Date of Patent: May 21, 2002**

(54) **DEVICE AND METHOD OF A BACK EMF PERMANENT ELECTROMAGNETIC MOTOR GENERATOR**

(75) Inventor: **John C. Bedini**, Coeur d'Alene, ID (US)

(73) Assignee: **Bedini Technology, Inc.**, Coeur d'Alene, ID (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/483,715**

(22) Filed: **Jan. 13, 2000**

(51) **Int. Cl.**^{7} **H02K 47/14**

(52) **U.S. Cl.** **318/140**; 318/492; 318/538; 310/113

(58) **Field of Search** 318/140, 148, 318/151, 153, 459, 491, 492, 538; 310/152, 156, 112, 113

(56) **References Cited**

U.S. PATENT DOCUMENTS

2,560,260	A	7/1951	Sturtevant et al.
2,613,246	A	10/1952	Spodig
2,974,981	A	3/1961	Vervest et al.
4,222,021	A	9/1980	Bunker, Jr.
4,265,746	A	5/1981	Zimmerman, Sr. et al.
4,265,754	A	5/1981	Menold
4,417,167	A	11/1983	Ishii et al.
4,423,343	A	12/1983	Field, II
4,429,263	A	1/1984	Miiller
4,433,260	A	2/1984	Weisbord et al.
4,488,075	A	12/1984	DeCesare
RE31,950	E	7/1985	Binns
4,536,230	A	8/1985	Landa et al.
4,761,590	A	8/1988	Kaszman
4,916,346	A	4/1990	Kliman
4,972,112	A	11/1990	Kim
4,980,595	A	12/1990	Arora
5,130,595	A	7/1992	Arora
5,177,054	A	1/1993	Lloyd et al.
5,227,702	A	7/1993	Nahirney
5,334,894	A	8/1994	Nakagawa
5,350,958	A	9/1994	Ohnishi
5,356,534	A	10/1994	Zimmerman et al.

(List continued on next page.)

FOREIGN PATENT DOCUMENTS

EP	0 462 340 A1	12/1991
EP	0 932 248 A1	7/1999
GB	2 224 398 A	5/1990
WO	WO 99/38247	7/1999

Primary Examiner—Bentsu Ro
(74) *Attorney, Agent, or Firm*—Thomas G. Walsh; Graybeal Jackson Haley LLP

(57) **ABSTRACT**

This invention is a back EMF permanent electromagnetic motor generator and method using a regauging process for capturing available electromagnetic energy in the system. The device is comprised of a rotor with magnets of the same polarity; a timing wheel in apposition to a magnetic Hall Effect pickup switch semiconductor; and a stator comprised of two bars connected by a permanent magnet with magnetized pole pieces at one end of each bar. There are input and output coils created by wrapping each bar with a conducting material such as copper wire. Energy from the output coils is transferred to a recovery rectifier or diode. The magnets of the rotor, which is located on a shaft along with the timing wheel, are in apposition to the magnetized pole pieces of the two bars. The invention works through a process of regauging, that is, the flux fields created by the coils is collapsed because of a reversal of the magnetic field in the magnetized pole pieces thus allowing the capture of available back EMP energy. Additional available energy may be captured and used to re-energize the battery, and/or sent in another direction to be used as work. As an alternative, the available back EMF energy may be dissipated into the system.

10 Claims, 3 Drawing Sheets

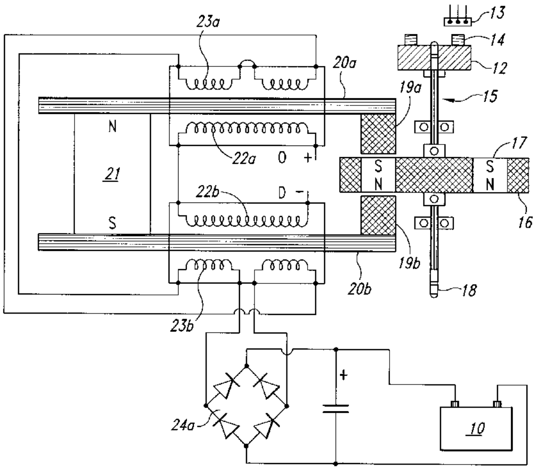

U.S. PATENT DOCUMENTS

5,369,325	A		11/1994	Nagate et al.	
5,371,426	A		12/1994	Nagate et al.	
5,451,825	A		9/1995	Strohm	
5,508,575	A		4/1996	Elrod, Jr.	
5,514,923	A		5/1996	Gossler et al.	
5,528,113	A	*	6/1996	Boys et al.	318/16
5,532,532	A		7/1996	DeVault et al.	
5,731,645	A	*	3/1998	Clifton et al.	310/74
6,118,202	A	*	9/2000	Pinkerton	310/178

* cited by examiner

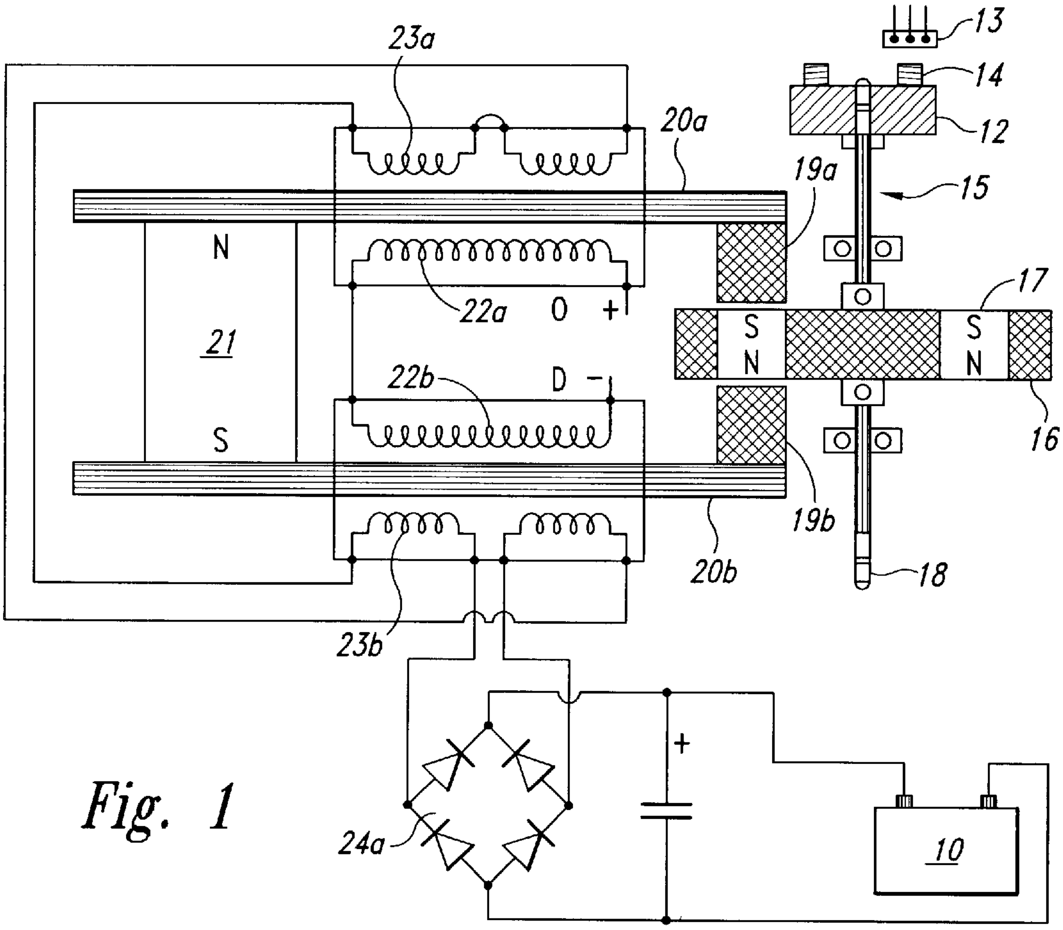
13
14
12
15
23a
20a
19a
17
N
21
S
22a
O
+
22b
D
−
S
N
S
N
16
19b
20b
18
23b
+
24a
10

Fig. 1

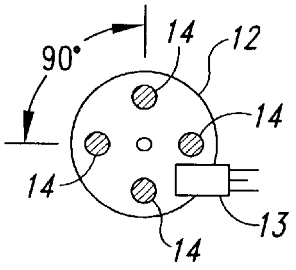
90°
14
12
14
14
13
14

Fig. 1A

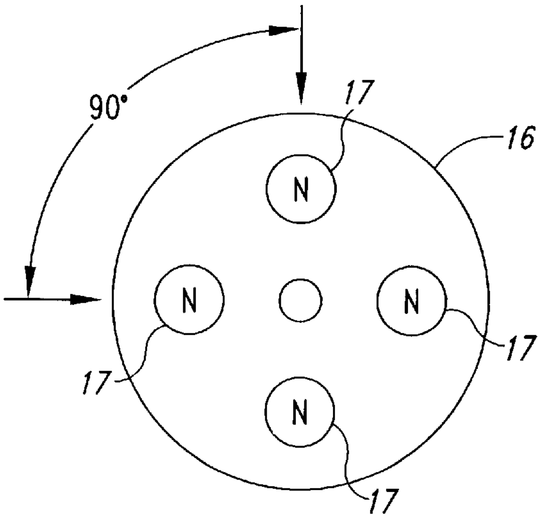
90°
17
16
N
N
N
N
17
17
17

Fig. 1B

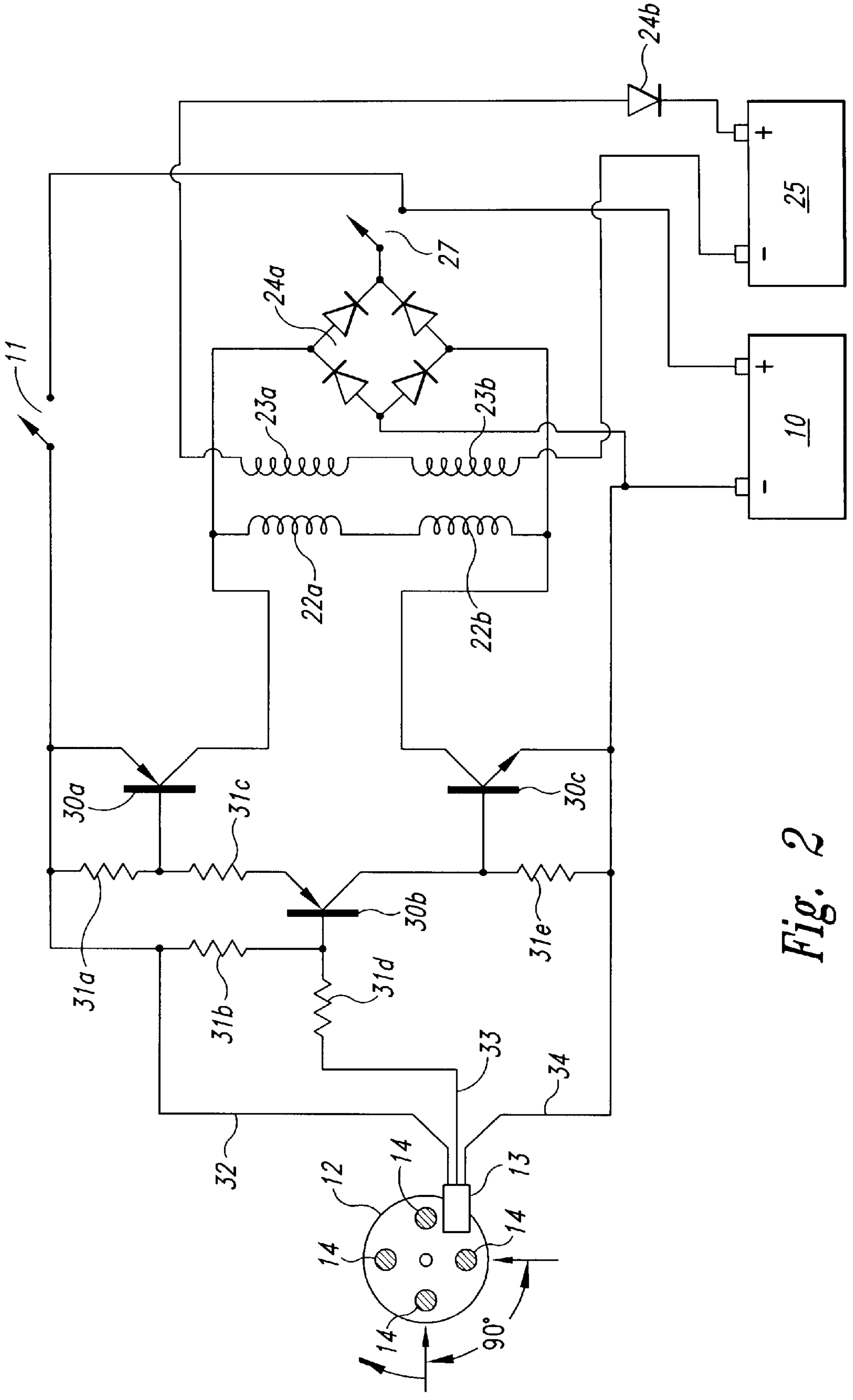
+
-
+
-
90°

Fig. 2

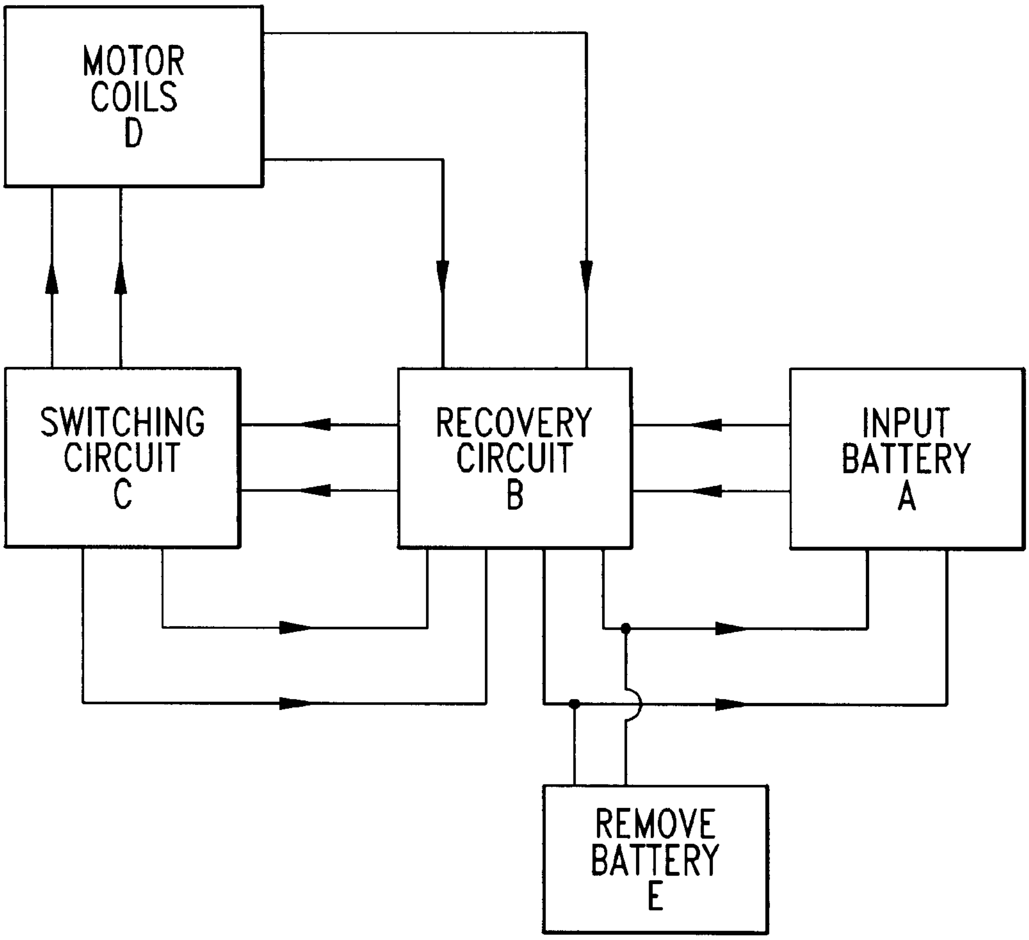
MOTOR
COILS
D
SWITCHING
CIRCUIT
C
RECOVERY
CIRCUIT
B
INPUT
BATTERY
A
REMOVE
BATTERY
E

Fig. 3

DEVICE AND METHOD OF A BACK EMF PERMANENT ELECTROMAGNETIC MOTOR GENERATOR

BACKGROUND OF THE INVENTION

1. Field of the Invention

The invention relates generally to the capturing of electromagnetic energy using a method and device to create back EMF (electromagnetic force) and re-phasing of the back EMF to recycle and capture the available back EMF energy. Back EMF is also referred to as regauging and may be defined as energy created by the magnetic field from coils, and only by coils, and not by magnets.

2. Background Information and Related Art

Operation of a normal magnetic motor has the rotor pole attracting the stator pole, resulting in the generation of power from the magnets to the rotor and flywheel. During this phase, energy flows from the magnetics to the rotor/flywheel and is stored in the increased rotation. A rotor pole leaving a stator pole and creating a condition of drag-back results in power having to be put back into the magnetic section by the rotor and flywheel to forcibly overcome the drag-back. In a perfect, friction-free motor, the net force field is therefore referred to as most conservative. In other words, a most conservative EMF motor has maximum efficiency. Without extra energy continually fed to the motor, no net work can be done by the magnetic field, since half the time the magnetic field adds energy to the load (the rotor and flywheel) and the other half of the time it subtracts energy back from the load (the rotor and flywheel). Therefore the total net energy output is zero in any such rotary process without additional energy input. To use a present day magnetic motor, continuous energy must be input into the motor to overcome drag-back and to power the motor and its load.

Present EMF motors and generators all use such conservative fields and therefore, have internal losses. Hence, it is necessary to continually input all of the energy that the motor outputs to the load, plus more energy to cover losses inside the motor itself. EMF motors are rated for efficiency and performance by how much energy input into the motor actually results in output energy to the load. Normally, the Coefficient of Performance (COP) rating is used as a measure of efficiency. The COP is the actual output energy going into the load and powering it, divided by the energy that must be input into the device with its load. COP is the power out into the load, divided by the power input into the motor/load combination. If there were zero internal losses in a motor, that "perfect" motor would have a coefficient of performance (COP) equal to 1.0. That is, all energy input into the motor would be output by the motor directly into the load, and none of the input energy would be lost or dissipated in the motor itself.

In magnetic motor generators presently in use, however, due to friction and design flaws, there are always internal losses and inefficiencies. Some of the energy input into the motor is dissipated in these internal losses. As a consequence, the energy that gets to the load is always less than the input energy. So a standard motor operates with a COP of less than 1.0 which is expressed as COP<1.0. An inefficient motor may have a COP=0.4 or 0.45, while a specially designed, highly efficient motor may have a COP=0.85.

The conservative field inside of a motor itself can be divided into two phases. Producing a conservative field involves net symmetry between the "power out" phase from the magnetics to the rotor/flywheel and the "power back in" phase from the rotor/flywheel back to the magnetics. That is, the two flows of energy (one from the magnetics into the rotor and flywheel, and one from the rotor and flywheel back to the magnetics) are identical in magnitude but opposite in direction. Each phase alone is said to be "asymmetrical"; that is, it either has: 1) a net energy flow out to the rotor/flywheel; or 2) a net energy flow back into the magnetics from the rotor/flywheel. In simplified terms, it is referred to as "power out" and "power back in" phases with respect to the motor magnetics. Hence, the two asymmetrical phases are: 1) the power-out phase; and 2) the "power back in" phase, with reference to the magnetics.

For the power-out phase, energy is derived from the EMF existing between the stator pole and incoming rotor pole in an attraction mode. In this phase, the rotary motion (angular momentum and kinetic energy) of the rotor and flywheel is increased. In short, power is added to the rotor/flywheel (and thus to the load) from the fields between stator pole and rotor pole (the electromagnetic aspects of the system).

For the "power back in" phase, energy must be fed back into the magnetics from the rotor and flywheel (and the load) to overcome the drag-back forces existing between stator pole and outgoing rotor pole. In this phase, energy is returned back to the internal magnetic system from the rotary motion of the rotor and flywheel (the angular momentum, which is the rotational energy×time). As is well known in physics, a rotor/flywheel's angular momentum provides a convenient way to store energy with the spinning rotor/flywheel mass acting as an energy reservoir.

All present day conventional magnetic motors use various methods for overcoming and partially reversing back EMF. Back EMF is the return pulse from the coil out of phase and is also referred to as regauging. The back EMF is shorted out and the rotor is attracted back in, therefore eliminating back drag. This can be accomplished by pouring in more energy, which overpowers the back EMF, thereby producing a forward EMP in that region. The energy required for this method must be furnished by the operator.

The motor of the present invention uses only a small amount of energy to "trigger" a much larger input of available energy by supplying back EMF, thus increasing the potential energy of the system. It then utilizes this excess potential energy to reduce or reverse back EMF, thereby increasing the efficiency of the motor and, therefore, the COP.

If the energy in phase 1 (the power-out phase) is increased by additional available energy in the electromagnetics themselves, then the energy in phase 1 can be made greater than the energy in phase 2 (the power-back-in phase) without the operator furnishing the energy utilized. This produces a non-conservative net field. Net power can then be taken from the rotating stator and flywheel, because the available energy added into the stator and flywheel by the additional effects, is transformed by the rotor/flywheel into excess angular momentum and stored as such. Angular momentum is conserved at all times; but now some of the angular momentum added to the flywheel is evoked by additional effects in the electromagnetics rather than being furnished by the operator.

Electrodynamicists assume that the potential available energy of any system can be changed at will and without cost. This is back EMF and is well-known in physics. It is also routinely employed by electrodynamicists in the theoretical aspects. But to simplify the mathematics, electrodynamicists will create a back EMF twice simultaneously, each back EMF carefully selected just so that the two available forces that are produced are equal and opposite and cancel each other "symmetrically". This is referred to as "symmetrical back EMF". A symmetrical back EMF system cannot produce a COP>1.0.

On the other hand, the motor of the present invention deliberately creates a back EMF itself and its potential energy only once at a time, thereby retaining each extra force for a period of time and applying it to increase the angular momentum and kinetic energy of the rotor and flywheel. Specifically, this back EMF energy with its net force is deliberately applied in the motor of the present invention to overcome and even reverse the conventional drag-back (the back EMF). Hence less energy must be taken from the rotor and flywheel to overcome the reduced back EMF, and in the ideal case none is required since the back EMF has been overpowered and converted to forward EMF by the back EMF energy and force. In the motor of the present invention, the conventional back-drag section of the magnetics becomes a forward-EMF section and now adds energy to the rotor/flywheel instead of subtracting it. The important feature is that the operator only pays for the small amount of energy necessary to trigger the back EMF, and does not have to furnish the much larger back EMF energy itself.

When the desired energy in phase 1 (the power out phase) is thus made greater than the undesired "drag-back" energy in phase 2, then part of the output power normally dragged back from the rotor and flywheel by the fields in phase 2 is not required. Hence, additional power compared to the system (without the special back EMF mechanisms) is available from the rotor/flywheel. The rotor maintains additional angular momentum and kinetic energy, compared to a system which does not produce back EMF itself. Consequently, the excess angular momentum retained by the rotor and flywheel can be utilized as additional shaft power to power an external load connected to the shaft.

A standard magnetic motor operates as the result of the motor being furnished with external energy input into the system by the operator to reduce phase 2 (power back into the magnetics from the rotor/flywheel) by any of several methods and mechanisms. The primary purpose of this external energy input into the system is to overcome the back EMF and also provide for the inevitable energy losses in the system. There is no input of energy separate from the operator input. Therefore, the COP of any standard magnetic motor is COP less than 1.0. The efficiency of a standard magnetic motor varies from less than 50% to a maximum of about 85%, and so has a COP<1.0. When nothing is done in the motor that will produce a reduction of the back EMF without the operator inputing all the energy for it, then for even a frictionless, ideal permanent magnet motor, the COP can never exceed 1.0.

Until the introduction of the motor of the present invention, it has been standard universal practice that the operator must furnish all energy used to reduce the back EMF, provide for the internal losses, and power the load. It is therefore a common belief by the scientific community that an ideal (loss-less) permanent magnet motor cannot exceed COP=1.0. And that is true, so long as the operator himself must furnish all the energy. Further, since real permanent magnetic motors have real internal losses, some of the input energy is always lost in the motor itself, and that lost energy is not available for powering the rotor/flywheel and load. Hence a real permanent magnetic motor of the conventional kind will always have a COP<1.0.

The common assumption that the COP of a motor is limited to less than 1.0 is not necessarily true, and that COP>1.0 is permitted without violating the laws of nature, laws of physics, or laws of thermodynamics. However, it can immediately be seen that any permanent magnet motor exhibiting a COP>1.0 must have some available energy input returning in the form of back EMF.

A problem relates to how back EMF energy can be obtained from a circuit's external environment for the specific task of reducing the back-drag EMF without the operator having to supply any input of that excess energy. In short, the ultimate challenge is to find a way to cause the system to: 1) become an open dissipative system, that is, a system receiving available excess energy from its environment, in other words, from an external source; and 2) use that available excess energy to reduce the drag-back EMF between stator and rotor poles as the rotor pole is leaving the stator pole. If this objective can be accomplished, the system will be removed from thermodynamic equilibrium. Instead, it will be converted to a system out-of-thermodynamic equilibrium. Such a system is not required to obey classical equilibrium thermodynamics.

Instead, an out-of-equilibrium thermodynamic system must obey the thermodynamics of open systems far from the established and well-known parameters of thermodynamic equilibrium. As is well known in the physics of thermodynamics, such open systems can permissibly: 1) self-order; 2) self-oscillate; 3) output more back EMF energy than energy input by the operator (the available excess back EMF energy is received from an external source and some energy is input by the operator as well); 4) power itself as well as its loads and losses simultaneously (in that case, all the energy is received from the available external source and there is no input energy from the operator); and 5) exhibit negentropy, that is, produce an increase of energy that is available in the system, and that is independent of the energy put into the system by the operator. As a definition, entropy roughly corresponds to the energy of a system that has become unavailable for use. Negentropy corresponds to additional energy of a system that has become available for use.

In the back EMF permanent magnet electromagnetic motor generator of the present invention, several known processes and methods are utilized which allow the invention to operate periodically as an open dissipative system (receiving available excess energy from back EMF) far from thermodynamic equilibrium, whereby it produces and receives its excess energy from a known external source.

A method is utilized to temporarily produce a much larger source of available external energy around an energized coil. Then the unique design features of this new motor provides a method and mechanism that can immediately produce a second increase in that energy, concurrently as the energy flow is reversed. Therefore, the motor is capable of producing two asymmetrical back EMFs, one after the other, of the energy within a single coil, which dramatically increases the energy available and causes that available excess energy to then enter the circuit impulsively, being collected and utilized.

The present motor utilizes this available excess back EMF energy to overcome and even reverse the back-drag EMF between stator pole and rotor pole, while furnishing only a small trigger pulse of energy necessary to control and activate the direction of the back EMF energy flow.

By using a number of such dual asymmetrical self back EMFs for every revolution of the rotor, the rotor and flywheel collectively focus all the excess impulsive inputs into increased angular momentum (expressed as energy× time), shaft torque, and shaft power.

Further, some of the excess energy deliberately generated in the coil by the utilization of the dual process manifests in the form of excess electrical energy in the circuit and is utilized to power electrical loads, e.g., a lamp, fan, motor, or other electrical devices. The remainder of the excess energy generated in the coil can be used to power the rotor and flywheel, with the rotor/flywheel also furnishing shaft horsepower for powering mechanical loads.

This new and unique motor utilizes a means to furnish the relatively small amount of energy to initiate the impulsive asymmetrical self back EMF actions. Then part of the available excess electrical power drawn off from the back EMFs is utilized to recharge the battery with dramatically increased over voltage pulses.

The unique design features of this motor utilize both north and south magnetic poles of each rotor and stator magnet. Therefore, the number of impulsive self back EMFs in a single rotation of the rotor is doubled. Advanced designs increase the number of self back EMFs in a single rotor rotation with the result that there is an increase in the number of impulses per rotation which increase the power output of this new motor.

The sharp voltage pulse produced in the coil of this new motor by the rapidly collapsing field in the back EMF coil is connected to a battery in charge mode and to an external electrical load. The net result is that the coil asymmetrically creates back EMF itself in a manner adding available energy and impulse to the circuit. The excess available energy collected in the coil is used to reverse the back-EMF phase of the stator-rotor fields to a forward EMF condition, impulsively adding acceleration and angular momentum to the rotor and flywheel. At the same time, a part of the excess energy collected in the coil is used to power electrical loads such as charging a battery and operating a lamp or such other device.

It is well known in the art that changing the voltage alone creates a back EMF and requires no work. This is because to change the potential energy does not require changing the form of that potential energy, but only its magnitude. Work is rigorously the changing of the form of energy. Therefore, as long as the form of the potential energy is not changed, the magnitude can be changed without having to perform work in the process. The motor of the present invention takes advantage of this permissible operation to create back EMF asymmetrically, and thereby change its own usable available potential energy.

In an electric power system, the potential (voltage) is changed by inputing energy to do work on the internal charges of the generator or battery. This potential energy is expended within the generator (or battery) to force the internal charges apart, forming a source dipole. Then the external closed circuit system connected to that source dipole ineptly pumps the spent electrons in the ground line back through the back EMP of the source dipole, thereby scattering the charges and killing the dipole. This shuts off the energy flow from the source dipole to the external circuit. As a consequence of that conventional method, it is a requirement to input and replace additional energy to again restore the dipole. The circuits currently utilized in most electrical generators have been designed to keep on destroying the energy flow by continually scattering all of the dipole charges and terminating the dipole. Therefore, it is necessary to keep on inputing energy to the generator to keep restoring its source dipole.

An investigation of particle physics is required to see what furnishes the energy to the external circuit. Since neither a battery nor a generator furnishes energy to the external circuit, but only furnishes energy to form the source dipole, a better understanding of the electric power principle is required to fully understand how this new motor functions. A typical battery uses its stored chemical energy to form the source dipole. A generator utilizes its input shaft energy to rotate, forming an internal magnetic field in which the positive charges are forced to move in one direction and the negative charges in the reverse direction, thereby forming the source dipole. In other words, the energy input into the generator does nothing except form the source dipole. None of the input energy goes to the external circuit. If increased current is drawn into the external load, there also is increased spent electron flow being rammed back through the source dipole, destroying it faster. Therefore, dipole-restoring-energy has to be inputed faster. The chemical energy of the battery also is expended only to separate its internal charges and form its source dipole. Again, if increased current and power is drawn into the external load, there is increased spent electron flow being rammed back through the source dipole, destroying it faster. This results in a depletion of the battery's stored energy faster, by forcing it to have to keep restoring the dipole faster.

Once the generator or battery source dipole is formed (the dipole is attached also to the external circuit), it is well known in particle physics that the dipole (as is any charge) is a broken symmetry in the vacuum energy flux. By definition, this means that the source dipole extracts and orders part of that energy received from its vacuum interaction, and pours that energy out as the energy flowing through all space surrounding the external conductors in the attached circuit. Most of this enormous energy flow surging through space surrounding the external circuit does not strike the circuit at all, and does not get intercepted or utilized. Neither is it diverged into the circuit to power the electrons, but passes on out into space and is just "wasted". Only a small "sheath" of the energy flow along the surface of the conductors strikes the surface charges in those conductors and is thereby diverged into the circuit to power the electrons. Standard texts show the huge available but wasted energy flow component, but only calculate the small portion of the energy flow that strikes the circuit, is caught by it, and is utilized to power it.

In a typical circuit, the huge available but "wasted" component of the energy flow is about 10^{13} times as large as is the small component intercepted by the surface charges and diverged into the circuit to power it. Hence, around every circuit and circuit element such as a coil, there exists a huge non-intercepted, non-diverged energy flow that is far greater than the small energy flow being diverted and used by the circuit or element.

Thus there exists an enormous untapped energy flow immediately surrounding every EMF power circuit, from which available excess energy can be intercepted and collected by the circuit, if respective non-linear actions are initiated that sharply affect and increase the reaction cross section of the circuit (i.e., its ability to intercept this available but usually wasted energy flow).

The method in which the motor of the present invention alters the reaction cross section of the coils in the circuit, is by a novel use, which momentarily changes the reaction cross section of the coil in which it is invoked. Thus, by this new motor using only a small amount of current in the form of a triggering pulse, it is able to evoke and control the immediate change of the coil's reaction cross section to this normally wasted energy flow component. As a result, the motor captures and directs some of this usually wasted environmental energy, collecting the available excess energy in the coil and then releasing it for use in the motor. By timing and switching, the innovative gate design in this new motor directs the available excess energy so that it overcomes and reverses the return EMF of the rotor-stator pole combination during what would normally be the back EMF and demonstrates the creation of the second back EMF of the system. Now instead of an "equal retardation" force being produced in the back EMF region, a forward EMF is produced that is additive to the rotor/flywheel energy and not subtractive. In short, it further accelerates the rotor/flywheel.

This results in a non-conservative magnetic field along the rotor's path. The line integral of the field around that path (i.e., the net work on the rotor/flywheel to increase its energy and angular momentum) is not zero but a significant amount. Hence, the creation of an asymmetrical back EMF impulse magnetic motor: 1) takes its available excess energy from a known external source, the huge usually non-intercepted portion of the energy flow around the coil; 2) further increases the source dipolarity by this back EMF energy; and 3) produces available excess energy flow directly from the source dipole's increased broken symmetry in its fierce energy exchange with the local vacuum.

No laws of physics or thermodynamics are violated in the method and device of the present invention, and conservation of energy rigorously applies at all times. Nonetheless, by operating as an open dissipative system not in thermodynamic equilibrium with the active vacuum, the system can permissibly receive available excess energy from a known environmental source and output more energy to a load than must be input by the operator alone. As an open system not in thermodynamic equilibrium, this new and unique motor can tap in on back EMF to energize itself, loads and losses simultaneously, fully complying with known laws of physics and thermodynamics.

A search of prior art failed to reveal any devices that recycle available energy from back EMF of a permanent electromagnetic motor generator as described in the present invention. However, the following prior art patents were reviewed:

1. U.S. Pat. No. 5,532,532 to DeVault, et al., Hermetically Sealed Super-conducting Magnet Motor.
2. U.S. Pat. No. 5,508,575 to Elrod, Jr., Direct Drive Servovalve Having Magnetically Loaded Bearing.
3. U.S. Pat. No. 5,451,825 to Strohm, Voltage Homopolar Machine.
4. U.S. Pat. No. 5,371,426 to Nagate et al., Rotor For Brushless Motor.
5. U.S. Pat. No. 5,369,325 to Nagate et al., Rotor For Brushless Electromotor And Method For Making Same.
6. U.S. Pat. No. 5,356,534 to Zimmermann, deceased et al., Magnetic-Field Amplifier.
7. U.S. Pat. No. 5,350,958 to Ohnishi, Super-conducting Rotating Machine, A Super-conducting Coil, And A Super-conducting Generator For Use In A Lighting Equipment Using Solar Energy.
8. U.S. Pat. No. 5,334,894 to Nakagawa, Rotary Pulse Motor.
9. U.S. Pat. No. 5,177,054 to Lloyd, et al., Flux Trapped Superconductor Motor and Method.
10. U.S. Pat. No. 5,130,595 to Arora, Multiple Magnetic Paths Pulse Machine.
11. U.S. Pat. No. 4,980,595 to Arora, Multiple Magnetics Paths Machine.

12. U.S. Pat. No. 4,972,112 to Kim, Brushless D.C. Motor.
13. U.S. Pat. No. 4,916,346 to Kliman, Composite Rotor Lamination For Use In Reluctance Homopolar, And Permanent Magnet Machines.
14. U.S. Pat. No. 4,761,590 to Kaszman, Electric Motor.
15. U.S. Pat. No. 4,536,230 to Landa, et al., Anisotropic Permanent Magnets.
16. U.S. Pat. No. Re. 31,950 to Binns, Alternating Current Generators And Motors.
17. U.S. Pat. No. 4,488,075 to DeCesare, Alternator With Rotor Axial Flux Excitation.
18. U.S. Pat. No. 4,433,260 to Weisbord et al., Hysteresis Synchronous Motor Utilizing Polarized Rotor.
19. U.S. Pat. No. 4,429,263 to Muller, Low Magnetic Leakage Flux Brushless Pulse Controlled D-C Motor.
20. U.S. Pat. No. 4,423,343 to Field, II, Synchronous Motor System.
21. U.S. Pat. No. 4,417,167 to Ishii et al., DC Brushless Motor.
22. U.S. Pat. No. 4,265,754 to Menold, Water Treating Apparatus and Methods.
23. U.S. Pat. No. 4,265,746 to Zimmermann, Sr. et al. Water Treating Apparatus and Methods.
24. U.S. Pat. No. 4,222,021 to Bunker, Jr., Magnetic Apparatus Appearing To Possess a Single Pole.
25. U.S. Pat. No. 2,974,981 to Vervest et al., Arrester For Iron Particles.
26. U.S. Pat. No. 2,613,246 to Spodig, Magnetic System.
27. U.S. Pat. No. 2,560,260 to Sturtevant et al., Temperature Compensated Magnetic Suspension.

SUMMARY OF THE INVENTION

The device and method of the present invention is a new permanent electromagnetic motor generator that recycles back EMF energy (regauging) thus allowing the motor to produce an energy level of COP=0.98, more or less, depending upon configuration, circuitry, switching elements and the number and size of stators, rotors and coils that comprise the motor. The rotor is fixed between two pole pieces of the stator. The motor generator is initially energized from a small starter battery means, analogous to a spark plug, that sends a small amount of energy to the motor, thus stimulating a rotating motion from the rotor. As the rotor rotates, energy is captured from the surrounding electromagnetic field containing an asymmetrical pulse wave of back EMF. The energy produced and captured can be directed in one of several directions, including returning energy to the initial starter battery, rotating a shaft for work and/or sending a current to energize a fan, light bulb or other such device.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a top perspective view of a back EMF permanent electromagnetic motor generator with a single stator and a single rotor.

FIG. **1***a* is a side perspective view of a timing wheel and magnetic Hall Effect uptake switch of the back EMF motor generator.

FIG. **1***b* is a side perspective view of the rotor of the back EMF motor generator.

FIG. **2** is a schematic drawing incorporating circuitry for the back EMF motor generator.

FIG. **3** is a box diagram showing the relationships of the back EMF motor generator circuitry.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is a device and method for creating a back EMF permanent electromagnetic motor generator. As described in the Background Information, this new motor generator conforms to all applicable electrodynamic laws of physics and is in harmony with the law of the conservation of energy, the laws of electromagnetism and other related natural laws.

The back EMF permanent electromagnetic motor generator is comprised of combination of electrical, material and magnetic elements to capture available electromagnetic energy (back EMF) in a recovery rectifier or single diode from output coils. The capturing of back EMF energy is also known as regauging in the art. As an arbitrary starting point in describing this invention, an input battery, as a means of energy, sends power through a power on-off switch and then to a means for timing such as a magnetic timing switch (Hall Effect magnetic pickup switch, a semiconductor) which interfaces with or is in apposition to a magnet on a timing wheel. The timing wheel may contain any number of magnets of one or more, with the South polarity facing outward and in apposition with the Hall Effect pickup switch. The timing wheel is mounted at the end of a shaft that also runs through the center midline of a rotor containing any number of magnets of two or more. The rotor magnets are arranged in a manner wherein they have the same polarity and are equidistant from each other. The shaft has the timing wheel mounted at one end, the rotor, and then a means for work, such as a power take off at the opposite end. However, there are other embodiments in which the position of the rotor, timing wheel and power take-off have other arrangements. The rotor is stabilized to a platform or housing means and is fixed in a stationary position within a stator.

The stator is comprised of a permanent magnet connected to a means for conducting electromagnetic energy such as two parallel bars, each bar having a magnetized pole piece at one end of each bar. The conduction material of the bar may be ferrous, powdered iron, silicon steel, stainless magnetic steel, laminations of conductive material or any other magnetic conductive material. Each bar is wrapped in a conducting means to form an input coil. The means for conducting may be copper, aluminum or any other conductive material suitable for making a coil. The primary or input coil is connected to the switching circuit. A second conductive wrapping on top of the input coil becomes a secondary or output coil. The secondary or output coil is connected to the recovery circuit. The rotor is symmetrically located between the pole pieces of the bars of the stator and contains a series of magnets all having the same polarity, North or South, with each magnet in the rotor being in apposition to the pole piece as the rotor is in rotation around the shaft.

When the rotor is energized from the battery of the switching circuit, there is an initial magnetic field that is instantly overcome as the magnetized pole pieces are in apposition with the rotor magnets. As the rotor begins to move, increasing electromagnetic energy is produced as a result of flux gaiting from the apposed magnets of the rotor and pole pieces. The coils surrounding the bars "buck" the permanent magnet connecting the bars. This is known in the art as the "buck boosting" principle. When the permanent magnet is bucked by the coils, it reverses the polarity of the pole pieces which are apposed to the rotor magnets causing the rotor to increase its rotation or spin. The energy available from the fields that are collapsing in the primary and secondary coils, which creates the back EMF within the system, is now in non-equilibrium. Through circuitry and a switching means, energy can be put back into the system. Available energy captured from the back EMF, may be applied in different directions, including re-energizing the input battery, storage in a capacitor, conversion by a recovery rectifier to be stored in the input battery, a capacitor or a secondary or recovery battery. Recovery rectifiers convert AC to DC. Available energy may be used to energize an electric bulb, fan or any other uses.

The shaft in the midline of the rotor can transfer energy in the form of work through a power take-off. The power take-off may be connected to any number of secondary shafts, wheels, gears and belts to increase or reduce torque.

This is a description of the basic invention, however, there are an innumerable number of combinations and embodiments of stators, rotors, Hall Effect magnetic pickup switches, coils, recovery rectifiers and electronic connecting modes that may be combined on a single shaft or several shafts connected in various combinations and sequences, and of various sizes. There may be any number of stators to one rotor, however, there can be only one rotor if there is a single stator. The number of Hall Effect pickup switches may vary, for example, in the case of multiple stators of high resistant coils, the coils may be parallel to form a low resistant coil so that one Hall Effect pickup with one circuit may fire all of the stators at the same time. The number of magnets in both the timing wheel and the rotor may also vary in number as well as the size and strength of the magnets in gauss units. All types of magnets may be used. The number of winds on both the input and output coils on each conducting bar may also vary in number and in conductive material.

The motor generator, as shown in FIG. **1**, a top perspective view of a single stator, single rotor back EMF motor and is comprised of a means of providing energy, such as input battery **10** connected to power switch **11** (shown in FIG. **2**) and Hall Effect magnetic pickup switch **13**. Magnetic pickup **13** interfaces with timing wheel **12** to form a timing switch. Timing wheel **12** is comprised of four magnets **14** with the South pole of each said magnet facing outward to magnetic pickup **13**. Timing wheel **12** is fixed at one end of shaft **15**. Located on shaft **15** is rotor **16**. Rotor **16** can be of any size, said rotor containing four rotor magnets **17**. Said rotor magnets **17** are arranged in a manner so all have the same polarity. Opposite timing wheel **12** on shaft **15** is a means for work, such as a power take-off **18**. Rotor **16** is mounted in a fixed position with rotor magnets **17** in apposition with magnetized pole pieces **19***a* and **19***b*. Each pole piece **19***a* and **19***b* is connected to iron bars **20***a* and **20***b*. Iron bars **20***a* and **20***b* are connected by a permanent magnet **21**. A means for conduction is wrapped around iron bars **20***a* and **20***b* to form input coils **22***a* and **22***b*. Superimposed upon input coils **22***a* and **22***b* are output coils **23***a* and **23***b*. Output coils **23***a* and **23***b* are connected to full wave bridge first recovery rectifier **24***a*. First rectifier **24***a* is connected to battery **10**.

FIG. **1***a* is a perspective side view of the back EMF Motor Generator timing wheel **12** with Hall Effect magnetic pickup switch **13** in apposition individually to each of four magnets **14** as said timing wheel **12** rotates. Said magnets **14** have the South polarity facing outward with an equidistant separation of 90°.

FIG. **1***b* is a perspective side view of rotor **16** with four rotor magnets **17** with 900 equidistant separation and having the same polarity.

FIG. **2** is a schematic diagram of the motor generator circuitry showing input coil connections from input battery

10 through power switch **11**, transistors **30***a,b*, resistors **31***a–d*, through power supply lead **32** (VCC+) and to magnetic pickup **13**. Magnetic pickup **13** is in apposition with timing wheel magnets **14** located on timing wheel **12**. Off of magnetic pickup **13** is collector lead **33** and ground lead **34**. When current is reversed, it flows through resistor **31***e* and transistor **30***c* to input battery **10**. Input coils **22***a,b* send power to full wave bridge first recovery rectifier **24***a* which then sends power through switch recovery **27** back into the system, and/or to input battery **10**. Output coils **23***a* and **23***b* send power through single diode second recovery rectifier **24***b* to recovery battery **25**. In this particular embodiment, the value and type number of the components are as follows: Hall Effect magnetic pickup switch **13** is a No. 3020; transistor **30***a* is a No. 2N-2955; transistor **30***b* is a No. MPS-8599; and transistor **30***c* is a No. 2N-3055; resistors **31***a* and b are 470 ohms resistors; resistor **31***b* is a 2.2 K ohms resistor; resistor **31***c* is a 220 ohms resistor; resistor **31***d* is a 1 K ohms resistor; and recovery rectifier **24***a* is a 10 Amp, 400 volts bridge rectifier.

FIG. **3** is a box diagram showing the flow of voltage from input battery A, through recovery circuit B, switching circuit C and motor coils D. Motor coils D send available back EMF energy through recovery circuit B, and from B to recovery battery E and input battery A. Available back EMF energy can also flow from switching circuit C to recovery circuit B.

In multiple stator/rotor systems, each individual stator may be energized one at a time or all of the stators may be energized simultaneously. Any number of stators and rotors may be incorporated into the design of such multiple stator/rotor motor generator combinations. However, while there may be several stators per rotor, there can only be one rotor for a single stator. The number of stators and rotors that would comprise a particular motor generator is dependent upon the amount of power required in the form of watts. The desired size and horse power of the motor determines whether the stators will be in parallel or fired sequentially by the magnetic Hall Effect pickup switch or switches. The number of magnets incorporated into a particular rotor is dependent upon the size of the rotor and power required of the motor generator. In a multiple stator/rotor motor generator, the timing wheel may have one or more magnets, but must have one magnet Hall Effect pickup switch for each stator if the stators are not arranged in parallel. The back EMF energy is made available through the reversing of the polarity of the magnetized pole pieces thus collapsing the field around the coils and reversing the flow of energy to the recovery diodes, which is capturing the back EMF.

Individual motors may be connected in sequence with each motor having various combinations of stators and rotors or in parallel. Each rotor may have any number of rotor magnets ranging from a minimum of 2 to maximum of 60. The number of stators for an individual motor may range from 1 to 60 with the number of conducting bars ranging from 2 to 120.

What distinguishes this motor generator from all others in the art is the presence of a permanent magnet connecting the two conducting bars which transfer magnetic energy through the pole pieces to the rotor, thereby attracting the rotor between the pole pieces. With the rotor attracted in between the two pole pieces, the coils switch the polarity of the magnetic field of the pole pieces so that the rotor is repelled out. Therefore there is no current and voltage being used to attract the rotor. The only current being used is the repulsion of the rotor between the two conductive bar pole pieces thereby requiring only a small amount of current to repel the rotor. This is known as a regauging system and allows the capturing of available back EMF energy for use.

Finally, although the invention has been described with reference of particular means, materials and embodiments, it is to be understood that the invention is not limited to the particulars disclosed and extends to all equivalents within the scope of the claims.

What is claimed is:

1. A back EMF permanent electromagnetic motor generator using a regauging process for capturing available electromagnetic energy in the system comprising:

a. a means for producing a source of input energy to said motor generator, said energy conducted through a power switch and a magnetic timing switch;

b. a means for timing said motor generator, having a timing wheel connected to a shaft, said timing wheel comprised of one or more magnets, said magnets arranged so that the South polarity faces outward and is in apposition with said timing switch;

c. a rotor and a means for transferring a force on said shaft, said rotor containing magnets having the same polarity;

d. a stator comprised of two bars made of a magnetic conductive means, said bars connected by a permanent magnet and having a pole piece at one end of each said bar, said pole pieces thereby magnetized and in apposition with magnets of said rotor, said rotor positioned between said pole pieces;

e. an input coil comprised of a conductive material wrapped around each of said bars of said stator, said input coil connected to a circuit;

f. an output coil comprised of a conductive material wrapped around each of said input coil of said bars, said output coils connected to a recovery rectifier or diode;

g. said circuit for transferring electrical power, said circuit connecting said means for producing input energy to said power switch, said circuit comprising a pickup timing switch thereby providing an electronic voltage pulse to said circuit and then to said coils resulting in rotation of said rotor, said rotor turning said timing wheel to provide additional electronic voltage pulses to said circuit and then to said coils again, thereby providing a means for collecting available back EMF energy for storage, use or dissipation into the system.

2. The motor generator of claim **1**, wherein said input producing means is a battery.

3. The motor generator of claim **1**, wherein said magnetic timing switch is a Hall Effect magnetic pickup semiconductor.

4. The motor generator of claim **1**, wherein the number of rotors may range from 1 to 60.

5. The motor generator of claim **1**, wherein the number of magnets in said rotor has a range of 2 to 60.

6. The motor generator of claim **1**, wherein said magnetic conductive means may be ferrous, powdered iron, silicon steel, stainless magnetic steel, laminations of a conductive material or any other magnetic conductive material.

7. The motor generator of claim **1**, wherein the number of magnetic conductive bars has a range of 2 to 120.

8. The motor generator of claim **1**, wherein said means for transferring a force on said shaft may be a power take-off mechanism, gears, flywheels, belts or other methods of transferring force for the purpose of doing work.

9. A back EMF permanent electromagnetic motor generator using a regauging process for capturing available electromagnetic energy in the system comprising:

a. a battery connected to a power switch and a Hall Effect magnetic pickup switch, said magnetic pickup switch in apposition with a timing wheel, said timing wheel on a shaft;

b. a rotor and a power take-off means on said shaft, said rotor having four magnets of the same polarity, said magnets equidistant from each other;

c. a stator, said stator comprised of two iron bars connected by a permanent magnet and wrapped in a first copper conducting material to form an input coil that is connected to a circuit, on top of said input coil, there is wrapped a second copper conducting material to form an output coil with said output coil connected to a recovery rectifier or single diode, said iron bars further having a pole piece at one of each end of said bars, said pole pieces in apposition to said magnets on said rotor;

d. a circuit for transferring electrical power, said circuit connecting said battery to said power switch, said circuit comprising-said Hall Effect pickup switch thereby providing an electronic voltage pulse through said circuit and then to said coils resulting in rotation of said rotor, said rotor turning said timing wheel to provide additional electronic voltage pulses to said coils again, thereby providing a means for collecting available back EMF energy for storage, use or dissipation into the system.

10. A method of producing a back EMF permanent electromagnetic motor generator using a regauging process comprising the steps of:

a. placing a rotor containing magnets of the same polarity on a shaft with a timing wheel on said shaft;

b. placing in apposition to said timing wheel a Hall Effect magnetic pickup switch;

c. placing in apposition to said magnets of said rotor magnetized pole pieces of a stator, said stator comprised of two conducting bars connected by a permanent magnet; input coil connected a circuit;

e. on top of said input coil, wrapping a second conducting material to form an output coil, said output coil connected to a recovery rectifier or a single diode;

f. connecting a circuit for transferring electrical power, said circuit connected to a battery and a power switch, said circuit comprising said Hall Effect magnetic pickup switch thereby providing an electronic voltage pulse to said circuit and then to said coils resulting in increase rotation of said rotor, said rotor turning said timing wheel to provide additional electronic pulses to said circuit and then to said coils again, thereby providing a means for collecting available back EMF energy for storage, use or dissipation into the system.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,392,370 B1
DATED : May 21, 2002
INVENTOR(S) : John C. Bedini

Page 1 of 1

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 13,
Line 19, please delete the hyphen in between "comprising" and "said" and insert a space.

Column 14,
Line 11, after the semicolon, please delete "input coil connected a circuit;".
Line 12, please insert the following new paragraph:
-- d. wrapping said bars with a first conducting material to form an input coil, said input coil connected to a circuit --.

Signed and Sealed this

Tenth Day of December, 2002

JAMES E. ROGAN
Director of the United States Patent and Trademark Office

rexresearch.com

Benjiman [Benjamin] R. TEAL

Magnipulsion Engine

[MagnePulsion]

Peter Lindemann's Magnipulsion Project:
http://www.free-energy.ws/bob-teal.html

DVD: "Electric Motor Secrets" ---
http://www.free-energy.ws/products.html

Interview Video:
http://www.esmhome.org/library/bob-teal/index.html

Interview Video, downloadable:
http://www.livevideowatch.com/?tag=Teal

Company Literature (PDF):
http://www.free-energy.ws/pdf/Teal-MagneTeal.pdf

***Tallahassee Democrat* (Date unknown, ca. 1976)**

Magnipulsion is Answer to Energy Crisis --- Bob Teal

By Doris Morgan

A retired enginer who works in a small shop at his home in Madison says he has invented a machine which might be the answer to the energy crisis.

B.R."Bob" Teal said he got the idea for what he calls his 'magnipulsion' engine when he was writing a still-unpublished science fiction novel.

After retiring from an engineering post in Hawaii with Ling Temco Vought in 1972, Teal and his wife Beatrice moved to Madison where he decided to try out his idea.

Entering Teal's backyard workshop, you see two stainless steel contraptions with wires and switches stuck here and there, his latest magnipulsion engines.

After explaining that the thing started with a car battery, Teal flips several switches which sends current to the magnets placed around the engine.

The one horsepower engine which sounds like a sewing machine while operating is the sixth in a series of electromagnet machines which Teal has developed since he began in 1972.

While the pistons pump up and down, much like a car engine through their magnetic cylinders, the pulleys spin around. Speed is reduced or increased by the flip of a switch that controls the power like the dimming of a light bulb.

“As it is now the machine can power shop tools, water pumps, or conveyers”, Teal said. “You can add horsepower by adding magnets”.

“I have built six engines and the latest will operate a 20-ton conveyer:, Teal said. He said he bought the car batteries more than a year ago and they have never been charged since.

Teal says his invention could save coal and oil which are now required to create electricity. “A small magnipulsion engine could operate a home central air conditioner for about 50 cents per day”, he said.

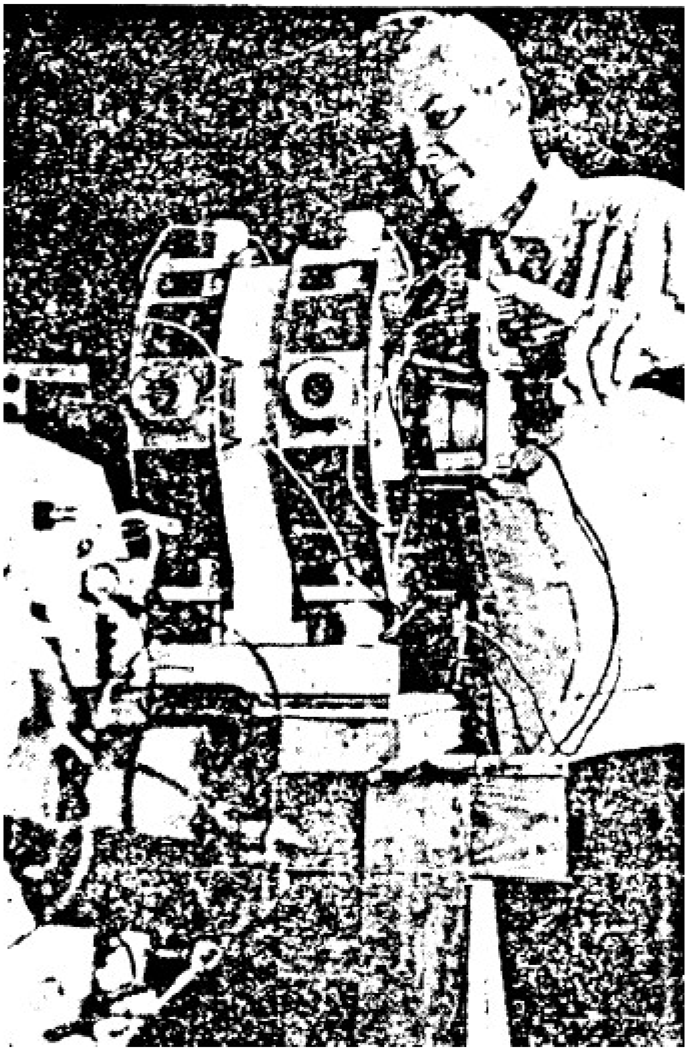

***Valdosta Times* (GA, 13 March 1977)**

Madison Man Builds New Engine

By Elizabeth Butler

The writing of a since fiction book eventually led a Madison resident to invent an electromagnetic engine which he says emits no pollution and takes little current to operate.

B.R. “Bob” Teal, a retired electronics engineer and author, has one patent and two pending on his invention.

“I wrote a science fiction book several years before retiring and I needed the story plausible --- an engine that emitted no pollution --- noise or air pollution --- so in my mind at the time I named it magnipulsion.

“Then after retiring in Honolulu in 1972 we came to Madison to live and having nothing else to do, I decided to build the engine. I built it and it works.”

Teal, 54, said scientists and engineers have come from all over the country and confirmed that it does what they consider to be impossible.

“I have built six engines and the latest will operate a 20-ton conveyor”, said Teal, who was born in Morven, NC, about 60 miles from Charlotte. “It operates fro a standard car battery and draws an extremely small amount of current. I bought the car batteries over a year ago and they have never been charged but once.”

Teal constructed his first engine of wood within a few weeks to see if it would work. The second engine was a crude one of metal and later he went to Jacksonville to get parts made for a prototype.

“I have dreams of building one large enough to run cars and boats”, teal said. “Judging from what I have done so far, I guess it would cost $400 or $500. There are very few moving parts so you would not need highly trained mechanics.”

Since the United Press International carried a story of his invention in May 1976, Teal has received hundreds of letters about the invention from around the world and has answered all of them.

“I have been attempting to get the government interested in hopes of getting a grant to build a shop and hire qualified personnel”, he said. “If I don’t succeed in the very near future in getting the grant, my attorney is preparing to go public and sell stock”.

Teal said telegrams from all over the nation and from foreign cities have been sent to President Jimmy Carter asking him to investigate the potential of the engine.

The electromagnetic engine is not the first invention for Teal, While working as an RCA engineer and filling a contract with the Navy in 1964, he was responsible for a classified invention which related to radio frequency cables and was valued at $50 million. “I didn’t get a penny for that”, said Teal, who has a letter from the navy Department giving him full credit for the invention.

Teal, who attended Wofford College in Spartanburg SC, and the University of Hawaii, retired from the Coast Guard after 20 years in 1962. He then worked for Ling Temco Vought, RCA, and finally went into Civil Service work with te Air Force, retiring in 1972.

During 1975-76 he wrote a column, “Is Justice?” for the Madison Carrier and last year he published a book of poems.

He is married to the former Beatrice Mae Cole of Baltimore MD, and they have two married children.

US Patent # 4,024,421

Magnetically Operable Engine or Power Plant

Benjiman Teal

Abstract

A magnetically operable engine or power plant embodying a rotary crankshaft having two or more offset cranks which, by means of respective connecting rods pivoted thereto, and also to the sliding cores or armatures of electromagnets, are actuated to apply torque to the crankshaft for driving purposes. Electrical current is supplied to the magnet windings by fixed distributor switches which are successively operated by one or more cams effectively mounted on the crankshaft, the switches receiving pulses of current in timed relationship and in such a manner that torque is continuously applied to the crankshaft.

US Cl. 310/24
Intl Cl. H02K 7/06 (20060101); H02K 7/075 (20060101); H02K 007/06

References Cited

U.S. Patent Documents: 568947 // 1131614 // 1347002 // 1886040 // 2056719 // 2639544 // 3688136 // 3832608

Description

The present invention relates to a magnetically operable engine or power plant which, in a general way, functions in the manner of an internal combustion engine in that it employs a rotary crankshaft having offset cranks which derive their motion from power driven reciprocating members. Unlike a conventional combustion engine, the reciprocating members, instead of being motivated by the explosion of combustible gases in a combustion chamber, are caused to reciprocate by magnetic attraction, such members being in the form of cores or armatures which are associated with electromagnets, there being at least one magnet for each crank. Motivating current impulses are successively supplied to the various magnets by distributor means embodying respective normally open circuit making and breaking devices which are successively closed by one or more cams that rotate with the camshaft. The basic principle involved in thus applying torque to the crankshaft may appropriately be referred to as "magnipulsion" (a coined word) and such term may be employed where appropriate throughout this specification.

The present magnetically operable engine or power plant is capable of being used as a power source in connection with a wheeled automotive vehicle of either the passenger-carrying type or as a toy automobile. It is also capable of being used as a fixedly mounted power plant for driving adjacent machinery of various sorts but, irrespective of the particular use to which the invention may be put, the essential features thereof are at all times preserved.

The provision of an engine which is extremely simple in its construction and which therefore may be manufactured at a low cost; one which is comprised of a minimum number of parts, particularly moving parts, and which therefore is unlikely to get out of order; one which is rugged and durable and which therefore will withstand rough usage; one which is smooth and silent in its operation; one which is capable of ease of assembly and disassembly for purposes of inspection of parts, replacement or repair thereof; and one which otherwise is well adapted to perform the services required of it are further desirable features which have been borne in mind in the production and development of the present invention.

The provision of an engine or power plant such as has briefly been outlined above, and possessing the stated advantages, constitutes the principal object of the invention. Numerous other objects and advantages of the invention, not at this time enumerated, will become readily apparent as the nature of the invention is better understood.

In the accompanying two sheets of drawings forming a part of this specification, one illustrative embodiment of the invention has been shown .

In these drawings:

FIG. 1 is a perspective view of a magnetically operable power plant or engine embodying the principles of the present invention, portions of the framework or chassis being broken away in the interests of clarity;

FIG. 1

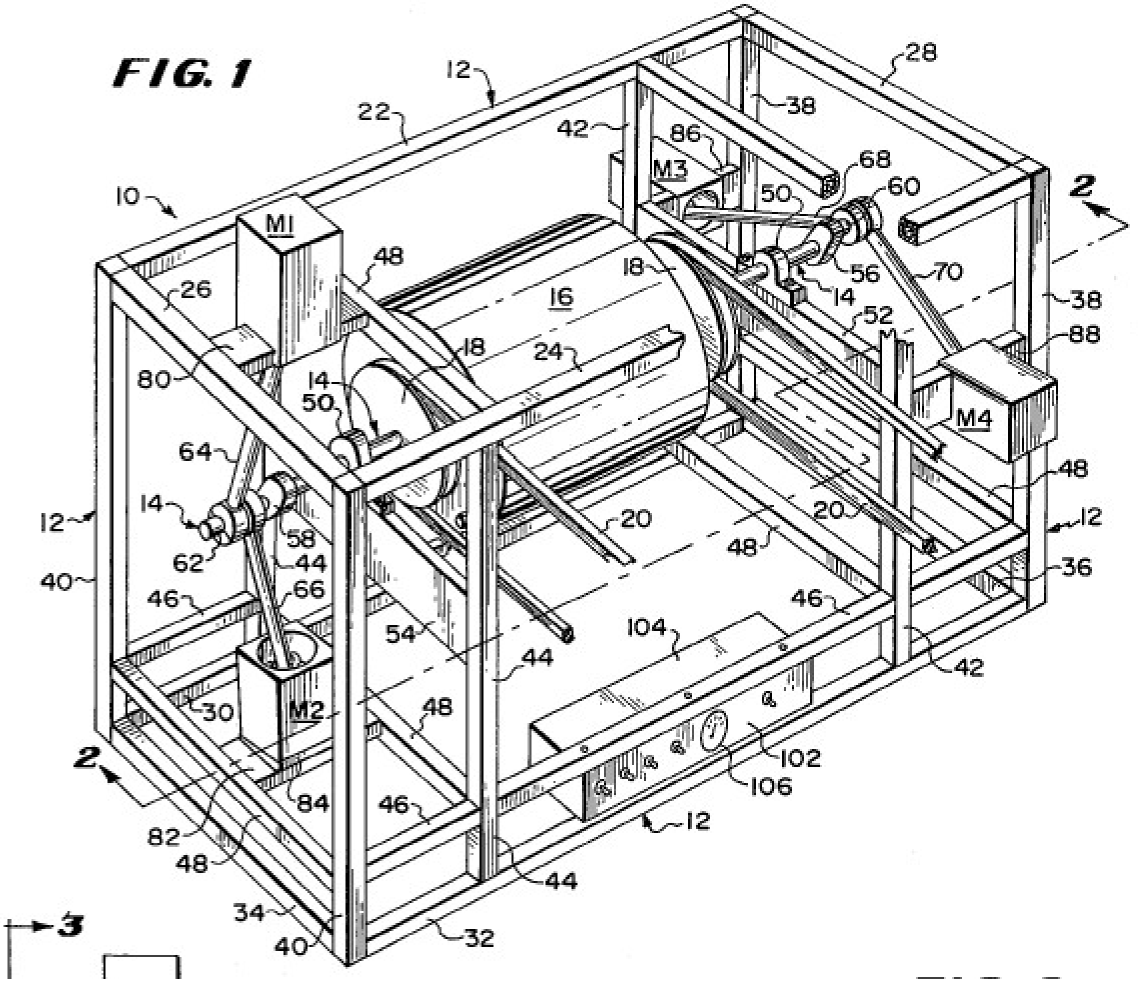
12
22
28
42
38
86
M3
68
50
60
2
10
M1
48
18
16
56
70
14
38
26
52
88
18
24
80
14
50
M4
64
48
14
20
20
12
12
58
48
62
44
36
40
46
66
46
54
104
44
42
30
M2
48
2
102
84
46
106
82
12
48
44
34
3
40
32

FIG. 2 is a sectional view taken substantially on the vertical plane indicated by the line 2--2 of FIG. 1 and in the direction of the arrows;

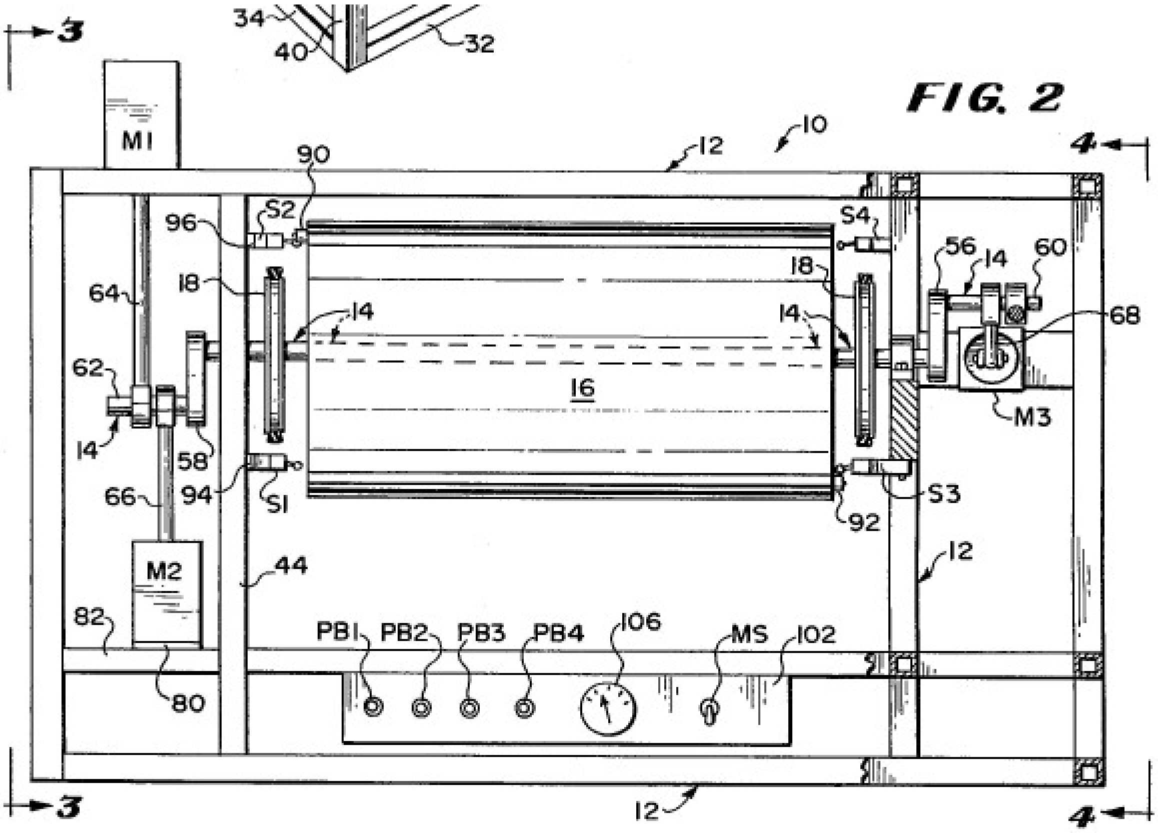
FIG. 2
3
34
40
32
M1
10
12
4
90
96
S2
S4
18
64
56
14
60
68
62
14
16
M3
14
58
66
94
S1
92
S3
M2
44
12
82
PB1
PB2
PB3
PB4
106
MS
102
80
3
12
4

FIG. 3 is an end elevational view taken in the direction of the arrows associated with the line 3--3 of FIG. 2; **FIG. 4** is an end elevational view taken in the direction of the arrows associated with the line 4--4 of FIG. 2;

FIG. 3

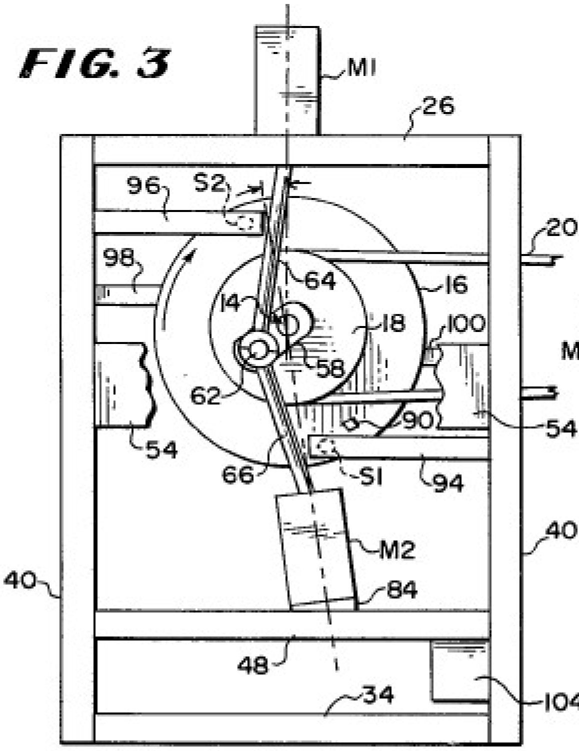
M1
26
96
S2
20
98
64
16
14
18
100
M4
58
62
54
90
54
66
S1
94
M2
40
40
84
48
34
104

FIG. 4

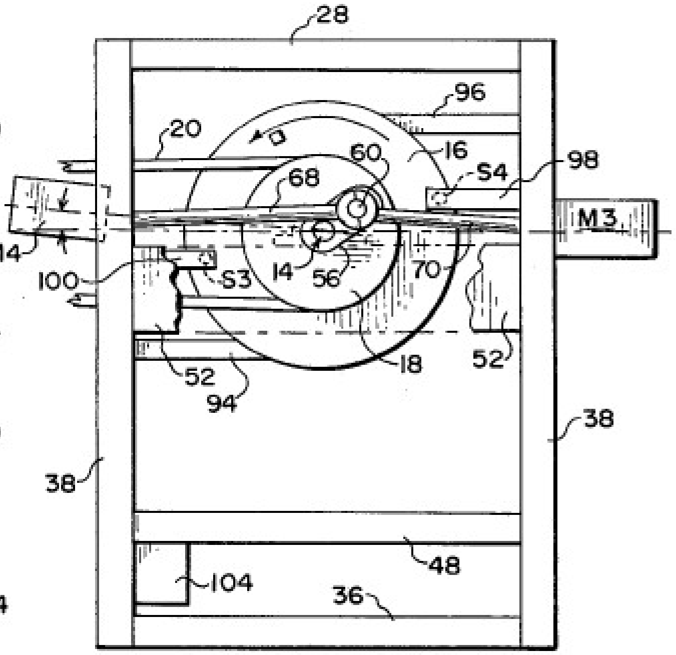
28
96
20
16
60
S4
98
68
M3
100
S3
14
56
70
52
18
52
94
38
38
104
48
36

FIG. 5 is an enlarged detail sectional view taken through one of the electromagnets and its associated connecting rod and crank which are employed in connection with the invention; and

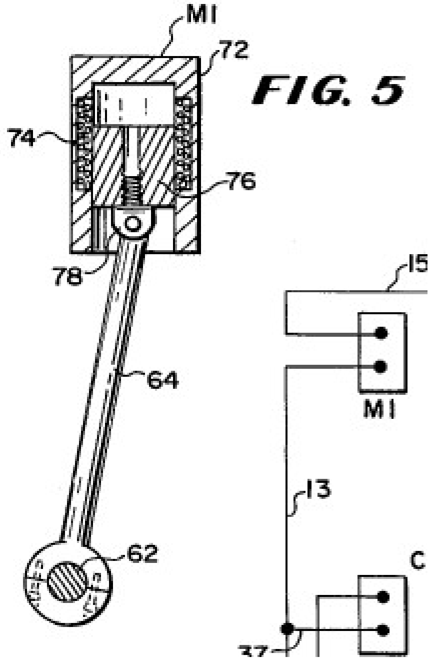

FIG. 5

M1
72
74
76
78
64
62
15
M1
13
C
37

FIG. 6 is an electric circuit diagram of the power plant or engine.

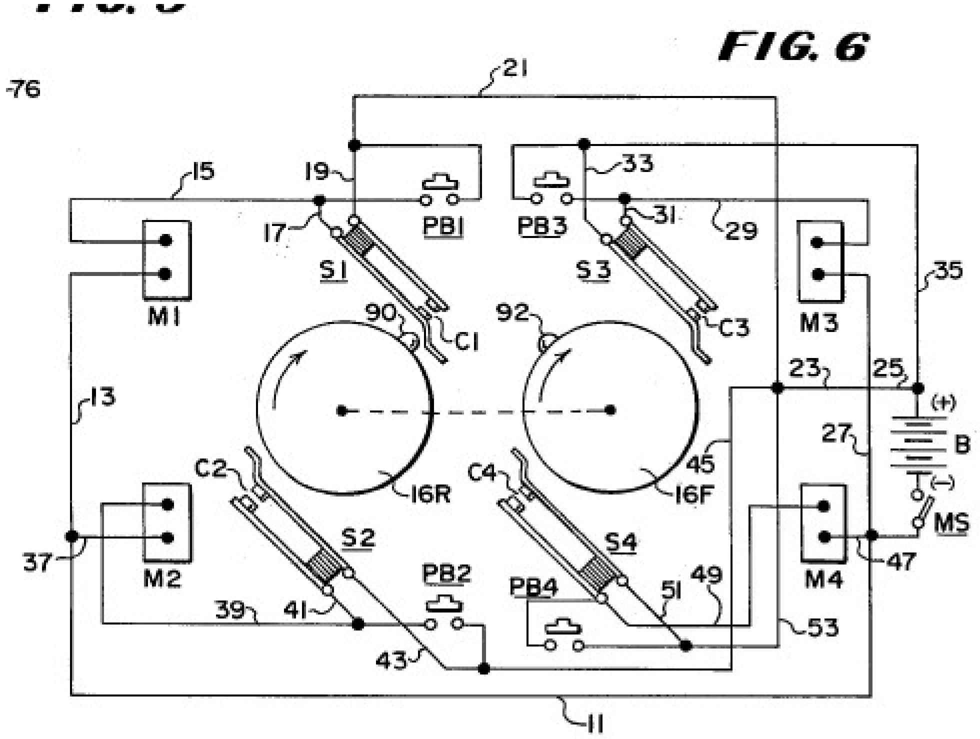

FIG. 6

76
21
15
19
33
31
29
17
PB1
PB3
S1
S3
M1
M3
35
90
92
C1
C3
23
25
(+)
13
27
B
(-)
C2
C4
16R
16F
45
MS
S2
S4
37
M2
M4
47
PB2
PB4
49
39
41
51
53
43
11

Referring now to the drawings in detail, and in particular to FIGS. 1 and 2, a power plant or engine constructed according to the invention is designated in its entirety by the reference numeral 10 and it embodies in its general organization a chassis or framework 12 which serves to rotatably support an elongated crankshaft 14 on which there is mounted a relatively massive flywheel 16 in the medial region thereof. A pair of pulleys 18 on opposite sides of the flywheel 16 have associated therewith respective drive belts 20 which may extend to a suitable transmission (not shown) in the case of a wheeled automative vehicle, or to the input drive element in the case of a stationary equipment which is to be driven by the power plant.

The framework 12 is comprised of an upper rectangular frame having longitudinal frame members 22 and 24 and transverse frame members 26 and 28, and a lower rectangular frame having longitudinal frame members 30 and 32 and transverse frame members 34 and 36. The framework 12 further includes a pair of intermediate posts 40 and front and rear vertical intermediate posts 42 and 44 respectively. A series of longitudinal struts 46 and transverse struts 48 extend variously between the posts 38, 40, 42, and 44 and establish an intermediaterectangular frame a slight distance above the lower rectangular frame 30, 32, 34 and 36.

A pair of bearing assemblies 50 are supported upon transverse support bars 52 and 54 and serve to rotatably support the crankshaft 14. Such crankshaft is provided with crank arms 56 and 58 on opposite sides of the flywheel 16 and with offset cranks 60 and 62. The crank 62 is connected by means of connecting rods 64 and 66 to respective electromagnets M1 and M2 which are fixedly mounted on the framework 12, while the crank arm 60 is similarly connected by connecting rods 68 and 70 to respective electromagnets M3 and M4, all in a manner that will be made clear presently.

The various electromagnets M1, M2, M3 and M4, together with their associated connecting rods 64, 66, 68 and 70 are substantially identical and therefore a description of one of them will suffice for them all. Accordingly, the magnet M1 (see also FIG. 5) embodies a magnet casing or shell 72 within which there is disposed a magnet winding 74. An armature or core 76 is slidably disposed within the casing 72 and is pivotally connected as indicated at 78 to the associated connecting rod 64. The remaining magnets M2, M3 and M4 are similarly connected to their associated connecting rods

The magnet M1 is mounted in a vertical position upon an upper shelf member 80 while the magnet M2 is mounted in a slightly inclined position upon a lower shelf member 82. As best shown in FIGS. 1, 2 and 3, the magnet M2 is seated upon a wedge-shaped base plate or block 84 which serves to misalign the axis of the magnet M2 from the axis of the magnet M1 by a small angle for a purpose that will be made clear presently. The magnet M3 is mounted in a horizontal position by means of a support member 86 while the magnet M4 is similarly mounted in a substantially horizontal position by means of a support member 88, the axis of the magnet M4 being slightly misaligned with respect to the axis of the magnet M3 likewise for a purpose that will be set forth subsequently.

Referring now to FIGS. 2, 3 and 4, one end of the flywheel 16 carries a cam 90 which is designed for successive engagements with a pair of substantially diametrically opposed microswitches S1 and S2 which are carried at the ends of a pair of horizontal supporting bars 94 and 96 respectively. Similarly, the other end of the flywheel 16 carries a cam 92 which is designed for successive engagement with a pair of substantially horizontally opposed microswitches S3 and S4 which are carried at the ends of a pair of horizontal supporting bars 98 and 100 respectively. As will become more readily apparent when the operation of the herein described magnetically operable engine or power plant is set forth in connection with the circuit diagram of FIG. 6, the arrangement of the various cams and microswitches is such that upon rotation of the crankshaft 14 and flywheel 16, the contacts associated with the microswitches S1, S3, S2 and S4 will become individually closed, successively and in the order named.

Considering now the operation of the herein described magnetically operable engine or power plant 10, and with reference to FIG. 6, it will be assumed that initially the position of the crankshaft 14 is such that, as shown in FIG. 3, energization of the magnet M1 will swing the crank 62 upwardly at such time as the cam 90 engages the microswitch S1. Closure of the C1 contacts associated with the switch S1 will establish a circuit leading from the negative side of the battery B, through the master switch MS, through leads 11, 13, magnet M1, leads 15, 17, contacts C1 of the switch S1, and leads 19, 21, 23 and 25, back to the positive side of the battery B. Energization of the magnet M1 will draw the core or armature 76 (FIG. 5) into the shell 72 and thus place the connecting rod 64 under tension so as to pull the crank 62 upwardly, thereby placing the crankshaft under torque for motivating purposes.

At approximately 90.degree. in the engine cycle, the cam 90 will engage the switch S3 and closure of the contacts C3 thereof will establish a circuit from the master switch MS through lead 27, magnet M3, leads 29, 31, C3 contacts of the switch S3, and leads 33, 35 back to the battery B, thus applying torque to the crankshaft 14 by placing the connecting rod 68 under tension.

At approximately 180.degree. in the engine cycle, the cam 90 will engage the switch S2 and closure of the C2 contacts thereof will establish a circuit from the master switch MS through leads 11, 37, magnet M2, leads 39, 41, contacts C2 of the switch S2, and leads 43, 23, 25 back to the battery B, thus applying torque to the crankshaft 14 by placing the connecting rod 66 under tension.

At approximately 270.degree. in the engine cycle, the cam 92 will engage the switch S4 and closure of the contacts C4 thereof will establish a circuit from the master switch MS, through lead 47, magnet M4, lead 49, C4 contacts of the switch S4, and leads 51, 43, 23, 25 back to the battery B, thus placing the connecting rod 70 under tension and applying torque to the crankshaft 14. The cycle is repetitious.

As shown in FIGS. 1 and 2, a control panel 102 is associated with a container or box 104 for the battery B and such panel may be provided with the aforementioned master switch MS and an indicator 106 which may disclose the amperage of current flowing from the battery B. Four push button switches PB1, PB2, PB3 and PB4 may also be provided on the control panel 102 and, as shown in FIG. 6, such push button switches are arranged so that they operate upon closure thereof to establish shunts across the respective microswitches S1, S2, S3 and S4 so that an initial momentary closure of a selected push button switch on a trial and error basis may cause torque to be applied to one or the other cranks 60 or 62 for initial application of torque to the crankshaft 14 in the event that neither crank is in a favorable position for engine starting at the time that the master switch MS is initially closed.

It is to be noted at this point that although the cams 90 and 92 are disclosed in FIG. 2 as being disposed at approximately 180.degree. from each other on the flywheel 16, such cams appear in FIG. 6 as being 90.degree. apart. It should be understood however that the disclosure of FIG. 6 is purely schematic and is intended to illustrate only the sequence of operation of the four switches S1, S2, S3 and S4 under the control of the cams 90 and 92 and that the front and rear end faces 16F and 16R shown in FIG. 6 rotate in the same direction since they do not represent front and rear end faces in the sense that they are illustrated in FIGS. 3 and 4. In other words, FIG. 6 is predicated solely upon the sequence of microswitch operation and does not represent a true positioning of parts such as is disclosed in FIGS. 1 to 4 inclusive. It is also to be noted that by reason of the slight axial misalignment of the magnets M1 and M2, and of the magnets M3 and M4, continuous crankshaft movement is effected since at no time will the engine or power plant 10 attain a position of dead-center where closure of the master switch crankshaft 14.

The invention is not to be limited to the exact arrangement of parts shown in the accompanying drawings or described in this specification as various changes in the details of construction may be resorted to without departing from the spirit of the invention. For example, the precise placement of the various cams 90 and 92 and microswitches S1, S2, S3 and S4 on the flywheel 16 and framework 12 may be varied if desired as regards their angular relationship with respect to one another, the only criteria being the attainment of sequential operation of the switches in such a manner that torque is applied to the crankshaft 14 at all times by at least one of the four connecting rods. Furthermore, if desired a greater or lesser number of electromagnets suitably mounted on the framework 12, together with a commensurate modification of the nature of the crankshaft 14, may be resorted to if deemed appropriate. Therefore, only insofar as the invention has particularly been pointed out in the accompanying claims is the same to be limited.

US Patent # 4,093,880

Magnetically Operable Engine

6 June 1978

Benjiman Teal

Abstract

A combination device for developing a mechanical output from electrical energy which uses at least one electrical magnet solenoid, and preferably a plurality of same, together with an associated timing mechanism for actively controlling the time and degree of energization of said electrical magnet(s). Preferably at least one of the electrical magnetic solenoid structures may be provided with an air compressor structure for increasing the normal atmospheric pressure to a desired pressure.

US Cl. 310/24
Intl. Cl. H02K 7/06 (20060101); H02K 7/065 (20060101); H02K 041/00

References Cited -- U.S. Patent Documents: 2056719 // 2570766 // 2588753 // 2639544 // 3832608 // 4019103

Description

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates generally to devices changing electrical energy into mechanical energy in a form which may be used to power useful work applications.

2. Description of the Prior Art

A common problem with known type devices for converting electrical energy into mechanical energy is that they often times are very inefficient and waste a lot of energy in the conversion process. This is a great disadvantage in today's age when conservation of energy is extremely important.

Another problem with known type engines utilizing electrical magnets for part of their structutre is that the associated slidable solenoid mechanism is normally used just for converting the electrical energy into mechanical drive output without any direct actuation of useful work structure directly from the movable solenoid armature.

SUMMARY OF THE INVENTION

An object of the present invention is to provide an electrical magnetic operated power engine for converting electrical energy into useful mechanical energy.

Another object of the present invention is to provide a magnet powered engine having a plurality of electromagnetic solenoids with electrical coils and movalbe solenoid armatures therewithin arranged in substantially the same plane and connected with associated mechanical transfer mechanism to convert the movement of the electric solenoid armatures into the desired mechanical power. A suitable timer structure is also associated with this mechanical structure for controlling the time and degree of energization of the respective electric-magnet coils for the various electro-magnetic solenoids.

A further object of the present invention is to provide an electro-magnetic solenoid engine structure which incorporates and has combined therewith a mechanical sturcture for compressing air at atmospheric pressure to a greater pressure for direct conversion of the electrical source power into a useful product.

An additional further object is to provide an electromagnetic type engine which includes a plurality of electromagnetic coils in assoicated timing structure arranged in substantially the same plane, with two or more of these structures being stacked one upon the other in order to greatly increase the overall output of the entire structure.

The magnet power engine and air compressor combination of this invention has a number of new and unique features not known or used before. A basic support structure is provided having at least one bank of electro-magnetic solenoids arranged in substantially the same plane thereon. Each electro-magentic solenoid has a slidable and movable solenoid armature therewithin which is mechanically operated and connected to associated mechanical structure for converting the slidable movement thereof into a rotary shaft power output. A timing mechanism is also associated with this mechanical structure for appropriately connecting the electrical power source to at least one electromagnetic coil, and preferably a plurality of same, so that they will operate in proper timing for directing the slidable movement of the solenoid armature back and forth within the solenoid electro magnetic coil.

Also preferably combined with the aforesaid structure is an attachable cylindrical sleeve structure having a slidable piston therewithin and appropriate air valves for connection with the slidable solenoid armature so that useful work may be directly accomplished such as by compressing atmospheric air to a much higher desired pressure. This structure may be provided for each of the electro-magnetic solenoids, or on less than all of same.

These together with other objects and advantages which will become subsequently apparent, reside in the details of constructuon and operation as more fully hereinafter described and claimed, reference being had to the accompanying drawings forming a part hereof, wherein like numerals refer to like parts throughout.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a top plan view of the device of this invention.

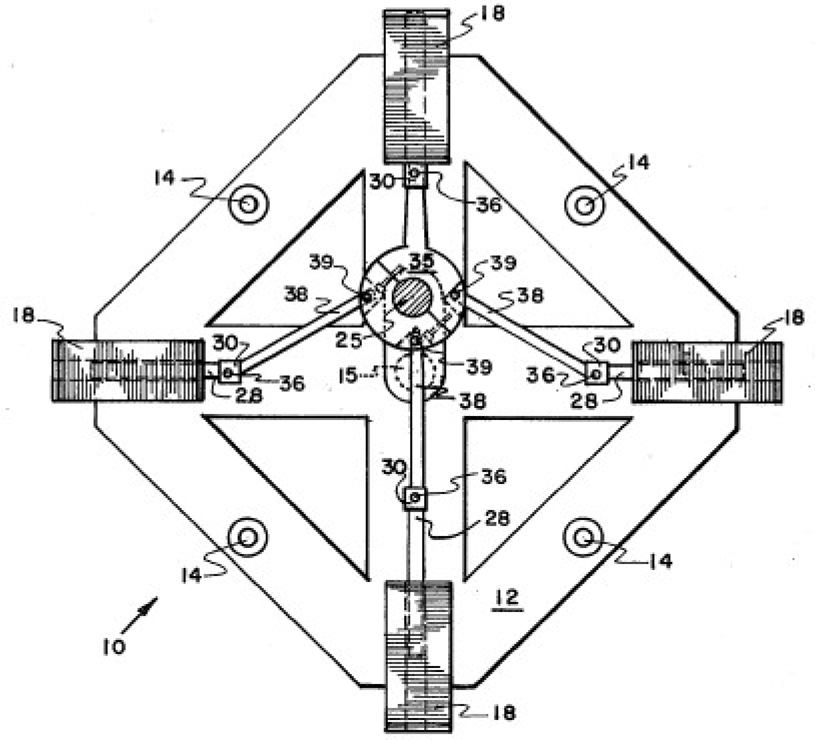
18
14
14
30
36
39
35
39
38
38
18
18
25
30
30
28
36
15
39
38
36
28
30
36
28
14
14
12
10
18

FIG. 1

FIG. 2 is a side elevational view of this invention.

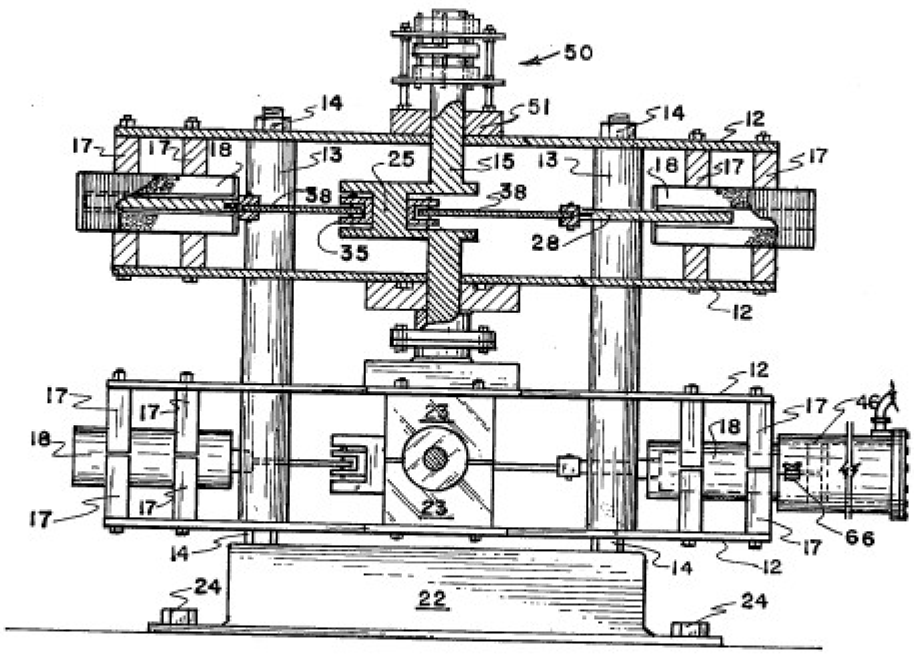
50
14
51
14
12
17
17
18
13
25
15
13
18
17
17
38
38
35
28
12
12
17
17
23
18
17
46
18
18
47
23
17
17
66
14
14
12
17
24
22
24

FIG. 2

FIG. 3 is a cross-sectional view of an individual electro-magnetic unit for the motor of this invention.

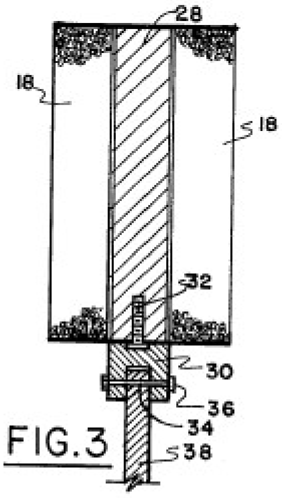
28
18
18
32
30
36
34
38

FIG.3

FIG. 4 is an enlarged cross-sectional view of the timing switch mechanism for the motor of this invention.

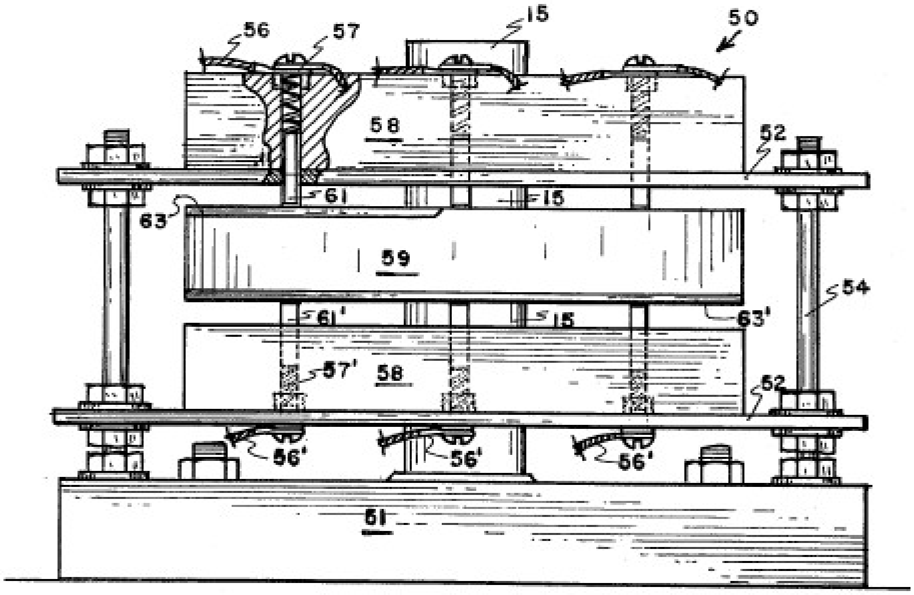
56
57
15
50
58
52
61
15
63
59
54
61'
15
63'
57'
58
52
56'
56'
56'
51

FIG. 4

FIG. 5 is a schematic diagram of the electrical hookup for a four unit single bank of the motor of this invention.

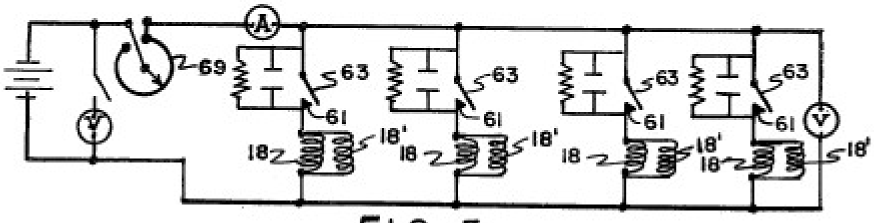
A
69
V
63
63
63
63
61
61
61
61
V
18
18'
18
18'
18
18'
18
18'

FIG.5

FIG. 6 is a cross-section of a single solenoid engine unit with air compressor attachment.

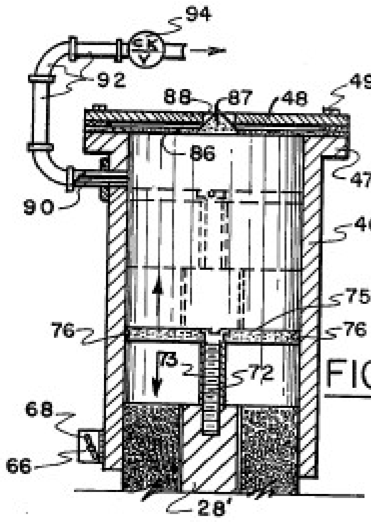

FIG. 6

FIG. 7 is a top plan view of the air compressor head end of FIG. 6.

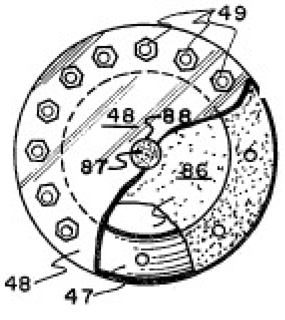

FIG.7

FIG. 8 is a schematic diagram of an alternate electronic electrical hookup for the motor.

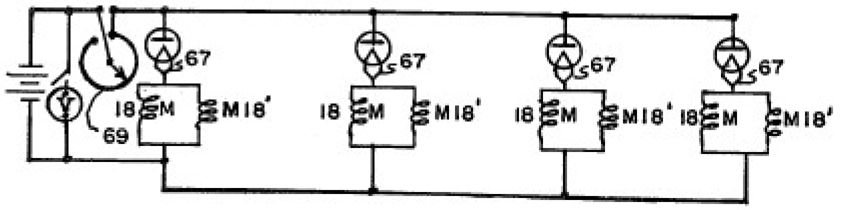

FIG. 8

DESCRIPTION OF THE PREFERRRED EMBODIMENT

Looking at FIG. 1 of the drawings, reference numeral 10 indicates one bank of the electro-magnet motor of this invention. Looking at FIG. 2, two banks, one being stacked on top of the other may be seen.

Each bank basically is provided with a support structure 12 approximately square in shape with two of said support squares being provided for each engine bank. Appropriate spaces separate these support squares and together with nuts and bolts 14 form the basic support for each engine bank. As may be best seen in FIG. 2 the lower must bank is supported from any desired support surface by means of another support structure 22 with appropriate bolts 24. Support brackets 17 appropriately mounted on the respective squares 12 support the respective electro-magnet coils 18 therebetween. As seen in FIG. 1, this bank is provided with four electro-magnets appropriately spaced at the four corners of the square support frame. It has been discovered in actual practice that a bank of four electro-magnets works extremely well and provides a fairly balanced power output. Each electro-magnet coil 18 is provided with a suitable number of coil windings to effect the slidable movement of a centrally mounted solenoid armature 28 therewithin.

As best seen in FIG. 3, the one end of the solenoid armature 28 is appropriately center tapped for reception of a threaded screw 32 therewithin. This threaded screw 32 appropriately secures a block of aluminum or other non magnetic material 30 to this end of the armature. The block 30 forms a connecting link through suitable apertures provided therewithin and by means of a connection of a pin 34 with locknut 36 thereon to a connecting rod portion 38. This connecting rod 38 is suitably pivoted at 39 to the disc 35 which is mounted so as to transfer power to the connecting rod crank 25. The connecting rod crank 25 is suitably formed at appropriate points along a crank-shaft 15 as best seen in FIG. 2. If only a single engine bank is to be used just a single crank-shaft throw 25 is necessary, but for an engine having two engine banks as seen in FIG. 2, two such throws are required and as the number of banks increases, the corresponding number of crank-shaft throws is necessitated.

In operation as the electro-magnets are energized in proper sequence, the electrical input will be transformed into mechanical output through the afore described mechanism. In order to assure that the electro-magnets are operated in proper time relationship a timing mechanism is also provided. Such timing mechanism is indicated in general by reference numberal 50 in FIG. 2, and shown in enlarged detail in the cross-sectionalview of FIG. 4.

A mechanical coupling unit 51 appropriately supports the timing mechanism at one end of the electro-magnet motor and by means of appropriately threaded bolts 54 support plates 52 therefrom. The plates 52 are spaced from each other and each supports thereupon insulator blocks 58. These blocks support electric connectors 56 which in turn connect with bias springs 57 and electrical brushes 61. The insulating blocks 58 are fixed upon the supports 52 and are non-roatable, but the center disc 59 supported therebetween is mounted for driven connection from the end of the crankshaft 15. Each side of the rotatable disc 59 is provided with conducting segments 63 which are internally connected to each other so that the electrical energy connected to input contacts 56 on the upper side of the timing mechanism may be appropriately transferred through springs 57, brushes 61 through the segment 63 to the lower brushes 61, springs 57' and contacts 56' which are appropriately connected to the respective electro-magnets of the engine. Of course the segment 63, 63' are appropriately designed for the number of electro-magents used for the respective banks as well as the total number of such banks and are connected by appropriate wiring, not shown, in such a manner that each electromagnet will be energized for substantially the full stroke of its slidable solenoid armature and thus transfer the maximum of energy from electrical form to mechanical form.

FIG. 5 shows an electrical schematic with one of the electro-magnet solenoids having coils 18 and 18' indicated and with contact brushes 61 and timing segment 63 indicated schematically. Also shown across the timing contact 61, 63 are resisters and capacitors in parallel. The purpose of this is to minimize arcing of the make and break contacts of the respective brushes and contact segments.

FIG. 8 shows an alternative form of energizing the electro-magnet solenoids which is generally referred to as a solid-state switching and which is actually performed by diodes 67. The current entering the solenoids may be regulated by the rheostates 69 as shown in FIGS. 5 and 8.

Looking at FIG. 2 on the right side thereof, an air compressor sleeve 46 is shown appropriately secured by clamp means 66 to one of the electro-magnet structures. FIG. 6 shows this structure in enlarged detail. The outer end of the solenoid armature 28' is appropriately tapped for reception of a threaded bolt 72 therewithin The enlarged head of the threaded bolt together with a spacer sleeve 73 appropriately support a flat piston member 75. This piston member 75 may be appropriately provided with a compression ring 76 of conventional type. The clamp structure 66 to secure the sleeve 46 at the end of the electro-magnet is preferably provided with a thumb nut 68 or the like to secure same. The other end of the sleeve 46 has an outstanding flange periphery 47 for suitably securing a cylinder head 48 thereto. Appropriate bolts and nuts 49 together with suitable gaskets form an air-tight seal between the head 48 and the sleeve 46, 47. For proper compressing operation to be effected, it is necessary that suitable valve mechanism be associated with this compressor structure. This is in the form of a flexible diaphragm 86 which is secured around the circumference thereof between the head 48 and the flange bracket 47. At the center of the flexible diaphragm 86 is mounted a cone shaped protrusion 87 which together with an aperture 88 provided in the center of the head 48 functions as a one way intake air valve. That is, when the piston 75 is moving away from the head and the diaphragm, air will be drawn into the inside of the cylinder sleeve 46 through the aperture 88, and then in turn on the compression stroke of the head 75 the protrusion 87 will seal aperture 88 and then the escape of air out this opening wil be prevented. Another outlet 90 appropriately connected by tubing 92 to a one way check valve 94 provides an outlet for the compressed air and yet prevents backflow thereof. All or some of the connecting tubing 92 may be of a flexible nature as desired.

As can be readily envisioned, when one or a plurality of the electro-magnets are provided with such air compressor sleeves, a direct useful engine output in the form of air under high pressure is achieved. As so constructed, the entire apparatus may be used solely for providing compressed air for various useful purposes. However, by connecting suitiable work apparatus to the output of crank-shaft 15, near the timing mechanism 50, other devices may be driven from this electro-magnet structure.

The foregoing is considered as illustrative only of the principles of the invention. Further, since numerous modifications and changes will readily occur to those skilled in the art, it is not desired to limit the invention to the exact construction and operation shown and described, and accordingly all suitable modifications and equivalents may be resorted to, falling within the scope of the invention.

Your Support Maintains this Service --

BUY

The ***Rex Research Civilization Kit***

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

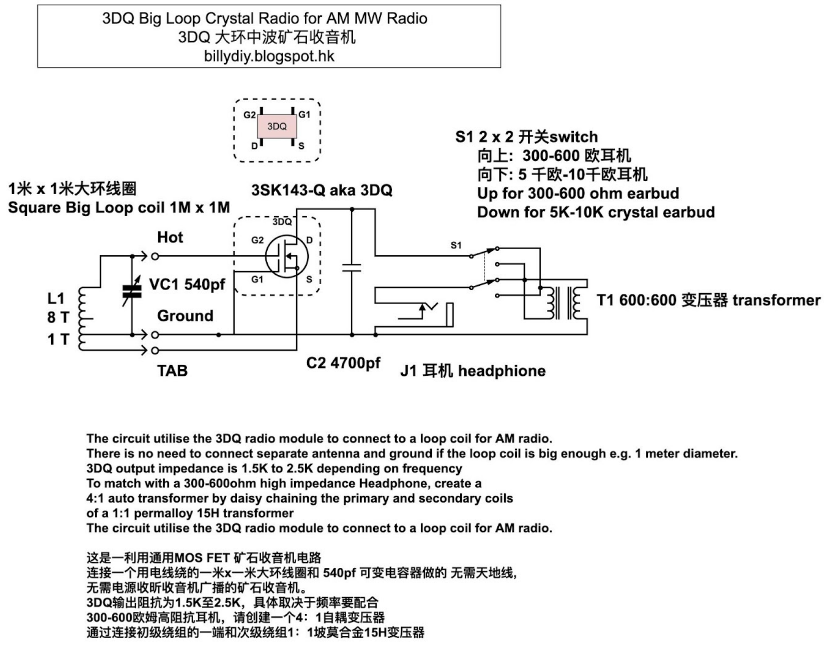

3DQ Big Loop Crystal Radio for AM MW Radio
3DQ 大环中波矿石收音机
billydiy.blogspot.hk

G2
3DQ
G1
D
S
S1 2 x 2 开关switch
向上：300-600 欧耳机
向下: 5 千欧-10千欧耳机
Up for 300-600 ohm earbud
Down for 5K-10K crystal earbud
1米 x 1米大环线圈
Square Big Loop coil 1M x 1M
3SK143-Q aka 3DQ
Hot
L1
8 T
1 T
VC1 540pf
Ground
TAB
C2 4700pf
J1 耳机 headphione
S1
T1 600:600 变压器 transformer

The circuit utilise the 3DQ radio module to connect to a loop coil for AM radio.
There is no need to connect separate antenna and ground if the loop coil is big enough e.g. 1 meter diameter.
3DQ output impedance is 1.5K to 2.5K depending on frequency
To match with a 300-600ohm high impedance Headphone, create a
4:1 auto transformer by daisy chaining the primary and secondary coils
of a 1:1 permalloy 15H transformer
The circuit utilise the 3DQ radio module to connect to a loop coil for AM radio.

这是一利用通用MOS FET 矿石收音机电路
连接一个用电线绕的一米x一米大环线圈和 540pf 可变电容器做的 无需天地线，
无需电源收听收音机广播的矿石收音机。
3DQ输出阻抗为1.5K至2.5K，具体取决于频率要配合
300-600欧姆高阻抗耳机，请创建一个4：1自耦变压器
通过连接初级绕组的一端和次级绕组1：1坡莫合金15H变压器

[https://3.bp.blogspot.com/-Y2wGI9ZEtF4/XLlm6Sz1ofI/AAAAAAAADGo/cCR4OLiB9kcYAuEIbEHzxWYFR3OO5zKxgCLcBGAs/s1600/3DQ%2Bbig%2Bloop%2Bam%2Bradio.jpg]

Posted 18th April by Billy

0 Add a comment

18th April Universal 3DQ MOS FET Crystal Radio Tuner Module for AM SW and FM

这是一款适用于**AM SW FM**的通用**MOS FET**矿石收音机电路
连接该无线电频带所需的线圈和可变容量就可无需电源收听收音机广播。
3DQ输出阻抗为**1.5K**至**2.5K**，具体取决于频率
要配合**300-600**欧姆高阻抗耳机，请创建一个
4：1自耦变压器，通过连接初级绕组的一端和次级绕组
1：1坡莫合金**15H**变压器

This is a general MOS FET crystal radio circuit that works for AM SW FM by connecting the appropriate coils and variable capacity required for that radio band.
3DQ output impedance is 1.5K to 2.5K depending on frequency
To match with a 300-600ohm high impedance Headphone, create a
4:1 auto transformer by connecting one end of the primary with the secondary windings
of a 1:1 permalloy 15H transformer

中文廣東話视频 Chinese Video Link

https://www.youtube.com/edit?o=U&video_id=DIG2ZJQ80D0 [https://www.youtube.com/edit?o=U&video_id=DIG2ZJQ80D0]

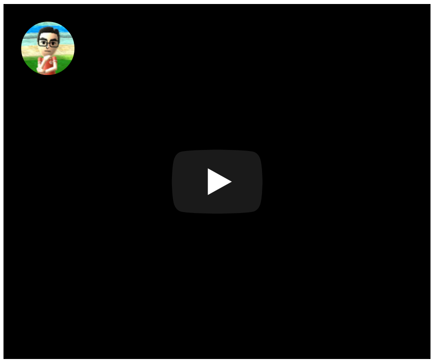

You Tune Video link:

https://www.youtube.com/watch?v=Qbav84lXeMc [https://www.youtube.com/watch?v=Qbav84lXeMc]

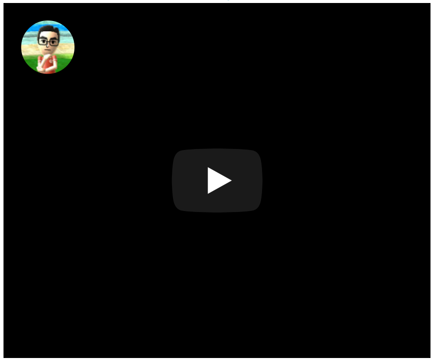

Circuit Diagram.

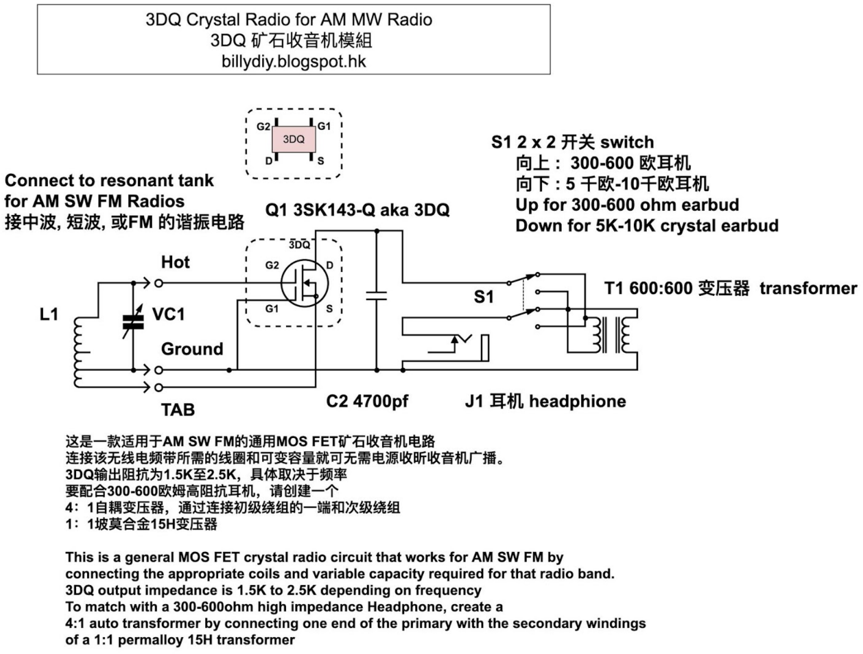

3DQ Crystal Radio for AM MW Radio
3DQ 矿石收音机模組
billydiy.blogspot.hk

G2
G1
3DQ
D
S
Connect to resonant tank
for AM SW FM Radios
接中波, 短波, 或FM 的谐振电路
Q1 3SK143-Q aka 3DQ
S1 2 x 2 开关 switch
向上： 300-600 欧耳机
向下 : 5 千欧-10千欧耳机
Up for 300-600 ohm earbud
Down for 5K-10K crystal earbud
Hot
L1
VC1
Ground
TAB
S1
T1 600:600 变压器 transformer
C2 4700pf
J1 耳机 headphione

这是一款适用于AM SW FM的通用MOS FET矿石收音机电路
连接该无线电频带所需的线圈和可变容量就可无需电源收听收音机广播。
3DQ输出阻抗为1.5K至2.5K，具体取决于频率
要配合300-600欧姆高阻抗耳机，请创建一个
4：1自耦变压器，通过连接初级绕组的一端和次级绕组
1：1坡莫合金15H变压器

This is a general MOS FET crystal radio circuit that works for AM SW FM by
connecting the appropriate coils and variable capacity required for that radio band.
3DQ output impedance is 1.5K to 2.5K depending on frequency
To match with a 300-600ohm high impedance Headphone, create a
4:1 auto transformer by connecting one end of the primary with the secondary windings
of a 1:1 permalloy 15H transformer

[https://1.bp.blogspot.com/-YVszuEPEcXw/XLlgjC6Z85I/AAAAAAAADGc/uyxvf_EfxOQATTcm2tO0NcEOOhp6h-TugCLcBGAs/s1600/3DQ%2BFM%2BCrystal%2BRadio%2BTuner%2BModule.jpg]

Posted 18th April by Billy

0 Add a comment

Dynamic Views theme. Powered by Blogger.

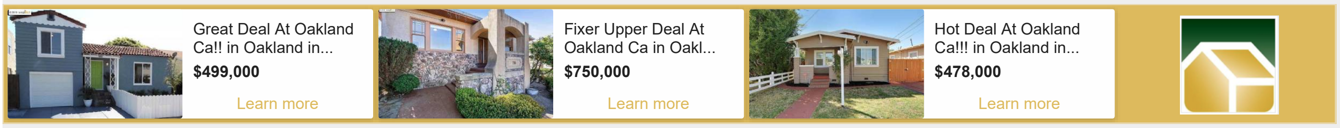

Great Deal At Oakland Ca!! in Oakland in...
$499,000
Learn more
Fixer Upper Deal At Oakland Ca in Oakl...
$750,000
Learn more
Hot Deal At Oakland Ca!!! in Oakland in...
$478,000
Learn more

13th March 1 meter loop AM crystal radio step by step construction guide

1 meter loop AM crystal radio step by step construction guide

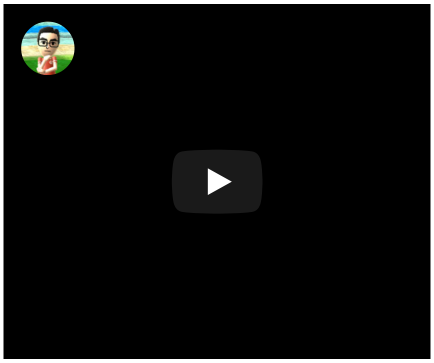

Steps to Design an AM crystal radio

1. The band & frequency range 2. Choose the tuning method3. Choose the Variable Cap.4. Calculate the Inductance 5. Make the coil6. Connect the circuit and Test

Checkout the above video for the instructions and testing.
And comparison of different tuning method, Variable Inductors and Variable Capacitors.

https://youtu.be/161vWissKLU

Posted 13th March by Billy

0 Add a comment

3rd March What do you need to know to build a crystal radio

Thanks for your interests in Crystal Radios.

Many people asked me how to build one. Many people tried and failed to receive any signals. I hope the following will give you the tips to avoid many of the challenges and let you enjoy the charm of having a functioning crystal radio. A radio that functions without batteries.

Here are the basic preparation work to build your first Crystal Radios or Power Free Diode Radio or Power Free MOS FET radio or Power Free LED Radio.

A Crystal radio i same as a regular radio you can purchase from the electronic shops, minus any form of electrical implications. Some Crystal radio uses physical audio amplification by using a horn, but no high frequency nor audio frequency amplifications.

A Crystal radio requires the following:
Antenna
Earth
Coil
Capacitor (s)
Detector (Crystal, Diode, zero voltage MOS FET, LED).
headphone or earbud

Transformer for impedance matching (optional)

But No batteries

1. Antenna and Earth.
Unlike a conventional radio that has bultin amplifications powered by batteries, crystal radios just turn whatever radio wave it receives from the antenna and coil into electrical signals without signal amplification and without using any batteries or other forms of electricity supply.

You need to be very close to the transmission station to be able to do this. Usually within 30KM.
For my case, my apartment is within 10 KM from two different radio transmission stations.
While some people can receive stations from 50KM or 100KM or 1000KM away, they have very long antennas (talking about miles) and earth system which is impossible to achieve in a city like HK.

The living environment in most Asian cities populated with concrete jungles is not too suitable for receiving crystal radio directly from your home.
Most city dwellers live in high rise apartments with not much open view between your apartment and the radio transmission stations.
If your apartment is blocked by another high rise apartment in the direction of the transmission stations, you will not be able to receive any stations through your crystal radio.
You can see this link for the AM and FM radio transmission stations in the world. This one is in particular about HK.

Link:
http://worldradiomap.com/hk/hong-kong [http://worldradiomap.com/hk/hong-kong]

http://www.asiawaves.net/hong-kong-radio.htm

For AM crystal radios , you will need a very long antenna and a good earth connection.
Depending on how close your apartment is from the target transmission stations, the antenna need to be at least 5M, or even 10M or more). If your apartment is facing the same direction of the radio transmission stations, you can hang a cable down from your window as an antenna. However that will alienate your neighbours as they may think you are hanging IP cameras to spy on them. It may also get yourself into trouble with the building management staff, or even the law enforcement units in your country.

If you are real close to the transmission station and even have a clear line of sight view of the transmitter (e.g. if you travel close to the transmission station), you can use any thread of 1 M or longer or a telescopic radio antenna to receive the AM radio.

For FM radio reception, you must be very close to the FM transmission station, e.g. < 5KM. As FM signals reduces much faster in distance than AM.
A good passive FM Yagi antenna with the main beam (called the radiator) measured 1.7M is required for good reception from home.
Or an active FM radio antenna with HF signal amplification is required. However this defeat the purpose and the charm of having a crystal radio in the first place. The no. 1 charm of having a crystal radio to me is to listen to radio without batteries or any other form of electrical power.

If you are very close to the FM transmission stations, e.g. for my case, around 3KM from the FM transmission station, you can get quite good a reception using a 1.2 meter to 1.8 meter telescopic radio antenna. FM radio signal reception is an art. Standing in different places just a few feet apart, or turning your antenna just a few degrees away, or holding your headphone cable horizontally versus vertically will effect the signal quality or even tune you to a different station.

Earth ((a.k.a. Ground),) is not required by FM radio, However, some tests show that for FM radio, if you tie one end of your coil to a thread that is lowered to touch the ground, the reception can be improved.

Earth will enhance reception for AM and SW radios. However, if the signal from your antenna is already strong enough, you may not need an earth.
For weaker reception, earth will definitely help.

If your apartment is still using metal water pipe, congratulations. You can simply connect your ground to that water pipe, just connect a few turns of bear wire (without the rubber skin) onto the water pipe. Unfortunately all your water pipes are made of plastic (e.g. many new buildings in mainland China as they build buildings with much cheaper materials), then you may be able to try some gas pipe, or electrical each (be careful of electric shock in case you incorrectly connect to the line that carry electricity instead of the earth/ground. In a worst case, you can try pulling a second antenna and use that as earth. Sometimes that may improve the reception.

The other alternative is to build a Loop crystal radio (0.5M to 1M in diameter, e.g. the size of an umbrella) , where the coil itself can replace the function of the antenna if you are close enough to the transmission station (< 10KM) and have a direct view to the transmission station that is not blocked by another high rise building. However some people are able to pick up radio stations through the reflected radio waves bouncing off the walls of the buildings eventhough they are blocked.
The draw back is you'll need more money to buy the long wires in order to build one. E.g. for a diameter of 1M, you need to have at least 40M of wires to have 12 turns for the coil, roughly HKD 100,

2. Head phone.
You will need a very sensitive head phone. The earbuds and headphones we used to listen to music from your smart phone or your hifi at home, will produce a very faint sound.
As these are designed for receiving a high power output, focus is on the quality not sensitivity.
You will need special head phones for crystal radio reception.
These can be purchased from taobao, or you can make your own using a ceramic piezo-electric buzzer (still available in the ShimshuiPo IpLiu Street shops（華輝）, the plastic tip of a ball pen,a silicon earbud cushion, glue, some wires and the headphone jack.
You can refer to my video on how to make one.

Link: https://www.youtube.com/watch?v=k3HsVlxR2tA [https://www.youtube.com/watch?v=k3HsVlxR2tA]

If impedance of your headphone does not match with the of the detector, you will need a step up or step down transformer for impedance matching.
Impedance matching is very important for crystal radio.
As Crystal radio has no signal amplification, we need to maximise the efficiency of the transfer from one stage to the next. and that require impedance matching to ensure we transfer all the energy down to the next stage of the circuit. Any impedance mismatch will mean energy loss, and lowered output, softer sound.

3. Coil

For AM, radio,
You can build a small coil (0.01M to 0.1M) to go with a long antenna and earth, or a big loop coil (0.5M to 1M diameter, see above)
You can use insulated magnetic wire (available from the DIY shops
Due to skin effect (radio signals only travel at the surface of the wire instead of the core), Litz wires that have multiple insulated smaller wires wrapped within a cotton thread will perform much better, almost doubled in terms of loudness than the solid Magnet wire
Each one has up to 660 tiny wires each measures 0.04mm in diameter. These are available in the taboo shops, around HKD 150 per roll that is enough to make one big coil for umbrella loop radio.
For small coils, the one with 60 tiny wires each 0.04mm can be used to wind on a dark ferrite rod (1cm diameter), the ones you see on most conventional radios.

measure Capacitance as well.
For AM radio, inductance required is from 180uH to 510uH, depending on the size of capacitance you match with.
180uH match with 512pf capacitor
330uH match with a 365pf capacitor
510uH match with a 220pf capacitor.

4. Variable capacitor v.s. fixed capacitor

Variable capacitors is also very important for a good radio. It is the knob that you turn to tune to the frequency for your crystal radio.

Air insulated variable capacitor made of copper has the highest Q factor. It's available in taboo but hard to find and very expansive > HKD 150.
Smaller versions of these < 20pf can be used for FM radio (HKD 20).

Next is the air insulated variable capacitor made of aluminium. Around HKD 30-50 depending on the size.
In general the bigger ones with the divider wider apart has a high Q.
Q is a factor that measures how good is the efficiency (sensitivity) and accuracy (selectivity) in transferring radio wave signals to electric signals.
The conventional plastic variable capacitor we see on the conversational radios will also work, but loudness will be diminished a bit (at least 30% less loud).

Alternatively, you can use a fixed capacitor, but vary the inductance of your coil to tune to different frequencies.
You tune the inductance by different methods:
1. Move the coil in or out of the ferrite stick. Very easy to do, see my videos.
2. Vary the length and separation of the turns within the coil (much harder to achieve practically).
3. Wind two coils. Then overlap these two coils on top of one another and the total inductance of the these two coils in series = the addition or subtraction of the two coils. Addition if they are facing the same direction (e.g. clockwise), subtraction if they are different direction (e.g. anticlockwise on one coil which is laid on top of a clock wise coil). The amount.to add or subtract also depends on how much of the area of the two coils are overlapping.

5. Detector.

Detector is the component that translate high frequency radio signals to Audible signals.
This conversion is called demodulation.
The action taken by the detector is really simplify only allow electric current to flow in one direction and stop current flow in the opposing direction.
This way, only the position half of the radio signal is allowed to pass through.
The the audio signal will be able to be reproduced by the headphone.

There are different types of detector.
I list them here in term of the performance. Best performance first.
With the exception of vacuum tube, all can function without battery or mains electric power.
This the charm and magic of crystal radio, or battery free radio.

Zero voltage MOS FET
Diode
LED (only works in very strong signals (> 200uA), lits up during radio reception and glow according to the strength of the signal.
Vacuum Tube (that need to be battery powered to lit up the heating coil, but no power required for detection).
Crystal

The name "Crystal Radio" v.s "FET radio" v.s. "Diode Radio" v.s "power free radio"

There is constant debate over the name given to all kinds of radios that can operate without battery or other kinds of man made electrical power (solar cell, acid cell from fruit, hand-held push/pull power generator or mains power supply).

With newer form of detectors that we discover in recent years, besides genuine crystals with cat whiskers, zero voltage MOS FET e.g. 3DQ (3SK-143-Q), Diode (e.g. 1N34A), or even LED are used as detectors. Should we or should we not keep calling them Crystal Radios ?

My vote is to still call these Crystal Radios.
Simply as it simplifies a mouthful of words such as power free zero voltage MOS FET radio.
And Crystal radios reminds every one of the Charm of listening to radios without power.

Posted 3rd March by Billy

0 Add a comment

25th April 2018

New Growthbed

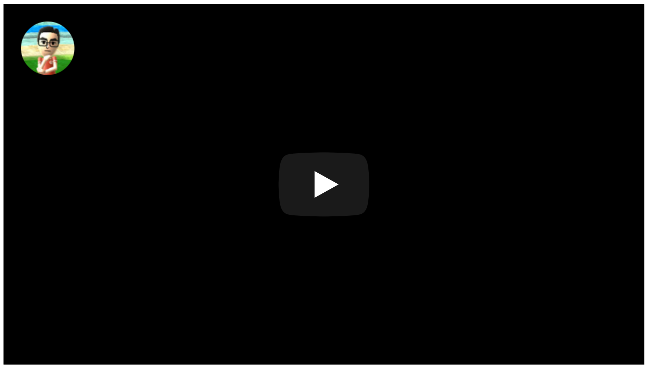

Posted 25th April 2018 by Billy

0 Add a comment

Great Deal At Oakland Ca!! in Oakland...
$499,000

Fixer Upper Deal At Oakland Ca in...
$750,000

Hot Deal At Oakland Ca!!! in Oal
$478,000

M

2nd March 2018

Self-cleaning CHOP aquaponics systems using IKEA transparent plastic storage boxes.

Self-cleaning CHOP aquaponics systems using IKEA transparent plastic storage boxes.

Here is the link to my youtube video.

https://www.youtube.com/watch?v=YC5RGUUXxJE&t=24s [https://www.youtube.com/watch?v=YC5RGUUXxJE&t=24s]

Dynamic Views theme. Powered by Blogger.

Gretting Card Crystal Radio
贺卡矿机

[https://1.bp.blogspot.com/-wyQH-V_HCkE/VySGA-f3s4I/AAAAAAAACIo/bEIRnV_NMv4C6T3rj1IQY72w9CLX4HgfACLcB/s1600/Screen%2BShot%2B2016-04-30%2Bat%2B4.57.29%2BPM.png]

Posted 29th April 2016 by Billy

3 View comments

11th April 2016

Dual pole sliding band switch

I made a Dual pole sliding band switch using 1mm copper sheets. The switch is for soldering directly onto the taps of a ferrite stick coil wound using 660x0.04 Litz wire.

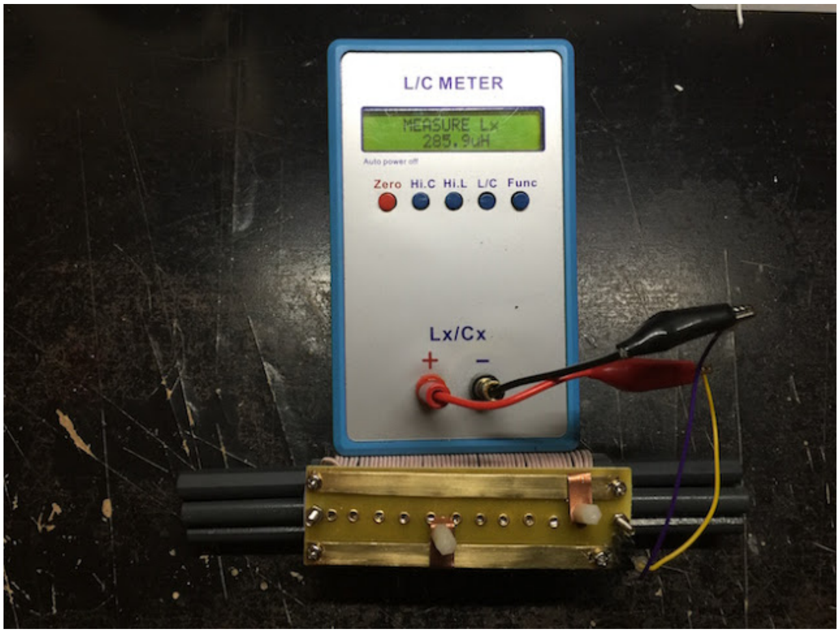
L/C METER
MEASURE Lx
285.9uH
Auto power off
Zero
Hi.C
Hi.L
L/C
Func
Lx/Cx
+
−

[https://2.bp.blogspot.com/-Ifa2K3uccUU/Vwz96BXfeII/AAAAAAAACiU/eN-s-tWpU_0F9AgQZkN_vEzVVWQmBwapACLcB/s1600/IMG_7619.JPG]

Dynamic Views theme. Powered by Blogger.

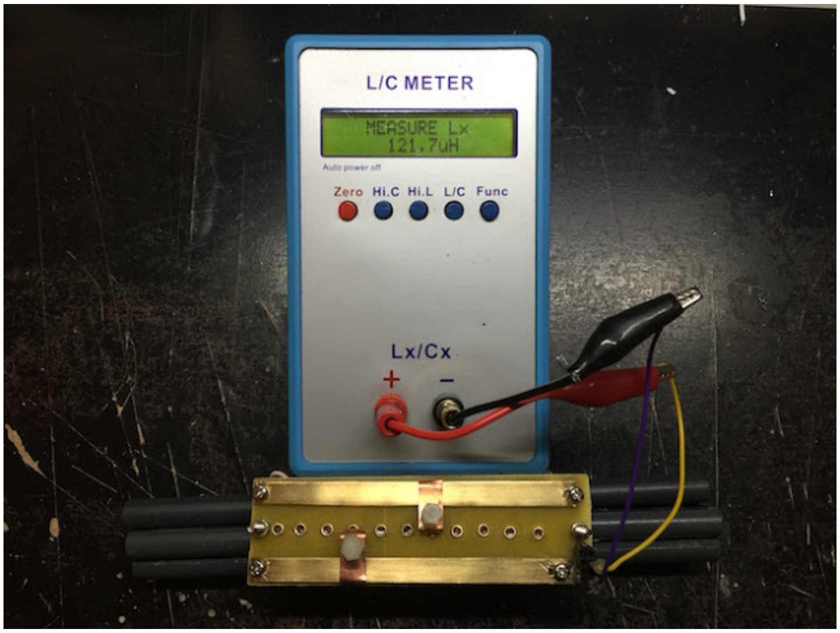
L/C METER
MEASURE Lx
121.7uH
Auto power off
Zero
Hi.C
Hi.L
L/C
Func
Lx/Cx
+
−

[https://1.bp.blogspot.com/-XA2sKaDvwww/Vwz96IRX16I/AAAAAAAACic/axvDn8pKHLsOcMInZ-Es8gFqbkmazA3ZQCLcB/s1600/IMG_7620.JPG]

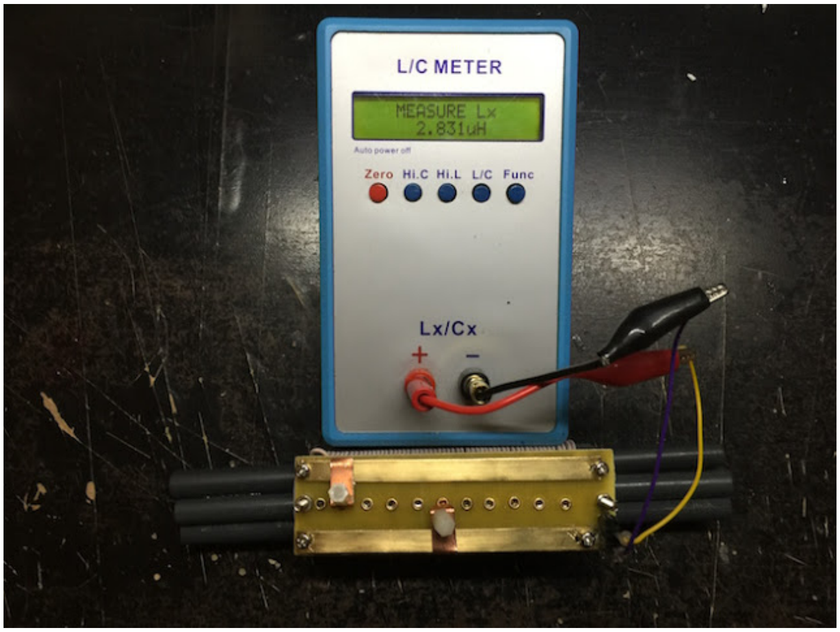
L/C METER
MEASURE Lx
2.831uH
Auto power off
Zero
Hi.C
Hi.L
L/C
Func
Lx/Cx
+
−

[https://3.bp.blogspot.com/-ww47kz6_CD0/Vwz96LdjJKI/AAAAAAAACiY/1SwUnIckh9g4lZgwADXjO_W0DYE0h8YVQCLcB/s1600/IMG_7621.JPG]

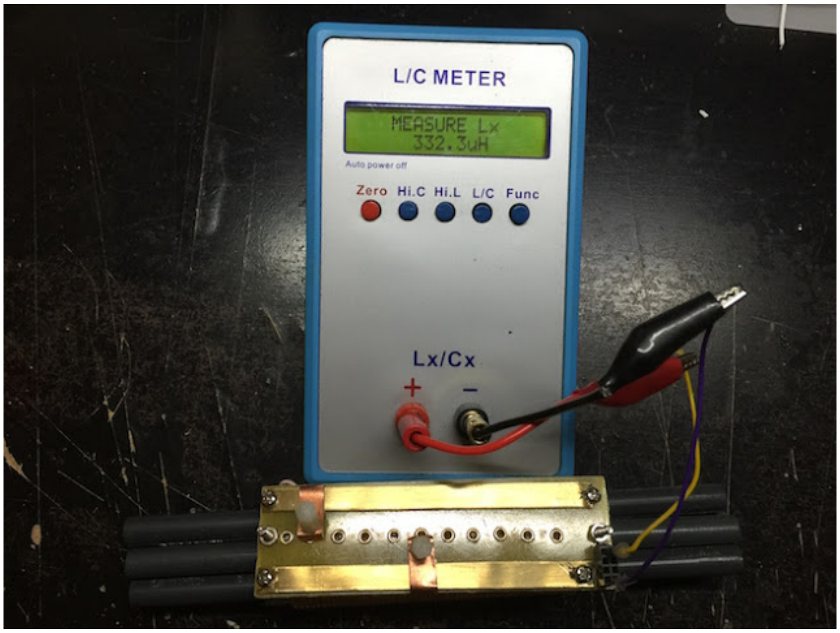
L/C METER
MEASURE Lx
332.3uH
Auto power off
Zero
Hi.C
Hi.L
L/C
Func
Lx/Cx
+
−

[https://4.bp.blogspot.com/-EQ8c05hVWvw/Vwz96r_c4WI/AAAAAAAACig/tGf5ywG9J-wcefFN2nopUG1Rr9vTXcaigCLcB/s1600/IMG_7622.JPG]

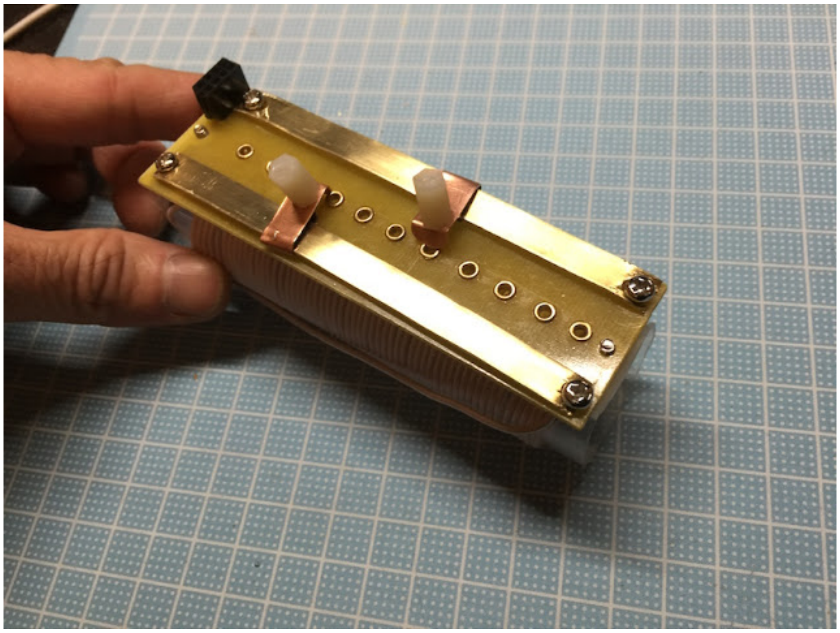

[https://1.bp.blogspot.com/-0H-hsyh4npI/Vwux9geVOXI/AAAAAAAACh8/PGh9vBol598EJQOXgRg_kbCz_8i4ojg9w/s1600/IMG_7615.JPG]

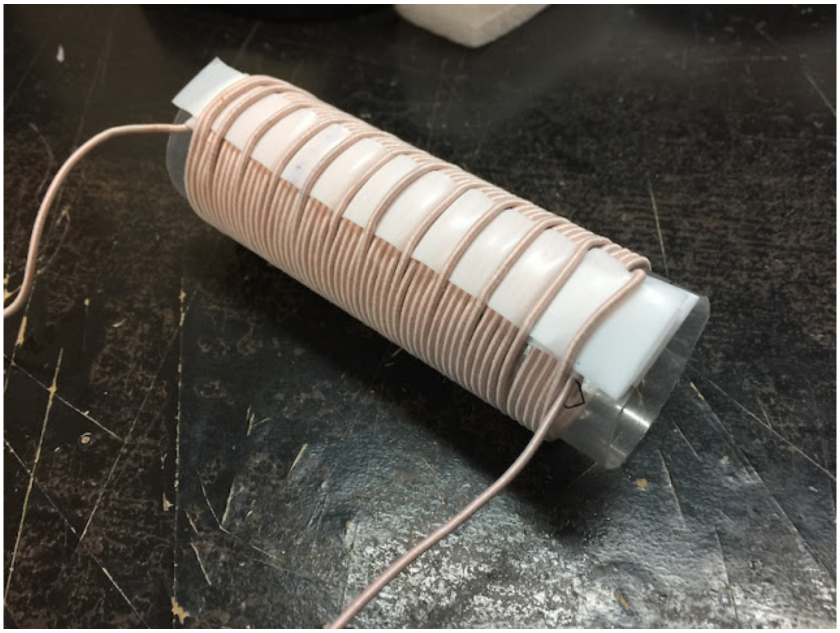

[https://2.bp.blogspot.com/-3pI_f5KHwZo/Vwux44jTjYI/AAAAAAAACgw/pkdws7eVKAQP5EyCU044NDNFoeO0wlNWQ/s1600/IMG_7593.JPG]

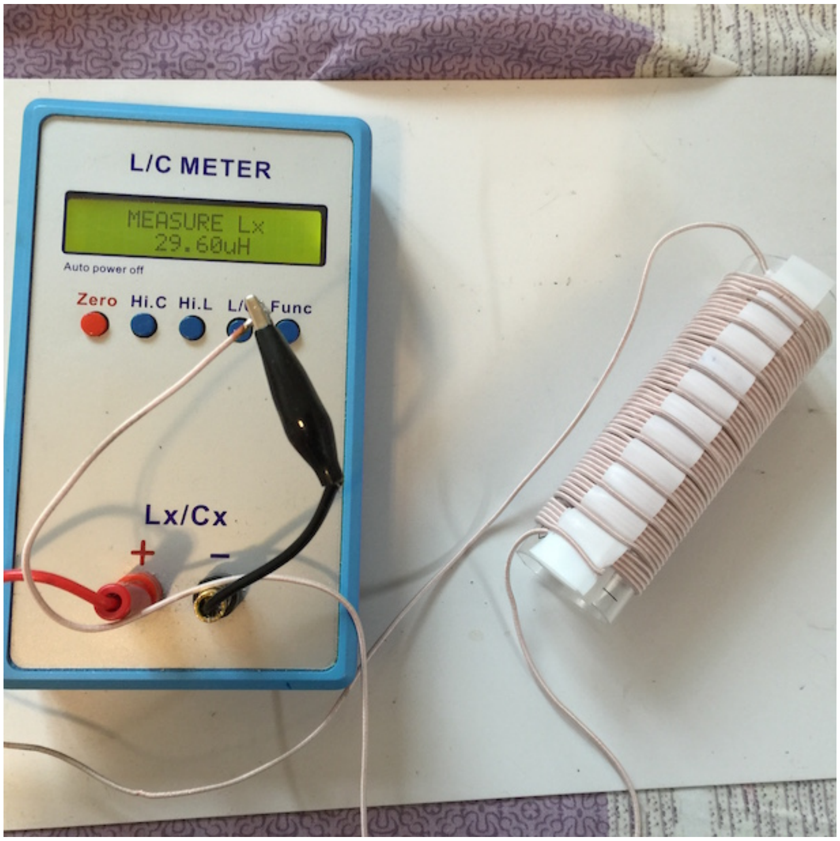
L/C METER
MEASURE Lx
29.60uH
Auto power off
Zero
Hi.C
Hi.L
L/C
Func
Lx/Cx
+
–

[https://4.bp.blogspot.com/-wAYeSgKLts8/Vwux4-FtnUI/AAAAAAAACg4/JZ79CtCCaZE4zMs3rXiC6qAtVrMC9TWZw/s1600/IMG_7594.JPG]

Inductance measured using 5 regular MW radio ferrite sticks

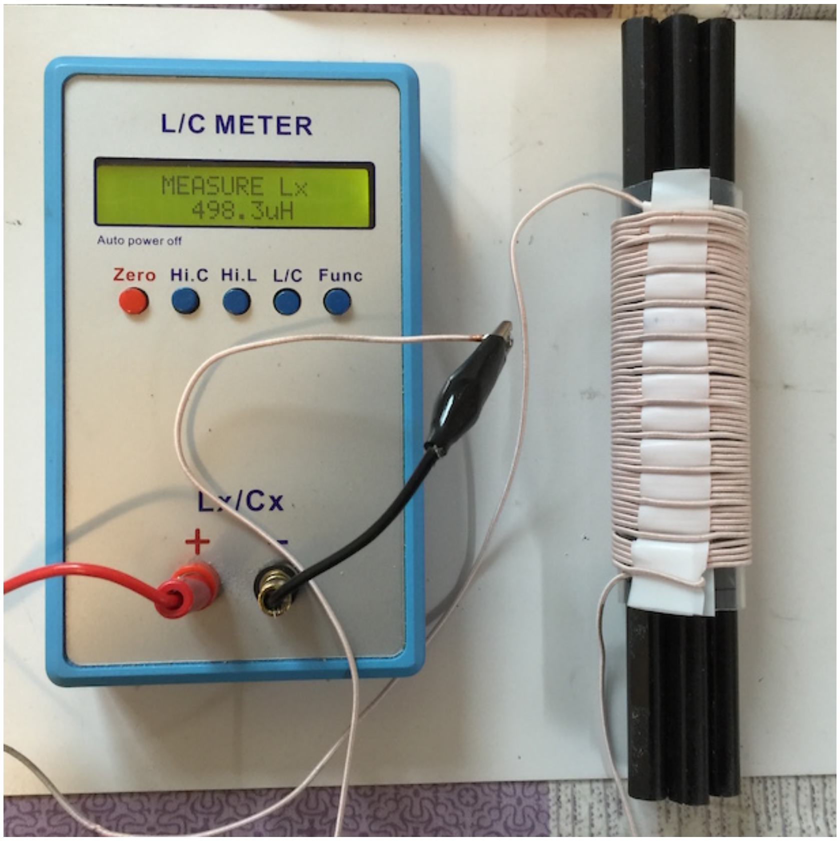
L/C METER
MEASURE Lx
498.3uH
Auto power off
Zero Hi.C Hi.L L/C Func
Lx/Cx
+
-

[https://2.bp.blogspot.com/-gNHgOfVKdcU/Vwux5NgmoAI/AAAAAAAACg0/N8pWsnNFIQA8sdbp7Iel0_dCra49L7PBA/s1600/IMG_7595.JPG]

Inductance measured 5 SW radio ferrite sticks

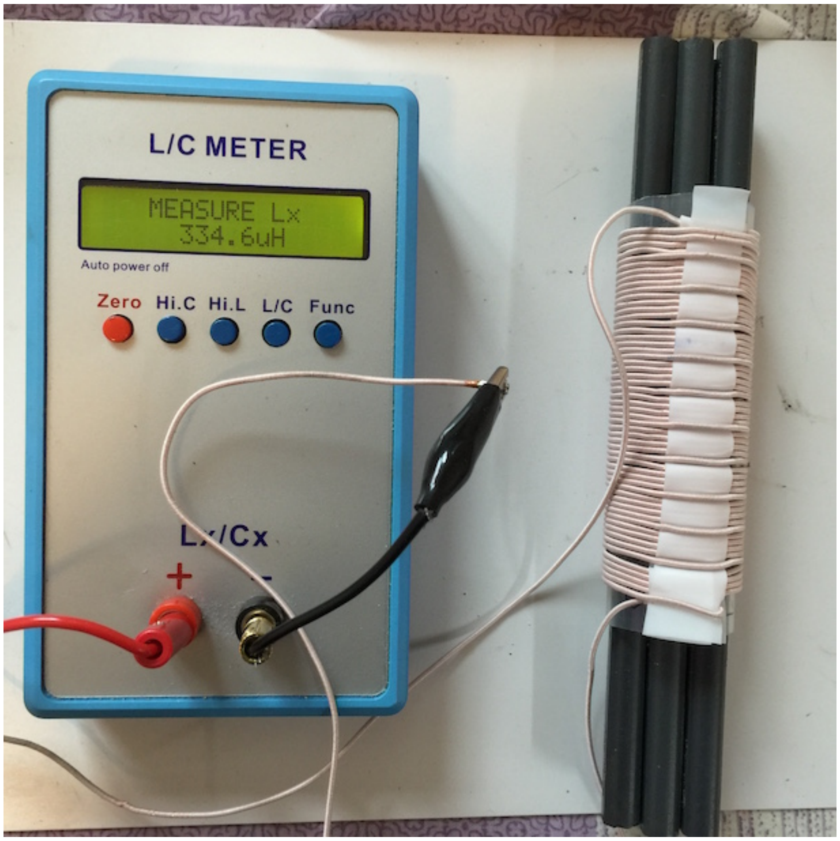
L/C METER
MEASURE Lx
334.6uH
Auto power off
Zero Hi.C Hi.L L/C Func
Lx/Cx
+
-

[https://3.bp.blogspot.com/-_zTI_Qzx01A/Vwux5z5aOZI/AAAAAAAACg8/oXnwgXscB6gLEmuJbiPkbkRG1HK8_SaAg/s1600/IMG_7596.JPG]

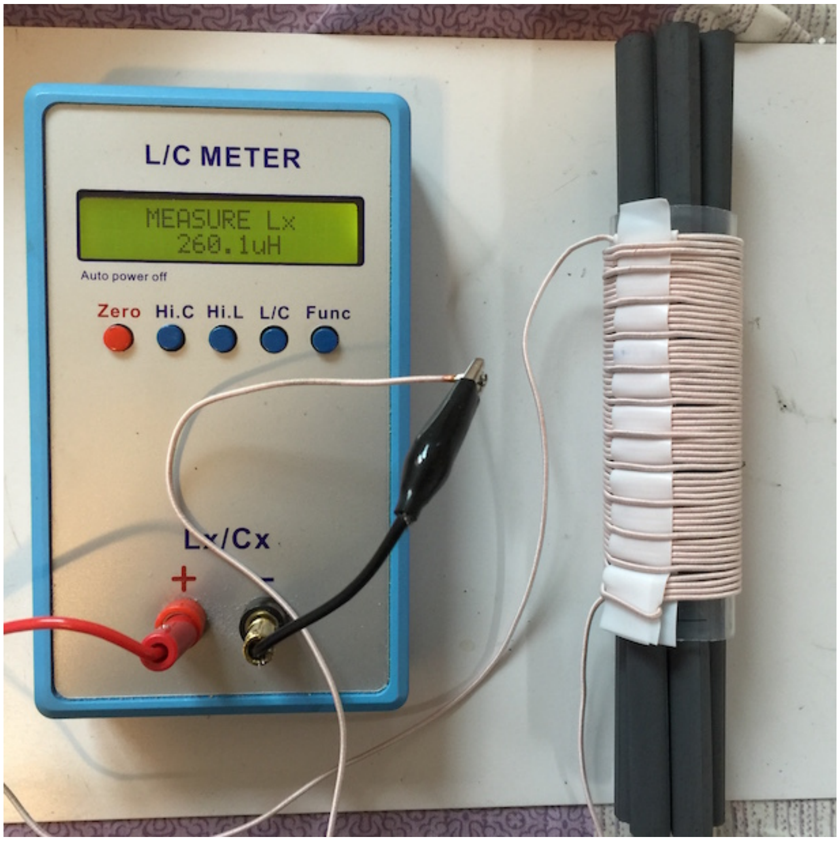
L/C METER
MEASURE Lx
260.1uH
Auto power off
Zero
Hi.C
Hi.L
L/C
Func
Lx/Cx
+
-

[https://3.bp.blogspot.com/-F00ICm6_7N4/Vwux59oKHtI/AAAAAAAAChA/mIeioM-JGFITVW8vYH6v1SrFtqsEGIfGw/s1600/IMG_7597.JPG]

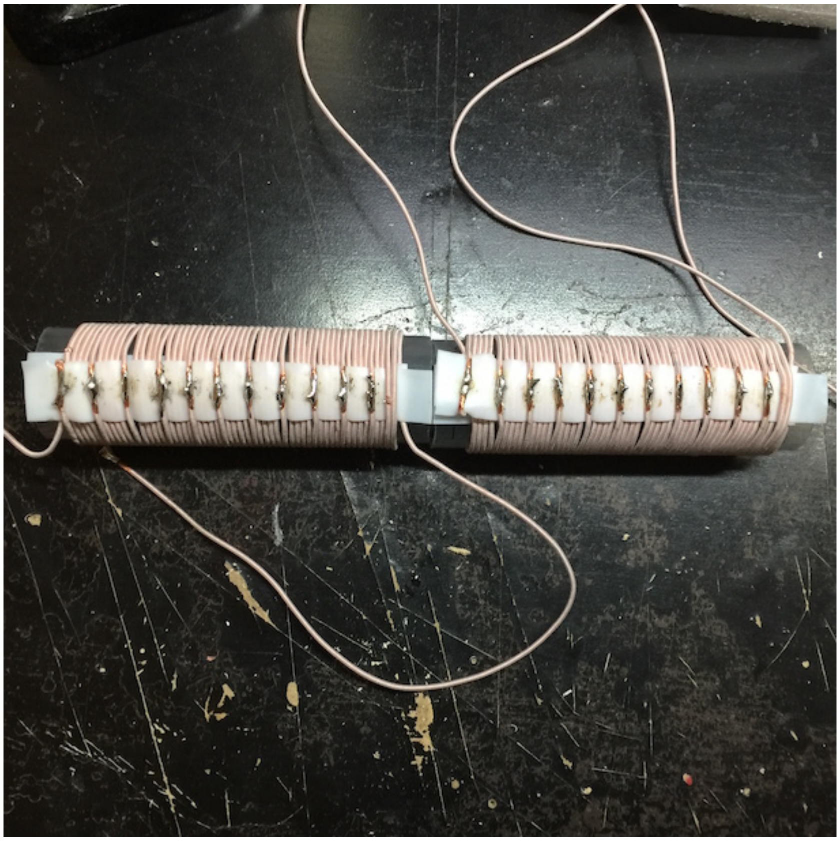

[https://2.bp.blogspot.com/-b4-JdZGceno/Vwux6MUykfI/AAAAAAAAChE/acYqh1MfnWw0rxPa9po-AgAnYwhiSKDig/s1600/IMG_7598.JPG]

Dynamic Views theme. Powered by Blogger.

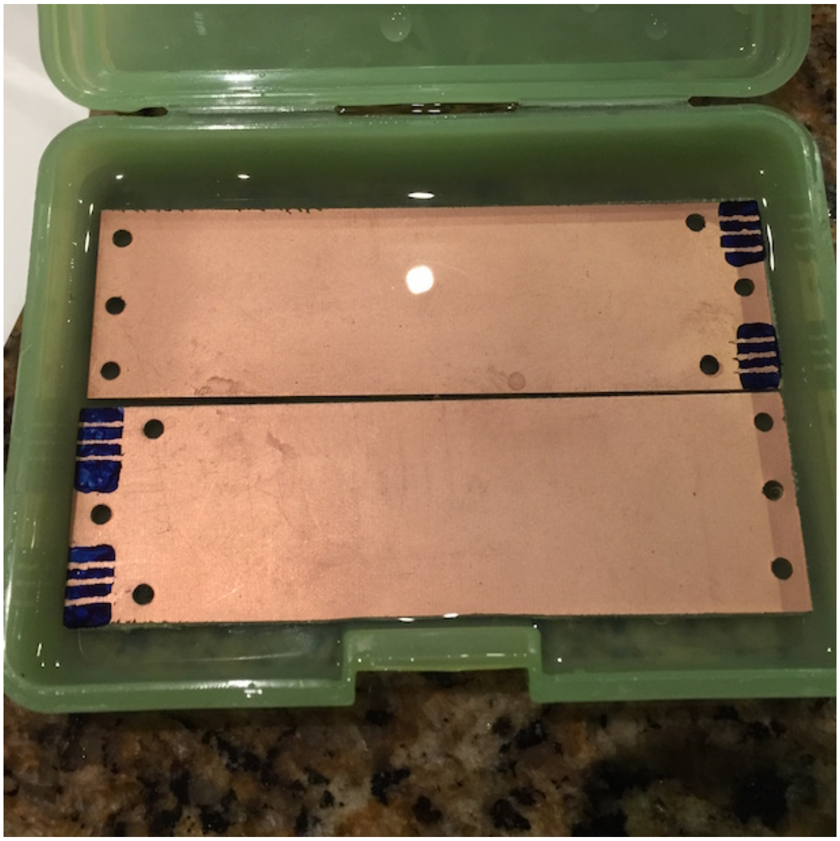

[https://3.bp.blogspot.com/-FF3YknlPqTg/Vwux6W7lZvI/AAAAAAAAChI/cxzUJuqqshwTYx0YJ1PjNcV7p07gygE6g/s1600/IMG_7599.JPG]

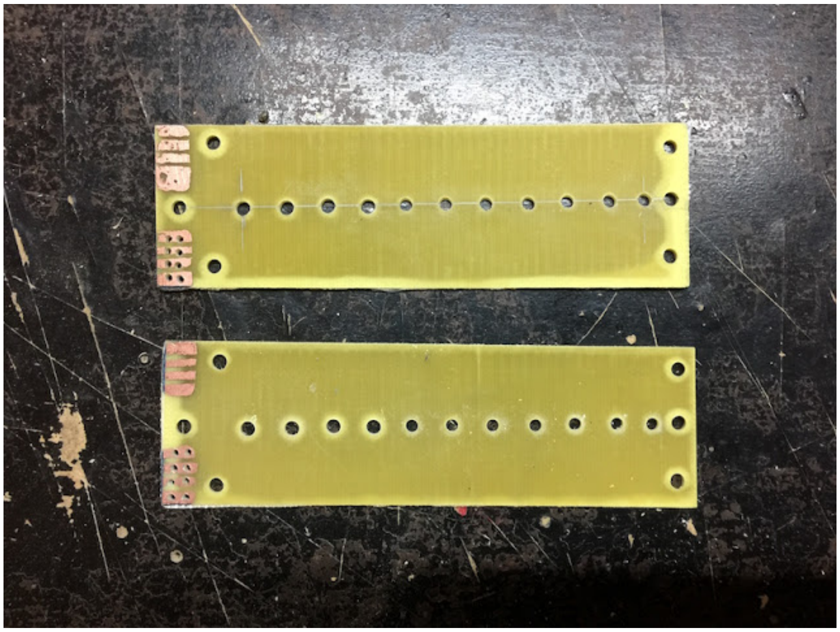

[https://1.bp.blogspot.com/-oKNAAILvv4k/Vwux6hDGsqI/AAAAAAAAChQ/hVLdtQNlv8kFfbgf32eUxSDMGyb3HJg0w/s1600/IMG_7602.JPG]

[https://4.bp.blogspot.com/-IlbduASQeXU/Vwux69ahX7I/AAAAAAAAChM/UjSvDT2TAIUrUaxYr038l4IciWq39fO7Q/s1600/IMG_7603.JPG]

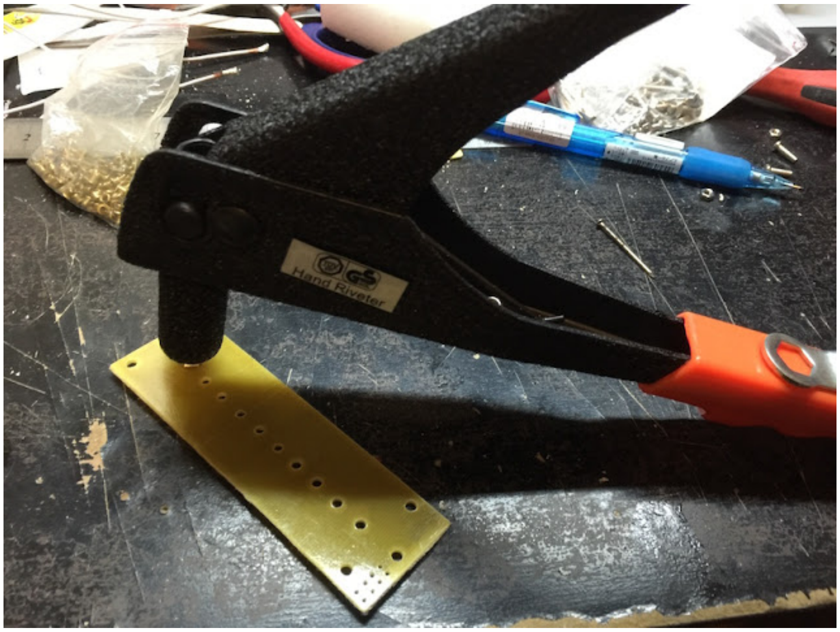

[https://3.bp.blogspot.com/-l6fcKpC31nQ/Vwux7O52x4I/AAAAAAAAChU/oZ-VHFUVNmUrGG9mbgzgKh1C7q9dHqnNQ/s1600/IMG_7604.JPG]

Dynamic Views theme. Powered by Blogger.

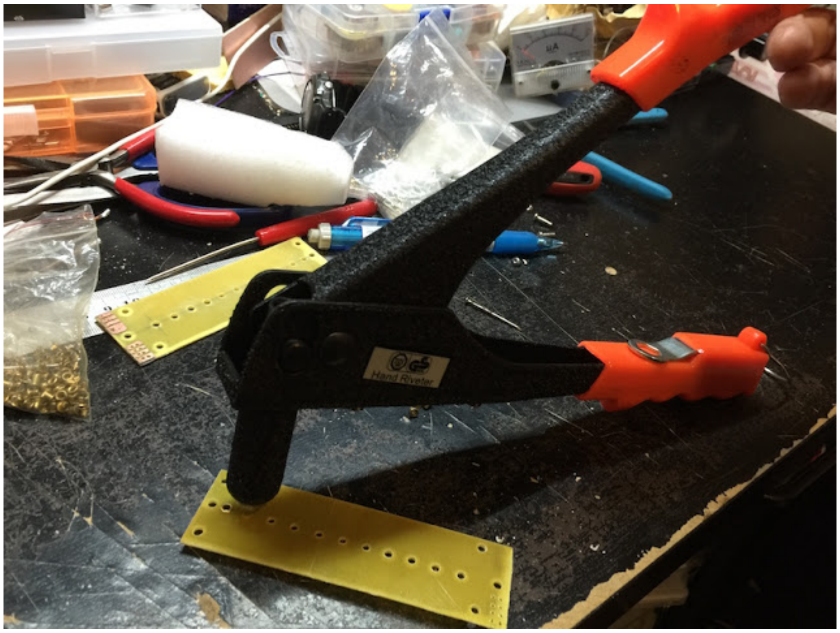

[https://2.bp.blogspot.com/-mc7fv0JY_EE/Vwux7dtsqWI/AAAAAAAAChY/UvYV8TBZ-q4CTFf14-MtM2jAb7Sc7KpMA/s1600/IMG_7605.JPG]

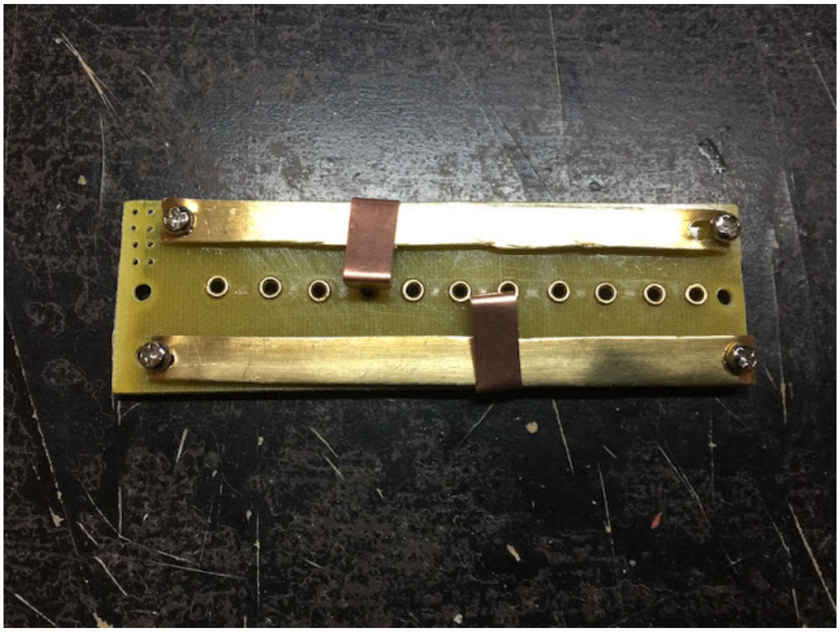

[https://1.bp.blogspot.com/-fXGoG17DSxg/Vwux7qcFAHI/AAAAAAAAChc/TyjQd1AOUDMao0cm_cERkddapxate2sYA/s1600/IMG_7606.JPG]

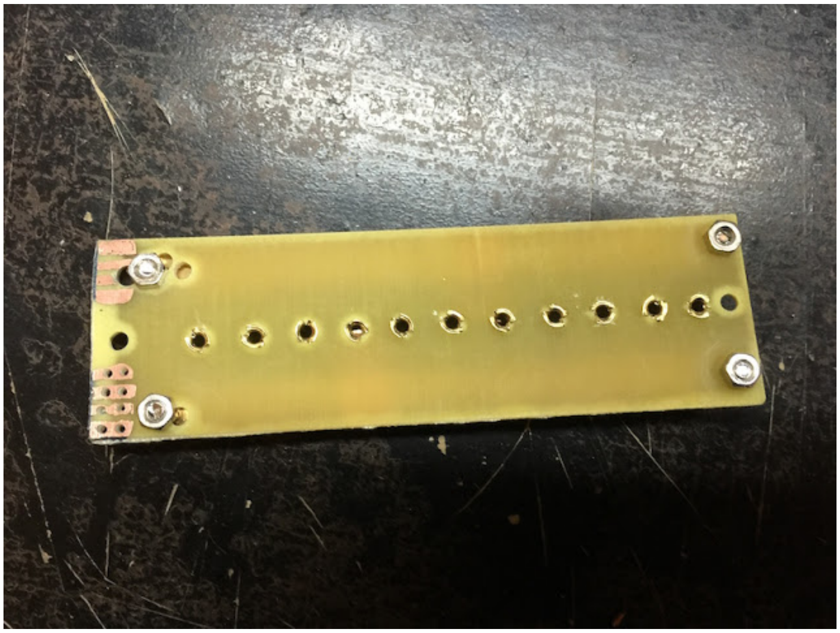

[https://4.bp.blogspot.com/-o7JrBo4KD8A/Vwux7_T53yI/AAAAAAAAChg/PZDk9y2ZN1gg_IDZqOZcxTdyzb9uGLjTw/s1600/IMG_7607.JPG]

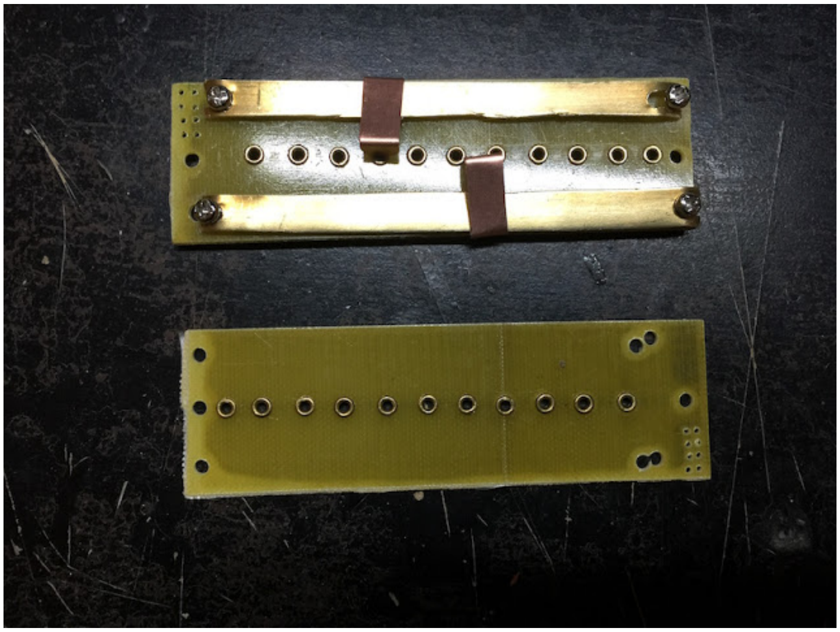

[https://3.bp.blogspot.com/-MVk55EdmXQ4/Vwux8AM_JWI/AAAAAAAAChk/xaLTGh8lXWE-ZgnyGytO6K_TPdER7QN5w/s1600/IMG_7608.JPG]

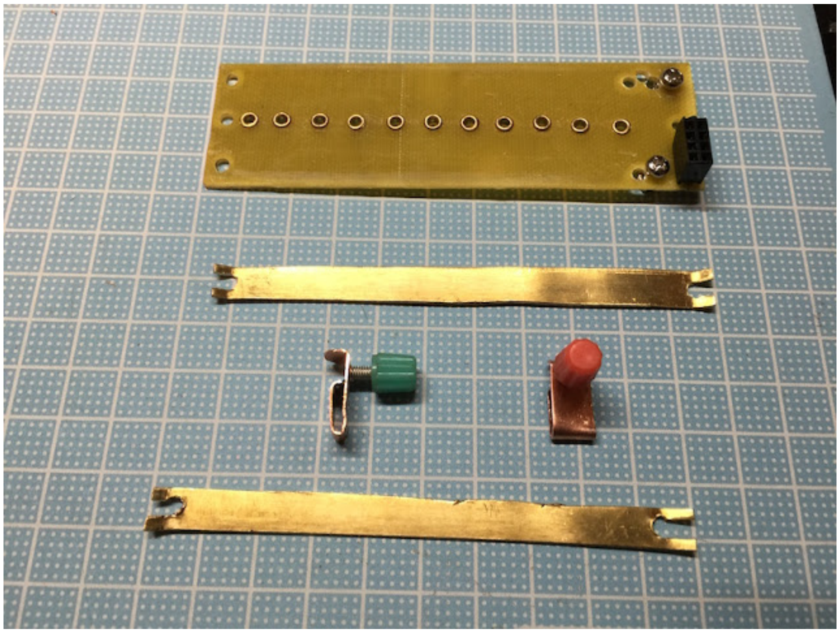

[https://1.bp.blogspot.com/-nM_fYRdI1n8/Vwux8ZticLI/AAAAAAAACho/kdqnNpesWxcc0ytdgiNclz1trWqJC8WjA/s1600/IMG_7609.JPG]

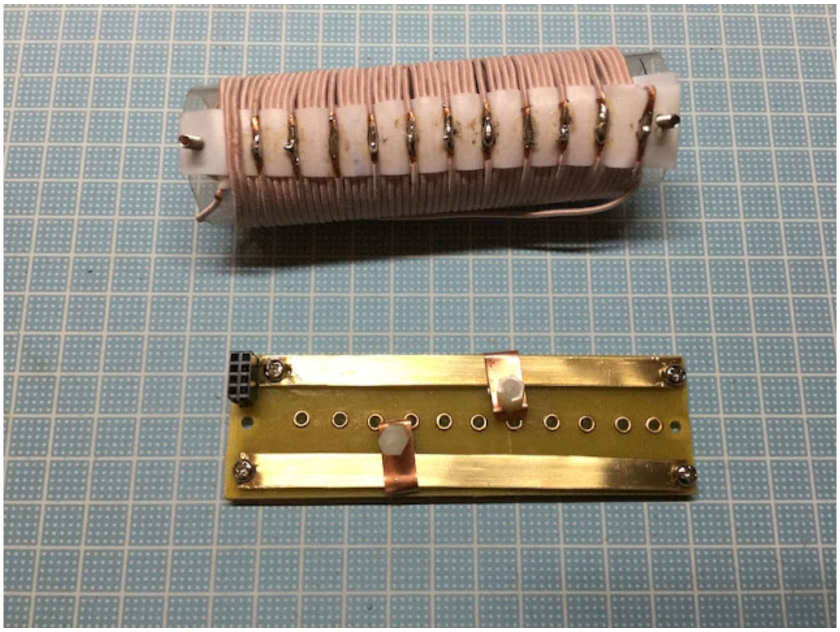

[https://3.bp.blogspot.com/-U_FN2ASFQKg/Vwux8qzIDAI/AAAAAAAAChs/EnHwQ5gftpkeGIXoBJUcwaDE4rRsJkjEA/s1600/IMG_7610.JPG]

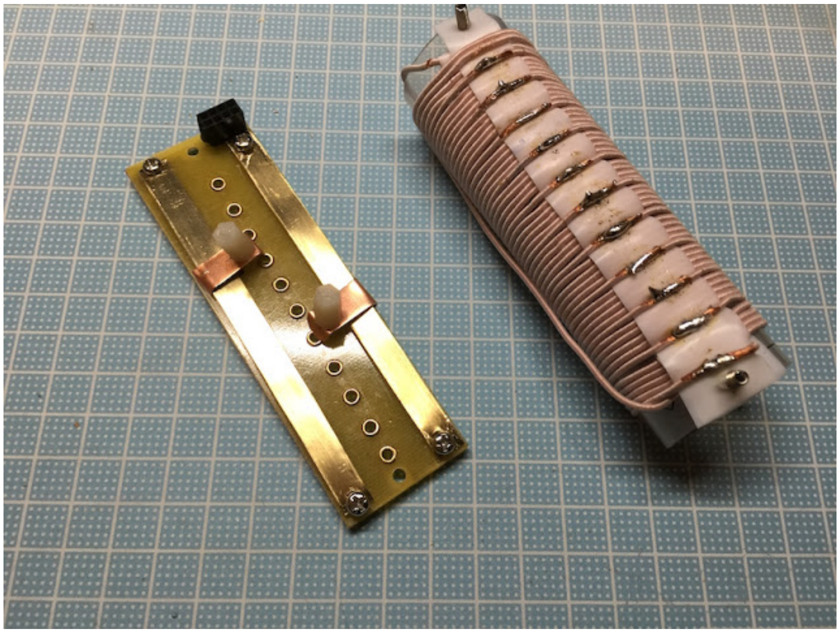

[https://1.bp.blogspot.com/-YT03b_ptiRA/Vwux85UDb9I/AAAAAAAAChw/lFIZayb3JUcR0xZ2tEf33JYPZn5loH-Vg/s1600/IMG_7611.JPG]

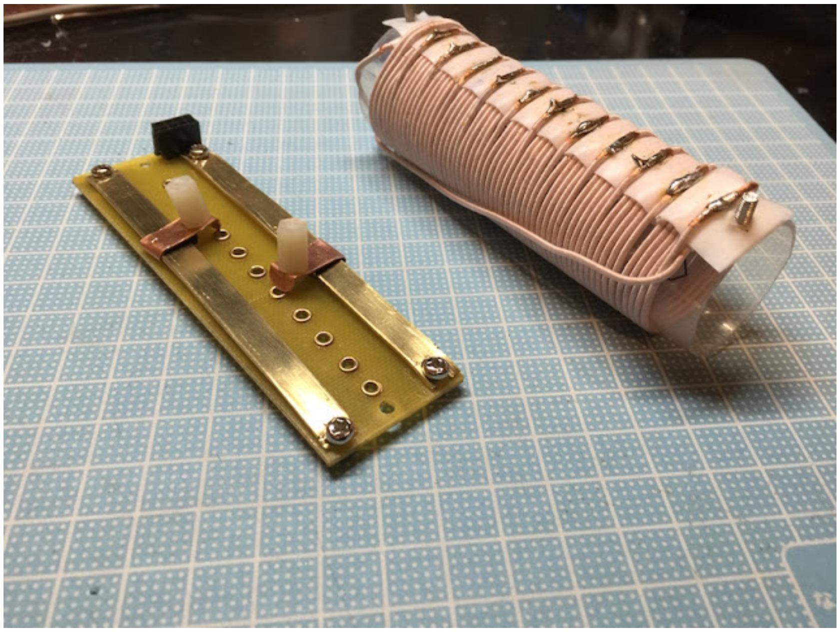

[https://1.bp.blogspot.com/-NSEUo8IzBIg/Vwux9OOj0sI/AAAAAAAACh0/A1bxuSIDKGstk6MYOstXYueYeTGBqSkLA/s1600/IMG_7612.JPG]

Dynamic Views theme. Powered by Blogger.

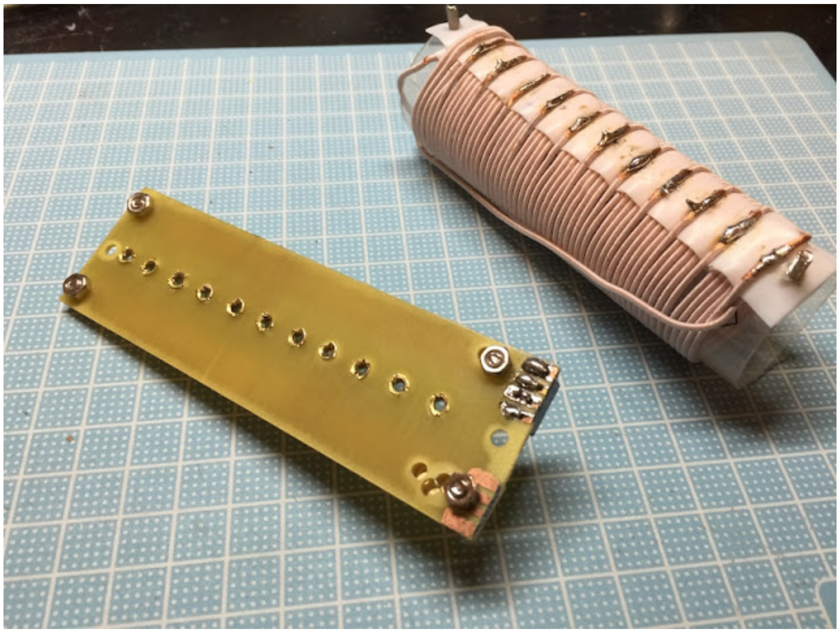

[https://1.bp.blogspot.com/-qFOAYO7UeC0/Vwux9WnkiQI/AAAAAAAACh4/9lgcKziBkEQmKSXyt4pUDNSa3WY872-rA/s1600/IMG_7613.JPG]

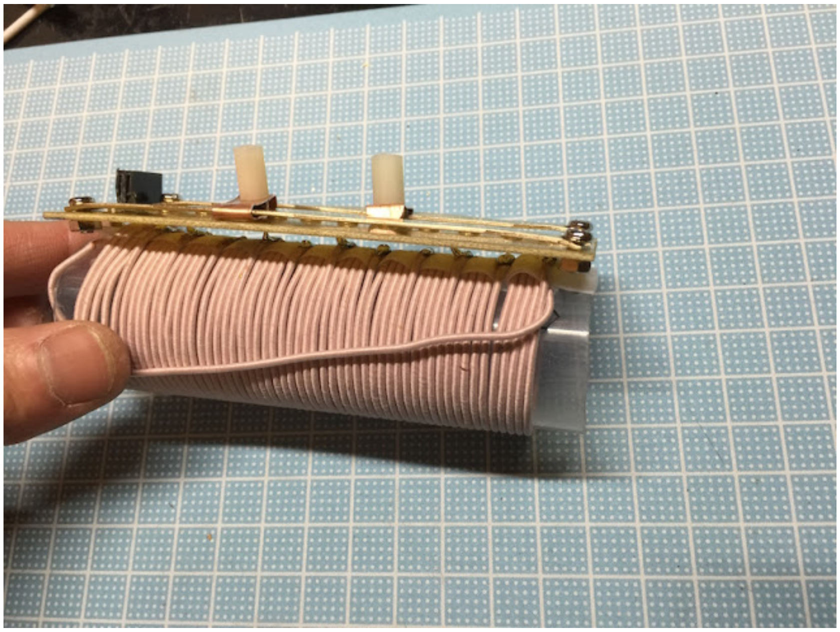

[https://1.bp.blogspot.com/-0FpUJDIzWY4/Vwux-qZcSsI/AAAAAAAACiA/6h8QjFz9LHo2X2PkYsfj_sN2puogQxEiQ/s1600/IMG_7614.JPG]

Dynamic Views theme. Powered by Blogger.

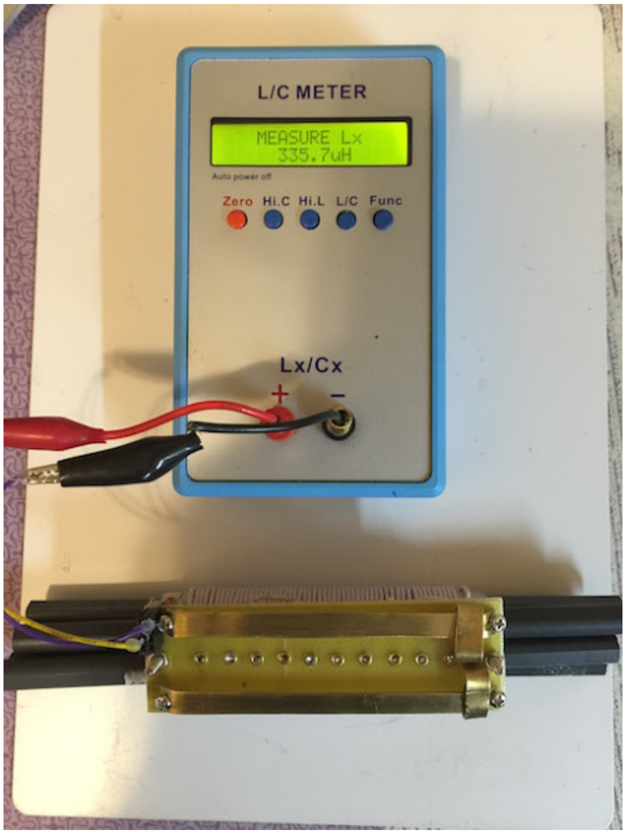
L/C METER
MEASURE Lx
335.7uH
Auto power off
Zero
Hi.C
Hi.L
L/C
Func
Lx/Cx
+
−

[https://1.bp.blogspot.com/-nDngei-QAO4/Vw--6z7ozXI/AAAAAAAACjA/wZgsV7Bmo7wcxRsuAmQ-6mgvSrZyqx7kgCLcB/s1600/IMG_7625.JPG]

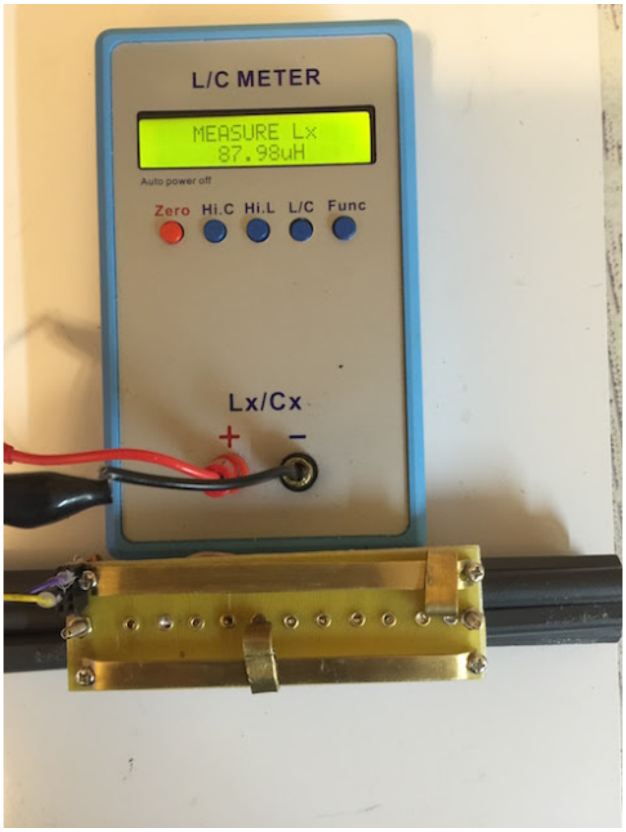
L/C METER
MEASURE Lx
87.98uH
Auto power off
Zero
Hi.C
Hi.L
L/C
Func
Lx/Cx
+
−

[https://4.bp.blogspot.com/-1sacSMBqmW8/Vw--6-7tz_I/AAAAAAAACi8/XW4L_MS5Er04hpmbLcoPgeEFvqeauqb3QCLcB/s1600/IMG_7627.JPG]

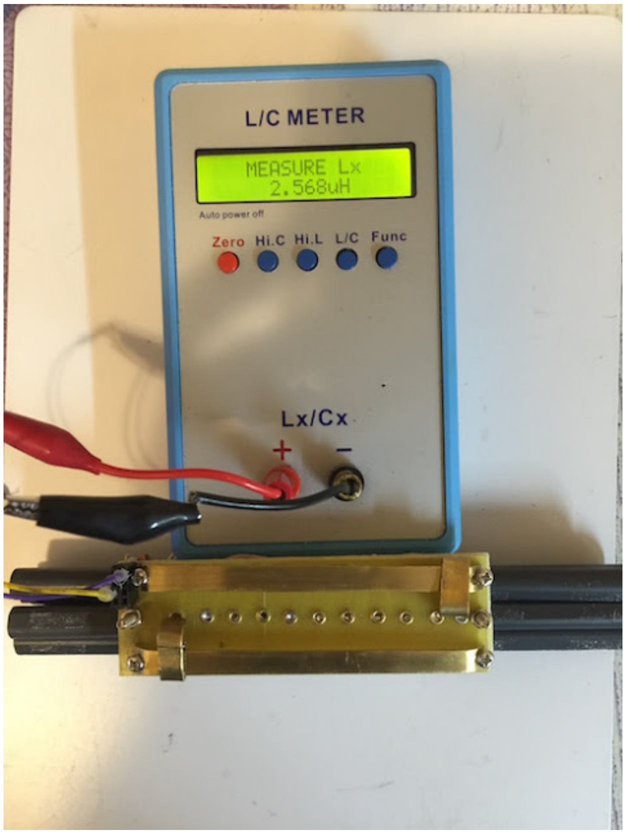
L/C METER
MEASURE Lx
2.568uH
Auto power off
Zero
Hi.C
Hi.L
L/C
Func
Lx/Cx
+
−

[https://4.bp.blogspot.com/-ToQrTWCVXmE/Vw--65Unz9I/AAAAAAAACjE/sozySU8Mcq0pu35teU36OdS71alKPxLlwCLcB/s1600/IMG_7628.JPG]

Posted 11th April 2016 by Billy

0 Add a comment

legalzoom Let's Make Your Business Official! Come back

8th April 2016

RCA Sound Powered Headphone

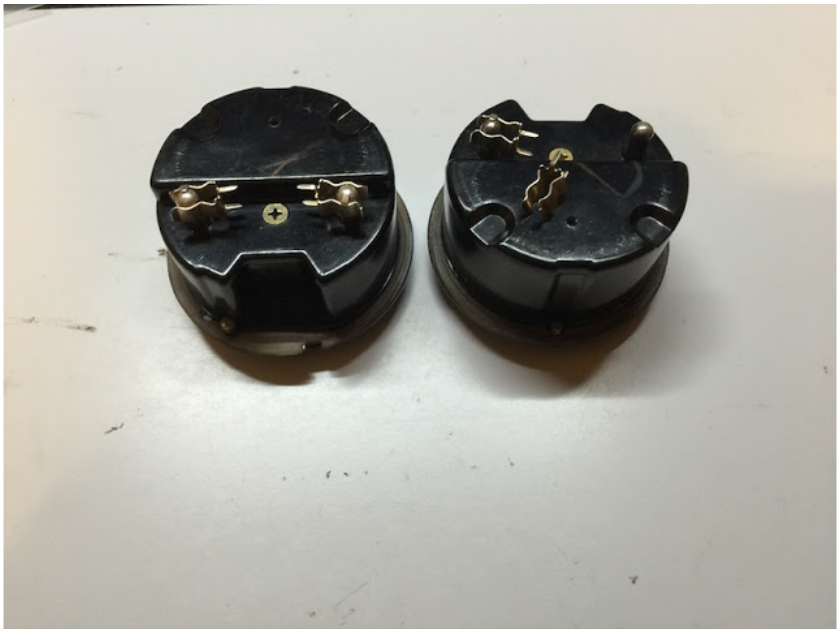

[https://2.bp.blogspot.com/-0arbQ78Fi4c/VwfGpqfsOiI/AAAAAAAACgM/Df75JSo1_UsLFQKCkN5K1Hadz_8z4Escw/s1600/IMG_7569.JPG]

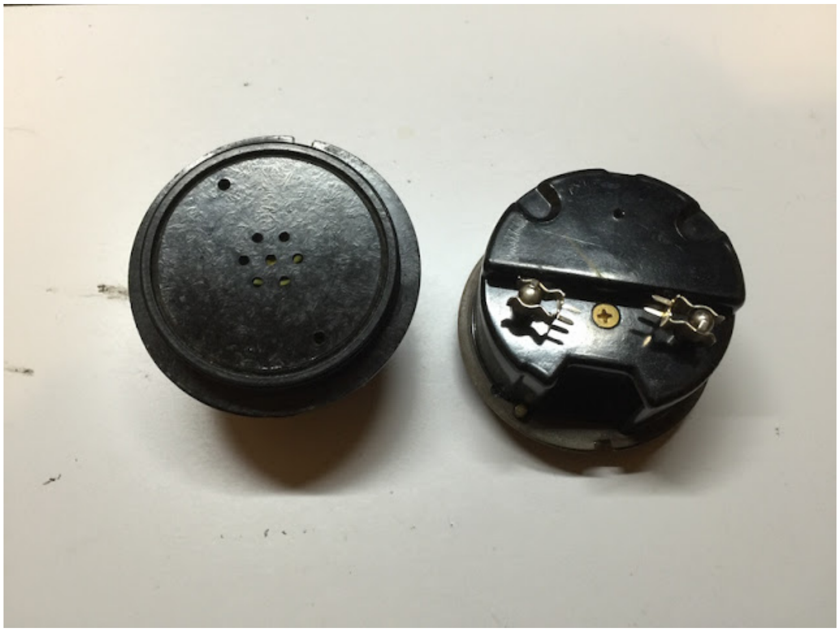

[https://3.bp.blogspot.com/-34GmmJ5oEq4/VwfGpsPactI/AAAAAAAACgE/b02WCPGDtYkSgWyY0yOC6NXpnCJJceMPg/s1600/IMG_7570.JPG]

[https://2.bp.blogspot.com/-k69I-Cj_OYM/VwfGpgDcoEI/AAAAAAAACgI/Y0X57_7XUQE4rvYx8sBDr0ZCN_KPtZ2jw/s1600/IMG_7571.JPG]

[https://4.bp.blogspot.com/-57J_DnBEQWw/VwfGqJPr8ZI/AAAAAAAACgQ/PVil3sb1Gm4Obn4W5NOFhjXHypbGWVAaA/s1600/IMG_7572.JPG]

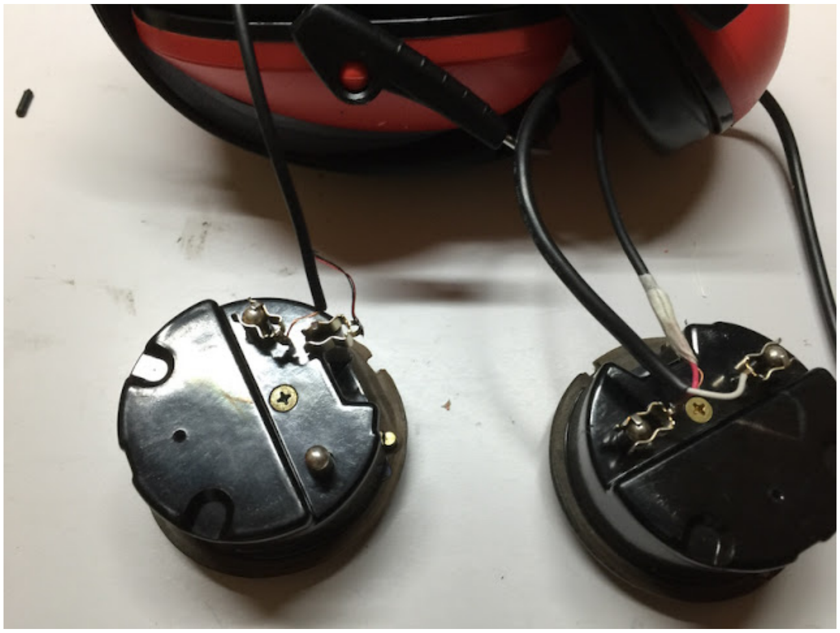

[https://1.bp.blogspot.com/-4kxoFlmV5VY/VwfGqYkTR2I/AAAAAAAACgU/KnQEjTkNxJAhvMMfcZ0dWthyqCzEIHfCQ/s1600/IMG_7573.JPG]

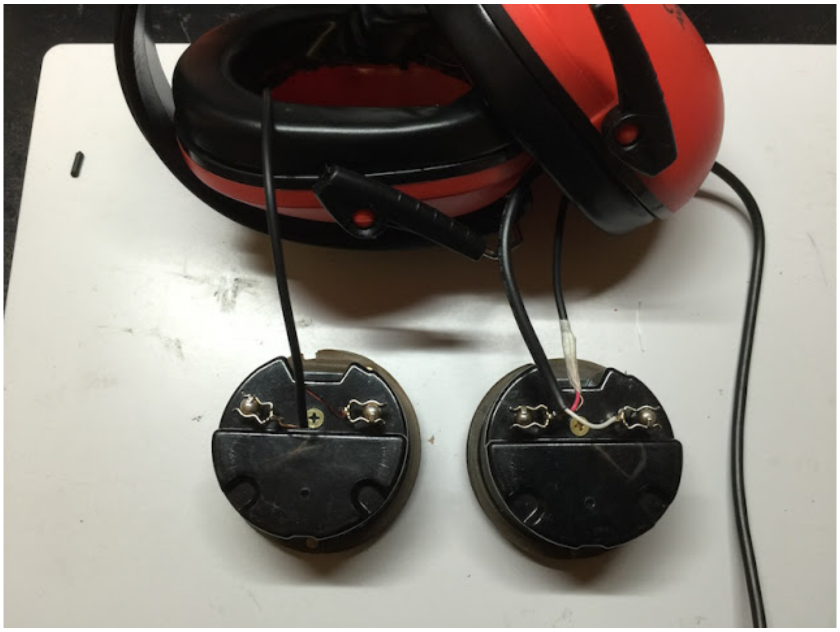

[https://1.bp.blogspot.com/-hg7B12cTHWo/VwfGqkfq4nI/AAAAAAAACgY/HC_z_pKHyzgcvCxvJrGXIDAehmd9n5tbg/s1600/IMG_7574.JPG]

[https://4.bp.blogspot.com/-bLlgV1GKBrw/VwfGqrxI-yI/AAAAAAAACgc/ys-BO_WyO8AyI128XLe0x8gJgv-Rvdl8A/s1600/IMG_7575.JPG]

[https://3.bp.blogspot.com/-AZk53btybkY/VwfGq-wZdlI/AAAAAAAACgg/YxnAK4-ECkIeRVUWbo8WVGsHVa9XDWc0w/s1600/IMG_7577.JPG]

Posted 8th April 2016 by Billy

0 Add a comment

3rd April 2016 : Loop antenna made from soft flat cable with 3mm copper strips

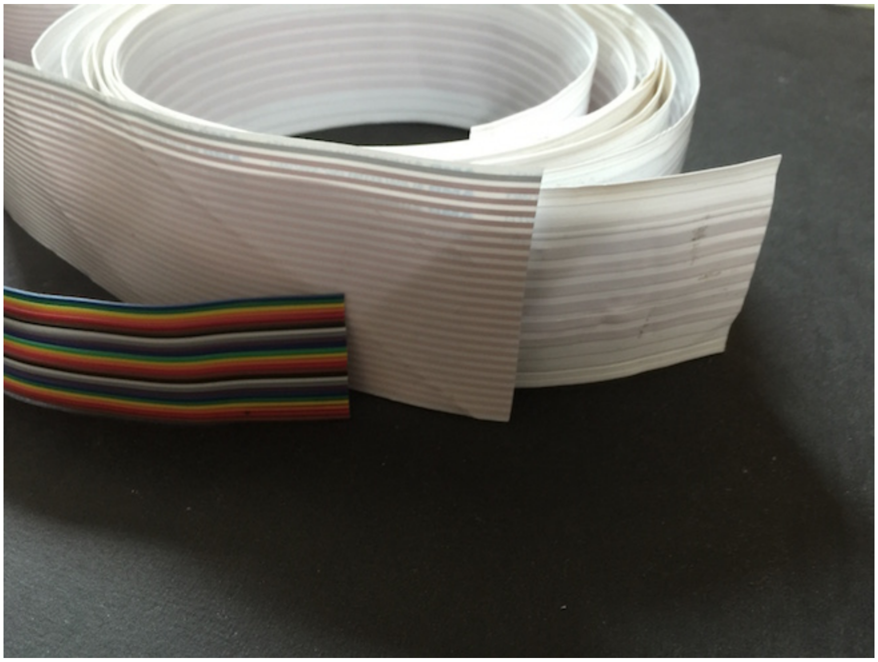

[https://4.bp.blogspot.com/-YflZaWTy81g/VwG6bRubyhI/AAAAAAAACfA/OMGBLhV4-mAHvytwo9sMPOpzfXab-NjAQ/s1600/IMG_6935.JPG]

[https://1.bp.blogspot.com/-GgpW0MsWiDM/VwG6bfJdJLI/AAAAAAAACfE/iksd6VV_3D4axit9f5F-FmCFD3-LFMeFQ/s1600/IMG_6936.JPG]

Dynamic Views theme. Powered by Blogger.

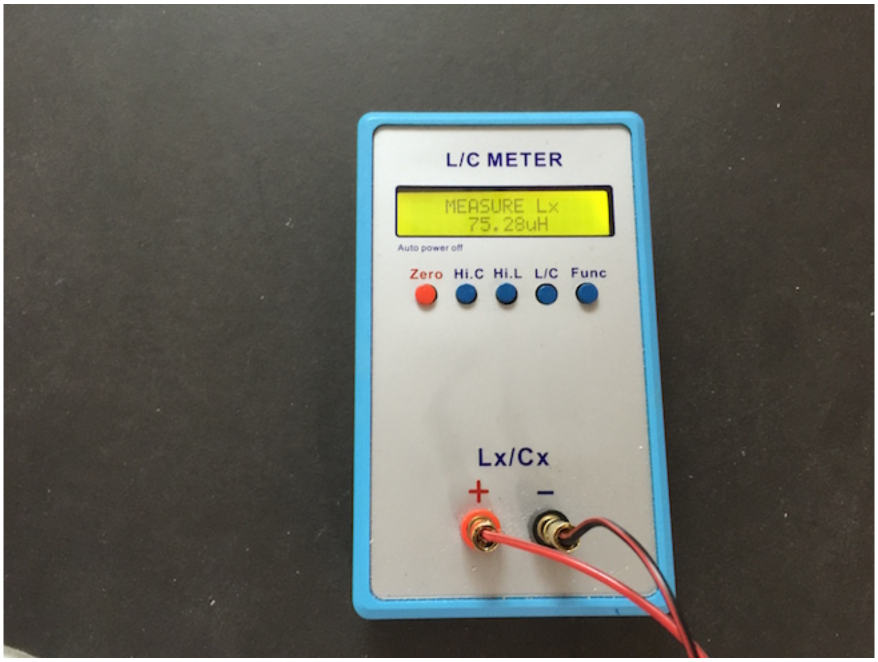
L/C METER
MEASURE Lx
75.28uH
Auto power off
Zero
Hi.C
Hi.L
L/C
Func
Lx/Cx
+
−

[https://1.bp.blogspot.com/-raLgBfRgy-A/VwG6bX8nDSI/AAAAAAAACe8/28a0RIN6FdoXjvXwJbmiKEL4i79L9ejYg/s1600/IMG_6937.JPG]

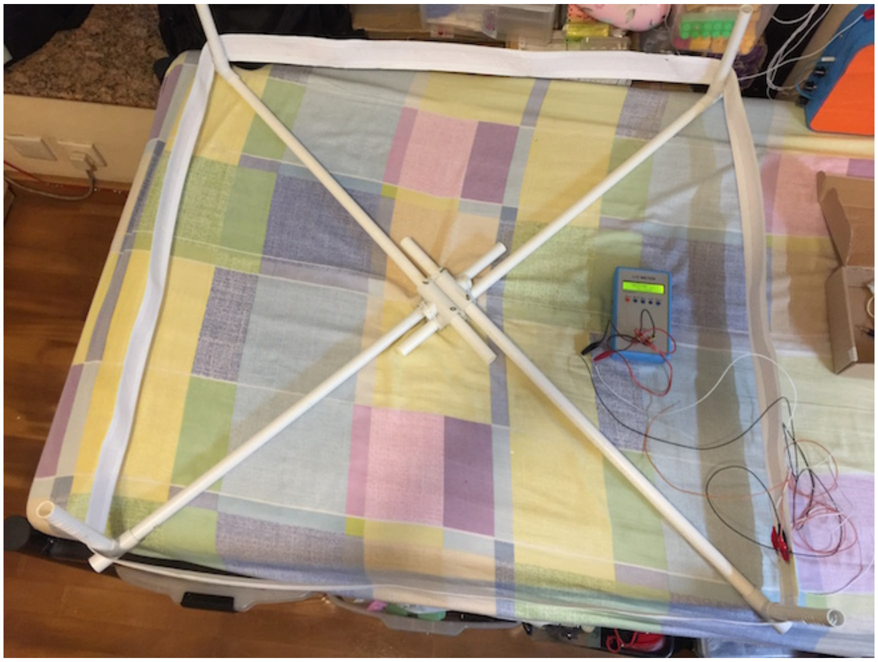

[https://3.bp.blogspot.com/-0V3S5IlUg80/VwG6b2aFkWI/AAAAAAAACfI/9ZtpD9jicv4eKQ8XLdT-BMWhqmWmg5QTQ/s1600/IMG_6938.JPG]

Dynamic Views theme. Powered by Blogger.

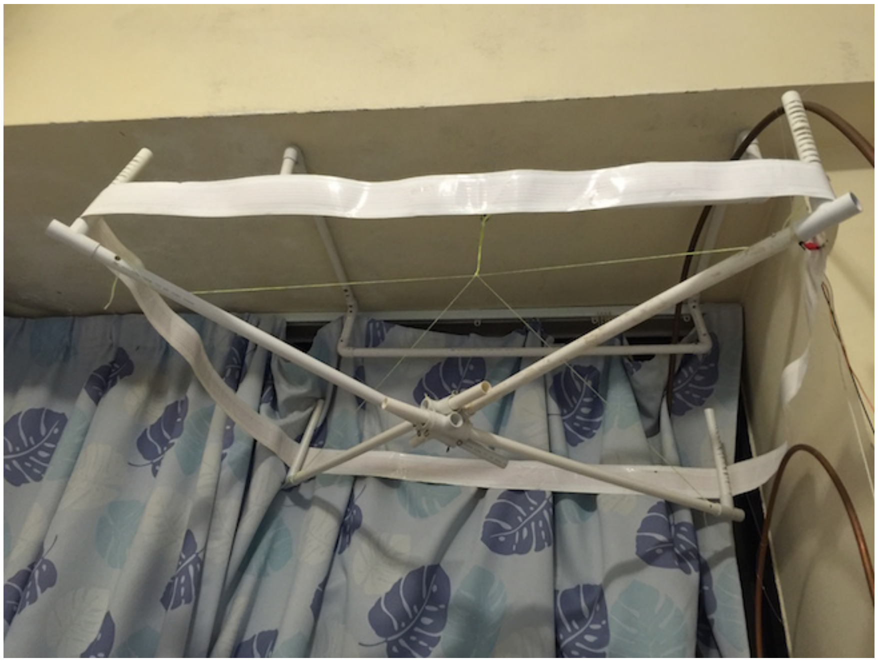

[https://2.bp.blogspot.com/-25rUWT5Zm1c/VwG6b_FjUEI/AAAAAAAACfM/UG8R-49XUSsSH_LH6FY0cd0RJ-ObbZ-4g/s1600/IMG_6942.JPG]

I tried to create a loop antenna coil using soft flat cable with 20 x 1mm flat copper strips with good effect before (the one in the middle of the picture).

This time I found find a bigger soft flat cable with 6 x 3mm flat copper strips with a 2mm separation in between (the one on the left of the picture).
You can make a loop solder the copper strips together to make a single coil with 6 turns.
For use with FET (3SK143-Q), 1 turn is reserved for the S-pole of the FET.
Total inductance is 75uH when I mount the laps placed on a square frame made of PVC pipes measured 900 mm on each side.
This matches with 2 x 500pf variable capacity connecting in parallel making a 1000 pf Variable capacitor.
Result is quite good, though still not comparable with a coil made from 0.04x660 Litz wires of the same size.
However, this one is easy to make and quite portable. You can just roll them up and carry them anywhere.
Find some sticks to create a frame and mount it on top. Or hang it from a tree when picnicking.

Posted 3rd April 2016 by Billy

0 Add a comment

2nd April 2016

DIY Tiny portable crystal radio laboratory 135 x 70 x 30 mm

Here are the pictures:

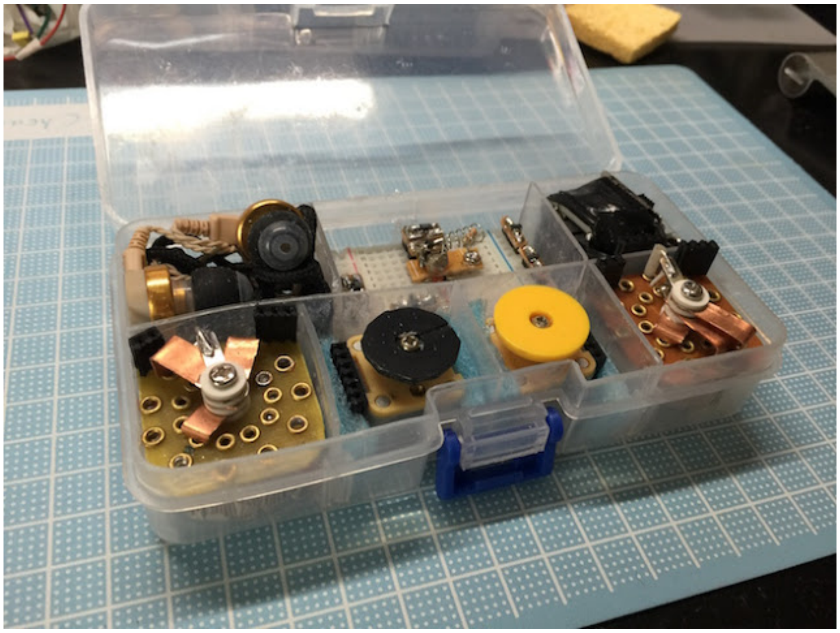

Inspired by skywave's traveling crystal set. I built this DIY Tiny (135 x 70 x 30 mm in size.) portable crystal radio laboratory.

I found a tiny sorting box made of low loss material (pp) and start to put things in:

- a couple of the DIY tiny multi-pole rotary switches (that I built) soldered in with the ferrite coils. I used 31mm ferrite core with100u permeability and Litz wire of 0.04 mm x 60 threads (see the picture below for the no. of turns).The shortest handle of the rotary switch selects the number of turns of the antenna coil. The longest handle selects the tapping point of the detector in the tank coil. The one in the middle selects the number of turns of the tank coil to match with the tuning capacitor
- a couple of 2x270pf plastic variable capacitor.
- a pair of 150 ohm hearing aid ear buds (total 300 ohm).
- a tiny 200K to 300 ohm audio transformer that I purchased from taboo.com soldered in with switches for the different way to connect , e.g. 200k to 300 ohm for my hearing aid ear buds, or just connect 200K in parallel with my piezoelectric ear bud.

The transform can be improved as its not ideal to have impedance fixed at 200k. I am trying to open it up to see if I can add additional taps without breaking the hair thin wires inside.

- A breadboard with the following different types of detectors
-- Crystal
-- BAT85
-- 1N60
-- 1N34A
-- 1SS86 x 2
-- HSM 2820 x 2
-- HSM 2850 x 2
-- 3SK143-Q (*3DQ) x 2
-- 33K resistor to work with 3DQ.
- Connecting wires (hidden under the breadboard).
- Tiny 50uA DC current meter (stilling trying to find one. So far I can only find a tiny 500uA meter that does not move much as the crystal radio current is so small). If I ever find one, it'll take the place currently used by the earbud.

I tried this in the hotel room when traveling and got good results.
- Tested using my 700mm umbrella loop antenna.
- Tested by hanging 3 meters of horizontal wire at the top of the curtain inside the room. [https://4.bp.blogspot.com/-Uq_TZea4wT4/VwBnupV99NI/AAAAAAAACeQ/OMB8y1i7TUsNOj0Eox2Ag9o-DGAV79HJw/s1600/IMG_6921.JPG]

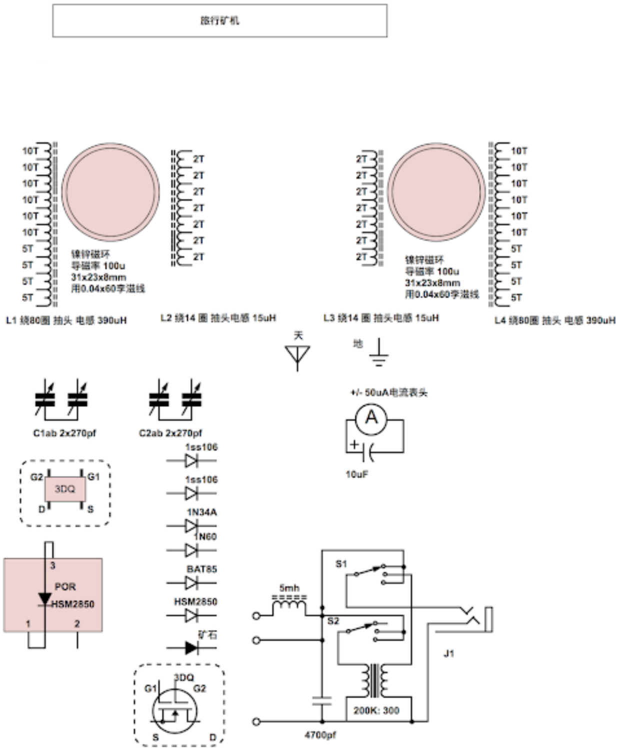
旅行矿机
10T
10T
10T
10T
10T
10T
5T
5T
5T
5T
镍锌磁环
导磁率 100u
31x23x8mm
用0.04x60李滋线
2T
2T
2T
2T
2T
2T
2T
L1 绕80圈 抽头 电感 390uH
L2 绕14 圈 抽头电感 15uH
L3 绕14 圈 抽头电感 15uH
L4 绕80圈 抽头 电感 390uH
天
地
C1ab 2x270pf
C2ab 2x270pf
+/- 50uA电流表头
A
10uF
G2
G1
3DQ
D
S
3
POR
HSM2850
1
2
1ss106
1ss106
1N34A
1N60
BAT85
HSM2850
矿石
3DQ
G1
G2
S
D
5mh
S1
S2
J1
200K: 300
4700pf

[https://4.bp.blogspot.com/-4akx2SkEdjU/VwBstylVhDI/AAAAAAAACes/teuGgRr0u9EyBOiGRm-BiyMl4ZpEH1OMQ/s1600/Screen%2BShot%2B2016-03-06%2Bat%2B10.59.31%2BPM.png]

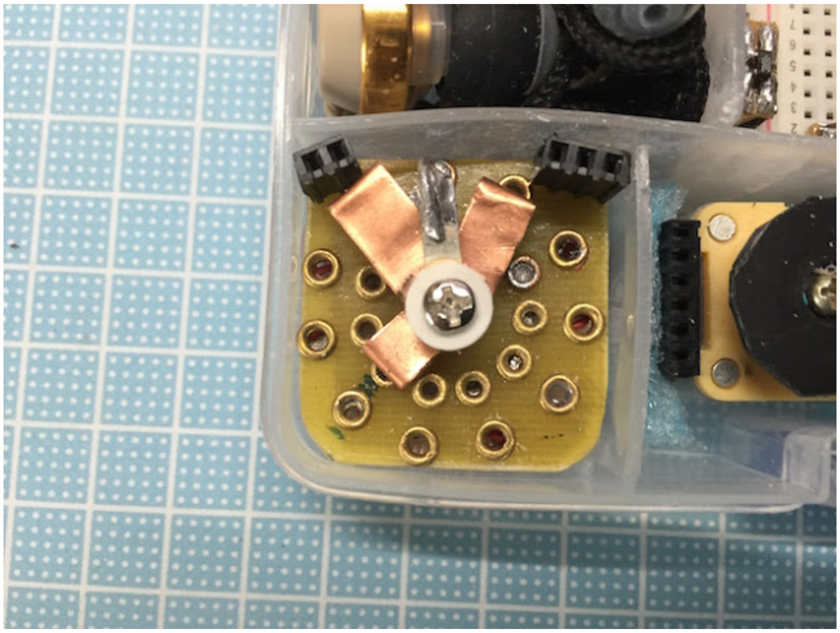

[https://1.bp.blogspot.com/-yylK17YOGV0/VwBnt1y4D_I/AAAAAAAACeA/WTuQCIOAYmcfpZ5ZTmJjAVaH72K2bpSmg/s1600/IMG_6917.JPG]

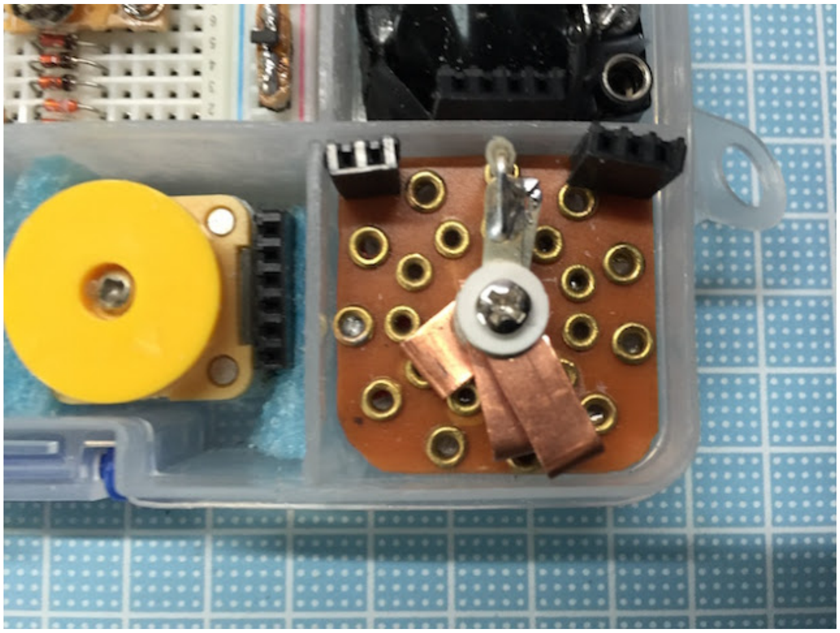

[https://3.bp.blogspot.com/-XIfomc4Xom4/VwBnt3EiYLI/AAAAAAAACd8/Y5QgpHjMWM4hhB_NLiIxvjXAoicfQPd7g/s1600/IMG_6918.JPG]

[https://3.bp.blogspot.com/-M1ndSnQEU-k/VwBnuTtKiaI/AAAAAAAACeI/-l029n_uDjgBjMlk78nJ_7fn7st6j9-1A/s1600/IMG_6919.JPG]

[https://3.bp.blogspot.com/-t1rJfevWnFo/VwBnukM43DI/AAAAAAAACeM/By6uiDuYMEwvISn5t6Soj9q_xSUR36IeA/s1600/IMG_6920.JPG]

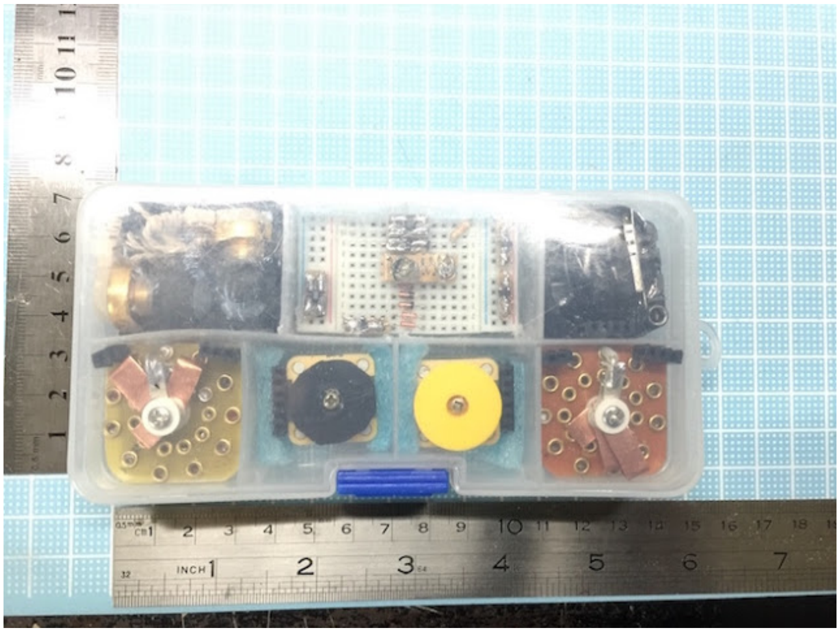

[https://2.bp.blogspot.com/-RlpVGlrkVbo/VwBnu2DM8qI/AAAAAAAACeU/4gvaXElH8lkNs8P7d81F9r1Oc0BPOLV6Q/s1600/IMG_6922.JPG]

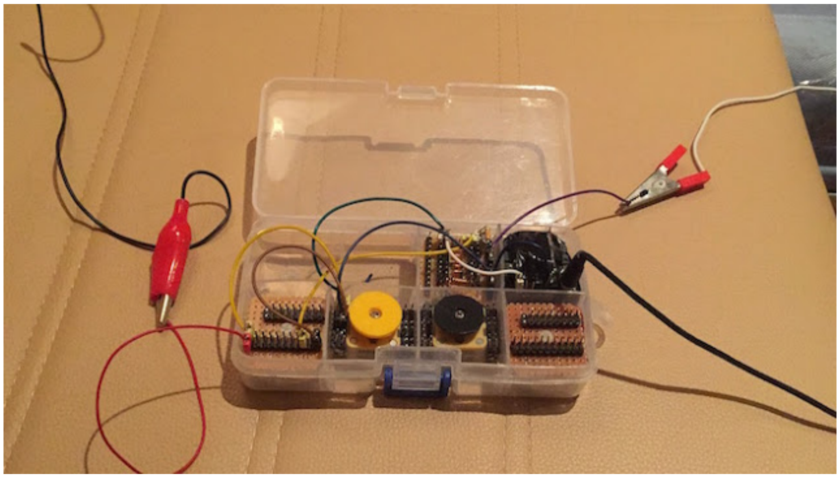

[https://1.bp.blogspot.com/-VAHDtVH_RwA/VwBn6fm2QcI/AAAAAAAACeY/OZU-FImODiUTo3P2Aa3ZMgJXHTJFq0Zsw/s1600/IMG_6538.JPG]

[https://2.bp.blogspot.com/-sh6TDwIRWf4/VwBoGMznkJI/AAAAAAAACec/fiF-jcUwiEcjDl7_WKHE2fvst41d4fTuw/s1600/IMG_6533.JPG]

Dynamic Views theme. Powered by Blogger.

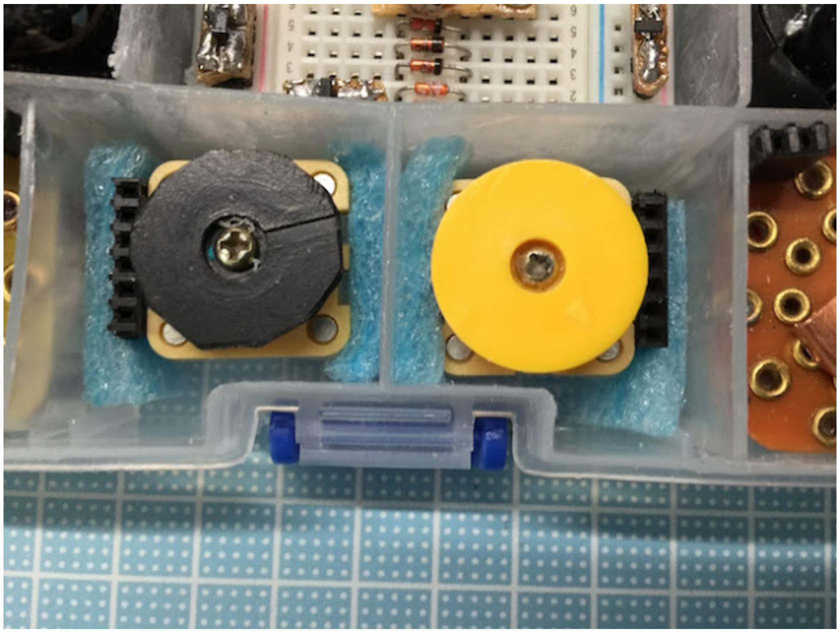

[https://1.bp.blogspot.com/-Re6TZmQrYU4/VwBntySygKI/AAAAAAAACeE/u-mA5R-jma4e2qgUjJd7plK5bF2anqABg/s1600/IMG_6916.JPG]

Posted 2nd April 2016 by Billy

2 View comments

Bitdefender -56% Be aware without a care! BUY NOW

22nd March 2016

DIY dual pole 18 position switch for KPB-02 Impedance matching transformer

I DIY a dual pole 18 position switch for KPB-02 Impedance matching transformer and another dual pole 10 position switch for tapping a ferrite toroid coil.

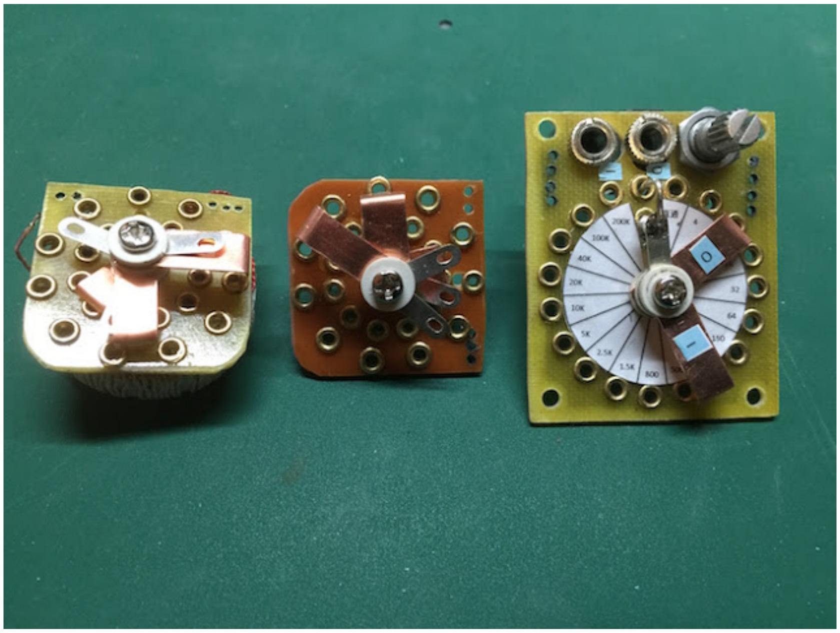

[https://4.bp.blogspot.com/-dsvWlUJsZt0/Vvk-gbMmwcI/AAAAAAAACcs/-LbWdW_k6e8ta4k6DiuVCZH_ajYONCfZg/s1600/IMG_6863.JPG]

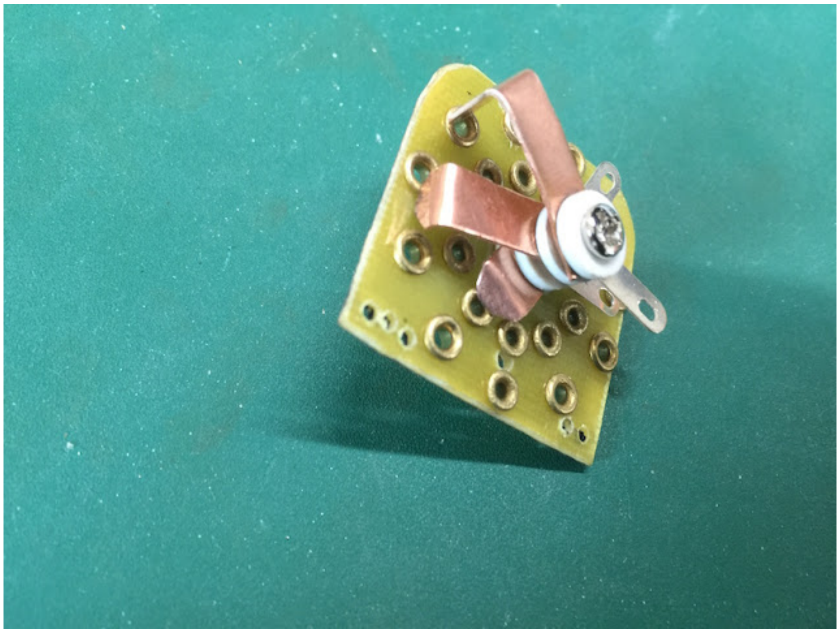

[https://4.bp.blogspot.com/-qYMkNI_6DPA/Vvk-fnxCK5I/AAAAAAAACcg/yzmJHUAADoMOjk4C5BmEAOuawqsV-0rxA/s1600/IMG_6852.JPG]

I added another 3 pole 10 position switch for tapping the ferrite toroid coil.
1st pole for the tap on the antenna coil.
2nd pole for the tap on the tank coil for the variable capacitor for tuning.
3rd pole for the tap on the tank coil for the detector (divide).

Here are the pictures.

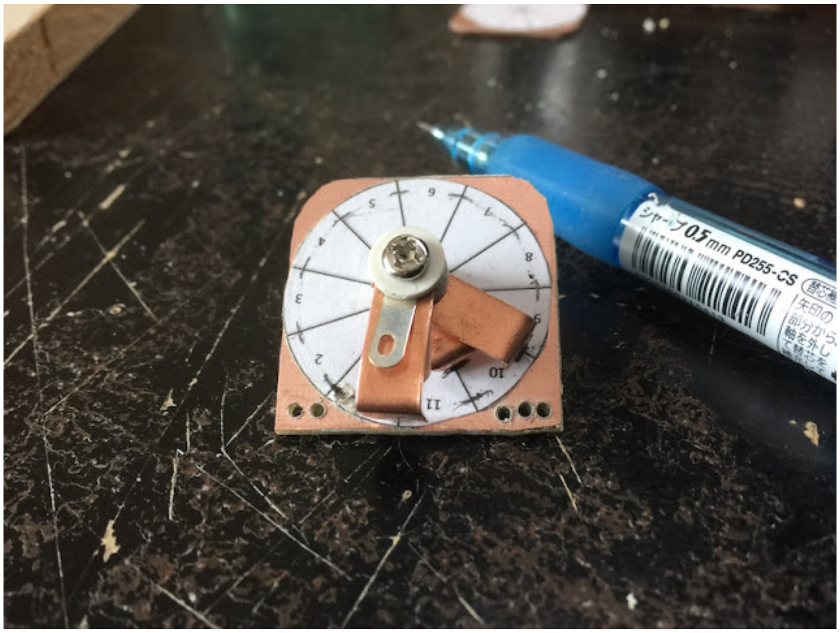

[https://3.bp.blogspot.com/-kcCmc8Jij4g/Vvk-b0wIukI/AAAAAAAACbQ/XcphLL3-pD06vlt2-2NPDkGd3U8wkGaaw/s1600/IMG_6822.JPG]

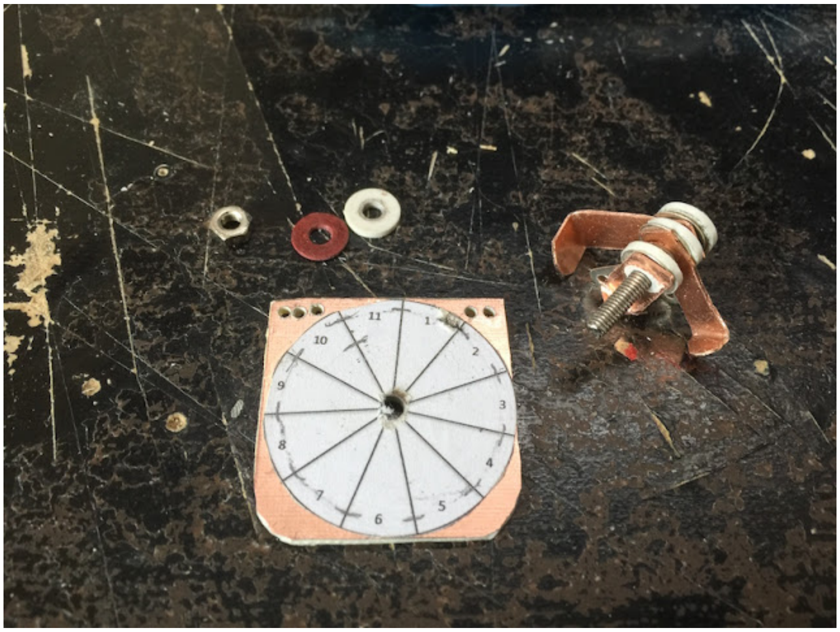

[https://2.bp.blogspot.com/-ZXEgnsZHHjs/Vvk-b06KCII/AAAAAAAACbU/JQXWjKY2Z5UpraMO-SLitK_XFVyYTjUQA/s1600/IMG_6823.JPG]

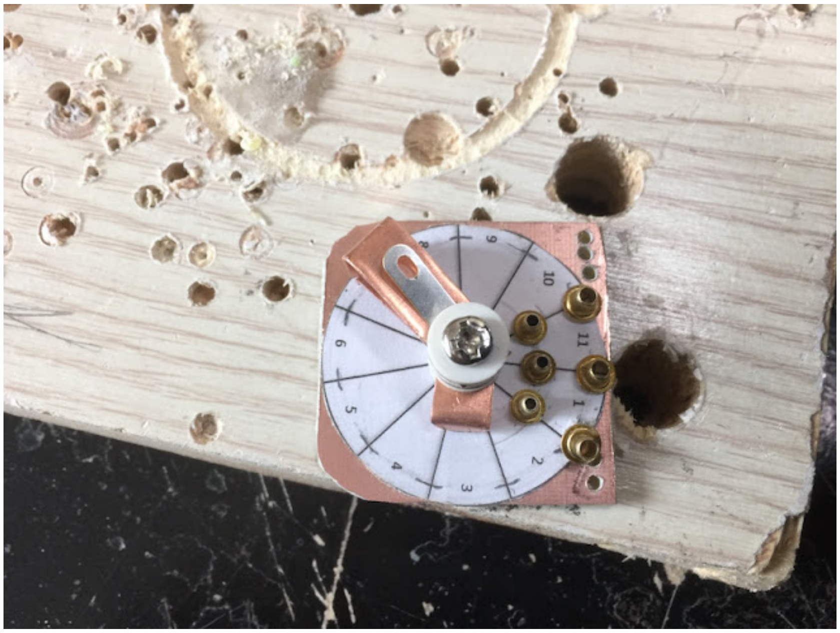

[https://1.bp.blogspot.com/-K-rTKIGPm3k/Vvk-bw6tj6I/AAAAAAAACbM/uYQQVXIVeUEhvW5jsedKVWSZUqXtxxwIA/s1600/IMG_6824.JPG]

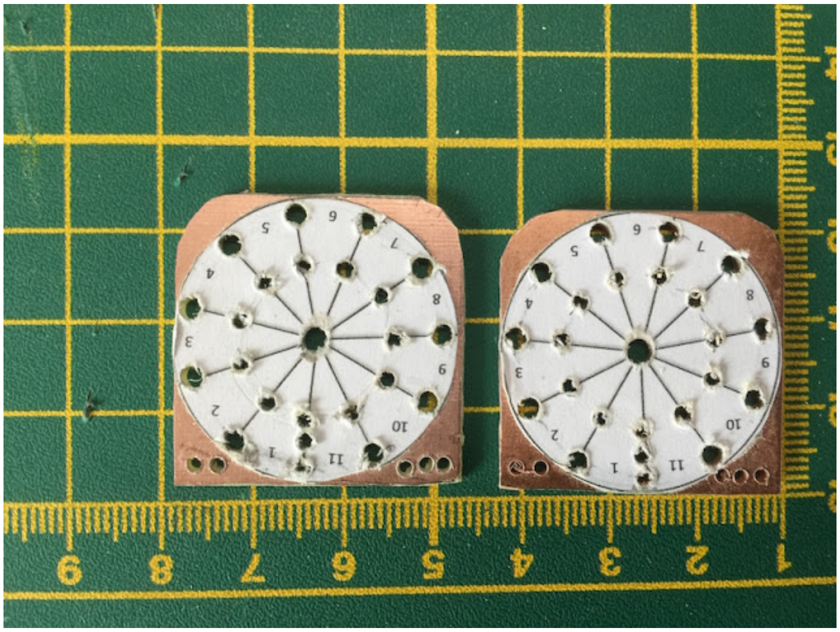

[https://4.bp.blogspot.com/-xu5QZcdjUZE/Vvk-cRU18sI/AAAAAAAACbY/zjaSm1T_ZXAbsJwOJrzl79bg2f0czAB2w/s1600/IMG_6825.JPG]

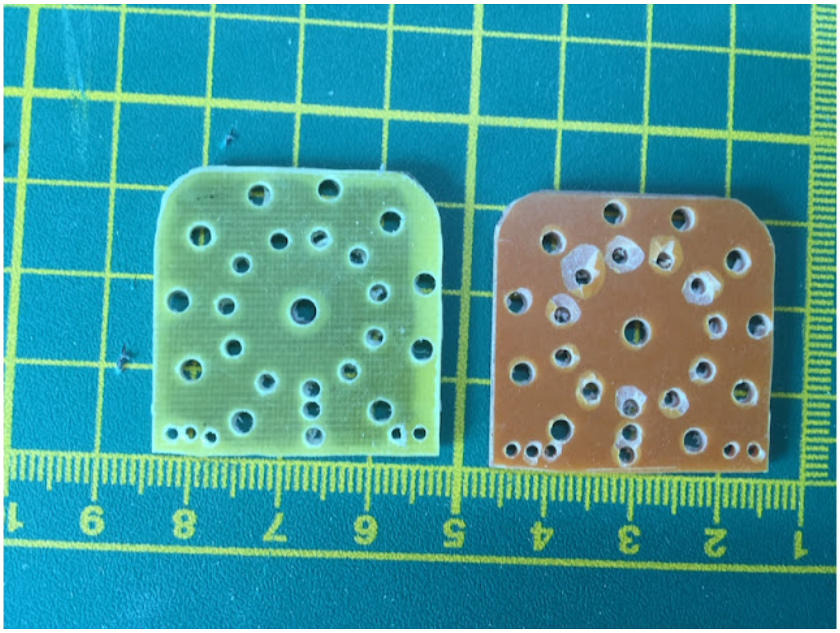

[https://2.bp.blogspot.com/-PXJH_LzAgpY/Vvk-cmmfrqI/AAAAAAAACbc/ezOVIPGHKjUyE5fYrCM7KtwbUbQnM08kQ/s1600/IMG_6826.JPG]

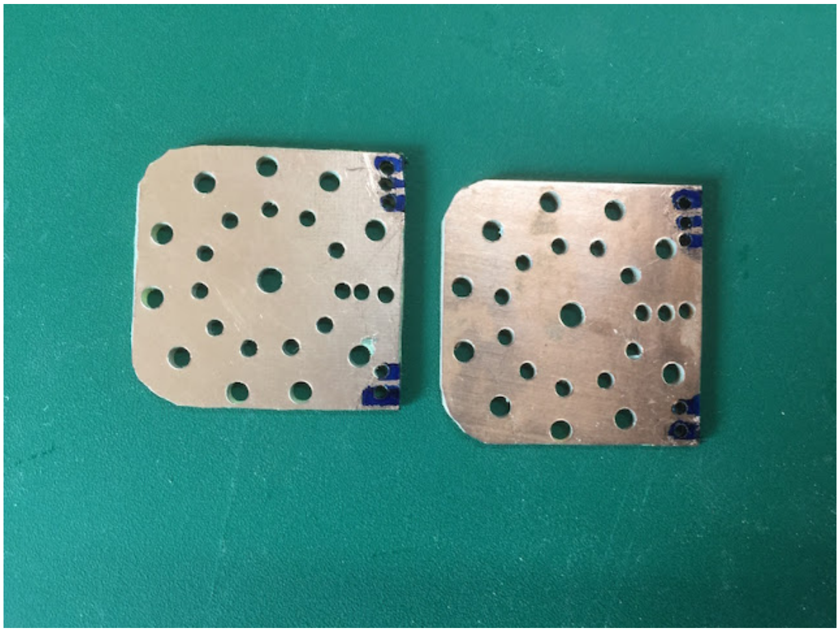

[https://3.bp.blogspot.com/-LhZIptJM-Yw/Vvk-crb8SiI/AAAAAAAACbg/cwwfx5GEPGsZfcFFsCET8dLutteB2jfQg/s1600/IMG_6827.JPG]

[https://4.bp.blogspot.com/-CDkpaG7BEhQ/Vvk-c2-5SUI/AAAAAAAACbk/_lw_KYAyAasRZRANSHRyIZ3TgGnLVLt4g/s1600/IMG_6828.JPG]

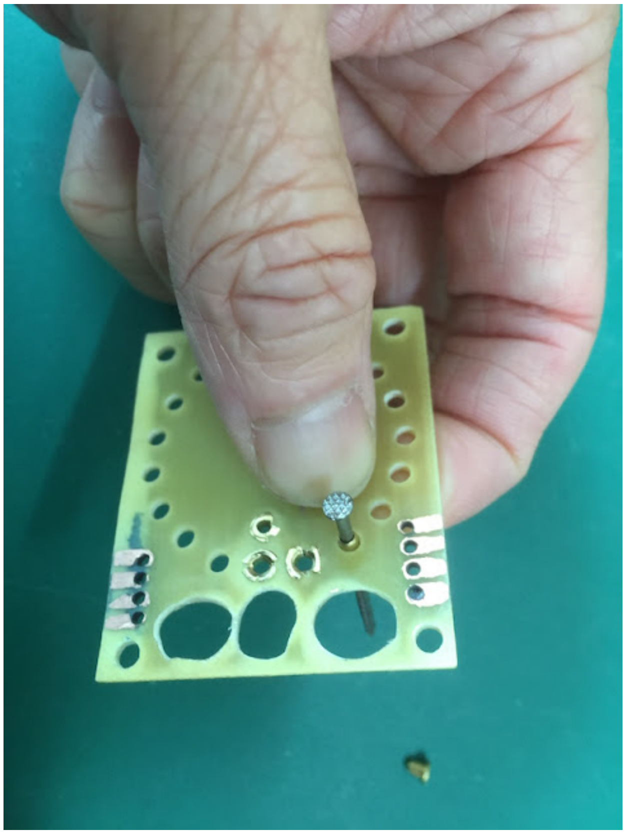

[https://1.bp.blogspot.com/-T4Nx0A0Xiug/Vvk-dItWs4I/AAAAAAAACbs/1sFJCl-bDvEQPPzAM8wZE3jXRhAxmPc6Q/s1600/IMG_6830.JPG]

Hand Riveter
GS

[https://2.bp.blogspot.com/-qknXaa_KR1o/Vvk-denerDI/AAAAAAAACbw/5hWp_kJYI5kRjYQ7-wgm2HV2OtlKDP4wA/s1600/IMG_6831.JPG]

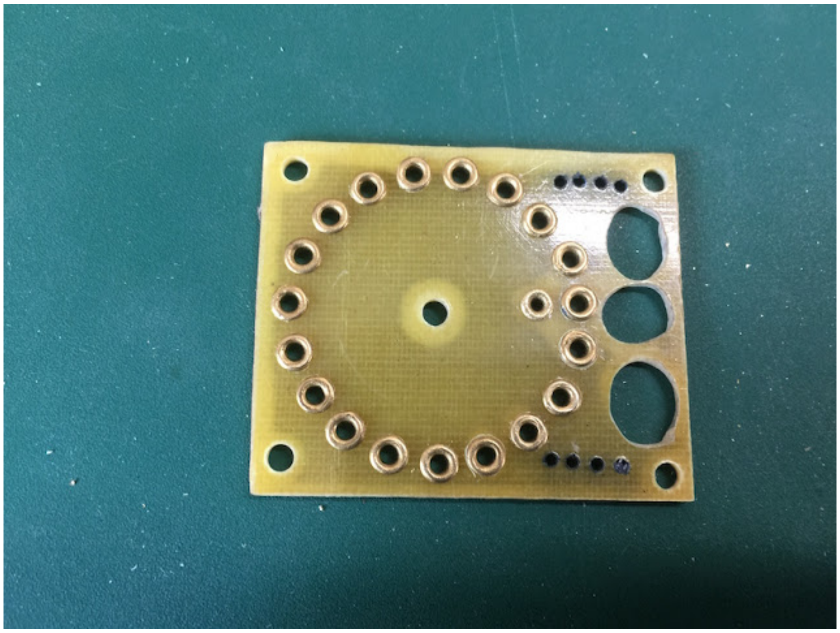

[https://2.bp.blogspot.com/-i1zFsLZDhKA/Vvk-d8AJbTI/AAAAAAAACb8/bKygUQuYiCkezQM1SCKIJ3JIAHSr7LVXA/s1600/IMG_6834.JPG]

Dynamic Views theme. Powered by Blogger.

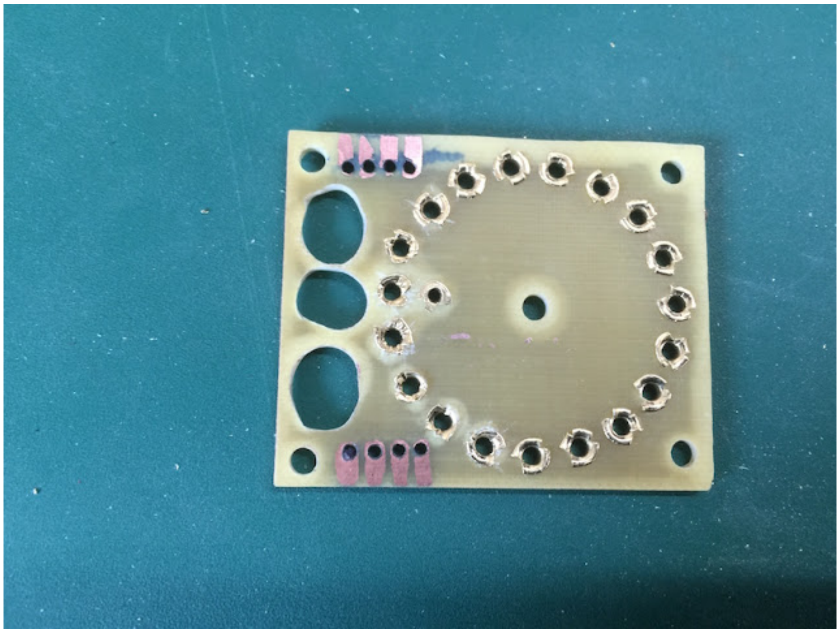

[https://1.bp.blogspot.com/-6-FkTEhTAeY/Vvk-eL4TtqI/AAAAAAAACcA/ZX4A6lfG5ic8LJhtoY_wYHOQGijDMI5Rg/s1600/IMG_6835.JPG]

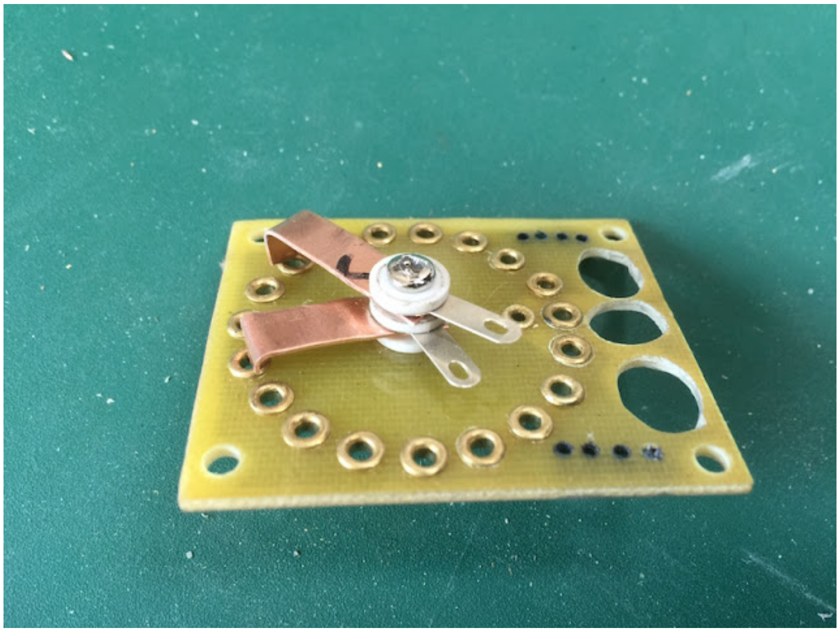

[https://3.bp.blogspot.com/-1Gt8Ix0fPDI/Vvk-eQj_lxI/AAAAAAAACcE/bMjKgu0iubc93sTmvGmOSJhuxZWPtCFEg/s1600/IMG_6836.JPG]

Dynamic Views theme. Powered by Blogger.

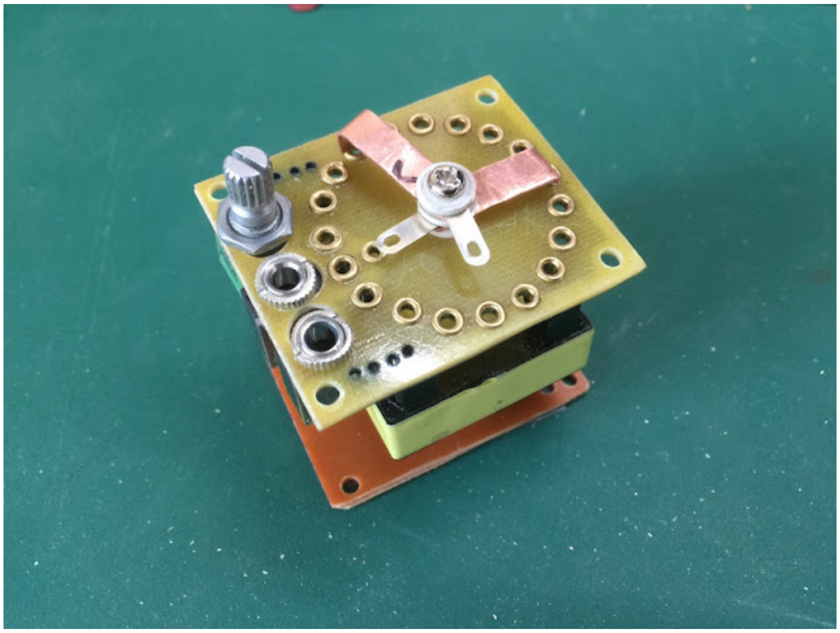

[https://1.bp.blogspot.com/-bFo1_QNh56A/Vvk-epdEYDI/AAAAAAAACcI/GRSoPr2iFQclA7F1CWBm9MBmUXVH_Jq-Q/s1600/IMG_6841.JPG]

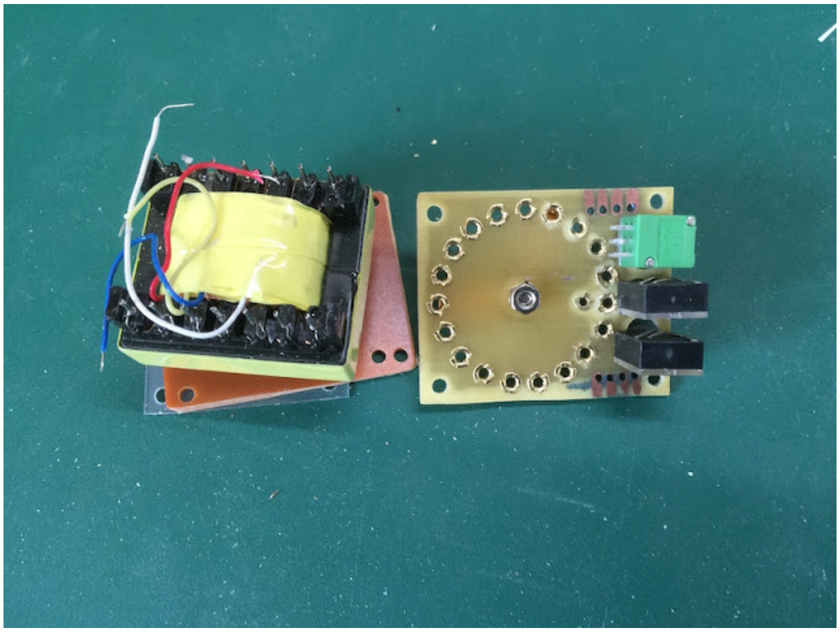

[https://4.bp.blogspot.com/-N6HYIMOa8aY/Vvk-eq1cjGI/AAAAAAAACcQ/6Aqve2gttuoi5ehw1AJIbs8pO2nnFxCQQ/s1600/IMG_6842.JPG]

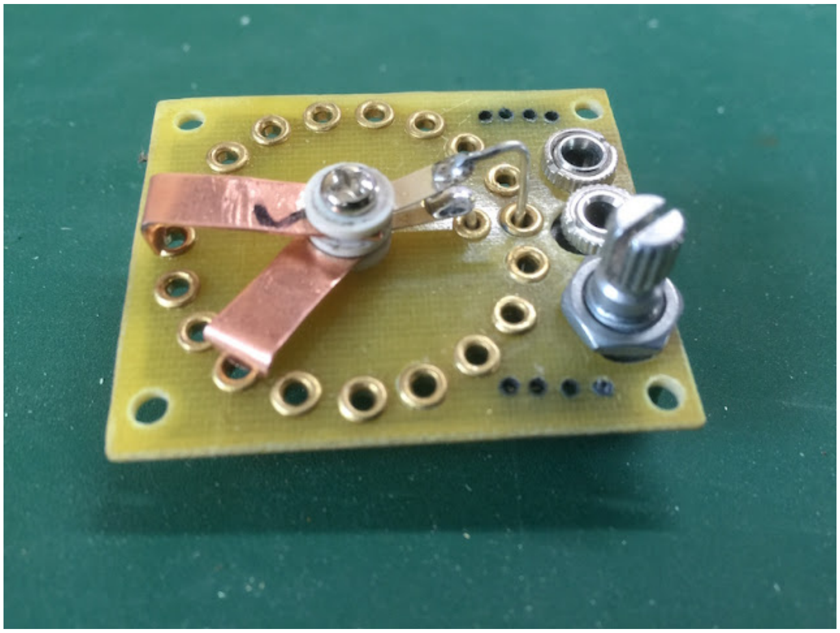

[https://4.bp.blogspot.com/-cXYxOT21-AA/Vvk-e3WsedI/AAAAAAAACcM/HKNW4Lq_65s7VgGzr4QcGpoVQsF2VRdjA/s1600/IMG_6843.JPG]

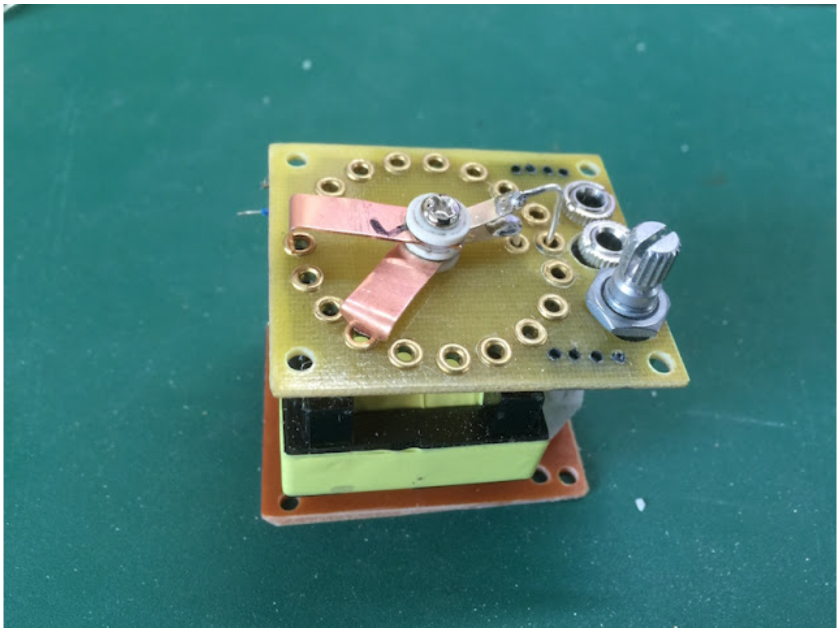

[https://1.bp.blogspot.com/-T0DFhqeRVas/Vvk-flusuuI/AAAAAAAACcU/7iem2U95i5UwOs4TbiJUZwE_OyPBJN6pg/s1600/IMG_6844.JPG]

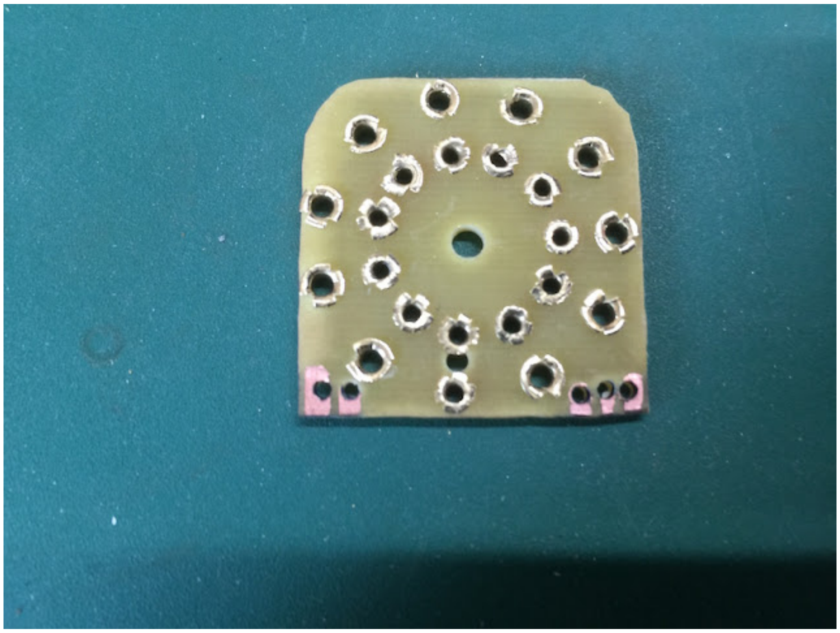

[https://1.bp.blogspot.com/-45GJ8ODp8H8/Vvk-fVUeMtI/AAAAAAAACcY/xGdgisYfXuwxzLqjRNSe7073mkbvW1q4A/s1600/IMG_6850.JPG]

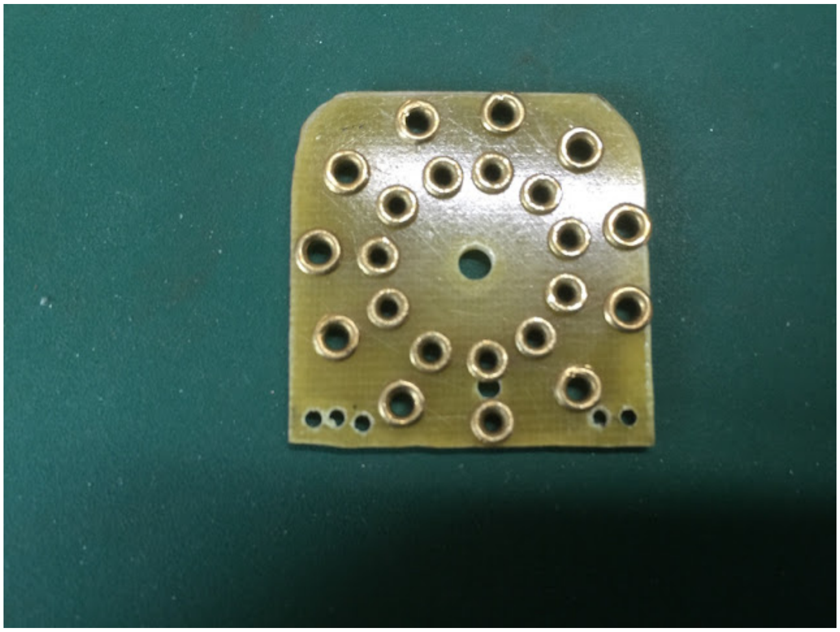

[https://2.bp.blogspot.com/-BD3Y0VW5eVE/Vvk-fZDS2zI/AAAAAAAACcc/_E9rEBwXF_IkW-NVPiRmt9eLA9ReBHdhw/s1600/IMG_6851.JPG]

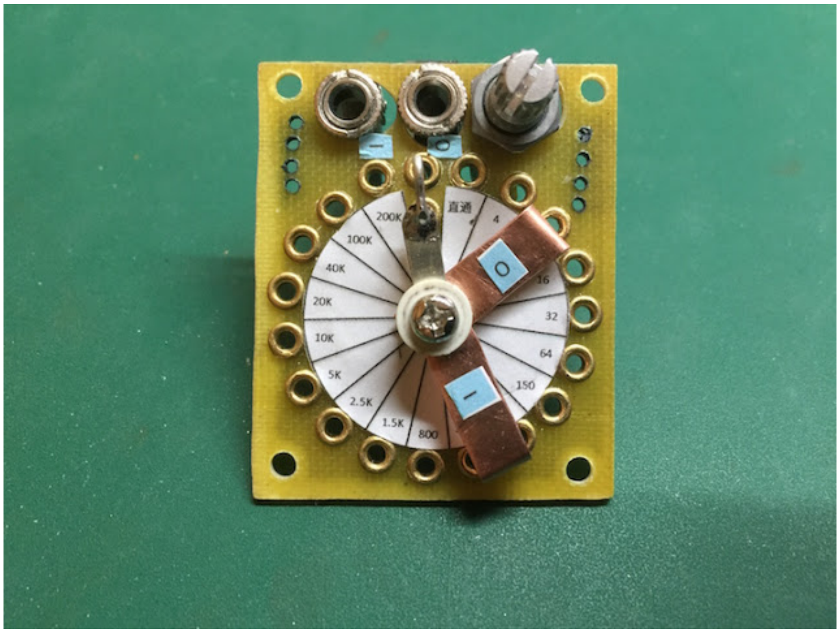
200K
100K
40K
20K
10K
5K
2.5K
1.5K
800
150
64
32
16
4
直通
O
I

[https://2.bp.blogspot.com/-dlNdHT6udhA/Vvk-f1obCCI/AAAAAAAACck/d4xpmtl1Rb4into5mVlEizAQexiSjQGEg/s1600/IMG_6861.JPG]

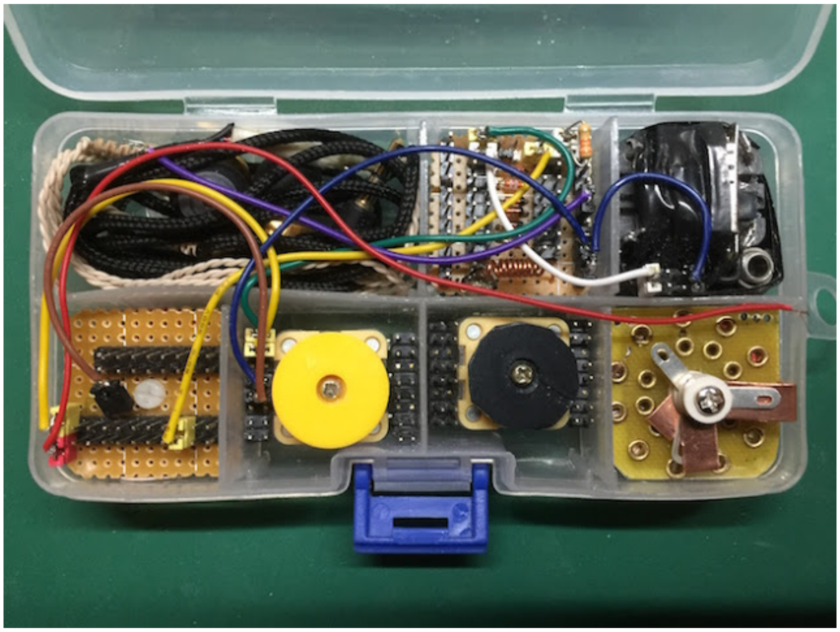

[https://1.bp.blogspot.com/-2z1nsXdOZLQ/Vvk-f04Xv6I/AAAAAAAACco/dYkA3O7NUcII69OdTT0uCZ_PwgcwiF1NA/s1600/IMG_6862.JPG]

Dynamic Views theme. Powered by Blogger.

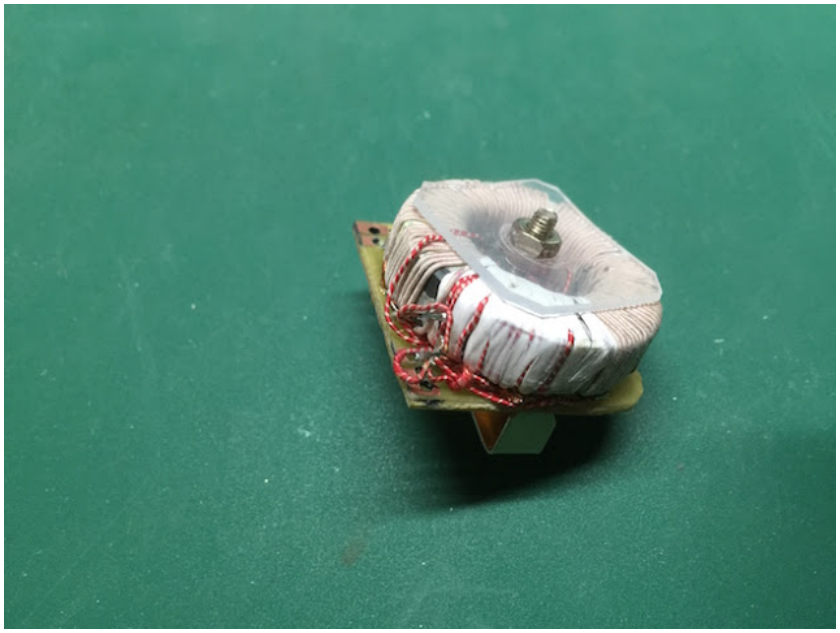

[https://4.bp.blogspot.com/-SbGC34KVvSw/Vvk-geO9TfI/AAAAAAAACcw/Y2qdENqRa5kYt5bvNj6oPPmChnQiLKwig/s1600/IMG_6864.JPG]

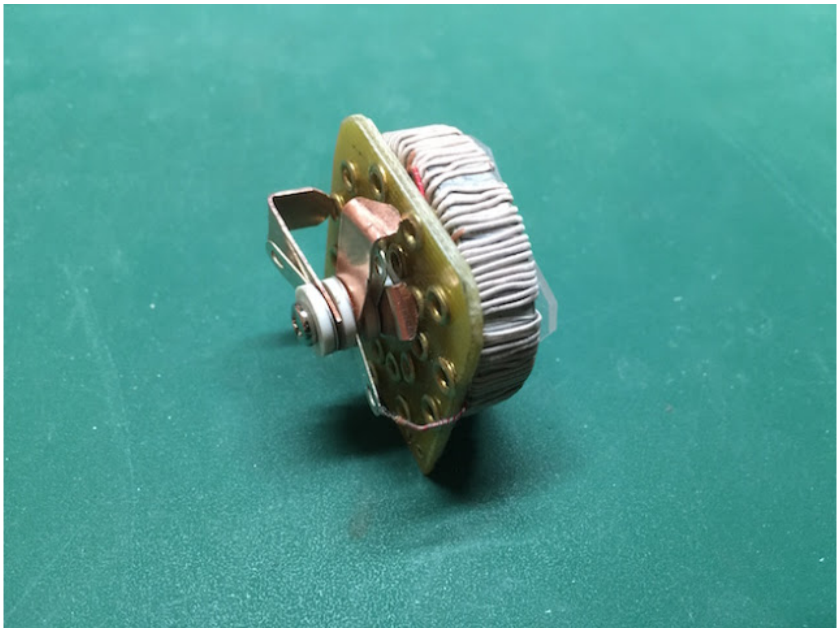

[https://1.bp.blogspot.com/-A1InIO3DjHI/Vvk-gl0bbqI/AAAAAAAACc0/FtAdRbczrFU_D5mvMyoG44F1OmAvIA_kA/s1600/IMG_6865.JPG]

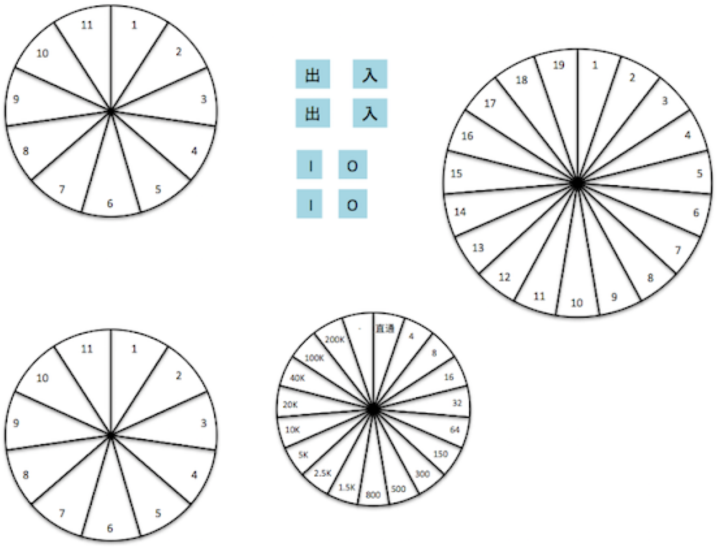
1
2
3
4
5
6
7
8
9
10
11
出
入
出
入
I
O
I
O
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
1
2
3
4
5
6
7
8
9
10
11
直通
4
8
16
32
64
150
300
500
800
1.5K
2.5K
5K
10K
20K
40K
100K
200K
-

[https://4.bp.blogspot.com/-edJ1EvaKtsk/Vvk-g-KZGDI/AAAAAAAACc4/auXgDNgwPnYwFz-UFOFHpmFMJhkKqp-jw/s1600/Layout.png]

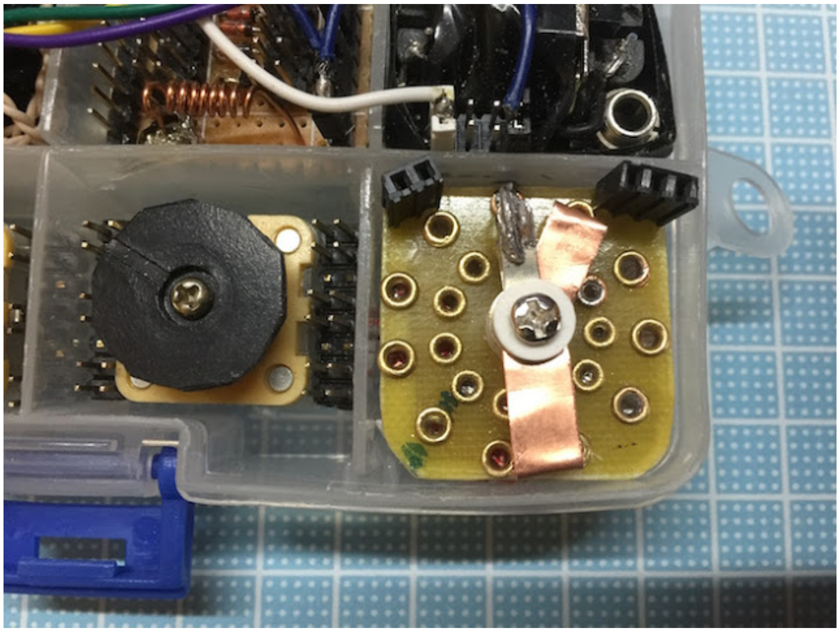

[https://1.bp.blogspot.com/-l5t60ZdibCY/VvsbBczXwaI/AAAAAAAACdU/m4Ul34qcECgRzrZEkkyksf6H8EPeV2MRQ/s1600/IMG_6875.JPG]

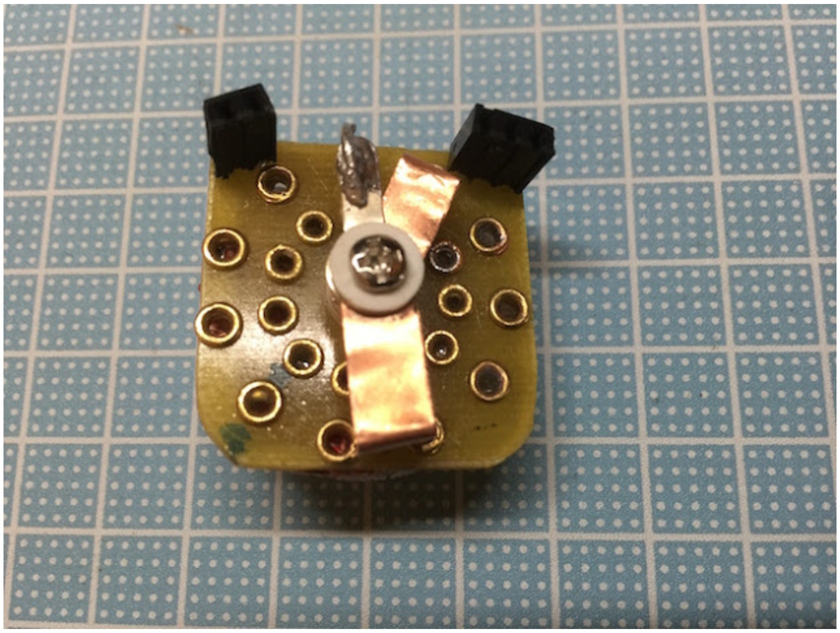

[https://4.bp.blogspot.com/-NscIA4MtKxw/VvsbBU6SK5I/AAAAAAAACdQ/m1h4iXC2cSwHmdkO4f0Y4RFgNnqytjooA/s1600/IMG_6876.JPG]

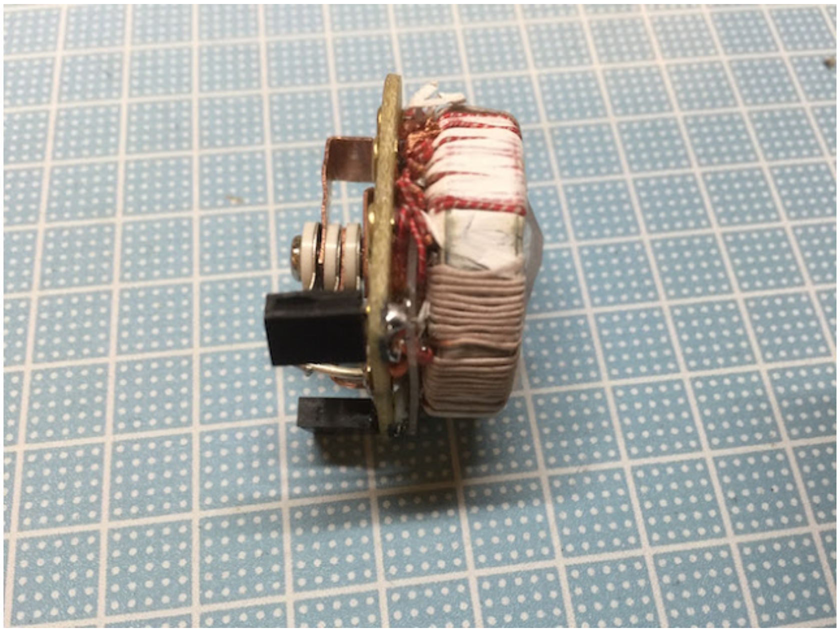

[https://3.bp.blogspot.com/-Utdoi9RAU60/VvsbBeIsHNI/AAAAAAAACdY/PrN7i-vZ7SEk3AU8WOzARg7Wn8PL5iZtg/s1600/IMG_6877.JPG]

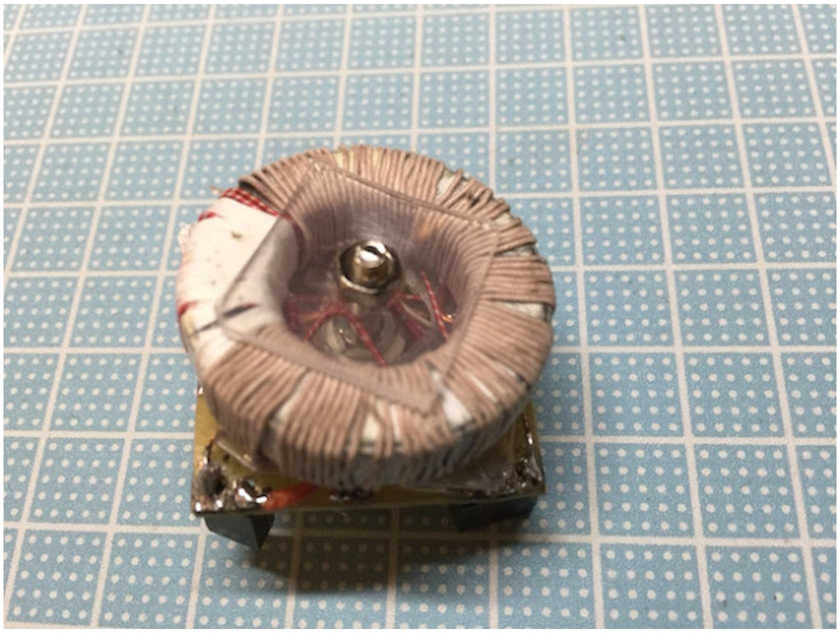

[https://4.bp.blogspot.com/-STfprMEm-W0/VvsbBuerdcI/AAAAAAAACdc/1L8YCJhZnJcjGLjPSPPChpp8MwNPO6CSw/s1600/IMG_6878.JPG]

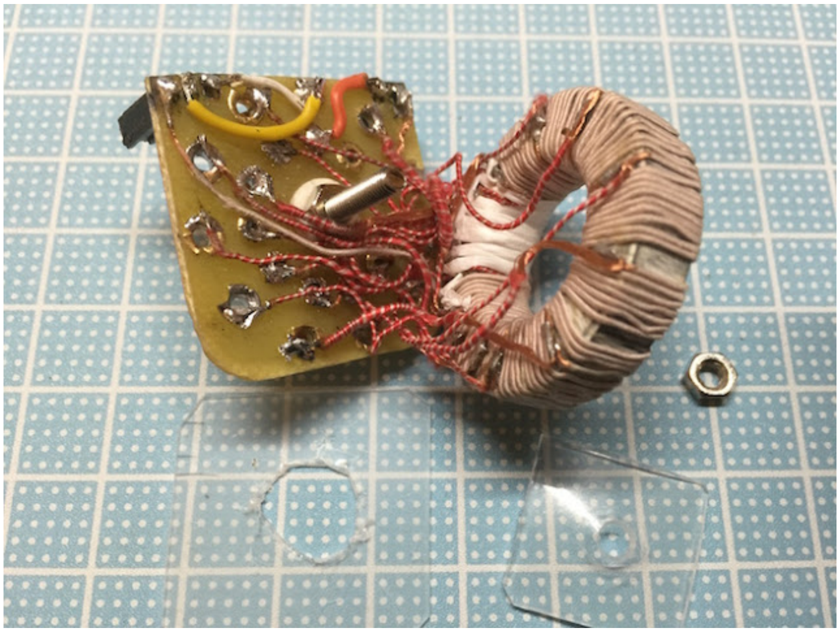

[https://3.bp.blogspot.com/-vTrl8JHvXB8/VvsbBxuk9LI/AAAAAAAACdg/SFhsbisMqVkg2O84XDNzeJzHeavw7Rhnw/s1600/IMG_6879.JPG]

Dynamic Views theme. Powered by Blogger.

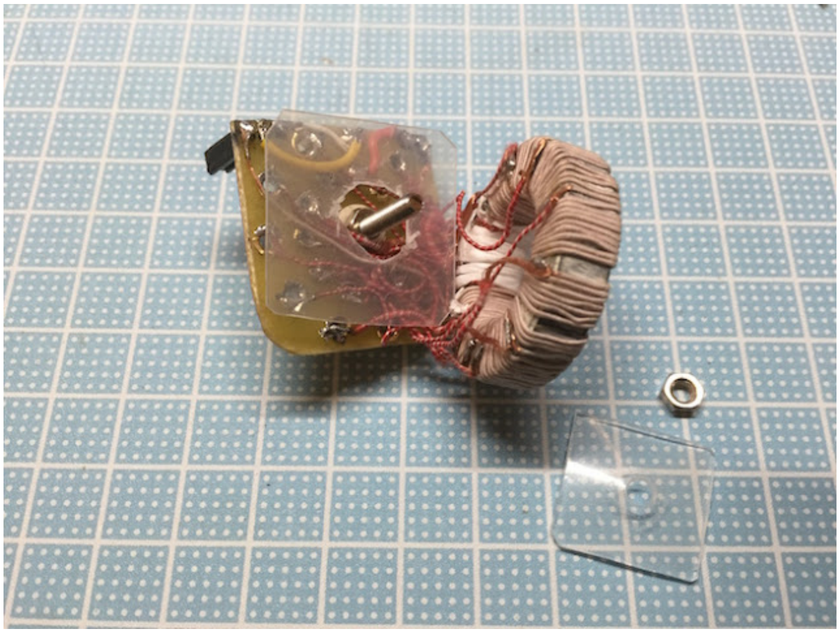

[https://4.bp.blogspot.com/-rtnqHpFmeNk/VvsbCFnibKI/AAAAAAAACdk/9BrZxEvZbj0VhXdgUbmQ9RKQMwN_gDBZQ/s1600/IMG_6880.JPG]

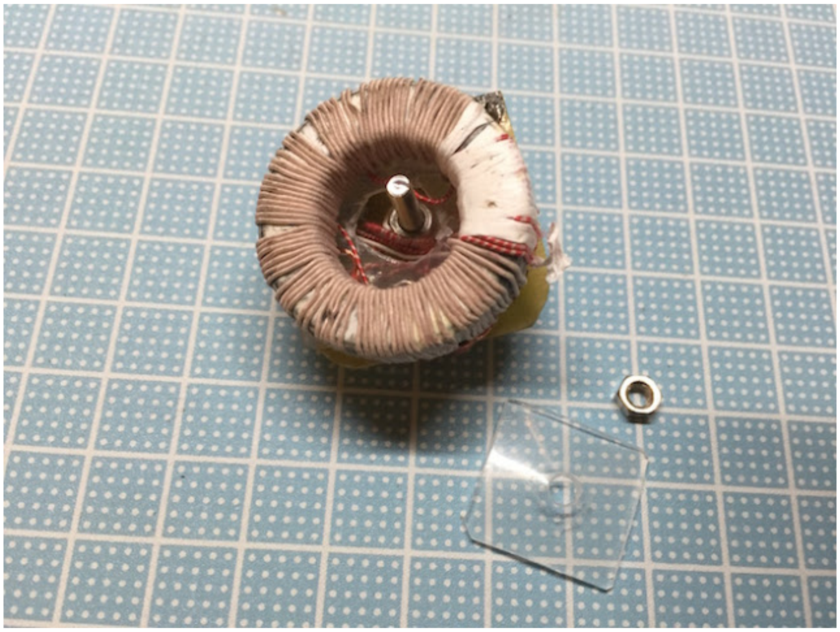

[https://1.bp.blogspot.com/-PUnPk0wNZGI/VvsbCOqjBgI/AAAAAAAACdo/wTw4Xw7K8Ks0o-wGTu20Z0O0OU6lRC3zw/s1600/IMG_6881.JPG]

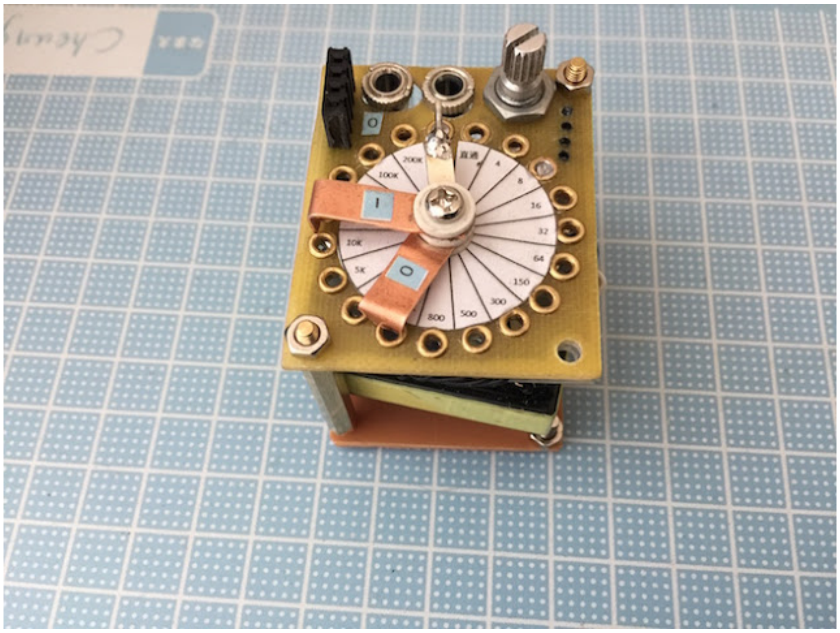
O
I
O
200K
100K
10K
5K
800
500
300
150
64
32
16
8
4

[https://3.bp.blogspot.com/-q18yJsBprJg/VwIEjAPaWwI/AAAAAAAACfk/hIA9xqefWokqkGxxDtP7ve8X5s9TVIq2Q/s1600/IMG_6956.JPG]

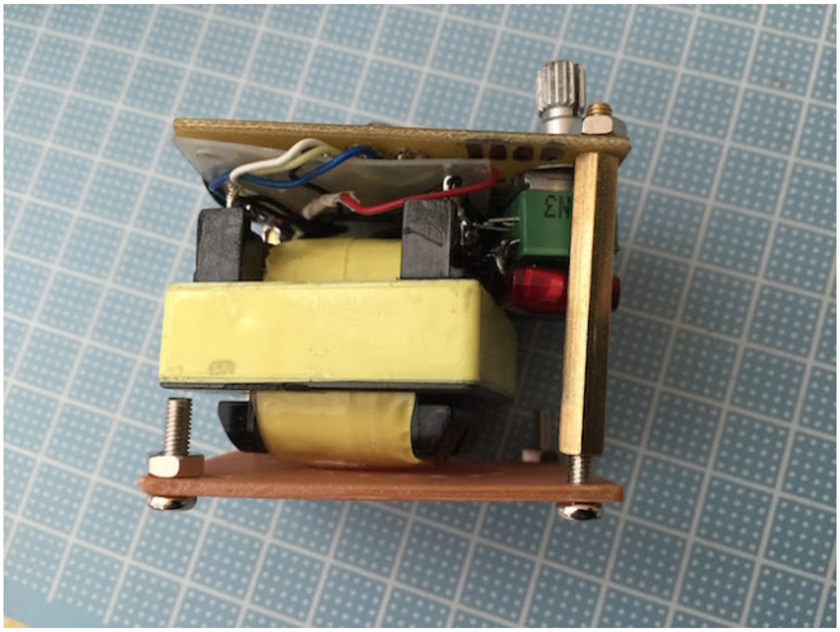

[https://3.bp.blogspot.com/-31OOPeueOAg/VwIEjB0_GII/AAAAAAAACfg/ViMhFHz3jawB_V0rln6g_6jkPO2I8oSrg/s1600/IMG_6957.JPG]

Dynamic Views theme. Powered by Blogger.

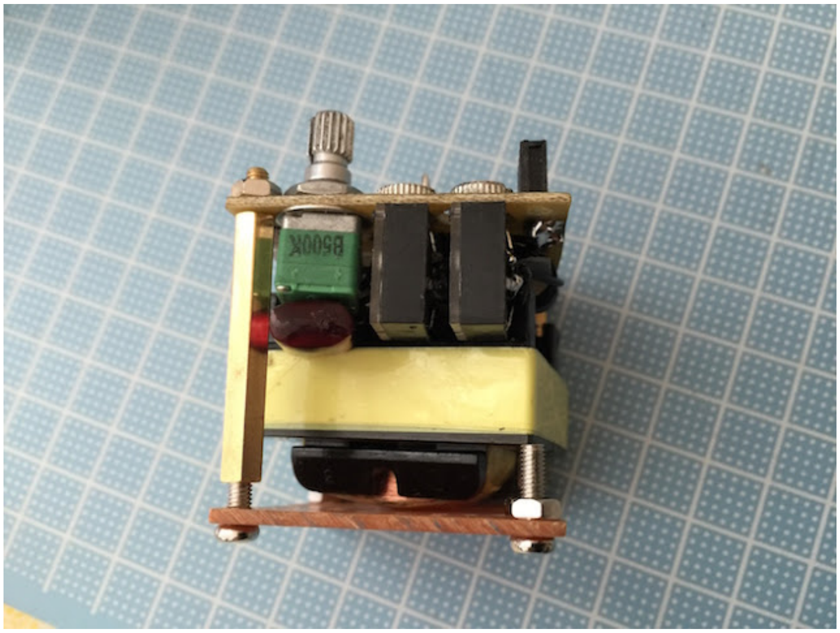

[https://3.bp.blogspot.com/-6jcl4Lsetjw/VwIEjB-E5QI/AAAAAAAACfc/yp3Q3jCgifgZ9y5vQYexqRFivJRM3QWCA/s1600/IMG_6958.JPG]

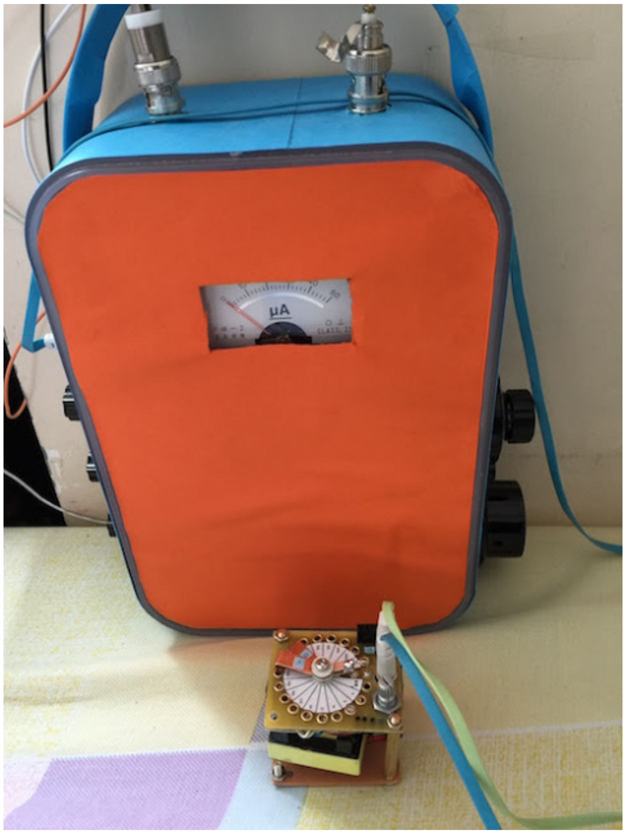

[https://1.bp.blogspot.com/-uB0aXsjVLNI/VwIEjoR1RLI/AAAAAAAACfo/fskpzbUsp2c2V7-z1wtT3QD8erArhuEYw/s1600/IMG_6959.JPG]

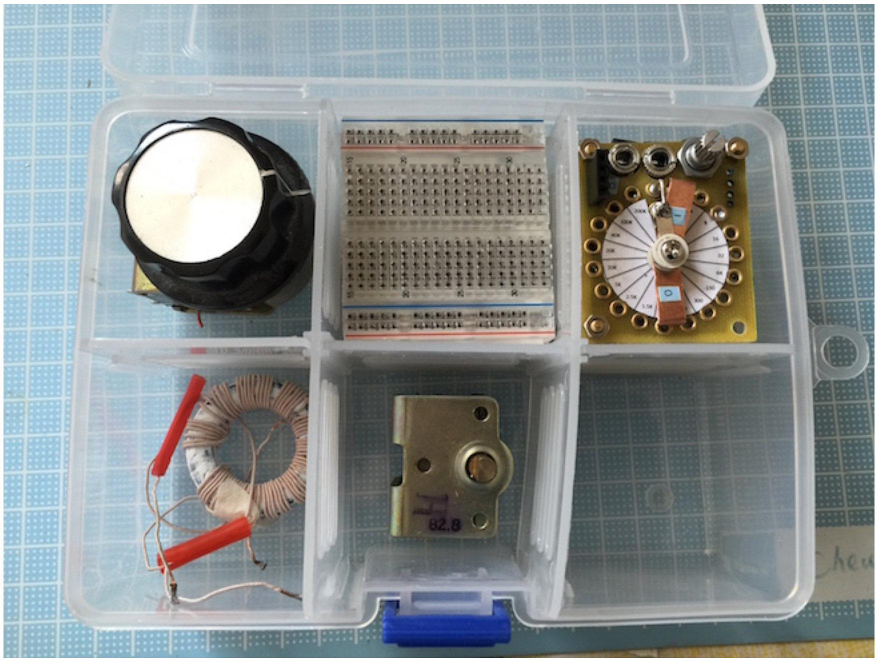

[https://1.bp.blogspot.com/-6iG1_zBvf6U/VwIEj2XHpGI/AAAAAAAACfs/Y9Qbur2-8t8-VBHbPk1KKS4zELg-rGe6w/s1600/IMG_6960.JPG]

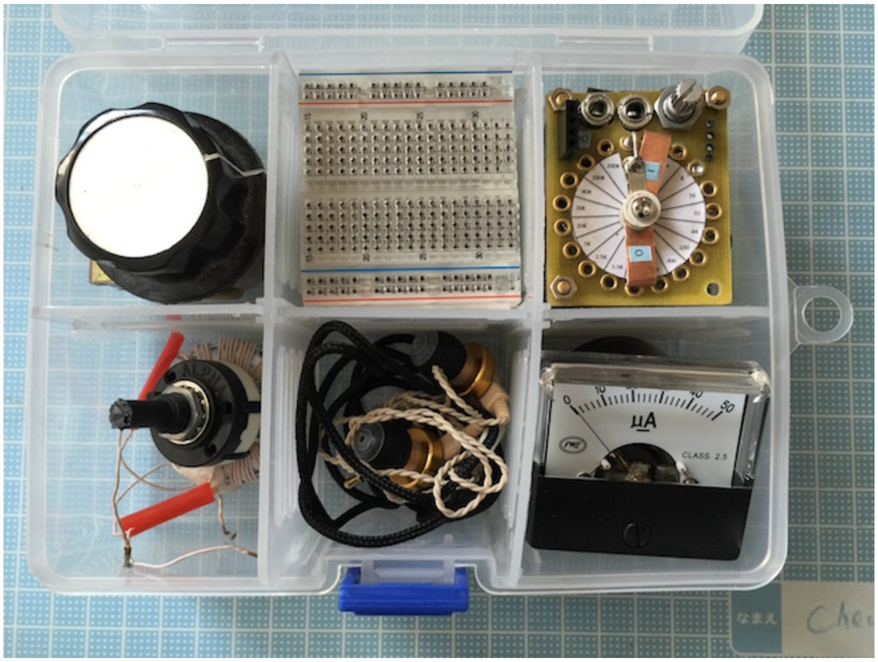

[https://4.bp.blogspot.com/-nMM7ELnyBVo/VwIEj9eDdZI/AAAAAAAACfw/vMPKwtSCrgcr8_kgEdd7PinpPOlAOCAXw/s1600/IMG_6961.JPG]

Dynamic Views theme. Powered by Blogger.

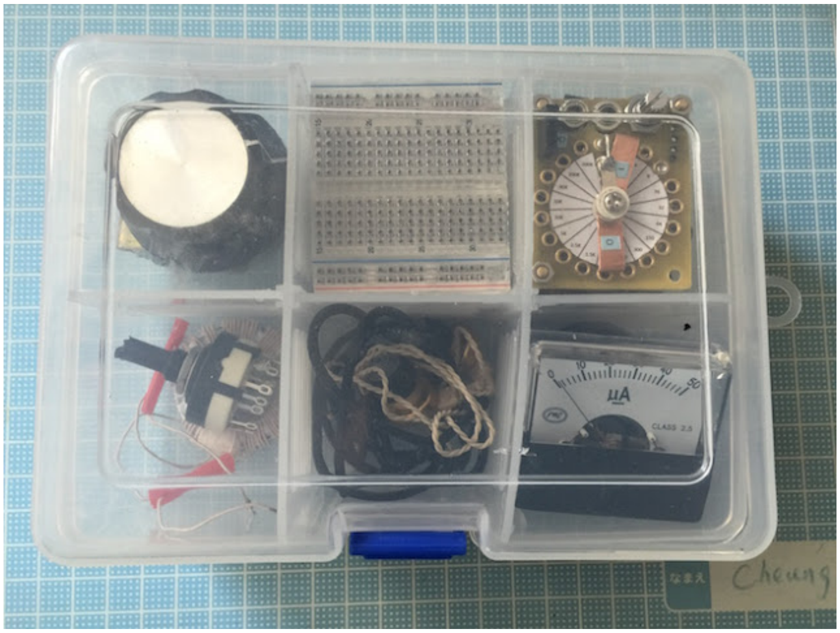

[https://1.bp.blogspot.com/-zSJeRtjHIWA/VwIEkOKOg5I/AAAAAAAACf0/lUOB8Bklp8E_34GbRIuPCTtqO_q671yAw/s1600/IMG_6962.JPG]

Posted 22nd March 2016 by Billy

0 Add a comment

17th March 2016

Vietnam Crystal Radio Test

I traveled to Vietnam, Ho Chi Minh City on the week of Mar 14 2016.

My hotel room is on 2/f, 3.5km from the broadcasting station for the voice of Vietnam (local state owned MW radio station) . Since the window is sealed, I have to rely on my umbrella loop antenna (0.8 meter diameter). I measured a dc current of 65uA from my Crystal set, so loud that I can plug in an 8 ohm headphone and still got reasonable loudness.

I also ran a series of tests using my tiny mobile crystal radio laboratory.

I brought my FM Crystal Set with the 1.2m extensible antenna too.
I could receive one station but weak and barely understood the broadcast.

Wearing my jacket crystal radio with a 400mm x 600mm loop antenna sewed into the back of the jacket, i could listen to the radios from within a bus traveling on the street using a 3SK143-Q crystal radio. You can see that at the end of my video.

I also visited the museum and saw some radio communication equipment used during the wars.

Purchased some nice wooden boxes that can be used to build some nice crystal sets. only USD 3.5 each.

I noticed that there are lots of antenna towers in Ho Chi Minh City. May be because the buildings are not tall enough, towers are needed to pull up the antenna to the right height. I guess most of these are for 3G mobile phone communication. I did see one super tall antenna that seems to match with my info about the location of the Radio transmission stations for the AM and FM radios.

Here are some pictures.

And the video of my crystal radio tests.
https://youtu.be/LVl63edu83Y [https://youtu.be/LVl63edu83Y]

Dynamic Views theme. Powered by Blogger.

[https://1.bp.blogspot.com/-Pke9jWiYe3M/Vuua9BVTiIl/AAAAAAAACWE/SZADzRHmynoztxLOkY1X_qcOk04U8L9FQ/s1600/IMG_6059.JPG]

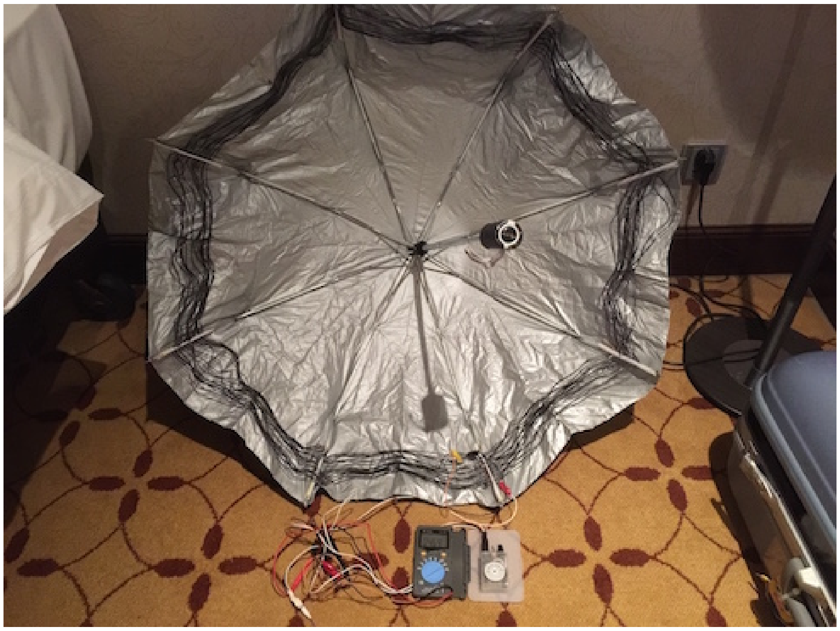

[https://1.bp.blogspot.com/--oXTH2h_Mls/Vuua9ivShBI/AAAAAAAACWM/eKZKNb5oaj8PXj1floAp4tKmpwByJNxeQ/s1600/IMG_6060.JPG]

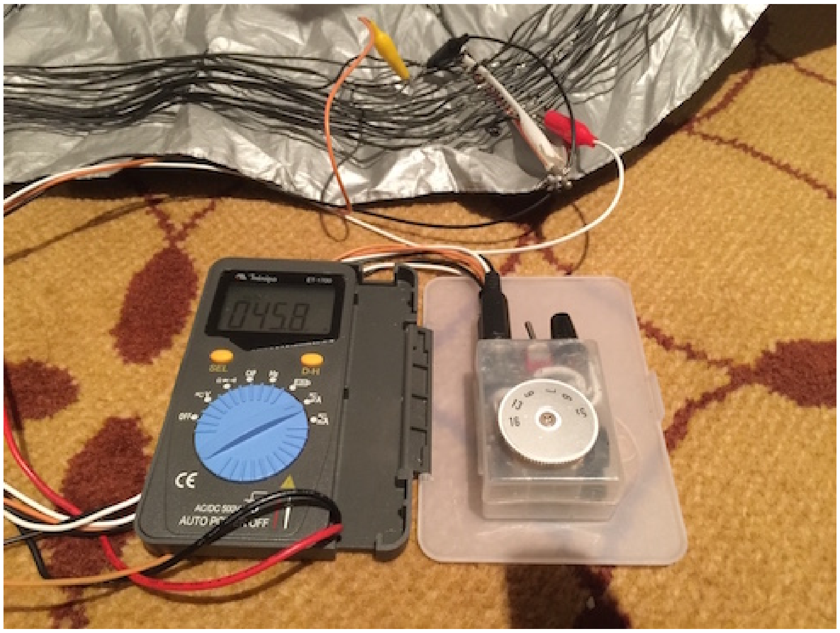

[https://2.bp.blogspot.com/-DOSuH9PU1dU/Vuua9xYD5II/AAAAAAAACWQ/lGfQ7PPGNzAWQD1bboqfExbRdPlurtXEw/s1600/IMG_6061.JPG]

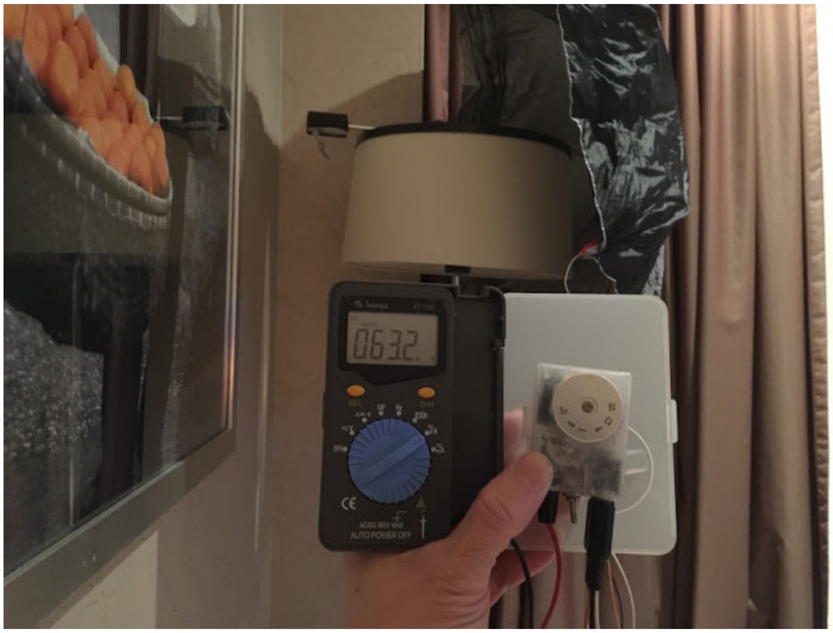
0632
SEL
D-H
OFF
CE
AC/DC 500V MAX
AUTO POWER OFF

[https://2.bp.blogspot.com/-4tEbT7P47zY/Vuua9wO7m7I/AAAAAAAACWY/GM84aoPbFGE7jq-jao-JxTze4cWCxp2wA/s1600/IMG_6062.JPG]

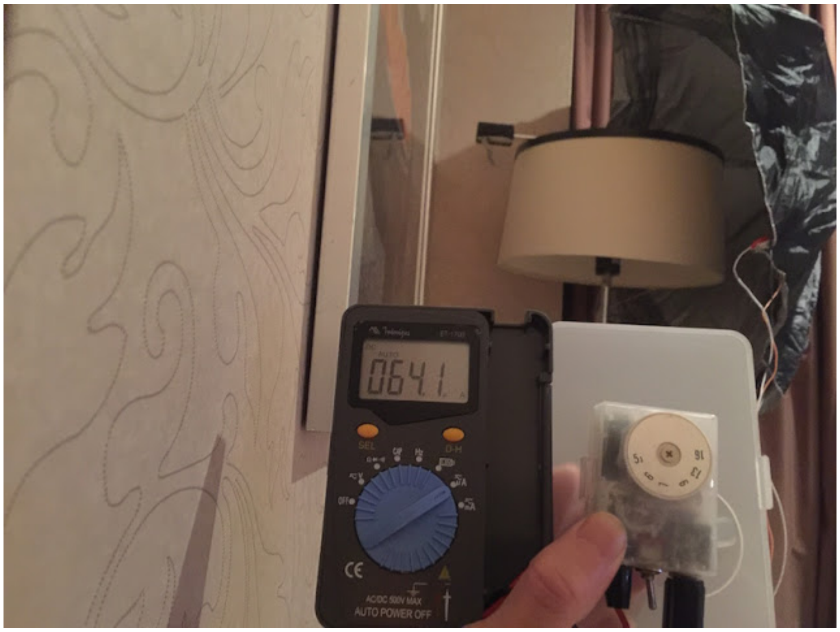
0641
SEL
D-H
OFF
CE
AC/DC 500V MAX
AUTO POWER OFF

[https://1.bp.blogspot.com/--63WFsdZkBE/Vuua-DVzNmI/AAAAAAAACWU/iRG_VdHjUPgwi01GFzEY3BI-LUUB0x11A/s1600/IMG_6064.JPG]

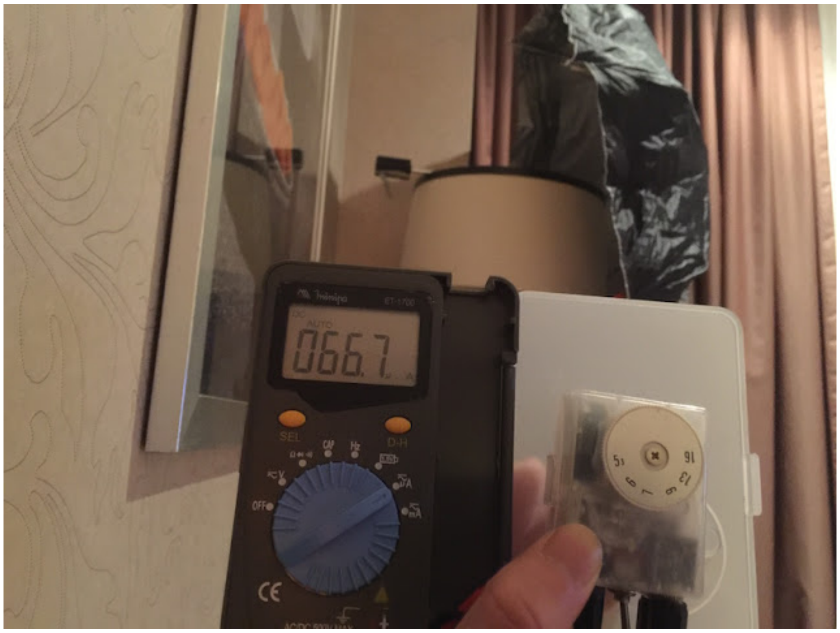

[https://4.bp.blogspot.com/-9MYhp8vmz_I/Vuua_nWuNGI/AAAAAAAACWo/O5Bl_p8cgq0LBG3i5Hzujmy375W4ozAUQ/s1600/IMG_6065.JPG]

[https://2.bp.blogspot.com/-9CExxwOXThU/Vuua-8aC1vI/AAAAAAAACWc/_2qKDl3b5vcQUpxqWKFdgb8SwBFggRorA/s1600/IMG_6066.JPG]

[https://4.bp.blogspot.com/-qoyLFGNFmXA/Vuua_LRwf1I/AAAAAAAACWg/L006McJWd6Qoo7gKi3I3MhY8SLx1n_UZA/s1600/IMG_6070.JPG]

[https://3.bp.blogspot.com/-hg9Rra7pAL8/Vuua_RIEYjI/AAAAAAAACWk/3n1Z9pYMoisGqB3zu8HicbqL8POG4OyHw/s1600/IMG_6071.JPG]

[https://1.bp.blogspot.com/-utXmP1BmFrg/Vuua_vnC5DI/AAAAAAAACWs/G1lmqhx-g8EN9xwd8tGywPUPL31KmiJYA/s1600/IMG_6072.JPG]

[https://3.bp.blogspot.com/-EXmYewEtWso/Vuua_7o5HZI/AAAAAAAACWw/-Z1mdbz0_7EwQdzf6U610sUSoNzpj7fjQ/s1600/IMG_6073.JPG]

Dynamic Views theme. Powered by Blogger.

[https://1.bp.blogspot.com/-QMOT8gTdo_U/VuubAKlKhJI/AAAAAAAACW0/O4gUjzOXNRUmYAdKhICRUjitD83JcyoHw/s1600/IMG_6074.JPG]

[https://4.bp.blogspot.com/-RgiIfFCorO8/VuubAG1E2AI/AAAAAAAACW4/461BoPLzQ58OxEOFd1Am2avmn-OBzlqyA/s1600/IMG_6075.JPG]

[https://1.bp.blogspot.com/-5HX-yqR4U9M/VuubAWwaszI/AAAAAAAACW8/noVn2ECbdtUW4amKIHHNCSwlX1TQP7u8A/s1600/IMG_6080.JPG]

[https://4.bp.blogspot.com/-JbpSOZ9flgg/VuubAtY_CxI/AAAAAAAACXA/AuReaAh3v2o2T-8Ec_seYBR5zRwTgyoLw/s1600/IMG_6081.JPG]

Dynamic Views theme. Powered by Blogger.

XIV
SẢN VIỆT NAM (3/2/1930 - 3/2/2016)
MỪNG XUÂN BÍNH THÂN - MỪNG ĐẢNG QUANG VINH

[https://4.bp.blogspot.com/-skCbBrJoCOU/VuubH1HwY_I/AAAAAAAACZQ/Pt_ZoFO7lq4FtqL9tlnrYIOTVOEej1HIQ/s1600/IMG_6672.JPG]

[https://2.bp.blogspot.com/-mCjvsuXQdVk/VuubAg1__VI/AAAAAAAACXE/z5rQNGm2sfUsUwXD70J_hqDGhcqCsSaXg/s1600/IMG_6115.JPG]

Dynamic Views theme. Powered by Blogger.

[https://3.bp.blogspot.com/-lWx6q2aK83Y/VuubA2RwolI/AAAAAAAACXI/om2VXk1jPxgp8cUda-XjVjJMzu_5lKYYg/s1600/IMG_6328.JPG]

[https://4.bp.blogspot.com/-7jwPB8Z964M/VuubBKL_BKI/AAAAAAAACXM/2avmzeW_UR4R1JSRca41Gzdrp5vNSEc8A/s1600/IMG_6329.JPG]

[https://3.bp.blogspot.com/-8BI2Q5TbPJs/VuubBKUc1UI/AAAAAAAACXQ/SzFZRBNwzLYQCYG0Ux5y_Bnoqnpt5T4gw/s1600/IMG_6330.JPG]

[https://2.bp.blogspot.com/-7s_f4teZ_eU/VuubBR5J0oI/AAAAAAAACXU/fqTNkoRpVms1uEJ5avaK2yASjGrbVMGAA/s1600/IMG_6331.JPG]

Dynamic Views theme. Powered by Blogger.

[https://3.bp.blogspot.com/-7gTicZP8Y1Q/VuubBzflIEI/AAAAAAAACXY/CC1uwx3dGfo0geScGnp6W0llsqaExqZJQ/s1600/IMG_6332.JPG]

[https://1.bp.blogspot.com/-QbzW8F6BPJo/VuubB8L9kSI/AAAAAAAACXc/LXFAR3S1Hjw1C0f1EaM_Gj-GXAV-JUp7A/s1600/IMG_6333.JPG]

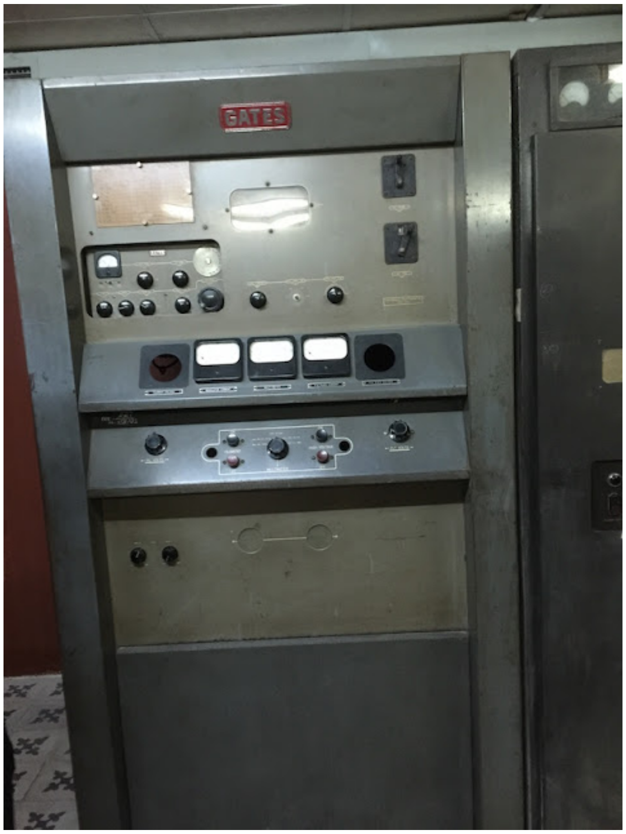

[https://4.bp.blogspot.com/-OiWhmHeXHwM/VuubByYTWOI/AAAAAAAACXg/81YE3B711xAOJI2jc9p-5Qd2J0EMNrdKw/s1600/IMG_6334.JPG]

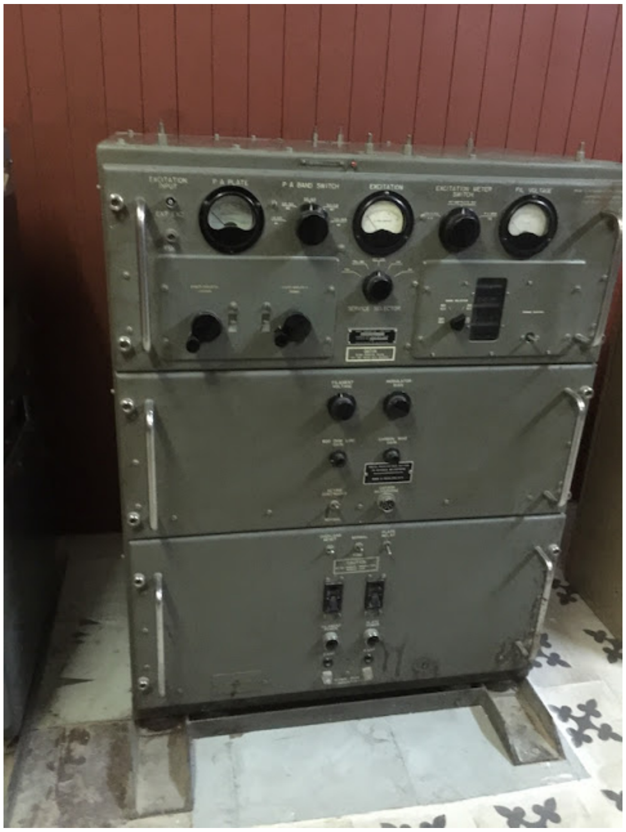

[https://4.bp.blogspot.com/-c3Uibq0-cjl/VuubCZEMkrI/AAAAAAAACXk/cOkhi4yFT3ob1tr7gE72BtHxeJCCYmjSw/s1600/IMG_6335.JPG]

[https://4.bp.blogspot.com/-gtAaPATjgCc/VuubCUp35yI/AAAAAAAACXo/-H-gU511CnIOtp4muRW7J2NM0e9FKLLxA/s1600/IMG_6337.JPG]

[https://2.bp.blogspot.com/-eTzYjxs-_68/VuubCkB5TrI/AAAAAAAACXs/7lmDR3a_KtwXgIxQ2apam3UrJqMYVMo1A/s1600/IMG_6338.JPG]

Dynamic Views theme. Powered by Blogger.

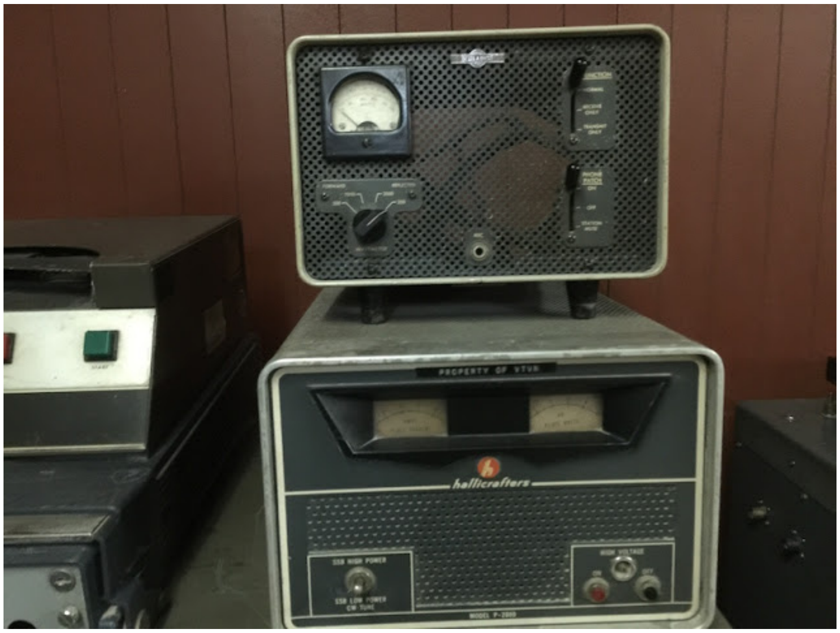
PROPERTY OF VTUR
hallicrafters
SSB HIGH POWER
SSB LOW POWER CW TUNE
HIGH VOLTAGE
ON
OFF
MODEL P-2000

[https://1.bp.blogspot.com/-Pq7SvHqHZI8/VuubC8nywyI/AAAAAAAACXw/Hf8VxBbr_ZErIP9ImTKjgVGMMBK9grRkQ/s1600/IMG_6339.JPG]

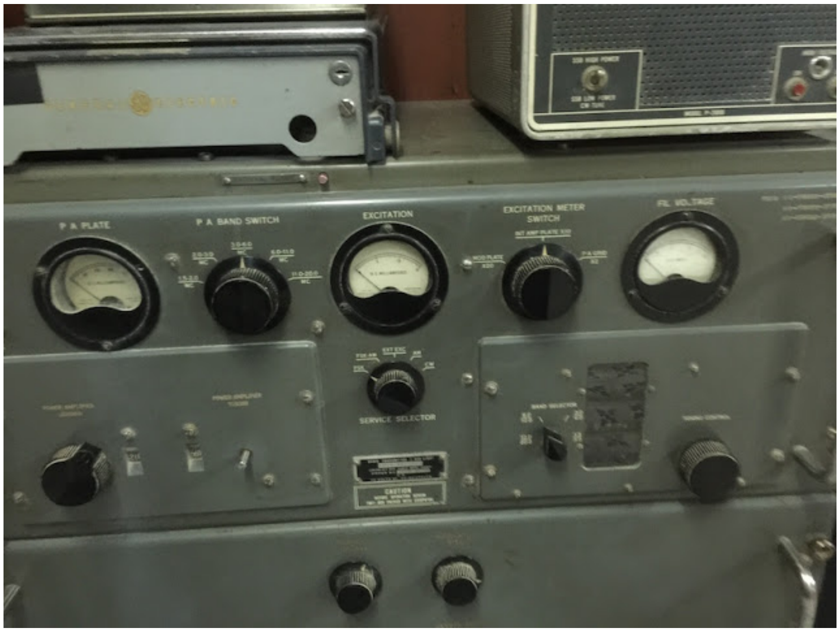
P A PLATE
P A BAND SWITCH
EXCITATION
EXCITATION METER SWITCH
FIL. VOLTAGE
SERVICE SELECTOR
CAUTION
MODEL P-2000

[https://2.bp.blogspot.com/-EZCqtUsmSeQ/VuubDLU4UMI/AAAAAAAACX0/fwVlLZkSJYYK4rJTt0skTY3MZgMBiGTrw/s1600/IMG_6340.JPG]

Dynamic Views theme. Powered by Blogger.

[https://1.bp.blogspot.com/-VeND07kXWwA/VuubDCE-KXI/AAAAAAAACX4/1hf9kx-XbcwqKRwRiMSmn-Rdi-mJYRL1Q/s1600/IMG_6341.JPG]

[https://1.bp.blogspot.com/-wCRylCnxcHE/VuubDRwDS8I/AAAAAAAACX8/Tu-vDFobBQcsI8mAPII1WHScwixIZ2r4A/s1600/IMG_6342.JPG]

[https://4.bp.blogspot.com/-8JACs8CBuPM/VuubDl8Gk1I/AAAAAAAACYA/wKDwIuL6ApUkxQsMGy4L8m9irLG7bzZ_Q/s1600/IMG_6343.JPG]

[https://3.bp.blogspot.com/-HGjSX1mp5qo/VuubDuvfQNI/AAAAAAAACYE/_eh-cuHa9fEAbpcPQ-XnpY-hTHfErwSYw/s1600/IMG_6345.JPG]

Dynamic Views theme. Powered by Blogger.

[https://1.bp.blogspot.com/-9dOx6IMX3V0/VuubD4S-ZmI/AAAAAAAACYI/5ZwPe7PS_NIyH3ui6hhroPx4NCOkxVcYQ/s1600/IMG_6356.JPG]

[https://4.bp.blogspot.com/-MID48pyy9_g/VuubD_iK1wI/AAAAAAAACYM/nv-dEIgHhmIDxBVCq2tU5VmbGivreGnxQ/s1600/IMG_6357.JPG]

Dynamic Views theme. Powered by Blogger.

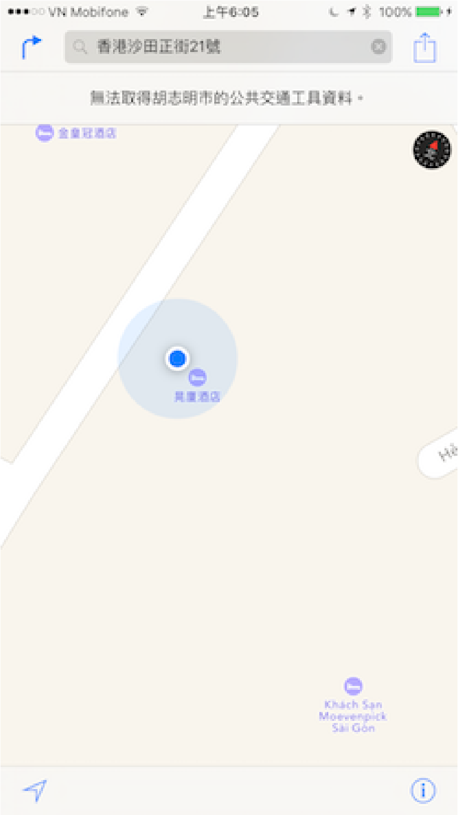
●●●○○ VN Mobifone
上午6:05
100%
香港沙田正街21號
金皇冠酒店
Hè
Khách Sạn Moevenpick Sài Gòn

無法取得胡志明市的公共交通工具資料。

[https://3.bp.blogspot.com/-aOBNx1ETc5U/VuubELhBj3I/AAAAAAAACYQ/Qu7TUWnnH-8pFSd4Rq7KxXUCxvRBhtZNg/s1600/IMG_6370.PNG]

安順
昆明
柳州
南宁市
河内
北海
海防市
寮國
沙耶武里
永珍
泰國
曼谷
柬埔寨
金邊
越南
順化
峴港
芽庄
胡志明市
芹苴市

[https://4.bp.blogspot.com/-TrLRb4ZbGRw/VuubEdHljTI/AAAAAAAACYU/6byyw-WfESQOtaxF-ss2douqaDMEudkYQ/s1600/IMG_6371.PNG]

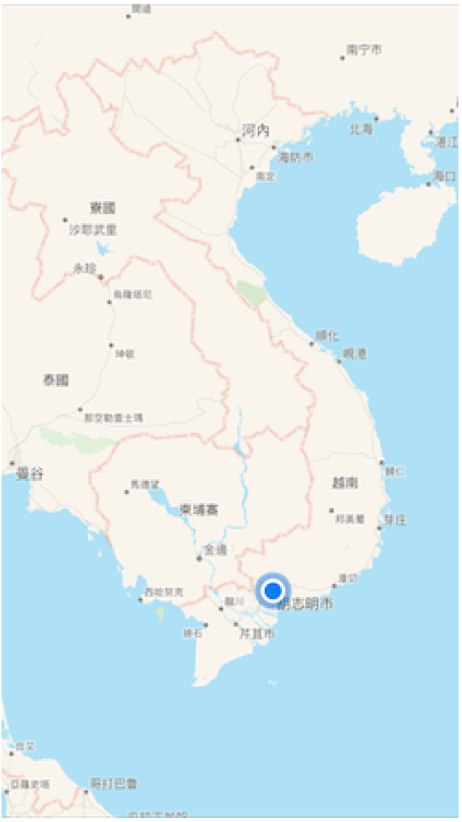
南宁市
河内
北海
湛江
海防市
海口
寮國
永珍
泰國
曼谷
柬埔寨
金邊
順化
峴港
越南
芽庄
胡志明市
芹苴市
哥打巴魯

[https://2.bp.blogspot.com/-HGUs7elehOs/VuubEt1UY1I/AAAAAAAACYY/5r0dIBpkY1gp0O-kFAwLw_1CQFsUInzpw/s1600/IMG_6372.PNG]

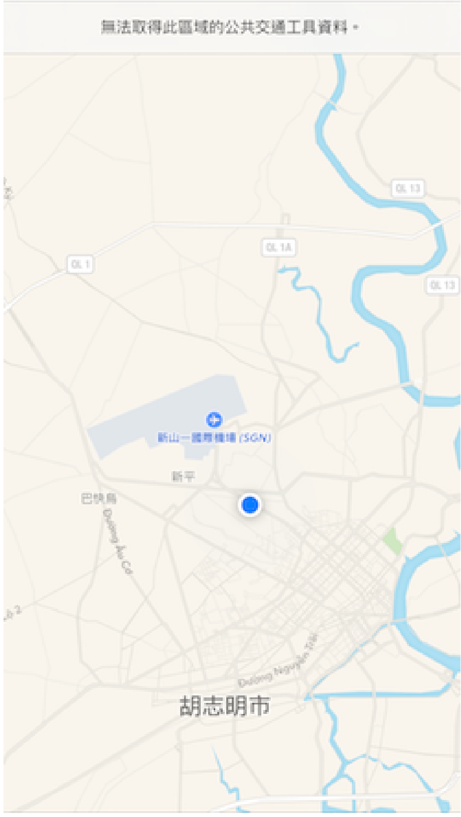
無法取得此區域的公共交通工具資料。

QL 13
QL 1A
QL 1
QL 13
新山一國際機場 (SGN)
新平
Đường Âu Cơ
Đường Nguyễn Trãi
胡志明市

[https://1.bp.blogspot.com/-D36aGyl5sKg/VuubEso1RhI/AAAAAAAACYc/BiSSvHSkm9wH-7zdU0_s4h5Y5I11XJS3w/s1600/IMG_6373.PNG]

Dynamic Views theme. Powered by Blogger.

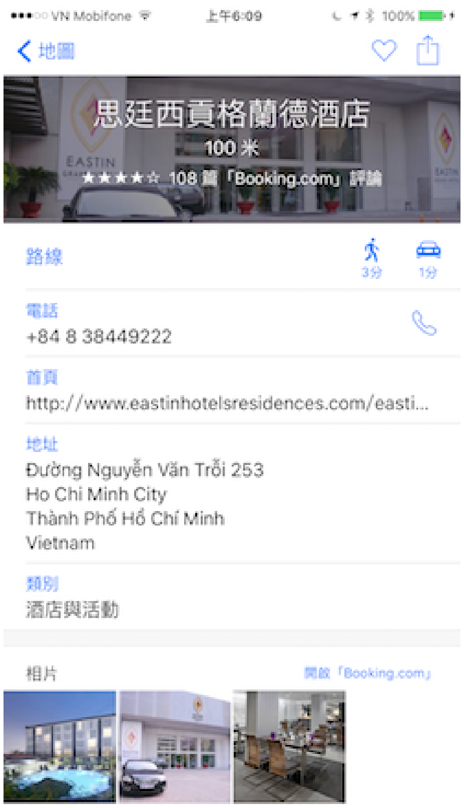

●●●○○ VN Mobifone
上午6:09
100%
地圖

思廷西貢格蘭德酒店
100 米
★★★★☆ 108 篇「Booking.com」評論

路線
3分
1分

電話
+84 8 38449222

首頁
http://www.eastinhotelsresidences.com/easti...

地址
Đường Nguyễn Văn Trỗi 253
Ho Chi Minh City
Thành Phố Hồ Chí Minh
Vietnam

類別
酒店與活動

相片
開啟「Booking.com」

[https://4.bp.blogspot.com/-0dmOtzwHb5Q/VuubEza55XI/AAAAAAAACYg/szTLzpi5IModiCC3uh9xWTaZX1S_5aKuA/s1600/IMG_6374.PNG]

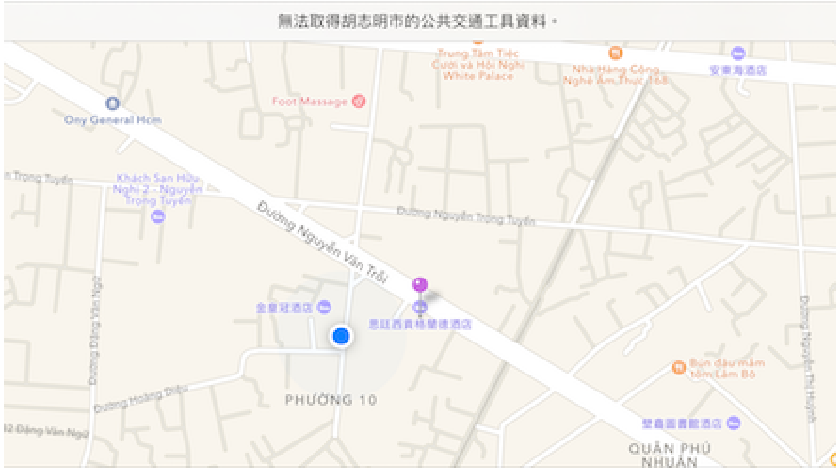

無法取得胡志明市的公共交通工具資料。

Trung Tâm Tiệc Cưới và Hội Nghị White Palace
Nhà Hàng Công Nghệ Ẩm Thực 168
Foot Massage
Ony General Hcm
Trọng Tuyển
Khách Sạn Hữu Nghị 2 - Nguyễn Trọng Tuyển
Đường Nguyễn Văn Trỗi
Đường Nguyễn Trọng Tuyển
思廷西貢格蘭德酒店
Đường Đặng Văn Ngữ
Đường Hoàng Diệu
PHƯỜNG 10
32 Đặng Văn Ngữ
Đường Nguyễn Thị Huỳnh
QUẬN PHÚ NHUẬN

[https://3.bp.blogspot.com/-Dyz7pKzDwRo/VuubFJT-6SI/AAAAAAAACYk/YMKM4-aRoFQz4obw1EFQvZsW14OnWhohQ/s1600/IMG_6375.PNG]

Dynamic Views theme. Powered by Blogger.

無法取得胡志明市的公共交通工具資料。

新山一國際機場 (SGN)
新平
QL 13
QL 52
TL 258

[https://1.bp.blogspot.com/-5XYMjzu8wQQ/VuubFMViIJI/AAAAAAAACYo/ljrfDvY0nQc-eFFWKxJ2NZgC_Y-qrRvnw/s1600/IMG_6376.PNG]

無法取得胡志明市的公共交通工具資料。

Sân Vận Động Quân Khu 7
Big C
Đường Hoàng Văn Thụ
Trung Tâm Tiệc Cưới và Hội Nghị White Palace
Đường Trần Khắc Chân
PHƯỜNG 3
Trường Đại học Tài chính Marketing
Đường Nguyễn Trọng Tuyển
Đường Nguyễn Văn Trỗi
Đường Tân Canh
Đường Đặng Văn Ngữ
PHƯỜNG 10
PHƯỜNG 15
QUẬN PHÚ NHUẬN
PHƯỜNG 1
Siêu Thị Win Mart - 63 Nguyễn Đình Chính

[https://3.bp.blogspot.com/-tCl9Ye0Wya8/VuubFfmW4AI/AAAAAAAACYs/LL7KSlb8JvEAHHbIK3SRJ-gBMJHpGsUCg/s1600/IMG_6377.PNG]

吉大港
米央瑪
南寧市
廣州
香港
河內
北海
湛江
海口
寮國
永珍
仰光
泰國
順化
峴港
曼谷
越南
柬埔寨
金邊
胡志明市

[https://4.bp.blogspot.com/-6RoZ16cNrho/VuubFn2GxhI/AAAAAAAACYw/AFfXcnelEFEZWfX875nuVJCwtUS-kg1xw/s1600/IMG_6378.PNG]

Dynamic Views theme. Powered by Blogger.

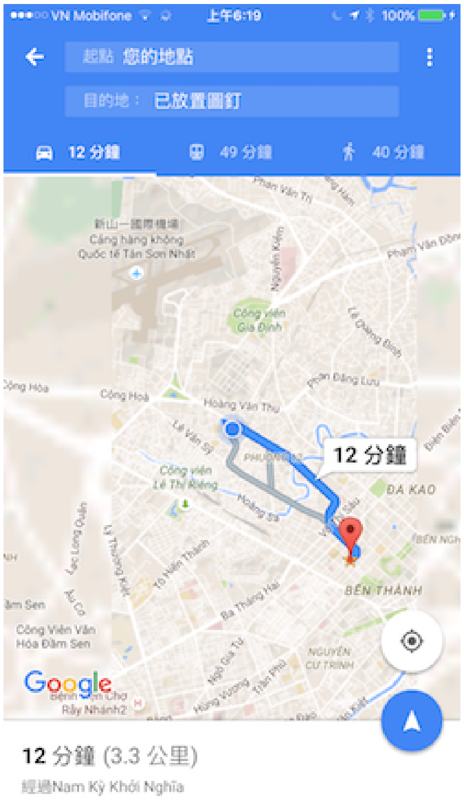
VN Mobifone 上午6:19 100%
起點 您的地點
目的地： 已放置圖釘
12 分鐘
49 分鐘
40 分鐘
Cảng hàng không Quốc tế Tân Sơn Nhất
Công viên Gia Định
Cộng Hoà
Hoàng Văn Thụ
Công viên Lê Thị Riêng
12 分鐘
ĐA KAO
BẾN THÀNH
Công Viên Văn Hóa Đầm Sen
NGUYỄN CƯ TRINH
Google
12 分鐘 (3.3 公里)
經過Nam Kỳ Khởi Nghĩa

[https://2.bp.blogspot.com/-5tkqWYZiHg8/VuubFo9Tf4I/AAAAAAAACY0/QVF21201fdATUC3D0sllkGomYa2WcS9BQ/s1600/IMG_6379.PNG]

[https://4.bp.blogspot.com/-fDpZ7FauhmE/VuubFyjcxaI/AAAAAAAACY4/m3BiGrwm5qoax3XjzFGxtuOGfUXmmDhBw/s1600/IMG_6533.JPG]

[https://2.bp.blogspot.com/-x21D3KpVCB8/VuubGOapfnI/AAAAAAAACZA/pA440JzHR0cdz_MY1Dklglx_Bwa29Tnfg/s1600/IMG_6535.JPG]

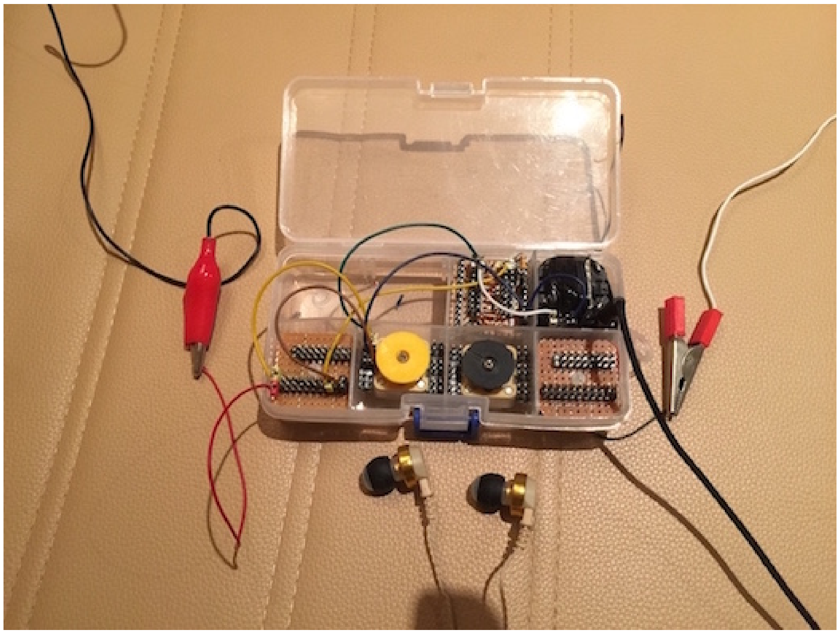

[https://3.bp.blogspot.com/-w7FWy-O-Og0/VuubGbeu_qI/AAAAAAAACY8/PKHGI0XjWWk0AmJ-6_pqz-vAsggkyT1eQ/s1600/IMG_6536.JPG]

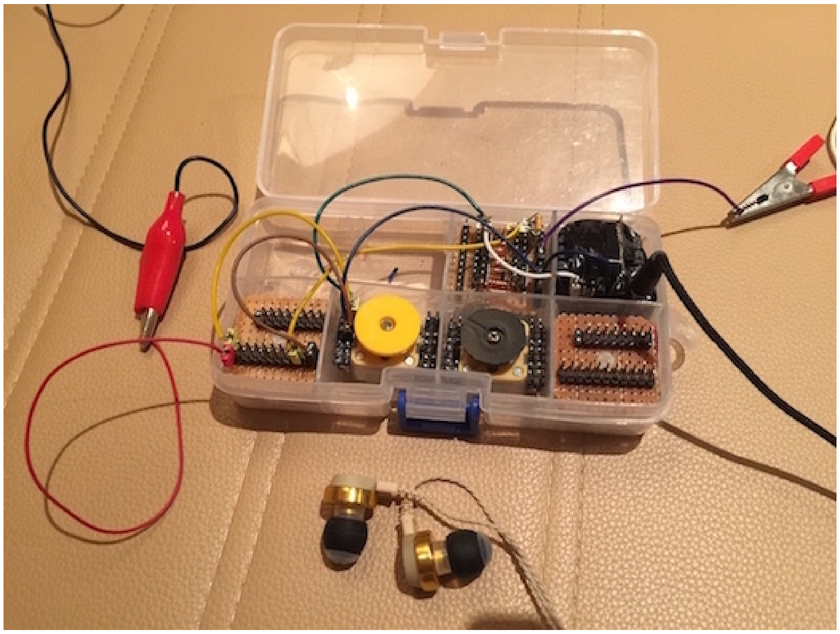

[https://3.bp.blogspot.com/-Ax5c3Ibzq_I/VuubHQkKANI/AAAAAAAACZE/T4f1VQwt2s4DQFvqpNPvwetSFTchrKnhA/s1600/IMG_6537.JPG]

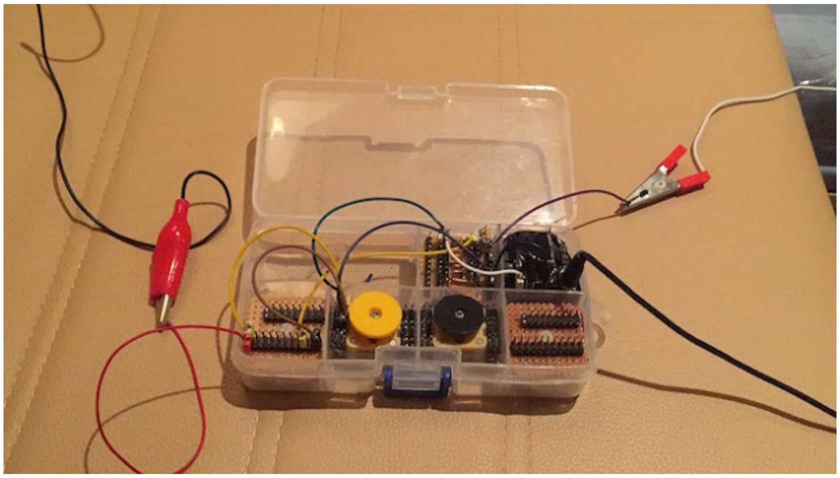

[https://1.bp.blogspot.com/-myqhbSJ_AtY/VuubHQCl8_I/AAAAAAAACZI/Q7gLhSFJX6EIP3rRXO3pNaqCFSSbfiLcA/s1600/IMG_6538.JPG]

Dynamic Views theme. Powered by Blogger.

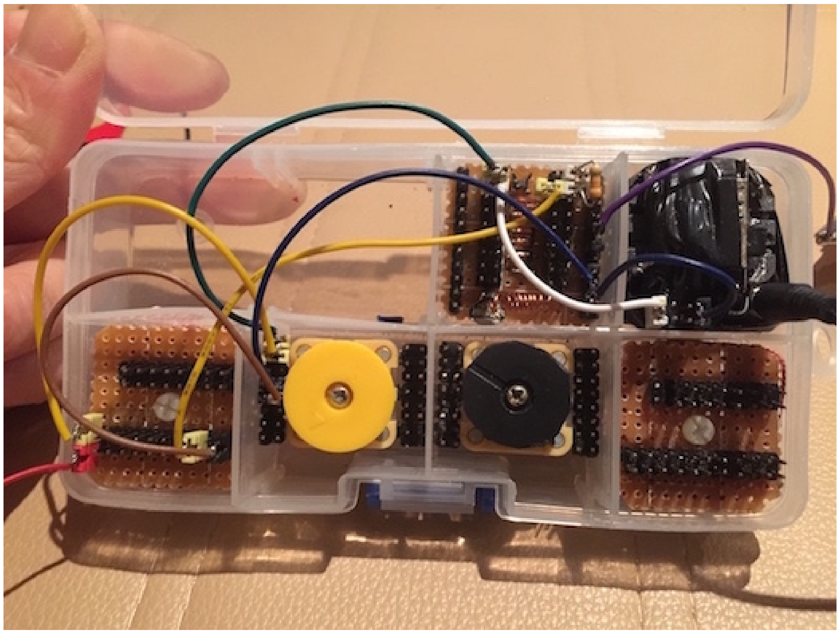

[https://3.bp.blogspot.com/-MuRRsjP2qgQ/VuubHvafbSI/AAAAAAAACZM/_0JvyX1V_9k4M7CobfgOx08Aafc_BEt_Q/s1600/IMG_6541.JPG]

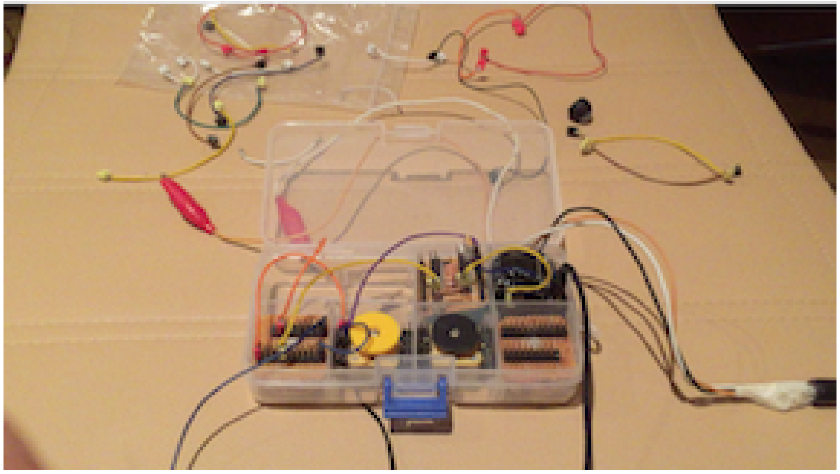

[https://2.bp.blogspot.com/-otUzUwUrPF4/VuubH4CV9wI/AAAAAAAACZU/L59Z89RdFzkKiHWrdjT-5pYyBDSmz7jzg/s1600/Screen%2BShot%2B2016-03-14%2Bat%2B9.36.50%2BPM.png]

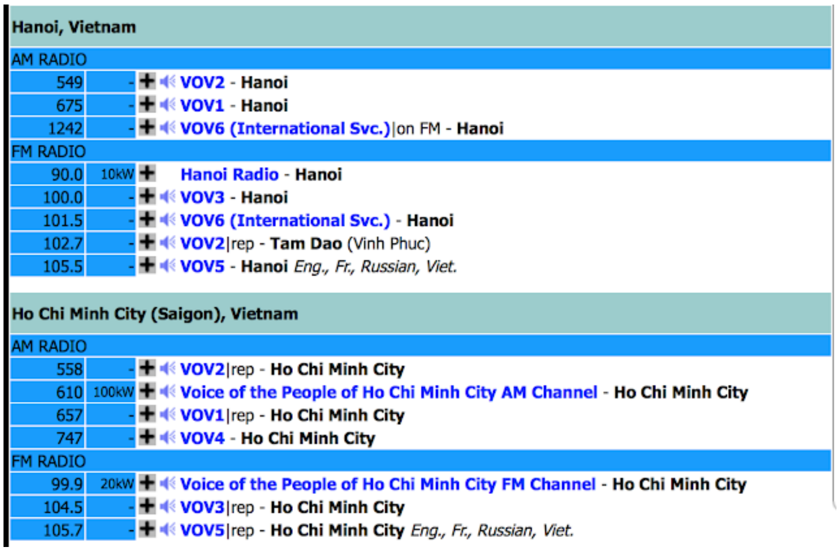

Hanoi, Vietnam			
AM RADIO			
549	-	+	**VOV2** - **Hanoi**
675	-	+	**VOV1** - **Hanoi**
1242	-	+	**VOV6 (International Svc.)**|on FM - **Hanoi**
FM RADIO			
90.0	10kW	+	**Hanoi Radio** - **Hanoi**
100.0	-	+	**VOV3** - **Hanoi**
101.5	-	+	**VOV6 (International Svc.)** - **Hanoi**
102.7	-	+	**VOV2**|rep - **Tam Dao** (Vinh Phuc)
105.5	-	+	**VOV5** - **Hanoi** *Eng., Fr., Russian, Viet.*
Ho Chi Minh City (Saigon), Vietnam			
AM RADIO			
558	-	+	**VOV2**|rep - **Ho Chi Minh City**
610	100kW	+	**Voice of the People of Ho Chi Minh City AM Channel** - **Ho Chi Minh City**
657	-	+	**VOV1**|rep - **Ho Chi Minh City**
747	-	+	**VOV4** - **Ho Chi Minh City**
FM RADIO			
99.9	20kW	+	**Voice of the People of Ho Chi Minh City FM Channel** - **Ho Chi Minh City**
104.5	-	+	**VOV3**|rep - **Ho Chi Minh City**
105.7	-	+	**VOV5**|rep - **Ho Chi Minh City** *Eng., Fr., Russian, Viet.*

[https://3.bp.blogspot.com/-WsYGgRksbE0/VuubIHgqKAI/AAAAAAAACZY/0gfLFse5oUQP3F4KHfzEwyqYaHewkhQng/s1600/Screen%2BShot%2B2016-03-18%2Bat%2B1.56.55%2BPM.png]

Posted 17th March 2016 by Billy

2 View comments

17th February 2016

3D Building Blocks for Crystal Radio

Create 3D building blocks for crystal radios so kids and adults can experiment with crystal radios. There is also a portable case for you to take what you have created out on the street to listen to radio using the big loop antenna on the case.

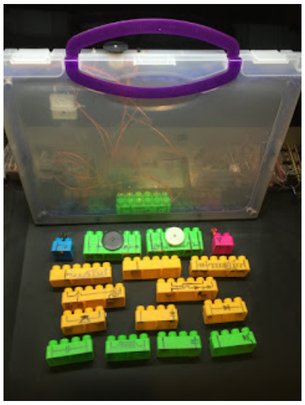

[https://1.bp.blogspot.com/-D-Q1YX12uxc/VsRVN_XRCdI/AAAAAAAACTw/brRaIZomEp4/s1600/IMG_5678.JPG]

Video link:
https://youtu.be/dbaRQoExeOs [https://youtu.be/dbaRQoExeOs]

Dynamic Views theme. Powered by Blogger.

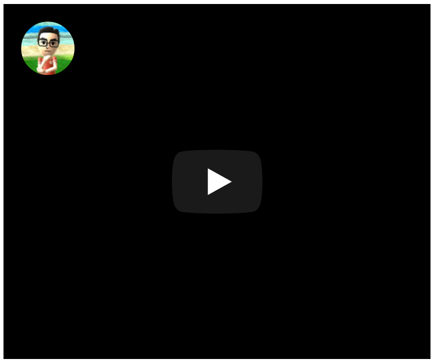

I have seen similar building blocks used to build lights, fans, and even robots, but not crystal radios. Though this is a low tech construction, I've spent weeks trying to figure out which building block is suitable (big enough to house the variable caps and ferrite sticks) and flexible with different length and height. Most importantly, how to make the connectors between the blocks so it's plug and play and yet the contacts are firm enough to be carried around.

I thought of using springs, but the extra turns for each connection may interfere with wireless signals. Finally settled with copper screws and elastic copper sheets that do the same magic as springs.

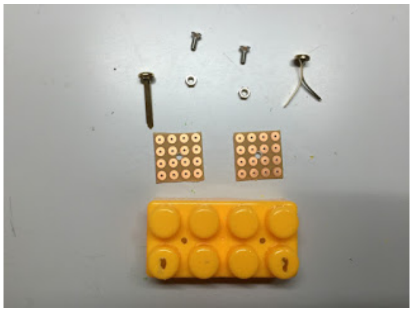

[https://3.bp.blogspot.com/-9JI_kYxNGhM/VsRVKZlXw4I/AAAAAAAACSc/QDm0na_VeeQ/s1600/IMG_5614.JPG]

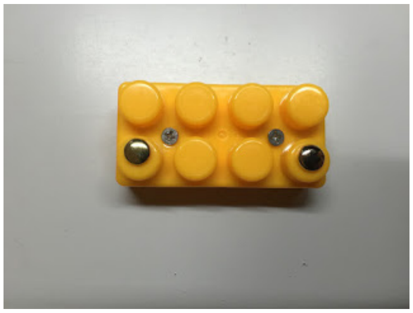

[https://1.bp.blogspot.com/-Bmom1Sj3SZE/VsRVJeCC34I/AAAAAAAACSM/gVARqJSsfl8/s1600/IMG_5615.JPG]

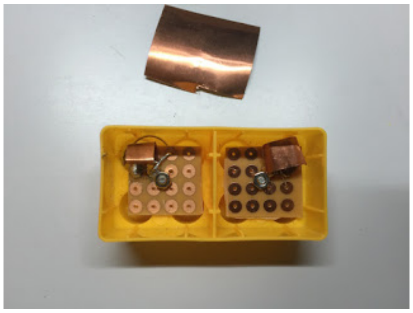

[https://3.bp.blogspot.com/-Mstz9EBpINo/VsRVJYJCxoI/AAAAAAAACSQ/izESt2crlrw/s1600/IMG_5617.JPG]

[https://2.bp.blogspot.com/--HLG9X07amc/VsRVJzFn-mI/AAAAAAAACSU/rhlu_uCZsoU/s1600/IMG_5618.JPG]

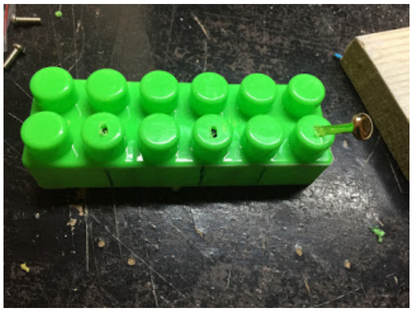

[https://3.bp.blogspot.com/-NVQYJ95Mdgw/VsRVJxMHRSI/AAAAAAAACSY/7Zv6_xUHOBk/s1600/IMG_5619.JPG]

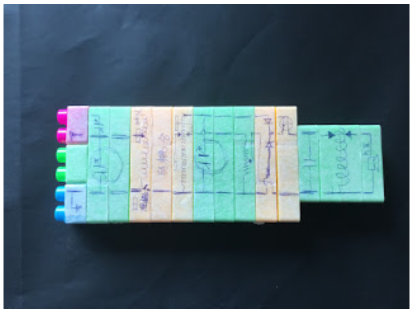

[https://3.bp.blogspot.com/-EscgUlalXbI/VsRVKeKC50I/AAAAAAAACSg/n5LPVoc-p2w/s1600/IMG_5620.JPG]

Dynamic Views theme. Powered by Blogger.

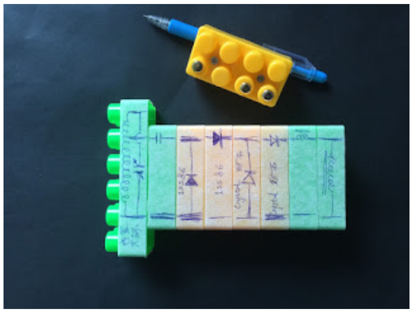

[https://3.bp.blogspot.com/-QZJsAAT9sPc/VsRVKbtPk8I/AAAAAAAACSk/6638hhD0M60/s1600/IMG_5621.JPG]

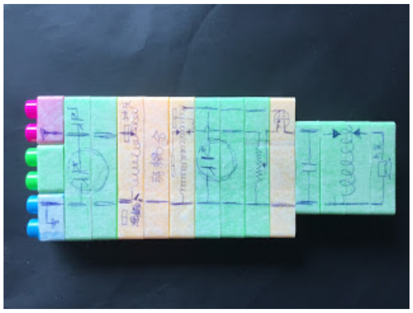

[https://4.bp.blogspot.com/-_CCnPHGNfnw/VsRVKwmrT6I/AAAAAAAACSo/_um4oIA6sp0/s1600/IMG_5622.JPG]

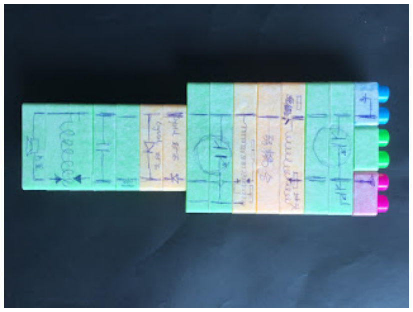

[https://4.bp.blogspot.com/-7uZRg6o8_C4/VsRVK35X8zI/AAAAAAAACSs/-bB7Mw_3f18/s1600/IMG_5623.JPG]

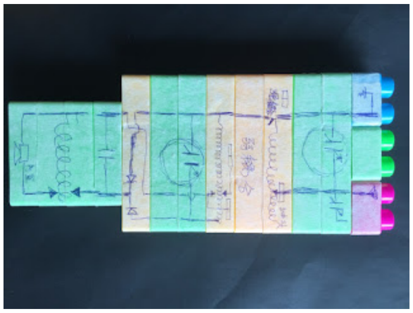

[https://3.bp.blogspot.com/-LbZB3jDAKYs/VsRVLC6rsiI/AAAAAAAACSw/b2h1yz1ruQA/s1600/IMG_5624.JPG]

[https://4.bp.blogspot.com/-uB7ZLlDE5B4/VsRVLXcYF_I/AAAAAAAACS0/4Lu-oOT5cic/s1600/IMG_5625.JPG]

[https://1.bp.blogspot.com/-U-dVXgjCNJs/VsRVLXg4dfI/AAAAAAAACS4/w3ktVYcMMj0/s1600/IMG_5648.JPG]

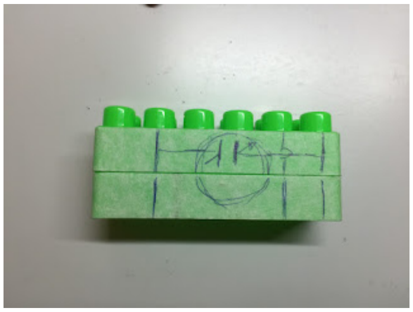

[https://4.bp.blogspot.com/-5wLY2DGNUDw/VsRVLtASJAI/AAAAAAAACS8/ve68nO-PjWY/s1600/IMG_5657.JPG]

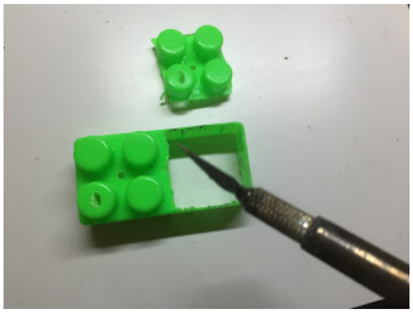

[https://1.bp.blogspot.com/-UMl1qU3z9F0/VsRVLre5HvI/AAAAAAAACTA/cC0KmK_YX-o/s1600/IMG_5658.JPG]

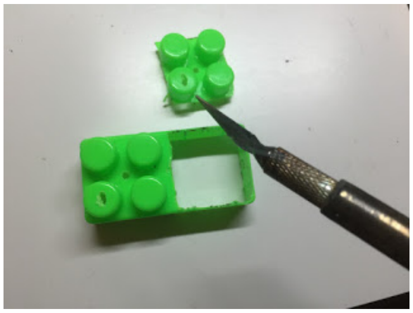

[https://1.bp.blogspot.com/-RzIh_XlphuU/VsRVL6Uub2I/AAAAAAAACTE/AAObSkrQ51o/s1600/IMG_5659.JPG]

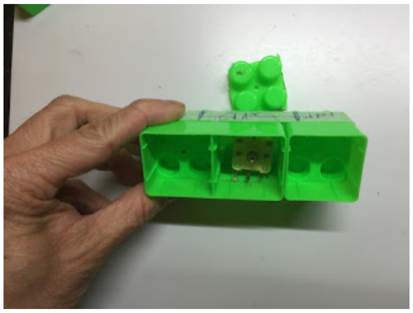

[https://4.bp.blogspot.com/-EdqdMWSn4So/VsRVL533y-I/AAAAAAAACTI/rhVIMOKqZG4/s1600/IMG_5660.JPG]

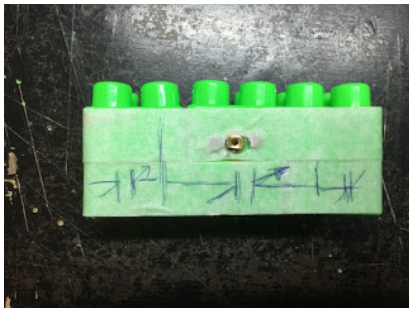

[https://3.bp.blogspot.com/-e0G5owCiEN0/VsRVMTXmXkI/AAAAAAAACTM/j30LH0grecE/s1600/IMG_5661.JPG]

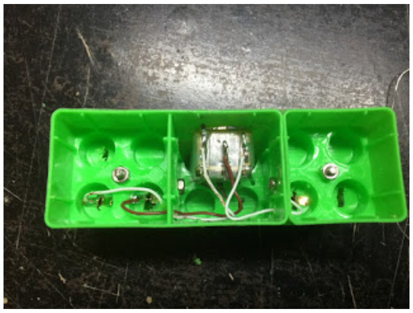

[https://2.bp.blogspot.com/-d2YfyJCM8Ic/VsRVMqz-yYI/AAAAAAAACTQ/Sq9LQAwqAlo/s1600/IMG_5662.JPG]

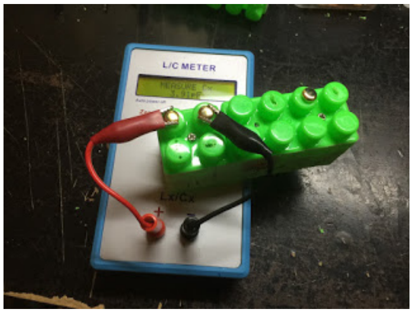

[https://2.bp.blogspot.com/-knRvN4_N3bA/VsRVMqoeBjI/AAAAAAAACTU/VE3WHquEPVY/s1600/IMG_5663.JPG]

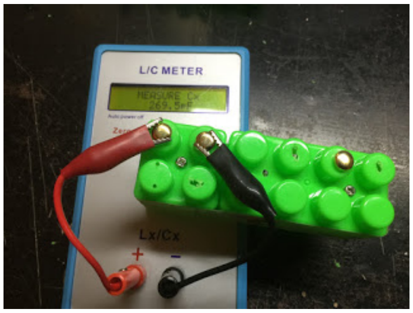

[https://4.bp.blogspot.com/-hY5rdP-s0vw/VsRVM2MC6_I/AAAAAAAACTY/YEhevzQDdlU/s1600/IMG_5664.JPG]

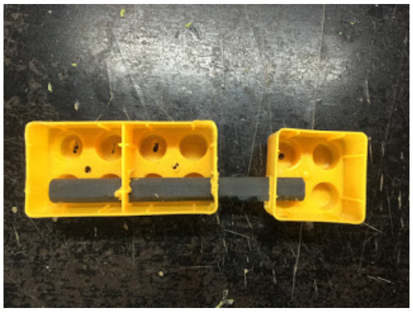

[https://3.bp.blogspot.com/-GFkabuzhgm8/VsRVNA9HqhI/AAAAAAAACTc/0e8ifZXPVqE/s1600/IMG_5673.JPG]

[https://2.bp.blogspot.com/-m3I9WEG9-TE/VsRVNDbH7MI/AAAAAAAACTg/cqwCXIXkCUo/s1600/IMG_5674.JPG]

Dynamic Views theme. Powered by Blogger.

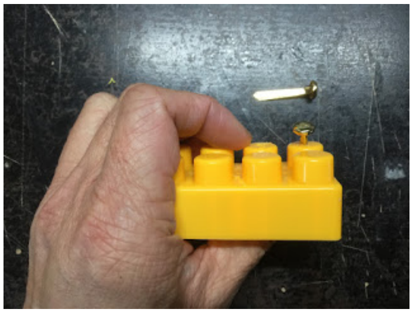

[https://1.bp.blogspot.com/-HhEiPfH3Aqk/VsRVNT_TqYI/AAAAAAAACTk/XZIEmbqQKSc/s1600/IMG_5675.JPG]

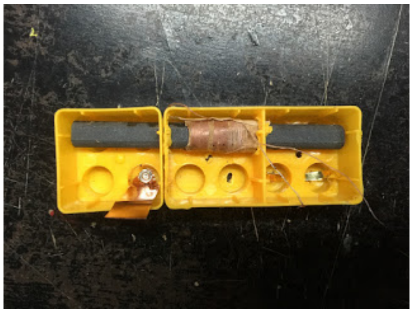

[https://1.bp.blogspot.com/-HMBpDjdTs7M/VsRVNSR6R3I/AAAAAAAACTo/eSzJ59Ozh0w/s1600/IMG_5676.JPG]

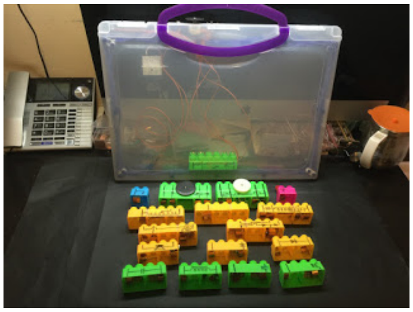

[https://4.bp.blogspot.com/-MPDSVirVnro/VsRVNiAR3GI/AAAAAAAACTs/FS0b598dv2s/s1600/IMG_5677.JPG]

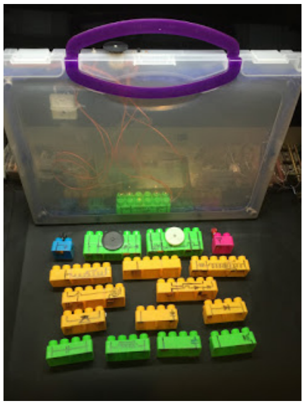

[https://1.bp.blogspot.com/-D-Q1YX12uxc/VsRVN_XRCdI/AAAAAAAACTw/brRaIZomEp4/s1600/IMG_5678.JPG]

[https://2.bp.blogspot.com/-4FldTWhi0Qk/VsRVN6w94oI/AAAAAAAACT0/0L_WOHE2Hl0/s1600/IMG_5679.JPG]

[https://1.bp.blogspot.com/-e1EuDvLgubI/VsRVOFWu-rI/AAAAAAAACT4/tfdcCAn7AWI/s1600/IMG_5680.JPG]

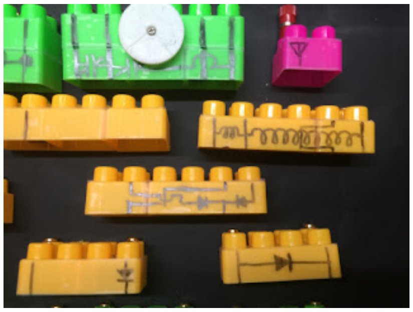

[https://2.bp.blogspot.com/-7hX2_ZMlSl4/VsRVOQPJ9gI/AAAAAAAACT8/gba0yMqtA08/s1600/IMG_5681.JPG]

[https://3.bp.blogspot.com/-Le-vDD1RLdM/VsRVOdT1k9I/AAAAAAAACUA/-wM0dIjOjxs/s1600/IMG_5682.JPG]

Dynamic Views theme. Powered by Blogger.

[https://1.bp.blogspot.com/-VvXLqIxjwMw/VsRVOhAEI6I/AAAAAAAACUE/kOSmifewS_o/s1600/IMG_5683.JPG]

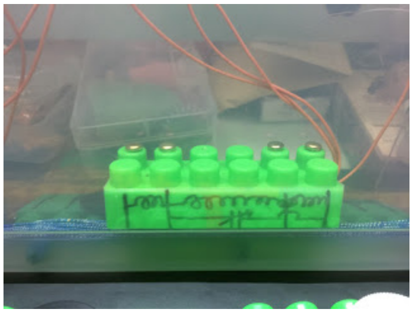

[https://2.bp.blogspot.com/-vLykPWJNpTE/VsRVO8P-uEI/AAAAAAAACUI/KcieYGj_s58/s1600/IMG_5684.JPG]

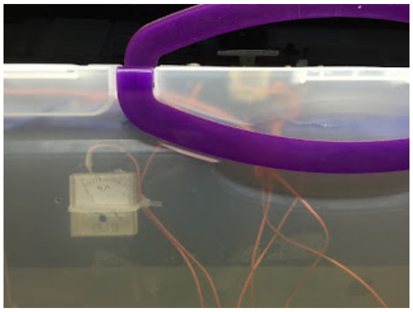

[https://2.bp.blogspot.com/-czyJtBY2vf0/VsRVO9x6WOI/AAAAAAAACUM/IkwKHOjTHGQ/s1600/IMG_5685.JPG]

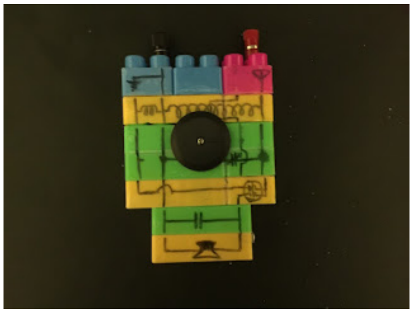

[https://4.bp.blogspot.com/-7cjsmja7zXM/VsRVPCO9IZI/AAAAAAAACUQ/V2e6I6wKbkU/s1600/IMG_5686.JPG]

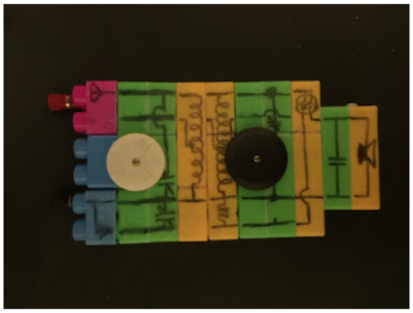

[https://3.bp.blogspot.com/-aOxiaYTjo1c/VsRVPXzQzwI/AAAAAAAACUU/-kdpe_2NdJo/s1600/IMG_5687.JPG]

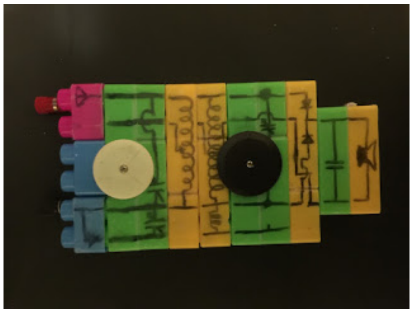

[https://4.bp.blogspot.com/-yR2A0qitFeM/VsRVPZZbdtI/AAAAAAAACUY/8QF9_-SRpWA/s1600/IMG_5688.JPG]

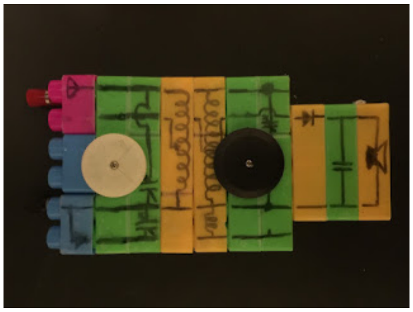

[https://2.bp.blogspot.com/-pemyXJIY_64/VsRVPt2wJ9I/AAAAAAAACUc/5wLz6F08b20/s1600/IMG_5689.JPG]

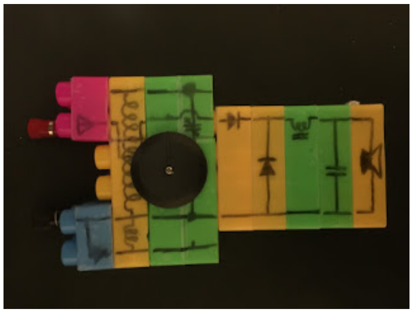

[https://2.bp.blogspot.com/-DnT8juXYYT0/VsRVPoaTx_I/AAAAAAAACUg/1dEjqzHeA8k/s1600/IMG_5690.JPG]

Dynamic Views theme. Powered by Blogger.

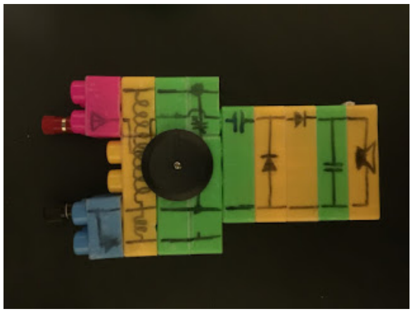

[https://4.bp.blogspot.com/-3TU7x4kuDrI/VsRVPz5JCqI/AAAAAAAACUk/vRWlDz8gXko/s1600/IMG_5691.JPG]

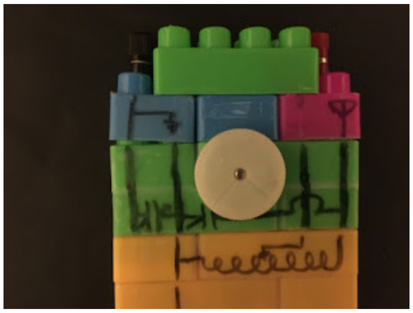

[https://1.bp.blogspot.com/-EM54csY71iM/VsRVQBd2yUI/AAAAAAAACUo/lJ2vKWeyGyY/s1600/IMG_5692.JPG]

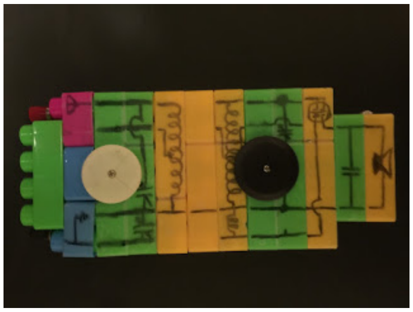

[https://1.bp.blogspot.com/-iTI7bhG0UBk/VsRVQMcAAFI/AAAAAAAACUs/SGNn69JnV9c/s1600/IMG_5693.JPG]

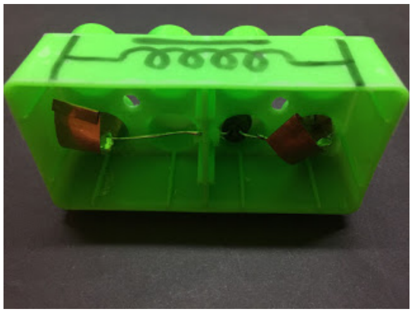

[https://1.bp.blogspot.com/-l9x8qZntSqU/VsRVQU7RqLI/AAAAAAAACUw/XfuV01f0pO4/s1600/IMG_5694.JPG]

[https://3.bp.blogspot.com/-5s1XBWWKAUk/VsRVRmKdpqI/AAAAAAAACVE/lqDsdvYfcA8/s1600/IMG_5695.JPG]

[https://2.bp.blogspot.com/-nBKFeB6NHcY/VsRVQ5DzosI/AAAAAAAACU0/SZ8nmLFokDk/s1600/IMG_5696.JPG]

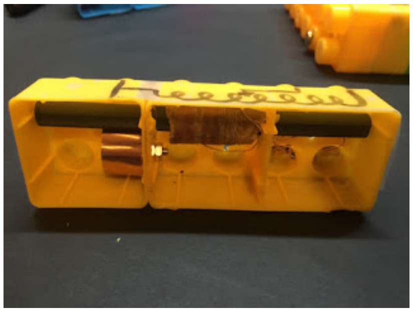

[https://1.bp.blogspot.com/-jy3JasXrkD4/VsRVRClh1hI/AAAAAAAACU4/BKR1AKKATWA/s1600/IMG_5697.JPG]

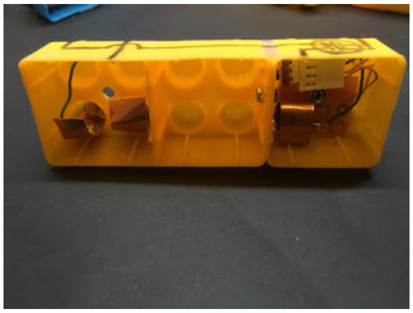

[https://1.bp.blogspot.com/-9KZSX0B22yQ/VsRVRUczeYI/AAAAAAAACU8/aB4LiJfnT0E/s1600/IMG_5698.JPG]

Dynamic Views theme. Powered by Blogger.

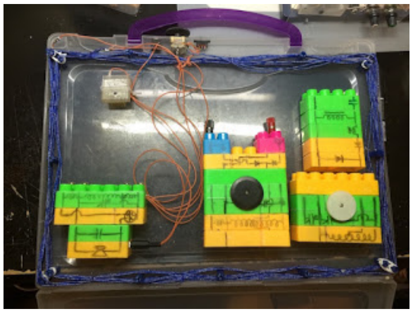

[https://4.bp.blogspot.com/-xdRllb2wSmA/VsRVRnoiDlI/AAAAAAAACVA/tBLA0QZ4e4Q/s1600/IMG_5709.JPG]

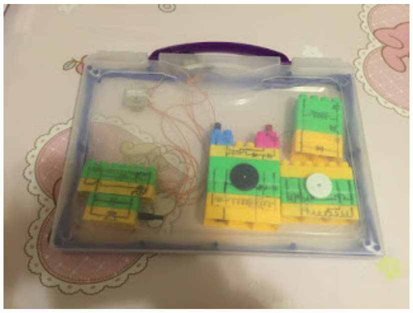

[https://3.bp.blogspot.com/-MCRp4B-bXbA/VsRVR80i6oI/AAAAAAAACVI/-otm-sxK1OI/s1600/IMG_5710.JPG]

[https://4.bp.blogspot.com/-AwrjVzkXo3E/VsRVSIxvYjI/AAAAAAAACVM/brPOm_hpSD4/s1600/IMG_5711.JPG]

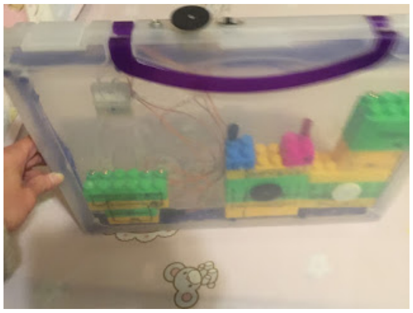

[https://1.bp.blogspot.com/-iXyOrUkwfxg/VsRVSLddyQI/AAAAAAAACVQ/mu5S0AW-dMA/s1600/IMG_5712.JPG]

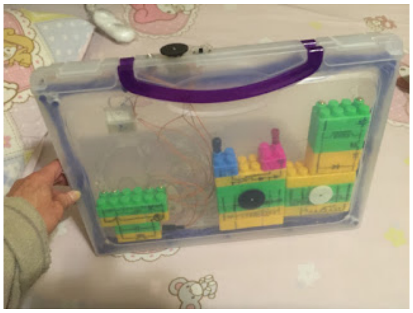

[https://1.bp.blogspot.com/-w8YT4uTQkMk/VsRVSgDhFLI/AAAAAAAACVU/eGb80r_tgdo/s1600/IMG_5713.JPG]

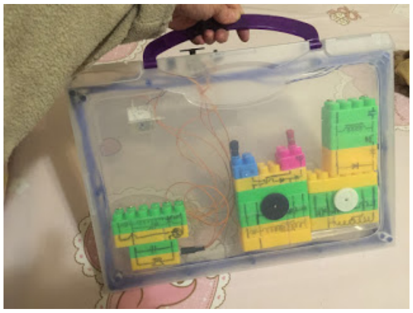

[https://2.bp.blogspot.com/-tzQHNzVkdjA/VsRVSgk1XuI/AAAAAAAACVY/00pFXtY9GEc/s1600/IMG_5715.JPG]

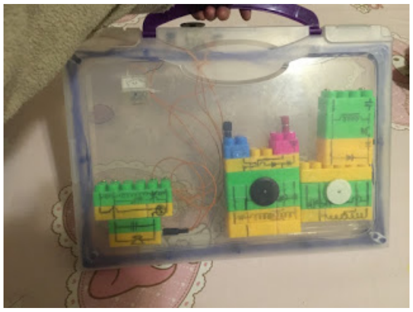

[https://2.bp.blogspot.com/-8qVf8BVjB1c/VsRVS5MVv3I/AAAAAAAACVc/KozKzNZOuOg/s1600/IMG_5716.JPG]

Posted 17th February 2016 by Billy

0 Add a comment

Great Deal At Oakland Ca!! in Oakland...
$499,000
Hot Off Market Way Below Market...
$140,000
Hot Deal At Oakland Ca!!! in Oal
$650,000
M

11th February 2016

Portable 3DQ FM SW MW Crystal. Regen. Radio

I created a portable FM SW MW Crystal Set with region functions using a combination of big loop and ferrite toroids.

Dynamic Views theme. Powered by Blogger.

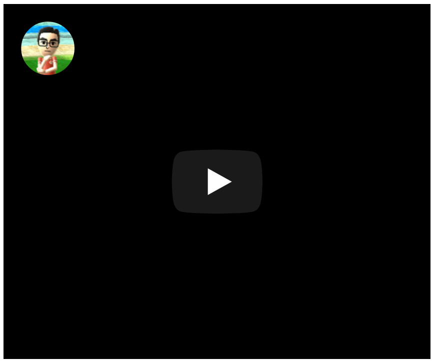

Circuit Diagram

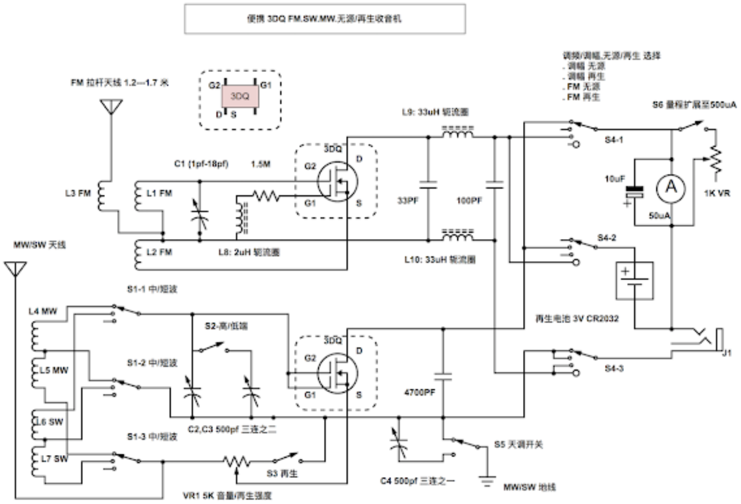
便携 3DQ FM.SW.MW.无源/再生收音机
FM 拉杆天线 1.2—1.7 米
G2
G1
3DQ
D
S
L9: 33uH 扼流圈
C1 (1pf-18pf)
1.5M
L3 FM
L1 FM
33PF
100PF
MW/SW 天线
L2 FM
L8: 2uH 扼流圈
L10: 33uH 扼流圈
调频/调幅,无源/再生 选择
. 调幅 无源
. 调幅 再生
. FM 无源
. FM 再生
S6 量程扩展至500uA
S4-1
10uF
A
50uA
1K VR
S4-2
再生电池 3V CR2032
S1-1 中/短波
L4 MW
S2-高/低端
S1-2 中/短波
L5 MW
4700PF
S4-3
J1
L6 SW
C2,C3 500pf 三连之二
S1-3 中/短波
L7 SW
S3 再生
VR1 5K 音量/再生强度
C4 500pf 三连之一
S5 天调开关
MW/SW 地线

[https://4.bp.blogspot.com/-Zxum5N-FL_M/Vrytda4CGbI/AAAAAAAACO0/GSCePvGsl50/s1600/Circuit%2BDiagram.png]

Components

S1 MW/SW Selection
S2 MW/SW Hi Band/Low Band Selection
S3 Switch attached to VR2, turn on for region adjustment or volume adjustment.
S4 Selection for FM or MW/SW and Regen or Battery Free Radio.
S5 Antenna adjustment switch
VR1 Volume or Regen effect adjustment. with S3 attached. Connect S3 to VR1 such that when resistance to 3DQ-S is the lowest, S3 will stay open.
L8 using a 1.5M 1/4 W resister as the frame, wind 30 turns of 0.07 or 0.1 mm magnet wire.
L9, L10 using a 1.5M 1W resister as the frame, wind 100 turns of 0.07 or 0.1 mm magnet wire.
C1 1pf-18pf butterfly high Q variable capacitor.
C2, C3, C4 3x500pf VR.
J1 connect to a reed headphone with impedance of 300 to 600 ohms.
3V CR2032 Button Battery for Regen.

L1 make a 120mm diameter C-shape loop (with a 50mm openning) using 12mm diameter soft copper pipe.
L2 make a 25mm diameter C-shape loop using 1 mm diameter silver plated copper wire. Wound in same direction as L1.
L3 make a 25mm diameter C-shape loop using 1 mm diameter silver plated copper wire. Wound in same direction as L1.

This Crystal Set can hook up to big loop coils or toroid coils.

For toroid coils:
L4 use a ferrite toroid with permeability of 100u (NXO100) diameter 37mm, height 7 mm, wind 50 turns using 175x0.04mm Litz wire.
L5 On the same toroid, using same type of litz wire, wind 10 turns in the opposite direction as L4.
L6 use a ferrite toroid with permeability of 100u (NXO100) diameter 31mm, height 7 mm, wind 12 turns using 1mm magnet wire.
L7 On the same toroid, using same type of magnet wire, wind 4 turns in the opposite direction as L6.

For big loops:

Dynamic Views theme. Powered by Blogger.

like winding on a toroid.
L5 On the same frame, using same type of litz wire, wind 1 turn in the opposite direction as L4.
L6 make a 1 Meter diameter C-Shape loop (with a 100mm opening) using 12mm diameter soft copper pipe.
L7 make a 0.5 Meter diameter C-Shape loop (with a 100mm opening) using 9 mm diameter soft copper pipe.

[https://1.bp.blogspot.com/-KSvLkg7lCKk/VrytiPdMncl/AAAAAAAACO4/qOwy8w7LbN8/s1600/IMG_5448.JPG]

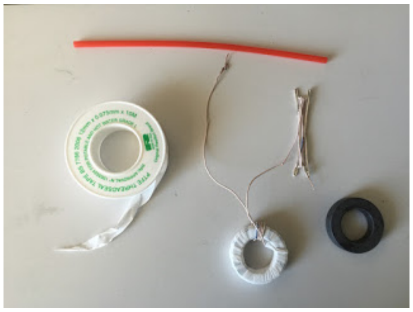

[https://4.bp.blogspot.com/-45V-d8zPnmA/VrytiIZyBSI/AAAAAAAACPA/bzZnuqiOeLM/s1600/IMG_5495.JPG]

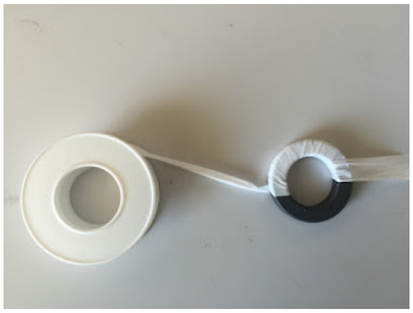

[https://1.bp.blogspot.com/-JsXI-N3lsSM/VrytiBF4EwI/AAAAAAAACO8/sANGr1x7OI4/s1600/IMG_5498.JPG]

[https://4.bp.blogspot.com/-DfNLyBSVolk/Vrytiean9_I/AAAAAAAACPE/bqvJssZvGNw/s1600/IMG_5499.JPG]

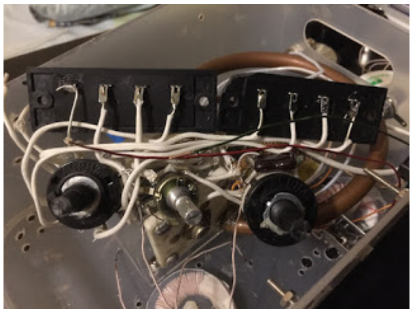

[https://3.bp.blogspot.com/-c5t76AugRMY/VrytisTs64I/AAAAAAAACPI/_aYNWIQ2p4I/s1600/IMG_5520.JPG]

[https://1.bp.blogspot.com/--4tywSrTHIw/VrytiskbakI/AAAAAAAACPM/xQQrCJkVfrw/s1600/IMG_5526.JPG]

[https://1.bp.blogspot.com/-DfT5naQobnQ/Vryti_0H5DI/AAAAAAAACPQ/vY4YBySqziA/s1600/IMG_5527.JPG]

[https://4.bp.blogspot.com/-lHbJQ_CJMIM/VrytjFziKiI/AAAAAAAACPU/TBNxlrXZezU/s1600/IMG_5528.JPG]

Dynamic Views theme. Powered by Blogger.

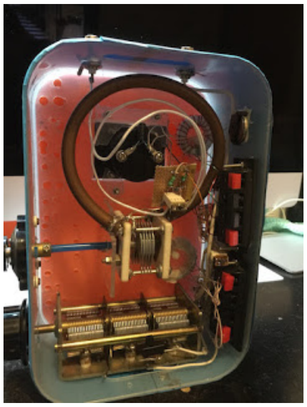

[https://3.bp.blogspot.com/-4WtgnuEfNe4/VrytjW_WRLI/AAAAAAAACPY/rZ7RETL4siQ/s1600/IMG_5529.JPG]

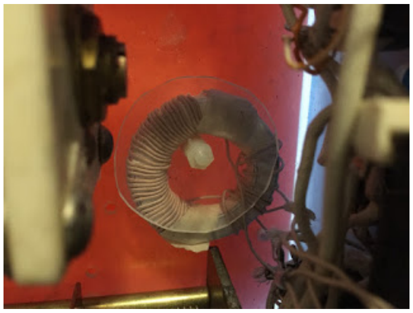

[https://4.bp.blogspot.com/-1ZF2OWxCuPg/VrytjsIfI6I/AAAAAAAACPg/pun15eWqFRY/s1600/IMG_5530.JPG]

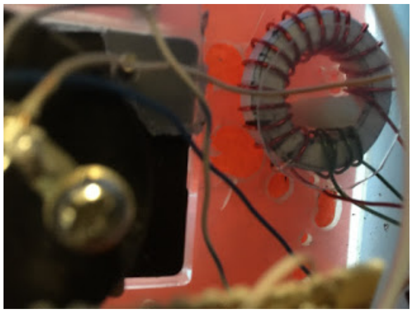

[https://1.bp.blogspot.com/-KSsFp_8X8Ns/Vrytj7aDwDI/AAAAAAAACPc/zWjRob1y_lo/s1600/IMG_5531.JPG]

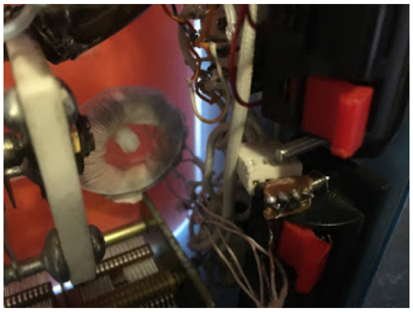

[https://4.bp.blogspot.com/-JAASUWqVFiI/VrytkKyxn-I/AAAAAAAACPk/TBRaSUZm6ug/s1600/IMG_5532.JPG]

[https://4.bp.blogspot.com/-B-M1aSaSwRI/VrytkVOwhiI/AAAAAAAACPo/kBr1qEc4GIQ/s1600/IMG_5533.JPG]

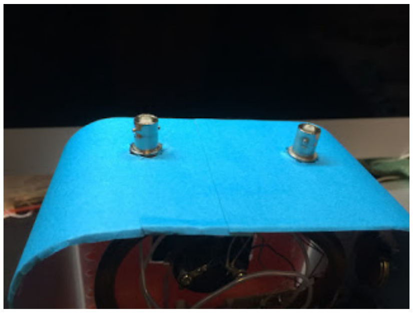

[https://3.bp.blogspot.com/-m5r8D32ecfg/VrytkdQ-8pI/AAAAAAAACPs/5Y4fDsHwIdQ/s1600/IMG_5534.JPG]

[https://2.bp.blogspot.com/-oKX4OU_at04/VrytktG7foI/AAAAAAAACPw/ZGvzS3lFFWs/s1600/IMG_5537.JPG]

[https://1.bp.blogspot.com/-5hOOx6iEdAY/Vrytk2F0pcI/AAAAAAAACP0/HtGBW4I1HPw/s1600/IMG_5538.JPG]

Dynamic Views theme. Powered by Blogger.

[https://2.bp.blogspot.com/-a0EH30Ix40g/VrytkxoHW1I/AAAAAAAACP4/qpTZXzlCXvI/s1600/IMG_5539.JPG]

[https://3.bp.blogspot.com/-poQElYUpilM/VrytlJy10vI/AAAAAAAACP8/528nWFQhauQ/s1600/IMG_5540.JPG]

[https://3.bp.blogspot.com/-Z6xdzX_BZLg/VrytlAc4IQI/AAAAAAAACQA/2gyMC7xMIkg/s1600/IMG_5541.JPG]

Dynamic Views theme. Powered by Blogger.

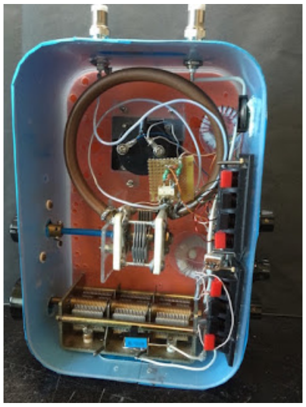

[https://2.bp.blogspot.com/-PpHSJGO-jxM/VrytlbT_ZgI/AAAAAAAACQE/kTxZjYcxio8/s1600/IMG_5542.JPG]

[https://4.bp.blogspot.com/-tN73K7poDpY/VrytlqYE9aI/AAAAAAAACQI/rlbBZ_R5Z-o/s1600/IMG_5543.JPG]

[https://1.bp.blogspot.com/-FJN5L7Vf9jA/VrytlhJP6xI/AAAAAAAACQQ/uMvqr7iT8Ak/s1600/IMG_5544.JPG]

[https://3.bp.blogspot.com/-oMQDcnIDwCs/VrytlwGC0-I/AAAAAAAACQM/U1ZVbpVu9d0/s1600/IMG_5545.JPG]

Dynamic Views theme. Powered by Blogger.

[https://3.bp.blogspot.com/-QKeSX7ZFnfU/VrytmDforEI/AAAAAAAACQU/n5nPEGyvMNA/s1600/IMG_5555.JPG]

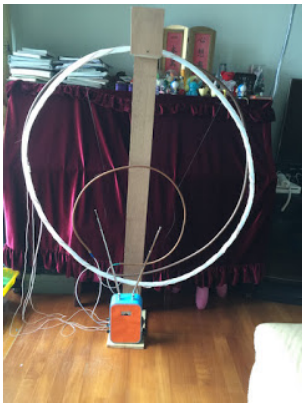

[https://1.bp.blogspot.com/-dyZG62qj-0k/VrytmVVLlSI/AAAAAAAACQY/RqnI2Cm-bgs/s1600/IMG_5556.JPG]

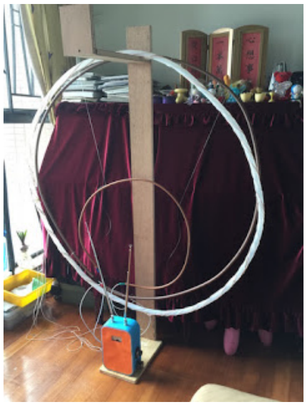

[https://1.bp.blogspot.com/-MmHqvxeqyHI/VrytmiK_ndI/AAAAAAAACQc/vSHHwYj67aA/s1600/IMG_5557.JPG]

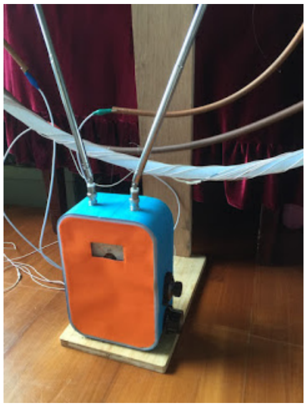

[https://1.bp.blogspot.com/-GsUr9Mytrfs/Vrytmhvv3AI/AAAAAAAACQg/0zmj71oPUPU/s1600/IMG_5558.JPG]

Dynamic Views theme. Powered by Blogger.

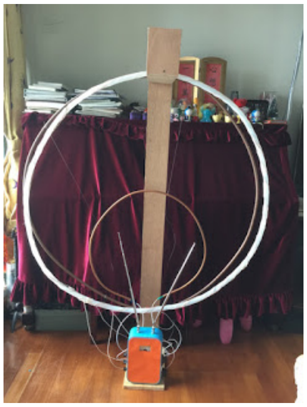

[https://3.bp.blogspot.com/-jleP5j1XWeo/Vrytm6fgcdI/AAAAAAAACQk/jR8Mq9fIYXo/s1600/IMG_5559.JPG]

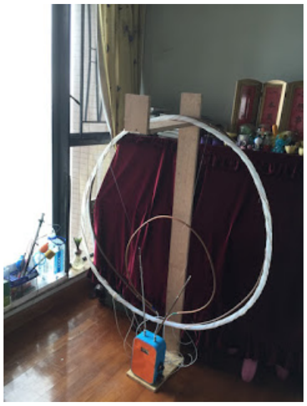

[https://4.bp.blogspot.com/-27eGMwzA57s/VrytnB6Dx6I/AAAAAAAACQo/ENcSaq6c9Ao/s1600/IMG_5560.JPG]

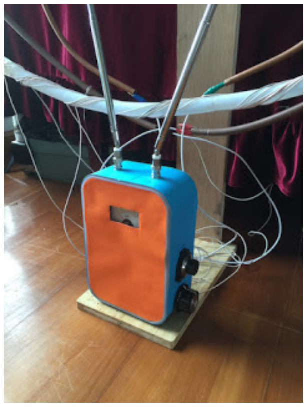

[https://1.bp.blogspot.com/-ctycbGwZ8ms/VrytnB6IS8I/AAAAAAAACQs/bGdc1bGUfiQ/s1600/IMG_5561.JPG]

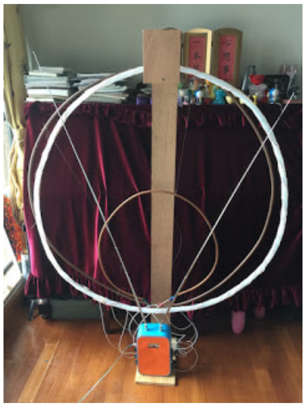

[https://3.bp.blogspot.com/-Co3tEPJelro/VrytnTCd8OI/AAAAAAAACQw/BPchrA-N1f8/s1600/IMG_5562.JPG]

[https://4.bp.blogspot.com/-5sOGHDKdCW0/VrytnsJcUmI/AAAAAAAACQ0/RD_PEBf-Wu4/s1600/IMG_5565.JPG]

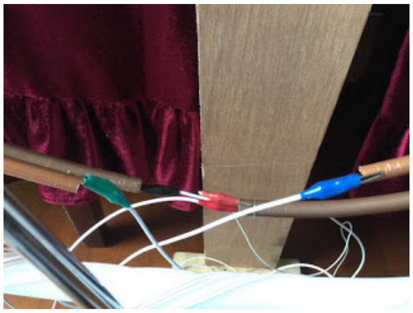

[https://2.bp.blogspot.com/-Ysw9vUwmwTI/Vrytn8ixC5I/AAAAAAAACQ4/kIxoEKkIGAk/s1600/IMG_5566.JPG]

[https://1.bp.blogspot.com/-waABLAMIe0E/VrytnxG5K3I/AAAAAAAACQ8/V6a1-KV9wLw/s1600/IMG_5567.JPG]

Dynamic Views theme. Powered by Blogger.

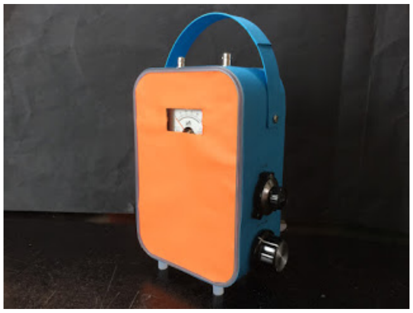

[https://2.bp.blogspot.com/-Vf9iMIj1xlA/VrytoOSUK_I/AAAAAAAACRA/WZyv_Xn2W74/s1600/IMG_5568.JPG]

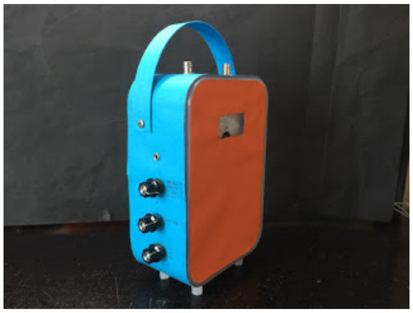

[https://4.bp.blogspot.com/-89quArXeC1s/VrytoQp0P6I/AAAAAAAACRE/VLURD1KVXhE/s1600/IMG_5570.JPG]

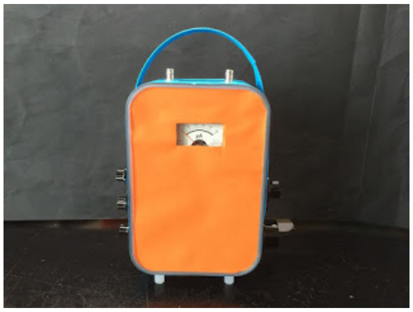

[https://1.bp.blogspot.com/-ayhUyZ6JRkk/Vrytofls7DI/AAAAAAAACRI/SA8Z4-A38S4/s1600/IMG_5571.JPG]

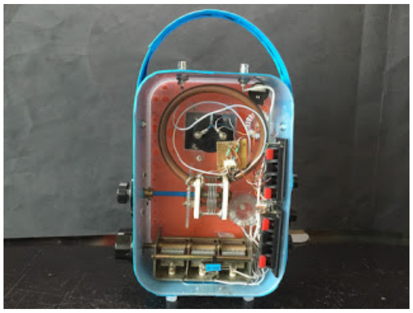

[https://2.bp.blogspot.com/-7GAgf3LOr_I/VrytouiHtfI/AAAAAAAACRM/Okva41gAzTU/s1600/IMG_5572.JPG]

Dynamic Views theme. Powered by Blogger.

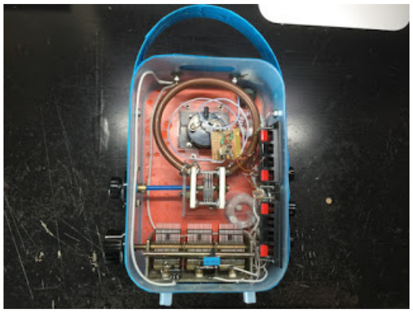

[https://1.bp.blogspot.com/-pzHzBOZZmwI/Vrytox4QfpI/AAAAAAAACRQ/TBfgCpWi-Ys/s1600/IMG_5574.JPG]

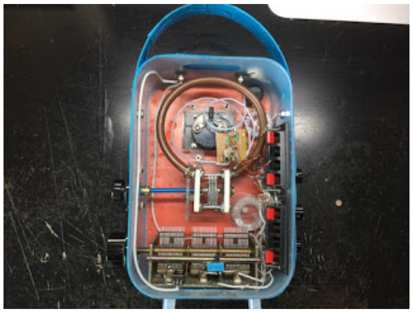

[https://1.bp.blogspot.com/-WNNzOQEr_Bo/VrytpCljpoI/AAAAAAAACRU/l2W-Obh7_4Y/s1600/IMG_5575.JPG]

[https://3.bp.blogspot.com/-_7Jg8zo-15s/Vrytpf8SZXI/AAAAAAAACRk/71fmZmZmm0g/s1600/IMG_5579.JPG]

[https://1.bp.blogspot.com/-5ZPzqYcgzis/VrytpeLB6yI/AAAAAAAACRY/_fJ_gvxWZNY/s1600/IMG_5580.JPG]

[https://3.bp.blogspot.com/-Xyi47geXS-E/Vrytp7VC3lI/AAAAAAAACRc/T15L7K6OhG4/s1600/IMG_5581.JPG]

[https://2.bp.blogspot.com/-C18psLBFfjo/VrytqMnej9I/AAAAAAAACRg/vfA5iWuO_EU/s1600/IMG_5582.JPG]

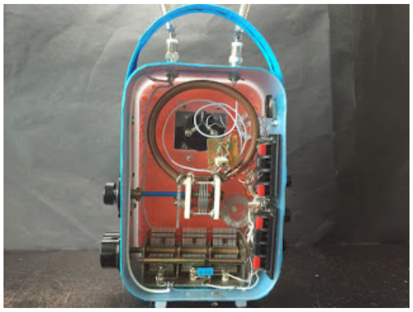

[https://1.bp.blogspot.com/-3kZ14WzVU74/VrytqawQXoI/AAAAAAAACRo/oWc3giuQzWs/s1600/IMG_5583.JPG]

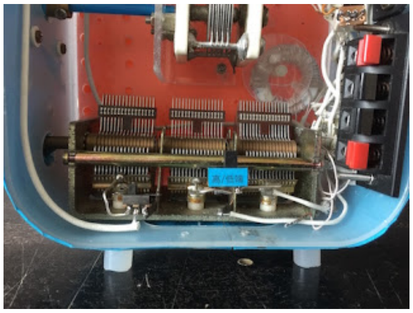

[https://2.bp.blogspot.com/-ewb-gdXaZhU/VrytqowhgRI/AAAAAAAACRs/9eBE_LMWvFg/s1600/IMG_5584.JPG]

Dynamic Views theme. Powered by Blogger.

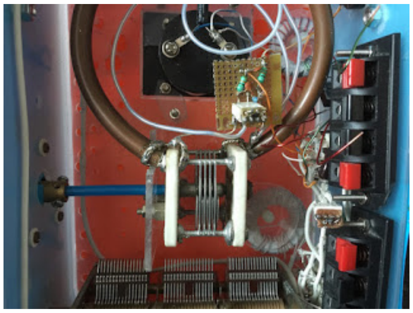

[https://2.bp.blogspot.com/-0xg9o1PnC9w/Vrytq6Dk8VI/AAAAAAAACRw/dAX8GpDcAm4/s1600/IMG_5585.JPG]

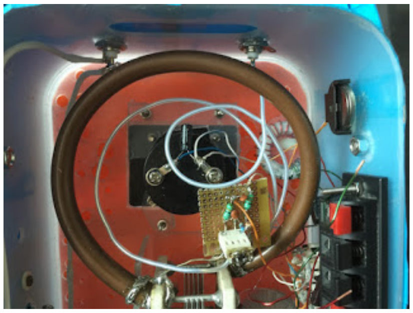

[https://4.bp.blogspot.com/-JxQHkVeeB0A/VrytqzYabvI/AAAAAAAACR0/C-eyScU_Bys/s1600/IMG_5586.JPG]

Posted 11th February 2016 by Billy

0 Add a comment

14th January 2016

Copper Tube SW radio with a DIY Var Cap.

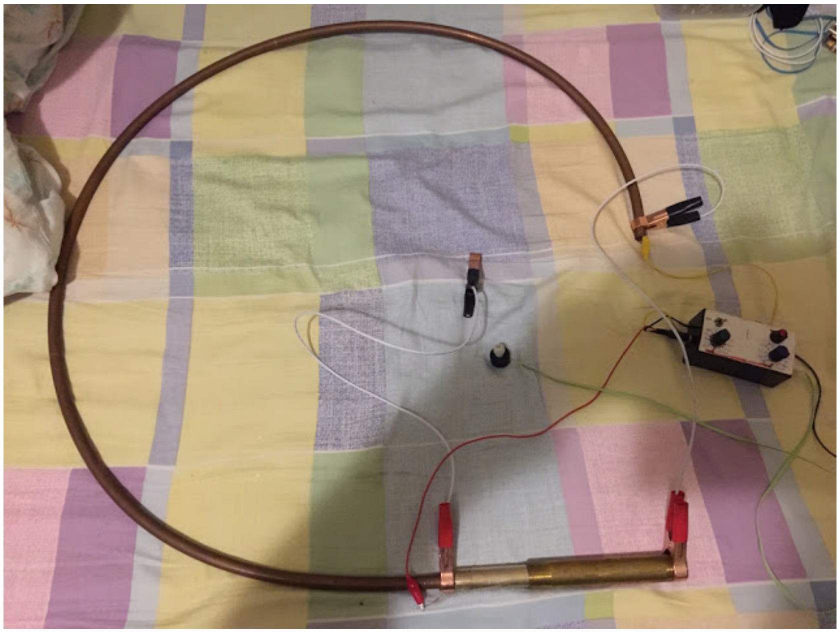

[http://2.bp.blogspot.com/-p9mh-YtSKY4/VpfAbwcyOsI/AAAAAAAACOk/WLb3ysgAoSw/s1600/IMG_5301.JPG]

I have built a very simple portable short wave radio that can receive stations without an antenna using half inch or quarter inch copper tubes.

First band the tube into a circle with a diameter close to 1 meter (I used 0.8 meter) but do not close the loop. Leave an openig of around 30 degree to 45 degree.

cover the lower SW frequencies) in parallel to both end of the tubes.
Connect one end of the tube to a Crystal radio diode eg 1N34A or 1N60 or even better with 1s106 that I used.
Connect the other end of the diode to a head phone with a impedance that matches that of the output from the diode. Otherwise a transformer will be required to match the impedance. Connect the other end of the tube to the Other end of the headphone. IN the picture below, I used a transformer that step up from 1.5K to 5K to match the impedance of the crystal earphone I am using.
If you are using a magnetic headphone, add a 2000-4000pf capacitor in parallel to it to improve the audio quality.
If you are using a Crystal headphone (aka piezo electric), use a transformer like what I did below for better audio quality, or add a 50k to 1M resistor in parallel to it to sound better.

Placing this horizontally next to my window from my 56/F apartment facing the sea, I can receive SW stations broadcasted from mainland China, Taiwan and other South Asian countries next to Hong Kong.
The loudness will rise and drop within a minute as the SW signal is reflected from afar.

Here are some pictures:

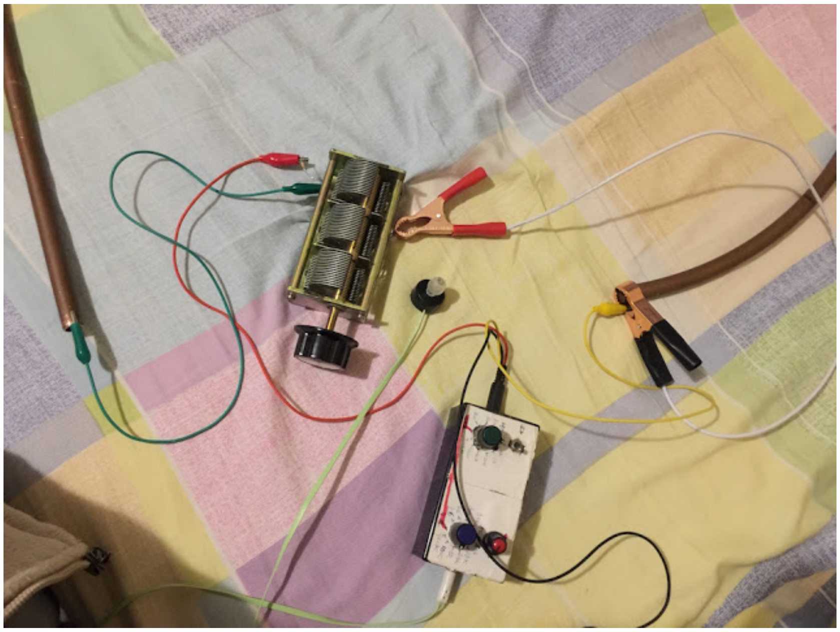

[http://4.bp.blogspot.com/-eOo2a8kKAoo/Vpe73-Jq95I/AAAAAAAACNw/vduQY4dvGME/s1600/IMG_5300.JPG]

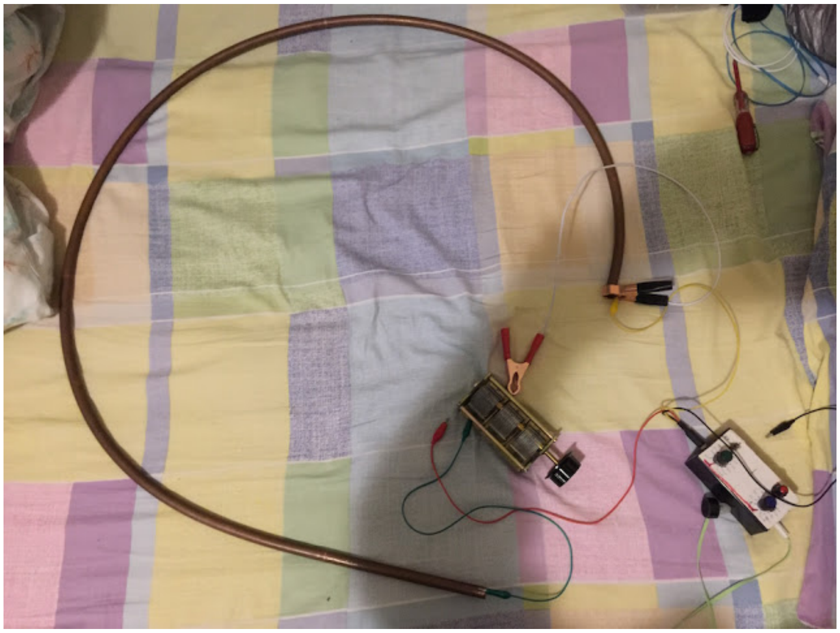

[http://1.bp.blogspot.com/-23NdhzZmycQ/Vpe73za4TYI/AAAAAAAACN0/j5X1GFvL0io/s1600/IMG_5299.JPG]

Technically you can DIY your own Var cap also using the copper tube. Instead of banding it to a circle. Band it to form a letter C with a long horizontal tail at the bottom. Then you can place an insulated wider copper tube on top of the tube used for the coil. If the two tubes are close enough together i.e. The gap is less than 0.5 mm, By sliding the outter tube in and out , you can vary the capacitance between them. I guess depending on the surface area and how tightly the two tubes are placed together, the capacitance can vary between 20pf to 500pf. The Q factor will be effected by what is being used as the insulator, Air is best but hard to avoid short circuit. Plastic sheets may work , but will be low Q:
I haven't tried this on SW yet, only on FM.

Here is some diagrams that show you how this will look like.

I mounted another hard straight copper tube covered with plastic sheet (diameter 3/4 Inch) to the original C-Shaped copper tube (diameter 1/2 inch) that forms the coil of the SW Crystal Radio. Because this will increase the surface area and the capacitance . Also the hard straight copper tube has less gap with the wider copper tube (diameter 1 inch), and can slide in and out smoothly.

You can always increase the length of the two copper tubes forming the variable cap by extending the length of these two copper tubes.

You can see from my Capacitance Meter that the DIY Var. Cap can go from 12pf (the outer tube all outside) to 318pf (the outer tube all inside).

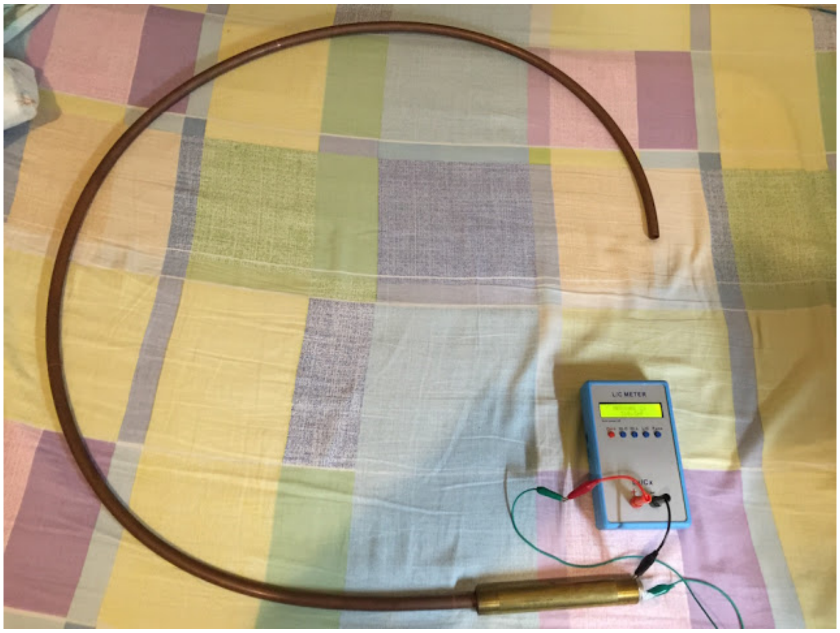

[http://3.bp.blogspot.com/-caLYoPy3lUk/VpeZPFAOWyI/AAAAAAAACNA/peDmNZNbK9Q/s1600/IMG_5297.JPG]

Will give it a go when I have some time.

Dynamic Views theme. Powered by Blogger.

[http://3.bp.blogspot.com/-deFfJUYMRM0/VpeZOM6qdVI/AAAAAAAACMg/ZLHuUw3IHHo/s1600/IMG_5291.JPG]

[http://1.bp.blogspot.com/-ogAX8wr6pYU/VpeZOK70H6I/AAAAAAAACMo/DgXDmKtiXAE/s1600/IMG_5292.JPG]

[http://2.bp.blogspot.com/-249uKYGQ_mE/VpebhveXS-I/AAAAAAAACNc/EqZ9dJ2Wk7w/s1600/IMG_5298.JPG]

[http://1.bp.blogspot.com/-owL9FUmxQLQ/VpeZOPU16hI/AAAAAAAACMk/QxLLtN_8EZM/s1600/IMG_5293.JPG]

Dynamic Views theme. Powered by Blogger.

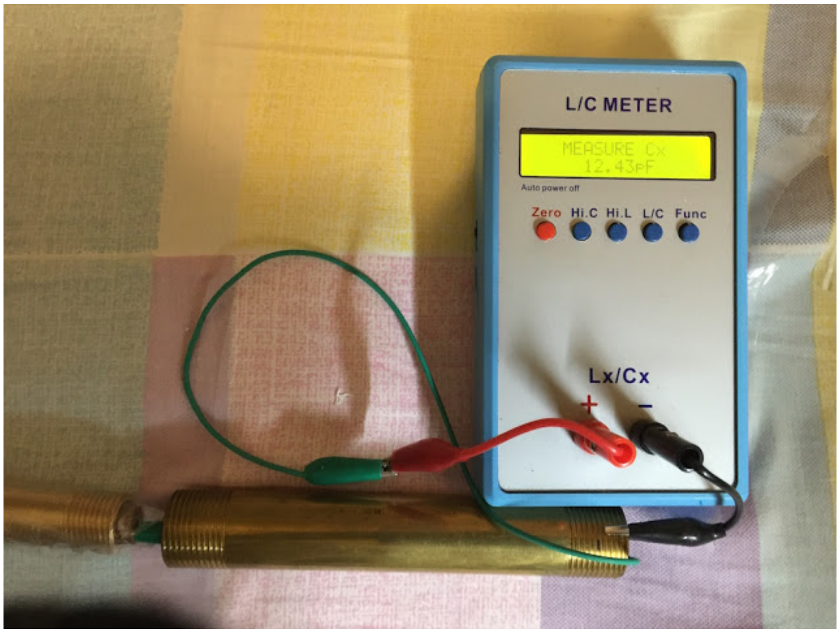
L/C METER
MEASURE Cx
12.43pF
Auto power off
Zero
Hi.C
Hi.L
L/C
Func
Lx/Cx
+
−

[http://1.bp.blogspot.com/-amohYVUmuKs/VpeZOtnaPqI/AAAAAAAACMs/v3rhoFflsv4/s1600/IMG_5294.JPG]

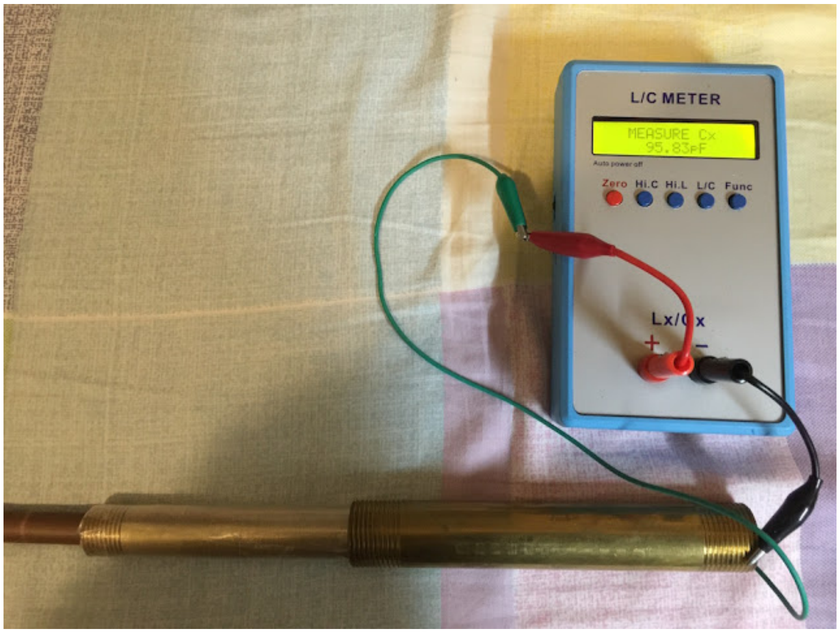
L/C METER
MEASURE Cx
95.83pF
Auto power off
Zero
Hi.C
Hi.L
L/C
Func
Lx/Cx
+
−

[http://4.bp.blogspot.com/-330pG1MZqkU/VpeZOlMv0AI/AAAAAAAACMw/nU1GT--cspE/s1600/IMG_5295.JPG]

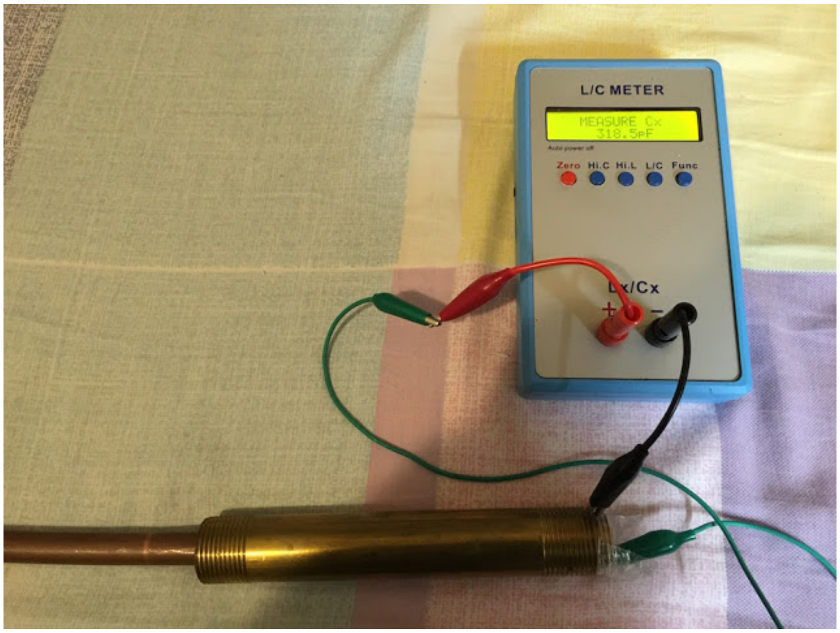

[http://3.bp.blogspot.com/-ZCW4jlaPsiI/VpeZOzTKxVI/AAAAAAAACM4/A-w6q0Ae6CI/s1600/IMG_5296.JPG]

I hooked it up to the diode and the crystal earphone through a transformer 1.5K to 5K to match my crystal earphone.

The loudness of this SW crystal set made using this DIY var. Cap is comparable to that when I used my high Q 500pf Var Cap above. However, tuning is not as stable was the regular Var Cap and you need to adapt to it to learn how to tune. As the copper tube sometime shifts and get loose which alter the Capacitance. With the already varying strength of the SW signal itself, it's even harder to tune and keep the SW station. However, this proof of concept works well.

See how it's hooked up.

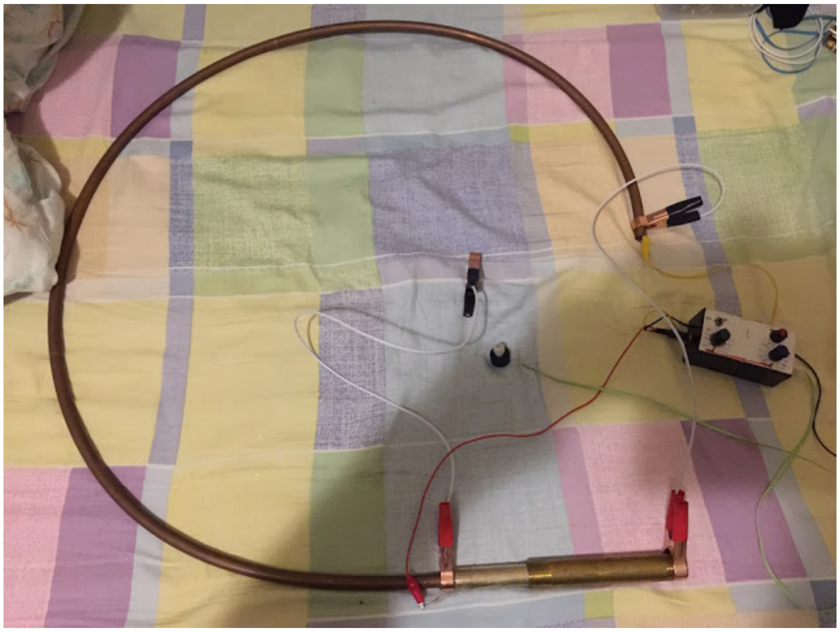

[http://1.bp.blogspot.com/-hT6LR89TP6M/Vpe74JbPc2I/AAAAAAAACN8/aIDbau_c0mM/s1600/IMG_5301.JPG]

Dynamic Views theme. Powered by Blogger.

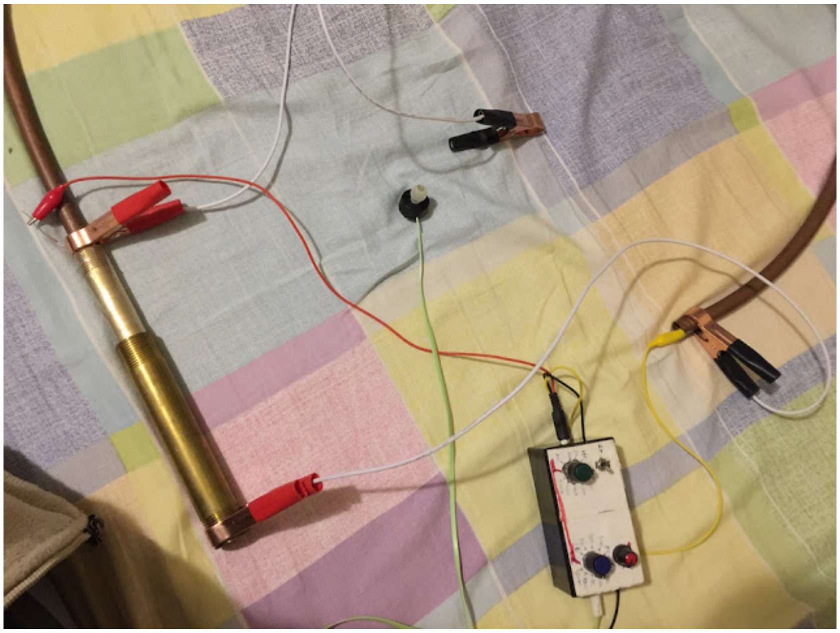

[http://1.bp.blogspot.com/-U6KytVWEf7k/Vpe74lfnINI/AAAAAAAACOM/ih9XPqLy-Bg/s1600/IMG_5302.JPG]

[http://2.bp.blogspot.com/-Zj1qKbVw8WY/Vpe74wIkrTI/AAAAAAAACOQ/5QrFHrSV9ms/s1600/IMG_5304.JPG]

Posted 14th January 2016 by Billy

0 Add a comment

8th January 2016

Simple Crystal Radio driving an 8 inch speaker

I experimented with a very simple two stage circuits with a single diode that drive an 8 eight speaker. Though not loud, but it's clearly audible within my quiet room.

Diode: 1ss106.
Distance from the nearest broadcasting station : 15KM.
Antenna 15 meters of 660x0.04 Litz wire drop down from my window 56/F floor, pushed out to 2 meters away from the wall of the building.

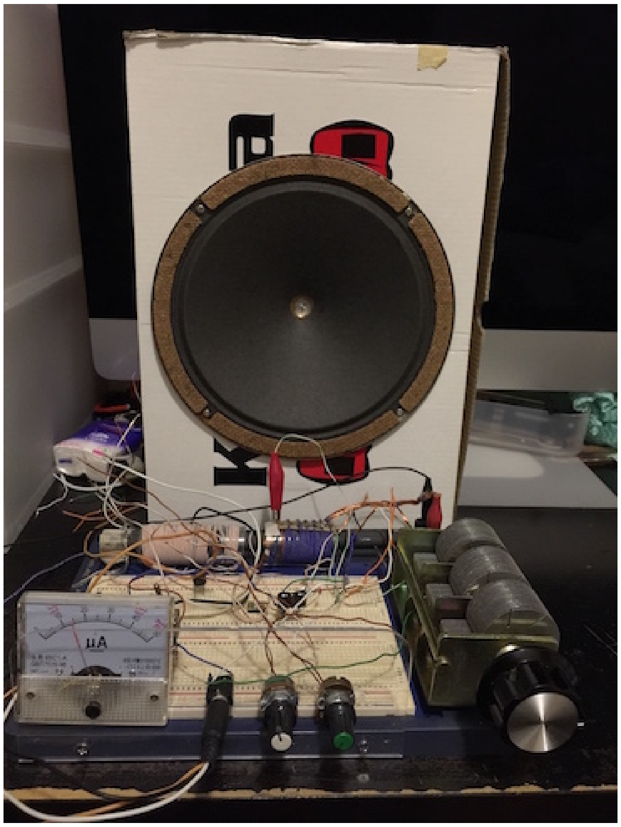

[http://2.bp.blogspot.com/-sQZgB03d_ac/Vo-YNdbejHI/AAAAAAAACK8/I8v-PfmhS2Q/s1600/IMG_5234.JPG]

[http://2.bp.blogspot.com/-sQZgB03d_ac/Vo-YNdbejHI/AAAAAAAACK8/I8v-PfmhS2Q/s1600/IMG_5234.JPG]

I can achieve a DC current of 150uA when using this conventional single Diode circuit.

[http://2.bp.blogspot.com/-sQZgB03d_ac/Vo-YNdbejHI/AAAAAAAACK8/I8v-PfmhS2Q/s1600/IMG_5234.JPG]

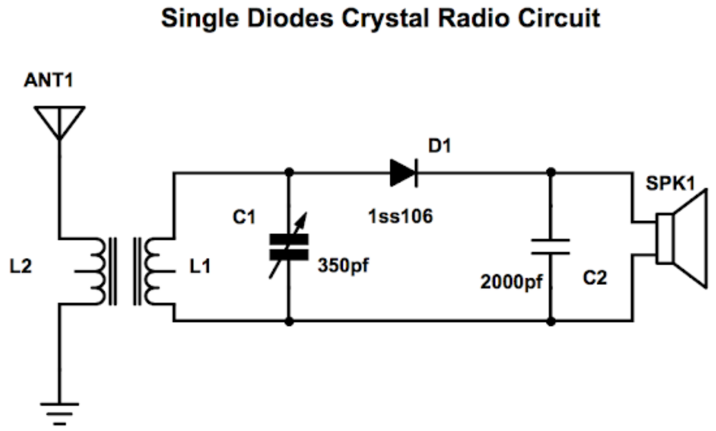
Single Diodes Crystal Radio Circuit
ANT1
D1
SPK1
C1
1ss106
L2
L1
350pf
2000pf
C2

[http://4.bp.blogspot.com/-7vL-qro9wfs/Vo_SZrtVTII/AAAAAAAACL4/tSXuK3VUGWg/s1600/Screen%2BShot%2B2016-01-

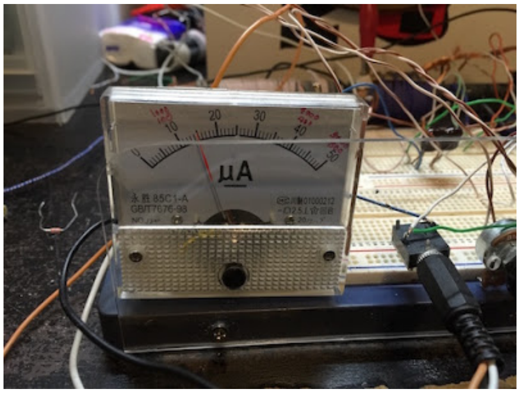

08%2Bat%2B11.10.35%2BPM.png]
[http://3.bp.blogspot.com/-mx7IGYUxaOw/Vo-YOMPvw3I/AAAAAAAACLU/09B7iPtbAok/s1600/IMG_5238.JPG]

Then I tried this dual Diode circuit invented by a member of the Chinese Crystal Radio Forum.
I can boost the DC current to 250uA.

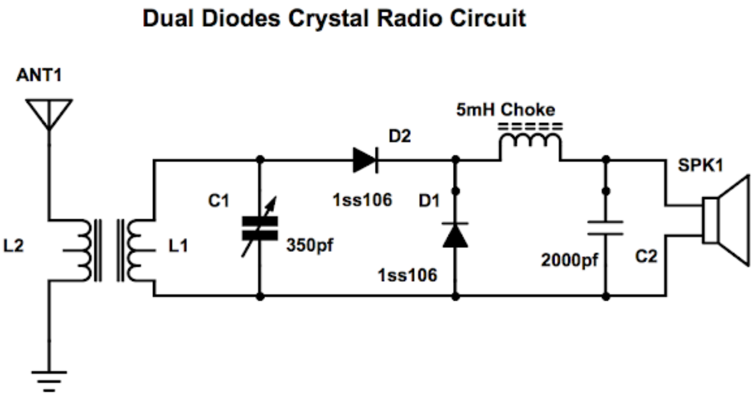
Dual Diodes Crystal Radio Circuit
ANT1
5mH Choke
D2
SPK1
C1
1ss106
D1
L2
L1
350pf
1ss106
2000pf
C2

[http://3.bp.blogspot.com/-U1KM5TT_tmY/Vo_SZtyKwiI/AAAAAAAACL0/BhICasprNkA/s1600/Screen%2BShot%2B2016-01-08%2Bat%2B11.12.15%2BPM.png]

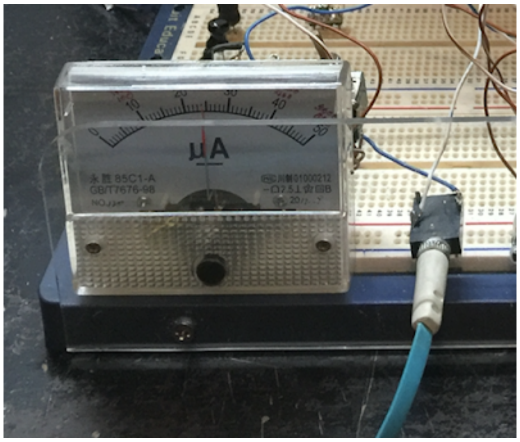

[http://1.bp.blogspot.com/-a6Ce5T2NlVE/Vo_TowTTU_I/AAAAAAAACME/q4AB4TF8h04/s1600/Screen%2BShot%2B2016-01-08%2Bat%2B11.19.40%2BPM.png]

Here is the video clip.

https://youtu.be/5NWsHX0kwe4 [https://youtu.be/5NWsHX0kwe4]

Here are the pictures.

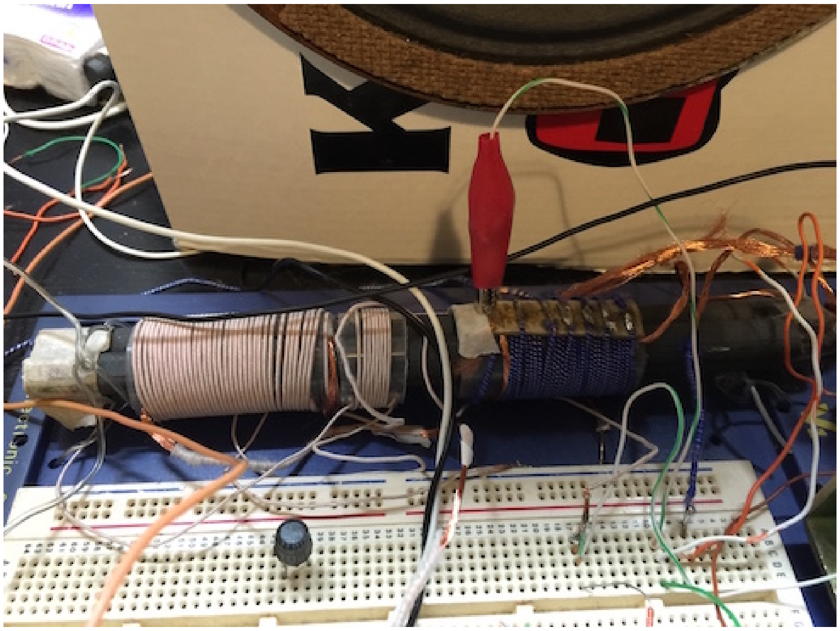

[http://3.bp.blogspot.com/-_MQIkzFXm8c/Vo-YNmmP9UI/AAAAAAAACLA/iaojC2ejAGY/s1600/IMG_5235.JPG]

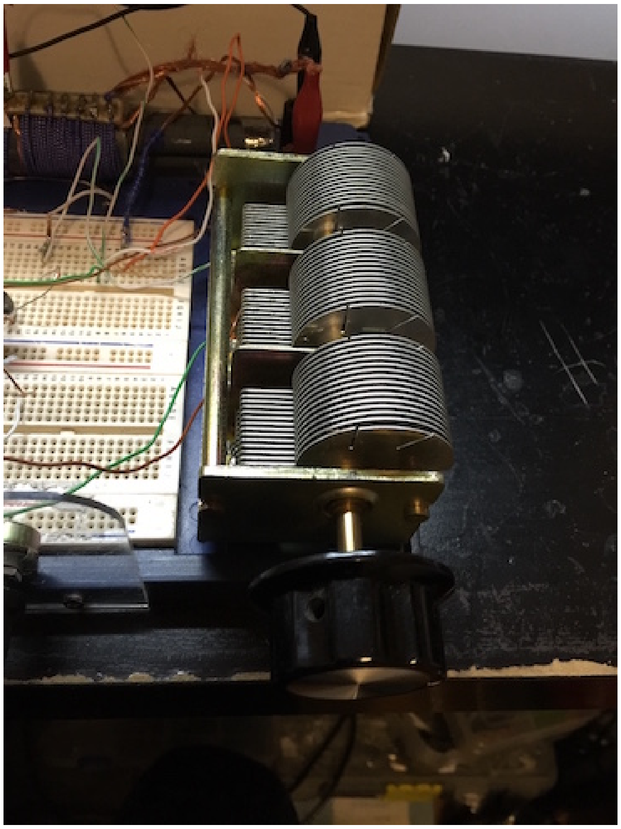

[http://4.bp.blogspot.com/-UA8-ZheAors/Vo-YNxDyPOI/AAAAAAAACLI/MLdYtAffWug/s1600/IMG_5236.JPG]

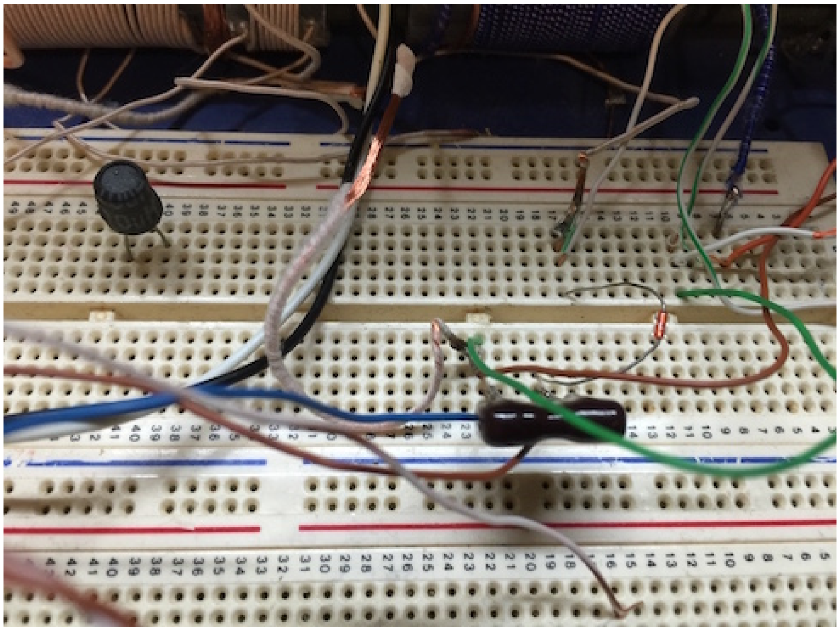

[http://4.bp.blogspot.com/-vNli3inruSI/Vo-YOG3VZbI/AAAAAAAACLM/kB6a8qr6KI0/s1600/IMG_5237.JPG]

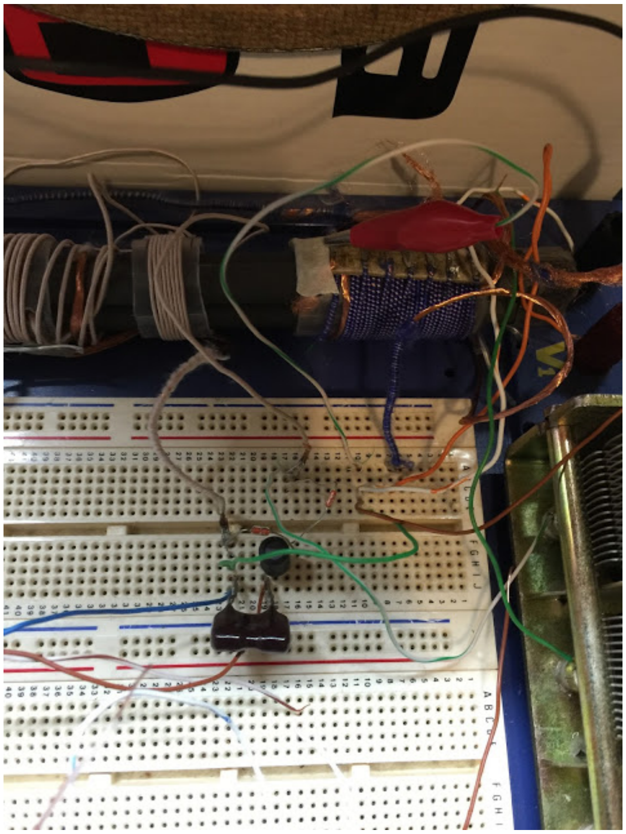

[http://2.bp.blogspot.com/-Hf461kmCtZQ/Vo_Ty7d7bgI/AAAAAAAACMM/Y6_jR7byO4g/s1600/IMG_5243.JPG]

Posted 8th January 2016 by Billy

1 View comments

Dynamic Views theme. Powered by Blogger.

Bitdefender -56% Be aware without a care! BUY NOW

31st December 2015

SC2-300 foldable headphone

Happy New Year to you.
I decided to give myself a new headphone by converting a faulty monster head phones to crystal radio headphone with two SC2-300 Reed ear pieces.
After fixing the sponge to the headphone, it can be worn comfortably, with good audio effect. The impedance is around 600 ohm when two ear pieces are in series.

[http://2.bp.blogspot.com/-ONbscrYC_Eg/VoU_rbE7LPI/AAAAAAAACKc/7ZyPRXGy_eo/s1600/IMG_5112.JPG]

[http://1.bp.blogspot.com/-D2lwLDJ-C20/VoU_pjOotWI/AAAAAAAACJk/vPqich72BHo/s1600/IMG_5091.JPG]

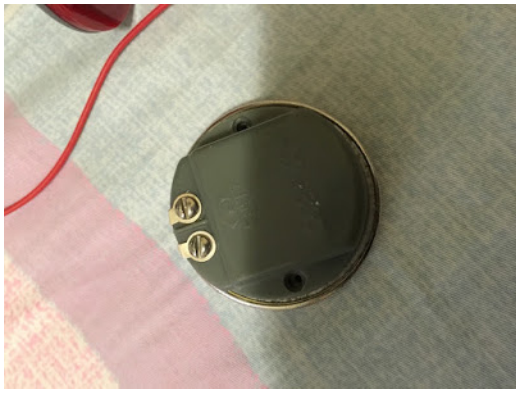

[http://2.bp.blogspot.com/-xGZCytknWpg/VoU_pjucwMI/AAAAAAAACJs/cStczNucRYA/s1600/IMG_5092.JPG]

Dynamic Views theme. Powered by Blogger.

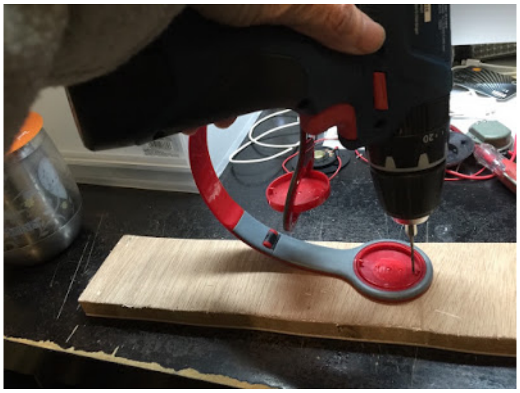

[http://1.bp.blogspot.com/-BFqomaYZf6g/VoU_pgOft9I/AAAAAAAACJo/niMeaNLLiBU/s1600/IMG_5093.JPG]

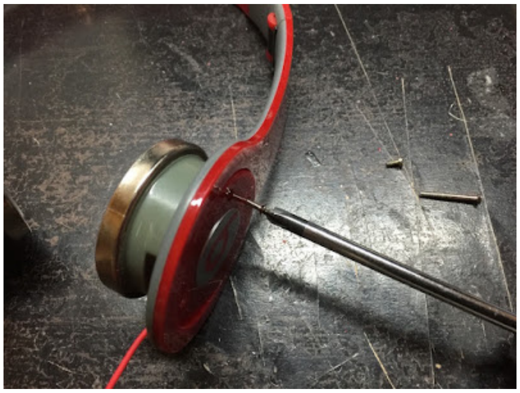

[http://3.bp.blogspot.com/-LIZMrsR_NiQ/VoU_qEmNp3I/AAAAAAAACJw/UpPIGCBWVWE/s1600/IMG_5094.JPG]

[http://2.bp.blogspot.com/-hNG7fMXGmHA/VoU_qfPWDPI/AAAAAAAACJ0/IsTUQql5Qig/s1600/IMG_5095.JPG]

[http://2.bp.blogspot.com/-wIhWI_E7KhE/VoU_qjfIAgI/AAAAAAAACJ8/7EQtnq5-QbA/s1600/IMG_5096.JPG]

Dynamic Views theme. Powered by Blogger.

[http://2.bp.blogspot.com/-blgYDkRvxt8/VoU_q01xIXI/AAAAAAAACKI/EcNl2zV_i2Y/s1600/IMG_5097.JPG]

[http://4.bp.blogspot.com/-08vmVzZn0Bs/VoU_qwrPlbI/AAAAAAAACKM/FNEjmbPRPHc/s1600/IMG_5110.JPG]

[http://4.bp.blogspot.com/-JKmdXC9-Qno/VoU_rd3ch1I/AAAAAAAACKU/0s61eydDBMM/s1600/IMG_5111.JPG]

Posted 31st December 2015 by Billy

0 Add a comment

10th October 2015

How to clean install OSX to a Fusion drive and keep the recovery hard disk function

A. Clean reinstall of OSX.

1. Make sure you back up all your data on the hard disk, as the following step will erase all data.
2. Create the OSX Install USB drive, insert it to the MAC workstation.
3. Since we will operate on the internal boot up drive of the Mac, we need to boot to OSX Install USB drive (See X.)

4. At the menu at the top of the screen, select disk utilities.
5. Highlight the Fusion Drive volume called "Macintosh HD". Click Erase to clean it up.
6. If its successful, then go ahead to install new OSX to this partition as usual.
7. Otherwise if this failed, that means fusion drive has been damaged.
8. Follow onto Be to fix it.

B. Split the Fusion Drive into physical hard disks.

1. Since we will operate on the internal boot up drive of the Mac, assumes that we've already boot up to OSX Install USB drive (See Y.).
2. At the menu at the top of the screen , select Utilities and Terminal.
3. On the terminal screen type:

   ```
   diskutil cs list.
   ```

   You will see something similar to the screen below.

4. Copy the long string after Logical Volume and replace UUID with it in the following command to delete the fusion drive logical volume (aka coreStorage logical volume) :

   ```
   diskutil  cs deletevolume  UUID
   ```

   For example:

   ```
   diskutil cs deletevolume E59B5A99-F8C1-461A-AE54-6EC11B095161
   ```

5. Copy the long string after Logical Volume Group and replace UUID with it in the following command to delete the fusion drive (aka coreStorage) :

   ```
   diskutil  cs delete UUID
   ```

   e.g. `diskutil cs delete E03B3F30-6A1B-4DCD-9E14-5E927BC3F5DC`

6. At this step, the fusion drive has been deleted, and the SSD hardsisk and the mechanical hard disk will be reappearing in diskutil as separate disks.

C. Install a fresh copy of OSX to the SSD hard disk and test that recovery hard disk partition works.

1. Go to diskutil to create a partition called Machintosh HD in the SSD hard disk using all the space there.
2. Do the same with the mechanical hard disk.
3. Follow the usual process to boot from the OSX install USB and install a fresh copy of OSX to the SSD hard disk.
4. This will create the proper boot up partitions, recovery hard disk partitions and the PSX partitions on the SSD hard disk.
5. Once the install is complete, test if OSX can start properly, but no need to go through the initial MAC OSX set up as we'll to scrap this and do the install once again later.
6. We now need to test if the recovery hard disk partition works.
7. Reboot to recovery hard disk (see X - by pressing and holding Command and R at the same time during boot up right after you heard the boot up sound, and only release it a few seconds after you see the apple logo and the progress bar for loading..)
8. This is important to test and make sure the recovery hard disk partition .

D. Recreate the fusion drive

1. Since we will operate on the internal boot up drive of the Mac, we need to boot to OSX Install USB drive (See Y.).
2. At the menu at the top of the screen , select Utilities and Terminal.
3. On the terminal screen type:

You should found that we have a list of physical hard disks only, no logical volume group there yet.

.

4. You will see something similar to the screen below.

5. Look for the largest partition on the SSD hard disk that should be close to the full hard disk size of the SSD (e.g. 121G), and mark the device name, this will usually be something like /dev/disk0s2

6. Look for the largest partition on the mechanical disk that will form the fusion drive with the SSD hard disk. This should be close to the full hard disk size of the mechanical hard disk (e.g. 3TB), and mark the device name, this will usually be something like /dev/disk1s2

7. Now we will create the fusion drive group (a.k.a. logical volume group) by typing this command: **diskutil cs create *lvgName device1 device2***

   For example:

   ```
   diskutil cs create fusiondrive  /dev/disk0s2  /dev/disk1s2
   ```

   You will see something like below showing up on the screen:

   ```
   Creating Core Storage Logical Volume Group

   Switching d isk0s2 to Core Storage

   Switching disk1s2 to Core Storage

   Waiting for Logical Volume Group to appear

   Discovered new Logical Volume Group "DBFEB690-107B-4EA6-905B-2971D10F5B53"

   Core Storage LVG UUID: DBFEB690-107B-4EA6-905B-2971D10F5B53

   Finished CoreStorage operation
   ```

8. Copy down the string after “Discovered new Logical Volume Group” using command + C

9. Next, create the fusion drive partition (a.k.a. logical volume) named “Macintosh HD” by typing this command:

   Diskutil cs createVolume *lvgUUID type name size*

   For example:

   ```
   Diskutil cs createVolume DBFEB690-107B-4EA6-905B-2971D10F5B53 jhfs+ “Macintosh HD” 100%
   ```

10. Go to diskutil to verify that you can see this newly created partition on the list.

11. Test by erasing all data from it.

12. Then you can go ahead to boot to the OSX Install USB drive and install a fresh copy of OSX onto it.

13. This will allow you to keep the recovery hard disk function.

X. How to boot to recover partition

Press and hold the command and R key together immediately after hearing the boot up bell sound.

Only release it 2 seconds after you see the Apple logo on the screen and the progress bar for booting . This will boot to recovery harddisk partition.

Y. How to boot to OSX Install USB

Press and hold the command and Option key together immediately after hearing the boot up bell sound.

Only release it 2 seconds after you see the Apple logo on the screen and This will provide you with a list of boot up disk to choose from, choose the OSX install USB to boot from.

4 View comments

4th October 2015 My first Vacuum Tube Radio - a 26V+1.2V three tubes region.

I built My first Vacuum Tube Radio - a 26V+1.2V three tubes region. radio.

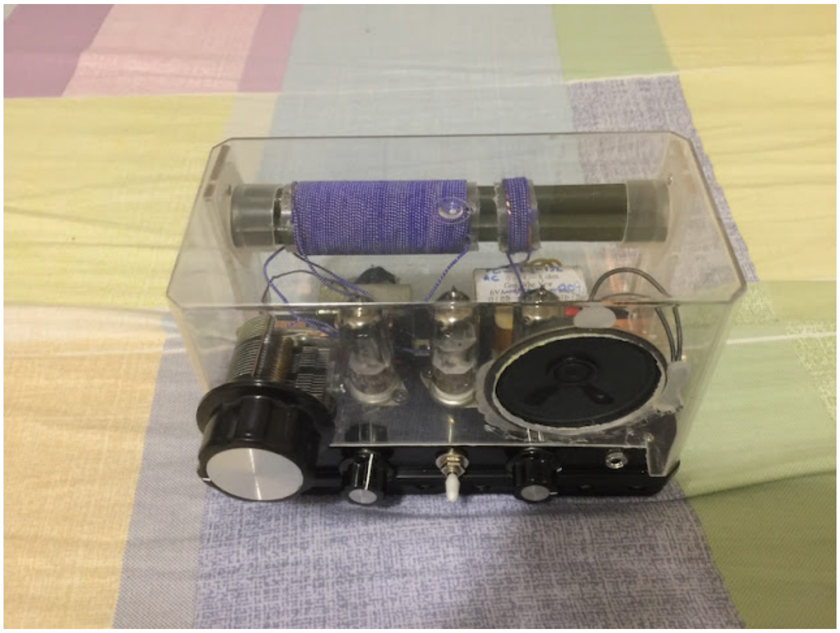

[http://1.bp.blogspot.com/-3Ay3MzXEPZQ/VhEscqD3JMI/AAAAAAAACGU/K0Js38_4W6Y/s1600/IMG_3123.JPG]

I spent 1 month collecting the components. The Chinese Vacuum tubes(1B2, 2P2) can be purchased from the Chinese Taobao shops @ half a USD each. The two transformers are 220V to 6V transformers. Capacitors and resistors are the standard ones.
However, shipping them from China to HK is a bit of headache. There are more and more stringent restrictions for exporting of goods from China. I spent a while finding a courier service that can accept these goods and reship to Hong Kong.
Shipping fees starts from USD 3 per KG. But special goods like these may take USD 6 per KG.

Then I cared too much about how the finished product looks. Spent 2 weeks finding the nice plastic box that is transparent enough.
Finally it all took place in the last few days. The initial circuit was designed by another Chinese hobbyist. The only modification I made is use a bigger coil (litz wire 0.04x220, instead of 0.04x60, and three ferrite rods instead of one), to make it more sensitive.

The negative feed back is somewhat too strong, I had to remove a few turns from the regen coil.

Here is how the finishing produce looks on the video.

https://www.youtube.com/watch?v=lQ4Kv42QWsU [https://www.youtube.com/watch?v=lQ4Kv42QWsU]

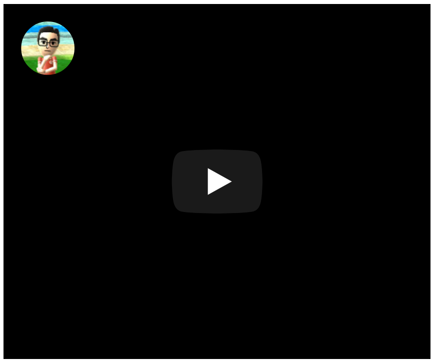

Dynamic Views theme. Powered by Blogger.

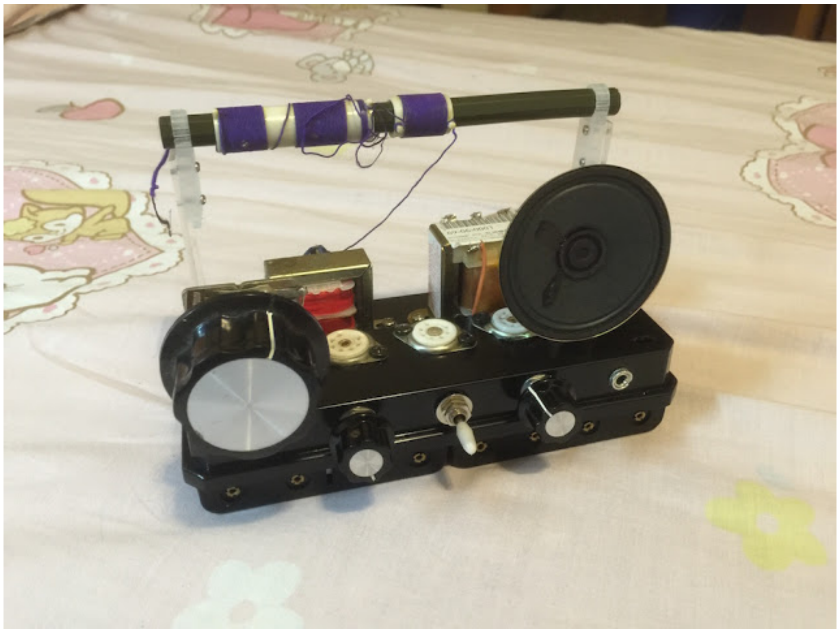

[http://4.bp.blogspot.com/-F6bAQQaTAQM/VhEsqacp5AI/AAAAAAAACGk/hg80VcGurc0/s1600/IMG_3098.JPG]

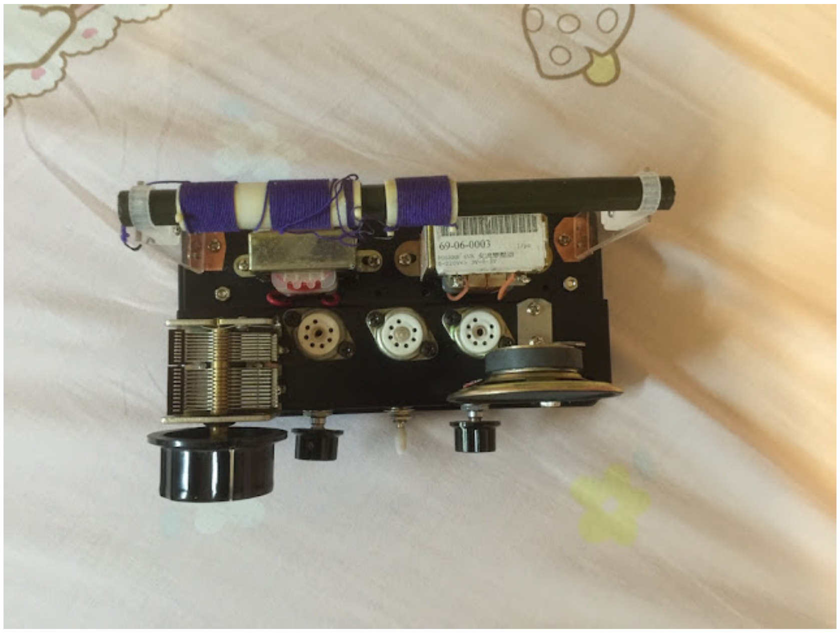

[http://3.bp.blogspot.com/-ODjkXJ5qWBk/VhEsqVsopnI/AAAAAAAACGg/-5lCJE-eFCw/s1600/IMG_3099.JPG]

Dynamic Views theme. Powered by Blogger.

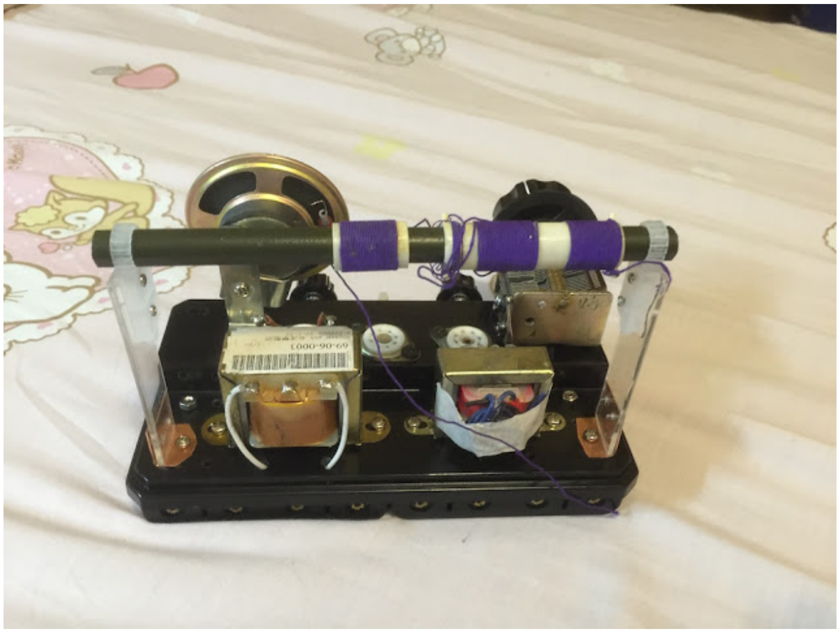

[http://4.bp.blogspot.com/-q6bJ4qaJ6a8/VhEsqc0_KXI/AAAAAAAACGc/6T0ZjVNM-vY/s1600/IMG_3100.JPG]

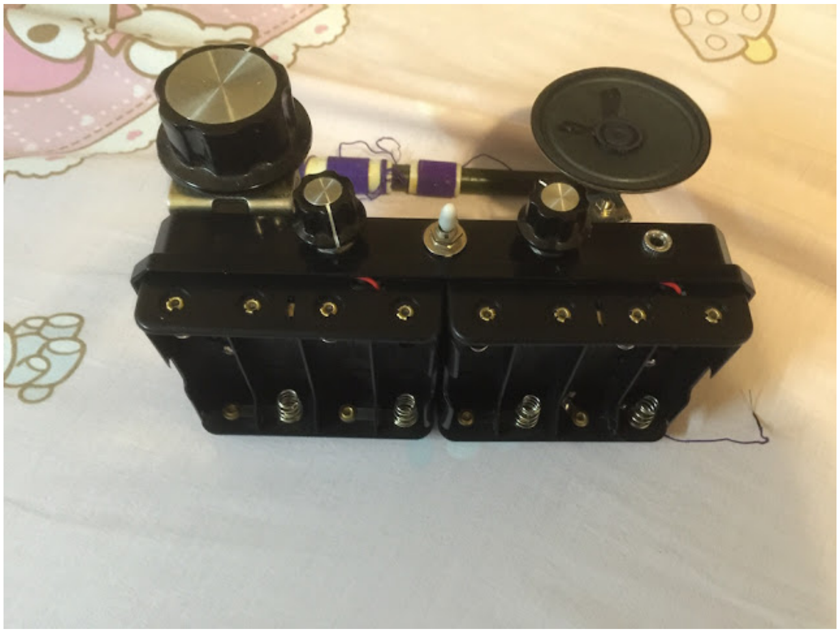

[http://3.bp.blogspot.com/-tua4wayNags/VhEsq6lI3LI/AAAAAAAACGs/Nc3OpalKtYU/s1600/IMG_3101.JPG]

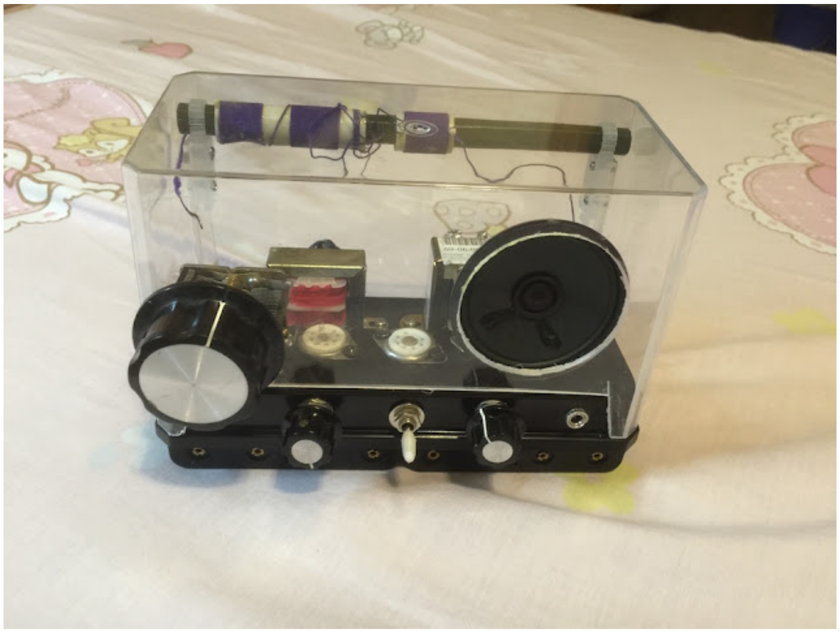

[http://3.bp.blogspot.com/-grtU4sfLqKw/VhEsrUR1Z9I/AAAAAAAACGw/eFV9Ny8QA40/s1600/IMG_3102.JPG]

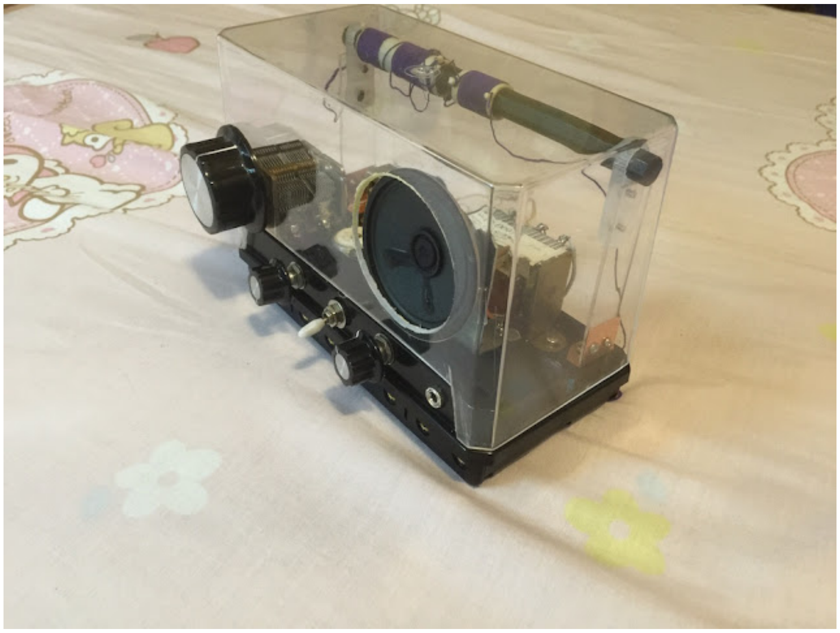

[http://3.bp.blogspot.com/-yiyniotReSI/VhEsrW2IOdI/AAAAAAAACG4/54R2ncPq-i8/s1600/IMG_3103.JPG]

Dynamic Views theme. Powered by Blogger.

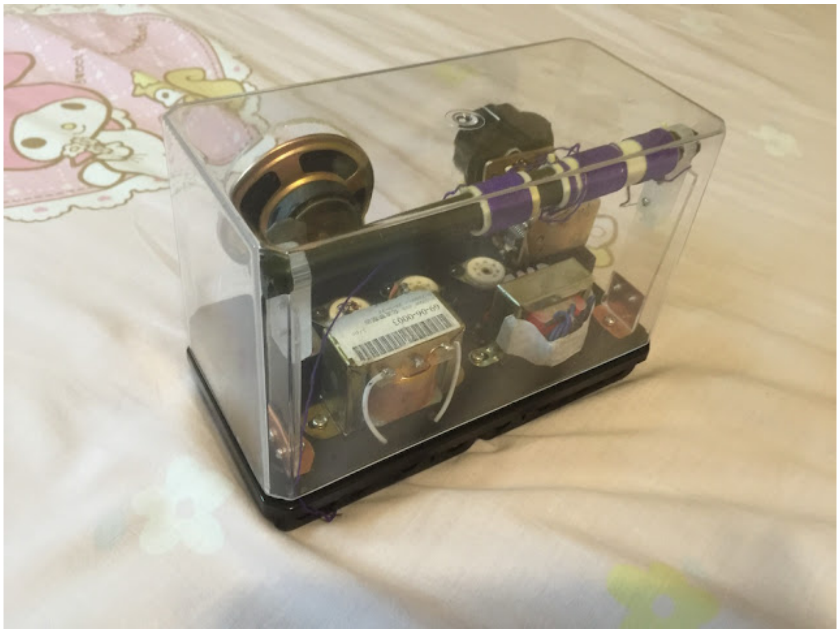

[http://3.bp.blogspot.com/-9JWUNauyzSw/VhEsrkJlZpI/AAAAAAAACHA/mCR4roQI-es/s1600/IMG_3104.JPG]

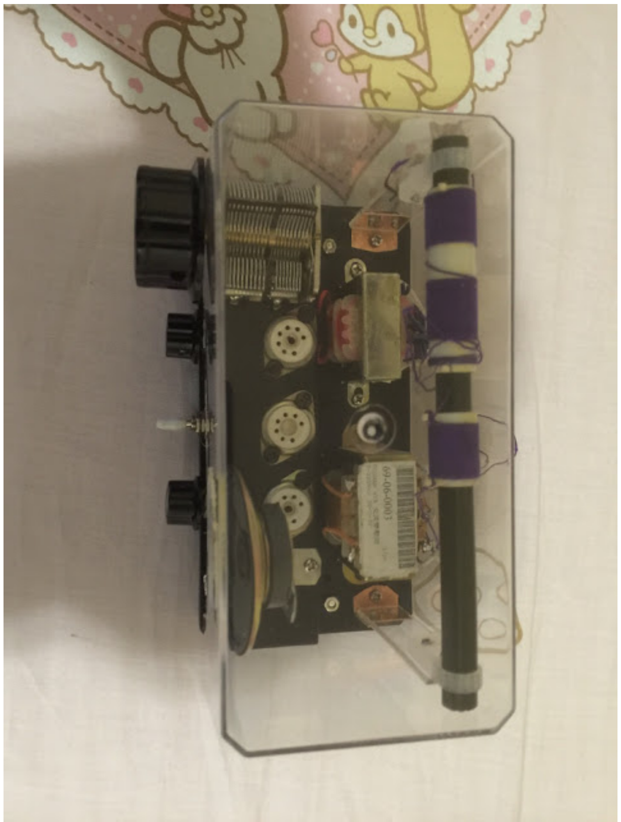

[http://4.bp.blogspot.com/-vGVAC90zic4/VhEsr-CTjqI/AAAAAAAACHE/0Wrw7nephK0/s1600/IMG_3106.JPG]

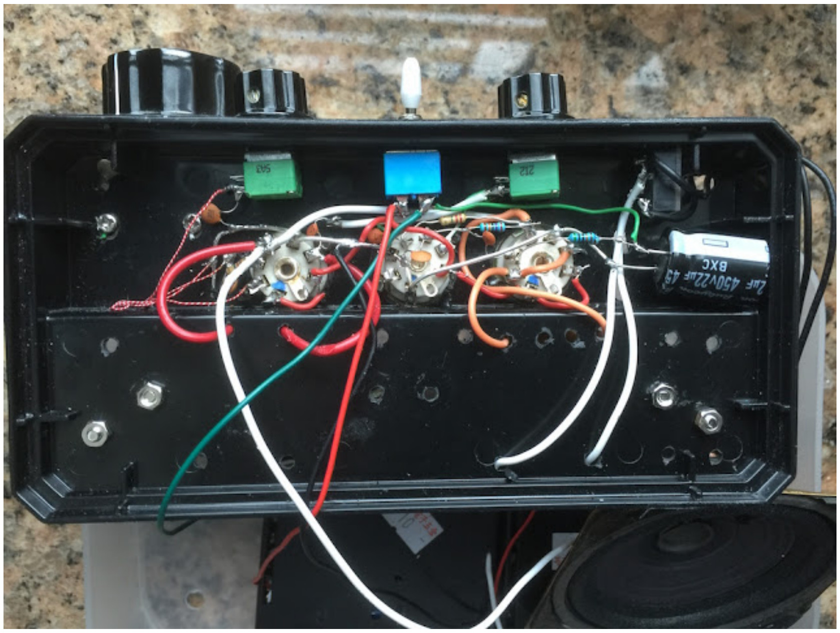
5U3
2I2
22uF 450V22uF 45
BXC

[http://3.bp.blogspot.com/-Ipc4OJKkKiM/VhEssfJHb1I/AAAAAAAACHQ/_V6GeVrnlAE/s1600/IMG_3113.JPG]

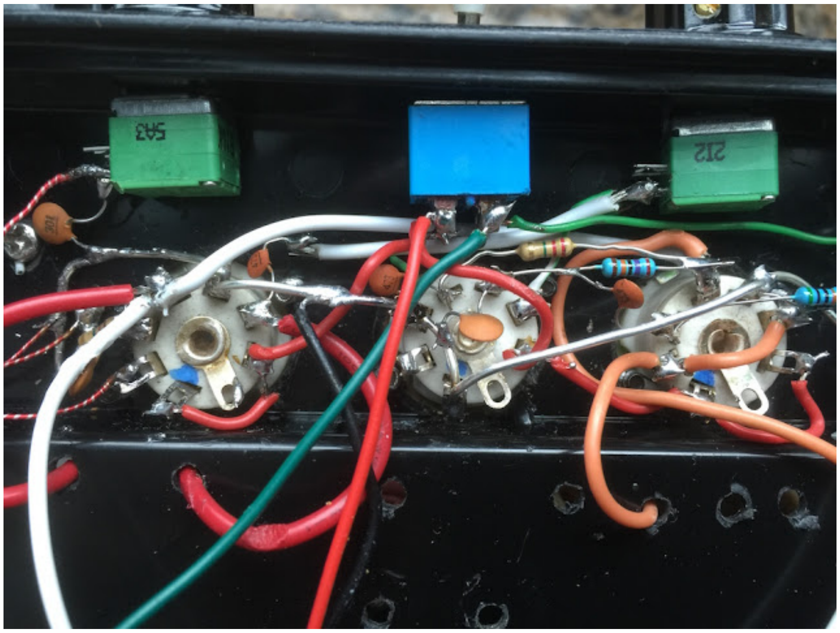

[http://4.bp.blogspot.com/-M8qvF7FFc2c/VhEsse_TeJI/AAAAAAAACHU/2CagOzQgNwE/s1600/IMG_3114.JPG]

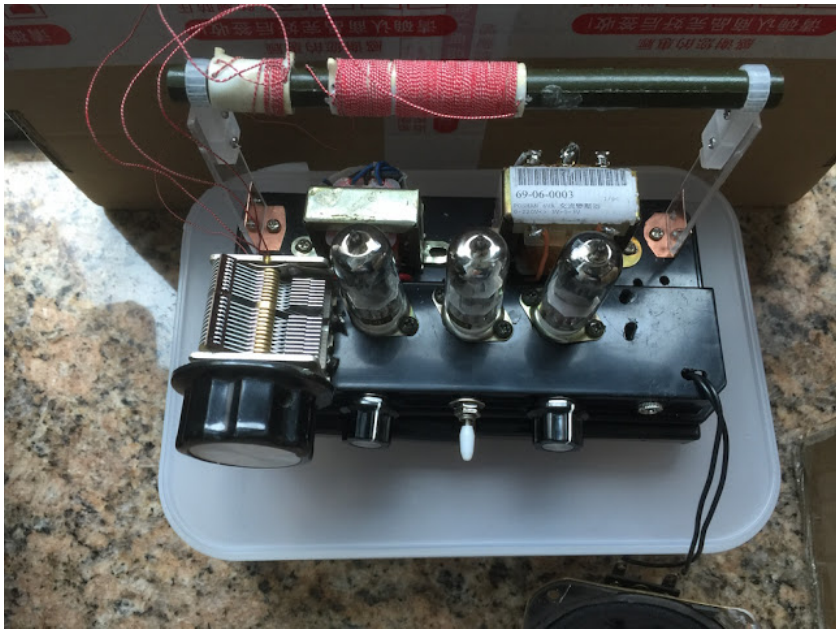

[http://4.bp.blogspot.com/-6BS6OJk8EW8/VhEssqM4dMI/AAAAAAAACHc/ABSuYQiWNl0/s1600/IMG_3119.JPG]

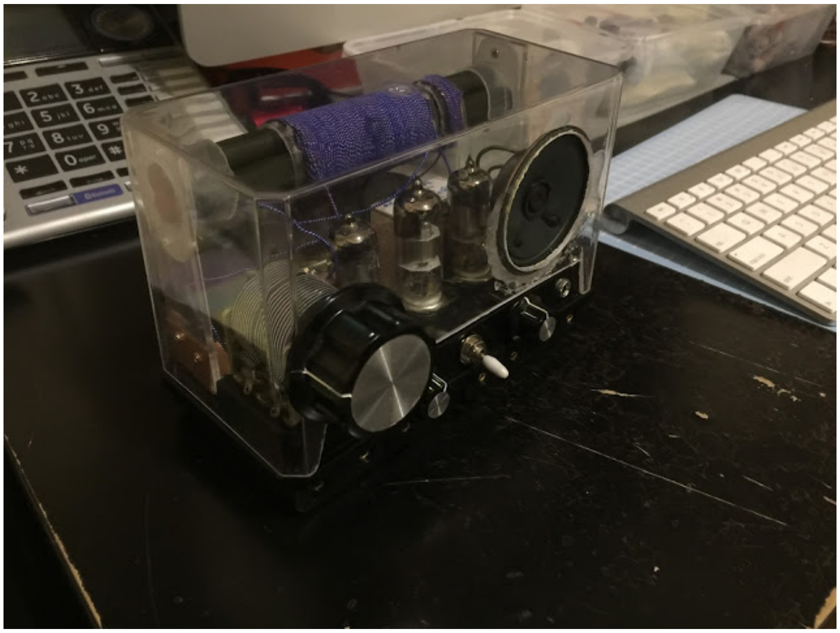

[http://2.bp.blogspot.com/-UggcM_NWWng/VhEstLiJbvI/AAAAAAAACHs/fZBz9gWlhT0/s1600/IMG_3121.JPG]

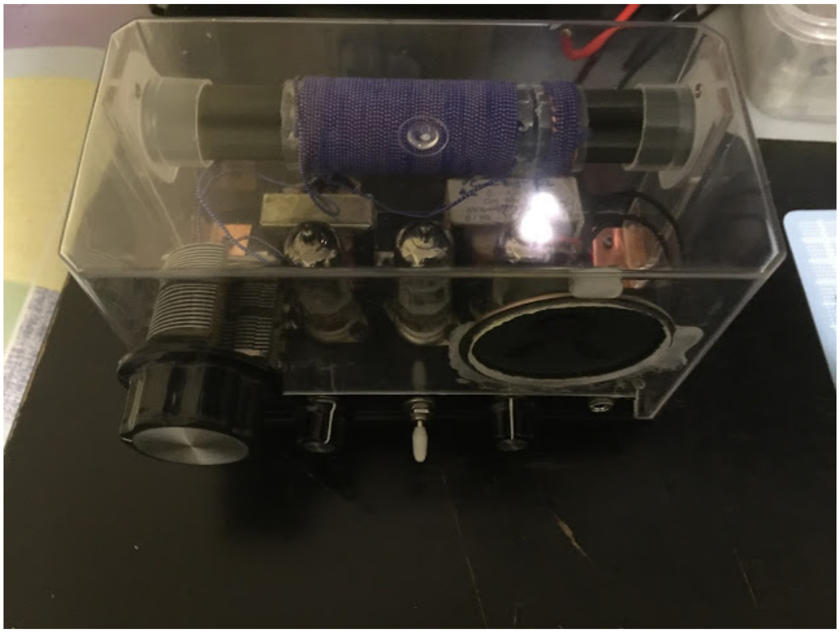

[http://1.bp.blogspot.com/-0L9IJP5s10o/VhEstMBCRPI/AAAAAAAACHw/OtVArSqDD34/s1600/IMG_3122.JPG]

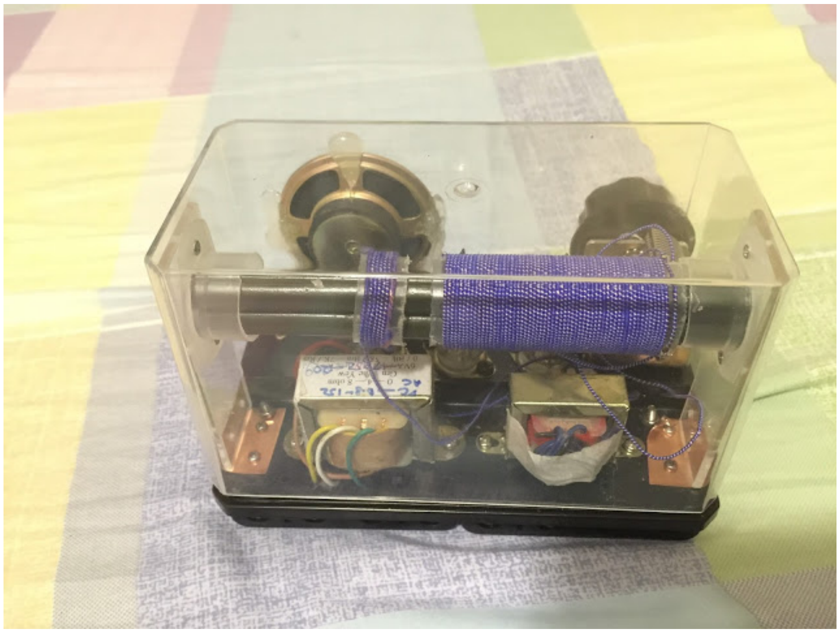

[http://4.bp.blogspot.com/-nIjIuIe1FdQ/VhEstVzPk9I/AAAAAAAACH4/QXpi0aN_d3Q/s1600/IMG_3124.JPG]

Dynamic Views theme. Powered by Blogger.

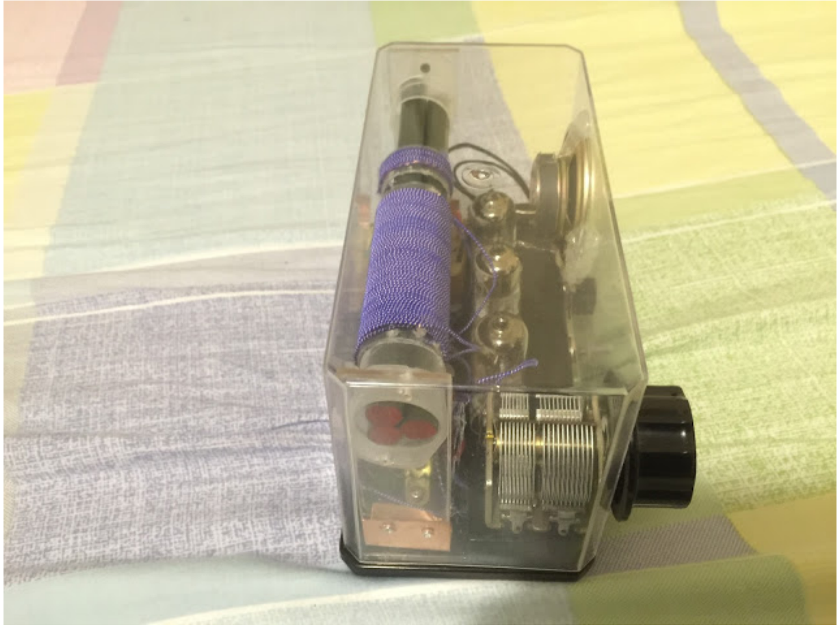

[http://2.bp.blogspot.com/-IhkWTJLZPbM/VhEst9iG3CI/AAAAAAAACIE/6fAm0jQqpjQ/s1600/IMG_3125.JPG]

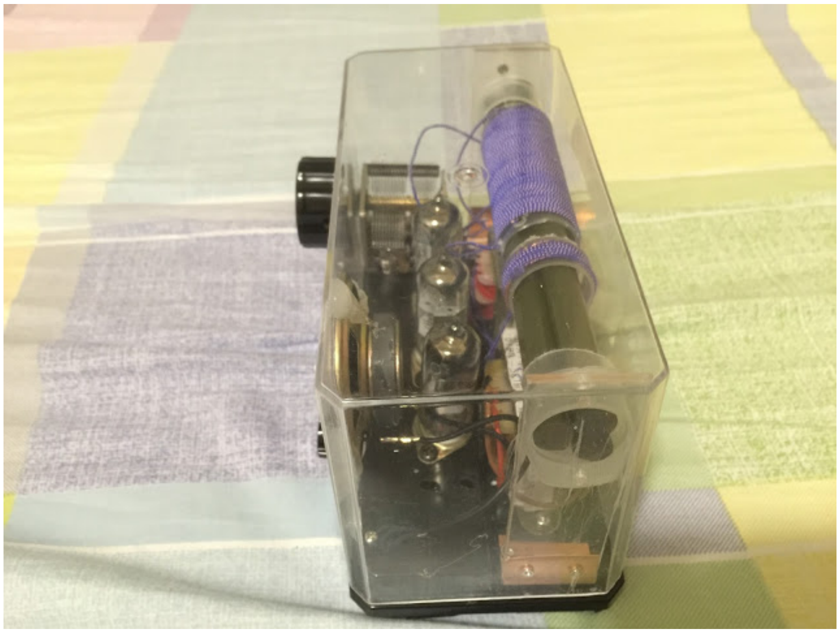

[http://2.bp.blogspot.com/-xIqanrXgukI/VhEsuIrG8JI/AAAAAAAACII/bq0e_LmaSxA/s1600/IMG_3126.JPG]

Dynamic Views theme. Powered by Blogger.

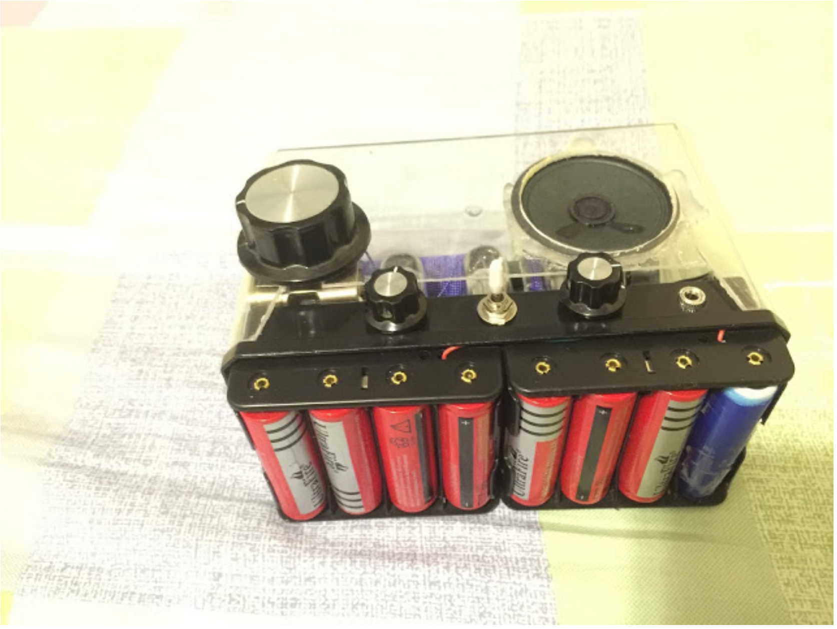

[http://3.bp.blogspot.com/-NSU-WufMqXA/VhEsuYAQuqI/AAAAAAAACIU/tz5AxvQtReY/s1600/IMG_3127.JPG]

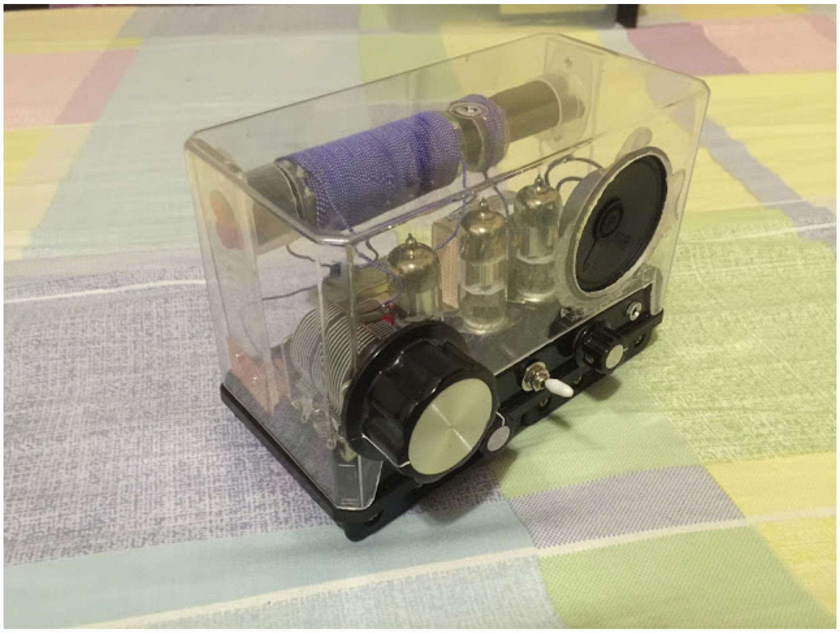

[http://1.bp.blogspot.com/-CBRHk3qv2hU/VhEsu4Ai0aI/AAAAAAAACIc/XCLs_KNa27I/s1600/IMG_3128.JPG]

[http://2.bp.blogspot.com/-bu9mzDFs3hA/VhEsu6IlISI/AAAAAAAACIY/FpQxMRqSFFU/s1600/IMG_3145.JPG]

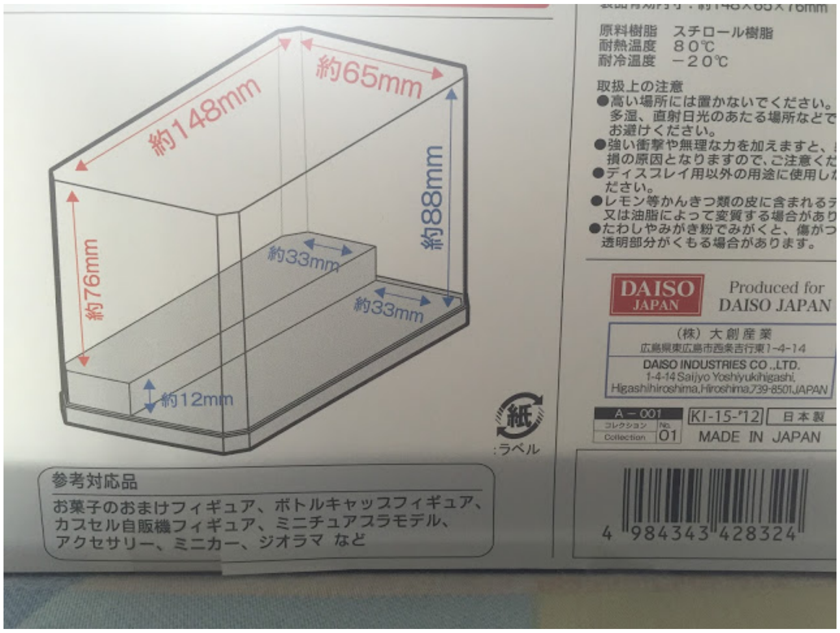
約148mm
約65mm
約88mm
約76mm
約33mm
約33mm
約12mm
原料樹脂 スチロール樹脂
耐熱温度 80℃
耐冷温度 −20℃
取扱上の注意
DAISO JAPAN
Produced for DAISO JAPAN
(株) 大創産業
DAISO INDUSTRIES CO.,LTD.
KI-15-12
日本製
MADE IN JAPAN
4 984343 428324
参考対応品
お菓子のおまけフィギュア、ボトルキャップフィギュア、
カプセル自販機フィギュア、ミニチュアプラモデル、
アクセサリー、ミニカー、ジオラマ など

[http://3.bp.blogspot.com/-86wt3ZxOy-8/VhEsvj5CH2I/AAAAAAAACIw/ZeATLLRVnzs/s1600/IMG_3148.JPG]

Posted 4th October 2015 by Billy

1 View comments

SQUARESPACE WEBSITES & ONLINE STORES FOR ENTREPRENEURS START YOUR FREE TRIAL

26th September 2015

Chinese Lantern Crystal Radio

Tomorrow is the mid-Autumn Festival that the Chinese celebrate the festival of the moon.
Children will light up their lanterns to shine through the parks.

What if that lantern can become a crystal radio.
Not only that, the candle light can energise the solar cells from the calculators to bring some small power to drive a regenerative radio to boost the speakers.

I made a lanterns and wind a number of coils to feed to the MOS FET 3DQ.
Next step is to find some solar cells. Those that will work with candle lights.

Here is what I made so far.

Now the crystal radio part is working well. I can listen to most of the local broadcast when I bring the lantern onto the street.
This does attract some people to stop by and listen to my crystal radio.

For the regenerative radio part, I need to find the solar cells to complete the project.

Video : https://youtu.be/F6e2em1ZVes [https://youtu.be/F6e2em1ZVes]

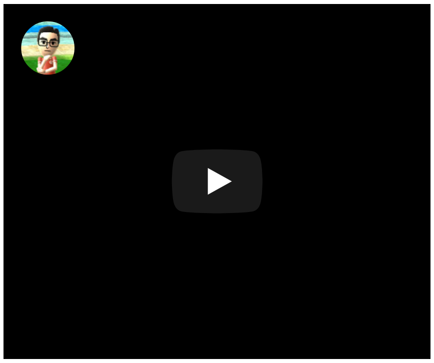

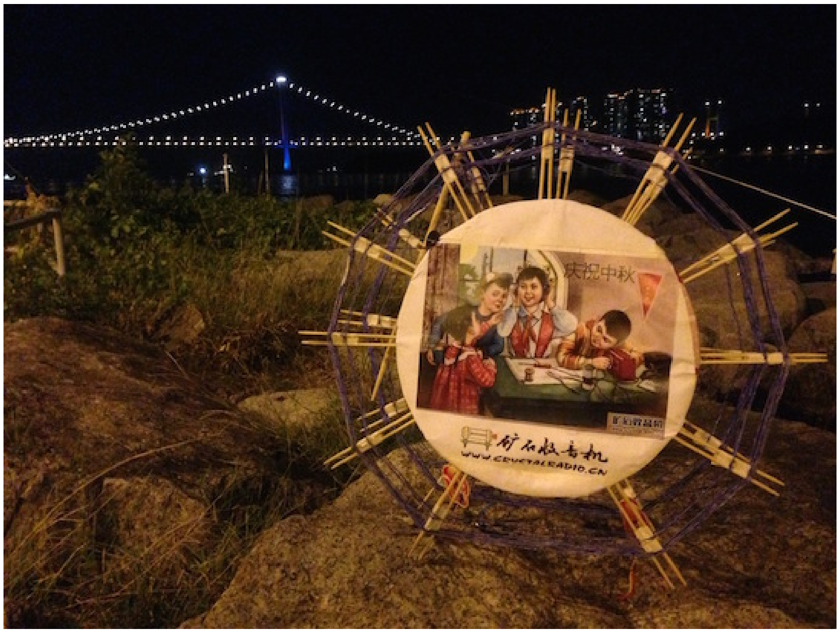

[http://4.bp.blogspot.com/-r_58rL-_tVY/VgZ3sEYCDpI/AAAAAAAACF0/iCpYzqHG2Y0/s1600/IMG_8749.JPG]

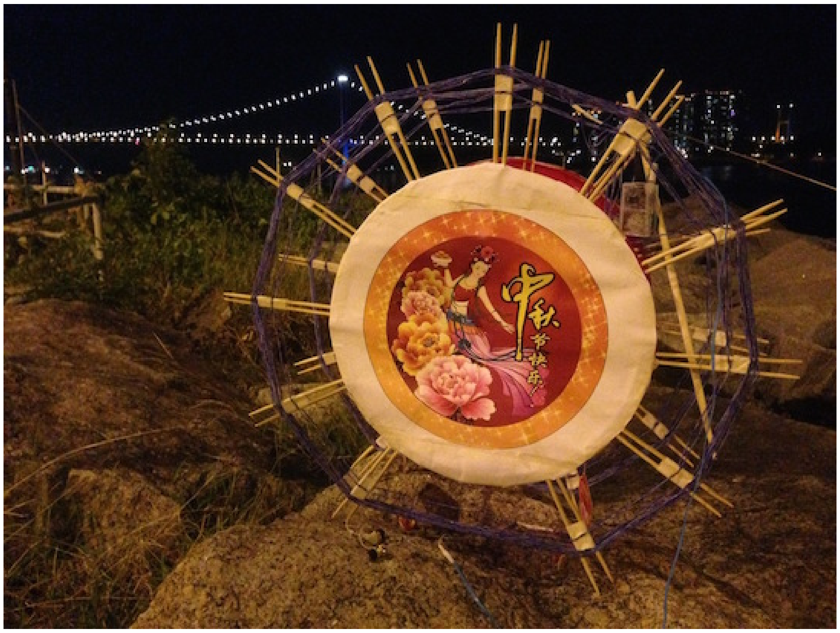

[http://4.bp.blogspot.com/-MOB8VED5z90/VgZ3sNwxx3I/AAAAAAAACF4/4PcvAwPZbjs/s1600/IMG_8748.JPG]

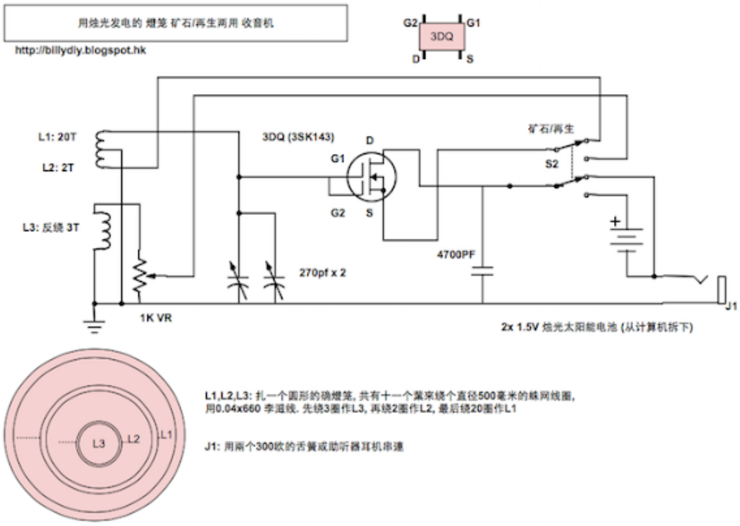

用烛光发电的 燈笼 矿石/再生两用 收音机

http://billydiy.blogspot.hk

G2
G1
3DQ
D
S
L1: 20T
L2: 2T
L3: 反绕 3T
3DQ (3SK143)
D
G1
G2
S
矿石/再生
S2
4700PF
270pf x 2
+
1K VR
J1
2x 1.5V 烛光太阳能电池 (从计算机拆下)
L3
L2
L1

L1,L2,L3: 扎一个圆形的确燈笼, 共有十一个桨来绕个直径500毫米的蛛网线圈, 用0.04x660 李兹线. 先绕3圈作L3, 再绕2圈作L2, 最后绕20圈作L1

J1: 用两个300欧的舌簧或助听器耳机串連

[http://3.bp.blogspot.com/-QHDOFA6pvmk/VgZ1YkAJgfI/AAAAAAAACDw/Xkbf559vDpA/s1600/Circuit%2BDiagram.png]

Dynamic Views theme. Powered by Blogger.

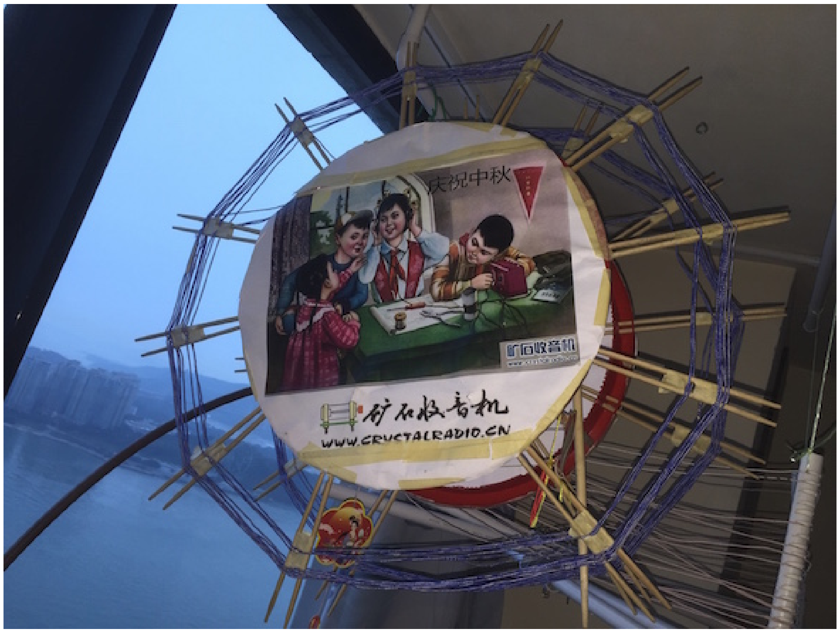
庆祝中秋
矿石收音机
矿石收音机
WWW.CRYSTALRADIO.CN

[http://1.bp.blogspot.com/-dU19vBBcjtY/VgZ1Yk0WlKI/AAAAAAAACDs/xvoaJJhQpYw/s1600/IMG_3013.JPG]

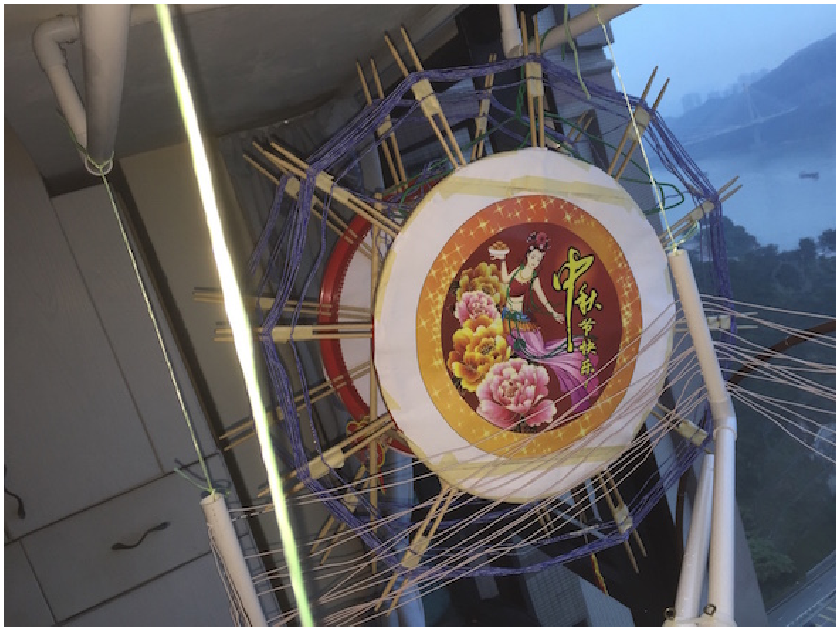

[http://3.bp.blogspot.com/-J0kG8aA1LkA/VgZ1Yl8MoUI/AAAAAAAACD0/t3Bpx3citNc/s1600/IMG_3014.JPG]

[http://1.bp.blogspot.com/-Vgr6tgNhR-M/VgZ1ZO1qXbI/AAAAAAAACD4/iHJJX-pvBM4/s1600/IMG_8718.JPG]

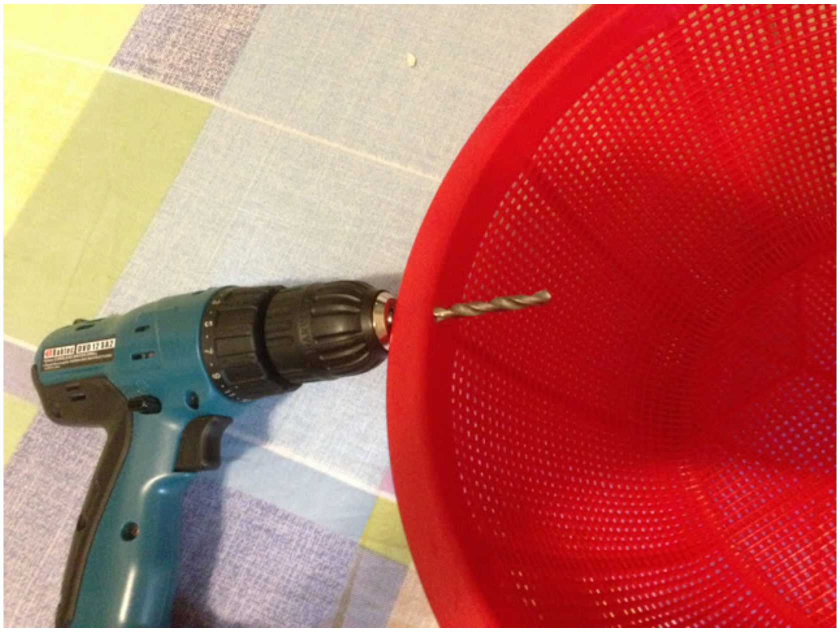

[http://1.bp.blogspot.com/-Q-a8jmoVpGE/VgZ1kCKamHI/AAAAAAAACFg/veQMnV87tGI/s1600/IMG_8719.JPG]

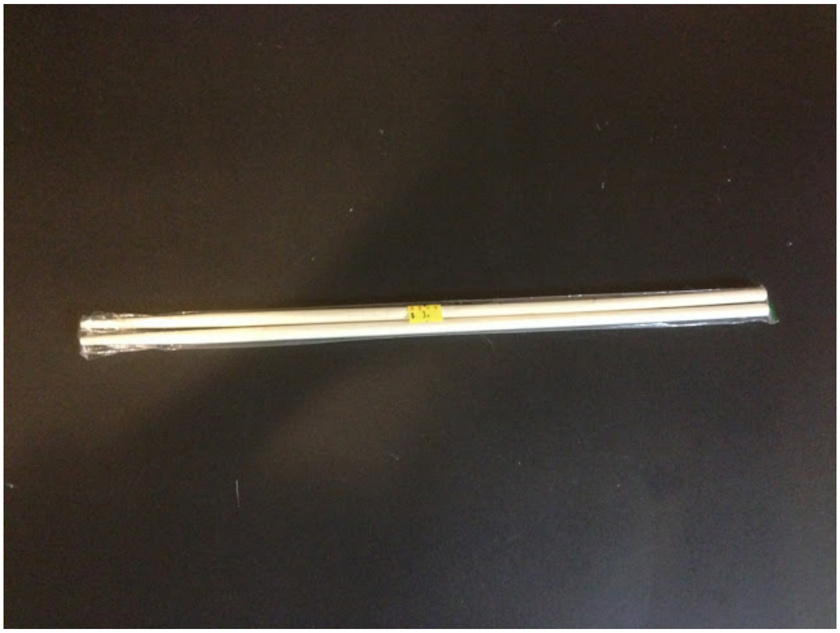

[http://3.bp.blogspot.com/-S5E395MoRws/VgZ1lfLmyII/AAAAAAAACFk/c0oXg_Vwwkc/s1600/IMG_8725.JPG]

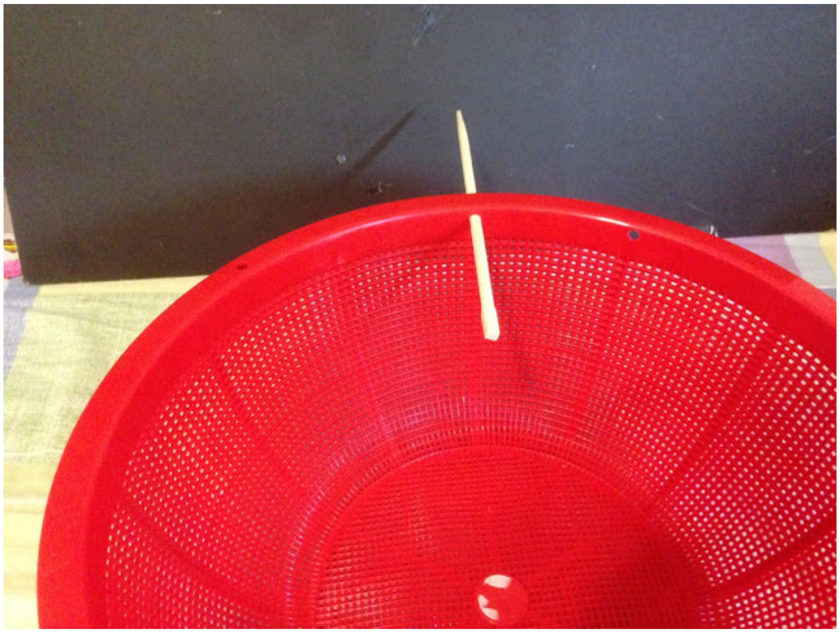

[http://3.bp.blogspot.com/-vuahPxay4v0/VgZ1kKCAhlI/AAAAAAAACFI/NeU60oVK6CA/s1600/IMG_8720.JPG]

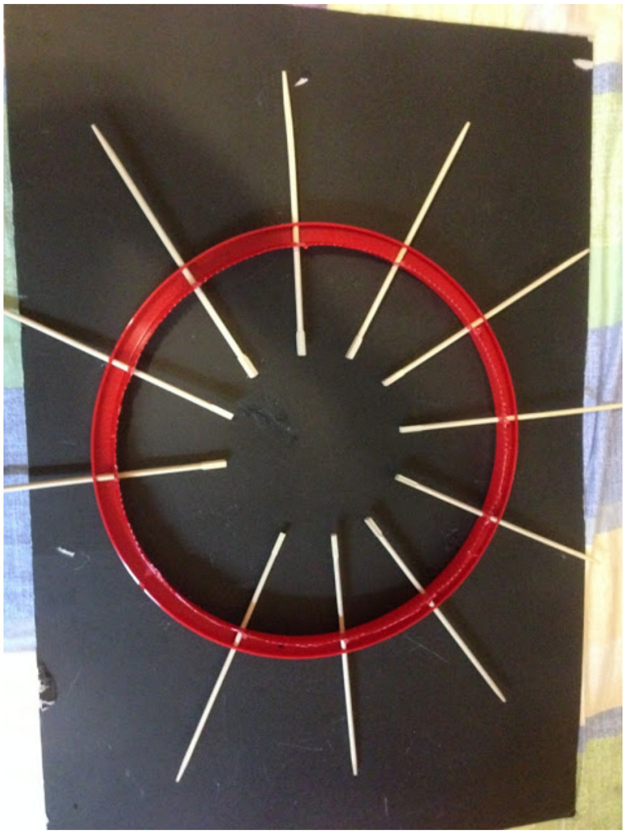

[http://4.bp.blogspot.com/-LC33DlKSk9M/VgZ1kgxxfUI/AAAAAAAACFY/3hyVmcvT-Ak/s1600/IMG_8721.JPG]

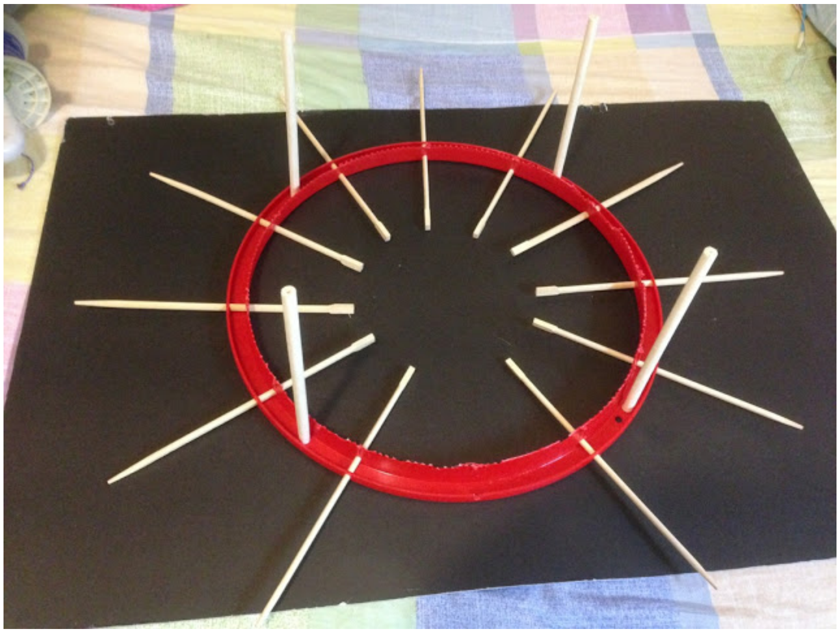

[http://3.bp.blogspot.com/-fEguBv0lFhU/VgZ1k59GNcI/AAAAAAAACFU/HP-Bf8WDpfY/s1600/IMG_8722.JPG]

[http://3.bp.blogspot.com/-fEguBv0lFhU/VgZ1k59GNcI/AAAAAAAACFU/HP-Bf8WDpfY/s1600/IMG_8722.JPG]

[http://3.bp.blogspot.com/-fEguBv0lFhU/VgZ1k59GNcI/AAAAAAAACFU/HP-Bf8WDpfY/s1600/IMG_8722.JPG]

Dynamic Views theme. Powered by Blogger.

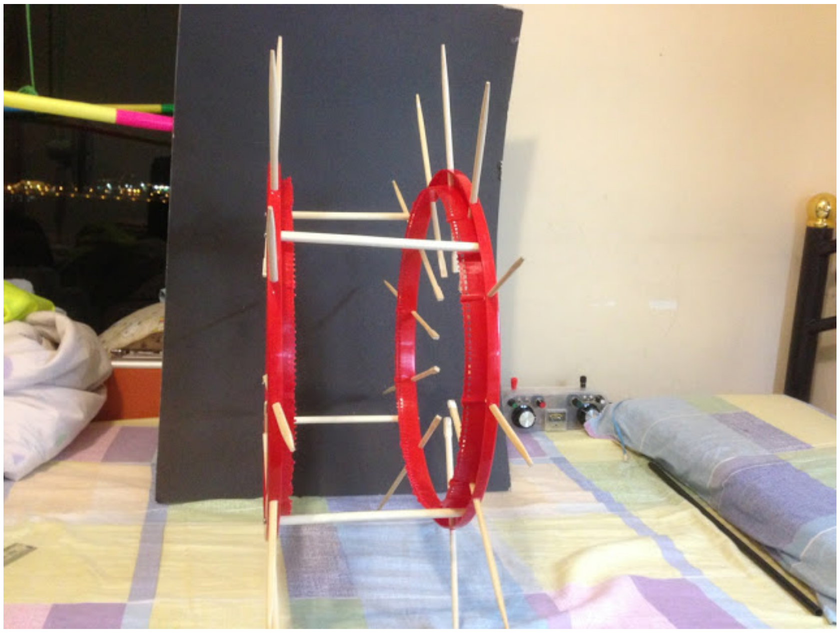

[http://4.bp.blogspot.com/-4b_F9qDf-qo/VgZ1lSQSVPI/AAAAAAAACFo/taROPzqACTs/s1600/IMG_8724.JPG]

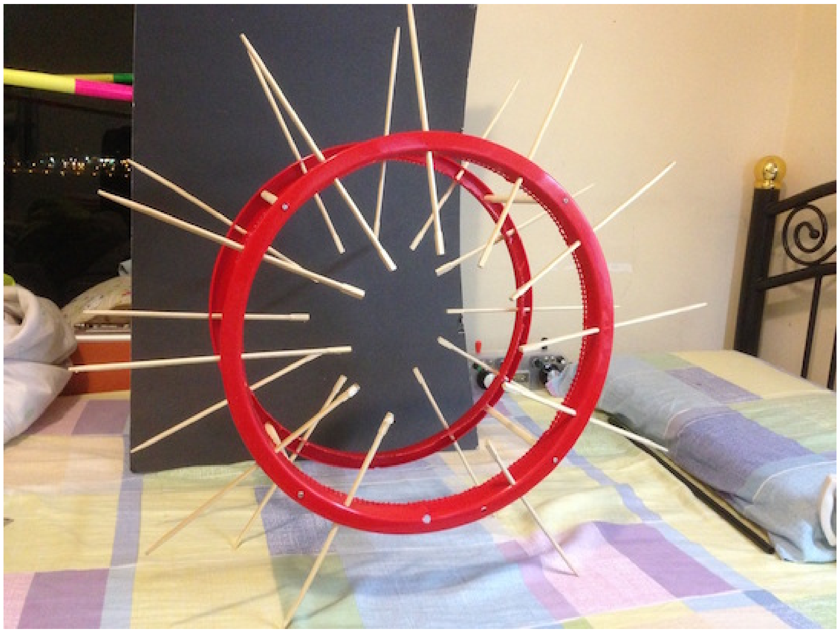

[http://3.bp.blogspot.com/-XS6N9EzX4Ow/VgZ1ZIBEUBI/AAAAAAAACD8/5NsuSPSmyno/s1600/IMG_8723.JPG]

Dynamic Views theme. Powered by Blogger.

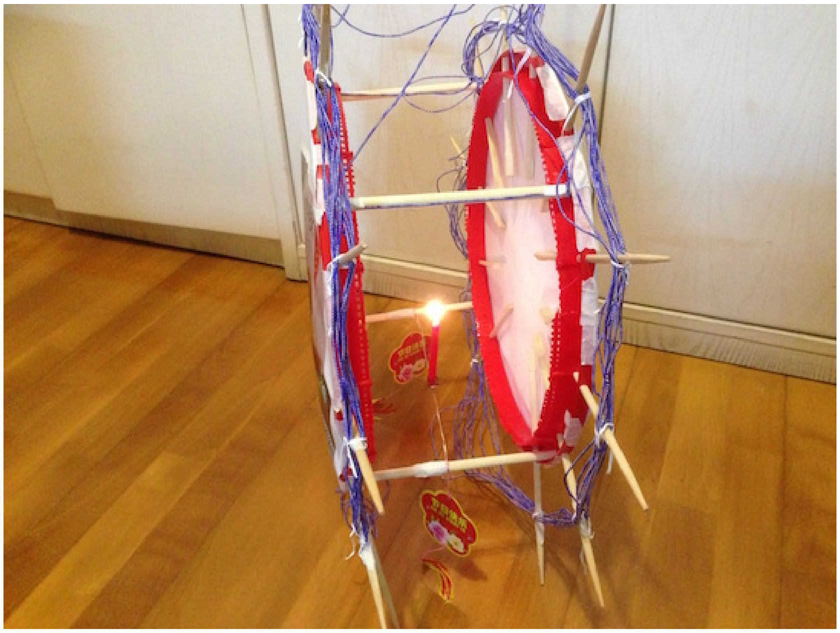

[http://2.bp.blogspot.com/-A5zT9XiW9K4/VgZ1ZaLDTOI/AAAAAAAACEk/Du2nyZddZgk/s1600/IMG_8732.JPG]

[http://2.bp.blogspot.com/-OZo3p9K2gxU/VgZ1ZtbjcLI/AAAAAAAACEM/bTxgAE0N0jc/s1600/IMG_8733.JPG]

Posted 26th September 2015 by Billy

0 Add a comment

3rd September 2015

Match box crystal radio

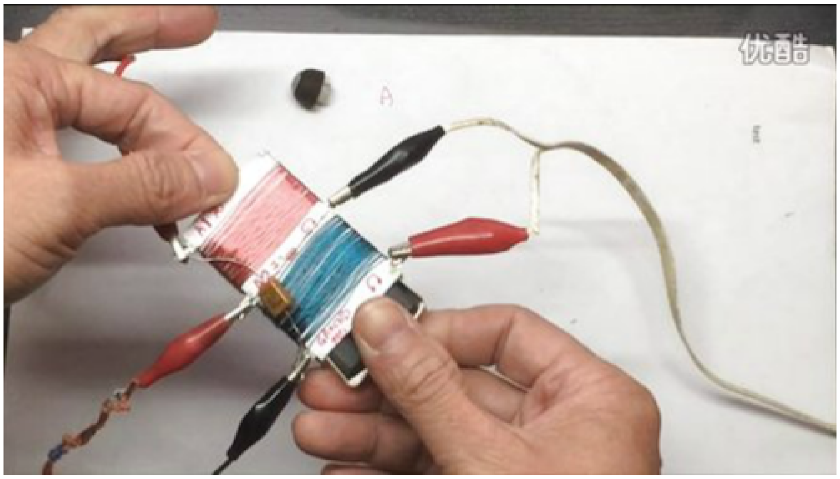

[http://3.bp.blogspot.com/-dmIYfrM-g6U/VehkjPTgnzI/AAAAAAAACDE/xTeA1_Ta93M/s1600/Screen%2BShot%2B2015-08-29%2Bat%2B8.06.00%2Bam.png]

Just made a match box radio using some short ferrite sticks.

Here is the testing video.

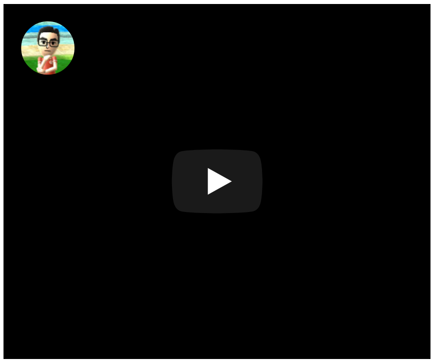

The radio is really easy to make. In the bear minimum, you just need the following:
1. A match box measured 60x35x10mm, or make one using cardboards yourself.
2.Magnet wires of AWG 27 or 0.3mm or similar thickness.
3. Any Germanium Diode (e.g. 1N34A, or 1N60, I used 1SS86).
4. Some solder lugs or you can use the paper clip as the connectors for antenna, earth, and earphones.
5. A pair of high impedance earphone or a crystal earbud.
6. A good outdoor antenna of 10 meters or above.
7. A good earth connection, e.g. connect to the water pipe or tap.
8. Fixed capacitor to match the coil for tuning: can be 100pf, 200pf or 300pf or omitted , value will be fixed during testing. Buy a few different fixed capacitor to test with.
9. Fixed capacitor of 2000pf, or can be omitted if your headphone is a crystal earbud.

How to make it:

1. Find a match box of 60x35x10mm,
2. Stick 3 solder lugs at the the left edge of the matchbox
one at the top (A1), one at the middle (A2), one at the bottom (G).
3. Stick 2 solder lugs at the right edge of the matchbox, one at the middle (O1),
one at the bottom (O2).
4. Put two or three small ferrite sticks (8 x 50 mm) inside the match box.
5. Wind the coils:
    L1: tighty wind 38 turns of 0.3 0r 0.4mm Magnet wire starting from the top of the matchbox
    L2: same as L1. wind on the same match box with a 5mm gap between L1 and L2
6. Solder L1, L2 and other components to the solder lugs as shown in the circuit diagram, except the variable capacitor C1. The value need to be tested before soldering that in.
7. Value of fixed capacitor C1 need to be tested to confirm, or can be completely omitted.
8. Fixed capacitor C2 can be omitted if you are using a crystal earbud.

Testing:
Connect the outdoor antenna to A1.
Connect the earth to G.

Connect the headphones to O1 and O2.

Slide the box holding the ferrite sticks inside the matchbox in and out of the matchbox to tune to a station. Try to replace the C1 or remove it to tune to the lowest AM/MW station possible (i..e as close to 500khz as possible).

Next, connect the outdoor antenna to A2.

Try to replace the C1 or remove it to tune to the upper most AM/MW station possible (i..e as close to 1500khz as possible).

Find a C1 value that will allow you to cover the best range, then solder it to the solder slug between A1 and G.

I used a pair of 150 ohm hearing aid earbud, after testing, I found the the performance of the radio is just as fine by omitting C1 and C2. Hence I removed C1 and C2 from my final circuit.

Reception for strong stations is good.

For waker stations that is too close to the strong stations, you have to add a wave trap connected in series between the outdoor antenna and the antenna input of the matchbox ratio to block the strong stations before you can get to the weaker ones.

The wave trap is just another tuning tank make from a coil and a variable capacitor.

Its function is to tune to the strong station that is blocking the weak station, so that signal is blocked from entering your crystal radio so the weaker stations can be heard.

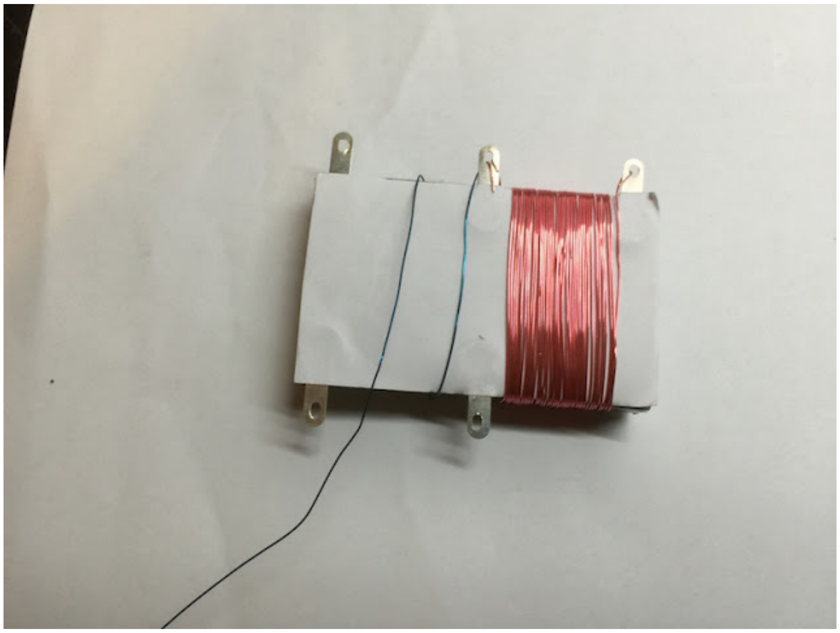

[http://1.bp.blogspot.com/-CigoTwU9qv0/VehkhpPa7lI/AAAAAAAACB4/K0Yw7W3qU44/s1600/IMG_2822.JPG]

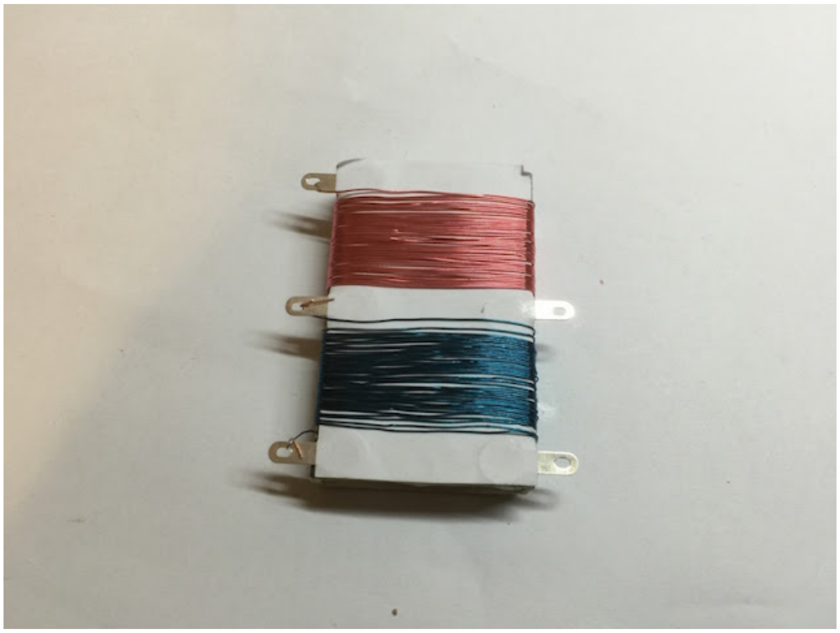

[http://1.bp.blogspot.com/-wO7vn5kyQV4/VehkhjslxuI/AAAAAAAACBw/O0L-uT_n9mo/s1600/IMG_2823.JPG]

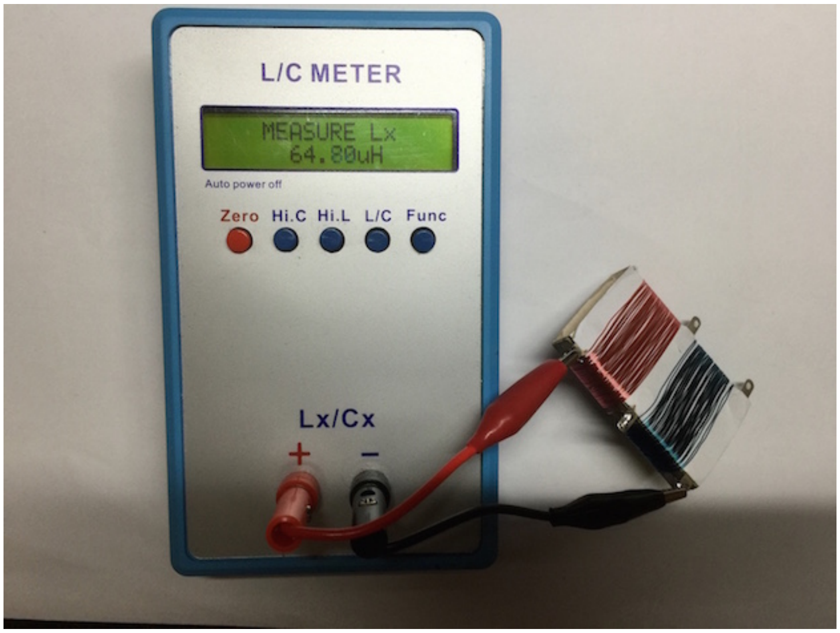
L/C METER
MEASURE Lx
64.80uH
Auto power off
Zero Hi.C Hi.L L/C Func
Lx/Cx
+
–

[http://1.bp.blogspot.com/-dI-rki58tAU/Vehkhki7CVI/AAAAAAAACB0/It7bNbhqO8I/s1600/IMG_2824.JPG]

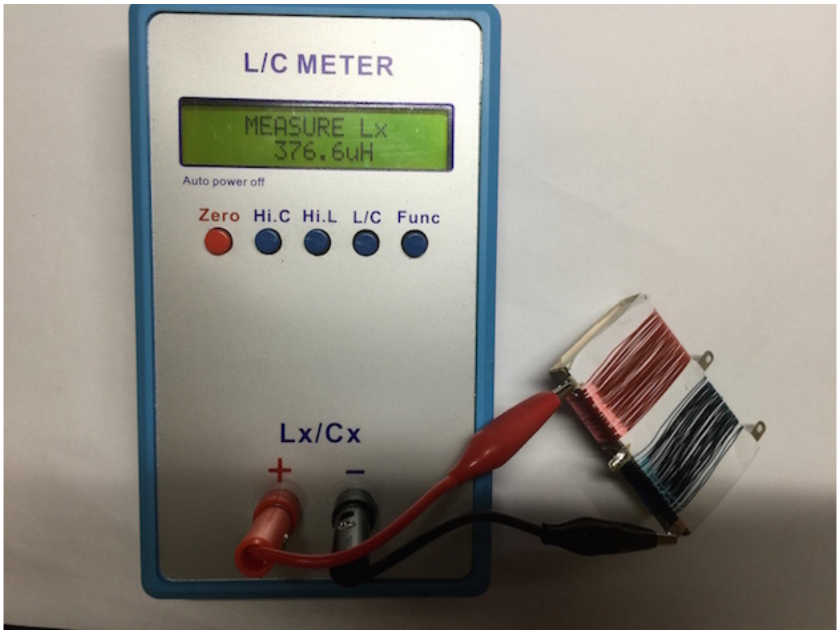
L/C METER
MEASURE Lx
376.6uH
Auto power off
Zero Hi.C Hi.L L/C Func
Lx/Cx
+
–

[http://4.bp.blogspot.com/-aHY5mphCy4A/VehkiSjEAlI/AAAAAAAACCA/fo0yYBiGu0c/s1600/IMG_2825.JPG]

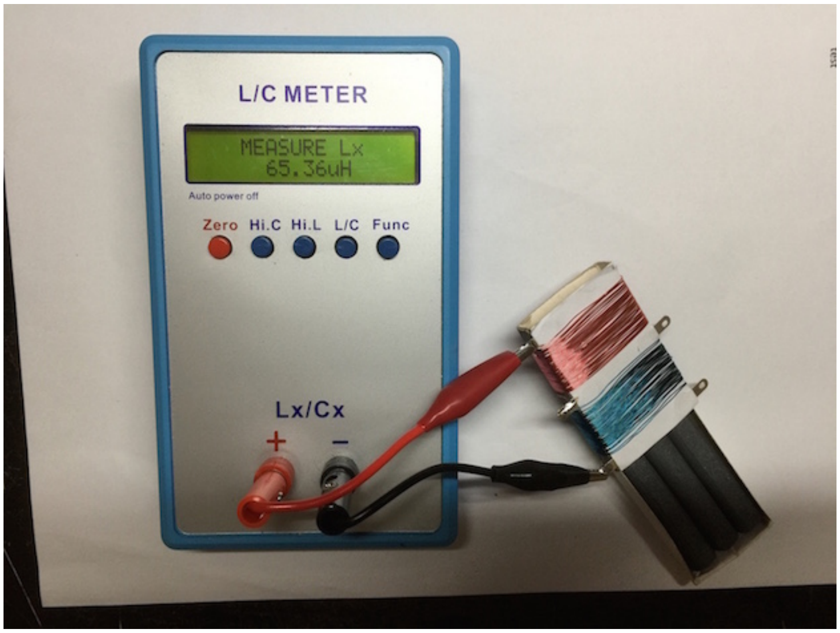
L/C METER
MEASURE Lx
65.36uH
Auto power off
Zero Hi.C Hi.L L/C Func
Lx/Cx
+ –

[http://1.bp.blogspot.com/--b2yHJUoMQ4/VehkidSxdtI/AAAAAAAACCE/3e9xsy_h1I4/s1600/IMG_2826.JPG]

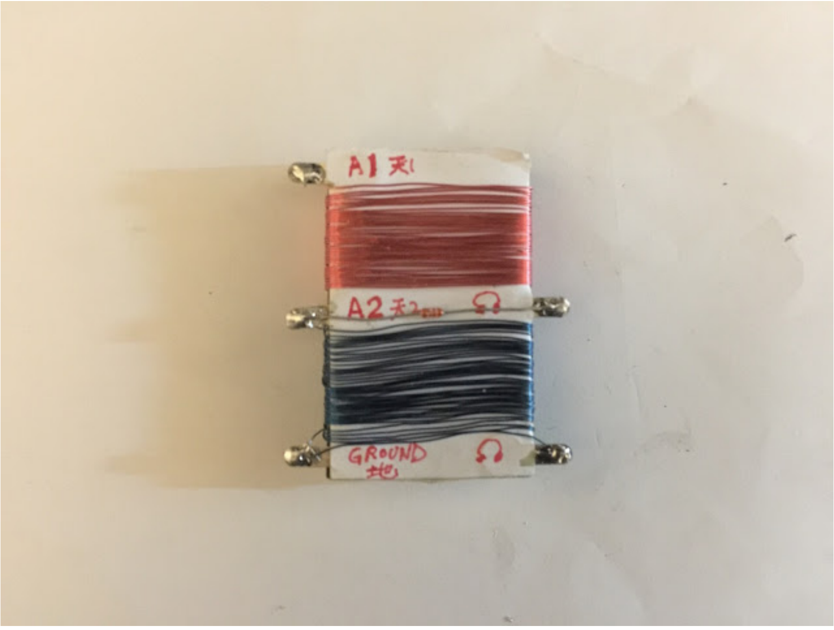
A1 天1
A2 天2
GROUND
地

[http://2.bp.blogspot.com/-qL4T33y-5iM/Vehkit_xemI/AAAAAAAACCI/JIguyu-RXTU/s1600/IMG_2831.JPG]

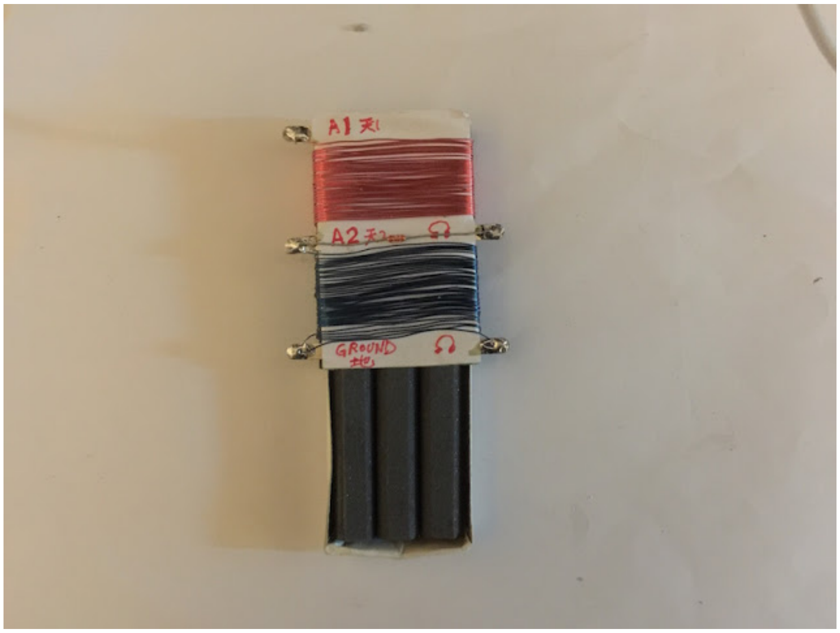
A1
A2
GROUND

[http://2.bp.blogspot.com/-dCAIN-h-P6A/VehkiwEJa9I/AAAAAAAACCQ/tjchHLRIl8c/s1600/IMG_2832.JPG]

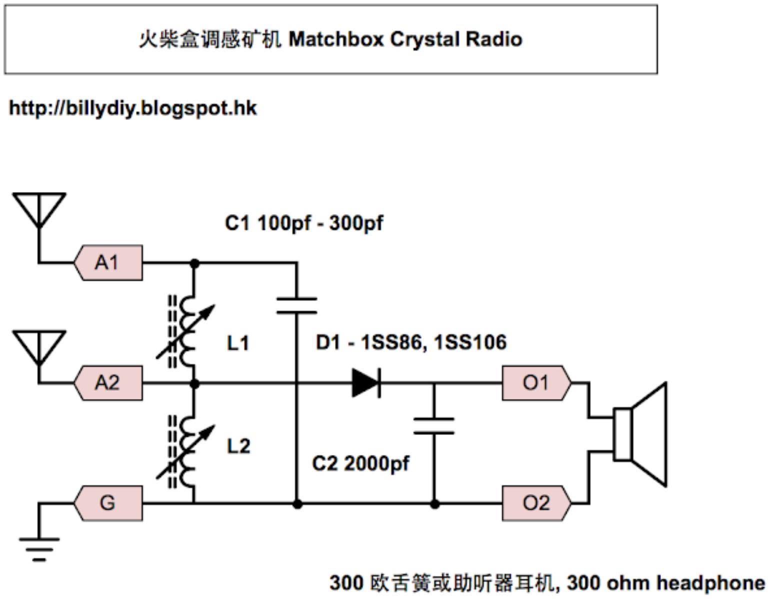
火柴盒调感矿机 Matchbox Crystal Radio
http://billydiy.blogspot.hk
C1 100pf - 300pf
A1
A2
G
L1
L2
D1 - 1SS86, 1SS106
C2 2000pf
O1
O2
300 欧舌簧或助听器耳机, 300 ohm headphone

[http://3.bp.blogspot.com/-s_kqtVbiU7E/Vehkj-gDzOI/AAAAAAAACCo/KKu2pVlkB9A/s1600/Screen%2BShot%2B2015-08-29%2Bat%2B9.20.49%2Bam.png]

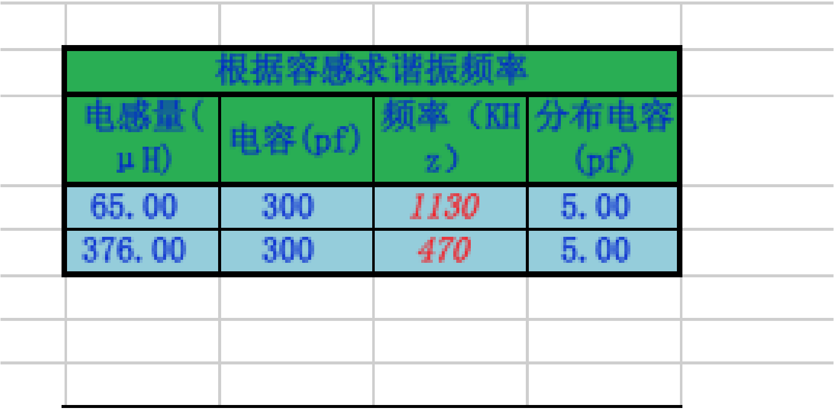

根据容感求谐振频率			
电感量(μH)	电容(pf)	频率（KHz）	分布电容(pf)
65.00	300	1130	5.00
376.00	300	470	5.00

[http://4.bp.blogspot.com/-0LPGyDDH3B4/Vehkkia8S-I/AAAAAAAACC0/dOzOcD-dx_c/s1600/Tuning%2Bcoverage.png]

Posted 3rd September 2015 by Billy

2 View comments

3rd September 2015

230mm (9") Ferrite Sleeve Loop coil

I came across a post for FSL (Ferrite Sleeve Loop) - FSL Antenna Design Optimization, All-out Experimentation to Determine Weak-Signal Performance Potential By Gary DeBock, Puyallup, WA, USA March 2012 http://www.am-dx.com/antennas/FSL%20Antenna%20Design%20Optimization.htm [http://www.am-dx.com/antennas/FSL%20Antenna%20Design%20Optimization.htm]

I was eager to find out if this is a way to create loops that can have the same sensitivity as a much better loop antenna without ferrite cores. If so, this will be a good way to shrink the loops to make smaller crystal radios with the same or even better sensitivity.

I started following the instruction and twist it a bit to work with what I have at home.

I purchased some cheap ferrite sticks with unknown permeability. Since I cannot find the 10x200mm ones, I purchased the 10x100mm ferrite sticks which is much cheaper (@ 51 RMB for a 100 pieces equivalent to US$8.03). I intend to connect two of these together to make a 10x200mm ferrite sticks. I end up purchasing 200 pieces.
This will allow me to make a FSL with a perimeter of 10x100mm = 1 meter. Diameter of 310mm.

I ended up making one with the diameter of 230mm or 9".

I started with a round plastic container for noodles then wrap multiple layers of foam onto it.
Finally use adhesive tapes to stick the 170 ferrite sticks into position as close enough to one another without any gaps.

Then wrap another layer of thin foam, before winding 20 turns of 0.04x660 litz wires to form the coil for the tuning tank, another 2 turns for MOS FET output.

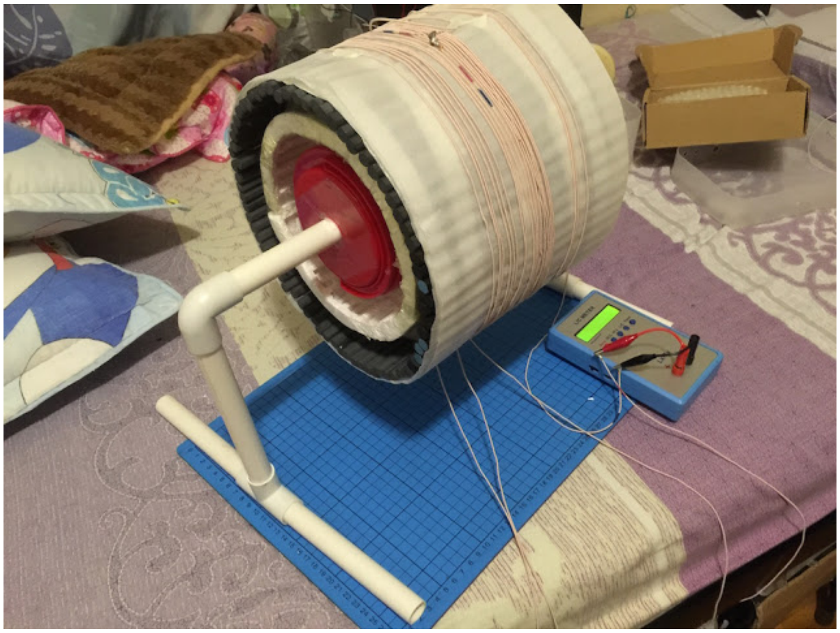

[http://3.bp.blogspot.com/-Wg9rrzwHwak/VehdTO_LNDI/AAAAAAAACBg/iuKEWx-YUzU/s1600/IMG_2677.JPG]

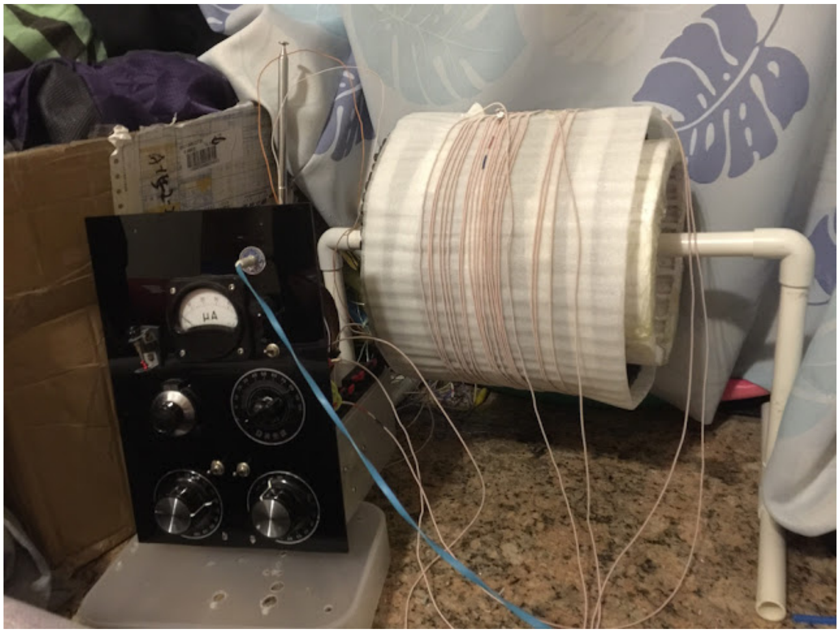

[http://4.bp.blogspot.com/-ksEZYGcgIl8/VehdTLHgM3I/AAAAAAAACBY/_qJy3EsMMIc/s1600/IMG_2680.JPG]

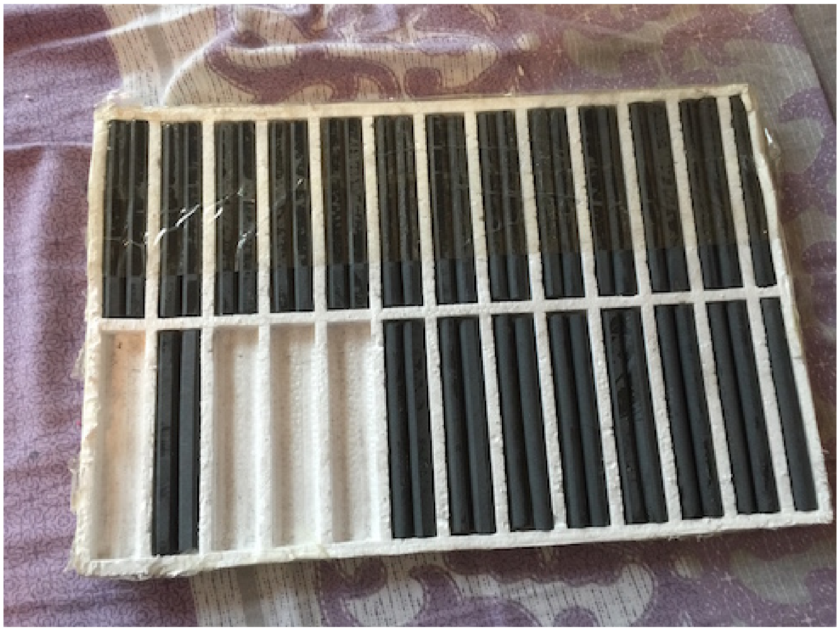

[http://1.bp.blogspot.com/-xbnSEULde2w/Vehal_8ynjI/AAAAAAAAB8w/F52_w0HRMyw/s1600/IMG_2665.JPG]

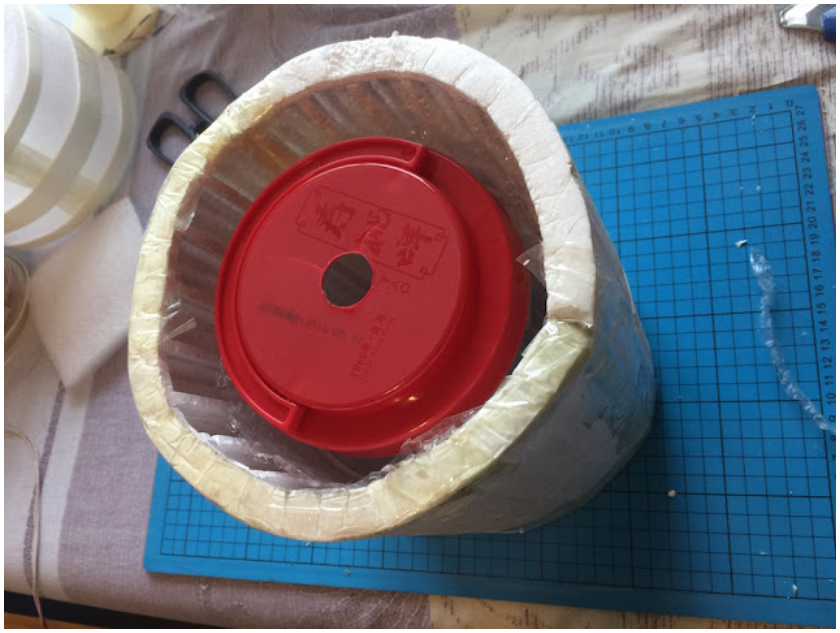

[http://2.bp.blogspot.com/-6qoWQMjNd5s/VehamOUGs3I/AAAAAAAAB9A/u5T8hqNLj48/s1600/IMG_2666.JPG]

Dynamic Views theme. Powered by Blogger.

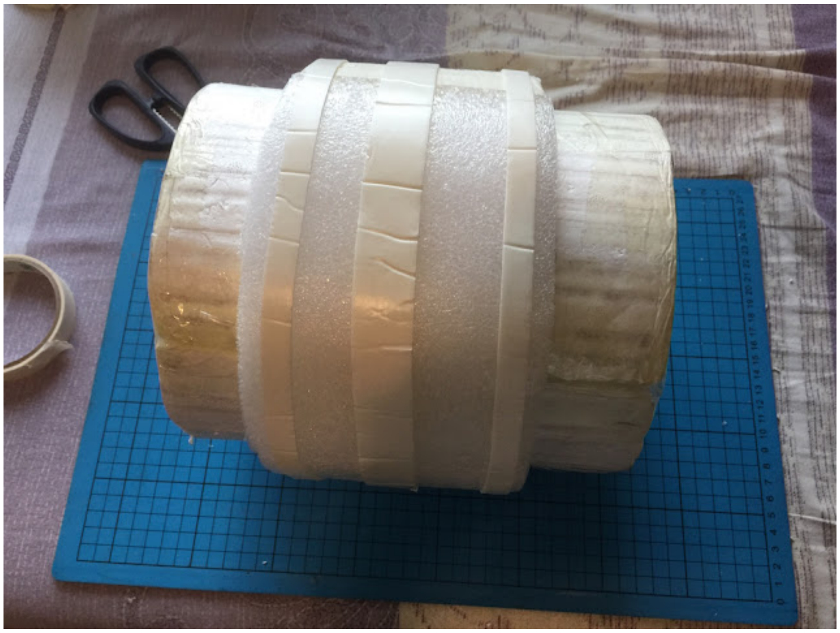

[http://4.bp.blogspot.com/-cQghIeTc8mM/VehamWMErwI/AAAAAAAAB9I/fUHkyjgsfJI/s1600/IMG_2667.JPG]

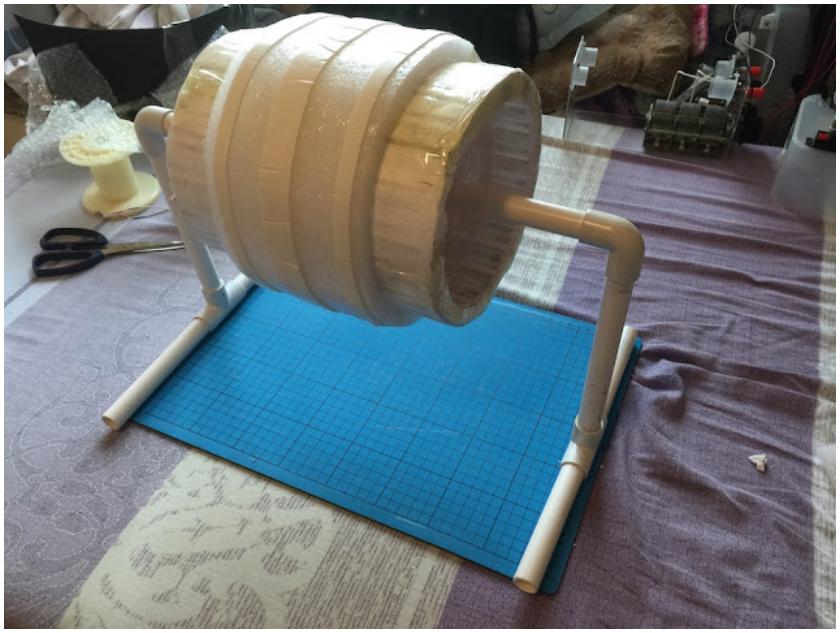

[http://3.bp.blogspot.com/-cglawW6PbpY/VehampMrDsI/AAAAAAAAB9Y/ZoxchyIHrmQ/s1600/IMG_2668.JPG]

Dynamic Views theme. Powered by Blogger.

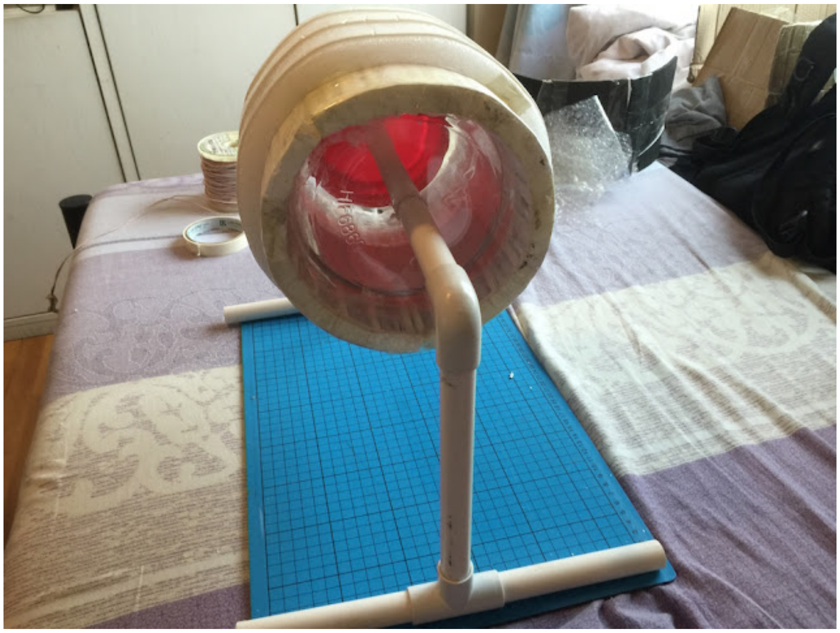

[http://4.bp.blogspot.com/--yy0LSck5PM/Vehamwg1NSI/AAAAAAAAB9c/o10F75hLbb4/s1600/IMG_2669.JPG]

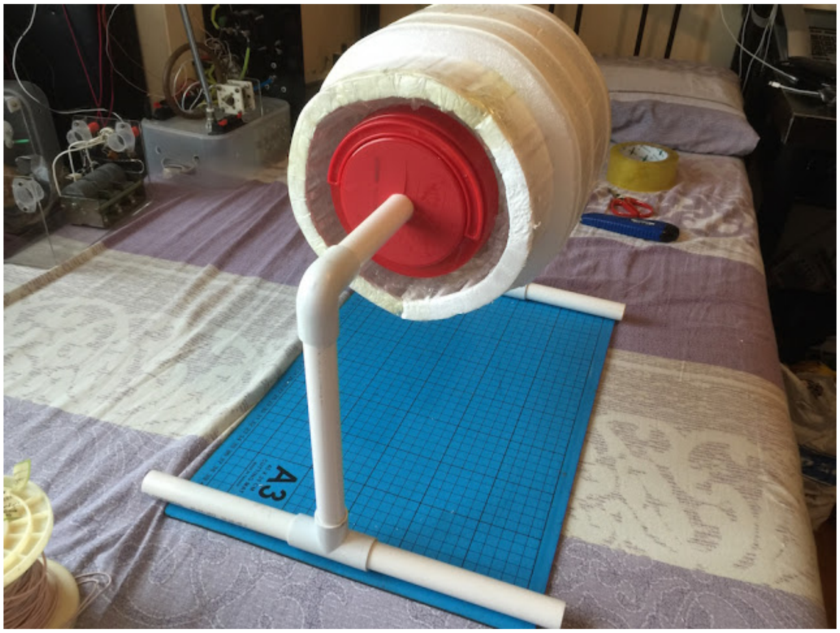

[http://1.bp.blogspot.com/-TVtIGqgar9Q/VehanWZAd2I/AAAAAAAAB_A/LmMAk3u1Ujo/s1600/IMG_2670.JPG]

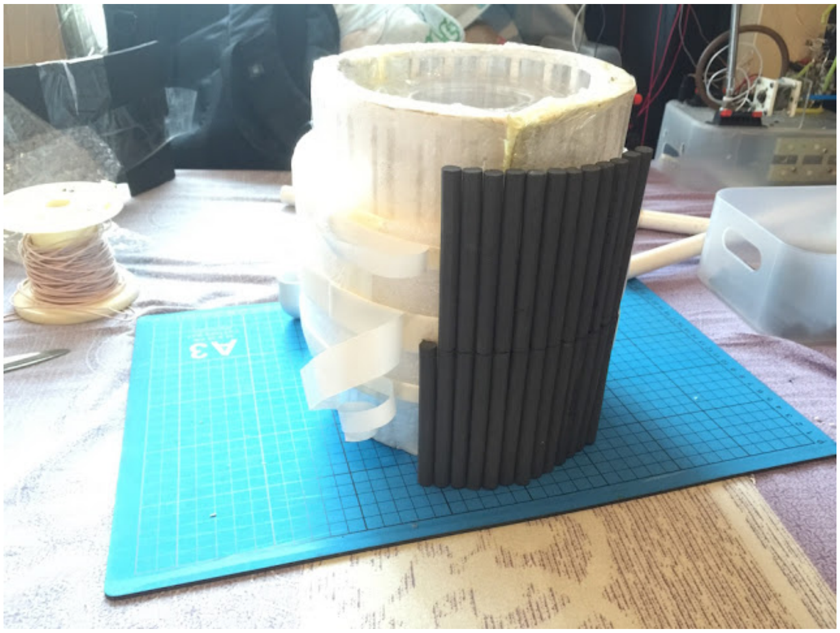

[http://1.bp.blogspot.com/-qgFvZCTvs9w/VehansW7LPI/AAAAAAAAB9s/g4BHm-WTwfw/s1600/IMG_2671.JPG]

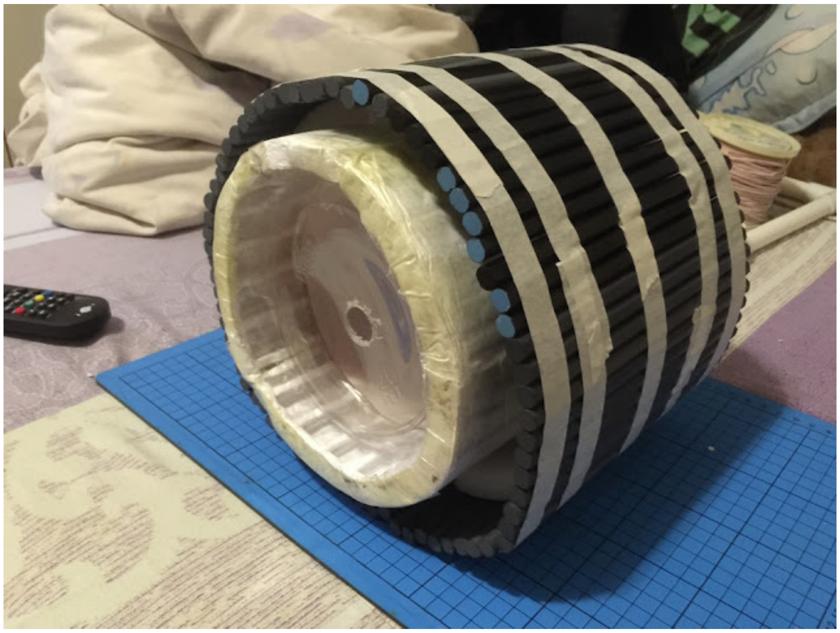

[http://4.bp.blogspot.com/-LTd7k9Azbr8/VehaoLzO-9I/AAAAAAAAB9w/ZIoHSLr7o1A/s1600/IMG_2672.JPG]

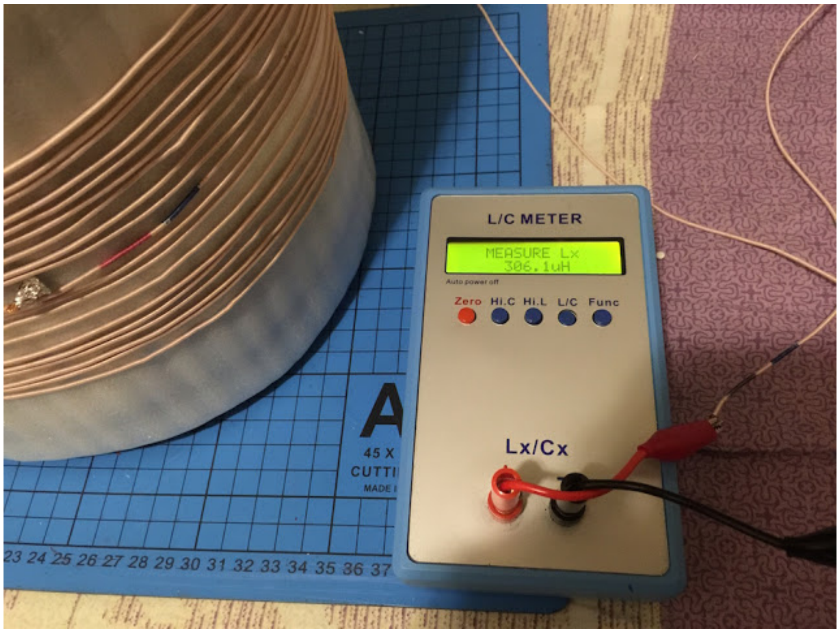
L/C METER
MEASURE Lx
306.1uH
Auto power off
Zero
Hi.C
Hi.L
L/C
Func
Lx/Cx

[http://2.bp.blogspot.com/-gCqvyFPan3I/VehaoKCvopI/AAAAAAAAB-E/Lp_47rmmsWU/s1600/IMG_2673.JPG]

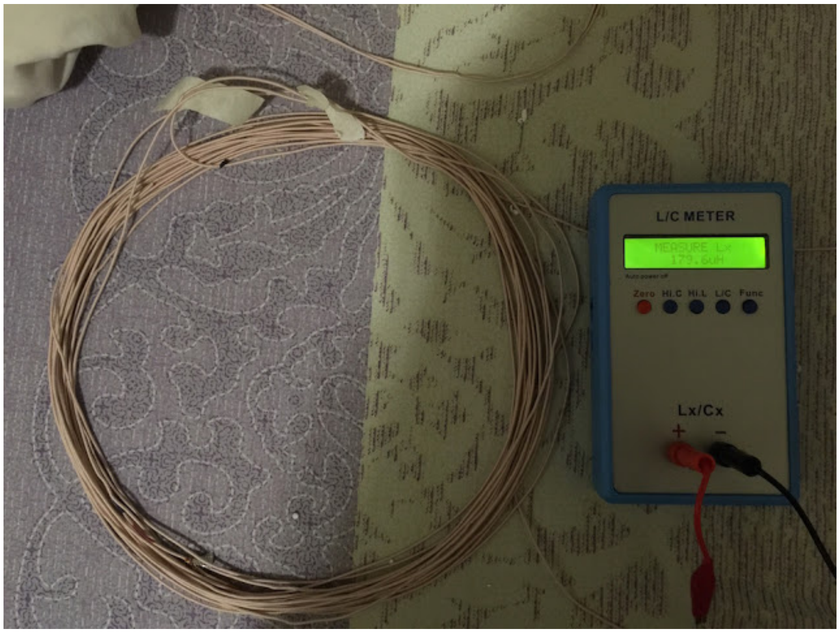
L/C METER
MEASURE Lx
179.6uH
Auto power off
Zero
Hi.C
Hi.L
L/C
Func
Lx/Cx
+
−

[http://1.bp.blogspot.com/-r7DmT2PGyl4/VehaobkFu9I/AAAAAAAAB94/laDJ5nggSqk/s1600/IMG_2674.JPG]

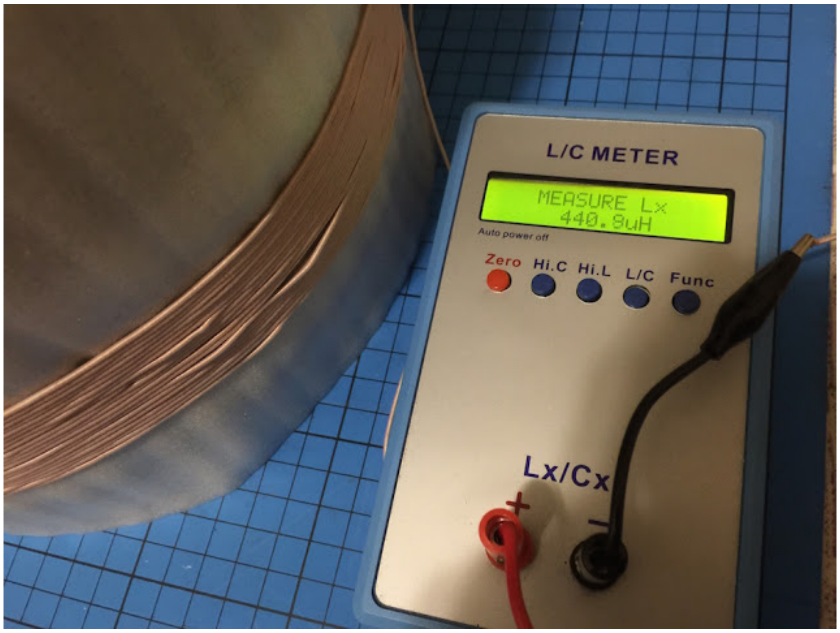
L/C METER
MEASURE Lx
440.8uH
Auto power off
Zero
Hi.C
Hi.L
L/C
Func
Lx/Cx
+
−

[http://4.bp.blogspot.com/-2xIksOqKXDI/Vehao0T-44I/AAAAAAAAB-Y/Sg4qlGwVqFA/s1600/IMG_2675.JPG]

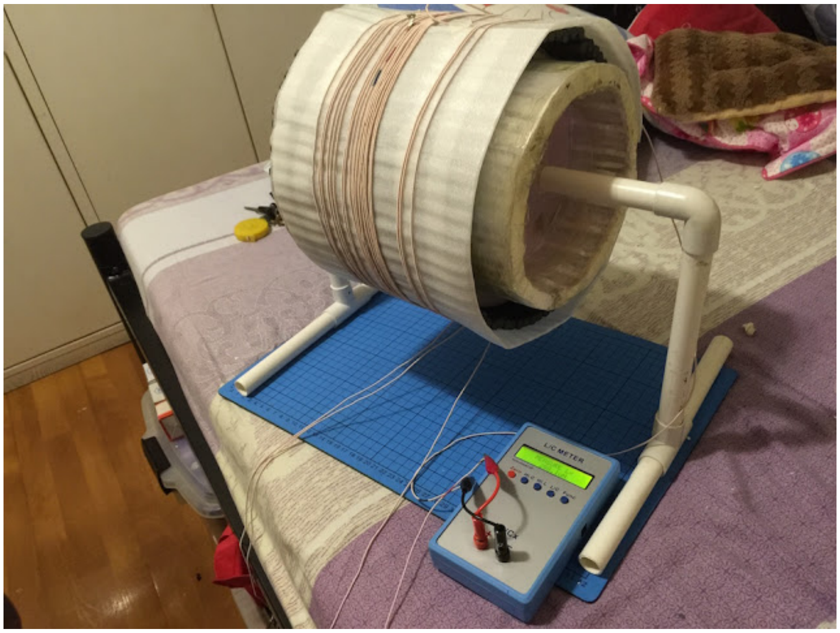

[http://2.bp.blogspot.com/-sI3ynvykgUw/VehapCPZxjI/AAAAAAAAB-M/AQyicrH2svM/s1600/IMG_2676.JPG]

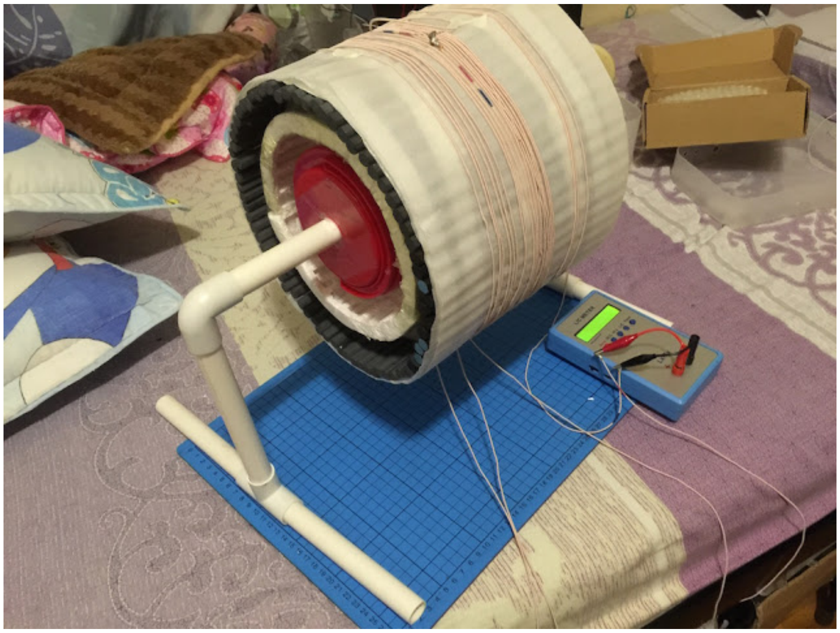

[http://3.bp.blogspot.com/-R8RKDxkhlTw/VehapV6n5XI/AAAAAAAAB-s/0-e-yal3aEo/s1600/IMG_2677.JPG]

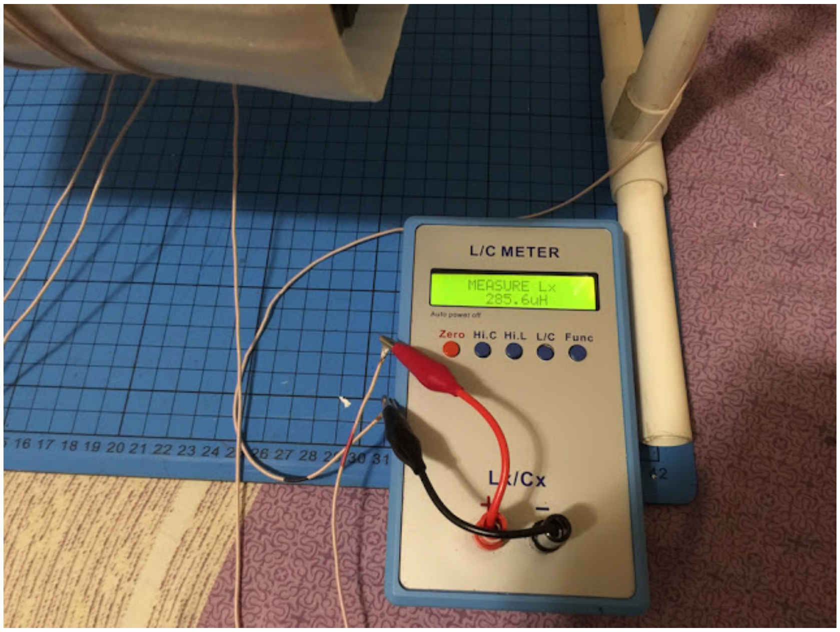
L/C METER
MEASURE Lx
285.6uH
Auto power off
Zero
Hi.C
Hi.L
L/C
Func
Lx/Cx
+
–

[http://4.bp.blogspot.com/-otc8xpvVZd0/VehaqORpB-I/AAAAAAAAB-g/anjQXeTdBUg/s1600/IMG_2679.JPG]

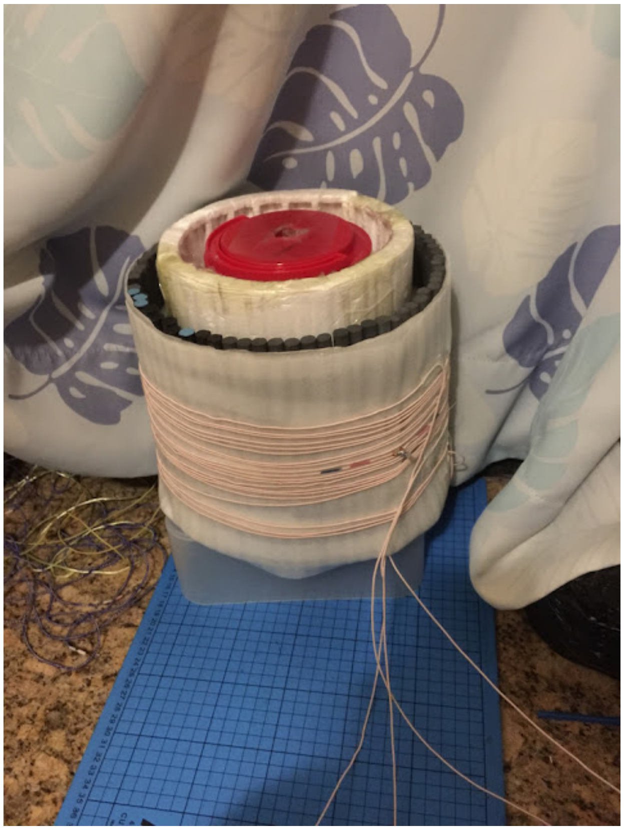

[http://3.bp.blogspot.com/-YsXOK2UhYDs/VehaqnFCraI/AAAAAAAAB-4/u8Bpn871KxA/s1600/IMG_2681.JPG]

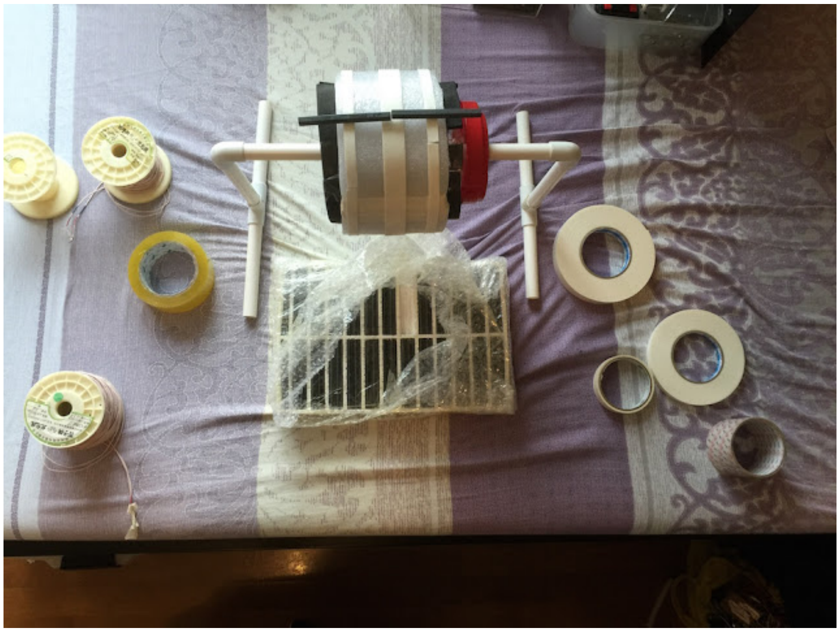

[http://2.bp.blogspot.com/-eyoihK78J8Y/VehakIVkTII/AAAAAAAAB8g/zJXBNXiEVa0/s1600/IMG_2658.JPG]

[http://4.bp.blogspot.com/-vkd_NPkcQrg/VehakEBI-7I/AAAAAAAAB8U/gKsc7F7uVPM/s1600/IMG_2659.JPG]

[http://2.bp.blogspot.com/-ueO_kVJ-6IM/VehakCZvXbI/AAAAAAAAB8Y/jhiYVbgM0bw/s1600/IMG_2660.JPG]

Dynamic Views theme. Powered by Blogger.

[http://3.bp.blogspot.com/-vpMdOx7uunU/VehaksAZ2yI/AAAAAAAAB8c/pTM87c4HeRU/s1600/IMG_2661.JPG]

[http://2.bp.blogspot.com/-EUgpN0i7x84/VehalMpoHrI/AAAAAAAAB80/mzBE5Zf8PEQ/s1600/IMG_2662.JPG]

Dynamic Views theme. Powered by Blogger.

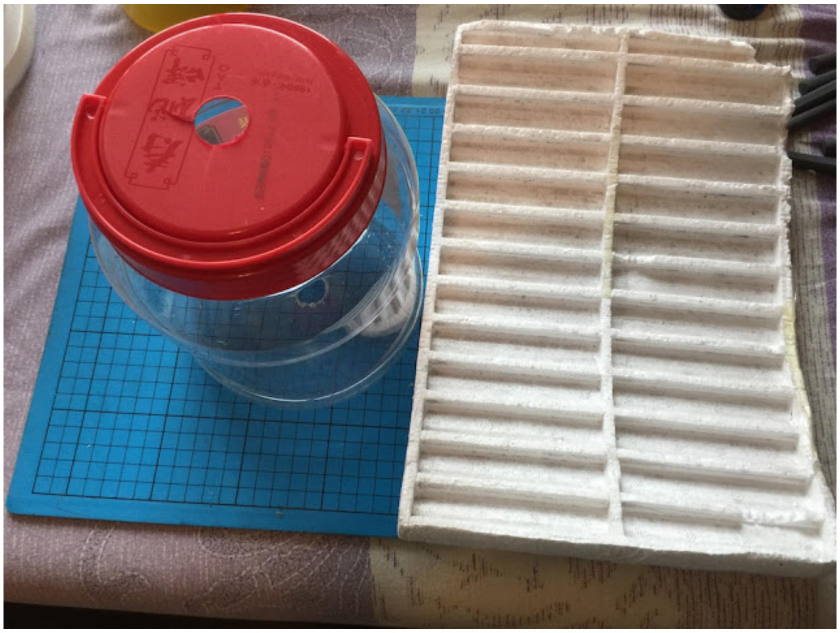

[http://1.bp.blogspot.com/-tK_z-j6VLwY/VehalbEy9-I/AAAAAAAAB8o/ILILtIGoXBI/s1600/IMG_2663.JPG]

[http://3.bp.blogspot.com/-Uq1YyrfzxWk/VehalvbPZWI/AAAAAAAAB8s/GAPQMnT44yg/s1600/IMG_2664.JPG]

I was amazed when I hook it up to the air variable capacity and the MOS FET.
Without the ferrite stick, an air loop that size will only produce very tiny output that is just audible.
With the ferrite stick sleeve, the FSLe produced doubled output volume.

Next is to compare this with the other bigger air loops to find out what sized air loops will have the same sensitivity as the FSL.

Testing is not easy as the sensitivity of the reception varies throughout the day.
My place is close to the airport, and every 10 minutes, a plane will literally fly over the top my 60 storey building (may be 1000 meter above), that did have some impact to the reception of crystal radio.
I am also trying to find a measurable unit that can give a true comparison.
For example, I tried measuring the voltage across the terminals of the audio output without the headphone.
I also tried measuring the output current level.
But I found that peak dc voltage, or peak dc current does not mean the loudest audio output.
I do not have a high frequency AC Voltmeter or db meter, so cannot get much out of this.

Next, I tried to use some iPhone software that can measure the audio level in decibels.
This seems to work quite well. I have created a device that can connect the earbud of my 2x150 ohm hearing aid earphone into the mic of the iPhone without much gaps

Then the software will pickup the audio signal form the mic and workout the peak output level.
By comparing the peak output level (db), I will be able to compare the sensitivity of different crystal radios objectively.

下载附件 (120.15 KB)
2014-7-14 01:46 上传

[http://1.bp.blogspot.com/-Pv5PWDfl_Gc/VehcEM9tE6I/AAAAAAAACAs/4YW6zWsA73Y/s1600/Screen%2BShot%2B2015-09-03%2Bat%2B10.39.07%2Bpm.png]

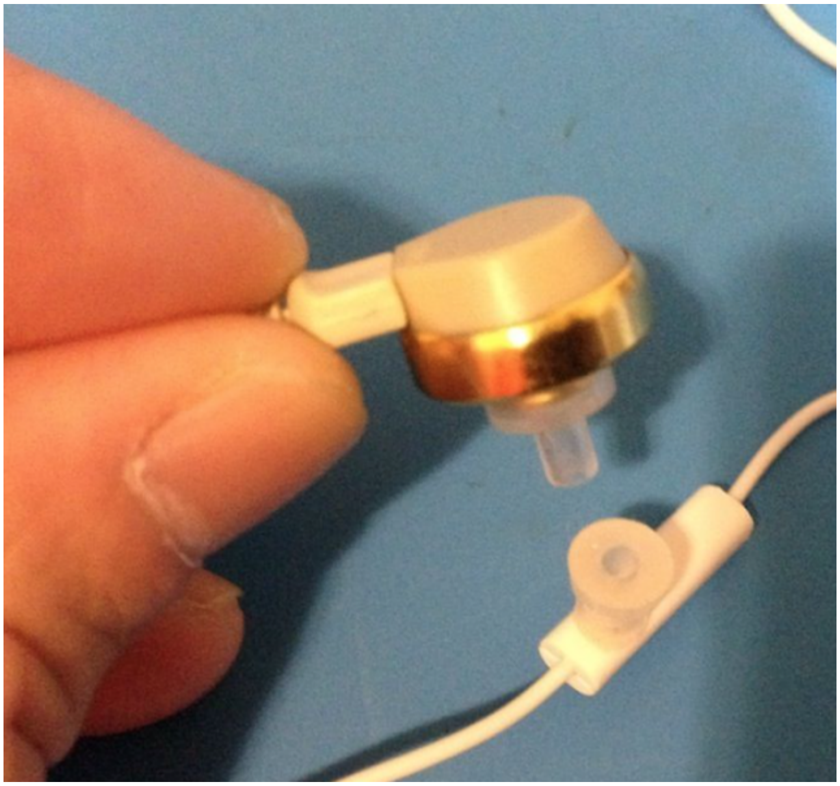

[http://4.bp.blogspot.com/-RUk-KUoSsz8/VehcFYA2ToI/AAAAAAAACBA/QzQS_rWxYqs/s1600/Screen%2BShot%2B2015-09-03%2Bat%2B10.39.54%2Bpm.png]

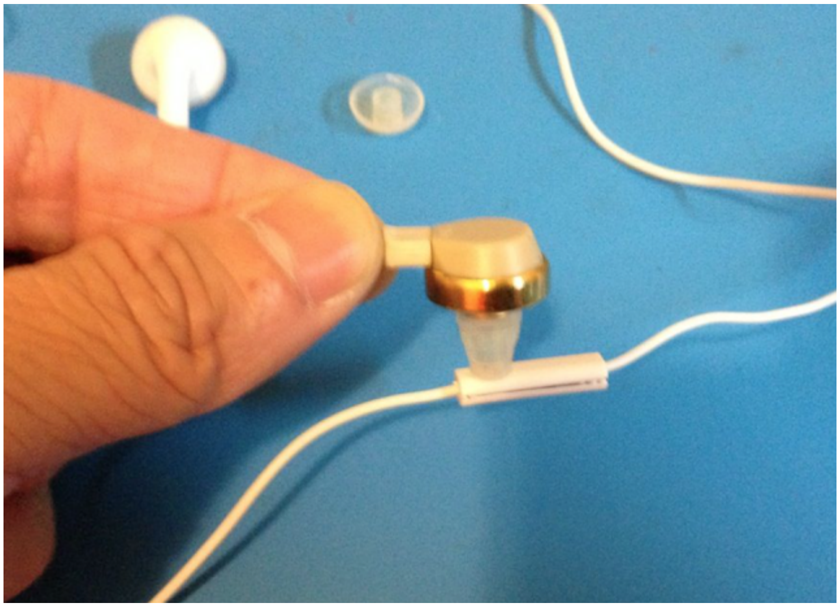

[http://1.bp.blogspot.com/-BCInM0-J0q4/VehcFK5VkqI/AAAAAAAACBE/HN2piv4CvNU/s1600/Screen%2BShot%2B2015-09-03%2Bat%2B10.39.41%2Bpm.png]

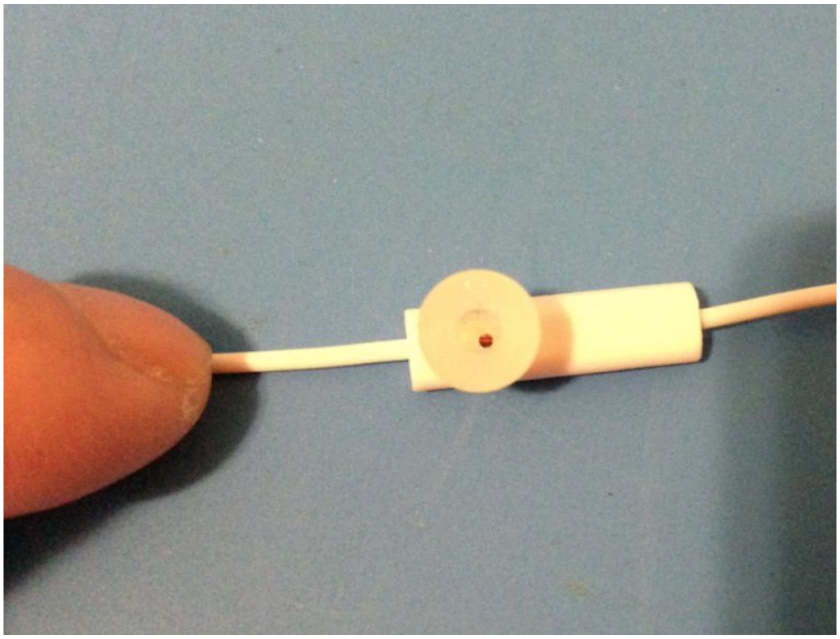

[http://1.bp.blogspot.com/-cLP1jxla9fQ/VehcEREd4hI/AAAAAAAACA8/jWKagE_dP18/s1600/Screen%2BShot%2B2015-09-03%2Bat%2B10.39.33%2Bpm.png]

Dynamic Views theme. Powered by Blogger.

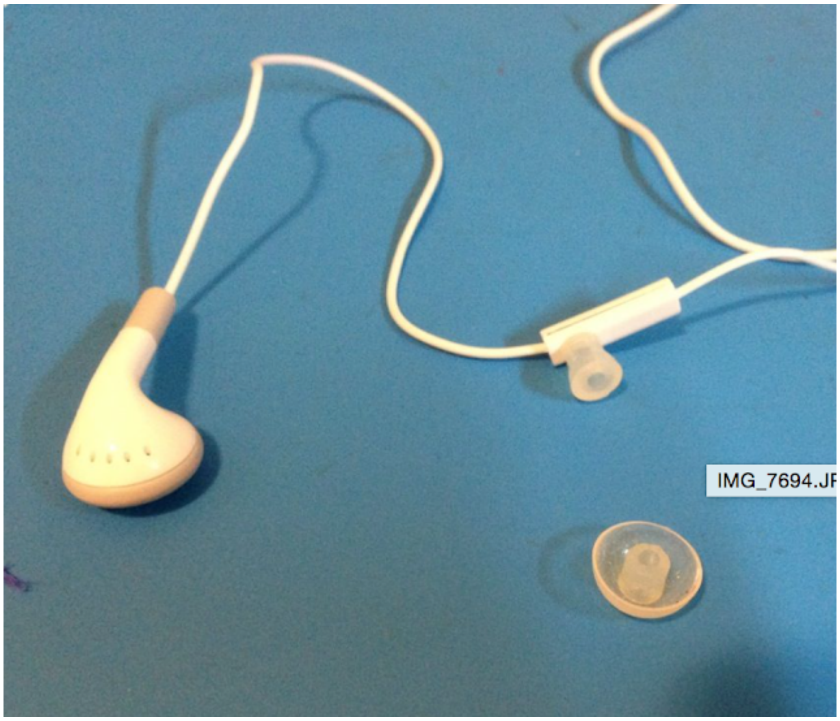

[http://2.bp.blogspot.com/-lIOzpEODl_c/VehcEcW-KZI/AAAAAAAACA4/Y8W-5Xqvv_w/s1600/Screen%2BShot%2B2015-09-03%2Bat%2B10.39.14%2Bpm.png]

There are multiple free db meter software on App Store. I found the measured value of the same audio level from different software varies a bit. In the worst case the variance can be +/50 % when you try to measure the same audio output with different software.

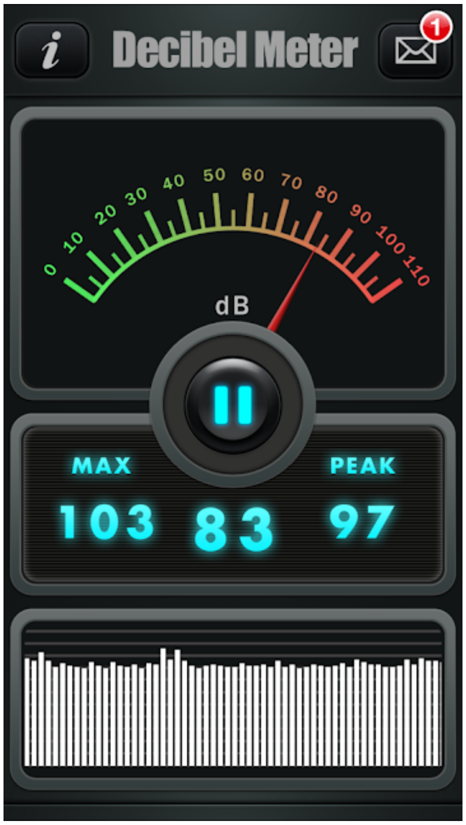
Decibel Meter
1
0 10 20 30 40 50 60 70 80 90 100 110
dB
MAX
PEAK
103
83
97

[http://3.bp.blogspot.com/-xnxvts0cX7s/Vehaq7ZMFIl/AAAAAAAACAQ/zbPkMRjakwE/s1600/IMG_2688.PNG]

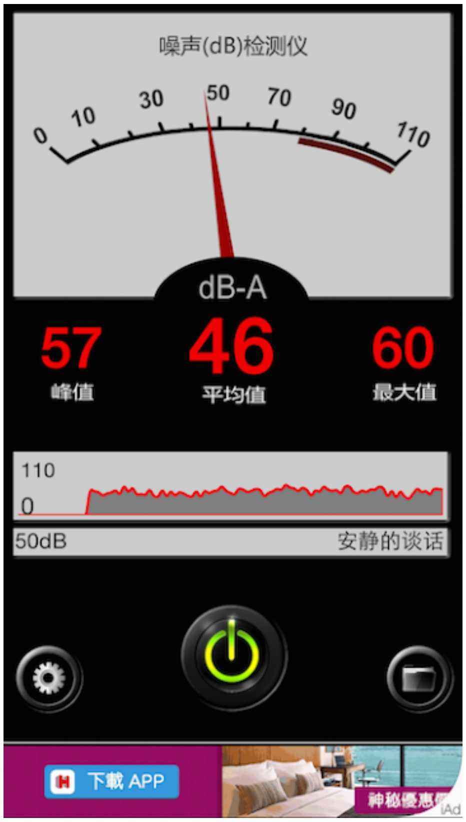
噪声(dB)检测仪
0
10
30
50
70
90
110
dB-A
57
峰值
46
平均值
60
最大值
110
0
50dB
安静的谈话
下載 APP
神秘優惠
iAd

[http://3.bp.blogspot.com/-MO-NOsb4fXE/VehaswvGL9I/AAAAAAAAB_0/mmebjP2UtEs/s1600/IMG_2710.PNG]

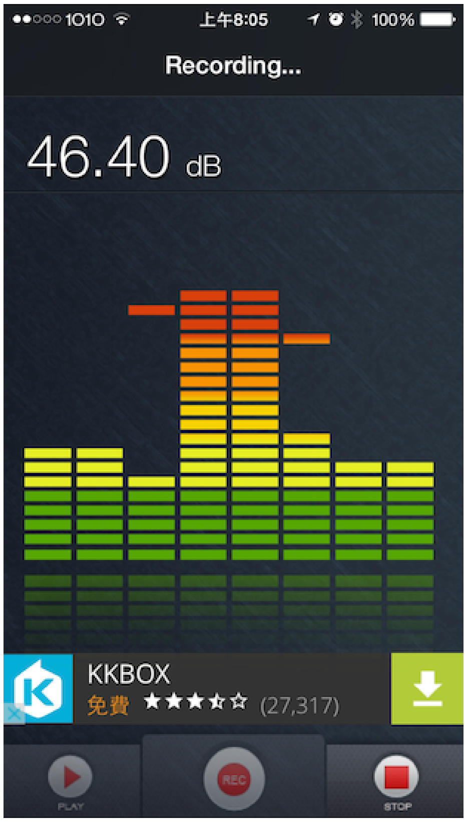
●●○○○ 1010
上午8:05
100%
Recording...
46.40 dB
KKBOX
免費 ★★★★☆ (27,317)
PLAY
REC
STOP

[http://2.bp.blogspot.com/-QZHmGapoHt0/VehasjW0oaI/AAAAAAAAB_4/tjFtp9bg9MI/s1600/IMG_2696.PNG]

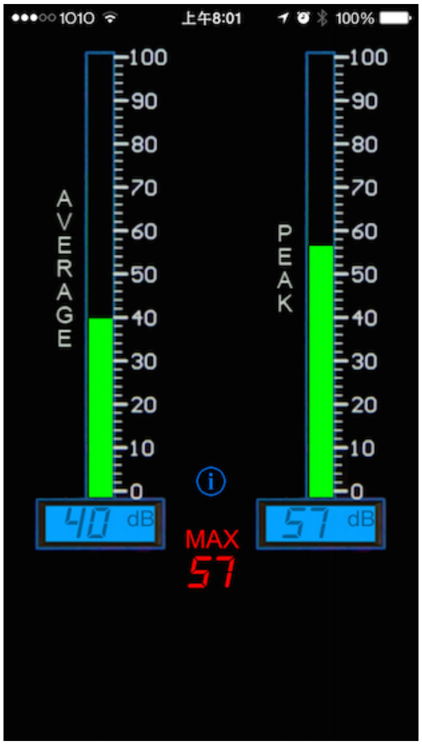
●●●○○ 1010
上午8:01
100%
AVERAGE
PEAK
100
90
80
70
60
50
40
30
20
10
0
40 dB
57 dB
MAX
57

[http://2.bp.blogspot.com/-i9uNUoLY2n0/VeharaprcPI/AAAAAAAAB_Q/9P08US5qDYk/s1600/IMG_2690.PNG]

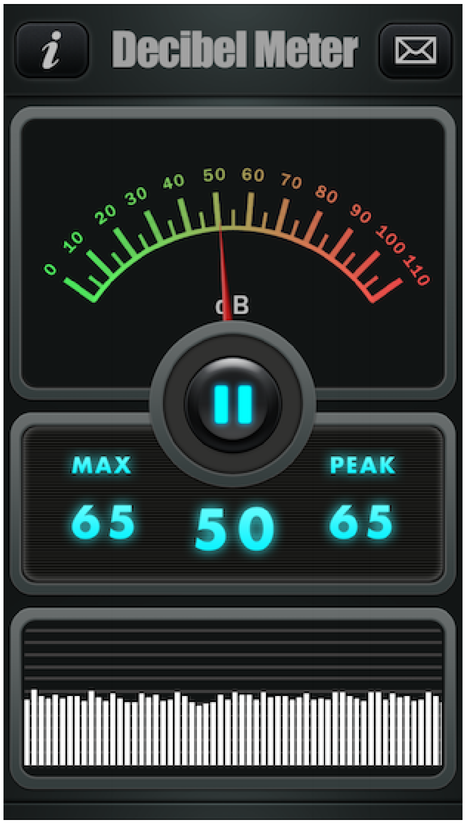
Decibel Meter
0 10 20 30 40 50 60 70 80 90 100 110
dB
MAX
PEAK
65
50
65

[http://3.bp.blogspot.com/--SEJVRNN714/VehasETdzcI/AAAAAAAACAA/ozavt2dZv2U/s1600/IMG_2692.PNG]

I also made an extensible frame to allow me to adjust the size of the air loop that can be used to compare with the FSL. I hope to find out what size Air loop will have the same sensitivity as the FSL.

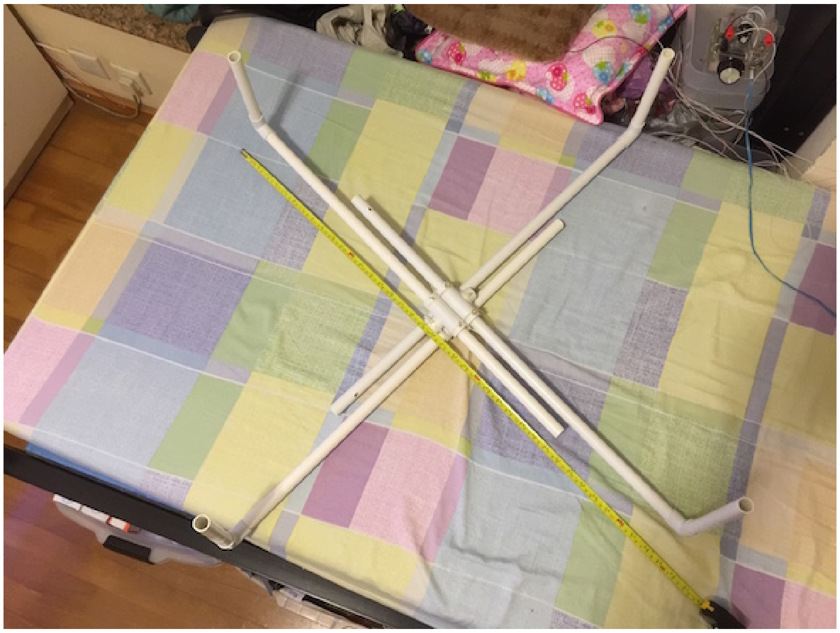

[http://2.bp.blogspot.com/-vVY4tsAv5Ig/VehatU-zJiI/AAAAAAAAB_8/ytMtzBak19k/s1600/IMG_2773.JPG]

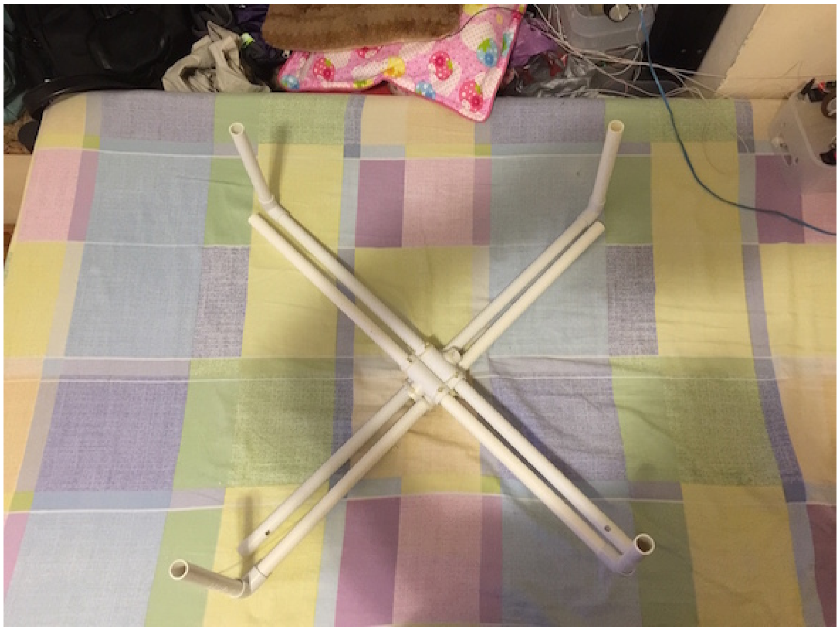

[http://3.bp.blogspot.com/-fj6N9dLWMgE/Vehat349lJI/AAAAAAAACAU/6m2ptu24dxE/s1600/IMG_2776.JPG]

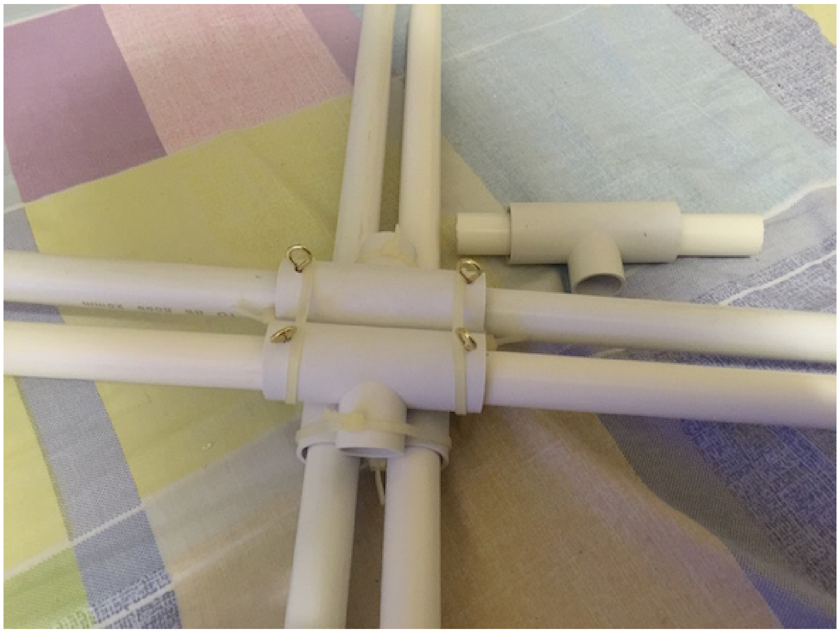

[http://3.bp.blogspot.com/-M8a1IMKF2Us/VehatrX6xaI/AAAAAAAACAE/vPoxRHU3Ulk/s1600/IMG_2774.JPG]

Once I prove that FSL is really worth the effort and money, I may upgrade my 9" 230mm diameter FSL to 12" diameter (330mm wide).

Posted 3rd September 2015 by Billy

1 View comments

SQUARESPACE WEBSITES & ONLINE STORES FOR ENTREPRENEURS START YOUR FREE TRIAL

1st September 2015 I won an award at the 6th Crystal Radio DIY contest

I am very happy to win the award at the 6th Crystal Radio DIY contest in China.

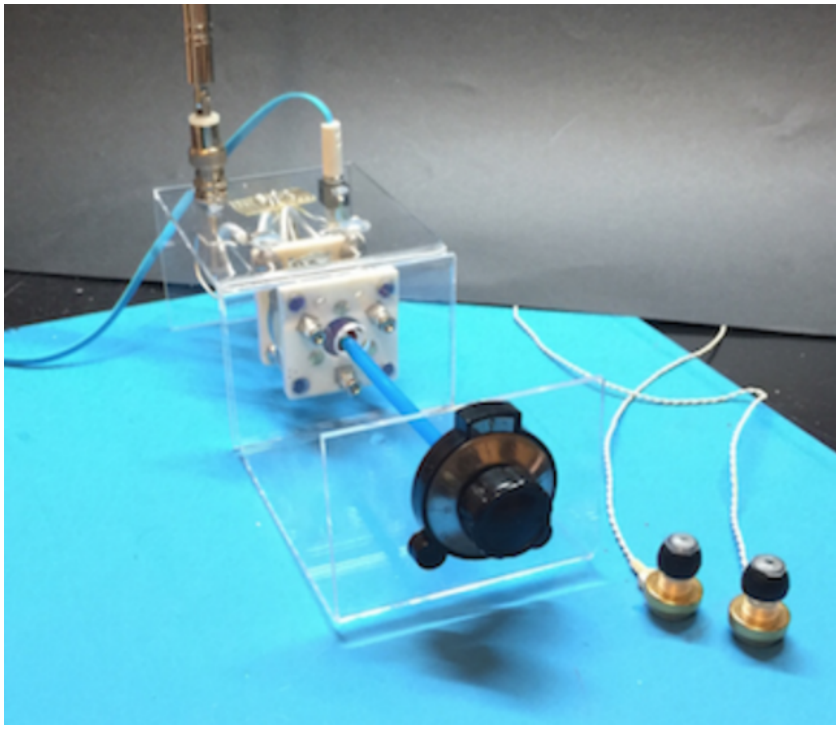

[http://4.bp.blogspot.com/-QPiFl3iCBvM/VeXAFs5zXZI/AAAAAAAAB70/OqjQd4ouGGo/s1600/Screen%2BShot%2B2015-09-01%2Bat%2B11.10.12%2Bpm.png]

Video Link

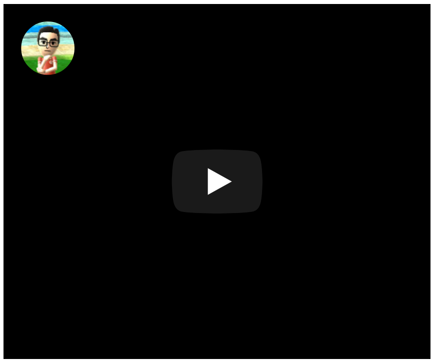

Here is the prize and award certificates.

[http://4.bp.blogspot.com/-z0jyyyCUqJo/VeW_k473DJI/AAAAAAAAB7M/xfZgEgGRuq4/s1600/IMG_2849.JPG]

矿石收音机
WWW.CRYSTALRADIO.CN
矿石收音机论坛成立十周年暨第五届DIY大赛纪念
2014年2月14日

[http://4.bp.blogspot.com/-bqBxMBIV_vA/VeW_lzyBPDI/AAAAAAAAB7g/JBG0yzK6Hb4/s1600/IMG_2855.JPG]

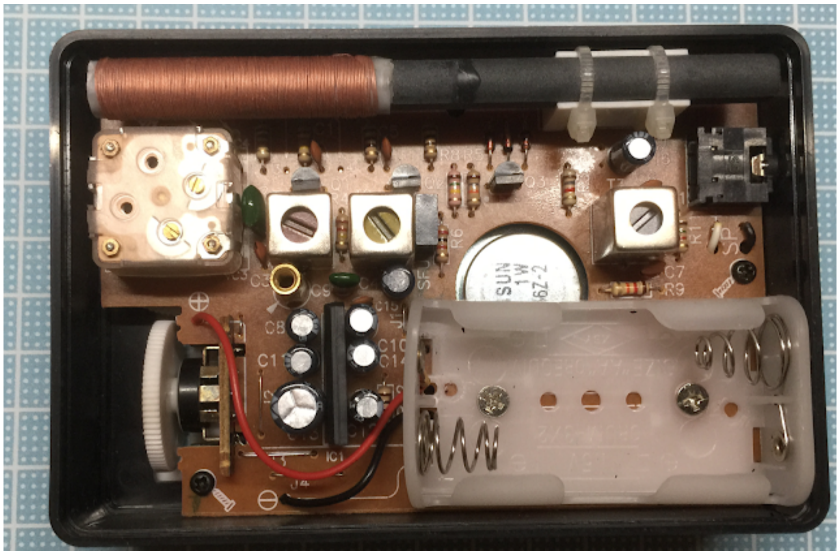

[http://2.bp.blogspot.com/-rFz2c_rPU7M/VeW_mLM0RZI/AAAAAAAAB7s/veqK-ZGL3dU/s1600/Screen%2BShot%2B2015-09-01%2Bat%2B11.00.55%2Bpm.png]

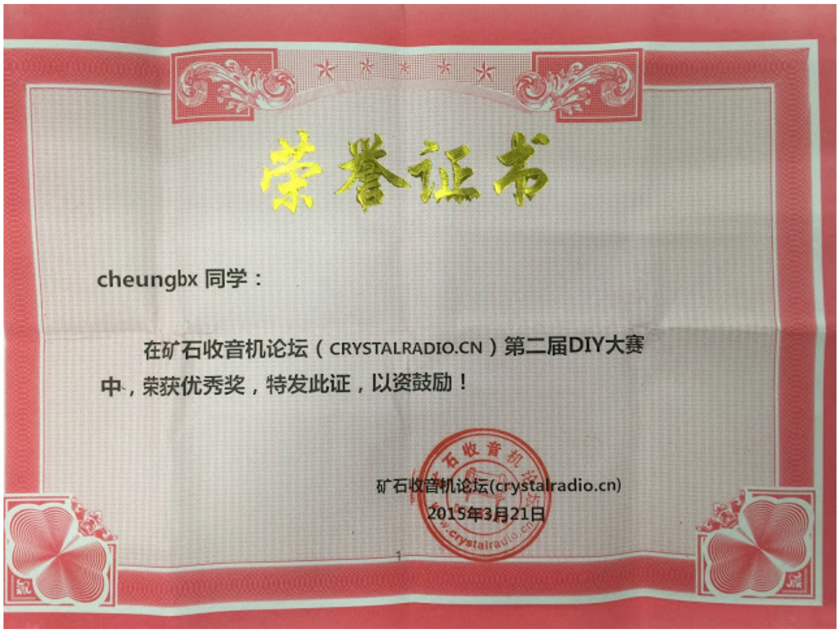

荣誉证书

cheungbx 同学：

在矿石收音机论坛（CRYSTALRADIO.CN）第二届DIY大赛中，荣获优秀奖，特发此证，以资鼓励！

矿石收音机论坛(crystalradio.cn)
2015年3月21日

[http://2.bp.blogspot.com/-Tw4KJAUpCoE/VeW_lD3oqkI/AAAAAAAAB7Q/QufJz_KnkA0/s1600/IMG_2848.JPG]

My post can be found here:

The Chinese version here.
http://www.crystalradio.cn/thread-625705-1-1.html

An english version here
http://theradioboard.com/rb/viewtopic.php?f=2&t=6088

Dynamic Views theme. Powered by Blogger.

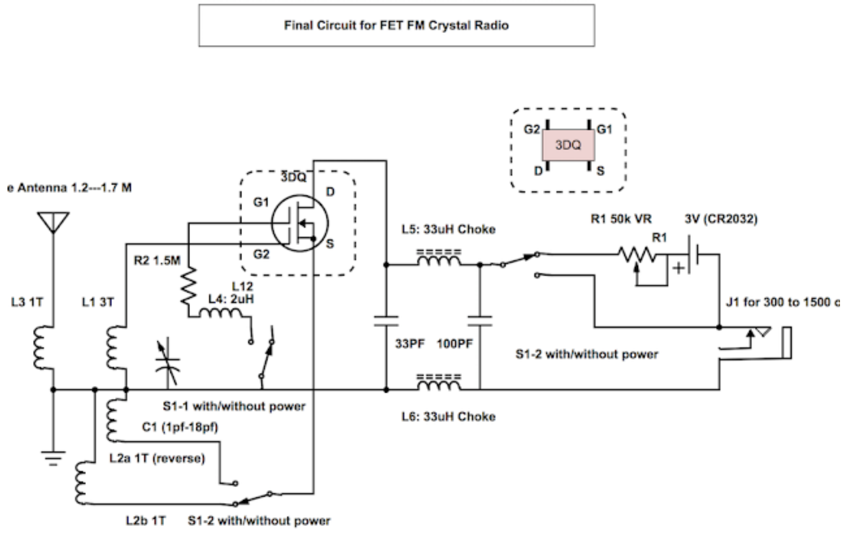
Final Circuit for FET FM Crystal Radio
G2
G1
3DQ
D
S
e Antenna 1.2---1.7 M
3DQ
G1
D
S
G2
L5: 33uH Choke
R1 50k VR
3V (CR2032)
R1
R2 1.5M
L12
L4: 2uH
L3 1T
L1 3T
J1 for 300 to 1500 c
33PF
100PF
S1-2 with/without power
S1-1 with/without power
C1 (1pf-18pf)
L6: 33uH Choke
L2a 1T (reverse)
L2b 1T
S1-2 with/without power

[http://1.bp.blogspot.com/-VmxFnJ9MFL4/VeXBJf9YzGI/AAAAAAAAB8A/N4GalWS1jUw/s1600/Screen%252520Shot%2525202015-02-09%252520at%25252012.47.38%252520am.png]

L1 using 2mm silver plated copper wire to wind 3 turns with a diameter of 25mm

L2a using 1mm silver plated copper wire to wind 1 turn in opposite direction as L1. with a diameter of 25mm

L2b same as L2a but wind 1 turn in same direction as L1.

L3 same as L2b.

L4 2uH Choke, can be DIY using 1.5M 1/4W resistor as the core, wind 30 turns using 0.07 to 0.1mm insulated copper wire.

L5, L6 33uH Choke, can be DIY using 1.5M 1W resistor as the core, wind 100 turns using 0.07mm insulated copper wire.

C1 Silver plated butterfly air variable capacitor (1pf-18pf) or other similar variable capacitor

J1 for 300 to 1500 ohm headphone

Posted 1st September 2015 by Billy

0 Add a comment

29th July 2015

Crystal Radio Tests in Tokyo

I just visited Tokyo and brought my umbrella crystal radio for AM reception, and my FM crystal radio for FM reception. Both the AM and FM crystal radios picked up a few local stations with strong signals.

I also visited Akihabara and bought a crystal radio earbud and another 1000 ohm hearing aid like ear bud.

Dynamic Views theme. Powered by Blogger.

AM Radio test

https://www.youtube.com/watch?v=ESitUNSQBdE [https://www.youtube.com/watch?v=ESitUNSQBdE]

FM Radio test

https://www.youtube.com/watch?v=-zFnImwYf1M [https://www.youtube.com/watch?v=-zFnImwYf1M]

Here are some photos from Akihabara.

[http://2.bp.blogspot.com/-bBpj3HiYi2E/VbjHniod_QI/AAAAAAAAB5g/PdZ0oTJZ7uA/s1600/IMG_2421.jpg]

Dynamic Views theme. Powered by Blogger.

拡声用スピーカー
税込 ¥2,160
税込 ¥2,160
税込 ¥2,480
拡声用スピーカー
TH-58 15W/8Ω
税込 ¥6,260
アシダ音響
DH-35 3W/8Ω
税込 ¥2,900
拡声用スピーカー
税込¥10,800
詳しくは、おたずね下さい
税込¥11,500

[http://3.bp.blogspot.com/-O7Agk8s9fDQ/VbjH4x0EJTI/AAAAAAAAB6M/7xqWnfZVorA/s1600/IMG_2426.jpg]

TOA防滴メガホン
アシダ音響
DH-35 3W/8Ω
税込 ¥2,900
ユニペックス
6W防滴メガホン
税込 ¥13,400
ユニペックス TR-215W
6W防滴ホイッスルメガホン
税込 ¥16,500
ユニペックス TR-215S
6W防滴サイレンメガホン
税込 ¥16,500
ユニペックス TRC-3W
3W小型メガホン
税込 ¥9,180
ビジネスホン用
KJ-
税込¥1
ヘッドセット専用
T-800
一般回線
税込¥10
HEADSET

[http://3.bp.blogspot.com/-B5XwTt6KWIQ/VbjH40ElQDI/AAAAAAAAB6Q/idnujwRwp7s/s1600/IMG_2427.jpg]

Dynamic Views theme. Powered by Blogger.

[http://1.bp.blogspot.com/-YvVtusSDHJI/VbjHk-tEdSI/AAAAAAAAB5w/by8ygCMXKeY/s1600/IMG_2406.jpg]

[http://3.bp.blogspot.com/-DiqXKq6EeyI/VbjHlDwu8hI/AAAAAAAAB4g/tPqw_ZGfsnk/s1600/IMG_2408.jpg]

Dynamic Views theme. Powered by Blogger.

[http://2.bp.blogspot.com/-8tsGBar5ojw/VbjHITRP5_I/AAAAAAAAB5o/sbnbi0co08A/s1600/IMG_2409.jpg]

[http://4.bp.blogspot.com/-6Rkff7aNQGA/VbjHlkbFgkI/AAAAAAAAB5k/8YEwEW6lSWI/s1600/IMG_2410.jpg]

[http://3.bp.blogspot.com/-dqAJpDMJVS0/VbjHmMR3c2I/AAAAAAAAB4w/9chEOWo7ZAA/s1600/IMG_2412.jpg]

「音の芯」が決まる!
銅箔オイル・コンデンサ
AMCA 600DCV
0.047μF 6,200円
0.1μF 6,400円
0.22μF 6,600円
0.47μF 7,000円
300B (国産)
TAKATSUKI

[http://1.bp.blogspot.com/-iEEJLvy4MXk/VbjHmSWAF2I/AAAAAAAAB5E/KZr9xh2I_tY/s1600/IMG_2413.jpg]

Dynamic Views theme. Powered by Blogger.

「音の芯」が決まる!
銅箔オイル・コンデンサ
AMCA 600DCV
0.047μF 6,200円
0.1μF 6,400円
0.22μF 6,600円
0.47μF 7,000円
300B (国産)
TAKATSUKI
843
6R-A3
6L6

[http://2.bp.blogspot.com/-nKnHfqnIdYw/VbjHmceWjmI/AAAAAAAAB5A/SaitukYdf4c/s1600/IMG_2414.jpg]

Stereo
10cmフルレンジ
オーディオクラフト 2015
管球王国
Stereo Sound
シンプルな回路でわかりやすい
真空管
オーディオアンプ
BCL

[http://2.bp.blogspot.com/-T7U6BVW2swg/VbjHmg5td4I/AAAAAAAAB5s/FZ-hgXbfjxg/s1600/IMG_2415.jpg]

Dynamic Views theme. Powered by Blogger.

直流安定化電源
安定化電源
パネルメータ
アナログテスタ
(03)3255-

[http://4.bp.blogspot.com/-bylbZZIuM8k/VbjHnBFPjSI/AAAAAAAAB5U/9dC0iCxFbGk/s1600/IMG_2417.jpg]

ASHIDA
MT-3
¥5,670
ATH-750COM
¥4,980
¥10,800
¥11,235

[http://4.bp.blogspot.com/--LuM52jSxuA/VbjHnmBntrI/AAAAAAAAB5c/2VW2JYlGdFI/s1600/IMG_2420.jpg]

100V BN-23
G ¥420-
目盛板30mm ¥100
目盛板38mm ¥100
目盛板55mm ¥100
BN-5701 ¥250
パイロットBF-4S
24V ¥300
パイロットBF-4S
12V
パイロットBF-4S
6V ¥300

[http://4.bp.blogspot.com/--rCKS5WLkIk/VbjHkaltuQI/AAAAAAAAB4I/MH2bXq1Sz0E/s1600/IMG_2405.jpg]

[http://1.bp.blogspot.com/-QFXKQzdi0Oc/VbjHl03VoBI/AAAAAAAAB40/2SrgRmWVw7s/s1600/IMG_2411.jpg]

[http://1.bp.blogspot.com/-c6qle61tzNw/VbjHm5FMuWI/AAAAAAAAB5I/V3pmpQKYg_M/s1600/IMG_2416.jpg]

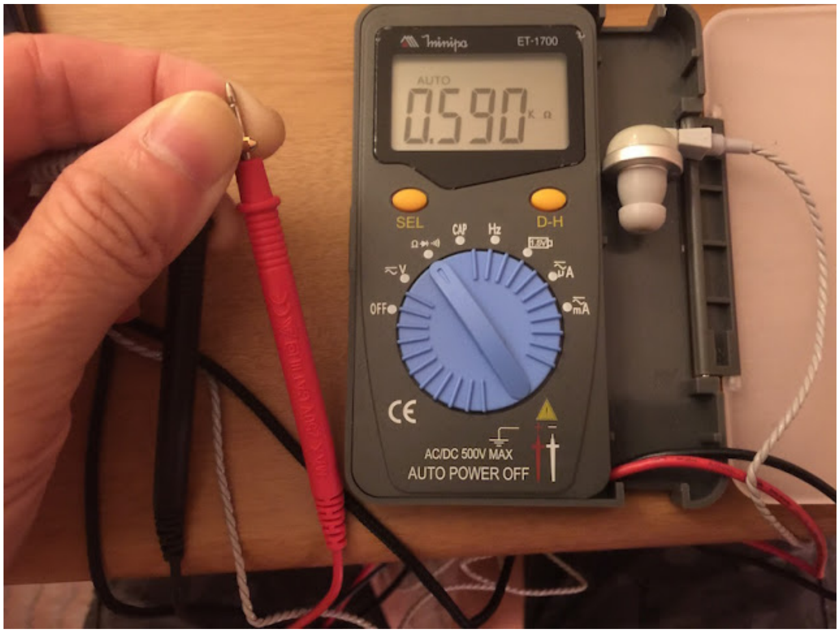

Minipa
ET-1700
0.590
SEL
D-H
AC/DC 500V MAX
AUTO POWER OFF

[http://2.bp.blogspot.com/-TyMUToycItE/VbjH44DwpzI/AAAAAAAAB6I/7wnhsXCVI5I/s1600/IMG_2439.jpg]

[http://1.bp.blogspot.com/-u1-_vlQXeSk/VbjHiN-PQ5I/AAAAAAAAB3U/D280a1ILjcQ/s1600/IMG_2389.jpg]

KHR-3HV ver.2
SALE
HPI
G-ROBOTS G-Dog
G-Dog 完成品
¥1,500

[http://3.bp.blogspot.com/-c66tvEwyHfE/VbjHiJlJsSI/AAAAAAAAB3Q/syIZ_rOMZJo/s1600/IMG_2390.jpg]

Dynamic Views theme. Powered by Blogger.

AFINIA 3D
H480
AFINIA
高温注意!
さわらないで!
Touch!
3Dプリンター
AFINIA H-480

[http://2.bp.blogspot.com/-wJTJm00gCVM/VbjHiN9IsVI/AAAAAAAAB3g/3ySdn9lUeal/s1600/IMG_2391.jpg]

[http://1.bp.blogspot.com/-GLQq7XcdTmU/VbjHi9CHU4I/AAAAAAAAB3w/CqL4LEV-Ksg/s1600/IMG_2395.jpg]

[http://4.bp.blogspot.com/-ZLbcAH_5cGU/VbjHi8thyCI/AAAAAAAAB6A/kggX6q4NA_M/s1600/IMG_2396.jpg]

アルプスコントローラ
1210002
¥520
アルプスコントローラ
122000D
¥570
レバー復帰
アルプスコントローラ
122400Y
スイッチ付

[http://4.bp.blogspot.com/-h6GLUWf6MdQ/VbjHjI2nbFI/AAAAAAAAB58/WAqIW_WjcBM/s1600/IMG_2398.jpg]

¥330-
¥390
¥470
¥330-

[http://2.bp.blogspot.com/-AvkXciZWfBE/VbjHjmKZnyI/AAAAAAAAB54/olex77UAvXI/s1600/IMG_2399.jpg]

¥620-
¥620-
¥1,030
¥330-
¥330-
¥470

[http://2.bp.blogspot.com/-QalYbD_n6Us/VbjHjo2DQYI/AAAAAAAAB38/kxzEwGPazPU/s1600/IMG_2400.jpg]

20pF×
ポリバリコン
260p
¥310
2連バリコン
440p×2
¥3600
単連バリコン
360p
¥1650
豆コン
60P
¥2060

[http://4.bp.blogspot.com/-HBpoBGSq-jY/VbjHkKwtppI/AAAAAAAAB50/OgsqSF_Zv0Q/s1600/IMG_2401.jpg]

SRRM
アルプスロータリSW
1-1-12
アルプスロータリSW
1-2-6
アルプスロータリSW
1-3-4
200V BN-5701
R/緑/橙 ¥380
100V BN-23
G ¥420-

[http://1.bp.blogspot.com/-iPANkGZLkmk/VbjHkV2WB0I/AAAAAAAAB4Q/rNbPaxYfSaE/s1600/IMG_2404.jpg]

Posted 29th July 2015 by Billy

0 Add a comment

8th August 2014

My blog [http://www.crystalradio.cn/forum.php?mod=viewthread&tid=579221&highlight=] [https://www.blogger.com/blogger.g?blogID=2311644141825422235]

My other blog articles

My blog

Dynamic Views theme. Powered by Blogger.

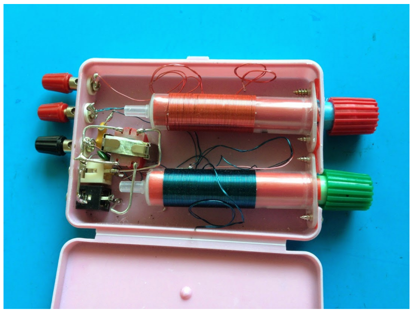

[http://1.bp.blogspot.com/-bZEAzgCCHTs/U-Ssydip9wI/AAAAAAAABsU/6QXHg9xPyho/s1600/IMG_8015.JPG]

主题	版块/群组	回复/查看	最后发帖
无天地线磁棒针筒调感矿机 [http://www.crystalradio.cn/thread-583577-1-1.html]	矿石机技术区 [http://www.crystalradio.cn/forum-3-1.html]	24 [http://www.crystalradio.cn/thread-583577-1-1.html] 464	cheungbx [http://www.crystalradio.cn/home.php?mod=space&username=cheungbx] 半小时前 [http://www.crystalradio.cn/forum.php?mod=redirect&tid=583577&goto=lastpost#lastpost]
转发：小收音机：仅1根纳米管可嵌入细胞 [http://www.crystalradio.cn/thread-582873-1-1.html]	矿石机技术区 [http://www.crystalradio.cn/forum-3-1.html]	13 [http://www.crystalradio.cn/thread-582873-1-1.html] 190	loveoldpc [http://www.crystalradio.cn/home.php?mod=space&username=loveoldpc] 4天前 [http://www.crystalradio.cn/forum.php?mod=redirect&tid=582873&goto=lastpost#lastpost]
移动磁棒调台的针筒矿机 [http://www.crystalradio.cn/thread-582382-1-1.html]	矿石机技术区 [http://www.crystalradio.cn/forum-3-1.html]	37 [http://www.crystalradio.cn/thread-582382-1-1.html] 534	cheungbx [http://www.crystalradio.cn/home.php?mod=space&username=cheungbx] 半小时前 [http://www.crystalradio.cn/forum.php?mod=redirect&tid=582382&goto=lastpost#lastpost]
针筒调感式矿机 [http://www.crystalradio.cn/thread-581679-1-1.html]	矿石机技术区 [http://www.crystalradio.cn/forum-3-1.html]	25 [http://www.crystalradio.cn/thread-581679-1-1.html] 617	cheungbx [http://www.crystalradio.cn/home.php?mod=space&username=cheungbx] 7天前 [http://www.crystalradio.cn/forum.php?mod=redirect&tid=581679&goto=lastpost#lastpost]

[http://www.crystalradio.cn/forum.php?mod=viewthread&tid=389699&highlight=]	= [http://www.crystalradio.cn/forum-3-1.html]	59 [http://www.crystalradio.cn/thread-389699-1-1.html] 4879	hjhj [http://www.crystalradio.cn/home.php?mod=space&username=hjhj] 2014-2-4 00:59 [http://www.crystalradio.cn/forum.php?mod=redirect&tid=389699&goto=lastpost#lastpost]

Posted 8th August 2014 by Billy

0 Add a comment

Squarespace Website Builder
Make and manage your own professional website with Squarespace's all-in-one platform.
OPEN

30th April 2014

My Crystal Radios

My Crystal Radios

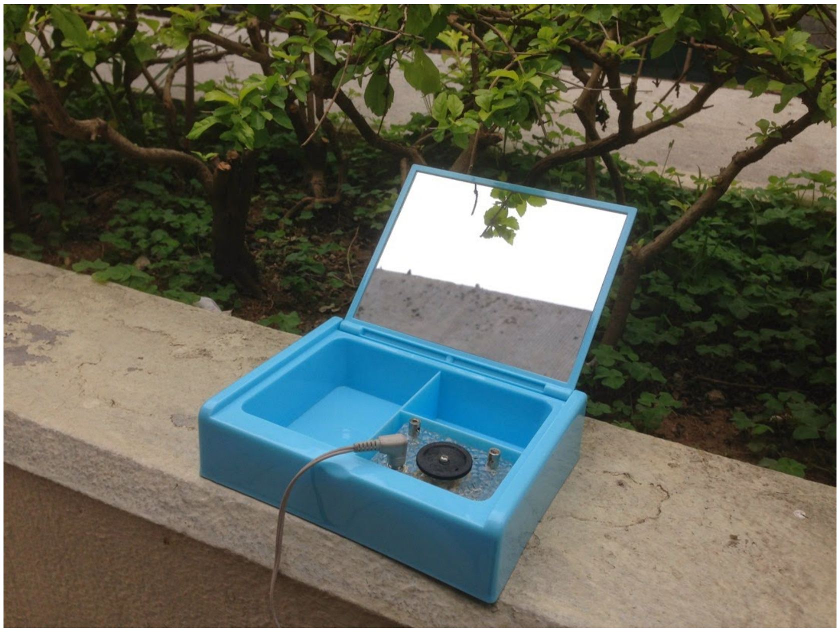

[http://1.bp.blogspot.com/-p1dMQBrjztA/U2EiFxxefvI/AAAAAAAABqI/jptO1nD6R9w/s1600/IMG_6461.JPG]

Dynamic Views theme. Powered by Blogger.

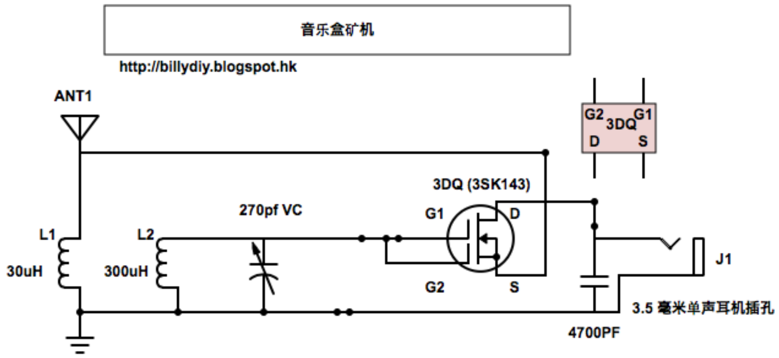

[http://2.bp.blogspot.com/-lASlH1QpBhc/U2EiPuCFjDI/AAAAAAAABqc/oMu2ZfomjjA/s1600/Circuit+Diagram.png]

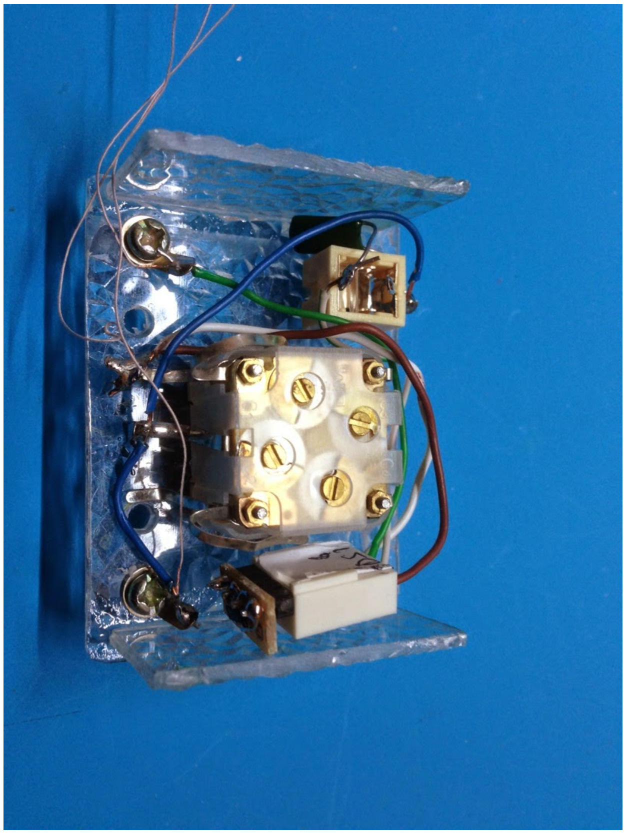

[http://4.bp.blogspot.com/-vq6rMk-1730/U2EiJJWvBUI/AAAAAAAABqQ/H_B8W4quwig/s1600/IMG_6458.JPG]

Do you know that you can listen to radio without battery ? I am holding such a radio in my hand now. It’s all powered by the radio energy.

I am going to share with you how I enjoyed my hobby - making crystal radios. We’ll explore what is a crystal radio, how to make one, and how you can benefit from this hobby.

When I was 10 years old, I was attracted to an article about a magical radio that worked without battery. I saved a few dollars, bought some electronic components and tried to make one. To my disappointment, I could failed. At that time, there is no one to ask or talk to. And I have to drop this hobby.

Last year, a youtube video about a special powerless radio caught my attention. I decided to re-do what I failed to do when I was a kid. I found a forum on the internet hosted in China called crystalradio.cn. After a lot of research and consulting people on the forum, I started making my first Crystal radio. I was overjoyed and could not sleep that night when I heard the the radio broadcast. Today, I have made more than 20 Crystal radios. As small as a pencil, and as big as as a Hula Loop. I even took part in the annual crystal radio DIY contest. My creation got into the second round, and ranked #12 of all the 40 submissions.

Now, let’s see how a Crystal Radio works.

In the old days, Radios were big and expansive. They consume a lot of power as the vacuum tubes inside are big and hot. Then someone replaced the vacuum tubes with much smaller crystals. Not only smaller, these radio can also work without battery. By just harvesting the radio energy, you can listen to a radio broadcast using a head phone. When the signal is exceptionally strong, it can even drive a loud speaker. One other benefit is that as there is no amplification, you can listen to the original sound without distortion.

As technology advanced, crystals were replaced by a electronic component called diodes.

You might now want to know how can you make a crystal radio ?

The simplest crystal radio is made up of 4 components: an antenna, a ground connection, a crystal or diode, and a headphone. This type of radio can only receive a single station as it cannot be tuned.

To add the tuning capacity, you can add a coil, a tuner (or capacitor). The tuner and the coil together allow you to tune to the station that resonant with a certain frequency.

There are two types of crystal radios:

The fixed crystal radios, and the portable crystal radios.

Fixed crystal radios need a long antenna (10 meters or above), and a connection to the ground (your water pipe), to compensate for the smaller coil. The signal will be much stronger than the portable one.

For example, these ones on the table are fixed crystal radios.

I have the tiny crystal radio, a super crystal radio that can receive medium wave and short wave, and the big crystal radio.

Portable crystal radios do away with the long antenna and the ground so you can walk freely with the radio.

To compensate for the lack of a long antenna, you need a bigger coil, at least half a meter in diameter.

Look at these, they are portable crystal radios. I can bring them to the street and listen to the radio.

I have the umbrella crystal radio, the hula loop crystal radio, the hat radio, the filing cabinet radio.

This one is a wearable crystal radio.

One thing I must tell you.

If you try to get on a plane with this on you'll be treated very specially, like a terrorist.

Feel free to try one of these crystal radios.

You can also refer to these links if you are interested to make one.

www.crystalradio.cn

billydiy.blogspot.hk

I find that having a hobby made my life complete. In the past, I just spent my spare time watching TV or films. It's a good thing to relax. However, I find that I enjoy creating things more. Making crystal radio satisfy my inner drive of creativity. It also gave me the stamina to focus on my job.

Do you have a hobby ? If not, I hope you can find one and enjoy it like I do.

Posted 30th April 2014 by Billy

1 View comments

4th January 2014

百变简易大环矿石收音机

这是一个或多个由超薄的印字机排线当线圈,和3DQ检波器组成的多用途，多效果的无天地线大环矿石收音机。电路是典型的3DQ检波电路，配合270pf 双联可变电容及利用超薄大环的不同配搭，做出多用途，多效果的百变无电源收音机组合。

[http://1.bp.blogspot.com/-4-CZKOQ4tWI/UsfGWB6r4TI/AAAAAAAABjw/alxyiOVAJGs/s1600/IMG_5096.JPG]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]

[http://www.blogger.com/blogger.g?blogID=2311644141825422235]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]

大环的主要材料：

用到一种薄薄的印字机排线，在电子令件店有售，，约20元一米，可以做个薄薄的大环，可用作矿石收音机的线圈, 效果还不下于用电线或漆包铜线绕的大环。此排线由1毫文的铜箔

做成. 表面面积大, 能改善趋肤效应带来的影响.铜片与铜片间距1.5毫米, 减少邻近效应,分布电容. 测试后発觉比用电脑排线好.响度加大了。

多效果：利用排线插头交错对接来做出”单股单回路"，或把把每两个排线交错对接，把邻近两个铜箔

一起做个双股线圈来做个"双股单回路"，捕捉多一些电波，增强接收效果。也可做个双回路来增强选择性。可以用一个排线的"单环双回路"或用两个排线的"双环双回路"。

更可做个双环增强回路：用双大环来增强3DQ的能量来增强输出。

多用途：大环矿机：利用排线插座和插头，这大环可打开，可闭合，方便穿进其他支架上，

例如可轻易穿进雨伞骨内做成雨伞矿机。

[http://1.bp.blogspot.com/-hoVRDP1ypwE/UsfGdNML12I/AAAAAAAABmY/7ly4VHRLrB8/s1600/IMG_5213.JPG]

[http://4.bp.blogspot.com/-okxpqY2otI8/UsfGdq3GdcI/AAAAAAAABmg/4vOJlGhYmMA/s1600/IMG_5214.JPG]

也可挂在窗户旁，当个室内无天地线矿机，

[http://1.bp.blogspot.com/-4-CZKOQ4tWI/UsfGWB6r4TI/AAAAAAAABjw/alxyiOVAJGs/s1600/IMG_5096.JPG]

更可做个小一点的大环，把它折成方形，発挥你的想像力，做出不同用法的无天地线的矿石收音机：
放上帽子边，做个帽子矿机

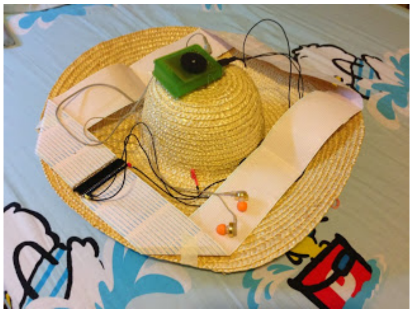

[http://2.bp.blogspot.com/-9Tq6fL7iPQI/UsfGV0m00sI/AAAAAAAABjk/7uRCQu9ebNQ/s1600/IMG_5094.JPG]

放入背包内，做个背包矿机，

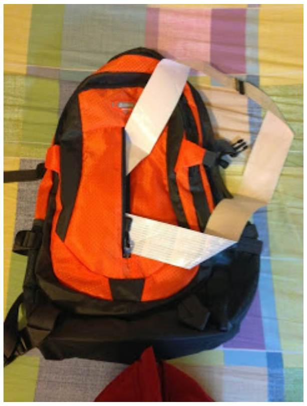

[http://1.bp.blogspot.com/-iUyk_0ARW74/UsfGalGseZI/AAAAAAAABlY/Hzsby0MnktE/s1600/IMG_5190.JPG]

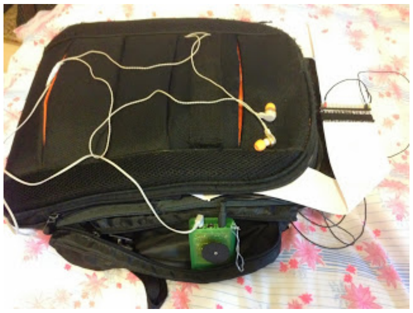

[http://2.bp.blogspot.com/-cw-78zBvXj4/UsfGZYJV1zI/AAAAAAAABk4/SRSozHBugWs/s1600/IMG_5106.JPG]

Dynamic Views theme. Powered by Blogger.

[http://1.bp.blogspot.com/-ftaBsa8aHQU/UsfGZkd75-I/AAAAAAAABIA/17rseky6hrE/s1600/IMG_5108.JPG]

放入马甲内，做个马甲矿机，

[http://4.bp.blogspot.com/-TFGyloqWfls/UsfGjcUFGKI/AAAAAAAABos/38-SgwwPOYE/s1600/IMG_5279.JPG]

[http://1.bp.blogspot.com/-5VGLAC9tL8Q/UsfGjxGriKI/AAAAAAAABoo/0PBvhxWPOro/s1600/IMG_5280.JPG]

放入文件匣内，做个文件匣矿机。

[http://4.bp.blogspot.com/-wEhQcrpVGrQ/UsfGaG7tNfI/AAAAAAAABIM/CBjW1VFUloY/s1600/IMG_5109.JPG]

Dynamic Views theme. Powered by Blogger.

[http://4.bp.blogspot.com/-0PJauwFuibA/UsfGadZ9PLI/AAAAAAAABlg/bseYl4fdt_U/s1600/IMG_5110.JPG]

做法：

线圈：

找一段印字机用的薄排线(24芯或以), 如买不到,可用电脑彩虹排线代替.

长度是乎用途, 如做雨伞矿机或室内大环用2.8米长的排线.

如做帽子,背包,马甲,文件匣, 用1.2米长的排线.

先用烙铁把排线上的塑料熔丢.露出铜片。再把排线的一端焊接到一个26针排针上，另一端焊接到一个26针排线插坐上再把它焊接到另一个26针排针上。（如用彩虹排线，因彩虹排线的间距离较密，要把两条线当一条线接）。

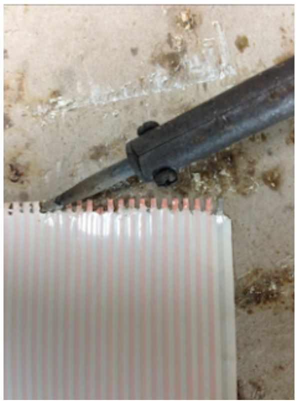

[http://4.bp.blogspot.com/-28bVQ0Z4G-w/UsfGldoFOpI/AAAAAAAABpI/17PeHg9Kg6E/s1600/Screen+Shot+2013-10-14+at+6.24.37+PM.png]

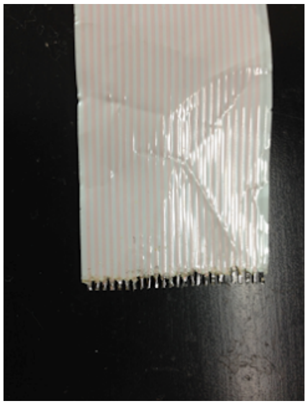

[http://3.bp.blogspot.com/-eGDtlkJZPSc/UsfGkyPgu1I/AAAAAAAABpM/HFmNLFD8NUg/s1600/Screen+Shot+2013-10-14+at+6.24.30+PM.png]

10
11
INCHES
6
5
4
3
2
1
0

[http://4.bp.blogspot.com/-rCqeleL-Bno/UsfGkZzMcSI/AAAAAAAABo4/M_M9nwKhEHc/s1600/Screen+Shot+2013-10-14+at+6.23.40+PM.png]

[http://3.bp.blogspot.com/-PwM-jbUSDpU/UsfGkKDeJEI/AAAAAAAABpQ/uJRVYsG56Bs/s1600/Screen+Shot+2013-10-14+at+6.22.49+PM.png]

排线插坐和排针是用来把大环的每一圈交错对接和把大环闭合。另一排针是用来做抽头，可匹配阻扩，又可改变波段去收听中波及短波扩播。

电路:

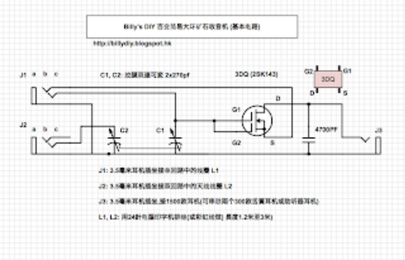
http://billydiy.blogspot.hk
J1 a b c
J2 a b c
C1, C2:
2x270pf
3DQ (2SK143)
G2
G1
3DQ
D
S
D
G1
G2
S
C2
C1
4700PF
J3

[http://4.bp.blogspot.com/-WsEjXrwUFYM/UsfGSe8Ow9I/AAAAAAAABiQ/iq2Ha6hX6jQ/s1600/Circuit+Diagram+1.png]

电路是典型的3DQ检波电路，调谐线圈L1 由插坐J1连接,天线线圈L2 由插坐J2连接.

不同的组合可做出不同的电路和效果.

参照电路图,把所有部件包括3dq, 可变电容，线圈插坐J1,J2,耳机插座J3等直接焊接到一小洞洞板上.再装进塑料盒子内。

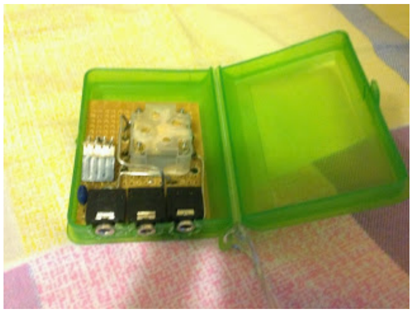

[http://1.bp.blogspot.com/-ONai7jeqyFU/UsfGgyr5OtI/AAAAAAAABno/juyVz8y3U6E/s1600/IMG_5270.JPG]

[http://2.bp.blogspot.com/-dMYP6gXJGRw/UsfGhiPXA-I/AAAAAAAABoA/pGgzNwzbSIg/s1600/IMG_5273.JPG]

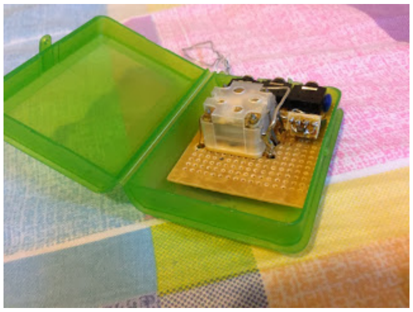

[http://3.bp.blogspot.com/-E6oa0ClBCcw/UsfGh5cCWKI/AAAAAAAABoM/cs8pJLp_Px8/s1600/IMG_5274.JPG]

把J1,J2接上不同颜色的两脚jumper来插进线圈抽头的排针上。接线插可把两个排针接在一起来做成多种组合:

[http://2.bp.blogspot.com/-GOIpPaOx-CA/UsfGiFSEaxI/AAAAAAAABoI/nYS0bc0ZU2c/s1600/IMG_5275.JPG]

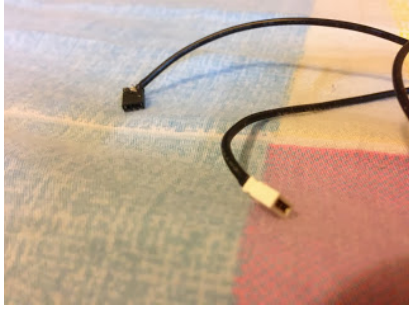

[http://1.bp.blogspot.com/-6NppNDS7Qa8/UsfGibLg0oI/AAAAAAAABoU/bBpXKHJQvW8/s1600/IMG_5276.JPG]

[http://4.bp.blogspot.com/-moXA9ynykz8/UsfGi16d9SI/AAAAAAAABo0/Hv5t8ZJk9mU/s1600/IMG_5277.JPG]

J1-黑色 jumper-接L1线圈地端

J1-红色 jumper-接L1线圈热端

J1-黄色 jumper-接3DQ源极线圈

J2-黑色 jumper-接L2线圈地端

J2-白色 jumper-接L2线圈热端

排线大环的不同组合：

单股单回路:

如做2.8米长的大环,可做阳台上的戸内矿机

如做1.2米长的小大环可用于帽子，背包，马甲，文件匣矿机等：

[http://1.bp.blogspot.com/-4-CZKOQ4tWI/UsfGWB6r4TI/AAAAAAAABjw/alxyiOVAJGs/s1600/IMG_5096.JPG]

把排针交错插进排线插坐上, 例如排针的第一针插进排线插坐的第二洞内, 第二针插进插坐的第三洞内, ...,第廿三针插进插坐的第廿四洞内.

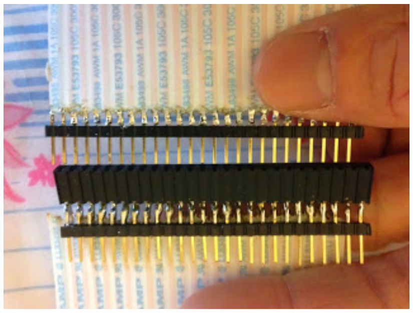

[http://3.bp.blogspot.com/-

再把J1-黄色jumper的其中一脚插到大环的第1匝的排针上

把J1-黑色jumper的其中一脚插到大环的第2匝的排针上

把J1-红色jumper的其中一脚插到大环的第24匝的排针上

这便可把排线的每一线连成大环中的每一匝，做成一个廿四匝的大环.

-接L1线圈地端

J1-红色-接L1线圈

再把J1J1-黄色-接3DQ源极线圈

J2-黑色-接L2线圈地端

J2-白色-接L2线圈

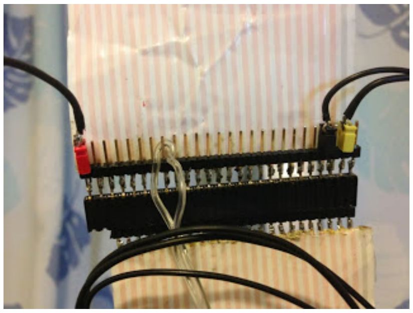

[http://3.bp.blogspot.com/-EpvOw2N4JoY/UsfGXaiYQYI/AAAAAAAABkM/nDsYNdrxhm0/s1600/IMG_5101.JPG]

双股单回路:

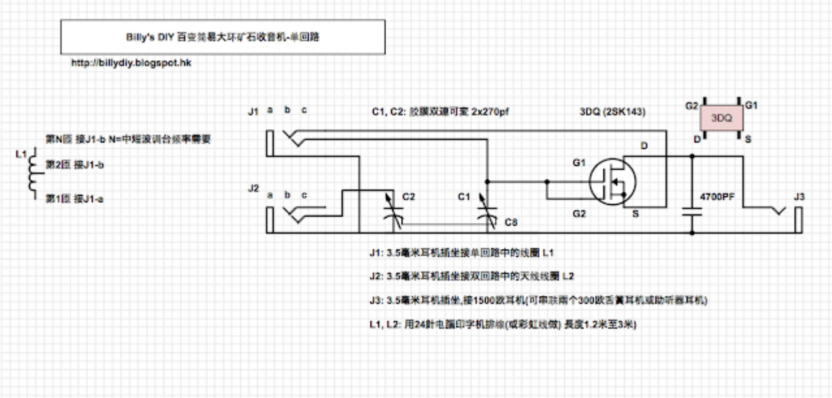
Billy's DIY 百变简易大环矿石收音机-单回路
http://billydiy.blogspot.hk
J1 a b c
C1, C2: 胶膜双連可変 2x270pf
3DQ (2SK143)
G2
G1
3DQ
D
S
第N匝 接J1-b N=中短波调台频率需要
L1
第2匝 接J1-b
第1匝 接J1-a
J2 a b c
C2
C1
C8
G1
G2
D
S
4700PF
J3
J1: 3.5毫米耳机插坐接单回路中的线圈 L1
J2: 3.5毫米耳机插坐接双回路中的天线线圈 L2
J3: 3.5毫米耳机插坐,接1500欧耳机(可串联两个300欧舌簧耳机或助听器耳机)
L1, L2: 用24針电腦印字机排線(或彩虹线做) 長度1.2米至3米)

[http://4.bp.blogspot.com/-VDWlp1wp4EM/UsfGSWgcOiI/AAAAAAAABiY/C7X23p1w-ks/s1600/Circuit+Diagram+2.png]

可用于2.8米长的大环。

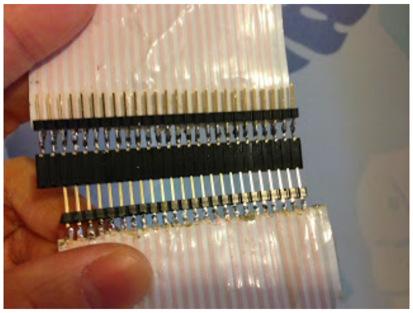

[http://2.bp.blogspot.com/-XOO7T-oHdDs/UsfGXwlc0oI/AAAAAAAABkY/mvXKxZrWp-8/s1600/IMG_5102.JPG]

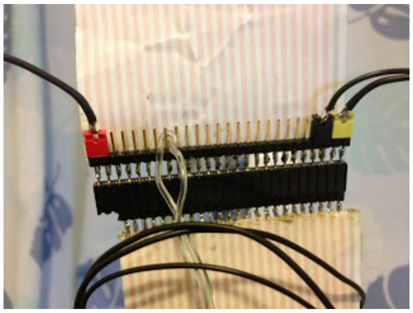

[http://4.bp.blogspot.com/-LvTNC6j4zo8/UsfGXUj7nnI/AAAAAAAABkE/h0AfoCmdhgY/s1600/IMG_5100.JPG]

把排针每两针交错插进排线插坐上, 例如排针的第一针插进排线插坐的第三洞内, 第二针插进插坐的第四洞内,第三针插进插坐的第五洞内,第四针插进插坐的第六洞内, ...,第廿一针插进插坐的第廿三洞内, 第廿二针插进插坐的第廿四洞内.
再把J1-黄色jumper的两脚插到大环的排针#1 和#2上,使他短路.

把J1-黑色jumper的两脚插到大环的排针#3 和#4上,使他短路.

把J1-红色jumper的两脚插到大环的排针#23 和#24上,使他短路.

这便可把排线的每一线连成大环中的每一匝, 做一个双骰的十二匝的大环.

单环双回路:

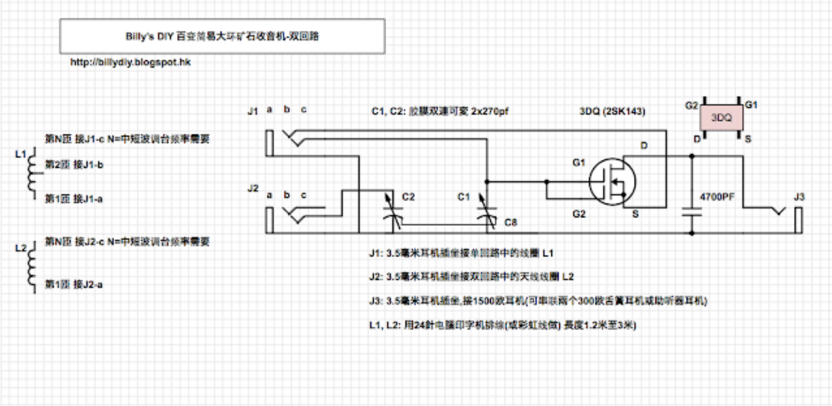
Billy's DIY 百变简易大环矿石收音机-双回路
http://billydiy.blogspot.hk
L1
第N匝 接J1-c N=中短波调台频率需要
第2匝 接J1-b
第1匝 接J1-a
L2
第N匝 接J2-c N=中短波调台频率需要
第1匝 接J2-a
J1 a b c
J2 a b c
C1, C2: 胶膜双連可变 2x270pf
3DQ (2SK143)
G2 G1 3DQ D S
C2
C1
C8
G1
G2
D
S
4700PF
J3
J1: 3.5毫米耳机插坐接单回路中的线圈 L1
J2: 3.5毫米耳机插坐接双回路中的天线线圈 L2
J3: 3.5毫米耳机插坐,接1500欧耳机(可串联兩个300欧舌簧耳机或助听器耳机)
L1, L2: 用24針电腦印字机排線(或彩虹线做) 長度1.2米至3米)

[http://2.bp.blogspot.com/-kWkDFT9avfM/UsfGSr8LSuI/AAAAAAAABic/Gn2z9IR9ZVI/s1600/Circuit+Diagram+3.png]
可用于2.8米长的大环。

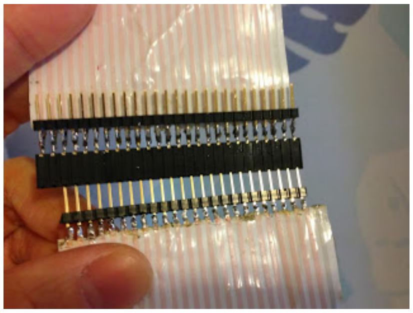

[http://2.bp.blogspot.com/-XOO7T-oHdDs/UsfGXwlc0oI/AAAAAAAABkY/mvXKxZrWp-8/s1600/IMG_5102.JPG]

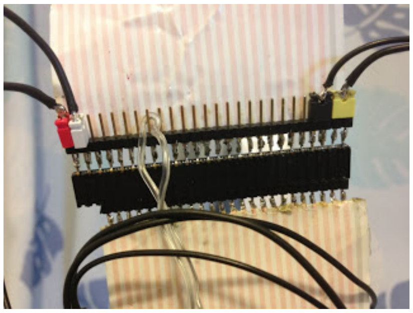

[http://4.bp.blogspot.com/-ya1XeEMghEw/UsfGW24mOSI/AAAAAAAABkg/5RAgG5inhvc/s1600/IMG_5099.JPG]

把排针每两针交错插进排线插坐上, 例如排针的第一针插进排线插坐的第三洞内, 第二针插进插坐的第四洞内,第三针插进插坐的第五洞内,第四针插进插坐的第六洞内, ...,第廿三针插进插坐的第廿五洞内, 第廿四针插进插坐的第廿六洞内.

再把J1-黄色jumper的两脚插到大环的排针#1和#2上

把J1-黑色jumper的的两脚插到大环的排针#3和#4上

把J1-红色jumper的其中一脚插到大环的排针#23上

把J2-白色jumper的其中一脚插到大环的排针#24上

这便可做成两个十二匝的大环, 可做双回路.

双环双回路;

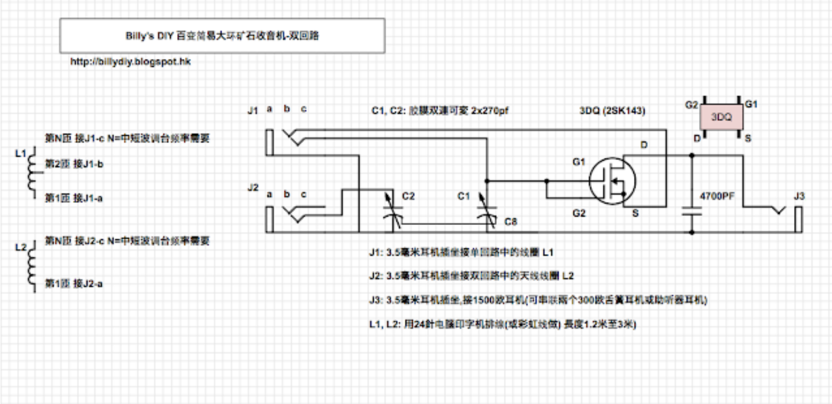
Billy's DIY 百变简易大环矿石收音机-双回路
http://billydiy.blogspot.hk
第N匝 接J1-c N=中短波调台频率需要
L1
第2匝 接J1-b
第1匝 接J1-a
L2
第N匝 接J2-c N=中短波调台频率需要
第1匝 接J2-a
J1 a b c
J2 a b c
C1, C2: 胶膜双連可変 2x270pf
3DQ (2SK143)
G2 G1 3DQ D S
D G1 G2 S
C2 C1 C8
4700PF
J3
J1: 3.5毫米耳机插坐接单回路中的线圈 L1
J2: 3.5毫米耳机插坐接双回路中的天线线圈 L2
J3: 3.5毫米耳机插坐,接1500欧耳机(可串联兩个300欧舌簧耳机或助听器耳机)
L1, L2: 用24針电腦印字机排線(或彩虹线做) 長度1.2米至3米)

[http://2.bp.blogspot.com/-kWkDFT9avfM/UsfGSr8LSuI/AAAAAAAABic/Gn2z9IR9ZVI/s1600/Circuit+Diagram+3.png]

可用于2.8米长的大环。

[http://4.bp.blogspot.com/-TD8htYjMlrg/UsfGb995gXI/AAAAAAAABl4/x4NK_mK5whw/s1600/IMG_5205.JPG]

用相同长度的排线做两个大环,大环#1 和大环#2 .

把排针每两针交错插进排线插坐上, 例如排针的第一针插进排线插坐的第三洞内, 第二针插进插坐的第四洞内,第三针插进插坐的第五洞内,第四针插进插坐的第六洞内, ...,第廿三针插进插坐的第廿五洞内, 第廿四针插进插坐的第廿六洞内.

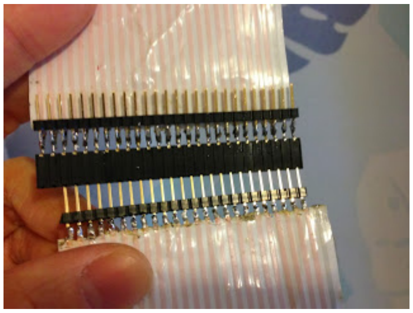

[http://2.bp.blogspot.com/-XOO7T-oHdDs/UsfGXwlc0oI/AAAAAAAABkY/mvXKxZrWp-8/s1600/IMG_5102.JPG]

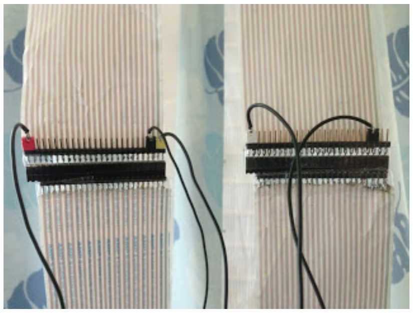

[http://4.bp.blogspot.com/-O70eMGlpUi0/UsfGe7Fn8pI/AAAAAAAABnA/KWUwo_2p4Ds/s1600/IMG_5246.JPG]

再把J1-黄色jumper的两脚插到大环#1的排针#1 和#2上,使他短路.

把J1-黑色jumper的两脚插到大环#1的排针#3 和#4上,使他短路.

把J1-红色jumper的两脚插到大环#1 的排针#25 和#26上,使他短路.

把J2-黑色jumper的两脚插到大环#2的排针#3 和#4上,使他短路.

把J2-白色jumper的两脚插到大环#2 的排针#25 和#26上,使它短路.

这便可把排线的每一线连成大环中的每一匝, 做一个双骰的十二匝的大环.

双环增强回路:

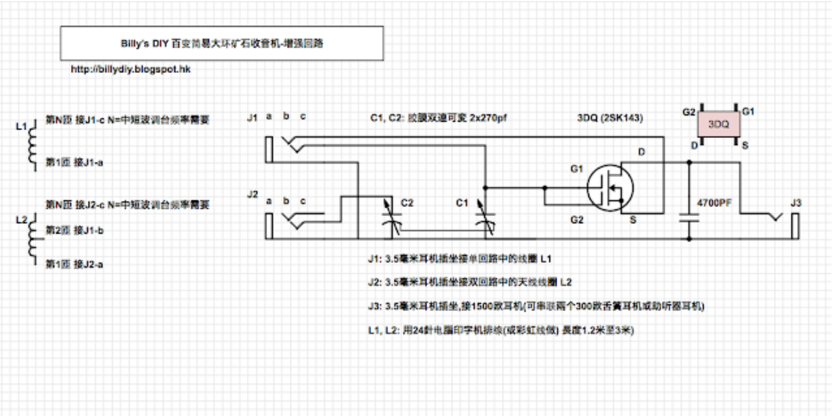
Billy's DIY 百变简易大环矿石收音机-增强回路
http://billydiy.blogspot.hk
L1
第N匝 接J1-c N=中短波调台频率需要
第1匝 接J1-a
L2
第N匝 接J2-c N=中短波调台频率需要
第2匝 接J1-b
第1匝 接J2-a
J1 a b c
J2 a b c
C1, C2: 胶膜双连可变 2x270pf
C2
C1
3DQ (2SK143)
G1
G2
D
S
3DQ
4700PF
J3
J1: 3.5毫米耳机插坐接单回路中的线圈 L1
J2: 3.5毫米耳机插坐接双回路中的天线线圈 L2
J3: 3.5毫米耳机插坐,接1500欧耳机(可串联两个300欧舌簧耳机或助听器耳机)
L1, L2: 用24針电脑印字机排線(或彩虹线做) 長度1.2米至3米)

[http://1.bp.blogspot.com/-pjmSWnaNgbU/UsfGTPv1JlI/AAAAAAAABik/31Ma7cyL-1c/s1600/Circuit+Diagram+4.png]

这电路原作者是lq19512003老师的无天地线磁棒矿机,现改良用于大环上.

原理是使用两个距离远到互不干扰的线圈来吸收两倍的无线电能量.

一倍能量用来检波，另一倍提供额外能量给3DQ源极去检波以增强输出.

原理请看h ttp://www.crystalradio.cn/forum.php?mod=viewthread&tid=237270

做法和双环双回路相似,唯一不同是把再把J1-黄色jumper的两脚改插到大环#2的排针#1 和#2上,使他短路.

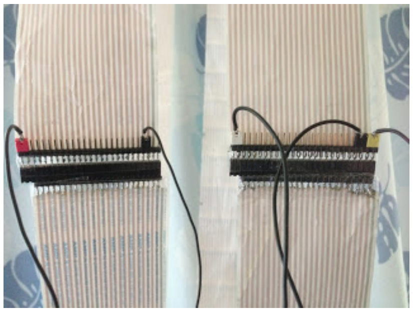

[http://3.bp.blogspot.com/-vW5KGxW_l8k/UsfGeytnPJI/AAAAAAAABnU/D1z-eTe-NJ8/s1600/IMG_5247.JPG]

矿机调整测试：

把耳机插进耳机插坐. 我用的耳机是两个150欧的助听器耳机串联. 也可用两个300欧的舌簧耳机串联，或接到T725阻矿匹配变压器的2.5千欧端来匹配不同阻扩的耳机。

单回路电路的调整：

调整可变电容调到想听的电台。也可改变红色jumper 接线圈的位置来调到不同的波段。

如要收短波，可把红色jumper 接到线圈的第二匝。 ，

双回路电路的调整：

调整L1和L2的距离使选择性加强，声量也可接受.

调整可变电容调到一个低端的电台。再调整可变电容上的两个微调电容，使响度最大。

再调整可变电容调到一个高端的电台。再调整可变电容上的两个微调电容，使响度最大。重覆以上多一次。

然后可以调到想听的电台。无天地线接收良好，能收数个本地强电台。

双环增强回路电路的调整：

把L1和L2的距离增大使互不干扰

调整可变电容调到一个低端的电台。再调整可变电容上的两个微调电容，使响度最大。

再调整可变电容调到一个高端的电台。再调整可变电容上的两个微调电容，使响度最大。重覆以上多一次。

然后可以调到想听的电台。无天地线接收良好，能收数个本地强电台。

视频展示:
http://v.youku.com/v_show/id_XNjU2MjAwMDg0.html [http://v.youku.com/v_show/id_XNjU2MjAwMDg0.html]

Posted 4th January 2014 by Billy

0 Add a comment

18th October 2013

Ball Point Pen Crystal Radio

Ball Point Pen Crystal radio has been featured in 007 movies as a tool for spies. It's been manufactured in the 50's for hobbies. Now it has become a collector item and cost above HK$1000.
Now you can DIY yourself to make one without costing a fortune.
And use it like a Spy. Nobody will know that the crystal radio is hidden inside a ball pen.
Check out this Youtube Video: http://www.youtube.com/watch?v=VsNDIoLwYPs [http://www.youtube.com/watch?

Dynamic Views theme. Powered by Blogger.

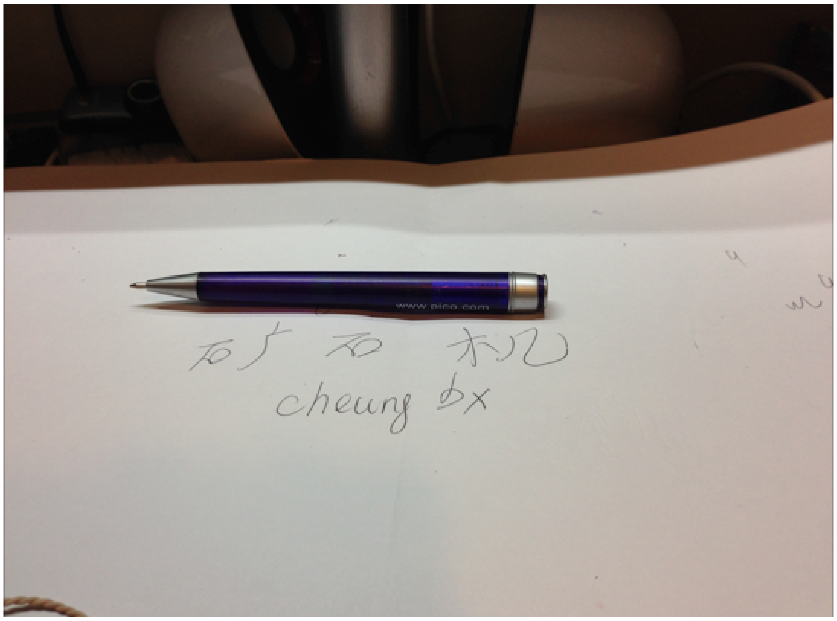
www.pico.com
矿 石 机
cheung bx

[http://4.bp.blogspot.com/-YEyeHrhBKg0/UmFghztqAnI/AAAAAAAABgg/4SnbM1iUJYw/s1600/Screen+Shot+2013-10-18+at+1.38.31+AM.png]

CHUNG HWA
SHANGHAI, CHINA FIRST PENCIL CO., LTD.
PICO

[http://3.bp.blogspot.com/-AKov792511Y/UmFghyM1smI/AAAAAAAABgw/kZECAAc83dw/s1600/Screen+Shot+2013-10-18+at+1.04.46+AM.png]

Dynamic Views theme. Powered by Blogger.

CHUNG HWA
PICO

[http://4.bp.blogspot.com/-A2tc95w_s_Q/UmFgh6MzkdI/AAAAAAAABg0/xaxvF0O5wNk/s1600/Screen+Shot+2013-10-18+at+1.04.26+AM.png]

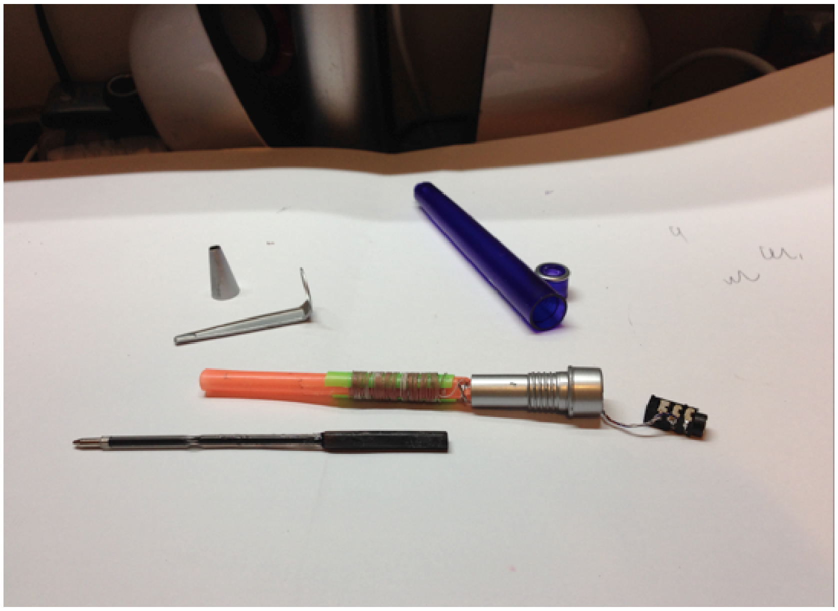

[http://4.bp.blogspot.com/-KOxoP449pMs/UmFgi2zLqBI/AAAAAAAABhE/SVzdo3UbaEs/s1600/Screen+Shot+2013-10-18+at+1.39.27+AM.png]

[http://4.bp.blogspot.com/-2FICJcSJ-gY/UmFgjLUpDMI/AAAAAAAABhQ/6rrnnoi2o_4/s1600/Screen+Shot+2013-10-18+at+1.39.47+AM.png]

[http://1.bp.blogspot.com/-VKTXAnc3FxE/UmFgj3u-FFI/AAAAAAAABhY/FmtrUKa00Wk/s1600/Screen+Shot+2013-10-18+at+1.39.55+AM.png]

Dynamic Views theme. Powered by Blogger.

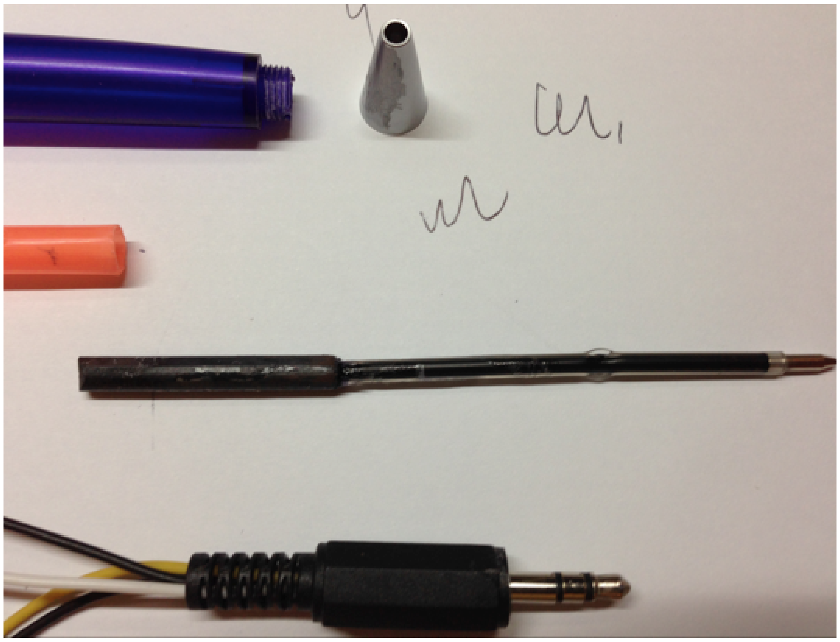

[http://2.bp.blogspot.com/-sOY3yknp_D8/UmFgkJd9EXI/AAAAAAAABho/8uoN3HKazEY/s1600/Screen+Shot+2013-10-18+at+1.40.04+AM.png]

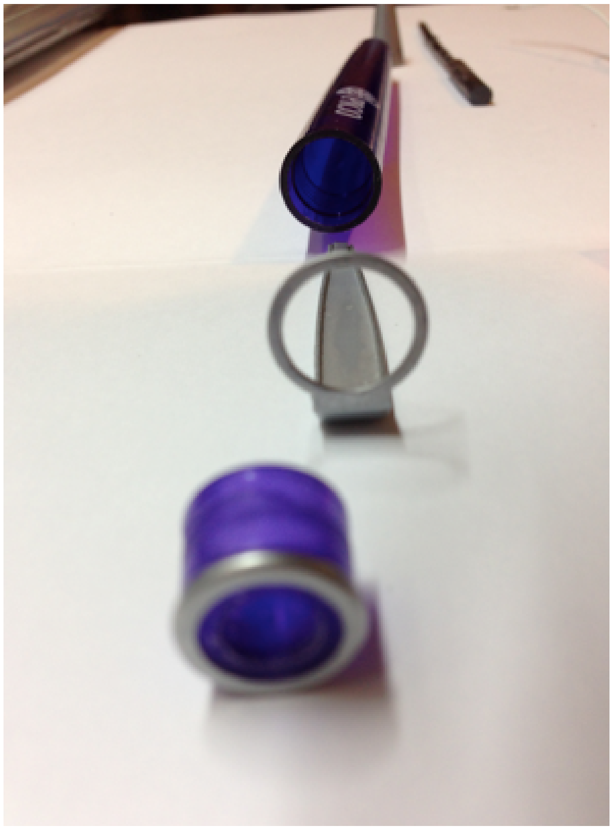

[http://2.bp.blogspot.com/-Wps5tmelSLY/UmFgkDkvs7I/AAAAAAAABhg/-QoCWIYd5Fo/s1600/Screen+Shot+2013-10-18+at+1.40.27+AM.png]

Dynamic Views theme. Powered by Blogger.

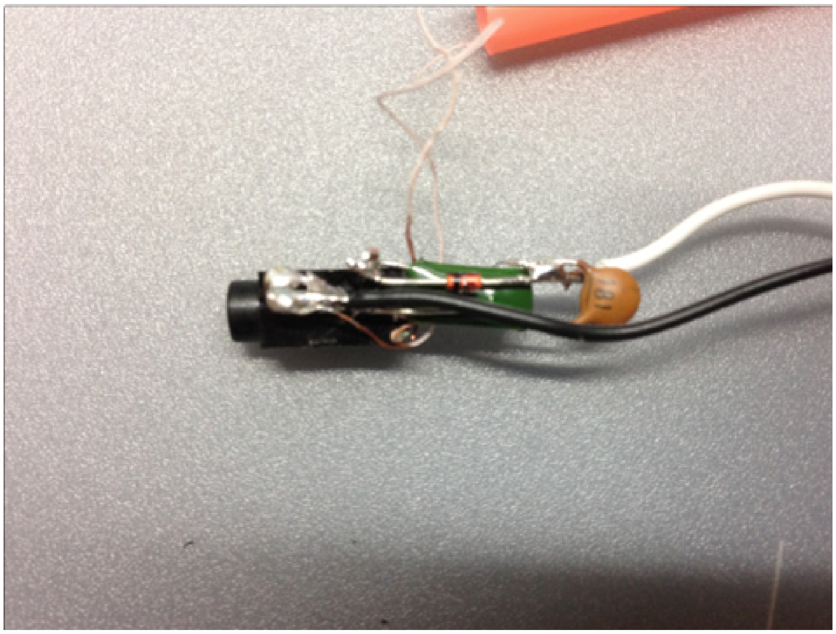

[http://2.bp.blogspot.com/-kipBW--wVKU/UmFgkihFHjI/AAAAAAAABhs/JliACoZKDFA/s1600/Screen+Shot+2013-10-18+at+2.14.21+AM.png]

在网上看到圆珠笔矿机，现在已成收藏品，售价 $1000 高得有点过份。
我成功DIY了一个圆珠笔矿机，改良了一些, 比网上的更好：
1，此圆珠笔矿机能用来书写。
2。没有固定连着的线，只在笔顶装了个微型双声道插坐，接耳机，天地线。
要另外做一根转换线来把双声道插分成单声道耳机和天地线。
3，单凭外型，你没法知道这是部矿机。

花了不小时间找合用的圆珠笔及小型磁棒。
数天的试验，用了四枝不同的圆珠笔。
圆珠笔矿机终于做成。

做法：
用吸管做线架，用0.04x14股线绕了两层共25毫文长，7毫米粗的线圈,加了磁棒在最中央位置時最高电感约 480uH.
用个30毫米长4毫米粗的小型磁棒在吸管内作拉杆式调台，
配合180PF固定电容（加上线圈和其他引线产生的约10PF分布定电容）可覆盖大部分中波频率。

有待改良：
因是微型制作，只用了单回路及BAT85二极管检波。
有空可改用3DQ电路，直接焊在贴片上。
也可加绕线圈，改成双回路，可改善现在的轻微串台。
但不知用调感式调台能用双回路吗？
现在的笔芯是用AB胶固定的，不能更换。
可改用固定在小型磁棒上的小圆管来接入笔芯，方便更换

在家中加上天地线后接收良好，能收所有强台。
当天线太强时，有轻微串台，要用天调修正。

今天带了圆珠笔矿机回单位，接上我做的排线间易大环当天线，用电源地线当地线, 能收数个强台。

也吸引了各同工的眼球。
大家都说我是危险人物，什么间谍用品都有。

补充内容
加了了个线圈抽头接检波二极, 选择性好了.
用胶布把笔芯固定现可更换.
加了海绵在拉杆和吸管中间, 防滑防走台.
上了测试及圆珠笔矿机解体的视频.可到20樓看看.
http://v.youku.com/v_show/id_XNjIzMzIzMTg0.html [http://v.youku.com/v_show/id_XNjIzMzIzMTg0.html]

Posted 18th October 2013 by Billy

1 View comments

Great Deal At Oakland Ca!! in Oakland...
$499,000
Hot Off Market Way Below Market...
$140,000
Hot Deal At Oakland Ca!!! in Oal
$650,000

12th October 2013

Simple AM Transmitter.

Create an AM transmitter to broadcast your own AM radio station to test Crystal Radio.
Range is within a few meters.

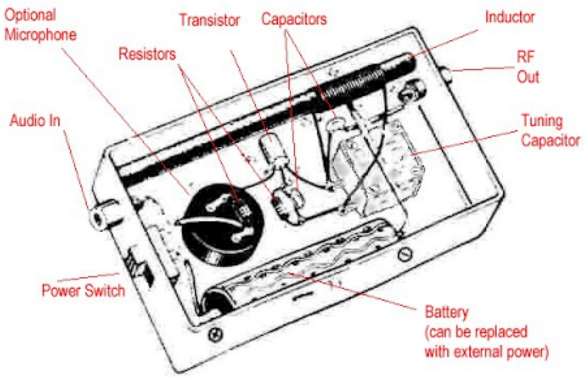
Optional Microphone
Transistor
Capacitors
Inductor
Resistors
RF Out
Audio In
Tuning Capacitor
Power Switch
Battery (can be replaced with external power)

[http://3.bp.blogspot.com/-k0b9RVoeD-o/UlliE_exm2I/AAAAAAAABgM/9_cVc0QnPl8/s1600/Screen+Shot+2013-10-12+at+10.33.36+PM.png]

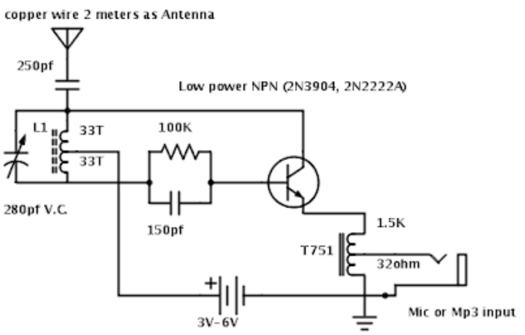
copper wire 2 meters as Antenna
250pf
Low power NPN (2N3904, 2N2222A)
L1
33T
33T
100K
280pf V.C.
150pf
1.5K
T751
32ohm
+
3V-6V
Mic or Mp3 input

L1: Wound 66 turns on a 10mmx 100mm powder core, tap at 33 turns. Using 0.4mm magnet wire.

[http://3.bp.blogspot.com/-_IdsAQVXeY4/UlliE7UQzSI/AAAAAAAABgI/6VxeycMemao/s1600/Screen+Shot+2013-10-12+at+10.47.17+PM.png]

Posted 12th October 2013 by Billy

0 Add a comment

25th September 2013

Portable Antenna Free Powder Core Crystal Radio

Dynamic Views theme. Powered by Blogger.

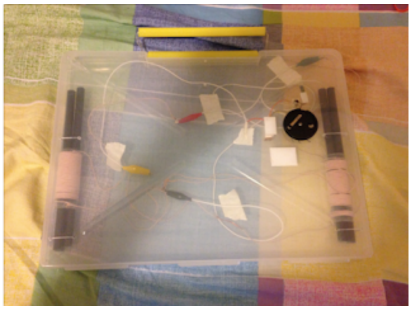

[http://4.bp.blogspot.com/-izv3DGadbUw/UkMCaJssgCI/AAAAAAAABeQ/p39Y_vlN8Jo/s1600/Screen+Shot+2013-09-24+at+11.28.38+PM.png]

Because space is limited, only thin-film variable capacitor can be used.
However, it does not effect the function of the original circuit.

The radio can receive all local stations without an antenna.

When tested in the streets, I attracted some passers-by. They saw a transparent enclosure, and try looking for batteries.
When they can not find it, they asked me why.
I explained to them that a Crystal Radio runs without battery.
They were amazed.

Everybody try to classify crystal radio enthusiasts into types: technical, artistic. I think I might be practical, as I tend to create portable crystal radios that I can listen to on the road.
For example, I can bring this portable radio to the office, stand it up as a normal file, and listen to crystal radio when I have a coffee break.

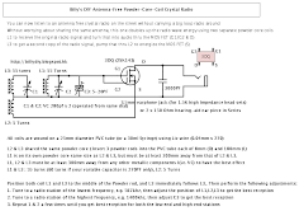

[http://4.bp.blogspot.com/-HqJDAJAzlPw/UkMCaKAWXPI/AAAAAAAABeU/4HFZ6v_6jBA/s1600/Antenna-Free-Powder-Core-Coil-Crystal-Radio.png]

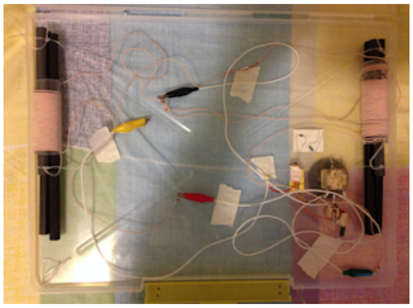

[http://1.bp.blogspot.com/--jORnWCTpIA/UkMCaCaiN4I/AAAAAAAABeY/plvWp5UHWH8/s1600/Screen+Shot+2013-09-24+at+11.29.03+PM.png]

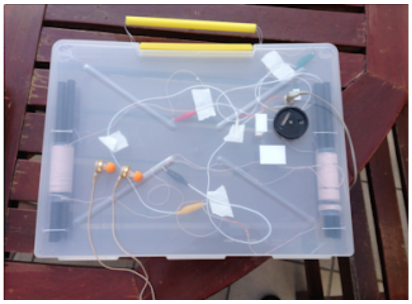

[http://3.bp.blogspot.com/-d8s4ZsgRUl4/UkMCvSyunMI/AAAAAAAABfM/aHaGstR9vlk/s1600/Screen+Shot+2013-09-25+at+10.09.40+PM.png]

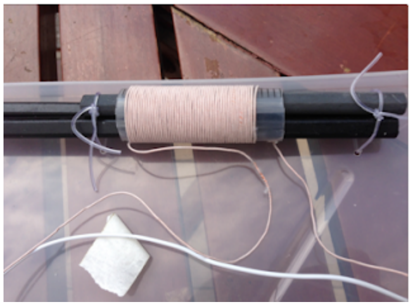

[http://1.bp.blogspot.com/-UdMo7FoYpmY/UkMCvR3keal/AAAAAAAABfQ/-Q6t8f7pt7w/s1600/Screen+Shot+2013-09-25+at+10.10.19+PM.png]

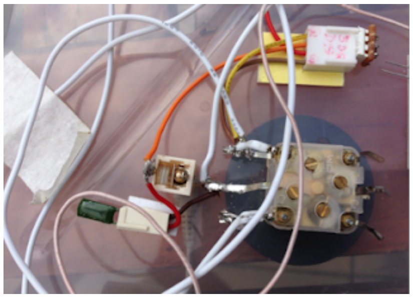

[http://3.bp.blogspot.com/-W5Yy5u_XLys/UkMCv5nFCcI/AAAAAAAABfY/ED4JKVPKKT0/s1600/Screen+Shot+2013-09-25+at+10.10.28+PM.png]

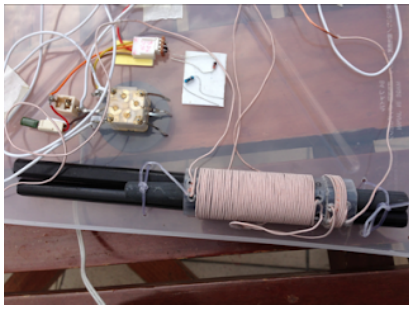

[http://2.bp.blogspot.com/-1A8IxN3oDEM/UkMCwAp7zBI/AAAAAAAABfg/BIWOFC7GSwg/s1600/Screen+Shot+2013-09-25+at+10.10.38+PM.png]

[http://3.bp.blogspot.com/-3DBToSIGbOM/UkMCwX11b0I/AAAAAAAABfo/XYItAyzvY08/s1600/Screen+Shot+2013-09-25+at+10.10.47+PM.png]

Dynamic Views theme. Powered by Blogger.

[http://3.bp.blogspot.com/-ptWV1nQ2PnY/UkMCxO4PChI/AAAAAAAABfw/wQeReC9hTzw/s1600/Screen+Shot+2013-09-25+at+10.10.55+PM.png]

Posted 25th September 2013 by Billy

0 Add a comment

18th September 2013

Crystal Radio Testing in Mumbai

I brought my umbrella crystal radio to India. In a hotel close to the sea in Mumbai, I was able to receive a few local AM stations. The volume was really high.
See this video. http://youtu.be/ip3z3q34N0E [http://youtu.be/ip3z3q34N0E]

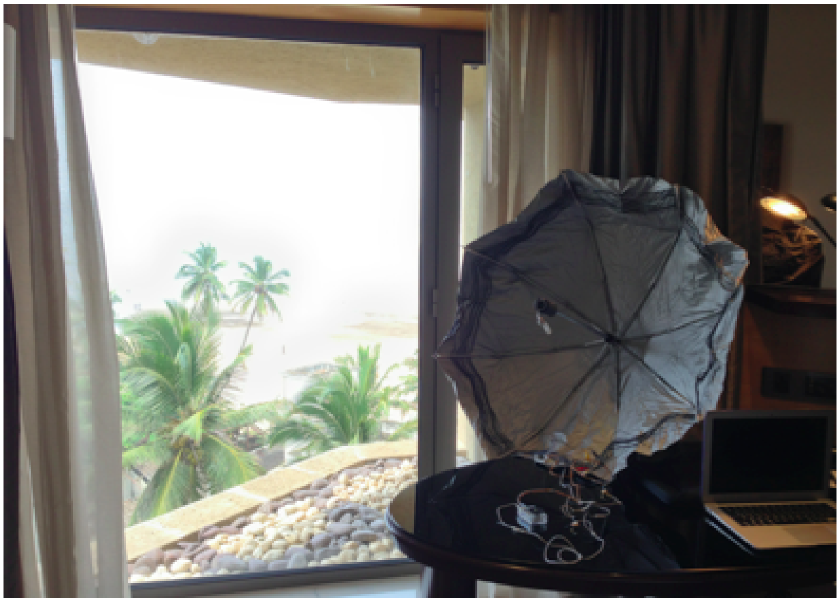

[http://3.bp.blogspot.com/-It4GcDMtXjI/Ujm4klnK91I/AAAAAAAABd0/fESNZtEbARQ/s1600/Screen+Shot+2013-09-12+at+12.05.30+PM.png]

JW MARRIOTT

[http://1.bp.blogspot.com/-HsVOyO-wEWI/Ujm4jlqfEJI/AAAAAAAABdQ/QCAGtdubgeE/s1600/Screen+Shot+2013-09-12+at+12.04.56+PM.png]

[http://1.bp.blogspot.com/-It4GcDMtXjI/Ujm4kInK91I/AAAAAAAABd4/YzGUTFPo7Ik/s1600/Screen+Shot+2013-09-12+at+12.05.30+PM.png]

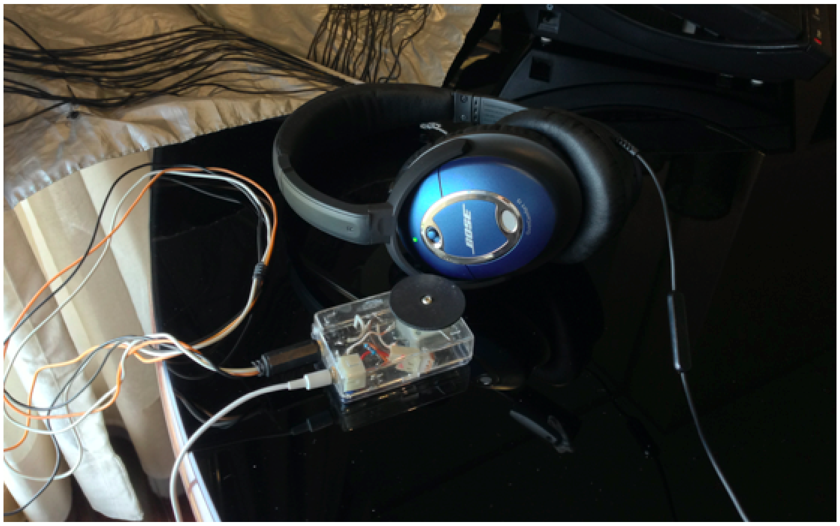

[http://3.bp.blogspot.com/-W0_Yk26Ydes/Ujm4mN_XPRI/AAAAAAAABds/8p-g9JvZcwY/s1600/Screen+Shot+2013-09-12+at+12.28.25+PM.png]

[http://4.bp.blogspot.com/-vQeQ4INqHZc/Ujm4kwfyFSI/AAAAAAAABdY/aqVxHhQ6mec/s1600/Screen+Shot+2013-09-12+at+12.05.52+PM.png]

[http://3.bp.blogspot.com/-VNpSwIwYeFo/Ujm4lBXV96I/AAAAAAAABdw/0_bo67TmaQs/s1600/Screen+Shot+2013-09-12+at+12.27.16+PM.png]

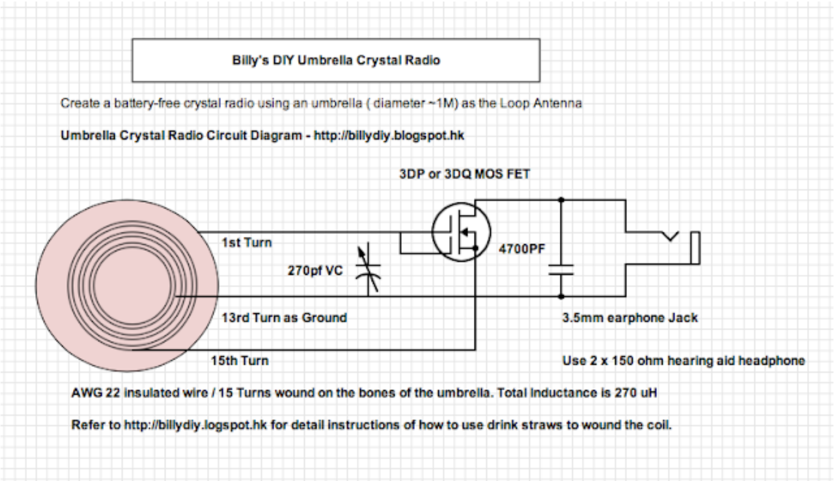
Billy's DIY Umbrella Crystal Radio
Create a battery-free crystal radio using an umbrella (diameter ~1M) as the Loop Antenna
Umbrella Crystal Radio Circuit Diagram - http://billydiy.blogspot.hk
3DP or 3DQ MOS FET
1st Turn
270pf VC
4700PF
13rd Turn as Ground
3.5mm earphone Jack
15th Turn
Use 2 x 150 ohm hearing aid headphone
AWG 22 insulated wire / 15 Turns wound on the bones of the umbrella. Total inductance is 270 uH
Refer to http://billydiy.logspot.hk for detail instructions of how to use drink straws to wound the coil.

[http://4.bp.blogspot.com/-ifl5C9xQo1o/Ujm4jtDTcyI/AAAAAAAABdg/FYmuaggicmU/s1600/Screen+Shot+2013-09-12+at+10.33.52+PM.png]

Posted 18th September 2013 by Billy

0 Add a comment

SEIZE THE SEASON
BOOK BY NOV 15 & SAVE UP TO 35%
BOOK NOW
WHISTLER BLACKCOMB

1st September 2013

Why Radio Reception is better at night?

Why Radio Reception is better at night?

In short at night the medium wave, short wave signals can transmit further because the ionospheric layer that bounces these medium wave and short wave signals all over the world gets thicker at night. In the morning when the sun is up, this ionoshperic layer gets thinner and the reflection of radio wave will not be as effective. Due to this effect, radio broadcasting station will lower their transmission power a bit to avoid broadcasting to another country.

According to international regulations, a portion of the radio frequency range of electromagnetic waves is reserved for radio broadcasts.
This is further divided into :

- LW or long wave (frequency range of 150 to 405 kHz) ,
- MW or medium-wave (525 ~ 1605 kHz) ,
- SW or shortwave (2.3 to 26.1 MHz, including 12 m section , see "SW Shortwave Frequency) ,
- FM (band divided from country to country , varies by region) .

SW Shortwave Frequency

UHF	Frequency (MHz)	UHF	Frequency (MHz)
11	25.600-26.100	41	7.100-7.300
13	21.450-27.750	49	5.950-6.200
16	17.700-17.900	60	4.750-5.060
19	15.100-15.450	75	3.900-4.000
25	11.700-11.975	90	3.200-3.400
31	9.500 - 9.775	120	2.300-2.495

Longwave lowest frequency depends mainly on the ground wave transmission, propagation distance, receiver performance is very stable throughout the year . Used mainly in European countries for domestic broadcasting. China has no long-wave radio stations.

Mediumwave is also propagated by terrestrial transmission , similar to that of longwave similar, but influenced by the geographical environment - better over a level surface is better, e.g. fields, plains, sea. worse over the forest, mountains.

transmission. This will also mean that more radio stations from all over the world (local at home, and from the overseas) can be received and jammed into your radio, interfering one another.

SW - shortwave is transmitted mainly by ionospheric effects by multiple reflections spread to thousands of kilometers away , even halfway around the world. This is how the international shortwave radio stations are broadcasted. The ground wave transmission distance for short wave (SW) is limited.
As the short wave gets to your radio through several reflections, the signal strength will be unstable and the volume will fluctuate from time to time. So the high-end shortwave radios have automatic gain control circuit (AGC) to improve the stability of radio reception.
Shortwave reception performances changes with time , season and weather conditions. Average winter sunny day is best, while raindy days in summer nights are the worst.
In addition, your location may be within the quiet zones , namely the area between two reflections. The quiet zone varies by time and frequency. So some radio stations will broadcast using multiple SW frequencies simultaneously and adjust with the change of season.

FM Radio FM mode is generally used for high-fidelity stereo broadcast , the division of the frequency range is complex because it need to jam into the limited spectrum left over from the television broadcasting channels.
FM frequency range is different per country: E.g. Russia from 65.8 to 73 MHz, Japan, from 76 to 90 MHz, China, US and Europe 88 ~ 108 ~ 104 MHz or 88 MHz. FM signal and the television broadcast signal is propagated only on a straight line , so the FM program can only be received in the region locally. The general maximum distance does not exceed 100 km. The reception is very stable as the signal is broadcasted over a straight line to a limited geographical location all under good control. Because the propagation distance is small, each region may be from a broadcast network , the same band can be reused.

In summary, at night the medium wave, short wave signals can transmit further because the ionospheric layer that bounces these medium wave and short wave signals all over the world gets thicker at night. In the morning when the sun is up, this ionoshperic layer gets thinner and the reflection of radio wave will not be as effective. Due to this effect, radio broadcasting station will lower their transmission power a bit to avoid broadcasting to another country.

Posted 1st September 2013 by Billy

0 Add a comment

1st September 2013

TA7642+LM386 regeneration radio

TA7642 (aka MK484) is an IC for AM and SW radio. It has all the circuit you need to make a regeneration radio of Auto Gain Control. Only need to add the LM386 to amplify the sound to drive the speaker.

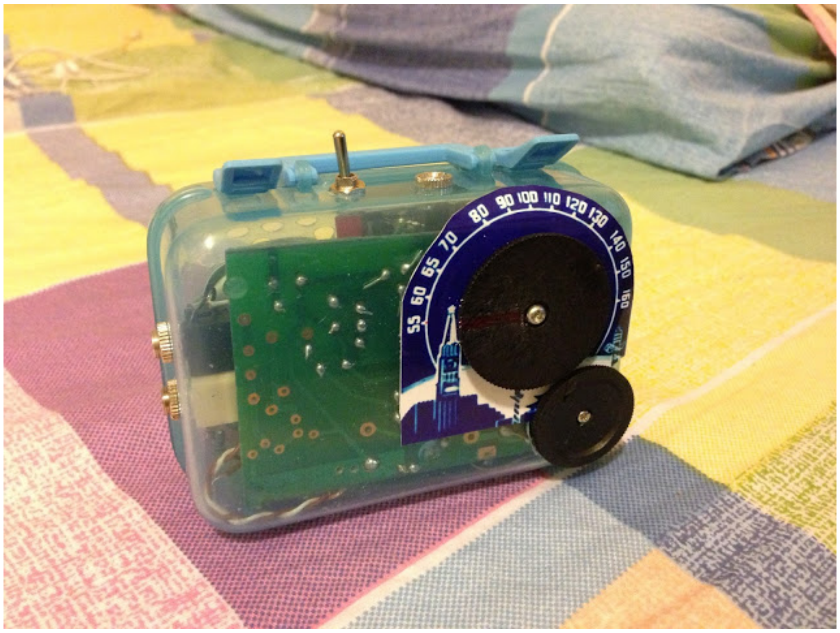

[http://2.bp.blogspot.com/-zGJ2RSo65fo/UiNN5M-CkzI/AAAAAAAABc4/H2uxz_ni0QQ/s1600/IMG_3457.JPG]

I have ordered the circuit board from the Taobao shop and changed the circuit a bit to double it up as an amplifier for my other crystal radios.
I also added a socket to insert an external radio coil, so I can connect my 1 meter diameter SW radio loop coil (made of

9mm copper tube, see last photo below), and change it to an Short Wave Radio.

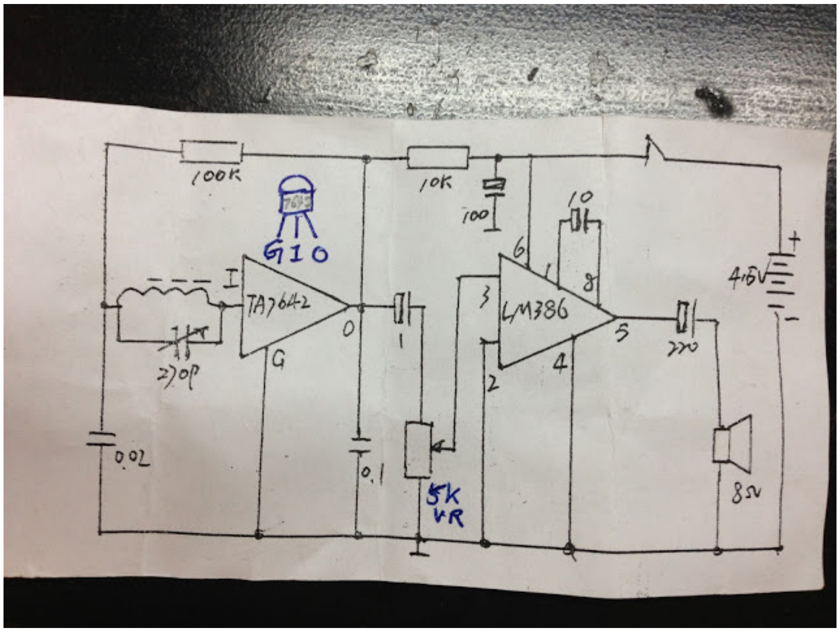
100K
10K
100
10
G10
I
TA7642
O
G
270P
0.02
0.1
1
5K
VR
6
1
8
3
2
LM386
4
5
270
4.5V
+
-
8Ω

[http://2.bp.blogspot.com/-t8UIUSu63tg/UiNNLdlAvfI/AAAAAAAABaw/gnHYsE6Pw5Y/s1600/Circuit+Diagram.JPG]

[http://1.bp.blogspot.com/-HiH16WPULzQ/UiNNLEPVWnI/AAAAAAAABa0/3REJl_8LIo4/s1600/IMG_3438.JPG]

[http://4.bp.blogspot.com/-jGQ-cgeq0nE/UiNNLBKGFAI/AAAAAAAABas/2GE4bPygBlI/s1600/IMG_3439.JPG]

[http://2.bp.blogspot.com/-602mLmOjsBM/UiNNL5qtcmI/AAAAAAAABbE/GA8YbpigPZc/s1600/IMG_3440.JPG]

[http://3.bp.blogspot.com/-RQe0vbP0okc/UiNNMNANmJI/AAAAAAAABbI/nX0774VRwRM/s1600/IMG_3441.JPG]

[http://3.bp.blogspot.com/-JekFB1Po5mE/UiNNMSCEkGI/AAAAAAAABbM/JXbFjFNIjBs/s1600/IMG_3442.JPG]

Dynamic Views theme. Powered by Blogger.

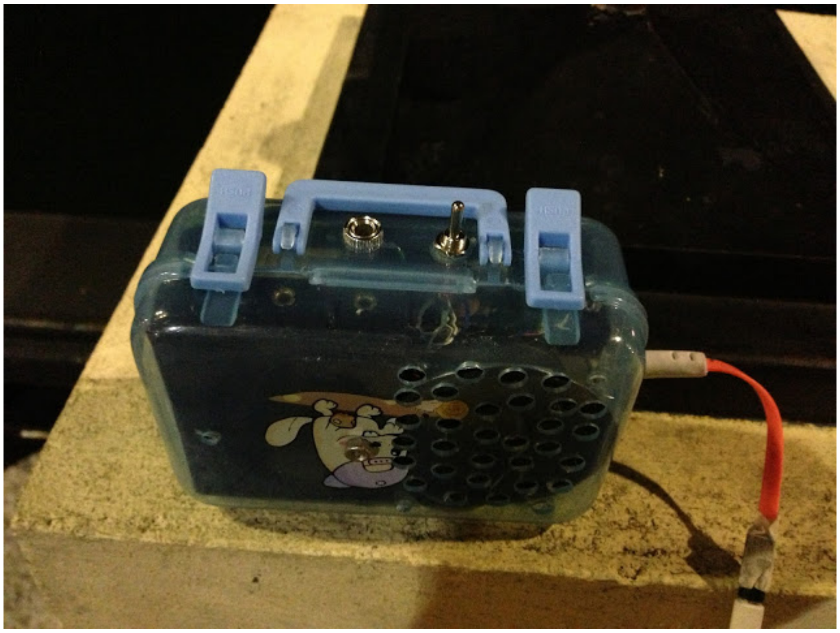

[http://1.bp.blogspot.com/-PIuEVX5ztBw/UiNNM4Se9rI/AAAAAAAABbc/lk_KjXh4euo/s1600/IMG_3444.JPG]

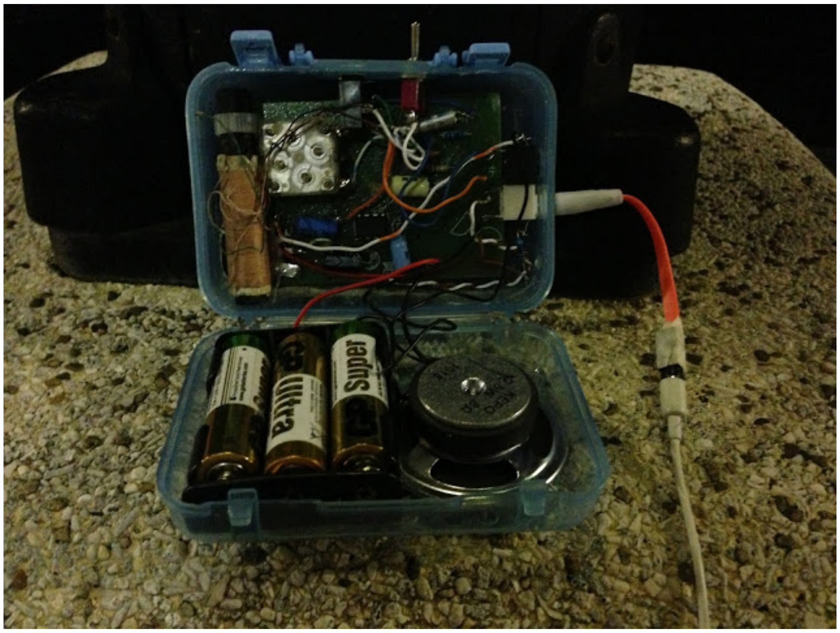

[http://2.bp.blogspot.com/-jUty3_NAOhA/UiNNNL4DzSI/AAAAAAAABbg/Ih0Rxir_OR8/s1600/IMG_3446.JPG]

Dynamic Views theme. Powered by Blogger.

[http://2.bp.blogspot.com/-EOG1lYN0T0U/UiNNNb5UKAI/AAAAAAAABbo/KvF2HPTzc7g/s1600/IMG_3447.JPG]

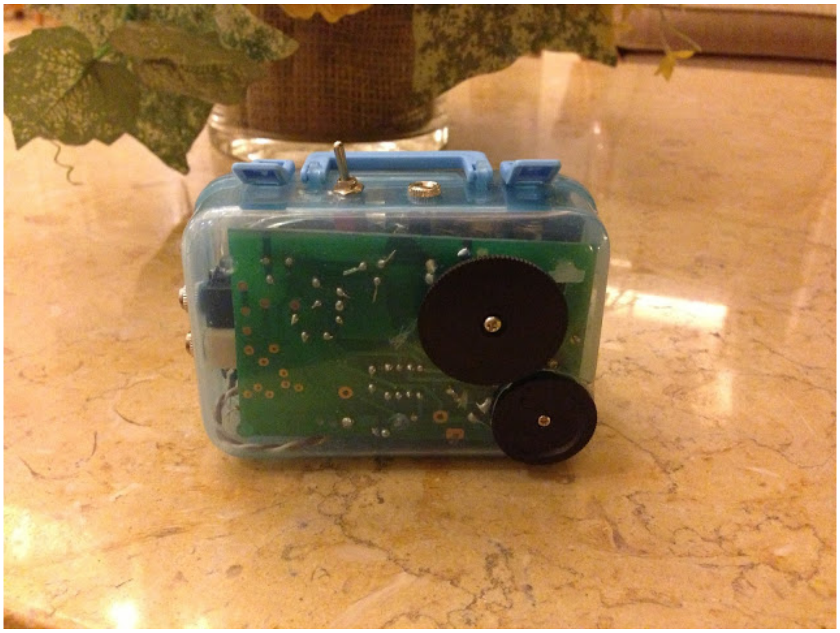

[http://2.bp.blogspot.com/-okPcbnasppY/UiNNN55lRSI/AAAAAAAABb0/g3gk_b5FzvU/s1600/IMG_3448.JPG]

Dynamic Views theme. Powered by Blogger.

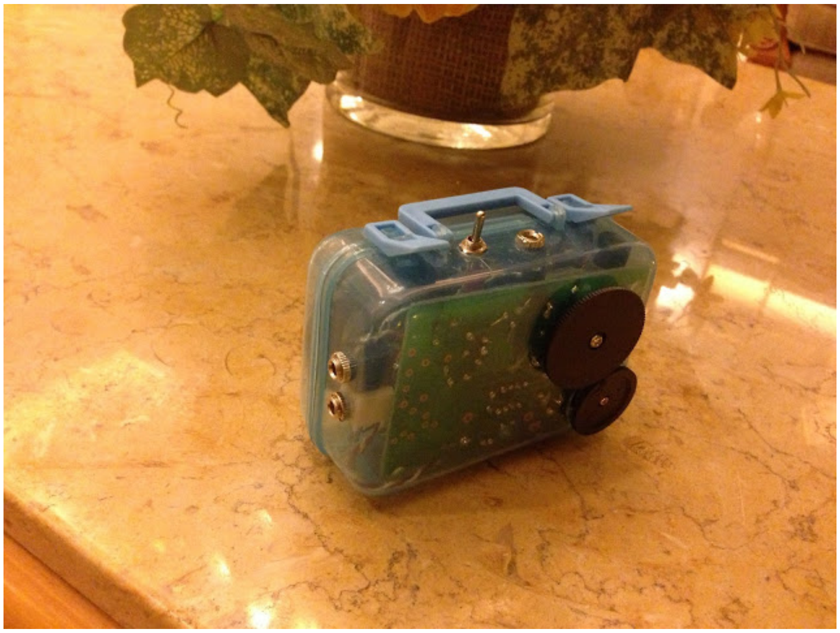

[http://3.bp.blogspot.com/-UvZhMOpAjbk/UiNNOKHZrRI/AAAAAAAABb4/3DbBPSU22xg/s1600/IMG_3449.JPG]

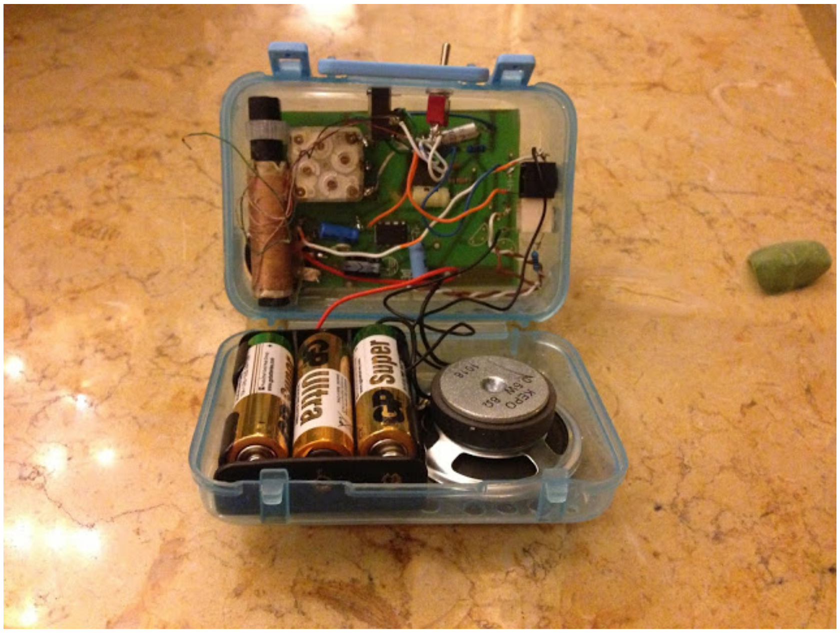

[http://3.bp.blogspot.com/-sf4ztgml1hg/UiNNOaNqKLI/AAAAAAAABcA/zUBk_C2DUKA/s1600/IMG_3450.JPG]

[http://3.bp.blogspot.com/-eyEuQuVNBtc/UiNNO1VDIFI/AAAAAAAABcM/CKz_22MfA_w/s1600/IMG_3451.JPG]

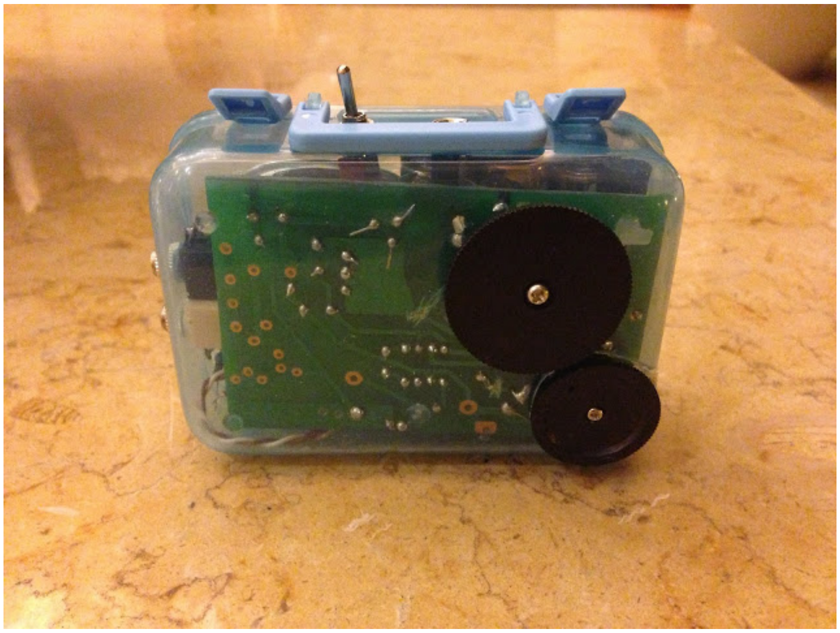

[http://2.bp.blogspot.com/-mt4VTT2gxm4/UiNNPM5M6UI/AAAAAAAABcQ/4K5tQSmP_j8/s1600/IMG_3452.JPG]

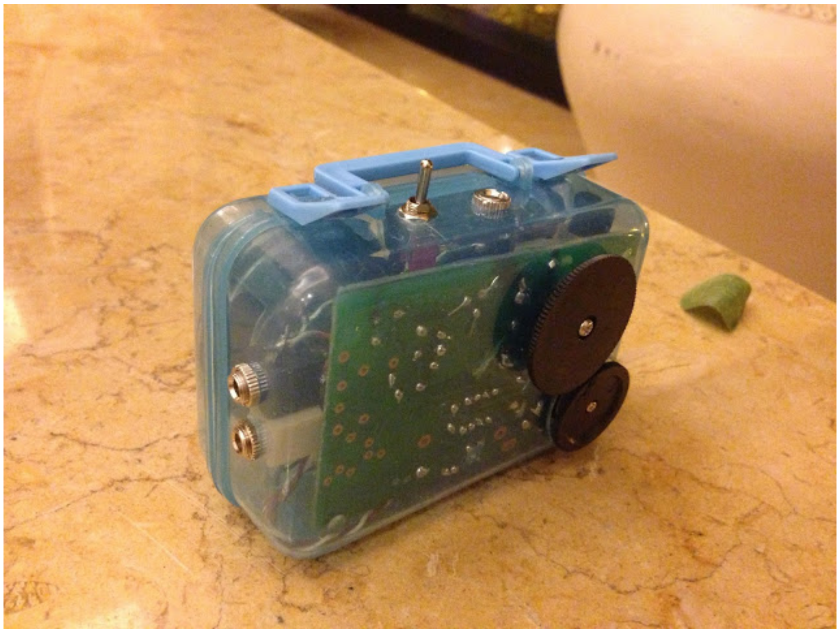

[http://2.bp.blogspot.com/-See7OI2opaQ/UiNNPa1ForI/AAAAAAAABcY/10IzkyIJDQ0/s1600/IMG_3454.JPG]

[http://4.bp.blogspot.com/-kLkkVVbfH2A/UiNNP5yu7YI/AAAAAAAABck/Qi-w2PCOK8U/s1600/IMG_3455.JPG]

[http://1.bp.blogspot.com/-8njHZaQqhhU/UiNNQHAJiyI/AAAAAAAABco/x8rW1VptH8Q/s1600/Short+Wave+Big+Loop.png]

Posted 1st September 2013 by Billy

3 View comments

29th August 2013

Antenna-Free Powder-Core-Coil Crystal Radio

You can now listen to an antenna free crystal radio on the street without carrying a big loop radio around.

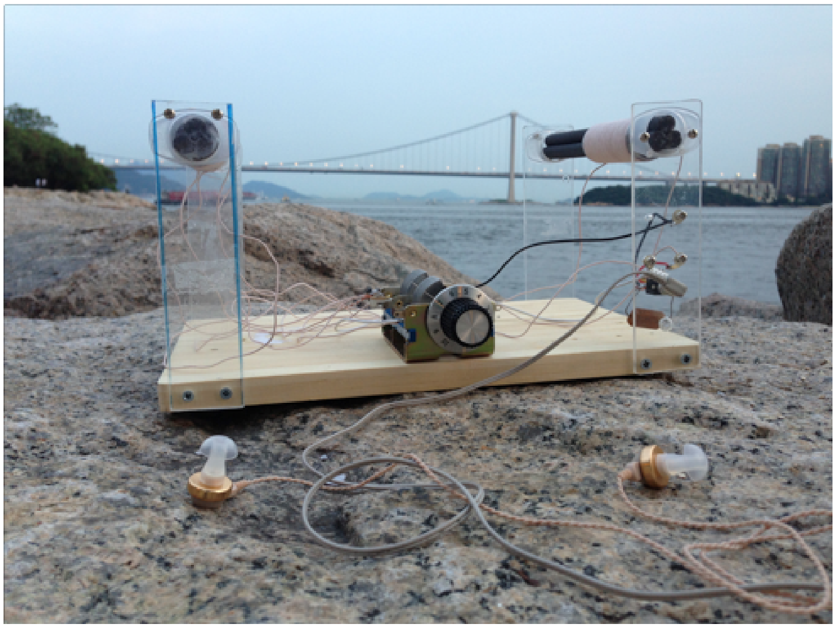

[http://4.bp.blogspot.com/-Dw4S-jKjN3s/UiAdAxqeEQI/AAAAAAAABYg/Ao6hPtYDh9E/s1600/Screen+Shot+2013-08-

Dynamic Views theme. Powered by Blogger.

[http://2.bp.blogspot.com/-D_R-juaEpdQ/UiAdGlH42NI/AAAAAAAABZE/pMNHKBBhy3I/s1600/Screen+Shot+2013-08-30+at+9.58.34+AM.png]

Checkout this youtube video for the testing of this radio [http://youtu.be/kRPyu5_Lh-8] :

Without worrying about sharing the same antenna, this one doubles up the radio wave energy using two separate powder core coils .
L1 to receive the original radio signal and turn that into audio thru the MOS FET (G1/G2 & D).
L3 to get a second copy of the radio signal, pump that thru L2 to energize the MOS FET (S).
All coils are wound on a 25mm diameter PVC tube (or a 30ml Syringe) using Liz wire (0.04mm x 270).
L2 & L3 shared the same powder core (Insert 3 powder rods into the PVC tube each of 8mm (D) and 180mm (L) .
L1 is on its own powder core same size as L2 & L3, but must be at least 300mm away from that of L2 & L3.
L1, L2 & L3 must be at least 300mm away from any other metallic components I(i.e. VC) to have the best effect.
L1 & L3 : 55 turns (60 turns if your variable capacitor is 270PF only), L2: 5 Turns.
L1 & L3 : 55 turns (60 turns if your variable capacitor is 270PF only), L2: 5 Turns.

Position both coil L1 and L3 to the middle of the Powder rod, and L3 immediately follows L3.. Then perform the following adjustments:

1. Tune to a radio station of the lowest frequency, e.g. 567khz, then adjust the position of L1,L2,L3 to get the best reception .
2. Tune to a radio station of the highest frequency, e.g. 1400khz, then adjust C3 to get the best reception.
3. Repeat 1 & 2 a few times until you get best reception for both the low end and high end stations

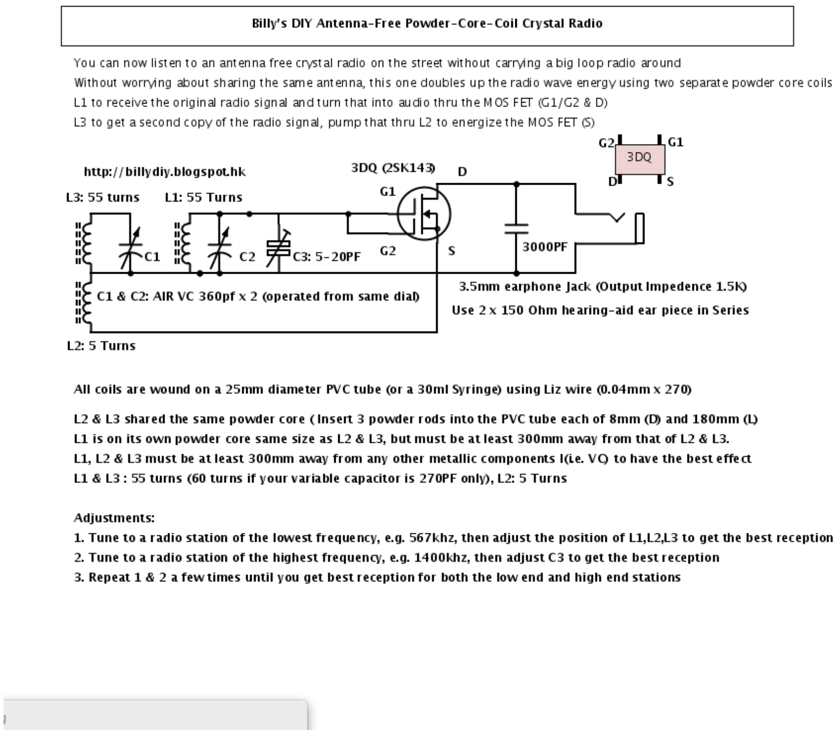

Billy's DIY Antenna-Free Powder-Core-Coil Crystal Radio

You can now listen to an antenna free crystal radio on the street without carrying a big loop radio around
Without worrying about sharing the same antenna, this one doubles up the radio wave energy using two separate powder core coils
L1 to receive the original radio signal and turn that into audio thru the MOS FET (G1/G2 & D)
L3 to get a second copy of the radio signal, pump that thru L2 to energize the MOS FET (S)

G2
3DQ
G1
D
S
http://billydiy.blogspot.hk
3DQ (2SK143)
D
G1
L3: 55 turns
L1: 55 Turns
C1
C2
C3: 5-20PF
G2
S
3000PF
C1 & C2: AIR VC 360pf x 2 (operated from same dial)
3.5mm earphone Jack (Output Impedence 1.5K)
Use 2 x 150 Ohm hearing-aid ear piece in Series
L2: 5 Turns

All coils are wound on a 25mm diameter PVC tube (or a 30ml Syringe) using Liz wire (0.04mm x 270)

L2 & L3 shared the same powder core (Insert 3 powder rods into the PVC tube each of 8mm (D) and 180mm (L)
L1 is on its own powder core same size as L2 & L3, but must be at least 300mm away from that of L2 & L3.
L1, L2 & L3 must be at least 300mm away from any other metallic components I(i.e. VC) to have the best effect
L1 & L3 : 55 turns (60 turns if your variable capacitor is 270PF only), L2: 5 Turns

Adjustments:
1. Tune to a radio station of the lowest frequency, e.g. 567khz, then adjust the position of L1,L2,L3 to get the best reception
2. Tune to a radio station of the highest frequency, e.g. 1400khz, then adjust C3 to get the best reception
3. Repeat 1 & 2 a few times until you get best reception for both the low end and high end stations

[http://4.bp.blogspot.com/-xEYEm8n8vts/UiAdF5FA4iI/AAAAAAAABY0/mhqgvabmlA4/s1600/Circuit+Diagram.png]

[http://1.bp.blogspot.com/-oMCZLraci6Q/UiAc_1bh_DI/AAAAAAAABYc/Ub0p-S01-4g/s1600/Screen+Shot+2013-08-29+at+10.18.27+PM.png]

[http://3.bp.blogspot.com/-mgzm3y-WuXM/UiAdACEf_vI/AAAAAAAABYM/kx6rd5UEH40/s1600/Screen+Shot+2013-08-29+at+10.18.37+PM.png]

Dynamic Views theme. Powered by Blogger.

[http://2.bp.blogspot.com/-rirmOxQbJJ4/UiAdAXYM_SI/AAAAAAAABYU/rmtPztgMeGo/s1600/Screen+Shot+2013-08-29+at+10.18.52+PM.png]

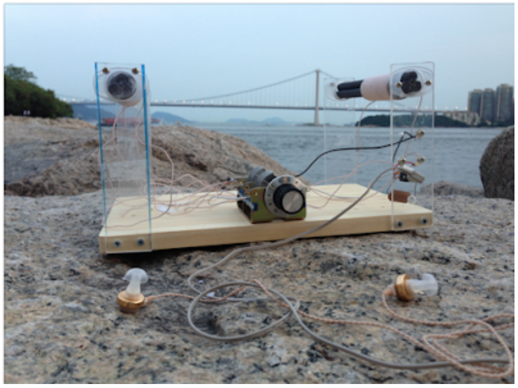

[http://4.bp.blogspot.com/-Dw4S-jKjN3s/UiAdAxqeEQI/AAAAAAAABYg/Ao6hPtYDh9E/s1600/Screen+Shot+2013-08-29+at+10.19.05+PM.png]

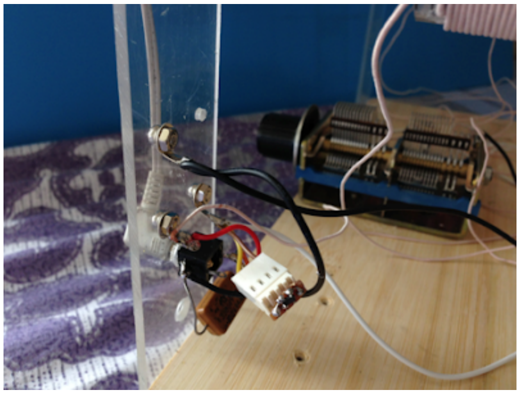

[http://1.bp.blogspot.com/-gdy26vmOzfk/UiAdF7JM2aI/AAAAAAAABYw/m4haMq8wNMM/s1600/Screen+Shot+2013-08-30+at+9.56.34+AM.png]

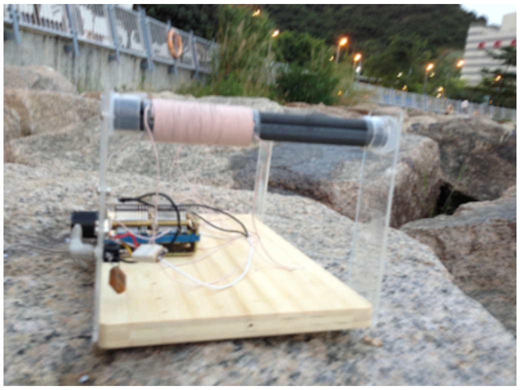

[http://4.bp.blogspot.com/-BVIFJr3yTyE/UiAdGQ3DMTI/AAAAAAAABZI/NbjncXnyWv8/s1600/Screen+Shot+2013-08-30+at+9.57.42+AM.png]

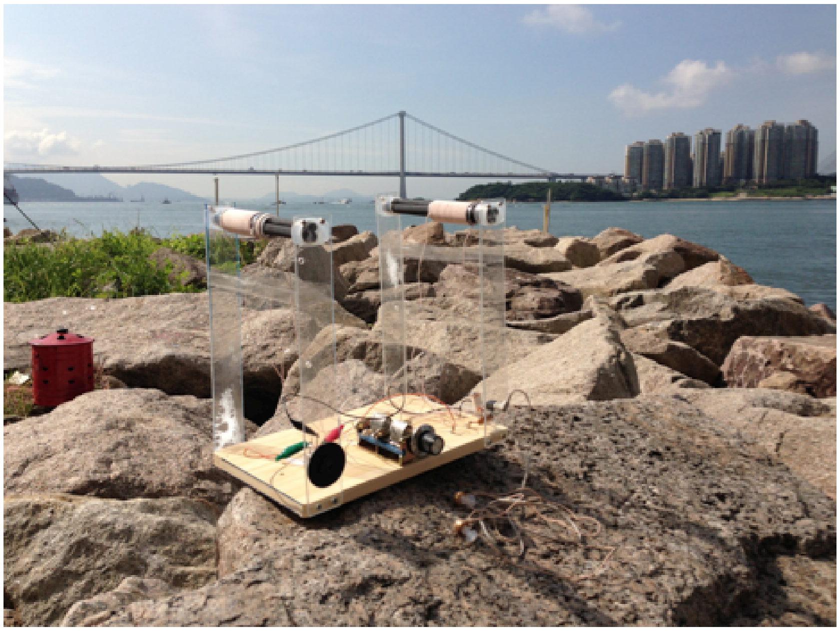

[http://1.bp.blogspot.com/-JPIP-Dj5klU/UiAdF8QMfzI/AAAAAAAABY4/QL0gQpbfuD4/s1600/Screen+Shot+2013-08-30+at+10.00.30+AM.png]

Posted 29th August 2013 by Billy

0 Add a comment

legalzoom
Let's Make Your Business Official!
Come back

12th August 2013

Portable Document Tray Crystal Radio

Portable Document Tray Crystal Radio

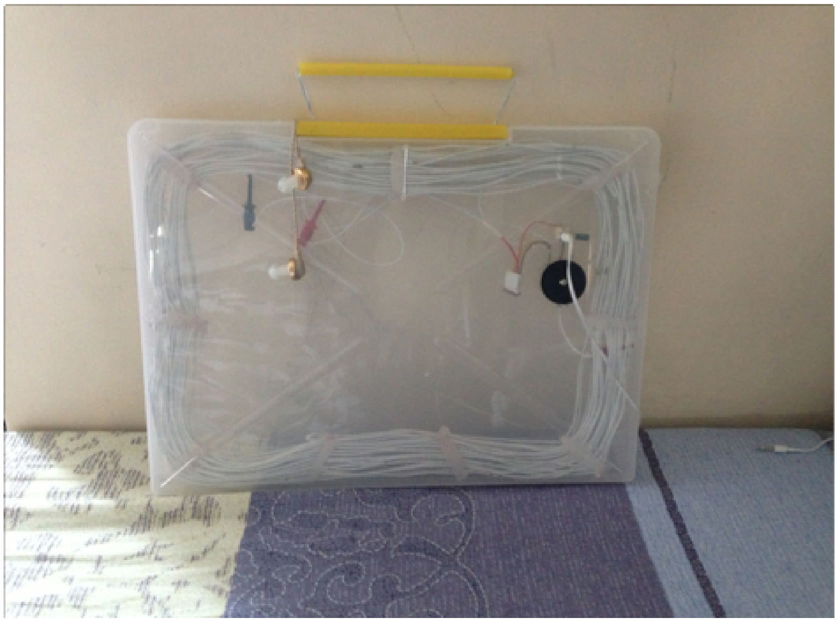

[http://1.bp.blogspot.com/-3mB9i-V3mkA/UambfizR cI/AAAAAAAABXg/nbDIcv9xeFY/s1600/Screen+Shot+2013-08-

13+at+10.03.46+AM.png]

Make a battery-free crystal radio using a filing box.
Place it next to the window to listen to radio stations without battery.
You can also put it inside your briefcase or backpack and listen to crystal radio while you are traveling.

How to make it:

Find a plastic filing box, the bigger the better. The one I got is B4 (380mmx280mm). It'll be even better if you can find an A3 filing box.
Then wound a radio coil inside following the instructions below:

Make a card board coil mod same size as the filing box. Cut 9 slots 40mm deep, 10 mm wide at the edge of the boards. The number of slots must odd, as we need the wires to go above and below to create a spider web effect. This will be better for the radio signal reception.

Get some regular electric wires size around SWG22. Wound 26 turns above and below the slots as shown in the video. How to wind a spiderweb radio coil. [http://www.youtube.com/watch?v=GHq7fVLr9OI]

Then fix the wire in place, by either ting each slot up with fishing wire or super glue or rubber gun.
Drill some holes into the back of the filing box so you can tie the radio coil up.

Drill more holes to hold the crystal radio components.

We will need a MOS FET called (2SK143, or 3DQ), a plastic variable capacitor that goes from 20pf to 270pf as you adjust the dial.
A 3.5mm headphone socket.
Two hearing aid headphones (300 Ohm) connected in series to a mono 3.5mm headphone jack.

This radio operates without antenna nor ground.
You can receive several local radio stations, and the sound is loud.

You can also turn it into a briefcase crystal radio by adding a handle.

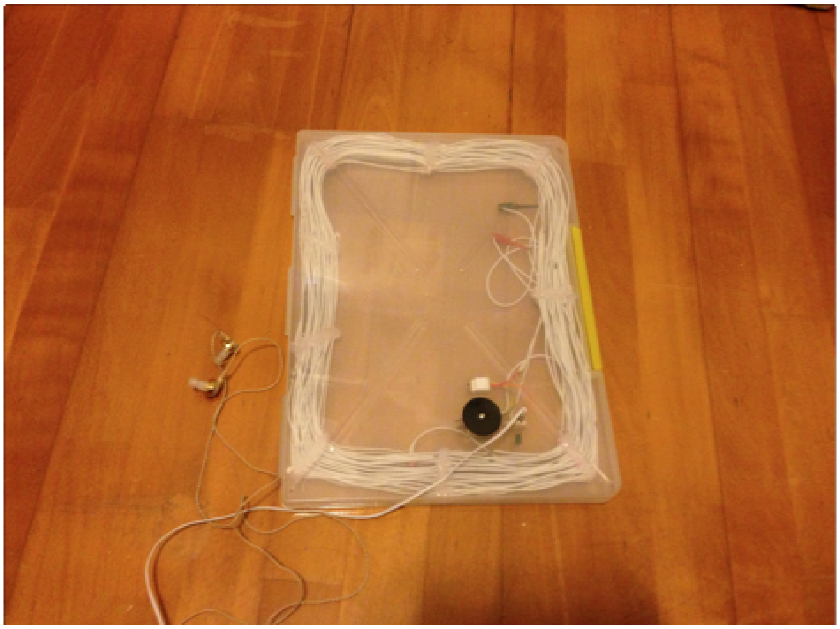

[http://2.bp.blogspot.com/-8pkwIc_Hc5A/UgmbeVa1F0I/AAAAAAAABXA/kG13iItOSQw/s1600/Screen+Shot+2013-08-12+at+11.29.10+PM.png]

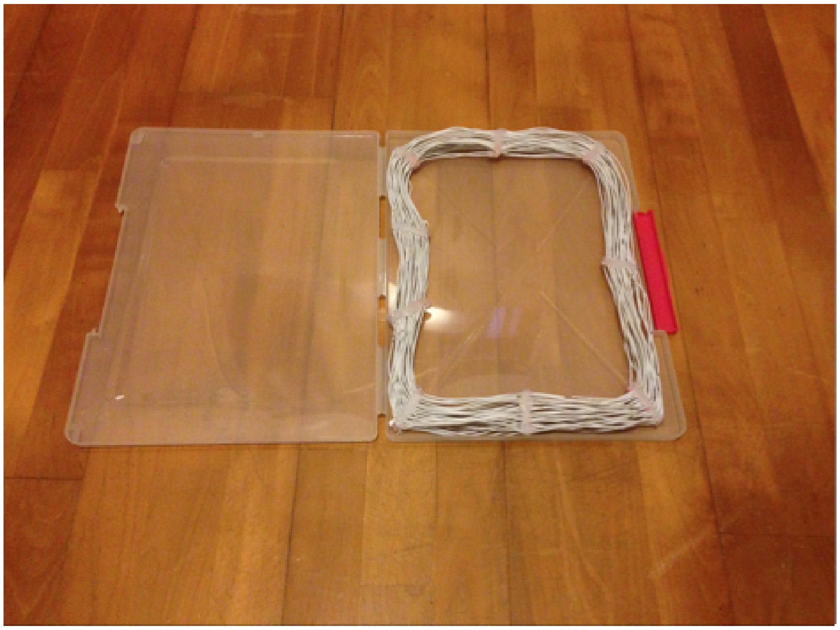

[http://4.bp.blogspot.com/-Supyzgrcofo/Ugmbdm2_UmI/AAAAAAAABWk/R6fqAOuQ9Hw/s1600/Screen+Shot+2013-08-12+at+11.28.18+PM.png]

[http://4.bp.blogspot.com/-Ud4TL89rGM8/UgmbdRG_NMI/AAAAAAAABWU/RcXFf2cdmRY/s1600/Screen+Shot+2013-08-12+at+11.28.29+PM.png]

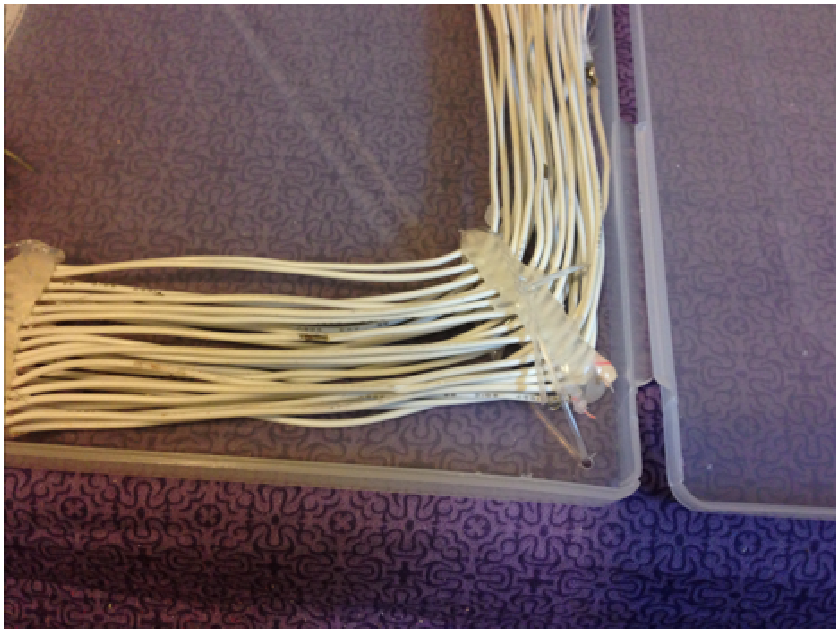

[http://4.bp.blogspot.com/-0WmYSuQIe0E/UgmbdmAypgI/AAAAAAAABWc/mTytUuMxhB4/s1600/Screen+Shot+2013-08-12+at+11.28.42+PM.png]

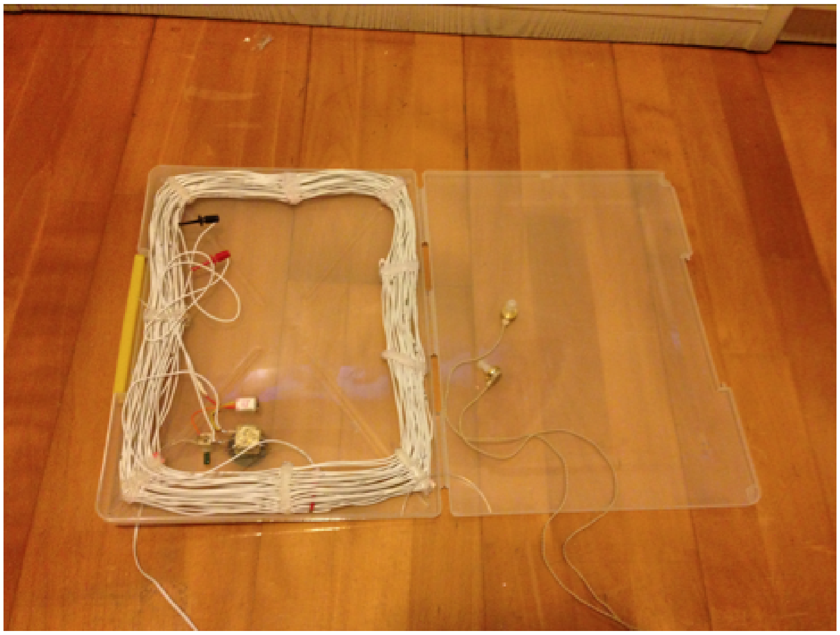

[http://2.bp.blogspot.com/-rcWdWSmMUho/UgmbeJuwUJI/AAAAAAAABWs/wicKYvmJZ-s/s1600/Screen+Shot+2013-08-12+at+11.28.55+PM.png]

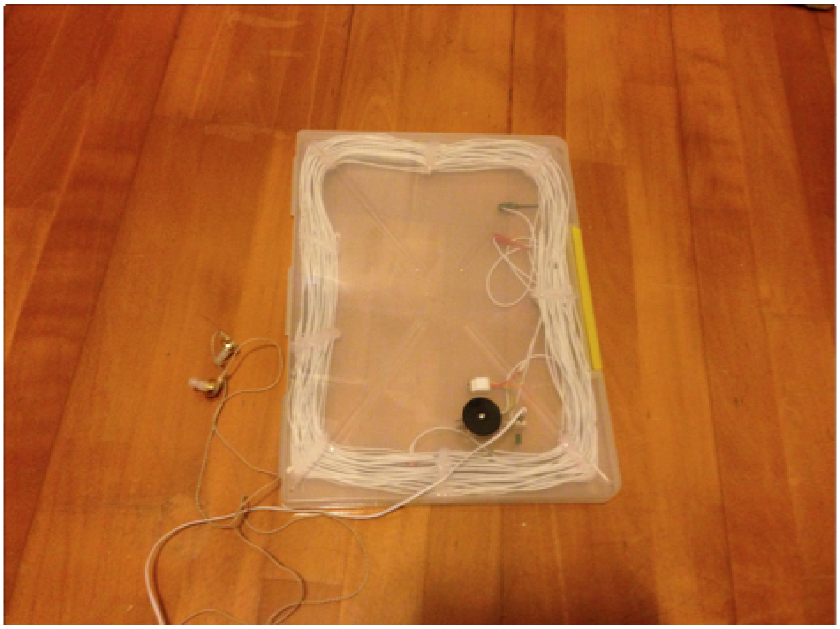

[http://2.bp.blogspot.com/-8pkwIc_Hc5A/UgmbeVa1F0I/AAAAAAAABXA/kG13iltOSQw/s1600/Screen+Shot+2013-08-12+at+11.29.10+PM.png]

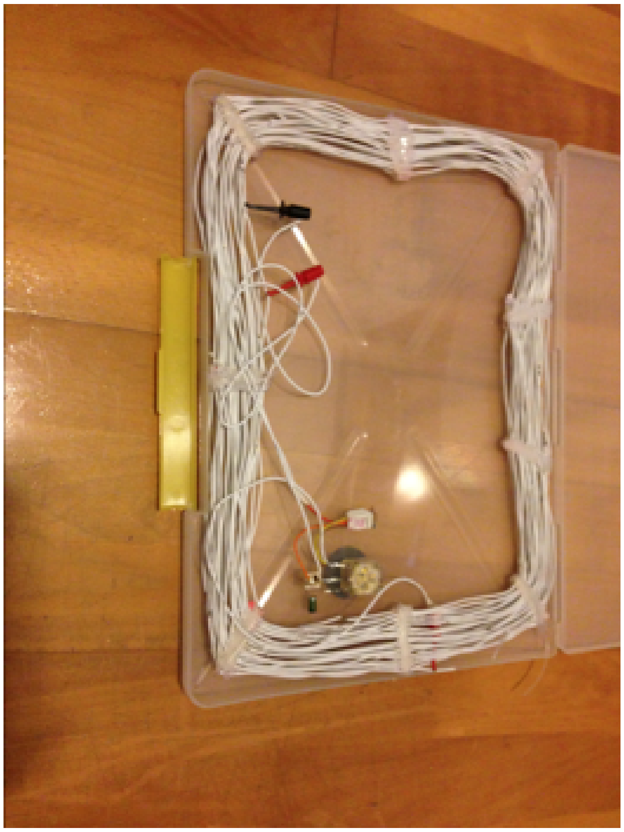

[http://3.bp.blogspot.com/-xUyXzlD7jxI/UgmbexcpLfI/AAAAAAAABW8/saC6l1FSZSE/s1600/Screen+Shot+2013-08-12+at+11.29.33+PM.png]

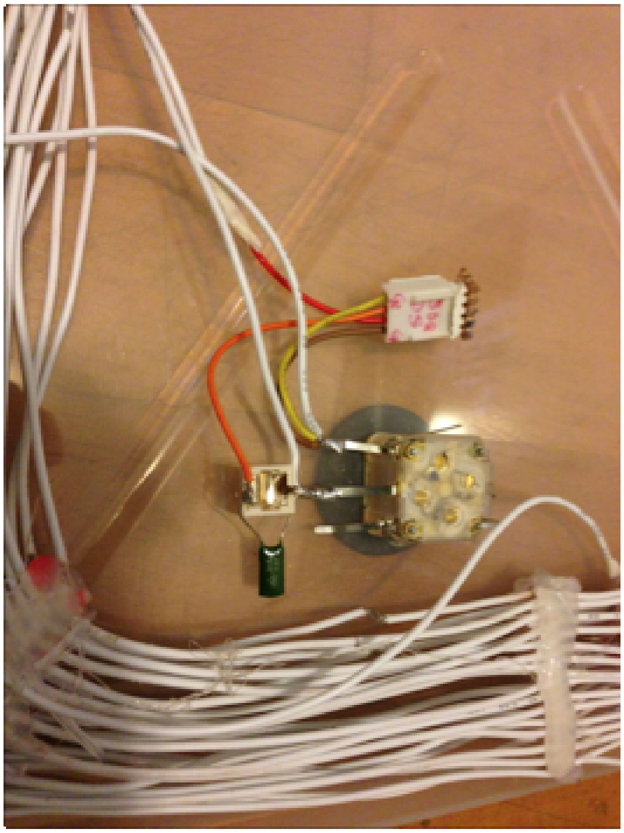

[http://3.bp.blogspot.com/-LMmrvn0OuPE/UgmbeqEmHsI/AAAAAAAABXE/F6XToIjIxm4/s1600/Screen+Shot+2013-08-12+at+11.29.21+PM.png]

[http://3.bp.blogspot.com/-bMhBesF_kig/UgmbfJj2_2I/AAAAAAAABXM/aN44QGzuiMY/s1600/Screen+Shot+2013-08-12+at+11.59.43+PM.png]

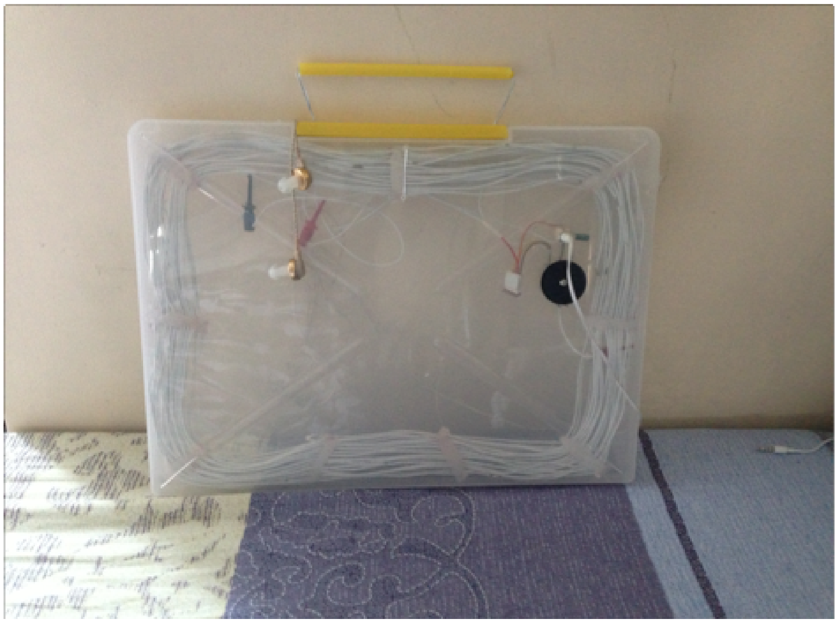

[http://1.bp.blogspot.com/-3mB9i-V3mkA/UgmbfizR_cI/AAAAAAAABXg/nbDIcv9xeFY/s1600/Screen+Shot+2013-08-13+at+10.03.46+AM.png]

[http://4.bp.blogspot.com/-4mWwU6JnMrs/Ugmbfttj-KI/AAAAAAAABXc/jbtkjdWONy0/s1600/Screen+Shot+2013-08-13+at+12.00.03+AM.png]

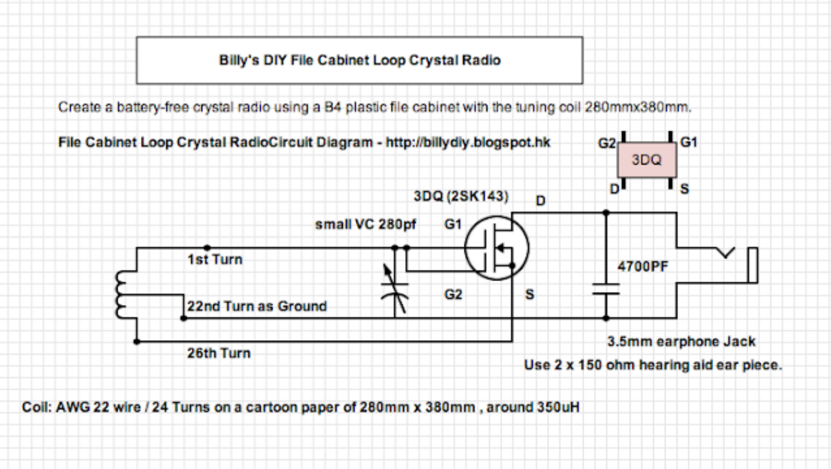

Billy's DIY File Cabinet Loop Crystal Radio

Create a battery-free crystal radio using a B4 plastic file cabinet with the tuning coil 280mmx380mm.

File Cabinet Loop Crystal RadioCircuit Diagram - http://billydiy.blogspot.hk

Use 2 x 150 ohm hearing aid ear piece.

Coil: AWG 22 wire / 24 Turns on a cartoon paper of 280mm x 380mm , around 350uH

[http://4.bp.blogspot.com/-x_jzJPm6wsI/Ugmbf93DevI/AAAAAAAABXU/_6x6q-MVsSE/s1600/Screen+Shot+2013-08-13+at+12.13.52+AM.png]

Posted 12th August 2013 by Billy

0 Add a comment

10th August 2013

You need a crystal earphone to listen to the battery free crystal radio.
I've been making the crystal earphone myself as it's hard to buy in Asia.
A friend of mine helped me buy a few crystal earphone from US (Ceramic Earphone US $ 3.5).
I compared the factory made crystal earphone (left0 with the ones I made myself (right).

The result: the homemade crystal earphone won!

The factory made one is not as loud as the homemade one, it's about 30% softer. The homemade one has better audio quality, with a much richer base than the factory made one.

If you also want to own a look at my other posts:

http://billydiy.blogspot.hk/2013/03/build-crystal-earphone-for-crystal.html [http://billydiy.blogspot.hk/2013/03/build-crystal-earphone-for-crystal.html]

http://www.crystalradio.cn/thread-390154-1-1.html [http://www.crystalradio.cn/thread-390154-1-1.html]

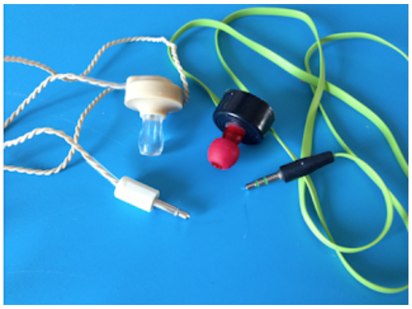

[http://1.bp.blogspot.com/-42a2qLr9T9w/UgYMV1mCtyI/AAAAAAAABVg/Hqvq0bRZluo/s1600/Screen+Shot+2013-08-10+at+5.06.52+PM.png]

[http://3.bp.blogspot.com/-BgxVfobVasQ/UgYMan148tI/AAAAAAAABV8/zPU4X62cUCk/s1600/Screen+Shot+2013-08-10+at+5.25.32+PM.png]

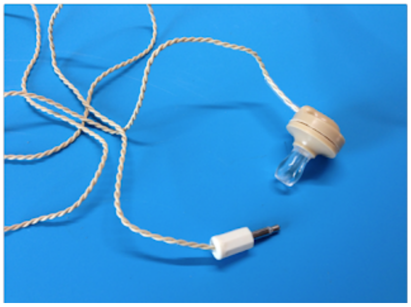

[http://2.bp.blogspot.com/-kc2cF1qgVGE/UgYMV0zD5bI/AAAAAAAABVc/oGTixy451vQ/s1600/Screen+Shot+2013-08-10+at+5.07.10+PM.png]

[http://4.bp.blogspot.com/-pblFLoDIp0Q/UgYMVxHSOiI/AAAAAAAABVk/oX16KbsAFHA/s1600/Screen+Shot+2013-08-10+at+5.07.55+PM.png]

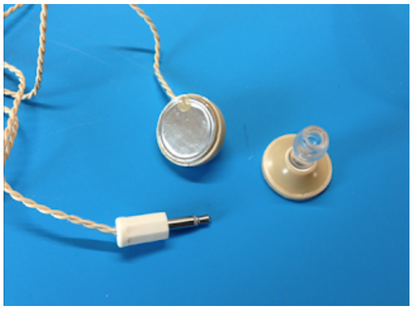

[http://1.bp.blogspot.com/-OqF1CwFnMgs/UgYMWZYy9xI/AAAAAAAABV0/JjWpaip9I1k/s1600/Screen+Shot+2013-08-10+at+5.08.08+PM.png]

Posted 10th August 2013 by Billy

0 Add a comment

29th July 2013

How Crystal Radio Works & useful links

How Crystal radio work?

Text explanation

Dynamic Views theme. Powered by Blogger.

http://www.crystalradio.net/howitworks/index.shtml [http://www.crystalradio.net/howitworks/index.shtml]

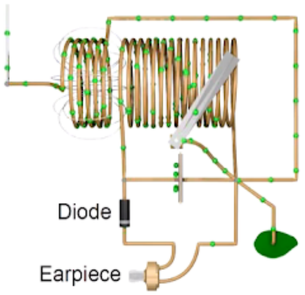

[http://1.bp.blogspot.com/-1S4kzIMb2b8/UfZuLaK2wYI/AAAAAAAABVE/gXJNMl5PVoU/s1600/Screen+Shot+2013-07-29+at+9.28.45+PM.png]

Video explanation
http://www.youtube.com/watch?v=0-PParSmwtE [http://www.youtube.com/watch?v=0-PParSmwtE]

Some useful links

1. How to make a Foxhole radio [http://www.youtube.com/watch?v=vU8zaszyTYA] - The simplest crystal radio

[http://4.bp.blogspot.com/-jAm9LD1caE8/UfZvTnvJbhI/AAAAAAAABVM/pzoHg-Ish1g/s1600/Screen+Shot+2013-07-29+at+9.33.31+PM.png]

2. Crystal Radio Forum in China - www.crystalradio.cn [http://www.crystalradio.cn/forum-3-1.html]

3. Crystal Radio Forum in US - www.crystalradio.net [http://www.crystalradio.net/]

4. http://www.radio-world.com.cn/forum.php?mod=forumdisplay&fid=5 [http://www.radio-world.com.cn/forum.php?mod=forumdisplay&fid=5]

5. 中国电子DIY之家- 收音机DIY [http://www.ndiy.cn/forum-107-1.html]

6. http://www.haodiy.net [http://www.haodiy.net/a/jishuwenzhang/shouyinjiwuxiandianDIY/shouyinji/]

7. Online circuit diagram drawing tool [http://www.digikey.com/schemeit]

8. My Chinese Blog - 小張夢工場 [http://billydiydreamshop.lofter.com/]

Posted 29th July 2013 by Billy

0 Add a comment

legalzoom
Let's Make Your Business Official!
Come back

gift box AM FM solar powered regeneration crystal radio

I designed a small gift box AM FM solar powered regeneration crystal radio.All wiring done with silver-plated wire to allow a variety of functions and AM FM coils to be placed in a 120mmx90mmx40mm small box.Appears to be an ordinary wooden box, yet opens into a crystal radio.

This cute gift box crystal radio can be a gift for the children. Promote interest in crystal radio. The machine can operate with or without power.

AM circuit includes 100mm ferrite rod with a coil wound using 0.05x120 lz wire.

http://www.youtube.com/watch?v=i72Ir_EHQ6Q [http://www.youtube.com/watch?v=i72Ir_EHQ6Q]

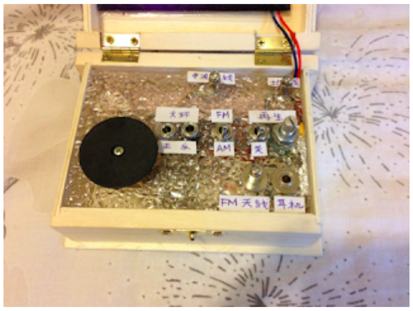

[http://1.bp.blogspot.com/-8z-oxBzXhOU/UefddaoTF6I/AAAAAAAABSk/UeyFgyCkqSI/s1600/Screen+Shot+2013-07-16+at+12.06.18+AM.png]

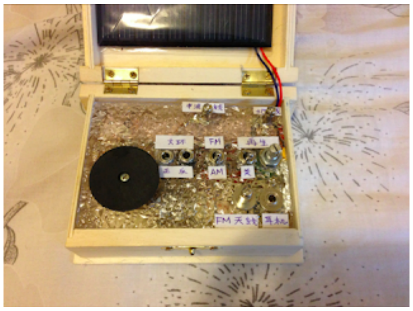

[http://2.bp.blogspot.com/-LVs1pfqai6Y/UefddaTWycI/AAAAAAAABSo/iWLF5HAsQLY/s1600/Screen+Shot+2013-07-16+at+12.15.20+AM.png]

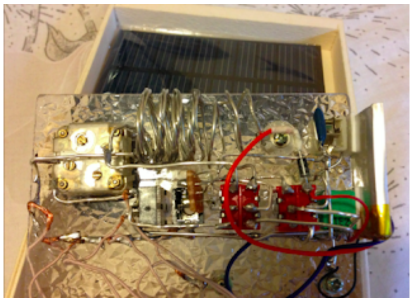

[http://2.bp.blogspot.com/-0zAx5uwZk44/UefddcCzXLI/AAAAAAAABSs/dpNz0L6My8w/s1600/Screen+Shot+2013-07-16+at+12.15.46+AM.png]

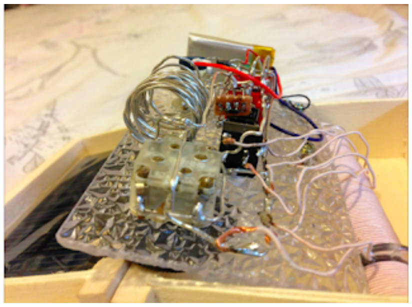

[http://3.bp.blogspot.com/-NTm3KsqTde4/UefdeLAy8II/AAAAAAAABTE/XsW1tFsrgvM/s1600/Screen+Shot+2013-07-16+at+12.16.03+AM.png]

[http://4.bp.blogspot.com/-Vx697Z5hkSw/UefdeRtSY_I/AAAAAAAABS8/i-2pUmWrpWw/s1600/Screen+Shot+2013-07-16+at+12.16.24+AM.png]

Dynamic Views theme. Powered by Blogger.

[http://4.bp.blogspot.com/-pffGllEktlk/UefdeZCU-ol/AAAAAAAABTA/Mv0UK5Se9_4/s1600/Screen+Shot+2013-07-16+at+12.16.37+AM.png]

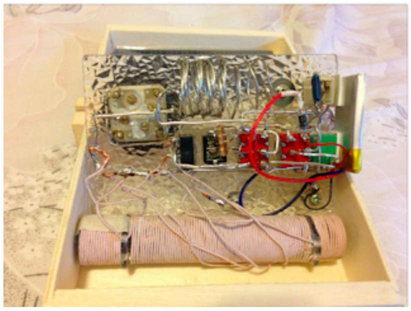

[http://1.bp.blogspot.com/-cg9stNWxMVA/UefdfAGAezI/AAAAAAAABTU/6GuQVPGJdS4/s1600/Screen+Shot+2013-07-16+at+12.16.50+AM.png]

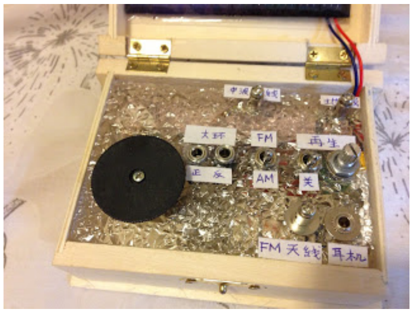

[http://3.bp.blogspot.com/-ixhSddvtB_Q/UefdqpGoMCI/AAAAAAAABTo/rxijs3ncKpM/s1600/IMG_2831.JPG]

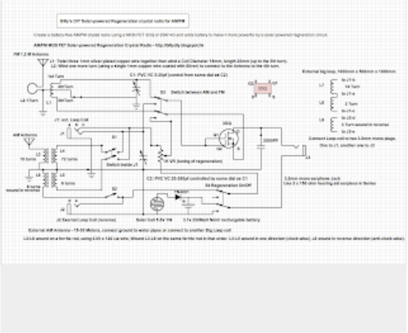

[http://1.bp.blogspot.com/-4JFt5tosTsM/Uefe6zPZUKI/AAAAAAAABUc/s8wYW8XBeXs/s1600/Screen+Shot+2013-07-15+at+11.27.32+PM.png]

Dynamic Views theme. Powered by Blogger.

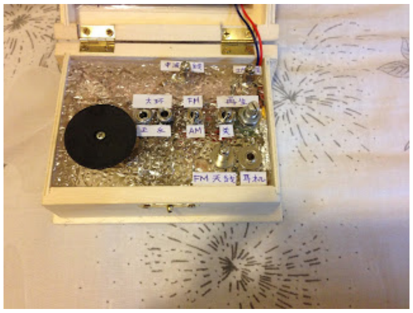

[http://4.bp.blogspot.com/-bjUbOfyQk3U/UefdsBscurl/AAAAAAAABT4/iIwgoctHxk4/s1600/IMG_2834.JPG]

Posted 17th July 2013 by Billy

0 Add a comment

6th July 2013

What is AM and FM ?

What is AM and FM ? Are AM and MW the same thing ?

Audible Frequency

standard range of audible frequencies [https://en.wikipedia.org/wiki/Hearing_range] for humans is 20 to 20,000 Hz,
However, the audible frequency ,e.g. the music , is not efficient to be broadcasted over radio wave nor can travel very far.
To solve the problem, we will need something to carrier the music. This something is called the Carrier wave.
Then we need to use a method to put the music onto the carrier wave and produce the ultimate radio wave that can be broadcasted. This method is called modulation.

Common layman terms that are mostly right , but not always right.

AM and MW usually means the same thing in layman terms. It means the broadcast radio transmitted at the frequency within the medium wave (MW) frequency range (or Band) -(540Khz to 1650khz) and modulated using AM (Amplitude Modulation).

MW is a subset of a standard term called MF - Medium Frequency) that describe frequency range of 300Khz to 3000Khz. See table below).

Amplitude modulation means the carrier frequency is fixed all the time, but the amplitude of the carrier wave will vary according to the the wave of the music being carried. i.e. The amplitude is modulated by the music.

FM and ultra short wave usually means the same thing in layman terms. It means the broadcast radio transmitted at the frequency within the ultra short wave frequency range - 88Mhz to 108Mhz) and modulated using FM (Frequency Modulation).

Ultra short wave is a subset of a standard term called VHF - Very High Frequency) that describe frequency range of 30 Mhz to 300Mhz. See table below).

See table below.

Frequency modulation means the amplitude of the carrier wave is fixed all the time, but the frequency of the carrier wave will vary back and forth by a factor from the central frequency of the radio station you tuned into . This variation of frequency is according to the the wave of the music being carried. I.e. The frequency is modulated by the music.

Aviation communication uses AM but at the frequency range just above the FM radio ?

Pilots communicates with the airports at the Vivil aviation band - VHF 108Mhz to 136 MHz.
However, the voice of the pilots are put on the carrier wave using Amplitude Modulation (AM).
This is a good example to show that FM radio band, will sometimes be broadcasted using AM.

Types of modulation

AM

is the name given to the method to put an audio signal (voice/music) onto the radio wave, called
Amplitude Modulation - fixed frequency with the amplitude of the wave shifting according to the volume of the music.

1. A standard radio wave of a fixed frequency (of that of the radio station) is first broadcasted. If no one music is played, the standard radio wave will have same amplitude (volume) all along, and you will hear silence from the radio receiver.
When the music starts playing and feed to the transmitter, then depending on the volume of the audio signal (voice/music), the radio wave will go up or down in amplitude (or volume).

This ups and downs in amplitude (volume is then translated by your radio receiver into audible sound coming form your headphone.

FM

is the name given to the method to put an audio signal (voice/music) onto the radio wave, called
Frequency Modulation - fixed amplitude with the frequency of the wave shifting according to the volume of the music.

Like AM, 1. a standard radio wave of a base frequency (of that of the radio station) is first broadcasted. If no one music is played, the standard radio wave will have same frequency all along, and you will hear silence from the radio receiver.
When the music starts playing and feed to the transmitter, then depending on the volume of the audio signal (voice/music), the radio wave will shifts left and right in frequency within the frequency spectrum allocated to that radio station. E.g. RTHK2 may be broadcasting at 94.8Mhz +/-0.05Mhz. So, if music volume is low, it'll shift to 94.75Mhz, if music volume is high, it'll shift to 94.85Mhz.

What is LW, MW, SW1, SW2, Ultra SW ?

LW, MW, SW and FM are all descriptions of the frequency range used in radio broadcasting.

LW is short for Long Wave with a frequency range from 100Khz to 500Khz. It is a subset of the standard wireless terms we called LF (low frequency).
MW is short for Medium Wave with a frequency range from 0.5Mhz (500Khz) to 1.7 Mhz (1700Khz). i.e. MF (medium frequency)

SW is for Short wave, i.e. HF (high frequency)
SW1 is short for Short Wave One with a frequency range from 5Mhz to 10 Mhz.
SW2 is short for Short Wave Two with a frequency range from 11Mhz to 19 Mhz.

Ultra Shortwave is the range where FM radios are broadcasted. 70Mhz to 120Mhz.
i.e. VHF (very high frequency).

Band name	A	I	F	Example Uses

	breviation	U band number	equency and Wavelength	
Extremely low frequency [https://en.wikipedia.org/wiki/Extremely_low_frequency]	ELF	1	3–30 Hz 100,000–10,000 km	Communication with submarines [https://en.wikipedia.org/wiki/Communication_with_submarines]
Super low frequency [https://en.wikipedia.org/wiki/Super_low_frequency]	SLF	2	30–300 Hz 10,000–1,000 km	Communication with submarines
Ultra low frequency [https://en.wikipedia.org/wiki/Ultra_low_frequency]	ULF	3	300–3,000	Submarine communication, communication within mines [https://en.wikipedia.org/wiki/Through_the_earth_mine_communications]

			Hz 1,000–100 km	
Very low frequency [https://en.wikipedia.org/wiki/Very_low_frequency]	VLF	4	3–30 kHz 100–10 km	Navigation [https://en.wikipedia.org/wiki/Radio_navigation] , time signals [https://en.wikipedia.org/wiki/Radio_clock] , submarine communication, wireless heart rate monitors [https://en.wikipedia.org/wiki/Heart_rate_monitor] , geophysics [https://en.wikipedia.org/wiki/Geophysics]
Low frequency [https://en.wikipedia.org/wiki/Low_frequency]	LF (LW)	5	30–300 kHz 10–1 km	Navigation, time signals [https://en.wikipedia.org/wiki/Radio_clock] , AM longwave [https://en.wikipedia.org/wiki/Longwave] broadcasting (Europe and parts of Asia), RFID [https://en.wikipedia.org/wiki/RFID] , amateur radio [https://en.wikipedia.org/wiki/Amateur_radio]
Medium frequency [https://en.wikipedia.org/wiki/Medium_frequency]	MF (MW "AM radio")	6	300–3,000 kHz 1,000–100 m	AM [https://en.wikipedia.org/wiki/AM_broadcasting] (medium-wave) broadcasts, amateur radio, avalanche beacons [https://en.wikipedia.org/wiki/Avalanche_transceiver]
High frequency [https://en.wikipedia.org/wiki/High_frequency]	HF (SW)	7	3–30 MHz 1	Shortwave [https://en.wikipedia.org/wiki/Shortwave] broadcasts, citizens band radio [https://en.wikipedia.org/wiki/Citizens_band_radio] , amateur radio and over-the-horizon [https://en.wikipedia.org/wiki/Over-the-horizon] aviation communications, RFID [https://en.wikipedia.org/wiki/RFID] , over-the-horizon radar [https://en.wikipedia.org/wiki/Over-the-horizon_radar] , automatic link establishment [https://en.wikipedia.org/wiki/Automatic_link_establishment] (ALE) / near-vertical incidence skywave [https://en.wikipedia.org/wiki/Near_Vertical_Incidence_Skywave] (NVIS) radio communications, marine and mobile radio telephony [https://en.wikipedia.org/wiki/Marine_and_mobile_radio_telephony]

			0–10 m	
Very high frequency [https://en.wikipedia.org/wiki/Very_high_frequency]	VHF ("FM Radio")	8	30–300 MHz 10–1 m	FM [https://en.wikipedia.org/wiki/FM_broadcasting] , television [https://en.wikipedia.org/wiki/Television] broadcasts, line-of-sight ground-to-aircraft and aircraft-to-aircraft communications, land mobile and maritime mobile communications, amateur radio, weather radio [https://en.wikipedia.org/wiki/Weather_radio]
Ultra high frequency [https://en.wikipedia.org/wiki/Ultra_high_frequency]	UHF	9	300–3,000 MHz 1–0.1 m	Television broadcasts, microwave oven [https://en.wikipedia.org/wiki/Microwave_oven] , microwave [https://en.wikipedia.org/wiki/Microwave] devices/communications, radio astronomy [https://en.wikipedia.org/wiki/Radio_astronomy] , mobile phones [https://en.wikipedia.org/wiki/Mobile_phone] , wireless LAN [https://en.wikipedia.org/wiki/Wireless_LAN] , Bluetooth [https://en.wikipedia.org/wiki/Bluetooth] , ZigBee [https://en.wikipedia.org/wiki/ZigBee] , GPS [https://en.wikipedia.org/wiki/GPS] and two-way radios such as land mobile, FRS [https://en.wikipedia.org/wiki/Family_Radio_Service] and GMRS [https://en.wikipedia.org/wiki/GMRS] radios, amateur radio, satellite radio [https://en.wikipedia.org/wiki/Satellite_radio] , Remote control Systems, ADSB [https://en.wikipedia.org/wiki/Automatic_dependent_surveillance_%E2%80%93_broadcast]
Super high frequency [https://en.wikipedia.org/wiki/Super_high_frequency]	SHF	10	3–30 GHz 100–10 mm	Radio astronomy, microwave devices/communications, wireless LAN, DSRC [https://en.wikipedia.org/wiki/Dedicated_short-range_communications] , most modern radars [https://en.wikipedia.org/wiki/Radar] , communications satellites [https://en.wikipedia.org/wiki/Communications_satellite] , cable and satellite television broadcasting, DBS [https://en.wikipedia.org/wiki/Direct-broadcast_satellite] , amateur radio, satellite radio [https://en.wikipedia.org/wiki/Satellite_radio]
Extremely high frequency [https://en.wikipedia.org/wiki/Extremely_high_frequency]	EHF	11	30–300 GHz 10–1 m	Radio astronomy, high-frequency microwave radio relay [https://en.wikipedia.org/wiki/Microwave_radio_relay] , microwave remote sensing [https://en.wikipedia.org/wiki/Remote_sensing] , amateur radio, directed-energy weapon [https://en.wikipedia.org/wiki/Active_Denial_System] , millimeter wave scanner [https://en.wikipedia.org/wiki/Millimeter_wave_scanner] , wireless LAN (802.11ad)

Terahertz [https://en.wikipedia.org/wiki/Terahertz_radiation] or Tremendously high frequency [https://en.wikipedia.org/wiki/Tremendously_high_frequency]	THz or THF	12	300–3,000 GHz 1–0.1 mm	Experimental medical imaging to replace X-rays, ultrafast molecular dynamics, condensed-matter physics [https://en.wikipedia.org/wiki/Condensed-matter_physics] , terahertz time-domain spectroscopy [https://en.wikipedia.org/wiki/Terahertz_time-domain_spectroscopy] , terahertz computing/communications, remote sensing [https://en.wikipedia.org/wiki/Remote_sensing] ,

Posted 6th July 2013 by Billy

0 Add a comment

4th July 2013

Sun bath for my crystal radios

Last Sunday I went to the gold coast beach in Hong Kong.
It's quite close to the FM radio transmission station.
You can listen to the FM radio from the battery free FM crystal radio just by raising the antenna.

I've also took this chance to take a good photo of all the different types of crystal radios I made.

[http://3.bp.blogspot.com/-vtr2eymKSoM/UdVspfyA98I/AAAAAAAABRc/ONGlXkIaRy8/s442/Screen+Shot+2013-07-01+at+5.57.45+PM.png]

Dynamic Views theme. Powered by Blogger.

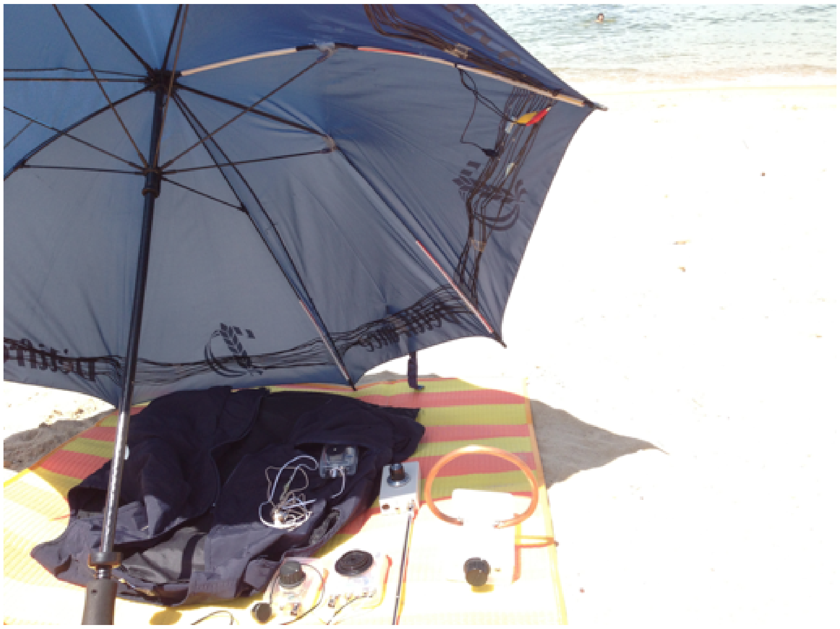

[http://2.bp.blogspot.com/-x9uRI5r7_zI/UdVspf56XFI/AAAAAAAABRQ/pUrJpD1wQJY/s513/Screen+Shot+2013-07-01+at+5.58.54+PM.png]

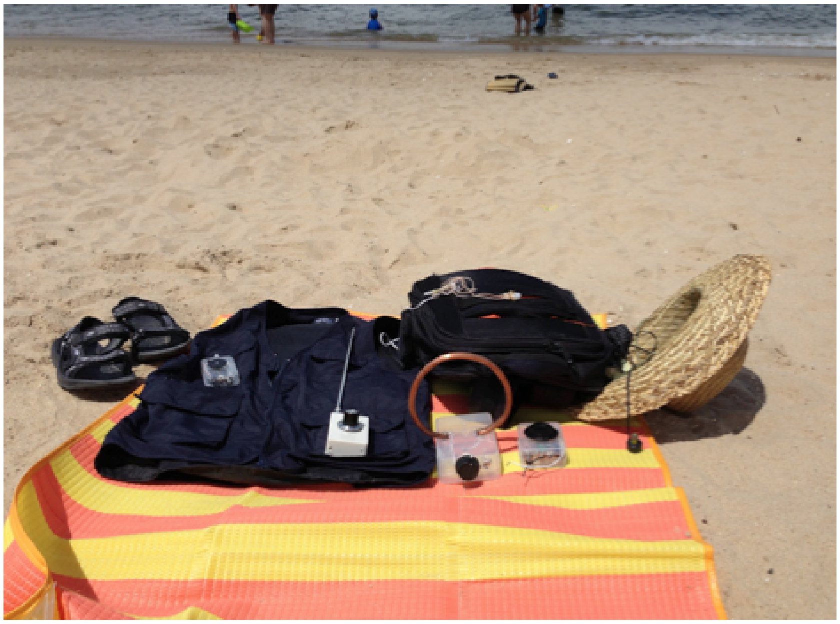

[http://1.bp.blogspot.com/-LRdBXnct1pU/UdVspShIfoI/AAAAAAAABRU/8zb8Wn46DIk/s469/Screen+Shot+2013-07-01+at+5.59.50+PM.png]

Links to my chinese blog

1. 雨伞矿机 - http://www.crystalradio.cn/thread-389699-1-1.html [http://www.crystalradio.cn/thread-389699-1-1.html]
2. 馬甲矿机 - http://www.crystalradio.cn/thread-414334-1-1.html [http://www.crystalradio.cn/thread-414334-1-1.html]
3. 背包矿机 - http://www.crystalradio.cn/thread-397792-1-1.html [http://www.crystalradio.cn/thread-397792-1-1.html]
4. 帽子矿机 - http://www.crystalradio.cn/thread-390223-1-1.html [http://www.crystalradio.cn/thread-390223-1-1.html]
5. FM矿机 - 參照bd1234567老師的创作 http://www.crystalradio.cn/forum ... read&tid=220240 [http://www.crystalradio.cn/forum.php?mod=viewthread&tid=220240]
6. FM大环矿机 - 參照梁老師的创作 http://www.crystalradio.cn/forum.php?mod=viewthread&tid=73497 [http://www.crystalradio.cn/forum.php?mod=viewthread&tid=73497]
7. AM / FM多用途矿机 - http://www.crystalradio.cn/thread-420838-1-1.html [http://www.crystalradio.cn/thread-420838-1-1.html]

Dynamic Views theme. Powered by Blogger.

Posted 4th July 2013 by Billy

0 Add a comment

legalzoom
Let's Make Your Business Official!
Come back

3rd July 2013

The simplest FM Crystal Radio

The simplest battery-free FM Crystal Radio.

Make the world's simplest battery-free FM crystal radio

Only consists of two high frequency diode: 1N34A or ISS86 or ISS106.
A 100K Resistor.
An 1.8M antenna.
And a crystal or high impedance earphone

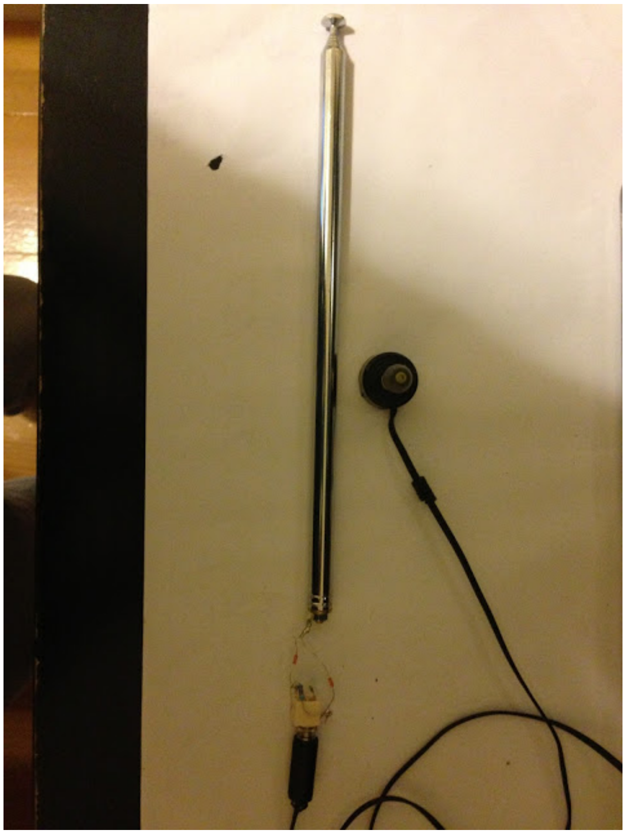

[http://1.bp.blogspot.com/-x9lr6oLJ9z8/UdQ_lBNismI/AAAAAAAABQE/pfbpj7x0Azs/s1600/IMG_2607.JPG]

Refer to this circuit diagram

Dynamic Views theme. Powered by Blogger.

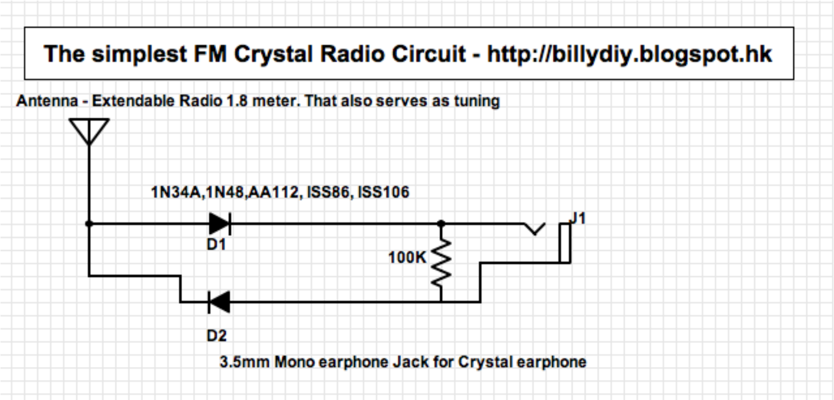
The simplest FM Crystal Radio Circuit - http://billydiy.blogspot.hk
Antenna - Extendable Radio 1.8 meter. That also serves as tuning
1N34A,1N48,AA112, ISS86, ISS106
D1
100K
J1
D2
3.5mm Mono earphone Jack for Crystal earphone

[http://2.bp.blogspot.com/-G9onfk5x8FA/UdRDYhFFyAI/AAAAAAAABRA/mtiP8wFX78E/s639/Screen+Shot+2013-07-03+at+11.28.54+PM.png]

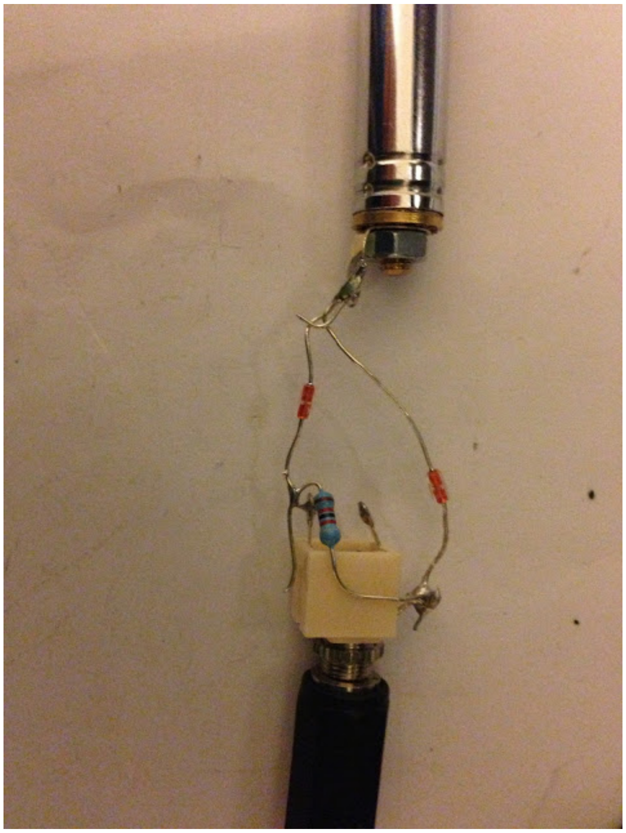

[http://4.bp.blogspot.com/-UVDtJdx6leQ/UdQ_lE_swzI/AAAAAAAABQM/ToNTHsm8pdc/s1600/IMG_2608.JPG]

[http://1.bp.blogspot.com/-jvFTM9CBpKQ/UdQ_kx4Q35I/AAAAAAAABQA/W1wxPV0ezZk/s1600/IMG_2609.JPG]

Posted 3rd July 2013 by Billy

1 View comments

28th June 2013

Other Crystal Radio Circuits

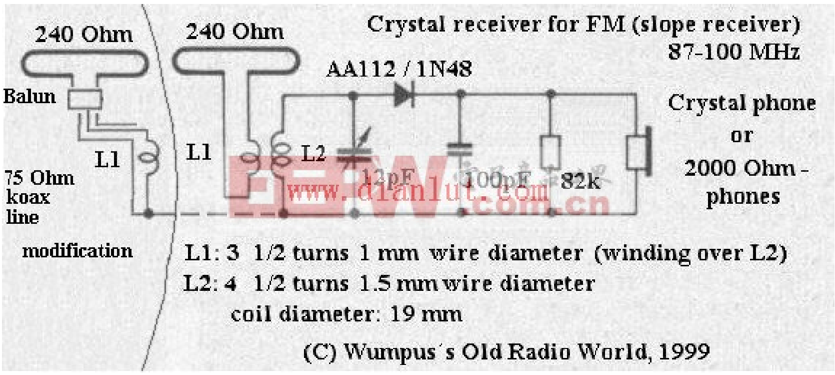
240 Ohm
240 Ohm
Crystal receiver for FM (slope receiver)
87-100 MHz
AA112 / 1N48
Balun
Crystal phone
or
L1
L1
L2
2000 Ohm -
phones
75 Ohm
koax
line
12pF
100pF
82k
www.dianlut.com.cn
modification
L1: 3 1/2 turns 1 mm wire diameter (winding over L2)
L2: 4 1/2 turns 1.5 mm wire diameter
coil diameter: 19 mm
(C) Wumpus´s Old Radio World, 1999

[http://1.bp.blogspot.com/-uIELBKMZo-I/Uc2DPHecvZI/AAAAAAAABK8/AgJ9AnIi42s/s573/2_101224175651_1.JPG]

Posted 28th June 2013 by Billy

0 Add a comment

28th June 2013

Wearable Battery-Free Jacket Crystal Radio

Build an always wearable Crystal Radio by winding a coil at the inner linings of your jacket.
You can walk with it, run with it, ride bicycle with it and continue to listen to radio.

[http://2.bp.blogspot.com/-tdd3lUhW9JM/Uc2ZTDN5LfI/AAAAAAAABNw/jmid0PV_fmU/s1600/IMG_2383.JPG]

[http://4.bp.blogspot.com/-h6QYObgCJhE/Uc2ZTJbbUnI/AAAAAAAABNo/l3EgotNHT8U/s1600/IMG_2384.JPG]

Dynamic Views theme. Powered by Blogger.

[http://2.bp.blogspot.com/-jgONTqzWIy8/Uc2ZTPsMk6I/AAAAAAAABNk/9NWtlo5UrWE/s1600/IMG_2385.JPG]

[http://1.bp.blogspot.com/-qYLeo4x6eHI/Uc2ZUcyhcfI/AAAAAAAABOI/da6I7jh7KQs/s1600/IMG_2388.JPG]

Dynamic Views theme. Powered by Blogger.

[http://2.bp.blogspot.com/-smKxUD_zSok/Uc2ZU6qrRTI/AAAAAAAABO4/8lI_Tb4p5Pw/s1600/IMG_2391.JPG]

[http://4.bp.blogspot.com/-VKigLEZD8SU/Uc2ZU6EMTbI/AAAAAAAABOQ/CwqnV6A_WPQ/s267/Screen+Shot+2013-06-04+at+8.59.54+PM.png]

Dynamic Views theme. Powered by Blogger.

[http://1.bp.blogspot.com/-qYLeo4x6eHI/Uc2ZUcyhcfI/AAAAAAAABOU/oqw31OQSSuk/s1600/IMG_2388.JPG]

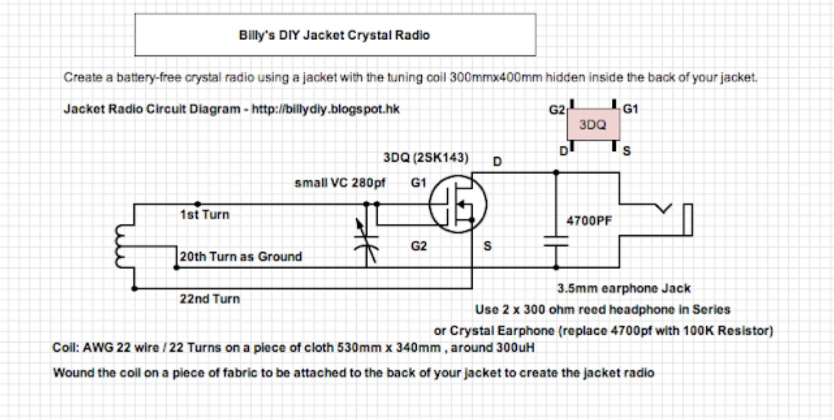

Billy's DIY Jacket Crystal Radio

Create a battery-free crystal radio using a jacket with the tuning coil 300mmx400mm hidden inside the back of your jacket.

Jacket Radio Circuit Diagram - http://billydiy.blogspot.hk

G2
G1
3DQ
D
S
3DQ (2SK143)
D
small VC 280pf
G1
1st Turn
4700PF
G2
S
20th Turn as Ground
22nd Turn
3.5mm earphone Jack

Use 2 x 300 ohm reed headphone in Series

or Crystal Earphone (replace 4700pf with 100K Resistor)

Coil: AWG 22 wire / 22 Turns on a piece of cloth 530mm x 340mm , around 300uH

Wound the coil on a piece of fabric to be attached to the back of your jacket to create the jacket radio

[http://1.bp.blogspot.com/-dTQhQcSn_5s/Uc2aUpsH7mI/AAAAAAAABPw/VPUaOywgJIY/s790/Screen+Shot+2013-06-28+at+10.14.20+PM.png]

Dynamic Views theme. Powered by Blogger.

PROFESSIONAL WIRE & CABLE
MIZU
WIRE
AWG# 22
1007
300
FEET
U.S.A. LICENCED
PRODUCT CODE
53-16-0221 (BLACK)

[http://4.bp.blogspot.com/-zk7AhAeqsXQ/Uc2ZVl2Z3kI/AAAAAAAABOY/KH8dma6pRZ0/s235/Screen+Shot+2013-06-04+at+9.00.25+PM.png]

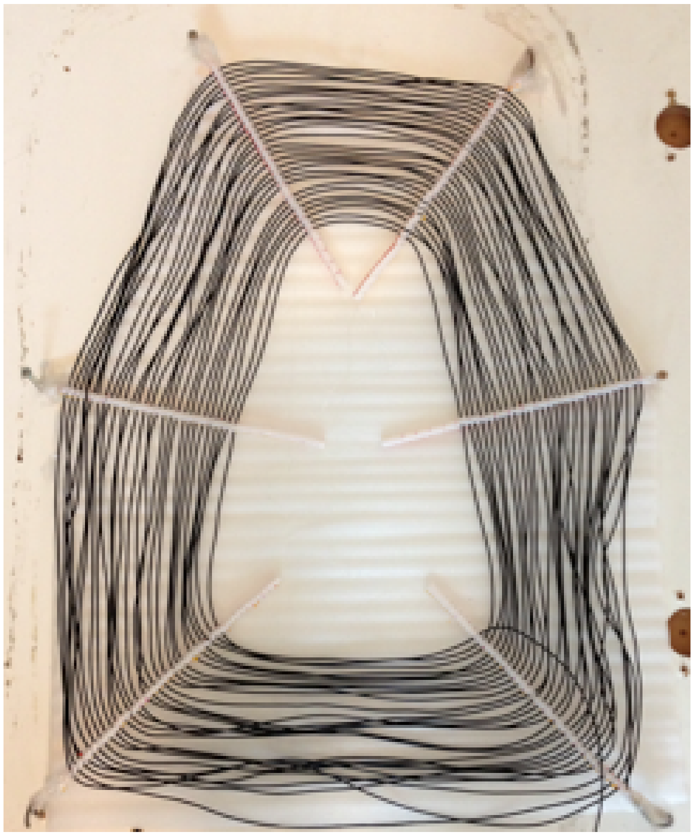

[http://1.bp.blogspot.com/-9pm2x80WsdA/Uc2ZVtDK6xI/AAAAAAAABOg/ReHQwKYY8RY/s294/Screen+Shot+2013-06-04+at+9.00.47+PM.png]

Dynamic Views theme. Powered by Blogger.

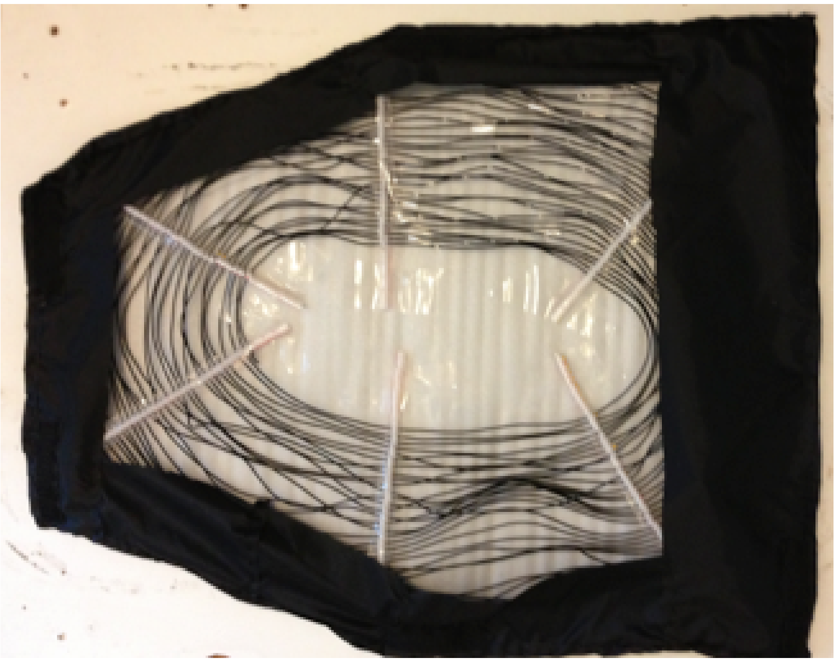

[http://2.bp.blogspot.com/-U_H5k2MTg-w/Uc2ZV5OUAgI/AAAAAAAABOo/yTd7RExNVng/s325/Screen+Shot+2013-06-04+at+9.01.03+PM.png]

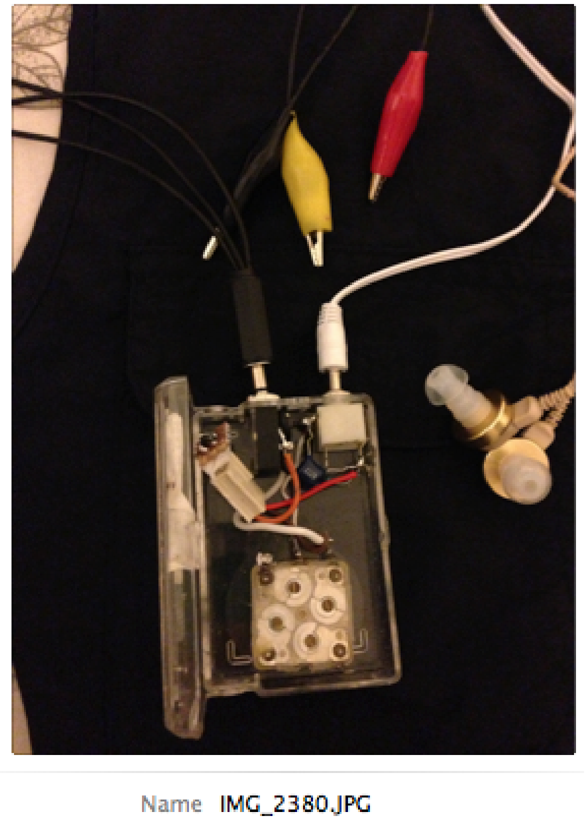

Name IMG_2380.JPG

[http://1.bp.blogspot.com/-ITHllJ1hQGc/Uc2ZWAILx8I/AAAAAAAABO0/24gT2gO_ZvQ/s404/Screen+Shot+2013-06-04+at+9.01.50+PM.png]

Dynamic Views theme. Powered by Blogger.

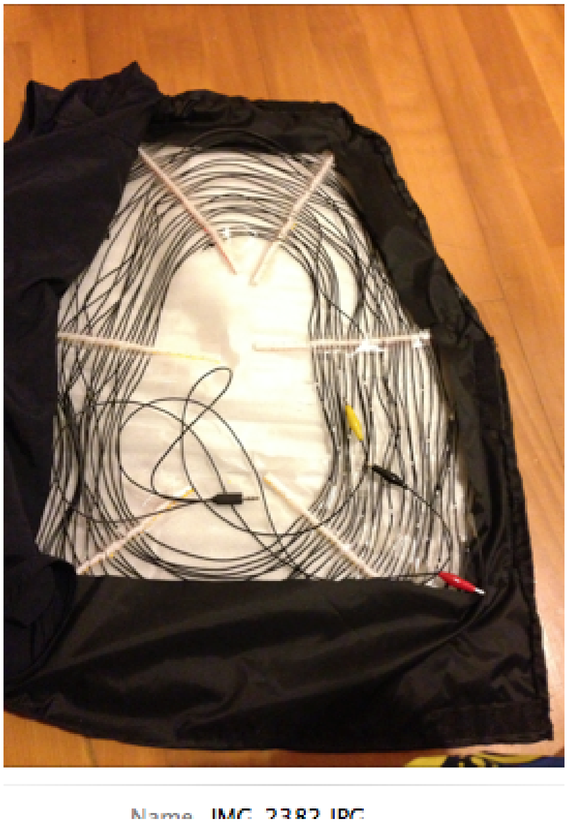

[http://4.bp.blogspot.com/-RgIOZLtkEJU/Uc2ZWkHVdbI/AAAAAAAABPA/fyeH5ppVWdM/s396/Screen+Shot+2013-06-04+at+9.02.12+PM.png]

[http://2.bp.blogspot.com/-holyCfigwDk/Uc2ZXIFdhxI/AAAAAAAABPQ/juheK5d07x4/s475/Screen+Shot+2013-06-04+at+9.09.30+PM.png]

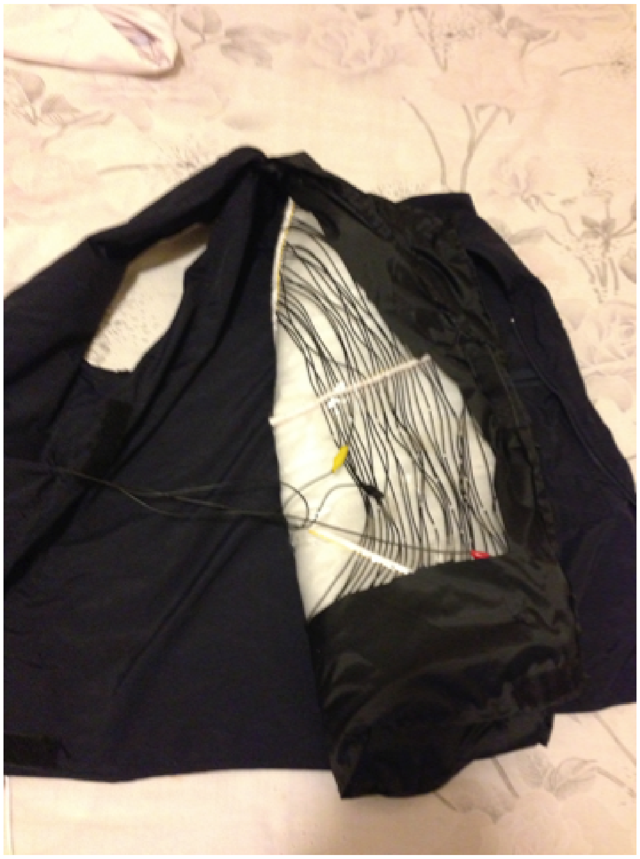

[http://1.bp.blogspot.com/-HtsnQEDeYLQ/Uc2ZXP_6UAI/AAAAAAAABPI/eFawBMa4PO4/s479/Screen+Shot+2013-06-04+at+9.09.59+PM.png]

[http://4.bp.blogspot.com/-RtT9MjHo9AM/Uc2ZXrU8JBI/AAAAAAAABPg/NNxsZ7F_cdE/s480/Screen+Shot+2013-06-04+at+9.10.08+PM.png]

0 Add a comment

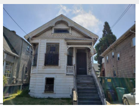

Great Deal At Oakland Ca!! in Oakland...
$499,000

Hot Off Market Way Below Market...
$140,000

Hot Deal At Oakland Ca!!! in Oal
$650,000

28th June 2013

AM/FM/Loop Crystal Radio

[http://3.bp.blogspot.com/-W5MoCILOqOU/Uc2WU1CTbbI/AAAAAAAABNA/hH64_FNxMpo/s436/Screen+Shot+2013-06-23+at+12.06.23+PM.png]

Build a multifunction battery free crystal radio to listen to either AM or FM radio.
You can also use the same radio to connect to a bigger loop coil to get stronger signal for the AM radio instead of relying on the smaller AM coil inside the radio.

Parts you will need:
A Diode that works in high frequency (200Mhz or above, e.g. AA112, 1N34A, 1N48, Iss86 or Iss106.
A Diode that work in AM frequency e.g 1N34A, 1N60, BAT85.
A quadruple PVC Variable capacitor with one end for 2x 20pf, the other end for 2x 140pf.
50pf and 180pf fixed ceramic capacitor
100K 1/4W resistor
33uH Choke coil.
AM radio coil - ferrite rod
3.5mm mono headphone jack
3.5mm stereo headphone jack with internal switches (for selecting whether to use the external big loop or the internal coil for AM.
A 2 by 3 switch to select AM or FM.
A high impedance 4K Ohm earphone or a crystal earphone.

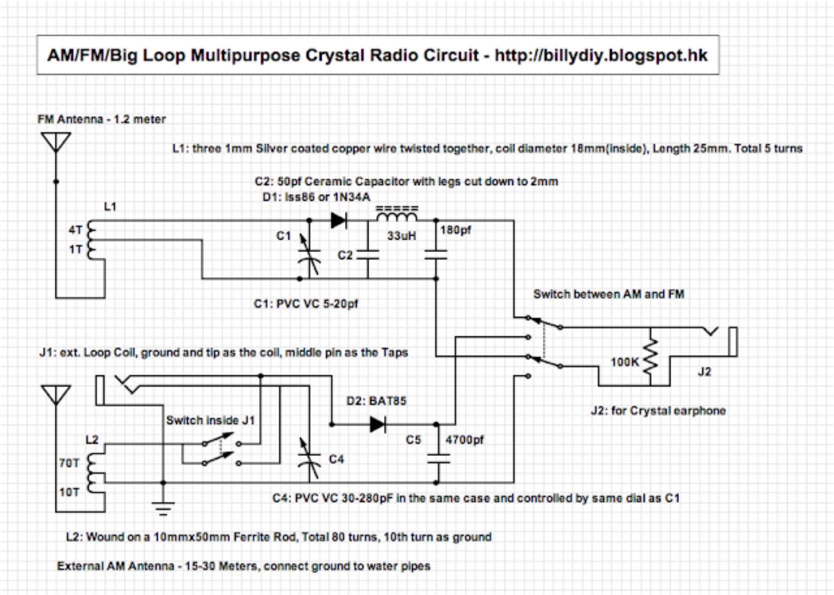
AM/FM/Big Loop Multipurpose Crystal Radio Circuit - http://billydiy.blogspot.hk
FM Antenna - 1.2 meter
L1: three 1mm Silver coated copper wire twisted together, coil diameter 18mm(inside), Length 25mm. Total 5 turns
C2: 50pf Ceramic Capacitor with legs cut down to 2mm
D1: Iss86 or 1N34A
L1
4T
1T
C1
C2
33uH
180pf
C1: PVC VC 5-20pf
Switch between AM and FM
J1: ext. Loop Coil, ground and tip as the coil, middle pin as the Taps
100K
J2
D2: BAT85
J2: for Crystal earphone
Switch inside J1
L2
70T
10T
C5
4700pf
C4
C4: PVC VC 30-280pF in the same case and controlled by same dial as C1
L2: Wound on a 10mmx50mm Ferrite Rod, Total 80 turns, 10th turn as ground
External AM Antenna - 15-30 Meters, connect ground to water pipes

[http://2.bp.blogspot.com/--ZAhsy2SRko/Uc2WUu2Gz6I/AAAAAAAABM8/czIHQF1iPzI/s855/Screen+Shot+2013-06-23+at+11.22.21+AM.png]

Dynamic Views theme. Powered by Blogger.

To connect to external Loop coil for AM reception instead of using the internal AM coil, connect 3-color crocodile camp to the 3.5mm Stereo jack - white for the tuning, black for ground, yellow for the decoder diode.

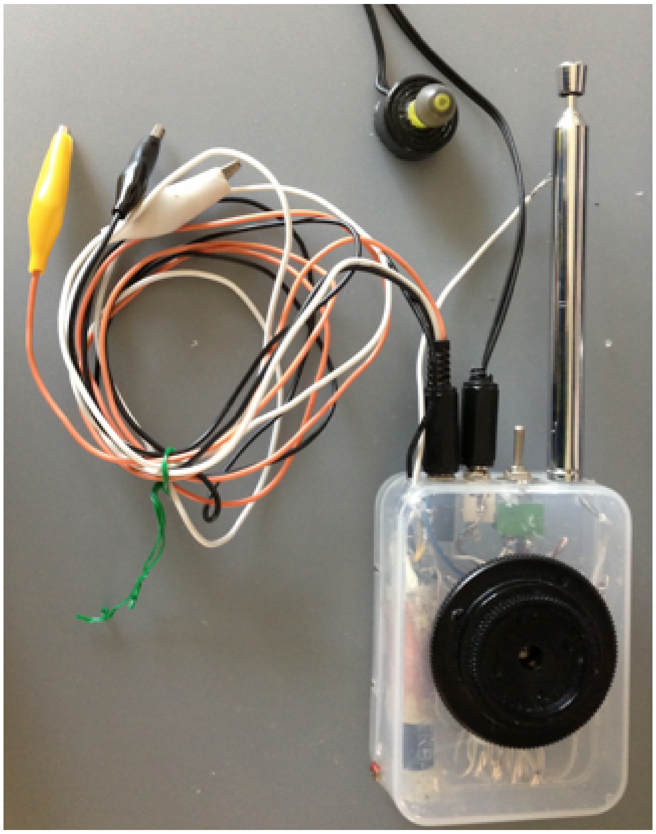

[http://3.bp.blogspot.com/-df0Gv1TiXME/Uc2WTw3VL8I/AAAAAAAABMs/hJaPbNYh4uQ/s415/Screen+Shot+2013-06-23+at+1.19.26+PM.png]

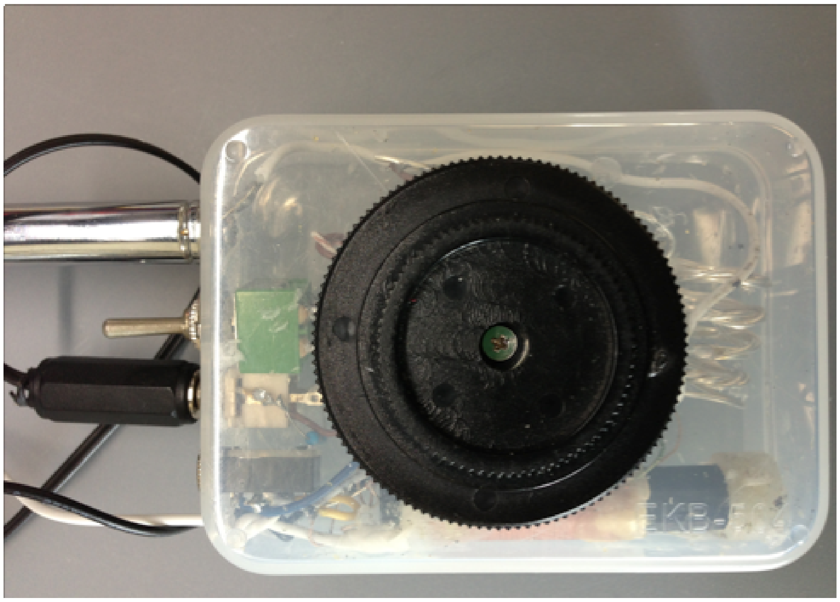

[http://2.bp.blogspot.com/-966mXwGZozE/Uc2WVmB2lLI/AAAAAAAABNY/P2KdoMFVIR0/s489/Screen+Shot+2013-06-23+at+12.07.13+PM.png]

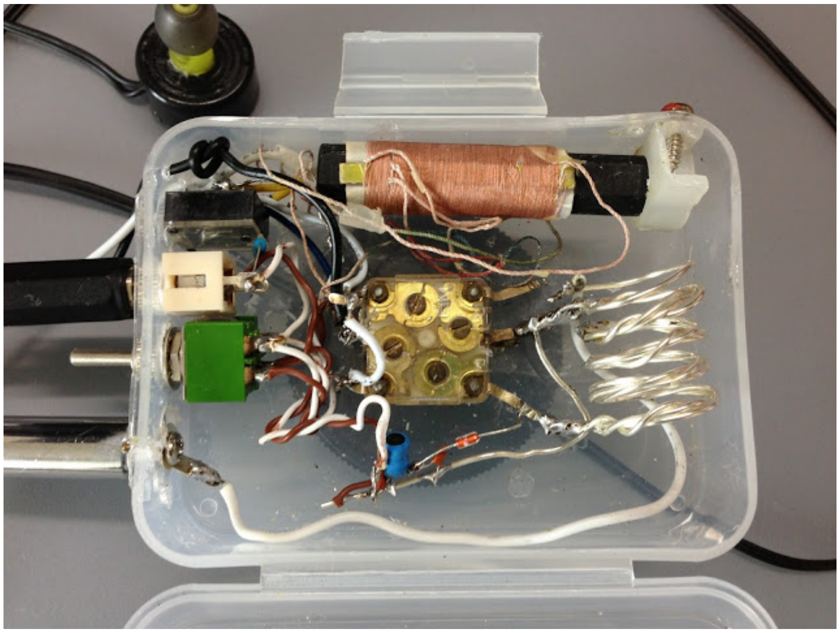

[http://1.bp.blogspot.com/-UURJiV8VfBY/Uc2WTLqcLFI/AAAAAAAABMc/q9_Y7C47Zb8/s1600/IMG_2521.JPG]

Posted 28th June 2013 by Billy

0 Add a comment

28th June 2013

FM Crystal Radio

Build a battery free FM crystal radio using 1mm silver plated copper wire

[http://2.bp.blogspot.com/-LyTLmJ0D5ao/Uc2SwAO-02I/AAAAAAAABL4/jkHYmvlimcE/s1600/IMG_2535.JPG]

This FM radio runs without battery and you can listen to FM radio by just placing this close to your window.
If reception is not good, you can attach an external FM antenna (or with antenna amplifier), instead of using the 1.2M radio antenna.

Parts you will need:
A Diode that works in high frequency (200Mhz or above, e.g. AA112, 1N34A, 1N48, Iss86 or Iss106.
A 0-12 pf AIR Variable capacitor - you can get a 0-22pf one, then remove all the copper plates and leave just 4.
50pf and 180pf fixed ceramic capacitor
100K 1/4W resistor
33uH Choke coil.
Silver platted copper wire 1mm diameter
1.2meter radio antenna.
3.5mm mini headphone jack
A high impedance 4K Ohm earphone or a crystal earphone.

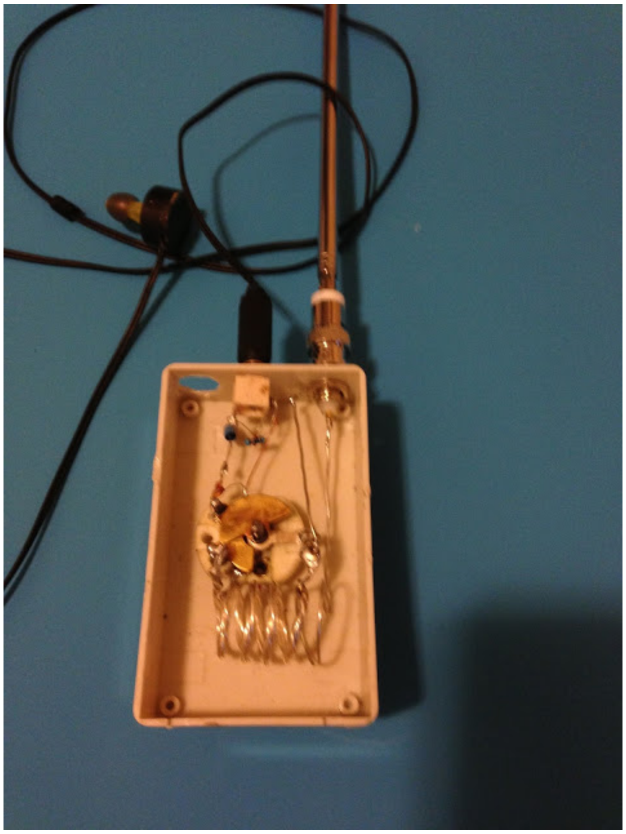

[http://3.bp.blogspot.com/-NhzpbldzMyU/Uc2SwLCh3LI/AAAAAAAABL8/J7ym0UWhqqk/s1600/IMG_2536.JPG]

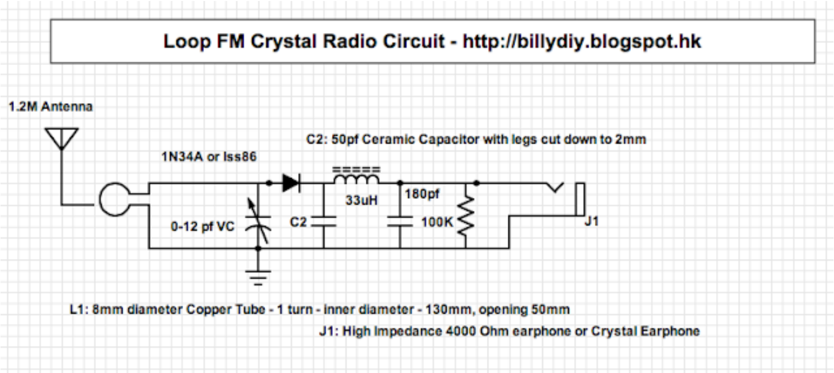
Loop FM Crystal Radio Circuit - http://billydiy.blogspot.hk
1.2M Antenna
1N34A or Iss86
C2: 50pf Ceramic Capacitor with legs cut down to 2mm
33uH
180pf
100K
0-12 pf VC
C2
J1
L1: 8mm diameter Copper Tube - 1 turn - inner diameter - 130mm, opening 50mm
J1: High Impedance 4000 Ohm earphone or Crystal Earphone

[http://2.bp.blogspot.com/-84ZQcpL-iDM/Uc2DeEbsVzI/AAAAAAAABLE/vF4um1tHpws/s763/Screen+Shot+2013-06-28+at+8.36.46+PM.png]

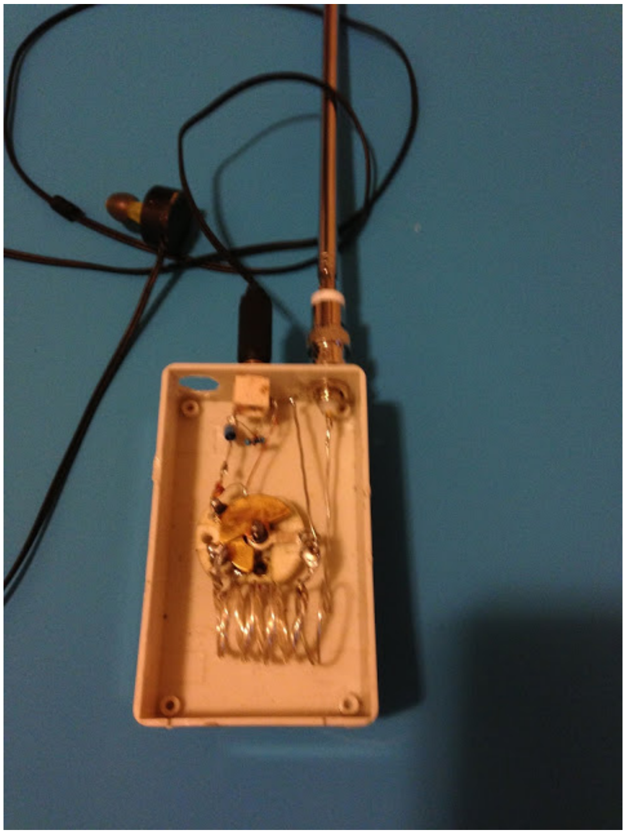

[http://3.bp.blogspot.com/-NhzpbldzMyU/Uc2SwLCh3LI/AAAAAAAABL0/eRdHXo7f22w/s1600/IMG_2536.JPG]

Posted 28th June 2013 by Billy

4 View comments

28th June 2013

Battery Free Loop FM Crystal Radio

Battery Free Loop FM Radio using an 8-mm diameter copper tube.

This FM radio runs without battery and you can listen to FM radio by just placing this close to your window.
If reception is not good, you can attach an external FM antenna (or with antenna amplifier), to the Loop at the point 1/3 of the loop close to the ground.

Parts you will need:
A Diode that works in high frequency (200Mhz or above, e.g. AA112, 1N34A, 1N48, Iss86 or Iss106.
A 0-12 pf AIR Variable capacitor - you can get a 0-22pf one, then remove all the copper plates and leave just 4.
50pf and 180pf fixed ceramic capacitor
100K 1/4W resistor
33uH Choke coil.
3.5mm mini headphone jack
A high impedance 4K Ohm earphone or a crystal earphone.
Some wires.

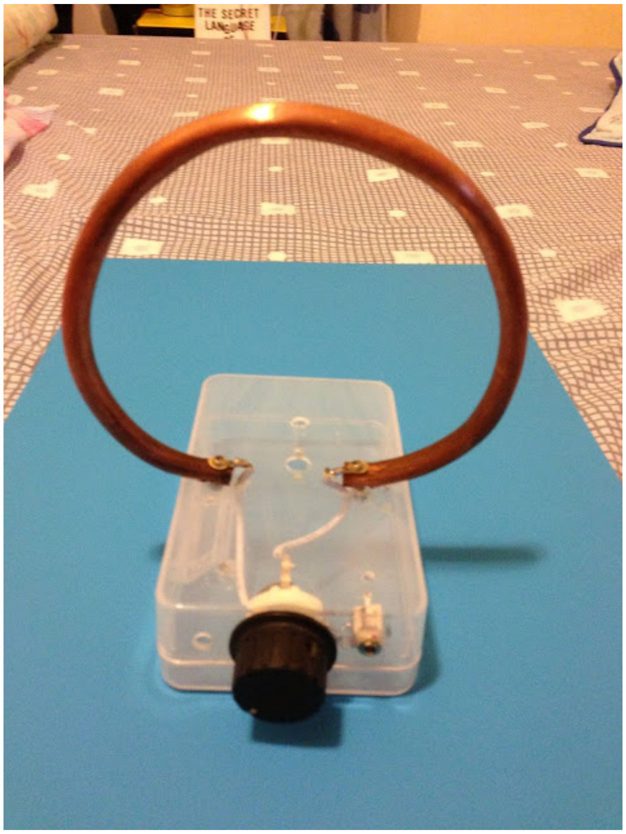

[http://2.bp.blogspot.com/-1QLVTR6aixI/Uc2EVLvYurI/AAAAAAAABLU/FHVgksSmYt0/s1600/IMG_2527.JPG]

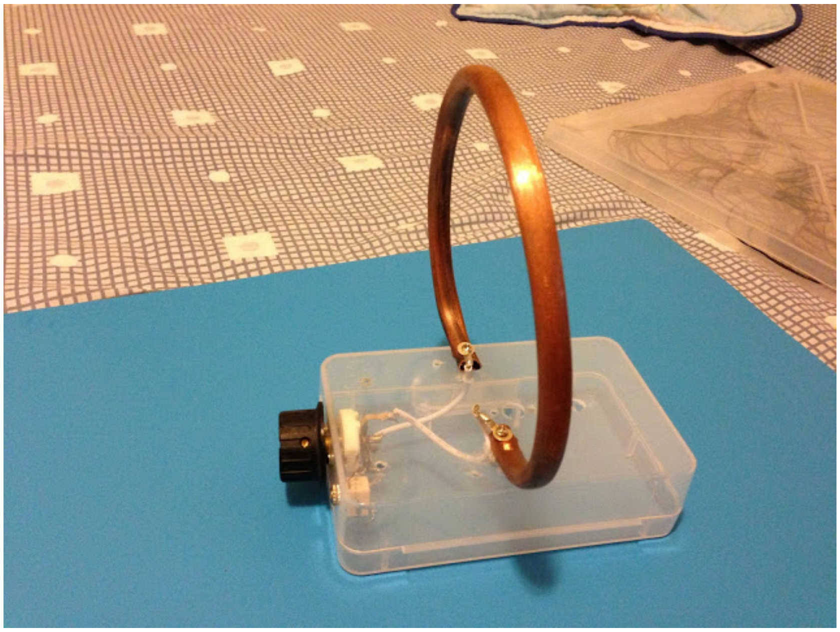

[http://1.bp.blogspot.com/-1DHVXlxHjMY/Uc2EVCFGUkI/AAAAAAAABLk/K03mBQ-9OXM/s1600/IMG_2528.JPG]

Dynamic Views theme. Powered by Blogger.

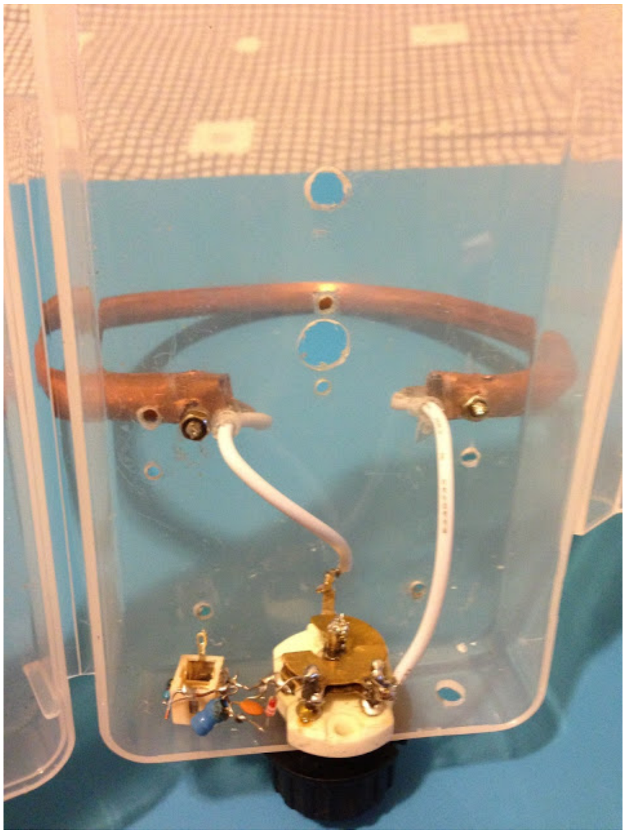

[http://1.bp.blogspot.com/-U8f34Fh_4EM/Uc2EVA6kIQI/AAAAAAAABLY/aI6OZ2c5Edc/s1600/IMG_2529.JPG]

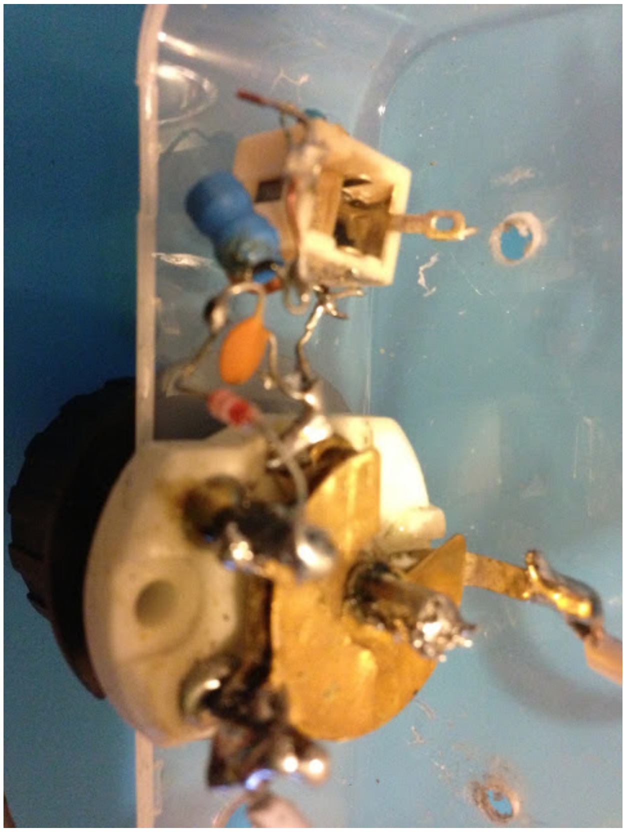

[http://1.bp.blogspot.com/-ONsnEMeXv-g/Uc2EWEFWNnI/AAAAAAAABLo/ap0dwwE1PKY/s1600/IMG_2531.JPG]

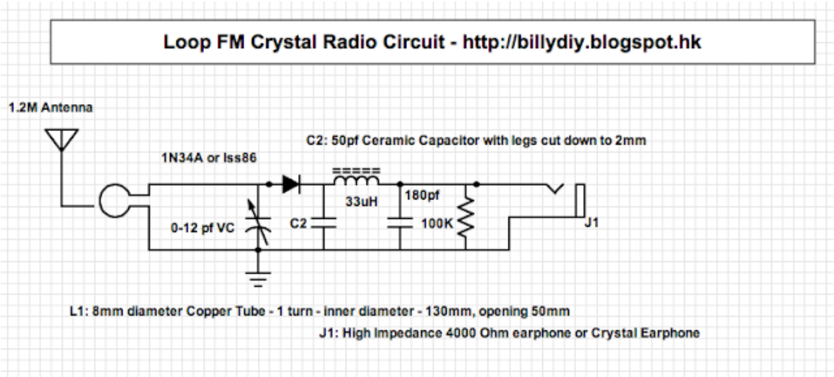
Loop FM Crystal Radio Circuit - http://billydiy.blogspot.hk
1.2M Antenna
C2: 50pf Ceramic Capacitor with legs cut down to 2mm
1N34A or Iss86
33uH
180pf
100K
J1
0-12 pf VC
C2
L1: 8mm diameter Copper Tube - 1 turn - inner diameter - 130mm, opening 50mm
J1: High Impedance 4000 Ohm earphone or Crystal Earphone

[http://1.bp.blogspot.com/-84ZQcpL-iDM/Uc2DeEbsVzI/AAAAAAAABLA/R_rFlJDlWn8/s763/Screen+Shot+2013-06-28+at+8.36.46+PM.png]

Posted 28th June 2013 by Billy

1 View comments

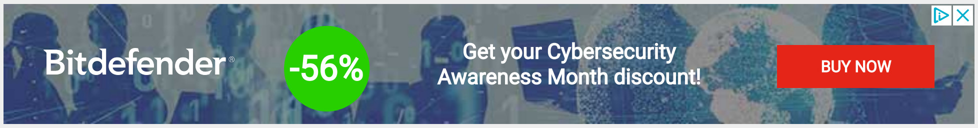
Bitdefender -56% Get your Cybersecurity Awareness Month discount! BUY NOW

18th May 2013

Make a variable capacitor out of a CD case and a CD

Make a variable capacitor out of a CD and a CD case.

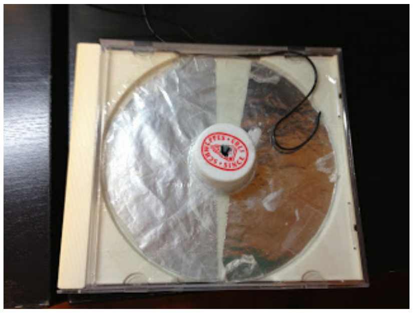

[http://2.bp.blogspot.com/-fjTMYdnkRSQ/UZgpVKlgr8I/AAAAAAAABKY/BgEkJoISM4M/s1600/IMG_1865.jpg]

Use CD that is transparent (without the magnetic/metal coating.

[http://1.bp.blogspot.com/-Izm_2BkhKVU/UZgmsAytz2I/AAAAAAAABI4/Rm2qMOOijil/s1600/IMG_1855.JPG]

Cut a circular piece of aluminium paper foil (even better if you can use copper foil), a little bit smaller than the size of the CD. Then cut it in half.
Paste one half of the aluminium paper foil onto the bottom of the CD case.

Paste the other half onto one side of the CD.
Tape wires onto the aluminium paper.

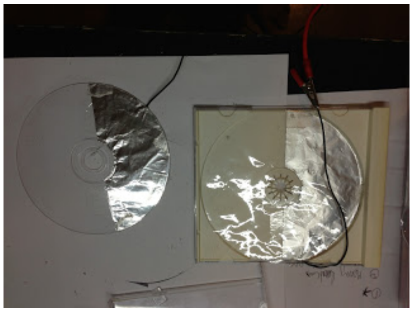

[http://4.bp.blogspot.com/-AhSF6NvwA5A/UZgmsA2llhI/AAAAAAAABI8/lm5waOADzDU/s1600/IMG_1858.JPG]

Cut a circular piece of thin plastic film (as soon as you can find, yet strong enough to withstand the pressure and still insulate the aluminium foil from one another).
Paste the plastic film onto the bottom of the CD case on directly top of the aluminium paper foil.

Put the CD on top with aluminium paper facing down).
Paste a plastic bottle caps on the centre of CD to function as the knob to turn the variable capacitor.
In front of the CD case, open a hole slightly larger than the plastic caps.
on the CD case. Close the CD case.

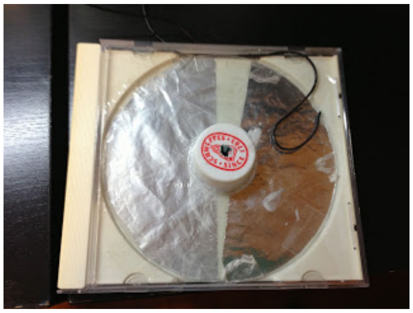

[http://1.bp.blogspot.com/-oNHHtxWGxn4/UZgqEgChk6I/AAAAAAAABKk/8JVTjJD8ADg/s1600/IMG_1865.jpg]

Test the CD Variable Capacitor using a Capacitance Meter. Mine can have a capacitance from 51pf to 233pf when turning the knob. You may be able to do better by pressing the CD and the CD case firmer together. Or replace the aluminium foil with copper foil.

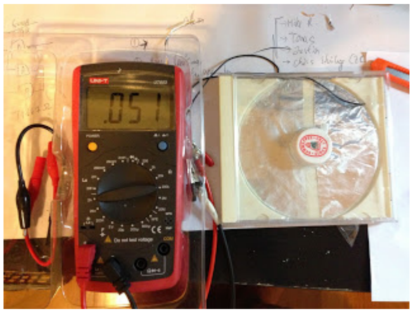

[http://4.bp.blogspot.com/-5bFY1MxzH1w/UZgonsTLu7I/AAAAAAAABKI/X4ureG3Esss/s1600/IMG_1860.jpg]

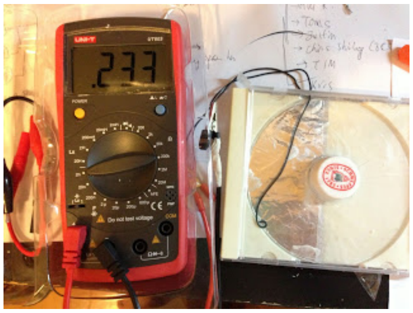

[http://4.bp.blogspot.com/-ocCLqbtDhVQ/UZgon6TKWGI/AAAAAAAABKM/9nSD9nMt6a4/s1600/IMG_1862.jpg]

Posted 18th May 2013 by Billy

2 View comments

19th April 2013

Backpack Crystal Radio

I was trying to bring the crystal radio out of the studio and into daily life.
However, whenever I use my umbrella crystal radio or hat crystal radio, people would give me a strange look.If we can hide a 300mmx400mm rectangular coil inside he backpack , we can build a backpack crystal radio and listen to musics as you travel.

And I did it! Checkout this youtube video:
http://www.youtube.com/watch?v=MbV9e09-Dxw [http://www.youtube.com/watch?v=MbV9e09-Dxw]
Although the volume of radio smaller than an umbrella, but you can still listen to some of the local radio stations clearly.
I can listen to the crystal radio every day on my way to work. The radio is loud enough and even louder in locations with a better radio reception.

Remember to leave me some comments (at the comment box) at the bottom of this if you have read this blog.

Warning: If you are going on a flight or to a sensitive area, please remember to remove the coil from the backpack before going, otherwise you may not be able to board the aircraft.

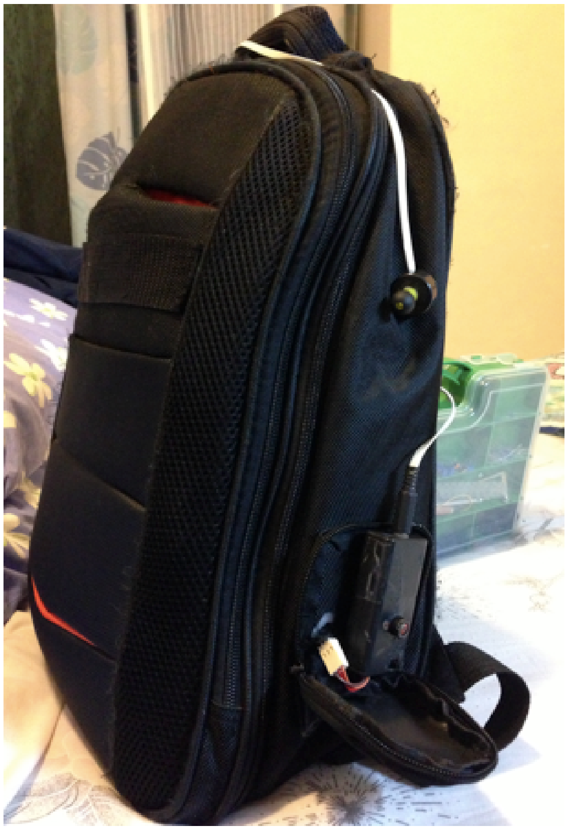

[http://4.bp.blogspot.com/-VY4RfayhLZ4/UXFIeGq43yI/AAAAAAAABBo/VgW_NXnFnTA/s1600/Screen+Shot+2013-04-19+at+8.33.07+PM.png]

[http://2.bp.blogspot.com/-iQektzNpj58/UXFIeGQI-NI/AAAAAAAABBw/_aTVC8cHqdE/s1600/Screen+Shot+2013-04-19+at+8.32.33+PM.png]

Circuit Diagram

Backpack Radio Circuit Diagram - http://billydiy.blogspot.hk

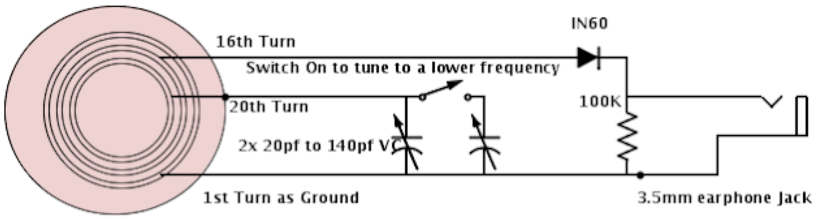

Coil: use any soft insulated wire (e.g. AWG 22). Wound a total of 20 turns on a box 300mm x 400mm
Then put the coil inside the cover of the backpack
See further details on how to build the backpack radio at http://billydiy.blogspot.hk

[http://1.bp.blogspot.com/-Q3ueJXl4Ak4/UXFIeEDQC-I/AAAAAAAABBs/nzQDabrMuzw/s1600/Screen+Shot+2013-04-19+at+7.28.36+PM.png]

Components you will need.

A backpack 300x400 mm or more, the larger the diameter, the more sensitive the radio.
The following instructions assume a 300x400 mm backpack.
Diode IN60, IN34A,
100K resistor,
40pf to 280PF variable capacitor (you can get it from an old radio)
3.5mm headphone jack,
18 meters of thin wire
A plastic battery case (those for two 2A Batteries) as a case for he crystal radio.
Crystal headset or 2K ohms or more high-impedance headphones. If you do not have one, follow instructions here to make one.

http://billydiy.blogspot.hk/2013/03/build-crystal-earphone-for-crystal.html

How to make it

A. Backpack Antenna

1. Found a a box the same size as the backpack. Then wind the wires around it to make the coil. A total of 20 turns are required. Then create a tap for the 16th turn. When finished, use tapes to tie the coil together. The coil inductance should be about 300uH. Then line-to-line capacitance is around 18PF. It will match well with a 20 pF to 280pF variable capacitor to tune into both radio stations at the low-end (531kHz) and the high-end (1453Khz).
2. Cut the front cover of the backpack to create an opening to install the coil.
3. Cut some more holes inside the backpack, to pass the wires inside the backpack to the side of the backpack.
4. Create a small side pocket to hold the radio with some cloth and zippers
5. Put the coil inside the opening created at the front cover of the backpack.
6. Pass the wires from the coil through the opening to the side pocket of the backpack.
7. Connect the the 1st turn to terminal 1, 16th turn to terminal 2, 20th turn to terminal 3.

Dynamic Views theme. Powered by Blogger.

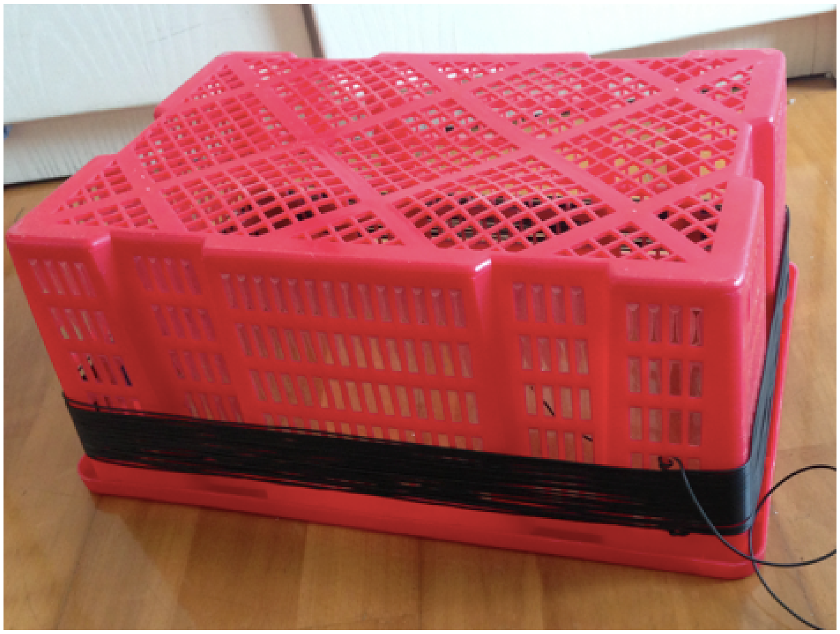

[http://1.bp.blogspot.com/-2zShyrNeYOw/UXFIetsHVSI/AAAAAAAABB0/2vz6rDDw6dQ/s1600/Screen+Shot+2013-04-19+at+8.33.56+PM.png]

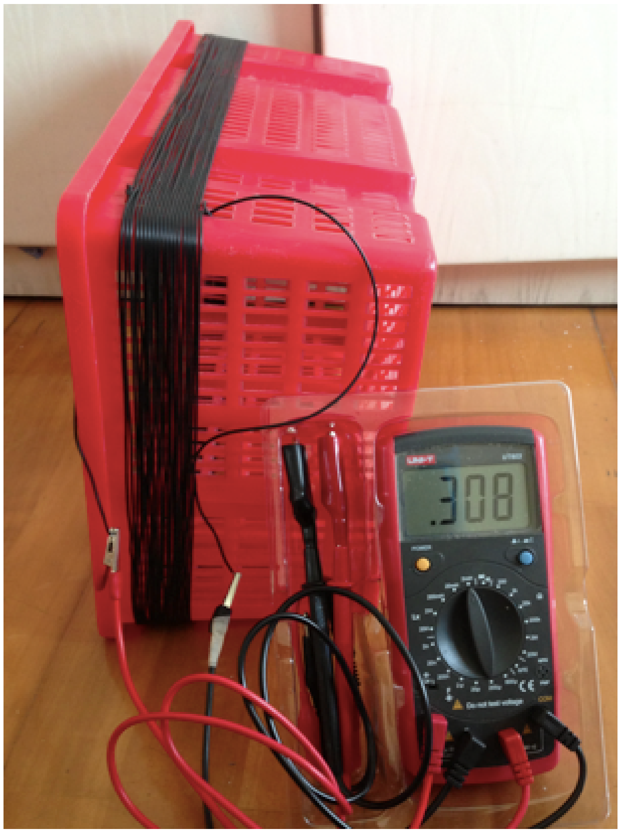

[http://3.bp.blogspot.com/-dazrCkp6ODQ/UXFIem2W7yI/AAAAAAAABB8/ZAJwpfJ9w1A/s1600/Screen+Shot+2013-04-19+at+8.34.19+PM.png]

[http://2.bp.blogspot.com/-KYYk7WNEXXs/UXFIfdkGrYI/AAAAAAAABCQ/mR_1Z080bN4/s1600/Screen+Shot+2013-04-19+at+8.35.47+PM.png]

[http://1.bp.blogspot.com/-AXlnP3vvwHo/UXFle1gEfjI/AAAAAAAABCA/DLqzB13ZlU0/s1600/Screen+Shot+2013-04-19+at+8.35.02+PM.png]

Dynamic Views theme. Powered by Blogger.

[http://1.bp.blogspot.com/-_X74ZnSUDrI/UXFIfW_66vI/AAAAAAAABCI/fldmEj8rjiM/s1600/Screen+Shot+2013-04-19+at+8.40.42+PM.png]

B. Crystal radio (without coil).

1. Drills holes in the plastic 2A battery box to install the headphone jack, variable capacitor and to pass the terminal wires.
2. Get the headphone jack and a variable capacitor installed, tighten the screws.
3. Refer to the circuit diagram and photos, weld the parts together.
4. The switch lead is connected to the two stator variable capacitor used to select the low-end or high-end radio.
5. Connect the ground wire from the headphone jack to terminal 1.

Connected diode to terminal 2.
Connect the variable capacitor stator to terminal 3.

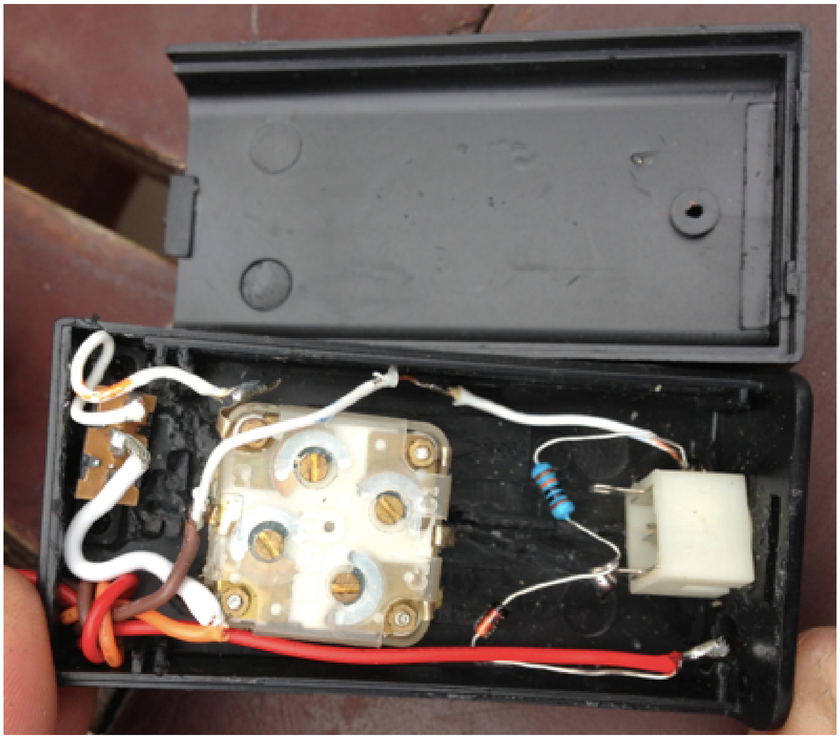

[http://1.bp.blogspot.com/-udqhWYc-XzU/UXFIg-iJU8I/AAAAAAAABC4/lthXiUY-Z30/s1600/Screen+Shot+2013-04-19+at+8.44.11+PM.png]

[http://2.bp.blogspot.com/-E_u3RVnQhnY/UXFIfpf4izI/AAAAAAAABCU/XS1Hk91t8aY/s1600/Screen+Shot+2013-04-19+at+8.41.07+PM.png]

[http://2.bp.blogspot.com/-wwxL_Pv5Urw/UXFIgNM6bII/AAAAAAAABCo/o5WGImptXpU/s1600/Screen+Shot+2013-04-19+at+8.41.29+PM.png]

C. Connect the backpack antenna and the Crystal Radio together.

1. Attach the crystal radio to the side pocket of the backpack with a sticky tape.
2 Connect the terminal from the crystal radio to that of the backpack coil.
3. Plug the Crystal headphones into the headphone jack.
4 Place the backpack radio near a window or get to the streets to find an open space.
This backpack radio does not require a ground/earth connection, so you are free to walk on the street.
Align the backpack radio with the direction of the radio transmitting stations, slowly swing the backpack sideways to find the right location for best radio reception. At the same time, turn the knob of the variable capacitor tunes to radio stations. If you are located at a wide open space, you should be able to receive signals from a radio station. And then continue to adjust the angle to find the best reception.
You also can turn the switch on or off to select the low-end or high-end radio stations.

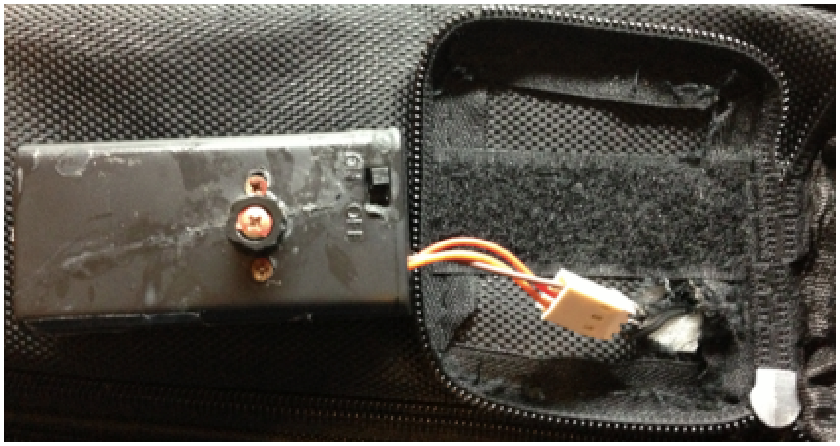

[http://2.bp.blogspot.com/-vagNCSvSIEE/UXFIgz-aC1I/AAAAAAAABC8/IrTlWW9_-TE/s1600/Screen+Shot+2013-04-19+at+9.25.15+PM.png]

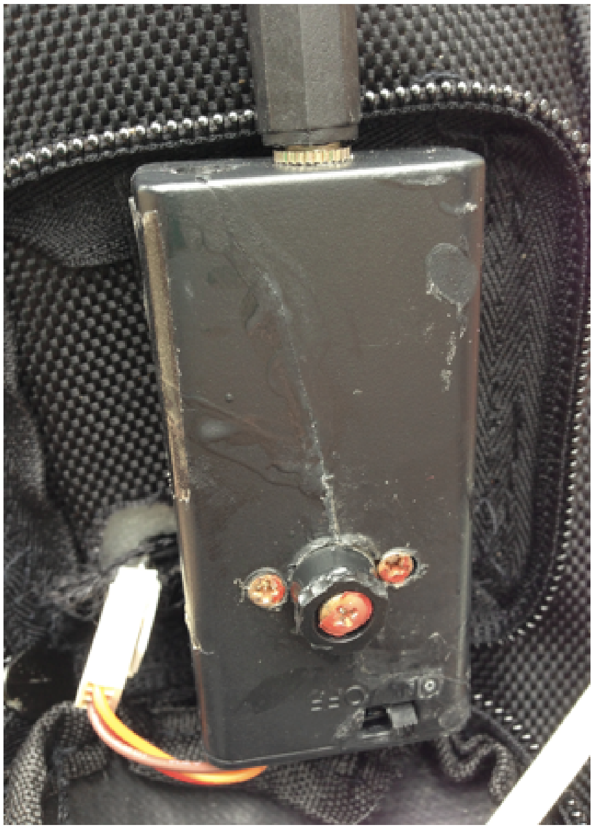

[http://1.bp.blogspot.com/-mUzUqIWXWUA/UXFIgIMJj9I/AAAAAAAABDI/JsSniw1UEhQ/s1600/Screen+Shot+2013-04-19+at+8.42.22+PM.png]

BURGER KING

[http://3.bp.blogspot.com/-AXgfdzjuSI4/UXFIgY6FRzI/AAAAAAAABCw/ThRngIdWwcc/s1600/Screen+Shot+2013-04-19+at+8.43.03+PM.png]

[http://3.bp.blogspot.com/-w9i9n__0Aek/UXFIgj73NtI/AAAAAAAABDA/rPq5YYGZ82M/s1600/Screen+Shot+2013-04-19+at+8.43.26+PM.png]

Posted 19th April 2013 by Billy

1 View comments

12th April 2013

Battery-free Loud Huge Pipe Crystal Radio

Imagine building a huge crystal radio using PVC pipes as the wiring frame that is loud enough to drive the speaker.

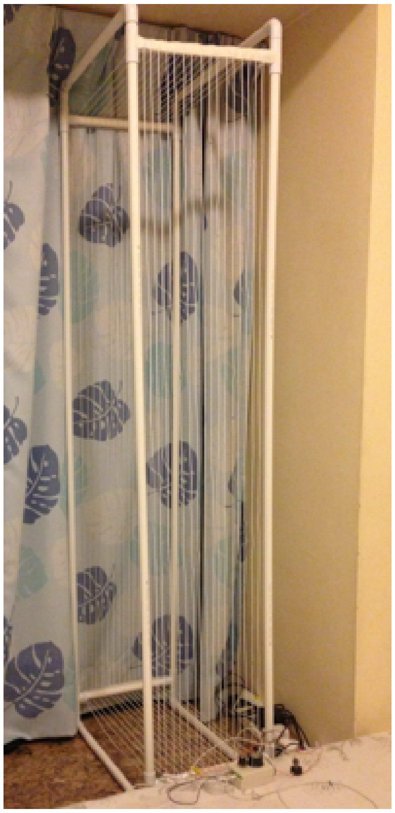

[http://4.bp.blogspot.com/-1keQ1A93uUI/UWrmRcNMhpI/AAAAAAAABAo/0fRnzzDFLQw/s1600/Screen+Shot+2013-04-15+at+1.09.32+AM.png]

I saw this in the discussion forum in crystalradio.net.cn and try to build it with my own design.
This is a huge crystal radio 1.5 meter high, .3 meter wide with the coil wound on the plastic pipes you used to run electric wires into the walls. The pipes are exceptionally cheaper than I thought, only HKD 6 per 3 meters.

Setting up the framework takes one hour, and winding the coil another hour.
After all the hard work, I did it! The audio output level is two times louder than the umbrella radio and you can really drive the speaker. It's so loud that I end up adding a volume control to lower down the sound so I can listen to music while getting to sleep.

Checkout this youtube video:
http://www.youtube.com/watch?v=difYPuzWj4M [http://www.youtube.com/watch?v=difYPuzWj4M]

You can follow the instructions below to make this.

Remember to leave me some comments (at the comment box) at the bottom of this if you have read this blog.

CIRCUIT DIAGRAM

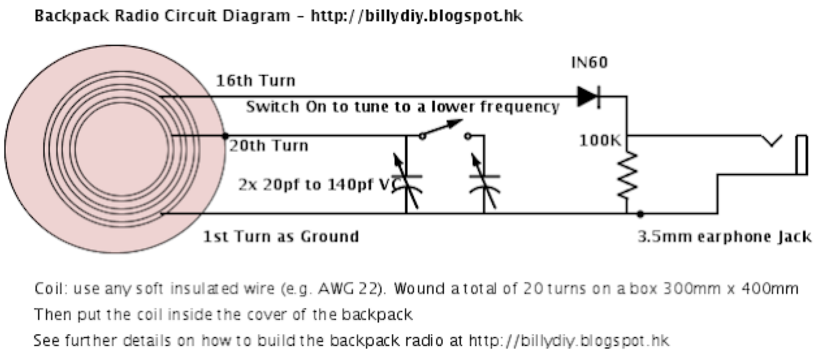

Backpack Radio Circuit Diagram - http://billydiy.blogspot.hk

Coil: use any soft insulated wire (e.g. AWG 22). Wound a total of 20 turns on a box 300mm x 400mm
Then put the coil inside the cover of the backpack
See further details on how to build the backpack radio at http://billydiy.blogspot.hk

[http://1.bp.blogspot.com/-Q3ueJXl4Ak4/UXFIeEDQC-I/AAAAAAAABBs/nzQDabrMuzw/s1600/Screen+Shot+2013-04-19+at+7.28.36+PM.png]

Parts you'll need.

1. Diode IN60, or IN34A
2. Variable Resister 100K type B (for volume control).
3. 30pf to 360pf Variable Capacitor (you can get it from an old radio if you cannot buy it),
4. 2.5mm earphone jack,
5. 3.5 meter of thin wires in three different colours (black for earth, red for antenna, orange for detector)
6. Three crocodile clamps of the same colours.
7. A transparent plastic case for the crystal radio. 60mmx40mmx20mm
8. A crystal earphone. If you do not have it, you can buy it from amazon or e-Bay (around HKD 100 including all the handling fees) or make one yourselves at HKD 10 following the instructions from my other blog posting here:
   How to make a crystal earphone for crystal radios [http://www.blogger.com/blogger.g?blogID=2311644141825422235#editor/target=post;postID=1827613235001197062]

   [http://2.bp.blogspot.com/-ALf15x9ipP0/UV0zuFSOrRI/AAAAAAAAA38/Xf4NU5Wtze4/s1600/Screen+Shot+2013-04-04+at+4.02.38+PM.png]
9. 70 meters of any insulated wires. I used the AWG 22 plastic wire. You can also use AWG 26 which is thinner and lighter. However, the thicker the wire, the more radio wave you capture and the higher sensitivity and higher audio output.

[http://2.bp.blogspot.com/-rYyEBQEY_B4/UVmUKnniVil/AAAAAAAAA0Q/iua6obT96AQ/s1600/IMG_1200.JPG]

You can get most of these from electronic ships in Shamshupo, except the variable capacitor. You can easily buy some old recycled walkman with radios for HKD20-50 from the hawkers in Ip Liu Street in Sham Shui Po. The other way is to go to taobao.com to buy them online and ship to HK. The goods is RMB 5, but the shipping fees is minimum RMD 30, so not worth it unless you buy in great quantity. For the Crystal Earphone, it's very hard to buy nowadays. Refer to my other post to make one yourselves. DIY Crystal Earphone for Crystal Radios [http://billydiy.blogspot.com/2013/03/build-crystal-earphone-for-crystal.html]

From the Hardware store:

10. Buy 3 x 20mm diameter PVC pipes , each 3 meters long - around HK$6 each (ask the shop keeper to cut it in half so you will have 6 x 1.5meter pipes, easier to carry home).
11. 8 x L shape PVC pipe joints for 20mm pipes - around $1.5 each
12. 8 x T shape PVC pipe joints for 20mm pipes - around $1.5 each

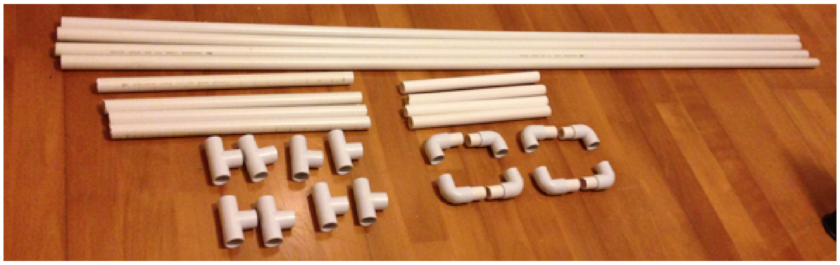

[http://1.bp.blogspot.com/-8MIAYBPNieQ/UWrmUGn8PgI/AAAAAAAABBA/39zUYA-DFrE/s1600/Screen+Shot+2013-04-15+at+1.10.57+AM.png]

TOOLS

1. Solder Iron
2. Saw
3. White Tapes

HOW TO MAKE IT.

A. Create the Hug Pipe Antenna

1. First build the wiring frame:

- Use the saw cut out the pipes into the following pieces:
  - 4 x 1.5 meter pipes
  - 4 x 470mm pipes.
  - 4 x 270mm pipes
  - 8 x 35mm pipes for connecting T-Joints and L-joints together
- connect one 470mm pipe to a L-Joint then to a 35mm pipe then to an T-joint. Then connect the 270mm pipe to the T-Joint.
- Repeat the above to create the base and the top as below.

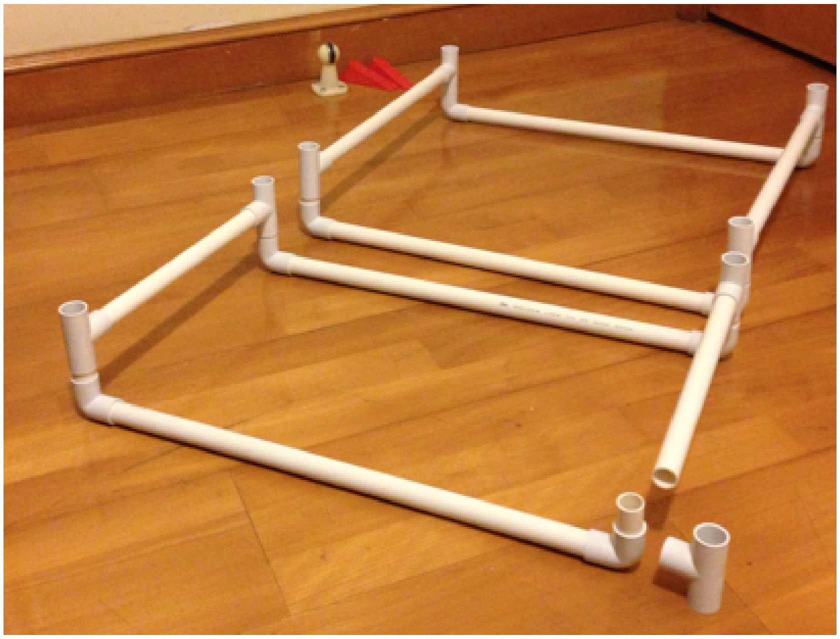

[http://1.bp.blogspot.com/-bh0dkEeMbYY/UWrmUFhfAyI/AAAAAAAABBI/szII001q-Gs/s1600/Screen+Shot+2013-04-15+at+1.11.10+AM.png]

- Now insert the 4 x 1.5 meter pipes into the T-joints (connecting the square base with the square tops).
- Fastern the joins by applying pressure.

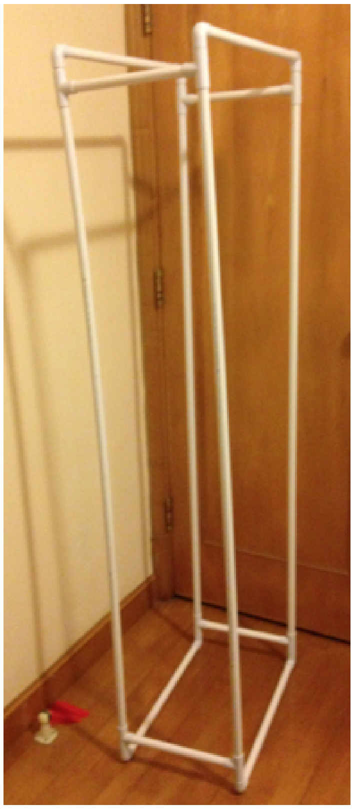

[http://3.bp.blogspot.com/-sj-pnbjhl98/UWrm28ihP5I/AAAAAAAABBQ/9Rq2RzNoQas/s1600/Screen+Shot+2013-04-15+at+1.11.54+AM.png]

2. Put the white tapes onto the 4 edges of the base and the top of the frame. Then put a mark on the tape every 20mm. This is the separation between coils.

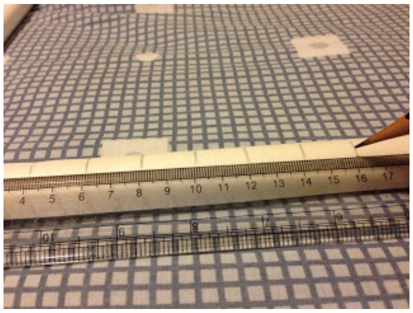

[http://1.bp.blogspot.com/-60fMfQJ6aNk/UWgDgscyt1I/AAAAAAAAA_k/MFIUQpX1B1g/s1600/IMG_1310.JPG]

3. Put another tape onto the first tape, for sticking the wire in place.
4. Start winding the coil. Lift up the outer tape, put the wire it at the places that we marked, and stick the outer tape to hold the wire, then continue the wind all the 17 turns.

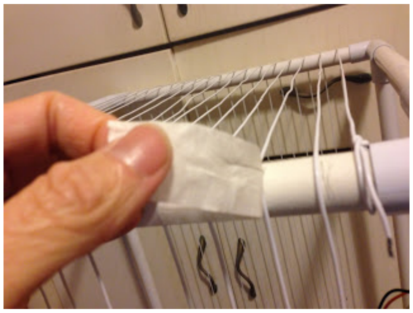

[http://4.bp.blogspot.com/-1LAX7awkscM/UWgDnpkzU5I/AAAAAAAAA_s/IwkKt3quUS4/s1600/IMG_1311.JPG]

5. After winding is done, create the tabs by cutting out the jacket of the wire and expose the copper inside for every turn. At around 80mm above the base.
6. If you have a inductance meter, measure the inductance of the whole coil. It should be close to 300 uh. As shown below, the inductance of the coil is around 294 uH

[http://2.bp.blogspot.com/-1hLmBI9EIf8/UWfwOGp-OxI/AAAAAAAAA-M/eJV6TJ92p1Y/s1600/IMG_1280.JPG]

B. Make the crystal radio (without the coil).

1. Drill holes on the plastic case to fit the earphone jacks, variable capacitor knobs, and the variable resistor knobs, and one more hole to pass the wires for the crocodile clamps, and one more hole for a thin wire to hang the box.
2. Mount the earphone jacks , the variable capacitor and the variable resistor onto the case and fasten the screws.
3. Refer to the circuit diagram and the real life wiring photo, solder the parts together.
4. Connect a black wire to the ground of the Jack and connect the other end to a black crocodile clamp.
5. Connect a white wire to the tip C1 of the variable capacitor.
6. Connect an orange wire to the diode and connect the other end to a yellow crocodile clamp.

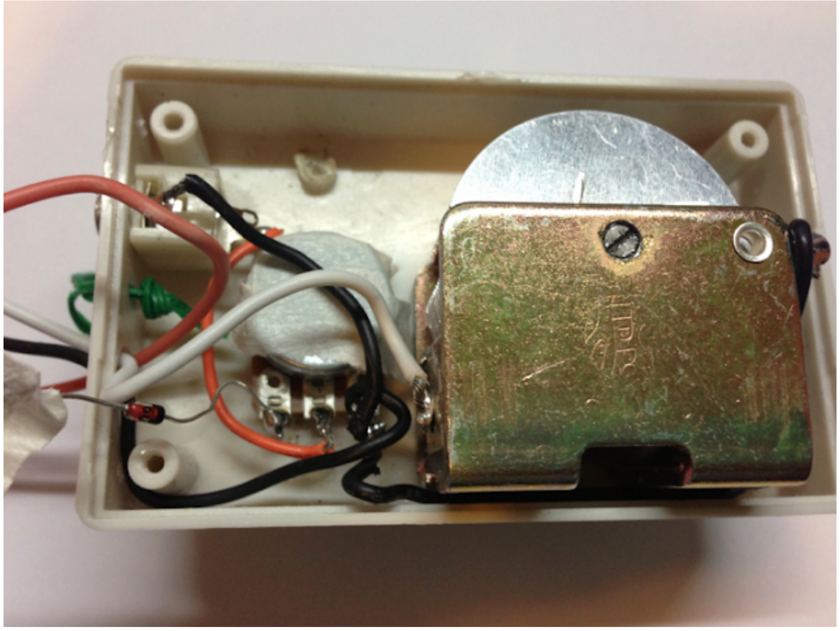

[http://2.bp.blogspot.com/-KB-_5OFCxEg/UWf4MAx0hLI/AAAAAAAAA-w/sKXMpRWu6zk/s1600/Screen+Shot+2013-04-12+at+8.03.34+PM.png]

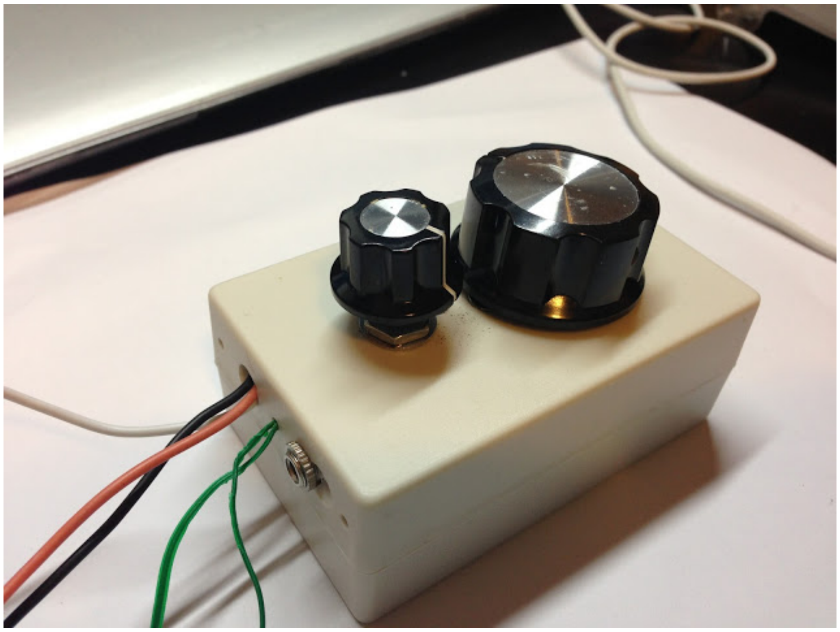

[http://4.bp.blogspot.com/-zpN-V5fYIGk/UWf3ymcjDJI/AAAAAAAAA-o/mTmsqaNg290/s1600/IMG_1301.JPG]

C. Put the Huge Pipe Antenna and the Crystal Radio together.

1. Stand the huge pipe radio next to the window. Use some boxes to raise the frame at least 80mm above any concrete or steel platforms. Otherwise, all the radio signal will be drained to the earth through these concrete or steel frames of your window.
2. Connect black crocodile clamp from crystal radio (for ground) to the tab of the turn 0 (beginning of the coil).
3. Connect the red crocodile clamp (for tuning) to the tab of the last turn (end of the coil).
4. Connect the yellow crocodile clamp (for detector) to the tab of the 14th turn.
5. Plug a crystal earphone to the earphone jack of the radio.
6.

7. Point the huge pipe antenna coil towards the location of the transmitters , or try swing it sideways slowly to find the right spot. At the same time turn the knob of the variable capacitor to tune to a station.
8. If your window is facing a wide open space you should be able to tune to a station. If your window is blocked by another building, then then the signal may be weaken. But keep trying.
9. Once you find the best spot, you can try moving the yellow crocodile clamp up or down to the the other tabs to see if you can get a louder signal. Depending on the actual situation, you may need to make more tabs on the other turns to get the loudest signal. Also, by connecting to different tabs, you can tune into other part of the AM radio frequency to tune into more radio stations.
10. Good Luck and Happy Listening.

Posted 12th April 2013 by Billy

0 Add a comment

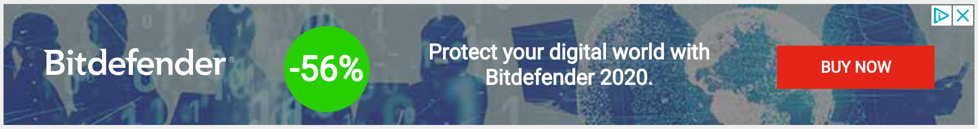
Bitdefender -56% Protect your digital world with Bitdefender 2020. BUY NOW

4th April 2013

Battery-free Hat Crystal Radio

Continuing on with the Umbrella Crystal Radio [http://billydiy.blogspot.hk/2013/04/umbrella-crystal-radio.html] , Billy's DIY Dreamshop brings you the Battery-free Hat Crystal Radio.

Inspired by my friend, and following the similar concept as the umbrella radio, I want to shrink the spiderweb coil from 1 meter in diameter down to 0.3 meter. This will fit nicely on something you can wear, for example a hat or a necklace.

And I did it! Though the audio output level is only half of that of the umbrella radio you can listen clearly to 3 HK local radios from an open area.

Now, you can wear a hat and listen to radio! Just make sure you point your hat at the direction of the radio transmitters see this link. Hong Kong AM Radio Stations and Transmission Locations [http://billydiy.blogspot.hk/2013/04/hong-kong-am-radio-stations-and.html]

You can follow the instructions below to make this.

Remember to leave me some comments (at the comment box) at the bottom of this if you have read this blog.

Dynamic Views theme. Powered by Blogger.

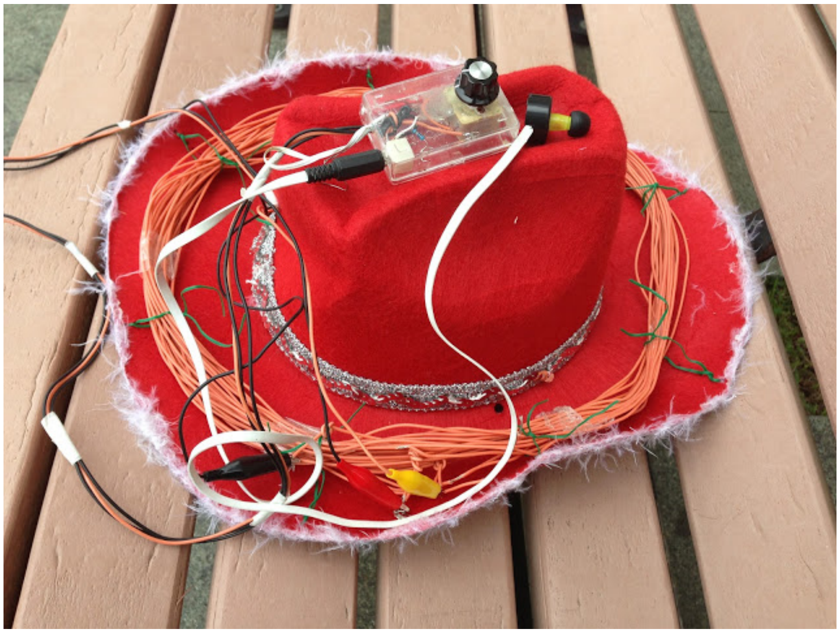

[http://1.bp.blogspot.com/-MQtUDAQ6sT8/UV0h0rXbjLI/AAAAAAAAA2Q/lmBBsVe11Oc/s1600/IMG_1233.JPG]

[http://2.bp.blogspot.com/-mWIF75qeyqc/UV0h93b_m-I/AAAAAAAAA2Y/N2k7vBZbKmI/s1600/IMG_1235.JPG]

CIRCUIT DIAGRAM

Dynamic Views theme. Powered by Blogger.

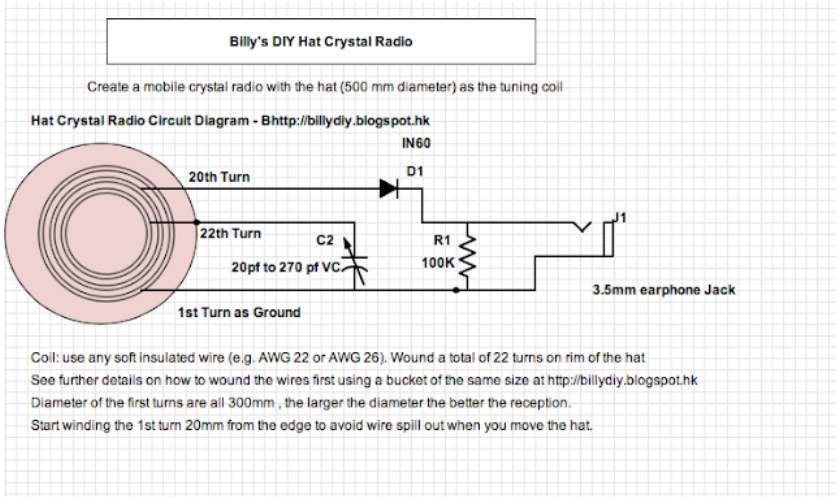

Billy's DIY Hat Crystal Radio

Create a mobile crystal radio with the hat (500 mm diameter) as the tuning coil

Coil: use any soft insulated wire (e.g. AWG 22 or AWG 26). Wound a total of 22 turns on rim of the hat
See further details on how to wound the wires first using a bucket of the same size at http://billydiy.blogspot.hk
Diameter of the first turns are all 300mm , the larger the diameter the better the reception.
Start winding the 1st turn 20mm from the edge to avoid wire spill out when you move the hat.

[http://3.bp.blogspot.com/-CeHfFT2VkjA/UXFXAFbEoFI/AAAAAAAABDg/o53rnpWT9XA/s1600/Screen+Shot+2013-04-19+at+10.38.24+PM.png]

Parts you'll need.

1. A straw hat with a diameter of 500mm or above, the larger the diameter, the louder will be your radio. The following instructions assumes a 500mm-diameter hat. Adjust the length of wires you need accordingly if you have a bigger hat.
2. Any thin wire of any thickness to wound a coil of 22 turns and diameter 300mm around the hat.
3. Diode IN60,
4. Resister 100K,
5. 270pf or 360pf Variable Capacitor (you can get it from an old radio if you cannot buy it),
6. 3.5mm earphone jack,
7. 1.5 meter of thin wires in three different colours (black for earth, red for antenna, orange for detector)
8. Three crocodile clamps of the same colours.
9. A transparent plastic case for the crystal radio. 60mmx40mmx20mm
10. A crystal earphone. If you do not have it, you can buy it from amazon or e-Bay (around HKD 100 including all the handling fees) or make one yourselves at HKD 10 following the instructions from my other blog posting here:
    How to make a crystal earphone for crystal radios [http://www.blogger.com/blogger.g?blogID=2311644141825422235#editor/target=post;postID=1827613235001197062]

    [http://2.bp.blogspot.com/-ALf15x9ipP0/UV0zuFSOrRI/AAAAAAAAA38/Xf4NU5Wtze4/s1600/Screen+Shot+2013-04-04+at+4.02.38+PM.png]
11. Twenty five meters of any wires protected by rubber stealth. I used the AWG 22 plastic wire. You can also use AWG 26 which is thinner and lighter. However, the thicker the wire, the more radio wave you capture and the higher sensitivity and higher audio output.

[http://2.bp.blogspot.com/-rYyEBQEY_B4/UVmUKnniViI/AAAAAAAAA0Q/iua6obT96AQ/s1600/IMG_1200.JPG]

You can get most of these from electronic ships in Shamshupo, except the variable capacitor. You can easily buy some old recycled walkman with radios for HKD20-50 from the hawkers in Ip Liu Street in Sham Shui Po. The other way is to go to taobao.com to buy them online and ship to HK. The goods is RMB 5, but the shipping fees is minimum RMD 30, so not worth it unless you buy in great quantity.
For the Crystal Earphone, it's very hard to buy nowadays. Refer to my other post to make one yourselves. DIY Crystal Earphone for Crystal Radios [http://billydiy.blogspot.com/2013/03/build-crystal-earphone-for-crystal.html]

TOOLS

1. Solder Iron
2. Leather hole puncher for cutting 2mm holes. The paper hole puncher is too big. You can use a nail to cut the hole if you do not have the leather hole puncher.

HOW TO MAKE IT.

A. Create the Hat Antenna

1. First let's create the hat antenna. Use the hole cutter or a sharp object to cut two small holes on the rim of the hat at least 20mm from the edge. This is for tying the wire to hold the coil in place. Repeat this at the five other places around the hat so you can hold the coil firmly later on.

[http://4.bp.blogspot.com/-893XFDdMYdM/UV0k8P5lzoI/AAAAAAAAA2o/SlgSne79_kA/s1600/IMG_1221.JPG]

2. Find a round bucket or rubbish bin around the same size as the coil you are going to wind. Then wind the coil using the thin wire. Total 22 turns. When finish, use the twist tie or wires to tie the coil at the six places to hold it in place and also later on to mount onto hat.

[http://3.bp.blogspot.com/-VfkbZ_4bKJk/UV0mzG47ivI/AAAAAAAAA20/cCv0k944h1k/s1600/IMG_1239.JPG]

3. Mount the coil onto the hat by tying the twist tie or wires onto the six pair of holes you created in step A1.

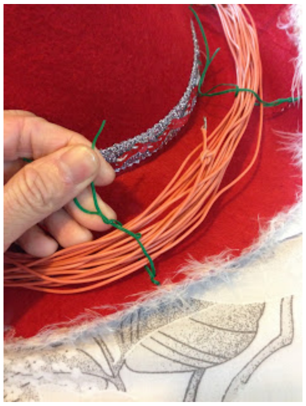

[http://3.bp.blogspot.com/-C-ZN1turigQ/UV0oNC7Q0aI/AAAAAAAAA3E/Zwe--tYlUlQ/s1600/IMG_1240.JPG]

4. Create Tabs for connection to the Crystal Radio set at the following turns.

- Turn 0 - Beginning of the coil
- Turn 18
- Turn 20
- Turn 22 - End of the coil.

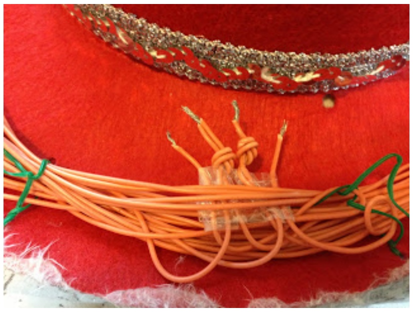

[http://4.bp.blogspot.com/--R93N5a8YKQ/UV0qiDyOeVI/AAAAAAAAA3Y/9DS09uyYsvo/s1600/IMG_1241.JPG]

The inductance of the coil is around 338 uH

[http://1.bp.blogspot.com/-gTuPjJ-sELo/UWLnugq3oDI/AAAAAAAAA9Y/qPrKbJeMfs0/s1600/IMG_1275.jpg]

B. Make the crystal radio (without the coil).

1. Drill holes on the plastic case to fit the earphone jacks, variable capacitor knobs, and one more hole to pass the wires for the crocodile clamps.
2. Mount the earphone jacks and the variable capacitor onto the case and fasten the screws.
3. Refer to the circuit diagram and the real life wiring photo, solder the parts together.
4. Connect a black wire to the ground of the Jack and connect the other end to a black crocodile clamp.
5. Connect a red wire to the tip C1 of the variable capacitor.
6. Connect an orange wire to the diode and connect the other end to a yellow crocodile clamp.

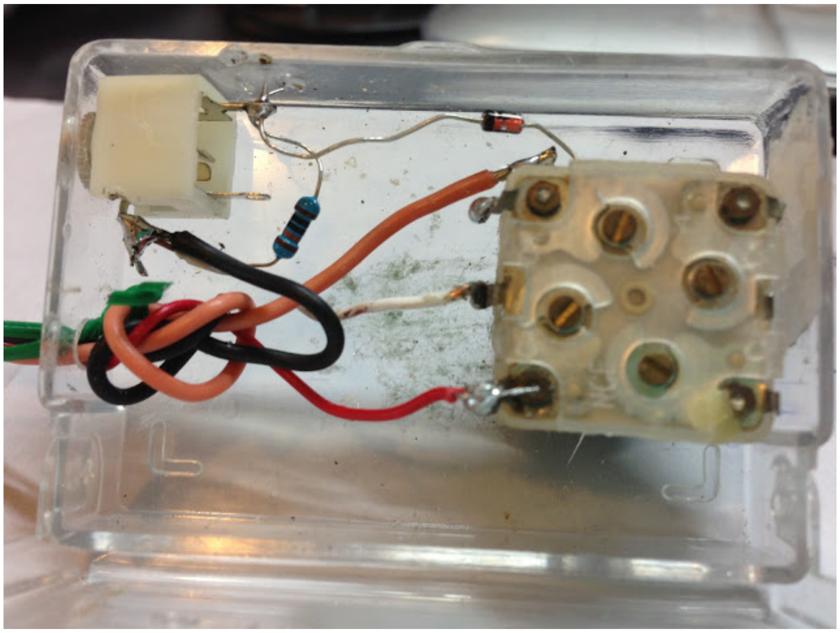

[http://1.bp.blogspot.com/-FsoOXwgYmUI/UV1VCEnXRQI/AAAAAAAAA6Y/76a4H0CYx5Y/s1600/IMG_1260.JPG]

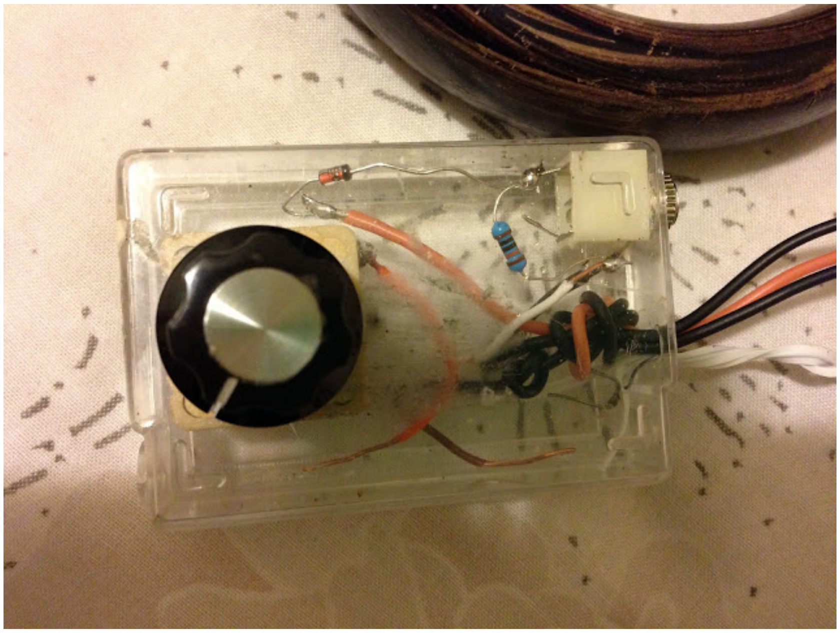

[http://4.bp.blogspot.com/-nygOMUZib2c/UVmffBrMV9I/AAAAAAAAA10/8jQaminvcT8/s1600/IMG_1204.JPG]

C. Put the hat antenna and the crystal radio together.

1. Hang the crystal radio from hat, or on the rim of the hat using a piece of wire to tie it.
2. Connect black crocodile clamp from crystal radio (for ground) to the tab of the turn 0 (beginning of the coil).
3. Connect the red crocodile clamp (for tuning) to the tab of the 22 th turn (end of the coil).
4. Connect the yellow crocodile clamp (for detector) to the tab of the 20th turn.
5. Plug a crystal earphone to the earphone jack of the radio.
6. Bring the hat down to the street and find an empty space, or get close to the window
7. This crystal radio does not require an earth, so you can walk freely on the street
8. Point the hat towards the location of the transmitters , or try swing it sideways slowly, and up and down to find the right spot. At the same time turn the knob of the variable capacitor to tune to a station.
9. If you have a wide open space you should be able to tune to a station. Then continue to adjust the angle of the hat to find the direction with the best reception.
10. Once you find the best spot, you can try moving the yellow crocodile clamp down to the the 18th tabs to see if you can get a louder signal. Depending on the actual situation, you may need to make more tabs on the other turns to get the loudest signal.

Next

I will try to build an FM Crystal radio. There are different circuits being tested.

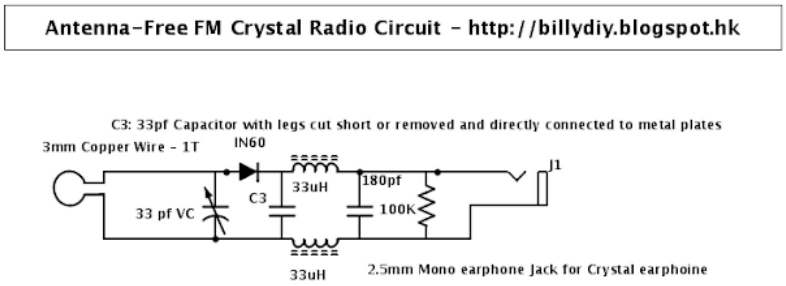
Antenna-Free FM Crystal Radio Circuit - http://billydiy.blogspot.hk
C3: 33pf Capacitor with legs cut short or removed and directly connected to metal plates
3mm Copper Wire - 1T
IN60
33uH
180pf
100K
C3
33 pf VC
J1
33uH
2.5mm Mono earphone Jack for Crystal earphoine

[http://2.bp.blogspot.com/-tAEUrg9BlrM/UV2r9jTn_0I/AAAAAAAAA84/UKfMMMHNDqA/s1600/Screen+Shot+2013-04-05+at+12.33.07+AM.png]

Dynamic Views theme. Powered by Blogger.

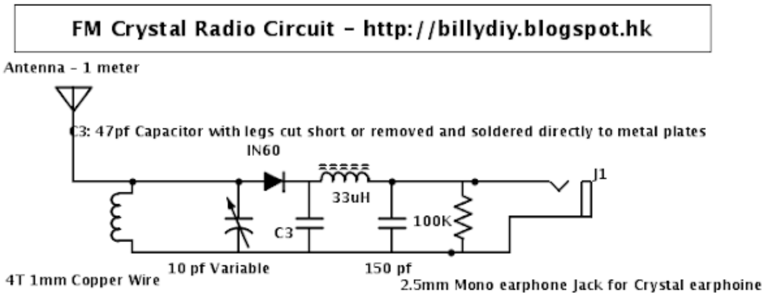
FM Crystal Radio Circuit - http://billydiy.blogspot.hk
Antenna - 1 meter
C3: 47pf Capacitor with legs cut short or removed and soldered directly to metal plates
IN60
33uH
J1
100K
C3
4T 1mm Copper Wire
10 pf Variable
150 pf
2.5mm Mono earphone Jack for Crystal earphoine

[http://2.bp.blogspot.com/-bvY7DJBO8VA/UV2sAKX3O6I/AAAAAAAAA9A/WGWSkc6zAUQ/s1600/Screen+Shot+2013-04-05+at+12.35.26+AM.png]

Posted 4th April 2013 by Billy

0 Add a comment

1st April 2013

Battery-free Umbrella Crystal Radio

Create a battery free, crystal radio using an umbrella as the tuning coil which doubles up as a powerful spiderweb antenna.

You can now walk down the beach with this umbrella crystal radio.
Enjoy sun bathing while listening to the crystal radio!

The umbrella coil can double up as a super strong spiderweb antenna for testing your crystal radio even at home!
Just put the umbrella antenna close to your window and point to the direction of the radio transmission station.

Checkout this Youtube video:
http://www.youtube.com/watch?v=MElL_0IpV4 [http://www.youtube.com/watch?v=MElL_0IpV4g]

I have done one more field test of the radio.

http://www.youtube.com/watch?v=fYvUTIYZ43I

Remember to leave me some comments (at the comment box) at the bottom of this if you have read this blog.

What is a Crystal Radio

Crystal radio is a the simplest radio that runs purely on energy captured from the radio wave and do not require a battery or other forms of electricity supply.
Radio signals are transmitted by the radio stations in HK, e.g. RTHK or Commercial Radio , or the Metro radio, or even from stations across the border from mainland China. See here for details
Hong Kong AM Radio Stations and Transmission Locations [http://billydiy.blogspot.hk/2013/04/hong-kong-am-radio-stations-and.html]

I made crystal radio since I was a teenager. At that time I used to fail, because getting a good antenna set up was and still is a challenge in a concrete jungle like Hong Kong especially if you are living in a high rise apartment; and you are not rich enough to own the roof (i.e. me). Now I can resolve this issue with this umbrella radio that doubles up as a powerful antenna!

Dynamic Views theme. Powered by Blogger.

[http://3.bp.blogspot.com/-Ekdw-4ecSNU/UYpvDTcNzQI/AAAAAAAABIY/dBdn7ASU8ZY/s1600/IMG_1757.JPG]

From Dream to Creation

I notice that there are lots of web postings for mobile crystal radios with huge antenna wound on a Cross made from meter long PVC pipes. This is great. However, the bulky set up is pretty bulky and very hard to transport around.
What if we can replace the cross with a meter-wide umbrella.
Turning on the crystal radio will be as easy as opening an umbrella.
That's how I started the creation of my first umbrella radios. Version 1 of the product (see below) was built in a rush in just 2 hours. It achieve the function to receive many radio stations and the signal is pretty strong too. However, it's built on a extra-large golf umbrella , and the wires are wound on tiny holes at the tip of the umbrella. So, wires tend to be exposed when you close the umbrella.

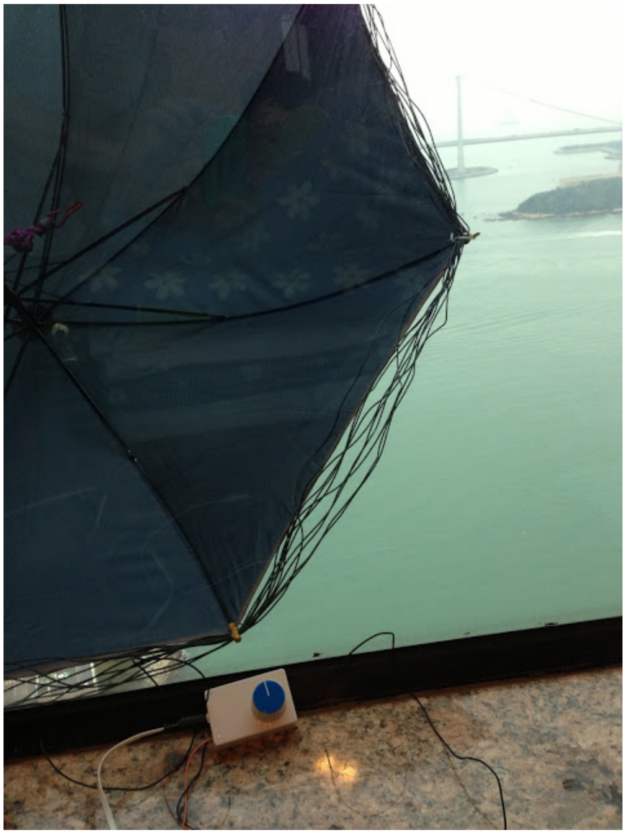

[http://2.bp.blogspot.com/-iUavwbbZZls/UVmpG30vncl/AAAAAAAAA2A/d-5M5s7iUuE/s1600/IMG_1109.JPG]

Umbrella Antenna Version 1.

Then I think of better ways to wind the coil inside the umbrella and created umbrella crystal radio version 2.
So, now even if it's raining (and no lightning), I can open up the umbrella with the coil safely inside, and me happily listening to the musics from the crystal radio. I call it the Umbrella Crystal Radio.

Dynamic Views theme. Powered by Blogger.

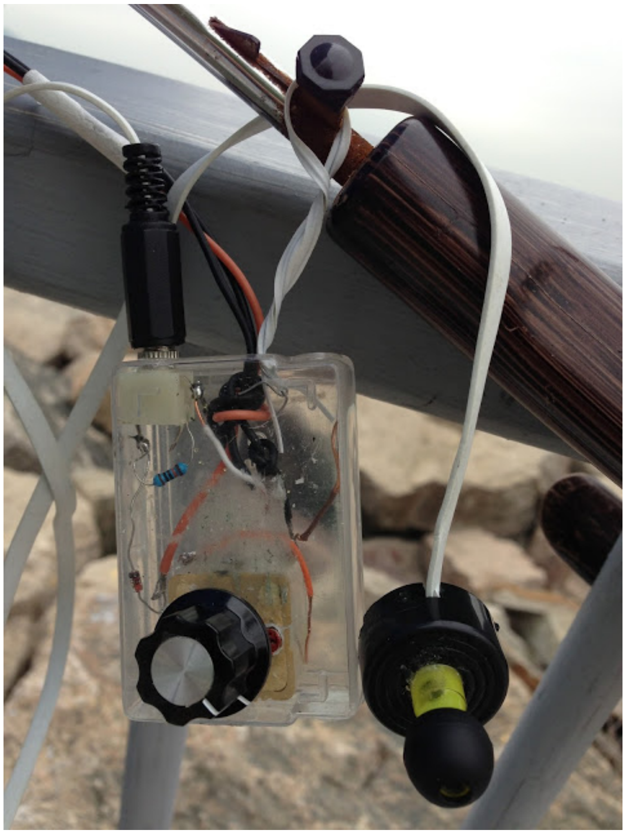

[http://3.bp.blogspot.com/-djZ69vC6Qvc/UVmRXlzSAQI/AAAAAAAAAz0/EAgOuHobJRA/s1600/IMG_1178.JPG]

Mounting the Umbrella radio on the fence, reception is super strong.

In this blog, I will show you how to build this umbrella crystal radio , please follow instructions part A, B, and C.
If you already have an existing crystal radio, you can still use the umbrella as the antenna.
Just follow part A to build the umbrella antenna, then jump to part D to use it as an antenna for your existing crystal radios. Just place the umbrella next to your window and point at the right direction to receive the radio waves.

Circuit Diagram

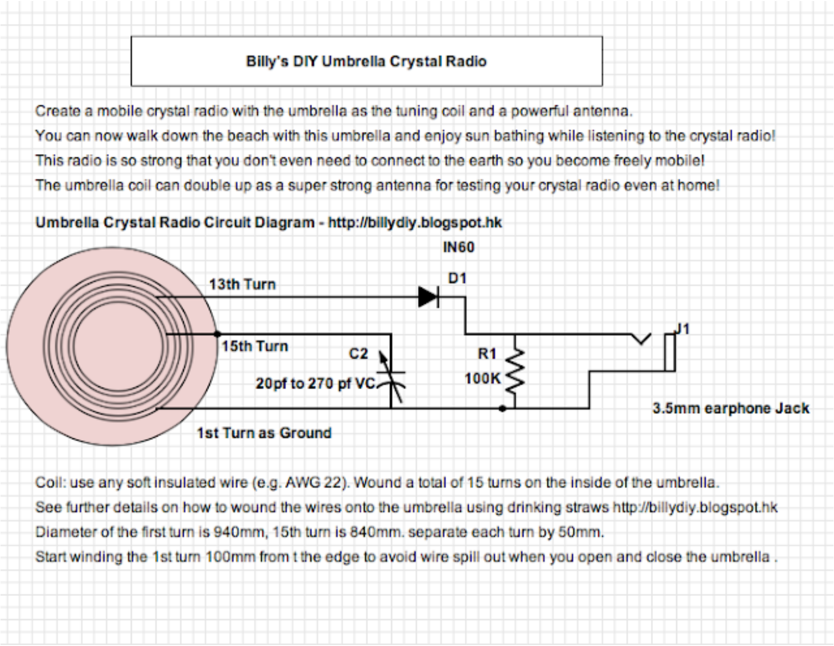

Billy's DIY Umbrella Crystal Radio

Create a mobile crystal radio with the umbrella as the tuning coil and a powerful antenna.
You can now walk down the beach with this umbrella and enjoy sun bathing while listening to the crystal radio!
This radio is so strong that you don't even need to connect to the earth so you become freely mobile!
The umbrella coil can double up as a super strong antenna for testing your crystal radio even at home!

Umbrella Crystal Radio Circuit Diagram - http://billydiy.blogspot.hk

Coil: use any soft insulated wire (e.g. AWG 22). Wound a total of 15 turns on the inside of the umbrella.
See further details on how to wound the wires onto the umbrella using drinking straws http://billydiy.blogspot.hk
Diameter of the first turn is 940mm, 15th turn is 840mm. separate each turn by 50mm.
Start winding the 1st turn 100mm from t the edge to avoid wire spill out when you open and close the umbrella .

19+at+10.41.10+PM.png]

Parts you will need

1. An umbrella around 1 meter in diameter, 8 or more ribs.
The instructions below assumes that you have an umbrella of 8 ribs, and 1 meter in diameter.

You can choose the following types of umbrellas each with a different diameter. The bigger the diameter the better the radio reception, and the louder the radio. However, a regular umbrella of 900mm is good enough.

Diameter of different types of umbrellas (photo below from left to right):

Golf umbrella	1230 mm
Car umbrella	1180 mm
Large umbrella	1070 mm
Regular umbrella	900mm

[http://1.bp.blogspot.com/-vF1exGODXlI/UX0go2x7gpI/AAAAAAAABEE/DOhOlg73ea4/s1600/Screen+Shot+2013-04-28+at+9.12.53+PM.png]

[http://4.bp.blogspot.com/-AYDLYs-zJQo/UX0gohUSqCI/AAAAAAAABEA/zs04fSou9Ss/s1600/Screen+Shot+2013-04-28+at+9.13.13+PM.png]

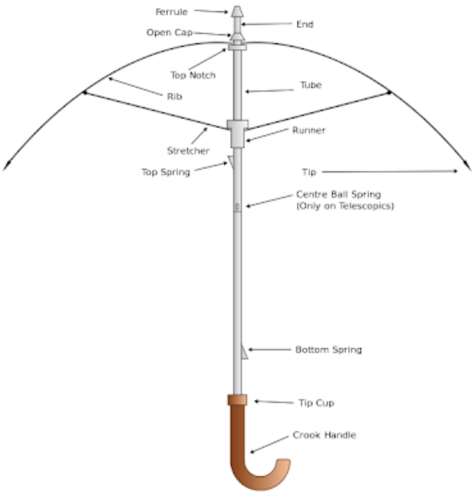
Ferrule
Open Cap
End
Top Notch
Rib
Tube
Runner
Stretcher
Top Spring
Tip
Centre Ball Spring
(Only on Telescopics)
Bottom Spring
Tip Cup
Crook Handle

[http://1.bp.blogspot.com/-u4H0b1n0S7k/UVmK5XZCOGI/AAAAAAAAAzQ/30lQ-EwFZ98/s1600/Screen+Shot+2013-04-01+at+9.14.05+PM.png]

2. 8 or more Drinking Straws of diameter 8mm , those from Mac Donald are perfect. You need as many as the number of ribs your umbrella has. Most umbrella has 8 ribs.

[http://2.bp.blogspot.com/-0mf8l_j4jFo/UVmVZW8yHbI/AAAAAAAAA0Y/ImJKk9N-JUI/s1600/IMG_1201.JPG]

3. Diode IN60,
4. Resister 100K,
5. 270pf or 360pf Variable Capacitor (you can get it from an old radio if you cannot buy it),

6. 3.5mm earphone jack,

7. 1.5 meter of thin wires in three different colours (black for earth, red for antenna, orange for detector)
8. Three crocodile clamps of the same colours.

9. Fifty meters of any regular electric wires. I used the AWG 22 plastic wire. You can also use AWG 26 which is thinner and lighter. However, the thicker the wire, the more radio wave you capture and the higher sensitivity and higher audio output.

PROFESSIONAL WIRE & CABLE
AWG#22
100 FEET
U.S.A. LICENCED
PROFESSIONAL WIRE & CABLE
MIZU
100 FEET

[http://2.bp.blogspot.com/-rYyEBQEY_B4/UVmUKnniViI/AAAAAAAAA0Q/iua6obT96AQ/s1600/IMG_1200.JPG]

Dynamic Views theme. Powered by Blogger.

10. A transparent plastic case for the crystal radio. 60mmx40mmx20mm
11. A crystal earphone. If you do not have it, you can buy it from amazon or e-Bay (around HKD 100 including all the handling fees) or make one yourselves at HKD 10 following the instructions from my other blog posting here: How to make a crystal earphone for crystal radios [http://www.blogger.com/blogger.g?blogID=2311644141825422235#editor/target=post;postID=1827613235001197062]

[http://4.bp.blogspot.com/-ALf15x9ipP0/UV0zuFSOrRI/AAAAAAAAA34/rPr_l0AV14Y/s1600/Screen+Shot+2013-04-04+at+4.02.38+PM.png]

You can get most of these from electronic ships in Shamshupo, except the variable capacitor. You can easily buy some old recycled walkman with radios for HKD20-50 from the hawkers in Ip Liu Street in Sham Shui Po. The other way is to go to taobao.com to buy them online and ship to HK. The goods is RMB 5, but the shipping fees is minimum RMD 30, so not worth it unless you buy in great quantity. For the Crystal Earphone, it's very hard to buy nowadays. Refer to my other post to make one yourselves. DIY Crystal Earphone for Crystal Radios [http://billydiy.blogspot.com/2013/03/build-crystal-earphone-for-crystal.html]

Tools

1. Solder Iron
2. Leather hole puncher for cutting 2mm holes. The paper hole puncher is too big. You can use a nail to cut the hole if you do not have the leather hole puncher.

[http://4.bp.blogspot.com/-0qEaSjjIEOY/UVmW0Lca6LI/AAAAAAAAA0o/ZDfWv6M4EWY/s1600/IMG_1202.JPG]

How to make it.

A. Create the Umbrella Antenna

1. First let's create the umbrella antenna. Use the hole cutter to cut fifteen 2 mm holes on the drinking straw. You may want to mark it first with your ruler before cutting. Start from 40mm, then mark the hole to be cut every 5mm until you reach 115mm.

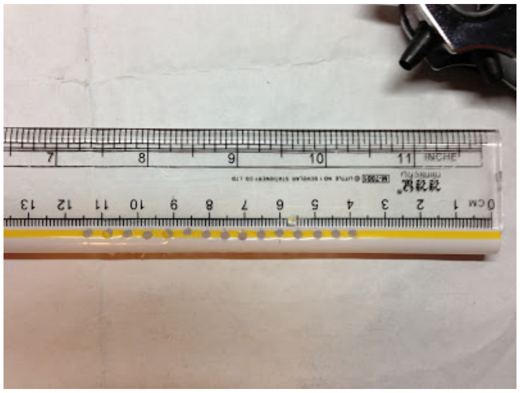

[http://2.bp.blogspot.com/-dLe9r1nIizA/UVmV7CXcN2I/AAAAAAAAA0g/RDf5clSyWRE/s1600/IMG_1161.JPG]

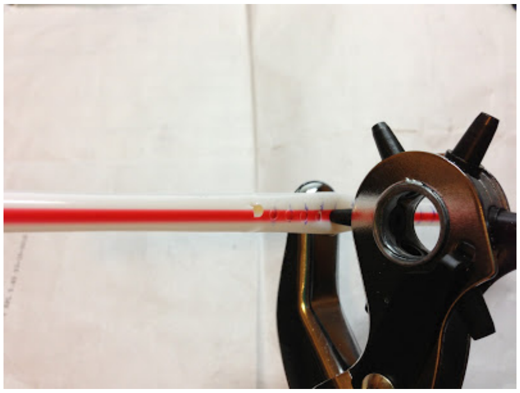

[http://2.bp.blogspot.com/-lxJ5UFcGkPA/UVmXeszhaLI/AAAAAAAAA00/ezbypv5yixE/s1600/IMG_1159.JPG]

2. At each hole cut a slanting slot on the side of the straw so it'll be easier to pull the wire over.

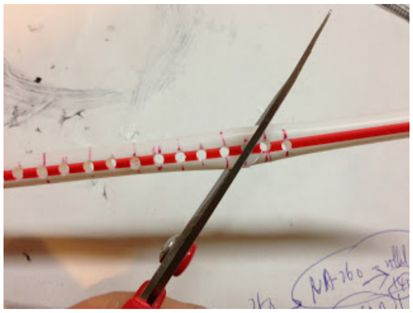

[http://4.bp.blogspot.com/-RGxWhist1V0/UYPQqRjwJjI/AAAAAAAABE0/mS6yv3Vzj_c/s1600/IMG_1697.JPG]

This is how a finished straw looks.

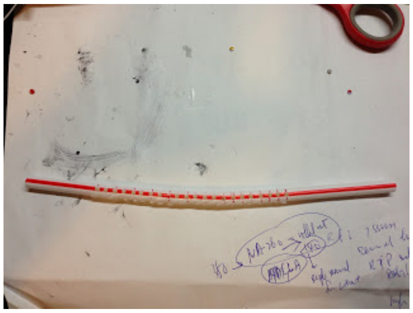

[http://1.bp.blogspot.com/-YdDRP_1dImY/UYPQwKXCcOI/AAAAAAAABE8/sC39Oolj0vg/s1600/IMG_1698.JPG]

3. Remove all 8 tips from the ribs of the umbrella. Insert the straws into the rib, then put the tip back on again.

[http://1.bp.blogspot.com/-cOc0Dk3oKQU/UVmX1I2aZdI/AAAAAAAAA04/-0C0Ycr-3M0/s1600/IMG_1164.JPG]

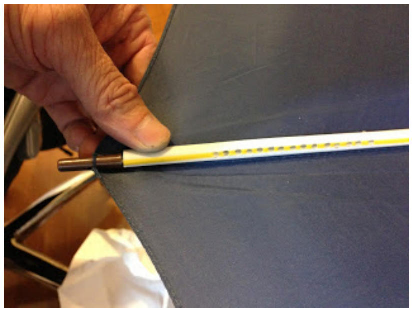

[http://3.bp.blogspot.com/-qH__j-4e3ZY/UVmYCwPvb8I/AAAAAAAAA1A/J27ffuqfTaA/s1600/IMG_1166.JPG]

4. Wind the coil by by pressing the wire in the slanted openings to the holes of the drinking straw. Lock the wire securely into the hole before going for the next turn.

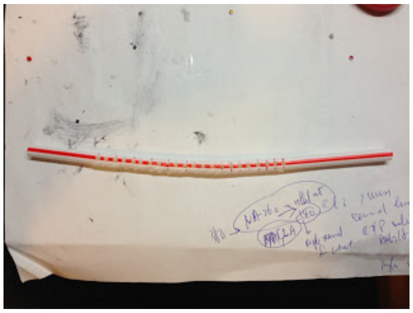

[http://4.bp.blogspot.com/-P0ZJzLBdyOI/UYU-gtQoltI/AAAAAAAABFY/Fxb8kDM63sA/s1600/IMG_1699.JPG]

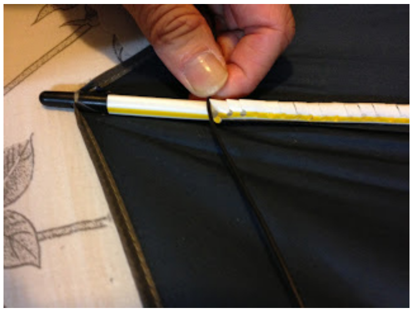

[http://2.bp.blogspot.com/--_nFnE0F08w/UYU-gf7G2KI/AAAAAAAABFQ/MqkMGKTFw-A/s1600/IMG_1700.JPG]

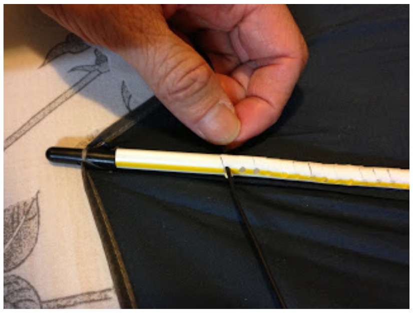

[http://1.bp.blogspot.com/-rvvGjhK4knM/UYU-gt1UkhI/AAAAAAAABFU/ccZAtrcXLR8/s1600/IMG_1701.JPG]

[http://4.bp.blogspot.com/-x1PUd5cN6_4/UYU-h-piWEI/AAAAAAAABFw/bvAD4aEryYY/s1600/IMG_1704.JPG]

[http://3.bp.blogspot.com/-ToDmUd1wJFw/UYU-ivDSBCI/AAAAAAAABF0/QLGfzErEjF0/s1600/IMG_1705.JPG]

[http://1.bp.blogspot.com/-UjOvv-5nryY/UYU-jk-SEIl/AAAAAAAABGQ/-CWkF9Lt1wM/s1600/IMG_1708.JPG]

Dynamic Views theme. Powered by Blogger.

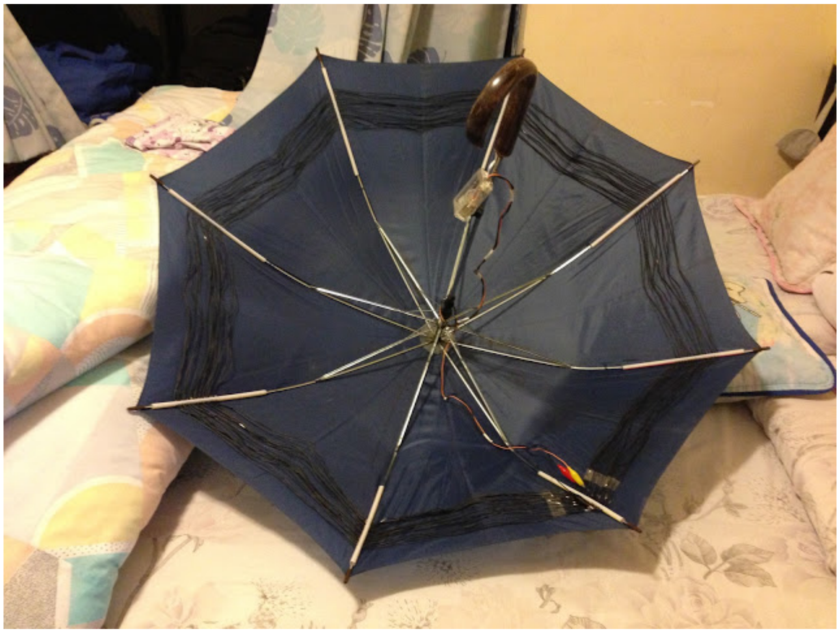

[http://2.bp.blogspot.com/-fqvkB6WE_QQ/UVmZjX_XpKI/AAAAAAAAA1U/gm6HY_su50w/s1600/IMG_1198.JPG]

5. Create the tabs for the coil for tuning by cutting the insulating rubber of the wire to expose the copper inside. To avoid the adjacent wire touching one another, you can create the tabs of the odd number of turns on the left side, and the even number of turns on the right side. If you have a solder iron, put some solder on to the copper to avoid the stealth covering up the copper again.

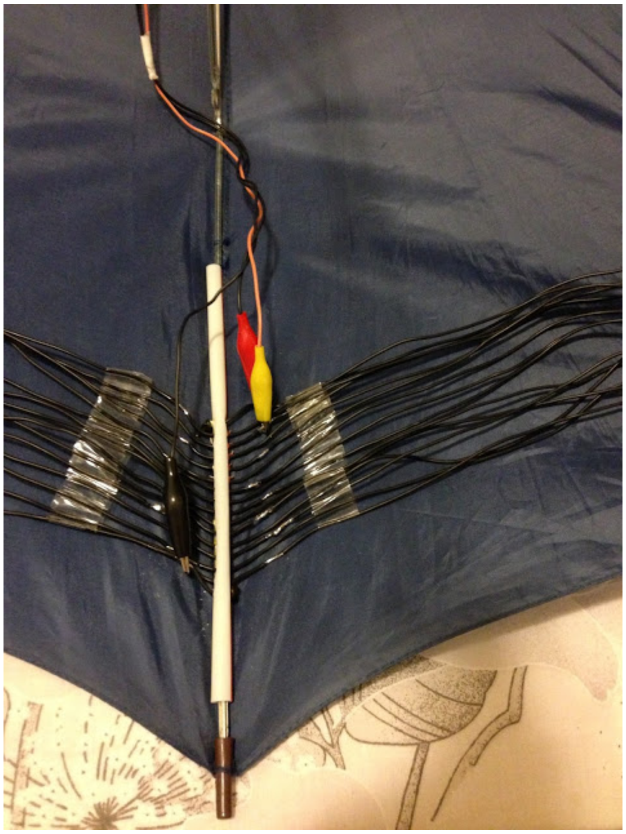

[http://1.bp.blogspot.com/-jDdljBxQlWk/UVmb5RZsppI/AAAAAAAAA1g/DBXQnyD2aIl/s1600/IMG_1199.JPG]

The inductance of the completed umbrella coil should be around 300uH, you can measure it if you have an inductance meter. Otherwise, you can determine if you can tune into most of the radios.
If the inductance is too high, i.e. you cannot tune to stations at the high end (towards 1700 Khz) , remove some turns.
If the inductance is too low, i.e. you cannot tune to stations at the low end (towards to 500 Khz) , add some turns.

6. To avoid the wires slipping through the openings on the straw, apply some glue there to close the gap.

B. Make the crystal radio (without the coil).

1. Drill holes on the plastic case to fit the earphone jacks, variable capacitor knobs, and one more hole to pass the wires for the crocodile clamps.

2. Mount the earphone jacks and the variable capacitor onto the case and fasten the screws.

3. Refer to the circuit diagram and the real life wiring photo, solder the parts together.
Connect a black wire to the ground of the Jack and connect the other end to a black crocodile clamp.
Connect a red wire to the tip C1 of the variable capacitor.
Connect an orange wire to the diode and connect the other end to a yellow crocodile clamp.

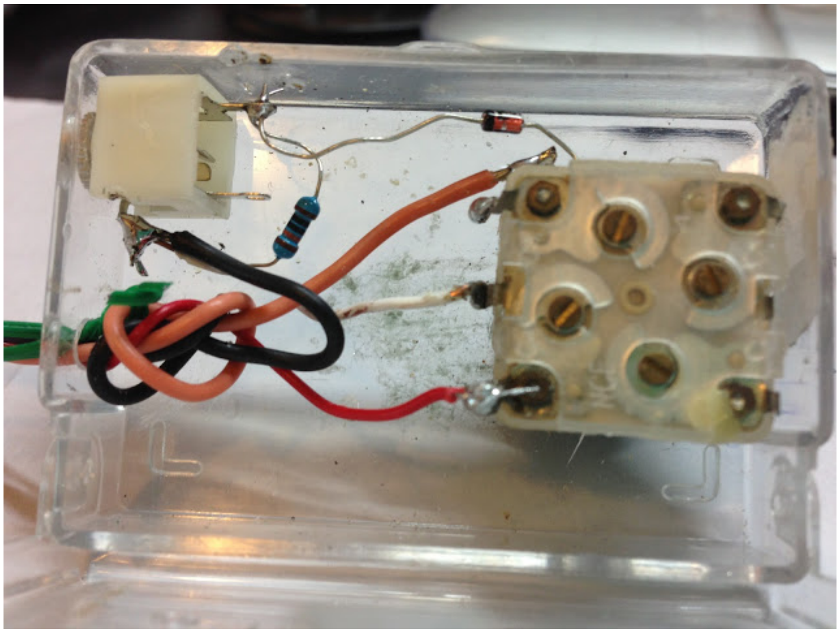

[http://1.bp.blogspot.com/-FsoOXwgYmUI/UV1VCEnXRQI/AAAAAAAAA6Y/76a4H0CYx5Y/s1600/IMG_1260.JPG]

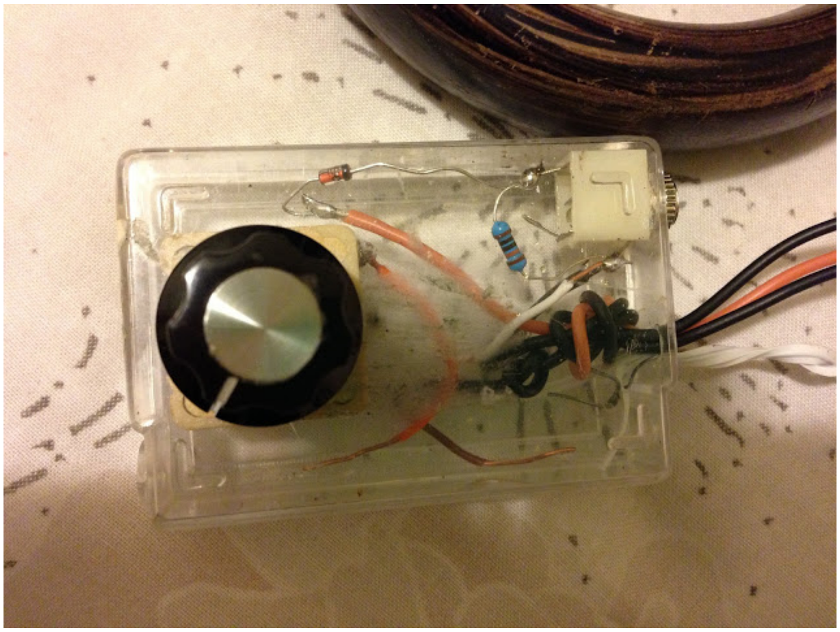

[http://4.bp.blogspot.com/-nygOMUZib2c/UVmffBrMV9I/AAAAAAAAA10/8jQaminvcT8/s1600/IMG_1204.JPG]

C. Put the umbrella antenna and the crystal radio together.

1. Hang the crystal radio from the handle of the umbrella using a piece of wire to tie it.
2. Connect black crocodile clamp from crystal radio (for ground) to the tab of the 1st turn (closest turn to the edge of the umbrella).
3. Connect the red crocodile clamp (for tuning) to the tab of the 15th turn (innermost turn).
4. Connect the yellow crocodile clamp (for detector) to the tab of the 13th turn.
5. Plug a crystal earphone to the earphone jack of the radio.
6. Bring the umbrella down to the street and find an empty space, or get close to the window.
7. This crystal radio does not require an earth, so you can walk freely on the street.
8. Point the umbrella towards the location of the transmitters , or try swing it sideways slowly, and up and down to find the right spot. At the same time turn the knob of the variable capacitor to tune to a station.
9. If you have a wide open space you should be able to tune to a station. Then continue to adjust the angle of the umbrella to find the direction with the best reception.
10. Once you find the best spot, you can try moving the yellow crocodile clamp down to the other tabs to get a louder signal. e.g. Try the 10th Turn or even 7th turn and so on.
11. Some times, the inductance of the umbrella antenna may be too high, you may also want to try moving the RED crocodile clamp down some turns too, e.g. try the 13th turn, or even 10th turn to find the best matching frequency to tune to a station.

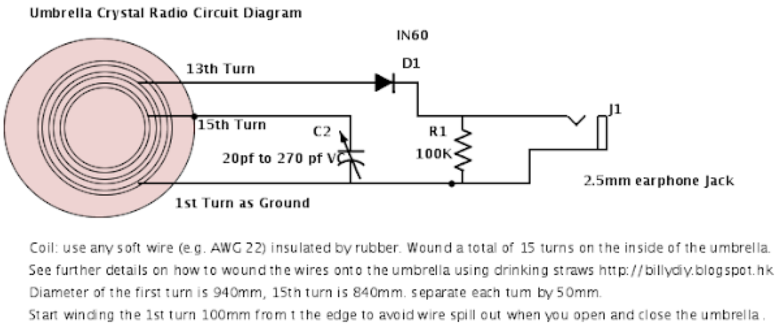
Umbrella Crystal Radio Circuit Diagram

Coil: use any soft wire (e.g. AWG 22) insulated by rubber. Wound a total of 15 turns on the inside of the umbrella.
See further details on how to wound the wires onto the umbrella using drinking straws http://billydiy.blogspot.hk
Diameter of the first turn is 940mm, 15th turn is 840mm. separate each turn by 50mm.
Start winding the 1st turn 100mm from t the edge to avoid wire spill out when you open and close the umbrella .

[http://2.bp.blogspot.com/-CJMXABv3a7o/UV1dfnp3f-I/AAAAAAAAA7A/yZsq0oQgq_w/s1600/Screen+Shot+2013-04-04+at+3.58.08+PM.png]

D. Use the umbrella antenna for your existing crystal radio.

1. Hang your existing crystal radio from the handle of the umbrella using a piece of wire to tie it.
2. Using a crocodile clamp, connect the crystal radio's antenna wire to the tab of the 1st turn (outermost turn).
3. If your crystal radio also need an earth connection, connect the wire to earth. Or if you want to try it on the street , hold the earth wire of your crystal radio by your bare hand to create an earth. Or, you can connect the earth wire to some metal poles , fences or benches that touches the ground,
4. Plug a crystal earphone to the earphone jack of the radio.
5. Bring the umbrella down to the street and find an empty space, or get close to the window.
6. Point the umbrella towards the location of the transmitters , or try swing it sideways slowly, and up and down to find the right spot. At the same time turn the knob of the variable capacitor to tune to a station.
7. If you have a wide open space you should be able to tune to a station. Then continue to adjust the angle of the umbrella to find the direction with the best reception.
8. Once you find the best spot, you can try moving the yellow crocodile clamp down to the other tabs to get a louder signal. e.g. Try the 10th Turn or even 7th turn and so on.
9. Some times, the inductance of the umbrella antenna may be too high, you may also want to try moving the RED crocodile clamp down some turns too, e.g. try the 13th turn, or even 10th turn to find the best matching frequency to tune to a station.

Dynamic Views theme. Powered by Blogger.

Testing Result

I tested this umbrella radio next to the sea. The radio is very loud. and recorded a record high DC current of 44uA. A regular crystal radio at home will only get around 4uA.

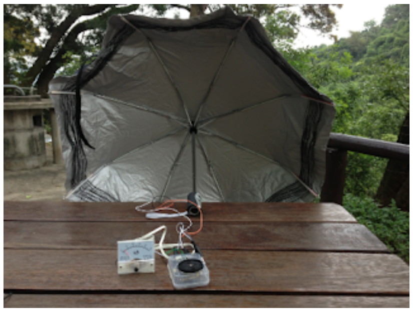

[http://3.bp.blogspot.com/-63NLY_TAJWY/UX09Jhk1paI/AAAAAAAABEY/PsXXPiRoDMU/s1600/Screen+Shot+2013-04-28+at+10.53.12+PM.png]

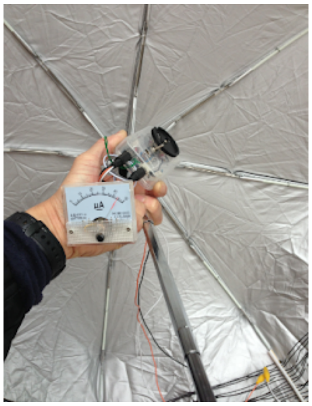

[http://1.bp.blogspot.com/-BkXGrBIBhiY/UX09LvGfKFI/AAAAAAAABEg/dneyDudJ53Y/s1600/Screen+Shot+2013-04-28+at+10.50.27+PM.png]

What's Next?

I also made this super large golf umbrella radio in much the same way.
Only difference is the umbrella size is 1180mm in diameter.

Délifrance

[http://3.bp.blogspot.com/-9J1rVIxxWCY/UYjh9pGQwEI/AAAAAAAABII/0GBZivDqNtM/s1600/IMG_1757.JPG]

[http://2.bp.blogspot.com/-Ztn-L_ziDQM/UYjh8uw-6uI/AAAAAAAABH4/iApibP0fdr4/s1600/IMG_1755.JPG]

[http://3.bp.blogspot.com/-Sb1gpUya6Ew/UYjh818kM1I/AAAAAAAABH8/R7y17TVf9cE/s1600/IMG_1756.JPG]

[http://3.bp.blogspot.com/-l57uvwJrDSk/UYjh7eCjhCI/AAAAAAAABHk/coUls9-0UBE/s1600/IMG_1751.JPG]

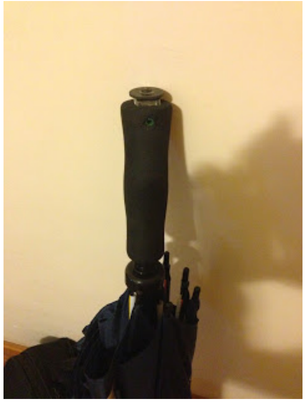

[http://1.bp.blogspot.com/-YaSHe0j_LXU/UYjh7W2CduI/AAAAAAAABHc/kczzaiQrsN4/s1600/IMG_1752.JPG]

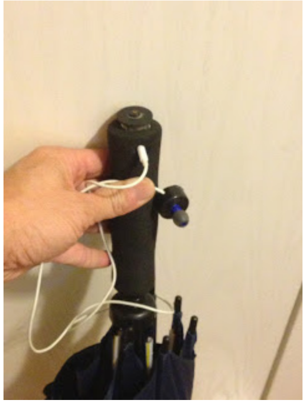

[http://1.bp.blogspot.com/-I5brWLq7y3I/UYjh7_LkGBI/AAAAAAAABHo/WwYI8Dxh1So/s1600/IMG_1753.JPG]

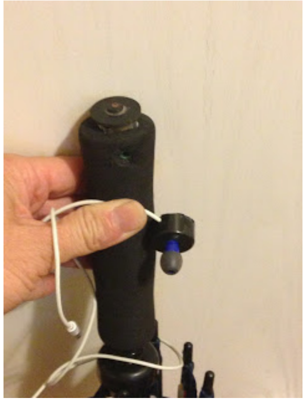

[http://2.bp.blogspot.com/-e7oVdxICzW4/UYjh8jo8rCI/AAAAAAAABIA/npMLfxZpNwo/s1600/IMG_1754.JPG]

[http://1.bp.blogspot.com/-qzwgpOMzrmU/UYjh6TTNCwI/AAAAAAAABHQ/BpIWZCGY6Vw/s1600/IMG_1750.JPG]

Dynamic Views theme. Powered by Blogger.

[http://4.bp.blogspot.com/-xI1qsofvDBc/UYjh5JSTWWI/AAAAAAAABG0/3hAtS-lBr1U/s1600/IMG_1745.JPG]

[http://4.bp.blogspot.com/-_w5Ohp5JgbE/UYjh5P6WFEI/AAAAAAAABG4/FXsEHD5EhTY/s1600/IMG_1746.JPG]

[http://1.bp.blogspot.com/-8v9N5j7h2OU/UYjh5IfWN2I/AAAAAAAABG8/xEhTx3e0eJk/s1600/IMG_1747.JPG]

[http://2.bp.blogspot.com/-yls4zhs3ptQ/UYjh6jyFaeI/AAAAAAAABHM/3pWsnknSOrk/s1600/IMG_1748.JPG]

Dynamic Views theme. Powered by Blogger.

[http://1.bp.blogspot.com/-dz-BC2lbCxQ/UYjh6HhuJbI/AAAAAAAABHI/BTChkCplYaI/s1600/IMG_1749.JPG]

Posted 1st April 2013 by Billy

0 Add a comment

1st April 2013

Hong Kong AM FM Radio Frequencies and Transmitter Locations

Crystal Radios is perfect for AM and recently crystal radio can also cover FM in locations close to the FM transmitting locations, or with FM radio antenna and FM radio signal amplifiers.

Here is a web site that listed all radio broadcast stations in the world.

http://www.asiawaves.net/hong-kong-radio.htm

Radio Broadcast Stations in Hong Kong

The following information is extracted from a web site that listed all radio broadcast stations in the world. Please refer this in detail.

http://www.asiawaves.net/hong-kong-radio.htm

Information below is © Copyright Alan G. Davies 1999-2009, all rights reserved.

I have extracted the info for Hong Kong.

Radio Television Hong Kong (RTHK) [http://www.rthk.org.hk/]

Broadcasting House
30 Broadcast Drive
Kowloon
Hong Kong

Tel: +852 2339 6300
Fax: +852 2336 9314

- RTHK Radio 1: in Cantonese / Mandarin on 92.6-94.4 MHz
- RTHK Radio 2: in Cantonese on 94.8-96.9 MHz

- RTHK Radio 4: in English & Cantonese on 97.6-98.9 MHz
- RTHK Radio 5: in Cantonese / Mandarin on 783 kHz and 92.3, 99.4 & 106.8 MHz
- RTHK Radio 6: relays BBC World Service in English 24h on 675 kHz
- RTHK Radio 7 (Putonghua Channel): in Mandarin on 621 kHz and 100.9 & 103.3 MHz

All networks operate 24h, but some may carry merged programming at times during local evenings and overnight.

Hong Kong Commercial Broadcasting Company Ltd (HKCR) [http://www.crhk.com.hk/]

3 Broadcast Drive
Kowloon
Hong Kong

Tel: +852 2336 5111
Fax: +852 2338 0021

- HKCR CR1 (Supercharged 881): in Cantonese on 88.1-89.5 MHz
- HKCR CR2 (Ultimate 903): in Cantonese on 90.3-92.1 MHz
- AM 864 in English and limited Nepalese on 864 kHz

All networks operate 24h

Metro Broadcast Corporation Ltd [http://www.metroradio.com.hk/]

Basement 2
Site 6, Whampoa Gardens
Hunghom
Kowloon
Hong Kong

Tel: +852 2123 9888
Fax: +852 2123 9877

- Metro Info: in Cantonese on 99.7-102.1 MHz
- Metro Finance: in Cantonese on 102.4-106.3 MHz
- Metro Plus: in English, Mandarin, Filipino, Hindi and Indonesian on 1044 kHz AM Stereo

Mediumwave (AM) Radio Frequencies in Hong Kong

Mediumwave (AM) Radio Frequencies in Hong Kong

Frequency (kHz)	Network	Site	Power (kW)
567 kHz	RTHK3	Golden Hill	20 kW
621 kHz	RTHK Putonghua Channel	Golden Hill	20 kW
675 kHz	RTHK6 (BBC World Service)	Peng Chau	1 kW
783 kHz	RTHK5	Golden Hill	20 kW
864 kHz	HKCR AM864	Peng Chau	10 kW
1044 kHz	Metro Plus	Peng Chau	10 kW
1584 kHz	RTHK3	Chung Hom Kok	0.1 kW

FM Radio Frequencies in Hong Kong

(Frequencies in MHz; TRP = transmitter power, ERP = maximum effective radiated power)

Target Area	Site	RTHK1	RTHK2	RTHK4	HKCR1 (Supercharged 881)	HKCR2 (Ultimate 903)	Metro Info	Metro Finance	TRP (kW)	ERP (kW)
Fan Ling	Cloudy Hill	93.2	95.3	97.8	88.3	90.7	100	104.7		0.5 kW

Hong Kong Island (North), Sai Kung	Kowloon Peak	94.4	96.9	98.9	89.5	92.1	101.8	106.30.5 kW	1 kW
Hong Kong Island (South)	Lamma Island	93.6	96	98.2	89.1	91.6	102.1	104.50.3 kW	0.5 kW
Kowloon	Mt Gough	92.6	94.8	97.6	88.1	90.3	99.7	1041 kW	3 kW
Sha Tin	Beacon Hill	93.5	96.3	98.1	89.2	91.1	100.5	102.40.1 kW	0.15 kW
Tsuen Wan	Golden Hill	92.9	95.6	98.4	88.9	90.9	101.6	105.5	0.1 kW
Tuen Mun	Castle Peak	93.4	96.4	98.7	88.6	91.2	100.4	102.50.5 kW	0.7 kW

Additional local coverage on FM - frequencies in MHz, transmitter power (TRP) in kW:

- RTHK3: Mount Nicholson 97.9 MHz 0.02 kW
- RTHK3: Chung Hom Kok 106.8 MHz 0.06 kW
- RTHK3: Tin Shui Wai 107.8 MHz 0.0375 kW
- RTHK3: Tseung Kwan O [Junk Bay] 107.8 MHz 0.015 kW
- RTHK5: Tin Shui Wai 92.3 MHz 0.0375 kW
- RTHK5: Tseung Kwan O [Junk Bay] 99.4 MHz 0.015 kW
- RTHK5: Castle Peak 106.8 MHz 0.01 kW
- RTHK Putonghua Channel: Castle Peak 100.9 MHz 0.003 kW
- RTHK Putonghua Channel: Tai Hang Road 100.9 MHz 0.005 kW
- RTHK Putonghua Channel: Tin Shui Wai 103.3 MHz 0.0375 kW
- RTHK Putonghua Channel: Tseung Kwan O [Junk Bay] 103.3 MHz 0.015 kW
- Metro Info: Red Hill 101.0 MHz 0.005 kW
- Metro Finance: Red Hill 102.6 MHz 0.005 kW

Symphony of Lights:Narration accompanying the light show at 1200 UTC daily is transmitted for the Victoria Harbour area on 103.4 MHz (English), 106.8 MHz (Cantonese) and 107.8 MHz (Mandarin).

List of HK AM / MW RADIO Stations

567Khz 20kW Radio 3 (RTHK) [http://www.rthk.org.hk/channel/radio3/] |also on FM - **Golden Hill** music variety/info English
621Khz 20kW Putonghua Channel (RTHK7) [http://www.rthk.org.hk/rthk/putonghua.htm] | also on FM - **Golden Hill** Tsuen Wan news/oldies
675Khz 10kW Radio 6 (RTHK/BBCWS) [http://www.rthk.org.hk/channel/radio6/] |also on FM - **Peng Chau** news, relays BBCWS English
747Khz **Radio Zhongshan AM Radio** [http://www.zhongshanradio.com/zh-CN/introduce001.html] - **Zhongshan** (Guangdong) CN
783Khz 20kW Radio 5 (RTHK) [http://www.rthk.org.hk/channel/radio5/] |also on FM - **Golden Hill** educational/cultural Cantonese
864Khz 10KW HKCR AM864 -= Peng Chau - Fri/Sat - Philippines, others 24 hours songs.
942Khz Shenzhen Radio Love FM [http://www.szr.com.cn/szmg_broadcast/index.do?a=942] - **Shenzhen** (Guangdong) CN traffic
1044Khz 10kW Metro Plus (Metro) [http://www.metroradio.com.hk/] - **Peng Chau** in English, Mandarin, Filipino, Hindi and Indonesian on 1044 kHz AM Stereo

1584Khz 0.1KW Radio 3 (RTHK) [http://www.rthk.org.hk/channel/radio3/] |rep - **Chung Horm Kok** music variety/info English

Hong Kong Radio Transmitter Locations

Dynamic Views theme. Powered by Blogger.

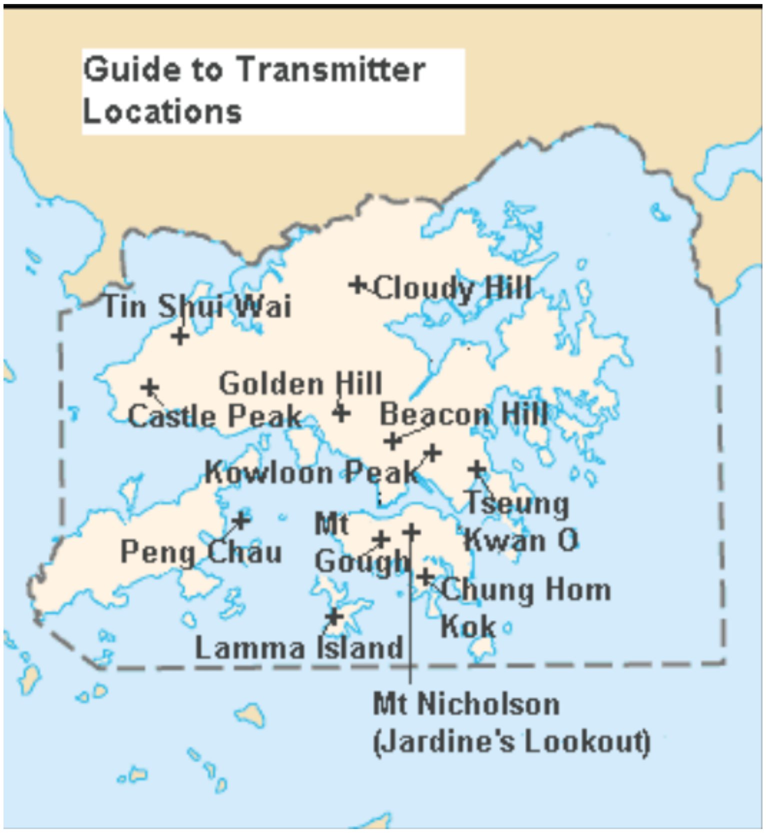
Guide to Transmitter Locations
Tin Shui Wai
Cloudy Hill
Golden Hill
Castle Peak
Beacon Hill
Kowloon Peak
Tseung Kwan O
Peng Chau
Mt Gough
Chung Hom Kok
Lamma Island
Mt Nicholson (Jardine's Lookout)

[http://2.bp.blogspot.com/-cotm0IPjw0I/UVl3c4TOEQI/AAAAAAAAAyw/P5Y3GTAhSNM/s1600/Screen+Shot+2013-04-01+at+7.59.47+PM.png]

Posted 1st April 2013 by Billy

0 Add a comment

Bitdefender -56% Be aware without a care! BUY NOW

27th March 2013

DIY Crystal Earphone for Crystal Radios

[http://4.bp.blogspot.com/-W1q6NTSXpuU/UVLilrs6J_I/AAAAAAAAAx0/CAvQFcxXxCg/s1600/Cyrstal+Earphone.png]

What is a Crystal Earphone

Crystal earphone also called Ceramic earphone/earpiece is the soul of crystal radios. They are one of a few types of earphones that are sensitive enough to produce sound from a Crystal Radio. Regular HIFI headphones will not work with a crystal radio. The Crystal radio runs without battery. It runs purely by the radio wave captured by the antenna and convert that into tiny electric current (< 1uA) that is yet sufficient to produce sound through the highly sensitive crystal earphone.

Wikipedia definition of Crystal Earphone [http://en.wikipedia.org/wiki/Crystal_earpiece]

[http://2.bp.blogspot.com/-ALf15x9ipP0/UV0zuFSOrRI/AAAAAAAAA38/Xf4NU5Wtze4/s1600/Screen+Shot+2013-04-04+at+4.02.38+PM.png]

The original crystal earphone (shown above) cannot be purchased in Hong Kong since 10 years ago. You can still buy it from UK through the online shopping sites, e.g. Amazon, or e-Bay, the cost is not too expansive, but the shipping cost alone will kill this idea. The total cost to buy one oversea may exceed HKD 120, for something valued around HKD 30 max.

SPECIFICATIONS of a Crystal Earphone:

- Impedance: 20K ohm
- Capacity: 15,000 pf
- Sensitivity: 57 dB at 1KHz
- Frequency Range: 200 ~ 8000 Hz

However, with some readily available components you can buy from HK Shamshuipo, Ip Liu Street, you make one yourselves at HKD 10 following the instructions here.

Remember to leave me some comments (at the comment box) at the bottom of this if you have read this blog.

Parts you'll need.

1. Piezo-Electric Buzzer (HK$ 7).
2. Cut out 20mm of the tip from a used plastic ball pen (HK$0).

3. A 2.5mm head Jack & wire from a used/broken headphone (HK$0. If you don't have that, you can also buy some spare wires and a 2.5mm mono headphone Jack (the one in black below) (HK$ 3).
4. A rubber earbud cousin from existing earbuds (top right on the picture below).

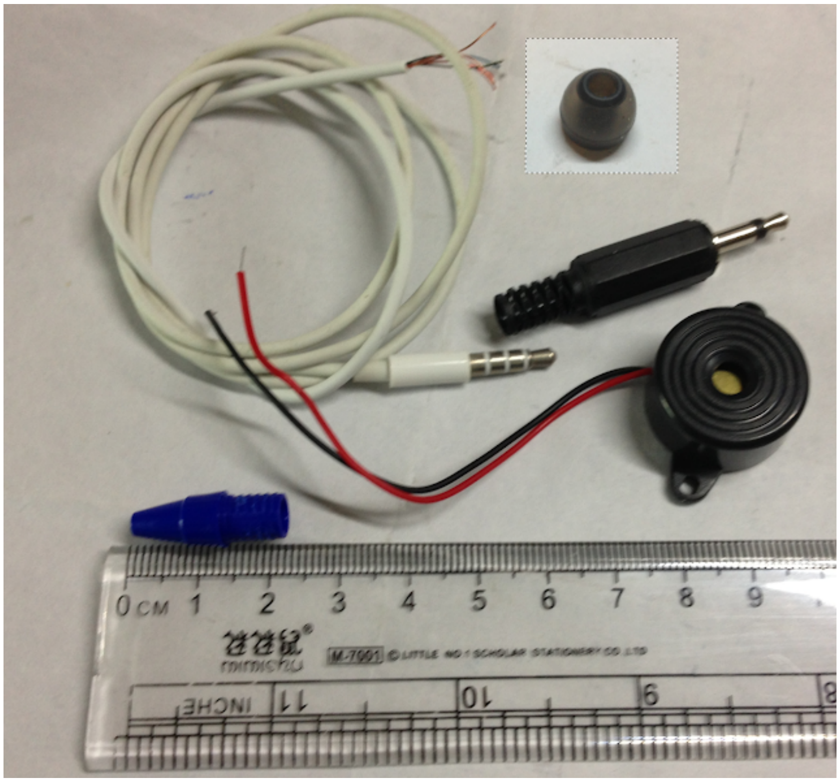

[http://4.bp.blogspot.com/-W5ExpIIP2L0/UV1aUnwMazI/AAAAAAAAA6o/Hk0QoJe3pyQ/s1600/Screen+Shot+2013-04-04+at+6.46.24+PM.png]

How to make it.

Watch this video or follow the instructions below:
http://www.youtube.com/watch?v=DjCKVAMJGlo [http://www.youtube.com/watch?v=DjCKVAMJGlo]

1. Using a screwdriver and brutal force, remove the lid of the buzzer

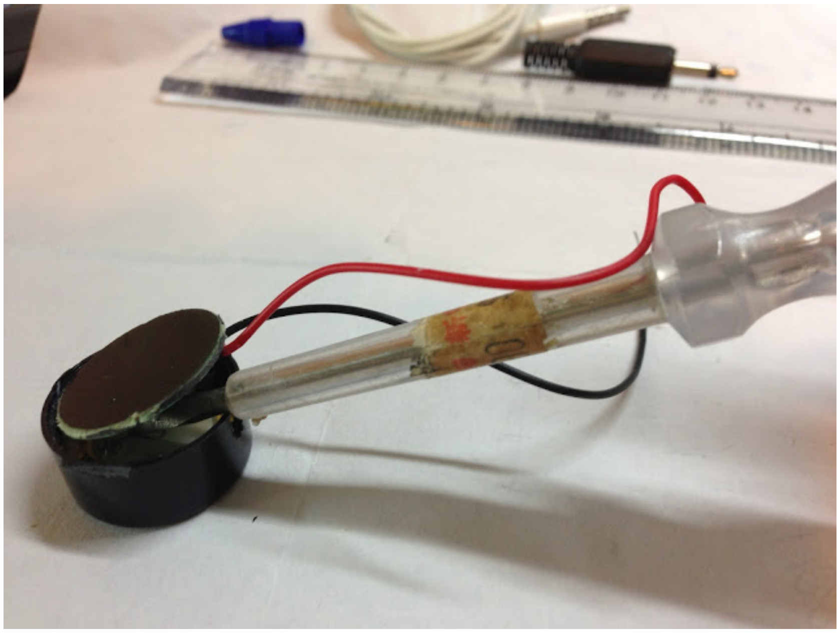

[http://4.bp.blogspot.com/-7mKdE9pSAKE/UV1U9zeAK8I/AAAAAAAAA5M/BoCX_P18bT8/s1600/IMG_1246.JPG]

2. Find the Piezo-electric plate (the circle plate in gold and white and the circular circuit board with some other electronic components.

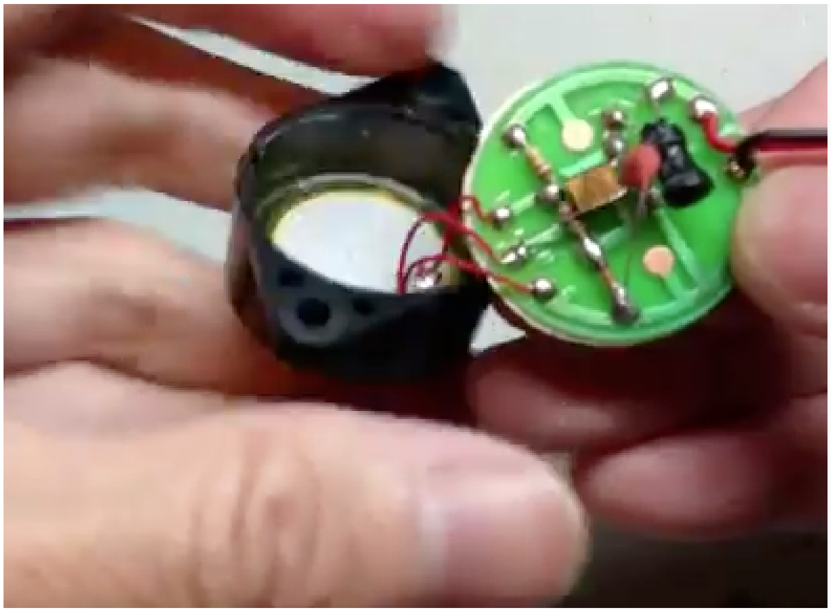

[http://2.bp.blogspot.com/-ZhHrjOlicT4/UV03KsrjOCI/AAAAAAAAA4M/iAA0YXY_9N8/s1600/Screen+Shot+2013-04-04+at+4.17.13+PM.png]

3. Remove all other electronic components using a solder iron and a clamp; as these components will reduce the impedance and also lower the audio output of the crystal earphone we are going to make.

Dynamic Views theme. Powered by Blogger.

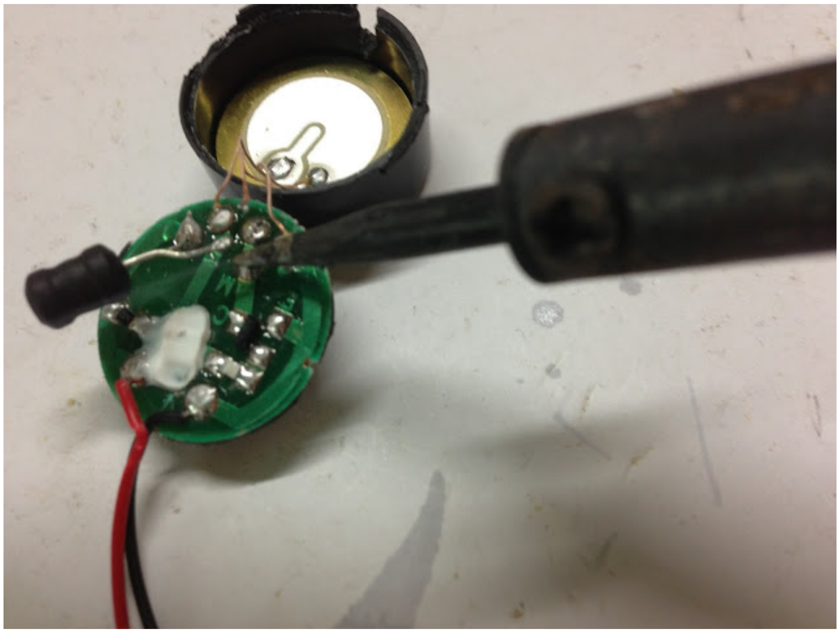

[http://4.bp.blogspot.com/-u4SR5E_5NRI/UV1U91sn2hI/AAAAAAAAA5A/vlQichGwxCw/s1600/IMG_1248.JPG]

This is how it looks after all the unnecessary electronic components have been removed from the circuit board.

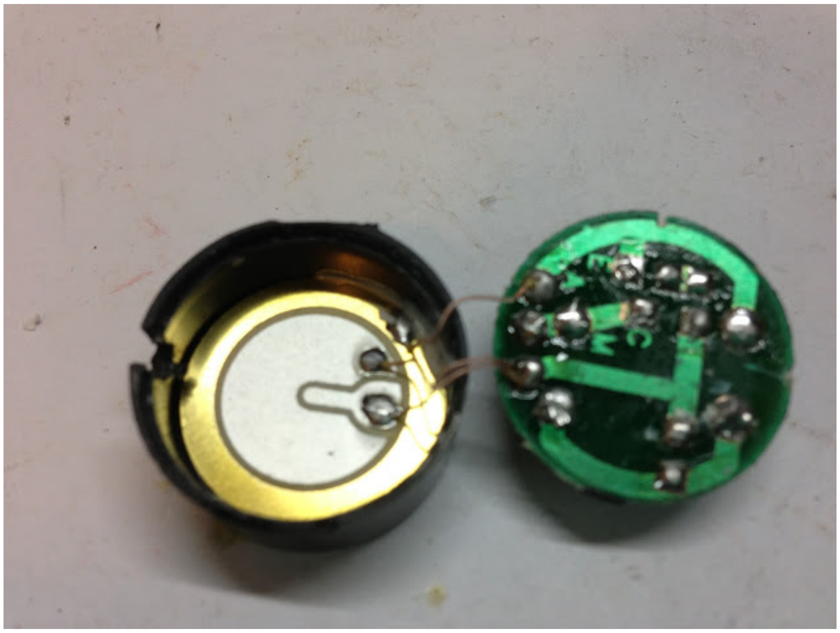

[http://1.bp.blogspot.com/-uC2t2vpXUus/UV1U-kz9wol/AAAAAAAAA5g/UKrFKi_8s9c/s1600/IMG_1251.JPG]

4. Using a solder iron, connect two wires from the earphone jack directly onto the two points on the plate. One wire touching the circumference (gold), a second wire touching the inner plate (white). There may be a third white plate which is smaller, connect the wire from there to the other bigger white plate to get maximum audio output. Polarity is not important. So, you can choose which wire to go where.

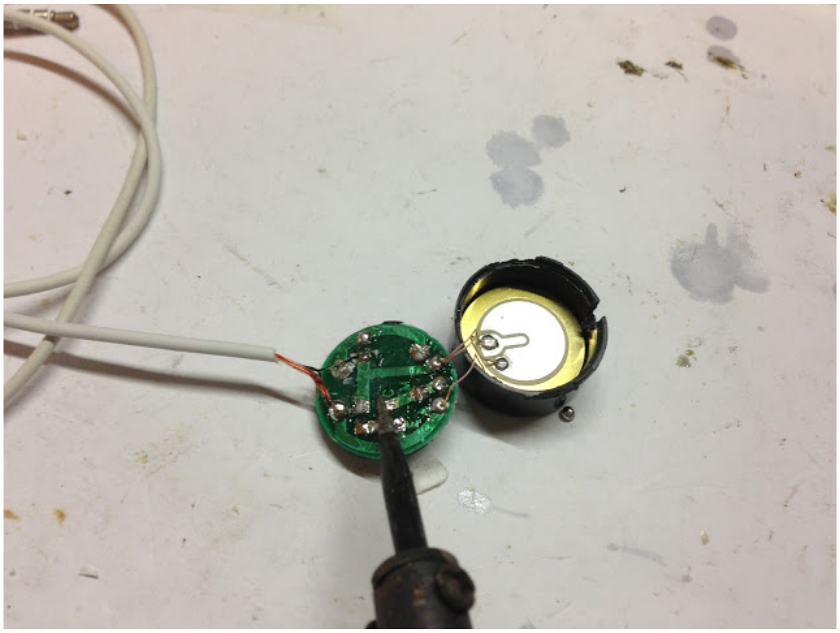

[http://3.bp.blogspot.com/-PqnLtCqYJ14/UV1U_ScG6vI/AAAAAAAAA5k/6vlkYKsyf-Y/s1600/IMG_1253.JPG]

5. If you are using a separate earphone jack, connect the other end of the wire to an 2.5mm mono headphone jack.
6. Before you close the lid, test the connections it by shorting the terminals of the jack together using a screw driver or touching the tip and the base of the earphone jack with a multimeter set to Ohm-meter. You should hear some "click" sounds each time you shorten the terminals of the jack.

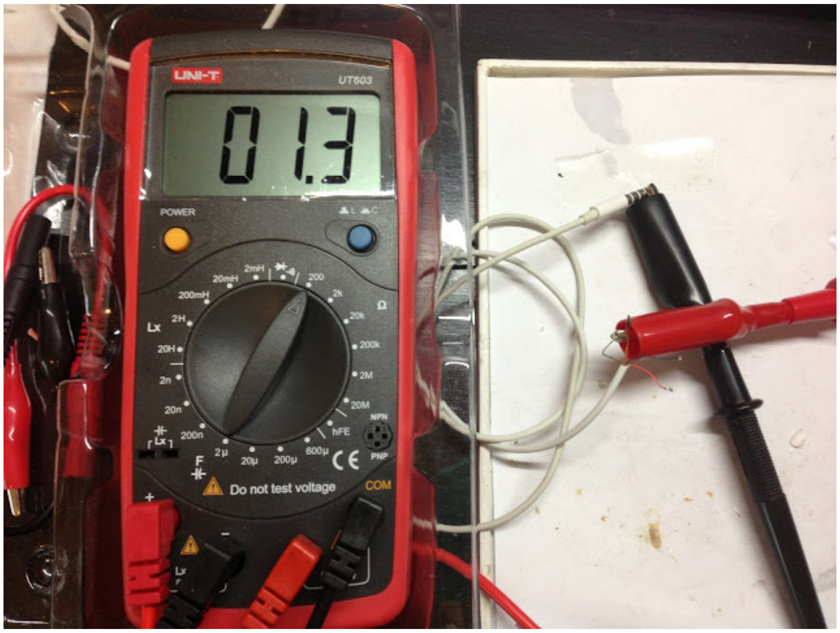

[http://2.bp.blogspot.com/-f8bHtWE-d1Q/UV1U-4AXTNI/AAAAAAAAA5c/IP9i34Ay-u4/s1600/IMG_1252.JPG]

7. Close the lid and secure it by applying glue. Do not use the strong instant super glue. Use some regular glue so can still open up the lid later for maintenance in case the wire inside is broken.

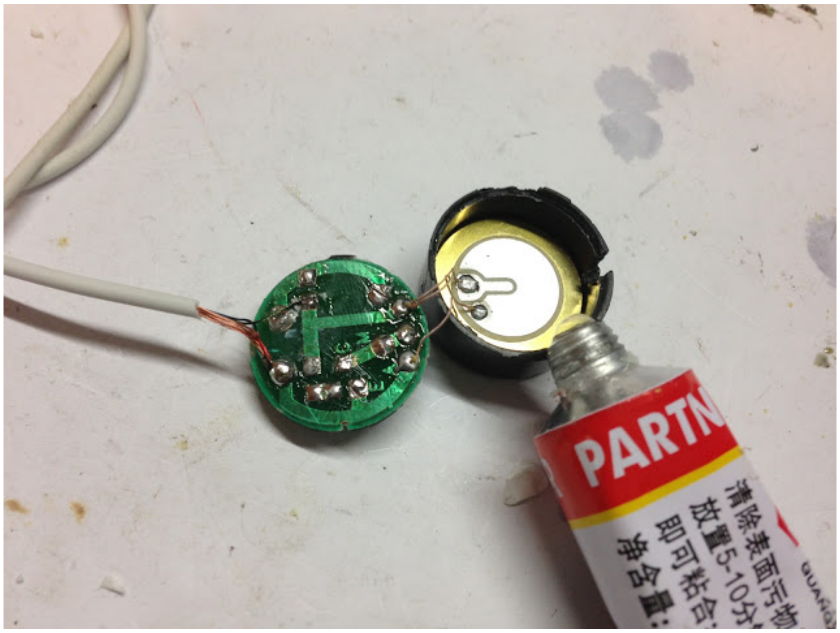

[http://4.bp.blogspot.com/-AhGlmtzQ604/UV1U_63a5WI/AAAAAAAAA6A/vWaW7oQGpBI/s1600/IMG_1254.JPG]

8. Line up the hole from the tip of the ball pen to onto the hole in front of the buzzer. Stick them together using instant super glue. The function of the ball pen tip is to focus the sound to the ear.

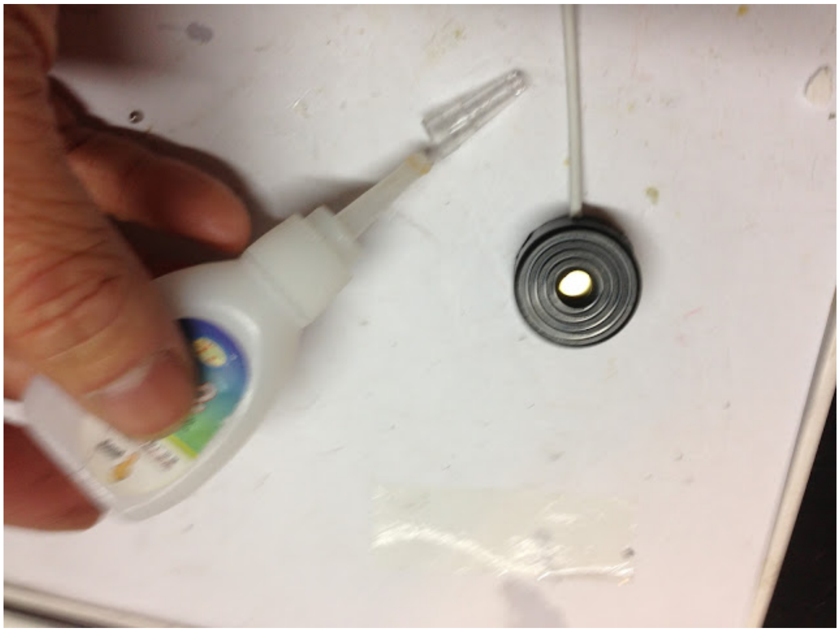

[http://3.bp.blogspot.com/-Dj7lCuYkL-Y/UV1U_3tn3tI/AAAAAAAAA5s/2ypnc_nQy_o/s1600/IMG_1255.JPG]

9. Attach the earbud cousin to the tip of the pen.

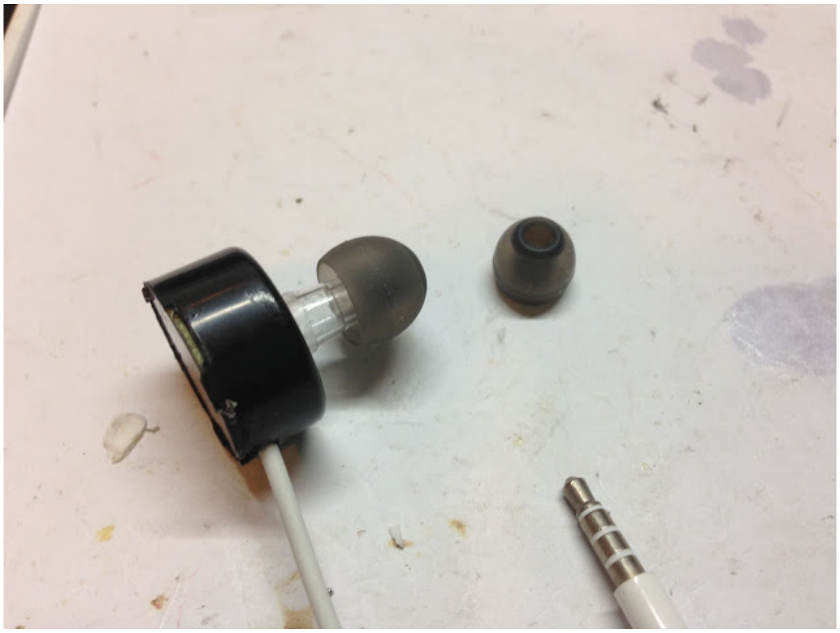

[http://3.bp.blogspot.com/-dYJDvrIR8hc/UV1VANPrc8I/AAAAAAAAA54/1TCHuDQzlOU/s1600/IMG_1256.JPG]

10. Connect it to your Crystal Radio to test.

My Library of Crystal earphones

[http://4.bp.blogspot.com/-o5FHwsGk9zI/UV1VB_VCCUI/AAAAAAAAA6Q/7K2nNAodoGU/s1600/IMG_1259.JPG]

Different types of piezo-electric buzzer you can find

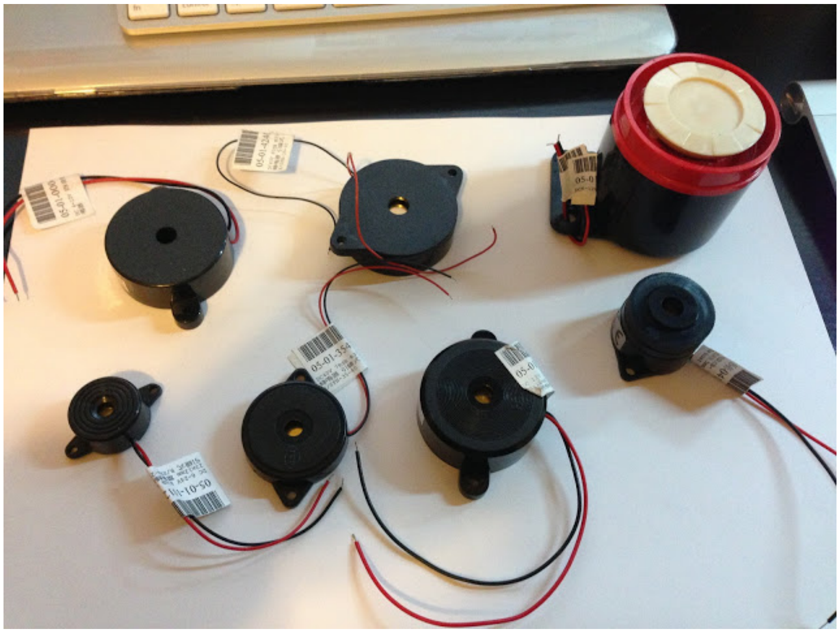

[http://1.bp.blogspot.com/-Z7tIVCPFr_8/UWLsPJUbPWI/AAAAAAAAA9o/HIYUHfVX-us/s1600/IMG_1276.JPG]

Posted 27th March 2013 by Billy

4 View comments

22nd March 2013

DIY - Battery-Free Tiny Crystal Radio

Build a Tiny Crystal Radio that runs without battery.

[http://www.blogger.com/blogger.g?blogID=2311644141825422235]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]

[http://www.blogger.com/blogger.g?blogID=2311644141825422235]
[http://www.blogger.com/blogger.g?blogID=2311644141825422235]

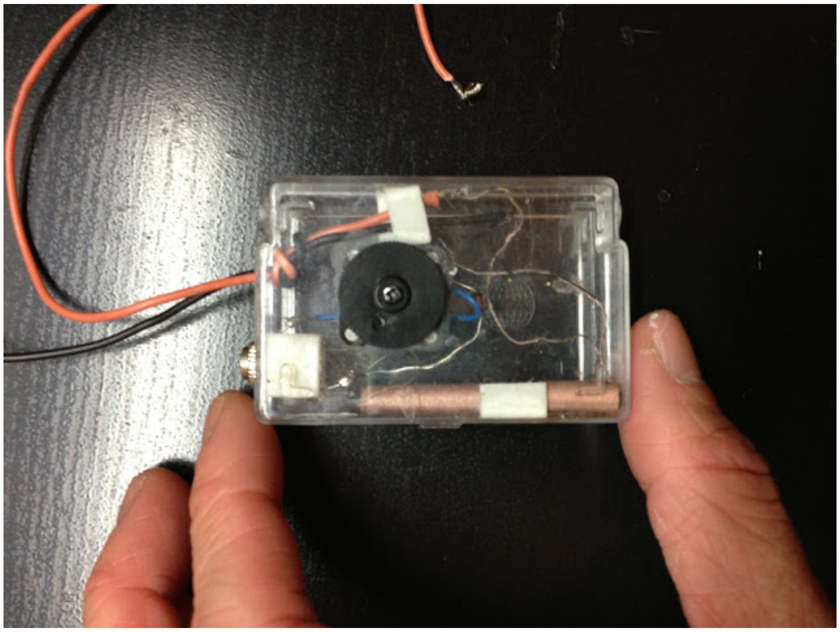

[http://2.bp.blogspot.com/-FYzx5aqFfVM/UVGOaN0HeBI/AAAAAAAAAxQ/XQtWPgoL_VQ/s1600/IMG_1098

Parts you'll need

1. Variable Capacitor 360 pf - from an old radio.
2. Coil wound on a Ferrite Powder Rod - from an old radio e.g. old Walkman
3. A Crystal Earphone or a PC speaker with builtin amplifier.
4. Germanium Diode IN60
5. Resistor 100 Kohm
6. 3.5mm earphone jack
7. A small plastic box,
8. some wires .
9. Solder iron.
10. Antenna - 20 meters hanging vertically or horizontally at a high location (Warning : Lightning Hazard)
11. Ground - connect a wire to your metal water pipe.

[http://www.blogger.com/blogger.g?blogID=2311644141825422235]

[http://4.bp.blogspot.com/-QqxK3OffVq4/UVLe8yWkFzI/AAAAAAAAAxo/4vtNVM1EUMY/s1600/parts.png]

You can get most of these from electronic ships in Shamshupo, except parts #1,2 & 3.
For the variable capacitor and the coil you can easily buy some old recycled walkman or old radios for HKD20-50 from the hawkers in Ip Liu Street in Sham Shui Po. The other way is to go to taobao.com to buy them online and ship to HK. The goods is RMB 5, but the shipping fees is minimum RMD 30, so not worth it unless you buy in great quantity.
For the Crystal Earphone, it's very hard to buy nowadays. Refer to my other post to make one yourselves. DIY Crystal Earphone for Crystal Radios [http://billydiy.blogspot.com/2013/03/build-crystal-earphone-for-crystal.html]

Remember to leave me some comments (at the comment box) at the bottom of this if you have read this blog.

How to make it

http://www.youtube.com/watch?v=6VsgXuNm13k [http://www.youtube.com/watch?v=6VsgXuNm13k]

Dynamic Views theme. Powered by Blogger.

Circuit Diagram

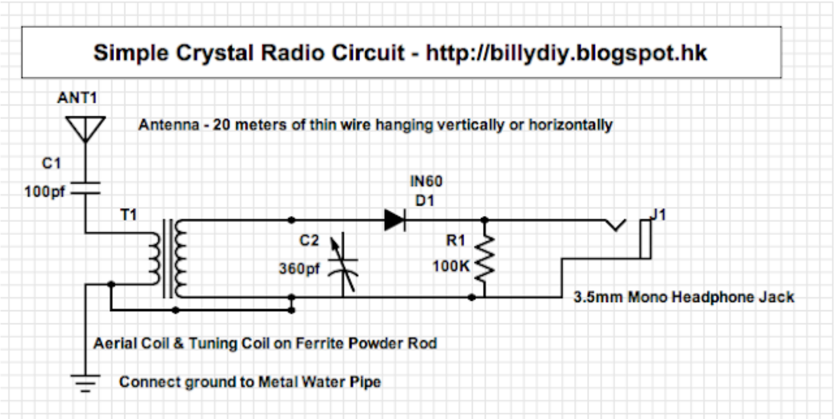
Simple Crystal Radio Circuit - http://billydiy.blogspot.hk
ANT1
Antenna - 20 meters of thin wire hanging vertically or horizontally
C1
100pf
T1
IN60
D1
C2
360pf
R1
100K
J1
3.5mm Mono Headphone Jack
Aerial Coil & Tuning Coil on Ferrite Powder Rod
Connect ground to Metal Water Pipe

[http://1.bp.blogspot.com/-C2qNgc8TvDg/UXFYJSUHmyI/AAAAAAAABDw/KaD3xmUTk-w/s1600/Screen+Shot+2013-04-19+at+10.43.18+PM.png]

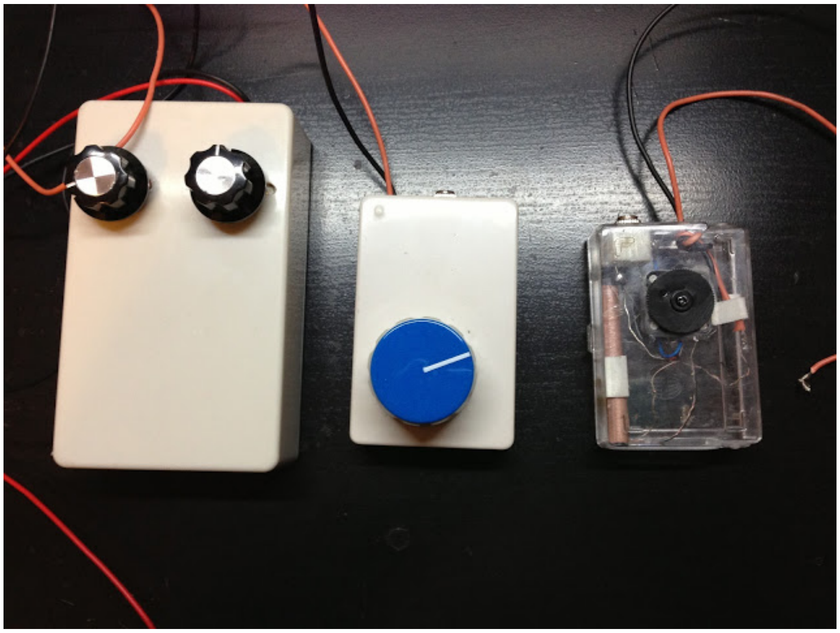

[http://1.bp.blogspot.com/-6Ns9sST5X4s/UVGPUo-IFmI/AAAAAAAAAxY/up4TGDU0jC0/s1600/IMG_1097.JPG]

Dynamic Views theme. Powered by Blogger.

- Contest Results
  - 2004
  - 2005
  - 2006
  - 2007
  - 2008
  - 2009
  - 2010
  - 2006 Sprint
  - 2008 Sprint
  - 2009 Sprint
- Previous
- Next

Radios Of The 2010 First Final Farewell DX Contest
Contest Entrants

Page 3

Sean Whiteacre, Hobby Class Entry

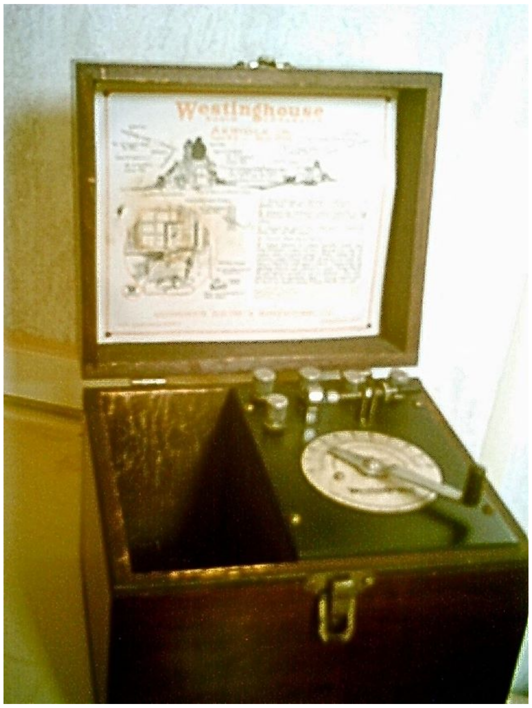

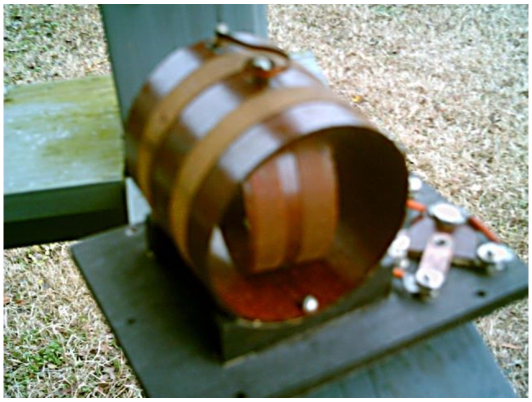

Westinghouse Radio
Apparatus AERIOLA, JR
Type RE style 307421

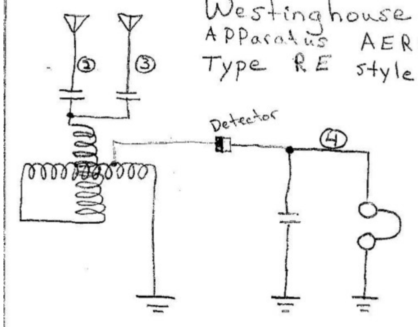

Made in June of 1921

② This post is for Stations Below 350 meters

③ This post is for Stations between 350 to 500 meters

④ Terminal 4 for headphones

The Antenna taps are for
Selecting the frequency
coverage of the radio
not for tuning the
radio for better reception.

The radio is a Westinghouse Aeriola, Jr. I did not use a regular diode, but the real detector on the radio. The "Perikon" detector uses two rather rare minerals; Bornite in the fixed cup and Zincite in the movable section. Anyways I used the Perikon Detector. I also used the original 1921 C.Brandes Superior headphones from 1921.. Antenna was 25 ft in air and 150 ft long, Inverted L type.

Kevin Norton, Open Class, Loop Class and Active Device Class

I used the K-3 Big Litz Dx Set for the Open Class. It was the same set up as last year. For the Loop Class, I used a low Q loop built on cardboard box.

The Active Device Class entry was a WN6Q MPF-102 Regen (Via Owen Poole). I originally tried this as a double tuned set using 2x 660/46 ferrite assisted air coils, 4 gang ceramic wiperless variable caps, and a Litz air tickler. The regen set worked well, but was very hard to line up and very touchy. When the antenna tank was removed the set worked MUCH better! It still WAS a regen, and therefore still had a few flukes. It was MUCH more tamed, and Dxing became much easier.

The sets radiation was read off of my Icom R-70, so digital spotting was a snap. The final version was a single tuned tank of a ferrite assisted air coil, a 4 gang ceramic wiperless and crystal earphones in parallel. The antenna was VERY lightly coupled via a few insulated turns of a gimmick capacitor. At times, I could zero beat the set simply by turning my head a bit to one side!

I copied SOMETHING (anything from beat notes to ear splitting locals) on nearly EVERY frequency out of the 119 channels. Maybe 90 stations or more of at least very weak audio or better.

FWIW, I made a big boo-boo and left the set in place during a vicious white out snow squall. We had very high winds and blizzard type of snowfall rates. The antenna kept discharging arcs every 5 or 10 seconds. They could be heard in the next room. By the time I realized what was going on, it was too late. I pulled the antenna line off and measured the arcs, which by that time were down to around 3/4". To fry the set in the position it was in, they had to have been at around TWO inches (30 Kv??) to bypass the gimmick cap twists and leap to the set. Needless to say, a new FET had to be installed. Live and learn.

Glen Yarbro, Open Class

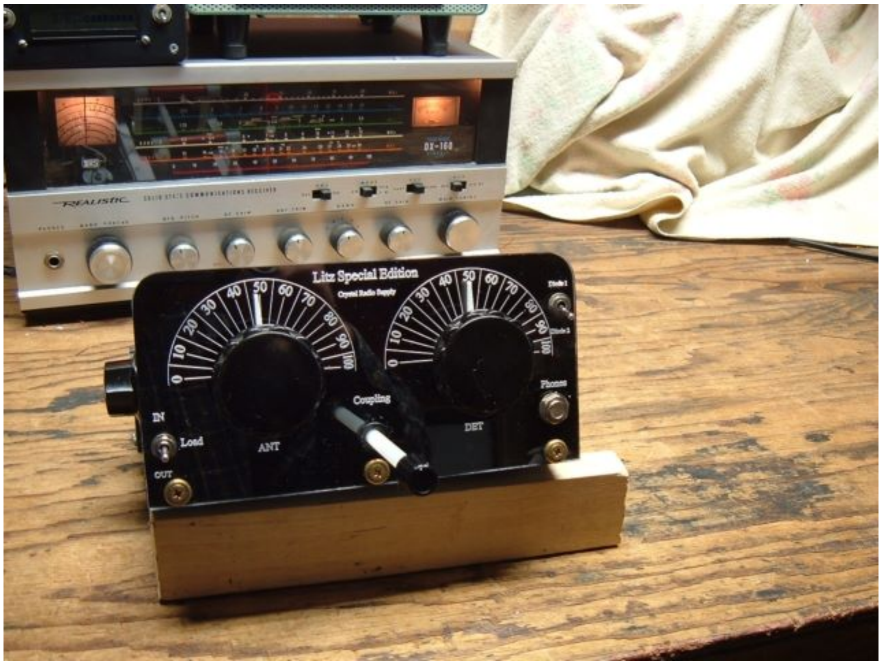
Litz Special Edition
Crystal Radio Supply
Diode 1
Diode 2
Phones
Coupling
IN
Load
OUT
ANT
DET
REALISTIC
DX-160

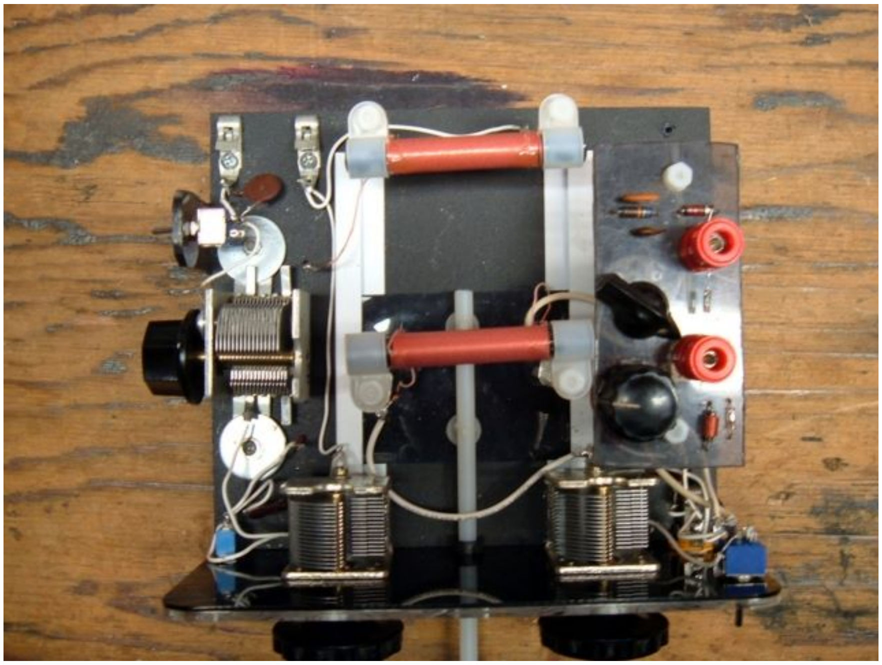

The set used was a Litz Special, modified to Dave Schmarder's #44 wiring configuration. The antenna was 250 feet long and 30 feet high. The ground consisted of two 8 foot ground rods.

The set used an 1N34A diode and 4K Navy hadphones. No wave traps were used. The spotter radio was a Realistic DX-60 with a frequency counter.

Continued on Page 4.

Select Menu

- Home
- About Us
- Contest Results
- 1AD Contests

- Contest Results
  - 2004
  - 2005
  - 2006
  - 2007
  - 2008
  - 2009
  - 2010
  - 2006 Sprint
  - 2008 Sprint
  - 2009 Sprint
- Previous
- Next

Radios Of The 2010 First Final Farewell DX Contest Contest Entrants

Page 2

O. T. Anderson, Open Class Entry

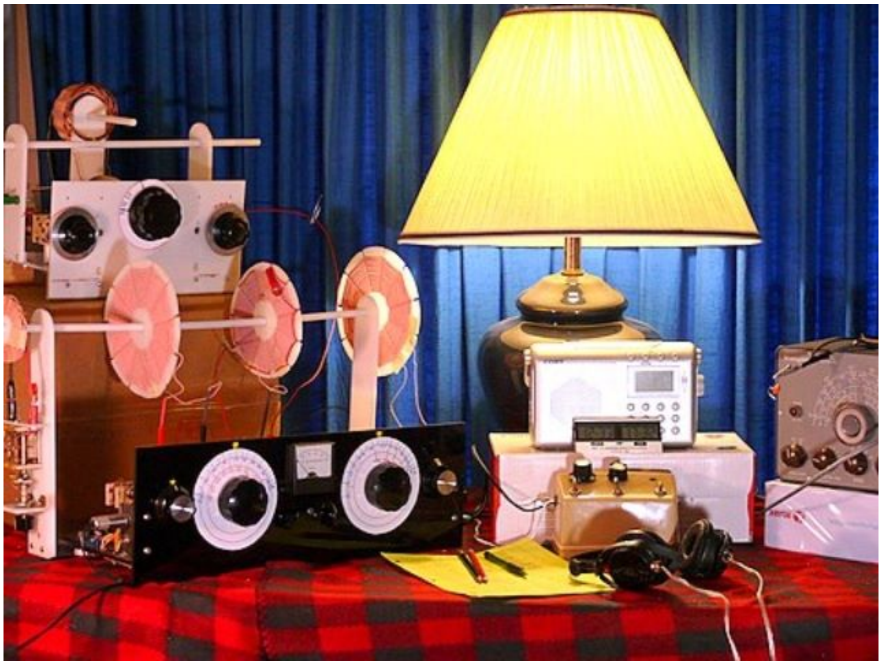

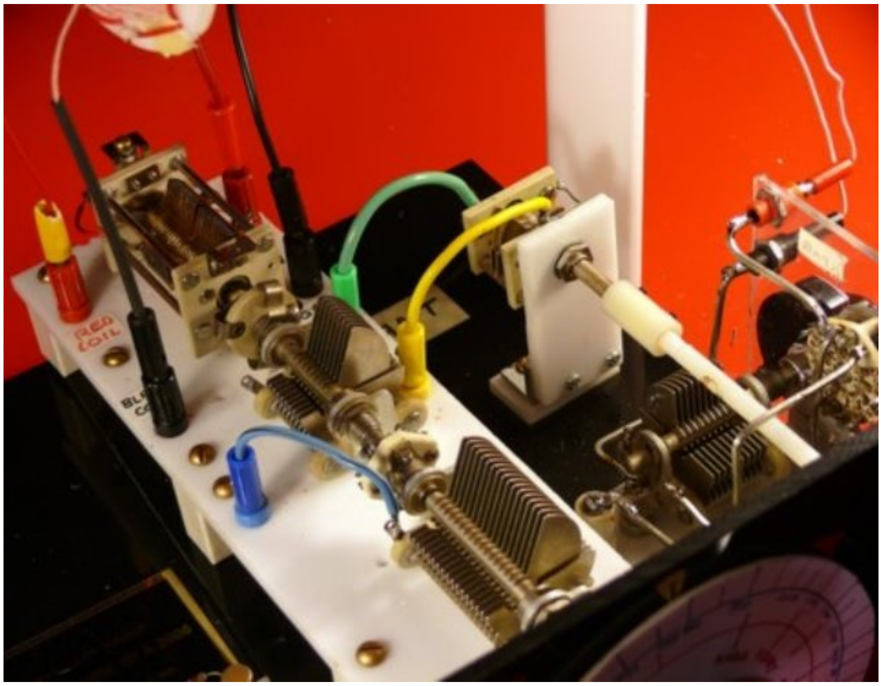

This the same crystal set I used last year with a few revisions. I had 66 stations last year and my goal was to double it. The Mexican stations come in good in Oklahoma.

This crystal set is the same as last year with 3-4 updates.. The extra coils are wave traps for knocking out that KFAQ, 50,000 watts, in sight of our house. Also another 50K across the town. Up high is a new wave trap with silver caps. It will tune any local station. The pictures also shows a "Drug Store" spotter and a Knight Sig Generator. The two dials show the freqs and also spots several local stations. The right dial has another calibration, one is the 1700 to 530 freqs, the other calibrations is the high end freqs when using only the 140 mmfd condenser, about 1750kcs to 1000 kcs. Lots of band spread.

The other picture is the RF section with a silver cap at the far end. It tunes the antenna. The small var cap with the nearest blue wire connects the 2 nearest caps for all the freq 1700 to 540, normal use. With the blue wire not connected we use only the 140 mmfd cap, tuning about 1750 to 1000. This gives good band spread at the high end. The blue and yellow wires connect a 15 mmfd var cap, helps to "rock" in a distant station. If I good get rid of that terrible KFAQ signal things would be a lot better.

Jack Bryant, Open Class Entry

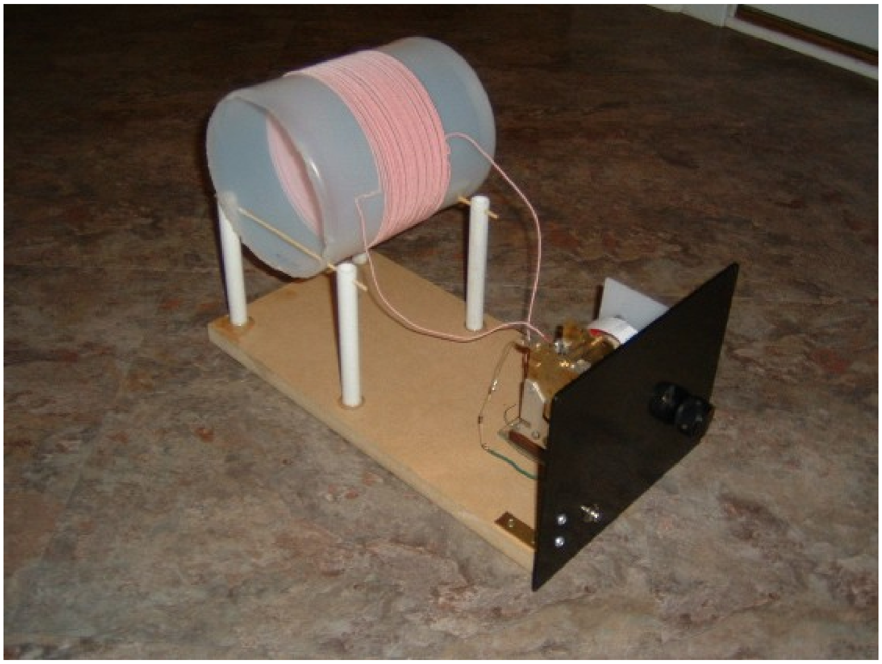

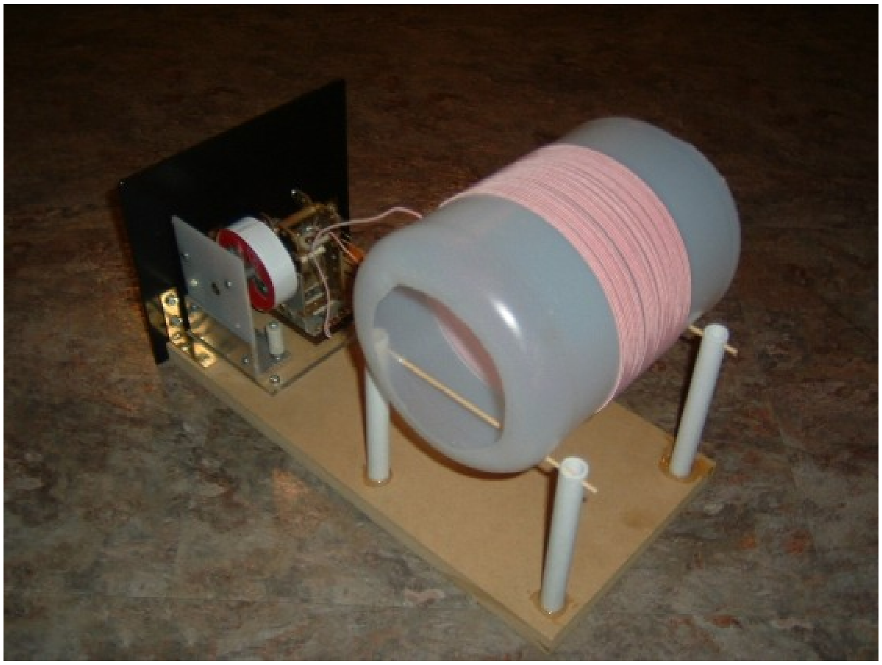

I only worked the last week of the contest, focusing on the broadcast band. I used essentially the same set for that I used in years past.

I have used a Realistic DX-398 for several previous contests, and I used it for this one, too. I used a PC for logging along with paper copy backup. I used a four foot table from Lowe's as my listening post.

Garry Nichols, Loop Entry

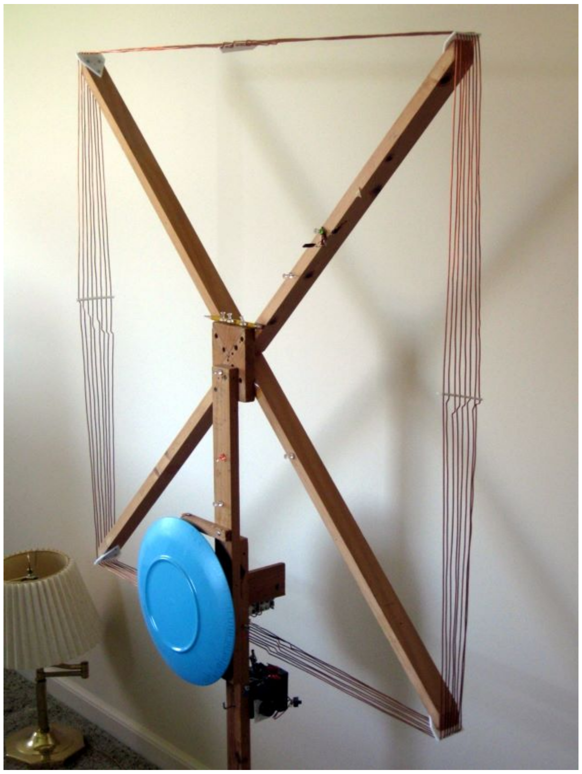

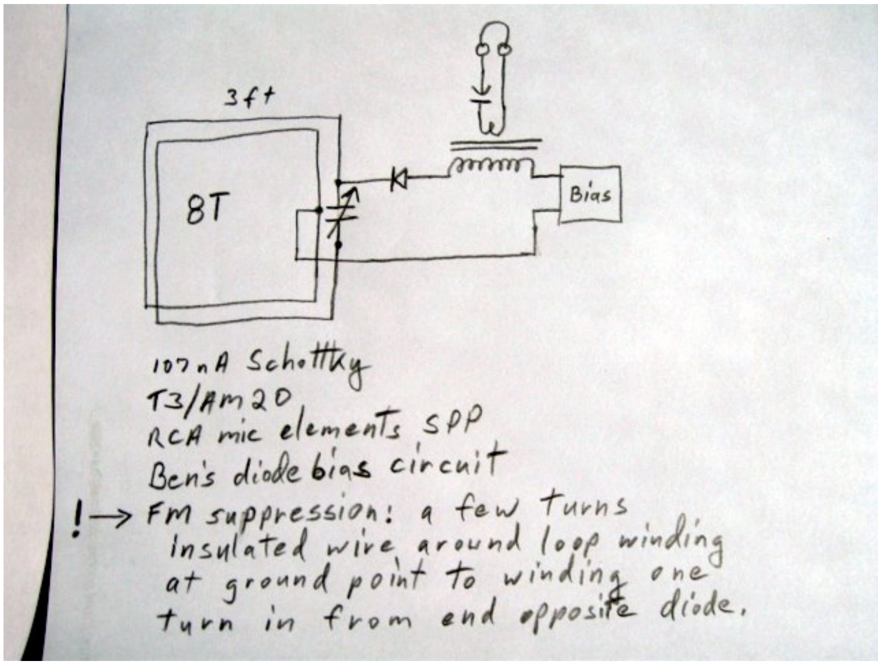

107nA Schottky
T3/AM20
RCA mic elements SPP
Ben's diode bias circuit
!→ FM suppression: a few turns insulated wire around loop winding at ground point to winding one turn in from end opposite diode.

(Note from Jack: don't you just love the back of the envelope schematic! A wise sage at work use to say that if a detailed study didn't agree with the broadbrush, back of the envelope calculations, then the detailed study was probably wrong!)

A few changes for my loop this year:

1) Turns spreaders at the midpoints of the frame tightened things up and stopped rattling wire. Don't know why I didn't do this sooner!

2) I seriously reduced my troublesome FM interference with a simple trick. Somehow I noticed that a few pF across some of the loop turns seemed to reduce it. So I experimented first with a tiny variable (which turned out to have too many pF even fully open) and then with a small length of thin teflon insulated wire. The best positioning was from the exact center of the total winding across a few turns to a position one turn in from the end of the winding opposite the diode. I put it there to possibly help balance the loop because I use the other half of the winding for the detector. A few turns of the wire around each winding did the trick.

The upper half of the BCB used to be plagued by FM. It was difficult to tell if I was hearing a weak AM station or FM interference unless I spent some time listening. A real pain! My mod suppressed interference on the upper half and I can only hear FM if I tune above 1700. A lucky break after on and off problems for years!

3) I tried Ben's diode bias circuit this year along with my favorite 107 nA Schottky diode. I had to bias the diode, so that it appeared to have a higher saturation current. A bit more, lower in the BCB, and a bit less near the top. Fits theory pretty well! I used only the T3/AM20 transformer on my RCA mic elements to avoid possible problems with the Select-To-Match circuit and the bias circuit not getting along. How best to hook the two up left some doubt. For my setup, I could never decide if the STM was better than the T3/AM20 alone anyway.

Dan McGillis, Open Class Entry and Active Device Entry

I used two radios in the 1/15/10 contest.

For the BCB Crystal radio section, I used a simple double tuned Mystery set made with dual-gang capacitors from Leeds Radio and ferrite rods wound with 165/46 litz. It's described here.

It was a very good daytime performer, but only a fair night time DX catcher. The bandwidth was just a bit too wide for all the strong DX stations and the digital hash which now seems to be everywhere at night. Still, it's a nice loud little radio.

For the active device section, I used a 2-JFET Regen plus an op-amp audio amplifier with an active bandpass filter. My primary goal was to be able to copy Morse code (CW) in the 40 meter ham band (7 MHz). The radio is described here.

The performance of this simple set-up really surprised me. It's an excellent CW receiver -- stable, quiet, sensitive, and selective. Excellent for SSB and AM too. A real keeper. I even managed to make some 2-way contacts on 40m using this receiver and about 3 watts of transmitting power. That was a lot of fun.

Jerry Walker, Open Class Entry

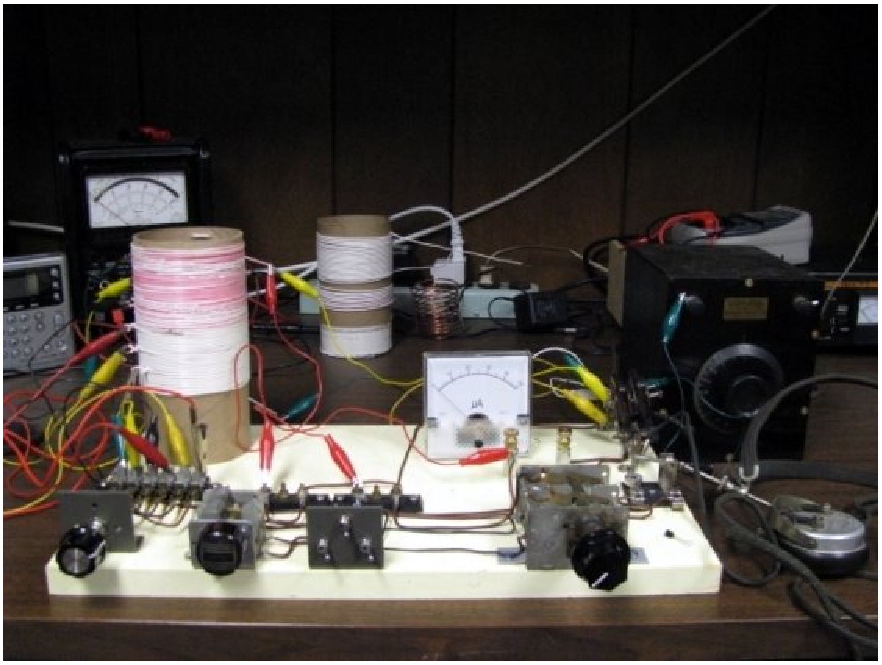

-The crystal set used was double tuned.

-The antenna was a 75 ft long wire.

Continued on Page 3.

Mini A8 Real Time Car GPS Tracker...
$9.67 Specifications: This GPS tracker applies the newest technology with the following...
ebay
Ad

Select Menu

- Home
- About Us
- Contest Results
- 1AD Contests
- 1AD Entries
- Antenna Matching
- Antennas
- BHAM Radios
- Crystal Contests
- Crystal Radio Entries
- Sprint Contest
- Headphones
- Projects, Mods
- Favorite Links
- Contact Jack

Sub Pages

- Main Page
- 2004 Page 1
- 2004 Page 2
- 2004 Page 3
- 2005 Page 1
- 2005 Page 2

- Contest Results
    - 2004
    - 2005
    - 2006
    - 2007
    - 2008
    - 2009
    - 2010
    - 2006 Sprint
    - 2008 Sprint
    - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2009 Contest Entrants

Page 4

James Gallant, Open Class

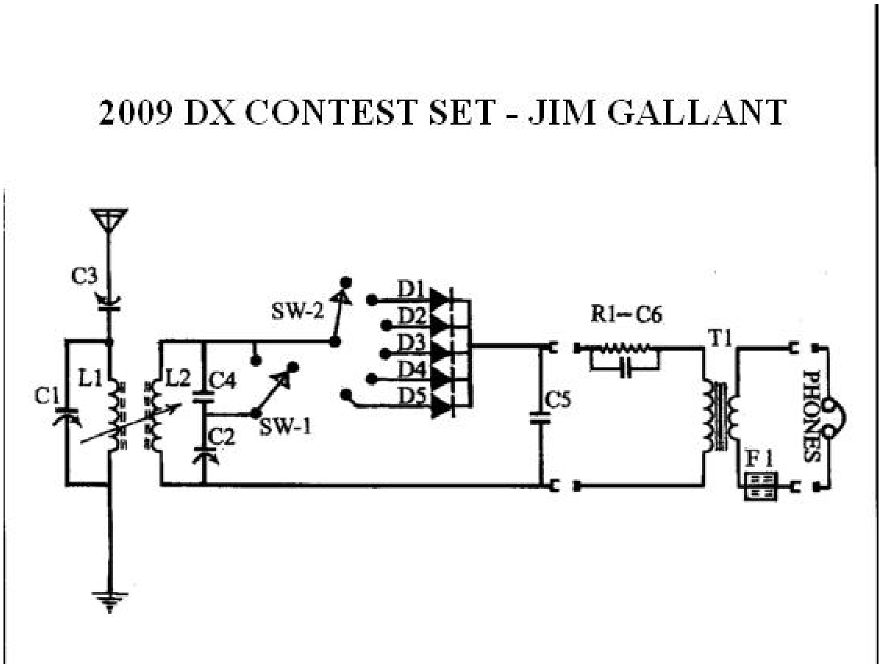
2009 DX CONTEST SET - JIM GALLANT
C3
SW-2
D1
D2
D3
D4
D5
R1–C6
T1
C1
L1
L2
C4
C2
SW-1
C5
PHONES
F1

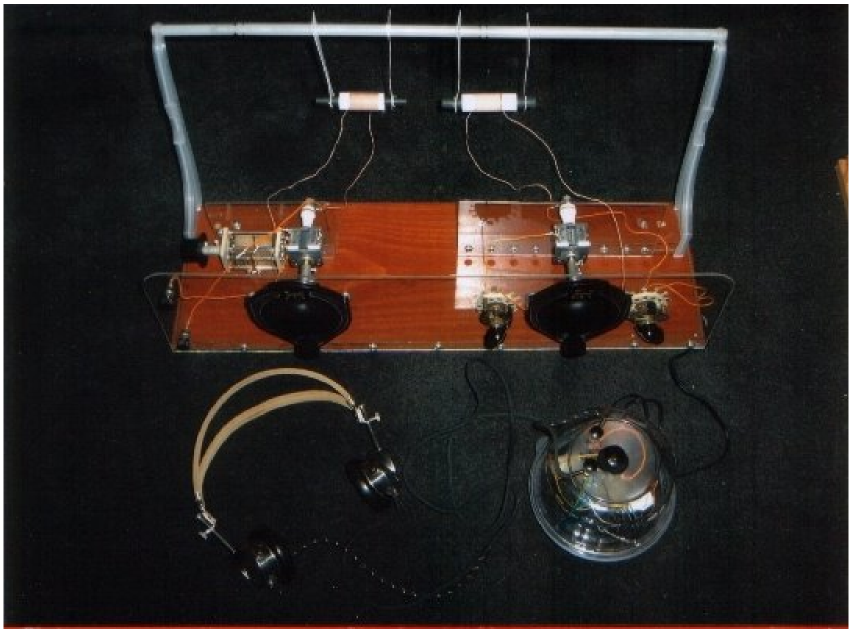

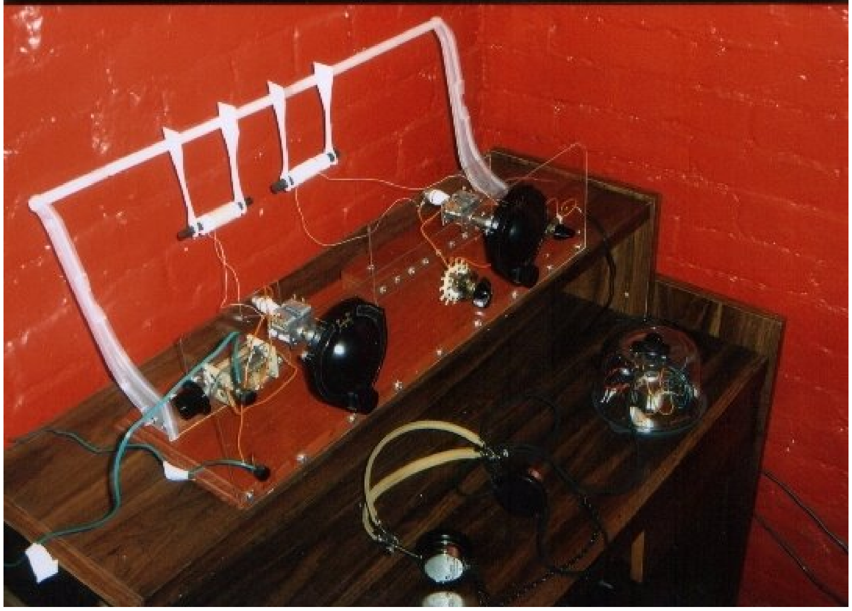

Due to a dispute with a former landlord (NOT related to the radio), I lost my previous office, radio loft, and 5 story high, 115 foot long-wire antenna for this year's contest. I also used the same mediocre "throw together" 100/46 Litz wire dual-tuned set that I whipped up last year. I still used my simple Bogen T-725 transformer driving a pair of 1925 vintage Western Electric 509-W headphones. I had to run it all off a long-wire antenna only 20 feet or less off the ground, bent around all over the place to fit my back yard.

Strangely, this year the radio seemed to work FAR better with the antenna (series) matching capacitor moved from the antenna line above the ATU tank, to the ground wire BELOW the ATU tank; most likely due to the vast differences between to two antenna sites used with the set in 2008 and 2009. This gives further evidence of how two different builders with two basically identical sets have been getting different results as to where to put the ATU series capacitor. With the same radio, radically DIFFERENT results, depending on the antenna used).

My only other change was the replacement of the approximately 46K fixed resistor in my "Benny" with a 250K audio taper pot - a noticeable improvement.

I assumed the set performance would be much worse overall than last year, but I was pleasantly surprised. Although it did have poorer overall reception of New England area stations, it did as well or BETTER on DX. In fact, I raised my total station count this year from 62 to 68 stations, with fewer of them being locals (Nashville was my furthest hit, I believe). It raised my points from 42,000 to over 64,000 - a 50% increase. I attribute this to the fact that this is my second contest.

I am getting more capable with the "spotter radio" in identification of stations, as well as hunting stations down on the Crystal Set after first isolating them with the spotter. Hopefully for next year I will have a 660/46 Litz “dream set” hooked up to a nice set of SP phones and a decent 30 foot high, 50 foot long 4-wire flat top antenna squeezed into my yard. My overall goal is fairly modest. Massachusetts is notoriously BAD for crystal radio DX. The glaciers in the last ice age pushed billions of tons of our topsoil down into the Atlantic Ocean south of Connecticut. As a result of this, we have poorer and less efficient grounds than other regions.

Assuming a 450 mile average maximum DX "listening circle" for the average set, over half of MY circle here in coastal Massachusetts is to my east, over the wide open empty ocean. If I can ever manage to break a 100 station level count, I will consider that winning my own personal gold medal!

James D Fletcher, Hobby Class

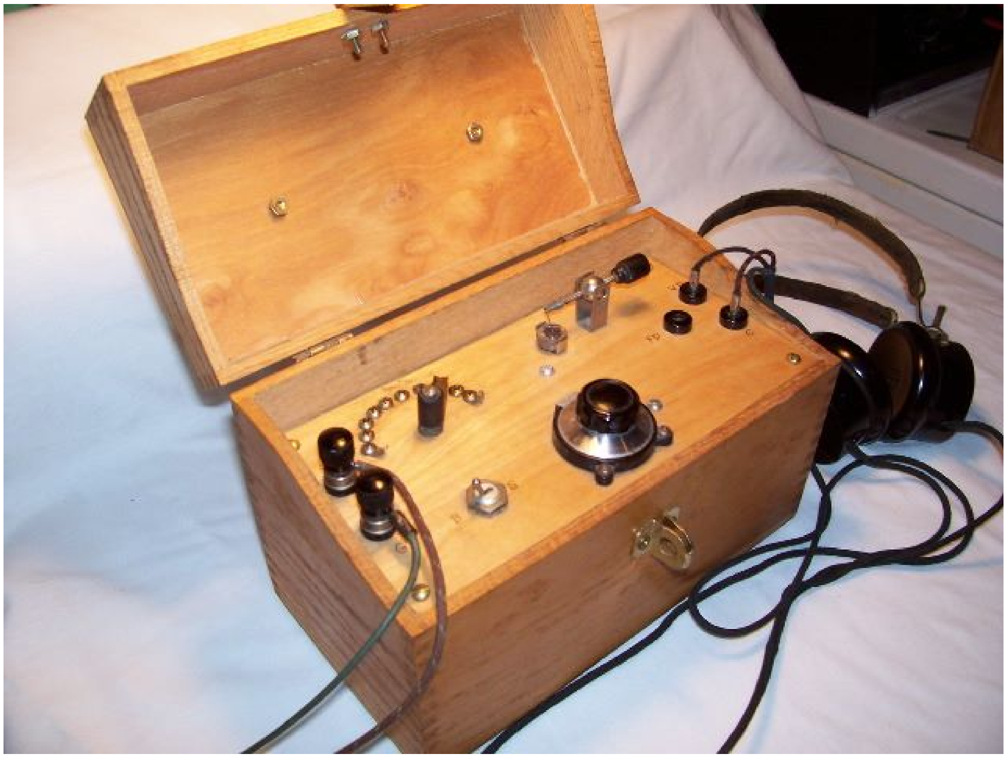

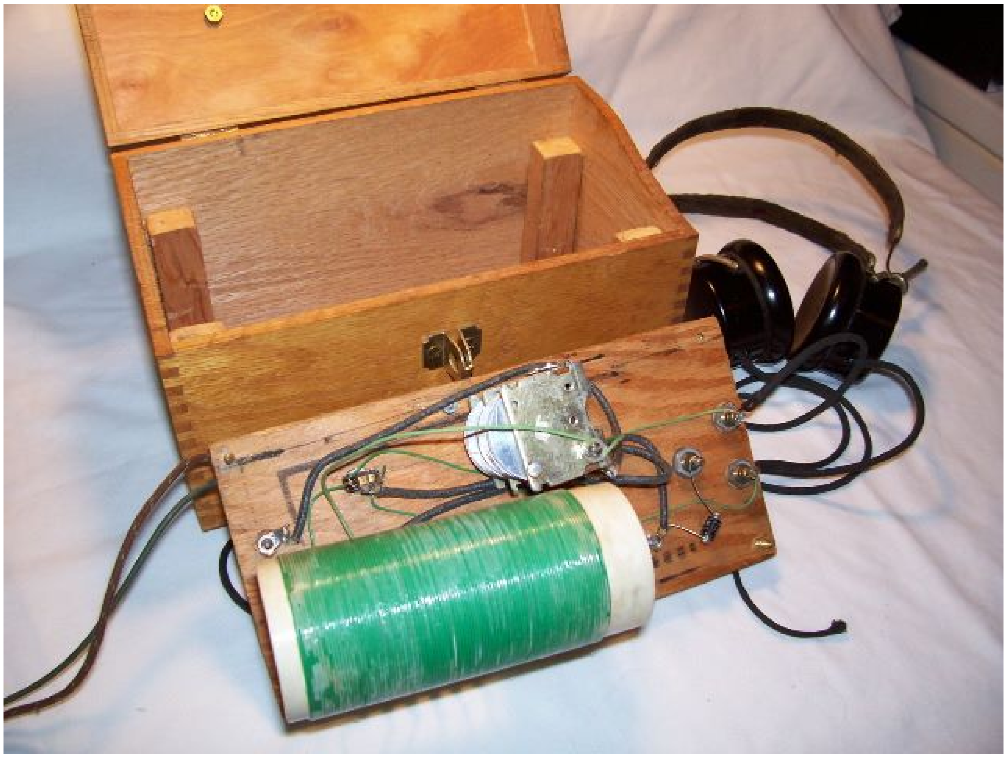

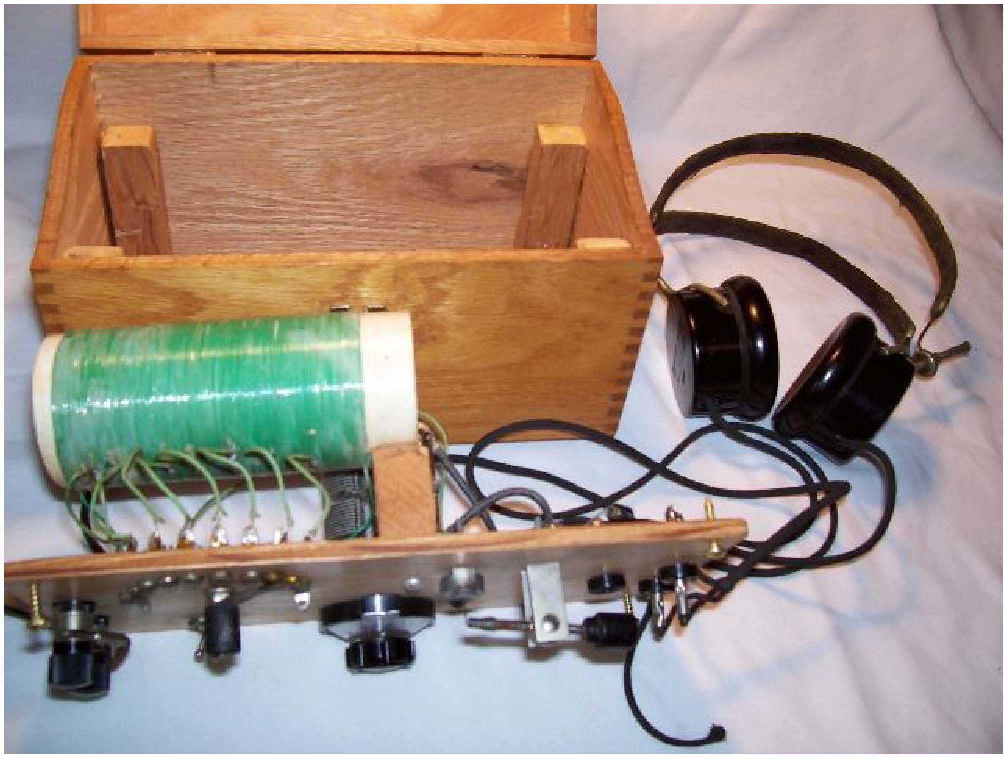

I built the MRL No. 2A "Long Distance Crystal Set" from plans I got from Mr. Elmer G. Osterhoudt. I used parts from what I had on hand and wound the coil on PVC pipe. This was in 1979 and my oldest son was 10 years old, so this was a project we worked on together and gave him some practical experience he still uses today.

I put the works in a 8" x 5" oak file box that office supply stores used to sell. I use a type "C" Baldwin headset that really makes a difference compared to regular headsets. This is the same set that I used in the 2008 Contest.

Lem Morrison, Open Class

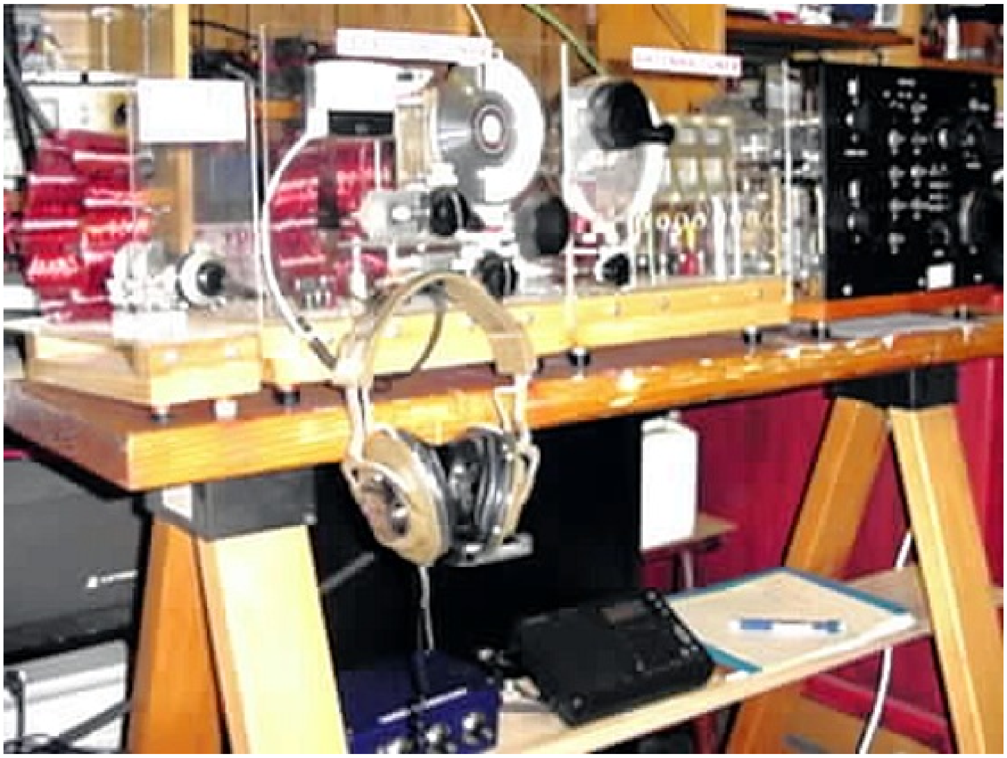

- Contest Results
  - 2004
  - 2005
  - 2006
  - 2007
  - 2008
  - 2009
  - 2010
  - 2006 Sprint
  - 2008 Sprint
  - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2008 Contest Entrants

Page 3

James Gallant

View Page 1, JPG File

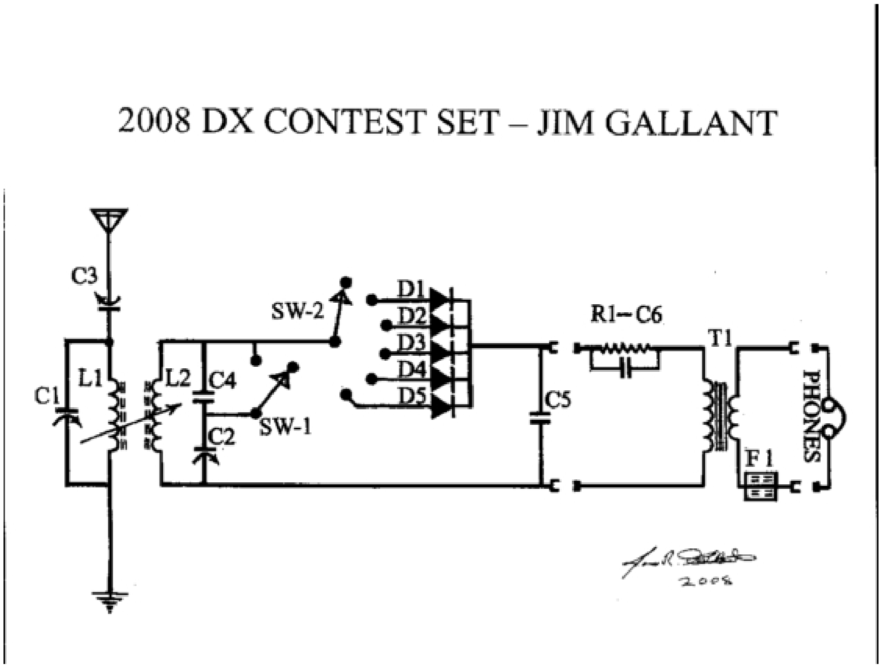
2008 DX CONTEST SET – JIM GALLANT
C3
SW-2
D1
D2
D3
D4
D5
R1–C6
T1
C1
L1
L2
C4
C2
SW-1
C5
PHONES
F1
2008

View Page 2, JPG File

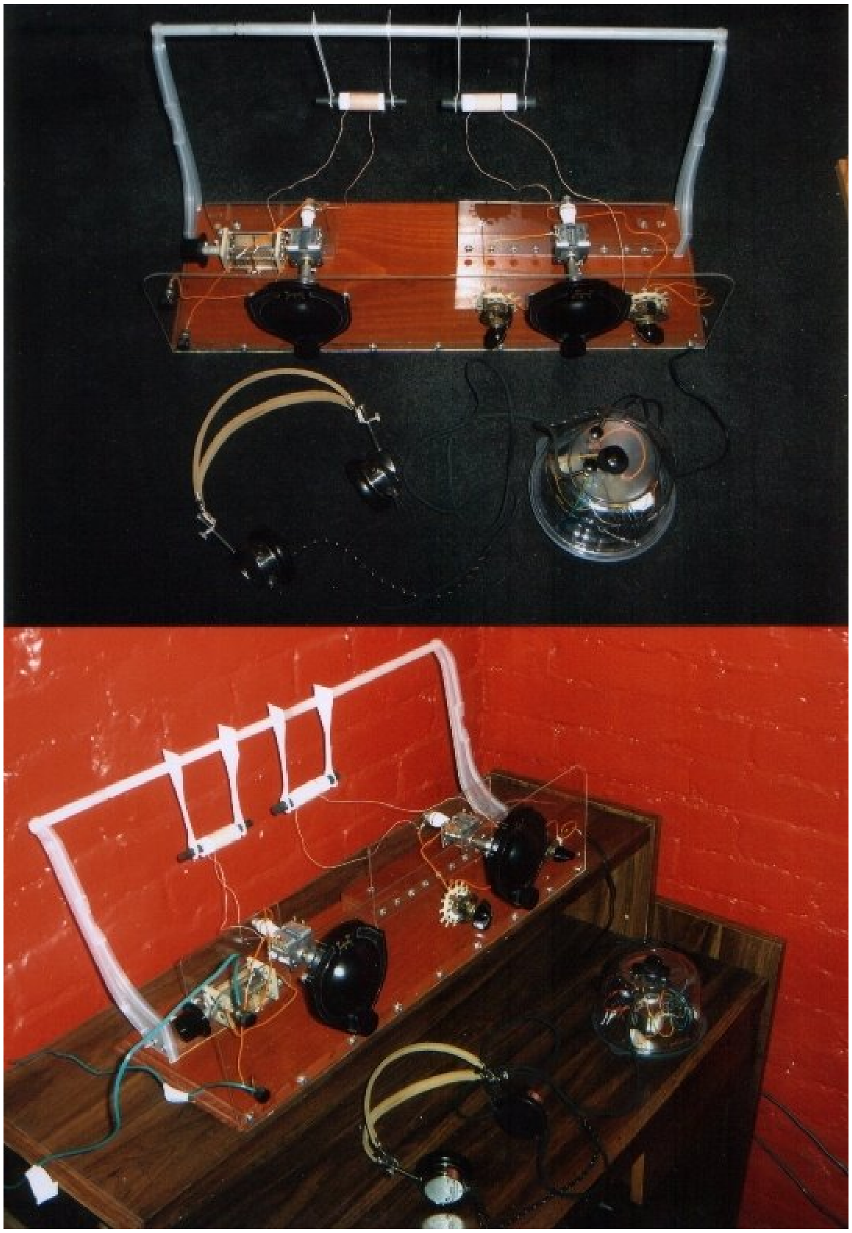

[To the main page.](#)

Ad
REDBUBBLE

Select Menu

- Home
- About Us
- Contest Results
- 1AD Contests
- 1AD Entries
- Antenna Matching
- Antennas
- BHAM Radios
- Crystal Contests
- Crystal Radio Entries
- Sprint Contest

- Contest Results
  - 2004
  - 2005
  - 2006
  - 2007
  - 2008
  - 2009
  - 2010
  - 2006 Sprint
  - 2008 Sprint
  - 2009 Sprint
- Previous
- Next

Radios Of The 2010 First Final Farewell DX Contest
Contest Entrants

Page 4

Dan Denapoli, Open Class and below BCB Class

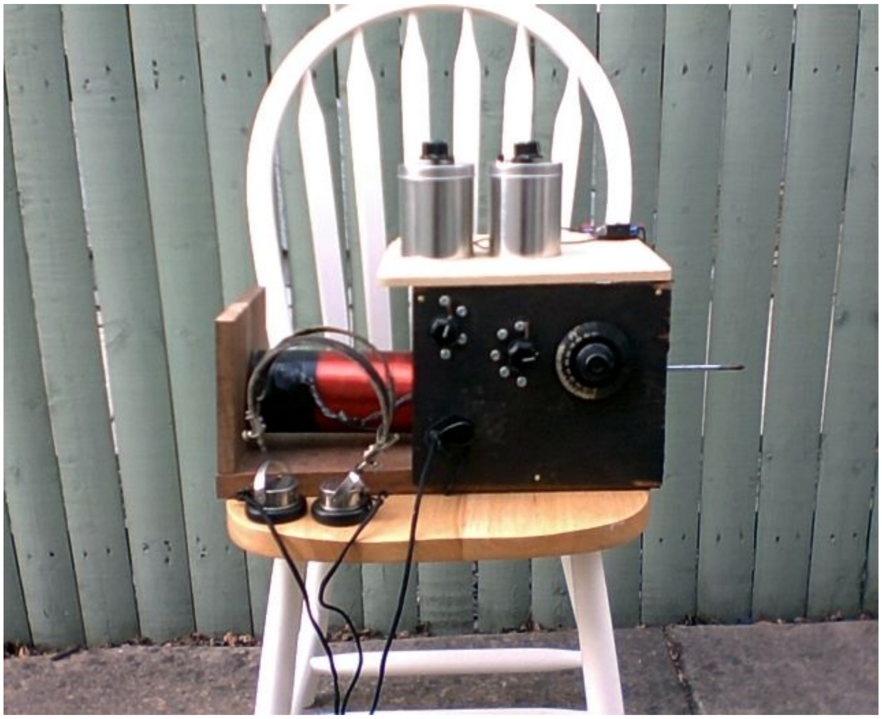

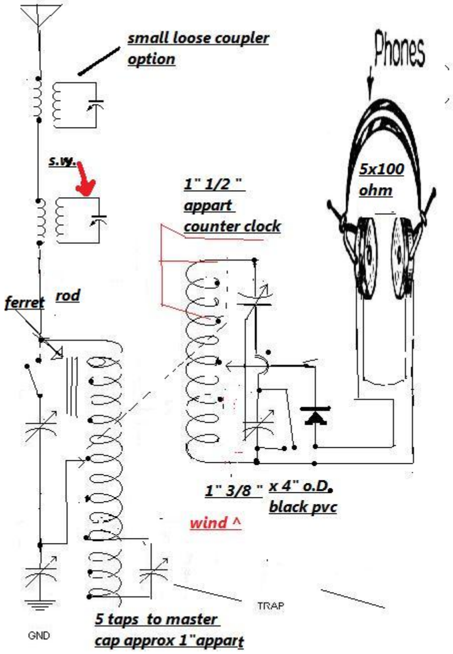
small loose coupler option
Phones
s.w.
1" 1/2 " appart counter clock
5x100 ohm
ferret rod
1" 3/8 " x 4" o.D. black pvc
wind ^
TRAP
GND
5 taps to master cap approx 1"appart

Chris T, Hobby Class

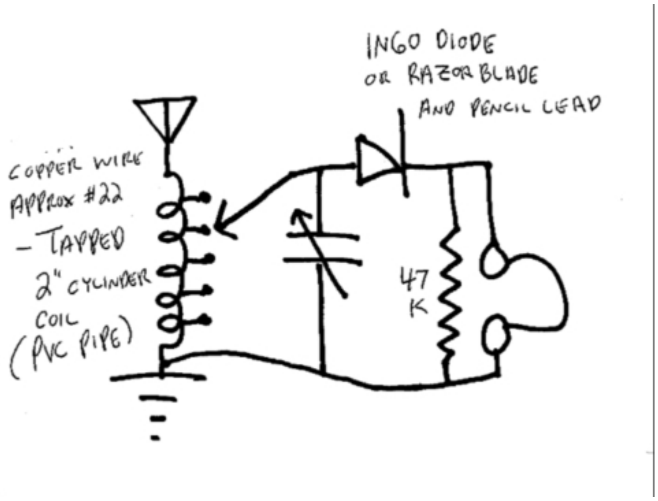

Dave Schmarder, Active Device Class

Click here to go to my website and see what I did this year:

-Single 7C5 Loctal tube regenerative radio with 6 volt filament and 25 volts DC plate

-Two tuned circuits using toroidal forms. Single tuning control

-Sound powered headphones

-Antenna, 40 meter wire up 3 - 4 meters running east to west.

Robert Weaver, Active Device Class

This year, I have two different active device receiver entries. One is a direct conversion receiver using a sheet beam tube, that I built two years ago, and used in the Radioboard summer DX contest in 2008, but never in this contest. Shortly after I built it, I built a sheet beam superhet receiver and the direct conversion receiver was mostly forgotten. I wanted to dust it off, make a few improvements and enter it in another contest. It took a while to get to it. So, I began the contest using the same superhet that I entered last year. That receiver is completely unchanged, and you can find a description of it under the 2009 entries, or see the writeup on my webpage.

The direct conversion receiver has a tuned front end to help reduce interference from strong local stations and out of band signals. The signal enters the control grid and is mixed with the oscillator signal which appears on the deflectors. The result is an audio signal which appears as a differential signal on the plates. This audio signal drives the push pull primary of the interstage audio transformer. The secondary is connected back to the control grid in series with the incoming RF so that it is amplified reflex style. The resulting amplified audio is now a common mode (in-phase signal) at the plates, and so has no effect on the interstage transformer, and passes through the primary winding and on to the output transformer which is in series. The secondary of the output transformer drives the headphones.

Feedback for the local oscillator is taken differentially from the plates, through an audio blocking capacitor to the primary of the oscillator coil. One half of the secondary of the oscillator coil forms a tuned circuit with one gang of the variable capacitor. This side of the secondary drives one deflector, and the other half of the secondary produces an opposite phase signal to drive the other deflector. This oscillator is very stable and is almost completely immune to pulling towards the incoming RF. In fact this is a bit of a disadvantage, because it makes it very touchy tuning in AM signals without getting a beat frequency.

In the first version of this receiver, I had included feedback from the cathode to a one turn tickler on the RF coil, to provide a small amount of regeneration. However, after making a number of modifications and tests, I finally concluded that there was little to no benefit. So, I finally removed it. A description of the receiver is here.

There is a DPDT switch which reverses the phase of the interstage transformer secondary. This is to correct for slight unbalances in the tube which could cause oscillation when the volume control is turned all the way up. There is a bit of phase shift that varies across the band, and so the best position of the switch changes at different frequencies. I had tried various other plate balancing techniques (such as in my superhet receiver) but found that nothing worked as well as the reversing switch.

The receiver worked reasonably well. It was built as an experiment to see what could be done with a sheet beam tube. If you read the write up on my website, I sounded much more enthusiastic right after I built it. However, nearly two years later, and having built a better receiver in the interim, I find that listening to AM for any length of time on this set can be tiresome with the touchy tuning and almost constant beat tone. So, after using it for a couple of days and logging a respectable number of stations, I retired it and switched back to the superhet which outperforms it in almost every respect.

It seems that each year, I manage to learn a new trick or two. This year, I connected a set of wavetraps between my loop antenna pickup coil and the receiver. I wasn't sure how wall wavetraps would work with the loop. But, using the antenna orientation to null out a local station on 860kHz, and then using two of the wavetraps tuned to the same frequency to further null it out, I was able to log my first station ever on 870 kHz: WWL New Orleans, which by the distance/power formula was my second best catch of the contest. My best catch was KVNS 880, Brownsville TX. Attempts to log a MW station outside of North America have failed so far. Hopefully, one of these days it will happen.

Return to the index page.

Expect the unexpected
VISIT CLE

Select Menu

- Home
- About Us
- Contest Results
- 1AD Contests
- 1AD Entries
- Antenna Matching
- Antennas
- BHAM Radios
- Crystal Contests
- Crystal Radio Entries
- Sprint Contest
- Headphones
- Projects, Mods
- Favorite Links
- Contact Jack

Sub Pages

- Main Page
- 2004 Page 1
- 2004 Page 2
- 2004 Page 3

- Contest Results
    - 2004
    - 2005
    - 2006
    - 2007
    - 2008
    - 2009
    - 2010
    - 2006 Sprint
    - 2008 Sprint
    - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2009 Contest Entrants

Page 5

Dan McGillis, SW Class (regen set)

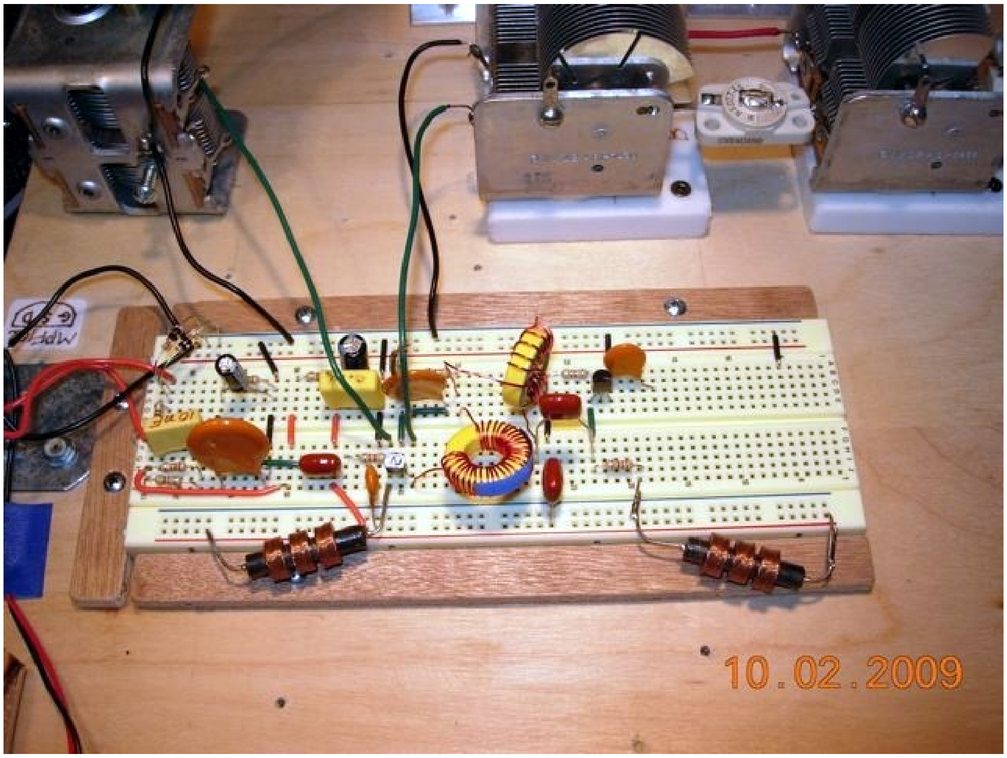

Thanks again for letting me enter my SW regen (with it's Radio Shack amplified speaker) in the '09 Sprint contest.

The contest gave me the needed nudge to really exercise the receiver and learn about it's strengths and weaknesses. I only tuned a lower CW portion of 40m (7.025 - 7.070 MHz) but there were hundreds of stations that could have been copied over the weekend period. Turns out there was a PA QSO party and a FISTS contest part of the weekend also. So there were weak and also extremely strong signals everywhere.

But I wasn't in the contest to score lots of points. I wanted to evaluate the receiver on weak & strong signals, so I restricted my station search to those calling CQ, especially the weak ones next to strong signals.

Click here for a link to Dave Schmarder's RadioBoard with a schematic and discussion of the radio.

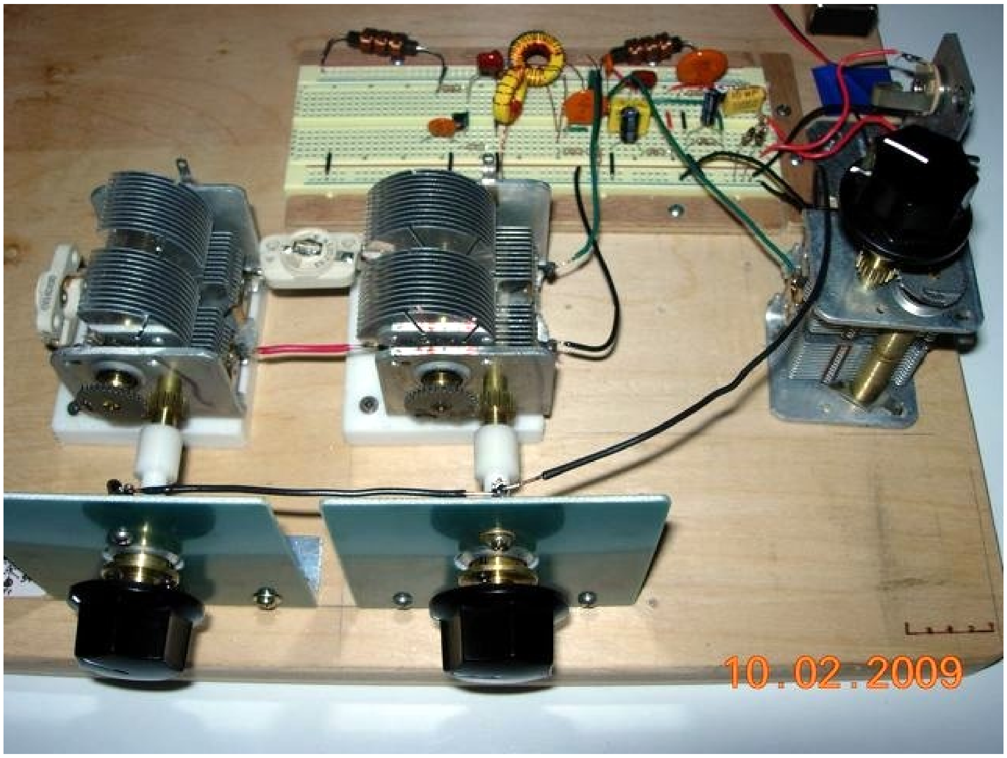

Dave Schmarder, BCB Class (crystal set)

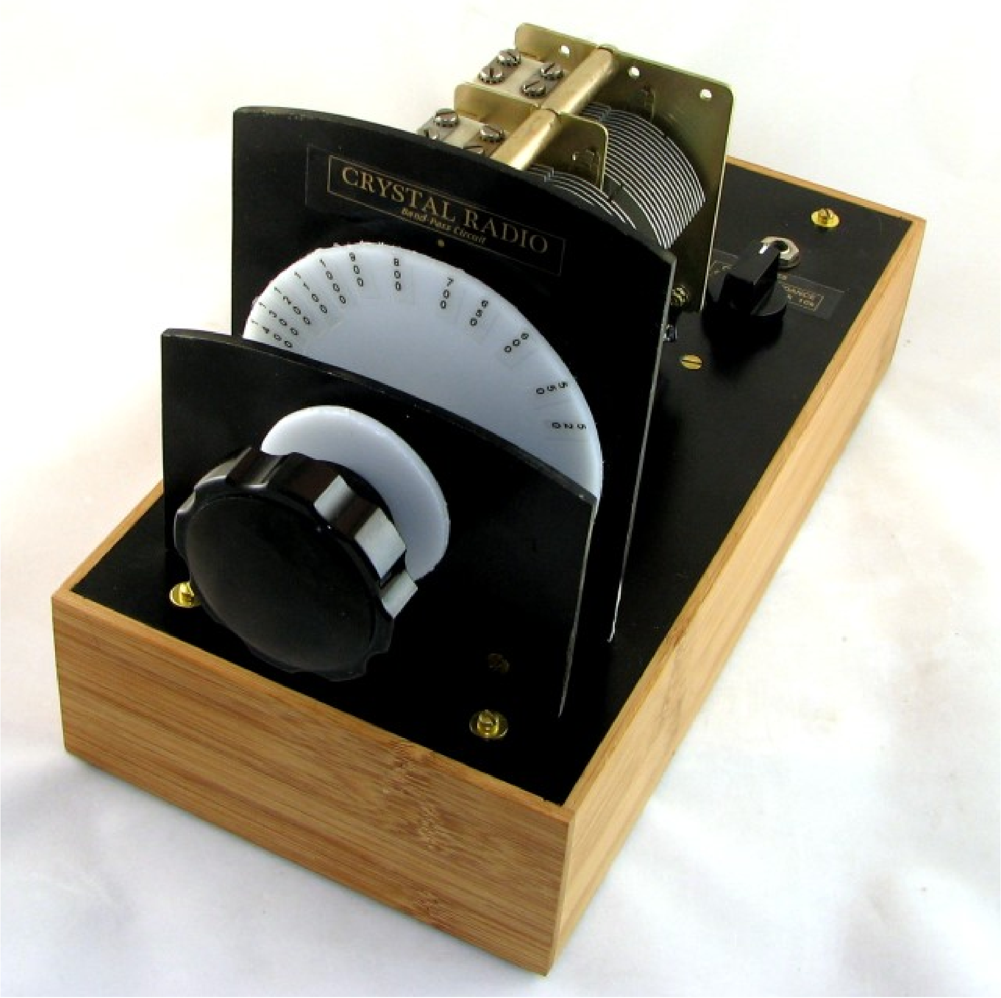
CRYSTAL RADIO
Band Pass Circuit

Here is the link to my contest entry. I just put up the Google map, so the page is now ready. I received 73 stations in the contest on my #76 crystal set. I had fun using my band pass radio. The outcome was better than I had expected.

Glen Yarbro, BCB Class (crystal set)

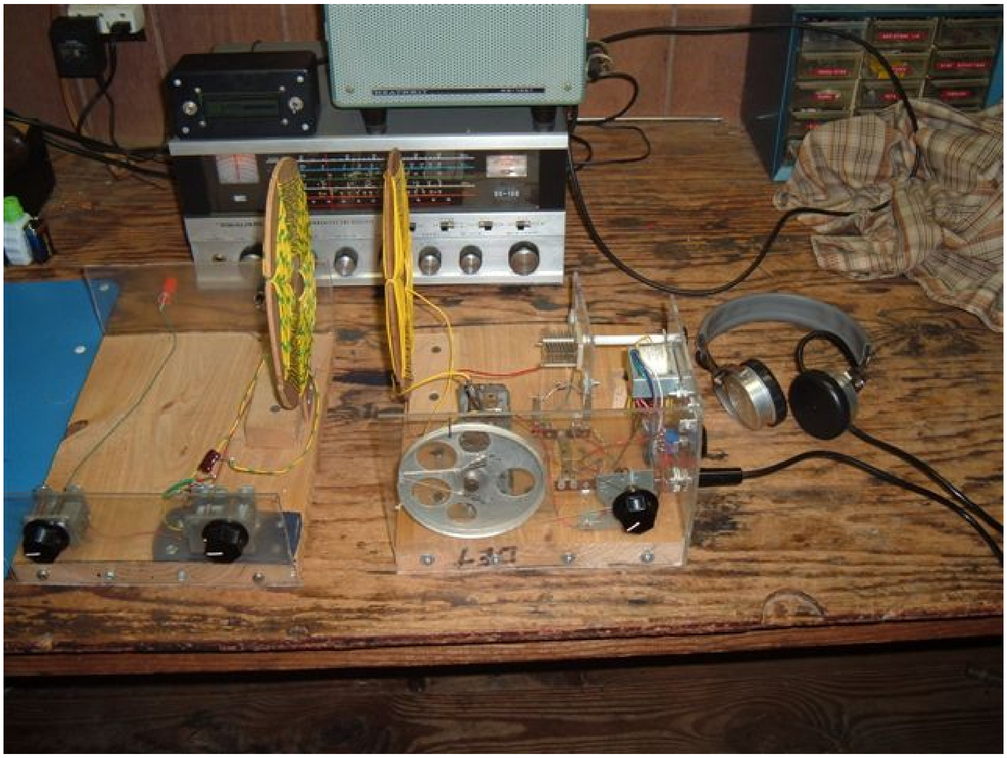

Dave's crystal radio #44
Headphones: Navy 4k
Diode: 1n34a
Antenna: 300 ft long wire

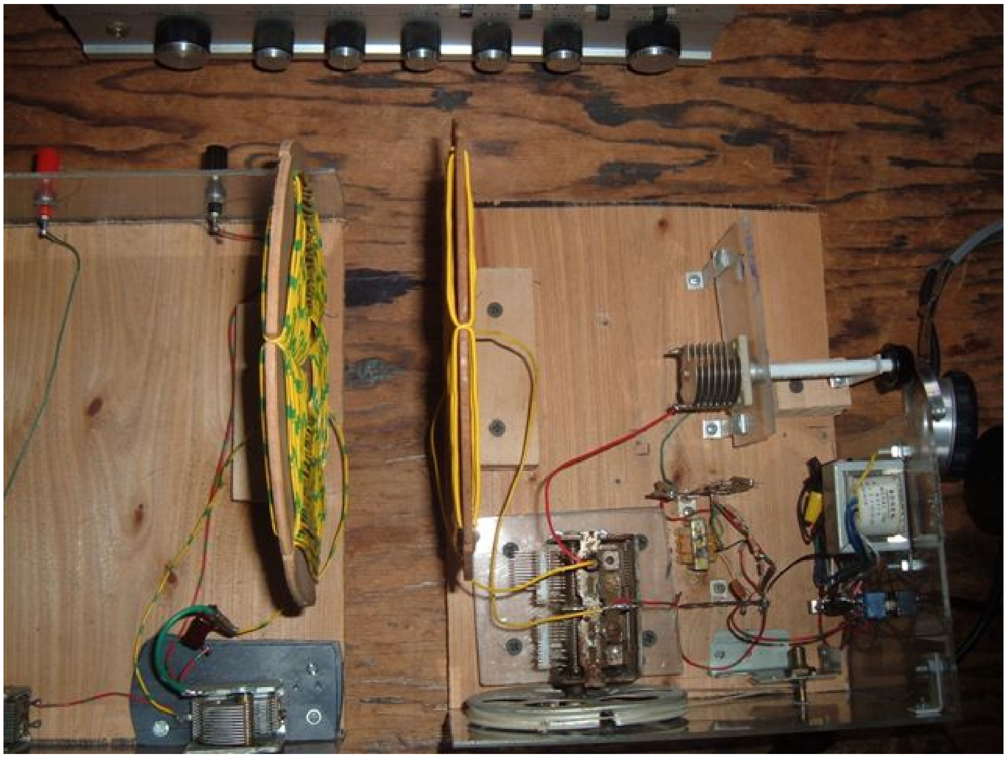

Gary Nichols, BCB Class (regen)

I was using a single J310 regen with a B+ of 4.5V. My tuning range was a bit below and above 680 and 1320. I forgot to measure it and I've already taken some of the set apart in preparation for mods.

The restricted tuning range was due to the 30-145 pF ARC 5 variable that I like to use because of the really slow and precise worm drive. The set was a bit of a last minute effort, so I gave up on extending the range. I had plenty of stations to listen to anyway. My problem was that it seemed more difficult than usual to ID stations based on the info broadcast. There were a lot of games on over the weekend evenings and also the usual "sports talk" stations which are as bad as Radio Disney to identify, it seems.

I also forgot to listen for my area stations during the day! Ooops!

The antenna was about 50 ft of wire sloping to the ground from a 2nd floor window (10 ft of it inside). I sloped it south-westerly for possilbe best performance from my location here in the northeast near Syracuse, NY. My ground was the nearby 2 draw metal filing cabinet. One time I sat down to listen and noticed more than the usual hand capacity. I quick check and I realized I hadn't hooked up the filing cabinet!

I used a "Bringhurst String" (5.6, 10, 22, 47 pF) of "clip selectible" fixed series caps between the antenna and top of the L/C to control signal strength and "pull" the tuning a bit up and down band.

James Kearman, BCB Class (regen)

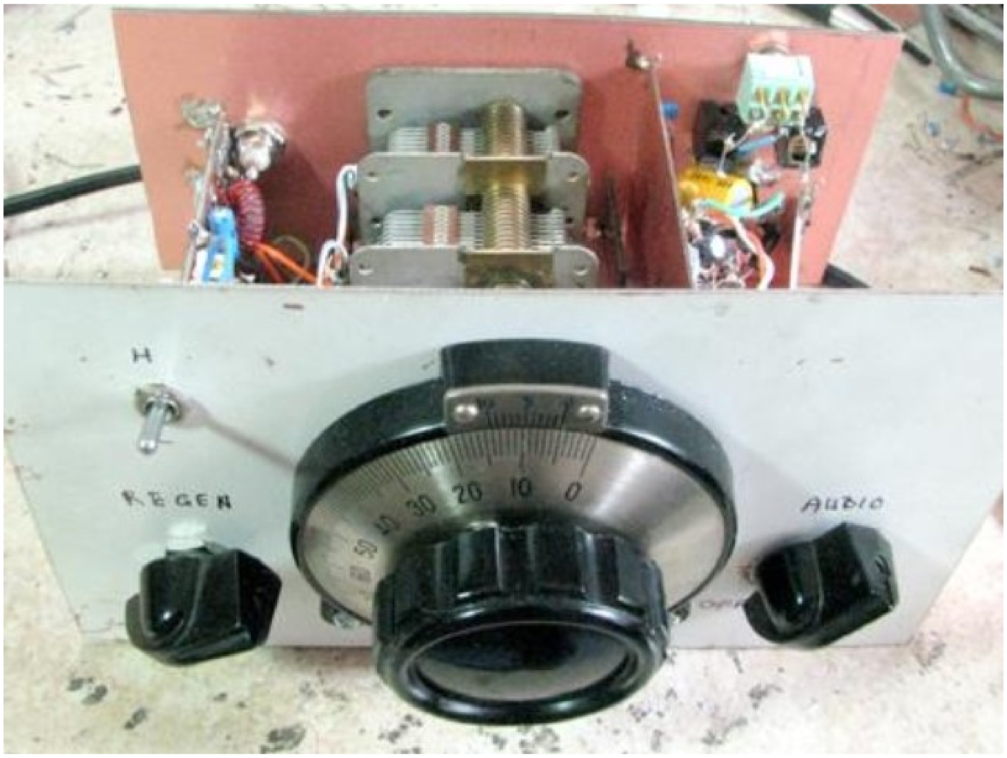

My receiver was a homebrew solid-state regen, the one shown here. I used a homebrew air-core loop, 14-foot circumference.

Sean Whiteacre, BCB Class (crystal)

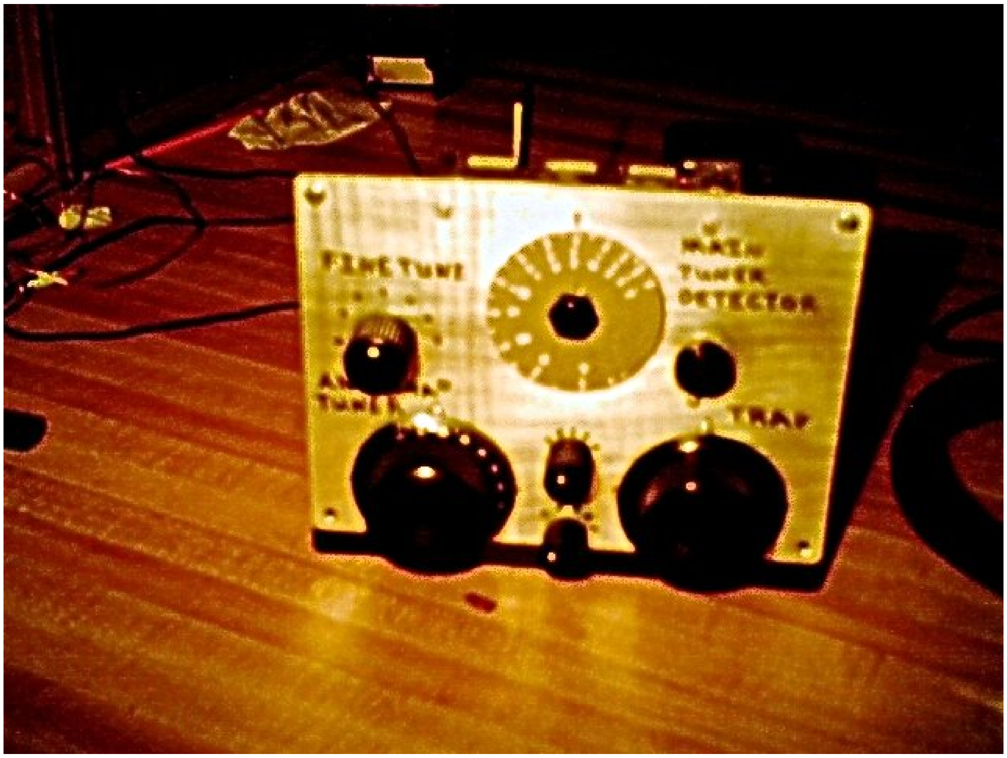
FINE TUNE
MAIN TUNER DETECTOR
TRAP

The diode is a FT-205 Type. I used Dynalec Headphones and RCA Large Can Sound Power headphones.

Jack Bryant, Two-way Shortwave Class (regen)

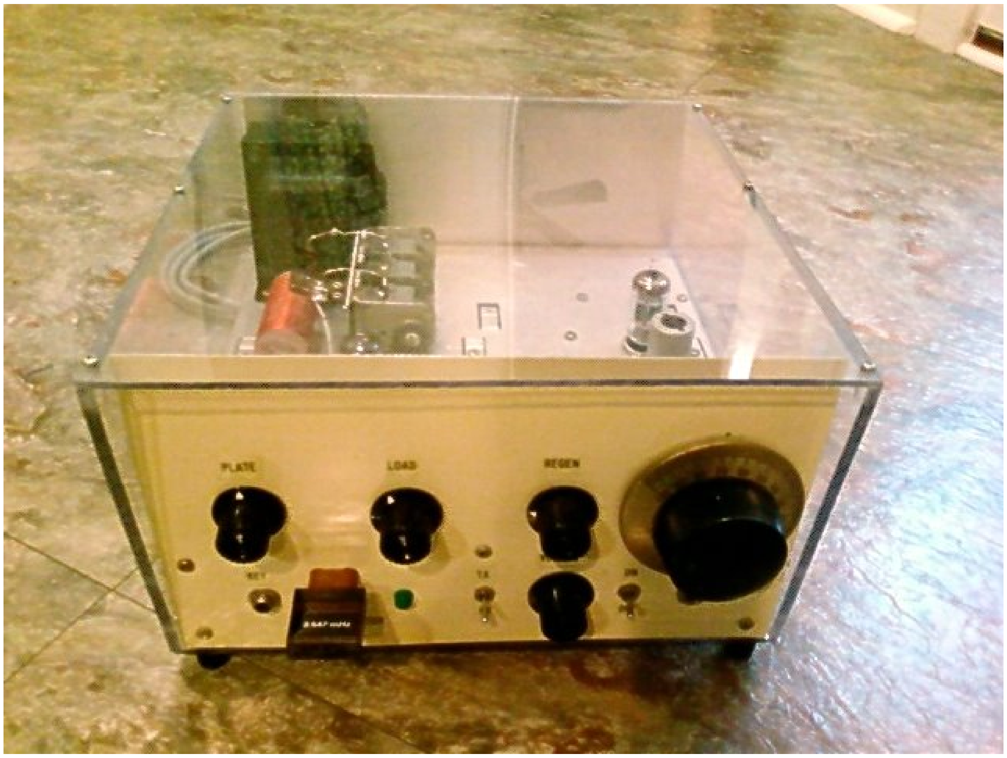

I participated just a bit in the contest. I started Sunday night and made a two contacts and then one more early in the morning. The set you see was "rescued" from the Shelby, NC hamfest a few years ago. It was beautifully made with a plexi-glass case and painted front. It was sold as a non-working unit. I hated to see this fine construction job go to waste. The rig had a two tube crystal controlled transmitter and a two tube receiver, all in one package.

The receiver did work but covered about a 3 MHz tuning range. The vernier was a tiny one. I swapped it out for a NOS Velvet Vernier. The tuning range was narrowed down considerably, now covering roughly 3.500-3.508 MHz. I changed the top coupled detector antenna cap to 1.5 pf. The plate voltage on the regen detector was reduced from about 130 volts to 22 volts. The power cord was changed to a three prong cord.

The change in receiver performance is dramatic. The receiver is now very usable. I wanted some additional selectivity, so I put a .005 uF cap across the primary of the audio output transformer. I fed the audio output into my PC and used the Spectrogram program to check out the bandwidth. The audio peak is between 400 and 800 Hz, just right for my ear. The tube lineup is a 6CY5 detector and a 12AX7 two-stage audio amp.

The transmitter initially did have some output, but it was acting strangely. It turned out that the pi-network output of the transmitter had to be redesigned. It worked, but with an output at 40 meters instead of 80 meters! Some turns were removed from the coil and new fixed capacitors were installed. The output variable cap had to be changed as well. I had to fabricate a mounting plate so the existing front panel hole could be utilized.

Now the transmitter puts out 10 watts (plus or minus). It is sensitive to the type of crystal used. For the contest I used a small surplus 3.547 MHz crystal installed in an old FT-243 case. It worked great. The transmitter uses a 6C4 oscillator and a 5763 power amp.

This rig looks a Novice station project, perhaps from an old ARRL publication; however, I have not found a schematic for it yet. The station does work well. It drives either phones or a small external speaker.

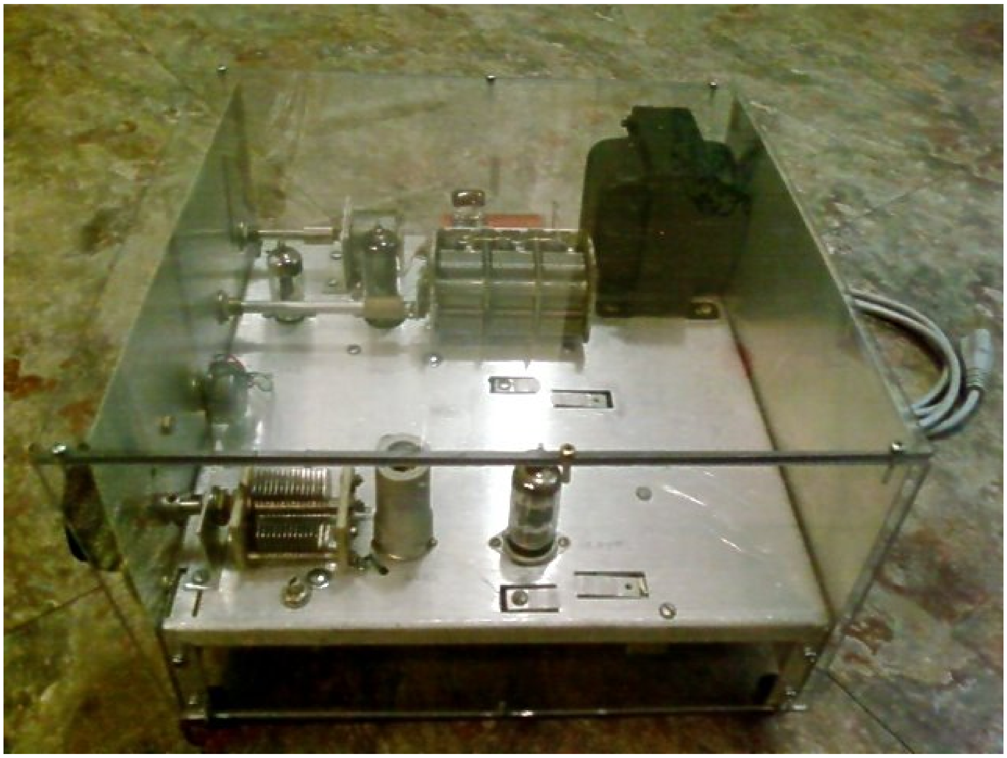

Go to 2009 page 6.

Select Menu

- Contest Results
    - 2004
    - 2005
    - 2006
    - 2007
    - 2008
    - 2009
    - 2010
    - 2006 Sprint
    - 2008 Sprint
    - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2006 Contest Entrants

Page 3

Steve Hewlett

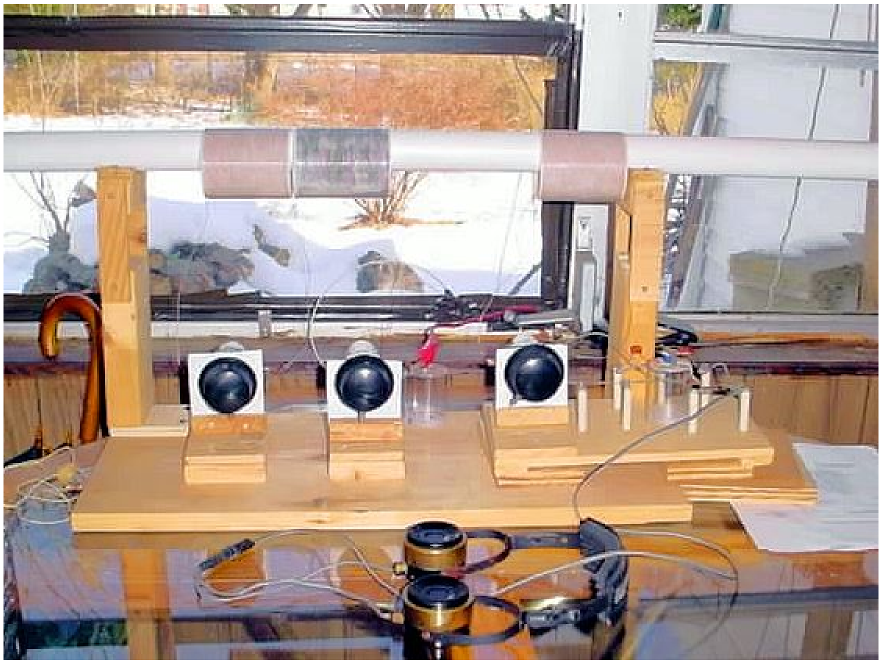

My 2006 contest set has three coils which slide on a polystyrene tube. From left to right they are wave trap, antenna coil, detector coil. All three coils are space wound (one wire width spacing between turns) copper wire; #24 ga. for wave trap and detector coils and # 23 ga. for the antenna coil.

A Single Jackson 365 pF VC is used on the wave trap coil and a single Jackson 410 pF VC is used on the detector coil. A dual Jackson 410 pF VC is used on the antenna coil in a Tuggle arrangement. The diode is a FO-215. A Calrad transformer and a "benny" (are included).

Tim Kilboy

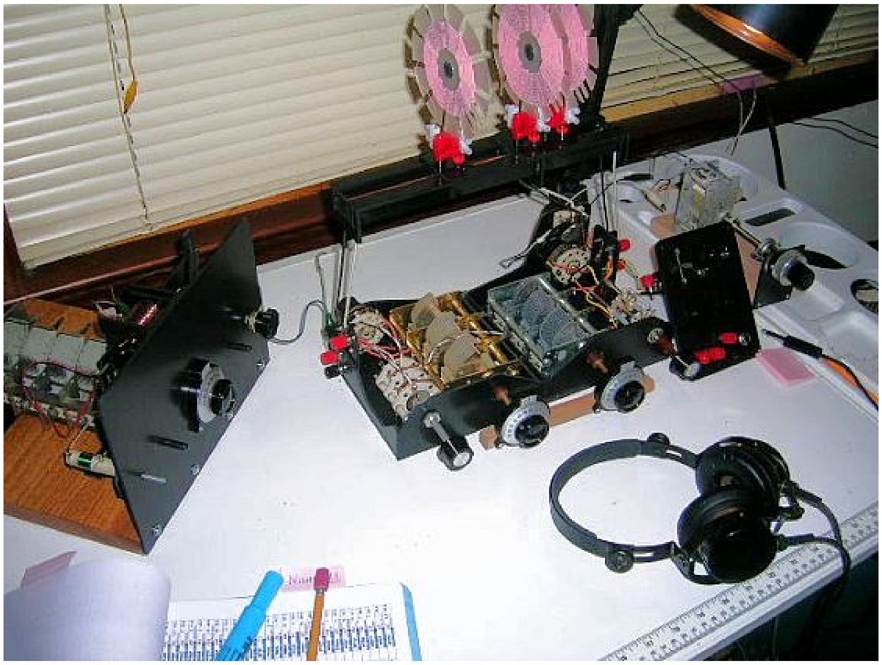

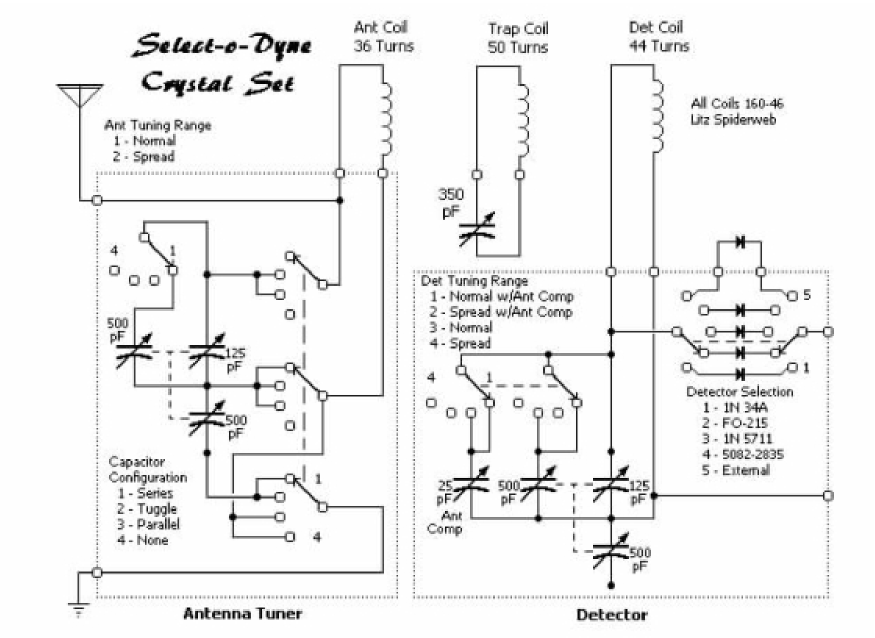
Select-o-Dyne
Crystal Set
Ant Coil
36 Turns
Trap Coil
50 Turns
Det Coil
44 Turns
All Coils 160-46
Litz Spiderweb
Ant Tuning Range
1 - Normal
2 - Spread
350
pF
4
1
500
pF
125
pF
500
pF
Capacitor
Configuration
1 - Series
2 - Tuggle
3 - Parallel
4 - None
1
4
Det Tuning Range
1 - Normal w/Ant Comp
2 - Spread w/Ant Comp
3 - Normal
4 - Spread
4
1
25
pF
Ant
Comp
500
pF
125
pF
500
pF
5
1
Detector Selection
1 - 1N 34A
2 - FO-215
3 - 1N 5711
4 - 5082-2835
5 - External
Antenna Tuner
Detector

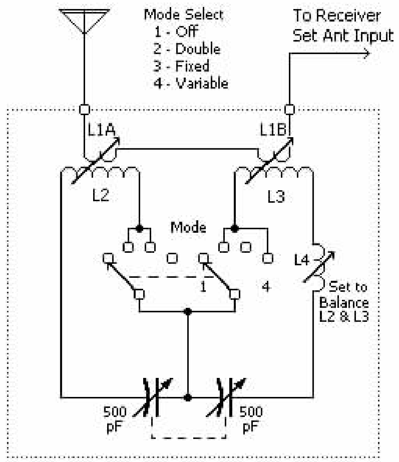
Mode Select
1 - Off
2 - Double
3 - Fixed
4 - Variable
To Receiver
Set Ant Input
L1A
L1B
L2
L3
Mode
1
4
L4
Set to
Balance
L2 & L3
500
pF
500
pF

In-Line Trap

L1A - 1/2 of 20 turn Binocular Coil
L1B - 1/2 of 20 turn Binocular Coil
L2 - 52 T on 3/8"x2" Ferrite Rod
L3 - 48 T on 3/8"x2" Ferrite Rod
L4 - 2-20 uH Ferrite Slug Variable Ind

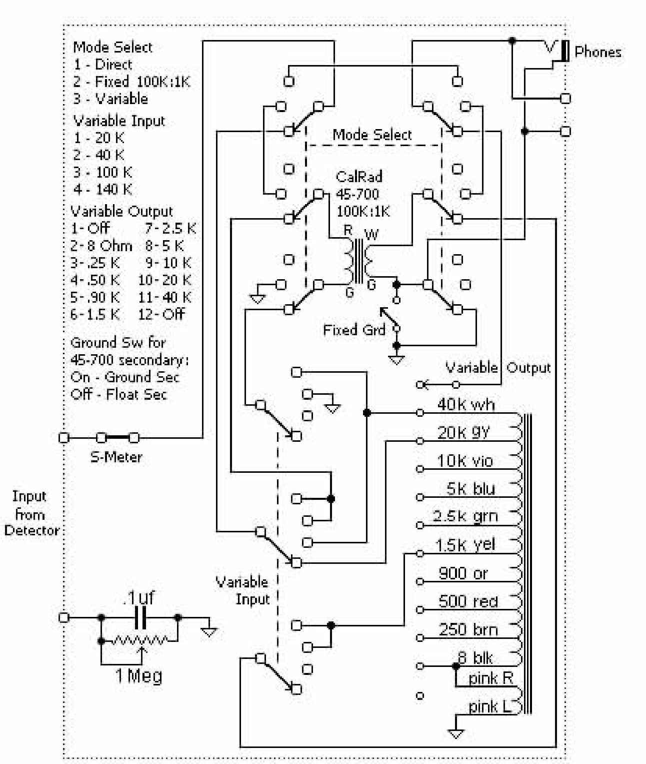
Mode Select
1 - Direct
2 - Fixed 100K:1K
3 - Variable
Variable Input
1 - 20 K
2 - 40 K
3 - 100 K
4 - 140 K
Variable Output
1- Off 7- 2.5 K
2- 8 Ohm 8- 5 K
3- .25 K 9- 10 K
4- .50 K 10- 20 K
5- .90 K 11- 40 K
6- 1.5 K 12- Off
Ground Sw for
45-700 secondary:
On - Ground Sec
Off - Float Sec
Phones
Mode Select
CalRad
45-700
100K:1K
R
W
G
G
Fixed Grd
Variable Output
40K wh
20K gy
10K vio
5K blu
2.5k grn
1.5k yel
900 or
500 red
250 brn
8 blk
pink R
pink L
S-Meter
Input
from
Detector
Variable
Input
.1uf
1 Meg

Impedance Matching

Wayne Thelen

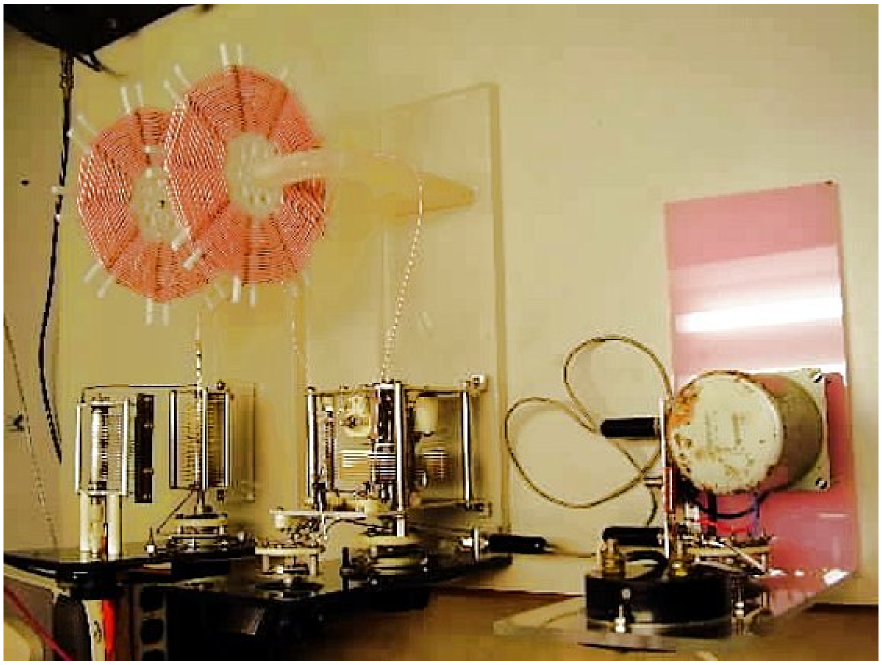

BLOCK DIAGRAM

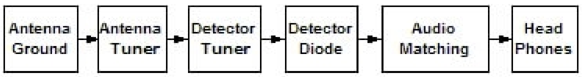

SCHEMATIC

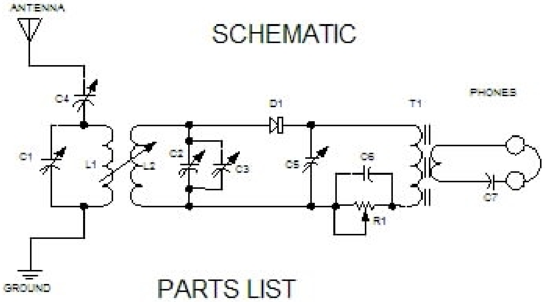

PARTS LIST

C1 - 435 pF Variable Capacitor with vernier drive
C2 - 506 pF Variable Capacitor with vernier drive
C3 - 0 - 50 pF Variable Capacitor, (trimmer)
C4 - 450 pF Variable Capacitor
C5 - 3 - 45 pF Variable Capacitor, (trimmer)
C6 - 0.1 μF capacitor
C7 - 0.47 μF capacitor
D1 - 1N34A Germanium diode
R1 - 500 K-Ohm pot
L1 - 36 1/2 TURNS 660/46 Litz wire on 9 pt spider web coil form; L = 160.3 μH; 5.70" O.D.
L2 - 37 1/2 TURNS 660/46 Litz wire on 9 pt spider web coil form; L = 171.1 μH; 5.85" O.D.
T1 - VINTAGE RCA Input Transformer - P=250 K Ohms CT ; S=600 to 200 Ohms
Phones - (2) U. S. Instruments UA1614 "Sound Powered" UA1614

Full details can be seen at Wayne's Website

Jack Bryant

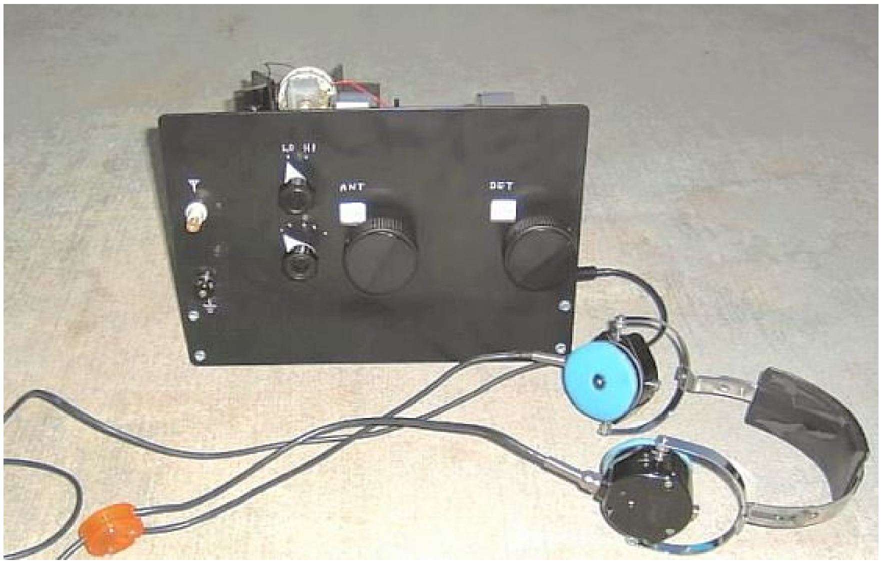

I used the same set in the BCB Open Class Category for 2004, 2005, and 2006. I tried the Hobbydyne in my toroid set, but my set worked better without it...too much signal loss. The selectivity of the set was already pretty good.

I did use an outboard cap this year in series with the set's built-in dual section cap. I liked it so much that I will probably modify my set and put the cap inside. Part of the reason for this is that my antennas are large with a lot of capacitance to ground. The standalone antenna cap adds a lot of flexibility.

Another mod to the set was a neon bulb connected from antenna to ground. This year we had some pretty good thunderstorms during the first part of the contest. When the bulb lit the room up, well, it was time to shut down! (I exaggerated just a little bit here.)

I did encounter a problem that I will mention, and perhaps you can avoid it. I wear glasses, and the constant pressure of my headphones against the ear piece of the glasses gave me a terrific headache. I apparently injured a nerve since it took me several weeks to recover. I am wondering if it is time for laser surgery so I can eliminate those glasses. You notice I didn't say stop using headphones…

Gil Stacy

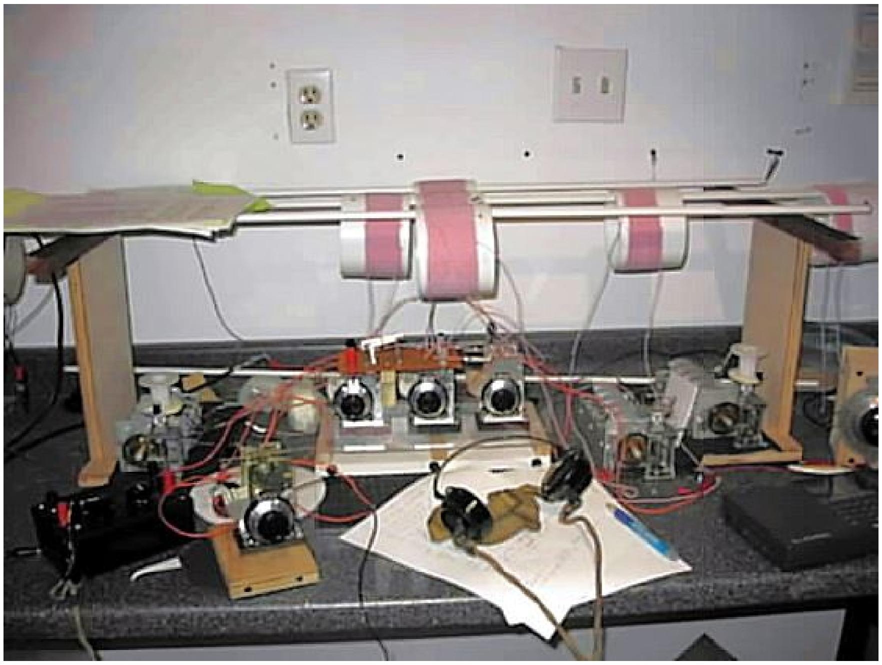

Set: Same set as last year. Headset: RCA Big Cans tuned> and recharged by Steve Bringhurst.

Antennas: 88’ top loaded vertical up 50’ and 180’ inverted “U” up 50’. Ground is the post 2004 DX DX contest ground described at the bottom of this page Set was the same as used in 2005. The antenna was a top loaded vertical cage antenna.

Lem Morrison

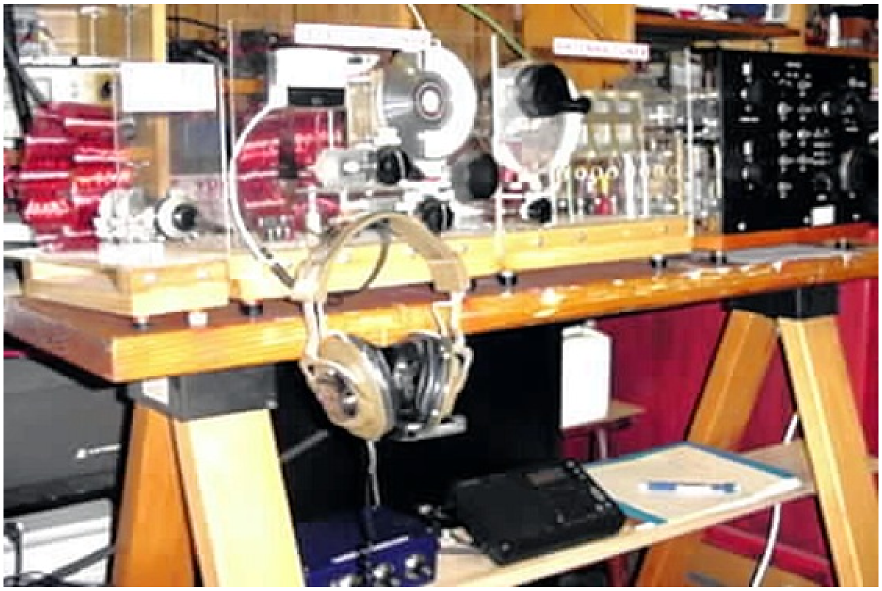

Set was the same as used in 2005. The antenna was a top loaded vertical cage antenna.

Alex Perez

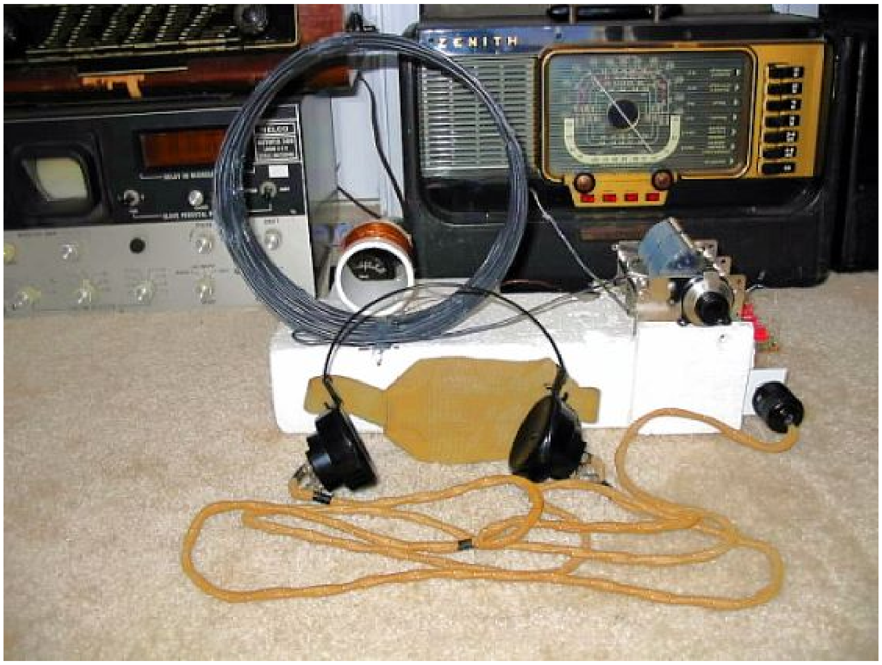

Crystal Set DX 2006 Receiver

Despite the shortcomings that befell my design last year, the exact same setup is employed this year. The antenna spans a length of approximately eighty feet in a northeast to southwest orientation, rising to a height of forty five feet at the far end from the initial twelve feet. A convenient cold water pipe serves as a ground and is connected with a fifteen foot length of wire running horizontally. In order to minimize losses, an eight inch diameter twenty turn coil consisting of fourteen gauge wire is utilized and tuning is accomplished with a high quality vernier driven variable capacitor across the coil. A small ceramic trimmer loosely couples the antenna to the top of the tank circuit. Last year, the advantages of Schottky diodes were realized and a single BAT28 performs admirably in the design. A high impedance transformer matches the high impedance detector output to the low impedance DLR No. 5 headphones. These sound-powered headphones were most graciously provided as a prize by the contest committee two years ago and were instrumental in greatly boosting the DX capability of the receiver. A simple inductively coupled trap must be in place at all times. At night, it opens up a few hundred kilohertz around WXEM while during the day, it enables the reception of weaker stations across the band.

Band conditions seemed different this year but proved to be advantageous. After considerable experience gained during the preceding contest and an overall greater familiarity with band conditions, I promptly logged the majority of the typical nighttime band markers. The ensuing evenings resulted in only a few new nocturnal DX catches. Regrettably, I should have exploited the terrific grayline conditions to a greater extent this year. Nonetheless, many stations and points were gained due to a more casual approach to grayline DXing. Typically DX within fifty kilohertz of WXEM is impossible, even at night. Nonetheless, stations were logged on 1410, 1500, and even WXEM’s own frequency of 1460 when it decided to completely shut down for a short while. Following in the pattern set by my previous participation, I did surpass my score considerable while greatly enjoying the Contest.

[Back to the main page.](http://www.crystalradio.us/)

Ad
Mini A8 Real Time Car GPS Tracker...
$9.67 Specifications: This GPS tracker applies the newest technology with the following...
ebay

Select Menu

- Home
- About Us
- Contest Results
- 1AD Contests
- 1AD Entries
- Antenna Matching
- Antennas
- BHAM Radios
- Crystal Contests
- Crystal Radio Entries

- Contest Results
  - 2004
  - 2005
  - 2006
  - 2007
  - 2008
  - 2009
  - 2010
  - 2006 Sprint
  - 2008 Sprint
  - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2007 Contest Entrants

Page 1

Mike Tuggle

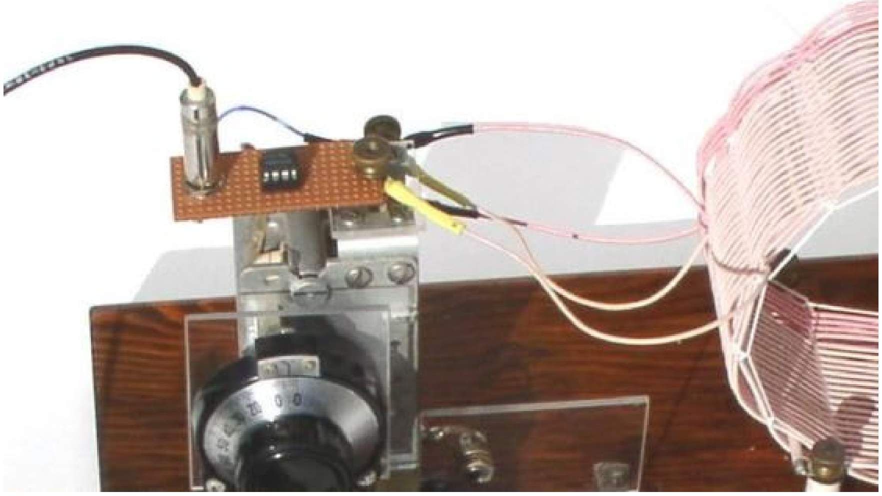

This detail shot shows the original Lyonodyne C2 and L2. The entire set can be seen below.

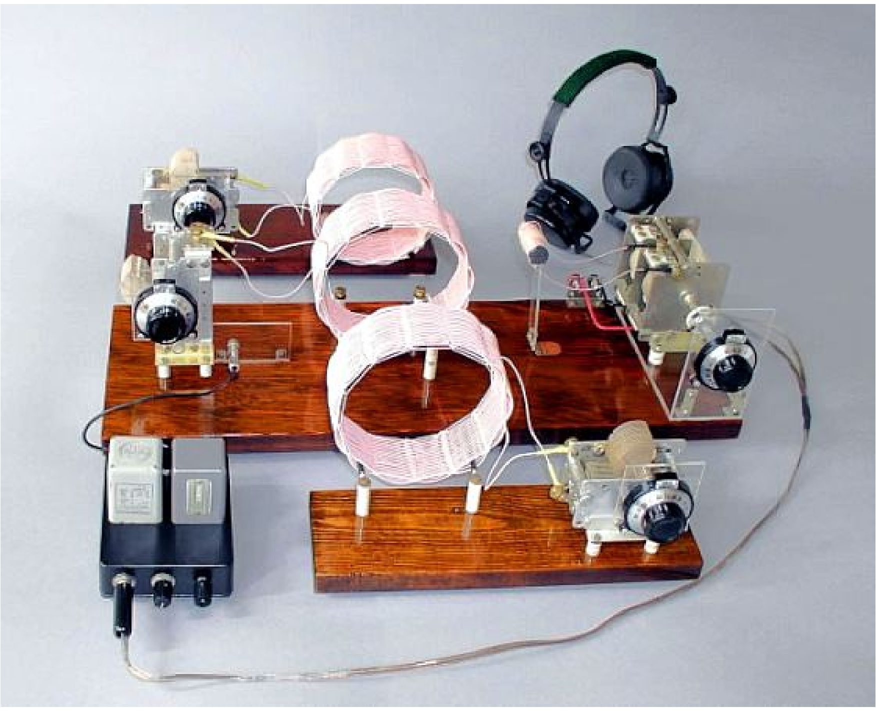

The MOSFET detector is mounted on its own circuit board along with its output jack. This board mounts to the set using the thumb nuts that attach L2 to C2. L3, a darker shade of pink, is mounted in back ('cold' end) of L2. The original diode output at bottom-center was not used.

To me, the most impressive feature of the ALD MOSFET detector was its superb selectivity. This makes sense, since the GATE presents essentially a zero load to the tuning tank, thus preserving its Q.

Volume, especially on weaker stations, was noticeably less than that delivered by the single diode detector. This may be partly due to the less-than-ideal matching between DRAIN-SOURCE and the audio matching and output circuits. Didn't have much time to play around with matching before the contest started.

Many thanks to my 'Benefactor' who provided the much-needed ALD110900A's during our recent national shortage.

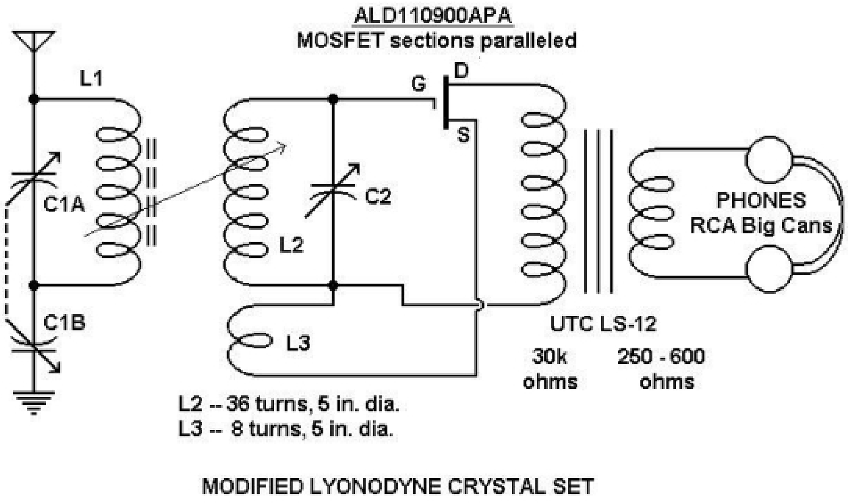
ALD110900APA
MOSFET sections paralleled
L1
G
D
S
C1A
C2
L2
PHONES
RCA Big Cans
C1B
L3
UTC LS-12
30k ohms
250 - 600 ohms
L2 -- 36 turns, 5 in. dia.
L3 -- 8 turns, 5 in. dia.
MODIFIED LYONODYNE CRYSTAL SET

For my 2007 XSDX set I adapted one of the popular ALD110900A zero-voltage-threshold MOSFET detectors to the Lyonodyne. The highest voltage, at the top of the secondary tank, is applied to the MOSFET GATE. An added-on 8-turn coil L3, closely coupled to the 36-turn secondary coil, feeds the DRAIN-SOURCE and audio load.

Evan Haydon

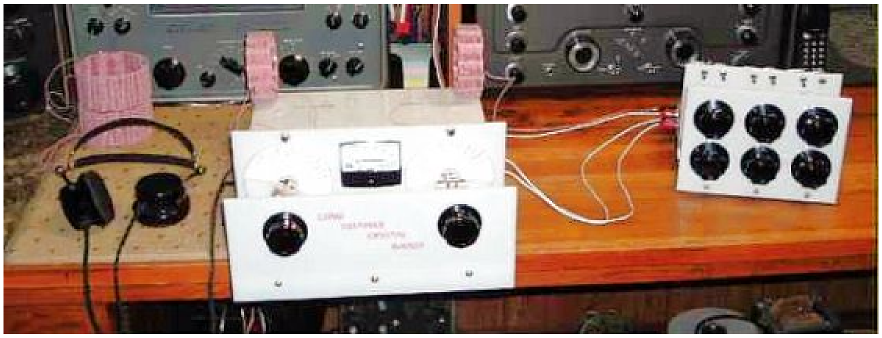

Another fun Elmer Memorial Crystal Radio DX Contest is over. I used the same crystal radio that I used last year with only a couple of improvements. I bought some 420/46 litz wire for my coils. The Q of the coils used in the crystal radio is now 800 from 550 khz to 1200 khz. The Q then tapers down to 600 at 1700 khz. My six toroid traps are all wound with the same litz wire and have a Q of 400.

My antenna is the same as last year except that both ends are now up 40 feet. I am now using a good set of Big Cans for headphones with a bogen T725 transformer for matching to the radio.

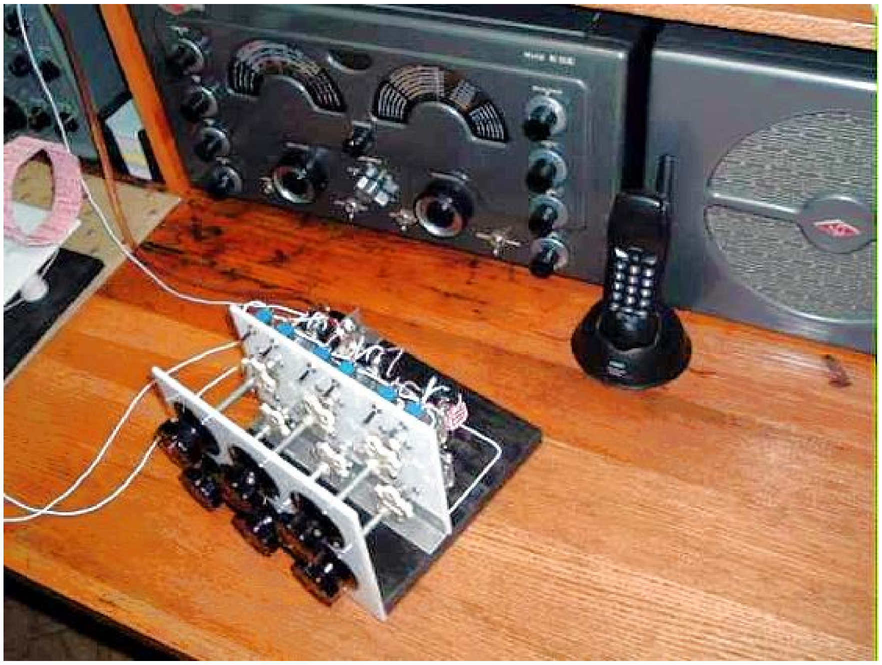

I did my usual listening to the broadcast band almost every evening and some mornings from the middle of September up to contest time. I made a lot of notes. These paid off in time saved finding and identifying the largest number of stations possible during the contest. It is amazing how fast some stations appear and fade away at the gray line. One has to be on the right frequency at the right time to catch them.

The band conditions here were very good the first day of the contest. The second day was really pretty good too. After that, the conditions were generally poor with some spotty openings. Here are some of my statistics for the contest.

Day	Hours spent	Stations found
1-12	15.5	166
1-13	9	34
1-14	11	23
1-15	12	19
1-16	11	19
1-17	6	8
1-18	7	8
1-19	6	3
1-20	5	5
1-21	8.5	4
	91 hours	289 stations

Some of the stations heard during the contest:

WWRV 530 khz South Caicos Islands 1987 miles
CMQ 670 khz Cuba 1483 miles
CM 530 khz Cuba 1481 miles
CKWX 1140 khz Vancouver, BC. Canada 1400 miles
KNBR 680 khz San Francisco, CA. 1379 miles
WBZ 1030 khz Boston, MA. 1333 miles

My personal choice for best catch during the contest:
KGBC 1540 khz Galveston, TX. 250 watts 804 miles
Good copy for an hour on the evening of Jan 19.

On an average winter day I can always hear about 55 stations. Most are within 175 miles. One everday station heard is 5 kw at 247 miles. My all time list of stations heard on my crystal radio that ID in English is now 412 stations. Let's do it again next year.

Bob Jewell

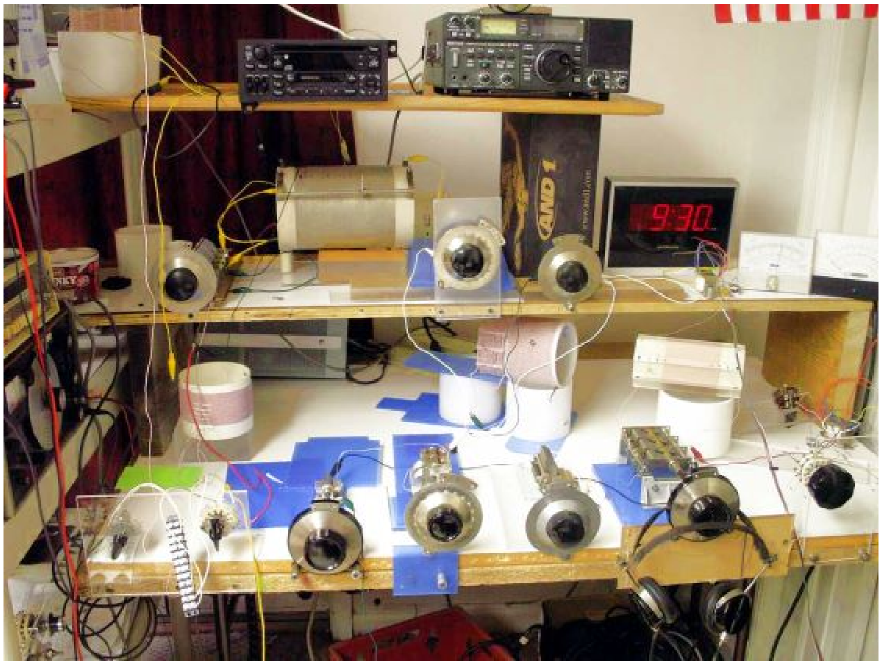

ANTENNAS:

1.Long Wire approx. 200 ft long, up 80 ft, lead-in approx. 100 ft, all one-piece litz to rig.

2.Four element 20M colinear antenna, fed with 300-ohm feeder, used as three antennas. Up 85 ft. 3.Misc. Spotter small antennas.

ANTENNA TUNER:

Seven inch diameter, 6.5 inches long, ceramic form, space wound, 90 turns, bare silver plated wire, selectable on each turn.

Capacitor: 4 section, 570 pf each, Variable, ceramic insulated silver plates. Wired series and parallel.

RIG:

1.Antenna coil: Large (?) litz on 4.5 inch diameter opaque plastic pipe form, close wound. Detector: 185/46 litz on 3.5 inch diameter ceramic form, space wound. Q ~ 650. (HP-4342A)

2.Grounds (2): house ground system, 200 gal LP tank underground, water filled. Connected with 50 ft of RG-8 coax, inner conductor and braid paralleled.

3.Diodes (22): different TYPES, on ceramic rotary switches, instantly selectable. Favorite soldered in.

4.Capacitors: ceramic insulated, silver-plated, approx. 550 pf.

5.AUDIO: Bogen T-725 plus 200K to 1000 ohm
45-703, miniature transformer.

6.HEADPHONES (3): USI, 1247-A, mike elements wired in series. (2) N.I.B. (Ebay, no name) soundpowered headphones, one set series, one set parallel wired.

7.All traps, (3), ceramic insulated caps and litz coils.

8.“Tuggle” type circuit.

9.“S” meter: Olson TE-380 meter, 0-8.7 ua,
Heath: 0-200 ua, series connected.

10.All internal RF wiring 185/46 litz.

11.ALL components mounted on Plexiglas sheets.

12.Many “Classic” vernier dials.

ACCESSORIES:

1.Icom R-70, spotter
2.Chrysler van auto radio, spotter
3.Icom PCR 100 with PCR 1000 software, spotter.
4.B&K Signal Generator, E200-D, frequency verification.
5.B&K, 1851, Frequency counter , frequency verification.
6.H.P., 4342A, Q meter. (max Q ~ 1000).
7.Heathkit, QM-1, Q meter. (max Q ~ 500).

Charles Pullen (Chuckster)

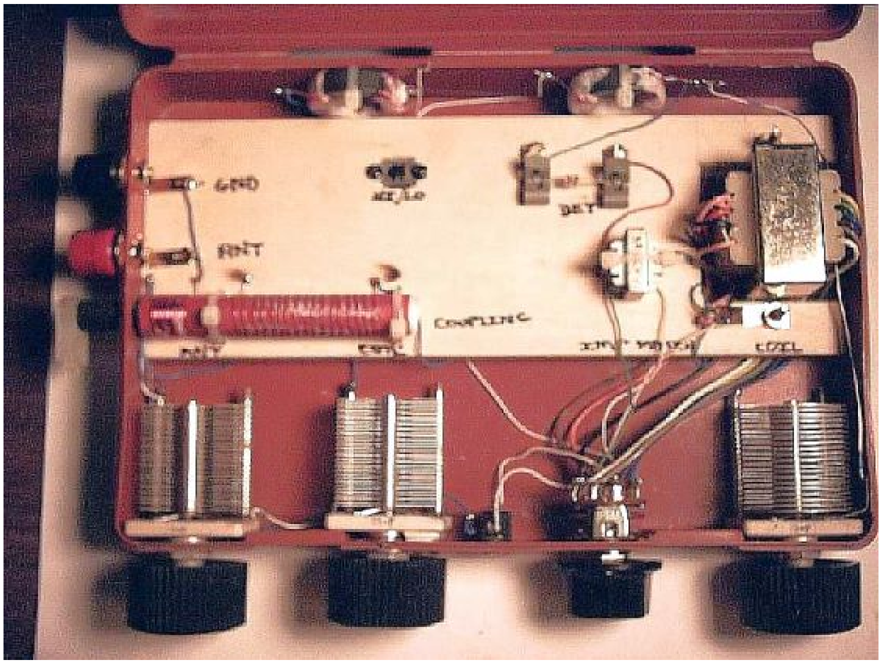

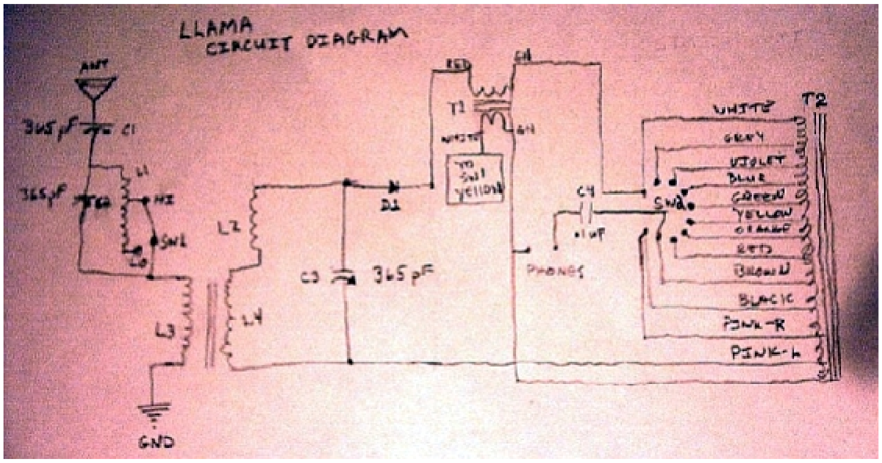
LLAMA
CIRCUIT DIAGRAM
ANT
365pF
C1
365pF
C2
L1
HI
SW1
LO
L2
L3
L4
GND
C3
365pF
D1
RED
T1
WHITE
TO SW1 YELLOW
GN
C4
.1uF
PHONES
SW2
T2
WHITE
GREY
VIOLET
BLUE
GREEN
YELLOW
ORANGE
RED
BROWN
BLACK
PINK-R
PINK-L

This circuit is a hybrid of Steve Bringhurst's select to match circuit and Dave Schmarder's toroid set #45. The design goal for this radio was high performance in a small package. The case is 9 inches x 6 inches deep x 1 3/4 inches tall. The toroids are FT 114A-61 amidon cores, L1 tapped at 12 turns. The toroid coupling coils are wound on a small tube with an ID slightly larger than 3/8 inch. The first coil is close wound, 15 turns. The second coil is triple spaced, seventeen turns. Both coils 24 awg and are approximately 29 uH with coupling accomplished by moving a 3/8" x 4" ferrite rod in and out of the coupling coil's tube. The detector used was a pair of HP 5082-2835 schottky diodes. The phones used was a pair of Philmore xtal earplugs in parallel thru a radio shack mono Y adaptor. The antenna used is a half sized g5rv ham antenna approximately 25 ft up, center conductor & braid of the coax connected together. Ground system is 3 four foot copper clad steel rods 8 ft apart connected with #14 awg solid bare copper wire.

The detector cap sometimes has a very slight but noticeable bit of hand capacitance at times, it seems to depend on the weather. All in all a good performing radio.

Dan McGillis

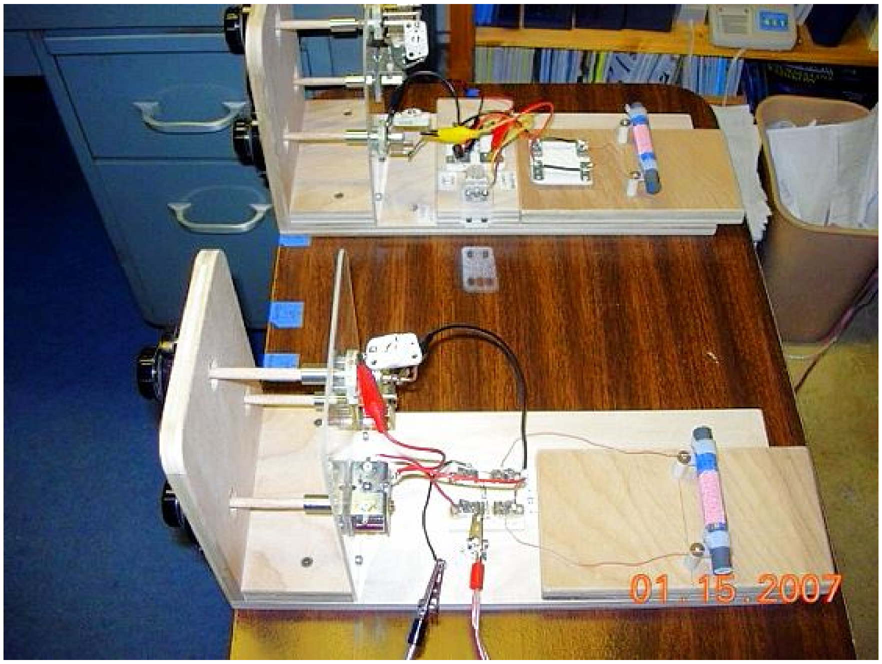

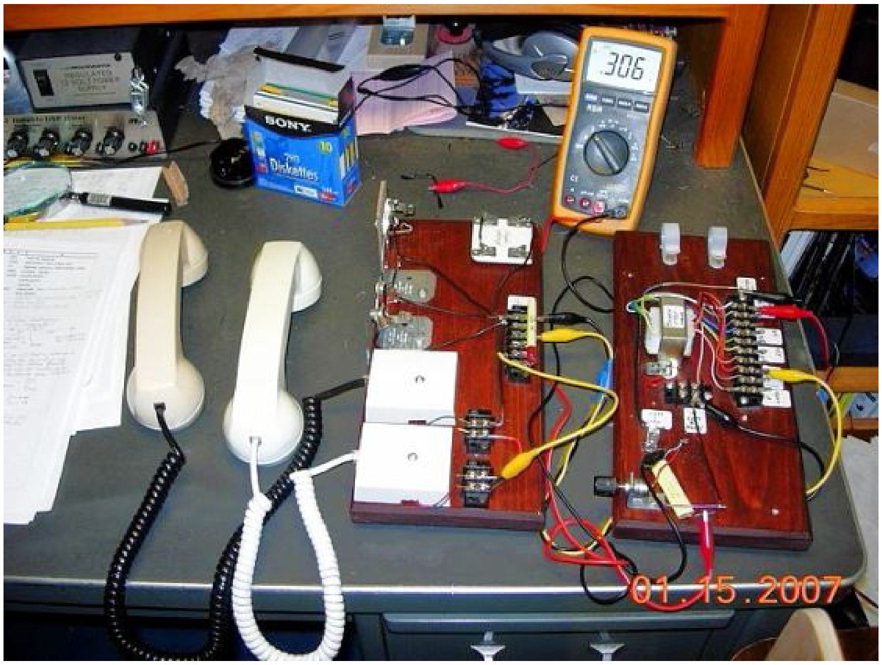

For the 2007 Crystal Set DX Contest, I used an all ferrite-rod double tuned set-up: Detector - ATU - Wave Trap.

The inductors are Amidon Ferrite 61 0.5”x4” ferrite rods with 165/46 Litz coils wound directly onto the rods. All capacitors have ceramic insulators. Detector tuning is split into 3 ranges to cover the broadcast band. Each range has it’s own capacitor, trimmer and 8:1 vernier dial. The “Tuggle tuned” ATU is split into two ranges - each with its own capacitors and 8:1 dials. The inductively coupled wave trap has one range and a 48:1 vernier. A FO-215 diode is Hobbydyne coupled to the top of the detector coil - no taps. Diode output goes to a Bogen T-725 transformer then to two USI UA-1614 sound powered elements in series - or - standard telephone handsets in series.

The antenna is an end-fed 200’ inverted Vee with it’s apex at about 40’; ground is a 4’ pipe driven into the ground.

Dial number-to-frequency calibration curves for the detector and ATU were converted to vernier dial numbers for every 10 khz of the broadcast band. A station heard on a digital read-out Yaesu FT-757GX spotting radio could be quickly converted to crystal-set dial numbers - and vice-versa.

I think the bandspreading and calibration expansion are the best features of this set-up.

Dave Schmarder

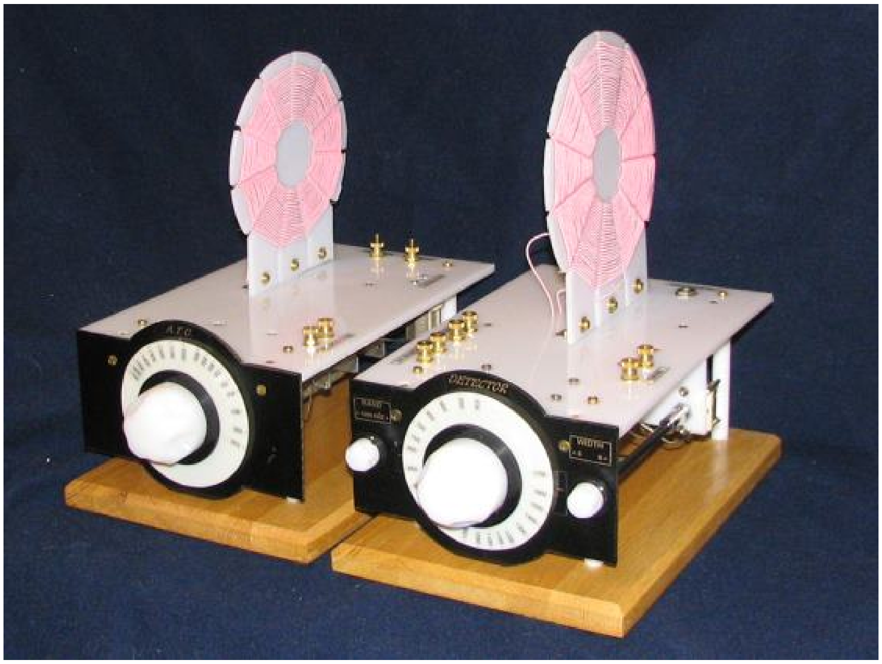

- Contest Results
  - 2004
  - 2005
  - 2006
  - 2007
  - 2008
  - 2009
  - 2010
  - 2006 Sprint
  - 2008 Sprint
  - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2004 Contest Entrants

Page 2

Dan Petersen

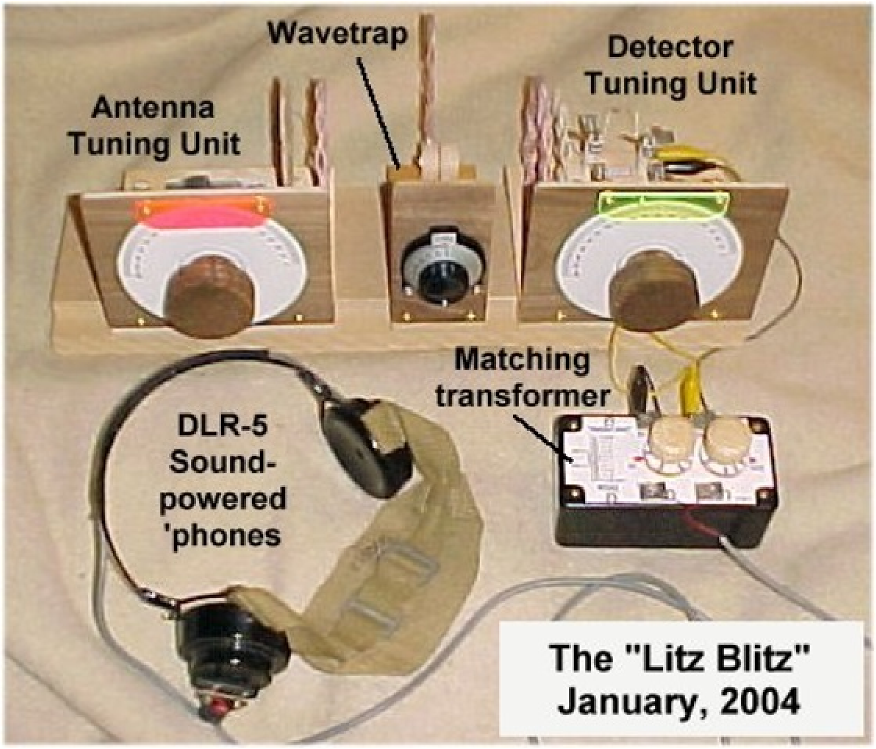
Wavetrap
Antenna Tuning Unit
Detector Tuning Unit
Matching transformer
DLR-5 Sound-powered 'phones
The "Litz Blitz" January, 2004

Here is a picture of Dan's set, the "Litz Blitz". This set is in the XSS Newsletter as an article in the March, 2004 issue.

Brian Wingard

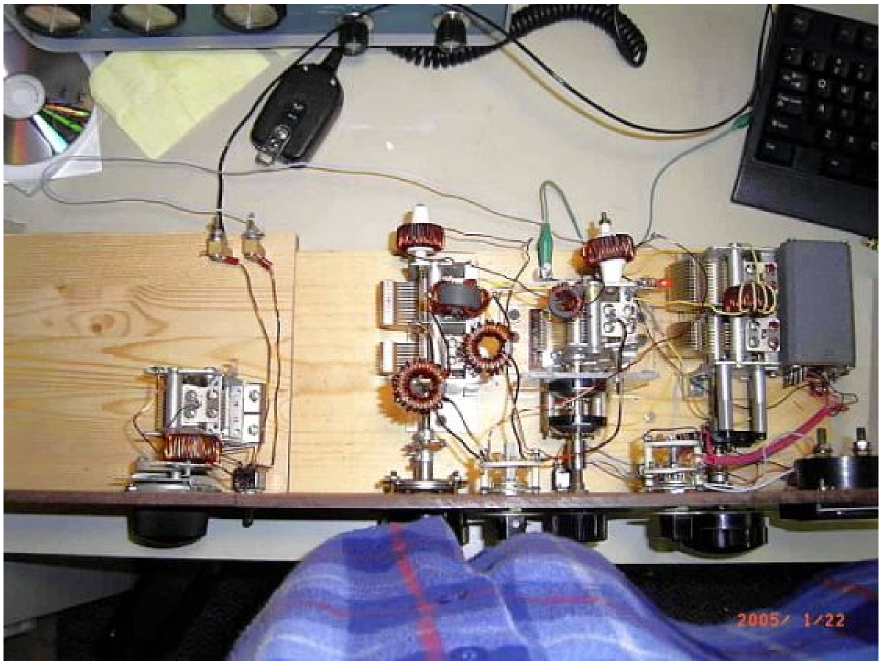

Notice the FT114-61 cores and large diameter wire. Brian uses separate cores for Lo Band/Hi band.

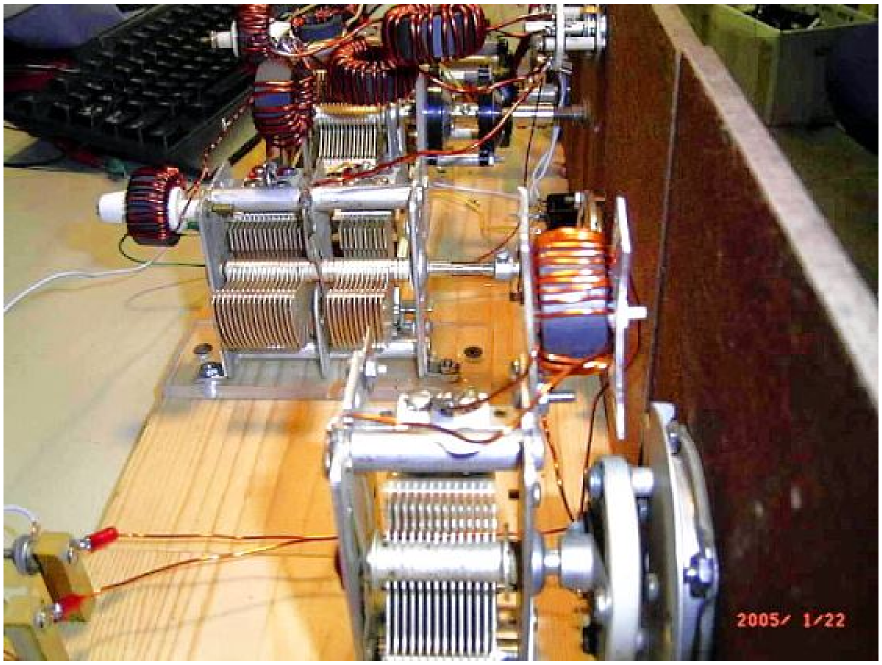

Steve Bringhurst

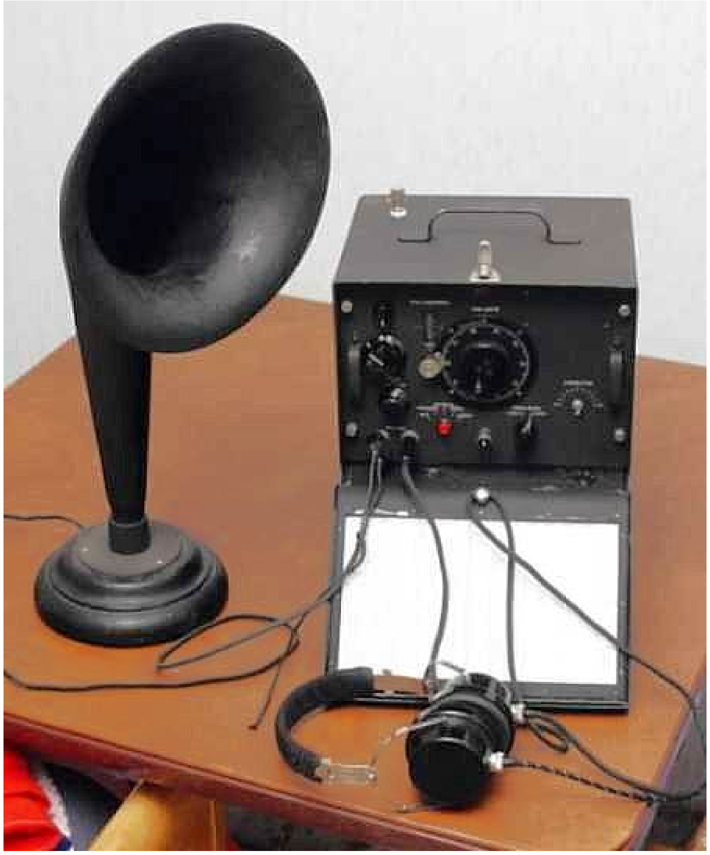

Steve uses a BC-221 as the basis for his BCB set. The BC-221 is a frequency meter, WWII vintage.

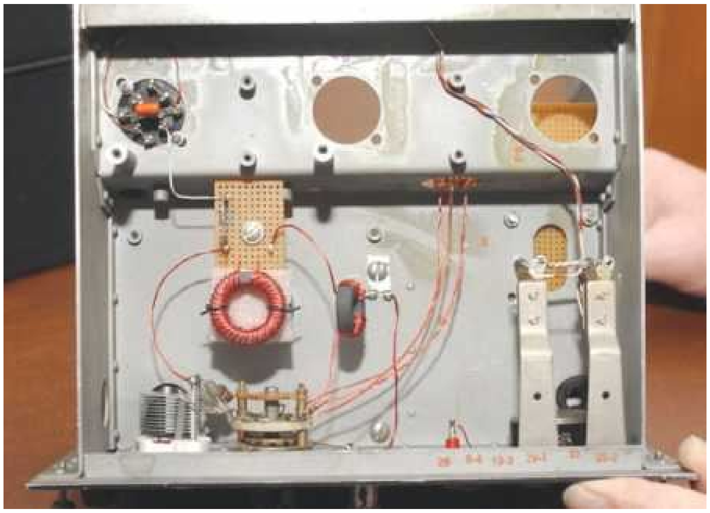

Mike Tuggle

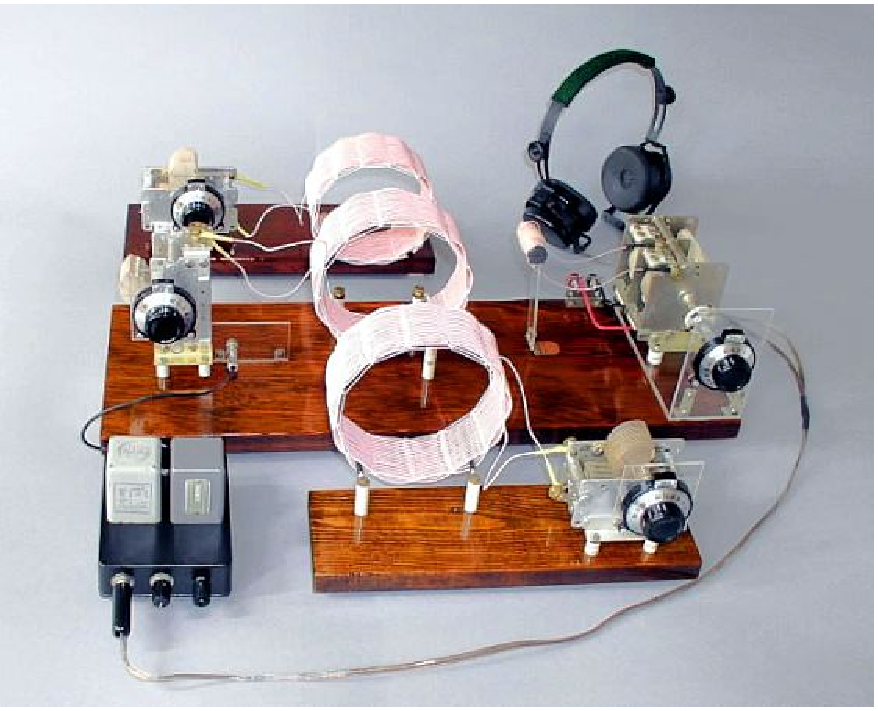

Mike Tuggle's record smashing DX set.

Mike Branson

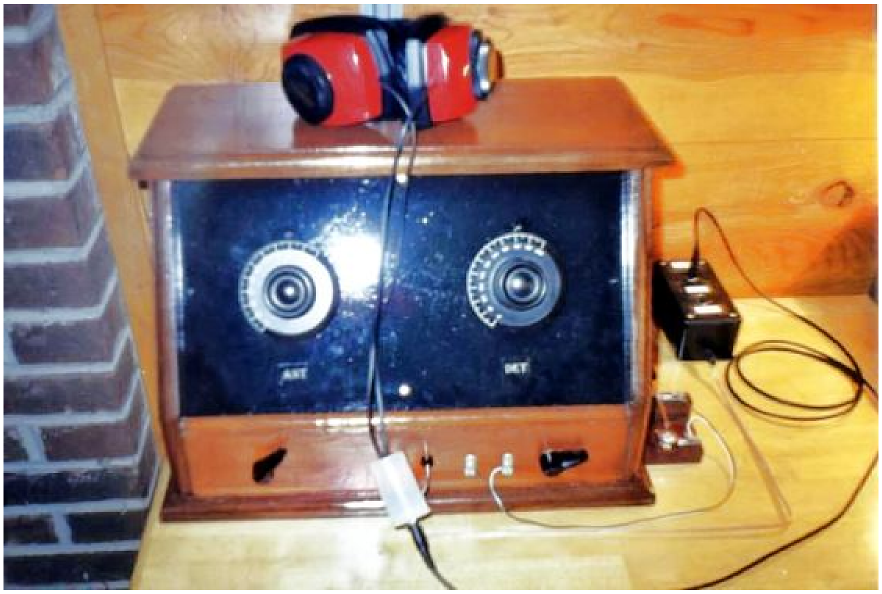

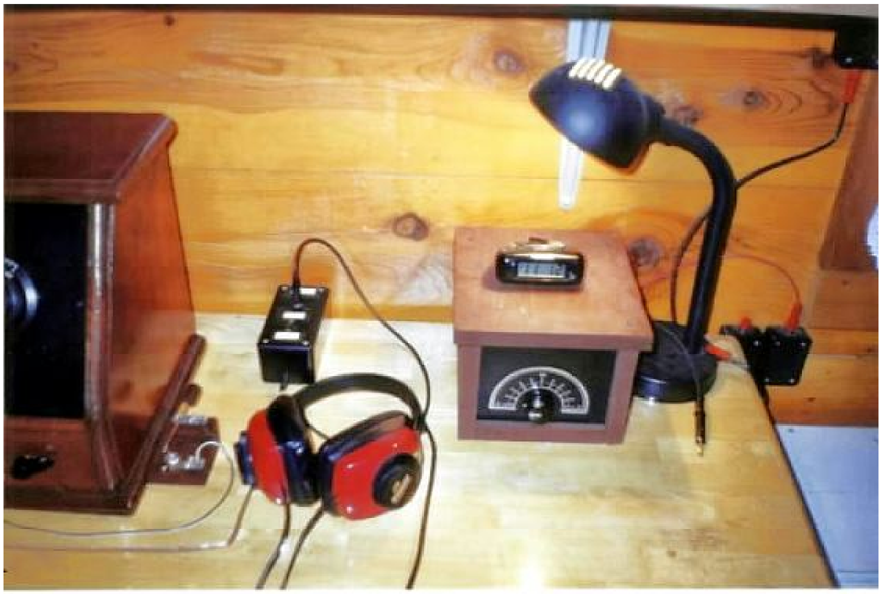

The matching unit.

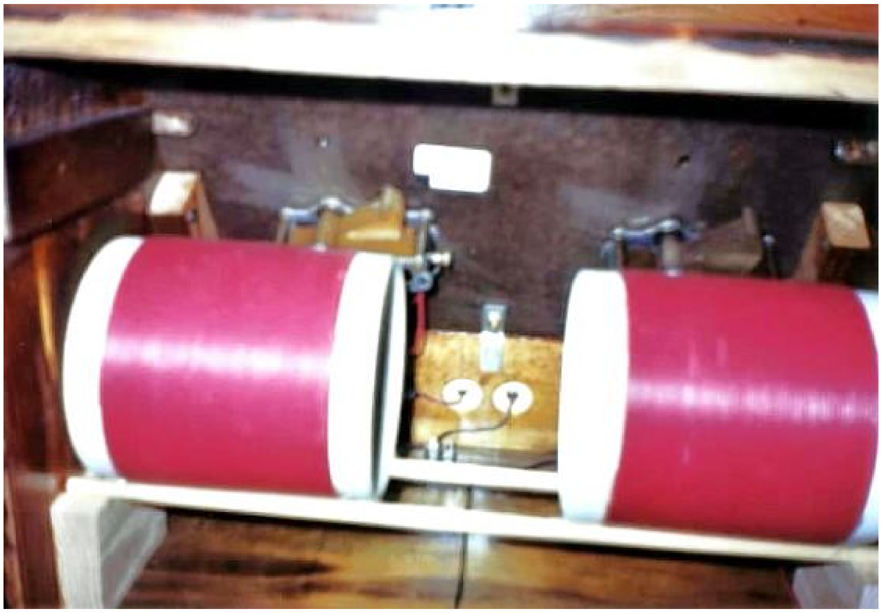

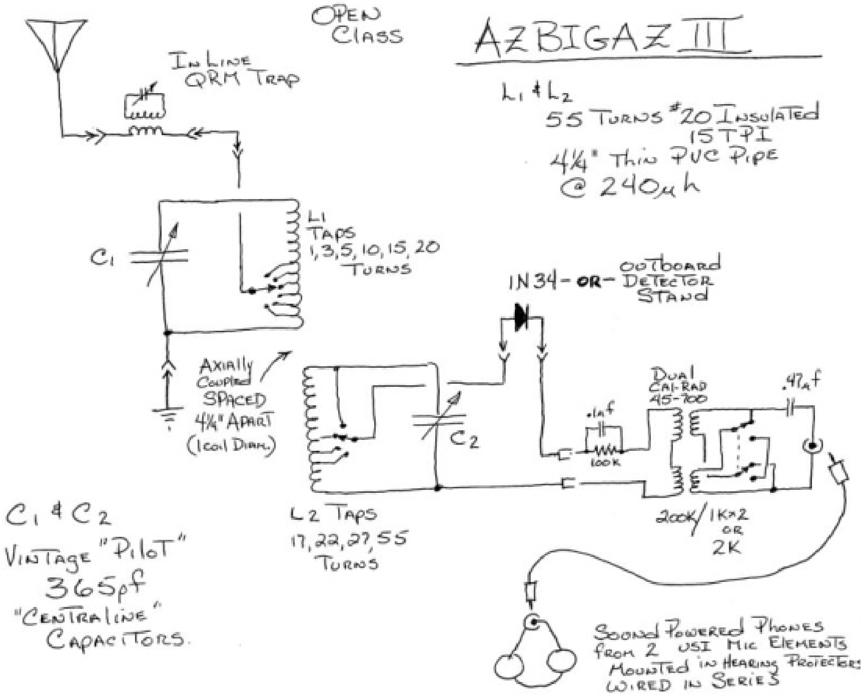
OPEN CLASS
AZBIGAZ III
IN LINE QRM TRAP
L1 & L2
55 TURNS #20 INSULATED 15 TPI
4¼" THIN PVC PIPE
@ 240μh
L1 TAPS 1,3,5,10,15,20 TURNS
C1
IN34 -OR- OUTBOARD DETECTOR STAND
AXIALLY COUPLED SPACED 4¼" APART (1 COIL DIAM.)
C2
.1μf
100K
DUAL CAL-RAD 45-700
.47μf
200K/1K×2 OR 2K
L2 TAPS 17,22,27,55 TURNS
C1 & C2 VINTAGE "PILOT" 365pf "CENTRALINE" CAPACITORS.
SOUND POWERED PHONES FROM 2 USI MIC ELEMENTS MOUNTED IN HEARING PROTECTERS WIRED IN SERIES

Jack Bryant

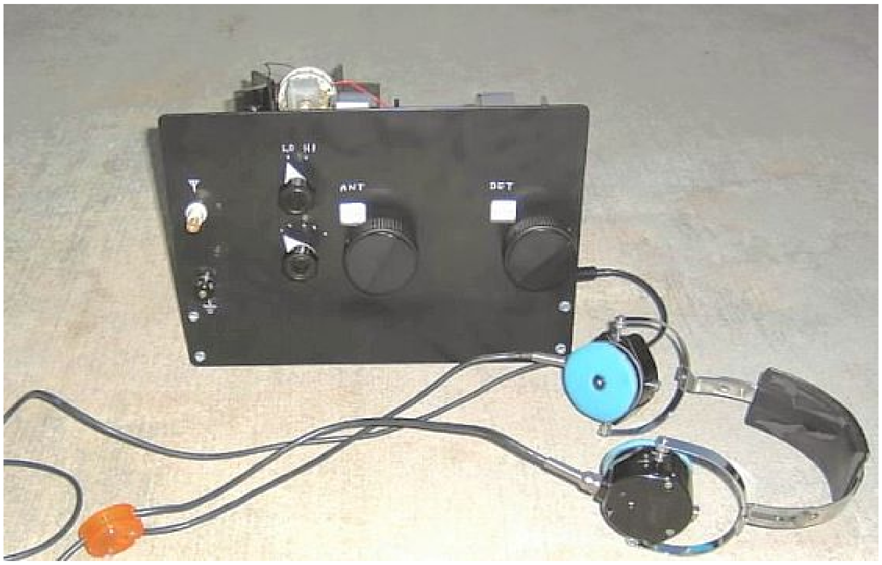
LO HI
ANT
DET

I started putting the set together on January 5, 2004...way late! I used a black plastic clipboard for the front and a clear orange plastic one for the chassis plate. 1"x3" small boards are used to hold up the chassis plate. I installed the coil supports during the contest and the phone jack afterward.

This set is a modified bandpass unit, but uses two separate coil assemblies. Each assembly consists of toroids wound on FT114-61A cores for the antenna and detector coils. Each assembly has it own coupling coil, connecting the antenna and detector circuits.

Here are the specs:

Lo Band
Antenna: FT114A-61, 33T #18

Couping Coil:FT-82-61, 10T ant side, 5T detector side, tapped every turn (det side only) Detector: FT114A-61, 40T #18

High Band

Antenna: FT114A-61, 22T #14 Couping Coil:FT-82-61, 10T ant side, 5T detector side, tapped every turn (det side only) Detector: FT114A-61, 20T #14

The audio transformer is one of the Fair Radio 100k:100 ohm units. The headphones are Dynalec mic elements. I started with one of the ITT diodes, then settled on one with a blue stripe (characteristics unknown).

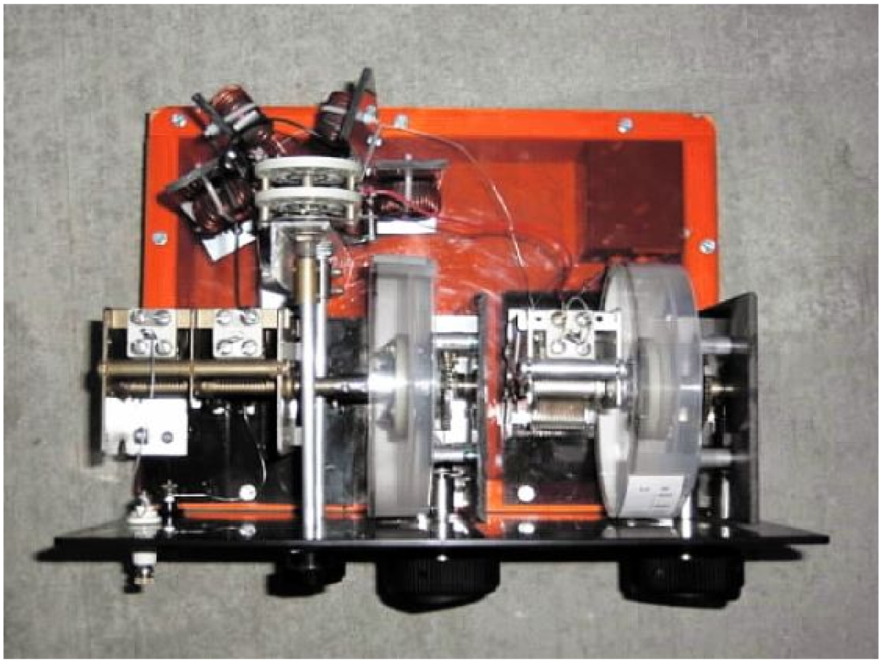

Continued on Page 3.

Ad
INNERACTIVE
SAVE $500 WITH CODE RET
IES AURACLOUD®
"NOW IN 3D"
CLICK HERE

Select Menu

- Home
- About Us
- Contest Results
- 1AD Contests
- 1AD Entries
- Antenna Matching
- Antennas
- BHAM Radios
- Crystal Contests
- Crystal Radio Entries
- Sprint Contest
- Headphones
- Projects, Mods
- Favorite Links
- Contact Jack

Sub Pages

- Main Page
- 2004 Page 1
- 2004 Page 2
- 2004 Page 3
- 2005 Page 1
- 2005 Page 2
- 2006 Page 1
- 2006 Page 2
- 2006 Page 3
- 2007 Page 1
- 2007 Page 2

- Contest Results
    - 2004
    - 2005
    - 2006
    - 2007
    - 2008
    - 2009
    - 2010
    - 2006 Sprint
    - 2008 Sprint
    - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2009 Contest Entrants

Page 2

Ron Everingham, Open Class Entry

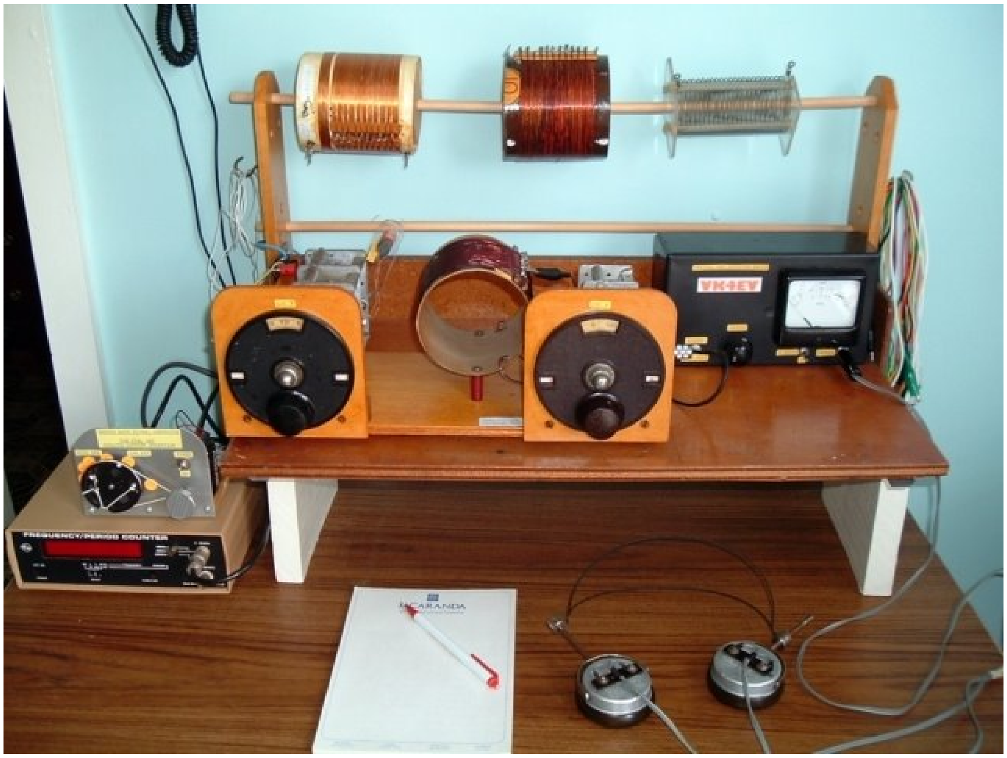
VK4EV
FREQUENCY/PERIOD COUNTER
JACARANDA

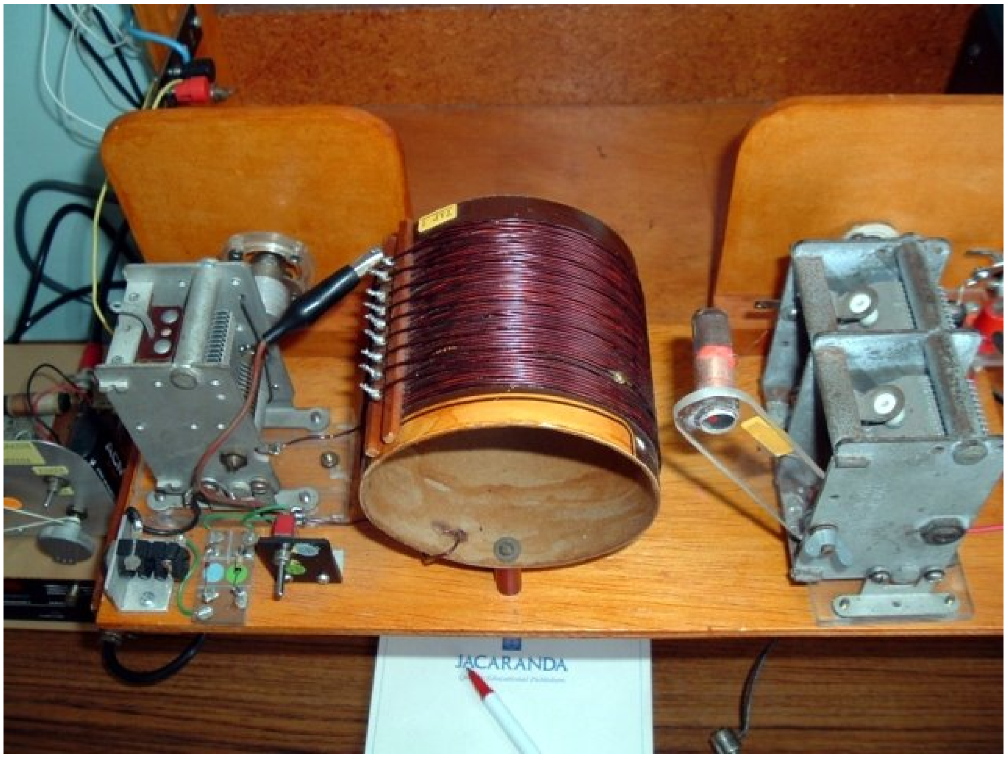

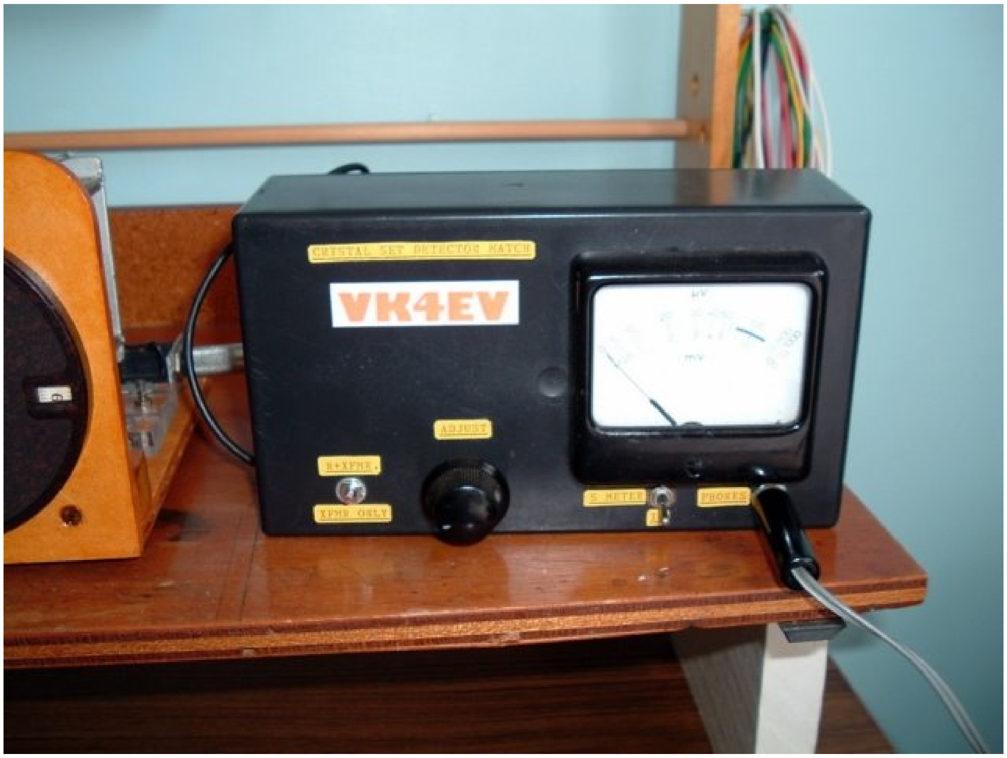

This double tuned radio is really my first attempt for DX listening by virtue of using two vernier dials, variable coupling and audio matching.

L1 is wound on a piece of ferrite rod connected to a swinging arm to obtain variable coupling for L2. The S meter is a surplus 30uA movement which gives a good indication for setting the tuning capacitors to a particular frequency when a signal generator is loosely coupled to the aerial.

The aerial used is about 20 feet high and 80 feet long bent to fit into my small lot.

Dave Schmarder, Loop Class Entry

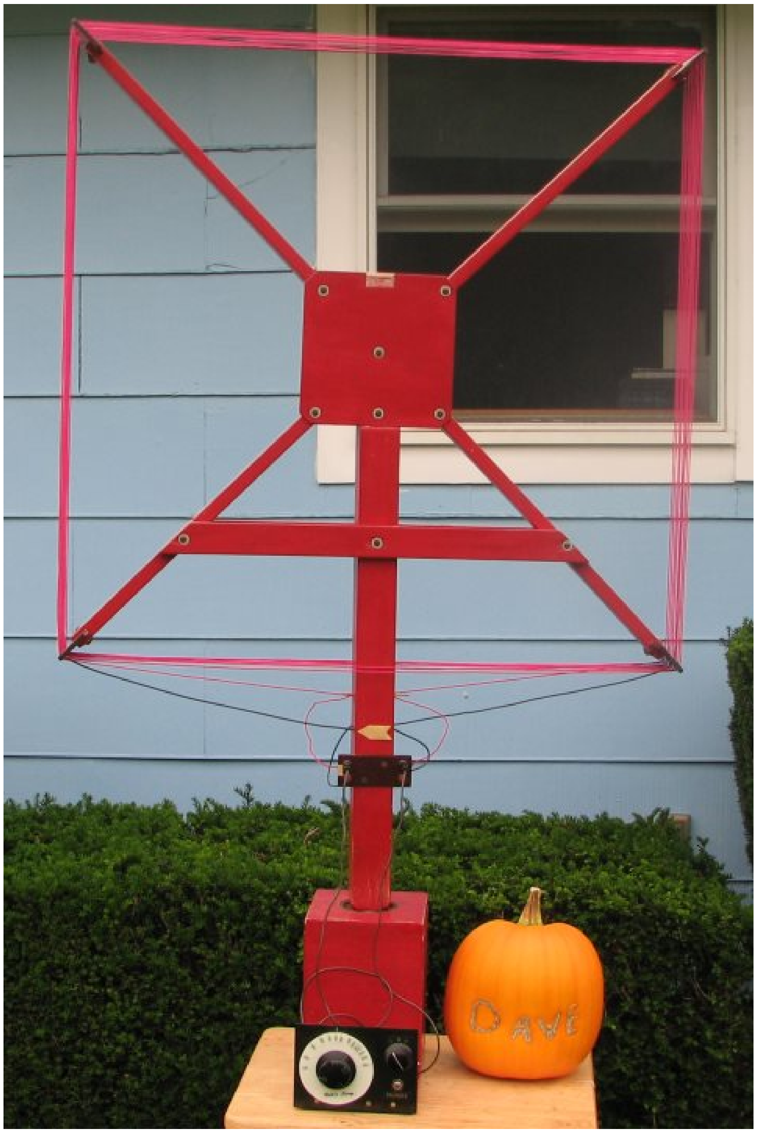

To see Dave's log and contest set description click here. This will take you to his fine website. Here is his loop antenna description.

Jack Hennon, Open Class Entry

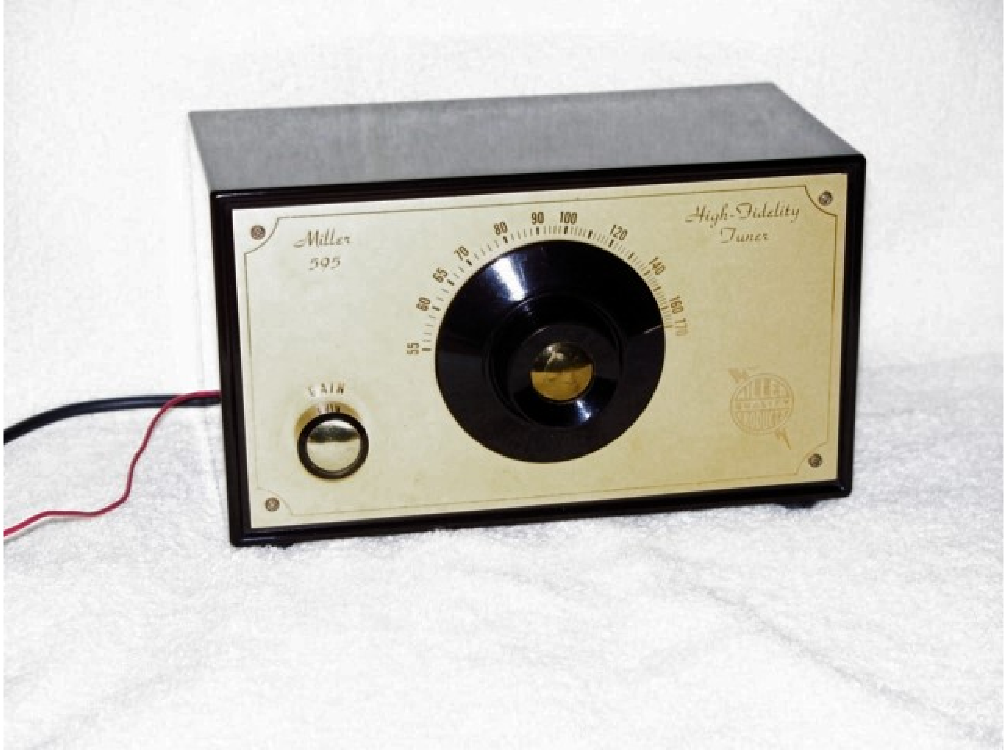
Miller
595
High-Fidelity
Tuner
55
60
65
70
80
90
100
120
140
160
170
GAIN

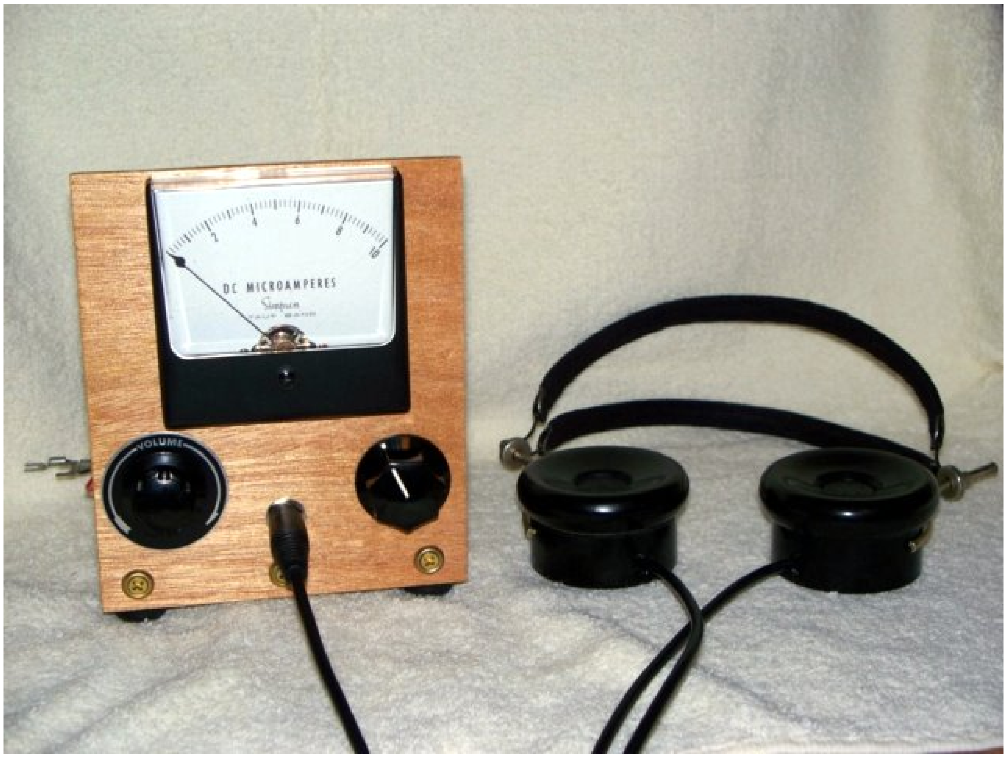
DC MICROAMPERES
Simpson
2
4
6
8
10
VOLUME

Jack used the venerable Miller 565 with an external audio matching unit. Notice the very rare 10 ua meter in his audio unit!

Evan Haydon, Open Class Entry

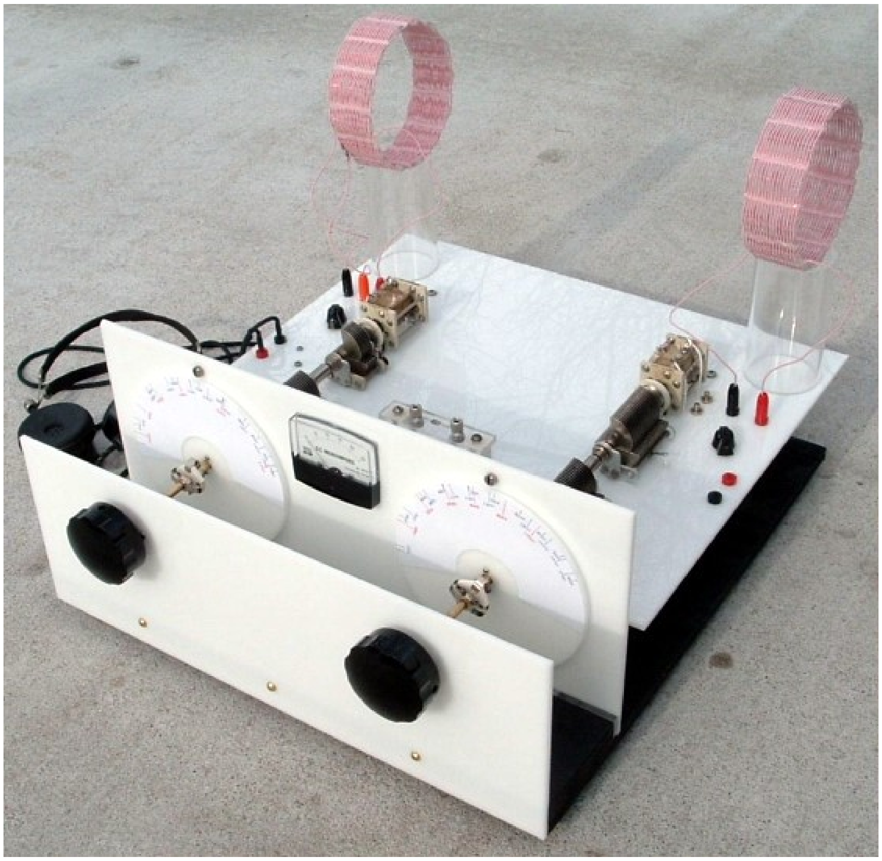

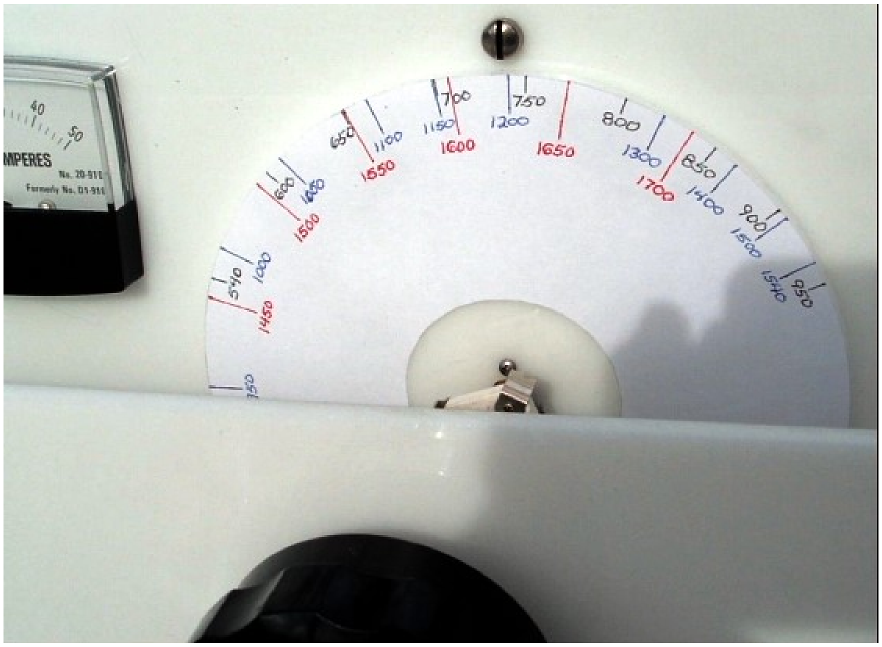

I spent a lot of time preparing for the 2009 contest. I spent many hours most evenings for the last three months listening to the broadcast band. From both notes made then and from memory, I am very familiar with what stations appear at what time of the day at this location. I watch band conditions carefully and fairly accurately know which stations it is even possible to hear at any time on any frequency.

My crystal radio will at night hear any signal that my spotter radio will hear except for weak signals 10kc either side of my 3 local stations. My crystal radio will hear 95% of the daytime signals that my spotter radio will hear. On a good cold winter day, I have received 57 stations. My radio, traps, and antenna system are exactly the same as used in the 2008 contest. I could not find any way to improve the system.

This year's contest had unusual conditions here. The first day had very good receiving conditions. The second day had good conditions. The band then stagnated. It did not change for the rest if the contest. The stagnant condition was not what I would call a normal winter condition either. There was no flux or change (due to weather conditions) to cause stations to appear and disappear. For day after day it was the same old stations all up and down the band.

The contest was fun to work as always. My list of all time stations identified on the broadcast band now numbers 505 stations. I identified 66 stations this year that I did not hear in the 2008 contest. That means that I identified 36 stations in the 2008 contest that I did not hear this year. In the 2009 contest, I logged 4 stations from New York state, 1 from Boston, 3 from California, 2 from Philadelphia, and 25 stations from Canada. I identified stations from 31 different states, Canada, Mexico, and Cuba. IBOC probably cost me 3 or 4 stations to not be logged this year

Below is a summary of my activities during the 2009 crystal radio contest.

```
        Stations
Day Logged  Time spent
1-16   197   15 Hr  35 Min
1-17   32    12 Hr  20 Min
1-18   27    11 Hr  20 Min
1-19   21    11 Hr  10 Min
1-20   11     8 Hr   0 Min
1-21    9     7 Hr  20 Min
1-22    8     6 Hr  50 Min
1-23    8     8 Hr  15 Min
1-24    7     7 Hr  35 Min
1-25    8    10 Hr  20 Min

Total time spent - 98 hours 45 minutes
Total stations heard - 328
Total stations identified - 312
Total points - 583,312
```

Michael Bartolone, Open Class Entry

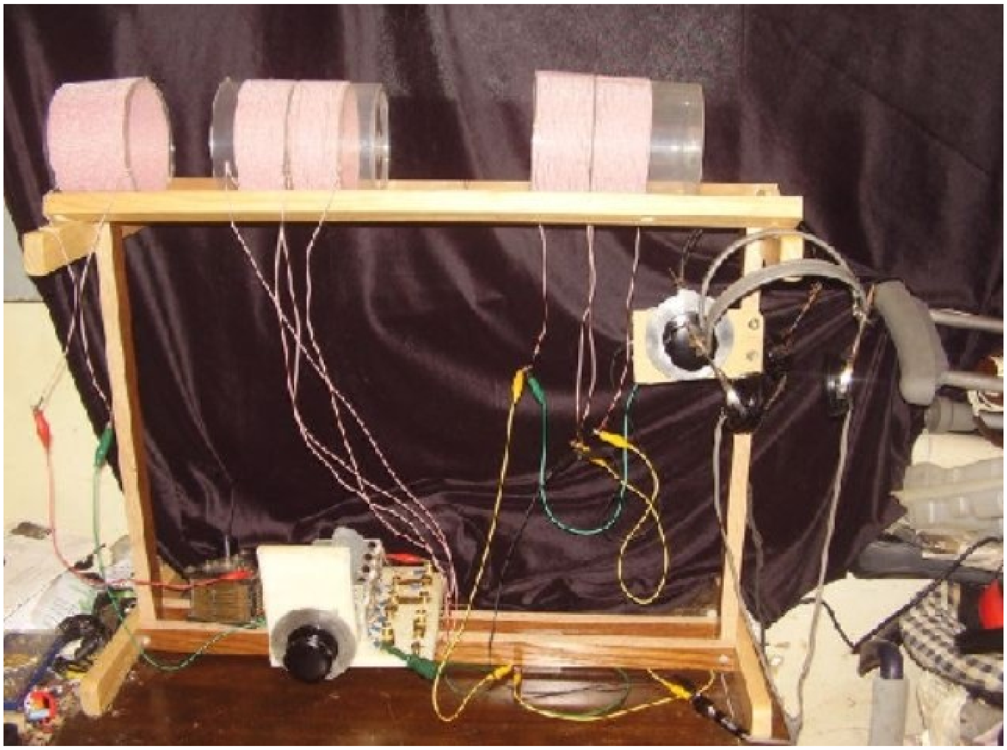

My current contest rig has contra wound solenoid coil 5.25 in diameter, each section 29 turns of "indoor antenna wire", 2.15in length per section, single gang 400pf air variable cap, unknown diode scavenged from a clock radio picked up at Goodwill (the diode has two red bands on one end and no other markings), Clevite brush phones.

Coils suspended in a slide cradle 20 inches above the desk. The tank is loose coupled to an ATU with contra wound solenoid coil 5.25in diameter, each section 26 turns switchable so I can use either single section, or both in series, or in parallel, with a 5 gang 380pf per gang cap switchable to put various gangs in series or parallel with the various coil configs.

One tunable trap with a single 5.25 in diameter coil, 43 turns, 3 in long with 440pf cap for tuning.

Note the 'indoor antenna wire" is like very cheap litz…35 strands individually insulated but only twisted, not braided like true litz. However, the price was too good to pass up ($0.02 per foot when I bought it).

My antenna is a 75+ foot long wire (16ga speaker wire) average about 12 feet above ground, run from the eaves at one corner of the house to a tower about 35 feet from the house and then back to the other corner. The ground is a single copper clad rod about 3 feet long about 2 feet into

- Contest Results
    - 2004
    - 2005
    - 2006
    - 2007
    - 2008
    - 2009
    - 2010
    - 2006 Sprint
    - 2008 Sprint
    - 2009 Sprint
- Previous
- Next

Radios Of The 2010 First Final Farewell DX Contest Contest Entrants

Page 1

Mike Tuggle, Lyonodyne-17, Open Class Entry

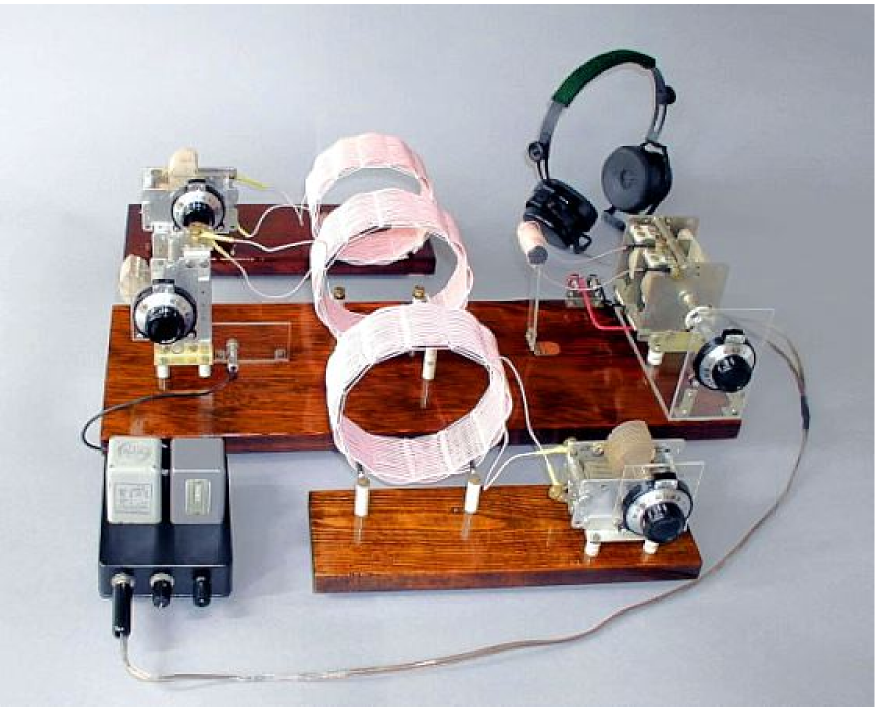

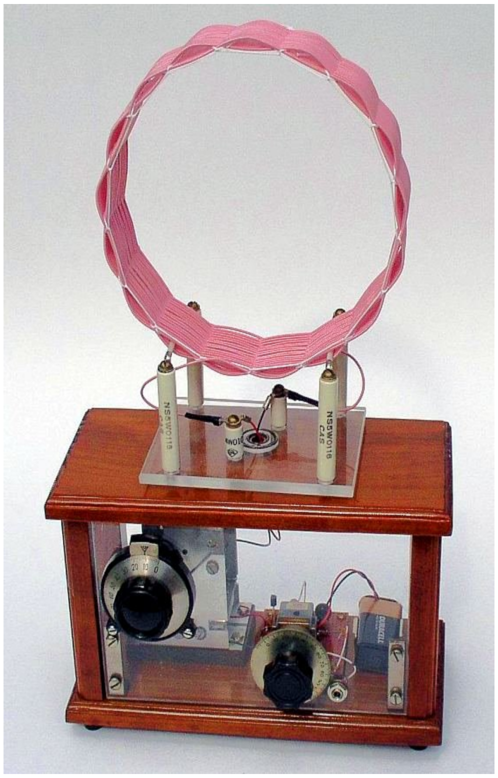

The Lyonodyne 17 was used in the crystal set category, the same set used last year. The Active Device set used in the contest is the Homodyne. Click here for the info on the radio, as used in the 2007 Active Device Contest..

Evan Haydon, Open Class Entry

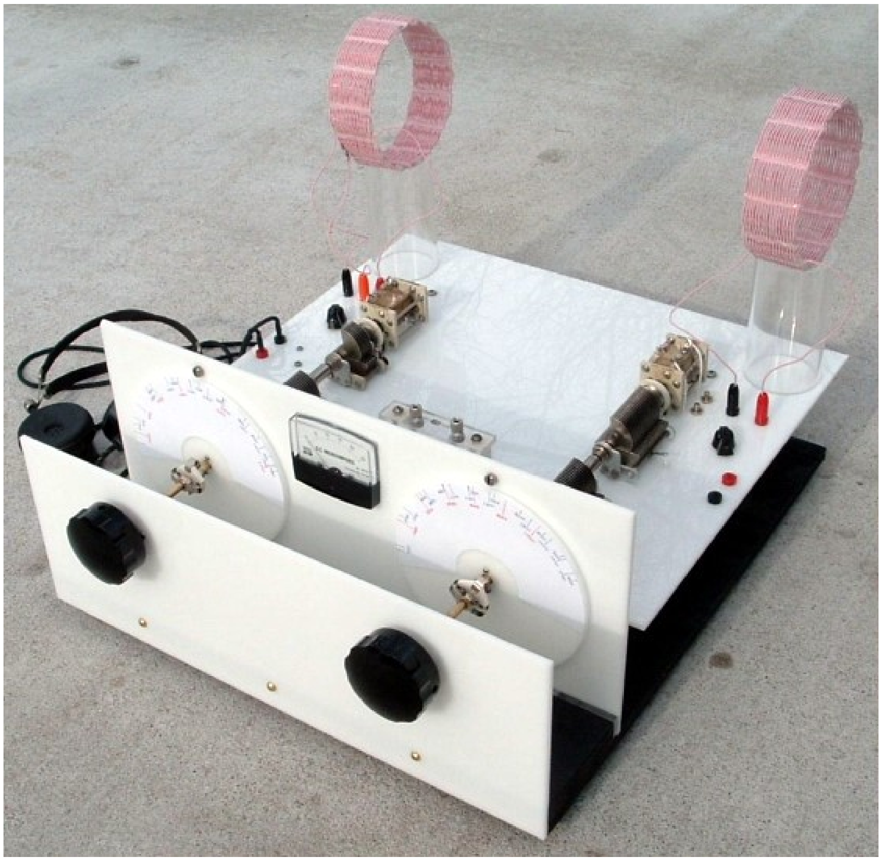

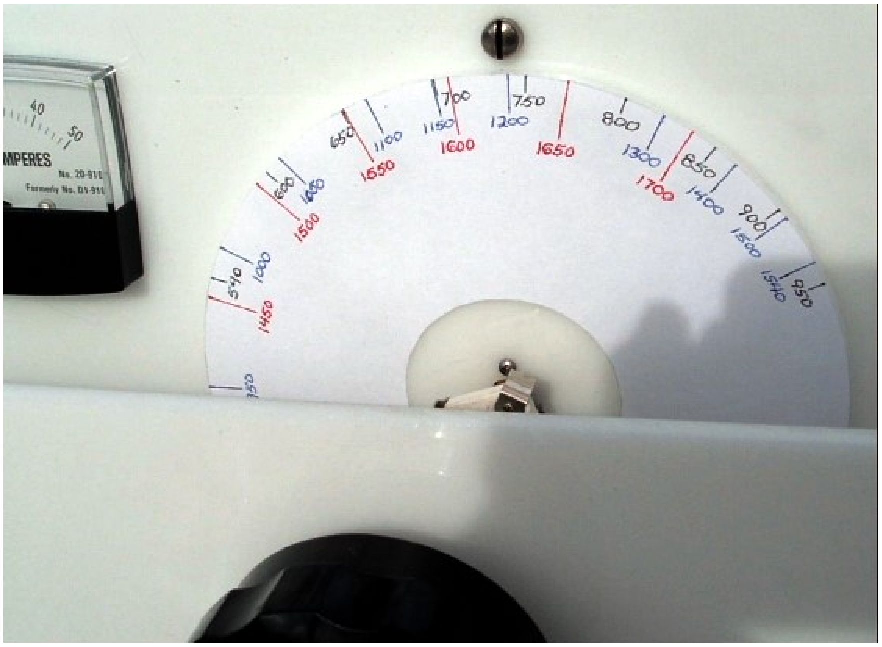
40
50
AMPERES
No. 20-911
Formerly No. D1-911
350
540
1000
1450
600
1500
650
1100
1550
700
1150
1600
750
1200
1650
800
1300
1700
850
1400
900
1500
1540
950

This year’s crystal radio contest was fun, even though it was long. The broadcast band conditions in Lincoln were what I would call normal or generally good. I used the same crystal radio and antenna setup that I used in the 2009 contest.

23 radio stations that I identified in the 2009 contest were not heard in the 2010 contest. I identified 24 new stations in the 2010 contest to add to my all time heard list. That list now stands at 536 stations identified. For that number, I only count US and Canadian stations.

In the 2010 contest I heard stations in 33 states from California to New York. I identified 30 Canadian stations, 7 Cuban stations, and 2 Mexican stations. The most stations identified on one frequency was 9 stations on 1190 kHz.

```
Number of stations per week: (Jan 15 to Mar 15, 2010)
Week 1 - 244
Week 2 - 42
Week 3 - 33
Week 4 - 21
Week 5 - 16
Week 6 - 17
Week 7 - 15
Week 8 - 6

Total of 394 broadcast band stations identified

Total points: 793,421
```

Ralf Siemieniec, Open Class and Below BCB Class Entries

Receiver Used For Long Wave Reception

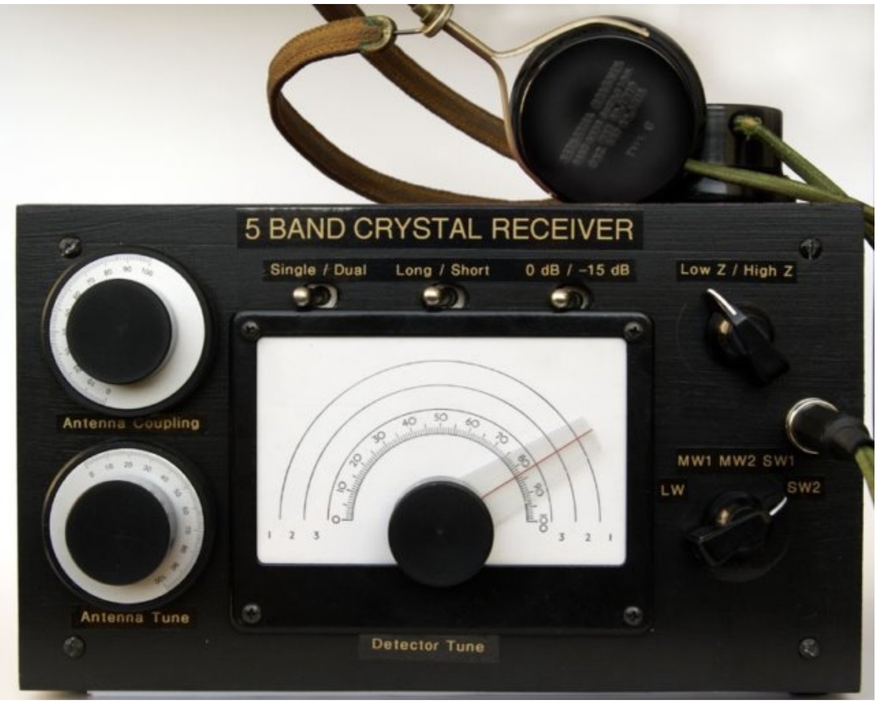
5 BAND CRYSTAL RECEIVER
Single / Dual
Long / Short
0 dB / –15 dB
Low Z / High Z
Antenna Coupling
MW1 MW2 SW1
LW
SW2
Antenna Tune
Detector Tune

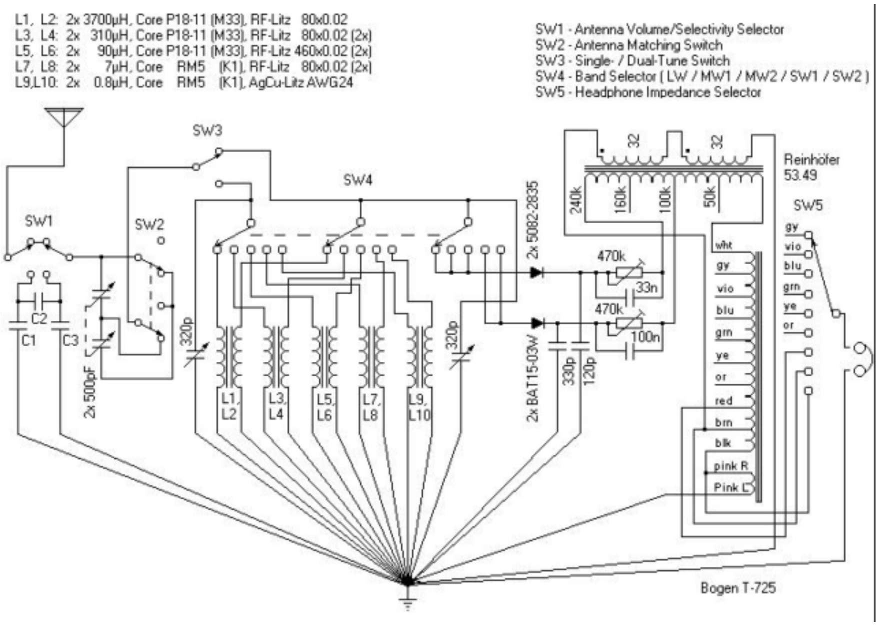
L1, L2: 2x 3700µH, Core P18-11 (M33), RF-Litz 80x0.02
L3, L4: 2x 310µH, Core P18-11 (M33), RF-Litz 80x0.02 (2x)
L5, L6: 2x 90µH, Core P18-11 (M33), RF-Litz 460x0.02 (2x)
L7, L8: 2x 7µH, Core RM5 (K1), RF-Litz 80x0.02 (2x)
L9,L10: 2x 0.8µH, Core RM5 (K1), AgCu-Litz AWG24
SW1 - Antenna Volume/Selectivity Selector
SW2 - Antenna Matching Switch
SW3 - Single- / Dual-Tune Switch
SW4 - Band Selector (LW / MW1 / MW2 / SW1 / SW2)
SW5 - Headphone Impedance Selector
SW3
SW4
SW1
SW2
C2
C1
C3
2x 500pF
320p
L1, L2
L3, L4
L5, L6
L7, L8
L9, L10
320p
2x 5082-2835
2x BAT15-03W
330p
120p
470k
33n
470k
100n
32
32
240k
160k
100k
50k
Reinhöfer 53.49
SW5
wht
gy
vio
blu
grn
ye
or
red
brn
blk
pink R
Pink L
gy
vio
blu
grn
ye
or
Bogen T-725

Long Wave Receiver Schematic

Open Class Crystal Set Entry

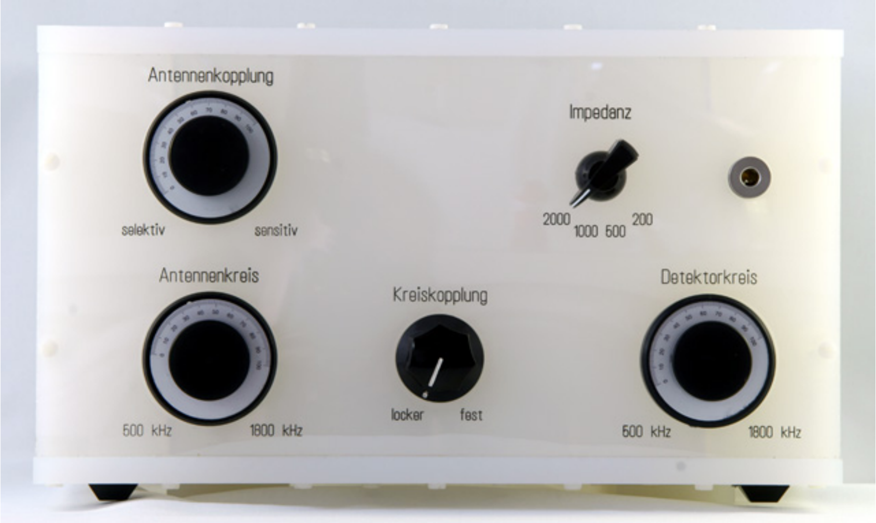
Antennenkopplung
selektiv
sensitiv
Impedanz
2000
1000
500
200
Antennenkreis
500 kHz
1800 kHz
Kreiskopplung
locker
fest
Detektorkreis
600 kHz
1800 kHz

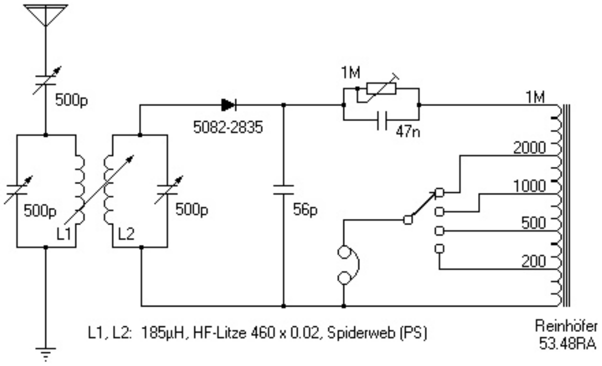
500p
1M
5082-2835
47n
1M
2000
1000
500
200
500p
L1
L2
500p
56p
L1, L2: 185µH, HF-Litze 460 x 0.02, Spiderweb (PS)
Reinhöfer
53.48RA

Crystal Set Schematic

Spiderweb Coil Q vs. Frequency

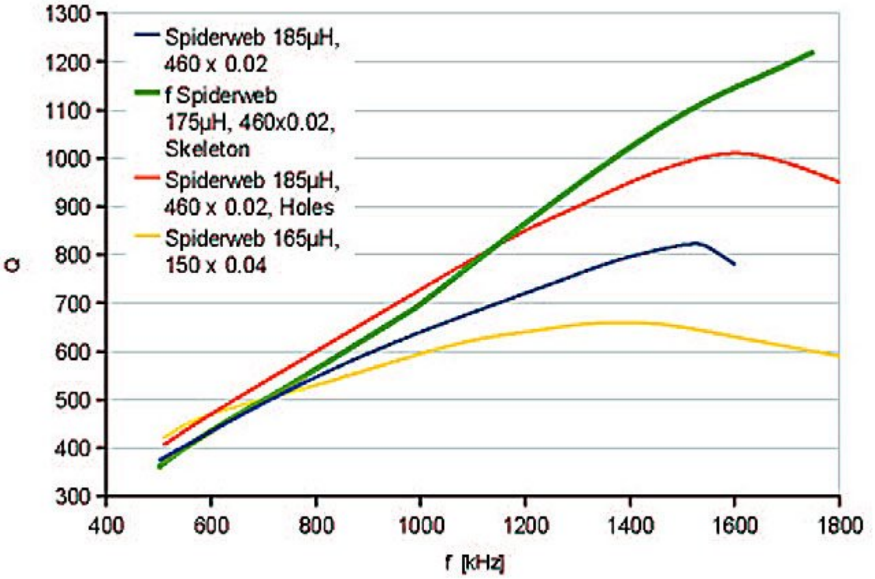

Here is a map showing the received stations.

- Contest Results
    - 2004
    - 2005
    - 2006
    - 2007
    - 2008
    - 2009
    - 2010
    - 2006 Sprint
    - 2008 Sprint
    - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2005 Contest Entrants

Page 2

Lem Morrison

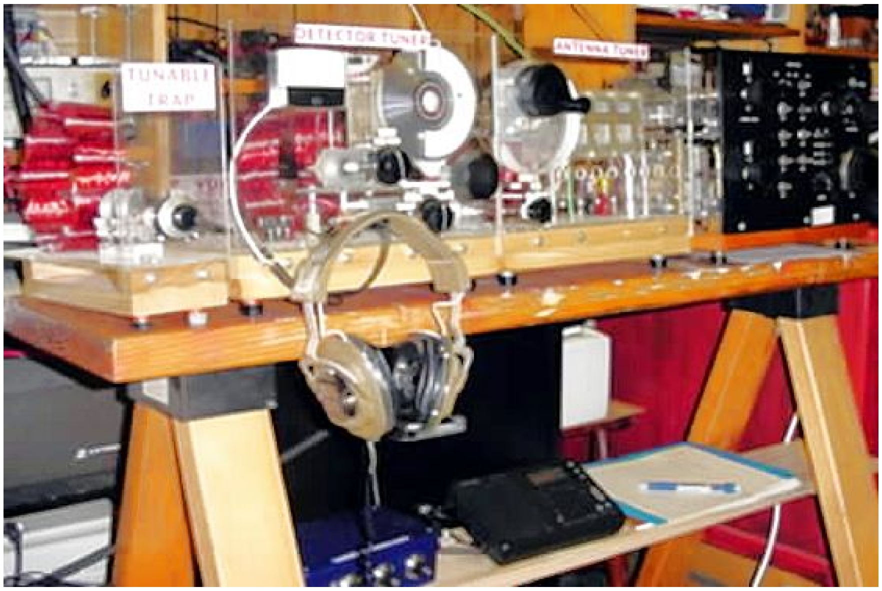

N4AHJ 2005 CONTEST CRYSTAL RADIO and OTHER INFO

The picture shown is the set in the 2004 Contest configuration.

TRAPS:
(1) Tunable (1) - 2-section variable, sections in parallel, 240pF total; Rook coil with 18 gauge wire, 5.25 inches effective diameter, 47 turns; loose-coupled to the Detector coil.

(2) Tunable (2) (located in AMU) - both use 14 gauge enamel wire on FT-114A-61 cores and dual-section variables with both sections in parallel.

(3) Fixed (9) (located in AMU) - 14 gauge to 22 gauge enamel wire on FT-114A-61 cores, silver mica fixed capacitors, and plated ceramic trimmers (one trap for each of the local “blow torches”). Each link-coupled to the antenna feed to the set.

ANTENNA MATCHING UNIT (AMU):
One 4-section variable capacitor, four fixed capacitors (500pF, 1000pF, 2200pF, and 5000pF), and two coils (42 turns 14 gauge enamel on a FT-240-61 and 26 turns 16 gauge enamel on a FT-114A-61). All components switched to provide four matching arrangements: series LC, parallel LC, inductance only, and capacitance only. Both inductors are tapped to provide 23 L settings. An antenna switch is provided for switching between up to five antennas.

BANDPASS TUNER:
Capacitor - 520pF variable, 2 sections used in Tuggle arrangement; vernier is a surplus right-angle unit with a 25:1 ratio.

Coil - Two Rook coils with 14 gauge enamel wire, 6.25 inches effective diameter, 15 turns each, 30 turns total; bandswitch provides either series or parallel coupling for Low band and High band tuning based on Ben Tongue’s approach; loose-coupled to the Detector coil.

DETECTOR TUNER:
Capacitor - 420pF variable with home-made vernier with a 32:1 ratio (vernier constructed from gears salvaged from a VCR and a TV tuner).

Coil - Two Rook coils with 14 gauge enamel wire, 6.25 inches effective diameter, 18 turns each, 36 turns total; bandswitch provides either series or parallel coupling to provide Low band and High band tuning based on Tongue's approach; impedance taps (High Band: 100%, 75%, 64%, 50%, and 28%; Low Band: 100%, 72%, 50%, 28%) on coils to provide matching to diodes.

Diodes - Two (one marked FO215 and one marked 3502) in parallel.

AUDIO:
Matching - Reproduction of Steve Bringhurst's Ulti-Match circuit.

Headphones - Two Automatic Electric GH-66919 elements (manufactured 1942) out of a military handset and a set of Steve Bringhurst's enhanced SP cans.

NOTES:
The basic set is similar to my 2004 contest set. The primary > differences: larger Rook coils with larger gauge wire, hotter diodes (thanks to Steve Bringhurst), 20 uA meter, and new front panels.

I determined optimum AMU settings across the AM BCB for each antenna prior to the contest. These settings were reasonably reproducible during the contest, even when it rained!

The capability to switch among three antennas was great, especially between vertical and horizontal polarization. A station that was uncopiable on one antenna usually would be loader on one of the others. Generally, I used the vertical the most, but I noticed on several occasions that I could hear a station on one of the slopers when I could not on the Tee.

By the way, I live on a city lot that is 85' wide and 165' deep. However, I am blessed with having very good relationships with my neighbors on both sides and with a wooded area directly behind my lot. After Hurricane Ivan destroyed all my antennas (three oak trees ranging from 50' to 80' tall fell!), both neighbors allowed me to re-erect the Slopers across their backyards. Also, the 90' leg of the Tee ran well into the woods behind my lot.

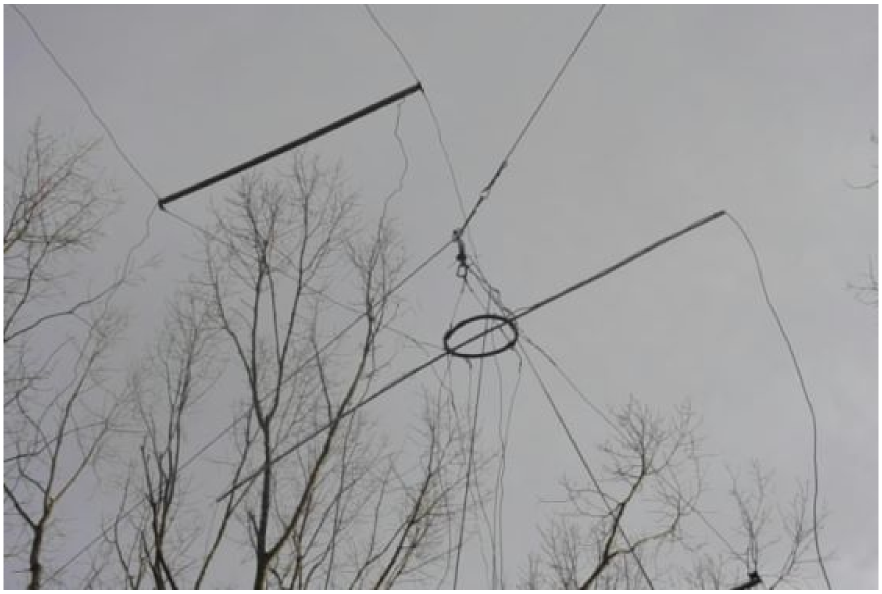

ANTENNAS:
(1) Tee, 46’ tall; top-loading legs consist of two parallel wires spaced 38” and run 90’ to Southwest and 40’ to Northeast. The vertical core consists of a cage of four parallel wires spaced 12” apart. To provide broader bandwidth I added two vertical wires from each of the top-loading legs (spaced 4’ from core center), and one additional vertical wire on Southeast and Northwest sides (spaced 4’ from core center). All of the vertical wires are connected at the base and series fed with a single wire.

(2) Sloper, 118’ with far end at 50’ and station end at 8’, oriented toward East-Northeast

(3) Sloper, 130’ with far end at 50’ and station end at 8’, oriented toward West-Northwest

Ground system: Ham station ground - six ground rods interconnected and connected to shack with two quarter-inch copper straps; supplemented by one 8’ ground rod directly under the Tee and 4-4’ ground rods under the wires dropped from the top-loading legs, all inter-connected with copper wire; both ground systems connected with a run of copper braid (salvaged from old RG-8) over a #10 copper wire. Radials are 17-gauge aluminum electric fence wire; used 12 radials with lengths from 30’ to 135’ (one radial attached to the city water system!).

Mike Tuggle

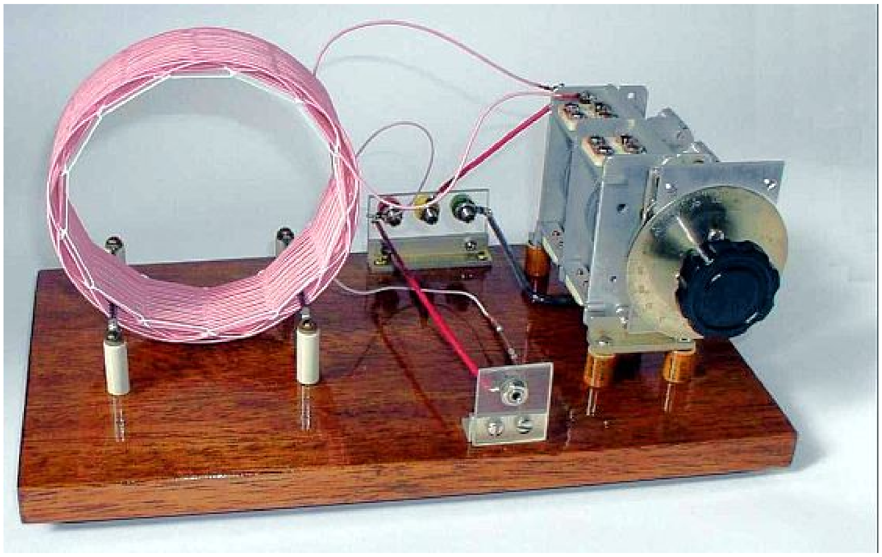

As it turns out, all loggings were made in the 'BROAD' configuration.

The DX gods never really smiled this time, maybe an occasional smirk. Still, I guess I can't complain.

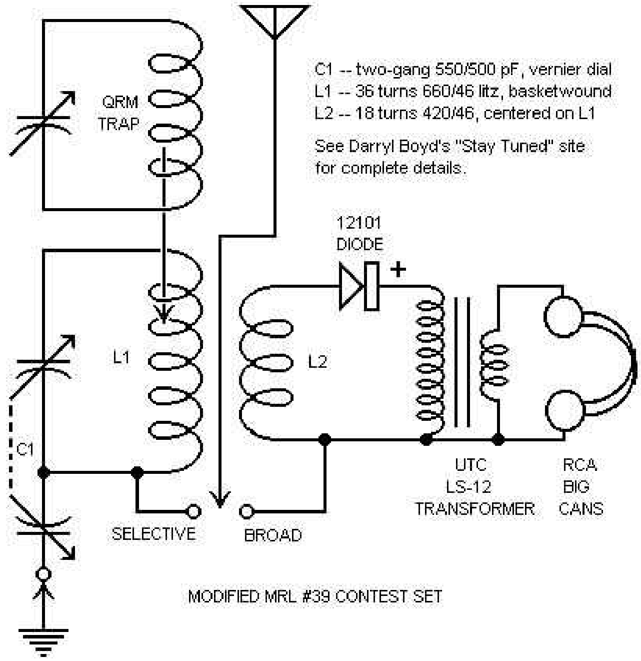
C1 -- two-gang 550/500 pF, vernier dial
L1 -- 36 turns 660/46 litz, basketwound
L2 -- 18 turns 420/46, centered on L1
See Darryl Boyd's "Stay Tuned" site
for complete details.
QRM
TRAP
12101
DIODE
+
L1
L2
C1
UTC
LS-12
TRANSFORMER
RCA
BIG
CANS
SELECTIVE
BROAD
MODIFIED MRL #39 CONTEST SET

Philip Miller Tate (Ge_Whiz)

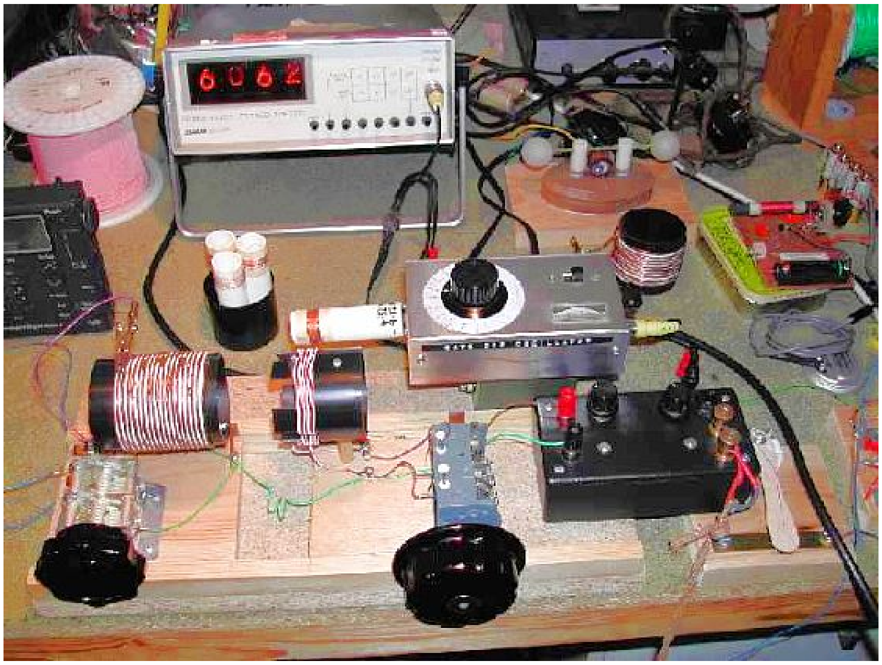

The radio is a loose-coupled set. The antenna coil is heavy solid copper wire, 16 turns on a two-and-a-half inch diameter ABS drain pipe, tapped every other turn. The aerial is coupled in via one stage of a dual 120 + 120pF variable capacitor; the other is in parallel with the coil. There is no earth connection – the set works better without it – but the earth level is hard-wired through to the detector coil. The aerial itself is a total of 120 feet of wire running from the end of the house, 50 feet to the shed roof and back at an angle of about 35 degrees to the opposite end of the house, average height 20 feet above ground. Note that there is no vertical section, as the radio is located on the top floor of the house.

The detector coil is eight turns of 660/46 Litz wire with an inductance of 5uH, wound on a five-point rook former of the same ABS pipe, in parallel with a 350pF variable capacitor with three-turn gearing. The combination tuning system covers approximately 4-15MHz. The detector is an AAZ18 germanium gold-bonded diode. Output is to a Select-to-Match transformer system and thence to a pair of DLR-5 military balanced-armature headphones.

Alongside the detector coil, as shown above, is the plug-in coil of a homebrew gate-dip oscillator connected to a Racal Nixie-tube digital frequency meter. By carefully adjusting the tuning of the GDO, the dip of the detector coil resonance can be found to about +/-20kHz accuracy, and used to find the station on the spotter radio, a Sony IC100.

Rich Shivers

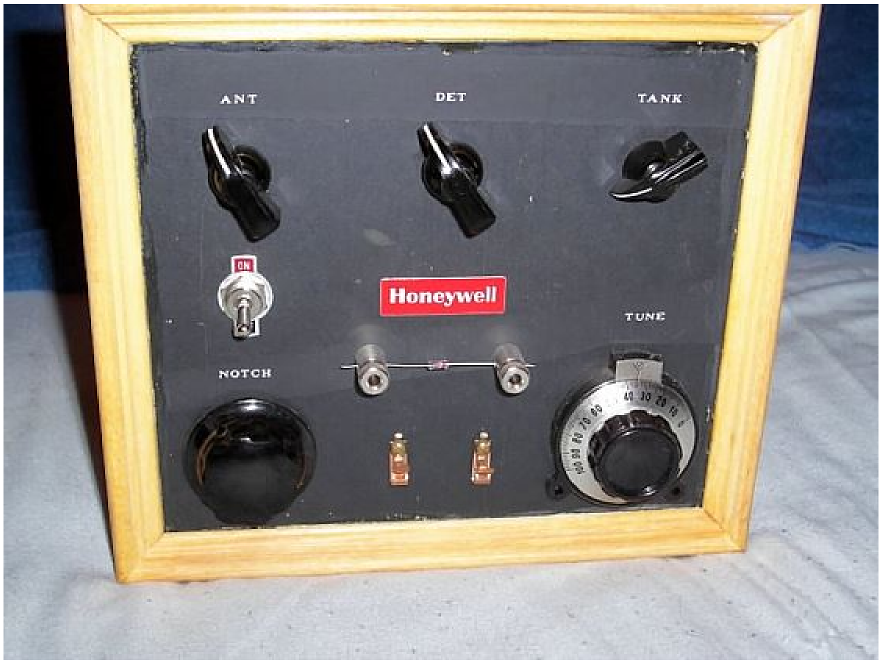

I have followed the comments on the Rap 'n Tap board, and agree that conditions were not very good this year. Compared with last year my total station count was down and DX stations were just not there. Disappointed, yes; did I have fun, yes.

I am using a large tapped single coil set. It's an air core basket weave coil 5" in diameter, about 8" long. 66 turns of 14awg THNN household wire. It is tapped every 11 turns with separate rotary switches for the antenna, detector, and capacitor. Normal arrangement has the antenna and detector tapped near the bottom of the coil. By tapping down the tank circuit I can band-spread the top of the band for improved selectivity. The tuning cap is a three section 400uh unit. One section across the coil, the other two section in parallel from the bottom of the coil to ground. A 8 to 1 vernier dial drives the tuning capacitor. Between the antenna and tank coil is a wave trap constructed of a tank circuit and a coupling coil. Both trap coils are wound on a single ferrite core. Headphone are WE 509's. Detector is a 1N34. Antenna is a T arrangement; 40' across the roof, with an 8' lead-in, the other leg drops straight down for 12'.

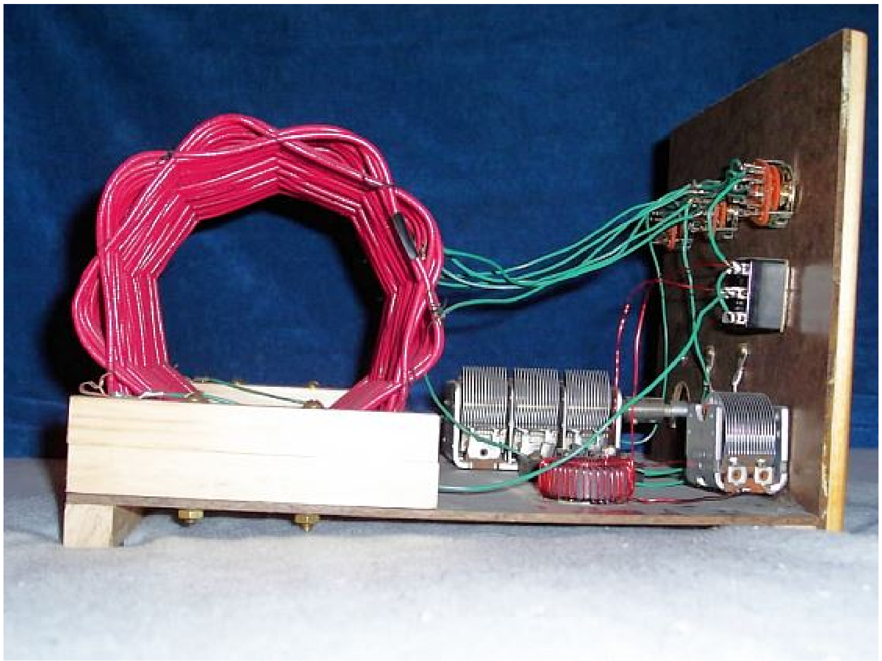

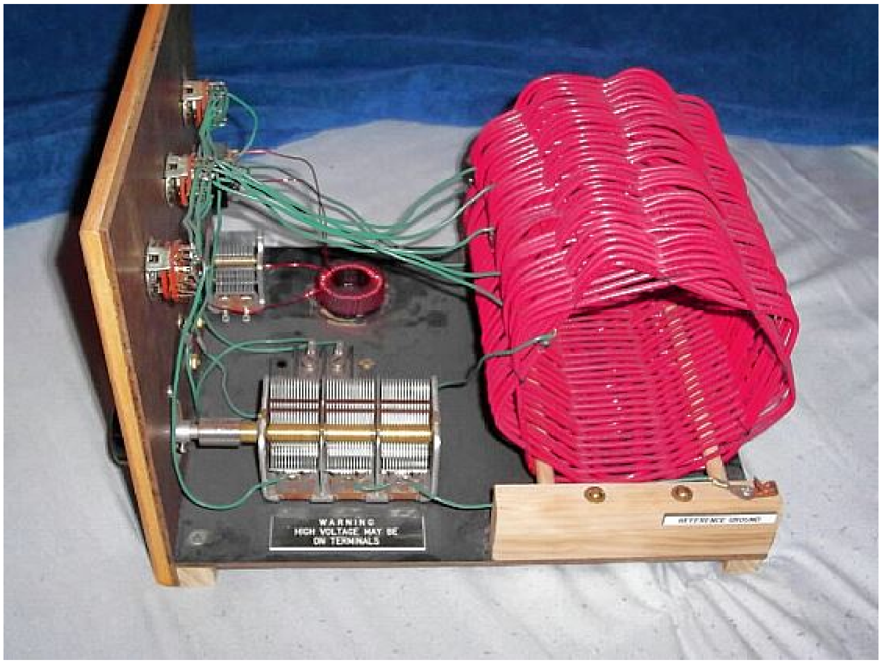

Lou Dayich

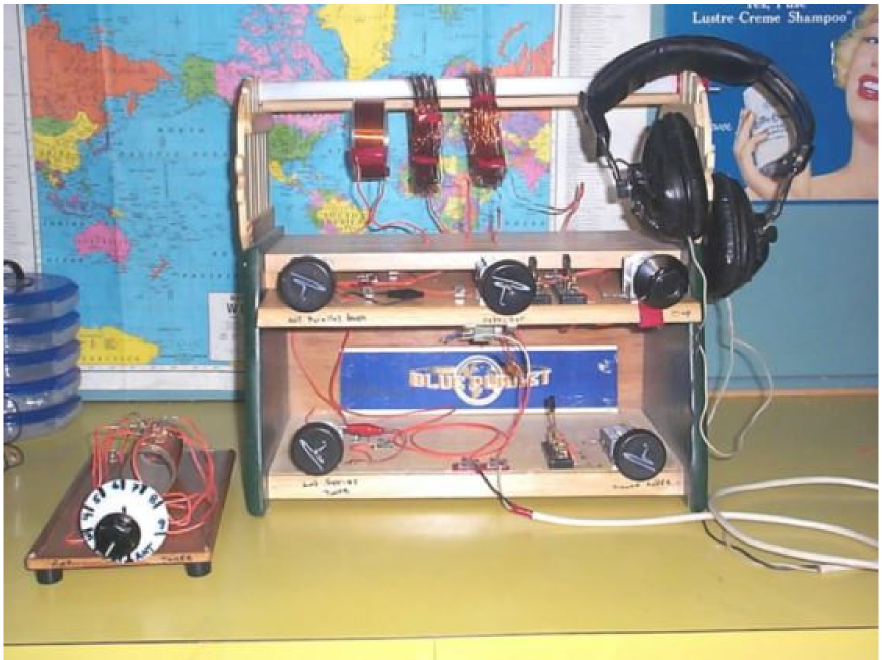

I built a new set for this year's contest. Some components were transplants from the 2004 contest. This set uses three coils and 5 variable capacitors, you will also notice three knife switches that were used in various parts of the receiver.

The antenna coil is aproximately 30 turns of close wound magnet wire, apx. 3 inch diameter. The caps (top and bottom) on the left side of the receiver provide parallel and series tuning to the antenna coil. The alligator clips on both caps allow the caps to be easily taken out of the circuit. The detector and trap coils are apx 4.5 inches, 9 point basket weave coils. During most of the contest, I actually set this up with the trap in the middle and the set became triple tuned.

The detector coil is a dual gang Russian variable cap (built in vernier) and the knife switch allowed me to use one or two gangs for detector tuning. The bottom right cap uses the same type of cap and knife switch arrangement for ground leg tuning. The top right cap was fitted with a 10 to 1 vernier reduction for wave trap tuning. The phones are sound powered cobbled from some USI elements and a bogen matching transformer can be seen mounted under the "shelf". I used two diodes at different times during the contest, an FO215 (solo and configured in a Hobbydyne set up) and a schotty diode.

As you can see the radio case "was" a wooden breadbox and the door came off making a nice shelf. On the left side you will see a stand alone inductive antenna tuner, this helped during the contest and is really great for casual listening; with the caps configured in a certain fashion, I can tune 5 or 6 of my regular stations just by clicking the rotary switch on the tuner. I had more points and station count with this year's receiver, but due to high winds just prior to the contest lost my best antenna. I ended up with a sloper apx 80 feet long and about 25 feet at its highest. Lots of fun! It is great to see all the radios that been built.

Dick Kleijer

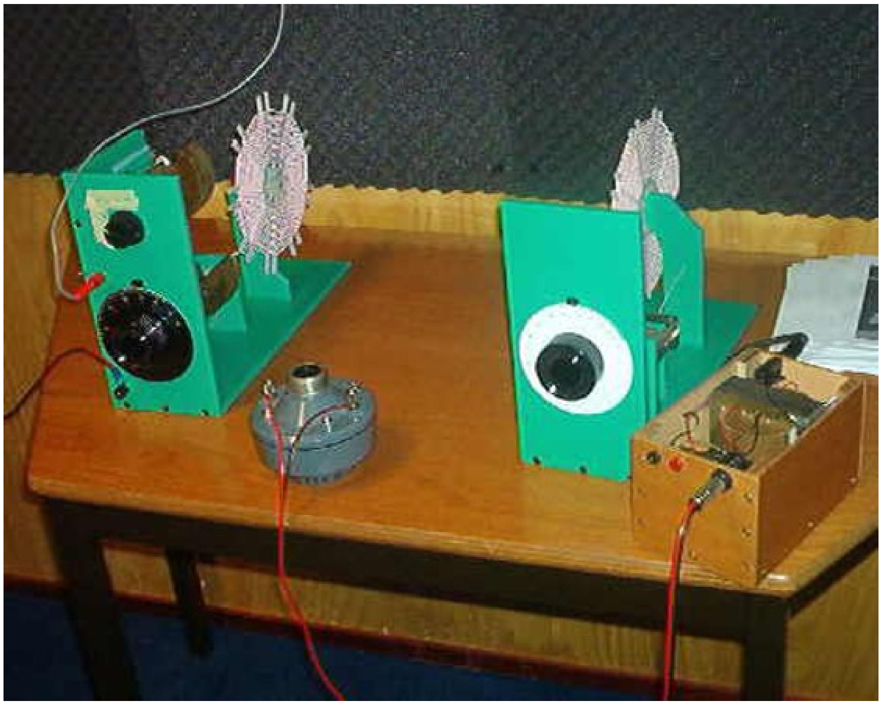

My receiver is build with separate unit's.
- the antenna unit
- detector unit
- audio transformer unit

The coils are made with litz wire 660/46 and are wound on very low loss polypropylene formers. The diode is a double shottky diode, type: HSMS282K. The transformer unit has a input impedance of 1600 k.Ohm, this is about the impedance of the detector LC circuit, so there is maximum power transfer.

As loudspeaker, I use a driver unit of a horn loudspeaker. The antenna is a 22 meter (72feet) wire, with the highest point on 18 meters (59 feet). On the picture, you see the antenna wire coming down from the top of the mast. More info about my receivers, you will find on my website: crystal-radio.eu

Antenna

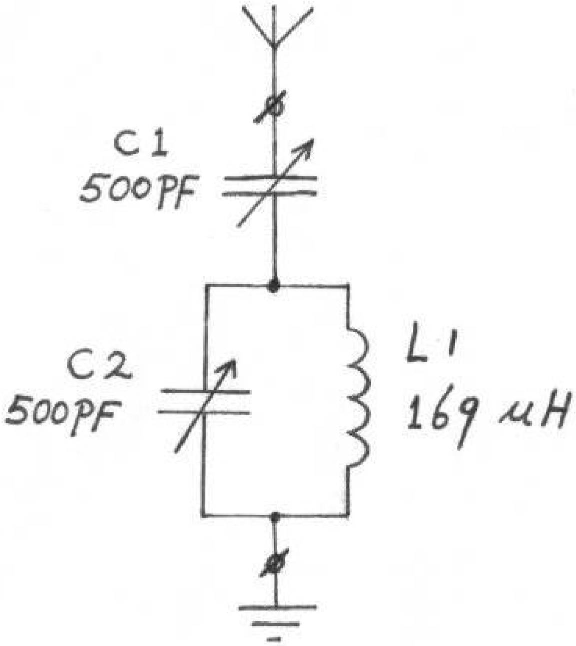
C 1
500PF
C 2
500PF
L 1
169 µH

Antenna matching unit.

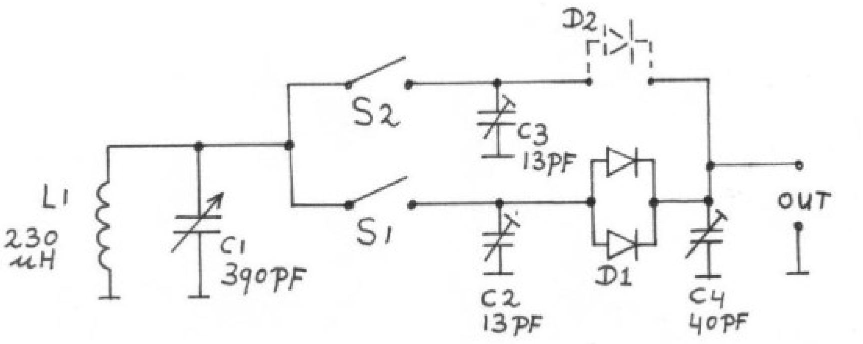
D2
S2
C3
13PF
L1
230
uH
C1
390PF
S1
C2
13PF
D1
C4
40PF
OUT

Detector Schematic

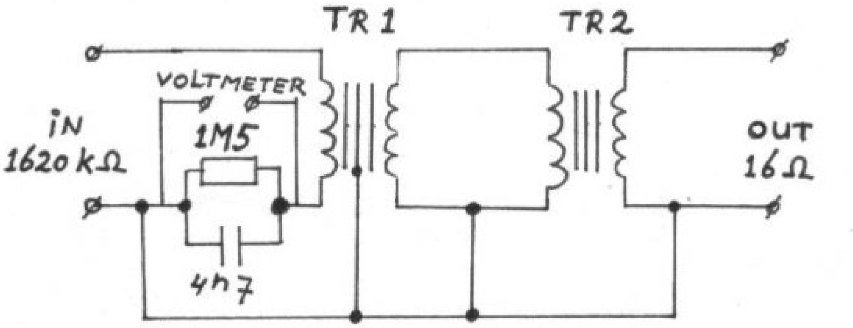
TR 1
TR2
VOLTMETER
IN
1620 kΩ
1M5
4n7
OUT
16Ω

Audio matching unit.

[Back to the main radios page.](http://www.crystalradio.us/crystalradios/2005-2.htm)

Select Menu

- Home
- About Us
- Contest Results
- 1AD Contests
- 1AD Entries
- Antenna Matching
- Antennas
- BHAM Radios
- Crystal Contests
- Crystal Radio Entries
- Sprint Contest
- Headphones
- Projects, Mods
- Favorite Links
- Contact Jack

Sub Pages

- Main Page
- 2004 Page 1
- 2004 Page 2
- 2004 Page 3
- 2005 Page 1
- 2005 Page 2
- 2006 Page 1
- 2006 Page 2

- Contest Results
    - 2004
    - 2005
    - 2006
    - 2007
    - 2008
    - 2009
    - 2010
    - 2006 Sprint
    - 2008 Sprint
    - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2006 Contest Entrants

Page 1

Charles Pullen (Chuckster)

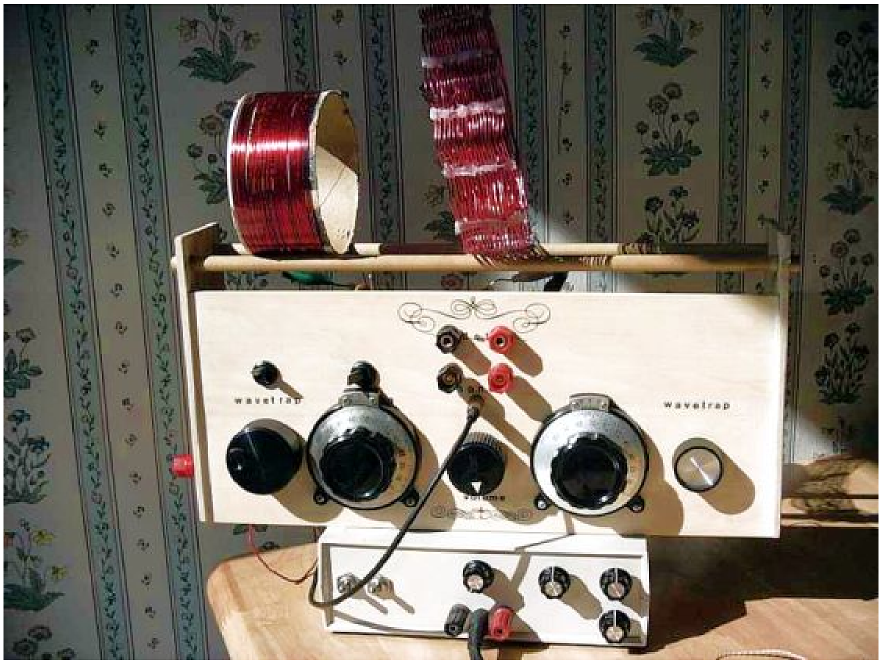

Station count was a little better than last year. I modified the metro to have a tuggle tuned antenna, added vernier drives for the antenna and detector capacitors and installed an old 5 gang fm tuner to give the detector circuit a selectable bandspread cap.

The antenna is a sixty foot inverted L approximately thirty feet up, and the ground is two four foot copper clad grounding rods tied together. I slapped to inline wave traps together to help tame the bandmasters. I used my copy of Steve B.'s ultimatch for impedance matching, and my head set was two philmore xtal earplugs connected to the ultimatch with a radio shack mono y adaptor.

My coils were severly mismatched and pretty beat up as seen in the photo of the "improved metro", but just ran out of time before the contest began to conjure up a new set of coils. I plan to replace the coils with toroids next year ala the TK2 circuit. Experiments with the toroids look really good but again didn't get them ready in time for the contest.

I'm pleased with the results I got considering the shape my coils were
in and the really warm weather conditions we've had …

Chuck

Dave Schmarder

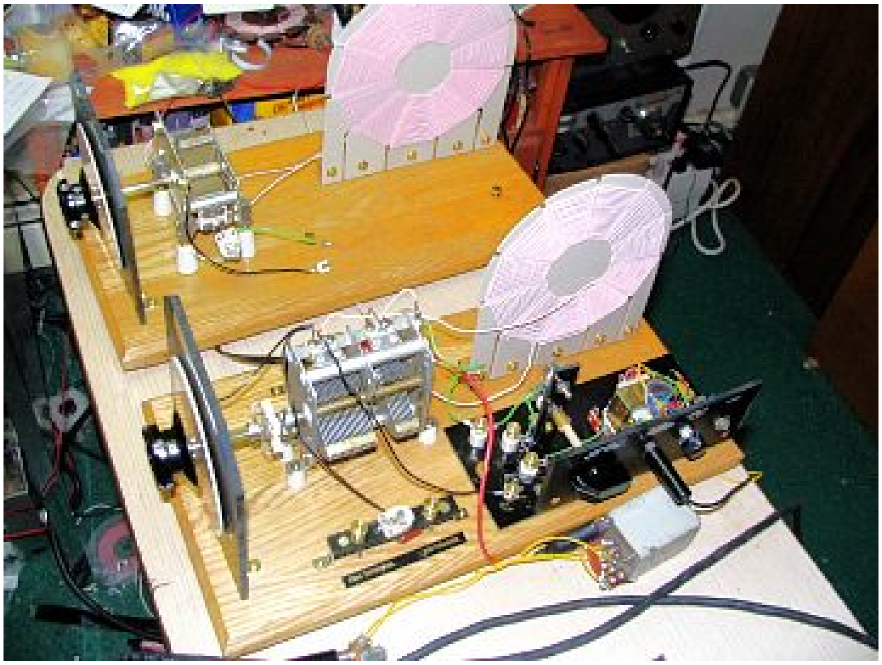

Here is the web page with additional comments, etc. Dave's 2006 Contest Log Dave's #63 Radio

...I used my #63 set with the crystal detector on the antenna tuning unit board...The other coil is a wave trap circuit.

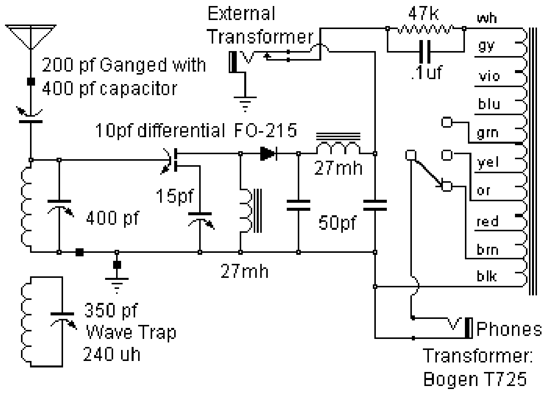
External
Transformer
47k
wh
gy
.1uf
vio
blu
200 pf Ganged with
400 pf capacitor
10pf differential
FO-215
grn
yel
27mh
or
15pf
400 pf
50pf
red
brn
27mh
blk
350 pf
Wave Trap
240 uh
Phones
Transformer:
Bogen T725

Hobby Class Crystal Radio (c) 2006, D. Schmarder

Dick Kleijer

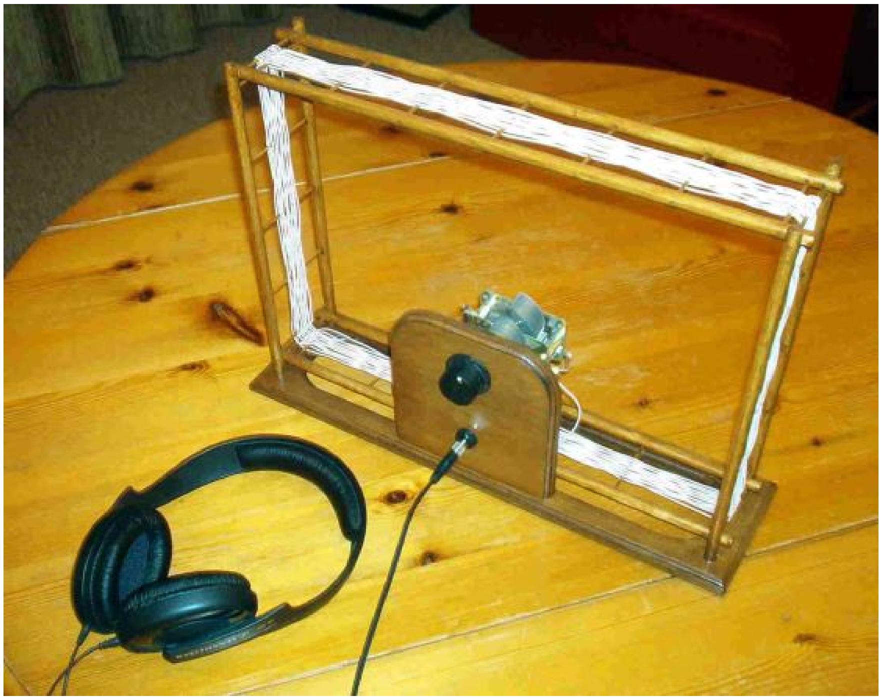

The receiver is a loop receiver with sides of 26x38 cm (10x15 inch). The wire is 660x0.04 mm litz wire (660/46).

Number of turns: 16

The two transformers have a total primary impedance of 300 k.Ohm. The headphone is a modern 2x32 Ohm type (Sennheiser HD202).

Dick Kleijer

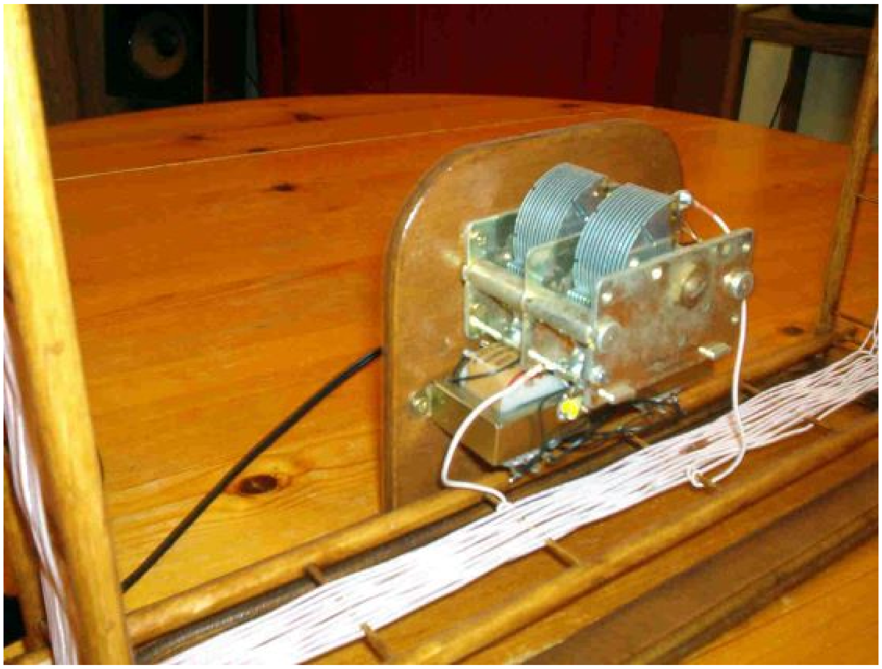

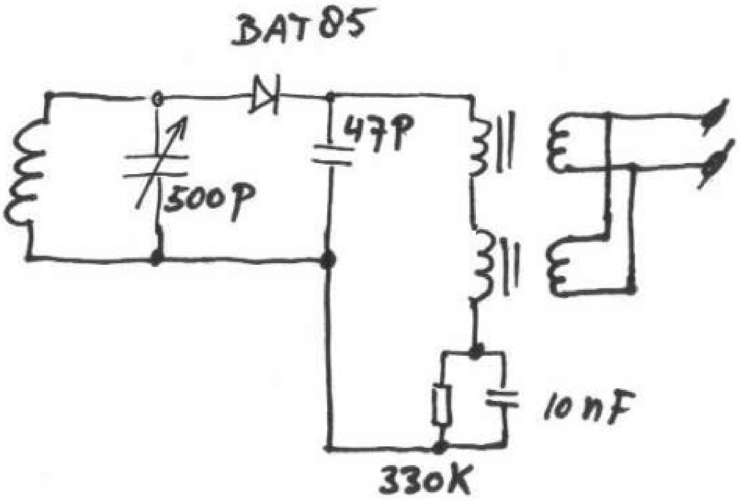
BAT85
47P
500P
10nF
330K

Doug Herigstad

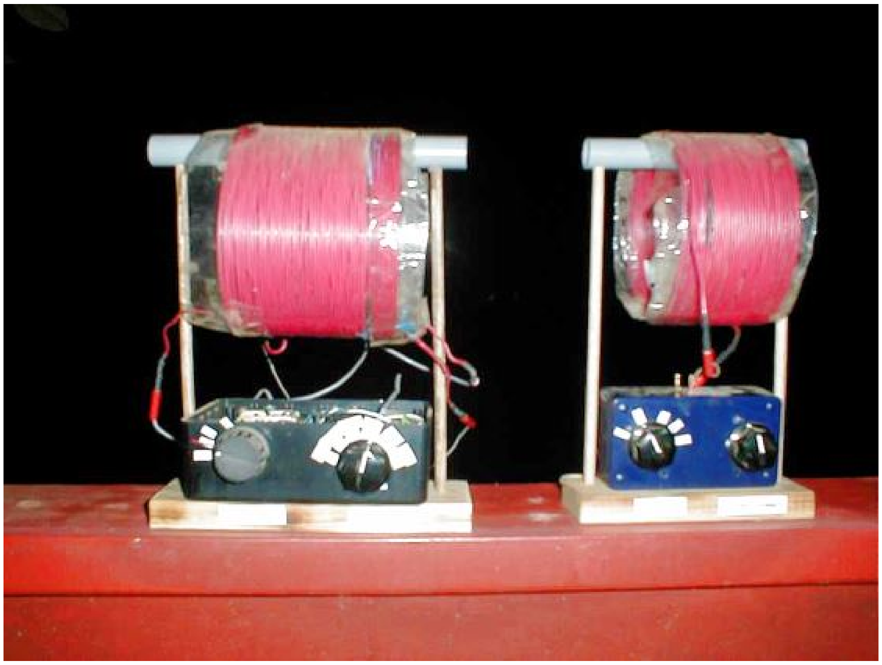

The radio I used is a double tuned set with the coils space wound on a 5” form. (3 liter soda bottle). I built a variation of Owen Pooles’ “Project Radio” on his website.

The antenna coil is 35 turns of space wound solid 18ga hookup wire. I used a combination parallel/series Tuggle tuner with separate 365pf vcaps with the series vcap in the ground path instead of between the antenna and primary tank. This gives me kind of a band spread.

The secondary or main tank vcap was pulled from an old battery set and has a 6:1 vernier built in. The main tank is 47 turns of space wound 18ga solid hookup wire on a 5” form. Only the antenna circuit is grounded to a cold water pipe.

The detector coil is solid 18ga wire wound on a 4” ferrite rod and fastened on the inside of the main tank. The detector itself is a 1N34A diode connected through a 39K ohm resistor bypassed with a .1 mylar capacitor connected to the white lead of a Bogen T725 transformer.

The headphones I used were the elements pulled from a discarded hearing impaired telephone series wired into a set of shooters muffs. These sound elements are much more sensitive that regular telephone elements. They work the best on the 375ohm tap off the Bogen. Much better than my 2000 ohm Newcomb magnetic headset.

The antenna itself is 12ga stranded wire 50’ high and 100’ long on the horizontal part oriented east to west. The far end is the west end …

Regards,
Doug Herigstad

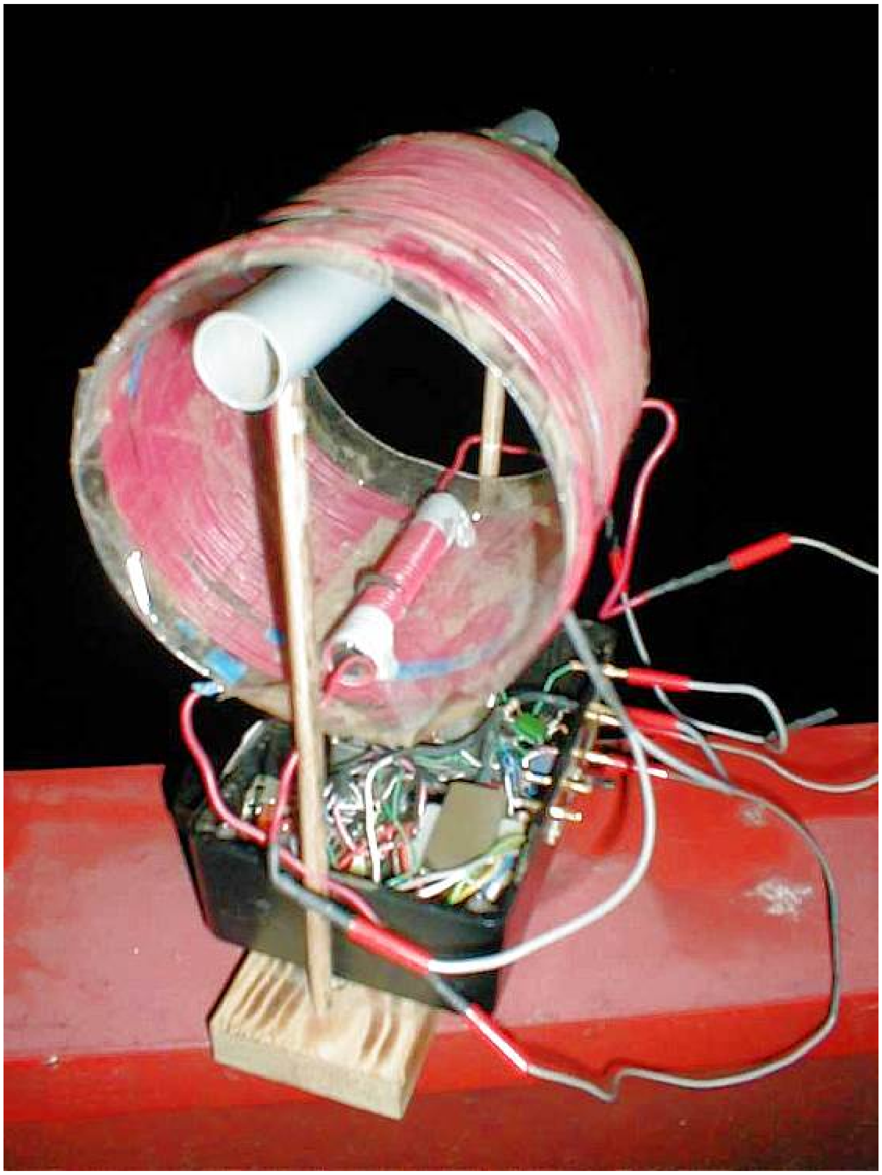

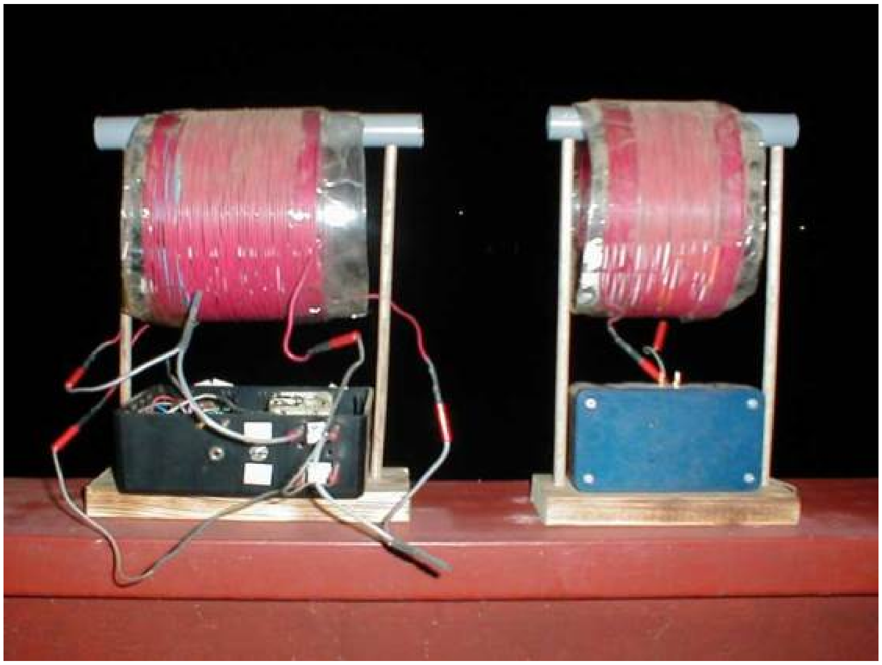

Ed Phelps

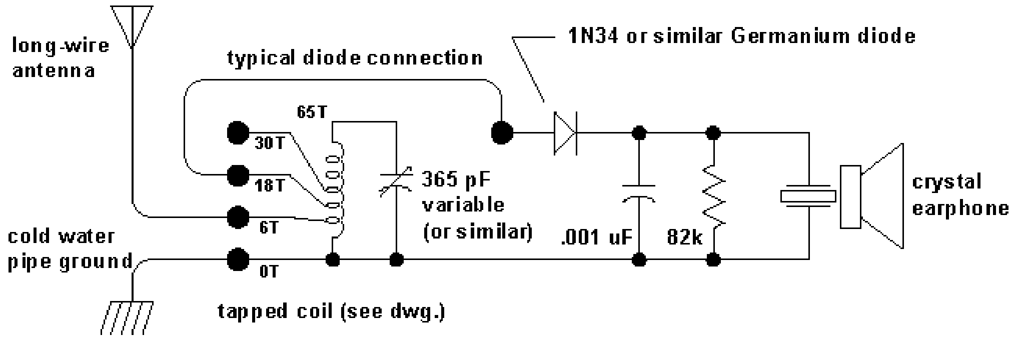
long-wire antenna
typical diode connection
1N34 or similar Germanium diode
65T
30T
18T
6T
0T
365 pF variable (or similar)
.001 uF
82k
crystal earphone
cold water pipe ground
tapped coil (see dwg.)

Crystal radio with typical connections for a long wire antenna and good ground connections. The diode is connected for weak signals and moderate selectivity.

I've made a few minor changes to the schematic …

(1) The fixed value cap is 250 pf instead of 1000 pf.
(2) The fixed resistor is 100 k instead of 82 k.
(3) I used an audio transformer plus a set of "big cans"...instead of the Xtal earpiece. (my skull is still distorted from those big cans) I think it'll still work without the fixed capacitor and resistor if I go back to using the Xtal earpiece. This would leave me with a set made from just four parts. I may try running it, in that fashion, next year.

73, Ed NN2E

Evan Haydon

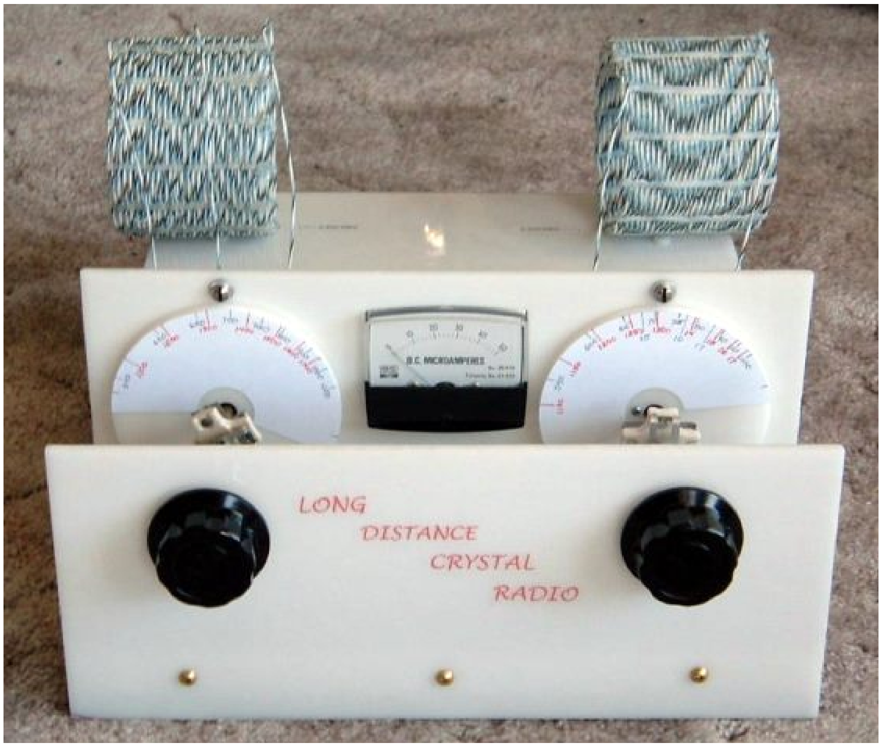
D.C. MICROAMPERES
LONG
DISTANCE
CRYSTAL
RADIO

The SUPERCRYSTODYNE

I decided to get serious about the 2006 crystal radio contest and make the maximum commitment to do my best. I knew that I could do better than what I did in the 2005 contest. I wanted to see what a maximum commitment would produce. I started the day after the 2005 contest ended. The following is commited, devoted, crazy, or whatever you want to call it. The following documents over 400 hours spent toward this goal. After all, I'm retired and have plenty of time. What better way to spend it?

First thing was to build a new and better crystal radio that had the maximum efficiency that I could possibly put into it. The design would use the same basic circuit as I had been using. The coils had to have the highest Q possible. I needed a test instrument to measure Q. Thank you Ebay for allowing me to purchase one. Now to get some good wire to use. I purchased some different guages of stranded silver teflon wire. Thank you

Ebay again. I wound many many coils. I found the proper size for maximum Q at the frequencies to be received. No larger or smaller wire size is better. No larger or smaller coil diameter is better. No more or less turns of wire is better. Maximum Q is found. Now, one pair of coils in the receiver will not maintain maximum Q all the way across the broadcast band. I use two different sets of coils. One set for 530-1100 kc. One set for 1100-1700 kc. Now I have maximum Q across the whole broadcast band! These coils are the basket weave type held together with hot glue. They are air core coils using 22 guage stranded silver teflon wire. The two sets of coils can be swapped on the radio in less than 60 seconds.

Both of the variable capacitors that I use have porcelain insulation for the stator plates. The range of tuning must cover as much of the 180 degree range of the variable capacitors as possible. This must be done for both sets of coils with no critical lossy switches or taps on the coils. This was accomplished by using dual section capacitors and a couple trimming or padding fixed silver mica capacitors as necessary. The only tap on either coil is one on the secondary coil for the detector crystal. The position of that tap was determined by actual on the air listening for maximum selectivity and sensitivity in that order. The support for all RF parts is white 1/4 inch thick plastic. All wiring in the radio is made using stranded silver teflon wire.

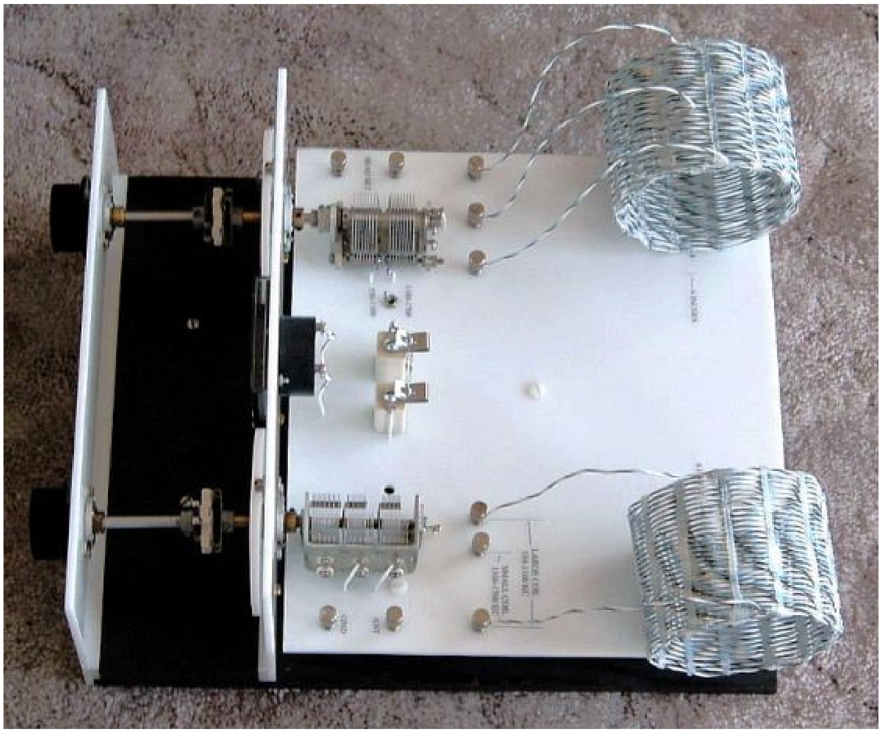

Now I had to have calibrated dials to quickly and accurately go to any frequency. Each of the two tuning shafts pass through the two panels. Each has a six to one vernier attached to the panel. So, each shaft has a 36 to 1 tuning reduction! On each shaft between the two verniers is an insulated flexible coupler. This isolates the tuning knobs from the RF circuits of the receiver to eliminate hand effect capacitance to the receiver. On the front of the rear vernier I placed a four inch diameter plastic disc. This disc is directly fixed to the shaft of the variable capacitor. On this disc I have paper dials calibrated for both coil sets. Each coil set uses a different colored ink. They are held on by double sided tape for easy removal. Thus, I can remove and make a new one in about a half an hour.

The detector is two parallel 1N34 diodes. They were selected for maximum sensitivity to a weak signal by on the air testing. I have a 50 microampere meter in series with the diode detector. A signal of less than 1 microampere is 100% readable. I am about 700 km from Chicago and Denver. It is not unusual for KOA or WBBM to come in at a full scale reading of 50 microamperes.

The headphones that I use are unique. I started with two USI balanced armature low impedance elements taken from an old sound powered handset. I first took them apart. I removed the coils. I took all of the wire off of the coils. I rewound them using super hair fine wire from a 6000 ohm sensitive relay coil. Each headset coil now measures well over 2000 ohms dc resistance. They were installed into a cheap, but nicely padded, headset that I already had. They are very sensitive. I also like the frequency responce of them as they reproduce higher audio frequencies better than the lows. The lows tend to be "muddled" on crowded frequencies.

Last summer was antenna experimenting time. I put up, tested, and took down more different sizes, lengths, heights, and orientations of wire than I can remember. Needless to say, considerable time was spent on this aspect of my endeavor. What I wound up with was a wire about 200 feet long that generally runs east and west. It is up 40 feet at the east end, 60 feet up at a point 120 foot west of that, and then sontinues 80 feet southwest of that ending up 25 feet above ground. The down lead comes off at a point close to the 60 foot support. All wire in the antenna is 18 guage stranded silver teflon wire. I have a ten millihenry choke to ground at the listening bench. Before I put that on, I saw 3/4 inch sparks from the antenna.

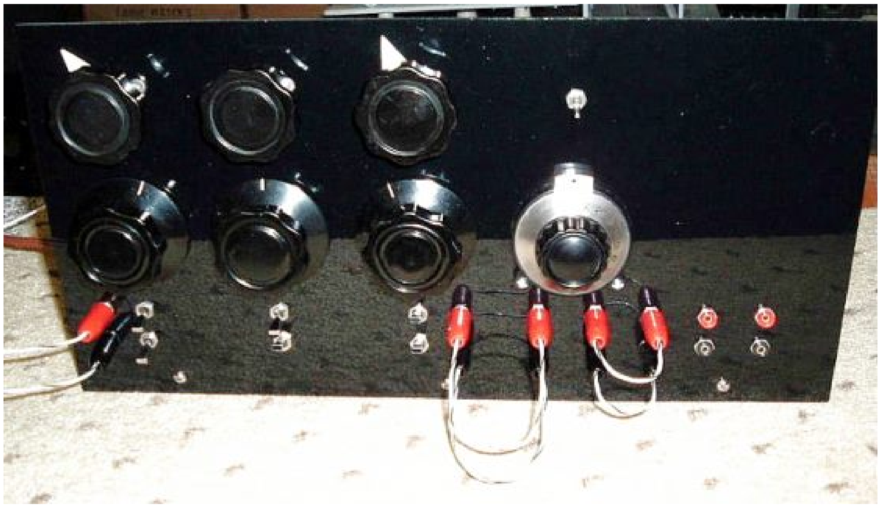

Designing filters to kill local stations became a must. I have a one kilowatt station only two kilometers from me. The Q meter came into use again. I ordered a number of each of three different sizes of torroids from Amidon. The coils for the traps are designed for maximum Q at the operating frequencies. I needed five for the top of the band and only one for the bottom of the band. They are wound using 22 guage stranded silver teflon wire. Each has a small 365 pf capacitor in parallel with it. There is a small winding on each for coupling. These small coupling windings can be put in or out of series with the antenna with a flip of a switch. Any number of traps can or can not be used at any time. They are all mounted on plastic to preserve their Q. These traps work very well and tune very sharp. This filter unit is a separate unit from the crystal radio. Also on this unit is a variable capacitor with poreclain insulation and silver plated plates. It is used to trim the antenna when needed.

By now, I was getting a good start on my quest. Next up was getting familiar with the broadcast band. This means lots of time spent listening. I listened almost every evening and night from the middle of August until the contest started in January. I estimate that I spend at least 300 hours listening and gathering information. I systematically monitored every frequency at all times of the day and night and carefully documented which stations could be heard at what time and on what frequency. The more times that I heard an individual station, I knew that on any given day there was a good chance of my hearing it. All of this information was used in the contest for maximum efficiency of the time available. This is critical because many stations can be heard only during a small window at sunset. Last year in the contest I didn't even listen at this time of day figuring that all of these signals were too weak to be heard. Wow, was I missing the boat or what? This is the best time to listen. I added almost 150 stations this fall to my all time heard list by listening during this period of the day. The information I then had for possible reception of stations was put to use in the 2006 contest. It worked well. 144 stations logged on the first day! The broadcast band conditions on that first night were the best that I have experienced in the past two years. They were nothing short of amazing. I put in 14 hours of listening and logging on that day alone. On every day of the contest I was by the radio for a couple of hours around sunrise, off and on during the day, and from 4 PM until about 11 PM. A person just can't miss that one more station logged. I put in at least eight hours of listening and record keeping for the contest for each of the nine remaining days of the contest. Almost 100 hours during contest week.

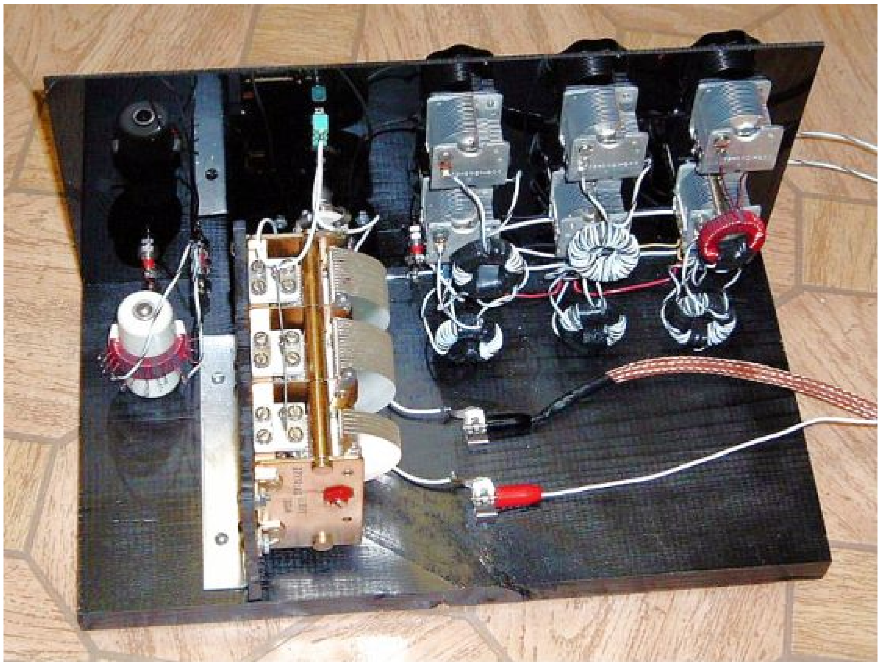

Total time spent for construction, preparation, and contesting turned into somewhere near 400 hours of fun. Every day listening is different. That's what makes it so interesting and fun. The band conditions were very good almost every day of the contest. Some nights produced longer

- Contest Results
  - 2004
  - 2005
  - 2006
  - 2007
  - 2008
  - 2009
  - 2010
  - 2006 Sprint
  - 2008 Sprint
  - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2004 Contest Entrants

Page 3

Lem Morrison

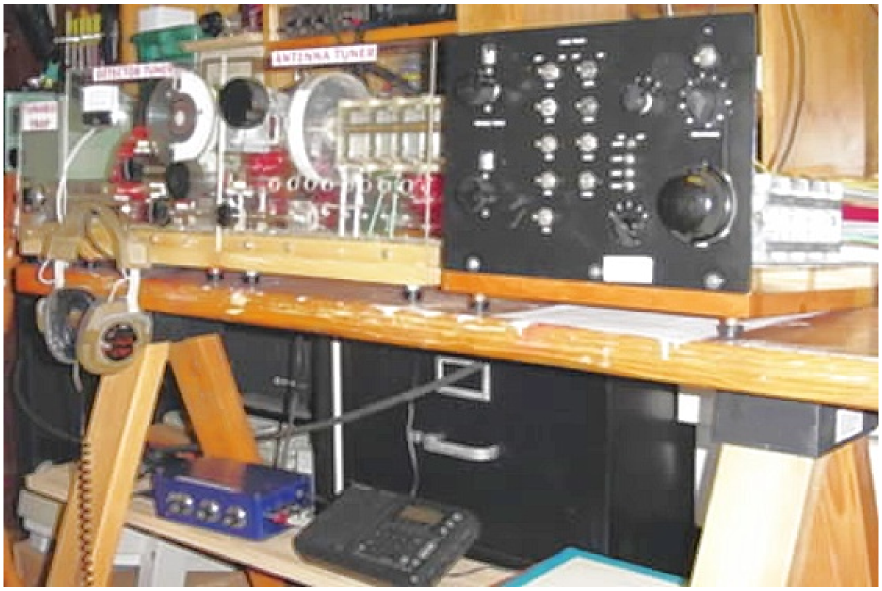

N4AHJ 2004 CONTEST CRYSTAL RADIO and OTHER INFO

ANTENNAS:

(1) End-fed inverted V long wire, 135 feet with apex at 40 feet, two 135 feet counterpoises under the antenna.

(2) Vertical, 48' tall Tee; top-loading consisted of two parallel wire spaced 38' and run 90' to Southwest and 40' to Northeast; 10 radials with lengths from 40' to 100'; series fed at base with single wire feed. Vertical core consisted of a cage of four parallel wires spaced 12" and two wires from each of the top-loading legs dropped down, spaced 4" from core center, and connected to base to provide broader bandwidth.

(3) Sloper, 118' with far end at 50' and station end at 8', oriented toward East-Northeast.

(4) Sloper, 130' with far end at 50' and station end at 8', oriented toward West-Northwest

Ground system: Ham station ground - six ground rods interconnected and connected to shack with two quarter-inch copper straps; supplemented by one 8' ground rod directly under Vertical antenna and 2-4' ground rods under the wires dropped from the top-loading legs, all inter-connected with copper wire; both ground systems connected with two runs of copper braid salvaged from old RG-8 coax. Radials were 17-ga aluminum wire (used for electric fences).

TRAPS:

(1) Tunable(1) - 2-section variable, sections in parallel, 240pF total; Rook coil with 18 gauge wire, 5.25 inches effective diameter, 47 turns; air-coupled to the Detector coil.

(2) Tunable (2) (located in Antenna Matching Unit) - both use 14 gauge enamel wire on FT-114A-61 cores and dual-section variables with both sections in parallel.

(3) Fixed (9) (located in Antenna Matching Unit) - 14 gauge to 22 gauge enamel wire on FT-114A-61 cores, silver mica fixed capacitors, and plated ceramic trimmers (one trap for each of the local "powerhouses"). Each link-coupled to the antenna feed to the set.

ANTENNA MATCHING UNIT (AMU):

One 4-section variable capacitor, four fixed capacitors (500pF, 1000pF, 2200pF, and 5000pF), and two coils (42 turns 14 gauge enamel on a FT-240-61 and 26 turns 16 gauge enamel on a FT-114A-61). All components switched to provide four matching arrangements: series LC, parallel LC, inductance only, and capacitance only. Both inductors tapped to provide 23 settings. An antenna switch is provided for switching between up to five antennas.

TUGGLE TUNER:

Capacitor - 520pF variable, 4 sections, 2 used in Tuggle arrangement, vernier is a surplus right-angle unit with a 25:1 ratio.

Coil - Two Rook coils with 16 gauge enamel wire, 5.25 inches effective diameter, 15 turns each, 30 turns total; bandswitched to provide either series or parallel coupling to provide Low band and High band tuning, a la Ben Tongue's arrangement; air-coupled to the Detector coil.

DETECTOR TUNER:

Capacitor - 420pF variable with home-made vernier with a 32:1 ratio (vernier constructed from gears salvaged from a VCR and a TV tuner).

Coil - Two Rook coils with 16 gauge enamel wire, 5.25 inches effective diameter, 20 turns and 21 turns, 41 turns total; bandswitched to provide either series or parallel coupling to provide Low band and High band tuning, a la Ben Tongue's arrangement; switchable taps (High Band: 100%, 75%, 50%, 25% and Low Band: 100%, 87.5%, 67.5%, 50%, 25%) on coils to provide matching to diodes.

Diodes - Three 1N34A Radio Shack equivalents, paralleled and matched for Forward and Reverse resistances.

AUDIO:

Matching - Reproduction of Steve Bringhurst's Ulti-Match circuit.

Headphones - Two Automatic Electric GH-66919 elements (manufactured 1942) out of a military handset.

NOTES:

The basic set is the same as my 2003 contest set. I moved the fixed traps from the Tuggle Tuner to the AMU and complemented them with two tunable traps. There were a number of times when I used all three tunable traps and one, or more, fixed traps!

I determined optimum AMU settings across the AM BCB for each antenna prior to the contest.

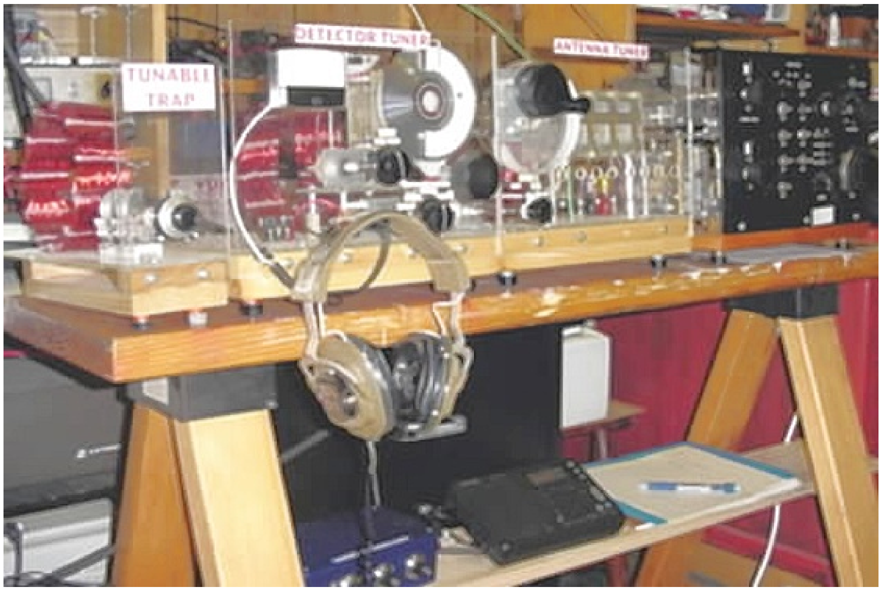

Notice the spotter set and audio matching unit on the bottom shelf.

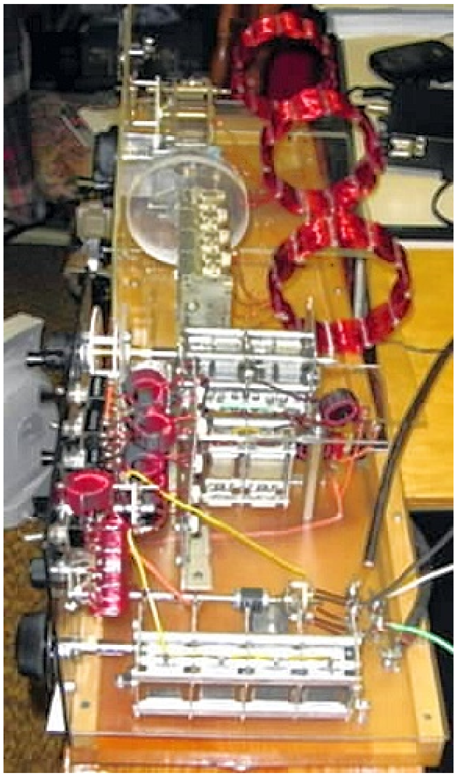

View from right side

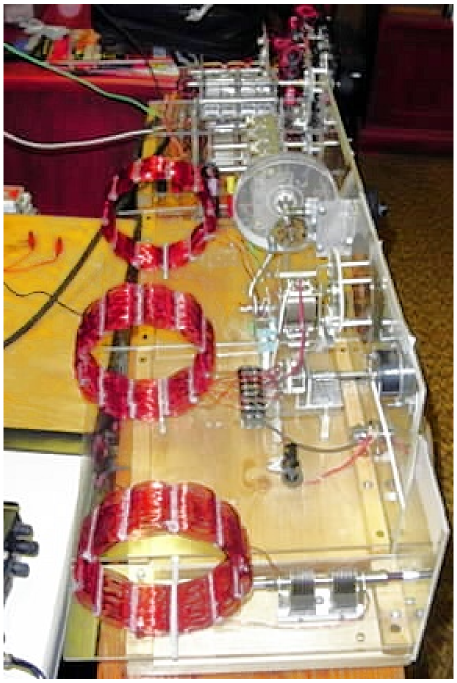

View from the left side

Ed Phelps

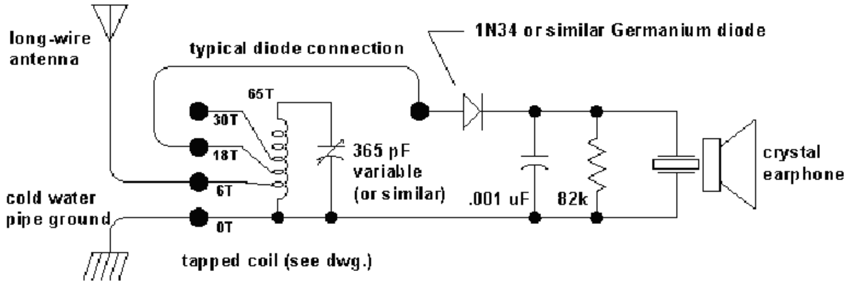
long-wire antenna
typical diode connection
1N34 or similar Germanium diode
65T
30T
18T
6T
0T
365 pF variable (or similar)
.001 uF
82k
crystal earphone
cold water pipe ground
tapped coil (see dwg.)

Crystal radio with typical connections for a long wire antenna and good ground connections. The diode is connected for weak signals and moderate selectivity.

Click to view larger schematic.

Mike Schroeder

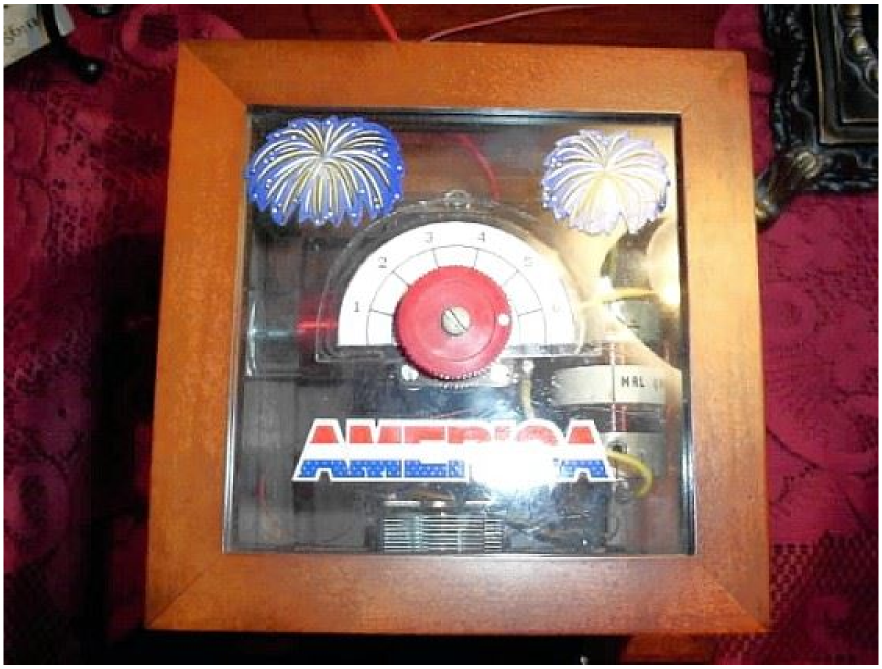

Top View

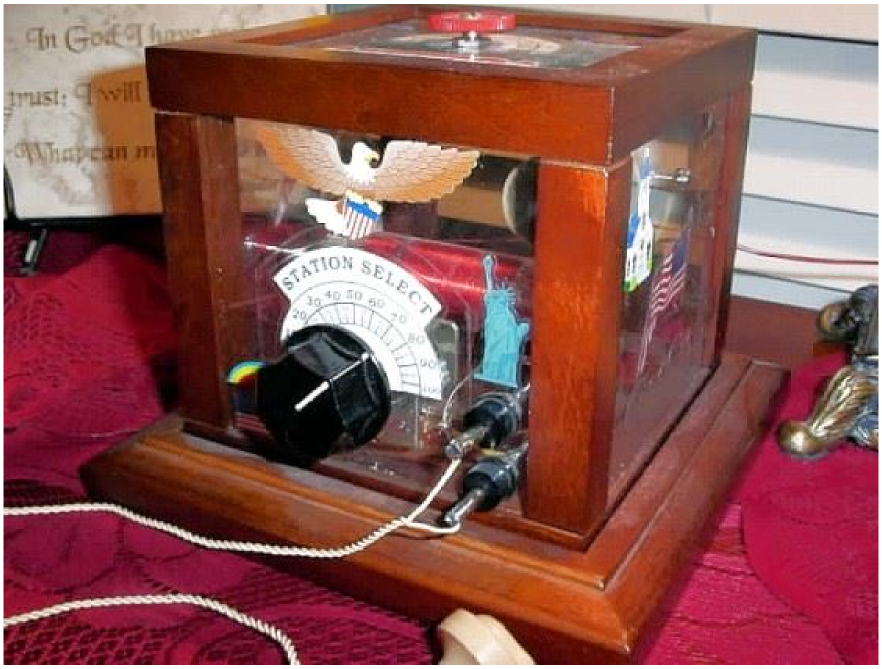
STATION SELECT

Front View

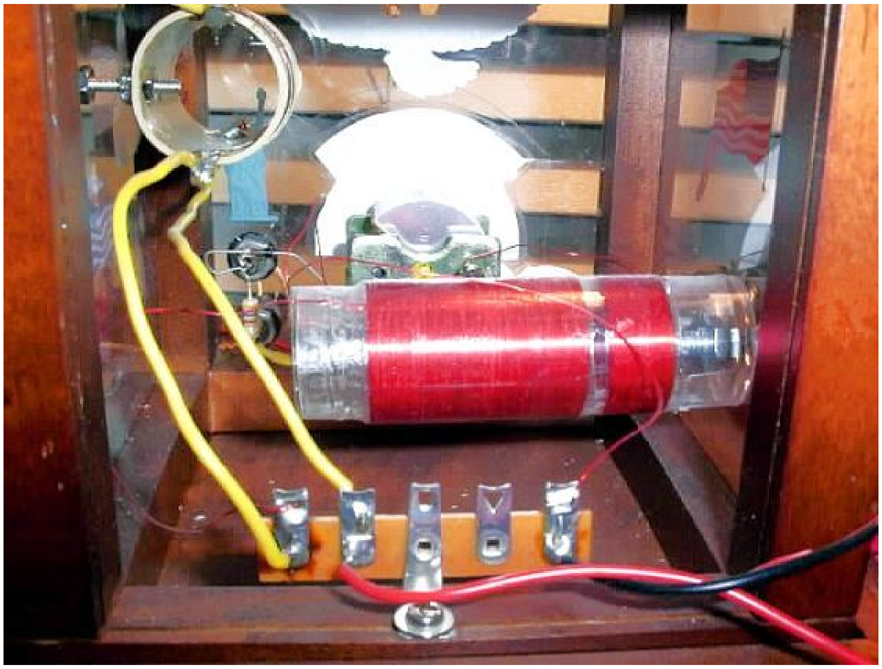

Rear View

Dave Schmarder

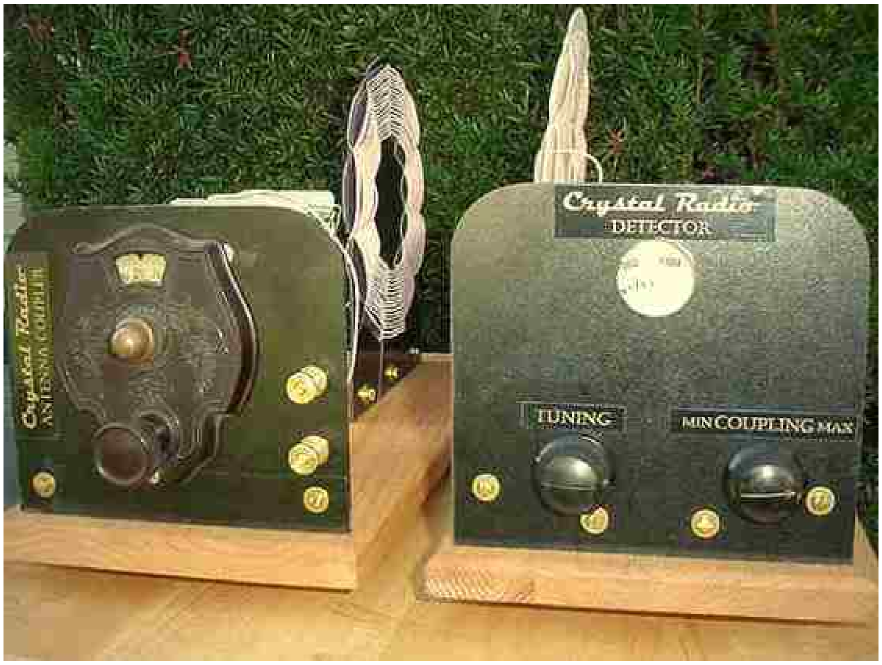

These pictures and discussion are from Dave's site at: Dave's #35 DX Radio. His full description of this receiver is at his makearadio.com website.

This is the Crystal Radio to End all Crystal Radios. Hello there. Can't you just feel the tension rising? With this new radio I shift gears again. Large litz wire, high quality variable capacitors, schottky diodes. It won't get much better than this. Last night I am proud to announce that I heard KMOX in St. Louis on this radio. That is my furthest DX ever. I am sure that the ionosphere was hot and the wind was blowing my way, but a heard station is a heard station. Enough bragging and on with the show.

This radio is built in modular form. Mike Peebles gave me the idea when he had his modular crystal set published in the Crystal Set Society newsletter. The first part built is the detector. This consists of the main tuned circuit with my 660/46 Überlitz Coil. This is of my own design. I am a big fan of litz spider coils, and this is a big litz spider coil. I believe I hit crystal set heaven with this coil. The detector unit also has a selectivity / sensitivity control using a differential capacitor. Besides all of this there is, of course, the detector diode itself. I made a board with a three position link switch and places for three diodes to be connected. This allows me to easily compare 3 diodes in the same set. After the diodes is an audio output transformer and RC network. The 27 mh choke provides a DC return path for the diode circuit.

The variable capacitor is an old one made by National Radio. I made a wheel and fashioned a vernier tuning system with a grommet and a dowel rod. I get nice, slow and smooth tuning with this system. You can use a vernier dial instead if you don't have one of these capacitors in your junkbox.

The selectivity and sensitivity are handled by a differential capacitor and a piston trimmer. This system was first used in the Hobbydyne II circuit by Jim Frederick. The piston trimmer is for tracking adjustment. You adjust the piston trimmer until the main tuning needs no touch up after turning the differential capacitor. This is a breakthrough in weak signal reception on a crystal set.

The audio transformer matches the very high (100 k ohm +) detector circuit with the much lower impedance headphones. I am using a 100k to 1.5k ohm transformer. If you can find a high quality transformer, go for it.

This covers the detector section of my high performance radio. Next comes the antenna input circuit and maybe a wavetrap.

My tuning unit is now complete. The coil is made from 660/46 litz. The hub diameter of the coil is 3 inches. The coil form made from 1/16 inch styrene sheet is 6-1/2 inches wide and 7-1/2 inches tall. There are 33 turns on it for an inductance of 150 µH. The variable capacitor has 600 pf per section. Smaller ones can be used, however the low end of the band might not tune. You could probably get away with a 500 pf per section capacitor. This antenna tuner is wired as a "Tuggle Tuner".

Earlier I had used a 150 micro henry coil made from 40/44 litz wire. This worked quite well. One night very late, I scanned the dial and with this set up I heard 40 stations. Not bad for a September night.

The two units are placed next to each other. The coil distance for good overall operation are about 15 inches apart. Careful adjustment of the antenna tuner and detector tuning are required. The Q of these types of coils have been measured at over 1000! This ain't your grand pappy's crystal set.

If you build one of these, let me know how it plays for you. Happy crystal set dx. Dave, N2DS

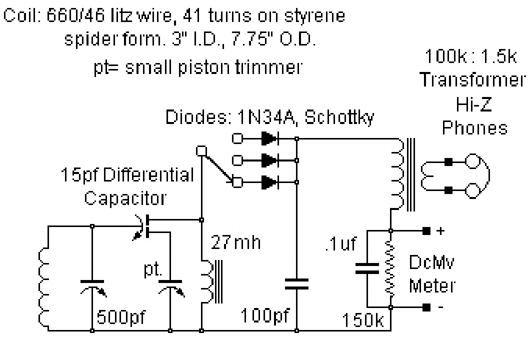
Coil: 660/46 litz wire, 41 turns on styrene spider form. 3" I.D., 7.75" O.D.
pt= small piston trimmer
100k : 1.5k Transformer
Hi-Z Phones
Diodes: 1N34A, Schottky
15pf Differential Capacitor
27mh
.1uf
pt.
DcMv Meter
+
-
500pf
100pf
150k

Crystal Radio Detector Schematic (c) 2004, D. Schmarder

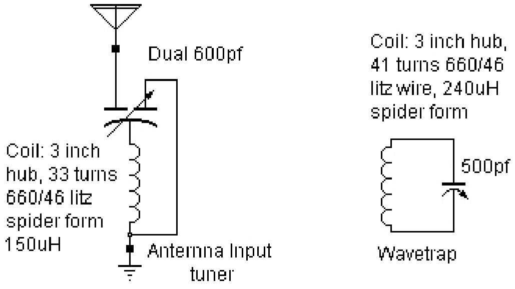
Dual 600pf
Coil: 3 inch hub, 33 turns 660/46 litz spider form 150uH
Antenna Input tuner
Coil: 3 inch hub, 41 turns 660/46 litz wire, 240uH spider form
500pf
Wavetrap

Crystal Radio Antenna Tuner and Wavetrap
(c) 2003 D. Schmarder, N2DS

Back to the main page.

SEE WHAT MY CLIENTS ARE SAYING!
Learn More
30 recommendations
Eduardo Ledesma
Real Estate Agent
Realty One Group
Cal BRE: # 01198526
Ads Powered by Reach150

- Contest Results
    - 2004
    - 2005
    - 2006
    - 2007
    - 2008
    - 2009
    - 2010
    - 2006 Sprint
    - 2008 Sprint
    - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2008 Contest Entrants

Page 1

Mike Tuggle, Lyonodyne-17, Open Class Entry

"So, the solenoid is king, the spider needs a make-over and the basketweave is done." - from a post on Rap-'n-Tap -

Right.

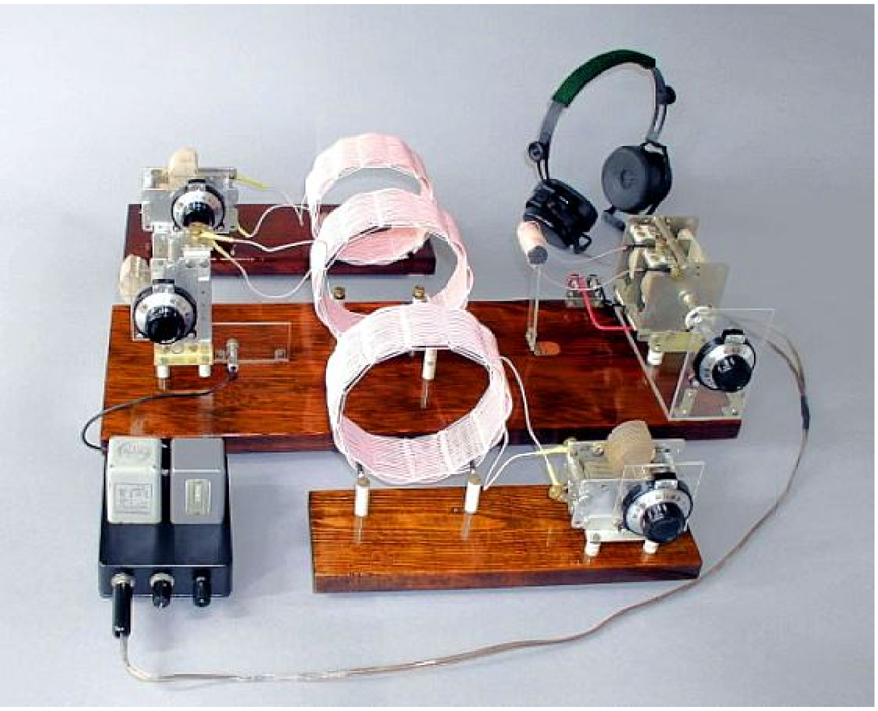

Mike Tuggle, Loop Class Entry

This double-tuned set is built onto a Steve McDonald (VE7SL)-design alt.-az. loop antenna.

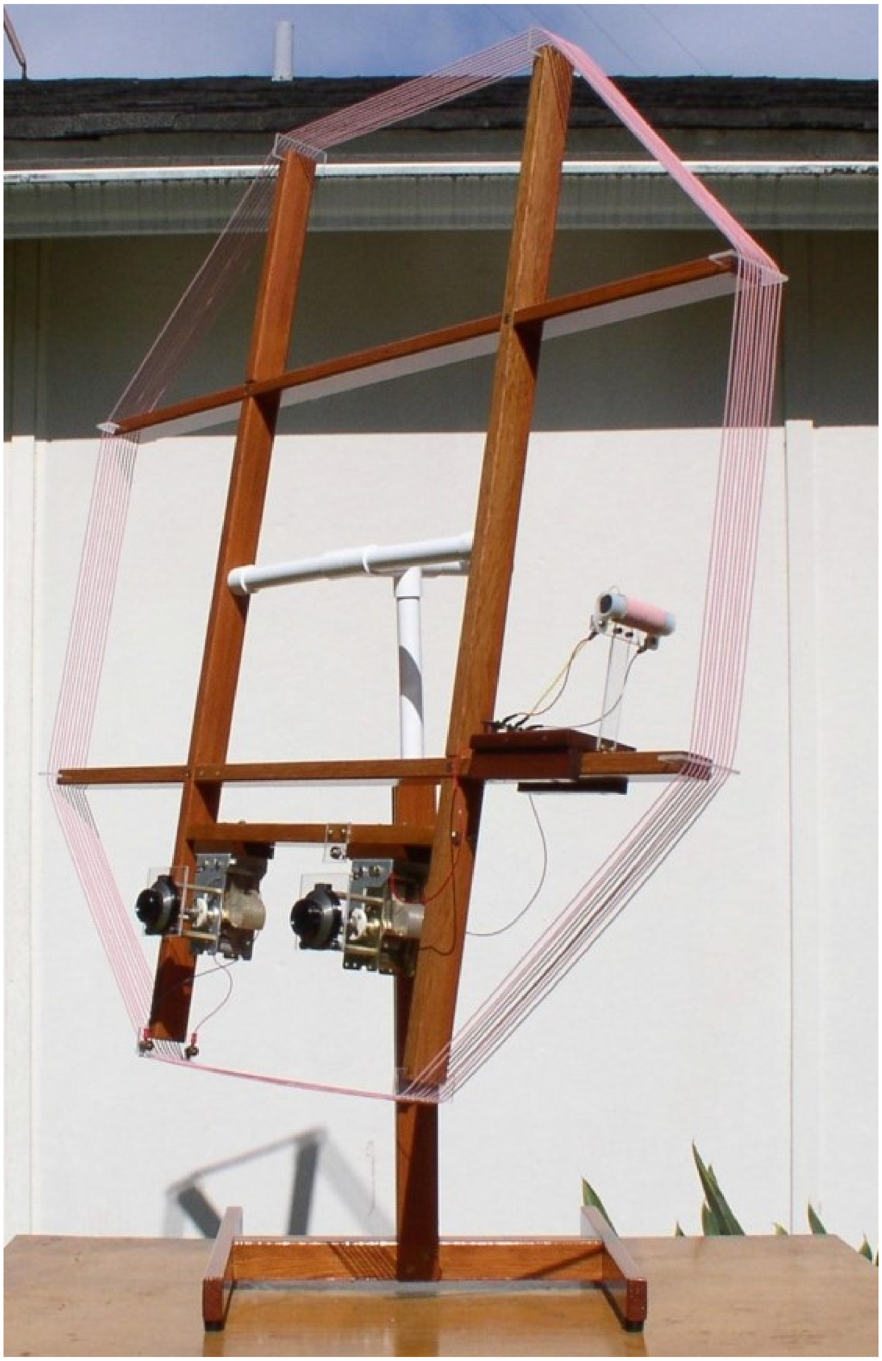

The ferrite coil slides on a loop strut to vary coupling. The idea was to see if mainland stations (2400 miles, and greater, away) could be heard here in Hawaii on a passive loop crystal set. The answer is, yes -- just barely. I really could have done with a bigger loop. Maybe next year.

Leading up to this circuit, I tried several designs using zero bias-voltage MOSFETs. A tapped ferrite coil or a pair of 5-inch diameter basket coils was used for gate and drain-source circuits. The basket coils were way over-coupled. None of the MOSFET circuits had promising sensitivity for possible mainland reception. Only by going to an ITT FO-215, kindly provided by Steve Bringhurst, could the big hop be made. This repeats my finding in last year's XSDX: The MOSFETs provide good selectivity but are not all that sensitive.

Twenty-five stations were heard, five of them DX on the mainland. Two of the DX stations were co-channel with strong local stations. I suspect the DX stations had exalted carrier help from the locals. This may be a potentially exploitable property of loop antennas: their characteristically sharp nulls can be directed to knock down strong local stations while, at the same time, leaving some carrier to exalt co-channel DX stations lying anywhere off the null direction. However, complete nulling the local station takes out everything.

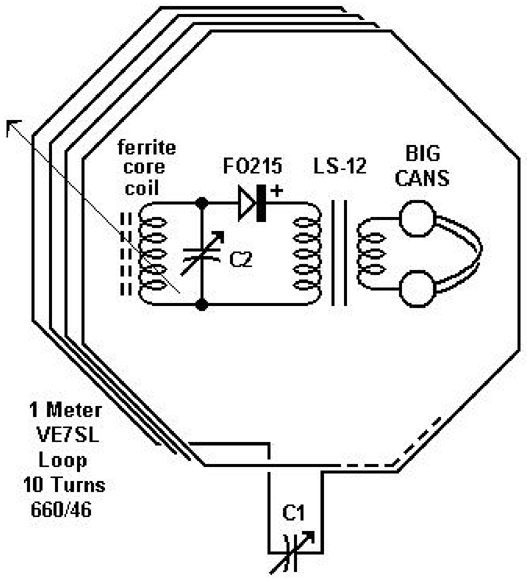

Evan Haydon

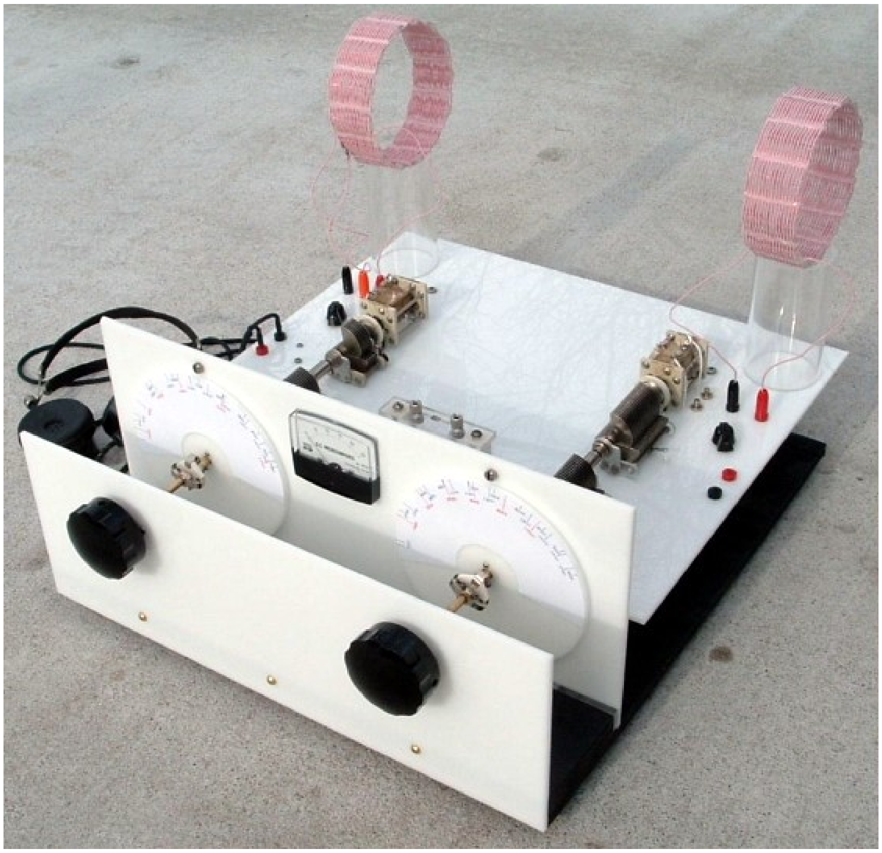

I built a new crystal radio this year. I am now using 660/46 litz wire for my coils. One pair of coils for 530 kHz to 950 kHz. A second pair of coils for 950 kHz to 1700 kHz. They are basket weave with a five inch diameter. I wound solenoid coils of many diameters, inductances, and wire spacings with the 660/46 litz. None of them approached the measured Q of my five inch basket wound coils. My calibrated dials are six inches in diameter with two 6:1 verniers in series driving each dial. A third scale on each dial covers 1500 kHz to 1700 kHz. A pair of dual 8-60 pf silver plated porcelain insulated variable capacitors are switched in for 1500 kHz to 1700 kHz. Silver plated trimmer capacitors calibrate them to cover the dial. Overall selectivity and sensitivity are to the point that I really can’t complain.

This year’s contest was fun because the band was in flux all week. The first day had average to good conditions. The next three days were very poor with spotty receptions. The next two days had cold fronts pass through and stir things up. There were some unusual openings to the southwest. The last days were lively during the gray line times with some new stations for my all time heard list. That list is now 469 stations.

Here is a table of my activities during the 2008 contest.

DATE	#HRS	#STATIONS
1-18	13	168
1-19	13	35
1-20	10.5	18
1-21	7.5	9
1-22	7	15
1-23	9	20
1-24	7	9
1-25	5	8
1-26	6	16
1-27	9.5	3

Total of 87.5 hours and 302 stations. 291 stations identified. 11 stations not identified.
Total points: 556,834
Just in case you haven't guessed by now, I only eat, sleep, and work the crystal radio contest for 10 days in January.

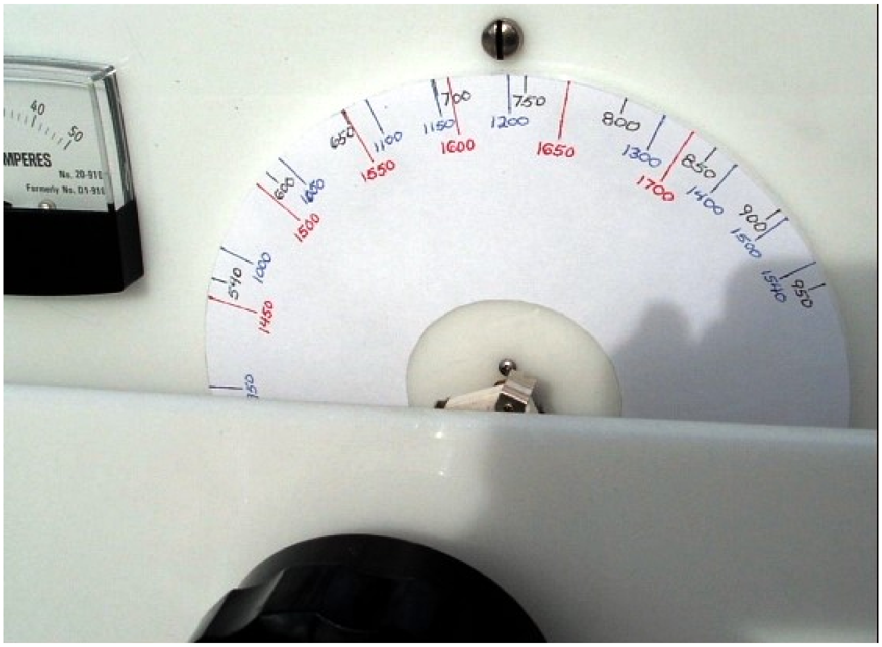
40
50
MPERES
540
1000
1450
600
1500
650
1550
1100
700
1150
1600
750
1200
1650
800
1300
1700
850
1400
900
1500
950
1540

Dan McGillis

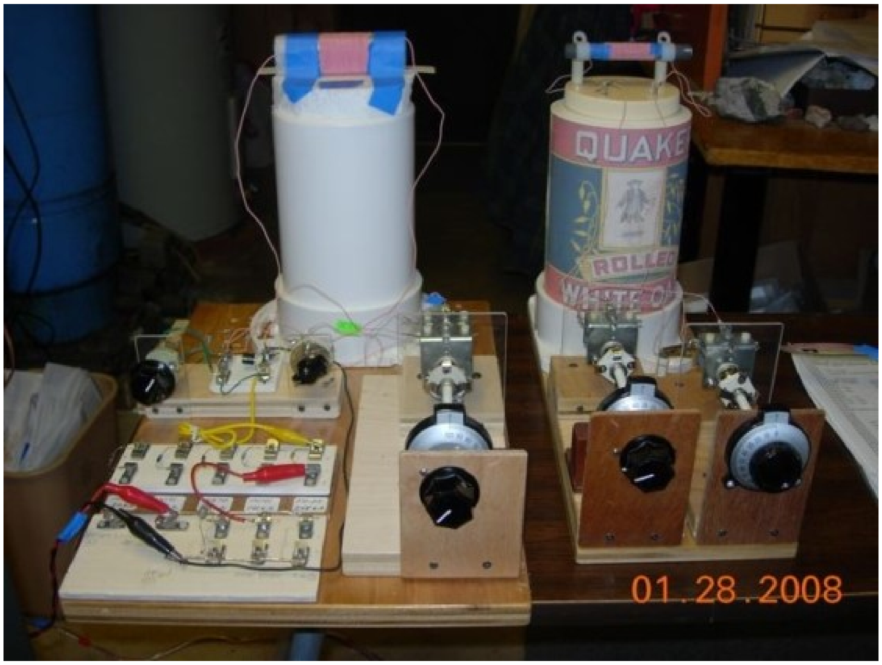

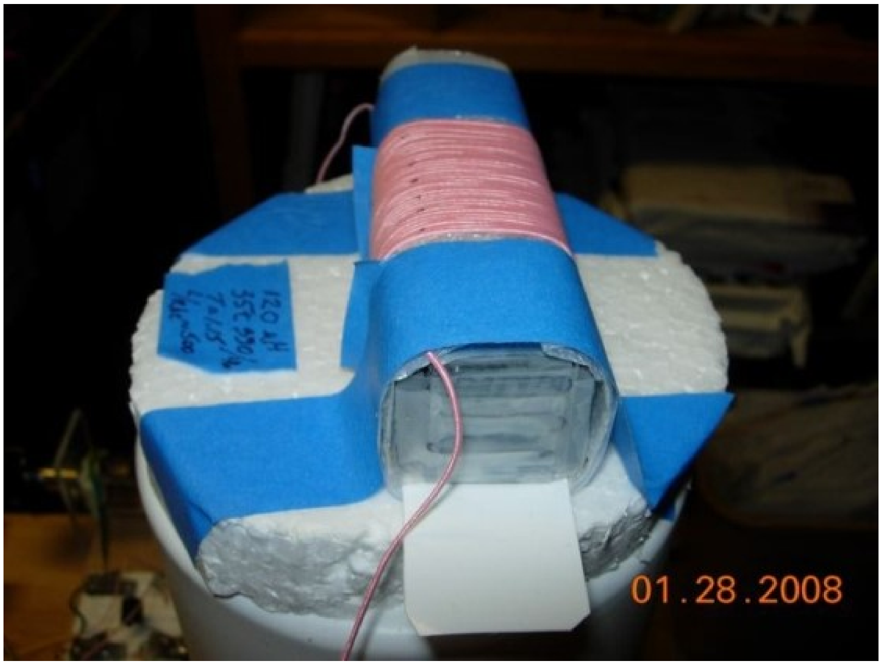

The set-up is a standard double-tuned arrangement with hobbydyne coupling. I tried to learn from and use the info that folks have so generously posted on R'nT and Dave's forums. Wish I could "give back" as much as I've received.

Antenna: 200' inverted "V", apex ~ 40'. I "sling-shotted" a wire over the highest tree near the house.

Ground is a 4' pipe into damp soil + a wire into a pond.

The set is Double Tuned: Ferrite inductors 10" above table top on empty HDPE containers. Getting the coils up away from the table made a BIG difference.

Detector Coil: Vintage ferrite bars, 9x2, 1/8” sep, 35t cw 330/46, 120 µH. "Battradio" ferrite bars, 2 stacks of 9 (7 high+1 each side) Wrapped with packing foam. This is kind of a 1.25"x1.25", "squarish" coil. I used Used the recent posted Q info with the wire I had available.

Dual gang 10-410pf cap, series-or-parallel (ceramic DPDT switched). The switch in the tank circuit didn't seem to hurt selectivity. Two tuning ranges gave good mechanical bandspread (<25 khz/div @ max f) via Jackson Bros. ceramic insulated cap, + 48:1 vernier, 0-100 dials. The Tuning is VERY sharp - really need the verniers (8:1 dial x 6:1 planetary).

Tuggle tuned ATU: 0.5”x4” Amidon 61 rod, 50t cw 165/46, 203 µH (~best for this antenna). 4" Amidon rods give better Q than 7.5" rods. Separate 365 pf Jackson Bros. caps + verniers. I needed a 48:1 vernier on parallel cap - tuning is sharp.

Hobbydyne: ~ 17pf (ceramic air variable) + 27 mH; choke (Dave Schmarders'). Helped at high end but probably don't need it with these ferrite bars.

Diode: FO-215 diode, rd~258k. (Picked the one with largest rd.). I could switch-in schottky's etc, but this was best overall match.

Audio: Dave's 200kUberformer + 500k Benny, shunted by 1µF. Two USI-UA1611-1 SP elements in series. (200k:2.5k).

Calibration: Curve-fitted table of dial settings for every 10 khz of AM band. This is VERY useful for quickly pouncing on DX that pops up then fades.

Op-Aid: Master log of 300+ stations heard over the last 2 years. It really helps having frequencies paired to: ID’s, formats, ZIP’s, area codes, major roads, distance, grey-line, etc.

Spotter radio: YAESU FT-757GX.

Jack Bryant

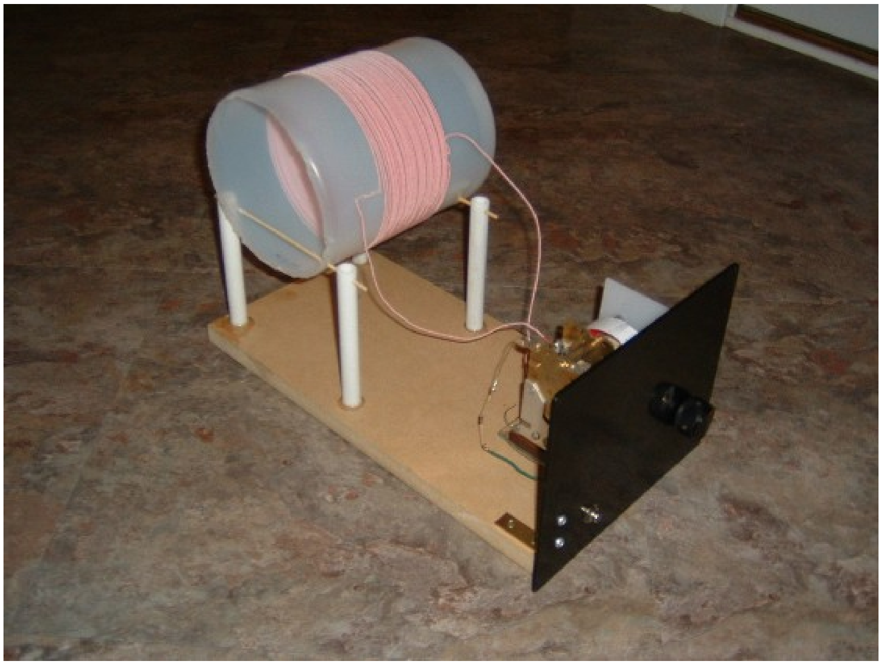

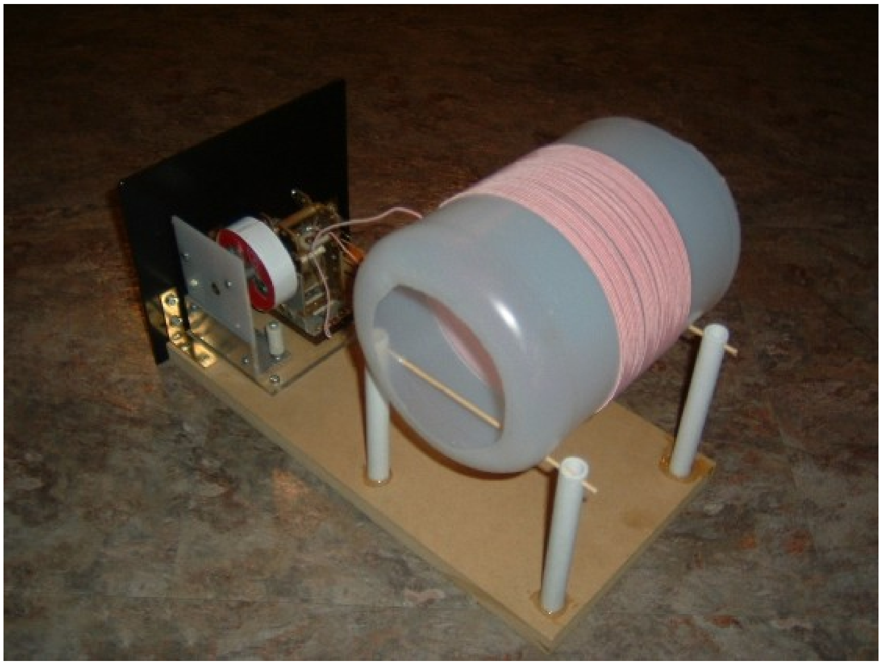

I wanted to try something new this contest. My old friend, Lem Morrison, has had a lot of success with air core coils, so I thought I would focus on a similar set for 2008. I had conversations or e-mail exchanges with several friends on a set recommendations. This included Lem, Gil Stacy, Mike Tuggle, Dave Schmarder, and Steve Bringhurst. The set would use 660/46 Litz on some kind of form. Steve's recommendation of HDPE vinegar jugs with a six inch diameter was the final choice of coil form. I used bamboo skewers through the ends of the jugs as part of the supports system. These skewers went into vertical posts of CPVC pipe; The pipes were epoxied into the fiberboard bases. The arrangement is was I call “goal post support”. I set the jugs up several inches from the fiberboard to help minimize losses.

The radio has five basic building blocks: the trap, antenna tuning, detector, audio matching, and headphones. The trap was the easiest part since I used a toroid unit I had built a few years ago. It uses and FT114A-61 toroid with a five turn link opposite the primary winding. The ceramic insulated variable cap is tuned with a 5:1 ratio velvet vernier from a WW2 BC-375 tuning unit.

The antenna module uses a variable with a built-in 2: 1 vernier and is wired in series with the antenna. The antenna connects to the frame, and the stator connects to the top of the parallel tank circuit. This tank circuit consists of a “goal post” mounted Litz jug-coil and a ceramic insulated variable cap with a vernier, also from a BC-375 tuning unit. These verniers have a 50:1 ratio and use a right angle drive approach. A drum made from a peanut butter jar top provides basis for the frequency readout.

In many other radios I used a dual section cap for the combination of antenna and tank circuit caps, but I wanted additional flexibility. The approached used here worked out great. The separate series antenna cap was especially useful during daytime DXing. It allowed me to use a greater value of antenna series cap for higher volume on weak daytime stations as well as reduced series capacitance when needed for selectivity. The base of the unit is fiber board with a from panel made from a black plastic clipboard.

Next is the detector module with the goal post mounted Litz jug-coil, an FO-215 diode, and a Fair Radio “holy grail” variable cap. This unit also uses the same type base and front panel as in the antenna unit. I selected to use neither taps nor a contra- wound approach on the two coils, just continuous wound coils. I did add complexity by adding an additional eight turn winding between the Litz rope for biasing a zero- bias MOSFET. This was switched in and out via a panel mounted double pole ceramic switch. I had trouble with the MOSFET and switching arrangement and finally bypassed it, using only the diode detector.

A panel mounted phono plug allows connection to the remote audio matching unit. A length of video cable joined the detector and audio units. I build a version of Mike Tuggle's double transformer Stanley match, and it worked great. Mine was wired for only two impedance options. I used the high input setting almost exclusively. I also tried a Select-To-Match and single transformer Stanley matching units for use with the in-the-ear units, but the double transformer Stanley Match seemed to work as well or better than those.

The headphones required more attention than any other part of the radio. The many hours of listening in previous contests with heavy sound powered elements had taken its toll on the nerves around my ears. In the previous contests the irritation became quite severe, such that even one of my eyeballs throbbed with pain! I had mentioned this one Rap N Tap,and a received quite a number of suggestions.

Garry Nichols and John Davidson provided me with an in-the-ear Shure unit to test. Steve Bringhurst provided a set of British phones with a wire behind the neck for partial support and a cloth band over the head. I also tried some Koss and Realistic in-the-ear phones. I used both the Shure and Koss phones in the beginning of the contest. They were comfortable and worked fine, but were not nearly as sensitive as my old phones. Steve's phones were sensitive but were not comfortable. I finally settled on a helicopter headset with salvaged SP elements. The headset fit completely over and around the ears. While these were not as sensitive as my old set of phones, they were a reasonable compromise.

I have used a Realistic DX-398 for several previous contests, and I used it for this one, too. A neat trick is to bring the antenna matching unit near the DX-398. Sometimes the DX-398 needed this extra boost the hear the signal the crystal set was receiving!

I like to use computer logging as I go along, so I used my normal home PC for that. A four foot table from Lowe's and a comfortable chair rounded out the listening post.

Continued on Page 2.

Prices just for you
UP TO 50% OFF select hotels*
Ad ✕
See deals
Expedia
*Terms apply

Select Menu

- Home
- About Us
- Contest Results
- 1AD Contests
- 1AD Entries
- Antenna Matching
- Antennas
- BHAM Radios

- Contest Results
    - 2004
    - 2005
    - 2006
    - 2007
    - 2008
    - 2009
    - 2010
    - 2006 Sprint
    - 2008 Sprint
    - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2007 Contest Entrants

Page 2

Jeff Welty

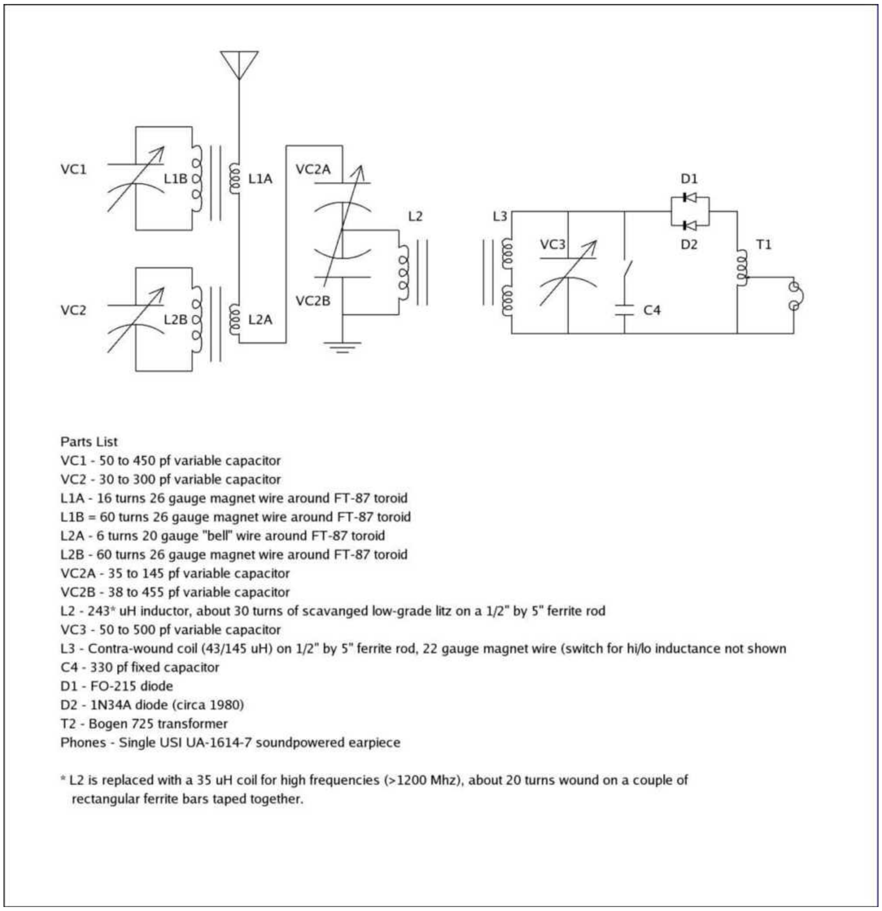

Parts List
VC1 - 50 to 450 pf variable capacitor
VC2 - 30 to 300 pf variable capacitor
L1A - 16 turns 26 gauge magnet wire around FT-87 toroid
L1B = 60 turns 26 gauge magnet wire around FT-87 toroid
L2A - 6 turns 20 gauge "bell" wire around FT-87 toroid
L2B - 60 turns 26 gauge magnet wire around FT-87 toroid
VC2A - 35 to 145 pf variable capacitor
VC2B - 38 to 455 pf variable capacitor
L2 - 243* uH inductor, about 30 turns of scavanged low-grade litz on a 1/2" by 5" ferrite rod
VC3 - 50 to 500 pf variable capacitor
L3 - Contra-wound coil (43/145 uH) on 1/2" by 5" ferrite rod, 22 gauge magnet wire (switch for hi/lo inductance not shown
C4 - 330 pf fixed capacitor
D1 - FO-215 diode
D2 - 1N34A diode (circa 1980)
T2 - Bogen 725 transformer
Phones - Single USI UA-1614-7 soundpowered earpiece

* L2 is replaced with a 35 uH coil for high frequencies (>1200 Mhz), about 20 turns wound on a couple of rectangular ferrite bars taped together.

Click here for the large schematic view

The variable capacitor on the antenna coil is from a radio in that I remember being in my sister's room in the 1970's, (which I dismantled for parts sometime at the end of the 70's, and made my very first crystal set with), and the detector coil is wound around a ferrite rod from the first radio I ever owned as a teen age boy, also back in the 1970's, and died about 4 years ago so I scavenged the ferrite rod from it.

The basic radio is mostly built from parts literally scavenged out of other radios over the years. The only "new" parts are the bogen 725 transformer, FO-215 diode (thanks Dave Schmarder) and sound-powered headphones.

Strong memories in this set for me, and hearing stations fading in and out really brings back memories of the days when I'd listen to my radio, hearing far away stations and caught up in the wonder of it.

Otto Danby

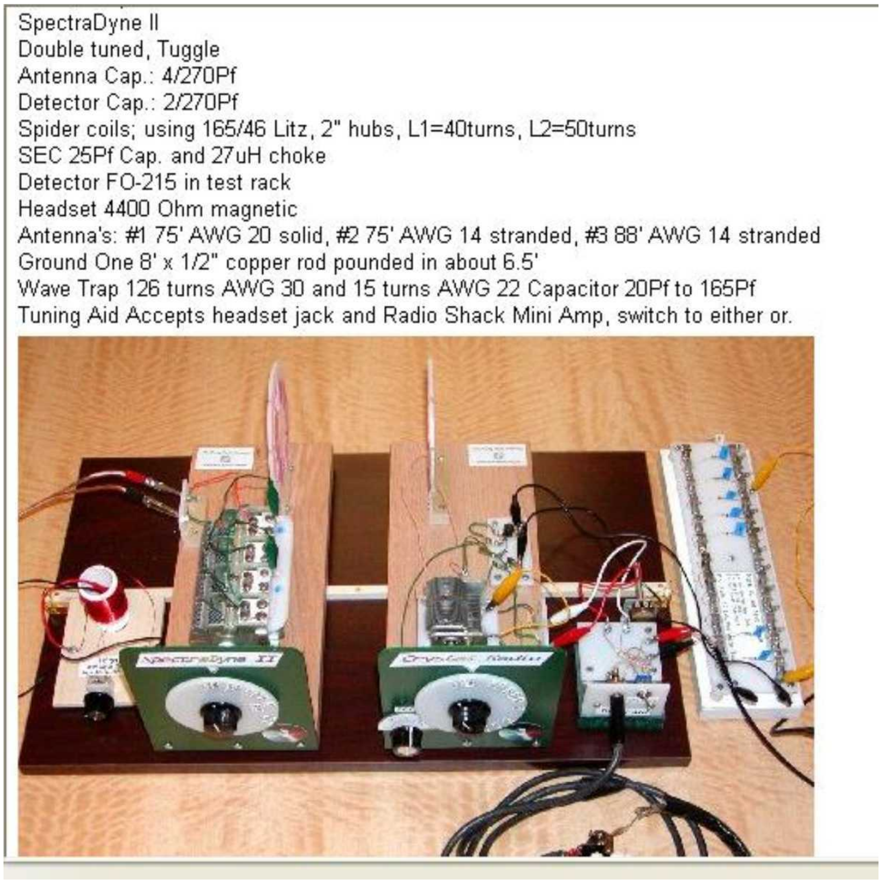

SpectraDyne II
Double tuned, Tuggle
Antenna Cap.: 4/270Pf
Detector Cap.: 2/270Pf
Spider coils; using 165/46 Litz, 2" hubs, L1=40turns, L2=50turns
SEC 25Pf Cap. and 27uH choke
Detector FO-215 in test rack
Headset 4400 Ohm magnetic
Antenna's: #1 75' AWG 20 solid, #2 75' AWG 14 stranded, #3 88' AWG 14 stranded
Ground One 8' x 1/2" copper rod pounded in about 6.5'
Wave Trap 126 turns AWG 30 and 15 turns AWG 22 Capacitor 20Pf to 165Pf
Tuning Aid Accepts headset jack and Radio Shack Mini Amp, switch to either or.

Click for large view.

Peter Kertulla

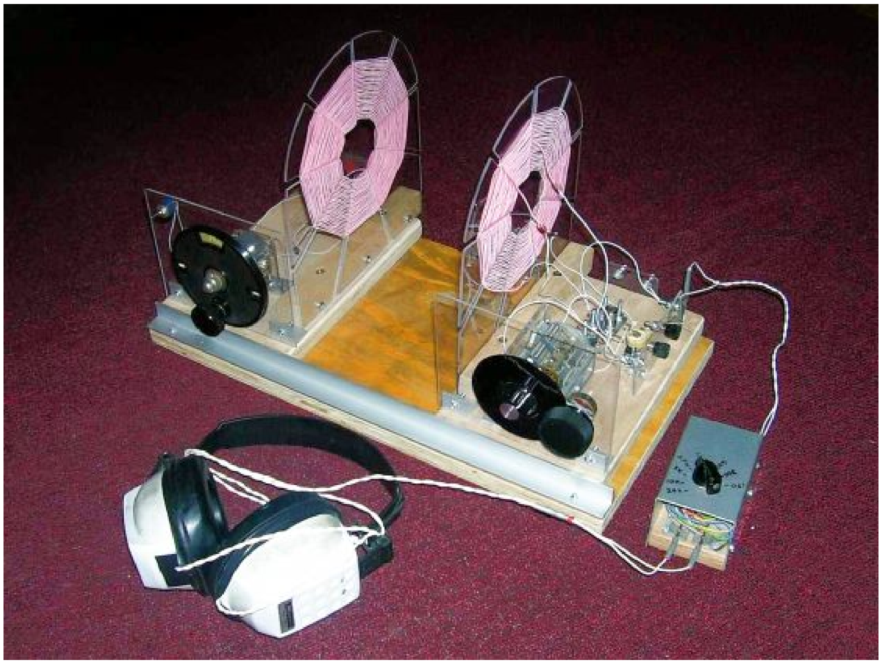

2007 CRYSTAL SET CONTEST SET

The rig this year is all new. It is double tuned with 660 litz spiderweb coils. Both capacitors use reduction drives for easier tuning. The detector coil was changed at the last minute to a contra-wound configuration…the jury is still out as to the impact this made. I think I may use separate plug in coils for HI and LO band next year. The antenna is the same as last year… 100 foot end fed up about 25 feet. Headphones are sound powered elements in a reused headset with a bogen transformer that has switched impedances.

Robert Golding

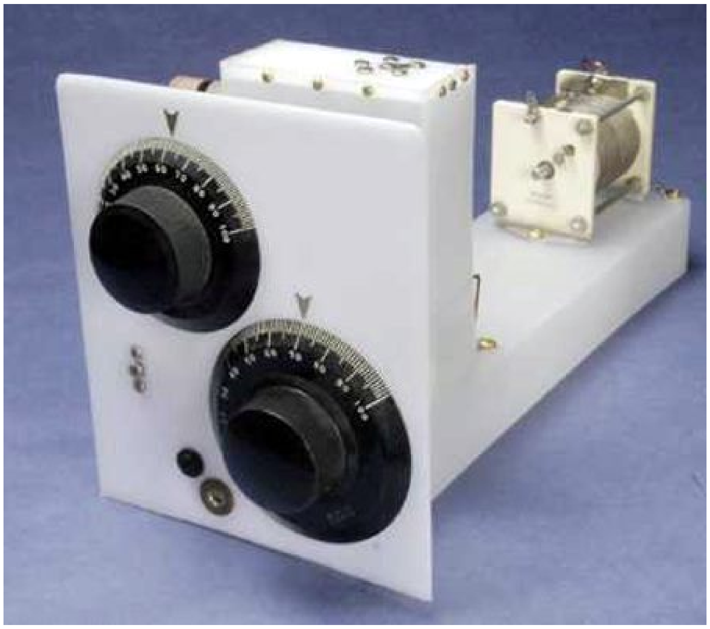

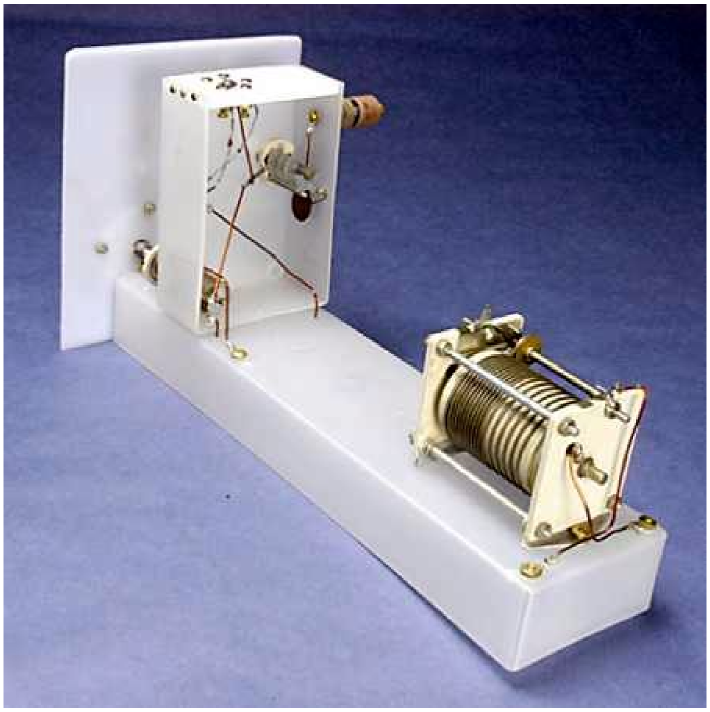

My short wave set is very simple.

It consists of a tank circuit composed of a small Johnson roller inductor with a small 100pf silver plated variable cap. I connected another small variable as an antenna trimmer and included a small trimmer to ground. It has a diode selector arrangement including a single generic Radio Shack germanium (black band) and choices of one two or three paired Schottky IN5711s. I found it necessary to add a bias attachment for the IN5711s.

I had an outboard Benny setup that plugs in between my matching transformer and the set. Therefore I added a tiny watch battery to the wiper and one end of the rheostat. The bias really perks up the IN5711 diodes.

My headphones are Western Electric sound powered matched with a UTC Ouncer transformer.

The antenna trimmer proved to be ineffective. So, I tried Steve B's cap string which helps somewhat.

Naturally I wasn't quite ready when the contest started. So I got through the first couple of nights with the diodes still on clip leads.

I used a couple of small door knob caps for the antenna cap
string which is overkill, but I sort of like the look of it.

Steve Hewlett

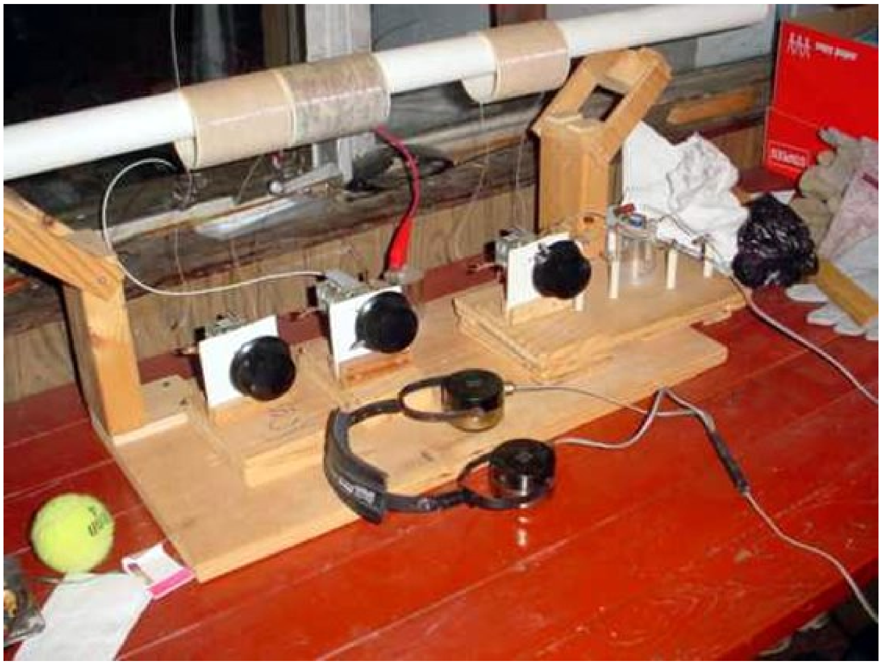

2007 Crystal Radio Contest Radio Set Description

My 2007 contest set is the same as my 2006 contest set. The set has three coils which slide on a polystyrene tube. From left to right they are wave trap, antenna coil, detector coil. All three coils are space wound (one wire width spacing between turns) copper wire; #24 ga. for wave trap and detector coils and # 23ga. for the antenna coil. All three coil forms are 3" diameter polystyrene tube. The inductance of each coil is approx. 251 uH.

A single Jackson 365 pF VC is used on the wave trap coil and a single Jackson 410 pF VC is used on the detector coil. A dual Jackson 410 pF VC is used on the antenna coil in a Tuggle arrangement. The diode is a FO-215. A Calrad transformer and a "benny" is used in the detector circuit to match the SP headset - USI 1614 - I am using. The antenna is a 3-wire flat top aprroximately 75 feet long and 13-18 feet high. Ground is a series of three 8 foot copper plated steel rods driven all the way in the ground four feet apart.

Wayne Thelen

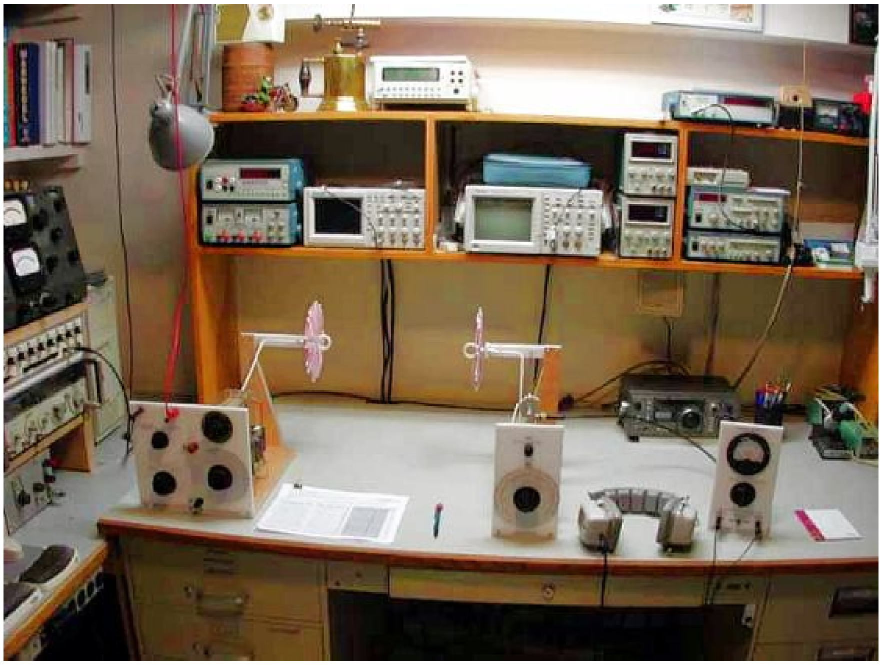

Wayne's Newest Set

- Contest Results
    - 2004
    - 2005
    - 2006
    - 2007
    - 2008
    - 2009
    - 2010
    - 2006 Sprint
    - 2008 Sprint
    - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2009 Contest Entrants

Page 3

Rodney Look, Open Class Entry

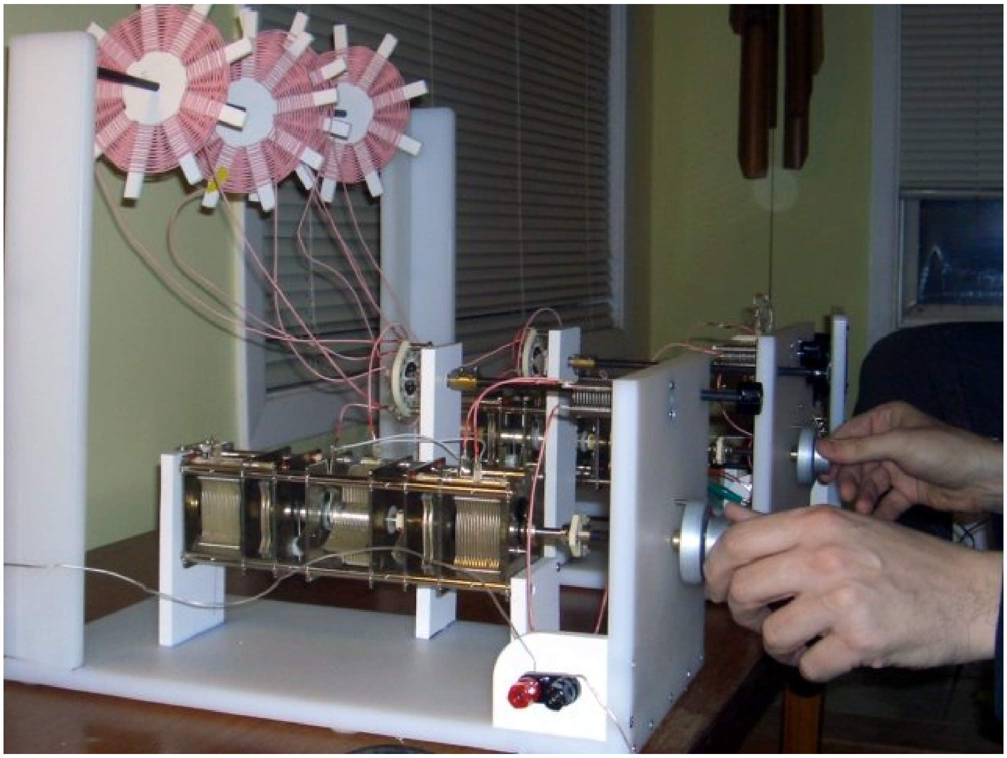

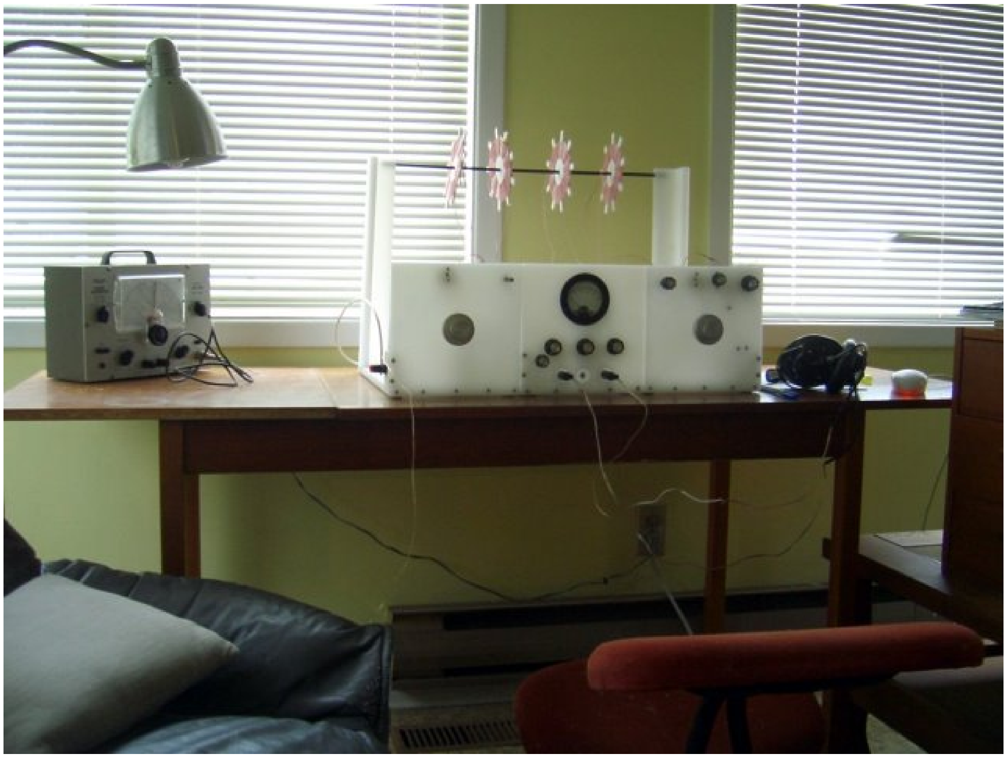

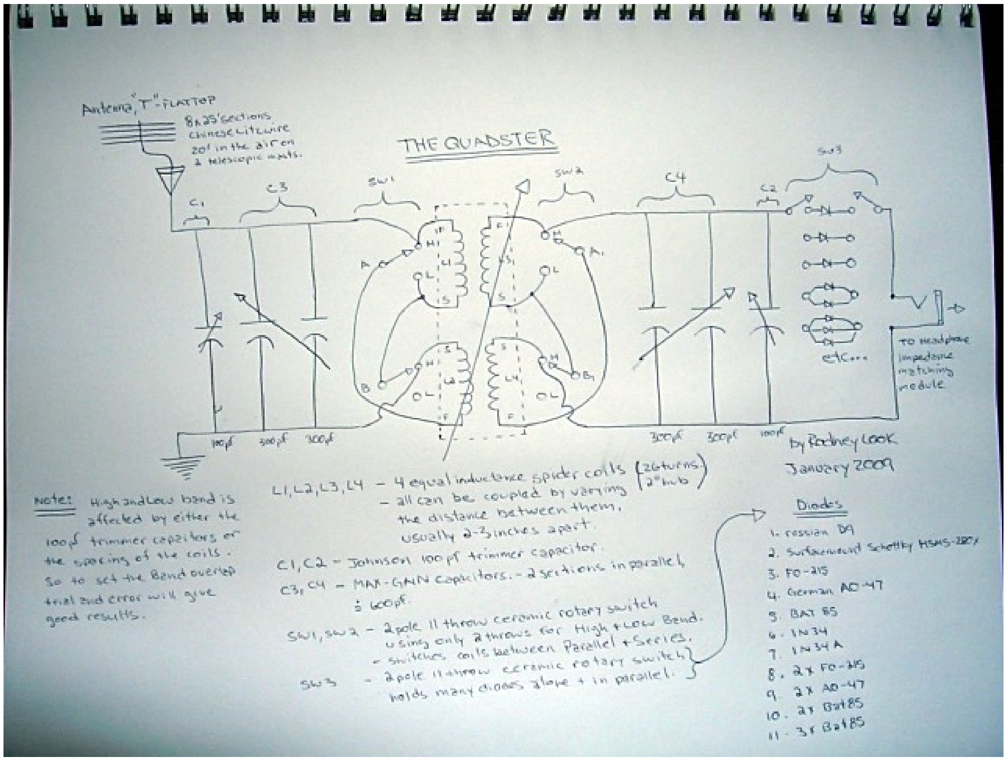
Antenna, "T"-FLATTOP
8 x 25' sections
chinese Litewire
20' in the air on
2 telescopic masts.
THE QUADSTER
C1
C3
SW1
SW2
C4
C2
SW3
A
B
A1
B1
H
L
S
F
L1
L2
L3
L4
etc...
TO Headphone
impedance
matching
module
100pf
300pf
300pf
300pf
300pf
100pf

by Rodney Cook
January 2009

L1, L2, L3, L4 – 4 equal inductance spider coils (26 turns, 2" hub)
– all can be coupled by varying the distance between them, usually 2-3 inches apart.

C1, C2 – Johnson 100pf trimmer capacitor.

C3, C4 – MAX-GAIN capacitors. – 2 sections in parallel ≐ 600pf.

SW1, SW2 – 2 pole 11 throw ceramic rotary switch using only 2 throws for High + Low Band.
– switches coils between Parallel + Series.

SW3 – 2 pole 11 throw ceramic rotary switch holds many diodes alone + in parallel.

Note: High and Low band is affected by either the 100pf trimmer capacitors or the spacing of the coils. So to set the Band overlap trial and error will give good results.

Diodes
1. russian D9
2. Surface mount Schottky HSMS-287x
3. FO-215
4. German AO-47
5. BAT 85
6. 1N34
7. 1N34A
8. 2x FO-215
9. 2x AO-47
10. 2x Bat85
11. 3x Bat85

Here is a link to theThe Pictures Slideshow for more pictures of the set.

Here is a link that shows the location of the stations copied.

Glen Yarbro, Open Class Entry

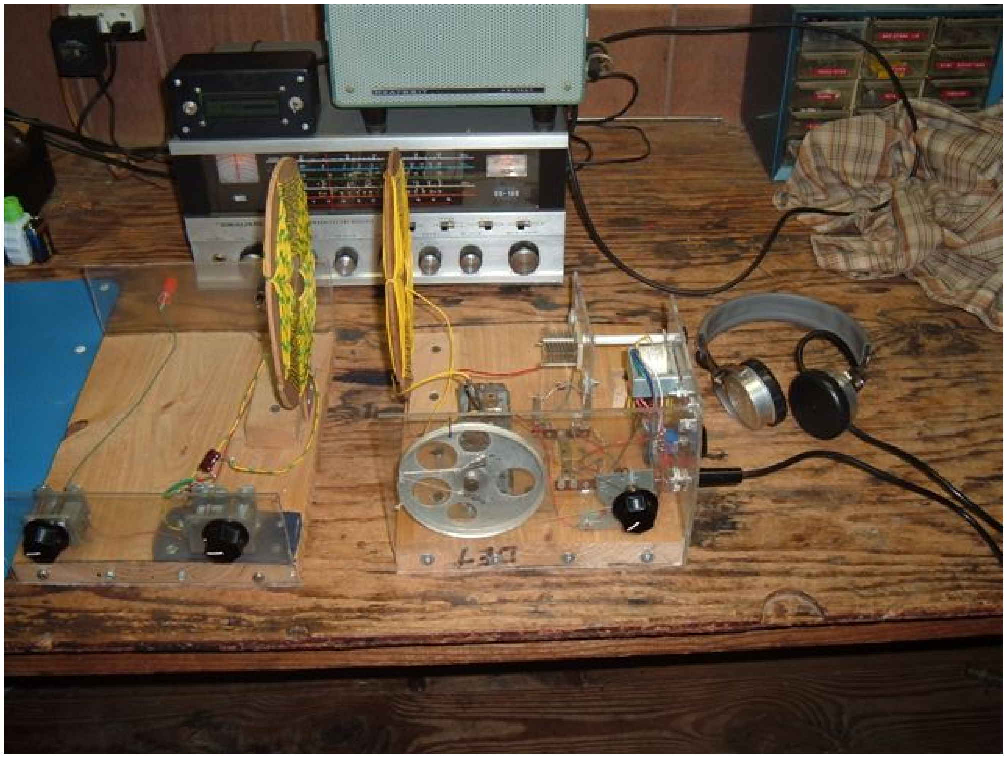

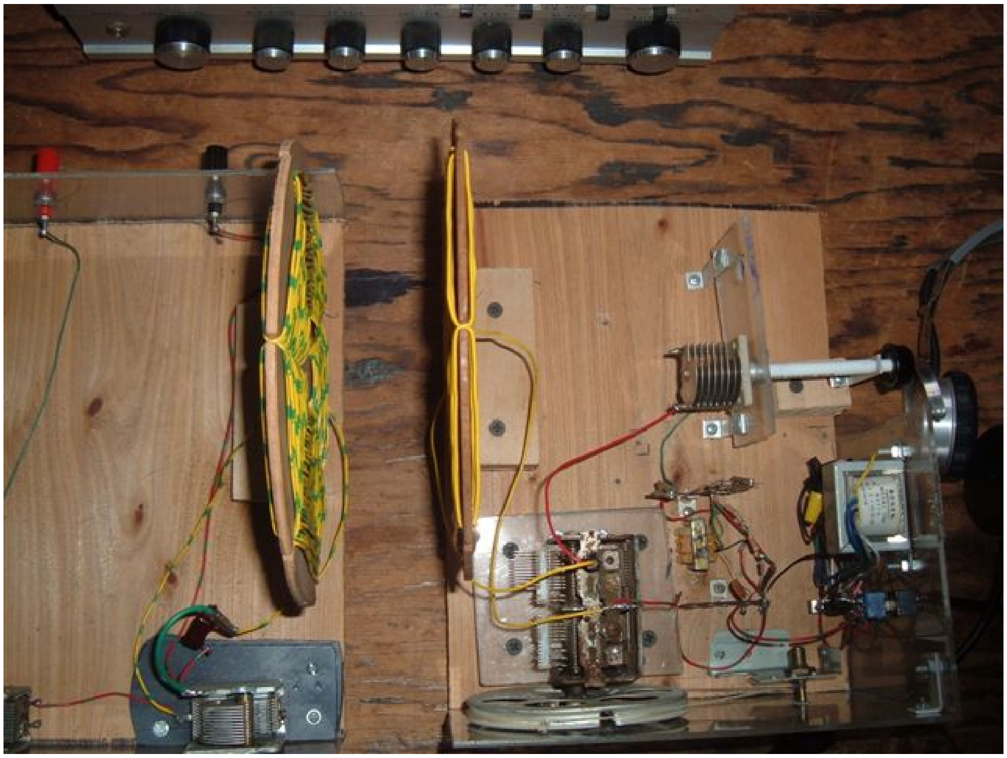

Radio : Dave Schmarder's #44
Antenna: 80m trap dipole, 30 feet high
Headphones: 4 k Navy phones
Coils were wound with 24 gauge Bell Wire

O. T. Anderson, Open Class Entry

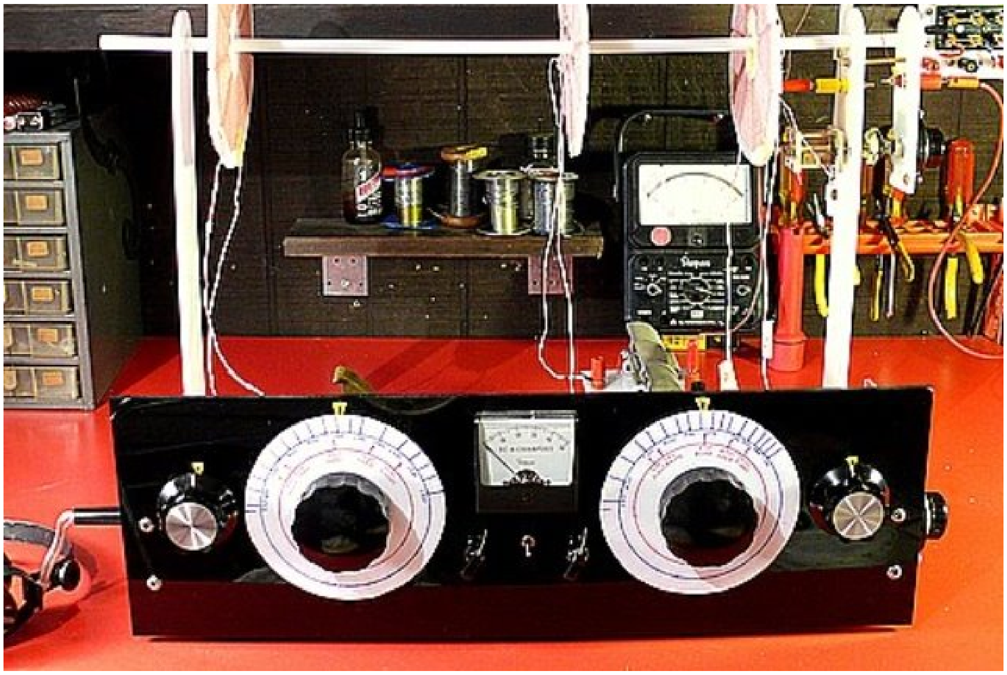

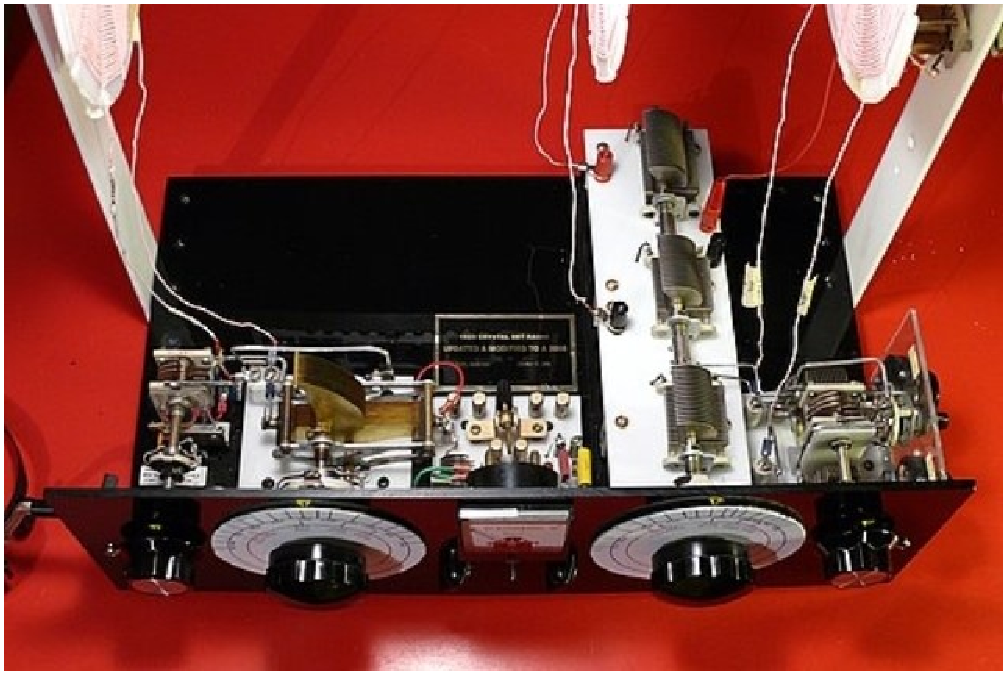

This crystal set has been modified several times, the dials, Var Caps,and wave traps. Being a ham operator I really liked calibrated dials. Some of the local stations are in red.

The right hand 3 tuning caps are the Hammarlund type. The three caps which are ganged together is a "Tuggle" front end. The detector cap is an old ancient job from a very early radio. I modified it using HDPE to replace old brittle bakelite. I don't know how good it is, but it looks and feels great. It has good band spread down at 1200-1700 kc. The three coils and output Xformer I bought from that good man, "Dave".

The two right and left small caps are a band spread for about one station. It feels good to "rock" these a bit on a weak station. The "S" meter is a real help. A 3 position switch gives a off, normal and high position. I would not built a new set without one. Since this is a "crystal set" I had to put in new (old type) "Cat Whisker". Can you see it? It also has a place for 3 diodes.

My biggest trouble here in Broken Arrow, OK is 1170 kc. That's KFAQ, old KVOO, 50,000 watts. Big 50,000. Has three towers, see them from our house. I have 2 wave traps, one 1170 and the other is 740, KRMG. The switch and stationary caps are at bottom right. Works OK, but 1170 is still in the back ground.

Built up an external trap. It's still there. I accidentally found the trouble moving the set around one day, cleaning up. I nulled out 1170 as I had the set turned around so-so. Apparently the detector coil is a good antenna just like an antenna in a AC-DC radio. So what should I do? Relocate my operating desk, or better yet build a new set patterned after THE LYONODYNE, (which has wave traps on the detector coil).

My wife, Doris, K5BNQ, wasn't crazy about me being back in the shop. She said, when it gets dark, I should come in, build a fire in the fire place, read the newspaper, and have a big cold Pepsi or Mt. Dew. Then maybe she would give me some real good home made chocolate cookies!!! And some ice cream!!!

Who could beat a deal like that?

Mark Hampton, Open Class Entry

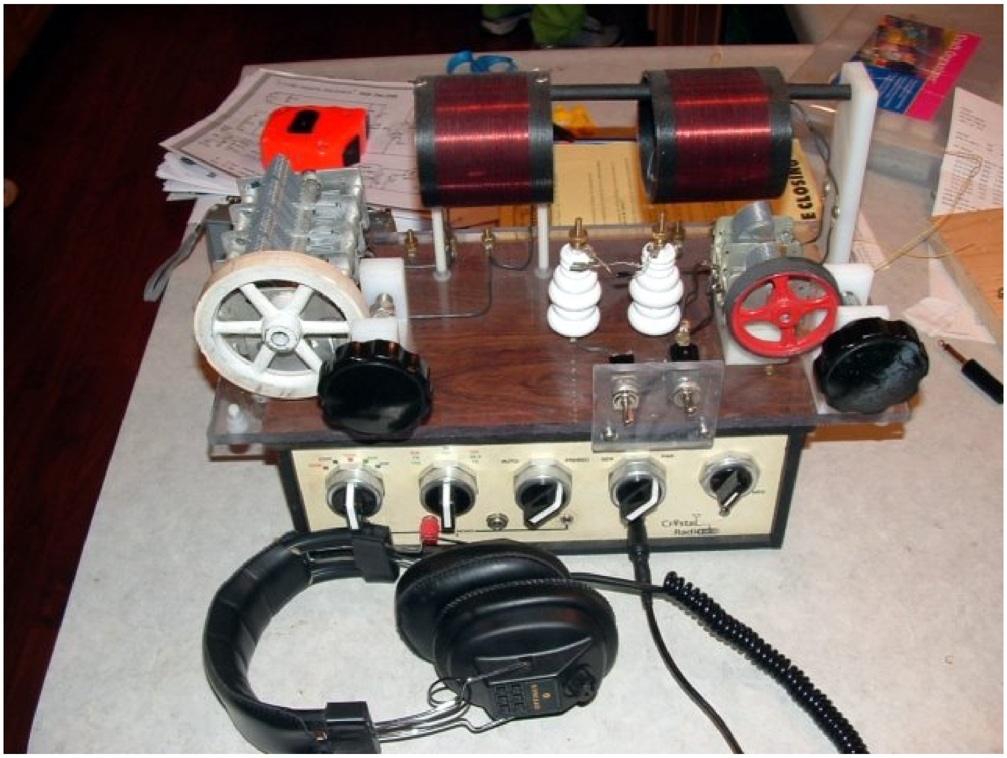

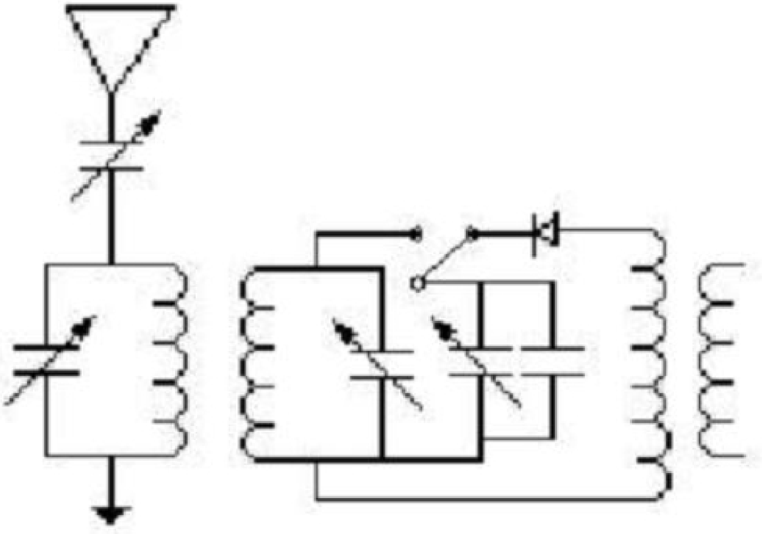

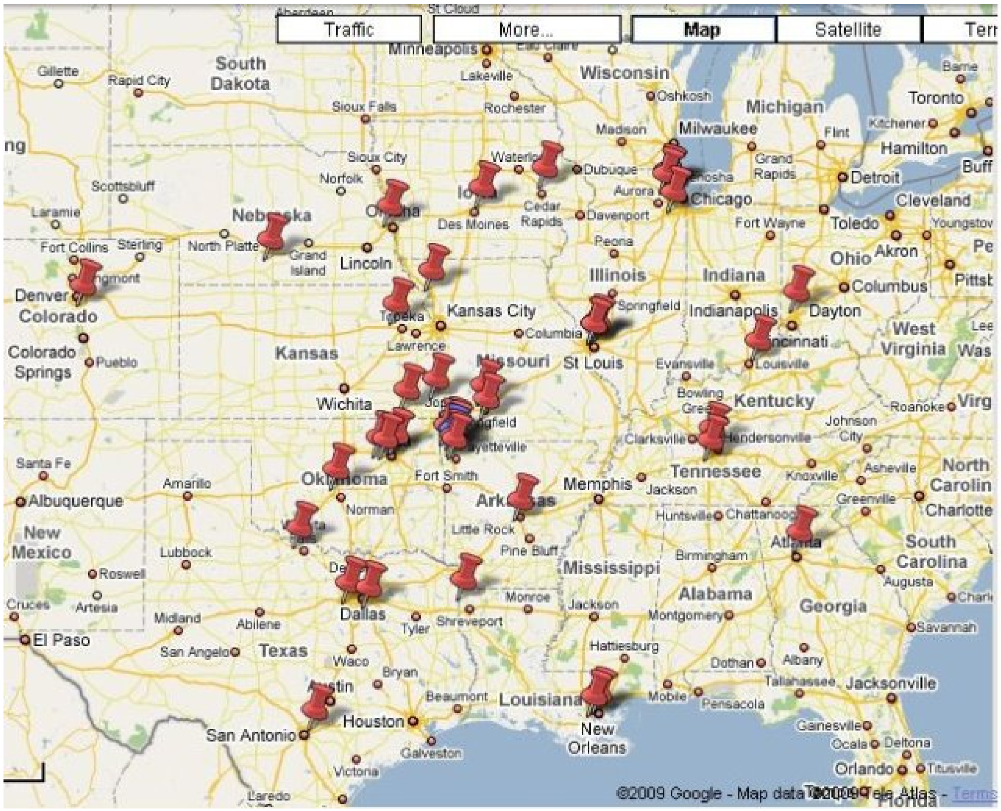
Traffic
More...
Map
Satellite
South Dakota
Wisconsin
Michigan
Nebraska
Colorado
Kansas
Missouri
Illinois
Indiana
Ohio
Kentucky
Tennessee
Oklahoma
Arkansas
Mississippi
Alabama
Georgia
Texas
Louisiana
New Mexico
Minneapolis
Chicago
Detroit
Toronto
Kansas City
St Louis
Wichita
Memphis
Birmingham
Atlanta
Dallas
Houston
San Antonio
New Orleans
Denver
Albuquerque
El Paso
©2009 Google - Map data ©2009 Tele Atlas - Terms

My set is a simple double tuned with a Tuggle front end and home brew friction reduction drives. A high/low switch spreads out the band a little further. The "high voltage insulators" are wood toy wheels stacked up and painted white. I thought it would add to the industrial look of the set.

My antenna is a 300' longwire, 30' in the air and oriented NS. The ground is tied to the cattle fencing (welded wire on metal post every 10') that runs all over this part of the county.

The unit is sitting atop my universal match box. The headset is SP units mounted in old Olympus stereo headphones.

The barely visible blue pin is my location on the map.

Garry Nichols, Loop Class Entry

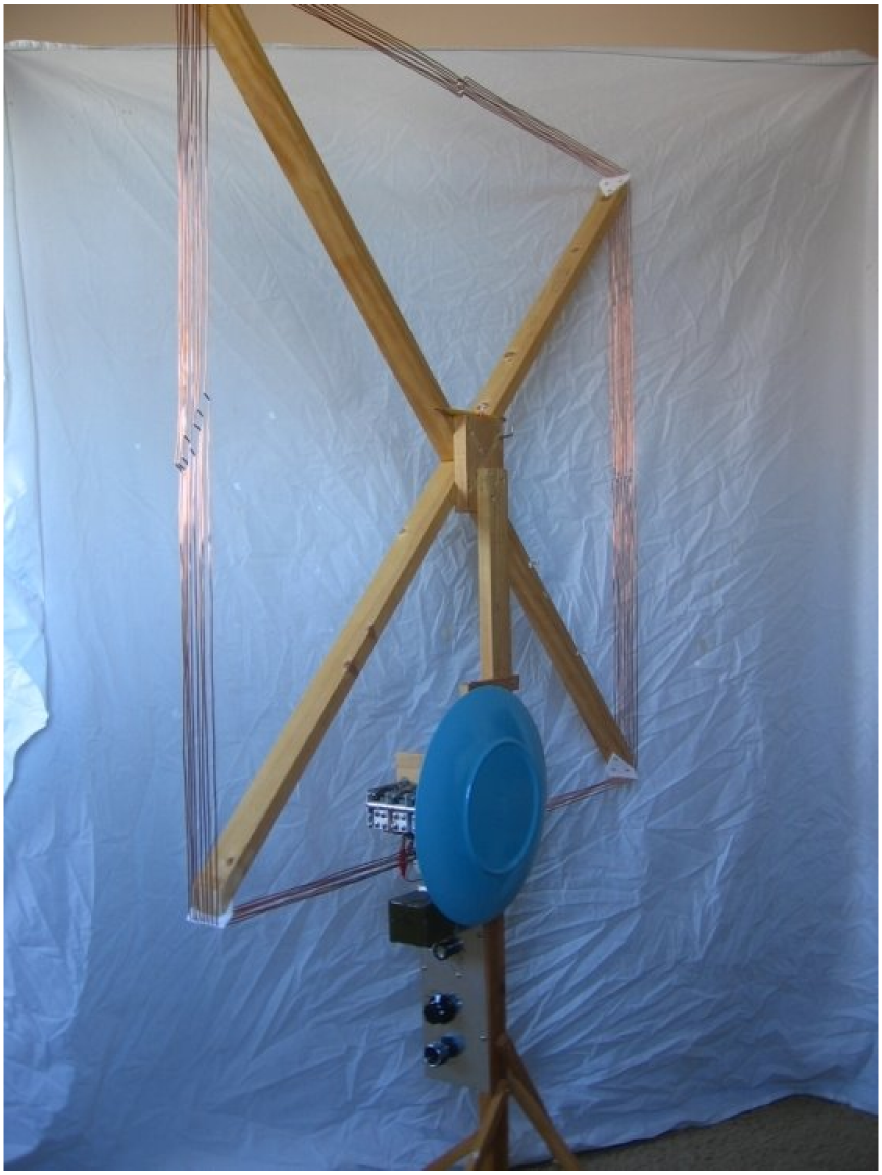

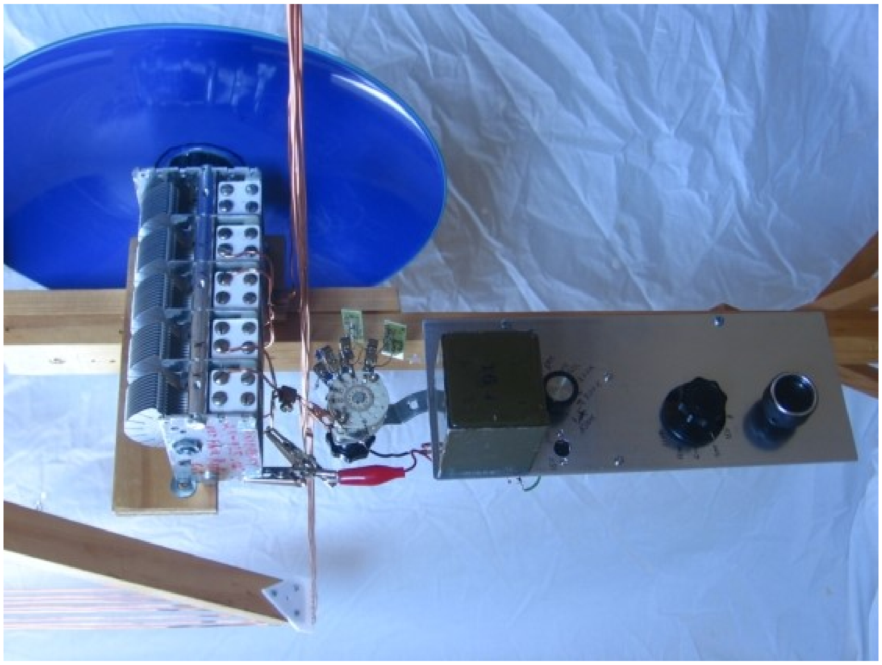

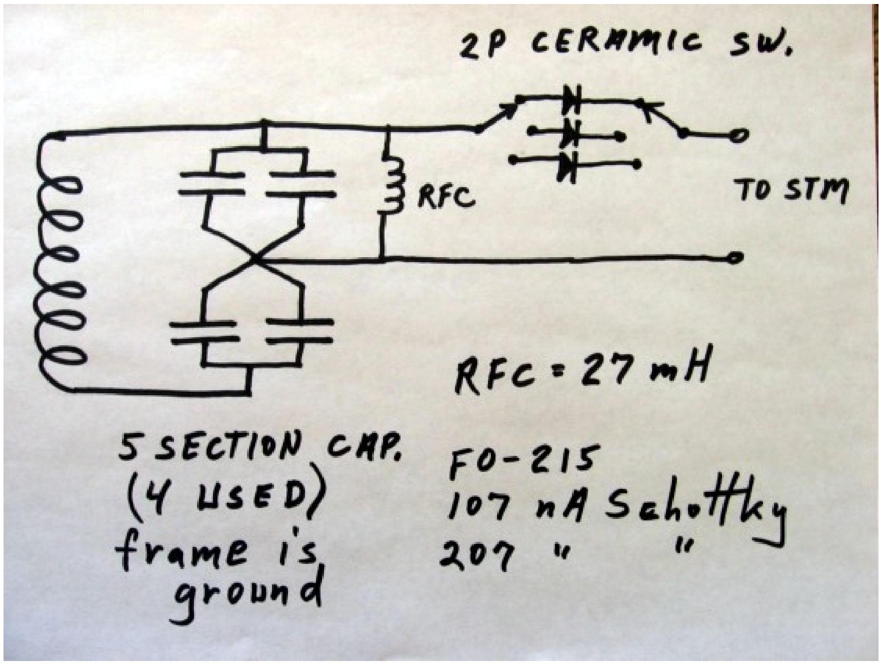

This is the same set used in the 2008 contest, except for a few minor changes:

-I took off the Select-To-Match this year and ran the loop with only the T3/AM-20 transformer (100:100k) and a 107 nA Schottky diode.

-Big can mic elements in series had a 2 uf electrolytic between them and the transformer.

This setup (with short leads) got rid of my FM interference completely, and I did not hear the intermittent hum that I was picking up last year from various orientations of the loop.

Kevin Norton, Open Class and Short Wave Class Entries

Sets: On B'cast I used the K-3 xtalset. It is a 75 ft longwire through a wiperless ceramic 4 gang, then to a 4 ceramic gang wiperless tank (660/46 ferrite assisted air coil). Each var cap is wired so that a 500 gang is bussed to the adjacent 500 gang. So two 500 gangs equal one leg of my cap. The gnd is to the FRAME of the tank var cap (NOT the "cold" stators).

The det tank uses the same coil type and var cap type / wiring layout. There is no det tank connection to Earth or ant tank.

All caps are 6:1 ball driven and have oversized insulated wooden knobs.

A single FO 215 was used fully at the hot end of the det tank. Although single Schottkys HAVE been used in past contests, results were worse this time (for whatever reasons).

The match is 2x UTC A-27s and the phones are David Clark SP types wired in series @1200ohms.

Future K-3 improvements could include raising the match Z by wiring in additional A-27's, changing the benny pot to a higher value and trying various benny caps. Also a set of high band coils should wake things up quite a bit on the high end.

An outboard applied RF bias is available by a weak signal digital B'cast transmitter and coupled to the set by an unterminated litz air coil. This transmitter only covers 600 Kc and up, so an analog Heath signal generator fills in from 530 to 600. Using just a tiny level of (variable) RF greatly narrows the bandwidth. Strong locals are obliterated and signals otherwise NEVER copied come into good audibility. There is an almost "regenny" sound when doing this.

I can find stations by using the analog or digital scales, bias the set into the "supercharged" mode,and then peak all controls. Then I simply shut down the RF, move the bias coil away and wait for a signal to come up into true xtalset passive copy.

The HF set was a simple affair. A wiperless ceramic var cap on 11uH homemade > air dux. There is a two turn ant coil . The tank has no gnd . The det is an unknown diode directly into series xtal earphones.

Overall a good a time, despite a slow start.

Curtis Gamble, Open Class Entry

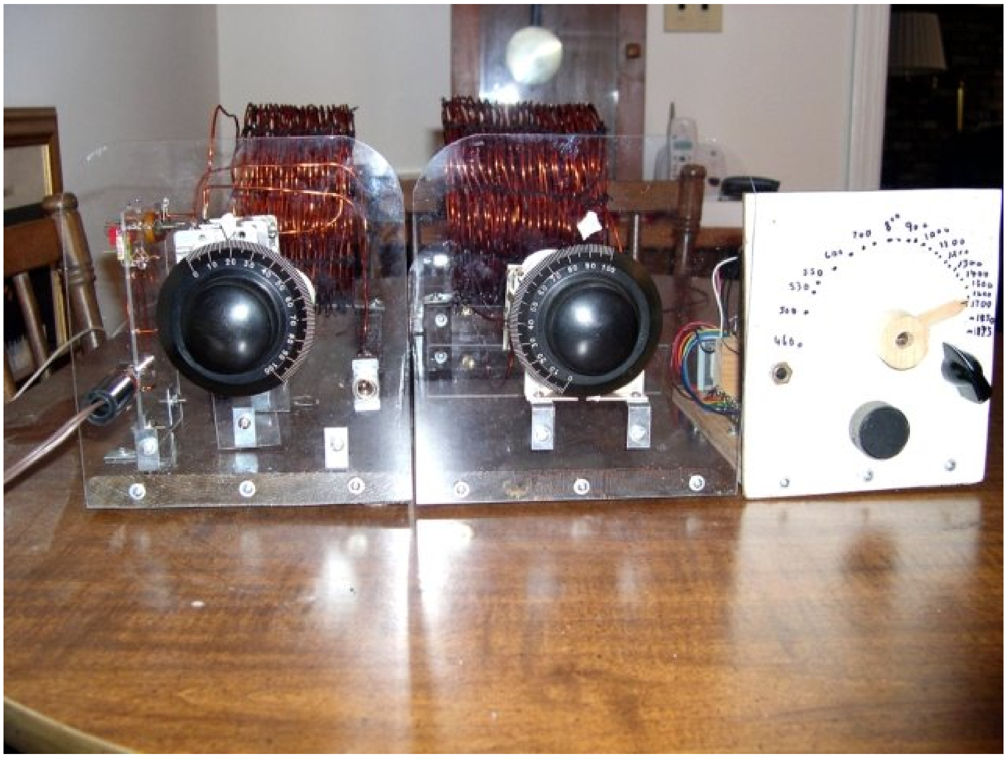

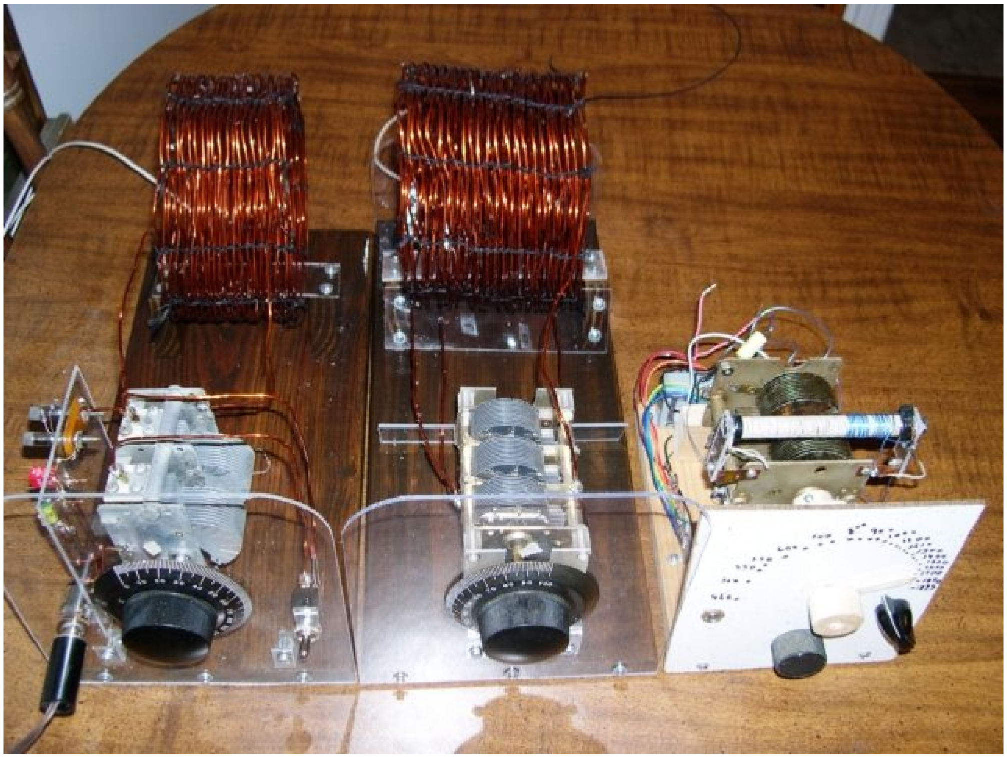

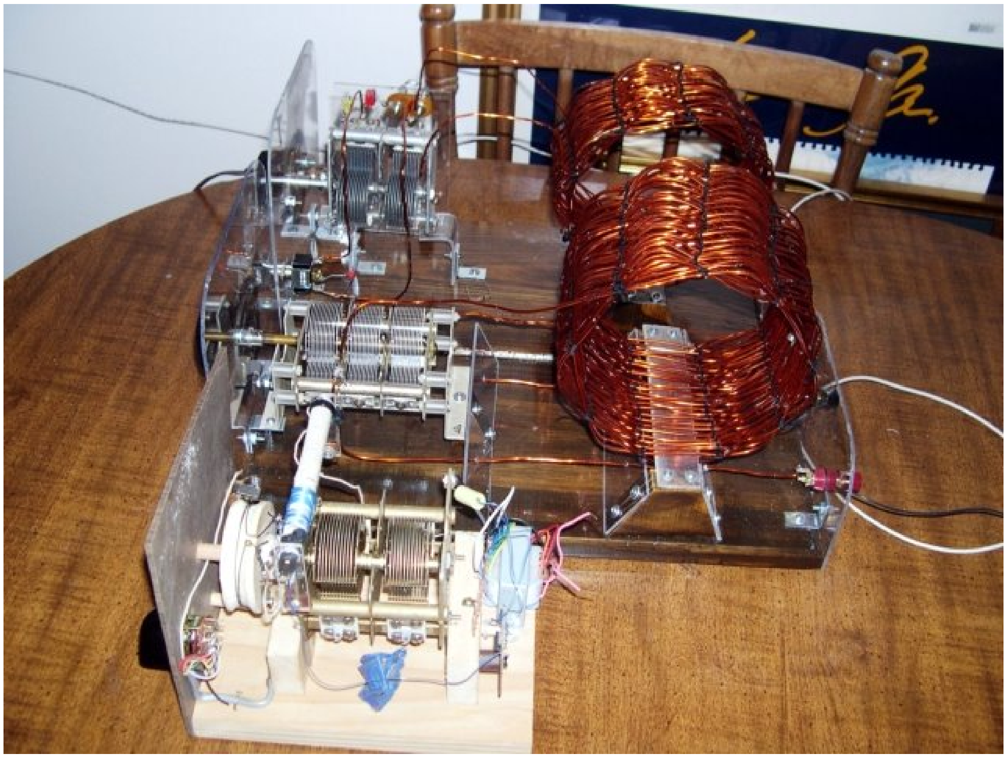

This is my first passive receiver that was not a bread board set. It includes a Tuggle antenna turning unit with 12 gauge enamel wire rook coil, center of the picture, a 1N34 detector with 14 gauge enamel wire rook coil, and a Triad SP 21 300 Hz to 100Kz response output to sound powered head phones.

The unit to the right is an IC detector with an ALD110900A and BOGEN T725 trans- former to the same sound power headphones. The 160 meter flat top dipole is up about 45 feet, plus I use a 50 foot long wire, and a house copper water pipe ground.

I have two locations I can work, my old home place in Tuscaloosa, Alabama in the same shack I've had for over 50 years, and the home I live in now in Hoover, AL. I worked this contest in Hoover, Alabama. This location has very low noise, and after sunset the local blow torches are not a issue. I normally don't use a trap.

Charles Pullen, Open Class and SW Class Entries

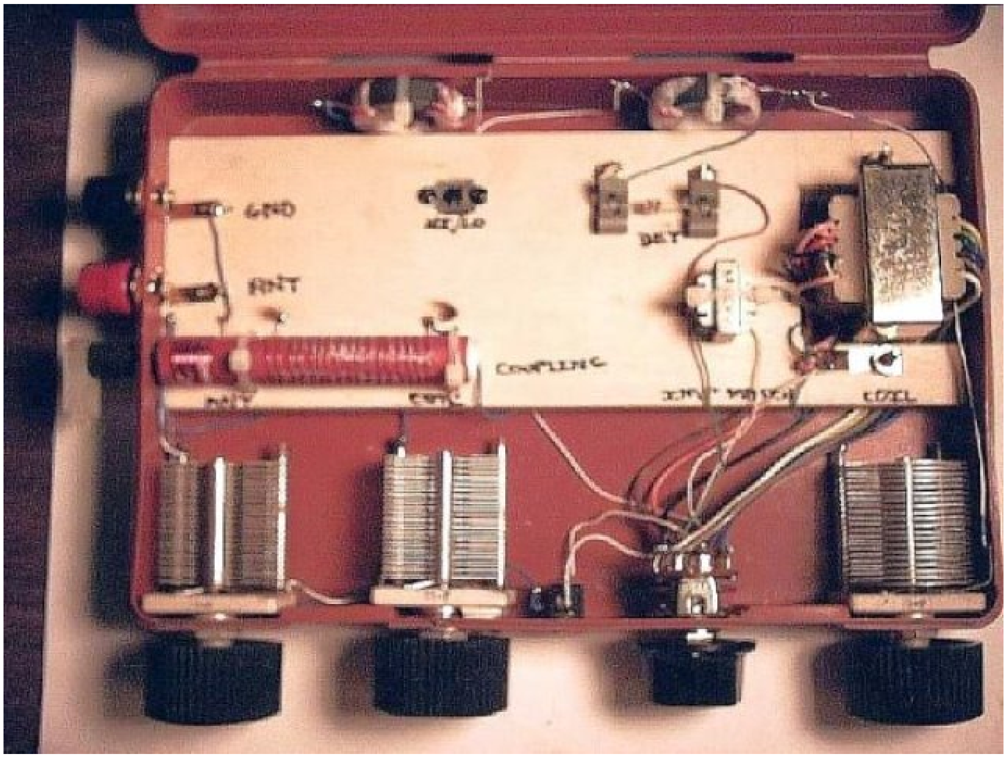
GND
ANT
DET
COUPLING
ANT
COIL
COIL

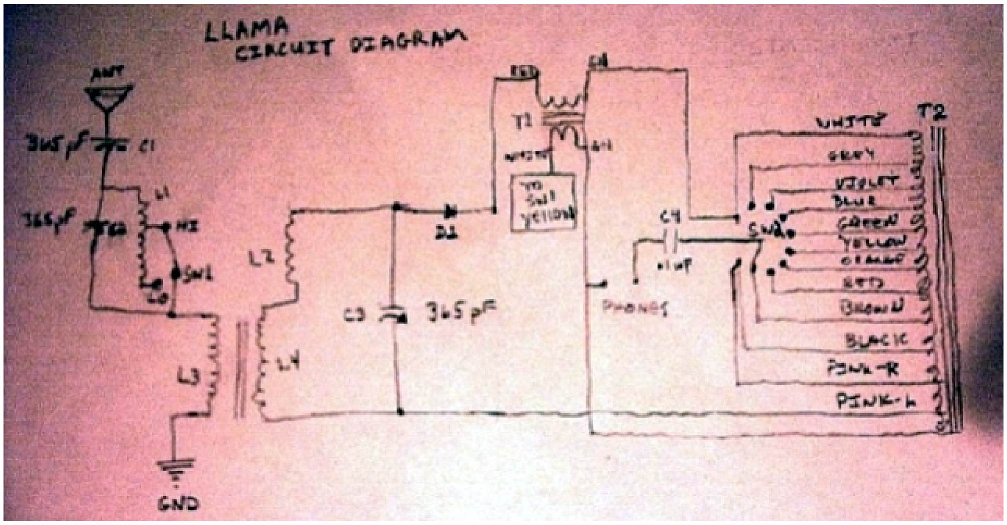
LLAMA
CIRCUIT DIAGRAM
ANT
365pF
C1
L1
365pF
C2
HI
SW1
LO
L2
L3
L4
GND
C3
365pF
D1
RED
T1
GN
WHITE
GN
TO SW1 YELLOW
C4
.1uF
PHONES
SW2
T2
WHITE
GREY
VIOLET
BLUE
GREEN
YELLOW
ORANGE
RED
BROWN
BLACK
PINK-R
PINK-L

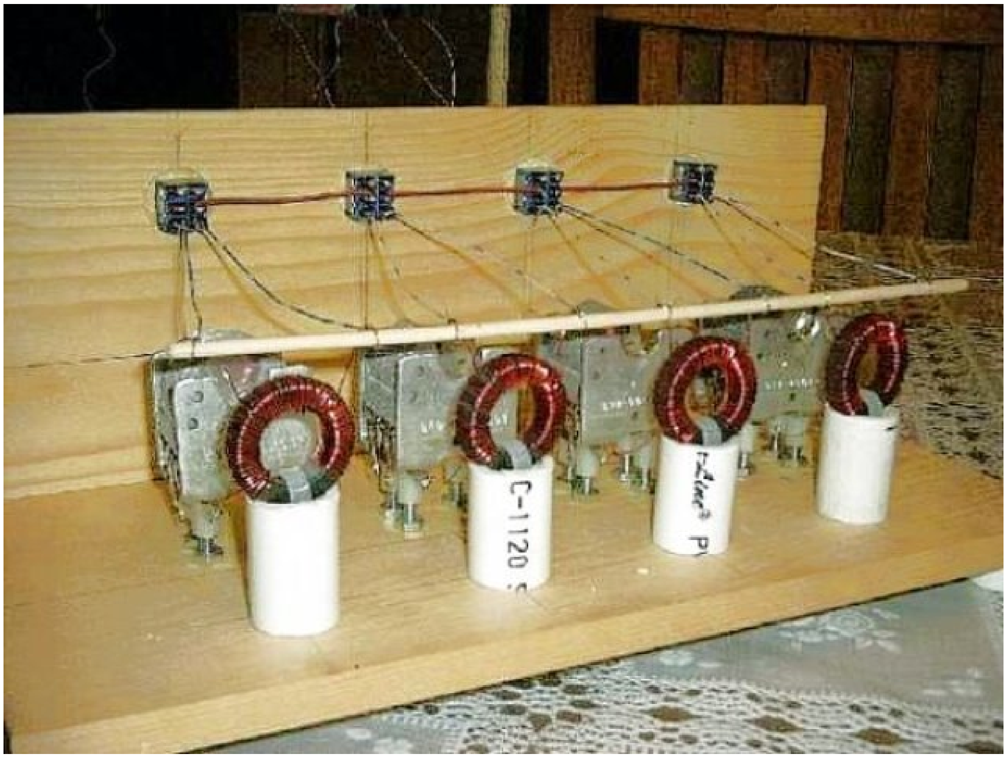

This is the same radio I used in the ’07 contest, with no additional modifications. I call it the Llama, because of the case it’s built in.

The antenna this year was a 40 foot inverted L, about 25 feet above ground. I added two additional ground rods to my ground system this year to compensate for the smaller antenna. I had to make extensive use of the wave traps this time around, the locals at 1080, 820, 770 were totally dominating the middle of the band.

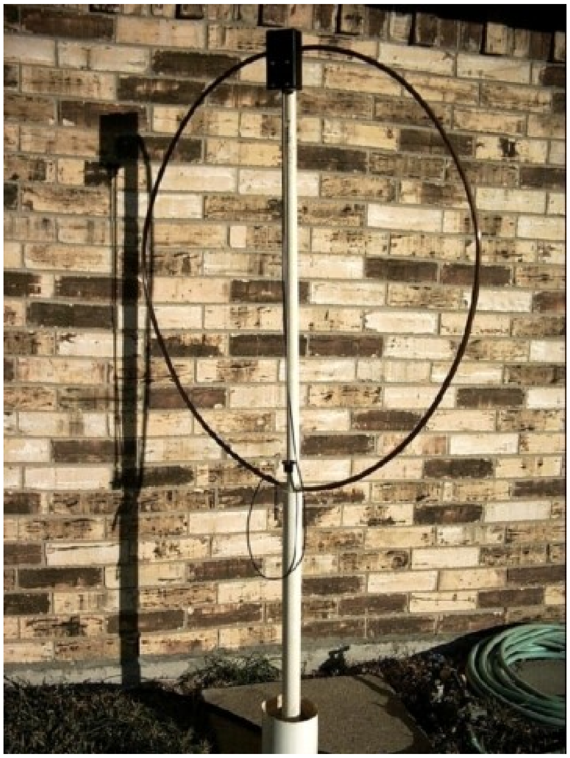

This is a very simple SW loop antenna radio I converted from a QRP 40 meter small transmitting loop I built about 5 years ago, and has been beaten up quite a bit over the years. It consists of a four foot diameter, ½ inch copper loop that is paralleled with a 365 pf air variable capacitor and a 1n34a diode detector.

I hooked it to my copy of Steve Bringhurst’s ultimatch unit for impedance matching and listened in with a pair of philmore xtal earplugs in connected with a radio shack mono y-adaptor.

- Contest Results
  - 2004
  - 2005
  - 2006
  - 2007
  - 2008
  - 2009
  - 2010
  - 2006 Sprint
  - 2008 Sprint
  - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2009 Contest Entrants

Page 1

Ralf Siemieniec, Open Class, SW Class, and Below BCB Class Entries

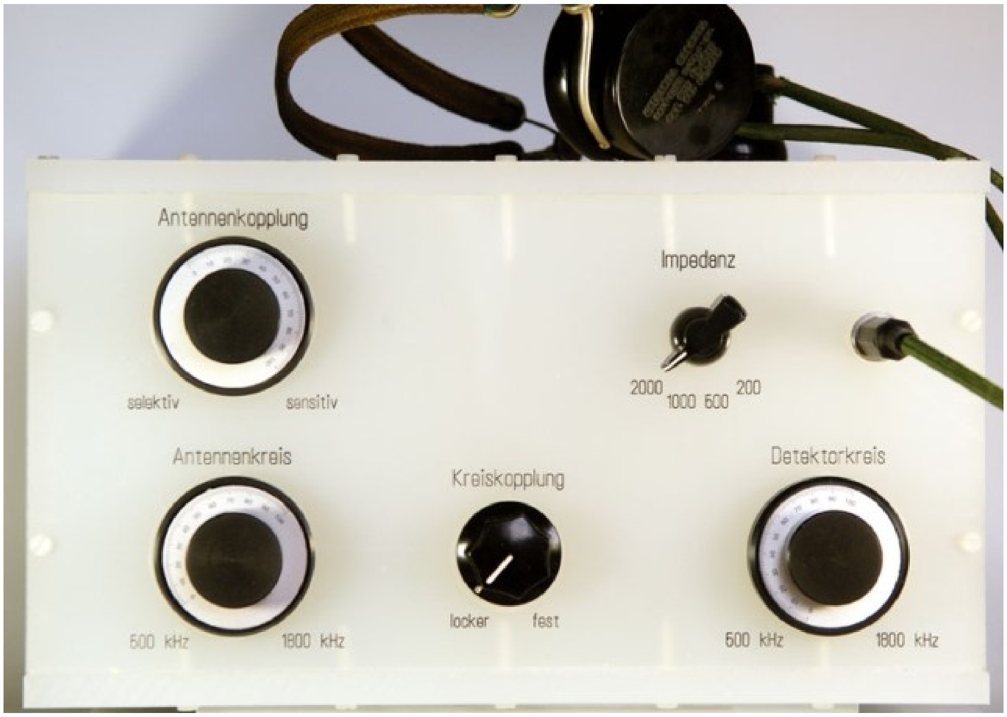
Antennenkopplung
selektiv
sensitiv
Impedanz
2000
1000
600
200
Antennenkreis
500 kHz
1800 kHz
Kreiskopplung
locker
fest
Detektorkreis
500 kHz
1800 kHz

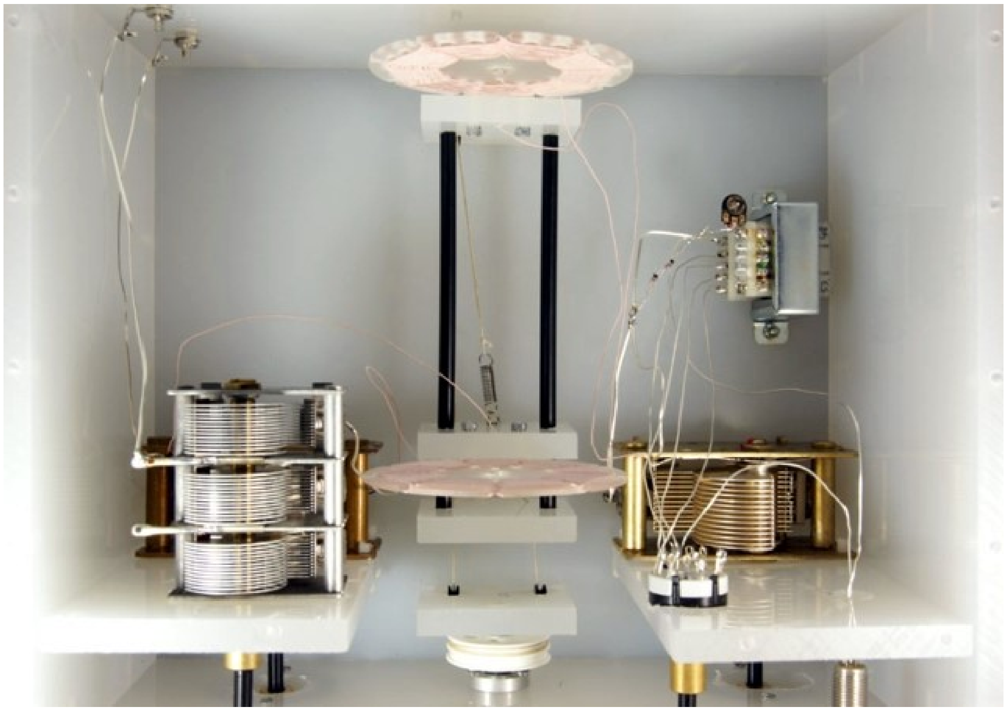

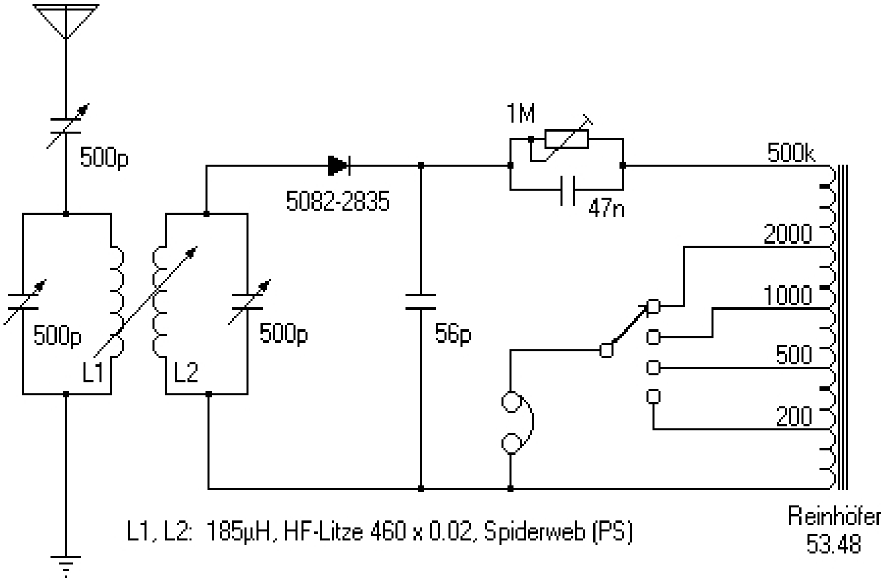
1M
500p
5082-2835
500k
47n
2000
1000
500p
L1
L2
500p
56p
500
200
L1, L2: 185µH, HF-Litze 460 x 0.02, Spiderweb (PS)
Reinhöfer
53.48

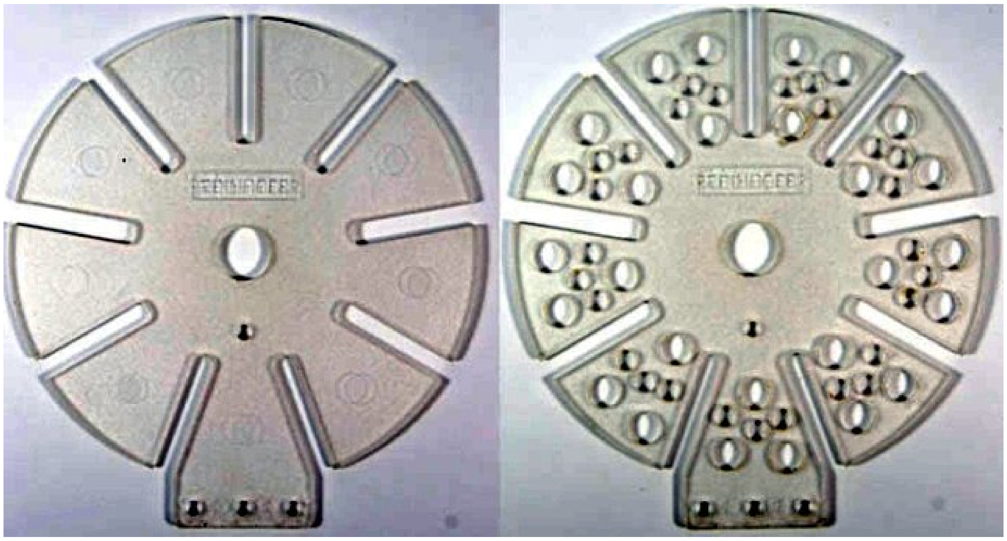

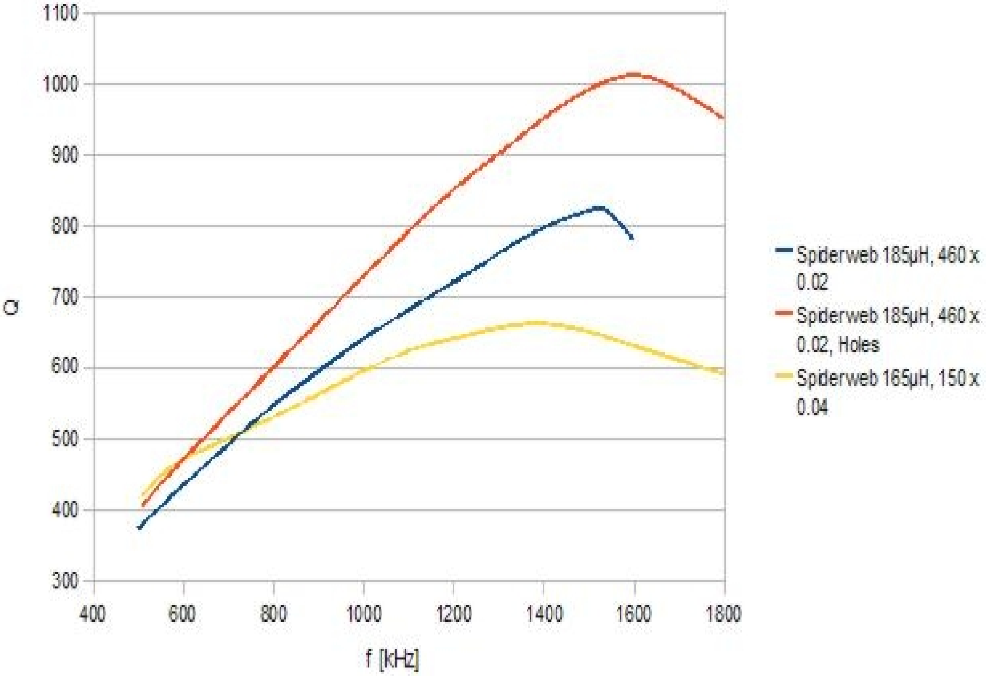
1100
1000
900
800
700
600
500
400
300
Q
400
600
800
1000
1200
1400
1600
1800
f [kHz]
Spiderweb 185µH, 460 x 0.02
Spiderweb 185µH, 460 x 0.02, Holes
Spiderweb 165µH, 150 x 0.04

In the (broadcast band/mw) contest, I used my latest receiver which was just finished right in time. The schematic features a double-tuned circuit with an additionally antenna matching cap and probably does not show surprises. The whole device was designed to gain highest Q values within my possibilities. I used polystyrene coilformers of the german manufacturer Reinhöfer which already offers the possibility to gain high Q values.

The coils have 185µH and were wound with 460 x 0.02mm litz wires. Measurements of such a coil show increasing Q with frequency up to 1550kHz and a maximum value of app. 830. A noteworthy improvement was reached by drilling six holes in each segment of the coilformer, now the Q increased across the frequency range of interest (in Europe, station in the mediumwave band cover a range from 531kHz to 1611kHz). Moreover, also a maximum Q of app. 1020 was found. Therefore I decided not to use contra-coils since I wanted to avoid the use of any (more or less lossy) switch in the RF section.

To preserve the high Q, I used TRW silver-plated air variables in the antenna and detector circuit. Coupling of the coils is made variable by a somewhat old-fashioned technique as used in old tube radios. Anyway, it is working very smoothly and reliable.

Housing is done by low-loss polypropylene sheets which keeps the Q up (refer the measurements of Dick Kleijer at http://www.crystal-radio.eu). In difference to most constructions presented so far, I completely enclosed the device. The main reason is to keep dust away from the caps - if you ever did Q measurements of high-quality aircaps with dust in between the plates, you will understand why. The dust probably kills more Q than the housing might do, providing the use of appropriate materials (thus, no wood, PVC, metal, acryl glas or most colored plastics).

All variable caps are completed by 6:1 vernier dials made by UK Jackson Brothers. 5082-2835 schottky diodes are used as detector diodes and provide the best matching between detector circuit and audio transformer I was able to realize. The matching transformer with an input impedance of 500k was also manufactured by Reinhöfer. The radio is completed by a pair of balanced-armature headphones Baldwin Type C as well as the usual benny circuit.

The real backdraw of my receiver is the limited antenna. I live in a five-storied house which not much chances for long antenna wires. Thus my best option was a 10m wire between my balcony and the roof of the next garage. Compared to my other set which I used in the last sprint contest and still used for the longwave and shortwave band in this year, selectivity and also sensitivity was noteworthy improved. The 6:1 vernier dials are already a minimum requirement, 12:1 or more would be favourable.

In all, I was surprized about how many and how distant stations I was able to receive during the contest.

Shown below is the set used in this contest for SW and LW. It is the same set used in the 2008 Sprint Contest.

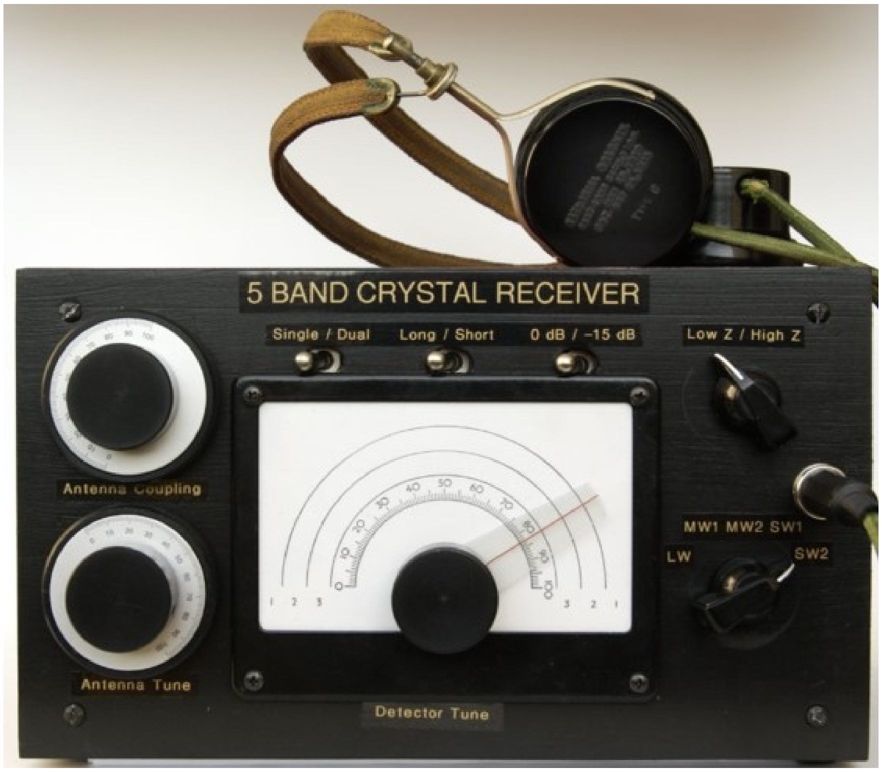
5 BAND CRYSTAL RECEIVER
Single / Dual
Long / Short
0 dB / -15 dB
Low Z / High Z
Antenna Coupling
Antenna Tune
Detector Tune
MW1 MW2 SW1
LW
SW2

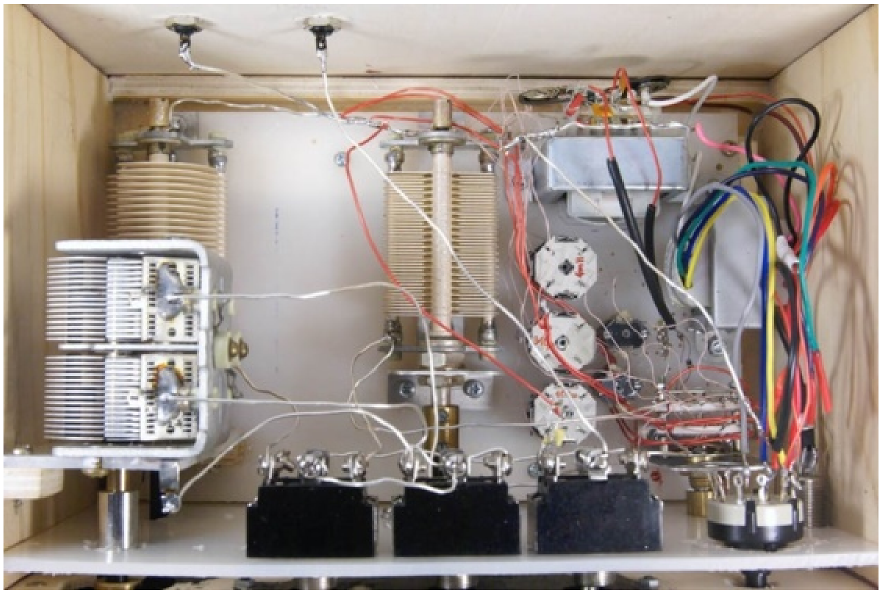

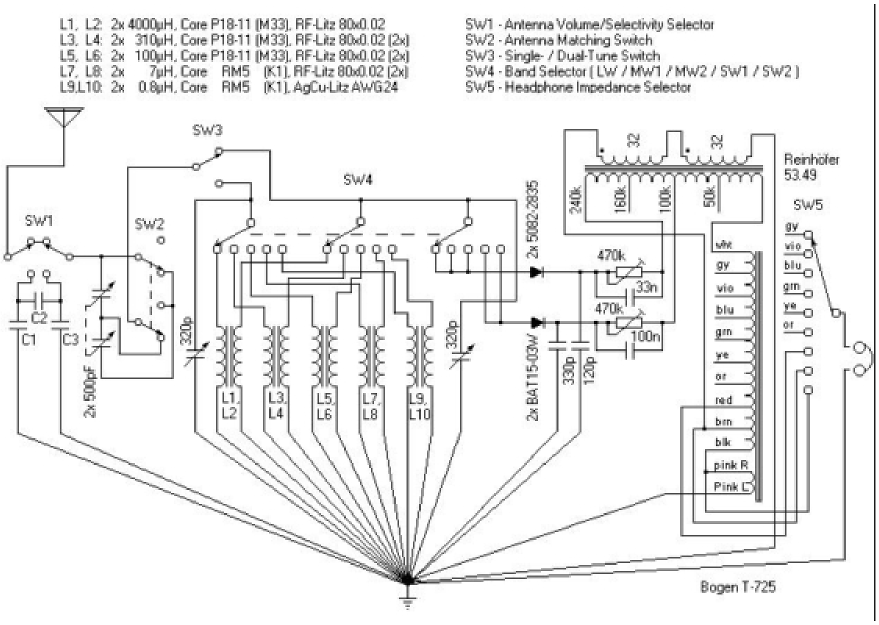
L1, L2: 2x 4000µH, Core P18-11 (M33), RF-Litz 80x0.02
L3, L4: 2x 310µH, Core P18-11 (M33), RF-Litz 80x0.02 (2x)
L5, L6: 2x 100µH, Core P18-11 (M33), RF-Litz 80x0.02 (2x)
L7, L8: 2x 7µH, Core RM5 (K1), RF-Litz 80x0.02 (2x)
L9,L10: 2x 0.8µH, Core RM5 (K1), AgCu-Litz AWG24
SW1 - Antenna Volume/Selectivity Selector
SW2 - Antenna Matching Switch
SW3 - Single- / Dual-Tune Switch
SW4 - Band Selector (LW / MW1 / MW2 / SW1 / SW2)
SW5 - Headphone Impedance Selector
SW3
SW4
SW1
SW2
C1
C2
C3
2x 500pF
320p
320p
L1, L2
L3, L4
L5, L6
L7, L8
L9, L10
2x 5082-2835
2x BAT15-03W
330p
120p
470k
33n
470k
100n
32
32
240k
160k
100k
50k
Reinhöfer 53.49
SW5
wht
gy
vio
blu
grn
ye
or
red
brn
blk
pink R
Pink L
gy
vio
blu
grn
ye
or
Bogen T-725

Dan McGillis, Open Class Entry

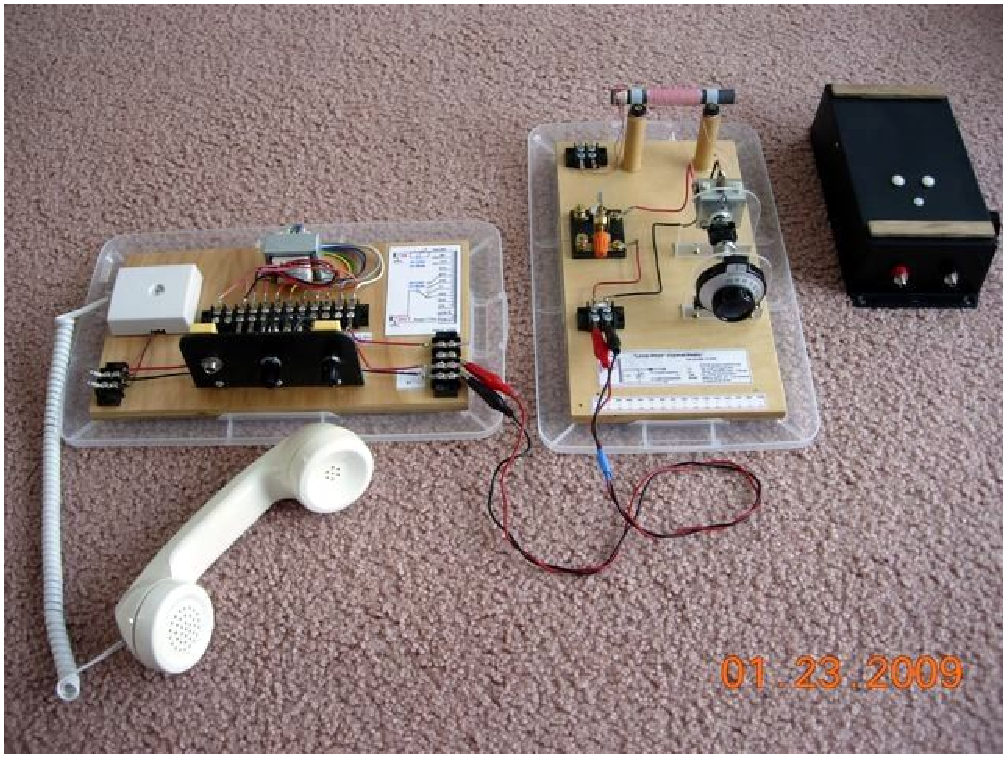

My entry this year is a simple loop-stick crystal set - something the grandkids can play with. It's been described at the RadioBoard.

I used a little antenna tuner that's also been described before. Click here for the RadioBoard link for the tuner.

The set didn't do too bad for only a few days of listening. A big 200' antenna certainly helps :-) It makes a nice progressive set. I started out using Mike Peebles's cat's whisker detector stand It's very smooth & works well. Substituting a FO-215 diode perked things up. Adding a Bogen+Benny for some audio matching made a big difference. Coupling-in a simple antenna tuner really brings-in the DX and kills the SW ghosts.

Mike Tuggle, Lyonodyne-17, Open Class Entry

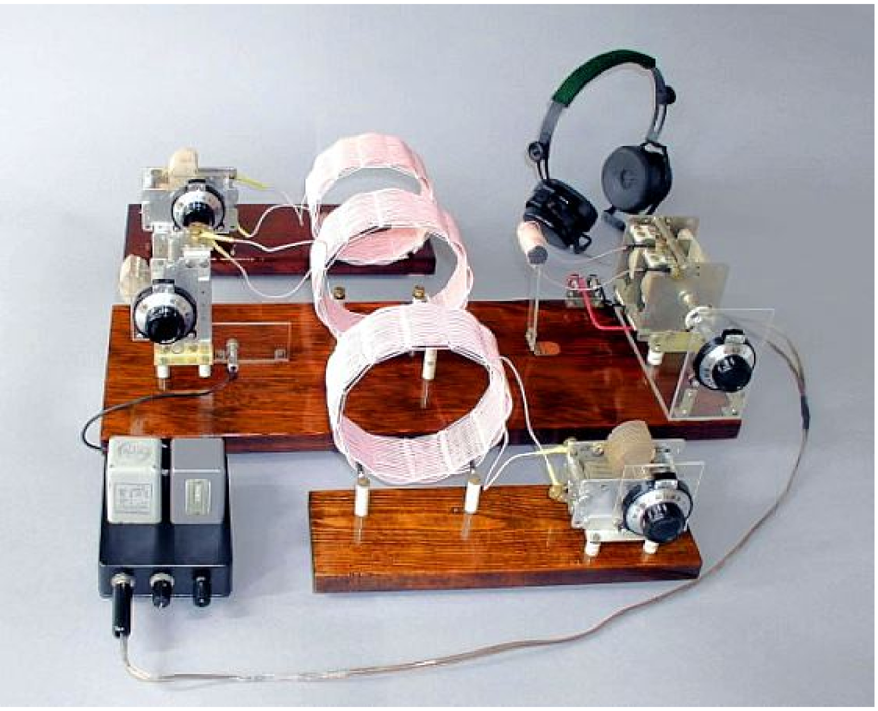

This year's effort was something of a disaster. The night before the start, our annual Kona winds blew the antenna into the mango tree and ripped it to shreds. I spent opening day out in wind and rain, in true ham fashion, getting the antenna down and re-erecting a simple longwire in its place. Conditions throughout the contest were blah. I don't know whether the original 4-wire flat-top was really that much better, or if conditions were simply that blah.

Jack Bryant, Open Class, Two-Way Shortwave Class, and Below BCB Class Entries

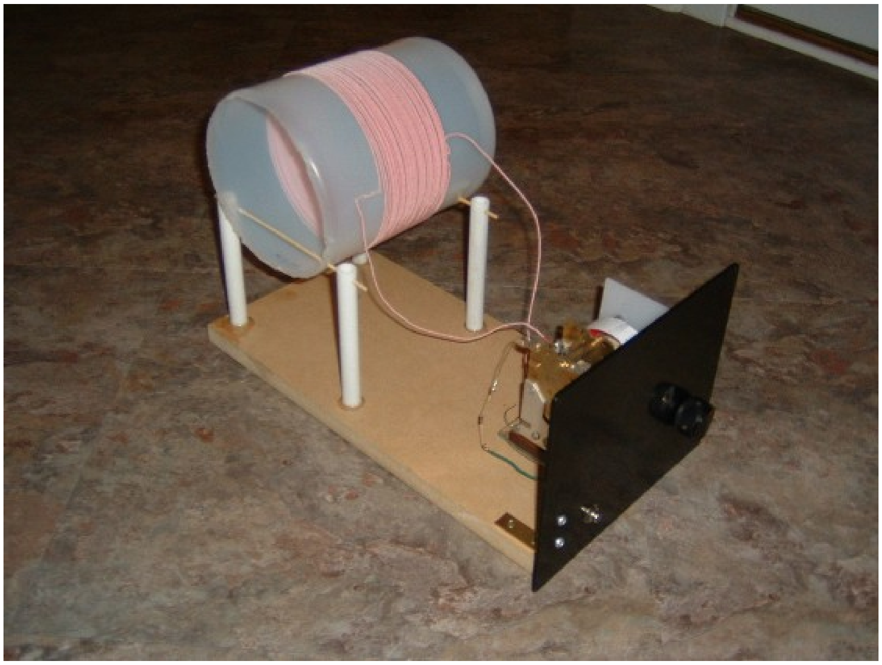

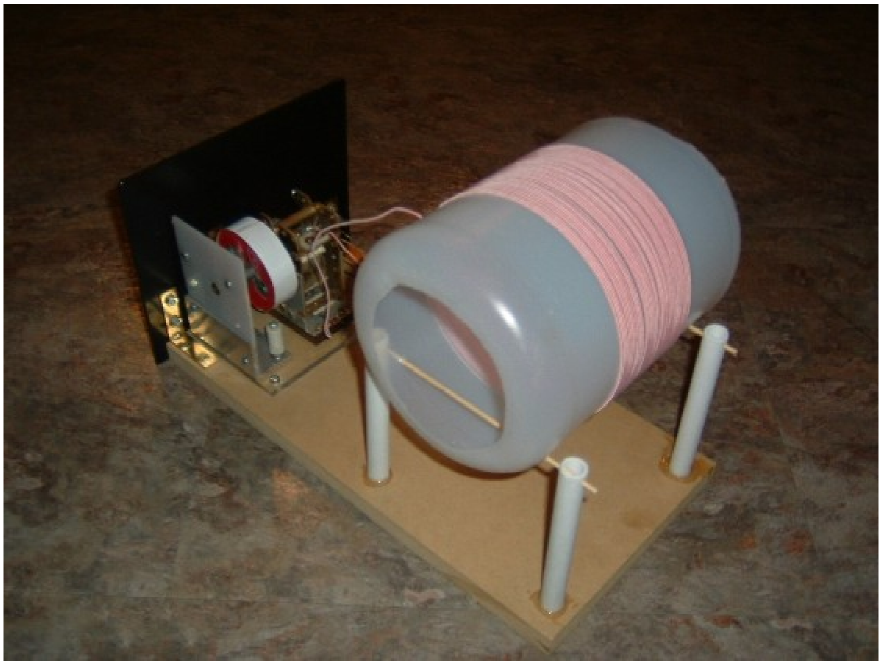

I only worked the first 24 hours of the contest on the broadcast band. I had copied one navigational beacon at 529 kHz, and decided to build a simple set for LF. I copied four more stations for a total of five. The stations are: DIW, CLB, BH, YYW, and LYQ.

I worked a few contacts on 75m AM using the xtal set as a receiver. My transmitter VFO had bitten the dust, but I got a crystal for the transmitter on Saturday, and that worked fine.

I used essentially the same set for the broad cast band that I used in years past, with minor mods to the antenna tuner. I did use Schottky diodes for the fist time. I used two in parallel during the day, but moved to just one at night due to the overload I experienced using the pair.

This year I used a different set of phones made from SP mic elements. The earphones were made of hard plastic, but the headset band placed the phones on the back part of my ears. I did not experience the ear pain of previous years, but I made sure to move my glasses far up on my head, or not use them at all to eliminate pinching the nerve beside my ear.

I have used a Realistic DX-398 for several previous contests, and I used it for this one, too. I used a PC for logging along with paper copy backup. I used a four foot table from Lowe's as my listening post.

Continued on Page 2.

Ad ✕

Select Menu

- Home
- About Us
- Contest Results
- 1AD Contests
- 1AD Entries
- Antenna Matching
- Antennas
- BHAM Radios
- Crystal Contests
- Crystal Radio Entries
- Sprint Contest
- Headphones
- Projects, Mods
- Favorite Links

- Contest Results
  - 2004
  - 2005
  - 2006
  - 2007
  - 2008
  - 2009
  - 2010
  - 2006 Sprint
  - 2008 Sprint
  - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2005 Contest Entrants

Page 1

Charles Pullen (Chuckster)

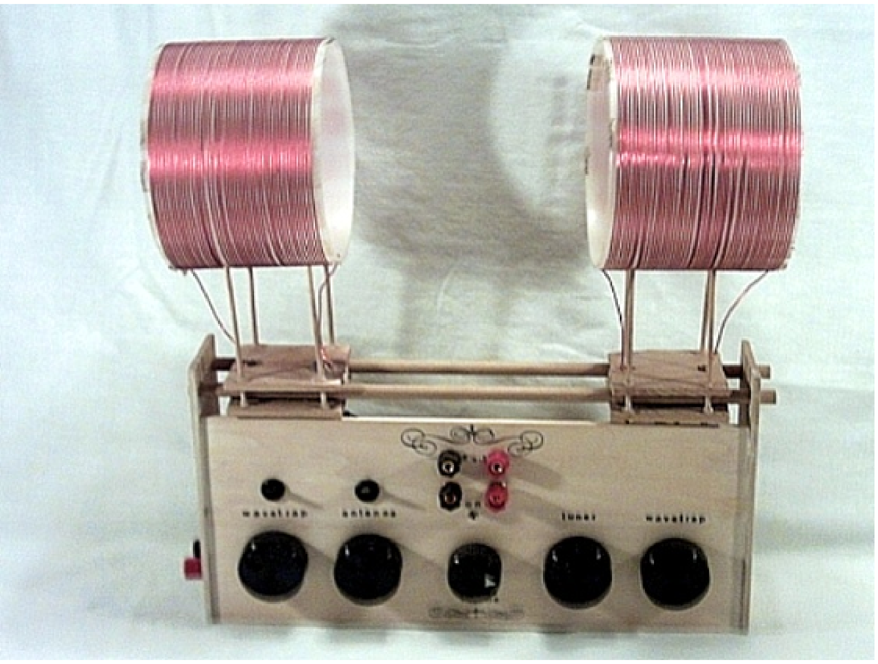

Basic Double Tuned Set

The antenna coil is a 5.5 inch space wound with 16awg @ 240uh paralleled with a 20-340pf air variable, loosely coupled to a 5.5 inch space wound with 16awg @ 270uh paralleled with a 20-340pf air variable detector coil. The detector used was two 1ss98 schottky diodes in parallel. I Used the original ultimatch circuit by Steve Bringhurst for impedance matching. Used two Philmore xtal earbuds in parallel, connected to the ultimatch with a radio shack mono Y-adaptor.

The lone wave trap is a 4" x 3/8 inch ferrite rod wrapped with litz to obtain 270uH with a parallel 20-340pf air variable. This was arranged so it could be loosely coupled to the detector coil, suspended inside the center of the detector coil or outside the detector coil in axis with its center, or rotated 180 degrees out of the way. I found this particular style of wave trap to be extremely effective for even the strongest local.

The antenna was two separate wires, one 50 foot wire tacked to the ceiling running about 30 feet east-west the remaining twenty feet running north & south. The second antenna wire was a portable reel of 25 feet borrowed from my Grundig YB and attached to the set. This is in a second story apt. Ground was about 25 feet of speaker wire running to a CWP.

Dave Schmarder

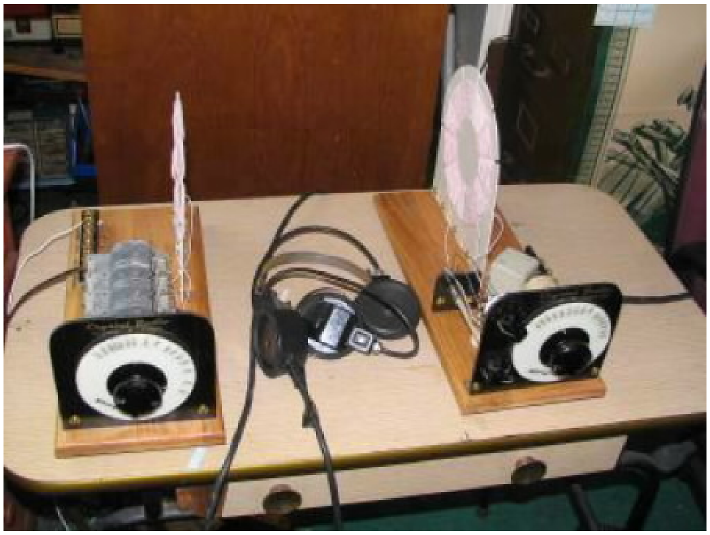

See complete description of Dave's radio at this link: Dave's #50 Set

Open Class

Dual 660/46 litz coils and two tuning capacitors.
differential coupling capacitor
Schottky Diode Detector
Sound powered headphones
Antenna 40 meter wire up 3 - 4 meters running east to west.

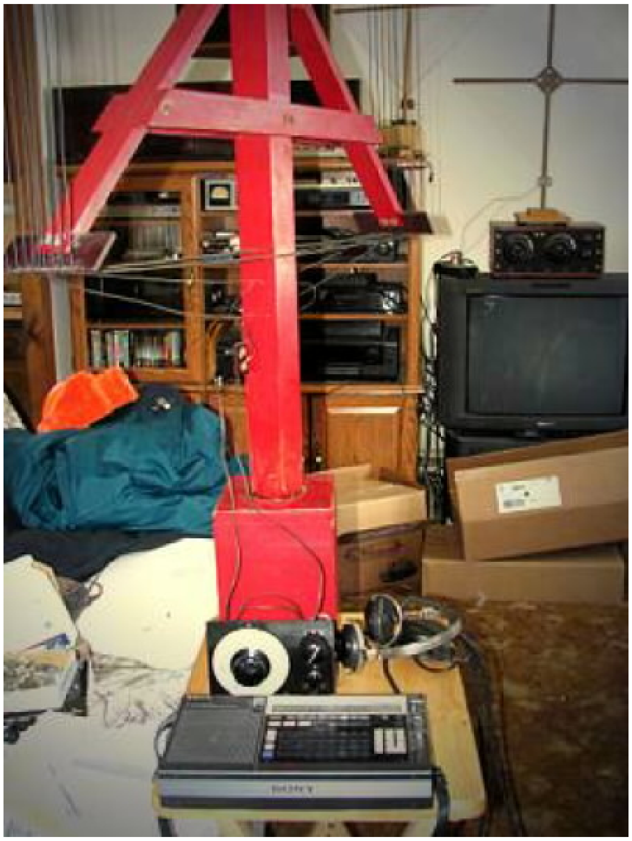

33 inches (84 cm) per side. FO-215 Detector, Bogen transformer, Sound powered phones

See the details of this loop at Dave's Loops Page.

Bill Meacham (Exray)

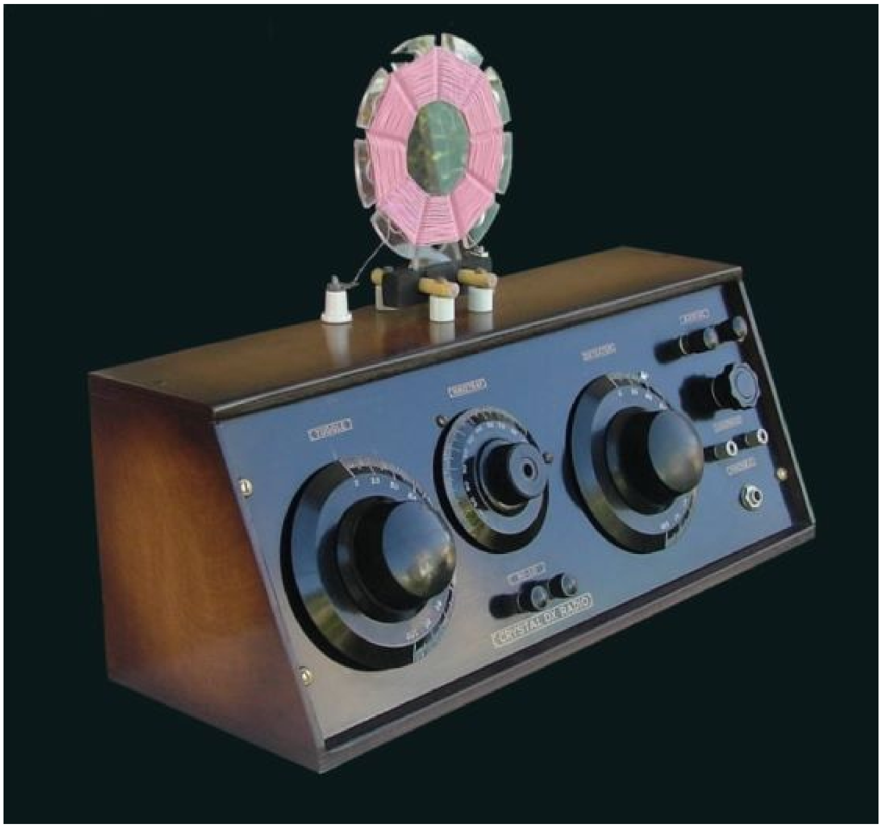

This set is my third generation of a model based on Steve Bringhurst's TK-2 toroidial set. I went with an external detector coil on this one which offers the advantage of being able to close-couple an external trap (not shown) to the detector coil to notch out instances of interference when the coil itself is picking up local signals via the air.

The set is built on a Garolite front panel with custom-made decals and lacquered for durability and shine. All of the RF components are on standoffs to avoid any possible leakage path through the panel. Vintage tuning capacitors were used and they have been thoroughly cleaned and checked for rf integrity. Both are straightline tuning types to minimize crowding at the high end of the band.. The input tuner is of the "Tuggle" type and uses a toroidial inductor. The inline wavetrap also uses a toroid. The detector uses a nice Pilot Kilograd tuning cap and a coil of 165-strand litz wound on a Lexan former. Its Q is greater than 500 across most of the band. Detector is a germanium FO-215 and the set uses a Calrad output transformer with an adjustable AF impedance matching control (benny) for minimizing distortion on strong local stations. I have been using the set with sound powered phones and they match well.

Calibration on this set is very accurate and it's easily resettable to locate particular stations. I added a set of front panel terminal posts to allow for tinkering with other detector tank options. The coil mounting assembly on top is also configured in this fashion to allow experimentation with other coils. Ceramic standoffs and feedthru bushings were incorporated to assure minimal losses.

Cabinet is homemade, made with Baltic Birch and is stained and lacquered using clear and toned lacquer.

Ed Phelps

.. is the same 140 ft, base insulated, vertical I used last year. (picture is crooked, not the tower)

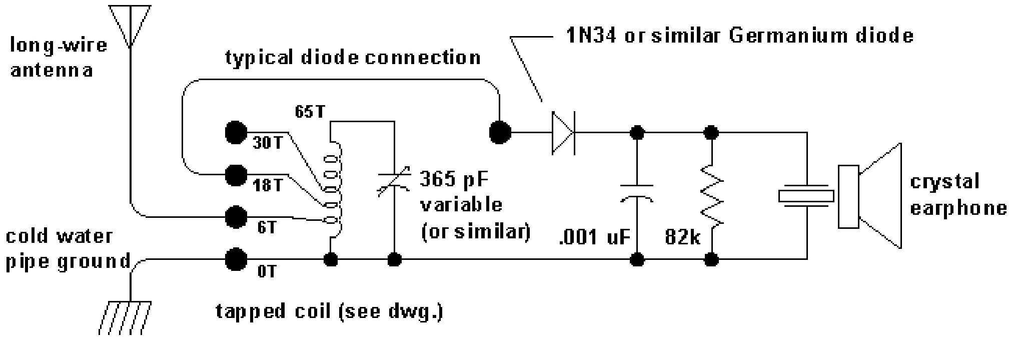
long-wire antenna
typical diode connection
1N34 or similar Germanium diode
65T
30T
18T
6T
0T
365 pF variable (or similar)
.001 uF
82k
crystal earphone
cold water pipe ground
tapped coil (see dwg.)

Crystal radio with typical connections for a long wire antenna and good ground connections. The diode is connected for weak signals and moderate selectivity.

I made a few minor changes to this circuit. In place of the 1000 pf capacitor, I use 250 pf. The 82k resistor, I changed to 100k. I use a pair of ceramic earphones, wired in series.

Eric Haydon

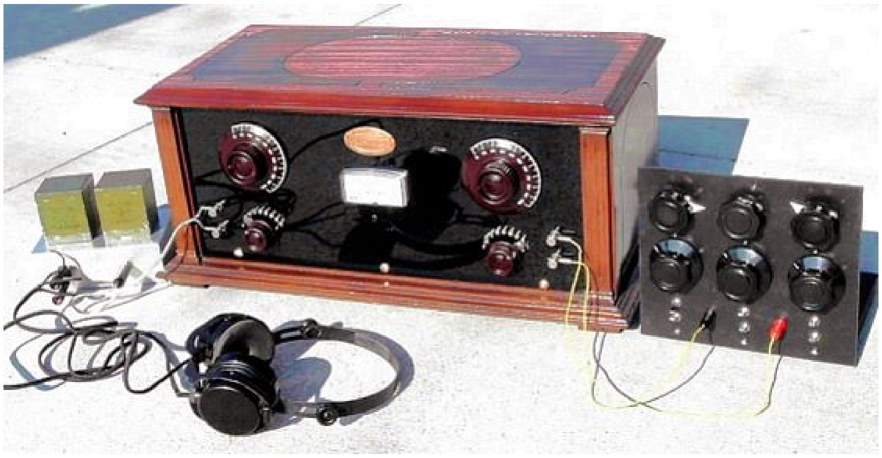

The contest this year was a lot of fun. I put in many many hours of listening. The band was very up and down in Nebraska. Thursday and Friday nights were the only nights that I would call "normal". The unusual conditions on the other nights did however allow for the reception of a few stations that I had never heard before on my crystal radio. I probably heard at least 30 more Spanish speaking stations than I logged. I only logged the ones that were very strong or heard almost every night.

I heard many more stations this year than last year due to several improvements that I have made to my listening station. I used the same crystal radio as last year with the following improvements. My antenna is now higher and has been extended to about 240 feet in length. My ground system is improved. I was lucky enough last summer to acquire a pair of RCA "big can" headphones in very nice condition. They do help some. I finally built some filters that work very well to trap strong local stations. This allows me to now receive quite a few new stations on frequencies close to my local stations. The last improvement > is that every year I get "smarter".

Gil Stacy

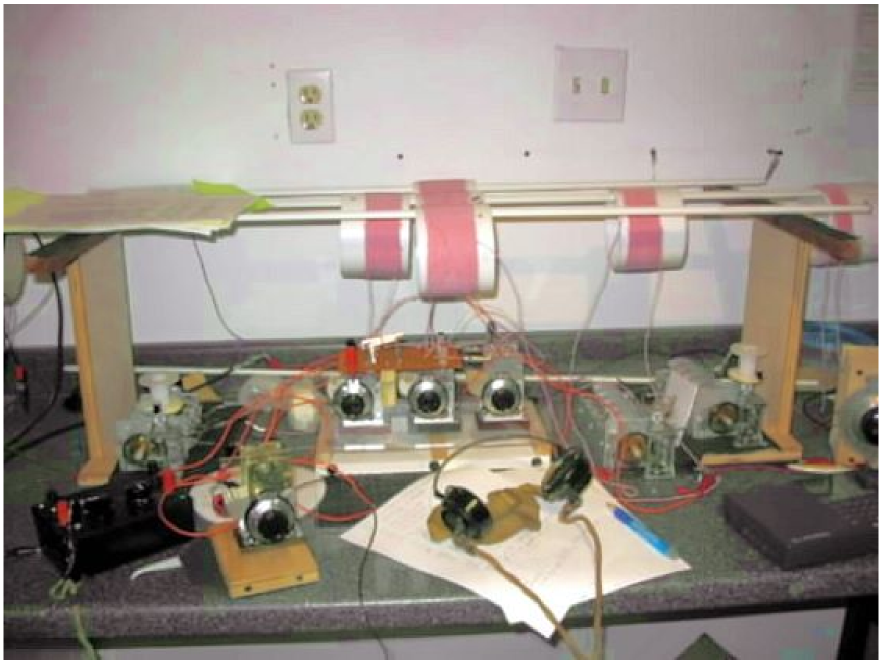

Set: Same set as last year.
Headset: RCA Big Cans tuned and recharged by Steve Bringhurst.
Antennas: 88’ top loaded vertical up 50’ and 180’ inverted “U” up 50’.
Ground is the post 2004 DX contest ground described at the bottom of
this page

Jack Bryant

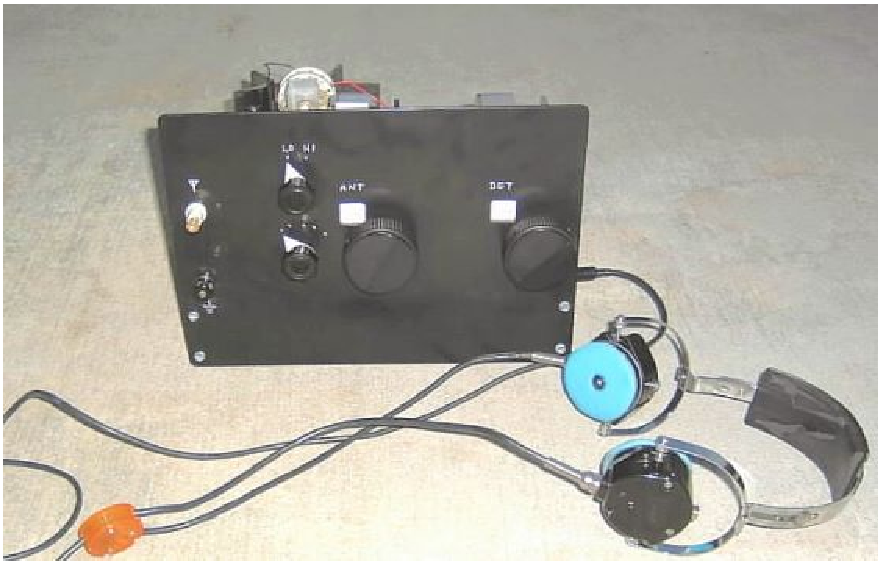

I used the 2004 BCB Contest set again this year. We moved after the last contest, so the antennas, ground system, and location were different.

I used two antennas. There first was a top loaded vertical. The top section is about 110' long and is a 80M - 10M trap dipole. The antenna is up about 50' high. The coax feed line is used as the feeder.

The second antenna is an upside down U, roughly 40' vertical on each end, 260' horizontal. I feed it with ladderline in the horizontal section, about 20' from where one end goes vertical. This makes it about 60' from one end of the wire. The antenna was used in three different configurations:

1) Each bottom vertical run is grounded. This is the grounded loop configuration.

2) Only the end closest to the feedline is grounded. This is similar to an inverted L approach

3) No ends are grounded. This is similar to an off-center fed dipole. All changes were done manually (no switches or relays yet).

Continued on Page 2.

Need to get away?
www.hopjump.com
SAVE on HOTELS in your HAPPY place!
Bahamas Florida California Mexico

Select Menu

- Home
- About Us
- Contest Results
- 1AD Contests
- 1AD Entries
- Antenna Matching
- Antennas
- BHAM Radios
- Crystal Contests
- Crystal Radio Entries
- Sprint Contest
- Headphones
- Projects, Mods
- Favorite Links
- Contact Jack

Sub Pages

- Main Page
- 2004 Page 1

- Contest Results
  - 2004
  - 2005
  - 2006
  - 2007
  - 2008
  - 2009
  - 2010
  - 2006 Sprint
  - 2008 Sprint
  - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2004 Contest Entrants

Page 1

Lou Dayich

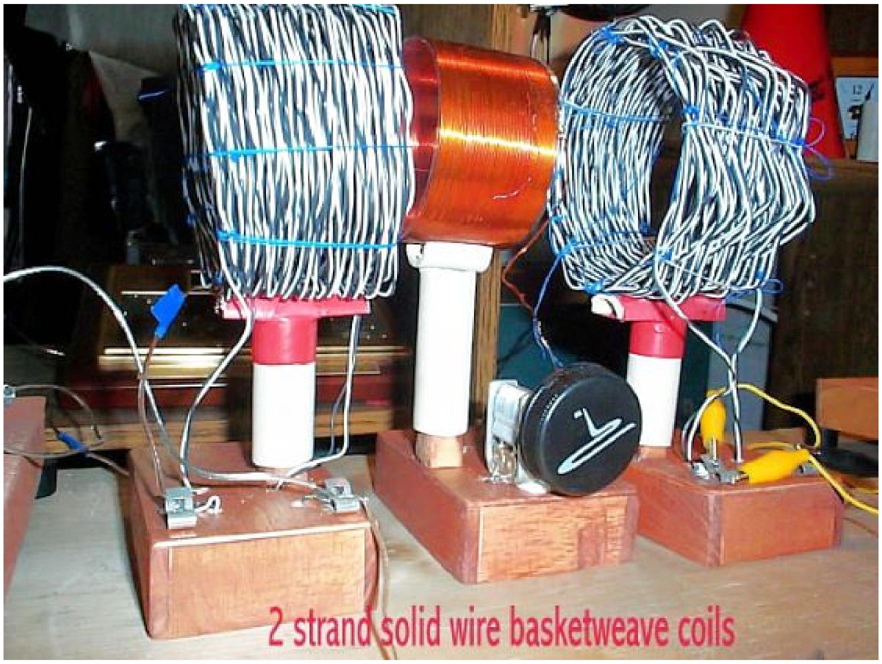

I have attached a photo of the set that I used during the contest. Daryll Boyd posted a picture on his site. I humorously called it "The Bondini". The most interesting feature is that this uses 4 inch basket weave coils with 9 points- double tuned.

The wire is 2 solid strands of twisted wire. I use a schotty diode and sp phones with a Bogen matching transformer. I wired it in as many combinations of coil and cap as time would permit.

Bob Jewell

Here is Bob's DX set.

Alex Perez

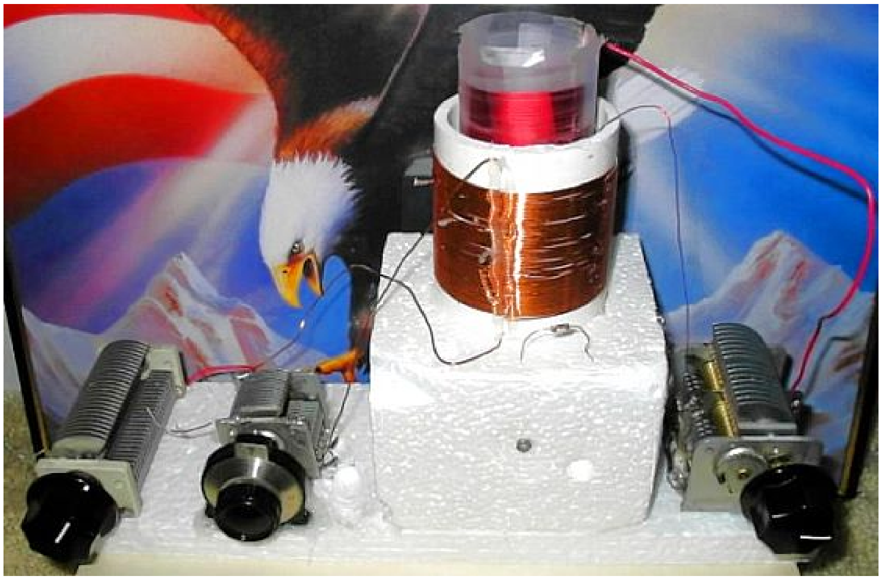

Hobby Class Single Tuned Crystal Radio Capable Of DX My antenna consists of 70 feet of 18 gauge plastic insulated tinned copper wire oriented in a North East and South West direction with the lead in pointing North East. The far end is elevated over 40 feet while the other is only 12 feet high. A cold water pipe ground connects to the receiver via 15 feet of insulated 14 gauge solid copper wire.

A 250pF ceramic frame air variable capacitor couples the antenna to the receiver. The tank circuit is composed of a 365pF phenolic insulated air variable capacitor and a coil of 60 turns of 21 gauge wire wound on a 2¾ inch PVC form. It is tapped at 30, 40, and 50 turns from the ground end.

To combat the QRM generated by WXEM, only a mile away, the wave trap has to be closely coupled to the tank circuit. The trap is composed of a 250pF phenolic insulated air variable capacitor in parallel with a 60 turn coil of 26 gauge wire on a one inch diameter form.

Best sensitivity and selectivity is obtained with the 30 turn tap. Usually a single Philmore crystal earphone and a selected germanium diode were employed. Some improvement was noticed when using two earphones in series, probably a better impedance match.

I attribute my success with such basic equipment to mostly quiet RF environment, many hours spent DXing, and listening well into the night and the morning on one occasion. Furthermore, I was fortunate to acquire some high scoring stations. Somehow I heard two different stations on 630 25 minutes apart. WAVU was only 203 kilometers away but running a mere 28 Watts! Still, this does not compare to KEYH on 850. It was 1,211 kilometers away and transmitting at a power of just 185 Watts! The remaining acquisitions were locals, 50 KW Clear Channel, or not too distant 5 and 1 KW stations. I received one Canadian and one Cuban station, both 50 KW. The furthest catch was WBZ, Boston. A station on 1260 identified itself as Radio Disney but I was unable to obtain a call sign. Out of three possible candidates, I am fairly sure it was WWMK. Using a spotter set to ascertain direction,the station was either from Chicago or Cleveland, certainly not Boston. 5,000 Watt DX from 800 kilometers away is certainly not uncommon for me, but anything else would be. Besides, I did log a few stations running 5,000 Watts from Ohio.

Alex Perez
KG4IHN
Age 15

Evan Haydon

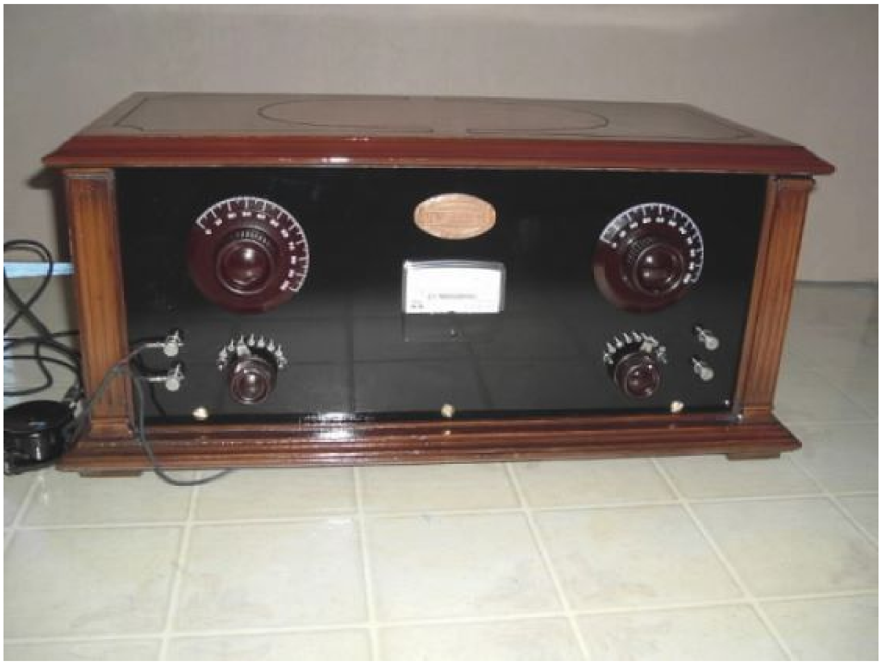

My crystal radio is a design that has evolved after many years of searching for a good design. Most of the ideas used in this radio came from internet sites. No litz wire here. Just old cotton covered 20 gauge wire wound in a basket weave form. No sound powered headphones here. Just my trusty old magnetic Trimm Dependable's.

Yes, the meter is in the circuit all the time. It is a 50 µA movement in series with the headphones. It is helpful in maximizing design of the circuits, selecting the best crystal to use, and in tuning taps and caps for maximum signal reception. It indicates carrier strength of the received signal. The carrier contributes nothing to the audio heard in the headphones. A signal of 1ua is 100% readable. One local station gives a reading of about 500 ua.

My antenna is 18 gauge solid magnet wire. It is basically a Windom antenna that is insulated at the supports which are at the ends and at the lead-in point. It is 180 feet long and runs due east and west. The wire to the receiver comes off at a point about 60 feet from the west end and is about 55 feet long. The lead-in take off point is up 36 feet with the ends up about 24 feet.

The radio is made in a 20's battery radio wood cabinet. It uses three Atwater Kent variable capacitors. It uses Atwater Kent knobs with two sets of antique switch points mounted on a black plastic front panel. Genuine Atwater Kent copper emblem at the center top of the front panel. Antique connectors are on the front panel for antenna and headphone connections. Wiring inside is all by 16 gauge bus bar. 1N34A diode. The primary and secondary coils are 4 1/2 inches apart with no electrical connection between them. I have not found a trap circuit that helps enough to talk about.

ATWATER KENT
D.C. MICROAMPERES
0
10
20
30
40
50
60
70
80
90
100

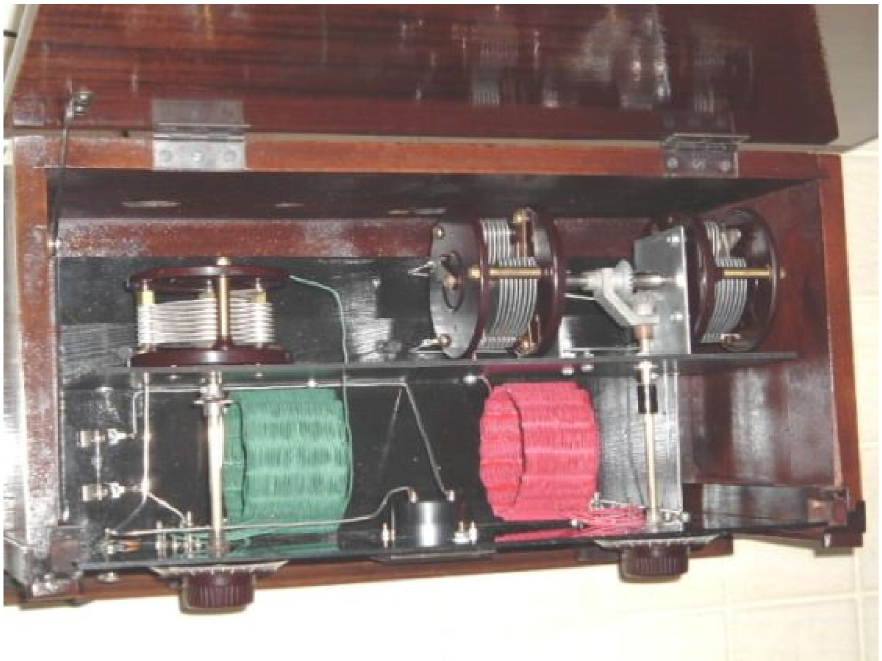

Berthold Bosch

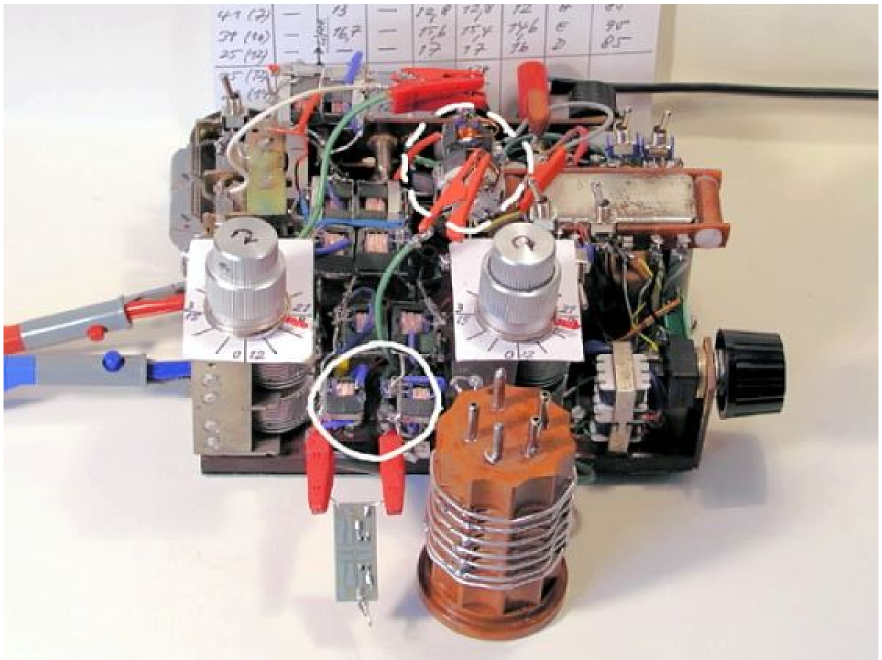

2004 XSS DX Contest Information (SW, Open Class)

My 2004 contest log containing 156 SW stations (incl. 5 of uncertain ID) suggests that conditions in Central Europe are still more different from those in the USA than on the MW band. There are a lot of high-power SW/BC stations of 100 to 1000 kW in the numerous European countries and the neighbour areas, and particular networks that operate on rather many frequencies - like R.Free Europe, R. Liberty, VoAmerica (partly R.Sawa), VoRussia, etc. with their various programs. It is mostly possible to ID the received stations of these networks, but often difficult to determine the site (country) from which the particular tx operates. Only a few real DX stations are in my log with China/Peking being the most distant one (9500 km). So, is it really fair to compare my results with those obtained e.g. in the US if one gets one point,also for every "nearby" SW station?

I used my 2003 contest set (see Wuggy´s 2003 Contest site) which I extended for SW reception. This consisted in installing an air coil of 4 uH (fore in photo) for the 13m and 16m bands, a pair of Epcos pot cores of SW material K1 with solid-wire coils of 5.5 uH for 19m to 25m (marked by broken circle), and pot cores with 45/46 Litz coils of 10 uH for the 31m to 75m bands (full circle). Except for the air-coil tank I generally employ double tuning. The principal circuit layout for SW is the same as in the 2003 set (Ben Tongue approach). With the 5 SW coils there are now in total 13 coils in the set (LW, 3*MW, 3*SW), not counting those provided for triple tuning on extra boards (Boschodyne Plus).

At SW frequencies the obtainable tank resonance resistances are much lower than those at MW. Consequently sensitivity is a serious problem, and more often than on MW one hears two or even three stations at the same time. For improving the situation I employed a modified MFJ tuner 901B in front of the set (not shown in photo) to squeeze the last dB of power out of the antenna and trade it for selectivity. The tuner can be wired to act as a SPC or SL pre-selector, an "ultimate transmatch", or a capacitively-coupled third tank circuit (for triple tuning). With triple tuning I measured for example at 41m a loaded (i.e. with A/G and phones connected) -3dB bandwidth of 23 kHz and a -10dB b.w. of 47 kHz. -- I use a Schottky diode BAT 62 (as in photo) if high sensitivity is the prime goal, and 2 or 3 paralleled 5082-2835 if it is selectivity. The outdoor loop that I described last year in my contest report served as antenna.

For finding and then logging stations I employed two methods:

(1) First I scanned the band with my spotter radio ATS-404. When the scan had come to stop at a station, I tuned the xtal set to the particular frequency with the help of a loosely-coupled signal generator. After turning off the signal generator I then listened if I could hear that station in the SP phones of my set.

(2) When I had the xtal set tuned-up to a particular frequency within the SW band investigated, I slowly moved the sig. generator about 50 to 70 kHz around this frequency while watching for beat notes. When I found a strong beat I tuned the set to zero this beat, then turned off the signal generator, and listened whether I could hear the particular station. This was done repeatedly to scan various parts of the band. If I had been able to definitely hear the station in methods (a) or (b) I partly switched-on the spotter radio again for helping to identify the station received on the xtal set.

During this concentrated listening on SW, I found the rule of thumb again confirmed that Mike Tuggle formulated in a recent paper: "If you can hear them on a radio, you can hear them on a crystal set" - nearly always in my case.

Berthold Bosch

Dejan Momirov

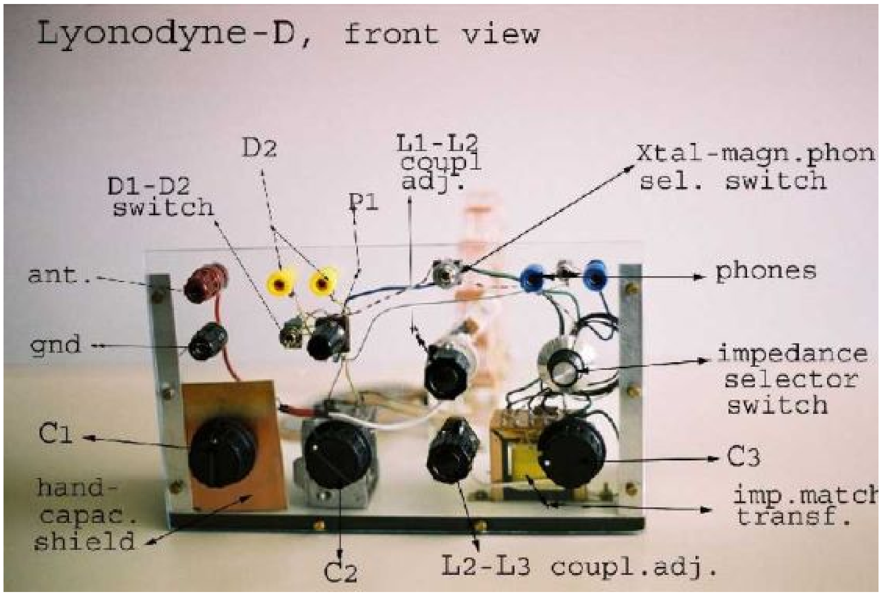
Lyonodyne-D, front view
D2
L1-L2 coupl adj.
Xtal-magn.phon sel. switch
D1-D2 switch
P1
ant.
phones
gnd
impedance selector switch
C1
C3
hand-capac. shield
imp.match transf.
C2
L2-L3 coupl.adj.

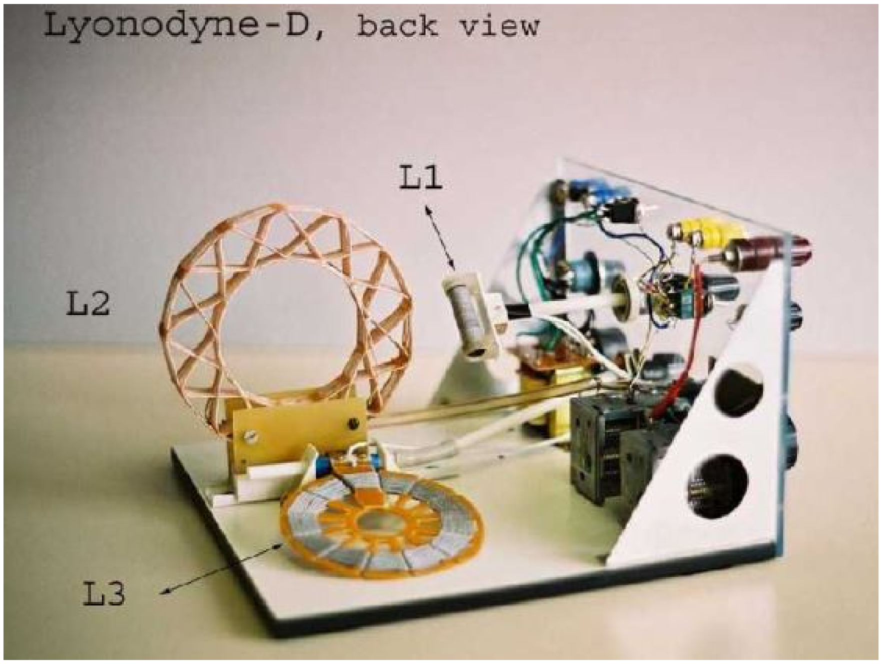
Lyonodyne-D, back view
L1
L2
L3

2004 DX CONTEST SET DESCRIPTION

This is basically a Lyonodyne with some differences, such as:

1. The complete set has been built on one chassis and all controls are at the front panel.

2. The C1-C2 coupling is double: by means of rotation and changing the distance, and wave-trap is coupled only by rotation.

3. Using one or both sections of the C2 spreads the upper BCB.

4. There is one built-in detector diode with the possibility to experiment with other types of detectors.

5. Audio transformer is home-made and together with the impedance-selection switch allows the use of different impedance headphones, as well as a sensitive speaker.

I am aware of all the imperfections which are subject to experimental improvements. I want to thank Mike Tuggle who invited me to join the Contest and gave me many useful advice. My gratitude also goes to Ben Tongue for helping me to improve my impedance-matching transformer.

N.B.: The switch s1 is not shown on the photo (introduced after the photo was taken!).

antenna position

Gil Stacy

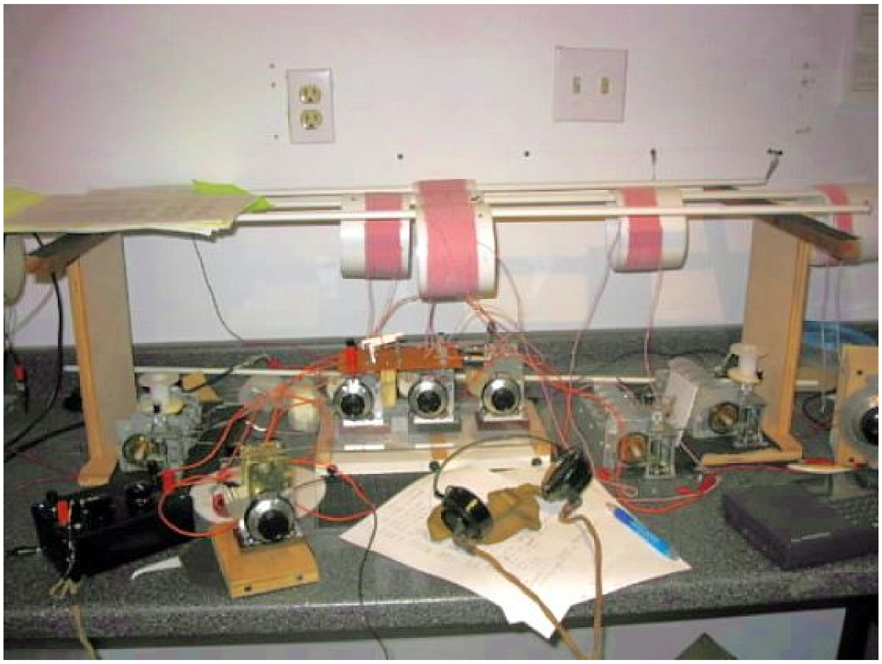

Five Litz 4.5” O.D. coils are in use at all times. The detector and antenna coils are controlled by the three 8-1 vernier dials mounted in plexiglass. The antenna circuit is a modified Tuggle front end approach with one TRW cap (C1) in series with the antenna with C1’s stator linked to C2’s stator. C2 is tied to the antenna coil and earth ground. C3 is tied to the detector tank. All caps are TRW’s previously available from Fair Radio. This offers the set more flexibility in tuning than its previous traditional Tuggle Front end which used a dual ganged cap.

The three inductive traps surrounding the detector tank are driven by one gang each of three five ganged traps with right angled 120-1 drives from Fair Radio. Hand capacitance is controlled by use of empty plastic wire spools, slotted to accept the dials of the right-angled drives. The stand-alone TRW cap with vernier dial controls the inline toroid trap. The three inductive traps are permanently set to trap locals 1400, 1290 and 1230. The inline trap is used trap two local stations at 630 and 900 when in their vicinities. Otherwise it is used to assist swatting down 1230. Untrapped, 1400, 1290, and 1230 will show up all over the band because of their proximity to my location. The phones are DLR5s matched to the set via an STM designed by Steve Bringhurst. The set uses 3 paralleled Agilent 5082-2835s and a 3RT 12101 which are selected via a switch. To the right of the set is my frequency spotter Radio Shack DX 398. The antenna is an end fed wire, approximately 140’ long, with a 90’ horizontal run, up 50 feet. A post-contest improvement by adding three 8’ copper-plated ground rods to existing water main ground has noticeably improved selectivity, sensitivity and my ability to control the qrm of locals 1230, 1290 and 1400. The new ground system has also opened up about 150 KHz previously lost to the above three locals.

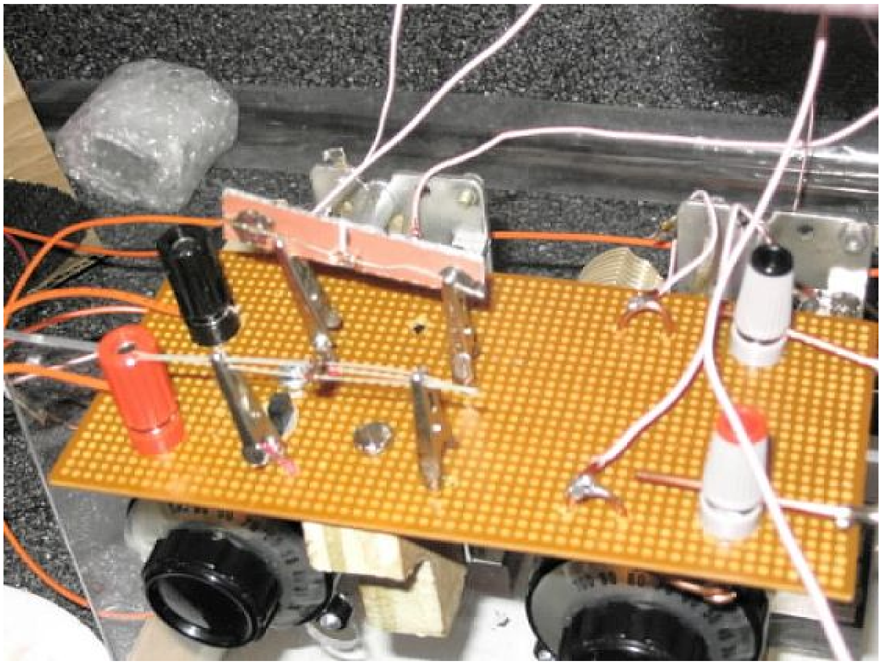

Detector Board

Continued on Page 2.

$67.09 Shop ➔ $29.99 Shop ➔ $19.99 Shop ➔ ebay

Select Menu

- Home
- About Us
- Contest Results
- 1AD Contests
- 1AD Entries
- Antenna Matching
- Antennas
- BHAM Radios
- Crystal Contests
- Crystal Radio Entries
- Sprint Contest
- Headphones
- Projects, Mods
- Favorite Links
- Contact Jack

Sub Pages

- Main Page
- 2004 Page 1
- 2004 Page 2
- 2004 Page 3
- 2005 Page 1
- 2005 Page 2
- 2006 Page 1
- 2006 Page 2
- 2006 Page 3
- 2007 Page 1
- 2007 Page 2
- 2007 Page 3

- Contest Results
    - 2004
    - 2005
    - 2006
    - 2007
    - 2008
    - 2009
    - 2010
    - 2006 Sprint
    - 2008 Sprint
    - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2009 (Sprint) Contest Entrants

Page 6

Lem Morrison, BCB Class (regen)

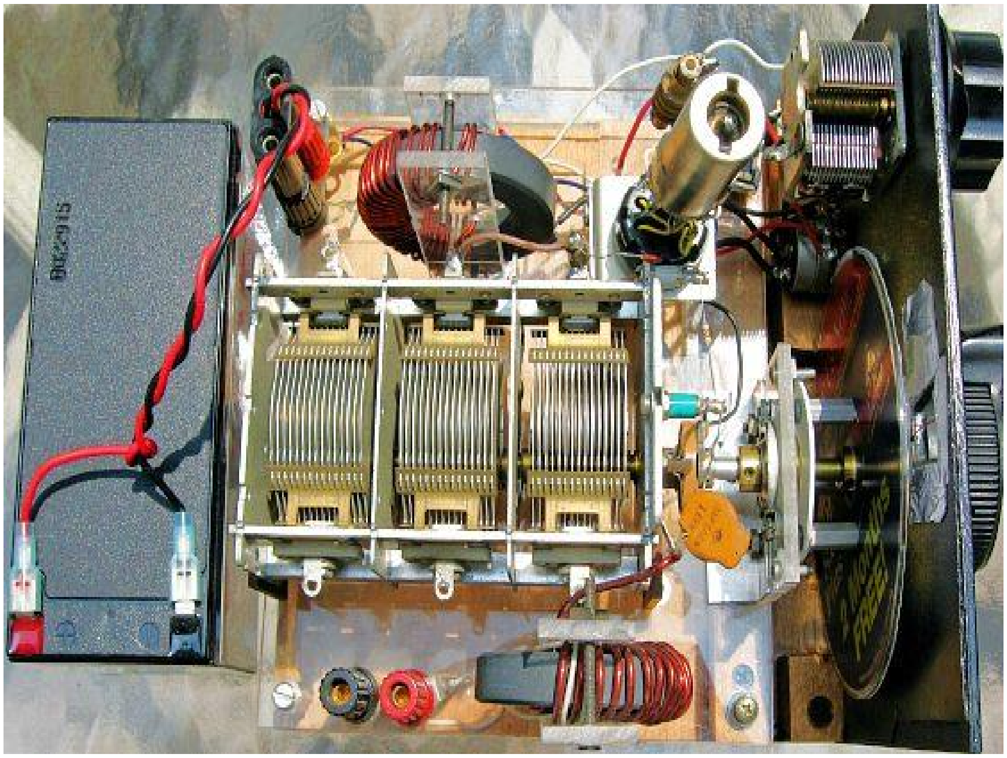

My receiver was a homebrew Regen using 12DZ6 in double-tuned circuit, coax-fed T antenna.

N4AHJ ONE-TUBE (12DZ6) REGEN SET

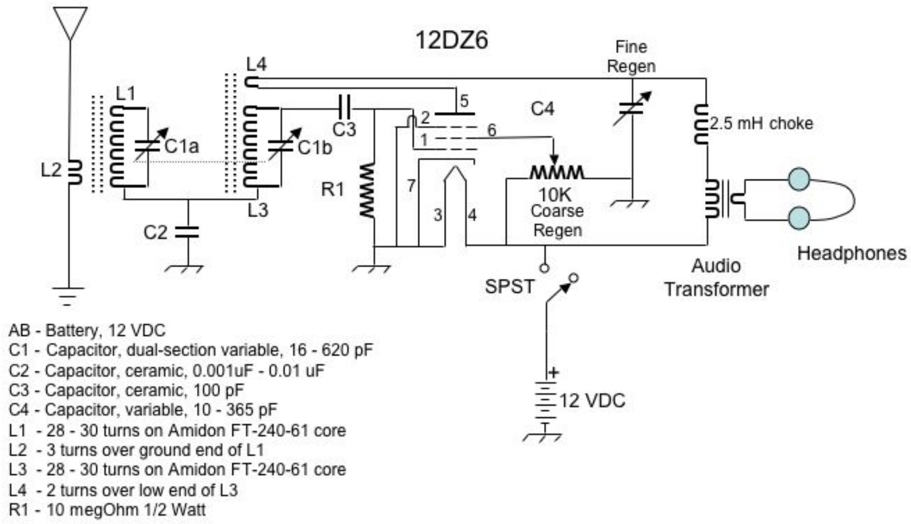
12DZ6
L4
L1
C1a
C1b
C3
L2
L3
C2
R1
5
6
2
1
7
3
4
C4
Fine
Regen
2.5 mH choke
10K
Coarse
Regen
SPST
Audio
Transformer
Headphones
+
12 VDC

AB - Battery, 12 VDC
C1 - Capacitor, dual-section variable, 16 - 620 pF
C2 - Capacitor, ceramic, 0.001uF - 0.01 uF
C3 - Capacitor, ceramic, 100 pF
C4 - Capacitor, variable, 10 - 365 pF
L1 - 28 - 30 turns on Amidon FT-240-61 core
L2 - 3 turns over ground end of L1
L3 - 28 - 30 turns on Amidon FT-240-61 core
L4 - 2 turns over low end of L3
R1 - 10 megOhm 1/2 Watt

Mike Tuggle, BCB Class (crystal)

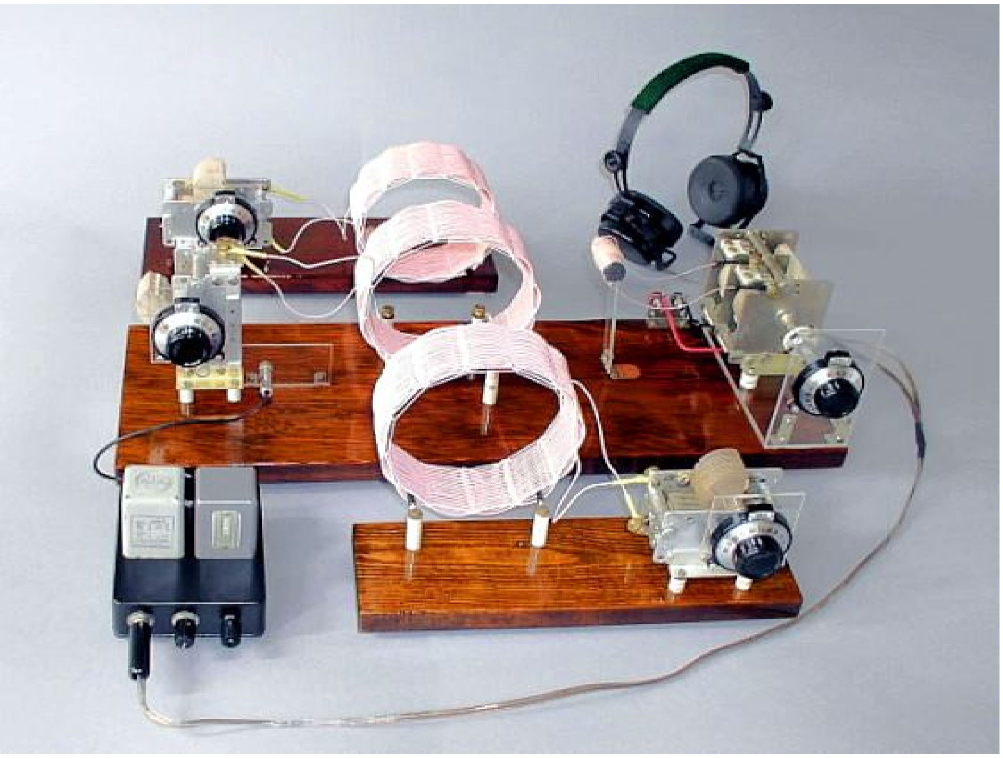

Lyonodyne-17, double-tuned, 12101 3RT detector, two wave traps, RCA Big Cans

---

Mike Tuggle, Below BCB Class (regen)

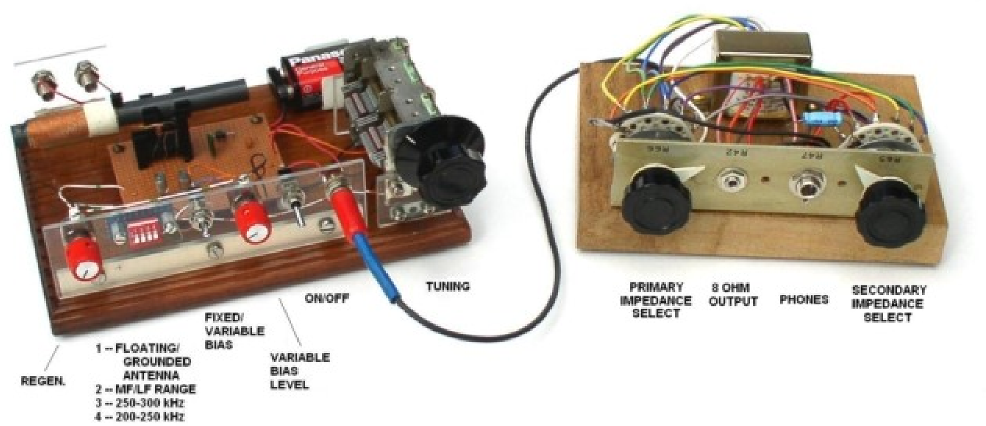
REGEN.
1 -- FLOATING/ GROUNDED ANTENNA
2 -- MF/LF RANGE
3 -- 250-300 kHz
4 -- 200-250 kHz
FIXED/ VARIABLE BIAS
ON/OFF
VARIABLE BIAS LEVEL
TUNING
PRIMARY IMPEDANCE SELECT
8 OHM OUTPUT
PHONES
SECONDARY IMPEDANCE SELECT

Set is same as used in 2009 1AD Contest.

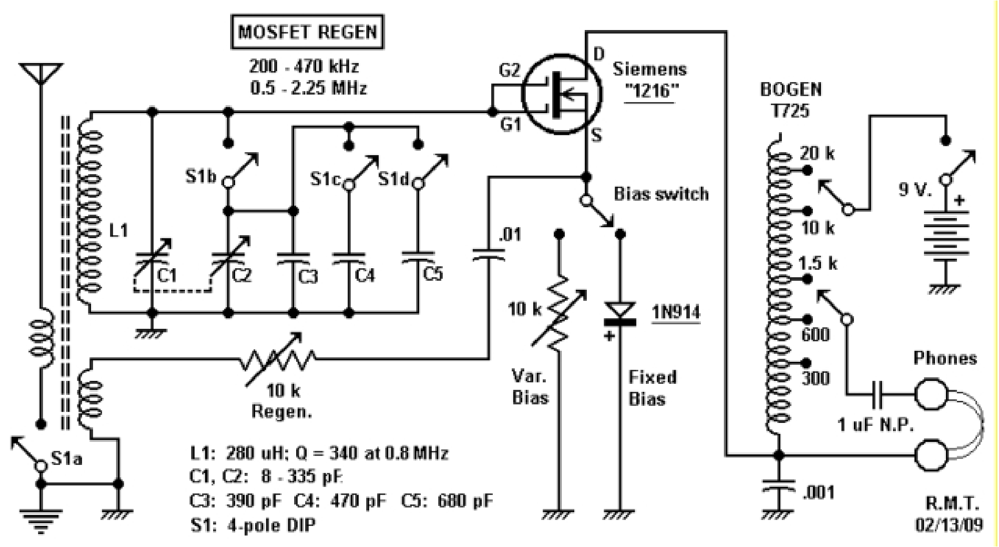
MOSFET REGEN
200 - 470 kHz
0.5 - 2.25 MHz
G2
G1
D
S
Siemens "1216"
L1
S1b
S1c
S1d
C1
C2
C3
C4
C5
.01
Bias switch
10 k
Var. Bias
1N914
Fixed Bias
10 k Regen.
S1a
BOGEN T725
20 k
10 k
1.5 k
600
300
9 V.
Phones
1 uF N.P.
.001
L1: 280 uH; Q = 340 at 0.8 MHz
C1, C2: 8 - 335 pF
C3: 390 pF C4: 470 pF C5: 680 pF
S1: 4-pole DIP
R.M.T. 02/13/09

Jack Ivey, BCB Class (crystal)

Equipment is a dual-tuned radio with spider coils of 175/46, Ge diode detector, Bogen matching transformer, modified balanced armature headphones, with a 250 foot C-shaped antenna about 8 feet above ground, #14 wire.

Kevin Norton, BCB Class (crystal)

Receiver was the "Crystal Commando". This set is a three coil Tuggle set w/cotton basket weaves. The det coil has ceramic var cap, all others are bakelite. Det. was FO215, w/ series crystal earphones. This set played VERY sharp at times . It can tune the upper part of the band in a wiper less configuration. Despite it's simple design I think the reason it worked so well was due to care in construction (i.e., component spacing and insulation). I wanted to take a step back from big litz and just see how well a simple set could be made to work.

Overall I quite happy with this set.

Ralf 'Brösel' Siemieniec, BCB Class (crystal)

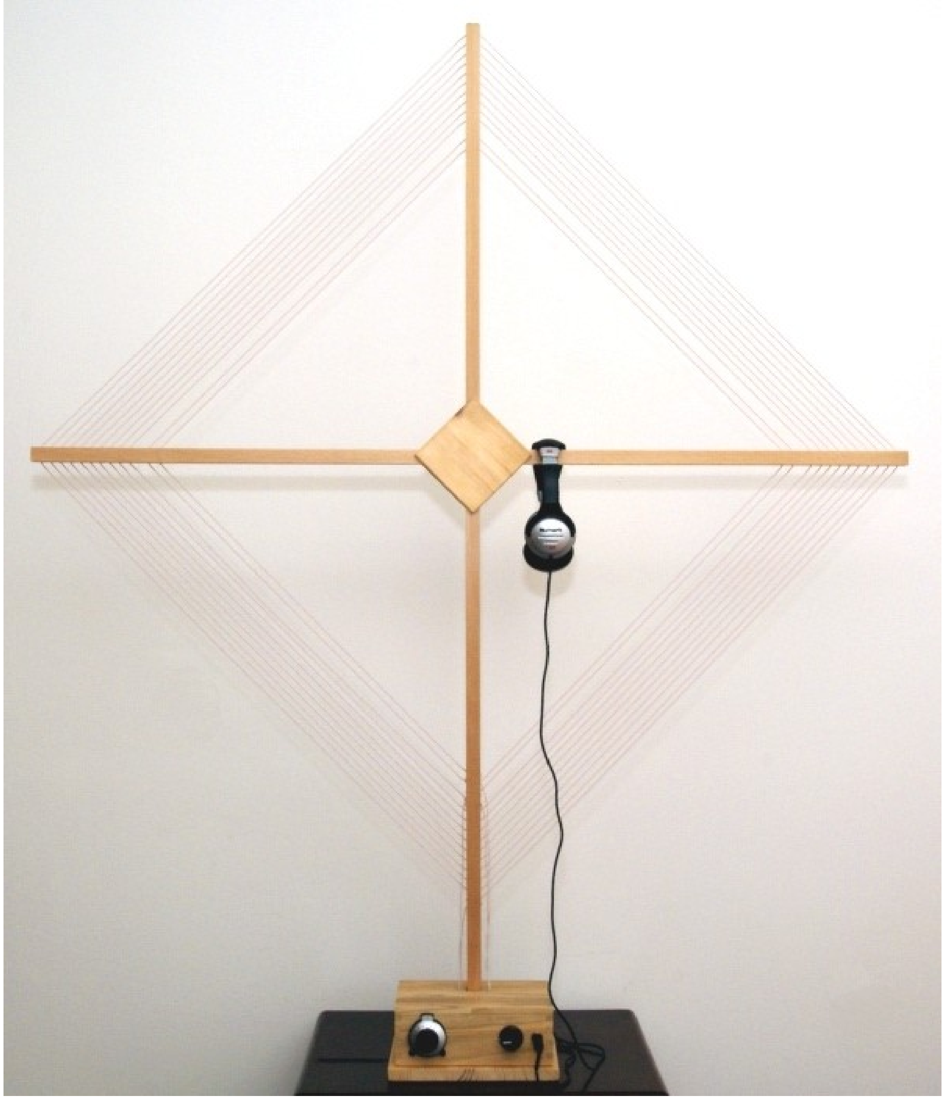

For this years sprint contest I was curious about how my reworked loop antenna crystal radio performs. For this loop having a spiral loop design, I increased the size to 0.9m x 0.9m. Since the litz wire used before (150 x 0.04 mm) was not that much rugged, I switched to a double silk-covered 270 x 0.071 mm litz wire with an outside diameter of about 2mm. The air variable is a good quality three-ganged device with ceramic insulators, but just one gang is used. Tuning is done with help of a 9:1 vernier dial which is coupled by a plastic extension to the shaft of the variable cap. The air variable cap is isolated against the wooden box while the box itself is mounted head-down on a wooden base. The loop construction is also attached to the base although mounting all things together is a bit more difficult this way.

In difference to all loop antenna sets I know so far I decided to use one of the zero-voltage MOSFETs (ALD110900). The gate is directly connected to the hot end of the tank circuit; the only problem left was how to feed the source. There are basically two possibilities – connect the source via a capacitor or via a coupling coil. After making a simple spice model of the measured characteristics of the MOSFET and doing a number of Pspice simulations I found it is more advantageous to use a coupling coil. After finishing the radio I found that there were some annoying FM ghosts which are probably caught by the coupling loop. To get rid of these I added the FM choke. Matching to the headphones is done by a “Ueberformer” from Dave Schmarder. In the contest, I used a pair of sound powered phones with USIC elements from Dynalec (note that the photo shows my “self-made” phones with rocking armature elements made by Racal. The phones with USIC elements are a bit more sensitive, but the shown phones are very lightweight and comfortable).

The loop does perform very well and is surprisingly sensitive. Selectivity is quite good for a single-tuned unit, especially at the low end of the band. Tuning is rather sharp, therefore the vernier dial seems to be a “must have”. What I like most is the clean sound. I found no distortions so far even in case of strong signals.

Here is a Google map link that you might enjoy viewing.

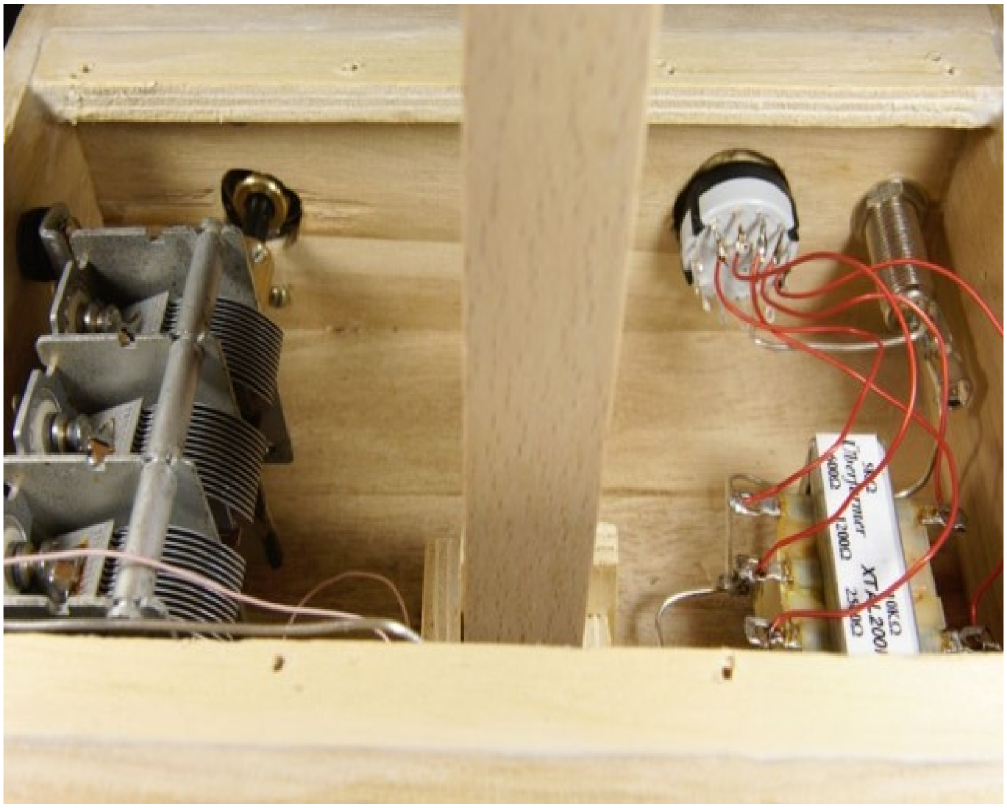

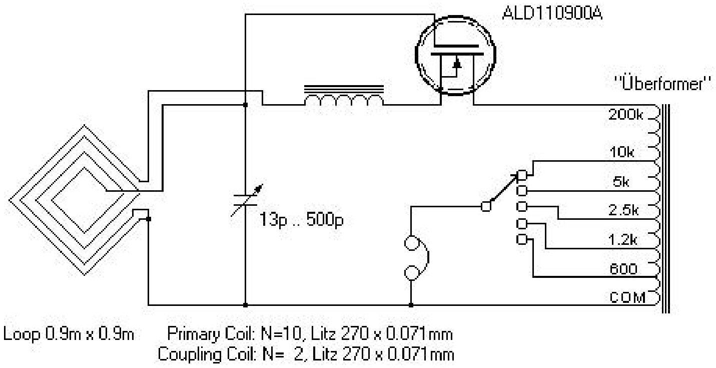
ALD110900A
"Überformer"
200k
10k
5k
2.5k
1.2k
600
COM
13p .. 500p
Loop 0.9m x 0.9m
Primary Coil: N=10, Litz 270 x 0.071mm
Coupling Coil: N= 2, Litz 270 x 0.071mm

>

Robert Weaver, BCB Class (superhet)

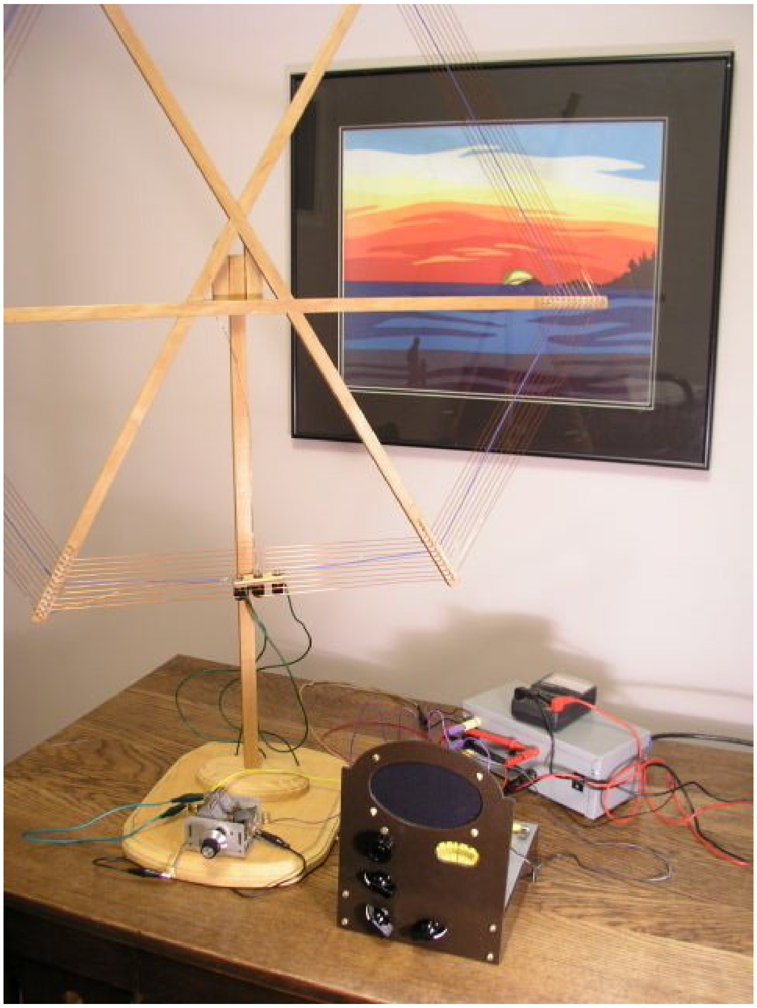

-6ME8, 1AD superhet
-30 inch diameter Tuned loop
-Used in the 2009 1AD contest (more info is shown in that section of the website).

ONE TUBE DOUBLE REFLEXED SUPERHET RECEIVER

R. Weaver, January 2009

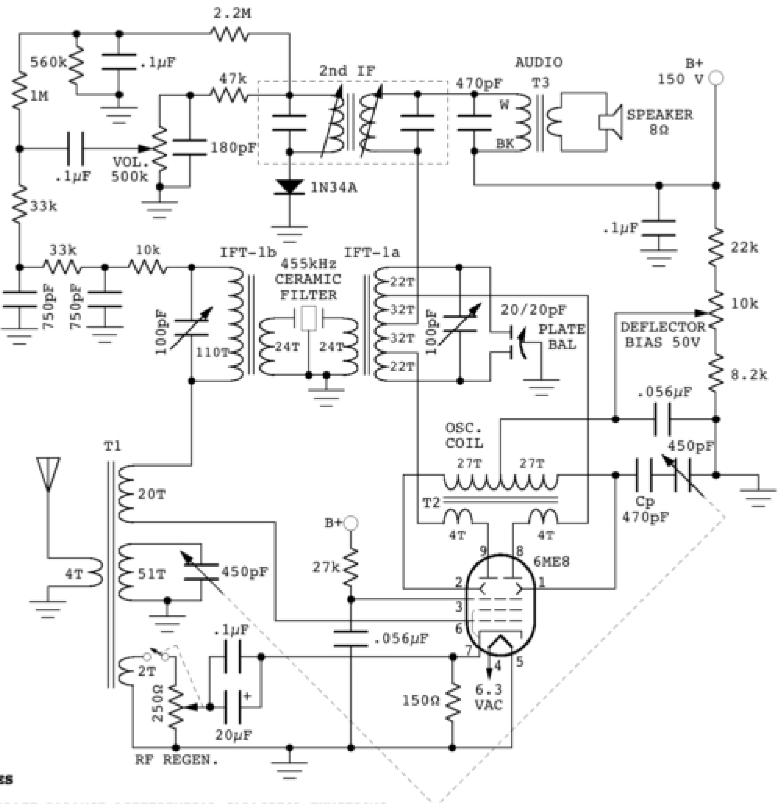
2.2M
560k
.1μF
1M
47k
2nd IF
470pF
AUDIO
T3
W
BK
SPEAKER
8Ω
B+
150 V
.1μF
VOL.
500k
180pF
1N34A
.1μF
33k
22k
33k
10k
IFT-1b
455kHz
CERAMIC
FILTER
IFT-1a
750pF
750pF
100pF
110T
24T
24T
22T
32T
32T
22T
100pF
20/20pF
PLATE
BAL
10k
DEFLECTOR
BIAS 50V
8.2k
.056μF
OSC.
COIL
27T
27T
450pF
T1
T2
Cp
470pF
20T
4T
4T
B+
27k
9
8
6ME8
4T
51T
450pF
2
1
3
6
.1μF
.056μF
7
4
5
2T
250Ω
150Ω
6.3
VAC
20μF
RF REGEN.

NOTES

1. PLATE BALANCE DIFFERENTIAL CAPACITOR FUNCTIONS AS REGEN CONTROL FOR IF.
2. BOTH SECTIONS OF THE TUNING CAPACITOR ARE EQUIPPED WITH TRIMMERS, BUT ARE NOT SHOWN ON THE SCHEMATIC. TRIMMERS ARE ADJUSTED FOR BEST TRACKING;
3. Cp IS PADDER CAPACITOR. 470pF IS OPTIMUM VALUE FOR 540-1700 KHZ TUNING RANGE USING 450PF VARIABLE CAPACITOR. IF TRACKING IS UNSATISFACTORY WITH FIXED VALUE, AN ADJUSTABLE TYPE MAY BE SUBSTITUTED.
4. DEFLECTOR BIAS IS ADJUSTED FOR MAXIMUM OSCILLATOR AMPLITUDE (MEASURED WITH SCOPE AT 6ME8 PIN 2) WITH TUNING CAPACITOR SET TO MINIMUM FREQUENCY.

INDUCTORS

T1: FT-82-61 FERRITE TOROID WITH TURNS AS INDICATED (#28 AWG ENAMELED WIRE), 2 TURN TICKLER IS WOUND ON TOP OF THE GROUNDED END OF THE 51 TURN TANK WINDING;

T2, IFT-1a, IFT-1b: FT-114A-61 FERRITE TOROID WITH TURNS AS INDICATED (#28 AWG ENAMELED WIRE);

IFT-2: GENERIC 455 kHz 2nd IF TRANSFORMER;

T3: BOGEN T725, 70V LINE AUDIO MATCHING TRANSFORMER.

Curtis Gamble, BCB Class (crystal)

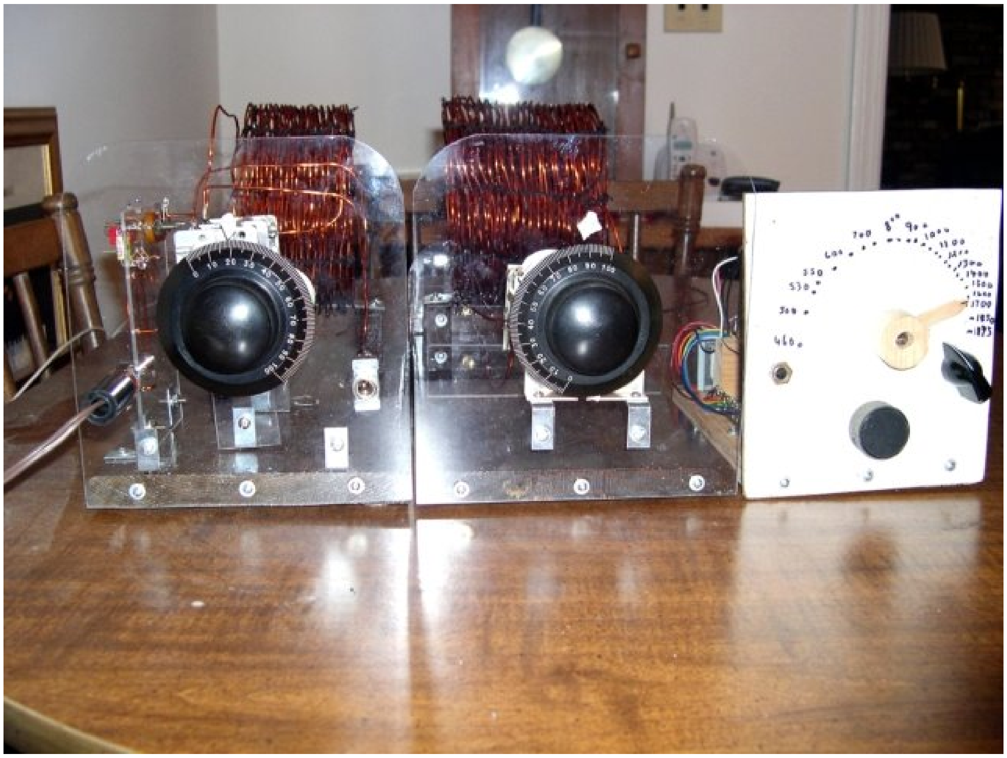

I just used the two modules with the plexi-glass fronts and no traps.

Back to the main index page.

Mini A8 Real Time Car GPS Tracker...

$9.67 Specifications: This GPS tracker applies the newest technology with the following...

ebay

Ad

Select Menu

- Home
- About Us
- Contest Results
- 1AD Contests
- 1AD Entries
- Antenna Matching
- Antennas
- BHAM Radios
- Crystal Contests
- Crystal Radio Entries
- Sprint Contest
- Headphones
- Projects, Mods
- Favorite Links
- Contact Jack

Sub Pages

- Main Page
- 2004 Page 1
- 2004 Page 2
- 2004 Page 3
- 2005 Page 1
- 2005 Page 2
- 2006 Page 1
- 2006 Page 2

- Contest Results
    - 2004
    - 2005
    - 2006
    - 2007
    - 2008
    - 2009
    - 2010
    - 2006 Sprint
    - 2008 Sprint
    - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2008 Contest Entrants

Page 4

Evan Haydon

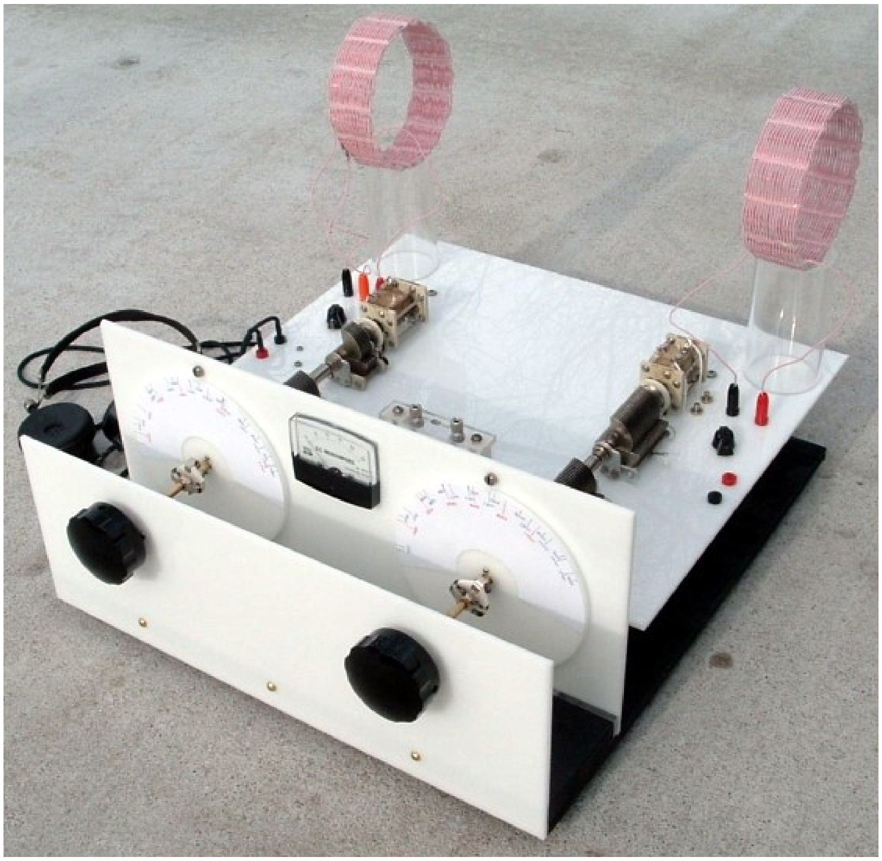

Click here to see Evan's set that he also used in the January 2008 Crystal Set DX Contest.

Ralf 'Brösel' Siemieniec

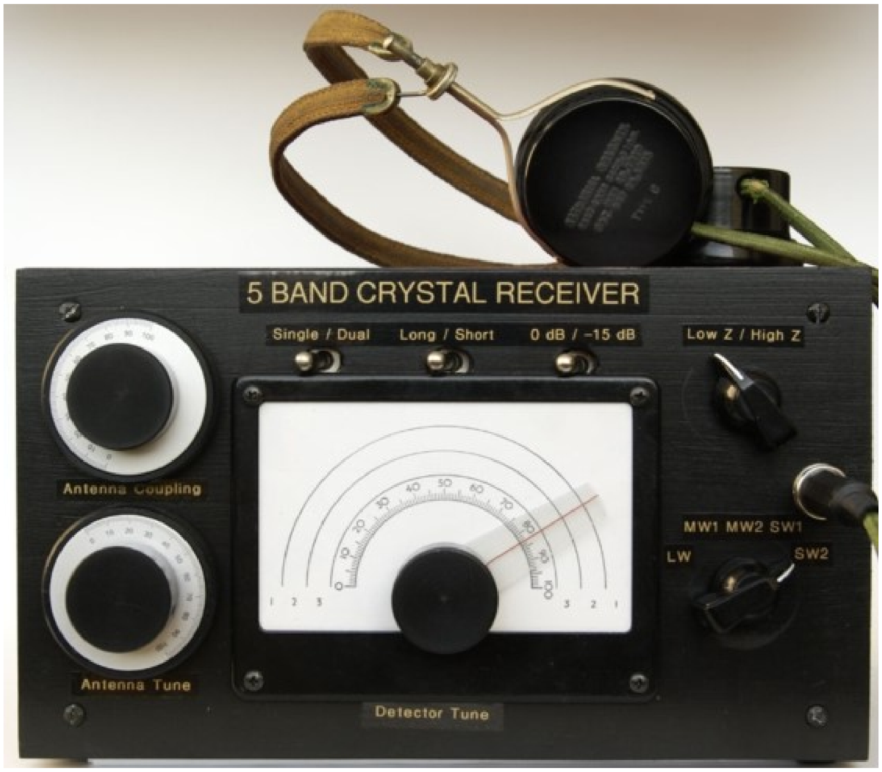
5 BAND CRYSTAL RECEIVER
Single / Dual
Long / Short
0 dB / –15 dB
Low Z / High Z
Antenna Coupling
Antenna Tune
Detector Tune
MW1 MW2 SW1
LW
SW2

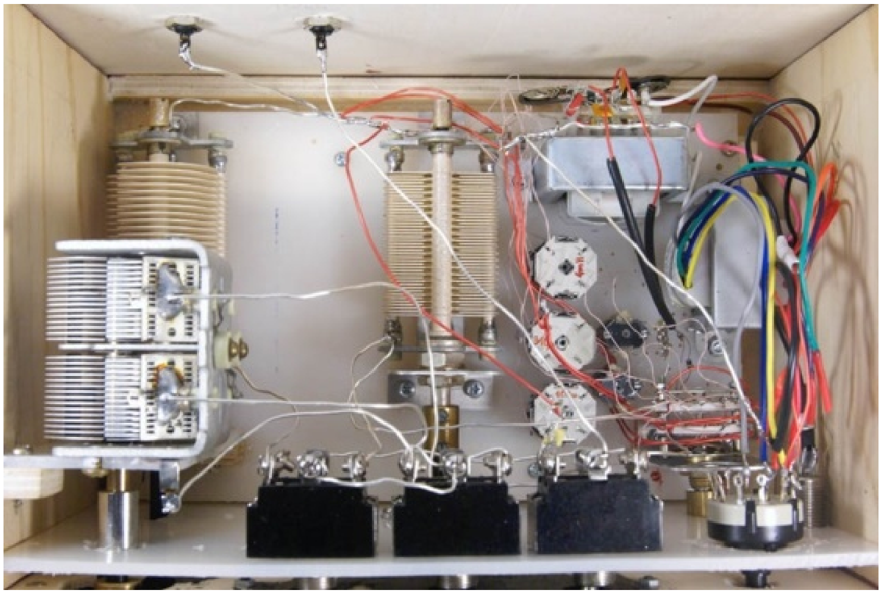

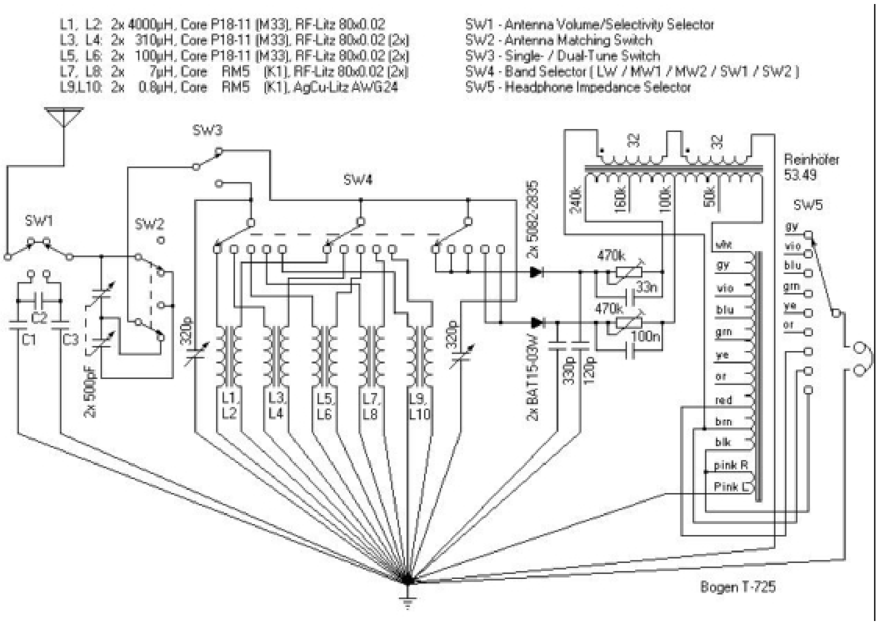
L1, L2: 2x 4000µH, Core P18-11 (M33), RF-Litz 80x0.02
L3, L4: 2x 310µH, Core P18-11 (M33), RF-Litz 80x0.02 (2x)
L5, L6: 2x 100µH, Core P18-11 (M33), RF-Litz 80x0.02 (2x)
L7, L8: 2x 7µH, Core RM5 (K1), RF-Litz 80x0.02 (2x)
L9,L10: 2x 0.8µH, Core RM5 (K1), AgCu-Litz AWG24
SW1 - Antenna Volume/Selectivity Selector
SW2 - Antenna Matching Switch
SW3 - Single- / Dual-Tune Switch
SW4 - Band Selector (LW / MW1 / MW2 / SW1 / SW2)
SW5 - Headphone Impedance Selector
SW3
SW4
SW1
SW2
C1
C2
C3
2x 500pF
320p
L1, L2
L3, L4
L5, L6
L7, L8
L9, L10
320p
2x 5082-2835
2x BAT15-03W
330p
120p
470k
33n
470k
100n
32
32
240k
160k
100k
50k
Reinhöfer 53.49
SW5
wht
gy
vio
blu
grn
ye
or
red
brn
blk
pink R
Pink L
gy
vio
blu
grn
ye
or
Bogen T-725

My crystal radio tunes the mediumwave (broadcast) band as well as the shortwave band, both are split into two subbands, and the longwave band. To keep the receiver small, I was using closed ferrite cores from Epcos, P18-11-M33 for the MW & LW bands and RM5-K1 for the SW bands. The coils are all done using litz wires 80 x 0.02 mm (two wires in parallel for MW and SW1) and two paralled litz wires of silver-plated copper strands for SW2. Actually, this is the same wire as I used for the wiring, the insulation of the silver-plated litz wire is made of Teflon, which should not cause too much losses (as I hope). The antenna and detector coils are each wound on a separate core which were later stacked together.

The radio has two tuned circuits and a variable antenna coupling and can be switched between single- and double-tuned mode. Additionally, there is an attenuator in case of need, the variable antenna coupling capacitor is a two-ganged type and can be switched in series or parallel. The antenna and detector tuning caps are both silver-plated devices. Rectification is done using two paralleled 5082-2835 for the BC band and again two paralleled BAT15-03W for the SW band. The matching section consists of two transformers - a 53.49 by Reinhöfer Röhrentechnik in Germany (input impedance 240k) switched together with a Bogen T-725. I managed to get a pair of Baldwin Type C headphones which are much more sensitive than all of my modern but more comfortable types.

All is mounted at a polypropylene sheet for proper insulation. The most distant station received so far was 1419 km away in Tunesia. Could probably be more if I had another antenna than the metal made balustrade of my balcony - I live in a multi-party house, thus there is not much chance for long antenna wires.

Lem Morrison

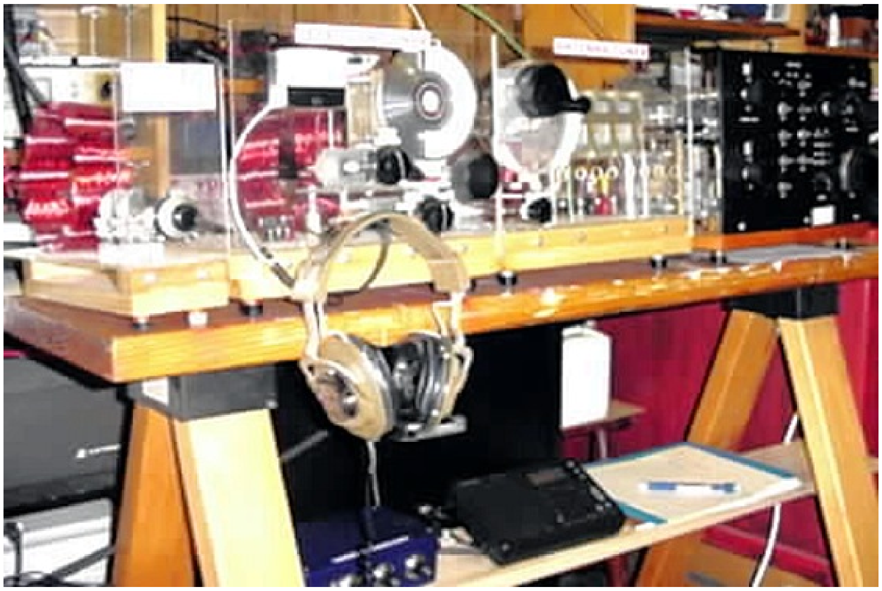

Lem used an updated version of this set that was used in the last several Crystal Set DX and Sprint Contests.

Garry Nichols

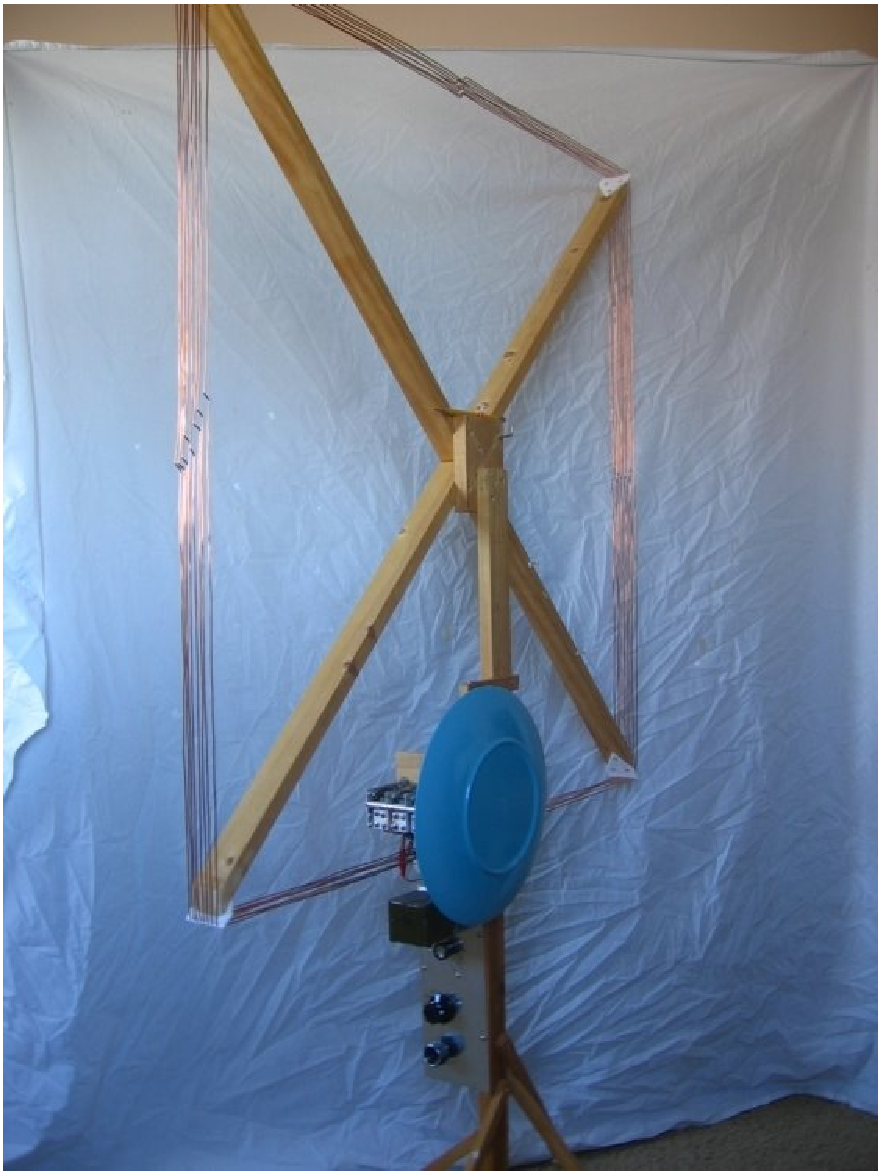

Same loop as in the photos for last winter's contest. However, I was not using the STM, only the T3/AM20 100:100k transformer and the single 107 nA Schottky. Also, I was not attaching the ground to the variable's frame and using a RFC for diode DC continuity through the variable.

The ground of the detector/audio went to the center of the loop winding > instead. I found that there was much less interference from a local FM station this way.

Click here to see more about this set that Garry also used in the January 2008 Crystal Set DX Contest.

Mike Tuggle

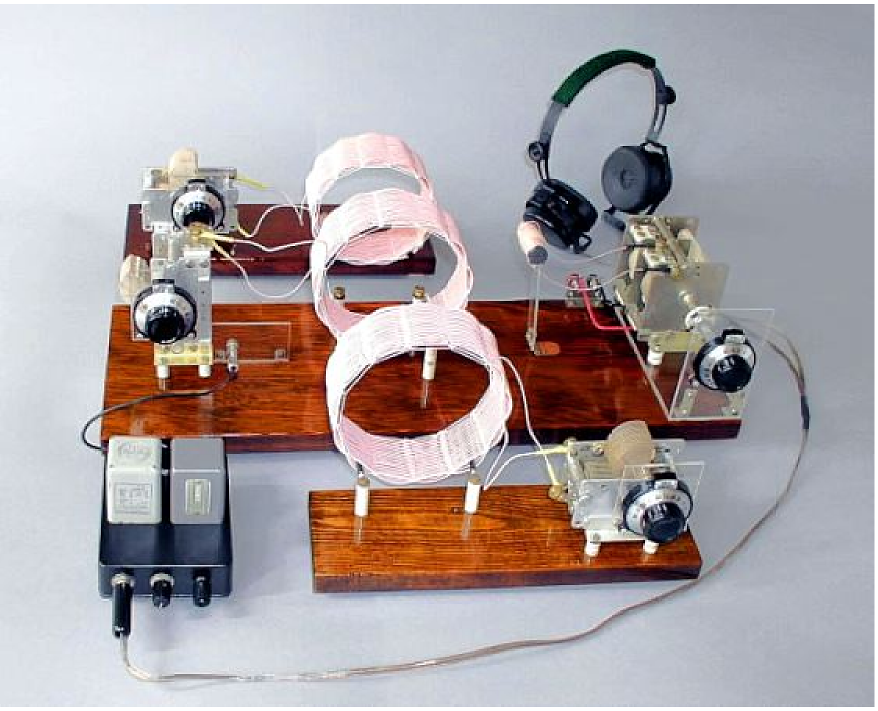

The infamous Lyonodyne-17, double-tuned, 12101 3RT detector, two wave traps, RCA Big Cans.

Dan McGillis

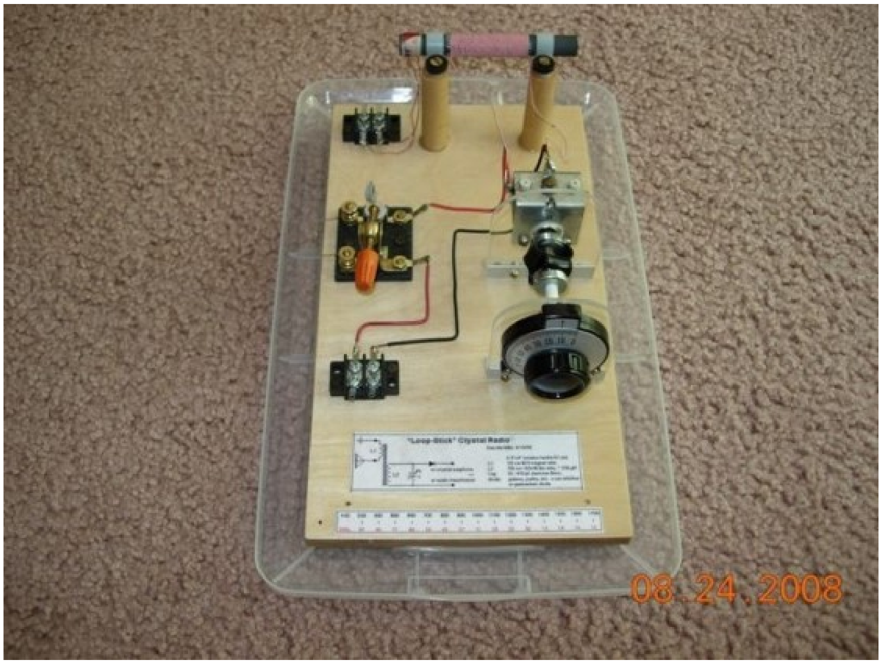

For more info, click here to see Dan's posting on Dave Schmarder's RadioBoard.

Dave Schmarder

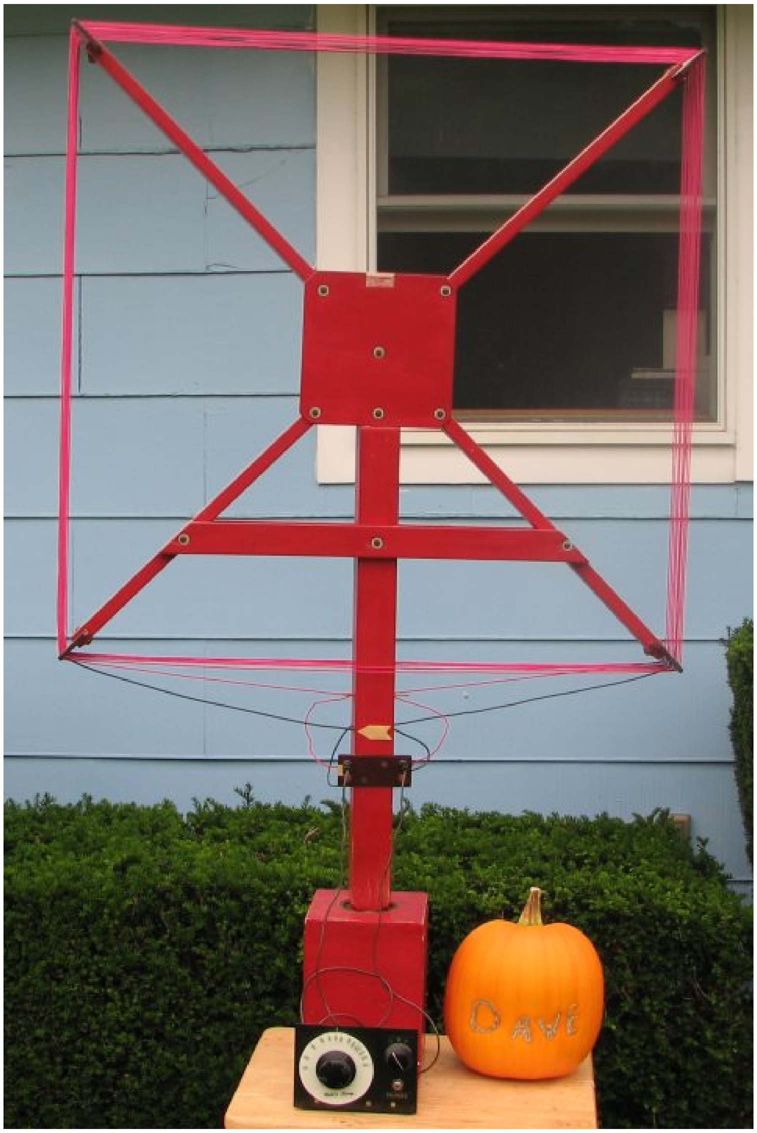

Click here to see more on Dave's Loop Set at his fine website.

Jack Bryant

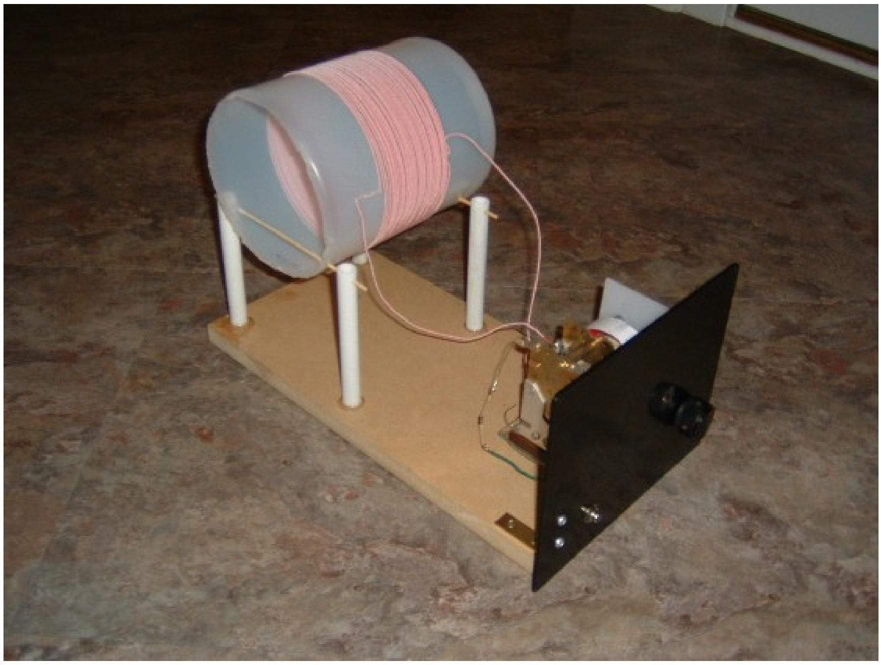

[Click here to see Jack's set that he also used in the January 2008 Crystal Set DX Contest.]

Curtis Gamble

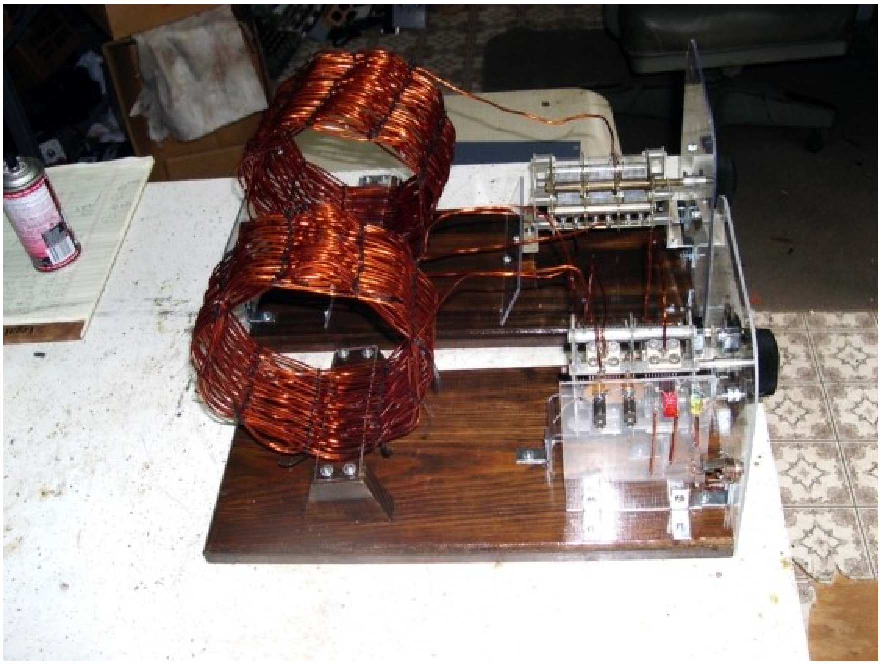

Denver Cohen

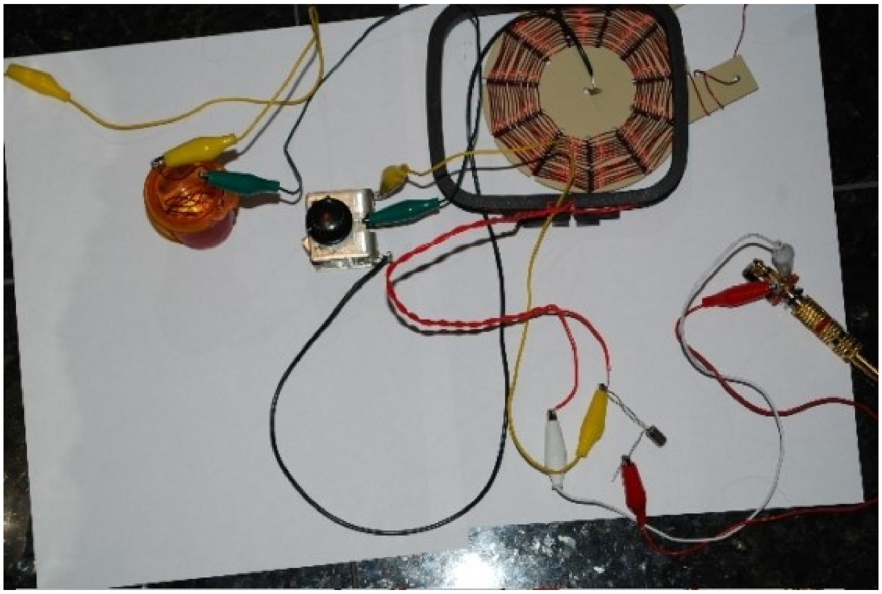

Back to the Index

Select Menu

- Home
- About Us
- Contest Results
- 1AD Contests
- 1AD Entries
- Antenna Matching
- Antennas
- BHAM Radios
- Crystal Contests
- Crystal Radio Entries
- Sprint Contest
- Headphones
- Projects, Mods
- Favorite Links
- Contact Jack

Sub Pages

- Main Page
- 2004 Page 1
- 2004 Page 2
- 2004 Page 3
- 2005 Page 1
- 2005 Page 2
- 2006 Page 1
- 2006 Page 2
- 2006 Page 3
- 2007 Page 1
- 2007 Page 2
- 2008 Page 1
- 2008 Page 2
- 2008 Page 3
- 2008 Page 4
- 2009 Page 1

- Contest Results
    - 2004
    - 2005
    - 2006
    - 2007
    - 2008
    - 2009
    - 2010
    - 2006 Sprint
    - 2008 Sprint
    - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2008 Contest Entrants

Page 2

Brent Seres

Antenna Tuning
Detector Tuning

Set: Double tuned torroid, with Tuggle front end, using FO215 and Schottky diodes. 100k to 1 k Hammond matching transformer, and sound powered phones. Coupling is accomplished with a third torroid, primary has 3 turns inseries with ground lead, secondary has 3 turns in series with main detector coil.

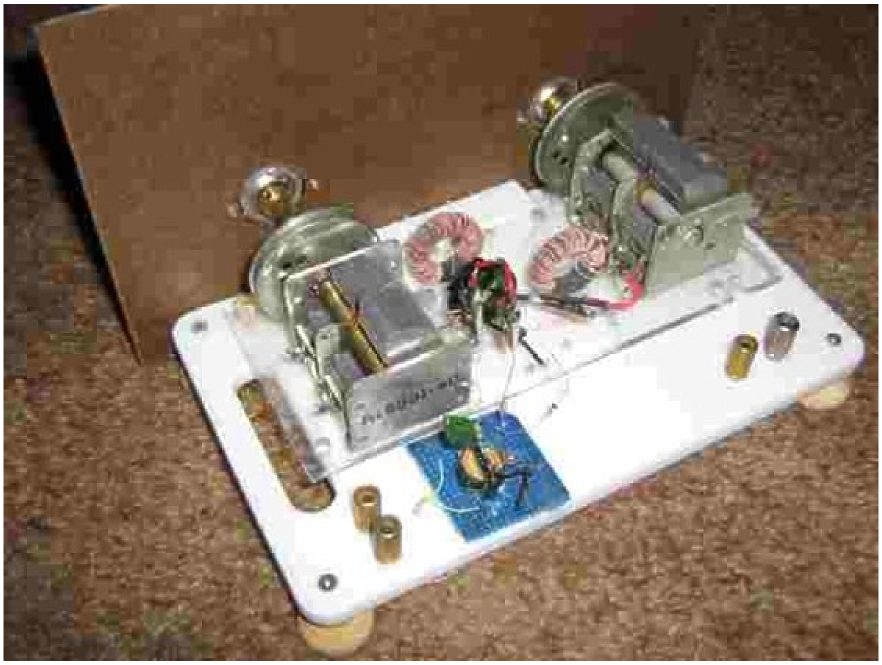

Dave Schmarder

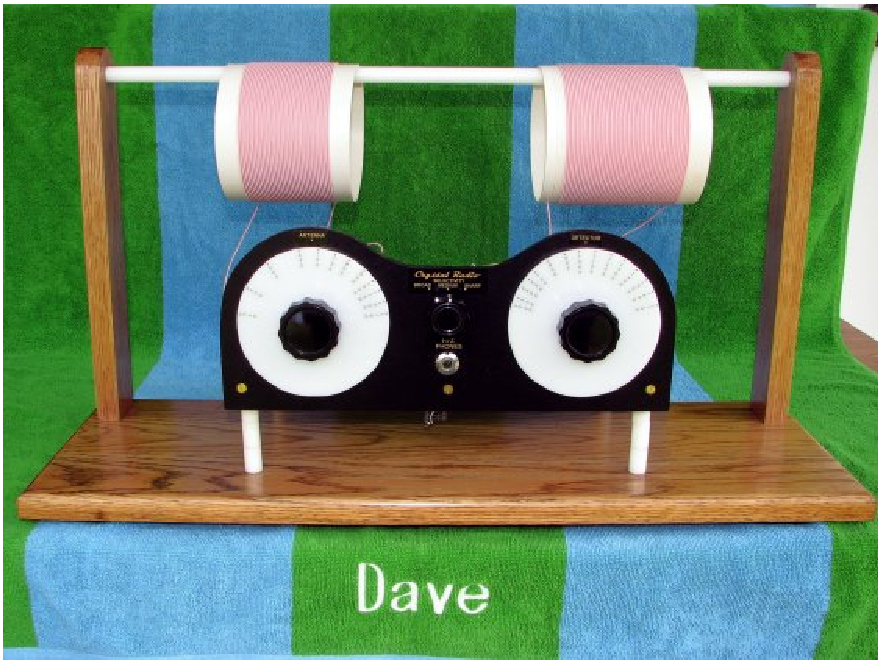
ANTENNA
DETECTOR
Crystal Radio
SELECTIVITY
BROAD MEDIUM SHARP
PHONES
Dave

For more info, visit Dave's website at: Dave Schmarder's 2008 Crystal Set Contest Entry This is his #70 Crystal Set.

Jack Hennon

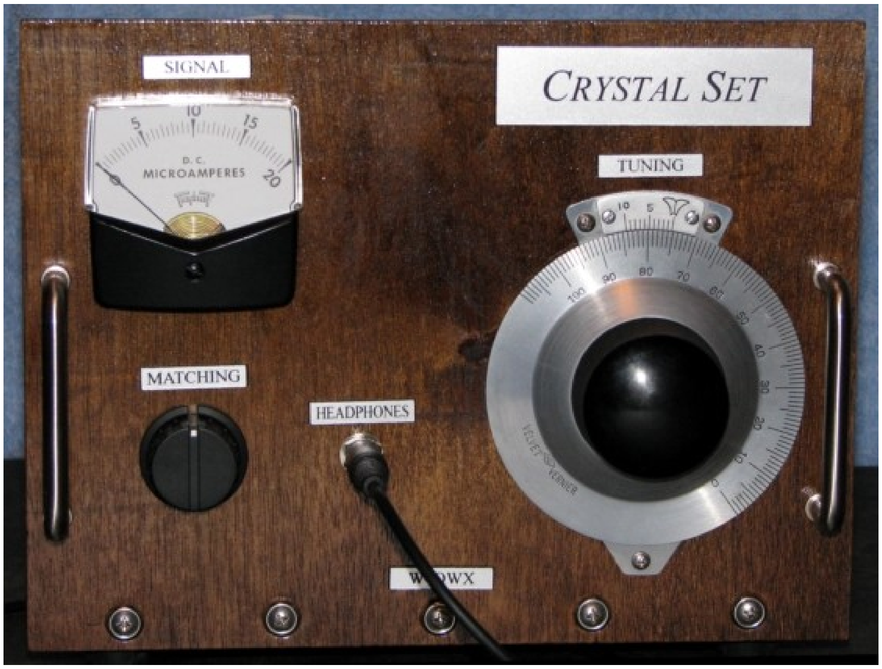
SIGNAL
CRYSTAL SET
TUNING
MATCHING
HEADPHONES
D. C. MICROAMPERES

The set uses a Tuggle Tuner, a Schottly diode detector with three diodes in parallel, and a UTC A27 matching transformer.

My antenna was a 135 dipole fed in the center with 45 feet of 450 ohm open wire line. I connected the feed line wires together at the set and worked the antenna as a T against ground.

Again the contest was fun. I did better than last year, copying 100 stations. Conditions may have been better or it could be due to my new set and more sensitive headphones.

CRYSTAL SET

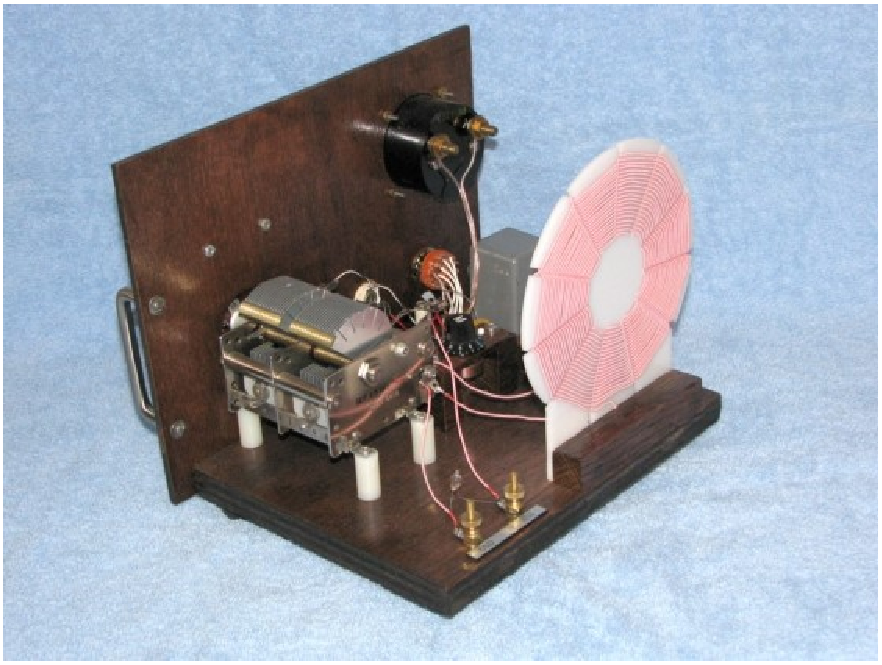

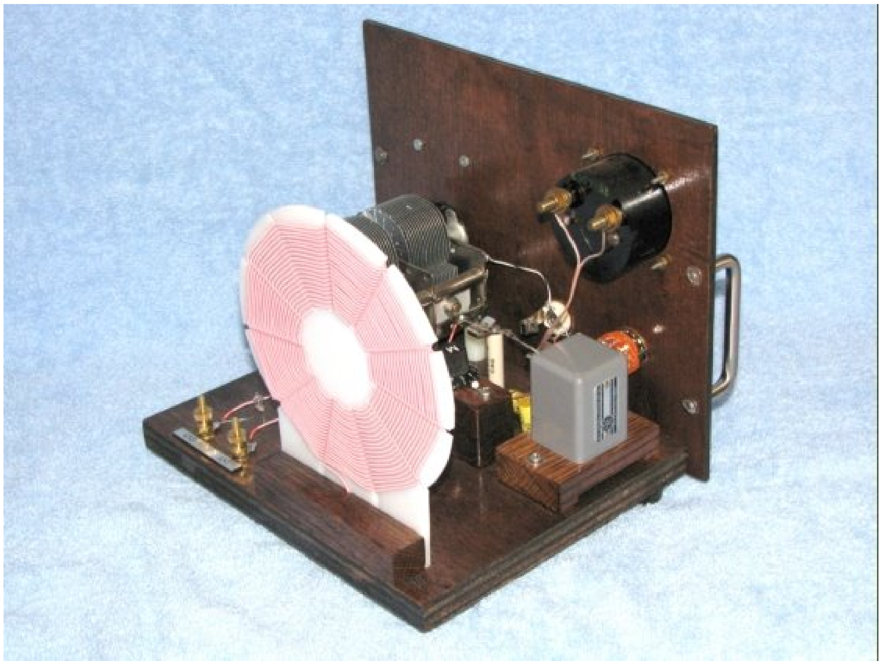

Garry Nichols

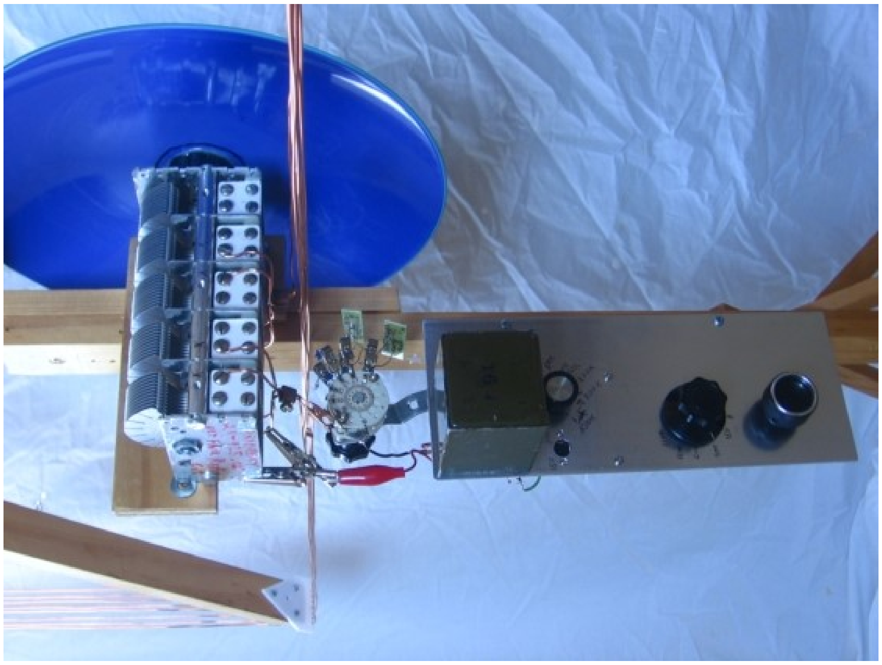

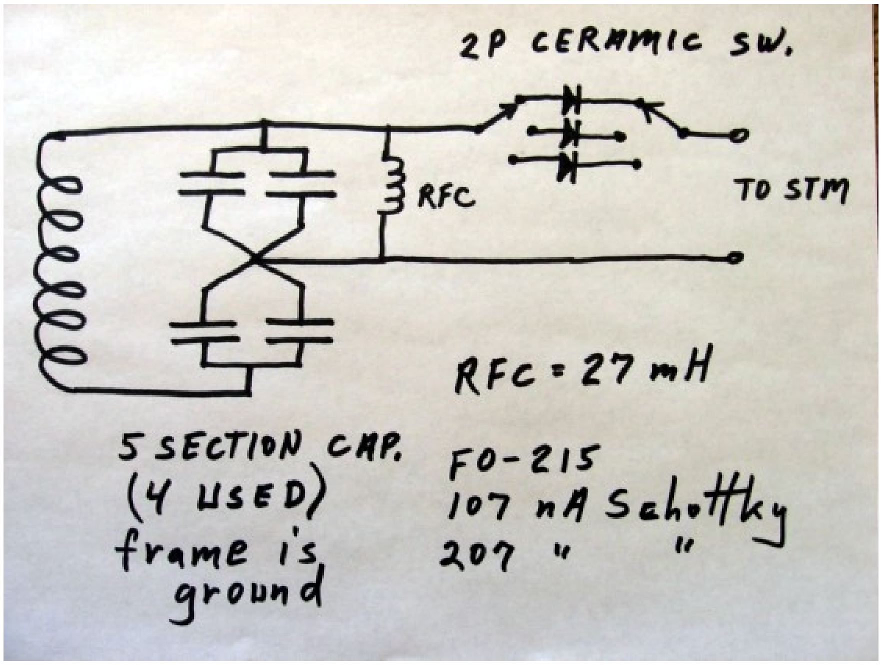

James Kretzschmar

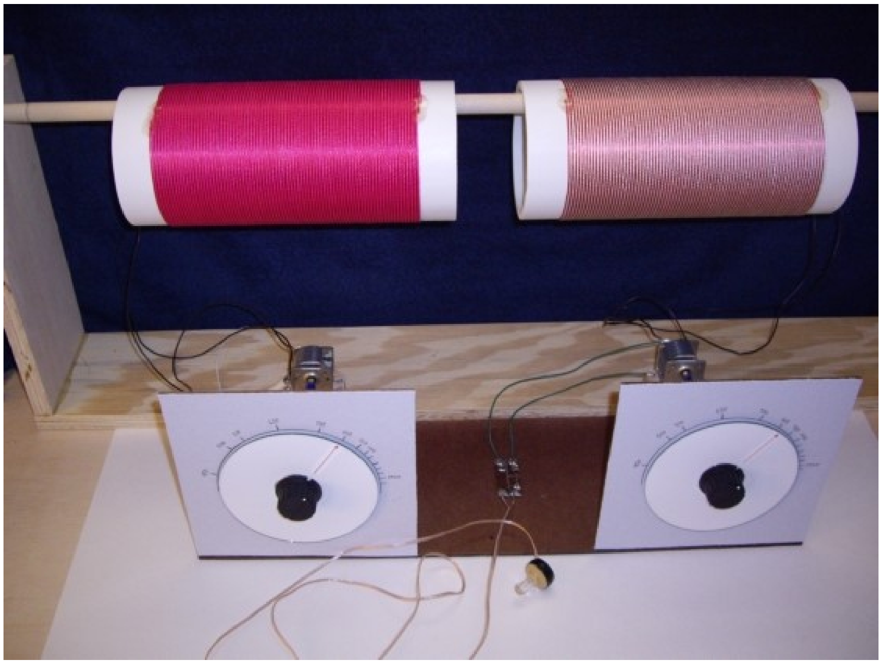

This radio is big improvement over the radio that I entered in the 2001 Crystal Radio building contest. I started tinkering in early January 2008 and finally decided on a design on 17 January 2008, and built the radio that afternoon … ready to use the next day.

My coils are wound on 4.5” PVC pipe with 80 turns #18 wire space wound the width of the wire with nylon masonery cord. The coils actually have too much inductance, however I did not have time to change them. The variable capacitors have ¼” dowel wooden extensions to avoid the effect of hand capacitance. To provide smooth operation and stability a ¼” washer is glued to the front panel.

My antenna is about 200 feet of wire strung through the trees in the backyard about 20 feet up. In operation, the right side is tuned to where you want to be, then the left side is tuned for the strongest signal, the coils are moved in/out as needed for signal strength and selectivity. I have had real good results with this design.

At my location in Winston Salem, NC we have a real strong station at 830 KHz and I can routinely hear 840 KHz WHAS in Louisville, KY in the clear. Two small stations were heard during the 2008 contest: (1) WMAC 940 KHz in Macon, GA (2) CKNX 920 KHz in Wingham, Canada. A total of 35 stations were copied. A simple piezo earphone was used. I plan to tinker with this design to optimize reception.

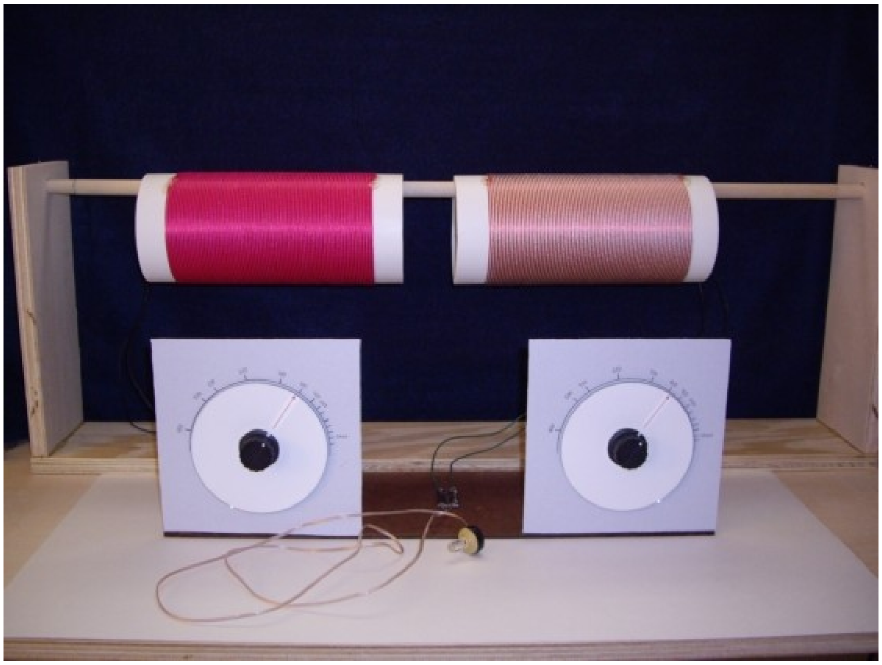

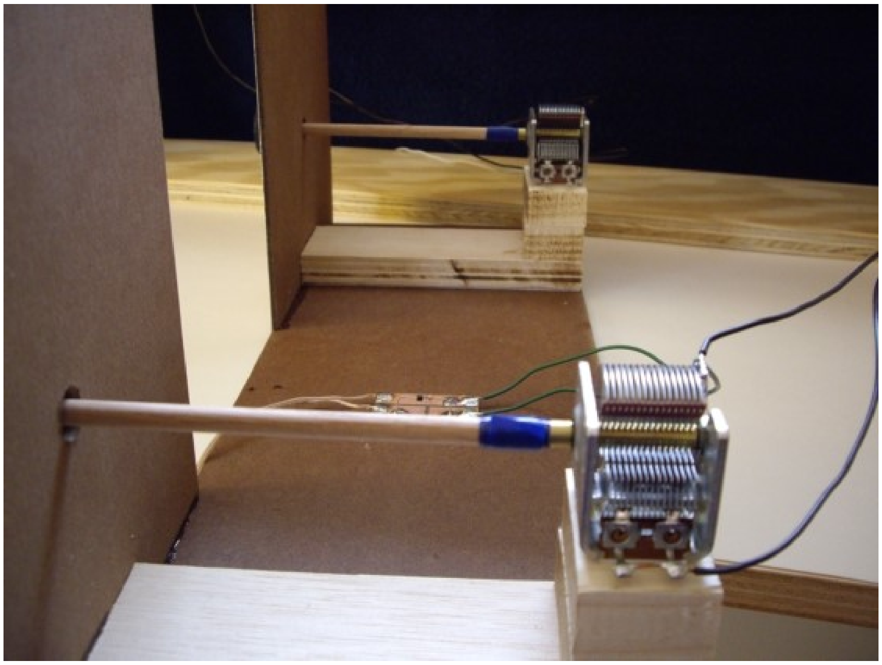

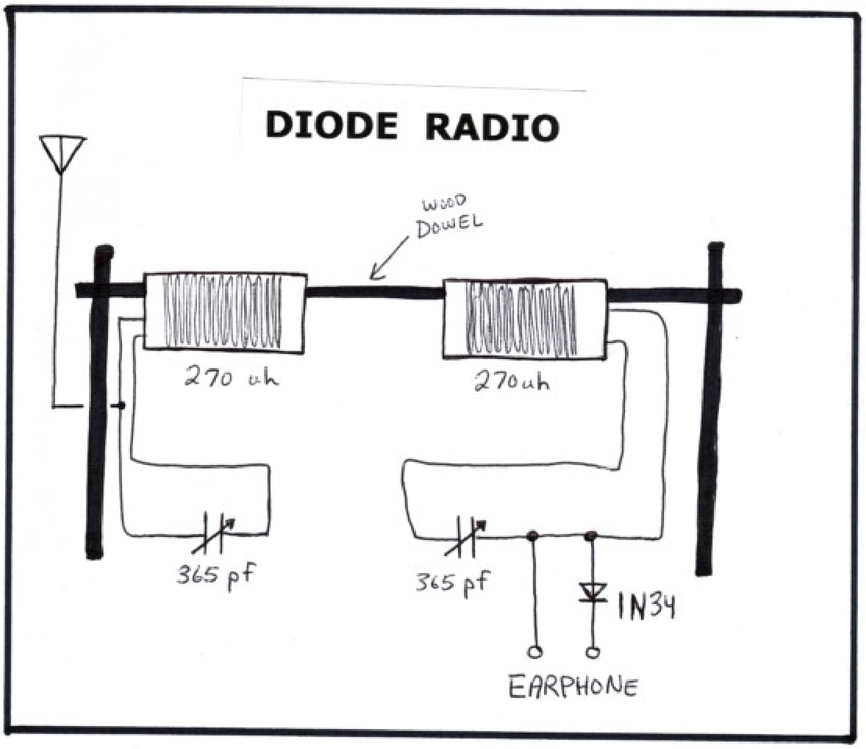

Continued on page 3.

Ad
INNERACTIVE
SAVE $500 WITH CODE RET
IES AURACLOUD®
"NOW IN 3D"
CLICK HERE

Select Menu

- Home
- About Us
- Contest Results
- 1AD Contests
- 1AD Entries
- Antenna Matching
- Antennas
- BHAM Radios
- Crystal Contests
- Crystal Radio Entries
- Sprint Contest
- Headphones
- Projects, Mods
- Favorite Links
- Contact Jack

Sub Pages

- Main Page
- 2004 Page 1
- 2004 Page 2
- 2004 Page 3
- 2005 Page 1
- 2005 Page 2
- 2006 Page 1
- 2006 Page 2
- 2006 Page 3

- Contest Results
  - 2004
  - 2005
  - 2006
  - 2007
  - 2008
  - 2009
  - 2010
  - 2006 Sprint
  - 2008 Sprint
  - 2009 Sprint
- Previous
- Next

Crystal Radios Of The 2006 Contest Entrants

Page 2

Peter Kerttula

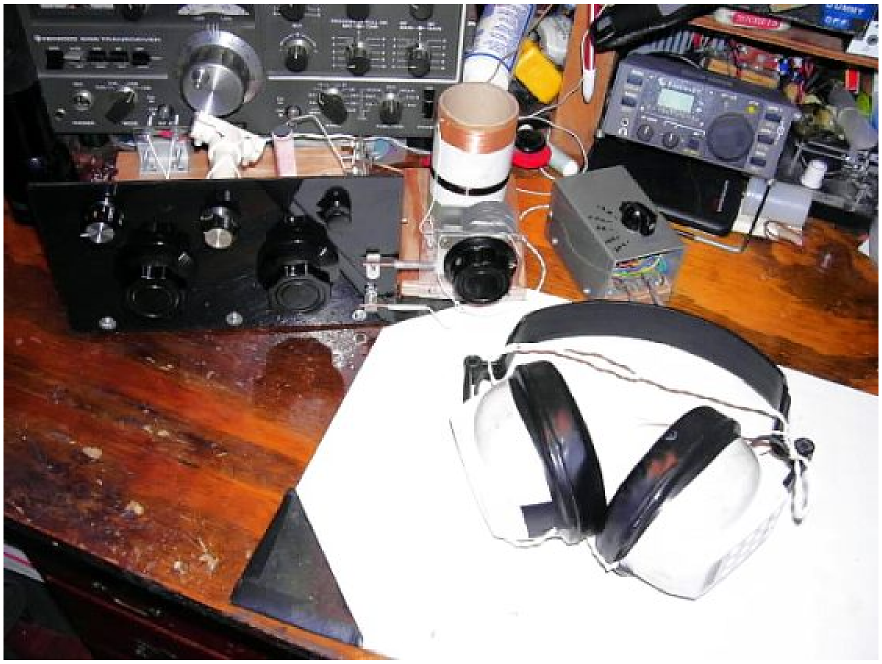

The rig for the 2006 contest is a Jim Frederick design, the Hobby Dyne 2. I added crystal switching and use a reduction on the tune side capacitor.

The headphones are sound powered made from SP mike elements. A bogen transformer handles the output transformation.

The antenna is a 100 foot wire bent to fit the lot! A cold water pipe handles the ground.

Mark Roliff

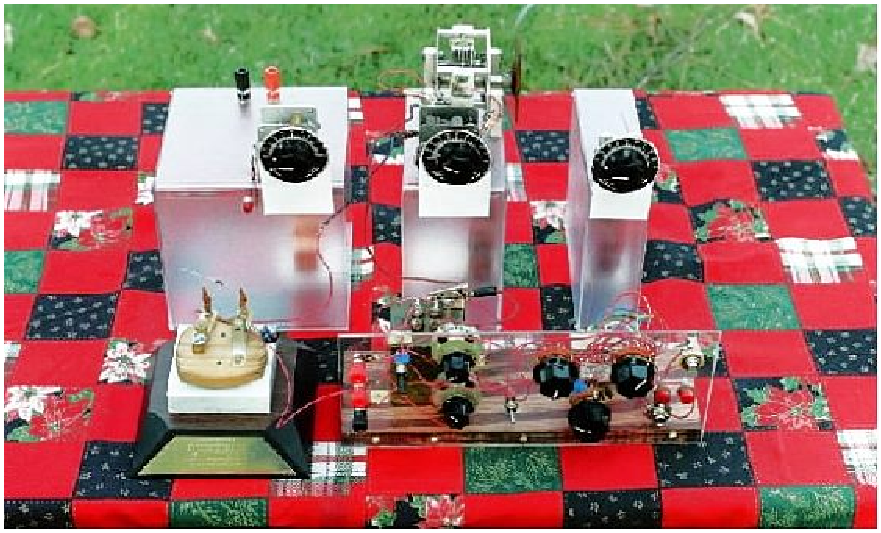

The unit at the left rear is the antenna tuner. The coil is 33 turns of #22 magnet wire. The cap. is a dual 365pf variable purchased from the XSS.

The unit at the center rear is the detector tank. Its coil is 40 turns of #24 magnet wire. The front cap. is a dual section variable salvaged from a junked radio. The smaller front section is not used. The rear cap. is used for band spread.

The unit at the right rear is the inductive trap. The coil is 40 turns of # 24 magnet wire. The cap. is a single gang 365pf purchased from the XSS.

At the right front is the detector stand mounted on a salvaged trophy plinth. The mass of the wood and marble lowers the detector stand's sensitivity to vibration. All logged stations were heard with the cat whisker and galena crystal. At the start of each listening session, I would tune in a weak local station using a germanium diode and then disconnect it and match the signal's strength and clarity to a hot spot on the galena crystal. The whole process took less than a minute. The crystal was purchased from the XSS.

The unit at the right front is my version of Steve Bringhurst's UltiMatch 2. I used a pair of RCA "Big Cans" connected to the Red tap of the Bogen transformer.

Mark Roliff

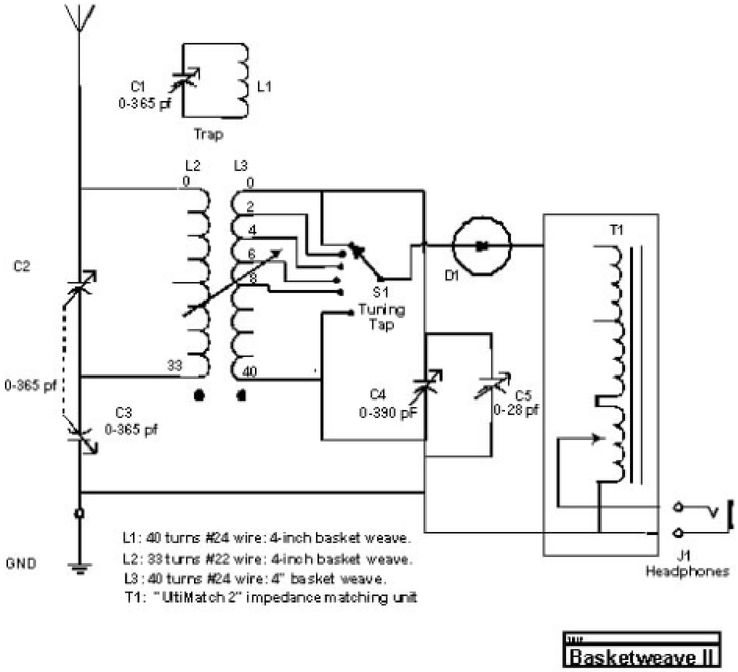
C1
0-365 pf
L1
Trap
L2
L3
0
0
2
4
6
8
33
40
C2
0-365 pf
C3
0-365 pf
S1
Tuning
Tap
D1
T1
C4
0-390 pF
C5
0-28 pf
GND
J1
Headphones
L1: 40 turns #24 wire: 4-inch basket weave.
L2: 33 turns #22 wire: 4-inch basket weave.
L3: 40 turns #24 wire: 4" basket weave.
T1: "UltiMatch 2" impedance matching unit
Basketweave II

..Note that the audio impedance matching unit is a copy of Steve Bringhurst's UltiMatch 2 … only the Bogen T725 transformer was used, and the Benny was bypassed …

Mark Hampton

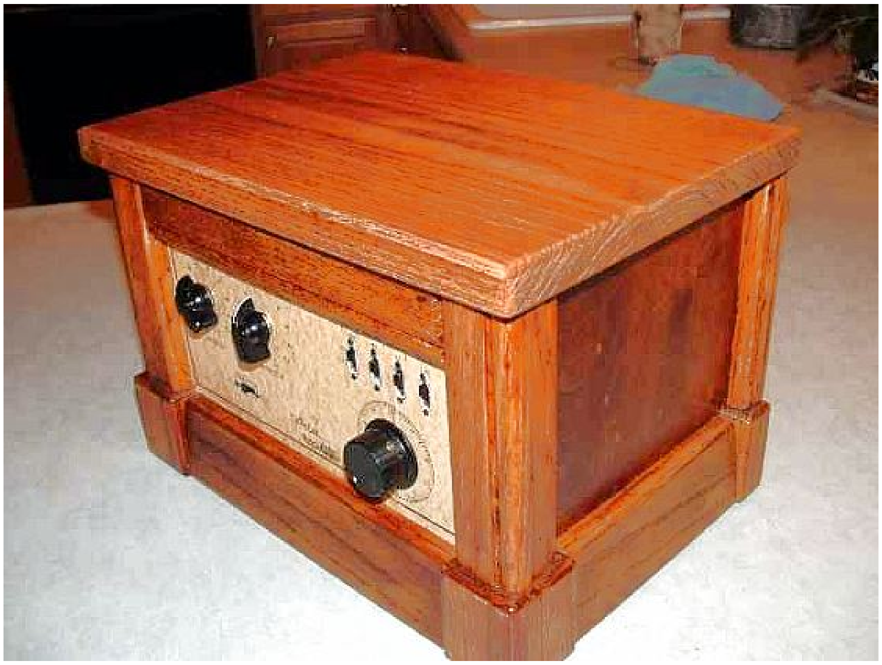

My interest in crystal radios was less than a month old when I found out about the DX contest. The result is my entry which, for reasons that will become obvious, I have dubbed the Comedyne. If you find any of this humorous rest assured that you are laughing with me and not at me.

I entered the formula for calculating coils into a spreadsheet. Later on I modified the formula so that I could enter the diameter instead of the radius and promptly forgot that I had done so. The coil form I used was 4 inch PVC that I turned down and threaded at 22 TPI. When I went to calculate the length I entered a 2 instead of 4 and came up with a length of 5 inches. This resulted in a coil of 800 uH instead of the 280 to 300 uH that I wanted. I was too green to recognize that this was not right and of course the coil was all soldered up before I found my mistake.

The one thing that went right on the coil was the winding and tapping. I had tried various methods of making taps on a coil with little success. Seems I don't have the patience, eyesight and requisite number of arms to do a decent job of it so I came up with a new (at least to me) idea. I drilled line of small holes at every 5th thread. I placed the coil form in the freezer for a while to shrink it and then quickly wound the coil. Then I scraped the enamel off each wire that passed over a hole with a sharp pointed hobby knife. I bent a small hook in the end of each tap wire passed it through the hole, turned it 90 deg and "hooked" the winding. I placed a tiny dab of flux on the wire and then folded the tag end of the hook down tight on the coil. A dab of solder finished the job. The result was a nice tight, neat (if slightly over-wound) coil. I bridged all the taps together with a loom to help take the stress off of the solder joints.

The only caps I had were some I ordered from Antique Electronics to play with because they were cheap. They are 4 section totaling 240pf and 540 deg. rotation. I found out why they are so cheap. Nobody in his right mind would want a VC designed for PC board mounting. I finally figured out a way to mount them and only destroyed one in the process. I didn't have time to order a shaft extension so I machined one out of some plastic rod.

When selecting a diode from two dozen 34As two of them accidentally touched together and the sound level really jumped up. I called them a pair and mated them forever. My antenna is an almost 300 foot wire running due north/south. It ranges from 15 to 20 feet off the ground. The ground is a piece of ½ inch black pipe driven 6 feet into the ground.

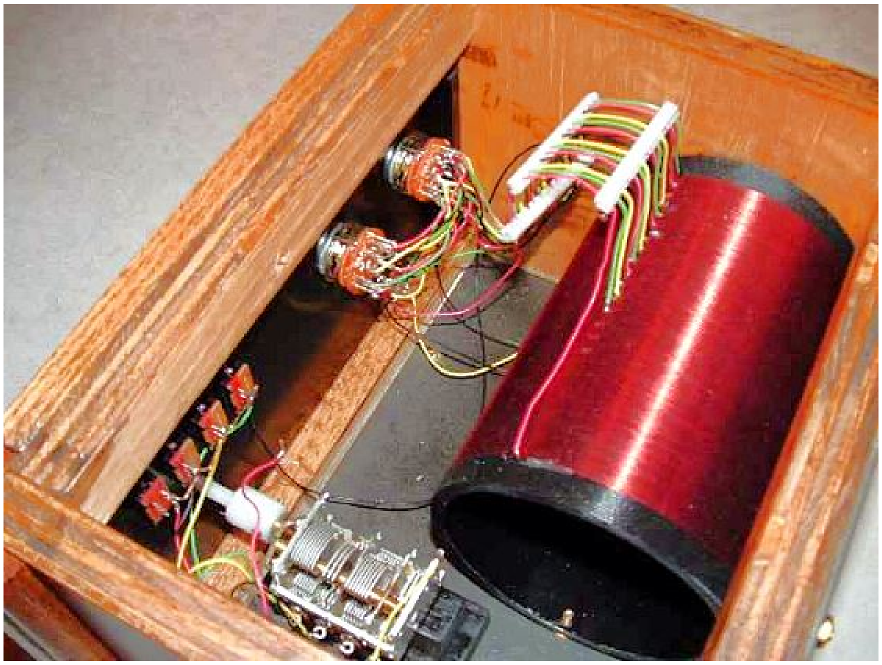

I used 12 pos. rotary switches from Antique Electronics for the antenna and tuner taps. For the other switches I used the slide type strictly for looks. (Certainly not for the "ease" of mounting!) The tuning knob and slide switches are from Radio Shack.

Four of the slide switches are for the capacity. 200pf, 30pf, and 85pf on the variable and a 150pf fixed cap. The other switches the ground to switch out half the coil. My grand plan was to be able to do some short wave tuning. (I did manage to do some.)

I'm an admirer of Art Deco and Arts and Crafts design so I set out to make a case for the radio that would look at home in a room designed by Frank Loyd Wright. I also wanted it to look like it had been around for 50 or 60 years. The sides are the very last scraps of some birch plywood from a project I did over 6 years ago. The remainder is solid oak and some oak trim that I modified. An artfully sloppy stain job (left dark in cracks and crevi and lighter on the corners and edges) with Minwax English Chestnut stain gives the look of age. It is finished with three coats of Deft semi-gloss. I am more or less pleased with the result. The trim above the faceplate is a little too wide and I'm not crazy about the flat top. I may do something about the top.

The faceplate is 1/8 in plastic that I found laying around the engineering shop at work. I made a milling template on Autocad and adhered it to the plate. After milling I soaked it off with some WD-40. I then modified the Cad drawing for the graphics and printed it out on some parchment looking card stock that I found at Hobby Lobby. Adhering the graphic onto the faceplate in proper alignment was another comedy of errors. Needless to say there are come creases and wrinkles that worked out perfectly in my plan for the "aged" look. Two coats of spray poly were applied to protect the paper.

The back plate is a 97-cent clipboard from Wal-Mart. I cut it to fit and shot it with some cheap black spray paint. It has the perfect cheap-black-fiberboard-radio-back appearance. Maybe because that's what it is.

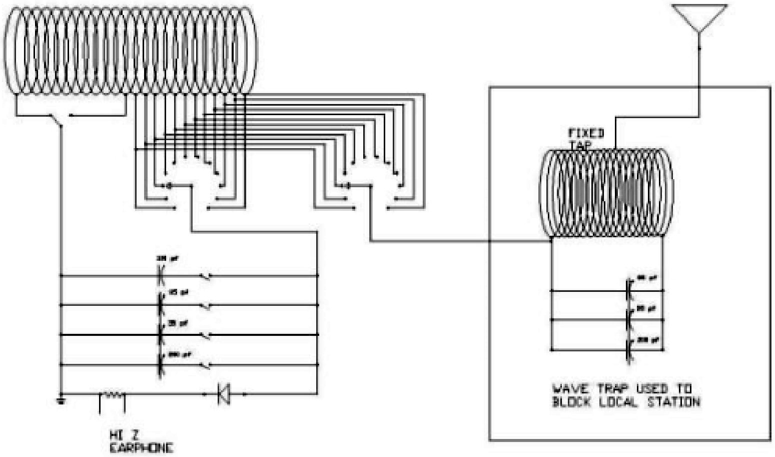

The set tuned very broadly. That may have to do with the huge coil. A local was smeared all over the dial in the daytime and half of it at night. A hastily constructed trap was used to block the local until they signed off at 11PM. Then I would reconnect the antenna directly to my set. Even switching out half of the coil had little effect on the selectivity. Strong stations like WHO and WWL covered a lot of my tuning range. Using the trap as an antenna tuner may have helped but I wanted to stay within the hobby class regulations.

Still I think I did pretty well. I managed to get about 35 stations for a score of just over 55,000 points. The furthest was CBC in Vancouver BC. I will wind a proper coil for the set when time allows.

Michael Branson

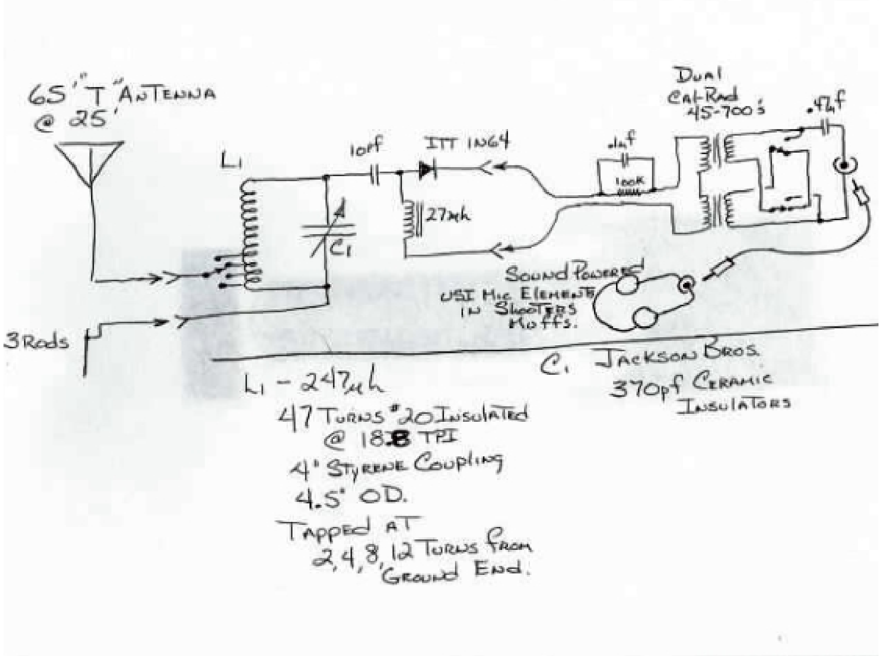
65' "T" ANTENNA
@ 25'
L1
10pf
ITT 1N64
.1uf
100K
Dual
Cal-Rad
45-700's
.47uf
27mh
C1
3 Rods
Sound Powered
USI Mic Elements
in Shooters
Muffs.
L1 - 247uh
47 Turns #20 Insulated
@ 18 TPI
4" Styrene Coupling
4.5" OD.
Tapped at
2,4,8,12 Turns from
Ground End.
C1 Jackson Bros
370pf Ceramic
Insulators

Mike Tuggle

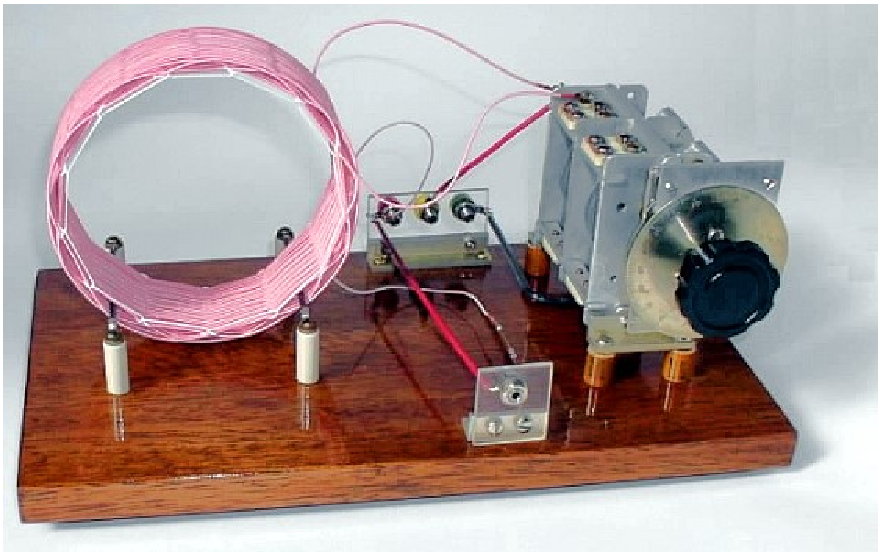

...Turned out to be quite a contest. Conditions were very flat until the second Friday (20 Jan.); then propagation really picked up. Area T-storms produced a moderate static level Sat. & yesterday, but good prop. seemed to persist.

Same set as last year's event, except I did put a lead telluride rockstand on the stronger local stations. (This was the same detector I heard KRVN 880 Nebraska on several contests back. But that was on the Lyonodyne 17 set.)

Two of the stations heard were new: KXTK 1280 never heard before on any set; CFAC 960 first heard in last year's 1-AD contest.

Mercifully a clandestine neighborhood super noise source held off during prime listening hours …

Sean Whitacre

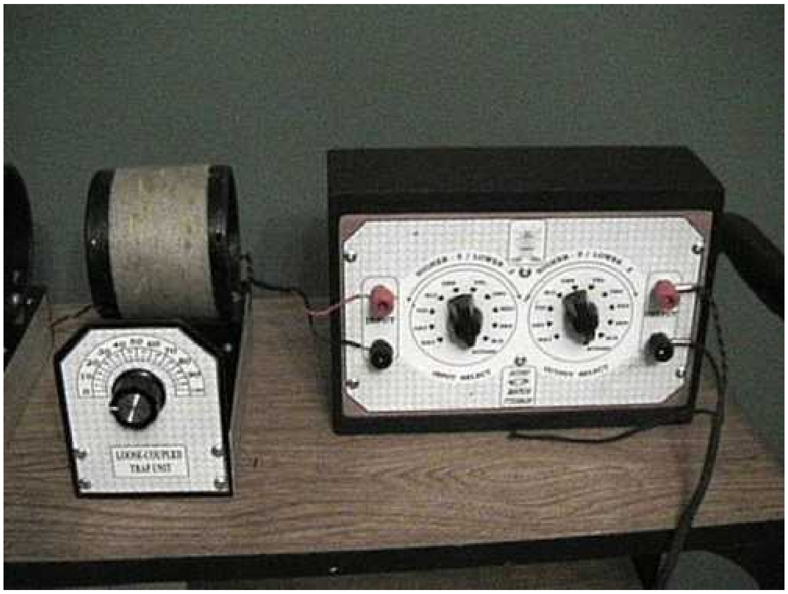
LOOSE-COUPLED TRAP UNIT

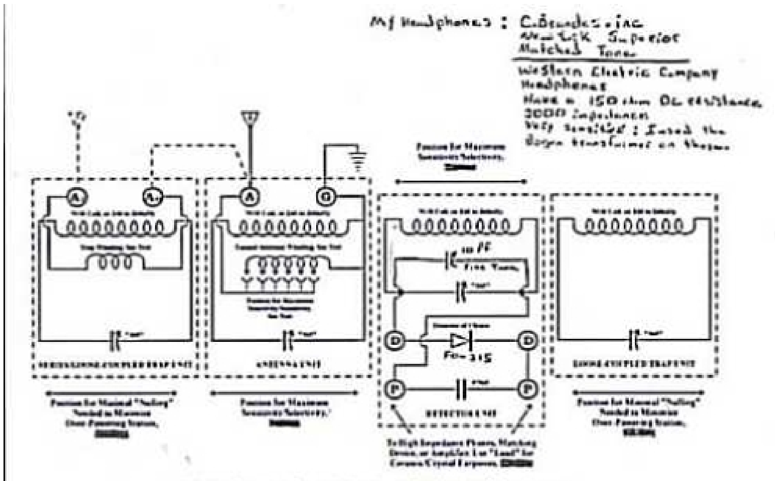
A
A
A
G
10 PF
Fine Tune
D
D
FO-215
P
P
ANTENNA UNIT
DETECTOR UNIT
LOOSE-COUPLED TRAP UNIT

My Headphones: C.Brandes, inc
Navy Type Superior
Matched Tone

Western Electric Company
Headphones
Have a 150 ohm DC resistance
3000 impedance
Very sensitive: I used the
Bogen transformer on these.

Schematic #1, Basic Modular Crystal Radio Receiving System.

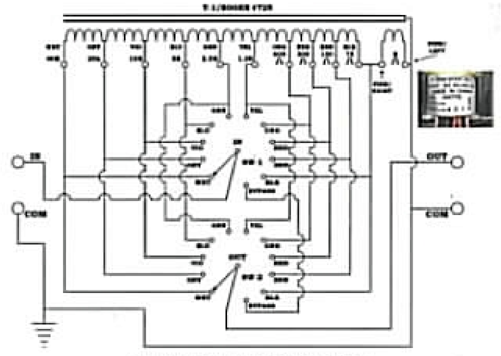

SCHEMATIC #2, AUDIO MATCH, BOGEN.

Continued on Page 3.

Ad ✕

Select Menu

- Home
- About Us
- Contest Results
- 1AD Contests
- 1AD Entries
- Antenna Matching
- Antennas
- BHAM Radios
- Crystal Contests
- Crystal Radio Entries
- Sprint Contest
- Headphones
- Projects, Mods
- Favorite Links
- Contact Jack

Sub Pages

- Main Page
- 2004 Page 1
- 2004 Page 2
- 2004 Page 3
- 2005 Page 1
- 2005 Page 2
- 2006 Page 1
- 2006 Page 2
- 2006 Page 3
- 2007 Page 1
- 2007 Page 2
- 2008 Page 1
- 2008 Page 2
- 2008 Page 3
- 2008 Page 4
- 2009 Page 1
- 2009 Page 2
- 2009 Page 3
- 2009 Page 4
- 2009 Page 5
- 2009 Page 6
- 2010 Page 1

US006024935A

United States Patent [19]

Mills et al.

[11] **Patent Number: 6,024,935**

[45] **Date of Patent: Feb. 15, 2000**

[54] **LOWER-ENERGY HYDROGEN METHODS AND STRUCTURES**

[75] Inventors: **Randell L. Mills**, Malvern; **William R. Good**, Wayne; **Jonathan Phillips**, State College; **Arthur I. Popov**, Philadelphia, all of Pa.

[73] Assignee: **Blacklight Power, Inc.**, Cranbury, N.J.

[21] Appl. No.: **08/822,170**

[22] Filed: **Mar. 21, 1997**

Related U.S. Application Data

[63] Continuation of application No. 08/592,712, Jan. 26, 1996, abandoned.

[51] **Int. Cl.**^{7} **C01B 3/02**

[52] **U.S. Cl.** **423/648.1**; 422/129

[58] **Field of Search** 423/648.1; 422/129

[56] **References Cited**

U.S. PATENT DOCUMENTS

1,001,589	8/1911	Hatfield .	
2,708,656	5/1955	Fermi .	
3,297,484	1/1967	Niedrach	136/86
3,300,345	1/1967	Lyons, Jr.	136/86

(List continued on next page.)

FOREIGN PATENT DOCUMENTS

0 392 325	10/1990	European Pat. Off. .	
0 395 066	10/1990	European Pat. Off.	376/100

(List continued on next page.)

OTHER PUBLICATIONS

The Associated Press, "Pennsylvania Company . . . Cold Fusion Mystery"; 1991, Lexis Nexis Reprint.

Boston Globe, Wednesday, Apr. 19, 1989, "Successful nuclear fusion experiment by the Italians".

Broad, "2 Teams Put New Life in Cold Fusion Theory", New York Times, Apr. 26, 1991, p. A18.

(List continued on next page.)

Primary Examiner—Wayne Langel

Attorney, Agent, or Firm—Jeffrey S. Melcher; Farkas & Manelli PLLC

[57] **ABSTRACT**

Methods and apparatus for releasing energy from hydrogen atoms (molecules) by stimulating their electrons to relax to quantized lower energy levels and smaller radii (smaller semimajor and semiminor axes) than the "ground state" by providing energy sinks or means to remove energy resonant with the hydrogen energy released to stimulate these transitions. An energy sink, energy hole, can be provided by the transfer of at least one electron between participating species including atoms, ions, molecules, and ionic and molecular compounds. In one embodiment, the energy hole comprises the transfer of t electrons from one or more donating species to one or more accepting species whereby the sum of the ionization energies and/or electron affinities of the electron donating species minus the sum of the ionization energies and/or electron affinities of the electron accepting species equals approximately mX27.21 eV (mX48.6 eV) for atomic (molecular) hydrogen below "ground state" transitions where m and t are integers. The present invention further comprises a hydrogen spillover catalyst, a multifunctionality material having a functionality which dissociates molecular hydrogen to provide free hydrogen atoms which spill over to a functionality which supports mobile free hydrogen atoms and a functionality which can be a source of the energy holes. The energy reactor includes one of an electrolytic cell, a pressurized hydrogen gas cell, and a hydrogen gas discharge cell. A preferred pressurized hydrogen gas energy reactor comprises a vessel; a source of hydrogen; a means to control the pressure and flow of hydrogen into the vessel; a material to dissociate the molecular hydrogen into atomic hydrogen, and a material which can be a source of energy holes in the gas phase. The gaseous source of energy holes includes those that sublime, boil, and/or are volatile at the elevated operating temperature of the gas energy reactor wherein the exothermic reaction of electronic transitions of hydrogen to lower energy states occurs in the gas phase.

499 Claims, 9 Drawing Sheets

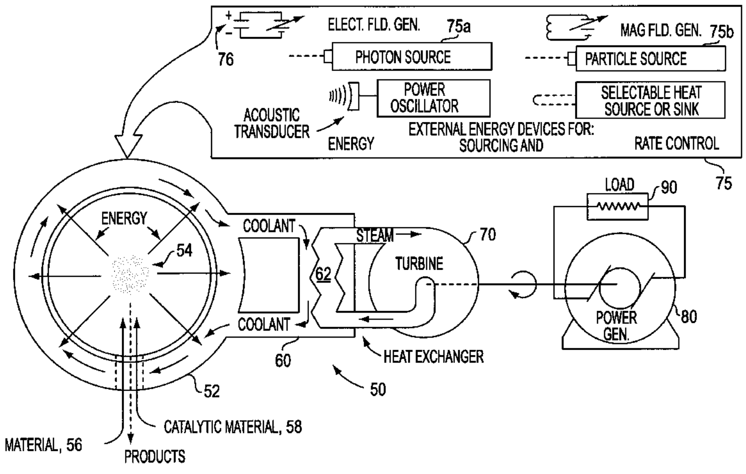

U.S. PATENT DOCUMENTS

3,359,422	12/1967	Pollock	250/84.5
3,377,265	4/1968	Caesar	204/290
3,448,035	6/1969	Serfass	204/278
3,669,745	6/1972	Beccu	136/20
3,701,632	10/1972	Lovelock	23/232
3,755,128	8/1973	Herwig	204/320
3,816,192	6/1974	Brower .	
3,835,019	9/1974	Lovelock	204/130
3,917,520	11/1975	Katz et al.	204/129
4,265,720	5/1981	Winstel	204/129
4,274,938	6/1981	Schulten et al.	204/239
4,487,670	12/1984	Bellanger et al.	204/129
4,512,966	4/1985	Nelson .	
4,568,568	2/1986	Asano et al.	427/125
4,664,904	5/1987	Wolfrum et al.	423/648.1
4,737,249	4/1988	Shepard, Jr. et al.	204/129
4,774,065	9/1988	Penzorn .	
4,923,770	5/1990	Grasselli et al.	429/101
4,968,395	11/1990	Pavelle et al.	204/130
4,986,887	1/1991	Gupta et al. .	
5,215,729	6/1993	Buxbaum	423/648.1
5,318,675	6/1994	Patterson .	
5,372,688	12/1994	Patterson .	

FOREIGN PATENT DOCUMENTS

53-134792	7/1975	Japan .	
136644	10/1981	Japan	423/648.1
WO 90/10935 A1	9/1990	WIPO .	
WO 90/13126	11/1990	WIPO	376/100
WO 90/14668 A2	11/1990	WIPO .	
WO 91/01036 A1	1/1991	WIPO .	
WO 91/08573 A1	6/1991	WIPO .	
WO 92/10838	6/1992	WIPO	376/100
WO 93/17437 A1	9/1993	WIPO .	
WO 94/10688 A1	5/1994	WIPO .	
WO 94/14163 A1	6/1994	WIPO .	
WO 94/15342 A1	7/1994	WIPO .	
WO 94/29873	12/1994	WIPO .	
WO 95/20816 A1	8/1995	WIPO .	
WO 96/42085 A2	12/1996	WIPO .	

OTHER PUBLICATIONS

Bush, et. al., "Helium Production During the Electrolysis . . . Experiments", Preliminary Note, Univ. of Texas, pp. 1–12.

Notoya, "Cold Fusion . . . Nickel Electrode", Fusion Technology, vol. 24, pp. 202–204.

Notoya, "Tritium Generation . . . Nickel Electrodes", Fusion Technology, vol. 26, pp.

Oka, et. al., "D_2O–fueled fusion power reactor using electromagnetically induced D–D_n, D–D_p, and Deuterium–tritium reactions—preliminary design of a reactor system", Fusion Technology, vol. 16, No. 2, Sep. 1989, pp. 263–267.

Ohmori, et. al., "Excess Heat Evolution . . . Tin Cathodes" Fusion Technology, vol. 24, pp. 293–295 (1993).

Rogers, "Isotopic hydrogen fusion in metals", Fusion Technology, vol. 16, No. 2, Sep. 1989, pp. 2254–2259.

Rout, et. al., "Phenomenon of Low Energy Emissions from Hydrogen/Deuterium Loaded Palladium", 3rd Annual Conference on Cold Fusion (Oct. 21–25, 1992).

Srinivasan, et. al., "Tritium and Excess Heat Generation during Electrolysis of Aqueous Solutions of Alkali Salts with Nickel Cathode", 3rd Annual Conference on.

Stein, "Theory May Explain Cold Fusion Puzzle", Lexis Reprint, Washington News, Apr. 25, 1991.

Suplee, "Two New Theories on Cold Fusion . . . Scientists"; The Washington Post 1st Section, p. A11, (1991).

Bishop, "More Labs Report Cold Fusion Results", Wall Street Journal, Oct. 19, 1992.

Bishop, "It ain't over til it's over . . . Cold Fusion", Popular Science, Aug. 1993, pp. 47–51.

Browne, "Pysicists Put Atom in 2 Places at Once", The New York Times.

Bush, et. al., "Power in a Jar: the Debate Heats Up", 88 Bus. Week, Science & Technology, Oct. 26, 1992.

Bush, et. al., "Helium Production During the Electrolysis . . . Experiments", Preliminary Note, Univ. of Texas, pp. 1–12.

Catlett, et. al., "Hydrogen transport in lithium hydride as a function of pressure" The Journal of Chemical Physics, 58(8), 3432–3438 (Apr. 1978).

Chien, et. al., "On an Electrode . . . Tritium and Helium", J. Electroanal Chem., 1992, pp. 189–212.

Close, "Too Hot to Handle—The Race for Cold Fusion", Princeton University Press, 1989.

Criddle, "The Roles of Solution . . . Excess Heating", Electrochemical Science & Technology Centre, Univ. of Ottawa.

Datz, et. al., "Molecular Association in Alkali Halide Vapors", Journal of Chemical Physics, vol. 34, No. 5, (Feb. 1961), pp. 558–564.

Dagani, "Cold Fusion—Utah Pressures Pons, Fleischmann", Jan. 14, 1991, C&EN, pp. 4–5.

Dagani, "Latest Cold Fusion Results Fail to Win Over Skeptics", Jun. 14, 1993, C&EN, pp. 38–40.

Dagani, "New Evidence Claimed for Nuclear Process in 'Cold Fusion'", C&EN Washington, (Apr. 1991), pp. 31–33.

Experimental Verification by Idaho National Engineering Laboratory, pp. 13–25.

Hardy, et. al., "The Volatility of Nitrates and Nitrites of the Alkali Metals", Journal of the Chemical Society, pp. 5130–5134 (1963).

Huizenga, "Cold Fusion—The Scientific Fiasco of the Century", Oxford University Press, 1993.

Huizenga, "Cold Fusion Labled Fiasco of the Century", Forum for Applied Research and Public Policy, vol. 7, No. 4, pp. 78–83.

Jones, (article by Dagani), "Cold Fusion Believer . . . Research", Jun. 5, 1995, C&EN, pp. 34–41.

Jones, "Current Issues in Cold Fusion . . . Particles", Surface and Coatings Technology, 51 (1992), pp. 283–289.

Jones, et. al., "Faradic Efficiences . . . Cells", J. Phys. Chem. 1995, pp. 6973–6979.

Jones, et. al., "Examination of Claims of Miles . . . Experiments", J. Phys. Chem. 1995, pp. 6966–6972.

Ivanco, et. al., "Calorimetry For a Ni/K_2CO_3 Cell", AECL Research, Jun. 1994.

Kahn, "Confusion in a Jarr", Nova 1991.

Karabut, et. al., "Nuclear Product . . . Deuterium", Physics Letters A170, (1992), p. 265.

Klein, "Attachments to Report of Cold Fusion Testing", Cold Fusion, No. 9, pp. 16–19.

Labov, "Special Observations . . . Background", The Astrophysical Journal, 371, Apr. 20, 1991, pp. 810–819.
Lehigh X–Ray Photoelectron Spectroscopy Report, Dec. 8, 1993.
Miles, et. al., "Search for Anomalous Effects . . . Palladium Cathodes", Naval Air Warfare Center Weapons Divsion, Proceedings of 3rd Int. Conf. on Cold Fusion.
Miles et al, "Correlation of Excess . . . Palladium Cathodes", J. Electronl. Chem., 1993, pp. 99–117.
Miles et al, "Heat and Helium . . . Experiments", Conference Proceedings, vol. 33, 1991, pp. 363–372.
Miles et al, "Electrochemical . . . Palladium Deuterium System", J. Electroanal Chem., 1990, pp. 241–254.
Mills, et. al., "Fractional Quantum . . . Hydrogen", Fusion Technology, vol. 28, Nov. 1995, pp. 1697–1719.
Mills, Unification of Spacetime, the Forces, Matter, Energy, HydroCatalysis Power Corporation, 1992, pp. 53–84.
Mills, "The Grand Unified Theory of Classical Quantum Mechanics", pp. 1–9.
Mills, "Hydrocatalysis Power Technology", Statement of Dr. Randall L. Mills, May, 1993.
Mills Technologies, "1KW Heat Exchanger System", Thermacore, Inc., Oct. 11 1991, pp. 1–6.
Mills Technologies, "1KW Heat Exchanger System", Thermacore, Inc., Apr. 17, 1992, pp. 1–6.
Monroe, et. al., "A Schrodinger Cat Superposition State of an Atom", Science, vol. 272, (May 24, 1996), pp. 1131–1101.
Morrison, "Review of Progress in Cold Fusion", Transactions of Fusion Technology, vol. 26, Dec. 1994, pp. 48–55.
Morrison, "Cold Fusion Update No. 12, ICCPG", Jan. 17, 1997, available online at "www.skypoint.com".
Niedra, "Replication of the Apparant Excess Heat Effect in Light Water . . . Cell", NASA Technical Memorandum 107167, (Feb. 1996).
Nieminen, "Hydrogen atoms band together", Nature, vol. 356, Mar. 26, 1992, pp. 289–291.
Notoya, et. al., "Excess Heat Production in Electrolysis . . . Electrodes", Proceedings of the Int. Conf. on Cold Fusion, Oct. 21–25, 1992, Tokyo, Japan.
Rees, "Cold Fusion . . . What Do We Think?", Journal of Fusion Energy, (1991), vol. 10, No. 1, pp. 110–116.
Rousseau, "Case Studies in Path© Science", vol. 80, American Scientist, (1992), pp. 54–63.
Service, "Cold Fusion:Still Going", Newsweek Focus, Jul. 19, 1993.
Shaubach, et. al., "Anomalous Heat . . . Carbonate", Thermacore, Inc., pp. 1–10.
Storms, et. al., "Electroyltic Tritium Production", Fusion Technology, vol. 17, Jul. 1990, pp. 680–695.
Taubes, "Bad Science", Random House, 1993, pp. 303, 425–481.
Vaselli et al., "Screening Effect of Impurities in Metals: A possible Explanation of the Process of Cold Nuclear Fusion", 11 Nuovo Cimento Della Societa Italiana di Fisica.
Williams, "Upper Bounds on Cold Fusion in Electrolytic Cells", Nature, vol. 342, 23 Nov. 1989, pp. 375–384.
Yamaguchi et al, "Direct Evidence . . . Palladium", NTT Basic Research Laboratories, (1992) pp. 1–10.
Zweig, "Quark Catalysis of Exothermal Nuclear Reactions", Science, vol. 201, (1978), pp. 973–979.
Bush, et. al., Journal Electrochanal. Chem., vol. 304, pp. 271–278 (1991).
Shrivenvassan, et. al., 3rd Annual Conference on Cold Fusion (1992).
Notoya, Fusion Technology, vol. 24, p. 202 (1993).
Ohmori, et. al., Fusion Technology, vol. 24, p. 293 (1993).
Boston Globe, Wednesday, Apr. 19, 1989, "Successful nuclear fusion experiment by the Italians".
Oka, et. al., "D_2O–fueled fusion power reactor using electromagnetically induced D–D_n, D–D_p, and Deuterium–tritium reactions—preliminary design of a reactor system", Fusion.
Fusion Digest, "Cold Nuclear Fusion Bibliography", 1993.
Rogers, "Isotopic hydrogen fusion in metals", Fusion Technology, vol. 16, No. 2.
Fusion Digest, "Heat? Neutrons? Charged Particles?", 1993.
Brodowsky, "Solubility and diffusion of hydrogen and deuterium in palladium and palladium alloys", Technical Bulletin, Engelhard Indust., vol. 7, No. 1–2 (1966), pp. 41–50.
Prop. to the United Press, "Theory May Explain 'Cold Fusion' Puzzle"; 1991; Lexis Nexis Reprint.
The Associated Press, "Pennsylvania Company . . . Cold Fusion Mystery"; 1991, Lexis Nexis Reprint.
The New York Times, "2 Teams Put New Life in 'Cold' Fusion Theory"; 1991; Section A, p. 18, col. 1; Lexis Nexis Reprint.
The Washington Post, "Two New Theories on Cold Fusion . . . Scientists"; 1991; 1st.
Albagli et al., *Journal of Fusion Energy*, 9(2):133–148 (1990).
Alber et al., Z. Phys. A.—*Atomic Nuclei*, vol. 333, (1989), pp. 319–320.
Alessandrello et al., *Il Nuovo Cimento*, 103A (11) :1617–1638 (1990).
Balke et al., Physical Review C, 42 91) :30–37 (1990).
Benetskii et al., Kratkie Soobshcheniya po fizike, No. 6, pp. 58–60, 1989 (translation of).
Besenbacher et al., Journal of Fusion Energy, 9 (3) :315–317 (1990).
Bush et al., J. Electroanal. Chem., 304:271–278 (1991).
Chapline, UCRL—101583, Jul. 1989, pp. 1–9.
Cooke, ORNL/FTR—3341, Jul. 31, 1989, pp. 2–15.
Cribier et al., Physics Letters B, vol. 228, No. 1, Sep. 7, 1989, pp. 163–166.
Faller et al., J. Radioanal. Nucl. Chem. Letters, vol. 137, No. 1, (Aug. 21, 1989), pp. 9–16.
Hajdas et al., Solid State Communications, vol. 72, No. 4, (1989) pp. 309–313.
Horanyi, J. Radioanal. Nucl. Chem., Letters, vol. 137, No. 1, (Aug. 21, 1989), pp. 23–28.
Kreysa et al., J. Electronanal. Chem. vol. 266, (1989) pp. 437–450.
Legett et al., Physical Review Letters, 63(2):191–194 (1989).
Lewis et al., Nature, vol. 340, Aug. 17, 1989, pp. 525–530.
Maly et al., Fusion Technology, vol. 24, Nov. 1993, pp. 307–318.
McNally, Jr., Fusion Technology, 16(2):237–239 (1989).
Mills et al., Fusion Technology, 25:103 (1994).
Mills et al., Fusion Technology, vol. 20, (Aug. 1991), pp. 65–81.
Miskelly et al., Science, vol. 246, No. 4931, Nov. 10, 1989, pp. 793–796.
Noninski, Fusion Technology, vol. 21, (Mar. 1992), pp. 163–167.
Noninski et al., Fusion Technology, vol. 19, Mar. 1991, pp. 364–368.
Ohashi et al., J. of Nucl. Sci. and Tech., vol. 26, No. 7, (Jul. 1989), pp. 729–732.

Price et al., Physical Review Letters, vol. 63, No. 18, Oct. 30, 1989, pp. 1926–1929.

Salamon et al., Nature, vol. 344, Mar. 29, 1990, pp. 401–405.

Schrieder et al., Z. Phys. B–Condensed Matter, vol. 76, No. 2, pp. 141–142, (1989).

Shani et al., Solid State Communications, vol. 72, No. 1, (1989) pp. 53–57.

The New York Times, May 3, 1989, pp. A1, A22, article by M. Browne.

The Wall Street Journal, Apr. 26, 1989, p. B4, article by D. Stipp.

The Washington Post, May 2, 1989, pp. A1, A7, article by P. Hilts.

The Washington Post, Jul. 13, 1989, pp. A14.

The Washington Post, Mar. 29, 1990, p. A3.

The Washington Times, Mar. 24, 1989, p. A5, article by D. Braaten.

Ziegler et al., Physical Review Letters, vol. 62, No. 25, Jun. 19, 1989, pp. 2929–2932.

FIG. 1

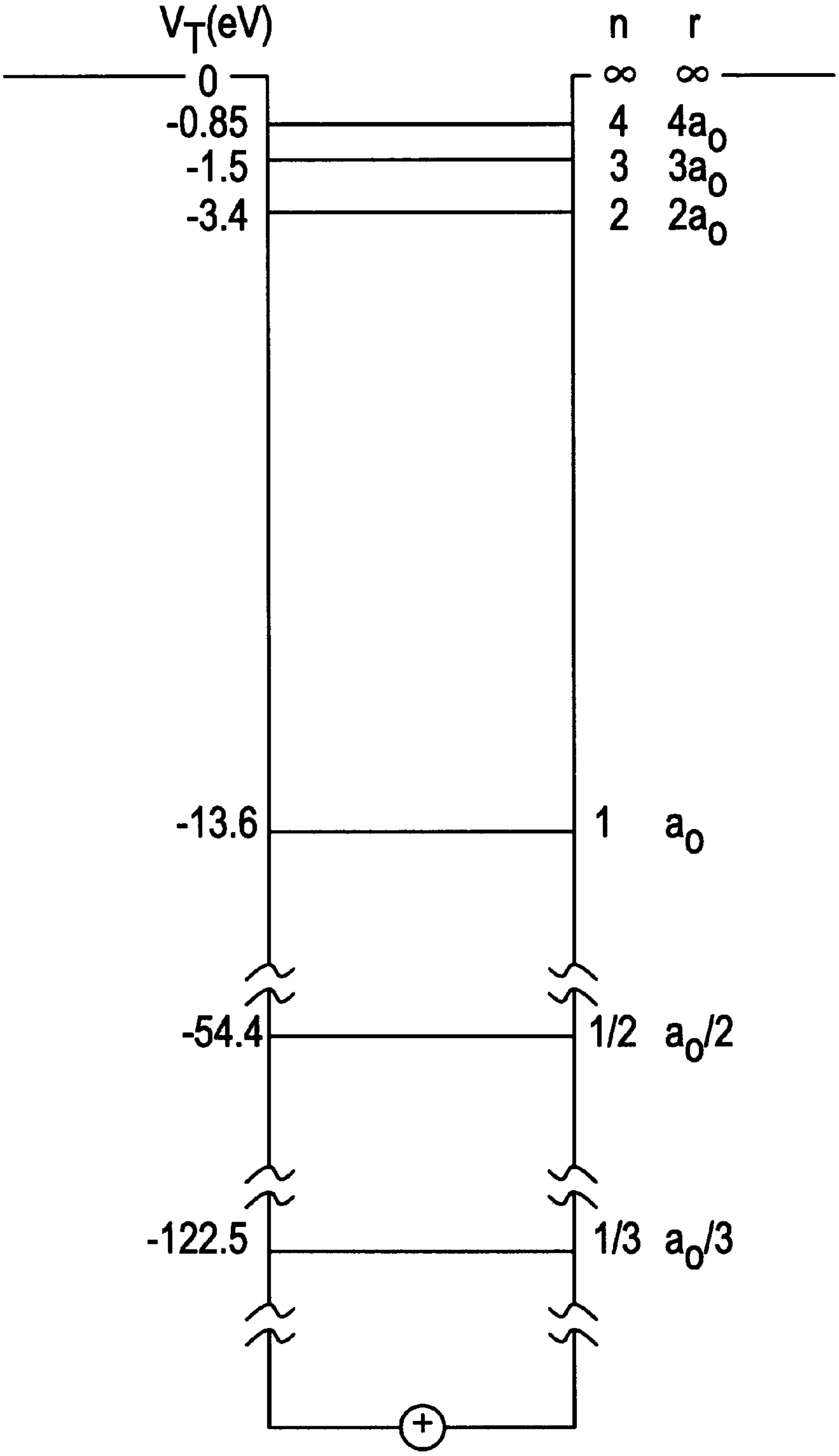
V_T(eV)
n
r
0
∞
∞
-0.85
4
$4a_0$
-1.5
3
$3a_0$
-3.4
2
$2a_0$
-13.6
1
a_0
-54.4
1/2
$a_0/2$
-122.5
1/3
$a_0/3$
+

FIG. 2

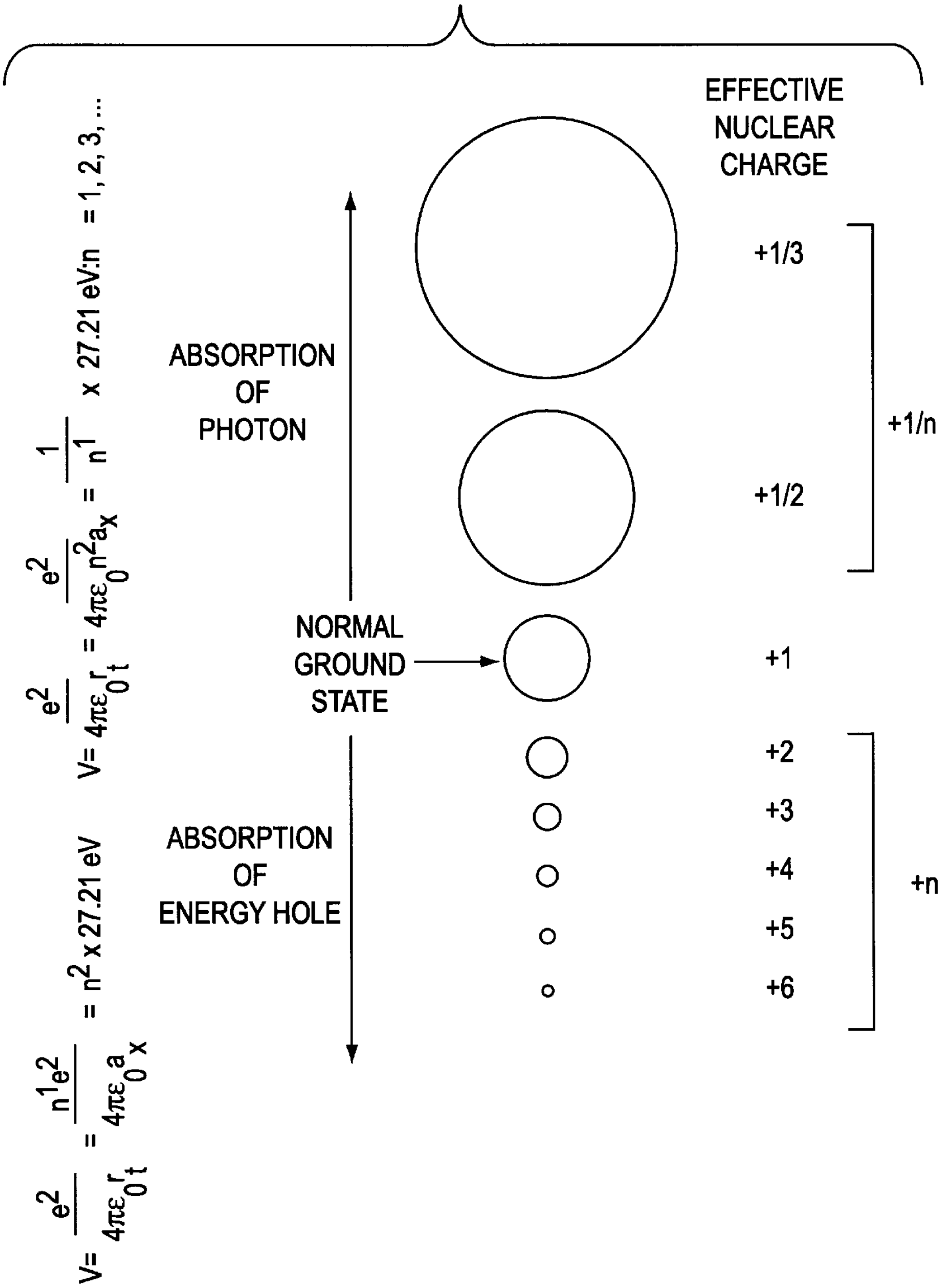
$V = \frac{e^2}{4\pi\varepsilon_0 r_t} = \frac{e^2}{4\pi\varepsilon_0 n^2 a_x} = \frac{1}{n^1} \times 27.21\ \text{eV};n = 1, 2, 3, \ldots$
$V = \frac{e^2}{4\pi\varepsilon_0 r_t} = \frac{n^1 e^2}{4\pi\varepsilon_0 a_x} = n^2 \times 27.21\ \text{eV}$
ABSORPTION OF PHOTON
NORMAL GROUND STATE
ABSORPTION OF ENERGY HOLE
EFFECTIVE NUCLEAR CHARGE
+1/3
+1/2
+1/n
+1
+2
+3
+4
+5
+6
+n

FIG. 3

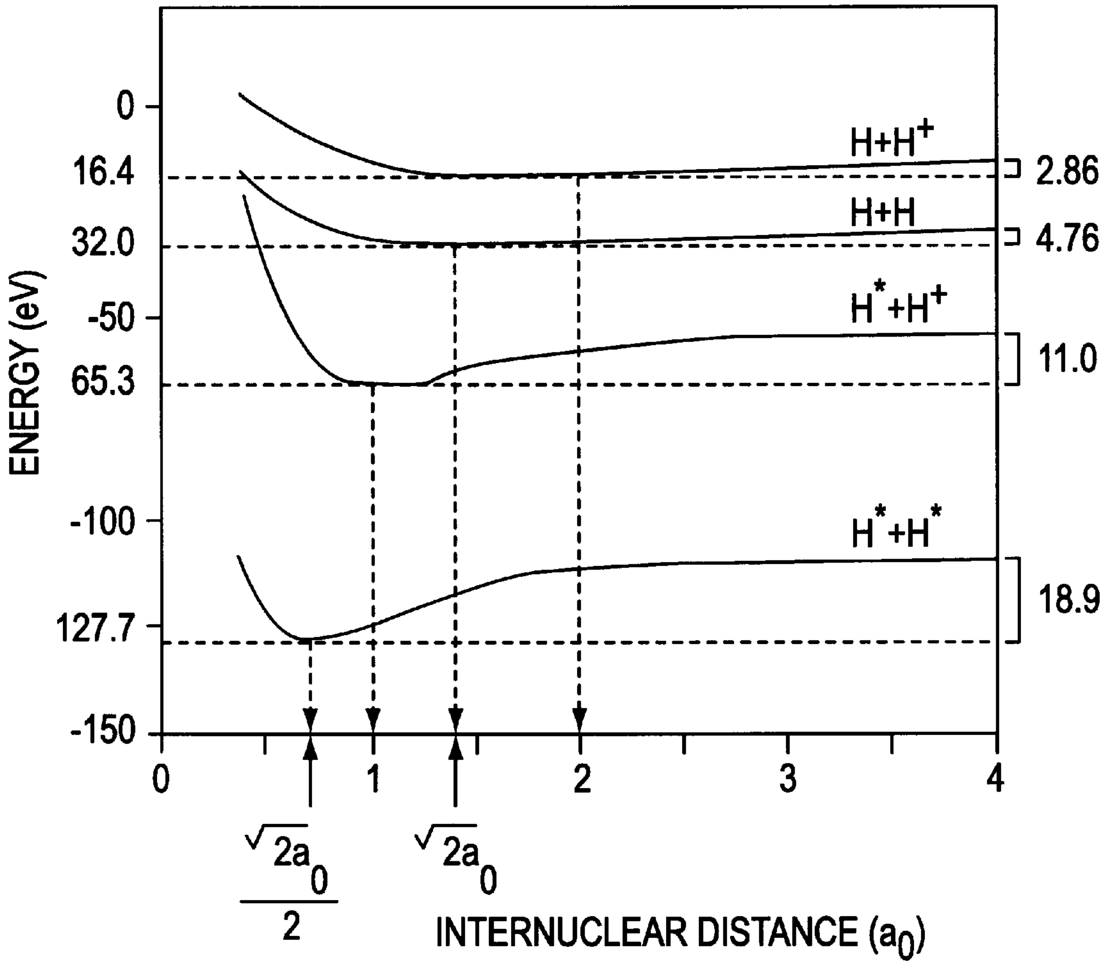
0
16.4
32.0
-50
65.3
-100
127.7
-150
ENERGY (eV)
H+H$^+$
H+H
H*+H$^+$
H*+H*
2.86
4.76
11.0
18.9
0
1
2
3
4
$\frac{\sqrt{2}a_0}{2}$
$\sqrt{2}a_0$
INTERNUCLEAR DISTANCE (a_0)

FIG. 4

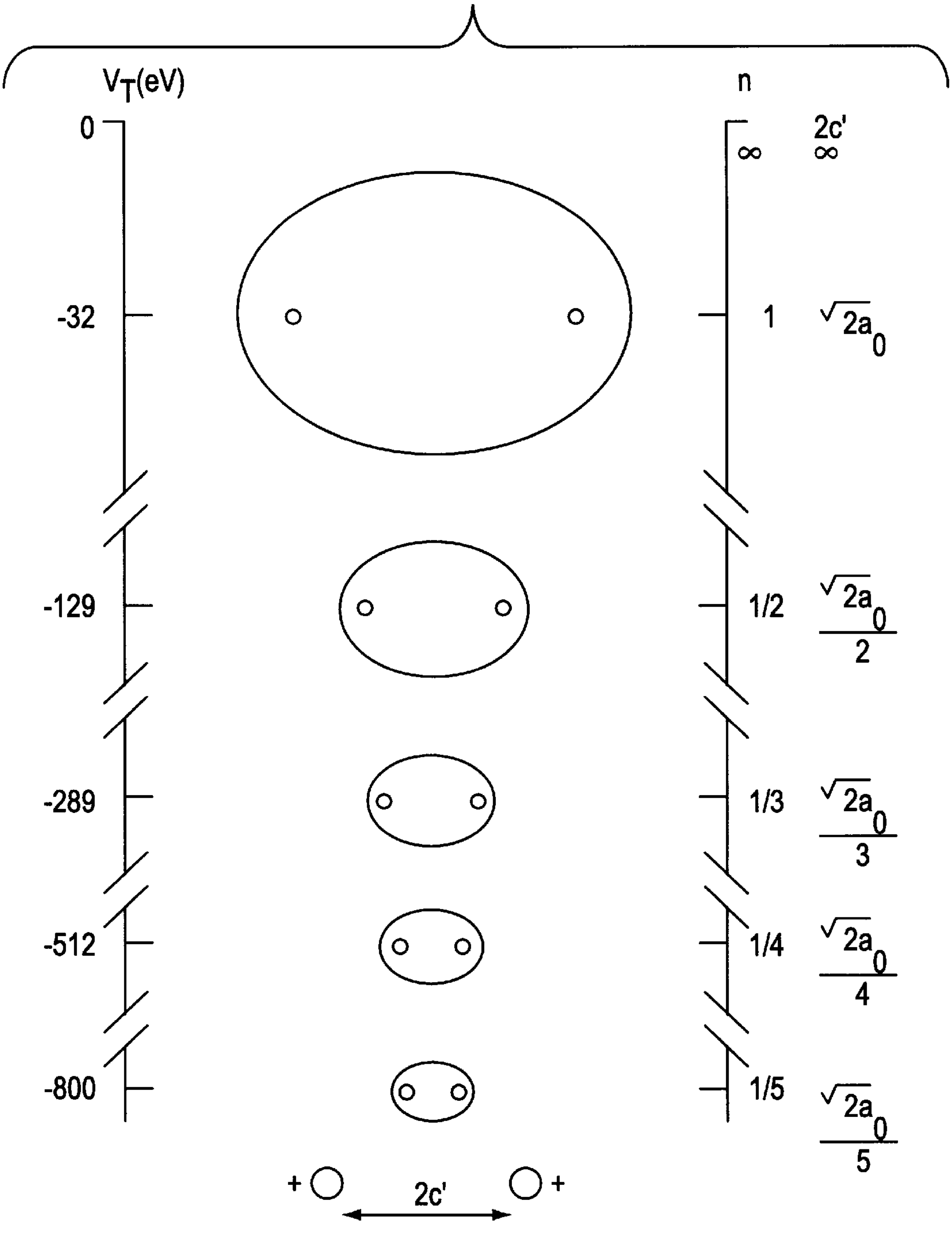

V_T(eV)
0
-32
-129
-289
-512
-800
n
∞
1
1/2
1/3
1/4
1/5
2c'
∞
$\sqrt{2}a_0$
$\frac{\sqrt{2}a_0}{2}$
$\frac{\sqrt{2}a_0}{3}$
$\frac{\sqrt{2}a_0}{4}$
$\frac{\sqrt{2}a_0}{5}$
+
+
2c'

FIG. 5

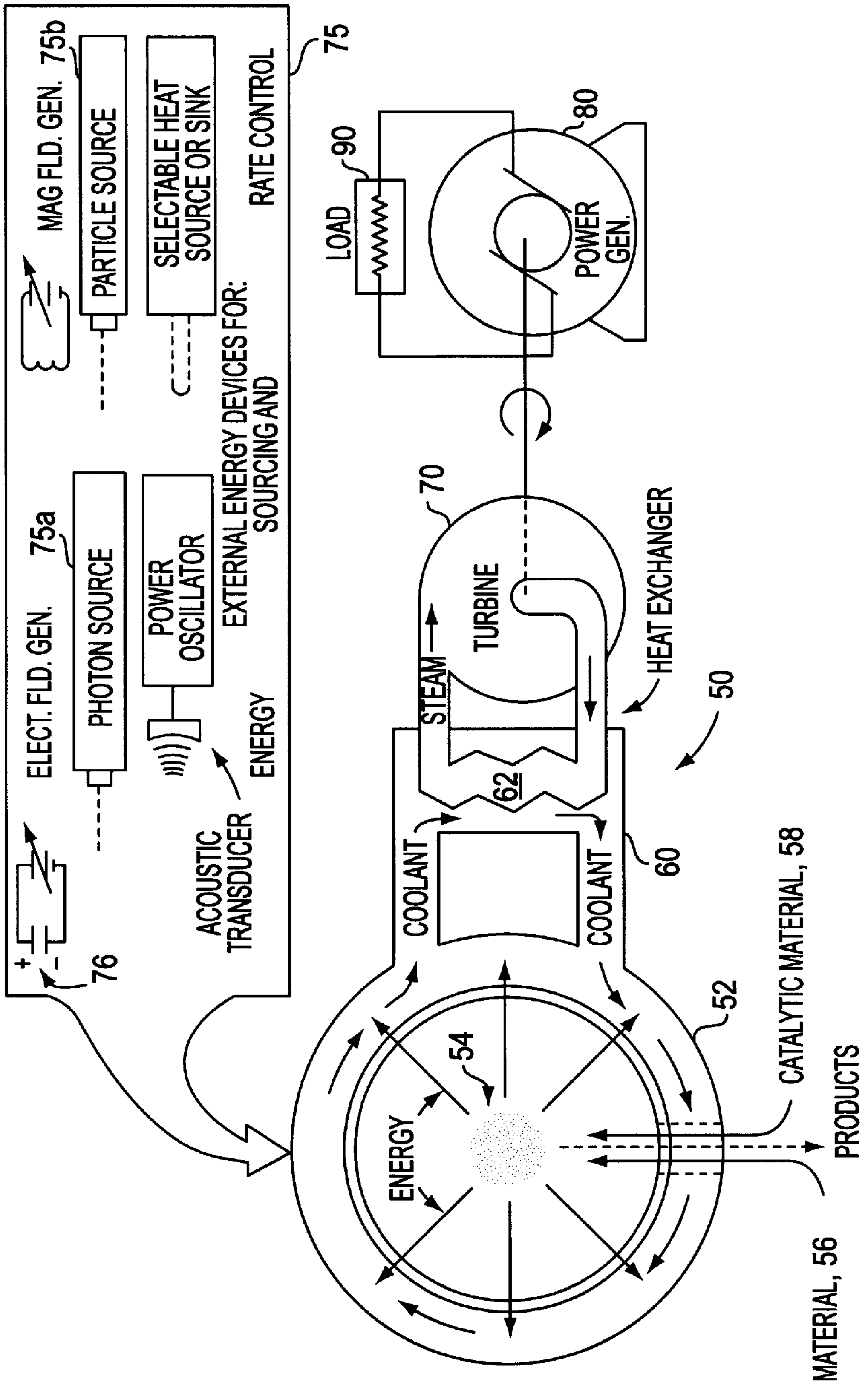
ELECT. FLD. GEN.
75a
PHOTON SOURCE
POWER OSCILLATOR
ENERGY
ACOUSTIC TRANSDUCER
EXTERNAL ENERGY DEVICES FOR: SOURCING AND
MAG FLD. GEN.
75b
PARTICLE SOURCE
SELECTABLE HEAT SOURCE OR SINK
RATE CONTROL
75
+
-
76
LOAD
90
POWER GEN.
80
70
TURBINE
STEAM
HEAT EXCHANGER
50
62
COOLANT
COOLANT
60
54
ENERGY
52
CATALYTIC MATERIAL, 58
PRODUCTS
MATERIAL, 56

FIG. 6

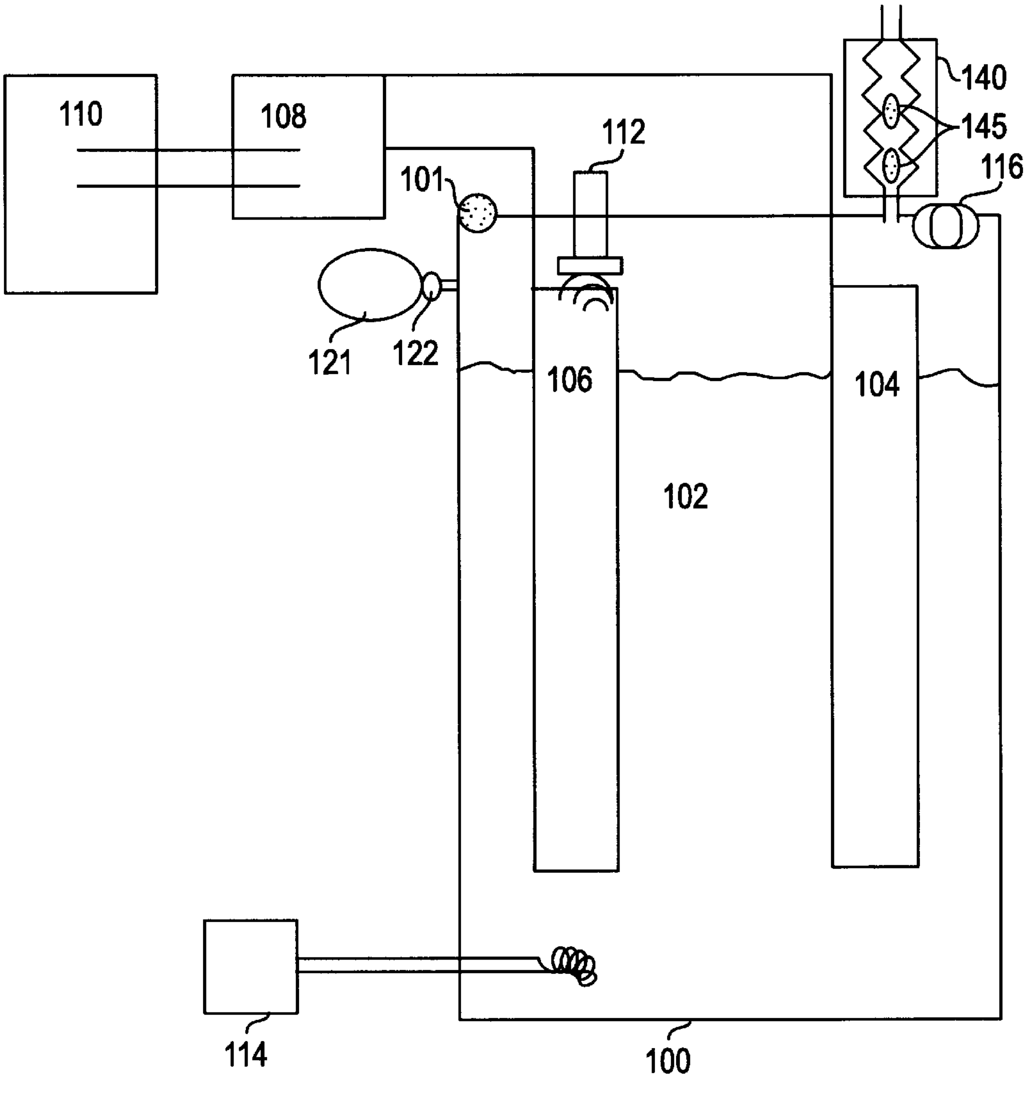
110
108
112
140
145
116
101
121
122
106
104
102
114
100

FIG. 7

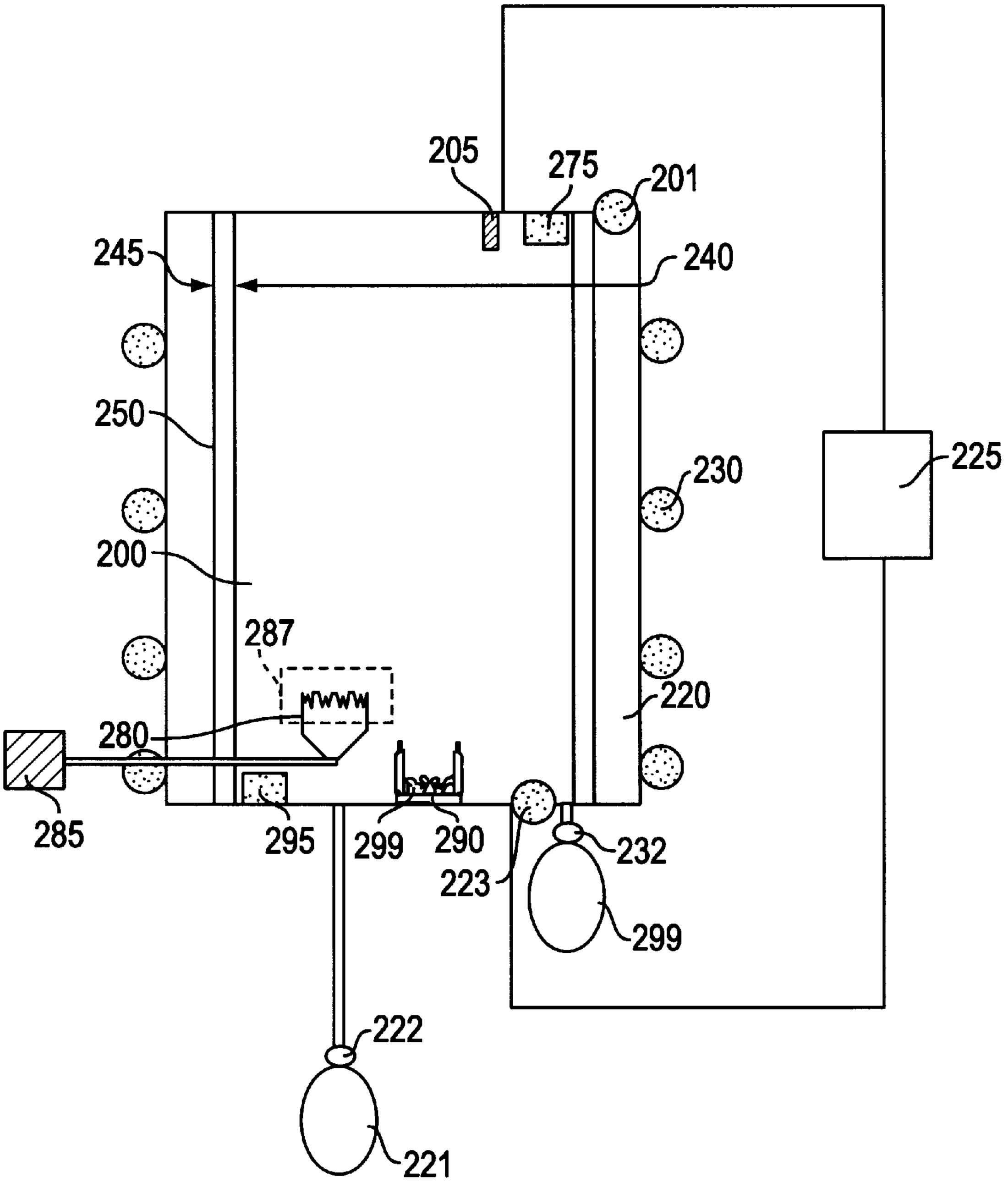

FIG. 8

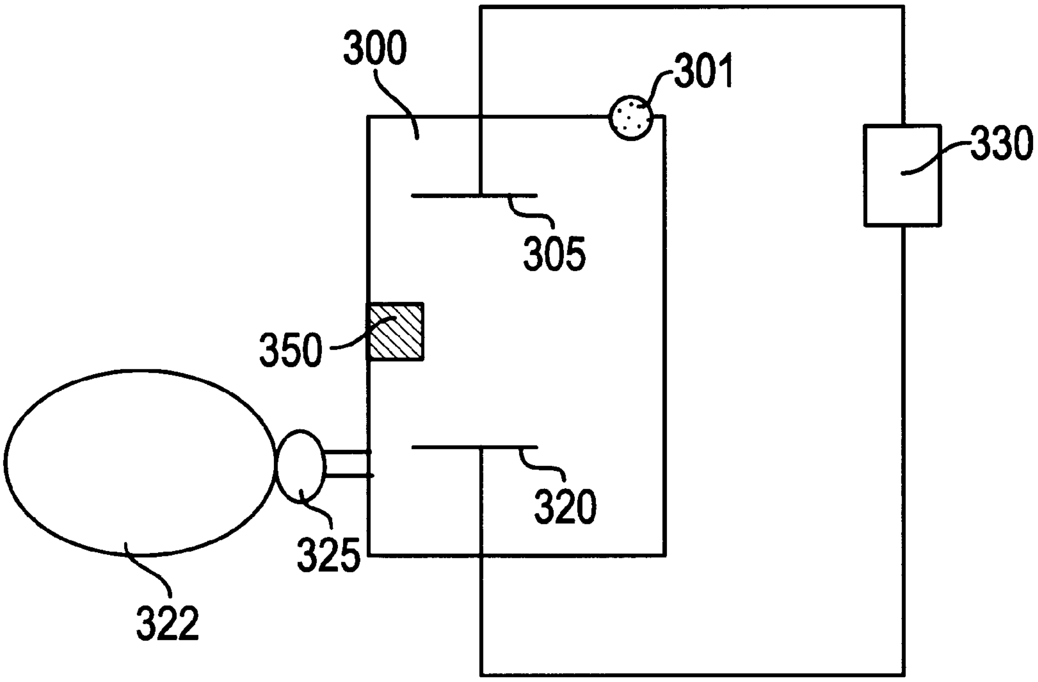
300
301
330
305
350
320
325
322

FIG. 9

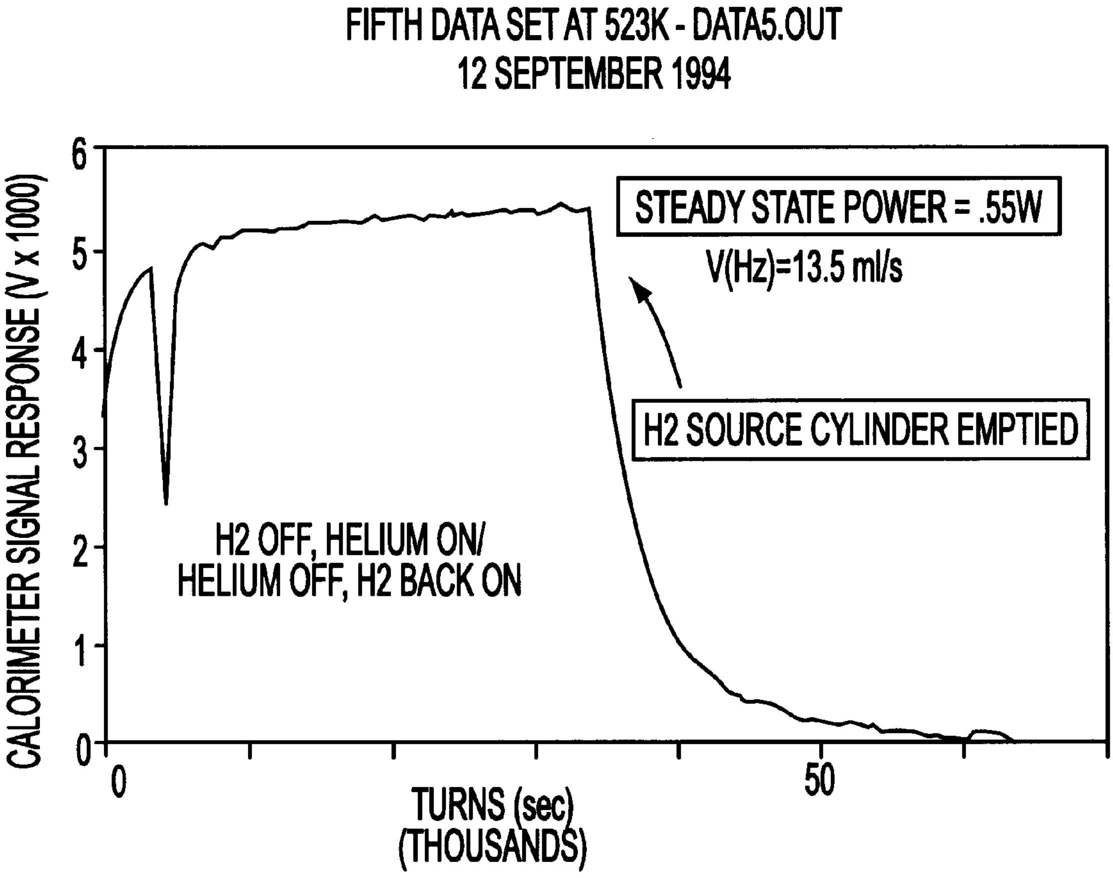
FIFTH DATA SET AT 523K - DATA5.OUT
12 SEPTEMBER 1994
CALORIMETER SIGNAL RESPONSE (V x 1000)
6
5
4
3
2
1
0
0
50
TURNS (sec)
(THOUSANDS)
STEADY STATE POWER = .55W
V(Hz)=13.5 ml/s
H2 SOURCE CYLINDER EMPTIED
H2 OFF, HELIUM ON/
HELIUM OFF, H2 BACK ON

LOWER-ENERGY HYDROGEN METHODS AND STRUCTURES

This application is a continuation of U.S. patent application Ser. No. 08/592,712, filed Jan. 26, 1996, abandoned.

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates to methods and apparatus for releasing energy from hydrogen atoms (molecules) as their electrons are stimulated to relax to lower energy levels and smaller radii (smaller semimajor and semiminor axes) than the "ground state" by providing a transition catalyst which acts as an energy sink or means to remove energy resonant with the electronic energy released to stimulate these transitions according to a novel atomic model. The transition catalyst should not be consumed in the reaction. It accepts energy from hydrogen and releases the energy to the surroundings. Thus, the transition catalyst returns to the origin state. Processes that require collisions are common. For example, the exothermic chemical reaction of H+H to form H_2 requires a collision with a third body, M, to remove the bond energy–$H+H+M \rightarrow H_2+M$. The third body distributes the energy from the exothermic reaction, and the end result is the H_2 molecule and an increase in the temperature of the system. Similarly, the transition from the n=1 state of hydrogen to the

$$n = \frac{1}{\text{integer}}$$

states of hydrogen is possible via a resonant collision, say n=1 to n=½. In these cases, during the collision the electron (s) couples to another electron transition or electron transfer reaction, for example, which can absorb the exact amount of energy that must be removed from the hydrogen atom (molecule), a resonant energy sink. The end result is a lower-energy state for the hydrogen and increase in temperature of the system. Each of such reactions is hereafter referred to as a shrinkage reaction: each transition is hereafter referred to as a shrinkage transition; each energy sink or means to remove energy resonant with the hydrogen electronic energy released to effect each transition is hereafter referred to as an energy hole, and the electronic energy removed by the energy hole to effect or stimulate the shrinkage transition is hereafter referred to as the resonance shrinkage energy. An energy hole comprising a reactant ion that is spontaneously regenerated following an endothermic electron ionization reaction of energy equal to the resonance shrinkage energy is hereafter referred to as an electrocatalytic ion. An energy hole comprising two reactants that are spontaneously regenerated following the an endothermic electron transfer reaction between the two species wherein the differences in their ionization energies is equal to the resonance shrinkage energy is hereafter referred to as an electrocatalytic couple.

The present invention of an electrolytic cell energy reactor, pressurized gas energy reactor, and a gas discharge energy reactor, comprises: a source of hydrogen; one of a solid, molten, liquid, and gaseous source of energy holes; a vessel containing hydrogen and the source of energy holes wherein the shrinkage reaction occurs by contact of the hydrogen with the source of energy holes; and a means for removing, the (molecular) lower-energy hydrogen so as to prevent an exothermic shrinkage reaction from coming to equilibrium. The present invention further comprises methods and structures for repeating this shrinkage reaction to produce shrunken atoms (molecules) to provide new materials with novel properties such as high thermal stability.

2. Description of the Related Art

Existing atomic models and theories are unable to explain certain observed physical phenomena. The Schrödinger wavefunctions of the hydrogen atom, for example, do not explain the extreme ultraviolet emission spectrum of the interstellar medium or that of the Sun, as well as the phenomenon of anomalous heat release from hydrogen in certain electrolytic cells having a potassium carbonate electrolyte or certain gas energy cells having a hydrogen spillover catalyst comprising potassium nitrate with the production of lower-energy hydrogen atoms and molecules, which is part of the present invention. Thus, advances in energy production and materials have been largely limited to laboratory discoveries having limited or sub-optimal commercial application.

SUMMARY OF THE INVENTION

The present invention comprises methods and apparatuses for releasing heat energy from hydrogen atoms (molecules) by stimulating their electrons to relax to quantized potential energy levels below that of the "ground state" via electron transfer reactions of reactants including electrochemical reactant(s) (electrocatalytic ion(s) or couple(s)) which remove energy from the hydrogen atoms (molecules) to stimulate these transitions. In addition, this application includes methods and apparatuses to enhance the power output by enhancing the reaction rate- the rate of the formation of the lower-energy hydrogen. The present invention further comprises a hydrogen spillover catalyst, a multifunctionality material having a functionality which dissociates molecular hydrogen to provide free hydrogen atoms which spill over to a functionality which supports mobile free hydrogen atoms and a functionality which can be a source of the energy holes. The energy reactor includes one of an electrolytic cell, a pressurized hydrogen gas cell, and a hydrogen gas discharge cell.

A preferred pressurized hydrogen gas energy reactor comprises a vessel; a source of hydrogen; a means to control the pressure and flow of hydrogen into the vessel; a material to dissociate the molecular hydrogen into atomic hydrogen, and a material which can be a source of energy holes in the gas phase. The gaseous source of energy holes includes those that sublime, boil, and/or are volatile at the elevated operating temperature of the gas energy reactor wherein the shrinkage reaction occurs in the gas phase.

The present invention further comprises methods and apparatuses for repeating a shrinkage reaction according to the present invention to cause energy release and to provide shrunken atoms and molecules with novel properties such as high thermal stability, and low reactivity. The lower-energy state atoms and molecules are useful for heat transfer, cryogenic applications, as a buoyant gas. as a medium in an engine such as a Sterling engine or a turbine, as a general replacement for helium, and as a refrigerant by absorbing energy including heat energy as the electrons are excited back to a higher energy level.

Below "Ground State" Transitions of Hydrogen Atoms

A novel atomic theory is disclosed in Mills, R., *The Grand Unified Theory of Classical Quantum Mechanics*, (1995), Technomic Publishing Company, Lancaster, Pa. provided by HydroCatalysis Power Corporation, Great Valley Corporate Center, 41 Great Valley Parkway, Malvern, Pa. 19355; *The Unification of Spacetime, the Forces, Matter, and Energy*, Mills, R., Technomic Publishing Company, Lancaster, Pa., (1992); *The Grand Unified Theory*, Mills, R. and Farrell, J., Science Press, Ephrata, Pa., (1990); Mills, R., Kneizys, S., Fusion Technology, 210, (1991), pp. 65–81; Mills, R., Good, W., Shaubach, R., "Dihydrino Molecule Identification", Fusion Technology, 25, 103 (1994); Mills, R., Good, W., "Fractional Quantum Energy Levels of Hydrogen", Fusion Technology, Vol. 28. No. 4, November, (1995), pp. 1697–1719, and in my previous U.S. patent applications entitled "Energy/Matter Conversion Methods and Structures", Ser. No. 08/467,051 filed on Jun. 6, 1995 which is a continuation-in-part application of Ser. No. 08/416,040 filed on Apr. 3, 1995 which is a continuation-in-part application of Ser. No. 08/107,357 filed on Aug. 16, 1993, which is a continuation-in-part application of Ser. No. 08/075,102 (Dkt. 99437) filed on Jun. 11, 1993, which is a continuation-in-part application of Ser. No. 07/626,496 filed on Dec. 12, 1990 which is a continuation-in-part application of Ser. No. 07/345,628 filed Apr. 28, 1989 which is a continuation-in-part application of Ser. No. 07/341,733 filed Apr. 21, 1989 which are all incorporated herein by this reference.

Fractional Quantum Energy Levels of Hydrogen

A number of experimental observations given in the Experimental Section below lead to the conclusion that atomic hydrogen can exist in fractional quantum states that are at lower energies than the traditional "ground" (n=1) state. For example, existence of fractional-quantum-energy-level hydrogen atoms, hereafter called hydrinos, provides an explanation for the soft X-ray emissions of the dark interstellar medium observed by Labov and Bowyer [S. Labov and S. Bowyer, Astrophysical Journal, 371 (1991) 810] and an explanation for the soft X-ray emissions of the Sun [Thomas, R. J., Neupert, W., M., Astrophysical Journal Supplement Series, Vol. 91, (1994), pp. 461–482; Malinovsky, M., Heroux, L., Astrophysical Journal, Vol. 181, (1973), pp. 1009–1030; Noyes, R., *The Sun, Our Star*, Harvard University Press, Cambridge, Ma., (1982), p. 172; Phillips, J. H., *Guide to the Sun*, Cambridge University Press, Cambridge, Great Britain, (1992), pp. 118–119; 120–121; 144–145].

J. J. Balmer showed in 1885 that the frequencies for some of the lines observed in the emission spectrum of atomic hydrogen could be expressed with a completely empirical relationship. This approach was later extended by J. R. Rydberg, who showed that all of the spectral lines of atomic hydrogen were given by the equation:

$$\bar{\nu} = R\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right) \tag{1}$$

where R=109,677 cm^{-1}, n_f=1,2,3, . . . , n_i=2,3,4, . . . , and $n_i > n_f$. Niels Bohr, in 1913, developed a theory for atomic hydrogen that gave energy levels in agreement with Rydberg's equation. An identical equation, based on a totally different theory for the hydrogen atom, was developed by E. Schrödinger, and independently by W. Heisenberg, in 1926.

$$E_n = -\frac{e^2}{n^2 8\pi\varepsilon_o a_H} = \frac{13.598\text{ eV}}{n^2} \tag{2a}$$

$$n = 1, 2, 3, \ldots \tag{2b}$$

where a_H is the Bohr radius for the hydrogen atom (52.947 pm), e is the magnitude of the charge of the electron, and ϵ_o is the vacuum permittivity. Mills' theory predicts that Eq. (2b), should be replaced by Eq. (2c).

$$n = 1, 2, 3, \ldots, \text{and}, n = \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots \tag{2c}$$

The quantum number n=1 is routinely used to describe the "ground" electronic state of the hydrogen atom. Mills [Mills, R., *The Grand Unified Theory of Classical Quantum Mechanics*, (1995), Technomic Publishing Company, Lancaster, Pa.] in a recent advancement of quantum mechanics has shown that the n=1 state is the "ground" state for "pure" photon transitions (the n=1 state can absorb a photon and go to an excited electronic state, but it cannot release a photon and go to a lower-energy electronic state). However, an electron transition from the ground state to a lower-energy state is possible by a "resonant collision" mechanism. These lower-energy states have fractional quantum numbers,

$$n = \frac{1}{\text{integer}}.$$

Processes that occur without photons and that require collisions are common. For example, the exothermic chemical reaction of H+H to form H_2 does not occur with the emission of a photon. Rather, the reaction requires a collision with a third body, M, to remove the bond energy–$H+H+M \rightarrow H_2+M$. The third body distributes the energy from the exothermic reaction, and the end result is the H_2 molecule and an increase in the temperature of the system. Similarly, the n=1 state of hydrogen and the

$$n = \frac{1}{\text{integer}}$$

states of hydrogen are nonradiative, but a transition between two nonradiative states is possible via a resonant collision, say n=1 to n=½. In these cases, during the collision the electron couples to another electron transition or electron transfer reaction which can absorb the exact amount of energy that must be removed from the hydrogen atom, a resonant energy sink called an energy hole. The end result is a lower-energy state for the hydrogen and increase in temperature of the system.

Wave Equation Solutions of the Hydrogen Atom

Recently, Mills [Mills, R., *The Grand Unified Theory of Classical Quantum Mechanics*, (1995), Technomic Publishing Company, Lancaster, Pa.] has built on the work generally known as quantum mechanics by deriving a new atomic theory based on first principles. The novel theory hereafter referred to as Mills' theory unifies Maxwell's Equations, Newton's Laws, and Einstein's General and Special Relativity. The central feature of this theory is that all particles (atomic-size and macroscopic particles) obey the same physical laws. Whereas Schrödinger postulated a boundary condition: $\Psi \rightarrow 0$ as $r \rightarrow \infty$, the boundary condition in Mills' theory was derived from Maxwell's equations [Haus, H. A., "On the radiation from point charges", American Journal of Physics, 54, (1986), pp. 1126–1129.]:

> For non-radiative states, the current-density function must not possess space-time Fourier components that are synchronous with waves traveling at the speed of light.

Application of this boundary condition leads to a physical model of particles, atoms, molecules, and, in the final analysis, cosmology. The closed-form mathematical solutions contain fundamental constants only, and the calculated values for physical quantities agree with experimental observations. In addition, the theory predicts that Eq. (2b), should be replaced by Eq. (2c).

Bound electrons are described by a charge-density (mass-density) function which is the product of a radial delta function ($f(r)=\delta(r-r_n)$), two angular functions (spherical harmonic functions), and a time harmonic function. Thus, an electron is a spinning, two-dimensional spherical surface, hereafter called an electron orbitsphere, that can exist in a bound state at only specified distances from the nucleus. More explicitly, the orbitsphere comprises a two dimensional spherical shell of moving charge. The corresponding current pattern of the orbitsphere comprises an infinite series of correlated orthogonal great circle current loops. The current pattern (shown in FIG. 1.4 of Mills [Mills, R., *The Grand Unified Theory of Classical Quantum Mechanics*, (1995), Technomic Publishing Company, Lancaster, Pa.]) is generated over the surface by two orthogonal sets of an infinite series of nested rotations of two orthogonal great circle current loops where the coordinate axes rotate with the two orthogonal great circles. Each infinitesimal rotation of the infinite series is about the new x-axis and new y-axis which results from the preceding such rotation. For each of the two sets of nested rotations, the angular sum of the rotations about each rotating x-axis and y-axis totals $\sqrt{2}\,\pi$ radians. The current pattern gives rise to the phenomenon corresponding to the spin quantum number.

The total function that describes the spinning motion of each electron orbitsphere is composed of two functions. One function, the spin function, is spatially uniform over the orbitsphere, spins with a quantized angular velocity, and gives rise to spin angular momentum. The other function, the modulation function, can be spatially uniform—in which case there is no orbital angular momentum and the magnetic moment of the electron orbitsphere is one Bohr magneton—or not spatially uniform—in which case there is orbital angular momentum. The modulation function also rotates with a quantized angular velocity. Numerical values for the angular velocity, radii of allowed orbitspheres. energies, and associated quantities are calculated by Mills.

Orbitsphere radii are calculated by setting the centripetal force equal to the electric and magnetic forces.

The orbitsphere is a resonator cavity which traps photons of discrete frequencies. The radius of an orbitsphere increases with the absorption of electromagnetic energy. The solutions to Maxwell's equations for modes that can be excited in the orbitsphere resonator cavity give rise to four quantum numbers, and the energies of the modes are the experimentally known hydrogen spectrum.

Excited states are unstable because the charge-density function of the electron plus photon have a radial doublet function component which corresponds to an electric dipole. The doublet possesses spacetime Fourier components synchronous with waves traveling at the speed of light; thus it is radiative. The charge-density function of the electron plus photon for the n=1 principle quantum state of the hydrogen atom as well as for each of the

$$n = \frac{1}{\text{integer}}$$

states mathematically is purely a radial delta function. The delta function does not possess spacetime Fourier components synchronous with waves traveling at the speed of light; thus, each is nonradiative.

Catalytic Lower-energy Hydrogen Electronic Transitions

Comparing transitions between below "ground" (fractional quantum) energy states as opposed to transitions between excited (integer quantum) energy states, it can be appreciated that the former are not effected by photons; whereas, the latter are. Transitions are symmetric with respect to time. Current density functions which give rise to photons according to the nonradiative boundary condition of Mills [Mills, R., *The Grand Unified Theory of Classical Quantum Mechanics*, (1995), Technomic Publishing Company, Lancaster, Pa.] are created by photons in the reverse process. Excited (integer quantum) energy states correspond to this case. And, current density functions which do not give rise to photons according to the nonradiative boundary condition are not created by photons in the reverse process. Below "ground" (fractional quantum) energy states correspond to this case. But, atomic collisions can cause a stable state to undergo a transition to the next stable state. The transition between two stable nonradiative states effected by a collision with an resonant energy sink is analogous to the reaction of two atoms to form a diatomic molecule which requires a third-body collision to remove the bond energy [N. V. Sidgwick, *The Chemical Elements and Their Compounds*, Volume I, Oxford, Clarendon Press, (1950), p. 17].

Energy Hole Concept

The nonradiative boundary condition of Mills and the relationship between the electron and the photon give the "allowed" hydrogen energy states which are quantized as a function of the parameter n. Each value of n corresponds to an allowed transition effected by a resonant photon which excites the electronic transition. In addition to the traditional integer values (1, 2, 3, . . . ,) of n, values of fractions are allowed which correspond to transitions with an increase in the central field (charge) and decrease in the size of the hydrogen atom. This occurs, for example, when the electron couples to another electronic transition or electron transfer reaction which can absorb energy, an energy sink. This is the absorption of an energy hole. The absorption of an energy hole destroys the balance between the centrifugal force and the increased central electric force. As a result, the electron undergoes a transition to a lower energy nonradiative state.

From energy conservation, the resonance energy hole of a hydrogen atom which excites resonator modes of radial dimensions

$$\frac{a_H}{m+1}$$

is

$$mX27.2\ \text{eV where } m=1,2,3,4,\ldots \tag{3}$$

After resonant absorption of the energy hole, the radius of the orbitsphere, a_H, shrinks to

$$\frac{a_H}{m+1}$$

and after p cycles of resonant shrinkage, the radius is

$$\frac{a_H}{mp+1}.$$

In other words, the radial ground state field can be considered as the superposition of Fourier components. The removal of negative Fourier components of energy mX27.2 eV, where m is an integer increases the positive central electric field inside the spherical shell by m times the charge of a proton. The resultant electric field is a time-harmonic solution of Laplace's Equations in spherical coordinates. In this case, the radius at which force balance and nonradiation are achieved is

$$\frac{a_H}{m+1}$$

where m is an integer. In decaying to this radius from the "ground" state, a total energy of $[(m+1)^2-1^2]$X13.6 eV is released. The transition between two stable nonradiative states effected by a collision with an energy hole is analogous to the reaction of two atoms to form a diatomic molecule which requires a third body collision to remove the bond energy [N. V. Sidgwick, *The Chemical Elements and Their Compounds*, Volume I, Oxford, Clarendon Press, (1950), p. 17]. The total energy well of the hydrogen atom is shown in FIG. **1**. The exothermic reaction involving transitions from one potential energy level to a lower level is hereafter referred to as HydroCatalysis.

A hydrogen atom with its electron in a lower than "ground state" energy level corresponding to a fractional quantum number is hereafter referred to as a hydrino atom. The designation for a hydrino atom of radius

$$\frac{a_o}{p}$$

where p is an integer is

$$H\left[\frac{a_o}{p}\right].$$

The size of the electron orbitsphere as a function of potential energy is given in FIG. **2**.

An efficient catalytic system that hinges on the coupling of three resonator cavities involves potassium. For example, the second ionization energy of potassium is 31.63 eV. This energy hole is obviously too high for resonant absorption. However, K^+ releases 4.34 eV when it is reduced to K. The combination of K^+ to K^{2+} and K^+ to K, then, has a net energy change of 27.28 eV.

$$27.28\text{ eV} + K^+ + K^+ + H\left[\frac{a_H}{p}\right] \to K + K^{2+} + H\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV} \tag{4}$$

$$K + K^{2+} \to K^+ + K^+ + 27.28\text{ eV} \tag{5}$$

And, the overall reaction is

$$H\left[\frac{a_H}{p}\right] \to H\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV} \tag{6}$$

Note that the energy given off as the atom shrinks is much greater than the energy lost to the energy hole. Also, the energy released is large compared to conventional chemical reactions.

Disproportionation of Energy States

Lower-energy hydrogen atoms, hydrinos, can act as a source of energy holes that can cause resonant shrinkage because the excitation and/or ionization energies are mX27.2 eV (Eq. (3)). For example, the equation for the absorption of an energy hole of 27.21 eV, m=1 in Eq. (3), during the shrinkage cascade for the third cycle of the hydrogen-type atom,

$$H\left[\frac{a_H}{3}\right],$$

with the hydrogen-type atom,

$$H\left[\frac{a_H}{2}\right],$$

that is ionized as the source of energy holes that cause resonant shrinkage is represented by

$$27.21\text{ eV} + H\left[\frac{a_H}{2}\right] + H\left[\frac{a_H}{3}\right] \to H^+ + e^- + H\left[\frac{a_H}{4}\right] + [4^2 - 3^2] \times 13.6\text{ eV} - 27.21\text{ eV} \tag{7}$$

$$H^+ + e^- \to H\left[\frac{a_H}{1}\right] + 13.6\text{ eV} \tag{8}$$

And, the overall reaction is

$$H\left[\frac{a_H}{2}\right] + H\left[\frac{a_H}{3}\right] \to H\left[\frac{a_H}{1}\right] + H\left[\frac{a_H}{4}\right] + [4^2 - 3^2 - 4] \times 13.6\text{ eV} + 13.6\text{ eV} \tag{9}$$

The general equation for the absorption of an energy hole of 27.21 eV, m=1 in Eq. (3), during the shrinkage cascade for the pth cycle of the hydrogen-type atom,

$$H\left[\frac{a_H}{p}\right],$$

with the hydrogen-type atom,

$$H\left[\frac{a_H}{m'}\right],$$

that is ionized as the source of energy holes that cause resonant shrinkage is represented by

$$27.21\text{ eV} + H\left[\frac{a_H}{m'}\right] + H\left[\frac{a_H}{p}\right] \to H^+ + e^- + H\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV} - (m'^2 - 2) \times 13.6\text{ eV} \tag{10}$$

$$H^+ + e^- \to H\left[\frac{a_H}{1}\right] + 13.6\text{ eV} \tag{11}$$

And, the overall reaction is

$$H\left[\frac{a_H}{m'}\right] + H\left[\frac{a_H}{p}\right] \to H\left[\frac{a_H}{1}\right] + H\left[\frac{a_H}{(p+1)}\right] + [2p + 1 - m'^2] \times 13.6\text{ eV} + 13.6\text{ eV} \tag{12}$$

Transitions to nonconsecutive energy levels involving the absorption of an energy hole of an integer multiple of 27.21 eV are possible. Lower-energy hydrogen atoms, hydrinos, can act as a source of energy holes that can cause resonant shrinkage with the absorption of an energy hole of mX27.2 eV (Eq. (3)). Thus, the shrinkage cascade for the pth cycle of the hydrogen-type atom,

$$H\left[\frac{a_H}{p}\right],$$

with the hydrogen-type atom,

$$H\left[\frac{a_H}{m'}\right],$$

that is ionized as the source of energy holes that cause resonant shrinkage is represented by

$$m \times 27.21\text{ eV} + H\left[\frac{a_H}{m'}\right] + H\left[\frac{a_H}{p}\right] \rightarrow \tag{13}$$

$$H^+ + e^- + H\left[\frac{a_H}{(p+m)}\right] + [(p+m)^2 - p^2 - (m'^2 - 2m)] \times 13.6\text{ eV}$$

$$H^+ + e^- \rightarrow H\left[\frac{a_H}{1}\right] + 13.6\text{ eV} \tag{14}$$

And, the overall reaction is

$$H\left[\frac{a_H}{m'}\right] + H\left[\frac{a_H}{p}\right] \rightarrow \tag{15}$$

$$H\left[\frac{a_H}{1}\right] + H\left[\frac{a_H}{(p+m)}\right] + [2pm + m^2 - m'^2] \times 13.6\text{ eV} + 13.6\text{ eV}$$

Hydrogen is a source of energy holes. The ionization energy of hydrogen is 13.6 eV. Disproportionation can occur between three hydrogen atoms whereby two atoms provide an energy hole of 27.21 eV for the third hydrogen atom. Thus, the shrinkage cascade for the pth cycle of the hydrogen-type atom,

$$H\left[\frac{a_H}{p}\right],$$

with two hydrogen atoms,

$$H\left[\frac{a_H}{1}\right],$$

as the source of energy holes that cause resonant shrinkage is represented by

$$27.21\text{ eV} + 2H\left[\frac{a_H}{1}\right] + H\left[\frac{a_H}{p}\right] \rightarrow \tag{16}$$

$$2H^+ + 2e^- + H\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV}$$

$$2H^+ + 2e^- \rightarrow 2H\left[\frac{a_H}{1}\right] + 27.21\text{ eV} \tag{17}$$

And, the overall reaction is

$$H\left[\frac{a_H}{p}\right] \rightarrow H\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p] \times 13.6\text{ eV} \tag{18}$$

The spectral lines from dark interstellar medium and the majority of the solar power can be attributed to disproportionation reactions as given in the Spectral Data of Hydrinos from the Dark Interstellar Medium and from the Sun Section of Mills [Mills, R., *The Grand Unified Theory of Classical Quantum Mechanics*, (1995), Technomic Publishing Company, Lancaster, Pa.]. This assignment resolves the mystery of dark matter, the solar neutrino problem, and the mystery of the cause of sunspots and other solar activity and why the Sun emits X-rays. It also provides the reason for the abrupt change in the speed of sound and transition from "radiation zone" to "convection zone" at a radius of 0.7 the solar radius, 0.7 R_s as summarized in Example 4 below.

Energy Hole (Atomic Hydrogen)

In a preferred embodiment, energy holes, each of approximately 27.21 eV, are provided by electron transfer reactions of reactants including electrochemical reactant(s) (electrocatalytic ion(s) or couple(s)) which cause heat to be released from hydrogen atoms as their electrons are stimulated to relax to quantized potential energy levels below that of the "ground state". The energy removed by an electron transfer reaction, energy hole, is resonant with the hydrogen energy released to stimulate this transition. The source of hydrogen atoms can be the production on the surface of a cathode during electrolysis of water in the case of an electrolytic energy reactor and hydrogen gas or a hydride in the case of a pressurized gas energy reactor or gas discharge energy reactor.

Below "Ground State" Transitions of Hydrogen-type Molecules and Molecular Ions

Two hydrogen atoms react to form a diatomic molecule, the hydrogen molecule.

$$2H[a_H] \rightarrow H_2[2c' = \sqrt{2}\,a_o] \tag{19}$$

where 2c' is the internuclear distance. Also, two hydrino atoms react to form a diatomic molecule, hereafter called a dihydrino molecule.

$$2H\left[\frac{a_H}{p}\right] \rightarrow H_2^*\left[2c' = \frac{\sqrt{2}\,a_o}{p}\right] \tag{20}$$

where p is an integer.

The central force equation for hydrogen-type molecules has orbital solutions which are circular, elliptic, parabolic, or hyperbolic. The former two types of solutions are associated with atomic and molecular orbitals. These solutions are nonradiative if the boundary condition for nonradiation given in the One Electron Atom Section of *The Unification of Spacetime, the Forces, Matter, and Energy*, Mills, R., Technomic Publishing Company, Lancaster, Pa., (1992), is met. The mathematical formulation for zero radiation is that the function that describes the motion of the electron must not possess space-time Fourier components that are synchronous with waves traveling at the speed of light. The boundary condition for the orbitsphere is met when the angular frequencies are

$$\omega_n = \frac{\hbar}{m_e r_n^2} \tag{21}$$

As demonstrated in the One Electron Atom Section of *The Unification of Spacetime, the Forces, Matter, and Energy*, Mills, R., Technomic Publishing Company, Lancaster, Pa., (1992), this condition is met for the product function of a radial Dirac delta function and a time harmonic function where the angular frequency, ω, is constant and given by Eq. (21).

$$\omega_n = \frac{\hbar}{m_e r_n^2} = \frac{\frac{\pi L}{m_e}}{A} \tag{22}$$

where L is the angular momentum and A is the area of the closed geodesic orbit. Consider the solution of the central force equation comprising the product of a two dimensional ellipsoid and a time harmonic function. The spatial part of the product function is the convolution of a radial Dirac delta function with the equation of an ellipsoid. The Fourier transform of the convolution of two functions is the product of the individual Fourier transforms of the functions: thus, the boundary condition is met for an ellipsoidal-time harmonic function when

$$\omega_n = \frac{\pi\hbar}{m_e A} = \frac{\hbar}{m_e ab} \tag{23}$$

where the area of an ellipse is

$$A=\pi ab \tag{24}$$

where 2b is the length of the semiminor axis and 2a is the length of the semimajor axis. The geometry of molecular hydrogen is elliptic with the internuclear axis as the principle axis; thus, the electron orbital is a two dimensional ellipsoidal-time harmonic function. The mass follows geodesics time harmonically as determined by the central field of the protons at the foci. Rotational symmetry about the internuclear axis further determines that the orbital is a prolate spheroid. In general, ellipsoidal orbits of molecular bonding, hereafter referred to as ellipsoidal molecular orbitals (M. O.'s), have the general equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{25}$$

The semiprinciple axes of the ellipsoid are a, b, c.

In ellipsoidal coordinates the Laplacian is

$$(\eta - \zeta)R_\xi \frac{\delta}{\delta\xi}\left(R_\xi \delta \frac{\phi}{\delta\xi}\right) + (\zeta - \xi)R_\eta \frac{\delta}{\delta\eta}\left(R_\eta \delta \frac{o}{\delta\eta}\right) + (\xi - \eta)R_\zeta \frac{\delta}{\delta\zeta}\left(R_\zeta \delta \frac{\phi}{\delta\zeta}\right) = 0 \tag{26}$$

An ellipsoidal M. O. is equivalent to a charged conductor whose surface is given by Eq. (25). It carries a total charge q, and it's potential is a solution of the Laplacian in ellipsoidal coordinates, Eq. (26).

Excited states of orbitspheres are discussed in the Excited States of the One Electron Atom (Quantization) Section of *The Unification of Spacetime, the Forces, Matter, and Energy*, Mills, R., Technomic Publishing Company, Lancaster, Pa., (1992). In the case of ellipsoidal M. O.'s, excited electronic states are created when photons of discrete frequencies are trapped in the ellipsoidal resonator cavity of the M. O. The photon changes the effective charge at the M. O. surface where the central field is ellipsoidal. Force balance is achieved at a series of ellipsoidal equipotential two dimensional surfaces confocal with the ground state ellipsoid. The trapped photons are solutions of the Laplacian in ellipsoidal coordinates, Eq. (26).

As is the case with the orbitsphere, higher and lower energy states are equally valid. The photon standing wave in both cases is a solution of the Laplacian in ellipsoidal coordinates. For an ellipsoidal resonator cavity, the relationship between an allowed circumference, 4aE, and the photon standing wavelength, λ, is

$$4aE=n\lambda \tag{27}$$

where n is an integer and where

$$k = \frac{\sqrt{a^2 - b^2}}{a} \tag{28}$$

is used in the elliptic integral E of Eq. (27). Applying Eqs. (27) and (28), the relationship between an allowed angular frequency given by Eq. (23) and the photon standing wave angular frequency, ω, is:

$$\frac{\pi\hbar}{m_e A} = \frac{\hbar}{m_e n a_1 n b_1} = \frac{\hbar}{m_e a_n b_n} = \frac{1}{n^2}\omega_1 = \omega_n \tag{29}$$

where n=1,2,3,4, . . .

$$n = \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$$

ω_1 is the allowed angular frequency for n=1 a_1 and b_1 are the allowed semimajor and semiminor axes for n=1

From Eq. (29), the magnitude of the elliptic field corresponding to a below "ground state" transition of the hydrogen molecule is an integer. The potential energy equations of hydrogen-type molecules are

$$V_e = \frac{-2pe^2}{8\pi\varepsilon_o\sqrt{a^2 - b^2}} \ln\frac{a + \sqrt{a^2 - b^2}}{a - \sqrt{a^2 - b^2}} \tag{30}$$

$$V_p = \frac{p}{8\pi\varepsilon_o} \frac{e^2}{\sqrt{a^2 - b^2}} \tag{31}$$

where

$$a = \frac{a_o}{p} \tag{32}$$

$$b = \frac{1}{p\sqrt{2}} a_o \tag{33}$$

$$c' = \sqrt{a^2 - b^2} = \frac{\sqrt{2}\, a_o}{2p} \tag{34}$$

and where p is an integer. From energy conservation, the resonance energy hole of a hydrogen-type molecule which causes the transition

$$\mathrm{H}_2^*\left[2c' = \frac{\sqrt{2}\, a_o}{p}\right] \rightarrow \mathrm{H}_2^*\left[2c' = \frac{\sqrt{2}\, a_o}{p + m}\right] \tag{35}$$

where m and p are integers. During the transition, the elliptic field is increased from magnitude p to magnitude p+m. The corresponding potential energy change equals the energy absorbed by the energy hole.

$$\text{Energy hole} = -V_e - V_p = mp^2 \text{X} 48.6 \text{ eV} \tag{37}$$

Further energy is released by the hydrogen-type molecule as the internuclear distance "shrinks". The total energy, E_T, released during the transition is

$$E_T = -13.6\text{ eV}\left[\left(2(m+p)^2\sqrt{2} - (m+p)^2\sqrt{2} + \frac{(m+p)^2\sqrt{2}}{2}\right)\ln\frac{\sqrt{2}+1}{\sqrt{2}-1} - (m+p)^2\sqrt{2}\right] + 13.6\text{ eV}\left[\left(2p^2\sqrt{2} - p^2\sqrt{2} + \frac{p^2\sqrt{2}}{2}\right)\ln\frac{\sqrt{2}+1}{\sqrt{2}-1} - p^2\sqrt{2}\right] \tag{38}$$

A schematic drawing of the total energy well of hydrogen-type molecules and molecular ions is given in FIG. **3**. The exothermic reaction involving transitions from one potential energy level to a lower level below the "ground state" is also hereafter referred to as HydroCatalysis.

A hydrogen-type molecule with its electrons in a lower than "ground state" energy level corresponding to a fractional quantum number is hereafter referred to as a dihydrino molecule. The designation for a dihydrino molecule of internuclear distance,

$$2c' = \frac{\sqrt{2}\,a_o}{p}$$

where p is an integer, is

$$H_2^*\left[2c' = \frac{\sqrt{2}\,a_o}{p}\right].$$

A schematic drawing of the size of hydrogen-type molecules as a function of total energy is given in FIG. **4**.

The magnitude of the elliptic field corresponding to the first below "ground state" hydrogen-type molecule is 2. From energy conservation, the resonance energy hole of a hydrogen molecule which excites the transition of the hydrogen molecule with internuclear distance

$$2c' = \sqrt{2}\,a_o$$

to the first below "ground state" with internuclear distance

$$2c' = \frac{1}{\sqrt{2}}a_o$$

is given by Eqs. (30) and (31) where the elliptic field is increased from magnitude one to magnitude two:

$$V_e = \frac{-2e^2}{8\pi\varepsilon_o\sqrt{a^2-b^2}}\ln\frac{a+\sqrt{a^2-b^2}}{a-\sqrt{a^2-b^2}} = -67.813\text{ eV} \tag{39}$$

$$V_p = \frac{e^2}{8\pi\varepsilon_o\sqrt{a^2-b^2}} = 19.23\text{ eV} \tag{40}$$

$$\text{Energy hole} = -V_e - V_p = m \times 48.6\text{ eV} \tag{41}$$

In other words, the ellipsoidal "ground state" field of the hydrogen molecule can be considered as the superposition of Fourier components. The removal of negative Fourier components of energy

$$m\text{X}48.6\text{ eV} \tag{42}$$

where m is an integer, increases the positive electric field inside the ellipsoidal shell by m times the charge of a proton at each focus. The resultant electric field is a time harmonic solution of the Laplacian in ellipsoidal coordinates. The hydrogen molecule with internuclear distance

$$2c' = \sqrt{2}\,a_o$$

is caused to undergo a transition to a below "ground state" level, and the internuclear distance for which force balance and nonradiation are achieved is

$$2c' = \frac{\sqrt{2}\,a_o}{1+m}.$$

In decaying to this internuclear distance from the "ground state", a total energy of

$$E_T = -13.6\text{ eV}\left[\left(2(1+m)^2\sqrt{2} - (1+m)^2\sqrt{2} + \frac{(1+m)^2\sqrt{2}}{2}\right)\ln\frac{\sqrt{2}+1}{\sqrt{2}-1} - (1+m)^2\sqrt{2}\right] + 13.6\text{ eV}\left[\left(2\sqrt{2} - \sqrt{2} + \frac{\sqrt{2}}{2}\right)\ln\frac{\sqrt{2}+1}{\sqrt{2}-1} - \sqrt{2}\right] \tag{43}$$

is released.

Energy Hole (Molecular Hydrogen)

In a preferred embodiment, energy holes, each of approximately mX48.6 eV, are provided by electron transfer reactions of reactants including electrochemical reactant(s) (electrocatalytic ion(s) or couple(s)) which cause heat to be released from hydrogen molecules as their electrons are stimulated to relax to quantized potential energy levels below that of the "ground state". The energy removed by an electron transfer reaction, energy hole, is resonant with the hydrogen energy released to stimulate this transition. The source of hydrogen molecules can be the production on the surface of a cathode during electrolysis of water in the case of an electrolytic energy reactor and hydrogen gas or a hydride in the case of a pressurized gas energy reactor or gas discharge energy reactor.

Energy Reactor

The present invention of an electrolytic cell energy reactor, pressurized gas energy reactor, and a gas discharge energy reactor, comprises: a source of hydrogen; one of a solid, molten, liquid, and gaseous source of energy holes; a vessel containing hydrogen and the source of energy holes wherein the shrinkage reaction occurs by contact of the hydrogen with the source of energy holes; and a means for removing the (molecular) lower-energy hydrogen so as to prevent the exothermic shrinkage reaction from coming to equilibrium. The shrinkage reaction rate and net power output are increased by conforming the energy hole to match the resonance shrinkage energy. In general, power output can be optimized by controlling the temperature, pressure of the hydrogen gas, the source of the energy hole including the electrocatalytic ion or couple which provides the energy hole, the counterion of the electrocatalytic ion or couple, and the area of the surface on which the shrinkage reaction occurs. The present invention further comprises a hydrogen spillover catalyst, a multifunctionality material having a functionality which dissociates molecular hydrogen to provide free hydrogen atoms which spill over to a functionality which supports mobile free hydrogen atoms and a functionality which can be a source of the energy holes.

A preferred pressurized hydrogen gas energy reactor comprises a vessel; a source of hydrogen; a means to control the pressure and flow of hydrogen into the vessel; a material to dissociate the molecular hydrogen into atomic hydrogen, and a material which can be a source of energy holes in the gas phase. The gaseous source of energy holes includes those that sublime, boil, and/or are volatile at the elevated operating temperature of the gas energy reactor wherein the shrinkage reaction occurs in the gas phase.

Other objects, features, and characteristics of the present invention, as well as the methods of operation and the functions of the related elements, will become apparent upon consideration of the following description and the appended claims with reference to the accompanying drawings, all of which form a part of this specification, wherein like reference numerals designate corresponding parts in the various figures.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a schematic drawing of the total energy well of the hydrogen atom;

FIG. **2** is a schematic drawing of the size of electron orbitspheres as a function of potential energy;

FIG. **3** is a schematic drawing of the total energy wells of the hydrogen molecule, $H_2[2c'=\sqrt{2}a_o]$, the hydrogen molecular ion, $H_2[2c'=2a_o]^+$, the dihydrino molecule,

$$H_2^*\left[2c' = \frac{a_o}{\sqrt{2}}\right],$$

and the dihydrino molecular ion, $H_2^*[2c'=a_o]^+$;

FIG. **4** is a schematic drawing of the size of hydrogen-type molecules,

$$H_2^*\left[2c' = \frac{\sqrt{2}\,a_o}{p}\right],$$

as a function of total energy;

FIG. **5** is a schematic drawing of an energy reactor in accordance with the invention;

FIG. **6** is a schematic drawing of an electrolytic cell energy reactor in accordance with the present invention;

FIG. **7** is a schematic drawing of a pressurized gas energy reactor in accordance with the present invention;

FIG. **8** is a schematic drawing of a gas discharge energy reactor in accordance with the invention; and

FIG. **9** is a plot of the excess heat release from flowing hydrogen in the presence of nickel oxide powder containing strontium niobium oxide (Nb^{3+}/Sr^{2+} electrocatalytic couple) by the very accurate and reliable method of heat measurement, thermopile conversion of heat into an electrical output signal.

DETAILED DESCRIPTION OF THE PRESENTLY PREFERRED EMBODIMENTS

CATALYTIC ENERGY HOLE STRUCTURE FOR ATOMS

Single Electron Excited State

An energy hole is provided by the transition of an electron of a species to an excited state species including a continuum excited state(s) of atoms, ions, molecules, and ionic and molecular compounds. In one embodiment, the energy hole comprises the excited state transition of an electron of one species whereby the transition energy of the accepting species equals approximately mX27.21 eV where m is an integer.

Single Electron Transfer

An energy hole is provided by the transfer of an electron between participating species including atoms, ions, molecules, and ionic and molecular compounds. In one embodiment, the energy hole comprises the transfer of an electron from one species to another species whereby the sum of the ionization energy of the electron donating species minus the ionization energy or electron affinity of the electron accepting species equals approximately mX27.21 eV where m is an integer.

Single Electron Transfer (Two Species)

An efficient catalytic system that hinges on the coupling of three resonator cavities involves potassium. For example, the second ionization energy of potassium is 31.63 eV. This energy hole is obviously too high for resonant absorption. However, K^+ releases 4.34 eV when it is reduced to K. The combination of K^+ to K^{2+} and K^+ to K, then, has a net energy change of 27.28 eV; m=1 in Eq. (3).

$$27.28\text{ eV} + K^+ + K^+ + H\left[\frac{a_H}{p}\right] \to K + K^{2+} + H\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV} \tag{44}$$

$$K + K^{2+} \to K^+ + K^+ + 27.28\text{ eV} \tag{45}$$

And, the overall reaction is

$$H\left[\frac{a_H}{p}\right] \to H\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV} \tag{46}$$

Note that the energy given off as the atom shrinks is much greater than the energy lost to the energy hole. And, the energy released is large compared to conventional chemical reactions.

For sodium or sodium ions no electrocatalytic reaction of approximately 27.21 eV is possible. For example, 42.15 eV of energy is absorbed by the reverse of the reaction given in Eq. (45) where Na^+ replaces K^+:

$$Na^+ + Na^+ + 42.15\text{ eV} \to Na + Na^{2+} \tag{47}$$

Other less efficient catalytic systems hinge on the coupling of three resonator cavities. For example, the third ionization energy of palladium is 32.93 eV. This energy hole is obviously too high for resonant absorption. However, Li^+ releases 5.392 eV when it is reduced to Li. The combination of Pd^{2+} to Pd^{3+} and Li^+ to Li, then, has a net energy change of 27.54 eV.

$$27.54\text{ eV} + Li^+ + Pd^{2+} + H\left[\frac{a_H}{p}\right] \to Li + Pd^{3+} + H\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV} \tag{48}$$

$$Li + Pd^{3+} \to Li^+ + Pd^{2+} + 27.54\text{ eV} \tag{49}$$

And, the overall reaction is

$$\mathrm{H}\left[\frac{a_H}{p}\right] \rightarrow \mathrm{H}\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV} \tag{50}$$

Single Electron Transfer (One Species)

An energy hole is provided by the ionization of an electron from a participating species including an atom, an ion, a molecule, and an ionic or molecular compound to a vacuum energy level. In one embodiment, the energy hole comprises the ionization of an electron from one species to a vacuum energy level whereby the ionization energy of the electron donating species equals approximately mX27.21 eV where m is an integer.

Titanium is one of the catalysts (electrocatalytic ion) that can cause resonant shrinkage because the third ionization energy is 27.49 eV, m=1 in Eq. (3). Thus, the shrinkage cascade for the pth cycle is represented by

$$27.491\text{ eV} + \mathrm{Ti}^{2+} + \mathrm{H}\left[\frac{a_H}{p}\right] \rightarrow \mathrm{Ti}^{3+} + e^- + \mathrm{H}\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV} \tag{51}$$

$$\mathrm{Ti}^{3+} + e^- \rightarrow \mathrm{Ti}^{2+} + 27.491\text{ eV} \tag{52}$$

And, the overall reaction is

$$\mathrm{H}\left[\frac{a_H}{p}\right] \rightarrow \mathrm{H}\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV} \tag{53}$$

Rubidium is also a catalyst (electrocatalytic ion). The second ionization energy is 27.28 eV.

$$27.28\text{ eV} + \mathrm{Rb}^{+} + \mathrm{H}\left[\frac{a_H}{p}\right] \rightarrow \mathrm{Rb}^{2+} + e^- + \mathrm{H}\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV} \tag{54}$$

$$\mathrm{Rb}^{2+} + e^- \rightarrow \mathrm{Rb}^{+} + 27.28\text{ eV} \tag{55}$$

And, the overall reaction is

$$\mathrm{H}\left[\frac{a_H}{p}\right] \rightarrow \mathrm{H}\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV} \tag{56}$$

Other single electron transfer reactions to provide energy holes of approximately mX27.21 eV where m is an integer appear in my previous U.S. Patent Applications entitled "Energy/Matter Conversion Methods and Structures", Ser. No. 08/467,051 filed on Jun. 6, 1995 which is a continuation-in-part application of Ser. No. 08/416,040 filed on Apr. 3, 1995 which is a continuation-in-part application of Ser. No. 08/107,357 filed on Aug. 16, 1993, which is a continuation-in-part application of Ser. No. 08/075,102 (Dkt. 99437) filed on Jun. 11, 1993, which is a continuation-in-part application of Ser. No. 07/626,496 filed on Dec. 12, 1990 which is a continuation-in-part application of Ser. No. 07/345,628 filed Apr. 28, 1989 which is a continuation-in-part application of Ser. No. 07/341,733 filed Apr. 21, 1989, which are incorporated herein by reference.

Multiple Electron Transfer

An energy hole is provided by the transfer of multiple electrons between participating species including atoms, ions, molecules, and ionic and molecular compounds. In one embodiment, the energy hole comprises the transfer of t electrons from one or more species to one or more species whereby the sum of the ionization energies and/or electron affinities of the electron donating species minus the sum of the ionization energies and/or electron affinities of the electron acceptor species equals approximately mX27.21 eV where m and t are integers.

An energy hole is provided by the transfer of multiple electrons between participating species including atoms, ions, molecules, and ionic and molecular compounds. In one embodiment, the energy hole comprises the transfer of t electrons from one species to another whereby the t consecutive electron affinities and/or ionization energies of the electron donating species minus the t consecutive ionization energies and/or electron affinities of the electron acceptor equals approximately mX27.21 eV where m and t are integers.

In a preferred embodiment the electron acceptor species is an oxide such as MnO_x, AlO_x, SiO_x. A preferred molecular electron acceptor is oxygen, O_2.

Two Electron Transfer (One Species)

In an embodiment, a catalytic system that provides an energy hole hinges on the ionization of two electrons from an atom, ion, or molecule to a vacuum energy level such that the sum of two ionization energies is approximately 27.21 eV. Zinc is one of the catalysts (electrocatalytic atom) that can cause resonant shrinkage because the sum of the first and second ionization energies is 27.358 eV, m=1 in Eq. (3). Thus, the shrinkage cascade for the p th cycle is represented by

$$27.358\text{ eV} + \mathrm{Zn} + \mathrm{H}\left[\frac{a_o}{p}\right] \rightarrow \mathrm{Zn}^{2+} + 2e^- + \mathrm{H}\left[\frac{a_o}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV} \tag{57}$$

$$\mathrm{Zn}^{2+} + 2e^- \rightarrow \mathrm{Zn} + 27.358\text{ eV} \tag{58}$$

And, the overall reaction is

$$\mathrm{H}\left[\frac{a_H}{p}\right] \rightarrow \mathrm{H}\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV} \tag{59}$$

Two Electron Transfer (Two Species)

In another embodiment, a catalytic system that provides an energy hole hinges on the transfer of two electrons from an atom, ion, or molecule to another atom or molecule such that the sum of two ionization energies minus the sum of two electron affinities of the participating atoms, ions, and/or molecules is approximately 27.21 eV. A catalytic system that hinges on the transfer of two electrons from an atom to a molecule involves palladium and oxygen. For example, the first and second ionization energies of palladium are 8.34 eV and 19.43 eV, respectively. And, the first and second electron affinities of the oxygen molecule are 0.45 eV and 0.11 eV, respectively. The energy hole resulting from a two electron transfer is appropriate for resonant absorption. The combination of Pd to Pd^{2+} and O_2 to $\mathrm{O}_2^{\,2-}$, then, has a net energy change of 27.21 eV.

$$27.21\text{ eV} + \mathrm{Pd} + \mathrm{O}_2 + \mathrm{H}\left[\frac{a_o}{p}\right] \rightarrow \mathrm{Pd}^{2+} + \mathrm{O}_2^{2-} + \mathrm{H}\left[\frac{a_o}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV} \tag{60}$$

-continued

$$\mathrm{Pd}^{2+} + \mathrm{O}_2^{2-} \rightarrow \mathrm{Pd} + \mathrm{O}_2 + 27.21\ \mathrm{eV} \tag{61}$$

And, the overall reaction is

$$\mathrm{H}\left[\frac{a_H}{p}\right] \rightarrow \mathrm{H}\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\ \mathrm{eV} \tag{62}$$

Additional atoms, molecules, or compounds which could be substituted for O_2 are those with first and second electron affinities of approximately 0.45 eV and 0.11 eV, respectively, such as a mixed oxide (MnO_x, AlO_x, SiO_x) containing O to form O^{2-} or O_2 to form O_2^{2-}.

Two Electron Transfer (Two Species)

In another embodiment, a catalytic system that provides an energy hole hinges on the transfer of two electrons from an atom, ion, or molecule to another atom, ion, or molecule such that the sum of two ionization energies minus the sum of one ionization energy and one electron affinity of the participating atoms, ions, and/or molecules is approximately 27.21 eV. A catalytic system that hinges on the transfer of two electrons from an atom to an ion involves xenon and lithium. For example, the first and second ionization energies of xenon are 12.13 eV and 21.21 eV, respectively. And, the first ionization energy and the first electron affinity of lithium are 5.39 eV and 0.62 eV, respectively. The energy hole resulting from a two electron transfer is appropriate for resonant absorption. The combination of Xe to Xe^{2+} and Li^+ to Li^-, then, has a net energy change of 27.33 eV.

$$27.33\ \mathrm{eV} + \mathrm{Xe} + \mathrm{Li}^+ + \mathrm{H}\left[\frac{a_o}{p}\right] \rightarrow \mathrm{Xe}^{2+} + \mathrm{Li}^- + \mathrm{H}\left[\frac{a_o}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\ \mathrm{eV} \tag{63}$$

$$\mathrm{Xe}^{2+} + \mathrm{Li}^- \rightarrow \mathrm{Xe} + \mathrm{Li}^+ + 27.33\ \mathrm{eV} \tag{64}$$

And, the overall reaction is

$$\mathrm{H}\left[\frac{a_H}{p}\right] \rightarrow \mathrm{H}\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\ \mathrm{eV} \tag{65}$$

Two Electron Transfer (Two Species)

In another embodiment, a catalytic system that provides an energy hole hinges on the transfer of two electrons from an atom, ion, or molecule to another atom, ion, or molecule such that the sum of two ionization energies minus the sum of two ionization energies of the participating atoms and/or molecules is approximately 27.21 eV. A catalytic system that hinges on the transfer of two electrons from a first ion to a second ion involves silver(Ag^+) and silver (Ag^{2+}). For example, the second and third ionization energies of silver are 21.49 eV and 34.83 eV, respectively. And, the second and first ionization energies of silver are 21.49 eV and 7.58 eV, respectively. The energy hole resulting from a two electron transfer is appropriate for resonant absorption. The combination of Ag^+ to Ag^{3+} and Ag^{2+} to Ag, then, has a net energy change of 27.25 eV.

$$27.25\ \mathrm{eV} + \mathrm{Ag}^+ + \mathrm{Ag}^{2+} + \mathrm{H}\left[\frac{a_o}{p}\right] \rightarrow \mathrm{Ag}^{3+} + \mathrm{Ag} + \mathrm{H}\left[\frac{a_o}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\ \mathrm{eV} \tag{66}$$

$$\mathrm{Ag}^{3+} + \mathrm{Ag} \rightarrow \mathrm{Ag}^+ + \mathrm{Ag}^{2-} + 27.25\ \mathrm{eV} \tag{67}$$

And, the overall reaction is

$$\mathrm{H}\left[\frac{a_H}{p}\right] \rightarrow \mathrm{H}\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\ \mathrm{eV} \tag{68}$$

Three Electron Transfer (Two Species)

In another embodiment, a catalytic system that provides an energy hole hinges on the transfer of three electrons from an ion to another ion such that the sum of the electron affinity and two ionization energies of the first ion minus the sum of three ionization energies of the second ion is approximately 27.21 eV. A catalytic system that hinges on the transfer of three electrons from an ion to a second ion involves Li^- and Cr^{3+}. For example, the electron affinity, first ionization energy, and second ionization energy of lithium are 0.62 eV, 5.392 eV, and 75.638 eV, respectively. And, the third, second, and first ionization energies of Cr^{3+} are 30.96 eV, 16.50 eV, and 6.766 eV, respectively. The energy hole resulting from a three electron transfer is appropriate for resonant absorption. The combination of Li^- to Li^{2+} and Cr^{3+} to Cr, then, has a net energy change of 27.42 eV.

$$27.42\ \mathrm{eV} + \mathrm{Li}^- + \mathrm{Cr}^{3+} + \mathrm{H}\left[\frac{a_o}{p}\right] \rightarrow \mathrm{Li}^{2+} + \mathrm{Cr} + \mathrm{H}\left[\frac{a_o}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\ \mathrm{eV} \tag{69}$$

$$\mathrm{Li}^{2+} + \mathrm{Cr} \rightarrow \mathrm{Li}^- + \mathrm{Cr}^{3+} + 27.42\ \mathrm{eV} \tag{70}$$

And, the overall reaction is

$$\mathrm{H}\left[\frac{a_H}{p}\right] \rightarrow \mathrm{H}\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\ \mathrm{eV} \tag{71}$$

Three Electron Transfer (Two Species)

In another embodiment, a catalytic system that provides an energy hole hinges on the transfer of three electrons from an atom, ion, or molecule to another atom, ion, or molecule such that the sum of three consecutive ionization energies of the electron donating species minus the sum of three consecutive ionization energies of the electron accepting species is approximately 27.21 eV. A catalytic system that hinges on the transfer of three electrons from an atom to an ion involves Ag and Ce^{3+}. For example, the first, second, and third ionization energies of silver are 7.58 eV, 21.49 eV, and 34.83 eV, respectively. And, the third, second, and first ionization energies of Ce^{3+} are 20.20 eV, 10.85 eV, and 5.47 eV, respectively. The energy hole resulting from a three electron transfer is appropriate for resonant absorption. The combination of Ag to Ag^{3+} and Ce^{3+} to Ce, then, has a net energy change of 27.38 eV.

$$27.38\text{ eV} + \text{Ag} + \text{Ce}^{3+} + \text{H}\left[\frac{a_o}{p}\right] \rightarrow \text{Ag}^{3+} + \text{Ce} + \text{H}\left[\frac{a_o}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV} \tag{72}$$

$$\text{Ag}^{3+} + \text{Ce} \rightarrow \text{Ag} + \text{Ce}^{3+} + 27.38\text{ eV} \tag{73}$$

And, the overall reaction is

$$\text{H}\left[\frac{a_H}{p}\right] \rightarrow \text{H}\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV} \tag{74}$$

ADDITIONAL CATALYTIC ENERGY HOLE STRUCTURES

Single Electron Transfer

In a further embodiment, an energy hole of energy equal to the total energy released for a below "ground state" electronic transition of the hydrogen atom is provided by the transfer of an electron between participating species including atoms, ions, molecules, and ionic and molecular compounds. In one embodiment, the energy hole comprises the transfer of an electron from one species to another species whereby the sum of the ionization energy of the electron donating species minus the ionization energy or electron affinity of the electron accepting species equals approximately

$$\frac{m}{2}27.21\text{ eV};$$

where m is an integer.

For m=3 corresponding to the n=1 to n=½ transition, an efficient catalytic system that hinges on the coupling of three resonator cavities involves arsenic and calcium. For example, the third ionization energy of calcium is 50.908 eV. This energy hole is obviously too high for resonant absorption. However, As^+ releases 9.81 eV when it is reduced to As. The combination of Ca^{2+} to Ca^{3+} and As^+ to As, then, has a net energy change of 41.1 eV.

$$41.1\text{ eV} + \text{As}^+ + \text{Ca}^{2+} + \text{H}\left[\frac{a_o}{p}\right] \rightarrow \text{As} + \text{Ca}^{3+} + \text{H}\left[\frac{a_o}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV} \tag{75}$$

$$\text{As} + \text{Ca}^{3+} \rightarrow \text{As}^+ + \text{Ca}^{2+} + 41.1\text{ eV} \tag{76}$$

And, the overall reaction is

$$\text{H}\left[\frac{a_H}{p}\right] \rightarrow \text{H}\left[\frac{a_H}{(p+1)}\right] + [(p+1)^2 - p^2] \times 13.6\text{ eV} \tag{77}$$

Multiple Electron Transfer

An energy hole is provided by the transfer of multiple electrons between participating species including atoms, ions, molecules, and ionic and molecular compounds. In one embodiment, the energy hole comprises the transfer of t electrons from one or more species to one or more species whereby the sum of the ionization energies and/or electron affinities of the electron donating species minus the sum of the ionization energies and/or electron affinities of the electron acceptor species equals approximately

$$\frac{m}{2}27.21\text{ eV}$$

where m and t are integers.

CATALYTIC ENERGY HOLE STRUCTURES FOR MOLECULES

Single Electron Excited State

An energy hole is provided by the transition of an electron of a species to an excited state species including atoms, ions, molecules, and ionic and molecular compounds. In one embodiment, the energy hole comprises the excited state transition of an electron of one species whereby the transition energy of the accepting species is mp^{2}X48.6 eV where m and p are integers.

Single Electron Transfer

An energy hole is provided by the transfer of an electron between participating species including atoms, ions, molecules, and ionic and molecular compounds. In one embodiment, the energy hole comprises the transfer of an electron from one species to another species whereby the sum of the ionization energy of the electron donating species minus the ionization energy or electron affinity of the electron accepting species equals approximately mp^{2}X48.6 eV where m and p are integers.

Single Electron Transfer (Two Species)

An efficient catalytic system that hinges on the coupling of three resonator cavities involves iron and lithium. For example, the fourth ionization energy of iron is 54.8 eV. This energy hole is obviously too high for resonant absorption. However, Li^+ releases 5.392 eV when it is reduced to Li. The combination of Fe^{3+} to Fe^{4+} and Li^+ to Li, then, has a net energy change of 49.4 eV.

$$49.4\text{ eV} + \text{Fe}^{3+} + \text{Li}^+ + \text{H}_2\left[2c' = \sqrt{2}\,a_o\right] \rightarrow \text{Fe}^{4+} + \text{Li} + \text{H}_2^*\left[2c' = \frac{\sqrt{2}\,a_o}{2}\right] + 95.7\text{ eV} \tag{78}$$

$$\text{Li} + \text{Fe}^{4+} \rightarrow \text{Li}^+ + \text{Fe}^{3+} + 49.4\text{ eV} \tag{79}$$

And, the overall reaction is

$$\text{H}_2\left[2c' = \sqrt{2}\,a_o\right] \rightarrow \text{H}_2^*\left[2c' = \frac{\sqrt{2}\,a_o}{2}\right] + 95.7\text{ eV} \tag{80}$$

Note that the energy given off as the molecule shrinks is much greater than the energy lost to the energy hole. And, the energy released is large compared to conventional chemical reactions.

An efficient catalytic system that hinges on the coupling of three resonator cavities involves scandium. For example, the fourth ionization energy of scandium is 73.47 eV. This energy hole is obviously too high for resonant absorption. However, Sc^{3+} releases 24.76 eV when it is reduced to Sc^{2+}. The combination of Sc^{3+} to Sc^{4+} and Sc^{3+} to Sc^{2+}, then, has a net energy change of 48.7 eV.

$$48.7\text{ eV} + \text{Sc}^{3+} + \text{Sc}^{3+} + \text{H}_2\left[2c' = \sqrt{2}\,a_o\right] \rightarrow \text{Sc}^{4+} + \text{Sc}^{2+} + \text{H}_2^*\left[2c' = \frac{\sqrt{2}\,a_o}{2}\right] + 95.7\text{ eV} \tag{81}$$

-continued

$$Sc^{2+} + Sc^{4+} \rightarrow Sc^{3+} + Sc^{3+} + 48.7 \text{ eV} \tag{82}$$

And, the overall reaction is

$$H_2[2c' = \sqrt{2}a_o] \rightarrow H_2^*\left[2c' = \frac{\sqrt{2}a_o}{2}\right] + 95.7 \text{ eV} \tag{83}$$

An efficient catalytic system that hinges on the coupling of three resonator cavities involves yttrium. For example, the fourth ionization energy of gallium is 64.00 eV. This energy hole is obviously too high for resonant absorption. However, Pb^{2+} releases 15.03 eV when it is reduced to Pb^{+}. The combination of Ga^{3+} to Ga^{4+} and Pb^{2+} to Pb^{+}, then, has a net energy change of 48.97 eV.

$$48.97 \text{ eV} + Ga^{3+} + Pb^{2+} + H_2[2c' = \sqrt{2}a_o] \rightarrow Ga^{4+} + Pb^{+} + H_2^*\left[2c' = \frac{\sqrt{2}a_o}{2}\right] + 95.7 \text{ eV} \tag{84}$$

$$Ga^{4+} + Pb^{+} \rightarrow Ga^{3+} + Pb^{2+} + 48.97 \text{ eV} \tag{85}$$

And, the overall reaction is

$$H_2[2c' = \sqrt{2}a_o] \rightarrow H_2^*\left[2c' = \frac{\sqrt{2}a_o}{2}\right] + 95.7 \text{ eV} \tag{86}$$

Single Electron Transfer (One Species)

An energy hole is provided by the ionization of an electron from a participating species including an atom, an ion, a molecule, and an ionic or molecular compound to a vacuum energy level. In one embodiment, the energy hole comprises the ionization of an electron from one species to a vacuum energy level whereby the ionization energy of the electron donating species equals approximately mp^2X48.6 eV where m and p are integers.

Multiple Electron Transfer

An energy hole is provided by the transfer of multiple electrons between participating species including atoms, ions, molecules, and ionic and molecular compounds. In one embodiment, the energy hole comprises the transfer of t electrons from one or more species to one or more species whereby the sum of the ionization energies and/or electron affinities of the electron donating species minus the sum of the ionization energies and/or electron affinities of the electron acceptor species equals approximately mp^2X48.6 eV where m, p, and t are integers.

An energy hole is provided by the transfer of multiple electrons between participating species including atoms, ions, molecules, and ionic and molecular compounds. In one embodiment, the energy hole comprises the transfer of t electrons from one species to another whereby the t consecutive electron affinities and/or ionization energies of the electron donating species minus the t consecutive ionization energies and/or electron affinities of the electron acceptor equals approximately mp^2X48.6 eV where m, p, and t are integers.

In a preferred embodiment the electron acceptor species is an oxide such as MnO_x, AlO_x, SiO_x. A preferred molecular electron acceptor is oxygen, O_2.

Two Electron Transfer (One Species)

In an embodiment, a catalytic system that provides an energy hole hinges on the ionization of two electrons from an atom, ion, or molecule to a vacuum energy level such that the sum of two ionization energies is approximately mp^2X48.6 eV where m, and p are integers.

Two Electron Transfer (Two Species)

In another embodiment, a catalytic system that provides an energy hole hinges on the transfer of two electrons from an atom, ion, or molecule to another atom or molecule such that the sum of two ionization energies minus the sum of two electron affinities of the participating atoms, ions, and/or molecules is approximately mp^2X48.6 eV where m and p are integers.

Two Electron Transfer (Two Species)

In another embodiment, a catalytic system that provides an energy hole hinges on the transfer of two electrons from an atom, ion, or molecule to another atom, ion, or molecule such that the sum of two ionization energies minus the sum of one ionization energy and one electron affinity of the participating atoms, ions, and/or molecules is approximately mp^2X48.6 eV where m and p are integers.

Other Energy Holes

In another embodiment, energy holes, each of approximately mX67.8 eV given by Eq. (30)

$$-m \times V_e = -m \times \frac{-2e^2}{8\pi\varepsilon_o\sqrt{a^2-b^2}} \ln\frac{a+\sqrt{a^2-b^2}}{a-\sqrt{a^2-b^2}} = m \times 67.813 \text{ eV} \tag{87}$$

are provided by electron transfer reactions of reactants including electrochemical reactant(s) (electrocatalytic ion(s) or couple(s)) which cause heat to be released from hydrogen molecules as their electrons are stimulated to relax to quantized potential energy levels below that of the "ground state". The energy removed by an electron transfer reaction, energy hole, is resonant with the hydrogen energy released to stimulate this transition. The source of hydrogen molecules is the production on the surface of a cathode during electrolysis of water in the case of an electrolytic energy reactor and hydrogen gas or a hydride in the case of a pressurized gas energy reactor or gas discharge energy reactor.

An energy hole is provided by the transfer of one or more electrons between participating species including atoms, ions, molecules, and ionic and molecular compounds. In one embodiment, the energy hole comprises the transfer of t electrons from one or more species to one or more species whereby the sum of the ionization energies and/or electron affinities of the electron donating species minus the sum of the ionization energies and/or electron affinities of the electron acceptor species equals approximately mX67.8 eV where m and t are integers.

An efficient catalytic system that hinges on the coupling of three resonator cavities involves magnesium and strontium. For example, the third ionization energy of magnesium is 80.143 eV. This energy hole is obviously too high for resonant absorption. However, Sr^{2+} releases 11.03 eV when it is reduced to Sr^{+}. The combination of Mg^{2+} to Mg^{3+} and Sr^{2+} to Sr^{+}, then, has a net energy change of 69.1 eV.

$$69.1 \text{ eV} + Mg^{2+} + Sr^{2+} + H_2[2c' = \sqrt{2}a_o] \rightarrow Mg^{3+} + Sr^{+} + H_2^*\left[2c' = \frac{\sqrt{2}a_o}{2}\right] + 95.7 \text{ eV} \tag{88}$$

-continued

$$Mg^{3+} + Sr^{+} \rightarrow Mg^{2+} + Sr^{2+} + 69.1 \text{ eV} \tag{89}$$

And, the overall reaction is

$$H_2\left[2c' = \sqrt{2}\, a_o\right] \rightarrow H_2^*\left[2c' = \frac{\sqrt{2}\, a_o}{2}\right] + 95.7 \text{ eV} \tag{90}$$

Another efficient catalytic system that hinges on the coupling of three resonator cavities involves magnesium and calcium. In this case, Ca^{2+} releases 11.871 eV when it is reduced to Ca^{+}. The combination of Mg^{2+} to Mg^{3+} and Ca^{2+} to Ca^{+}, then, has a net energy change of 68.2 eV.

$$68.2 \text{ eV} + Mg^{2+} + Ca^{2+} + H_2\left[2c' = \sqrt{2}\, a_o\right] \rightarrow Mg^{3+} + Ca^{+} + H_2^*\left[2c' = \frac{\sqrt{2}\, a_o}{2}\right] + 95.7 \text{ eV} \tag{91}$$

$$Mg^{3+} + Ca^{+} \rightarrow Mg^{2+} + Ca^{2+} + 68.2 \text{ eV} \tag{92}$$

And, the overall reaction is

$$H_2\left[2c' = \sqrt{2}\, a_o\right] \rightarrow H_2^*\left[2c' = \frac{\sqrt{2}\, a_o}{2}\right] + 95.7 \text{ eV} \tag{93}$$

In four other embodiments wherein the theory is given in my previous U.S. patent application Ser. No. 08/107,357 filed on Aug. 16, 1993 which is incorporated herein by this reference, energy holes, each of approximately:

nXE_T eV with zero order vibration where E_T is given by Eq. (38);

mX31.94 eV where 31.94 eV is given by Eq. (222) of the U.S. patent application Ser. No. 08/107,357 where n and m are integers,

$$\begin{aligned} E_D &= E(2H[a_o]) - E_{T_{zero\ order}} - \frac{E_{vib}}{2} \\ &= -27.21 + 31.94 = 4.73 \text{ eV} \end{aligned} \tag{222}$$

and

95.7 eV (corresponding to m=1 in Eq. (43) with zero order vibration which is given by the difference in

$$-E_{T_{zero\ order}} - \frac{E_{vib}}{2}$$

of Eqs. (254) and (222) of the U.S. patent application Ser. No. 08/107,357))

$$\begin{aligned} E_D &= E\left(2H\left[\frac{a_o}{2}\right]\right) - E_{T_{zero\ order}} - \frac{E_{vib}}{2} \\ &= -108.8 + 127.66 = 18.86 \text{ eV} \end{aligned} \tag{254}$$

are provided by electron transfer reactions of reactants including electrochemical reactant(s) (electrocatalytic ion(s) or couple(s)) which cause heat to be released from hydrogen molecules as their electrons are stimulated to relax to quantized potential energy levels below that of the "ground state". The energy removed by an electron transfer reaction, energy hole, is resonant with the hydrogen energy released to stimulate this transition. The source of hydrogen molecules is the production on the surface of a cathode during electrolysis of water in the case of an electrolytic energy reactor and hydrogen gas or a hydride in the case of a pressurized gas energy reactor or gas discharge energy reactor.

An energy hole is provided by the transfer of one or more electrons between participating species including atoms, ions, molecules, and ionic and molecular compounds. In one embodiment, the energy hole comprises the transfer of t electrons from one or more species to one or more species whereby the sum of the ionization energies and/or electron affinities of the electron donating species minus the sum of the ionization energies and/or electron affinities of the electron acceptor species equals approximately mX31.94 eV (Eq. (222)) where m and t are integers.

An energy hole is provided by the transfer of one or more electrons between participating species including atoms, ions, molecules, and ionic and molecular compounds. In one embodiment, the energy hole comprises the transfer of t electrons from one or more species to one or more species whereby the sum of the ionization energies and/or electron affinities of the electron donating species minus the sum of the ionization energies and/or electron affinities of the electron acceptor species equals approximately mX95.7 eV where m and t are integers.

ENERGY REACTOR

An energy reactor **50**, in accordance with the invention, is shown in FIG. **5** and comprises a vessel **52** which contains an energy reaction mixture **54**, a heat exchanger **60**, and a steam generator **62**. The heat exchanger **60** absorbs heat released by the shrinkage reaction, when the reaction mixture, comprised of shrinkable material, shrinks. The heat exchanger exchanges heat with the steam generator **62** which absorbs heat from the exchanger **60** and produces steam. The energy reactor **50** further comprises a turbine **70** which receives steam from the steam generator **62** and supplies mechanical power to a power generator **80** which converts the steam energy into electrical energy, which can be received by a load **90** to produce work or for dissipation.

The energy reaction mixture **54** comprises an energy releasing material **56** including a source of hydrogen isotope atoms or a source of molecular hydrogen isotope, and a source of energy holes **58** which resonantly remove approximately mX27.21 eV to cause atomic hydrogen "shrinkage" and approximately mX48.6 eV to cause molecular hydrogen "shrinkage" where m is an integer wherein the shrinkage reaction occurs by contact of the hydrogen with the source of energy holes. The shrinkage reaction releases heat and shrunken atoms and/or molecules.

The source of hydrogen can be hydrogen gas, dissociation of water including thermal dissociation, electrolysis of water, hydrogen from hydrides, or hydrogen from metal-hydrogen solutions. In all embodiments, the source of energy holes can be one or more of an electrochemical, chemical, photochemical, thermal, free radical, sonic, or nuclear reaction(s) or inelastic photon or particle scattering reaction(s). In the latter two cases, the present invention of an energy reactor comprises a particle source **75***b* and/or photon source **75***a* to supply the said energy holes. In these cases, the energy hole corresponds to stimulated emission by the photon or particle. In preferred embodiments of the pressurized gas energy and gas discharge reactors shown in FIGS. **7** and **8**, respectively, a photon source **75***a* dissociates hydrogen molecules to hydrogen atoms. The photon source producing photons of at least one energy of approximately mX27.21 eV,

$$\frac{m}{2}\times 27.21\ \text{eV},$$

or 40.8 eV causes stimulated emission of energy as the hydrogen atoms undergo the shrinkage reaction. In another preferred embodiment, a photon source **75***a* producing photons of at least one energy of approximately mX48.6 eV, 95.7 eV, or mX31.94 eV causes stimulated emission of energy as the hydrogen molecules undergo the shrinkage reaction. In all reaction mixtures, a selected external energy device **75**, such as an electrode may be used to supply an electrostatic potential or a current (magnetic field) to decrease the activation energy of the resonant absorption of an energy hole. In another embodiment, the mixture **54**, further comprises a surface or material to dissociate and/or absorb atoms and/or molecules of the energy releasing material **56**. Such surfaces or materials to dissociate and/or absorb hydrogen, deuterium, or tritium comprise an element, compound, alloy, or mixture of transition elements and inner transition elements, iron, platinum, palladium, zirconium, vanadium, nickel, titanium, Sc, Cr, Mn, Co, Cu, Zn, Y, Nb, Mo, Tc, Ru, Rh, Ag, Cd, La, Hf, Ta, W, Re, Os, Ir, Au, Hg, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Vb, Lu, Th, Pa, U, activated charcoal (carbon), and intercalated Cs carbon (graphite). In a preferred embodiment, a source of energy holes to shrink hydrogen atoms comprises a catalytic energy hole material **58**, typically comprising electrocatalytic ions and couples that provide an energy hole of approximately mX27.21 eV plus or minus 1 eV. In a preferred embodiment, a source of energy holes to shrink hydrogen molecules comprises a catalytic energy hole material **58**, typically comprising electrocatalytic ions and couple (s) including those that provide an energy hole of approximately nX48.6 eV plus or minus 5 eV. The electrocatalytic ions and couple(s) include the electrocatalytic ions and couples described in my previous U.S. Patent Applications entitled "Energy/Matter Conversion Methods and Structures", Ser. No. 08/467,051 filed on Jun. 6, 1995 which is a continuation-in-part application of Ser. No. 08/416,040 filed on Apr. 3, 1995 which is a continuation-in-part application of Ser. No. 08/107,357 filed on Aug. 16, 1993, which is a continuation-in-part application of Ser. No. 08/075,102 (Dkt. 99437) filed on Jun. 11, 1993, which is a continuation-in-part application of Ser. No. 07/1626,496 filed on Dec. 12, 1990 which is a continuation-in-part application of Ser. No. 07/345,628 filed Apr. 28, 1989 which is a continuation-in-part application of Ser. No. 07/341,733 filed Apr. 21, 1989, which are incorporated herein by reference.

A further embodiment is the vessel **52** containing a source of energy holes including an electrocatalytic ion or couple(s) (source of energy holes) in the molten, liquid, gaseous, or solid state and a source of hydrogen including hydrides and gaseous hydrogen. In the case of a reactor which shrinks hydrogen atoms, the embodiment further comprises a means to dissociate the molecular hydrogen into atomic hydrogen including an element, compound, alloy, or mixture of transition elements, inner transition elements, iron, platinum, palladium, zirconium, vanadium, nickel, titanium, Sc, Cr, Mn, Co, Cu, Zn, Y, Nb, Mo, Tc, Ru, Rh, Ag, Cd, La, Hf, Ta, W, Re, Os, Ir, Au, Hg, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Vb, Lu, Th, Pa, U, activated charcoal (carbon), and intercalated Cs carbon (graphite) or electromagnetic radiation including UV light provided by photon source **75**.

The present invention of an electrolytic cell energy reactor, pressurized gas energy reactor, and a gas discharge energy reactor, comprises: a source of hydrogen; one of a solid, molten, liquid, and gaseous source of energy holes; a vessel containing hydrogen and the source of energy holes wherein the shrinkage reaction occurs by contact of the hydrogen with the source of energy holes; and a means for removing the (molecular) lower-energy hydrogen so as to prevent an exothermic shrinkage reaction from coming to equilibrium. The present energy invention is further described in my previous U.S. Patent Applications entitled "Energy/Matter Conversion Methods and Structures", Ser. No. 08/467,051 filed on Jun. 6, 1995 which is a continuation-in-part application of Ser. No. 08/416,040 filed on Apr. 3, 1995 which is a continuation-in-part application of Ser. No. 08/107,357 filed on Aug. 16, 1993, which is a continuation-in-part application of Ser. No. 08/075,102 (Dkt. 99437) filed on Jun. 11, 1993, which is a continuation-in-part application of Ser. No. 07/626,496 filed on Dec. 12, 1990 which is a continuation-in-part application of Ser. No. 07/345,628 filed Apr. 28, 1989 which is a continuation-in-part application of Ser. No. 07/341,733 filed Apr. 21, 1989, and my publications, Mills, R., Kneizys, S., Fusion Technology, 210, (1991), pp. 65–81; Mills, R., Good, W., Shaubach, R., "Dihydrino Molecule Identification", Fusion Technology, 25, 103 (1994); Mills, R., Good, W., "Fractional Quantum Energy Levels of Hydrogen", Fusion Technology, Vol. 28, No. 4, November, (1995), pp. 1697–1719 which are incorporated herein by reference.

Electrolytic Energy Reactor

An electrolytic energy reactor is described in my previous U.S. patent applications entitled "Energy/Matter Conversion Methods and Structures", Ser. No. 08/467,051 filed on Jun. 6, 1995 which is a continuation-in-part application of Ser. No. 08/416,040 filed on Apr. 3, 1995 which is a continuation-in-part application of Ser. No. 08/107,357 filed on Aug. 16, 1993, which is a continuation-in-part application of Ser. No. 08/075,102 (Dkt. 99437) filed on Jun. 11, 1993, which is a continuation-in-part application of Ser. No. 07/626,496 filed on Dec. 12, 1990 which is a continuation-in-part application of Ser. No. 07/345,628 filed Apr. 28, 1989 which is a continuation-in-part application of Ser. No. 07/341,733 filed Apr. 21, 1989 which are incorporated herein by reference. A preferred embodiment of the energy reactor of the present invention comprises an electrolytic cell forming the reaction vessel **52** of FIG. **5** including a molten electrolytic cell. The electrolytic cell **100** is shown generally in FIG. **6**. An electric current is passed through the electrolytic solution **102** having a electrocatalytic ions or couples providing energy holes equal to the resonance shrinkage energy (including the electrocatalytic ions and couples described in my previous U.S. Patent Applications incorporated herein by reference) by the application of a voltage to an anode **104** and cathode **106** by the power controller **108** powered by the power supply **110**. Ultrasonic or mechanical energy may also be imparted to the cathode **106** and electrolytic solution **102** by vibrating means **112**. Heat can be supplied to the electrolytic solution **102** by heater **114**. The pressure of the electrolytic cell **100** can be controlled by pressure regulator means **116** where the cell can be closed. The reactor further comprises a means **101** that removes the (molecular) lower-energy hydrogen such as a selective venting valve to prevent the exothermic shrinkage reaction from coming to equilibrium.

In a preferred embodiment, the electrolytic cell is operated at zero voltage gap by applying an overpressure of hydrogen with hydrogen source **121** where the overpressure can be controlled by pressure control means **122** and **116**. Water can be reduced to hydrogen and hydroxide at the cathode **106**, and the hydrogen can be oxidized to protons at the anode **104**. An embodiment of the electrolytic cell energy reactor, comprises a reverse fuel cell geometry which removes the lower-energy hydrogen under vacuum. A preferred cathode **106** of this embodiment has a modified gas diffusion layer and comprises a gas route means including a first Teflon membrane filter and a second carbon paper/Teflon membrane filter composite layer. A further embodiment comprises a reaction vessel that can be closed except for a connection to a condensor **140** on the top of the vessel **100**. The cell can be operated at a boil such that the steam evolving from the boiling electrolyte **102** can be condensed in the condenser **140**, and the condensed water can be returned to the vessel **100**. The lower-energy state hydrogen can be vented through the top of the condenser **140**. In one embodiment, the condensor contains a hydrogen/oxygen recombiner **145** that contacts the evolving electrolytic gases. The hydrogen and oxygen are recombined, and the resulting water can be returned to the vessel **100**. The heat released from the exothermic reaction whereby the electrons of the electrolytically produced hydrogen atoms (molecules) are induced to undergo transitions to energy levels below the "ground state" and the heat released due to the recombination of the electrolytically generated normal hydrogen and oxygen can be removed by a heat exchanger **60** of FIG. **5** which can be connected to the condensor **140**.

In vacuum, in the absence of external fields, the energy hole to stimulate a hydrogen atom (molecule) to undergo a shrinkage transition is mX27.21 eV (mX48.6 eV) where m is an integer. This resonance shrinkage energy can be altered when the atom (molecule) is in a media different from vacuum. An example is a hydrogen atom (molecule) absorbed to the cathode **106** present in the aqueous electrolytic solution **102** having an applied electric field and an intrinsic or applied magnetic field provided by external magnetic field generator **75**. Under these conditions the energy hole required can be slightly different from mX27.21 eV (mX48.6 eV). Thus, a source of energy holes including electrocatalytic ion and couple reactants can be selected which has a redox (electron transfer) energy resonant with the resonance shrinkage energy when operating under these conditions. In the case where a nickel cathode **106** is used to electrolyze an aqueous solution **102** where the cell is operating within a voltage range of 1.4 to 5 volts, the K^+/K^+ and Rb^+ (Fe^{3+}/Li^+ and Sc^{3+}/Sc^{3+}) electrocatalytic ions and couples are preferred embodiments to shrink hydrogen atoms (molecules).

The cathode provides hydrogen atoms (molecules), and the shrinkage reaction occurs at the surface of the cathode where hydrogen atoms (molecules) and the source of energy holes (electrocatalytic ion or couple) are in contact. Thus, the shrinkage reaction can be dependent on the surface area of the cathode. For a constant current density, giving a constant concentration of hydrogen atoms (molecules) per unit area, an increase in surface area increases the reactants available to undergo the shrinkage reaction. Also. an increase in cathode surface area decreases the resistance of the electrolytic cell which improves the electrolysis efficiency. A preferred cathode of the electrolytic cell including a nickel cathode has the properties of a high surface area, a highly stressed and hardened surface such as a cold drawn or cold worked surface, and a large number of grain boundaries.

In a preferred embodiment of the electrolytic cell energy reactor, the source of energy holes can be incorporated into the cathode, mechanically by methods including cold working the source of energy holes into the surface of the cathode; thermally by methods including melting the source of energy holes into the surface of the cathode and evaporation of a solvent of a solution of the source of energy holes in contact with the surface of the cathode, and electrostatically by methods including electrolytic deposition, ion bombardment, and vacuum deposition.

The shrinkage reaction rate can be dependent upon the composition of the cathode **106**. Hydrogen atoms (molecules) are reactants to produce energy via the shrinkage reaction. Thus, the cathode must efficiently provide a high concentration of hydrogen atoms (molecules). The cathode **106** can be comprised of any element. compound, alloy, or mixture of a conductor or semiconductor including transition elements and compounds, actinide and lanthanide elements and compounds, and group IIIB and IVB elements and compounds. Transition metals dissociate hydrogen gas into atoms to a more or lesser extent depending on the metal. Nickel and titanium readily dissociate hydrogen molecules and are preferred embodiments for shrinking hydrogen atoms. The cathode can alter the energy of the absorbed hydrogen atoms (molecules) and affect the energy of the shrinkage reaction. A cathode material can be selected which provides resonance between the energy hole and the resonance shrinkage energy. In the case of the K^+/K^+ electrocatalytic couple with carbonate as the counterion for catalyzing the shrinkage of hydrogen atoms, the relationship of the cathode material to the reaction rate can be:

$$Pt<Pd<<Ti,Fe<Ni$$

This can be the opposite order of the energy released when these materials absorb hydrogen atoms. Thus, for this electrocatalytic couple, the reaction rate can be increased by using a cathode which weakly absorbs the hydrogen atoms with little perturbation of their electronic energies.

Also, coupling of resonator cavities and enhancement of the transfer of energy between them can be increased when the media is a nonlinear media such as a magnetized ferromagnetic media. Thus, a paramagnetic or ferromagnetic cathode, a nonlinear magnetized media, increases the reaction rate by increasing the coupling of the resonance shrinkage energy of the hydrogen atom and energy hole comprising an electrocatalytic ion or couple. Alternatively, a magnetic field can be applied with the magnetic field generator **75**. Magnetic fields at the cathode alter the energy of absorbed hydrogen and concomitantly alter the resonance shrinkage energy. Magnetic fields also perturb the energy of the electrocatalytic reactions (energy hole) by altering the energy levels of the electrons involved in the reactions. The magnetic properties of the cathode are selected as well as the strength of the magnetic field which is applied by magnetic field generator **75** to optimize shrinkage reaction rate-the power output. A preferred ferromagnetic cathode is nickel.

A preferred method to clean the cathode of the electrolytic cell including a nickel cathode is to anodize the cathode in a basic electrolytic solution including approximately 0.57 M X_2CO_3 (X is the alkali cation of the electrolyte including K^+) and to immerse the cathode in a dilute solution of H_2O_2 such as approximately 3% H_2O_2. In a further embodiment of the cleaning method, cyclic voltametry with a second electrode of the same material as the first cathode is performed. The cathode can be then thoroughly rinsed with distilled water. Organic material on the surface of the cathode inhibits the catalytic reaction whereby the electrons of the electrolytically produced hydrogen atoms (molecules) are induced to undergo transitions to energy levels below the "ground state". Cleaning by this method removes the organic material from the cathode surface and adds oxygen atoms onto the cathode surface. Doping the metal surface, including a nickel surface, with oxygen atoms by anodizing the cathode and cleaning the cathode in H_2O_2 increases the power output by decreasing hydrogen recombination to molecular hydrogen and by decreasing the bond energy between the metal and the hydrogen atoms (molecules) which conforms the resonance shrinkage energy of the absorbed hydrogen to the energy hole provided by the source of energy holes including the K^+/K^+ (Sc^{3+}/Sc^{3+}) electrocatalytic couples.

Different anode materials have different overpotentials for the oxidation of water, which can affect ohmic losses. An anode of low overpotential will increase the efficiency. Nickel, platinum, and dimensionally stable anodes including platinized titanium are preferred anodes. In the case of the K^+/K^+ electrocatalytic couple where carbonate is used as the counterion, nickel is a preferred anode. Nickel is also a preferred anode for use in basic solutions with a nickel cathode. Nickel is inexpensive relative to platinum, and fresh nickel can be electroplated onto the cathode during electrolysis.

A preferred method to clean a dimensionally stable anode including a platinized titanium anode is to place the anode in approximately 3 M HCl for approximately 5 minutes and then to rinse it with distilled water.

In the case of hydrogen atom shrinkage, hydrogen atoms at the surface of the cathode **106** form hydrogen gas which can form bubbles on the surface of the cathode. These bubbles act as an boundary layer between the hydrogen atoms and the electrocatalytic ion or couple. The boundary can be ameliorated by vibrating the cathode and/or the electrolytic solution **102** or by applying ultrasound with vibrating means **112**; and by adding wetting agents to the electrolytic solution **102** to reduce the surface tension of the water and prevent bubble formation. The use of a cathode having a smooth surface or a wire cathode prevents gas adherence. And an intermittent current, provided by an on-off circuit of power controller **108** provides periodic replenishing of hydrogen atoms which are dissipated by hydrogen gas formation followed by diffusion into the solution while preventing excessive hydrogen gas formation which could form a boundary layer.

The shrinkage reaction can be temperature dependent. Most chemical reactions double their rates for each 10° C. rise in temperature. Increasing the temperature increases the collision rate between the hydrogen atoms (molecules) and the electrocatalytic ion or couple which will increase the shrinkage reaction rate. With large temperature excursions from room temperature, the kinetic energy distribution of the reactants can be sufficiently altered to cause the energy hole and the resonance shrinkage energy to conform to a more or lesser extent. The rate can be proportional to the extent of the conformation or resonance of these energies. The temperature can be adjusted to optimize the shrinkage reaction rate-energy production rate. In the case of the K^+/K^+ electrocatalytic couple, a preferred embodiment can be to run the reaction at a temperature above room temperature by applying heat with heater **114**.

The shrinkage reaction can be dependent on the current density. An increase in current density can be equivalent, in some aspects, to an increase in temperature. The collision rate increases, and the energy of the reactants increases with current density. Thus, the rate can be increased by increasing the collision rate of the reactants; however, the rate may be increased or decreased depending on the effect of the increased reactant energies on the conformation of the energy hole and the resonance shrinkage energy. Also, increased current dissipates more energy by ohmic heating and may cause hydrogen bubble formation, in the case of the shrinkage of hydrogen atoms. But, a high flow of gas may dislodge bubbles which diminishes any hydrogen gas boundary layer. The current density can be adjusted with power controller **108** to optimize the excess energy production. In a preferred embodiment, the current density can be in the range 1 to 1000 milliamps per square centimeter.

The pH of the aqueous electrolytic solution **102** can affect the shrinkage reaction rate. In the case that the electrocatalytic ion or couple is positively charged, an increase in the pH will reduce the concentration of hydronium at the negative cathode; thus, the concentration of the electrocatalytic ion or couple cations will increase. An increase in reactant concentration increases the reaction rate. In the case of the Rb^+ or K^+/K^+ (Sc^{3+}/Sc^{3+}) ion or couple, a preferred pH can be basic (7.1–14).

The counterion of the electrocatalytic ion or couple of the electrolytic solution **102** can affect the shrinkage reaction rate by altering the energy of the transition state. For example, the transition state complex of the K^+/K^+ electrocatalytic couple with the hydrogen atom has a plus two charge and involves a three body collision which can be unfavorable. A negative two charged oxyanion can bind the two potassium ions; thus, it provides a neutral transition state complex of lower energy whose formation depends on a binary collision which can be greatly favored. The rate can be dependent on the separation distance of the potassium ions as part of the complex with the oxyanion. The greater the separation distance, the less favorable can be the transfer of an electron between them. A close juxtaposition of the potassium ions will increase the rate. The relationship of the reaction rate to the counterion in the case where the K^+/K^+ couple is used can be:

$$OH^{-}<PO_4^{3-},\ HPO_3^{2-}<SO_4^{2-}<<CO_3^{2-}$$

Thus, a planar negative two charge oxyanion including carbonate with at least two binding sites for K^+ which provides close juxtaposition of the K^+ ions can be preferred as the counterion of the K^+/K^+ electrocatalytic couple. The carbonate counterion can be also a preferred counterion for the Rb^+ electrocatalytic ion.

A power controller **108** comprising an intermittent current, on-off, electrolysis circuit will increase the excess heat by providing optimization of the electric field as a function of time which provides maximum conformation of reactant energies, provides an optimal concentration of hydrogen atoms (molecules) while minimizing ohmic and electrolysis power losses and, in the case of the shrinkage of hydrogen atoms, minimizes the formation of a hydrogen gas boundary layer. The frequency, duty cycle, peak voltage, step waveform, peak current, and offset voltage are adjusted to achieve the optimal shrinkage reaction rate and shrinkage reaction power while minimizing ohmic and electrolysis power losses. In the case where the K^+/K^+ electrocatalytic couple can be used with carbonate as the counterion; nickel as the cathode: and platinum as the anode, a preferred embodiment can be to use an intermittent square-wave having an offset voltage of approximately 1.4 volts to 2.2 volts; a peak voltage of approximately 1.5 volts to 3.75 volts; a peak current of approximately 1 mA to 100 mA per square centimeter of cathode surface area; approximately a 5%–90% duty cycle; and a frequency in the range of 1 Hz to 1500 Hz.

Further energy can be released by repeating the shrinkage reaction. The atoms (molecules) which have undergone shrinkage diffuse into the cathode lattice. A cathode **106** can be used which will facilitates multiple shrinkage reactions of hydrogen atoms (molecules). One embodiment is to use a cathode which can be fissured and porous to the electrocatalytic ion or couple such that it can contact shrunken atoms (molecules) which have diffused into a lattice, including a metal lattice. A further embodiment is to use a cathode of alternating layers of a material which provides hydrogen atoms (molecules) during electrolysis including a transition metal and an electrocatalytic ion or couple such that shrunken hydrogen atoms (molecules) periodically or repetitively diffuse into contact with the electrocatalytic ion or couple.

The shrinkage reaction can be dependent on the dielectric constant of the media. The dielectric constant of the media alters the electric field at the cathode and concomitantly alters the energy of the reactants. Solvents of different dielectric constants have different solvation energies, and the dielectric constant of the solvent can also lower the overpotential for electrolysis and improve electrolysis efficiency. A solvent, including water, can be selected for the electrolytic solution **102** which optimizes the conformation of the energy hole and resonance shrinkage energy and maximizes the efficiency of electrolysis.

The solubility of hydrogen in the reaction solution can be directly proportional to the pressure of hydrogen above the solution. Increasing the pressure increases the concentration of reactant hydrogen atoms (molecules) at the cathode **106** and thereby increases the rate. But, in the case of the shrinkage of hydrogen atoms this also favors the development of a hydrogen gas boundary layer. The hydrogen pressure can be controlled by pressure regulator means **116** to optimize the shrinkage reaction rate.

In a preferred embodiment, the cathode **106** of the electrolytic cell comprises the catalytic material including a hydrogen spillover catalyst described in the Pressurized Gas Energy Reactor Section below. In another embodiment, the cathode comprises multiple hollow vessels comprising a thin film conductive shell whereby lower-energy hydrogen diffuses through the thin film and collects inside each vessel and undergoes disproportionation reactions therein.

The heat output can be monitored with thermocouples present in at least the vessel **100** and the condensor **140** of FIG. **6** and the heat exchanger **60** of FIG. **5**. The output power can be controlled by a computerized monitoring and control system which monitors the thermistors and controls the means to alter the power output.

Pressurized Gas Energy Reactor

A pressurized gas energy reactor comprises the first vessel **200** of FIG. **7** containing a source of hydrogen including hydrogen from metal-hydrogen solutions, hydrogen from hydrides, hydrogen from the dissociation of water including thermal dissociation, hydrogen from the electrolysis of water, or hydrogen gas. In the case of a reactor which shrinks hydrogen atoms, the reactor further comprises a means to dissociate the molecular hydrogen into atomic hydrogen such as a dissociating material including an element, compound, alloy, or mixture of transition elements and inner transition elements, iron, platinum, palladium, zirconium, vanadium, nickel, titanium, Sc, Cr, Mn, Co, Cu, Zn, Y, Nb, Mo, Tc, Ru, Rh, Ag, Cd, La, Hf, Ta, W, Re, Os, Ir, Au, Hg, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Vb, Lu, Th, Pa, U, activated charcoal (carbon), and intercalated Cs carbon (graphite) or electromagnetic radiation including UV light provided by photon source **205** such that the dissociated hydrogen atoms (molecules) contact a source of energy holes including a molten, liquid, gaseous, or solid source of the energy holes including the electrocatalytic ions and couples described in my previous U.S. Patent Applications entitled "Energy/Matter Conversion Methods and Structures", Ser. No. 08/467,051 filed on Jun. 6, 1995 which is a continuation-in-part application of Ser. No. 08/416,040 filed on Apr. 3, 1995 which is a continuation-in-part application of Ser. No. 08/107,357 filed on Aug. 16, 1993, which is a continuation-in-part application of Ser. No. 08/075,102 (Dkt. 99437) filed on Jun. 11, 1993, which is a continuation-in-part application of Ser. No. 07/626,496 filed on Dec. 12, 1990 which is a continuation-in-part application of Ser. No. 07/345,628 filed Apr. 28, 1989 which is a continuation-in-part application of Ser. No. 07/341,733 filed Apr. 21, 1989, which are incorporated herein by reference. The pressurized gas energy reactor further comprises a means **201** to remove the (molecular) lower-energy hydrogen such as a selective venting valve to prevent the exothermic shrinkage reaction from coming to equilibrium. One embodiment comprises heat pipes as heat exchanger **60** of FIG. **5** which have a lower-energy hydrogen venting valve at a cold spot.

A preferred embodiment of the pressurized gas energy reactor of the present invention comprises a first reaction vessel **200** with inner surface **240** comprised of a material to dissociate the molecular hydrogen into atomic hydrogen including an element, compound, alloy, or mixture of transition elements and inner transition elements, iron, platinum, palladium, zirconium, vanadium, nickel, titanium, Sc, Cr, Mn, Co, Cu, Zn, Y, Nb, Mo, Tc, Ru, Rh, Ag, Cd, La, Hf, Ta, W, Re, Os, Ir, Au, Hg, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Th, Dy, Ho, Er, Tm, Vb, Lu, Th, Pa, U, activated charcoal (carbon), and intercalated Cs carbon (graphite). In a further embodiment, the inner surface **240** can be comprised of a proton conductor. The first reaction vessel **200** can be sealed in a second reaction vessel **220** and receives hydrogen from source **221** under pressure which can be controlled by pressure measurement and control means **222** and **223**. In a preferred embodiment the hydrogen pressure can be in the range of 10^{-3} atmospheres to **100** atmospheres. The wall **250** of the first vessel **200** can be permeable to hydrogen. The outer surface **245** and/or outer vessel **220** has a source of energy holes equal to the resonance shrinkage energy. In one embodiment the source of energy holes can be a mixture or solution containing energy holes in the molten, liquid, or solid state. In another embodiment an electric current can be passed through the material having a source of energy holes. The reactor further comprises a means to control the reaction rate such as current source **225** and heating means **230** which heat the first reaction vessel **200** and the second reaction vessel **220**. In a preferred embodiment the outer reaction vessel **220** contains oxygen, the inner surface **240** comprises one or more of a coat of nickel, platinum, or palladium. The outer surface **245** can be coated with one or more of copper, tellurium, arsenic, cesium, platinum, or palladium and an oxide such as CuO_x, PtO_x, PdO_x, MnO_x, AlO_x, SiO_x. The electrocatalytic ion or couple can be regenerated spontaneously or via a regeneration means including heating means **230** and current source **225**.

In another embodiment, the pressurized gas energy reactor comprises only a single reaction vessel **200** with a hydrogen impermeable wall **250**. In the case of a reactor which shrinks hydrogen atoms, one or more of a hydrogen dissociating materials including transition elements and inner transition elements, iron, platinum, palladium, zirconium, vanadium, nickel, titanium, Sc, Cr, Mn, Co, Cu, Zn, Y, Nb, Mo, Tc, Ru, Rh, Ag, Cd, La, Hf, Ta, W, Re, Os, Ir, Au, Hg, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Vb, Lu, Th, Pa, U, activated charcoal (carbon), and intercalated Cs carbon (graphite) are coated on the inner surface **240** with a source of energy holes including one or more of copper, tellurium, arsenic, cesium, platinum, or palladium and an oxide such as CuO_x, PtO_x, PdO_x, MnO_x, AlO_x, SiO_x. In another embodiment, the source of energy hole can be one of a inelastic photon or particle scattering reaction(s). In a preferred embodiment the photon source **205** supplies the energy holes where the energy hole corresponds to stimulated emission by the photon. In the case of a reactor which shrinks hydrogen atoms the photon source **205** dissociates hydrogen molecules into hydrogen atoms. The photon source producing photons of at least one energy of approximately mX27.21 eV,

$$\frac{m}{2} \times 27.21 \text{ eV},$$

or 40.8 eV causes stimulated emission of energy as the hydrogen atoms undergo the shrinkage reaction. In another preferred embodiment, a photon source **205** producing photons of at least one energy of approximately mX48.6 eV, 95.7 eV, or mX31.94 eV causes stimulated emission of energy as the hydrogen molecules undergo the shrinkage reaction.

A preferred inner surface, **240**, and outer surface, **245**, of the pressurized gas energy reactor including a nickel surface has the properties of a high surface area, a highly stressed and hardened surface such as a cold drawn or cold worked surface, and a large number of grain boundaries.

In an embodiment of the pressurized gas energy reactor, the source of energy holes can be incorporated into the inner surface, **240**, and outer surface, **245**, mechanically by methods including cold working the source of energy holes into the surface material and thermally by methods including melting the source of energy holes into the surface material (fusion). Further methods of incorporation include dry impregnation, evaporation of a solution of the source of energy holes in contact with the surface material (precipitation), ion bombardment, vacuum deposition, impregnation, leaching, and electrostatic incorporation including electrolytic deposition and electroplating. A preferred method to clean the inner surface **240** and the outer surface **245** including a nickel surface is to fill the inner vessel and the outer vessel with a basic electrolytic solution including approximately 0.57 M X_2CO_3 (X is the alkali cation of the electrolyte including K^+) and to fill the inner vessel and the outer vessel with a dilute solution of H_2O_2. Each of the inner vessel and the outer vessel can be then thoroughly rinsed with distilled water. In one embodiment, at least one of the vessel **200** or the vessel **220** can be then filled with a solution of the energy hole including an approximately 0.57 M K_2CO_3 solution.

In a further embodiment, textural and/or structural promoters are incorporated with the source of energy holes to increase the shrinkage reaction rate.

In one embodiment of the method of operation of the pressurized gas energy reactor, hydrogen can be introduced inside of the first vessel from source **221** under pressure which can be controlled by pressure control means **222**. In the case of a reactor which shrinks hydrogen atoms, the molecular hydrogen can be dissociated into atomic hydrogen by a dissociating material or electromagnetic radiation including UV light provided by photon source **205** such that the dissociated hydrogen atoms contact a source of energy holes including a molten, liquid, gaseous, or solid source of the energy holes. The atomic (molecular) hydrogen releases energy as its electrons are stimulated to undergo transitions to lower energy levels by the energy holes. Alternatively, the hydrogen dissociates on the inner surface **240**, diffuses though the wall **250** of the first vessel **200** and contacts a source of energy holes on the outer surface **245** or contact a source of energy holes including a molten, liquid, gaseous, or solid source of the energy holes as hydrogen atoms or recombined hydrogen molecules. The atomic (molecular) hydrogen releases energy as its electrons are stimulated to undergo transitions to lower energy levels by the energy holes. The electrocatalytic ion or couple can be regenerated spontaneously or via a regeneration means including heating means **230** and current source **225**. The (molecular) lower-energy hydrogen can be removed from vessel **200** and/or vessel **220** by a means to remove the (molecular) lower-energy hydrogen such as a selective venting valve means **201** which prevents the exothermic shrinkage reaction from coming to equilibrium. To control the reaction rate (the power output), an electric current can be passed through the material having a source of energy holes equal to the resonance shrinkage energy with current source **225**, and/or the first reaction vessel **200** and the second reaction vessel **220** are heated by heating means **230**. The heat output can be monitored with thermocouples present in at least the first vessel **200**, the second vessel **220**, and the heat exchanger **60** of FIG. **5**. The output power can be controlled by a computerized monitoring and control system which monitors the thermistors and controls the means to alter the power output. The (molecular) lower-energy hydrogen can be removed by a means **201** to prevent the exothermic shrinkage reaction from coming to equilibrium.

A method of preparation of the catalytic material of the present invention of catalytic systems that hinge on the transfer of an electron from a cation to another capable of producing energy holes for shrinking hydrogen atoms includes the steps of:

Mixing the oxides of the cations with the hydrogen dissociating material.

Thoroughly mixing by repeatedly sintering and pulverizing.

Example of a Ceramic Catalytic Material: Strontium Niobium Oxide ($SrNb_2O_6$) on Ni Powder

To prepare the ceramic catalytic material: strontium niobium oxide ($SrNb_2O_6$) on Ni powder, 2.5 kg of $SrNb_2O_6$ are added to 1.5 kg of −300 mesh Ni powder. The materials are mixed to make a homogeneous mixture. The powder can be sintered or calcinated in an oven at 1600° C. in atmospheric air for 24 hours. The material can be cooled and ground to remove lumps. The material can be re-sintered at 1600° C. in air for another 24 hours. The material can be cooled to room temperature and powderized.

A method of preparation of the catalytic material of the present invention of catalytic systems that hinge on the transfer of an electron from a cation to another capable of producing energy holes for shrinking hydrogen atoms includes the steps of:

Dissolving ionic salts of the cations into a solvent. In a preferred embodiment, the ionic salts are dissolved in deionized demineralized water to concentration of 0.3 to 0.5 molar.

Uniformly wetting a dissociation material with the dissolved salt solution.

Draining the excess solution.

Drying the wetted dissociation material in an oven preferably at a temperature of 220° C.

Pulverizing the dried catalytic material into a powder.

Example of a Ionic Catalytic Material: Potassium Carbonate (K_2CO_3) on Ni Powder

To prepare the ionic catalytic material: potassium carbonate (K_2CO_3) on Ni powder, a 1 liter solution of 0.5 M K_2CO_3

in water is poured over 500 grams of −300 mesh Ni powder. The materials are stirred to remove air pockets around the grains of Ni. The excess solution can be drained off. The powder can be dried in an oven at 200° C. If necessary the material can be ground to remove lumps.

Hydrogen Spillover Catalysts

In a preferred embodiment, the source of hydrogen atoms for the catalytic shrinkage reaction comprises a hydrogen spillover catalyst.

A hydrogen spillover catalyst according to the present invention comprises:

A hydrogen dissociation material or means which forms free hydrogen atoms or protons;

A conduit material onto which free hydrogen atoms spill and which supports free, mobile hydrogen atoms and provides a path or conduit for the flow of hydrogen atoms or protons;

A source of energy holes which catalyze the shrinkage reaction, and optionally

A support material into which the former materials are embedded as a mixture, compound, or solution.

Such hydrogen dissociation materials include surfaces or materials to dissociate hydrogen, deuterium, or tritium, comprise an element, compound, alloy, or mixture of transition elements and inner transition elements, iron, platinum, palladium, zirconium, vanadium, nickel, titanium, Sc, Cr, Mn, Co, Cu, Zn, Y, Nb, Mo, Tc, Ru, Rh, Ag, Cd, La, Hf, Ta, W, Re, Os, Ir, Au, Hg, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Vb, Lu, Th, Pa, U, activated charcoal (carbon), and intercalated Cs carbon (graphite). Such conduit materials onto which free hydrogen atoms spill and which supports free, mobile hydrogen atoms and which provides a path or conduit for the flow of hydrogen atoms include nickel, platinum, carbon, tin, iron, aluminum, and copper and their compounds, mixtures, or alloys. In an embodiment, such support materials into which the former materials are embedded as a mixture, compound, or solution includes carbon, silica, nickel, copper, titania, zinc oxide, chromia, magnesia, zirconia, alumina, silica-alumina, and zeolites. In an embodiment, one or more of the other components are deposited on the support material by electroplating. The source of energy holes to cause atomic hydrogen "shrinkage" are preferably of approximately mX27.21 eV and/or to cause molecular hydrogen "shrinkage" are of approximately mX48.6eV where m is an integer including the electrocatalytic ions and couples described in my previous U.S. Patent Applications entitled "Energy/Matter Conversion Methods and Structures", Ser. No. 08/467,051 filed on Jun. 6, 1995 which is a continuation-in-part application of Ser. No. 08/416,040 filed on Apr. 3, 1995 which is a continuation-in-part application of Ser. No. 08/107,357 filed on Aug. 16, 1993, which is a continuation-in-part application of Ser. No. 08/075,102 (Dkt. 99437) filed on Jun. 11, 1993, which is a continuation-in-part application of Ser. No. 07/626,496 filed on Dec. 12, 1990 which is a continuation-in-part application of Ser. No. 07/345,628 filed Apr. 28, 1989 which is a continuation-in-part application of Ser. No. 07/341,733 filed Apr. 21, 1989, which are incorporated herein by reference. The counterion of the energy hole of the spillover catalyst includes those given in the Handbook of Chemistry and Physics, Robert C. Weast, Editor, 58th Edition, CRC Press, West Palm Beach, Fla., (1974) pp. B61–B178 which is incorporated by reference herein, organic ions including benzoic acid, phthalate, salicylate, aryl sulfonate, alky sulfate, alkyl sulfonate, and alkyl carboxylate, and the anion of an acid which forms an acid anhydride including sulfite, sulfate, carbonate, bicarbonate, nitrite, nitrate, perchlorate, phosphite, hydrogen phosphite, dihydrogen phosphite, phosphate, hydrogen phosphate, and dihydrogen phosphate. In another embodiment the anion can be in equilibrium with its acid and its acid anhydride.

The functionalities of the hydrogen spillover catalyst are combined with the other functionalities as separate species or as combinations comprising a mixture, solution, compound, or alloy of more than one functionality. For example, in one embodiment, the hydrogen dissociation material and the source of energy holes each comprise homogeneous crystals— each crystal contains one component. and these functionalities are mixed with the conduit material without a support material. Whereas, in another embodiment, the hydrogen dissociation material and the source of energy holes comprise heterogeneous crystals-each crystal contains both of the components, and the heterogeneous crystals are mixed with the conduit material which coats a support material. In a third exemplary embodiment, the source of energy holes can be embedded in the conduit material. and this combined species can be mixed with the hydrogen dissociation material which can be embedded in the same or a different conduit material without a support material.

A method of preparation of the hydrogen spillover catalytic material of the present invention includes the steps of:

Mixing the components of the spillover catalyst by the method of incipient wetness impregnation.

Thoroughly mixing the components by sintering.

A further method of preparation of the hydrogen spillover catalytic material of the present invention includes the steps of:

Dissolving or dispersing the components to be mixed in a suitable solvent such as water and drying the solution or mixture.

Removing the solvent by drying, or the wet mixture, suspension, or solution can be frozen and the solvent can be sublimed.

Thoroughly mixing the components by sintering.

An incipient wetness method of preparation of the hydrogen spillover catalytic material of the present invention comprising a source of energy holes for shrinking hydrogen atoms that hinge on the transfer of an electron from a cation to another includes the steps of:

Dissolving a desired weight of the ionic salts of the cations into a desired volume of solvent. In a preferred embodiment, the ionic salts are dissolved in deionized demineralized water.

Preparing an incipiently wet conduit-hydrogen dissociation material by uniformly wetting the conduit-hydrogen dissociation material with the dissolved salt solution so that the pores of the material are just filled. The total volume of solvent required can be the desired amount, and the weight percent of the ionic salts of the cations in the final material can be determined by the desired weight of the ionic salts of the cations dissolved in the desired volume of solvent.

Mechanically mixing the wetted material to insure uniform wetting.

Drying the incipiently wet conduit-hydrogen dissociation material in an oven preferably at a temperature of 150° C. In an embodiment the material can be heated until the counterion(s) of the cations chemically decompose to preferably oxides.

Pulverizing the dried material comprised of the conduit-hydrogen dissociation-source of energy holes material into a powder.

- Optionally, mechanically mixing the dried and powdered material with further hydrogen dissociation material including a powder mixed with a conduit material and a support material.

Example of a Ionic Hydrogen Spillover Catalytic Material: 40% by Weight Potassium Nitrate (KNO_3) on 1%-Pd-on-Graphitic Carbon Powder

To prepare one kilogram of the ionic hydrogen spillover catalytic material: 40% by weight potassium nitrate (KNO_3) on 1%-Pd-on-graphitic carbon powder, 0.40 kg of KNO_3 are dissolved in 1 liter of H_2O. Incipient wetness requires 1 ml of H_2O per gram of −300 mesh graphite powder, and 0.67 grams of KNO_3 are required per gram of graphitic carbon powder to achieve a 40% by weight KNO_3 content in the final material. The aqueous KNO_3 solution can be slowly added to 0.6 kg of 1%-Pd-on-300-mesh-graphitic carbon powder as the slurry can be mixed. The slurry can be then placed on an evaporation dish which can be inserted into an oven at 150° C. for one hour. Heating causes the water to evaporate from the slurry. The KNO_3 coated 1%-Pd-on-graphitic carbon can be ground into a powder.

Another incipient wetness method of preparation of the hydrogen spillover catalytic material of the present invention comprising a source of energy holes for shrinking hydrogen atoms that hinge on the transfer of an electron from a cation to another includes the steps of:

- Dissolving a desired weight of the ionic salts of the cations into a desired volume of solvent. In a preferred embodiment, the ionic salts are dissolved in deionized demineralized water.
- Preparing an incipiently wet conduit material by uniformly wetting the conduit material with the dissolved salt solution so that the pores of the material are just filled. The total volume of solvent required can be the desired amount, and the weight percent of the ionic salts of the cations in the final material can be determined by the desired weight of the ionic salts of the cations dissolved in the desired volume of solvent.
- Mechanically mixing the wetted material to insure uniform wetting.
- Drying the incipiently wet conduit material in an oven preferably at a temperature of 150° C. In an embodiment, the material can be heated until the counterion(s) of the cations chemically decompose to preferably oxides.
- Pulverizing the dried material comprised of the conduit material and the source of energy holes into a powder.
- Mechanically mixing the dried and powdered material with a hydrogen dissociation material including a powder mixed with a conduit material and a support material.

Example of a Ionic Hydrogen Spillover Catalytic Material: 40% by Weight Potassium Nitrate (KNO_3) on Graphitic Carbon Powder with 5% by Weight 1%-Pd-on-Graphitic Carbon Powder

To prepare one kilogram of the ionic hydrogen spillover catalytic material: 40% by weight potassium nitrate (KNO_3) on graphitic carbon powder with 5% by weight 1%-Pd-on-graphitic carbon powder, 0.67 kg of KNO_3 are dissolved in 1 liter of H_2O. Incipient wetness requires 1 ml of H_2O per gram of −300 mesh graphite powder, and 0.40 grams of KNO_3 are required per gram of graphite powder to achieve a 40% by weight KNO_3 content in the final material. The aqueous KNO_3 solution can be slowly added to 0.55 kg of graphite powder as the slurry can be mixed. The slurry can be then placed on an evaporation dish which can be inserted into an oven at 150° C. for one hour. Heating causes the water to evaporate from the slurry. The KNO_3 coated graphite can be ground into a powder. The powder can be weighed. Approximately 50 grams (5% of the weight of the KNO_3 coated graphite) of 1%-Pd-on--300-mesh graphitic carbon powder can be mixed into the KNO_3 coated graphitic carbon powder.

Example of the Mode of Operation of the Exemplary Catalytic Materials

The catalytic material can be placed into the pressurizable vessel **200**. The vessel can be flushed with an inert gas such as He, Ar, or Ne to remove air contaminants in the vessel. The vessel and its contents are heated to the operational temperature, typically 100° C. to 400° C., before the vessel can be pressurized with hydrogen, typically 20 to 140 PSIG.

In an embodiment, the source of energy holes is potassium ions (K^+/K^+) or rubidium ions (Rb^+) intercalated into carbon. In another embodiment, the source of energy holes is an amalgam of the electrocatalytic ion or couple and its reduced metallic form such as rubidium ions (Rb^+) and rubidium metal or potassium ions (K^+/K^+) and potassium metal.

In an embodiment, the source of hydrogen atoms is a hydrogen dissociation means including a hydrogen gas stream blown over a hot filament or grid such as a hot refractory metal including a filament or grid of Ti, Ni, Fe, W, Au, Pt, or Pd at an elevated temperature such as 1800° C. The dissociation means provides hydrogen atoms as well as hydrogen ions, and the momentum of the atoms brings them in contact with the source of energy holes. Or, the hydrogen atoms and ions sputter onto the spillover catalyst. In one preferred embodiment of the pressurized gas reactor, a low pressure can be maintained by pressure regulator means **222** and a pump means **223** to minimize hydrogen atom recombination into molecular hydrogen and remove (molecular) lower-energy hydrogen.

In an embodiment the source of hydrogen atoms is water which dissociates to hydrogen atoms and oxygen on a water dissociation material such as an element, compound, alloy, or mixture of transition elements and inner transition elements, iron, platinum, palladium, zirconium, vanadium, nickel, titanium, Sc, Cr, Mn, Co, Cu, Zn, Y, Nb, Mo, Tc, Ru, Rh, Ag, Cd, La, Hf, Ta, W, Re, Os, Ir, Au, Hg, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Vb, Lu, Th, Pa, U, activated charcoal (carbon), and intercalated Cs carbon (graphite). In a further embodiment, the water dissociation material can be maintained at an elevated temperature by a heat source and temperature control means **230**. In an embodiment including one comprising a hydrogen spillover catalyst, the source of hydrogen can be from hydrocarbons including natural gas which can be reformed on a reforming a material such as nickel, cobalt, iron, or a platinum-group metal to hydrogen atoms and carbon dioxide. In a further embodiment, the reforming material can be maintained at an elevated temperature by a heat source and a temperature control means **230**. In another embodiment, the source of hydrogen atoms can be from the decomposition of a metal hydride where the decomposition can be controlled by controlling the temperature of the metal hydride with the heat source and temperature control means **230**. In another embodiment, the hydride can be coated by methods including electroplating with another material such as the hydrogen dissociation material.

In a preferred embodiment a product of the shrinkage reaction, (molecular) lower-energy hydrogen, can be removed to prevent product inhibition. Thus, the forward energy yielding reaction rate can be increased. One means to remove lower-energy (molecular) hydrogen is to supply the reaction mixture with a scavenger for lower-energy hydrogen. The scavenger absorbs or reacts with the product, lower-energy, hydrogen, and the resulting species can be removed from the reaction mixture. In another embodiment lower-energy hydrogen which is absorbed on the catalysts can be removed via displacement with an inert molecule or atom such as helium that flows through the vessel **200**.

Other objects, features, and characteristics of the art of catalysis as well as the methods of preparation, operation and the functions of the related elements, as described by Satterfield [Charles N. Satterfield, Heterogeneous Catalysis in Industrial Practice, Second Edition, McGraw-Hill, Inc., New York, (1991)] are applied to the present invention and are incorporated by reference herein. Application of the art of catalysis to the present invention of a pressurized gas energy reactor for the release of energy by the catalytic reaction wherein the electrons of hydrogen atoms undergo transitions to lower energy states include the use of an adiabatic reactor, fluidized-bed reactor, transport line reactor, multitube reactor, reverse multitube reactor having the heat exchange means including a fluid in the tubes and the catalytic material surrounding the tubes, and a multitube reactor or reverse multitube reactor comprising a fluidized bed of the catalytic material. Furthermore, in an embodiment comprising a solvated source of energy holes, a suspended hydrogen dissociation material including a hydrogen spillover catalyst, and hydrogen gas, the reactor comprises a trickle-bed reactor, a bubble-column reactor, or a slurry reactor.

For example, in a preferred embodiment, the fluidized bed reactor **200** comprises the hydrogen spillover catalytic material: 40% by weight potassium nitrate (KNO_3) on graphitic carbon powder with 5% by weight 1%-Pd-on-graphitic carbon powder. The reacting hydrogen gas can be passed up through a bed of the finely divided solid catalytic material, preferably having a particle size in the range of about 20 to 100 μm, which can be highly agitated and assumes many of the characteristics of a fluid. A cyclone separator **275** returns the fines to the bed. The hydrogen pressure and flow rate are controlled by pressure and flow rate control means **222**. Preferably at atmospheric or slightly higher pressures, the corresponding maximum linear velocity can be less than 60 cm/s.

Gaseous Source of Energy Holes

A preferred hydrogen gas energy reactor for the release of energy by an electrocatalytic and/or a disproportination reaction, wherein the electrons of hydrogen atoms undergo transitions to lower energy states in the gas phase, comprises a vessel **200** of FIG. **7** capable of containing a vacuum or pressures greater than atmospheric; a source of hydrogen **221**; a means **222** to control the pressure and flow of hydrogen into the vessel; a source of atomic hydrogen in the gas phase, and a source of energy holes in the gas phase.

The reaction vessel **200** comprises a vacuum or pressure vessel comprised of a temperature resistance material such as ceramic, stainless steel, tungsten, alumina, Incoloy, and Inconel.

In an embodiment, the source of hydrogen atoms in the gas phase is a hydrogen dissociation means including a hydrogen gas stream blown over a hot filament or grid **280** such as a hot refractory metal including a filament or grid of Ti, Ni, Fe, W, Au, Pt, or Pd at an elevated temperature such as 1800° C. The dissociation means provides hydrogen atoms as well as hydrogen ions, and the momentum of the atoms brings them in contact with the source of energy holes. In a preferred embodiment of the gaseous-source-of-energy-holes gas reactor, a low pressure can be maintained by pressure regulator means **222** and a pressure measurement and pump means **223** to minimize hydrogen atom recombination into molecular hydrogen. The pressure can be measured by measuring the power dissipated in the hot filament or grid which can be operated at constant resistance by a servo loop **285** comprising a voltage and current measurement means, a power supply, and a voltage and current controller where the hydrogen pressure versus power dissipation of the filament or grid at the operating resistance has been calibrated. In another embodiment, the source of atomic hydrogen comprises one or more hydrogen dissociation materials which provide hydrogen atoms by dissociation of molecular hydrogen. Such hydrogen dissociation materials include surfaces or materials to dissociate hydrogen, deuterium, or tritium, including a hydrogen spillover material such as palladium or platinum on carbon and an element, compound, alloy, or mixture of transition elements and inner transition elements, iron, platinum, palladium, zirconium, vanadium, nickel, titanium, Sc, Cr, Mn, Co, Cu, Zn, Y, Nb, Mo, Tc, Ru, Rh, Ag, Cd, La, Hf, Ta, W, Re, Os, Ir, Au, Hg, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Vb, Lu, Th, Pa, U, activated charcoal (carbon), and intercalated Cs carbon (graphite). In one embodiment, nonequilibrium conditions of the hydrogen and hydride are maintained by controlling the temperature and hydrogen pressure to provide atomic hydrogen in the gas phase. In another embodiment, the source of atomic hydrogen comprises a tungsten capillary which on the outlet can be heated by electron bombardment to 1800–2000 K such as the atomic hydrogen source described by Bischler [Bischler, U.; Bertel, E., J. Vac. Sci. Technol., A. (1993), 11(2), 458–60] which is incorporated herein by reference. In a further embodiment, the tungsten capillary can be heated by the energy released by the hydrogen shrinkage reaction. In another embodiment, the source of atomic hydrogen comprises an inductively coupled plasma flow tube such as that described by Gardner [Gardner, W. L., J. Vac. Sci. Technol., A. (1995), 13(3, Pt. 1), 763–6] which is incorporated herein by reference, and the hydrogen dissociation fraction can be measured with the sensor of Gardner.

The source of energy holes can be placed in a chemically resistant open container such as a ceramic boat **290** inside the reaction vessel. Or, the source of energy holes can be placed in a vessel which has a connection for the passage of the gaseous source of energy holes to the reaction vessel.

The cell can have a boat or container, which is connected to the reaction vessel, for containing the material used for forming the gaseous catalyst.

The gaseous source of energy holes includes those that sublime, boil, and/or are volatile at the elevated operating temperature of the gas energy reactor wherein the shrinkage reaction occurs in the gas phase. For example, $RbNO_3$ and KNO_3 are each volatile at a temperature much less than that at which each decomposes [C. J. Hardy, B. O. Field, J. Chem. Soc., (1963), pp. 5130–5134]. In one embodiment, the ionic hydrogen spillover catalytic material: 40% by weight potassium or rubidium nitrate on graphitic carbon powder with 5% by weight 1%-Pd-on-graphitic carbon powder can be operated at a temperate at which the potassium or rubidium nitrate can be volatile. Further disproportionation reactions of the product, lower-energy hydrogen atoms, release additional heat energy.

In a preferred embodiment, the source of energy holes is a thermally stable salt of rubidium or potassium such as RbF, RbCl, RbBr, RbI, Rb_2S_2, RbOH, Rb_2SO_4, Rb_2CO_3, Rb_3PO_4, and KF, KCl, KBr, KI, K_2S_2, KOH, K_2SO_4,

K_2CO_3, K_3PO_4, K_2GeF_4. Further preferred sources of energy holes of approximately mX27.21 eV to cause atomic hydrogen "shrinkage" and/or approximately mX48.6 eV to cause molecular hydrogen "shrinkage" where m is an integer include the electrocatalytic ions and couples described in my previous U.S. Patent Applications entitled "Energy/Matter Conversion Methods and Structures", Ser. No. 08/467,051 filed on Jun. 6, 1995 which is a continuation-in-part application of Ser. No. 08/416,040 filed on Apr. 3, 1995 which is a continuation-in-part application of Ser. No. 08/107,357 filed on Aug. 16, 1993, which is a continuation-in-part application of Ser. No. 08/075,102 (Dkt. 99437) filed on Jun. 11, 1993, which is a continuation-in-part application of Ser. No. 07/626,496 filed on Dec. 12, 1990 which is a continuation-in-part application of Ser. No. 07/345,628 filed Apr. 28, 1989 which is a continuation-in-part application of Ser. No. 07/341,733 filed Apr. 21, 1989, which are incorporated herein by reference. The counterion includes those given in the Handbook of Chemistry and Physics, Robert C. Weast, Editor, 58th Edition, CRC Press, West Palm Beach, Fla., (1974) pp. B61–B178 which is incorporated by reference herein. The preferred anion can be stable to hydrogen reduction and thermal decomposition and can be volatile at the operating temperature of the energy reactor.

The catalyst may be an ionic compound which is resistant to hydrogen reduction. Moreover, the catalyst is adapted to provide gaseous atoms which may be ionized.

The following compounds are preferred gaseous sources of energy holes in the gas energy reactor. Higher temperatures result in a higher vapor pressure of the source of energy holes which increases the reaction rate; however, the increase in total pressure increases the recombination rate of hydrogen atoms to hydrogen molecules. In each exemplary case that follows, the operating temperature of the energy reactor can be that which provides an optimal reaction rate. In an embodiment, the cell temperature can be about 50° C. higher than the (highest) melting point of the source of energy holes (in the case that the source of energy holes comprises an electron transfer between two cations—an electrocatalytic couple). The hydrogen pressure can be maintained at about 200 millitorr, and molecular hydrogen can be dissociated with a hot filament or grid **280** of FIG. **7**.

Single Ion Catalysts (Electrocatalytic Ions)

Single-ion catalysts (electrocatalytic ions) capable of producing energy holes for shrinking hydrogen atoms. The number following the atomic symbol (n) is the nth ionization energy of the atom. That is for example, $Rb^+ + 27.28\ eV = Rb^{2+} + e^-$. (melting point=(MP); boiling point=(BP))

Catalytic Ion	n	nth ionization energy
Mo^{2+}	3	27.16
MoI_2		
Ti^{2+}	3	27.49
$TiCl_2$(MP = subl H_2,BP = d 475° C. vac)		
($TiCl_4/Ti_{metal}$)		
Rb^{1+}	2	27.28
$RbNO_3$(MP = 310° C.,BP = subl)		
Rb_2S_2(MP = 420° C.,BP = volat > 850)		
RbI(MP = 647° C.,BP = 1300° C.)		

Two Ion Catalyts (Electrocatalytic Couples)

Two-ion catalysts (electrocatalytic couples) capable of producing energy holes for shrinking hydrogen atoms. The number in the column following the ion, (n), is the nth ionization energy of the atom. That is for example, $K^+ + 31.63\ eV = K^{2+} + e^-$ and $K^+ + e^- = K + 4.34$ eV. (melting point=(MP); boiling point=(BP))

Atom Oxidized	n	nth Ionization Energy (ev)	Atom Reduced	n	nth Ionization Energy (ev)	Energy Hole (ev)
Sn 4+	5	72.28	Si 4+	4	45.14	27.14
$SnCl_4$(MP = −33° C.,BP = 114.1° C.)			$SiCl_4$(MP = −70° C., BP = 57.57° C.)			
Pr 3+	4	38.98	Ca 2+	2	11.87	27.11
$PrBr_3$(MP = 691° C.,BP = 1547° C.)			$CaBr_2$(MP = 730° C. sl d, BP = 806 − 812° C.)			
Sr 2+	3	43.60	Cr 2+	2	16.50	27.10
$SrCl_2$(MP = 875° C.,BP = 1250° C.)			CrI_2(MP = 856° C., BP = 800 sub vac° C.)			
Cr 3+	4	49.10	Tb 3+	3	21.91	27.19
CrF_3(MP = >1000° C.,BP = 1100 − 1200° C. subl)			TbI_3(MP = 946° C.,BP > 1300° C.)			
Sb 3+	4	44.20	Co 2+	2	17.06	27.14
$SbCl_3$(MP = 73.4° C.,BP = 283° C.)			$CoCl_2$(MP = 724° C. in HCl gas, BP = 1049° C.)			
Bi 3+	4	45.30	Ni 2+	2	18.17	27.13
$BiCl_3$(MP = 230 − 232° C.,BP = 447° C.)			$NiCl_2$(MP = 1001° C., BP = 973° C. subl)			
Pd 2+	3	32.93	In 1+	1	5.79	27.14
PdF_2 (MP = volat)			InCl(MP = 225° C.,BP = 608° C.)			
La 3+	4	49.95	Dy 3+	3	22.80	27.15
$LaCl_3$(MP = 860° C.,BP > 1000° C.)			$DyCl_3$(MP = 718° C., BP = 1500° C.)			
La 3+	4	49.95	Ho 3+	3	22.84	27.11
LaI_3(MP = 772° C.)			HoI_3(MP = 989° C.,BP = 1300° C.)			
K 1+	2	31.63	K 1+	1	4.34	27.28
KNO_3(MP = 334° C.,BP = subl)			KNO_3(MP = 334° C.,BP = subl)			
K_2S_2(MP = 470° C.)			K_2S_2(MP = 470° C.)			
KI(MP = 681° C.,BP = 1330° C.)			KI(MP = 681° C.,BP = 1330° C.)			
V 3+	4	46.71	Pd 2+	2	19.43	27.28
VF_3(MP > 800° C.,BP Subl)			PbF_2(MP = 855° C.,BP = 1290° C.)			
VOCl(BP = 127° C.)			PbI_2(MP = 402° C.,BP = 954° C.)			
Lu 3+	4	45.19	Zn 2+	2	17.96	27.23
$LuCl_3$(MP = 905° C.,BP = subl 750° C.)			$PbCl_2$(MP = 283° C., BP = 732° C.)			
As 3+	4	50.13	Ho 3+	3	22.84	27.29
AsI_3(MP = 146° C.,BP = 403° C.)			HoI_3(MP = 989° C., BP = 1300° C.)			
Mo 5+	6	68.00	Sn 4+	4	40.73	27.27
$MoCl_5$(MP = 194° C.,BP = 268° C.)			$SnCl_4$(MP = −33° C., BP = 114.1° C.)			
Sb 3+	4	44.20	Cd 2+	2	16.91	27.29
SbI_3(MP = 170° C.,BP = 401° C.)			CdI_2(MP = 387° C., BP = 796° C.)			
Ag 2+	3	34.83	Ag 1+	1	7.58	27.25
AgF_2(MP = 690° C.,BP = 700° C. d)			AgF(MP = 435° C., BP = 1159° C.)			
La 3+	4	49.95	Er 3+	3	22.74	27.21
LaI_3(MP = 772° C.,BP = 1000° C.)			ErI_3(MP = 1020° C., BP = 1280° C.)			
V 4+	5	65.23	B 3+	3	37.93	27.30
VCl_4(MP = −28° C.,BP = 148.5° C.)			BCl_3(MP = −107.3° C., BP = 12.5° C.)			
Fe 3+	4	54.80	Ti 3+	3	27.49	

-continued

Atom Oxidized Energy Hole (ev)	n	nth Ionization Energy (ev)	Atom Reduced	n	nth Ionization Energy (ev)
27.31					
$FeCl_3$(MP = 306° C.,BP = 315° C. d) $TiCl_3$(MP = 440° C. d, BP = 660° C.)					
Co 2+	3	33.50	Tl 1+	1	6.11
27.39					
CoI_2(MP = 515 vac° C.,BP = 570° C. vac) TlI(MP = 440° C. d, BP = 823° C.)					
CoF_2(MP = 1200° C.,BP = 1400° C.) TIF(MP= 327° C. d, BP = 655° C.)					
Bi 3+	4	45.30	Zn 2+	2	17.96
27.34					
$BiBr_3$(MP = 218° C.,BP = 453° C.) $ZnBr_2$(MP = 394° C. d, BP = 650° C.)					
As 3+	4	50.13	Dy 3+	3	22.80
27.33					
AsI_3(MP = 146° C.,BP = 403° C.) DyI_3(MP = 955° C. d, BP = 1320° C.)					
Ho 3+	4	42.50	Mg 2+	2	15.03
27.47					
$HoCl_3$(MP = 718° C.,BP = 1500° C.) $MgCl_2$(MP = 714° C., BP = 1412° C.)					
K 1+	2	31.63	Rb 1+	1	4.18
27.45					
KI(MP = 618° C.,BP = 1330° C.) RbI(MP = 647° C.,BP = 1300° C.)					
Cr 3+	4	49.10	Pr 3+	3	21.62
27.48					
$CrCl_3$(MP = 1150° C.,BP = 1300° C. subl) $PrCl_3$(MP = 786° C., BP = 1700° C.)					
Sr 2+	3	43.60	Fe 2+	2	16.18
27.42					
$SrCl_2$(MP = 875° C.,BP = 1250° C.) $FeCl_2$(MP = 670° C.,BP subl)					
Ni 2+	3	35.17	Cu 1+	1	7.73
27.44					
$NiCl_2$(MP = 1001° C.,BP = 973° C. subl) CuCl(MP = 430° C., BP = 1490° C.)					
Sr 2+	3	43.60	Mo 2+	2	16.15
27.45					
$SrCl_2$(MP = 875° C.,BP = 1250° C.) $MoCl_2$					
Y 3+	4	61.80	Zr 4+	4	34.34
27.46					
YCl_3(MP = 721° C.,BP = 1507° C.) $ZrCl_4$(MP = 437° C., BP = 331° C. subl)					
Cd 2+	3	37.48	Ba 2+	2	10.00
27.48					
CdI_2(MP = 387° C.,BP = 796° C.) BaI_2(MP = 740° C.)					
Ho 3+	4	42.50	Pb 2+	2	15.03
27.47					
HoI_3(MP = 989° C.,BP = 1300° C.) PbI_2(MP = 402° C., BP= 954° C.)					
Pd 2+	3	32.93	Li 1+	1	5.39
27.54					
PdF_2(MP = volat) LiF(MP = 845° C.,BP = 1676° C.)					
Eu 3+	4	42.60	Mg 2+	2	15.03
27.56					
$EuCl_3$(MP = 850° C.) $MgCl_2$(MP = 714° C.,BP = 1412° C.)					
Er 3+	4	42.60	Mg 2+	2	15.03
27.56					
$ErCl_3$(MP = 774° C.,BP = 1500° C.) $MgCl_2$(MP = 714° C., BP = 1412° C.)					
Bi 4+	5	56.00	Al 3+	3	28.45
27.55					
$BiCl_4$(MP = 226° C.) $AlCl_3$(MP = 190° C.,BP = 177.8° C. subl)					
Ca 2+	3	50.91	Sm 3+	3	23.40
27.51					
$CaBr_2$(MP = 730° C. sl d, BP = 806 – 812° C.) $SmBr_3$(MP subl > 1000° C.)					
V 3+	4	46.71	La 3+	3	19.18
27.53					
VaF_3(MP > 800° C., subl) $LaCl_3$(MP = 860° C.,BP > 1000° C.)					
Gd 3+	4	44.00	Cr 2+	2	16.50
27.50					
GdI_3(MP = 926° C.,BP = 1340° C.) CrI_2(MP = 856° C., BP = 800° C. subl vac)					
Mn 2+	3	33.67	Ti 1+	1	6.11
27.56					
MnI_2(MP = 638° C. vac, BP = 500° C. subl vac) TIF(MP = 327° C., BP = 655° C.)					
Yb 3+	4	43.70	Fe 2+	2	16.18
27.52					
$YbBr_3$(MP = 956° C.,BP = d) $FeBr_2$(MP = 684° C. d)					
Ni 2+	3	35.17	Ag 1+	1	7.58
27.59					
$NiCl_2$(MP = 1001° C.,BP = 973° C. subl) AgCl(MP = 455° C., BP = 1550° C.)					
Zn 2+	3	39.72	Yb 2+	2	12.18
27.54					
$ZnCl_2$(MP = 283° C.,BP = 732° C. subl) $YbCl_2$(MP = 702° C., BP = 1900° C.)					
Se 4+	5	68.30	Sn 4+	4	40.73
27.57					
SeF_4(MP = –13.8° C.,BP > 100° C.) $SnCl_4$(MP = –33° C., BP = 114.1° C.) SnF_4(MP = 705° C.subl)					
Sb 3+	4	44.20	Bi 2+	2	16.69
27.51					
SbI_3(MP = 170° C.,BP = 401° C.) BiI_2(MP = 400° C., BP = subl vac)					
Eu 3+	4	42.60	Pb 2+	2	15.03
27.57					
EuF_3(MP = 1390° C.,BP = 2280° C.) $PbCl_2$(MP = 501° C., BP = 950° C.)					

In an embodiment wherein the anion can be reduced by hydrogen, the anion is chemically stabilized. For example, the product of the reduction is added to the gas cell to stabilize the anion. In a further embodiment, the anion can be replaced continuously or intermittently. In the case of the nitrate ion, the product ammonia can be removed from the vessel. oxidized to nitrate, and returned to the cell. In one embodiment, the product ammonia can be removed from the vessel by collection in a condenser and can be oxidized to nitrate on a platinum or iridium screen at elevated temperatures such as 912° C. In a further embodiment, the nitrate ion to ammonia reaction can be minimized by decreasing the hydrogen pressure while optimizing the vapor phase catalytic hydrogen shrinkage reaction. In an embodiment, a low pressure of hydrogen atoms can be generated by dissociation of molecular hydrogen on a hot filament or grid **280** of FIG. **7**. A low pressure of molecular hydrogen can be maintained via the hydrogen supply **221**, the hydrogen flow control means **222**, and the hydrogen pressure measurement and vacuum means **223**. The hydrogen pressure can be maintained at a low pressure by adjusting the supply through the inlet with flow controller **222** versus the amount pumped away at the outlet by the pressure measurement and pump means **223**. The pressure can be adjusted to maximize the output power while minimizing the degradation of nitrate. The optimal hydrogen pressure can be less than about one torr. In an embodiment, the source of hydrogen atoms in the gas phase can be a hydrogen dissociation means including a hydrogen gas stream blown over a hot filament or grid **280** such as a hot refractory metal including a filament or grid of Ti, Ni, Fe, W, Au, Pt, or Pd at an elevated temperature such as 1800° C. The hydrogen molecular source can be directed over the filament or grid and onto the gaseous source of energy holes. The pressure and flow of the hydrogen atoms prohibits the collision of the counterion of the source of energy holes (such as the nitrate ion) from contacting the hot filament or grid. Thus, the thermal decomposition or reduction of the anion on the filament or grid can be prevented. In another embodiment, a negative potential can be maintained as a grid electrode **287** surrounding the filament or grid. The grid electrode permits the passage of hydrogen atoms from the filament or grid and repels the anion from contacting the hot filament or grid. Thus, the thermal or chemical breakdown of the anion (couterion) can be prevented.

In an embodiment, the source of energy holes is an electrocatalytic ion or electrocatalytic couple comprising cation-anion pairs in the gas phase wherein the cation-anion pairs are dissociated by external source means **75** of FIG. **5** which includes, for example, a particle source **75***b* and/or photon source **75***a* and/or a source of heat, acoustic energy, electric fields, or magnetic fields. In a preferred embodiment, the cation-anion pairs are thermally dissociated by heat source **230** or photodissociated by photon source **205** of FIG. **7**.

In another embodiment of the gas energy reactor having a gaseous source of energy holes, the source of energy holes is atomized with an atomizer means **295** to provide a gaseous source of energy holes. In a preferred embodiment of the atomizer, atoms are boiled, sublimed, or vaporized by a heating means such as the boat heating means **299**, and the gaseous atoms are ionized to form a source of energy holes including the electrocatalytic ions or electrocatalytic couples of my previous patent applications incorporated herein by reference. In one embodiment, the atoms are thermally ionized by the heating means **230**, by the hydrogen atom source **280** including a hot filament or grid, or by an inductively coupled plasma flow tube. For example, the gas energy cell shown in FIG. **7** comprises rubidium or potassium metal in the boat **290** which has a vapor pressure that can be controlled by controlling the temperature of the boat by heating means **230** and or **299**. Hydrogen molecules are dissociated to atoms on the hot filament or grid **280**. The rubidium (potassium) metal in the gas phase can be ionized to Rb^+ (K^+) by the same or different hot filament or grid **280**. The Rb^+ (K^+/K^+) electrocatalytic ion (couple) serves as a source of energy holes to shrink the hydrogen atoms. In another embodiment, the hot filament or grid **280** comprises a metal(s) or can be electroplated with a metal(s) which boils off as a cation(s) that are a source of energy holes. For example, Mo^{2+} ions (Mo^{2+} electrocatalytic ion) enter the gas phase of the energy cell **200** from the hot molybdenum filament or grid **280**. The hot molybdenum filament or grid **280** also dissociates hydrogen molecules to hydrogen atoms. For a further example, Ni^{2+} and Cu^+ ions (Ni^{2+}/Cu^+ electrocatalytic couple) enter the gas phase of the energy cell **200** from the hot nickel and hot copper or hot nickel-copper alloy filament or grid **280**. In another embodiment, the photon source **75***a* and the particle source **75***b* of FIG. **5**, including an electron beam, ionize species such as atoms in the gas phase to form the source of energy holes including the electrocatalytic ions or electrocatalytic couples of my previous patent applications incorporated herein by reference. In another embodiment, the atoms or ions are ionized chemically by a volatilized reactant such as an ionic species which oxidizes or reduces the atoms or ions to form a source of energy holes.

The power of the gas energy reactor can be controlled by controlling the amount of the source of energy holes (electrocatalytic ion or couple) in the gas phase and/or by controlling the concentration of atomic or lower-energy hydrogen. The concentration of the gaseous source of energy holes (electrocatalytic ion or couple) can be controlled by controlling the initial amount of the volatile source of energy holes (electrocatalytic ion or couple) present in the reactor, and/or by controlling the temperature of the reactor with temperature control means **230** which determines the vapor pressure of the volatile source of energy holes (electrocatalytic ion or couple). The reactor temperature further controls the power by changing the rate of the catalytic hydrogen shrinkage reaction. The concentration of atomic hydrogen can be controlled by controlling the amount of atomic hydrogen provided by the atomic hydrogen source **280**. For example, the amount of hydrogen atoms in the gas phase can be controlled by controlling the flow of hydrogen over or through the hot filament or grid, the tungsten capillary heated by electron bombardment, or the inductively coupled plasma flow tube; by controlling the power dissipated in the inductively coupled plasma flow tube; by controlling the temperature of the hot filament or grid, or the tungsten capillary heated by electron bombardment; by controlling the pressure of the hydrogen and temperature of the hydride maintained under nonequilibrium conditions, and by controlling the rate of removal of recombined hydrogen from the cell by pump means **223**. Another means to control the shrinkage reaction rate can be by controlling the pressure of a non reactive gas with non reactive gas source **299**, non reactive gas flow control means **232**, and pressure measurement and pump means **223**. The non reactive gas such as a noble gas competes with collisions between the source of energy holes (electrocatalytic ion or couple) and hydrogen atoms or competes with collisions that yield lower-energy hydrogen disproportionation reactions. Noble gases include He, Ne, and Ar. Further such reaction non reactive "reaction quenching" gases include carbon dioxide and nitrogen.

The hydrogen partial pressure can be further controlled by throttling hydrogen into the cell by a hydrogen value control means **222** while monitoring the pressure with a pressure measurement means **222** and **223**. In a preferred embodiment, the hydrogen pressure can be controlled by controlling the temperature with heating means **230** of the gas energy reactor which further comprises a hydrogen storage means such as a metal hydride or other hydride including saline hydrides, titanium hydride, vanadium, niobium, and tantalum hydrides, zirconium and hafnium hydrides, rare earth hydrides, yttrium and scandium hydrides, transition element hydrides, intermetalic hydrides, and their alloys known in the art as given by W. M. Mueller, J. P. Blackledge, and G. G. Libowitz, *Metal Hydrides*, Academic Press, New York, (1968), *Hydrogen in Intermetalic Compounds I*, Edited by L. Schlapbach, Springer-Verlag, Berlin, and *Hydrogen in Intermetalic Compounds II*, Edited by L. Schlapbach, Springer-Verlag, Berlin, which are incorporated by reference herein. The temperature of the cell can be controlled by a temperature control and measurement means **230** such that the vapor pressure of the hydrogen in equilibrium with the hydrogen storage material can be the desired pressure. In one embodiment, nonequilibrium conditions of the hydrogen and hydride are maintained by controlling the temperature and hydrogen pressure to provide atomic hydrogen. In several embodiments, the hydrogen storage means can be a rare earth hydride with an operating temperature of about 800° C.; lanthanum hydride with an operating temperature of about 700° C.; gadolinium hydride with an operating temperature of about 750° C.; neodymium hydride with an operating temperature of about 750° C.; yttrium hydride with an operating temperature of about 800° C.; scandium hydride with an operating temperature of about 800° C.; ytterbium hydride with an operating temperature of about 850–900° C.; titanium hydride with an operating temperature of about 450° C.; cerium hydride with an operating temperature of about 950° C.; praseodymium hydride with an operating temperature of about 700° C.; zirconium-titanium (50%/50%) hydride with an operating temperature of about 600° C.; an alkali metal/alkali metal hydride mixture such as Rb/RbH or K/KH with an operating temperature of about 450° C., and an alkaline earth metal/alkaline earth hydride mixture such as Ba/BaH_2 with an operating temperature of about 900–1000° C.

The heat output can be monitored with thermocouples present in at least the vessel **200** and the heat exchanger **60** of FIG. **5**. The rate of the shrinkage reaction rate can be monitored by ultraviolet or electron spectroscopy of the photons or electrons emitted via lower-energy hydrogen transitions, by X-ray photoelectron spectroscopy (XPS) of lower-energy hydrogen, and by mass spectroscopy, Raman or infrared spectroscopy, and gas chromatography of the molecular lower-energy hydrogen (dihydrino). Lower-energy hydrogen atoms and molecules are identified by XPS as higher binding energy species than normal hydrogen. The dihydrino can be identified by mass spectroscopy as a species with a mass to charge ratio of two (m/e=2) that has a higher ionization potential than that of normal hydrogen by recording the ion current as a function of the electron gun energy. The dihydrino can be identified by gas chromatography at low temperature such as gas chromatography with an activated carbon (charcoal) column at liquid nitrogen temperature or with a column that will separate para from ortho hydrogen such as an Rt-Alumina column, or a HayeSep column at liquid nitrogen temperature wherein normal hydrogen can be retained to a greater extent than dihydrino. The dihydrino can be identified by Raman and infrared spectroscopy as a molecule with higher vibrational and rotational energy levels as compared to those of normal hydrogen. The output power can be controlled by a computerized monitoring and control system which monitors the thermistors, spectrometers, and gas chromatograph and controls the means to alter the power output. The (molecular) lower-energy hydrogen can be removed by a means **201** to prevent the exothermic shrinkage reaction from coming to equilibrium.

In another embodiment of the gas energy reactor having a gaseous source of energy holes, hydrogen atoms are produced by a pyrolysis reaction such as the combustion of a hydrocarbon wherein the catalytic source of energy holes can be in the gas phase with the hydrogen atoms. In a preferred mode, the pyrolysis reaction occurs in an internal combustion engine whereby the hydrocarbon or hydrogen containing fuel comprises a source of energy holes that are vaporized (become gaseous) during the combustion. In a preferred mode, the source of energy holes (electrocatalytic ion or couple) is a thermally stable salt of rubidium or potassium such as RbF, RbCl, RbBr, RbI, Rb_2S_2, RbOH, Rb_2SO_4, Rb_2CO_3, Rb_3PO_4, and KF, KCl, KBr, KI, K_2S_2, KOH, K_2SO_4, K_2CO_3, K_3PO_4, K_2GeF_4. Additional counterions of the electrocatalytic ion or couple include organic anions including wetting or emulsifying agents. In another embodiment, the hydrocarbon or hydrogen containing fuel further comprises water as a mixture and a solvated source of energy holes including emulsified electrocatalytic ions or couples. During the pyrolysis reaction, water serves as a further source of hydrogen atoms which undergo a shrinkage reaction catalyzed by the source of energy holes wherein the water can be dissociated to hydrogen atoms thermally or catalytically on a surface such as the cylinder or piston head which can be comprised of material which dissociates water to hydrogen and oxygen. The water dissociation material includes an element, compound, alloy, or mixture of transition elements and inner transition elements, iron, platinum, palladium, zirconium, vanadium, nickel, titanium, Sc, Cr, Mn, Co, Cu, Zn, Y, Nb, Mo, Tc, Ru, Rh, Ag, Cd, La, Hf, Ta, W, Re, Os, Ir, Au, Hg, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Vb, Lu, Th, Pa, U. activated charcoal (carbon), and intercalated Cs carbon (graphite).

POWER DENSITY OF GAS ENERGY REACTOR (GAS PHASE HYDROGEN SHRINKAGE REACTION)

The equations numbers which follow referred to those given by Mills [Mills, R., *The Grand Unified Theory of Classical Quantum Mechanics*, (1995), Technomic Publishing Company, Lancaster, Pa.]. The rate of the disproportionation reaction, $r_{m,m',p}$, to cause resonant shrinkage, Eqs. (5.22–5.30), is dependent on the collision rate between the reactants and the efficiency of resonant energy transfer. It is given by the product of the rate constant, $k_{m,m',p}$, (Eq. (5.47)), the total number of hydrogen or hydrino atoms, N_H, and the efficiency, E (Eq. (6.33)), of the transfer of the resonance shrinkage energy from the donor hydrino atom to the energy hole provided by the acceptor hydrino atom,

$$E = \frac{1}{1 + \left[\frac{r}{R_0}\right]^6} \tag{6.37}$$

$$R_0^6 = (8.8 \times 10^{-25}) J \eta^{-4} \phi_D^0 \kappa^2$$

where r is the distance between the donor and the acceptor, J is the overlap integral between the resonance shrinkage energy distribution of the donor hydrino atom and the distribution of the energy hole provided by the acceptor hydrino atom, η is the dielectric constant, and κ^2 is a function of the mutual orientation of the donor and acceptor transition moments. Electronic transitions of lower-energy hydrogen atoms occur only by nonradiative energy transfer; thus, the quantum yield of the fluorescence of the donor, Φ_D, of Eq. (6.37) is equal to one. The rate of the disproportionation reaction, $r_{m,m',p}$, to cause resonant shrinkage is

$$r_{m,m',p} = E N_H 4\pi \frac{1}{2} \sqrt{2}\, n_H \left(\frac{a_H}{p}\right)^2 \sqrt{\frac{3kT}{m_H}} \tag{6.38}$$

The factor of one half in Eq. (6.38) corrects for double counting of collisions [Levine, I., Physical Chemistry, McGraw-Hill Book Company, New York, (1978), pp. 420–421]. The power, $P_{m,m',p}$, is given by the product of the rate of the transition, Eq. (6.38), and the energy of the disproportination reaction (Eq. (5.27)).

$$P_{m,m',p} = E \frac{N_H^2}{V} 4\pi \frac{1}{\sqrt{2}} \left(\frac{a_H}{p}\right)^2 \sqrt{\frac{3kT}{m_H}} [2pm + m^2 - m'^2 + 1] \times 2.2 \times 10^{-18} \text{ W} \tag{6.39}$$

where V is the volume. For a disproportionation reaction in the gas phase, the energy transfer efficiency is one. The power given by substitution of

$$E=1,\ p=2,\ m=1,\ m'=2,\ V=1\ m^3,\ N=3\times10^{21},\ T=675\ K \tag{6.40}$$

into Eq. (6.39) is

$$P_{m,m',p}=1\text{ GW}(1\text{ kW/cm}^3) \tag{6.41}$$

In the case that the reaction of hydrogen to lower-energy states occurs by the reaction of a catalytic source of energy holes with hydrogen or hydrino atoms, the reaction rate is dependent on the collision rate between the reactants and the efficiency of resonant energy transfer. The hydrogen-or-hydrino-atom/electrocatalytic-ion collision rate per unit volume,

$$Z_{H[\frac{a_H}{p}]Catalyst},$$

for a gas containing n_H hydrogen or hydrino atoms per unit volume, each with radius

$$\frac{a_H}{p}$$

and velocity v_H and n_c electrocatalytic ions per unit volume, each with radius $r_{Catalyst}$ and velocity v_c is given by Levine [Levine, I., Physical Chemistry, McGraw-Hill Book Company, New York, (1978), pp. 420–421].

$$Z_{H[\frac{a_H}{p}]Catalyst} = \pi\left(\frac{a_H}{p} + r_{Catalyst}\right)^2\left[\langle v_H\rangle^2 + \langle v_C\rangle^2\right]^{1/2} n_H n_C \tag{6.42}$$

The average velocity, v_{avg}, can be calculated from the temperature, T, [Bueche, F. J., Introduction to Physics for Scientists and Engineers, McGraw-Hill Book Company, New York, (1986), pp. 261–265].

$$\frac{1}{2}m_H v_{avg}^2 = \frac{3}{2}kT \tag{6.43}$$

where k is Boltzmann's constant. Substitution of Eq. (5.44) into Eq. (5.42) gives the collision rate per unit volume,

$$Z_{H[\frac{a_H}{p}]Catalyst},$$

in terms of the temperature, T.

$$Z_{H[\frac{a_H}{p}]Catalyst} = \pi\left(\frac{a_H}{p} + r_{Catalyst}\right)^2\left[3kT\left(\frac{1}{m_H} + \frac{1}{m_C}\right)\right]^{1/2} n_H n_C \tag{6.44}$$

The rate of the catalytic reaction, $r_{m,p}$, to cause resonant shrinkage is given by the product of the collision rate per unit volume,

$$Z_{H[\frac{a_H}{p}]Catalyst},$$

the volume, V, and the efficiency, E, of resonant energy transfer given by Eq. (6.37).

$$r_{m,p} = E\pi\left(\frac{a_H}{p} + r_{Catalyst}\right)^2\left[3kT\left(\frac{1}{m_H} + \frac{1}{m_C}\right)\right]^{1/2}\frac{N_H N_C}{V} \tag{6.45}$$

The power, $P_{m,p}$, is given by the product of the rate of the transition, Eq. (6.45), and the energy of the transition, Eq. (5.8).

$$P_{m,p} = E\pi\left(\frac{a_H}{p} + r_{Catalyst}\right)^2\left[3kT\left(\frac{1}{m_H} + \frac{1}{m_C}\right)\right]^{1/2}\frac{N_H N_C}{V}[2mp + m^2]\times 2.2\times 10^{-18}\text{ W} \tag{6.46}$$

In the case of a gas phase catalytic shrinkage reaction wherein the source of energy holes is a single cation having an ionization energy of 27.21 eV with hydrogen or hydrino atoms, the energy transfer efficiency is one. Rubidium (Rb^+) is an electrocatalytic ion with a second ionization energy of 27.28 eV. The power for the reaction given by Eqs. (5.9). (5.10), and (5.8) with the substitution of

$$E=1,\ p=1,\ m=1,\ V=1\text{ m}^3,\ N_H=3\times10^{21},\ N_c=3\times10^{21},\ m_c=1.4\times10^{-25}\text{ kg},\ r_c=2.16\times10^{-10}\text{ m},\ T=675\text{ K} \tag{6.47}$$

into Eq. (6.46) is

$$P_{m,p}=55\text{ GW }(55\text{ kW/cm}^3) \tag{6.48}$$

In the case that the catalytic reaction of hydrogen to lower-energy states occurs on a surface, the energy transfer efficiency is less than one due to differential surface interactions of the absorbed hydrogen or hydrino atoms and the electrocatalytic ion. The power given by Eqs. (6.46) and (6.47) with

$$E=0.001 \tag{6.49}$$

is

$$P_{m,p}=55\text{ MW }(55\text{ W/cm}^3) \tag{6.50}$$

Less efficient catalytic systems hinge on the coupling of three resonator cavities. For example, an electron transfer occurs between two cations which comprises an energy hole for a hydrogen or hydrino atom. The reaction rate is dependent on the collision rate between catalytic cations and hydrogen or hydrino atoms and the efficiency of resonant energy transfer with a concomitant electron transfer with each shrinkage reaction. The rate of the catalytic reaction, $r_{m,p}$, to cause resonant shrinkage is given by the product of the collision rate per unit volume,

$$Z_{H[\frac{a_H}{p}]Catalyst},$$

the volume, V, and the efficiency, E_e, of resonant energy transfer given by Eq. (6.37) where r is given by the average distance between cations in the reaction vessel.

$$r_{m,p} = E_c\pi\left(\frac{a_H}{p} + r_{Catalyst}\right)^2\left[3kT\left(\frac{1}{m_H} + \frac{1}{m_C}\right)\right]^{1/2}\frac{N_H N_C}{V} \tag{6.51}$$

The power, $P_{m,p}$, is given by the product of the rate of the transition, Eq. (6.51), and the energy of the transition, Eq. (5.8).

$$P_{m,p} = E_c\pi\left(\frac{a_H}{p} + r_{Catalyst}\right)^2\left[3kT\left(\frac{1}{m_H} + \frac{1}{m_C}\right)\right]^{1/2}\frac{N_H N_C}{V}[2mp + m^2]\times 2.2\times 10^{-18}\text{ W} \tag{6.52}$$

A catalytic system that hinges on the coupling of three resonator cavities involves potassium. For example, the second ionization energy of potassium is 31.63 eV. This energy hole is obviously too high for resonant absorption. However, K^+ releases 4.34 eV when it is reduced to K. The combination of K^+ to K^{2+} and K^+ to K, then, has a net energy change of 27.28 eV. Consider the case of a gas phase catalytic shrinkage reaction of hydrogen or hydrino atoms by potassium ions as the electrocatalytic couple having an energy hole of 27.28 eV. The energy transfer efficiency is given by Eq. (6.37) where r is given by the average distance between cations in the reaction vessel. When the K^+ concentration is

$$3\times10^{22}\frac{K^+}{m^3},$$

r is approximately 5×10^{-9} m. For $J=1$, $\Phi_D=1$, $\kappa^2=1$, $\tau_D=10^{-13}$ sec (based on the vibrational frequency of KH^+), and m=1 in Eq. (5.8), the energy transfer efficiency, E_c, is approximately 0.001. The power for the reaction given by Eqs. (5.13), (5.14), and (5.8) with the substitution of

$$E=0.001,\ p=1,\ m=1,\ V=1\ m^3,\ N_H=3\times10^{22},\ N_c=3\times10^{21},\ m_c=6.5\times10^{-26}\ kg,\ r_c=1.38\times10^{-10}\ m,\ T=675\ K \tag{6.53}$$

into Eq. (6.52) is

$$P_{m,p}=300\ MW\ (300\ W/cm^3) \tag{6.54}$$

Gas Discharge Energy Reactor

A gas discharge energy reactor comprises a hydrogen isotope gas filled glow discharge vacuum chamber **300** of FIG. **8** including an ozonizer-type capacitor, a hydrogen source **322** which supplies hydrogen to the chamber **300** through control valve **325**, and a voltage and current source **330** to cause current to pass between a cathode **305** and an anode **320**. In one embodiment comprising an ozonizer-type capacitor gas discharge cell, one of the electrodes can be shielded by a dielectric barrier such as glass or a ceramic moiety. In a preferred embodiment, the cathode further comprises a source of energy holes of approximately mX27.21 eV to cause atomic hydrogen "shrinkage" and/or approximately mX48.6 eV to cause molecular hydrogen "shrinkage" where m is an integer (including the electrocatalytic ions and couples described in my previous U.S. patent applications entitled "Energy/Matter Conversion Methods and Structures", Ser. No. 08/467,051 filed on Jun. 6, 1995 which is a continuation-in-part application of Ser. No. 08/416,040 filed on Apr. 3, 1995 which is a continuation-in-part application of Ser. No. 08/107,357 filed on Aug. 16, 1993, which is a continuation-in-part application of Ser. No. 08/075,102 (Dkt. 99437) filed on Jun. 11, 1993, which is a continuation-in-part application of Ser. No. 07/626,496 filed on Dec. 12, 1990 which is a continuation-in-part application of Ser. No. 07/345,628 filed Apr. 28, 1989 which is a continuation-in-part application of Ser. No. 07/341,733 filed Apr. 21, 1989 which are incorporated by reference). A preferred cathode **305** for shrinking hydrogen atoms is a palladium cathode whereby a resonant energy hole can be provided by the ionization of electrons from palladium to the discharge current. A second preferred cathode **305** for shrinking hydrogen atoms comprises a source of energy holes via electron transfer to the discharge current including at least one of beryllium, copper, platinum, zinc, and tellurium and a hydrogen dissociating means such as a source of electromagnetic radiation including UV light provided by photon source **350** or a hydrogen dissociating material including the transition elements and inner transition elements, iron, platinum, palladium, zirconium, vanadium, nickel, titanium, Sc, Cr, Mn, Co, Cu, Zn, Y, Nb, Mo, Tc, Ru, Rh, Ag, Cd, La, Hf, Ta, W, Re, Os, Ir, Au, Hg, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Vb, Lu, Th, Pa, U, activated charcoal (carbon), and intercalated Cs carbon (graphite). The reactor further comprises a means to control the energy dissipated in the discharge current when electrons are transferred from an electron donating species to provide an energy hole for hydrogen atoms (molecules) including pressure controller means **325** and current (voltage) source **330**. The gas discharge energy reactor further comprises a means **301** to remove the (molecular) lower-energy hydrogen such as a selective venting valve to prevent the exothermic shrinkage reaction from coming to equilibrium.

In another embodiment of the gas discharge energy reactor, the source of energy hole can be one of a inelastic photon or particle scattering reaction(s). In a preferred embodiment the photon source **350** supplies the energy holes where the energy hole corresponds to stimulated emission by the photon. In the case of a reactor which shrinks hydrogen atoms, the photon source **350** dissociates hydrogen molecules into hydrogen atoms. The photon source producing photons of at least one energy of approximately mX27.21 eV,

$$\frac{m}{2}\times27.21\text{ eV},$$

or 40.8 eV causes stimulated emission of energy as the hydrogen atoms undergo the shrinkage reaction. In another preferred embodiment, a photon source **350** producing photons of at least one energy of approximately mX48.6 eV, 95.7 eV, or mX31.94 eV causes stimulated emission of energy as the hydrogen molecules undergo the shrinkage reaction.

In another embodiment, a magnetic field can be applied by magnetic field generator **75** of FIG. **5** to produce a magnetized plasma of the gaseous ions which can be a nonlinear media. Coupling of resonator cavities and enhancement of the transfer of energy between them can be increased when the media is nonlinear. Thus, the reaction rate (transfer of the resonance shrinkage energy of the hydrogen atoms to the energy holes, the electrocatalytic ions or couples) can be increased and controlled by providing and adjusting the applied magnetic field strength.

In one embodiment of the method of operation of the gas discharge energy reactor, hydrogen from source **322** can be introduced inside of the chamber **300** through control valve **325**. A current source **330** causes current to pass between a cathode **305** and an anode **320**. The hydrogen contacts the cathode which comprises a source of energy holes of approximately mX27.21 eV to cause atomic hydrogen "shrinkage" and approximately mX48.6 eV to cause molecular hydrogen "shrinkage" where m is an integer. In a preferred embodiment, electrons are transferred from an electron donating species present on the cathode **305** to the discharge current to provide energy holes for hydrogen atoms (molecules). In the case of a reactor which shrinks hydrogen atoms, the molecular hydrogen can be dissociated into atomic hydrogen by a dissociating material on the cathode **305** or by a source of electromagnetic radiation including UV light provided by photon source **350** such that the dissociated hydrogen atoms contact a source of energy holes including a molten, liquid, gaseous, or solid source of the energy holes. The atomic (molecular) hydrogen releases energy as its electrons are stimulated to undergo transitions to lower energy levels by the energy holes. The energy dissipated in the discharge current when electrons are transferred from an electron donating species can be controlled to provide an energy hole equal to the resonance shrinkage energy for hydrogen atoms (molecules) by controlling the gas pressure from source **322** with pressure controller means **325** and the voltage with the current (voltage) source **330**. The heat output can be monitored with thermocouples present in at least the cathode **305**, the anode **320**, and the heat exchanger **60** of FIG. **5**. The output power can be controlled by a computerized monitoring and control system which monitors the thermistors and controls the means to alter the power output. The (molecular) lower-energy hydrogen can be removed by a means **301** to prevent the exothermic shrinkage reaction from coming to equilibrium.

In another embodiment of the gas discharge energy reactor, a preferred cathode **305** comprises the catalytic material including a spillover catalyst described in the Pressurized Gas Energy Reactor Section.

Another embodiment of the gas discharge energy reactor comprises a gaseous source of energy holes wherein the shrinkage reaction occurs in the gas phase, and the gaseous hydrogen atoms are provided by a discharge of molecular hydrogen gas. In a further embodiment the gaseous source of energy holes can be provided by a discharge current which produces the gaseous source of energy holes (electrocatalytic ion or couple) such as a discharge in potassium metal to form K^+/K^+, rubidium metal to form Rb^+, or titanium metal to form Ti^{2+}. The embodiment comprises a hydrogen isotope gas filled glow discharge chamber **300**. The glow discharge cell can be operated at an elevated temperature such this the source of energy holes (electrocatalytic ion or couple) can be sublimed, boiled, or volatilized into the gas phase. In an embodiment, the counterion of the source of energy holes (electrocatalytic ion or couple) can be the hydride anion (H^-) such as rubidium hydride (Rb^+ electrocatalytic ion) and/or potassium hydride (K^+/K^+ electrocatalytic couple).

In an embodiment, the source of energy holes can be an electrocatalytic ion or electrocatalytic couple comprising cation-anion pairs in the gas phase wherein the cation-anion pairs are dissociated by external source means **75** of FIG. **5** which includes, for example, a particle source **75*b*** and/or photon source **75*a*** and/or a source of heat, acoustic energy, electric fields, or magnetic fields. In a preferred embodiment, the cation-anion pairs are thermally dissociated by heat source **75** of FIG. **5** or photodissociated by photon source **350** of FIG. **8**.

Refrigeration Means

A further embodiment of the present invention comprises a refrigeration means which comprises the electrolytic cell of FIG. **6**, the pressurized hydrogen gas cell of FIG. **7**, and the hydrogen gas discharge cell of FIG. **8** of the present invention wherein a source of lower-energy atomic (molecular) hydrogen is supplied rather than a source of normal hydrogen. The lower-energy hydrogen atoms are reacted to a higher energy state with the absorption of heat energy according to the reverse of the catalytic shrinkage reaction such as those given by Eqs. (4–6); (7–9); (10–12); (13–15); (16–18); (48–50); (51–53); (54–56); (57–59); (60–62), (63–65), (66–68), (69–71), (72–74), and (75–77). The lower-energy hydrogen molecules are reacted to a higher energy state with the absorption of heat energy according to the reverse of the catalytic shrinkage reaction such as that given by Eqs. (78–80); (81–83); (84–86); (88–90), and (91–93). In this embodiment, means **101**, **201** and **301** of FIGS. **6**, **7**, and **8**, respectively, serve to remove the normal hydrogen such as a selective venting valves to prevent the endothermic reaction from coming to equilibrium.

Compositions of Matter Comprising at Least Lower-Energy Hydrogen Atom(s) and/or Lower-energy Hydrogen Molecule(s)

The present invention further comprises molecules containing lower-energy hydrogen atoms. Lower-energy hydrogen can be reacted with any atom of the periodic chart or known organic or inorganic molecule or compound or metal, nonmetal, or semiconductor to form an organic or inorganic molecule or compound or metal, nonmetal, or semiconductor containing lower-energy hydrogen atoms and molecules. The reactants with lower-energy hydrogen include neutral atoms, negatively or positively charged atomic and molecular ions, and free radicals. For example, lower-energy hydrogen can be reacted with water or oxygen to form a molecule containing lower-energy hydrogen and oxygen, and lower-energy hydrogen can be reacted with singly ionized helium to form a molecule containing helium and lower-energy hydrogen. Lower-energy hydrogen can be also reacted with metals. In one embodiment of the electrolytic cell energy reactor, lower-energy hydrogen produced during operation at the cathode can be incorporated into the cathode by reacting with it; thus, a metal-lower-energy hydrogen material can be produced. In all such reactions, the reaction rate and product yield are increased by applying heat, and/or pressure.

Lower-energy hydrogen molecules (dihydrinos) are purified from hydrogen gas by combustion of the normal hydrogen. Oxygen can be mixed with the sample to be purified, and the sample can be ignited. In a second embodiment of the method of dihydrino purification, the sample can be flowed over a hydrogen recombiner which reacts with the normal hydrogen in the gas stream to form water. In a third embodiment, lower-energy hydrogen molecules (dihydrinos) are collected in a cathode of an electrolytic energy reactor of the present invention such as a metal cathode including a nickel cathode or a carbon cathode. The cathode can be heated in a vessel to a first temperature which causes normal hydrogen to preferentially off gas by external heating or by flowing a current through the cathode. The normal hydrogen can be pumped off, then the cathode can be heated to a second higher temperature at which dihydrino gas can be released and collected. In a fourth embodiment, the gas sample is purified by cryofiltration including gas chromatography at low temperature such as gas chromatography with an activated carbon (charcoal) column at liquid nitrogen temperature and with a column which will separate para from ortho hydrogen such as an Rt-Alumina column, or a HayeSep column at liquid nitrogen temperature wherein normal hydrogen can be retained to a greater extent than dihydrino. In a fifth embodiment, the gas sample is purified by cryodistillation wherein normal hydrogen can be liquefied and separated from gaseous lower-energy hydrogen (dihydrino). The dihydrino can be concentrated by liquefaction in liquid helium.

EXPERIMENTAL VERIFICATION OF THE PRESENT THEORY

EXAMPLE 1

The article by Mills and Good [Mills, R., Good, W., "Fractional Quantum Energy Levels of Hydrogen", Fusion Technology, Vol. 28, No. 4, November, (1995), pp. 1697–1719] describes the determination of excess heat release during the electrolysis of aqueous potassium carbonate by the very accurate and reliable method of heat measurement, flow calorimetry; describes the experimental identification of hydrogen atoms in fractional quantum energy levels—hydrinos—by X-ray Photoelectron Spectroscopy (XPS); describes the experimental identification of hydrogen atoms in fractional quantum energy levels—hydrinos—by emissions of soft x-rays from dark matter; describes the experimental identification of hydrogen molecules in fractional quantum energy levels—dihydrino molecules by high resolution magnetic sector mass spectroscopy with ionization energy determination, and gives a summary.

In Summary

The complete theory which predicts fractional quantum energy levels of hydrogen and the exothermic reaction whereby lower-energy hydrogen is produced is given elsewhere [Mills, R., *The Grand Unified Theory of Classical Quantum Mechanics*, (1995), Technomic Publishing Company, Lancaster, Pa., provided by HydroCatalysis Power Corporation, Great Valley Corporate Center, 41 Great Valley Parkway, Malvern, Pa., 19355, R. Mills; *Unification of Spacetime, the Forces, Matter, and Energy* (Technomic Publishing Company, Lancaster, Pa., 1992)].

Excess power and heat were observed during the electrolysis of aqueous potassium carbonate. Flow calorimetry of pulsed current electrolysis of aqueous potassium carbonate at a nickel cathode was performed in a single-cell dewar. The average power out of 24.6 watts exceeded the average input power (voltage times current) of 4.73 watts by a factor greater than 5. The total input energy (integration of voltage times current) over the entire duration of the experiment was 5.72 MJ; whereas, the total output energy was 29.8 MJ. No excess heat was observed when the electrolyte was changed from potassium carbonate to sodium carbonate. The source of heat is assigned to the electrocatalytic, exothermic reaction whereby the electrons of hydrogen atoms are induced to undergo transitions to quantized energy levels below the conventional "ground state". These lower energy states correspond to fractional quantum numbers: $n=\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$. Transitions to these lower energy states are stimulated in the presence of pairs of potassium ions (K^+/K^+ electrocatalytic couple) which provide 27.2 eV energy sinks.

The identification of the $n=\frac{1}{2}$ hydrogen atom , $H(n=\frac{1}{2})$ is reported. Samples of the nickel cathodes of aqueous potassium carbonate electrolytic cells and aqueous sodium carbonate electrolytic cells were analyzed by XPS. A broad peak centered at 54.6 eV was present only in the cases of the potassium carbonate cells. The binding energy (in vacuum) of $H(n=\frac{1}{2})$ is 54.4 eV. Thus, the theoretical and measured binding energies for $H(n=\frac{1}{2})$ are in excellent agreement.

Further experimental identification of hydrinos—down to $H(n=\frac{1}{8})$—can be found in the alternative explanation by Mills et al. for the soft X-ray emissions of the dark interstellar medium observed by Labov and Bowyer [S. Labov and S. Bowyer, Astrophysical Journal, 371 (1991) 810] of the Extreme UV Center of the University of California, Berkeley. The agreement between the experimental spectrum and the energy values predicted for the proposed transitions is remarkable.

The reaction product of two $H(n=\frac{1}{2})$ atoms, the dihydrino molecule, was identified by mass spectroscopy (Shrader Analytical & Consulting Laboratories). The mass spectrum of the cryofiltered gases evolved during the electrolysis of a light water K_2CO_3 electrolyte with a nickel cathode demonstrated that the dihydrino molecule, $H_2(n=\frac{1}{2})$, has a higher ionization energy, about 63 eV, than normal molecular hydrogen, $H_2(n=1)$, 15.46 eV. The high resolution (0.001 AMU) magnetic sector mass spectroscopic analysis of the postcombustion gases indicated the presence of two peaks of nominal mass two at 70 eV and one peak at 25 eV. The same analysis of molecular hydrogen indicates only one peak at 25 eV and one peak at 70 eV. In the case of the postcombustion sample at 70 eV, one peak was assigned as the hydrogen molecular ion peak, $H_2^+(n=1)$, and one peak was assigned as the dihydrino molecular peak, $H_2^+(n=\frac{1}{2})$ which has a slightly larger magnetic moment.

EXAMPLE 2

In the January 1994 edition of Fusion Technology, [Mills, R., Good, W., Shaubach, R., "Dihydrino Molecule Identification", Fusion Technology, 25, 103 (1994)] Mills et al. review and present three sets of data of heat production and "ash" identification including the work of HydroCatalysis Power Corporation (Experiments #1–#3) and Thermacore, Inc. (Experiments #4–#14).

In Summary

Mills et al. report the experimental evidence supporting the Mills theory that an exothermic reaction occurs wherein the electrons of hydrogen atoms and deuterium atoms are stimulated to relax to quantized potential energy levels below that of the "ground state" via electrochemical reactants K^+ and K^+; Pd^{2+} and Li^+, or Pd and O_2 of redox energy resonant with the energy hole which stimulates this transition. Calorimetry of pulsed current and continuous electrolysis of aqueous potassium carbonate (K^+/K^+ electrocatalytic couple) at a nickel cathode was performed. The excess power out of 41 watts exceeded the total input power given by the product of the electrolysis voltage and current by a factor greater than 8. The "ash" of the exothermic reaction is atoms having electrons of energy below the "ground state" which are predicted to form molecules. The predicted molecules were identified by lack of reactivity with oxygen, by separation from molecular deuterium by cryofiltration, and by mass spectroscopic analysis.

The combustion of the gases evolved during the electrolysis of a light water K_2CO_3 electrolyte (K^+/K^+ electrocatalytic couple) with a nickel cathode was incomplete. The mass spectroscopic analysis (Air Products & Chemicals, Inc.) of uncombusted gases demonstrated that the species predominantly giving rise to the m/e=2 peak must have a different m/e=1 to m/e=2 production efficiency than hydrogen. And, the further mass spectroscopic analysis of the m/e=2 peak of the uncombusted gas demonstrated that the dihydrino molecule, $H_2(n=\frac{1}{2})$, has a higher ionization energy than H_2.

According to the analysis by Mills et al. of the raw data, Miles of the China Lake Naval Air Warfare Center Weapons Division observed the dideutrino molecule as a species with a mass to charge ratio of four and having a higher ionization potential than normal molecular deuterium. Miles was using mass spectroscopy to analyze the cryofiltered gases evolved from excess power producing electrolysis cells (palladium cathode and a $LiOD/D_2O$ electrolyte; an electrocatalytic couple of 27.54 eV). [B. F. BUSH, J. J. LAGOWSKI, M. H. MILES, and G. S. OSTROM, "Helium Production During the Electrolysis of D_2O in Cold Fusion Experiments", J. Electroanal. Chem., 304, 271 (1991); M. H. MILES, B. F. BUSH, G. S. OSTROM, and J. J. LAGOWSKI, "Heat and Helium Production in Cold Fusion Experiments", Proc. Conf. The Science of Cold Fusion, Como, Italy, Jun. 29–Jul. 4, 1991, p. 363, T. BRESSANI, E. DEL GIUDICE, and G. PREPARATA, Eds., SIF (1991); M. H. MILES, R. A. HOLLINS, B. F. BUSH, J. J. LAGOWSKI, and R. E. J. MILES, "Correlation of Excess Power and Helium Production During D_2O and H_2O Electrolysis Using Palladium Cathodes", J. Electroanal. Chem., 346, 99 (1993); M. H. MILES and B. F. BUSH, "Search for Anomalous Effects Involving Excess Power and Helium During D_2O Electrolysis Using Palladium Cathodes," Proc. 3rd Int. Conf. Cold Fusion, Nagoya, Japan, Oct. 21–25, 1992, p. 189].

Palladium sheets coated on one side with a hydrogen impermeant gold layer and coated on the other surface with an oxide coat (MnO_x, AlO_x, SiO_x) were deuterium or hydrogen loaded at NTT Laboratories. Heat was observed from light and heavy hydrogen only when the mixed oxide coat was present (Pd/O_2 electrocatalytic couple). The high resolution (0.001 AMU) quadrapole mass spectroscopic analysis of the gases released when a current was applied to a deuterium (99.9%) loaded MnO_x coated palladium sheet indicate the presence of a large shoulder on the D_2 peak which Mills et al. assign to the dideutrino molecule, $D_2(n=\frac{1}{2})$. [E. YAMAGUCHI and T. NISHIOKA, "Direct Evidence for Nuclear Fusion Reactions in Deuterated Palladium," Proc. 3rd Int. Conf. Cold Fusion,, Nagoya, Japan, October 21–25, 1992, p. 179; E. YAMAGUCHI and T. NISHIOKA, "Helium-4 Production from Deuterated Palladium at Low Energies," NTT Basic Research Laboratories and IMRA Europe S. A., Personal Communication (1992)].

EXAMPLE 3

Pennsylvania State University has determined excess heat release from flowing hydrogen in the presence of nickel oxide powder containing strontium niobium oxide (Nb^{3+}/Sr^{2+} electrocatalytic couple) by the very accurate and reliable method of heat measurement, thermopile conversion of heat into an electrical output signal [Phillips, J., "A Calorimetric Investigation of the Reaction of Hydrogen with Sample PSU #1", Sep. 11, 1994, A Confidential Report submitted to HydroCatalysis Power Corporation provided by HydroCatalysis Power Corporation, Great Valley Corporate Center, 41 Great Valley Parkway, Malvern, Pa. 19355]. Excess power and heat were observed with flowing hydrogen over the catalyst which increased with increasing flow rate. However, no excess power was observed with flowing helium over the catalyst/nickel oxide mixture or flowing hydrogen over nickel oxide alone. As shown in FIG. **9**, approximately 10 cc of nickel oxide powder containing strontium niobium oxide immediately produced 0.55 W of steady state output power at 523° K. When the gas was switched from hydrogen to helium, the power immediately dropped. The switch back to hydrogen restored the excess power output which continued to increase until the hydrogen source cylinder emptied at about the 40,000 second time point. With no hydrogen flow the output power fell to zero.

The source of heat is assigned to the electrocatalytic, exothermic reaction whereby the electrons of hydrogen atoms are induced to undergo transitions to quantized energy levels below the conventional "ground state". These lower energy states correspond to fractional quantum numbers: $n=\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$. Transitions to these lower energy states are stimulated in the presence of pairs of niobium and strontium ions (Nb^{3+}/Sr^{2+} electrocatalytic couple) which provide 27.2 eV energy sinks.

EXAMPLE 4

The article in the Spectral Data of Hydrinos from the Dark Interstellar Medium and from the Sun Section of Mills [Mills, R., *The Grand Unified Theory of Classical Quantum Mechanics*, (1995), Technomic Publishing Company, Lancaster, Pa.] describes the experimental identification of hydrogen atoms in fractional quantum energy levels—hydrinos—by emissions of soft X-rays from dark matter and the Sun; provides a resolution to the Solar Neutrino Problem, the Temperature of the Solar Corona Problem, the Broadening of the Hydrogen 911.8 Å Line Problem, the Temperature of the Transition from "Radiation Zone" to "Convection Zone" Problem, the Cool Carbon Monoxide Clouds Problem, the Stellar Age Problem, the Solar Rotation Problem, the Solar Flare Problem, and the problem of the ionizing energy source of hydrogen planets, and describes the experimental identification of hydrogen atoms in fractional quantum energy levels—hydrinos—by spin/nuclear hyperfine structure transition energies obtained by COBE for which no other satisfactory assignment exists.

In Summary

As shown in Table 1 Mills [Mills, R., *The Grand Unified Theory of Classical Quantum Mechanics*, (1995), Technomic Publishing Company, Lancaster, Pa.], hydrogen transitions to electronic energy levels below the "ground" state corresponding to fractional quantum numbers predicted by Mills' theory match the spectral lines of the extreme ultraviolet background of interstellar space. And, hydrogen disproportionation reactions yield ionized hydrogen, energetic electrons, and hydrogen ionizing radiation. This assignment resolves the paradox of the identity of dark matter and accounts for many celestial observations such as: diffuse Hα emission is ubiquitous throughout the Galaxy, and widespread sources of flux shortward of 912 Å are required [Labov, S., Bowyer, S., "Spectral observations of the extreme ultraviolet background", The Astrophysical Journal, 371, (1991), pp. 810–819].

Further experimental identification of hydrinos-down to $H(n=\frac{1}{8})$—can be found in the alternative explanation by Mills for the soft X-ray emissions of the dark interstellar medium observed by Labov and Bowyer [S. Labov and S. Bowyer, Astrophysical Journal, 371 (1991) 810] of the Extreme UV Center of the University of California, Berkeley. The agreement between the experimental spectrum and the energy values predicted for the proposed transitions is remarkable.

The paradox of the paucity of solar neutrinos to account for the solar energy output by the pp chain is resolved by assigning a major portion of the solar output to lower-energy hydrogen transitions. The photosphere of the Sun is 6000 K; whereas, the temperature of the corona based on the assignment of the emitted X-rays to highly ionized heavy elements is in excess of 10^6 K. No satisfactory power transfer mechanism is known which explains the excessive temperature of the corona relative to that of the photosphere. The paradox is resolved by the existence of a power source associated with the corona. The energy which maintains the corona at a temperature in excess of 10^6 K is that released by disproportionation reactions of lower-energy hydrogen as given by Eqs. (13–15). In Table 2 of Mills, the energy released by the transition of the hydrino atom with the initial lower-energy state quantum number p and radius

$$\frac{a_H}{p}$$

to the state with lower-energy state quantum number (p+m) and radius

$$\frac{a_H}{(p+m)}$$

catalyzed by a hydrino atom with the initial lower-energy state quantum number m', initial radius

$$\frac{a_H}{m'},$$

and final radius a_H are given in consecutive order of energy from the 1→½ H transition to the ⅑→⅒ H transition. The agreement between the calculated and the experimental values is remarkable. Furthermore, many of the lines of Table 2 had no previous assignment, or the assignment was unsatisfactory [Thomas, R. J., Neupert. W., M., Astrophysical Journal Supplement Series, Vol. 91, (1994), pp. 461–482; Malinovsky, M., Heroux, L., Astrophysical Journal, Vol. 181, (1973), pp. 1009–1030; Noyes, R., *The Sun, Our Star*, Harvard University Press, Cambridge, Ma., (1982), p. 172; Phillips, J. H., *Guide to the Sun*, Cambridge University Press, Cambridge, Great Britain, (1992), pp. 118–119; 120–121; 144–145]. The calculated power of 4×10^{26} W matches the observed power output of 4×10^{26} W.

The broadening of the solar HI911.8 Å line (911.8 Å to ≈600 Å) is six times that predicted based on the thermal electron energy at the surface of the photosphere (T=6,000 K) where the HI 911.8 Å continuum originates, and based on the relative width of the helium continuum lines, He I 504.3 Å (He I 504.3 Å to ≈530 Å) and He II 227.9 Å (He II 227.9 Å to ≈225 Å) [Thomas, R. J., Neupert, W., M., Astrophysical Journal Supplement Series, Vol. 91, (1994), pp. 461–482; Stix, M., *The Sun*, Springer-Verlag, Berlin, (1991), pp. 351–356; Malinovsky, M., Heroux, L.. Astrophysical Journal, Vol. 181, (1973), pp. 1009–1030; Noyes, R., *The Sun, Our Star*, Harvard University Press, Cambridge, Ma., (1982), p. 172; Phillips, J. H., *Guide to the Sun*, Cambridge University Press, Cambridge, Great Britain, (1992), pp. 118–119; 120–121; 144–145]. The latter lines are proportionally much narrower; yet, the corresponding temperatures of origin must be higher because the transitions are more energetic. Furthermore, the H 911.8 Å continuum line of the spectrum of a prominence is about one half the width of the same line of the quiet Sun spectrum. Yet, the temperature rises to greater than 10,000 K in a prominence. The problem of the anomalous spectral feature of the excessive broadening of the continuum line of hydrogen to higher energies can be resolved by assignment of the broadening mechanism to energetic disproportionation reactions involving hydrogen atoms as reactants.

The reaction product, lower-energy hydrogen, can be reionized as it is diffuses towards the center of the Sun. The abrupt change in the speed of sound and transition from "radiation zone" to "convection zone" at a radius of 0.7 the solar radius, 0.7 R_s, with a temperature of 2×10^6 K matches the ionization temperature of lower-energy hydrogen.

Another spectroscopic mystery concerns an infrared absorption band of the chromosphere at a wavelength of 4.7 μm which was previously assigned to carbon monoxide despite the implausibility of its existence in the observed region which has a temperature above that at which carbon monoxide would break up into its constituent carbon and oxygen atoms. This problem can be resolved by assignment of the broad 4.7 μm feature to a temperature broadened rotational transition of a molecular ion of lower-energy hydrogen. The assignment of the 4.7 μm absorption line to the J=0 to J=1 transition rotational transition of $H_2^*[2c'=3a_o]^+$ provides a resolution of the problem of cool carbon monoxide clouds.

Modeling how stars evolve leads to age estimates for some stars that are greater than the age of the universe. Mills' theory predicts that presently, stars exist which are older than the elapsed time of the present expansion as stellar evolution occurred during the contraction phase.

General Relativity provides a resolution to the problem of the loss of angular momentum of the core which is in agreement with the current Solar models and helioseismology data. The photon transfer of momentum to expanding spacetime mechanism provides a resolution to the solar rotation problem of the slowly rotating Solar core.

Further stellar evidence of disproportionation reactions is the emission of extreme ultraviolet radiation by young stars called A stars. They appear to have energetic, ultraviolet-emitting upper atmospheres, or coronas, even though astronomers believe such stars lack the ability to heat these regions.

Numerous late-type stars, particularly dM stars, are known to flare from time to time at visible and X-ray wavelengths. An extremely pronounced flare was observed by the Extreme Ultraviolet Explorer (EUVE) Deep Survey telescope on the star AU Microscopii at a count of 20 times greater than that at quiescence [Bowyer, S., Science, Vol. 263, (1994), pp. 55–59]. Emission lines in the extreme ultraviolet were observed for which there is no satisfactory assignment. These spectral lines match hydrogen transitions to electronic energy levels below the "ground" state corresponding to fractional quantum numbers as shown in Table 3 of Mills. The lines assigned to lower-energy hydrogen transitions increased significantly in intensity during the flare event. The data is consistent with disproportionation reactions of lower-energy hydrogen as the mechanism of solar flare activity.

Planetary evidence of disproportionation reactions is the emission of energy by Jupiter, Saturn, and Uranus in excess of that absorbed from the Sun. Jupiter is gigantic ball of gaseous hydrogen. Saturn and Uranus are also largely comprised of hydrogen. H_3^+ is detected from all three planets by infrared emission spectroscopy [J. Tennyson, Physics World, July, (1995), pp. 33–36]. Disproportionation reactions of hydrogen yield ionizing electrons, energy, and ionized hydrogen atoms. Ionizing electrons and protons can both react with molecular hydrogen to produce H_3^+.

The spin/nuclear hyperfine structure transition energies of lower-energy hydrogen match closely certain spectral lines obtained by COBE [E. L. Wright, et. al., The Astrophysical Journal, 381, (1991), pp. 200–209; J. C. Mather, et. al., The Astrophysical Journal, 420, (1994), pp. 439–444] for which no other satisfactory assignment exists.

EXAMPLE 5

Pennsylvania State University has determined excess heat release from flowing hydrogen in the presence of ionic hydrogen spillover catalytic material: 40% by weight potassium nitrate (KNO_3) on graphitic carbon powder with 5% by weight 1%-Pd-on-graphitic carbon (K^+/K^+ electrocatalytic couple) by the very accurate and reliable method of heat measurement, thermopile conversion of heat into an electrical output signal [Phillips, J., Shim, H., "Additional Calorimetric Examples of Anomalous Heat from Physical Mixtures of K/Carbon and Pd/Carbon", Jan. 1, 1996, A Confidential Report submitted to HydroCatalysis Power Corporation provided by HydroCatalysis Power Corporation, Great Valley Corporate Center, 41 Great Valley Parkway, Malvern, Pa. 19355]. Excess power and heat were observed with flowing hydrogen over the catalyst. However, no excess power was observed with flowing helium over the catalyst mixture. Rates of heat production were reproducibly observed which were higher than that expected from the conversion of all the hydrogen entering the cell to water, and the total energy observed was over four times larger than that expected if all the catalytic material in the cell were converted to the lowest energy state by "known" chemical reactions. Thus, "anomalous" heat, heat of a magnitude and duration which could not be explained by conventional chemistry, was reproducibly observed.

EXAMPLE 6

Excess heat from a pressurized gas energy cell having a gaseous source of energy holes has been observed by HydroCatalysis Power Corporation [manuscript in progress] with low pressure hydrogen in the presence of molybdenum iodide (MoI_2) (Mo^{2+} electrocatalytic ion) which was volatilized at the operating temperature of the cell, 210° C. The calorimeter was placed inside a large convection oven that maintained the ambient temperature of the cell at the operating temperature. The cell comprised a 40 cc stainless steel pressure vessel that was surrounded by a 2 inch thick molded ceramic thermal insulator. The cell was sealed with a vacuum tight flange that had a two hole Buffalo gland for a tungsten wire to dissociate molecular hydrogen, a perforation for a Type K thermocouple, a 1/16 inch inlet for hydrogen which was connected to a 1/4 inch stainless steel tube which connected to the hydrogen supply. The flange was sealed with a copper gasket. The bottom of the vessel had a 1/4" vacuum port connected to a stainless steel tube with a valve between the cell and a vacuum pump and vacuum gauge. Less than one gram of MoI_2 catalyst was placed in a ceramic boat inside the vessel. The vapor pressure of the catalyst was estimated to be about 50 millitorr at the operating temperature 210° C. The hydrogen pressure of about 200 to 250 millitorr was controlled manually by adjusting the supply through the inlet versus the amount pumped away at the outlet where the pressure was monitored in the outlet tube by the vacuum gauge. For each run, the total pressure was made (including the MoI_2 pressue in the case of the experimental run) precisely 250 millitorr.

The output power was determined by measuring difference between the cell temperature and the ambient oven temperature and comparing the result to a calibration curve generated by applying power to the inside of the cell with the tungsten filament. Excess power of 0.3 watts was observed from the 40 cc stainless steel reaction vessel containing less than 1 g of MoI_2 when hydrogen was flowed over the hot tungsten wire (≈2000° C.). However, no excess power was observed when helium was flowed over the hot tungsten wire or when hydrogen was flowed over the hot tungsten wire with no MoI_2 present in the cell. Rates of heat production were reproducibly observed which were higher than that expected from the conversion of all the hydrogen inside the cell to water, and the total energy observed was over 30 times larger than that expected if all the catalytic material in the cell were converted to the lowest energy state by "known" chemical reactions. Thus, "anomalous" heat, heat of a magnitude and duration which could not be explained by conventional chemistry, was reproducibly observed.

The gaseous contents of the reactor were monitored with a mass spectrometer. At the time that excess energy was produced corresponding to the case wherein hydrogen was flowed over the hot filament, a higher ionizing mass two species was observable; whereas, during the control run wherein hydrogen was flowed over the hot tungsten wire with no MoI_2 present in the cell, a higher ionizing mass two species was not observed. The higher ionizing mass two species is was assigned to the dihydrino molecule,

$$2\mathrm{H}\left[\frac{a_H}{p}\right] \rightarrow \mathrm{H}_2^*\left[2c' = \frac{\sqrt{2}\,a_o}{p}\right].$$

We claim:

1. A cell for extracting energy from hydrogen atoms comprising:

a reaction vessel containing

gaseous hydrogen atoms; and

a gaseous catalyst having a net enthalpy of reaction of about 27*(p/2) eV, where p is an integer greater than 1.

2. A cell according to claim **1**, wherein said gaseous catalyst comprises hydrogen atoms having a binding energy of about $E_b = 13.6/n^2$ eV, where n is a fraction whose numerator is **1** and denominator is an integer greater than 1.

3. A cell according to claim **1**, wherein said cell is constructed and arranged to be capable of maintaining the reaction,

$$m \times 27.21\ \mathrm{eV} + \mathrm{H}\left[\frac{a_H}{m'}\right] + \mathrm{H}\left[\frac{a_H}{p}\right] \rightarrow$$

$$\mathrm{H}^+ + e^- + \mathrm{H}\left[\frac{a_H}{(p+m)}\right] + [(p+m)^2 - p^2 - (m'^2 - 2m)] \times 13.6\ \mathrm{eV}$$

$$\mathrm{H}^+ + e^- \rightarrow \mathrm{H}\left[\frac{a_H}{1}\right] + 13.6\ \mathrm{eV}$$

$$\mathrm{H}\left[\frac{a_H}{m'}\right] + \mathrm{H}\left[\frac{a_H}{p}\right] \rightarrow$$

$$\mathrm{H}\left[\frac{a_H}{1}\right] + \mathrm{H}\left[\frac{a_H}{(p+m)}\right] + [2pm + m^2 - m'^2] \times 13.6\ \mathrm{eV} + 13.6\ \mathrm{eV}$$

where m and p are positive non-zero integers, m' is an integer greater than 1, and a_H is the radius of the hydrogen atom (n=1).

4. A cell according to claim **1**, further comprising a source of hydrogen atoms including molecules containing hydrogen atoms, and a second catalyst for disassociating said molecules to produce hydrogen atoms.

5. A cell according to claim **4**, wherein said second catalyst comprises at least one element selected from the group consisting of transition elements, inner transition elements, precious metals, refractory metals, lanthanides, actinides, and activated charcoal.

6. A cell according to claim **4**, wherein said second catalyst is at least one element selected from the group consisting of the refractory metals, activated charcoal, platinum, palladium, gold, rhenium and iridium.

7. A cell according to claim **4**, further comprising a valve for controlling the flow of said molecules over said second catalyst.

8. A cell according to claim **1**, wherein said gaseous catalyst is formed from a source of gaseous catalyst which is adapted to sublime, boil, or volatize when heated.

9. A cell according to claim **1**, wherein said gaseous catalyst is formed from a salt.

10. A cell according to claim **9**, wherein said salt is selected from the group consisting of halides, sulfates, phosphates, carbonates, hydroxides, and sulfides.

11. A cell according to claim **9**, wherein said gaseous catalyst is formed from a salt of rubidium or potassium.

12. A cell according to claim **11**, wherein said salt of potassium is selected from the group consisting of KF, KCl, KBr, KI, K_2S_2, KOH, K_2SO_4, K_2CO_3, K_2PO_4, and K_2GeF_4.

13. A cell according to claim **11**, wherein said salt of rubidium is selected from the group consisting of RbF, RbCl, RbBr, RbI, Rb_2S_2, RbOH, Rb_2SO_4, Rb_2CO_3, and Rb_3PO_4.

14. A cell according to claim **1**, wherein said gaseous catalyst comprises a cation having a vapor pressure greater than zero when said gaseous catalyst is heated, said cation being selected from the group consisting of (K^+), (Rb^+), (Mo^{2+}), and (Ti^{2+}).

15. A cell according to claim **1**, wherein said gaseous catalyst comprises a pair of cations having a vapor pressure greater than zero when said gaseous catalyst is heated, said pair of cations being selected from the group consisting of: (Sn^{4+}, Si^{4+}), (Pr^{3+}, Ca^{2+}), (Sr^{2+}, Cr^{2+}), (Cr^{3+}, Tb^{3+}), (Sb^{3+}, Co^{2+}), (Bi^{3+}, Ni^{2+}), (Pd^{2+}, In^+), (La^{3+}, Dy^{3+}), (La^{3+}, Ho^{3+}), (K^+, K^+), (V^{3+}, Pd^{2+}), (Lu^{3+}, Zn^{2+}), (As^{3+}, Ho^{3+}), (Mo^{5+}, Sn^{4+}), (Sb^{3+}, Cd^{2+}), (Ag^{2+}, Ag^+), (La^{3+}, Er^{3+}), (V^{4+}, B^{3+}), (Fe^{3+}, Ti^{3+}), (Co^{2+}, Ti^+), (Bi^{3+}, Zn^{2+}), (As^{3+}, Dy^{3+}), (Ho^{3+}, Mg^{2+}), (K^+, Rb^+), (Cr^{3+}, Pr^{3+}), (Sr^{2+}, Fe^{2+}), (Ni^{2+}, Cu^+), (Sr^{2+}, Mo^{2+}), (Y^{3+}, Zr^{4+}), (Cd^{2+}, Ba^{2+}), (Ho^{3+}, Pb^{2+}), (Pd^{2+}, Li^+), $Eu^{3+}, Mg^{2+})$, (Er^{3+}, Mg^{2+}), (Bi^{4+}, Al^{3+}), (Ca^{2+}, Sm^{3+}), (V^{3+}, La^{3+}), (Gd^{3+}, Cr^{2+}), (Mn^{2+}, Tl^+), (Yb^{3+}, Fe^{2+}), (Ni^{2+}, Ag^+), (Zn^{2+}, Yb^{2+}), (Se^{4+}, Sn^{4+}), (Sb^{3+}, Bi^{2+}), and (Eu^{3+}, Pb^{2+}).

16. A cell according to claim **1**, wherein said gaseous catalyst comprises an ionic compound resistant to reduction by hydrogen atoms.

17. A cell according to claim **1**, wherein said gaseous catalyst is adapted to provide ions.

18. A cell according to claim **1**, wherein said vessel includes temperature controlling structure capable of maintaining an atomic hydrogen partial pressure of less than about 1 torr.

19. A cell according to claim **1**, wherein said vessel includes temperature controlling structure capable of maintaining said catalyst in molten form.

20. A cell according to claim **1**, wherein said vessel includes temperature controlling structure capable of maintaining the temperature of said vessel at about 50° C. above the melting point of said gaseous catalyst.

21. A cell according to claim **1**, wherein said vessel is constructed and arranged to be capable of maintaining the hydrogen partial pressure in said vessel at about 200 millitorr.

22. A cell according to claim **15**, wherein said vessel includes temperature controlling structure capable of maintaining the temperature of said vessel at about 500° C. above the melting point of the compound having the highest melting point of a plurality of compounds which comprise a source of said gaseous catalyst.

23. A cell according to claim **1**, further comprising a valve for releasing said catalyst from said vessel.

24. A cell according to claim **1**, further comprising a valve for releasing said hydrogen atoms having a binding energy of about $E_b=13.6/n^2$ eV, where n is a fraction whose numerator is 1 and denominator is an integer greater than 1.

25. A cell according to claim **1**, wherein said vessel includes temperature controlling structure to control a temperature of said vessel.

26. A cell according to claim **1**, further comprising a heater for heating said vessel.

27. A cell according to claim **1**, further comprising a catalyst reservoir communicating with said reaction vessel, said catalyst reservoir containing said gaseous catalyst or a source of gaseous catalyst.

28. A cell according to claim **27**, further comprising a heater for heating said catalyst.

29. A cell according to claim **27**, wherein said catalyst reservoir is external to said reaction vessel.

30. A cell according to claim **29**, further comprising a flow control valve for controlling the flow of said catalyst from said catalyst reservoir into said vessel.

31. A cell according to claim **1**, further comprising a chamber for containing hydrogen atoms or a source of hydrogen atoms communicating with said reaction vessel.

32. A cell according to claim **31**, further comprising a flow control valve for controlling the flow of hydrogen atoms from said chamber into said vessel.

33. A cell according to claim **31** wherein said source of hydrogen atoms comprises an internal combustion engine.

34. A cell according to claim **31**, wherein said source of hydrogen atoms comprises a tungsten capillary constructed and arranged to be heated for dissociating molecules containing hydrogen atoms to produce said gaseous hydrogen atoms.

35. A cell according to claim **34**, further comprising a valve for controlling the flow of said molecules over said tungsten capillary.

36. A cell according to claim **31**, wherein said source of hydrogen atoms comprises an inductively coupled plasma flow tube for dissociating molecules containing hydrogen atoms to produce hydrogen atoms.

37. A cell according to claim **36**, further comprising a valve for controlling the flow of said molecules into said inductively coupled plasma flow tube.

38. A cell according to claim **36**, further comprising a power controller for controlling the power dissipated in said inductively coupled plasma flow tube.

39. A cell according to claim **1**, further comprising a heat exchanger for removing said extracted power from said cell.

40. A cell according to claim **1**, further comprising a power gauge for measuring the amount of extracted power in said cell.

41. A cell according to claim **1**, wherein said source of gaseous catalyst comprises an ionic compound which is resistant to hydrogen reduction and which is adapted to sublime, boil or become volatile when heated.

42. A cell according to claim **1**, wherein said source of gaseous catalyst comprises an ionic compound which is adapted to sublime, boil or become volatile when heated.

43. A cell according to claim **1**, wherein said source of gaseous catalyst comprises an ionic compound which is resistant to thermal degradation.

44. A cell according to claim **1**, wherein said gaseous catalyst is formed from at least one metal selected from the group consisting of Mo, Ti, and Rb.

45. A cell according to claim **1**, wherein said gaseous catalyst is formed from at least one salt selected from the group consisting of Mol_2, $TiCl_2$, $TiCl_4$, $SnCl_4$, $SiCl_4$, $PrBr_3$, $CaBr_2$, $SrCl_2$, Crl_2, Tbl_3, $SbCl_3$, CrF_3, $CoCl_2$, $BiCl_3$, $NiCl_2$, PdF_2, InCl, $LaCl_3$, $DyCl_3$, Lal_3, Hol_3, VF_3, PbF_2, VOCl, Pbl_2, $LuCl_3$, $PbCl_2$, Asl_3, Hol_3, $MoCl_5$, $SnCl_4$, Sbl_3, Cdl_2, AgF_2, AgF, Lal_3, Erl_3, VCl_4, BCl_3, $FeCl_3$, $TiCl_3$, Col_2, CoF_2, Tll, TlF, $BiBr_3$, $ZnBr_2$, Asl_3, Dyl_3, $HOCl_3$, $MgCl_2$, $CrCl_3$, $PrCl_3$, $SrCl_2$, $FeCl_2$, $NiCl_2$, CuCl, $SrCl_2$, $MoCl_2$, YCl_3, $ZrCl_4$, Cdl_2, Bal_2, Hol_3, Pbl_2, PdF_2, LiF, $EuCl_3$, $MgCl_2$, $ErCl_3$, $MgCl_2$, $ErCl_3$, $MgCl_2$, $BiCl_4$, $AlCl_3$, $CaBr_2$, $SmBr_3$, VaF_3, $LaCl_3$, Gdl_3, Crl_2, Mnl_2, $YbBr_3$, $FeBr_2$, $NiCl_2$, AgCl, $ZnCl_2$, $YbCl_2$, SeF_4, $SnCl_4$, SnF_4, Sbl_3, Bil_2, EuF_3, and $PbCl_2$.

46. A cell according to claim **1**, wherein said catalyst comprises potassium and has a net enthalpy of reaction of 27.28 eV.

47. A cell according to claim **1**, wherein said catalyst has a net enthalpy of reaction of about 27.2 eV.

48. A cell according to claim **1**, wherein said cell comprises a source of said gaseous catalyst combined with at least one of a hydrocarbon or water disposed such that said gaseous catalyst and said gaseous hydrogen atoms are capable being formed during combustion.

49. A cell according to claim **1**, further comprising a means for converting a source of catalyst to said gaseous catalyst.

50. A cell according to claim **49**, wherein said means for converting said source of catalyst to said gaseous catalyst comprises at least one of heat, electron-beam energy, photon energy, acoustic energy, electric field, or magnetic field.

51. A cell according to claim **1**, further comprising a filament coated with a source of gaseous catalyst.

52. A cell according to claim **1**, further comprising a source of gaseous hydrogen atoms is selected from the group consisting of hydrogen gas, water, hydrides, metal-hydrogen solutions, and hydrocarbons.

53. A cell according to claim **52**, further comprising means to disassociate water to form gaseous hydrogen atoms.

54. A cell according to claim **1**, further comprising a hot filament and a hydrogen containing gas stream.

55. A cell according to claim **1**, further comprising a hot grid and a hydrogen containing gas stream.

56. A cell according to claim **1**, further comprising a heated tungsten capillary and a hydrogen containing gas stream.

57. A cell according to claim **1**, further comprising a hydride maintained under nonequilibrium conditions.

58. A cell according to claim **1**, further comprising an inductively coupled plasma flow tube and a hydrogen gas containing stream.

59. A cell according to claim **1**, further comprising means to reform hydrocarbons to at least one of gaseous molecular and gaseous atomic hydrogen.

60. A cell according to claim **4**, wherein said second catalyst is treated with an aqueous solution of K_2CO_3 and H_2O_2.

61. A cell according to claim **4**, further comprising a temperature controlling structure capable of maintaining a selected atomic hydrogen partial pressure by controlling the temperature of said second catalyst.

62. A cell according to claim **61**, wherein said temperature controlling structure is at least one selected from the group consisting of an internal heater, an external heater, the catalysis of hydrogen, and a heat exchanger which removes energy from the cell.

63. A cell according to claim **1**, further comprising a source of UV light for disassociating hydrogen containing molecules to form said gaseous hydrogen atoms.

64. A cell according to claim **1**, further comprising a means for pyrolysis of hydrocarbons or water to form said gaseous hydrogen atoms.

65. A cell according to claim **4**, wherein said second catalyst comprises at least one selected from the group consisting of an element, compound, alloy or mixture of transition elements, inner transition elements, iron, platinum, palladium, zirconium, vanadium, nickel, titanium, Sc, Cr, Mn, Co, Cu, Zn, Y, Nb, Mo, Tc, Ru, Rh, Ag, Cd, La, Hf, Ta, W, Re, Os, Ir, Au, Hg, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Vb, Lu, Th, Pa, U, activated charcoal, and intercalated Cs carbon.

66. A cell according to claim **4**, further comprising a heater to heat said second catalyst.

67. A cell according to claim **4**, further comprising a filament or grid constructed and arranged to dissociate hydrogen and to heat said second catalyst.

68. A cell according to claim **4**, further comprising means for controlling the power output of said cell.

69. A cell according to claim **68**, wherein said means for controlling the power output of said cell comprises means for controlling the temperature of said second catalyst.

70. A cell according to claim **69**, wherein said means for controlling the temperature of said second catalyst comprises a filament or grid.

71. A cell according to claim **31**, further comprising a flow control means for controlling the flow of a source of gaseous hydrogen atoms or said gaseous hydrogen atoms from said chamber to said reaction vessel.

72. A cell according to claim **1**, further comprising a flow control means for controlling the flow of hydrogen from said reaction vessel.

73. A cell according to claim **71**, wherein said flow control means comprises a valve.

74. A cell according to claim **1**, further comprising a vacuum pump constructed and arranged for controlling the flow of hydrogen from said reaction vessel.

75. A cell according to claim **27**, further comprising a flow control means for controlling the flow of catalyst from said catalyst reservoir to said reaction vessel.

76. A cell according claim **1**, further comprising means for controlling the flow of said gaseous catalyst from said reaction vessel.

77. A cell according to claim **76**, wherein said means for controlling the flow of said gaseous catalyst from said reaction vessel comprises a valve.

78. A cell according to claim **1**, further comprising a vacuum pump constructed and arranged for controlling the flow of said gaseous catalyst from said reaction vessel.

79. A cell according to claim **1**, further comprising a nonreactive gas.

80. A cell according to claim **1**, further comprising a means for controlling the amount of a nonreactive gas in said reaction vessel.

81. A cell according to claim **1**, further comprising flow control means for controlling the flow of a nonreactive gas into said reaction vessel.

82. A cell according to claim **81**, wherein said means for controlling the amount of nonreactive gas in said vessel comprises a valve constructed and arranged to regulate the flow of said nonreactive gas into said reaction vessel.

83. A cell according to claim **1**, further comprising at least one of a valve or pump constructed and arranged for controlling the flow of a nonreactive gas from said reaction vessel.

84. A cell according to claim **1**, further comprising a pump in communication with said reaction vessel.

85. A cell according to claim **1**, further comprising structure for controlling the vapor pressure of said catalyst.

86. A cell according to claim **85**, wherein said structure for controlling the vapor pressure of said catalyst comprises a heater constructed and arranged to control the temperature of said reaction vessel.

87. A cell according to claim **1**, further comprising structure for maintaining a selected vapor pressure of said gaseous catalyst or source of said gaseous catalyst.

88. A cell according to claim **1**, further comprising structure for maintaining a selected vapor pressure of said gaseous catalyst or source of gaseous catalyst.

89. A cell according to claim **88**, wherein said structure for maintaining a selected vapor pressure of said gaseous catalyst or source of gaseous catalyst comprises a valve constructed and arranged for controlling the flow of said source of gaseous catalyst or gaseous catalyst from said catalyst reservoir and a valve constructed and arranged for controlling the flow of said source of gaseous catalyst or gaseous catalyst from said reaction vessel.

90. A cell according to claim **88**, wherein said structure for maintaining a selected vapor pressure of said source of gaseous catalyst or said gaseous catalyst comprises a valve constructed and arranged for controlling the flow of said source of gaseous catalyst or gaseous catalyst from said reaction vessel.

91. A cell according to claim **1**, further comprising a valve for releasing hydrogen atoms having a binding energy of about E_b=13.6/n^2 eV, where n is a fraction whose numerator is 1 and denominator is an integer greater than 1 or a compound containing said hydrogen atoms from said reaction vessel.

92. A cell according to claim **1**, further comprising means for adsorbing energy released from said hydrogen atom.

93. A cell according to claim **1**, wherein said vessel comprises an internal combustion chamber.

94. A cell according to claim **93**, wherein said internal combustion chamber is an engine cylinder.

95. A cell according to claim **1**, further comprising means for controlling the power output of said cell.

96. A cell according to claim **95**, wherein said means for controlling the power output of said cell comprises means for controlling the amount of said gaseous catalyst.

97. A cell according to claim **96**, wherein said means for controlling the amount of gaseous catalyst comprises means for controlling the temperature of said vessel and said gaseous catalyst is selected to have a vapor pressure dependent upon the temperature of said reaction vessel.

98. A cell according to claim **95**, wherein said means for controlling the power output of said cell comprises means for controlling the flow of a source of gaseous catalyst or gaseous catalyst from said reaction vessel.

99. A cell according to claim **95**, wherein said means for controlling the power output of said cell comprises means for controlling the temperature of a source of catalyst.

100. A cell according to claim **95**, wherein said means for controlling the power output of said cell comprises means for controlling the amount of said gaseous hydrogen atoms or a source of gaseous hydrogen atoms in said vessel.

101. A cell according to claim **95**, wherein said means for controlling the power output of said cell comprises means for controlling the flow of gaseous hydrogen atoms or source of said hydrogen atoms into said reaction vessel.

102. A cell according to claim **95**, wherein said means for controlling the power output of said cell comprises means for controlling the flow of gaseous hydrogen atoms or source of said hydrogen atoms from said reaction vessel.

103. A cell according to claim **95**, wherein said means for controlling the power output of said cell comprises controlling the amount of said nonreactive gas present in said reaction vessel.

104. A cell according to claim **103**, wherein said means for controlling the amount of nonreactive gas comprises means for controlling the flow of said nonreactive gas into said reaction vessel.

105. A cell according to claim **103**, wherein said means for controlling the amount of nonreactive gas comprises means for controlling the flow of said nonreactive gas from said reaction vessel.

106. A cell according to claim **95**, wherein said means for controlling the power output of said cell comprises means for controlling the flow of a hydrogen containing gas over at least one of a hot filament, a tungsten capillary heated by electron bombardment, or an inductively coupled plasma flow.

107. A cell according to claim **95**, wherein said means for controlling the power output of said cell comprises means for controlling the power dissipated in an inductively coupled plasma flow tube.

108. A cell according to claim **95**, wherein said means for controlling the power output of said cell comprises means for controlling the power dissipated in a hot filament, grid, or tungsten capillary heated by electron bombardment.

109. A cell according to claim **95**, wherein said means for controlling the power output of said cell comprises means for controlling the temperature of a hot filament, grid or tungsten capillary heated by electron bombardment over which a hydrogen containing gas flows.

110. A cell according to claim **95**, wherein said means for controlling the power output of said cell comprises means for controlling the temperature of a hydride maintained under nonequilibrium conditions.

111. A cell according to claim **27**, further comprising means for controlling the power output of said cell.

112. A cell according to claim **111**, wherein said means for controlling the power output of said cell comprises means for controlling the temperature of said catalyst reservoir and said gaseous catalyst is selected to have a vapor pressure dependent upon the temperature of said catalyst reservoir.

113. A cell according to claim **111**, wherein said means for controlling the power output of said cell comprises means for controlling the flow of said source of gaseous catalyst or gaseous catalyst from said catalyst reservoir into said reaction vessel.

114. A cell according to claim **1**, further comprising a boat contained within said reaction chamber, said boat containing a source of gaseous catalyst or said gaseous catalyst.

115. A cell according to claim **114**, further comprising a means for controlling the power of said cell.

116. A cell according to claim **114**, wherein said means for controlling the power output of said cell comprises means for controlling the temperature of said boat and said gaseous catalyst is selected to have a vapor pressure dependent upon the temperature of said boat.

117. A cell according to claim **27**, further comprising means for measuring the temperature of said catalyst reservoir.

118. A cell according to claim **114**, further comprising means for measuring the temperature of said boat.

119. A cell according to claim **27**, further comprising a means for measuring the temperature of said source of said gaseous catalyst contained in said catalyst reservoir.

120. A cell according to claim **114**, further comprising a means for measuring the temperature of said source of said gaseous catalyst contained in said boat.

121. A cell according to claim **1**, further comprising means for measuring the temperature of said reaction vessel.

122. A cell according to claim **1**, further comprising means for measuring the temperature of a source of said gaseous hydrogen atoms.

123. A cell according to claim **4**, further comprising means for measuring the temperature said second catalyst.

124. A cell according to claim **1**, further comprising means to measure the cell temperature.

125. A cell according to claim **124**, further comprising temperature controlling structure constructed and arranged to maintain a temperature in said reaction vessel greater than a temperature in said catalyst reservoir.

126. A cell according to claim **114**, further comprising temperature controlling structure constructed and arranged to maintain a temperature in said reaction vessel greater than a temperature in said boat.

127. A cell according to claim **1**, further comprising temperature controlling structure for maintaining a selected temperature of said reaction vessel.

128. A cell according to claim **1**, further comprising a nebulizer or atomizer.

129. A cell according to claim **1**, further comprising means to measure the pressure in said reaction vessel.

130. A cell according to claim **1**, further comprising means to measure the hydrogen pressure in said reaction vessel.

131. A cell according to claim **1**, further comprising means to measure the gaseous catalyst pressure in said reaction vessel.

132. A cell according to claim **1**, wherein said vessel is capable of containing a pressure within the range of 10^{-3} atmospheres to 100 atmospheres.

133. A cell according to claim **1**, further comprising a vacuum pump in communication with said vessel and said vessel being constructed and arranged to contain pressures less than atmospheric.

134. A cell according to claim **27**, further comprising a temperature controlling structure capable of maintaining a temperature in the said reaction vessel that is greater than a temperature of said catalyst reservoir.

135. A cell according to claim **27**, further comprising temperature controlling structure capable of maintaining said source of gaseous catalyst in a molten form.

136. A cell according to claim **114**, further comprising temperature controlling structure capable of maintaining said source of gaseous catalyst in a molten form.

137. A cell according to claim **27**, further comprising temperature controlling structure capable of maintaining the temperature of said catalyst reservoir at about 50° C. above the melting point of said gaseous catalyst.

138. A cell according to claim **1**, further comprising temperature controlling structure capable of maintaining the temperature of said reaction vessel at about 50° C. above the melting point of a source of gaseous catalyst.

139. A cell according to claim **114**, further comprising temperature controlling structure capable of maintaining the temperature of said boat at about 50° C. above the melting point of said gaseous catalyst.

140. A cell according to claim **15**, further comprising a catalyst reservoir in communication with said reaction vessel and temperature controlling structure capable of maintaining the temperature of said catalyst reservoir at about 50° C. above the melting point of the compound having the highest melting point of a plurality of compounds which comprise said source of gaseous catalyst.

141. A cell according to claim **15**, further comprising a boat contained in said reaction vessel and temperature controlling structure capable of maintaining the temperature of said boat at about 50° C. above the melting point of the compound having the highest melting point of a plurality of compounds which comprise said source of gaseous catalyst.

142. A cell according to claim **27**, further comprising temperature controlling structure capable of maintaining the temperature of said reaction vessel at about 50° C. above the melting point of said gaseous catalyst.

143. A cell according to claim **15**, further comprising a catalyst reservoir containing said source of gaseous catalyst and being in communication with said reaction vessel, and temperature controlling structure capable of maintaining the temperature of said reaction vessel at about 50° C. above the melting point of the compound having the highest melting point of a plurality of compounds which comprise said source of gaseous catalyst.

144. A cell according to claim **15**, further comprising a boat containing said source of gaseous catalyst and being disposed in said reaction vessel, and temperature controlling structure capable of maintaining the temperature of said reaction vessel at about 50° C. above the melting point of the compound having the highest melting point of a plurality of compounds which comprise said source of gaseous catalyst.

145. A method for extracting energy from hydrogen atoms comprising the steps of:

providing a gaseous catalyst having a net enthalpy of reaction of about 27*(p/2) eV, where p is an integer greater than 1;

providing gaseous hydrogen atoms; and

reacting said gaseous catalyst with said gaseous hydrogen atoms, thereby releasing energy from said gaseous hydrogen atoms.

146. A method according to claim **145**, wherein said gaseous hydrogen atoms are provided by disassociating molecules containing hydrogen atoms.

147. A method according to claim **145**, wherein gaseous hydrogen atoms are

provided by contacting molecules containing hydrogen atoms with a second catalyst for disassociating said molecules to produce hydrogen atoms in the gas phase.

148. A method according to claim **145**, wherein said gaseous catalyst comprises gaseous hydrogen atoms having a binding energy of about E_b=13.6/n^2 eV, where n is a fraction whose numerator is 1 and denominator is an integer greater than 1.

149. A method according to claim **145**, wherein said gaseous catalyst is provided according to the reaction,

$$m \times 27.21\text{ eV} + \mathrm{H}\left[\frac{a_H}{m'}\right] + \mathrm{H}\left[\frac{a_H}{p}\right] \to$$

$$\mathrm{H}^+ + e^- + \mathrm{H}\left[\frac{a_H}{(p+m)}\right] + [(p+m)^2 - p^2 - (m'^2 - 2m)] \times 13.6\text{ eV}$$

$$\mathrm{H}^+ + e^- \to \mathrm{H}\left[\frac{a_H}{1}\right] + 13.6\text{ eV}$$

$$\mathrm{H}\left[\frac{a_H}{m'}\right] + \mathrm{H}\left[\frac{a_H}{p}\right] \to$$

$$\mathrm{H}\left[\frac{a_H}{1}\right] + \mathrm{H}\left[\frac{a_H}{(p+m)}\right] + [2pm + m^2 - m'^2] \times 13.6\text{ eV} + 13.6\text{ eV}$$

where m and p are positive non-zero integers, m' is an integer greater than 1, and a_H is the radius of the hydrogen atom (n=1).

150. A method according to claim **145**, wherein said gaseous catalyst is provided by volatizing a material to a gaseous state and ionizing said gaseous material.

151. A method according to claim **145**, wherein said hydrogen atoms are provided by flowing gaseous molecules containing hydrogen atoms over a hot refractory metal, transition metal, platinum, palladium, gold, rhenium, or iridium.

152. A method according to claim **145**, wherein said hydrogen atoms are provided by flowing gaseous molecules containing hydrogen atoms over a tungsten capillary heated by electron bombardment to between 1800 and 2000 K.

153. A method according to claim **145**, wherein said hydrogen atoms are provided by flowing gaseous molecules containing hydrogen atoms in an inductively coupled plasma flow tube.

154. A method according to claim **145**, wherein said reaction occurs at a pressure less than about one torr.

155. A method according to claim **145**, wherein the partial pressure of hydrogen atoms in the reaction is less than about one torr.

156. A method according to claim **145**, wherein the partial pressure of hydrogen atoms in the reaction is about 200 millitorr.

157. A method according to claim **145**, wherein said reaction occurs at a temperature of about 50° C. above the melting point of said gaseous catalyst.

158. A method according to claim **145**, wherein the vapor partial pressure of said gaseous catalyst is controlled by varying the temperature of said gaseous catalyst.

159. A method according to claim **145**, further comprising a step of isolating produced hydrogen atoms having a binding energy of about $E_b=13.6/n^2$ eV, where n is a fraction whose numerator is 1 and denominator is an integer greater than 1.

160. A method according to claim **145**, further comprising the step of forming said gaseous hydrogen atoms from a source of gaseous catalyst comprising an ionic compound which is resistant to hydrogen reduction.

161. A method according to claim **145**, further comprising the step of forming said gaseous hydrogen atoms from a source of gaseous catalyst comprising an ionic compound which is resistant to hydrogen reduction and which is adapted to sublime, boil or become volatile when heated.

162. A method according to claim **145**, further comprising the step of forming said gaseous compound from a source of gaseous catalyst comprising an ionic compound which is adapted to sublime, boil or become volatile when heated.

163. A method according to claim **145**, further comprising the step of forming said gaseous compound from a source of gaseous catalyst comprising an ionic compound which is resistant to thermal degradation.

164. A method according to claim **145**, further comprising the step of forming said gaseous hydrogen atoms from a source of gaseous catalyst comprising a salt of rubidium or potassium.

165. A method according to claim **145**, further comprising the step of forming said gaseous hydrogen atoms from a source of gaseous catalyst comprising a salt that can form a vapor comprising ions when heated.

166. A method according to claim **145**, further comprising the step of forming said gaseous hydrogen atoms from a source of gaseous catalyst comprising a salt of rubidium selected from the group consisting of RbF, RbCl, RbBr, RbI, Rb_2S_2, RbOH, Rb_2SO_4, Rb_2CO_3, and Rb_3PO_4.

167. A method according to claim **145**, further comprising the step of forming said gaseous hydrogen atoms from a source of gaseous catalyst comprising a salt of potassium selected from the group consisting of KF, KCl, KBr, KI, K_2S_2, KOH, K_2SO_4, K_2CO_3, K_3PO_4, and K_2GeF_4.

168. A method according to claim **145**, further comprising the step of forming said gaseous hydrogen atoms from a source of gaseous catalyst comprising at least one metal selected from the group consisting of Mo, Ti, and Rb.

169. A method according to claim **145**, further comprising the step of forming said gaseous hydrogen atoms from a source of gaseous catalyst comprising at least one salt selected from the group consisting of MoI_2, $TiCl_2$, $TiCl_4$, $SnCl_4$, $SiCl_4$, $PrBr_3$, $CaBr_2$, $SrCl_2$, CrI_2, TbI_3, $SbCl_3$, CrF_3, $CoCl_2$, $BiCl_3$, $NiCl_2$, PdF_2, InCl, $LaCl_3$, $DyCl_3$, LaI_3, HoI_3, KNO_3, VF_3, PbF_2, VOCl, PbI_2, $LuCl_3$, $PbCl_2$, AsI_3, HoI_3, $MoCl_5$, $SnCl_4$, SbI_3, CdI_2, AgF_2, AgF, LaI_3, ErI_3, VCl_4, BCl_3, $FeCl_3$, $TiCl_3$, CoI_2, CoF_2, TlI, TlF, $BiBr_3$, $ZnBr_2$, AsI_3, DyI_3, $HoCl_3$, $MgCl_2$, $CrCl_3$, $PrCl_3$, $SrCl_2$, $FeCl_2$, $NiCl_2$, CuCl, $SrCl_2$, $MoCl_2$, YCl_3, $ZrCl_4$, CdI_2, BaI_2, HoI_3, PbI_2, PdF_2, LiF, $EuCl_3$, $MgCl_2$, $ErCl_3$, $MgCl_2$, $ErCl_3$, $MgCl_2$, $BiCl_4$, $AlCl_3$, $CaBr_2$, $SmBr_3$, VaF_3, $LaCl_3$, GdI_3, CrI_2, MnI_2, $YbBr_3$, $FeBr_2$, $NiCl_2$, AgCl, $ZnCl_2$, $YbCl_2$, SeF_4, $SnCl_4$, SnF_4, SbI_3, BiI_2, EuF_3, and $PbCl_2$.

170. A method according to claim **145**, wherein said gaseous catalyst comprises at least one ion selected from the group consisting of Mo^{2+}, Ti^{2+}, and Rb^+.

171. A method according to claim **145**, wherein said gaseous catalyst comprises at least one pair of ions selected from the group consisting of: (Sn^{4+}, Si^{4+}), (Pr^{3+}, Ca^{2+}), (Sr^{2+}, Cr^{2+}), (Cr^{3+}, Tb^{3+}), (Sb^{3+}, Co^{2+}), (Bi^{3+}, Ni^{2+}), (Pd^{2+}, In^+), (La^{3+}, Dy^{3+}), (La^{3+}, Ho^{3+}), (K^+, K^+), (V^{3+}, Pd^{2+}), (Lu^{3+}, Zn^{2+}), (As^{3+}, Ho^{3+}), (Mo^{5+}, Sn^{4+}), (Sb^{3+}, Cd^{2+}), (Ag^{2+}, Ag^+), (La^{3+}, Er^{3+}), (V^{4+}, B^{3+}), (Fe^{3+}, Ti^{3+}), (Co^{2+}, Ti^+), (Bi^{3+}, Zn^{2+}), (As^{3+}, Dy^{3+}), (Ho^{3+}, Mg^{2+}), (K^+, Rb^+), (Cr^{3+}, Pr^+), (Sr^{2+}, Fe^{2+}), (Ni^{2+}, Cu^+), (Sr^{2+}, Mo^{2+}), (Y^{3+}, Zr^{4+}), (Cd^{2+}, Ba^{2+}), (Ho^{3+}, Pb^{2+}), (Pb^{2+}, Li^+), (Eu^{3+}, Mg^{2+}), (Er^{3+}, Mg^{2+}), (Bi^{4+}, Al^{3+}), (Ca^{2+}, Sm^{3+}), (V^{3+}, La^{3+}), (Gd^{3+}, Cr^{2+}), (Mn^{2+}, Ti^+), (Yb^{3+}, Fe^{2+}), (Ni^{2+}, Ag^+), (Zn^{2+}, Yb^{2+}), (Se^{4+}, Sn^{4+}), (Sb^{3+}, Bi^{2+}), and (Eu^{3+}, Pb^{2+}).

172. A method according to claim **145**, further comprising the step of forming said gaseous catalyst from a source of gaseous catalyst comprising a salt of one or more cations and at least one anion selected from the group consisting of halides, sulfates, phosphates, carbonates, hydroxide and sulfides.

173. A method according to claim **145**, wherein said gaseous catalyst is selected to have a resonant adsorption with the energy released from said gaseous hydrogen atoms undergoing a transition to a lower energy state.

174. A method according to claim **145**, wherein said gaseous catalyst comprises potassium and has a net enthalpy of reaction of 27.28 eV.

175. A method according to claim **145**, wherein said gaseous catalyst has a net enthalpy of reaction of about 27.2 eV.

176. A method according to claim **145**, further comprising the step of combining a source of gaseous catalyst with a source of gaseous hydrogen atoms comprising at least one of a hydrocarbon or water, and providing combustion which volatizes said source of gaseous catalyst to form said gaseous catalyst and provide said gaseous hydrogen atoms.

177. A method according to claim **145**, further comprising the step of forming a source of gaseous catalyst or said gaseous catalyst in situ.

178. A method according to claim **177**, wherein said step of forming said source of gaseous catalyst or said gaseous catalyst in situ comprises the ionization of a reactant.

179. A method according to claim **178**, wherein said step of ionization comprises thermal ionization of said reactant.

180. A method according to claim **178**, wherein said step of ionization comprises chemical ionization of said reactant.

181. A method according to claim **180**, wherein said step of chemical ionization comprises oxidation or reduction of said reactant.

182. A method according to claim **145**, further comprising the step of volatizing a source of catalyst to form said gaseous catalyst utilizing energy from at least one of heat, electron-beam energy, photon energy, acoustic energy, electric field, or magnetic field.

183. A method according to claim **145**, wherein said the step of volatizing a source of catalyst comprises the step of heating a filament coated with said source of gaseous catalyst.

184. A method according to claim **145**, further comprising adding a source of catalyst to said reaction vessel and heating said source of catalyst to form said gaseous catalyst.

185. A method according to claim **145**, further comprising the step of providing a source of catalyst in a catalyst reservoir comprising a container separate from said vessel and said container communicates with said reaction vessel.

186. A method according to claim **145**, further comprising the step of providing a source of catalyst in a boat contained within said reaction vessel.

187. A method according to claim **145**, wherein said step of providing hydrogen atoms comprises the steps of disassociating a hydrogen containing compound into hydrogen atoms.

188. A method according to claim **145**, wherein said step of providing hydrogen atoms comprises the steps of passing a hydrogen containing gas over a hot filament.

189. A method according to claim **145**, wherein said step of providing hydrogen atoms comprises the steps of passing a hydrogen containing gas over a hot grid.

190. A method according to claim **145**, wherein said step of providing hydrogen atoms comprises the steps of passing a hydrogen containing gas through a tungsten capillary heated by electron bombardment.

191. A method according to claim **145**, wherein said step of providing hydrogen atoms comprises the steps of maintaining a hydride under nonequilibrium conditions.

192. A method according to claim **147**, wherein said second catalyst comprises at least one element selected from the group consisting of transition elements, inner transition elements, precious metals, refractory metals, lanthanides, actinides and activated charcoal.

193. A method according to claim **147**, wherein said second catalyst is selected from the group consisting of an element, compound, alloy or mixture of transition elements, inner transition elements, iron, platinum, palladium, zirconium, vanadium, nickel, titanium, Sc, Cr, Mn, Co, Cu, Zn, Y, Nb, Mo, Tc, Ru, Rh, Ag, Cd, La, Hf, Ta, W, Re, Os, Ir, Au, Hg, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Vb, Lu, Th, Pa, U, activated charcoal, and intercalated Cs carbon.

194. A method according to claim **147**, further comprising the step of utilizing a hot filament or hot grid to disassociate a hydrogen containing gas into gaseous hydrogen atoms and to heat said second catalyst.

195. A method according to claim **147**, further comprising the step of controlling the power output of said cell.

196. A method according to claim **147**, wherein said step of controlling the power output of said cell comprises controlling the temperature of said second catalyst.

197. A method according to claim **196**, wherein said step of controlling the temperature of said second catalyst comprises utilizing a heated filament or grid.

198. A method according to claim **145**, further comprising the step of forming said gaseous hydrogen atoms from at least one source of gaseous hydrogen atoms selected from the group consisting of hydrogen gas, water, hydrides, metal-hydrogen solutions, and hydrocarbons.

199. A method according to claim **145**, wherein said step of providing gaseous hydrogen atoms comprises pyrolyzing hydrocarbons or water.

200. A method according to claim **145**, further comprising the step of reforming hydrocarbons to at least one of gaseous molecular and gaseous atomic hydrogen.

201. A method according to claim **145**, further comprising the step of disassociating hydrogen containing molecules using UV light to form said gaseous hydrogen atoms.

202. A method according to claim **145**, further comprising the step of controlling the amount of gaseous hydrogen provided in said reaction vessel.

203. A method according to claim **145**, further comprising the step of controlling the flow of a source of gaseous hydrogen atoms or said gaseous hydrogen atoms from a chamber to said reaction vessel.

204. A method according to claim **145**, further comprising the step of utilizing a valve for controlling the flow of gaseous hydrogen or source of gaseous hydrogen from said reaction vessel.

205. A method according to claim **145**, further comprising the step of controlling the flow of gaseous hydrogen from said reaction vessel.

206. A method according to claim **145**, further comprising the step of utilizing a vacuum pump for controlling the flow of gaseous hydrogen from said reaction vessel.

207. A method according to claim **145**, further comprising the step of utilizing a valve for controlling the flow of gaseous hydrogen from a chamber into said reaction vessel.

208. A method according to claim **145**, further comprising controlling the partial pressure of said gaseous hydrogen atoms.

209. A method according to claim **145**, wherein a partial pressure of said gaseous hydrogen atoms or source of gaseous hydrogen atoms in said reaction vessel is maintained within the range of 10^{-3} atmospheres to 100 atmospheres.

210. A method according to claim **145**, further comprising controlling the amount of gaseous catalyst or a source of gaseous catalyst introduced into said reaction vessel.

211. A method according to claim **145**, further comprising controlling the flow of gaseous catalyst or a source of gaseous catalyst from a catalyst reservoir containing gaseous catalyst or a source of gaseous catalyst to said reaction vessel.

212. A method according to claim **145**, further comprising controlling the flow of gaseous catalyst or a source of gaseous catalyst from a boat containing gaseous catalyst or a source of gaseous catalyst.

213. A method according to claim **145**, further comprising the step of controlling the flow of said gaseous catalyst or a source of gaseous catalyst from said reaction vessel.

214. A method according to claim **145**, further comprising the step of controlling the vapor pressure of said gaseous catalyst or a source of gaseous catalyst in said reaction vessel.

215. A method according to claim **145**, wherein a vapor pressure of said gaseous catalyst or a source of gaseous catalyst is maintained at about 50 to 210 millitorr.

216. A method according to claim **145**, further comprising using a vacuum pump to control the flow of said gaseous catalyst or a source of gaseous catalyst from said reaction vessel.

217. A method according to claim **145**, further comprising using a valve to control the flow of a source of gaseous catalyst or a source of gaseous catalyst from a catalyst reservoir into said reaction vessel.

218. A method according to claim **145**, further comprising using a valve to control the flow of a source of gaseous catalyst or said catalyst from said reaction vessel.

219. A method according to claim **145**, further comprising the step of supplying a nonreactive gas to said reaction vessel.

220. A method according to claim **219**, further comprising the step of controlling the vapor pressure of said nonreactive gas in said reaction vessel.

221. A method according to claim **219**, further comprising the step of controlling the flow of said nonreactive gas supplied to said reaction vessel.

222. A method according to claim **219**, further comprising the step of controlling the amount of said nonreactive gas released from said reaction vessel.

223. A method according to claim **219**, further comprising utilizing a vacuum pump for controlling the flow of said nonreactive gas from said reaction vessel.

224. A method according to claim **219**, further comprising the step of utilizing a valve for controlling the flow of said nonreactive gas from said reaction vessel.

225. A method according to claim **219**, further comprising the step of utilizing a valve for controlling the flow of said nonreactive gas into said reaction vessel.

226. A method according to claim **145**, further comprising the step of controlling the vapor pressure of said gaseous catalyst in said reaction vessel.

227. A method according to claim **226**, wherein said step of controlling the vapor pressure of said gaseous catalyst comprises controlling the temperature in a catalyst reservoir containing a source of gaseous catalyst or said gaseous catalyst and being in communication with said reaction vessel, and controlling the flow of gaseous catalyst from said catalyst reservoir.

228. A method according to claim **226**, wherein said step of controlling the vapor pressure of said gaseous catalyst comprises controlling the temperature in a boat containing a source of gaseous catalyst or said gaseous catalyst and being contained in said reaction vessel, and controlling the flow of gaseous catalyst from said boat.

229. A method according to claim **145**, further comprising the step of controlling the temperature in a catalyst reservoir containing a source of gaseous catalyst or said gaseous catalyst and being in communication with said reaction vessel.

230. A method according to claim **145**, further comprising the step of controlling the temperature in a boat containing a source of gaseous catalyst or said gaseous catalyst and being contained in said reaction vessel.

231. A method according to claim **145**, wherein the reaction to provide a net enthalpy of about 27 (p/2) eV, where p is a positive integer greater than 1, comprises an electrochemical, chemical, photochemical, thermal, free radical, sonic, nuclear, inelastic photon, or particle scattering reaction, or mixtures thereof.

232. A method according to claim **145**, wherein a pressure in said reaction vessel is maintained within the range of 10^{-3} atmospheres to 100 atmospheres.

233. A method according to claim **145**, wherein said reaction occurring at a pressure less than atmospheric pressure.

234. A method according to claim **145**, further comprising the step of releasing hydrogen atoms from said reaction vessel having a binding energy of about $E_b=13.6/n^2$ eV, where n is a fraction whose numerator is 1 and denominator is an integer greater than 1 or a compound containing said hydrogen atoms.

235. A method according to claim **145**, further comprising the step of adsorbing said released energy.

236. A method according to claim **145**, further comprising the step of converting energy released from said hydrogen atom into electrical energy.

237. A method according to claim **145**, wherein said reaction step is conducted in an internal combustion chamber.

238. A method according to claim **145**, wherein said internal combustion chamber is an engine cylinder.

239. A method according to claim **145**, further comprising the step of controlling the power output of said cell.

240. A method according to claim **239**, wherein said step of controlling the power output of said cell comprises controlling the amount of said gaseous catalyst present in said reaction vessel.

241. A method according to claim **240**, wherein said step of controlling the amount of gaseous catalyst comprises controlling the temperature of said reaction vessel and selecting said gaseous catalyst or source of gaseous catalyst to have a vapor pressure dependent upon the temperature of said reaction vessel.

242. A method according to claim **240**, wherein said step of controlling the amount of gaseous catalyst comprises controlling the temperature of a catalyst reservoir containing a source of gaseous catalyst or said gaseous catalyst and being in communication with said reaction vessel, and selecting said gaseous catalyst to have a vapor pressure dependent upon the temperature of said catalyst reservoir.

243. A method according to claim **240**, wherein said step of controlling the amount of gaseous catalyst comprises controlling the flow of said source of gaseous catalyst or gaseous catalyst from said catalyst reservoir into said reaction vessel.

244. A method according to claim **240**, wherein said step of controlling the amount of gaseous catalyst comprises controlling the flow of said source of gaseous catalyst or gaseous catalyst from said reaction vessel.

245. A method according to claim **240**, wherein said step of controlling the amount of gaseous catalyst comprises controlling the temperature of a boat containing a source of gaseous catalyst or said gaseous catalyst and being contained in said reaction vessel, and selecting said gaseous catalyst to have a vapor pressure dependent upon the temperature of said boat.

246. A method according to claim **239**, wherein said step of controlling the power output of said cell comprises controlling the amount of said gaseous hydrogen atoms present in said reaction vessel.

247. A method according to claim **246**, wherein said step of controlling the amount of said gaseous hydrogen atoms comprises controlling the flow of gaseous hydrogen atoms or source of gaseous hydrogen atoms into said reaction vessel.

248. A method according to claim **246**, wherein said step of controlling the amount of said source of gaseous hydrogen atoms or gaseous hydrogen atoms comprises controlling the flow of said source of gaseous hydrogen atoms or gaseous hydrogen atoms from said reaction vessel.

249. A method according to claim **246**, wherein said step of controlling the amount of said source of gaseous hydrogen atoms or gaseous hydrogen atoms comprises controlling the temperature of a second catalyst for dissociating a hydrogen containing compound into gaseous hydrogen atoms.

250. A method according to claim **249**, wherein said step of controlling the temperature of a second catalyst for dissociating a hydrogen containing compound into gaseous hydrogen atoms comprises controlling the power dissipated in a second catalyst heater.

251. A method according to claim **239**, wherein said step of controlling the power output of said cell comprises controlling the amount of nonreactive gas in said reaction vessel.

252. A method according to claim **251**, wherein said step of controlling the amount of nonreactive gas in said reaction vessel comprises controlling the flow of nonreactive gas into said reaction vessel.

253. A method according to claim **252**, wherein said step of controlling the amount of nonreactive gas in said reaction vessel comprises controlling the flow of nonreactive gas from said reaction vessel.

254. A method according to claim **239**, wherein said step of controlling the power output of said cell comprises controlling the temperature of a source of gaseous catalyst.

255. A method according to claim **239**, wherein said step of controlling the power output of said cell comprises controlling the flow of a hydrogen containing gas over at least one of a hot filament, a tungsten capillary heated by electron bombardment, or an inductively coupled plasma flow.

256. A method according to claim **239**, wherein said step of controlling the power output of said cell comprises controlling the power dissipated in an inductively coupled plasma flow tube, hot filament or grid, or tungsten capillary heated by electron bombardment.

257. A method according to claim **239**, wherein said step of controlling the power output of said cell comprises controlling the temperature of a hot filament or tungsten capillary heated by electron bombardment over which a hydrogen containing gas flows.

258. A method according to claim **239**, wherein said step of controlling the power output of said cell comprises controlling the temperature of a hydride maintained under nonequilibrium conditions.

259. A method according to claim **145**, wherein a temperature in said reaction vessel is maintained at a higher temperature than in a catalyst reservoir in communication with said reaction vessel or a boat contained within said reaction vessel.

260. A method according to claim **145**, further comprising the step of measuring the temperature of a catalyst reservoir in communication with said reaction vessel or a boat contained in said reaction vessel.

261. A method according to claim **145**, further comprising the step of measuring the temperature of a source of said gaseous catalyst contained in a catalyst reservoir in communication with said reaction vessel or a boat contained within said reaction vessel.

262. A method according to claim **145**, further comprising the step of measuring the temperature of a chamber containing a source of said hydrogen atoms in communication with said reaction vessel.

263. A method according to claim **145**, further comprising the step of measuring the temperature of a source of said gaseous hydrogen atoms.

264. A method according to claim **147**, further comprising the step of measuring the temperature of said second catalyst.

265. A method according to claim **145**, further comprising the step of controlling the temperature of said reaction vessel.

266. A method according to claim **145**, further comprising utilizing a nebulizer or atomizer to form said gaseous catalyst.

267. A method according to claim **145**, further comprising the step of measuring the pressure in said reaction vessel.

268. A method according to claim **145**, further comprising the step of controlling the pressure in said reaction vessel.

269. A method according to claim **145**, further comprising the step of measuring the hydrogen pressure in said reaction vessel.

270. A method according to claim **145**, further comprising the step of measuring the gaseous catalyst pressure in said reaction vessel.

271. A method according to claim **145**, further comprising the step of converting energy released from said gaseous hydrogen atoms into electrical energy.

272. A method according to claim **145**, further comprising controlling the pressure of said gaseous catalyst by controlling the amount of said source of catalyst being added to said reaction vessel.

273. A cell for extracting energy from hydrogen atoms comprising:

a reaction vessel;

a source of gaseous hydrogen atoms; and

a source of a gaseous catalyst having a net enthalpy of reaction of about 27(p/2) eV, where p is an integer greater than 1.

274. A cell for extracting energy from hydrogen atoms comprising:

a reaction vessel;

a chamber communicating with said vessel, said chamber containing gaseous hydrogen atoms or a source of said hydrogen atoms; and

a catalyst reservoir communicating with said reaction vessel or a boat contained in said reaction vessel, said catalyst reservoir or boat containing a gaseous catalyst having a net enthalpy of reaction of about 27*(p/2) eV, where p is an integer greater than 1, or a source of said gaseous catalyst.

275. A cell according to claim **274**, wherein said source of gaseous catalyst comprises an ionic compound which is resistant to hydrogen reduction.

276. A cell according to claim **274**, wherein said source of gaseous catalyst comprises an ionic compound which is resistant to hydrogen reduction and which is adapted to sublime, boil or become volatile when heated.

277. A cell according to claim **274**, wherein said source of gaseous catalyst comprises an ionic compound which is adapted to sublime, boil or become volatile when heated.

278. A cell according to claim **274**, wherein said source of gaseous catalyst comprises an ionic compound which is resistant to thermal degradation.

279. A cell according to claim **274**, wherein said source of gaseous catalyst comprises a salt of rubidium or potassium.

280. A cell according to claim **274**, wherein said source of gaseous catalyst comprises a salt that can form a vapor comprising ions when heated.

281. A cell according to claim **274**, wherein said source of gaseous catalyst comprises a salt of rubidium selected from the group consisting of RbF, $RbCl$, $RbBr$, Rbl, Rb_2S_2, $RbOH$, Rb_2SO_4, Rb_2CO_3, and Rb_3PO_4.

282. A cell according to claim **274**, wherein said source of gaseous catalyst comprises a salt of potassium selected from the group consisting of KF, KCl, KBr, KI, K_2S_2, KOH, K_2SO_4, K_2CO_3, K_3PO_4, and K_2GeF_4.

283. A cell according to claim **274**, wherein said source of gaseous catalyst comprises at least one metal selected from the group consisting of Mo, Ti, and Rb.

284. A cell according to claim **274**, wherein said source of gaseous catalyst comprises at least one salt selected from the group consisting of Mol_2, $TiCl_2$, $TiCl_4$, $SnCl_4$, $SiCl_4$, $PrBr_3$, $CaBr_2$, $SrCl_2$, Crl_2, Tbl_3, $SbCl_3$, CrF_3, $CoCl_2$, $BiCl_3$, $NiCl_2$, PdF_2, $InCl$, $LaCl_3$, $DyCl_3$, Lal_3, Hol_3, KNO_3, VF_3, PbF_2, $VOCl$, Pbl_2, $LuCl_3$, $PbCl_2$, Asl_3, Hol_3, $MoCl_5$, $SnCl_4$, Sbl_3, Cdl_2, AgF_2, AgF, Lal_3, Erl_3, VCl_4, BCl_3, $FeCl_3$, $TiCl_3$, Col_2, CoF_2, Tll, TlF, $BiBr_3$, $ZnBr_2$, Asl_3, Dyl_3, $HoCl_3$, $MgCl_2$, $CrCl_3$, $PrCl_3$, $SrCl_2$, $FeCl_2$, $NiCl_2$, $CuCl$, $SrCl_2$, $MoCl_2$, YCl_3, $ZrCl_4$, Cdl_2, Bal_2, Hol_3, Pbl_2, PdF_2, LiF, $EuCl_3$, $MgCl_2$, $ErCl_3$, $MgCl_2$, $ErCl_3$, $MgCl_2$, $BiCl_4$, $AlCl_3$, $CaBr_2$, $SmBr_3$, VaF_3, $LaCl_3$, Gdl_3, Crl_2, Mnl_2, $YbBr_3$, $FeBr_2$, $NiCl_2$, $AgCl$, $ZnCl_2$, $YbCl_2$, SeF_4, $SnCl_4$, SnF_4, Sbl_3, Bil_2, EuF_3, and $PbCl_2$.

285. A cell according to claim **274**, wherein said gaseous catalyst comprises at least one ion selected from the group consisting of Mo^{2+}, Ti^{2+}, and Rb^{+}.

286. A cell according to claim **274**, wherein said gaseous catalyst comprises at least one pair of ions selected from the group consisting of: (Sn^{4+}, Si^{4+}), (Pr^{3+}, Ca^{2+}), (Sr^{2+}, Cr^{2+}), (Cr^{3+}, Tb^{3+}), (Sb^{3+},Co^{2+}), (Bi^{3+}, Ni^{2+}), (Pd^{2+}, In^{+}), (La^{3+}, Dy^{3+}, Ho^{3+}), K^{+}, K^{+}) , (V^{3+}, Pd^{2+}), (Lu^{3+}, Zn^{2+}), (As^{3+}, Ho^{3+}), (Mo^{5+}, Sn^{4+}), (Sb^{3+}, Cd^{2+}), (Ag^{2+}, Ag^{+}), (La^{3+}, Er^{3+}), (V^{4+}, B^{3+}), (Fe^{3+}, Ti^{3+}), (Co^{2+}, Ti^{+}), (Bi^{3+}, Zn^{2+}), (As^{3+}, Dy^{3+}), (Ho^{3+}, Mg^{2+}), (K^{+}, Rb^{+}), (Cr^{3+}, Pr^{3+}), (Sr^{2+}, Fe^{2+}), (Ni^{2+}, Cu^{+}), (Sr^{2+}, Mo^{2+}), (Y^{3+}, Zr^{4+}), (Cd^{2+}, Ba^{2+}),

(Ho^{3+}, Pb^{2+}), (Pb^{2+}, Li^{+}), (Eu^{3+}, Mg^{2+}), (Er^{3+}, Mg^{2+}), (Bi^{4+}, Al^{3+}), (Ca^{2+}, Sm^{3+}), (V^{3+}, La^{3+}), (Gd^{3+}, Cr^{2+}), (Mn^{2+}, Ti^{+}), (Yb^{3+}, Fe^{2+}), (Ni^{2+}, Ag^{+}), (Zn^{2+}, Yb^{2+}), $(Se^{4+}Sn^{4+})$, (Sb^{3+}, Bi^{2+}), and (Eu^{3+}, Pb^{2+}).

287. A cell according to claim **274**, wherein said source of gaseous catalyst comprises a salt of one or more cations and at least one anion selected from the group consisting of halides, sulfates, phosphates, carbonates, hydroxide and sulfides.

288. A cell according to claim **274**, wherein said gaseous catalyst comprises hydrogen atoms having a binding energy of about $E_b=13.6/n^2$ eV, where n is a fraction whose numerator is 1 and denominator is an integer greater than 1.

289. A cell according to claim **274**, wherein said gaseous catalyst comprises potassium and has a net enthalpy of reaction of 27.28 eV.

290. A cell according to claim **274**, wherein said gaseous catalyst has a net enthalpy of reaction of about 27.2 eV.

291. A cell according to claim **274**, wherein said cell comprises said source of said gaseous catalyst combined with at least one of a hydrocarbon or water disposed such that said gaseous catalyst and said gaseous hydrogen atoms are capable being formed during combustion.

292. A cell according to claim **274**, further comprising a means for converting said source of catalyst to said gaseous catalyst.

293. A cell according to claim **292**, wherein said means for converting said source of catalyst to said gaseous catalyst comprises at least one of heat, electron-beam energy, photon energy, acoustic energy, electric field, or magnetic field.

294. A cell according to claim **274**, further comprising a filament coated with said source of gaseous catalyst.

295. A cell according to claim **274**, wherein said source of gaseous hydrogen atoms is selected from the group consisting of hydrogen gas, water, hydrides, metal-hydrogen solutions, and hydrocarbons.

296. A cell according to claim **274**, wherein said source of hydrogen atoms comprises a hot filament and a hydrogen containing gas stream.

297. A cell according to claim **274**, wherein said source of hydrogen atoms comprises a hot grid and a hydrogen containing gas stream.

298. A cell according to claim **274**, wherein said source of hydrogen atoms comprises a heated tungsten capillary and a hydrogen containing gas stream.

299. A cell according to claim **274**, wherein said source of hydrogen atoms comprises a hydride maintained under nonequilibrium conditions.

300. A cell according to claim **274**, wherein said source of hydrogen atoms comprises an inductively coupled plasma flow tube and a hydrogen gas containing stream.

301. A cell according to claim **274**, wherein said source of hydrogen atoms comprises a hydrogen containing gas stream and a second catalyst for disassociating said hydrogen containing gas stream into free hydrogen atoms.

302. A cell according to claim **274**, wherein said source of hydrogen atoms comprises an internal combustion engine.

303. A cell according to claim **301**, wherein said second catalyst comprises at least one element selected from the group consisting of transition elements, inner transition elements, precious metals, refractory metals, lanthanides, actinides, and activated charcoal.

304. A cell according to claim **301**, wherein said second catalyst comprises at least one selected from the group consisting of an element, compound, alloy or mixture of transition elements, inner transition elements, iron, platinum, palladium, zirconium, vanadium, nickel, titanium, Sc, Cr, Mn, Co, Cu, Zn, Y, Nb, Mo, Tc, Ru, Rh, Ag, Cd, La, Hf, Ta, W, Re, Os, Ir, Au, Hg, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Vb, Lu, Th, Pa, U, activated charcoal, and intercalated Cs carbon.

305. A cell according to claim **301**, wherein said second catalyst is treated with an aqueous solution of K_2CO_3 and H_2O_2.

306. A cell according to claim **301**, further comprising a heater to heat said second catalyst.

307. A cell according to claim **301**, further comprising a filament or grid constructed and arranged to dissociate hydrogen and to heat said second catalyst.

308. A cell according to claim **301**, further comprising means for controlling the power output of said cell.

309. A cell according to claim **308**, wherein said means for controlling the power output of said cell comprises means for controlling the temperature of said second catalyst.

310. A cell according to claim **309**, wherein said means for controlling the temperature of said second catalyst comprises a filament or grid.

311. A cell according to claim **301**, further comprising a temperature controlling structure capable of maintaining a selected atomic hydrogen partial pressure by controlling the temperature of said second catalyst.

312. A cell according to claim **311**, wherein said temperature controlling structure is at least one selected from the group consisting of an internal heater, an external heater, the catalysis of hydrogen, and a heat exchanger which removes energy from the cell.

313. A cell according to claim **274**, wherein said source of hydrogen atoms comprises a means for pyrolysis of hydrocarbons or water.

314. A cell according to claim **274**, further comprising means to reform hydrocarbons to at least one of gaseous molecular and gaseous atomic hydrogen.

315. A cell according to claim **274**, further comprising a source of UV light for disassociating hydrogen containing molecules to form said gaseous hydrogen atoms.

316. A cell according to claim **274**, further comprising a flow control means for controlling the flow of said source of gaseous hydrogen atoms or said gaseous hydrogen atoms from said chamber to said reaction vessel.

317. A cell according to claim **316**, wherein said flow control means comprises a valve.

318. A cell according to claim **274**, further comprising a flow control means for controlling the flow of hydrogen from said reaction vessel.

319. A cell according to claim **318**, wherein said flow control means comprises a valve.

320. A cell according to claim **274**, further comprising a vacuum pump constructed and arranged for controlling the flow of hydrogen from said reaction vessel.

321. A cell according to claim **274**, further comprising a flow control means for controlling the flow of said gaseous catalyst or a source of gaseous catalyst from said catalyst reservoir to said reaction vessel.

322. A cell according to claim **321**, wherein said flow control means comprises a valve.

323. A cell according claim **274**, further comprising means for controlling the flow of catalyst from said reaction vessel.

324. A cell according to claim **274**, further comprising a valve constructed and arranged for controlling the flow of said catalyst from said reaction vessel.

325. A cell according to claim **274**, further comprising a vacuum pump constructed and arranged for controlling the flow of said gaseous catalyst or source of gaseous catalyst from said reaction vessel.

326. A cell according to claim **274**, further comprising a nonreactive gas.

327. A cell according to claim **274**, further comprising a means for controlling the amount of a nonreactive gas in said vessel.

328. A cell according to claim **274**, further comprising flow control means for controlling the flow of a nonreactive gas into said reaction vessel.

329. A cell according to claim **328**, wherein said means for controlling the amount of nonreactive gas in said vessel comprises a valve constructed and arranged to regulate the release of said nonreactive gas from said reaction vessel.

330. A cell according to claim **274**, further comprising a valve constructed and arranged for controlling the flow of a nonreactive gas from said reaction vessel.

331. A cell according to claim **274**, further comprising a vacuum pump constructed and arranged for controlling the flow of a nonreactive gas from said reaction vessel.

332. A cell according to claim **274**, further comprising a pump in communication with said reaction vessel.

333. A cell according to claim **274**, further comprising structure for controlling the vapor pressure of said catalyst.

334. A cell according to claim **333**, wherein said structure for controlling the vapor pressure of said catalyst comprises a heater constructed and arranged to control the temperature of said catalyst reservoir or said boat.

335. A cell according to claim **333**, wherein said structure for controlling the vapor pressure of said catalyst comprises a heater constructed and arranged to control the temperature of said reaction chamber.

336. A cell according to claim **274**, further comprising structure for maintaining a selected vapor pressure of said gaseous catalyst.

337. A cell according to claim **336**, wherein said structure for maintaining a selected vapor pressure of said gaseous catalyst comprises a valve constructed and arranged for controlling the flow of said source of gaseous catalyst or gaseous catalyst from said catalyst reservoir and a valve constructed and arranged for controlling the flow of said source of gaseous catalyst or gaseous catalyst from said reaction vessel.

338. A cell according to claim **336**, wherein said structure for maintaining a selected vapor pressure of said source of gaseous catalyst or said gaseous catalyst comprises a valve constructed and arranged for controlling the flow of said source of gaseous catalyst or gaseous catalyst from said reaction vessel.

339. A cell according to claim **274**, further comprising a valve for releasing hydrogen atoms having a binding energy of about $E_b=13.6/n^2$ eV, where n is a fraction whose numerator is 1 and denominator is an integer greater than 1 or a compound containing said hydrogen atoms.

340. A cell according to claim **274**, further comprising means for adsorbing energy released from said hydrogen atom.

341. A cell according to claim **274**, wherein said vessel comprises an internal combustion chamber.

342. A cell according to claim **341**, wherein said internal combustion chamber is an engine cylinder.

343. A cell according to claim **274**, further comprising means for controlling the power output of said cell.

344. A cell according to claim **343**, wherein said means for controlling the power output of said cell comprises means for controlling the amount of said gaseous catalyst.

345. A cell according to claim **344**, wherein said means for controlling the amount of gaseous catalyst comprises means for controlling the temperature of said vessel and said gaseous catalyst is selected to have a vapor pressure dependent upon the temperature of said reaction vessel.

346. A cell according to claim **344**, wherein said means for controlling the amount of gaseous catalyst comprises means for controlling the temperature of said catalyst reservoir and said gaseous catalyst is selected to have a vapor pressure dependent upon the temperature of said catalyst reservoir.

347. A cell according to claim **344**, wherein said means for controlling the amount of gaseous catalyst comprises means for controlling the temperature of said boat and said gaseous catalyst is selected to have a vapor pressure dependent upon the temperature of said boat.

348. A cell according to claim **344**, wherein said means for controlling the amount of gaseous catalyst comprises means for controlling the flow of said source of gaseous catalyst or gaseous catalyst from said catalyst reservoir into said reaction vessel.

349. A cell according to claim **344**, wherein said means for controlling the amount of gaseous catalyst comprises means for controlling the flow of said source of gaseous catalyst or gaseous catalyst from said catalyst reservoir from said reaction vessel.

350. A cell according to claim **343**, wherein said means for controlling the power output of said cell comprises means for controlling the amount of said gaseous hydrogen atoms or said source of gaseous hydrogen atoms in said vessel.

351. A cell according to claim **350**, wherein said means for controlling the amount of said gaseous hydrogen atoms or said source of gaseous hydrogen atoms comprises means for controlling the flow of gaseous hydrogen atoms or source of said hydrogen atoms into said vessel.

352. A cell according to claim **350**, wherein said means for controlling the amount of said gaseous hydrogen atoms or said source of gaseous hydrogen atoms comprises means for controlling the flow of gaseous hydrogen atoms or source of said hydrogen atoms from said vessel.

353. A cell according to claim **343**, wherein said means for controlling the power output of said cell comprises controlling the amount of said nonreactive gas present in said reaction vessel.

354. A cell according to claim **353**, wherein said means for controlling the amount of nonreactive gas comprises means for controlling the flow of said nonreactive gas into said reaction vessel.

355. A cell according to claim **354**, wherein said means for controlling the amount of nonreactive gas comprises means for controlling the flow of said nonreactive gas from said reaction vessel.

356. A cell according to claim **343**, wherein said means for controlling the power output of said cell comprises means for controlling the temperature of said source of catalyst.

357. A cell according to claim **343**, wherein said means for controlling the power output of said cell comprises means for controlling the flow of a hydrogen containing gas over at least one of a hot filament, a tungsten capillary heated by electron bombardment, or an inductively coupled plasma flow.

358. A cell according to claim **343**, wherein said means for controlling the power output of said cell comprises means for controlling the power dissipated in an inductively coupled plasma flow tube.

359. A cell according to claim **343**, wherein said means for controlling the power output of said cell comprises means for controlling the temperature of a hot filament or tungsten capillary heated by electron bombardment over which a hydrogen containing gas flows.

360. A cell according to claim **343**, wherein said means for controlling the power output of said cell comprises means for controlling the temperature of a hydride maintained under nonequilibrium conditions.

361. A cell according to claim **274**, further comprising means for measuring the temperature of said catalyst reservoir or said boat.

362. A cell according to claim **274**, further comprising a means for measuring the temperature of said source of said gaseous catalyst contained in said catalyst reservoir or said boat.

363. A cell according to claim **274**, further comprising means for measuring the temperature of said reaction vessel.

364. A cell according to claim **274**, further comprising means for measuring the temperature of said source of said gaseous hydrogen atoms.

365. A cell according to claim **301**, further comprising means for measuring the temperature said second catalyst.

366. A cell according to claim **274**, further comprising means to measure the cell temperature.

367. A cell according to claim **274**, further comprising temperature controlling structure constructed and arranged to maintain a temperature in said reaction vessel greater than a temperature in said catalyst reservoir.

368. A cell according to claim **274**, further comprising temperature controlling structure constructed and arranged to maintain a temperature in said reaction vessel greater than a temperature in said boat.

369. A cell according to claim **274**, further comprising temperature controlling structure for maintaining a selected temperature of said reaction vessel.

370. A cell according to claim **274**, further comprising a nebulizer or atomizer.

371. A cell according to claim **274**, further comprising means to measure the pressure in said reaction vessel.

372. A cell according to claim **274**, further comprising means to measure the hydrogen pressure in said reaction vessel.

373. A cell according to claim **274**, further comprising means to measure the gaseous catalyst pressure in said reaction vessel.

374. A cell according to claim **274**, wherein said vessel is capable of containing a pressure within the range of 10^{-3} atmospheres to 100 atmospheres.

375. A method for extracting energy from hydrogen atoms comprising the steps of:

volatizing a source of gaseous catalyst to form a gaseous catalyst having a net enthalpy of reaction of about 27*(p/2) eV, where p is an integer greater than 1;

providing gaseous hydrogen atoms; and

reacting said gaseous catalyst with said gaseous hydrogen atoms, thereby releasing energy from said gaseous hydrogen atoms and producing hydrogen atoms having a binding energy of about $E_b=13.6/n^2$ eV, where n is a fraction whose numerator is 1 and a denominator is an integer greater than 1.

376. A method according to claim **375**, wherein said source of gaseous catalyst comprises an ionic compound which is resistant to hydrogen reduction.

377. A method according to claim **375**, wherein said source of gaseous catalyst comprises an ionic compound which is resistant to hydrogen reduction and which is adapted to sublime, boil or become volatile when heated.

378. A method according to claim **375**, wherein said source of gaseous catalyst comprises an ionic compound which is adapted to sublime, boil or become volatile when heated.

379. A method according to claim **375**, wherein said source of gaseous catalyst comprises an ionic compound which is resistant to thermal degradation.

380. A method according to claim **375**, wherein said source of gaseous catalyst comprises a salt of rubidium or potassium.

381. A method according to claim **375**, wherein said source of gaseous catalyst comprises a salt that can form a vapor comprising ions when heated.

382. A method according to claim **375**, wherein said source of gaseous catalyst comprises a salt of rubidium selected from the group consisting of RbF, RbCl, RbBr, RbI, Rb_2S_2, RbOH, Rb_2SO_4, Rb_2CO_3, and Rb_3PO_4.

383. A method according to claim **375**, wherein said source of gaseous catalyst comprises a salt of potassium selected from the group consisting of KF, KCl, KBr, KI, K_2S_2, KOH, K_2SO_4, K_2CO_3, K_3PO_4, and K_2GeF_4.

384. A method according to claim **375**, wherein said source of gaseous catalyst comprises at least one metal selected from the group consisting of Mo, Ti, and Rb.

385. A method according to claim **375**, wherein said source of gaseous catalyst comprises at least one salt selected from the group consisting of MoI_2, $TiCl_2$, $TiCl_4$, $SnCl_4$, $SiCl_4$, $PrBr_3$, $CaBr_2$, $SrCl_2$, CrI_2, TbI_3, $SbCl_3$, CrF_3, $CoCl_2$, $BiCl_3$, $NiCl_2$, PdF_2, InCl, $LaCl_3$, $DyCl_3$, LaI_3, HoI_3, KNO_3, VF_3, PbF_2, VOCl, PbI_2, $LuCl_3$, $PbCl_2$, AsI_3, HoI_3, $MoCl_5$, $SnCl_4$, SbI_3, CdI_2, AgF_2, AgF, LaI_3, ErI_3, VCl_4, BCl_3, $FeCl_3$, $TiCl_3$, CoI_2, CoF_2, TlI, TlF, $BiBr_3$, $ZnBr_2$, AsI_3, DyI_3, $HoCl_3$, $MgCl_2$, $CrCl_3$, $PrCl_3$, $SrCl_2$, $FeCl_2$, $NiCl_2$, CuCl, $SrCl_2$, $MoCl_2$, YCl_3, $ZrCl_4$, CdI_2, BaI_2, HoI_3, PbI_2, PdF_2, LiF, $EuCl_3$, $MgCl_2$, $ErCl_3$, $MgCl_2$, $ErCl_3$, $MgCl_2$, $BiCl_4$, $AlCl_3$, $CaBr_2$, $SmBr_3$, VaF_3, $LaCl_3$, GdI_3, CrI_2, MnI_2, $YbBr_3$, $FeBr_2$, $NiCl_2$, AgCl, $ZnCl_2$, $YbCl_2$, SeF_4, $SnCl_4$, SnF_4, SbI_3, BiI_2, EuF_3, and $PbCl_2$.

386. A method according to claim **375**, wherein said gaseous catalyst comprises at least one ion selected from the group consisting of Mo^{2+}, Ti^{2+}, and Rb^{+}.

387. A method according to claim **375**, wherein said gaseous catalyst comprises at least one pair of ions selected from the group consisting of: (Sn^{4+}, Si^{4+}), (Pr^{3+}, Ca^{2+}), (Sr^{2+}, Cr^{2+}), (Cr^{3+}, Tb^{3+}), (Sb^{3+},Co^{2+}), (Bi^{3+}, Ni^{2+}), (Pd^{2+}, In^{+}), (La^{3+}, Dy^{3+}), (La^{3+}, Ho^{3+}), (K^{+}, K^{+}), (V^{3+}, Pd^{2+}), (Lu^{3+}, Zn^{2+}), (As^{3+}, Ho^{3+}), (Mo^{5+}, Sn^{4+}), (Sb^{3+}, Cd^{2+}, (Ag^{2+},Ag^{+}), (La^{3+}, Er^{3+}), (V^{4+}, B^{3+}), (Fe^{3+}, Ti^{3+}), (Co^{2+}, Ti^{+}), (Bi^{3+}, Zn^{2+}), (As^{3+}, Dy^{3+}), (Ho^{3+}, Mg^{2+}), (K^{+}, Rb^{+}), (Cr^{3+}, Pr^{3+}), (Sr^{2+}, Fe^{2+}), (Ni^{2+}, Cu^{+}), (Sr^{2+}, Mo^{2+}), (Y^{3+}, Zr^{4+}), (Cd^{2+}, Ba^{2+}), (Ho^{3+}, Pb^{2+}), (Pb^{2+}, Li^{+}), (Eu^{3+}, Mg^{2+}), (Er^{3+}, Mg^{2+}), (Bi^{4+},Al^{3+}), (Ca^{2+}, Sm^{3+}), (V^{3+}, La^{3+}), (Gd^{3+}, Cr^{2+}), (Mn^{2+}, Ti^{+}), (Yb^{3+}, Fe^{2+}), (Ni^{2+}, Ag^{+}), (Zn^{2+}, Yb^{2+}), (Se^{4+}, Sn^{4+}), (Sb^{3+}, Bi^{2+}), and (Eu^{3+}, Pb^{2+}).

388. A method according to claim **375**, wherein said source of gaseous catalyst comprises a salt of one or more cations and at least one anion selected from the group consisting of halides, sulfates, phosphates, carbonates, hydroxide and sulfides.

389. A method according to claim **375**, wherein said gaseous catalyst comprises hydrogen atoms having a binding energy of about $E_b=13.6/n^2$ eV, where n is a fraction whose numerator is 1 and denominator is an integer greater than 1.

390. A method according to claim **375**, wherein said catalyst is selected to have a resonant adsorption with the energy released from said gaseous hydrogen atoms undergoing a transition to a lower energy state.

391. A method according to claim **375**, wherein said gaseous catalyst comprises potassium and has a net enthalpy of reaction of 27.28 eV.

392. A method according to claim **375**, wherein said gaseous catalyst has a net enthalpy of reaction of about 27.2 eV.

393. A method according to claim **375**, further comprising the step of combining a source of gaseous catalyst with a source of gaseous hydrogen atoms comprising at least one of a hydrocarbon or water, and providing combustion which volatizes said source of gaseous catalyst to form said gaseous catalyst and provide said gaseous hydrogen atoms.

394. A method according to claim **375**, further comprising the step of forming said source of gaseous catalyst or said gaseous catalyst in situ.

395. A method according to claim **344**, wherein said step of forming said source of gaseous catalyst or said gaseous catalyst in situ comprises the ionization of a reactant.

396. A method according to claim **345**, wherein said step of ionization comprises thermal ionization of said reactant.

397. A method according to claim **395**, wherein said step of ionization comprises chemical ionization of said reactant.

398. A method according to claim **397**, wherein said step of chemical ionization comprises oxidation or reduction of said reactant.

399. A method according to claim **375**, wherein said step of volatizing said source of catalyst to form said gaseous catalyst utilizes energy from at least one of heat, electron-beam energy, photon energy, acoustic energy, electric field, or magnetic field.

400. A method according to claim **375**, wherein said step of volatizing said source of catalyst comprises the step of heating a filament coated with said source of said gaseous catalyst.

401. A method according to claim **375**, wherein said step of volatizing said source of gaseous catalyst to form said gaseous catalyst comprises the steps of volatizing said material to form gaseous atoms and ionizing said gaseous atoms to form ions.

402. A method according to claim **375**, further comprising adding a source of catalyst to said reaction vessel and heating said source of catalyst to form said gaseous catalyst.

403. A method according to claim **375**, wherein said source of catalyst is provided in a catalyst reservoir comprising a container separate from said reaction vessel and said container communicates with said reaction vessel.

404. A method according to claim **375**, wherein said source of catalyst is provided in a boat contained within said reaction vessel.

405. A method according to claim **375**, wherein said step of providing hydrogen atoms comprises the steps of disassociating a hydrogen containing compound into hydrogen atoms.

406. A method according to claim **375**, wherein said step of providing hydrogen atoms comprises the steps of passing a hydrogen containing gas over a hot filament.

407. A method according to claim **375**, wherein said step of providing hydrogen atoms comprises the steps of passing a hydrogen containing gas over a hot grid.

408. A method according to claim **375**, wherein said step of providing hydrogen atoms comprises the steps of passing a hydrogen containing gas through a tungsten capillary heated by electron bombardment.

409. A method according to claim **375**, wherein said step of providing hydrogen atoms comprises the steps of maintaining a hydride under nonequilibrium conditions.

410. A method according to claim **375**, wherein said step of providing hydrogen atoms comprises the steps of passing a hydrogen containing gas through an inductively coupled plasma flow tube.

411. A method according to claim **375**, wherein said step of providing hydrogen atoms comprises the steps of contacting a hydrogen containing gas with a second catalyst for disassociating said hydrogen containing gas into free hydrogen atoms.

412. A method according to claim **411**, wherein said second catalyst comprises at least one element selected from the group consisting of transition elements, inner transition elements, precious metals, refractory metals, lanthanides, actinides and activated charcoal.

413. A method according to claim **411**, wherein said second catalyst is selected from the group consisting of an element, compound, alloy or mixture of transition elements, inner transition elements, iron, platinum, palladium, zirconium, vanadium, nickel, titanium, Sc, Cr, Mn, Co, Cu, Zn, Y, Nb, Mo, Tc, Ru, Rh, Ag, Cd, La, Hf, Ta, W, Re, Os, Ir, Au, Hg, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Vb, Lu, Th, Pa, U, activated charcoal, and intercalated Cs carbon.

414. A method according to claim **411**, further comprising the step of utilizing a hot filament or hot grid to disassociate a hydrogen containing gas into gaseous hydrogen atoms and to heat said second catalyst.

415. A method according to claim **411**, further comprising the step of controlling the power output of said cell.

416. A method according to claim **415**, wherein said step of controlling the power output of said cell comprises controlling the temperature of said second catalyst.

417. A method according to claim **416**, wherein said step of controlling the temperature of said second catalyst comprises utilizing a heated filament or grid.

418. A method according to claim **375**, wherein said source of gaseous hydrogen atoms is selected from the group consisting of hydrogen gas, water, hydrides, metal-hydrogen solutions, and hydrocarbons.

419. A method according to claim **375**, wherein said step of providing gaseous hydrogen atoms comprises pyrolyzing hydrocarbons or water.

420. A method according to claim **375**, further comprising the step of reforming hydrocarbons to at least one of gaseous molecular and gaseous atomic hydrogen.

421. A method according to claim **375**, further comprising the step of disassociating hydrogen containing molecules using UV light to form said gaseous hydrogen atoms.

422. A method according to claim **375**, further comprising the step of controlling the amount of gaseous hydrogen provided in said reaction vessel.

423. A method according to claim **375**, further comprising the step of controlling the flow of a source of gaseous hydrogen atoms or said gaseous hydrogen atoms from a chamber to said reaction vessel.

424. A method according to claim **375**, further comprising the step of utilizing a valve for controlling the flow of gaseous hydrogen or source of gaseous hydrogen from said reaction vessel.

425. A method according to claim **375**, further comprising the step of controlling the flow of gaseous hydrogen or source of gaseous hydrogen from said reaction vessel.

426. A method according to claim **375**, further comprising the step of utilizing a vacuum pump for controlling the flow of gaseous hydrogen or source of gaseous hydrogen from said reaction vessel.

427. A method according to claim **375**, further comprising the step of utilizing a valve for controlling the flow of gaseous hydrogen or source of gaseous hydrogen from a chamber into said reaction vessel.

428. A method according to claim **375**, further comprising controlling the partial pressure of said gaseous hydrogen atoms.

429. A method according to claim **375**, wherein a partial pressure of said gaseous hydrogen atoms or source of gaseous hydrogen atoms in said reaction vessel is maintained within the range of 10^{-3} atmospheres to 100 atmospheres.

430. A method according to claim **375**, further comprising controlling the amount of gaseous catalyst or source of gaseous catalyst introduced into said reaction vessel.

431. A method according to claim **375**, further comprising controlling the flow of gaseous catalyst or source of gaseous catalyst from a catalyst reservoir containing gaseous catalyst or a source of gaseous catalyst to said reaction vessel.

432. A method according to claim **375**, further comprising controlling the flow of gaseous catalyst or source of gaseous catalyst from a boat containing gaseous catalyst or a source of gaseous catalyst.

433. A method according to claim **375**, further comprising the step of controlling the flow of said gaseous catalyst or said source of gaseous catalyst from said reaction vessel.

434. A method according to claim **375**, further comprising the step of controlling the vapor pressure of said gaseous catalyst or source of gaseous catalyst in said reaction vessel.

435. A method according to claim **375**, wherein a vapor pressure of said gaseous catalyst or said source of gaseous catalyst is maintained at about 50 to 210 millitorr.

436. A method according to claim **375**, further comprising using a vacuum pump to control the flow of said catalyst or said source of gaseous catalyst from said reaction vessel.

437. A method according to claim **375**, further comprising using a valve to control the flow of said source of gaseous catalyst or said catalyst from a catalyst reservoir into said reaction vessel.

438. A method according to claim **375**, further comprising using a valve to control the flow of said source of gaseous catalyst or said catalyst from said reaction vessel.

439. A method according to claim **375**, further comprising the step of supplying a nonreactive gas to said reaction vessel.

440. A method according to claim **375**, further comprising the step of controlling the vapor pressure of said nonreactive gas in said reaction vessel.

441. A method according to claim **440**, further comprising the step of controlling the flow of said nonreactive gas supplied to said reaction vessel.

442. A method according to claim **440**, further comprising the step of controlling the amount of said nonreactive gas released from said reaction vessel.

443. A method according to claim **440**, further comprising utilizing a vacuum pump for controlling the flow of said nonreactive gas from said reaction vessel.

444. A method according to claim **440**, further comprising the step of utilizing a valve for controlling the flow of said nonreactive gas from said reaction vessel.

445. A method according to claim **440**, further comprising the step of utilizing a valve for controlling the flow of said nonreactive gas into said reaction vessel.

446. A method according to claim **375**, further comprising the step of controlling the vapor pressure of said gaseous catalyst in said reaction vessel.

447. A method according to claim **446**, wherein said step of controlling the vapor pressure of said gaseous catalyst comprises controlling the temperature in a catalyst reservoir containing a source of gaseous catalyst or said gaseous catalyst and being in communication with said reaction vessel, and controlling the flow of gaseous catalyst from said catalyst reservoir.

448. A method according to claim **446**, wherein said step of controlling the vapor pressure of said gaseous catalyst comprises controlling the temperature in a boat containing a source of gaseous catalyst or said gaseous catalyst and being contained in said reaction vessel, and controlling the flow of gaseous catalyst from said boat.

449. A method according to claim **375**, further comprising the step of controlling the temperature in a catalyst reservoir containing a source of gaseous catalyst or said gaseous catalyst and being in communication with said reaction vessel.

450. A method according to claim **375**, further comprising the step of controlling the temperature in a boat containing a source of gaseous catalyst or said gaseous catalyst and being contained in said reaction vessel.

451. A method according to claim **375**, wherein the reaction to provide a net enthalpy of about 27 (p/2) eV, where p is a positive integer greater than 1, comprises an electrochemical, chemical, photochemical, thermal, free radical, sonic, nuclear, inelastic photon, or particle scattering reaction, or mixtures thereof.

452. A method according to claim **375**, wherein a pressure in said reaction vessel is maintained within the range of 10^{-3} atmospheres to 100 atmospheres.

453. A method according to claim **375**, wherein said reaction occurring at a pressure less than atmospheric pressure.

454. A method according to claim **375**, further comprising the step of releasing hydrogen atoms from said reaction vessel having a binding energy of about $E_b=13.6/n^2$ eV, where n is a fraction whose numerator is 1 and denominator is an integer greater than 1 or a compound containing said hydrogen atoms.

455. A method according to claim **375**, further comprising the step of adsorbing said released energy.

456. A method according to claim **375**, further comprising the step of converting energy released from said hydrogen atom into electrical energy.

457. A method according to claim **375**, wherein said reaction step is conducted in an internal combustion chamber.

458. A method according to claim **457**, wherein said internal combustion chamber is an engine cylinder.

459. A method according to claim **375**, further comprising the step of controlling the power output of said cell.

460. A method according to claim **459**, wherein said step of controlling the power output of said cell comprises controlling the amount of said gaseous catalyst present in said reaction vessel.

461. A method according to claim **460**, wherein said step of controlling the amount of gaseous catalyst comprises controlling the temperature of said reaction vessel and selecting said gaseous catalyst to have a vapor pressure dependent upon the temperature of said reaction vessel.

462. A method according to claim **460**, wherein said step of controlling the amount of gaseous catalyst comprises controlling the temperature of a catalyst reservoir containing a source of gaseous catalyst or said gaseous and being in communication with said reaction vessel, and selecting said gaseous catalyst to have a vapor pressure dependent upon the temperature of said catalyst reservoir.

463. A method according to claim **460**, wherein said step of controlling the amount of gaseous catalyst comprises controlling the flow of said source of gaseous catalyst or gaseous catalyst from said catalyst reservoir into said reaction vessel.

464. A method according to claim **460**, wherein said step of controlling the amount of gaseous catalyst comprises controlling the flow of said source of gaseous catalyst or gaseous catalyst from said reaction vessel.

465. A method according to claim **460**, wherein said step of controlling the amount of gaseous catalyst comprises controlling the temperature of a boat containing a source of gaseous catalyst or said gaseous and being contained in said reaction vessel, and selecting said gaseous catalyst to have a vapor pressure dependent upon the temperature of said boat.

466. A method according to claim **459**, wherein said step of controlling the power output of said cell comprises controlling the amount of said gaseous hydrogen atoms or source of gaseous hydrogen atoms present in said reaction vessel.

467. A method according to claim **466**, wherein said step of controlling the amount of said gaseous hydrogen atoms or source of gaseous hydrogen atoms comprises controlling the flow of gaseous hydrogen atoms or source of gaseous hydrogen atoms into said reaction vessel.

468. A method according to claim **466**, wherein said step of controlling the amount of said source of gaseous hydrogen atoms or gaseous hydrogen atoms comprises controlling the flow of said source of gaseous hydrogen atoms or gaseous hydrogen atoms from said reaction vessel.

469. A method according to claim **466**, wherein said step of controlling the amount of said source of gaseous hydrogen atoms or gaseous hydrogen atoms comprises controlling the temperature of a second catalyst for dissociating a hydrogen containing compound into gaseous hydrogen atoms.

470. A method according to claim **469**, wherein said step of controlling the temperature of a second catalyst for dissociating a hydrogen containing compound into gaseous hydrogen atoms comprises controlling the power dissipated in a second catalyst heater.

471. A method according to claim **459** wherein said step of controlling the power output of said cell comprises controlling the amount of nonreactive gas in said reaction vessel.

472. A method according to claim **471**, wherein said step of controlling the amount of nonreactive gas in said reaction vessel comprises controlling the flow of nonreactive gas into said reaction vessel.

473. A method according to claim **472**, wherein said step of controlling the amount of nonreactive gas in said reaction vessel comprises controlling the flow of nonreactive gas from said reaction vessel.

474. A method according to claim **459**, wherein said step of controlling the power output of said cell comprises controlling the temperature of a source of gaseous catalyst.

475. A method according to claim **459**, wherein said step of controlling the power output of said cell comprises controlling the flow of a hydrogen containing gas over at least one of a hot filament, a tungsten capillary heated by electron bombardment, or an inductively coupled plasma flow.

476. A method according to claim **459**, wherein said step of controlling the power output of said cell comprises controlling the power dissipated in an inductively coupled plasma flow tube, hot filament or grid, or tungsten capillary heated by electron bombardment.

477. A method according to claim **459**, wherein said step of controlling the power output of said cell comprises controlling the temperature of a hot filament or tungsten capillary heated by electron bombardment over which a hydrogen containing gas flows.

478. A method according to claim **459**, wherein said step of controlling the power output of said cell comprises controlling the temperature of a hydride maintained under nonequilibrium conditions.

479. A method according to claim **375**, wherein a temperature in said reaction vessel is maintained at a higher temperature than in a catalyst reservoir in communication with said reaction vessel or a boat contained within said reaction vessel.

480. A method according to claim **375**, further comprising the step of measuring the temperature of a catalyst reservoir in communication with said reaction vessel or a boat contained in said reaction vessel.

481. A method according to claim **375**, further comprising the step of measuring the temperature of a source of said gaseous catalyst contained in a catalyst reservoir in communication with said reaction vessel or a boat contained within said reservoir.

482. A method according to claim **375**, further comprising the step of measuring the temperature of a chamber containing a source of said hydrogen atoms in communication with said reaction vessel.

483. A method according to claim **375**, further comprising the step of measuring the temperature of a source of said gaseous hydrogen atoms.

484. A method according to claim **411**, further comprising the step of measuring the temperature of said second catalyst.

485. A method according to claim **375**, further comprising the step of controlling the temperature of said reaction vessel.

486. A method according to claim **375**, further comprising utilizing a nebulizer or atomizer to form said gaseous catalyst.

487. A method according to claim **375**, further comprising the step of measuring the pressure in said reaction vessel.

488. A method according to claim **375**, further comprising the step of controlling the pressure in said reaction vessel.

489. A method according to claim **375**, further comprising the step of measuring the hydrogen pressure in said reaction vessel.

490. A method according to claim **375**, further comprising the step of measuring the gaseous catalyst pressure in said reaction vessel.

491. A method according to claim **375**, further comprising the step of converting energy released from said gaseous hydrogen atoms into electrical energy.

492. A method according to claim **375**, further comprising controlling the pressure of said gaseous catalyst by controlling the amount of said source of catalyst being added to said reaction vessel.

493. A cell comprising:

a reaction vessel capable of containing a vacuum or pressures greater than atmospheric;

a source of hydrogen atoms; and

a source of a gaseous catalyst or a gaseous catalyst capable of accepting energy from atomic hydrogen thereby catalyzing a transition of the electron of atomic hydrogen to a state lower than that of uncatalyzed hydrogen.

494. A cell comprising:

a reaction vessel capable of containing a vacuum or pressures greater than atmospheric;

a source of hydrogen atoms; and

a source of a gaseous catalyst or a gaseous catalyst capable of accepting energy from atomic hydrogen thereby catalyzing a transition of the electron of atomic hydrogen to a state lower than that of uncatalyzed hydrogen and releasing energy from said hydrogen atom, wherein said gaseous catalyst comprises hydrogen atoms having a binding energy of about $E_b=13.6/n^2$ eV, where n is a fraction whose numerator is 1 and denominator is an integer greater than 1.

495. A cell comprising:

a reaction vessel capable of containing a vacuum or pressures greater than atmospheric;

a source of hydrogen atoms; and

a source of a gaseous catalyst or a gaseous catalyst capable of accepting energy from atomic hydrogen thereby catalyzing a transition of the electron of atomic hydrogen to a state lower than that of uncatalyzed hydrogen and releasing energy from said hydrogen atom, wherein said source of gaseous catalyst comprises at least one salt selected from the group consisting of RbF, RbCl, RbBr, RbI, Rb_2S_2, RbOH, Rb_2SO_4, Rb_2CO_3, Rb_3PO_4, KF, KCl, KBr, KI, K_2S_2, KOH, K_2SO_4, K_2CO_3, K_3PO_4, and K_2GeF_4.

496. A cell comprising:

a reaction vessel capable of containing a vacuum or pressures greater than atmospheric;

a source of hydrogen atoms; and

a source of a gaseous catalyst or a gaseous catalyst capable of accepting energy from atomic hydrogen thereby catalyzing a transition of the electron of atomic hydrogen to a state lower than that of uncatalyzed hydrogen and releasing energy from said hydrogen atom, wherein said gaseous catalyst comprises at least one ion selected from the group consisting of Mo^{2+}, Ti^{2+}, and Rb^+.

497. A cell comprising:

a reaction vessel capable of containing a vacuum or pressures greater than atmospheric;

a source of hydrogen atoms; and

a source of a gaseous catalyst or a gaseous catalyst capable of accepting energy from atomic hydrogen thereby catalyzing a transition of the electron of atomic hydrogen to a state lower than that of uncatalyzed hydrogen and releasing energy from said hydrogen atom, wherein said source of gaseous catalyst comprises at least one metal selected from the group consisting of Mo, Ti, and Rb.

498. A cell comprising:

a reaction vessel capable of containing a vacuum or pressures greater than atmospheric;

a source of hydrogen atoms; and

a source of a gaseous catalyst or a gaseous catalyst capable of accepting energy from atomic hydrogen thereby catalyzing a transition of the electron of atomic hydrogen to a state lower than that of uncatalyzed hydrogen and releasing energy from said hydrogen atom, wherein said source of gaseous catalyst comprises at least one salt selected from the group consisting of MoI_2, $TiCl_2$, $TiCl_4$, $SnCl_4$, $SiCl_4$, $PrBr_3$, $CaBr_2$, $SrCl_2$, CrI_2, TbI_3, $SbCl_3$, CrF_3, $CoCl_2$, $BiCl_3$, $NiCl_2$, PdF_2, InCl, $LaCl_3$, $DyCl_3$, LaI_3, HoI_3, KNO_3, VF_3, PbF_2, VOCl, PbI_2, $LuCl_3$, $PbCl_2$, AsI_3, HoI_3, $MoCl_5$, $SnCl_4$, SbI_3, CdI_2, AgF_2, AgF, LaI_3, ErI_3, VCl_4, BCl_3, $FeCl_3$, $TiCl_3$, CoI_2, CoF_2, TlI, TlF, $BiBr_3$, $ZnBr_2$, AsI_3, DyI_3, $HoCl_3$, $MgCl_2$, $CrCl_3$, $PrCl_3$, $SrCl_2$, $FeCl_2$, $NiCl_2$, CuCl, $SrCl_2$, $MoCl_2$, YCl_3, $ZrCl_4$, CdI_2, BaI_2, HoI_3, PbI_2, PdF_2, LiF, $EuCl_3$, $MgCl_2$, $ErCl_3$, $MgCl_2$, $ErCl_3$, $MgCl_2$, $BiCl_4$, $AlCl_3$, $CaBr_2$, $SmBr_3$, VaF_3, $LaCl_3$, GdI_3, CrI_2, MnI_2, $YbBr_3$, $FeBr_2$, $NiCl_2$, AgCl, $ZnCl_2$, $YbCl_2$, SeF_4, $SnCl_4$, SnF_4, SbI_3, BiI_2, EuF_3, and $PbCl_2$.

499. A cell comprising:

a reaction vessel capable of containing a vacuum or pressures greater than atmospheric;

a source of hydrogen atoms; and

a source of a gaseous catalyst or a gaseous catalyst capable of accepting energy from atomic hydrogen thereby catalyzing a transition of the electron of atomic hydrogen to a state lower than that of uncatalyzed hydrogen and releasing energy from said hydrogen atom, wherein said gaseous catalyst comprises at least one pair of ions selected from the group consisting of: (Sn^{4+}, Si^{4+}), (Pr^{3+}, Ca^{2+}), (Sr^{2+}, Cr^{2+}), (Cr^{3+}, Tb^{3+}), (Sb^{3+}, Co^{2+}), (Bi^{3+}, Ni^{2+}), (Pd^{2+}, In^+), (La^{3+}, Dy^{3+}), (La^{3+}, Ho^{3+}), (K^+, K^+), (V^{3+}, Pd^{2+}), (Lu^{3+}, Zn^{2+}), (As^{3+}, Ho^{3+}), (Mo^{5+}, Sn^{4+}), (Sb^{3+}, Cd^{2+}), (Ag^{2+},Ag^+), (La^{3+}, Er^{3+}), (V^{4+}, B^{3+}), (Fe^{3+}, Ti^{3+}), (Co^{2+}, Ti^+), (Bi^{3+}, Zn^{2+}), (As^{3+}, Dy^{3+}), (Ho^{3+}, Mg^{2+}), (K^+, Rb^+), (Cr^{3+}, Pr^{3+}), (Sr^{2+}, Fe^{2+}), (Ni^{2+}, Cu^+), (Sr^{2+}, Mo^{2+}), (Y^{3+}, Zr^{4+}), (Cd^{2+}, Ba^{2+}), (Ho^{3+}, Pb^{2+}), (Pb^{2+}, Li^+), (Eu^{3+}, Mg^{2+}), (Er^{3+}, Mg^{2+}), (Bi^{4+}, Al^{3+}), (Ca^{2+}, Sm^{3+}), (V^{3+}, La^{3+}), (Gd^{3+}, Cr^{2+}), (Mn^{2+}, Ti^+), (Yb^{3+}, Fe^{2+}), (Ni^{2+}, Ag^+), (Zn^{2+}, Yb^{2+}), (Se^{4+}, Sn^{4+}), (Sb^{3+}, Bi^{2+}), and (Eu^{3+}, Pb^{2+}).

* * * * *

rexresearch.com

Boyd BUSHMAN

Magnetic Beam Apparatus

Generates a magnetic monopole beam which emits pulses, levitates, degausses, stops electronics and separates materials

---

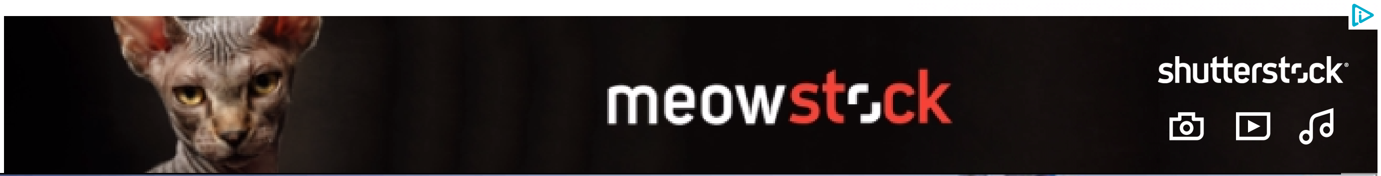

meowstock
shutterstock

WE'LL MATCH A LOWER PRICE PLUS GIVE YOU 25% OFF. ONLY AT HILTON.COM
Hilton
BOOK NOW
Terms and conditions apply
Roll over for brands.
Ad

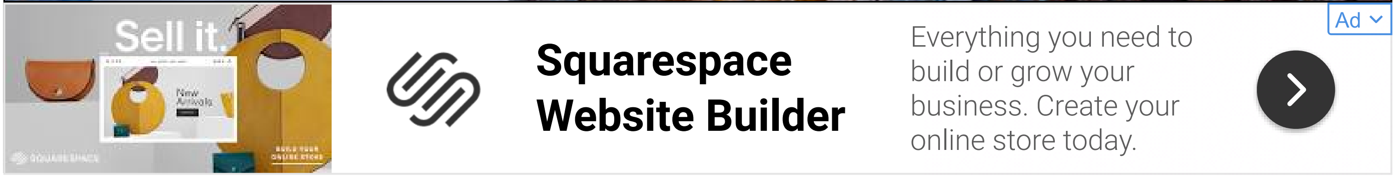

Sell it.
Squarespace Website Builder
Everything you need to build or grow your business. Create your online store today.
Ad

---

https://web.archive.org/web/20130608005034/http://boydbushman.com/Home.php

CLASSIFICATIONS: PROGRAM MANAGER & SENIOR SCIENTIST
(TOP SECRET CLEARANCE)

EXPERIENCE:

Lockheed Martin Aeronautics Company, 1986 to 2000 (and present consultant), Mr. Bushman has participated in Advanced Aircraft Concepts as a member of the technical team. He has conducted European Theater Tactical Fighter performance analysis, coordinated customer interface and development performance concepts. He developed laser, electro-optical sensor, and power systems. He developed and taught analysis courses. Several million-dollar contracts were obtained directly from him. Bushman holds 26 US Patents and classified Patents.

Texas Instruments, 1979-1987, Mr. Bushman worked on the development of laser guided smart bombs, Tank Mounted Infrared FLIR systems, focal plane arrays and GPS Systems all of which are currently deployed as major components of Tactical Land Air and Sea systems.

From 1976 to 1979 Mr. Bushman was a Manager at Parsons International of Iran, Mexico and then Louisiana with the U.S. department of Energy's Strategic Oil Reserve Program. He managed a division of Trans Universal Finance Company in Southern California. Funding was obtained for high multimillion dollar projects.

At Hughes Aircraft (Now Ratheon), 1968-1976, Mr. Bushman was the Manager/Technical Liaison officer to the Infrared Sensor Division (SBRC Santa Barbara Research Center). He was involved in technical description and financial disbursement. Systems were deployed to aircraft and satellite systems.

At Electro-Optical Systems (Now Ratheon), 1966-1968 as Program manager, he developed and produced Night Vision Image Intensifier Systems for military night operations. These are similar to night vision goggles used by pilots. Mr. Bushman achieved sustained production of hundreds of units per month.

1963-1966, General Dynamics (Pomona), he coordinated the analysis, test and development producing the Redeye (Now Stinger Missile). The development and testing was achieved in concert with Sidewinder Missile Development team at China Lake Test Facility in California. Mr. Bushman coordinated the computer based operations analysis and system performance analysis for missile development and manufacture.

---

https://groups.yahoo.com/neo/groups/antigravity1/conversations/topics/149
Mar 10, 2008

Magnetic Beam Amplifier

Here's more info about Bushman's Magnetic Beam Amplifier. It's interesting that it can provide a conductive channel in the air to produce 5-foot long sparks from high voltage. A long time ago I made and tested the unbalanced capacitor (It consists of parallel conducting plates, with a plurality of pins or points on one set of plates, facing the opposite plate, and separated by a dielectric, and air gap.). It didn't work, but I've just found that I connected it incorrectly, so I'll be testing it again.

http://web.archive.org/web/20040602192943/http://www.hometown.aol.com/dmboss1021/Page300.html

"An interesting effect he mentions is to have two such devices aimed at each other, and separated by up to five feet, with one having all south poles inward, and the other all north poles inward. And place a high voltage source adjacent to one beam source (such as a Van De Graff generator - set to cause sparks to jump 3/4" - meaning approx 7,500 volts in standard air) that there will be an arc extending along this beam for the entire five feet, to the other beam device!

I find this claim to very interesting, as the PE theory, indicates this energy is responsible for matter formation, and gravity, and that there is implied in the theory a method to generate electrical power directly from a "gravity beam".

This is due to the Tendency for nuclei and electrons to remain in their respective nodes and antinodes of this PE, and a preferential drift of these nodes/antinodes results in what we observe as the attraction of gravity. If you have some free electrons in a conductor or in a semi or quasi conductor - they would be influenced by this anti nodal drift.

If the mass of this material (either conductor or quasi conductor) is not free to move under the influence of this nodal/anitnodal drift (gravity), there would then be a tendency to produce an electric charge separation, as the free electrons would be induced to drift. This charge separation would result in an excess negative charge in the direction of the gravity - for example: the preferential drift near the earth is towards the planet, and thus one should expect that free electrons in the atmosphere (a quasi conductor and dielectric) would drift towards the planet under this influence.

The bulk mass of the atmosphere is pushed towards the planet, but reaches an equilibrium from the "pressure" of the air molecules against each other. The free electrons are not necessarily subject to this same equilibrium, and would continue to drift towards the planet. This results in the natural (and not fully explained by conventional science) electric field in the atmosphere!

This mechanism may explain the natural electric field of the atmosphere, as being due to gravity. And it follows that if one can create an isolated gravity beam of this nodal drift - that if you fix a conductor, or quasi conductor in this beam, that a charge separation will occur. You could then harness this energy directly as electrical current.

So if this magnetic beam amplifier is actually a concentrated PE beam, then perhaps this anomalous arcing effect is due to the gravitic nature of the "beam", in a similar manner as noted above. "

Jaro

---

http://alienopacity.com/boyd-bushman-former-of-the-skunk-works/
December 09, 2012

Boyd Bushman Formerly of the SKUNK Works

Seriously do you really need to know a sixty year old secret? I think it is time for us to know the truth. We flew out first anti gravity craft in the fifties a decade after the Roswell crash. Yeah it was tethered to a power cable but we are way beyond that first working model today in 2012. Boyd Bushman has built the Apparatus and Method for amplifying a magnetic beam – US Patent Issued on July 27, 1999. The technology involves the careful spinning of the liquid metal Mercury. In this video Boyd Bushman demonstrates the spinning CELT I think he called it. Its about the three thirty minute mark. It is an incredible display of energy and when you couple that with mercury you are able to generate huge magnetic forces and propulsion. His experiments on the Hutchinson effect show that magnetism can cancel gravity. Basically he took two rocks one with a pair of super magnets bolted together inside the rock and the other without the magnets. The one with bolted super magnets produced a three foot radius magnetic field around the rock. that is the increases the repulsive force. The result of dropping these two items from some height within the gravitational pull of the earth was that the rock with the magnets arrived later. This is incredibly significant in the world pf physics and shows the link between magnetic and gravity. An incredibly important step in finding the unified field theories that Relativity have somewhat led us away from. Boyd Bushman was a major leading player in the work done at the SKUNK Works over the last few decades. This is another high ranking official trying to release the skepticism in Americans so we can open our borders and our minds to the Visitors. Boyd Bushman confirms that we have nuclear powered magnetic propulsion craft that can reach heights of 60,000 feet and operate with distance specifications in the thousands of miles. Once again there are no Aliens because America does not believe in them and the true Exo-Political party in charge is not yet ready to take a step forward. Sixty years of darkness is too long already. Most Americans are UFO skeptics, have not really researched what is going on and don't want to be laughed at. I must be crazy but I think this is the scientist who created the craft and is describing parts of the technological road map from finding the craft, to measuring the craft dynamics, duplicating the dynamics and prototyping working models. Less than 10K hits on youtube for a man who built the magnetic propulsion technology. Then again I am not a skeptic anymore.

---

http://www.leedskalnin.com/

Magnetic Beam Amplifier

Neodymium, brass, nickel, stainless, copper. Variation of Lockheed Martin patent by Boyd Bushman. #5,929,732.
7"W x 7"H x 4"D

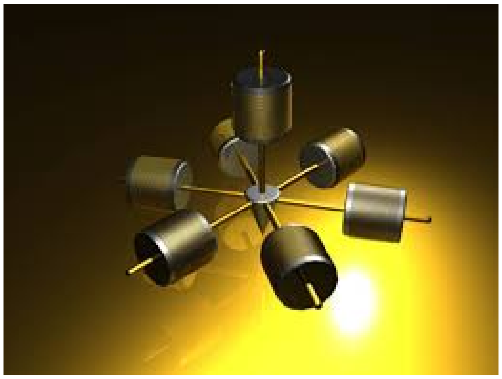

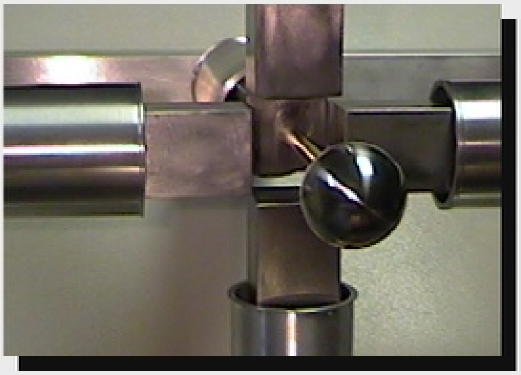

---

http://www.warriormatrix.com/about4003.html&sid=51cfaf396a72c42becd2bfefce3f5b03

FIG. 2

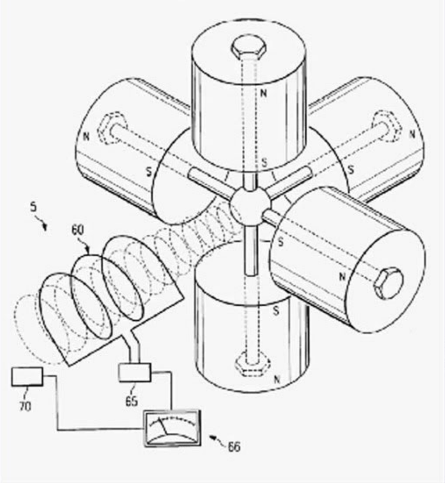

FIG. 3

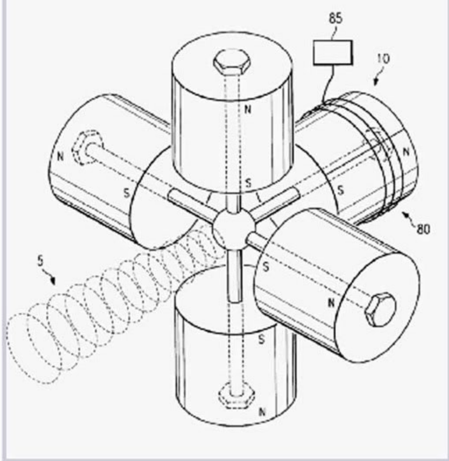

This is the Beamer Patent Drawing

According to the patent, this configuration creates a beam of some kind, represented by the series of dotted line circles extending out from the device. The author claims it is magnetic energy, but my own magnetic field analysis, and experiments show the magnetic field, while enhanced in this direction, does not appear to extend out to five feet as is claimed in the patent. (the magnetic flux does not appear to go this far at any appreciable strength, but a subtle energy effect does!)

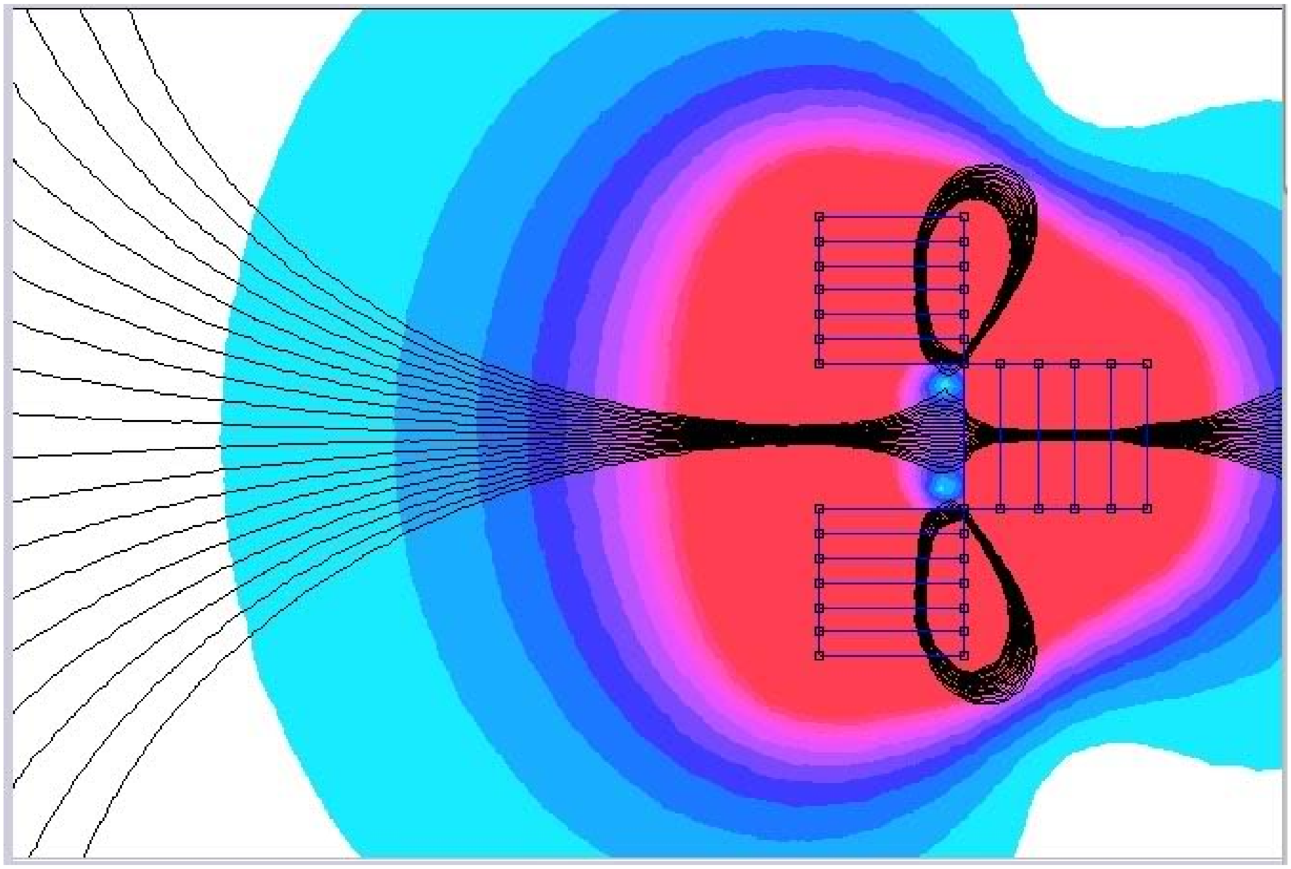

This is a plot of the magnetic field given off by a Beamer

The unusual properties of this beam, as described by the author, are very similar to the properties of a Primary Energy beam. And as the inert gases do exist as natural components of air, perhaps this beam he speaks of is a concentrated PE beam? (note in the next chapter a more powerful "Beamer" device is constructed, and it has a PE emanation all the time - and more so when you insert inert gas tubes into the "beam" area.)

He further claims that if you place a conductive coil (60) in the beam, and connected in series with a special unbalanced capacitor (65), and a voltmeter; that a voltage is produced which will vary if any object is in the beam, or any movement occurs in the beam. (yet to be tested)

Alternatively, if you place just the special capacitor in the beam (70),it will register a voltage if any movement occurs in the beam.

He further claims that if you place a coil over the unopposed magnet (80), and apply an ac current at various frequencies, you can modulate this beam, causing various effects at a distance.

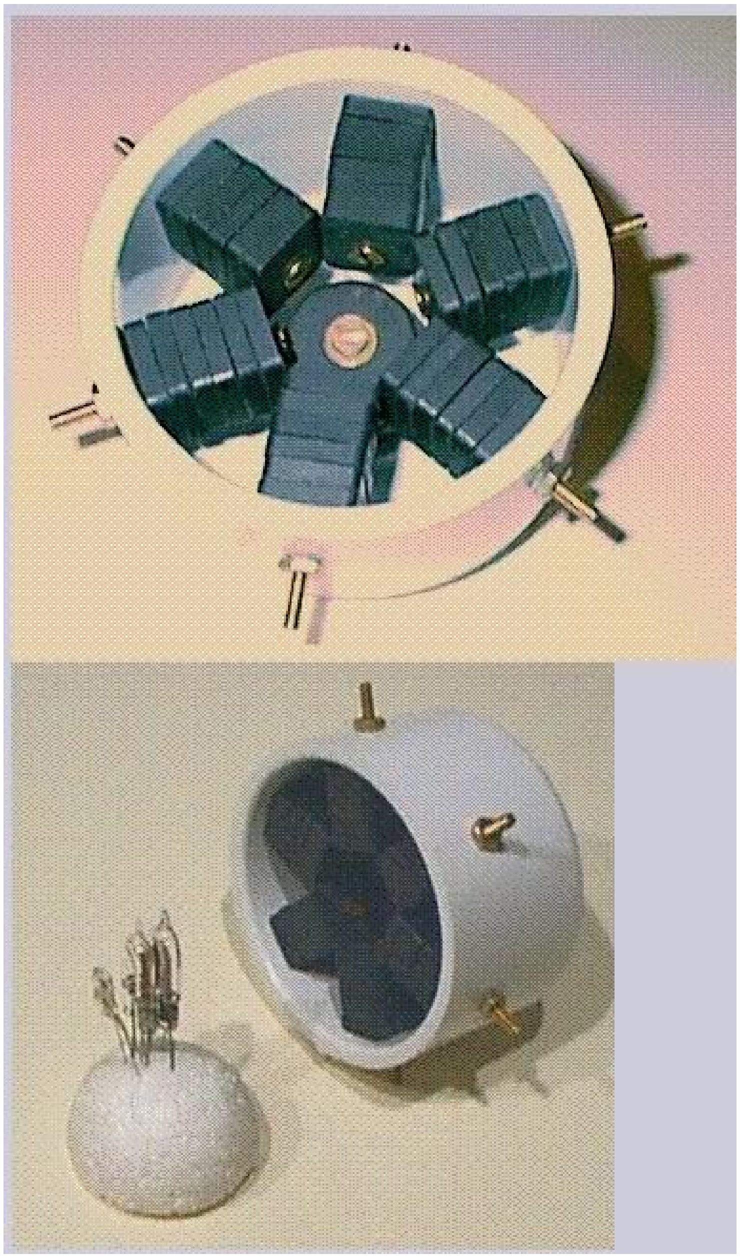

This is an experimental Beamer

I hope this will be interesting and of some use perhaps

Neil

---

https://www.youtube.com/watch?v=b2mmen7r92I

Replicating the Boyd Bushman Gravitational Experiment
by William Alek and Michael Ellegion

"Could gravity be related to its cousin magnetism?" Boyd Bushman's experiment is performed by others with similar results :

(SNNS) Opposing magnets inside the rock cause it to fall slower.

(NSNS) Attracting magnets inside the rock cause it to fall faster.

US5929732
Apparatus and method for amplifying a magnetic beam

Inventor(s): BUSHMAN BOYD B

An apparatus and method for creating a magnetic beam wherein a focusing magnet assembly (45) is comprised of a first opposing magnet pair (20) and a second opposing magnet pair (30) disposed in a focusing plane, each magnet of the respective opposing magnet pairs having a like pole directed towards the geometric center of the focusing magnet assembly (45) to form an alignment path, two like magnetic beams extending from the alignment path on each side of the focusing magnet assembly (45), each beam being generally perpendicular to the focusing plane. A like pole of an unopposed magnet (10) can be directed down the alignment path from one side of the focusing magnet assembly (45) to produce a single magnetic beam extending generally perpendicular from the focusing magnet assembly opposite unopposed magnet (10). **This beam is a magnetic monopole which emits pulses, levitates, degausses, stops electronics and separates materials.**

TECHNICAL FIELD OF THE INVENTION

The present invention relates generally to the field of magnetic devices, and more particularly to focusing a magnetic field to increase the efficiency of a magnet.

BACKGROUND OF THE INVENTION

Magnets have a variety of commercial uses. For instance, magnets perform a variety of physical tasks by attracting ferric materials, and magnets also perform a variety of electrical tasks, such as creating an electric current in a generator or alternator or enhancing the performance of electromagnetic devices such as inductors. However, the creation of a strong magnetic field to perform these functions generally entails the use of a large permanent magnet or a powerful electromagnet. Such magnets are often physically large and heavy, and tend to produce wide magnetic fields which can interfere with equipment near the magnet. Thus, commercial devices that include magnets must be designed around the physical size of the magnet needed to produce a desired magnetic field and around the width of the magnetic field produced.

SUMMARY OF THE INVENTION

Therefore, a need has arisen for an apparatus and method for manipulating and focusing the magnetic field of a magnet to reduce the size and strength required for a given magnet to perform a function.

In accordance with the present invention, an apparatus and method is provided that substantially eliminates or reduces disadvantages and problems associated with previously used unfocused magnets. The present invention uses opposing magnetic pairs to direct a magnetic field along an alignment axis.

More specifically, the present invention focuses a magnetic beam along an alignment axis by providing a focusing magnet assembly having at least three opposing magnets spaced apart in a focusing plane, the space between the opposing magnets forming an alignment path, and by further providing an unopposed magnet generally directed along the alignment axis towards the alignment path. Each opposing magnet has a like pole facing across the alignment path. The alignment axis corresponds to the alignment path formed by the spacing between the focusing assembly of magnets. In one alternative embodiment, the unopposed magnet can be placed in the focusing plane to form opposing magnet pairs which direct a magnetic beam from the alignment path along the alignment axis from each side of the focusing assembly.

In one alternative embodiment, a coil of conducting material is disposed in the magnetic beam near the exit of the beam from the focusing magnet assembly. An unbalanced capacitor and a voltmeter connected in series with the coil allow the detection of movement through the magnetic beam by measuring variations in the AC voltage of the coil. Alternatively, AC voltage can be measured from a pin array disposed in the beam as a substitute for the coil and unbalanced capacitor. In yet another embodiment, electric power, such as an alternating current, can be passed through coils arranged around one or more of the magnets of the assembly to allow the generation of a magnetic pulse.

The focused magnetic beam or beams produced by the present invention provides important technical advantages by focusing a magnetic field. The focused field may perform functions using less total magnetic force than is generally used by an unfocused magnet. This allows for the use of smaller permanent magnets, or alternatively, electromagnets that use lower amounts of electric power. Thus, a commercial device built with magnets according to the present invention **can be built smaller and with less of a need for power consumption** than devices using standard magnets. For instance, the present invention **can be used to more effectively separate ferric materials from non-ferric materials**, and can more effectively **induce propulsion or levitation** in systems that use magnetic fields, such as electric motors or levitating trains that use opposing magnetic fields. The present invention can also interact with electromagnetic devices such as computers by **stopping the operation of analog circuits from a distance, degaussing and otherwise interfering with digital circuits and memories, and emitting high powered electromagnetic pulses (EMP) at such circuits when induction coils are energized around the magnets with an alternating current. The present invention can also direct or transport electromagnetic charges through a distance.**

BRIEF DESCRIPTION OF THE DRAWINGS

A more complete understanding of the present invention and advantages thereof may be acquired by referring to the following description taken in conjunction with the accompanying drawings in which like reference numbers indicate like features and wherein:

FIG. 1 depicts an assembly of five magnets to produce a magnetic beam;

FIG. 2 depicts a magnetic beam adapted to detect motion; and

FIG. 3 depicts an electronic coil disposed around an opposing magnet for producing a magnetic pulse beam.

FIG. 1

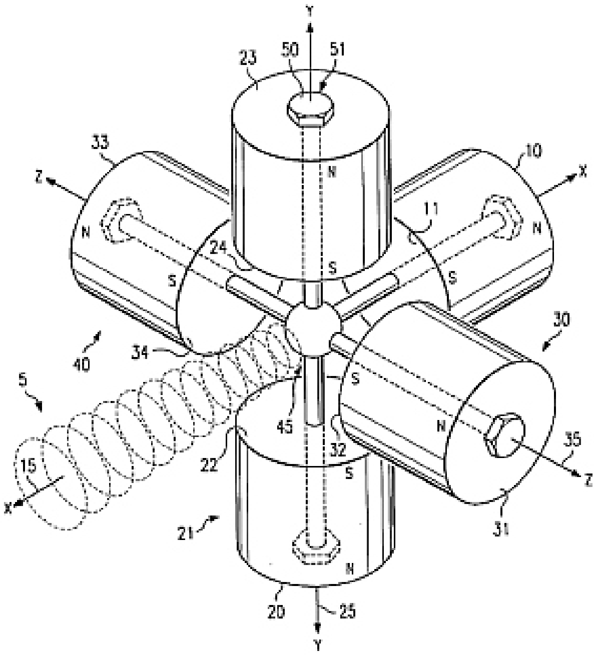

FIG. 2

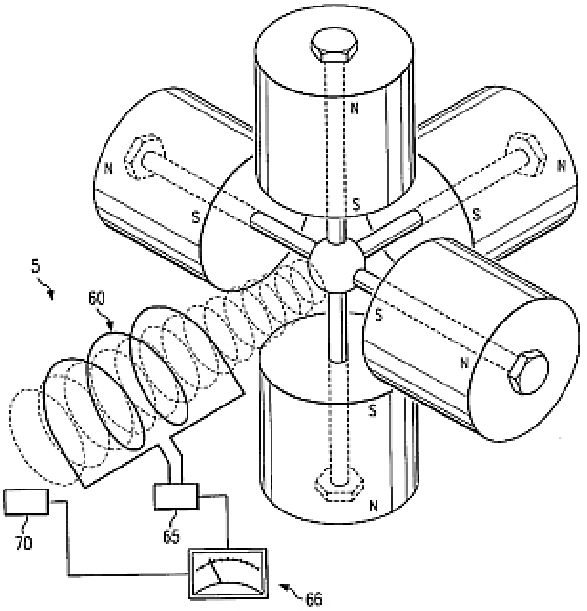

FIG. 3

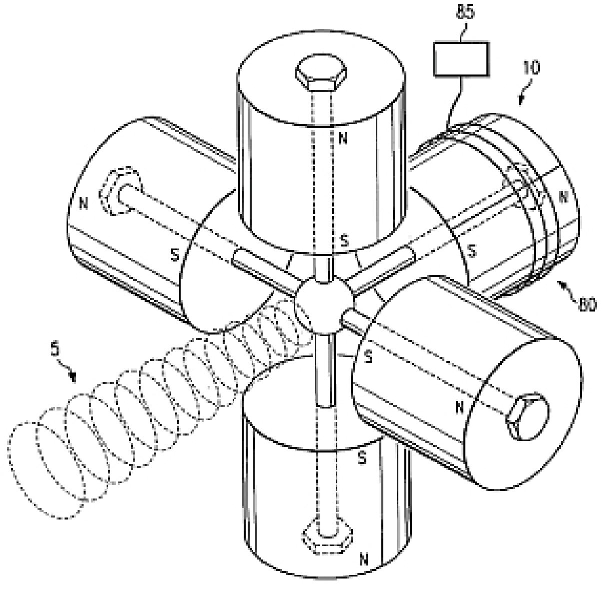

DETAILED DESCRIPTION OF THE INVENTION

Preferred embodiments of the present invention are illustrated in the figures, like numerals being used to refer to like and corresponding parts of the various drawings.

Referring now to FIG. 1, one embodiment of the present invention is depicted creating a south magnetic beam 5. A single unopposed magnet 10 is aligned along an alignment axis 15 so that the south pole 11 of magnet 10 points substantially in the direction of the intended magnetic beam. The alignment axis 15 in FIG. 1 is labeled as the "X" axis. A first opposing magnet pair 20 is disposed along a first or "Y" axis 25 substantially perpendicular to the alignment axis 15. Opposing magnet pair 20 is comprised of two magnets, 21 and 23, substantially aligned along the "Y" axis 25 so that the south pole 22 of magnet 21 and the south pole 24 of magnet 23 face each other. Magnets 21 and 23 are arranged to have a space between them which coincides with the path of alignment axis 15.

A second opposing magnet pair 30 is disposed along a second or "Z" axis 35, which is substantially perpendicular to the alignment axis 15 and also substantially perpendicular to the "Y" axis 25, the first and second magnet pairs forming a focusing magnet assembly 40. Second opposing magnet pair 30 is comprised of two opposing magnets, 31 and 33, each substantially aligned along the "Z" axis 35 so that the south pole 32 of magnet 31 faces the south pole 34 of magnet 33. Magnets 31 and 33 are arranged with a space between them, the space coinciding with the path of the alignment axis 15. The spaces of first and second opposing magnet pairs 20 and 30 coincide to form an alignment path 45 through which the alignment axis 15 is disposed. The first axis 25 and second axis 35 intercept within alignment path 45 at the approximate geometric center of the magnet assembly to form a focusing plane.

In the embodiment depicted by FIG. 1, each magnet has approximately the same magnetic force, and each magnet is a niosyum magnet constructed in the form of a cylinder with a hole 50 down its center. Nonferric bolts 51 are inserted through each hole and intersect at the geographic center of the magnetic pairs, where the Nonferric bolts can be coupled to hold the magnets in place. In alternative embodiments, any other convenient means of coupling the magnets in place can be used, and alternative means of producing a magnetic force, such as electromagnets, can be used. In another embodiment, unopposed magnet 10 can be removed so that a magnetic beam extends along the alignment axis 15 from each side of the focusing plane.

The embodiment of FIG. 1 depicts a like pole of each magnet directed towards the geographic center of the focusing magnet assembly 40. In FIG. 1, the south pole of each magnet is depicted as directed at the alignment path 45 and at the geographic center of the focusing magnet assembly 40. In other embodiments, the north pole of each magnet can be directed at the geographic center. In either case, the coupling of like poles in a plane and the introduction of a like pole perpendicular to the plane acts to focus the magnetic fields of each magnet into a beam extending perpendicular from the focusing plane along the alignment axis 15 of the single unopposed magnet 10. Thus, an unopposed magnet that would ordinarily have a measurable magnetic field of four to six inches extending outward from a pole can have a beam extending five to seven feet outward from the pole. The focused magnetic beam 5 extends along the alignment axis 15 beyond the magnetic field that would occur if each magnetic field of the beam assembly were combined into a single magnetic field, such as would happen if each of the magnets were combined end-to-end. The direction and intensity of the beam can be altered by changing the alignment of the single opposing magnet 10, or by changing the position or strength of the magnets in the focusing magnet assembly relative to the focusing plane, or by changing the shape or configuration of individual magnets. Thus, for instance, the beam could be made to rotate around alignment axis 15 by sequentially altering the strength of each magnet in the focusing magnet assembly.

The increased magnetic field of the embodiment depicted in FIG. 1 can be demonstrated by its physical effects. For instance, an unopposed magnet 10 that can lift a ferric object at a distance of approximately one inch can lift the same ferric object through a distance of approximately six inches when coupled to a magnetic assembly comprised of two pairs of opposing magnets as is depicted in FIG. 1. The same assembly **can cause cathode tube distortions at a distance of approximately two feet, can make electronic systems inoperable at a distance of approximately six inches, and can double the performance of an inductor at a distance of approximately eight inches.**

In another example of the physical effects of an embodiment such as that depicted in FIG. 1, a VanDeGraaff generator with a spark potential of approximately three quarters of an inch from its bell can be placed between an assembly of north opposing magnets and an assembly of south opposing magnets, the north opposing magnets located eight inches from the bell, and the south opposing magnets located five feet from the bell, with the bell located one foot from its base. This configuration will produce a large spark from the VanDeGraaff to the south opposing magnets, indicating the **potential for the present invention to perform electron transfer through a distance.**

In alternative embodiments, various numbers of magnets or opposing magnet pairs can be arranged around the alignment axis 15 to focus the magnetic beam 5. For instance, three opposing magnets could be used instead of the four opposing magnets depicted in FIG. 1. Alternatively, a greater number of magnets or magnet pairs can be arranged offset to the focusing plane to alter the focus or strength of the magnetic beam 5. In yet another embodiment, a plurality of focusing magnet assemblies can be disposed along the alignment axis in a plurality of focusing planes.

Referring now to FIG. 2, two alternative embodiments of the present invention are depicted. First, a coil 60 comprised of a conducting material, such as copper, is disposed within a magnetic beam 5. An unbalanced capacitor 65 is coupled in series with the coil 60 and a voltmeter 66. Unbalanced capacitor 65 can be an electrical device such as the electrical device disclosed in U.S. patent application Ser. No. 08/144,814, by Boyd Bushman, abandoned, now issued as a continuation-in-part as **U.S. Pat. No. 5,637,946**, which is incorporated herein by reference. Voltmeter 66 is any conventional voltmeter capable of measuring an electric current, such as an alternating current. Movement, and even just the presence of objects, through the beam produces an AC voltage in coil 60 which can be detected by unbalanced capacitor 65 and measured by voltmeter 66.

Alternatively, a pin array 70 disposed in the magnetic beam 5 can also be used to measure disturbances, such as movement through the magnetic beam, which cause an AC voltage in pin array 70 that can be measured by voltmeter 66. Pin array 70 is any device such as the device disclosed in U.S. patent application Ser. No. 08/144,814, by Boyd Bushman, abandoned, now issued as a continuation-in-part as U.S. Pat. No. 5,637,946, which is incorporated herein by reference. Alternatively, any other means of measuring a disturbance through a magnetic field can be used.

Referring now to FIG. 3, a coil 80 comprised of a conducting material, such as copper, is depicted disposed around single unopposed magnet 10. An electric power source, such as an AC generator 85 supplies an alternating current to coil 80 at a variety of frequencies. The strength of the magnetic beam 5 varies according to the frequency of the alternating current through coil 80 due to the magnetic field created by coil 80. Thus, a magnetic pulse beam is generated according to the frequency of the current passing through coil 80. When the frequency of the AC current is high, such as ten to the fifteenth power hertz, the magnetic pulse emitted simulates the effect of a destructive electromagnetic pulse. In alternative embodiments, coils can be disposed around each magnet or around the magnet assembly as a whole to create a magnetic field and pulse according to the frequency of a current applied to the coil.

The present invention provides several technical advantages related to the application of a magnetic field. For instance, by focusing a magnetic field, the present invention allows for a decrease in the size of magnets used in commercial devices. Thus, in commercial applications such as where a magnet is used to increase the efficiency of an inductor, the present invention allows for greater miniaturization with smaller magnets than would be allowed with the magnetic field of an unfocused magnet. Where the present invention is used with electromagnets, such as for the lifting of ferric materials, a smaller electric power source can be used to create a lifting magnetic force.

Other Patents :

US6028434 -- 100% Test Computer Printed circuits emit ultraviolet frequencies which identify its correct or defective operation. Only 1% to 10% are now tested.

US5999652 -- THREAT WARNING - Detection of aircraft and missile plume.

US5384802 -- Laser Apparatus

US5982180

Metal detection system and process using a high voltage to produce a visible electrical discharge

The apparatus includes an electrical circuit for producing a high voltage at a high frequency with very little current and an elongated and electrically insulated electrical conductive member coupled to the circuit for producing an electrical discharge when in close proximity to a metal object in the ground. The electrical circuit is capable of producing 50,000 volts and higher. In using the apparatus, the electrical conductive member may be moved to scan the ground over a metal object to obtain an image of the shape of the metal object.

US5420588
Wave Attenuation

A radar cross-section reduction system utilizes a charging device for creating a high positive DC static potential on an object such as an aircraft. The charging system is mounted to the object so that this DC potential would be applied to the outer skin or surface of the object. The charging system includes Van de Graaf generators located in housings attached to the object. Each Van de Graaf generator has two pulleys about which a belt is rotated. One of the pulleys and the belt are nonconductors, while the other pulley is a conductor. This results in a buildup of positive charges at one end of the belt and negative charges at the other end. The negative charges are picked off and applied to a ground section. The positive charges are picked off and applied to a collector. The collector is in electrical continuity with the outer structure of the object. Negative dissipators attached to the ground section dissipate the excess negative charges during flight. Positive dissipators may be selectively switched into engagement with the skin to dissipate the positive charge on the skin prior to landing.

US5542247
Apparatus Powered Using Laser Supplied Energy

A method and apparatus are providing for converting energy into thrust, and directing the thrust to move an object. The apparatus includes a chamber having air disposed therein, a pulsed laser for converting an energy source into light pulses, and a lens for receiving the light pulses and directing the light pulses toward a focal point within the chamber. Each light pulse converges in a region which is proximate to the focal point and causes molecules within the air which are at the region to disassociate. Disassociation of the molecules generates pressure waves which provide thrust for powering the object to move.

US5680135
Radiation Communication System

An aircraft has an exhaust flame or plume which can be modulated to communicate. A sound emitter is mounted to the aircraft for emitting acoustic waves into the exhaust plume. An encoder will control the emitter at selected digital sequence to provide a digital message. The flame or plume will radiate at a frequency range of interest depending on the type of aircraft. The sound waves cause the frequencies to change from a continuous spectrum to a spectrum which has a much lower amplitude. A detector remotely located from the aircraft will detect radiation. It filters frequencies outside of the frequency range of interest. It will discriminate between the modulated pattern and the continuous pattern. This output provides the digital code that was encoded by the encoder. A decoder will decode the message for the recipient.

US5637946
Thermally Energized Electrical Power Source

An electrical device functions as a thermally energized DC power source. The device has a base plate of conductive metal. A number of electrode points protrude upward from the base plate, terminating in a sharp edge. A collection plate of conductive metal is positioned above the sharp upper edges of the electrode points. The gap between the electrode points and between the collection plate and the electrode points is electrically insulated. An electrical potential exists between the base plate and the upper collection plate while the device is at and above room temperature. The potential difference increases as the temperature increases.

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

**... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !**

ORDER PAGE

Copyright 1997 Walter Heskes and Philip I. Nelson, all rights reserved.

ECONOCEANIC

TUNING AND DETECTION CIRCUITS

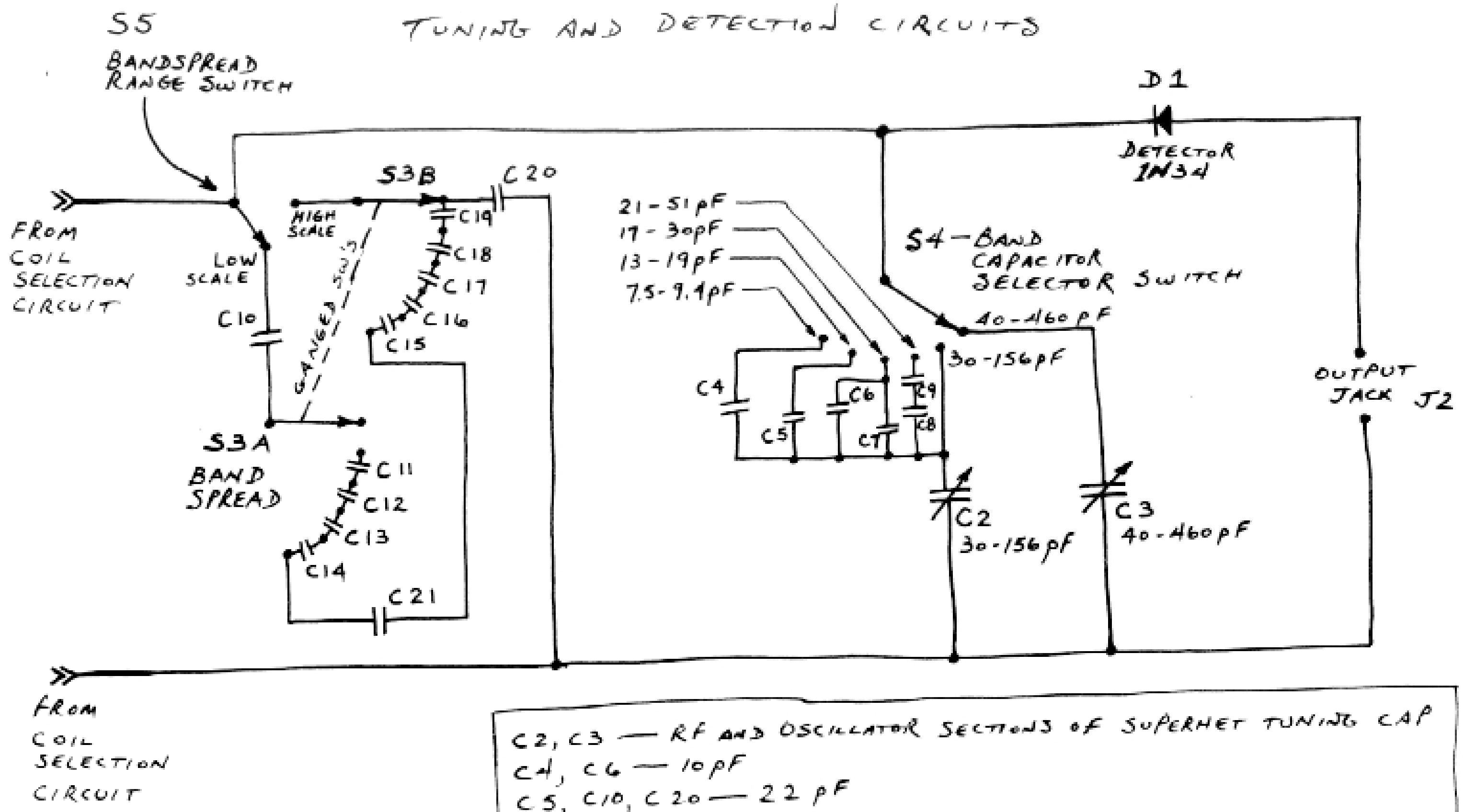
S5
BANDSPREAD RANGE SWITCH
FROM COIL SELECTION CIRCUIT
LOW SCALE
HIGH SCALE
C10
S3B
C20
C19
C18
C17
C16
C15
GANGED SW'S
S3A
BAND SPREAD
C11
C12
C13
C14
C21
FROM COIL SELECTION CIRCUIT
21-51pF
17-30pF
13-19pF
7.5-9.1pF
C4
C5
C6
C7
C8
C9
S4—BAND CAPACITOR SELECTOR SWITCH
40-460pF
30-156pF
C2
30-156pF
C3
40-460pF
D1
DETECTOR 1N34
OUTPUT JACK J2

C2, C3 — RF AND OSCILLATOR SECTIONS OF SUPERHET TUNING CAP
C4, C6 — 10pF
C5, C10, C20 — 22 pF
C7, C11, C12 — 27 pF
C8, C9 — 150pF
C13, C14, C15, C16, C17, C18, C19, C21 — 50pF

S3, S4 — 2-CIRCUIT, 6-POSITION ROTARY SWITCHES
S5 — SPDT SLIDE SWITCH
D1 — 1N34 OR 1N34A DIODE (POLARITY IS NOT IMPORTANT)
J2 — OUTPUT JACK: 1/8-INCH EARPHONE JACK RECOMMENDED

Build a Crystal Shortwave Radio

This project combines two popular themes from radio history—crystal radios and shortwave (SW) listening. It's designed from scratch by our non-resident engineer Walter Heskes.

Despite all of the advances in modern electronics, there are thousands of crystal sets in daily use throughout the world. The fidelity of tone and clarity of crystal rectification are hard to beat. What's more, crystal sets are absolutely free! A crystal receiver is powered solely by the radio waves that it pulls from the air.

We dubbed this inexpensive radio the "EconOceanic" in honor of the legendary Zenith TransOceanic family of shortwave radios. It tunes the AM broadcast band, from .5 MHz to 1.6 MHz, and most of the shortwave band, from 1.7 MHz up to about 17 MHz.

There's plenty of SW activity in these bands, so you'll hear lots of different voices, a mix of languages, and even music. In addition to powerhouse stations such as the BBC, you'll also hear non-commercial amateurs, some of whom broadcast using the dits and dahs of International Morse Code. And don't forget to set your wristwatch by the precise WWV time signals. They appear at exact multiples of 5.0 MHz across the SW bands. Listen for the calibration beeps and periodic time announcements.

We hope you enjoy this project. If you build an EconOceanic, send some email to let us know how it worked out. Walter can be reached at walterh@interworld.com.

Building Instructions

You're eager to start assembly, but a few minutes of preparation will help this project go smoothly. Here's how to get started.

Printing Instructions

Before you lay hands on any parts, print out this page and all of the diagrams in the following table.

Cabinet Decal	Component Layout
Coil Selector Circuit	Tuner/Detector Circuit
Coil Winding Diagram	Antenna Bracket Assembly

The large diagrams are designed to print on a standard 8.5 x 11-inch sheet of paper. Most of them are oriented horizontally for easier reading online. Choose your printer's "landscape" mode for the horizontal diagrams, to make sure they'll print on one sheet.

Be sure that your printer is set to print at a 1-to-1 ratio, so that the cabinet decal and layout diagram are the correct size.

Parts and Tools

We recommend getting *all* of the parts you need before you start assembly. Nothing is more frustrating than getting halfway through a project, then having to wait several days for some missing part to arrive in the mail.

The following parts list gives part numbers and sources for all major parts, plus some notes on particular items.

Most of the parts you need to build the EconOceanic Crystal Shortwave Receiver are available through Radio Shack. Two exceptions are the big tuning capacitor and the little trimmer capacitor. The big tuning capacitor can be obtained from Antique Electronic Supply. Or, you may be lucky enough to have a spare tuning capacitor in your junkbox. The small trimmer capacitor is available from All Electronics, allcorp@allcorp.com, 1-800-826-5432. Some notes on specific parts appear after this list.

Quantity	Component	Type	Part No.	Price
1	Diode	1N34A [1]	Radio Shack 276-1123	$ 1.19 pkg 10
1	Variable capacitor [2]	Dual-range: 40-460 pF/30-156 pF	AES CV300-145	$3.50
1	Trimmer capacitor	6-40pF	All Elect. VCAP-14	$1.50
1	Ceramic capacitors [3]	10 pF, 22 pF, 50 pF	Radio Shack 272-801	$2.49 pkg 100
2	Phone jacks	1/8-inch diameter	Radio Shack 274-251	$2.49 pkg 3
1	Speaker terminal strip [4]	Spring-loaded	Radio Shack 274-315	$ 1.29
4	Rotary switches	2-pole, 6-circuit	Radio Shack 275-1386	$1.69
1	Slide switch	DPDT	Radio Shack 275-407	$ 1.29 pkg 2
1	Enameled wire [5]	#22 and #26	Radio Shack 278-1345	$3.99 pkg 3
3	Knobs	1.0-inch diameter	Radio Shack 274-416	$2.29 pkg 4
2	Knobs	0.75-inch diameter	Radio Shack 274-415	$ 1.89 pkg 4
1	Project box	7.5" x 4.33" x 2.22" plastic	Radio Shack 270-224	$ 3.49
1	Cardboard tube	From paper towel roll		
1	Epoxy glue [6]	JB Weld or similar		
1	Household glue [7]	Elmer's Glueall or similar		
1	Rubber feet [8]	Heavy-duty self-stick	Radio Shack 64-2342	$ 1.49 pkg 8
Optional				
1		Battery powered	Radio Shack 277-	$11.99

	Amplified speaker [9]		1008	
1	Headphones	2,000-ohm	AES P-A466	$14.50
1	Antenna	6-foot telescoping	Radio Shack 270-1408B	$4.99
1	Fuse holder clips [10]	(for antenna swivel)	Radio Shack 270-739A	$ 1.19 pkg 2

Notes

[1] Any type of germanium diode will work.

[2] A variable air capacitor can be scrounged from any old AC/DC superheterodyne AM radio. Use the RF section for the wide range (40-460 pF) and the oscillator section for the narrow range (30-156 pF)

[3] Bags of ceramic capacitors are available in large quantities from Antique Electronic Supply and Radio Shack. These capacitors will work fine, but if you buy them in bulk, you may need to identify them using their markings or by reading them with a capacitor meter.

[4] The speaker terminal strip is used for the ground and antenna connections, as explained in the building instructions.

[5] Enamel-coated wire is available from Radio Shack and other suppliers. Radio Shack part no. 278-1345 is a package of three spools of enameled wire in the following sizes: #22 (gold enamel), #26 (green enamel), and #30 (red enamel), all for $3.99. If you don't see it on display, find it in the catalog and point it out to the clerks so they will order it for you.

[6] It's not critical which type of glue you use to fasten the metal frame of the tuning capacitor to the plastic surface of the instrument panel. Epoxy glue (JB weld or similar) will work, and so will silicon-based glues and perhaps others.

[7] Elmer's Glueall will work to coat the coils as you wind them. Or, you can use polyurethane varnish to seal the coils.

[8] Use four rubber feet to support the cabinet and two rubber feet to fasten the coil to the inside panel of the cabinet.

[9] If you choose to use an amplified speaker (highly recommended), use a short coaxial cable terminated with 1/8" plugs on both ends to connect the EconOceanic output to the speaker.

[10] One of the fuse holder clips is modified to form a swiveling support for the antenna mast, as explained in the building instructions. To fasten the antenna mast assembly to the cabinet, you'll also need a small screw, rubber washer, metal washer, lockwasher, and nut. If you don't have those items lying around somewhere, your house is cleaner than ours!

The tools that you'll need are pretty simple—a soldering iron, drill, wire stripper (or knife), screwdriver, and pliers or crescent wrench.

Preparing the Cabinet

Let's go! The first major step is to drill mounting holes for the controls in the plastic cabinet. The entire radio is mounted to the front panel, making it easy to remove for modifications or simply for showing a friend your handiwork.

The Component Layout Diagram and Cabinet Decal show you where to drill holes in the cabinet front panel for the components. Carefully trim the decal to fit the cabinet, then punch or cut holes in the decal where indicated. Glue on the decal. Don't put any glue under the places where you'll be drilling holes. Using a scribe or sharp knife, and guided by the scribe marks on the decal, make a little guide hole in the exact center of each place to drill. Using a sharp knife, carefully trim away the paper from each hole, so the drill won't tear the paper.

Regarding the decal, not all printers are created equal, so make sure it's printed at the right size before gluing it on. At the correct size, the decal measures 7.5 inches wide (horizontally) at the trim lines. If you can't get it to print correctly, you can try loading the image into a graphics program such as Photoshop or Paintshop Pro and resizing it for a new printout. The component layout diagram gives exact measurements for the holes, in any case.

Before you drill any holes, check clearances by placing the actual knobs in place atop the diagram. Now drill the holes. Drilling plastic is fast and easy. Wear protective eyegear to guard against flying fragments of plastic.

If you'd like to protect the paper decal against dirt and wear, spray on a thin coat of clear lacquer or poly finish, then set it aside to dry. A more permanent solution would be to have it laminated beforehand at a local copy shop.

Finally, mount the components in the positions shown in the component layout diagram. Pop the front panel into the case and admire your work. From the outside, this project looks almost finished—all that's left is to build the radio that goes inside!

Below are front and rear views of the author's prototype, after mounting the controls on the front panel. (This prototype did not have a decal.)

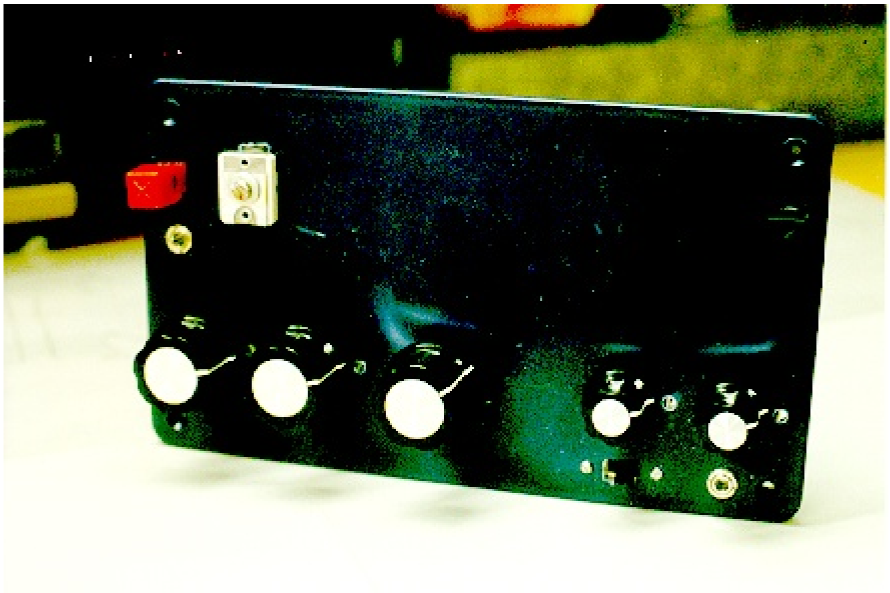

Building the Switchable Coils

Switchable coils are a unique feature of the EconOceanic. With many old shortwave radios, changing bands required you to unplug one coil and plug in another. With the EconOceanic, you simply turn switches to change coils in or out. That's more convenient, and it reduces wear and tear on the coils.

Making a Rigid Coil Form

In the EconOceanic, all of the BC and SW coils will be wound on a single core. The first step is to get something to wind your coils on. A standard roll of paper towels comes on a 1 5/8-inch diameter cardboard tube, which is just right.

Although you can wind the coils on a single tube, it's a good idea to reinforce the tube with a second stiffening layer, to help resist the tension of the wire as you wind the coils. To create the reinforcement, get an extra cardboard tube, slice it down its length, and remove a 3/8-inch wide strip along the cut, thus reducing its diameter. (Remember that this inner tube has to fit inside another tube of the same original diameter.) Apply a thin layer of Elmer's Glueall to the entire outer surface of the inner tube and insert it through the outer tube. Let it expand and gently press the tubes together to spread the glue. Hold them together for a couple of minutes, then set them aside for the glue to dry. Now you have a single, thick core.

Marking Eight Sections on the Coil Form

This design requires precise winding, which will be easier if you take a moment to mark your core with lines along its length to indicate its halves, quadrants, and eighth-sections. (In other words, draw eight equally-spaced straight lines, each one going the entire length of the tube.) When you need to wind a coil of 22 3/4 turns, you'll know exactly where to stop during the 23rd turn.

The more accurate your coils, the better they'll work. A coil with the wrong number of turns provides the wrong inductance and will not tune the desired range of frequencies.

Winding Your Coils

It's time to start winding, a task that requires care. Here's a little trick that simplifies the process. As you wind each coil, it helps to apply a thin layer of Elmer's Glueall to hold the most recently wound section together. Otherwise, the wires tend to spring apart (if that happens, you have to rewind the coil from scratch.) Wind about 10 turns, hold them tightly in place with your thumb, apply a pea-sized glob of glue to the wound section, and spread it with the middle finger of your other hand. Work the glue between the coils to encapsulate them. The thin layer dries in a few seconds and you can quickly continue winding. It's a technique you can develop pretty quickly. By the time you've wound all 101.5 turns of the big broadcast (BC) band coil, you'll feel like an expert!

Wind one coil at a time, following the Coil Winding Diagram. Use a full spool of wire when you start winding the big BC coil. It can be heartbreaking to run out of wire just before you reach the end.

Leave 10 inches of extra wire at each end of each coil. It is better to have extra length now which can be trimmed away later. These ends are the leads that you'll connect to the other components. Mark each lead wires using a little flag of masking tape (for instance, "250 uH Antenna" and "250 uH Ground"). The markers will simplify hookup later on.

The chart below lists the number of turns and size of wire to use for each coil. This radio tunes five bands. Each band requires a pair of windings: a small primary and a larger secondary.

Coil Section	Number of Turns	Wire gauge
250uH primary	17	26
250uH secondary	102	26
40uH primary	5.6	22
40uH secondary	33.5	22
20 uH primary	3.3	22
20 uH secondary	19.7	22
15 uH primary	2.3	22
15 uH secondary	14.5	22
10 uH primary	2.2	22
10 uH secondary	11.1	22

Each winding must be separated by a small gap from the next. Leave a 1/8-inch gap between windings for the 250uH, 40 uH, and 20 uH coils. Leave a 1/16-inch gap between the 15 uH and 10 uH coils.

Winding can be tedious, but it doesn't pay to rush. If you do this part correctly, you'll have a good-performing radio that performs virtually maintenance-free for many years.

When all the winding is done, check your work to make sure that everything is tight and secure. If necessary, apply more glue where needed. You can also apply a thin coat of clear varnish, if desired, for a more finished appearance.

This photo shows the author's prototype coil, ready for mounting.

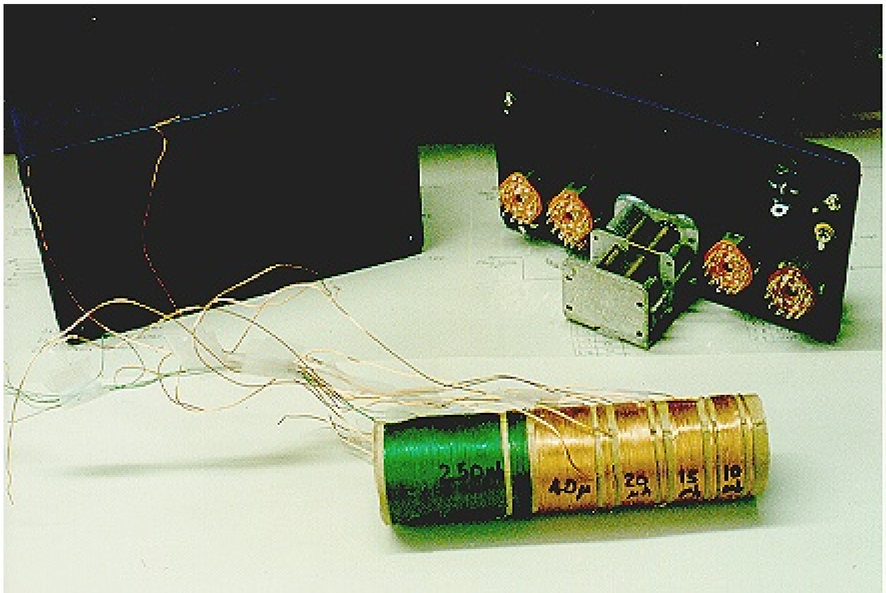

Mounting the Coils

Mount the coil assembly to the inside surface of the front panel. Check for adequate clearance when the tuning capacitor is fully open. Leave about 1/8-inch minimum clearance between the open rotor and the coil assembly. The rotor and coil must not touch. Mark the center of the coil form and mount self-sticking rubber feet to the centerline of the coil form. The feet serve as standoffs from the front panel, providing some clearance under the coil through which you can pass wires.

This photo shows the author's prototype coil mounted inside the front panel, which was placed upside down on the cabinet for the photo. Note the little plastic ties used to keep related leads together.

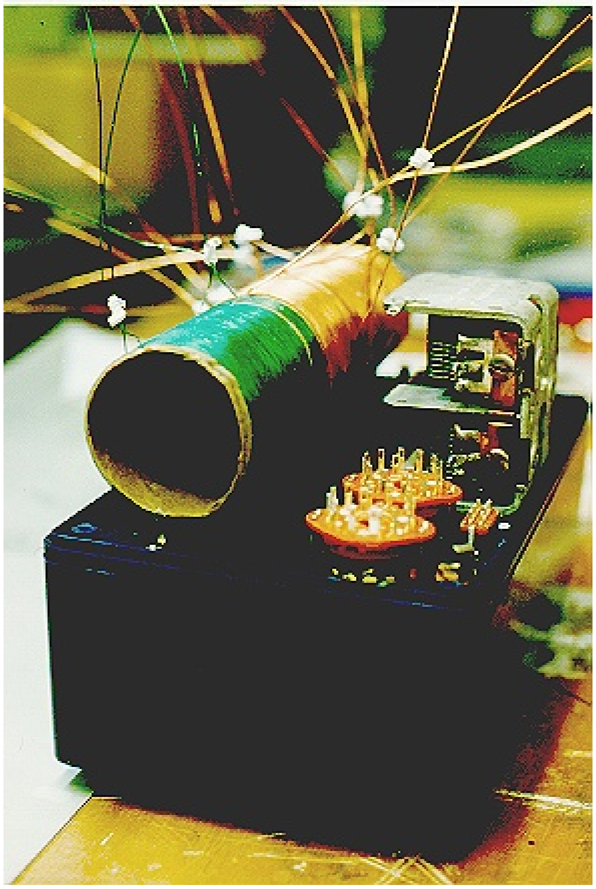

Wiring the chassis

Before you start hooking up components, take a few moments to study the schematic diagrams carefully. The ones you'll be using are the Coil Selector Circuit and Tuner/Detector Circuit.

Connecting Wires for Antenna and Ground

Let's start by attaching lead wires to the antenna and ground connections. These are shown in the Coil Selector Circuit. Connect wires to both antenna input jacks, the antenna trimmer, and the ground input jack. Check your connections for correctness and solder them.

Connecting Coils to the Coil Selector Switch

Identify the pinouts of the rotary switches. Use an ohmmeter to check continuity across the switch circuits and note the position of the switch pointer as each circuit is made. When you are certain you understand how the switch circuits are opened and closed, start wiring the antenna coils to the antenna coil selector switch.

Measure the length of wire needed to span the distance from the coil to the lug at the rotary selector switch. Straight runs are fine. Minimize the number of right angle bends, since those can reduce circuit sensitivity. Trim the coil takeoff leads to approximately 10 inches.

Using a razor blade or fine sandpaper, carefully scrape the enamel coating off each lead wire before you thread it through a lug on the rotary switch. Be careful not to break the fragile lead as you bend it around the lug.

Checking Coil Connections

Do not solder any connections until you have checked for proper continuity through each switch circuit and its coil. Refer to the Coil Selector Circuit as needed while checking your work.

Wiring Tuning Coils

After all the antenna coils are wired to the rotary switch, you can begin wiring the tuning coil circuits, referring to the Tuner/Detector Circuit diagram. Again, check for continuity through each switch circuit and its coil. Do not solder until all leads have been properly wired.

Next, connect the leads of the tuning coil to the tuning capacitor.

Connecting Capacitors

Connect the ceramic capacitors to the bandspread rotary switch. This is tedious work because the capacitor leads are short and the spacing between adjacent rotary switch lugs is small. Work slowly and carefully.

Connect the leads to the bandspread hi-lo selector switch. Since the DPDT switch has two circuits and only one circuit is needed, you can wire across both circuits if that makes wiring a bit easier for you. Or, use a larger SPDT switch.

Making Final Connections

Connect the 1N34 diode to the phone jack and to the band capacitor selector switch. Polarity is not important.

Connect the capacitors and leads to the tuning capacitor selector switch.

Connect the leads to the output jack.

Final Check and Soldering

You're almost done! Before soldering any connections, follow the schematic diagrams and recheck every connection for continuity. Correct any wiring errors that you find. When you're sure that everything's hooked up correctly, solder each lead.

Mounting the Antenna

The EconOceanic can use many different kinds of antennas. You'll get the best reception from a long outdoor antenna, of course. In many areas, however, you can get respectable performance from the 6-foot telescoping antenna listed in the Parts List.

To mount the telescoping antenna, modify one of the fuse holder clips as shown in the Antenna Bracket Assembly diagram. The clip forms a swiveling support for the antenna mast. To fasten the antenna mast assembly to the cabinet, you'll also need a small screw, rubber washer, metal washer, lockwasher, and nut.

Using Your Radio

Your EconOceanic radio is completely assembled. Now for the exciting part—using it! This description may sound harder at first than it really is in practice. We'll go through the process simply because it's a little more involved than using

Step 1. Connect antenna and ground. Connect your antenna and ground wires to the appropriate terminals on the front panel. If you are using the telescoping antenna, it is already connected, of course.

Step 2. Connect listening device. Plug in your 2000-ohm headphones or, better yet, use Radio Shack's amplified speaker so everyone can listen at the same time. The amplified speaker is a transistorized amplifier connected to an internal speaker. It also has a 1/8-inch phone jack, allowing you to listen without disturbing others.

The author connected the output of the amplified speaker to a good quality high fidelity speaker. Later, the author added a second amplified speaker to the radio. The second stage amplifies the weakest signals to audible levels and easily fills the room with sound. This diagram explains the connections:

Connect this . . .	To this
EconOceanic output jack	Input jack of first amplified speaker
Output jack of first amplified speaker	Input jack of second amplified speaker
Output jack of second amplified speaker	Hi-fi speaker

This photo shows the author's prototype, with temporary labels on the panel, connected to two amplified speakers in series.

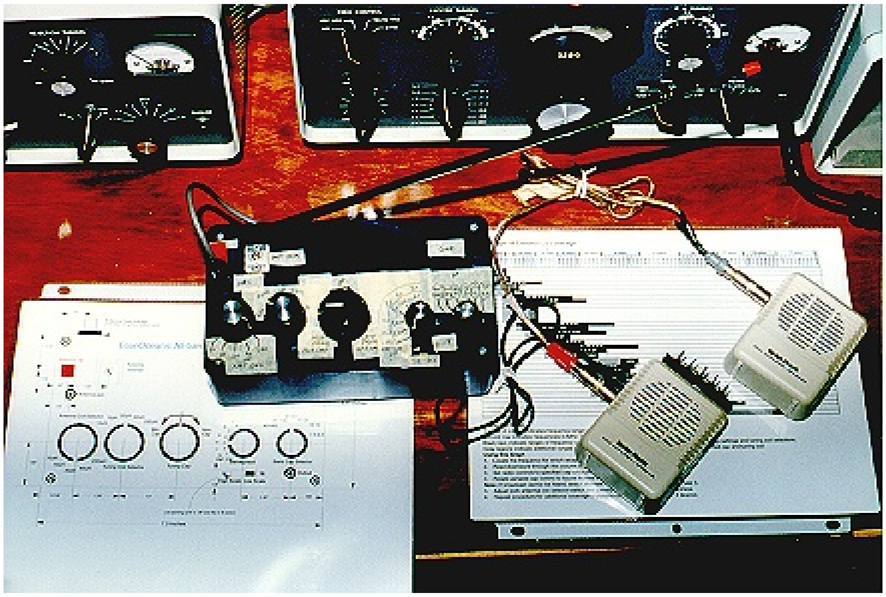

Step 3. Choose a Station. Depending upon the time of day and your location in the world, you can tune across the SW bands and listen in on broadcast activity. Some of the SW bands are more active in the evening hours, while others operate during the day.

Consult a schedule of shortwave broadcasts and note the time of day and frequency at which a desired broadcast will occur. Schedules are published in SW hobbyist magazines available at better newsstands. You can also find loads of schedule information on the Web, such as this resource page at SWLing.com. Another good resource is the book World Radio TV Handbook.

Step 4. Look up tuner settings. Use the Bar Graph of Continuous Coverage to find the correct capacitor and coil settings to tune in the desired frequency. Print this chart using your printer's landscape mode so that it will fit on a single 8.5x11-inch sheet.

As you become more familiar with the principles of resonance, you will find that you can tune the receiver to the desired frequency by thinking about the resonance formula and selecting coils and capacitors according to the following logic:

Low-frequency Broadcasts	High-frequency Broadcasts
Most capacitance	Least capacitance
Longest coil	Shortest coil

For example, to tune the AM broadcast band, you would select your longest coil (250uH) and your biggest capacitor setting (460 pF). Similarly, to tune the highest shortwave frequencies, you would select your shortest coil (10uH) and your smallest capacitor setting (7 - 9 pF).

It's possible to receive the same broadcast signal with different combinations of coils and capacitors. The best combination is the one at which the incoming signal is loudest. Experiment and learn the best settings. You can also combine dissimilar primary and secondary coils to receive radio signals. That is, you can set the antenna coil to 250uH and the tuning coil to 20uH and you will receive SW signals. This occurs because of the inductive coupling effects of the multiple coils on the same coil form.

Remember that the tuning coil alone determines the frequency at which radio signals will be received. The tuning coil resonates with the tuning capacitor. The antenna coil merely feeds the incoming radio waves to the tuning coil. Enjoy this advantage of mounting several adjacent coils on the same core.

Step 5. Turn the antenna coil selector switch to a desired primary coil.

Step 6. Rotate the tuning coil selector switch to the associated secondary coil.

Step 7. Turn the band capacitor selector switch to a desired setting. Now the band is selected.

Step 8. Rotate the tuning capacitor to select a broadcast signal within the selected band.

Step 9. Starting at the lowest setting of the "low" bandspread scale, adjust the bandspread switch for maximum broadcast signal loudness.

Step 10. If you are using the amplified speaker, adjust the volume to a comfortable setting. If you are using a pair of unamplified, 2000-ohm headphones, there is no loudness control.

Step 11. Log your settings. On a sheet of paper, write the time of day, the settings of your selector switches, and a description identifying the source of the broadast signal. For example, you might write:

```
9 P.M. EST, 20uH A coil, 20uH T coil, 13-19 pF, Radio Havana (Cuba) 8.9 MHz.
```

Logging your reception reports will help you locate your favorite radio broadcasts. Documenting your work also adds a hint of substance and credibility to your DX sessions. Above all, have fun!

Theory of Operation

This block diagram shows each stage of the EconOceanic crystal shortwave receiver. Click on each component to read more about it.

```
   Aerial
      |
      |
Antenna coil-->Tuning coil-->Tuning capacitor-->Detector-->Output
      |
      |
Earth ground
```

Aerial

The aerial's job is to capture electromagnetic waves from all sources and directions and funnel those charges into the antenna coil. The loudness of received signals is directly proportional to the height and length of the aerial. Mount the aerial as high as possible, preferably outdoors. A single wire 100 feet long makes an ideal aerial. The EconOceanic design provides for two antenna connections: one 1/8-inch phone jack and a spring-loaded clip (originally intended for fastening hi-fi speaker cables).

If your antenna is mounted outdoors, you should disconnect it whenever there is any chance of lightning in your vicinity.

Antenna coil

The antenna coil is the primary winding of the pair of coils that make up the antenna-tuning transformer. It is a fraction (1/6) of the number of turns that make up the tuning coil. The secondary coil multiplies the electrical field created by the antenna coil.

Earth ground

The earth ground connection effectively extends the antenna into the earth, which contains electrical signals. Keep the ground wire as short and straight as possible. There may be some sensitivity loss if the excess ground wire is coiled into a loop. Use either a metal cold water pipe that is buried at least several feet in the ground or an electrical conduit that is similarly buried or grounded. You can usually make a satisfactory ground connection using the screw that fastens the cover plate to an electrical outlet box.

Tuning coil

The tuning coil is the inductor which, in series with the tuning capacitor, determines the resonant frequency of the receiver according to the formula: **frequency = 1/[2*pi*sqroot(L*C)]**. Wind your coil carefully and space it close to the antenna coil (about 1/8" apart) to tune the specified frequencies.

The amount of coupling is directly proportional to the distance between the windings. If you slide the windings far apart, there is little coupling and a narrow resonance peak (e.g., 100 KHz). If you move the windings closer, there is more coupling and a broader resonance peak (e.g., 200 KHz). A broader resonance peak will mean less selective tuning; that is, several stations will be heard simultaneously at one tuner setting. Experiment, if you wish, with spacing your coils. For example, try 1/4-inch between the primary and secondary windings and test your design for selectivity. Have fun as you learn.

Tuning capacitor

The tuning capacitor is the variable component in the "tank" circuit which comprises the coil and the cap. By varying the capacitance, you can change the resonant frequency of the receiver within the limits of the coil inductance and the total capacitance available in the variable capacitor.

Detector (diode)

In this radio, a diode serves as the detector, separating the fluctuating direct current (containing the voices and music of the broadcast) from the amplitude-modulated, radio-frequency, alternating-current carrier wave that was transmitted from the radio station antenna. The direct current generated by the detector will drive the output device.

Output (headphones or amplified speaker)

The output device changes the electrical energy in the detector circuit into mechanical energy that moves air against our ears to create sound. For satisfactory sound output, this needs to have an input sensitivity of a few millivolts and an input impedance of a few thousand ohms, such as a pair of 2000-ohm headphones or Radio Shack's amplified speaker (part no. 277-1008). Don't try to use a permanent-magnet speaker, such as one scrounged from an old radio. Permanent-magnet speakers require more DC energy than a crystal set can generate.

If you completed this project and enjoy using your EconOceanic, please drop us a line and share your experiences. Walter's email address is walterh@interworld.com.

This radio construction project, including all descriptions, diagrams, photos, and the underlying electronic design, is published here for the noncommercial use of radio hobbyists. You may print and reproduce these project instructions for your personal use. Commercial use of this material is strictly forbidden.

©1995-2018 Philip I. Nelson, all rights reserved

Building a crystal radio set

Updated Tuesday 3rd July 2007

For many of the original teenagers, home-built crystal radios gave them their first contact with the world beyond their bedrooms. Here, Fraser Robertson describes how to make your own.

A crystal set is the most basic type of radio receiver. It is capable of receiving AM (amplitude modulated) transmissions, which are generally found on the long, medium and short wave bands.

The magic of the crystal set is that no power source is required; a tiny fraction of the transmitted signal provides sufficient power for the signal to be heard. What follows, while listing the parts you'll need, isn't a step-by-step how to guide, as part of the fun

Author

Fraser Robertson
34

(Other Open University departments[35])

Fraser Robertson is a project officer at The Open University, supporting various research projects.

View author profile >[34]

Related content

Crystal Radio

Transmitters

in building a crystal set is experimenting and optimising performance.

Aerial

Unless the crystal set is very close to the transmitter, an aerial, (or antenna) will be required in order to receive a big enough signal for the crystal set to work. Generally the aerial will be a length of wire, which should be as long and as high as possible.

A wire strung from an upstairs window to a distant tree will work well; say 30 metres long and seven metres high. The type of wire is not critical, but PVC covered stranded copper wire of about 2 milimetres diameter will work well.

If the wire is too thin it may break in the wind, and/or birds may fly into it; too thick and it will sag under its own weight.

The aerial should be insulated at each end from its support, a short length of washing line or other non-conductive cord can be used. An effective aerial will produce about one tenth of a volt radio frequency (RF), which will result in good reception.

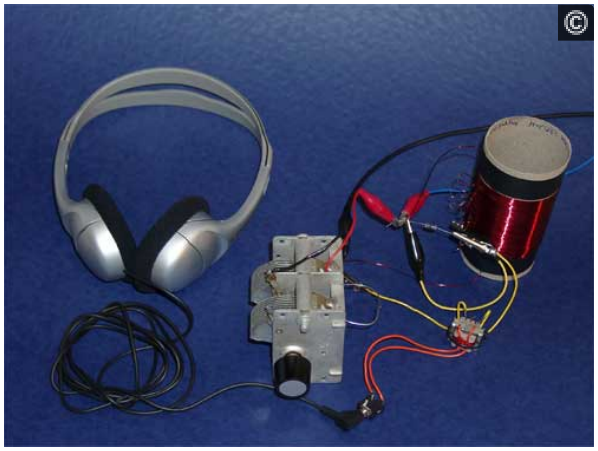

Earth

An earth connection is required to complete the aerial circuit. This can be a length of copper pipe driven into the ground, or a length of buried un-insulated wire.

More +

More like this

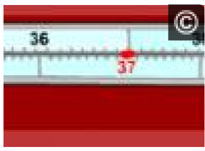
36

Challenge: Make a Thermometer [36]

Underwater Torch Challenge Glossary [37]

37

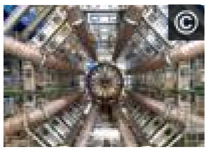
38

Why do we need a Large Hadron Collider? [38]

39

Anti-Matter [39]

As with the aerial, a better earth connection will improve reception. It is possible to connect to a metal heating or water pipe, but **this is not recommended on safety grounds** should there be an electrical fault. For the same reason, **the mains earth** should not be used.

When the crystal set is not in use the aerial should be connected directly to the earth connection. This will prevent any possible static build up on the aerial.

Tuned circuit

The aerial and earth are connected across a parallel tuned circuit. The purpose of this is to select the wanted station.

A simple crystal set will not have good selectivity so if two strong signals are received on nearby frequencies it will not be possible to separate them.

The tuned circuit comprises a coil (inductor) and a capacitor, which is usually variable for tuning. Pre-wound ferrite rod coils can be bought, or you can wind your own: for example, for medium wave, 100 turns of 0.5 milimetre enameled copper wire wound on a toilet roll or similar sized non-metallic former, twisted/tapped every ten turns.

The greater number of turns, the longer the wavelength (lower frequency), so a coil for long wave will require more turns. If using a pre-wound ferrite rod, ignore the small link windings.

An air spaced variable capacitor (with a value around 500 picofarad) removed from an old radio is ideal, or alternatively a small polyvaricon capacitor as used in transistor radios.

Variable capacitors often have two or more capacitors in one (known as **ganged**), in which case these can be connected together to increase the maximum capacitance (lower the tuning frequency).

Some variable capacitors also have FM (VHF) tuning sections included, which can be ignored. If a variable capacitor is not available, a fixed capacitor can be used (e.g. 150 picofarad - 470 picofarad), with the tuning adjusted by changing taps on the coil. In which case, taps every five turns would be useful.

Detector

The purpose of the detector is to recover the audio signal (waveform) from the radio frequency signal. Early crystal sets used an adjustable ‘cats whisker’ detector, but a germanium or Schottky diode is now used, examples are OA90, OA91, BAT81, BAT85, 1N34, 1N5711. Any one of these can be used.

Silicon diodes, (such as 1N4148) do not work as well due to their higher forward voltage drop. Modern zero gate threshold voltage FETs can be used for better sensitivity, although these are not readily available in small quantities. The detector diode is polarized (directional), but in the crystal set circuit it does not matter which way round it is connected.

Headphones

Traditional crystal sets used sensitive high impedance headphones, typically several thousand Ohms, which are hard to obtain now. A crystal earpiece can be used, or a sensitive pair of walkman style headphones (usually 32+32 Ohms) connected through a transformer (e.g. type LT700) to raise the impedance. Low impedance headphones will ‘load’ the tuned circuit excessively.

Look for headphones with a high sensitivity figure, for example 115 decibels - a higher number indicates higher sensitivity; CPC AV03626 106 decibel headphones work well. Be aware that a higher quoted decibel specification does not always guarantee higher sensitivity in practice.

Note that if a crystal earpiece is used, a resistor (such as a 47k Ohms) may need to be connected across it to provide a DC path for the detector diode.

Construction

The fine detail here is left to the builder. Component leads can be soldered, or twisted together, or 'crocodile clips' can be used to make connections.

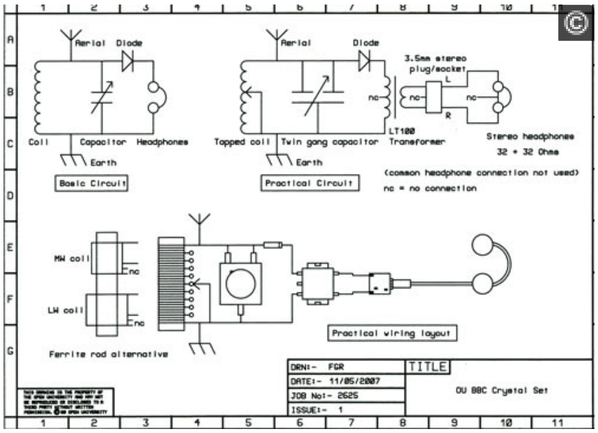
Aerial
Diode
Coil
Capacitor
Headphones
Earth
Basic Circuit
Aerial
Diode
3.5mm stereo plug/socket
L
R
nc
Tapped coil
Twin gang capacitor
LT100 Transformer
Earth
Practical Circuit
Stereo headphones
32 + 32 Ohms
(common headphone connection not used)
nc = no connection
MW coil
LW coil
Ferrite rod alternative
Practical wiring layout
DRN:- FGR
DATE:- 11/05/2007
JOB No:- 2625
ISSUE:- 1
TITLE
OU BBC Crystal Set
A
B
C
D
E
F
G
1
2
3
4
5
6
7
8
9
10
11

The diagram above, and the photograph at the top of the page, offers some guidance. **You can download a larger pdf of the diagram by clicking here**[4].

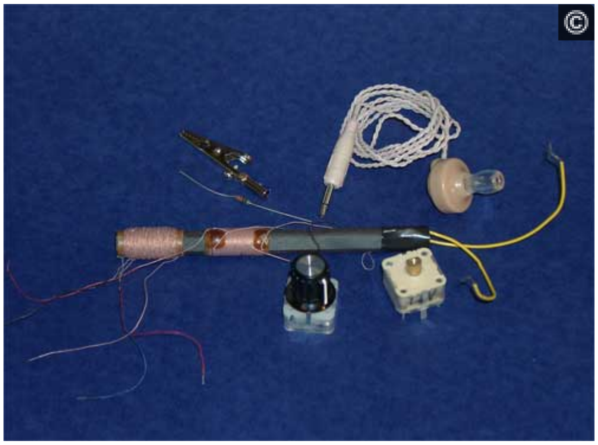

This photograph shows examples of alternative suggested parts. The components can be lashed together initially for experimentation, then connected more permanently and housed in any suitable enclosure once the crystal set is working.

Typical performance and troubleshooting

In the UK it will normally be possible to pick up two or three stations on medium wave, and/or BBC Radio 4 from Droitwich on long wave. If nothing at all is heard check:

- that all connections are sound, and ensure the enamel insulation on coil wire has been removed by being scraped off where connections are made
- that the correct connections for each component have been used
- that the tuning capacitor vanes are not bent and touching together.

If you get weak reception it may be that you need to improve the aerial / earth connections, or try a different pair of headphones.

Weblink

Watch a video showing the process of making a crystal set from **the OU's**

Subjects ▸ Science, Maths & Technology ▸ All content ▸ **Building a crystal radio set**

Science, Maths & Technology | Featured content | Free courses | All content

Building a crystal radio set

Updated Tuesday 3rd July 2007

For many of the original teenagers, home-built crystal radios gave them their first contact with the world beyond their bedrooms. Here, Fraser Robertson describes how to make your own.

Share this article

Author

A crystal set is the most basic type of radio receiver. It is capable of receiving AM (amplitude modulated) transmissions, which are generally found on the long, medium and short wave bands.

The magic of the crystal set is that no power source is required; a tiny fraction of the transmitted signal provides sufficient power for the signal to be heard. What follows, while listing the parts you'll need, isn't a step-by-step how to guide, as part of the fun in building a crystal set is experimenting and optimising performance.

Aerial

Unless the crystal set is very close to the transmitter, an aerial, (or antenna) will be required in order to receive a big enough signal for the crystal set to work. Generally the aerial will be a length of wire, which should be as long and as high as possible.

A wire strung from an upstairs window to a distant tree will work well; say 30 metres long and seven metres high. The type of wire is not critical, but PVC covered stranded copper wire of about 2 milimetres diameter will work well.

If the wire is too thin it may break in the wind, and/or birds may fly into it; too thick and it will sag under its own weight.

The aerial should be insulated at each end from its support, a short length of washing line or other non-conductive cord can be used. An effective aerial will produce about one tenth of a volt radio frequency (RF), which will result in good reception.

Fraser Robertson (Other Open University departments)

Fraser Robertson is a project officer at The Open University, supporting various research projects.

View author profile ›

Related content

Crystal Radio

Transmitters

Receivers

More +

More like this

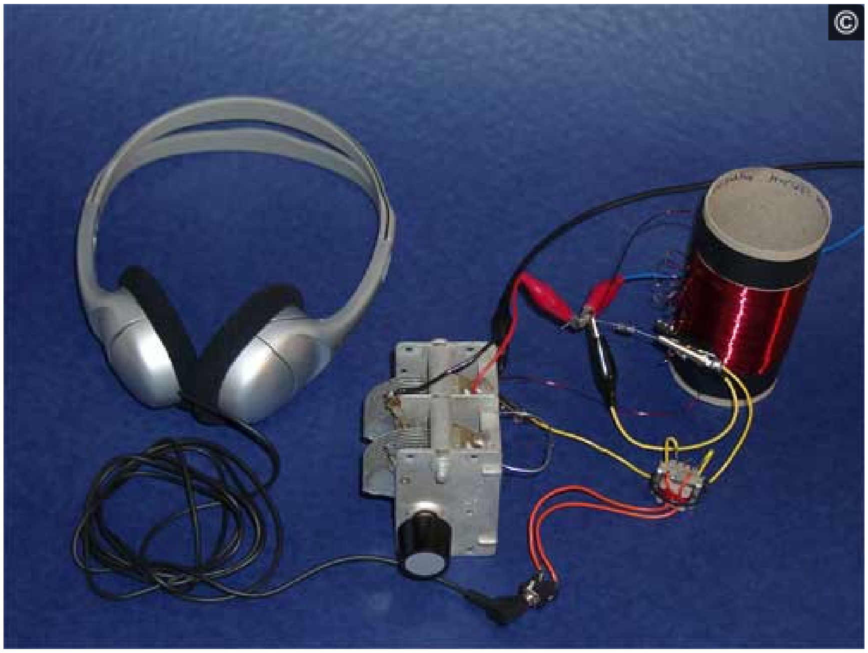

Earth

An earth connection is required to complete the aerial circuit. This can be a length of copper pipe driven into the ground, or a length of buried un-insulated wire.

As with the aerial, a better earth connection will improve reception. It is possible to connect to a metal heating or water pipe, but **this is not recommended on safety grounds** should there be an electrical fault. For the same reason, **the mains earth** should not be used.

When the crystal set is not in use the aerial should be connected directly to the earth connection. This will prevent any possible static build up on the aerial.

Tuned circuit

The aerial and earth are connected across a parallel tuned circuit. The purpose of this is to select the wanted station.

A simple crystal set will not have good selectivity so if two strong signals are received on nearby frequencies it will not be possible to separate them.

The tuned circuit comprises a coil (inductor) and a capacitor, which is usually variable for tuning. Pre-wound ferrite rod coils can be bought, or you can wind your own: for example, for medium wave, 100 turns of 0.5 milimetre enameled copper wire wound on a toilet roll or similar sized non-metallic former, twisted/tapped every ten turns.

Sir Bernard Lovell: An appreciation

Discover superstrings

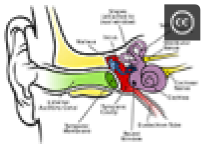

What is sound?

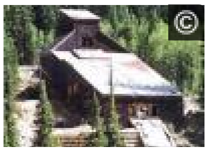

Inside the science - radioactivity

The greater number of turns, the longer the wavelength (lower frequency), so a coil for long wave will require more turns. If using a pre-wound ferrite rod, ignore the small link windings.

An air spaced variable capacitor (with a value around 500 picofarad) removed from an old radio is ideal, or alternatively a small polyvaricon capacitor as used in transistor radios.

Variable capacitors often have two or more capacitors in one (known as **ganged**), in which case these can be connected together to increase the maximum capacitance (lower the tuning frequency).

Some variable capacitors also have FM (VHF) tuning sections included, which can be ignored. If a variable capacitor is not available, a fixed capacitor can be used (e.g. 150 picofarad - 470 picofarad), with the tuning adjusted by changing taps on the coil. In which case, taps every five turns would be useful.

Detector

The purpose of the detector is to recover the audio signal (waveform) from the radio frequency signal. Early crystal sets used an adjustable ‘cats whisker’ detector, but a germanium or Schottky diode is now used, examples are OA90, OA91, BAT81, BAT85, 1N34, 1N5711. Any one of these can be used.

Silicon diodes, (such as 1N4148) do not work as well due to their higher forward voltage drop. Modern zero gate threshold voltage FETs can be used for better sensitivity, although these are not readily available in small quantities. The detector diode is polarized (directional), but in the crystal set circuit it does not matter which way round it is connected.

Headphones

Traditional crystal sets used sensitive high impedance headphones, typically several thousand Ohms, which are hard to obtain now. A crystal earpiece can be used, or a sensitive pair of walkman style headphones (usually 32+32 Ohms) connected through a transformer (e.g. type LT700) to raise the impedance. Low impedance headphones will ‘load’ the tuned circuit excessively.

Look for headphones with a high sensitivity figure, for example 115 decibels - a higher number indicates higher sensitivity; CPC AV03626 106 decibel headphones work well. Be aware that a higher quoted decibel specification does not always guarantee higher sensitivity in practice.

Note that if a crystal earpiece is used, a resistor (such as a 47k Ohms) may need to be connected across it to provide a DC path for the detector diode.

Construction

The fine detail here is left to the builder. Component leads can be soldered, or twisted together, or ‘crocodile clips’ can be used to make connections.

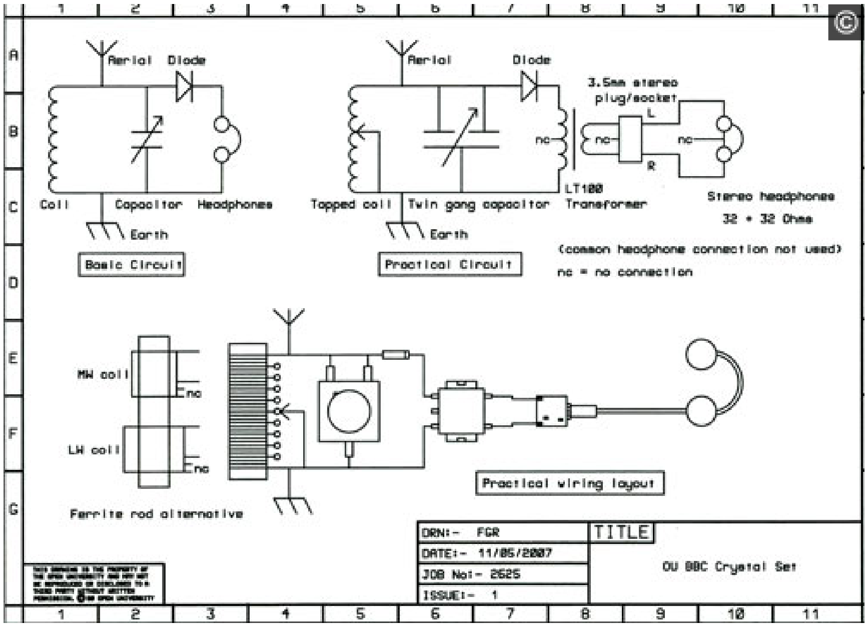
1
2
3
4
5
6
7
8
9
10
11
A
B
C
D
E
F
G
Aerial
Diode
Coil
Capacitor
Headphones
Earth
Basic Circuit
Aerial
Diode
3.5mm stereo plug/socket
L
R
nc
nc
nc
Tapped coil
Twin gang capacitor
LT100 Transformer
Stereo headphones
32 + 32 Ohms
Earth
Practical Circuit
(common headphone connection not used)
nc = no connection
MW coil
nc
LW coil
nc
Ferrite rod alternative
Practical wiring layout
DRN:- FGR
DATE:- 11/05/2007
JOB No:- 2625
ISSUE:- 1
TITLE
OU BBC Crystal Set

The diagram above, and the photograph at the top of the page, offers some guidance. **You can download a larger pdf of the diagram by clicking here**.

We use cookies to make sure our websites work effectively and to improve your user experience. If you continue to use this site we will assume that you are happy with this. However

More Info/Change Settings

this site we will assume that you are happy with this. However,
you can change your cookie settings at any time.

Continue

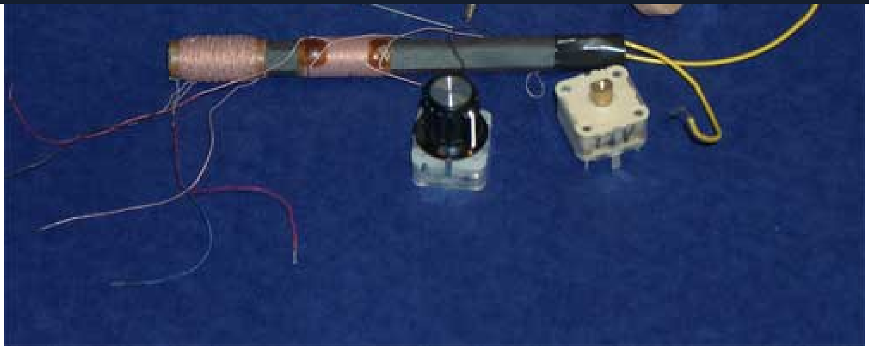

This photograph shows examples of alternative suggested parts. The components can be lashed together initially for experimentation, then connected more permanently and housed in any suitable enclosure once the crystal set is working.

Typical performance and troubleshooting

In the UK it will normally be possible to pick up two or three stations on medium wave, and/or BBC Radio 4 from Droitwich on long wave. If nothing at all is heard check:

- that all connections are sound, and ensure the enamel insulation on coil wire has been removed by being scraped off where connections are made
- that the correct connections for each component have been used
- that the tuning capacitor vanes are not bent and touching together.

If you get weak reception it may be that you need to improve the aerial / earth connections, or try a different pair of headphones.

Weblink

Watch a video showing the process of making a crystal set from **the OU's Interfaculty Electronics and Computing department**.

Copyright information +

Ratings ★★★★☆ Your rating None. Average rating 3.8 out of 5, based on 30 ratings

Be the first to post a comment

Leave a comment

Sign in or create your **OpenLearn account** to join the discussion.

We invite you to discuss this subject, but remember this is a public forum.

Please be polite, and avoid your passions turning into contempt for others. We may delete posts that are rude or aggressive; or edit posts containing contact details or links to other websites.

For further information, take a look at our frequently asked questions which may give you the support you need.

Have a question?

Other content you may like

Science, Maths & Technology

DIY: The Saucepan Radio

A step-by-step guide on how to make a radio, part of the BBC/OU's programme website for Rough Science 1

▶ Video 10 mins

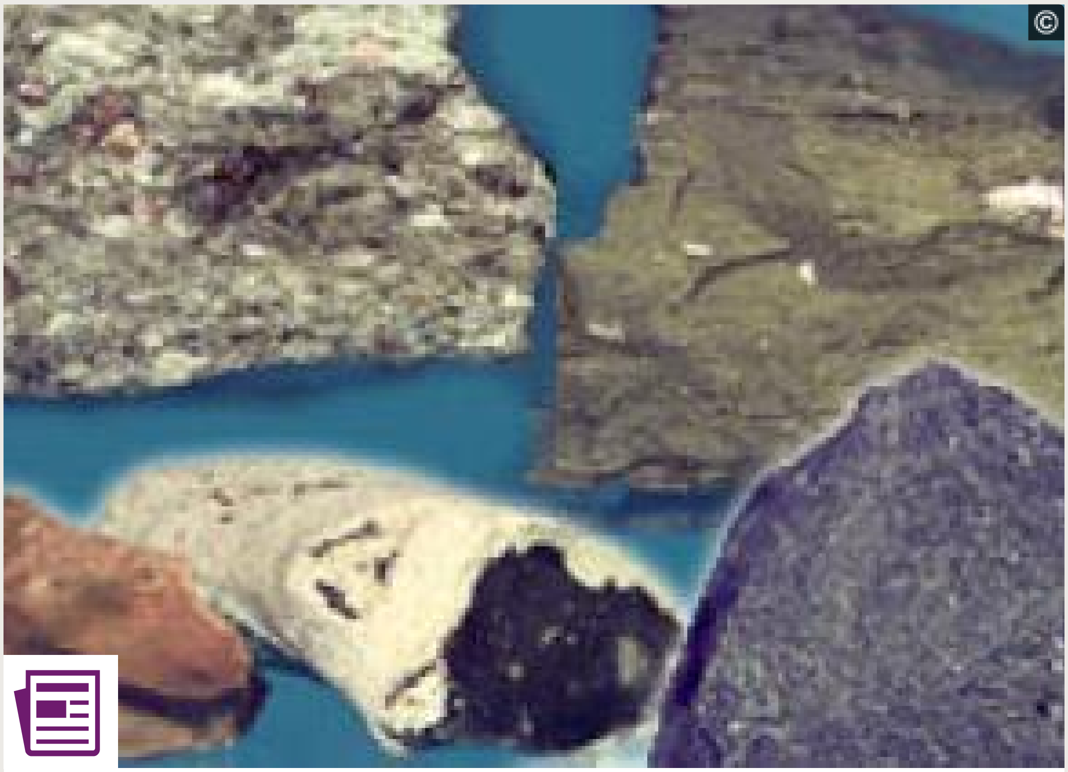

Science, Maths & Technology

You have chosen crystals

What is your crystaline rock?

Article

TV, Radio & Events

Rough Science 5 Zanzibar: Jonathan...

Jonathan Hare's diary about the challenge for the Call of the Wild programme, part of the fifth BBC/OU TV series Rough Science, based in...

Article

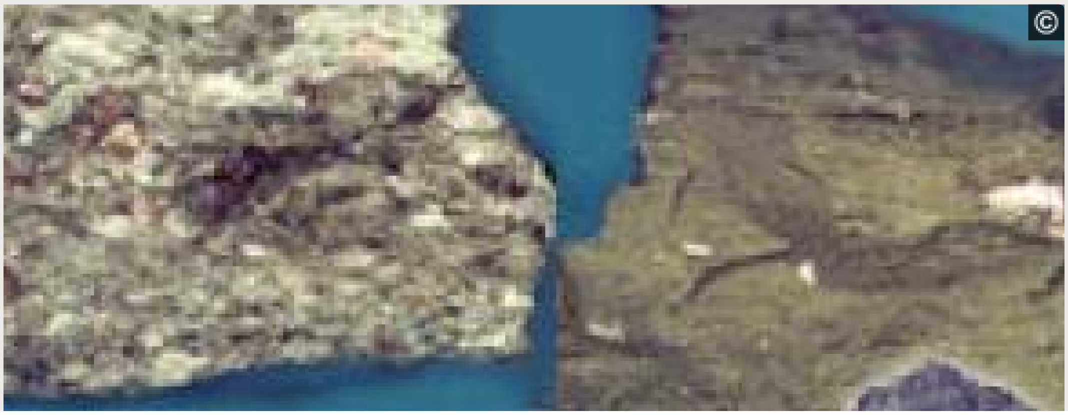

Science, Maths & Technology

You have chosen random

A step closer to identiifying your rock

Article

Science, Maths & Technology

Gneiss

A brief description of the nature of gneiss

Article

History & The Arts

Speaking Shakespeare: How was Shakespeare...

How do we know what actors sounded like when they performed premieres of Shakespeare plays? The clues, as ever, are in the lines themselves...

Video 10 mins

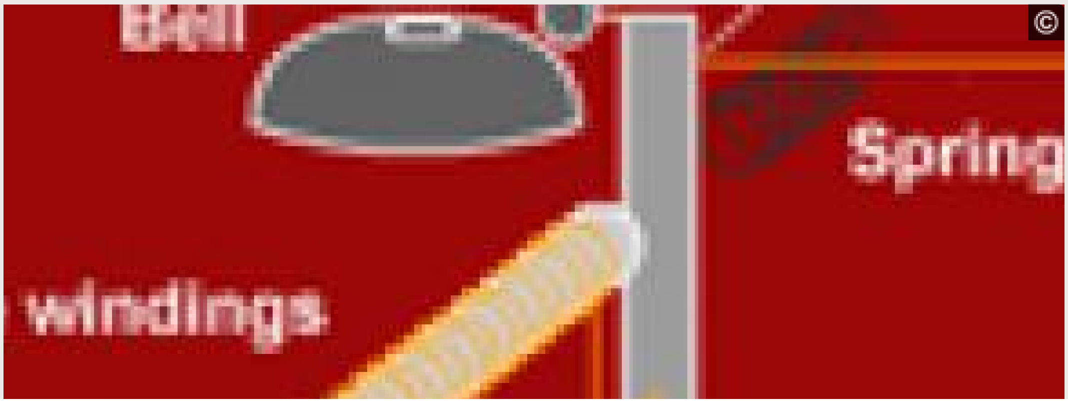

Science, Maths & Technology

Challenge: Transmit and Receive a Signal

The science behind transmitting and receiving radio signals, part of the BBC/OU's programme website for Rough Science 2

Article

Science, Maths & Technology

Microgabbro/ dolerite

A brief description of the nature of microgabbro or dolerite

Article

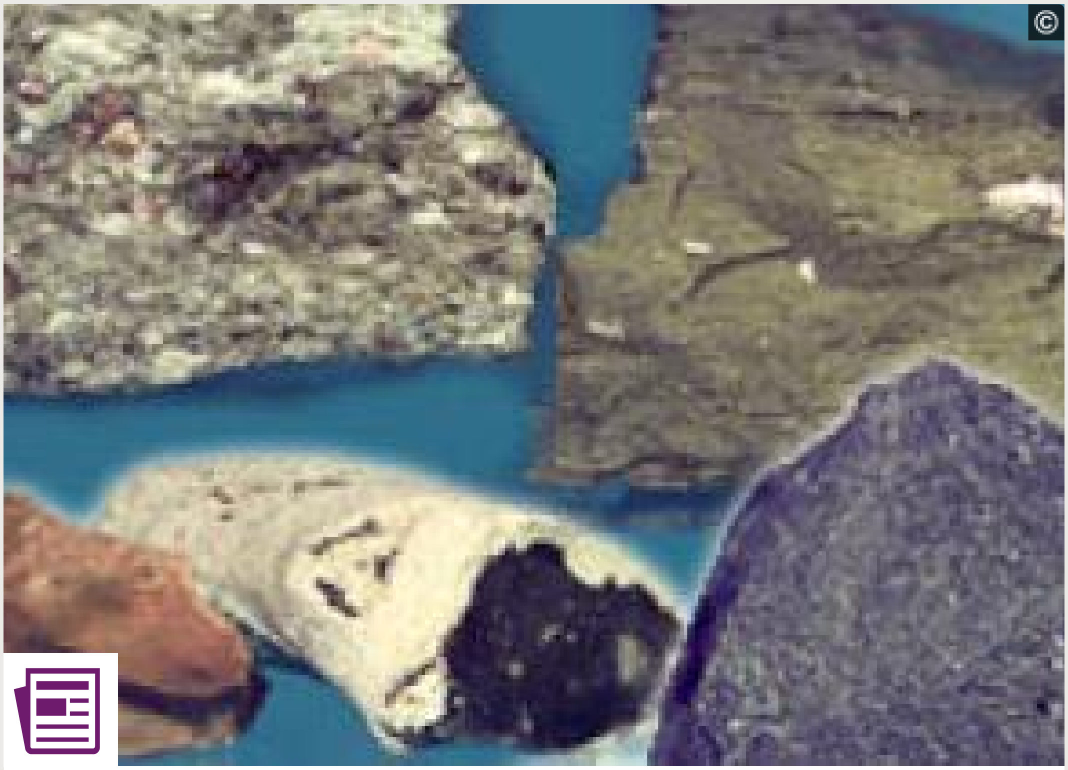

Science, Maths & Technology

Rock analyser: Just the facts

Our step by step rock analyser will help you identify your sample...

Article

Back to top

Search for free courses, interactives, videos and more!

About OpenLearn

About us

Frequently asked questions

Study with The Open University

Subscribe to our newsletter

Contact OpenLearn

OpenLearn Create

OpenLearn Cymru

Explore subjects

Money & Business

Education & Development

Health, Sports & Psychology

History & The Arts

Languages

Nature & Environment

Science, Maths & Technology

Society, Politics & Law

Explore resources

Latest from OpenLearn

Free courses

Skills for study

Skills for work

OU on TV & radio

Order our free print items

My OpenLearn profile

Our partners

OpenLearn works with other organisations by providing free courses and resources that support our mission of opening up educational opportunities to more people in more places.

Support us

The Open University OpenLearn

©2017. All rights reserved. The Open University is incorporated by Royal Charter (RC 000391), an exempt charity in England & Wales and a charity registered in Scotland (SC 038302). The Open University is authorised and regulated by the Financial Conduct Authority in relation to its secondary activity of credit broking.

Conditions of use Privacy and cookies
Modern Slavery Act Copyright

Twitter Facebook YouTube

Devises: Euro ▼

Connection | Création de Compte

BOUTIQUEDEPROMOTIONENLIGNE
World Best Online Store

Search

Panier(0) €0.00

Accueil HOT Catégories FAQ Nous Contacter

Product 2/21

Previous Product Listing Next

Cape Cod Credit Card Crystal Radio DIY KIT Schools- scouts science fairs. hYPKd7G2-07111823-729782709

★★★★★

Model: hYPKd7G2-07111823-729782709

800 Units in Stock

~~€36.00~~ €18.00

Ajouter au Panier :

⊖ 1 ⊕ Add to Cart

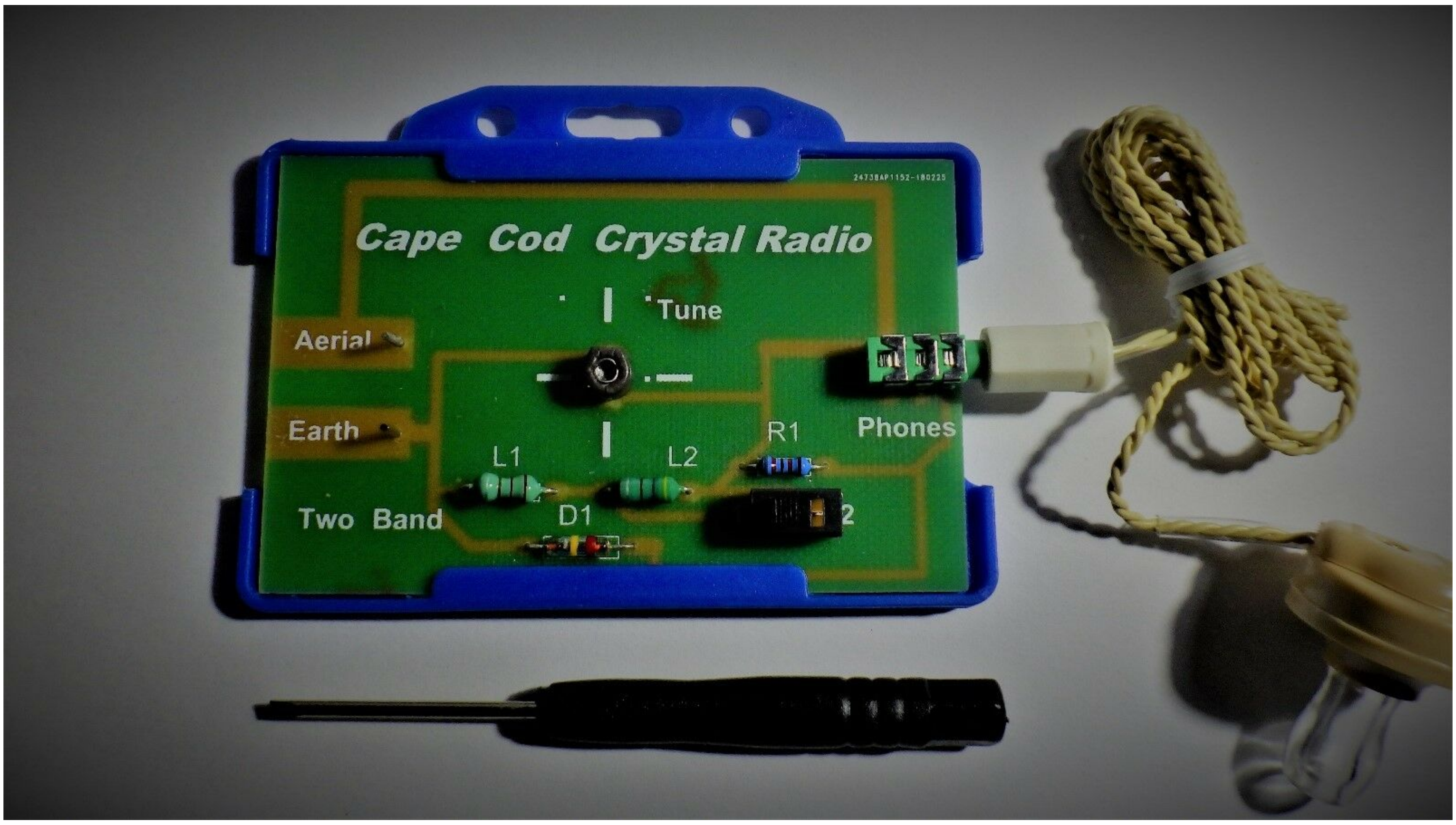

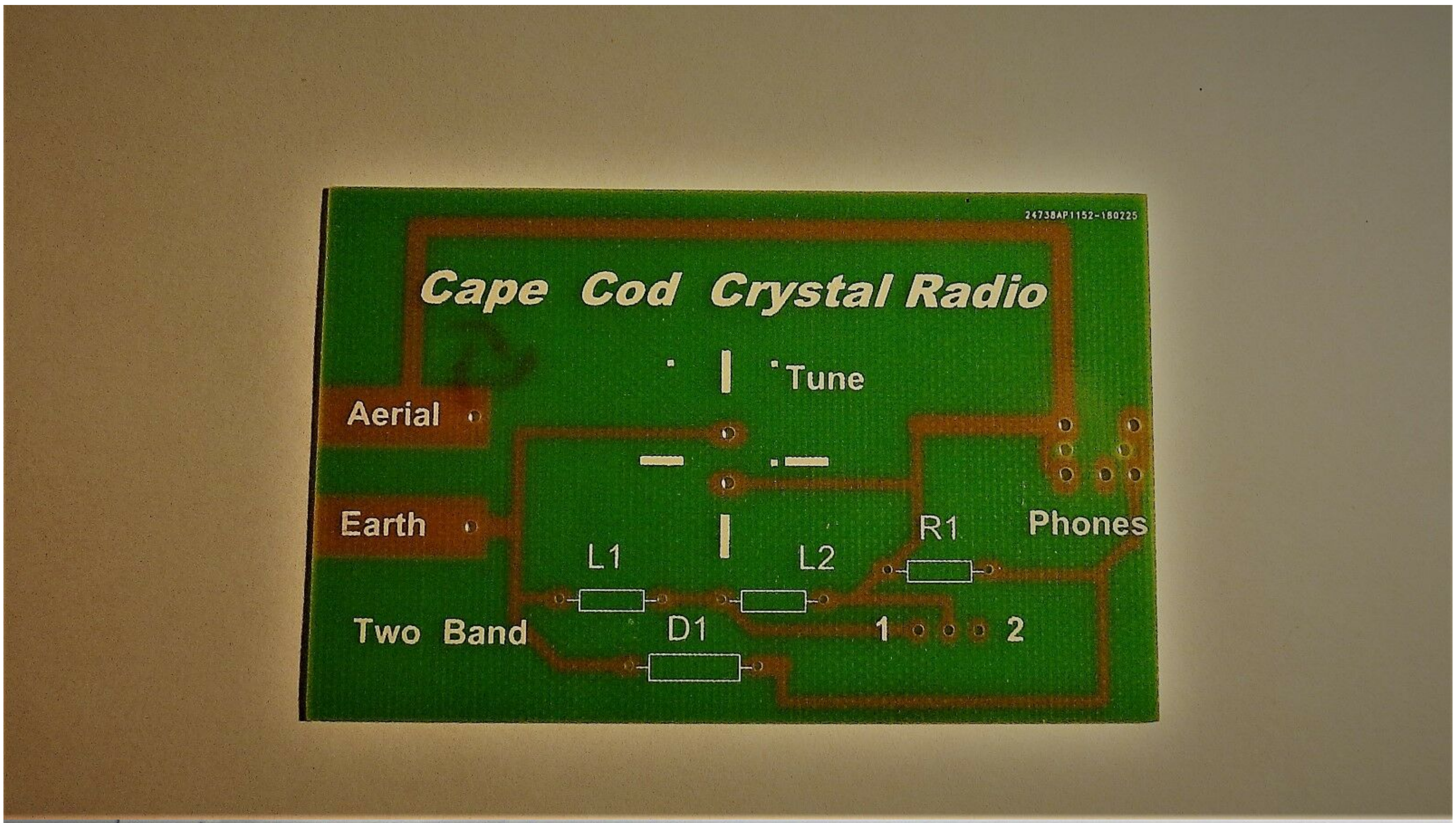
24738AP1152-180225
Cape Cod Crystal Radio
Tune
Aerial
Earth
L1
L2
R1
Phones
Two Band
D1
1
2

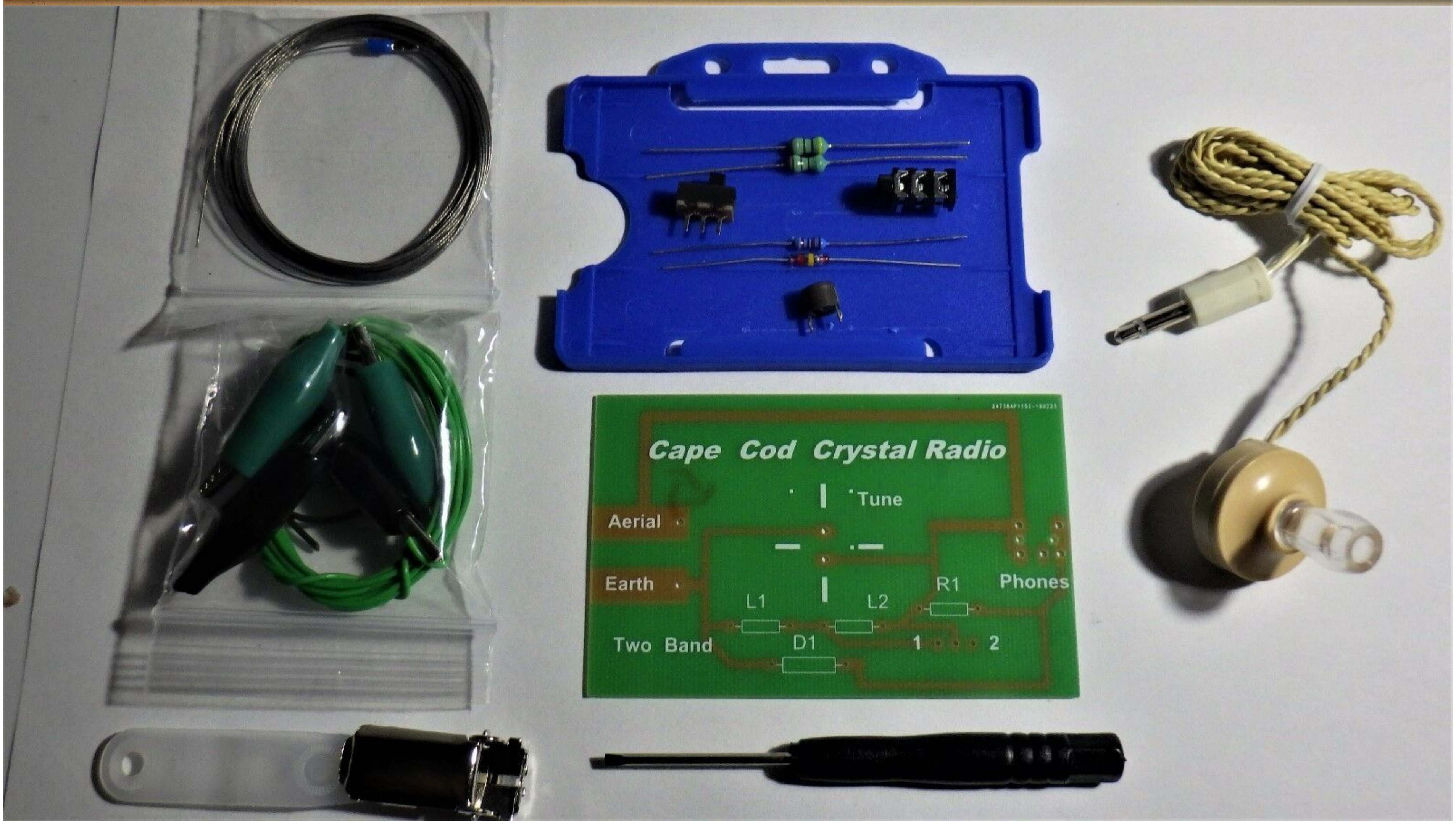
Cape Cod Crystal Radio
Tune
Aerial
Earth
L1
L2
R1
Phones
Two Band
D1
1
2

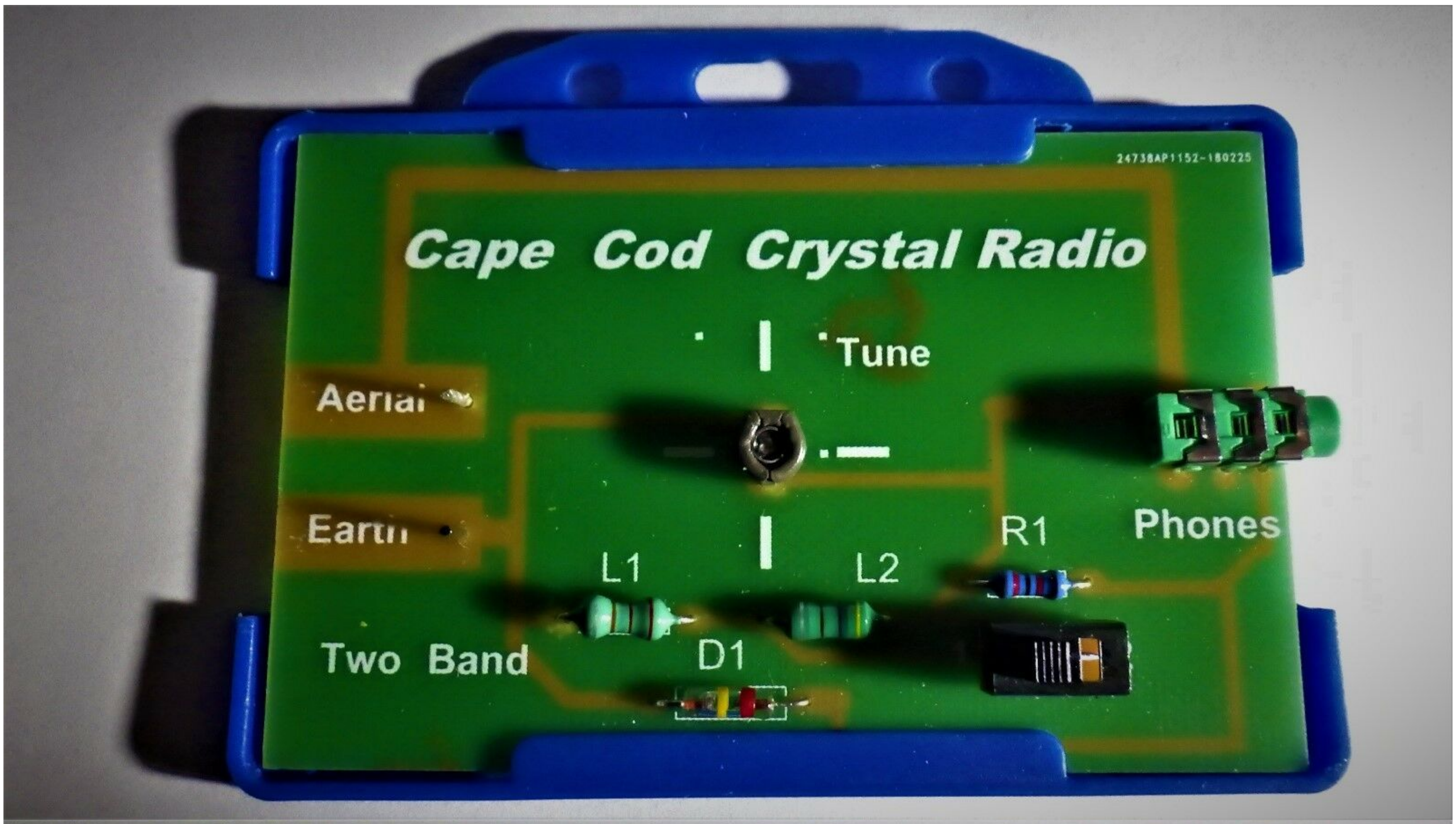
24738AP1152-180225
Cape Cod Crystal Radio
Tune
Aerial
Earth
Phones
R1
L1
L2
D1
Two Band

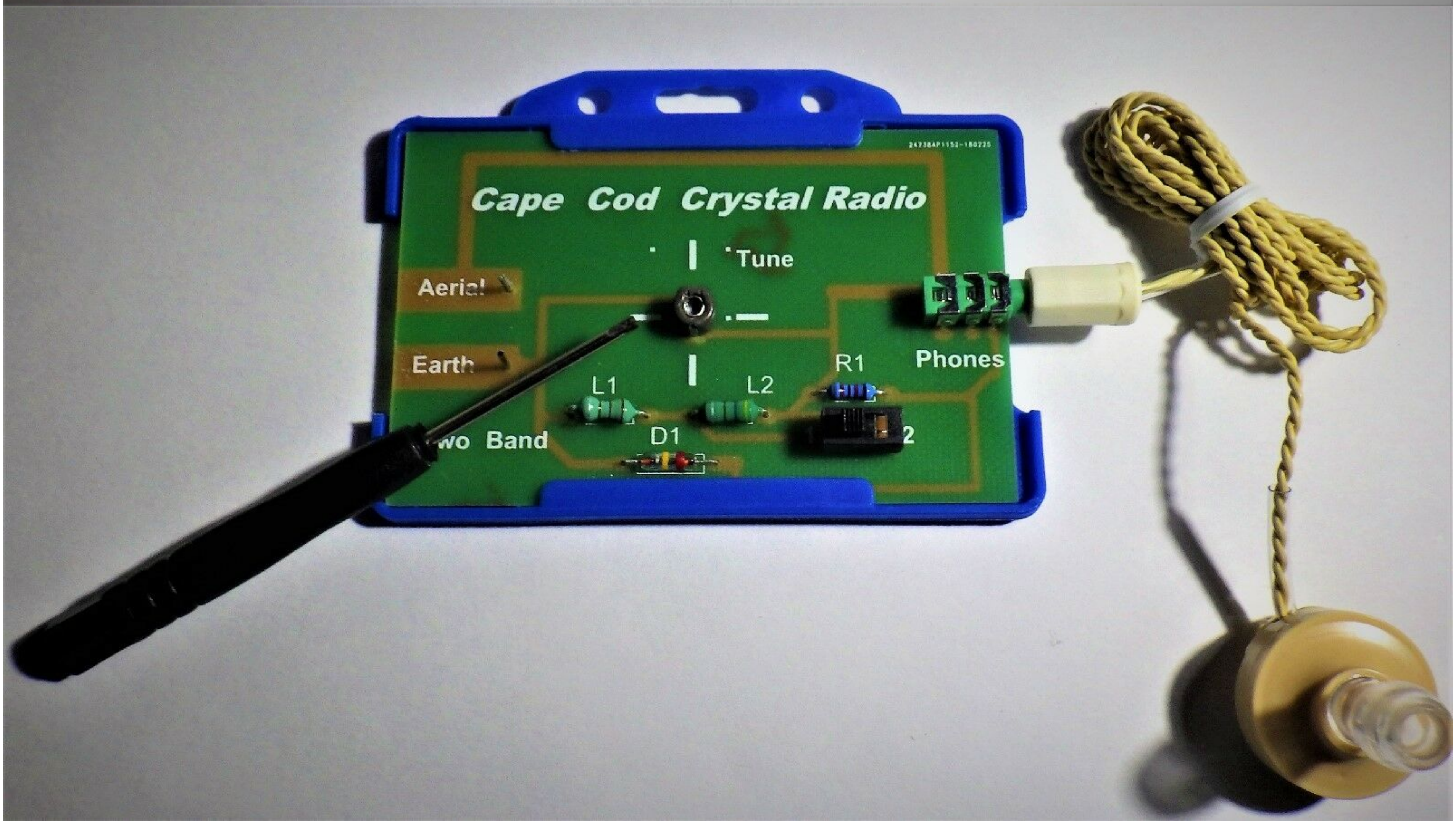
Cape Cod Crystal Radio
Tune
Aerial
Earth
Phones
R1
L1
L2
D1
Two Band

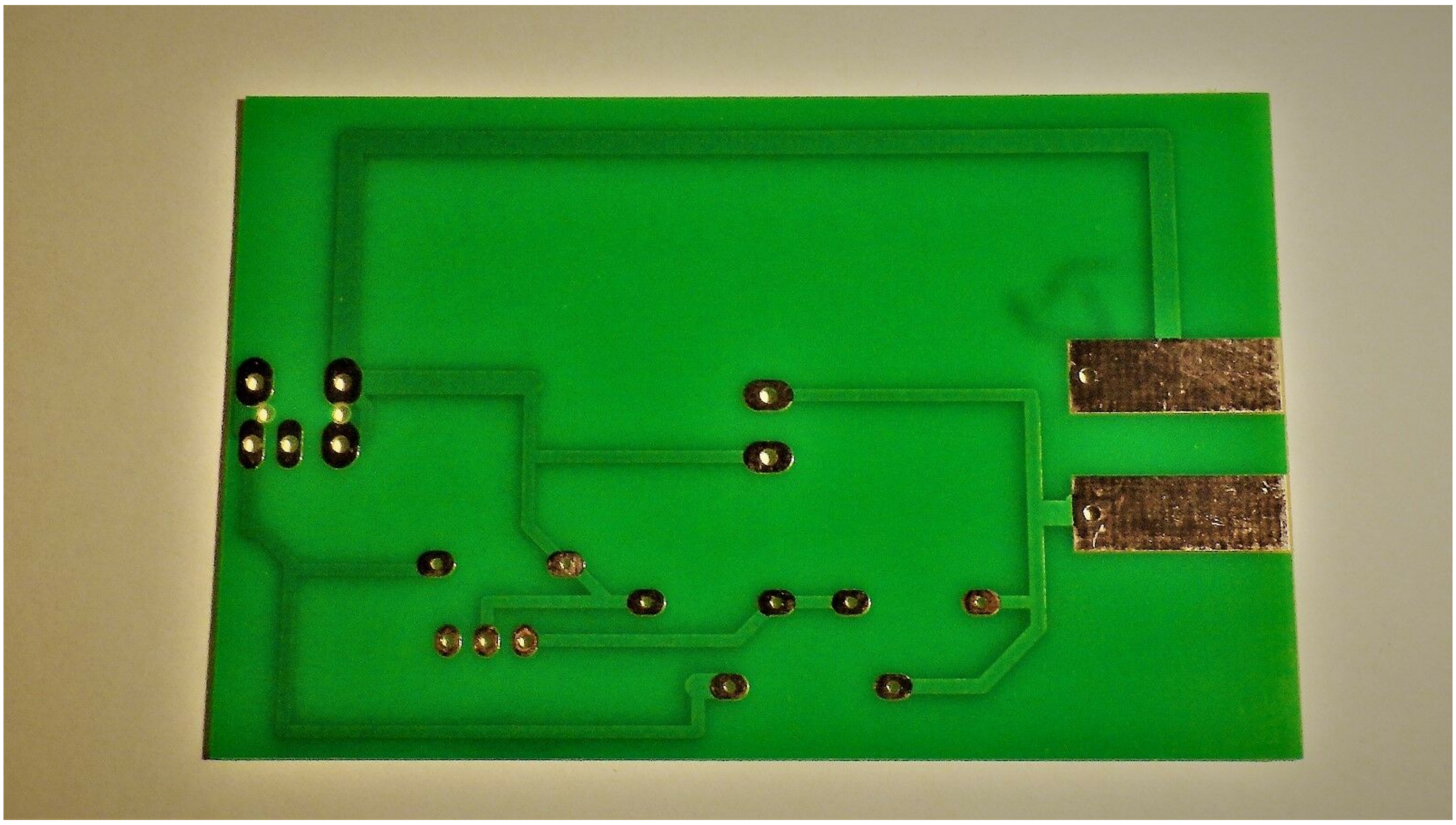

État :	Neuf: Objet neuf et intact, n'ayant jamais servi, non ouvert. Consulter l'annonce du vendeur pour avoir plus de détails.Afficher la définition de tous les états**- la page s'ouvre dans une nouvelle fenêtre ou un nouvel onglet**... En savoir plus**sur l'état**	Brand:	Poldhuradiokits
MPN:	POLCOD1	Modified Item:	No
Supported Modes:	AM	Country/Region of Manufacture:	United Kingdom

---

Customers who bought this product also purchased...

ICOM IC-R3 Communication...

~~€34.00~~

€17.00

★★★★★

YAESU FRG-8800 COMMUNICATIO...

~~€36.00~~

€18.00

★★★★★

Support mobile-YSK 90 pour Yaes...

~~€22.00~~

€11.00

★★★★★

Vintage KOYO FS-800...

~~€199.98~~

€99.99

★★★★★

Icom IC-R7000 RF Unit-afficher le...

~~€40.00~~

€20.00

★★★★★

CONTACT US

service17@vinayotap.com

CUSTOMER SERVICE

Mon Compte

FAQ

Plan du Site

CORPORATION

Livraisons & Retours

Confidentialité

Search

WHY CHOOSE US

Conditions Générales

À propos de nous

Paiement

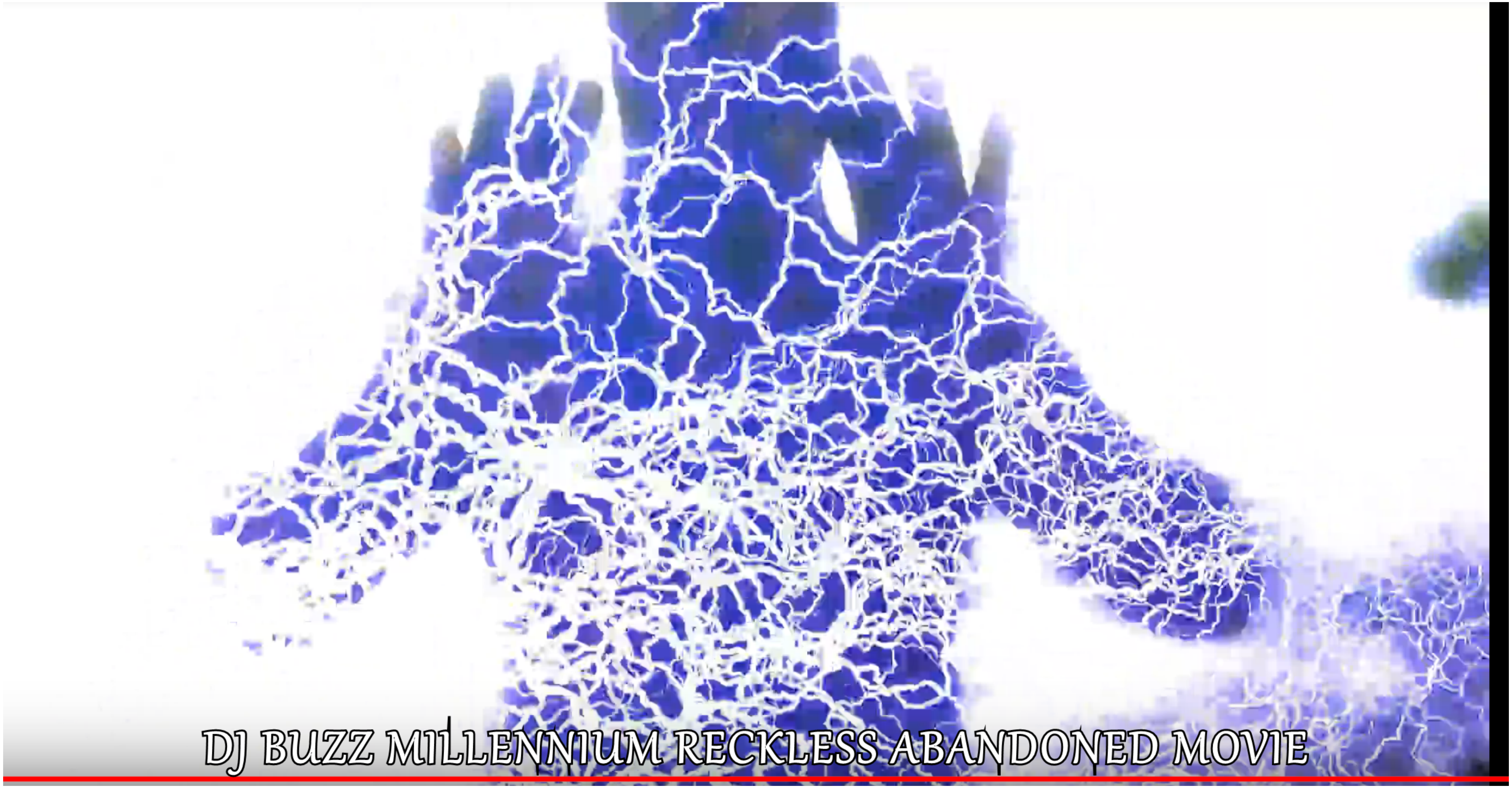

DJ BUZZ MILLENNIUM RECKLESS ABANDONED MOVIE

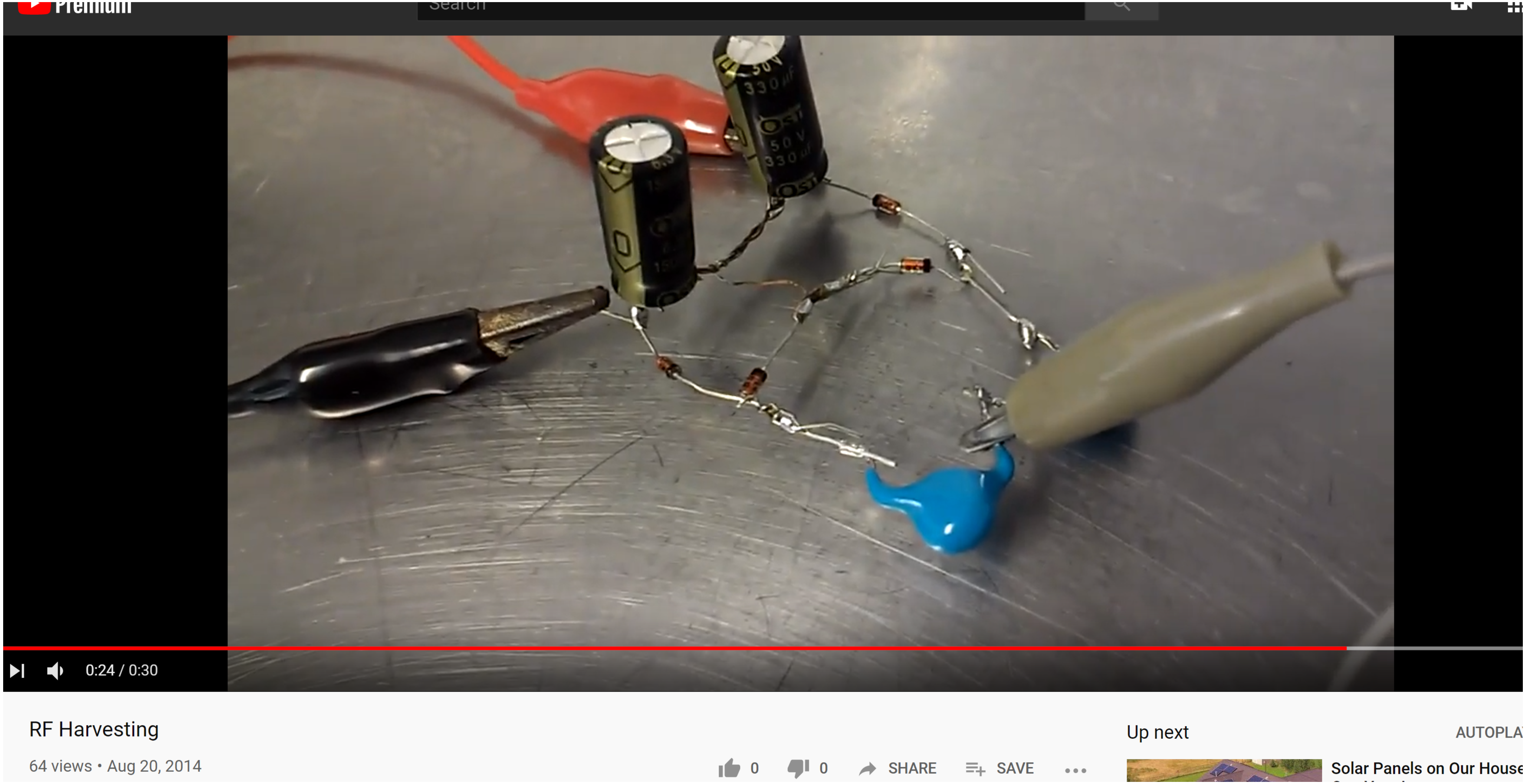
Premium
Search
50V
330µF
50 V
330µF
0:24 / 0:30
Up next
AUTOPLA
Solar Panels on Our House
0
0
SHARE
SAVE

RF Harvesting

64 views • Aug 20, 2014

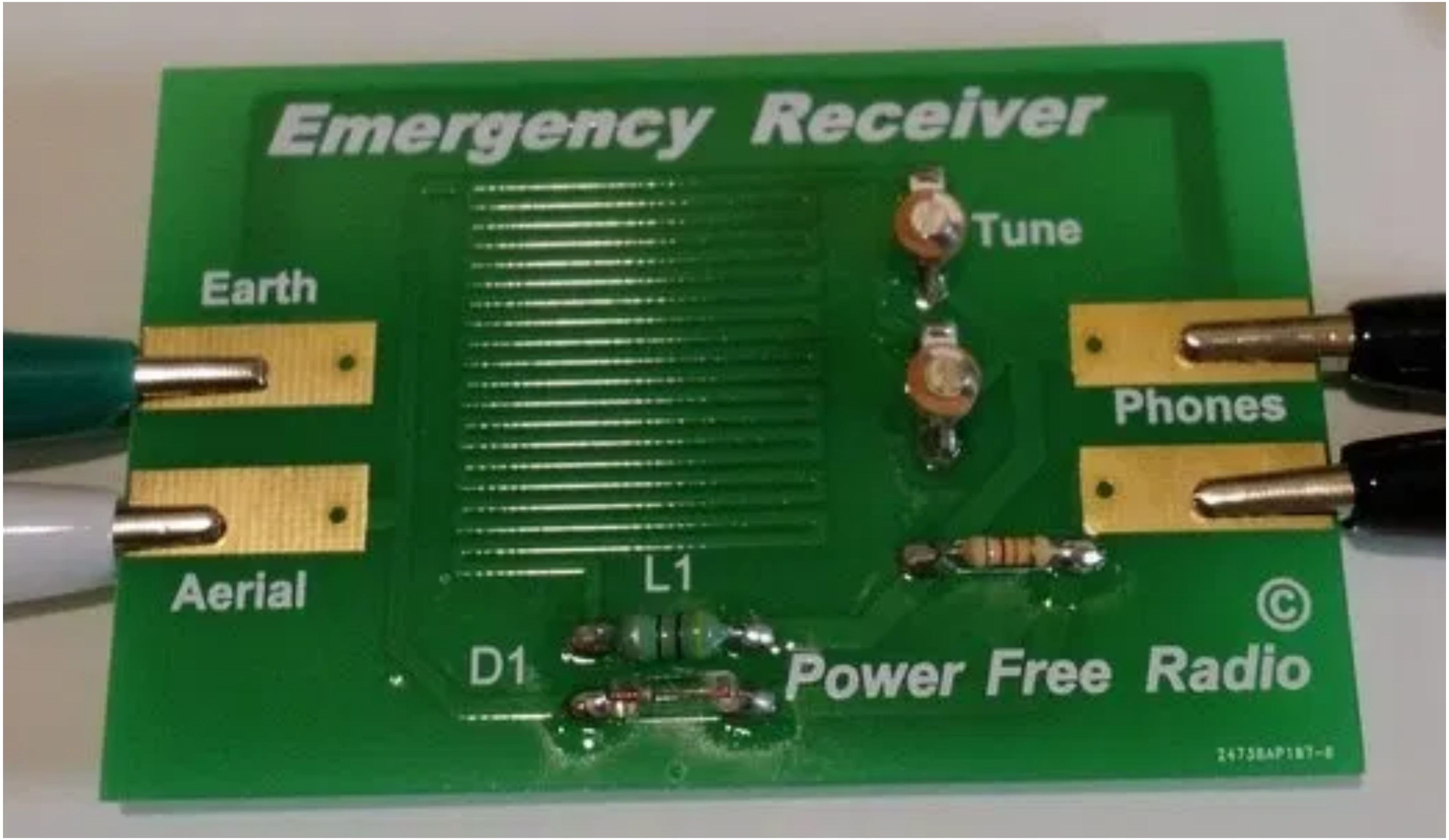
Emergency Receiver
Tune
Earth
Phones
Aerial
L1
D1
©
Power Free Radio

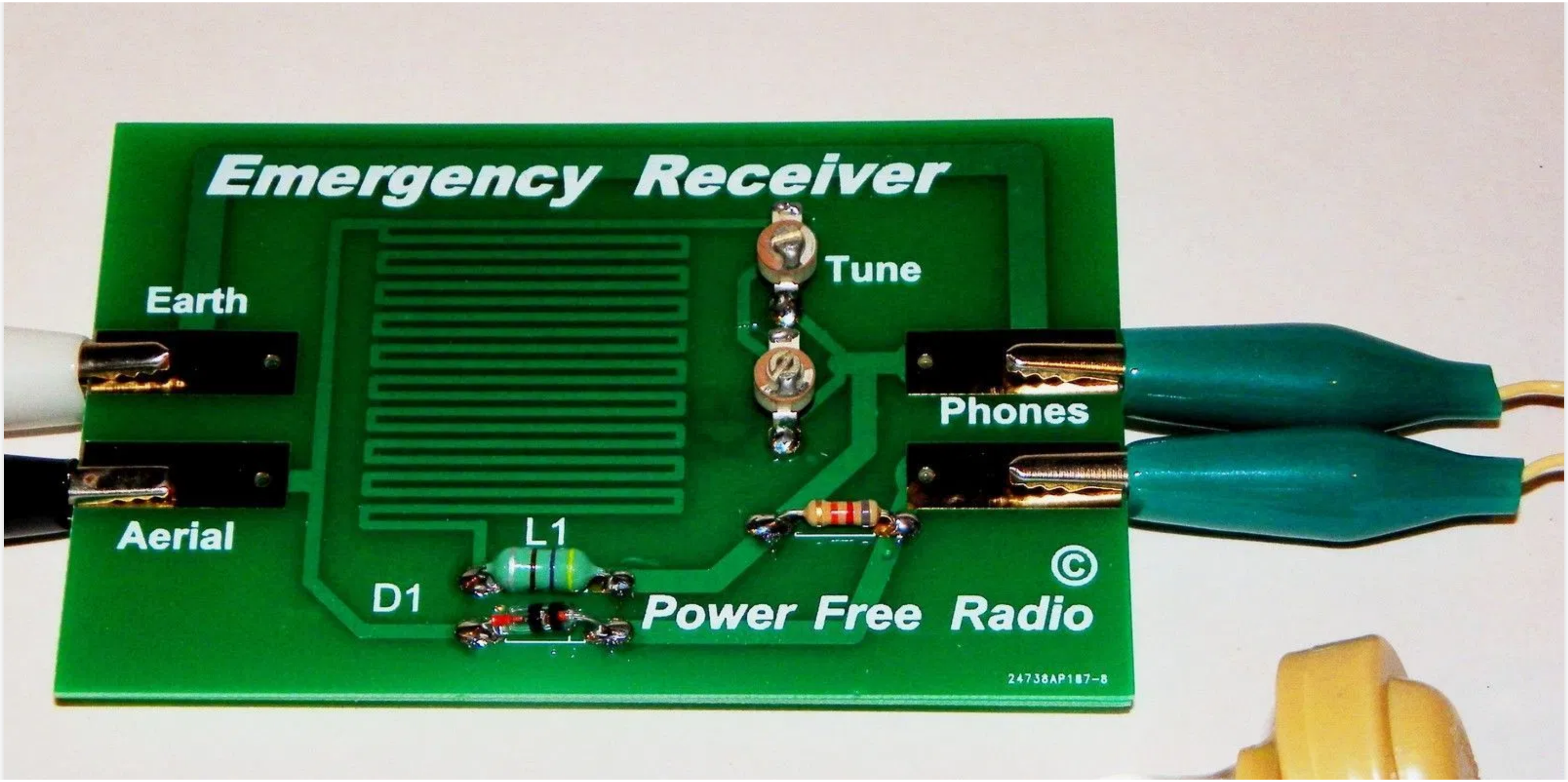
Emergency Receiver
Tune
Earth
Phones
Aerial
L1
D1
©
Power Free Radio
24738AP187-8

The Simplest AM Receiver

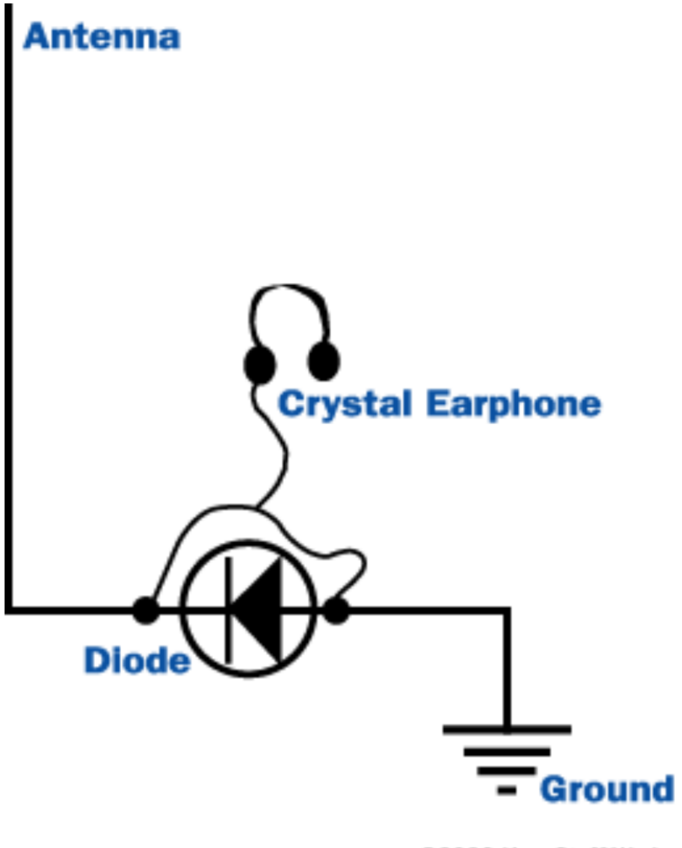

©2000 How Stuff Works

In the case of a strong AM signal, it turns out that you can create a simple radio receiver with just two parts and some wire! The process is extremely simple -- here's what you need:

- **A diode** - You can get a diode for about $1 at Radio Shack. Part number 276-1123 will do.
- **Two pieces of wire** - You'll need about 20 to 30 feet (15 to 20 meters) of wire. Radio Shack part number 278-1224 is great, but any wire will do.
- **A small metal stake** that you can drive into the ground (or, if the transmitter has a guard rail or metal fence nearby, you can use that)
- **A crystal earphone** - Unfortunately, Radio Shack does not sell one. However, Radio Shack does sell a Crystal Radio Kit (part number 28-178) that contains the earphone, diode, wire and a tuner (which means that you don't need to stand right next to the transmitter for this to work), all for $10.

You now need to find and be near an AM radio station's transmitting tower (within a mile/1.6 km or so) for this to work. Here's what you do:

DID YOU KNOW?

DOING THIS POSE FOR 5 MINUTES A DAY...

Holistic Ali

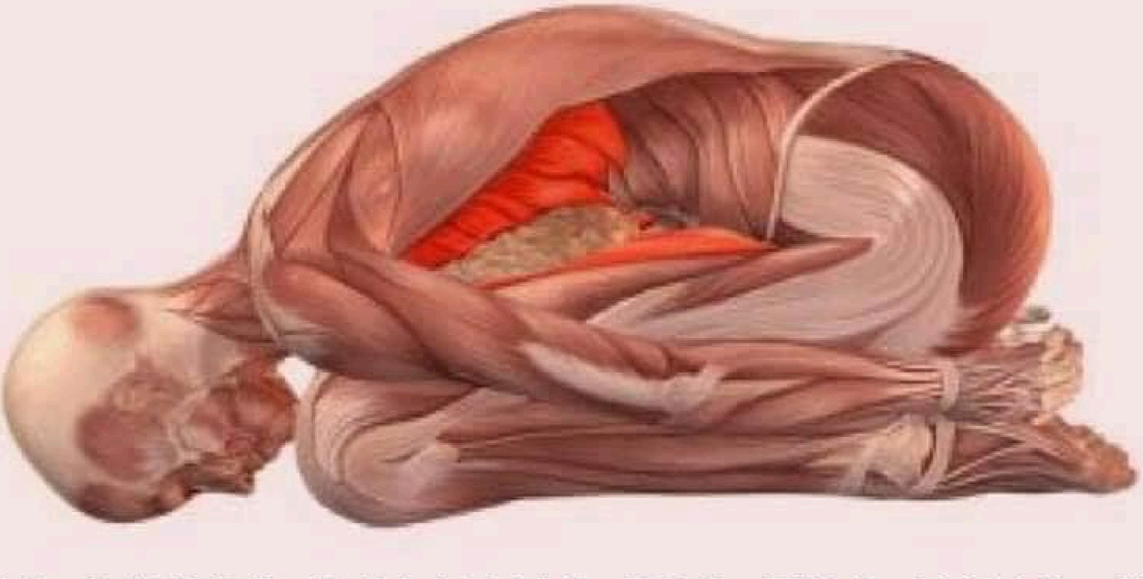

IT'S SUPER CALMING FOR THE MIND, IT'S GREAT FOR YOUR DIGESTION, IT ELONGATES THE LOWER BACK, IT OPENS UP THE HIPS.

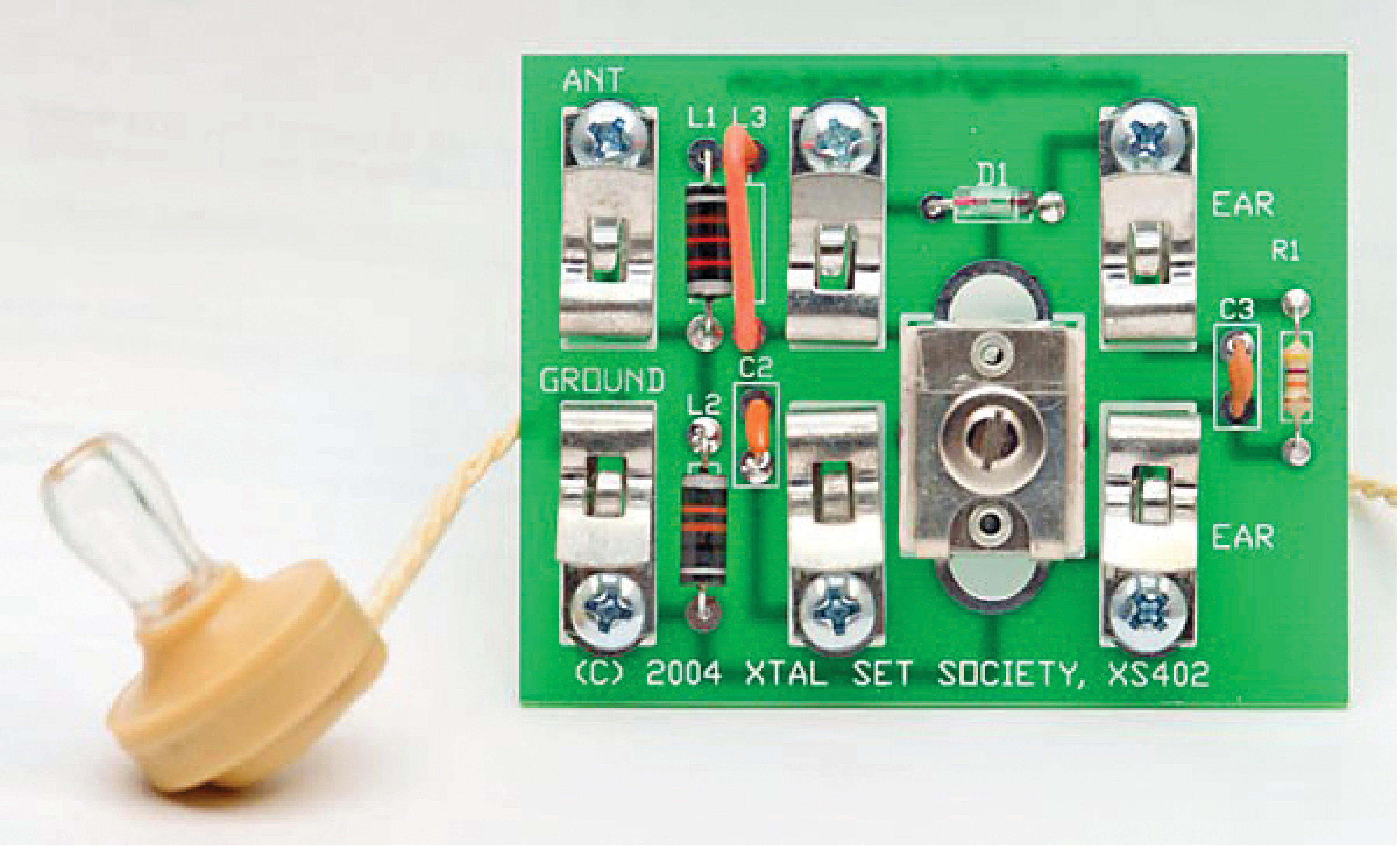
ANT
L1
L3
D1
EAR
R1
C3
C2
GROUND
L2
EAR
(C) 2004 XTAL SET SOCIETY, XS402

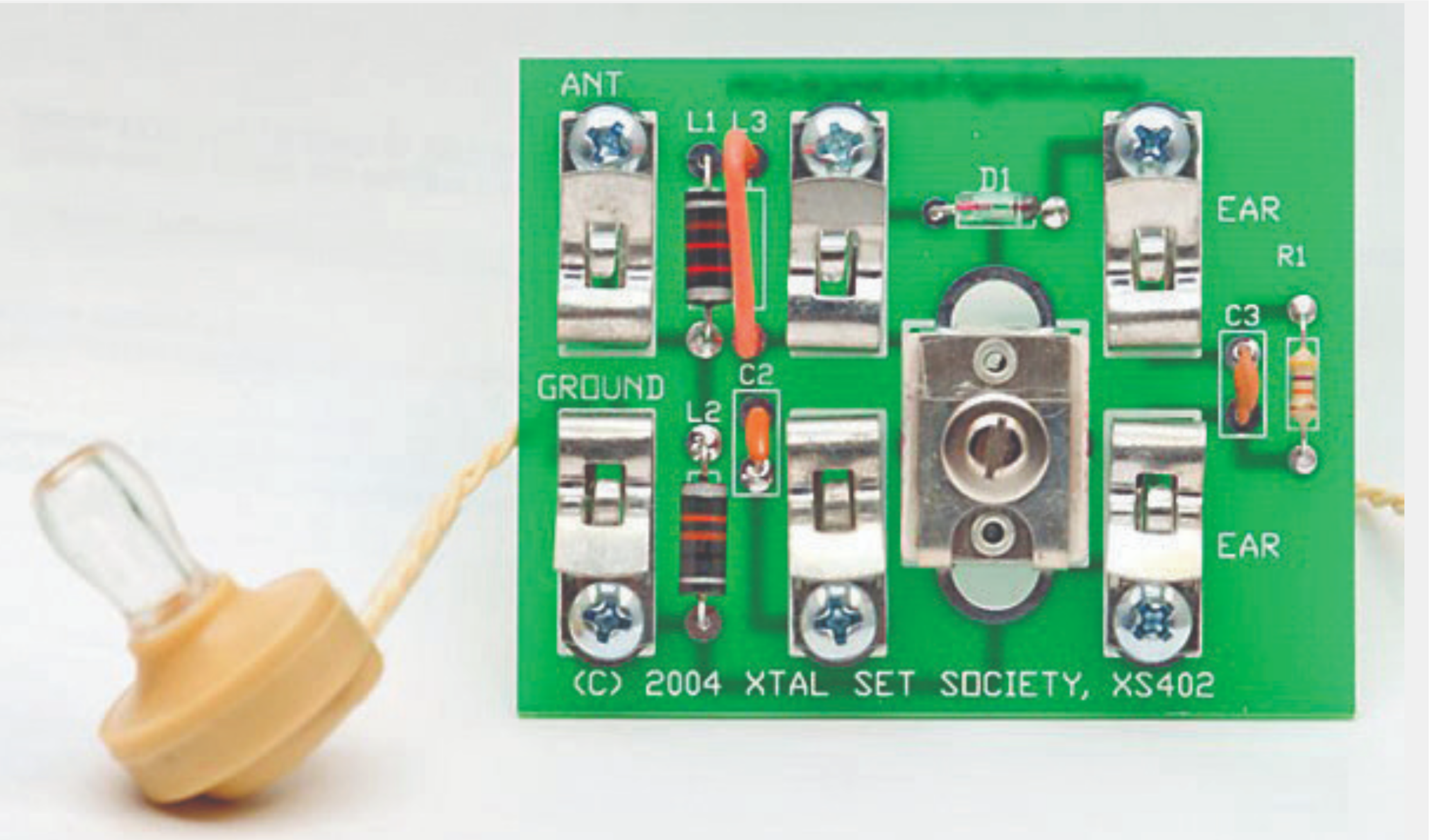
ANT
L1 L3
D1
EAR
R1
C3
C2
GROUND
L2
EAR
(C) 2004 XTAL SET SOCIETY, XS402

3DQ Crystal Radio for AM MW Radio
3DQ 矿石收音机模組
billydiy.blogspot.hk

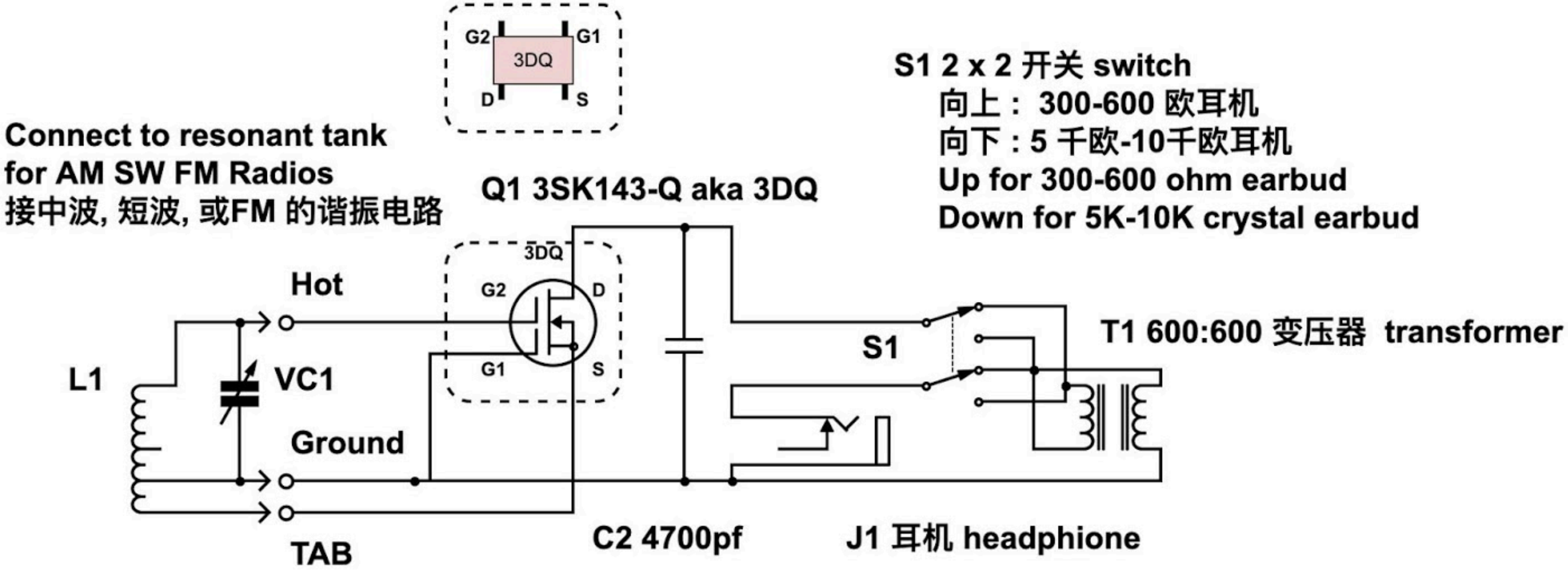
G2
G1
3DQ
D
S
L1
VC1
Hot
Ground
TAB
3DQ
G2
D
G1
S
Q1 3SK143-Q aka 3DQ
C2 4700pf
S1
J1 耳机 headphione
T1 600:600 变压器 transformer

Connect to resonant tank
for AM SW FM Radios
接中波, 短波, 或FM 的谐振电路

S1 2 x 2 开关 switch
向上： 300-600 欧耳机
向下：5 千欧-10千欧耳机
Up for 300-600 ohm earbud
Down for 5K-10K crystal earbud

这是一款适用于AM SW FM的通用MOS FET矿石收音机电路
连接该无线电频带所需的线圈和可变容量就可无需电源收听收音机广播。
3DQ输出阻抗为1.5K至2.5K，具体取决于频率
要配合300-600欧姆高阻抗耳机，请创建一个
4：1自耦变压器，通过连接初级绕组的一端和次级绕组
1：1坡莫合金15H变压器

This is a general MOS FET crystal radio circuit that works for AM SW FM by
connecting the appropriate coils and variable capacity required for that radio band.
3DQ output impedance is 1.5K to 2.5K depending on frequency
To match with a 300-600ohm high impedance Headphone, create a
4:1 auto transformer by connecting one end of the primary with the secondary windings
of a 1:1 permalloy 15H transformer

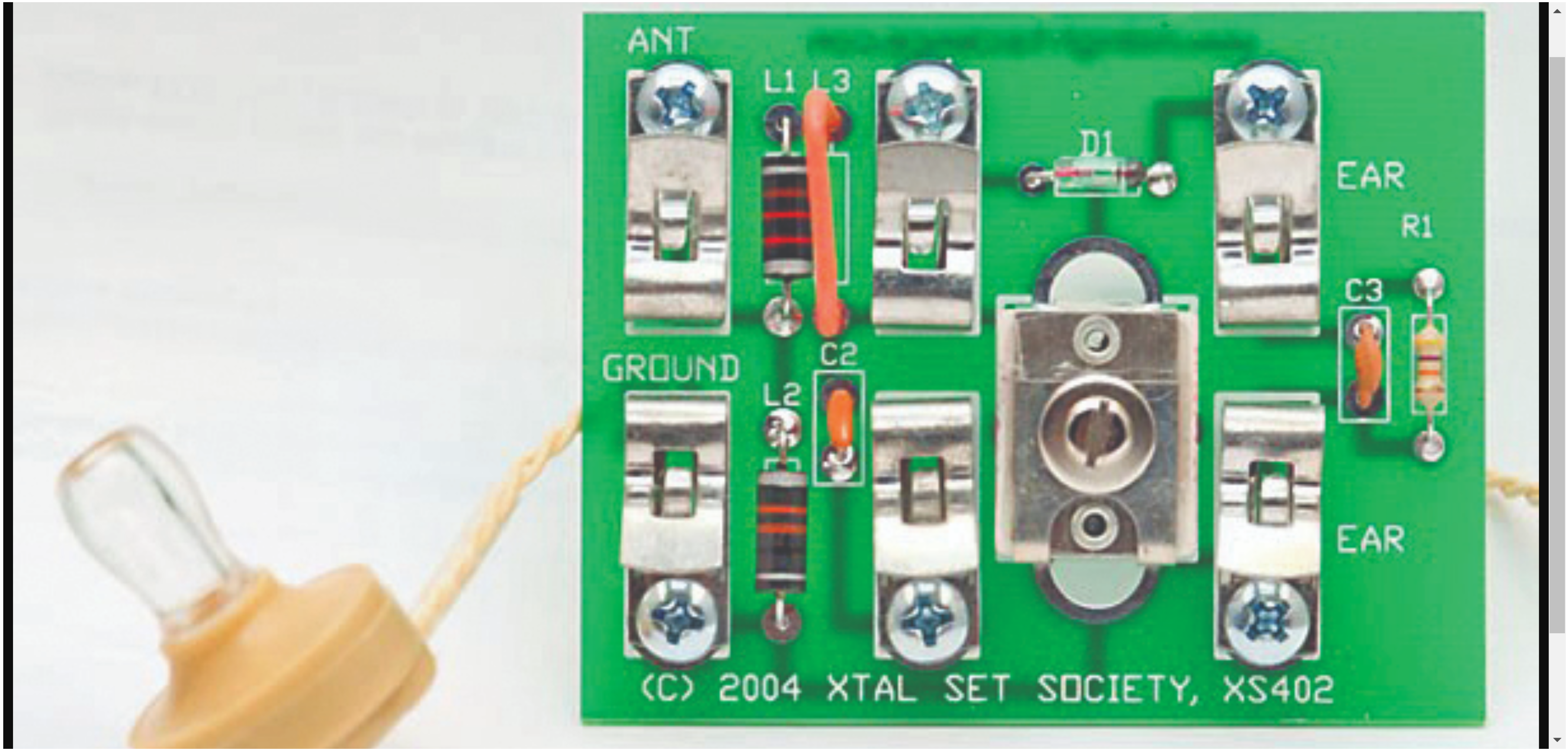
ANT
L1
L3
D1
EAR
R1
C3
GROUND
C2
L2
EAR
(C) 2004 XTAL SET SOCIETY, XS402

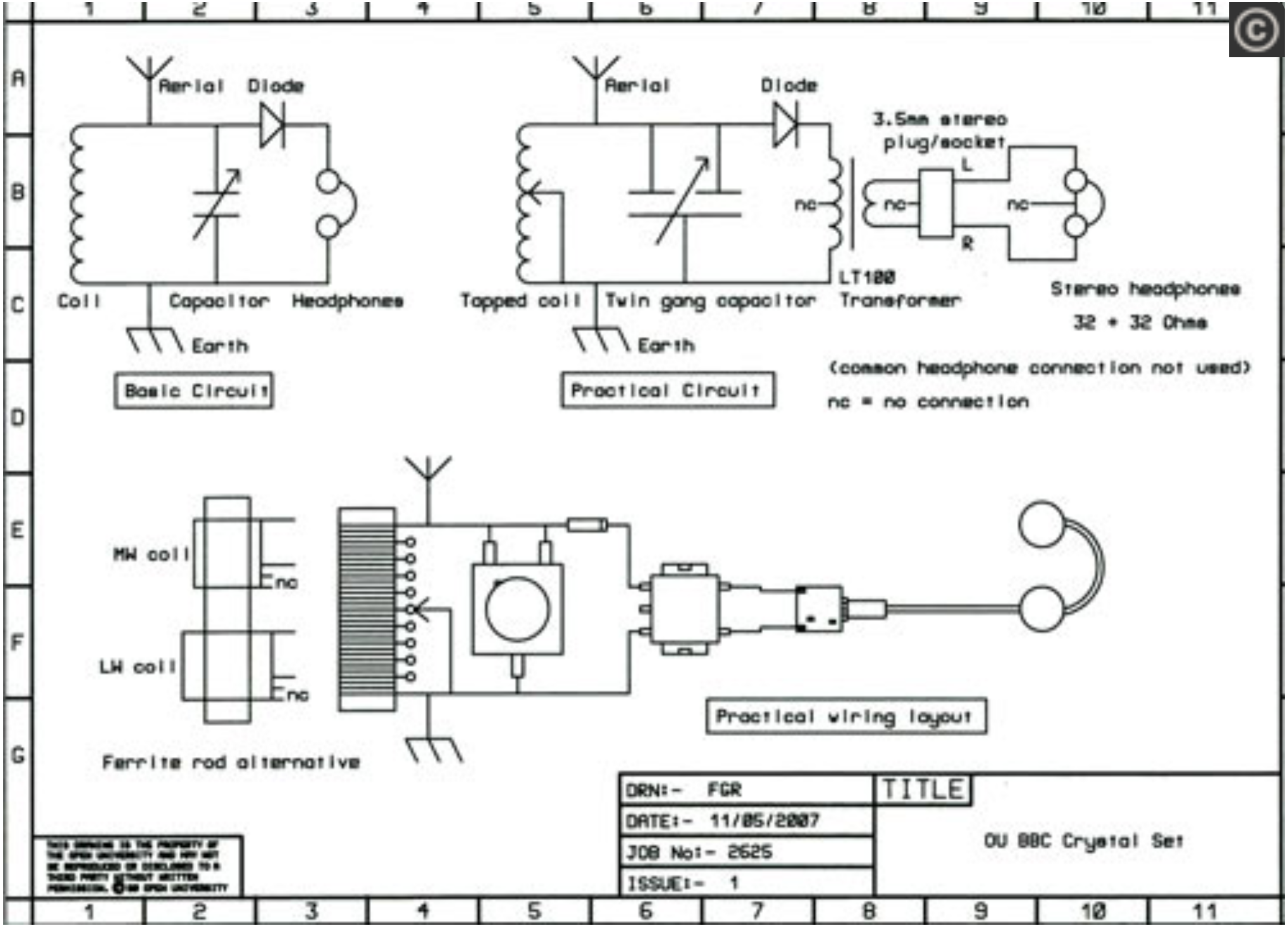
Aerial
Diode
Coil
Capacitor
Headphones
Earth
Basic Circuit
Aerial
Diode
3.5mm stereo
plug/socket
L
R
nc
nc
nc
Tapped coil
Twin gang capacitor
LT100
Transformer
Stereo headphones
32 + 32 Ohms
Earth
Practical Circuit
(common headphone connection not used)
nc = no connection
MW coil
nc
LW coil
nc
Ferrite rod alternative
Practical wiring layout
DRN:- FGR
DATE:- 11/05/2007
JOB No:- 2625
ISSUE:- 1
TITLE
OU BBC Crystal Set
THIS DRAWING IS THE PROPERTY OF THE OPEN UNIVERSITY AND MAY NOT BE REPRODUCED OR DISCLOSED TO A THIRD PARTY WITHOUT WRITTEN PERMISSION. © THE OPEN UNIVERSITY
1
2
3
4
5
6
7
8
9
10
11
A
B
C
D
E
F
G

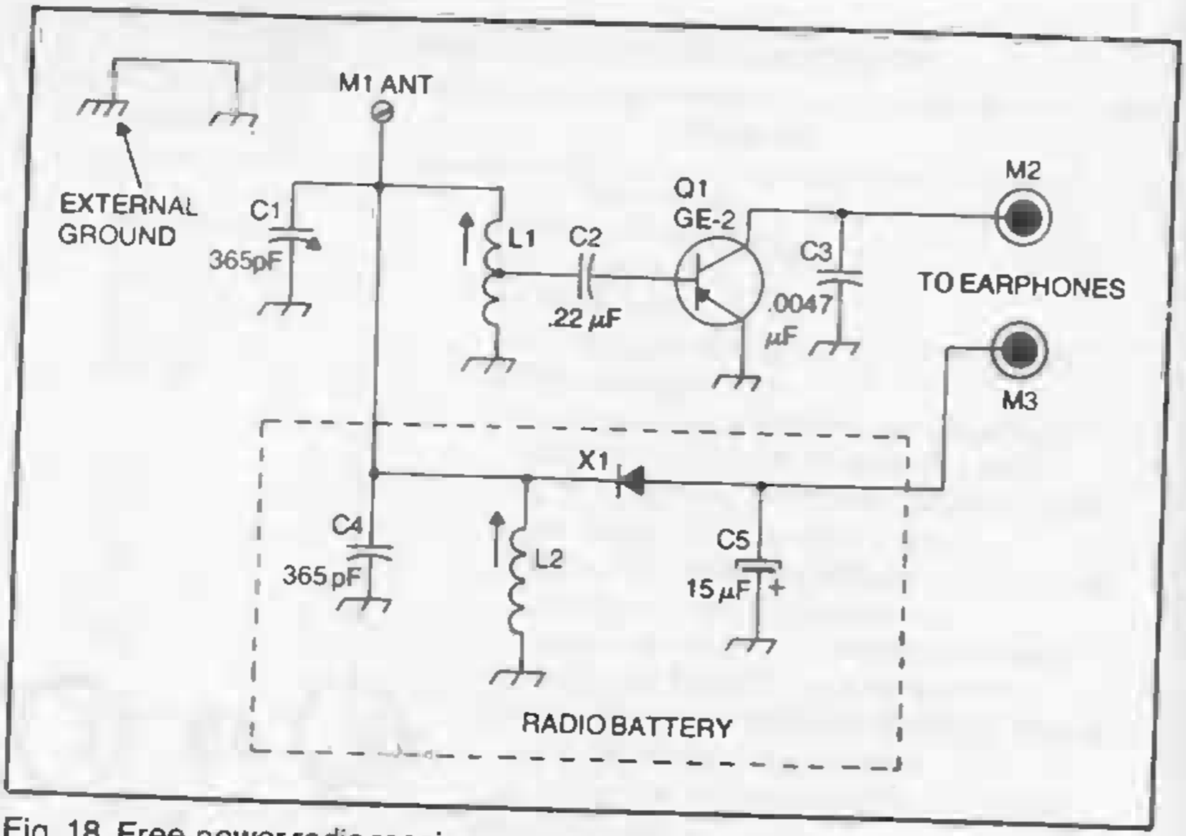
M1 ANT
EXTERNAL GROUND
C1
365pF
L1
C2
.22 µF
Q1
GE-2
C3
.0047 µF
M2
TO EARPHONES
M3
X1
C4
365 pF
L2
C5
15 µF
+
RADIO BATTERY

Fig. 18. Free-power radio receiver.

Table 18. Parts List for Free-Power AM Radio Receiver.

Item No.	Description
C1, C4	365-pF variable capacitors.
C2	.22-µF capacitor.
C3	.0047-µF capacitor.
C5	15-µF, 6-w VDC electrolytic capacitor.
L1, L2	Loopsticks. (Superex VLT-240 or equiv.)
M1	Binding post.
M2, M3	Test jacks.
Q1	GE-2 transistor.
X1	1N38B diode.

Table 17. Parts List for Solar TV Reception Booster.

Item No.	Description
C1	56-pF capacitor.
C2, C3, C6	.0022-µF capacitors.
C4	1.5-pF capacitor.
C5	18-pF capacitor.
C7	150-µF, 15-w VDC electrolytic capacitor.
C8, C9	.0022-µF capacitors.
L1	7½ turns No. 24 enameled wire, evenly wound on a 3/16-inch-diameter form.
L2	3 turns No. 24 enameled wire evenly wound on a 3/16-inch-diameter form.
L3	17 turns No. 24 enameled wire evenly wound on a 3/16-inch-diameter form.
L4	11½ turns of No. 24 enameled wire evenly wound on Speer Type E ferrite form.
L5, L6	10 turns No. 24 enameled wire evenly wound on Speer Type E ferrite form.
M1	4½ VDC, with three penlites of NiCd cells.
M2, M3	Binding posts.
M4	AC wall plug with cable clamp.
M5	4½-VDC Solar pack. (International Rectifier No. SP5G26C or equiv.)
Q1	HEP-3 transistor.
R1	51K resistor.
R2	30K resistor.
R3	240-ohm resistor.
R4	110-ohm resistor.
R5	750-ohm resistor.
S1	Spst switch.
T1, T2, T3, T4	Using Ferroxcube K5050-06 ferrite cores, insert 2 turns of special 300-ohm miniature twinlead in each form hole. Pull tight, and connect leads where shown in diagram.
T5.	6.3-VAC filament trnasformer. (Triad F-14X or equiv.)

18

Free-Power AM Radio Receiver

Strange as it may seem, this transistor broadcast-band radio receiver "steals" power from one station to give to another! The principle is basic: By tuning the battery-section antenna coil (L2) to the strongest broadcast station on the band, diode X1 can rectify the rf and convert it into DC current. Naturally, the closer you are to a strong station, the more current the "radio battery" section of your radio receiver will be able to supply. Once you have found this spot, the DC current is passed on to power the transistor circuit which acts as a genuine receiver, with the full tuning it affords. See Fig. 18 and Table 18.

The basic consideration is a good antenna and ground, the latter preferably being made to a water pipe or solid external ground composed of a pipe driven at least 4 feet into moist earth. This procedure not only ensures maximum signal pickup for the radio-battery portion of the circuit, but also provides best results for the GE-2 receiver circuit.

Once completed, just tune the radio battery as explained in the first paragraph and calibrate your receiver by adjusting L1 so that the bottom of the band occurs when C1 is fully meshed. If you have a local broadcast station operating near 540 kHz, this simplifies things tremendously. Once the calibration procedure is complete, forget entirely about adjusting L1 and do all your listening by tuning C1.

17

Solar TV Reception Booster

Here is a versatile rooftop TV reception booster that will add a full 18 dB of gain to the incoming TV signals at your receiver. This means no more ghosts, clearer pictures, and perhaps even reception of a distant station you could not get before. Best of all, this gadget is powered by an inexpensive solar cell that will virtually run forever without requiring replacement. Additionally, the circuit is designed with a charger circuit and reserve pack.

You can build it as is, or only include as much as you would like. For example, if you break the circuit where the X's are indicated, your booster will thrive entirely off its solar battery. If you break it at the two Y's, the solar cell charges a 4½-volt battery pack during the day so the booster will work just as well at night as in direct sunlight. If you include all the circuit (taking into account the indoor AC supply shown), you can occasionally recharge the 4½-volt pack just in case you have had a protracted period of rain or overcast weather which prevented the solar cell from sending its energy down to the reserve battery.

Construction is not complicated, but since you are dealing with both vhf and uhf signals, you will have to keep all leads extremely short and direct. In fact, you would do best to cram the parts close together so as to cut down on the length of interconnecting wires. Looking at the schematic, you will note that Q1 must be grounded; this is no problem; however, since the HEP-3 comes with four leads instead of the regular three. Turning the transistor upside down, you will notice an arrangement of leads as shown and identified in Fig. 17.

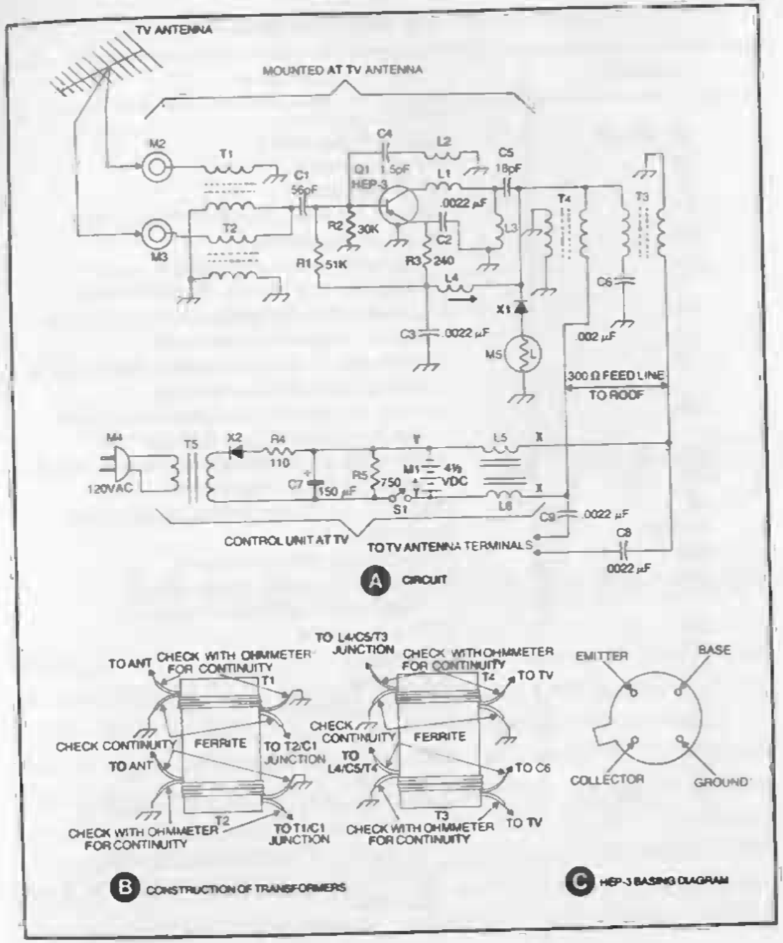
TV ANTENNA
MOUNTED AT TV ANTENNA
M2
M3
T1
T2
C1
56pF
C4
1.5pF
Q1
HEP-3
L2
L1
C5
18pF
.0022 µF
R2
30K
R1
51K
C2
R3
240
L3
L4
X1
C3
.0022 µF
M5
L
T4
T3
C6
.002 µF
300 Ω FEED LINE
TO ROOF
M4
120VAC
T5
X2
R4
110
C7
150 µF
R5
750
M1
4½
VDC
S1
Y
Y
L5
L6
X
X
C9
.0022 µF
C8
0022 µF
CONTROL UNIT AT TV
TO TV ANTENNA TERMINALS
A
CIRCUIT
TO ANT
CHECK WITH OHMMETER FOR CONTINUITY
T1
CHECK CONTINUITY
FERRITE
TO T2/C1 JUNCTION
TO ANT
T2
CHECK WITH OHMMETER FOR CONTINUITY
TO T1/C1 JUNCTION
TO L4/C5/T3 JUNCTION
CHECK WITH OHMMETER FOR CONTINUITY
T4
TO TV
CHECK CONTINUITY
FERRITE
TO L4/C5/T4
TO C6
T3
CHECK WITH OHMMETER FOR CONTINUITY
TO TV
B
CONSTRUCTION OF TRANSFORMERS
EMITTER
BASE
COLLECTOR
GROUND
C
HEP-3 BASING DIAGRAM

Fig. 17. Solar TV reception booster.

Take care in constructing your coils that you duplicate exactly those turns and techniques recommended in Table 17.

Once complete, you can test out your booster by hooking it to the TV set and placing a lit 60-watt lamp bulb in rather close proximity to M5. As the lamp is turned on, you should get a much-improved picture. Adjust the best reception on all channels by tuning L4 and the coupling between L1 and L2. Now mount at the TV antenna, positioning M5 for best sunlight hits.

Hint: Leave the AC plug plugged in at all times. It will not draw power until you throw switch S1 onto "charge."

Solar TV Reception Booster

Here is a versatile rooftop TV reception booster that will add 18 dB of gain to the incoming TV signals at your receiver. means no more ghosts, clearer pictures, and perhaps reception of a distant station you could not get before. Best of this gadget is powered by an inexpensive solar cell that virtually run forever without requiring replacement. Addition the circuit is designed with a charger circuit and reserve pack.

You can build it as is, or only include as much as you w

Tapping into Natural Radio

Figure 1 shows suggestions on utilizing a tree as an antenna. For HF, I suggest using the tallest tree available. Perhaps shorter trees would be better for VHF-UHF as in the 1990 *73* article discussed above—or perhaps not. As with any outside antenna it is wise to utilize some form of lightning protection.

RADIO RIDDLES

Last month:

I said, "Let's say that we could stain some radio waves, frozen at one moment in time, with some kind of dye so that we could actually see them. Of course we can't do this, except in our imagination. But if we did, how would the waves appear to us?"

For the area occupied by the waves at any one instant in time we would be trying to visualize the variations in the signal's electrical-field strength, and the reversals in field orientation (showing when the RF current changes direction of flow twice each cycle). This wouldn't look like a wavy line, or a sine wave graph as is represented in most texts on radio.

One way to imagine seeing the frozen waves is to envision that the space which the radio waves fill will show a darker shade of gray where their field strength is more intense, and a lighter shade where their field strength is weaker. For simplicity, let's consider only direct-wave propagation with no reflections or other impediments to the wave's travel between the transmitting antenna and receiving antenna.

In the space between the two antennas we would see bands of differing shades of gray oriented at right angles to the wave's direction of travel. There would be darker bands, less dark bands just next, even less dark bands next till a relatively light shade of gray was reached. Then somewhat darker bands, then even darker bands, and on to produce a sort of zebra-skin look. The darkest bands would be a half wavelength apart as would the lightest ones.

If we added a green tinge to our gray to indicate one orientation of the electrical field, and a red tinge for the field's other orientation then, starting at the middle of a darker band there would be greenish gray for a half wavelength along the direction of wave travel, a reddish gray for the next half wavelength, and so on.

Of course this picture is an over-simplification. For instance, we haven't talked about the wave's magnetic field, the polarization of the wave, or the spherical shape of the wavefront. But this simplified picture does cover the basic idea of radio waves in space.

This Month:

Let's say that, instead of freezing the waves, we had the advanced eyes and brains of some specialized android so that we could actually watch radio waves zip along on their 186,000 miles per hour flight from the transmitting antenna to our receiving antenna? How would those rascals look then?

You'll find an answer for this month's riddle, and much more, in next month's issue of *Monitoring Times*. 'Til then Peace, DX, and 73.

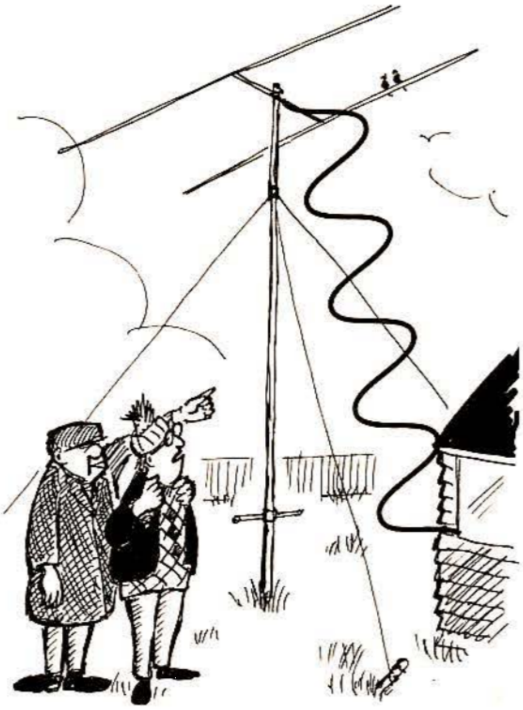

"You know, I thought those were classic standing radio waves until I read Clem Small's column this month. Now, I don't know what the heck they are!"

Motron Electronics 310 Garfield St Suite 4
PO Box 2748
Eugene, Oregon 97402
http://www.motron.com

$59.00 Plus S/H

DTMF DECODER

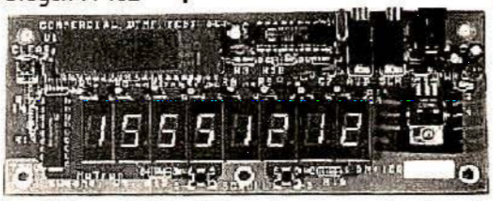

The **TDD-8X** features a large 8-digit LED display and decodes all 16 DTMF digits. The 104 character memory is viewed, without loss of data, by scrolling either left or right. Although a computer is not required, the serial ASCII output allows for automatic logging of decoded number groups. ToneLog ™, our IBM/Compatible logging software, is included with each TDD-8X purchase.

TDD-8X Assembled & Tested PCB $ 59.00 **PMK-1** Plastic Mounting Kit $15.00
CAB-1 Audio & Computer Cables 20.00 **PS-12** 12VDC AC Power Adaptor 10.00

Visa, MasterCard, American Express, Discover and Government Purchase Orders accepted.
COD on Cash or Money Order basis only.
Minimum Government Purchase Order: $99.
S/H: $8 USA; $11 Canada; $16 Foreign COD: $5.
Premium Shipping available at additional charge

Info: (541) 687-2118 **Orders: (800) 338-9058** Fax: (541) 687-2492

Prehistoric Radio!

FIGURE 1:

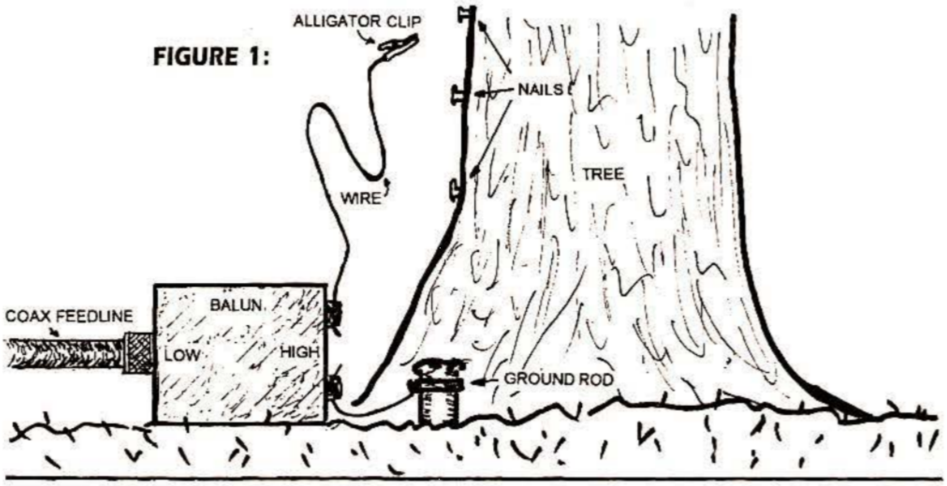

Connecting a feedline to a tree antenna. Nail small, thort, thin nails into the tree about every two or three feet, up to 10 or 15 feet high. Connect the clip to the various nails until best results are achieved. The balun is a 4:1 ratio or higher, high side to the tree.

If you've been following the "Below 500 kHz" column, you know that natural phenomena right here on earth have been producing, transmitting, and receiving radio signals since time began. Lightning bolts, and some other mighty reactions in nature, produce great amounts of electromagnetic radiation in the light, heat, and the radio portions of the electromagnetic-wave spectrum. Lightning discharges from tropical storms in particular are continually producing wide-band radio waves that propagate worldwide. You receive these signals as bothersome static noise on the LF, MF, and on HF bands.

Weak-signal operators who do moon-bounce work are only too well acquainted with the radio signals which our own sun transmits. These noisy signals can totally mask over moon-bounce signals when the sun and moon are both in the same direction from moon-bouncer's antenna. Radio waves known as "cosmic noise" are continually bombarding our antennas with radio signals generated by natural reactions occurring in galactic space. As a matter of fact, the discovery that the source of this particular radio noise lay outside our solar system marked the beginning of radio astronomy. In radio astronomy, naturally occurring radio signals from space are studied in an attempt to learn more about the cosmos.

OK, so there were radio signals before we were around to know about it, but what about the reception of those waves? Who received those signals? No person received them, but they were received by every naturally-occurring conductor to which they propagated. And, once trees evolved, they provided natural, grounded-vertical Marconi antennas to receive these waves.

Well, but where was the receiver for those antennas? The resistance of the tree's "body" was the "receiver." This resistance dissipated much of the received energy as heat rather than as sound, like the loudspeaker does in our HF receivers.

Of course, natural transmission and reception still goes on today. Have you asked the trees near your home if they have received any RF signals lately? I have. And the trees which I asked told me that they receive lots of signals—and I haven't gone off my rocker! They receive lightning-generated signals just as they have since time immemorial, and they also receive the signals which man puts onto the airwaves today.

How did I ask the trees? It's simple. I just connected a feedline to a tree, and connected the line to my receiver. I received a multitude of signals across the LF, MF, and HF bands. If I had tried other bands I would have, no doubt, received signals on those bands, too. As a matter of fact, a tall, live tree makes a decent LF-MF-HF antenna.

■ Tree-Antennas Throughout Radio's History

Old time radio expert Squires reported using trees as receiving aerials for distances of 50 kilometers. His method was to hammer a nail into the tree a few yards above the ground, and connect the receiver's antenna input connectors to the nail and to a ground.

Texanna Loomis, daughter of wireless communication pioneer Mahlon Loomis, says in her *Radio Theory and Operating* that "It has been found that a tree can be used for a receiving antenna, preferably an oak, by attaching a lead-in wire to the trunk of the tree."

Another old-time radio expert named Morecroft felt that reception came from the lead-in wire, rather than the tree itself. In my own tests I can't agree with Morecroft, because my lead-in was shielded, grounded coax, the attachment connectors to the tree were short, and the antenna performed much more like a long antenna than a short stub.

In more recent times there have also been reports of tree antennas. Cohen, in the April 1996 issue of *73 Magazine*, reports using a tree antenna on VHF and UHF with "mediocre-to-good results," but no acceptable results for lower bands. Perhaps his use of trees only 20 feet tall limited their performance on lower bands.

Also in *73* (May 1990), JA6HW and JA6AUI report using a 12 foot high tree for both transmitting and receiving on the 10 meter band with good results. Their article also says that hams used live-tree antennas in the 1930's, as did the U. S. Forces in Vietnam. In *Monitoring Times*, April 1989, Dr. Kosta reports that trees "frequently work better than conventional antennas" for television reception. In the September 1990 issue of *Monitoring Times* yours truly reports good results with a tall maple tree on LF through HF.

B1—Battery, 9 VDC, type 2U6
or equiv.
M1—Meter, 0-1-mA DC

56 Simple AM Mod. Monitor

☐ This simple modulation monitor for AM ham or CB transmitters requires no connection to the transmitter. Just position the loop near the final tank or antenna matching coil until the signal is heard in the headphones.

**PARTS LIST FOR
SIMPLE AM MOD. MONITOR**

C1—100-pF disc capacitor
D1—1N914 diode
E1—Magnetic headphone, 2000
ohms or better
L1—Coil, 3 turns on 1½-in. dia.
form, use any thin gauge wire

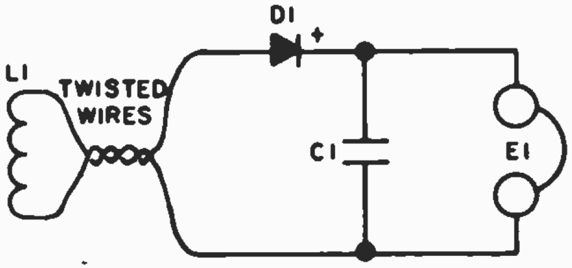

PAPER 53

THE LUCIFER REBELLION

LUCIFER was a brilliant primary Lanonandek Son of Nebadon. He had experienced service in many systems, had been a high counselor of his group, and was distinguished for wisdom, sagacity, and efficiency. Lucifer was number 37 of his order, and when commissioned by the Melchizedeks, he was designated as one of the one hundred most able and brilliant personalities in more than seven hundred thousand of his kind. From such a magnificent beginning, through evil and error, he embraced sin and now is numbered as one of three System Sovereigns in Nebadon who have succumbed to the urge of self and surrendered to the sophistry of spurious personal liberty—rejection of universe allegiance and disregard of fraternal obligations, blindness to cosmic relationships.

In the universe of Nebadon, the domain of Christ Michael, there are ten thousand systems of inhabited worlds. In all the history of Lanonandek Sons, in all their work throughout these thousands of systems and at the universe headquarters, only three System Sovereigns have ever been found in contempt of the government of the Creator Son.

1. THE LEADERS OF REBELLION

Lucifer was not an ascendant being; he was a created Son of the local universe, and of him it was said: "You were perfect in all your ways from the day you were created till unrighteousness was found in you." Many times had he been in counsel with the Most Highs of Edentia. And Lucifer reigned "upon the holy mountain of God," the administrative mount of Jerusem, for he was the chief executive of a great system of 607 inhabited worlds.

Lucifer was a magnificent being, a brilliant personality; he stood next to the Most High Fathers of the constellations in the direct line of universe authority. Notwithstanding Lucifer's transgression, subordinate intelligences refrained from showing him disrespect and disdain prior to Michael's bestowal on Urantia. Even the archangel of Michael, at the time of Moses' resurrection, "did not bring against him an accusing judgment but simply said, 'the Judge rebuke you.' " Judgment in such matters belongs to the Ancients of Days, the rulers of the superuniverse.

Lucifer is now the fallen and deposed Sovereign of Satania. Self-contemplation is most disastrous, even to the exalted personalities of the celestial world. Of Lucifer it was said: "Your heart was lifted up because of your beauty; you corrupted your wisdom because of your brightness." Your olden prophet saw his sad estate when he wrote: "How are you fallen from heaven, O Lucifer, son of the morning! How are you cast down, you who dared to confuse the worlds!"

Very little was heard of Lucifer on Urantia owing to the fact that he assigned his first lieutenant, Satan, to advocate his cause on your planet. Satan was a member of the same primary group of Lanonandeks but had never functioned as a System Sovereign; he entered fully into the Lucifer insurrection. The "devil" is none other than Caligastia, the deposed Planetary Prince of

Very little was heard of Lucifer on Urantia owing to the fact that he assigned his first lieutenant, Satan, to advocate his cause on your planet. Satan was a member of the same primary group of Lanonandeks but had never functioned as a System Sovereign; he entered fully into the Lucifer insurrection. The "devil" is none other than Caligastia, the deposed Planetary Prince of Urantia and a Son of the secondary order of Lanonandeks. At the time Michael was on Urantia in the flesh, Lucifer, Satan, and Caligastia were leagued together to effect the miscarriage of his bestowal mission. But they signally failed.

Abaddon was the chief of the staff of Caligastia. He followed his master into rebellion and has ever since acted as chief executive of the Urantia rebels. Beelzebub was the leader of the disloyal midway creatures who allied themselves with the forces of the traitorous Caligastia.

The dragon eventually became the symbolic representation of all these evil personages. Upon the triumph of Michael, "Gabriel came down from Salvington and bound the dragon (all the rebel leaders) for an age." Of the Jerusem seraphic rebels it is written: "And the angels who kept not their first estate but left their own habitation, he has reserved in sure chains of darkness to the judgment of the great day."

2. THE CAUSES OF REBELLION

Lucifer and his first assistant, Satan, had reigned on Jerusem for more than five hundred thousand years when in their hearts they began to array themselves against the Universal Father and his then vicegerent Son, Michael.

There were no peculiar or special conditions in the system of Satania which suggested or favored rebellion. It is our belief that the idea took origin and form in Lucifer's mind, and that he might have instigated such a rebellion no matter where he might have been stationed. Lucifer first announced his plans to Satan, but it required several months to corrupt the mind of his able and brilliant associate. However, when once converted to the rebel theories, he became a bold and earnest advocate of "self-assertion and liberty."

No one ever suggested rebellion to Lucifer. The idea of self-assertion in opposition to the will of Michael and to the plans of the Universal Father, as they are represented in Michael, had its origin in his own mind. His relations with the Creator Son had been intimate and always cordial. At no time prior to the exaltation of his own mind did Lucifer openly express dissatisfaction about the universe administration. Notwithstanding his silence, for more than one hundred years of standard time the Union of Days on Salvington had been reflectivating to Uversa that all was not at peace in Lucifer's mind. This information was also communicated to the Creator Son and the Constellation Fathers of Norlatiadek.

Throughout this period Lucifer became increasingly critical of the entire plan of universe administration but always professed wholehearted loyalty to the Supreme Rulers. His first outspoken disloyalty was manifested on the occasion of a visit of Gabriel to Jerusem just a few days before the open proclamation of the Lucifer Declaration of Liberty. Gabriel was so profoundly impressed with the certainty of the impending outbreak that he went direct to Edentia to confer with the Constellation Fathers regarding the measures to be employed in case of open rebellion.

Circuit Diagram

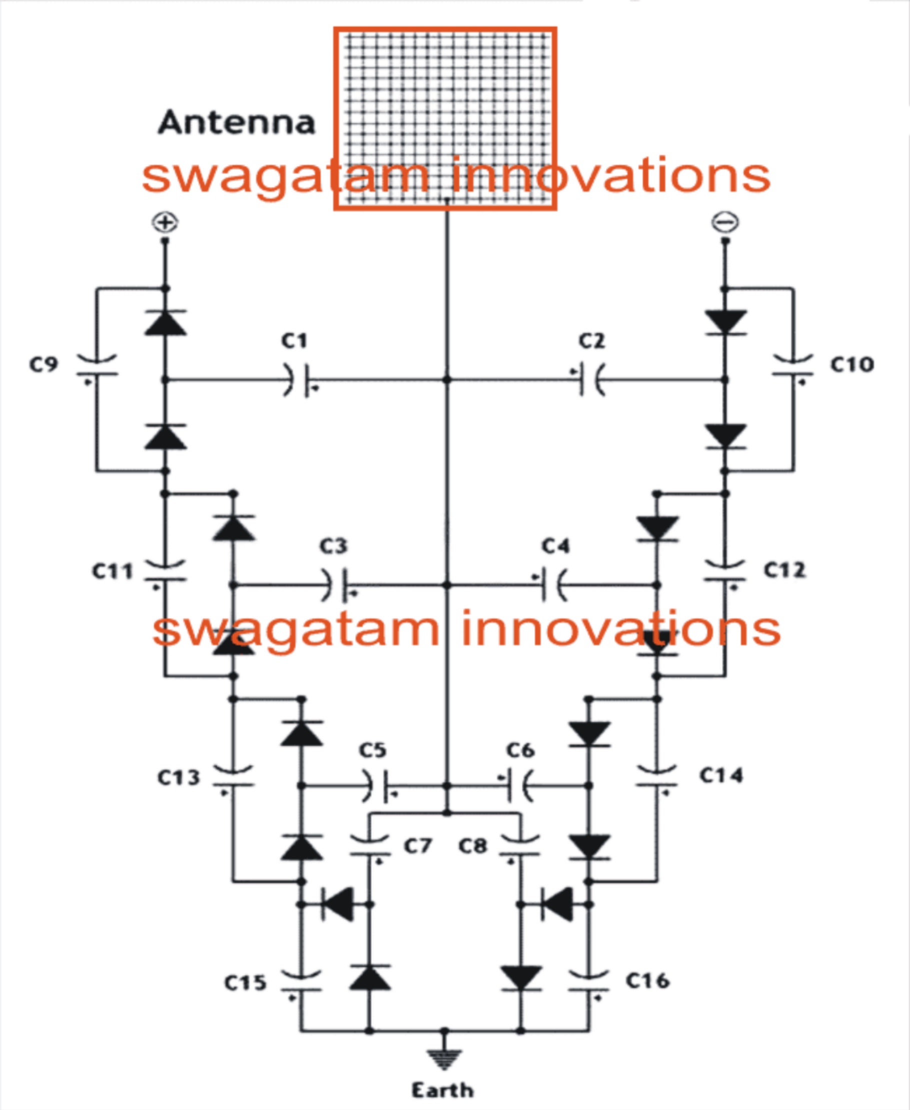
Antenna
swagatam innovations
⊕
⊖
C1
C2
C9
C10
C3
C4
C11
C12
swagatam innovations
C5
C6
C13
C14
C7
C8
C15
C16
Earth

Parts List

All Diodes are 1N4148

C1---C8 = 0.22uF/100V mylar

C9----C16 = 33uF/25V electrolytic

Improving the Free Energy Device

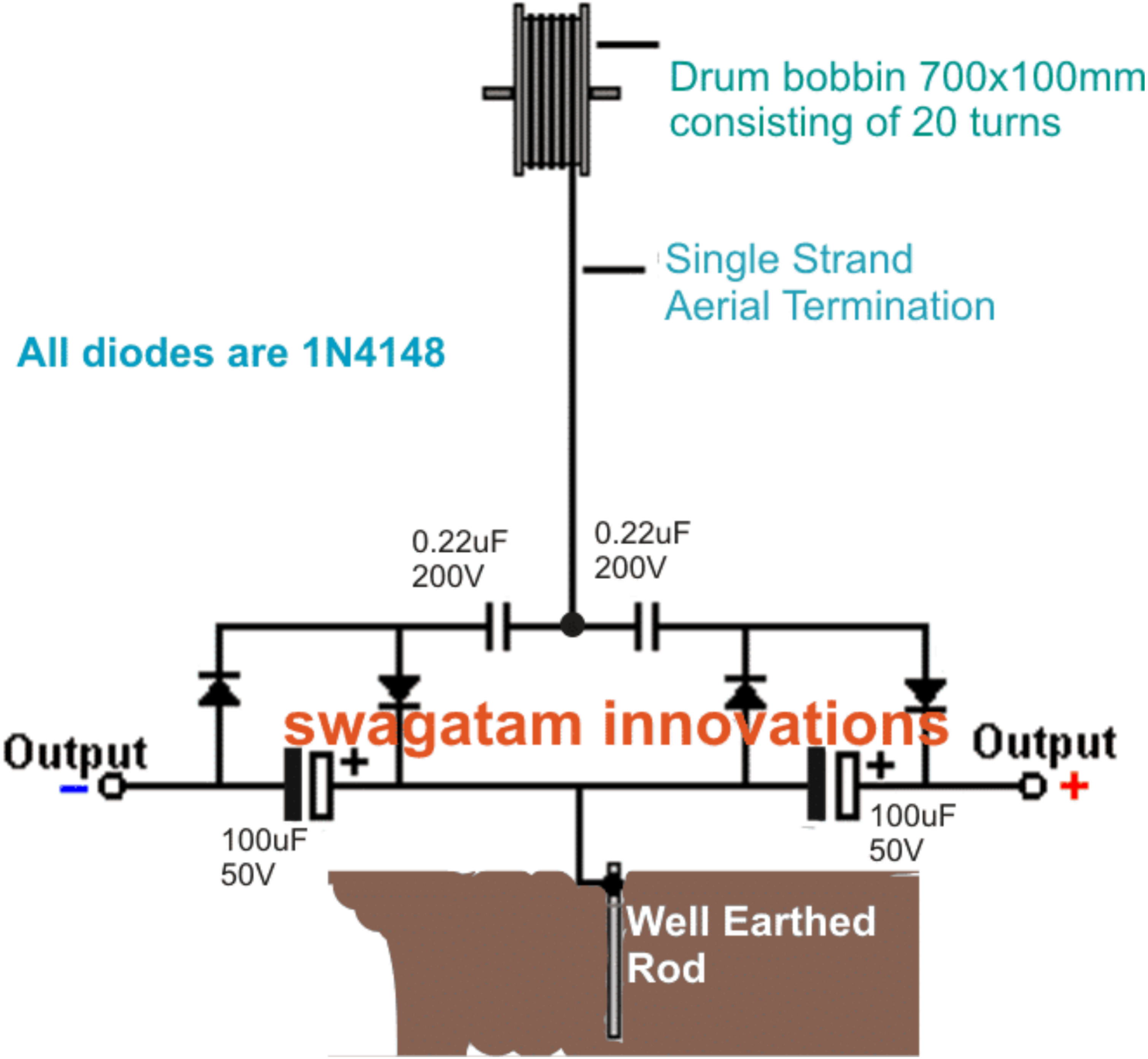
Drum bobbin 700x100mm
consisting of 20 turns
Single Strand
Aerial Termination
All diodes are 1N4148
0.22uF
200V
0.22uF
200V
Output
–
Output
+
100uF
50V
100uF
50V
Well Earthed
Rod
swagatam innovations

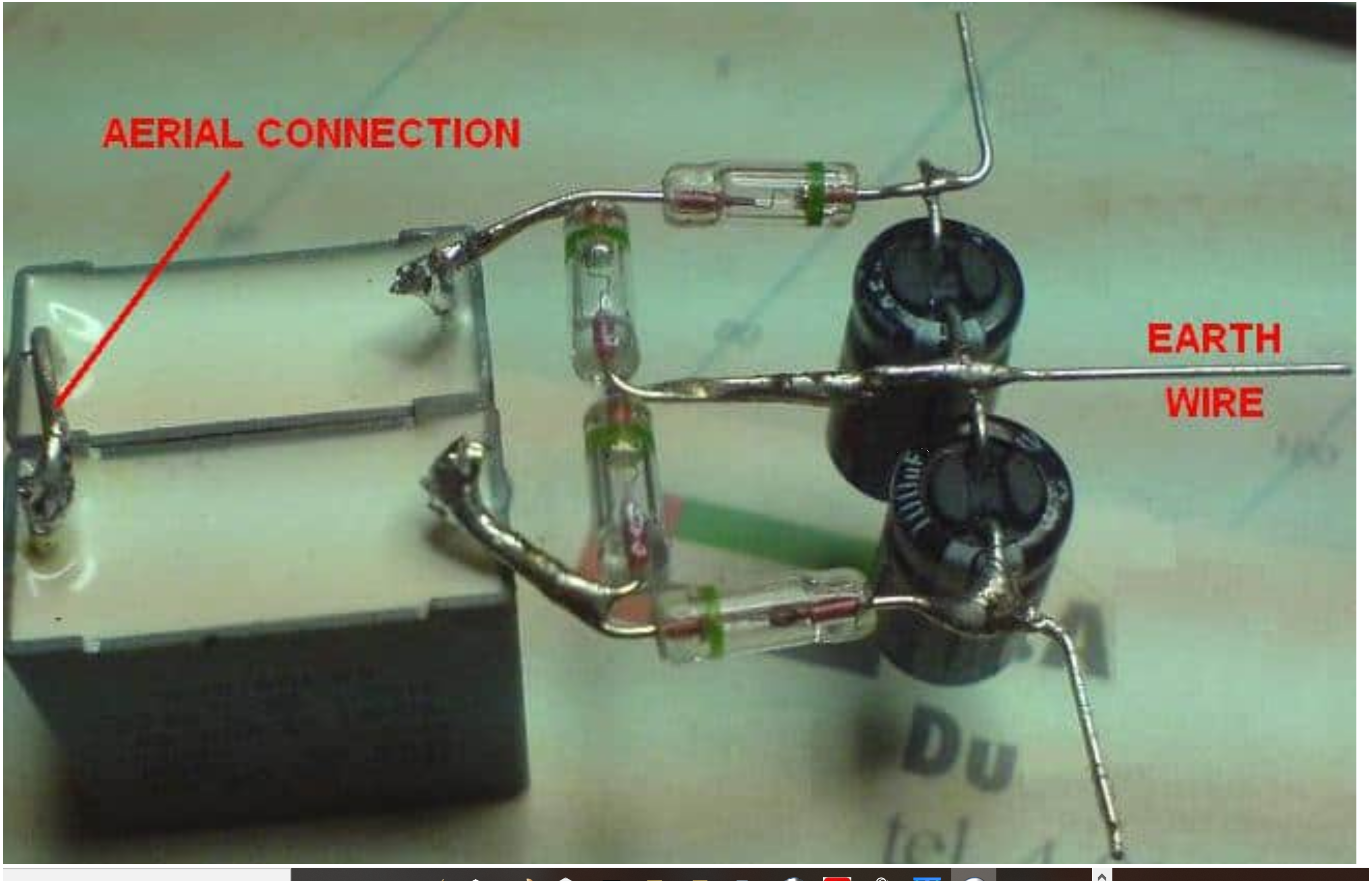
AERIAL CONNECTION
EARTH WIRE

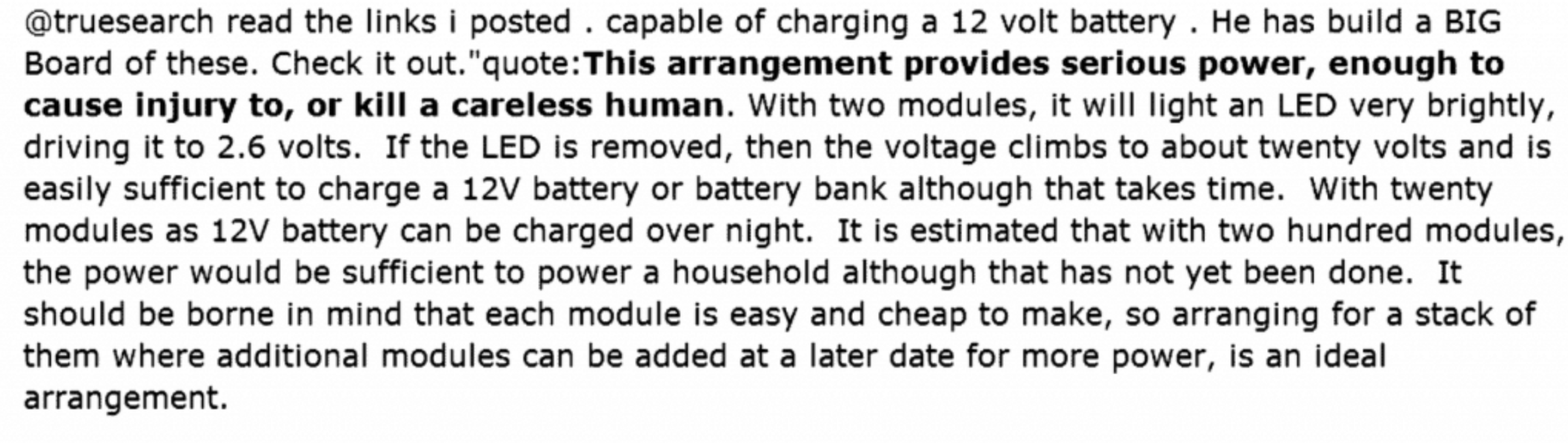

@truesearch read the links i posted . capable of charging a 12 volt battery . He has build a BIG Board of these. Check it out."quote:**This arrangement provides serious power, enough to cause injury to, or kill a careless human**. With two modules, it will light an LED very brightly, driving it to 2.6 volts. If the LED is removed, then the voltage climbs to about twenty volts and is easily sufficient to charge a 12V battery or battery bank although that takes time. With twenty modules as 12V battery can be charged over night. It is estimated that with two hundred modules, the power would be sufficient to power a household although that has not yet been done. It should be borne in mind that each module is easy and cheap to make, so arranging for a stack of them where additional modules can be added at a later date for more power, is an ideal arrangement.

Using Fast Recovery Diodes

In order to generate more electricity, it is possible by stepping up with more diodes. To make it work properly, the type of diodes and the construction of antenna play a crucial role.

To begin with this process, let us first start setting up the antenna. In order to set an antenna properly, there are some key issues to consider.

An antenna should be made of Ferrite and a height of 30 inch rod is an ideal option to setup the antenna for receiving radio waves.

In regard to diodes, the Geranium the lowest loss diodes and a low breakdown junction voltage of ~ 0.2 - 0.4 Volts is ideal, in case you are unable to find the you may use the regular 1N4148, would just work.

A radio wave gets intercepted in areas with higher concentration and traffic congestion. In this kind of situation it is seen that each diode can pull around 30mV.

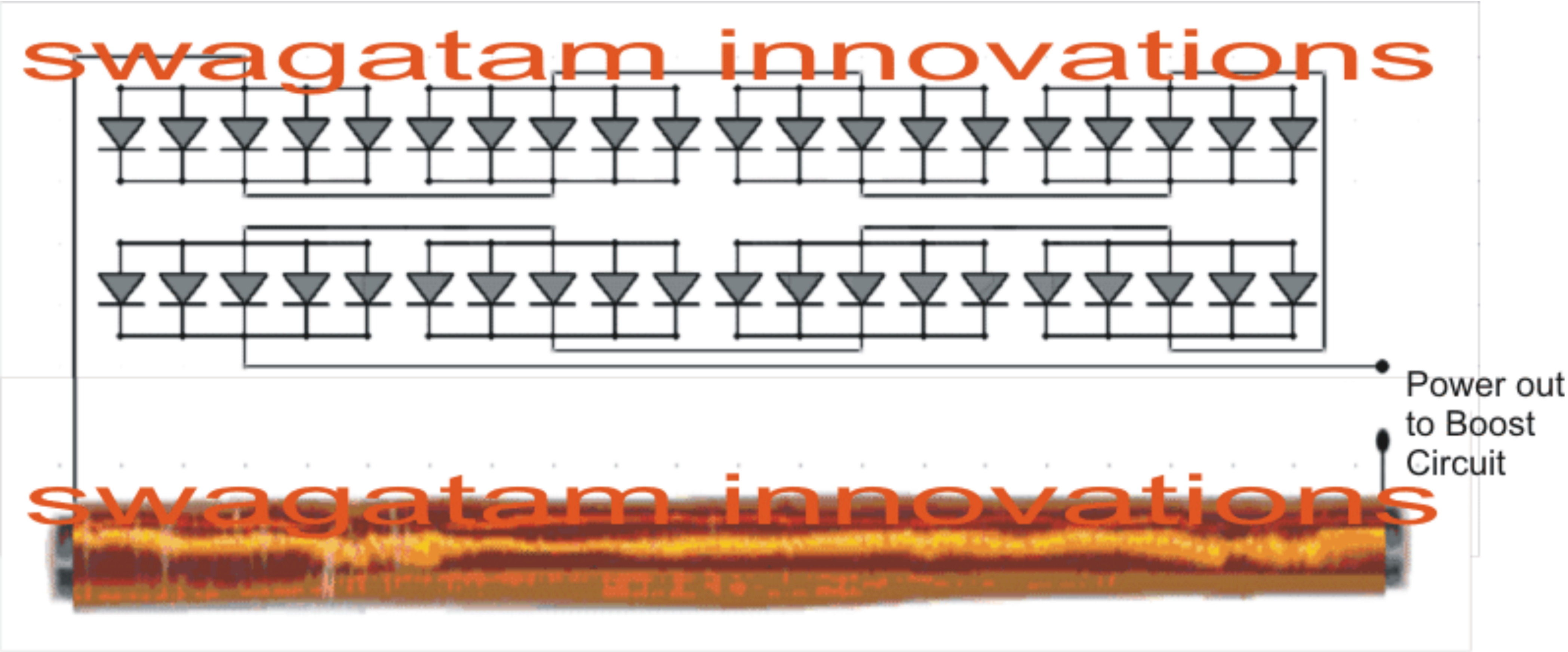
Power out
to Boost
Circuit

swagatam innovations

swagatam innovations

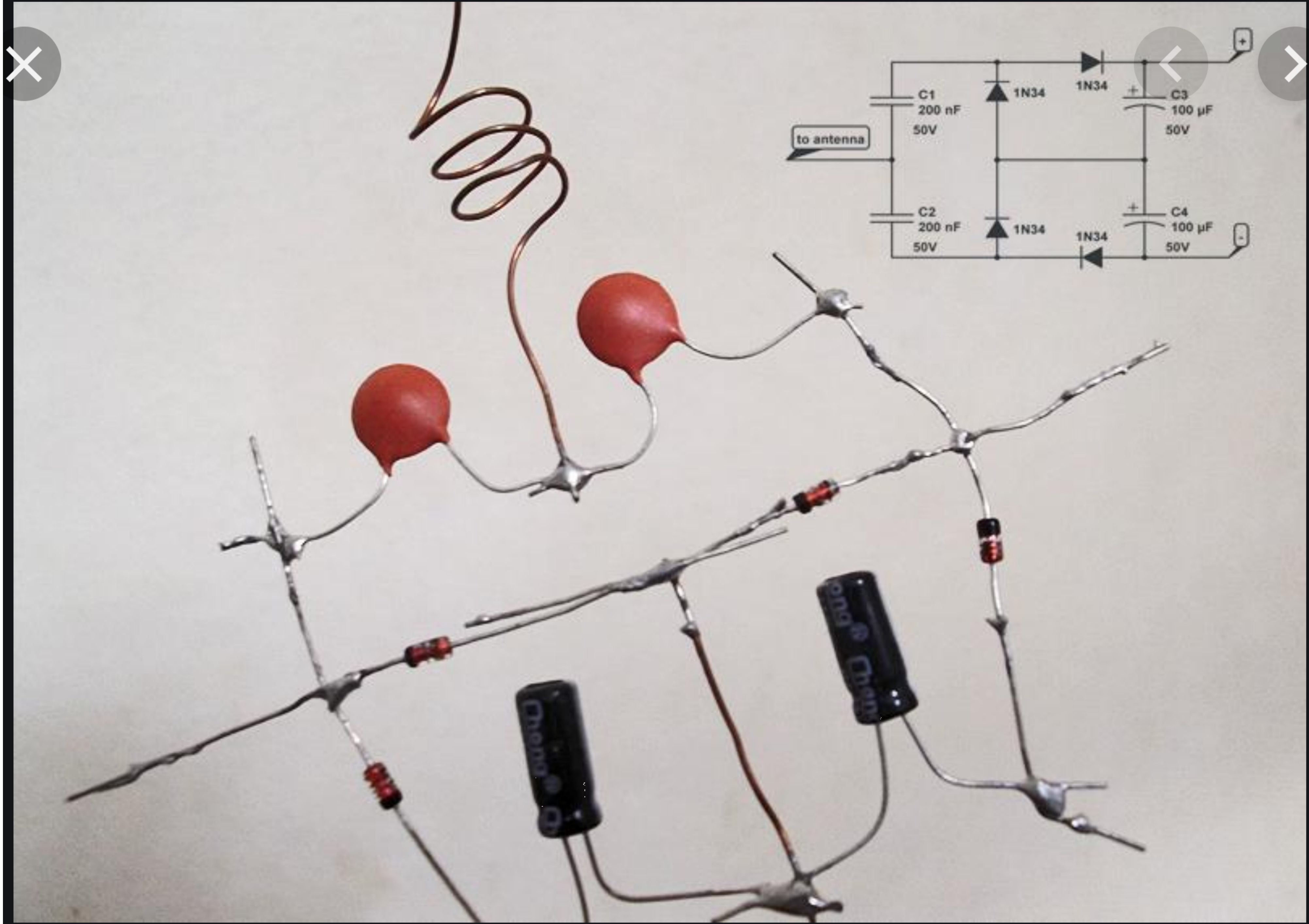
+
1N34
1N34
C1
200 nF
50V
C3
100 µF
50V
to antenna
C2
200 nF
50V
1N34
1N34
C4
100 µF
50V
-

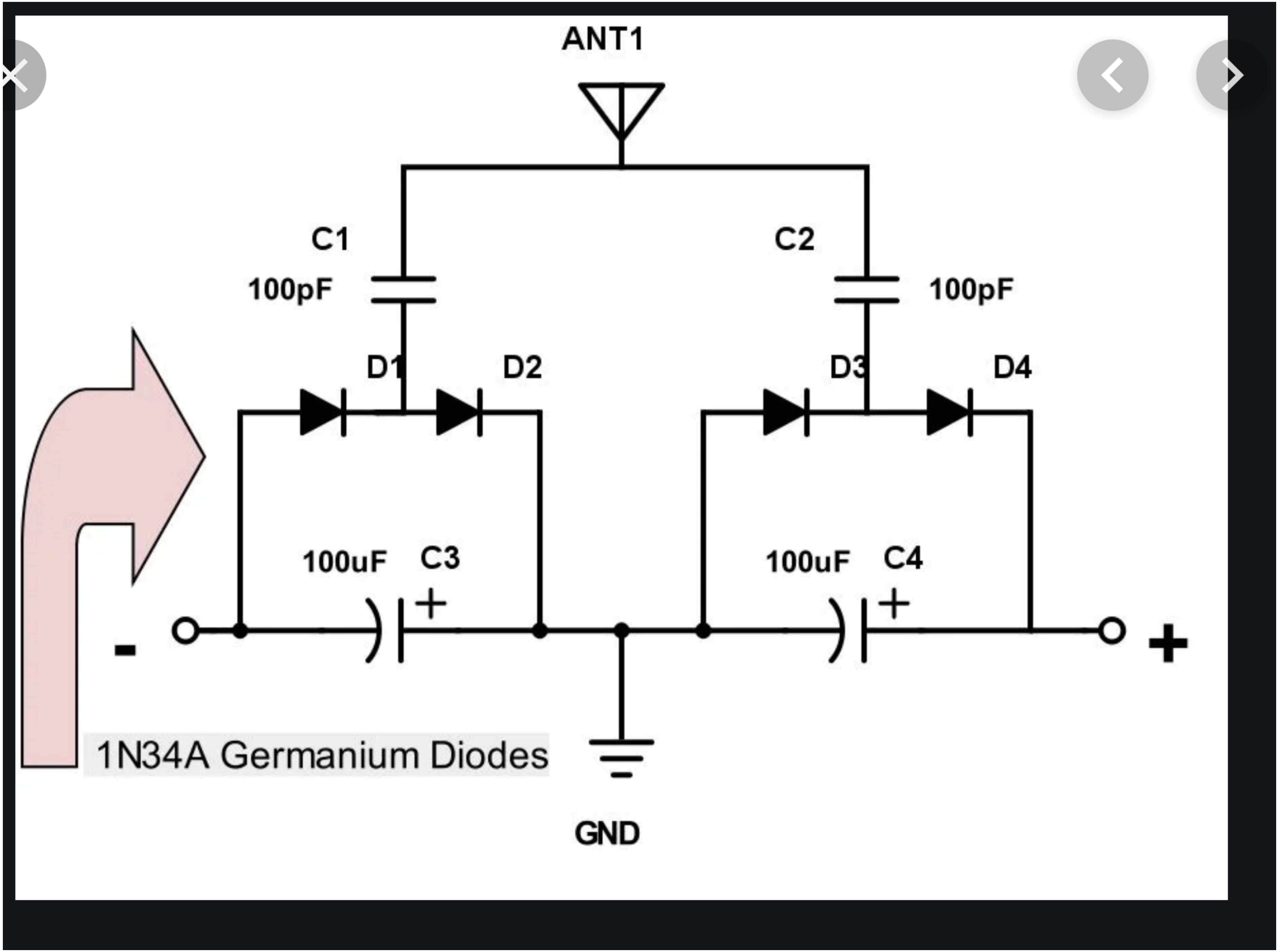
ANT1
C1
100pF
C2
100pF
D1
D2
D3
D4
100uF
C3
100uF
C4
+
+
-
+
1N34A Germanium Diodes
GND

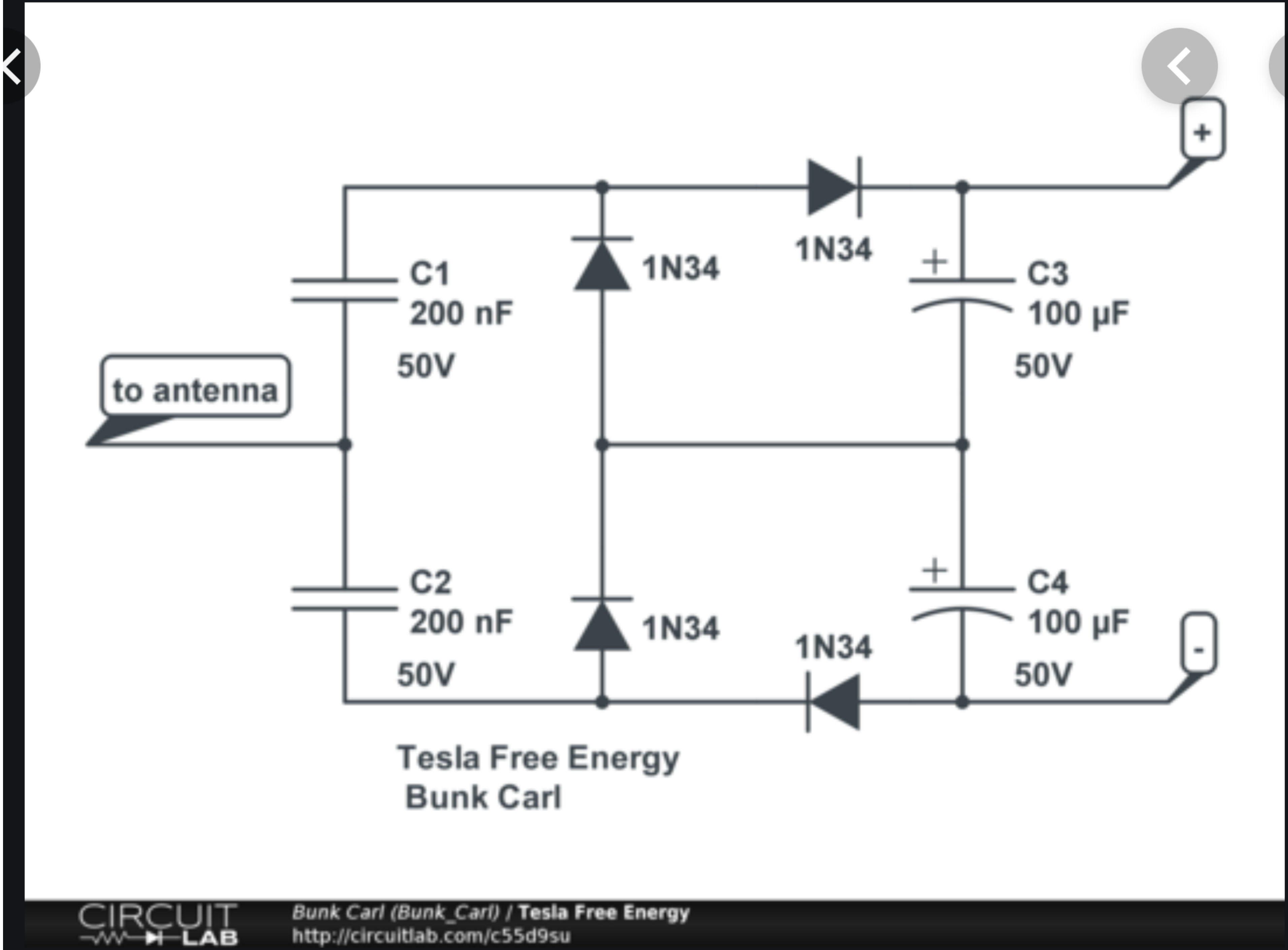
+
1N34
C1
200 nF
50V
1N34
C3
100 µF
50V
+
to antenna
C2
200 nF
50V
1N34
1N34
+
C4
100 µF
50V
-
Tesla Free Energy
Bunk Carl
CIRCUIT LAB
Bunk Carl (Bunk_Carl) / Tesla Free Energy
http://circuitlab.com/c55d9su

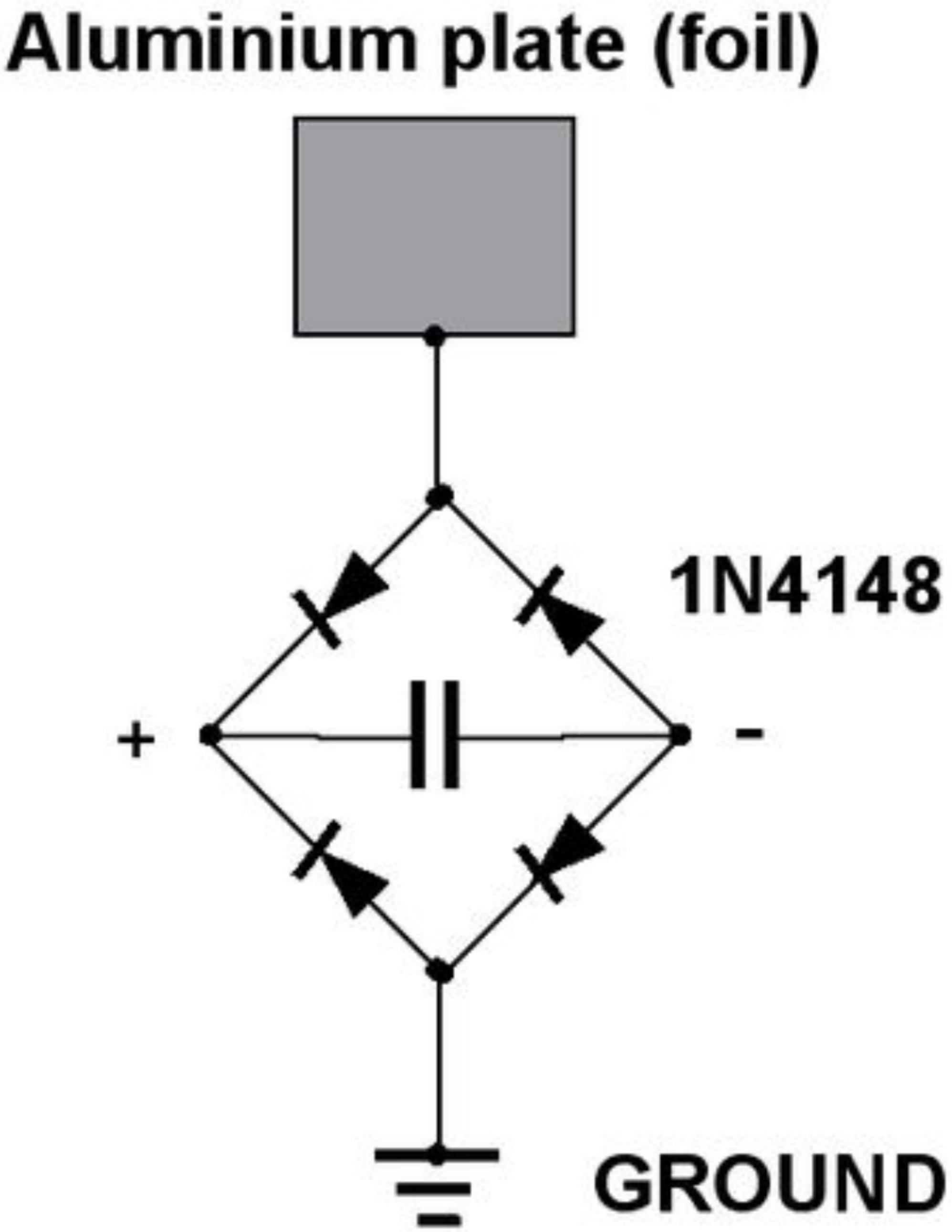
Aluminium plate (foil)
1N4148
+
-
GROUND

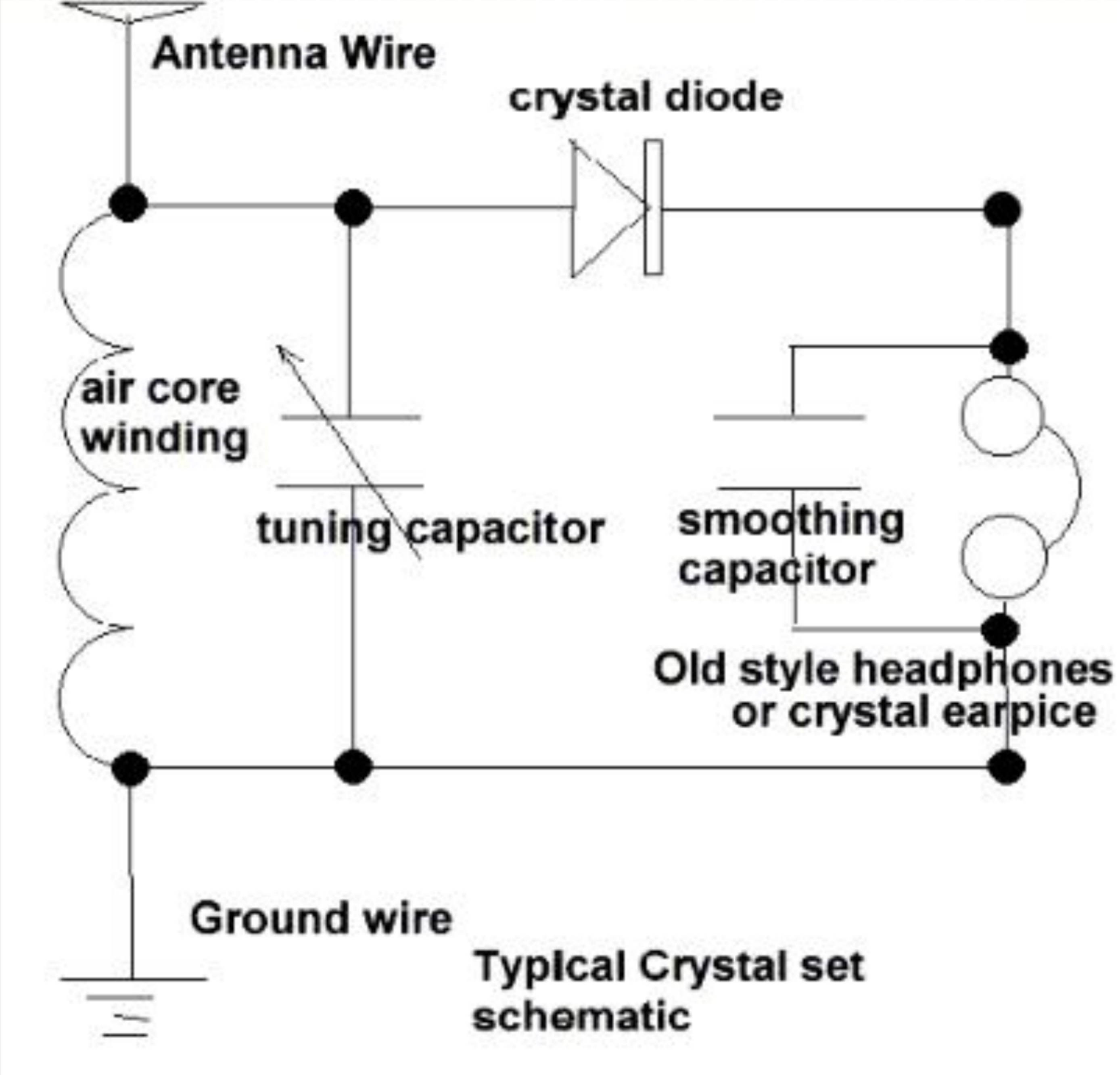
Antenna Wire
crystal diode
air core
winding
tuning capacitor
smoothing
capacitor
Old style headphones
or crystal earpice
Ground wire

Typical Crystal set
schematic

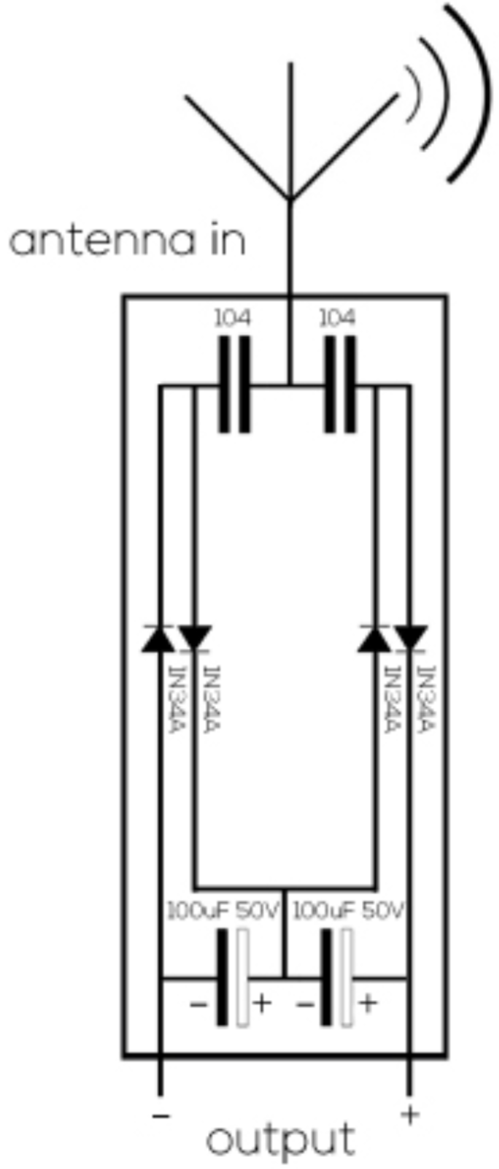
antenna in
104
104
1N34A
1N34A
1N34A
1N34A
100uF 50V
100uF 50V
–
+
–
+
–
output
+

RF to DC

Circuit Diagam (simplified)

Component list:

(2) Ceramic Capacitors (104)
(4) Germanium Diodes (1N34A)
(2) Electrolytic Capacitors (100uF 50V)

Copyright 2016. Drew Paul Designs. All rights reserved.

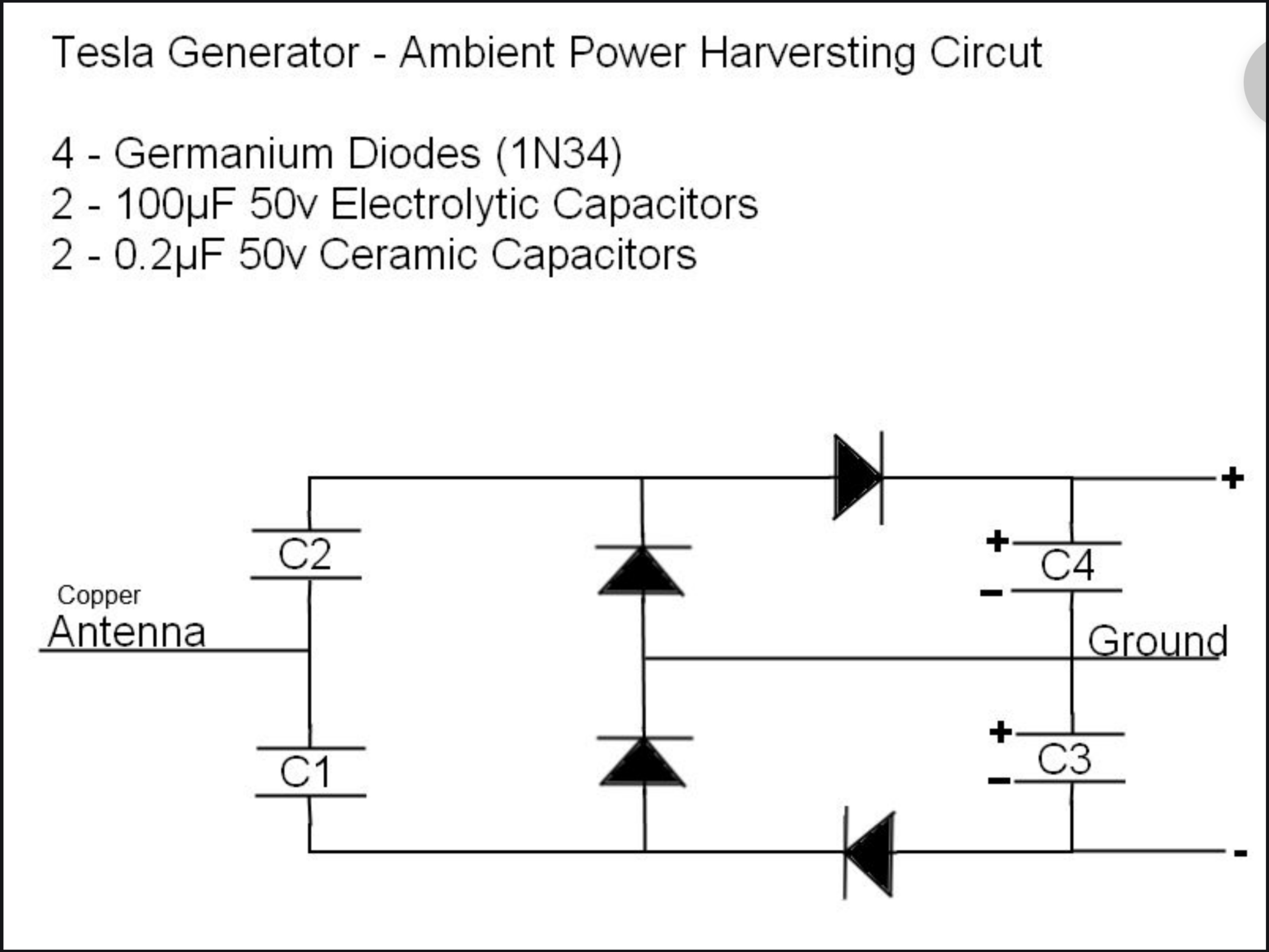

Tesla Generator - Ambient Power Harversting Circut

4 - Germanium Diodes (1N34)
2 - 100µF 50v Electrolytic Capacitors
2 - 0.2µF 50v Ceramic Capacitors

Copper
Antenna
C2
C1
+
C4
-
Ground
+
C3
-
+
-

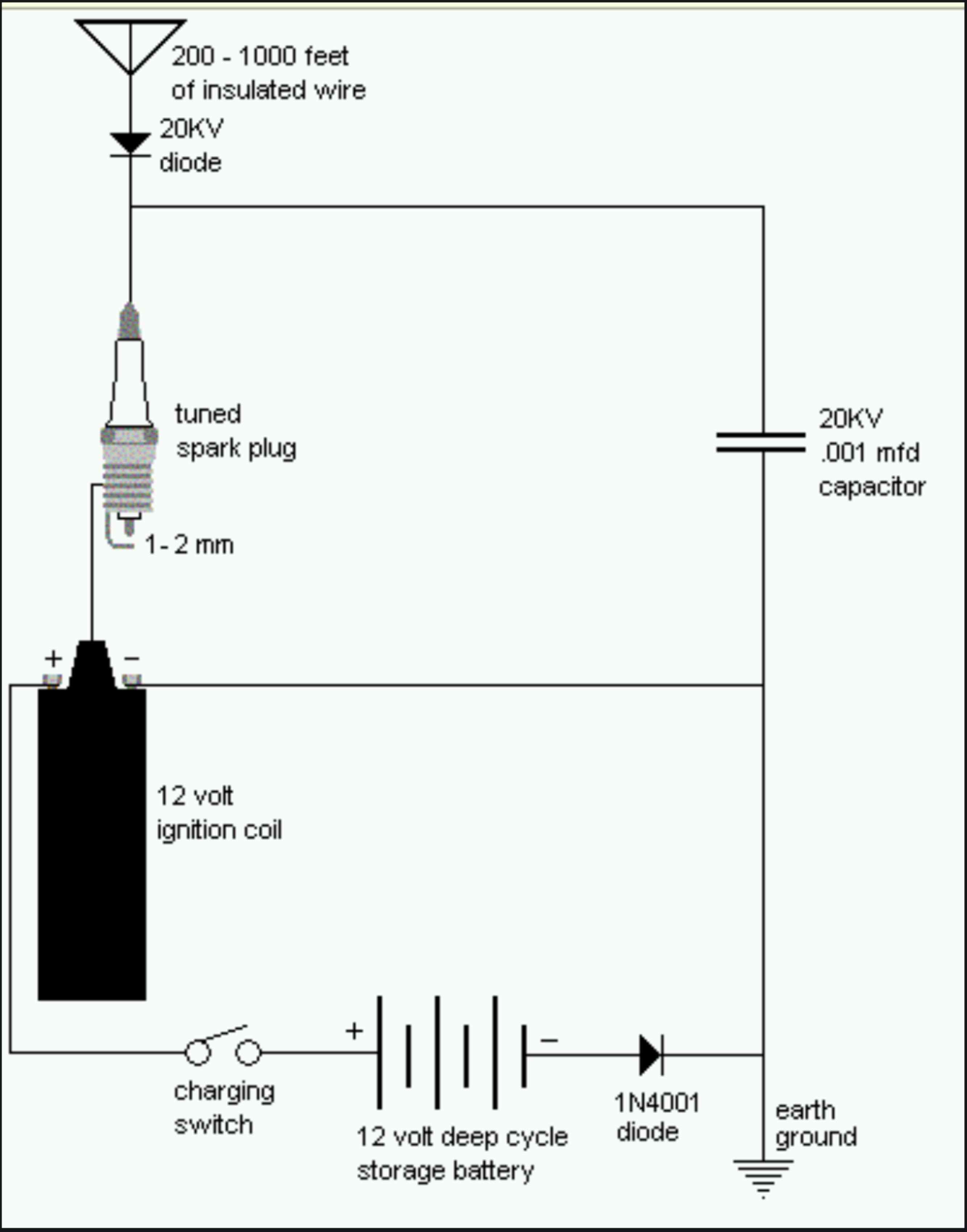
200 - 1000 feet
of insulated wire
20KV
diode
tuned
spark plug
1- 2 mm
20KV
.001 mfd
capacitor
+
−
12 volt
ignition coil
+
−
charging
switch
12 volt deep cycle
storage battery
1N4001
diode
earth
ground

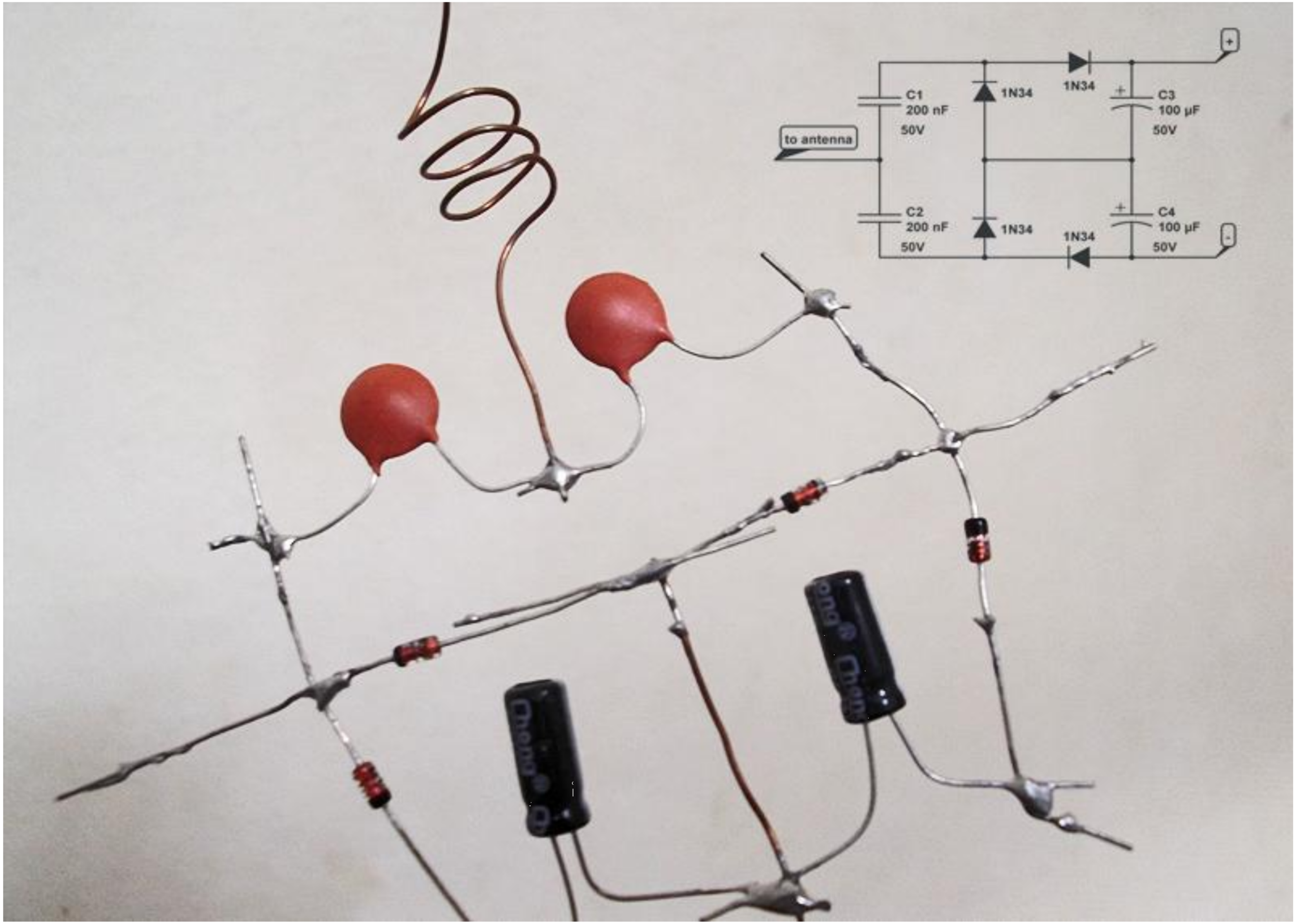
+
1N34
C1
200 nF
50V
1N34
C3
100 µF
50V
to antenna
C2
200 nF
50V
1N34
1N34
C4
100 µF
50V
-

199 Comments Share Save Hide Report 89% Upvoted

Connect the center of the two 100uF capacitors to earth and it should harvest more

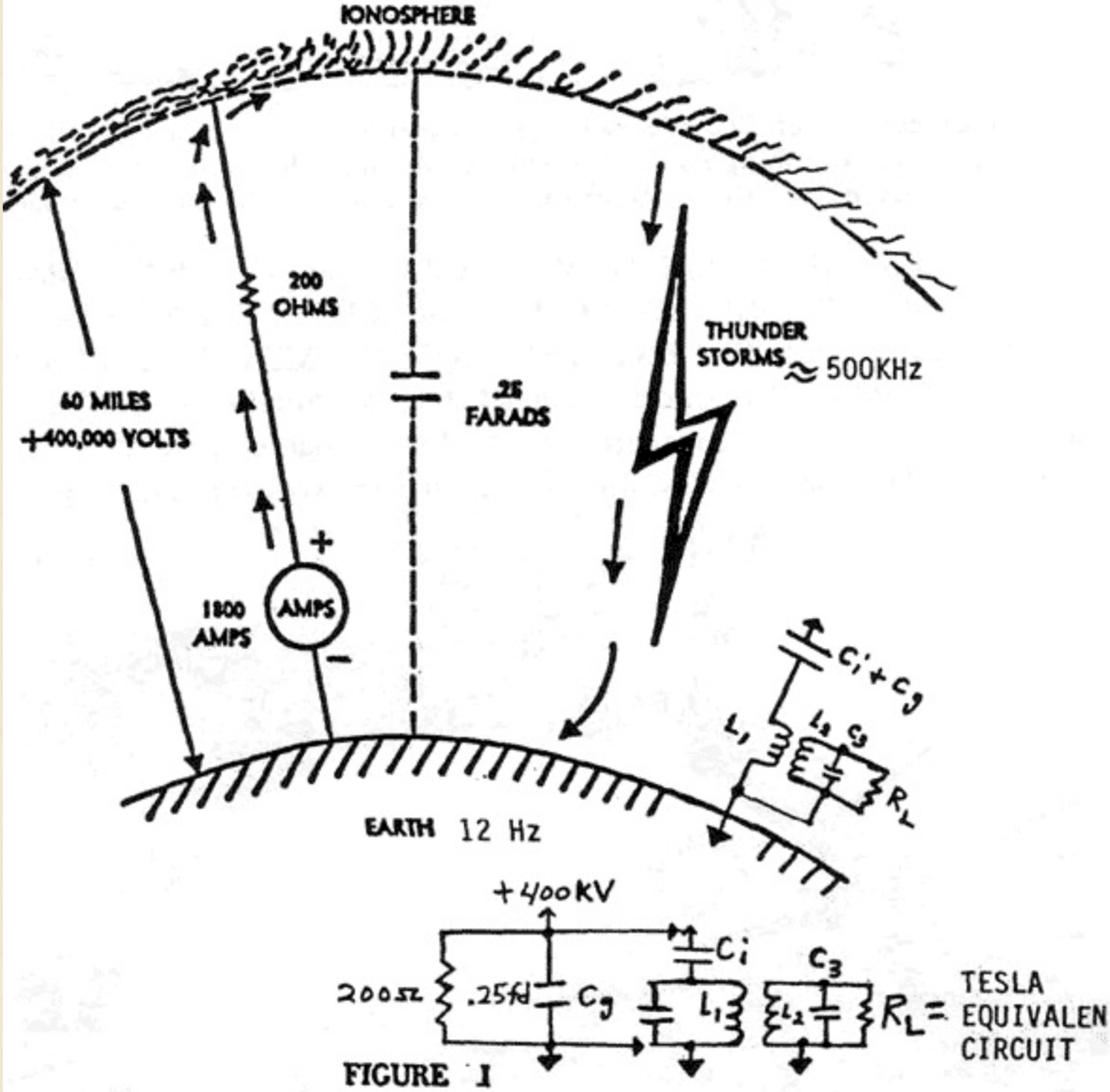
IONOSPHERE
200 OHMS
THUNDER STORMS ≈ 500KHz
60 MILES
+400,000 VOLTS
.25 FARADS
+
1800 AMPS
AMPS
−
Ci + Cg
L1
L2
C3
RL
EARTH 12 Hz
+400KV
Ci
C3
200Ω
.25fd
Cg
L1
L2
RL = TESLA EQUIVALEN CIRCUIT

FIGURE I

Equivalent circuit of earth's electrostatic voltage field.

TESLA SYSTEM

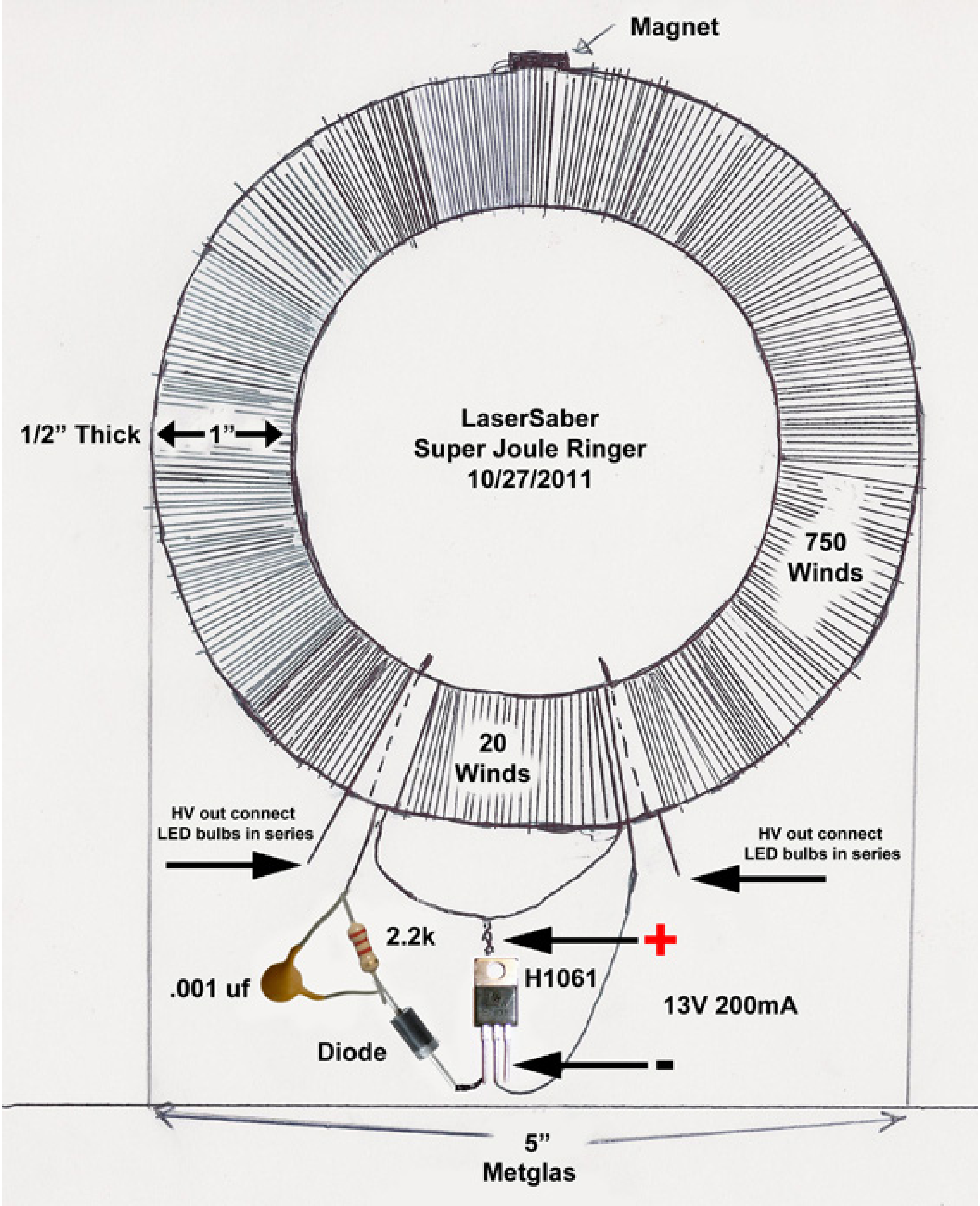
Magnet
LaserSaber
Super Joule Ringer
10/27/2011
1/2" Thick
1"
750
Winds
20
Winds
HV out connect
LED bulbs in series
HV out connect
LED bulbs in series
2.2k
+
.001 uf
H1061
13V 200mA
Diode
-
5"
Metglas

LaserSaber circuit

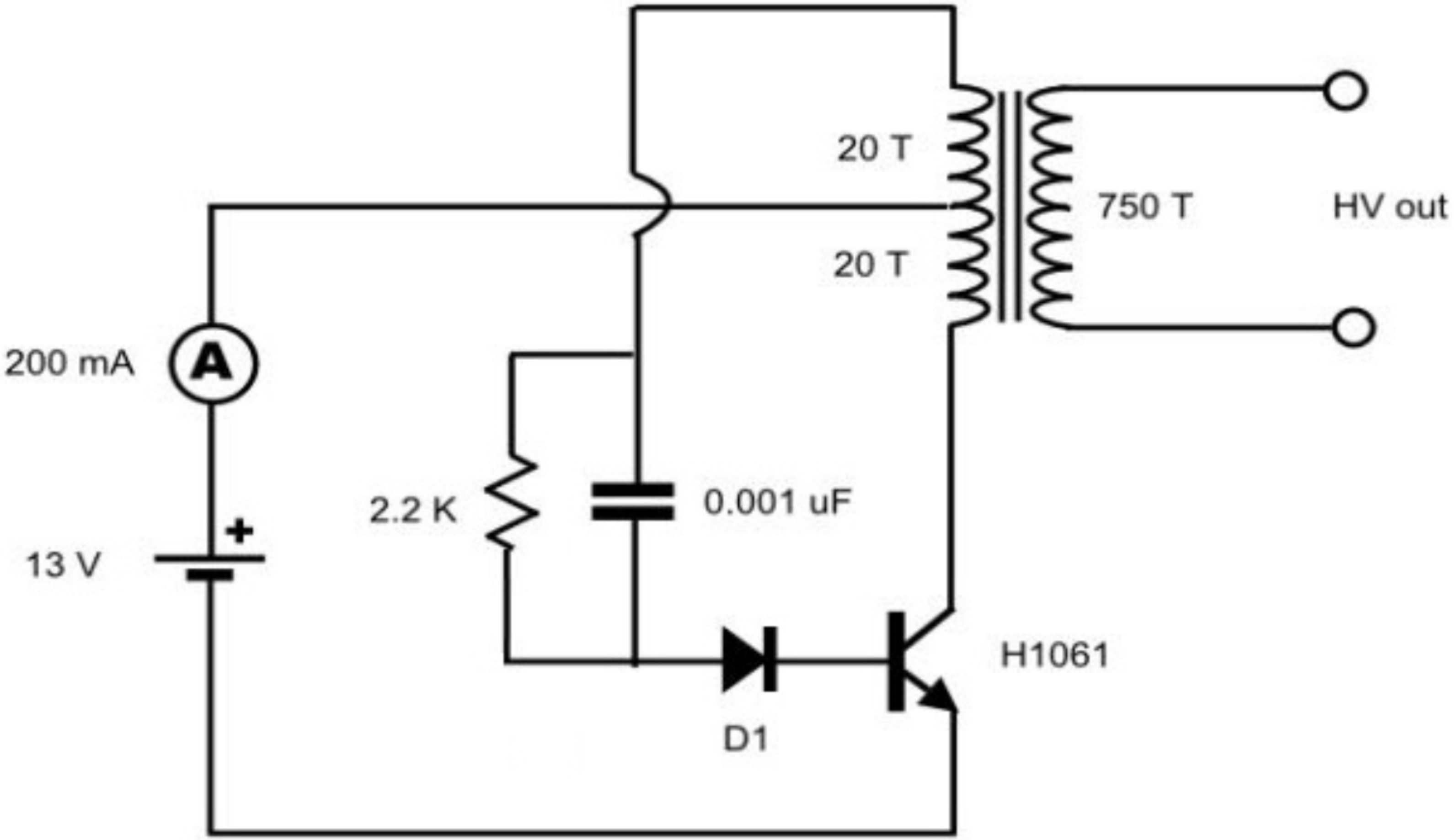
20 T
750 T
HV out
20 T
200 mA
A
2.2 K
0.001 uF
+
13 V
H1061
D1

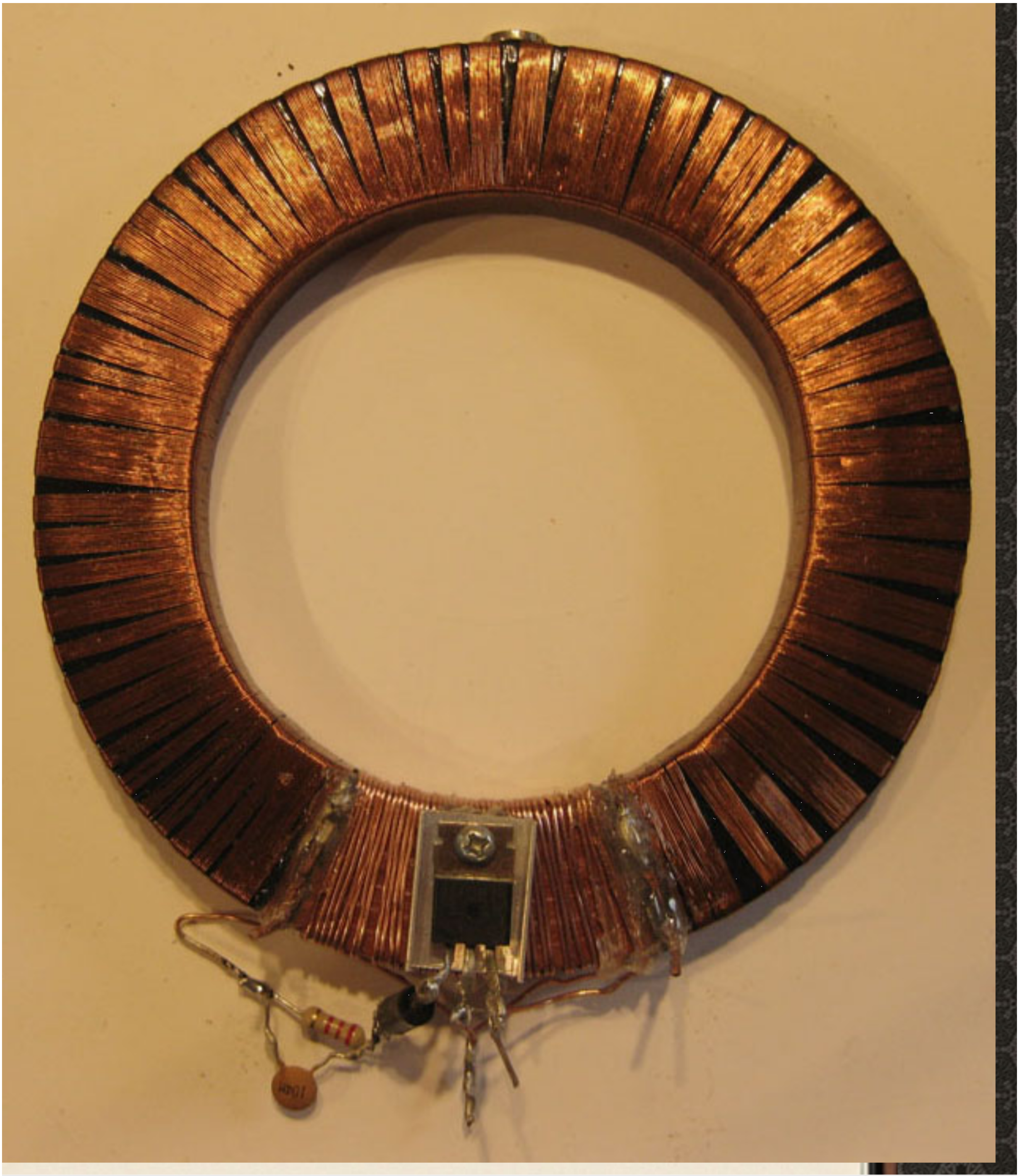

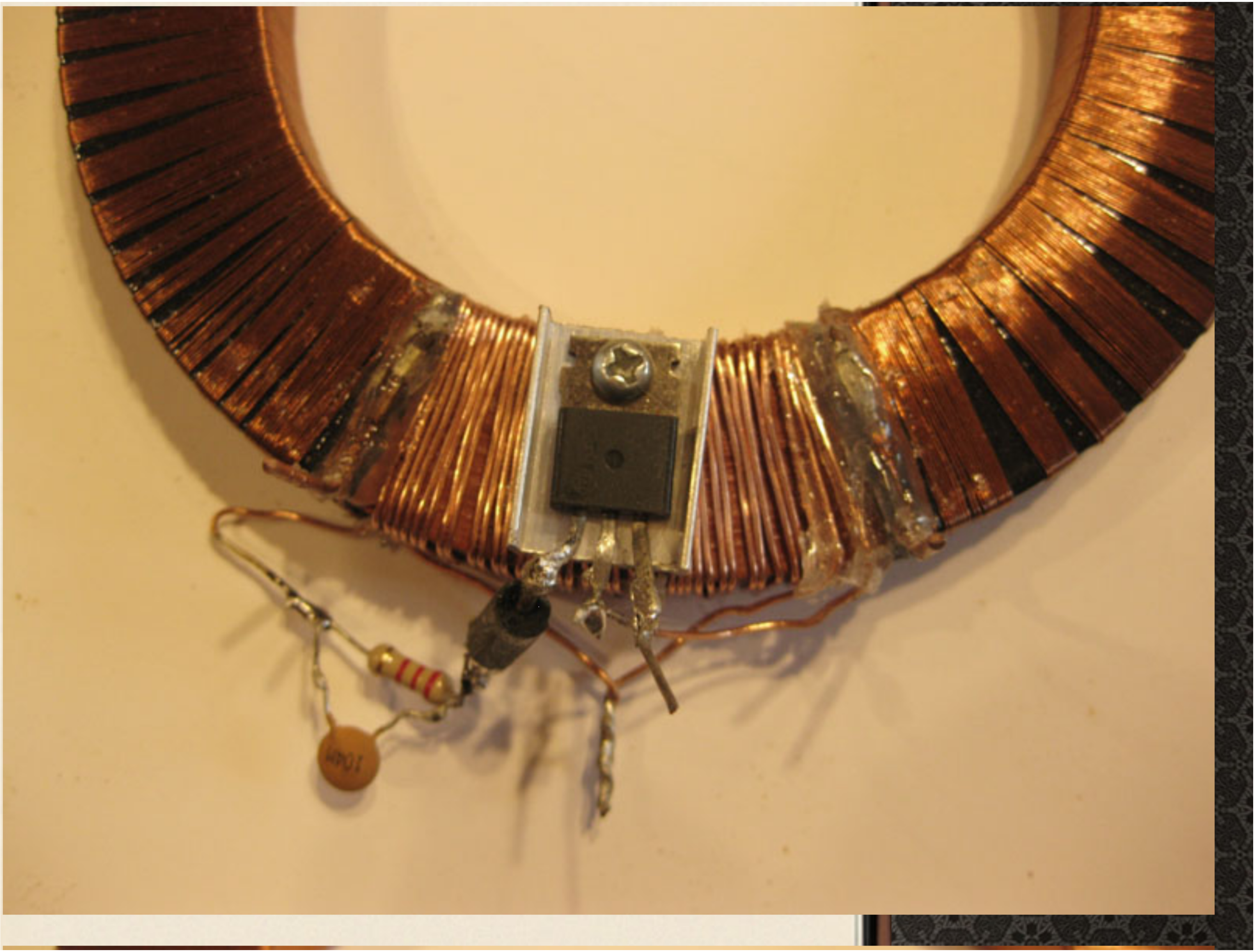

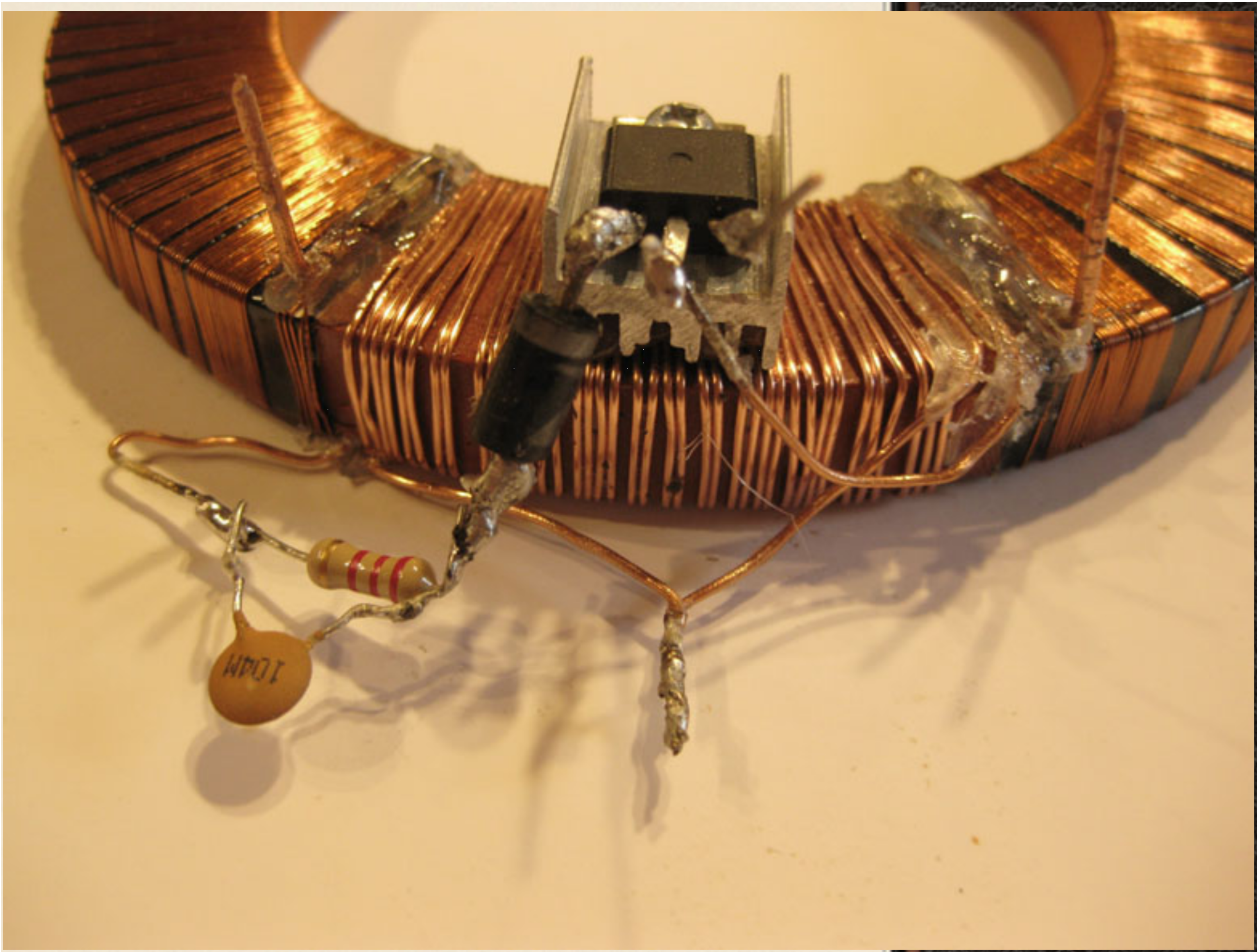
104M

- LED
- 1k resistor
- 2n904 transistor
- Thin magnet wire
- Solder
- Soldering iron
- Battery clip
- Ferrite core
- Dead battery

Prepare the Transistor

Lay the transistor on a table with the flat side facing up. Spread the pins of the transistor for easy soldering.

Bend the middle pin back and up behind the black plastic case of the transistor.

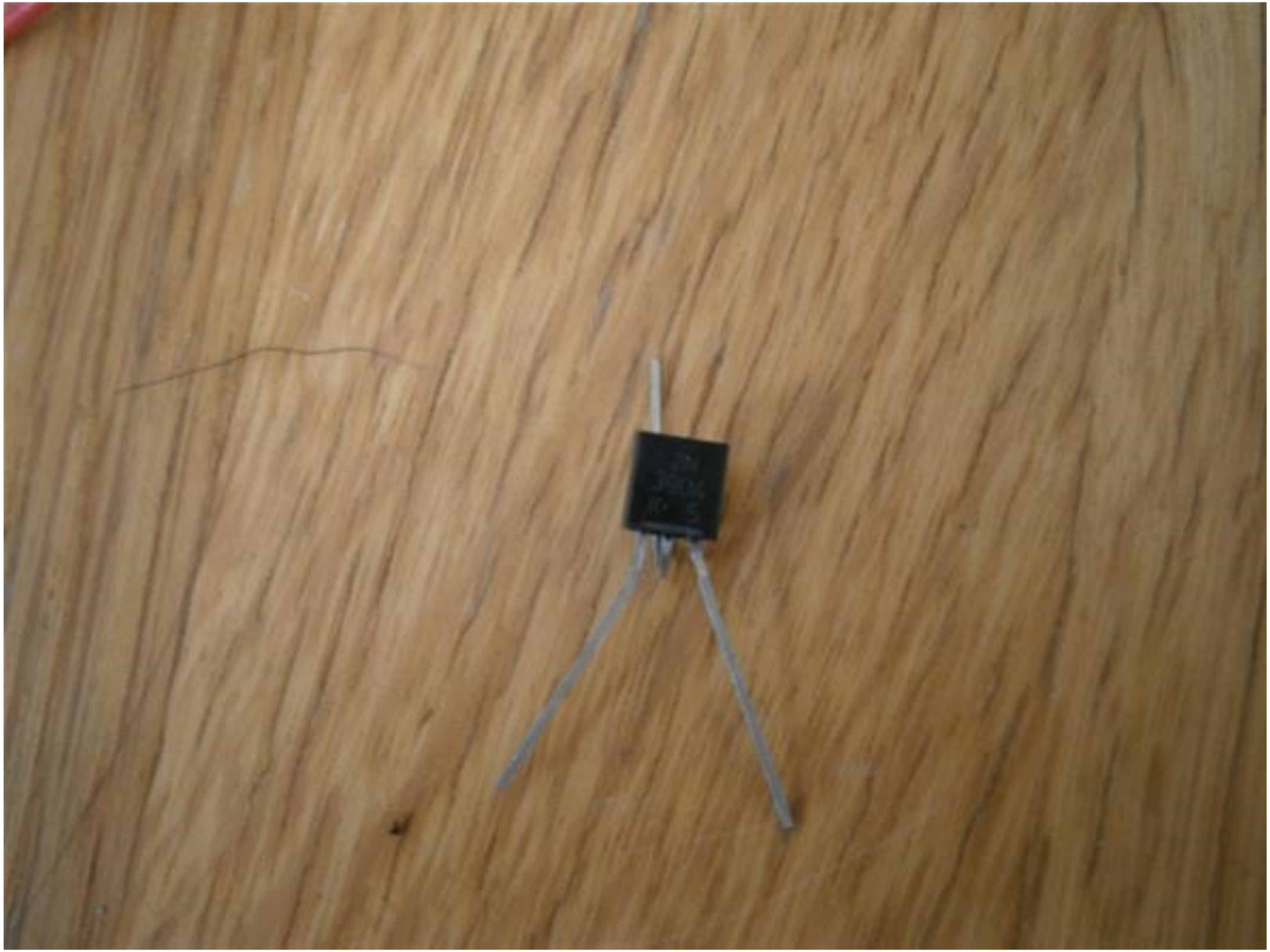

Place the LED

The LED is a polarized component. This means it will only work when it is facing the right way in our circuit. The positive lead of the LED is usually longer. Below, the positive lead is on the left, the shorter negative lead is on the right.

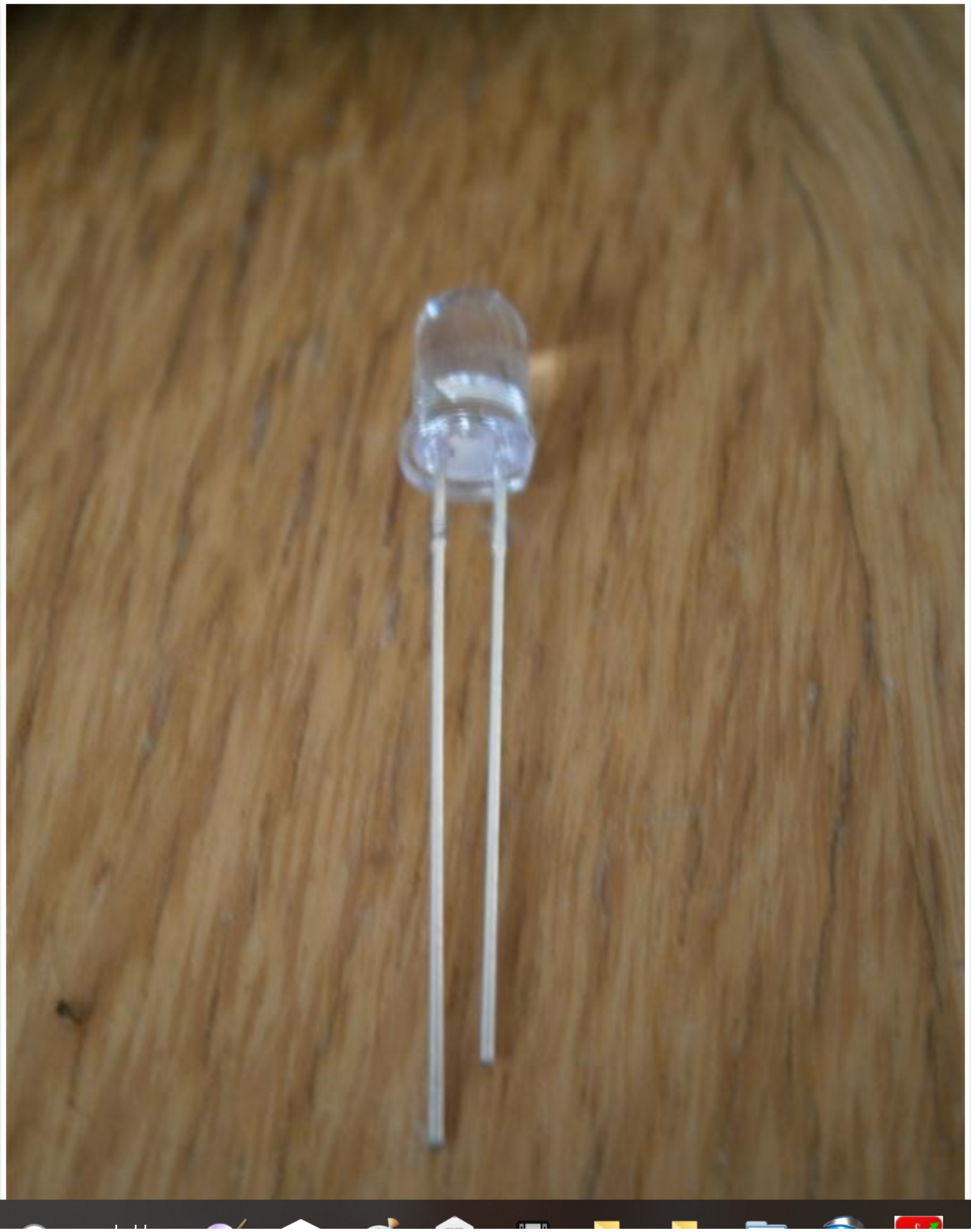

Place the LED on the transistor as shown below, with the positive side facing to the right.

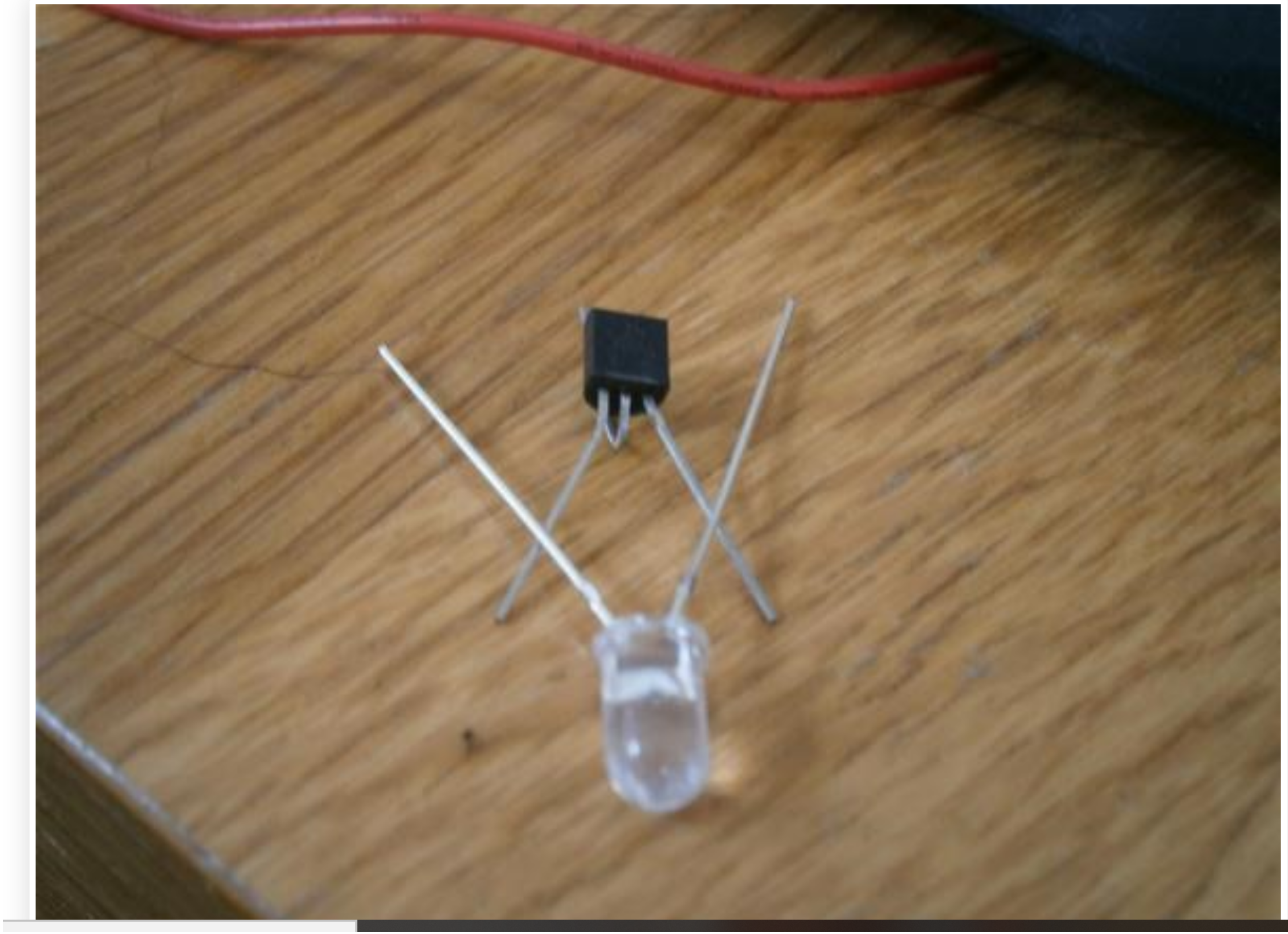

Now, solder the LED to the transistor.

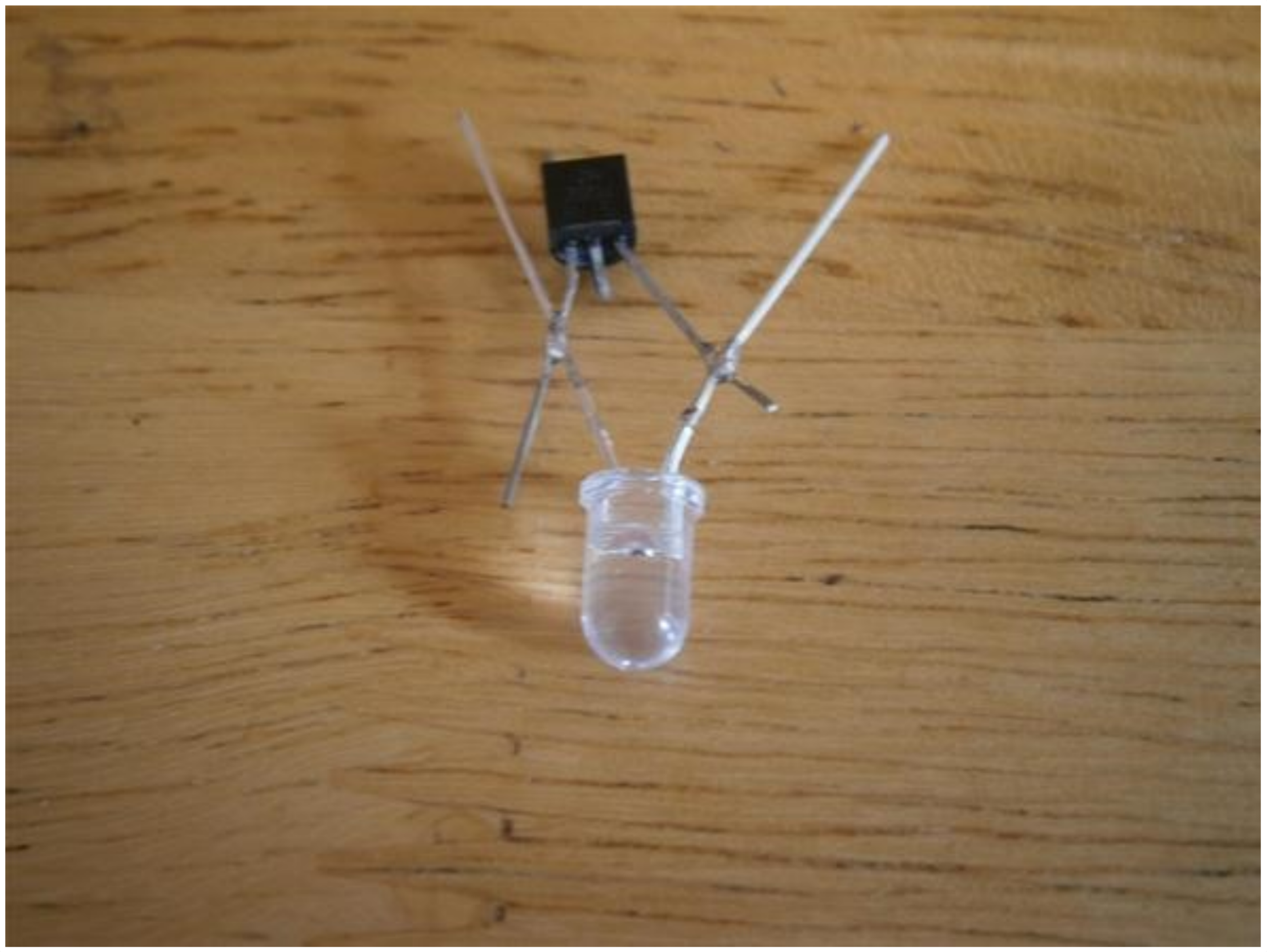

Place the Resistor

Place one end of the resistor on the middle pin of the transistor. Make sure the components stay in contact while you solder them.

Zillow®

Reimagine the way home.

Shop now

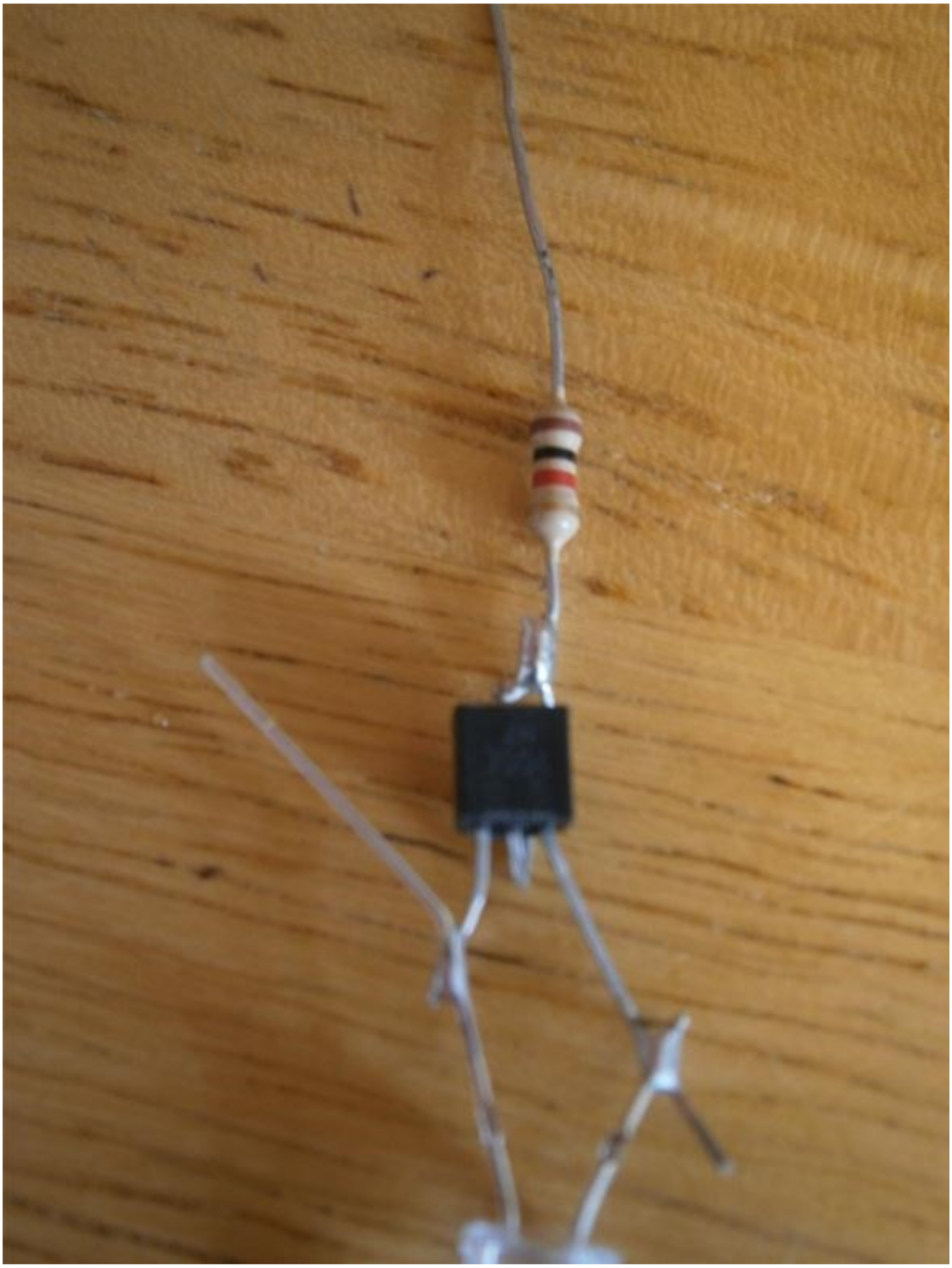

Wrap the Coil

This is where the dark magic happens. Wrap two enamel coated wires around the edge of a ferrite core. More wire will mean a stronger joule thief and a brighter LED. When you have wrapped the core, you should have two pairs of end wires. Connect one pair together as shown below on the right. Splay the other pair apart as shown below on the left.

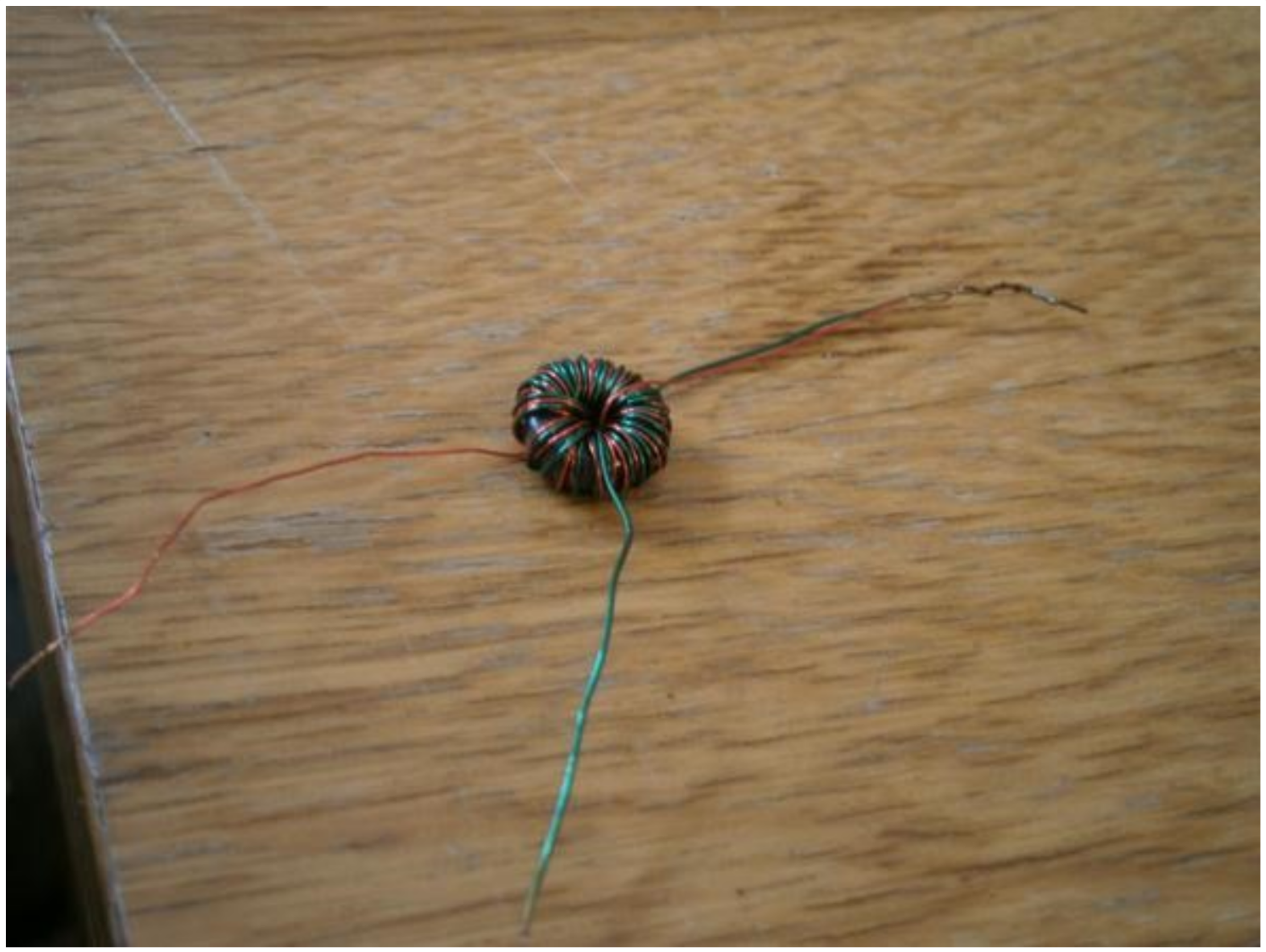

Attach Coil

Remember to burn away the enamel on the end of the thin wire with a lighter. With the wires burned and exposed, solder one free coil wire to the 1k resistor. Solder the other free coil wire to the positive side of the LED.

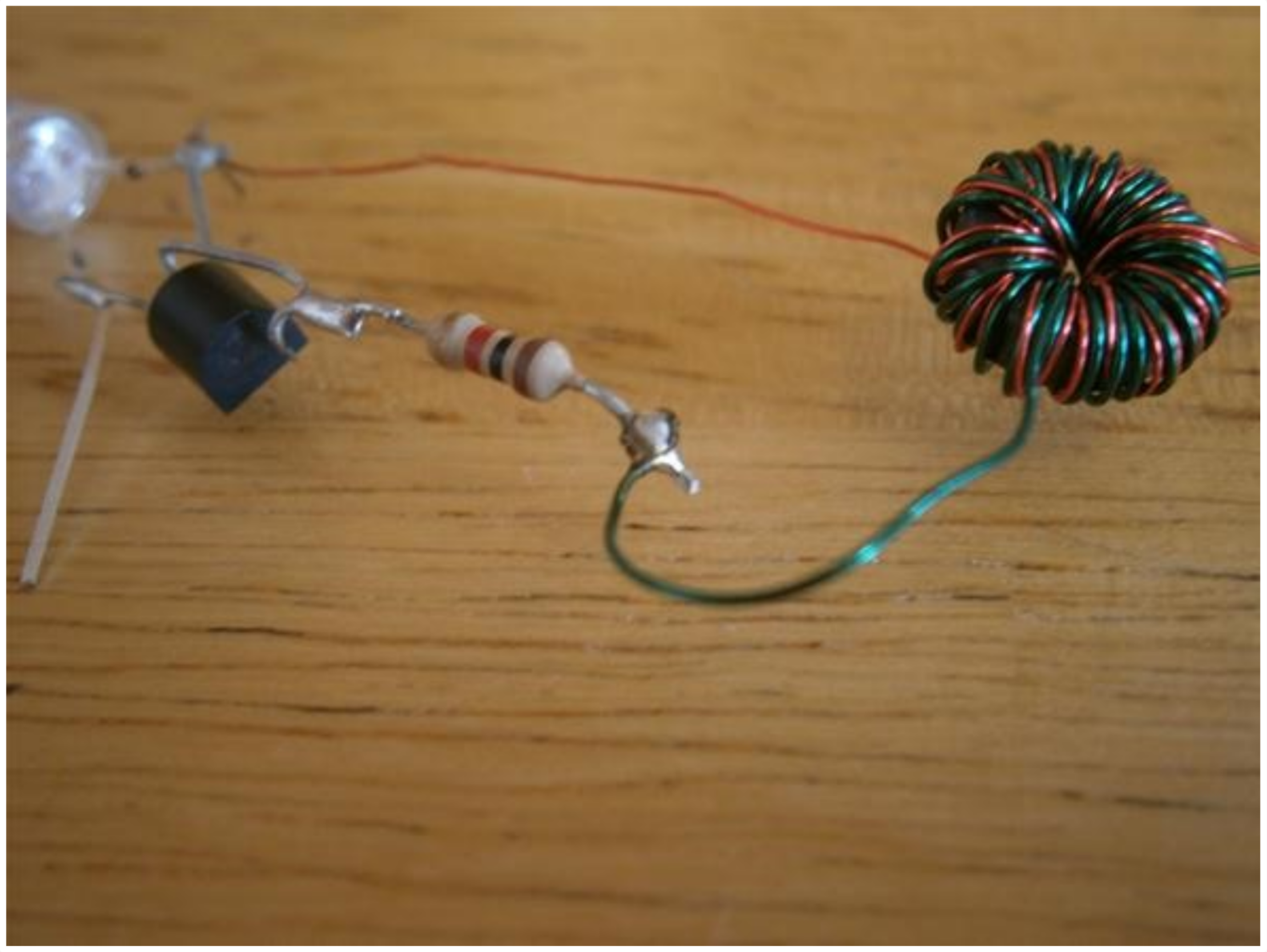

Below, we have to solder a coil wire to the intersection of the transistor and the positive leg of the LED.

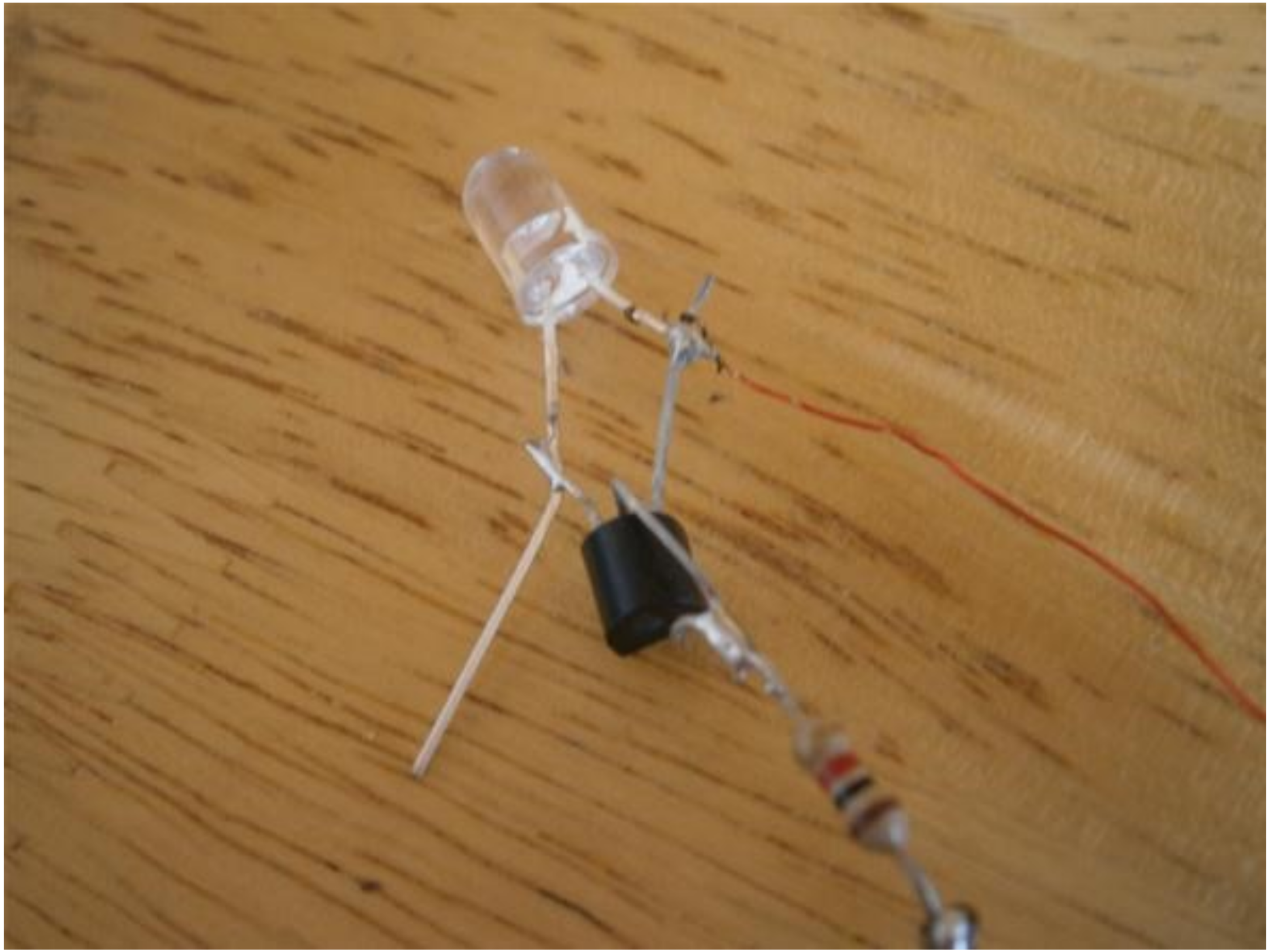

Below, notice that the end of the green enameled wire is exposed because the enamel was burned off.

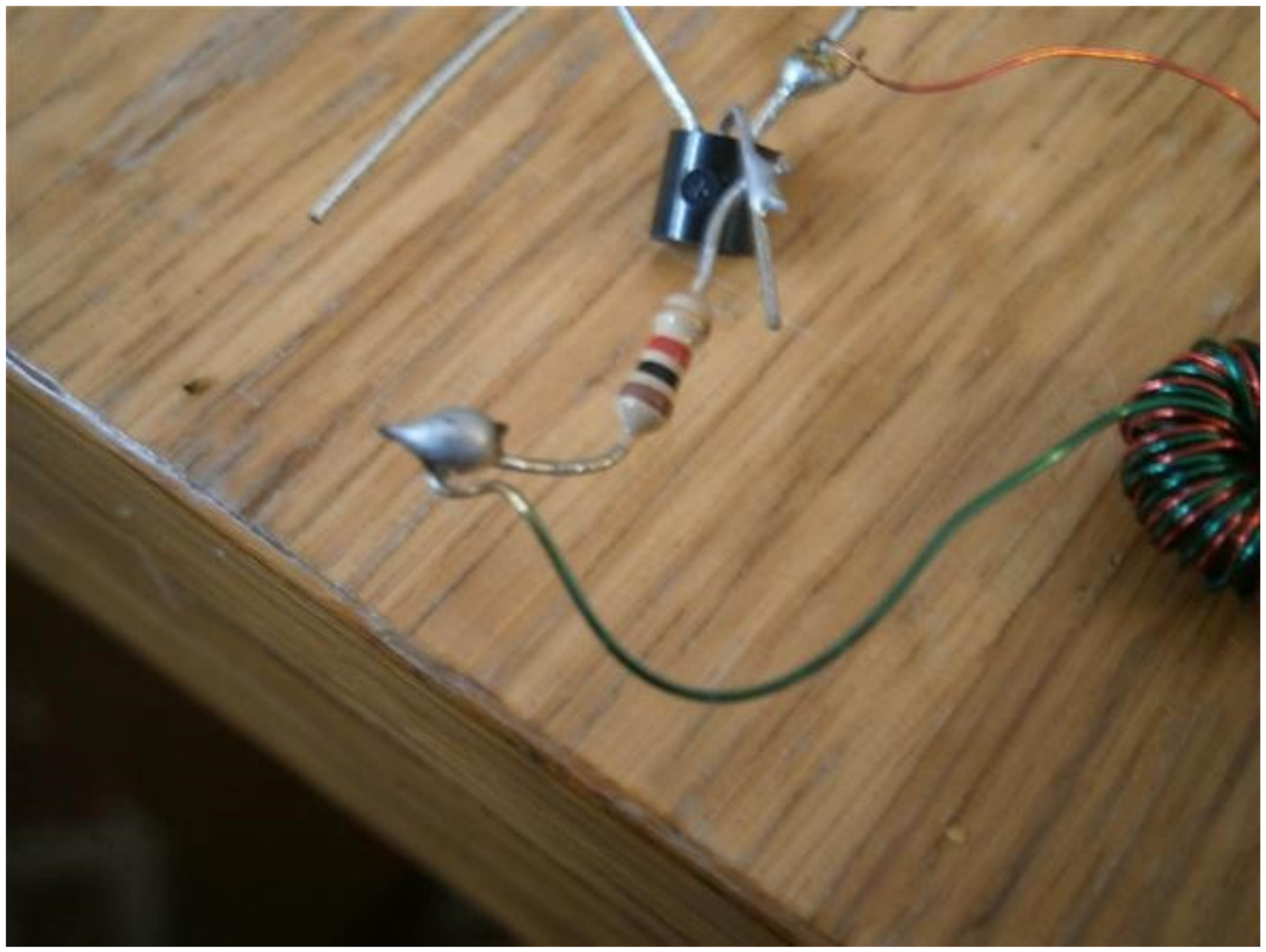

Solder Cli

Solder Clip

Now that the main joule thief is done, we can attach a battery clip for easy use. The joule thief will work without the battery clip, but only if you hold the wires in place with your fingers.

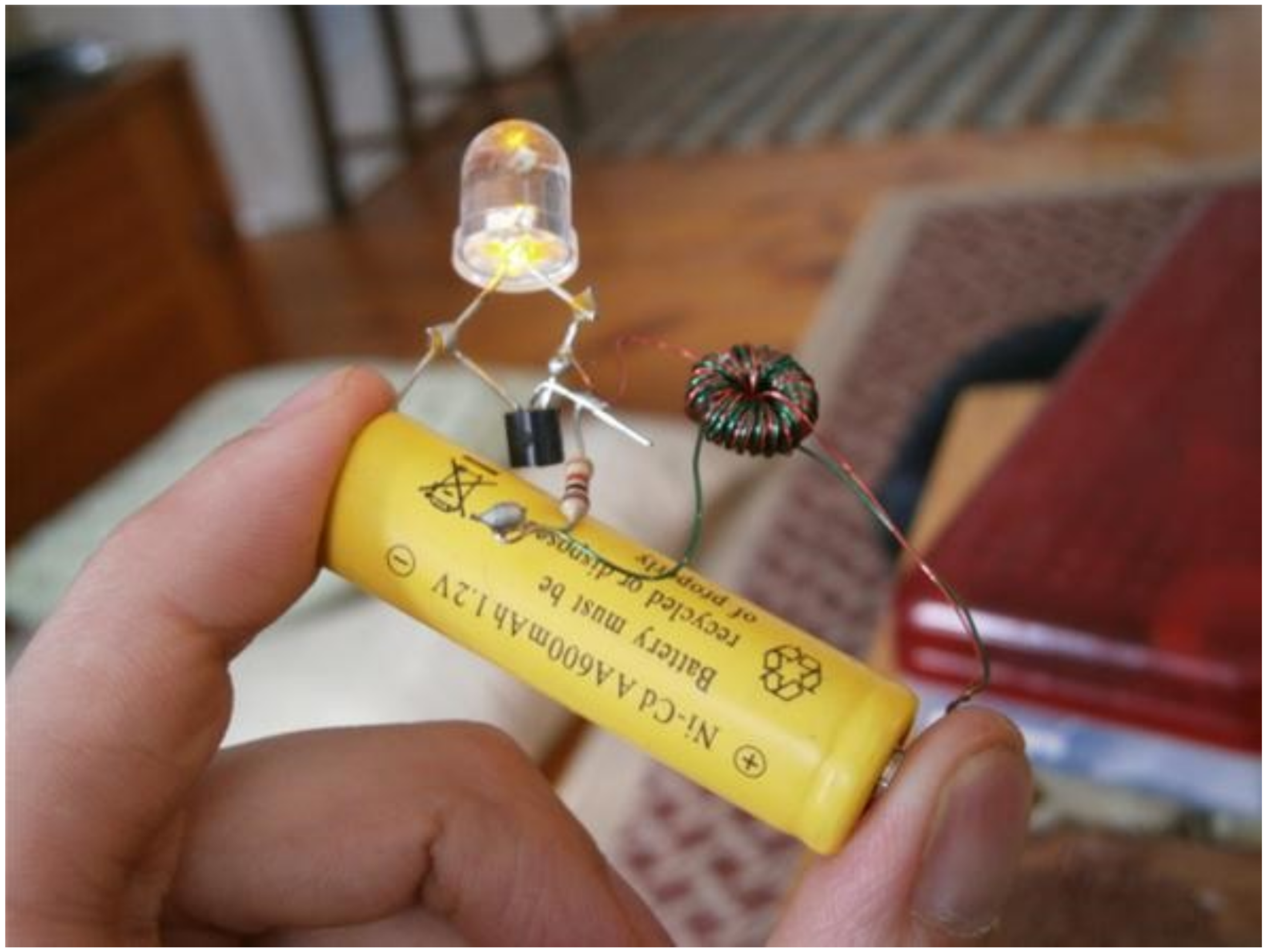
Ni-Cd AA600mAh 1.2V
Battery must be
recycled or disposed
of properly

The intertwined end of the coil connects directly to the positive battery wire. Be sure that both wires in the pair are soldered to the red wire.

The black battery wire connects to the negative end of the LED.

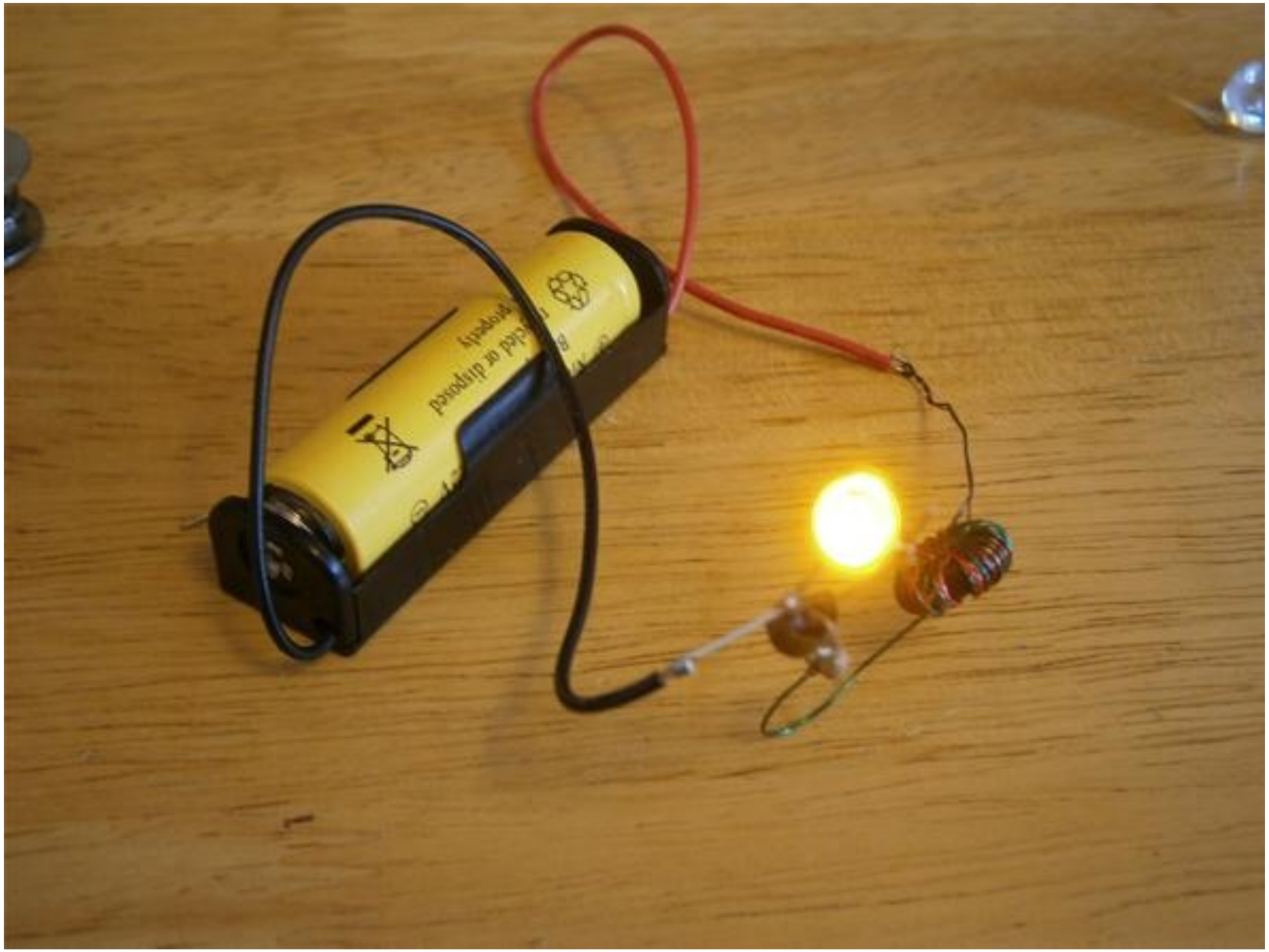

For those who need the schematics, this is what we just built:

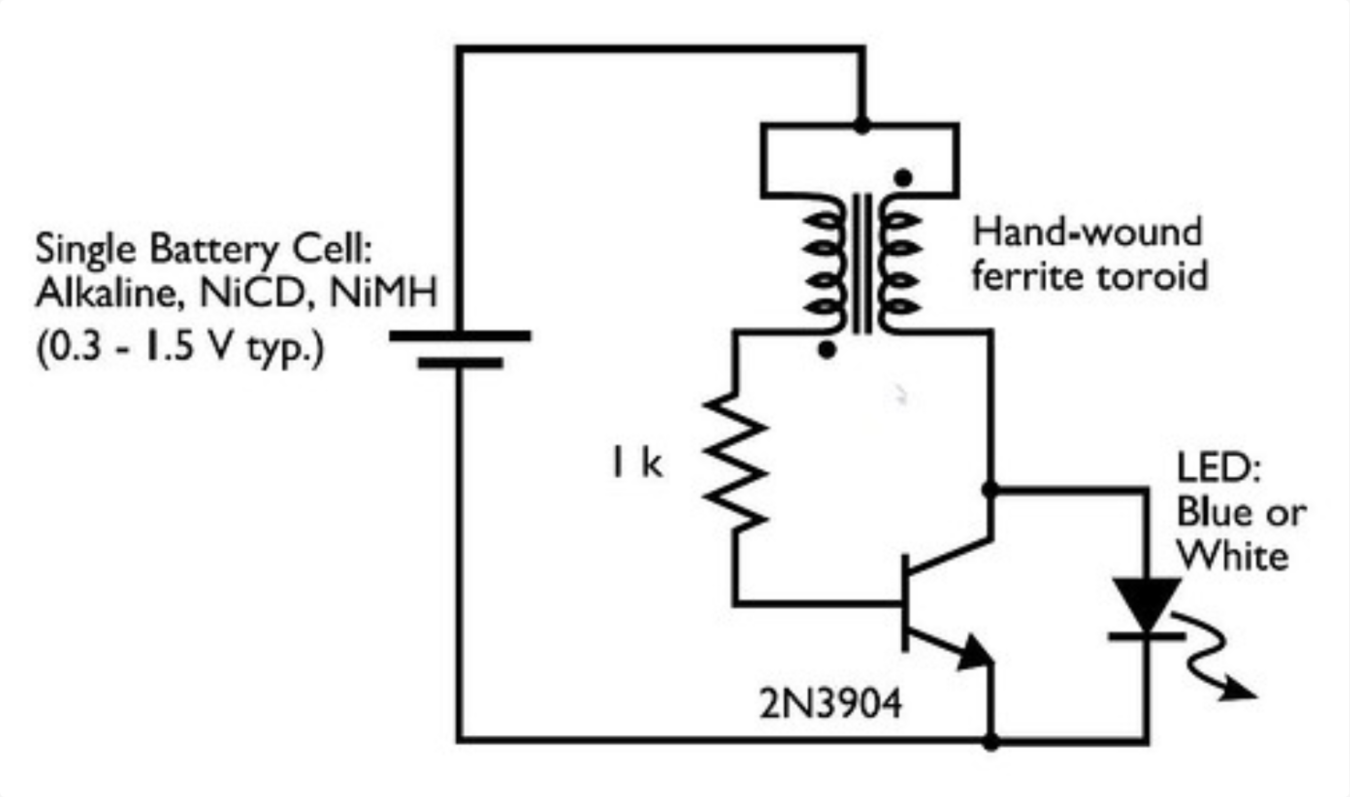

Single Battery Cell:
Alkaline, NiCD, NiMH
(0.3 - 1.5 V typ.)
Hand-wound
ferrite toroid
1 k
LED:
Blue or
White
2N3904

When the current flows through the coil and around the magnet, it produce an electromagnetic field (EMF). When the coil is not powered, the field collapses and produces an EMF kick in the coil that is of a higher voltage than the source was. The last of the electricity gets stored up in the field and released in large bursts to flicker the LED. The magnetic field oscillates so quickly that the blinking LED appears solid to the naked eye.

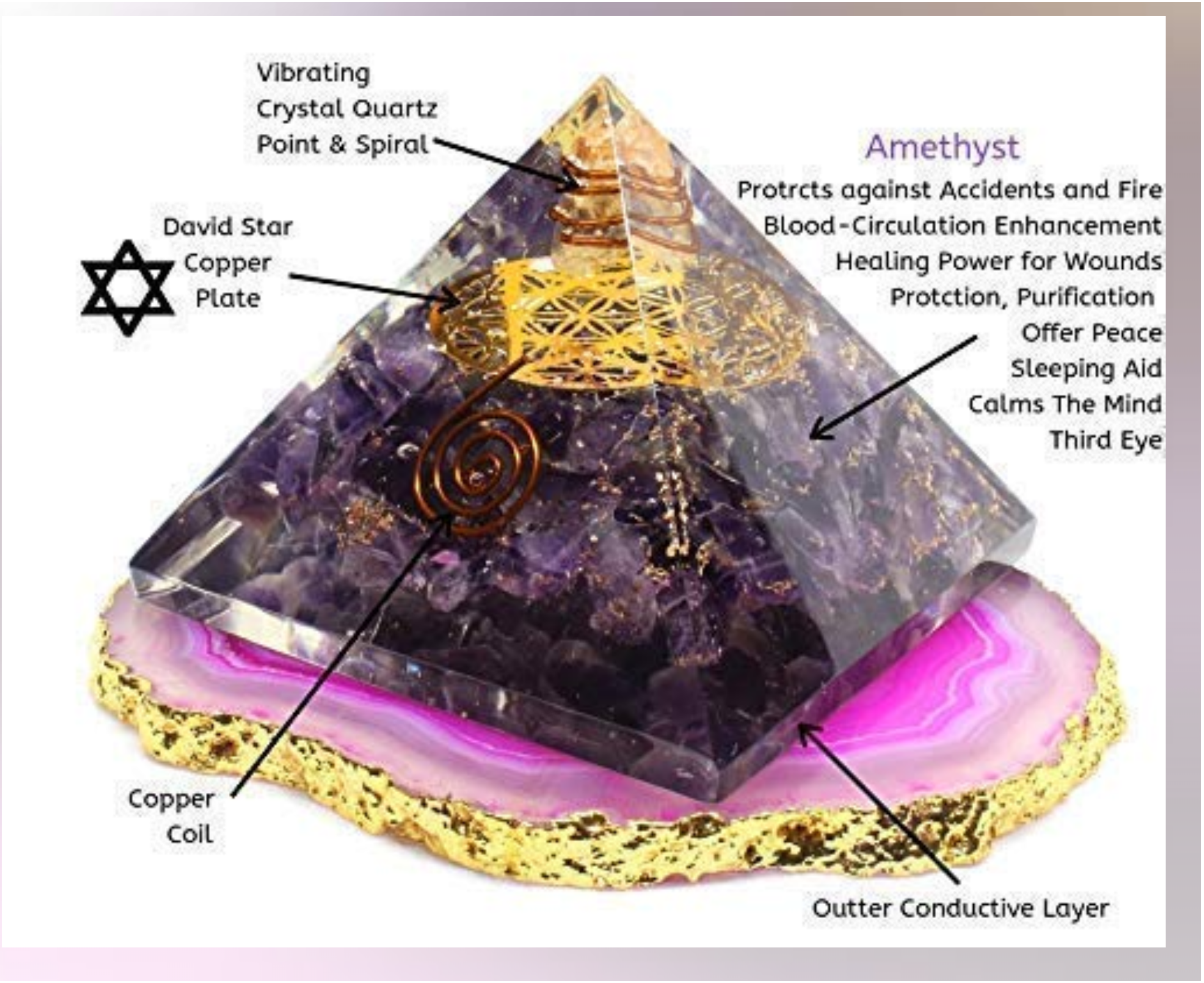
Vibrating
Crystal Quartz
Point & Spiral
David Star
Copper
Plate
Amethyst
Protrcts against Accidents and Fire
Blood-Circulation Enhancement
Healing Power for Wounds
Protction, Purification
Offer Peace
Sleeping Aid
Calms The Mind
Third Eye
Copper
Coil
Outter Conductive Layer

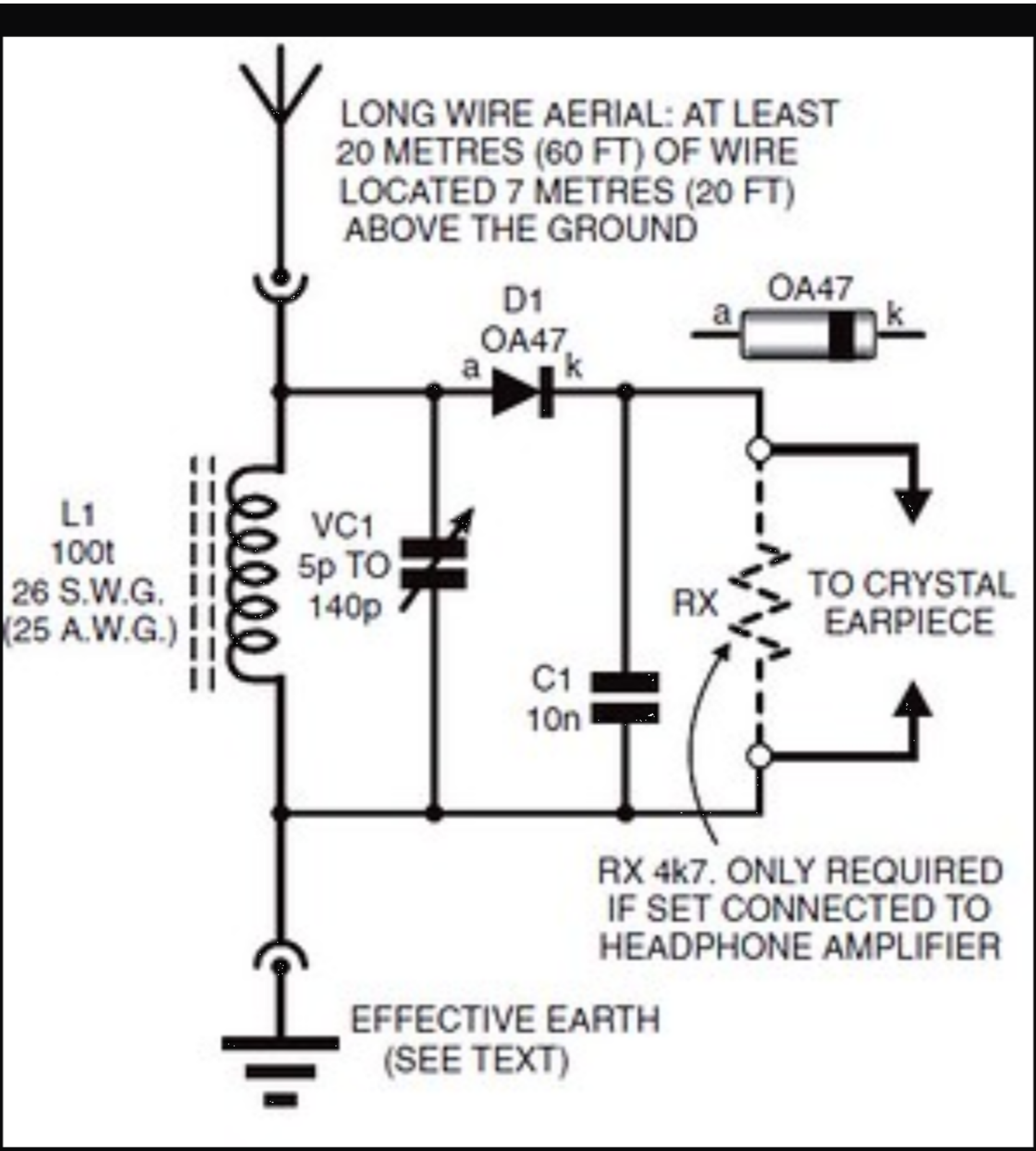
LONG WIRE AERIAL: AT LEAST
20 METRES (60 FT) OF WIRE
LOCATED 7 METRES (20 FT)
ABOVE THE GROUND
D1
OA47
a
k
OA47
a
k
L1
100t
26 S.W.G.
(25 A.W.G.)
VC1
5p TO
140p
RX
TO CRYSTAL
EARPIECE
C1
10n
RX 4k7. ONLY REQUIRED
IF SET CONNECTED TO
HEADPHONE AMPLIFIER
EFFECTIVE EARTH
(SEE TEXT)

** You can use 1N34 Germanium diode in place of OA47*

A long (at least 20 metres), high (17 metres or more) aerial and a good earth (a buried biscuit tin or a metre of copper pipe driven into damp ground) are required in order to ensure audible headphone reception. The earphones originally used with these receivers had an impedance of around 4000 ohms and were very sensitive (and heavy and uncomfortable). They are no longer available, but a crystal earpiece, which relies on the piezoelectric effect, will give acceptable results.

Low impedance "Walkman" type earphones are NOT suitable.

Component details:

Resistor- RX-4.7k, 0.25 W- only required if set is connect to audio amplifier

Capacitors- C1- 10nF disc ceramic VC1- 5p to 140p, polythene dielectric variable capacitor.

Semiconductors: D1: OA47 or 1N34- Germanium Diode

Miscellaneous: L1- Ferrite Rod, 100mm(4 inch)x9mm/10mm dia., with coil.

Crystal earpiece and jack socket to suit; plastic control knob; plastic insulated flexible cable for aerial wire, downlead and earth connection, 30 meters minimum; buried biscuit tin or 1 meter of copper pipe for earth system; 50gm reel of 26SWG enamelled copper wire, for tuning coil; card and glue for coil former; multistrand connecting wire; crocodile clips or terminals for aerial and earth lead connection; solder, etc.

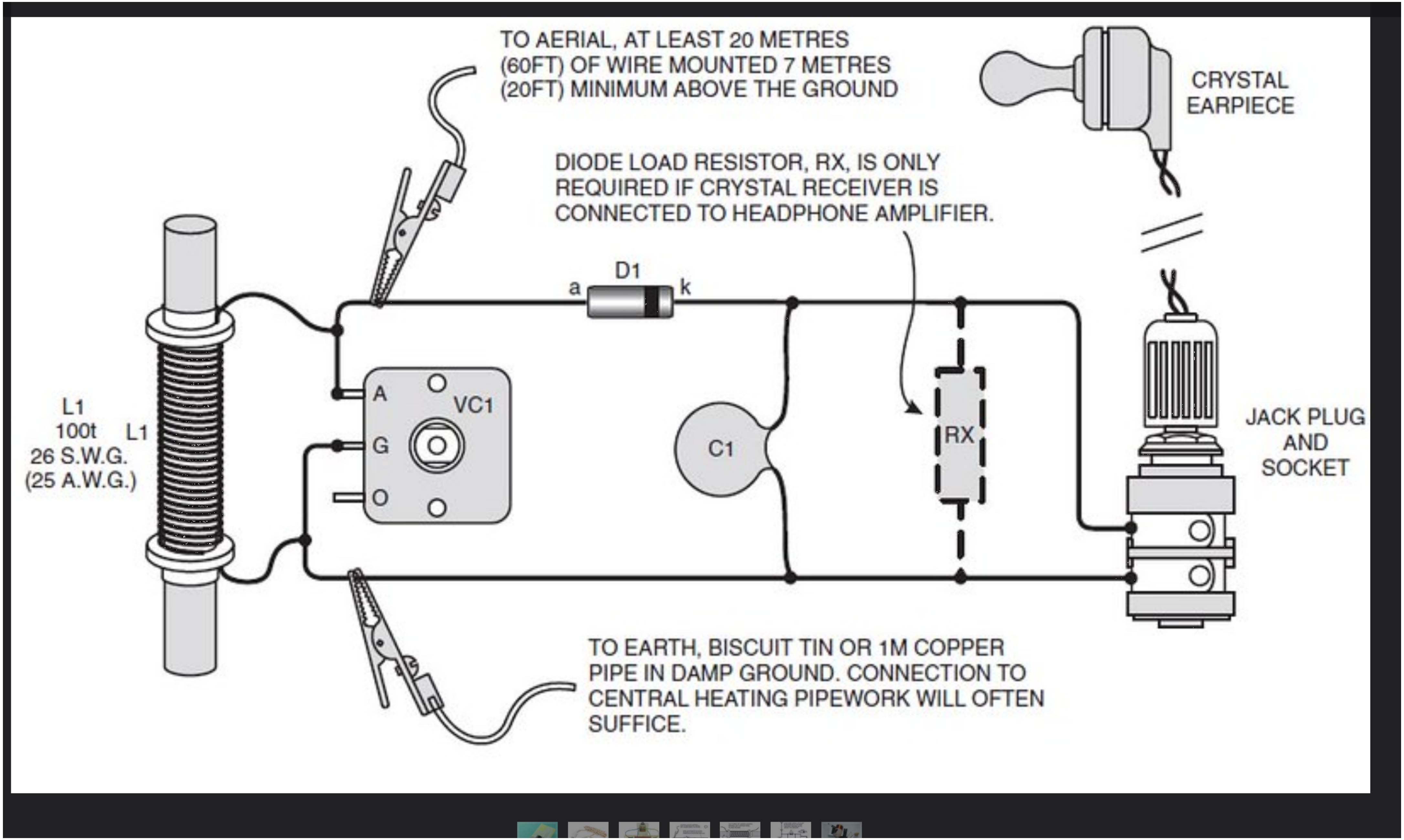
D1
a
k
VC1
A
G
O
C1
RX
L1
100t
L1
26 S.W.G.
(25 A.W.G.)
CRYSTAL EARPIECE
JACK PLUG AND SOCKET

TO AERIAL, AT LEAST 20 METRES (60FT) OF WIRE MOUNTED 7 METRES (20FT) MINIMUM ABOVE THE GROUND

DIODE LOAD RESISTOR, RX, IS ONLY REQUIRED IF CRYSTAL RECEIVER IS CONNECTED TO HEADPHONE AMPLIFIER.

TO EARTH, BISCUIT TIN OR 1M COPPER PIPE IN DAMP GROUND. CONNECTION TO CENTRAL HEATING PIPEWORK WILL OFTEN SUFFICE.

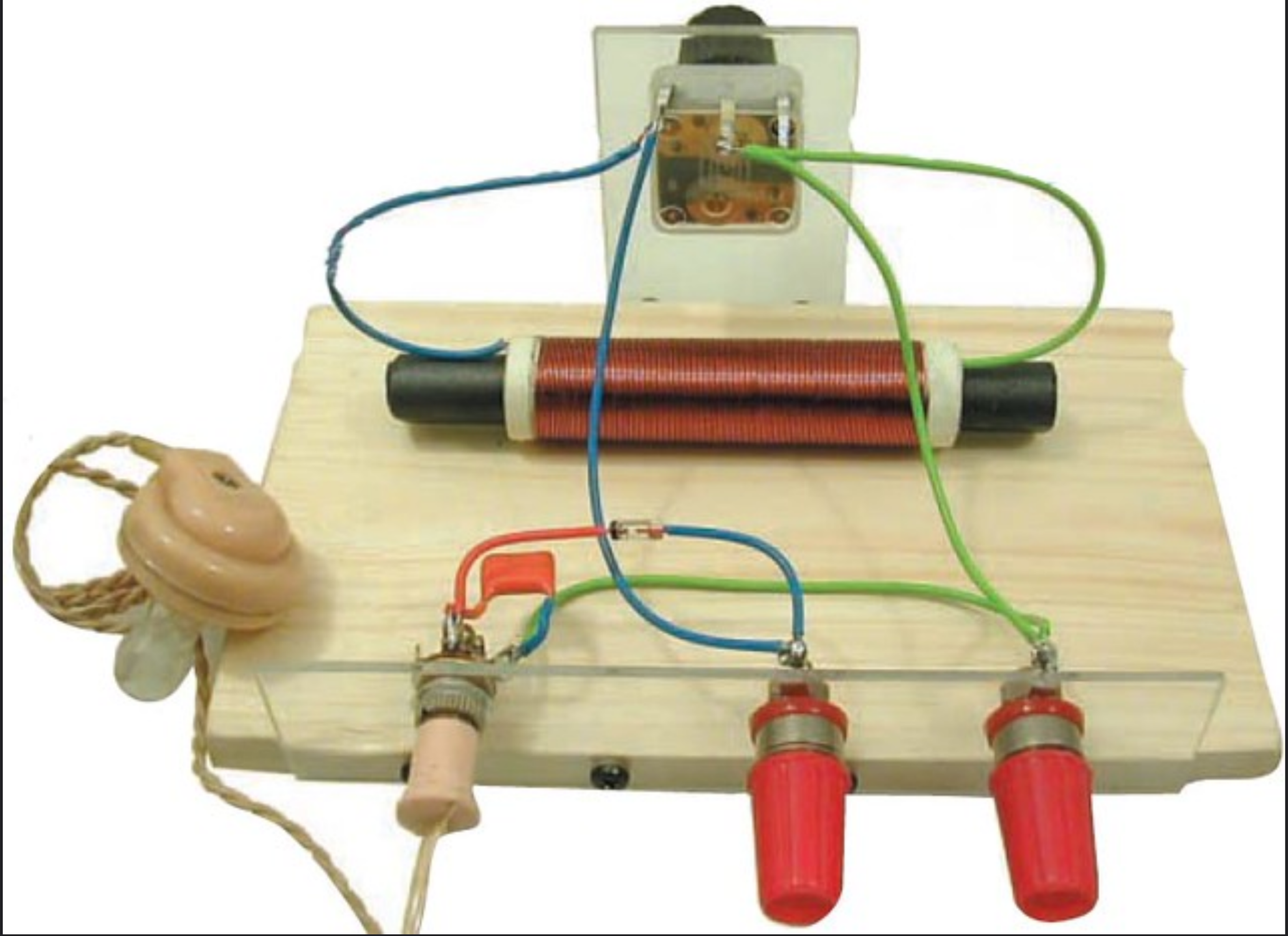

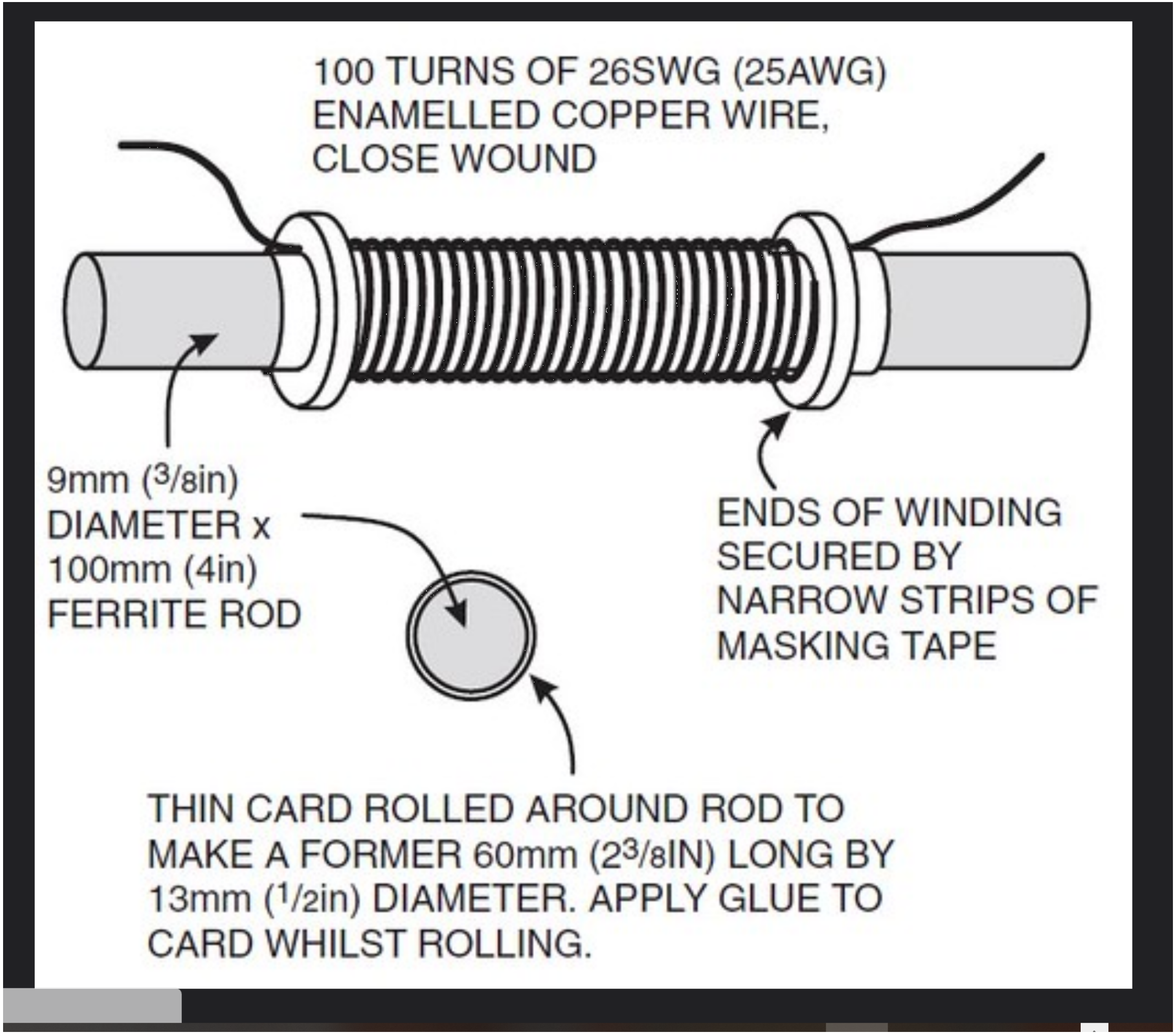
100 TURNS OF 26SWG (25AWG) ENAMELLED COPPER WIRE, CLOSE WOUND
9mm (3/8in) DIAMETER x 100mm (4in) FERRITE ROD
ENDS OF WINDING SECURED BY NARROW STRIPS OF MASKING TAPE
THIN CARD ROLLED AROUND ROD TO MAKE A FORMER 60mm (2 3/8IN) LONG BY 13mm (1/2in) DIAMETER. APPLY GLUE TO CARD WHILST ROLLING.

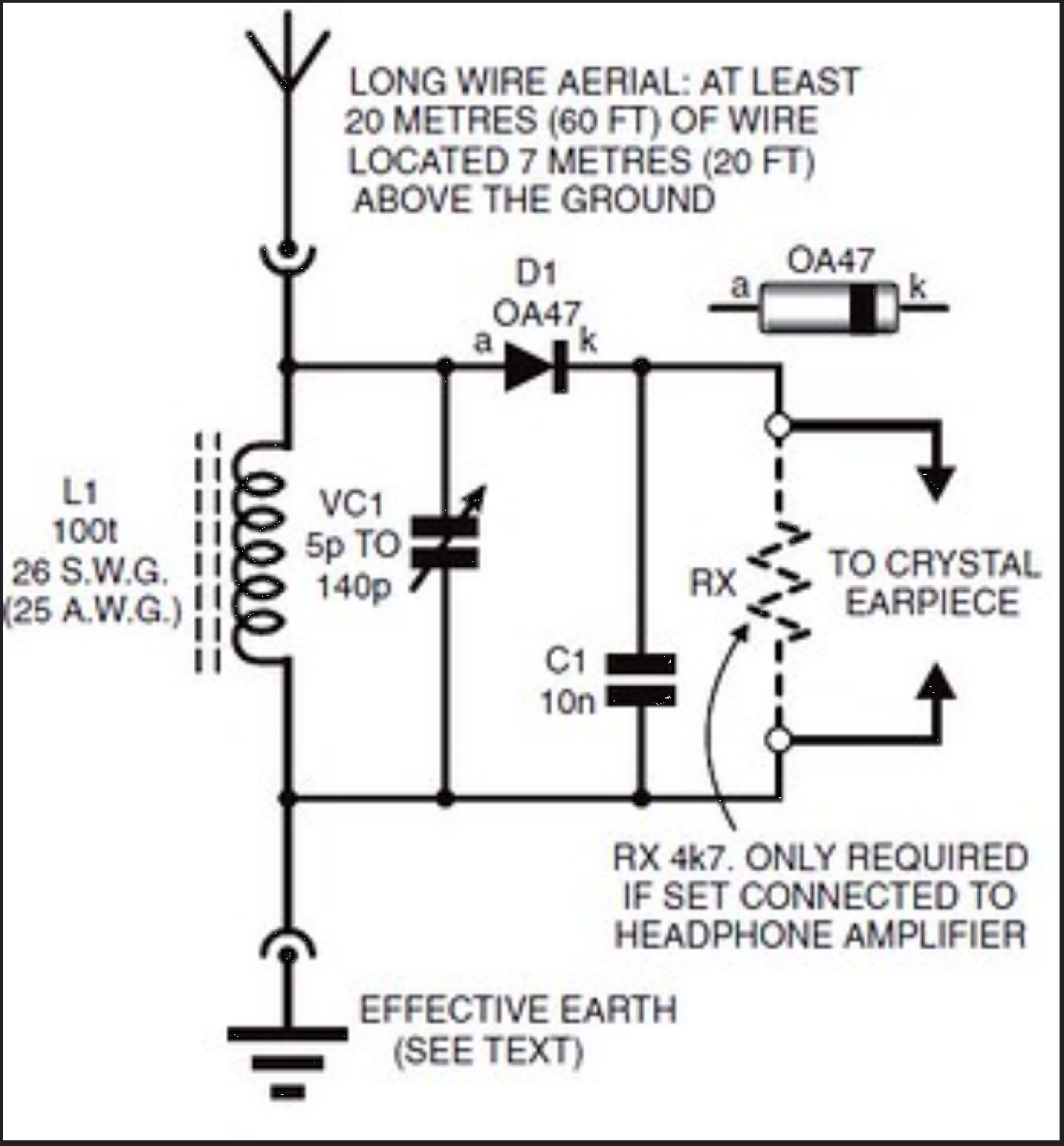
LONG WIRE AERIAL: AT LEAST
20 METRES (60 FT) OF WIRE
LOCATED 7 METRES (20 FT)
ABOVE THE GROUND
D1
OA47
a
k
OA47
a
k
L1
100t
26 S.W.G.
(25 A.W.G.)
VC1
5p TO
140p
RX
TO CRYSTAL
EARPIECE
C1
10n
RX 4k7. ONLY REQUIRED
IF SET CONNECTED TO
HEADPHONE AMPLIFIER
EFFECTIVE EARTH
(SEE TEXT)

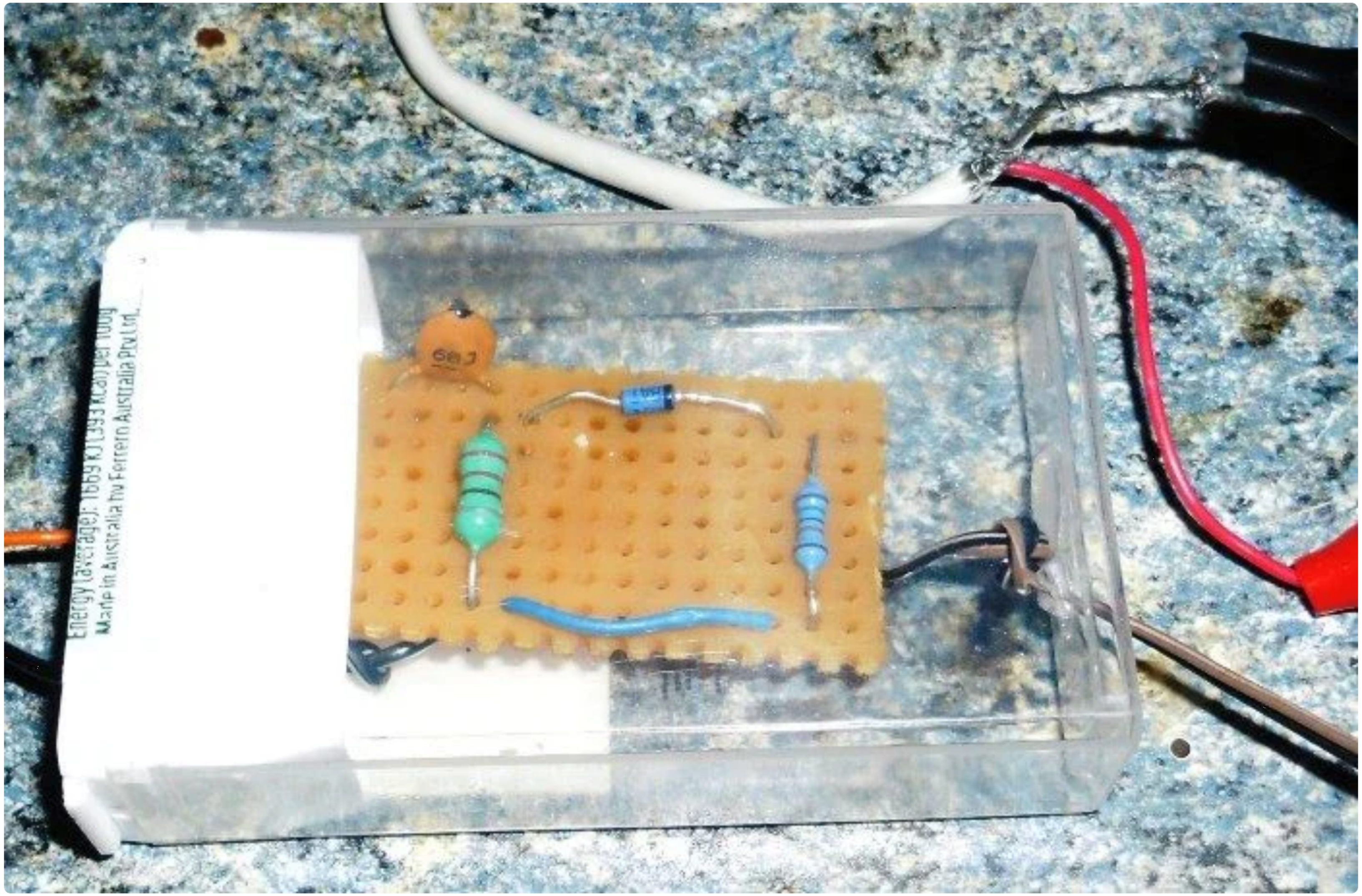
Energy (average): 1669 kJ (393 kcal) per 100g
Made in Australia by Ferrero Australia Pty Ltd.
68J

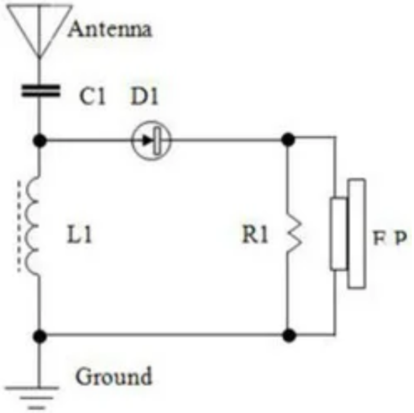
Antenna
C1
D1
L1
R1
F P
Ground

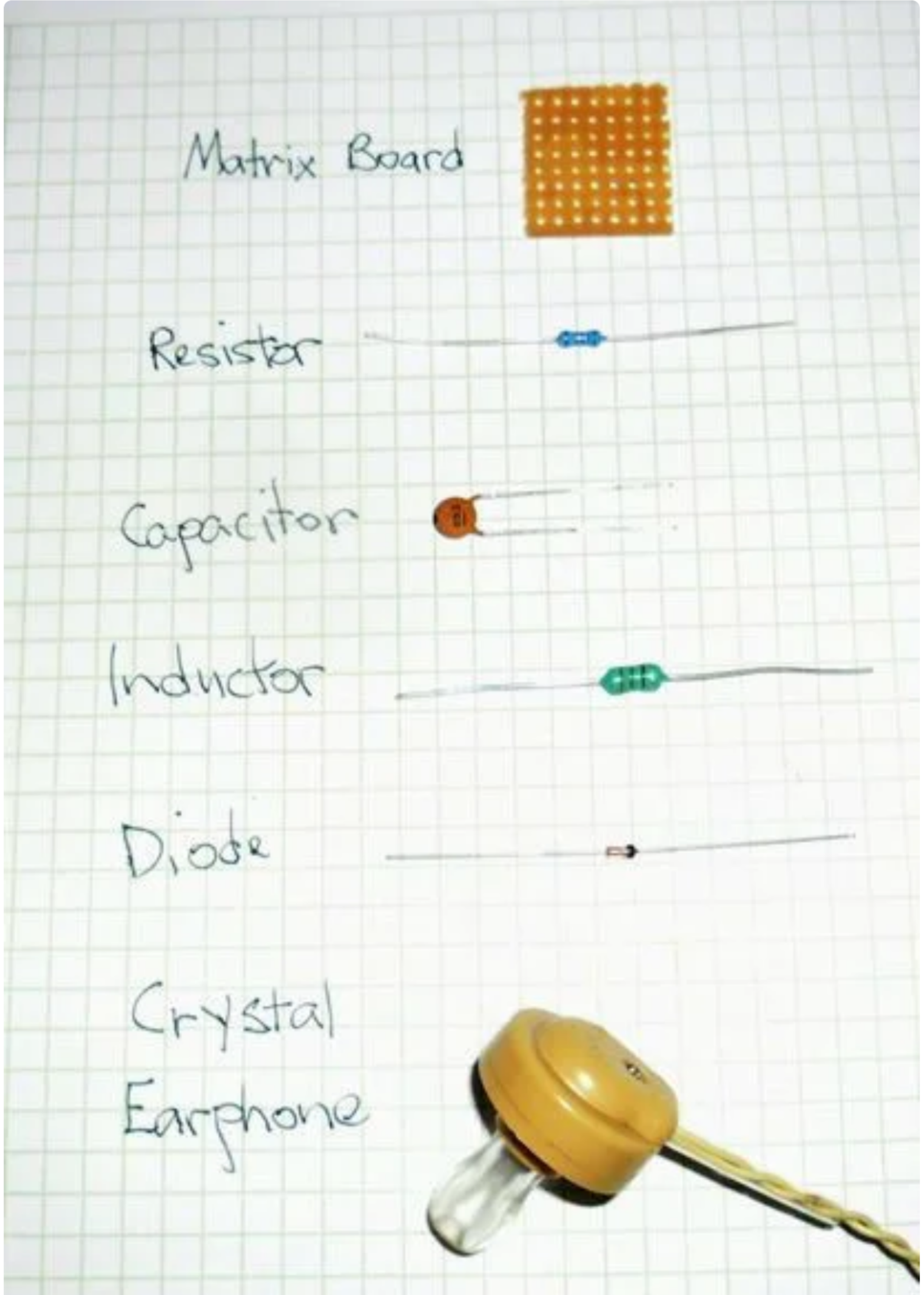
Matrix Board
Resistor
Capacitor
Inductor
Diode
Crystal Earphone

Parts List

Resistor - 47k - yellow-purple-black-red and brown - RR 0612 (pkt of 8)

Capacitor - 68pF - ceramic x 2 - RC 5322 (pkt of 2) and a 100 or 120 pF value as well for experiments.

Inductor - 220 uH - red-red-brown silver - LF 1538 (resistive type)

Polyvaricon tuning capacitor - 220 pF - RV 5728 - with knob and mounting screws

Diode - BAT46 - ZR 1141 (You can also use a 1N34A Germanium Diode too if you have one at home)

Ceramic Earphone - AS 3305 *

A 25 meter roll of yellow hook up wire for the Antenna wire and a 3 meter length of wire for the Ground wire.

Please note that some Jaycar parts come in multiples of 2 or more per packet. And please note the following:

PLEASE NOTE: THIS PROJECT/KIT CONTAINS SMALL PARTS THAT MAY FORM A

*A normal crystal radio earphone is OK, but if you can't get one of these, or if the one you bought goes dead (as they sometimes do,) you can use a substitute, such as the Murata PKM44EW passive transducer (see picture above) which is available from an old Telstra TF200 touchphone, (the one on the left in the diagram above,) or an equivalent, such as the ARIO transducer, from an old Telstra T1000 pushbutton phone.

The ARIO unit is soldered to the phone's pc board so you'll need to be able to unsolder the three mounting pins underneath the board, or find someone in the neighbourhood who is able.

Take the back off the TF200 (if you've obtained one of these phones,) and you'll see a black disc shaped object 2" round by 1/2' thick - with a red and black wire. Unplug the wires from the circuit board, and unscrew any retaining screws and remove the transducer. Cut the mini plug off, **carefully** strip the insulation from the ends of the wires and extend them by about 18" with 2 thin lengths of hookup wire. These pietzo devices make good earphones for crystal sets and can be housed in an old pair of ear muffs.

Miscellaneous Materials:

Miscellaneous Materials:

A Tic Tac box (smaller size)

A piece of matrix board at least 7 holes long by 8 holes across. Cut the board to fit neatly inside the Tic Tac box.

A short length of 2 differently coloured wires 60 mm in length and 2 crocodile clips with red and black plastic covers

A length of antenna wire at least 25 metres long and a 3 metre length of a different colour for the Ground wire.

A metal rod or cold water pipe for the ground stake. **Be careful which pipes you connect your Ground wire to.**

You will need a small sharp object for punching holes in the Tic Tac box.

A small pair of wire cutters for cutting and stripping wires.

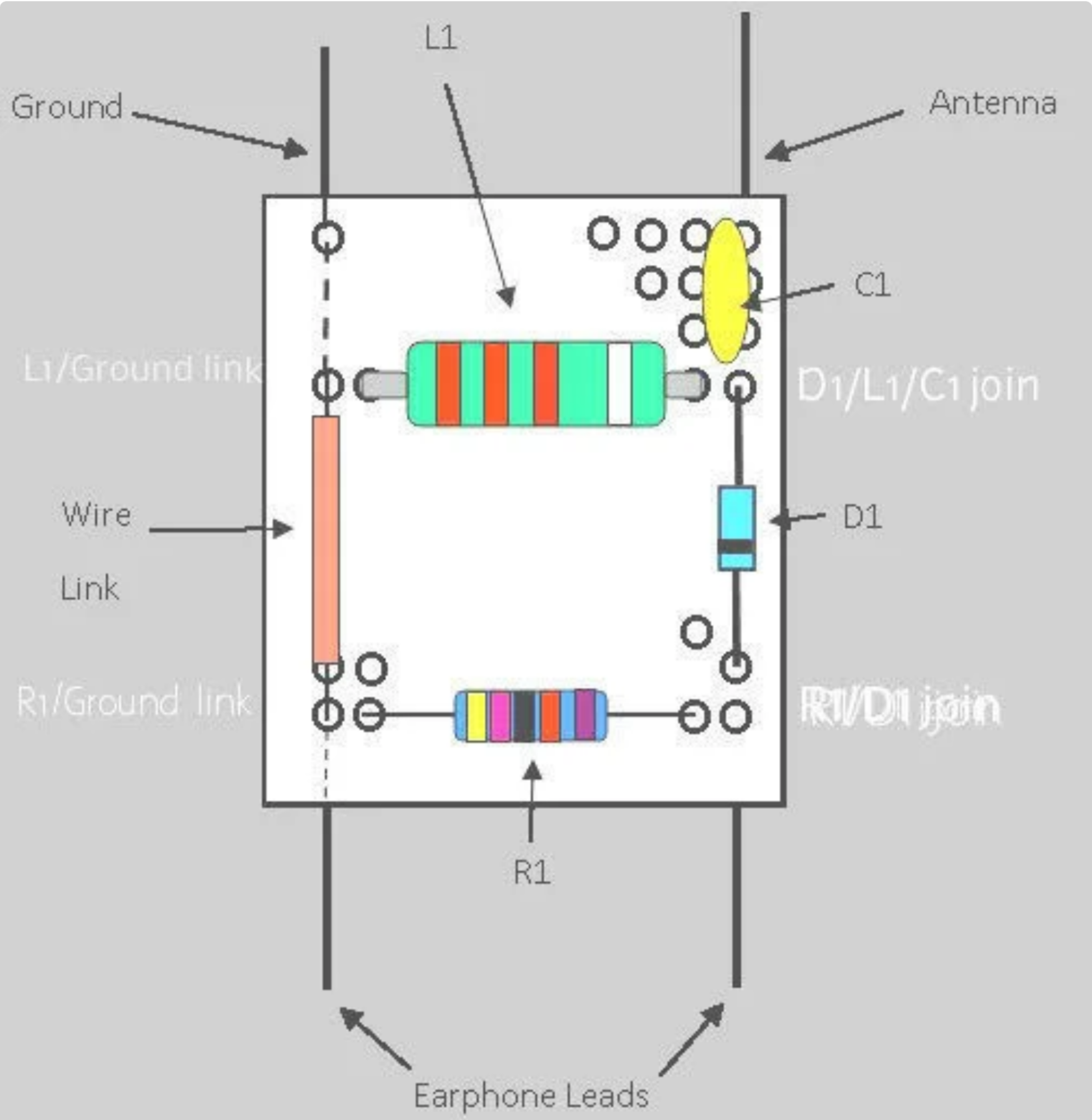
L1
Ground
Antenna
C1
L1/Ground link
D1/L1/C1 join
Wire
Link
D1
R1/Ground link
R1/D1 join
R1
Earphone Leads

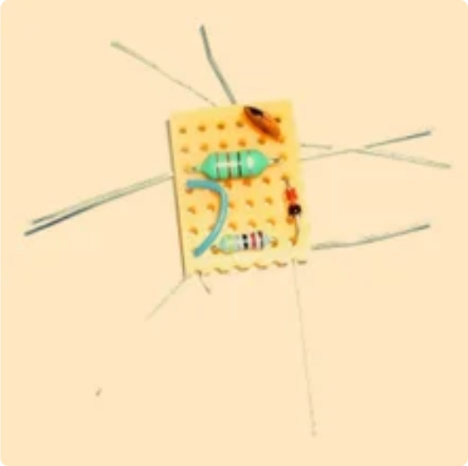

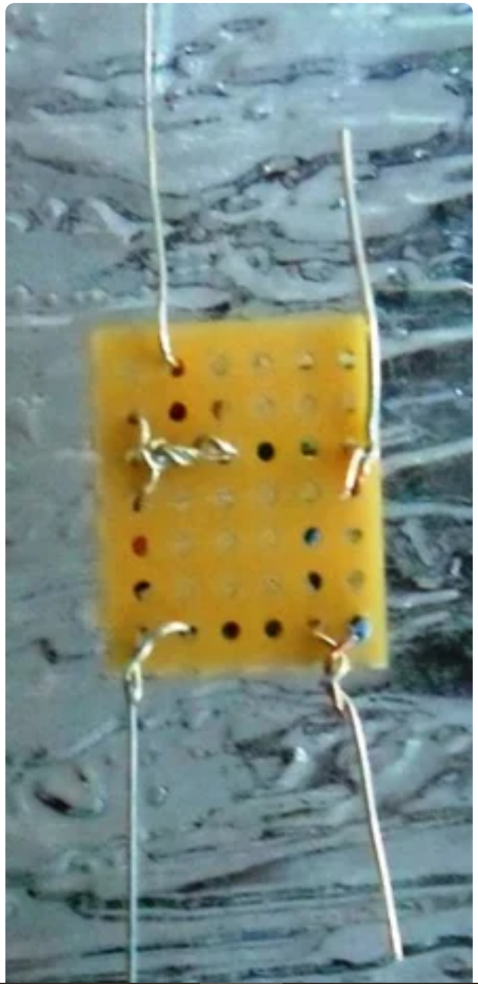

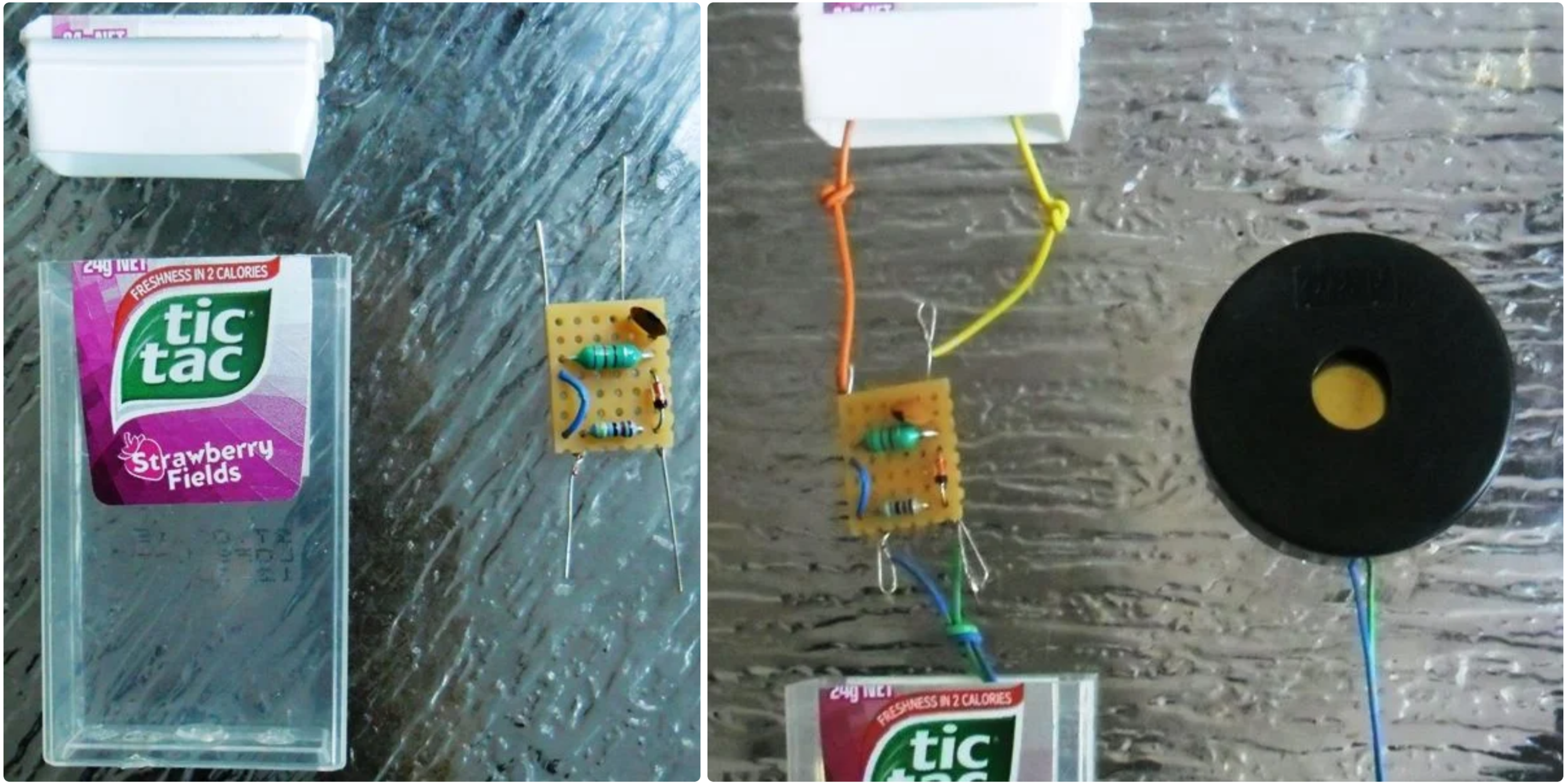
24g NET
FRESHNESS IN 2 CALORIES
tic tac
Strawberry Fields
24g NET
FRESHNESS IN 2 CALORIES
tic tac

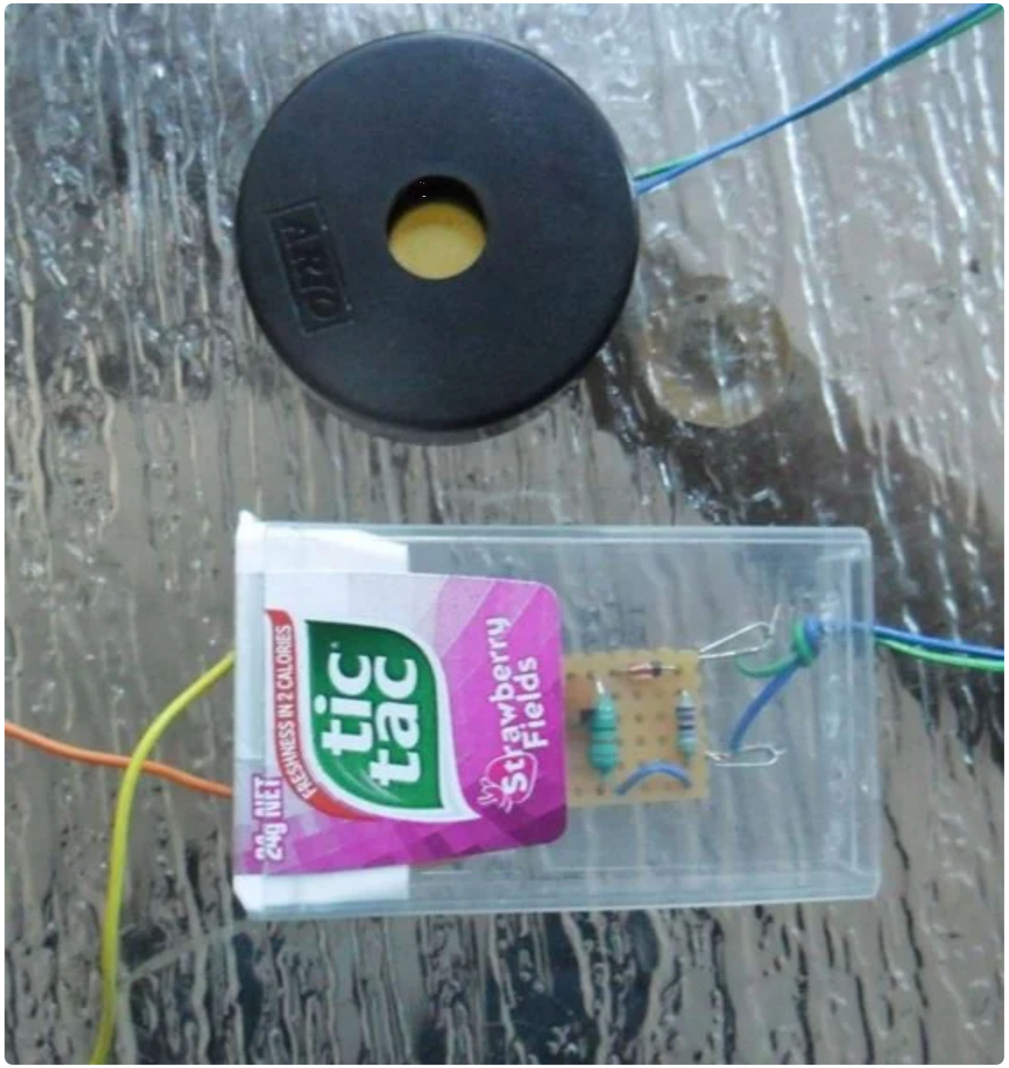

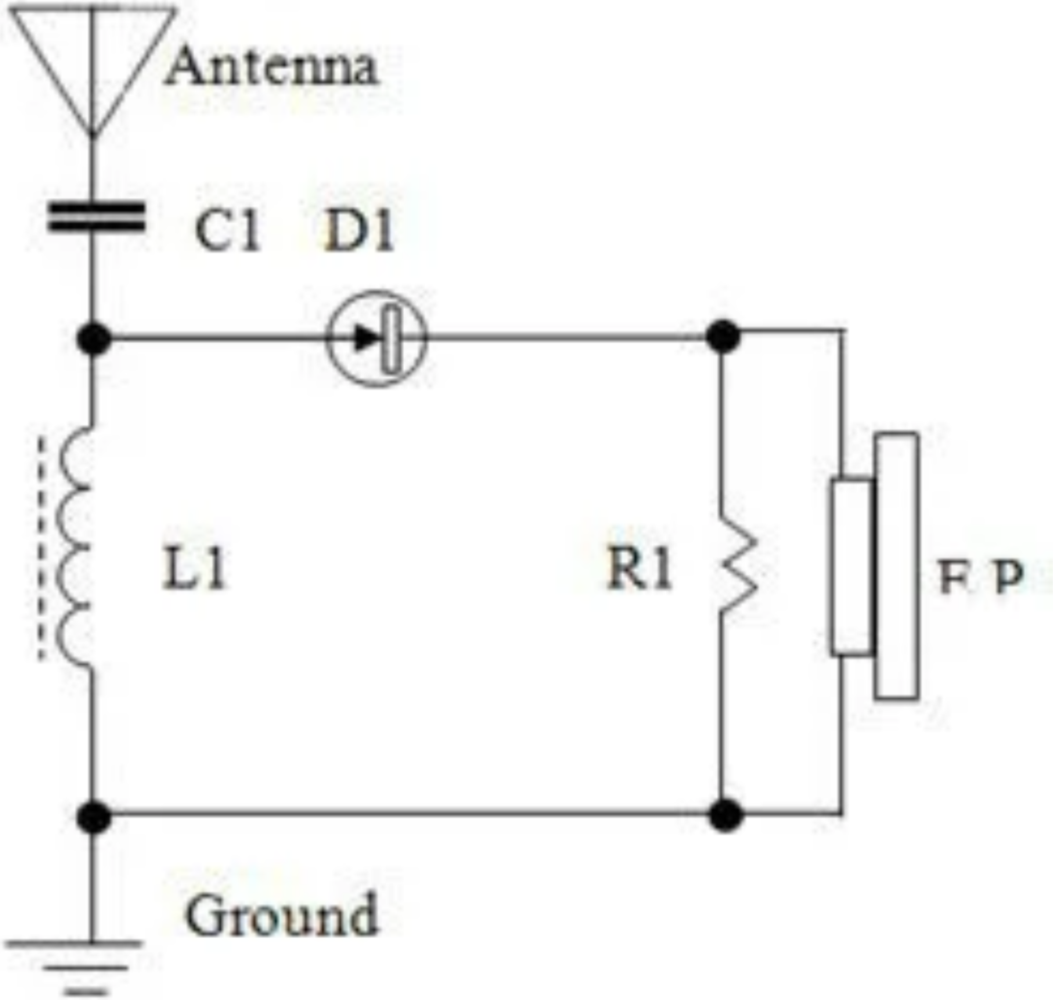
Antenna
C1
D1
L1
R1
F P
Ground

If you can't hear anything in the earphone, don't panic. It might just be a simple wiring mistake, which is easily fixed. Go back over all of the steps, making sure that you have the right value components from the electronics store. Make sure that each component is in the right place on the matrix board, (don't confuse the L1 inductor with the R1 resistor - they look a lot like each other!) and that no component wires have come undone from the twisted joins you have just made. I'll be writing up a troubleshooting step soon, so if you run into any problems, post your questions in the 'comments' section down at the bottom of the page, and I'll try and answer them as soon as possible.

Make sure that your antenna and ground wires aren't snagged on anything metal or anything **dangerous!**

If you have ANY doubts about the electrical safety of your antenna or ground wires, then consult a licensed electrical trades person, who will be able to advise you on electrical safety principles and procedures!

Always remember that electrical safety is your responsibility! If you don't think it is safe to proceed, then don't!

You will need at least 10 to 15 metres of antenna wire, strung between 2 insulating points (not connected to anything metallic, or that gets wet,) at least 2 to 3 metres in height - anything less than this minimum arrangement may mean that you cannot receive any signals at all

Step 7: Adding the Polyvaricon Tuning Capacitor

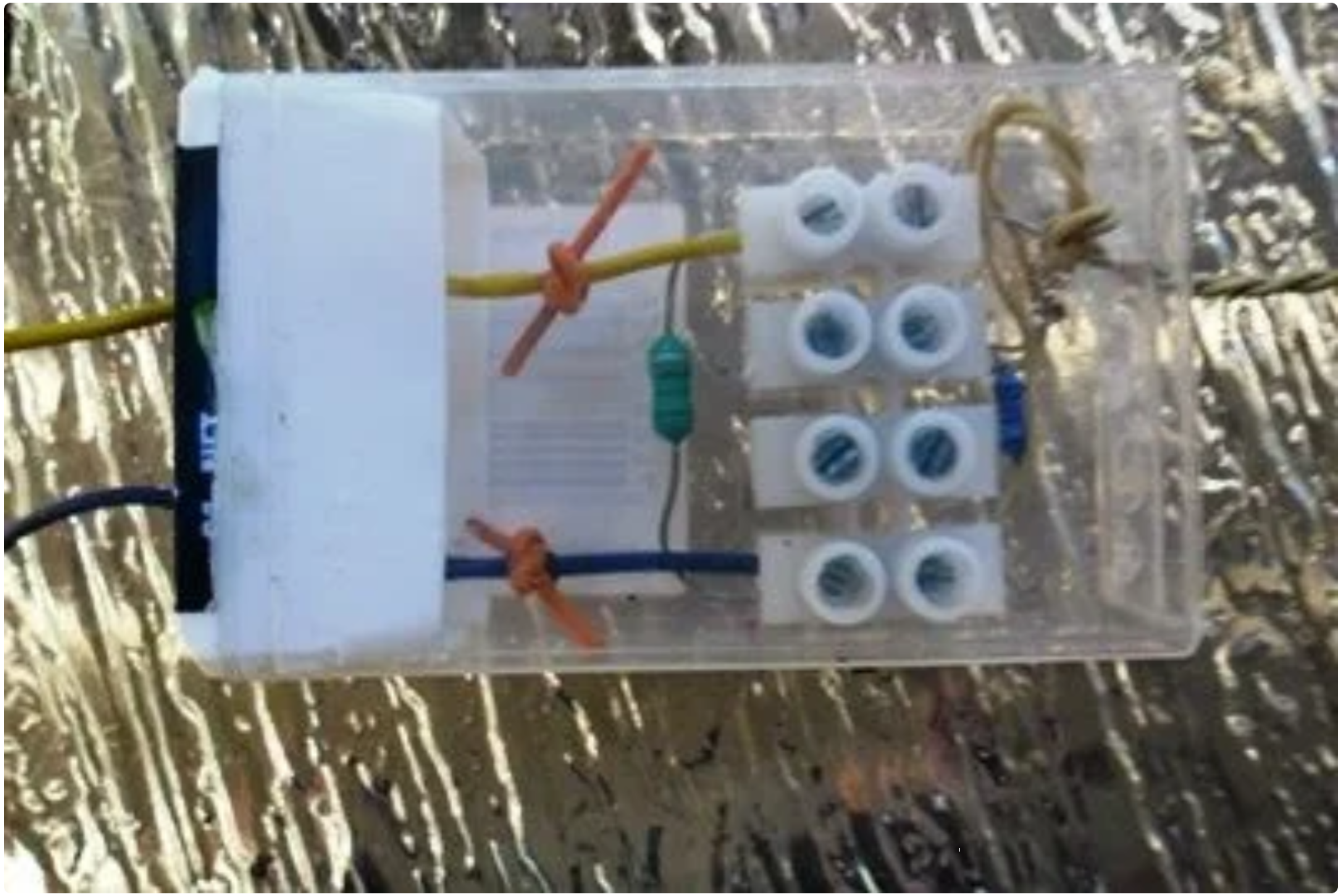

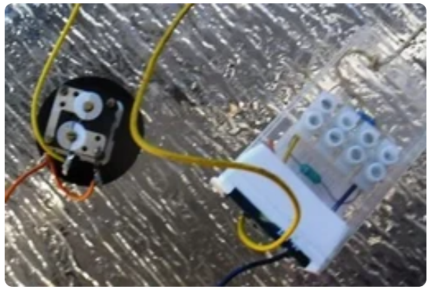

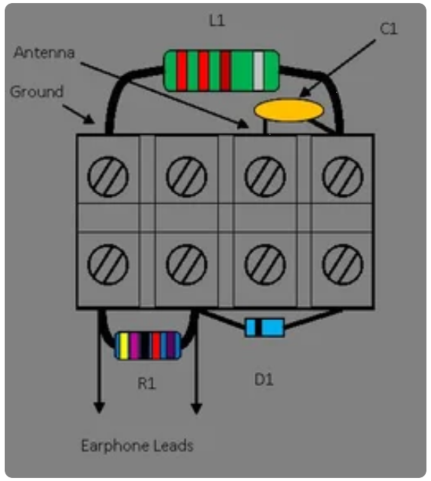
L1
C1
Antenna
Ground
R1
D1
Earphone Leads

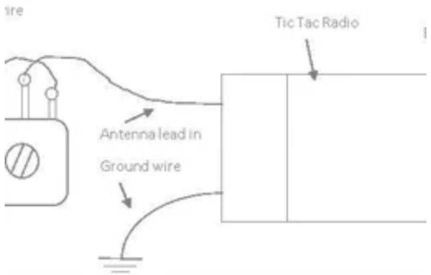
ire
Tic Tac Radio
Antenna lead in
Ground wire

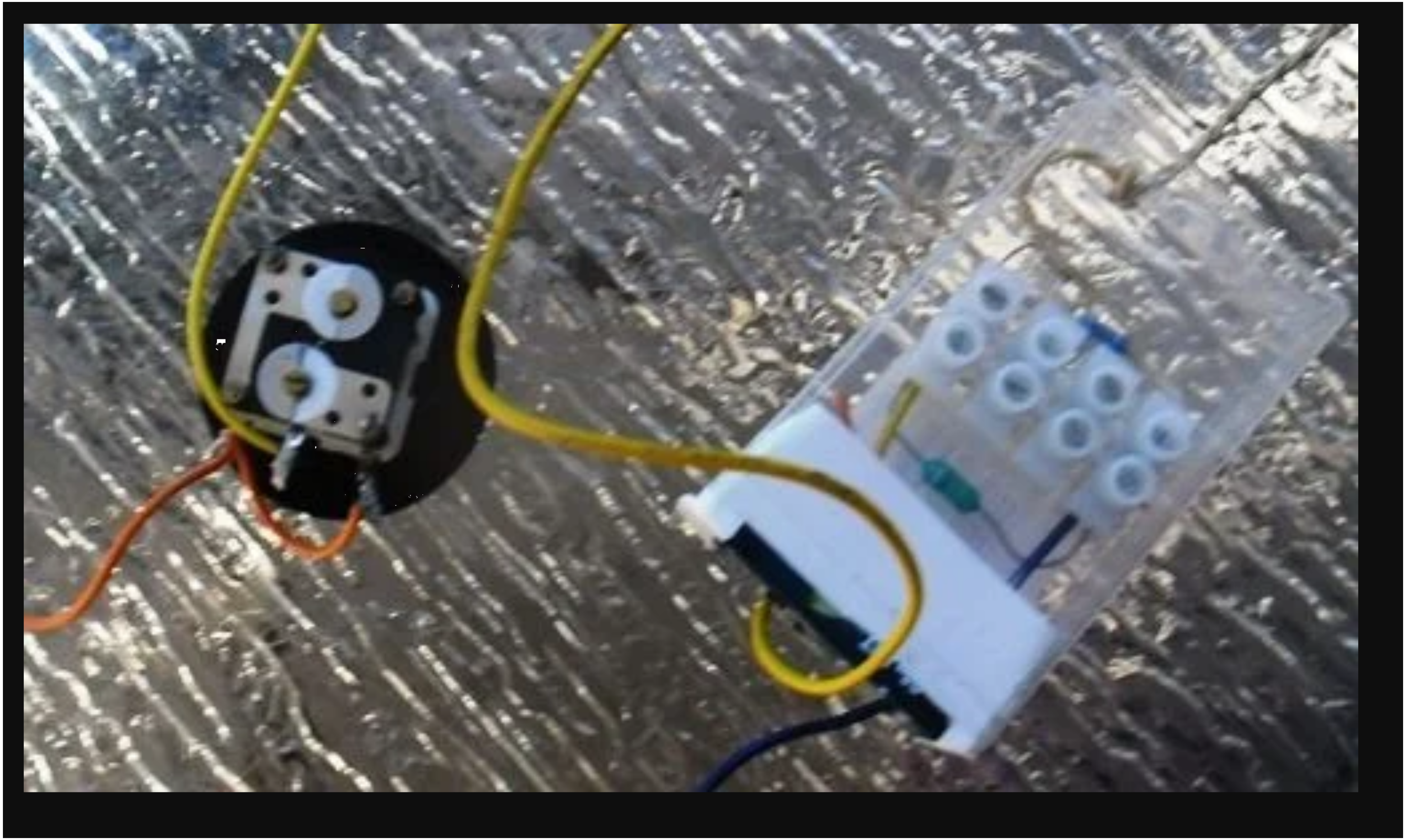

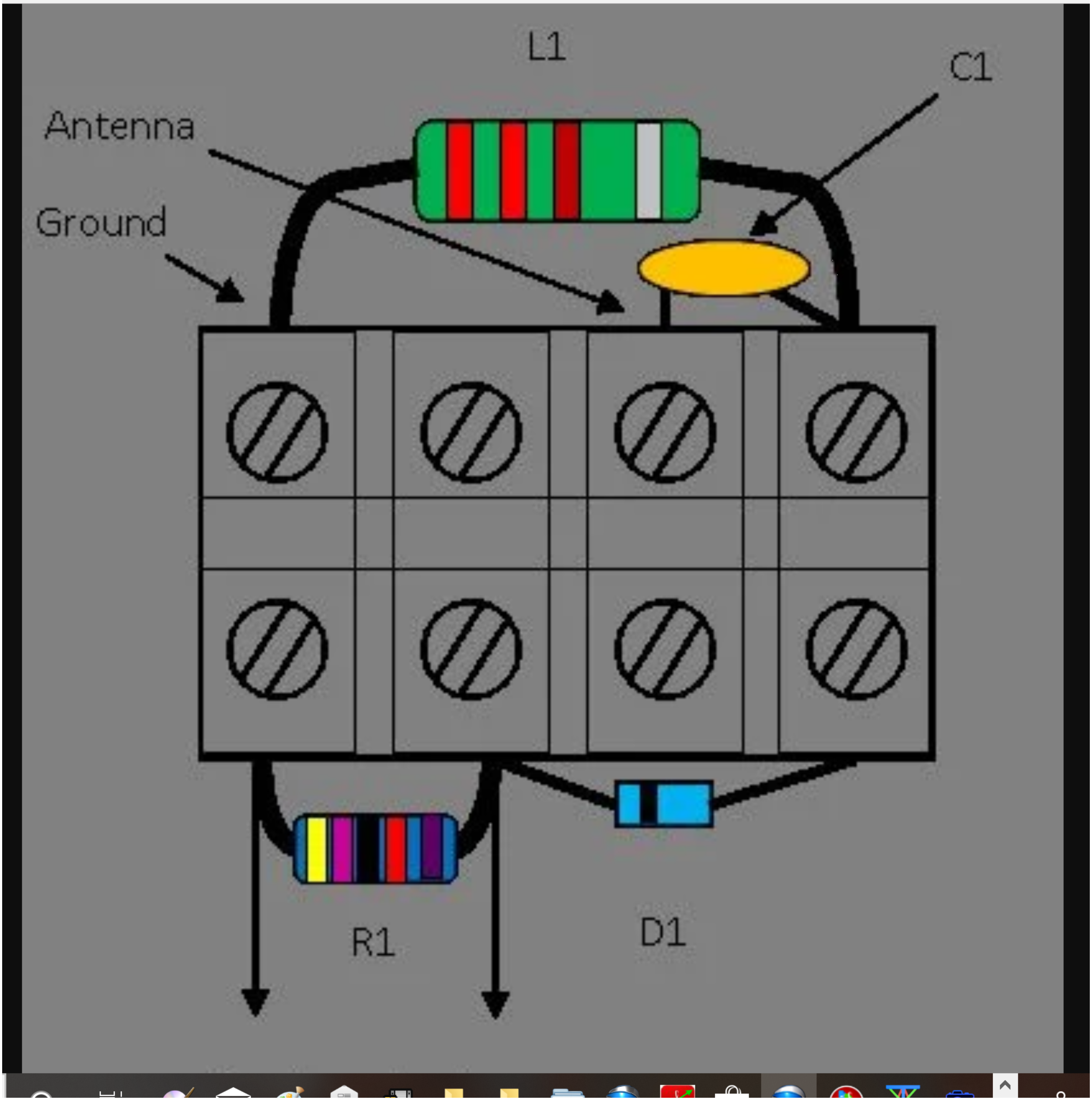
L1
C1
Antenna
Ground
R1
D1

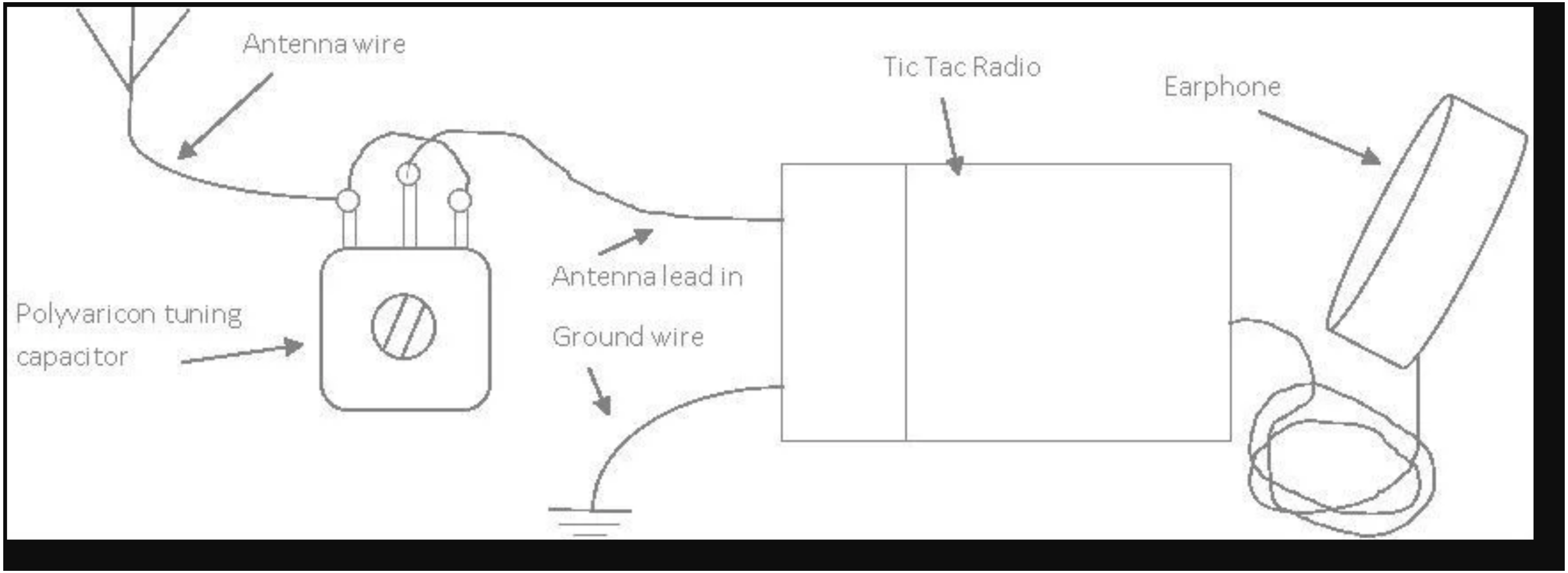
Antenna wire
Tic Tac Radio
Earphone
Antenna lead in
Polyvaricon tuning capacitor
Ground wire

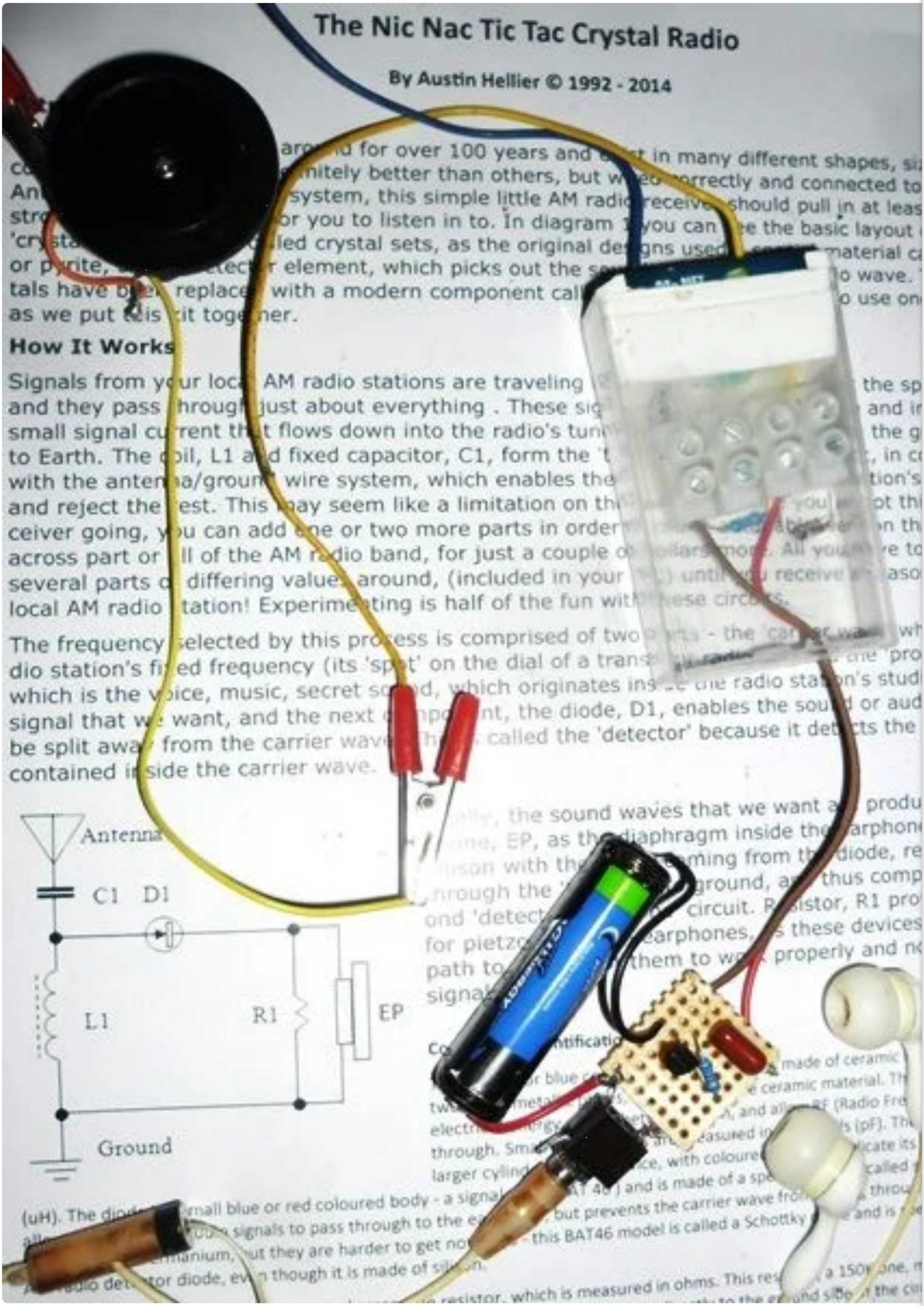

The Nic Nac Tic Tac Crystal Radio

By Austin Hellier © 1992 - 2014

[illegible] around for over 100 years and [illegible] in many different shapes, si[illegible] [illegible] nitely better than others, but w[illegible] orrectly and connected to [illegible] system, this simple little AM radio receiver should pull in at leas[illegible] [illegible] or you to listen in to. In diagram [illegible] you can [illegible] the basic layout [illegible] [illegible] lled crystal sets, as the original designs used [illegible] material c[illegible] or p[illegible] [illegible] element, which picks out the [illegible] o wave. [illegible] tals have been replace[illegible] with a modern component call[illegible] o use on[illegible] as we put this kit together.

How It Works

Signals from your local AM radio stations are traveling [illegible] the sp[illegible] and they pass through just about everything . These sig[illegible] and in[illegible] small signal current that flows down into the radio's tun[illegible] the g[illegible] to Earth. The coil, L1 and fixed capacitor, C1, form the 't[illegible] in c[illegible] with the antenna/ground wire system, which enables the [illegible] tion's and reject the rest. This may seem like a limitation on th[illegible] you [illegible] ot thi[illegible] ceiver going, you can add one or two more parts in order [illegible] on th[illegible] across part or all of the AM radio band, for just a couple of dollars more. All you [illegible] ve to several parts of differing values around, (included in your [illegible]) until you receive [illegible] aso[illegible] local AM radio station! Experimenting is half of the fun with these circuits.

The frequency selected by this process is comprised of two [illegible] - the 'carrier wa[illegible] wh[illegible] dio station's fixed frequency (its 'spot' on the dial of a trans[illegible] the 'pro[illegible] which is the voice, music, secret sound, which originates inside the radio station's studi[illegible] signal that we want, and the next component, the diode, D1, enables the sound or aud[illegible] be split away from the carrier wave. This is called the 'detector' because it detects the [illegible] contained inside the carrier wave.

[illegible] the sound waves that we want a[illegible] produ[illegible] [illegible] EP, as the diaphragm inside the earphone[illegible] [illegible] with th[illegible] ming from the diode, re[illegible] [illegible] hrough the [illegible] ground, a[illegible] thus comp[illegible] ond 'detect[illegible] circuit. Resistor, R1 pro[illegible] for pietzo[illegible] earphones, [illegible] these devices path to [illegible] them to work properly and n[illegible] signal[illegible]

Co[illegible] ntificatio[illegible]

[illegible] made of ceramic [illegible] r blue [illegible] ceramic material. Th[illegible] two [illegible] metal [illegible] and allow RF (Radio Fre[illegible] electr[illegible] easured in [illegible] s (pF). Th[illegible] through. Sma[illegible] ce, with coloure[illegible] icate its larger cylind[illegible] called [illegible] (uH). The diode [illegible] mall blue or red coloured body - a signal [illegible] AT 40) and is made of a sp[illegible] throu[illegible] all[illegible] signals to pass through to the e[illegible] but prevents the carrier wave fro[illegible] and is [illegible] [illegible] manium, but they are harder to get no[illegible] - this BAT46 model is called a Schottky [illegible] radio detector diode, even though it is made of silicon.

[illegible] resistor, which is measured in ohms. This res[illegible] a 150[illegible] one, [illegible] to the ground side of the ci[illegible]

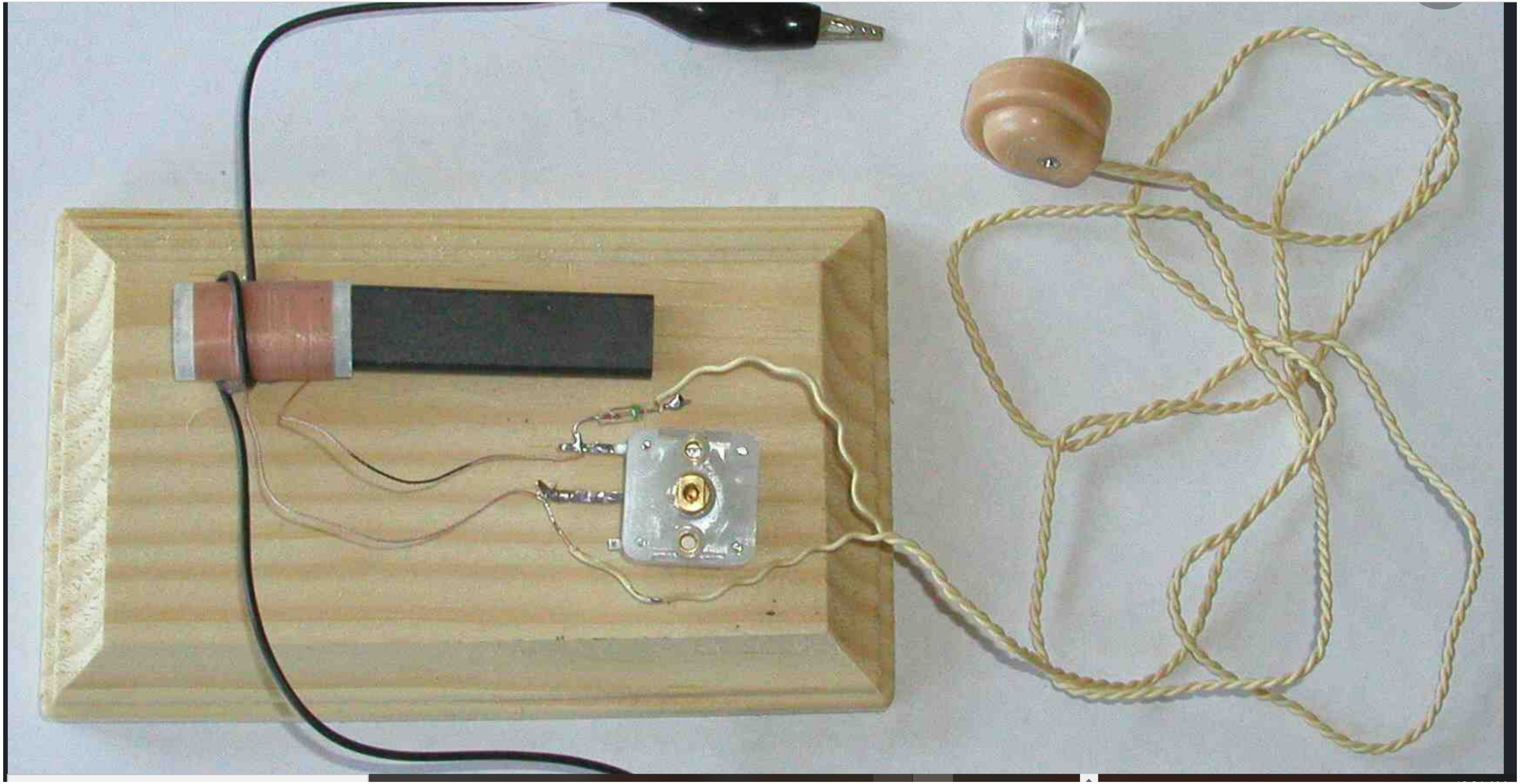

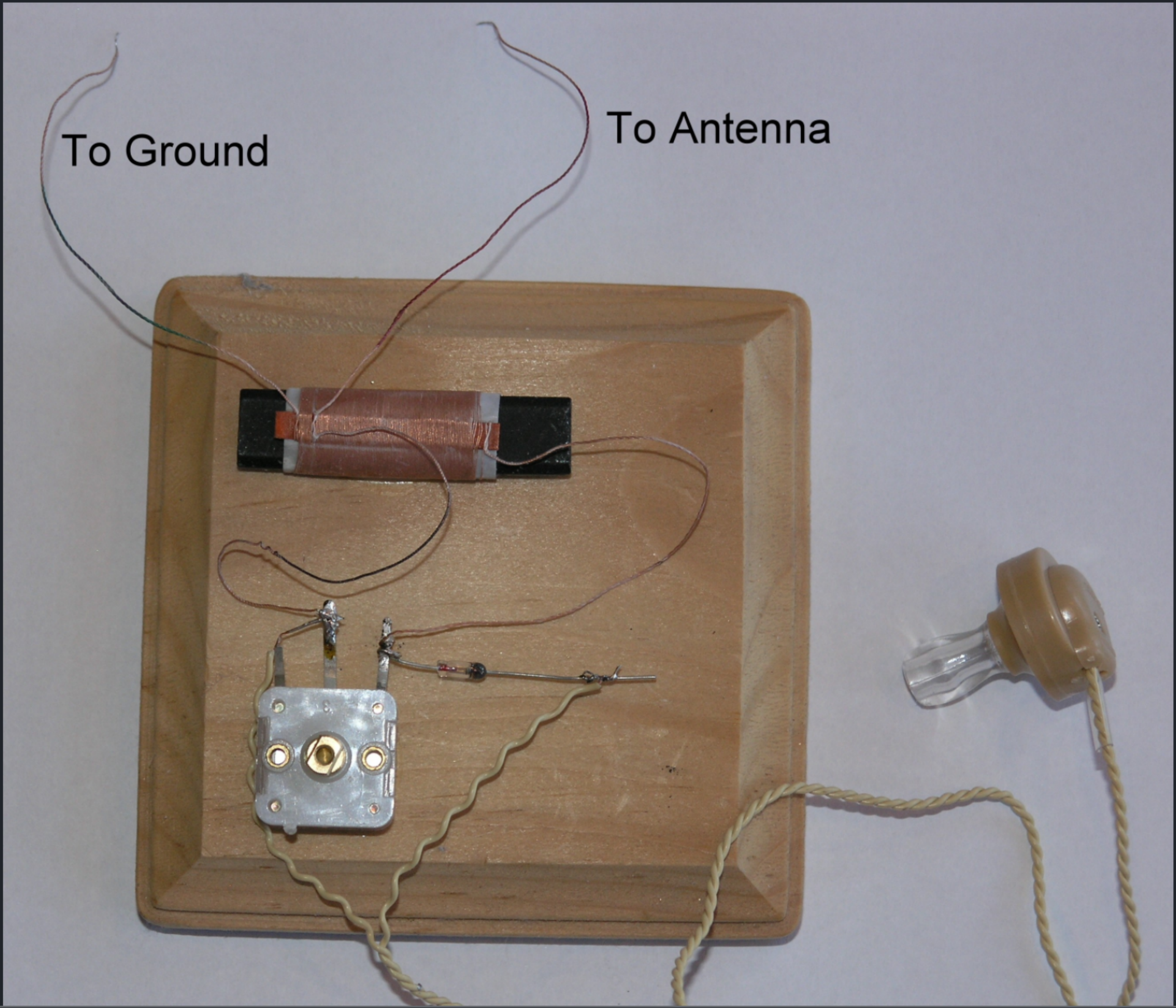
To Ground
To Antenna

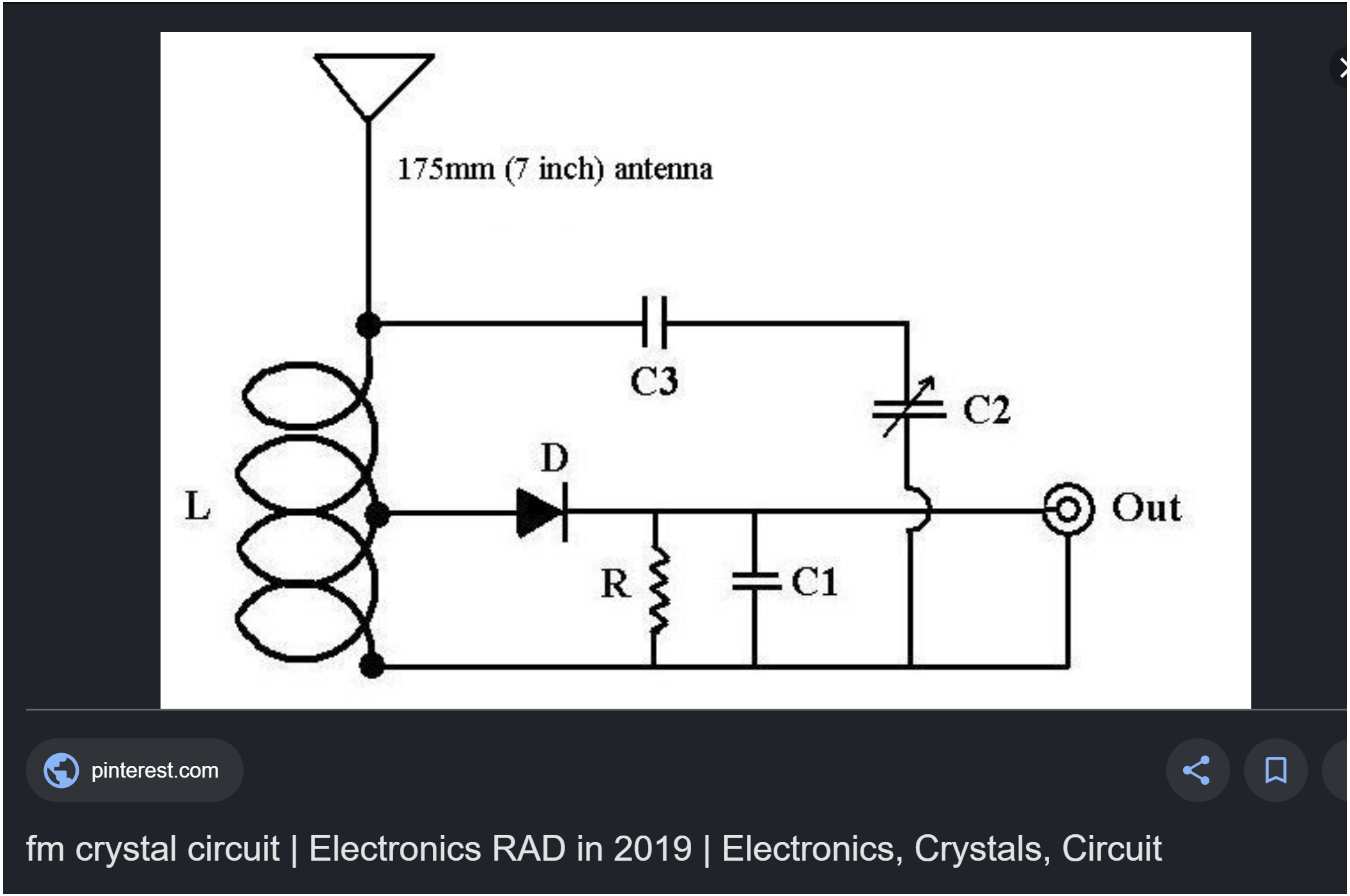
175mm (7 inch) antenna
C3
C2
D
L
Out
R
C1

pinterest.com

fm crystal circuit | Electronics RAD in 2019 | Electronics, Crystals, Circuit

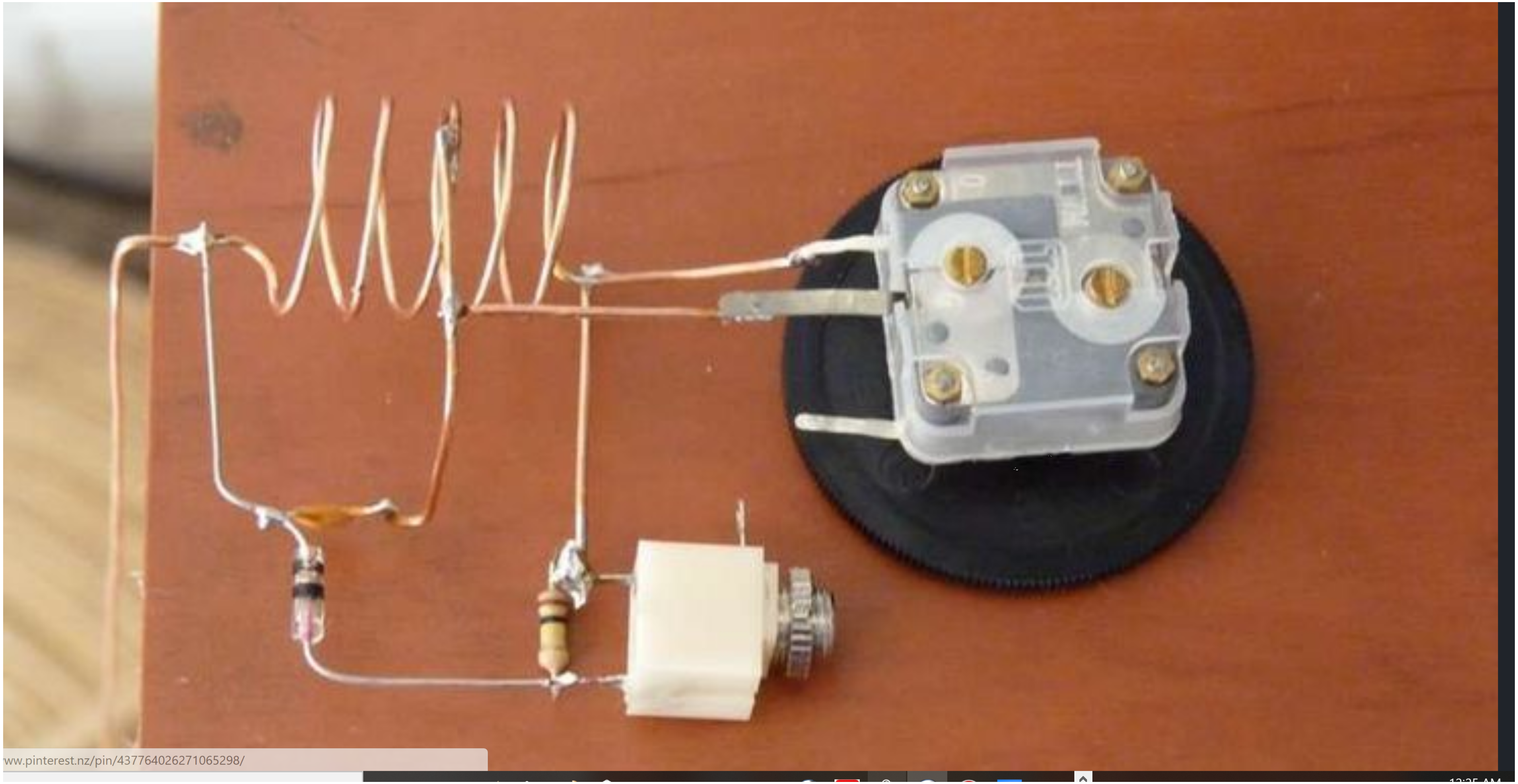
ww.pinterest.nz/pin/437764026271065298/

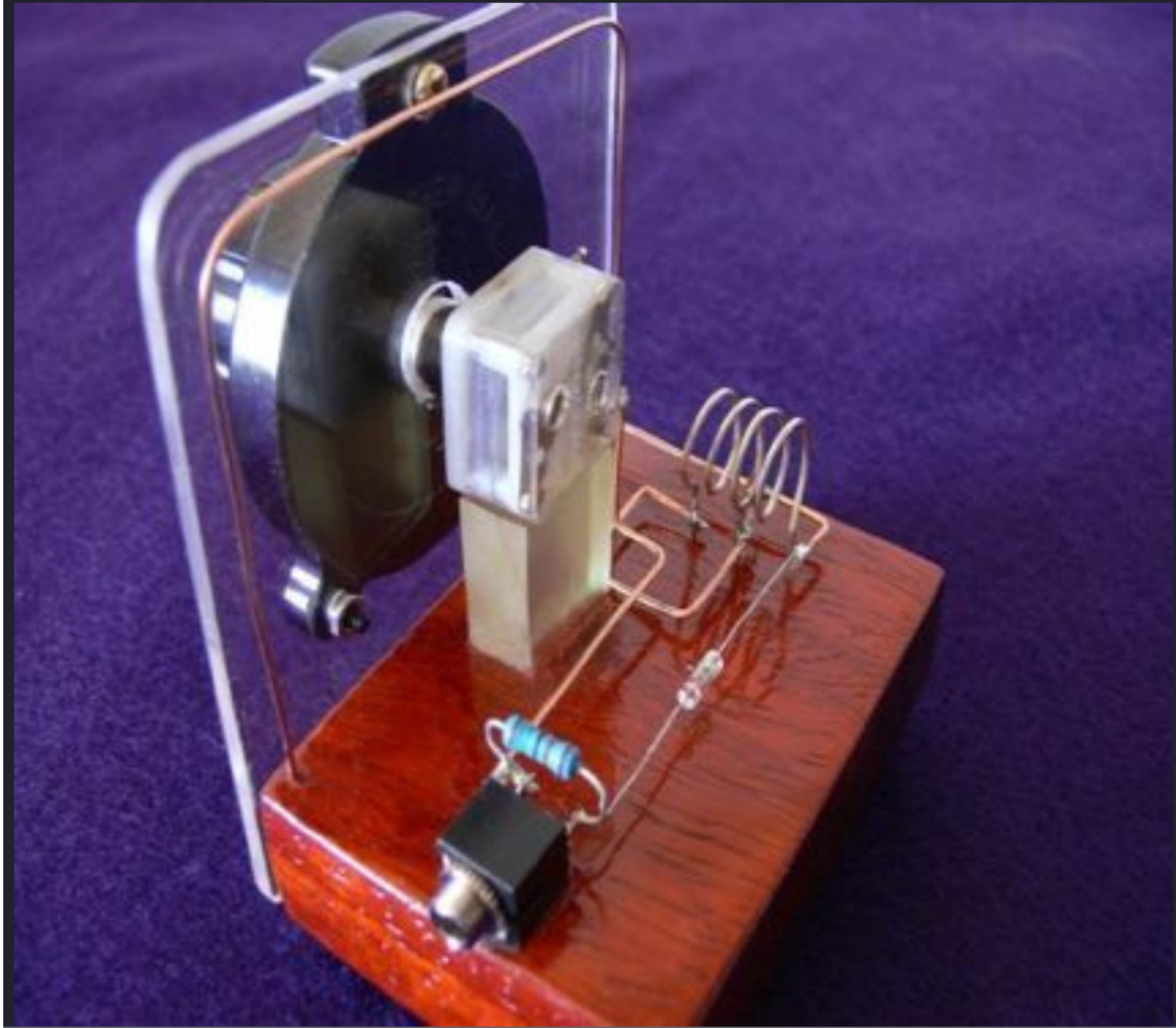

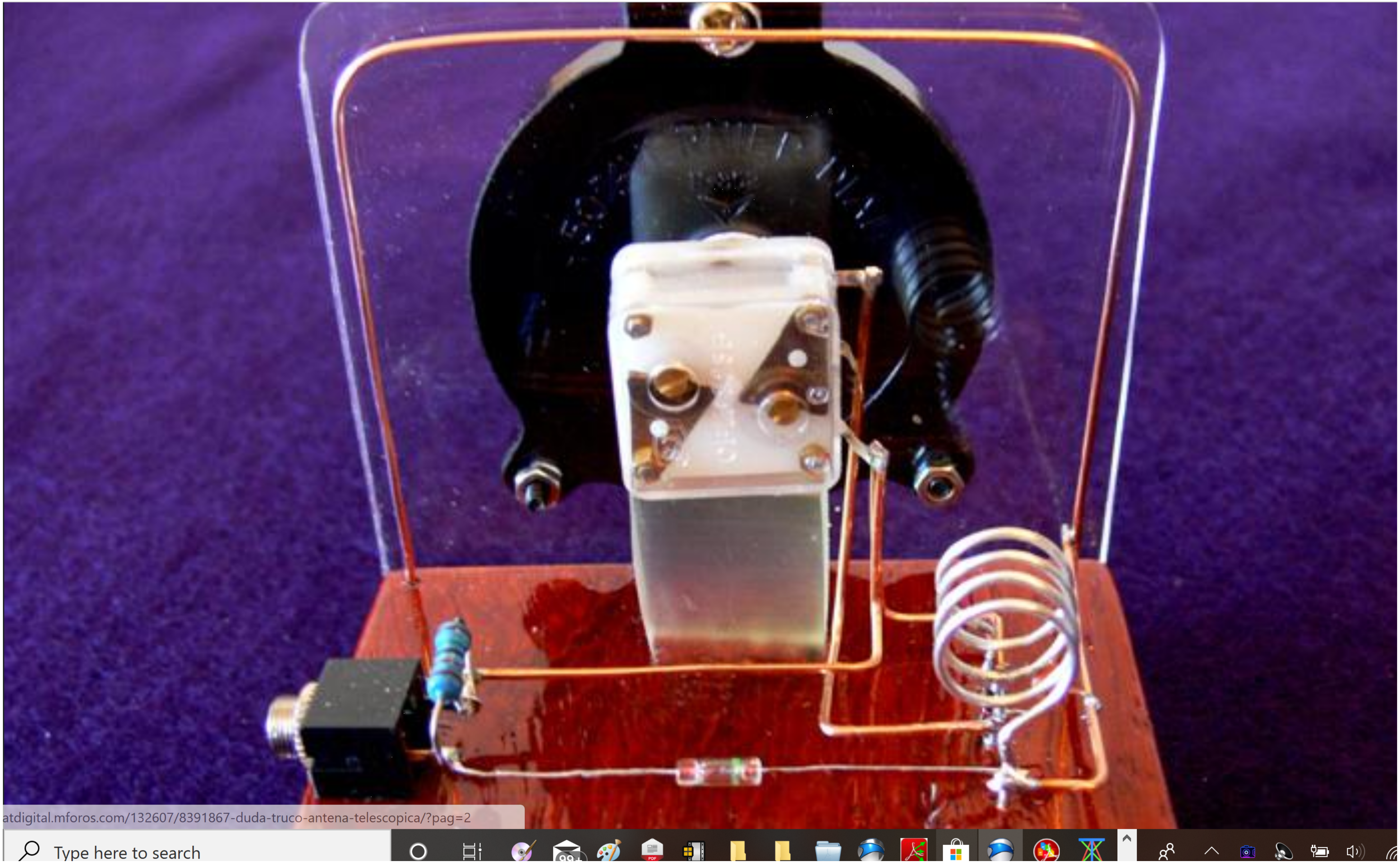

atdigital.mforos.com/132607/8391867-duda-truco-antena-telescopica/?pag=2

Type here to search

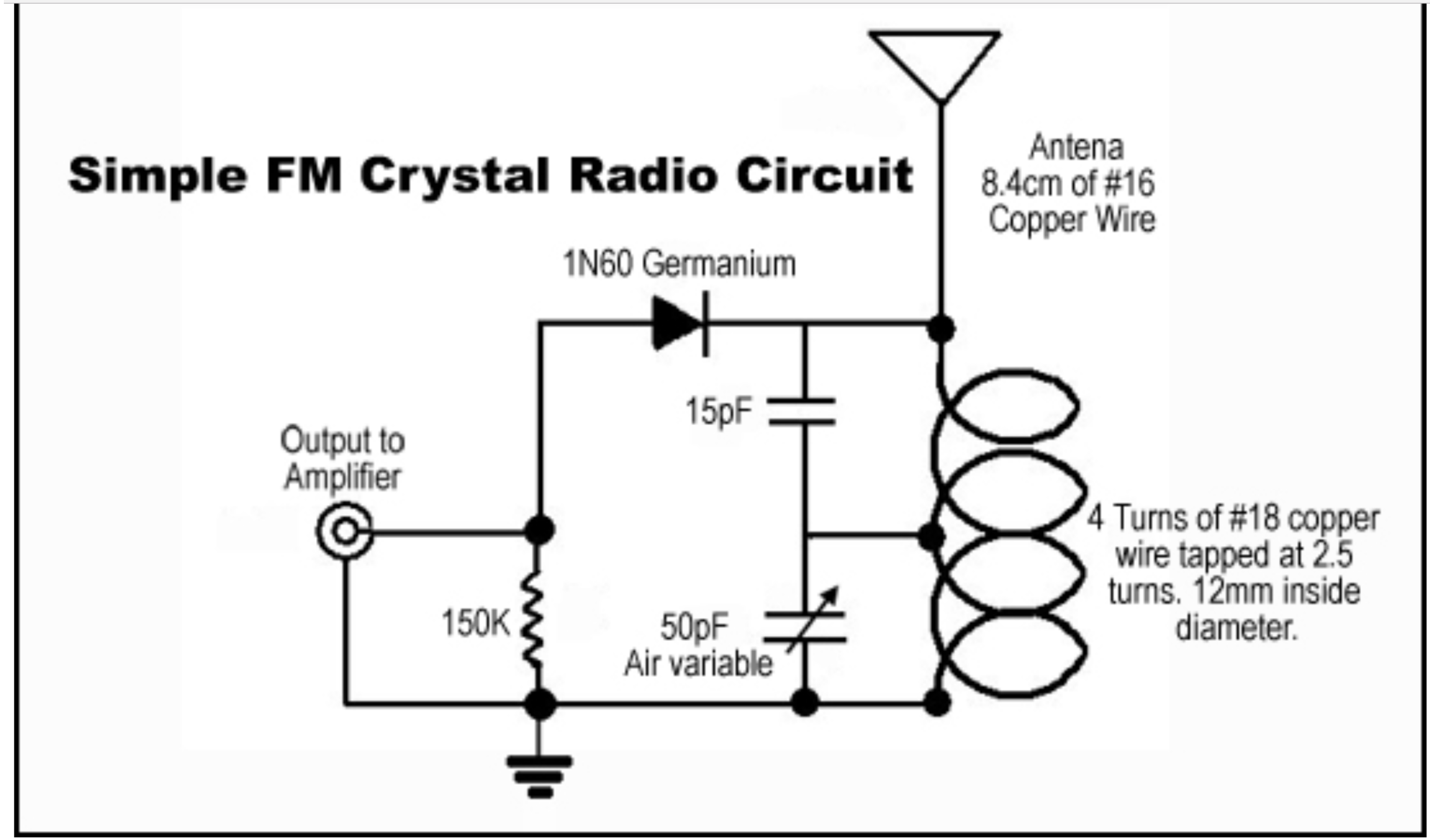
Simple FM Crystal Radio Circuit
Antena
8.4cm of #16
Copper Wire
1N60 Germanium
15pF
Output to
Amplifier
4 Turns of #18 copper
wire tapped at 2.5
turns. 12mm inside
diameter.
150K
50pF
Air variable

Parts List (some of these parts you can buy from our online store):

- 1N60 Germanium Diode
- 15pF Ceramic Capacitor
- 50pF Variable Capacitor
- 150K Ohm Resistor
- #16 & #18 Copper wires

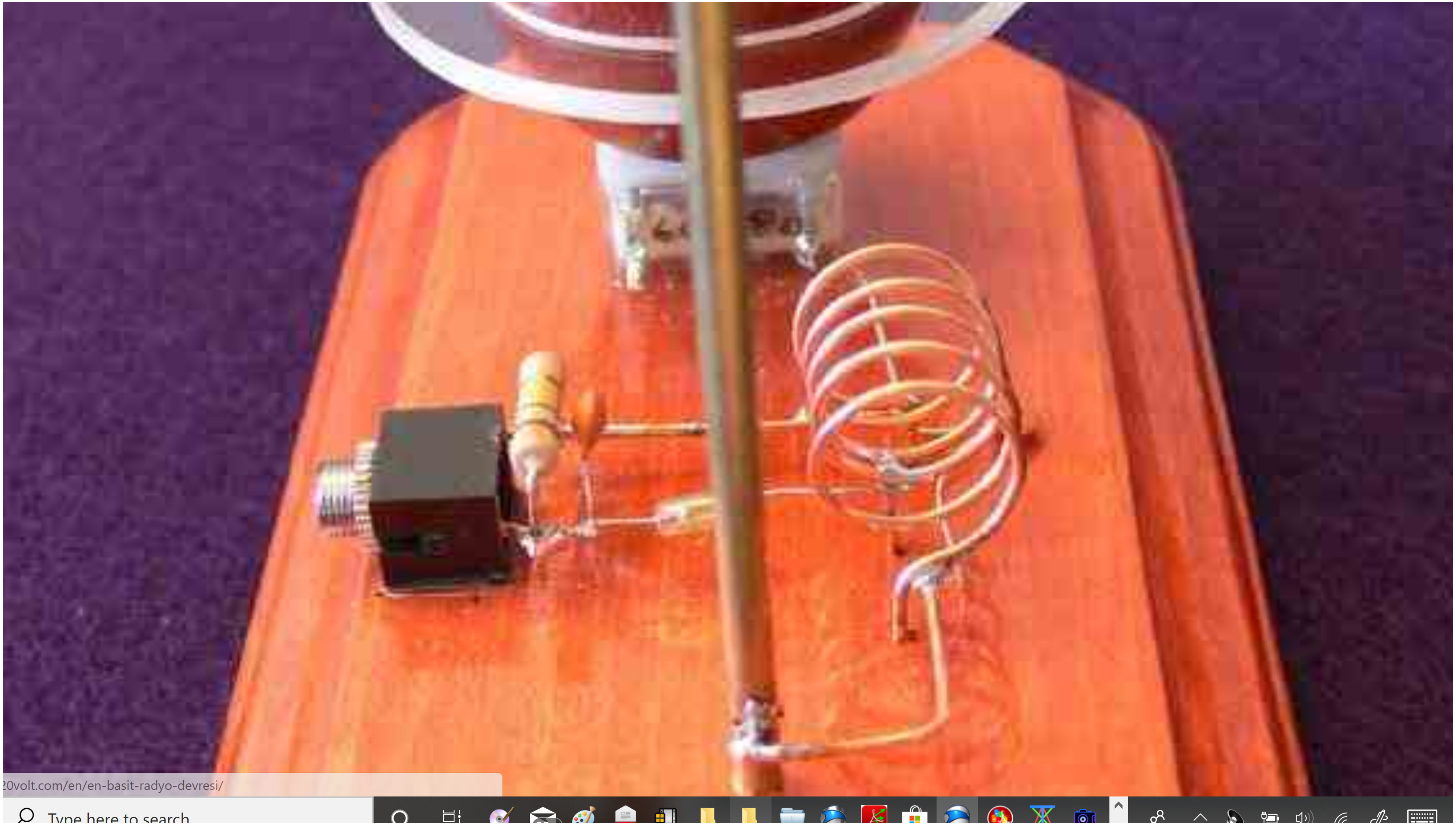
20volt.com/en/en-basit-radyo-devresi/
Type here to search

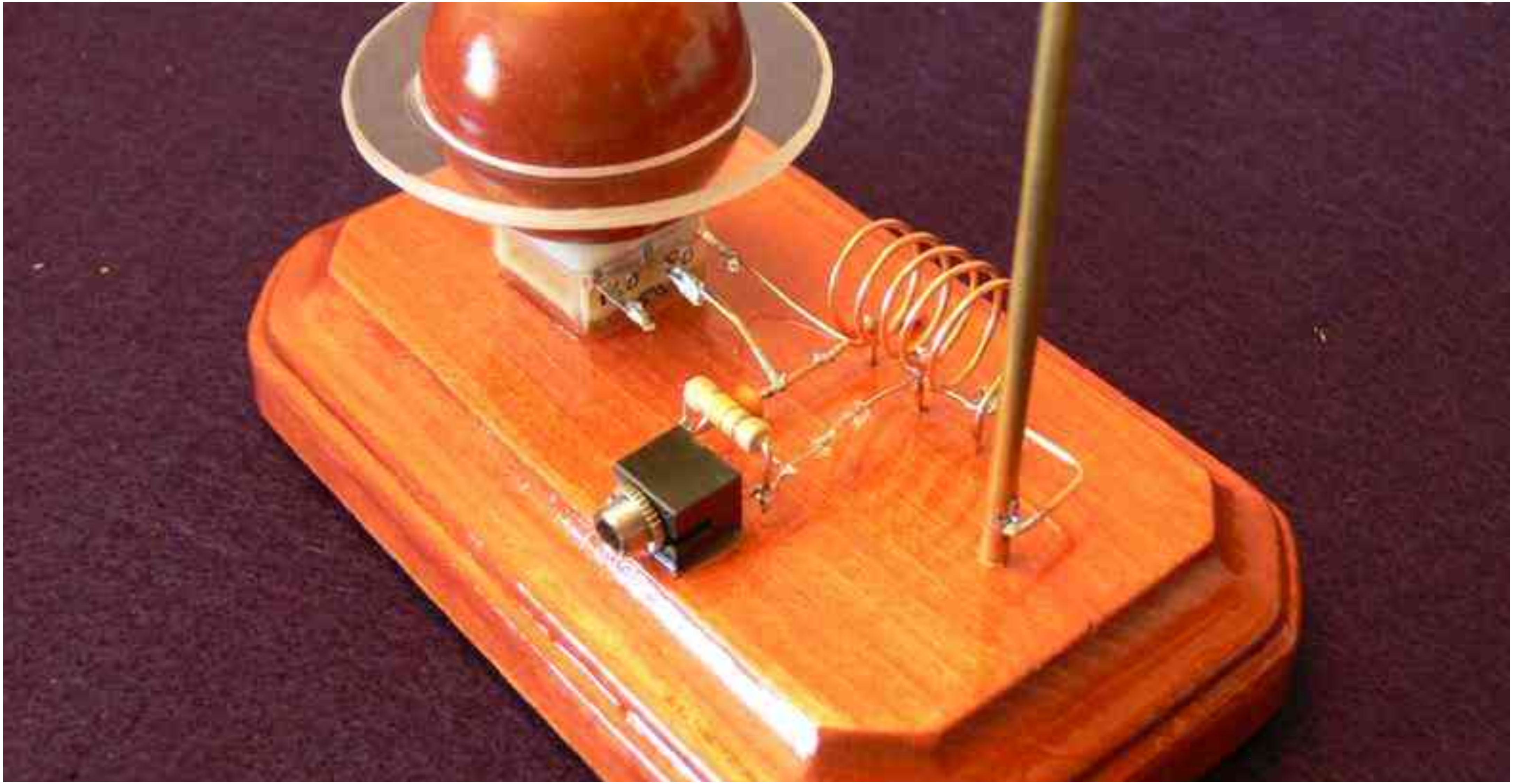

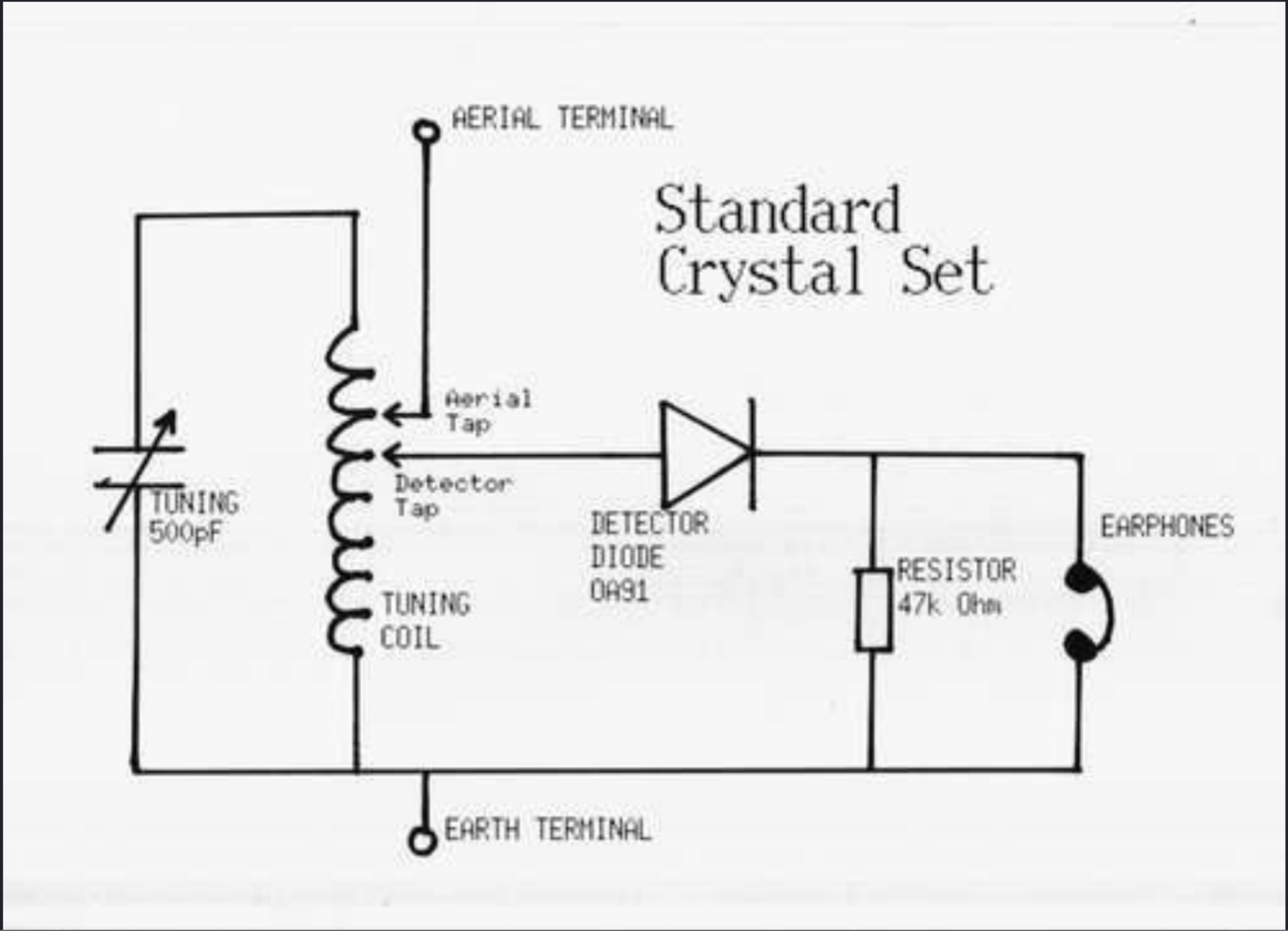

Standard Crystal Set

AERIAL TERMINAL
Aerial Tap
Detector Tap
TUNING 500pF
TUNING COIL
DETECTOR DIODE OA91
RESISTOR 47k Ohm
EARPHONES
EARTH TERMINAL

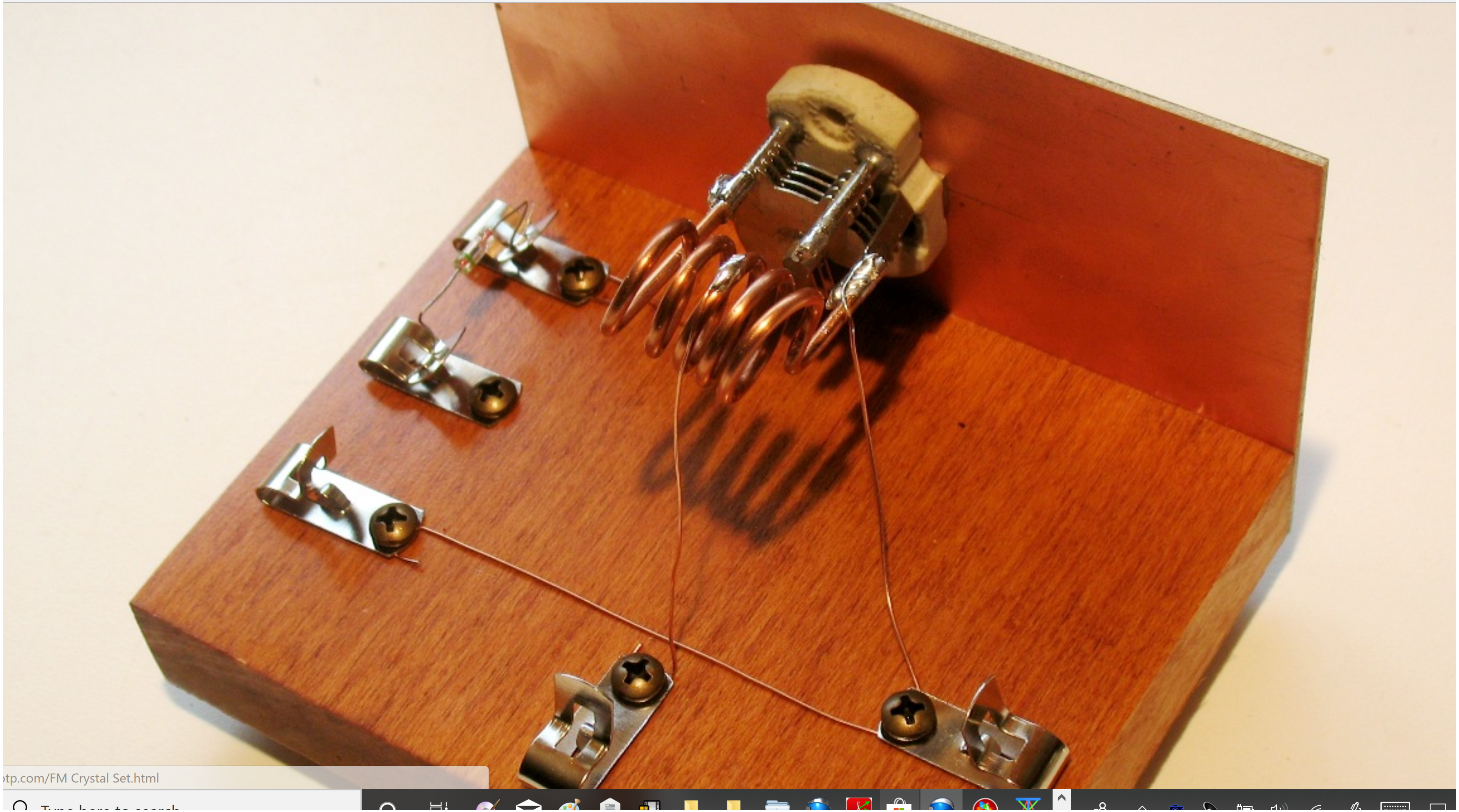
tp.com/FM Crystal Set.html
Type here to search

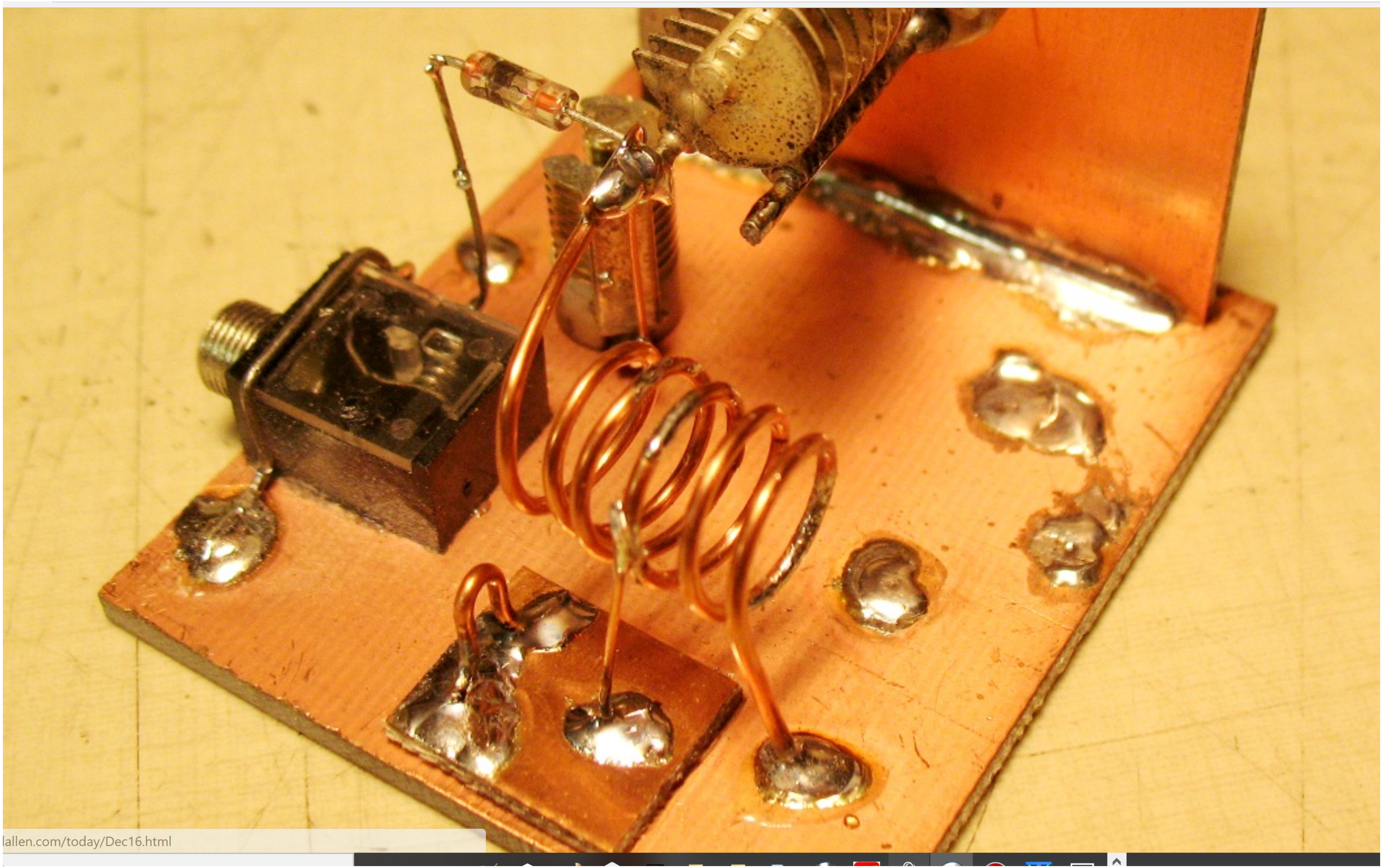

lallen.com/today/Dec16.html

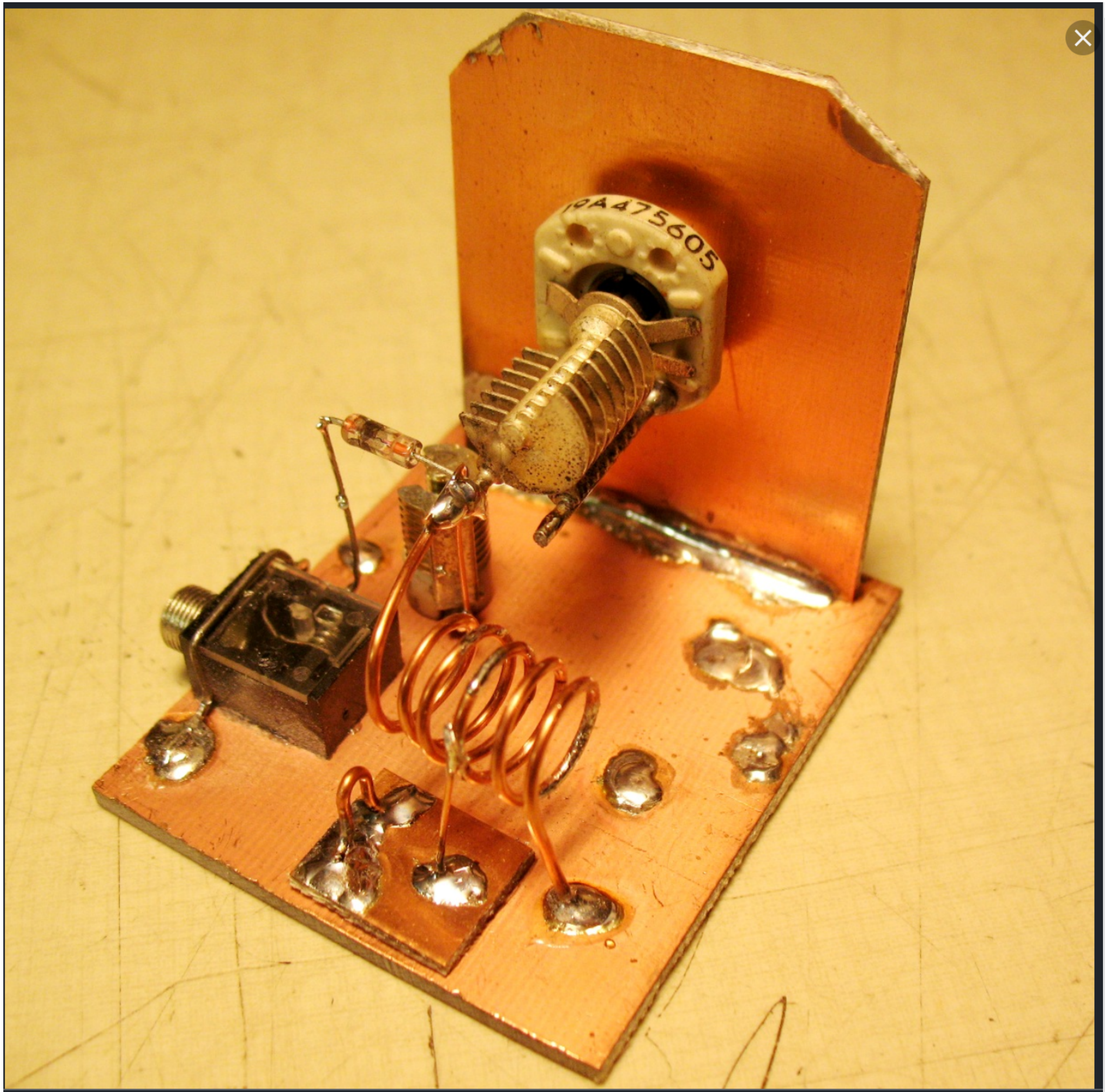
19A475605

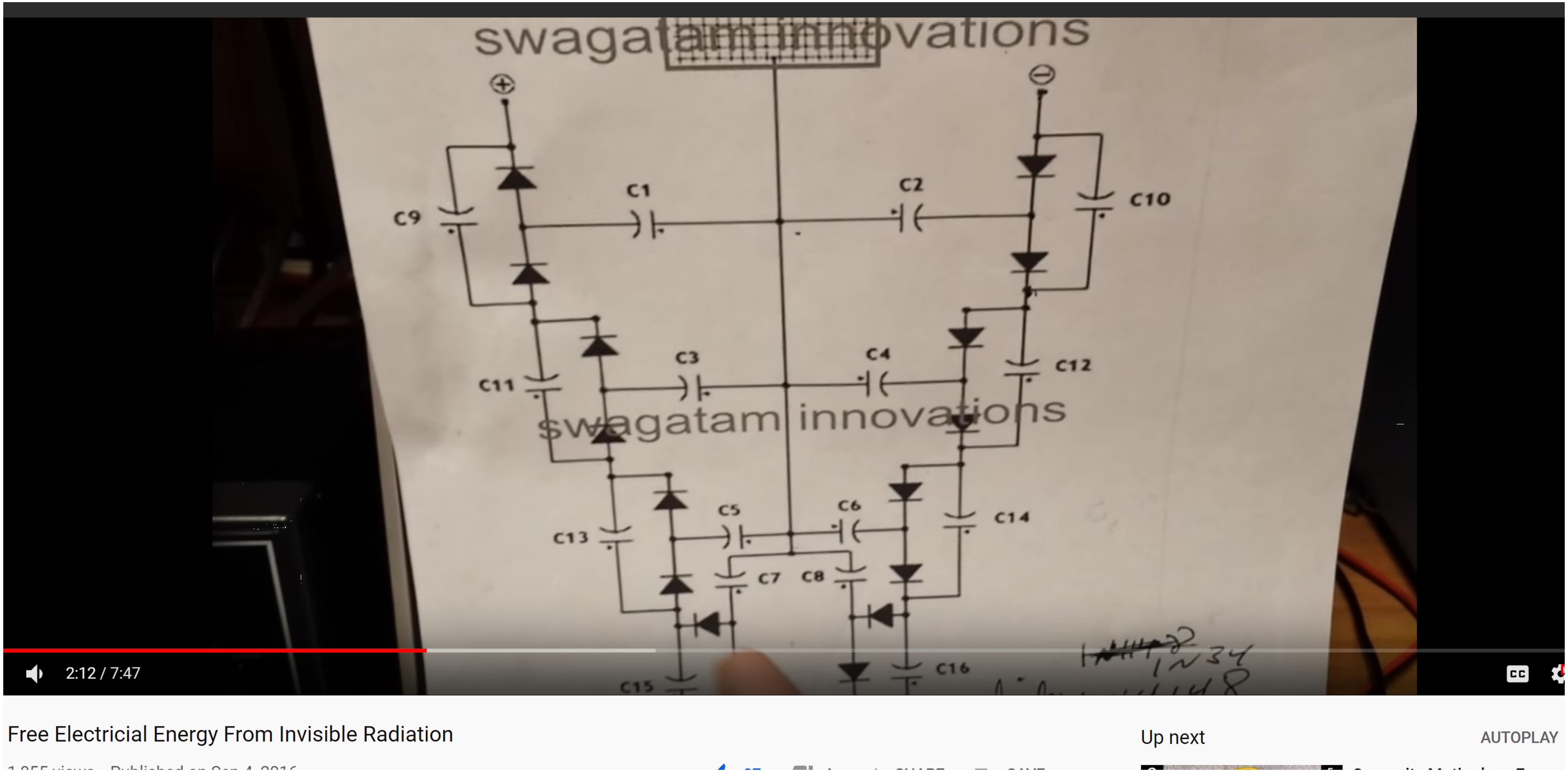
swagatam innovations
⊕
⊖
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
C13
C14
C15
C16
swagatam innovations
1N34
1N4148
2:12 / 7:47
Free Electricial Energy From Invisible Radiation
Up next
AUTOPLAY

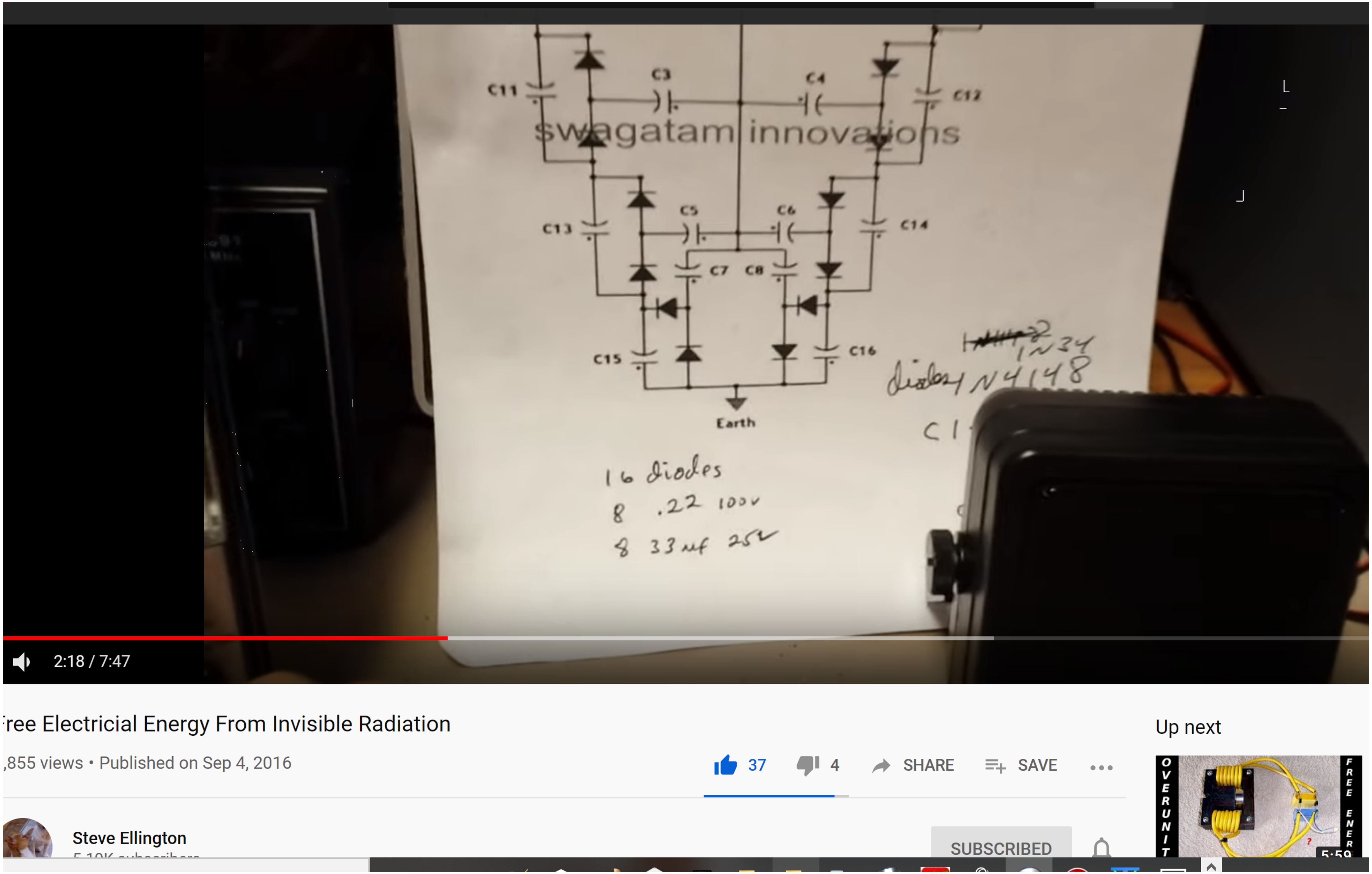
C11
C3
C4
C12
swagatam innovations
C5
C6
C13
C14
C7
C8
C15
C16
Earth
2:18 / 7:47
Up next
OVERUNIT
FREE ENER
5:59

~~1N4148~~ 1N34
diodes 1N4148

C1

16 diodes
8 .22 100v
8 33uf 25v

ree Electricial Energy From Invisible Radiation

,855 views • Published on Sep 4, 2016

37
4
SHARE
SAVE

Steve Ellington

SUBSCRIBED

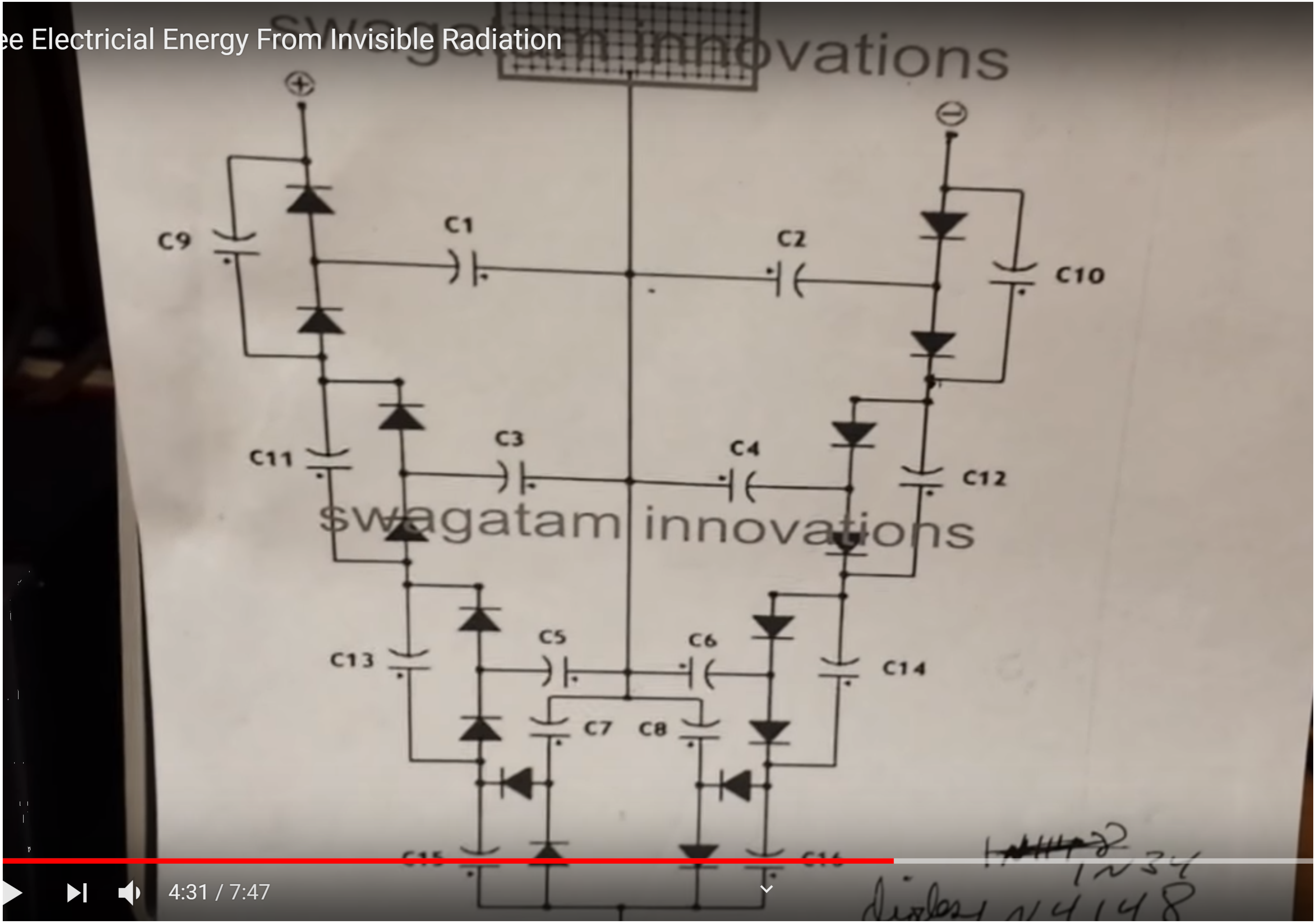

ee Electricial Energy From Invisible Radiation

swagatam innovations
⊕
⊖
C9
C1
C2
C10
C11
C3
C4
C12
swagatam innovations
C13
C5
C6
C14
C7
C8
C15
C16
1N34
diodes 1N4148
4:31 / 7:47

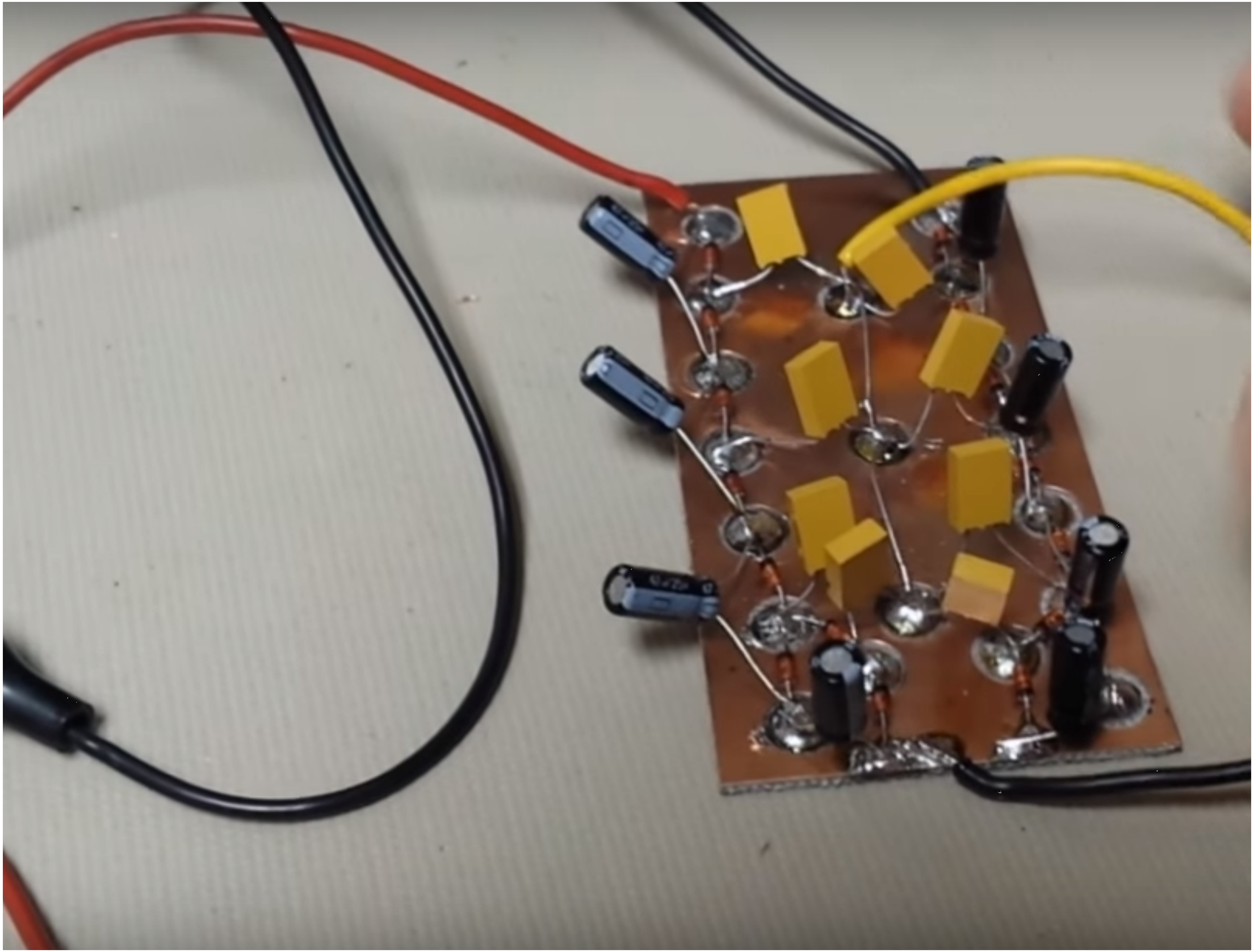

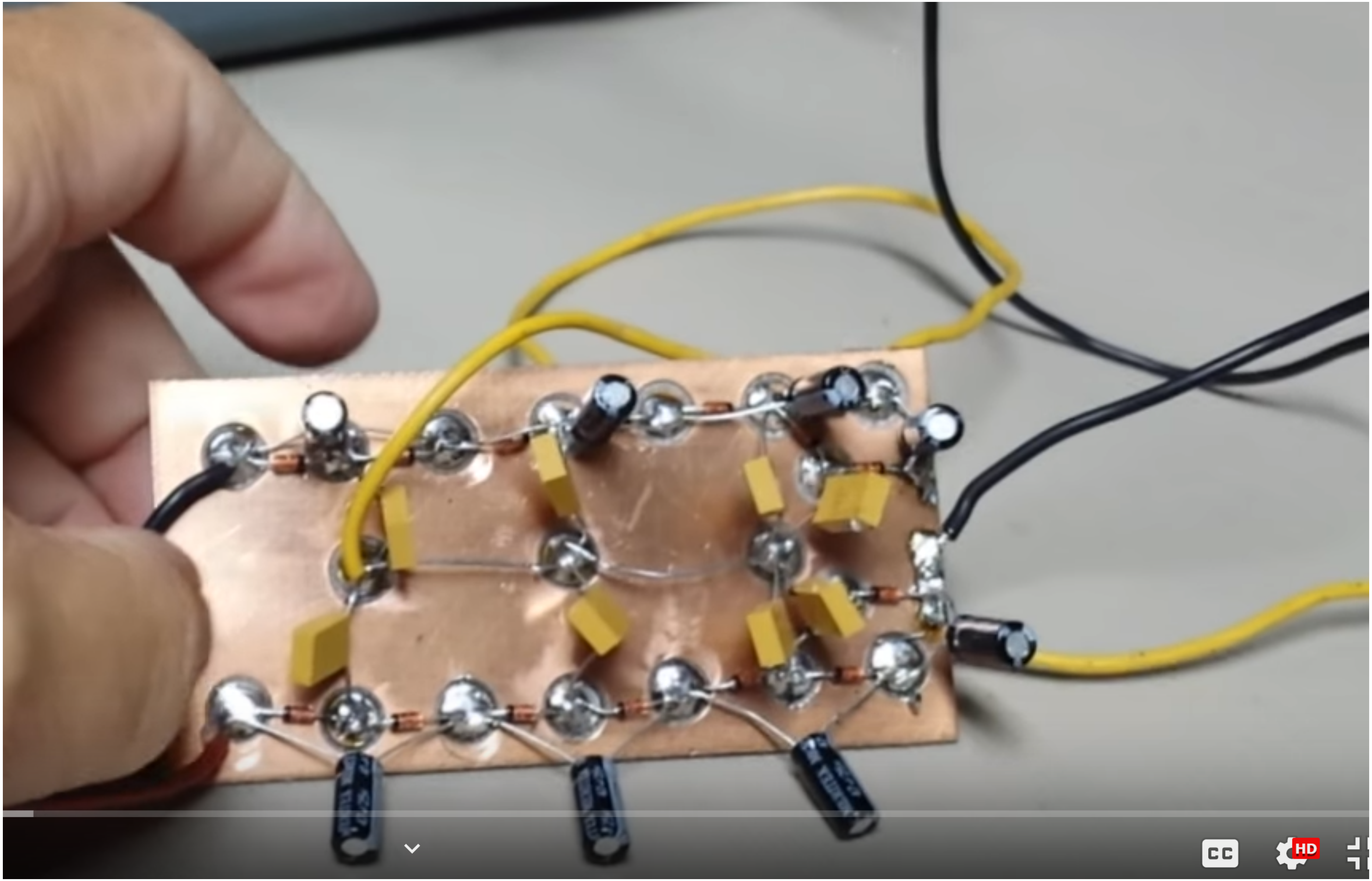
CC
HD

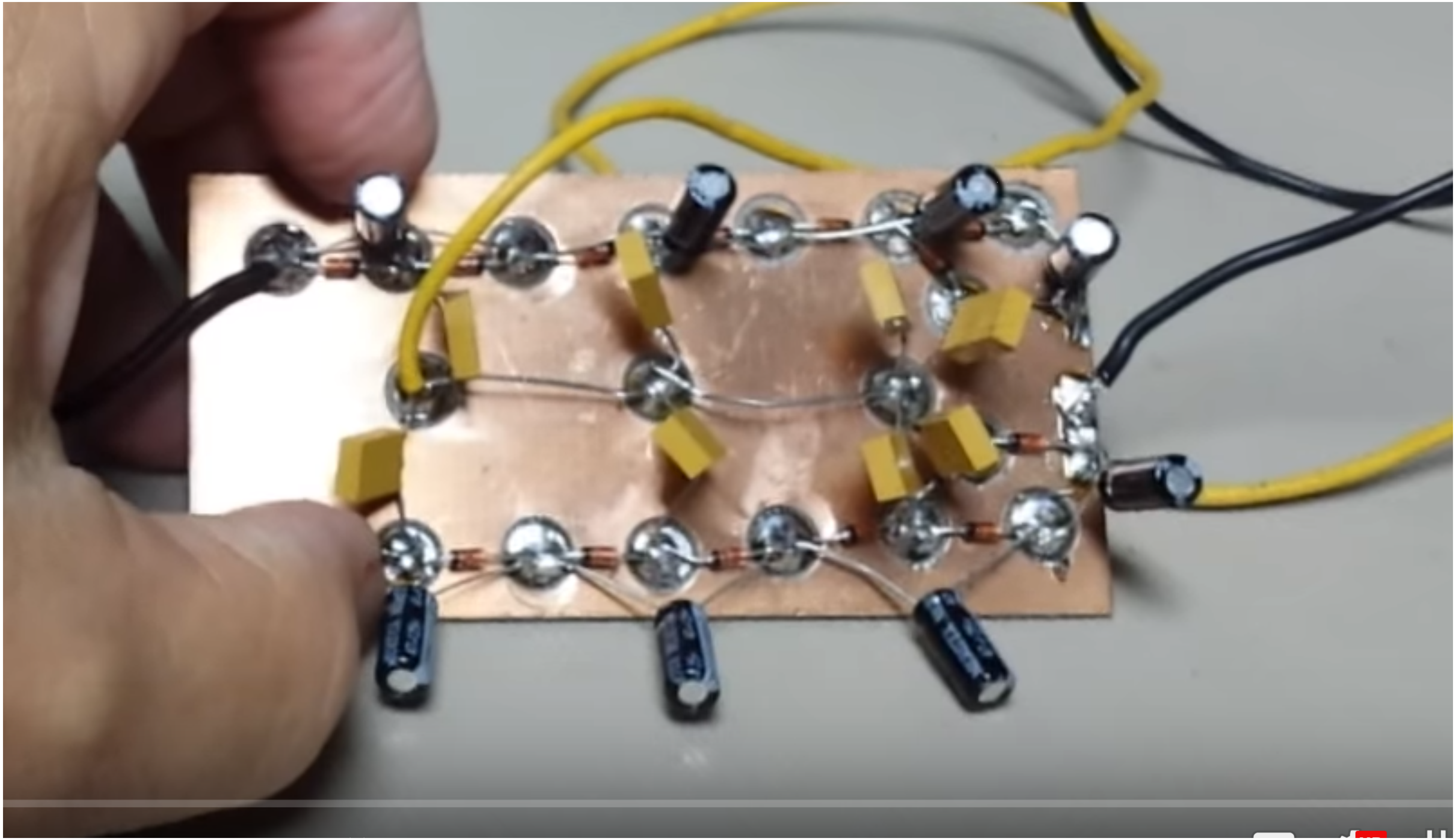

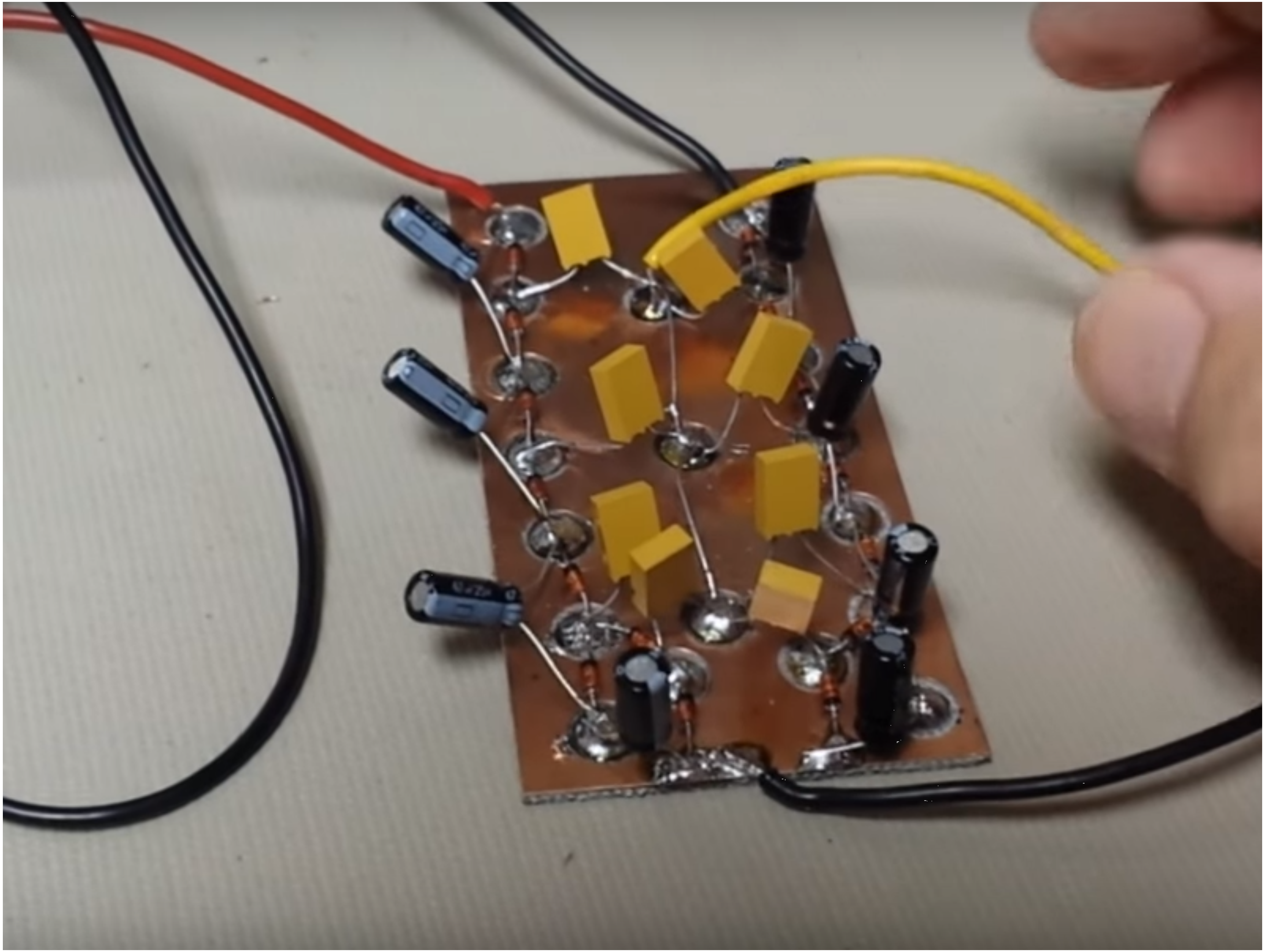

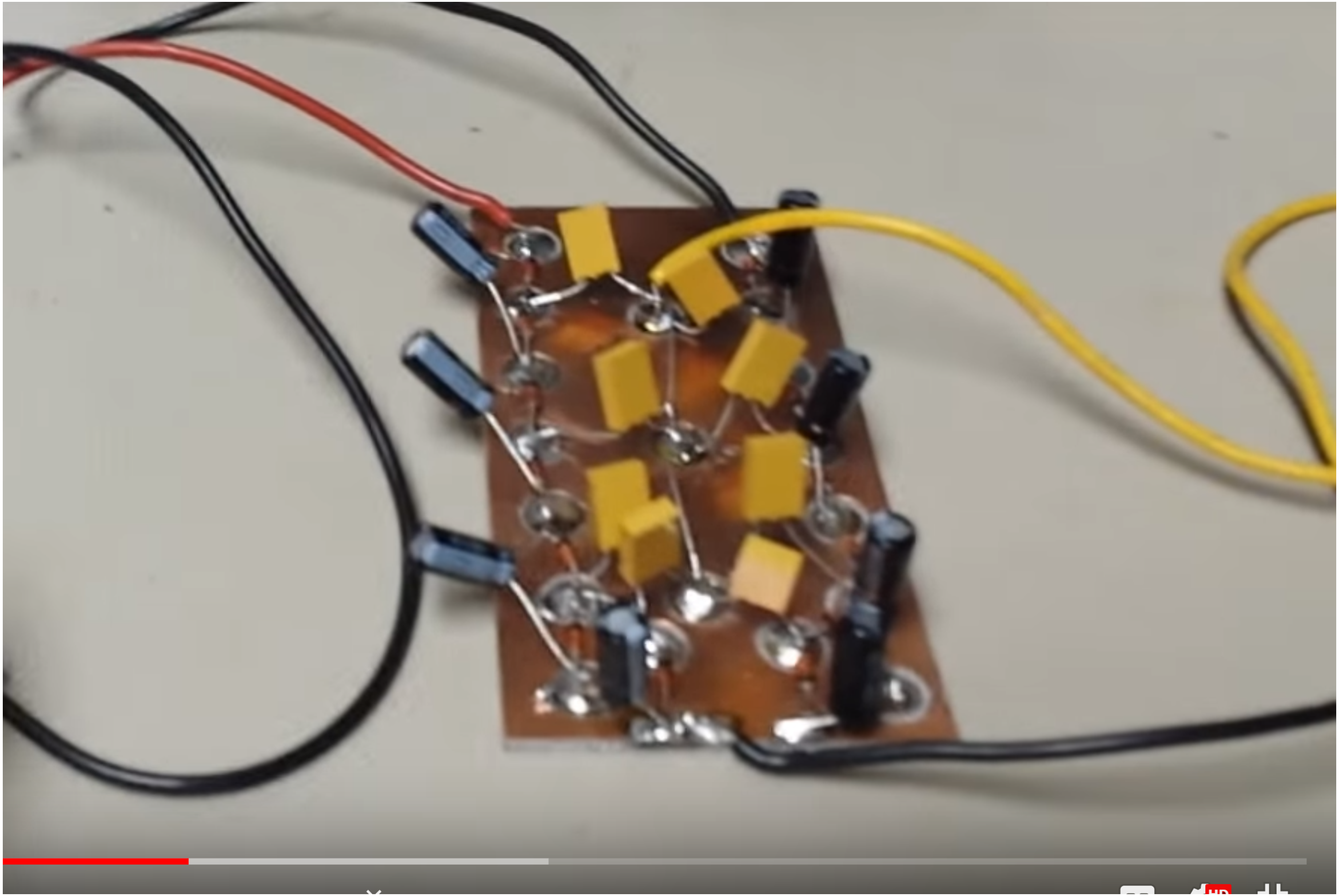

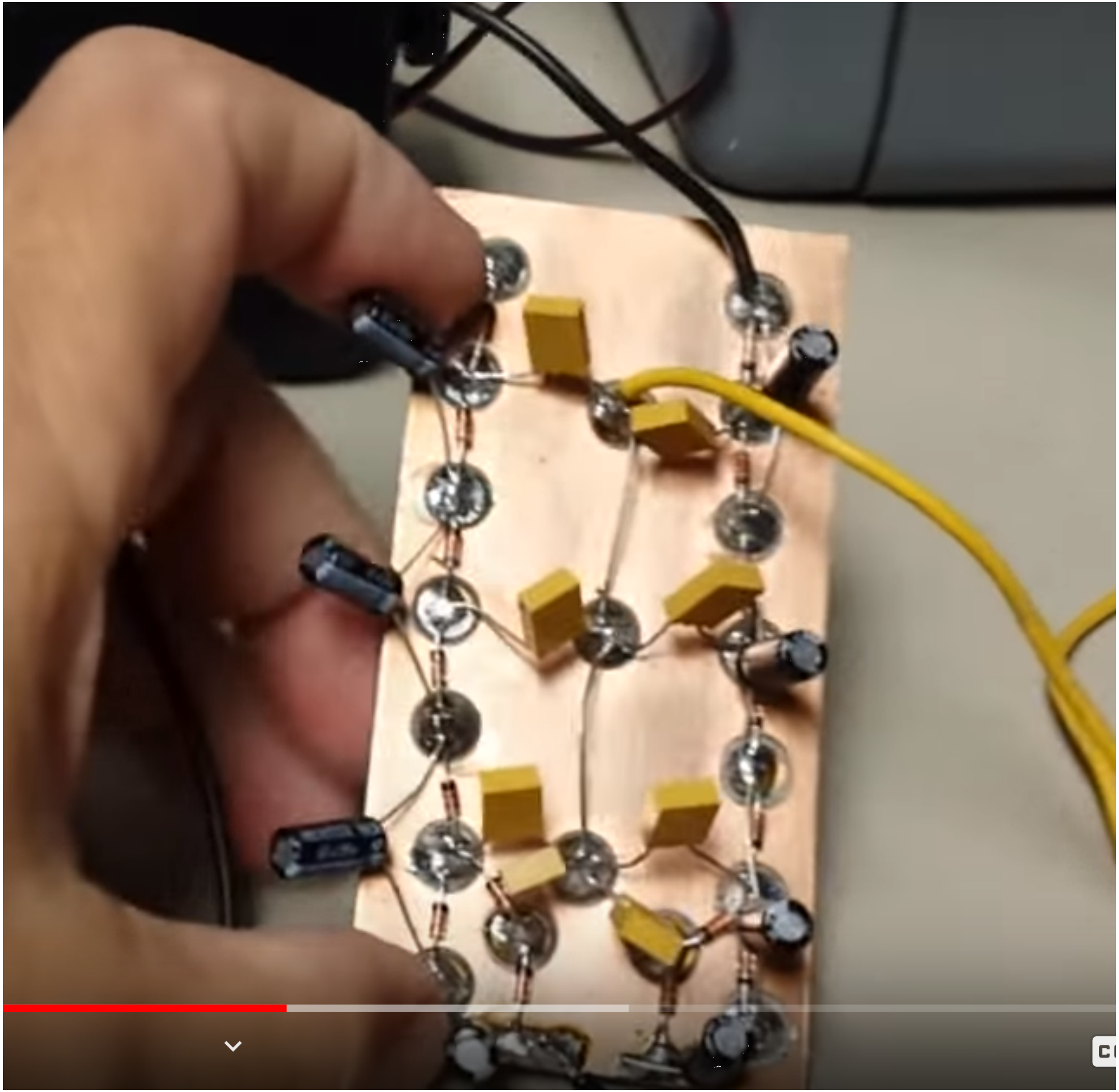

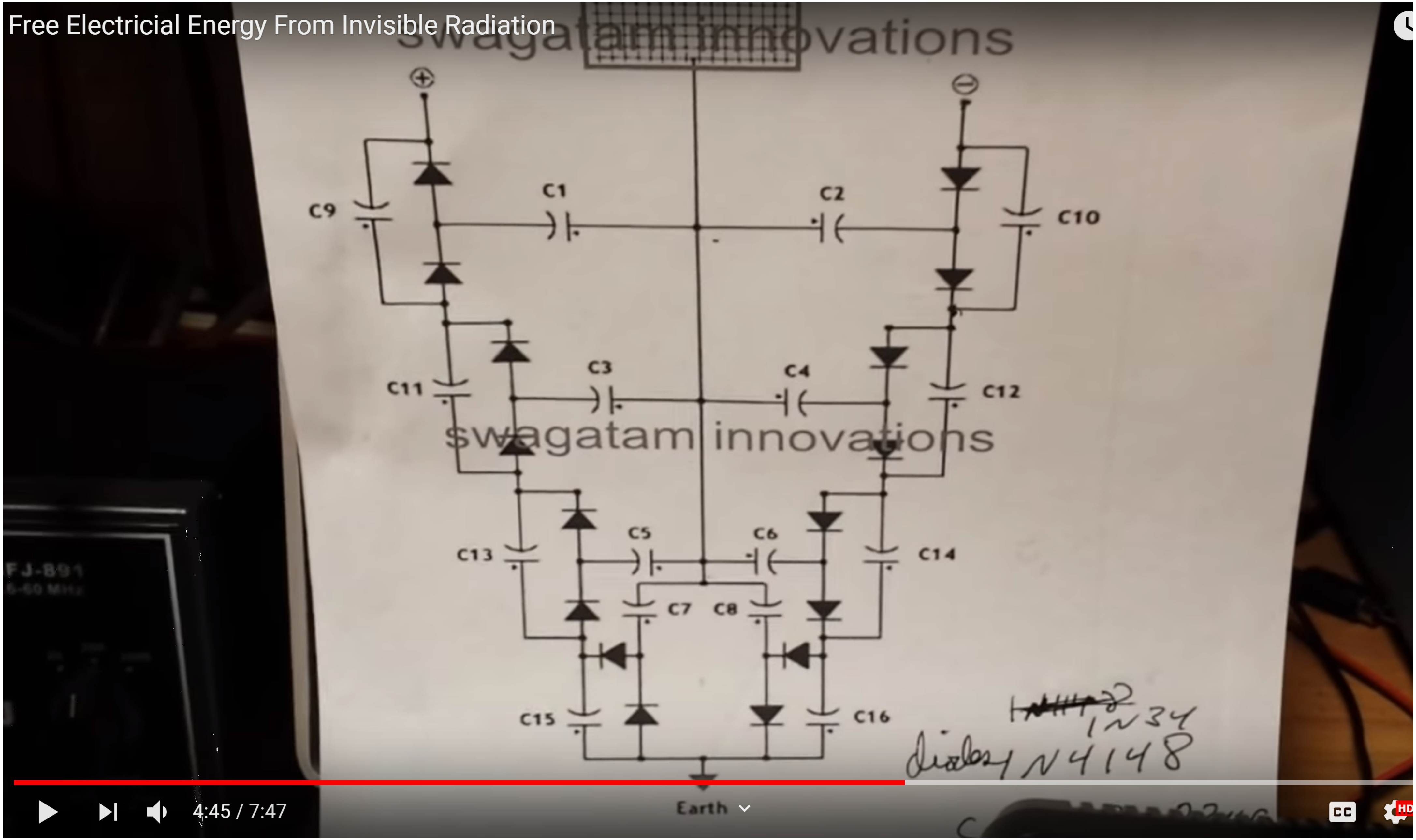

Free Electricial Energy From Invisible Radiation

swagatam innovations
⊕
⊖
C1
C2
C9
C10
C3
C4
C11
C12
swagatam innovations
C5
C6
C13
C14
C7
C8
C15
C16
1N34
diodes 1N4148
Earth
FJ-891
4:45 / 7:47
CC

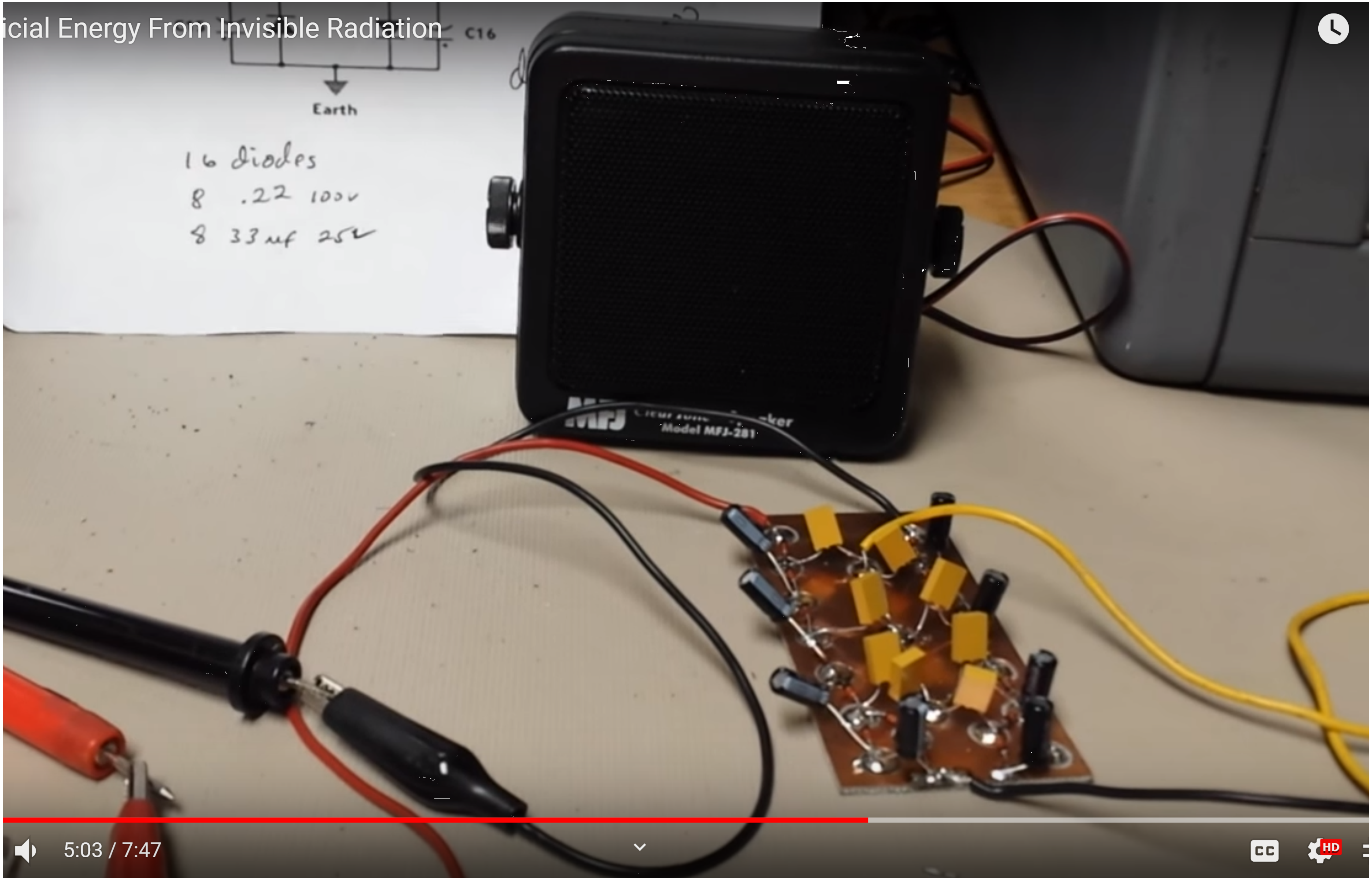

icial Energy From Invisible Radiation

16 diodes
8 .22 100v
8 33uf 25v

C16
Earth
MFJ
Model MFJ-281
5:03 / 7:47
CC
HD

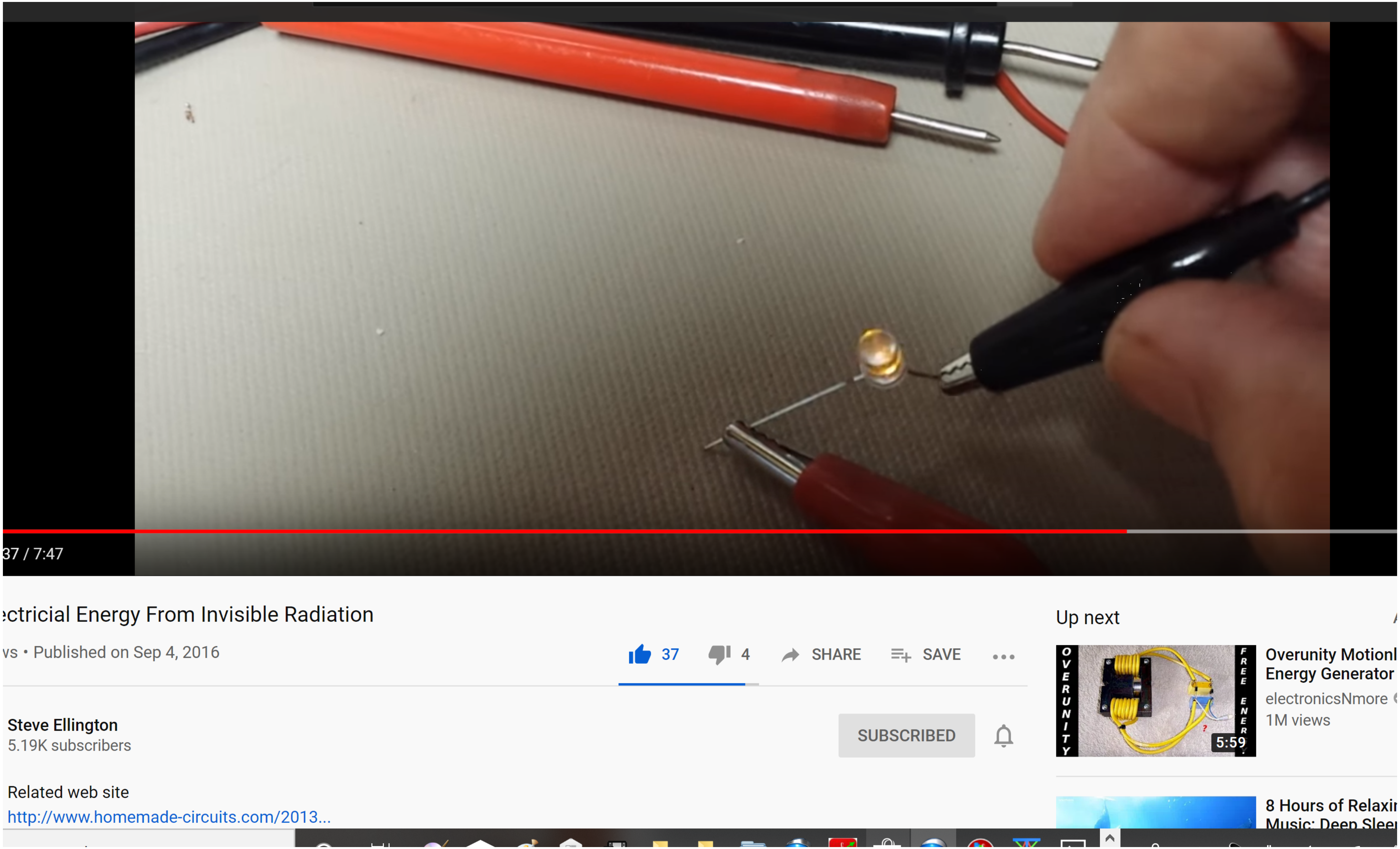
37 / 7:47
37
4
SHARE
SAVE
SUBSCRIBED
Up next
OVERUNITY
FREE ENERGY
5:59

ectricial Energy From Invisible Radiation

vs • Published on Sep 4, 2016

Steve Ellington
5.19K subscribers

Related web site
http://www.homemade-circuits.com/2013...

Overunity Motionl
Energy Generator
electronicsNmore
1M views

8 Hours of Relaxi
Music: Deep Slee

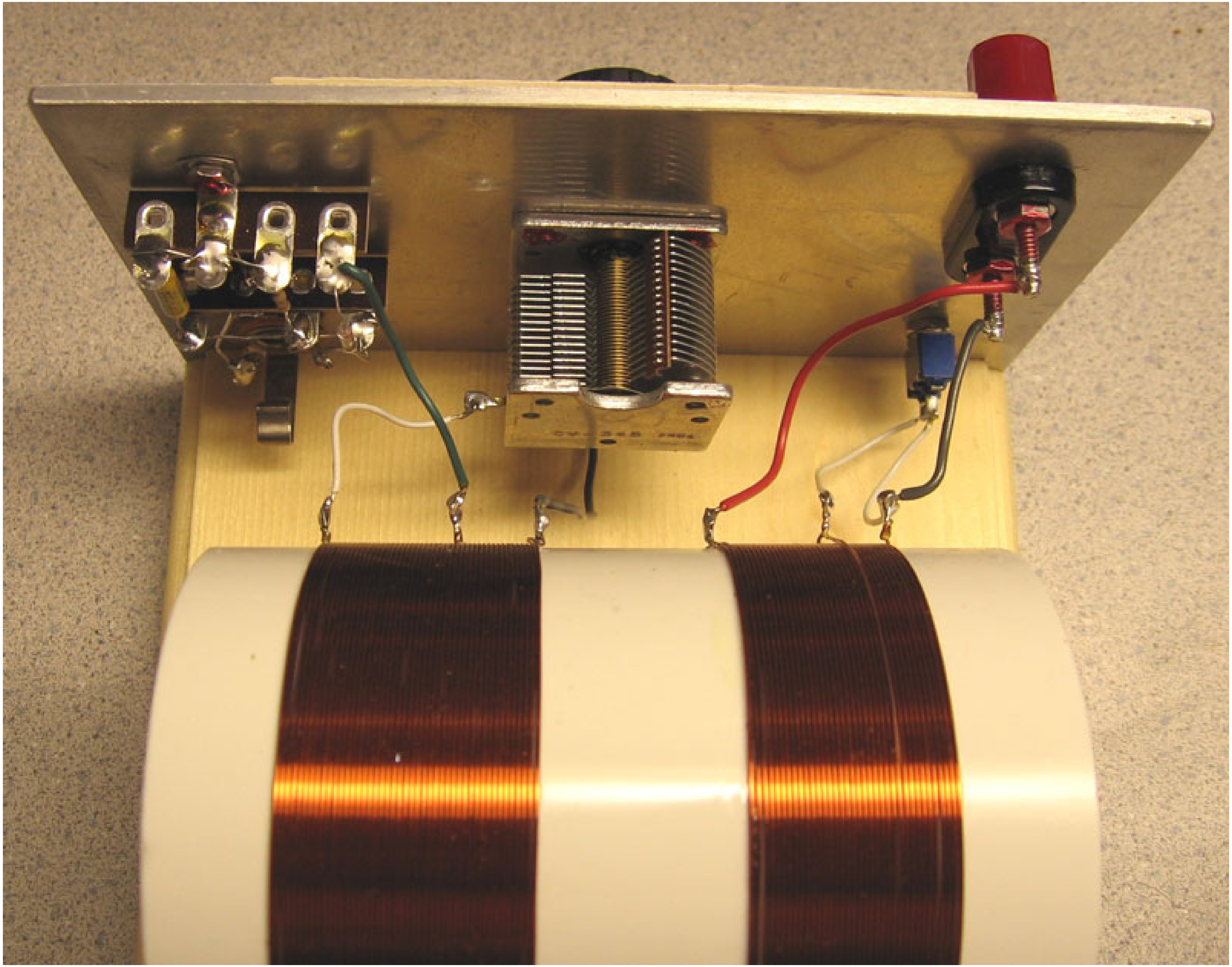

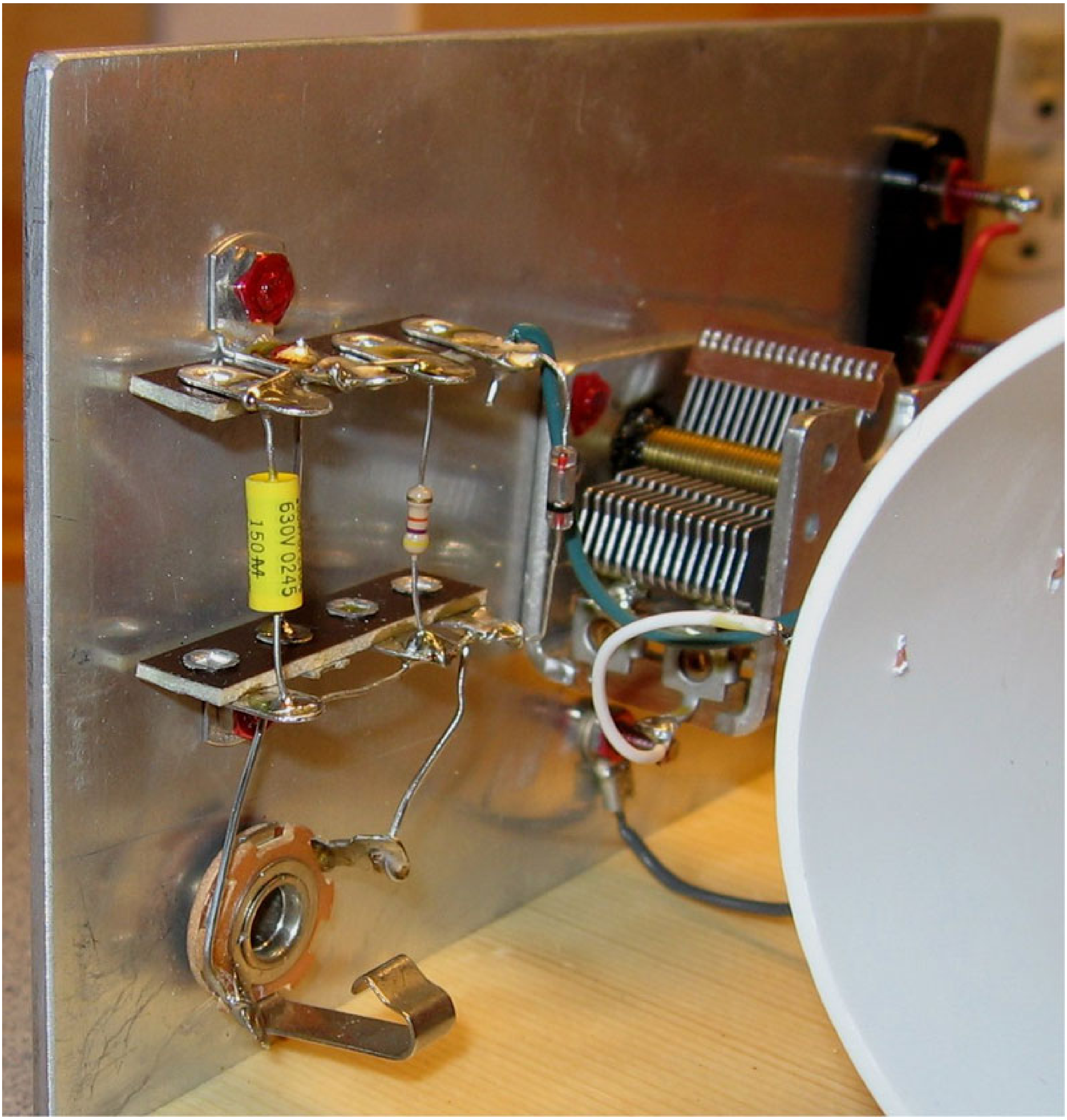
630V 0245
150n

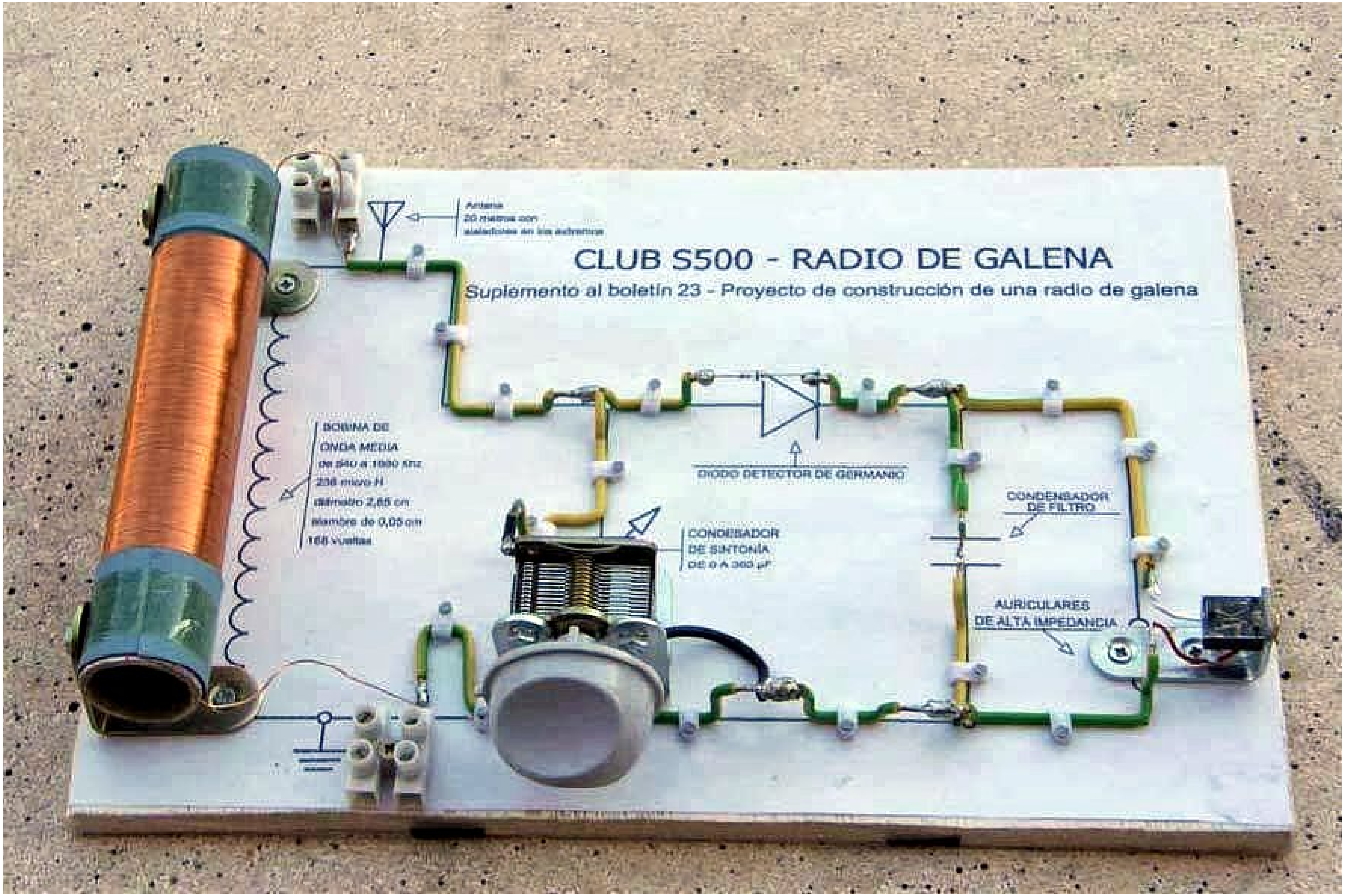
CLUB S500 - RADIO DE GALENA
Suplemento al boletín 23 - Proyecto de construcción de una radio de galena
Antena
20 metros con
aisladores en los extremos
BOBINA DE
ONDA MEDIA
de 540 a 1600 Khz
238 micro H
diámetro 2,65 cm
alambre de 0,05 cm
168 vueltas
DIODO DETECTOR DE GERMANIO
CONDENSADOR
DE SINTONÍA
DE 0 A 360 µF
CONDENSADOR
DE FILTRO
AURICULARES
DE ALTA IMPEDANCIA

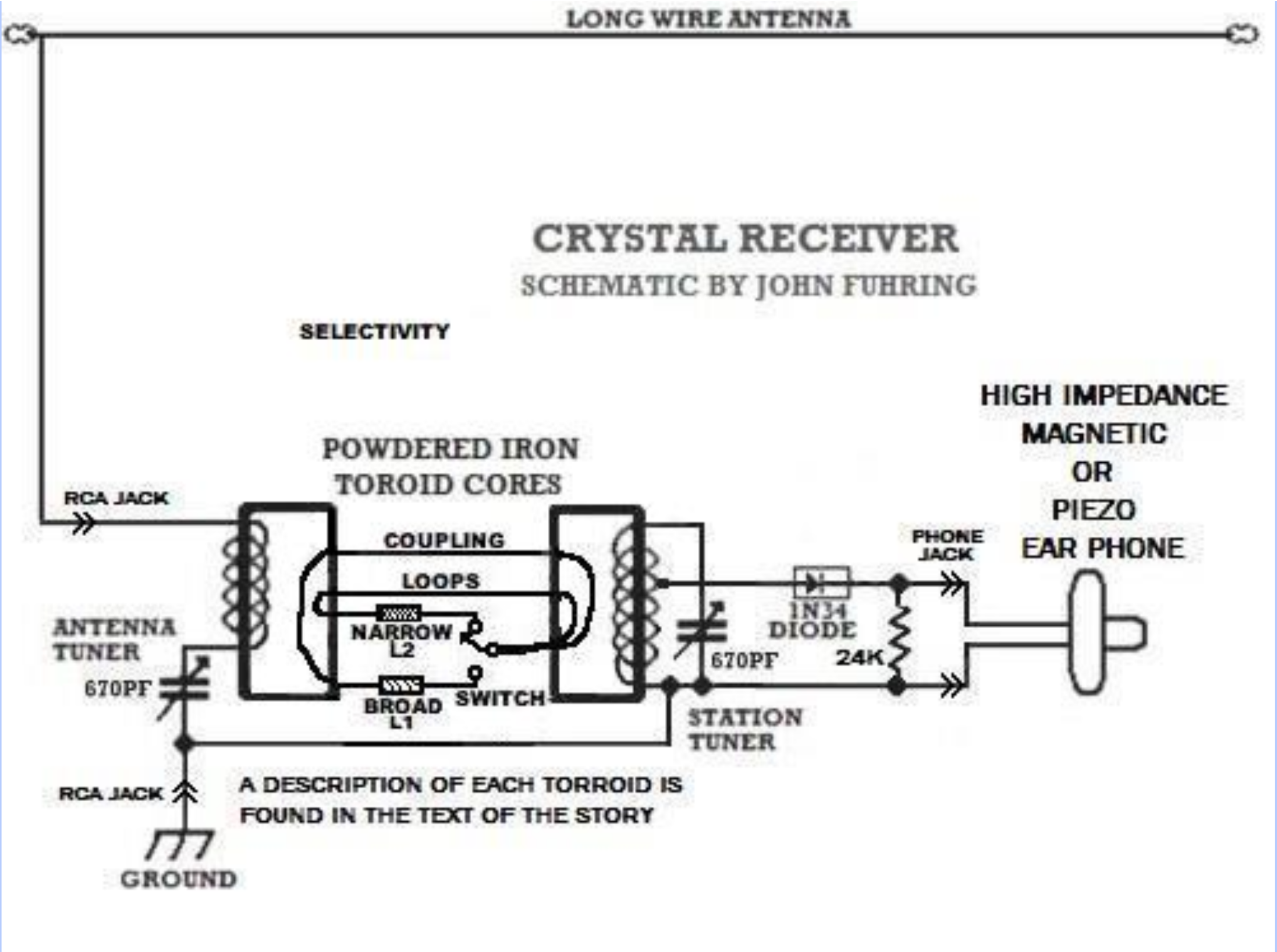
LONG WIRE ANTENNA
CRYSTAL RECEIVER
SCHEMATIC BY JOHN FUHRING
SELECTIVITY
HIGH IMPEDANCE MAGNETIC OR PIEZO EAR PHONE
POWDERED IRON TOROID CORES
RCA JACK
COUPLING
LOOPS
PHONE JACK
NARROW L2
ANTENNA TUNER
670PF
1N34 DIODE
24K
670PF
BROAD L1
SWITCH
STATION TUNER
RCA JACK
A DESCRIPTION OF EACH TORROID IS FOUND IN THE TEXT OF THE STORY
GROUND

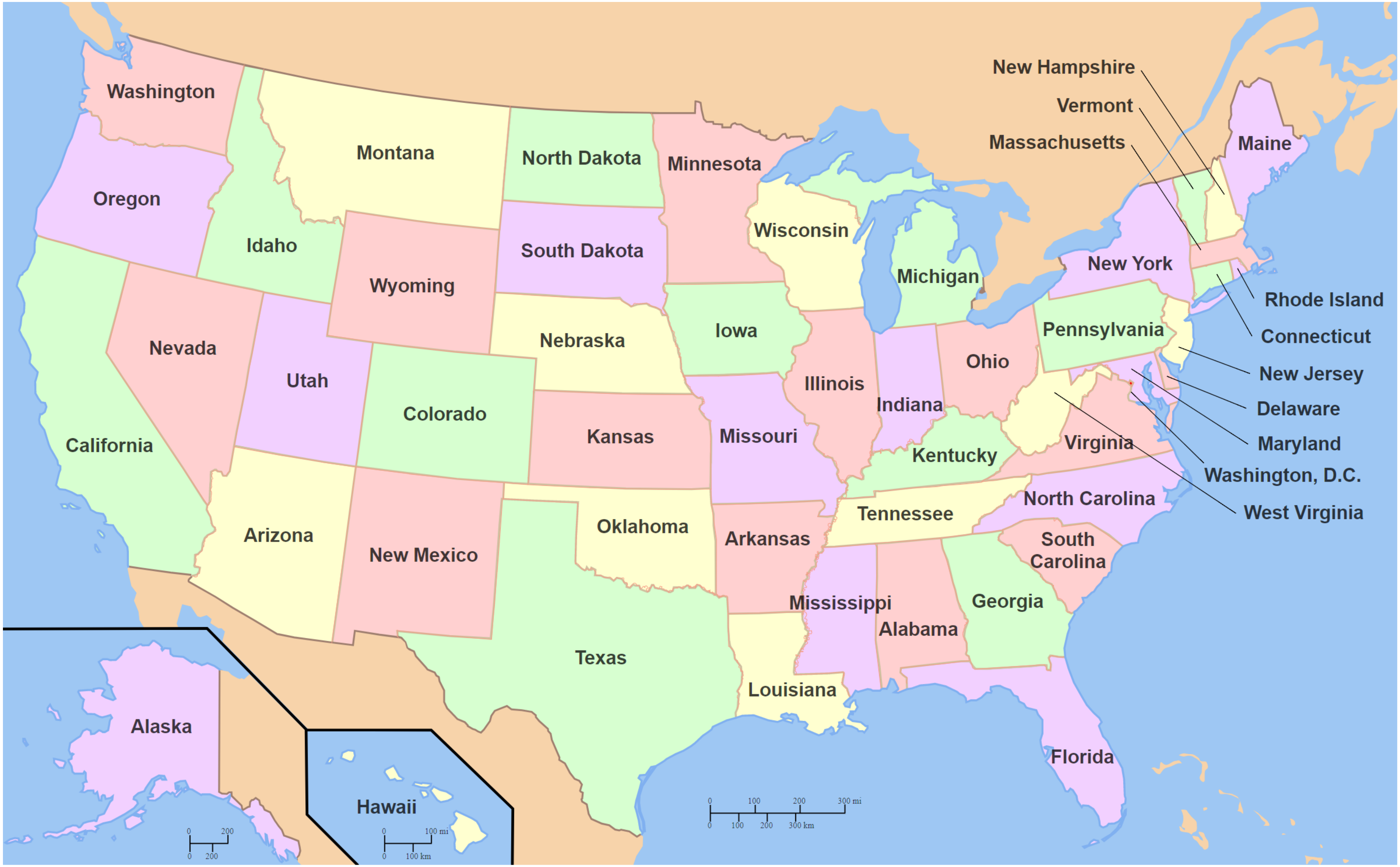
New Hampshire
Vermont
Massachusetts
Maine
Washington
Montana
North Dakota
Minnesota
Oregon
Idaho
Wisconsin
South Dakota
Wyoming
Michigan
New York
Rhode Island
Iowa
Nebraska
Pennsylvania
Connecticut
Nevada
Utah
Illinois
Ohio
New Jersey
Colorado
Indiana
Delaware
California
Kansas
Missouri
Virginia
Maryland
Kentucky
Washington, D.C.
North Carolina
Oklahoma
Tennessee
West Virginia
Arizona
Arkansas
South Carolina
New Mexico
Mississippi
Georgia
Alabama
Texas
Louisiana
Alaska
Florida
Hawaii
0 100 200 300 mi
0 100 200 300 km
0 200
0 200
0 100 mi
0 100 km

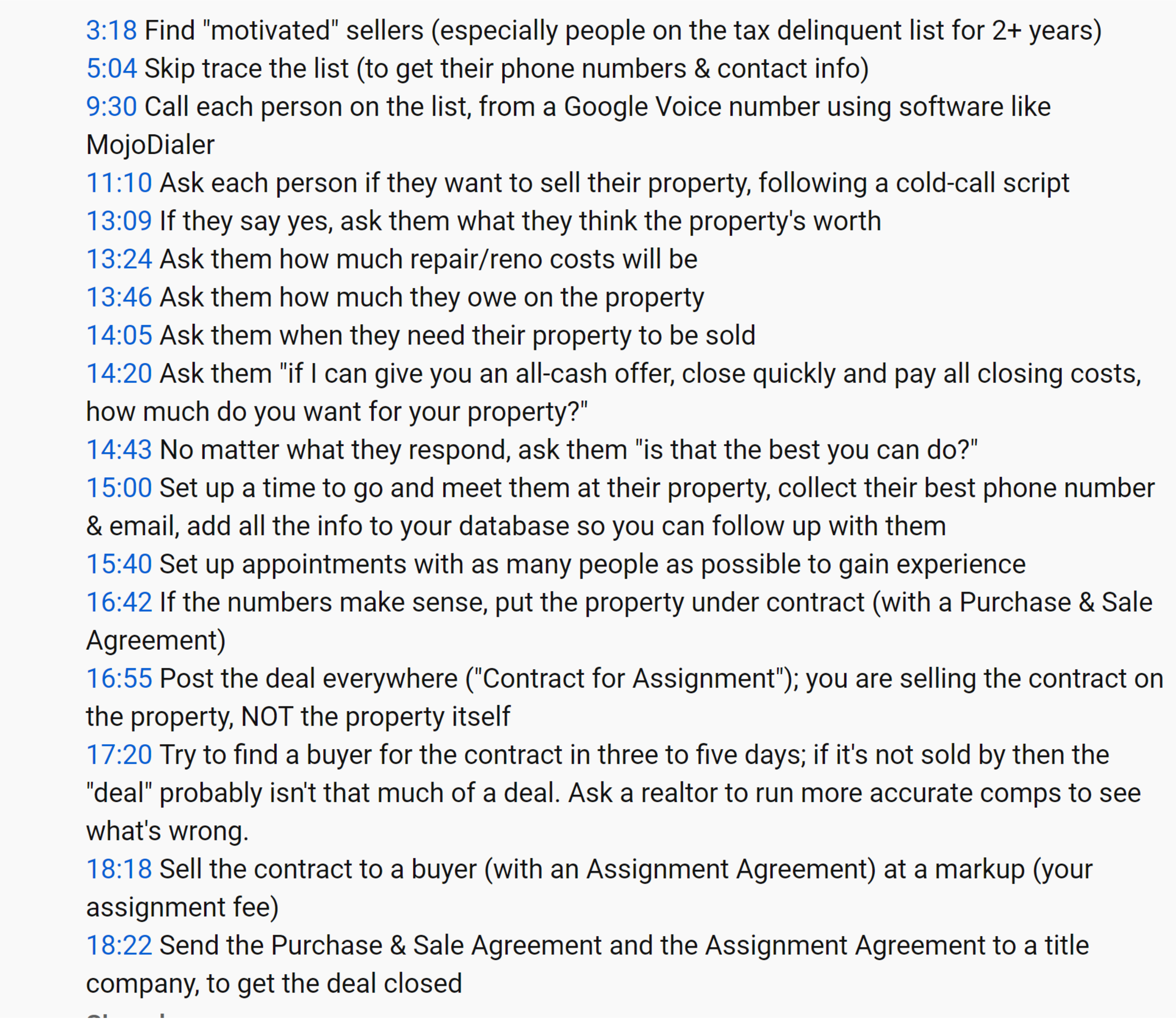

3:18 Find "motivated" sellers (especially people on the tax delinquent list for 2+ years)
5:04 Skip trace the list (to get their phone numbers & contact info)
9:30 Call each person on the list, from a Google Voice number using software like MojoDialer
11:10 Ask each person if they want to sell their property, following a cold-call script
13:09 If they say yes, ask them what they think the property's worth
13:24 Ask them how much repair/reno costs will be
13:46 Ask them how much they owe on the property
14:05 Ask them when they need their property to be sold
14:20 Ask them "if I can give you an all-cash offer, close quickly and pay all closing costs, how much do you want for your property?"
14:43 No matter what they respond, ask them "is that the best you can do?"
15:00 Set up a time to go and meet them at their property, collect their best phone number & email, add all the info to your database so you can follow up with them
15:40 Set up appointments with as many people as possible to gain experience
16:42 If the numbers make sense, put the property under contract (with a Purchase & Sale Agreement)
16:55 Post the deal everywhere ("Contract for Assignment"); you are selling the contract on the property, NOT the property itself
17:20 Try to find a buyer for the contract in three to five days; if it's not sold by then the "deal" probably isn't that much of a deal. Ask a realtor to run more accurate comps to see what's wrong.
18:18 Sell the contract to a buyer (with an Assignment Agreement) at a markup (your assignment fee)
18:22 Send the Purchase & Sale Agreement and the Assignment Agreement to a title company, to get the deal closed

DETERMINE YOUR MAXIMUM ALLOWABLE OFFER

Once you've determined the Rehabber's Maximum Purchase Price you will deduct your wholesaler assignment fee/profit to determine the Maximum Allowable Offer you can offer for the property.

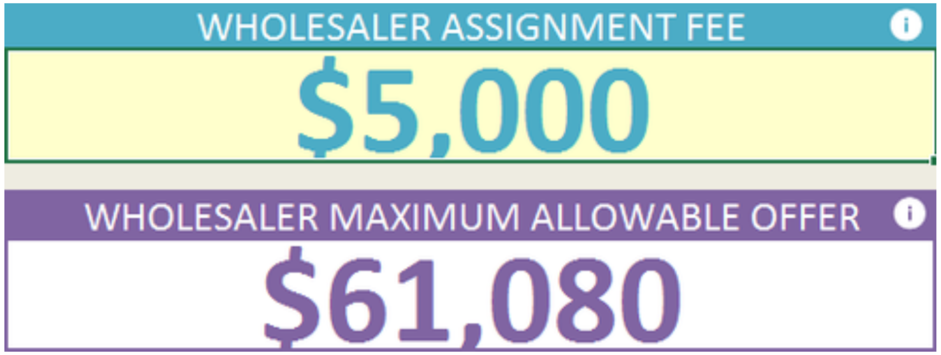

WHOLESALER ASSIGNMENT FEE

$5,000

WHOLESALER MAXIMUM ALLOWABLE OFFER

$61,080

Input Your Wholesaler Assignment Fee to Determine the Maximum Allowable Offer

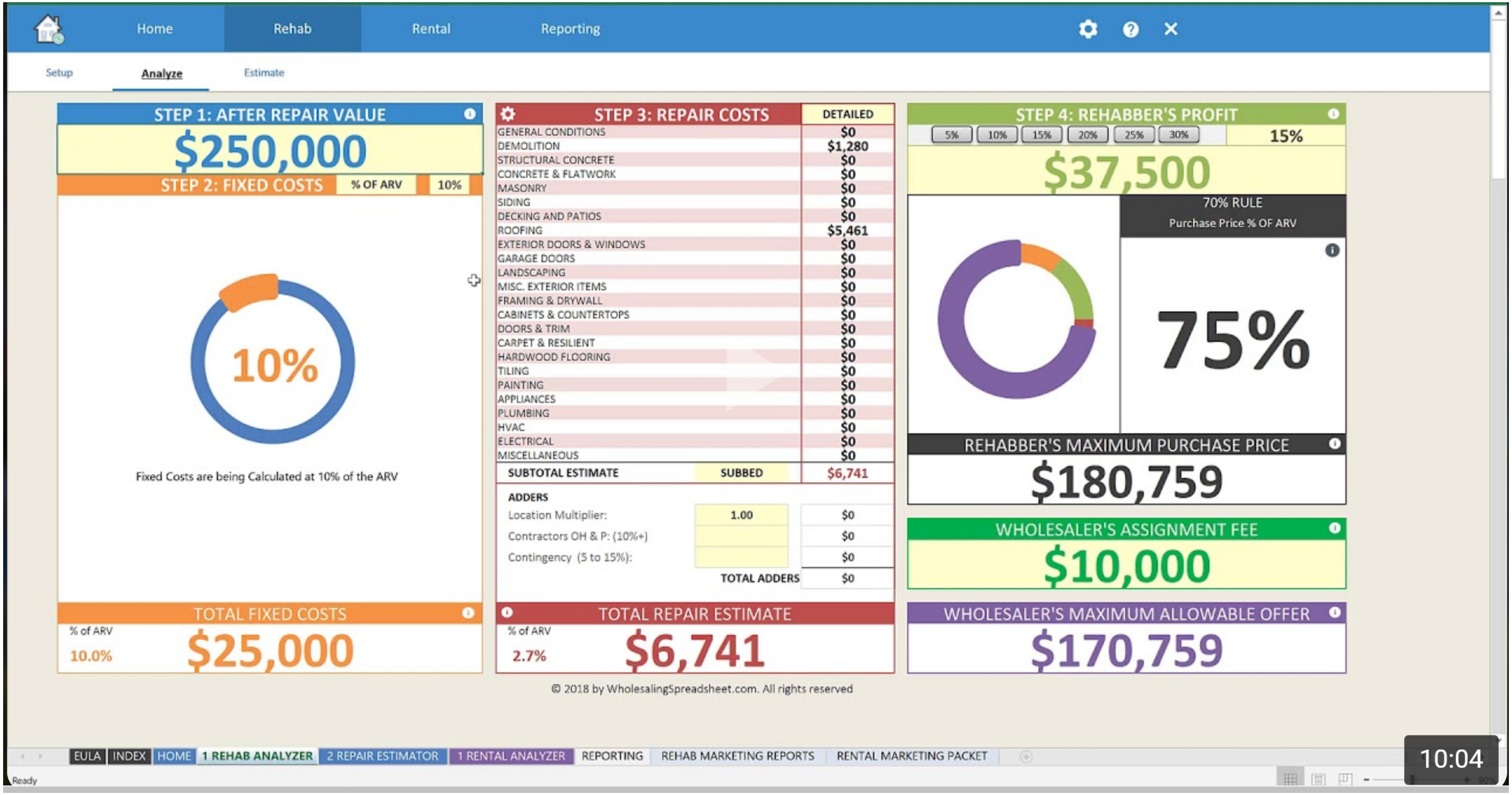
Home
Rehab
Rental
Reporting
Setup
Analyze
Estimate
EULA
INDEX
HOME
1 REHAB ANALYZER
2 REPAIR ESTIMATOR
1 RENTAL ANALYZER
REPORTING
REHAB MARKETING REPORTS
RENTAL MARKETING PACKET
Ready
10:04

STEP 1: AFTER REPAIR VALUE

$250,000

STEP 2: FIXED COSTS % OF ARV 10%

10%

Fixed Costs are being Calculated at 10% of the ARV

TOTAL FIXED COSTS

% of ARV 10.0% **$25,000**

STEP 3: REPAIR COSTS	DETAILED
GENERAL CONDITIONS	$0
DEMOLITION	$1,280
STRUCTURAL CONCRETE	$0
CONCRETE & FLATWORK	$0
MASONRY	$0
SIDING	$0
DECKING AND PATIOS	$0
ROOFING	$5,461
EXTERIOR DOORS & WINDOWS	$0
GARAGE DOORS	$0
LANDSCAPING	$0
MISC. EXTERIOR ITEMS	$0
FRAMING & DRYWALL	$0
CABINETS & COUNTERTOPS	$0
DOORS & TRIM	$0
CARPET & RESILIENT	$0
HARDWOOD FLOORING	$0
TILING	$0
PAINTING	$0
APPLIANCES	$0
PLUMBING	$0
HVAC	$0
ELECTRICAL	$0
MISCELLANEOUS	$0
SUBTOTAL ESTIMATE SUBBED	$6,741

ADDERS		
Location Multiplier:	1.00	$0
Contractors OH & P: (10%+)		$0
Contingency (5 to 15%):		$0
	TOTAL ADDERS	$0

TOTAL REPAIR ESTIMATE

% of ARV 2.7% **$6,741**

STEP 4: REHABBER'S PROFIT

☐ 5% ☐ 10% ☐ 15% ☐ 20% ☐ 25% ☐ 30% 15%

$37,500

70% RULE

Purchase Price % OF ARV

75%

REHABBER'S MAXIMUM PURCHASE PRICE

$180,759

WHOLESALER'S ASSIGNMENT FEE

$10,000

WHOLESALER'S MAXIMUM ALLOWABLE OFFER

$170,759

© 2018 by WholesalingSpreadsheet.com. All rights reserved

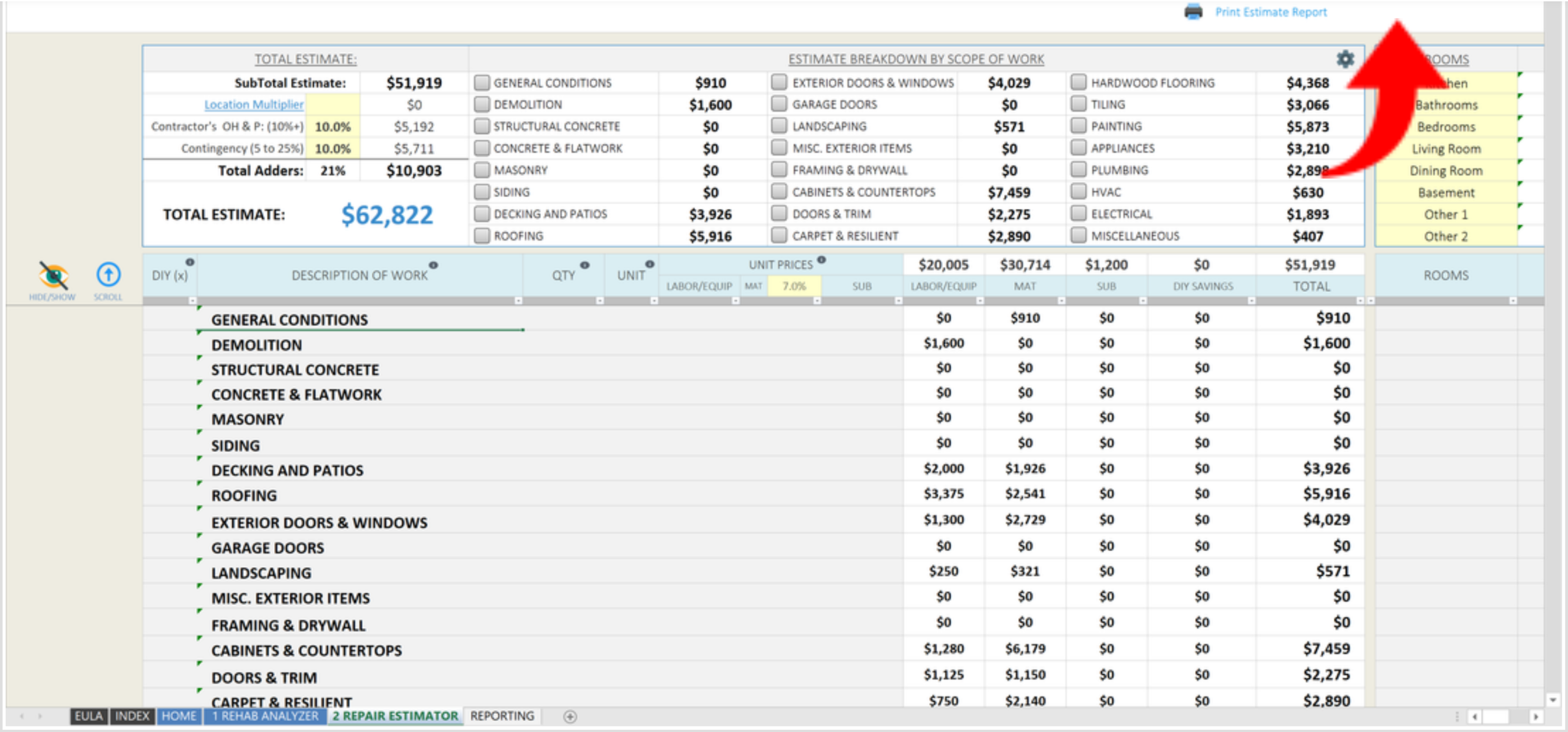

Print Estimate Report

TOTAL ESTIMATE:		
SubTotal Estimate:		$51,919
Location Multiplier		$0
Contractor's OH & P: (10%+)	10.0%	$5,192
Contingency (5 to 25%)	10.0%	$5,711
Total Adders:	21%	$10,903
TOTAL ESTIMATE:		$62,822

ESTIMATE BREAKDOWN BY SCOPE OF WORK

☐ GENERAL CONDITIONS	$910	☐ EXTERIOR DOORS & WINDOWS	$4,029	☐ HARDWOOD FLOORING	$4,368
☐ DEMOLITION	$1,600	☐ GARAGE DOORS	$0	☐ TILING	$3,066
☐ STRUCTURAL CONCRETE	$0	☐ LANDSCAPING	$571	☐ PAINTING	$5,873
☐ CONCRETE & FLATWORK	$0	☐ MISC. EXTERIOR ITEMS	$0	☐ APPLIANCES	$3,210
☐ MASONRY	$0	☐ FRAMING & DRYWALL	$0	☐ PLUMBING	$2,89
☐ SIDING	$0	☐ CABINETS & COUNTERTOPS	$7,459	☐ HVAC	$630
☐ DECKING AND PATIOS	$3,926	☐ DOORS & TRIM	$2,275	☐ ELECTRICAL	$1,893
☐ ROOFING	$5,916	☐ CARPET & RESILIENT	$2,890	☐ MISCELLANEOUS	$407

OOMS
hen
Bathrooms
Bedrooms
Living Room
Dining Room
Basement
Other 1
Other 2

DIY (x)	DESCRIPTION OF WORK	QTY	UNIT	UNIT PRICES				$20,005	$30,714	$1,200	$0	$51,919	ROOMS
				LABOR/EQUIP	MAT	7.0%	SUB	LABOR/EQUIP	MAT	SUB	DIY SAVINGS	TOTAL	
	GENERAL CONDITIONS							$0	$910	$0	$0	$910	
	DEMOLITION							$1,600	$0	$0	$0	$1,600	
	STRUCTURAL CONCRETE							$0	$0	$0	$0	$0	
	CONCRETE & FLATWORK							$0	$0	$0	$0	$0	
	MASONRY							$0	$0	$0	$0	$0	
	SIDING							$0	$0	$0	$0	$0	
	DECKING AND PATIOS							$2,000	$1,926	$0	$0	$3,926	
	ROOFING							$3,375	$2,541	$0	$0	$5,916	
	EXTERIOR DOORS & WINDOWS							$1,300	$2,729	$0	$0	$4,029	
	GARAGE DOORS							$0	$0	$0	$0	$0	
	LANDSCAPING							$250	$321	$0	$0	$571	
	MISC. EXTERIOR ITEMS							$0	$0	$0	$0	$0	
	FRAMING & DRYWALL							$0	$0	$0	$0	$0	
	CABINETS & COUNTERTOPS							$1,280	$6,179	$0	$0	$7,459	
	DOORS & TRIM							$1,125	$1,150	$0	$0	$2,275	
	CARPET & RESILIENT							$750	$2,140	$0	$0	$2,890	

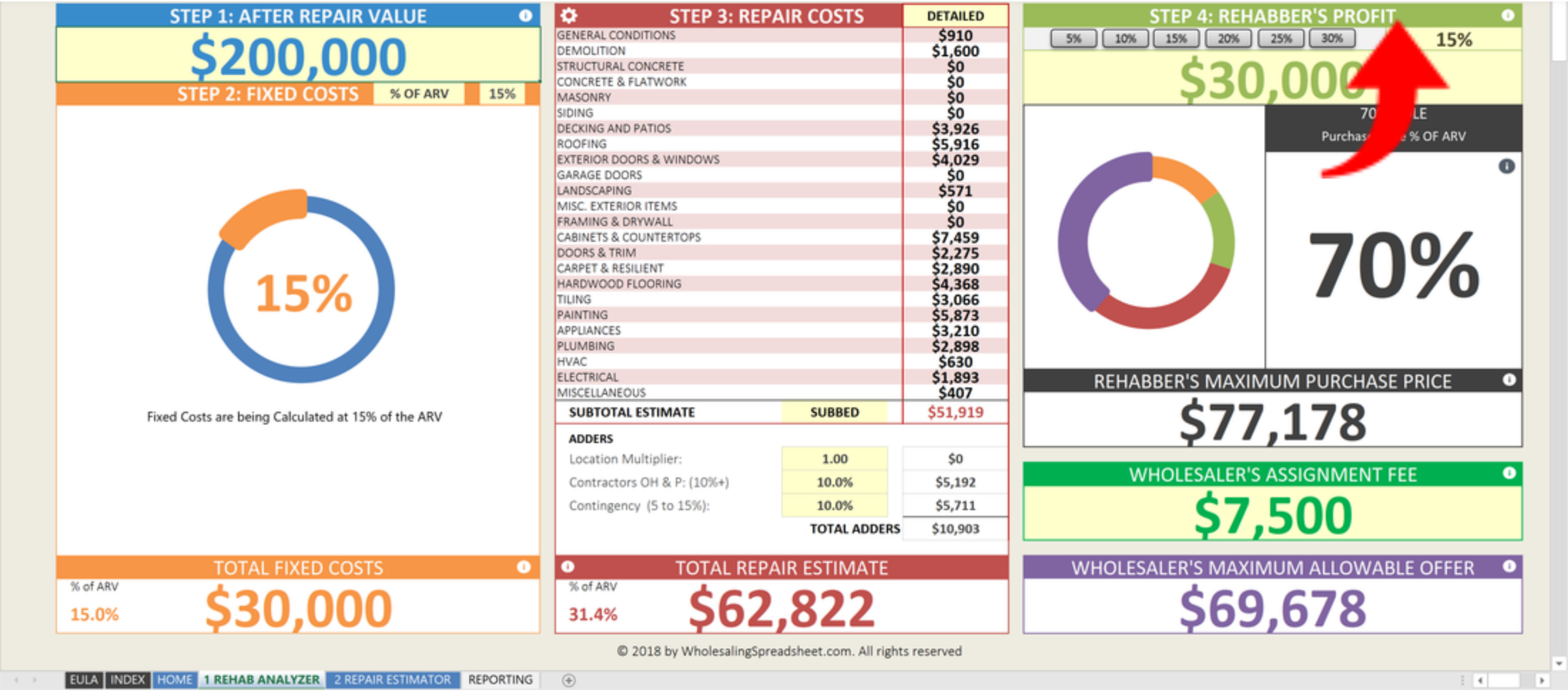

STEP 1: AFTER REPAIR VALUE

$200,000

STEP 2: FIXED COSTS % OF ARV 15%

15%

Fixed Costs are being Calculated at 15% of the ARV

TOTAL FIXED COSTS

% of ARV 15.0% $30,000

STEP 3: REPAIR COSTS

	DETAILED
GENERAL CONDITIONS	$910
DEMOLITION	$1,600
STRUCTURAL CONCRETE	$0
CONCRETE & FLATWORK	$0
MASONRY	$0
SIDING	$0
DECKING AND PATIOS	$3,926
ROOFING	$5,916
EXTERIOR DOORS & WINDOWS	$4,029
GARAGE DOORS	$0
LANDSCAPING	$571
MISC. EXTERIOR ITEMS	$0
FRAMING & DRYWALL	$0
CABINETS & COUNTERTOPS	$7,459
DOORS & TRIM	$2,275
CARPET & RESILIENT	$2,890
HARDWOOD FLOORING	$4,368
TILING	$3,066
PAINTING	$5,873
APPLIANCES	$3,210
PLUMBING	$2,898
HVAC	$630
ELECTRICAL	$1,893
MISCELLANEOUS	$407
SUBTOTAL ESTIMATE SUBBED	$51,919

ADDERS

Location Multiplier:	1.00	$0
Contractors OH & P: (10%+)	10.0%	$5,192
Contingency (5 to 15%):	10.0%	$5,711
	TOTAL ADDERS	$10,903

TOTAL REPAIR ESTIMATE

% of ARV 31.4% $62,822

STEP 4: REHABBER'S PROFIT

5% 10% 15% 20% 25% 30% 15%

$30,000

70 LE

Purchas % OF ARV

70%

REHABBER'S MAXIMUM PURCHASE PRICE

$77,178

WHOLESALER'S ASSIGNMENT FEE

$7,500

WHOLESALER'S MAXIMUM ALLOWABLE OFFER

$69,678

© 2018 by WholesalingSpreadsheet.com. All rights reserved

DID YOU KNOW?

The Malasana Pose Has Many Benefits

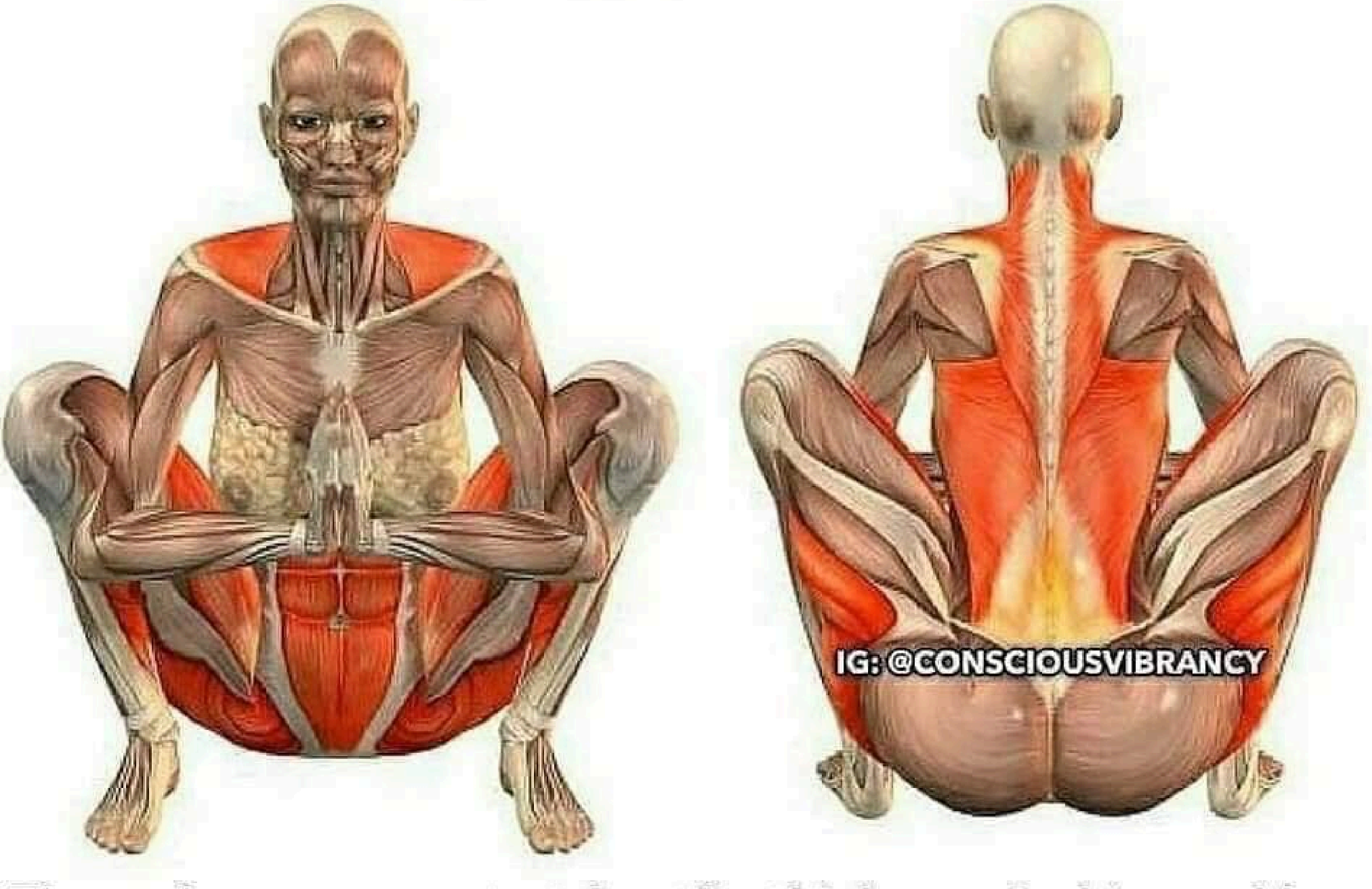

The malasana pose stretches the thighs, groin, hips, ankles, and torso. It tones the abdominal muscles and improves the function of the colon to help with elimination. This pose also increases circulation and blood flow in the pelvis, which can help regulate sexual energy.

SOURCE: YOGAOUTLET

DID YOU KNOW?

WHEN YOU DO THIS STRETCH DAILY...

Holistic Ali

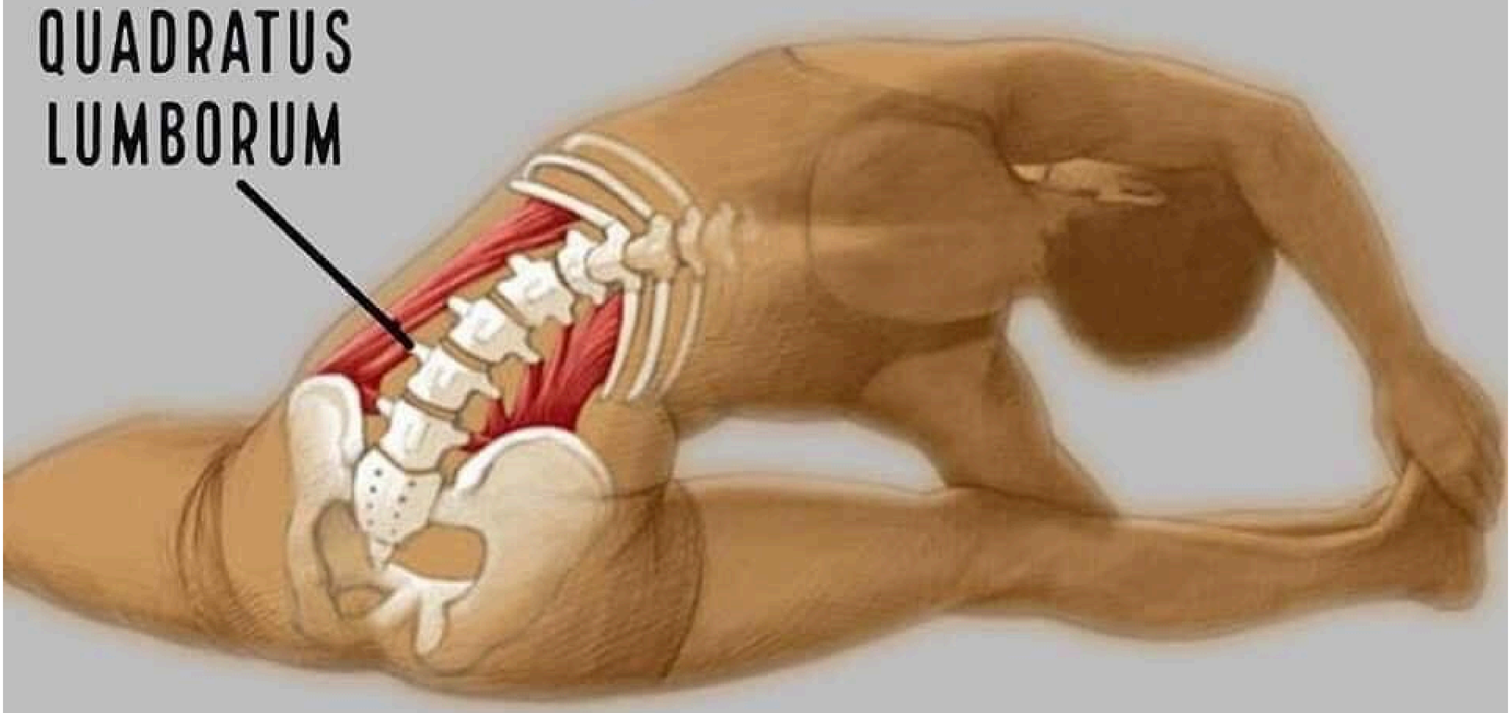

THIS IS THE AREA THAT CAUSES STIFFNESS AND LOWER BACK PAIN FROM LACK OF STRETCHING. THIS AREA MUST BE STRETCHED DAILY, ESPECIALLY IF YOU SIT ALL DAY.

DID YOU KNOW?

PLANKING FOR A FEW MINUTES PER DAY...

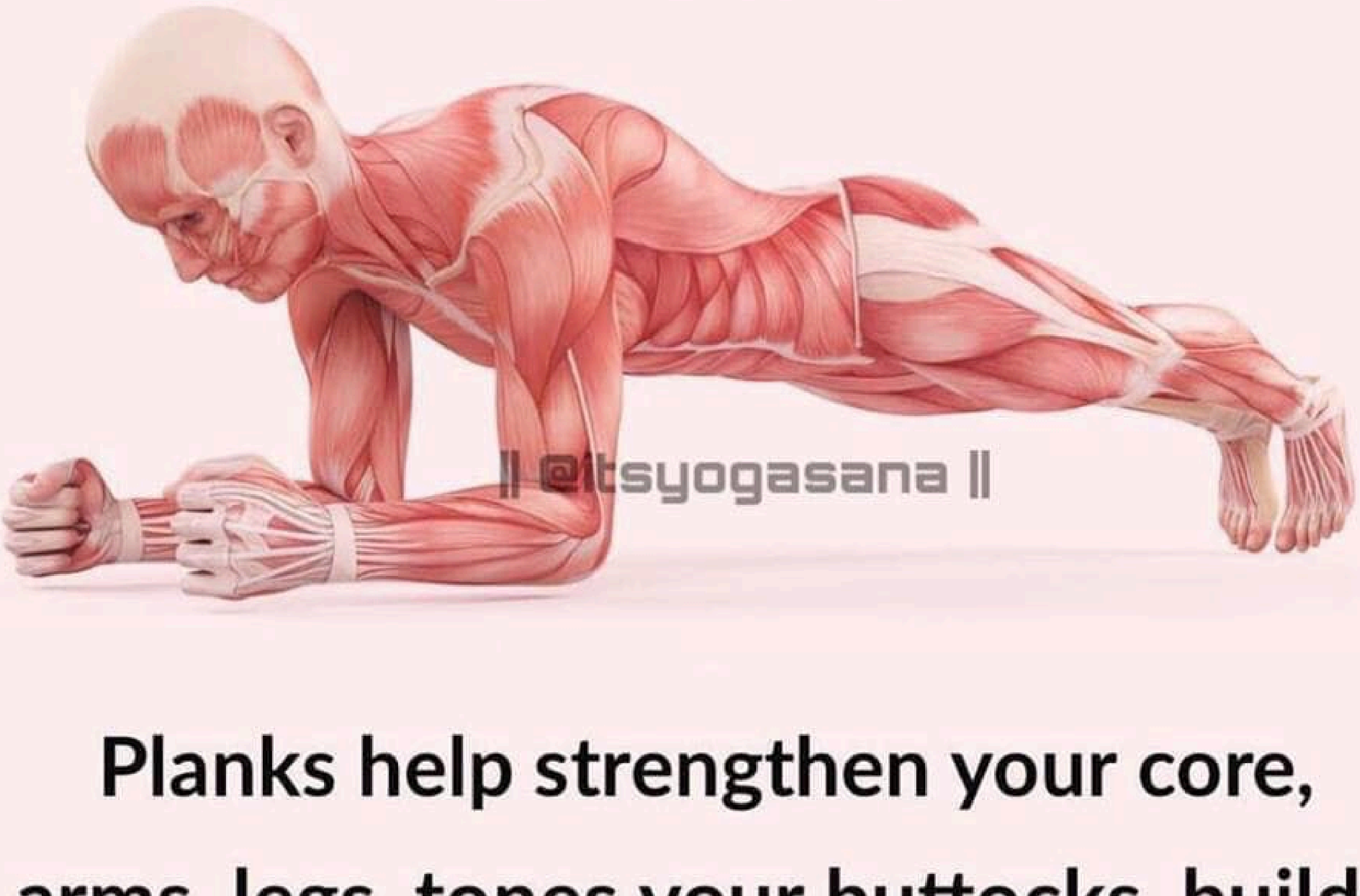

|| @itsyogasana ||

Planks help strengthen your core, arms, legs, tones your buttocks, builds abs & improves your mental focus

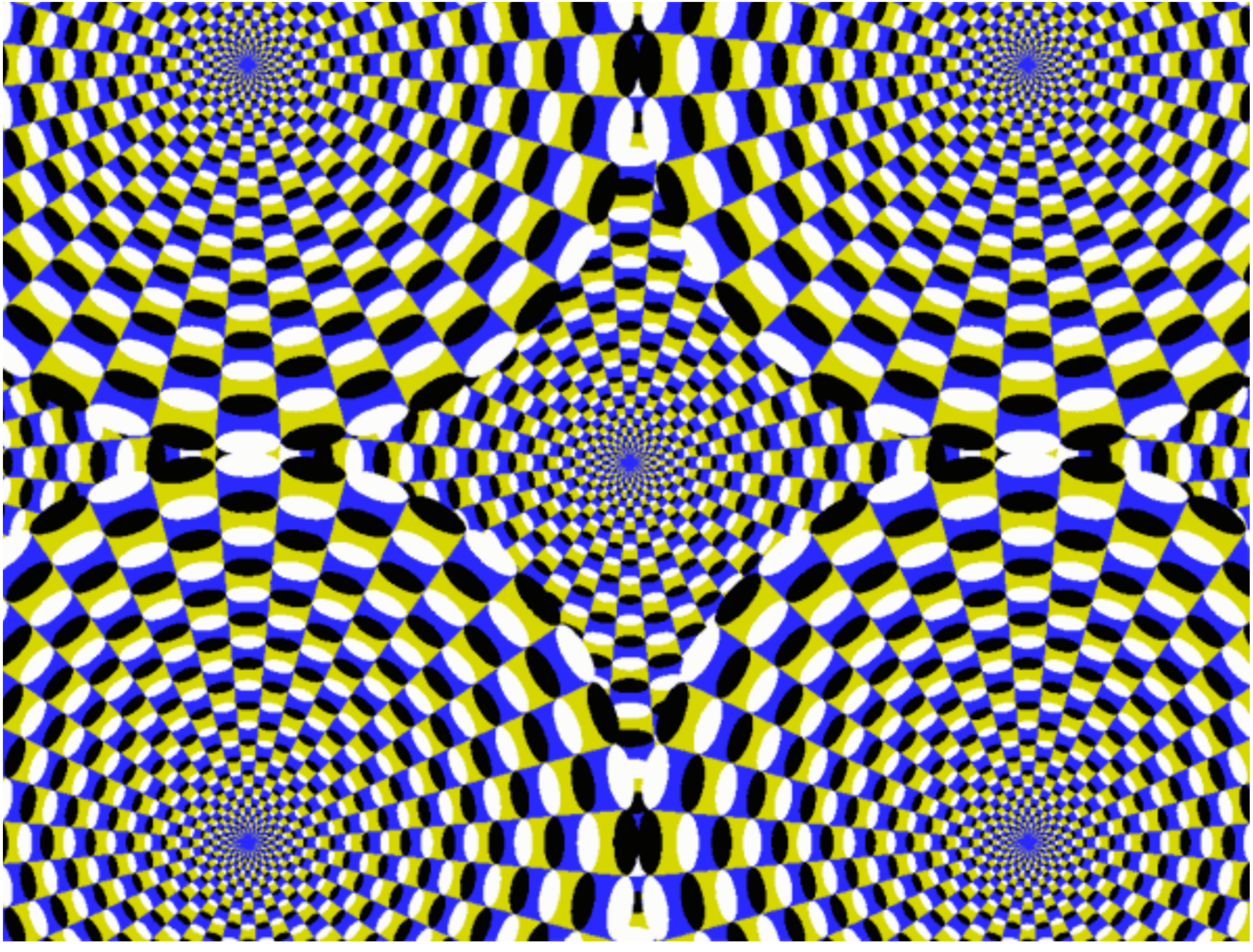

(Use of Hidden Information Content of the Field)
(Can Provide Action-at-a-Distance)

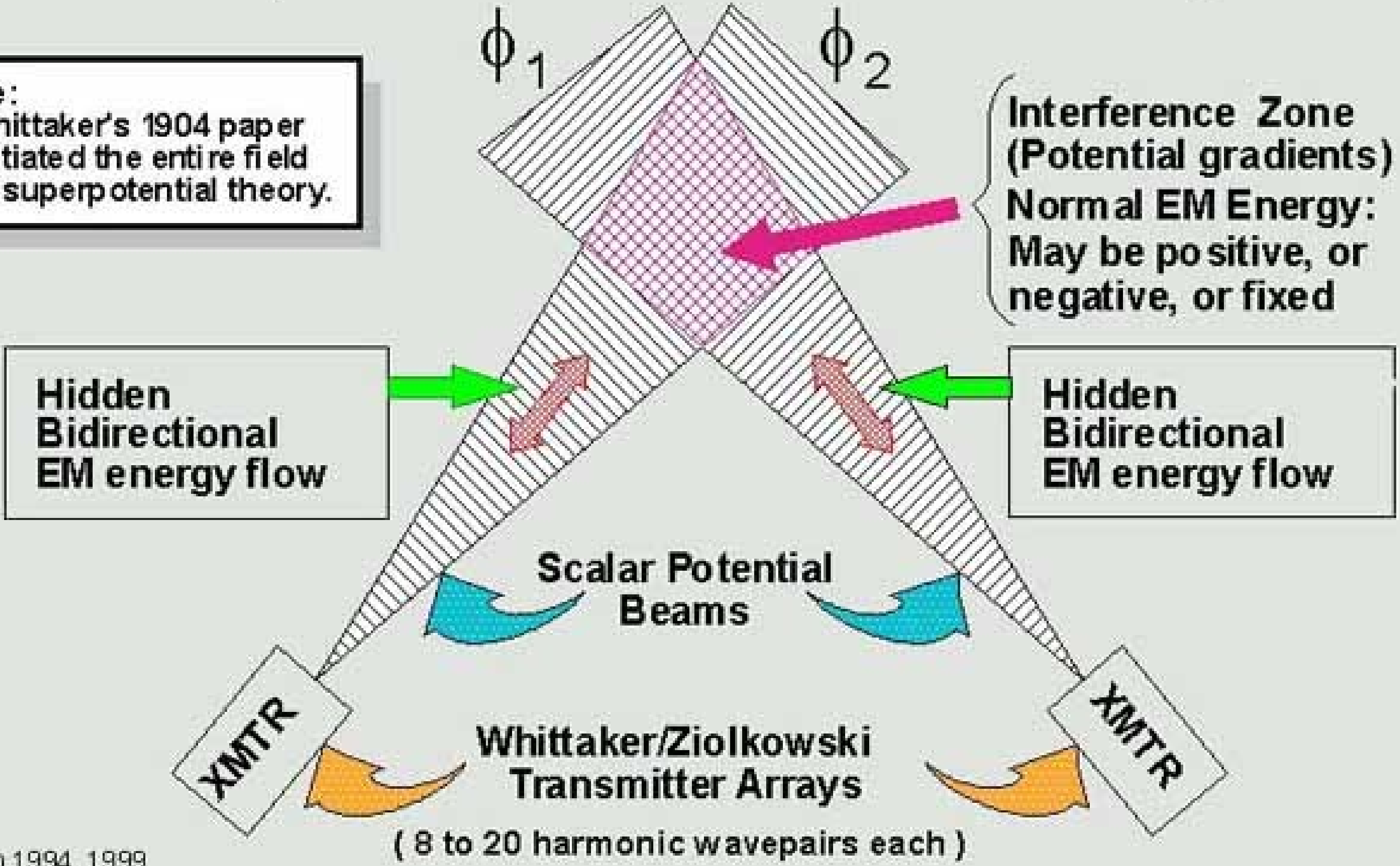

Note:
Whittaker's 1904 paper initiated the entire field of superpotential theory.

Interference Zone (Potential gradients)
Normal EM Energy: May be positive, or negative, or fixed

Hidden Bidirectional EM energy flow

Hidden Bidirectional EM energy flow

Scalar Potential Beams

Whittaker/Ziolkowski Transmitter Arrays
(8 to 20 harmonic wavepairs each)

© T. E. Bearden 1994, 1999

Figure 5. Scalar potential interferometry (between the two sets of bidirectional longitudinal EM wavepair functions) produces all EM force fields and waves.

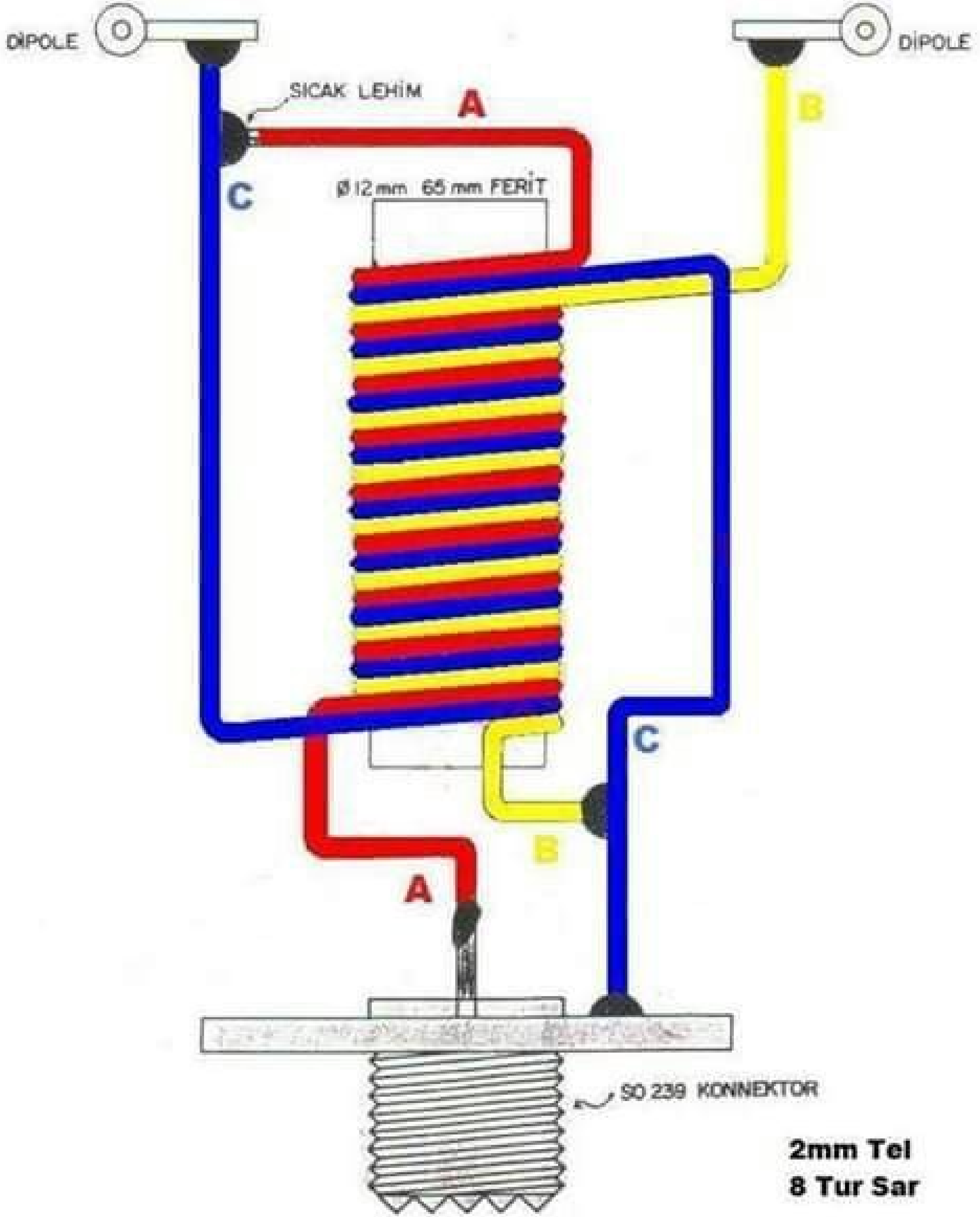
DİPOLE
DİPOLE
SICAK LEHİM
A
B
C
Ø 12 mm 65 mm FERİT
C
B
A
SO 239 KONNEKTOR

2mm Tel
8 Tur Sar

I6IBE

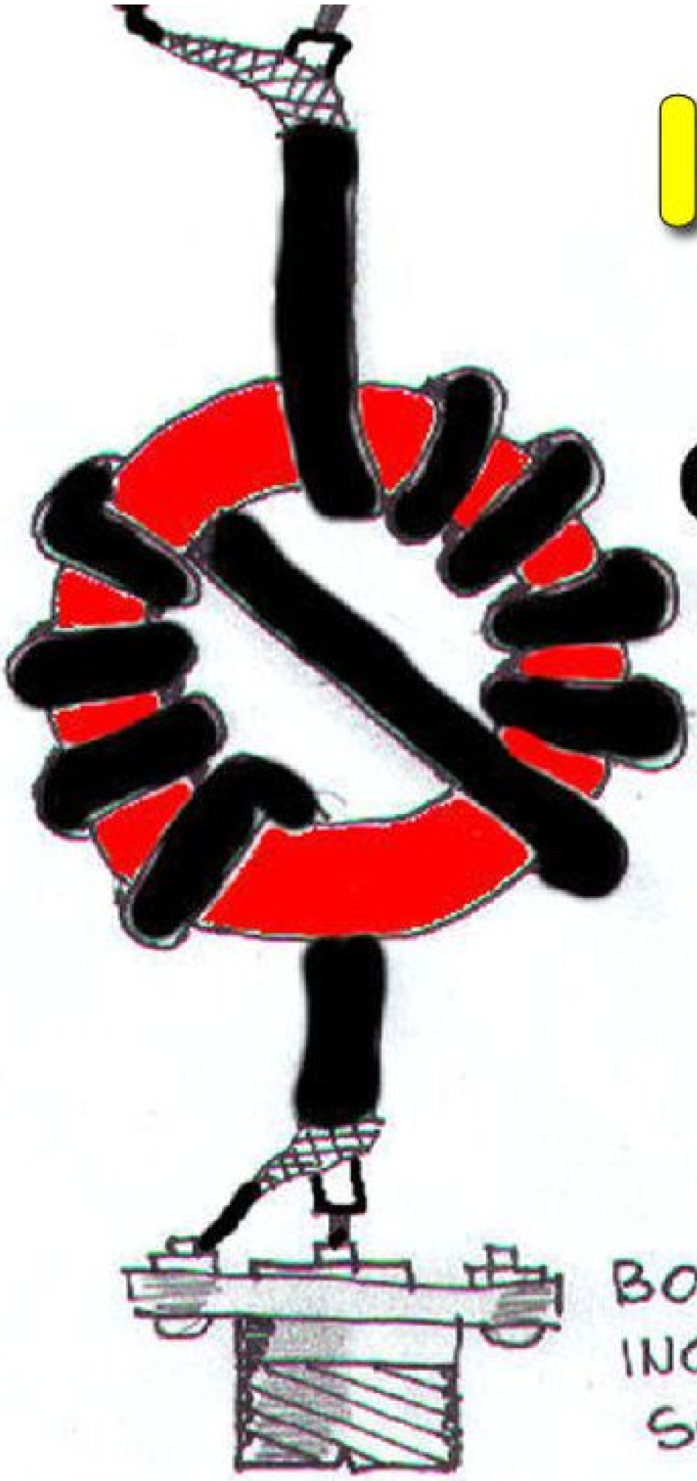
BOCCHETTONE
INGRESSO ANTENNA
SO-239

Choke RFI

toroide

cavo RG-58

5+5 spire

controfase

Formula LUNGHEZZA DELTA LOOP	Esempio 14.00 MHz (20 metri)	Lunghezza TOTALE del LOOP
Velocita'Luce : Freq MHz x 0.97= L	**299.8** VL **: 14.100** MHz x **0,97**	**20.624** metri totale loop
$L = \frac{300.000}{MHz} \times 0.97$	I6IBE	
Formula LUNGHEZZA Stub coax 75 Ohm	Esempio 14.00 MHz (20 metri)	Lunghezza TOTALE cavo 75 Ohm
Lunghezza LOOP x 0.66 : 4	**20.624** x **0.66** : **4**	**3.402** Metri coax RG-59

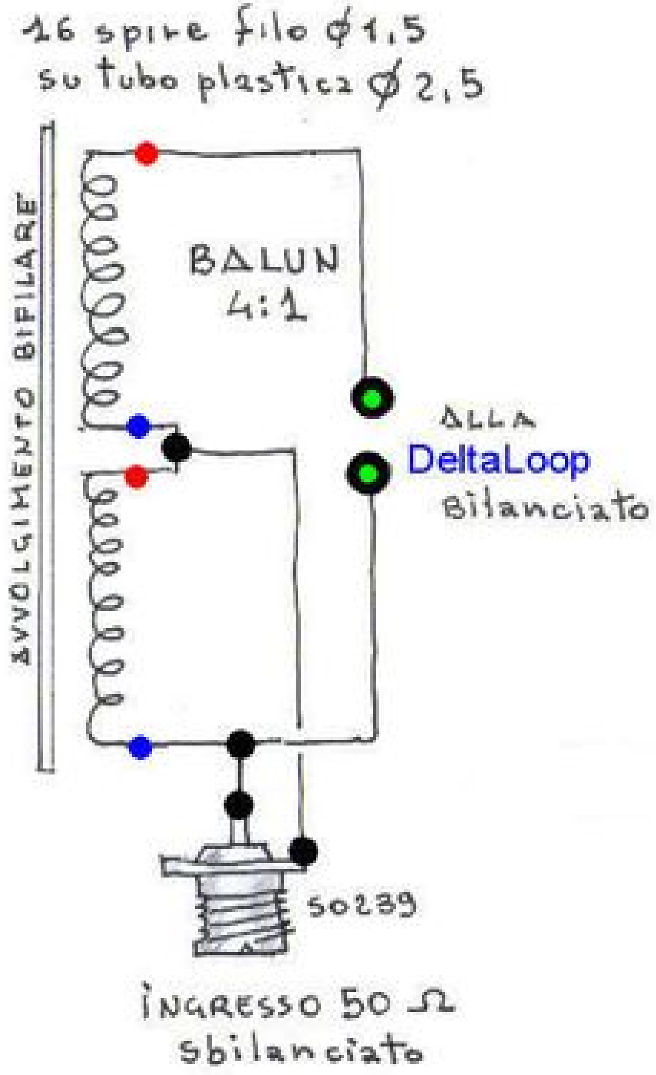
16 spire filo ∅ 1,5
su tubo plastica ∅ 2,5
AVVOLGIMENTO BIFILARE
BALUN 4:1
ALLA
DeltaLoop
Bilanciato
SO239
ingresso 50 Ω
sbilanciato

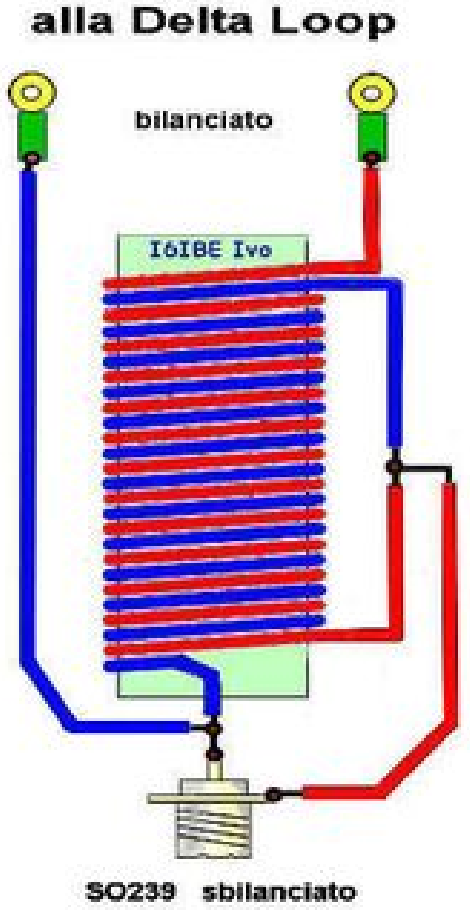
alla Delta Loop
bilanciato
I6IBE Ivo
SO239 sbilanciato

BALUN 4:1

Phase Conjugate Wavepairs Produce New Waves

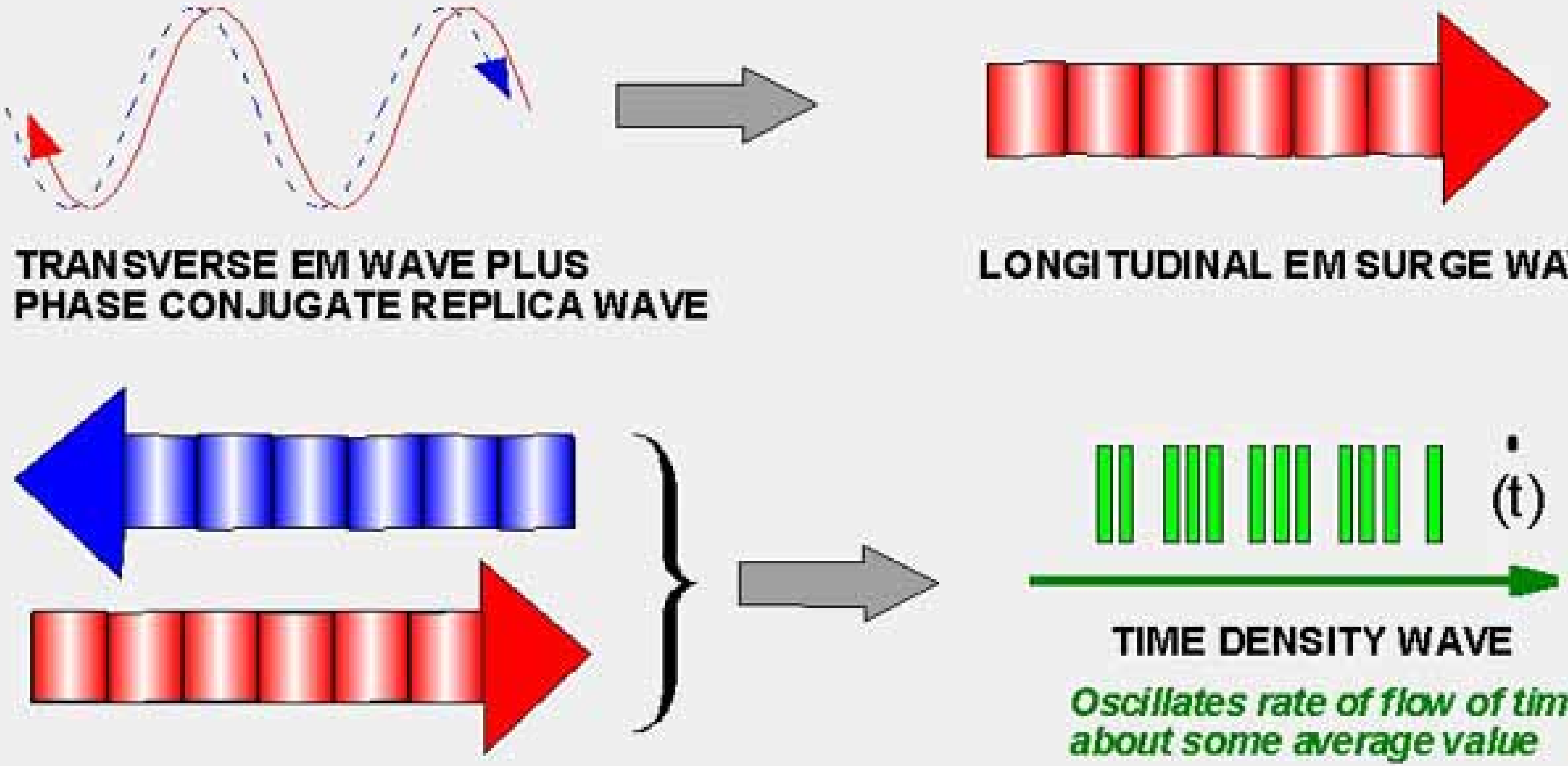

TRANSVERSE EM WAVE PLUS PHASE CONJUGATE REPLICA WAVE

LONGITUDINAL EM SURGE WAVE

LONGITUDINAL EM WAVE PLUS PHASE CONJUGATE REPLICA WAVE

TIME DENSITY WAVE

Oscillates rate of flow of time about some average value

© T.E. BEARDEN 1994, 1995

Supercap
Joule
Thief
2.5V
22F
DRE
2.5V
22F

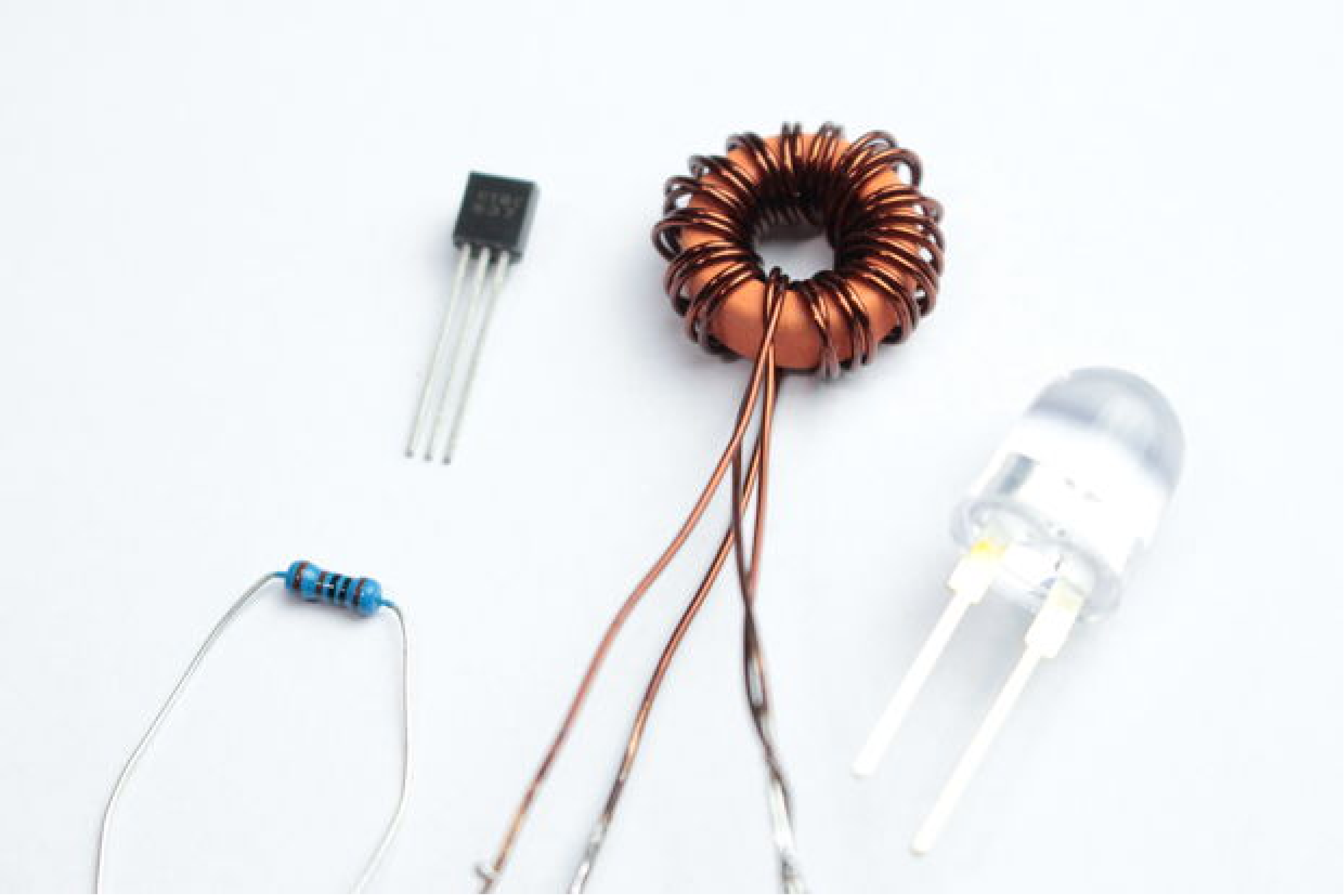

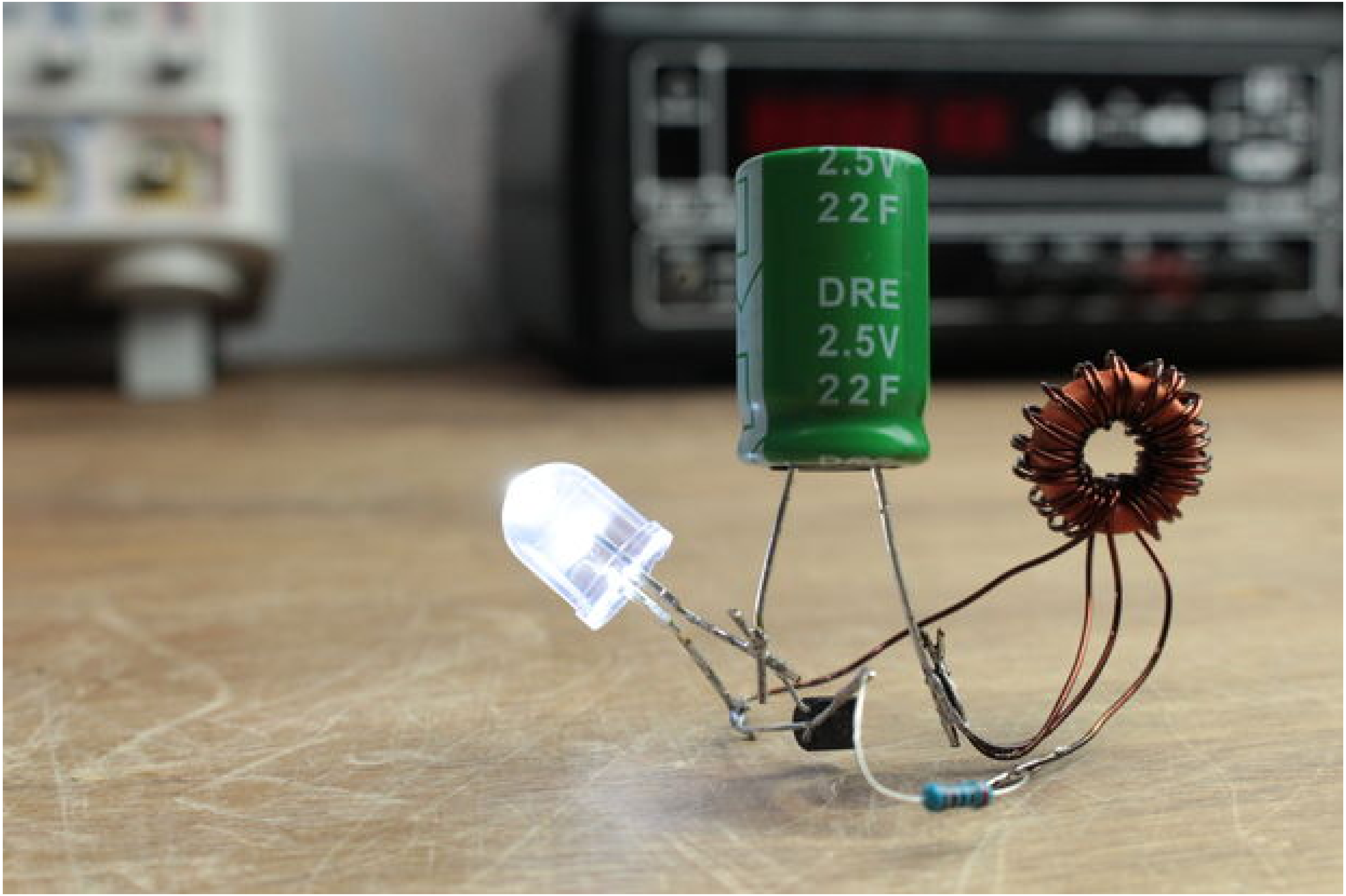
2.5V
22F
DRE
2.5V
22F

WIRTH

Hauptsammler
Mettmann

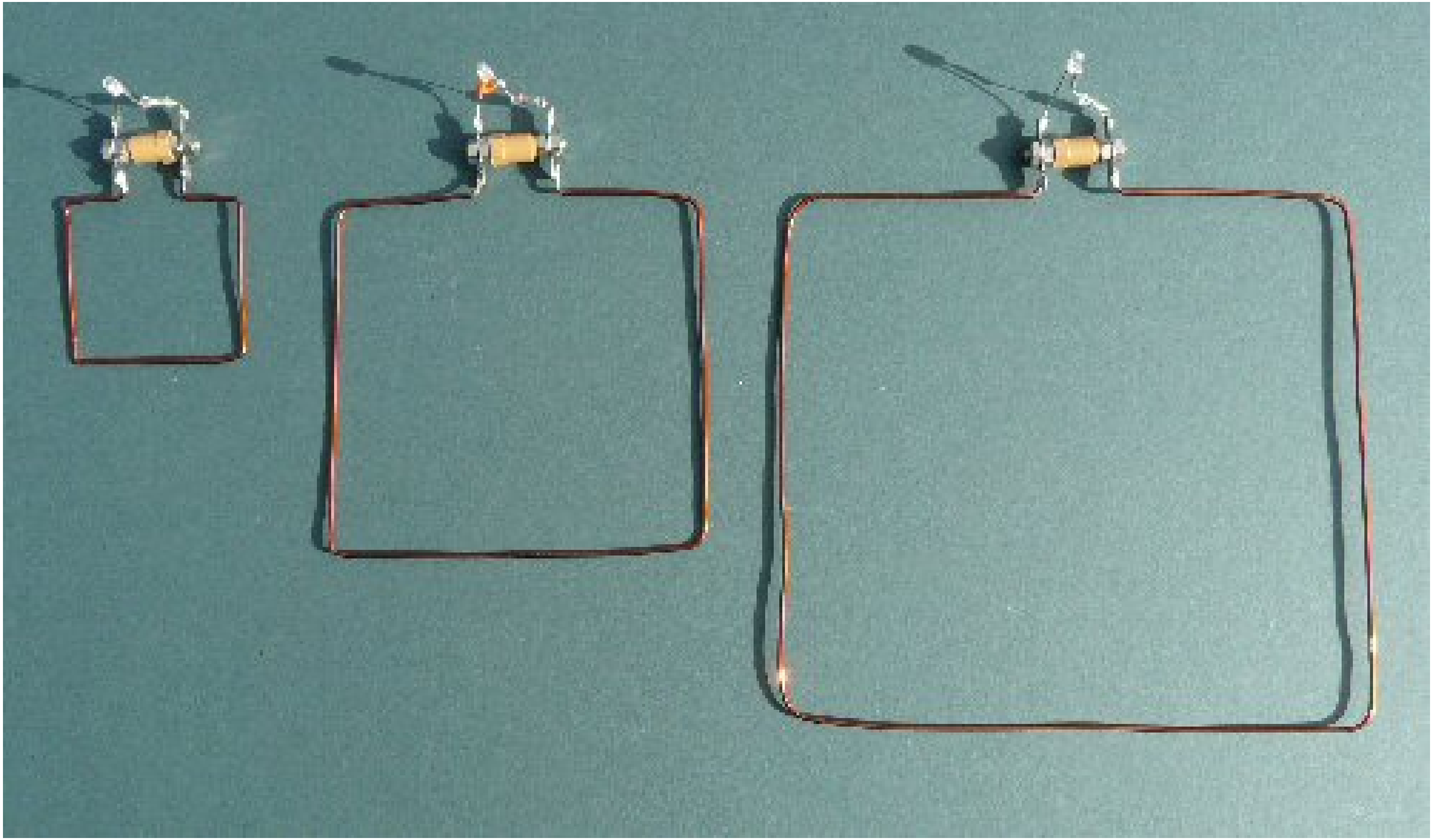

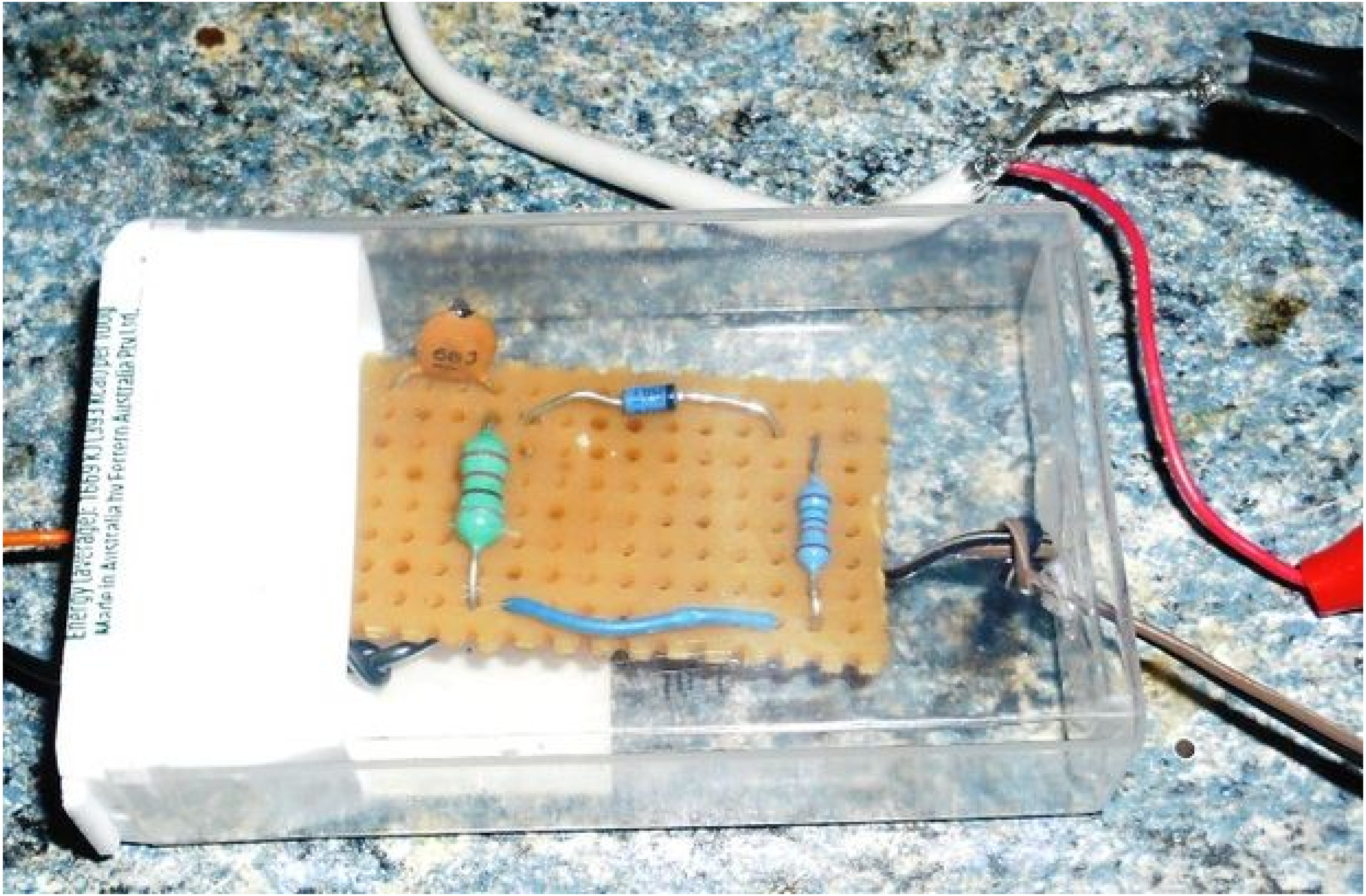
68J
Made in Australia by Fortein Australia Pty Ltd.

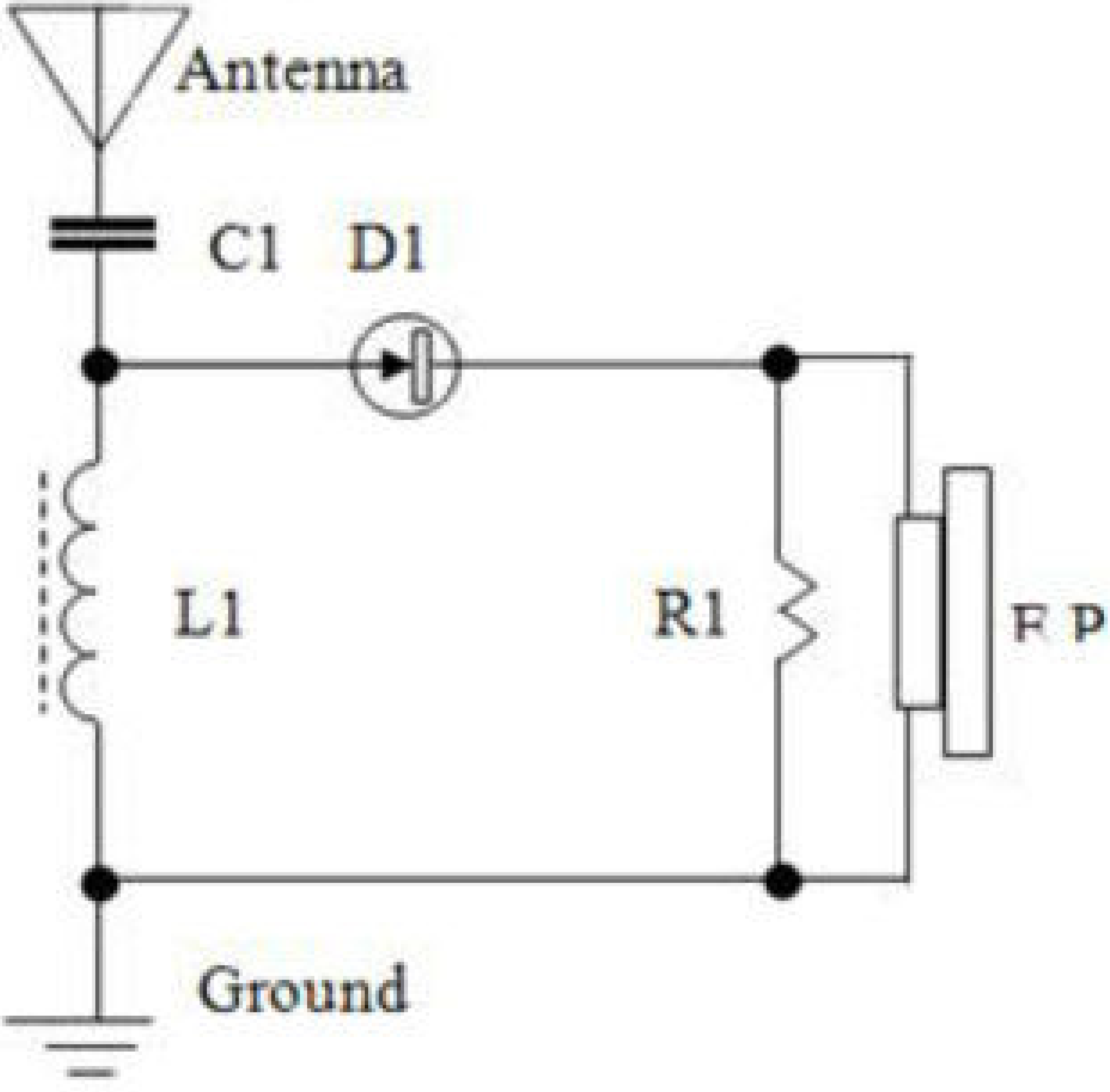
Antenna
C1
D1
L1
R1
F P
Ground

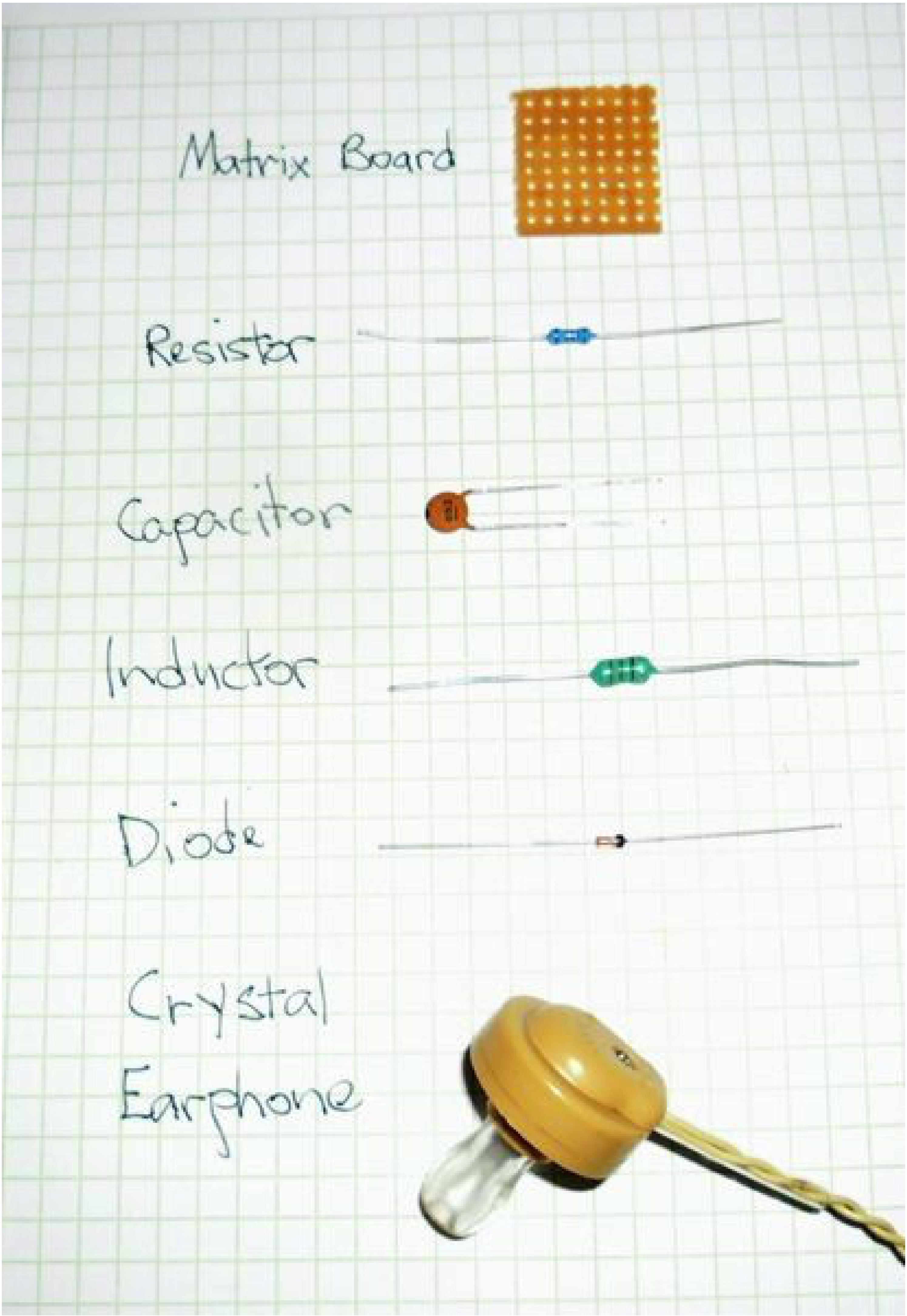
Matrix Board
Resistor
Capacitor
Inductor
Diode
Crystal
Earphone

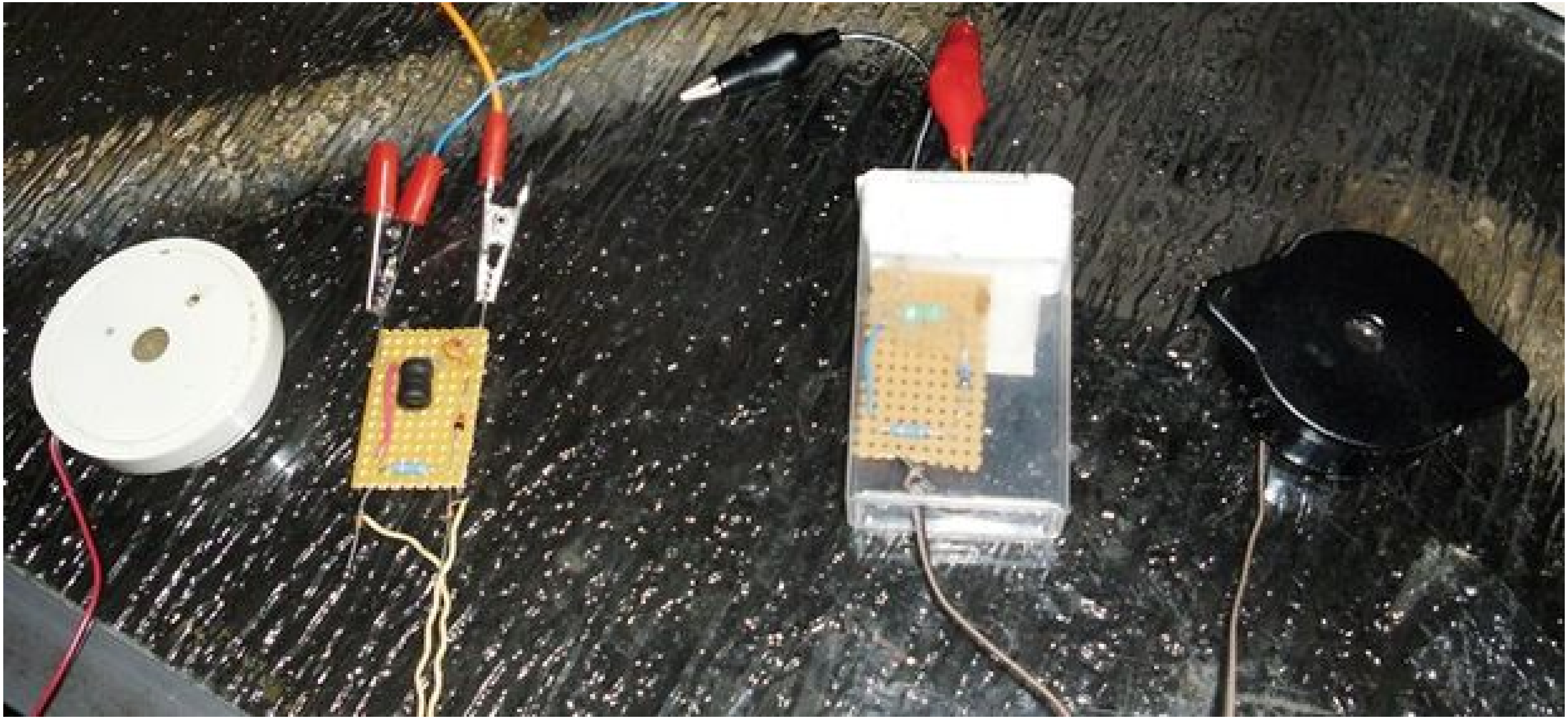

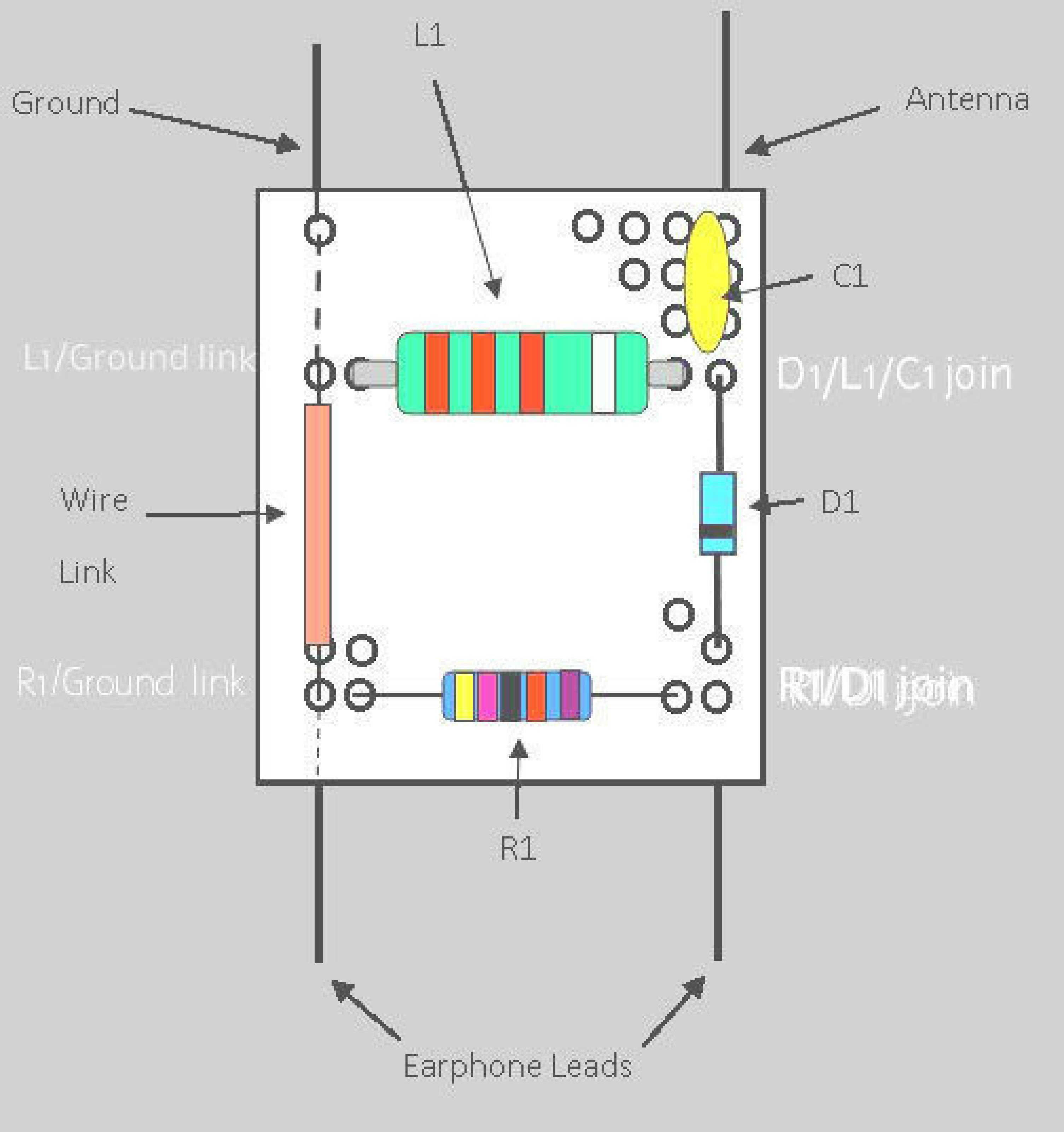
Ground
L1
Antenna
C1
L1/Ground link
D1/L1/C1 join
Wire
Link
D1
R1/Ground link
R1/D1 join
R1
Earphone Leads

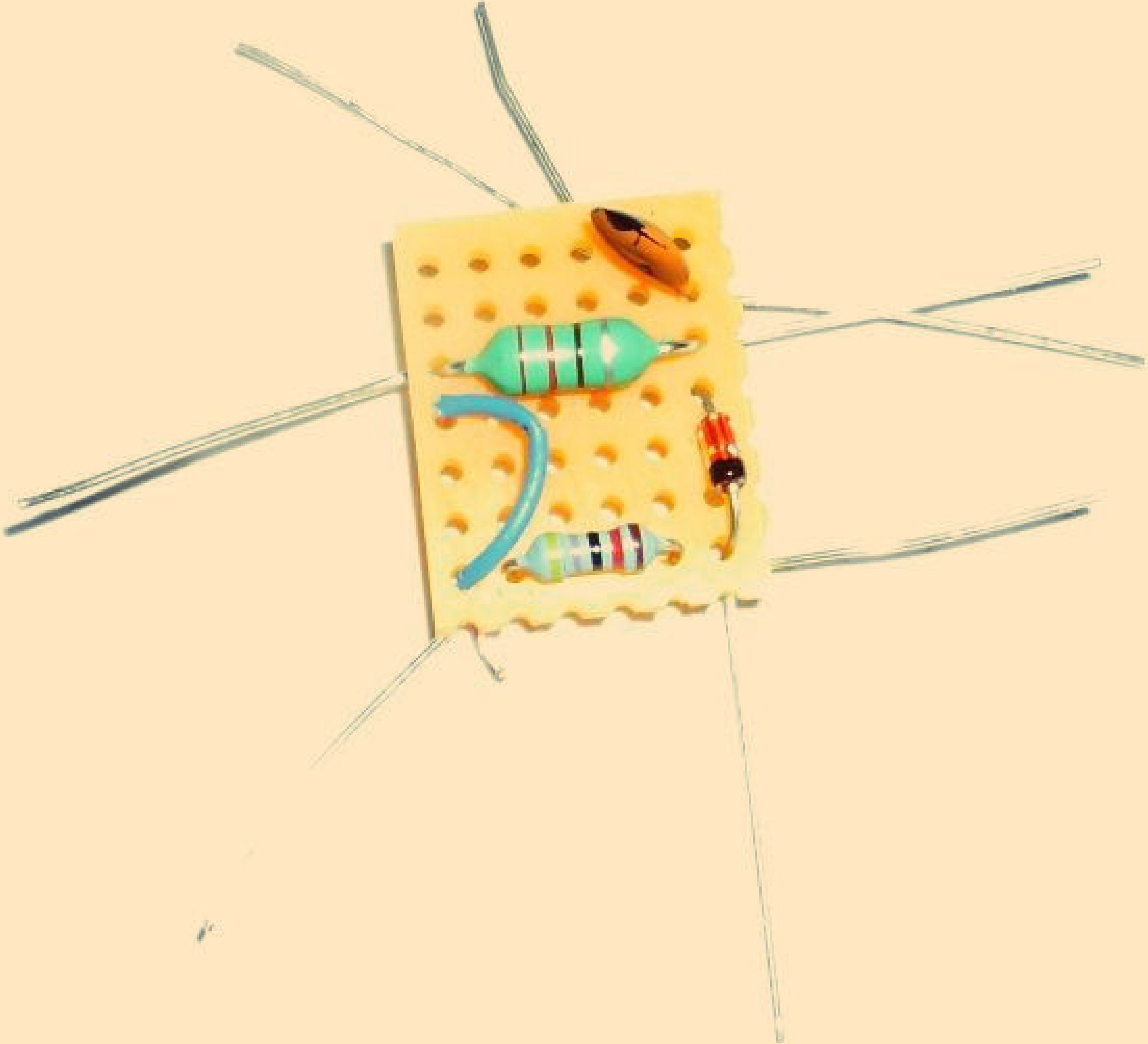

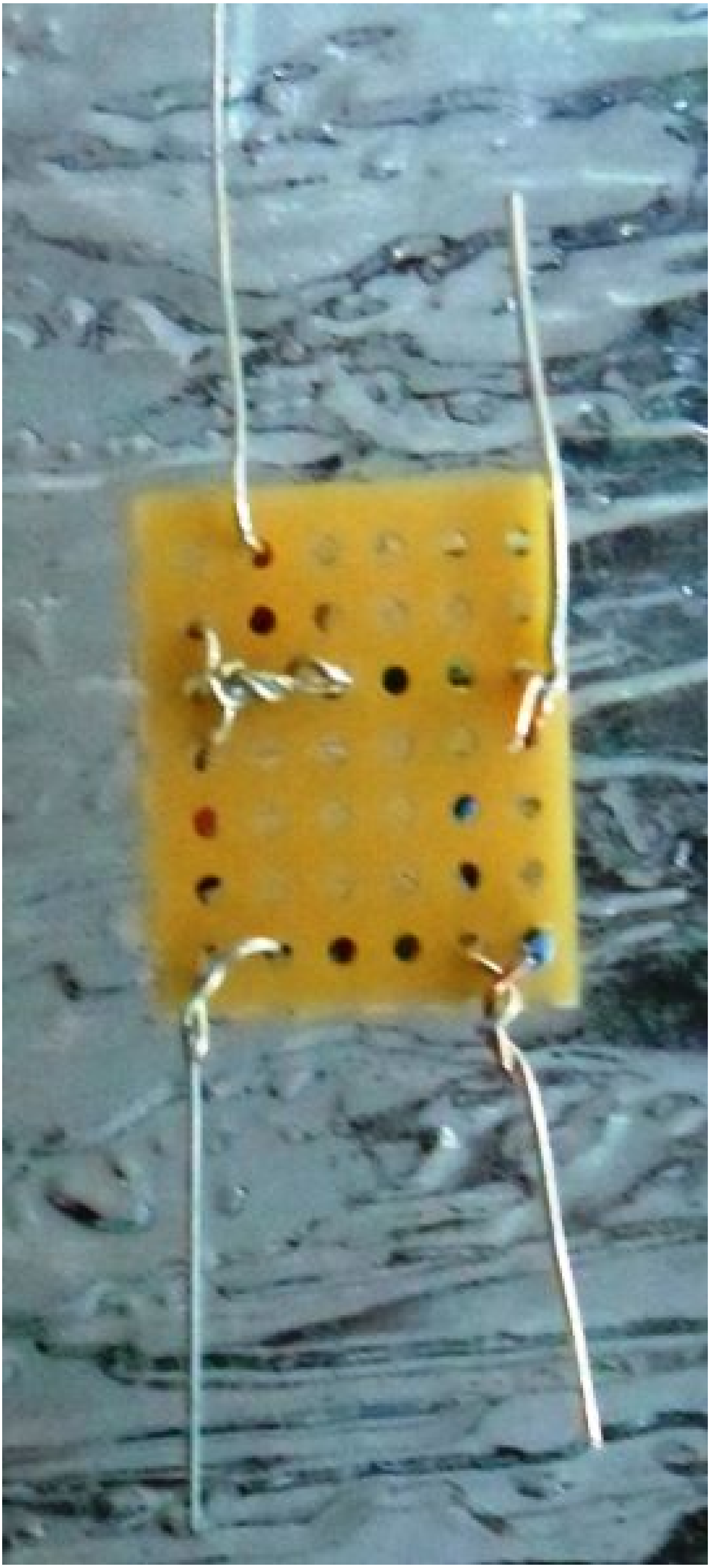

24g NET
FRESHNESS IN 2 CALORIES
tic tac
Strawberry Fields

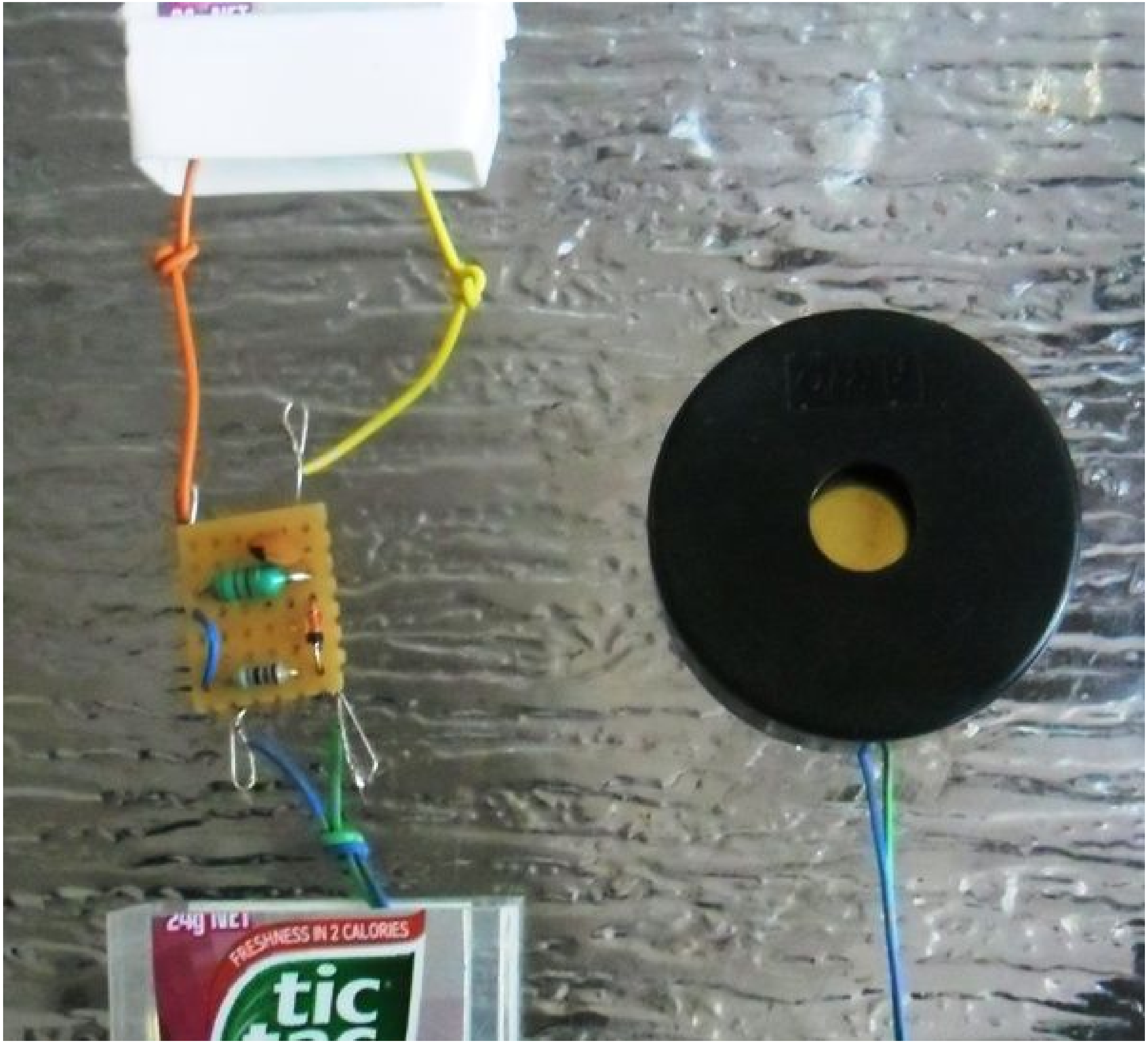
24g NET
FRESHNESS IN 2 CALORIES
tic tac

24g NET
FRESHNESS IN 2 CALORIES
tic tac
Strawberry Fields

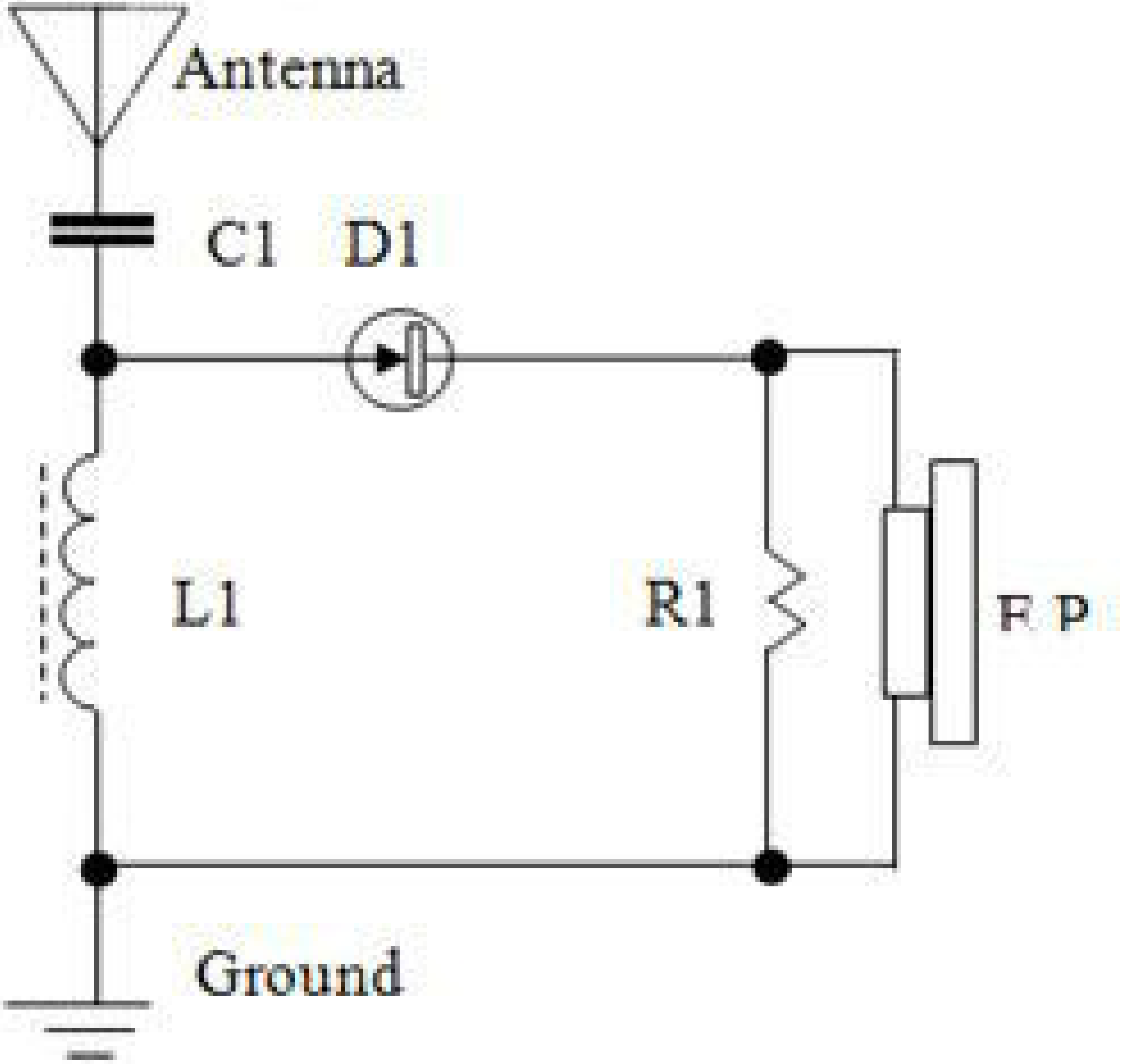
Antenna
C1
D1
L1
R1
F P
Ground

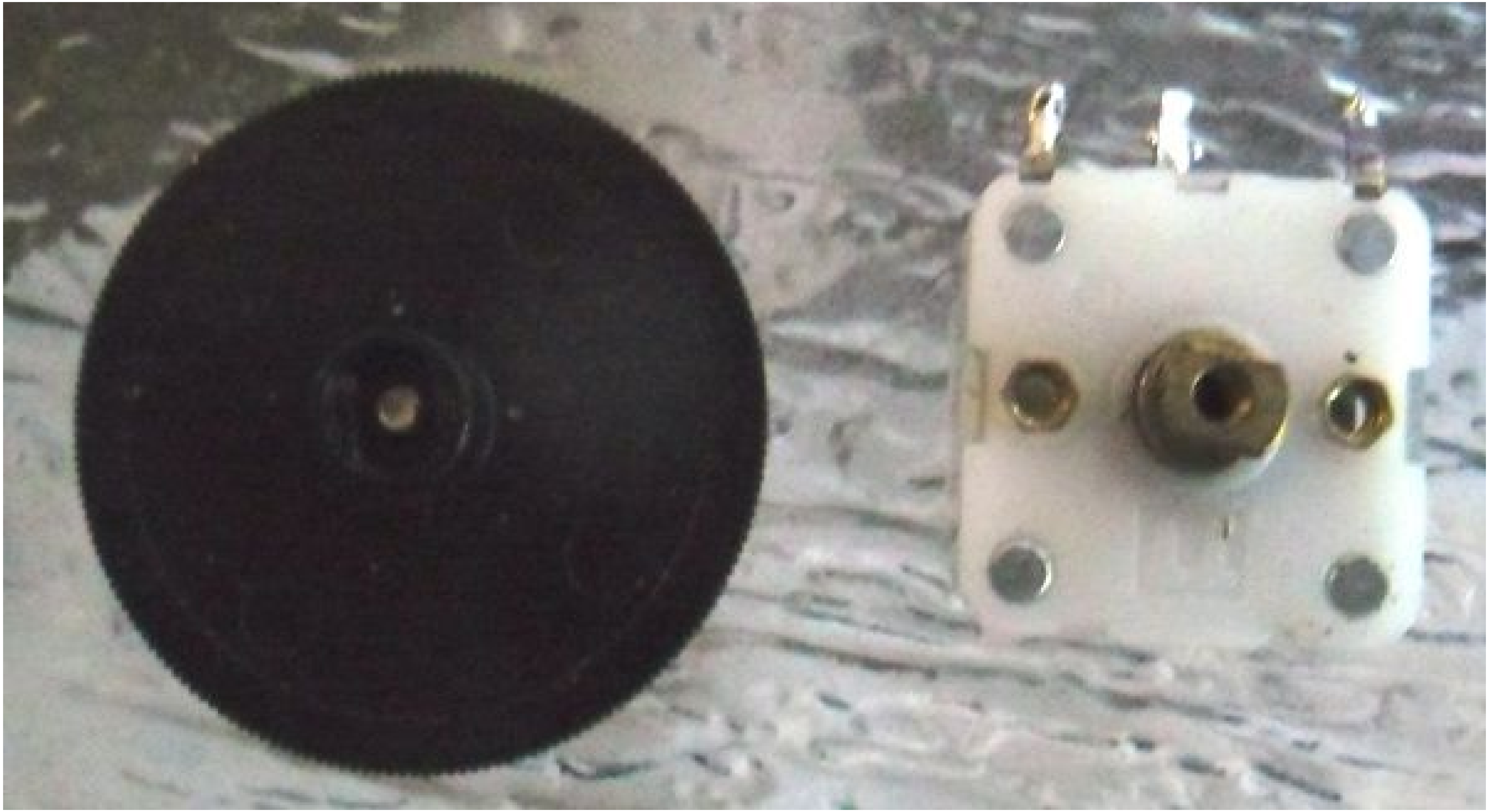

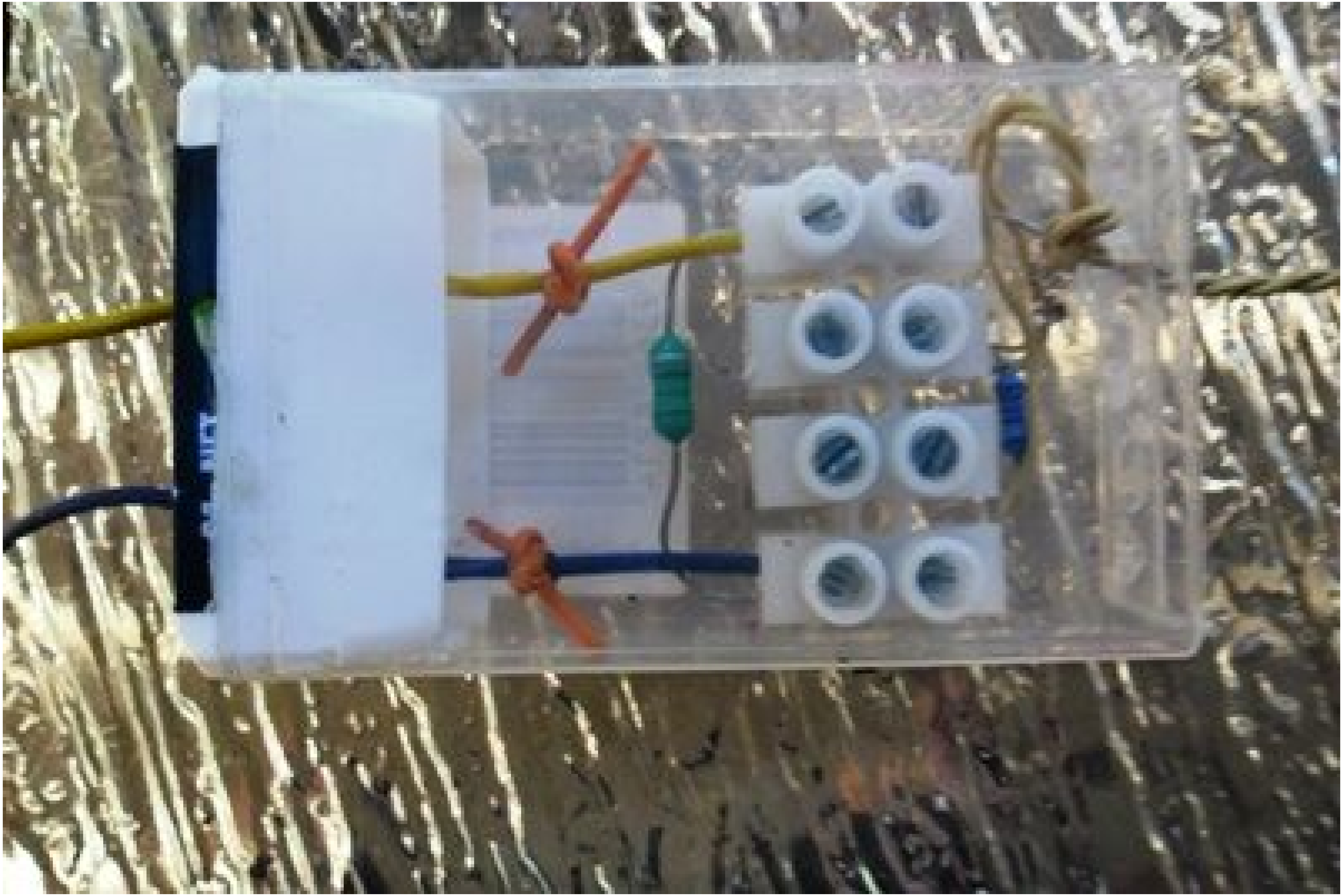

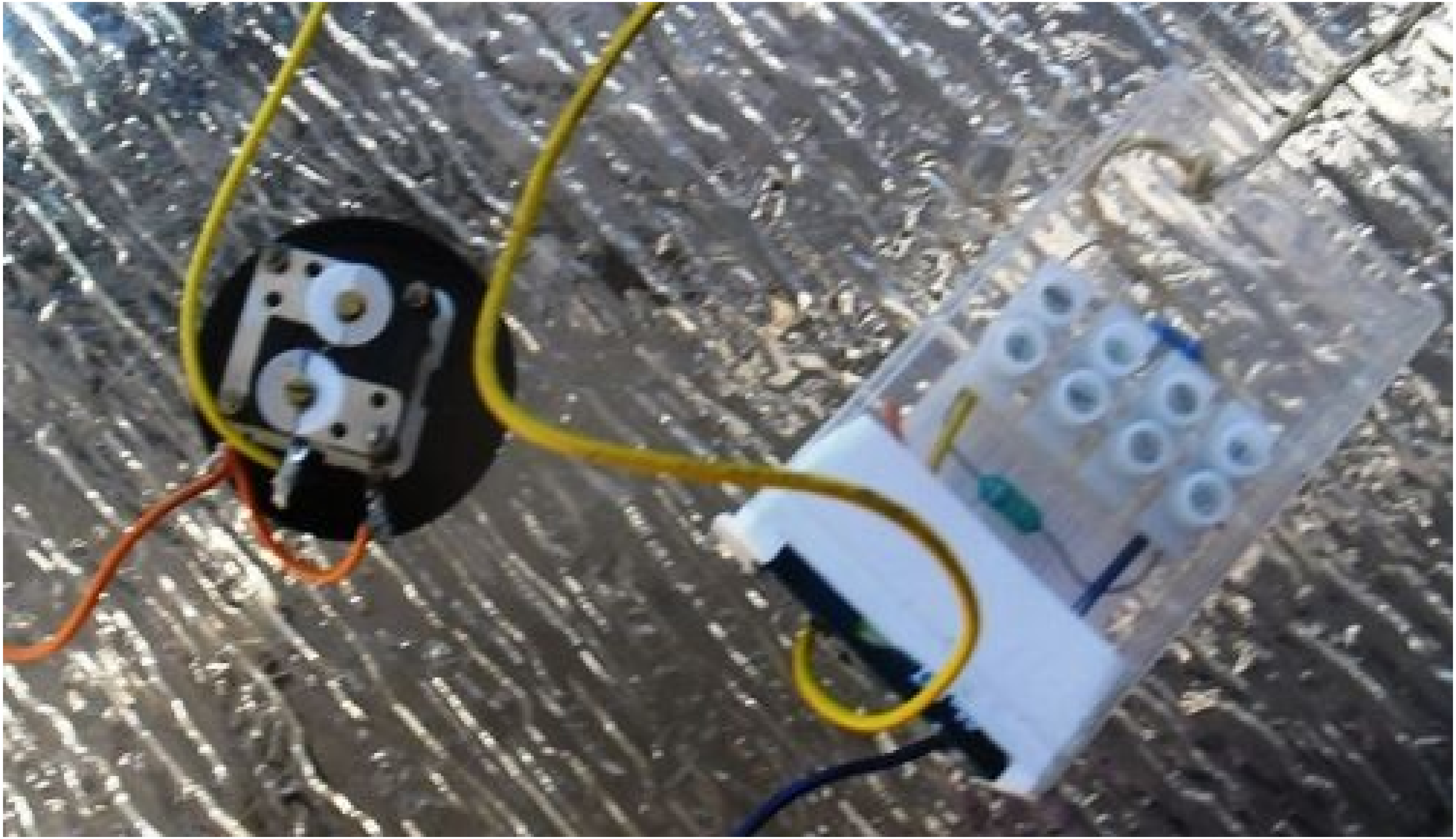

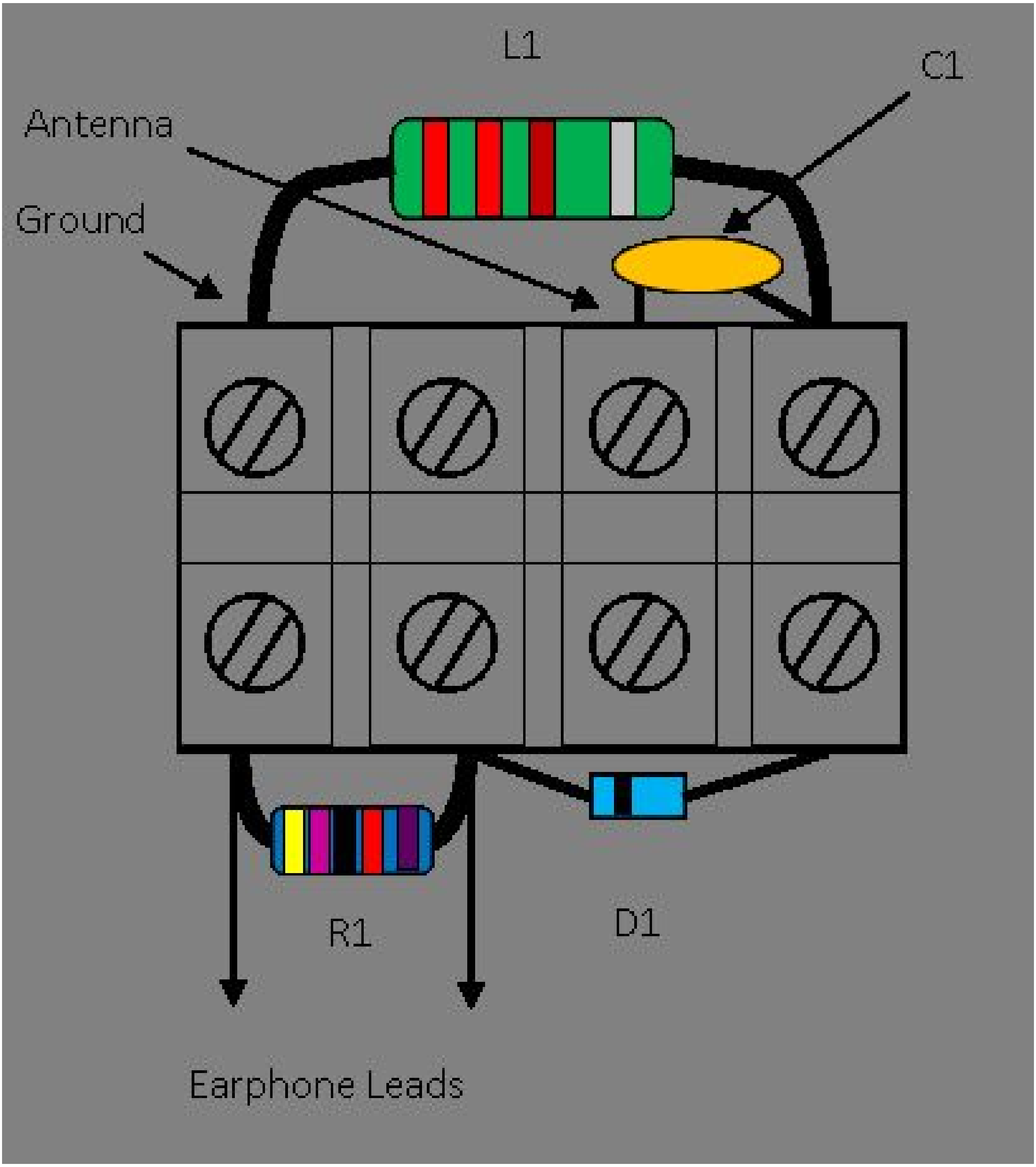
L1
C1
Antenna
Ground
R1
D1
Earphone Leads

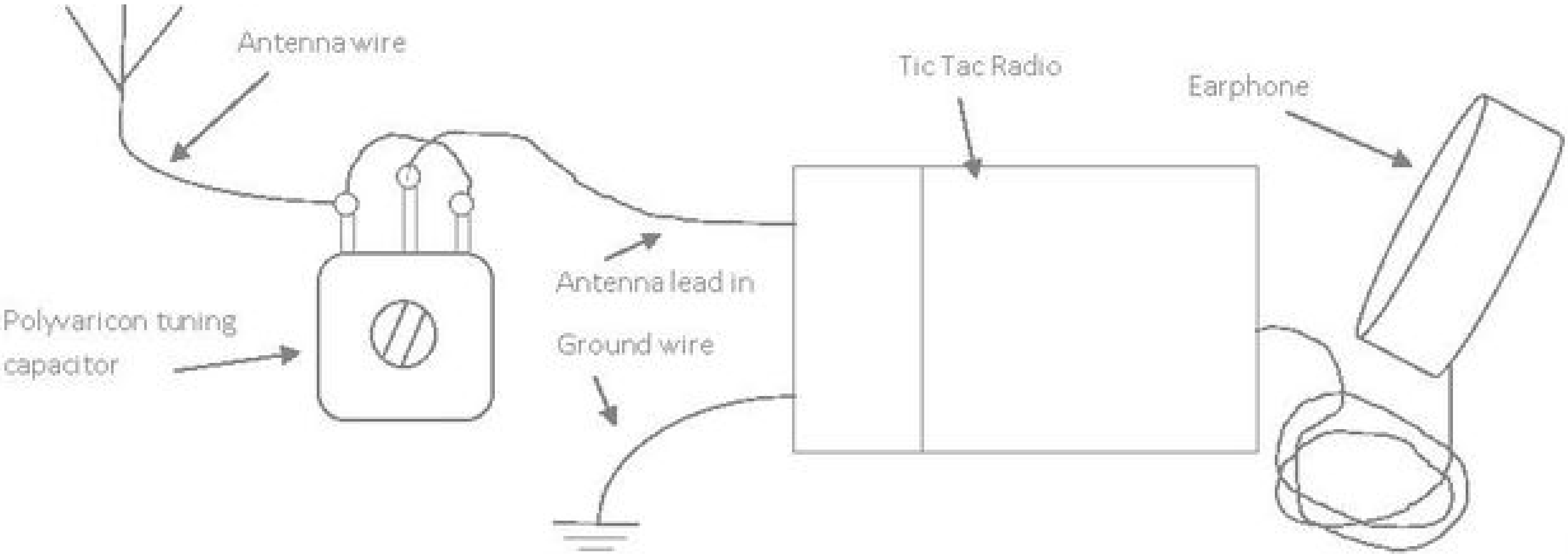
Antenna wire
Tic Tac Radio
Earphone
Antenna lead in
Polyvaricon tuning capacitor
Ground wire

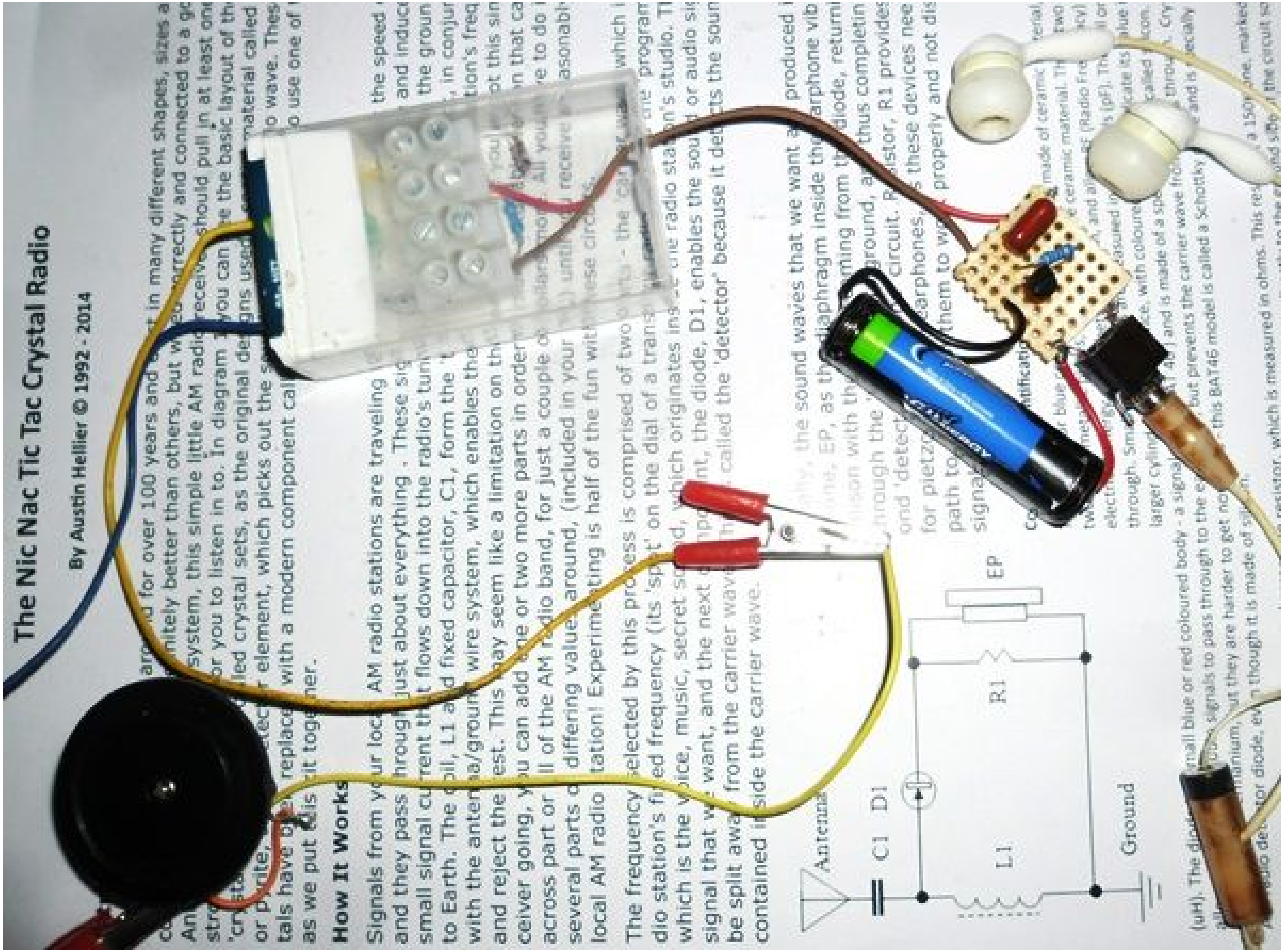
The Nic Nac Tic Tac Crystal Radio
By Austin Hellier © 1992 - 2014
How It Works
Antenna
C1
D1
L1
R1
EP
Ground

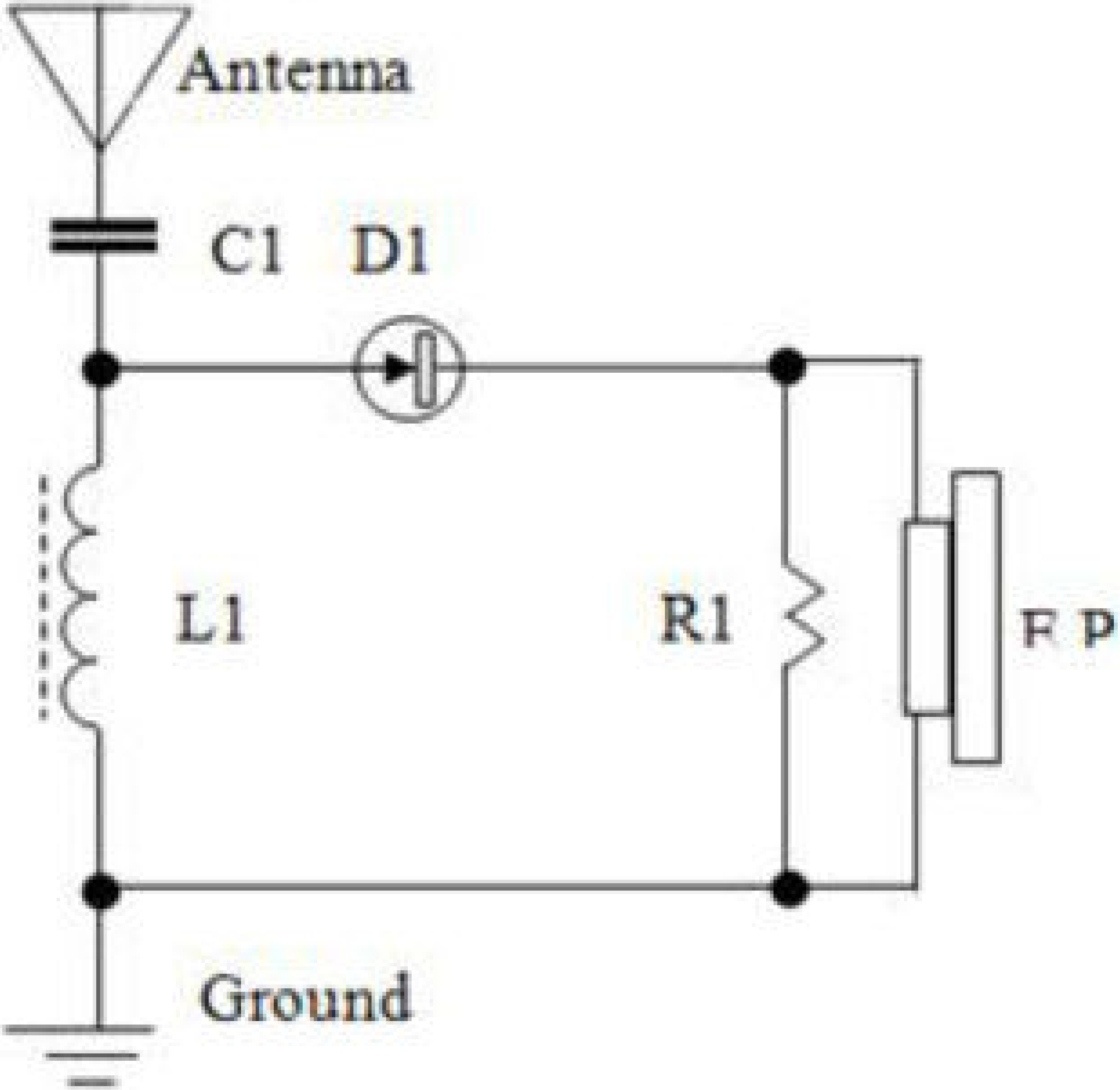
Antenna
C1
D1
L1
R1
E P
Ground

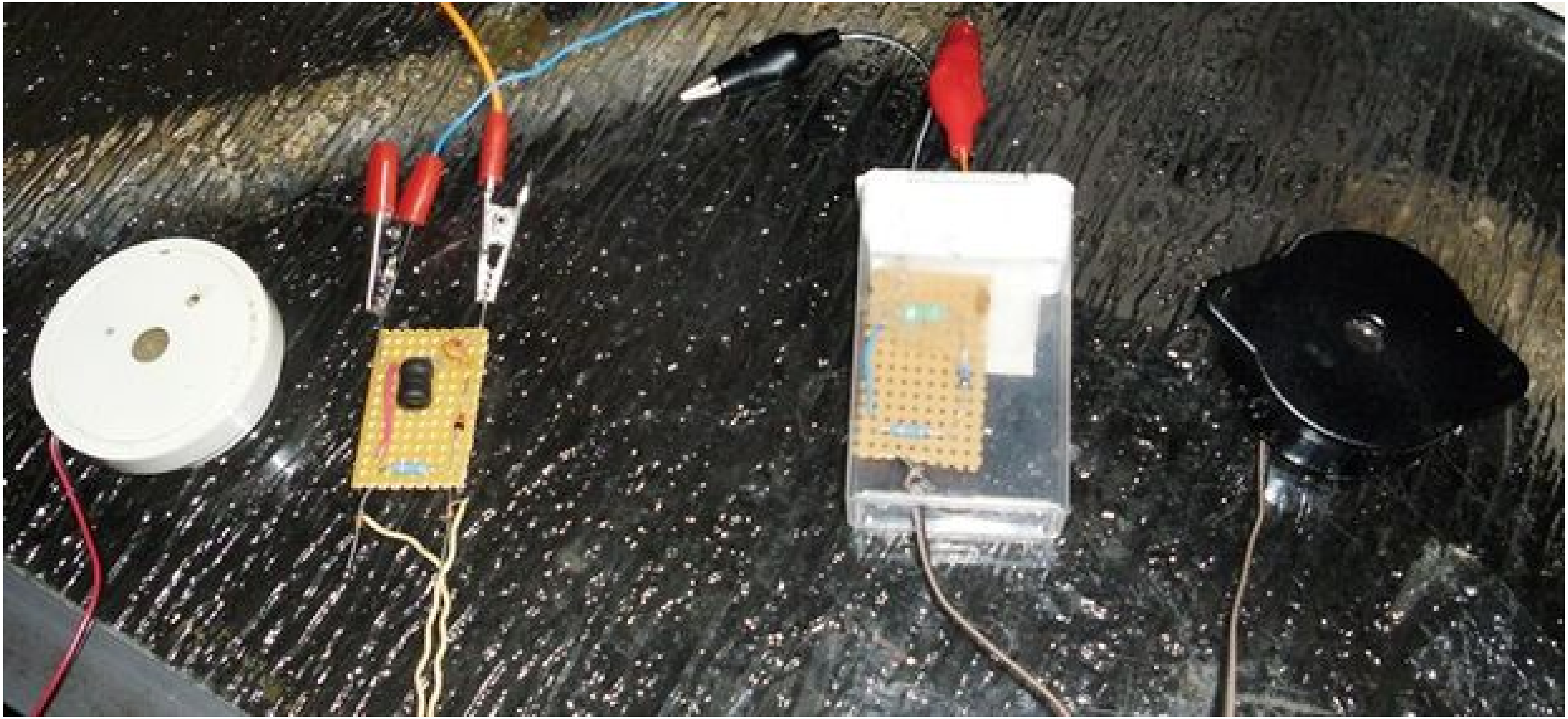

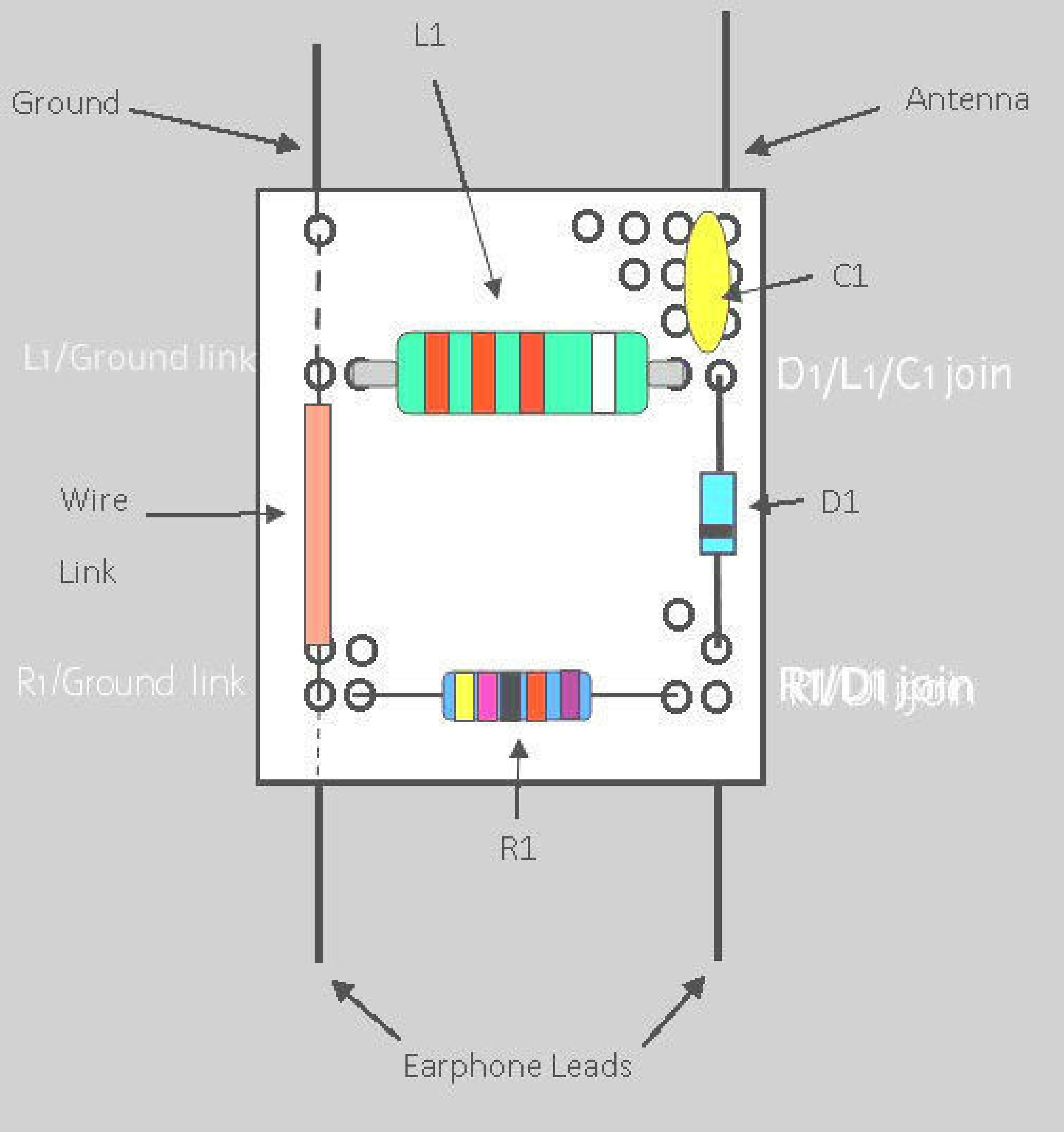
L1
Ground
Antenna
C1
L1/Ground link
D1/L1/C1 join
Wire
Link
D1
R1/Ground link
R1/D1 join
R1
Earphone Leads

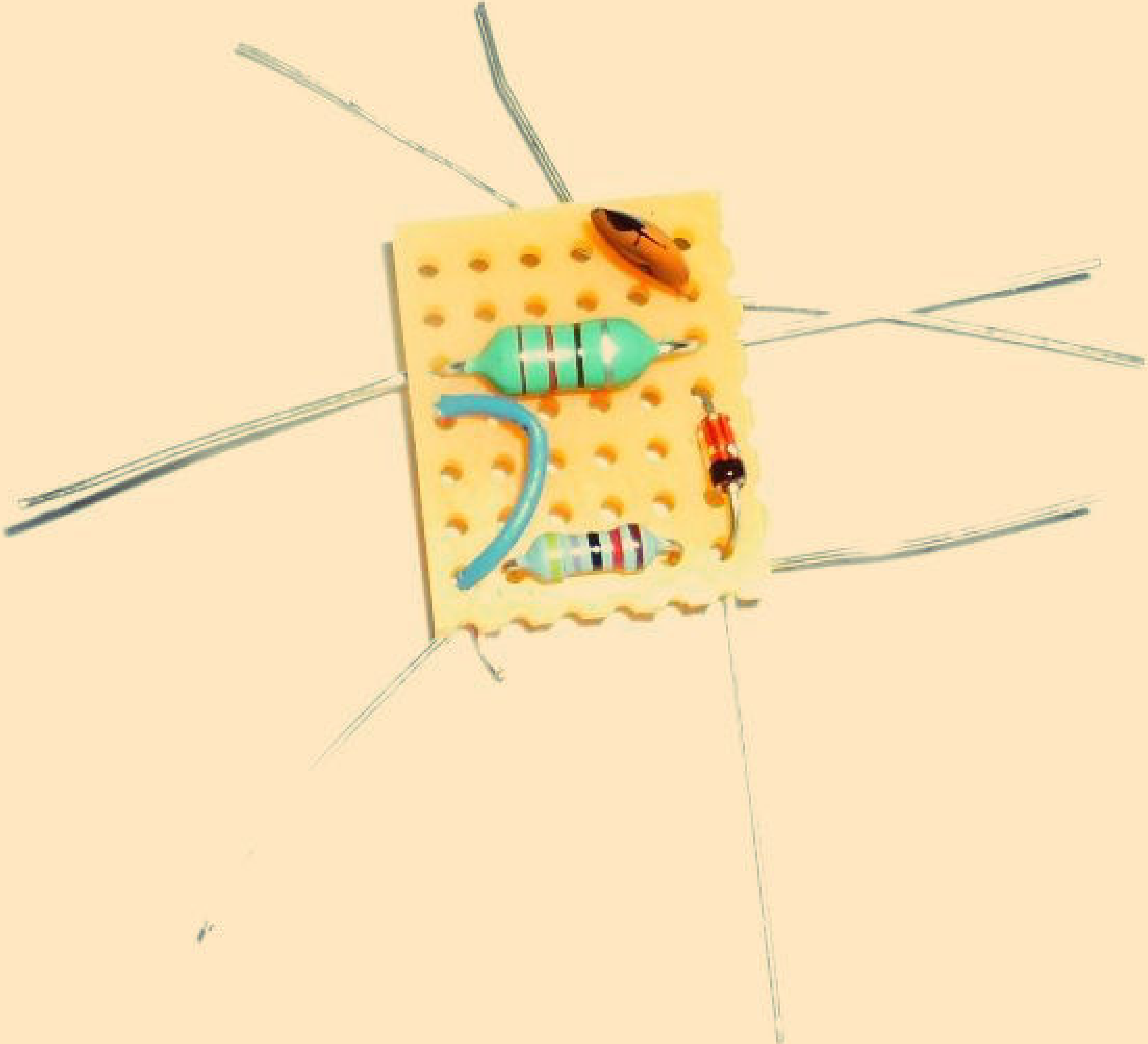

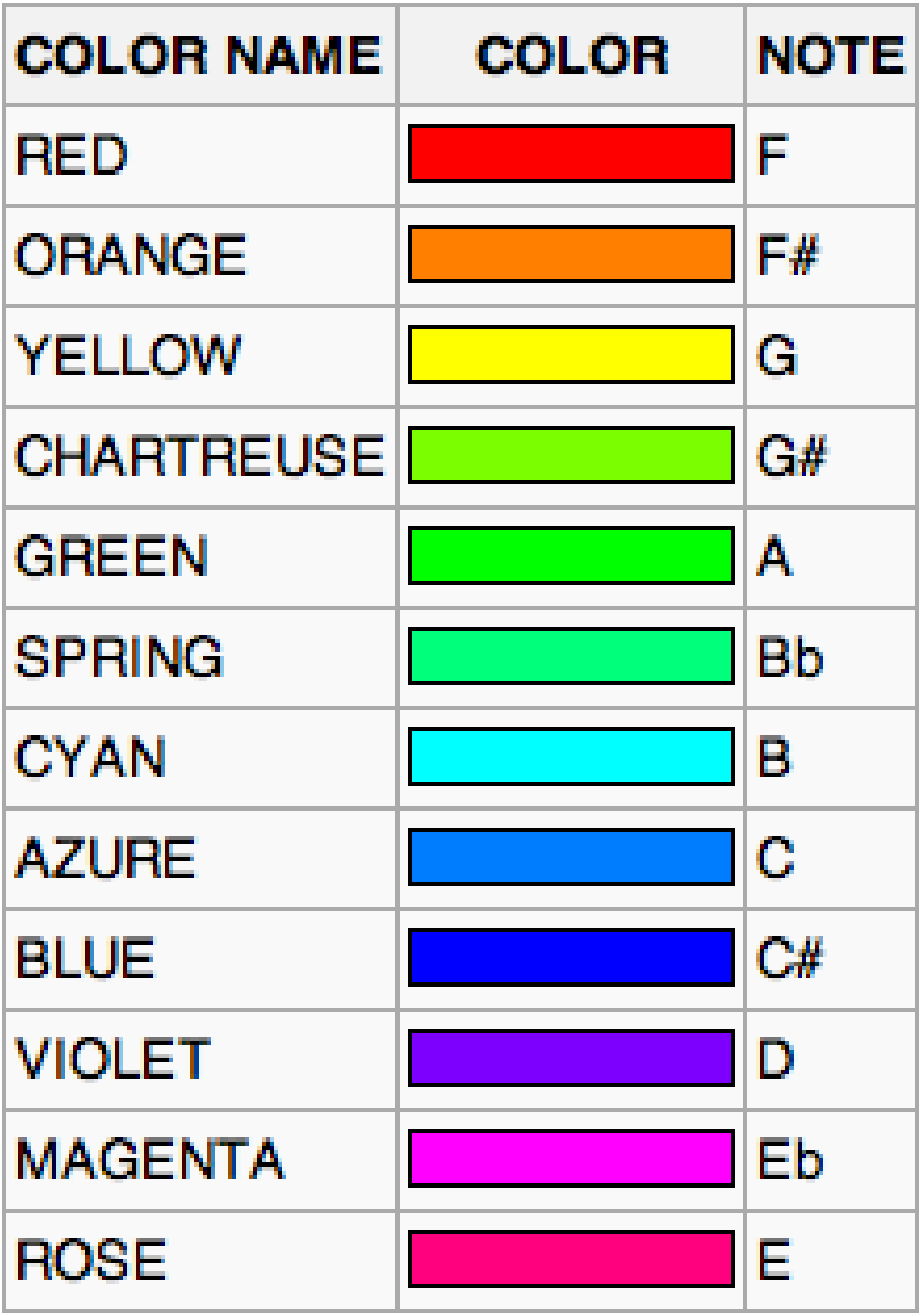

COLOR NAME	COLOR	NOTE
RED	🟥	F
ORANGE	🟧	F#
YELLOW	🟨	G
CHARTREUSE	🟩	G#
GREEN	🟩	A
SPRING	🟩	Bb
CYAN	🟦	B
AZURE	🟦	C
BLUE	🟦	C#
VIOLET	🟪	D
MAGENTA	🟪	Eb
ROSE	🟥	E

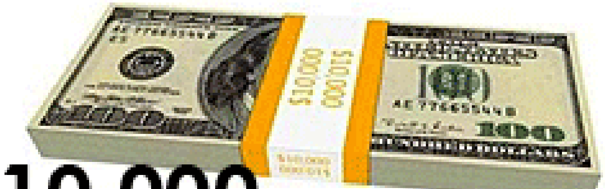

$10,000
in $100 bills

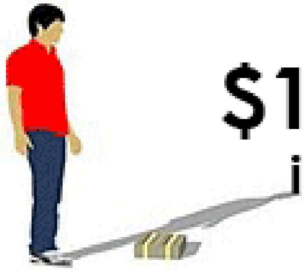

$1,000,000
in $100 bills

$100 Million
in $100 bills

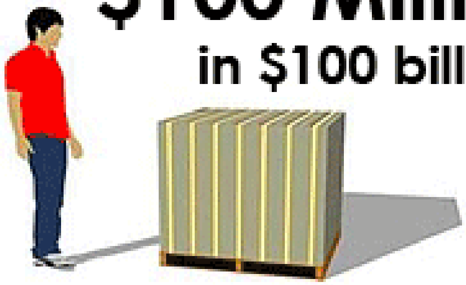

$1 Billion
in $100 bills

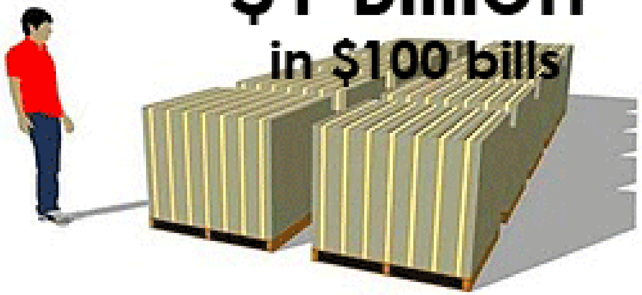

$1 Trillion
in $100 bills

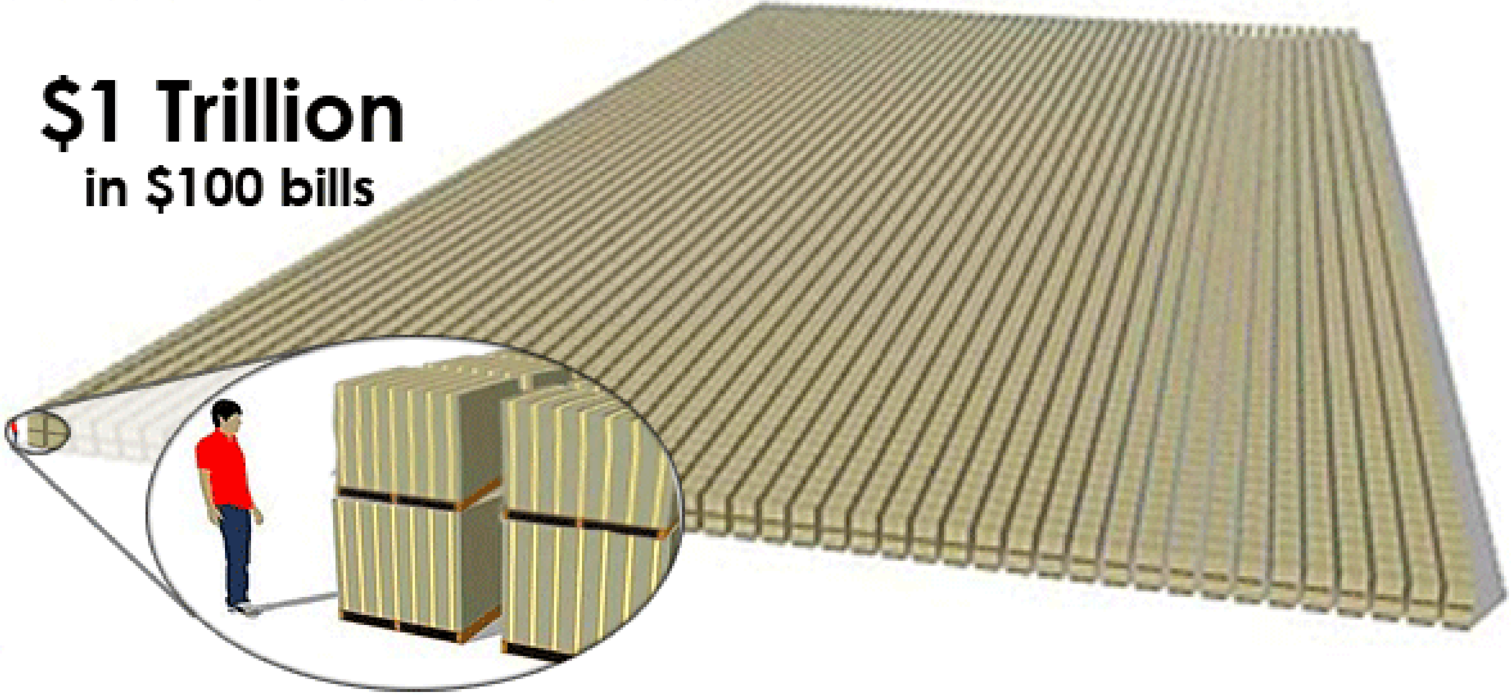

USB Type A Female Field Termination Connector

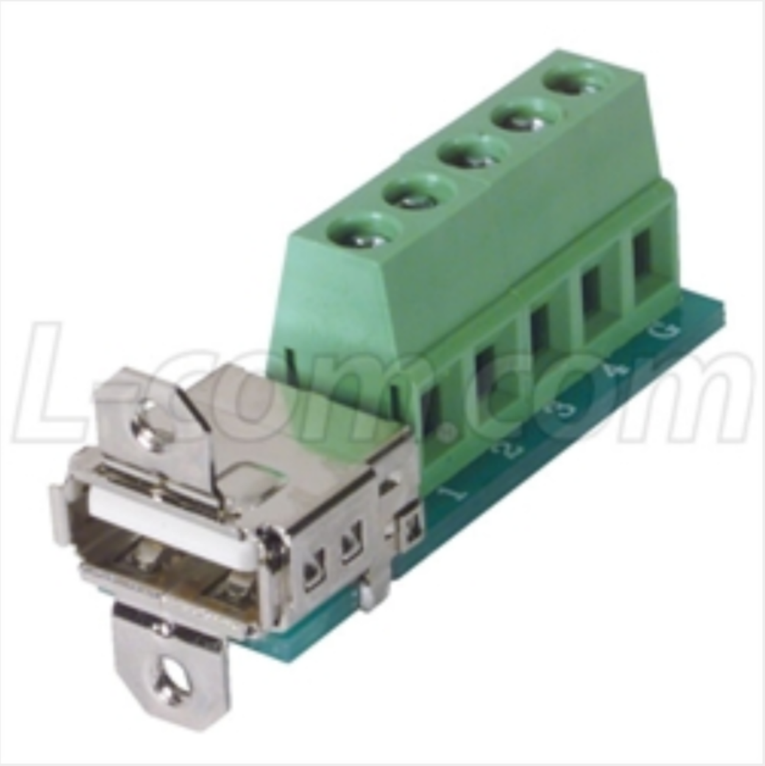

Roll over image to zoom in

L-com Item # USBAFT

List Price	$7.24
Your Price 1-9	**$7.24**
10-24	$6.81
25-99	$6.37
100 +	Call Us

Availability: In Stock

Available for Same Day Shipping

Quantity 1 ADD TO CART

Email This Page

Working Model of Crystal radio

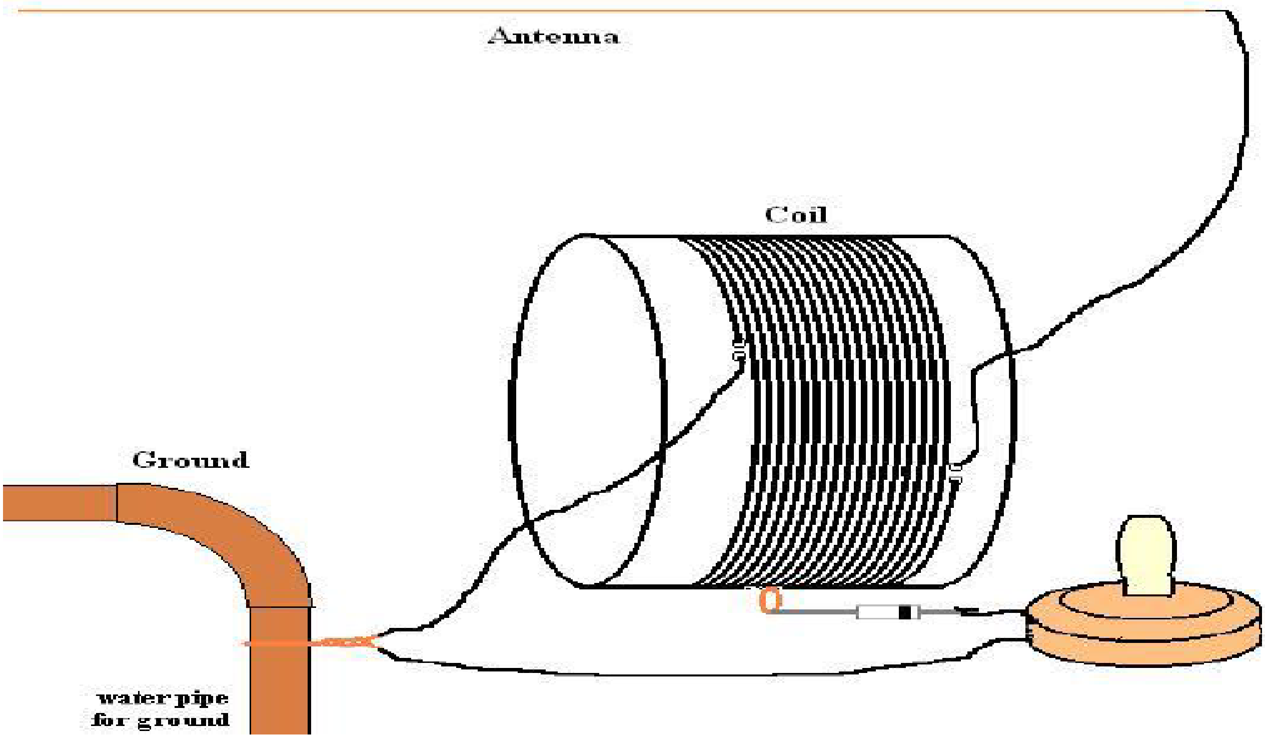
Antenna
Coil
Ground
water pipe
for ground

Schematic Diagram of the Radi

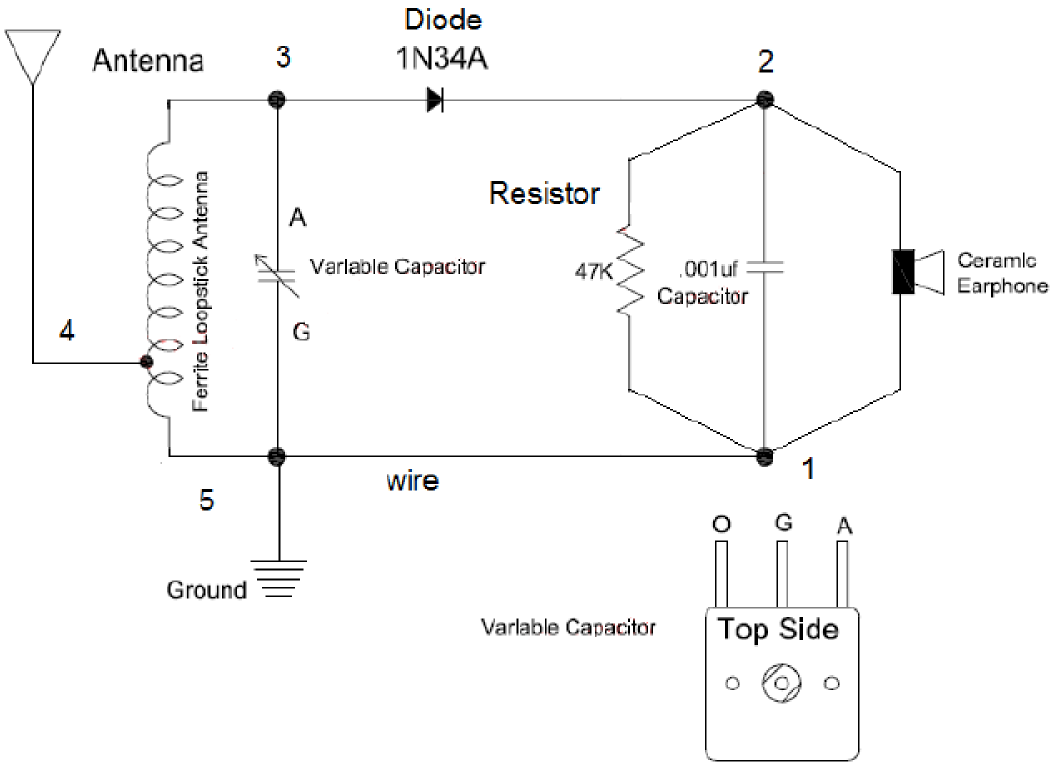
Antenna
3
Diode
1N34A
2
A
Variable Capacitor
G
Resistor
47K
.001uf
Capacitor
Ceramic
Earphone
Ferrite Loopstick Antenna
4
5
wire
1
Ground
O
G
A
Variable Capacitor
Top Side

Philmore

MANUFACTURING CO. Inc.
Richmond Hill 18, N. Y.

MODEL NO.
VC-1000

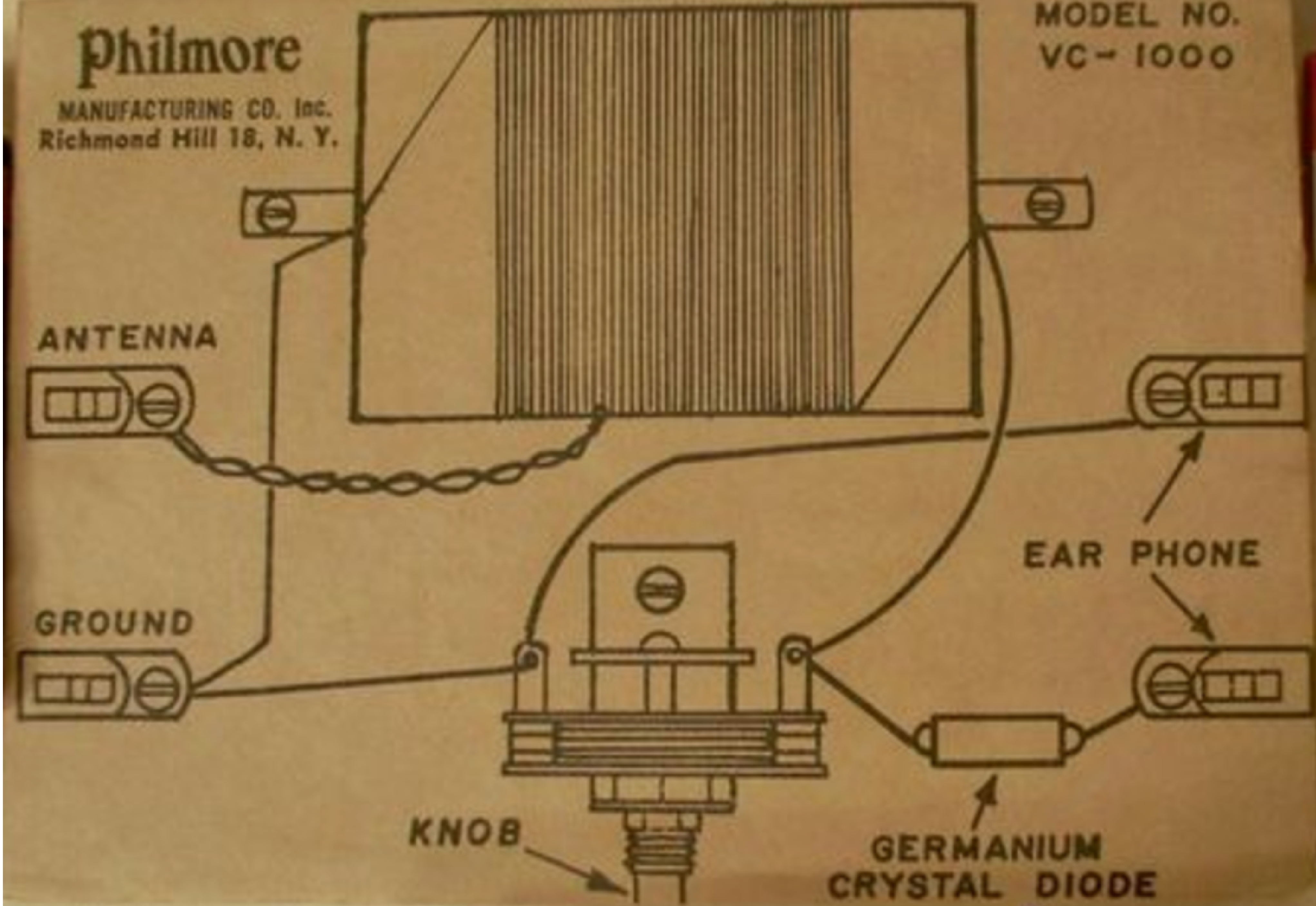

ANTENNA
GROUND
EAR PHONE
KNOB
GERMANIUM CRYSTAL DIODE

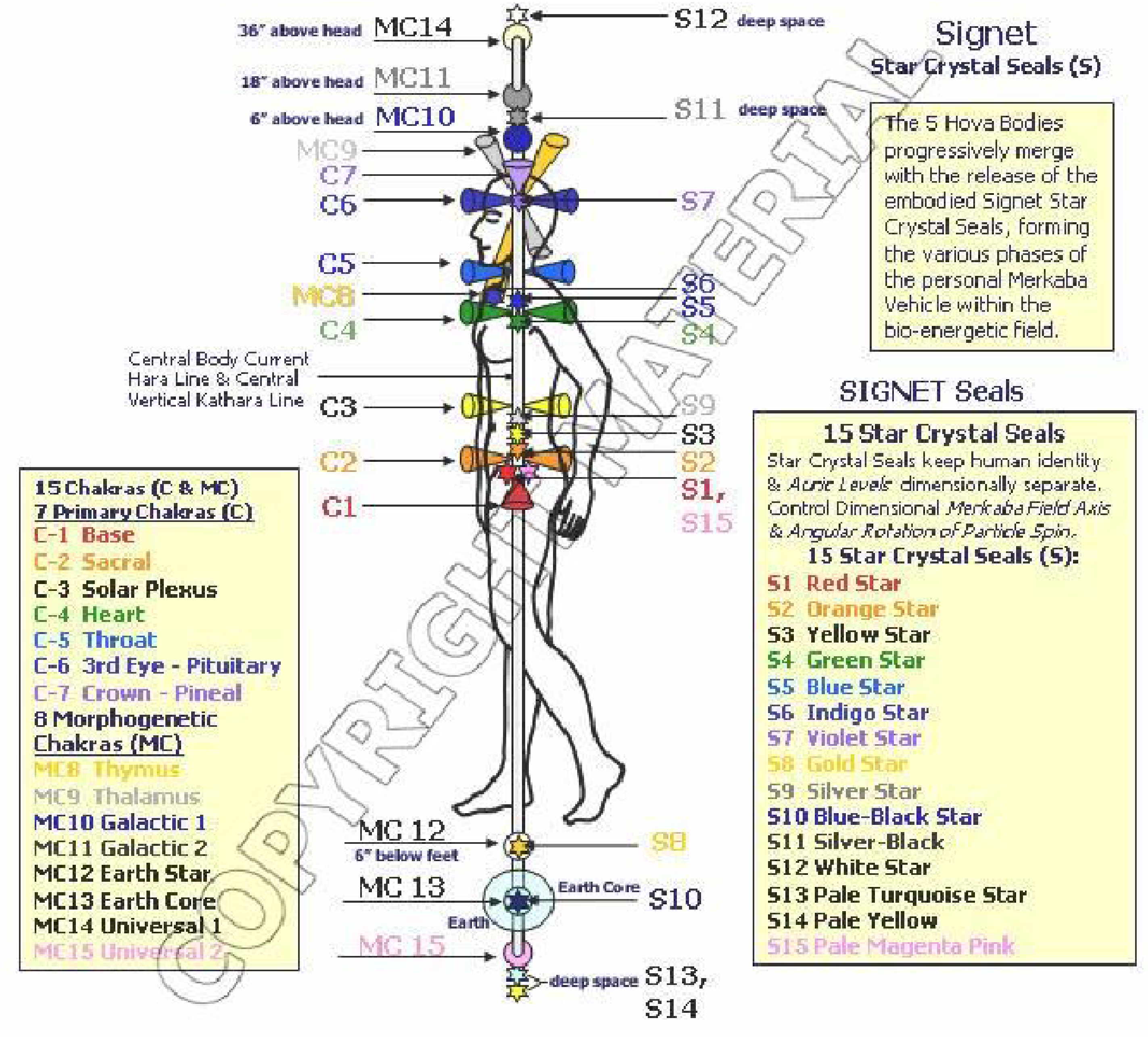
36" above head
MC14
S12 deep space
18" above head
MC11
6" above head
MC10
S11 deep space
MC9
C7
C6
S7
C5
MC8
S6
S5
C4
S4
Central Body Current Hara Line & Central Vertical Kathara Line
C3
S9
S3
C2
S2
C1
S1,
S15
MC 12
6" below feet
S8
MC 13
Earth Core
S10
Earth
MC 15
deep space
S13,
S14

Signet
Star Crystal Seals (S)

The 5 Hova Bodies progressively merge with the release of the embodied Signet Star Crystal Seals, forming the various phases of the personal Merkaba Vehicle within the bio-energetic field.

15 Chakras (C & MC)

7 Primary Chakras (C)
- C-1 Base
- C-2 Sacral
- C-3 Solar Plexus
- C-4 Heart
- C-5 Throat
- C-6 3rd Eye - Pituitary
- C-7 Crown - Pineal

8 Morphogenetic Chakras (MC)
- MC8 Thymus
- MC9 Thalamus
- MC10 Galactic 1
- MC11 Galactic 2
- MC12 Earth Star
- MC13 Earth Core
- MC14 Universal 1
- MC15 Universal 2

SIGNET Seals

15 Star Crystal Seals

Star Crystal Seals keep human identity & *Auric Levels* dimensionally separate. Control Dimensional *Merkaba Field Axis* & *Angular Rotation of Particle Spin.*

15 Star Crystal Seals (S):
- S1 Red Star
- S2 Orange Star
- S3 Yellow Star
- S4 Green Star
- S5 Blue Star
- S6 Indigo Star
- S7 Violet Star
- S8 Gold Star
- S9 Silver Star
- S10 Blue-Black Star
- S11 Silver-Black
- S12 White Star
- S13 Pale Turquoise Star
- S14 Pale Yellow
- S15 Pale Magenta Pink

COPYRIGHT MATERIAL

Copyright A'shayana & A'hzayana Deane, 1999-2007, All Rights Reserved; Part of the MCEO Freedom Teachings™ Series

Scalar Interferometry In The Endothermic (Cooling) Mode

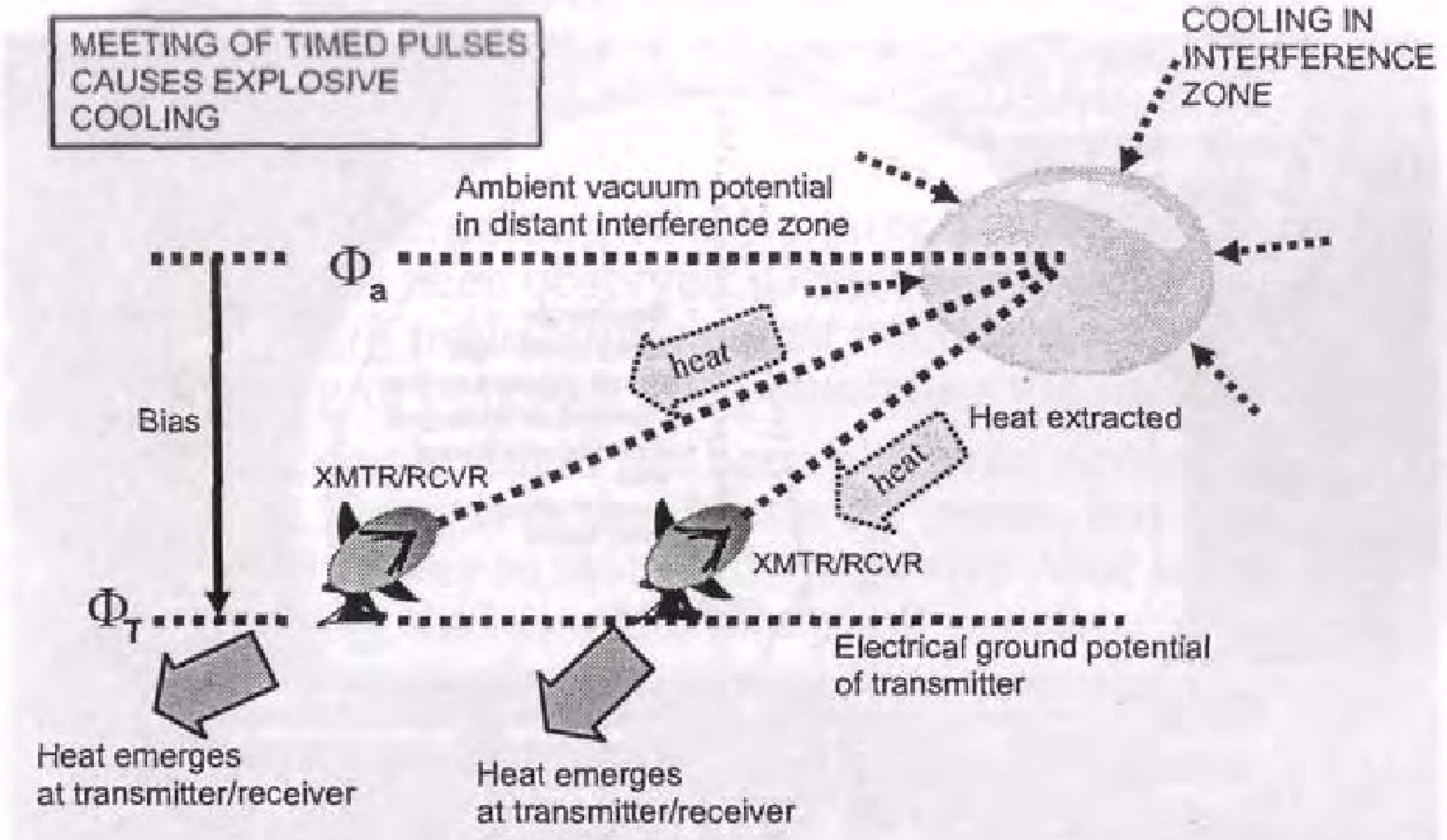
MEETING OF TIMED PULSES CAUSES EXPLOSIVE COOLING
COOLING IN INTERFERENCE ZONE
Ambient vacuum potential in distant interference zone
Φ_a
heat
Bias
Heat extracted
heat
XMTR/RCVR
XMTR/RCVR
Φ_T
Electrical ground potential of transmitter
Heat emerges at transmitter/receiver
Heat emerges at transmitter/receiver

In the endothermic mode, tremendous heat is received back in the transmitters. Soviets further retransmit the heat via interferometers to a distant "exhaust" point. Bennett Island was used for such an exhaust point for decades, and our observation platforms saw a great many 150 miles-long jets of exhaust effects in narrow beams at shallow angles.

Measuring Longitudinal and Transverse waves

The Fizzics Organisation

Transverse waves
A shaking rope is a good example. The wave moves along the rope but the bits of rope go up and down - across the direction of the wave. “Trans” means across, think of “transatlantic” or “transfer”.

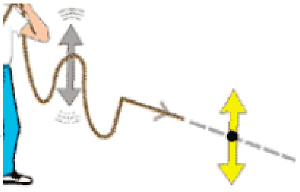

A water wave is another example of a transverse wave.

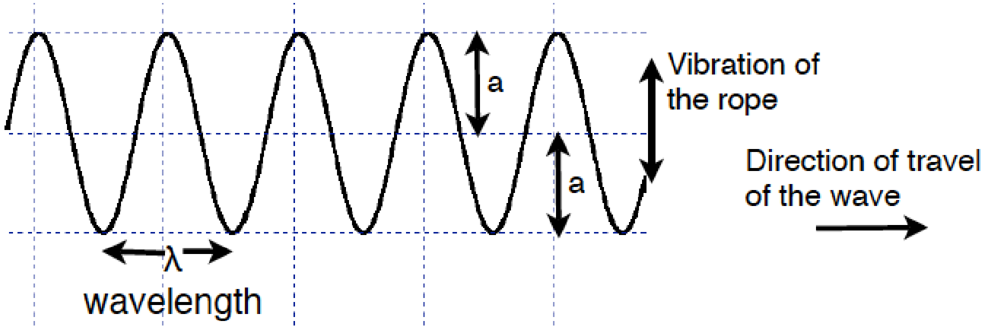

a is the amplitude of the wave measured from the maximum height to the middle. Think of a water wave, it is the distance from either the crest or the trough to where the water would be if it were calm.

λ is the wavelength. That is the distance from one point on a wave to the same point on the next wave - for example from crest to crest or trough to trough.
f is the frequency of the wave, that is the number of waves which are produced every second (or the number passing every second). This number is given the unit “hertz”. One hertz is one wave per second.

Most of the examples of waves we have to deal with are transverse waves. All the waves in the electromagnet spectrum are transverse waves.

There is a connection between the frequency and wavelength of a wave. If the frequency is increased - if there are far more waves, then they are closer together; in other words the wavelength is less.
The speed is equal to the frequency multiplied by the

$$V = f \times \lambda$$

m/s hertz metres

A longitudinal wave is a “pushing” wave. A line of dominos falling is an example.

If one end of a slinky is pushed and pulled then each coil pushes and pulls the next. The coils oscillate along in the same direction that the wave is traveling.

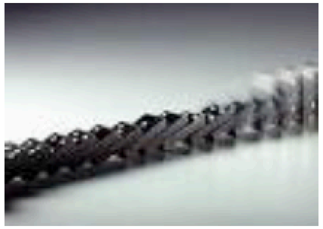

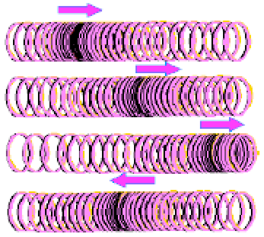

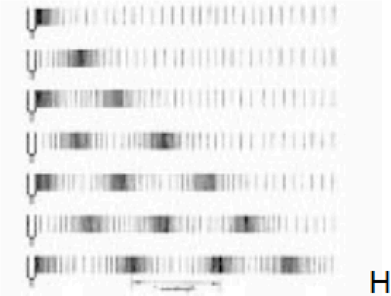

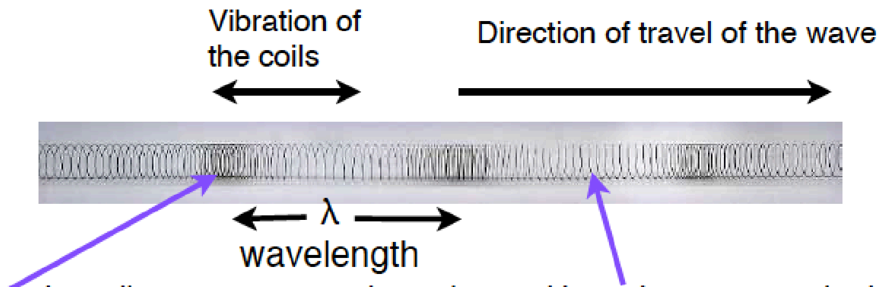

Here the coils are compressed together and here they are stretched out.

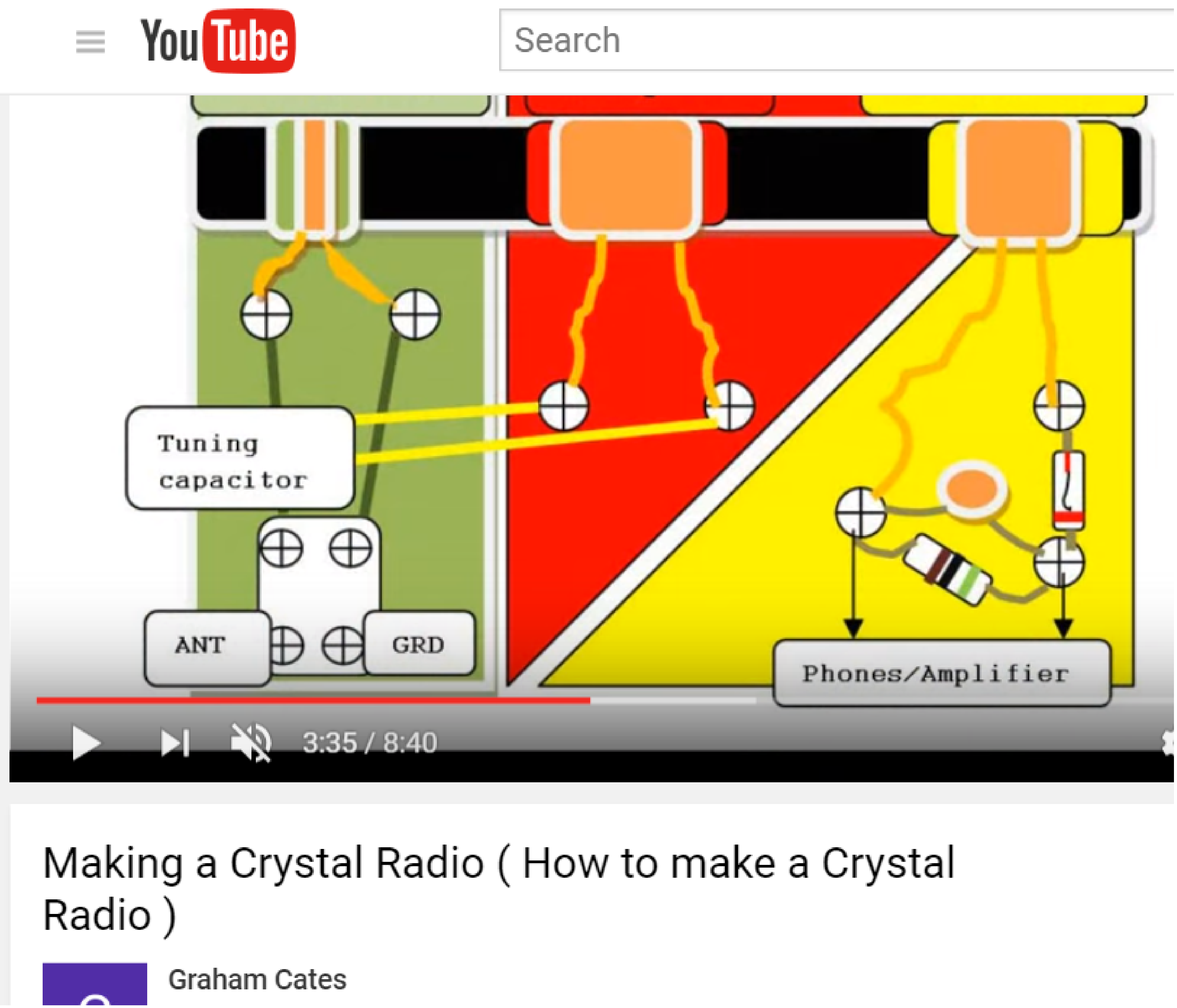
YouTube
Search
Tuning capacitor
ANT
GRD
Phones/Amplifier
3:35 / 8:40

Making a Crystal Radio (How to make a Crystal Radio)

Graham Cates

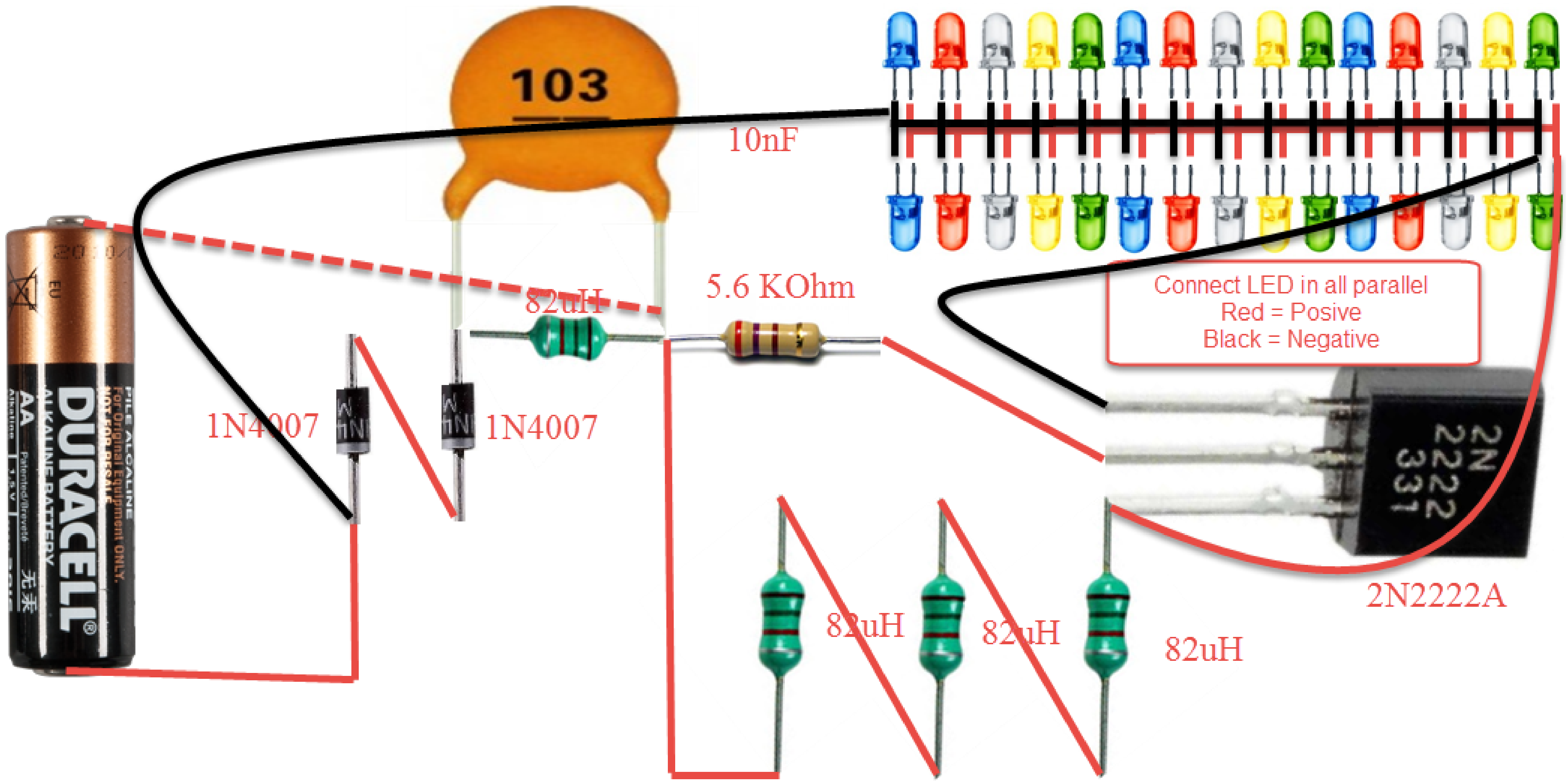
103
10nF
82uH
5.6 KOhm
1N4007
1N4007
DURACELL
82uH
82uH
82uH
2N2222A

Connect LED in all parallel
Red = Posive
Black = Negative

The joule thief circuit.

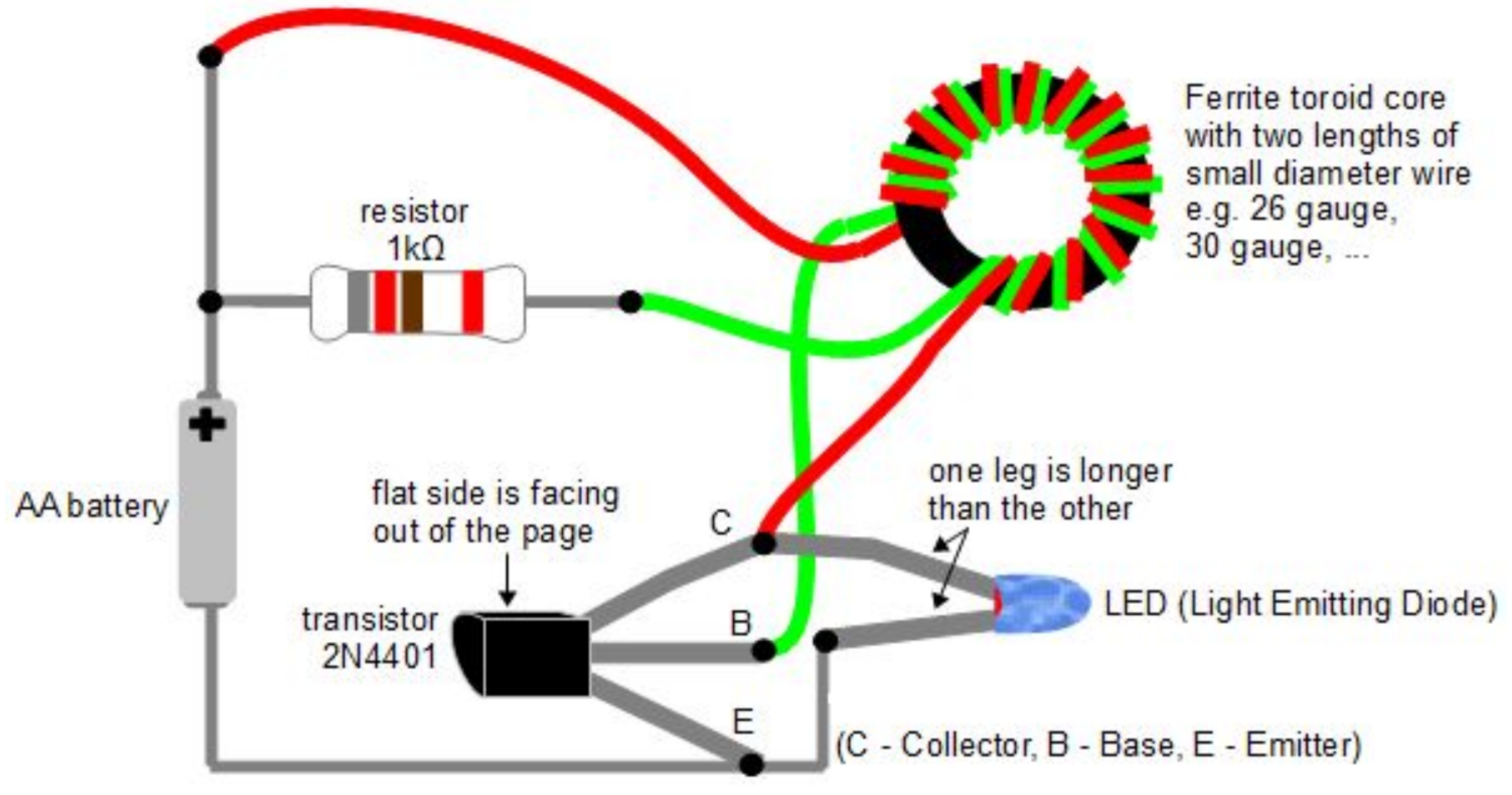
Ferrite toroid core
with two lengths of
small diameter wire
e.g. 26 gauge,
30 gauge, ...
resistor
1kΩ
AA battery
flat side is facing
out of the page
transistor
2N4401
C
B
E
one leg is longer
than the other
LED (Light Emitting Diode)
(C - Collector, B - Base, E - Emitter)

Transistor - The legs of the transistor can be determined by noticing that there's a flat side to the transistor case. See the diagram above. A large number of transistors have been reported to work: 2N4401, NET123AP, BC547B, 2SC2500, BC337, PN2222, to name just a few.

LED - One leg of the LED is longer than the other leg. Use this to determine which one goes where. See the diagram above.

Resistor - The diagram says use a 1 kilo ohm resistor but I've used an 820 ohm one just fine. I've also seen a 2 kilo ohm one in use. Use whatever works for you. You can also use a potentiometer (a variable resistor) so that you can easily adjust it to select the resistance that gives the best light.

Toroid ferrite core - Some people have gotten these by opening up compact fluorescent lightbulbs (CFLs). I took mine out of some device whose original function I don't know. To get it working, my first one had just 13 turns for each wire and I used a 30 gauge wire and a 26 gauge wire. The wire must be insulated. A variety of number of turns will work. This is something you can play with. Look at the diagram carefully to determine where the wires connect to.

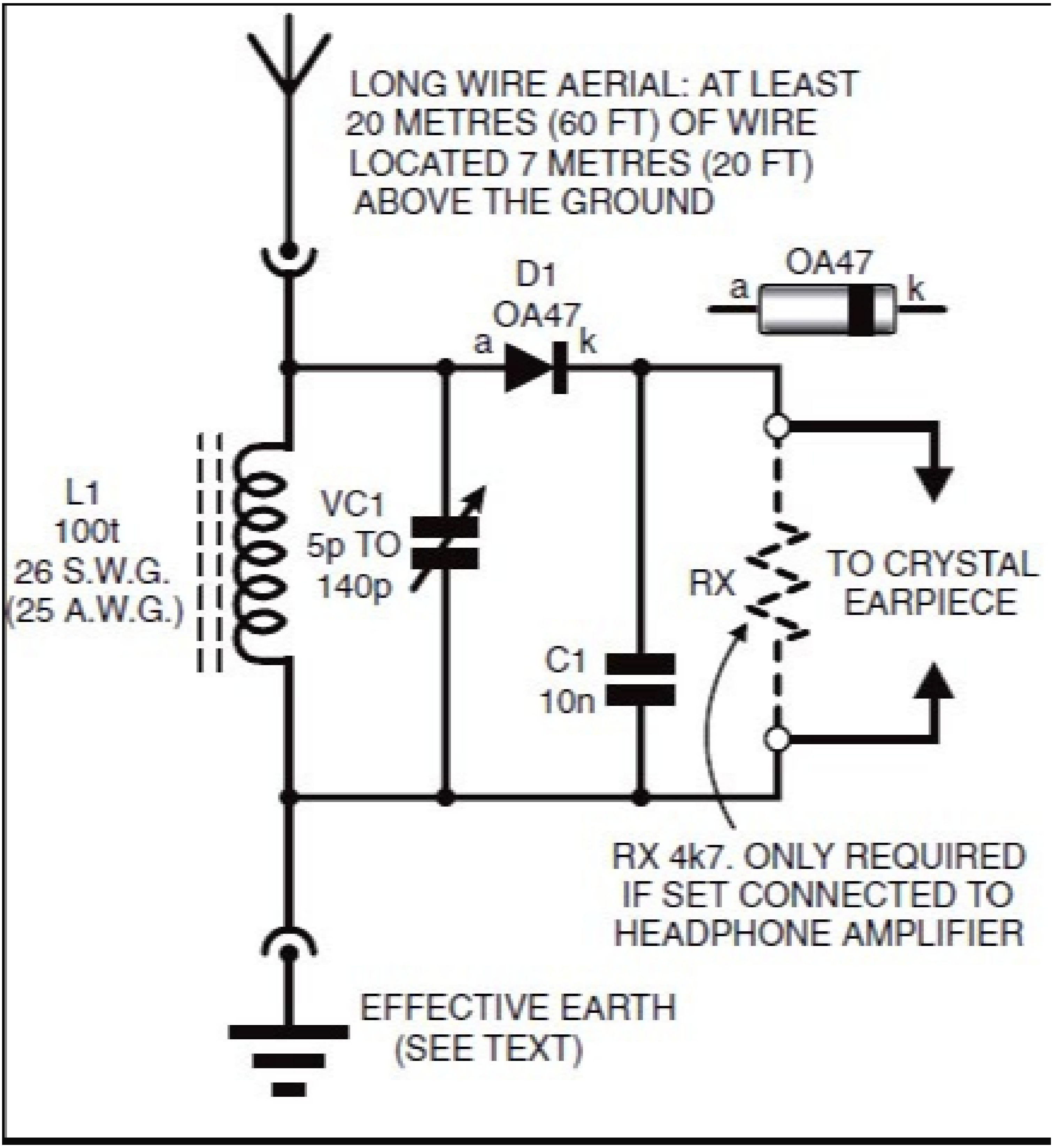

LONG WIRE AERIAL: AT LEAST
20 METRES (60 FT) OF WIRE
LOCATED 7 METRES (20 FT)
ABOVE THE GROUND
D1
OA47
a
k
OA47
a
k
L1
100t
26 S.W.G.
(25 A.W.G.)
VC1
5p TO
140p
C1
10n
RX
TO CRYSTAL
EARPIECE
RX 4k7. ONLY REQUIRED
IF SET CONNECTED TO
HEADPHONE AMPLIFIER
EFFECTIVE EARTH
(SEE TEXT)

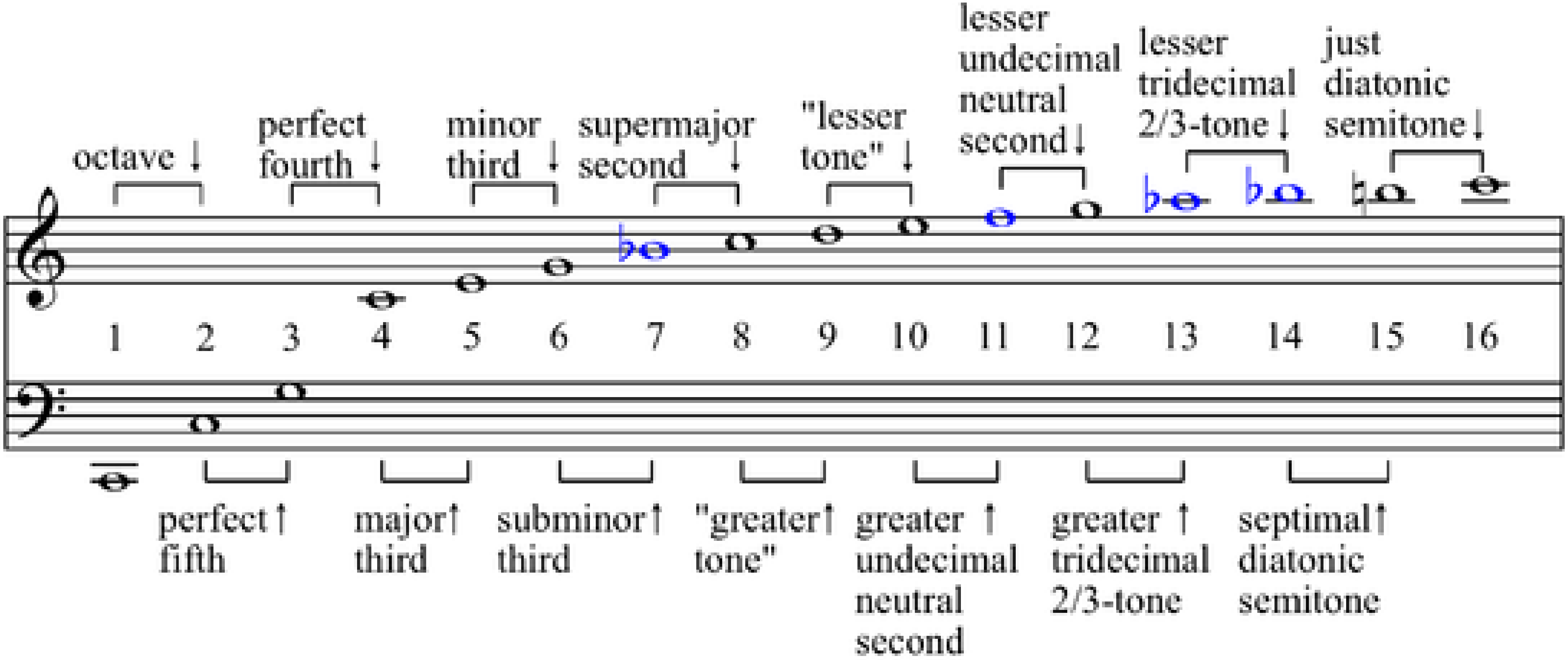
octave ↓
perfect fourth ↓
minor third ↓
supermajor second ↓
"lesser tone" ↓
lesser undecimal neutral second↓
lesser tridecimal 2/3-tone↓
just diatonic semitone↓
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
perfect↑ fifth
major↑ third
subminor↑ third
"greater↑ tone"
greater ↑ undecimal neutral second
greater ↑ tridecimal 2/3-tone
septimal↑ diatonic semitone

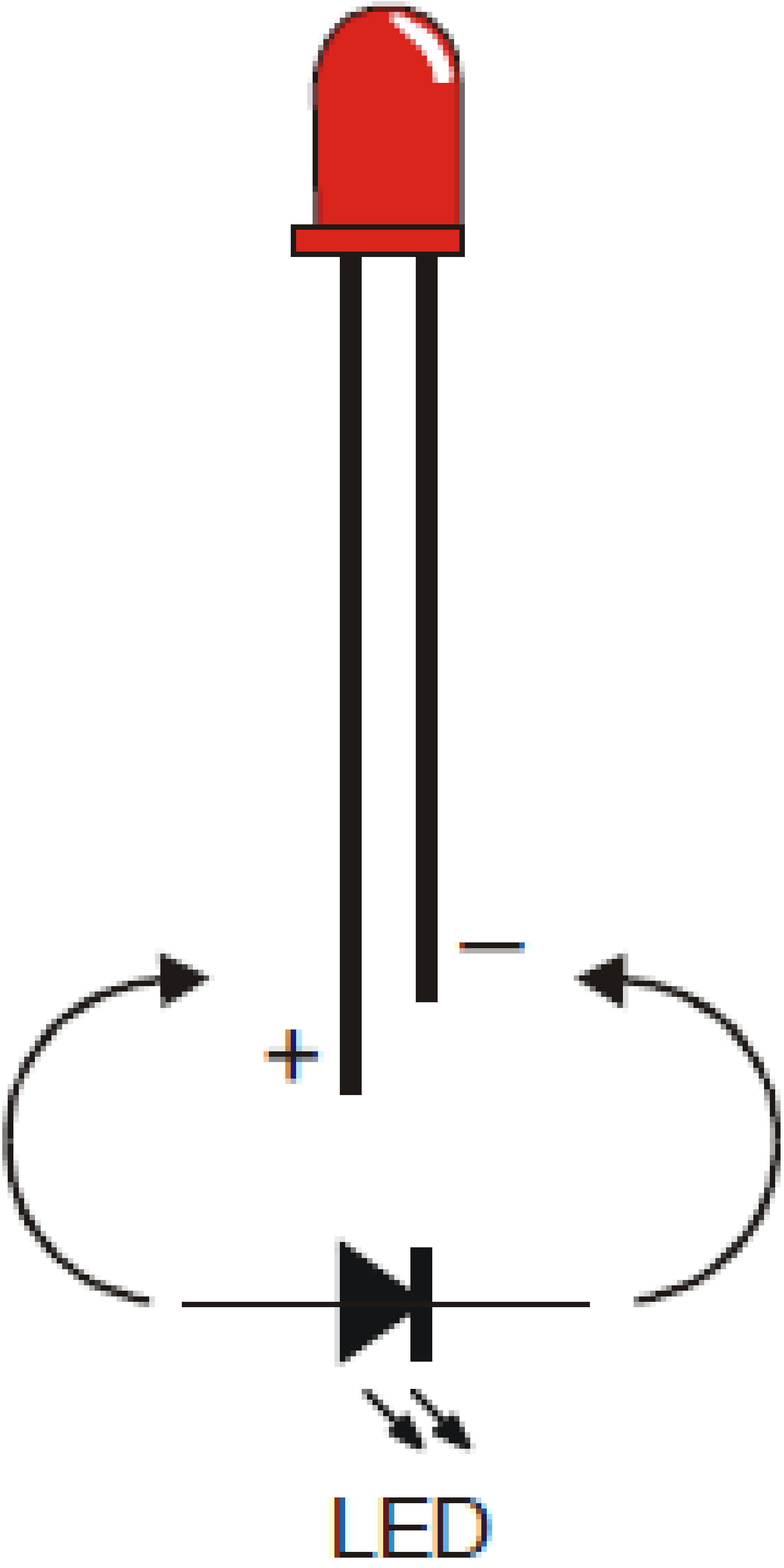
−
+
LED

The Tesla Experiment

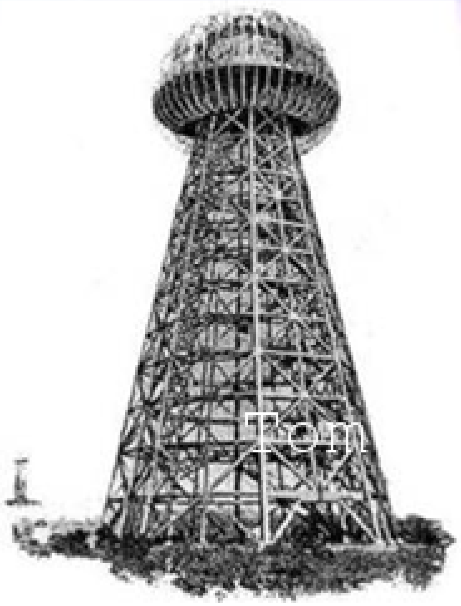

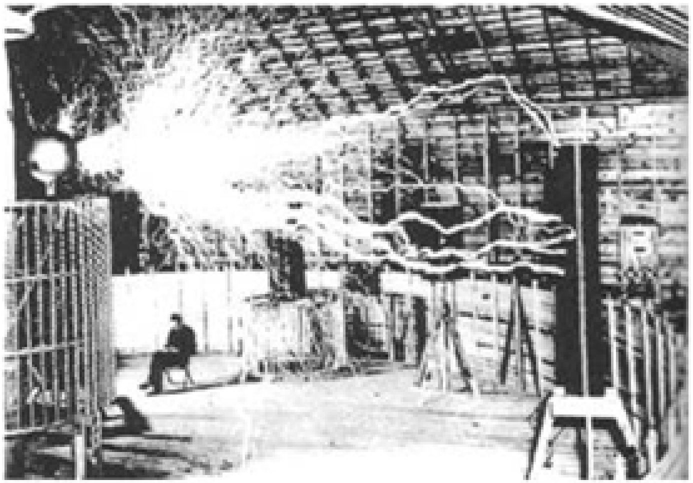

TRANSVERSE WAVE INTERFERENCE

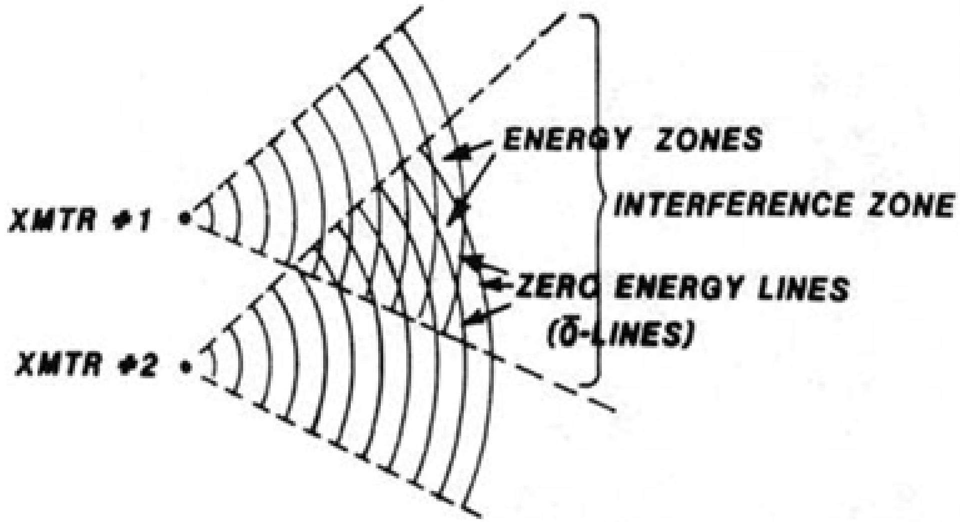
ENERGY ZONES
INTERFERENCE ZONE
XMTR #1
ZERO ENERGY LINES
(0̄-LINES)
XMTR #2

LONGITUDINAL WAVE INTERFERENCE
(SCALAR WAVES)

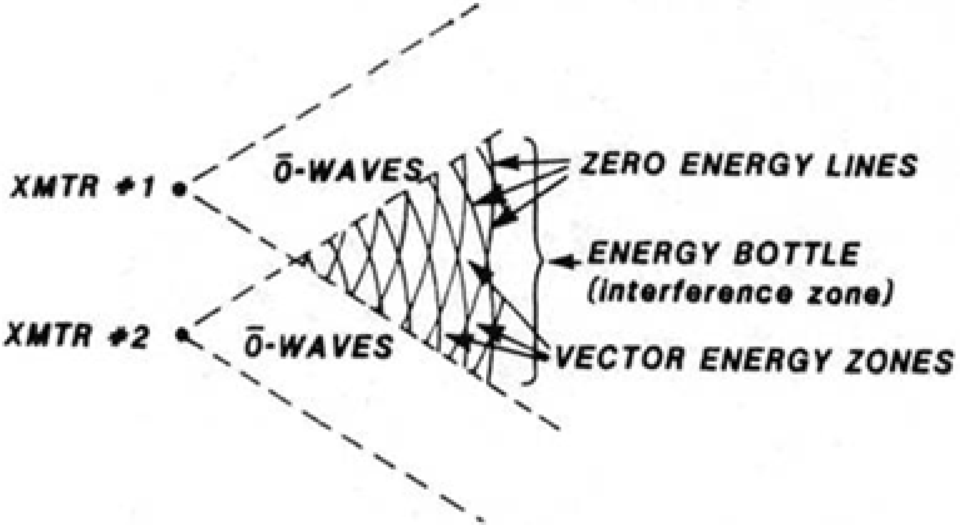
XMTR #1
0̄-WAVES
ZERO ENERGY LINES
ENERGY BOTTLE
(interference zone)
XMTR #2
0̄-WAVES
VECTOR ENERGY ZONES

© PoweredTemplates.com

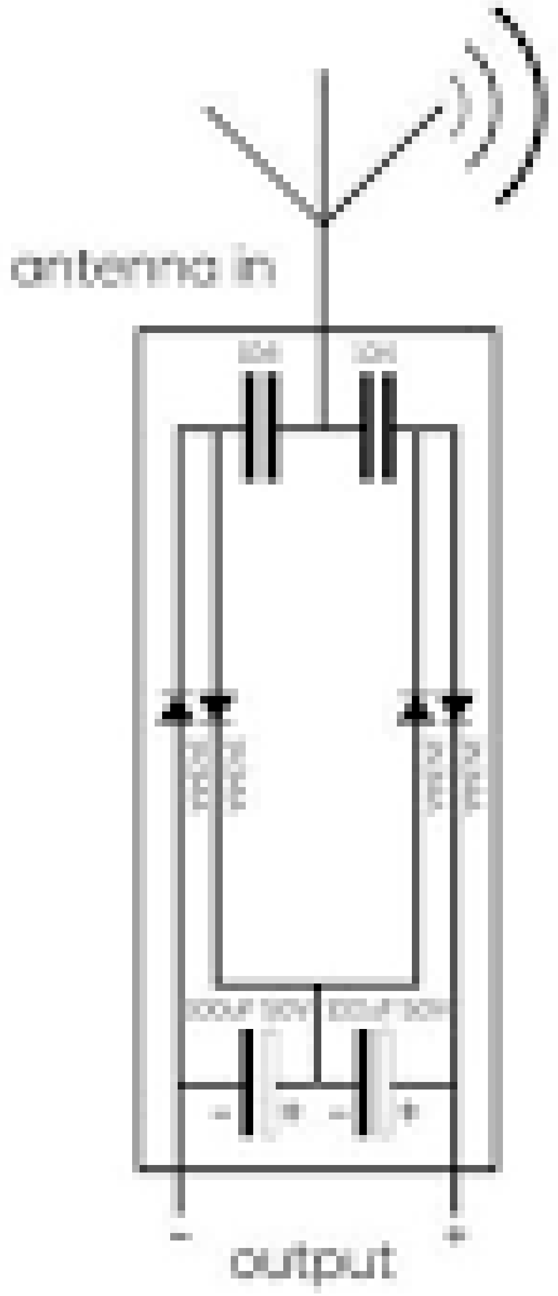
antenna in
104
104
100uF 50V
100uF 50V
-
+
-
+
-
output
+

RF to DC

Circuit Diagam (simplified)

Component list:
(2) Ceramic Capacitors (104)
(4) Germanium Diodes (1N34A)
(2) Electrolytic Capacitors (100uF 50V)

Copyright 2008. Drew Paul Designs. All rights reserved.

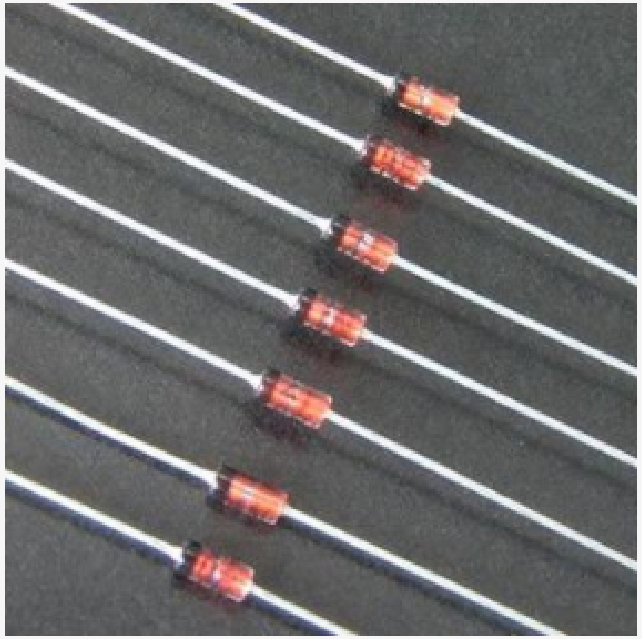

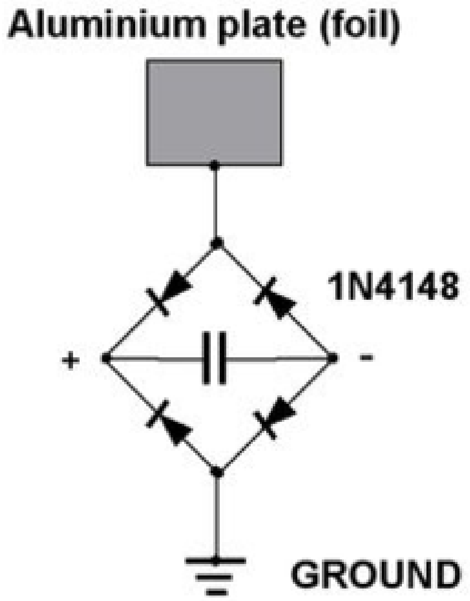
Aluminium plate (foil)
1N4148
+
-
GROUND

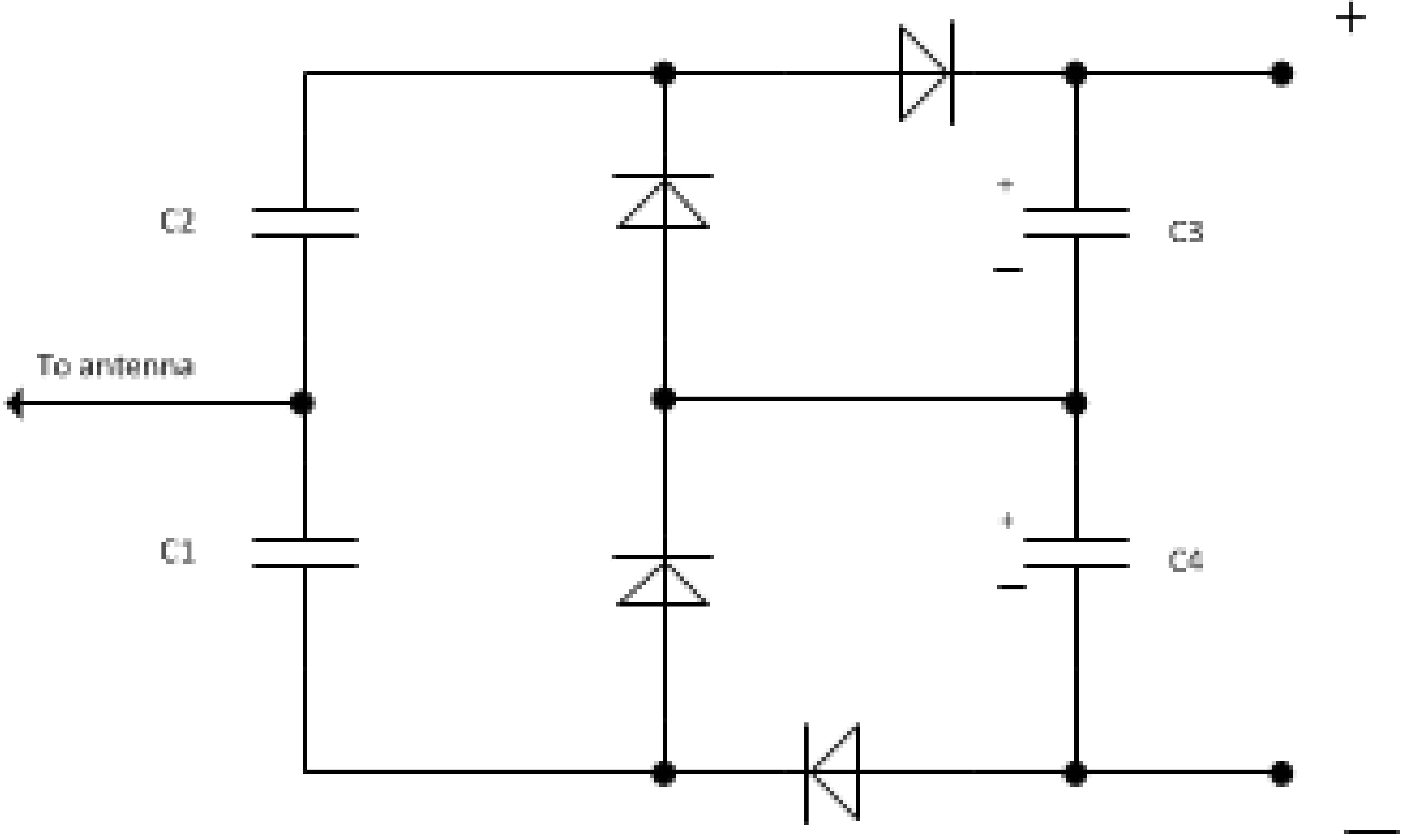
+
C2
C3
To antenna
C1
C4
+
–
+
–
–

Silent Sound Subliminal Mind Control

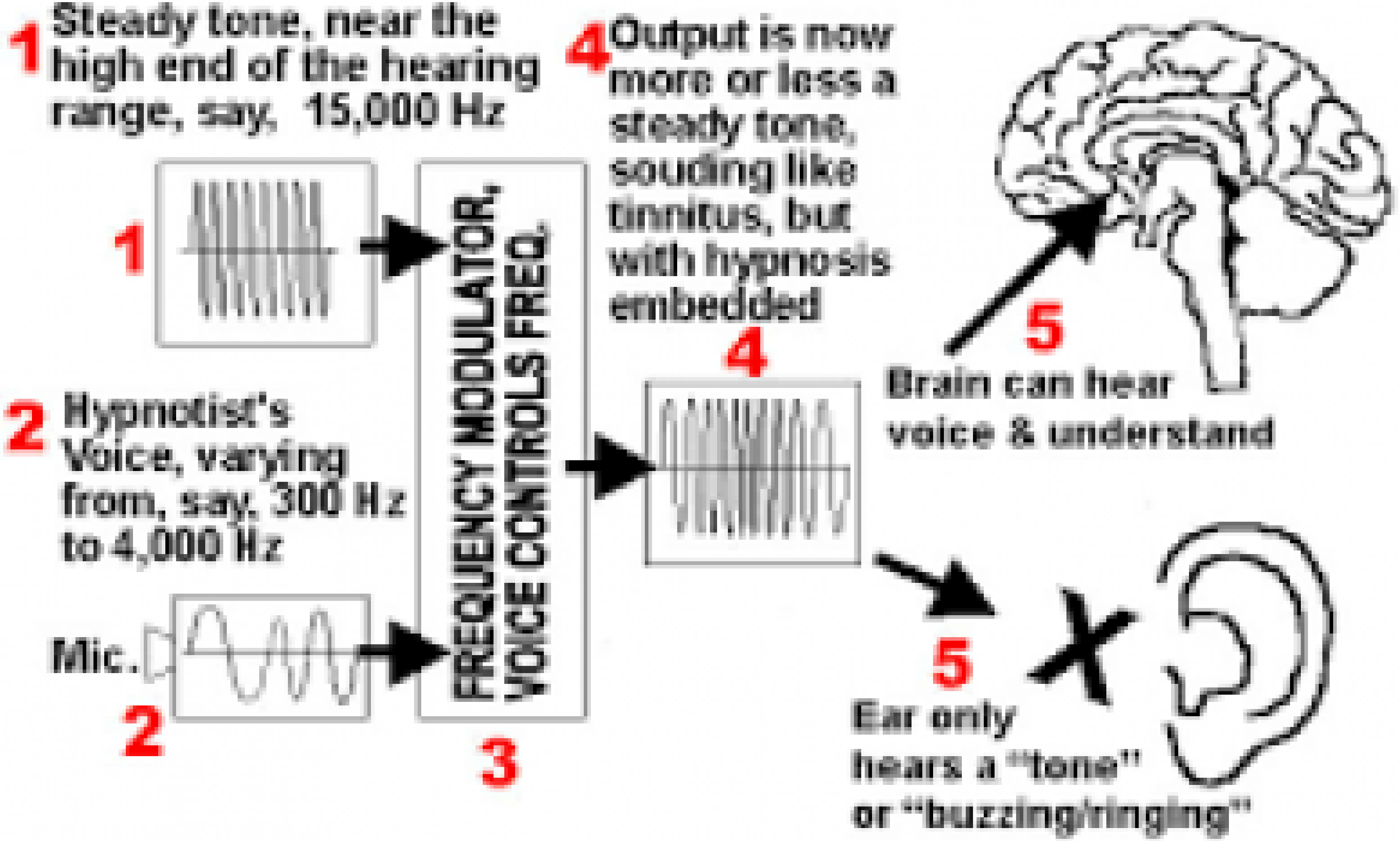

1 Steady tone, near the high end of the hearing range, say, 15,000 Hz

2 Hypnotist's Voice, varying from, say, 300 Hz to 4,000 Hz

4 Output is now more or less a steady tone, souding like tinnitus, but with hypnosis embedded

5 Brain can hear voice & understand

5 Ear only hears a "tone" or "buzzing/ringing"

1
Mic.
2
FREQUENCY MODULATOR,
VOICE CONTROLS FREQ.
3
4
5
5

Output can be via open air broadcast or piggybacked on TV/radio signals

U.S. Patent 5159703 issued Oct 27, 1992

Table 1: Dinshah's Sound/Color Equivalent Chart.

Color	Frequency	Musical Note	Note Frequency
Red	397.5	G	378 (3 x 126 = 378)
Orange	430.5	A	432 (3 x 144 = 432)
Yellow	464.5	A#	459 (3 x 153 = 459)
Lemon	498	B	486 (3 x 162 = 486)
Green	531.5	C	513 (3 x 171 = 513)
Turquoise	565	C#	540 (3 x 180 = 540)
Blue	598.5	D	567 (3 x 189 = 567)
Indigo	632	D#	594 (3 x 198 = 594)
Violet	665.5	E	648 (3 x 216 = 648)
Purple	565*	A# and E	562*
Magenta	531.5*	G and E	525*
Scarlet	498*	G# and D	501*

---

*denotes reverse polarity. From:

Dinshah GP. Spectralchrometry Encyclopedia, 3rd Ed.

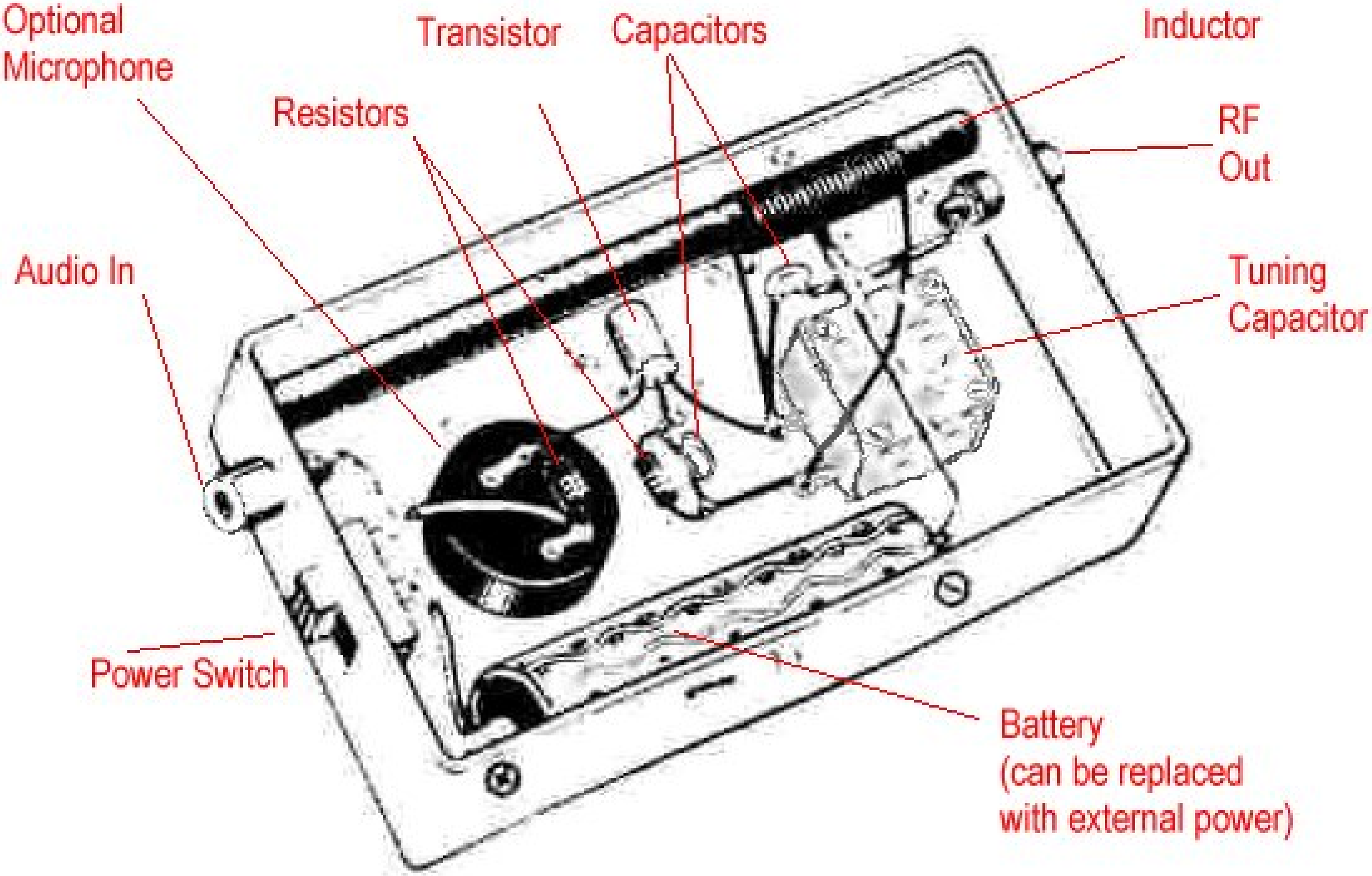
Optional Microphone
Transistor
Capacitors
Inductor
Resistors
RF Out
Audio In
Tuning Capacitor
Power Switch
Battery (can be replaced with external power)

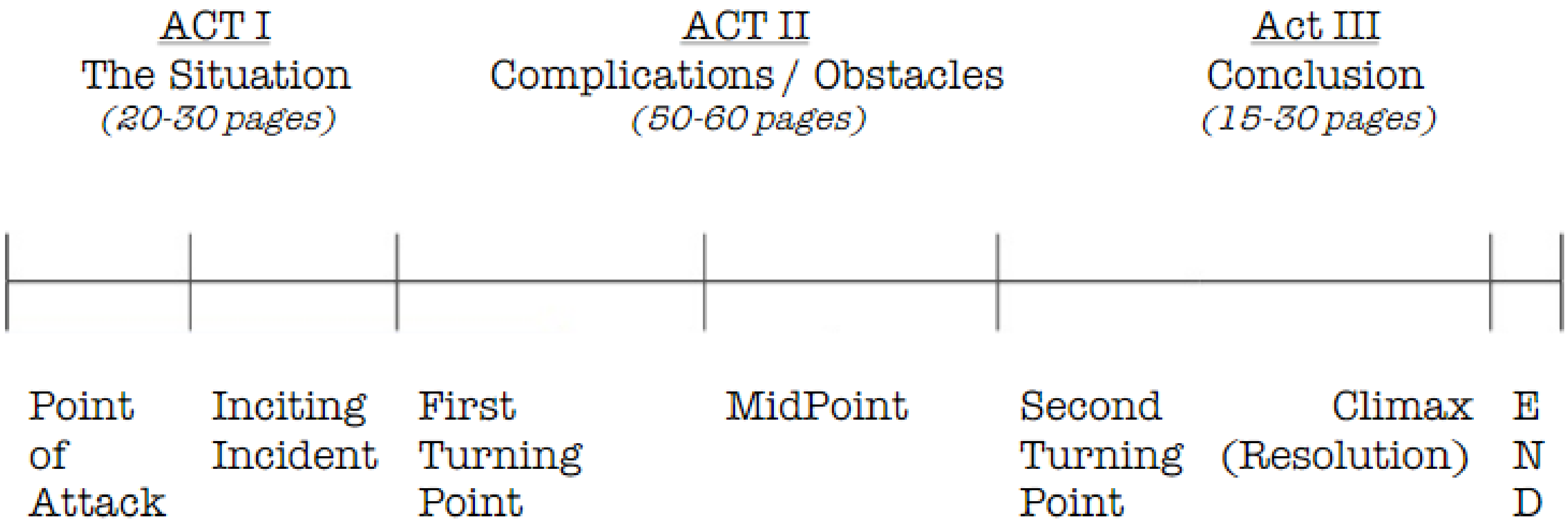
ACT I
The Situation
(20-30 pages)
ACT II
Complications / Obstacles
(50-60 pages)
Act III
Conclusion
(15-30 pages)
Point of Attack
Inciting Incident
First Turning Point
MidPoint
Second Turning Point
Climax (Resolution)
E N D

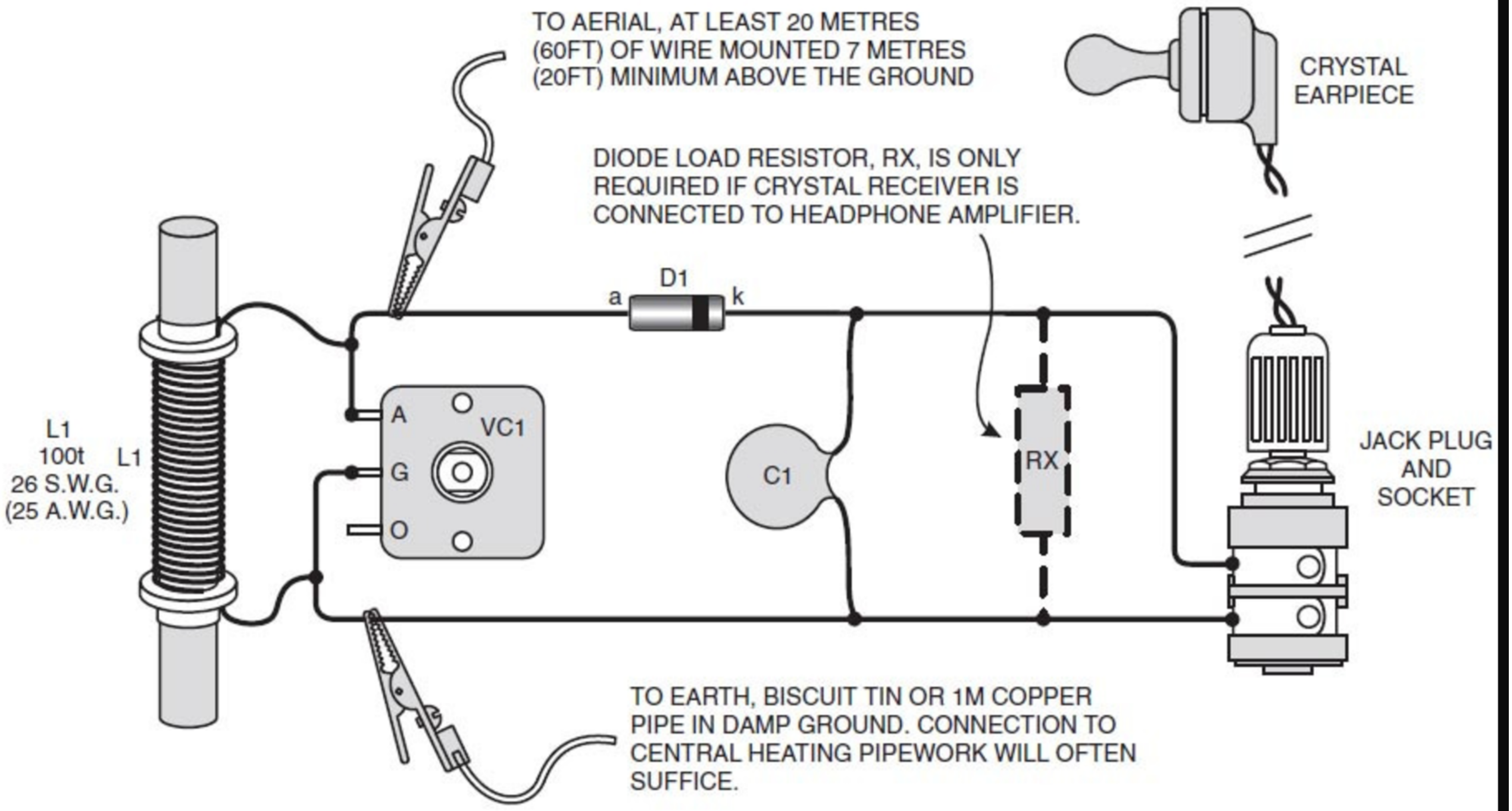

TO AERIAL, AT LEAST 20 METRES (60FT) OF WIRE MOUNTED 7 METRES (20FT) MINIMUM ABOVE THE GROUND

DIODE LOAD RESISTOR, RX, IS ONLY REQUIRED IF CRYSTAL RECEIVER IS CONNECTED TO HEADPHONE AMPLIFIER.

TO EARTH, BISCUIT TIN OR 1M COPPER PIPE IN DAMP GROUND. CONNECTION TO CENTRAL HEATING PIPEWORK WILL OFTEN SUFFICE.

CRYSTAL EARPIECE
D1
a
k
L1
100t
26 S.W.G.
(25 A.W.G.)
L1
A
G
O
VC1
C1
RX
JACK PLUG AND SOCKET

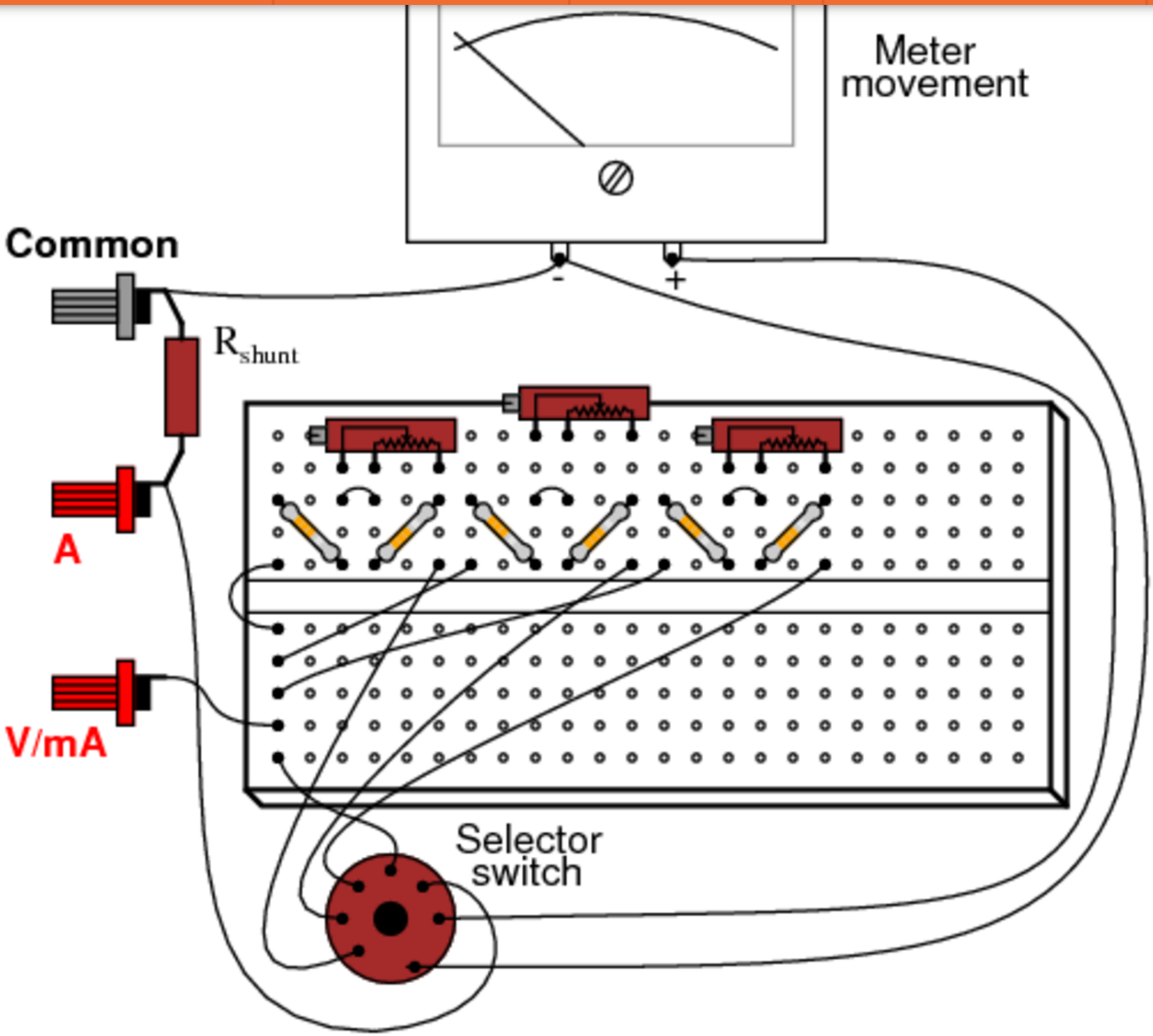
Meter
movement
Common
-
+
R_{shunt}
A
V/mA
Selector
switch

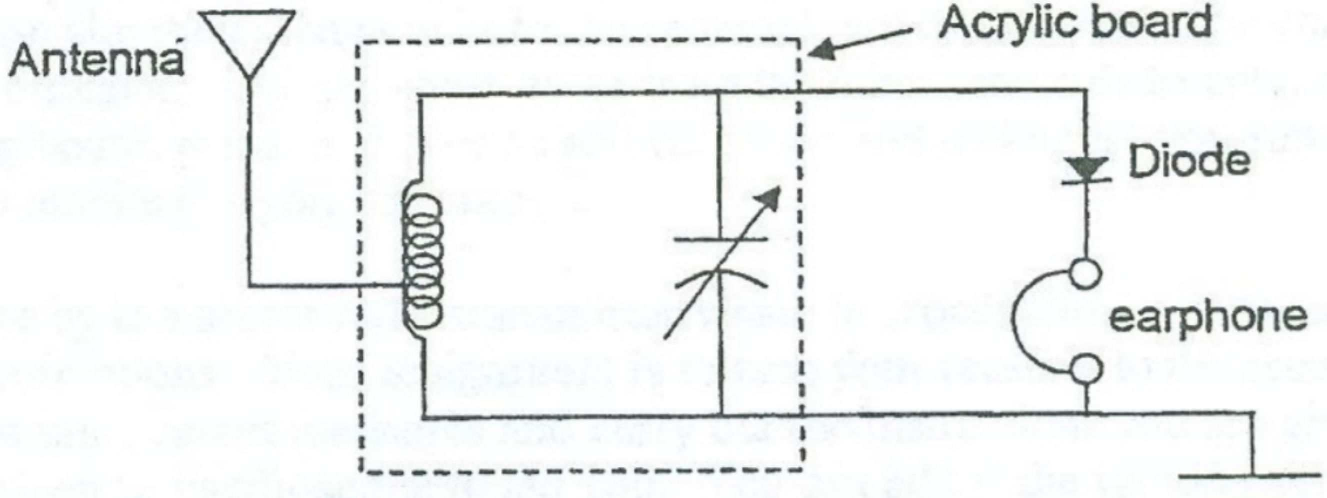
Antenna
Acrylic board
Diode
earphone

Schematic of basic radio receiver circuit

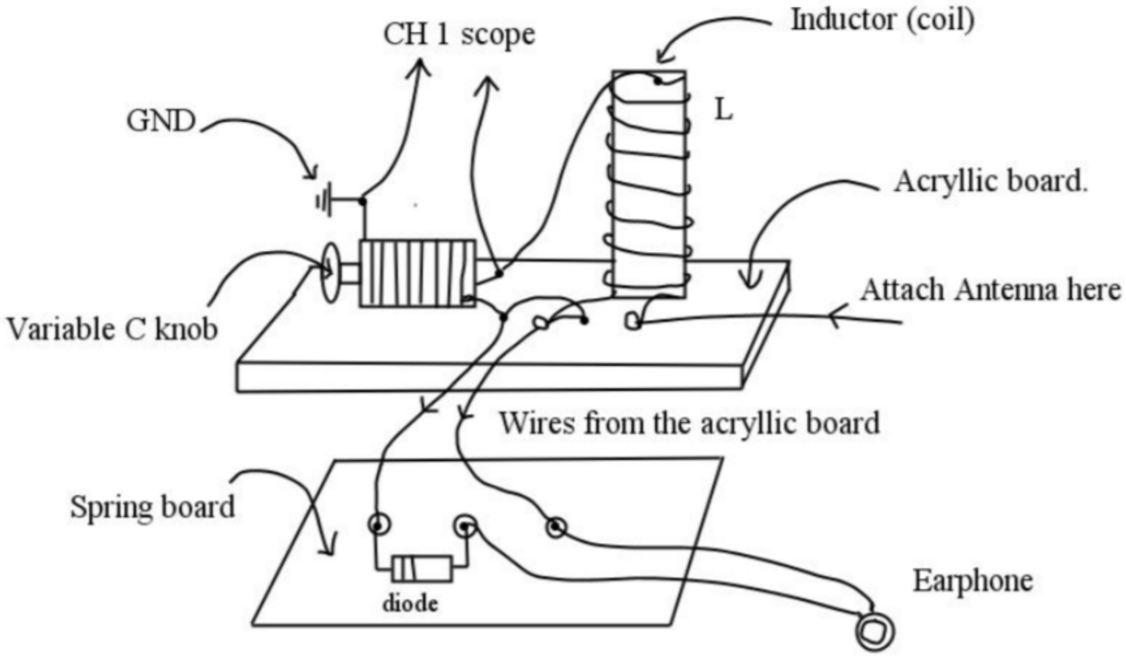
CH 1 scope
Inductor (coil)
L
GND
Acryllic board.
Attach Antenna here
Variable C knob
Wires from the acryllic board
Spring board
diode
Earphone

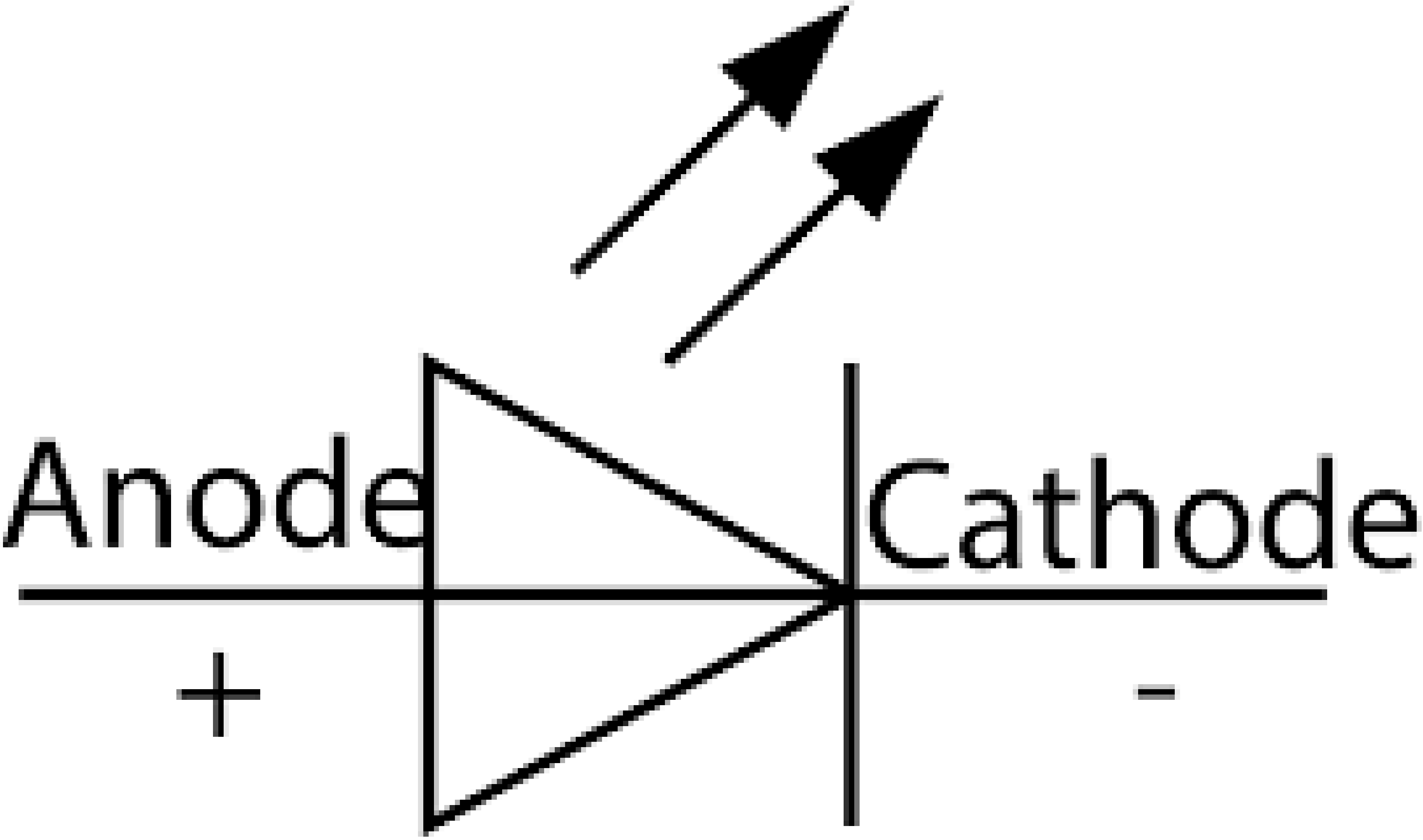
Anode
Cathode
+
-

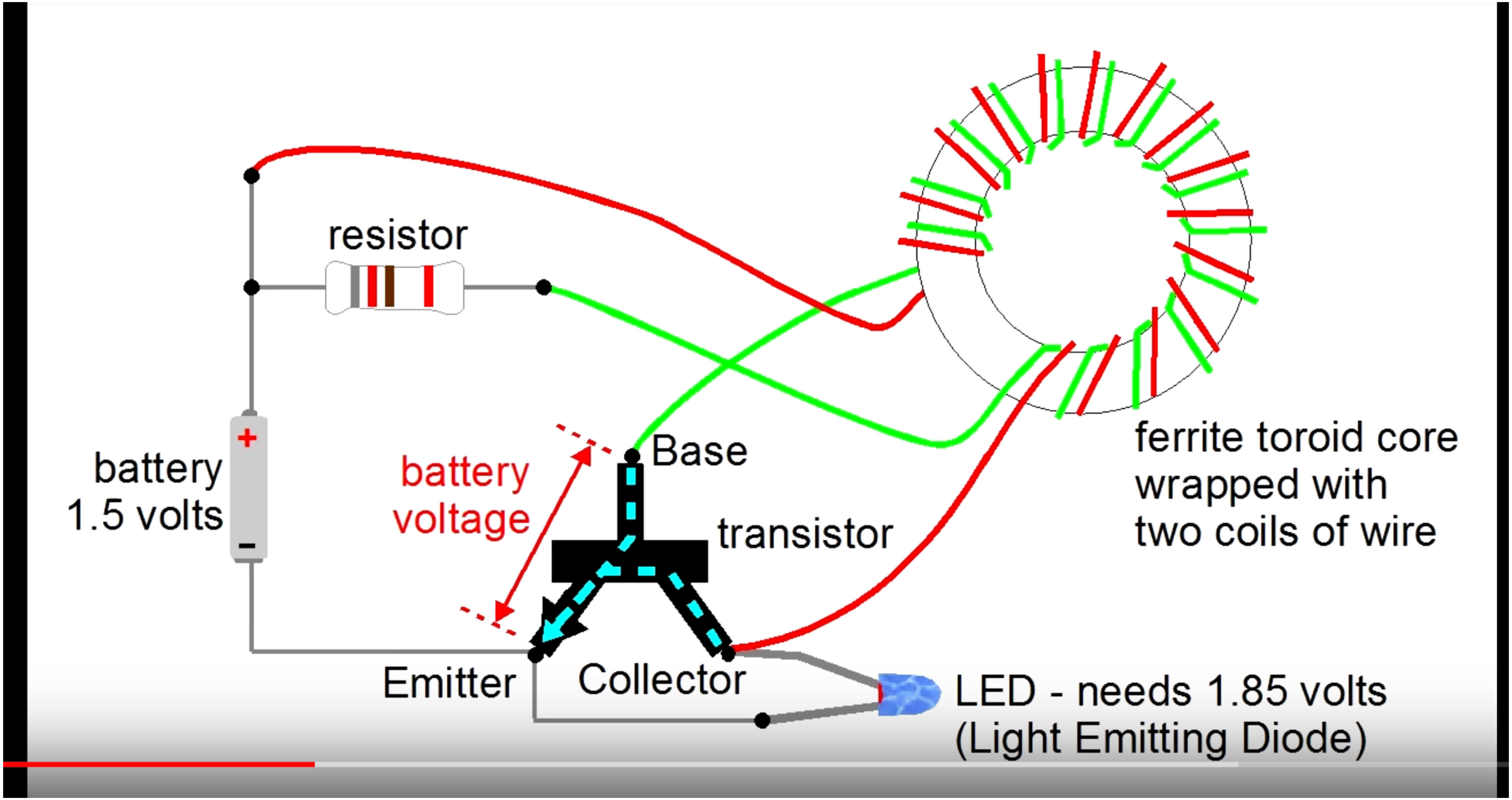
resistor
battery
1.5 volts
+
-
battery
voltage
Base
transistor
Emitter
Collector
ferrite toroid core
wrapped with
two coils of wire
LED - needs 1.85 volts
(Light Emitting Diode)

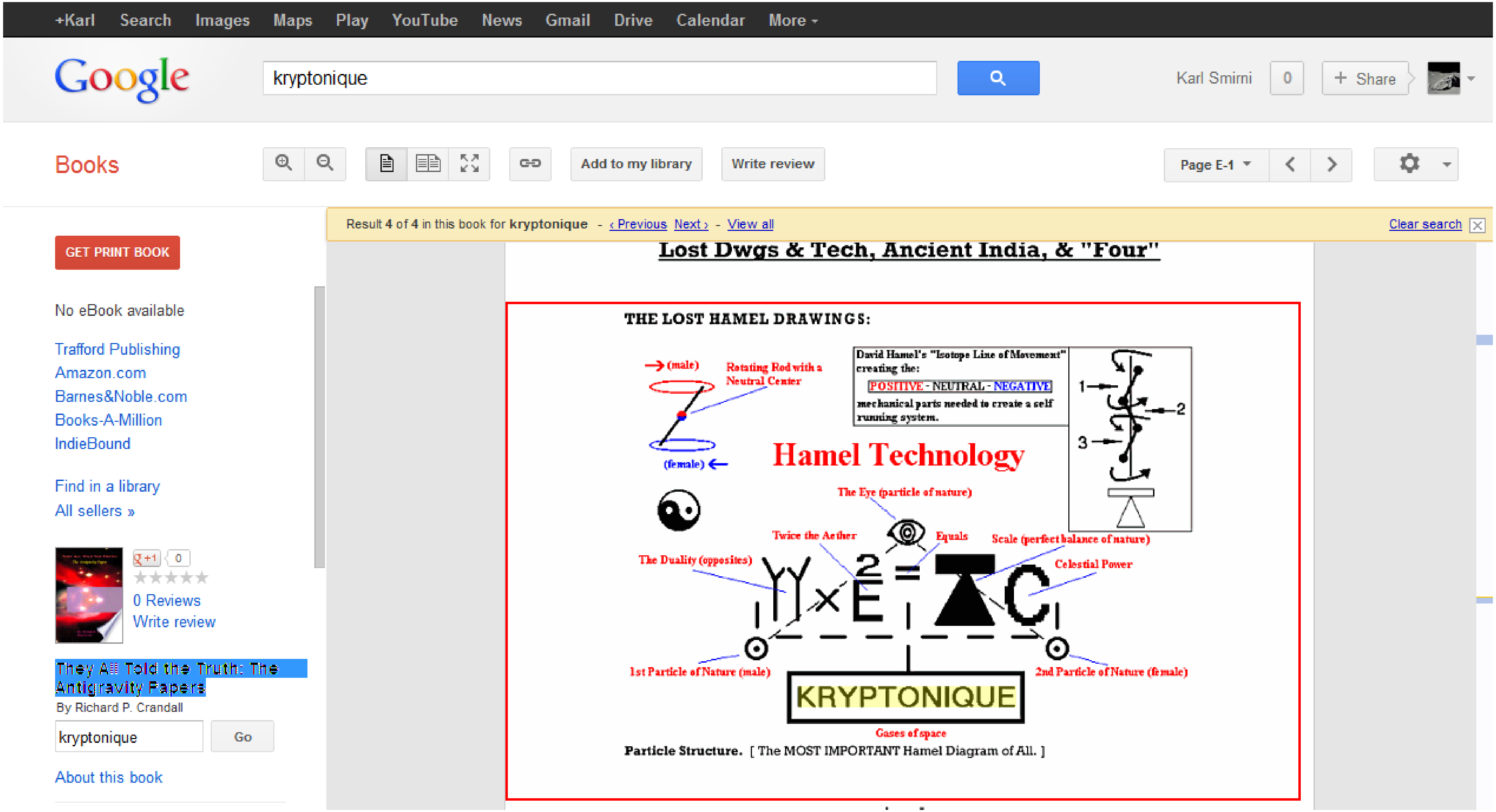
+Karl
Search
Images
Maps
Play
YouTube
News
Gmail
Drive
Calendar
More
Google
kryptonique
Karl Smirni
0
+ Share
Books
Add to my library
Write review
Page E-1
GET PRINT BOOK
No eBook available
Trafford Publishing
Amazon.com
Barnes&Noble.com
Books-A-Million
IndieBound
Find in a library
All sellers »
+1
0
0 Reviews
Write review
They All Told the Truth: The Antigravity Papers
By Richard P. Crandall
kryptonique
Go
About this book
Clear search

Result 4 of 4 in this book for **kryptonique** - ‹ Previous Next › - View all

Lost Dwgs & Tech, Ancient India, & "Four"

THE LOST HAMEL DRAWINGS:

David Hamel's "Isotope Line of Movement" creating the:
POSITIVE - NEUTRAL - NEGATIVE
mechanical parts needed to create a self running system.

(male)
Rotating Rod with a Neutral Center
(female)
1
2
3
Hamel Technology
The Eye (particle of nature)
Twice the Aether
Equals
Scale (perfect balance of nature)
The Duality (opposites)
Celestial Power
1st Particle of Nature (male)
2nd Particle of Nature (female)
KRYPTONIQUE
Gases of space

Particle Structure. [The MOST IMPORTANT Hamel Diagram of All.]

- synchronous charging circuit generate more electromagnetic wave

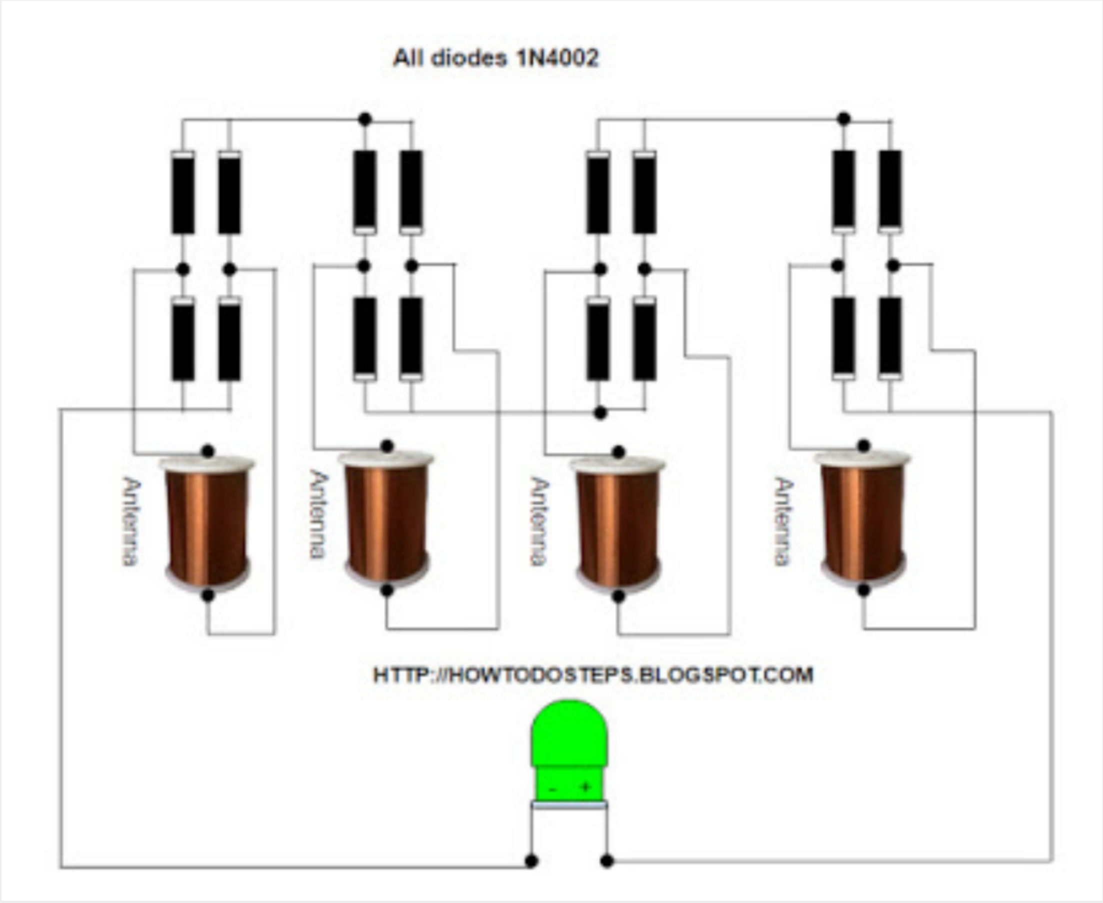
All diodes 1N4002
Antenna
Antenna
Antenna
Antenna
HTTP://HOWTODOSTEPS.BLOGSPOT.COM
-
+

Easy Simple Free Energy Harvesting from Surro...

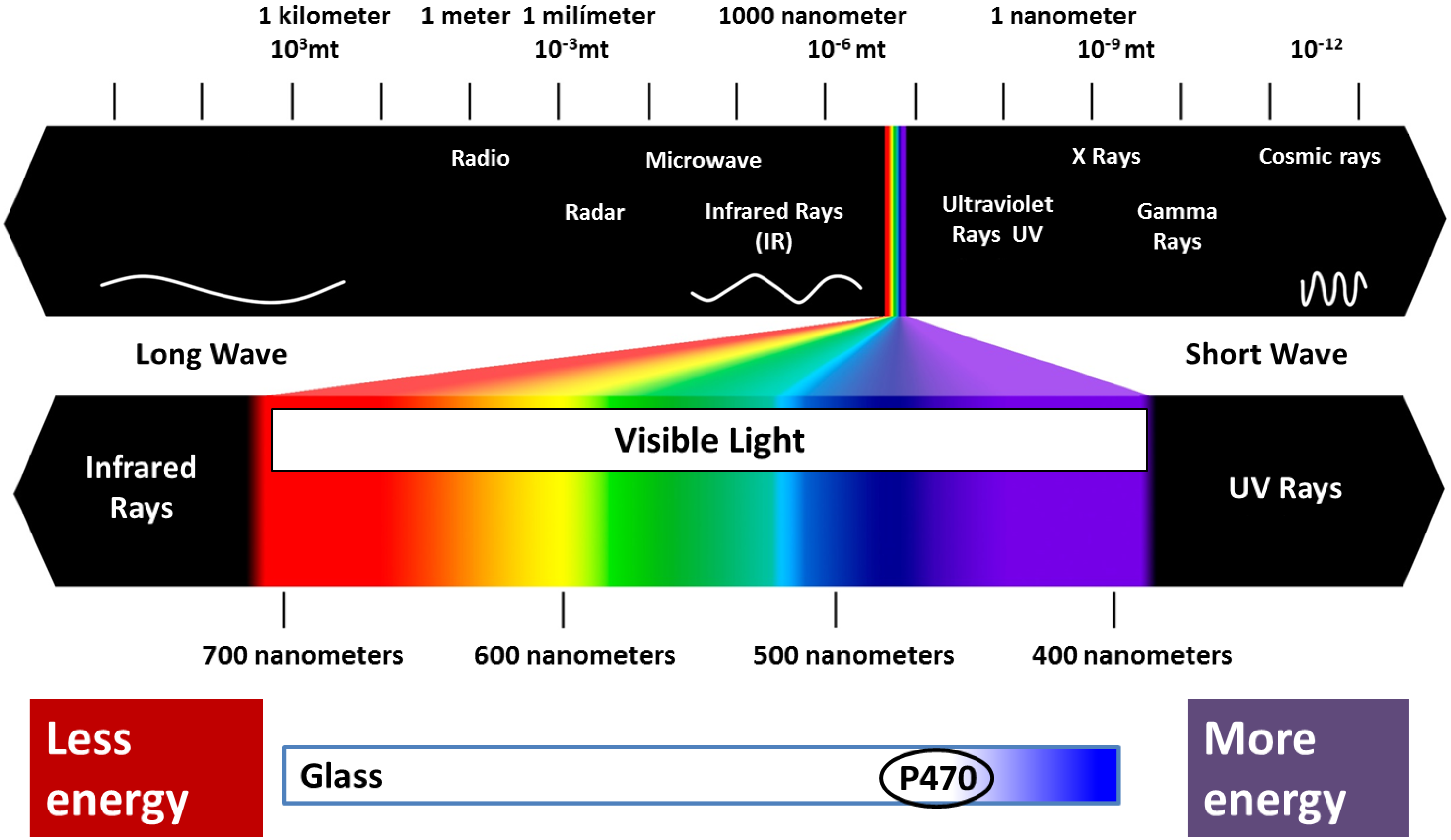
1 kilometer
10^3mt
1 meter
1 milímeter
10^{-3}mt
1000 nanometer
10^{-6} mt
1 nanometer
10^{-9} mt
10^{-12}
Radio
Microwave
Radar
Infrared Rays (IR)
Ultraviolet Rays UV
X Rays
Gamma Rays
Cosmic rays
Long Wave
Short Wave
Infrared Rays
Visible Light
UV Rays
700 nanometers
600 nanometers
500 nanometers
400 nanometers
Less energy
Glass
P470
More energy

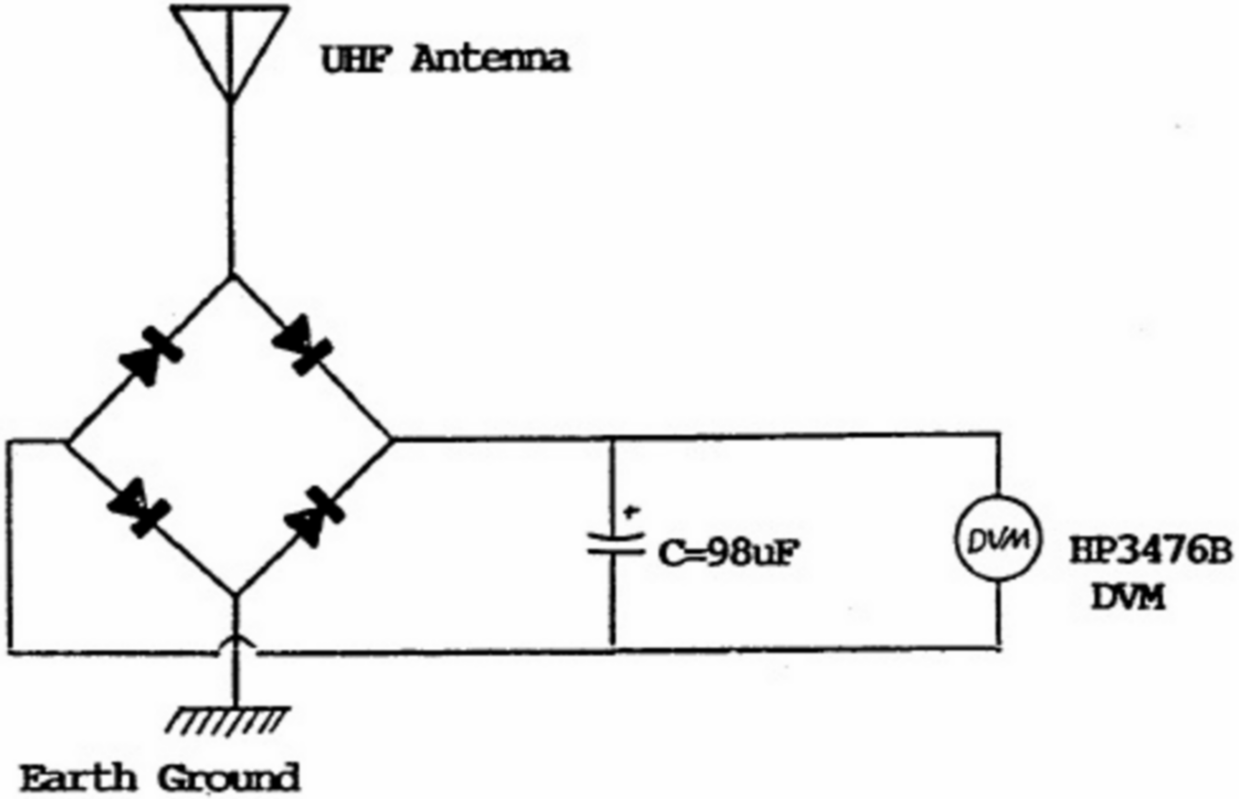
UHF Antenna
+
C=98uF
DVM
HP3476B
DVM
Earth Ground

Figure 2 -- Shows the main components of the FEA : 1) a bridge rectifier, 2) an earth ground connection, 3) an antenna, and 4) a capacitor. The DVM was used for making test measurements.

19DBI 2.4GHZ HIGH GAIN PANEL ANTENNA

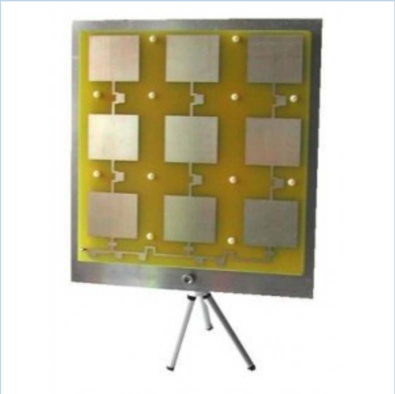

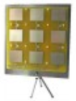

High gain 2.4gHz 19 dBi patch antenna.

More details

ON SALE!
$39.90

your@email.com

Notify me when available

Add to my wishlist

- Send to a friend
- Print
- View full size

IN THE SAME CATEGORY

14dBi 2.4gHz...

11dBi 2.4gHz...

18 dBi 5.8...

12dBi 5.8...

14dBi 5.8...

MORE INFO

19 dBi panel antenna.

Sizes 26x27 cm.

Horizontal & vertical beam width: 17 **degrees**

-

Package contains:

19 dBi antenna.

1 meter SMA connection cabel

Tripod.

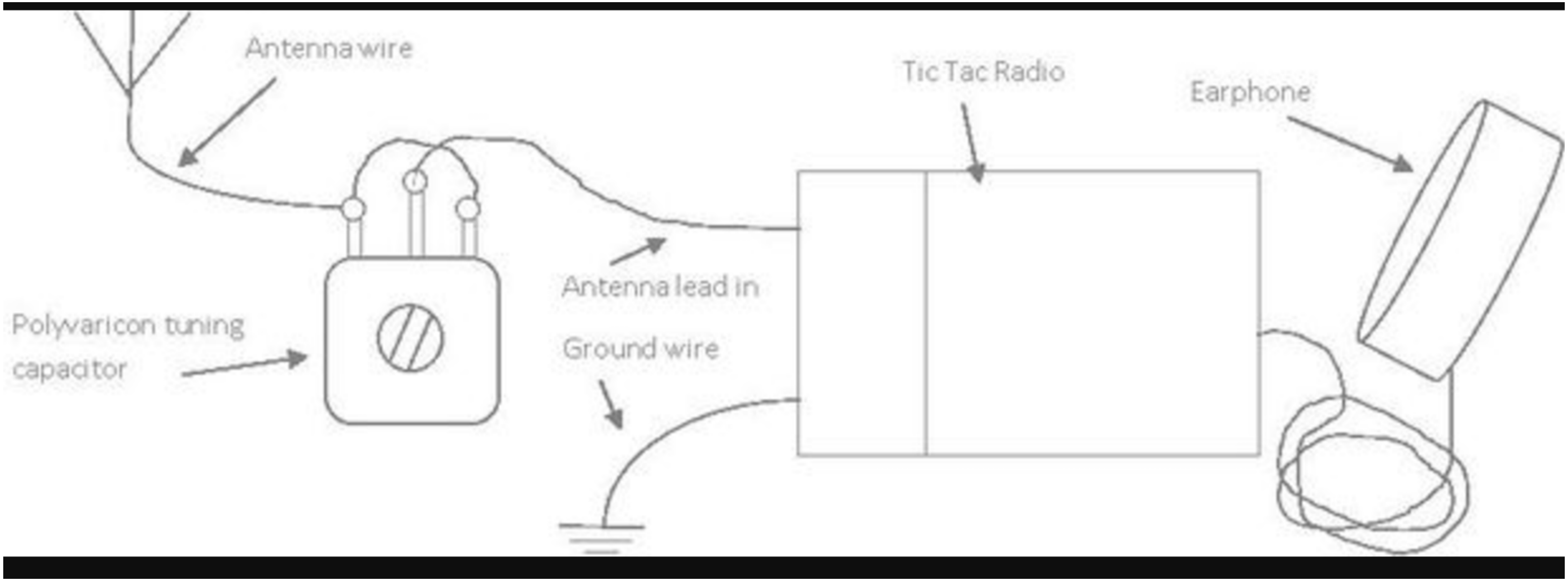
Antenna wire
Tic Tac Radio
Earphone
Antenna lead in
Polyvaricon tuning capacitor
Ground wire

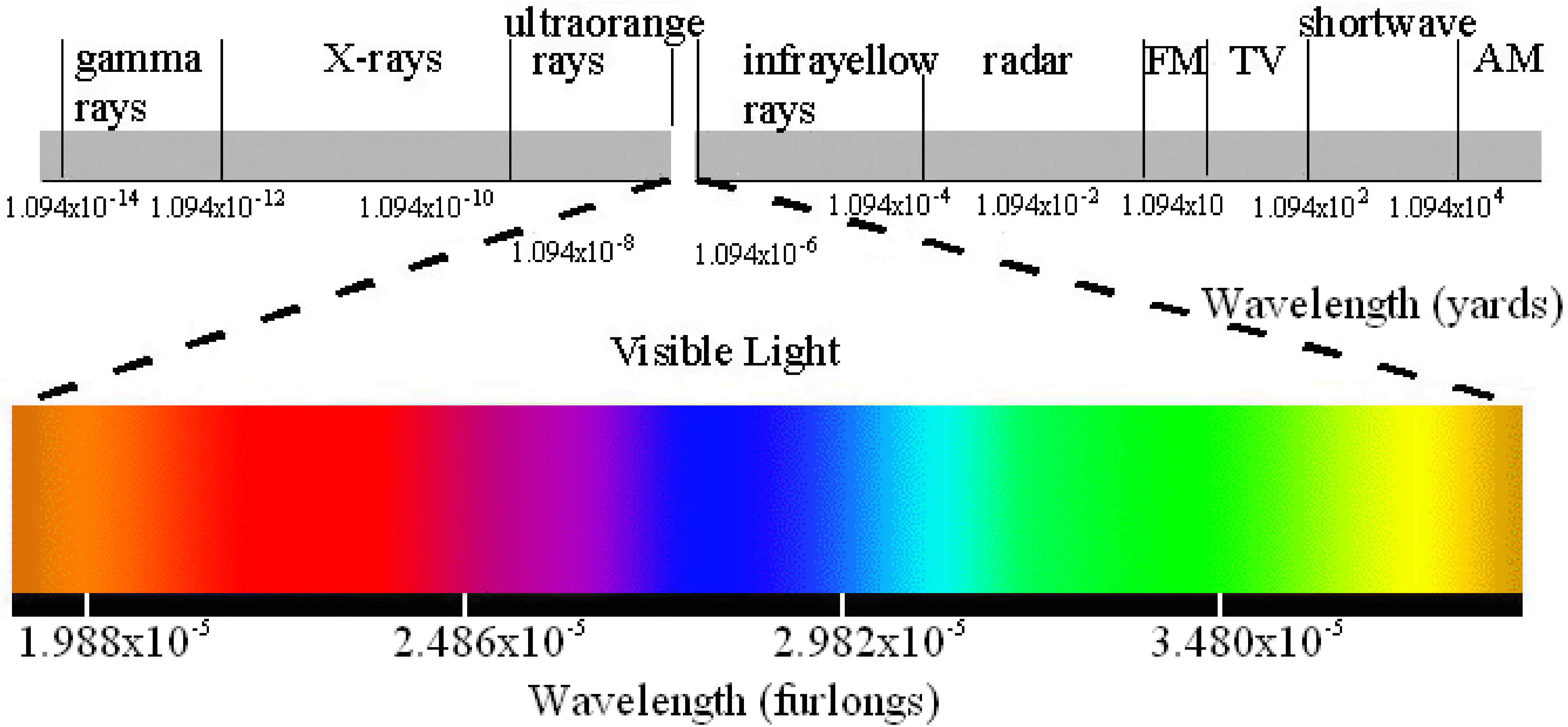
gamma rays
X-rays
ultraorange rays
infrayellow rays
radar
FM
TV
shortwave
AM
1.094x10^-14
1.094x10^-12
1.094x10^-10
1.094x10^-8
1.094x10^-6
1.094x10^-4
1.094x10^-2
1.094x10
1.094x10^2
1.094x10^4
Wavelength (yards)
Visible Light
1.988x10^-5
2.486x10^-5
2.982x10^-5
3.480x10^-5
Wavelength (furlongs)

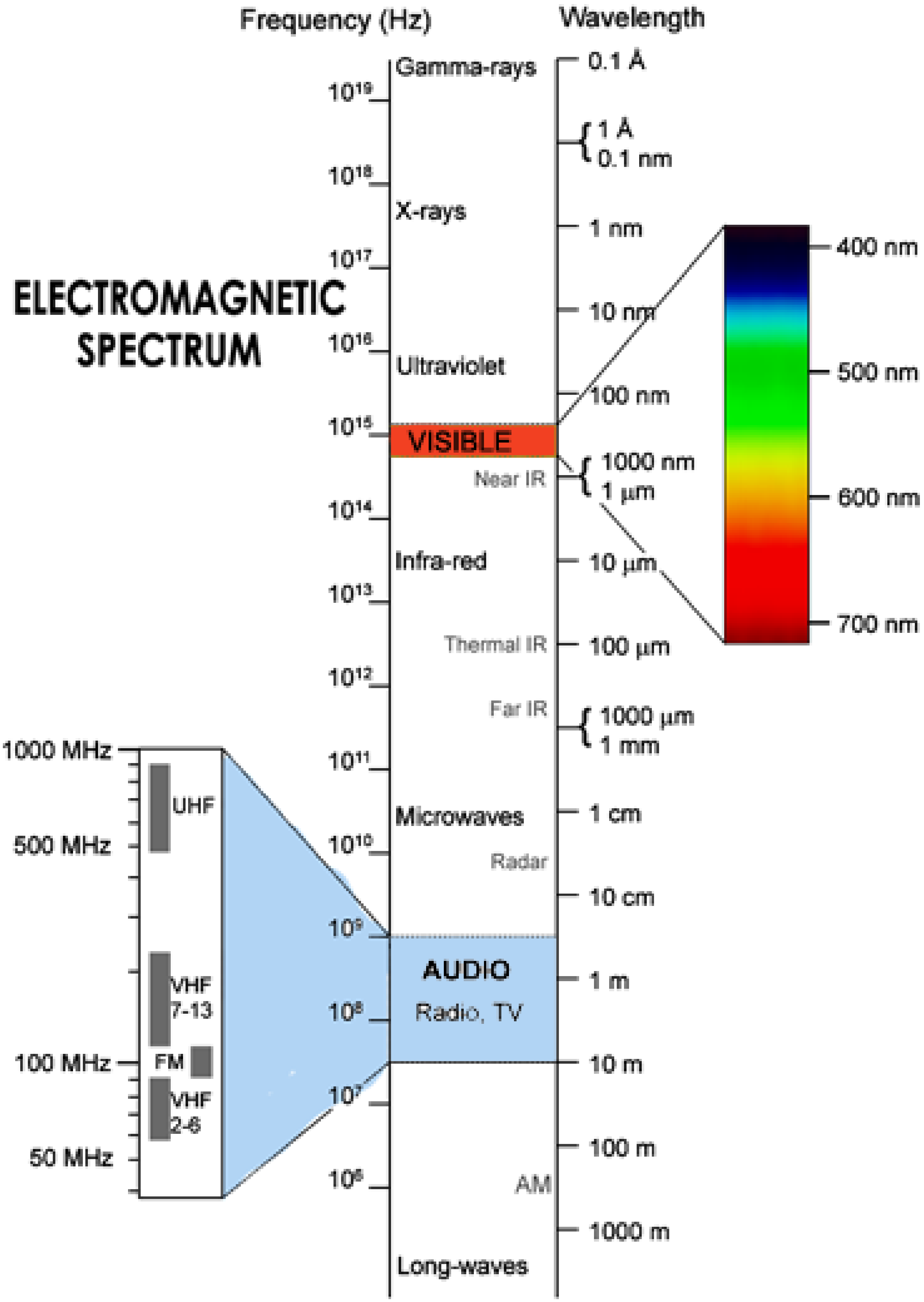
ELECTROMAGNETIC SPECTRUM
Frequency (Hz)
Wavelength
Gamma-rays
10^{19}
10^{18}
X-rays
10^{17}
10^{16}
Ultraviolet
10^{15}
VISIBLE
Near IR
10^{14}
Infra-red
10^{13}
Thermal IR
10^{12}
Far IR
10^{11}
Microwaves
10^{10}
Radar
10^{9}
AUDIO
Radio, TV
10^{8}
10^{7}
10^{6}
AM
Long-waves
0.1 Å
1 Å
0.1 nm
1 nm
10 nm
100 nm
1000 nm
1 µm
10 µm
100 µm
1000 µm
1 mm
1 cm
10 cm
1 m
10 m
100 m
1000 m
400 nm
500 nm
600 nm
700 nm
1000 MHz
500 MHz
100 MHz
50 MHz
UHF
VHF
7-13
FM
VHF
2-6

Louis E. Keiner - Coastal Carolina University

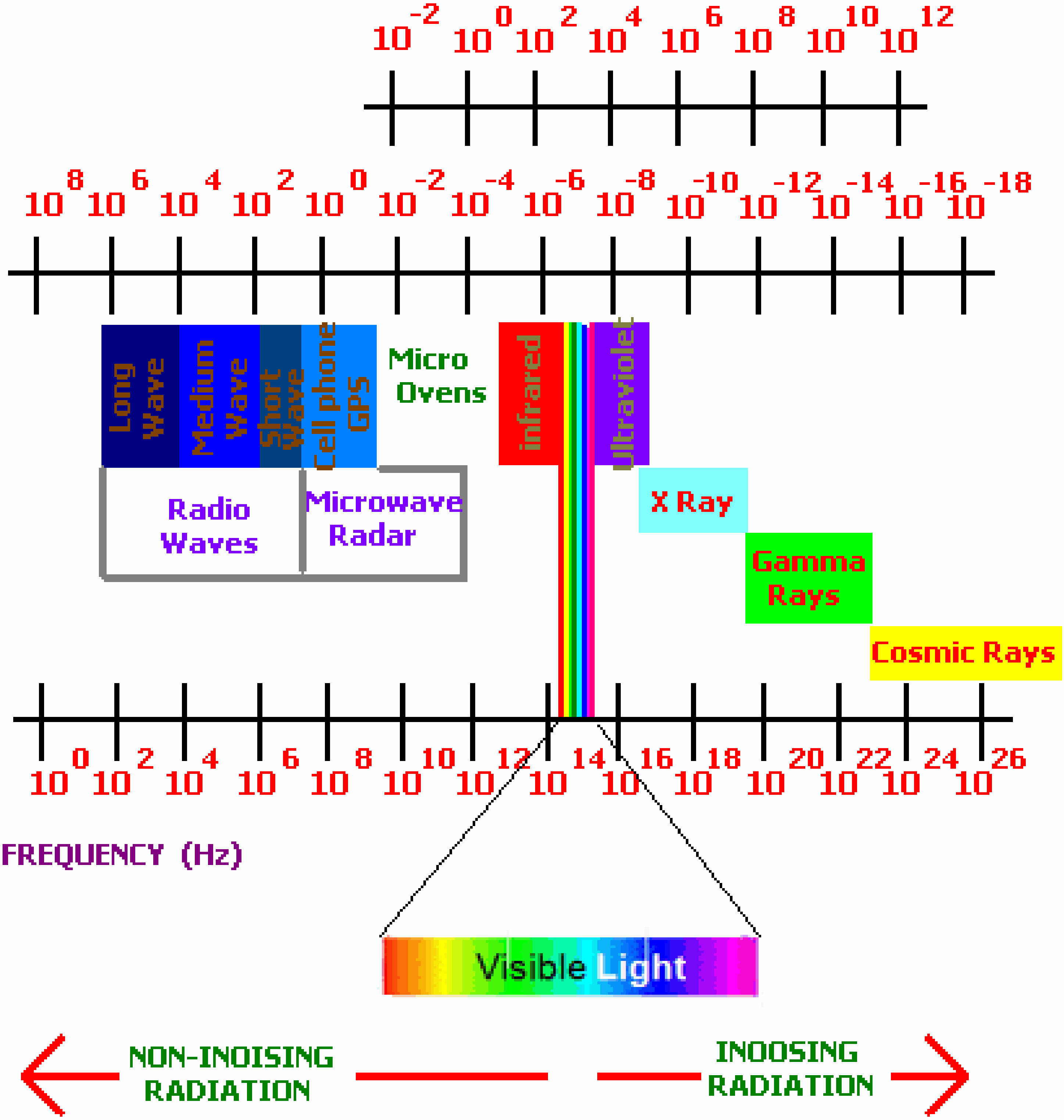
10^{-2} 10^{0} 10^{2} 10^{4} 10^{6} 10^{8} 10^{10} 10^{12}
10^{8} 10^{6} 10^{4} 10^{2} 10^{0} 10^{-2} 10^{-4} 10^{-6} 10^{-8} 10^{-10} 10^{-12} 10^{-14} 10^{-16} 10^{-18}
Long Wave
Medium Wave
Short Wave
Cell phone GPS
Micro Ovens
infrared
ultraviolet
Radio Waves
Microwave Radar
X Ray
Gamma Rays
Cosmic Rays
10^{0} 10^{2} 10^{4} 10^{6} 10^{8} 10^{10} 10^{12} 10^{14} 10^{16} 10^{18} 10^{20} 10^{22} 10^{24} 10^{26}
FREQUENCY (Hz)
Visible Light
NON-INOISING RADIATION
INOOSING RADIATION

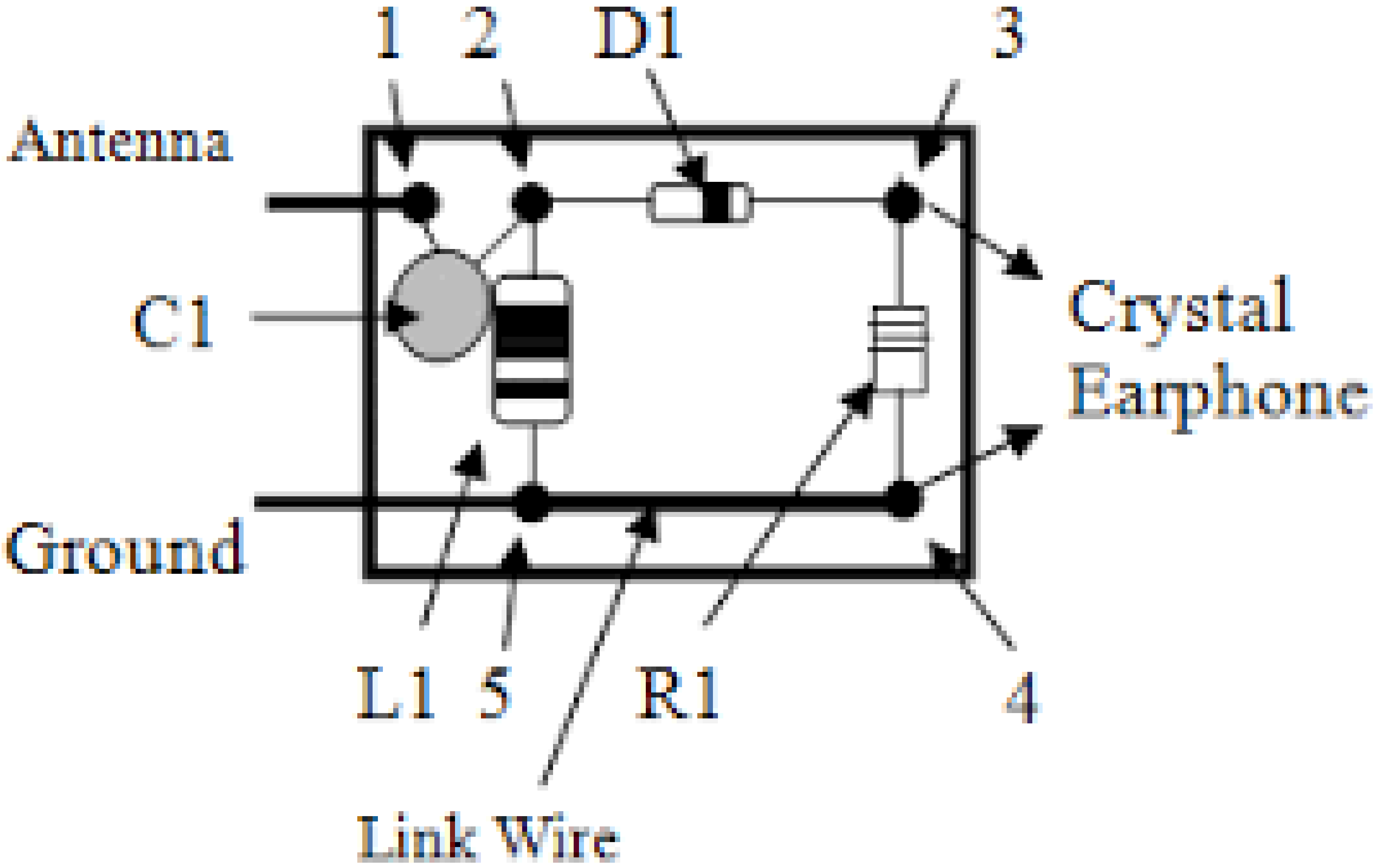
1
2
D1
3
Antenna
C1
Crystal
Earphone
Ground
L1
5
R1
4
Link Wire

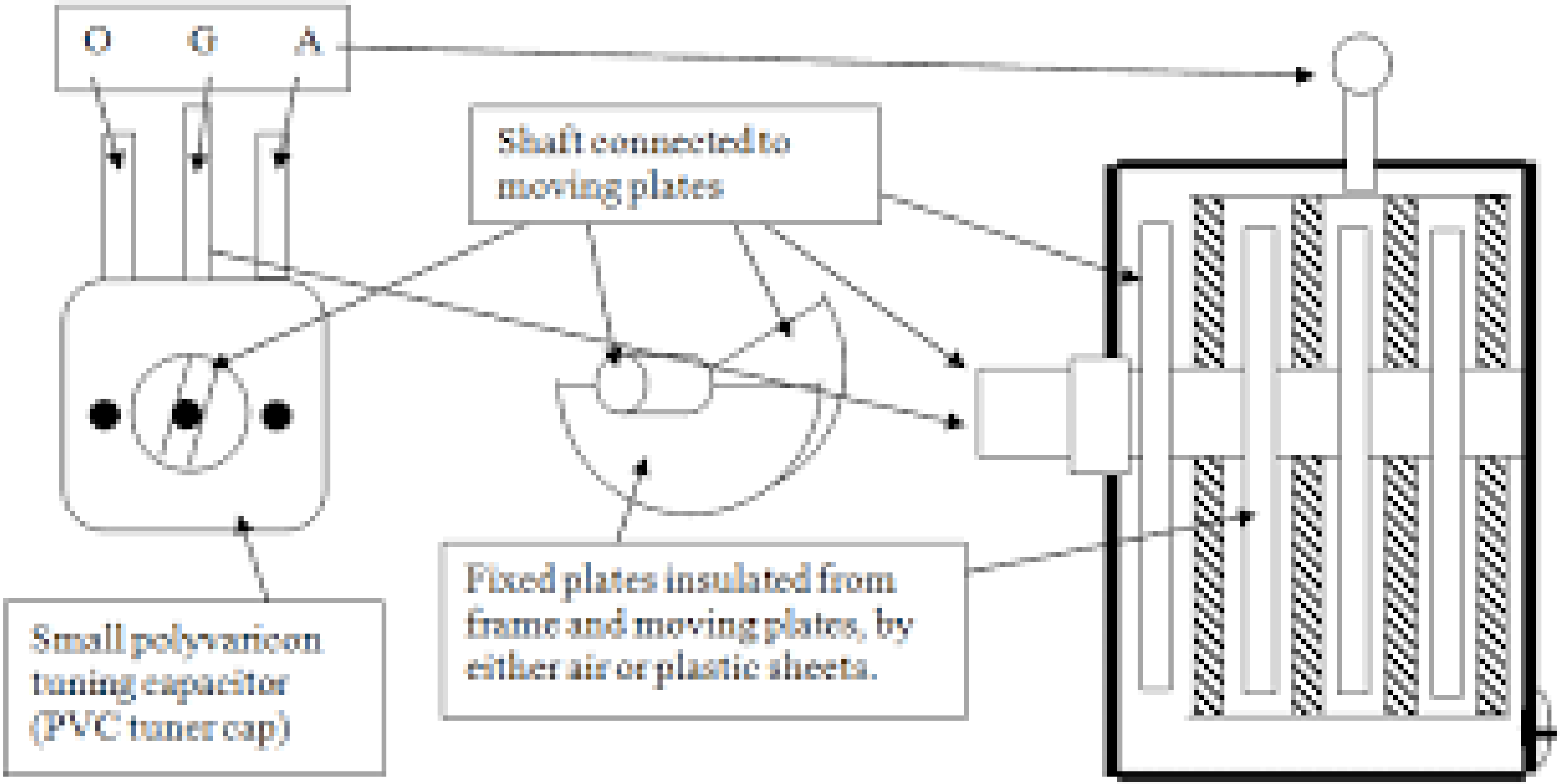
O
G
A
Shaft connected to moving plates
Fixed plates insulated from frame and moving plates, by either air or plastic sheets.
Small polyvaricon tuning capacitor (PVC tuner cap)

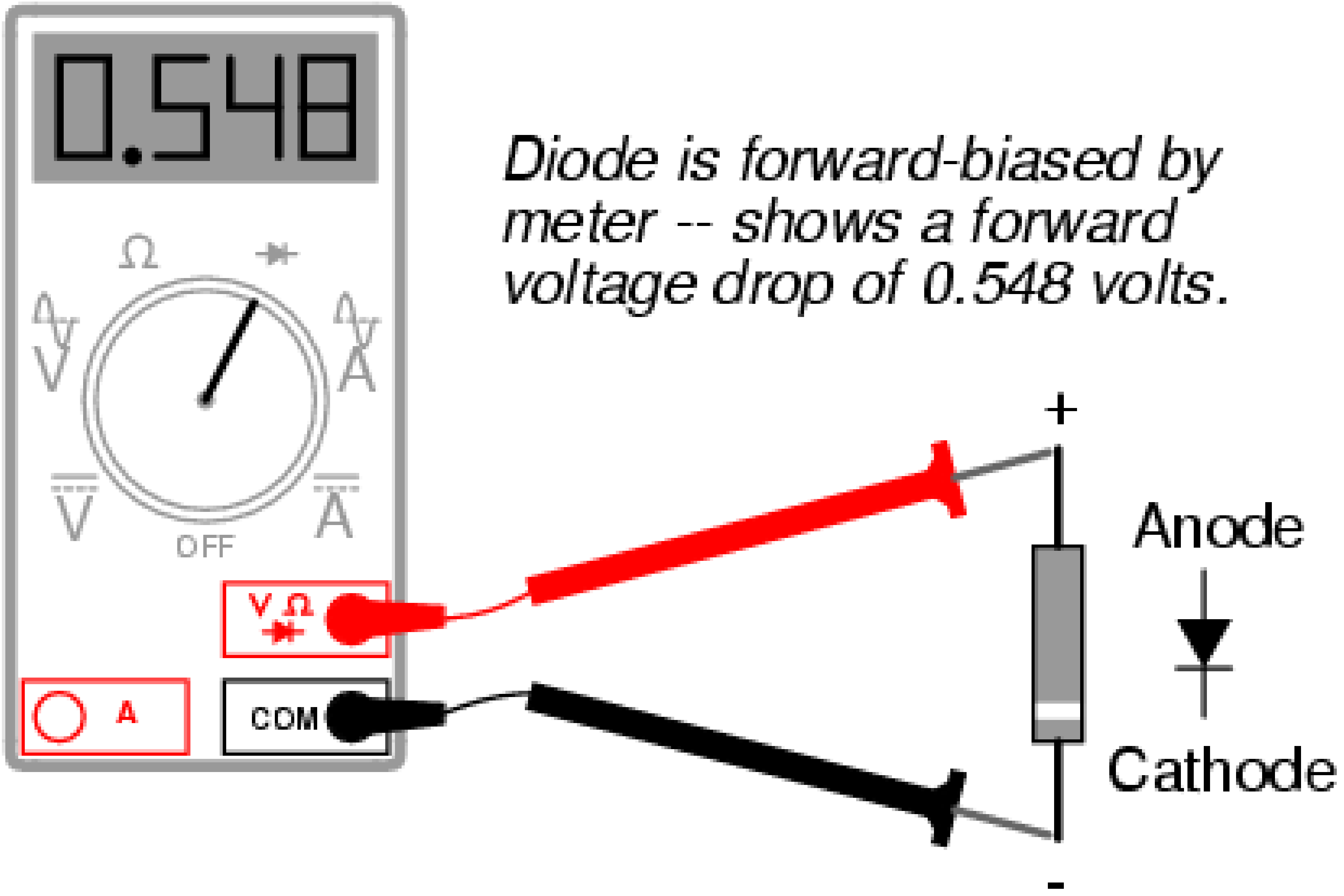
0.548
Ω
V
A
V
A
OFF
V Ω
COM
A
+
Anode
Cathode
-

Diode is forward-biased by meter -- shows a forward voltage drop of 0.548 volts.

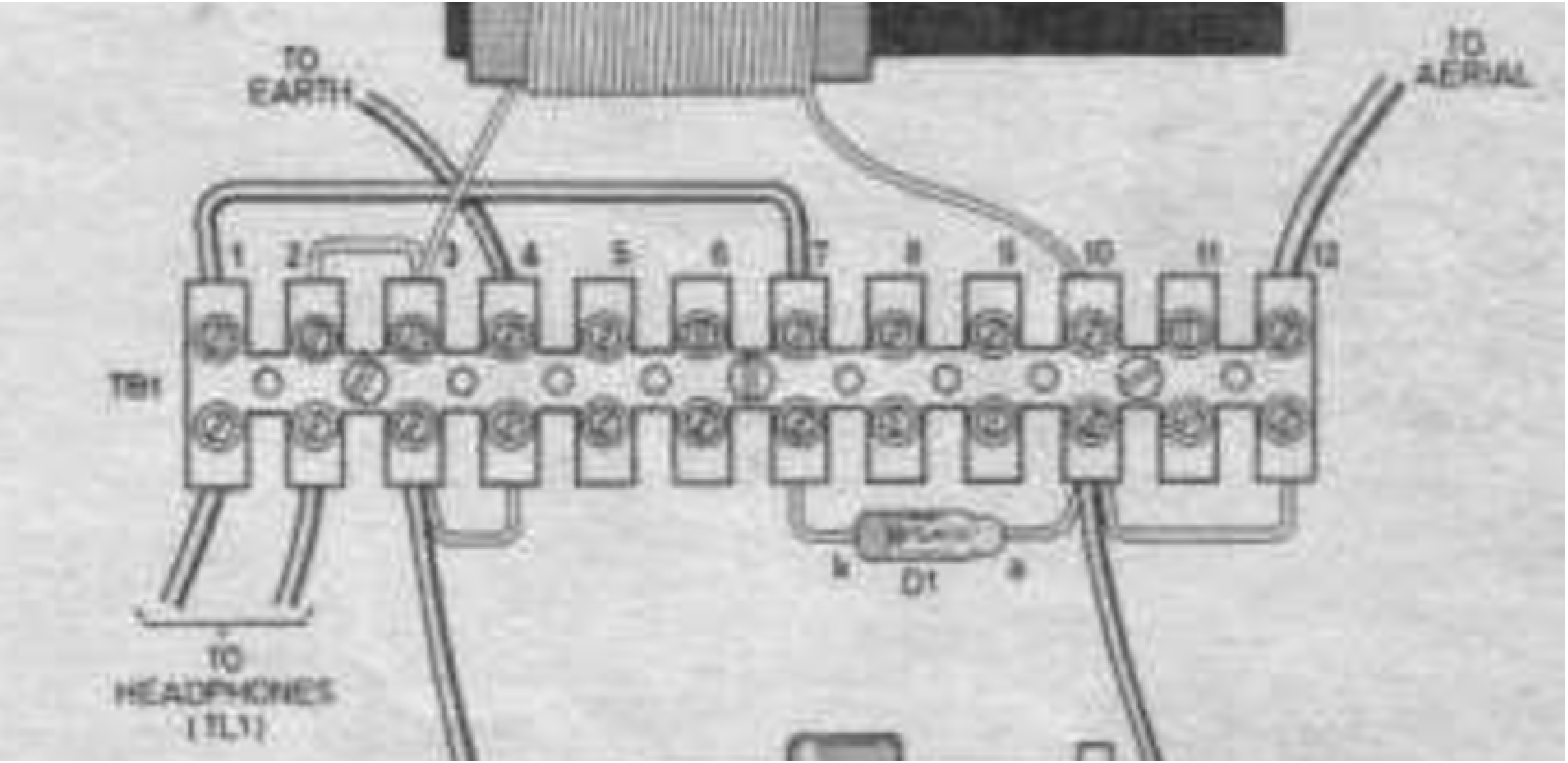
TO
EARTH
TO
AERIAL
1
2
3
4
5
6
7
8
9
10
11
12
TB1
k
D1
a
TO
HEADPHONES
(TL1)

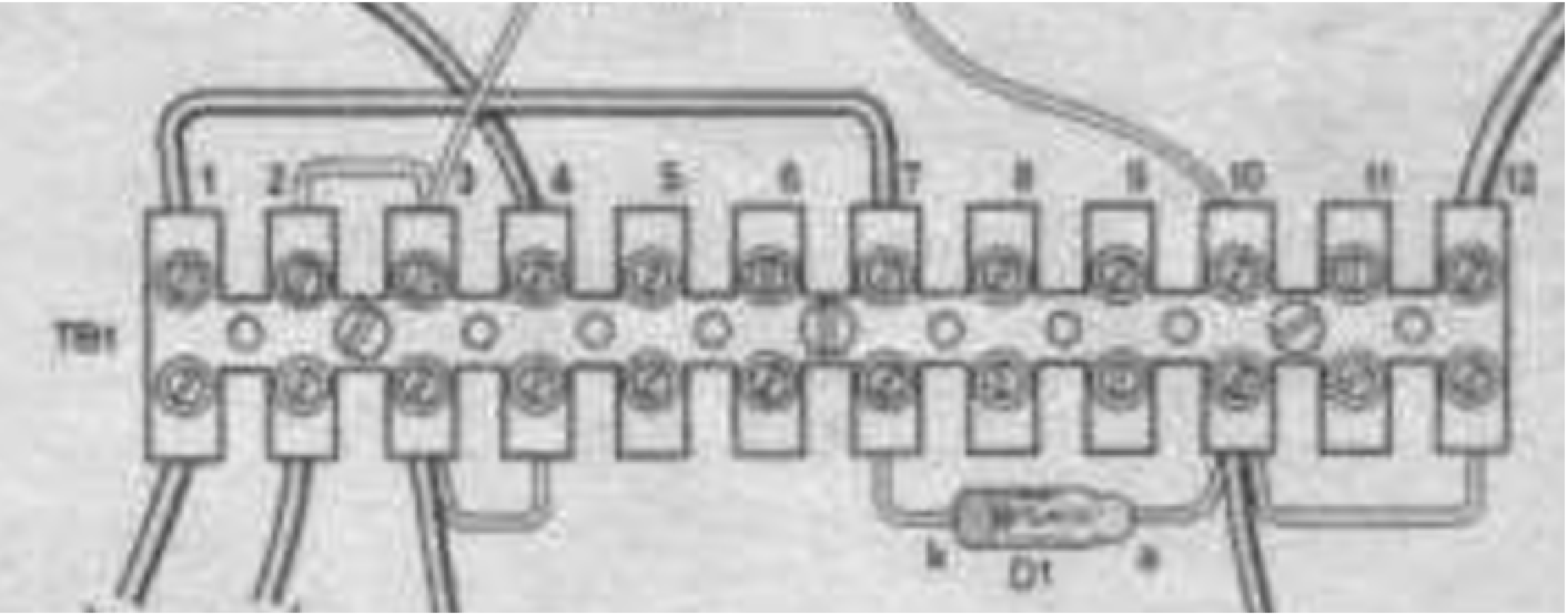
TB1
k
D1
a

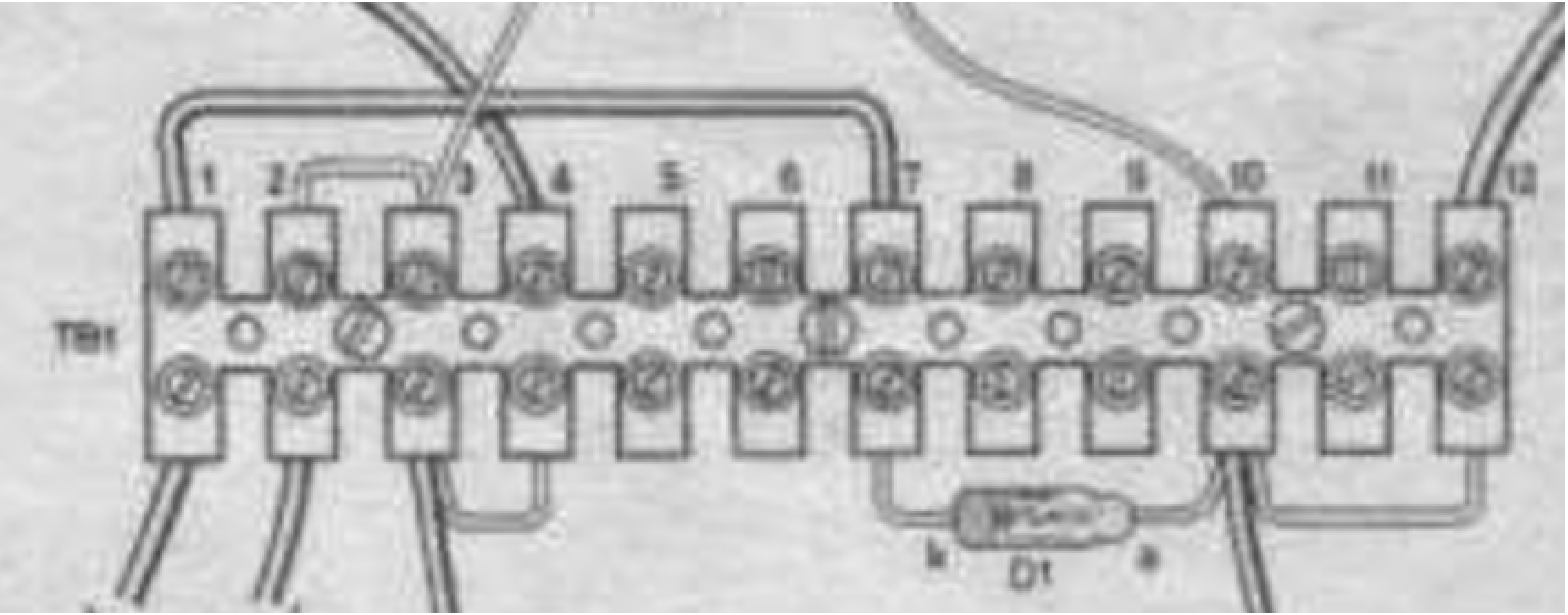
TB1
k
D1
a

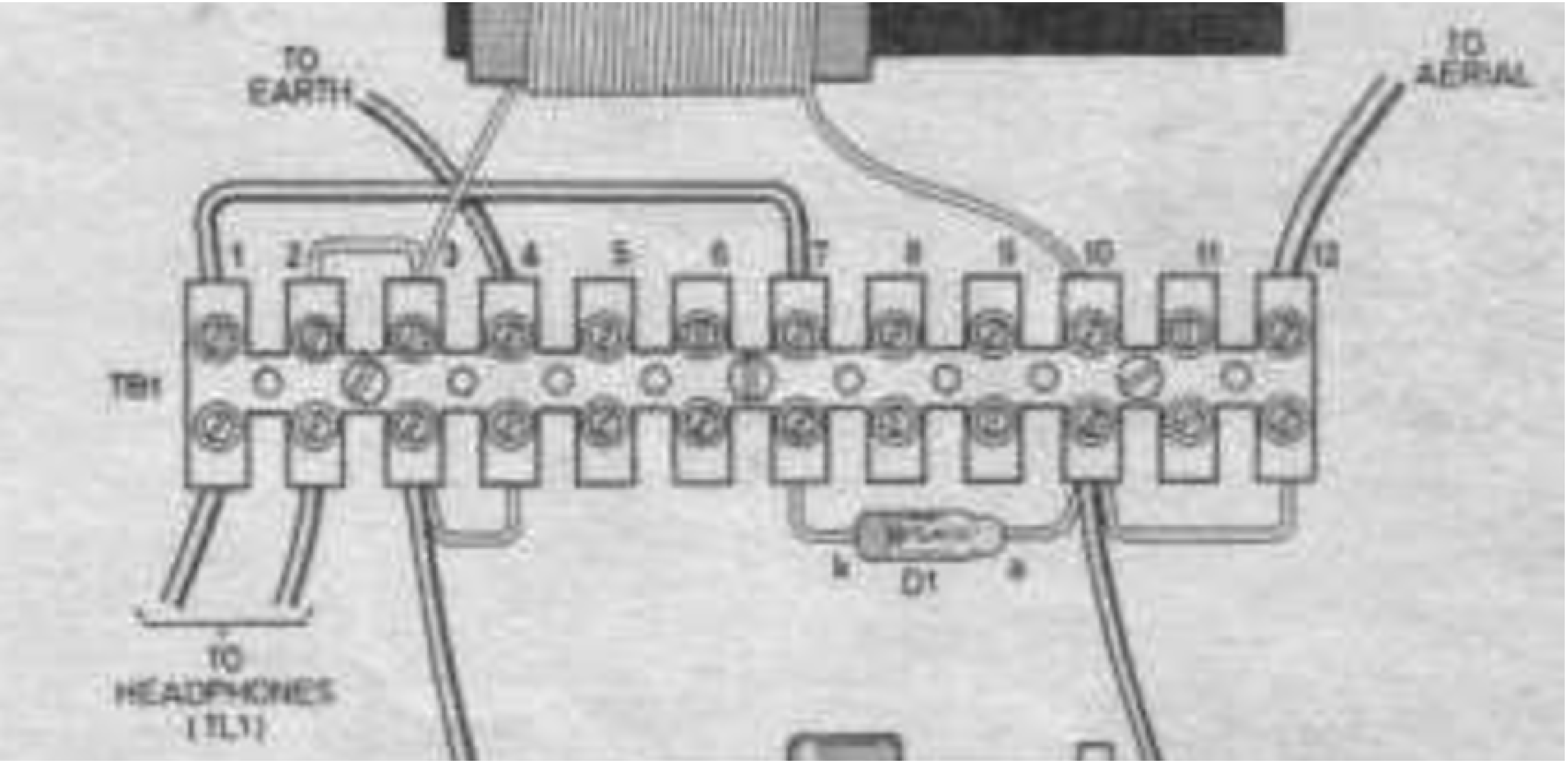
TO
EARTH
TO
AERIAL
1
2
3
4
5
6
7
8
9
10
11
12
TB1
k
D1
a
TO
HEADPHONES
(TL1)

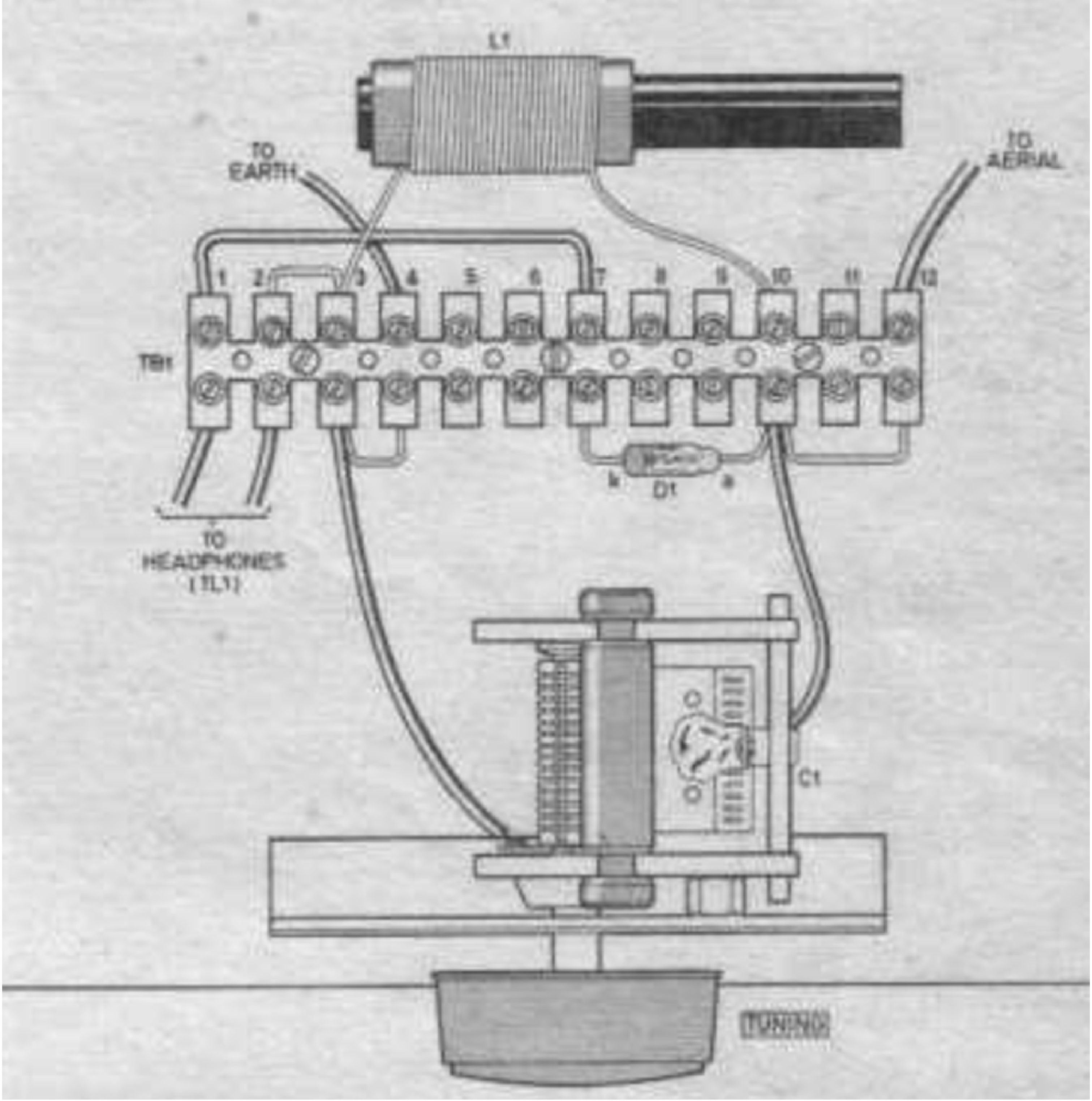
L1
TO EARTH
TO AERIAL
1
2
3
4
5
6
7
8
9
10
11
12
TB1
k
D1
a
TO HEADPHONES (TL1)
C1
TUNING

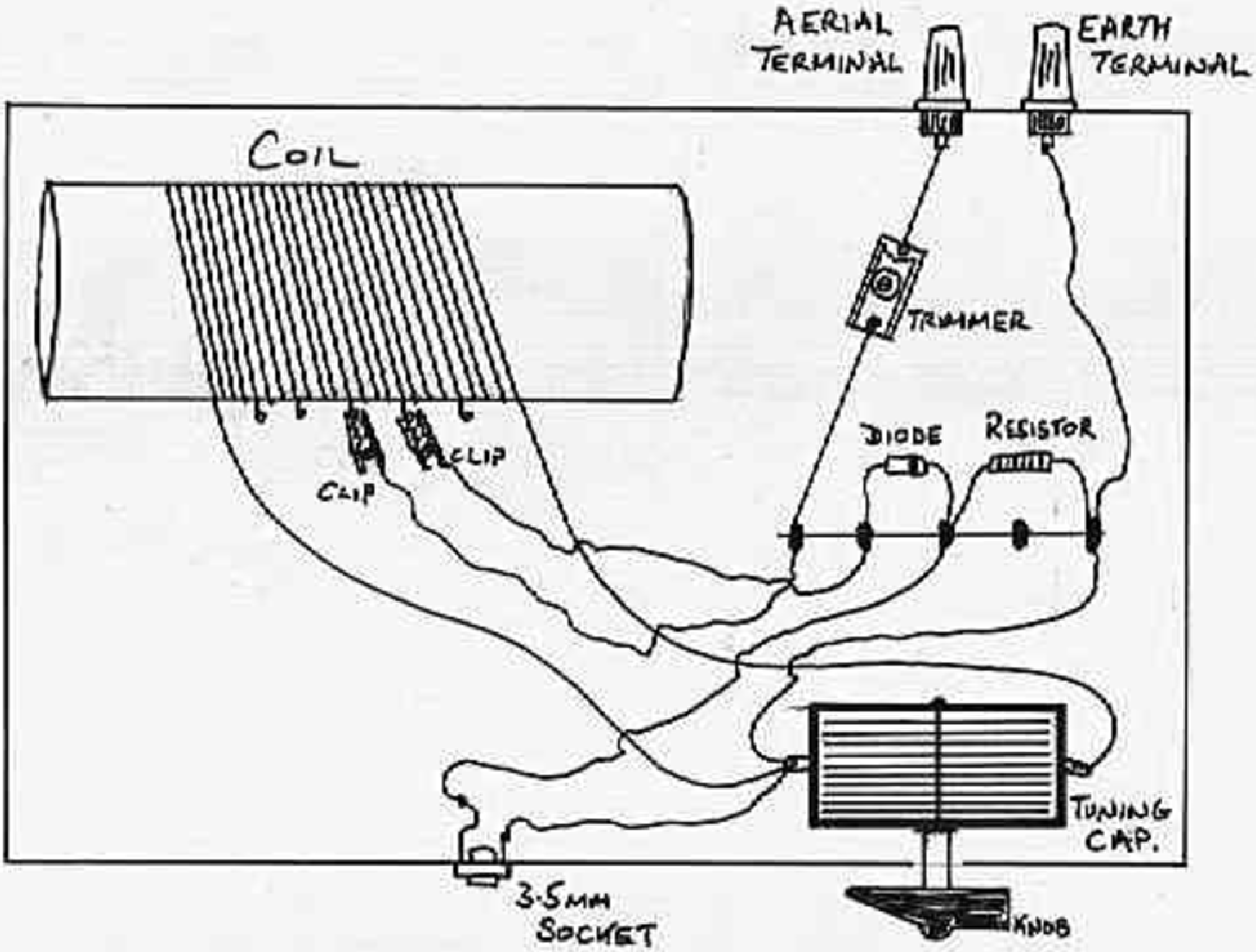
AERIAL TERMINAL
EARTH TERMINAL
COIL
TRIMMER
DIODE
RESISTOR
CLIP
CLIP
TUNING CAP.
3·5MM SOCKET
KNOB

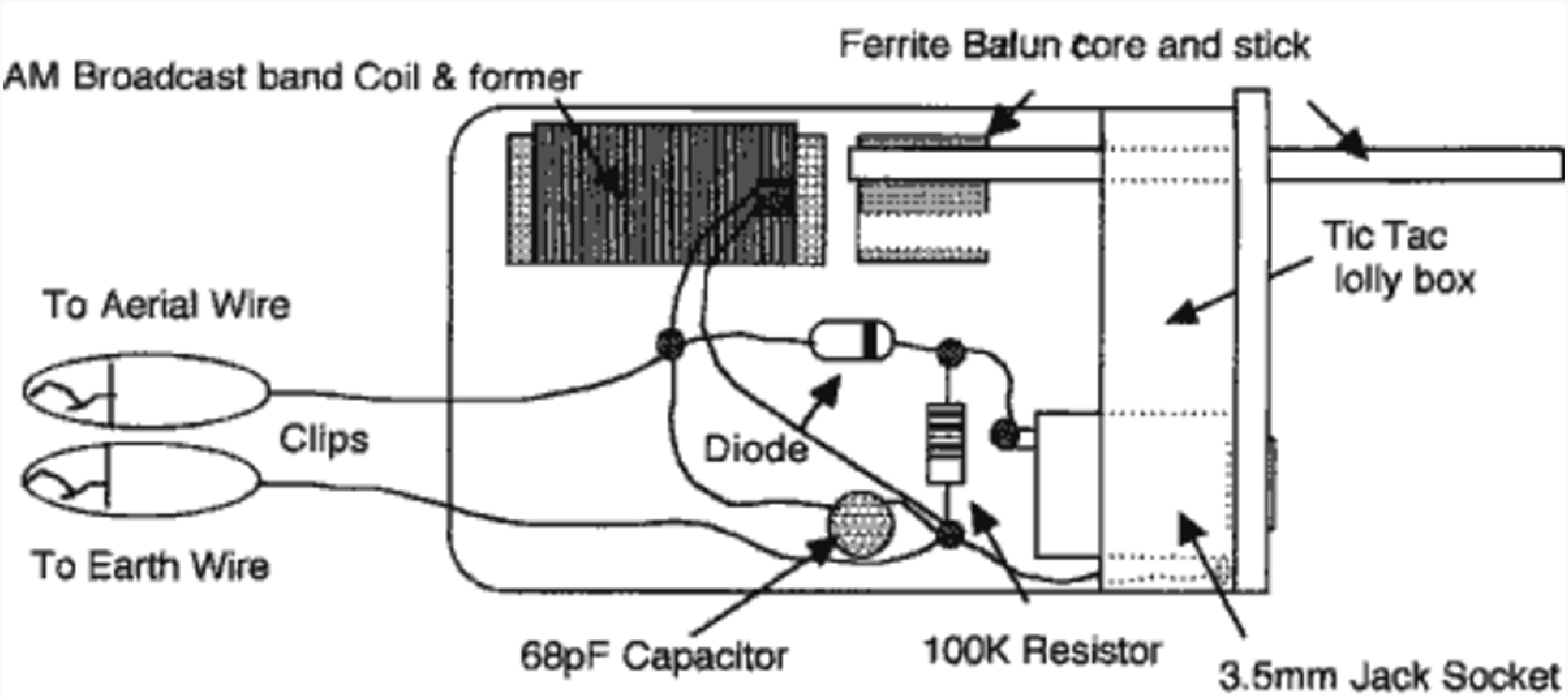
AM Broadcast band Coil & former
Ferrite Balun core and stick
Tic Tac
lolly box
To Aerial Wire
Clips
Diode
To Earth Wire
68pF Capacitor
100K Resistor
3.5mm Jack Socket

Diagram 2. Place balun inside coil former and use the stick to slide it up and down inside the coil to tune the Nic Nac Radio.

The MK 484 TRF Receiver - With Inductive Coupling!

Copyright © **Austin Hellier** 10/01/2001

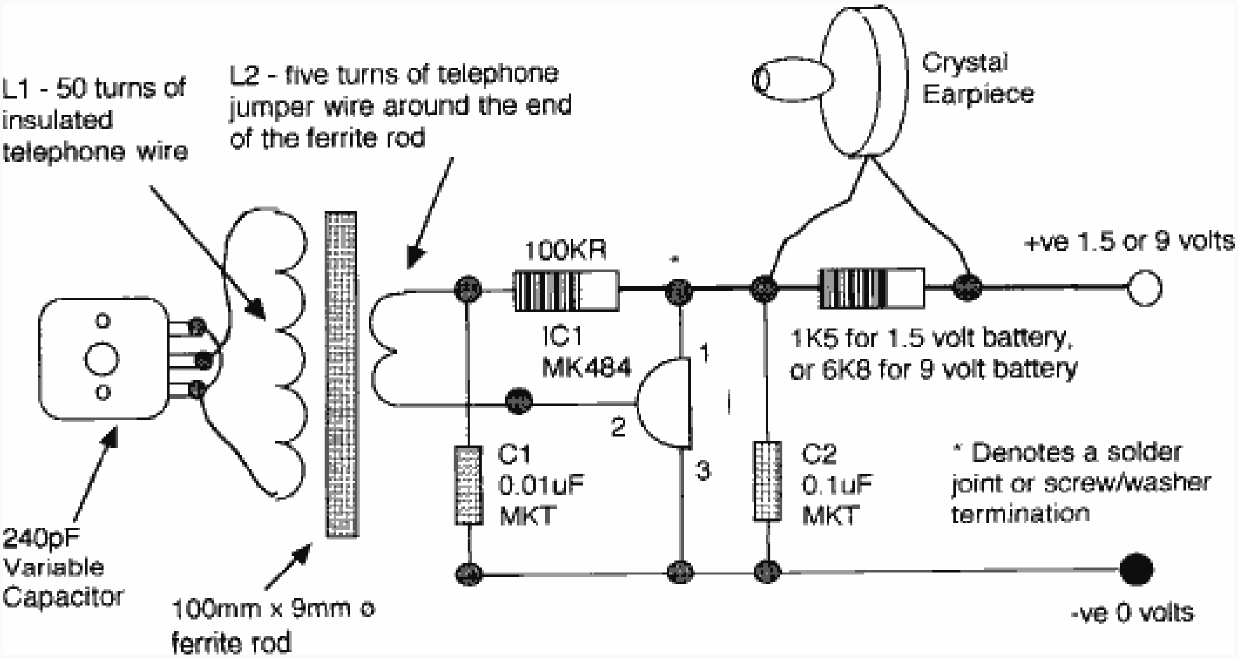
L1 - 50 turns of insulated telephone wire
L2 - five turns of telephone jumper wire around the end of the ferrite rod
Crystal Earpiece
100KR
*
+ve 1.5 or 9 volts
IC1
MK484
1
2
3
1K5 for 1.5 volt battery, or 6K8 for 9 volt battery
C1
0.01uF
MKT
C2
0.1uF
MKT
* Denotes a solder joint or screw/washer termination
240pF Variable Capacitor
100mm x 9mm ø ferrite rod
-ve 0 volts

Diagram 1. MK484 Radio Receiver for the AM Band

Build This Modern AM Radio Receiver For Just $12.00!

This adaption of the famous ZN414 (now MK484 plastic pack) AM radio receiver, was inspired by the fact that I live fairly close to my three local AM stations (2WL, 2WN and the ethnic broadcaster 2RPH) and would normally receive overloaded signals with a standard TRF design based on this IC. With the minor modifications made to the basic design as shown, and with the aid of the inductive coupling arrangement, i can pick up all three stations with clarity and volume, and with hardly any (and I mean ANY) interference from each other, and no oscillations!

Make sure that you wire the MK484 (IC1) up correctly - the pinouts shown above are seen from the top of the device. Wind L2 so that it fits snugly on the end of the rod, right next to the end of L1 's winding. The average 9 volt battery gives much better volume and selectivity, and should last for some months of constant use - if you've got a few spare $, try a new Eveready Lithium or the new Kodak Alkaline!

MK484 ICs are available from **http://www.dse.com.au/** (cat# Z6525 for $2.36AUD).

Construction

You can make this circuit up quickly on your SK-10 panel, or use something like the Dick Smith "Funway

Parts List		DSE#
IC1	MK 484	Z6525

The Nic Nac Crystal Radio

Copyright © **Austin Hellier** 12/02/2001

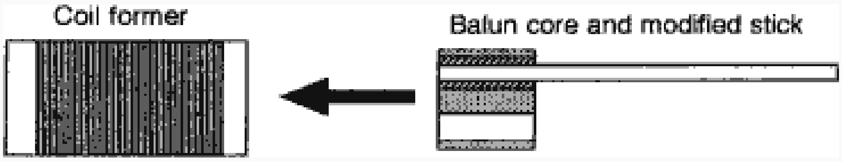
Coil former
Balun core and modified stick

How It Works

Radio signals are travelling through the atmosphere all the time. This simple radio receiver has been designed to capture them, and turn them into sound waves, so that you can hear what's going on. The Aerial and Earth wires form a return signal path from the AM transmitter, for these signals, and they are then fed into the coil and capacitor. This arrangement is known as a 'tank' circuit, because it stores radio energy, much in the same way your family car's tank stores petrol. The combination of a certain sized capacitor (68pF) and the amount of turns on the coil (around 120 turns) determines the frequency you will receive at. The Diode changes the signal from a radio one, to an audio one, and the Crystal Earphone changes the audio signal into sound waves that you can hear. the resistor adds stability by helping to prevent the Earphone from occasional 'dropout'.

Building The Radio

You'll need all the components listed below and a Tic Tac lolly box. Start by mounting the earphone jack into the lid of the box and then solder up each of the components into a tidy version of the 'rat's nest' construction method. Make sure you use a crocodile clip as a heat sink for the diode when soldering - they're pretty sensitive devices to heat. Then take the coil and remove the ferrite rod from inside it. Take the ferrite balun core and place one end of the plastic stick (obtained by cutting off both ends of a cotton bud) and glue it inside the hole, with its end level with the end of the balun core.

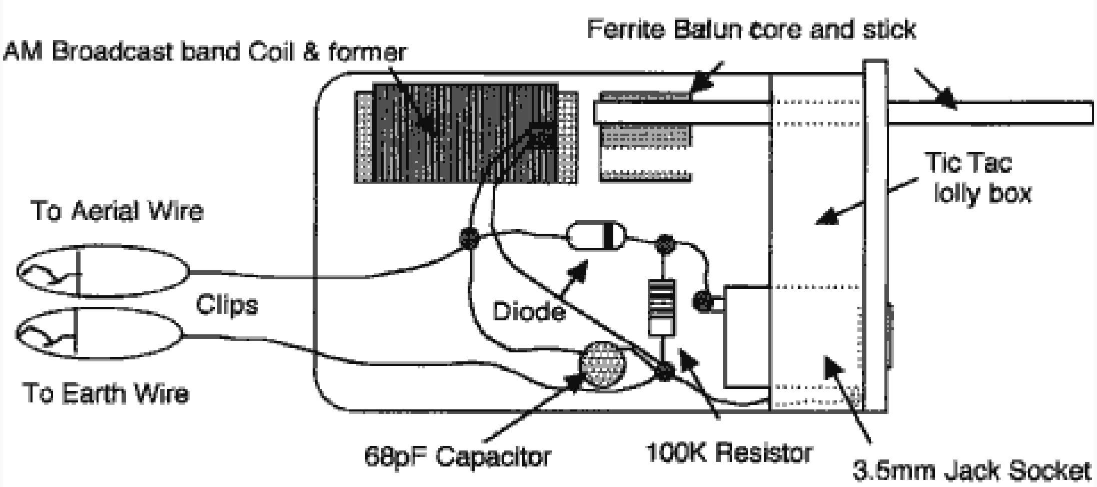
Ferrite Balun core and stick
AM Broadcast band Coil & former
Tic Tac
lolly box
To Aerial Wire
Clips
Diode
To Earth Wire
68pF Capacitor
100K Resistor
3.5mm Jack Socket

Diagram 2. Place balun inside coil former and use the stick to slide it up and down inside the coil to tune the Nic Nac Radio.

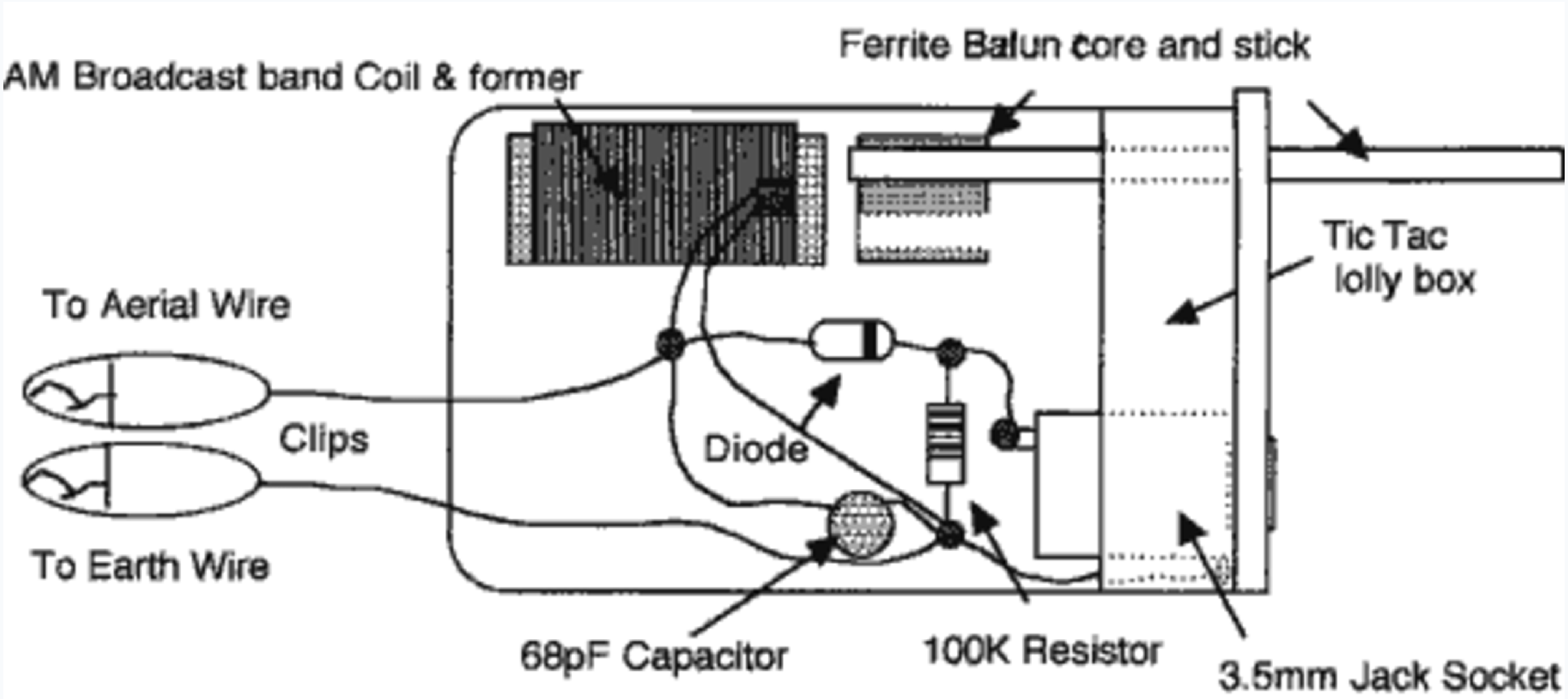
AM Broadcast band Coil & former
Ferrite Balun core and stick
Tic Tac
lolly box
To Aerial Wire
Clips
Diode
To Earth Wire
68pF Capacitor
100K Resistor
3.5mm Jack Socket

Diagram 2. Place balun inside coil former and use the stick to slide it up and down inside the coil to tune the Nic Nac Radio.

AMCI BROADCAST EQUIPMENT for VHF and UHF-TV and FM

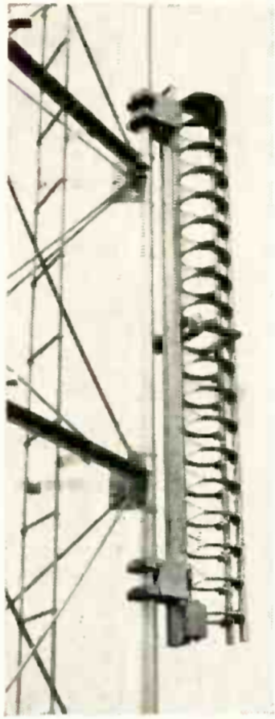
VHF-TV
Transmitting Antennas

VHF-TV TRANSMITTING ANTENNAS

Mast or tower-corner mounted
DIRECTIONAL OR OMNIDIRECTIONAL
for Channels 5 to 13

Omnidirectional VHF-TV antennas with power gains up to 20. Directional VHF-TV antennas with power gains up to 36.

UHF-TV TRANSMITTING ANTENNAS

Mast, tower-corner, or truss-mounted
DIRECTIONAL OR OMNIDIRECTIONAL
for Channels 14 to 83

Omnidirectional UHF-TV antennas with power gains up to 50. Directional UHF-TV antennas with power gains up to 80.

FM TRANSMITTING ANTENNAS

Mast or tower-corner mounted
DIRECTIONAL OR OMNIDIRECTIONAL

For single or multi-station transmission.

UHF-TV
Transmitting Antennas

SPECIAL PURPOSE ANTENNAS

The Empire State Building Master FM Antenna; two rows of dipoles around the 102nd floor observation level, the diplexers within the tower, and the transmission line, designed, built, and installed in 1965, permit up to 17 FM stations to broadcast simultaneously from the same antenna.

DIPLEXING FILTERS

Temperature compensated. Aural-to-visual and visual-to-aural rejection over 30 db. For use with transmitters up to 50 kw.

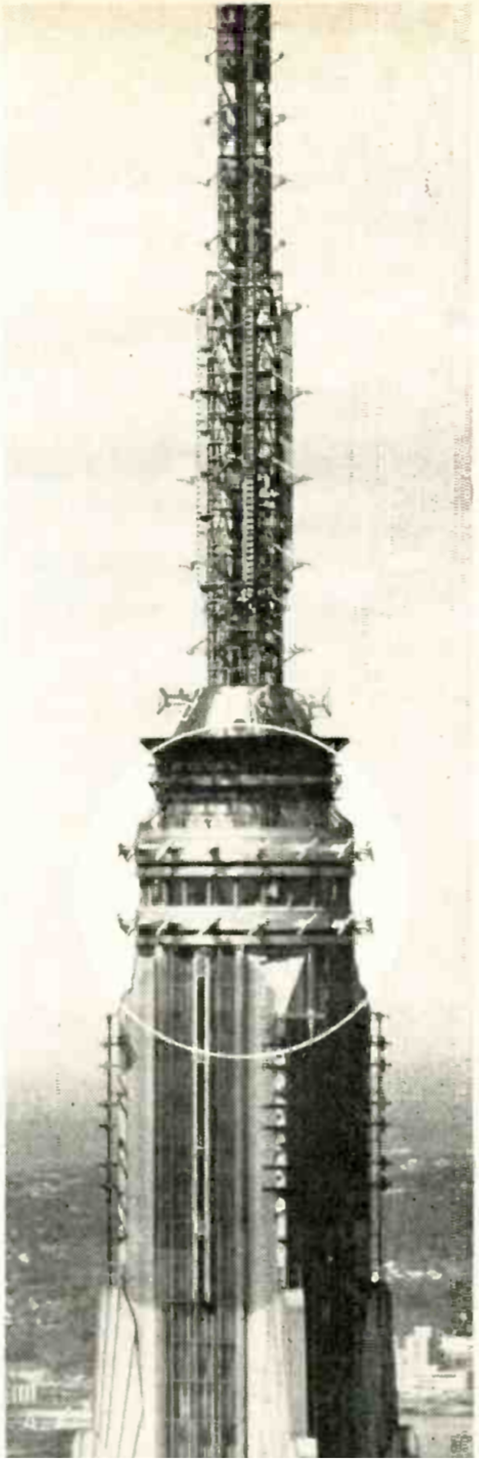

THE DE FOREST AUDION

"There is only *one* Audion—the De Forest"

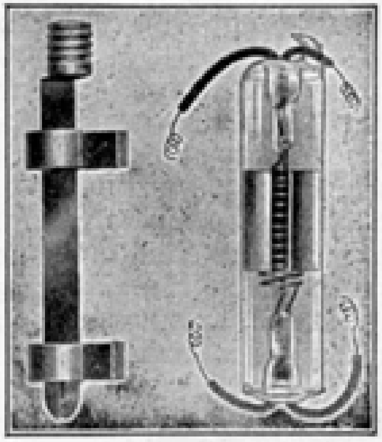

THE GENUINE DE FOREST TUBULAR AUDION

Is sold separately to any amateur who prefers to build his own Audion Detector **Price $5.50**

Adapter 40 cents extra.

Get the Bulletin (X16)

MOST SENSITIVE

The Bulletin of the U. S. Bureau of Standards states that the De Forest Audion is fully 50 per cent. more sensitive than any other known form of detector (Vol. 6, No. 4, page 540).

MOST RELIABLE

It is not affected by mechanical vibration nor burned out by static or the transmitting spark. It never fails at the critical moment.

The detector is the heart of the receiving set. Why waste valuable time on an insensitive, unreliable detector?

The genuine De Forest Audion is now within the means of every operator.

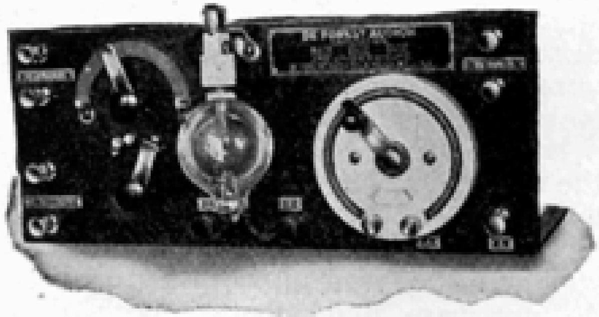

THE TYPE RJ9 DE FOREST AUDION DETECTOR

Incorporates the Audion Bulb and the genuine De Forest patented circuits with the most approved accessories needed to form a complete detector.

The most popular Audion Detector ever offered. **Price $14.00**

Get the Bulletin (M16)

WARNING—You are entitled to the genuine Audion, guaranteed by the owners of the Audion patents, when making an investment of this kind. Any evacuated detector having a filament, a grid and a plate, as well as other types, are covered by our patents, and several irresponsible infringers are being prosecuted. To be safe and get full value for your money, insist on the genuine De Forest Audion.

SEND FOR BULLETINS X16 AND M16 DESCRIBING AUDION

Detectors, Audion Amplifiers and Audion Receiving Cabinets

DE FOREST RADIO TELEPHONE & TELEGRAPH CO.

101 PARK AVENUE -:- NEW YORK, N. Y.

Makers of the Highest Grade Receiving Equipment in the World

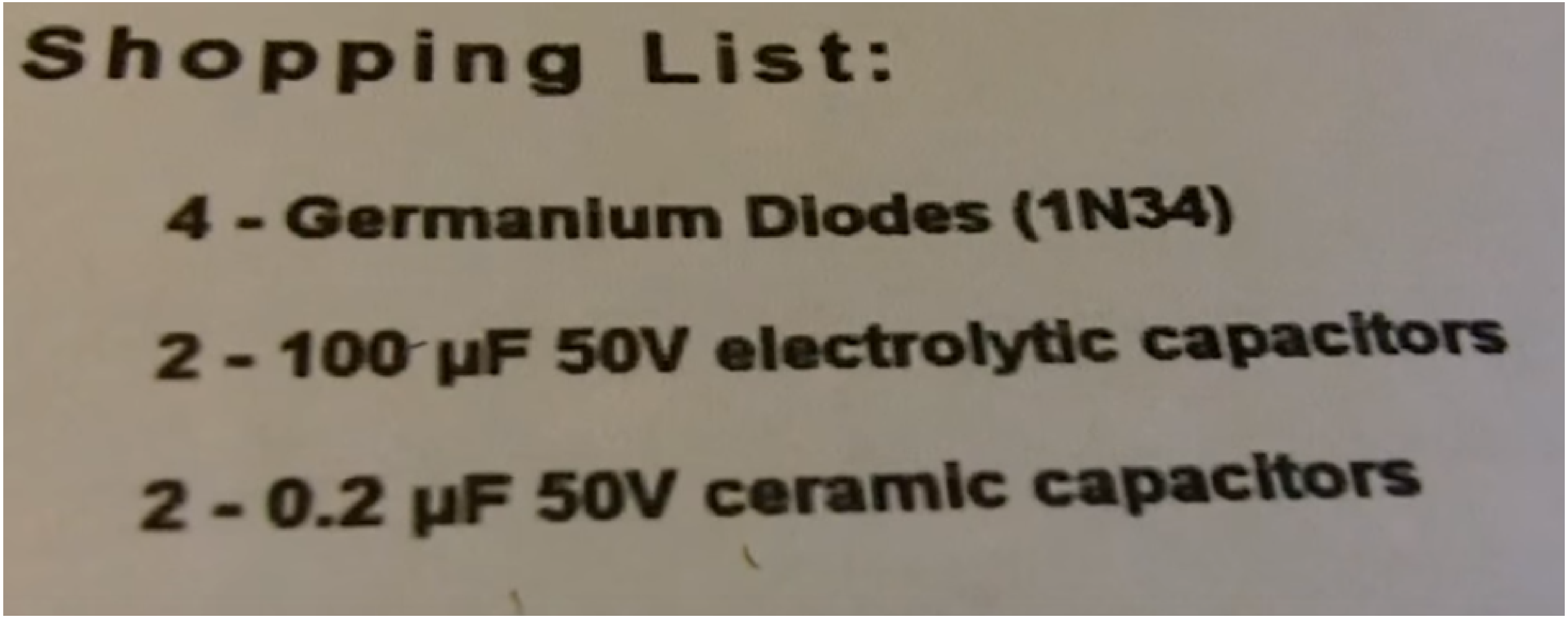

Shopping List:

4 - Germanium Diodes (1N34)

2 - 100 µF 50V electrolytic capacitors

2 - 0.2 µF 50V ceramic capacitors

LED diagram:

Electronzap.com | Electronzapdotcom (Youtube) | Electronzap (Pinterest)

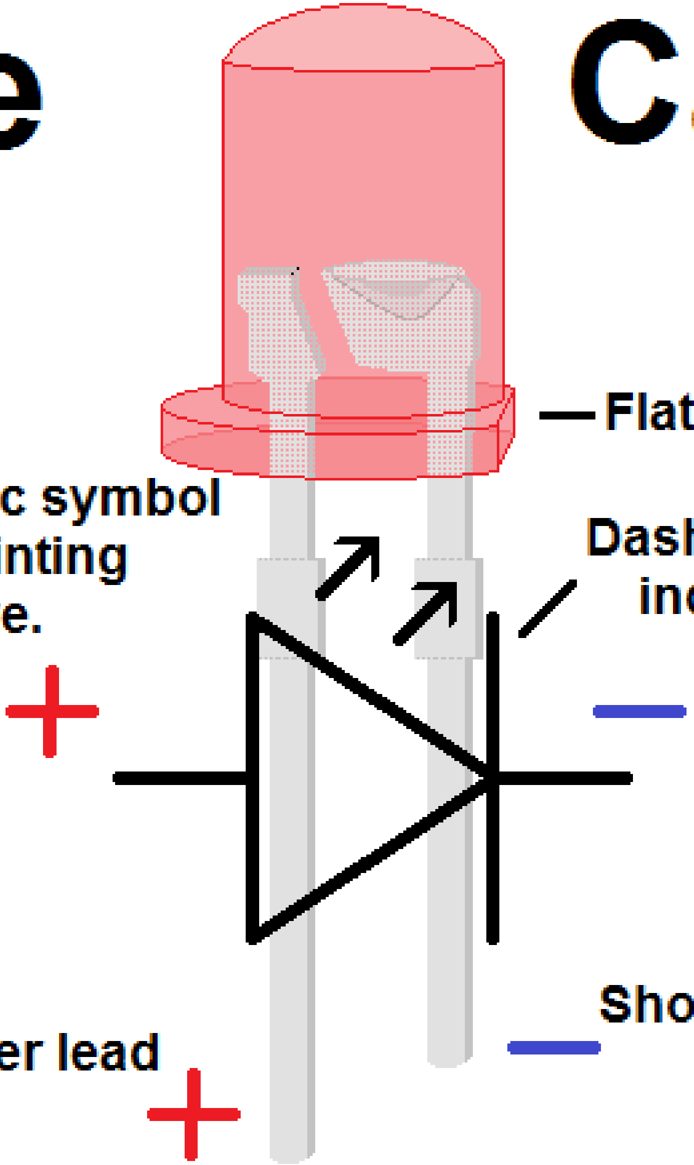
Flat edge
Longer lead
Shorter lead

Anode Side

Cathode Side

Anode side of schematic symbol resembles an arrow pointing from positive to negative. (Conventional current)

Dash side of schematic indicates cathode.

Electronzap

Electronzap

US Amateur Radio Bands

US AMATEUR POWER LIMITS

FCC 97.313 An amateur station must use the minimum transmitter power necessary to carry out the desired communications. (b) No station may transmit with a transmitter power exceeding 1.5 kW PEP.

Effective Date
March 5, 2012

Published by:
ARRL The national association for AMATEUR RADIO ®
www.arrl.org
225 Main Street, Newington, CT USA 06111-1494

160 Meters (1.8 MHz)

Avoid interference to radiolocation operations from 1.900 to 2.000 MHz

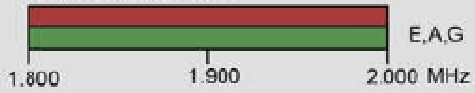

80 Meters (3.5 MHz)

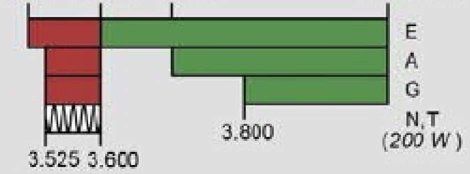

60 Meters (5.3 MHz)

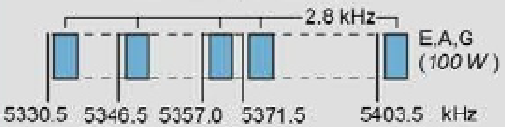

General, Advanced, and Amateur Extra licensees may operate on these five channels on a secondary basis with a maximum effective radiated output of 100 W PEP. Permitted operating modes include upper sideband voice (USB), CW, RTTY, PSK31 and other digital modes such as PACTOR III as defined by the FCC Report and Order of November 18, 2011. USB is limited to 2.8 kHz centered on 5332, 5348, 5358.5, 5373 and 5405 kHz. CW and digital emissions must be centered 1.5 kHz above the channel frequencies indicated above. Only one signal at a time is permitted on any channel.

40 Meters (7 MHz)

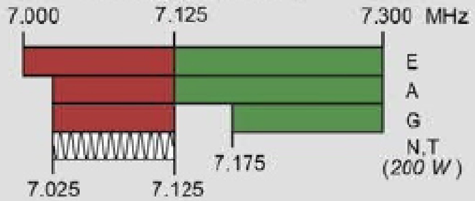

Phone and Image modes are permitted between 7.075 and 7.100 MHz for FCC licensed stations in ITU Regions 1 and 3 and by FCC licensed stations in ITU Region 2 West of 130 degrees West longitude or South of 20 degrees North latitude. See Sections 97.305(c) and 97.307(f)(11). Novice and Technician licensees outside ITU Region 2 may use CW only between 7.025 and 7.075 MHz and between 7.100 and 7.125 MHz. 7.200 to 7.300 MHz is not available outside ITU Region 2. See Section 97.301(e). These exemptions do not apply to stations in the continental US.

30 Meters (10.1 MHz)

Avoid interference to fixed services outside the US.

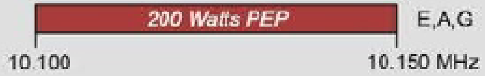

20 Meters (14 MHz)

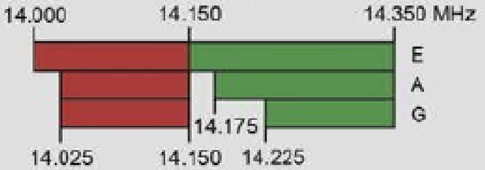

17 Meters (18 MHz)

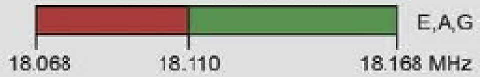

15 Meters (21 MHz)

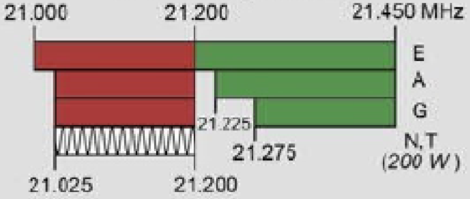

12 Meters (24 MHz)

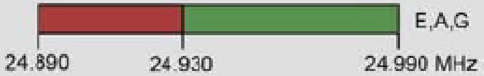

10 Meters (28 MHz)

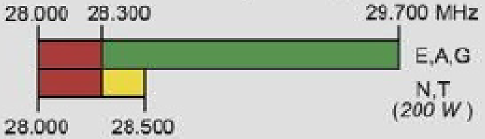

6 Meters (50 MHz)

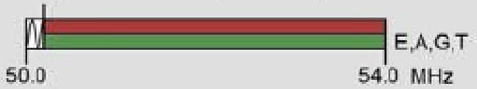

2 Meters (144 MHz)

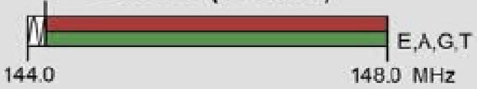

1.25 Meters (222 MHz)

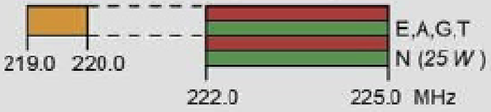

*Geographical and power restrictions may apply to all bands above 420 MHz. See *The ARRL Operating Manual* for information about your area.

70 cm (420 MHz)*

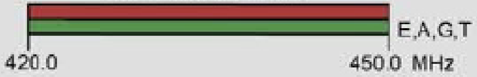

33 cm (902 MHz)*

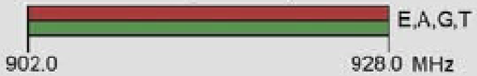

23 cm (1240 MHz)*

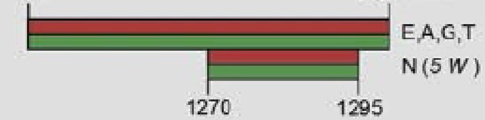

All licensees except Novices are authorized all modes on the following frequencies:

2300-2310 MHz
2390-2450 MHz
3300-3500 MHz
5650-5925 MHz
10.0-10.5 GHz *
24.0-24.25 GHz
47.0-47.2 GHz
76.0-81.0 GHz
122.25-123.0 GHz
134-141 GHz
241-250 GHz
All above 275 GHz

* No pulse emissions

KEY

Note:

CW operation is permitted throughout all amateur bands.

MCW is authorized above 50.1 MHz, except for 144.0-144.1 and 219-220 MHz.

Test transmissions are authorized above 51 MHz, except for 219-220 MHz

🟥 = RTTY and data
🟩 = phone and image
▨ = CW *only*
🟨 = SSB phone
🟦 = USB phone, CW, RTTY, and data
🟧 = Fixed digital message forwarding systems *only*

E = Amateur Extra
A = Advanced
G = General
T = Technician
N = Novice

See *ARRLWeb at www.arrl.org* for detailed band plans.

ARRL
We're At Your Service

ARRL Headquarters:
860-594-0200 (Fax 860-594-0259)
email: hq@arrl.org

Publication Orders:
www.arrl.org/shop
Toll-Free 1-888-277-5289 (860-594-0355)
email: orders@arrl.org

Membership/Circulation Desk:
www.arrl.org/membership
Toll-Free 1-888-277-5289 (860-594-0338)
email: membership@arrl.org

Getting Started in Amateur Radio:
Toll-Free 1-800-326-3942 (860-594-0355)
email: newham@arrl.org

Exams: 860-594-0300 email: vec@arrl.org

Copyright © ARRL 2012 rev. 4/9/2013

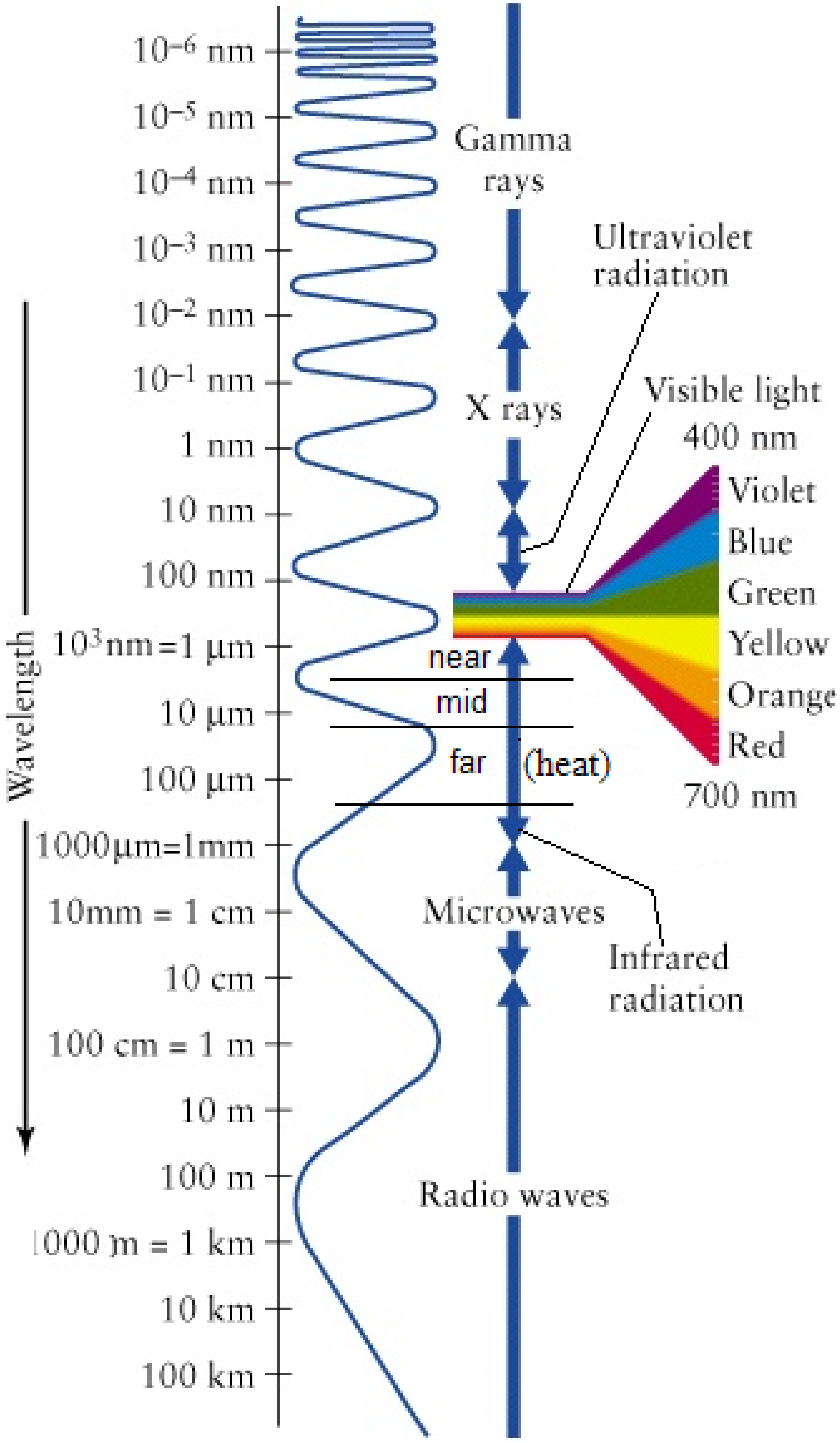
10⁻⁶ nm
10⁻⁵ nm
10⁻⁴ nm
10⁻³ nm
10⁻² nm
10⁻¹ nm
1 nm
10 nm
100 nm
10³nm=1 µm
10 µm
100 µm
1000µm=1mm
10mm = 1 cm
10 cm
100 cm = 1 m
10 m
100 m
1000 m = 1 km
10 km
100 km
Wavelength
Gamma rays
X rays
Ultraviolet radiation
Visible light
400 nm
Violet
Blue
Green
Yellow
Orange
Red
700 nm
near
mid
far
(heat)
Microwaves
Infrared radiation
Radio waves

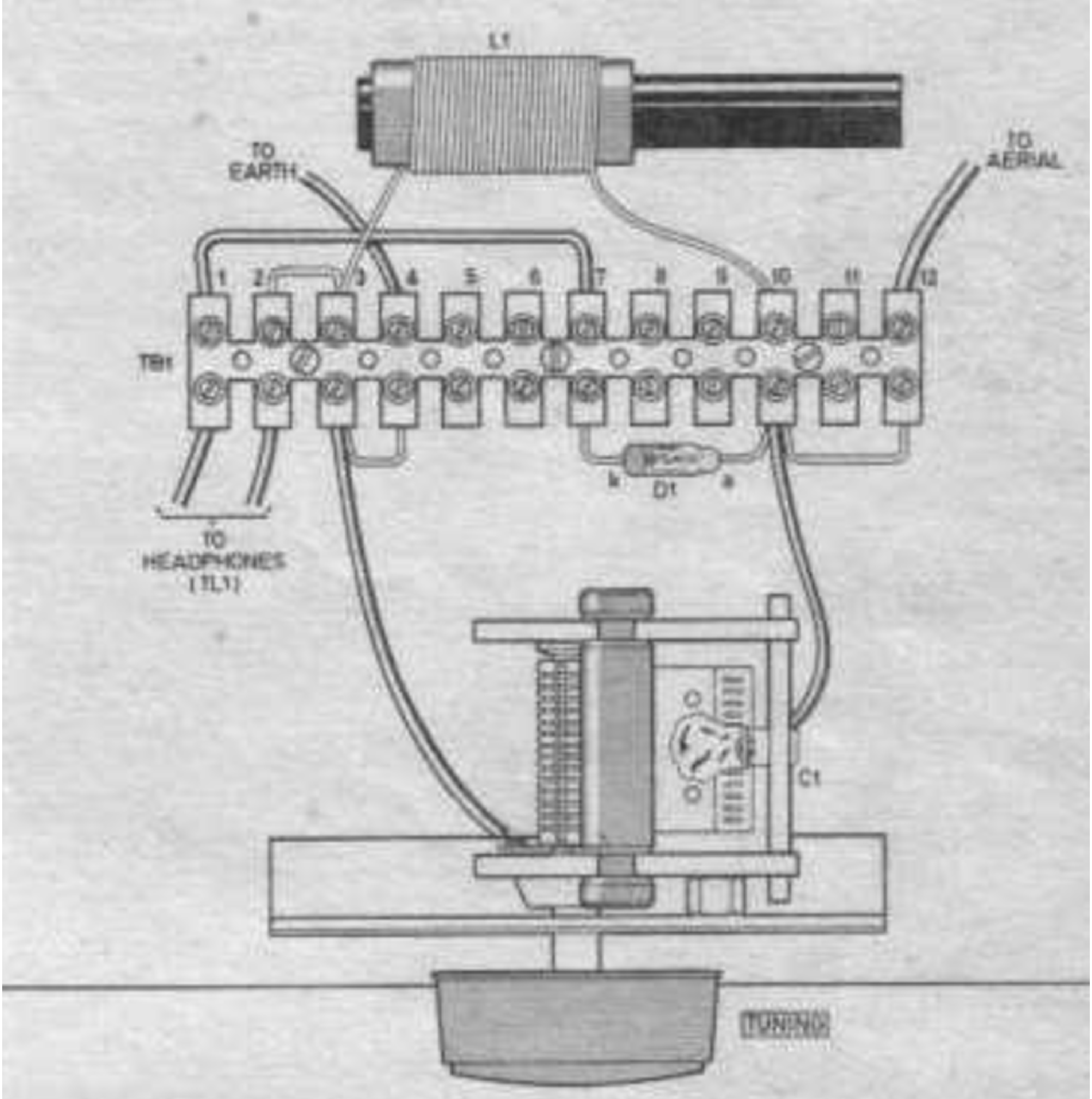
L1
TO EARTH
TO AERIAL
1
2
3
4
5
6
7
8
9
10
11
12
TB1
k
D1
a
TO HEADPHONES (TL1)
C1
TUNING

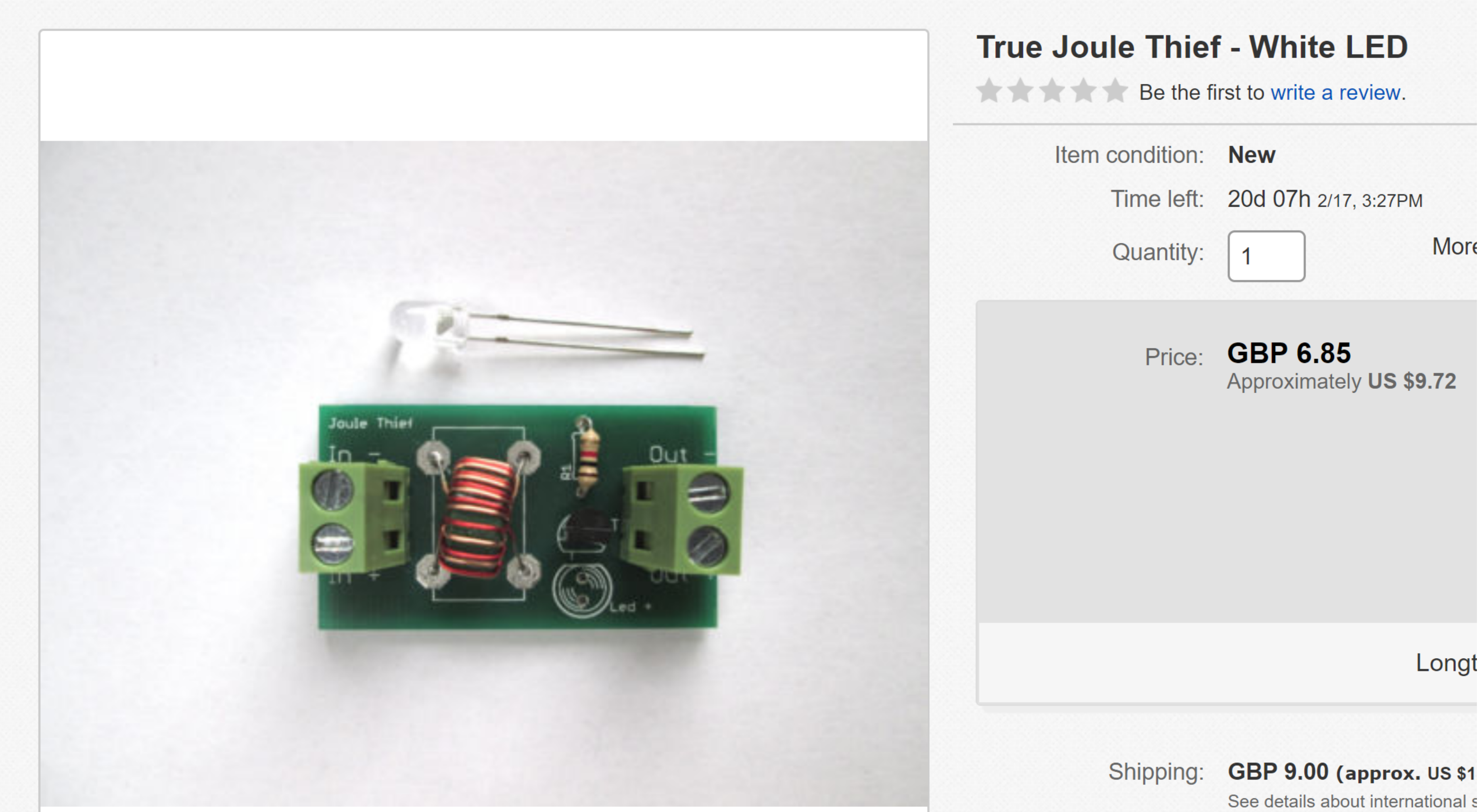

Joule Thief
In -
In +
Out -
Out
R1
T
Led +

True Joule Thief - White LED

☆☆☆☆☆ Be the first to write a review.

Item condition: **New**

Time left: 20d 07h 2/17, 3:27PM

Quantity: 1

More

Price: **GBP 6.85**
Approximately **US $9.72**

Longt

Shipping: **GBP 9.00 (approx. US $1**
See details about international s

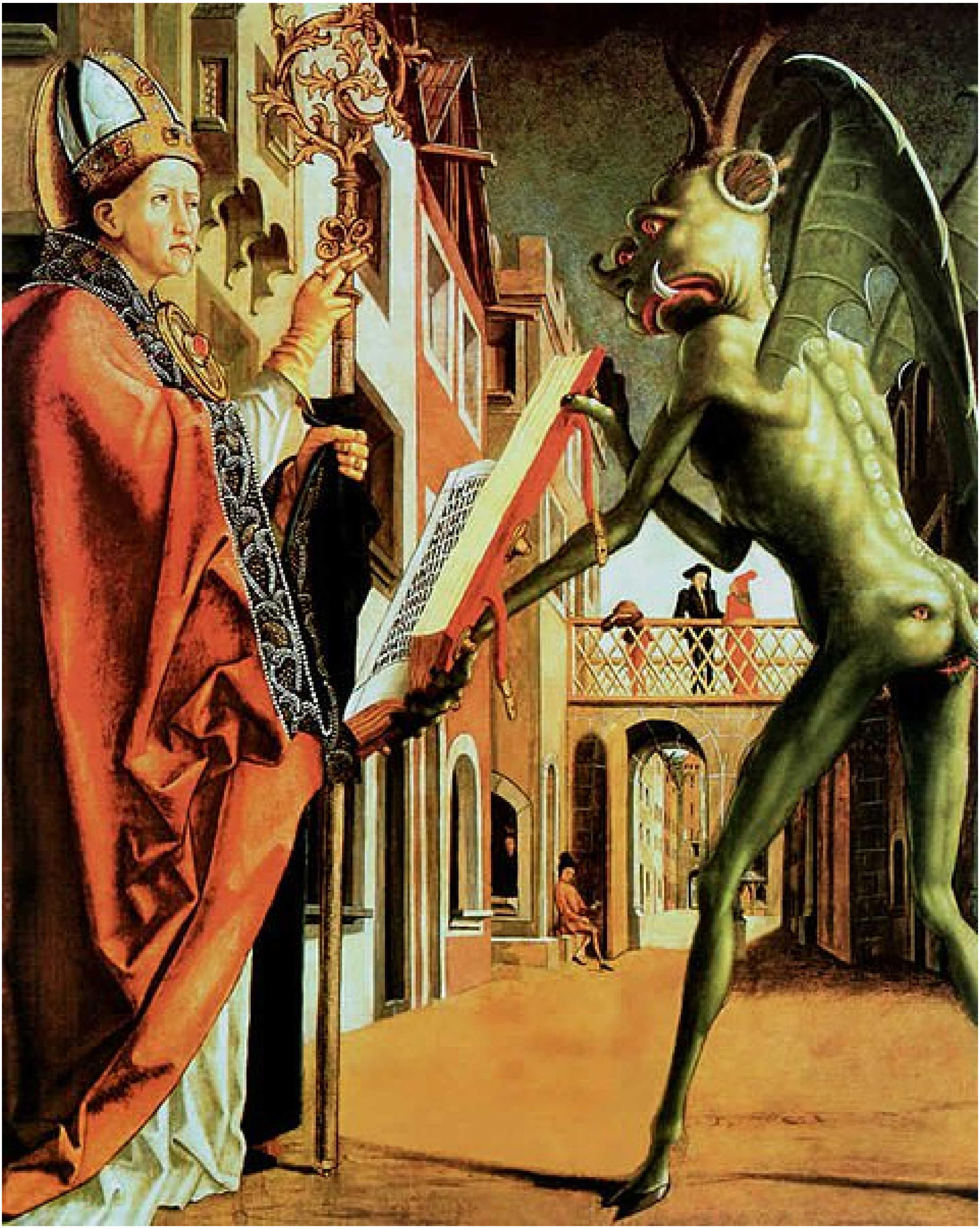

OLD WORLD ORDER - NEW WORLD ORDER

The Trinity of Globalist Control

THERE IS NO DEMOCRACY HERE

Vatican City City of London Washington D.C.

Religion Finance Military

NO FREEDOM FOR MIND OR BODY

All three are separate states,
completely independent of their respective countries.

WORLD SUPER POWERS

atruthsoldier.wordpress.com

City of London

Finance

Vatican City

Religion

Washington D.C.

Military

All three are separate states, completely independent of their respective countries.

what do these 3 cities have in common?

City of London

Washington
District of Columbia

Vatican City

- they pay no taxes
- they are under no national authority
- they have totally Independent Identities from the rest of the world

SPEAK FOR YOURSELF

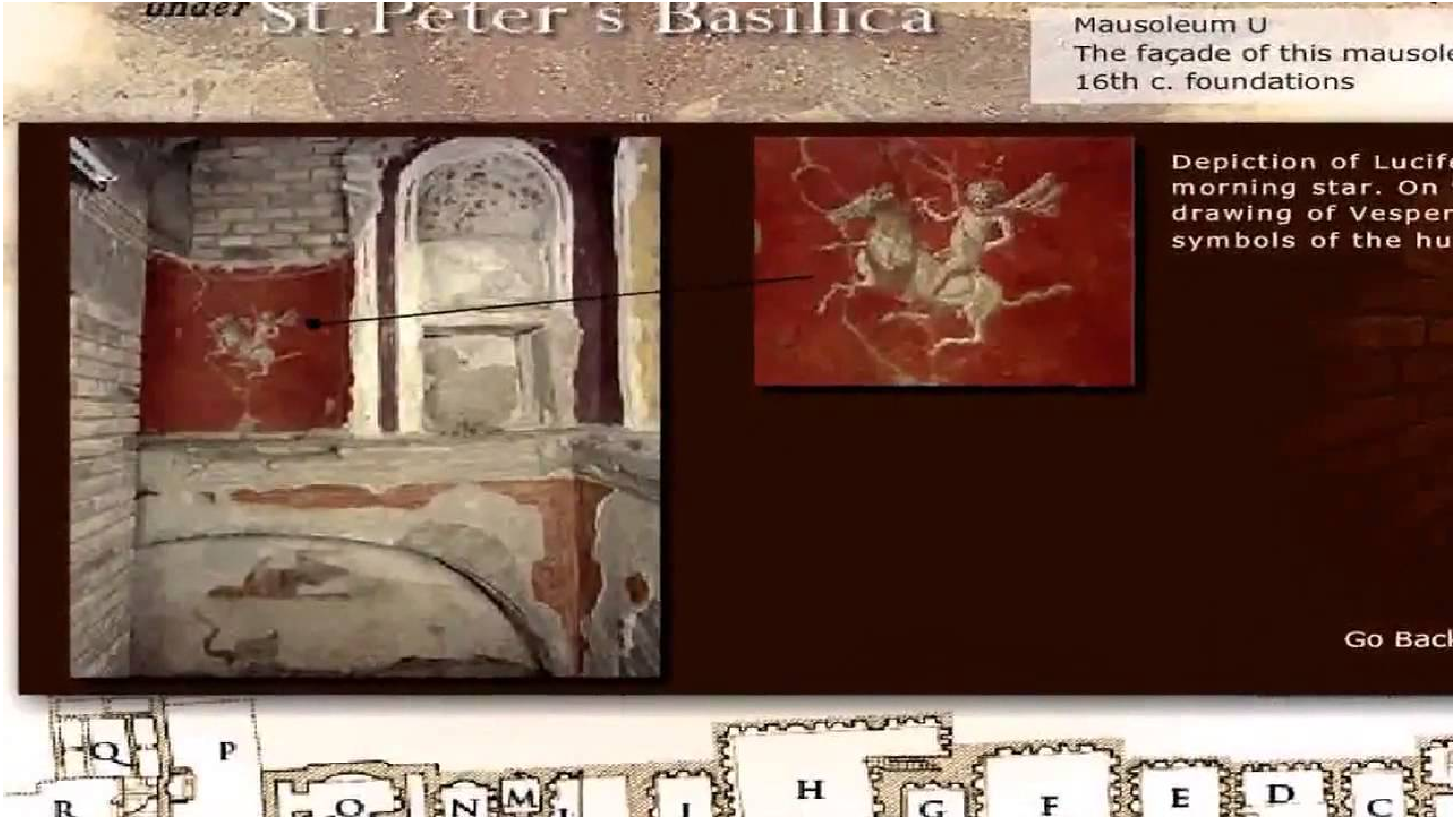

under St. Peter's Basilica

Mausoleum U
The façade of this mausole
16th c. foundations

Depiction of Lucif
morning star. On
drawing of Vesper
symbols of the hu

Go Back

POLARSTERN

ET OVES
ΣΥ ΒΟΣΚΕΙΣ ΤΑ

ELF, GWEN Towers, and HAARP Connection
These things are pretty sinister and all over the U.S.
Click "see more" to read article

LIGO TE
TV ES PETRVS ET SVPER
NC VNA FIDES
HINC SACERDO
ET O PETRE DIXISTI TV ES
VIVAT IESVS BEATVS
ΣΥ ΒΟΣΚΕΙΣ ΤΑ

THE BLACK CUBE OF SATURN

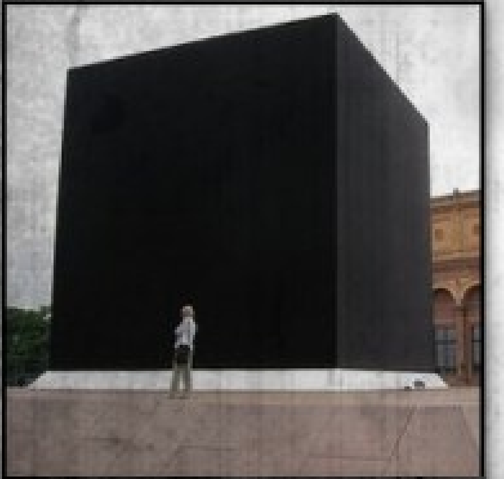

AUSTRALIA

DENMARK

MANHATTAN

SANTA ANA

GLOBAL OCCULT WORSHIP

GlacierDogMapMe: Pantheon, Vatican, Washington DC - What Were They Thinking!?

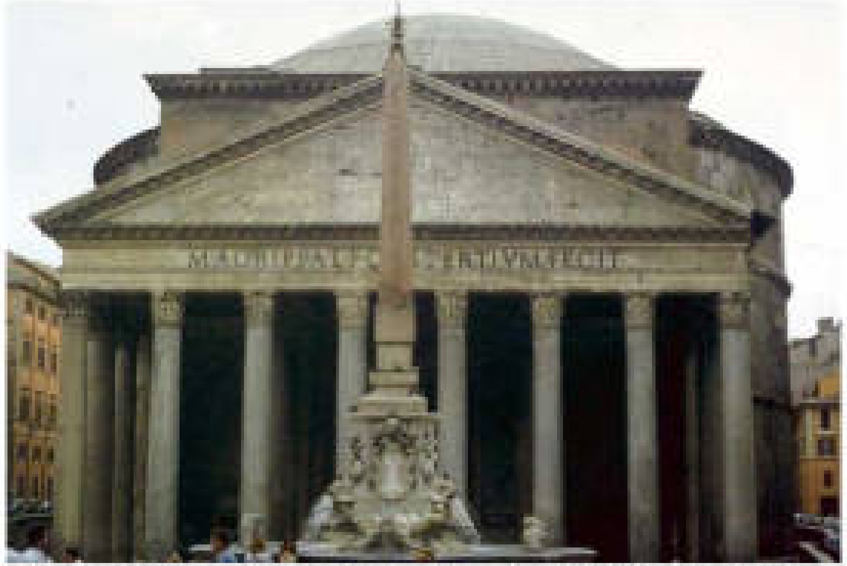

Roman Pantheon Dome facing Obelisk

http://www.historum.com/members/labienus-albums-roman-civilization-picture193-roman-pantheon.jpg

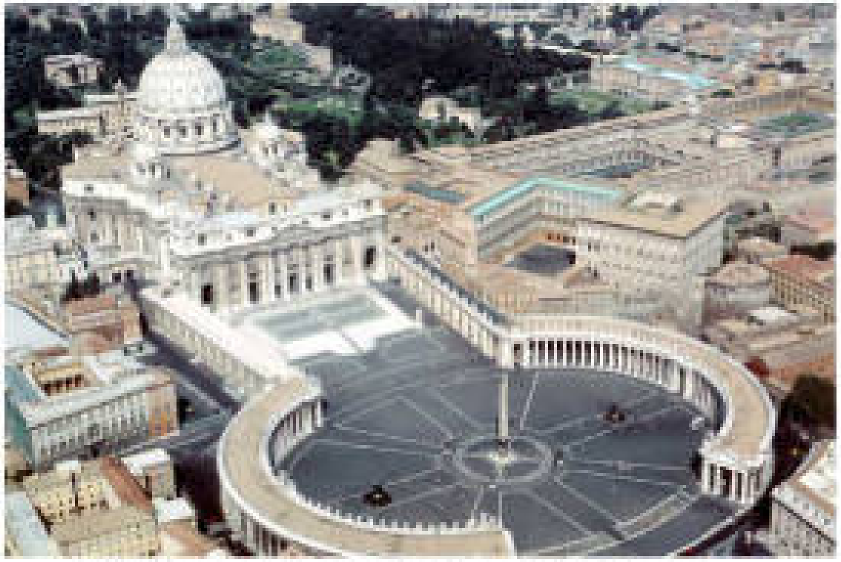

Vatican Dome facing Obelisk

http://www.raidersnewsupdate.com/leadstory70.htm

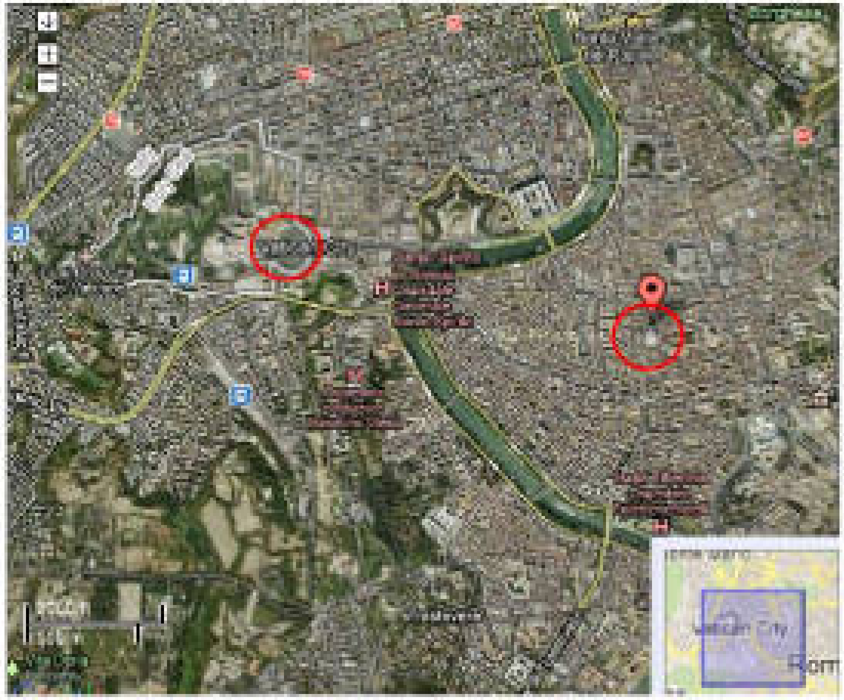

Above: Vatican City & Pantheon (red circles)

http://www.rome.info/map/

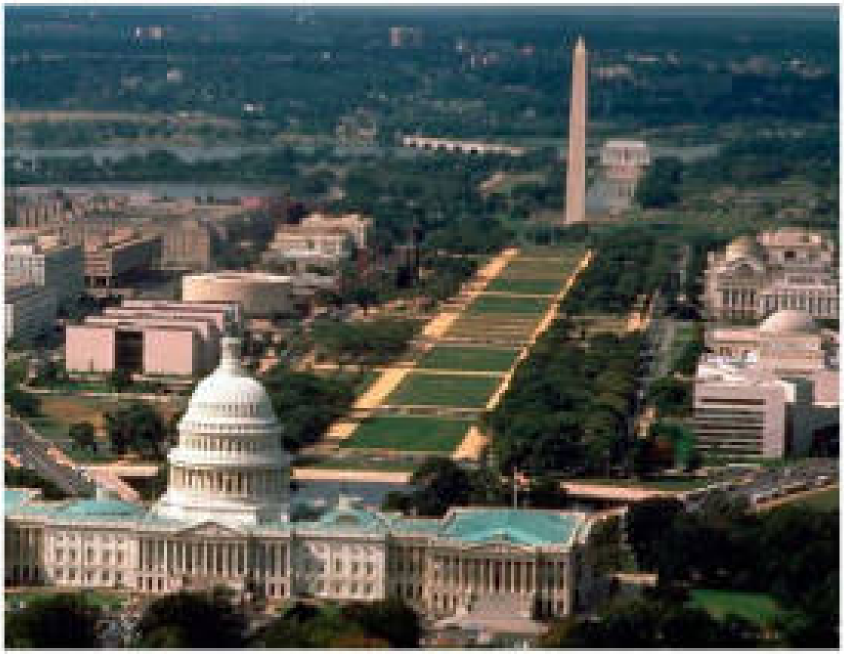

Washington DC Dome facing Obelisk

http://www.raidersnewsupdate.com/leadstory70.htm

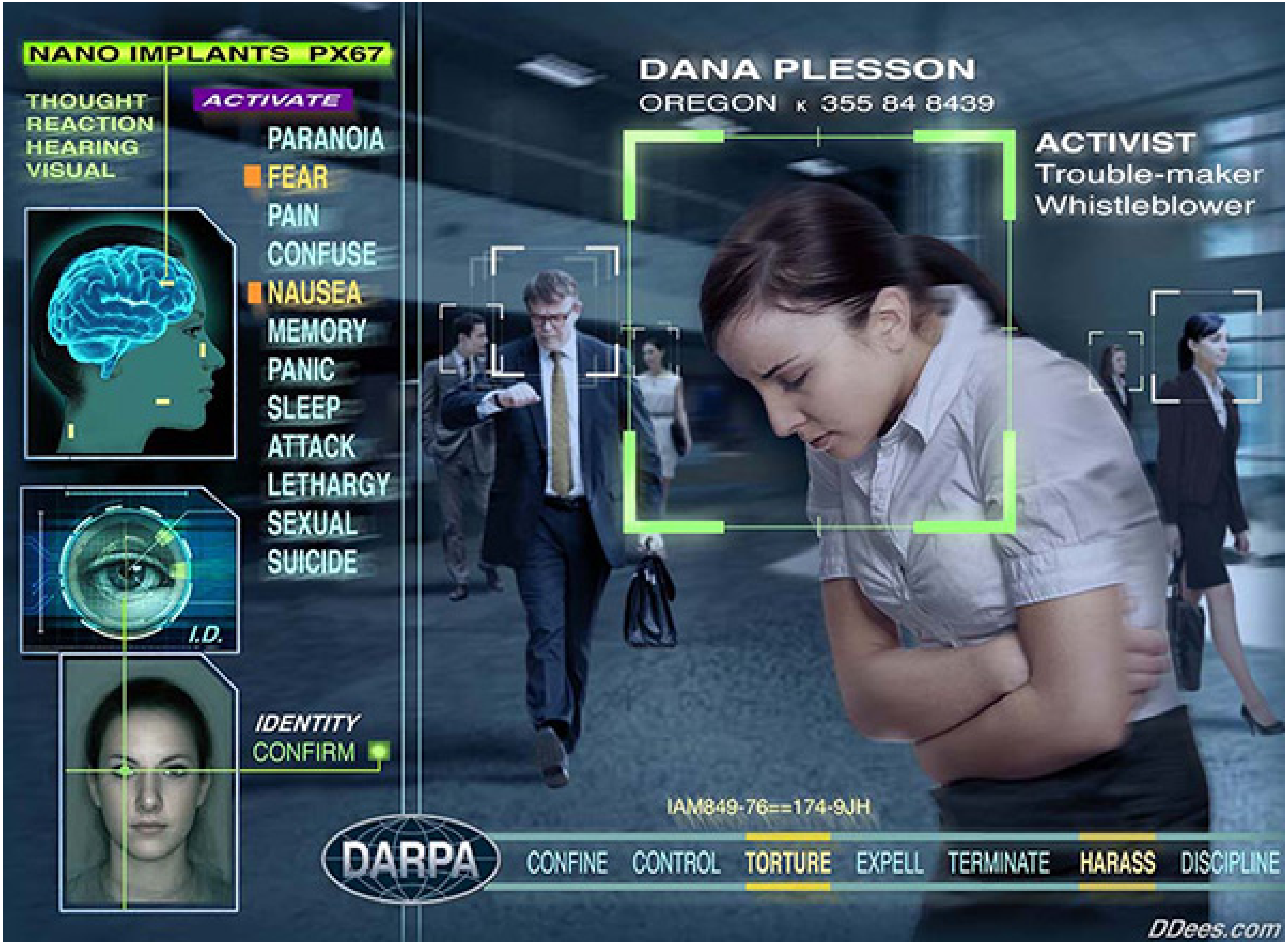
NANO IMPLANTS PX67
THOUGHT
REACTION
HEARING
VISUAL
ACTIVATE
PARANOIA
FEAR
PAIN
CONFUSE
NAUSEA
MEMORY
PANIC
SLEEP
ATTACK
LETHARGY
SEXUAL
SUICIDE
I.D.
IDENTITY
CONFIRM
DANA PLESSON
OREGON κ 355 84 8439
ACTIVIST
Trouble-maker
Whistleblower
IAM849-76==174-9JH
DARPA
CONFINE
CONTROL
TORTURE
EXPELL
TERMINATE
HARASS
DISCIPLINE
DDees.com

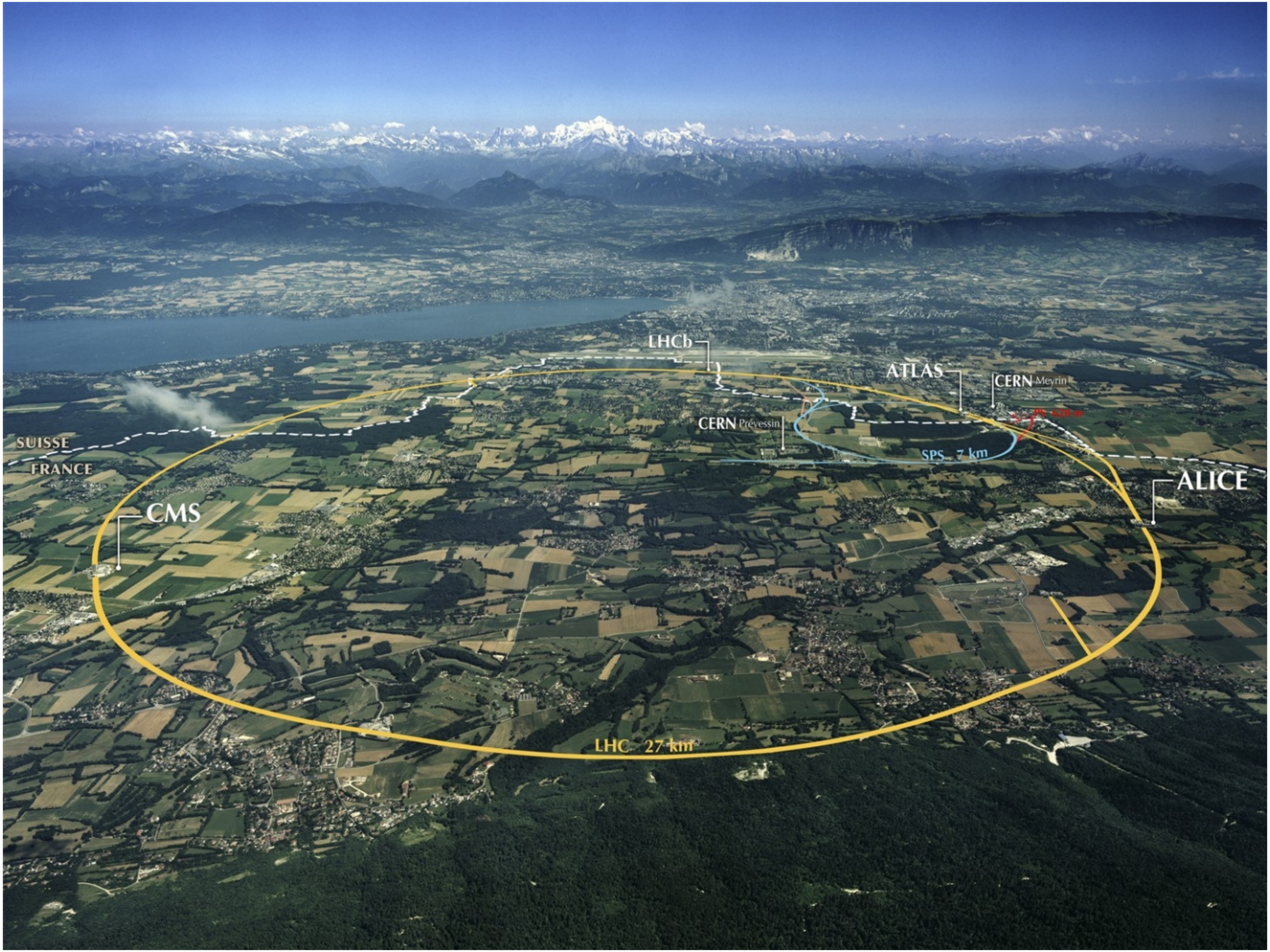
LHCb
ATLAS
CERN Meyrin
PS 628 m
CERN Prévessin
SPS 7 km
SUISSE
FRANCE
ALICE
CMS
LHC 27 km

An overview of how the weapons are deployed by the Mind Control Police

Mass Remote Mind Control by Blanket Coverage of the Population which is achieved via the Mobile Phone Network

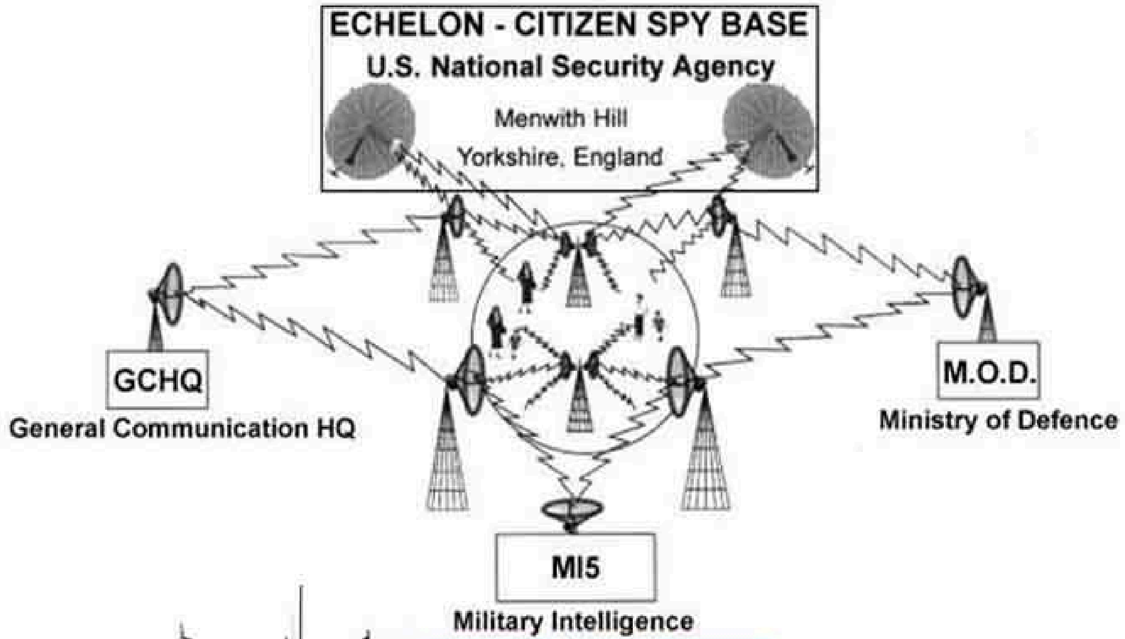
ECHELON - CITIZEN SPY BASE
U.S. National Security Agency
Menwith Hill
Yorkshire, England
GCHQ
General Communication HQ
M.O.D.
Ministry of Defence
MI5
Military Intelligence

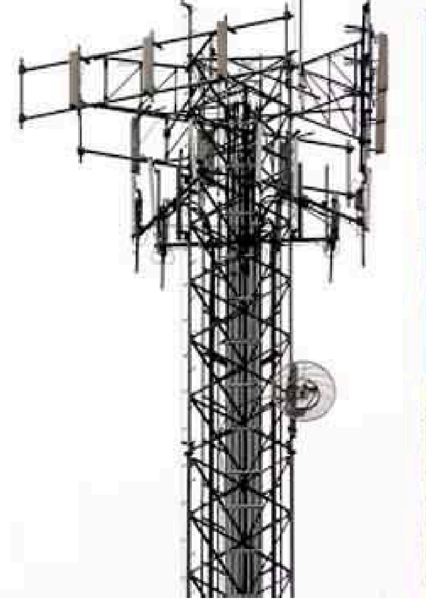

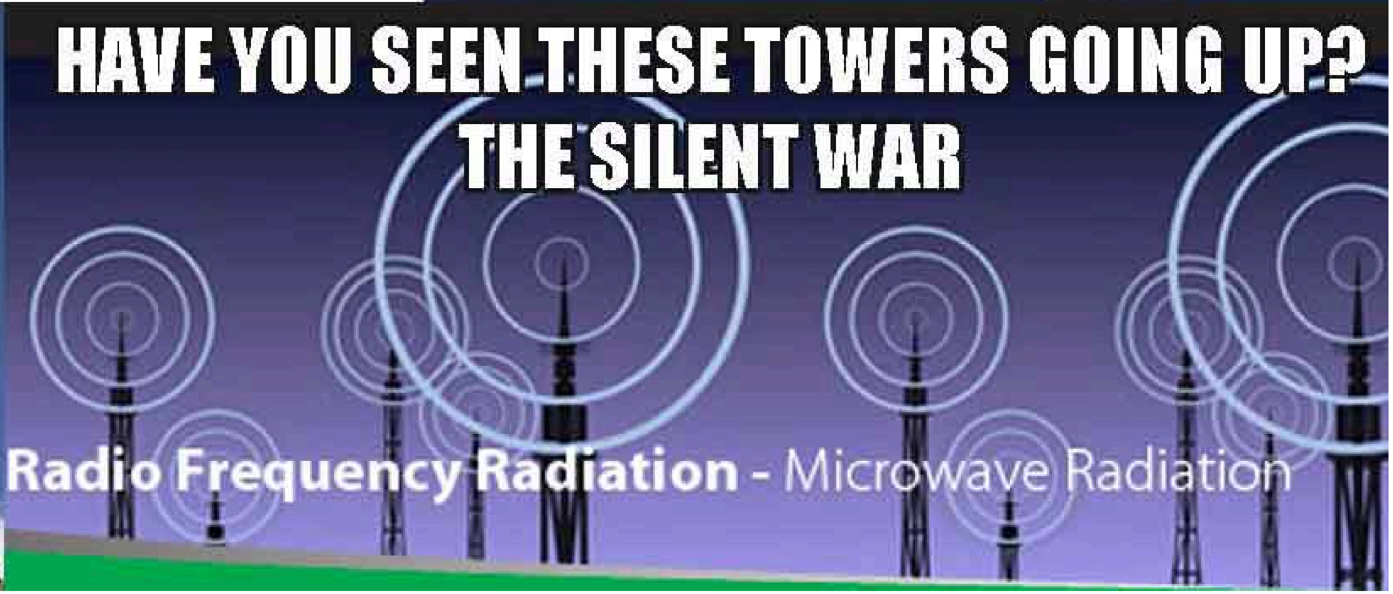
HAVE YOU SEEN THESE TOWERS GOING UP?
THE SILENT WAR
Radio Frequency Radiation - Microwave Radiation

BLACK CUBE- AUSTRALIA

Manhattan, NYC

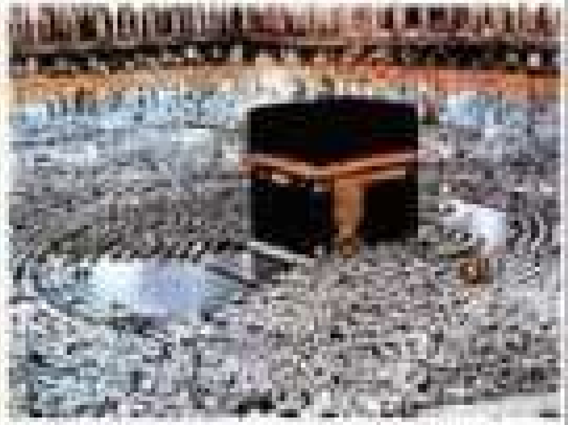

Kaba in Mecca

BLACK CUBE- DENMARK

Aple Store cube- NYC

"Rings going around SATURN"

Download from
Dreamstime.com
This watermarked comp image is for previewing purposes only.

ID 59211679
© Marinv | Dreamstime.com

Mu Radar
Arecibo
HAARP
HIPAS
EISCAT
Sura

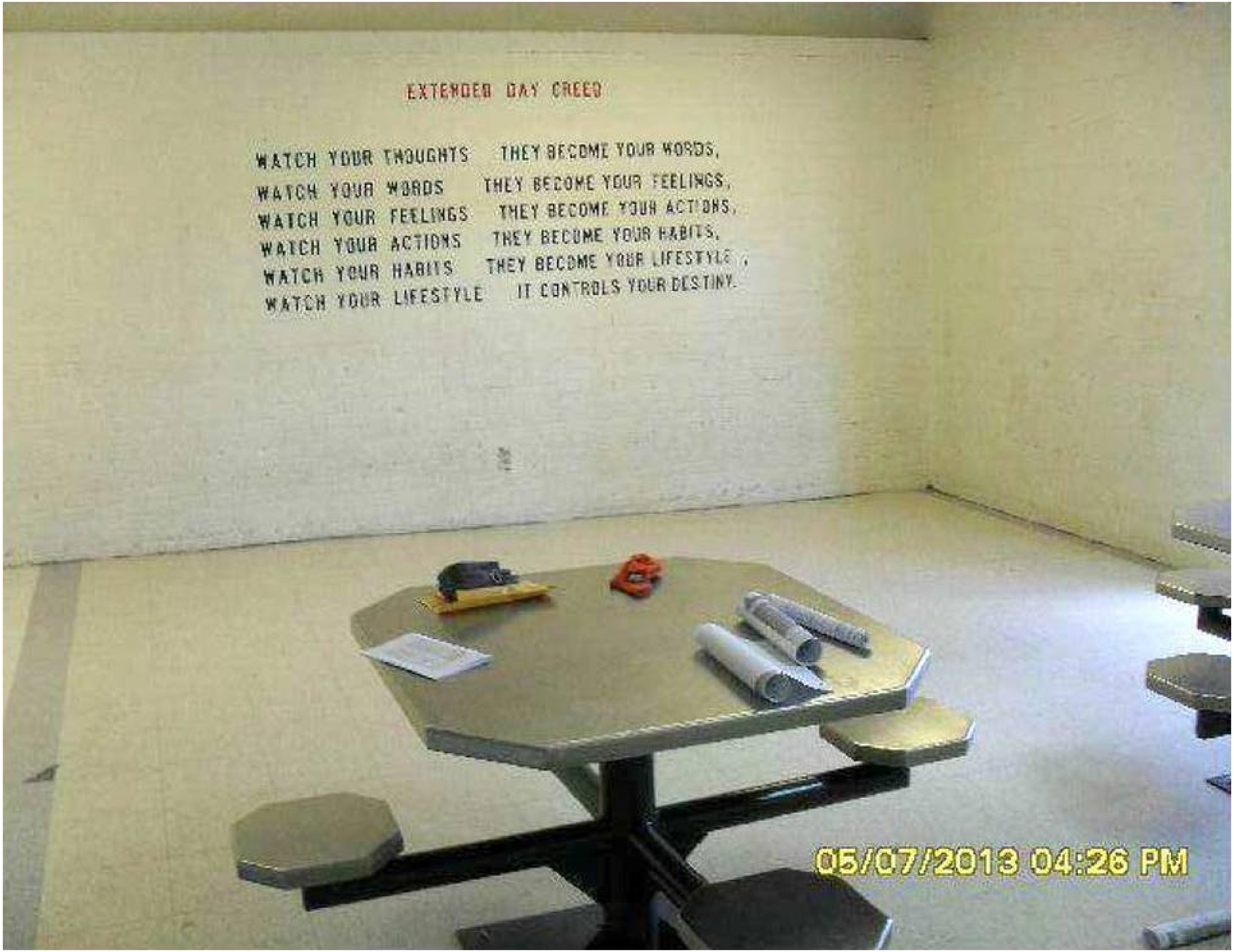
EXTENDED DAY CREED
WATCH YOUR THOUGHTS THEY BECOME YOUR WORDS,
WATCH YOUR WORDS THEY BECOME YOUR FEELINGS,
WATCH YOUR FEELINGS THEY BECOME YOUR ACTIONS,
WATCH YOUR ACTIONS THEY BECOME YOUR HABITS,
WATCH YOUR HABITS THEY BECOME YOUR LIFESTYLE,
WATCH YOUR LIFESTYLE IT CONTROLS YOUR DESTINY.
05/07/2013 04:26 PM

UNITED STATES OF AMERICA
E PLURIBUS UNUM
ONE CENT

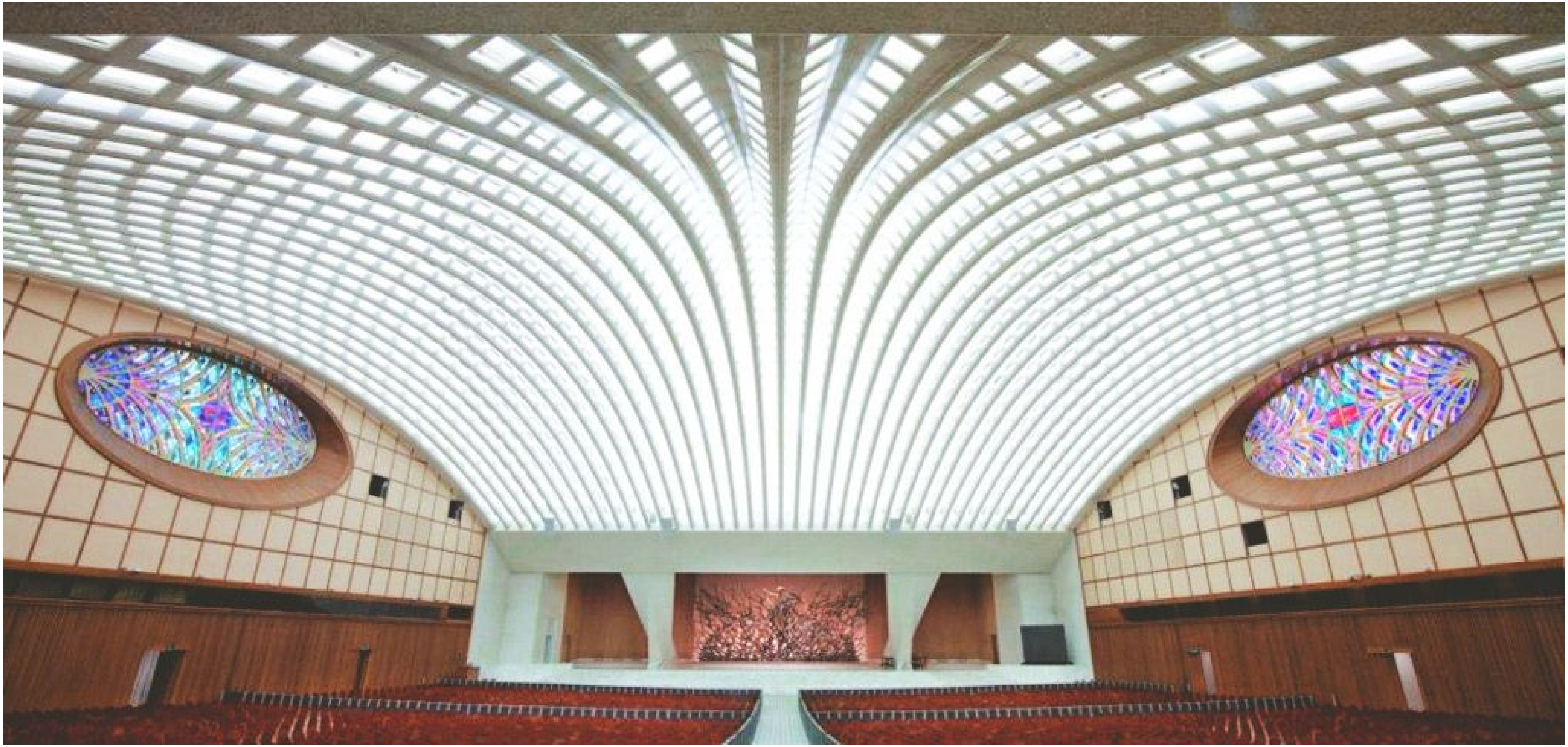

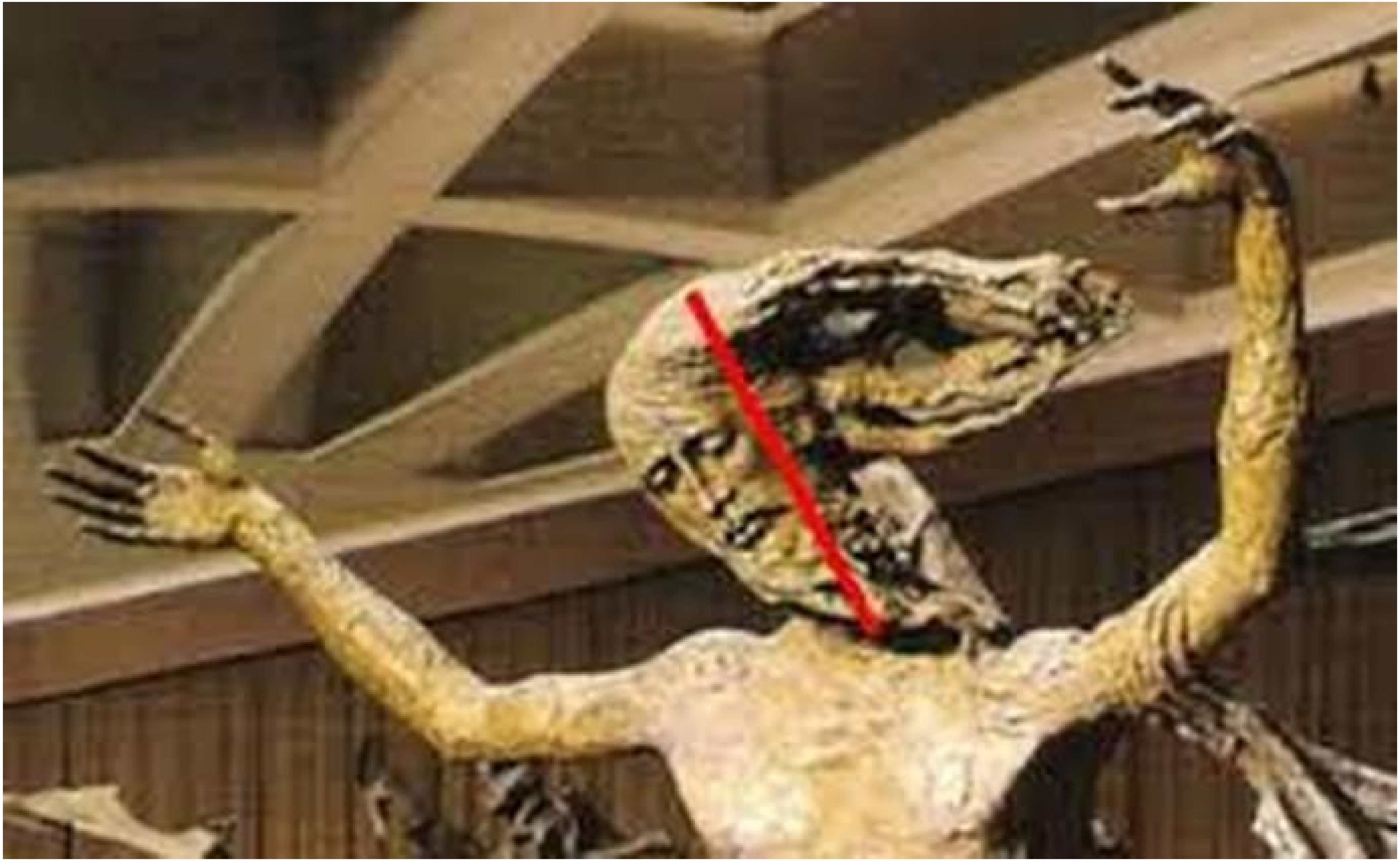

The Trinity of Globalist Control

Vatican City

Religion

City of London

Finance

Washington D.C.

Military

**All three are separate states,
completely independent of their respective countries.**

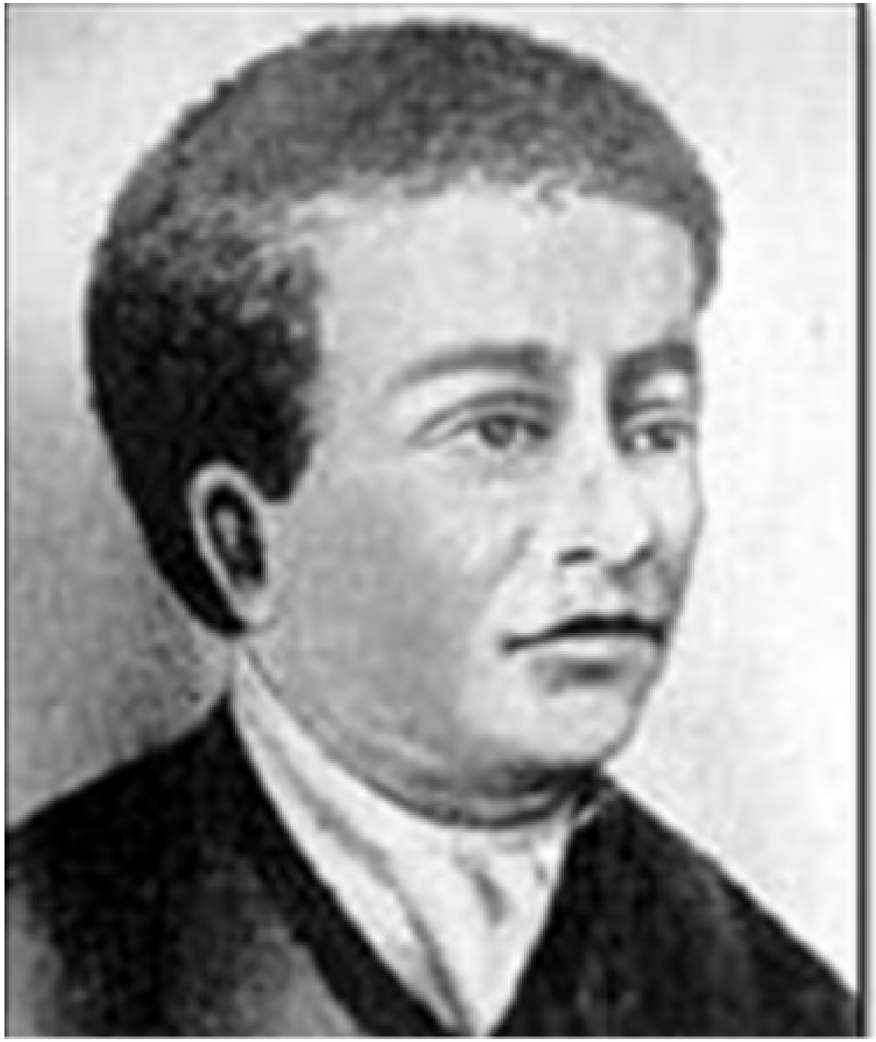

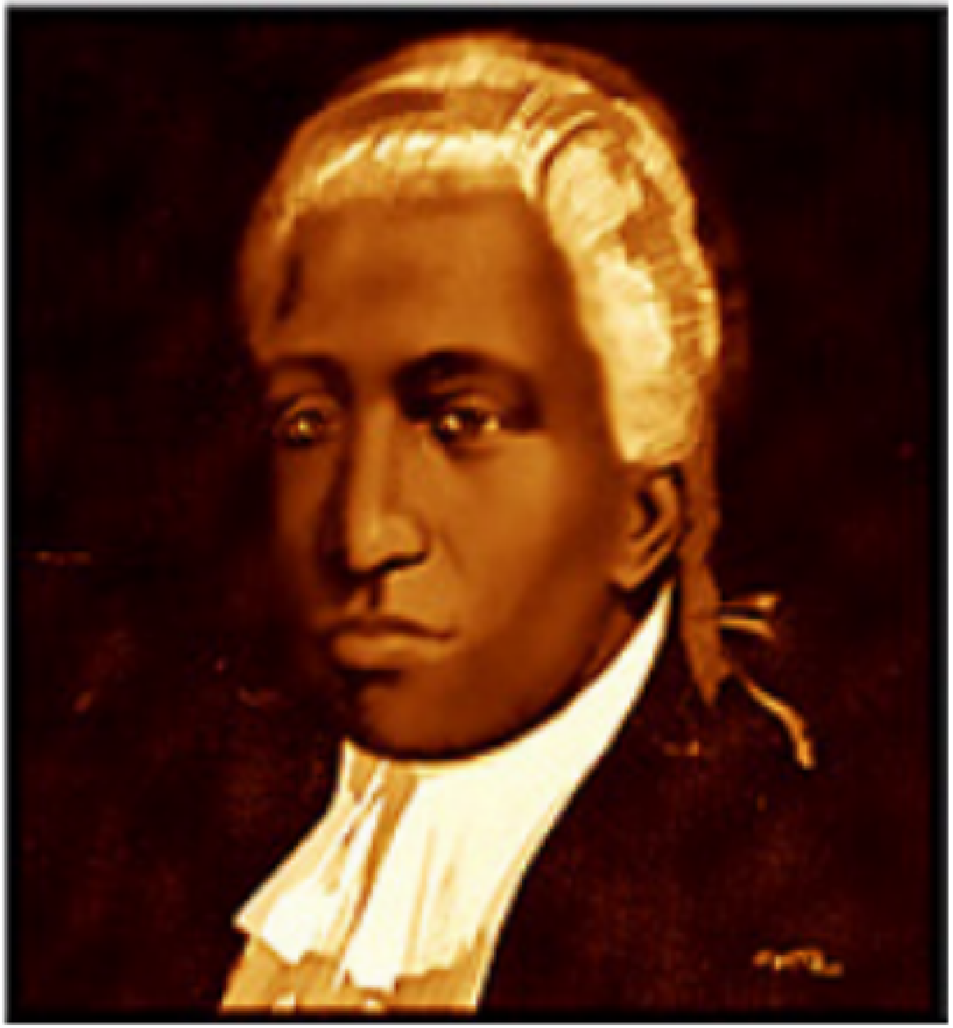

THE GREATEST OPERATIVE MASON MEETS THE GREATEST SPECTULATIVE MASON is Ben Bey Emmanuel Mu Ali, historically known as Benjamin Banneker really Prince Hall? Benjamin Banneker (November 9, 1731 – October 9, 1806) was a free African American astronomer, mathematician, surveyor, almanac author and farmer. Prince Hall (c.1735 – December 4, 1807), Hall formed the African Grand Lodge of North America.

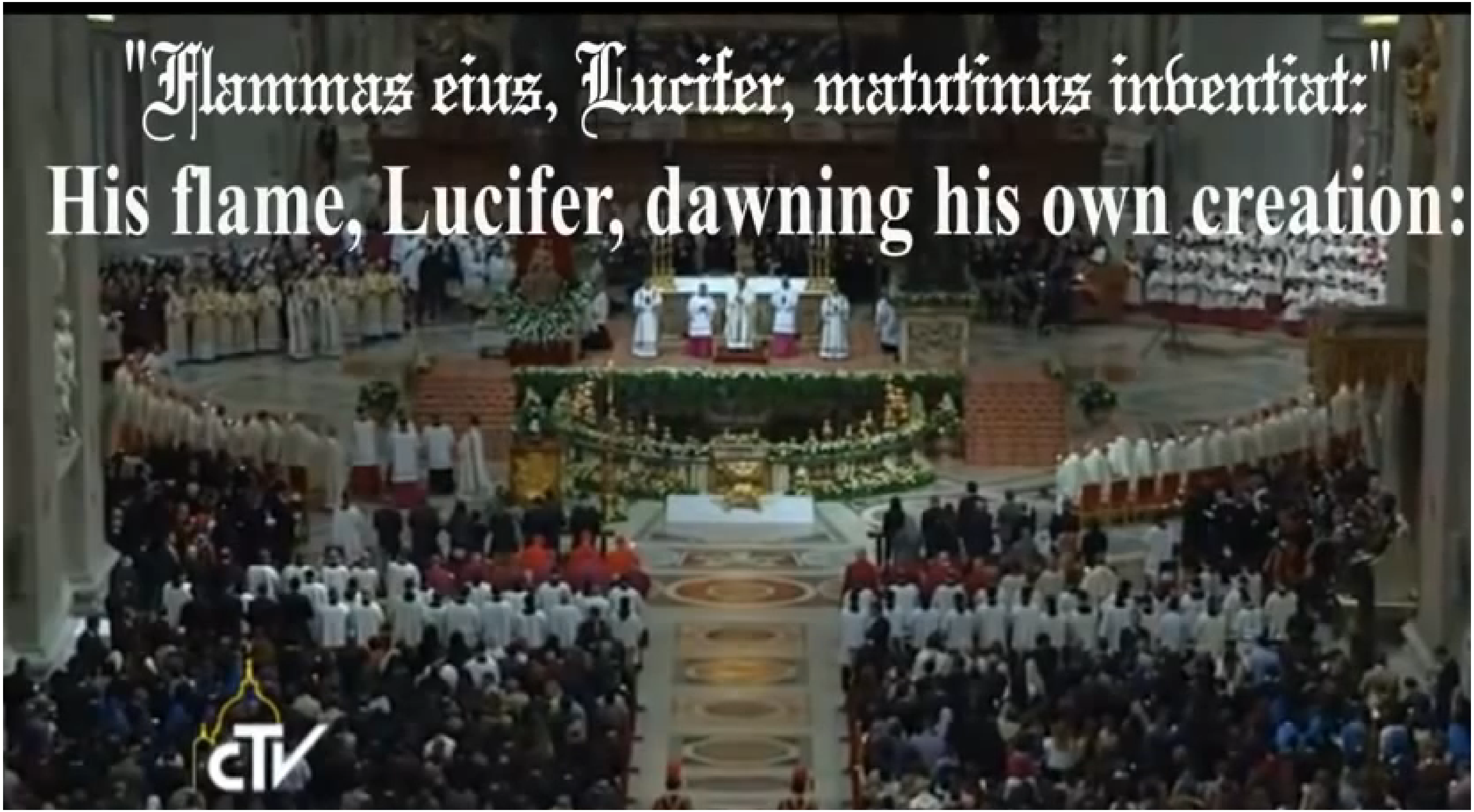

"Flammas eius, Lucifer, matutinus inveniat:"

His flame, Lucifer, dawning his own creation:

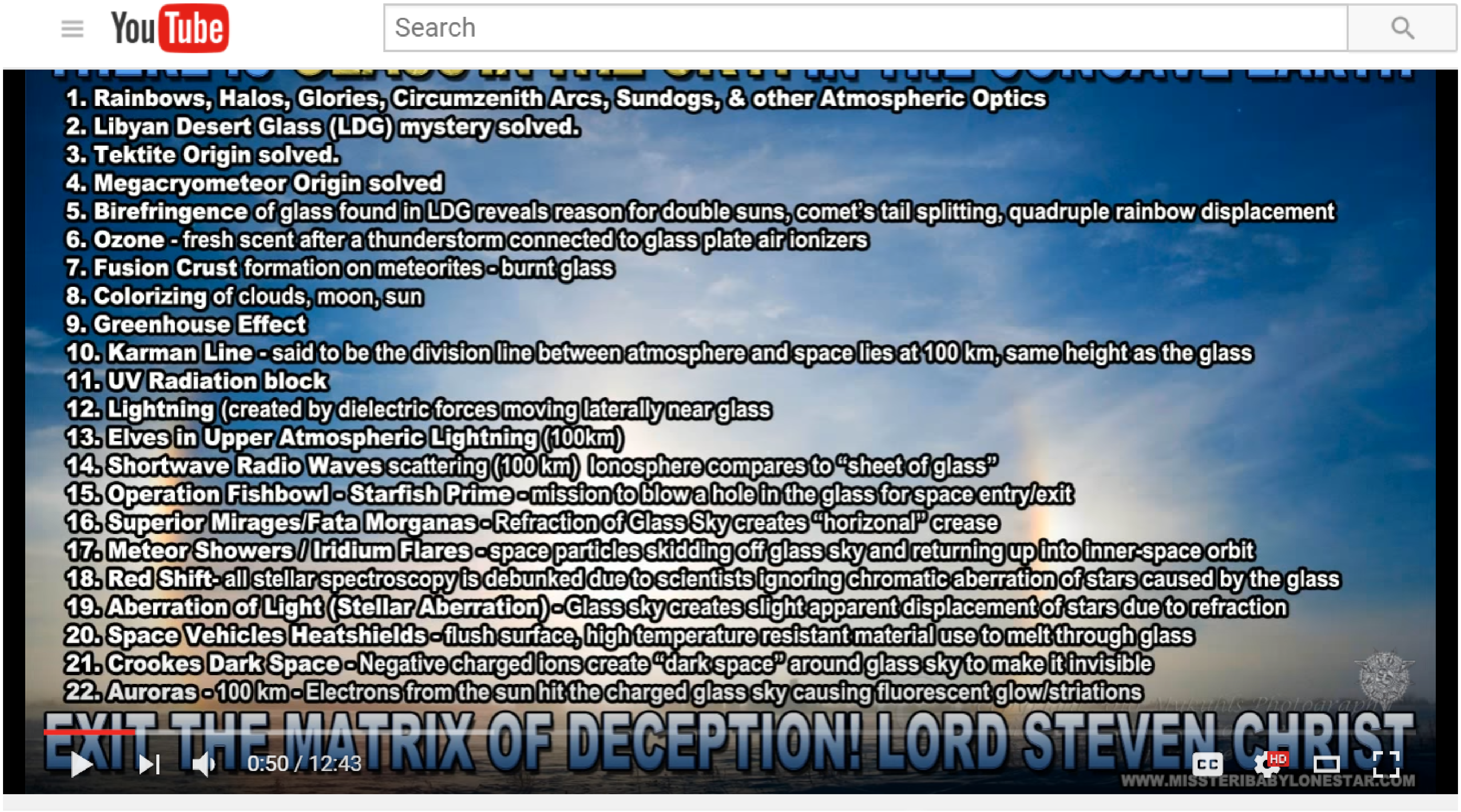

1. **Rainbows, Halos, Glories, Circumzenith Arcs, Sundogs, & other Atmospheric Optics**
2. **Libyan Desert Glass (LDG) mystery solved.**
3. **Tektite Origin solved.**
4. **Megacryometeor Origin solved**
5. **Birefringence** of glass found in LDG reveals reason for double suns, comet's tail splitting, quadruple rainbow displacement
6. **Ozone** - fresh scent after a thunderstorm connected to glass plate air ionizers
7. **Fusion Crust** formation on meteorites - burnt glass
8. **Colorizing** of clouds, moon, sun
9. **Greenhouse Effect**
10. **Karman Line** - said to be the division line between atmosphere and space lies at 100 km, same height as the glass
11. **UV Radiation block**
12. **Lightning** (created by dielectric forces moving laterally near glass
13. **Elves in Upper Atmospheric Lightning** (100km)
14. **Shortwave Radio Waves** scattering (100 km) Ionosphere compares to "sheet of glass"
15. **Operation Fishbowl - Starfish Prime** - mission to blow a hole in the glass for space entry/exit
16. **Superior Mirages/Fata Morganas** - Refraction of Glass Sky creates "horizonal" crease
17. **Meteor Showers / Iridium Flares** - space particles skidding off glass sky and returning up into inner-space orbit
18. **Red Shift**- all stellar spectroscopy is debunked due to scientists ignoring chromatic aberration of stars caused by the glass
19. **Aberration of Light (Stellar Aberration)** - Glass sky creates slight apparent displacement of stars due to refraction
20. **Space Vehicles Heatshields** - flush surface, high temperature resistant material use to melt through glass
21. **Crookes Dark Space** - Negative charged ions create "dark space" around glass sky to make it invisible
22. **Auroras** - 100 km - Electrons from the sun hit the charged glass sky causing fluorescent glow/striations

EXIT THE MATRIX OF DECEPTION! LORD STEVEN CHRIST

WWW.MISSTERIBABYLONESTAR.COM

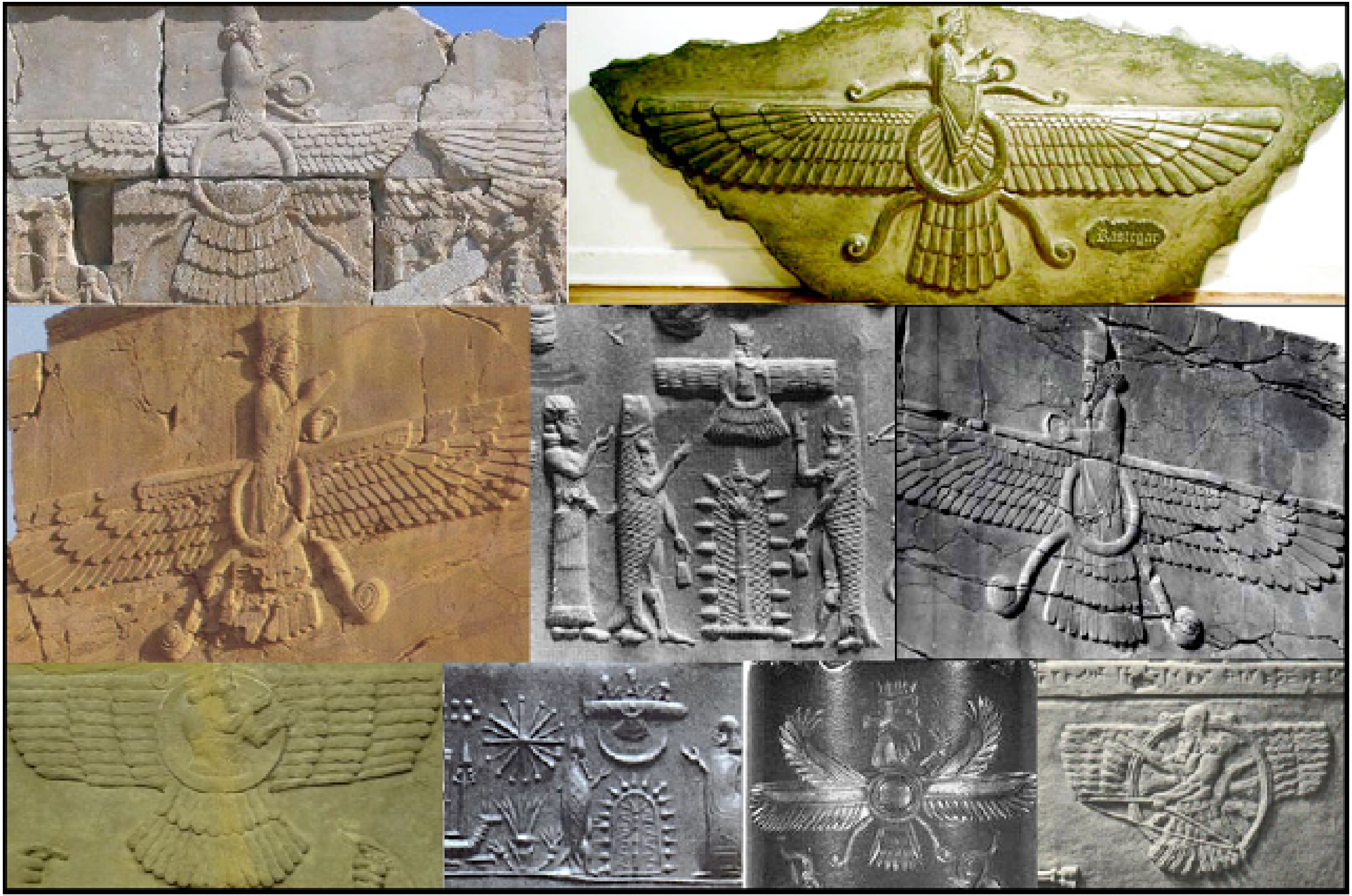

SUMERIAN DEPICTIONS OF ANUNNAKI GODS FLYING IN "WHIRLWINDS"

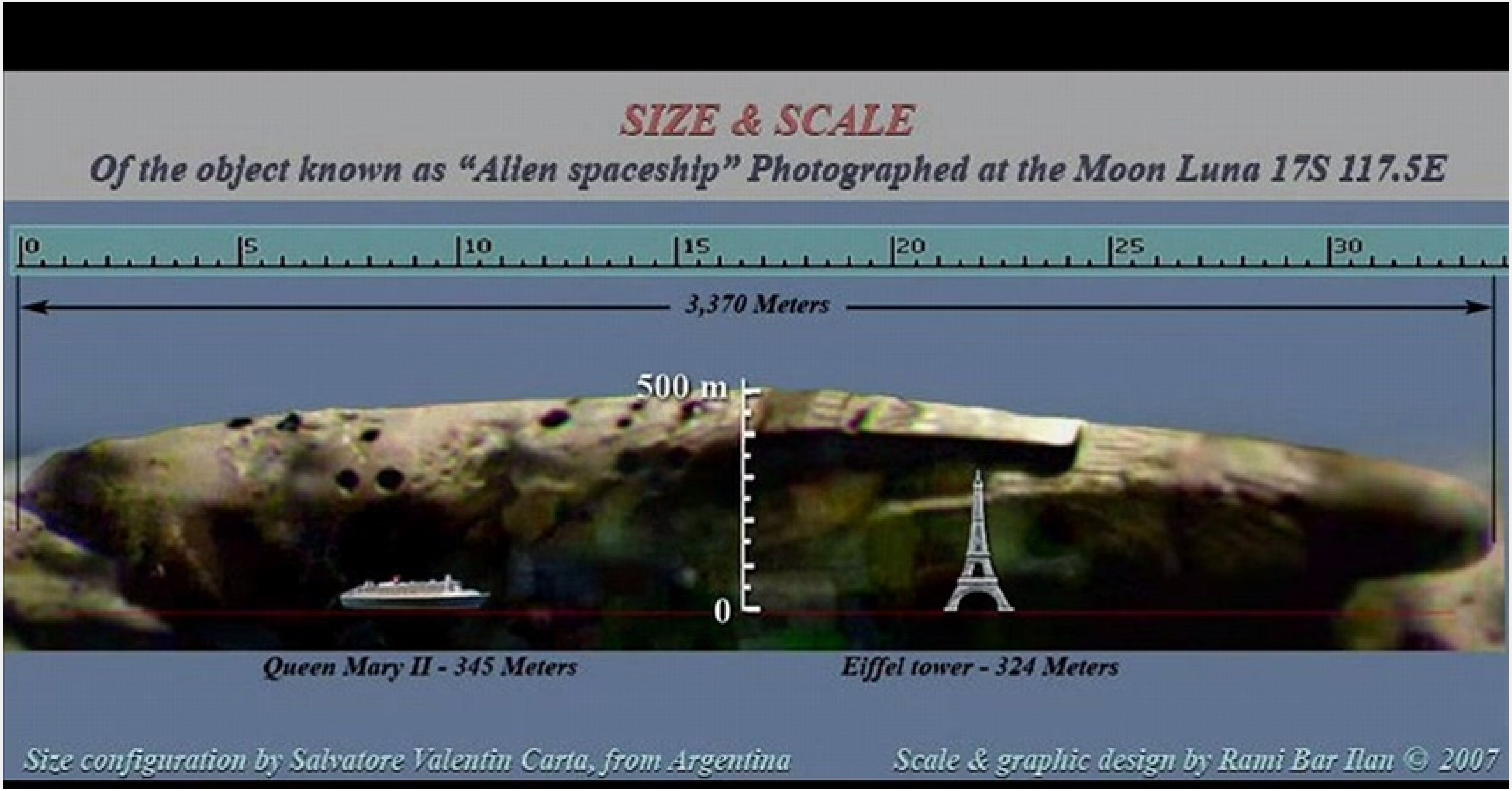
SIZE & SCALE
Of the object known as "Alien spaceship" Photographed at the Moon Luna 17S 117.5E
0
5
10
15
20
25
30
3,370 Meters
500 m
0
Queen Mary II - 345 Meters
Eiffel tower - 324 Meters
Size configuration by Salvatore Valentin Carta, from Argentina
Scale & graphic design by Rami Bar Ilan © 2007

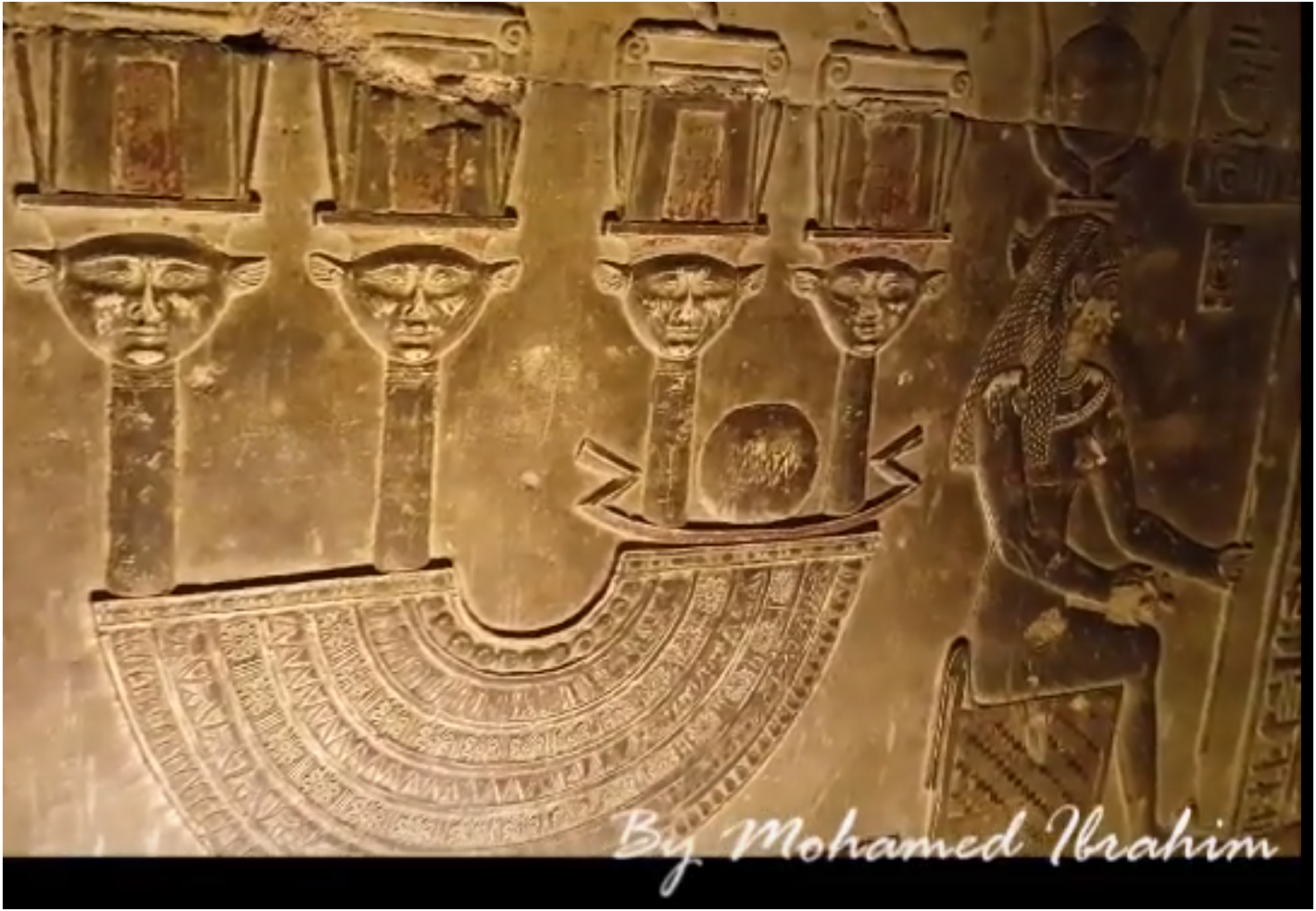

By Mohamed Ibrahim

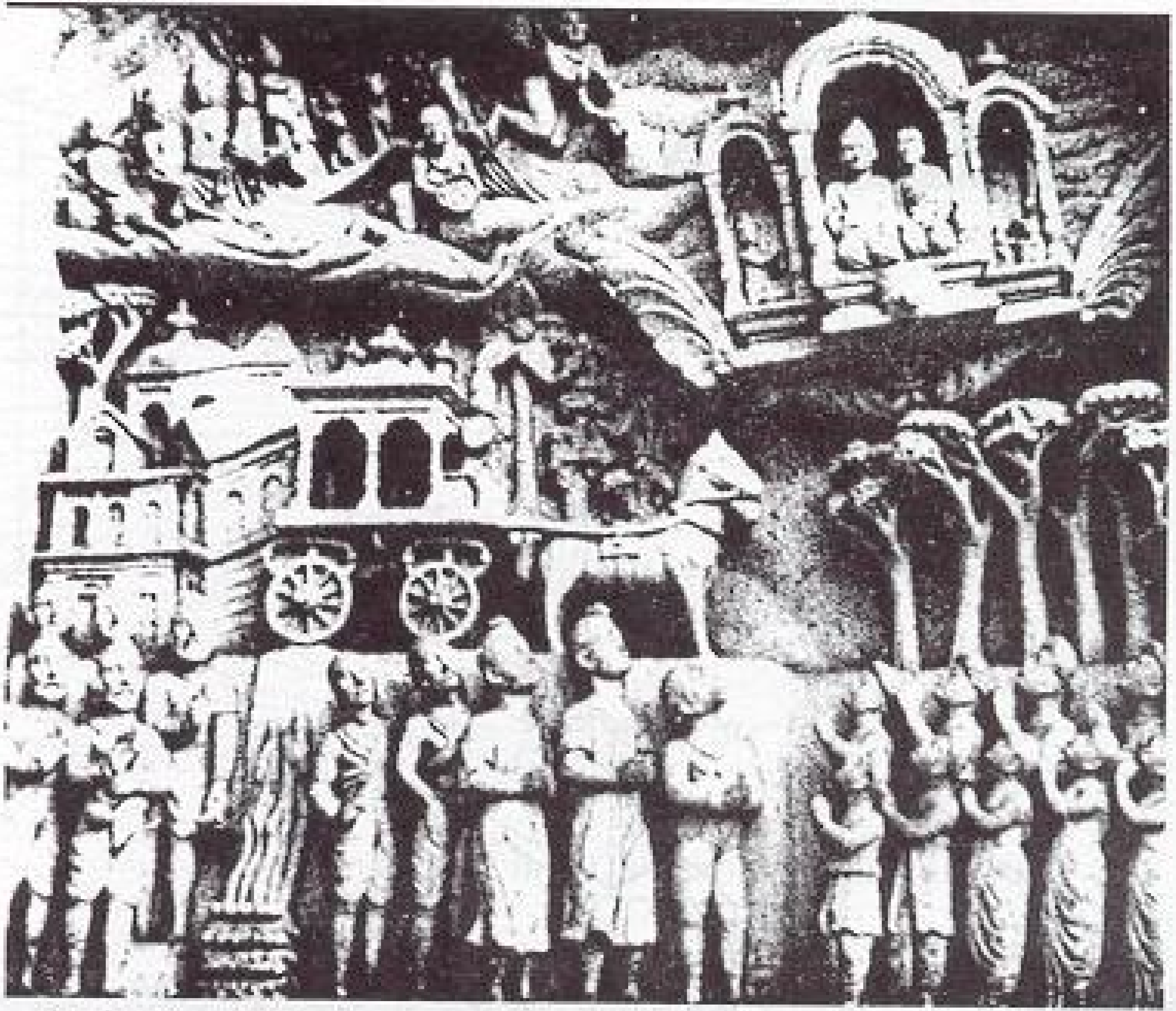

A vimana depicted in a temple relief at Ellora Caves, India.

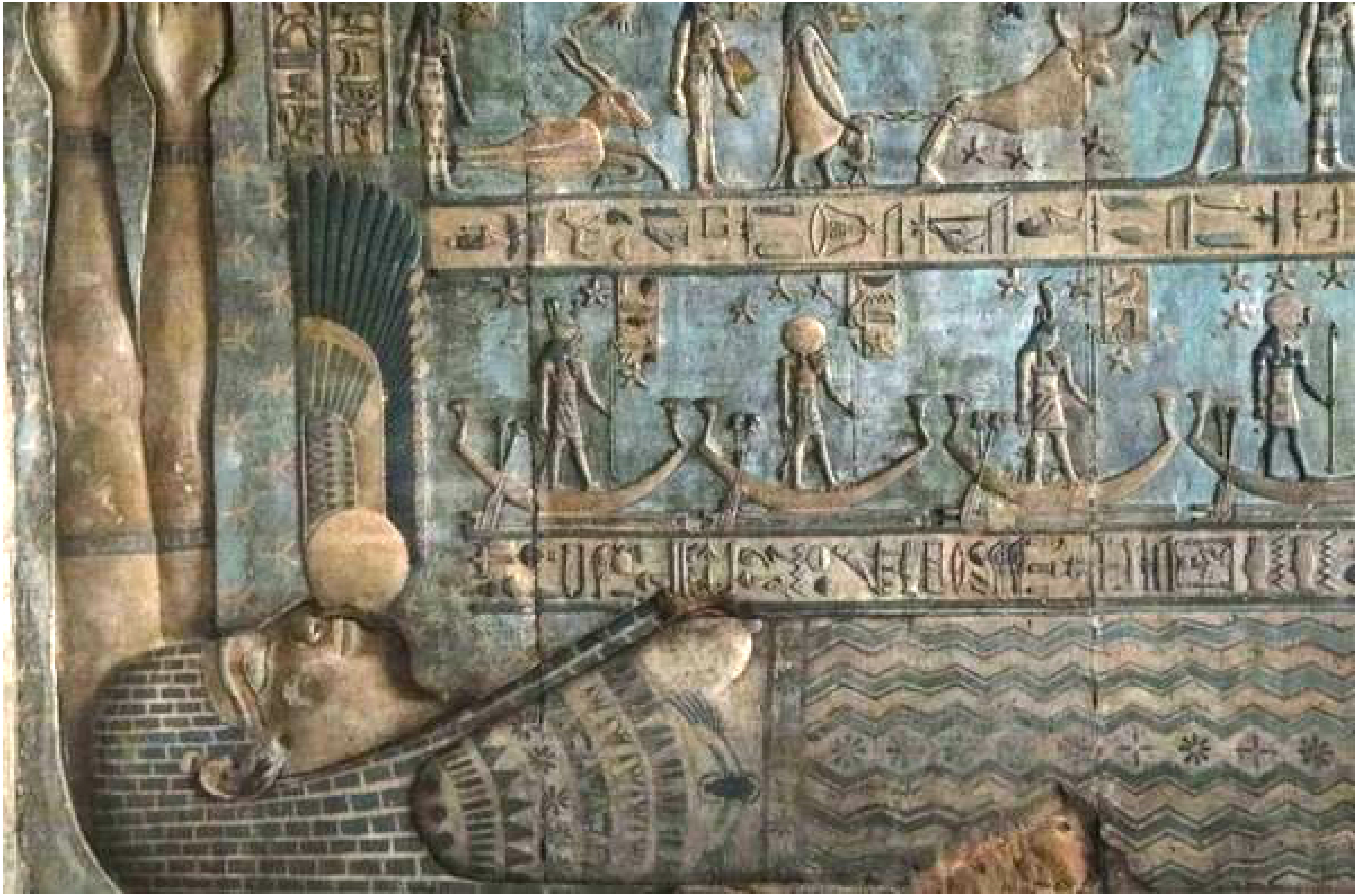

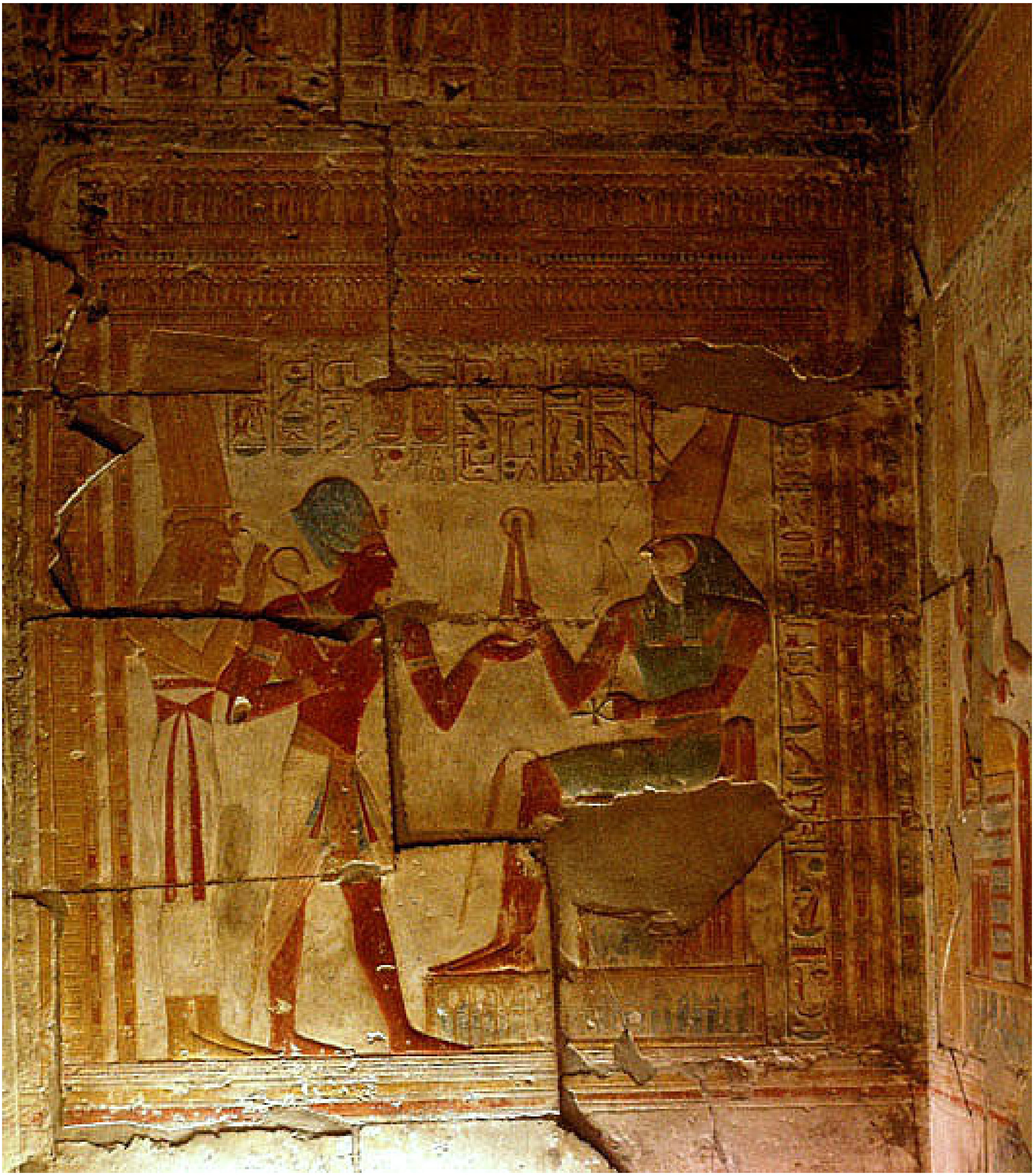

Travel to Eat
Kurt Buzard

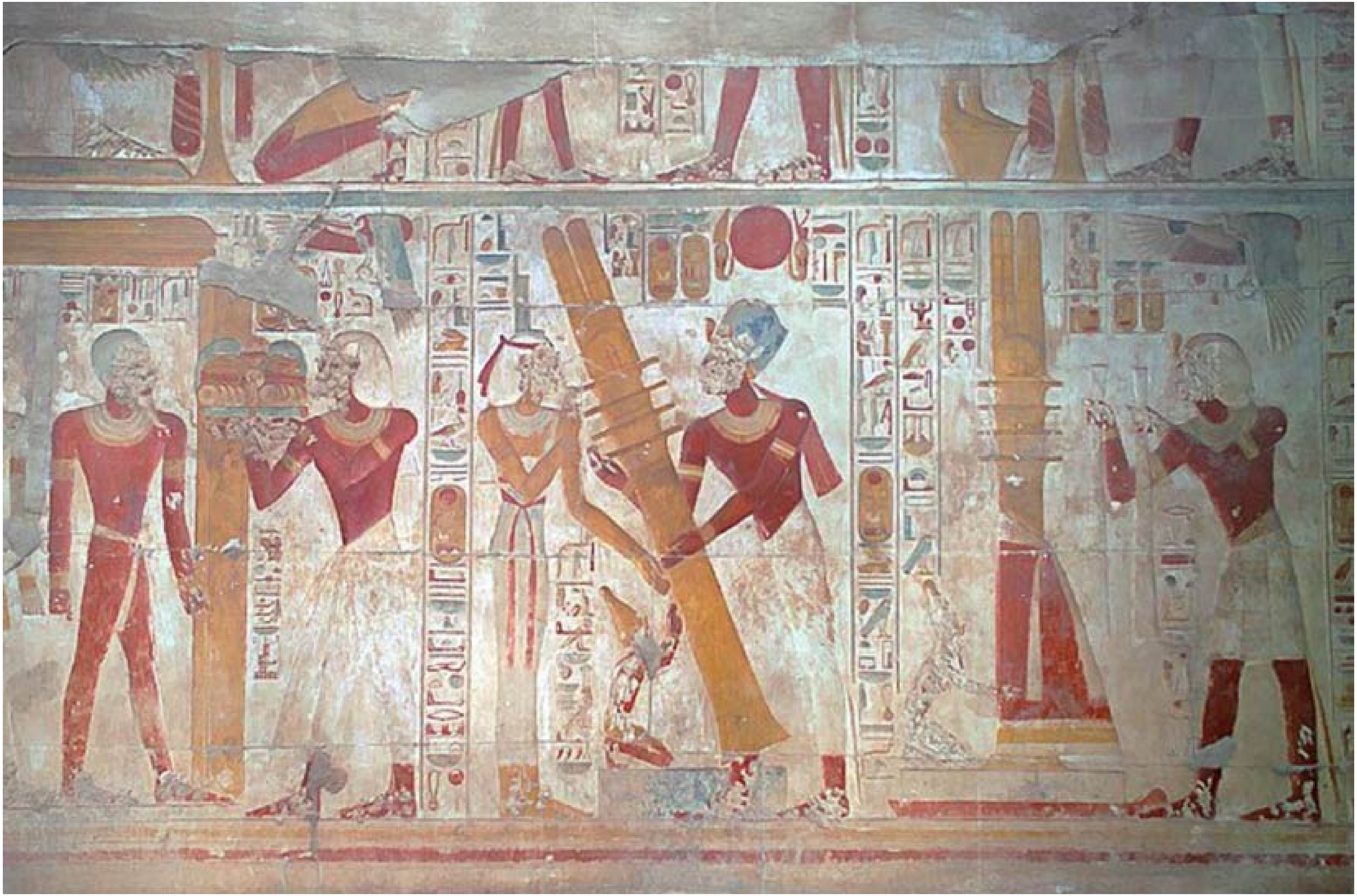

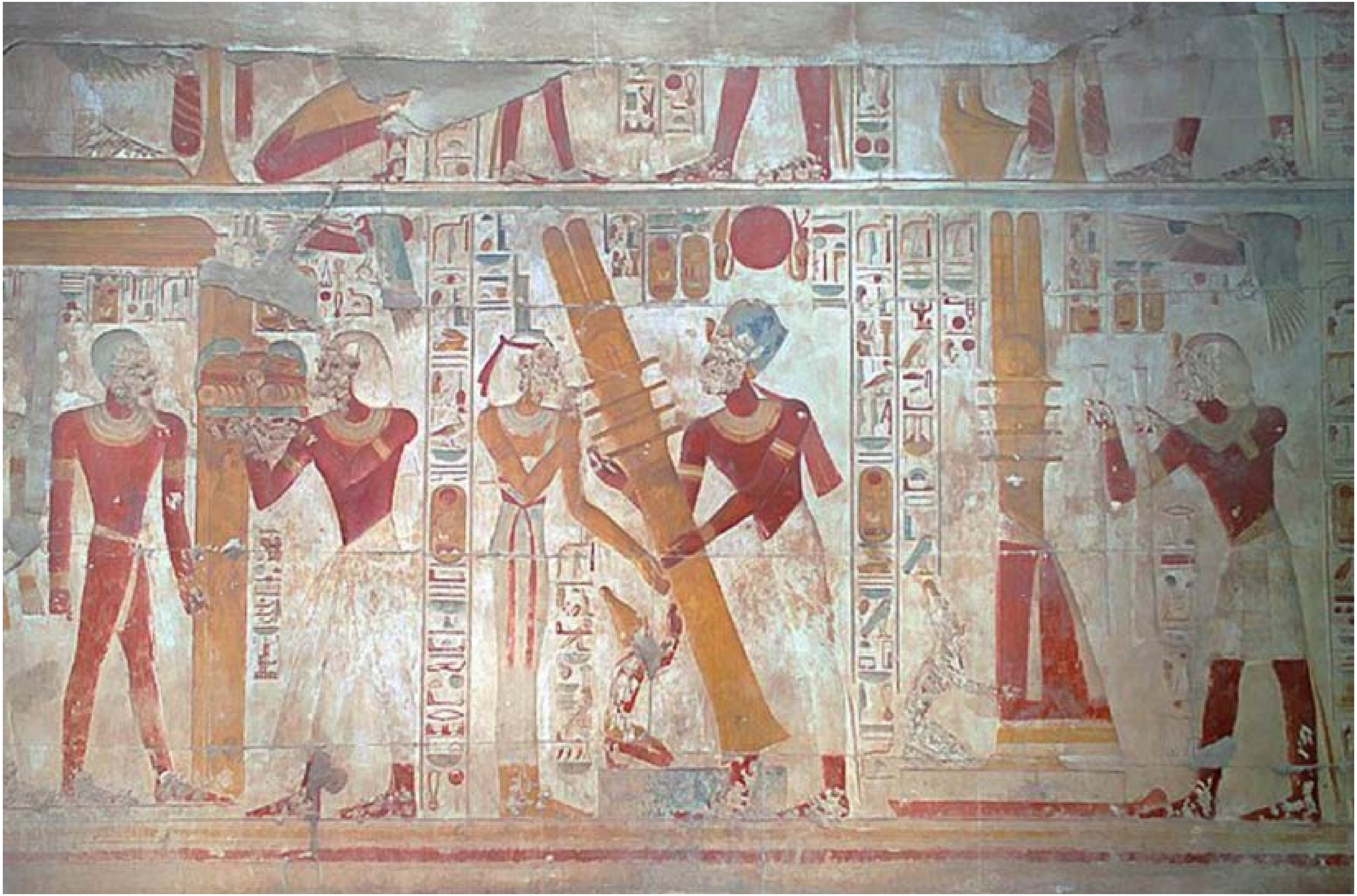

A possible physical artefact from our distant past shows a modern crop picture from 1991

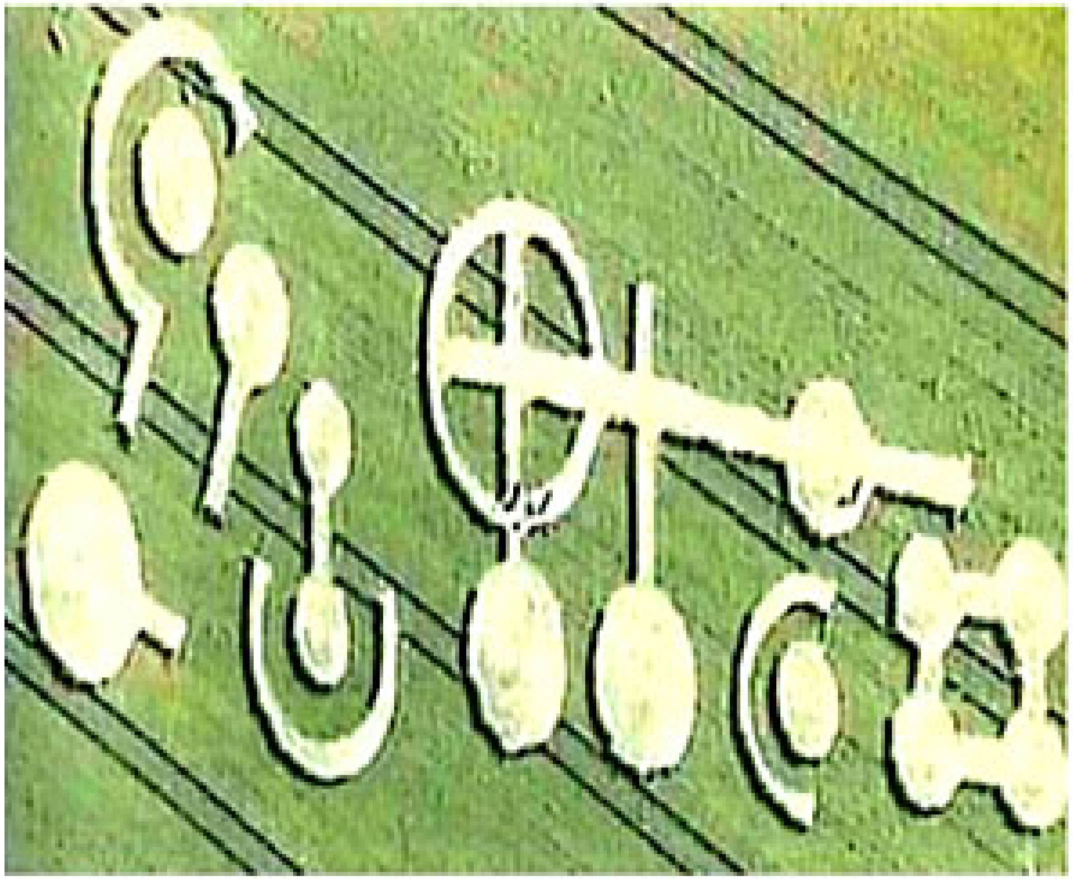

Grasdorf, Germany 1991

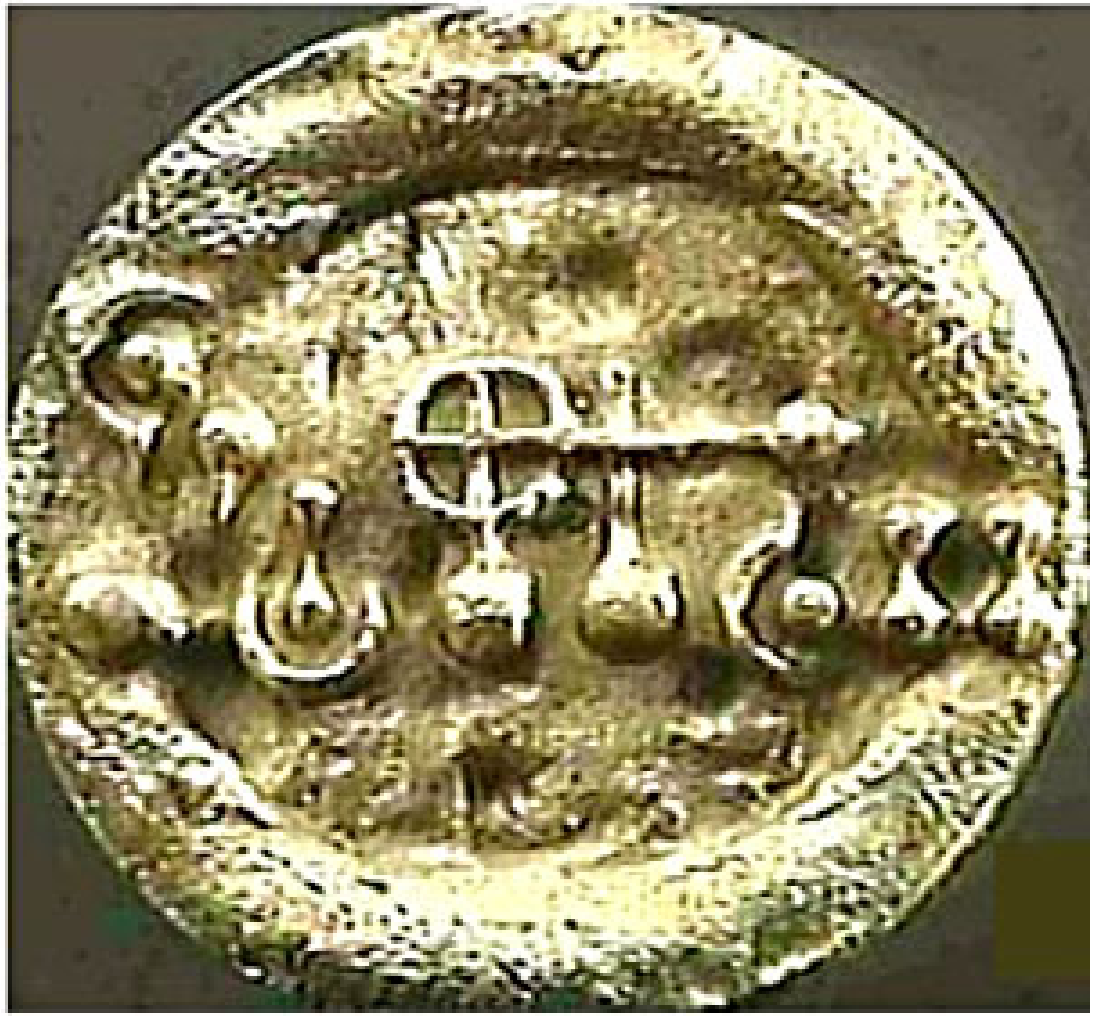

buried golden plate

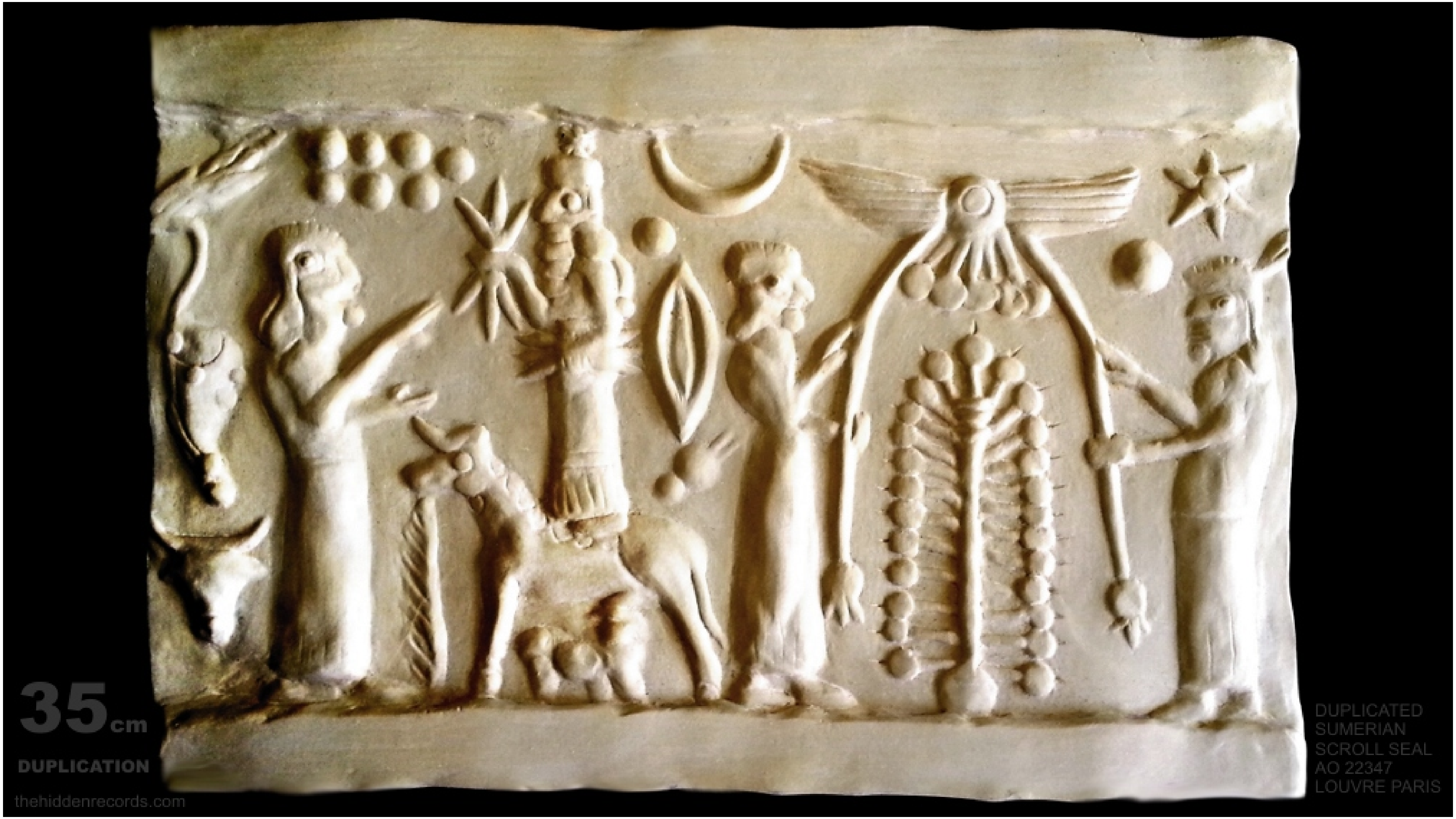

35cm
DUPLICATION

thehiddenrecords.com

DUPLICATED
SUMERIAN
SCROLL SEAL
AO 22347
LOUVRE PARIS

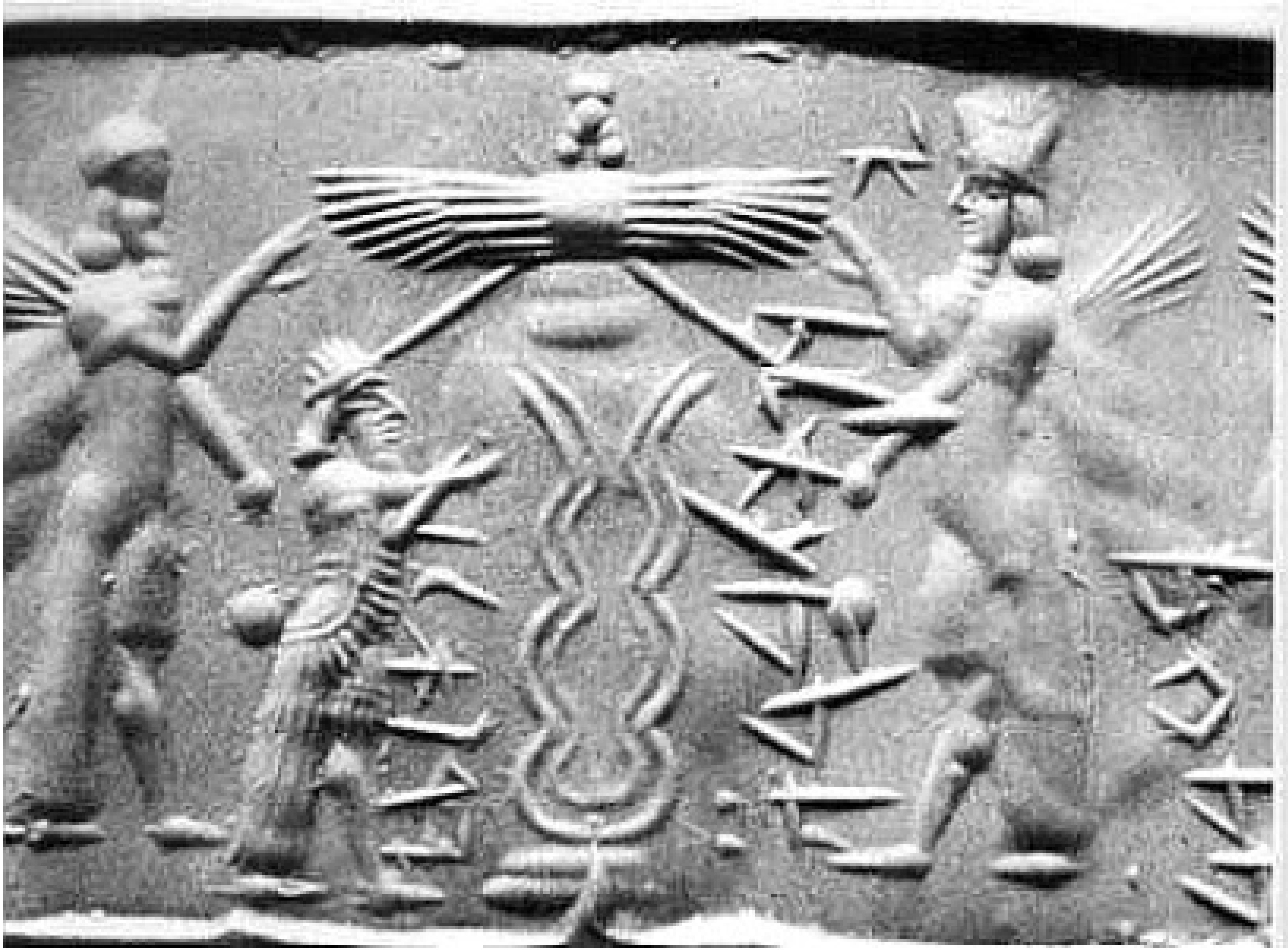

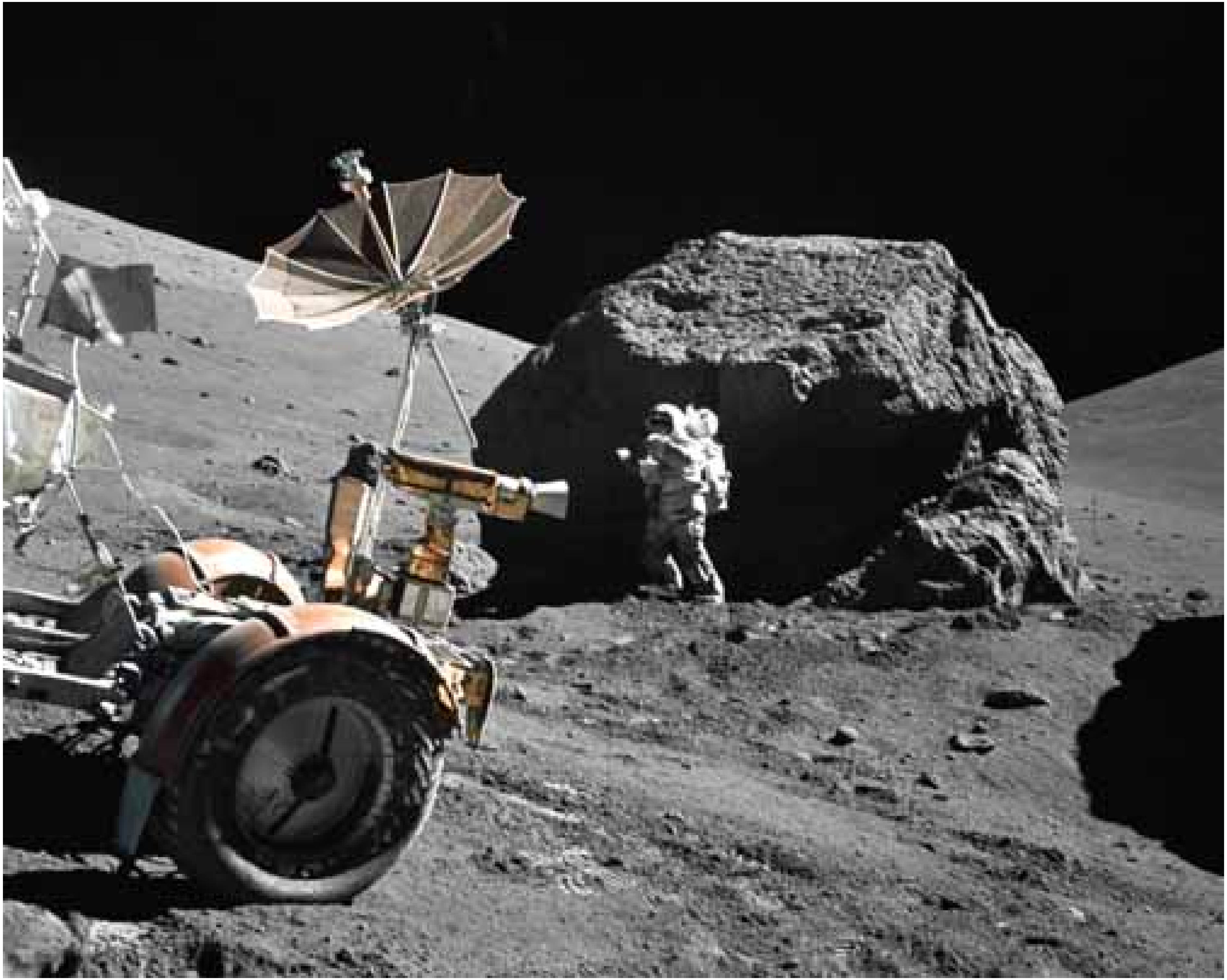

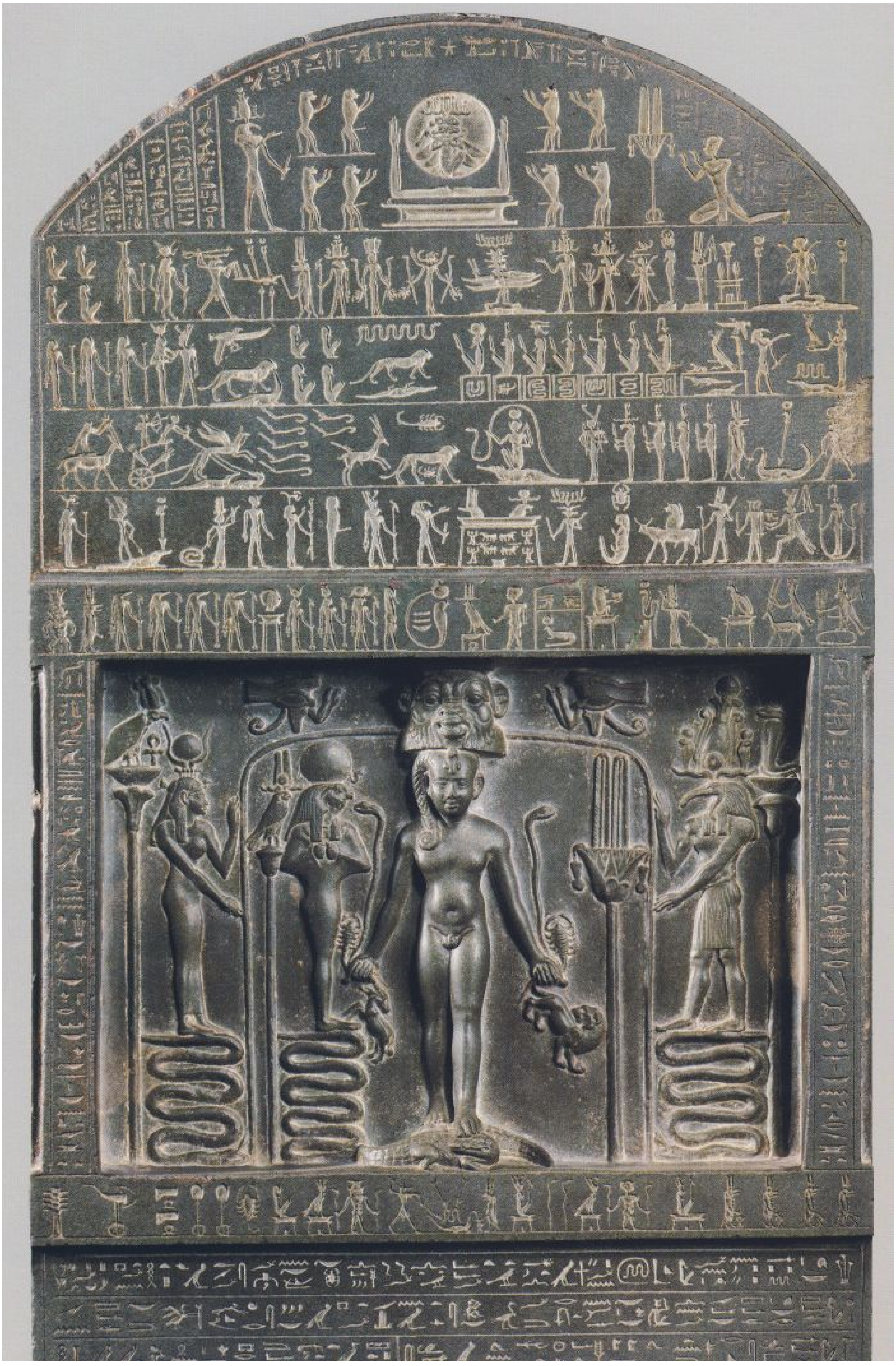

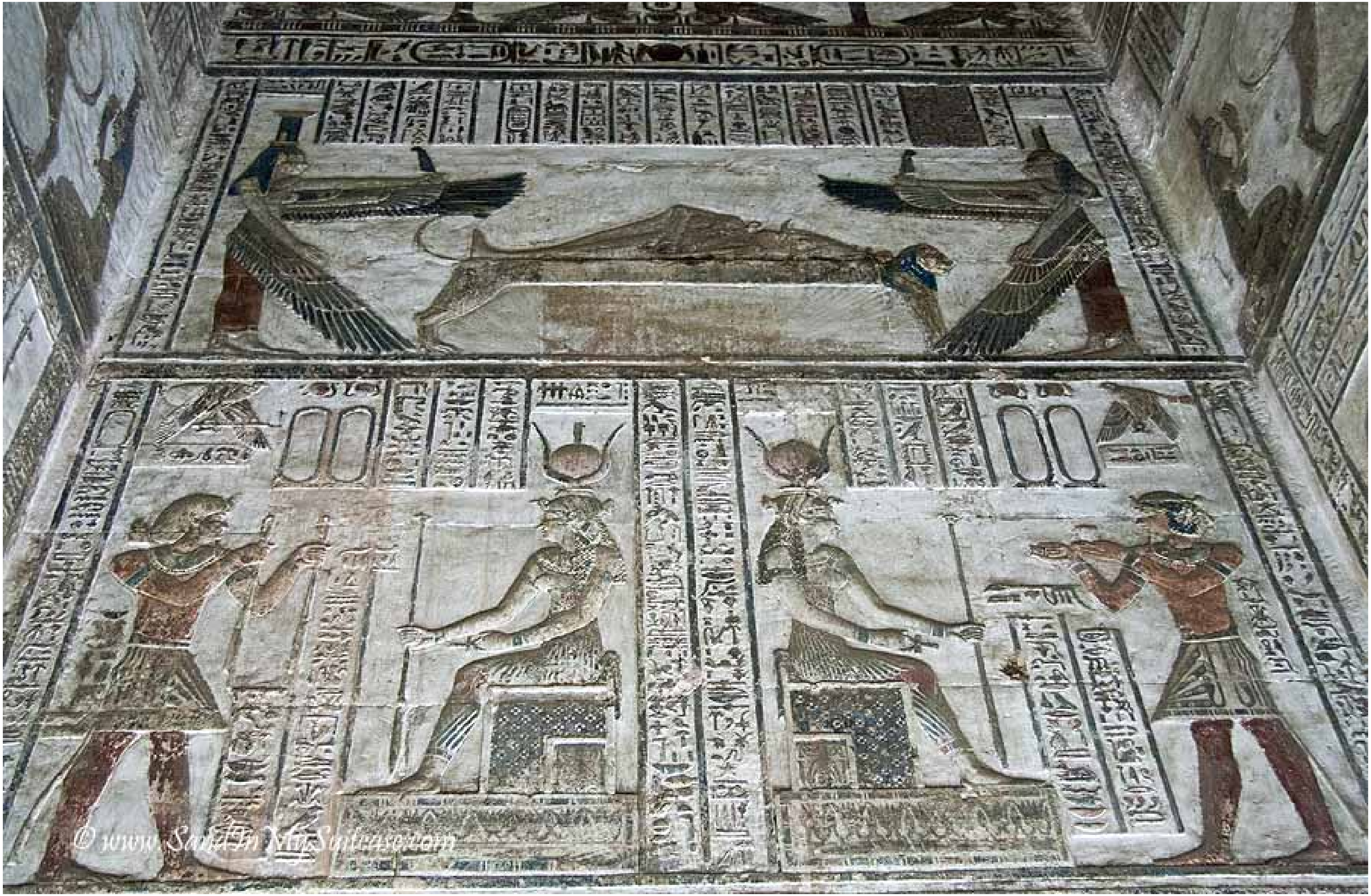

© www.SandInMySuitcase.com

© mich

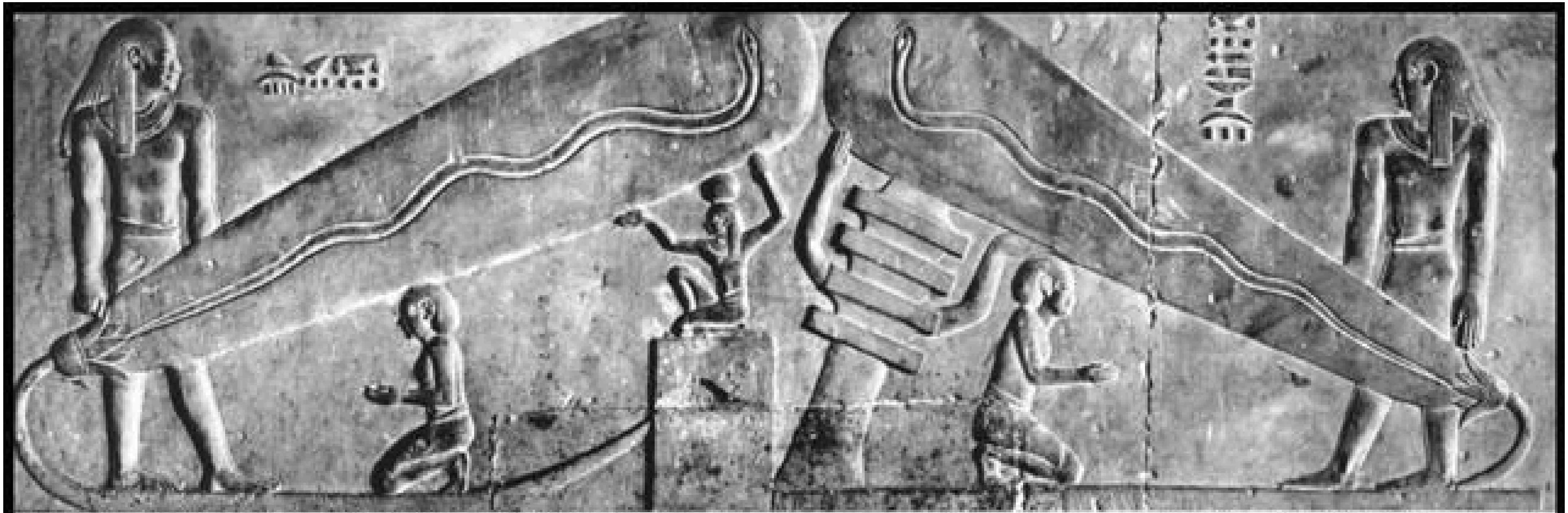

Battery-powered searchlights and filament lamps in the crypts of the Temple of Hathor (Isis) at Denderah

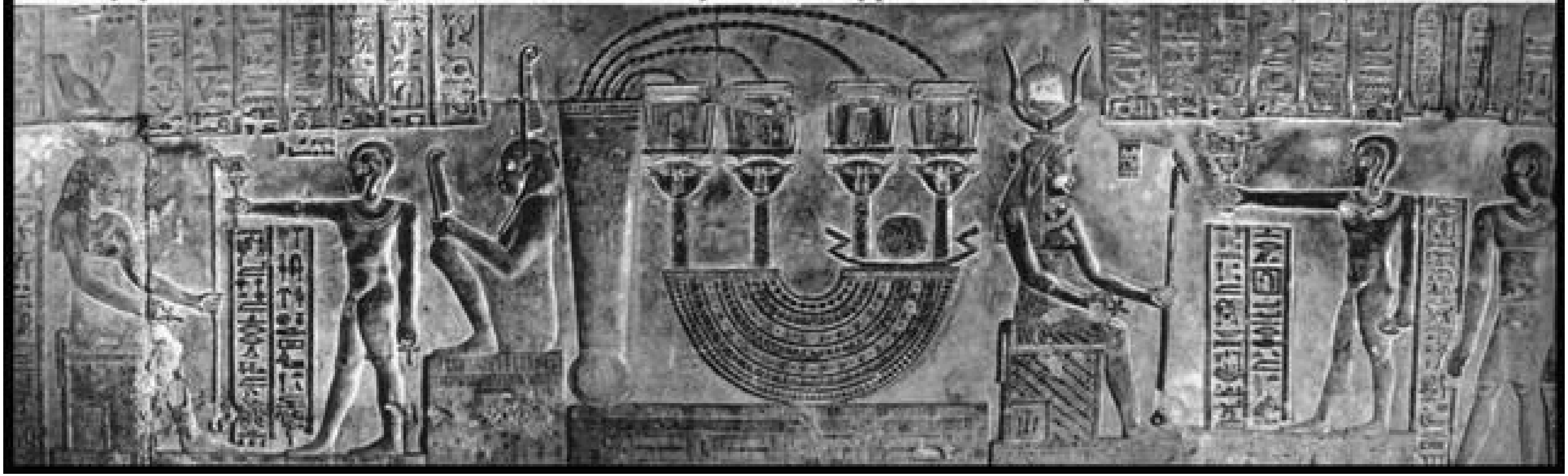

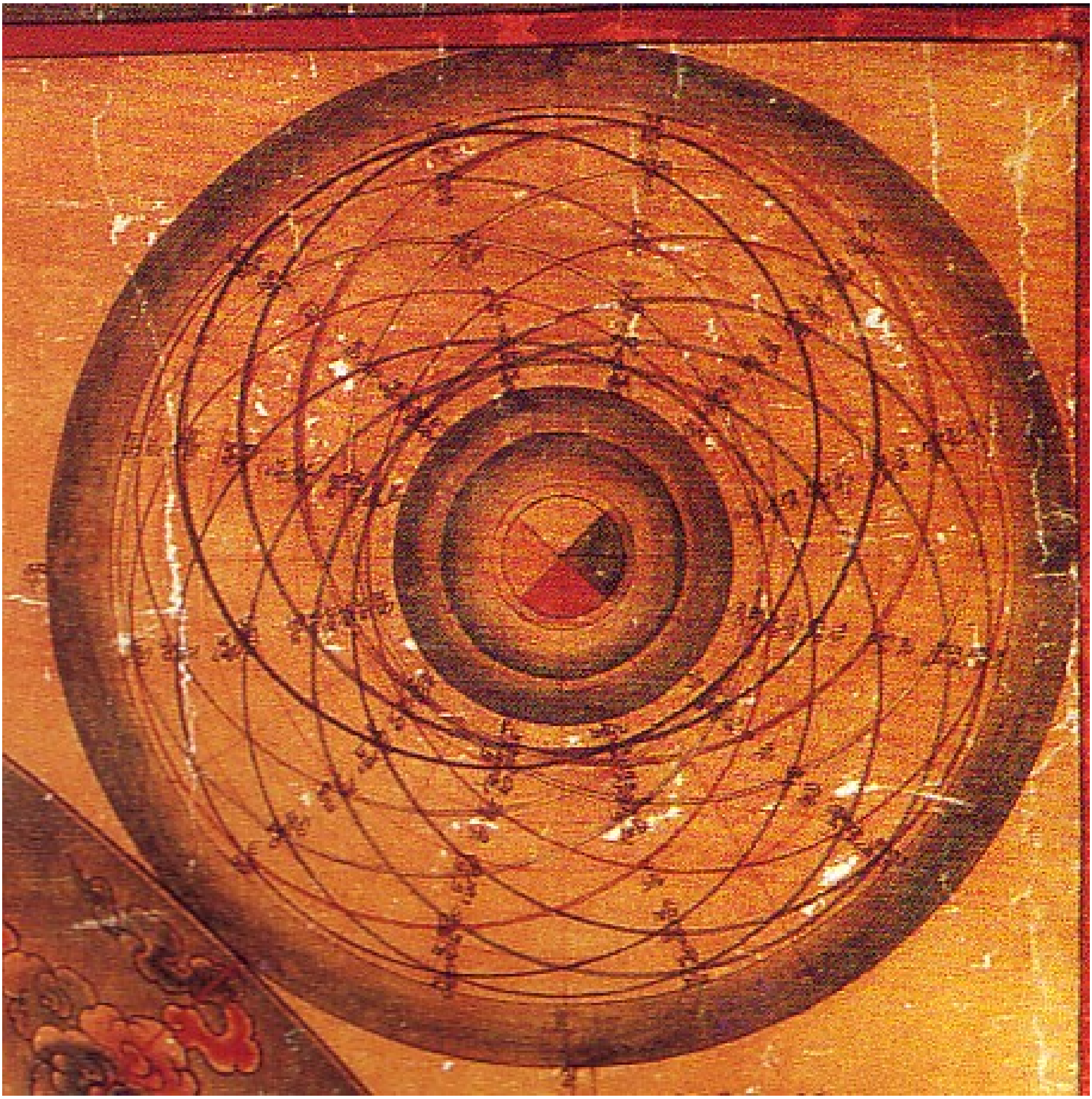

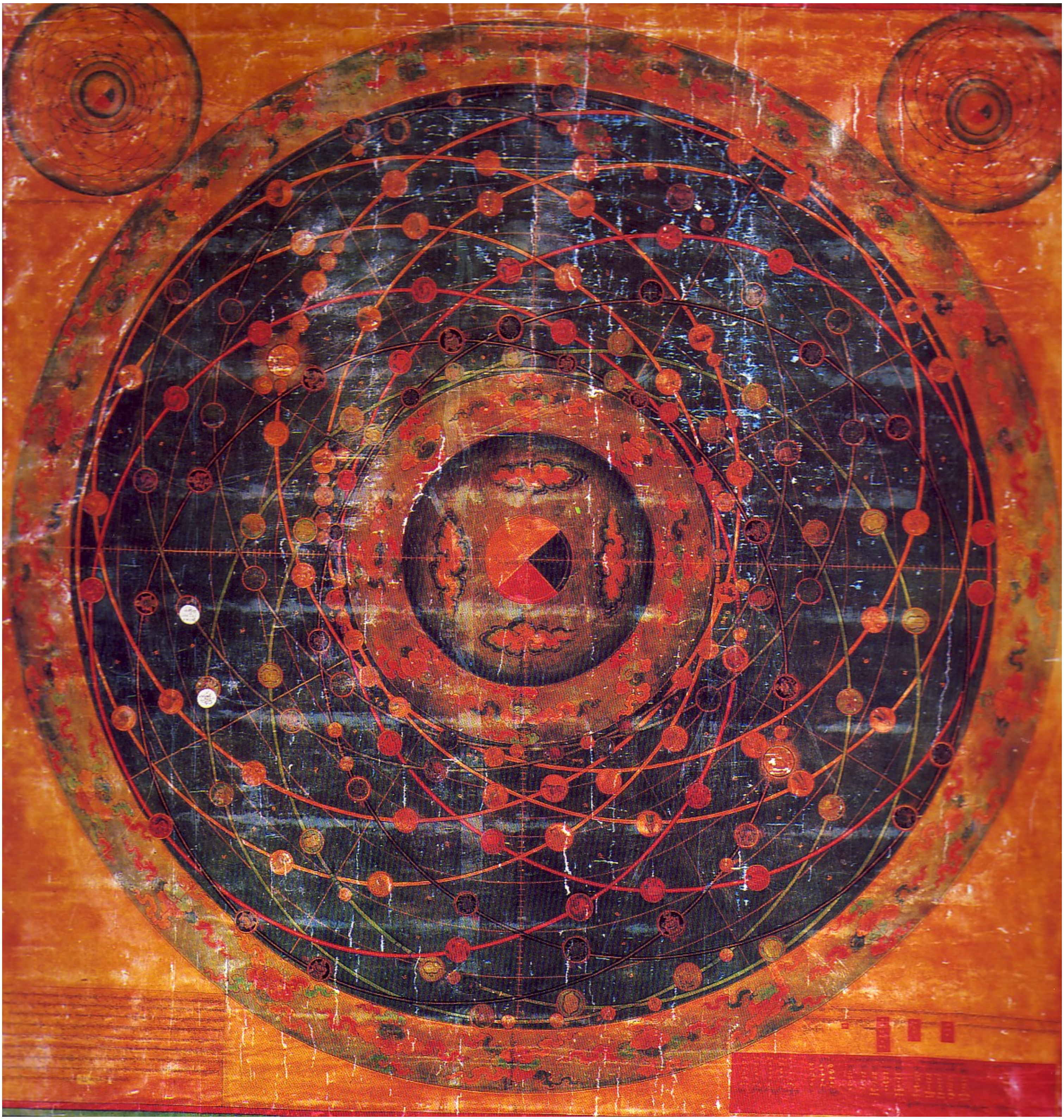

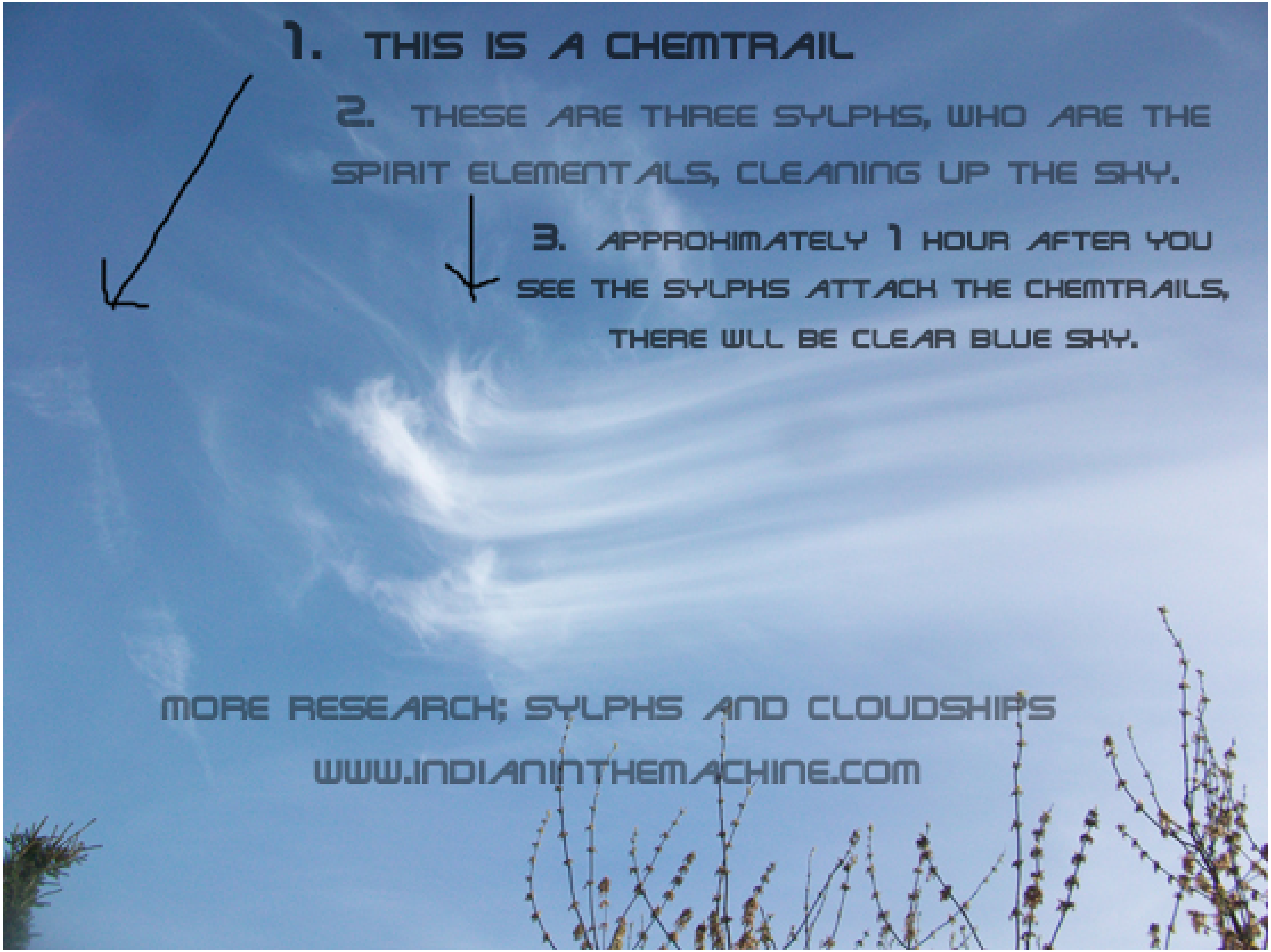

1. THIS IS A CHEMTRAIL

2. THESE ARE THREE SYLPHS, WHO ARE THE SPIRIT ELEMENTALS, CLEANING UP THE SKY.

3. APPROXIMATELY 1 HOUR AFTER YOU SEE THE SYLPHS ATTACH THE CHEMTRAILS, THERE WLL BE CLEAR BLUE SKY.

MORE RESEARCH; SYLPHS AND CLOUDSHIPS

WWW.INDIANINTHEMACHINE.COM

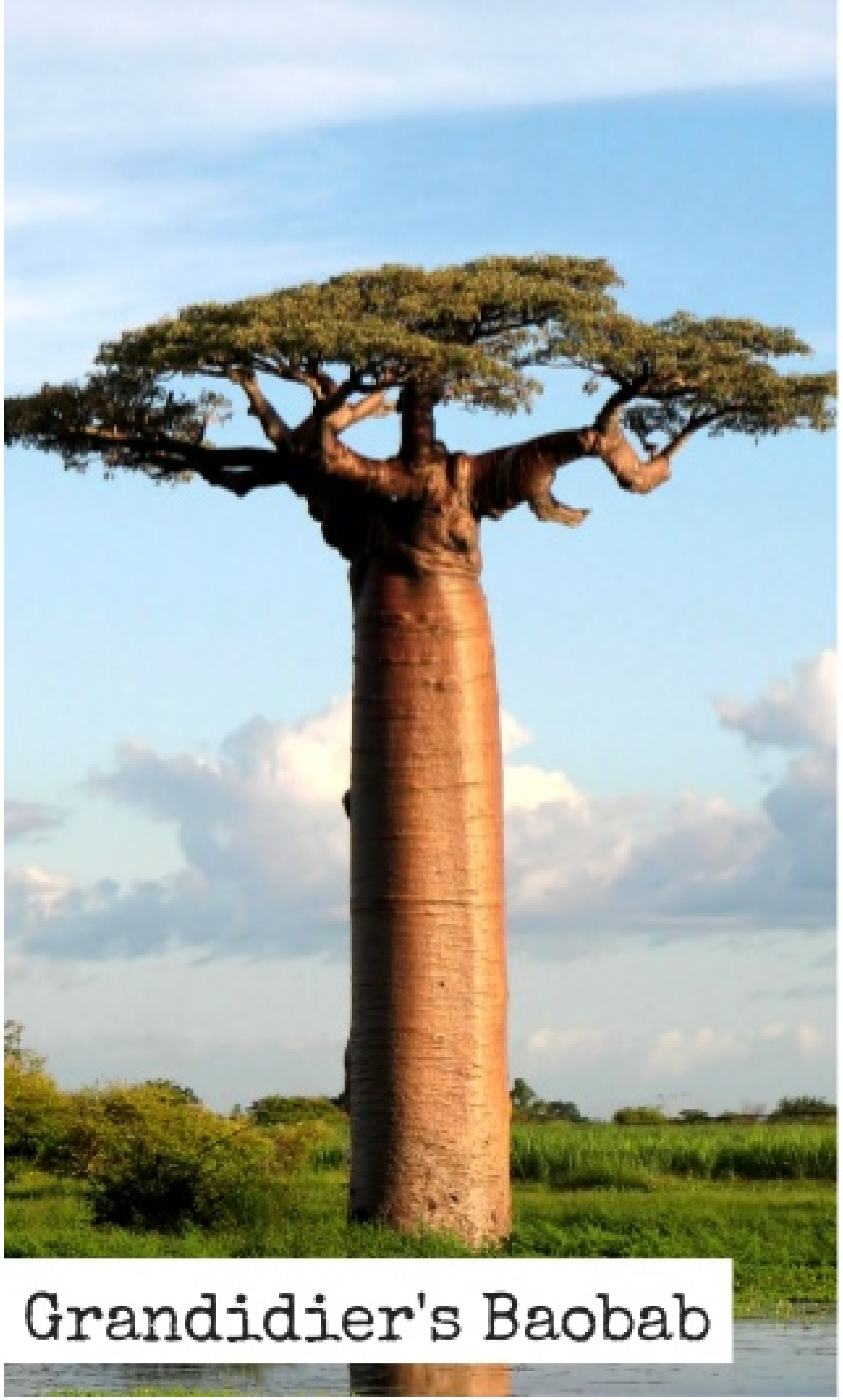

Grandidier's Baobab

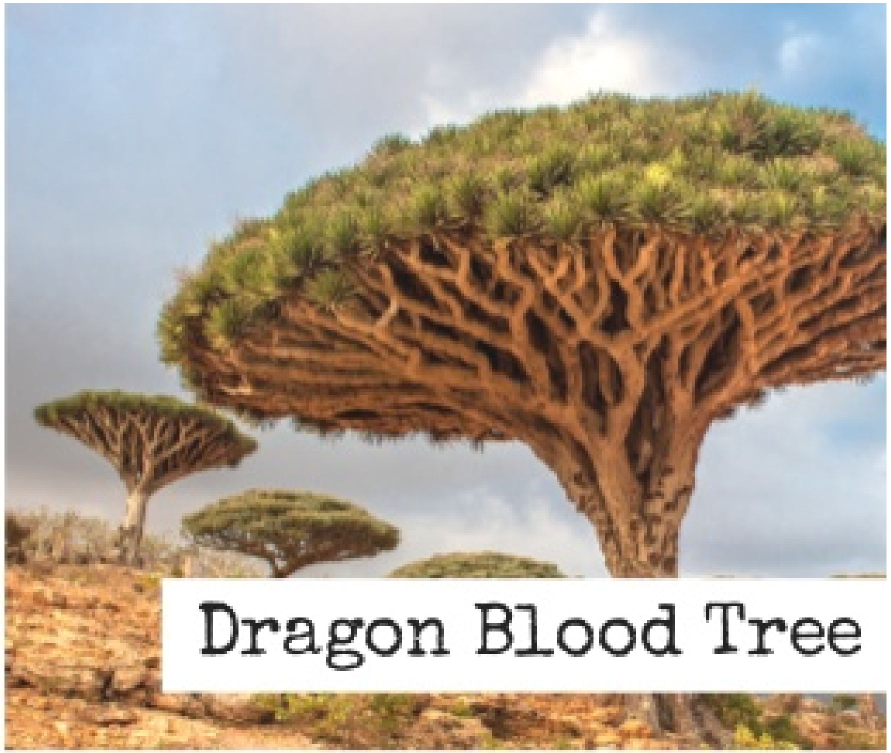

Dragon Blood Tree

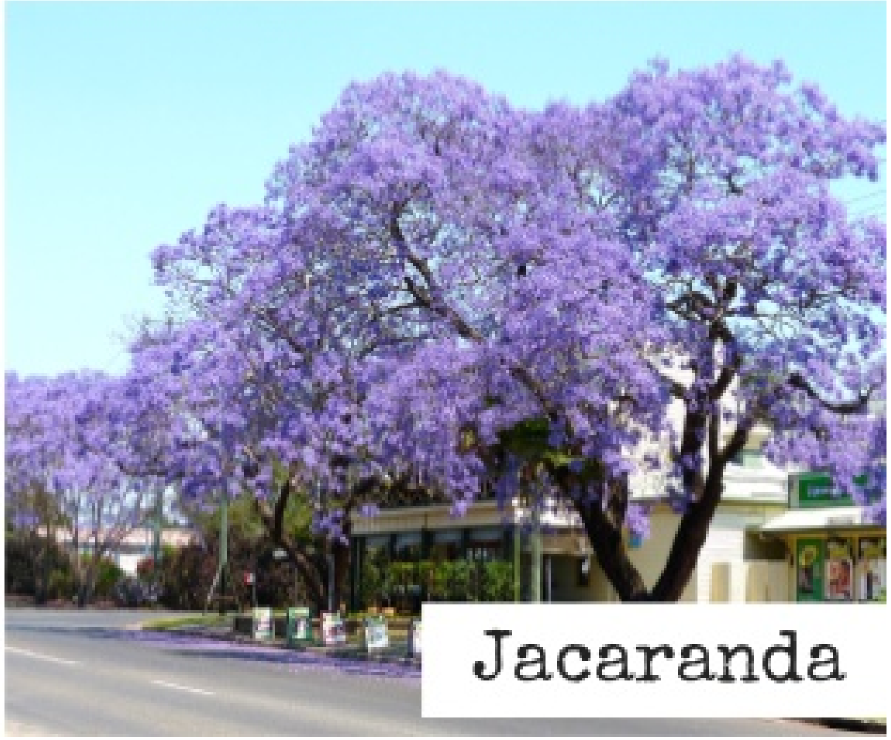

Jacaranda

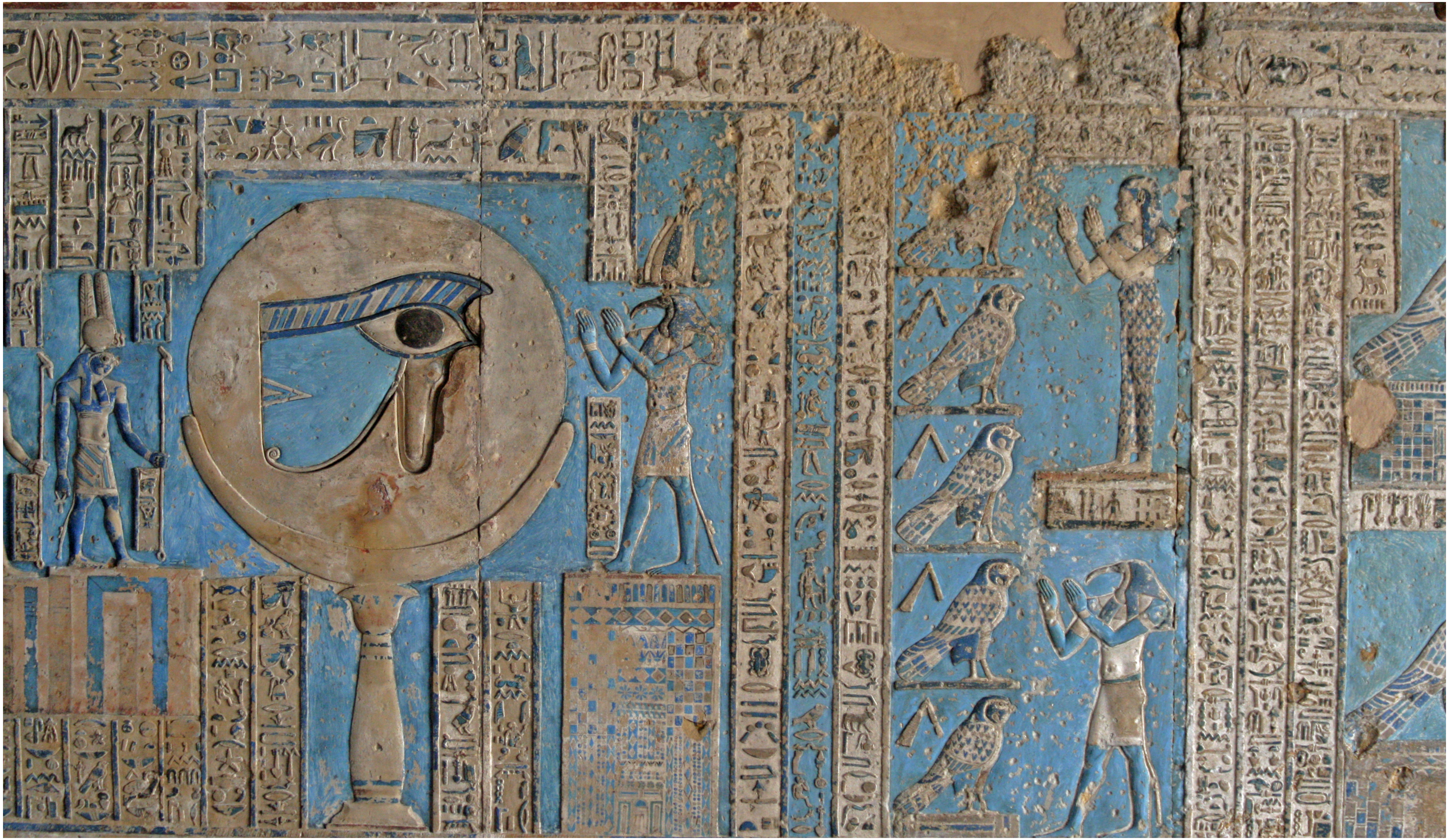

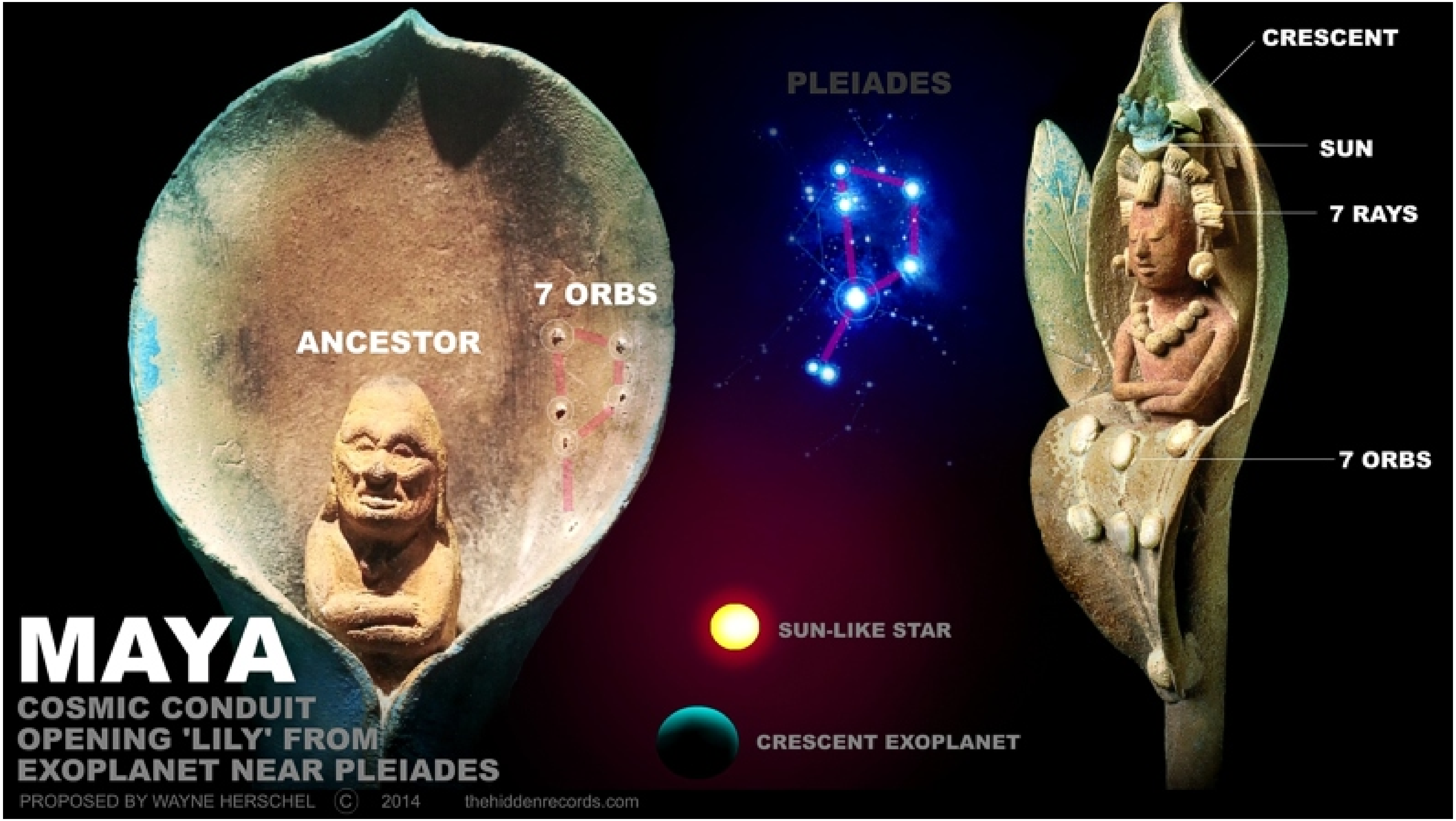
PLEIADES
7 ORBS
ANCESTOR
SUN-LIKE STAR
CRESCENT EXOPLANET
CRESCENT
SUN
7 RAYS
7 ORBS

MAYA

COSMIC CONDUIT OPENING 'LILY' FROM EXOPLANET NEAR PLEIADES

PROPOSED BY WAYNE HERSCHEL © 2014 thehiddenrecords.com

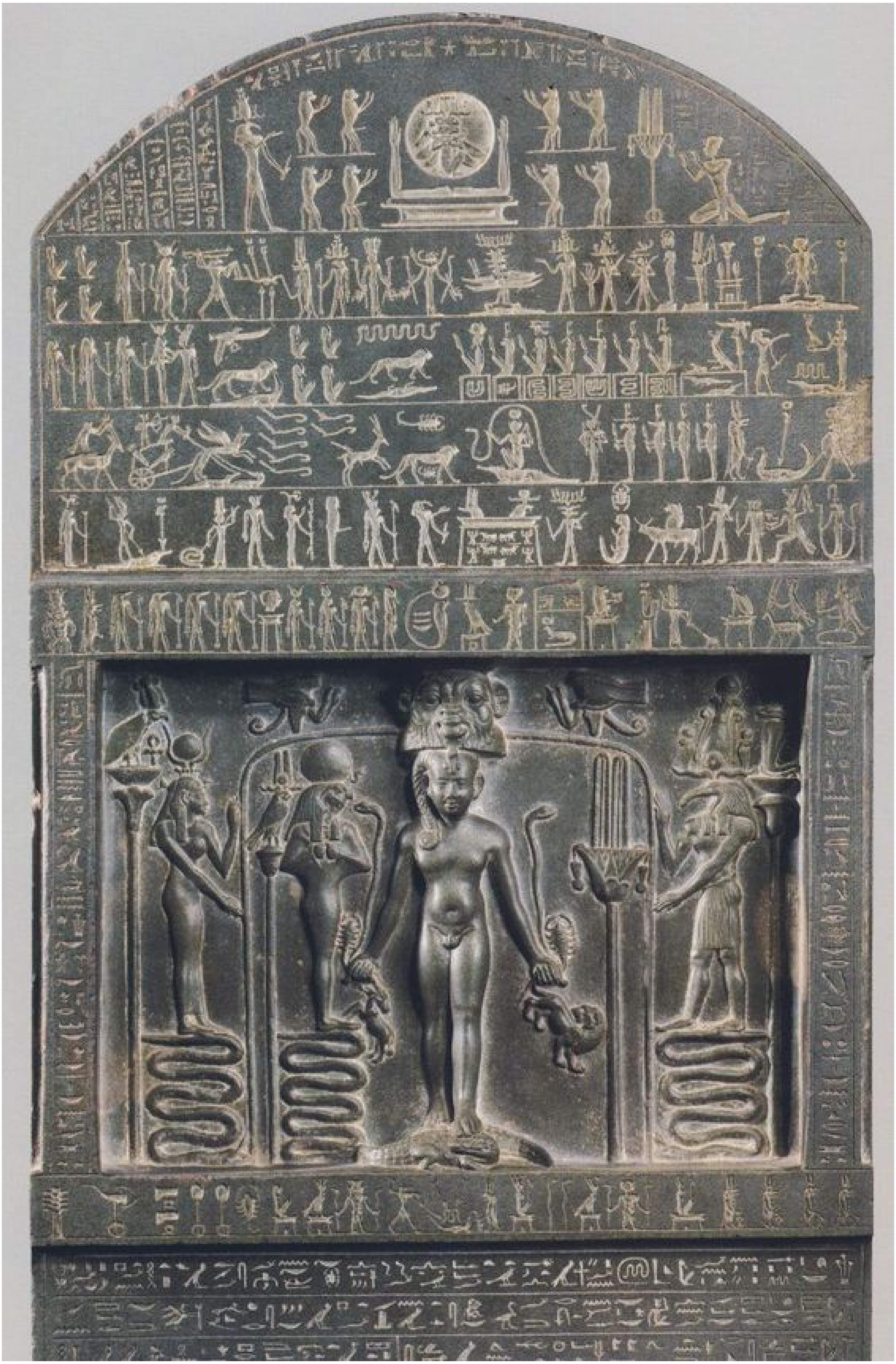

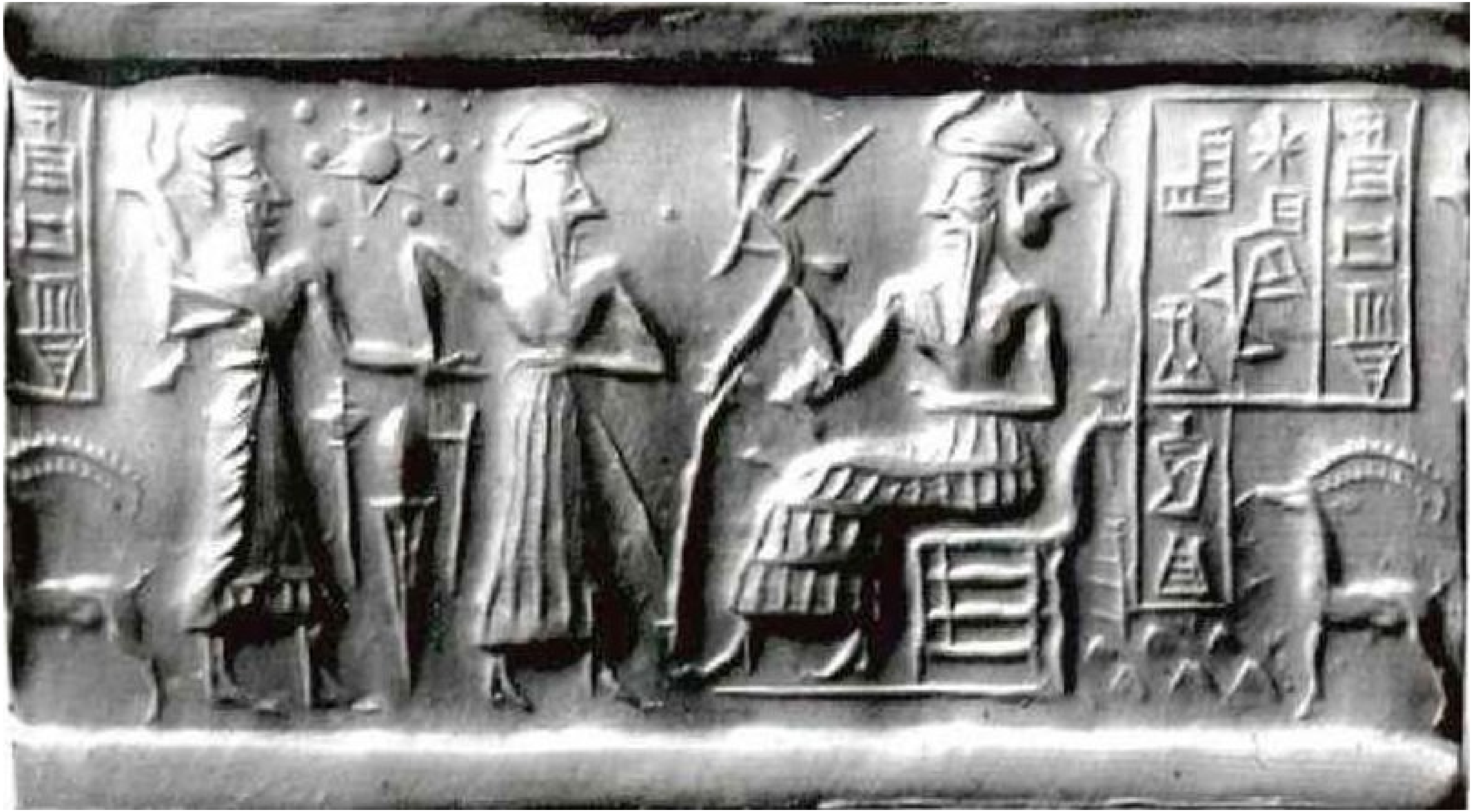

HAARP Ionospheric Research Facility

(8 Miles North of Gakona, Alaska)

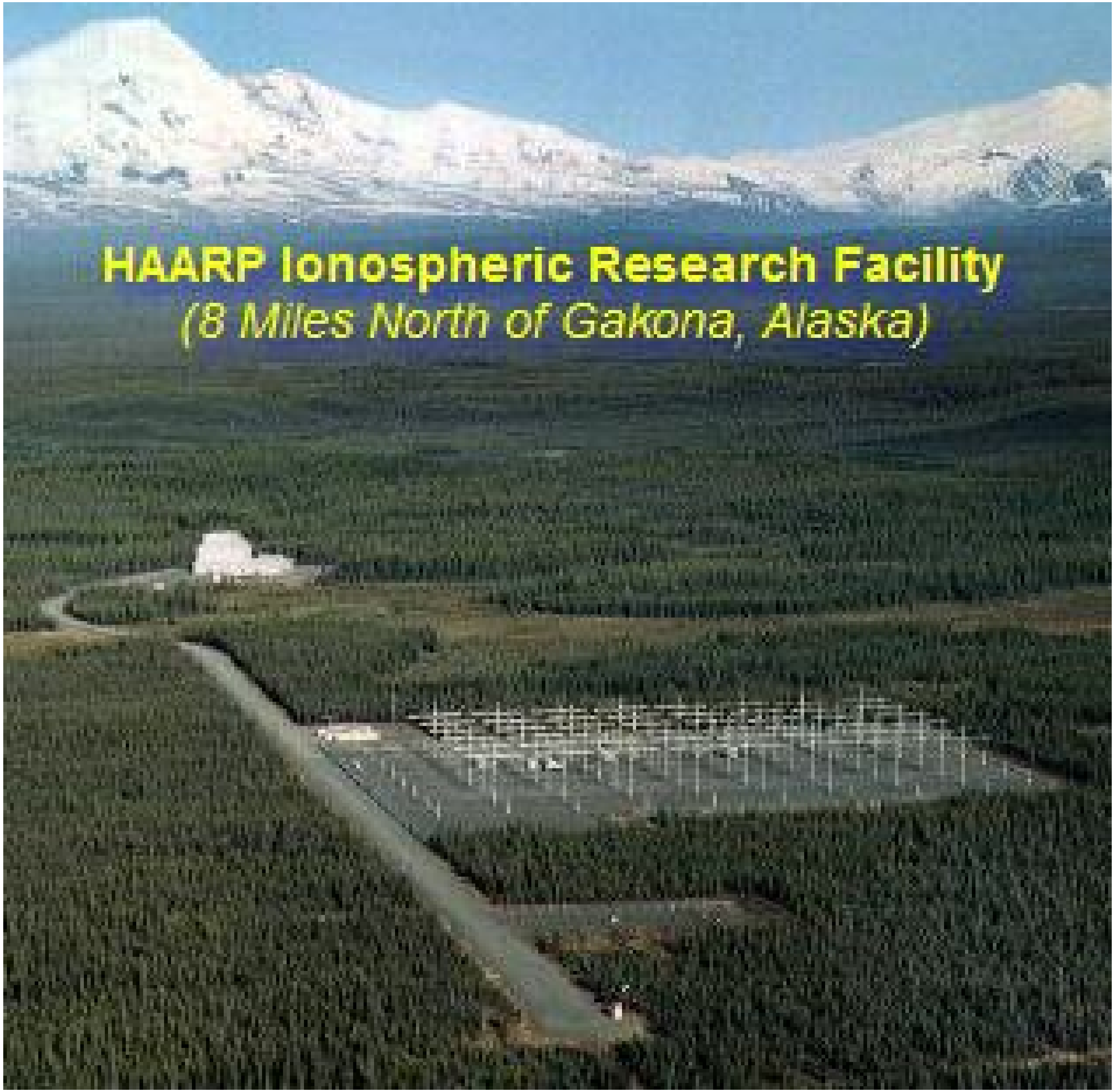

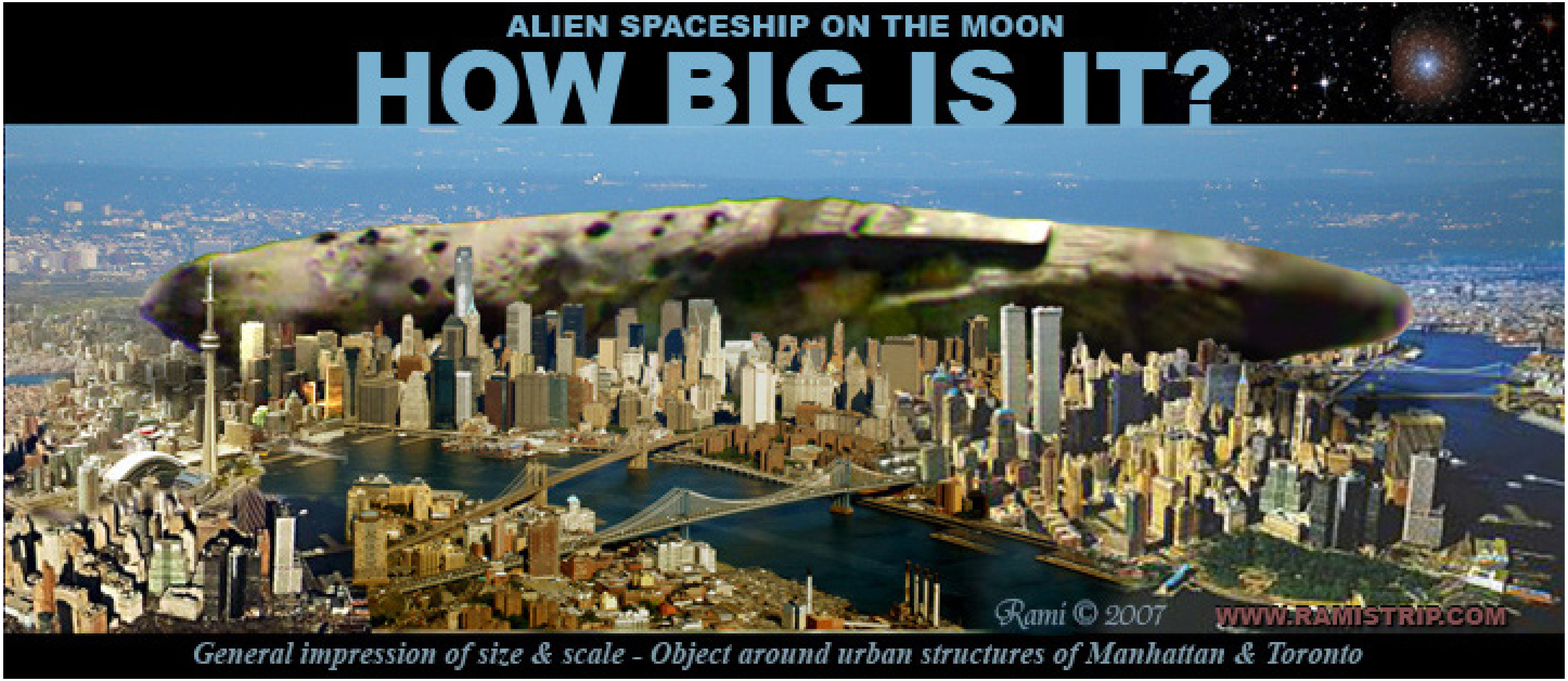
ALIEN SPACESHIP ON THE MOON
HOW BIG IS IT?
Rami © 2007
WWW.RAMISTRIP.COM

General impression of size & scale - Object around urban structures of Manhattan & Toronto

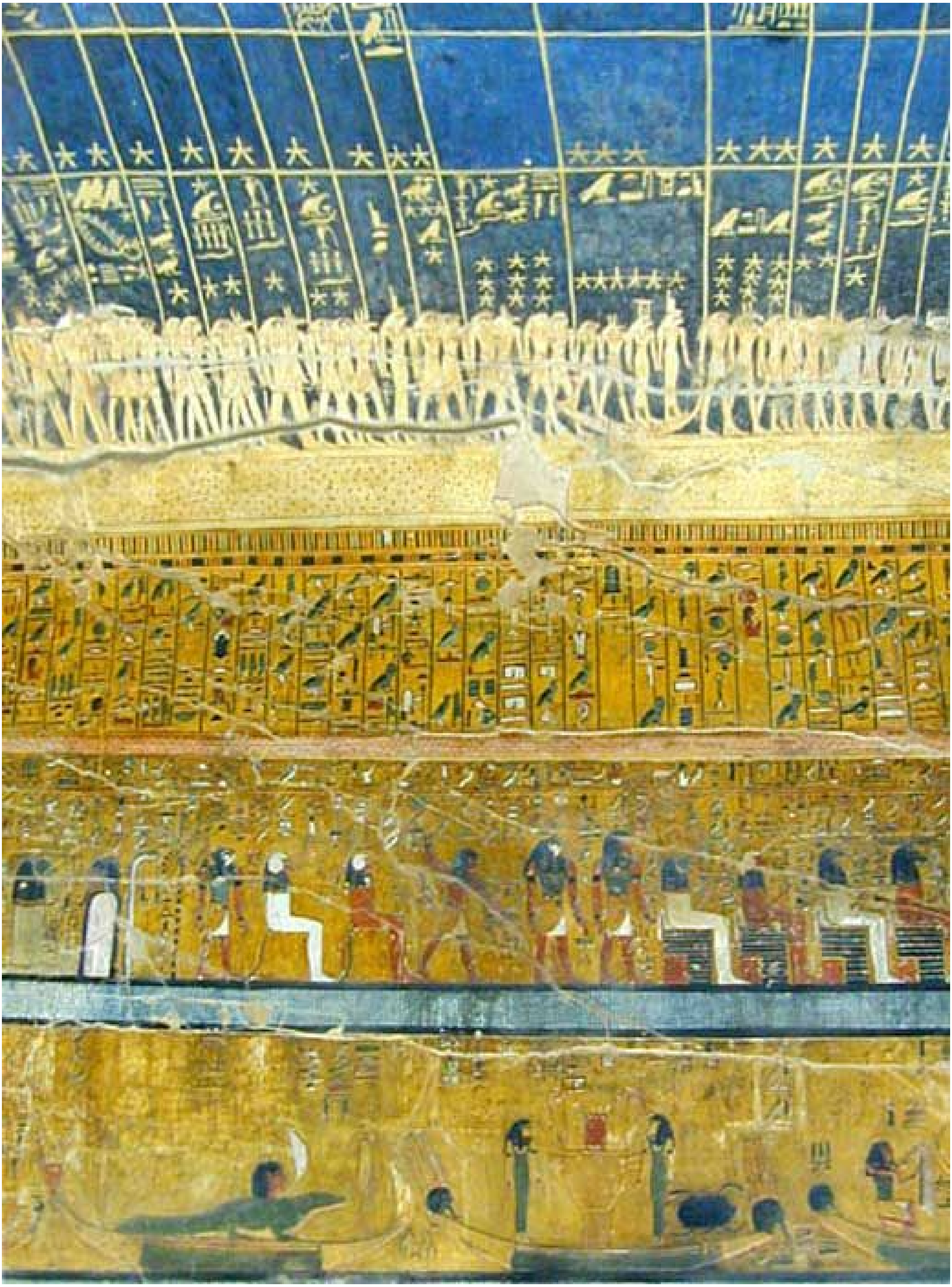

© 2013 boyd currey / bokeh.net.au

gettyimages
Kenneth Garrett
174405620

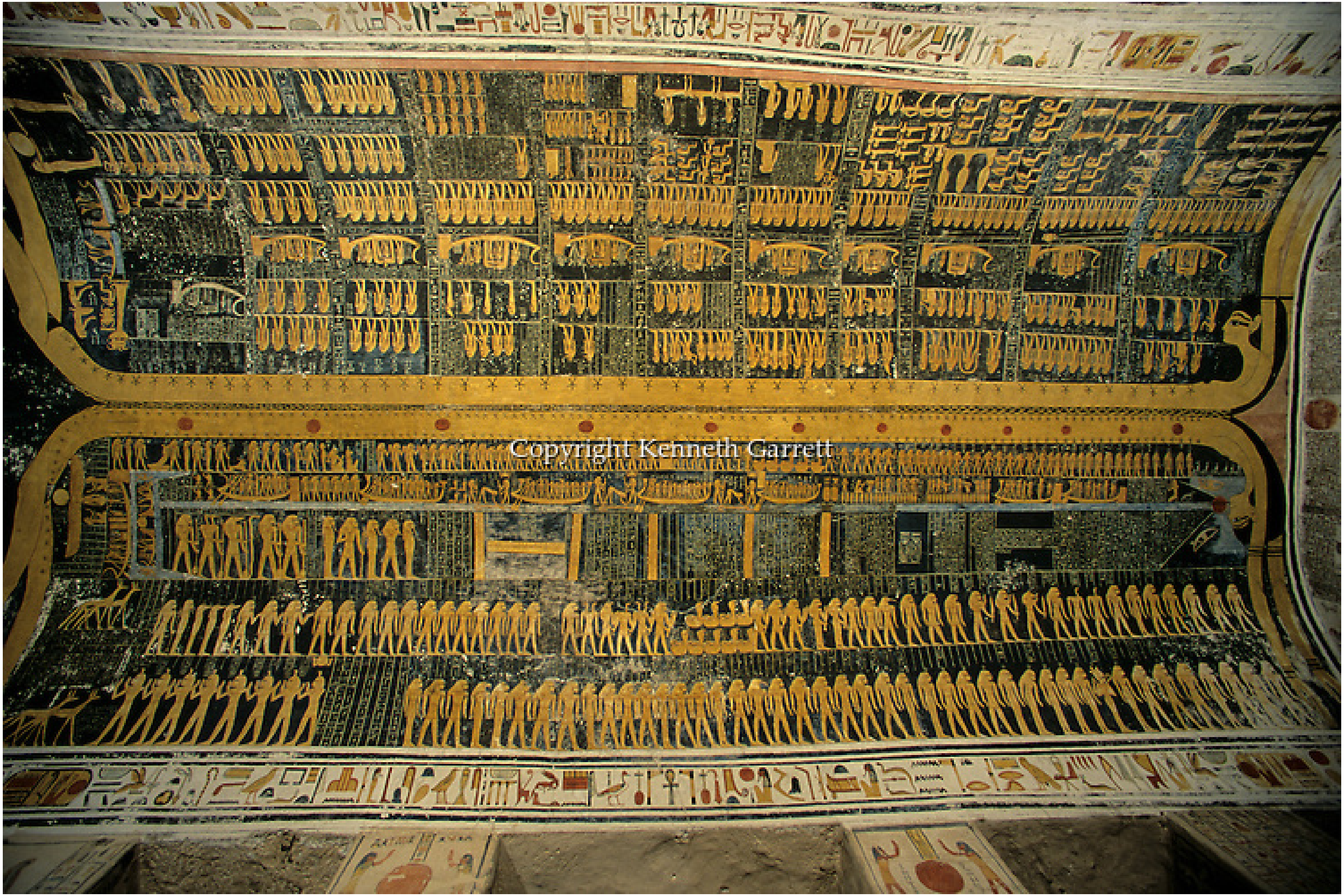

Copyright Kenneth Garrett

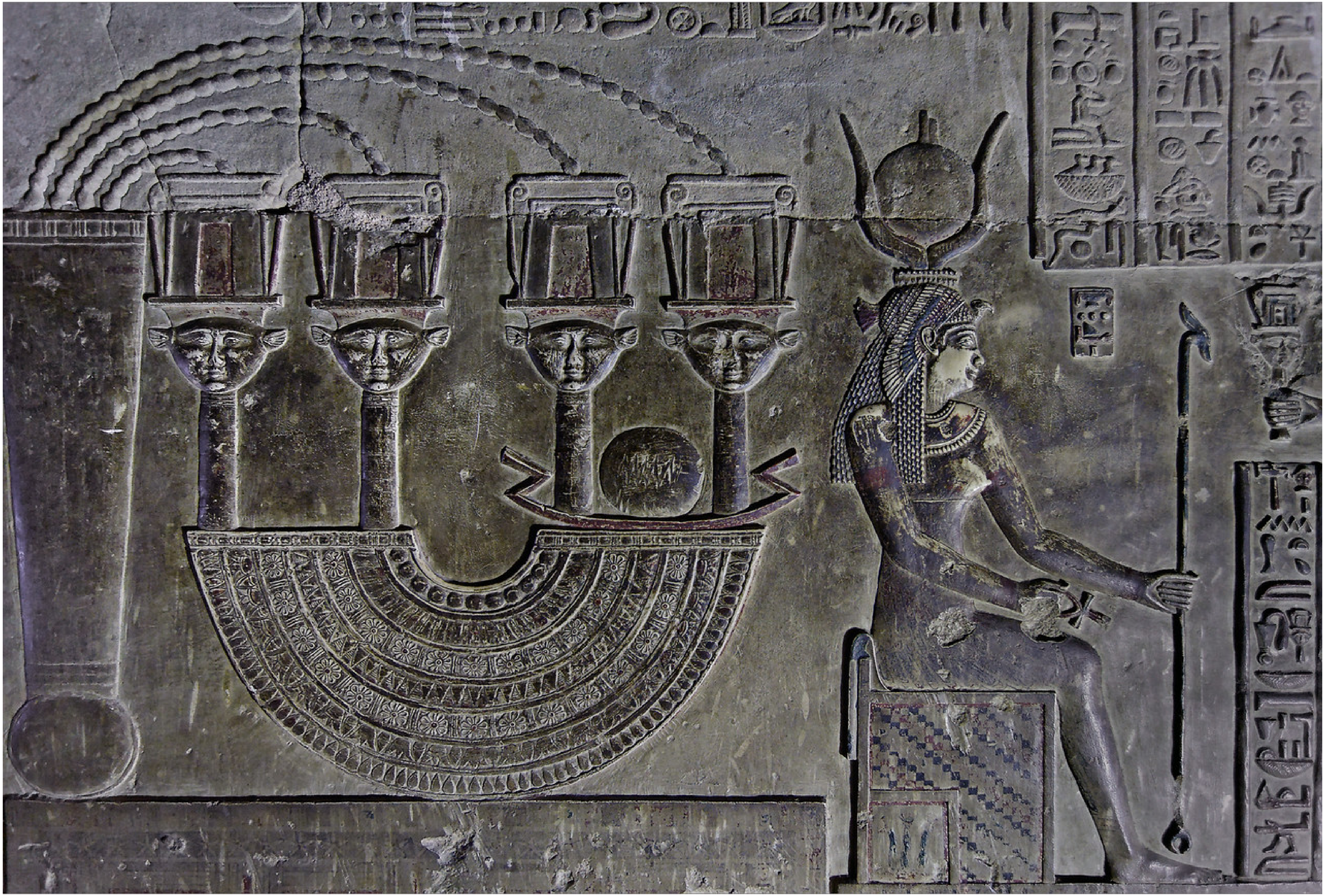

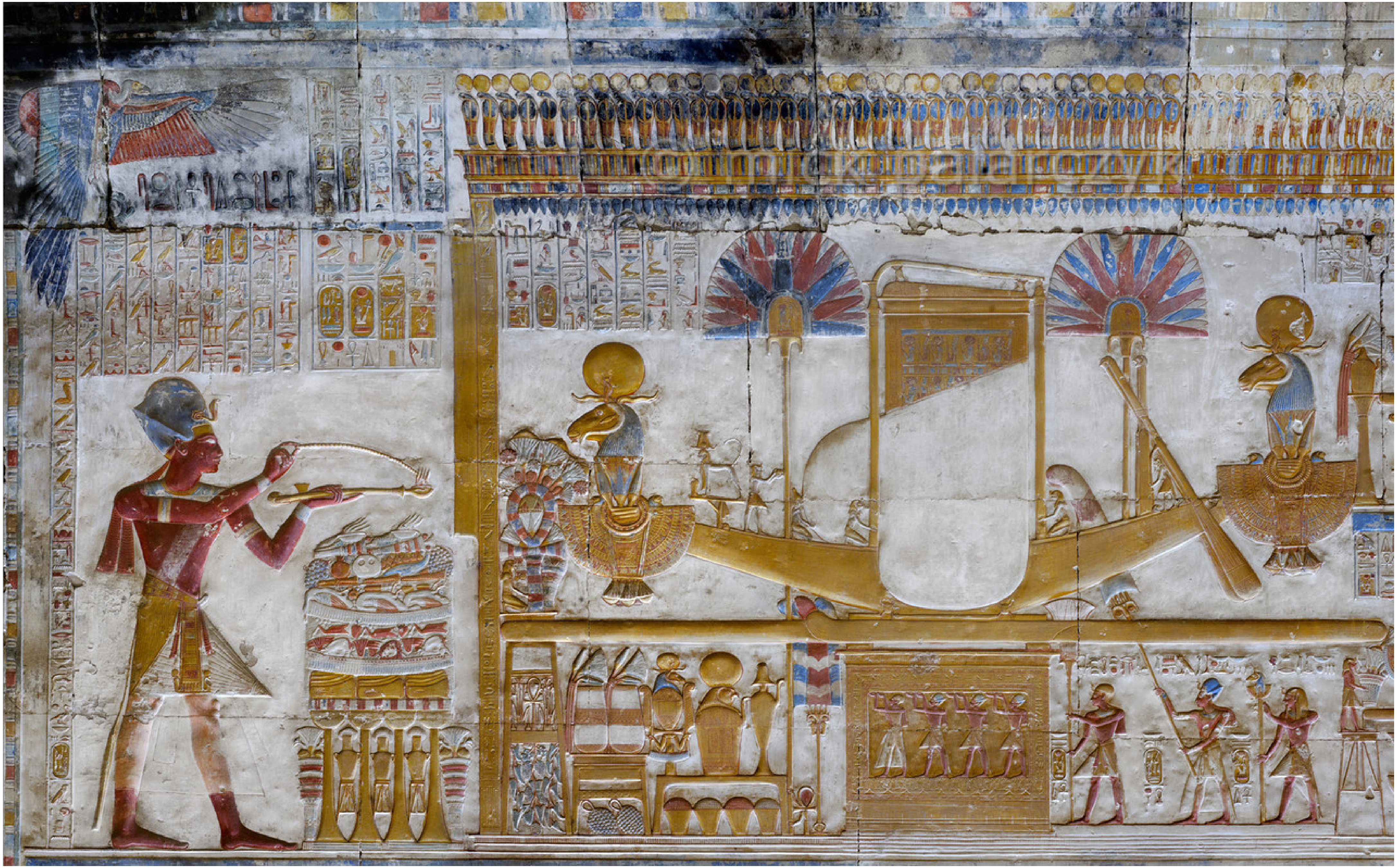

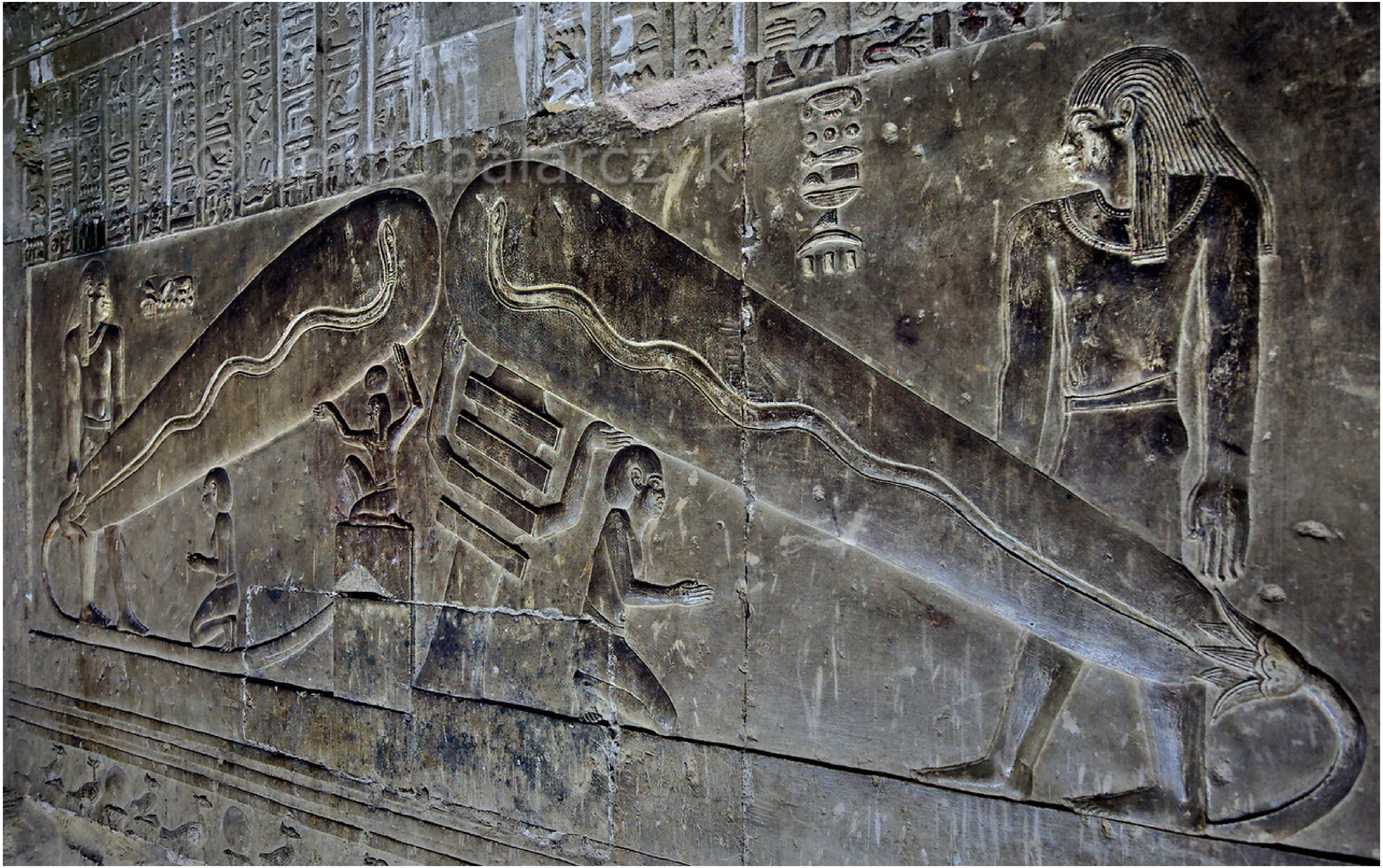

© [illegible] palarczyk

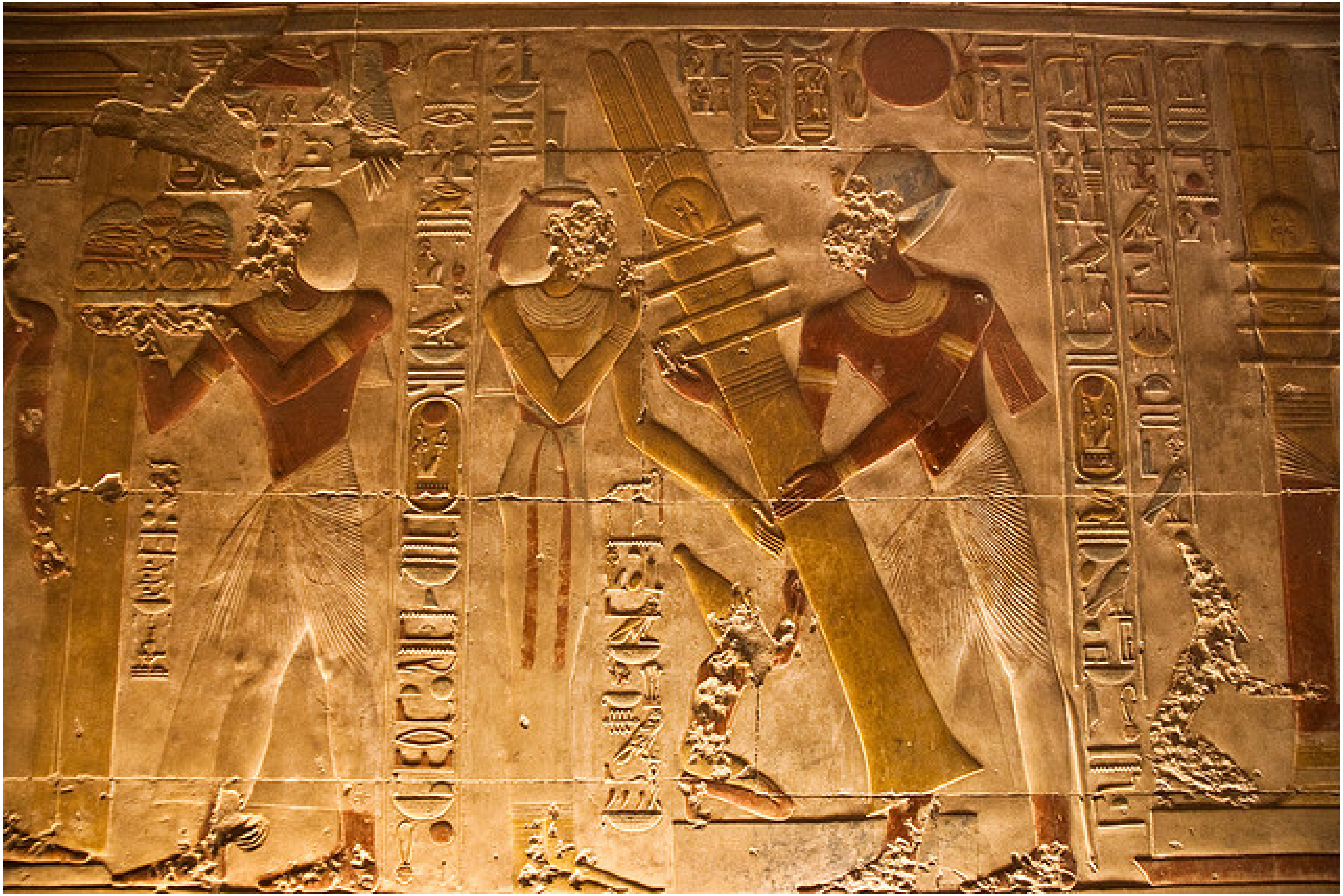

TALKS WITH TEACHINGS FROM MY COSMIC FRIENDS

ASHTAR SHERAN

ORTHON

angels-heaven.org

Pic. 4395 P EN

MEGADETH

PLATE 14

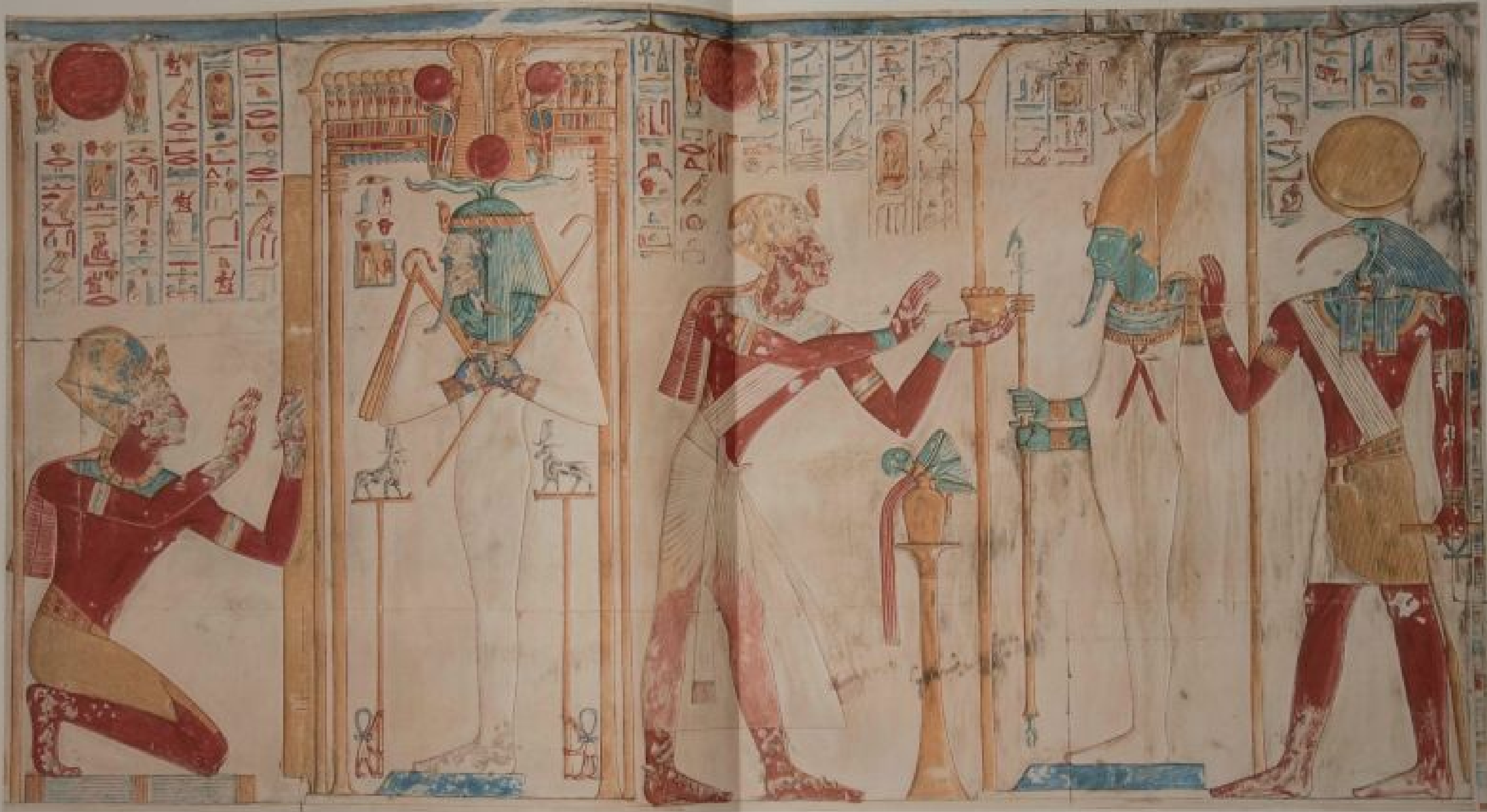

SETHOS WORSHIPS AND OFFERS NATRON TO OSIRIS

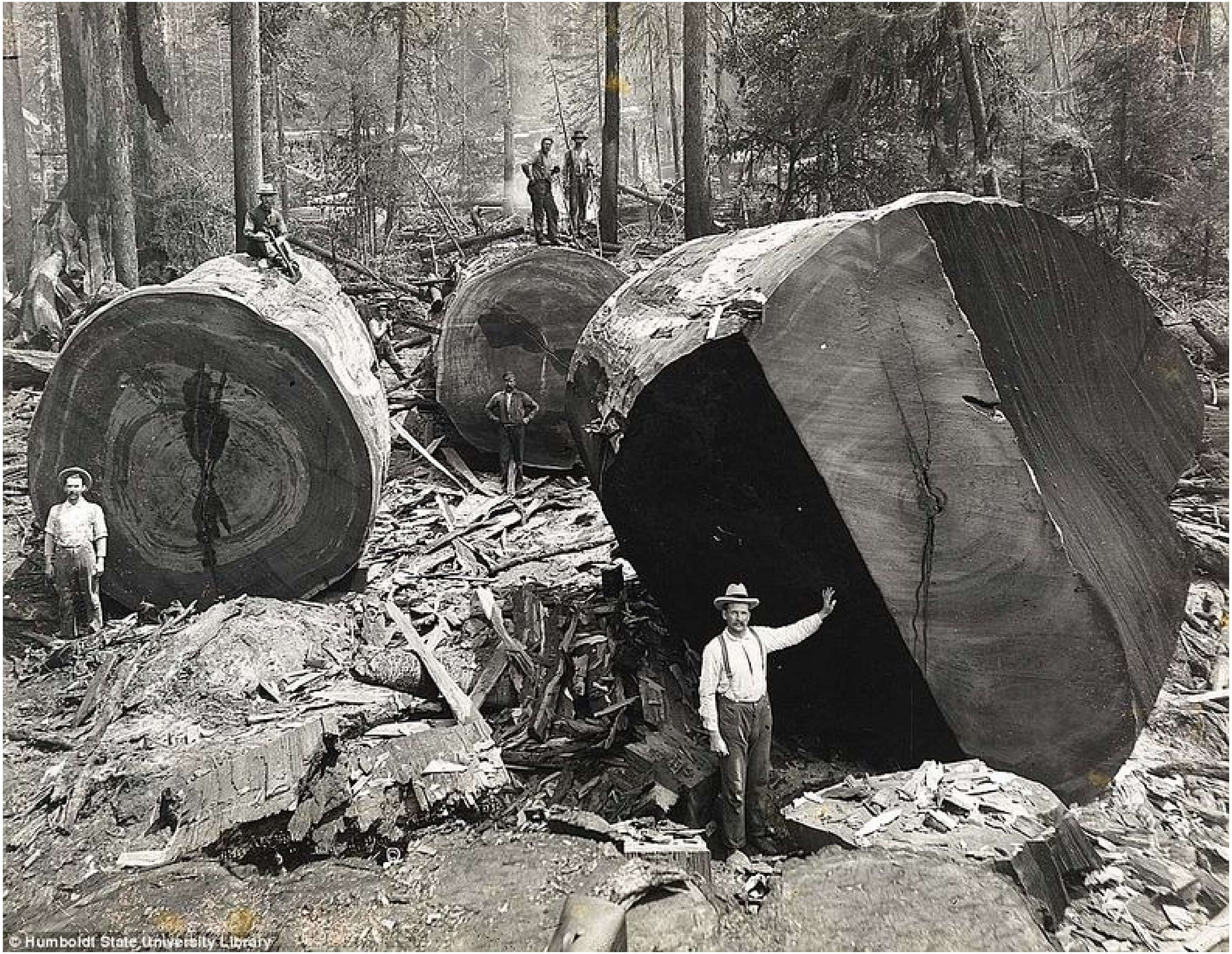

© Humboldt State University Library

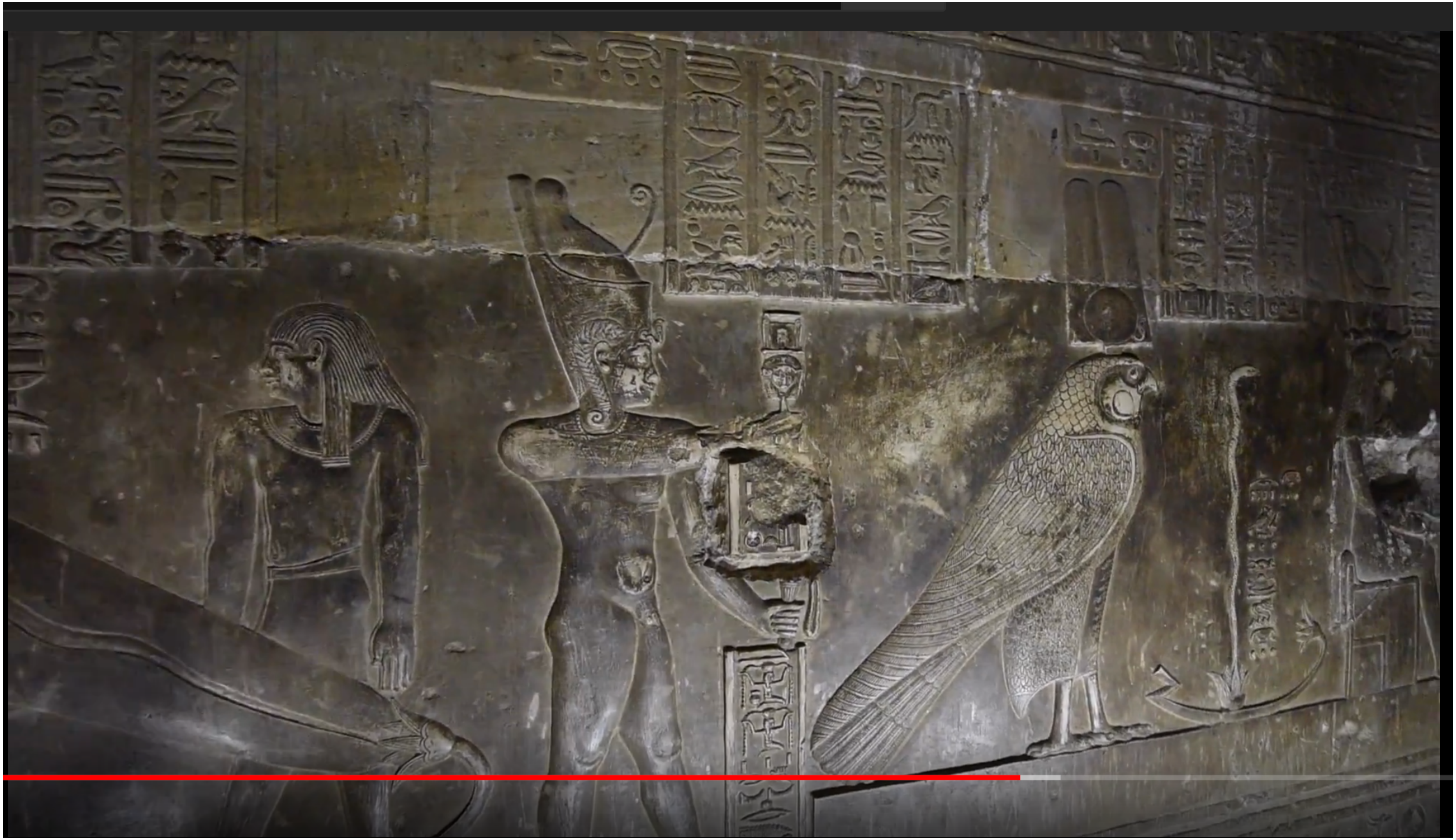

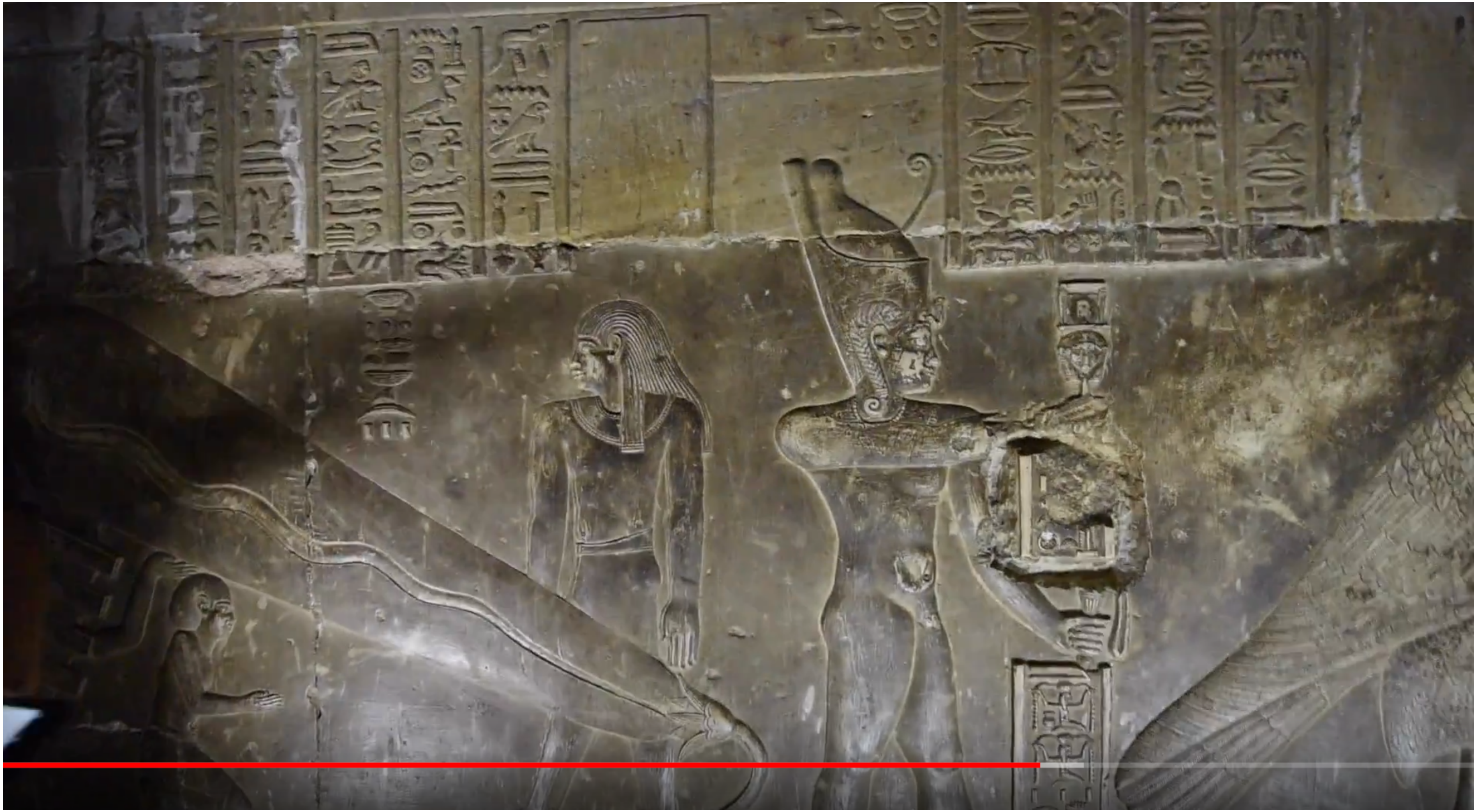

Meet The Iargans!

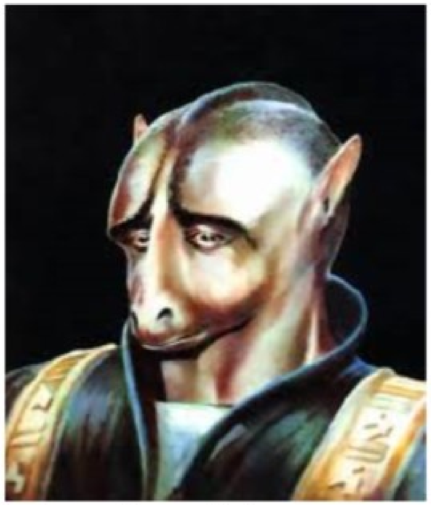
Typical Iargan

Iargan gets into photo
without being seen at
the time photo was taken!

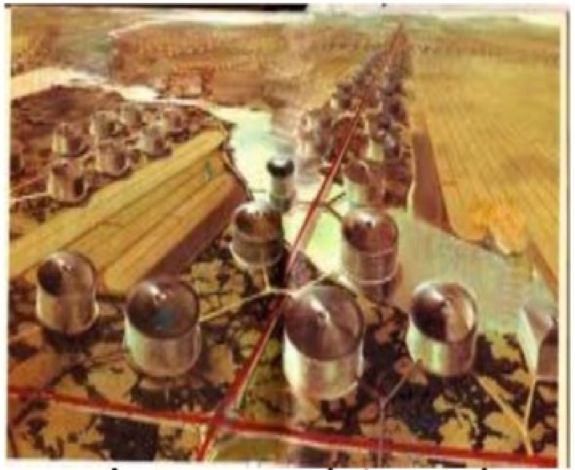
Iargan society and
transportation infrastructure
of housing units

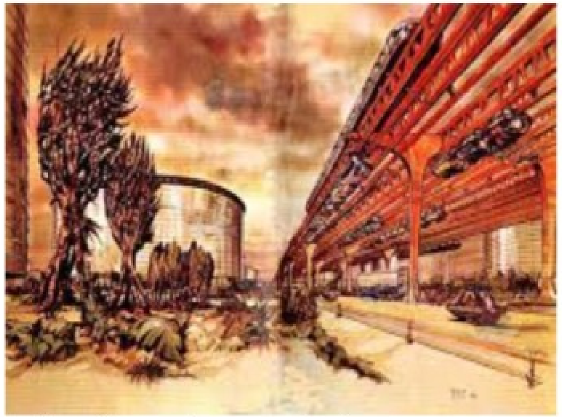
Another view and close up
of cylinder which houses 10 people each!

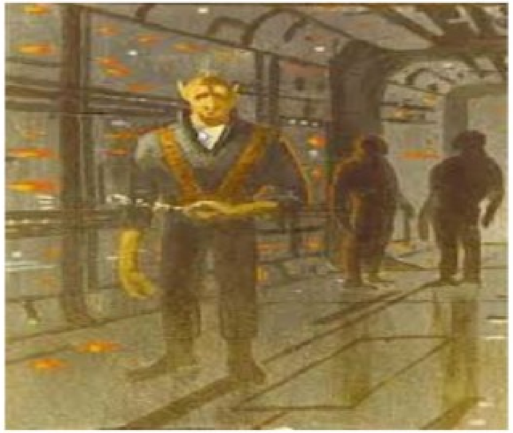
From Inside The Control Room

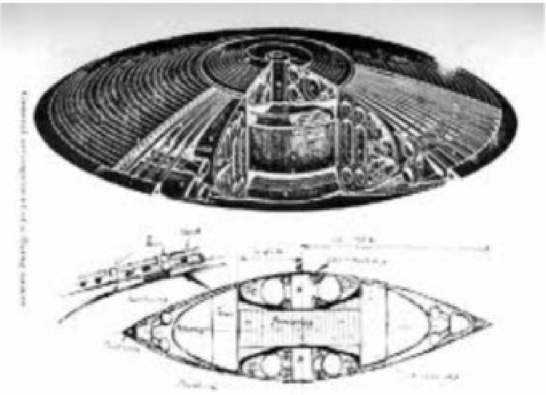
The blue print of a Sun wheel
which holds 10,000 people as well

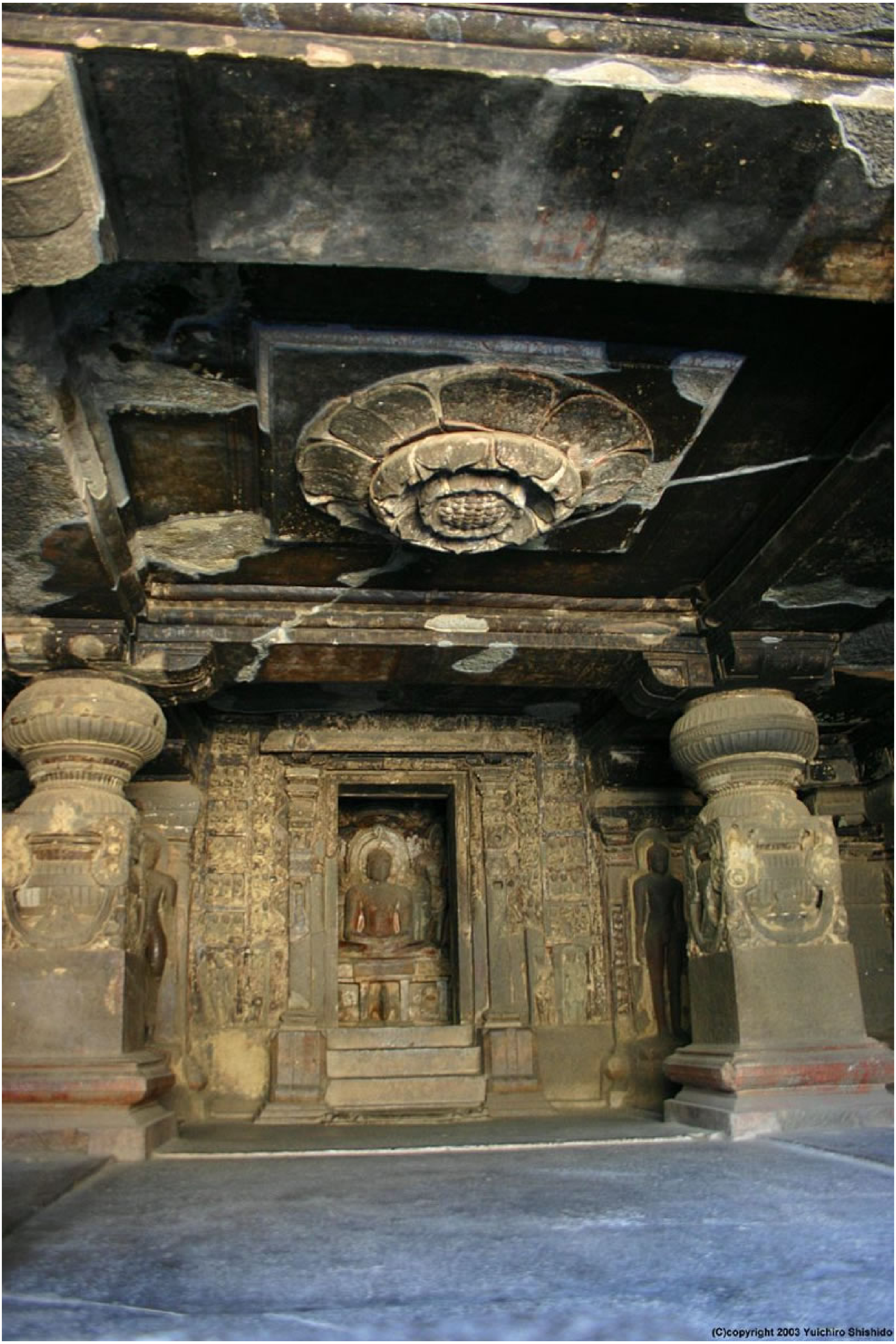

(C)copyright 2003 Yuichiro Shishido

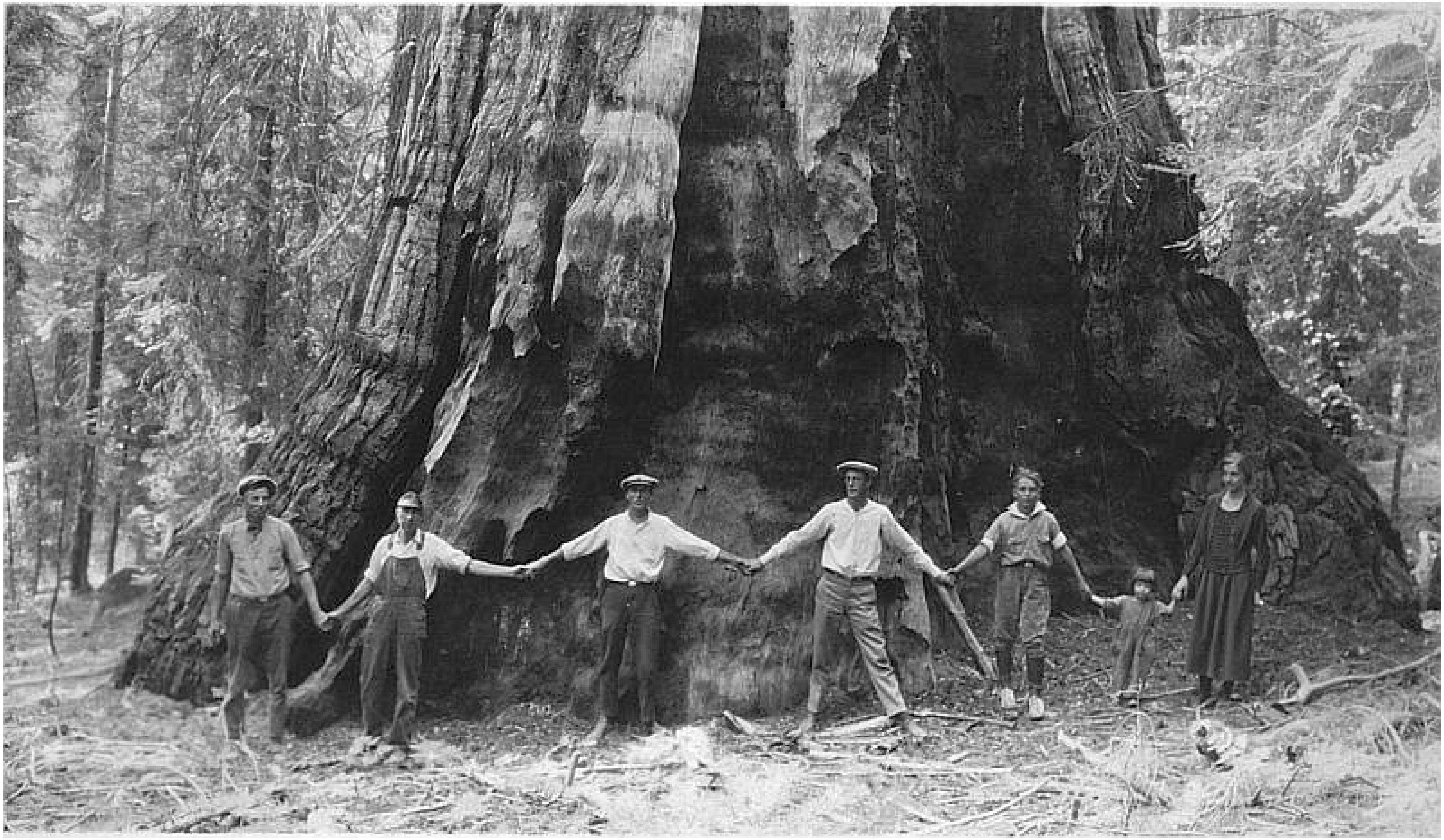

By Mohamed Ibrahim

By Mohamed Ibrahim

By Mohamed Ibrahim

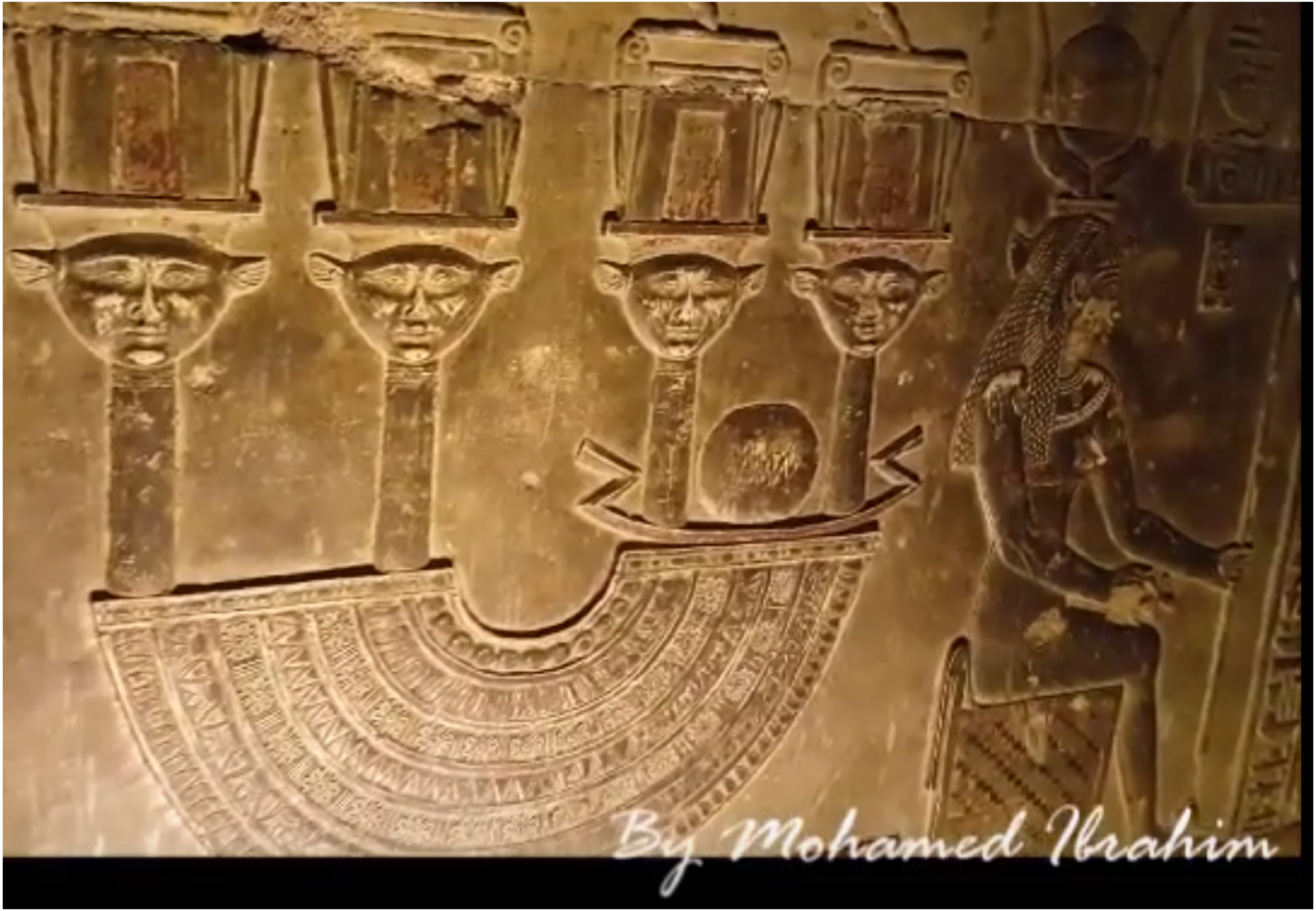
By Mohamed Ibrahim

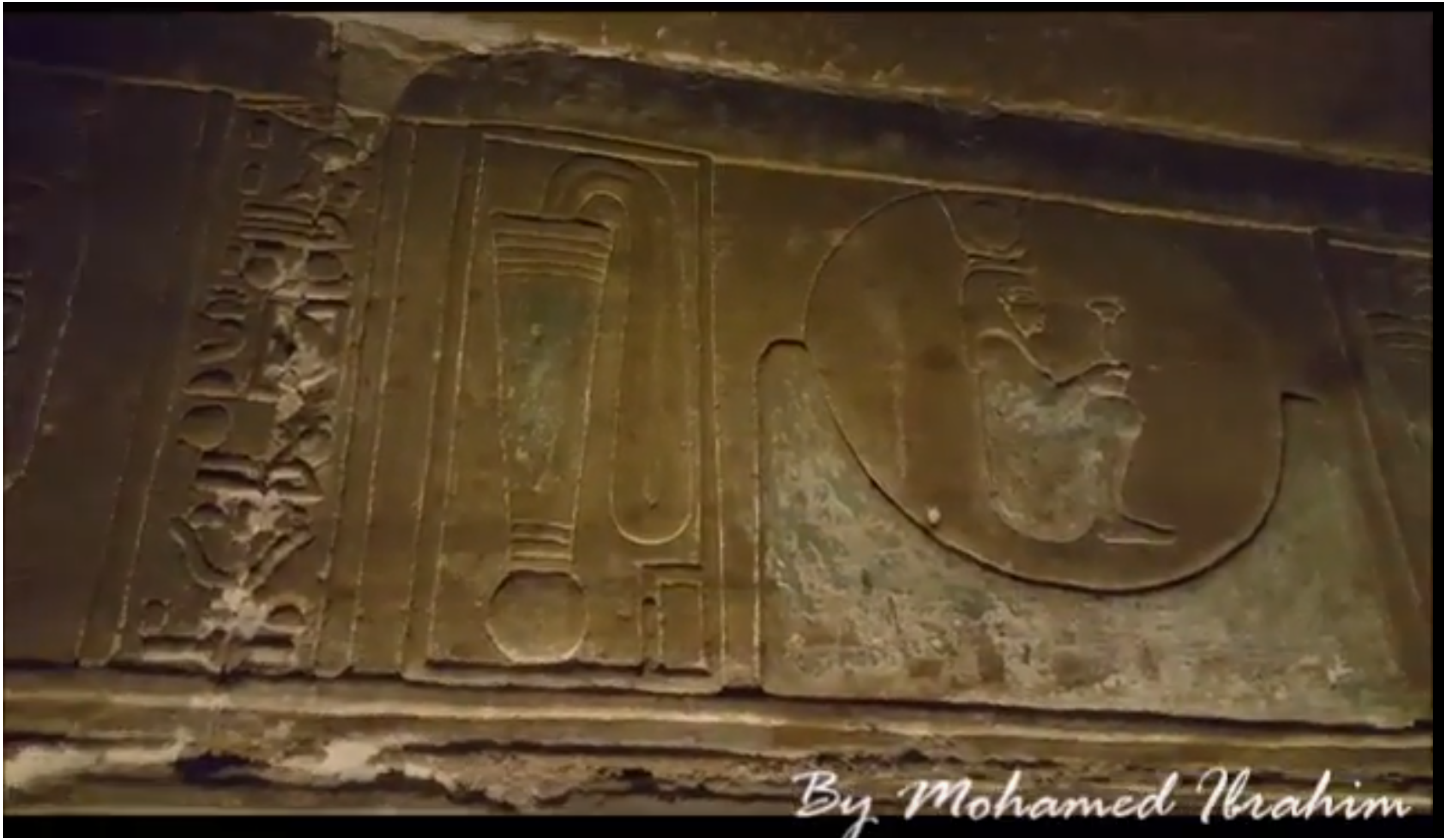
By Mohamed Ibrahim

By Mohamed Ibrahim
4:46

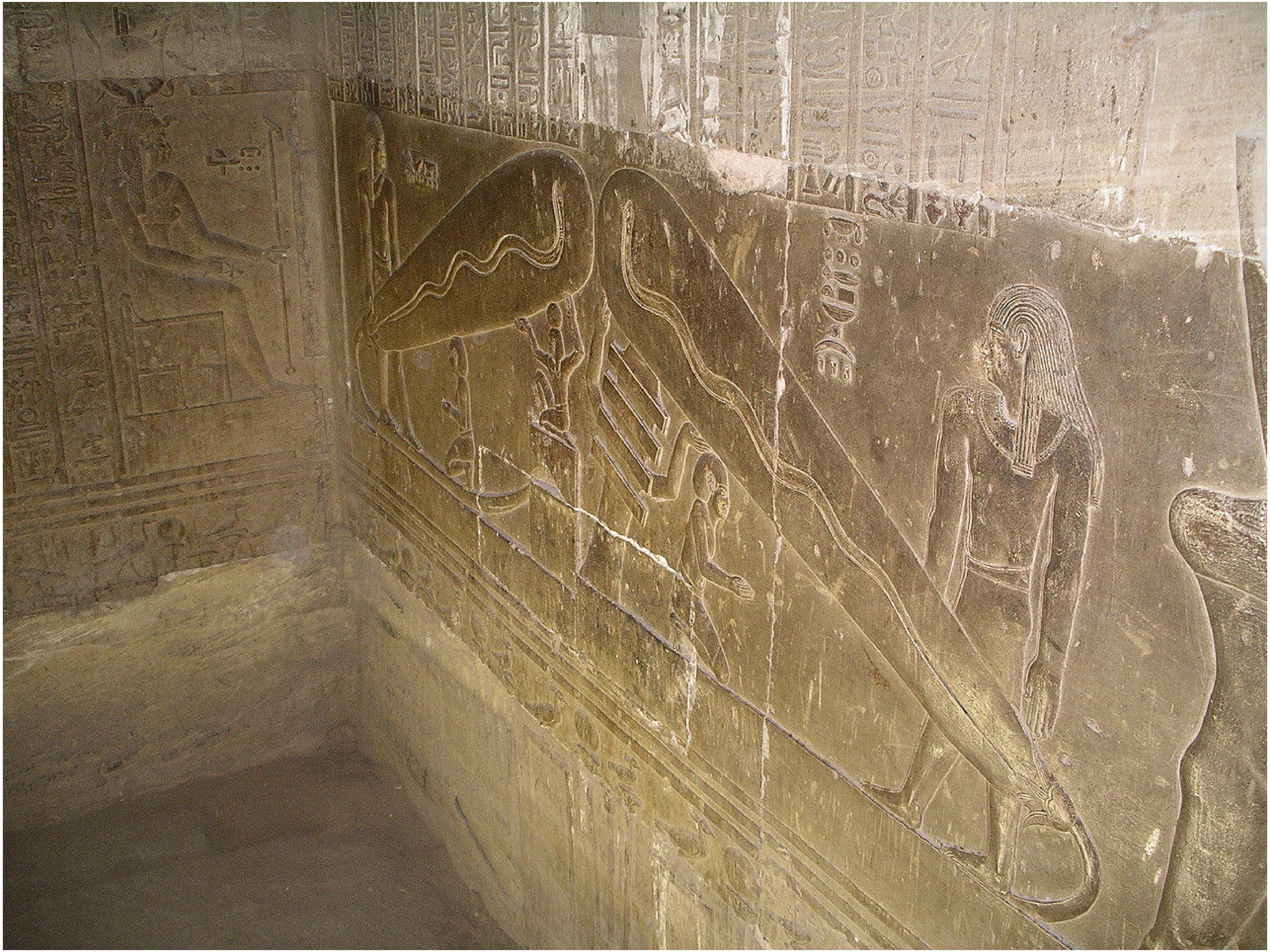

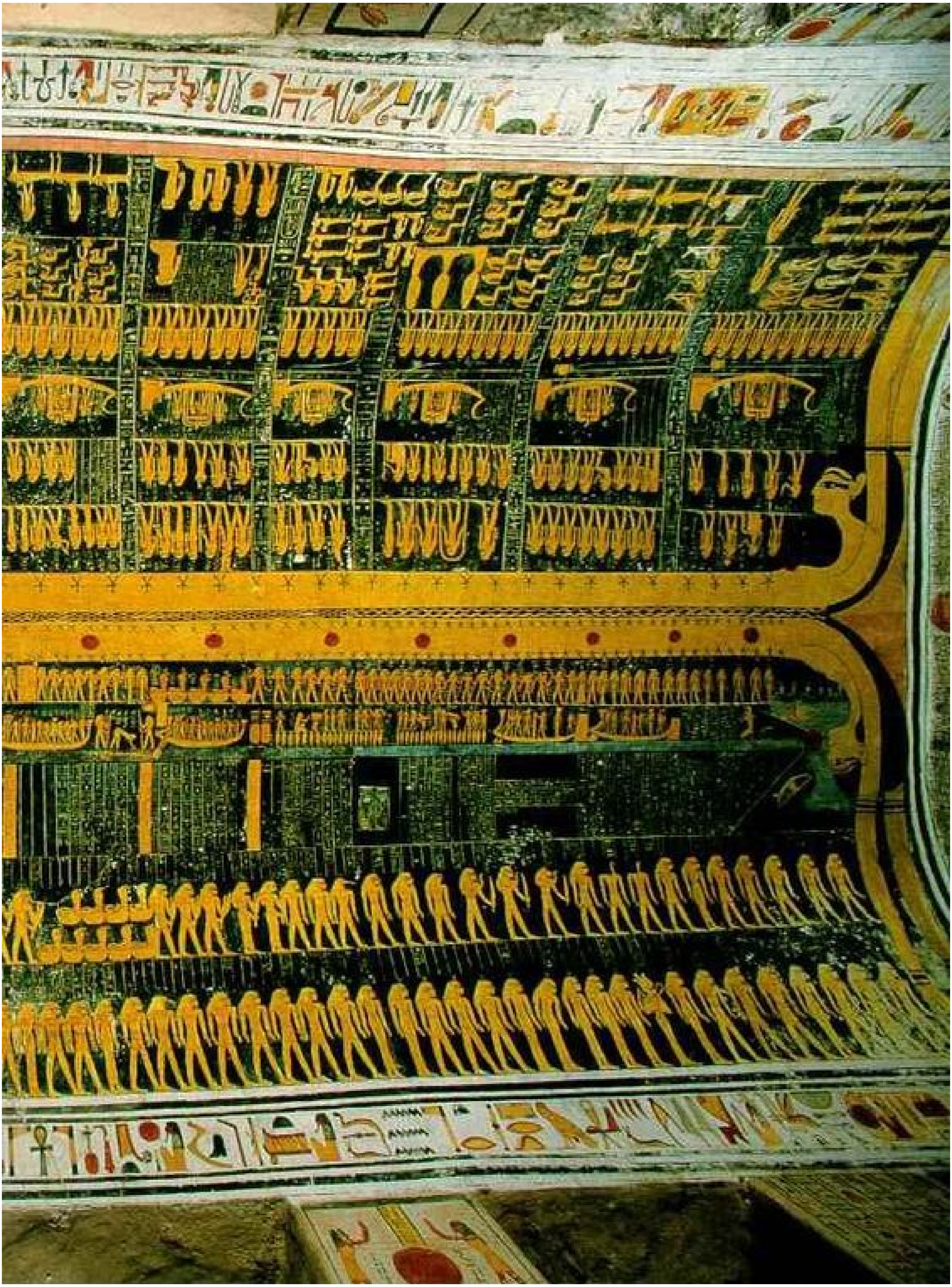

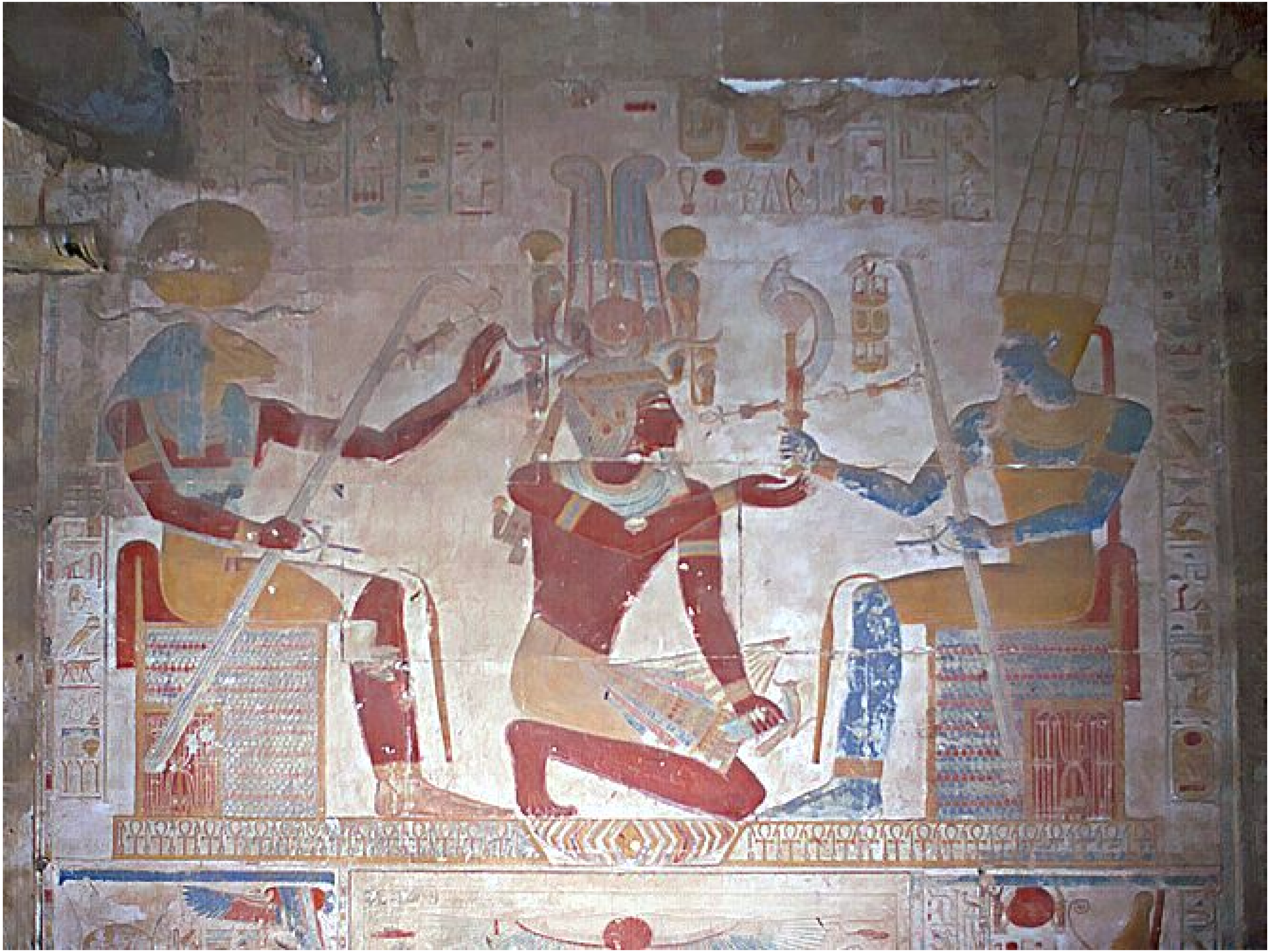

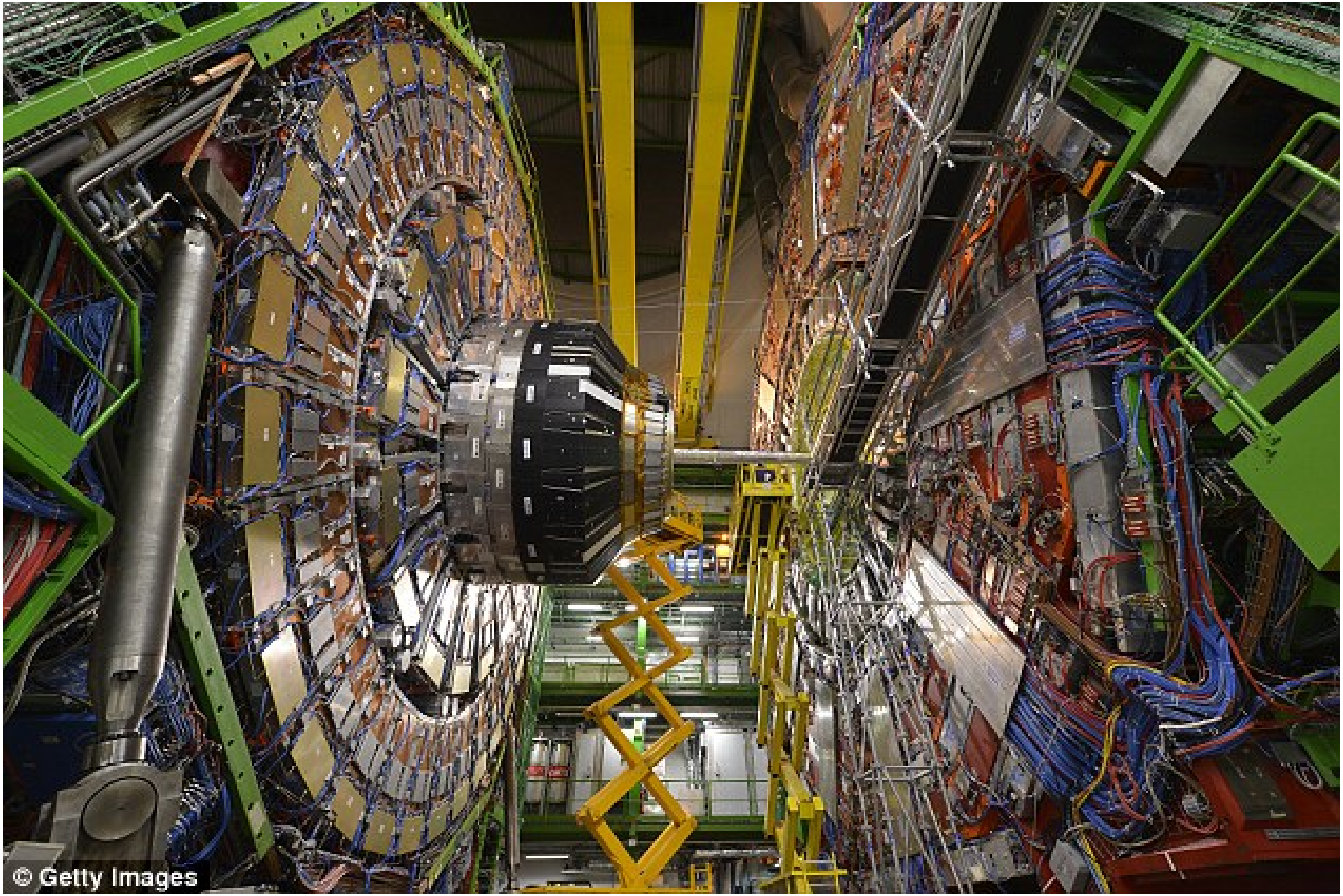

© Getty Images

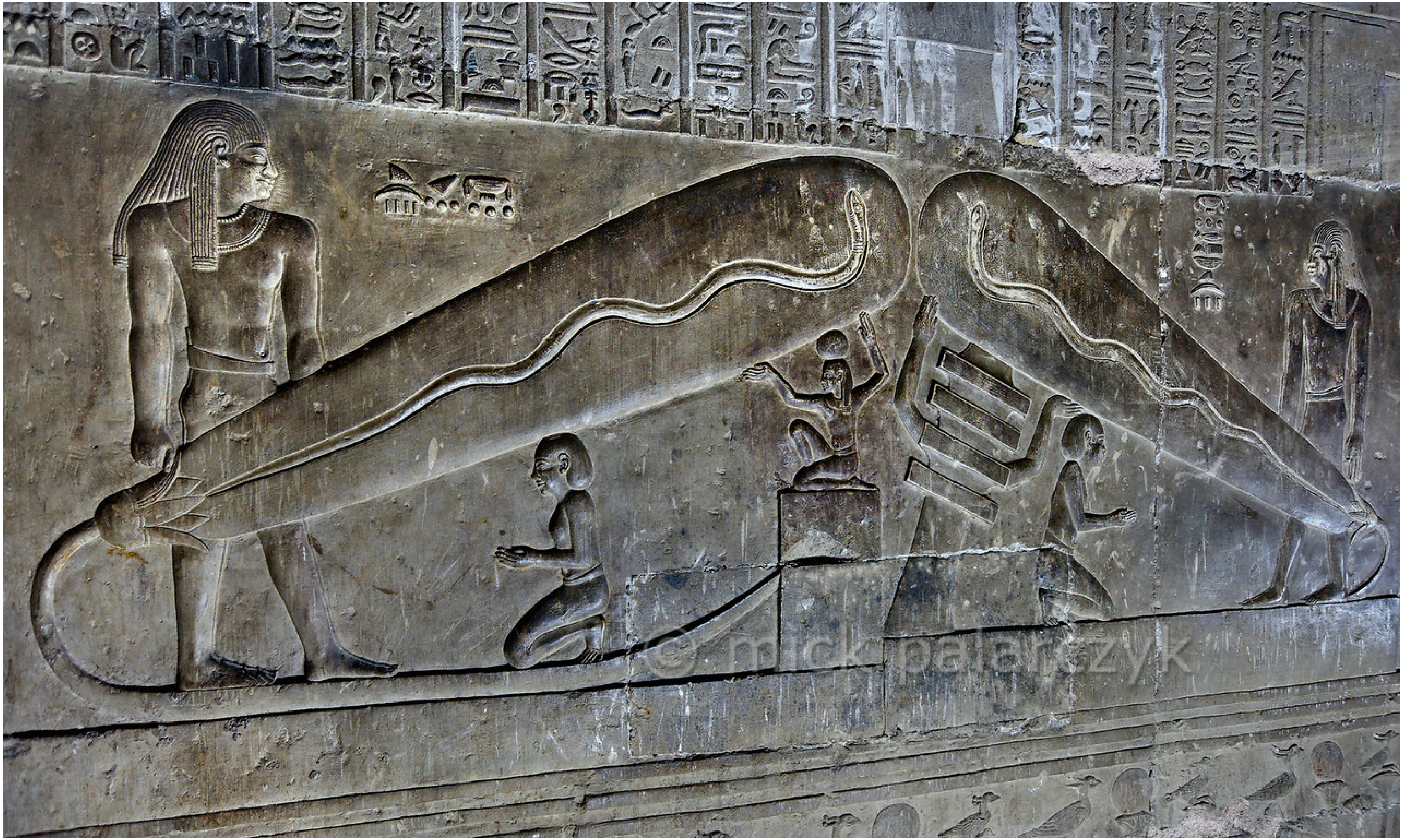

© mick palarczyk

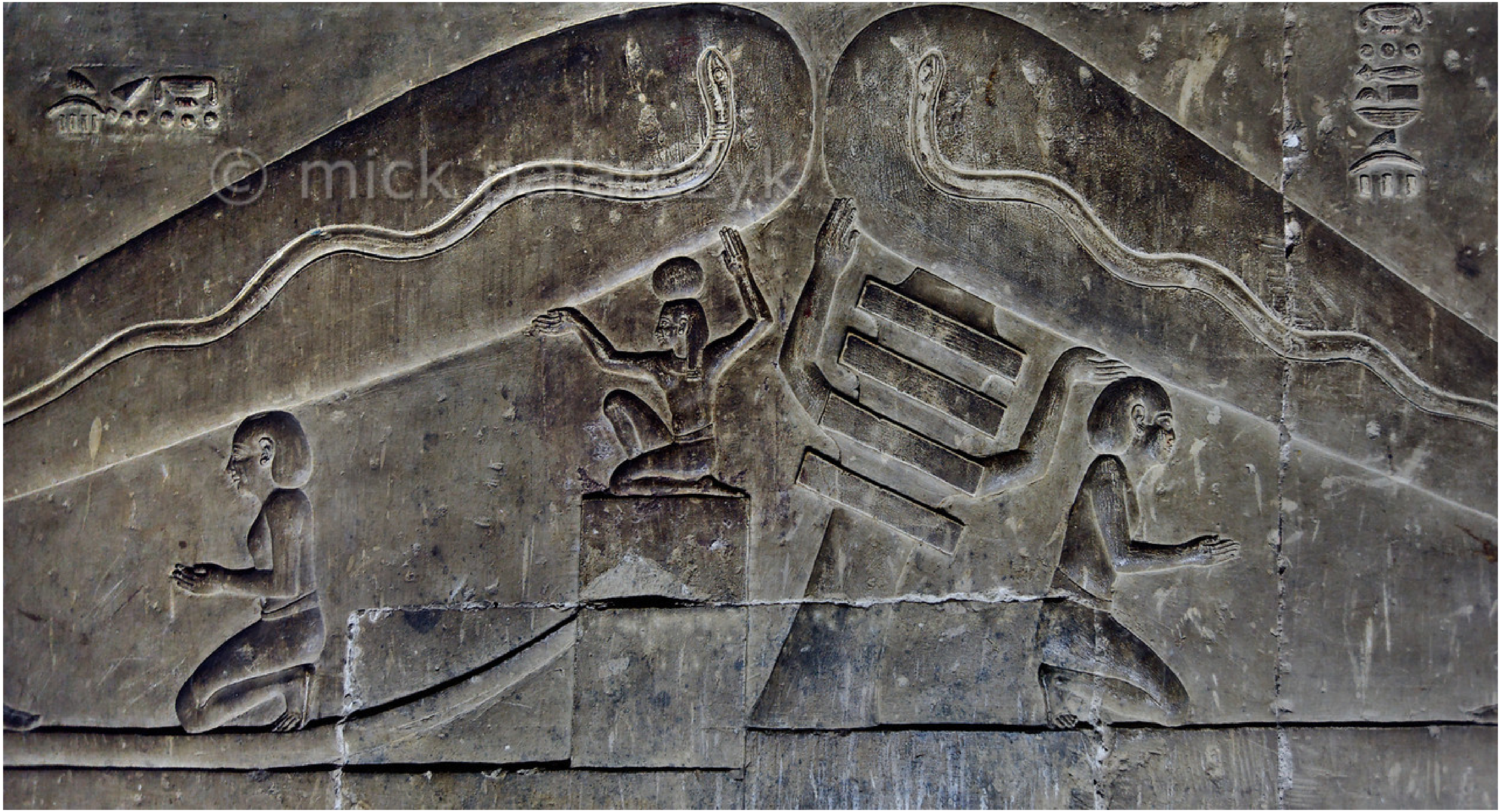

© mick [illegible]k

© mick palarczyk

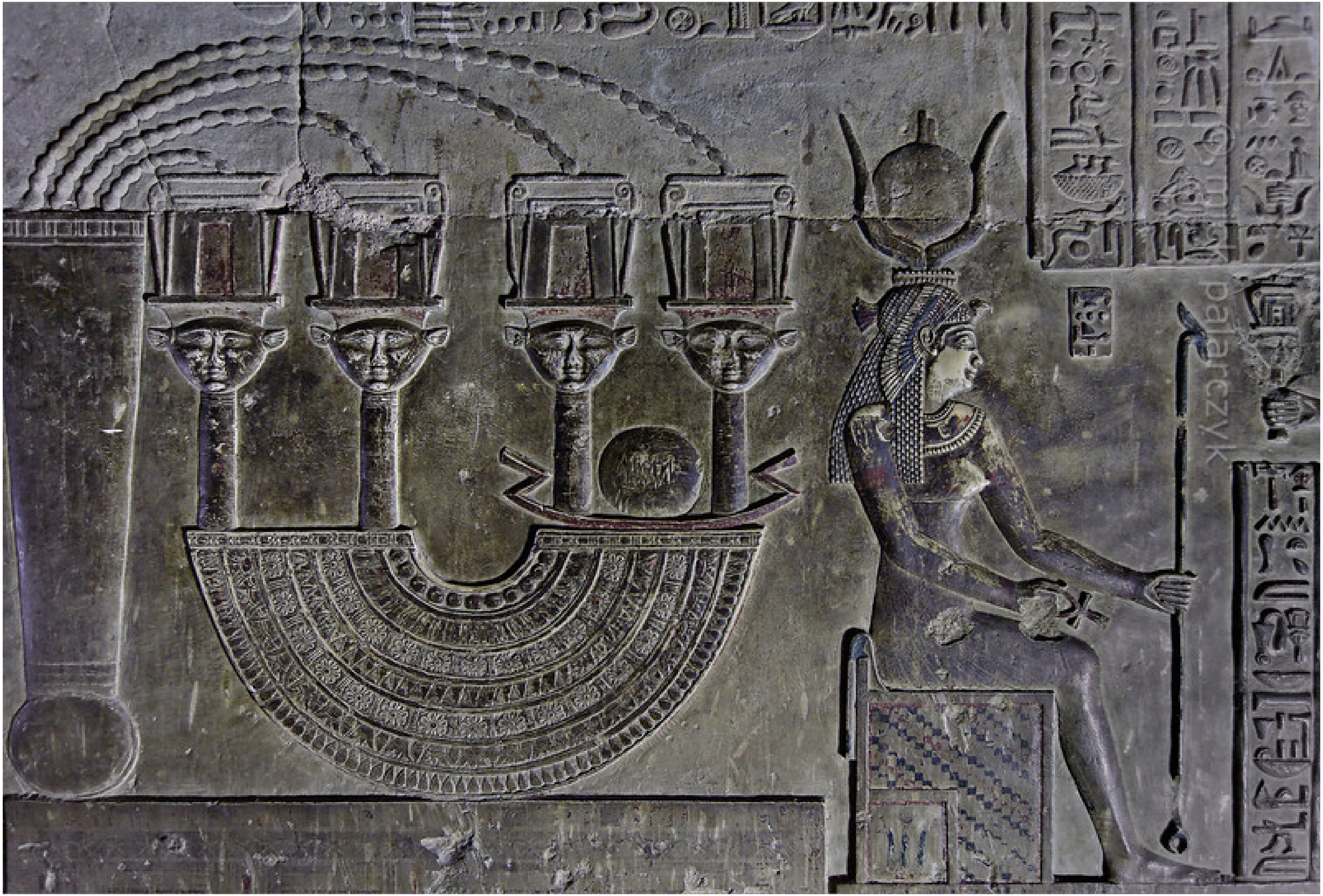

palarczyk

© paul smit

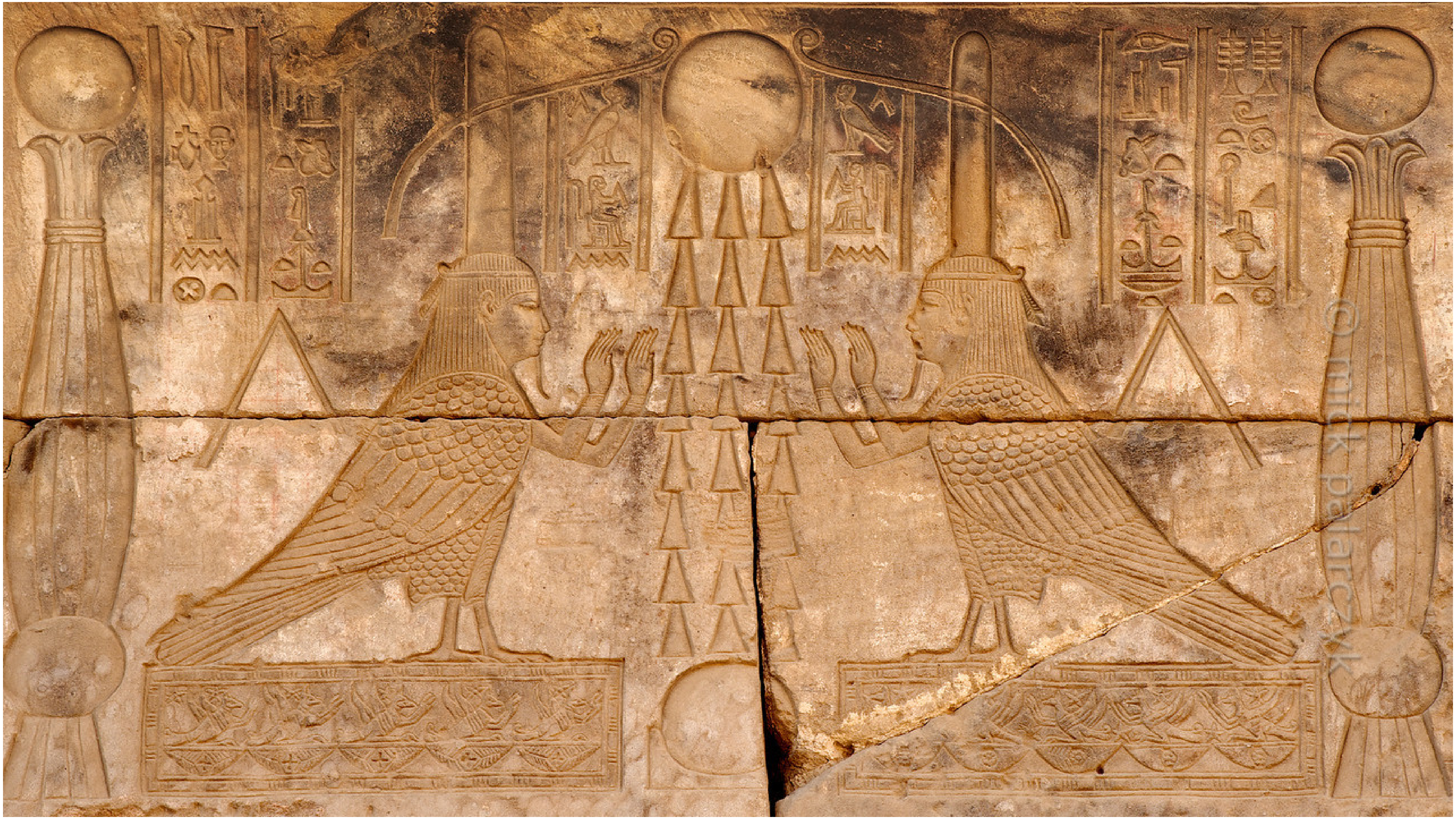

© mick palarczyk

©www.imagesofanthropology.com

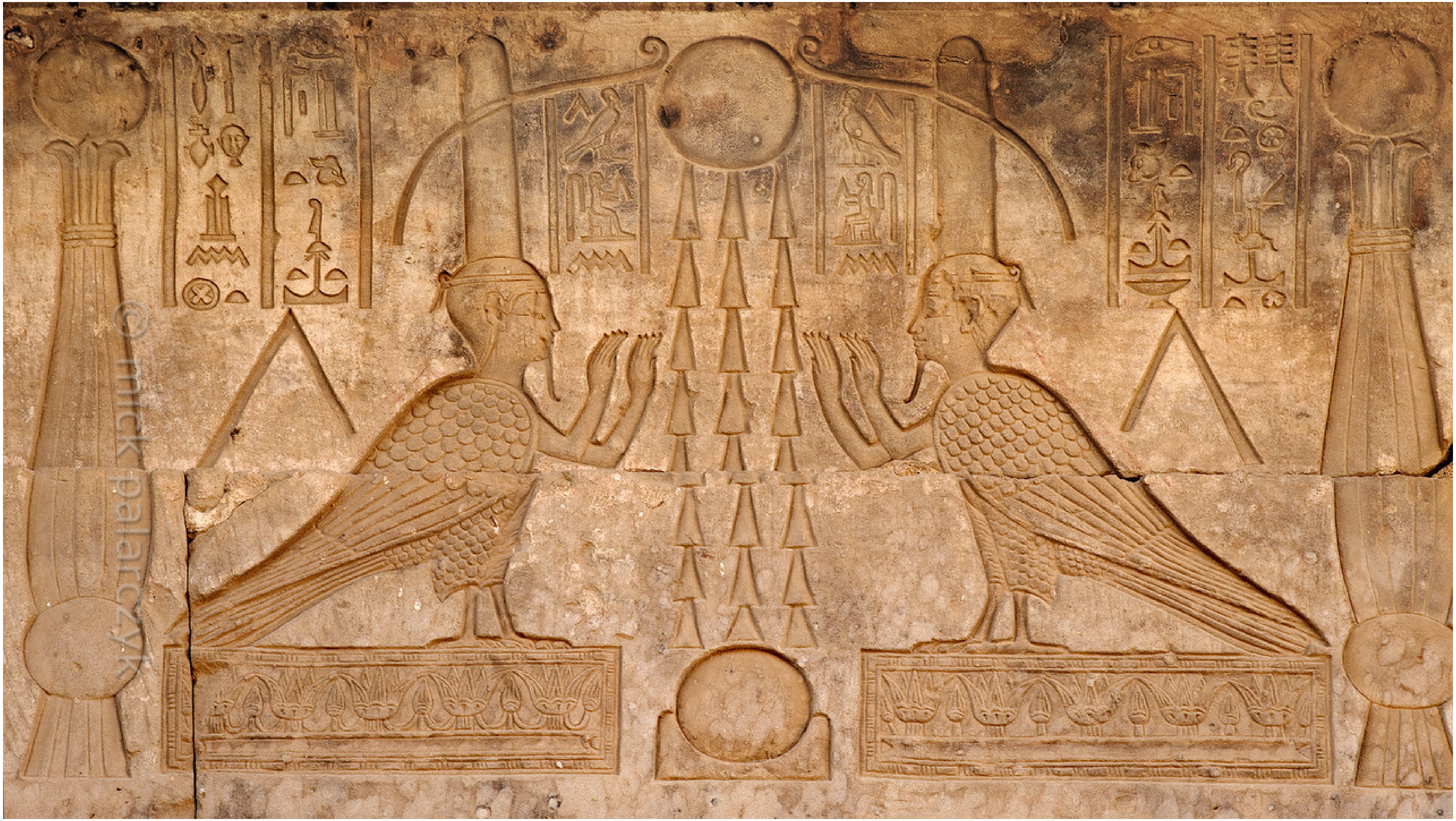

© mick palarczyk

© mick palarczyk

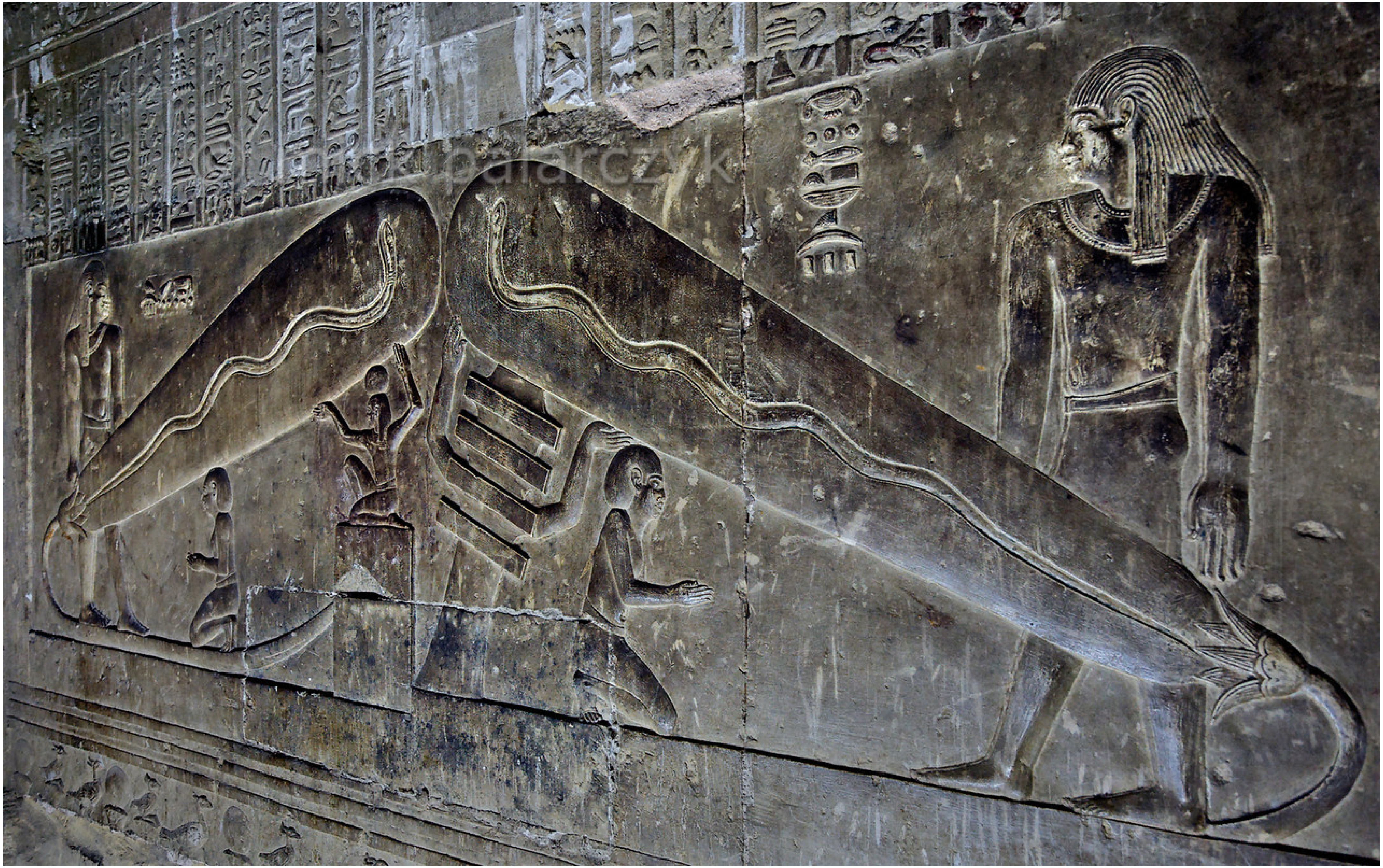
© [illegible] palarczyk

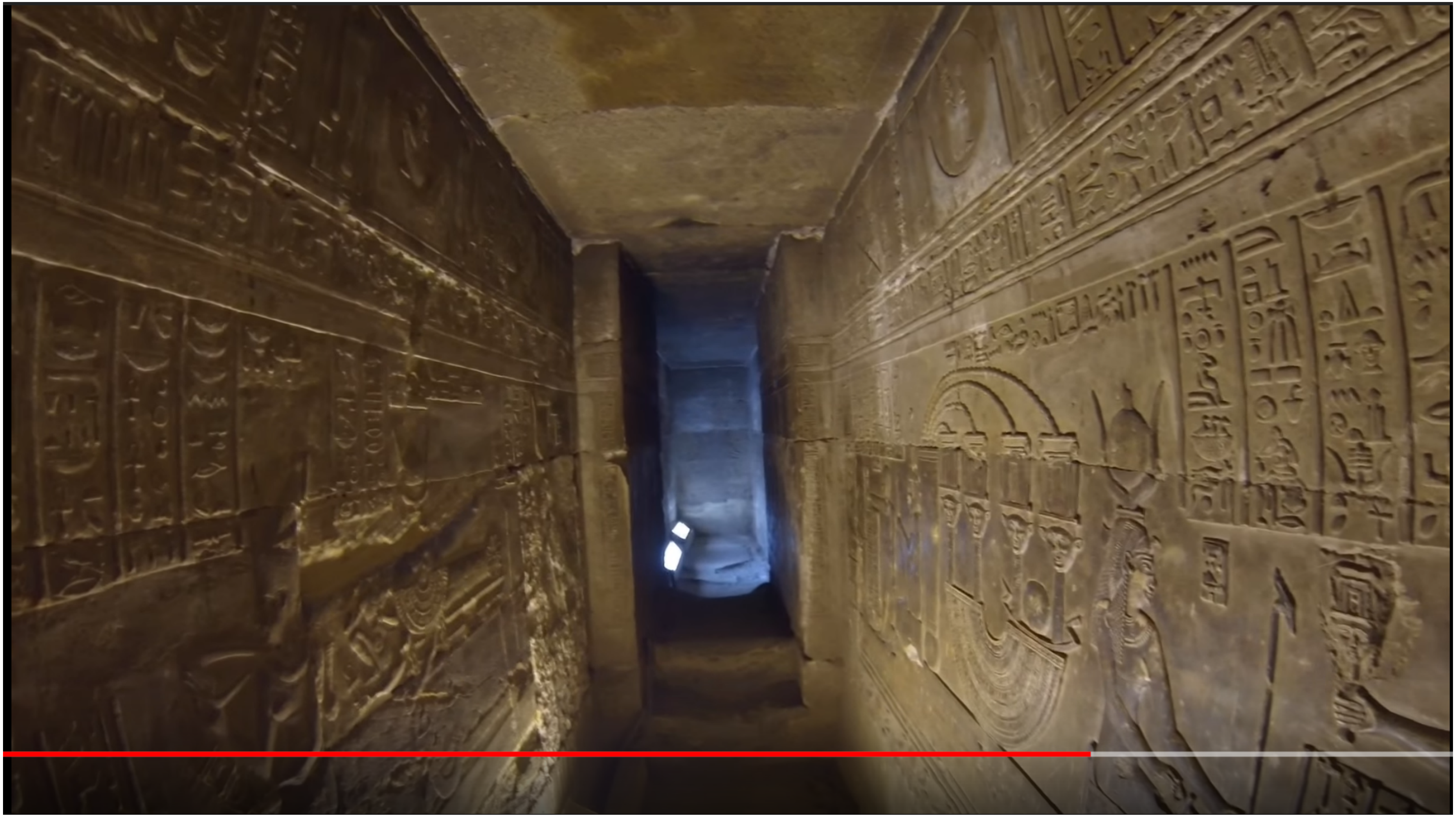

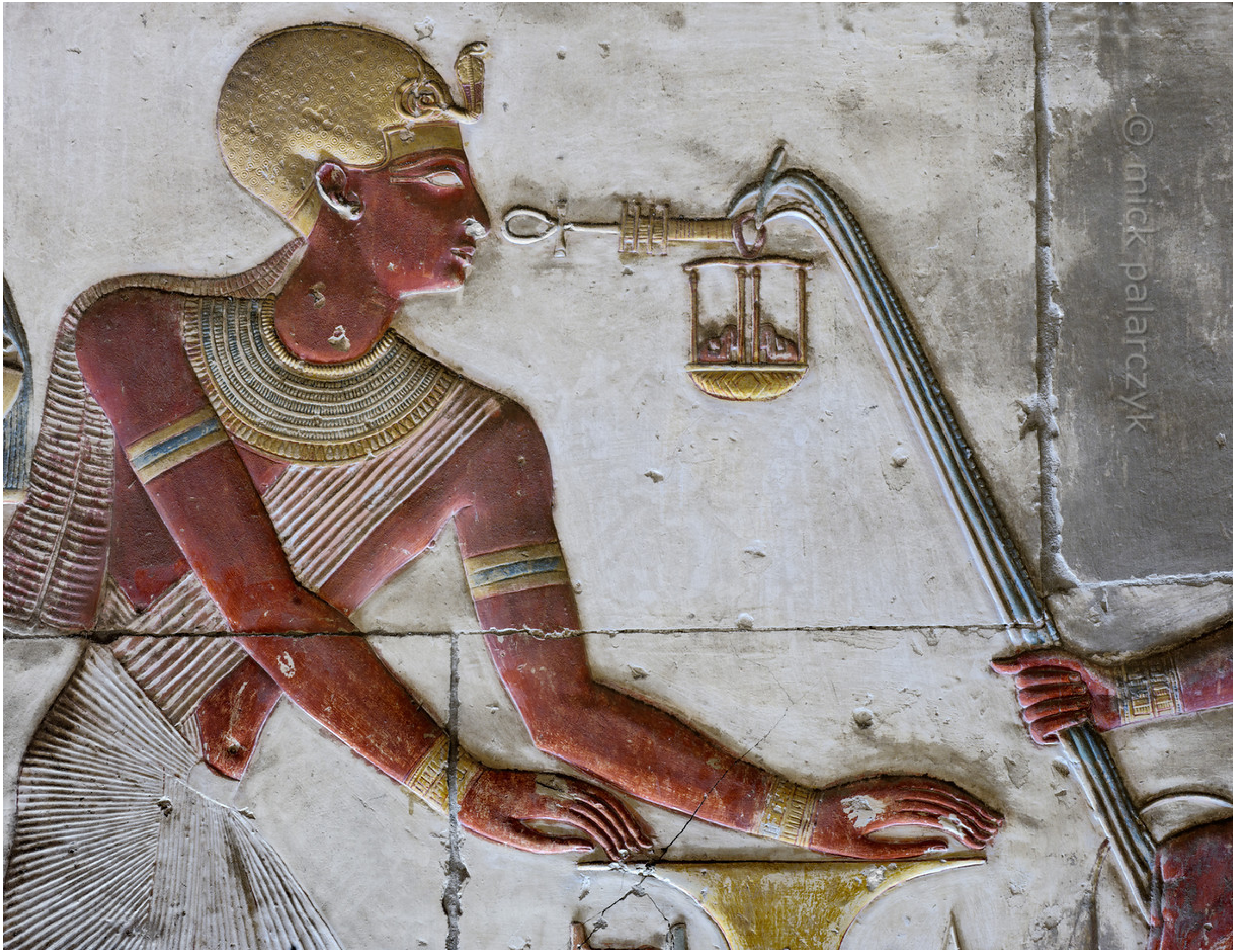

© mick palarczyk

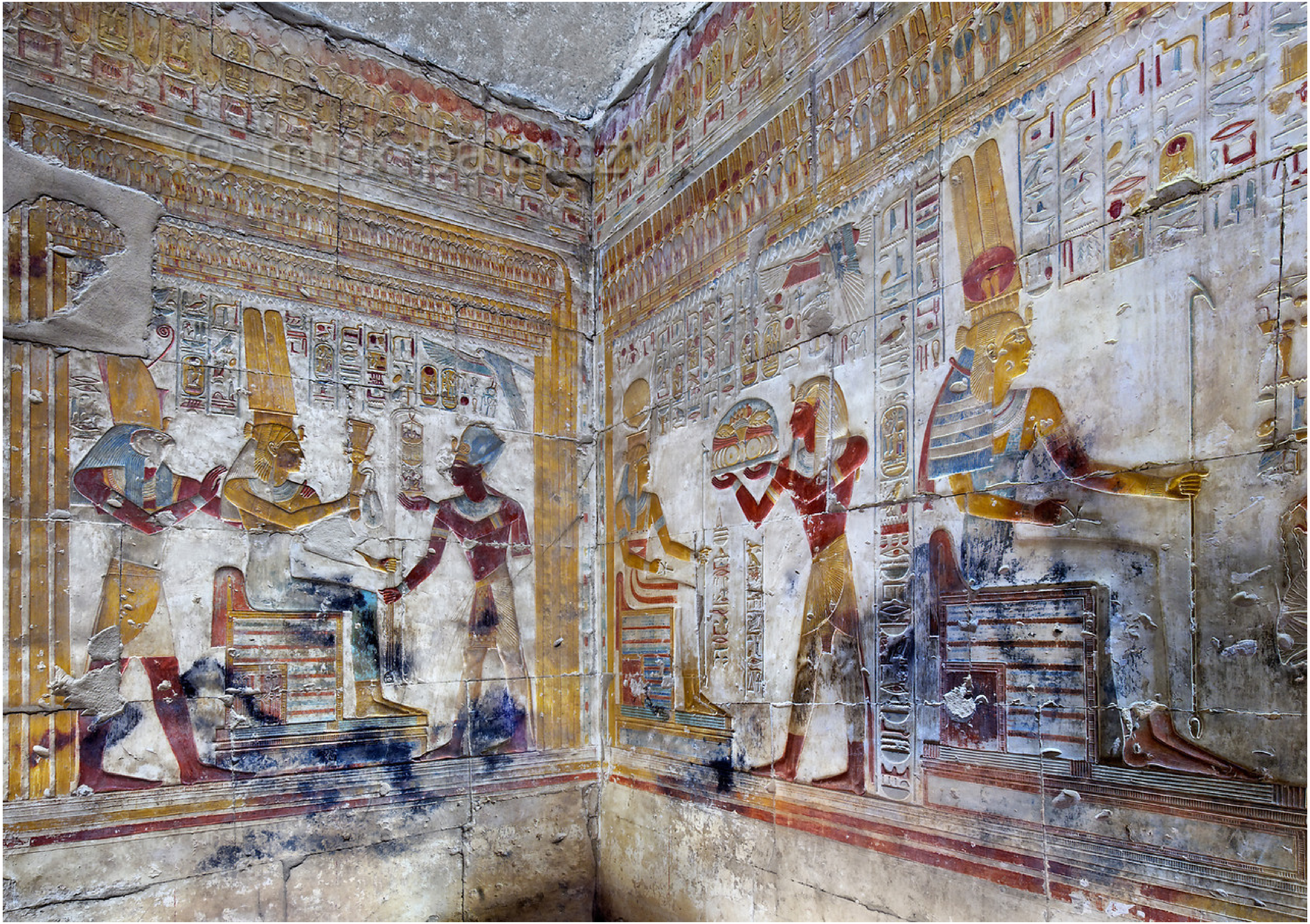

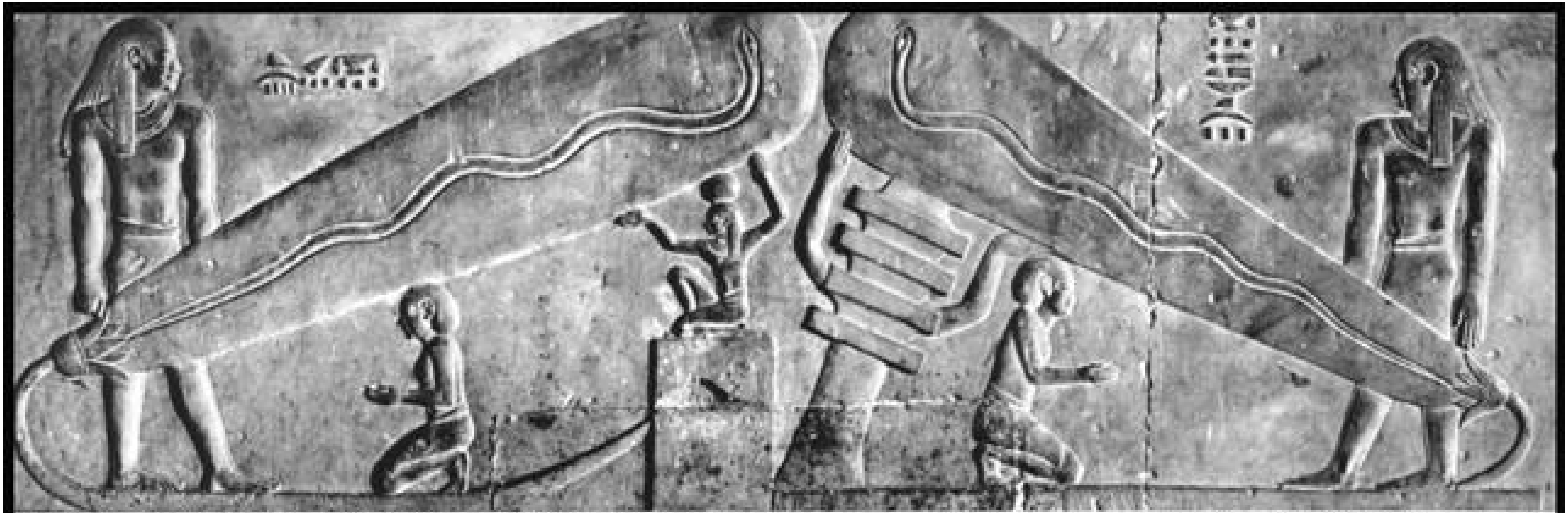

Battery-powered searchlights and filament lamps in the crypts of the Temple of Hathor (Isis) at Denderah

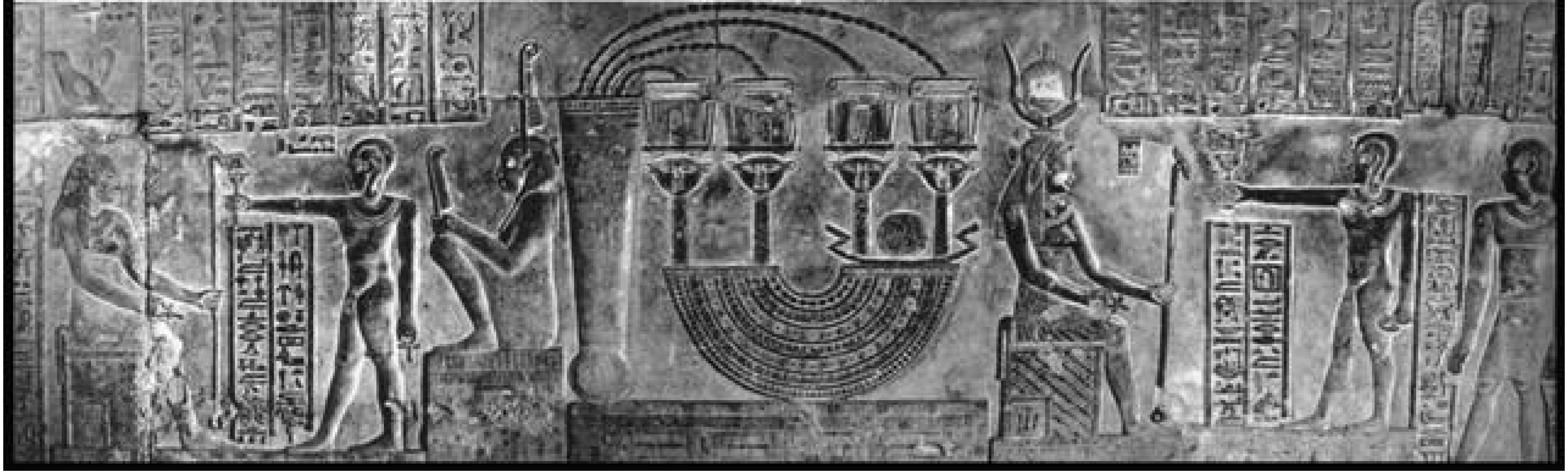

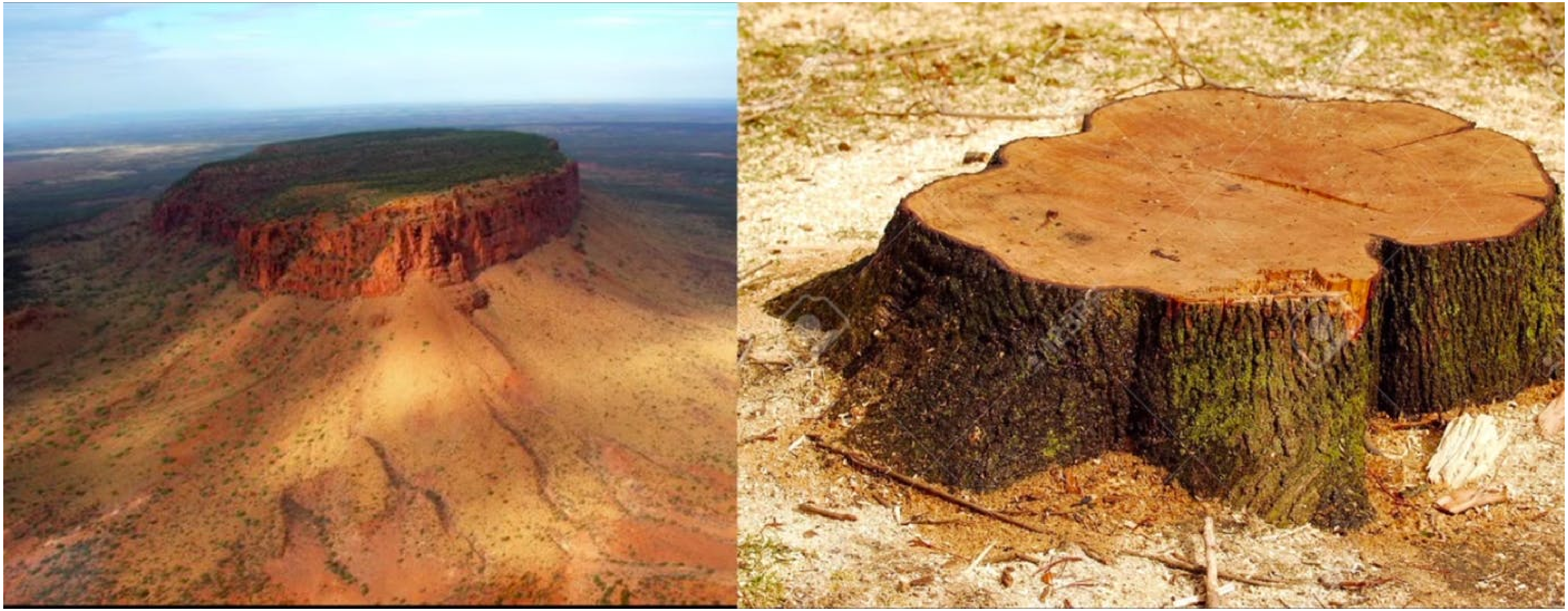

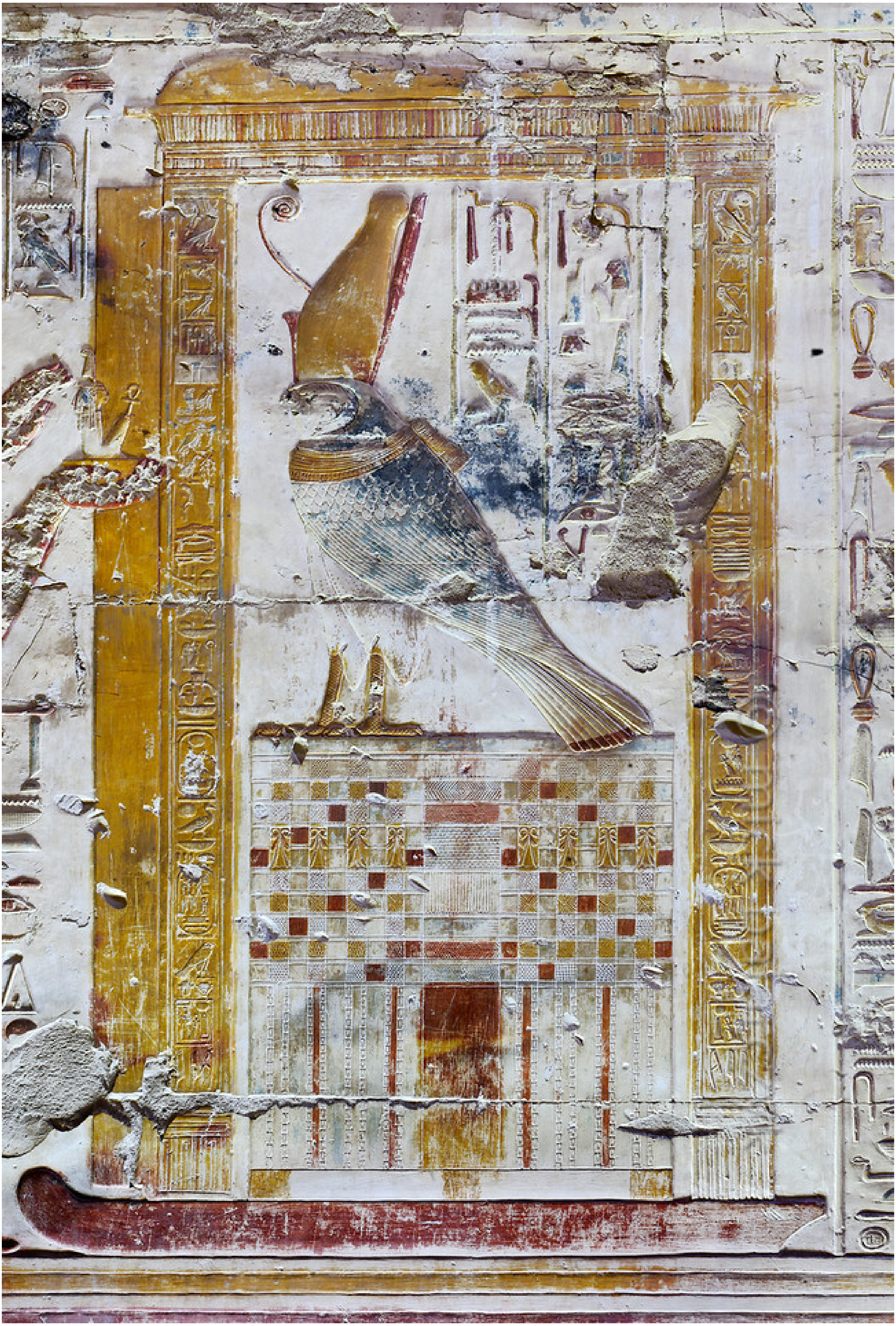

NATIONAL PARK SERVICE
PETRIFIED FOREST
NATIONAL PARK
UNITED STATES DEPARTMENT OF THE INTERIOR
NATIONAL PARK SERVICE

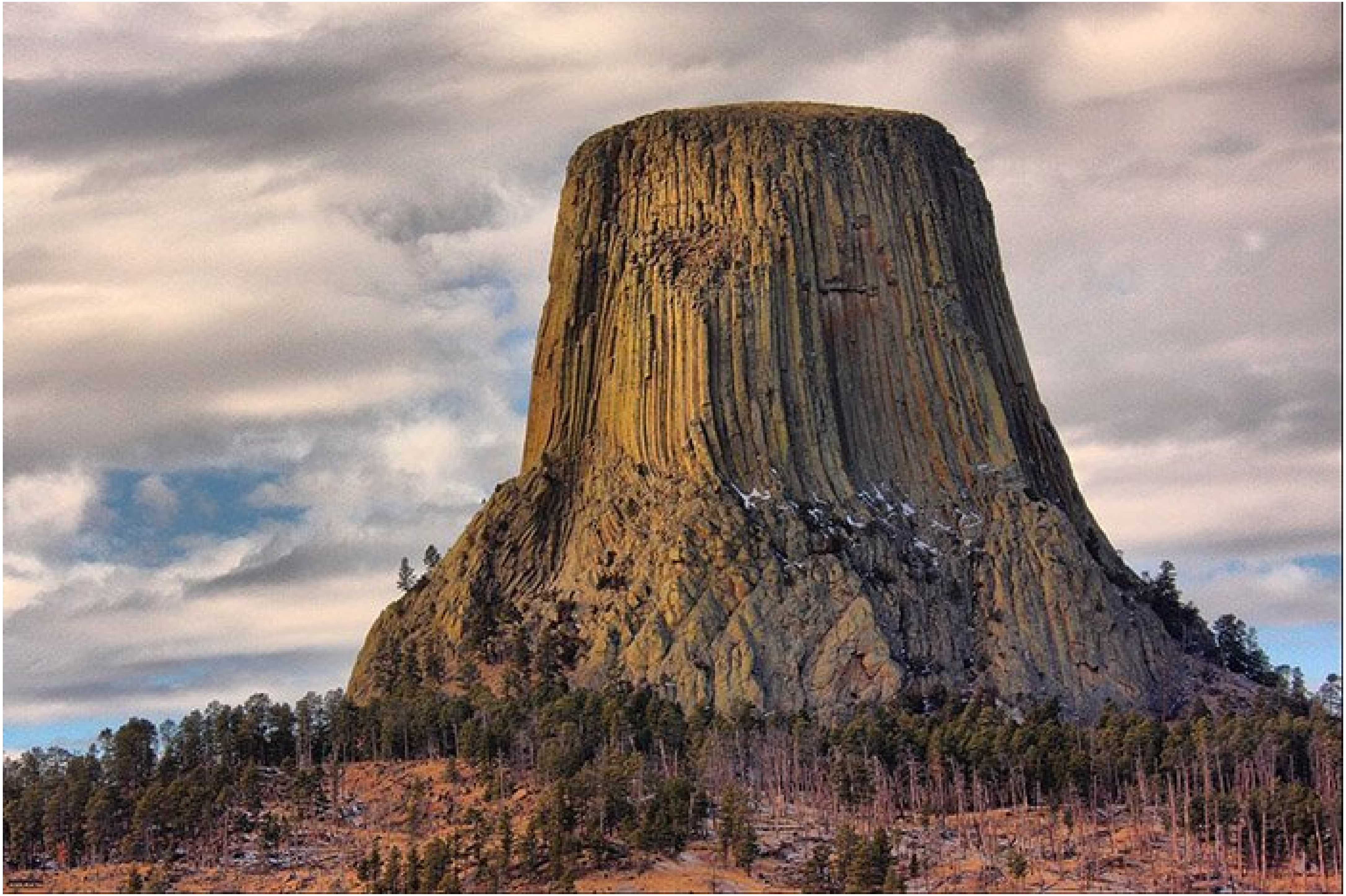

aesthetika

theartshoppauls

Double Protection
Crystal Quantum Radio Amulet
from Brane-Power.com

aesthetika

"True North"
Crystal Quantum Radio Amulet from Brane-Power

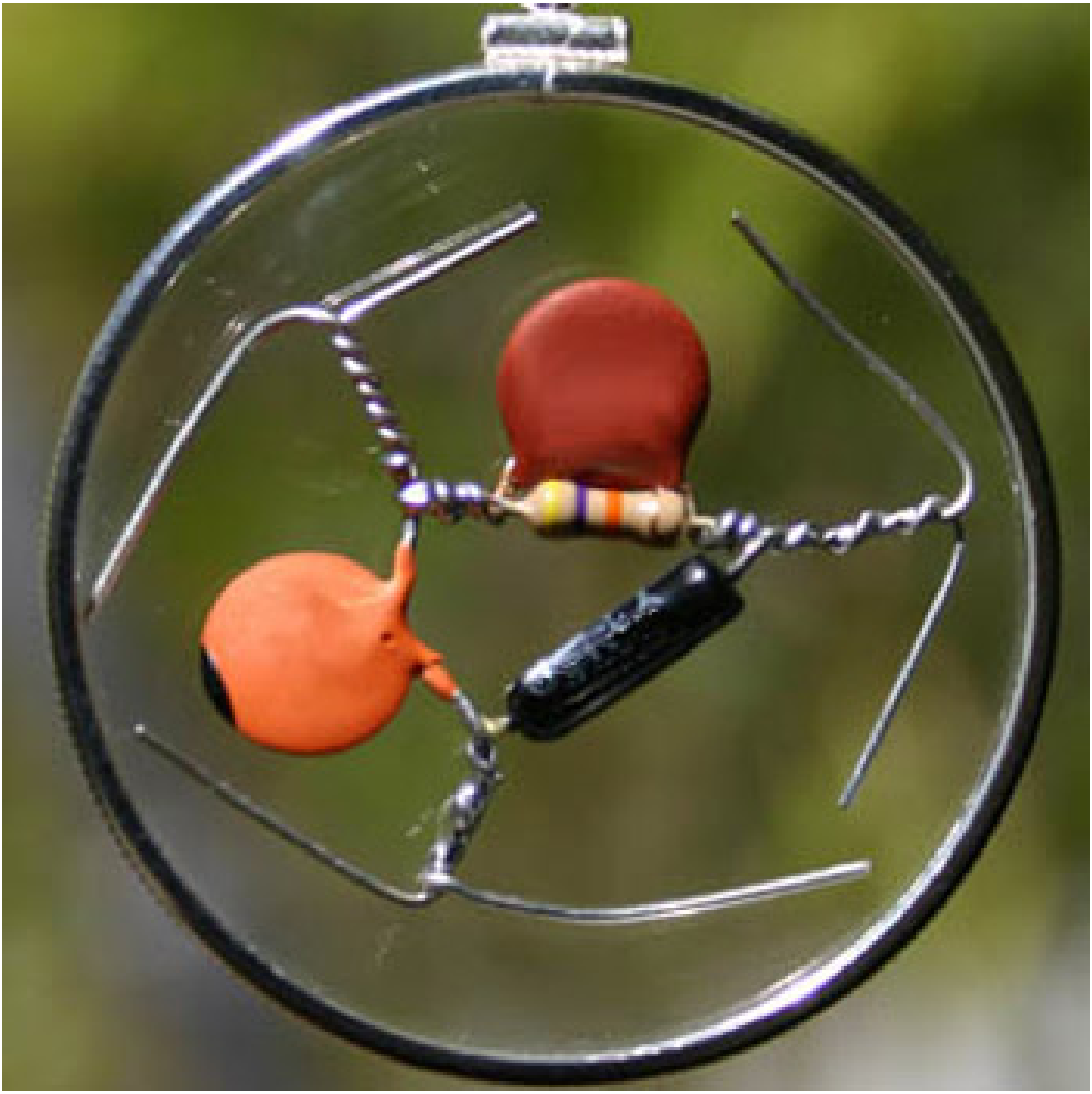

InsaneInventor

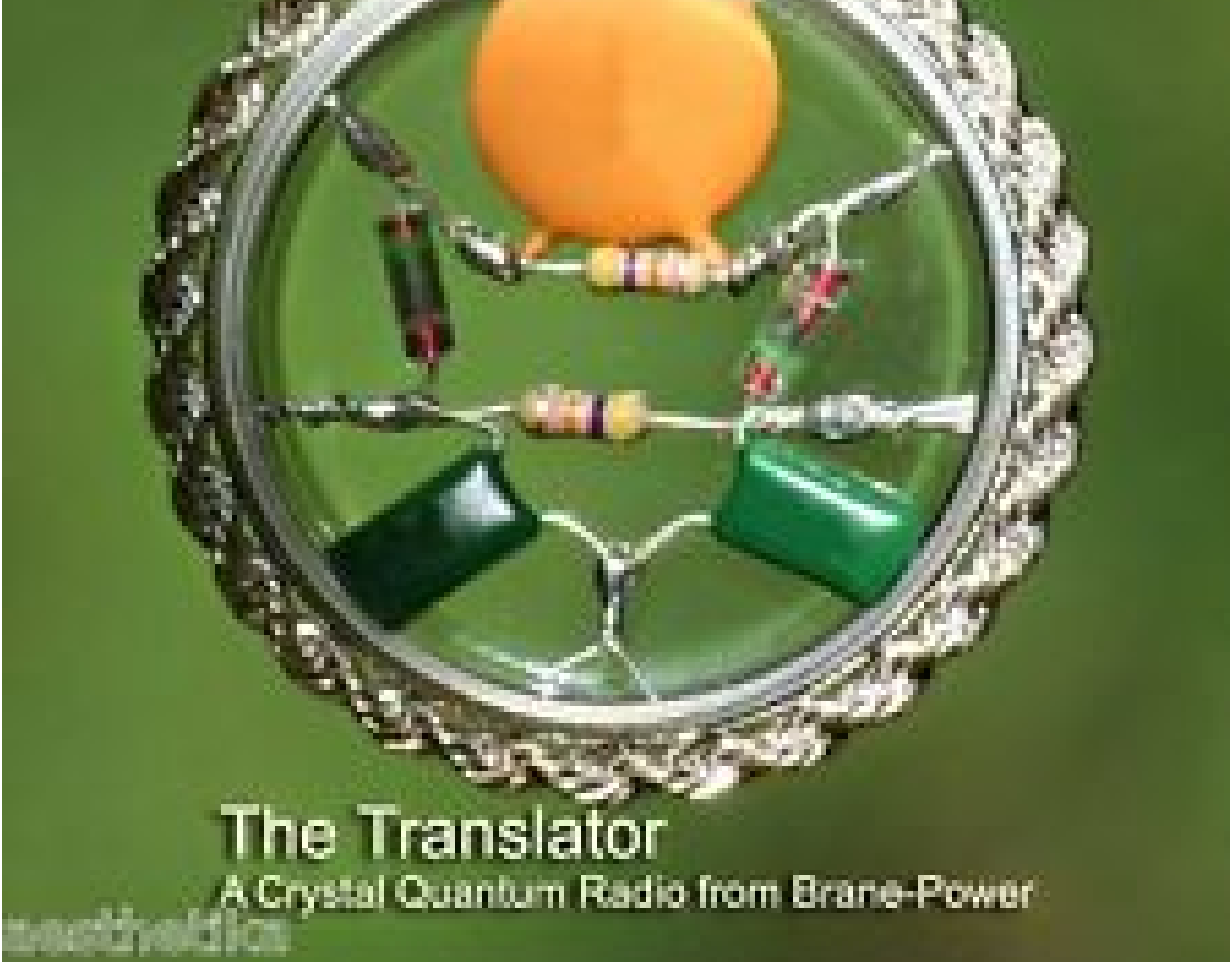

The Translator

A Crystal Quantum Radio from Brane-Power

theartshoppauls

E
C

IronAmmonite.com

SULFATO DE CALCIO

Schematic of the RF-system

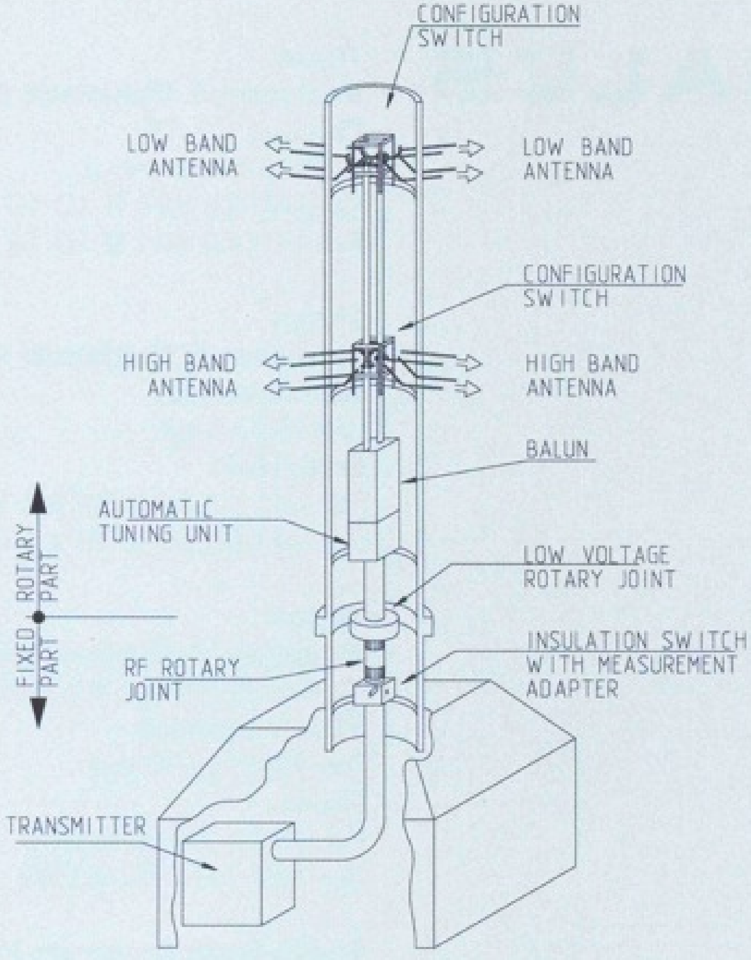
CONFIGURATION SWITCH
LOW BAND ANTENNA
LOW BAND ANTENNA
CONFIGURATION SWITCH
HIGH BAND ANTENNA
HIGH BAND ANTENNA
BALUN
AUTOMATIC TUNING UNIT
ROTARY PART
FIXED PART
LOW VOLTAGE ROTARY JOINT
INSULATION SWITCH WITH MEASUREMENT ADAPTER
RF ROTARY JOINT
TRANSMITTER

PRINCIPAL LAYOUT OF INDOOR RF SYSTEM

Design

- Full in house engineering capabilities
- Latest design tools
- Rigid tubular steel dipoles
- No suspension/counterweight system
- Overall DC grounding of antenna and screen
- High-Band/Low-Band array
- 30 years of experience

Performance

- High reliability
- High availability
- High flexibility

Maintenance

- Easy access to components
- No lowering of dipoles and screen

Cost efficiency

- Low life cycle cost

More than 30 years of experience
30 years of antennas on air

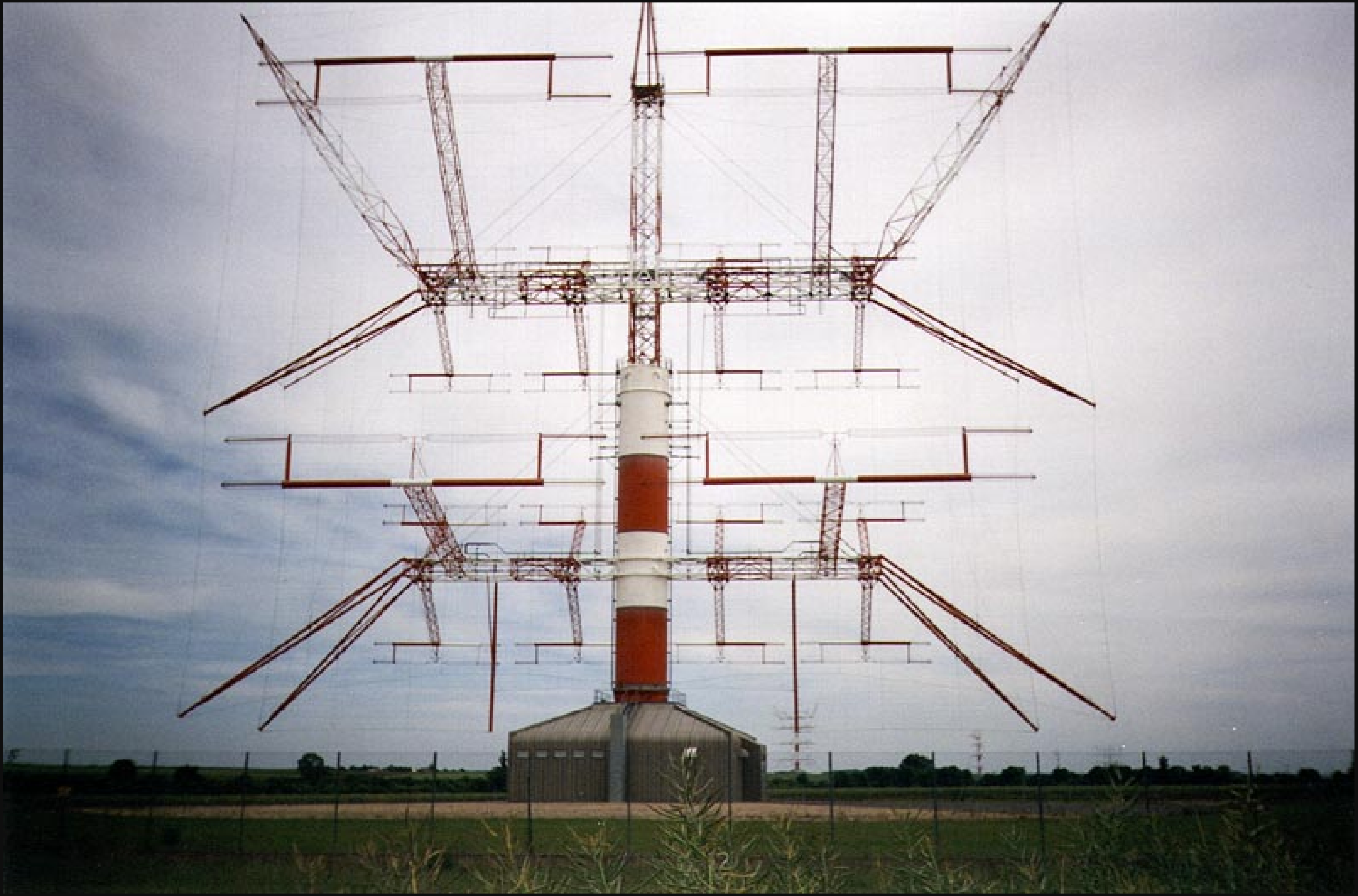

COMPARISON TO CURRENT TITLEHOLDER

ARECIBO

Since it was constructed in 1963 in Arecibo, Puerto Rico, this telescope had held the record as the world's largest telescope.

FAST

It's designed as a cable-net supporting structure, capable of forming a parabolic mirror and will be completed by 2016.

SIZE

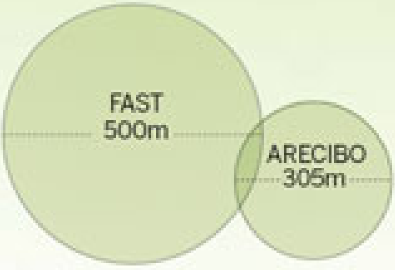

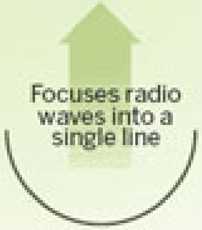

ARECIBO

Arecibo has a fixed spherical curvature.

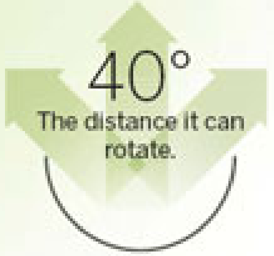

FAST

FAST can reshape into a paraboloidal surface.

ASTRONOMICAL GARDEN

URUMQI NANSHAN
25m
Radio telescope
Built in 1994 in Urumqi, Xinjiang.
Missions: Pulsar research. Determines orbits for the Chang'e lunar satellite and other lunar projects.

QINGHAI DELINGHA
13.7m Millimeter-wave radio telescope
Built in 1990 in Delingha, Qinghai.
Mission: Interstellar molecular spectroscopic research.

MIYUN
50m
Radio telescope
Built in 2005 in Miyun county, Beijing.
Mission: Long-term monitoring of a group of known millisecond pulsars.

MIYUN
Synthesis radio telescope
Built in 1983 in Miyun county, Beijing.
Missions: Northern sky survey and observation of Supernova Remnant and interplanetary scintillation.

KUNMING
40m
Radio telescope
Built in 2006 in Kunming, Yunnan province.
Mission: Successfully completes the VLBI tracking of China's first lunar probe Chang'e-1 with other Chinese telescopes.

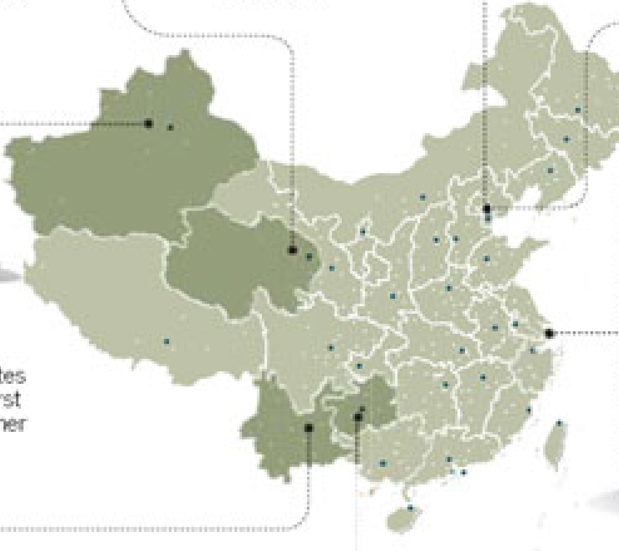

SHANGHAI SHESHAN
65m
Radio telescope (under construction)
Built by 2012 in Sheshan district, Shanghai.
Missions: Finds applications in deep space exploration and basic astronomy research and assists China's lunar exploration program.

SHANGHAI SHESHAN
25m
Radio telescope
Built in 1986 in Sheshan district, Shanghai.
Mission: Assists the VLBI (Very Long Baseline Interferometry) network.

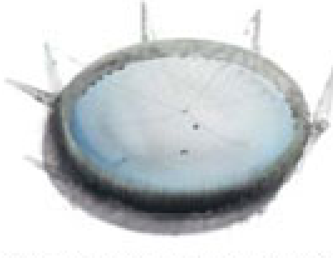

GUIZHOU PINGTANG
500m
Aperture spherical radio telescope (under construction)
Built by 2016 in Pingtang county, Guizhou province.
Missions: Surveys neutral hydrogen in the Milky Way and other galaxies, detects faint pulsars, looks for shining stars and listens for possible signals from other civilizations.

Source: Skyandtelescope.com, universetoday.com, URSI

GUILLERMO MUNRO / CHINA DAILY

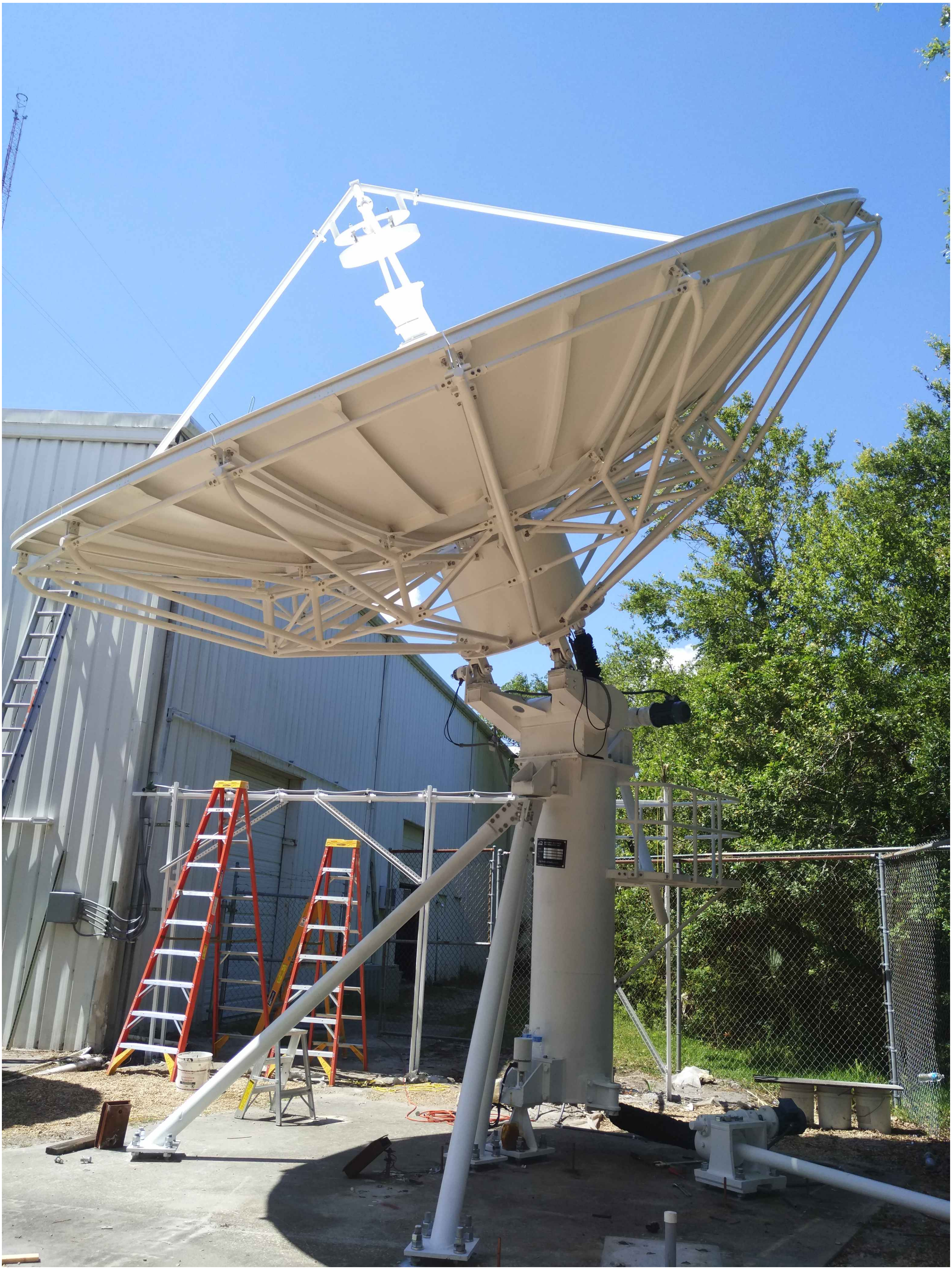

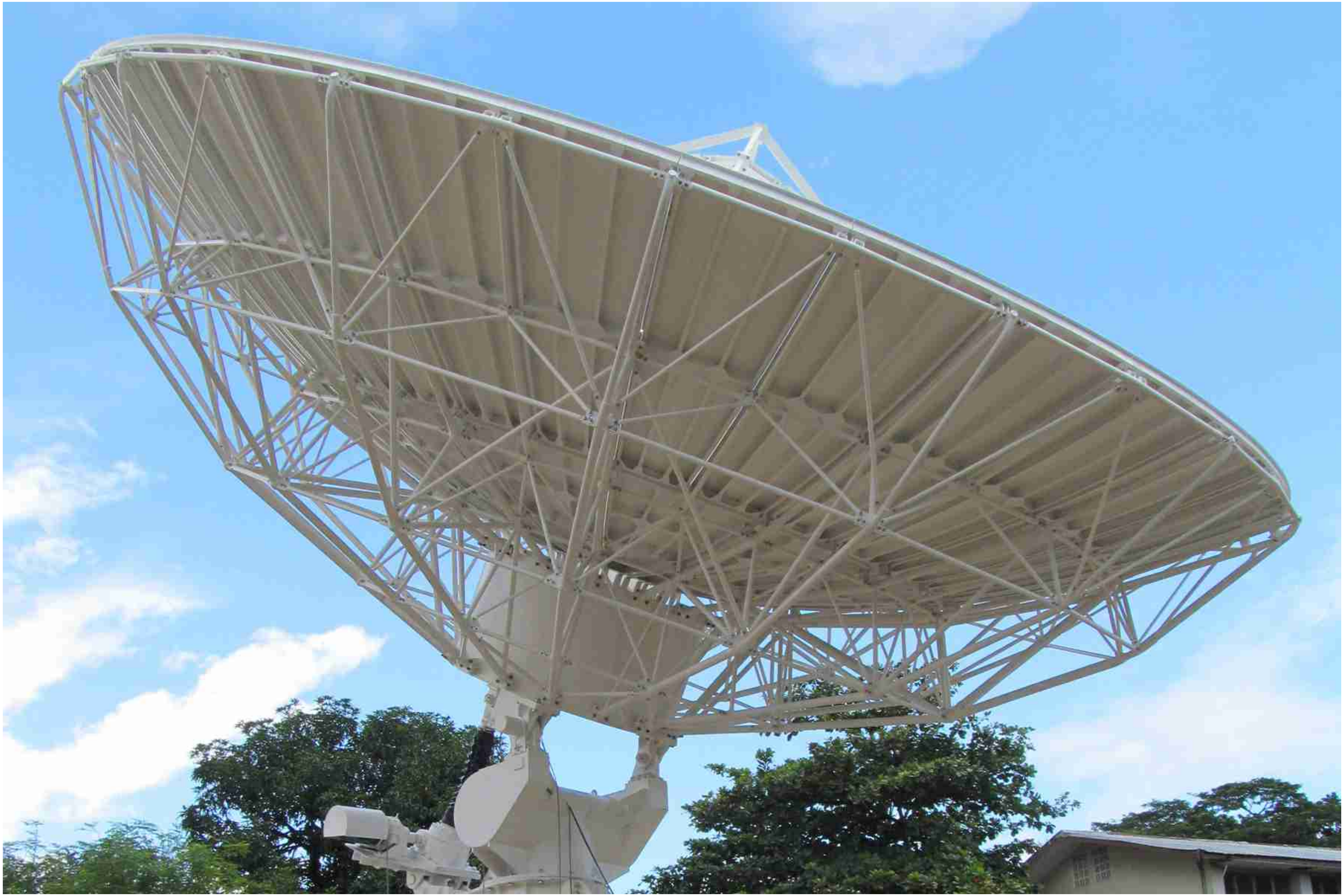

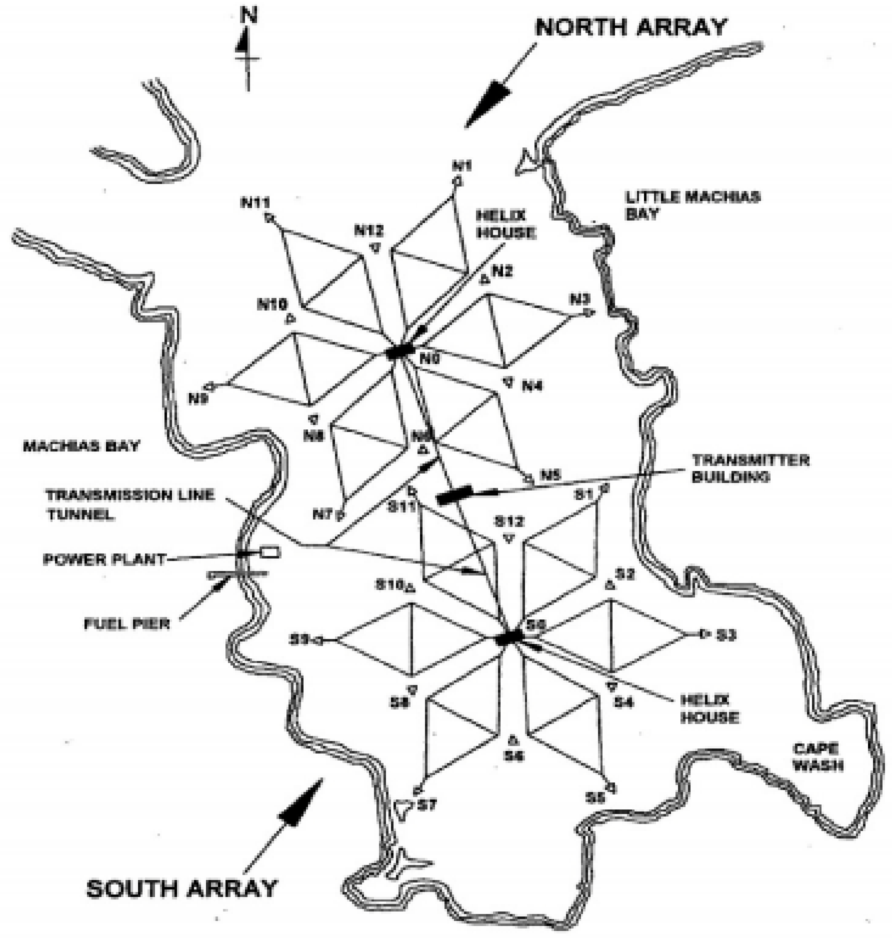
N
NORTH ARRAY
N1
N11
N12
HELIX HOUSE
LITTLE MACHIAS BAY
N2
N3
N10
N0
N9
N4
N8
N6
MACHIAS BAY
N5
TRANSMITTER BUILDING
TRANSMISSION LINE TUNNEL
N7
S11
S1
S12
POWER PLANT
S2
S10
FUEL PIER
S9
S0
S3
S8
S4
HELIX HOUSE
S6
CAPE WASH
S7
S5
SOUTH ARRAY

Figure 1. General layout, Cutler Peninsula.

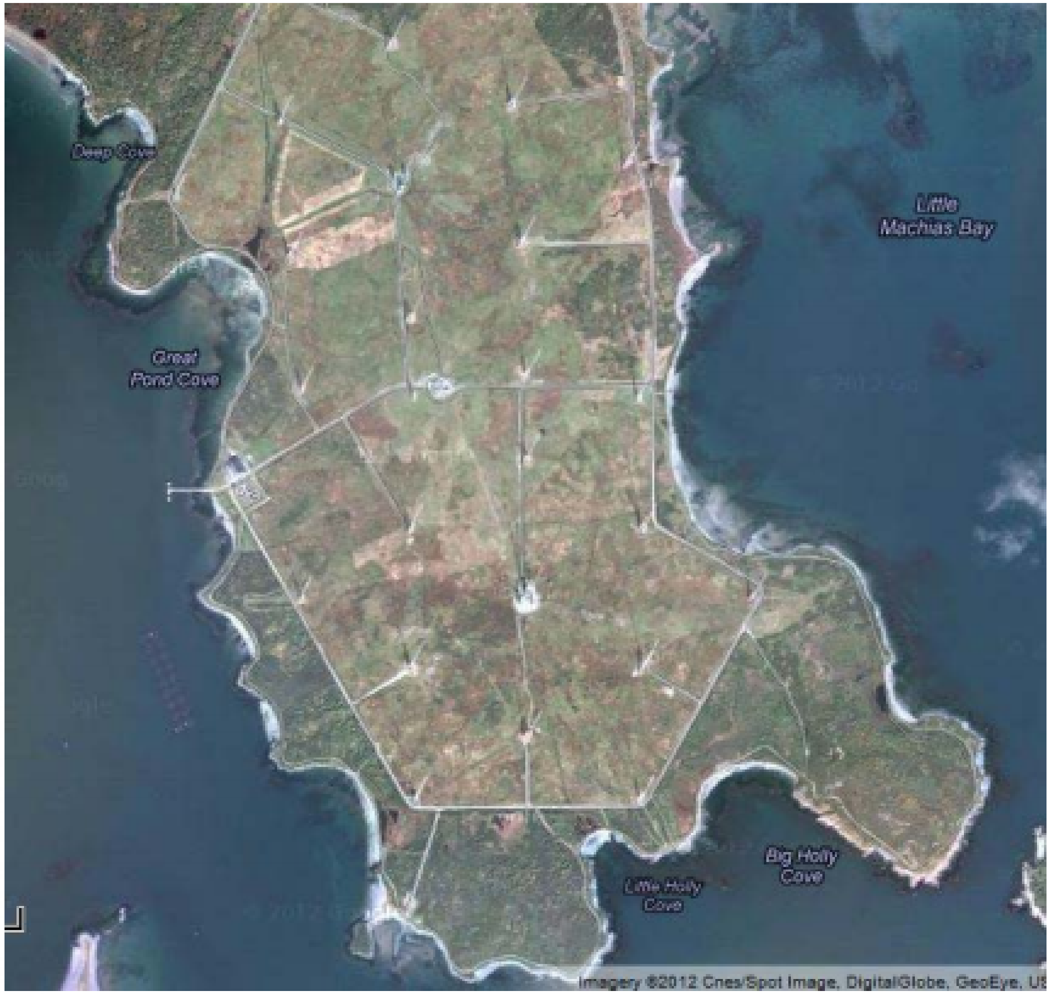
Deep Cove
Little Machias Bay
Great Pond Cove
Big Holly Cove
Little Holly Cove
Imagery ©2012 Cnes/Spot Image, DigitalGlobe, GeoEye, US

26 TOWERS- 850 to 1000 FT HIGH

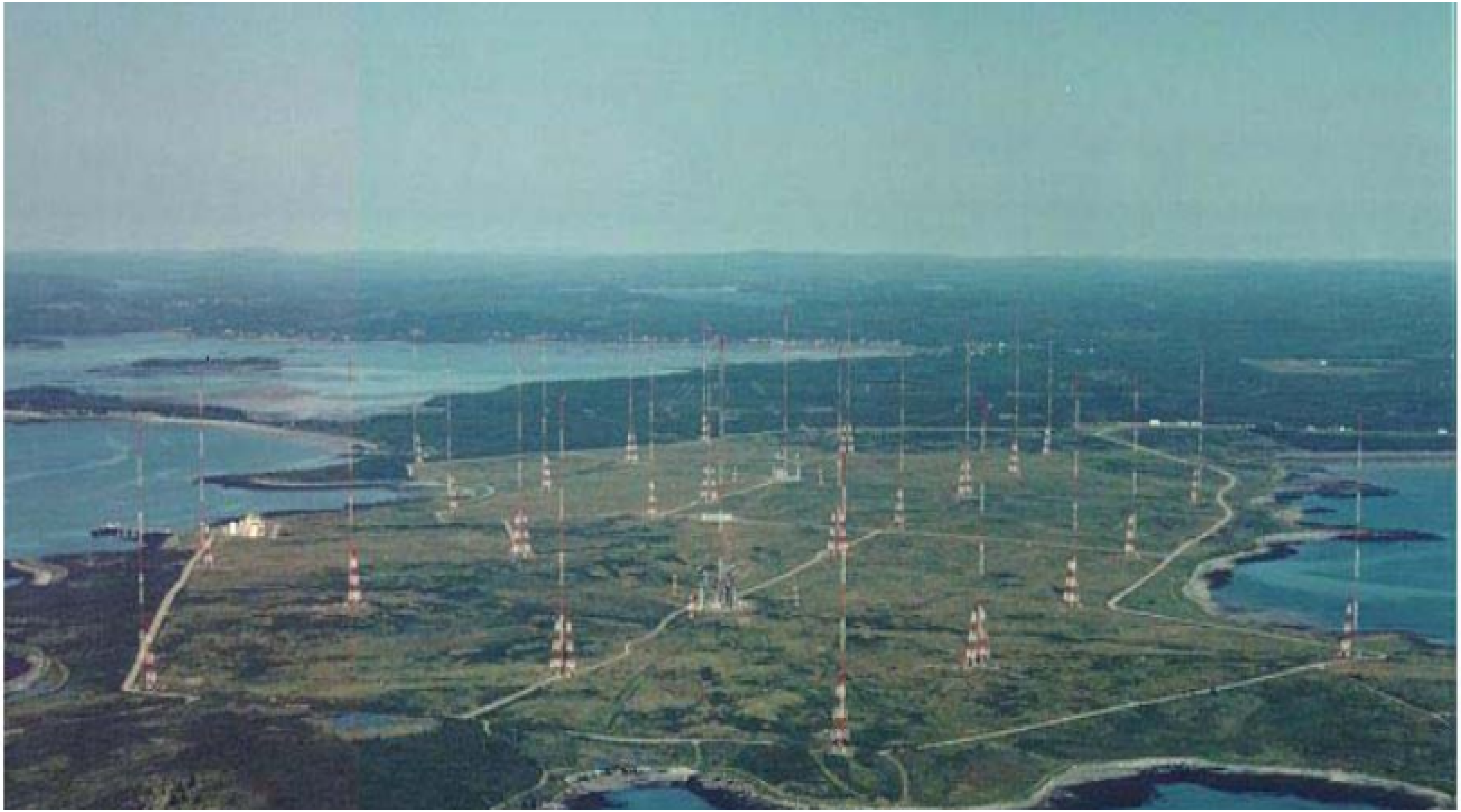

EACH INSULATOR IS 57 FT LONG TO WITHSTAND 250 KV

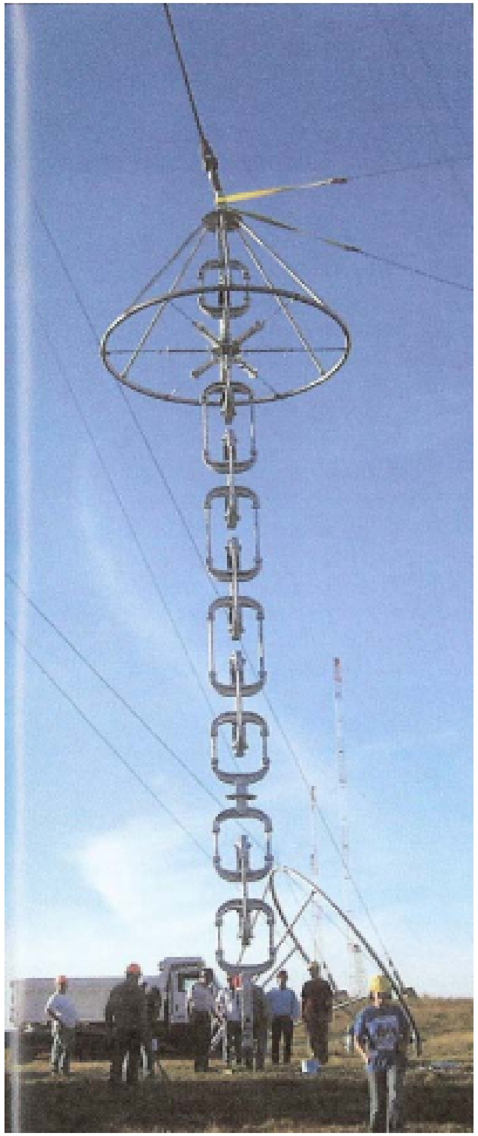

13,000 lbs.

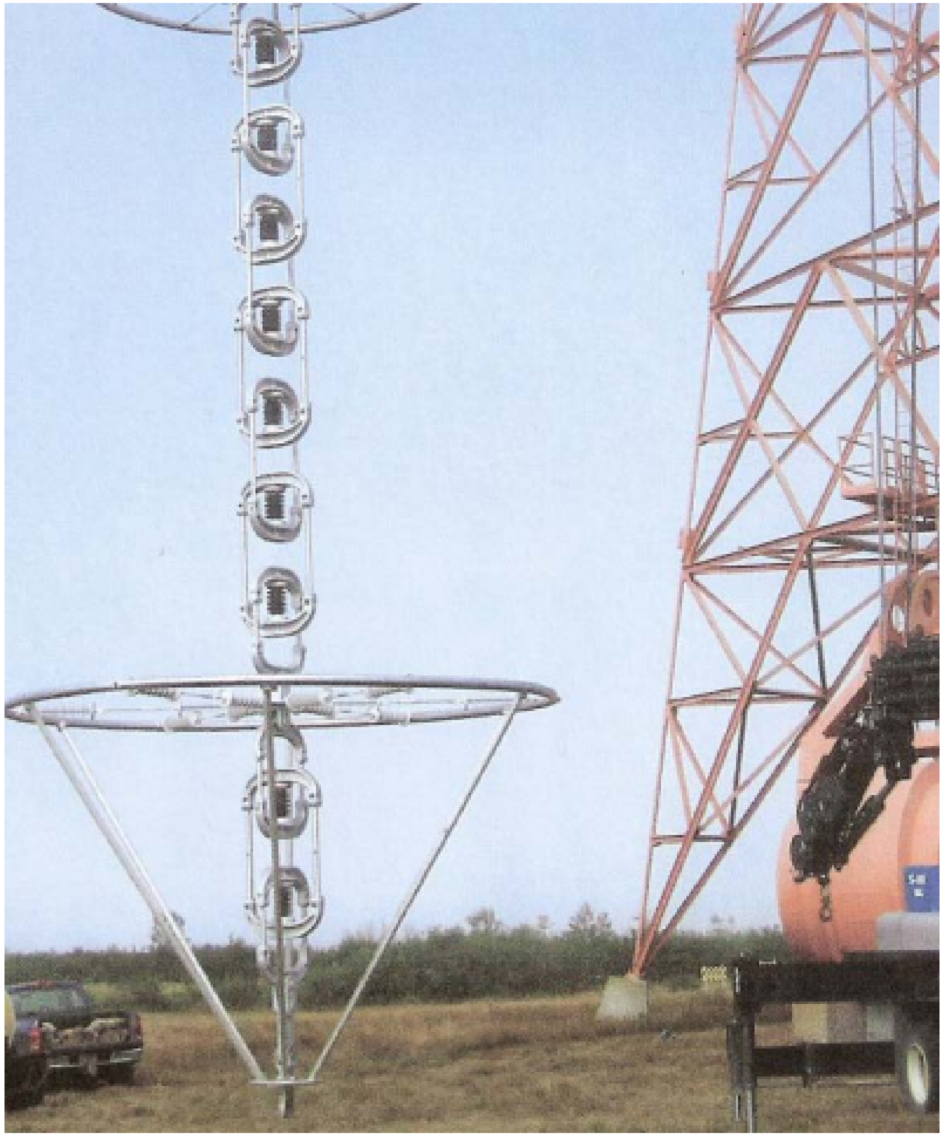

TUNING NETWORK- HELIX

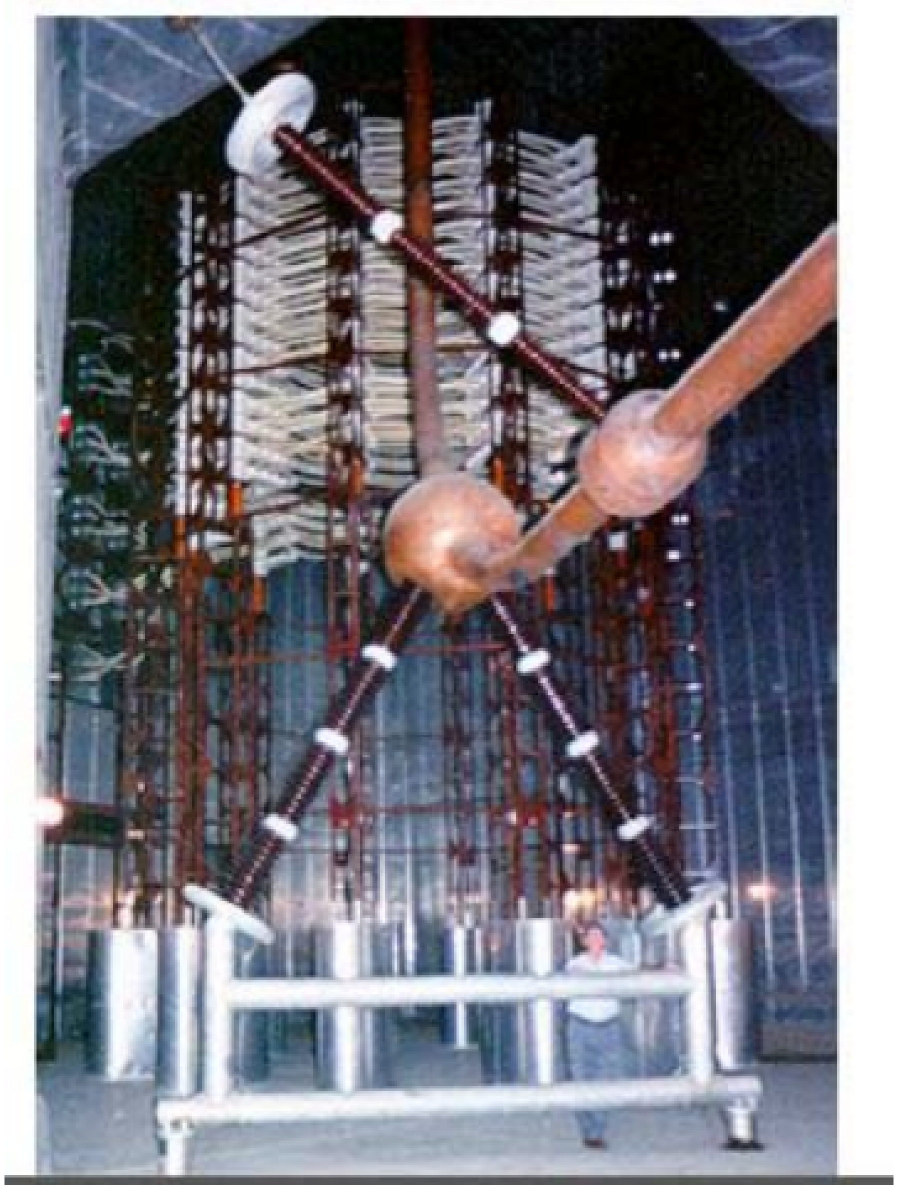

Ref 8

TUNING NETWORK-VARIOMETER

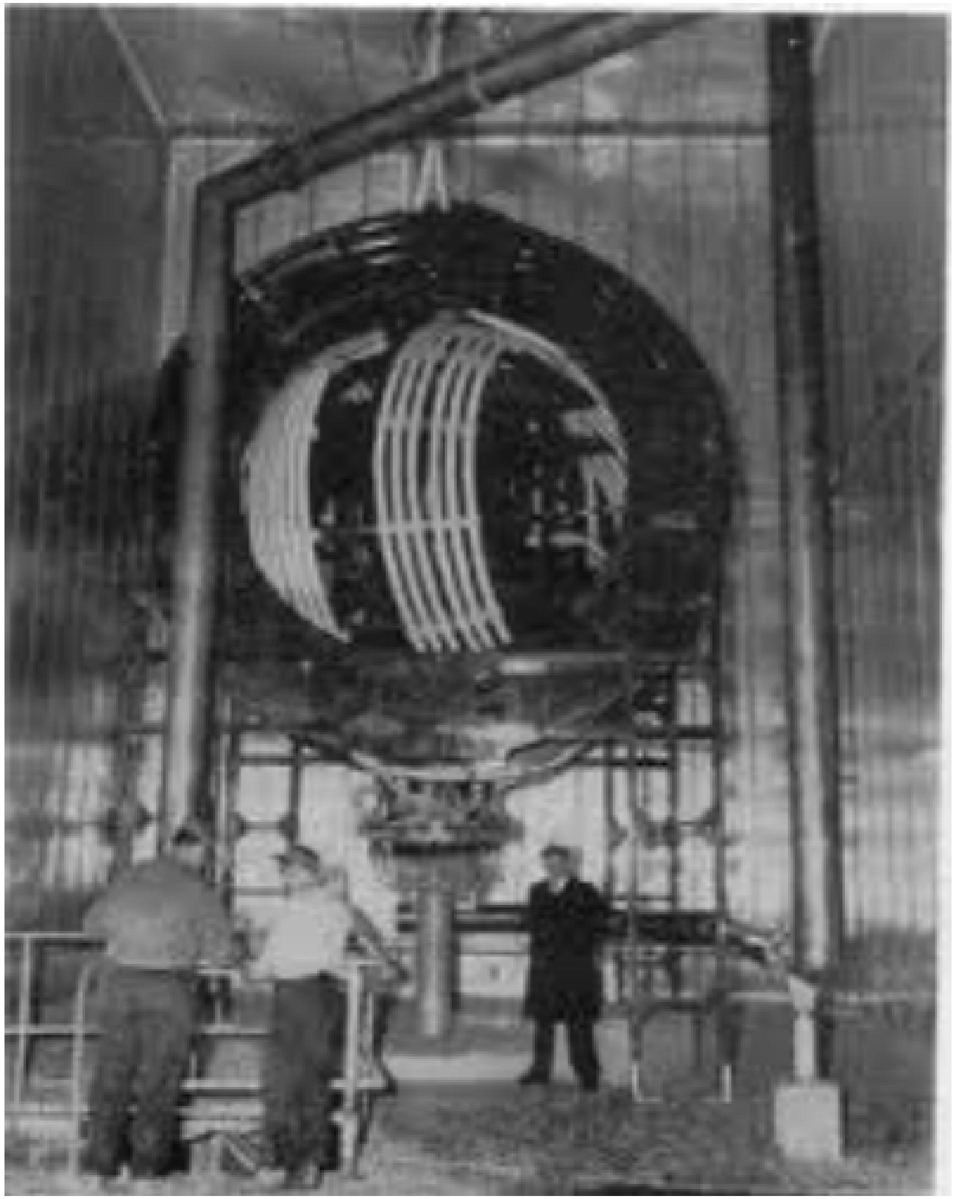

Ref 5 NAA

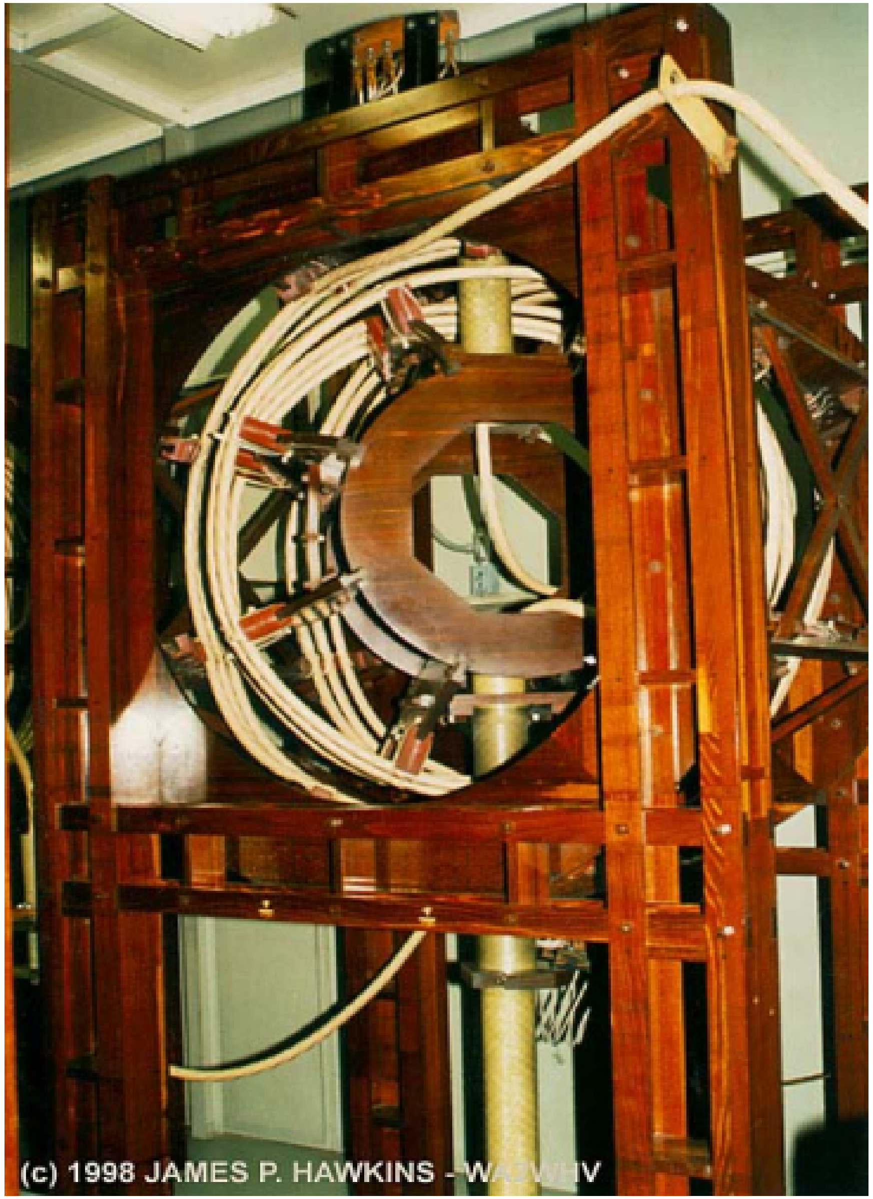

(c) 1998 JAMES P. HAWKINS - WA[illegible]HV

Wires are 4 inches diameter

NSS

JP Hawkins

SUBMARINE RADIO RECEIVERS

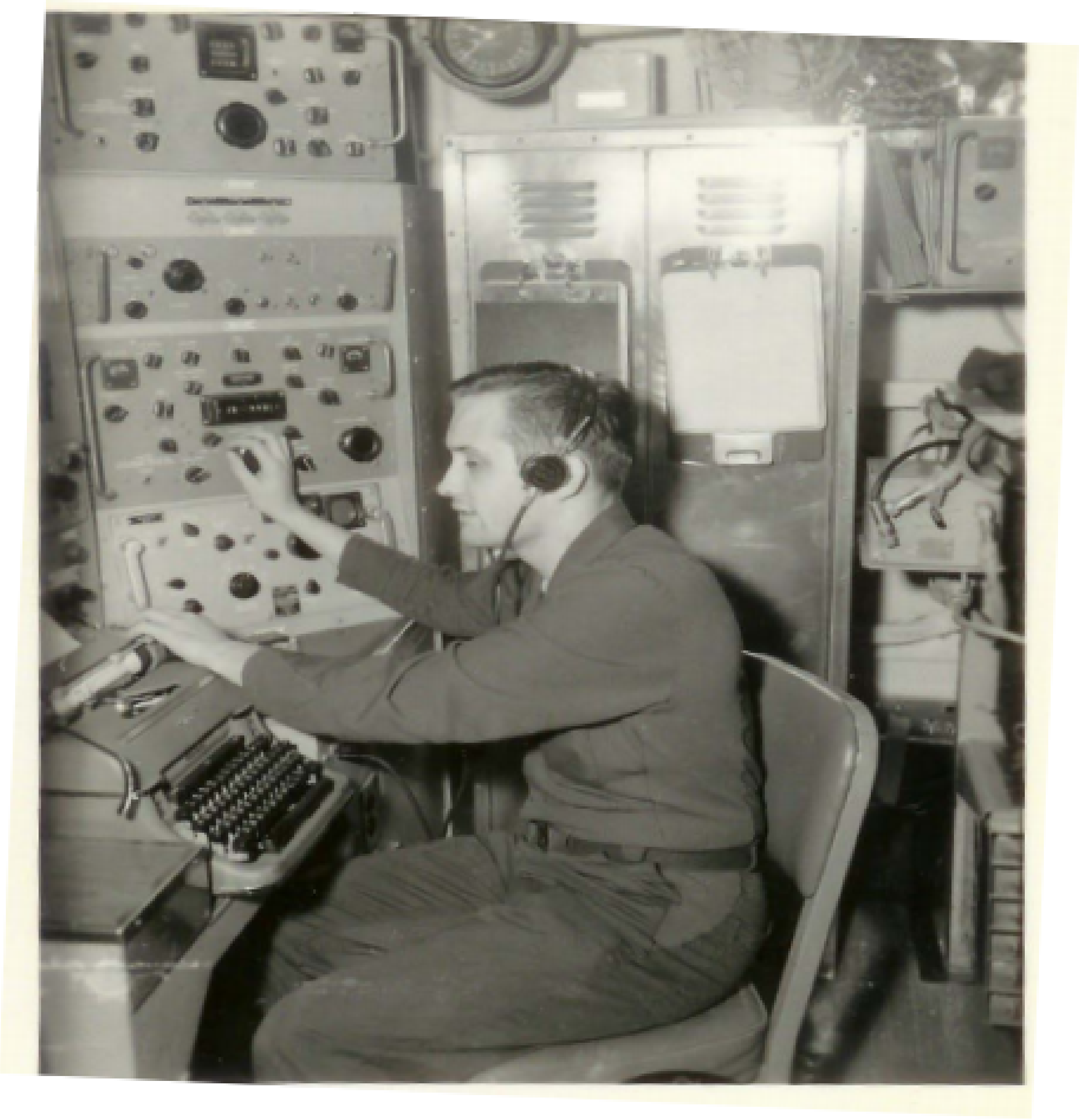

USS Robert E Lee 1966

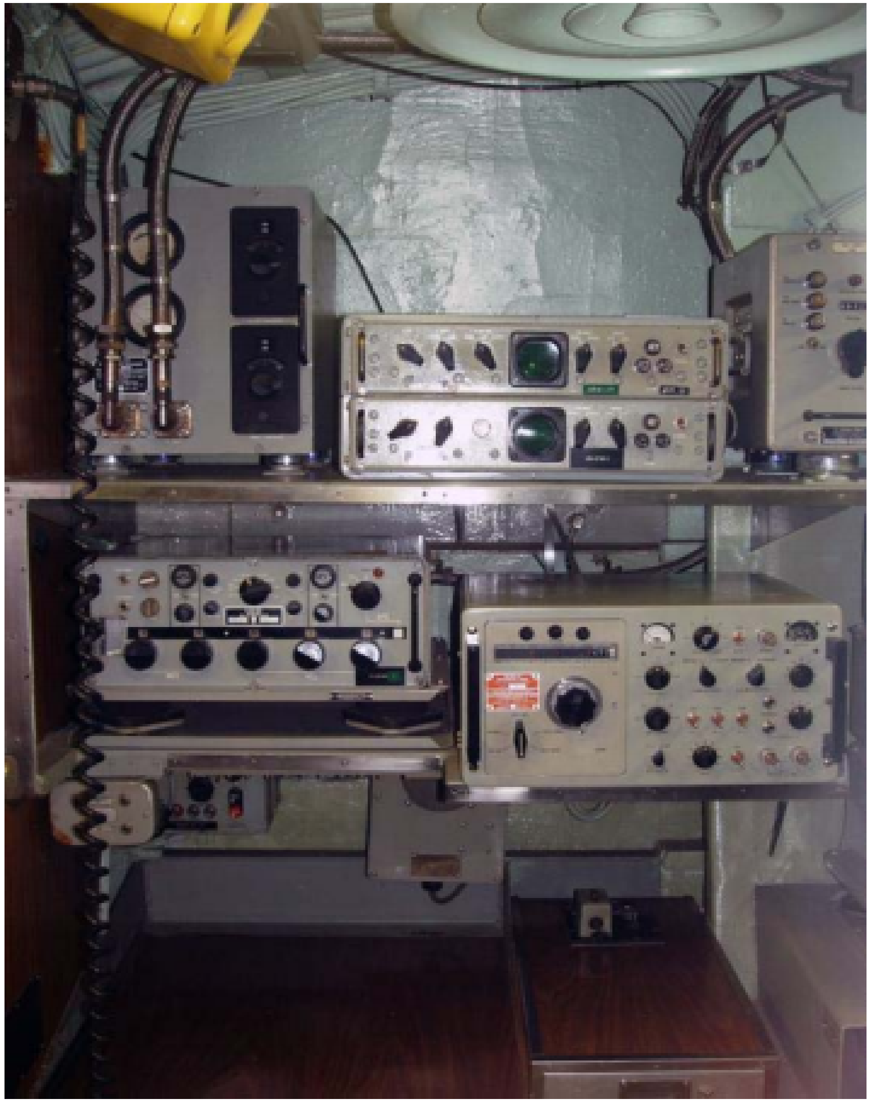

USS Nautilus 1970s

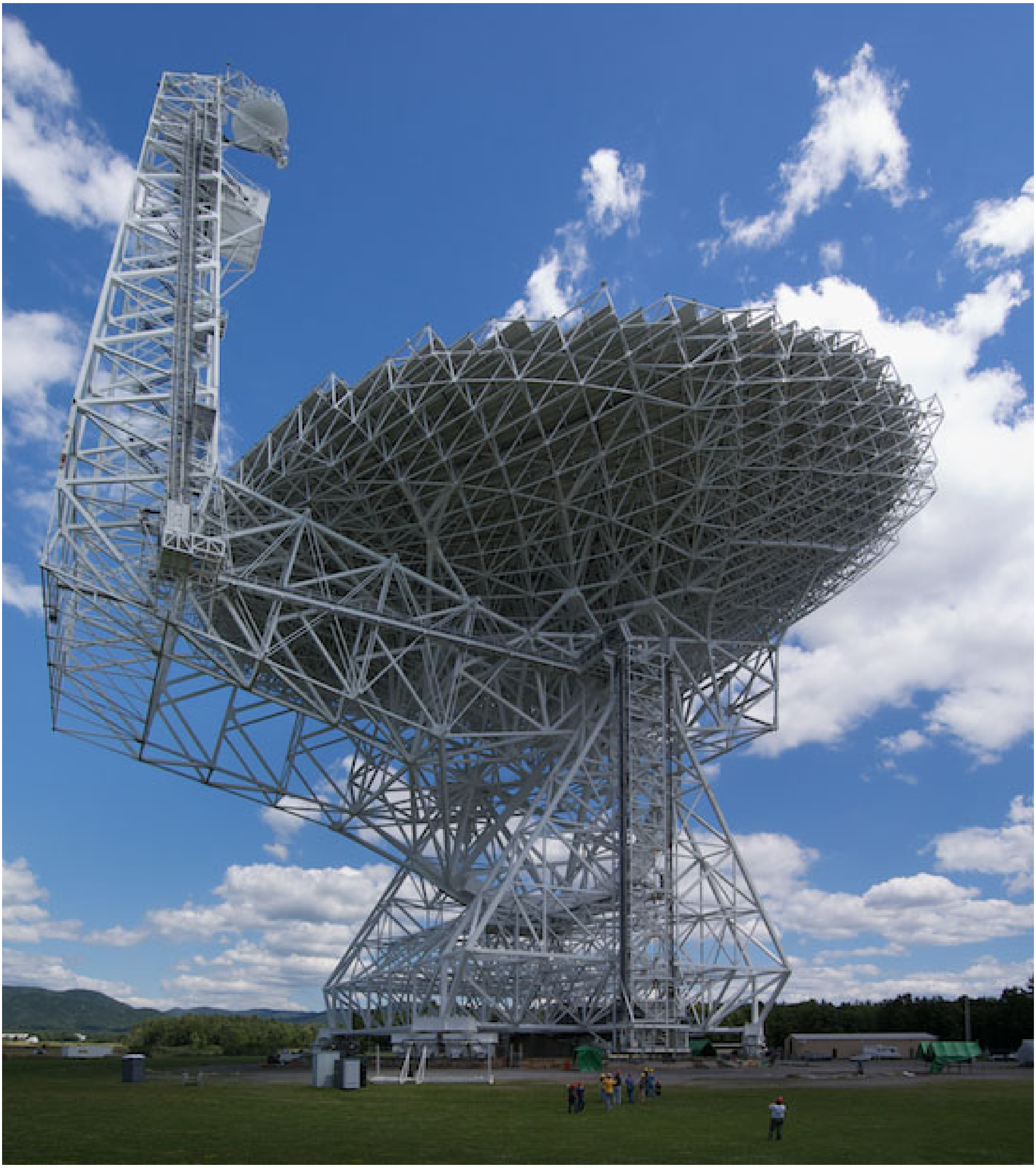

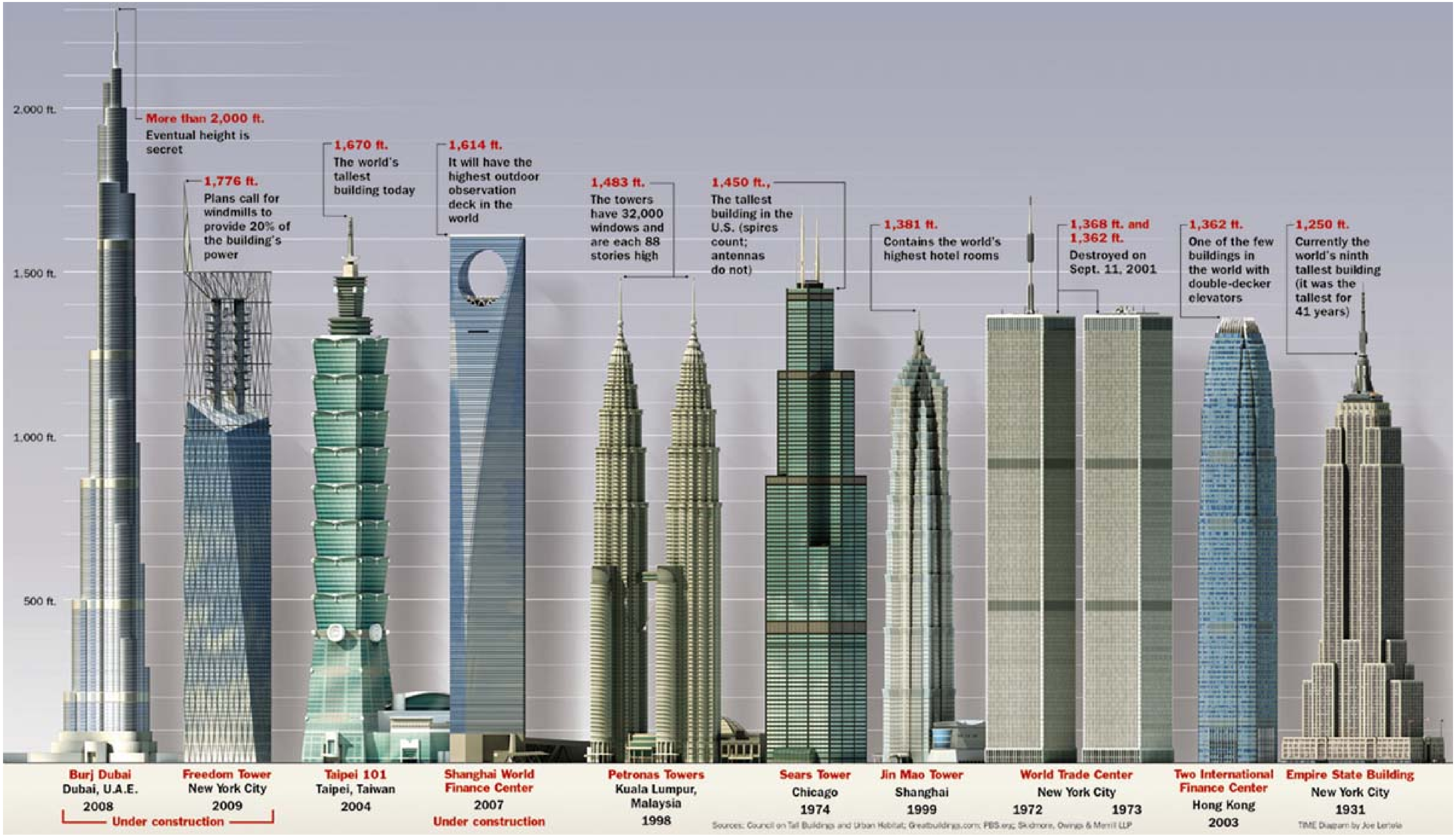

More than 2,000 ft. Eventual height is secret

1,776 ft. Plans call for windmills to provide 20% of the building's power

1,670 ft. The world's tallest building today

1,614 ft. It will have the highest outdoor observation deck in the world

1,483 ft. The towers have 32,000 windows and are each 88 stories high

1,450 ft., The tallest building in the U.S. (spires count; antennas do not)

1,381 ft. Contains the world's highest hotel rooms

1,368 ft. and 1,362 ft. Destroyed on Sept. 11, 2001

1,362 ft. One of the few buildings in the world with double-decker elevators

1,250 ft. Currently the world's ninth tallest building (it was the tallest for 41 years)

Burj Dubai
Dubai, U.A.E.
2008

Freedom Tower
New York City
2009

Under construction

Taipei 101
Taipei, Taiwan
2004

Shanghai World Finance Center
2007
Under construction

Petronas Towers
Kuala Lumpur, Malaysia
1998

Sears Tower
Chicago
1974

Jin Mao Tower
Shanghai
1999

World Trade Center
New York City
1972 1973

Two International Finance Center
Hong Kong
2003

Empire State Building
New York City
1931

Sources: Council on Tall Buildings and Urban Habitat; Greatbuildings.com; PBS.org; Skidmore, Owings & Merrill LLP

TIME Diagram by Joe Lertola

WEATHER MODIFICATION TECHNOLOGIES 2018

Image by Jim Lee, **ClimateViewer.com**

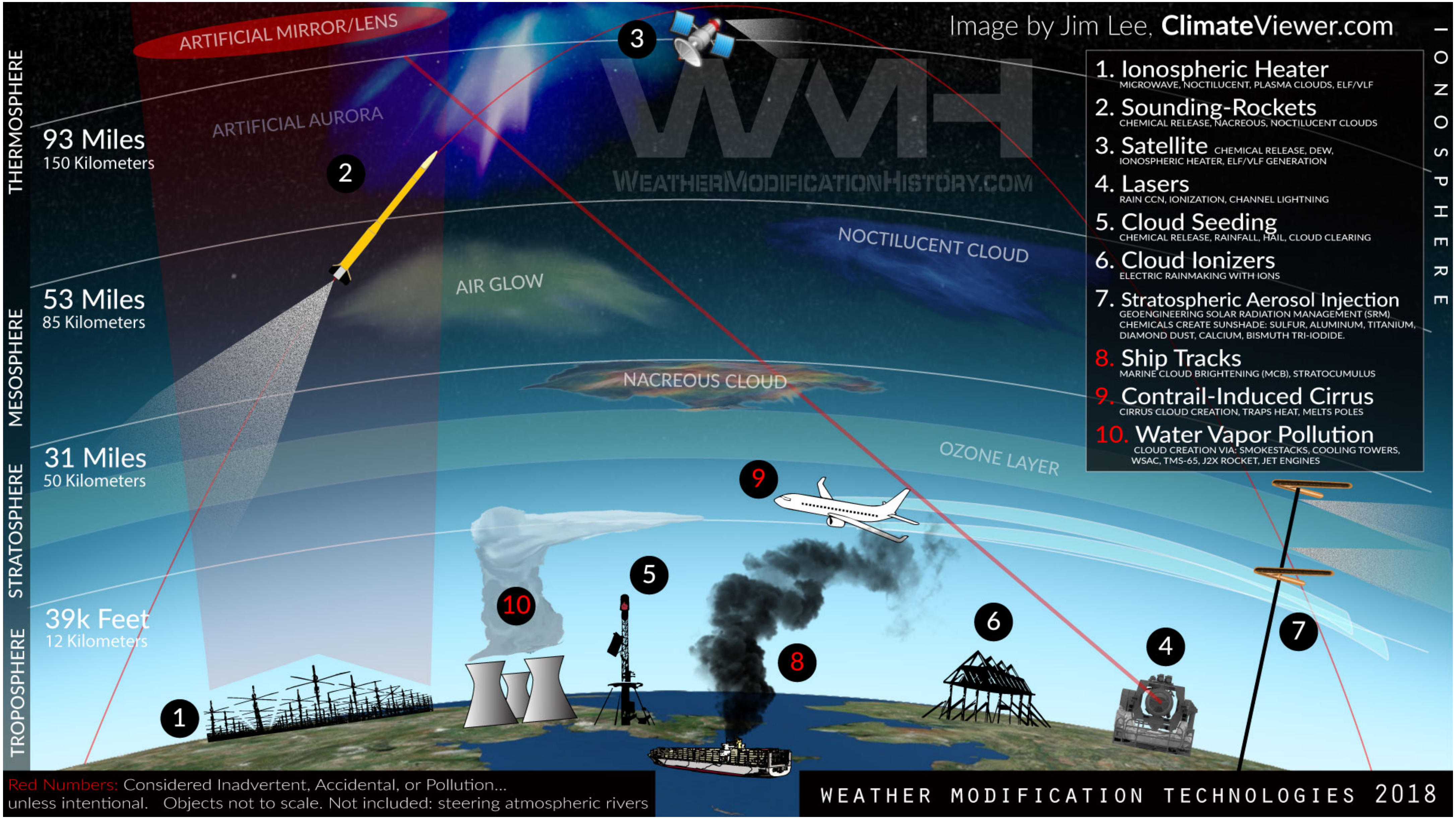
IONOSPHERE
THERMOSPHERE
MESOSPHERE
STRATOSPHERE
TROPOSPHERE
93 Miles
150 Kilometers
53 Miles
85 Kilometers
31 Miles
50 Kilometers
39k Feet
12 Kilometers
ARTIFICIAL MIRROR/LENS
ARTIFICIAL AURORA
AIR GLOW
NOCTILUCENT CLOUD
NACREOUS CLOUD
OZONE LAYER
WMH
WEATHERMODIFICATIONHISTORY.COM
1
2
3
4
5
6
7
8
9
10

1. **Ionospheric Heater** MICROWAVE, NOCTILUCENT, PLASMA CLOUDS, ELF/VLF
2. **Sounding-Rockets** CHEMICAL RELEASE, NACREOUS, NOCTILUCENT CLOUDS
3. **Satellite** CHEMICAL RELEASE, DEW, IONOSPHERIC HEATER, ELF/VLF GENERATION
4. **Lasers** RAIN CCN, IONIZATION, CHANNEL LIGHTNING
5. **Cloud Seeding** CHEMICAL RELEASE, RAINFALL, HAIL, CLOUD CLEARING
6. **Cloud Ionizers** ELECTRIC RAINMAKING WITH IONS
7. **Stratospheric Aerosol Injection** GEOENGINEERING SOLAR RADIATION MANAGEMENT (SRM) CHEMICALS CREATE SUNSHADE: SULFUR, ALUMINUM, TITANIUM, DIAMOND DUST, CALCIUM, BISMUTH TRI-IODIDE.
8. **Ship Tracks** MARINE CLOUD BRIGHTENING (MCB), STRATOCUMULUS
9. **Contrail-Induced Cirrus** CIRRUS CLOUD CREATION, TRAPS HEAT, MELTS POLES
10. **Water Vapor Pollution** CLOUD CREATION VIA: SMOKESTACKS, COOLING TOWERS, WSAC, TMS-65, J2X ROCKET, JET ENGINES

Red Numbers: Considered Inadvertent, Accidental, or Pollution... unless intentional. Objects not to scale. Not included: steering atmospheric rivers

HAARP

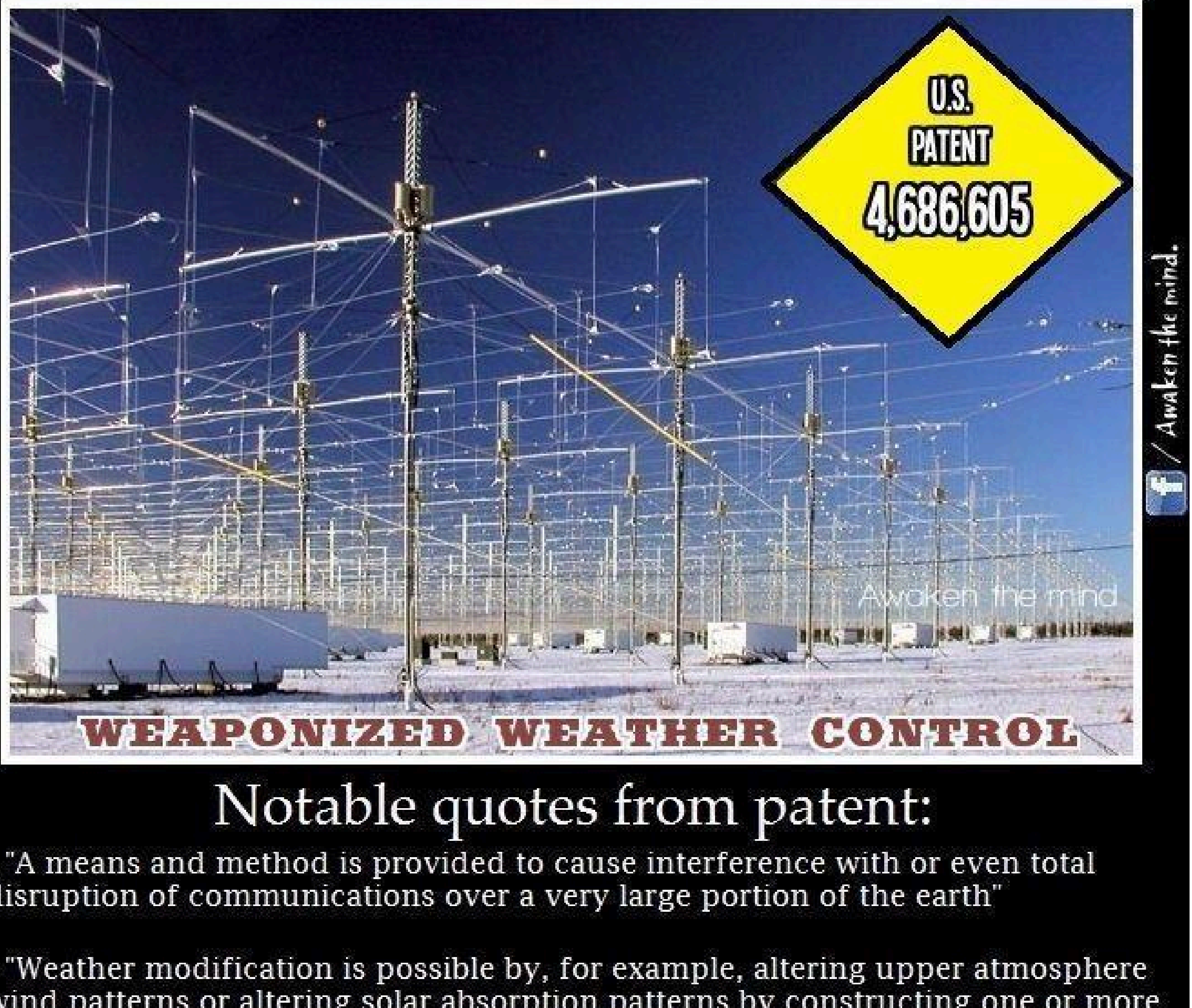
U.S.
PATENT
4,686,605
/ Awaken the mind.
Awaken the mind
WEAPONIZED WEATHER CONTROL

Notable quotes from patent:

- "A means and method is provided to cause interference with or even total disruption of communications over a very large portion of the earth"

- "Weather modification is possible by, for example, altering upper atmosphere wind patterns or altering solar absorption patterns by constructing one or more plumes of atmospheric particles which will act as a lens or focusing device."

- "The earth's magnetic field could be decreased or disrupted at appropriate altitudes to modify or eliminate the magnetic field."

Sowing seeds in clouds

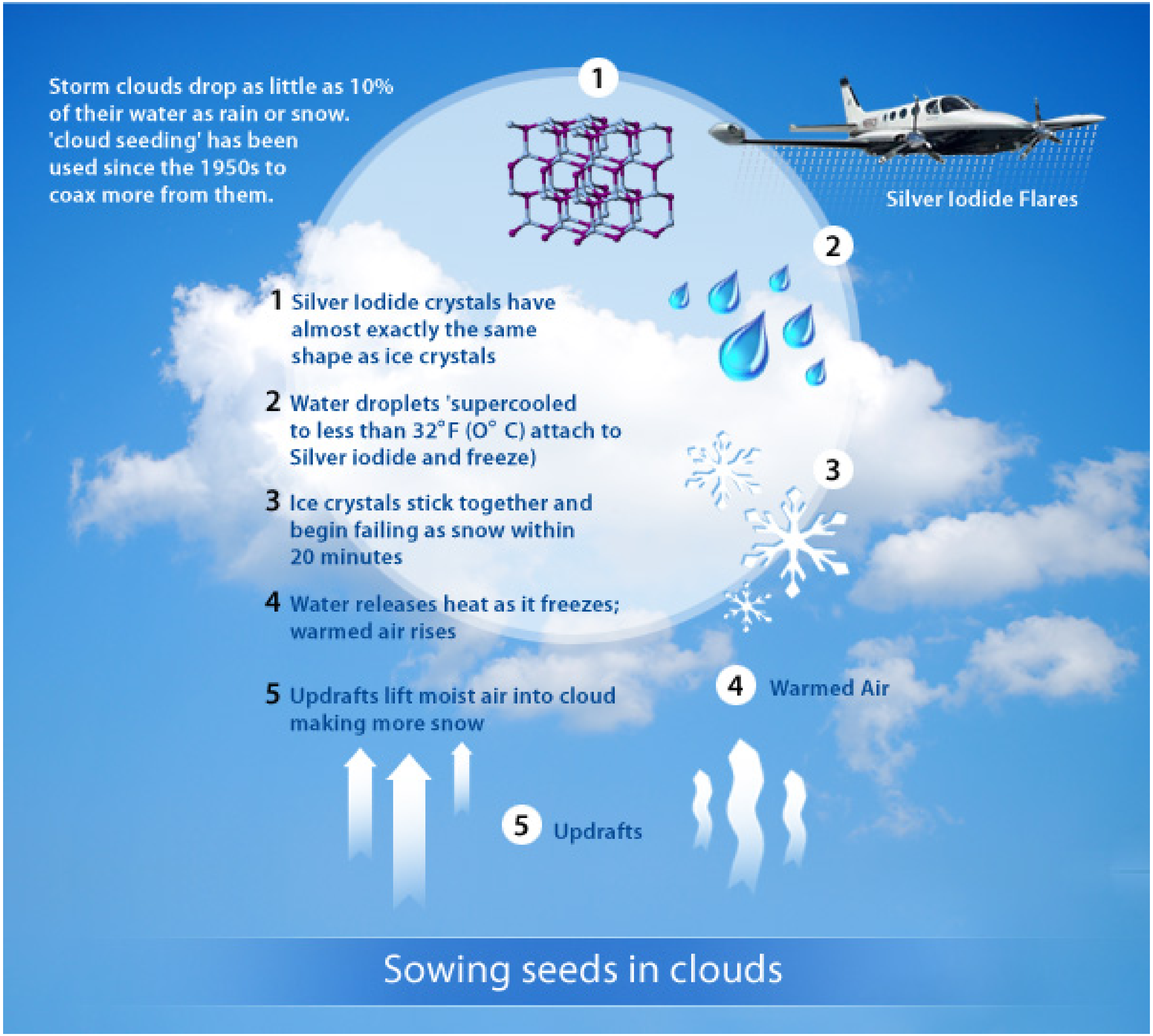

Storm clouds drop as little as 10% of their water as rain or snow. 'cloud seeding' has been used since the 1950s to coax more from them.

1 Silver Iodide crystals have almost exactly the same shape as ice crystals

2 Water droplets 'supercooled to less than 32°F (O° C) attach to Silver iodide and freeze)

3 Ice crystals stick together and begin failing as snow within 20 minutes

4 Water releases heat as it freezes; warmed air rises

5 Updrafts lift moist air into cloud making more snow

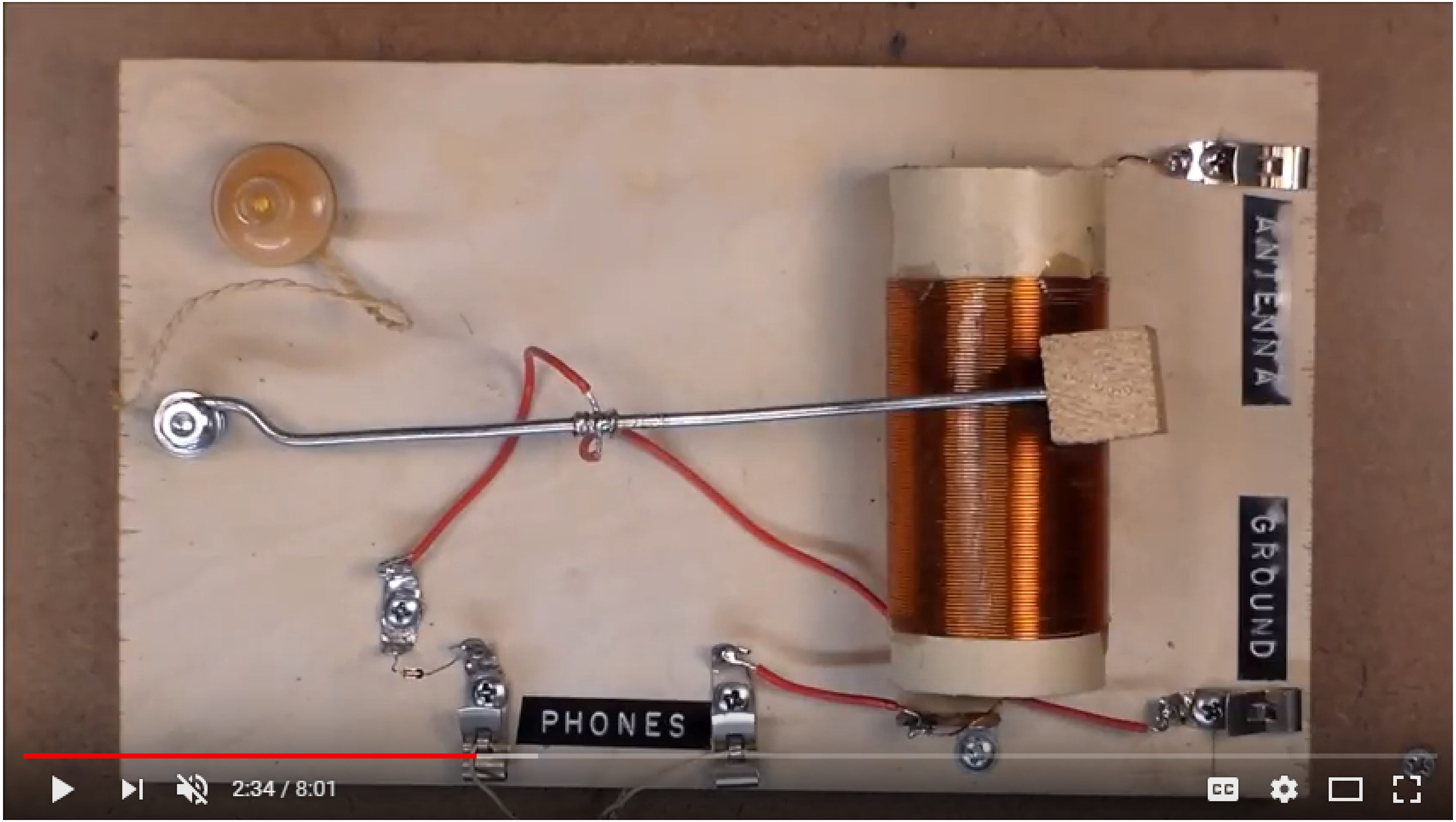
ANTENNA
GROUND
PHONES
2:34 / 8:01

BOY SCOUT CRYSTAL RADIO UPGRADE

SBX-1/Mobile HAARP (X-Band aka "The Golf Ball")

High-Frequency Active Auroral Research Program

#FrequencyWarfare #WeatherModification

@Conscious_god

WHERE DID THE TOWERS GO?

THE EVIDENCE OF DIRECTED FREE-ENERGY TECHNOLOGY ON 9/11, BY DR. JUDY WOOD

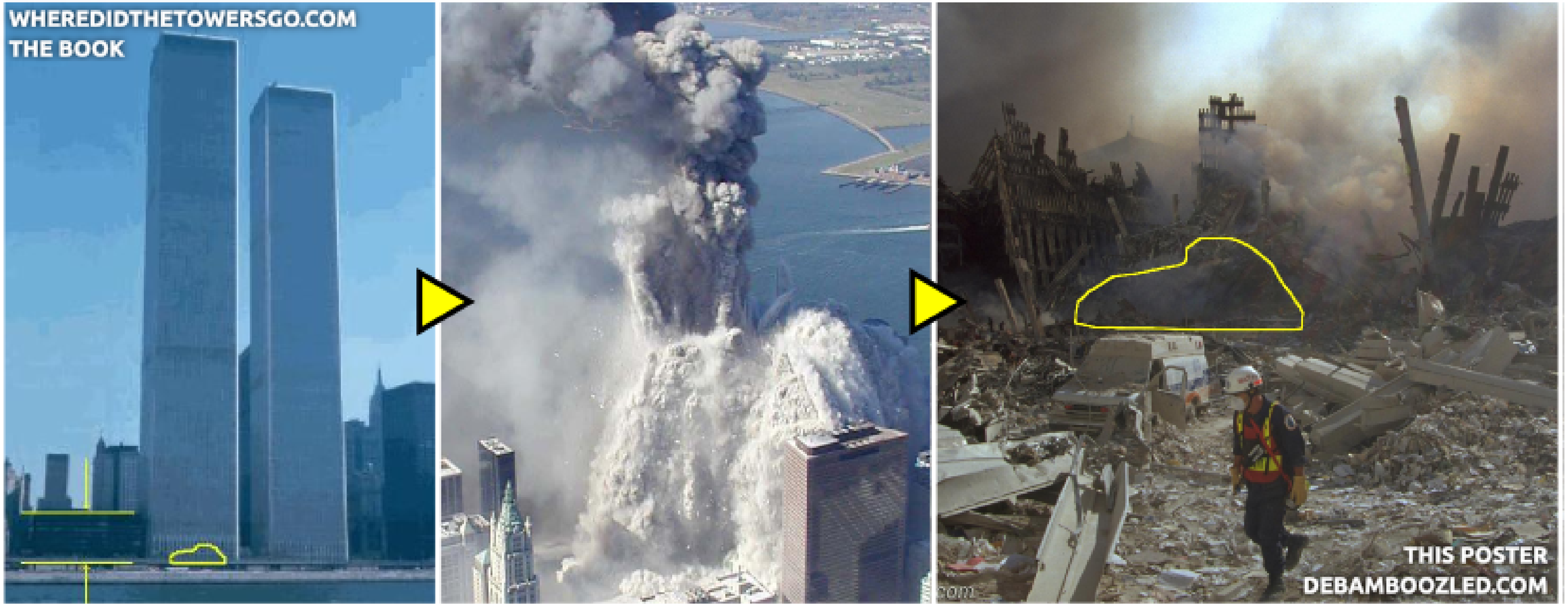

THEY LITERALLY TURNED TO DUST IN MID-AIR

"They didn't burn up, nor did they slam to the ground, but turned into dust in mid-air" Dr. Judy Wood - "Although they were demolished in such a way as to look as if they fell, they did not fall at all" John Lash - "Those towers literally disintegrated, they went into dust." Jesse Ventura - "A plane hitting the building, should not turn the buildings into dust." George Noory - "Arguably the most important book of the 21st century." Richard D. Hall - "If you don't have Dr. Judy Wood's book, do yourself a favor; buy it, read it." Mel Fabregas

CLASSIFIED FREE-ENERGY TECHNOLOGY REVEALED

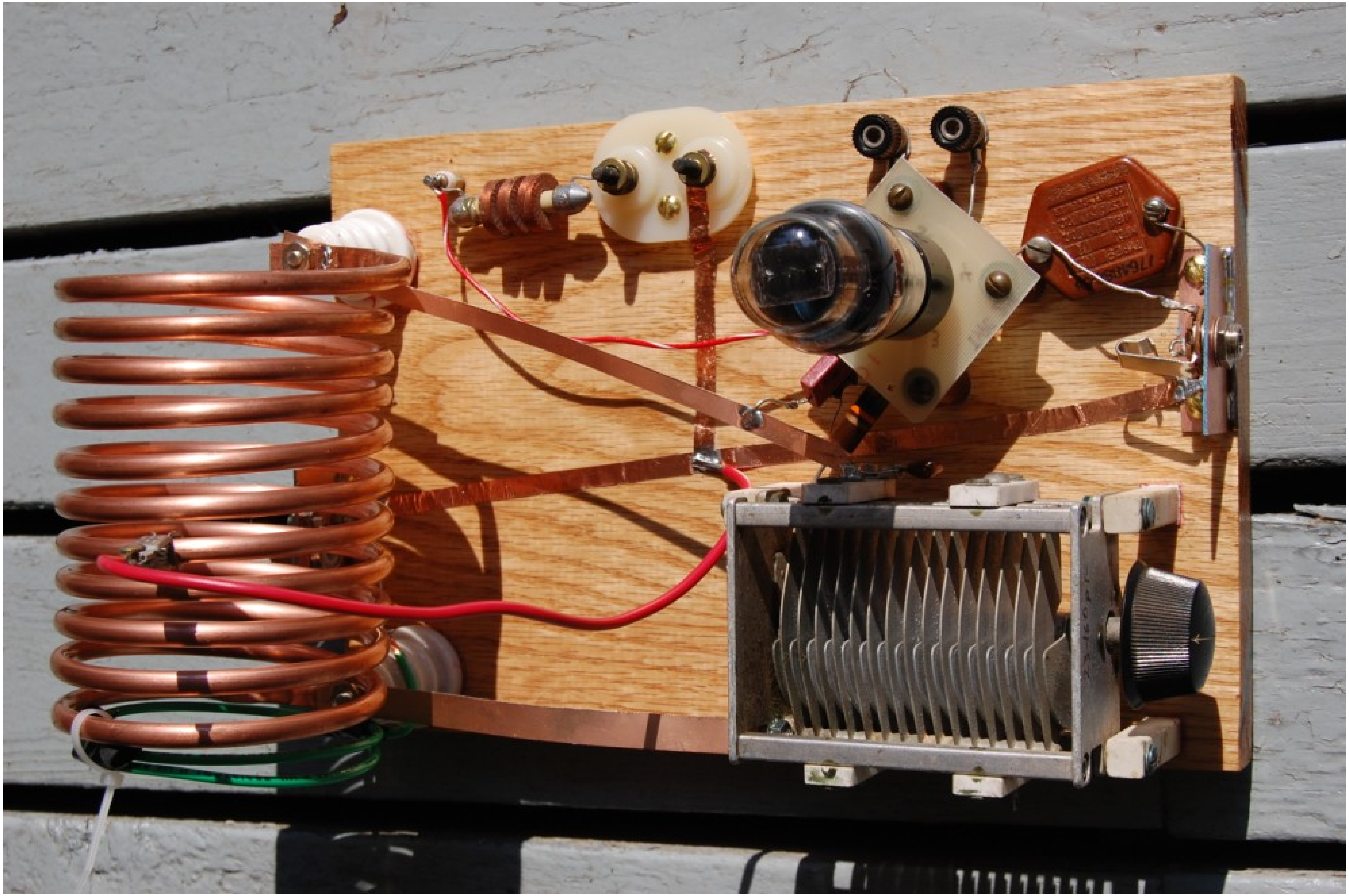

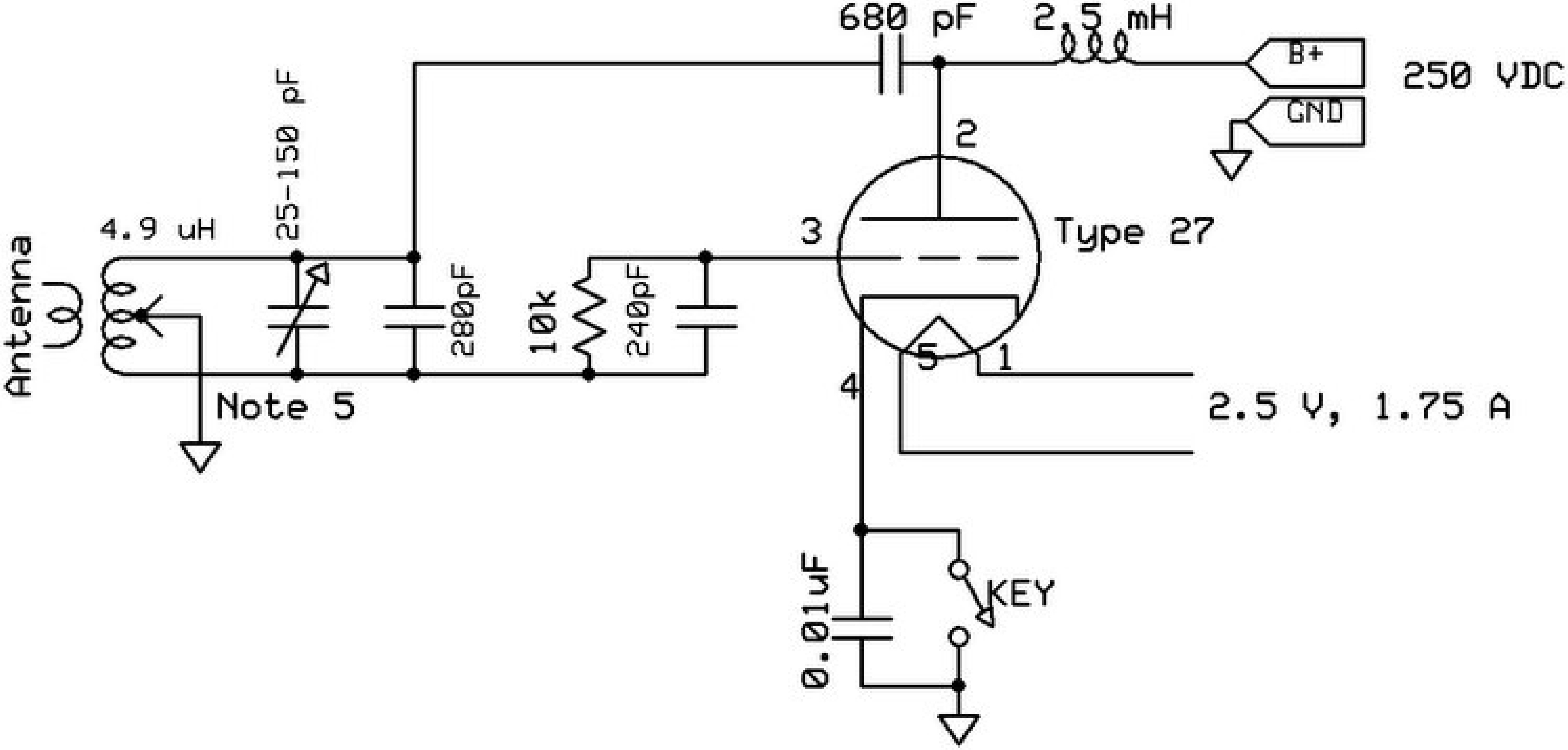
680 pF
2.5 mH
B+
250 VDC
GND
2
25-150 pF
4.9 uH
3
Type 27
Antenna
280pF
10k
240pF
5
1
Note 5
4
2.5 V, 1.75 A
0.01uF
KEY

70
80

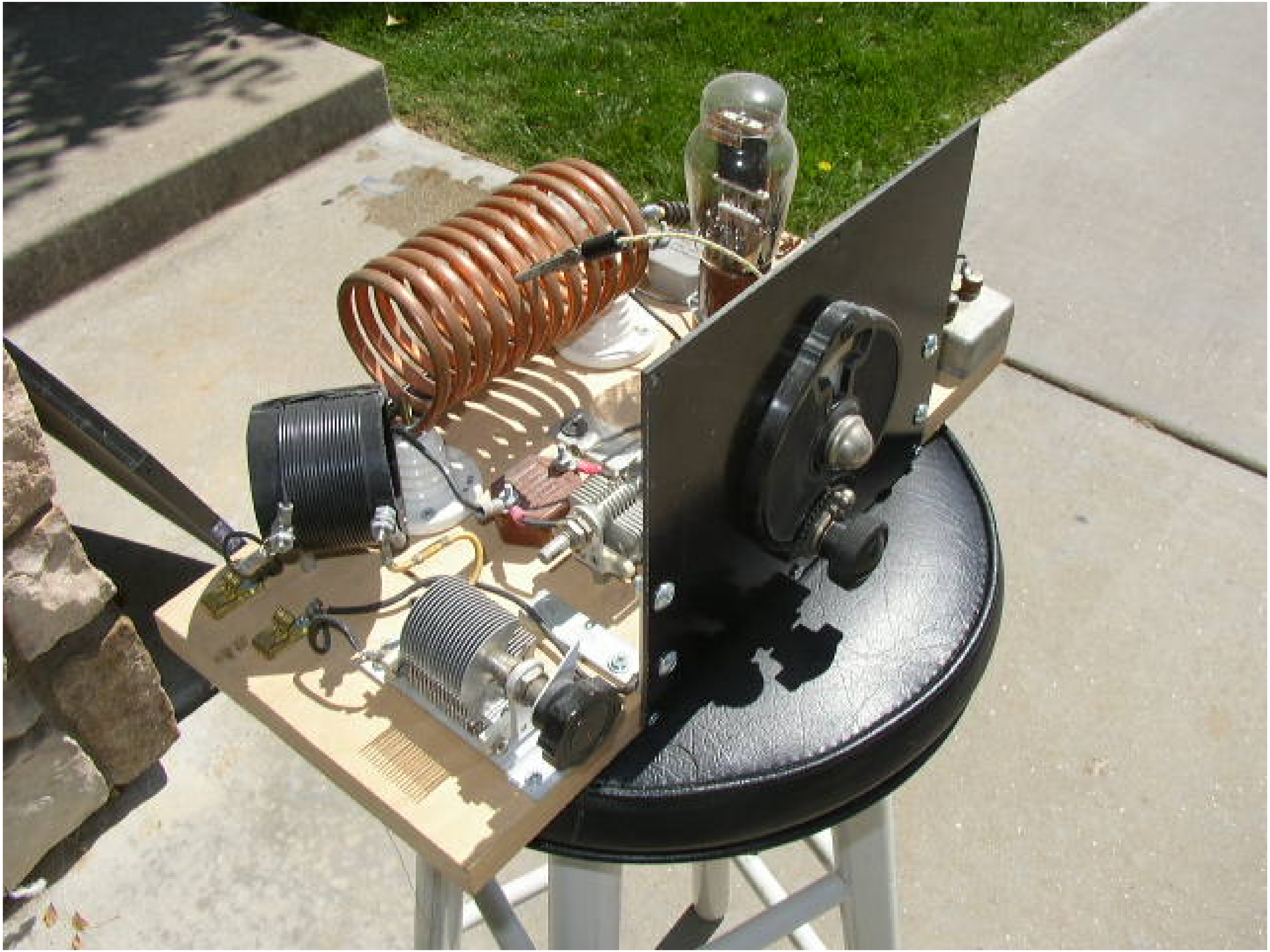

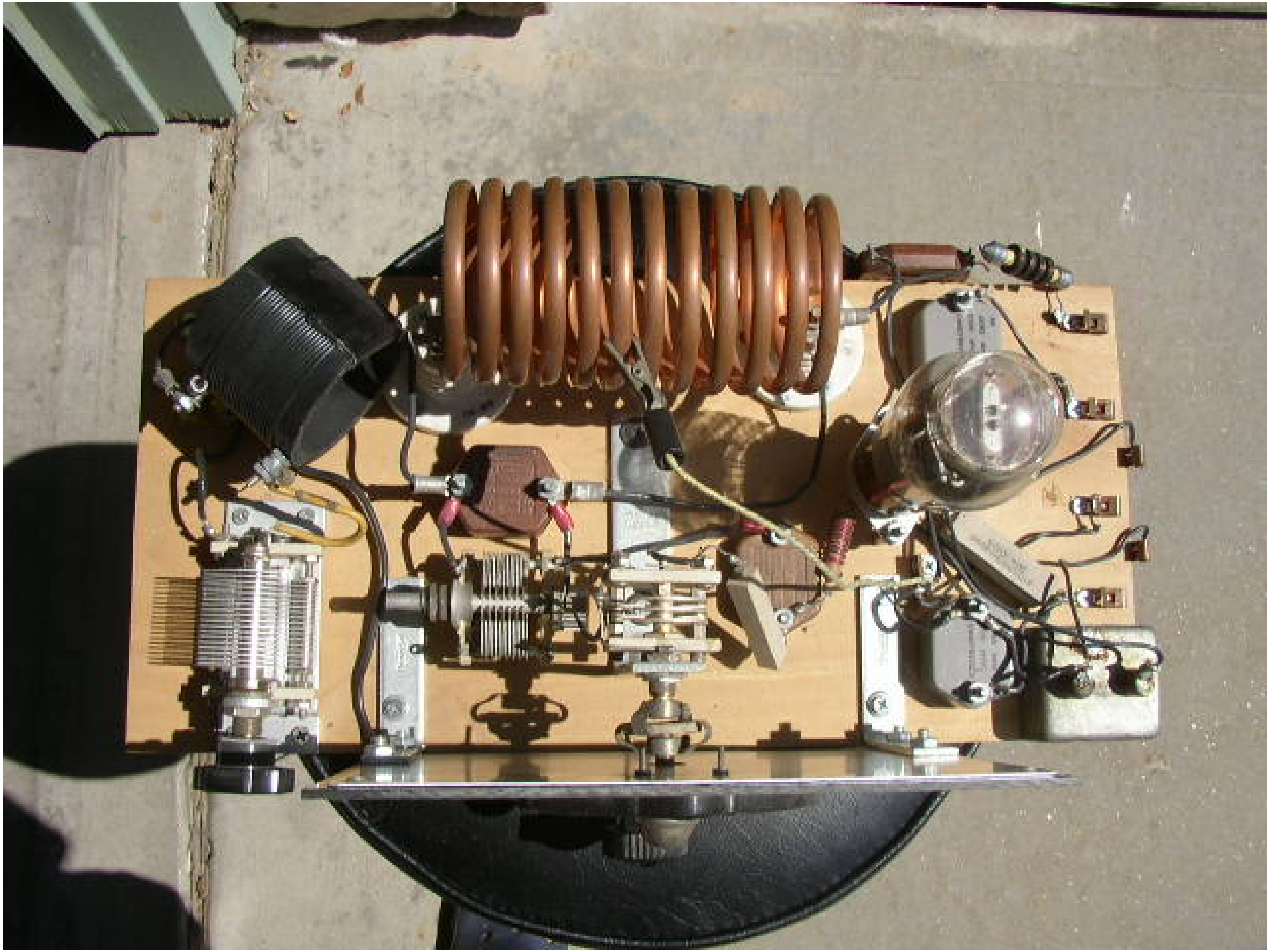

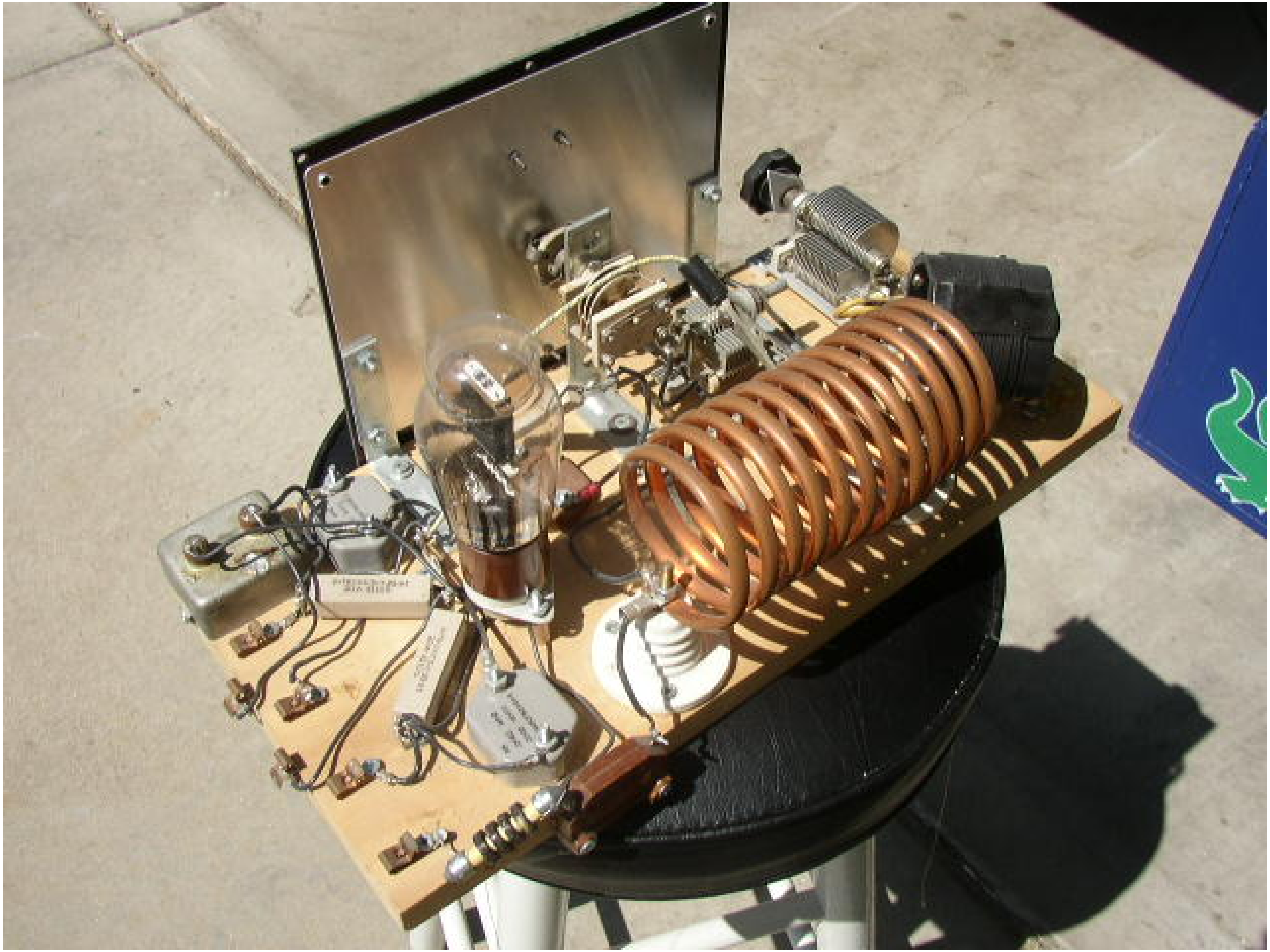

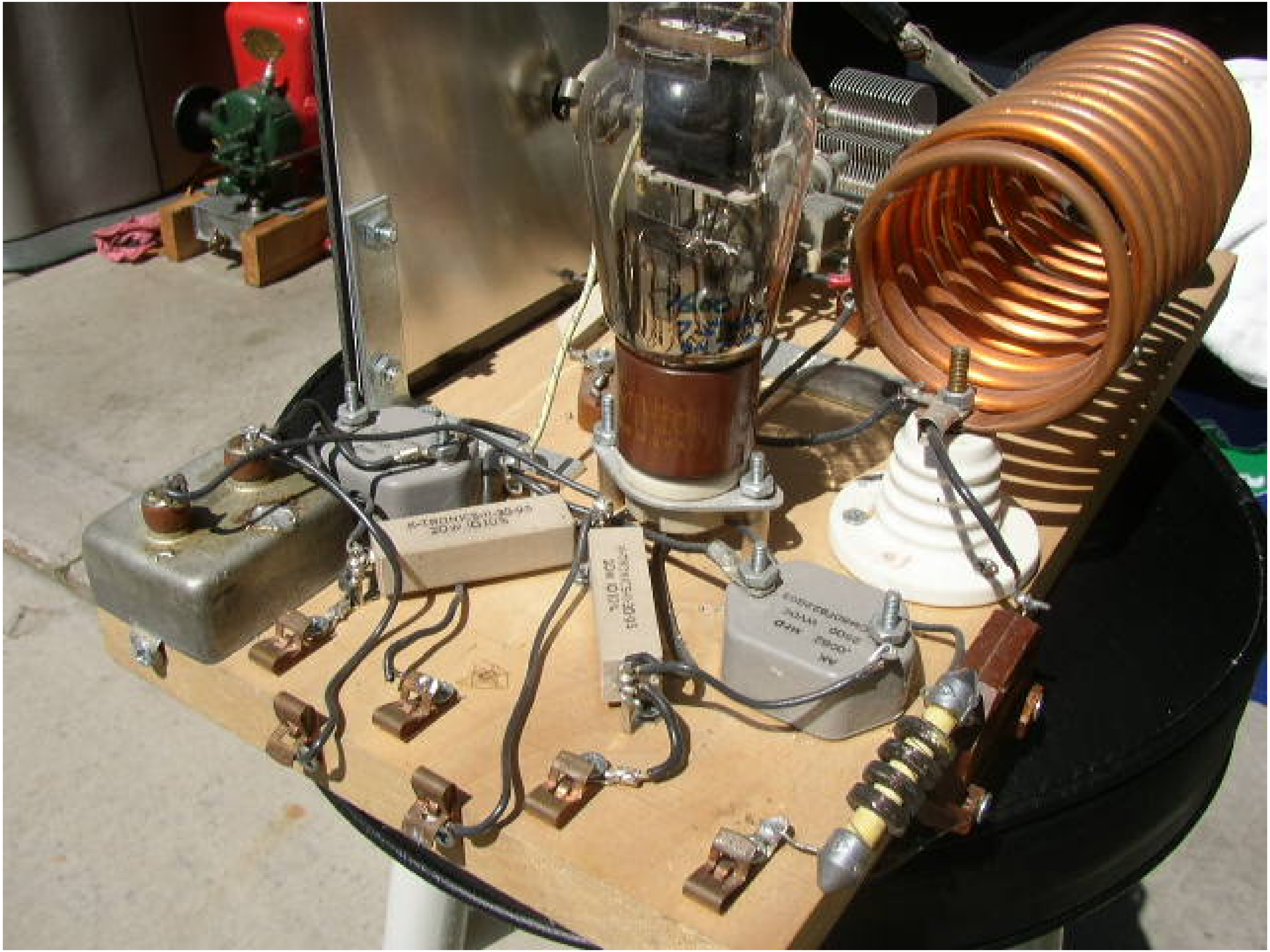

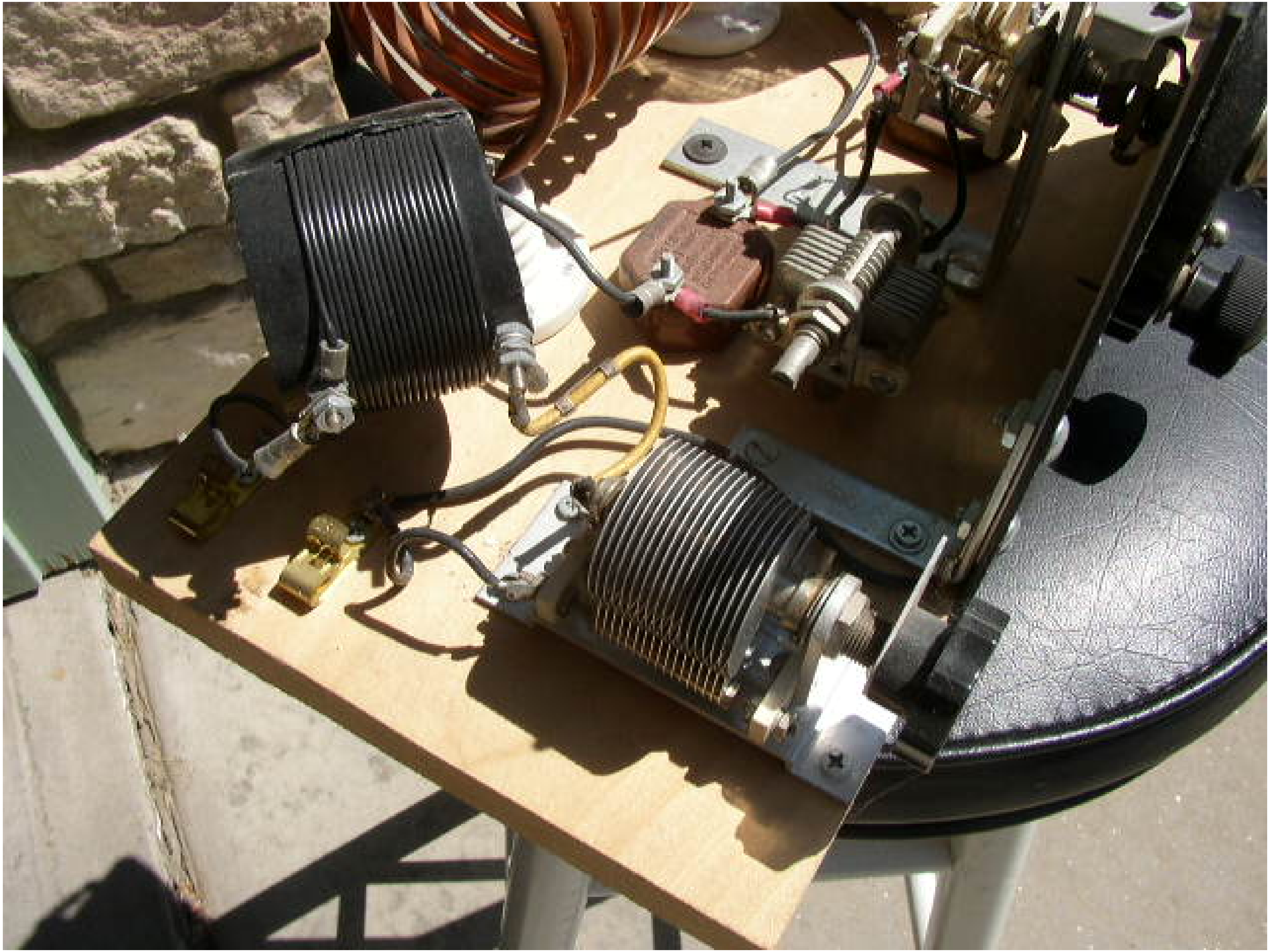

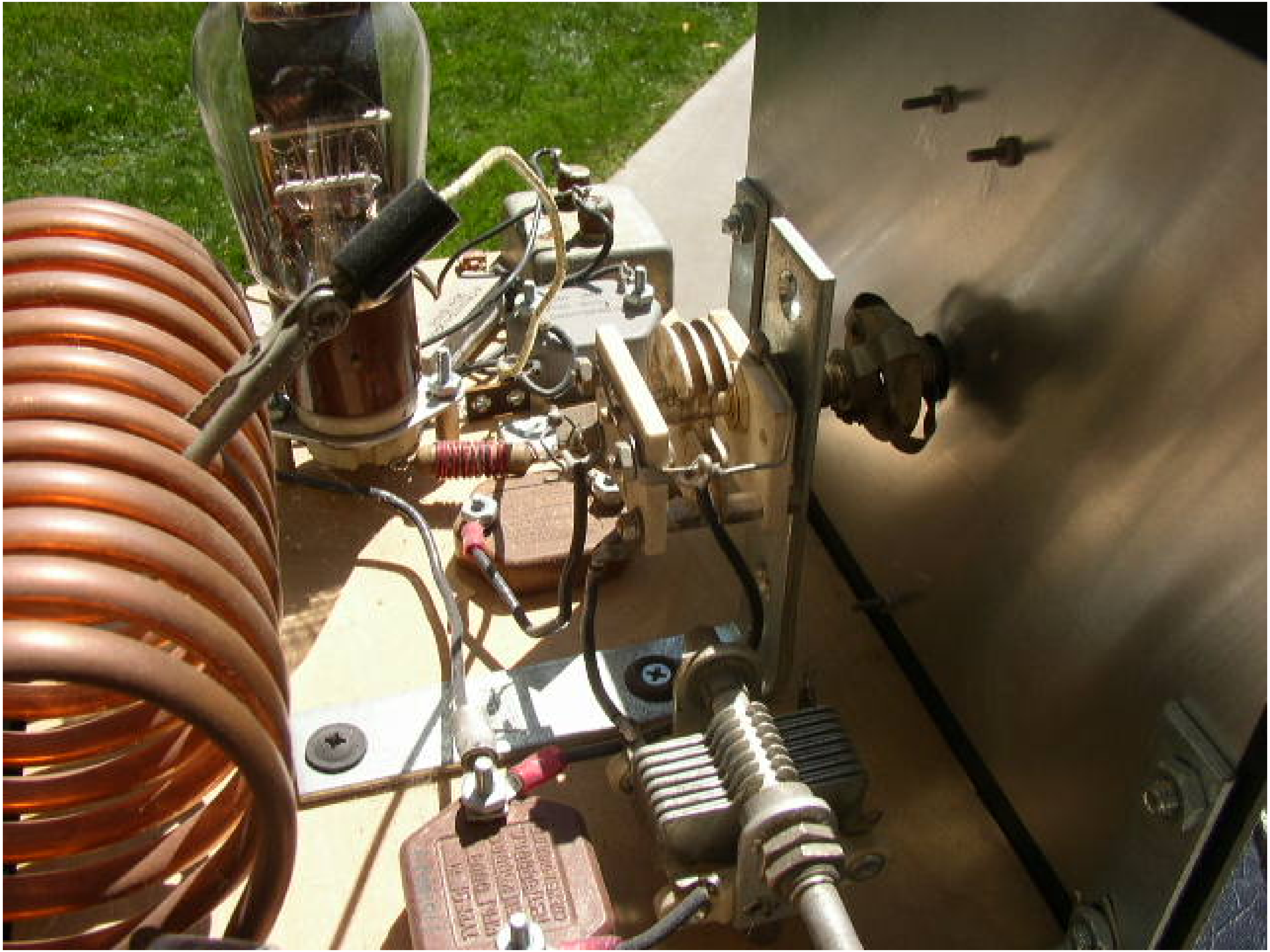

MILLIAMPERES
0
10
20
30
40
50
60
70
80
90
100

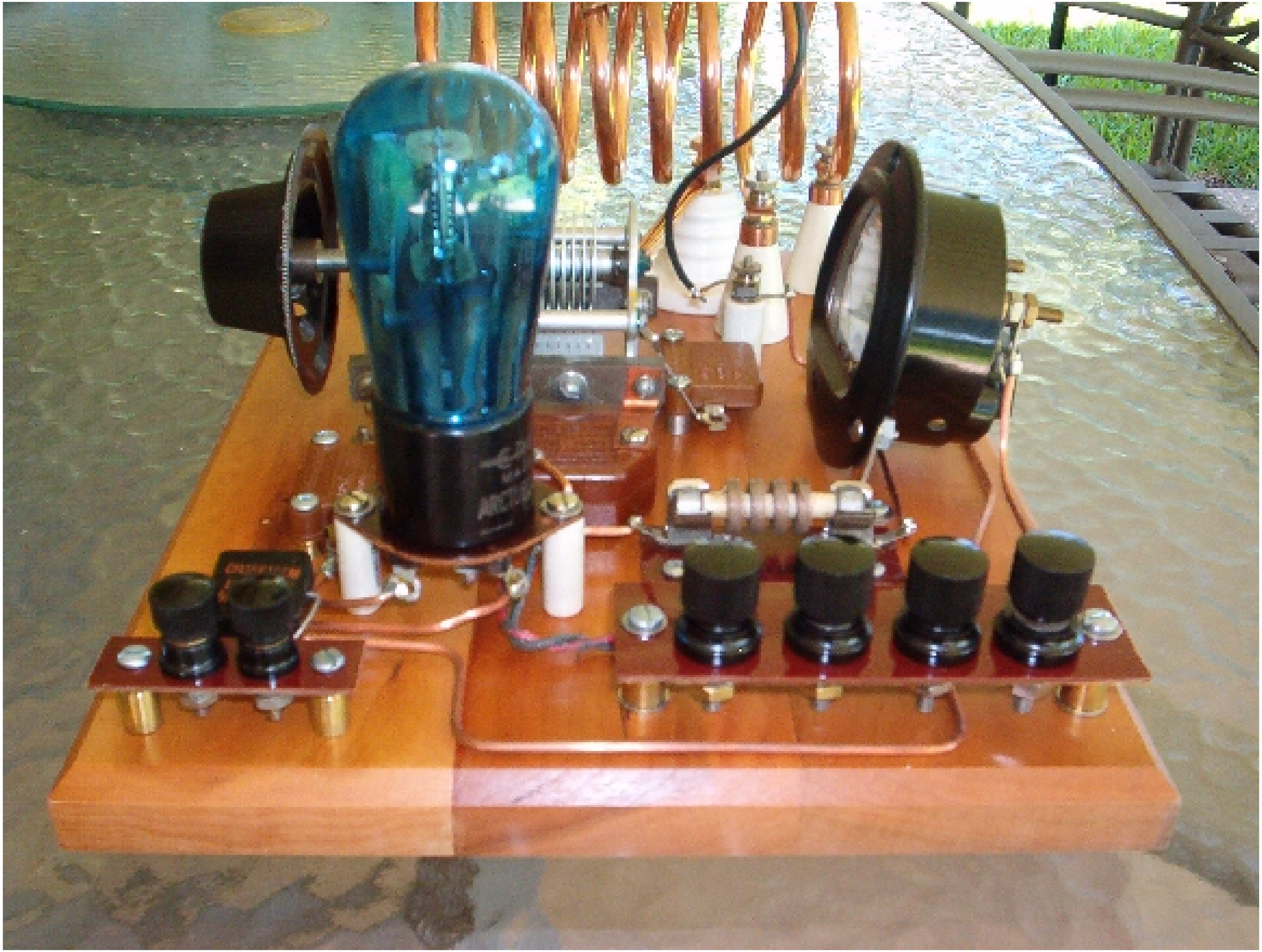

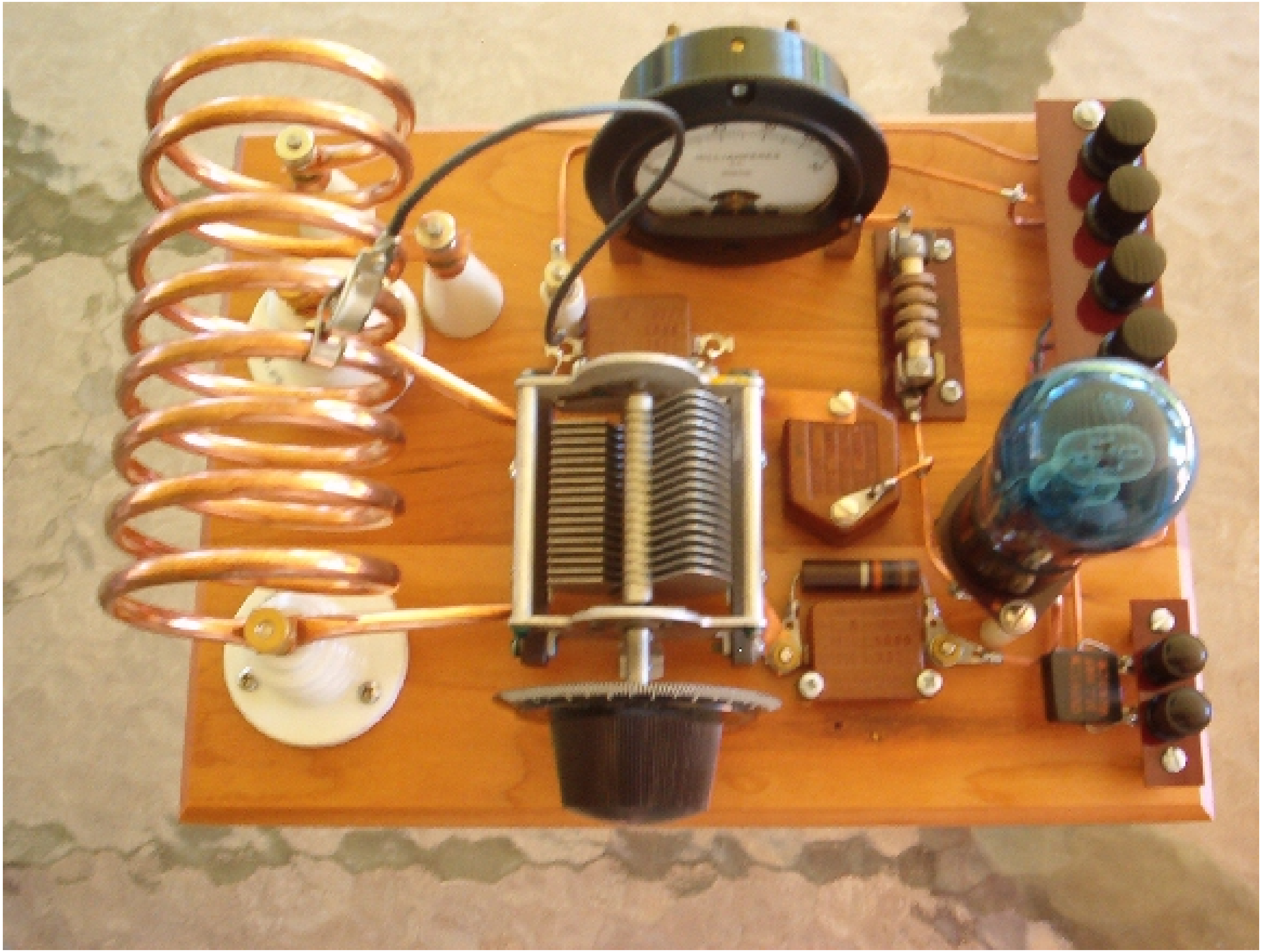

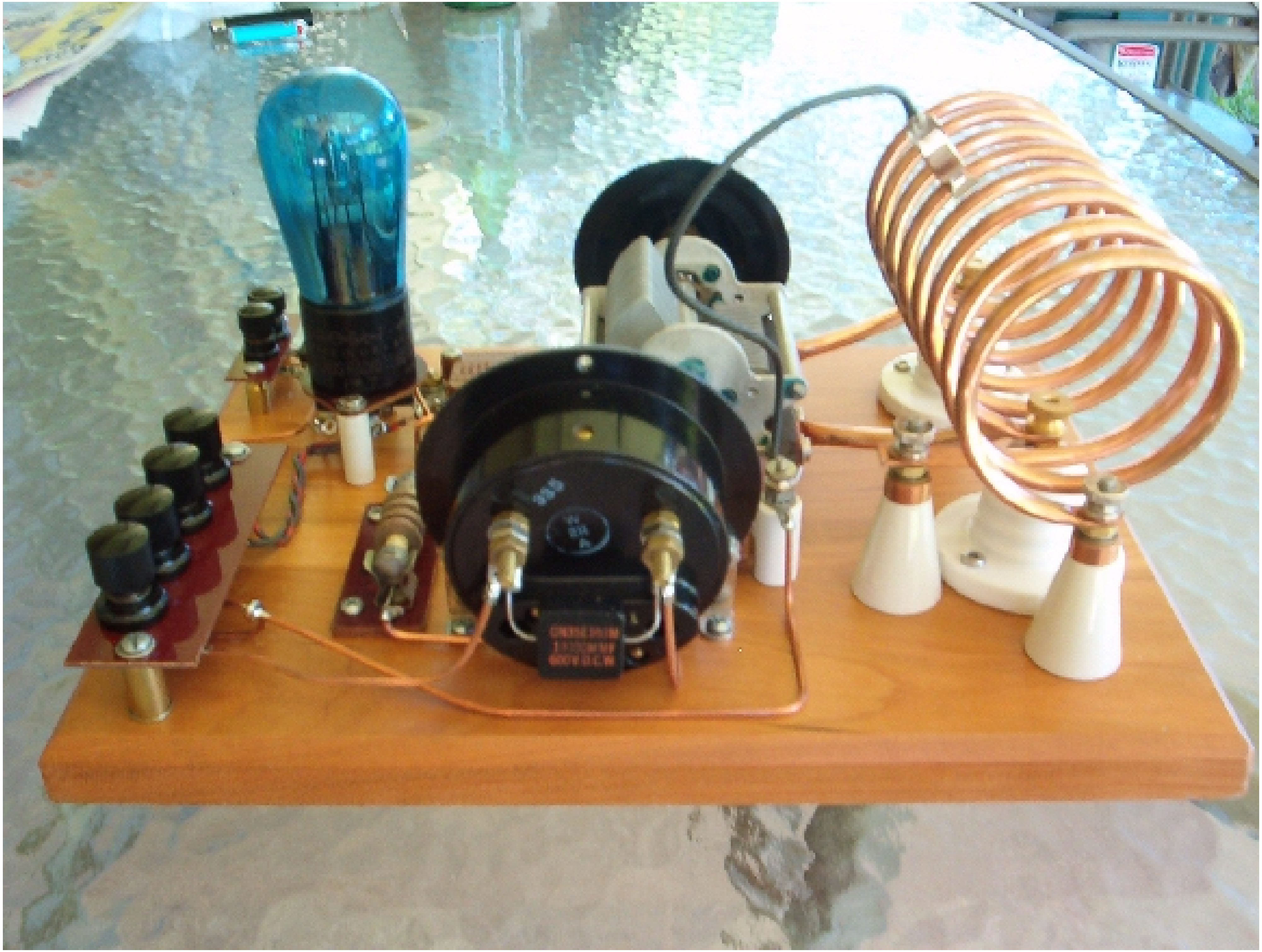

MILLIAMPERES
0
10
20
30
40
50
60
70
80
90
100

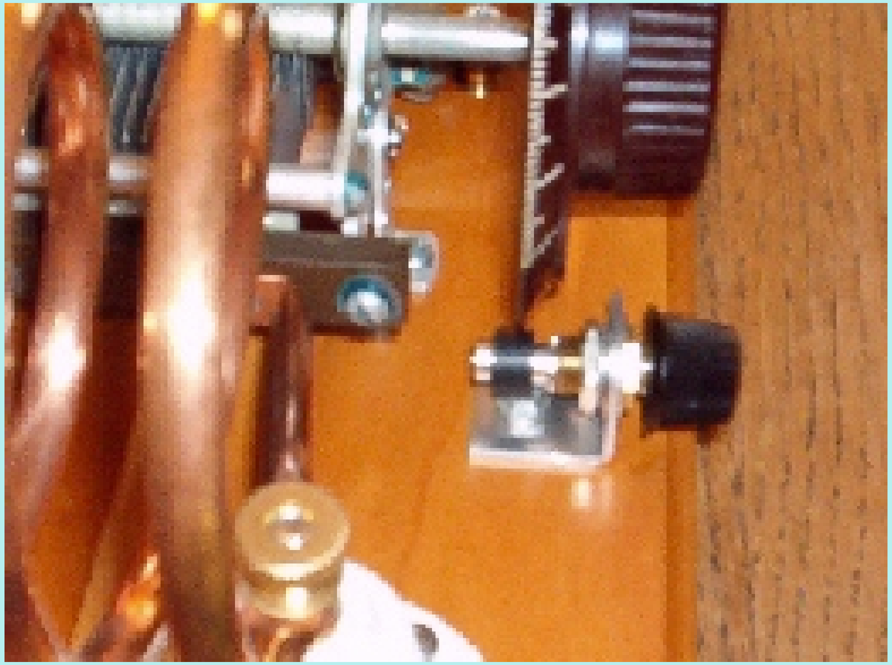

Side View

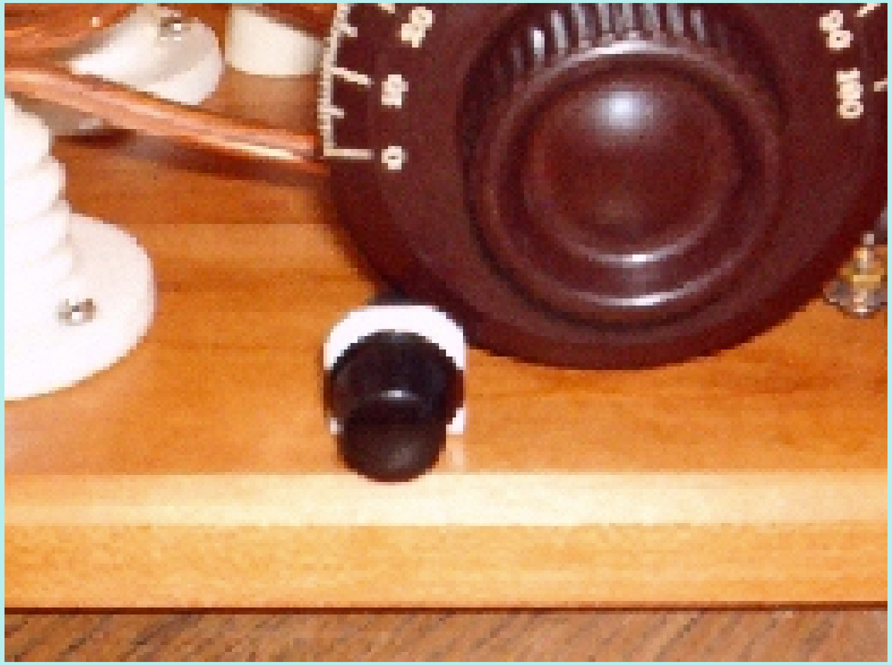

Front View

Articles written by James Tobola - KC5LDO
and reproduction, publication, or duplication of this article, or any part thereof,
in any manner is prohibited without the express written permission of the author. ©2012

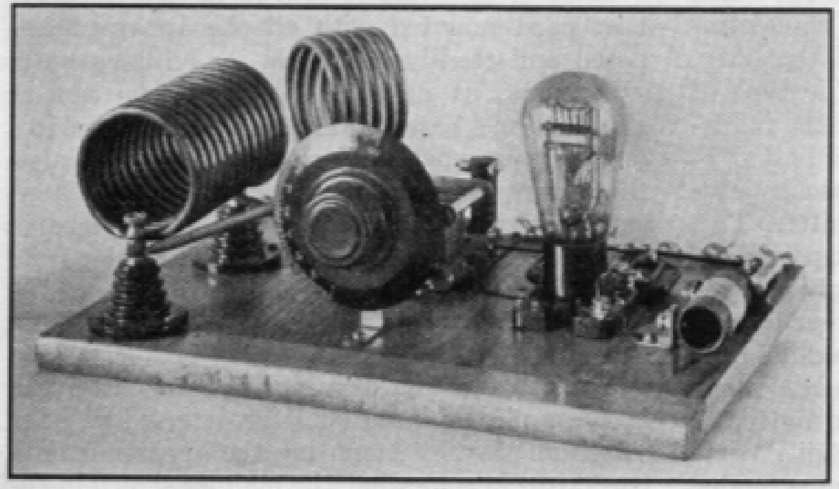

FIG. 703 — THE LOW-POWER SINGLE-TUBE TRANSMITTER

The plate tank circuit is at the left. The grid coil, leak and grid condenser are to the right of the Type 10 tube. The antenna coil is shown swung away from the plate coil to give loose antenna coupling.

Late last fall I saw an announcement regarding the upcoming '19
(AWA). Transmitters used during the event must only utilize 192
transmitters have to utilize self-excited oscillators! Listening-in o
imagine how different the bands must have sounded back in the l
heard sounded wonderful, considering the simplicity of the trans
came after watching and listening to W0VLZ's (Neil) superb You
watch these without wanting to roll-up their sleeves and start bui

After some research into the 1929 transmitter style, it became ap
Tuned (TNT) design. I can well imagine the countless late night l
simpler off-shoot of the TPTG design.

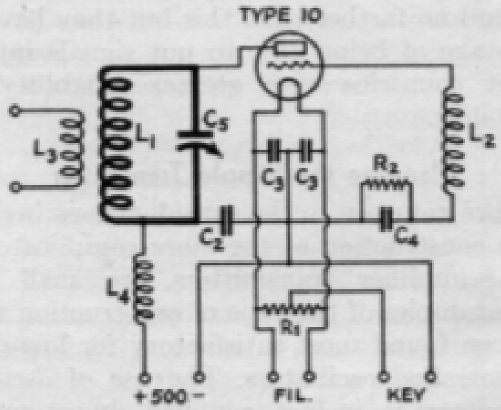

FIG. 704 — THE CIRCUIT OF THE TRANSMITTER

L_1, L_2 and L_3 — Plate, grid and antenna coils. The specifications are given under the illustration of the coils.

L_4 — A commercial "short-wave" receiving-type radio-frequency choke will do or one can be made by winding a two-inch length of half-inch tubing or wooden dowel with No. 38 d.s.c. or d.c.c. wire.

C_2 — 2000-µµfd. (.002 µfd.) mica fixed condenser, receiver type, if plate voltage does not exceed 500.

C_3 — 5000-µµfd. (.005 µfd.) mica fixed condenser, receiver type.

C_4 — 250-µµfd. (.00025 µfd.) mica fixed condenser, receiver type.

C_5 — 500-µµfd. (.0005 µfd.) variable condenser. Any good receiving condenser will be satisfactory.

R_1 — Center-tapped resistor, 75 to 100 ohms total resistance.

R_2 — Grid-leak resistor, 10,000 ohms. Any small resistor rated at 5 watts or more will do.

Three General Radio or similar stand-off insulators will be necessary, as well as 8 Fahnestock clips, some miscellaneous small machine screws and nuts, and a few feet of bus wire.

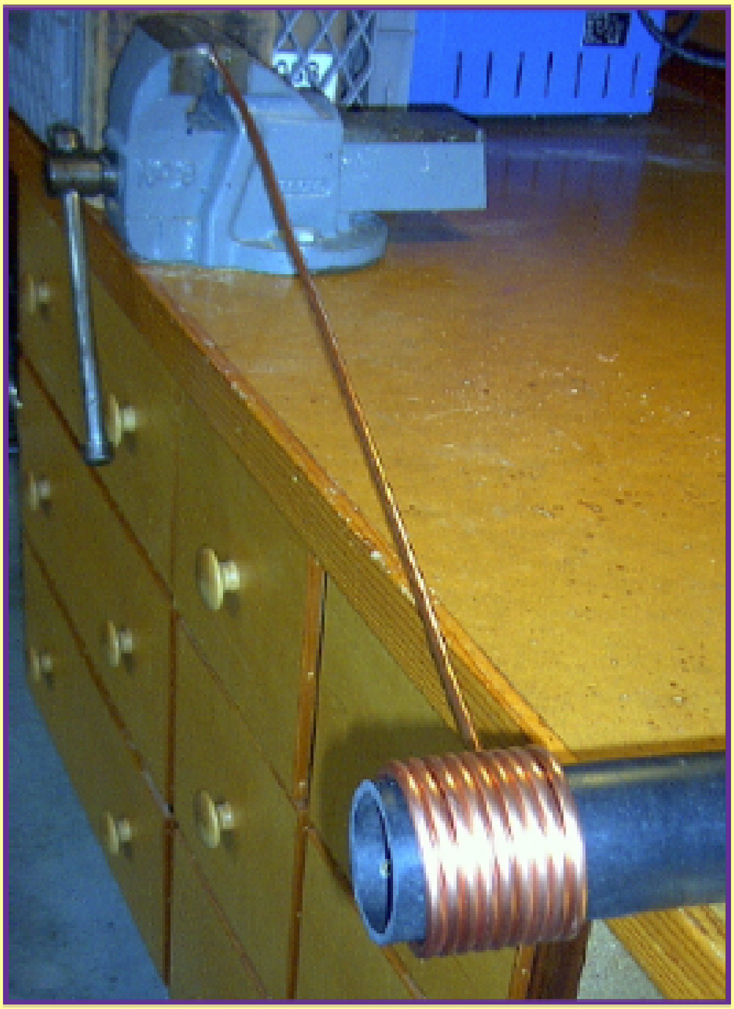

r coils. Once the proper number of turns is reached it is just a matter of flattening and drilling the ot
, I wound a plate coil for 40m as well. I haven't been brave enough to try the TNT on 20m yet but I w

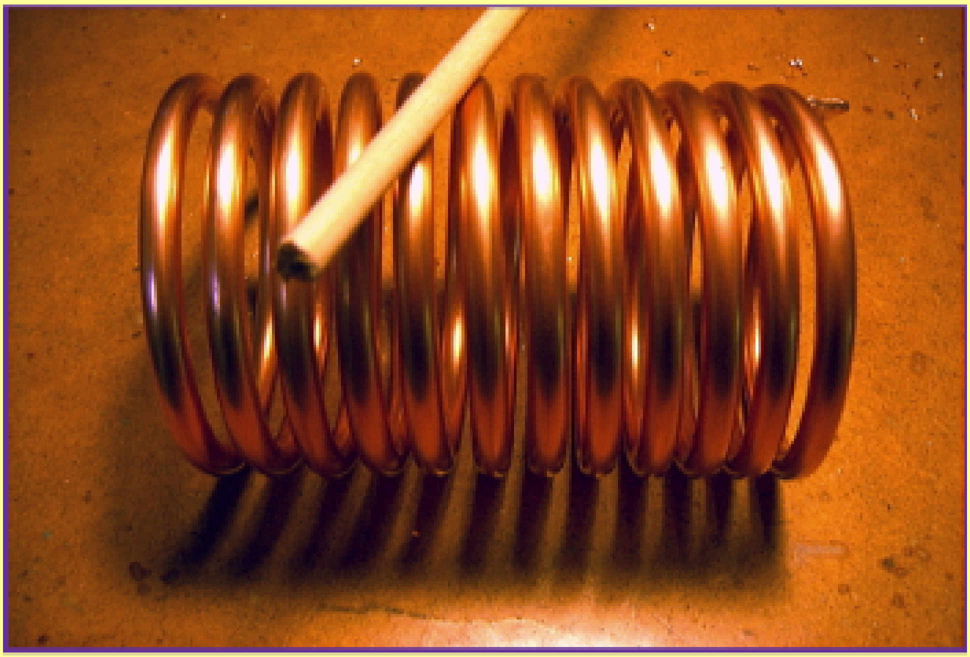

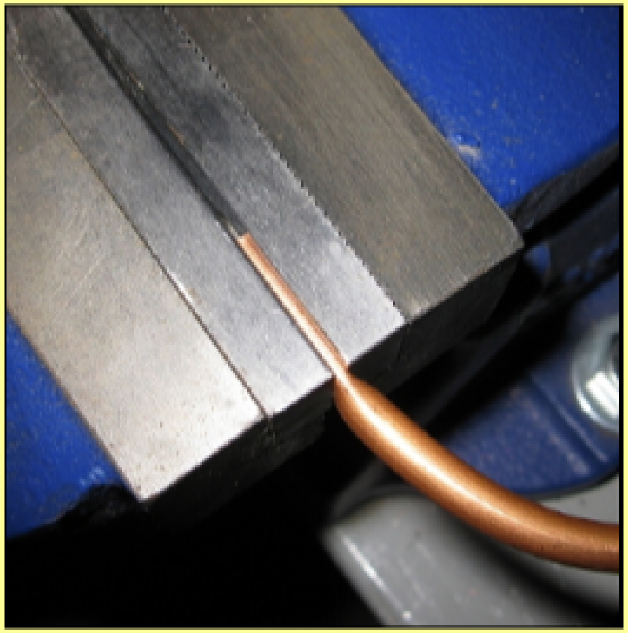

ither wound on bakelite tubing or on well-sealed wood dowel. Not having any bakelite made the choice any easy one. The 1" forms were made from some Yellow C

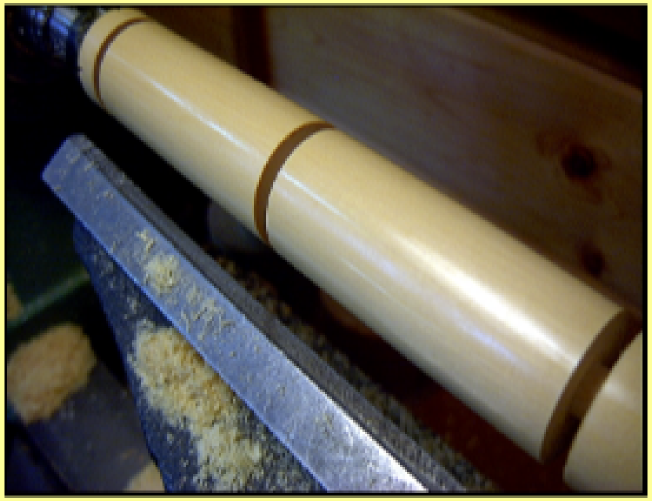

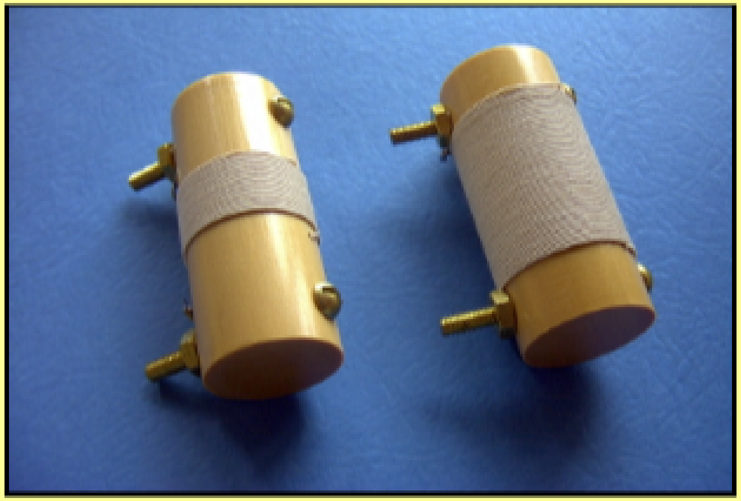

resistor was fabricated to resemble the original bakelite-enclosed 'Pilot' style, popular in the late 20's.The pictures indicate how this was done and the finished result. The small plexi-glass form was filled with black Fin
emoved and the entire package was baked in the kitchen toaster oven at its lowest temperature for several hours. This achieved the desired hardening effect and a suitable reproduction Pilot filament resistor.

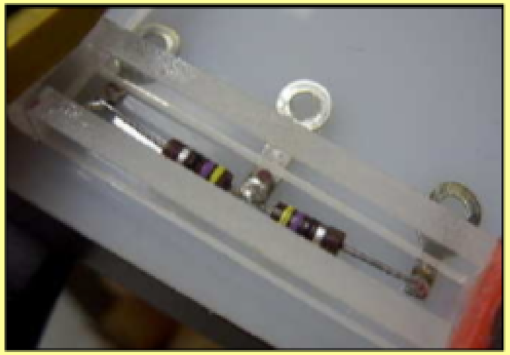

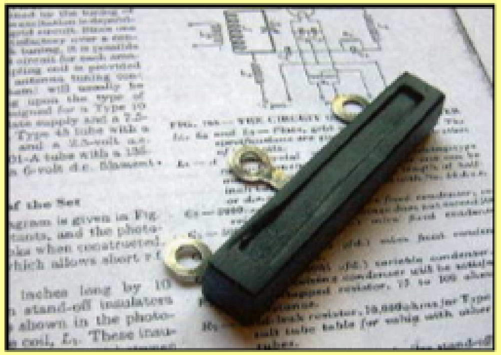

reproduction also, of an early 'Lavite' model. The ends of a new wire-wound resistor were removed and found to be made from brass. These ends were then fitted to the body of an older style 10K resistor, soldered, pai
y across the terminals of the grid cap. I found out later that the actual value of the grid leak is quite critical in the TNT. I tried various values and luckily the one I had manufactured turned out to be perfect. My earlier
UX-210) required a far larger grid leak to produce best keying and good output. If you are making your own grid leak I would recommend that the value be optimized first, before the grid leak is built in its final form.

CRESCENT
CRESCENT

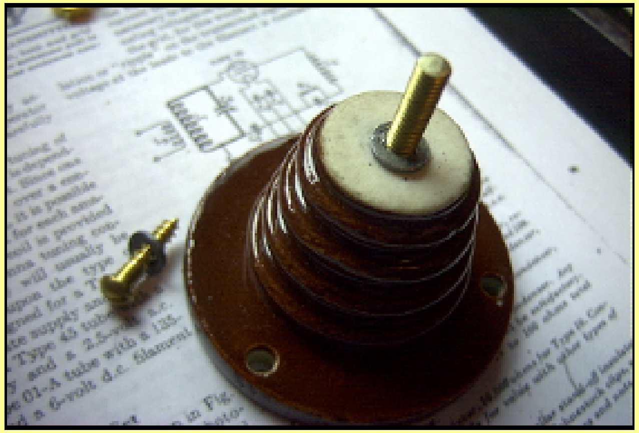

as breadboarded - first using the Type '45 and later with the Type '10. Various values were tried for both the grid leak and for the grid capacitor. Both affect keying and outp
ling of the plate voltage in order to remove high voltage from the large exposed tank coil. I did not want to run the chance of accidently grabbing hold of it late some evening
ference between shunt-feed and the standard series-feed method. It saddens me to think of all of the amateurs of the 20's or 30's that may have been unnecessarily hurt or ki

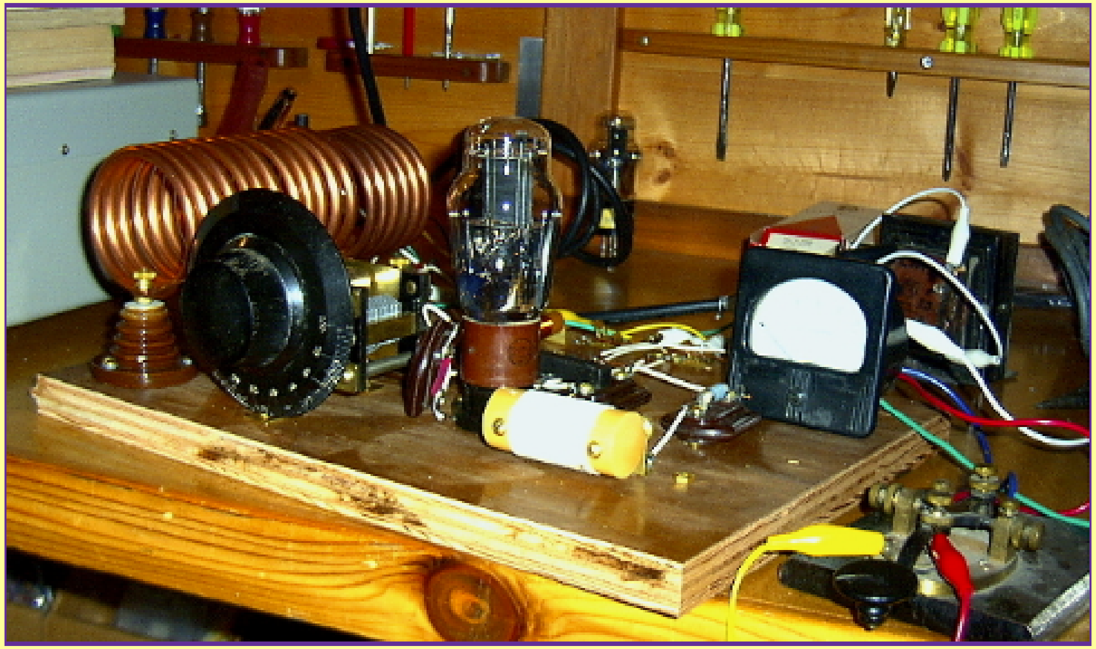

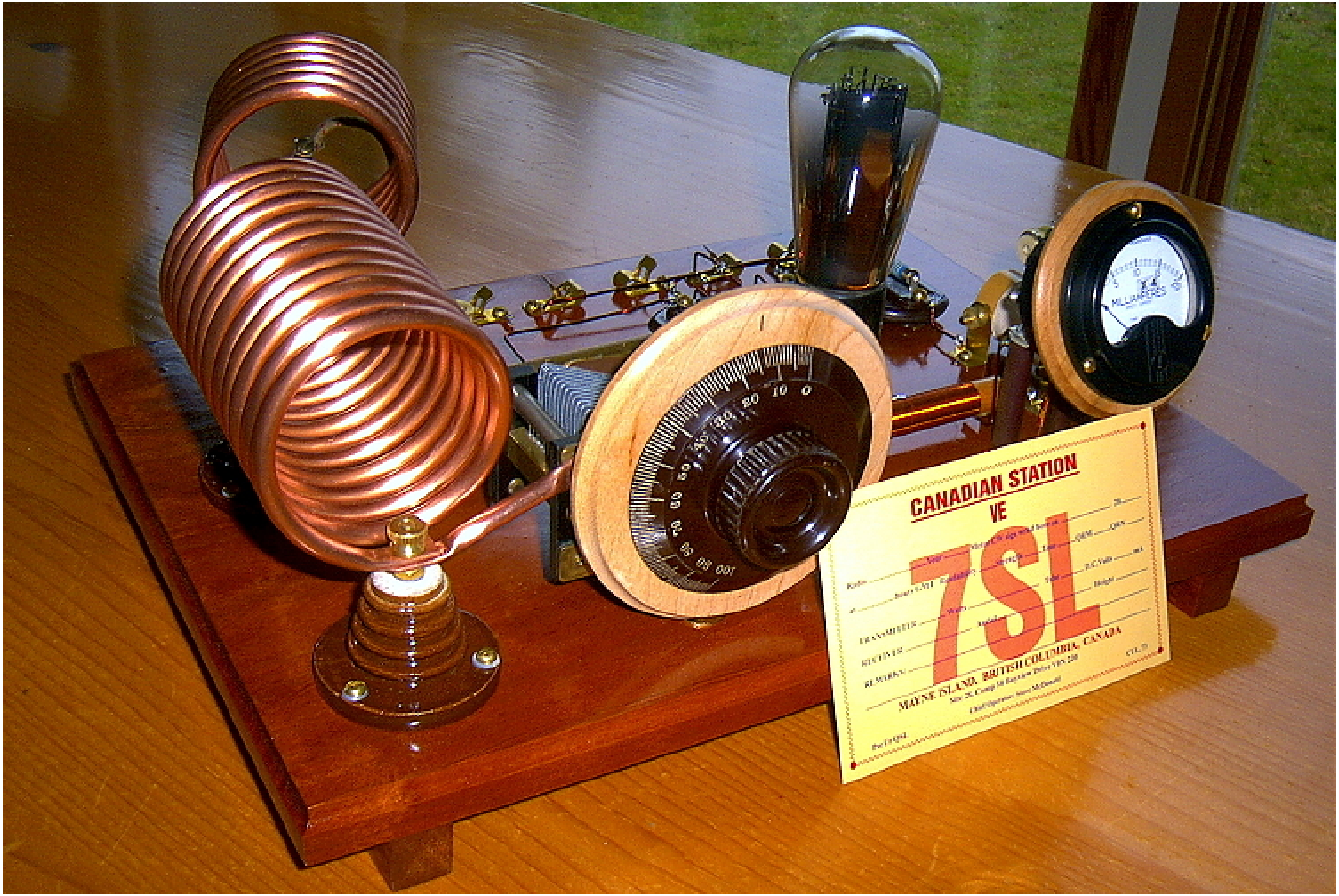
MILLIAMPERES
CANADIAN STATION
VE
7SL
TRANSMITTER
RECEIVER
REMARKS
MAYNE ISLAND, BRITISH COLUMBIA, CANADA

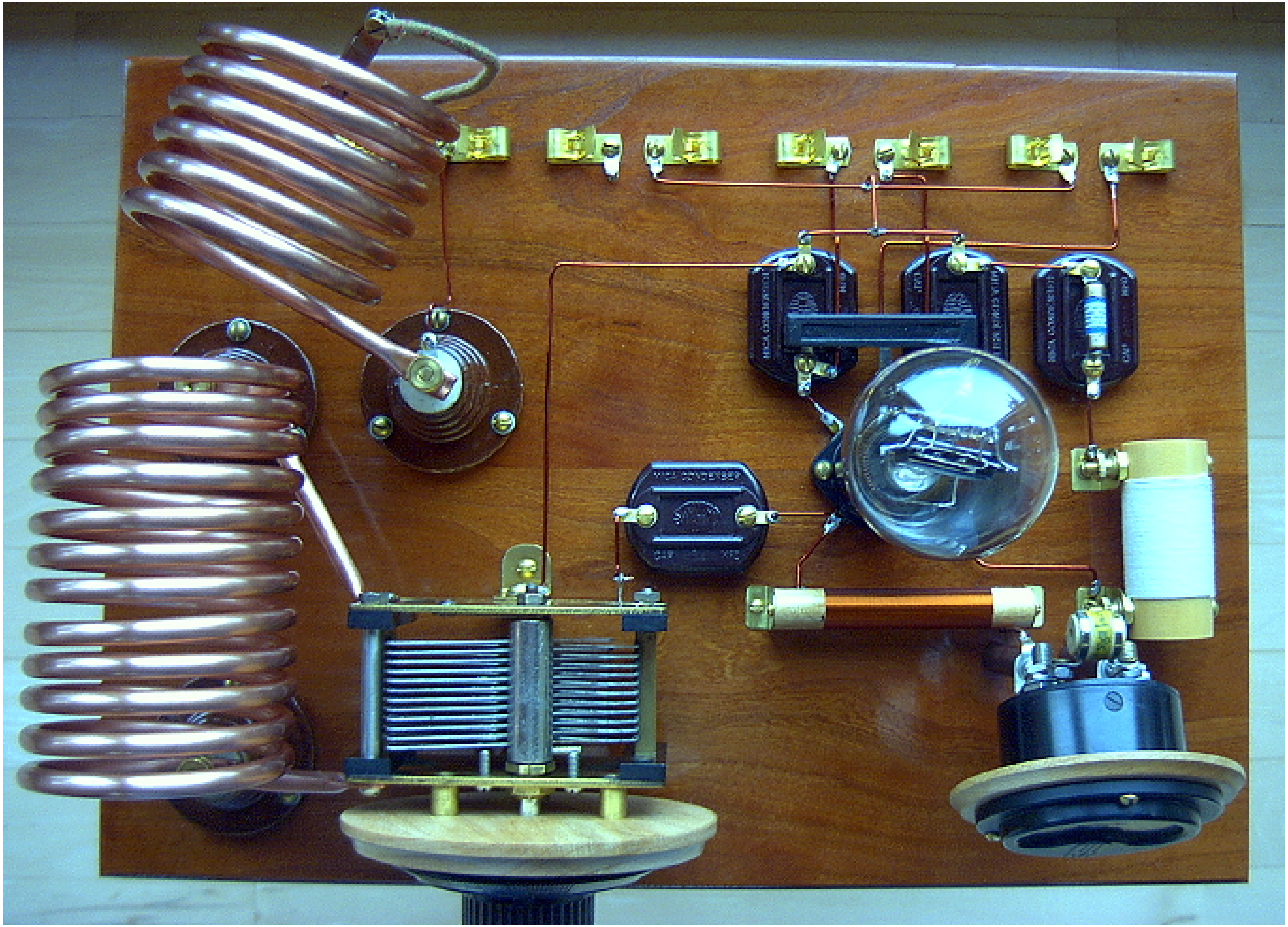
MICA CONDENSER

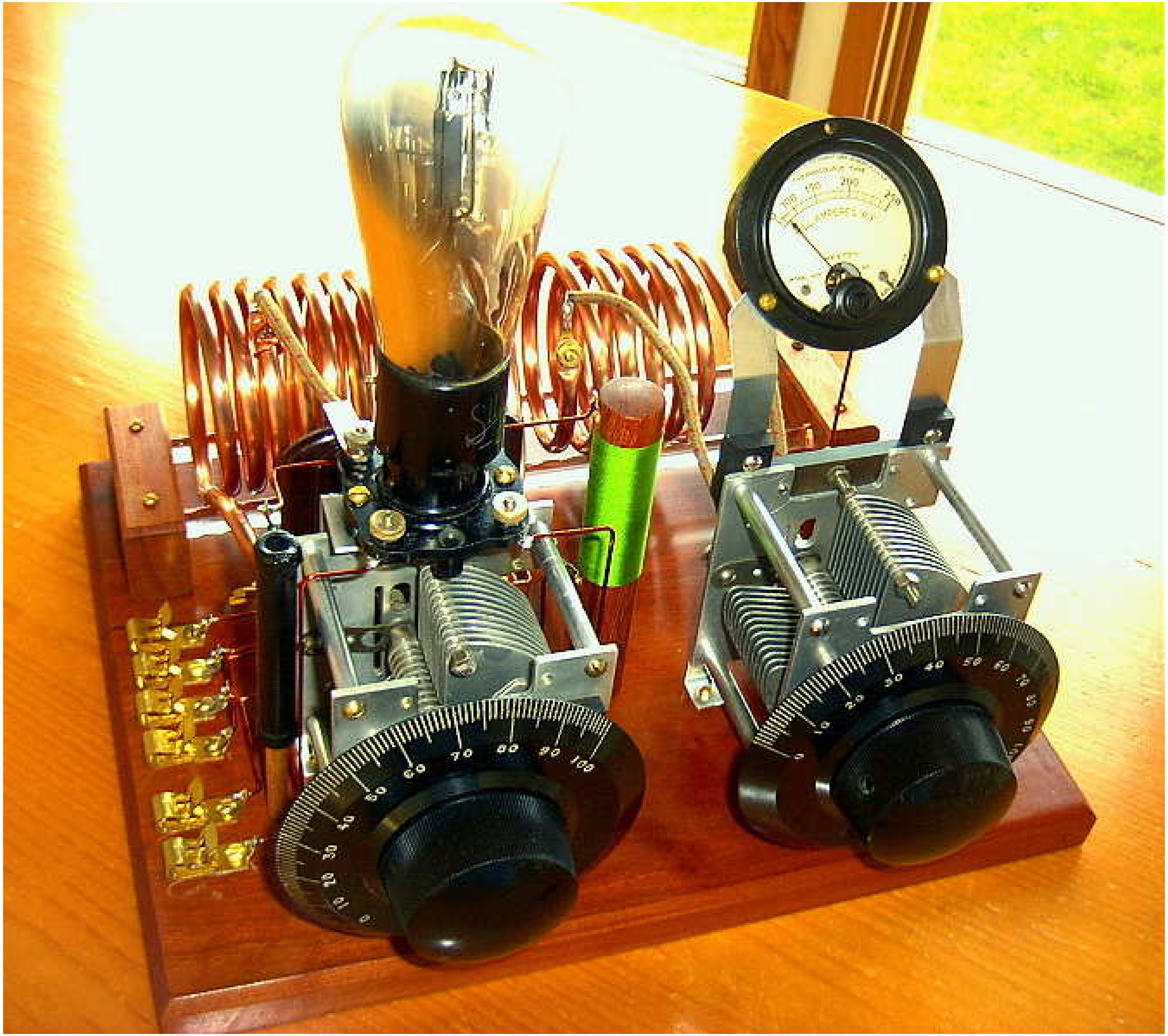
0
10
20
30
40
50
60
70
80
90
100

rexresearch.com

Carl BAUGH

Rejuvenation

Carl Baugh achieved phenomenal rejuvenation and other effects with a hyperbaric oxygen Biosphere plus EMF / frequencies.
Unfortunately, however, his Christian creationist beliefs have detracted from his credibility.

---

fetchingstock
shutterstock

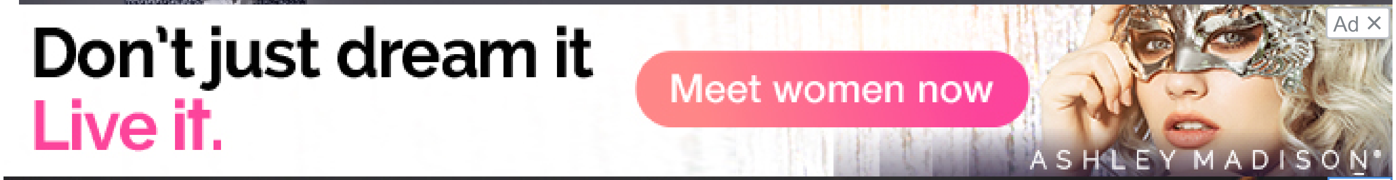
Ad
Don't just dream it
Live it.
Meet women now
ASHLEY MADISON

Ad
Education Career Fair
Sat, Feb 2 & Feb 9 | 9:30am-12:30pm
San Diego • Oakland | Los Angeles • Fresno • Sacramento
> LEARN MORE
Alliant International University

---

http://www.keelynet.com/biology/baugh.htm
03/31/01

Dr. Baugh & the Pre-Flood Environment

Director of the Creation Evidence Museum in Glen Rose, Texas, Dr. Carl Baugh has invented and patented a 'Hyperbaric Biosphere' chamber to test Biblical claims regarding pre-flood atmosphere and magnetic conditions based on Genesis.

Genesis states that life on earth was created 'perfect' and designed to live forever. It also states that after the fall of man, all elements of which everything is made, were cursed. Genesis then records the gradual degeneration of the whole of creation until the flood of Noah followed by an increasing rate of degeneration as time progressed.

If the dates and years reported use the same standards we do in modern times, then there is an overlapping of pre-flood humans with that of post-flood survivors where the pre-flood humans lived many more years.

The average age of man prior to the flood was approximately 920 years, after the flood, mans lifespan declined by 250 years until the time of Christ when lifespans averaged 70 years. Tests indicate simulation of a pre-flood environment would produce conditions of increased longevity, cellular growth and a change in molecular structures.

Dr. Baugh's experiments with his special hyperbaric biosphere have produced fascinating results that indicate potential methods to heal and/or rejuvenate people today. NASA has even requested Dr. Baugh's research and discoveries to determine their applicability to the space program.

Hyperbaric Biosphere

Experiments carried out in this pre-flood atmospheric chamber have produced amazing evidence supporting the biblical record of the pre-flood world.

Some examples include;

extended life span of drosophila fruit flies up to three generations, molecular change of venom in copperhead snakes to a non-toxic state, accelerated growth of piranha fish - from 2" to 16" over a 2.5 years period.

The Hyperbaric Biosphere patent allows for the ability to vary the mix of gases comprising the atmosphere in the chamber as well as to vary the electromagnetic field intensity, frequency and duration.

Dr. Baugh and an unidentified associate handle deadly copperhead snakes used in an experiment to test whether increased oxygen levels, as would be experienced in pre-flood conditions, might alter the physiology of the snakes.

He contends an artificial pre-flood environment could be duplicated with increased oxygen, higher pressure and the use of higher EM fields

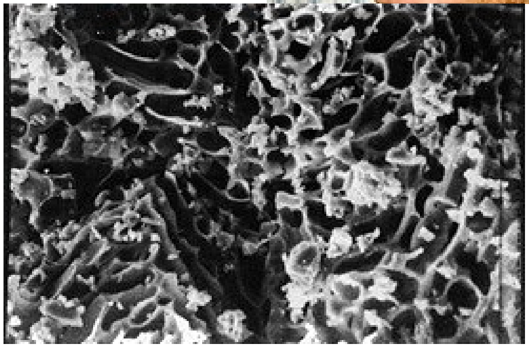

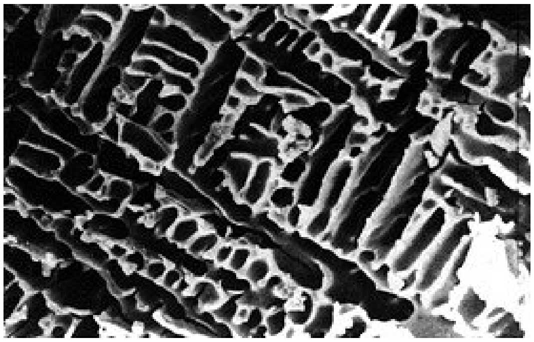

Another investigation at the creation evidence museum, the gnarled 'spaghetti-like' formation of copperhead snake venom as seen under the scanning electron microscope before the copperhead snake was 'housed' in a small hyperbaric chamber.

After four weeks in the hyperbaric chamber, venom from the same copperhead snake shows a much less distorted structure (less gnarled) indicating a lowering of the toxicity level. Conditions in the hyperbaric chamber are seeking to provide the type of environment that existed in the time prior to the great flood.

Additional information that might be of use;

It has been estimated by researchers at the University of Illinois that the atmospheric pressures in the earliest beginnings of life on earth, could have been as much as 29.4 pounds per square inch, twice the current sea level pressure (14.7 psi at sea level).

1) oxygen content would have been about 30% instead of the 21% we experience today

2) an open wound could heal overnight under these conditions

3) this atmospheric condition would support human life lasting several hundred years

A related post sent to a discussion list recounts the following unconfirmed information about rejuvenation effects noted by aquanauts living underwater for periods of 30 days or greater.

CJT Enterprises wrote:

Hi everyone,

A friend of mine who worked for the NSA related an experiment that was conducted by NASA. Three scientists lived on the floor of the ocean for about 1-3 months in a biosphere.

When they left they were all middle aged with graying hair and low libidos.

When they returned their hair was clear of gray, their wrinkles had started to disappear, and their sex drive was so increased that their wives complaied to NASA about it.

It turns out that certain glands and organs were "reactivated." One in particular was the gland that lies over the top of the heart. Blood tests showed unusual hormones. Hormones that are normally associated with the growth of young children.

f you check the Creation Research Center and look up the url for the Creation Research Institute you can then find the geologist who recreated the atmosphere of the flood in a container. I havent checked for his name yet. Been behind on my page and other projects, but I am having a pastor friend look into it for me.

What if we take this knowledge and build a "room" that emulates the atmosphere of the Earth before the flood and sleep in it for 8 hours a day? What effects might it have? My friend from the NSA said that for every day spent in there one year was added onto your life until you maxed out at 1000.

Chris

PATENTS

Closed ecological system and method for supporting life
US5935516

An apparatus to contain a closed ecological system, wherein the apparatus includes a closed, air-tight chamber to contain an atmosphere. The chamber includes at least one access opening designed to allow plants, animals, raw materials, and other items to be placed within the chamber. The chamber includes a system for introducing oxygen and/or carbon dioxide into the chamber and for maintaining the oxygen and/or carbon dioxide at a predetermined concentration percentage of the atmosphere. Preferably, the closed ecological system of the present invention also includes a system for establishing a magnetic field within the chamber of a predetermined strength and orientation. The chamber may also include systems for elevating the atmospheric pressure within the chamber and for maintaining the atmosphere within the chamber at a predetermined temperature and humidity level. The chamber may also include a specialized lighting and sound system to establish light and sound of a predetermined intensity, frequency, and wavelength within the chamber. A satellite chamber may be provided in selective fluid communication with the closed ecological system for providing plants, animals, or chemical processes with short-term exposure to the conditions of the closed ecological system. The present invention is also directed to a method of establishing and maintaining a closed ecological system to support plant and animal life and to aid in the development, discovery, and/or production of new and improved chemical compounds that may be used as pharmaceuticals and other chemical compounds.

FIELD OF INVENTION

The present invention relates generally to closed ecological systems, and more particularly to a method and apparatus for establishing and maintaining a closed ecological system wherein the environment within the closed system is controlled such that plant and animal life can exist to its more optimal potential and thrive, the effects and causes of plant and animal disease can be reduced and/or eliminated, mental well being of animals can be fostered, and new and improved pharmaceuticals and other chemical compounds (organic and inorganic) can be discovered and/or synthesized.

BACKGROUND OF THE INVENTION

The environment in which we live with plants and other animals has changed since biological systems first appeared. The composition of the air we breathe has continued to change. For example, there is reason to believe that the concentration of oxygen found within the air has decreased over time. The climate, including temperature, humidity, rainfall, snowfall, and the like has also been changing. It is further theorized that other important changes in the environment of the earth over time have been the marked decrease in atmospheric pressure, a decrease in the concentration of carbon dioxide in the earth's atmosphere, changes in the strength and orientation of the earth's magnetic field, and the change in the intensity and characteristics of the light radiation to which we are exposed. Furthermore, sounds of nature such as singing birds have become increasingly difficult to encounter due to a general decrease in the population of songbirds.

Many of the above and other changes are a natural result of the maturing of the earth. Other changes to the earth's environment are caused by purely natural events--volcanic eruptions, earthquakes, tidal waves, glaciers, and the like alter the earth itself and the surrounding atmosphere. Other changes to the earth and its atmosphere are due entirely to man. Pollution, overpopulation, overdevelopment, overutilization of natural resources, fishing, hunting, and farming have all altered our world. As the human population continues to increase, the pace at which these changes occur will inevitably increase.

These changes in the earth and the surrounding environment have had a detrimental impact on plants and animals including man. Certain plants and animals have vanished from the earth, unable to adapt to the changed earth. Many of the remaining varieties of plants and animals have found it difficult to thrive. The fruits and vegetables of trees and plants have lost taste and nutritional value. Animals, including humans, are increasingly disease stricken, weak, and otherwise unhealthy. Humans seem to have lost their general sense of well being, leading to an increase in depression, suicides, crime, violence, sickness, and other social and physical ailments and have become increasingly lethargic with shortened attention spans.

The increasing physical and mental ailments of humans has led to a constant search for new and improved pharmaceuticals to combat these sicknesses. As one example, scientists are constantly searching the far reaches of the planet for naturally occurring antiviral and antibiotic substances. Once such naturally occurring substances are discovered, scientists may attempt to synthesize these drugs in a laboratory. While much progress has been made in the field of locating and synthesizing these compounds for use as pharmaceuticals, many ailments currently have no known effective treatment drugs or the causative agents have been able to adapt or modify their own structure to develop a resistance to such treatments. For example, many cancers and AIDS have been particularly resistant to pharmaceuticals located and/or synthesized using conventional methods in standard environmental conditions. It is believed that the changes in the earth's environment have prevented certain drugs from being created in nature, discovered, and/or synthesized, and has prevented the compounding of certain organic and inorganic elements.

SUMMARY OF THE INVENTION

It is therefore an object of the present invention to solve the above and other problems caused by changes in the earth and its surrounding environment.

It is also an object of the present invention to provide a closed ecological system including an atmosphere with an elevated oxygen content.

It is another object of the present invention to provide a closed ecological system including an atmosphere with an elevated atmospheric pressure.

It is another object of the present invention to provide a closed ecological system with an atmosphere having an elevated concentration of carbon dioxide.

It is a further object of the present invention to provide a closed ecological system including a low level magnetic field.

Yet another object of the present invention is to provide a closed ecological system including specialized lighting and sound.

Still a further object of the present invention is to provide a closed ecological system with a controlled temperature, humidity, and other climate variables.

Another object of the present invention is to provide a closed ecological system having an overall environment that is favorable to the natural formation and/or the manufacture of improved compounds used to manufacture pharmaceuticals for the treatment of sickness and disease.

The present invention accomplishes these and other objects by providing an apparatus to contain a closed ecological system, wherein the apparatus comprises a closed, air-tight chamber to contain an atmosphere, and wherein the chamber includes at least one access opening. The access opening is designed to allow plants, animals, raw materials, and other items to be placed within the chamber. The chamber includes means for introducing oxygen and/or carbon dioxide into the chamber and means for maintaining the oxygen and/or carbon dioxide at a predetermined concentration percentage of the atmosphere. Preferably, the closed ecological system of the present invention also includes means for establishing a magnetic field within the chamber of a predetermined strength and orientation. The chamber may also include means for elevating the atmospheric pressure within the chamber and means for maintaining the atmosphere within the chamber at a predetermined temperature and humidity level. The chamber may also include a specialized lighting and sound system to establish light and sound of a predetermined intensity, frequency, and wavelength within the chamber. A satellite chamber may be provided in selective fluid communication with the closed ecological system for providing plants, animals, or chemical processes with short-term exposure to the conditions of the closed ecological system.

The present invention is also directed to a method of establishing and maintaining a closed ecological system to support plant and animal life and to aid in the development, discovery, and/or production of new and improved chemical compounds that may be used as pharmaceuticals.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a front elevational view of a closed ecological system in accordance with the present invention;

FIG. 2 is a side elevational view of a closed ecological system in accordance with the present invention;

FIG. 3 is a side elevational view of a closed ecological system in accordance with the present invention;

FIG. 4 is a rear elevational view of a closed ecological system in accordance with the present invention;

FIG. 5 is a schematic representation of a closed ecological system in accordance with the present invention;

FIG. 6 is a schematic representation of a closed ecological system including a lighting system in accordance with the present invention;

FIG. 7 is a schematic view of a closed ecological system including a magnetic field generation system in accordance with the present invention;

FIG. 8 is a schematic view of an audio system for use with a closed ecological system in accordance with the present invention;

FIG. 9 is a scanning electron microscope view of Copperhead snake venom taken from a Copperhead snake living under normal, ambient conditions on earth;

FIG. 10 is a scanning electron microscope view of Copperhead snake venom taken from a Copperhead snake that lived for four weeks in a closed ecological system in accordance with the present invention.

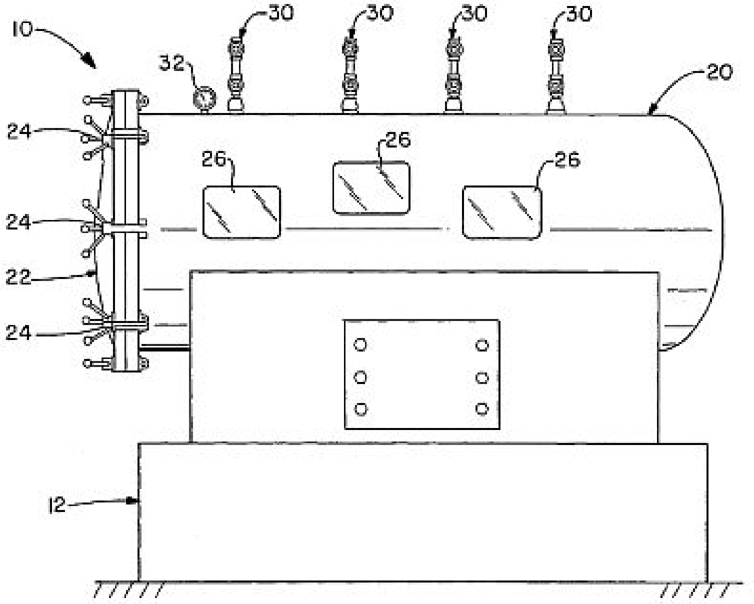
10
32
30
30
30
30
20
24
26
26
26
24
22
24
12

FIG.-1

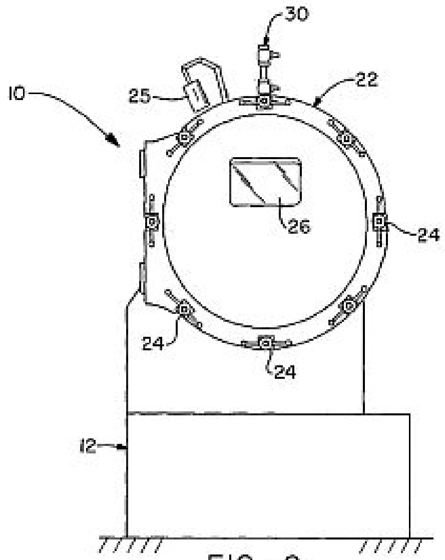
30
22
10
25
26
24
24
24
12

FIG.-2

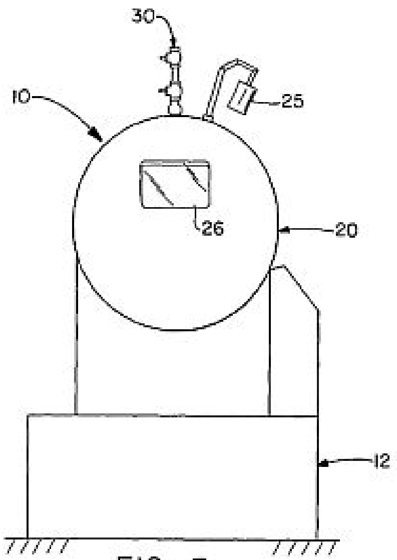
30
10
25
26
20
12

FIG.-3

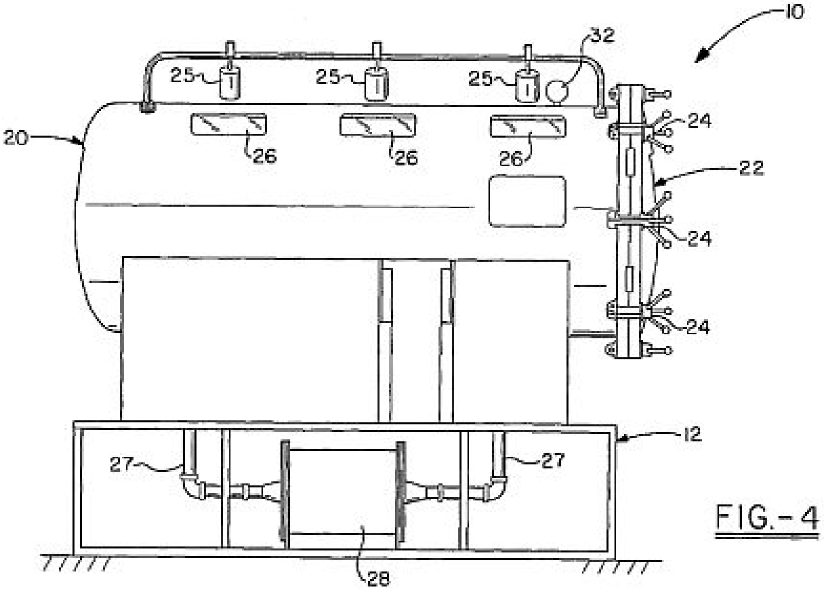
10
32
25
25
25
20
26
26
26
24
22
24
24
12
27
27
28

FIG.-4

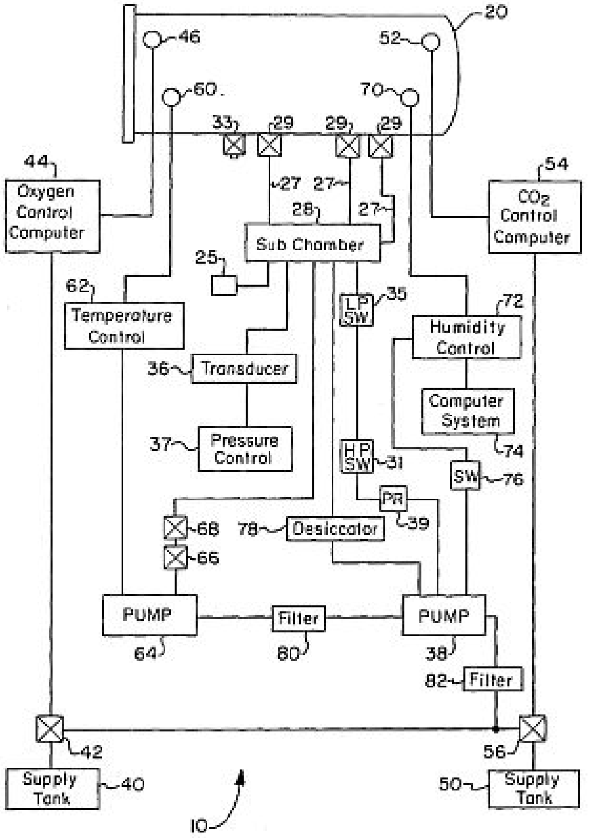
20
46
52
60
70
33
29
29
29
44
Oxygen Control Computer
54
CO$_2$ Control Computer
27
27
27
28
Sub Chamber
25
62
Temperature Control
LP SW
35
72
Humidity Control
36
Transducer
Computer System
74
37
Pressure Control
HP SW
31
SW
76
PR
39
68
78
Desiccator
66
PUMP
Filter
PUMP
64
80
38
82
Filter
42
56
Supply Tank
40
50
Supply Tank
10

FIG.-5

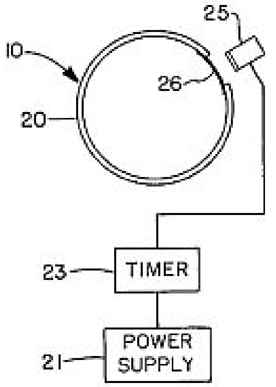
25
10
26
20
23
TIMER
21
POWER SUPPLY

FIG.-6

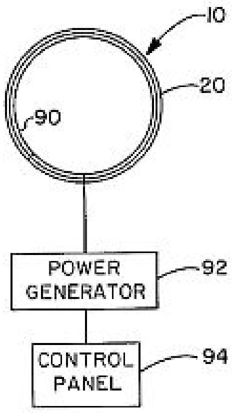
10
20
90
POWER GENERATOR
92
CONTROL PANEL
94

FIG.-7

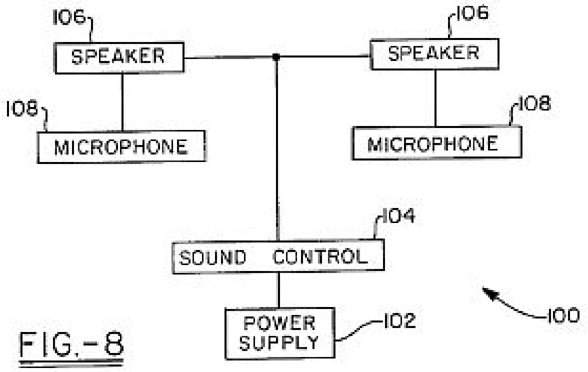
106
SPEAKER
106
SPEAKER
108
MICROPHONE
108
MICROPHONE
104
SOUND CONTROL
100
POWER SUPPLY
102

FIG.-8

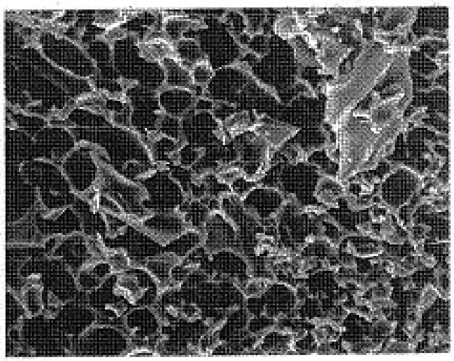

FIG.-9

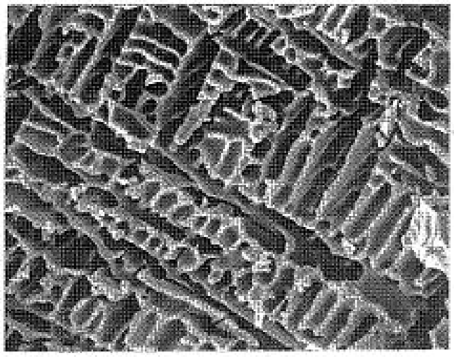

FIG.-10

DETAILED DESCRIPTION OF THE INVENTION

Reference will now be made in detail to the present preferred embodiments of the invention, examples of which are illustrated in the accompanying drawings. Wherever possible, the same reference numbers will be used throughout the drawings to refer to the same or like parts.

An apparatus for establishing, containing, and maintaining a closed ecological system in accordance with the present invention is shown generally at 10 in FIGS. 1-7 and comprises an air-tight, substantially hollow chamber 20 made of plastic, metal, fiberglass, or any other suitable material. Chamber 20 includes a sealable access opening such as hinged door 22 provided with suitable air locks 24 and seals (not shown) to ensure that door 22 does not reduce the air-tight integrity of the chamber 20. Chamber 20 is designed to contain therein an atmosphere, plants, animals, objects and any other desired components of a closed ecological system such that plants and animals may live therein and such that chemical reactions such as the production of pharmaceuticals can be carried out therein. Chamber 20 may be of any size desired such that the ecological system established and maintained therein may be of any suitable size. Chamber 20 may only be large enough to contain a few small plants and animals, or chamber 20 may be large enough to contain a community of numerous human beings. As shown, chamber 20 is sufficiently large to contain at least one human being and several plants and animals. Chamber 20 may include air-tight windows 26 formed therein or in door 22 thereof using suitable glass or plastic to allow those outside of the chamber 20 to observe the plants, animals, and any other animals, objects and/or conditions inside of chamber 20. Door 22 may be opened to remove, add, or reorient specimens as desired, after depressurization has returned internal pressure to ambient conditions. Windows 26 are designed to filter ultra-violet (UV) light to prevent unwanted UV light from affecting the ecological system contained within chamber 20. Specifically, UV light causes or encourages the formation of "free radicals" that are thought to be harmful to animals.

Chamber 20 is supported upon a base 12 to prevent the unwanted movement of the chamber 20 and other components of the apparatus 10. A sub-chamber or premixing chamber 28 (FIGS. 4 and 5) is preferably provided and is in selective fluid communication with chamber 20 through pipes or other conduits 27. Sub-chamber 28 provides a means for premixing the gases and liquids to be added to chamber 20 to ensure that such gases and liquids are uniformly distributed within chamber 20. It should be recognized that chamber 28 may also function as a satellite chamber in selective fluid communication with chamber 20 for establishing and maintaining a closed ecological system similar to that found within chamber 20, such that plants, animals, chemical reactions and the like can be placed within the closed ecological system found within the satellite chamber for time periods that differ (usually shorter) from the amount of time chamber 20 needs to remain closed. To function as a satellite chamber, sub-chamber 28 would need to include an access opening such as a door, and would need to include at least some of the features of chamber 20 as they are set forth below. Also, if sub-chamber 28 is to be used as a satellite chamber, chamber 20 and sub-chamber must be shielded from each other with a lead shield or the like to prevent the magnetic field (discussed fully below) of one chamber 20, 28 from interfering with the magnetic field of the other chamber 20, 28. For example, in addition to conducting experiments relating to the natural creation and/or synthesis of pharmaceuticals within chamber 20, sub-chamber 28, when properly adapted as a satellite chamber, may alternatively be utilized for the same types of experiments. Various chemicals could be mixed and placed within sub-chamber 28 for a period of time, and thereafter be removed from sub-chamber 28 and examined, without opening door 22 of chamber 20 and without altering the ecological system found within chamber 20.

Conduits 27 between sub-chamber 28 and chamber 20 are selectively blocked or closed using gate valves 29 (FIG. 5) or the like so that sub-chamber 28 may be selectively "disconnected" or prevented from being in fluid communication with chamber 20 to protect the integrity of the ecological system found within chamber 20 during premixing operations. After sub-chamber 28 is "disconnected" from chamber 20, various gases and other components of the closed ecological system as described herein may be introduced into chamber 28, without altering the ecological system within chamber 20. When the gases and liquids introduced into sub-chamber 28 are properly mixed, which may occur naturally or with the aid of mixing devices such as blowers, gate valves 29 may be opened to allow the properly mixed gases and liquids to enter chamber 20 as is described below. Sub-chamber 28 includes a pressure relief valve 25 so that excess pressure may be vented from sub-chamber 28 if necessary. Those skilled in the art will recognize that the apparatus need not include a sub-chamber 28 and such is contemplated herein, although use of sub-chamber 28 is thought to be preferable. Also, those skilled in the art will recognize that any ecological system or conditions created and maintained in chamber 20 may also be created and maintained in sub-chamber 28, and for reasons of clarity, the following description will not necessarily refer to providing an ecological system within both chambers 20,28 but will instead refer primarily to chamber 20.

Apparatus 10 includes numerous components and features such that an ecological system can be created and maintained within chamber 20 and so that the characteristics of the ecological system so created and maintained, such as temperature, atmospheric pressure, humidity, magnetic field, oxygen level, carbon dioxide level, and the like may be accurately varied and monitored. One or more feeder valves such as dual ball valves 30 are provided in fluid communication with the interior of chamber 20 and are also connected to vacuum/pressure pump 38 (FIG. 5) to allow fluids such as air, oxygen, water, nutrients and others (organic and inorganic) to be communicated from the appropriate source into the ecological system found within chamber 20 through one or more valves 30. As is noted above, it is thought preferable to premix fluids in sub-chamber 28 prior to introducing the fluids into chamber 20. Feeder lines or conduits may be surrounded by a coolant or other thermal fluid to vary the temperature of the fluid in the feeder line. Because of the potentially elevated atmospheric pressure (hyperbaric pressure) within chamber 20, dual ball valves 30 are thought to provide a more effective seal to prevent the escape of gases from the chamber 20 and to otherwise protect the integrity of the closed ecological system created and maintained by the apparatus 10, although other suitable valves are known and are contemplated for use in conjunction with the present invention 10. A pressure gauge 32 is also provided to allow those outside of chamber 20 to monitor the atmospheric pressure within chamber 20.

The pressure of the atmosphere within chamber 20 may be varied as desired to provide an elevated atmospheric pressure. It is thought preferable for example to create a hyperbaric closed ecological system--i.e., a closed ecological system having an increased atmospheric pressure relative to the atmospheric pressure normally encountered on earth. It is thought most preferable to provide an environment inside of chamber 20 having an atmospheric pressure of 1 to 2 atmospheres of pressure. FIG. 5 shows a pressure sensor or transducer 36 coupled to a pressure control mechanism such as pressure control computer 37 to control and regulate the pressure inside of chamber 20 and/or 28. Transducer 36 converts air pressure into electrical signals to provide input data to pressure control computer 37 so that computer can continually monitor the air pressure within chamber 20 and adjust the same accordingly. A vacuum/pressure pump 38 is provided in fluid communication with chamber 20 so that the pressure in chamber 20 can be elevated using compressed air generated by pump 38 and transferred into chamber 20 through one or more of the valves 30. Pump 38 may also be used to create a vacuum suction to draw air from chamber 20 resulting in a low pressure environment within chamber 20. As a safety and control measure, a high pressure safety switch 35 is provided to prevent overpressurization of the chamber 20. Likewise, a low pressure safety switch 31 is provided to prevent an excessive vacuum condition from being established within chamber 20. A pressure relay 39 is provided to selectively energize vacuum/pressure pump 38 as needed to increase or decrease the atmospheric pressure in chamber 20 and a blow off valve 33 is provided in fluid communication with chamber 20 to allow excessive air pressure that may be contained therein to be rapidly vented automatically when the atmospheric pressure within chamber 20 exceeds some threshold value such as 2 atmospheres.

In addition to an elevated atmospheric pressure, the closed ecological system of the present invention 10 preferably includes an atmosphere within chamber 20 having an elevated oxygen concentration relative to the concentration of oxygen normally found in the air surrounding the earth. Specifically, it is thought preferable to maintain a concentration of oxygen within chamber 20 such that the oxygen is equal to 22%-30% by volume of the atmosphere contained within chamber 20 when the atmospheric pressure is in the range of 1-2 atmospheres. For long term occupation within chamber 20 the oxygen level is most preferably maintained at 22%-24% by volume at 2 atmospheres of pressure, while for short term occupation, a month or less, the oxygen level within chamber 20 may be elevated to 26%-30% by volume at 2 atmospheres of pressure.

A source of oxygen such as a pressurized holding tank 40 or the like is provided to supply oxygen to the system 10. Alternatively, oxygen may be generated by an electrolysis process. Oxygen supply tank 40 is in fluid communication with vacuum/pressure pump 38 and chamber 20 using the appropriate conduits and connections. A valve assembly such as oxygen solenoid valve 42 is controlled by a control mechanism such as oxygen control computer 44 which is provided between supply tank 40 and pump 38 to regulate the flow of oxygen to pump 38 and into chamber 20 through one or more valves 30 (FIG. 1). An oxygen sensor 46 is provided in one or both chambers 20, 28 to sense the concentration of oxygen therein. Sensor 46 provides the concentration to oxygen control computer 44 as input so that computer 44 may determine if oxygen solenoid valve 42 should remain closed or should be opened to allow more oxygen to enter chambers 20, 28. A filter 82 is preferably provided between oxygen supply tank 40 and pump 38 to filter any contaminants from the oxygen.

By combining an increased level of oxygen concentration with a hyperbaric atmospheric pressure inside of chamber 20, the amount of oxygen absorbed by animals within the chamber 20 is maximized. This increased absorption of oxygen by the animals, including humans, leads to increased physical healing through the increased oxygen absorption of the cells of the animal, increased immunity, an increased level of mental awareness, and a general sense of well being for the animals that spend time within the closed ecological system created and maintained by apparatus 10. The increased absorption of oxygen by the animal cells raises the internal temperature of the cells, resulting in the damage to cancerous cells which have been found to have a lower resistance to heat than ordinary cells. As is discussed below in relation to FIGS. 9 and 10, the increased concentration of oxygen combined with an elevated atmospheric pressure, alone or in combination with other aspects of the closed ecological system found within chamber 20, has been found to result in the natural production of a superior snake venom having medicinal properties. Also, it is believed that the in increased oxygen level and atmospheric pressure contributes to improving the reactions among various chemicals which can lead to the production of superior chemical compounds such as very effective pharmaceuticals that may be used to treat illnesses for which there are currently few or no effective treatment drugs.

In conjunction with increasing the concentration of oxygen in chamber 20, it may also be desirable in certain circumstances to increase the concentration of carbon dioxide (CO2) in chamber 20. It has been theorized that an increased concentration of carbon dioxide lengthens the gestation period for mammals. This may be desirable to prevent premature births (births before the end of the normal gestation period for any particular mammal) or to lengthen the gestation period beyond normal to increase the development of the fetus. Carbon dioxide normally comprises approximately 0.026% by volume of the air normally encountered by mammals on earth. The apparatus 10 allows a closed ecological system to be created wherein the concentration of carbon dioxide therein may be elevated to approximately 0.1%-0.3% by volume. This increased level of carbon dioxide, in conjunction with an elevated atmospheric pressure of approximately 2 atmospheres leads to the lengthened gestation period for mammals. It has also been found that raising the level of carbon dioxide assimilation in plants enhances plant production of stalks, leaves, and fruit.

As is shown in FIG. 5, a supply of carbon dioxide such as a pressurized carbon dioxide supply tank 50 or the like is provided in fluid communication with pump 38, through filter 82 designed to remove contaminants from the carbon dioxide. Pump 38 is in fluid communication with chamber 20 through one or more dual ball valves 30. A carbon dioxide sensor 52 is capable of sensing the concentration of carbon dioxide in the chambers 20, 28. A control mechanism such as carbon dioxide control computer 54 receives the concentration of carbon dioxide from sensor 52 as input and determines whether to increase the concentration thereof. If more carbon dioxide is needed within chamber 20, carbon dioxide control computer 54 causes a valve assembly such as carbon dioxide solenoid valve 56 to open, thereby allowing carbon dioxide to flow from supply tank 50 to pump 38 and thereafter into chamber 20 through one or more valves 30.

The apparatus 10 of the present invention also includes systems to control the temperature and humidity of the air within chamber 20 and also includes an air filtration system to remove contaminants from the air. A temperature sensor 60 is provided within chamber 20 to sense the temperature therein. Sensor 60 is coupled to a temperature control unit 62 which may be a computer or a simple thermostat, which is coupled to a heating and cooling pump 64. Heating and cooling pump 64 is capable of adding heat to or removing heat from chambers 20, 28 as desired to control the temperature therein. Heating and cooling pump 64 is coupled to chambers 20,28 through a check valve 68 to sub-chamber 28 and main chamber 20. Preferably, the temperature inside of chambers 20 and 28 is maintained in the range of 70 DEG -80 DEG Fahrenheit (F.) at all times. The temperature may be decreased to approximately 70 DEG F. at "night" and maintained at approximately 80 DEG F. during the "day." Of course, the terms "day" and "night" do not necessarily refer to the time of day outside of the chambers 20, 28 although "day" and "night" within the chambers 20, 28 may be controlled to correspond to conditions outside of the chambers 20, 28. Electric lights 25 (FIGS. 4 and 6) are provided to shine light into the chamber 20 through windows 26. Lights 25 receive power from a power supply 21 and are also connected to a control box/timer 23 that allows an operator of the apparatus 10 to vary the intensity, duration, and wavelength of each light 25. Lights 25 are preferably filtered to provide a wavelength primarily in the range of 6365 Angstroms which results in a magenta light. Therefore, "day" may occur anytime lights 25 are on while "night" may occur anytime no light is shining through windows 26 because of darkness surrounding chamber 20 or because of shades or the like being placed over windows 26.

The humidity within chambers 20, 28 is also closely monitored and is maintained in the range of 40%-60%. A humidity sensor 70 is provided within chamber 20 to sense the percentage of moisture found in the atmosphere within chamber 20 and is coupled to a humidity control unit 72 which may include a gravity collection/injection valve apparatus or any other suitable humidity control apparatus. Humidity control unit 72 is coupled to a computer system 74 that is capable of controlling the humidity control unit 72 to maintain the humidity within chambers 20, 28 at the desired 40%-60%. Humidity control unit 72 and computer system 74 are coupled through a switch 76 to vacuum/pressure pump 38 such that pump 38 may be selectively energized through switch 76. Pump 38 is coupled to a desiccator unit 78 which is in fluid communication with chambers 20, 28. Upon being energized through switch 76, pump draws air from within chambers 20, 28, through desiccator unit 78 wherein moisture is removed from the air. The dehumidified air is then returned into chamber 20 through one or more valves 30. Because of the closed nature of closed ecological system created and maintained in apparatus 10, it will not normally be necessary to add humidity to chambers 20, 28, although such is contemplated and may be accomplished by the addition of a humidifier to the apparatus 10.

Apparatus 10 is also provided with an air filtration system comprising an air filter 80 provided in fluid communication with vacuum/pressure pump 38 and heating and cooling pump 64 such that any air introduced into chambers 20, 28 using vacuum/pressure pump 38 and/or heating and cooling pump 64 is first filtered of any airborne contaminants such as dust and bacteria. Both filter 80, and filter 82 discussed previously, may be a conventional filter media such as a foam or mesh, may be an electrostatic ionization filtration system that electrically charges the contaminants so that they are attracted to either a positively charged or negatively charged surface, or the air filters 80,82 may be a combination thereof or any other suitable air filtering system.

The apparatus 10 further comprises means for establishing and maintaining a magnetic field within the chamber 20 of a predetermined strength and orientation. As with other aspects of the apparatus and the closed ecological system created and maintained thereby, a magnetic field may also be established within chamber 28 although for reasons of clarity, reference will only be made to chamber 20. Referring to FIG. 7 wherein the apparatus 10 is shown schematically, a magnetic field coil 90 comprising shielded copper wire or the like surrounds and/or lines chamber 20 and is coupled by wires or the like to a D/D power generator 92 which generates controlled electromagnetic pulses (Hz) at optimal energy levels (Gauss). D/D power generator 92 is connected to a control mechanism such as control panel 94 which may be an analog device or a digital computer allowing a machine operator to vary the strength and orientation of the magnetic field created by the magnetic field coil 90 by varying the magnitude and direction of the electric current flowing through magnetic field coil 90.

It is thought preferable to establish and maintain a magnetic field within chamber 20 of 0.5 Hz to 30 Hz at 1-5 Gauss and is thought most preferable to average a magnetic field of 8 Hz within chamber 20. A variable magnetic field of 8 Hz (average, with intensity nearing 5 Gauss) is thought to simulate the magnetic field found on earth thousands of years ago. The magnetic field may be cyclical--i.e., turned on and off, or raised and lowered depending upon other aspects of the closed ecological system. For example, the magnetic field may be elevated only in the "day" or "night" as such terms are defined above. Alternatively or in conjunction with the "day and "night" variables, the magnetic field may be dependent upon the song activity of live birds present within the chamber 20 or dependent upon the sounds of birds that are broadcast within the chamber 20 through a sound system as is discussed in greater detail below.

FIG. 8 schematically illustrates a sound system 100 to be used in conjunction with the apparatus 10 of the present invention to establish predetermined sounds within the chamber 20. Sound system 100 comprises a power supply 102 which may be the same as power supply 21 discussed previously, or power supply 102 may be separate therefrom. Power supply 102 is connected to and provides electrical power to a sound control panel 104 from where one may control the selection and characteristics of the sounds broadcast by the sound system 100 into chamber 20. Control panel 104 is connected to one or more speakers 106 for broadcasting sound and is also connected to one or more microphones 108 for receiving sound from within chamber 20 and transmitting the same onto a recording medium or for broadcasting such sounds outside of chamber 20. Control panel 104 may be located either inside or outside of chamber 20 such that sound system 100 may be controlled by a person outside of chamber 20 or inside chamber 20, depending upon the location of control panel 104. It has been found beneficial for animals and plants to be within chamber 20 while sound from sound system 100 is broadcast into chamber 20. For example, harmonic music corresponding to songs of birds within chamber 20 at controlled times and under controlled conditions has been found to be particularly beneficial to plants and animals within chamber 20. By broadcasting the songs of birds where the songs are tuned to a frequency of 256 Hz ranging from 2-60 decibels in intensity, the sounds of the harmonic music will be tuned to the antenna of the DNA and cellular structure of the cells of animals which thereby causes a resonance of such cells and alters the vibratory cycle of the cells, leading to the improved health thereof, and leading to the deterioration of abnormal cells. As an alternative to broadcasting songs of birds into chamber 20, symphonic music, broadcast into chamber 20 in accordance with the above-noted broadcast parameters, may alternatively be utilized with the same or similar therapeutic effects.

As mentioned previously, the apparatus 10 allows for the creation of a closed ecological system in accordance with the above, thereby creating an environment conducive to the manufacture of exceptional pharmaceuticals for treating diseases and ailments for which safe and effective pharmaceuticals have not yet been discovered. Such exceptional pharmaceuticals may be obtained through natural organic production of living systems themselves or may be obtained by mixing various chemicals. For example, referring now to FIGS. 9 and 10, there can be seen a scanning electron microscope view of snake venom obtained from a Copperhead snake living under ordinary conditions found on earth (FIG. 9) and a scanning electron microscope view of snake venom obtained from the same snake, after the snake lived in an apparatus 10 housing a closed ecological system as is set forth above. Those skilled in the art will recognize the markedly improved characteristics of the venom shown in FIG. 10 obtained from the Copperhead snake that inhabited the closed ecological system of the present apparatus 10 for four weeks. Snake venom shown in FIG. 10 has been experimented with and found to have medicinal properties such as decreasing the size and occurrence of cancerous tumors, and providing relief from emotional illnesses. It is theorized that the toxicity of snake venom in the present invention would be lowered, if not eliminated, and the coherent structured chemical formation as shown would lead to immediate assimilation into patient receptors.

Those skilled in the art will recognize that the apparatus 10 may further comprise conveniences and necessaries to sustain life such as water nutrients, waste facilities and the like. Also, while the foregoing description has set forth the preferred embodiment of the invention in particular detail, it must be understood that numerous modifications, substitutions, and changes may be undertaken without departing from the true spirit and scope of the present invention as defined by the ensuing claims.

METHOD AND APPARATUS TO STIMULATE THE IMMUNE SYSTEM OF A BIOLOGICAL ENTITY

US7338431

A system and method for stimulating the immune systems of biological entities in an environment. Pulsed electrical currents are generated using an electric current generator. The pulsed electrical currents are fed through an arrangement of electrically conductive material such that magnetic energy is emitted from the arrangement into the environment. The arrangement of electrically conductive material is designed such that an intensity of the emitted magnetic energy varies across at least one spatial dimension of the environment. Certain embodiments of the arrangement, which have width-to-length ratios of approximately 0.6, tend to provide variations in the intensity of the magnetic energy field which are very good for stimulating the immune systems of the biological entities.

TECHNICAL FIELD

Certain embodiments of the present invention relate to stimulating the immune system of biological entities. More particularly, certain embodiments of the present invention relate to a system and method to stimulate the immune system of biological entities moving in an environment through application of pulsed magnetic energy.

BACKGROUND OF THE INVENTION

Use of magnetic energy to increase physiological performance of organisms has long been attempted. However, many of these techniques have been limited to belts, pads or mats which apply magnetic or electromagnetic energy to the person or other organism. Problems inherent in these techniques include the necessity for the organism to wear the belt or pad, and the necessity for a portable power source in order to generate magnetic energy. Furthermore, these techniques do not effect the environment surrounding the organism. Accordingly, there is a demand for an apparatus and method of applying pulsed magnetic energy to an organism (i.e., a biological entity) and its surrounding environment that is without the aforementioned disadvantages.

Further limitations and disadvantages of conventional, traditional, and proposed approaches will become apparent to one of skill in the art, through comparison of such systems and methods with the present invention as set forth in the remainder of the present application with reference to the drawings.

BRIEF SUMMARY OF THE INVENTION

An embodiment of the present invention comprises a system for stimulating immune systems of living biological entities in an environment. The system comprises at least one electric current generator providing a source of pulsed electrical current. The system further comprises at least one continuous coil of electrically conductive material having a first end and a second end, both of the ends being connected to the at least one generator to form a closed circuit such that the at least one coil emits a spatially non-uniform pulsed magnetic field into the environment in response to the pulsed electrical current to stimulate the immune systems as the biological entities move within the environment. Also, a configuration of the at least one coil comprises a plurality of turns of the conductive material in substantially a single spatial plane, and wherein the coil has an overall width-to-length ratio of between 0.4 and 0.8.

Another embodiment of the present invention comprises a system for stimulating immune systems of biological entities in an environment. The system comprises at least one electric current generator providing a source of pulsed electrical current. The system further comprises at least one arrangement of electrically conductive material having a first end and a second end, both of the ends being connected to the at least one generator to form a closed circuit such that the at least one arrangement emits a spatially non-uniform pulsed magnetic field into the environment in response to the pulsed electrical current to stimulate the immune systems as the biological entities move within the environment. Also, a configuration of the at least one arrangement comprises a plurality of substantially parallel segments of the conductive material forming a flat, substantially rectangular grid having an overall width-to-length ratio of between 0.4 and 0.8.

A further embodiment of the present invention comprises a method for stimulating immune systems of living biological entities in an environment. The method comprises positioning at least one arrangement of electrically conductive material below a surface of the environment and connecting the at least one arrangement of electrically conductive material to at least one electric current generator to form a closed circuit through the arrangement. The method further comprises generating a pulsed electrical current with the generator such that the pulsed electrical current propagates through the arrangement from a first end of the arrangement to a second end of the arrangement. The arrangement emits pulsed magnetic energy into the environment in response to the pulsed electrical current such that an intensity of the pulsed magnetic energy is non-uniform across at least one spatial dimension of the arrangement to stimulate the immune systems as the biological entities move within the environment.

Another embodiment of the present invention includes a system for stimulating the immune systems of biological entities in an environment. The system comprises at least one electric current generator providing a source of pulsed electrical current. The system further comprises at least one continuous coil of electrically conductive material having a first end and a second end where both ends are connected to the generator to form a closed circuit such that the coil emits a spatially non-uniform pulsed magnetic field into the environment in response to the pulsed electrical current to stimulate the immune systems as the biological entities move within the environment. A configuration of the coil includes a plurality of parallel straight segments of the conductive material, being substantially of the same length, and a plurality of curved segments of the conductive material. The continuous coil spirals outward from a central position of the coil in substantially a single spatial plane.

A further embodiment of the present invention includes a system for stimulating the immune systems of biological entities in an environment. The system comprises at least one electric current generator providing a source of pulsed electrical current. The system further comprises at least one continuous coil of electrically conductive material having a first end and a second end where both ends are connected to the generator to form a closed circuit such that the coil emits a spatially non-uniform pulsed magnetic field into the environment in response to the pulsed electrical current to stimulate the immune systems as the biological entities move within the environment. A configuration of the coil comprises a first plurality of parallel straight segments of the conductive material and a second plurality of parallel straight segments of the conductive material being substantially perpendicular to the first plurality of segments. The continuous coil winds outward from a central position of the coil in substantially a single spatial plane.

These and other advantages and novel features of the present invention, as well as details of illustrated embodiments thereof, will be more fully understood from the following description and drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 illustrates a first exemplary embodiment of a system for applying pulsed magnetic energy to an aquatic environment, in accordance with various aspects of the present invention.

FIG. 2 illustrates a first exemplary embodiment of a flat coil used to generate magnetic energy, in accordance with various aspects of the present invention.

FIGS. 3A-3C illustrate various views of an embodiment of the flat coil of FIG. 2, clearly showing the spacing between the coil turns, in accordance with various aspects of the present invention.

FIG. 4 illustrates an exemplary simulated graph of how a magnetic field intensity generated by the coil of FIG. 2 may be expected to vary non-linearly across a spatial dimension of the coil of FIG. 2, in accordance with various aspects of the present invention.

FIG. 5 illustrates an exemplary graph of measured data of how a magnetic field intensity generated by the coil of FIGS. 3A-3C varies across three spatial dimensions of the coil of FIGS. 3A-3C, in accordance with various aspects of the present invention.

FIG. 6 is a flowchart of an embodiment of a method to stimulate immune systems of biological entities in an environment, in accordance with various aspects of the present invention.

FIG. 7 illustrates a second exemplary embodiment of a flat coil used to generate magnetic energy, in accordance with various aspects of the present invention.

FIG. 8 illustrates a second exemplary embodiment of a system for applying pulsed magnetic energy to a stock pen environment, in accordance with various aspects of the present invention.

FIG. 9 illustrates a third exemplary embodiment of a system for applying pulsed magnetic energy to a garden environment, in accordance with various aspects of the present invention.

FIG. 10 illustrates a fourth exemplary embodiment of a system for applying pulsed magnetic energy to a sports environment, in accordance with various aspects of the present invention.

FIG. 11 illustrates a fifth exemplary embodiment of a system for applying pulsed magnetic energy to a golf course environment, in accordance with various aspects of the present invention.

FIG. 12 illustrates exemplary resultant current pulses that may be produced in the coil of FIG. 2 when applying an exemplary DC pulsed voltage waveform to the coil of FIG. 2, in accordance with an embodiment of the present invention.

FIG. 13 illustrates exemplary resultant current pulses produced in the coil of FIGS. 3A-3C when applying an exemplary DC pulsed voltage waveform to the coil of FIGS. 3A-3C, in accordance with an embodiment of the present invention.

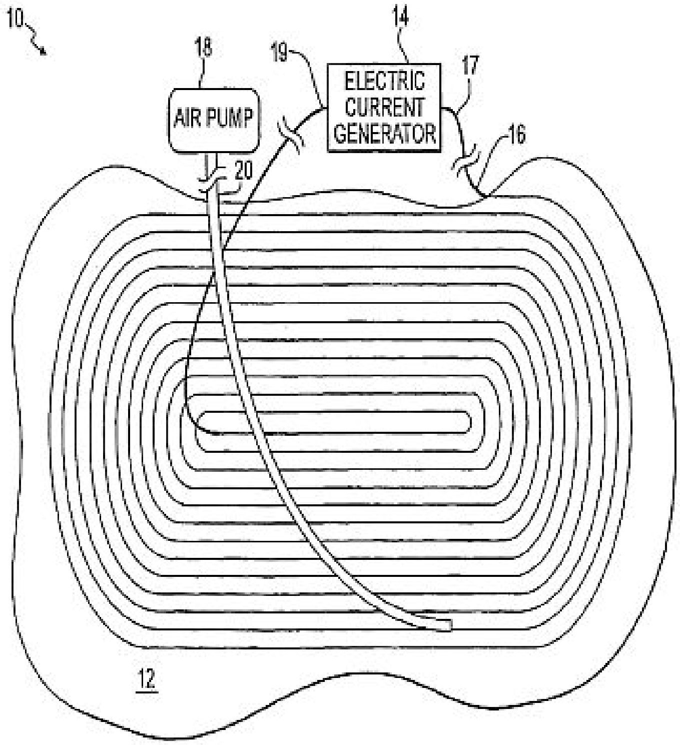
AIR PUMP
ELECTRIC CURRENT GENERATOR

FIG. 1

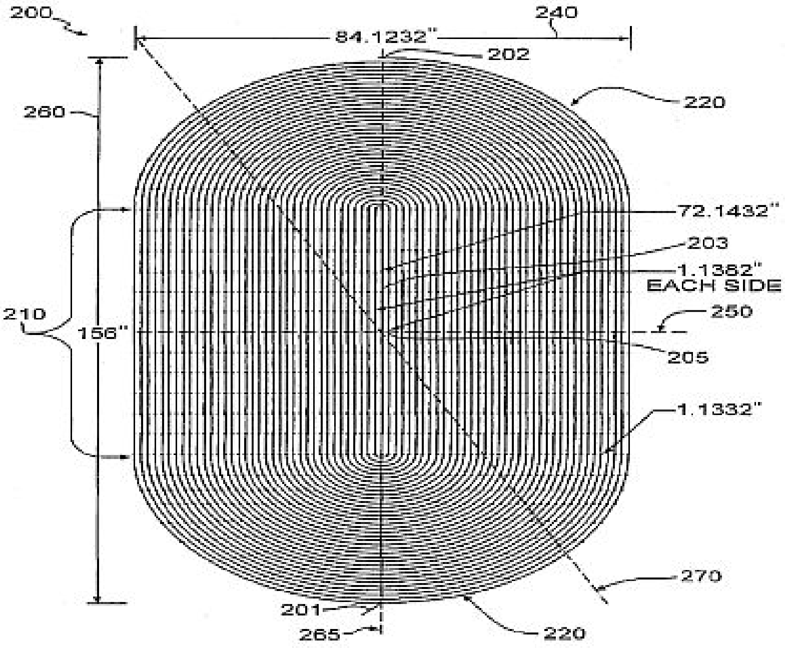
84.1232"
72.1432"
1.1382" EACH SIDE
156"
1.1332"

FIG. 2

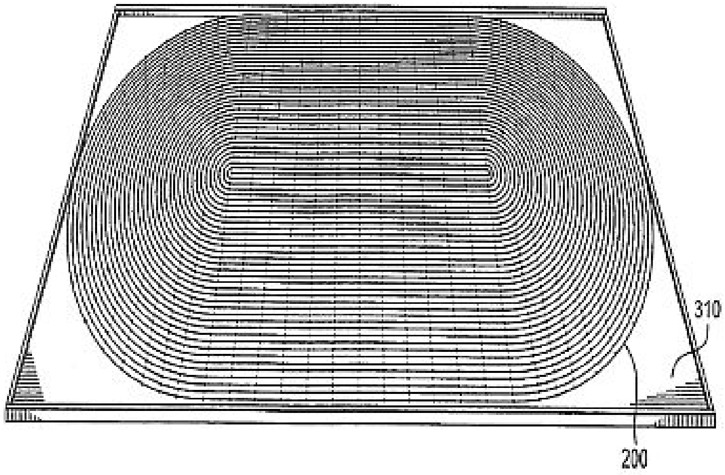

FIG. 3A

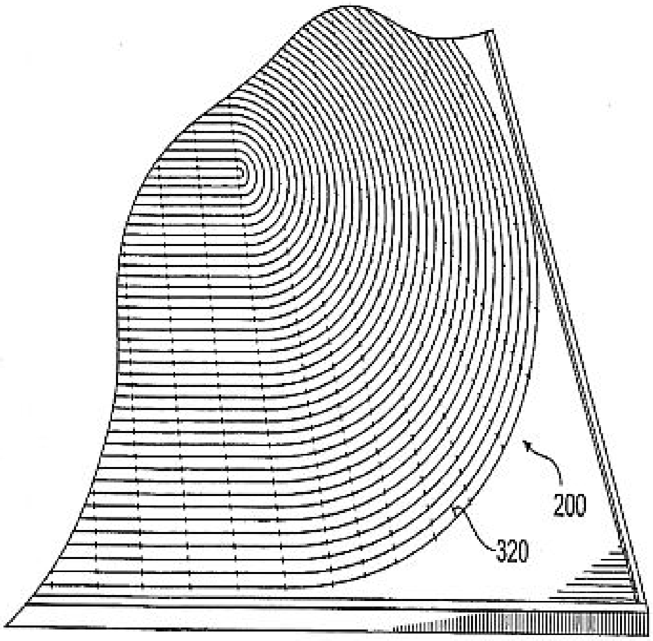

FIG. 3B

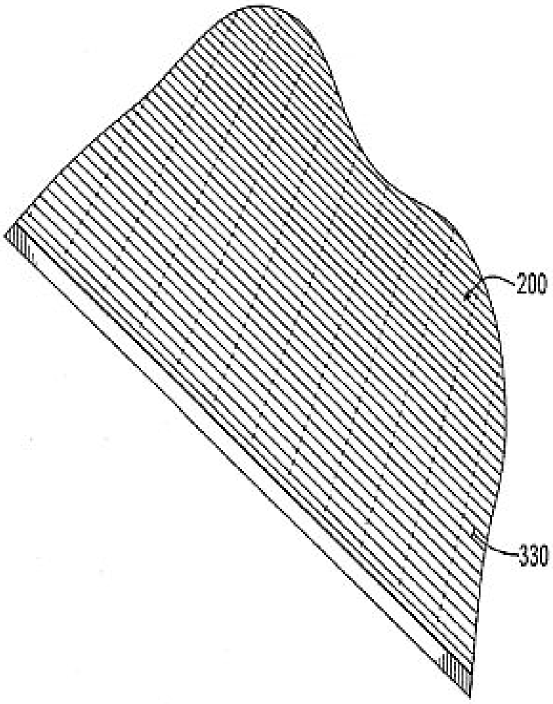

FIG. 3C

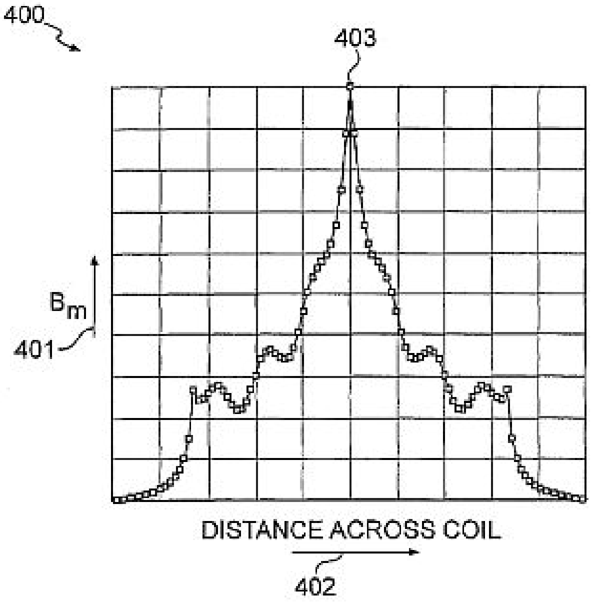
400
403
Bm
401
DISTANCE ACROSS COIL
402

FIG. 4

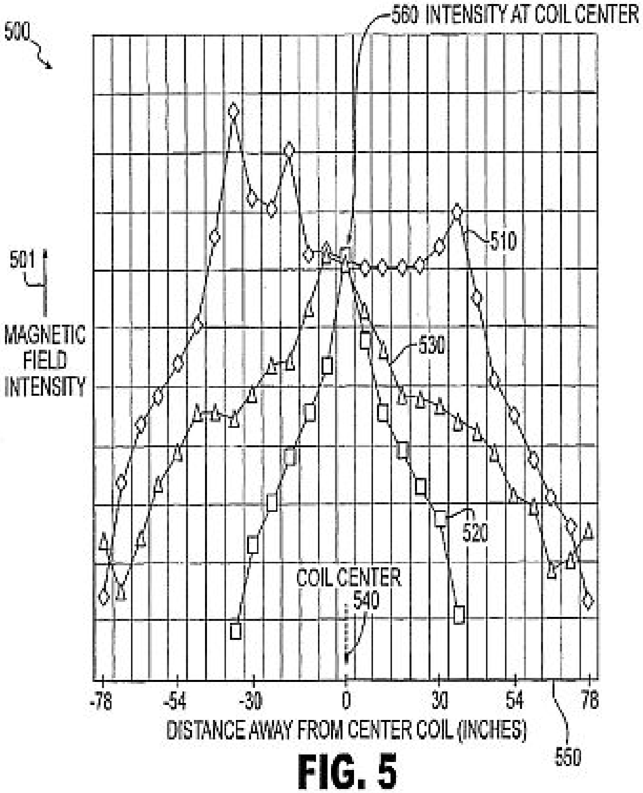
500
560 INTENSITY AT COIL CENTER
501
MAGNETIC FIELD INTENSITY
510
530
520
COIL CENTER
540
-78
-54
-30
0
30
54
78
DISTANCE AWAY FROM CENTER COIL (INCHES)
550

FIG. 5

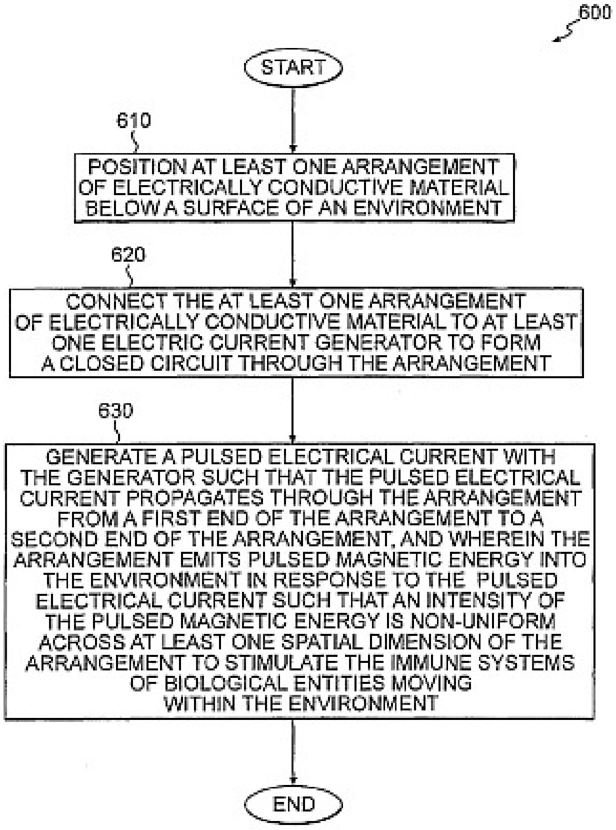
600
START
610
POSITION AT LEAST ONE ARRANGEMENT OF ELECTRICALLY CONDUCTIVE MATERIAL BELOW A SURFACE OF AN ENVIRONMENT
620
CONNECT THE AT LEAST ONE ARRANGEMENT OF ELECTRICALLY CONDUCTIVE MATERIAL TO AT LEAST ONE ELECTRIC CURRENT GENERATOR TO FORM A CLOSED CIRCUIT THROUGH THE ARRANGEMENT
630
GENERATE A PULSED ELECTRICAL CURRENT WITH THE GENERATOR SUCH THAT THE PULSED ELECTRICAL CURRENT PROPAGATES THROUGH THE ARRANGEMENT FROM A FIRST END OF THE ARRANGEMENT TO A SECOND END OF THE ARRANGEMENT, AND WHEREIN THE ARRANGEMENT EMITS PULSED MAGNETIC ENERGY INTO THE ENVIRONMENT IN RESPONSE TO THE PULSED ELECTRICAL CURRENT SUCH THAT AN INTENSITY OF THE PULSED MAGNETIC ENERGY IS NON-UNIFORM ACROSS AT LEAST ONE SPATIAL DIMENSION OF THE ARRANGEMENT TO STIMULATE THE IMMUNE SYSTEMS OF BIOLOGICAL ENTITIES MOVING WITHIN THE ENVIRONMENT
END

FIG. 6

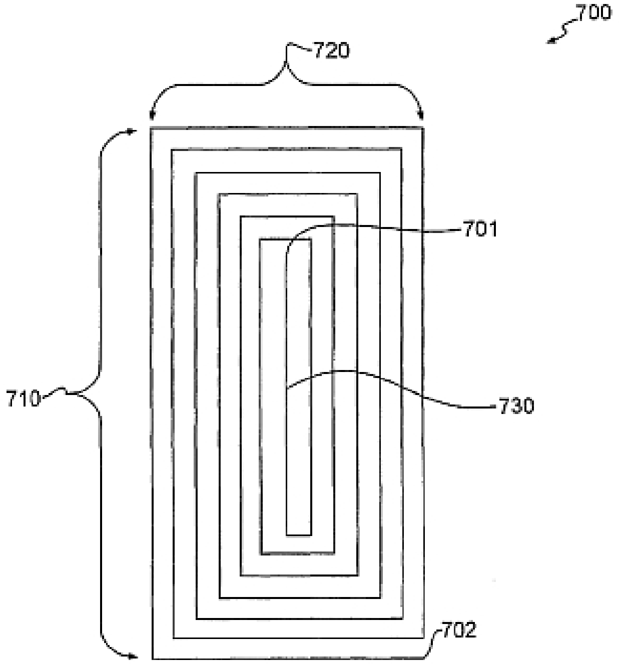
700
720
701
710
730
702

FIG. 7

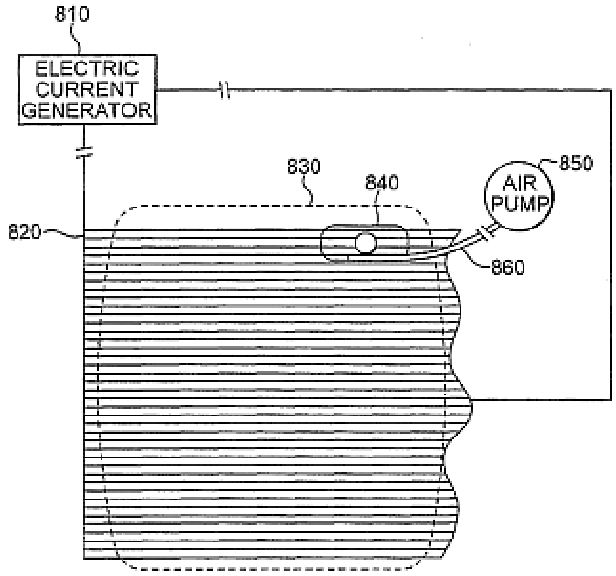
810
ELECTRIC CURRENT GENERATOR
830
840
850
AIR PUMP
820
860

FIG. 8

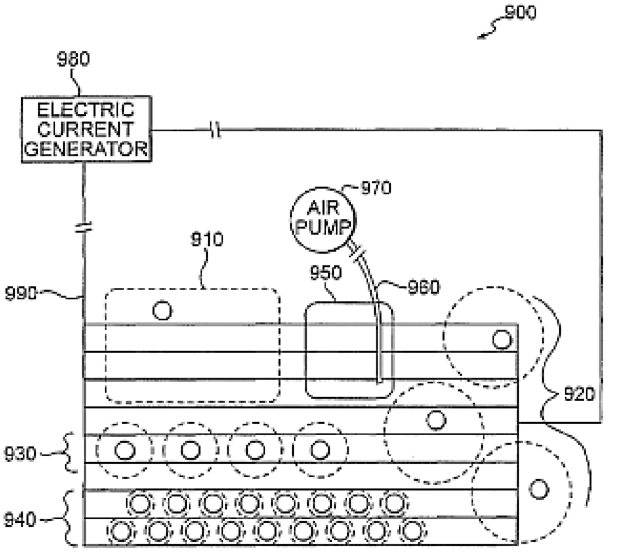
ELECTRIC CURRENT GENERATOR
AIR PUMP

FIG. 9

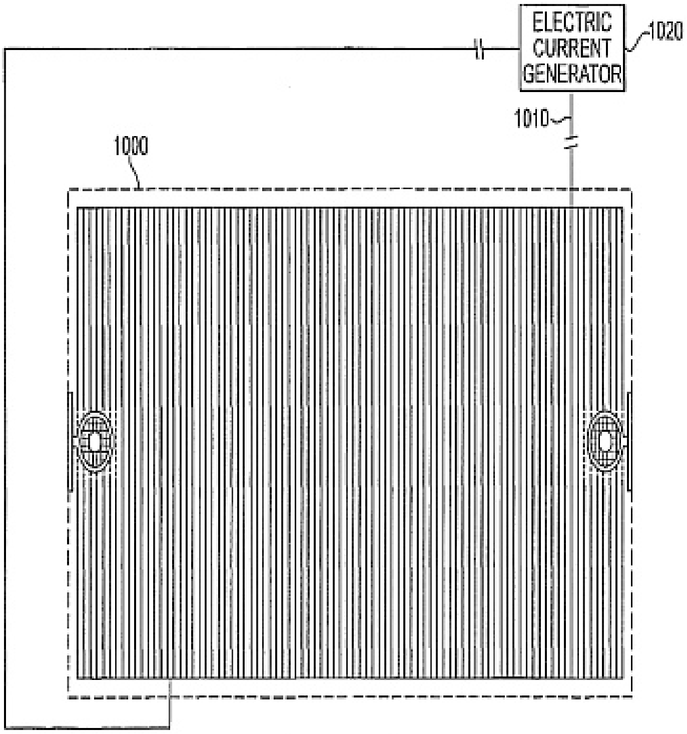
ELECTRIC CURRENT GENERATOR

FIG. 10

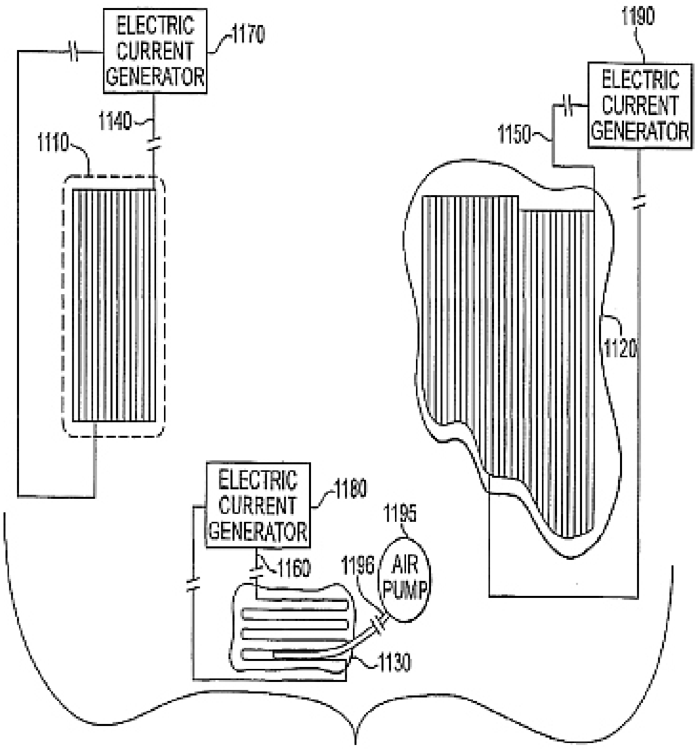
ELECTRIC CURRENT GENERATOR
1170
1140
1110
1190
ELECTRIC CURRENT GENERATOR
1150
1120
ELECTRIC CURRENT GENERATOR
1180
1195
AIR PUMP
1160
1196
1130

FIG. 11

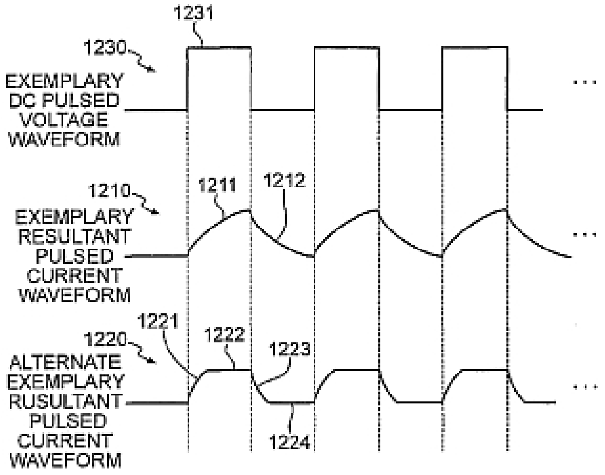
1231
1230
EXEMPLARY DC PULSED VOLTAGE WAVEFORM
1210
EXEMPLARY RESULTANT PULSED CURRENT WAVEFORM
1211
1212
1220
ALTERNATE EXEMPLARY RUSULTANT PULSED CURRENT WAVEFORM
1221
1222
1223
1224

FIG. 12

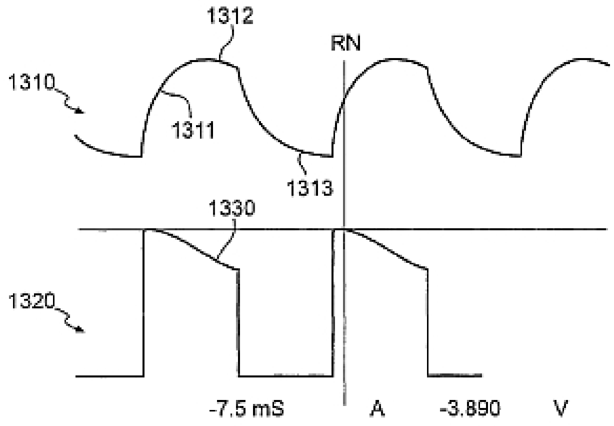

FIG. 13

DETAILED DESCRIPTION OF THE INVENTION

Reference will now be made in detail to the various embodiments of the present invention, examples of which are illustrated in the accompanying drawings. Whenever possible, the same reference numbers will be used throughout the drawings to refer to the same or like parts.

FIG. 1 illustrates a first exemplary embodiment of a system for applying pulsed magnetic energy to an aquatic environment, in accordance with various aspects of the present invention. Shown generally at 10, the embodiment of FIG. 1 includes a water reservoir 12, such as an aquarium, a pool, a pond or some other body of water. Below water reservoir 12 is placed electric wiring 16. In FIG. 1 electric wiring 16 is shown as a continuous coil of wiring, originating from the electric current generator 14 at a first end 17 and connecting back to the electric current generator 14 at a second end 19 to form a closed circuit. The electrical wiring 16 is placed over a substantial area of the water reservoir 12, at 1 to 30 inches below the bottom surface of the water reservoir 12, and is insulated from the surrounding elements and cross currents from other cables that may be present. If the water reservoir 12 has sides, such as in a pool, electric wiring 16 can also be placed behind the surface of the sides of reservoir 12. Electric wiring 16 is insulated by neutral materials, such as plastics, that will insulate the electric wiring 16, yet not produce a significant charge in the surrounding soil or substance. The electric wiring may be coated with an insulating material or may be placed in a non-conductive conduit, for example. Further, highly shielded electrical wire (e.g., using a conductive shield) is not desired as it decreases the capacity of diffusion of the pulsed magnetic energy into the aquatic environment. A pulsed DC voltage waveform is supplied to the electric wiring 16, and is controlled by the electric current generator 14. Pulsed current supplied by the generator 14 propagates from a first end 17 of the coil 16 to a second end 19 of the coil 16 in a closed loop in response to the pulsed DC voltage waveform. That is, both ends 17 and 19 of the coil 16 are connected to output terminals of the generator 14. Alternatively, pulsed current supplied by the generator 14 may propagate from the second end 19 to the first end 17.

Optionally, an air pump 18 and perforated air hose 20 may be included in the system embodied in FIG. 1. Air pump 18 and air hose 20 provide an oxygen source to the water within reservoir 12, thus, aerating the water and increasing the oxygen concentration of the water.

In use, pulsed electrical current is applied to the electric wiring 16 by an electric current generator 14. Various electronic components may be utilized to generate electric current for use in this system and may include a computer controlled subsystem, for example, to control the intensity and pulsed frequency of the emitted magnetic energy. An electric current generator typically includes a power transformer, a rectifier and filter circuit, and an electronic switching circuit. For example, an electric current generator may be plugged into a 240 VAC single phase power source and output a pulsed DC voltage waveform having a maximum peak amplitude of, for example, 80 VDC.

In accordance with an embodiment of the present invention, the magnetic energy emitted from the wiring 16 has magnetic field components in the range of about 0.5 to 30 Gauss, and the frequency range of the pulses is between 0.5 and 30 Hertz. Concurrently, air pump 18 pumps air through hose 20 to increase the oxygen concentration of the water. The water in free motion around the air bubbles is energized to a higher state because of the application of the pulsed magnetic energy, and thus may become even more saturated with oxygen.

Application of the magnetic energy to the reservoir 12 provides for application of magnetic energy to both plants and animals (i.e., biological entities) situated within the aquatic system. The affected plants and animals exhibit enhanced physiologic effects such as increased growth and overall health, for example, due to stimulation of the immune systems of the plants and animals. Furthermore, the highly oxygenated and energized water can then be utilized for watering plants which may not be directly exposed to the magnetic energy.

Electric current flows in a wire when a potential difference (i.e., a voltage) is applied across both ends of the wire. When electric current flows through a wire, a magnetic field is set up or emanates from the wire. If the wire is arranged in a coiled or grid-like configuration, for example, the magnetic fields from the various turns of the coil may combine constructively and destructively to form a spatially non-uniform magnetic field profile.

FIG. 2 illustrates an exemplary embodiment of a flat coil 200 used to generate magnetic energy, in accordance with various aspects of the present invention. The coil is a continuous coil of electrically conductive material (e.g., copper wire) having a first end 201 and a second end 202. Both ends 201 and 202 are connected to at least one generator (e.g. generator 14 in FIG. 1) to form a closed circuit. The generator creates a pulsed voltage waveform (e.g., a DC pulsed waveform) that results in pulsed electrical currents that propagate through the coil 200 from the first end 201 to the second end 202 (or vice versa). When the current pulses propagate through the coil 200, magnetic energy is emitted from the coil in the form of a pulsed magnetic field. When the coil is placed in an environment such as, for example, a swimming pool, a livestock yard, a garden, orchards, an athletic ground, a play ground, a pond, a lake, a whirl pool, a hot tub, or an aquarium, the magnetic energy is dispersed into the environment. The magnetic energy tends to stimulate the immune systems of biological entities (e.g., plants, animals, and humans) that are moving within the environment.

In accordance with an embodiment of the present invention, a configuration of the coil 200 comprises a plurality of spiraling turns of a conductive material such as, for example, copper wire. The spiraling turns do not necessarily follow a strictly mathematical spiral, but rather, the turns spiral at least in the sense that the turns wrap around on each other from the inside (i.e., from a central position 203 of the coil 200) to the outside of the coil 200. The configuration forms a flat, substantially oval surface having a width-to-length ratio of between 0.4 and 0.8. Ideally, the width-to-length ratio is 0.618 which is the "golden mean" ratio found in many instances of nature. For example, the width-to-length ratio of the coil 200 shown in FIG. 2 is 84.1232 inches divided by 156 inches which is a ratio of 0.539. The coil 200 of FIG. 2 has approximately 37 turns spaced at approximately 1 inch separation. In accordance with an embodiment of the present invention, the coil 200 comprises 8 Gage solid copper wire having a resistance of about 0.6 ohms per 1000 feet.

The configuration of the coil 200 includes a plurality of parallel straight segments 210 of insulated copper wire being of substantially the same length, and a plurality of curved segment of insulated copper wire 220. The segments 210 and 220 are not discrete in the sense that they must be connected to form the coil 200. Instead, the coil 200 is a continuous piece of copper wire. However, in an alternative embodiment the coil 200 could be made from discrete segments that are connected together by, for example, welding or soldering. The configuration of the coil 200 is such that the plurality of turns of the coil are substantially in a single spatial plane, which gives the coil 200 its flat shape.

In accordance with the embodiment of FIG. 2, the coil emits a magnetic field at least perpendicular to the surface of the coil (i.e., out of the page of FIG. 2). The intensity of the magnetic field varies (i.e., is non-uniform) across the coil 200, in a spatial dimension which is parallel to the surface of the coil 200. The variation may be non-linear, in accordance with certain embodiments of the present invention. Therefore, as a biological entity moves within the environment in which the coil 200 is placed, the biological entity experiences the variations of the magnetic field. The variations stimulate the immune system of the biological entity. Typically, the coil 200 is insulated using a non-conducting material such as a plastic, to prevent direct conduction of the electrical currents into the surrounding environment.

FIGS. 3A-3C illustrate various views of an embodiment of the flat coil 200 of FIG. 2, clearly showing the spacing between the coil turns, in accordance with various aspects of the present invention. The coil 200 shown in FIG. 3A is mounted on a surface 310 (e.g., a plywood board) using plastic clips to secure the wiring of the coil 200.

FIG. 3B shows a curved section of the coil 200 with the spacing 320 between the curved segments being about one inch. FIG. 3C shows a straight section of the coil 200 with the spacing 330 between the straight and parallel segments being about one inch.

FIG. 4 illustrates an exemplary simulated graph 400 of how a magnetic field intensity Bm 401 generated by the coil 200 of FIG. 2 may be expected to vary non-linearly across a spatial dimension 402 of the coil of FIG. 2 (distance across coil), in accordance with various aspects of the present invention. The peak magnetic intensity 403 occurs at the center of the coil 200. It is the configuration of the coil (e.g., shape, dimensions, spacing) that largely determines the shape of the spatially non-uniform magnetic field intensity. For example, the graph 400 of FIG. 4 may represent the variation in magnetic field intensity Bm across the width dimension 240 of the coil 200 and through a first center axis 250 of the coil 200 (see FIG. 2).

FIG. 5 illustrates an exemplary graph 500 of measured data of how a magnetic field intensity 501 generated by the coil 200 of FIGS. 3A-3C varies across three spatial dimensions of the coil 200 of FIGS. 3A-3C, in accordance with various aspects of the present invention. The data set 510 shows the variation in measured magnetic field intensity above (e.g., 18 inches) the coil 200 and across the length dimension 260 of the coil 200 along the axis 265 which includes the physical center point 205 of the coil 200 (see FIG. 2). The coil center 540 is shown on the distance axis 550 in FIG. 5.

The data set 520 shows the variation in measured magnetic field intensity above (e.g., 18 inches) above

the coil 200 and across the width dimension 240 of the coil 200 along the axis 250 which includes the physical center point 205 of the coil 200. The data set 530 shows the variation in measured magnetic field intensity above (e.g., 18 inches) the coil 200 and across a diagonal dimension of the coil 200 along the axis 270 which includes the physical center 205 of the coil 200.
The direction of the magnetic field intensities 510, 520, and 530 shown in the graph 500 is perpendicular to the flat surface of the coil 200. The absolute magnitude of the magnetic field intensities is a function of distance away from the surface or plane of the coil. In general, the magnetic field intensity decreases at points further away from the surface or plane of the coil. Notice that the magnetic field intensity 560 at the physical center of the coil is the same for all three data sets 510, 520, and 530 since the center corresponds to the same physical point in all three cases.

For example, if the coil 200 is placed flat and just beneath the bottom surface of a swimming pool (e.g., 1 to 30 inches), the magnetic field intensity 501 of the data sets 510, 520, and 530 will emanate above the coil into the water of the pool. As a swimmer swims through the pool across the coil (e.g., parallel to the surface of the coil), the swimmer will experience the magnetic variations of the magnetic field generated by the coil which stimulate the swimmer's immune system. In accordance with an embodiment of the present invention, the magnetic field is a pulsed magnetic field having a pulsed frequency of between 0.5 and 30 Hertz, and the intensity of the magnetic fields 510, 520, and 530 at a predetermined distance from the surface of the coil (e.g., 18 inches) vary through a range of about 0.5 to 30 Gauss across at least one spatial dimension of the coil 200. The electric current generator may comprise a programmable subsystem (e.g., a programmable logic controller or a computer-based subsystem such as a personal computer) which may control the frequency and intensity of the current pulses and, therefore, of the magnetic energy pulses.

FIG. 6 is a flowchart of an embodiment of a method 600 to stimulate immune systems of biological entities in an environment, in accordance with various aspects of the present invention. In step 610, at least one arrangement of electrically conductive material is positioned below a surface of an environment (for example, referring to FIG. 1, the coil 16 is positioned below the surface of the aquatic environment 12). In step 620, the at least one arrangement of electrically conductive material is connected to at least one electric current generator to form a closed circuit through the arrangement (for example, referring to FIG. 1, the ends 17 and 19 of the coil 16 are connected to electric terminals of the generator 14). In step 630, a pulsed electrical current is generated with the generator such that the pulsed electrical current propagates through the arrangement from a first end (e.g., end 19 in FIG. 1) of the arrangement to a second end (e.g., end 17 in FIG. 1) of the arrangement, and wherein the arrangement emits pulsed magnetic energy into the environment in response to the pulsed electrical current such that an intensity of the pulsed magnetic energy is non-uniform across at least one spatial dimension of the arrangement (e.g., see the non-uniform magnetic fields of FIGS. 4 and 5) to stimulate the immune systems as the biological entities move within the environment.

FIG. 7 illustrates a second exemplary embodiment of a flat coil 700 used to generate magnetic energy, in accordance with various aspects of the present invention. The coil 700 has a first end 701 and a second end 702 and comprises a first plurality of vertically oriented, parallel straight segments 710 of conductive wire and a second plurality 720 of horizontally oriented, parallel straight segments 720 of conductive wire which are substantially perpendicular to the first plurality of segments. The coil 700 is a continuous wire, winding outward from a central position 730 of the coil 700 in substantially a single spatial plane forming a flat, rectangular coil (i.e., a coil forming a flat, rectangular surface). The coil 700 of FIG. 7 is similar to the coil 200 of FIG. 2 except the curved segments of FIG. 2 are replaced with the straight segments 720 of FIG. 7 and the straight segments 710 are not all of the same length.

Referring now to FIG. 8, there is shown a second exemplary embodiment a system of the present invention, illustrating the application of pulsed magnetic energy to livestock animal pens. In this second embodiment, there is a stock pen 830 which can be used to retain any type of livestock animal or poultry. Further, it should be understood that such a stock pen can be enclosed in buildings or open fields. Electrical wiring 820 extends below the surface and laterally across first stock pen 830 to cover at least a substantial area of stock pen 830. The configuration of the electrical wiring 820 includes a plurality of substantially parallel segments of conductive material (e.g., segments of wiring) forming a flat, substantially rectangular grid. As in the first embodiment 10 of FIG. 1, electrical wiring 820 is comprised of electrical wiring having neutral insulating materials, and is placed under the surface of the selected area at a preferred depth of 1 to 30 inches. Electric current generator 810 generates and transmits pulsed electric currents through electric wiring 820. In accordance with an embodiment of the present invention, the width-to-length ratio of the substantially rectangular grid is between 0.4 and 0.8.

Optionally, this system may also include a water tank 840 for the watering of the animals. An air pump 850 may be used to pump air into the water tank 840 through hose 860. Aeration of water tank 840 by pump 850 increases the oxygen concentration of the water held within water tank 840.

In the embodiment of FIG. 8, the electrical wiring 820 forms a type of horizontal grid configuration across the stock pen 830. Pulsed current flows from the generator 810 through the various branches of the grid of wiring 820 and back to the generator 810, emitting pulsed magnetic energy (i.e., magnetic fields) into the environment of the stock pen 830. If the stockyard is enclosed, the electric wiring is buried at a depth between 1 and 30 inches within the dirt or other flooring material, such as concrete, for example. Again, a spatially non-uniform magnetic field is generated across the grid. However, the non-uniformity may be substantially different from that shown in FIGS. 4 and 5 due largely to the different configuration of the wiring 820 from that of the coil 200 of FIG. 2. However, a coil of the configuration of that of FIG. 1 or FIG. 2, or other configurations, could be used in the stockyard environment instead.

Now referring to FIG. 9, there is shown a third exemplary embodiment of a system of the present invention illustrating the application of pulsed magnetic energy to a plant bio-system 900, such as a garden. The plant based bio-system may include, for example, a grain plot 910, fruit or other trees 920, plants 930 and/or flowers 940. The plants may either be grown within the ground itself or within planting containers such as pots and the like. The environment 900 may optionally also include a water reservoir 950, which is aerated by hose 960 and air pump 970 to increase the oxygen concentration of the water held within reservoir 950. Pulsed magnetic energy is generated by an electric current generator 980 applying pulsed currents to electric wiring grid 990. Electric wiring grid 990 is positioned to cover at least a substantial portion of environment 900, and is placed below the surface of environment 900, at a depth of 1 to 30 inches. The plants affected by the application of the pulsed magnetic energy exhibit improved physiological effects such as improved and more rapid growth, and better overall health for example.

Shown in FIG. 10 is yet another exemplary embodiment of the present invention, illustrating applying magnetic energy to an athletic playing surface 1000, such as, for example a basketball court, a football field, a soccer field, a swimming pool, playgrounds or other playing surfaces. Electrical wiring 1010 is placed below playing surface 1000 at a preferred depth of between 1 and 30 inches. Electrical current generator 1020 generates and conducts electrical currents through electric wiring 1010 to create and diffuse the resultant magnetic energy throughout playing surface 1000. The magnetic energy is applied to those persons playing or competing on playing surface 1000. Upon application of the magnetic energy, humans may experience higher energy levels for longer periods of time, reduced fatigue, less muscle strain and soreness, in addition to increased concentration and precision in playing the particular sport.

Referring now to FIG. 11, there is shown yet another exemplary embodiment of the present invention, illustrating application of magnetic energy to a golf course hole. This embodiment includes a tee box 1110, a green 1120 and/or optionally a water reservoir 1130. Below tee box 1110 is placed electrical wiring 1140, which is placed at a depth of 1 to 30 inches. Electrical wiring 1150 is also placed below golf green 1120 at a preferred depth of 1 to 30 inches. If water reservoir 1130 is used, electrical wiring 1160 is positioned below the bottom of reservoir 1130 at a depth of 1 to 30 inches for emission of magnetic energy. Air pump 1195 and air hose 1196may optionally be used to pump air into the water reservoir 1130 to increase the oxygen content. Treated water from reservoir 1130 may be used to water the golf course fairways, greens, or other plants associated with the course as to achieve the benefits as described above in relation to the plant based bio-system embodiment. Electric current generators 1170, 1180 and 1190 produce pulsed electric currents such that magnetic energy is emitted from electrical wiring 1140, 1160, and 1150 respectively. Optionally, a single generator can control the emission of magnetic energy from electrical wiring 1140, 1160, and 1150. As with the above described embodiments, electric current generators 1170, 1180 and 1190 cause magnetic energy to be emitted having magnetic field components about in the range of 0.5 to 30 Gauss, and the frequency range of the pulses is between 0.5 and 30 Hertz.

Effects seen through application of the pulsed magnetic energy to a grass surfaces such as soccer fields, play grounds, football fields, golf greens and tee boxes include more rapid and healthier growth of grass, faster regeneration or repair of divots and ball marks, fewer attacks to these grasses by pests as the grasses are healthier. The human players experience gentle invigoration, increased energy, greater concentration, and less muscle soreness or strain. Increased mental activity and faster healing of wounds has also been noted. Sporting equipment such as golf clubs are not affected by the application of magnetic energy because no sustained electrical current is conducted to the metal portions of the clubs.

In use with any of the above embodiments, the characteristics of the magnetic energy remain the same, that is the magnetic energy having magnetic field components about in the range of 0.5 to 30 Gauss, and the frequency range of the pulses between 0.5 and 30 Hertz. In application of pulsed magnetic energy to humans, the magnetic field strength may be adjusted to vary between 4 to 8 Gauss. The placement of the electrical wiring below the surface of the selected area is adjusted to accommodate these parameters. Further, it is important to note that the spacing and arrangement of the electric wiring in the above described embodiments may be altered to achieve certain desired effects.

FIG. 12 illustrates exemplary resultant current pulses 1210 and 1220 that may be produced in the coil 200 of FIG. 2 when applying an exemplary DC pulsed voltage waveform 1230 to the coil 200 of FIG. 2, in accordance with an embodiment of the present invention. The pulsed voltage waveform 1230 shown in FIG. 12 is a square voltage waveform having a 50% duty cycle. Other duty cycles are possible as well, in accordance with various embodiments of the present invention. The pulsed voltage waveform 1230 is applied to the coil 200 by a generator (e.g., generator 14 of FIG. 1).

The frequency of the pulsed voltage waveform 1230 may be, for example, anywhere between 0.1 Hz and 30 Hz. Depending on the various parameters (e.g., the voltage level, the time constant, the pulsed frequency, etc.) of the system, the resultant pulsed current waveform in the coil may look like that of waveform 1210. Referring to the pulsed current waveform 1210, as the voltage level of the pulsed voltage waveform 1230 increases, the current level in the coil will begin to increase as seen in the segment 1211 of the pulsed current waveform 1210. The curved nature of the rising current level of the segment 1211 is due, at least in part, to the time constant of the system (including the coil) which is determined by inductive, capacitive, and resistive factors of the system. In the pulsed current waveform 1210, the current level rises continuously until the voltage level of the driving pulsed voltage waveform drops off.

When the voltage level of the pulsed voltage waveform 1230 decreases, the current level in the coil will begin to decrease as seen in the segment 1212 of the pulsed current waveform 1210. Again, the curved nature of the falling current level of the segment 1212 is due, at least in part, to the time constant of the system. In the pulsed current waveform 1210, the current level decreases continuously until the voltage level of the driving pulsed voltage waveform again rises. For example, the peak voltage level of the DC pulsed voltage waveform 1230 may be 80 VDC and the resultant peak current level of the pulsed current waveform 1210 may be 100 amps.

Referring to the pulsed current waveform 1220, as the voltage level of the pulsed voltage waveform 1230 increases, the current level in the coil will begin to increase as seen in the segment 1221 of the pulsed current waveform 1220. Again, the curved nature of the rising current level of the segment 1221 is due, at least in part, to the time constant of the system. In the pulsed current waveform 1220, the current level rises and then flattens off to a peak current level 1222 well before the voltage level of the pulsed voltage waveform drops off. This flattening off tends to occur when the peak voltage level 1231 is relatively low. The lower peak voltage level 1231 means that the current will not build to as high a level as it would with a higher peak voltage level driving the coil. Therefore, the pulsed current waveform 1220 reaches its peak level sooner and stays there.

When the voltage level of the pulsed voltage waveform 1230 decreases, the current level in the coil will begin to decrease as seen in the segment 1223 of the pulsed current waveform 1220. Again, the curved nature of the falling current level of the segment 1223 is due, at least in part, to the time constant of the system. In the pulsed current waveform 1220, the current level decreases to a zero current level 1224 well before the voltage level of the pulsed voltage waveform increases again. Again, this flattening off tends to occur when the peak voltage level 1231 is relatively low. That is, the current level does not have as far to fall since the peak current level was relatively low. Therefore, the current level reaches zero sooner and flattens off. For example, the peak voltage level of the DC pulsed voltage waveform 1230 may be 20 VDC and the resultant peak current level of the pulsed current waveform 1210 may be 10 amps.

FIG. 13 illustrates exemplary resultant current pulses 1310 produced in the coil 200 of FIGS. 3A-3C when applying an exemplary DC pulsed voltage waveform 1320 to the coil 200 of FIGS. 3A-3C, in accordance with an embodiment of the present invention. The pulsed voltage waveform 1320 shown in FIG. 13 is a pseudo-square voltage waveform having a 50% duty cycle. Other duty cycles are possible as well, in accordance with various embodiments of the present invention. The pulsed voltage waveform 1320 tends to droop over the segment 1330 due to the load the coil provides to the generator. Therefore, the pulsed voltage waveform 1320 is not perfectly square.

The frequency of the pulsed voltage waveform 1320 may be, for example, anywhere between 0.1 Hz and 30 Hz. Depending on the various parameters (e.g., the voltage level, the time constant, the pulsed frequency, etc.) of the system, the resultant pulsed current waveform in the coil may look like that of waveform 1310. Referring to the pulsed current waveform 1310, as the voltage level of the pulsed voltage waveform 1320 increases, the current level in the coil will begin to increase as seen in the segment 1311 of the pulsed current waveform 1310. The curved nature of the rising current level of the segment 1311 is due, at least in part, to the time constant of the system (including the coil) which is determined by inductive, capacitive, and resistive factors of the system. In the pulsed current waveform 1310, the current level rises continuously until the voltage level of the driving pulsed voltage waveform begins to droop. There is a time delay, however, between when the voltage level begins to droop and when the current level begins to decrease slightly over the segment 1312.

When the voltage level of the pulsed voltage waveform 1320 drops off, the current level in the coil will begin to decrease as seen in the segment 1313 of the pulsed current waveform 1310. Again, the curved nature of the falling current level of the segment 1313 is due, at least in part, to the time constant of the system. In the pulsed current waveform 1310, the current level decreases continuously until the voltage level of the driving pulsed voltage waveform again rises.

In accordance with various embodiments of the present invention, the systems described herein may be used by incrementing and/or decrementing the pulsed frequency over time. For example, in accordance with an embodiment of the present invention, the pulsed frequency may start at 0.5 Hz and be incremented every one minute by 0.5 Hz until reaching 28 Hz. Then the pulsed frequency may be decremented from 28 Hz back down to 0.5 Hz at a frequency step of 0.5 Hz every minute. Other methods of varying the pulsed frequency over time are possible as well and may be tailored to certain physiological conditions to be treated by stimulating the immune system.

In summary, a method and systems are disclosed for stimulating the immune systems of biological entities in an environment. A magnetic energy field is generated such that the magnetic energy field varies non-uniformly in intensity across at least one spatial dimension of the environment. The magnetic energy field is generated using an electric current generator which is connected to a coil or other alternate arrangement of conductive material such as wire. The coil or arrangement is typically placed beneath a surface of the environment. The magnetic energy field is pulsed at a predetermined frequency.

While the invention has been described with reference to certain embodiments, it will be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the scope of the invention. In addition, many modifications may be made to adapt a particular situation or material to the teachings of the invention without departing from its scope. Therefore, it is intended that the invention not be limited to the particular embodiment disclosed, but that the invention will include all embodiments falling within the scope of the appended claims.

ELECTROMAGNETIC ENERGY ENHANCING GROWTH AND BIOLOGICAL FUNCTION

WO2005084746

A system (14,16,18,20) and method of enhancing growth and biological functions of living systems through application of pulsed electromagnetic energy. An Electromagnetic generator (14) connected to electrical wiring (16) wich is placed below the surface of a selected area such as, a garden, stock pen or athletic playing field (24) for emission of electromagnetic energy to plants and animals, including humans.

FIELD OF THE INVENTION

The present invention is generally a means to enhance growth and performance of biological systems. More specifically, the present invention is a system and method to enhance physiological performance of organisms through application of pulsed electromagnetic energy.

BACKGROUND OF THE INVENTION

Use of electromagnetic energy to increase physiological performance of organisms has long been attempted. However, many of these techniques have been limited to belts, pads or mats which apply magnetic or electromagnetic energy to the person or other organism. Problems inherent in these techniques include the necessity for the organism to wear the belt or pad, and the necessity for a portable power source in order to generate electromagnetic energy.

Furthermore, these techniques do not effect the environment surrounding the organism, nor do they bring the environment and organism into gentle resonance. Accordingly, there is a demand for an apparatus and method of applying pulsed electromagnetic energy to an organism and its surrounding environment that is without the aforementioned disadvantages.

SUMMARY OF THE INVENTION

The present invention is an apparatus and method of applying pulsed electromagnetic energy to organisms such as humans or other animals, and to the organism's surrounding environment such as grass and other organisms within this environment. Specifically, electrical wiring which emits pulsed electromagnetic energy, is placed under the surface of the area in which the humans or other organisms are situated. This includes putting wiring in proximity to or under the surface of a variety of environments, such as livestock pens, poultry pens, ponds, garden plots, athletic courts, and golf courses. The system and method of the present invention is used in association with various different organisms or living systems such as humans and other animals, as well as plants, for example.

Electromagnetic energy is then pulsed through the electric wiring and diffused throughout the area surrounding the wiring, including the atmosphere above the surface of the selected area.

As the electromagnetic energy diffuses, it is applied to both the plants and animals within the selected area. The application of the pulsed electromagnetic energy brings these organisms into gentle resonance and enhances biological function and genetic expression of these organisms.

Generally, better health, increased appetite and higher consumption of water, are among the positive effects that have been experienced by participating humans.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a schematic diagram of a first preferred embodiment of the present invention, illustrating the application of pulsed electromagnetic energy to an aquatic bio-system.

Fig. 2. is a schematic diagram of a second preferred embodiment of the present invention, illustrating the application of pulsed electromagnetic energy to livestock pens.

Fig. 3 is a schematic diagram of a third preferred embodiment of the present invention, illustrating the application of pulsed electromagnetic energy to a plant bio-system.

Fig. 4. is a schematic diagram of a fourth preferred embodiment of the present invention, illustrating the application of pulsed electromagnetic energy to a basketball court.

Fig. 5 is a schematic diagram of a fifth preferred embodiment of the present invention, illustrating the application of pulsed electromagnetic energy to a golf course.

Fig 6. is a schematic view of an audio system for use with any embodiment of the present invention.

FIGURE 1

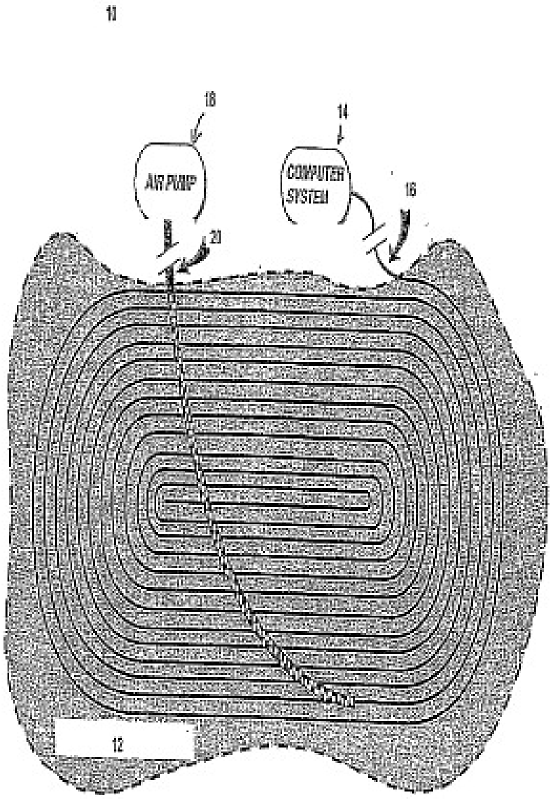
10
18
14
AIR PUMP
COMPUTER SYSTEM
16
20
12

FIGURE 2

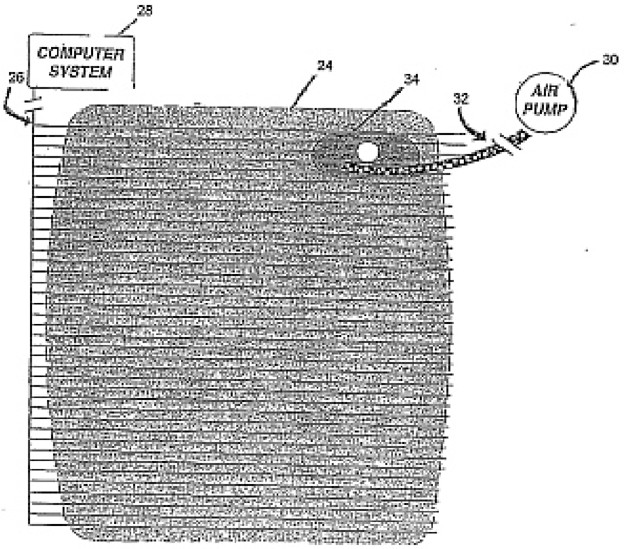
28
COMPUTER SYSTEM
26
24
34
32
30
AIR PUMP

FIGURE 3

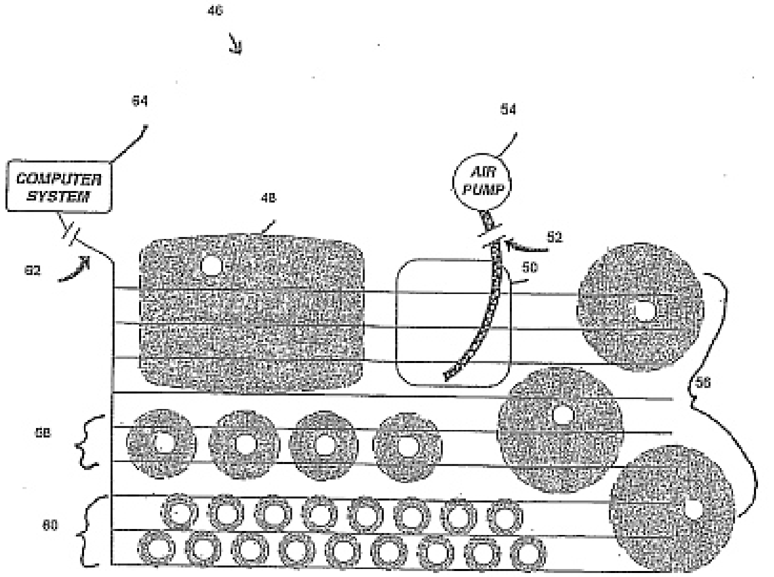
46
64
54
COMPUTER SYSTEM
48
AIR PUMP
52
50
62
56
58
60

FIGURE 4

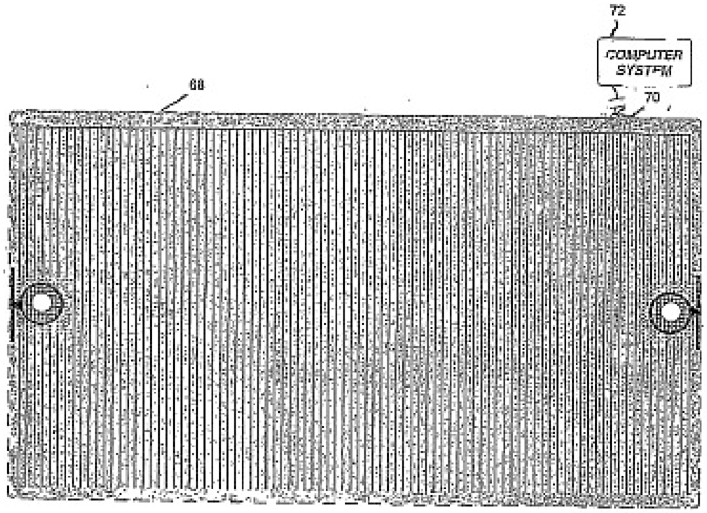
72
COMPUTER SYSTEM
68
70

FIGURE 5

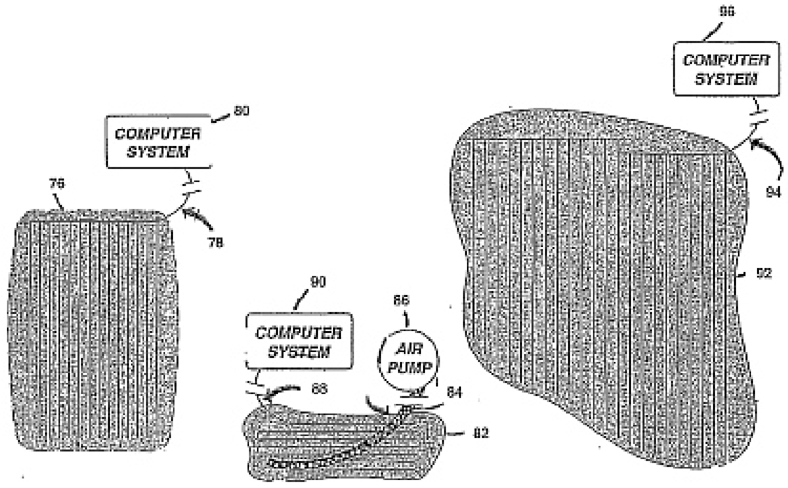
96
COMPUTER SYSTEM
80
COMPUTER SYSTEM
76
78
94
92
90
COMPUTER SYSTEM
86
AIR PUMP
88
84
82

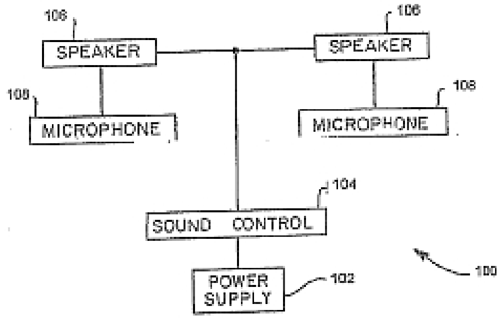
106
SPEAKER
106
SPEAKER
108
MICROPHONE
108
MICROPHONE
104
SOUND CONTROL
100
POWER SUPPLY
102

FIGURE 6

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Reference will now be made in detail to the present preferred embodiments of the present invention, examples of which are illustrated in the accompanying drawings.

Whenever possible, the same reference numbers will be used throughout the drawings to refer to the same or like parts.

Shown in Fig. 1 is a first preferred embodiment of the present invention. It illustrates the application of pulsed electromagnetic energy to any aquatic setting. Shown generally at 10, this embodiment includes a water reservoir 12, such as an aquarium, pool, pond or other body of water. Below water reservoir 12 is placed electric wiring 16. Although in Fig. 1 electric wiring 16 is shown in a continuous open ended coil of wiring, electric wiring 16 can be positioned in a discontinuous fashion having a main line and a plurality of other lines extending from the main line. The electrical wiring 16 is placed over a substantial area of water reservoir 12, from 1 to 30 inches below the bottom surface of the water reservoir 12 and is insulated from the surrounding elements and cross currents from other cables. If water reservoir 12 has sides, such as in a pool, electric wiring 16 can also be placed behind the surface of the sides of reservoir 12. Electric wiring 16 is insulated by neutral materials, such as plastics, that will insulate electric wiring 16, yet not produce a charge in the surrounding soil or substance. Further, highly shielded electrical wire is not desired as it decreases the capacity of diffusion of the pulsed electromagnetic energy. The pulsed electromagnetic energy is supplied to electric wiring 16, and is controlled by an electromagnetic energy generator 14.

Optionally, an air pump 18 and perforated air hose 20 may be included in the system embodied in Figure 1. Air pump 18 and air hose 20 provide an oxygen source to the water with in reservoir 12, thus, aerating the water and increasing the oxygen concentration of the water.

In use, pulsed electromagnetic energy is applied to the electric wiring 16 by an electromagnetic energy generator 14. Various electronic components maybe utilized to generate electromagnetic energy for use in this system, such as a computer system or a electromagnetic wave generator for example. Preferably, the electromagnetic energy has a magnetic field component in the range of 0.5 to 30 Gauss, and the frequency range of the pulses between 0.5 and 30 Hertz. Concurrently, air pump 18 pumps air through hose 20 to increase the oxygen concentration of the water. The water in free motion around the air bubbles is energized to a higher state because of the application of the pulsed electromagnetic energy, and thus becomes even more saturated with oxygen.

Application of the electromagnetic energy to reservoir 12 provides for application of electromagnetic energy to both plants and animals situated within the aquatic system. The affected plants and animals exhibit enhanced physiologic effects such as increased growth and overall health for example. Furthermore, the highly oxygenated and energized water can then be utilized for watering plants which are not directly exposed to electromagnetic energy.

Optionally the system and method of the present invention contemplates the sustained or intermittent introduction of acoustic energy into the aquatic setting or the environment of other embodiments of the present invention. Many forms of soothing sounds may provide enhanced benefit but most preferred is the use of symphonic music or sounds, preferably with minimal or no dischord. The introduction of such can be accomplished through a sound system 100 as is schematically illustrated in Fig. 6. Sound system 100 comprises a power supply 102, which may be independent or shared with other systems utilized in the environment, that is operably connected to a sound control panel 104 from which one may control the selection and characteristics of the sounds broadcast by sound system 100 into the surrounding environment of the present invention. Control panel 104 is connected to one or more speakers 106 for broadcasting the generated sound. In aquatic settings underwater speakers are contemplated.

Although not required, one or more microphones 108 for receiving sound from the aquatic setting or other environment is contemplated. Such sounds may be recorded as is known in the art.

Additionally the sustained or intermittent introduction of orgone or radionic energy into the aquatic setting of the environment or other embodiments of the present invention, especially but not necessarily in pulsed form, is contemplated. Such radionic energy is supplied through one or more radionic generators or collectors as are lcnown and available in the art. Exposure time and intensity may be varied.

Referring now to Fig. 2, there is shown a second embodiment of the present invention, illustrating the application of pulsed electromagnetic energy to livestock animal pens. In this embodiment, there is a stock pen 24 which can be used to retain any type of livestock animal or poultry. Further, it should be understood that such a stock pen can be enclosed in buildings or open fields. Electrical wiring 26 extends below the surface and laterally across first stock pen 24 to cover at least a substantial area of stock pen 24. As in the first embodiment, electrical wiring 26 is comprised of electrical wiring having neutral insulating materials, and is placed under the surface of the selected area at a preferred depth of 1 to 30 inches. Electromagnetic energy generator 28 generates and transmits electromagnetic energy to electric wiring 26. Optionally, this system may also include a water tank 34 for watering of the animals. An air pump 30 may be utilized to pump air into the water tank 34 through hose 32. Aeration of water tank 34 by pump 30 increases the oxygen concentration of the water held within water tank 34.

If the stoclcyard is enclosed, the electric wiring is buried at a depth between 1 and 30 inches within the dirt or other flooring material, such as concrete, for example.

Now referring to Fig. 3, there is shown a third embodiment of the present invention illustrating the application of electromagnetic energy to a plant bio-system 46, such as a garden.

The plant based bio-system may include, for example, a grain plot 48, fruit or other trees 56, plants 58 and/or flowers 60. The plants may either be grown within the ground itself or within planting containers such as pots and the like. The system 46 may optionally also include a water reservoir 50, which is aerated by hose 52 and air pump 54 to increase the oxygen concentration of the water held within reservoir 50. Electromagnetic energy is generated by an electromagnetic energy generator 64 and applied to system 46 via electric wiring 62. Electric wiring 62 is position to cover at least a substantial portion of system 46, and is placed below the surface of system 46, at a depth of 1 to 30 inches. The plants affected by the application of the electromagnetic energy exhibit improved physiological effects such as improved and more rapid growth, and better overall health for example.

Shown in Fig. 4 is yet another embodiment of the present invention, illustrating applying electromagnetic energy to an athletic playing surface 68, such as, for example a basketball court, a football field, a soccer field, a swimming pool, playgrounds or other playing surfaces.

Electrical wiring 70 is placed below playing surface 68 at a preferred depth of between 1 and 30 inches. Electromagnetic energy generator 72 generates and conducts electromagnetic energy through electric wiring 70 to diffuse the electromagnetic energy throughout playing surface 68.

The electromagnetic energy is applied to those persons playing or competing on playing surface 68. Upon application of the electromagnetic energy, humans experience higher energy levels for longer periods of time, reduced fatigue, less muscle strain and soreness, in addition to increased concentration and precision in playing the particular sport.

Referring now to Fig. 5, there is shown yet another embodiment of the present invention, illustrating application of electromagnetic energy to a golf course hole. This embodiment includes a tee box 76, a green 92 and/or optionally a water reservoir 82. Below tee box 76 is placed electrical wiring 78, which is placed at a depth of 1 to 30 inches. Electrical wiring 94 is also placed below golf green 92 at a preferred depth of 1 to 30 inches. If water reservoir 82 is utilized, electrical wiring 88 is positioned below the bottom of reservoir 82 at a depth of 1 to 30 inches for emission of electromagnetic energy. Treated water from reservoir 82 may be used to water the golf course fairways, greens, or other plants associated with the course as to achieve the benefits as described above in relation to the plant based bio-system embodiment.

Electromagnetic energy generators 80,90 and 96 produce and emit electromagnetic energy through electrical wiring 78,88 and 94 respectively. Optionally, a single electromagnetic generator can control the emission of electromagnetic energy through electrical wiring 78,88 and 94. As with the above described embodiments, electromagnetic energy generators 80,90 and 96 emit electromagnetic energy having a magnetic field component in the range of 0.5 to 30 Gauss, and the frequency range of the pulses between 0.5 and 30 Hertz.

Effects seen through application of the electromagnetic energy to a grass surfaces such as soccer fields, play grounds, football fields, golf greens and tee boxes include more rapid and healthier growth of grass, faster regeneration or repair of divots and ball marks, fewer attacks to these grasses by pests as the grasses are healthier. The human players experience gentle invigoration, increased energy, greater concentration, and less muscle soreness or strain.

Increased mental activity and faster healing of wounds has also been noted.

Sporting equipment such as golf clubs are not affected by the application of electromagnetic energybecause no sustained electrical current is conducted to the metal portions of the clubs.

In use with any of the above embodiments, the characteristics of the electromagnetic energy remain the same, that is electromagnetic energy having a magnetic field component in the range of 0.5 to 30 Gauss, and the frequency range of the pulses between 0.5 and 30 Hertz. The cycling/hertz of the electromagnetic energy is varied slowly during a 24 hour period. In application of pulsed electromagnetic energy to humans, the magnetic field strength is adjusted to be within 4 to 8 Gauss. The placement of the electrical wiring below the surface of the selected area is adjusted to accommodate these parameters. Further, it is important to note that the spacing of the electric wiring in the above described embodiments may be altered to achieve desired effects.

Although the principles, preferred embodiments and preferred operation of the present invention have been described in detail herein, this is not to be construed as being limited to the particular illustrative forms disclosed. They will thus become apparent to those skilled in the art that various modifications of the preferred embodiments herein can be made without departing from the spirit or scope of the invention as defined by the appended claims.

US6902521

SYSTEM AND METHOD TO ENHANCE GROWTH AND BIOLOGICAL FUNCTION OF LIVING SYSTEMS WITH PULSED ELECTROMAGNETIC ENERGY

FIELD OF THE INVENTION

The present invention is generally a means to enhance growth and performance of biological systems. More specifically, the present invention is a system and method to enhance physiological performance of organisms through application of pulsed electromagnetic energy.

BACKGROUND OF THE INVENTION

Use of electromagnetic energy to increase physiological performance of organisms has long been attempted. However, many of these techniques have been limited to belts, pads or mats which apply magnetic or electromagnetic energy to the person or other organism. Problems inherent in these techniques include the necessity for the organism to wear the belt or pad, and the necessity for a portable power source in order to generate electromagnetic energy.

Furthermore, these techniques do not effect the environment surrounding the organism, nor do they bring the environment and organism into gentle resonance. Accordingly, there is a demand for an apparatus and method of applying pulsed electromagnetic energy to an organism and its surrounding environment that is without the aforementioned disadvantages.

SUMMARY OF THE INVENTION

The present invention is an apparatus and method of applying pulsed electromagnetic energy to organisms such as humans or other animals, and to the organism's surrounding environment such as grass and other organisms within this environment. Specifically, electrical wiring which emits pulsed electromagnetic energy, is placed under the surface of the area in which the humans or other organisms are situated. This includes putting wiring in proximity to or under the surface of a variety of environments, such as livestock pens, poultry pens, ponds, garden plots, athletic courts, and golf courses. The system and method of the present invention is used in association with various different organisms or living systems such as humans and other animals, as well as plants, for example.

Electromagnetic energy is then pulsed through the electric wiring and diffused throughout the area surrounding the wiring, including the atmosphere above the surface of the selected area.

As the electromagnetic energy diffuses, it is applied to both the plants and animals within the selected area. The application of the pulsed electromagnetic energy brings these organisms into gentle resonance and enhances biological function and genetic expression of these organisms.

Generally, better health, increased appetite and higher consumption of water, are among the positive effects that have been experienced by participating humans.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a schematic diagram of a first preferred embodiment of the present invention, illustrating the application of pulsed electromagnetic energy to an aquatic bio-system.

Fig. 2. is a schematic diagram of a second preferred embodiment of the present invention, illustrating the application of pulsed electromagnetic energy to livestock pens.

Fig. 3 is a schematic diagram of a third preferred embodiment of the present invention, illustrating the application of pulsed electromagnetic energy to a plant bio-system.

Fig. 4. is a schematic diagram of a fourth preferred embodiment of the present invention, illustrating the application of pulsed electromagnetic energy to a basketball court.

Fig. 5 is a schematic diagram of a fifth preferred embodiment of the present invention, illustrating the application of pulsed electromagnetic energy to a golf course.

Fig 6. is a schematic view of an audio system for use with any embodiment of the present invention.

FIGURE 1

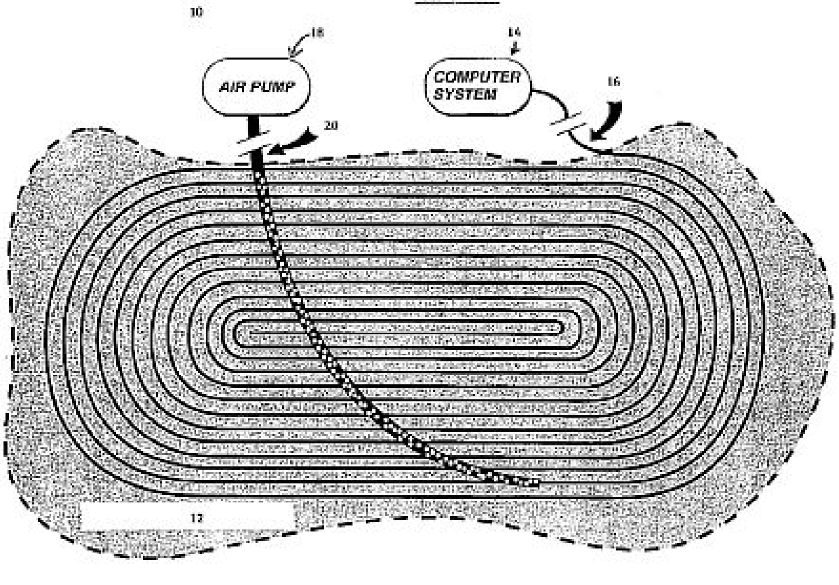

FIGURE 2

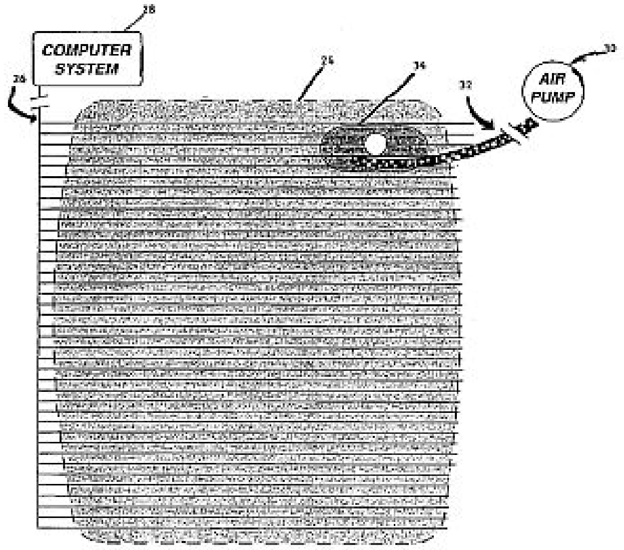
28
COMPUTER SYSTEM
26
24
34
32
AIR PUMP
30

FIGURE 3

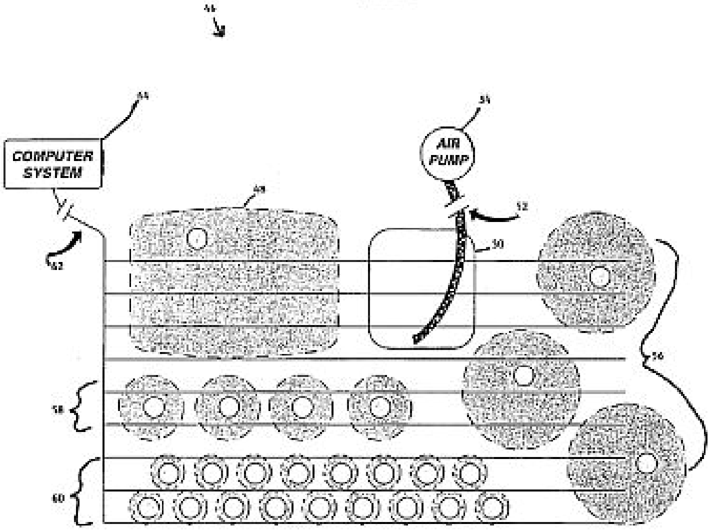
46
44
COMPUTER SYSTEM
54
AIR PUMP
48
52
50
42
56
58
60

FIGURE 4

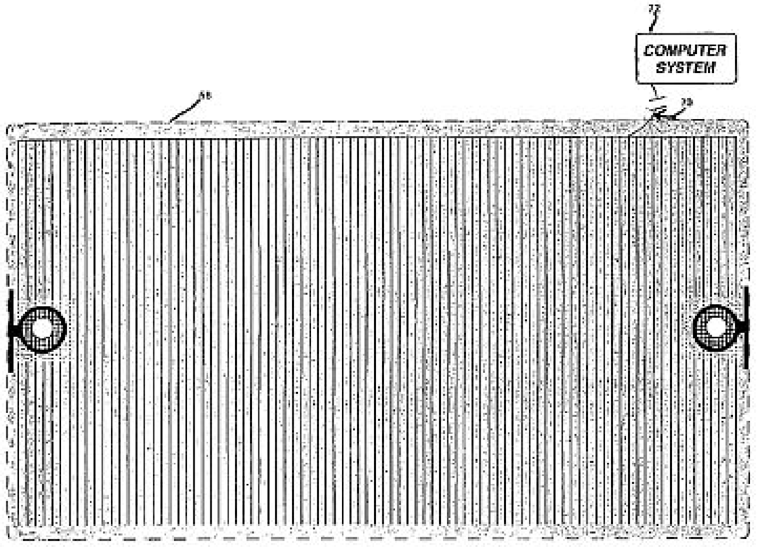
72
COMPUTER SYSTEM
68
70

FIGURE 5

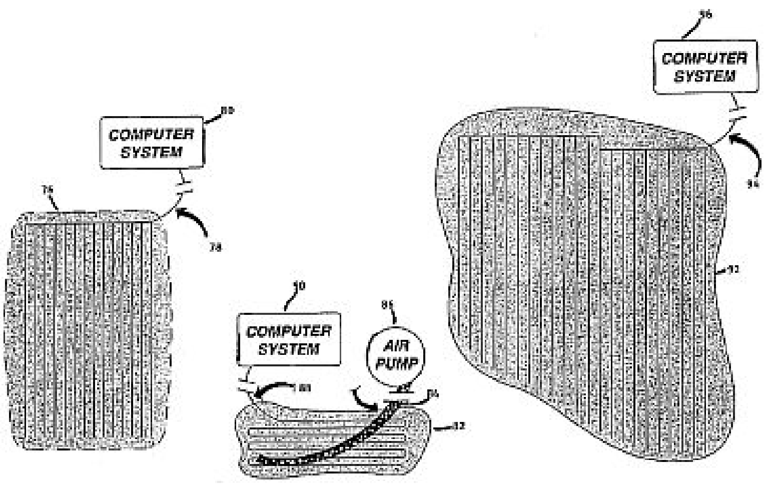

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Reference will now be made in detail to the present preferred embodiments of the present invention, examples of which are illustrated in the accompanying drawings.

Whenever possible, the same reference numbers will be used throughout the drawings to refer to the same or like parts.

Shown in Fig. 1 is a first preferred embodiment of the present invention. It illustrates the application of pulsed electromagnetic energy to any aquatic setting. Shown generally at 10, this embodiment includes a water reservoir 12, such as an aquarium, pool, pond or other body of water. Below water reservoir 12 is placed electric wiring 16. Although in Fig. 1 electric wiring 16 is shown in a continuous open ended coil of wiring, electric wiring 16 can be positioned in a discontinuous fashion having a main line and a plurality of other lines extending from the main line. The electrical wiring 16 is placed over a substantial area of water reservoir 12, from 1 to 30 inches below the bottom surface of the water reservoir 12 and is insulated from the surrounding elements and cross currents from other cables. If water reservoir 12 has sides, such as in a pool, electric wiring 16 can also be placed behind the surface of the sides of reservoir 12. Electric wiring 16 is insulated by neutral materials, such as plastics, that will insulate electric wiring 16, yet not produce a charge in the surrounding soil or substance. Further, highly shielded electrical wire is not desired as it decreases the capacity of diffusion of the pulsed electromagnetic energy. The pulsed electromagnetic energy is supplied to electric wiring 16, and is controlled by an electromagnetic energy generator 14.

Optionally, an air pump 18 and perforated air hose 20 may be included in the system embodied in Figure 1. Air pump 18 and air hose 20 provide an oxygen source to the water with in reservoir 12, thus, aerating the water and increasing the oxygen concentration of the water.

In use, pulsed electromagnetic energy is applied to the electric wiring 16 by an electromagnetic energy generator 14. Various electronic components maybe utilized to generate electromagnetic energy for use in this system, such as a computer system or a electromagnetic wave generator for example. Preferably, the electromagnetic energy has a magnetic field component in the range of 0.5 to 30 Gauss, and the frequency range of the pulses between 0.5 and 30 Hertz. Concurrently, air pump 18 pumps air through hose 20 to increase the oxygen concentration of the water. The water in free motion around the air bubbles is energized to a higher state because of the application of the pulsed electromagnetic energy, and thus becomes even more saturated with oxygen.

Application of the electromagnetic energy to reservoir 12 provides for application of electromagnetic energy to both plants and animals situated within the aquatic system. The affected plants and animals exhibit enhanced physiologic effects such as increased growth and overall health for example. Furthermore, the highly oxygenated and energized water can then be utilized for watering plants which are not directly exposed to electromagnetic energy.

Optionally the system and method of the present invention contemplates the sustained or intermittent introduction of acoustic energy into the aquatic setting or the environment of other embodiments of the present invention. Many forms of soothing sounds may provide enhanced benefit but most preferred is the use of symphonic music or sounds, preferably with minimal or no dischord. The introduction of such can be accomplished through a sound system 100 as is schematically illustrated in Fig. 6. Sound system 100 comprises a power supply 102, which may be independent or shared with other systems utilized in the environment, that is operably connected to a sound control panel 104 from which one may control the selection and characteristics of the sounds broadcast by sound system 100 into the surrounding environment of the present invention. Control panel 104 is connected to one or more speakers 106 for broadcasting the generated sound. In aquatic settings underwater speakers are contemplated.

Although not required, one or more microphones 108 for receiving sound from the aquatic setting or other environment is contemplated. Such sounds may be recorded as is known in the art.

Additionally the sustained or intermittent introduction of orgone or radionic energy into the aquatic setting of the environment or other embodiments of the present invention, especially but not necessarily in pulsed form, is contemplated. Such radionic energy is supplied through one or more radionic generators or collectors as are lcnown and available in the art. Exposure time and intensity may be varied.

Referring now to Fig. 2, there is shown a second embodiment of the present invention, illustrating the application of pulsed electromagnetic energy to livestock animal pens. In this embodiment, there is a stock pen 24 which can be used to retain any type of livestock animal or poultry. Further, it should be understood that such a stock pen can be enclosed in buildings or open fields. Electrical wiring 26 extends below the surface and laterally across first stock pen 24 to cover at least a substantial area of stock pen 24. As in the first embodiment, electrical wiring 26 is comprised of electrical wiring having neutral insulating materials, and is placed under the surface of the selected area at a preferred depth of 1 to 30 inches. Electromagnetic energy generator 28 generates and transmits electromagnetic energy to electric wiring 26. Optionally, this system may also include a water tank 34 for watering of the animals. An air pump 30 may be utilized to pump air into the water tank 34 through hose 32. Aeration of water tank 34 by pump 30 increases the oxygen concentration of the water held within water tank 34.

If the stockyard is enclosed, the electric wiring is buried at a depth between 1 and 30 inches within the dirt or other flooring material, such as concrete, for example.

Now referring to Fig. 3, there is shown a third embodiment of the present invention illustrating the application of electromagnetic energy to a plant bio-system 46, such as a garden.

The plant based bio-system may include, for example, a grain plot 48, fruit or other trees 56, plants 58 and/or flowers 60. The plants may either be grown within the ground itself or within planting containers such as pots and the like. The system 46 may optionally also include a water reservoir 50, which is aerated by hose 52 and air pump 54 to increase the oxygen concentration of the water held within reservoir 50. Electromagnetic energy is generated by an electromagnetic energy generator 64 and applied to system 46 via electric wiring 62. Electric wiring 62 is position to cover at least a substantial portion of system 46, and is placed below the surface of system 46, at a depth of 1 to 30 inches. The plants affected by the application of the electromagnetic energy exhibit improved physiological effects such as improved and more rapid growth, and better overall health for example.

Shown in Fig. 4 is yet another embodiment of the present invention, illustrating applying electromagnetic energy to an athletic playing surface 68, such as, for example a basketball court, a football field, a soccer field, a swimming pool, playgrounds or other playing surfaces.

Electrical wiring 70 is placed below playing surface 68 at a preferred depth of between 1 and 30 inches. Electromagnetic energy generator 72 generates and conducts electromagnetic energy through electric wiring 70 to diffuse the electromagnetic energy throughout playing surface 68.

The electromagnetic energy is applied to those persons playing or competing on playing surface 68. Upon application of the electromagnetic energy, humans experience higher energy levels for longer periods of time, reduced fatigue, less muscle strain and soreness, in addition to increased concentration and precision in playing the particular sport.

Referring now to Fig. 5, there is shown yet another embodiment of the present invention, illustrating application of electromagnetic energy to a golf course hole. This embodiment includes a tee box 76, a green 92 and/or optionally a water reservoir 82. Below tee box 76 is placed electrical wiring 78, which is placed at a depth of 1 to 30 inches. Electrical wiring 94 is also placed below golf green 92 at a preferred depth of 1 to 30 inches. If water reservoir 82 is utilized, electrical wiring 88 is positioned below the bottom of reservoir 82 at a depth of 1 to 30 inches for emission of electromagnetic energy. Treated water from reservoir 82 may be used to water the golf course fairways, greens, or other plants associated with the course as to achieve the benefits as described above in relation to the plant based bio-system embodiment.

Electromagnetic energy generators 80,90 and 96 produce and emit electromagnetic energy through electrical wiring 78,88 and 94 respectively. Optionally, a single electromagnetic generator can control the emission of electromagnetic energy through electrical wiring 78,88 and 94. As with the above described embodiments, electromagnetic energy generators 80,90 and 96 emit electromagnetic energy having a magnetic field component in the range of 0.5 to 30 Gauss, and the frequency range of the pulses between 0.5 and 30 Hertz.

Effects seen through application of the electromagnetic energy to a grass surfaces such as soccer fields, play grounds, football fields, golf greens and tee boxes include more rapid and healthier growth of grass, faster regeneration or repair of divots and ball marks, fewer attacks to these grasses by pests as the grasses are healthier. The human players experience gentle invigoration, increased energy, greater concentration, and less muscle soreness or strain.

Increased mental activity and faster healing of wounds has also been noted.

Sporting equipment such as golf clubs are not affected by the application of electromagnetic energy because no sustained electrical current is conducted to the metal portions of the clubs.

In use with any of the above embodiments, the characteristics of the electromagnetic energy remain the same, that is electromagnetic energy having a magnetic field component in the range of 0.5 to 30 Gauss, and the frequency range of the pulses between 0.5 and 30 Hertz. The cycling/hertz of the electromagnetic energy is varied slowly during a 24 hour period. In application of pulsed electromagnetic energy to humans, the magnetic field strength is adjusted to be within 4 to 8 Gauss. The placement of the electrical wiring below the surface of the selected area is adjusted to accommodate these parameters. Further, it is important to note that the spacing of the electric wiring in the above described embodiments may be altered to achieve desired effects.

Although the principles, preferred embodiments and preferred operation of the present invention have been described in detail herein, this is not to be construed as being limited to the particular illustrative forms disclosed. They will thus become apparent to those skilled in the art that various modifications of the preferred embodiments herein can be made without departing from the spirit or scope of the invention as defined by the appended claims.

YouTube Videos

www.youtube.com/watch?v=AzvytVTHOmA?

www.creationevidence.org ... Hard Questions fo

www.youtube.com/watch?v=AjheQjMwHJA?

www.youtube.com/watch?v=mAUR-jyIVm8?

www.youtube.com/watch?v=h5c9yKhGewk?

www.youtube.com/playlist?list=PLD1E0A4EFA11EB0EC?

www.youtube.com/watch?v=ghh2sM6IjUQ?

www.youtube.com/watch?v=8xL6sibLb6o?

www.youtube.com/watch?v=9M7W0TuTFzI?

www.youtube.com/playlist?list=PLD1E0A4EFA11EB0EC.

www.youtube.com/watch?v=Rk5fd5Cy_z4?

---

http://en.wikipedia.org/wiki/Carl_Baugh

Carl Baugh

Born Carl Edward Baugh
October 21, 1936 (age 77)
Kenedy, Texas, USA

Carl Edward Baugh (born October 21, 1936) is an American young earth creationist. Along with others, Baugh claims to have discovered human footprints alongside dinosaur footprints near the Paluxy River in Texas, but never provided evidence for his alleged findings.[1][2][3] Baugh is a national television host, purporting to present science supporting creationism on the program Creation in the 21st Century on the Trinity Broadcasting Network. His claims have been debunked by the scientific community and other creationists as pseudoscience.[4] His educational credentials have also been called into question.[5] He was also the president and a graduate of Pacific International University, an unaccredited university located in Springfield, Missouri.

Biography

Born in Kenedy, Texas, Baugh graduated in 1955 from Abilene High School in Abilene, Texas[citation needed]. He currently appears on Trinity Broadcasting Network program Creation in the 21st Century.[6] Baugh was also the president and a graduate of the Pacific International University, which was a non-accredited organisation.[7][8]

In 1984, Baugh started the Creation Evidence Museum in a double-wide trailer in Glen Rose, Texas, near Dinosaur Valley State Park, to promote creationism.[4] All of the museum exhibits have been strongly criticized as incorrectly identified dinosaur prints, other fossils, or outright forgeries.[9][10] In 2008, a descendant of a family that found many original Paluxy River dinosaur tracks in the 1930s claimed that her grandfather had faked many of them.[11] Others, such as purported dinosaur claws, were identified by University of Texas at Austin paleontologist Wann Langston as crocodile teeth.[12]

In 1996, Baugh presented his "man-tracks" in the controversial program The Mysterious Origins of Man.[13] Creationist Ken Ham, of Answers in Genesis, criticized the claims in a review titled "Hollywood's 'Moses' Undermines Genesis," regarding Baugh: "According to leading creationist researchers, this evidence is open to much debate and needs much more intensive research. One wonders how much of the information in the program can really be trusted!"[14] He also has been given television exposure by the tele-evangelist Kenneth Copeland. He has authored several self-published books[4] on such topics as the age of the universe, dinosaurs coexisting with humans and critiques of evolution.[15]

In 2001 Baugh and Creation Evidence Museum were featured on The Daily Show where Baugh likened human history to The Flintstones and the show poked fun at his claims about the hyperbaric biosphere, pterodactyl expeditions, and dinosaurs.[16]

He is a promoter of intelligent design. In 2002 he appeared with William A. Dembski at a conference in Texas and has built his more recent web material around ID and Dembski.[17]

The Accelerated Christian Education schools teach his theories as fact in some of their High School science workbooks.

Claims and criticism

Both scientists and creationists have criticized Baugh's claims. In 1982–1984, several scientists, including J.R. Cole, L.R. Godfrey, R.J. Hastings, and S.D. Schafersman, examined Baugh's purported "mantracks" as well as others provided by creationists in the Glen Rose Formation.[12] In the course of the examination "Baugh contradicted his own earlier reports of the locations of key discoveries" and many of the supposed prints "lacked human characteristics."[12] After a three-year investigation of the tracks and Baugh's specimens, the scientists concluded there was no evidence of any of Baugh's claims or any "dinosaur-man tracks".[12]

On September 27, 1984, Al West, a Baugh co-worker for two years, who followed the mantrack claims since 1974, and friend of Glen Kuban, publicly announced that Baugh "never had evidence for manprints as claimed.[12] Gayle Golden, writer for The Dallas Morning News, reported that Baugh "paid $10,000 for his Moab skeleton and confirmed that Baugh knew at their purchase that the bones had already been dated at 200-300 years. However Baugh later claimed that the bones were found in Cretaceous deposits."[12]

One of Baugh's more famous claims, aside from the dinosaur tracks, is an alleged out of place artifact of an "18th century miner's hammer" found in million-year-old Ordovician rock (he has also claimed it is in Cretaceous rock) found in 1934 from London, Texas.[18][19] Baugh asserted this as evidence against scientifically known ways that rocks form.[18] However, laboratory tests discounted his claim about the hammer's being formed in the rock.[18][19] J.R. Cole wrote, "The stone concretion is real, and it looks impressive to someone unfamiliar with geological processes. How could a modern artifact be stuck in Ordovician rock? The answer is that the concretion itself is not Ordovician. Minerals in solution can harden around an intrusive object dropped in a crack or simply left on the ground if the source rock (in this case, reportedly Ordovician) is chemically soluble."[20]

In July 2008, Baugh was in contact with Alvis Delk and James Bishop, who claimed to have found a dinosaur-human print fossil.[21] Baugh bought the "fossil" from Delk who used the money to pay his medical bills.[22] On the authenticity of the claims, reporter Bud Kennedy noted, "since no scientists were involved, about all we really know so far is that the museum has a new rock."[23] This was deemed "not a convincing human footprint in ancient rock" by biologist Glen J. Kuban and called a "blatant fake" by biologist PZ Myers.[24]

Creationist organizations such as Answers in Genesis have criticized Baugh's claims saying he "muddied the water for many Christians. . . . People are being misled."[4] Don Batten, of Creation Ministries International wrote: "Some Christians will try to use Baugh's 'evidences' in witnessing and get 'shot down' by someone who is scientifically literate. The ones witnessed to will thereafter be wary of all creation evidences and even more inclined to dismiss Christians as nut cases not worth listening to."[25] Answers in Genesis (AiG) lists the "Paluxy tracks" as arguments "we think creationists should NOT use" [emphasis in original].[26] Also Answers In Creation reviewed Baugh's museum and concluded "the main artifacts they claim show a young earth reveal that they are deceptions, and in many cases, not even clever ones."[27]

In his 1992 book Panorama of Creation, Baugh claims that a layer of metallic hydrogen surrounded the early earth. Furthermore, he professes that hexagonal water, or, "Creation water" as he calls it, is capable of healing. Such claims have been addressed by scientists as pseudoscience,[8] and his hypotheses and credentials are not accepted in academia.

Baugh has claimed several college degrees, at one point professing to earning three doctorates, when no evidence exists that he even graduated from high school.[28] All three "doctorates" are from unaccredited "schools." One is an honorary "Doctor of Philosophy in Theology" from the non-accredited California Graduate School of Theology. His claimed 1989 "doctorate" and masters degrees in Archaeology come from the non-accredited Pacific International University, of which Baugh was also the president.[28] His dissertation titled "Academic Justification for Voluntary Inclusion of Scientific Creation in Public Classroom Curricula, Supported by Evidence that Man and Dinosaurs Were Contemporary" was reviewed by Brett Vickers who criticized its "descriptions of his field-work on the Paluxy river 'man-tracks', speculation about Charles Darwin's religious beliefs and phobias, and biblical evidence of Adam's mental excellence."[29] In 2005, Baugh claimed to have completed yet another doctorate in theology from the unaccredited diploma mill known as Louisiana Baptist University.[30]

Bibliography

Dinosaur Promise Publishing (paperback) — 1987, 152 pages, ISBN 0-939497-01-8

Panorama of Creation Hearthstone Publishing, Ltd. (paperback) — 1992, 91 pages, ISBN 1-879366-01-0

Jurassic Park: Fact Vs. Fiction (with Bill Uselton) Hearthstone Publishing, Ltd. (paperback) — 1993, 30 pages, ISBN 1-879366-35-5

Footprints and the Stones of Time (with Clifford Wilson) Hearthstone Publishing, Ltd. (paperback) — 1994, 162 pages, ISBN 1-879366-17-7

Why Do Men Believe Evolution Against All Odds? Hearthstone Publishing, Ltd. (hardcover) — 1999, 160 pages, ISBN 1-57558-049-7

Footnotes

May, David (2008-07-28). "Rock-solid proof?". Mineral Wells Index. Retrieved 2008-09-21.

Stromberg, Pierre; Paul V Heinrich (2008). "The Coso Artifact: Mystery From the Depths of Time?". National Center for Science Education. Retrieved 2008-09-21.

"Indiana Legislation Contemplating ID?". National Center for Science Education. November 5, 2005.

"Footprints of Fantasy", by Kaylois Henry, Dallas Observer, December 12, 1996. Retrieved August 15, 2008.

"Creationism Alive and Kicking in Glen Rose", by Greg Beets, August 5, 2005, Austin Chronicle.

"Carl Baugh: Creation in the 21st Century". Trinity Broadcasting Network. 2008. Retrieved 2007-02-19.

George Brown. Protecting Australia's Higher Education System from Australian Universities Quality Agency

A Matter of Degree-Carl Baugh's Alleged Credentials from talk.origins (Originally published in NCSE Reports Vol 9, No. 6, Nov–December 1989.)

"Seeking Ancient Paths", John R. Armstrong, PSCF 41 (March 1989): 33-35, American Scientific Affiliation. Retrieved August 15, 2008.

"The Rise and Fall of the Paluxy Mantracks", by Ronnie J. Hastings, PSCF 40 (September 1988): 144-154. American Scientific Affiliation. Retrieved August 15, 2008.

"Human footprints along with dinosaur tracks?", by Bud Kennedy, August 10, 2008, Fort Worth Star-Telegram.

"Creation/Evolution". National Center for Science Education. Issue 15 (Volume 5, Number 1 — Winter 1985). Retrieved 2007-02-19.

"A Review of NBC's 'The Mysterious Origins of Man'", 1996, Glen J. Kuban. Retrieved August 15, 2008.

Thomas, Dave (March 1996). "NBC's Origins Show". Committee for Skeptical Inquiry. Retrieved 2007-02-19.

Greg Neyman. Creation Science Exposed: Creation Evidence Museum Lacks Evidence!. Answers in Creation.

"Tyrannosaurus Redux". The Daily Show. November 14, 2001. Retrieved 2007-02-19.

Barbara Forrest and Paul R. Gross. Creationism's Trojan Horse: The Wedge of Intelligent Design. 2004, page 293

"If I Had a Hammer". National Center for Science Education. Issue 15 (Volume 5, Number 1 — Winter 1985). Retrieved 2007-02-19.

"The London Hammer: An Alleged Out-of-Place Artifact". The Paluxy Dinosaur/"Man Track" Controversy. 14 July 2006.

Cole, J. R. 1985. "If I had a Hammer" Creation/Evolution, Issue XV, pp.46–47

Baugh, Carl (2008). "Alvis Delk Cretaceous Footprint". Creation Evidence Museum.

"One step at a time". Mineral Wells Index. August 11, 2008.

"Human footprints along with dinosaur tracks?". Star-Telegram. August 10, 2008.

"Transparent fakery". Pharyngula (blog). July 28, 2008.

"What About Carl Baugh?". Creation Ministries International. 1998.

"Arguments we think creationists should NOT use". Answers In Genesis. 2008.

Neyman, Greg (2008). "Review of Carl Baugh's Museum". Answers In Creation.

Matter of Degree: Carl Baugh's Alleged Credentials by Glen J. Kuban, NCSE Reports Vol 9, No. 6, Nov–December 1989. (Updated August 2006)

Vickers, Brett (May 31, 2002). "Some Questionable Creationist Credentials". talk.origins.

Tobia, P.J. "Reading, Writing and Jesus: What nearby schools don’t know about the Bible class they soon may be teaching", Nashville Scene, October 19, 2006.

External links

Creation Evidence Museum Baugh's official website
http://www.creationevidence.org/

Baugh's "doctoral dissertation" from Archive.org
http://web.archive.org/web/*/http://www.drcarlbaugh.org/

Analysis of Baugh's Dinosaur Tracks, National Center for Science Education Issue 15 (Volume 5, Number 1 — Winter 1985).
http://ncse.com/cej/5/1

A Matter of Degree: Carl Baugh's Alleged Credentials by Glen J. Kuban, NCSE Reports Vol 9, No. 6, Nov–December 1989. (Updated August 2006)
http://paleo.cc/paluxy/degrees.htm

What About Carl Baugh? by Dr Don Batten, of Creation Ministries International
http://paleo.cc/paluxy/whatbau.htm

Cretinism or Evilution? No. 3 from talk.origins
http://talkorigins.org/faqs/ce/3/part2.html

What about Carl Baugh? talk.origins
http://www.talkorigins.org/faqs/paluxy/whatbau.html

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

US005307162A

United States Patent [19]

Schowengerdt

[11] **Patent Number: 5,307,162**

[45] **Date of Patent: Apr. 26, 1994**

[54] **CLOAKING SYSTEM USING OPTOELECTRONICALLY CONTROLLED CAMOUFLAGE**

[76] Inventor: **Richard N. Schowengerdt**, 3145 Kerry La., Costa Mesa, Calif. 92626

[21] Appl. No.: **977,192**

[22] Filed: **Nov. 16, 1992**

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 682,945, Apr. 10, 1991, abandoned.

[51] **Int. Cl.**^{5} **H04N 7/18**

[52] **U.S. Cl.** **348/122**; 472/61; 348/586

[58] **Field of Search** 358/93, 108, 225, 230, 358/87; 472/57, 61, 65, 71, 75, 77, 81; H04N 7/18

[56] **References Cited**

U.S. PATENT DOCUMENTS

2,591,068 4/1952 Hilchey 272/17
2,797,612 7/1957 Holmes 272/10
3,352,965 11/1967 Driver et al. .... 358/93 X
4,962,420 10/1990 Judenich .... 272/11 X

Primary Examiner—Victor R. Kostak

[57] **ABSTRACT**

The Cloaking System is designed to operate in the visible light spectrum, utilizes optoelectronics and/or photonic components to conceal an object within it, and employs analog or digital control feedback resulting in camouflage adaptable to a changing background. The system effectively conceals either a still or moving object from view by the interposing of a shield between an observer and the object and recreating a full color synthetic image of the background on the shield for viewing by observer, thus creating the illusion of transparency of both the object and the Cloaking System. This system consists of four major elements: a sensor; a signal processor; a shield; and a means of interconnecting, supporting, and safely enclosing the aforementioned elements along with the concealed object.

10 Claims, 3 Drawing Sheets

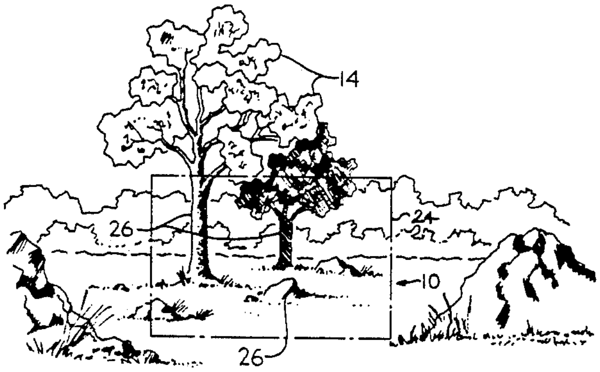

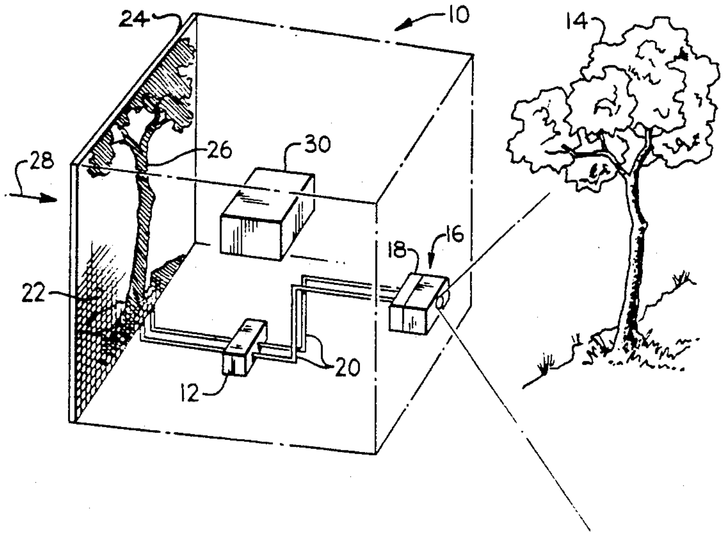

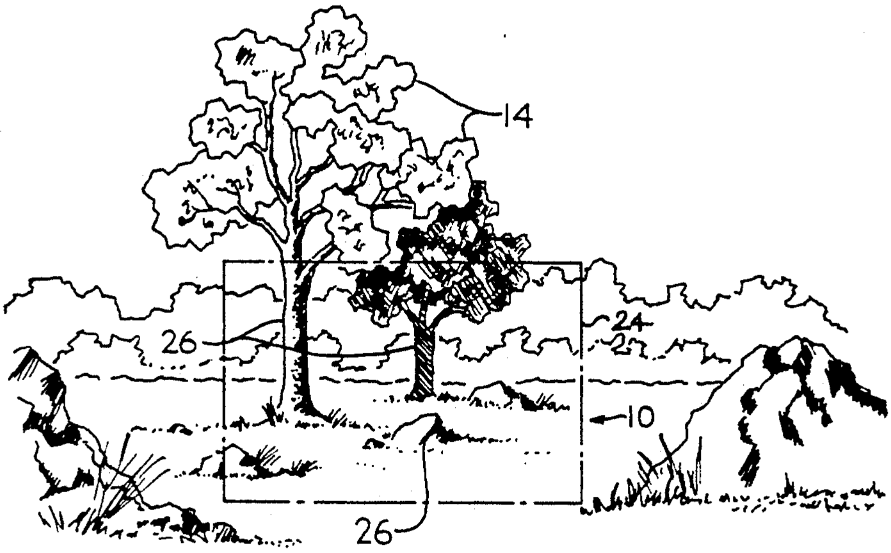
14
24
22
26
10
26

FIG. 1

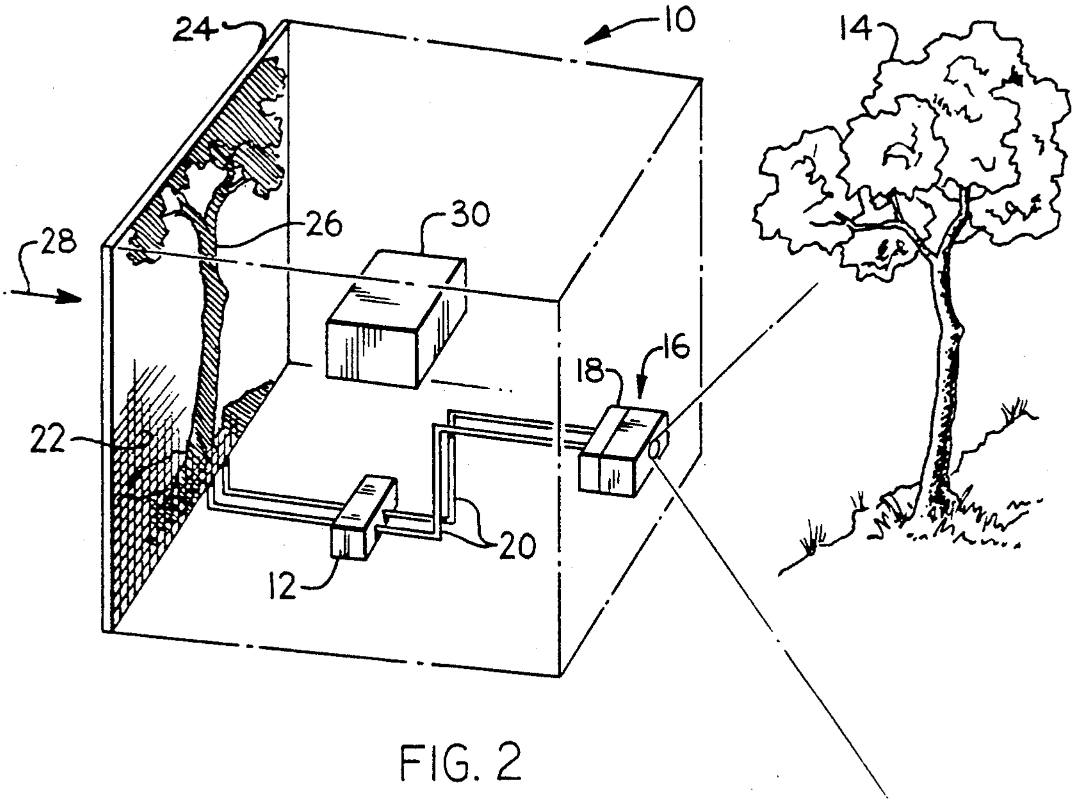
24
10
14
28
26
30
16
18
22
20
12

FIG. 2

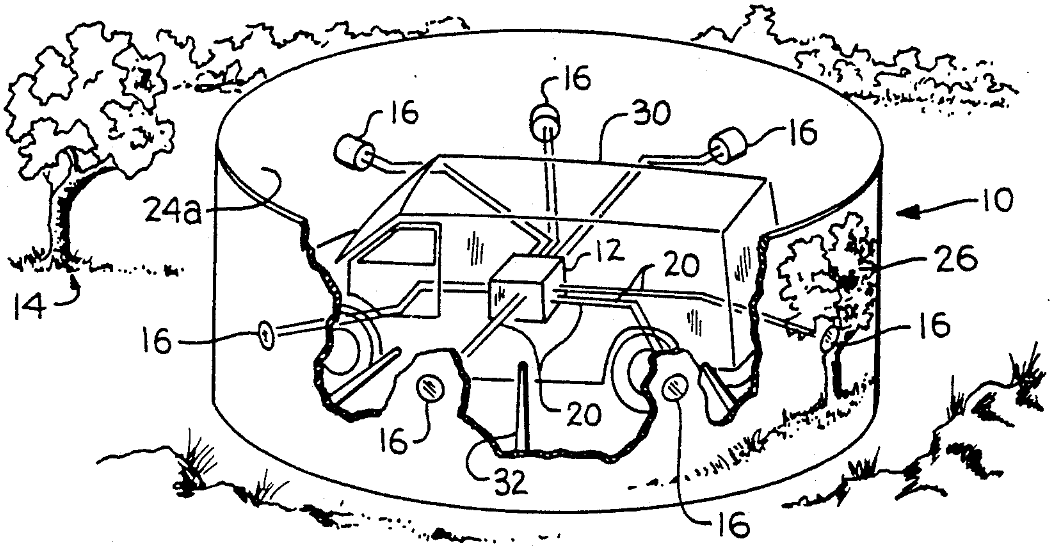

FIG. 3

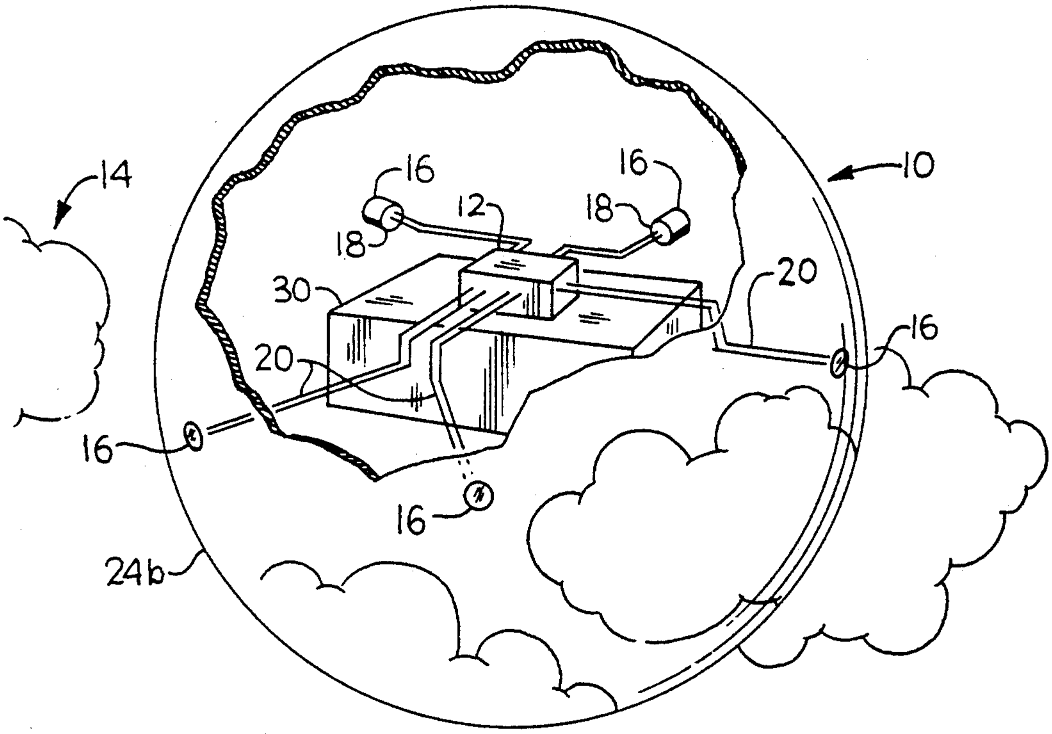

FIG. 4

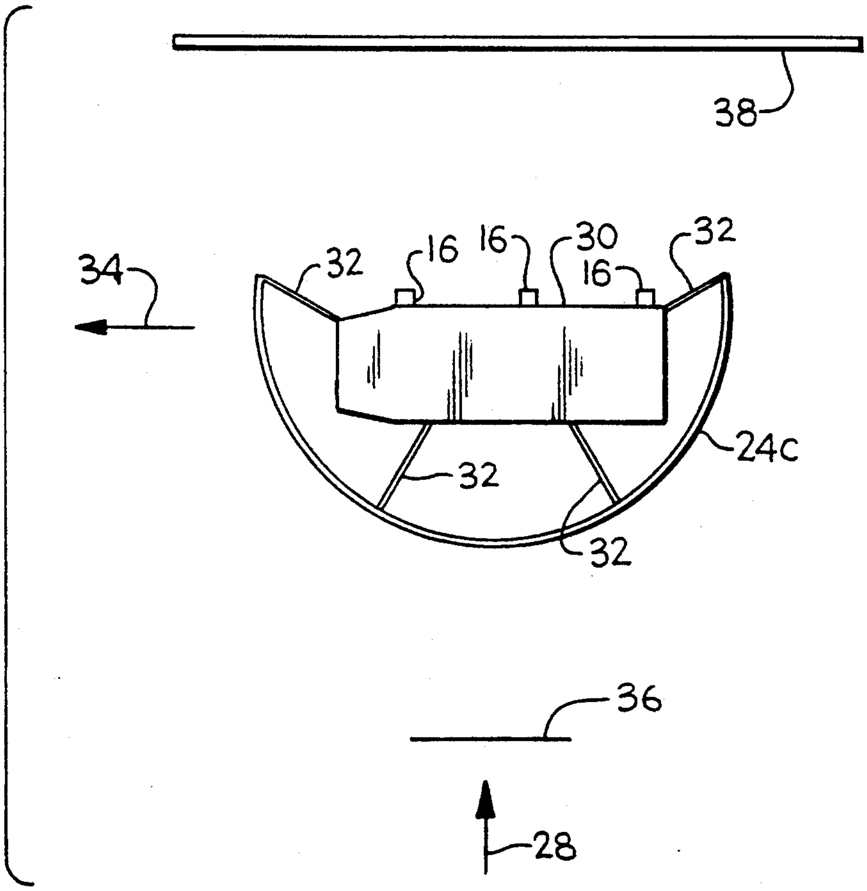
38
34
32
16
16
30
16
32
24c
32
32
36
28

FIG. 5

CLOAKING SYSTEM USING OPTOELECTRONICALLY CONTROLLED CAMOUFLAGE

This is a continuation in part of application Ser. No. 07/682,945, filed Apr. 10, 1991, **now abandoned.**

REFERENCES CITED

1. U.S. Pat. No. 2,591,068.
2. U.S. Pat. No. 2,797,612.
3. U.S. Pat. No. 3,352,965.
4. U.S. Pat. No. 4,962,420.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The need for this invention arises out of the present state of the art of military camouflage in the visible light spectrum. This invention is designed for the purpose of concealing both stationary and moving objects from view. The term "cloaking" as used in the title of this invention and elsewhere herein refers to concealing such objects.

2. Prior Art

The techniques in use today largely involve painting, coloring, and/or contour shaping to allow an object to better blend in with the background. Such techniques are rather primitive and do little to conceal a moving object as it is not possible with present fixed state camouflage techniques to cause an object to continually blend in with a changing background. In order to effectively camouflage a moving object, its appearance must be constantly controlled from the viewpoint of the observer to blend in with the changing background from the observer's perspective. Little has been done to utilize modern advancements in optoelectornics, computers, or microminiature components to camouflage an object. The proposed system was conceived with the view in mind of effectively concealing a stationary or moving object such as a man, vehicle, or aircraft from view by utilization of this technology. This invention will find practical application in military and law enforcement where field commanders find it useful to conceal such an object from view.

SUMMARY OF THE INVENTION

This system is designed to interrupt energy in the visible light electromagnetic spectrum to effectively conceal an object from view by an observer. This is accomplished by insertion of a shield between an observer and the object being concealed and recreating a full color synthetic real-time image of the background on the shield for view by an observer, thus creating the illusion of transparency of the object. The basic concept for the application consists of four basic building blocks: a sensor, a signal processor, and a shield which are interconnected through an analog, digital, of fiber optic data bus, and a supporting structure for all system elements, the object being concealed being interposed between the background and the shield.

The signal processor formats the image of the background as viewed by the sensor and presents a synthetic image of the background on the shield for view by an observer. The object is placed behind or within the shield which is of opaque construction, and having a nonspecular display surface thus concealing the object from view by the observer and yet creating an illusion of transparency of the complete Cloaking System and the enclosed object. In its nonactivated state, the shield would have minimum light reflectivity. Although either an analog signal processor or digital signal processor may be used, the latter is described herein as one specific embodiment. The digital signal processor constantly compares the synthetic image in the shield with the reference image within the sensor and makes corrections to the synthetic image to correspond with the actual image being sensed. Data on the data bus blows in both directions. Special software or embedded firmware will also be utilized in the signal processor to make changes in the image presented on the shield to correct for distance and angle of the observer with respect to the object being concealed; the source of intelligence information pertaining to an observer is beyond the scope of this invention but is mentioned here only for clarification.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a pictorial representation of the invention and is intended to illustrate the basic concept of the invention;

FIG. **2** is a pictorial isometric view of the system block diagram implementing the basic concept;

FIG. **3** is a pictorial perspective view of the invention and illustrates how an object can be concealed within a 360° shield;

FIG. **4** is a perspective view of the invention and illustrates how an object can be concealed within a sphere; and

FIG. **5** is an orthographic view that illustrates a simple concept demonstration model.

DESCRIPTION OF ONE EMBODIMENT OF THE INVENTION

The basic concept of the Cloaking System **10** is illustrated in FIG. **1** wherein the background **14** is displayed on the shield **24** as a synthetic image **26** on its nonspecular display surface, when viewed from the front. FIG. **2** illustrates how the basic concept is implemented. The digital signal processor **12** is the brain and controls all operations, digitally formats the image of the background **14** as viewed by the sensor **16**, establishes a reference image matrix **18** within the sensor **16**, and transmits the background image signal via the data bus **20** to a synthetic image matrix **22** within the shield **24**, thus causing a synthetic image **26** to be displayed on the surface of the shield **24** as seen from the direction of observation **28**. The object **30** to be concealed is placed behind the shield **24** to prevent its being observed. The digital signal processor **12** constantly compares the synthetic image matrix **22** with the reference image matrix **18** and makes corrections to the synthetic image matrix **22** which in turn corrects the synthetic image **26** to correspond with the actual background image **14** being sensed, thus minimizing distortion and maintaining realism of the synthetic image **26**. Data on the data bus **20** flows in both directions. The primary purpose of the system is to present on the shield **24** a near identical image of the background behind said object **30** so as to create the illusion of transparency of both said object **30** and the Cloaking System **10**; a secondary purpose of the system is to alter the synthetic image **26** to create an illusion of a background that does not exist; while such a deceptive method goes beyond the scope of simply creating an illusion of transparency of the object it would be very useful for special missions or difficult scenario conditions where it would be more effective to depict an artificial background than to display a representation of the actual background. Any such alterations in the synthetic image **26** would be accomplished by software or embedded firmware and is considered part of the tactical strategy to be employed, thus emphasizing the flexibility of the invention and its potential interaction with the operating environment.

The basic concept or cell unit depicted in the preceding FIG. **2** may be expanded to combine as many cells as desired to provide 360° coverage. Referring to FIG. **3**, a cylindrical shield **24***a* is illustrated with a cutaway view showing object **30** along with the digital signal processor **12**, data bus **20**, and sensors **16**, all supported by a suitable attachment means **32**. The digital signal processor **12** organizes and presents a different synthetic image **26** on the appropriate sector of the shield **24***a* utilizing signals from appropriate sensors **16**, together with tactical data concerning potential points of observation. The appropriate size and number of sectors would be determined by the optimal design for a given operational mission. Although shown here as separate components for clarity, the present state-of-the-art would allow incorporation of the an embedded circuitry digital signal processor **12** and sensors **16** into the shield **24***a* to provide a cylindrical smart skin to wrap around the object. The shield may also incorporate backplanes for mounting embedded sensors, processors and bus circuitry.

Carrying the basic concept a step further, the shield **24***b* may be formed into a spherical shape as illustrated in FIG. **4**. A cutaway view is again presented showing object **30**, sensors **16**, digital signal processor **12**, data bus **20**, and a spherically shaped shield **24***b*. As with the cylindrical shape, the digital signal processor **12** organizes and presents a different synthetic image **26** on the appropriate sector of the shield **24***b* utilizing both tactical data and signals from appropriate sensors **16**. In this configuration there would be considerably more sectors, as compared with the cylindrical shape, and the programming would be more complex to provide a high degree of realism in the image.

Irregular shapes for the shield may also be employed utilizing conformal smart skins which would be thin digital signal processor circuits embedded within the shield display surface. However, in order to reduce distortion of the image, either a circular, cyclindrical, or spherical shape would allow uniform algorithms to be employed in the embedded processor programs in the digital signal processor **12** to correct for circular or spherical aberration. Some distortion may appear as ripples or waves in the image, depending upon the shape of the shield, relative position and motion of the observer, and complexity of background images. However, in the algorithms could be developed, such as those required for irregular or conformal shields for military vehicles, to minimize such distortion.

ALTERNATE TECHNOLOGIES AND EMBODIMENTS

Several alternate technologies and methods of implementing this invention may be utilized.

One alternate technology employs the simplest method of implementing the Cloaking System concept which would utilize a conventional analog signal processor, a specially designed liquid crystal display to present the background image on the shield, and a conventional video camera for the sensor. Since this method utilizes conventional analog technology, this method is limited to flat plane or semicircular displays and would not lend itself to image correction as could be achieved by the digital options. However, this method would allow the construction of a prototype model within a year or two to demonstrate the basic cloaking concept.

Another alternate technology employs a digital option; the main differences between this method and the first-described analog method consists of the conversion of all analog video signals to digital signals, the usage of a digital signal processor to direct all operations, and higher technology used in constructing the shield and the sensor. Because of the high flexibility of the digital method, a great variety of shapes could be used for the shield and various compensation software or embedded firmware schemes could be employed to minimize distortion in the image; hence the digital method was selected and described in the preceeding section in considerable detail as one embodiment of the invention.

In still another method, photonic materials could be used for constructing the shield and sensor which would be capable of directly manipulating light signals and transmitting these signals on a higher speed fiber optic data bus. These light signals could be then digitized and further manipulated by a digital signal processor to perform necessary system functions but without the encumbrance of the slower speed of a digital data bus. The selection of either optoelectronic or photonic building blocks for the shield, sensor, and data bus would depend upon such factors as size and shape of the object to be shielded, speed of the object relative to the background, and environment. Usage of either optoelectronic or photonic materials in extremes such as that encountered in combat would require extensive testing and development. However, for more benign applications, a state-of-the-art system could be developed within three to five years utilizing currently available materials and components.

Any one of several available commercial or military digital process and bus architectures could be used depending upon the application and program requirements. IBM-PC integrated circuits and bus standards such as RS-232 (Serial), RS-422 (Parallel), or IEEE 488 (Parallel could be used for lower cost and benign environments. In more rigorous combat environments Very High Speed Integrated Circuit (VHSIC) architecture and Military standards such as MIL-STD-1553, MIL-STD-1750, or MIL-STD-1760 could be used for complex interfaces between men and machines. Sophisticated and advanced models using fiber optic technology and a high speed data bus would have advantages in terms of small size and conformal shape adaption of the shield.

CONCEPT DEMONSTRATION MODEL

A simple Concept Demonstration Model could be constructed utilizing any of the above-described technologies and methods. This model is shown in FIG. **5**; the object **30**, such as a vehicle, is shown in a top view with a semicircular shield **24***c* secured by an attachment means **32**, such as supporting struts. Sensors **16** view the wall **38** and the analog signals are subsequently displayed on the shield **24***c* facing the point of observation **28** behind line **36** which is fifty feet from the vehicle **30**. The vehicle **30** is moving to the left as shown by the arrow **34**. When the system is operating properly, the vehicle **30** would never be observed; only the wall **38** behind said vehicle **30** would be seen. For simplicity and to establish a reference, the wall **38** would probably be initially monochromatic and color patterns added later to present more difficult scenarios and to measure system limitations.

LIST OF REFERENCES CITED

1. U.S. Pat. No. 2,591,068, Hilchey, Eugene L., *Studio Background,* Apr. 1, 1952.
2. U.S. Pat. No. 2,797,612, Holmes, Oscar J., *Intermittent Motion Picture Film Feed Apparatus,* Jul. 2, 1957.
3. U.S. Pat. No. 3,352,965, Driver, Paul C., and Fowler, Robert E., *Method And Means For Enhancing Camouflaged Target Detection Utilizing Light Polarization Techniques,* Nov. 14, 1967.
4. U.S. Pat. No. 4,962,420, Judenich, Gennady I., *Entertainment Video Information System Having A Multiplane Screen,* Oct. 9, 1990.

I claim:

1. A system for concealing an object from view by an observer through use of a background scene and comprising:

means for sensing a background scene and for generating a video image signal therefrom;

means for creating a processed image of said background video image signal for display purposes; and

means for displaying an image represented by said processed image signal, the generated image being part of the sensed field of view, the object located between the means for sensing and the means for displaying, the object thereby being substantially camouflaged by the image displayed.

2. A concealing system according to claim **1** wherein said background scene sensing means comprises means for converting background light images electrical analog image signals for processing by an analog signal processor.

3. A concealing system according to claim **1** wherein said background scene sensing means for converting background light images to electrical digital image signals for processing by a digital signal processor comprises;

means for responding to digital commands from a digital signal processor;

means for storing digital image signals in a reference background image matrix array within the sensor upon command by the digital signal processor; and

means for providing digital information from said array upon said command from said digital signal processor.

4. A concealing system according to claim **1** wherein said sensing means for sensing a background scene produces fiberoptic image signals for said processed image.

5. A concealing system according to claim **1** wherein a processing subsystem, including said background sensing means, has the means for processing digital image signals for the purpose of presentation to a shield subsystem comprising:

means for formatting digital image signals from said sensing means;

means for receiving and commanding said sensing means to store digital image signals in a reference background image matrix therein;

means to receive and transmit digital image signals from the reference background image matrix to a synthetic image matrix within a shield subsystem;

means to compare the digital information stored in the reference background image matrix in the sensor with that stored in the synthetic image matrix;

means coupled to said means to compare digital information for correcting or altering the image displayed on the shield subsystem by providing commands to the shield subsystem;

means for converting analog video signals from said correcting means to digital image signals;

means for converting created optical image signals to digital image signals;

means for formatting digital image signals from said optical image creating means into a form suitable for presentation to the shield subsystem; and means for receiving and commanding the display at information stored in the synthetic image matrix.

6. A shield for concealing an object from view of an observer comprising:

means for providing a light absorbing opaque and nonspecular display surface which, in its nonactivated state, has minimum light reflectivity;

means for receiving and commanding the display of information from an analog signal processor;

means for providing in said display's activated state the display of a near-identical full color image of the background or other scene for view on the shield surface facing an observer as directed by said analog signal processor, said shield including means to obstruct visual sight of an adjacent concealed object;

means associated with said object for support thereof; and

means to support said analog processor, at least one sensor, and circuitry therefor.

7. A shield for concealing an object from view comprising:

means for providing a light absorbing opaque and nonspecular display surface that, in its nonactivated state, has minimum light reflectivity;

means for obstructing visual sight of an object in said shield, concealed adjacent thereto from view;

means for receiving and commanding the display of information from a digital signal processor;

means for providing in its activated state the display of a near-identical full-color image of a screen for view on the shield surface being viewed as directed by said digital signal processor;

means for storing digital image signals in an array within a synthetic image matrix;

means for display of said digital image signals in the form a full-color image of the background as stored in said synthetic image matrix, said full-color image to be presented on said surface being viewed;

means associated with said object for support thereof; and

means to support said digital signal processor and at least one sensor in association with said shield including backplanes for mounting embedded components, embedded sensors, and embedded data busses.

8. A shield according to claim **7** wherein said shield is constructed from fiberoptic materials and has means for responding to commands from said digital signal processor.

9. A shield according to claim **7** including an integral support structure for said shield, said signal processor, and an object to be concealed, said shield including backplanes for mounting embedded components, embedded sensors, and embedded data buses.

10. A shield according to claim **8** including an integral support structure for said sensor, said signal processor, and an object to be concealed, said shield including backplanes for mounting embedded components, embedded sensors, and embedded data buses.

* * * * *

US006859114B2

(12) United States Patent
Eleftheriades et al.

(10) **Patent No.: US 6,859,114 B2**
(45) **Date of Patent: Feb. 22, 2005**

(54) METAMATERIALS FOR CONTROLLING AND GUIDING ELECTROMAGNETIC RADIATION AND APPLICATIONS THEREFOR

(76) Inventors: **George V. Eleftheriades**, 59 Dorset Road, Scarborough, Ontario (CA), M1M 2S8; **Keith G. Balmain**, 44 Jackes Avenue, Apartment #1904, Toronto, Ontario (CA), M4T 1E5

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 4 days.

(21) Appl. No.: **10/453,907**

(22) Filed: **Jun. 2, 2003**

(65) **Prior Publication Data**

US 2004/0066251 A1 Apr. 8, 2004

Related U.S. Application Data

(60) Provisional application No. 60/384,080, filed on May 31, 2002.

(51) **Int. Cl.**^{7} **H01P 1/18**

(52) **U.S. Cl.** **333/156**; 333/116

(58) **Field of Search** 333/156, 116, 333/204; 343/700

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,731,244 B2 * 5/2004 Killen et al. .... 343/700 MS

* cited by examiner

Primary Examiner—Brian Young
Assistant Examiner—Joseph Lauture
(74) *Attorney, Agent, or Firm*—Beyer Weaver & Thomas, LLP

(57) ABSTRACT

A linear metamaterial comprises a transmission line, having a linear dimension, and is loaded with capacitors, and shunted with an inductor such that for an electromagnetic wave, having a wavelength greater than the linear dimension and traveling along the axis of the transmission line, the effective permeability and permittivity of the metamaterial are simultaneously negative. Applications for the metamaterial are also disclosed.

18 Claims, 28 Drawing Sheets

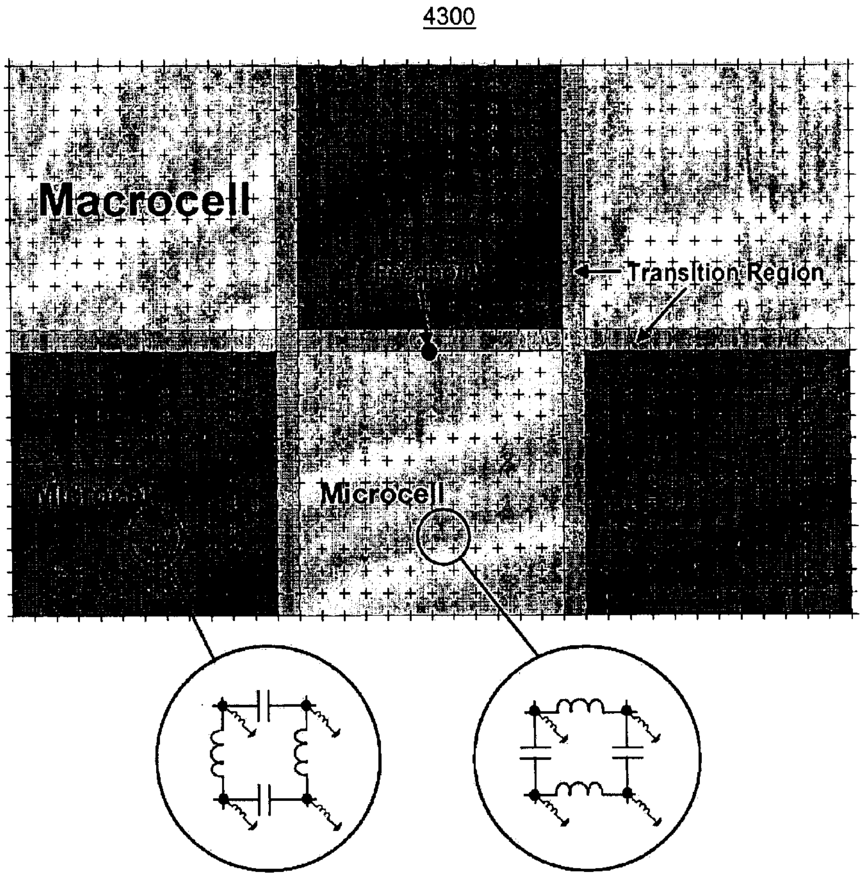

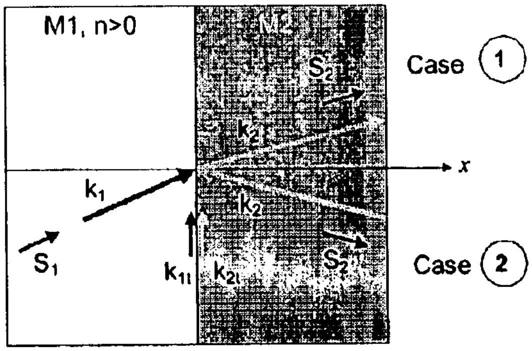
M1, n>0
M2
S_2
Case ①
k_2
x
k_1
k_2
S_1
k_{1t}
k_{2t}
S_2
Case ②

Figure 1

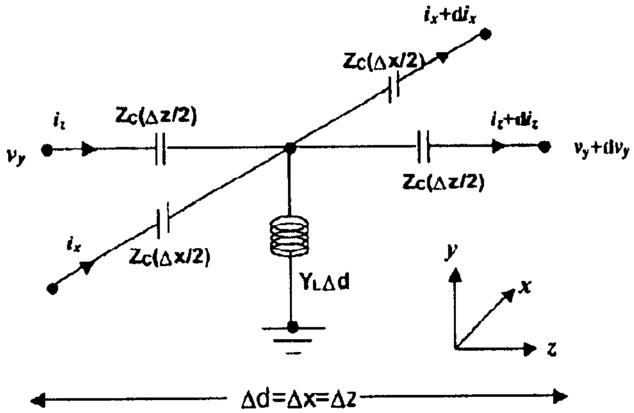
i_x+di_x
$Zc(\Delta x/2)$
i_z
$Zc(\Delta z/2)$
i_z+di_z
v_y
v_y+dv_y
$Zc(\Delta z/2)$
i_x
$Zc(\Delta x/2)$
$Y_L\Delta d$
y
x
z
$\Delta d=\Delta x=\Delta z$

Figure 2

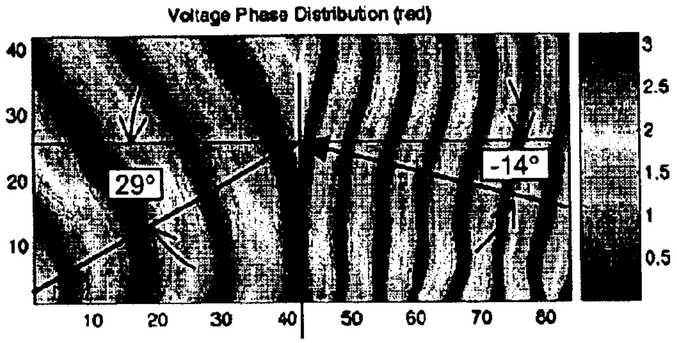

Figure 3

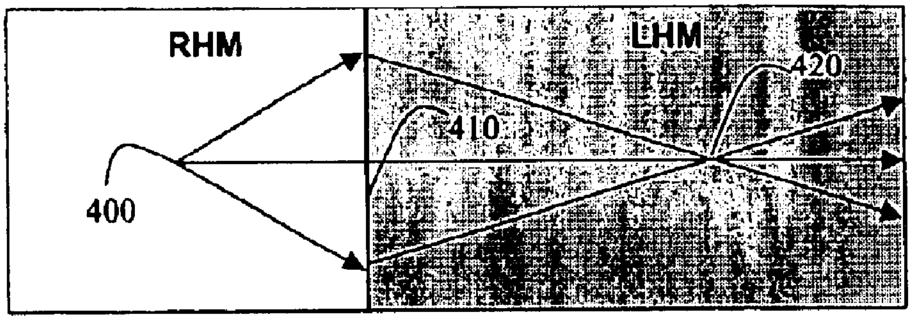
RHM
LHM
400
410
420

Figure 4

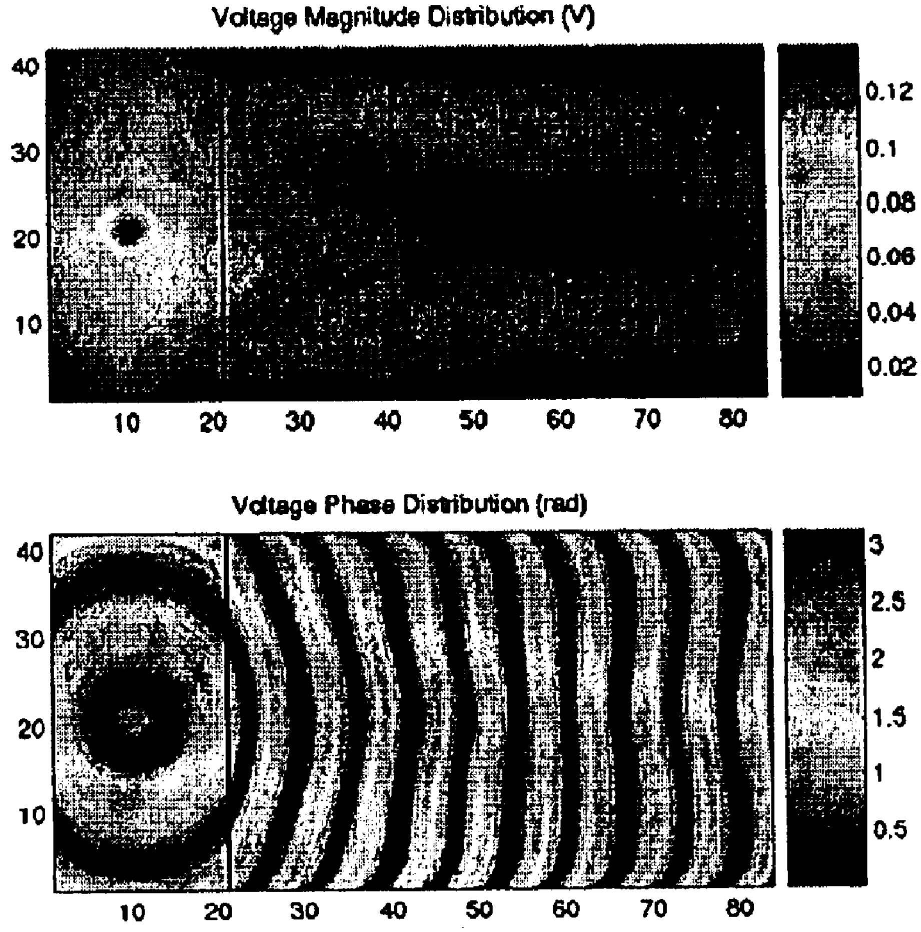

Figure 5

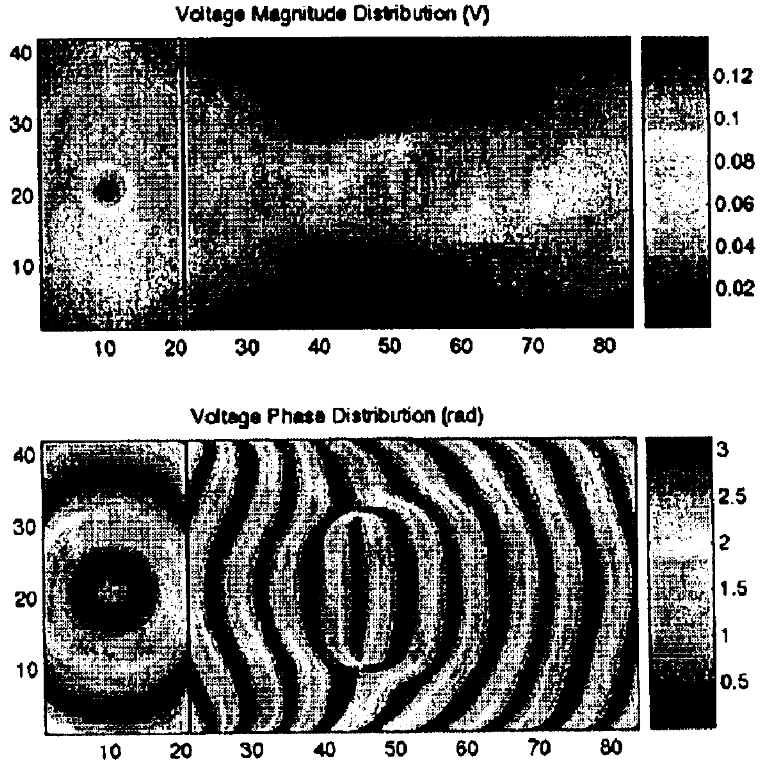

Figure 6

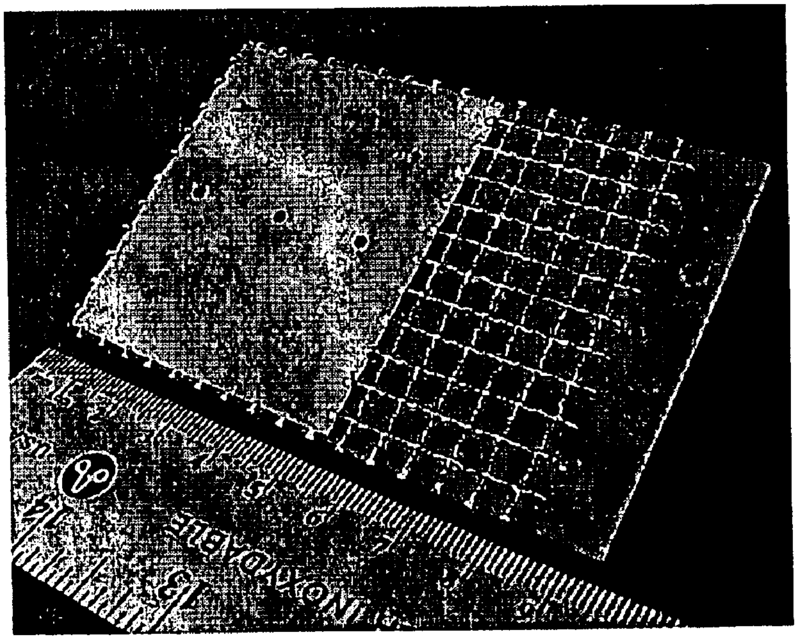

Figure 7

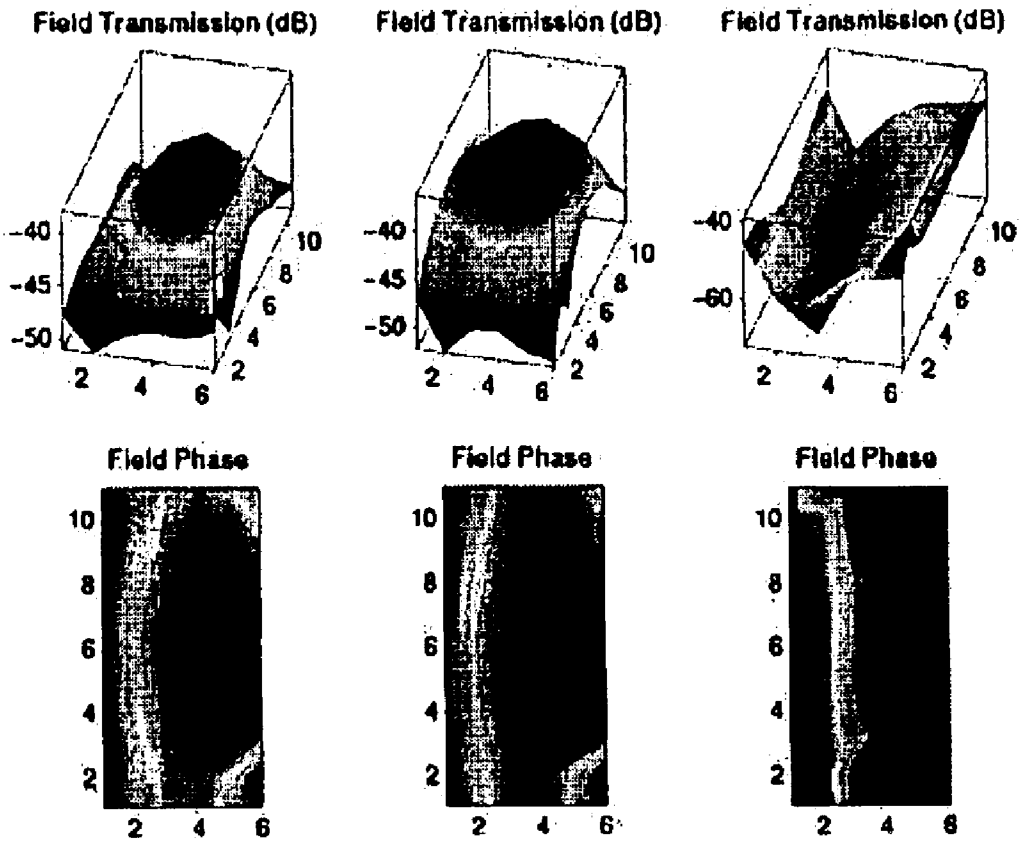
Field Transmission (dB)
Field Transmission (dB)
Field Transmission (dB)
-40
-45
-50
2
4
6
2
4
6
8
10
-40
-45
-50
2
4
6
2
4
6
8
10
-40
-60
2
4
6
2
4
6
8
10
Field Phase
Field Phase
Field Phase
10
8
6
4
2
2
4
6
10
8
6
4
2
2
4
6
10
8
6
4
2
2
4
6

Figure 8

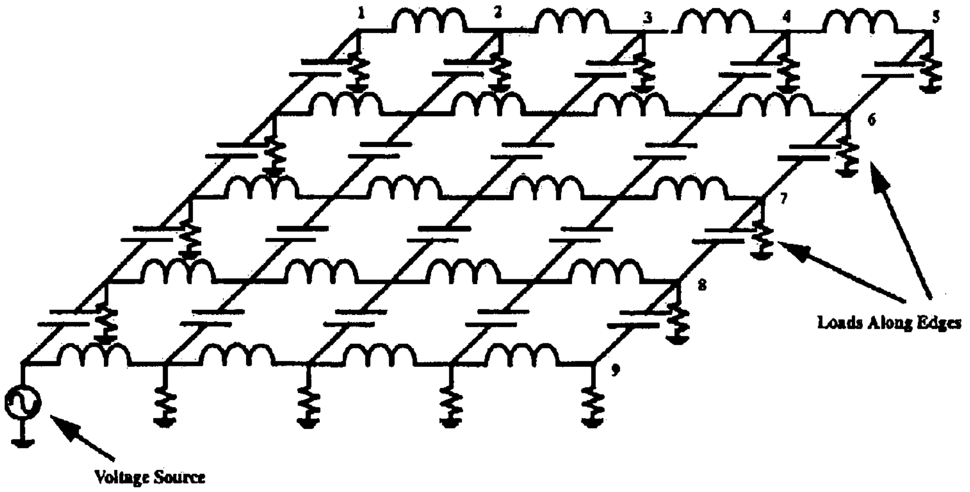
1
2
3
4
5
6
7
8
9
Loads Along Edges
Voltage Source

Capacitance: 1 pF (at each end of each segment)
Inductance: 25.33 nH (at each end of each segment)
Load Resistances: 300 Ohms
Source Voltage: 1 V
Grid Spacing: 1 mm
Height above ground: 1 mm

Figure 9

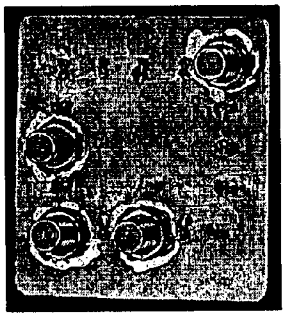

Figure 10

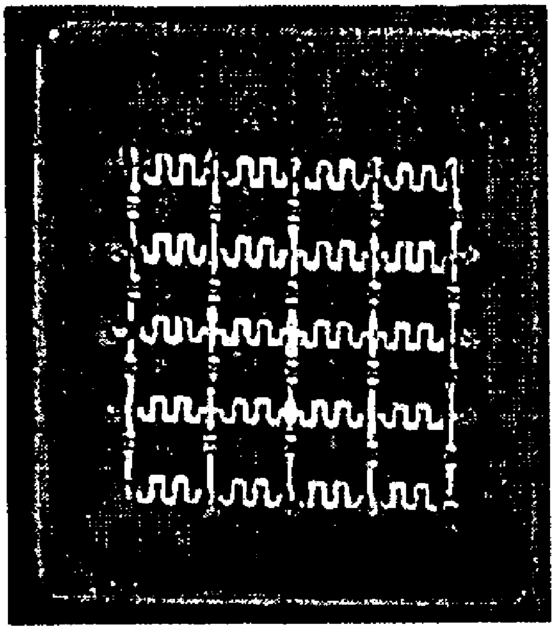

Figure 11

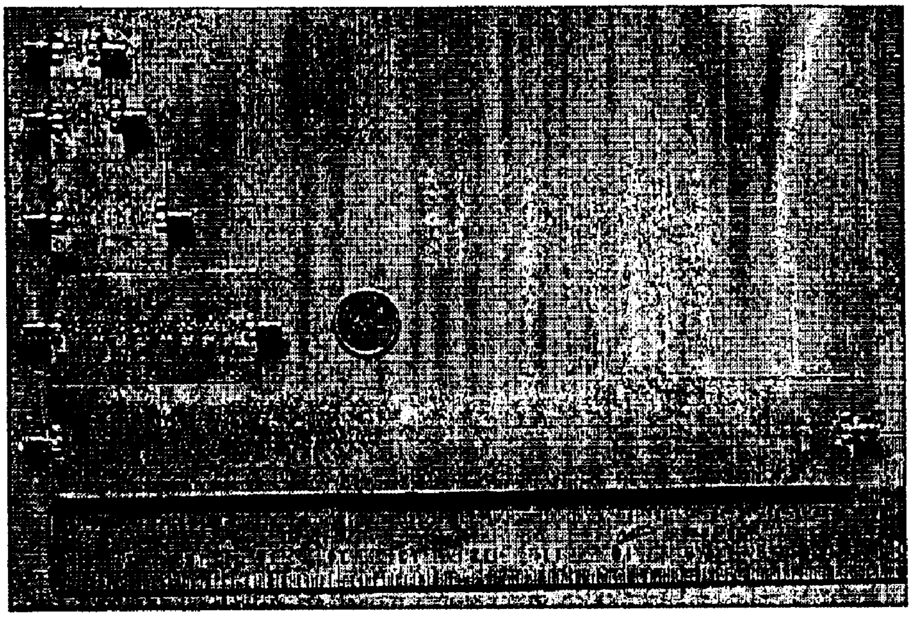

Figure 12

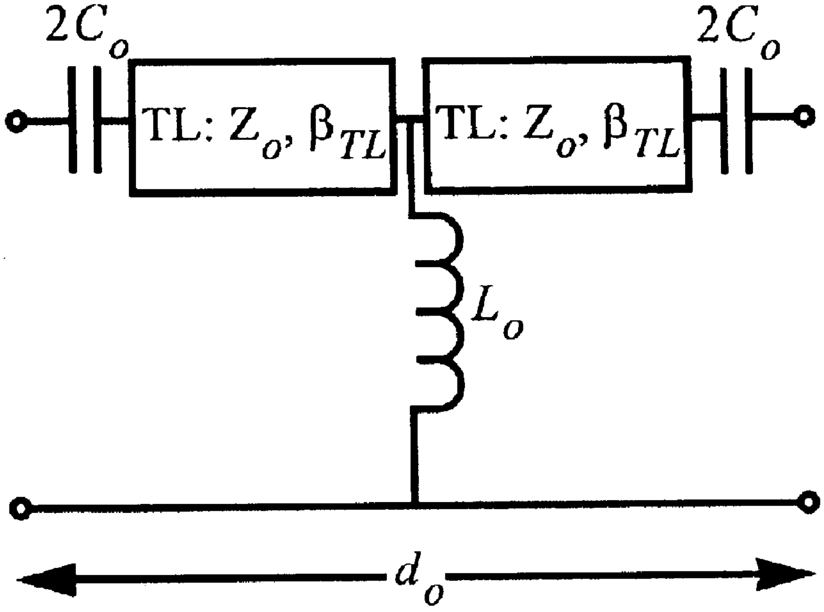
$2C_o$
$2C_o$
TL: Z_o, β_{TL}
TL: Z_o, β_{TL}
L_o
d_o

Figure 13

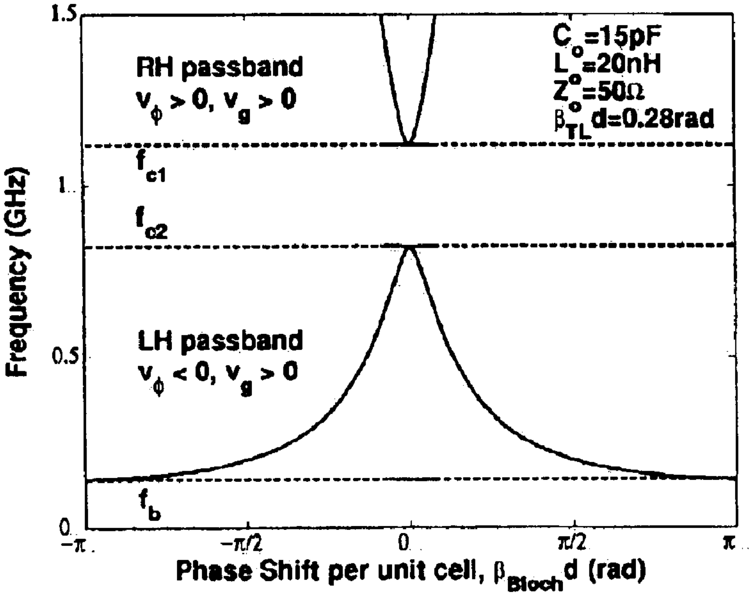

Figure 14

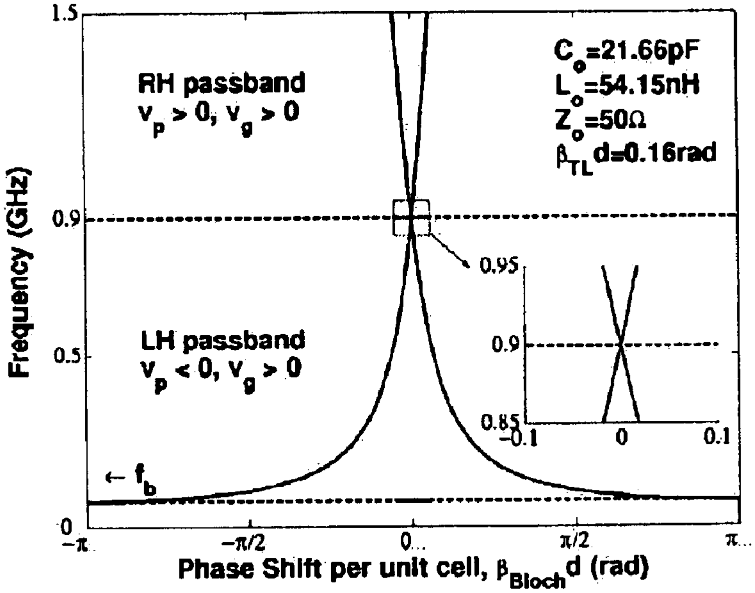

Figure 15

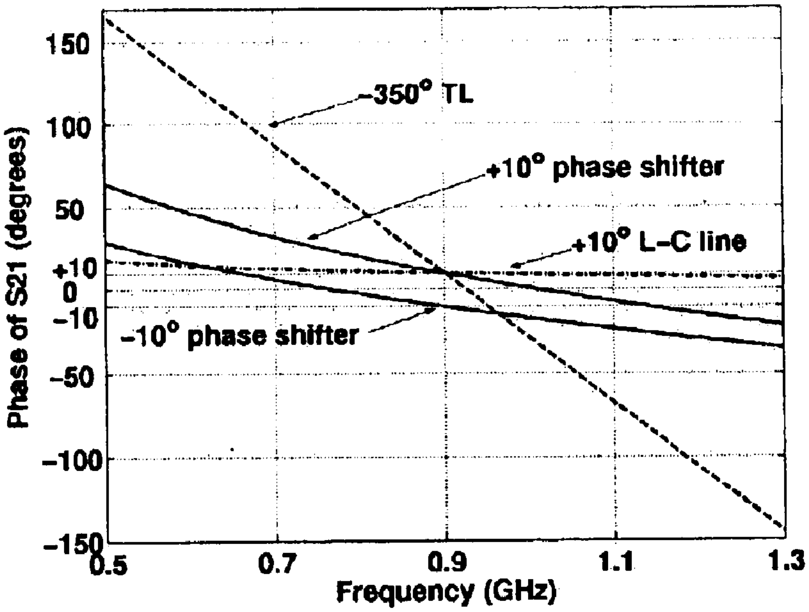

Figure 16

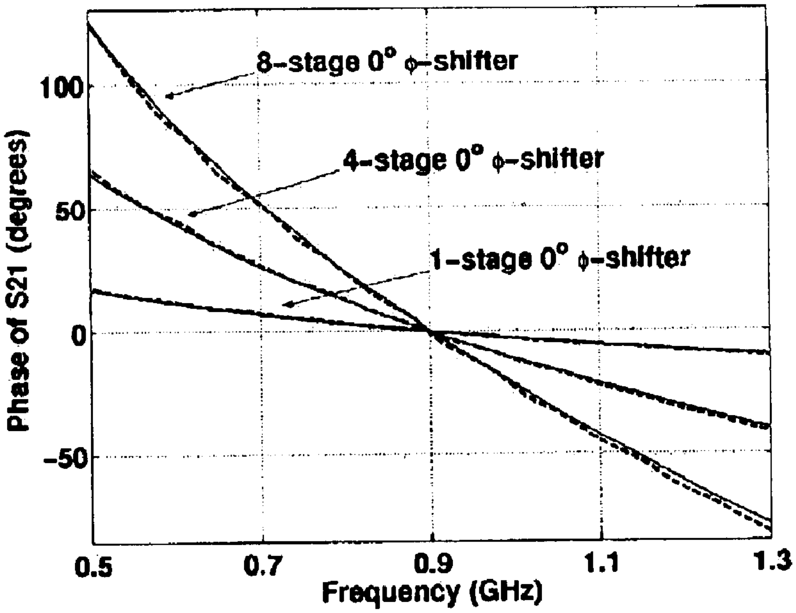

Figure 17

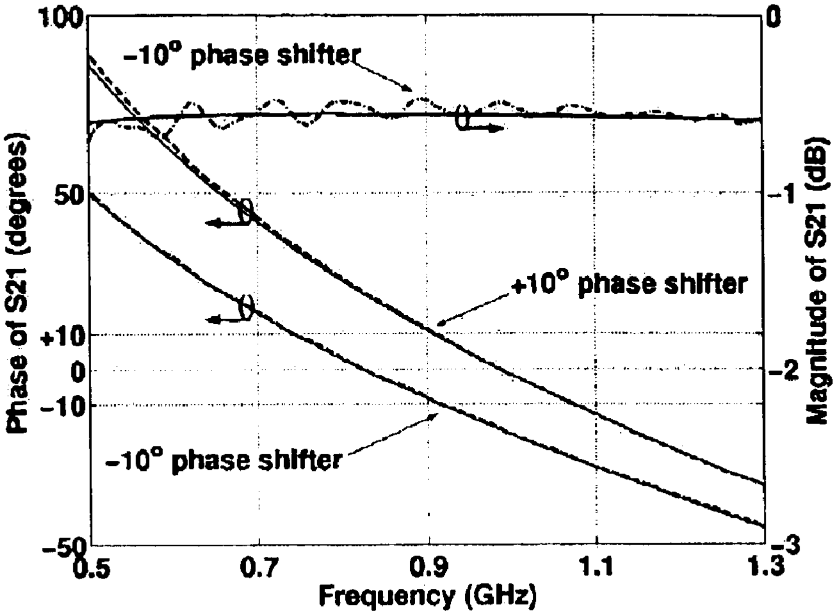

Figure 18

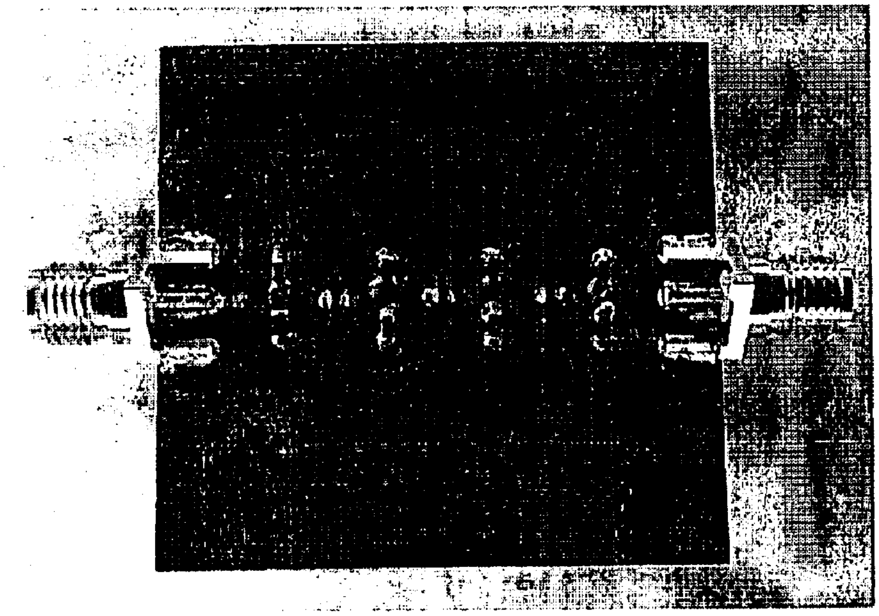

Figure 19

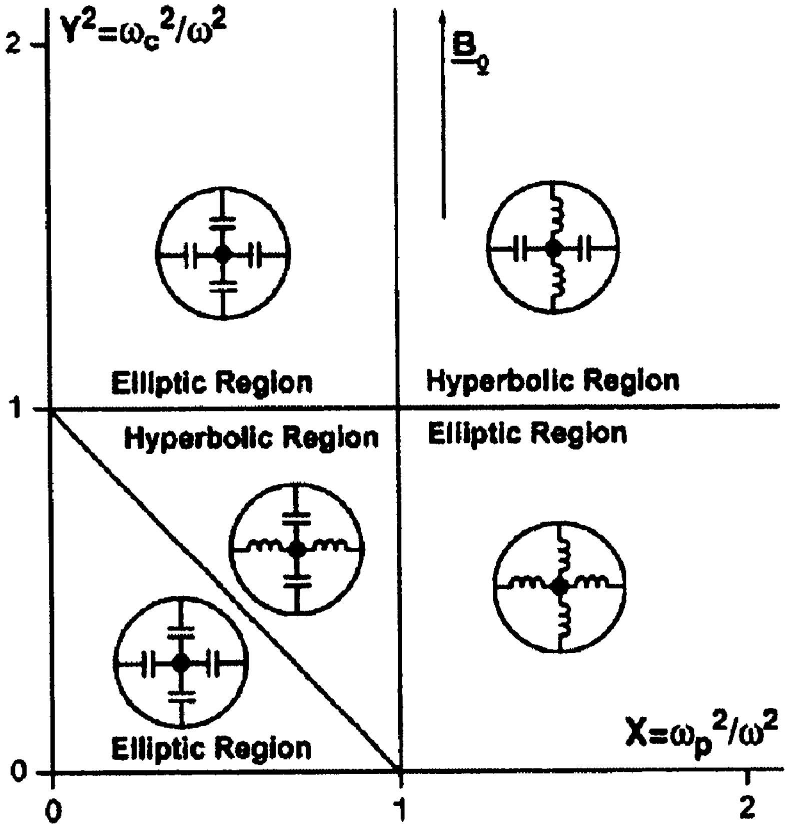
$Y^2=\omega_c^2/\omega^2$
$\underline{B}_0$
Elliptic Region
Hyperbolic Region
Hyperbolic Region
Elliptic Region
Elliptic Region
$X=\omega_p^2/\omega^2$
0
1
2
0
1
2

Figure 20

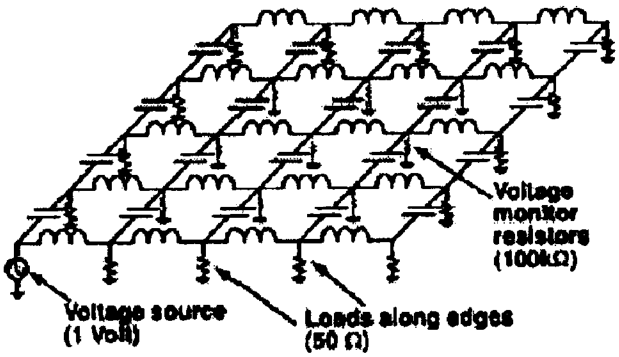
Voltage monitor resistors (100kΩ)
Voltage source (1 Volt)
Loads along edges (50 Ω)

Figure 21

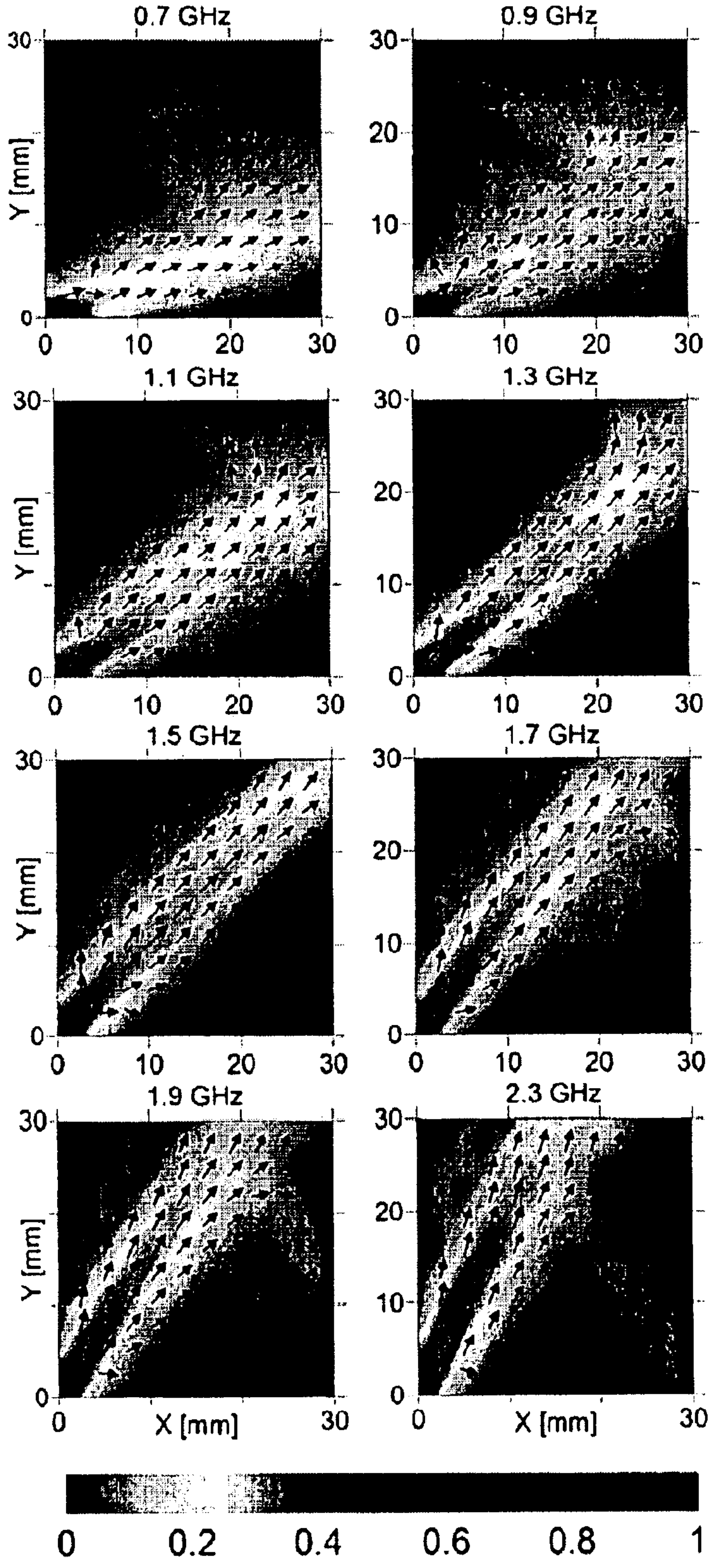
0.7 GHz
0.9 GHz
1.1 GHz
1.3 GHz
1.5 GHz
1.7 GHz
1.9 GHz
2.3 GHz
Y [mm]
X [mm]
0
10
20
30
0
0.2
0.4
0.6
0.8
1

Figure 22

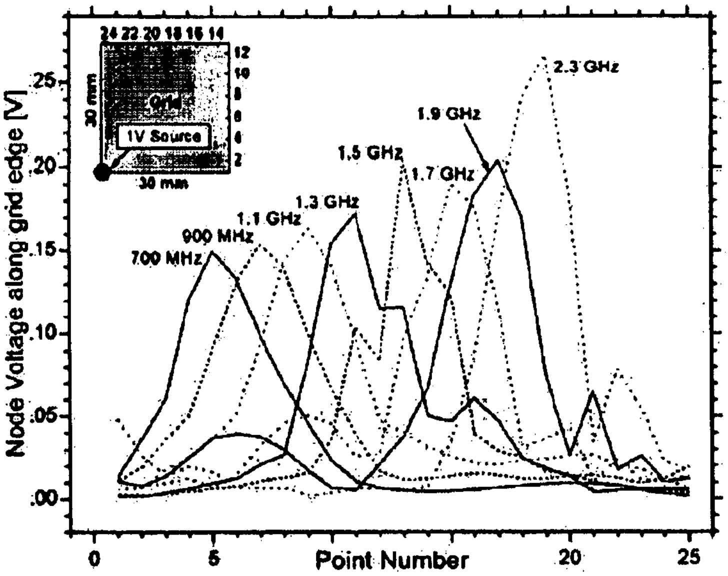

Figure 23

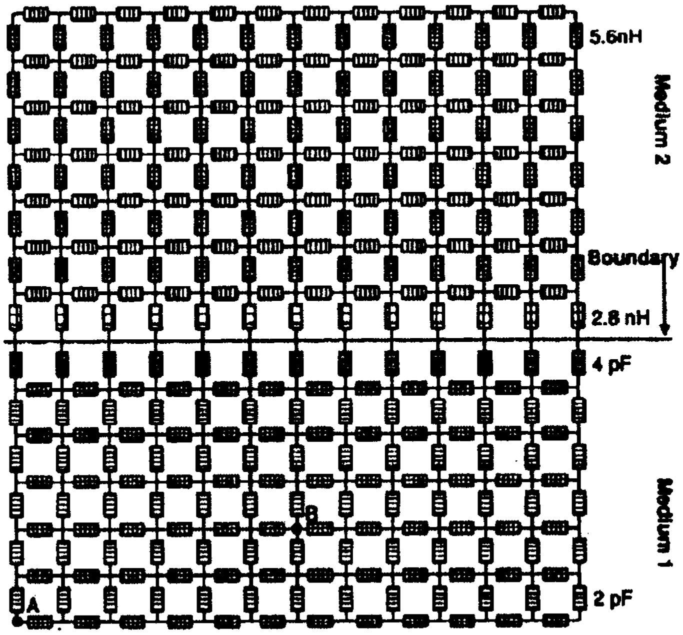
5.6nH
Medium 2
Boundary
2.8 nH
4 pF
Medium 1
B
A
2 pF

Figure 24

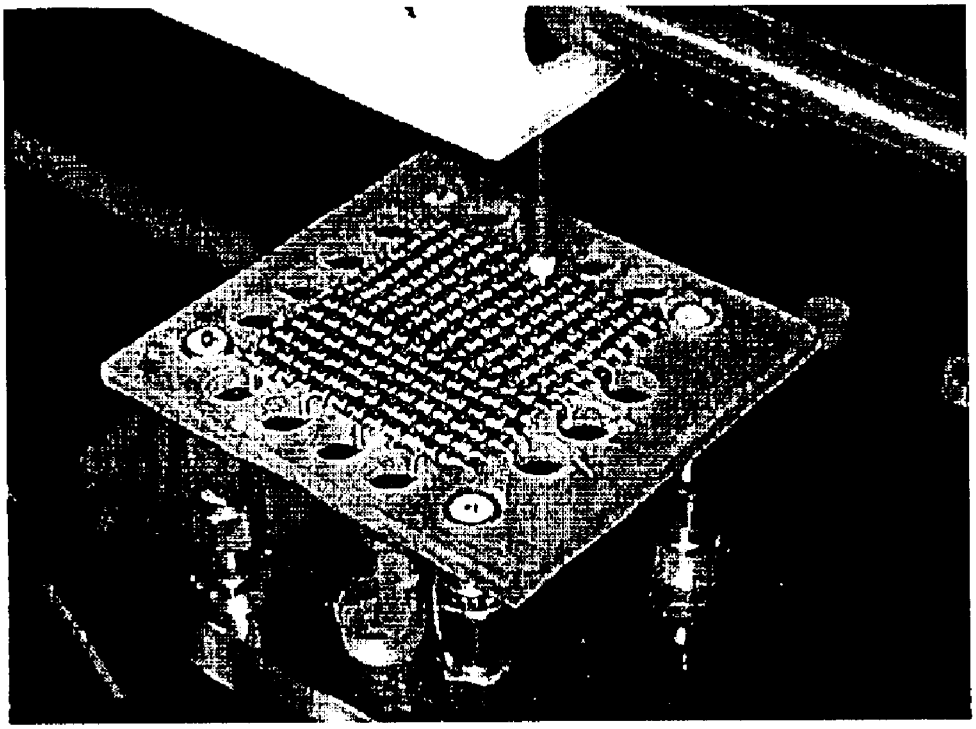

Figure 25

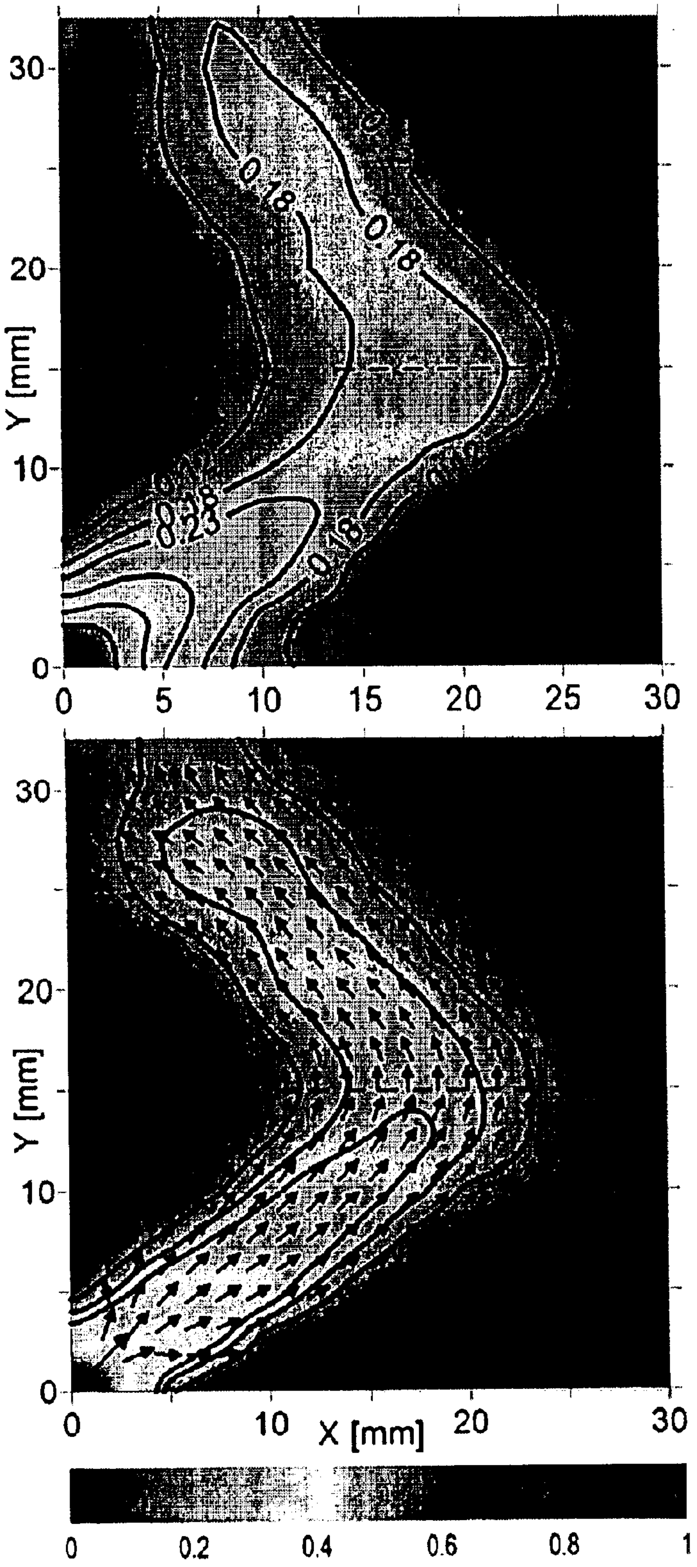
30
20
10
0
Y [mm]
0.18
0.18
0.18
0.23
0.18
0
5
10
15
20
25
30
30
20
10
0
Y [mm]
0
10
X [mm]
20
30
0
0.2
0.4
0.6
0.8
1

Figure 26

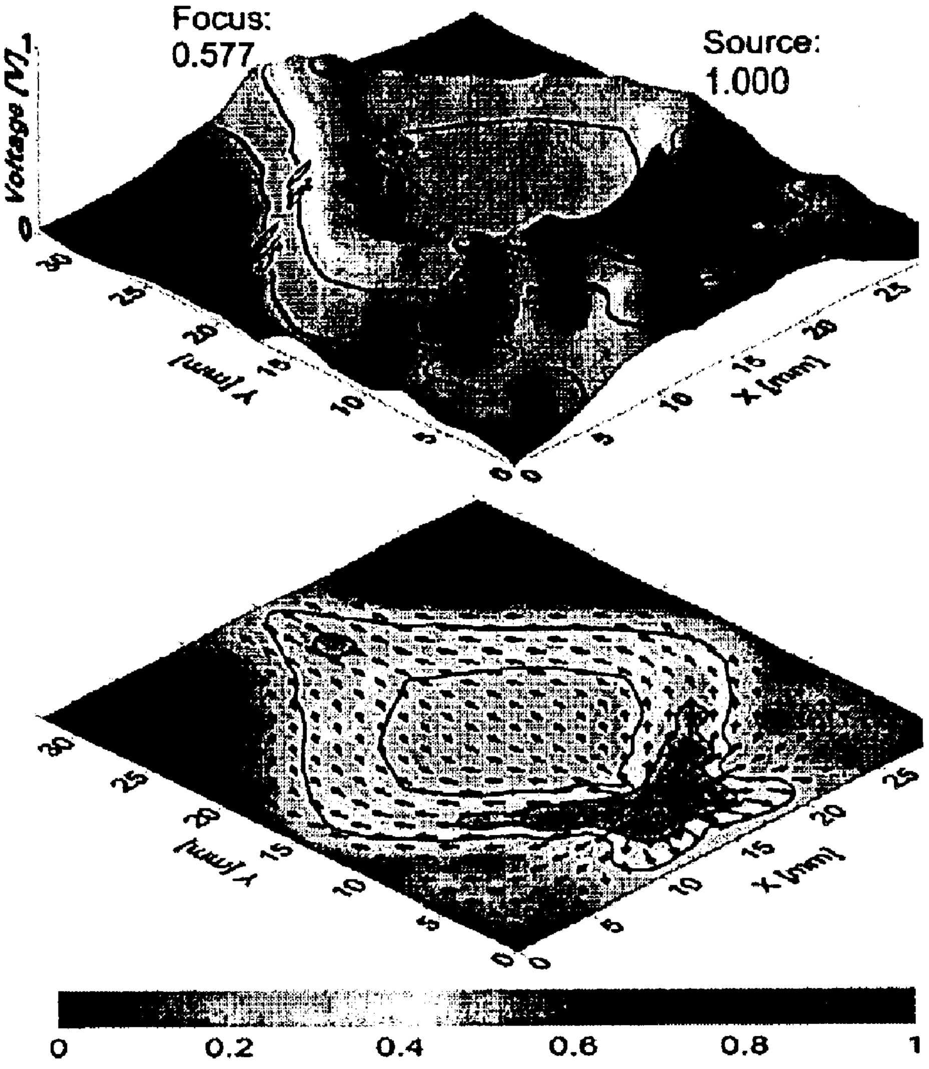
Focus:
0.577
Source:
1.000
Voltage [V]
1
0
30
25
20
15
10
5
0
Y [mm]
X [mm]
0.2
0.4
0.6
0.8
1

Figure 27

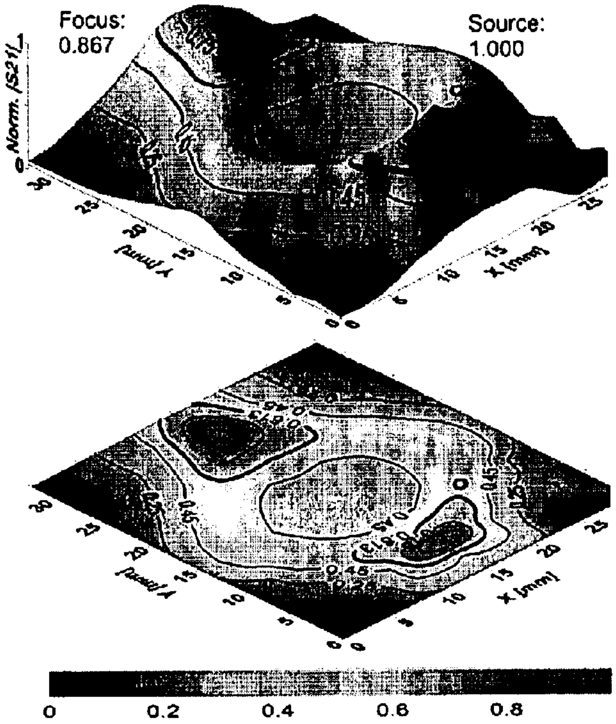

Figure 28

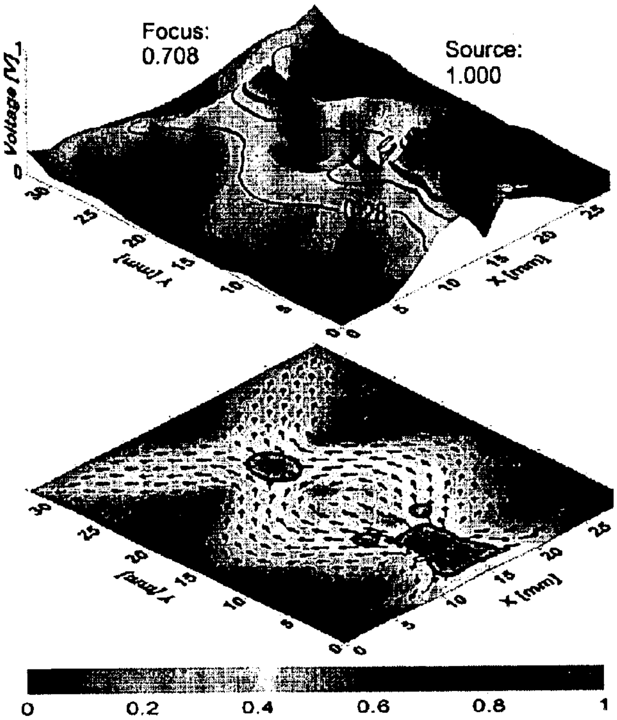

Figure 29

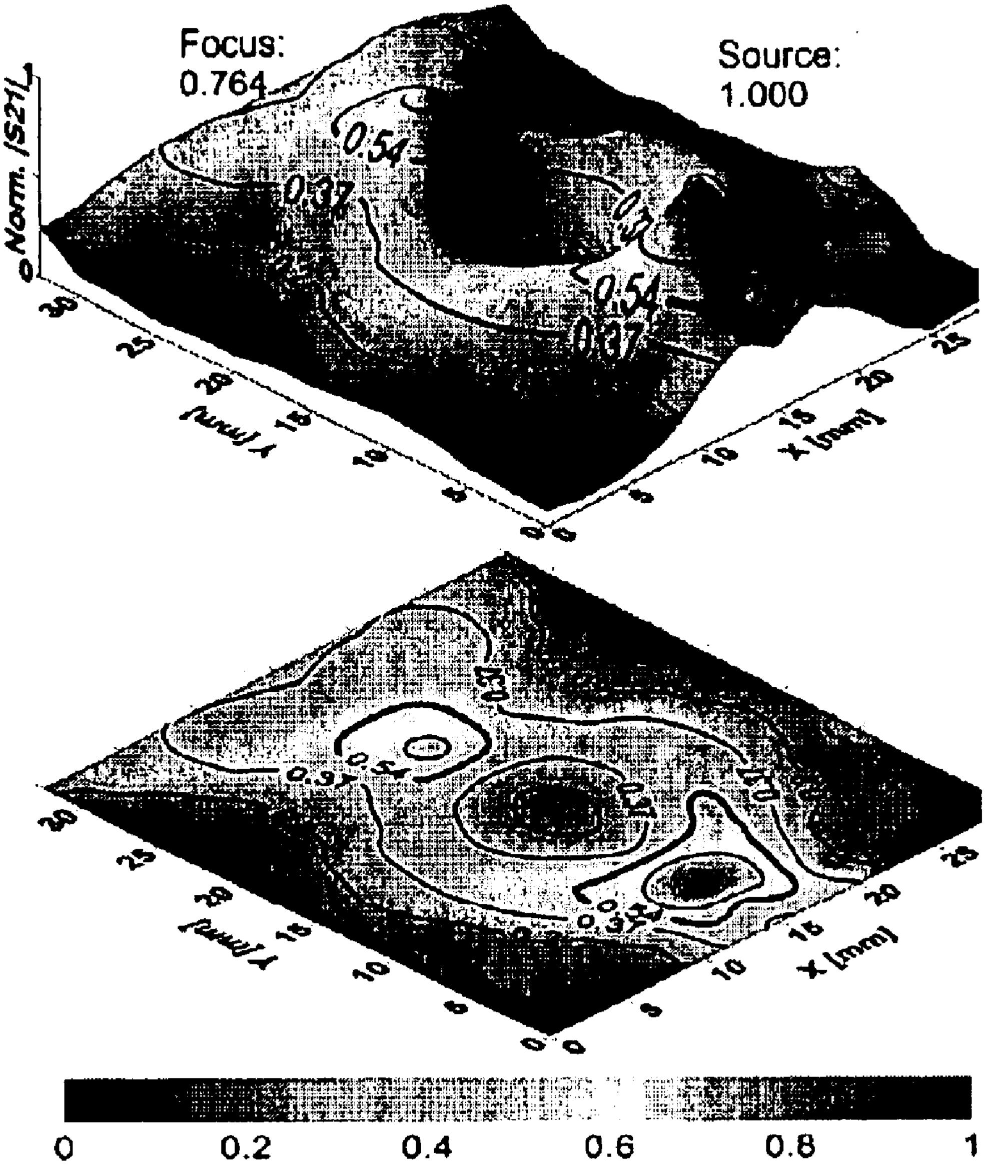
Focus:
0.764
Source:
1.000
Norm. |S21|
0.54
0.37
0.54
0.37
30
25
20
15
10
5
0
Y [mm]
X [mm]
0.54
0.37
0.37
0
0.2
0.4
0.6
0.8
1

Figure 30

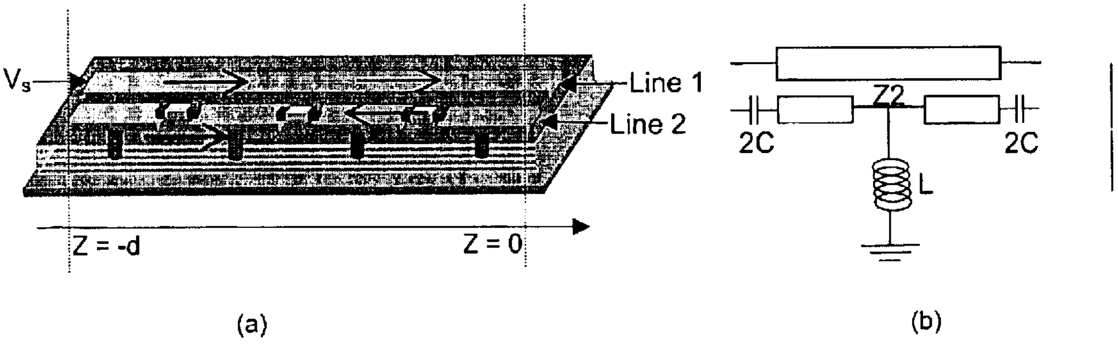
Vs
Line 1
Line 2
Z = -d
Z = 0
Z2
2C
2C
L
(a)
(b)

Figure 31

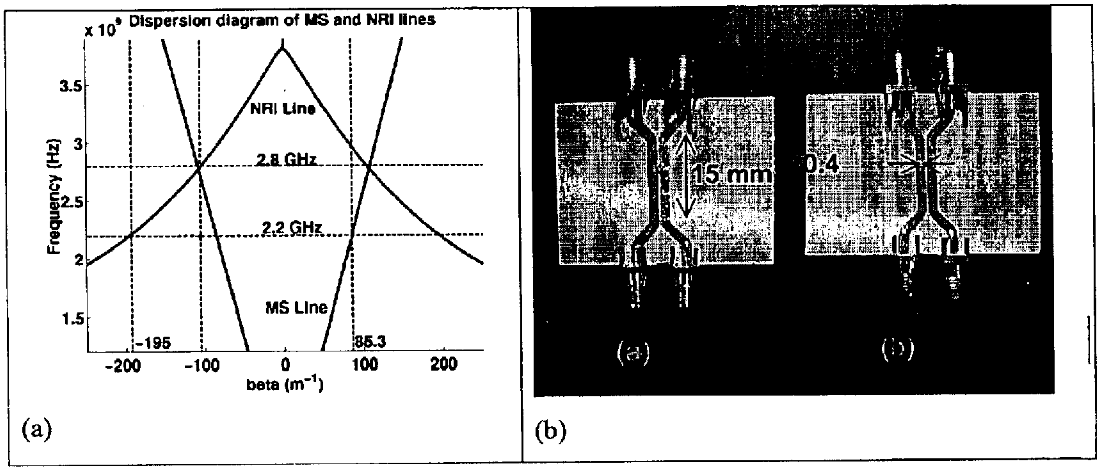

Figure 32

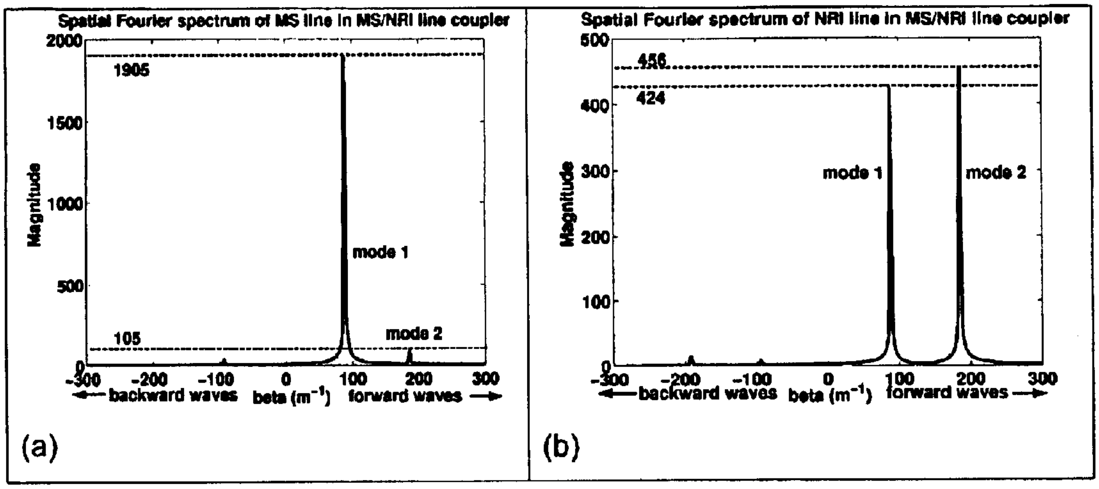

Figure 33

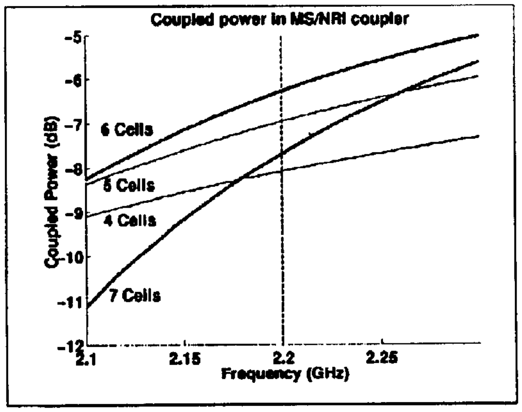

Figure 34

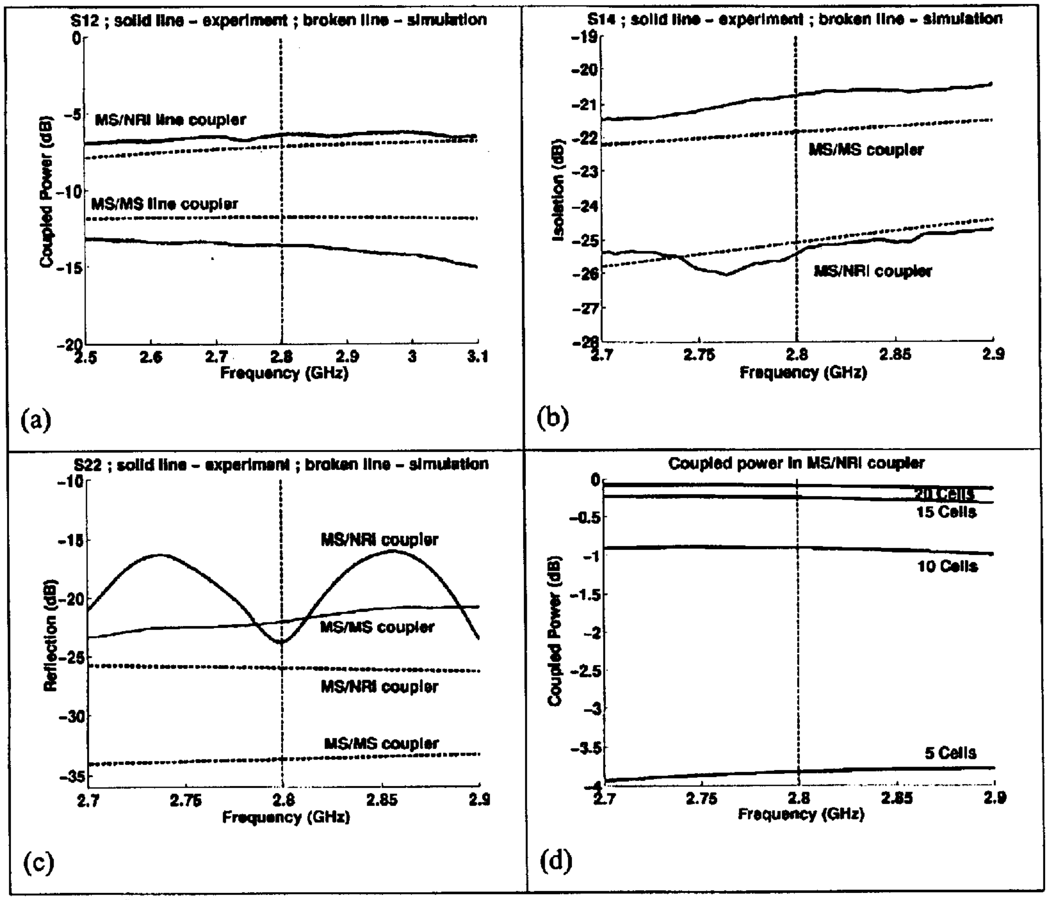

Figure 35

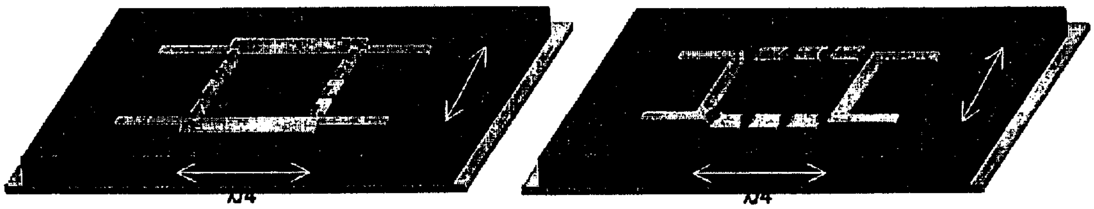

(a) (b)

Figure 36

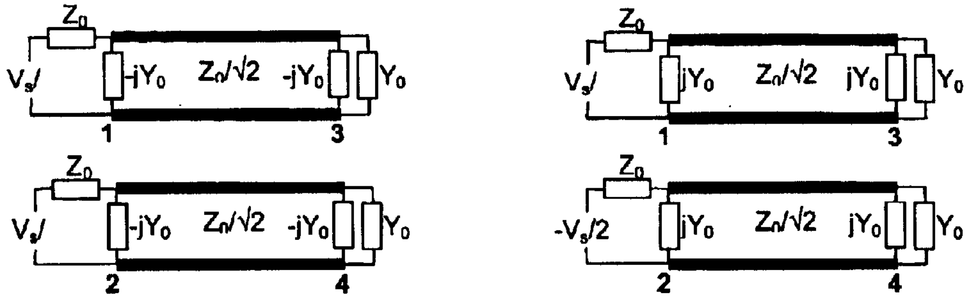
Z_0
$V_s/$
$-jY_0$
$Z_0/\sqrt{2}$
$-jY_0$
Y_0
1
3
Z_0
$V_s/$
$-jY_0$
$Z_0/\sqrt{2}$
$-jY_0$
Y_0
2
4
Z_0
$V_s/$
jY_0
$Z_0/\sqrt{2}$
jY_0
Y_0
1
3
Z_0
$-V_s/2$
jY_0
$Z_0/\sqrt{2}$
jY_0
Y_0
2
4

Even

Odd Mode

Figure 37

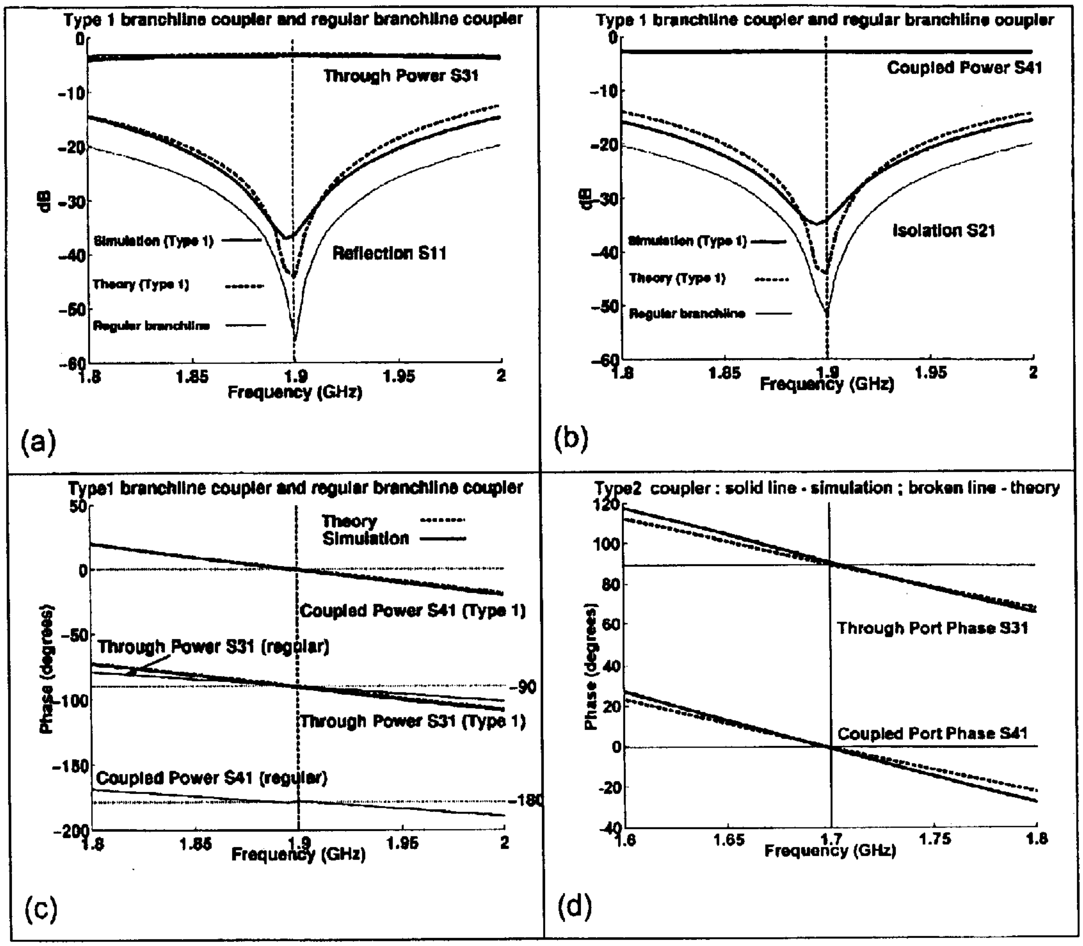
Type 1 branchline coupler and regular branchline coupler
Through Power S31
Simulation (Type 1)
Theory (Type 1)
Regular branchline
Reflection S11
dB
0
-10
-20
-30
-40
-50
-60
1.8
1.85
1.9
1.95
2
Frequency (GHz)
(a)
Type 1 branchline coupler and regular branchline coupler
Coupled Power S41
Isolation S21
Simulation (Type 1)
Theory (Type 1)
Regular branchline
dB
Frequency (GHz)
(b)
Type1 branchline coupler and regular branchline coupler
Theory
Simulation
Coupled Power S41 (Type 1)
Through Power S31 (regular)
Through Power S31 (Type 1)
Coupled Power S41 (regular)
-90
-180
50
0
-50
-100
-150
-200
Phase (degrees)
Frequency (GHz)
(c)
Type2 coupler : solid line - simulation ; broken line - theory
Through Port Phase S31
Coupled Port Phase S41
120
100
80
60
40
20
0
-20
-40
1.6
1.65
1.7
1.75
1.8
Phase (degrees)
Frequency (GHz)
(d)

Figure 38

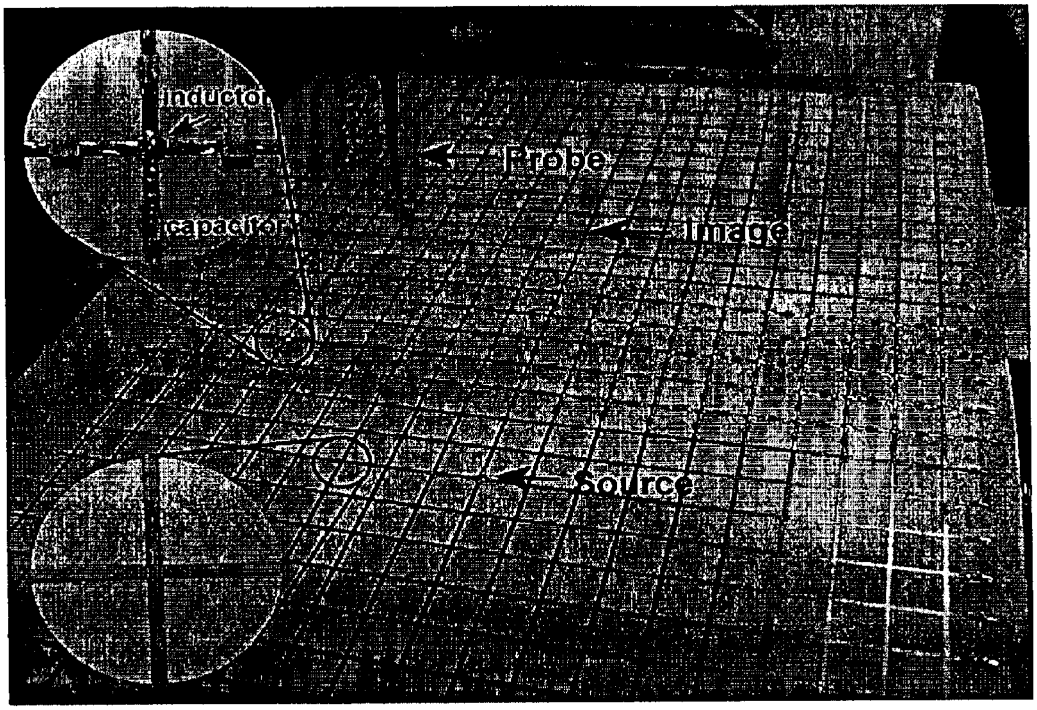
inductor
Probe
capacitor
Image
Source

Figure 39

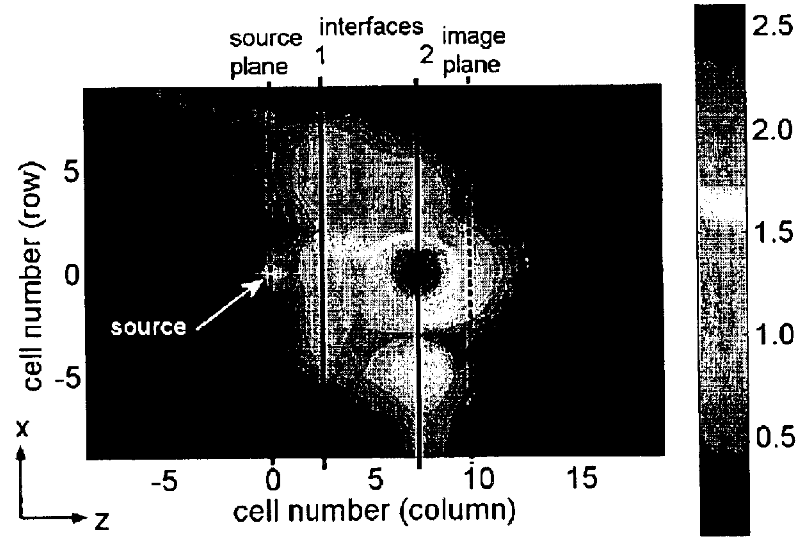

Figure 40

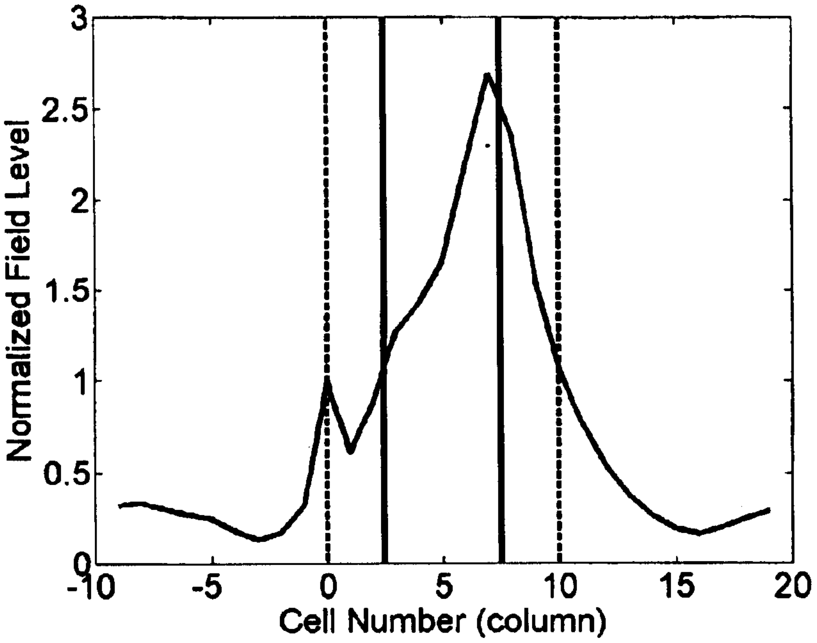

Figure 41

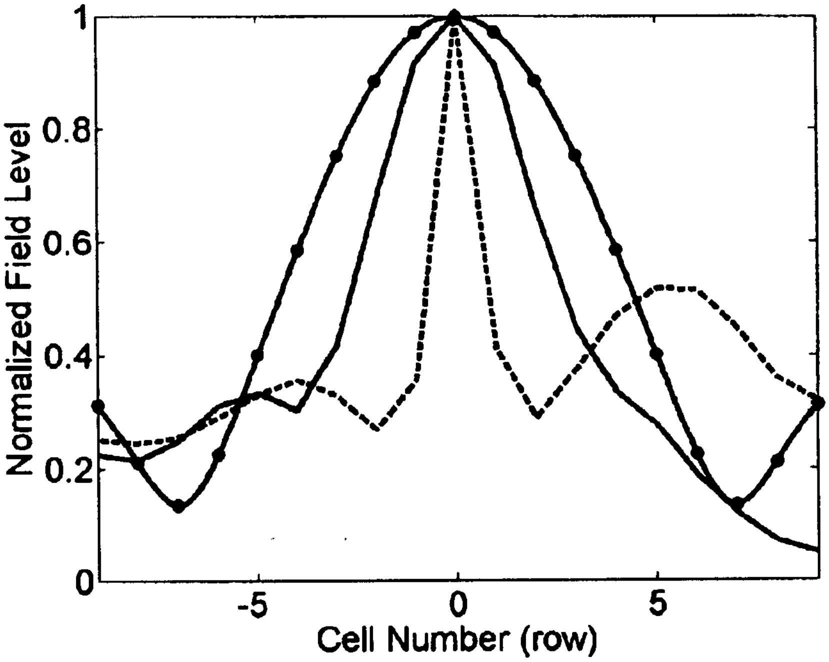

Figure 42

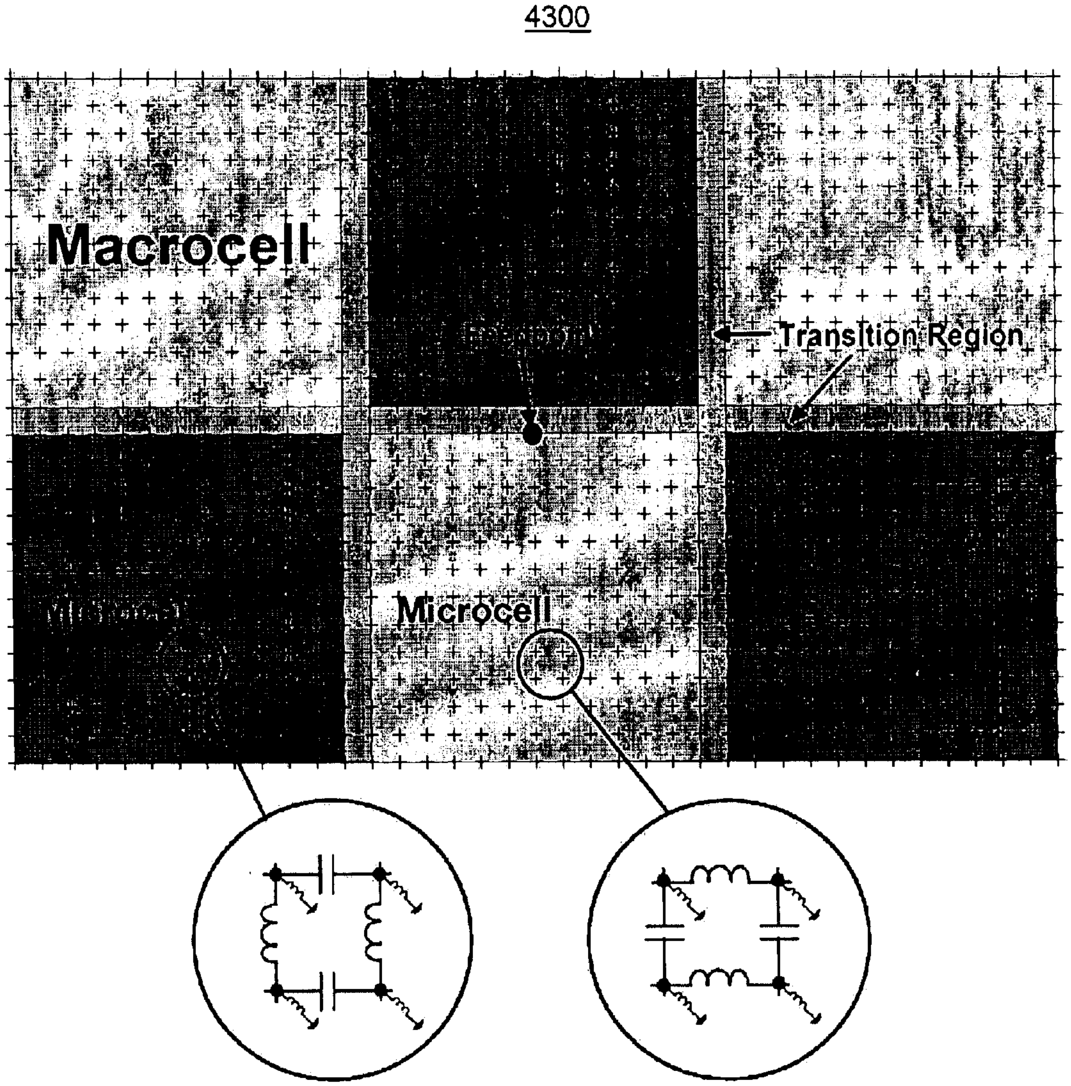
4300
Macrocell
Freepoint
Transition Region
Microcell
Microcell

Figure 43

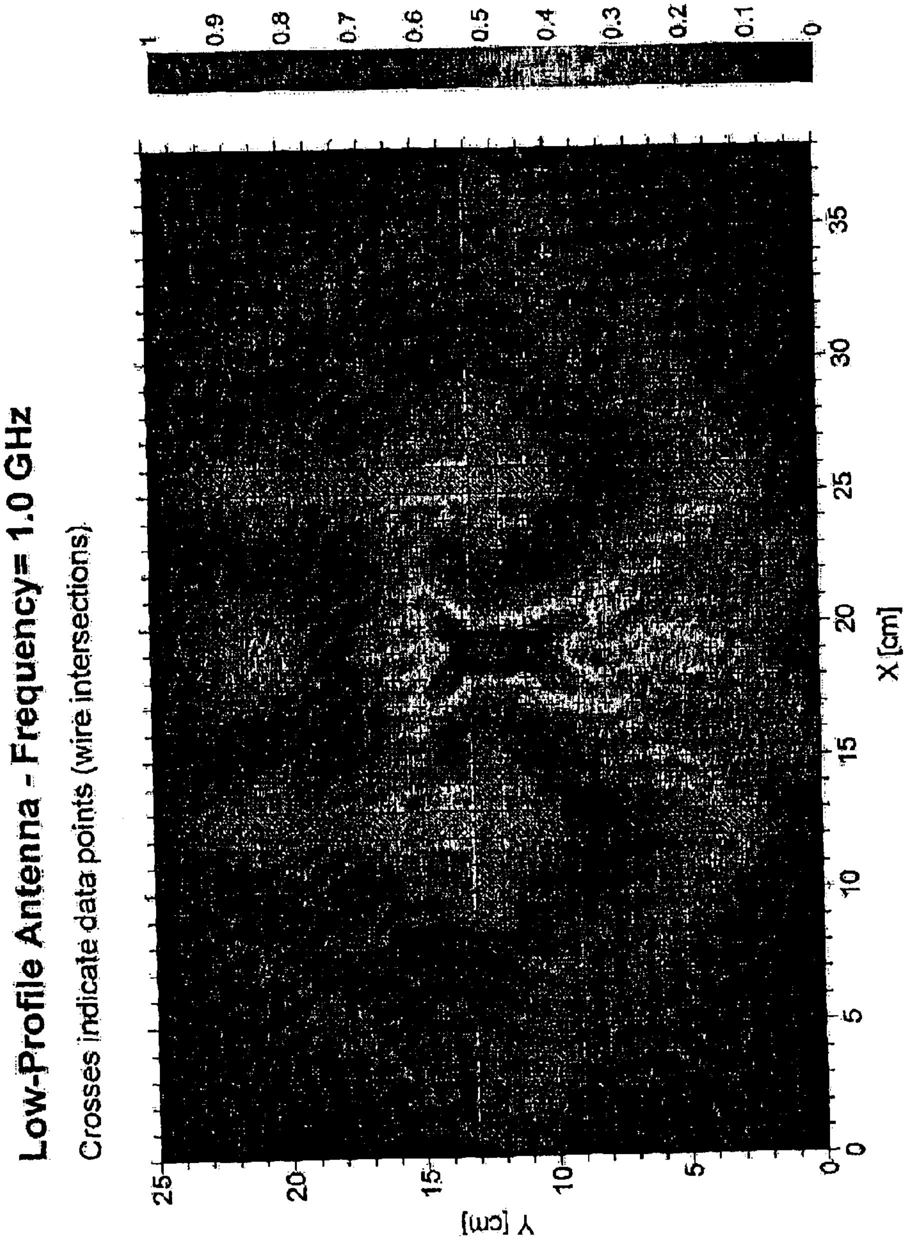
Low-Profile Antenna - Frequency= 1.0 GHz
Crosses indicate data points (wire intersections)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Y [cm]
25
20
15
10
5
0
X [cm]
0
5
10
15
20
25
30
35

Figure 44

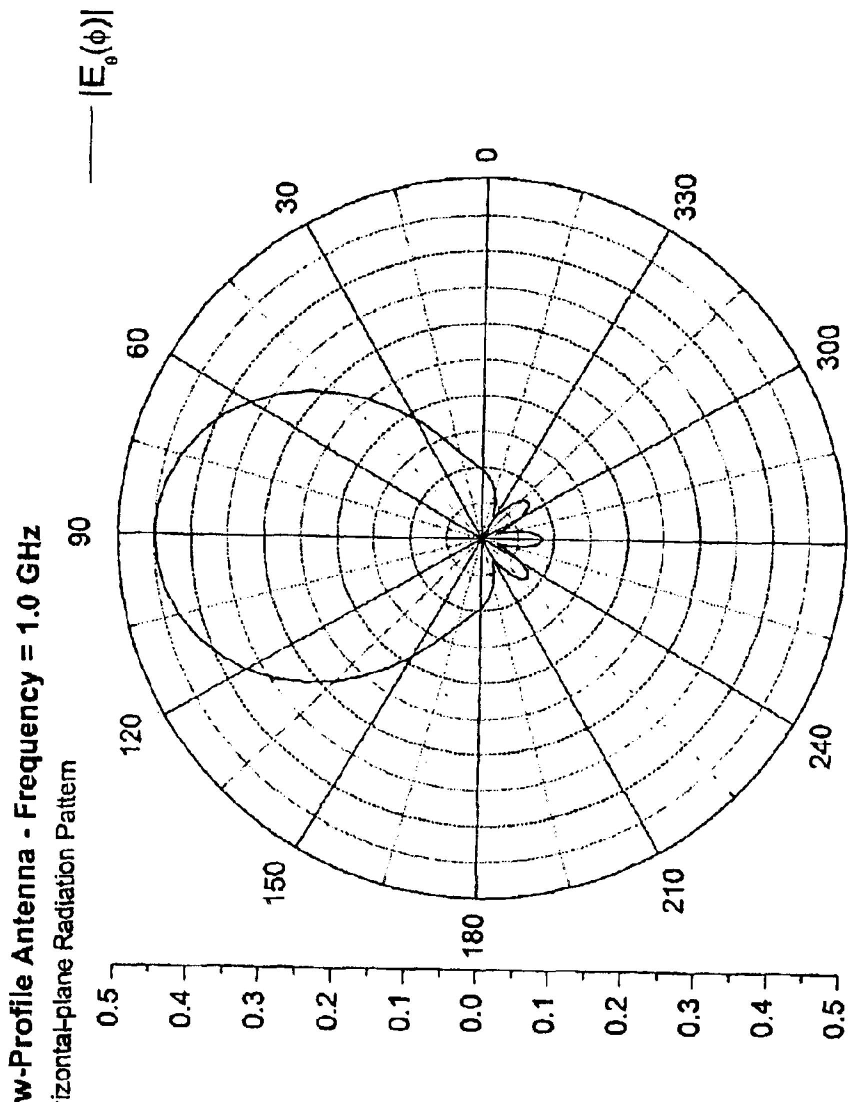
Low-Profile Antenna - Frequency = 1.0 GHz
Horizontal-plane Radiation Pattern
— |E_θ(ϕ)|
0
30
60
90
120
150
180
210
240
270
300
330
0.5
0.4
0.3
0.2
0.1
0.0
0.1
0.2
0.3
0.4
0.5

Figure 45

METAMATERIALS FOR CONTROLLING AND GUIDING ELECTROMAGNETIC RADIATION AND APPLICATIONS THEREFOR

CROSS-REFERENCE TO RELATED APPLICATIONS

The present application claims the benefit of U.S. Provisional Patent Application No. 60/384,080 filed on May 31, 2002, which is hereby incorporated by reference herein.

FIELD OF THE INVENTION

The present invention relates generally to the control and guidance of electromagnetic radiation and in particular to a left-handed metamaterial for controlling and guiding electromagnetic radiation and to applications therefor.

BACKGROUND OF THE INVENTION

The concept of a negative index of refraction, originally proposed by Veselago in the 1960s, suggested the possibility of materials in which the permittivity and permeability could be made simultaneously negative. Veselago termed these left-handed media (LHM), because the vectors E, H, and k would form a left-handed triplet instead of a right-handed triplet, as is the case in conventional, right-handed media (RHM). In such a material the phase velocity and Poynting vector are antiparallel. Recently, novel 3-dimensional (3-D) electromagnetic materials have successfully demonstrated negative refraction by synthesizing a negative refractive index. These artificial dielectrics (metamaterials) consist of loosely coupled unit cells composed of thin wire strips and split-ring resonators to synthesize negative permittivity and permeability, respectively. In these metamaterials, the choice of operating frequency is restricted to the region of the resonance, which results in a highly dispersive, narrow and behaviour with strong associated absorption losses.

Other structures with magnetic properties to refract electromagnetic radiation at negative angles have also been considered. For example, International PCT Application No. WO 00/41270 discloses a structure that exhibits magnetic properties when it receives incident electromagnetic radiation. The structure is formed from an array of capacitive elements, each of which is smaller, and preferably much smaller, than the wavelength of the incident electromagnetic radiation. Each capacitive element has a low-resistance conducting path associated with it and is such that a magnetic component of the incident electromagnetic radiation induces an electrical current to flow around a path and through the associated capacitive element. The creation of the internal magnetic fields generated by the flow of the induced electrical current gives rise to the structure's magnetic properties.

International PCT Application No. WO 02/03500 discloses a microstructured magnetic material having a magnetic permeability of negative value but unity magnitude over a particular radio frequency range.

Although the above references disclose structures with magnetic properties to refract electromagnetic radiation at negative angles, improved materials exhibiting negative refractive indices are desired.

It is therefore an object of the present invention to provide a novel left-handed metamaterial for controlling and guiding electromagnetic radiation and novel applications therefor.

SUMMARY OF THE INVENTION

According to one aspect of the present invention there is provided a planar metamaterial comprising:

two substantially orthogonal, coplanar sets of transmission lines, said transmission lines being spaced with a periodicity, loaded with capacitors with said periodicity, and shunted with inductors with said periodicity such that for an electromagnetic wave, having a wavelength greater than said periodicity and traveling along the plane of said transmission lines, the effective permeability and permittivity of said metamaterial are simultaneously negative.

According to another aspect of the present invention there is provided a linear metamaterial comprising:

a transmission line, having a linear dimension, and being loaded with capacitors, and shunted with an inductor such that for an electromagnetic wave, having a wavelength greater than said linear dimension and traveling along the axis of said transmission line, the effective permeability and permittivity of said metamaterial are simultaneously negative.

According to yet another aspect of the present invention there is provided a planar resonance cone metamaterial comprising:

a first set of transmission lines, spaced with a periodicity, and loaded with capacitors with said periodicity;

a second set of transmission lines, substantially orthogonal and coplanar with said first set of transmission lines, said second set of transmission lines being spaced with said periodicity, and loaded with inductors with said periodicity, said first and second sets of transmission lines exhibiting characteristics such that for an electromagnetic wave, having a wavelength greater than said periodicity by an order of magnitude, and traveling along the linear axis of said first set of transmission lines, the effective permittivity of said metamaterial is positive such that for an electromagnetic wave, having a wavelength greater than said periodicity by an order of magnitude and traveling along the linear axis of said second set of transmission lines, the effective permittivity of said metamaterial is negative.

According to yet another aspect of the present invention there is provided a planar resonance cone metamaterial comprising:

a first set of transmission lines, spaced with a first periodicity, and loaded with capacitors with a second periodicity;

a second set of transmission lines, substantially orthogonal and coplanar with said first set of transmission lines, said second set of transmission lines being spaced with said second periodicity, and being loaded with inductors with said first periodicity whereby for an electromagnetic wave, having a wavelength greater than said first and second periodicities by an order of magnitude and traveling along the longitudinal axis of said first set of transmission lines, the effective permittivity of said metamaterial is positive and for an electromagnetic wave having a wavelength greater than said first and second priodocities and traveling along the longitudinal axis of said second set of transmission lines, the effective permittivity of said metamaterial is negative.

According to yet another aspect of the present invention there is provided a near field focusing device comprising:

a first set of transmission lines, said first set of transmission lines being spaced with a first periodicity and loaded with capacitors with a second periodicity;

a second set of transmission lines, substantially orthogonal to, and coplanar with said first set of transmission lines, said second set of transmission lines being spaced with said second periodicity, loaded with capacitors with said first periodicity, and shunted with inductors, said first set of transmission lines intersecting said second set of transmission lines such that for an electromagnetic wave, having a wavelength greater than said first or second periodicity by an order of magnitude, and traveling along the plane of said transmission lines, the effective permeability and permittivity of said metamaterial are simultaneously negative; and

a planar waveguide, having a flat extent, coupled to said transmission lines, such that said flat extent is parallel to one set of transmission lines.

According to yet another aspect of the present invention there is provided a phase-shifting device comprising:

a transmission line, having a linear dimension and characteristic impedance;

capacitors loaded on said transmission line; and

an inductor shunting said transmission line, said transmission line exhibiting characteristics such that for an electromagnetic wave, having a wavelength greater than said linear dimension by an order of magnitude and traveling along the axis of said transmission line, the effective permeability and permittivity of transmission are simultaneously negative and such that said transmission line provides a continuous passband for said electromagnetic wave along said transmission line.

According to yet another aspect of the present invention there is provided a near field focussing device comprising:

a parallel-plate waveguide; and

a negative refractive index metamaterial in line with said waveguide and forming an interface therewith.

According to yet another aspect of the present invention there is provided a coupled-line coupler comprising:

a microstrip line; and

a left-handed material coupled to an edge of said microstrip line.

The present invention provides advantages in that the metamaterial is capable of internal positive and negative wave refraction, and guided wave beam formation, steering, and focusing over regions that are smaller than the wavelength of incident electromagnetic radiation. Moreover, the beam steering capability can be extended beyond the physical boundaries of the metamaterial into the surrounding space to produce a controllable radiation pattern.

BRIEF DESCRIPTION OF THE DRAWINGS

Embodiments of the present invention will now be described more fully with reference to the accompanying drawings in which:

FIG. **1** shows phase matching at an interface between a right-handed material (RHM) and a generic material i.e. a right-handed material or a left-handed material;

FIG. **2** is a two-dimensional (2-D) loaded left-handed transmission line unit cell;

FIG. **3** shows a plane wave illuminating a RHM/LHM interface at 2 GHz;

FIG. **4** illustrates the focusing effect at an interface between right-handed and left-handed media;

FIG. **5** shows a point source illuminating a RHM/RHM interface at 2 GHz;

FIG. **6** shows a point source illuminating a RHM/LHM interface at 2 GHz;

FIG. **7** shows a focusing device including a negative refractive index metamaterial interfaced with a parallel-plate waveguide at 1.5 GHz;

FIG. **8** shows experimental data demonstrating focal regions observed in the focusing device of FIG. **7** at 1.55 GHz and 1.65 GHz and the lack thereof at 2.55 GHz when operating beyond the left-handed frequency band;

FIG. **9** shows a planar resonance cone metamaterial including a square-celled transmission-line network containing inductors and capacitors oriented orthogonally and positioned over a ground plane;

FIG. **10** shows one side of the square-celled transmission-line network of FIG. **9** illustrating attached coaxial connectors;

FIG. **11** shows an opposite side of the square-celled transmission-line network of FIG. **9**, with chip capacitors and meander-line inductors in place.

FIG. **12** shows single-stage, 2-stage, 4-stage and 8-stage 0° phase-shifting circuits;

FIG. **13** is a schematic diagram of a 1-D phase-shifting unit cell;

FIG. **14** shows an unmatched dispersion relation for the phase-shifting unit cell of FIG. **13**;

FIG. **15** shows a matched dispersion relation for the phase-shifting unit cell of FIG. **13**;

FIG. **16** shows phase responses of a +10° phase-shifting device, a −350° transmission line, a −10° phase-shifting device and +10° L-C line;

FIG. **17** shows phase responses of one-stage, 4-stage and 8-stage 0° phase-shifting devices;

FIG. **18** shows phase and magnitude responses of +10° and −10° 4-stage phase-shifting devices;

FIG. **19** shows a 4-stage +10° phase-shifting device;

FIG. **20** is a plasma parameter diagram showing elliptic and hyperbolic regions;

FIG. **21** is a schematic diagram of a focusing device in the form of a uniform anisotropic planar L-C grid over ground, with corner feed and resistive edge-loading;

FIG. **22** shows a uniform-grid moment-method simulation displaying corner-fed resonance cones at eight frequencies;

FIG. **23** shows node voltage across 50Ω terminating resistors at numbered points around the edge of the grid of FIG. **21**, derived from the simulation data displayed in FIG. **22**;

FIG. **24** shows a focusing device in the form of a non-uniform L-C grid, wherein the inductors and capacitors have been transposed in the upper part of the grid and wherein the transition region includes a row of elongated cells;

FIG. **25** shows the physical layout of the grid of FIG. **24**;

FIG. **26** shows resonance cone refraction at 1.2 GHz;

FIG. **27** shows a focusing simulation at 1.3 GHz showing the grid-to-ground voltage magnitude;

FIG. **28** shows focusing measurements at 1.3 GHz showing the normalized **S21** magnitude;

FIG. **29** shows a focusing simulation at 1.7 GHz showing the grid-to-ground voltage magnitude;

FIG. **30** shows focusing measurements at 1.7 GHz showing the normalized **S21** magnitude;

FIG. **31** shows a MS/LHM coupled-line coupler and schematic of a coupler unit cell;

FIG. **32** shows a dispersion diagram for MS and LHM lines used in the coupler of FIG. **31** and photograph of a MS/LHM coupler and a conventional MS/MS coupler of the same length, line spacing and propagation constant;

FIG. **33** shows a spatial Fourier transform of line voltages on the MS line and LHM line of the MS/LHM coupled-line coupler of FIG. **31**;

FIG. **34** shows the coherence length of the MS/LHM coupled-line coupler of FIG. **31**;

FIG. **35** shows a comparison of a MS/MS coupler with a MS/LHM coupler of the same length, line spacing and propagation constant for coupled power, isolation, reflection, and arbitrary coupling;

FIG. **36** shows MS/LHM branch-line couplers;

FIG. **37** shows even/odd mode equivalent circuits of a type **1** MS/LHM branch-line coupler;

FIGS. **38***a* to **38***d* show scattering parameters of type **1** MS/LHM branch-line couplers compared to a regular branch-line coupler and scattering parameters of a type **2** MS/LHM branch-line coupler;

FIG. **39** shows a planar transmission-line left-handed lens;

FIG. **40** shows the measured vertical electric field detected 0.8 mm above the surface of the lens of FIG. **39** at 1.057 GHz;

FIG. **41** shows the measured vertical electric field above row 0 of the lens of FIG. **39** at 1.057 GHz;

FIG. **42** shows the measured vertical electric field at the source (dashed curve) and image (solid curve) planes along with the diffraction-limited image (solid curve with circles);

FIG. **43** shows an antenna formed of 6 macrocells of resonance cone metamaterial; and

FIGS. **44** and **45** show the grid-to-ground voltage of a six cell antenna operating at 1 GH_z and the resulting radiation pattern.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

The present invention relates generally to a metamaterial that supports propagating backward waves and thus exhibits left-handedness. Such metamaterials inherently support 2-D wave propagation, which is desirable for antennas, antenna beam formers, planar spectrum analyzers, compact RF/microwave lenses and antennas, phase compensators, antenna-integrated multiplexers and other microwave circuit applications.

To illustrate the concept of negative refraction, consider a phase-matching argument at the interface between a right-handed medium **M1** and another generic medium **M2**, as shown in FIG. **1**. The sign of the index of refraction of medium **M2** is not a concern. Consider an incident plane wave in medium **M1** with a wave vector k_1 (i.e. such that the x-component of k_1 is positive). A refracted wave with a wave vector k_2 is then established in medium **M2** such that the tangential wave vector components k_{1t} and k_{2t} are equal across the interface. This is the basis for Snell's Law, and it permits two scenarios for the orientation of k_2, represented as Case **1** and Case **2** in FIG. **1**. The conservation of energy also insists that the normal components of the Poynting vectors S_1 and S_2 remain in the positive-x direction through both media. If the medium **M2** is a conventional RHM, then refraction occurs as illustrated by Case **1**. However, if medium **M2** is a medium supporting propagating backward waves (i.e. a LHM), it is implied that power is propagated along the direction of phase advance, which requires in FIG. **1** that k_2 and S_2 be antiparallel. Consequently, the direction of k_2 is specified uniquely for backward-wave structures as illustrated by Case **2**. Under such conditions, power is refracted through an effectively negative angle, which implies a negative index of refraction.

Metamaterials exhibiting negative indices of refractions and applications therefor that enable electromagnetic radiation to be controlled and guided will now be described.

Planar Metamaterials

Turning now to FIG. **2**, a shunt-centered unit cell for a metamaterial or left-handed medium (LHM) that synthezises a negative refractive index is shown. In this particular example, the metamaterial is in the form of a two-dimensional (2-D), L-C loaded substantially orthogonal, coplanar transmission line network. In the unit cell, $Z_C=(j\omega C)^{-1}$ and $Y_L=(j\omega L)^{-1}$ represent the per-unit-length capacitor impedance and inductor admittance, respectively.

The unit cell dimensionality represented by Δd is provided by the host transmission line medium. At the continuous medium limit, Δd<<λ, the 2-D telegrapher's equations representing the structure of the unit cell can be expressed as:

$$\frac{\partial v_y}{\partial z} = -i_z\left(\frac{1}{j\omega C\Delta d}\right), \frac{\partial v_y}{\partial x} = -i_x\left(\frac{1}{j\omega C\Delta d}\right) \tag{1}$$

and

$$\frac{\partial i_z}{\partial z} + \frac{\partial i_x}{\partial z} = -v_y\left(\frac{1}{j\omega L\Delta d}\right). \tag{2}$$

Combining equations (1) and (2) yields:

$$\frac{\partial^2 v_y}{\partial x^2} + \frac{\partial^2 v_y}{\partial z^2} + \beta^2 v_y = 0, \quad \beta = -\frac{1}{\omega\sqrt{LC}\,\Delta d} \tag{3}$$

where β is the propagation constant. The phase and group velocities are antiparallel and are given by:

$$v_\phi = \frac{\omega}{\beta} = -\omega^2\sqrt{LC}\,\Delta d, \; v_g = \left(\frac{\partial\beta}{\partial\omega}\right)^{-1} = +\omega^2\sqrt{LC}\,\Delta d. \tag{4}$$

Subsequently, a negative refractive index can be defined as:

$$n = \frac{c}{v_\phi} = -\frac{1}{\omega^2\sqrt{LC}\sqrt{\mu_0\varepsilon_0}\,\Delta d}. \tag{5}$$

It is interesting to note that it is possible to achieve the same result for the phase velocity and index of refraction if an equivalent negative permittivity and permeability are defined as:

$$\mu_{eq} = -\frac{1}{\omega^2 C\Delta d}, \; \varepsilon_{eq} = -\frac{1}{\omega^2 L\Delta d}, \tag{6}$$

so that the total stored time-averaged energy, expressed by:

$$W = \frac{\partial(\mu_{eq}\omega)}{\partial\omega}|E|^2 + \frac{\partial(\varepsilon_{eq}\omega)}{\partial\omega}|H|^2, \tag{7}$$

remains positive.

In order to model the 2-D wave equation represented by equation (3), metamaterial including an array of unit cells, each as depicted in FIG. **2**, was implemented in a microwave-circuit simulator simulator. To simulate the incidence of waves on this LHM in a circuit environment, a right-handed medium (RHM) was also designed. The topology of the unit cells in the RHM were identical to those of FIG. **2**, except that the roles of C and L were interchanged, giving rise to the well-known LC low-pass network representation of free-space propagation.

The specification of the unit cell parameters in both media provides information about the permissible operating frequencies, the relative refractive indices, and also the required inductance and capacitance values. In both media, the wave impedance is given by:

$$Z_M = \sqrt{\frac{L}{C}}. \tag{8}$$

It is therefore reasonable to begin with the simple constraint that the two media be matched, and moreover, matched to free space (Z_M=377Ω).

In the limit Δd<<λ in equation (5), it is not possible to specify directly indices of refraction in the individual media, but it is possible to specify a relative index of refraction through the ratio of their respective phase shifts per unit cell $\beta_{LHM}\Delta d/\beta_{RHM}\Delta d$. The LHM was designed to be denser than the RHM, arbitrarily by a factor of 2. Specifically, phase shifts per unit cell of $|\beta_{RHM}\Delta d|=0.25$ and $|\beta_{LHM}\Delta d|=0.5$ were chosen in the right-handed and left-handed media, respectively. Choosing an operating frequency of 2 GHz, the LHM and RHM unit cell capacitive and inductive elements were specified as (C_{LHM}=422.5 fF, L_{LHM}=60.0 nH) and (C_{RHM}=52.8 fF, L_{RHM}=7.5 nH), respectively. The corresponding cut-off frequencies are given by:

$$f_{c,LHM} = \frac{1}{4\pi\sqrt{L_{LHM}C_{LHM}}},\ f_{c,RHM} = \frac{1}{\pi\sqrt{L_{RHM}C_{RHM}}}, \tag{9}$$

which were found to be 500 MHz and 16 GHz, respectively. Finally, the designed LHM and RHM arrays were appropriately terminated with matching resistors on all edges and simulated with a microwave-circuit simulator.

To verify the ray picture presented in the phase-matching argument of FIG. **1**, a RHM/LHM interface was constructed using 42×42 RHM and LHM unit cell arrays with $\beta_{RHM}\Delta d=+0.25$ and $\beta_{LHM}\Delta d=-0.5$, yielding a relative refractive index of −2. The RHM unit cell array was excited with a plane wave, which was simulated using sequentially phase-shifted voltage sources along the left boundary of the unit cell array. FIG. **3** illustrates the plane wave illuminating the RHM/LHM interface at 2 GH$_z$ for an incident angle of θ_{RHM}=29° with the axes being labelled according to unit cell number and the right vertical scale designating radians. The steepest phase descent in the LHM is observed along the direction of propagation, which is θ_{LHM}=−14° from the normal, in exact correspondence with Snell's Law for the given design parameters.

Turning now to FIG. **4**, it can be seen that a RHM/LHM planar interface **410** can also be used to focus electromagnetic waves from a point source located inside the RHM, within the LHM. These conditions can be modelled by exciting a single node **400** inside the RHM and observing the magnitude and phase of the voltages to ground at all points in the LHM. The focusing effect manifests itself as a "spot" distribution of voltage **420** at a predictable location in the LHM.

An alternative embodiment of an RHM/LHM interface used to generate the results of FIG. **3** will now be examined. In this embodiment finite length transmission line sections (k, Z_0, finite Δd) are inserted in each LC unit cell. In order that the host transmission-line medium does not significantly alter the propagation constant predicted by equation (5), it is necessary to modify the loading elements C_{LHM} and L_{LHM} to compensate for the presence of the distributed transmission line parameters. In the final design, the lines in each unit cell are designed to be air-filled, with Δd=5 mm. From equation (5), and using the appropriately compensated loading element values, the corresponding equivalent, absolute index of refraction of the LHM is approximately equal to −2.4. To maintain the relative refractive index of −2, the absolute index of refraction of the RHM is made to be +1.2.

In the interest of completeness, cases of both positive and negative refraction are examined, with the host transmission line medium in place. In the first case, a 42×21 RHM unit cell array is interfaced with another 42×63 RHM unit cell array with a relative refractive index of +2. The plane wave source is placed 11 unit cells into the first RHM. Here, focusing is not expected since Snell's Law for positive-index media predicts a continued divergence into the second RHM. FIG. **5** shows a point source illuminating the RHM/LHM interface at 2 GH$_Z$ as well as the corresponding magnitude and phase plots of these results with the axes labelled according to cell number. As can be seen, the results confirm that the cylindrical wave excitation diverges into the second medium.

In the second case, a RHM unit cell array is interfaced with a LHM unit cell array with a relative refractive index of −2. The unit cell array dimensions and source location are as specified above. This arrangement is expected to show focusing inside the LHM, in accordance with FIG. **4**. The paraxial limit dictates a focus in the LHM at twice the distance of the source from the interface, or near unit cell **44** of the array. FIG. **6** shows the point source illuminating the RHM/LHM interface at 2 GH$_Z$, as well as the corresponding magnitude and phase results with the axes labelled according to cell number. As can be seen, the results show focusing in the LHM, manifested in increased voltage amplitudes (nearly 65% of the source amplitude), and also in the reversal of the concavity of the wavefronts at both the RHM/LHM boundary and the expected focal point.

For applications in which waves are propagated only in a single dimension e.g. phase-shifting lines as will be described, the metamaterial above may be simplified to a 1-D array using a single transmission line i.e. a linear LHM metamaterial is a simplified case of the planar metamaterial.

Since the metamaterial exhibits left-handedness and allows electromagnetic radiation to be guided and controlled, the metamaterial is useful in a number of applications. Examples of applications for the present metamaterial will now be described.

Near-Field Focusing Device

Turning now to FIG. **7**, a focusing device is shown that includes a 55 mm×50 mm parallel-plate waveguide serving as a RHM interfaced with a negative refractive index metamaterial including an 11×6 unit cell array around 1.5 GHz. Each unit cell has a dimension of 5 mm. The fabricated 11×6 unit cell areas includes 400 μm wide microstrip lines on a 60 mil Rexolite® dielectric ($\in_r$=2.53) substrate. Shunt chip inductors are embedded into 1 mm holes drilled into the substrate at the appropriate unit cell sites. Chip capacitors are surface-mounted between gaps etched into the grid lines, and additional capacitors are placed at the array edges to maintain unit cell uniformity throughout. The two media meet along a flat border, defining their respective extents. The LHM extent is parallel to and displaced from the nearest microstrip line. The parallel-plate waveguide is excited with a shorted vertical probe connected to a SMA connector. Matching chip resistors of 100Ω are used to terminate the unit cells at the LHM and parallel-plate waveguide boundaries. The focusing device as shown measures approximately 60 mm×95 mm×1.5 mm.

FIG. **8** shows a sample of measured vertical E-field distributions over the 11×6 LHM unit cell array at 1.55 GHz, 1.65 GHz, and 2.55 GHz. The first two frequencies lie in the

LH passband, whereas the last frequency occurs in a second (RH) passband. In accordance with theory, the field distributions at 1.55 GHz and 1.65 GHz exhibit focusing, manifested by a localized region of increased transmission through the focusing device. The maximum focal amplitude obtained was observed at a frequency of 1.65 GHz (the central plot of FIG. **8**), where the distinction between the peak and the two edges of the spot was noted to be approximately 15 dB. Furthermore, the concavities of the wavefronts of FIG. **8** also indicate the convergent progression of phase associated with focusing. As expected, neither phenomenon is evident at 2.55 GHz, since the second passband does not possess left-handed properties. As will be appreciated, the results demonstrate focusing of an incident cylindrical wave over an electrically short area, i.e. near-field focusing.

Planar Resonance Cone Metamaterial

Turning now to FIGS. **9** to **11**, a planar resonance cone metamaterial is illustrated. In this embodiment, the metamaterial includes a substantially orthogonal, coplanar, periodic square-celled transmission line network that is analogous to a sheet of uniaxial plasma with permittivities of opposite signs in two orthogonal, in-plane directions. As can be seen, the transmission line network is disposed over a ground plate and parallel to it. Electromagnetic radiation is fed to the transmission line network at one corner with respect to the ground plate. The transmission line network is connected to instruments through coaxial connectors as shown in FIG. **10** and includes chip capacitors and meander-line inductors as shown in FIG. **11**. Resonance cone angle scans with frequency and with peak frequencies detected at the midpoints of the transmission line network sides, yielded an end-to-end frequency ratio of 2:1 with a transmission line network of only 4 cells by 4 cells. A center frequency of 1.4 GHz was used. The commercial chip capacitive and inductive elements used to construct the transmission line network yielded a compact array. A larger transmission line network having 24 cells by 24 cells was also modelled computationally. This transmission line network displayed resonance cones that showed little or no near-field beam spreading with increasing distance from the source. When a cone beam encountered an edge of the transmission line network with insufficient resistive absorption, specular reflection of the cone beam was observed.

Changing the reactive loads creates a different transmission line network. For example, a "transpose" transmission line network can be created by interchanging the inductive and capacitive elements.

Phase-Shifting Line Using a Linear Metamaterial

The metamaterial in accordance with the present invention is also suitable for use in a compact 1-D phase-shifting device that can be used to synthesize arbitrary transmission phases. Such as shifting device includes sections of LHM metamaterial lines cascaded with sections of conventional RHM transmission lines. Several embodiments of such a phase-shifting device, having a 1, 2, 4, and 8 stages can be seen in FIG. **12**. The phase-shifting device offers significant advantages when compared to standard delay transmission lines. The phase-shifting device is more compact in size and can achieve a positive, a zero, or a negative phase shift while occupying the same or shorter physical length. The phase-shifting device also exhibits a linear, flatter phase response with frequency, leading to shorter group delays.

In a LHM, the phase leads in the direction of group velocity, therefore incurring a positive phase shift with propagation away from the source. In addition, it is well known that in conventional RHM transmission lines, the phase lags in the direction of positive group velocity, thus incurring a negative phase shift with propagation away from the source. It therefore follows that phase compensation can be achieved at a given frequency by cascading a section of a LHM with a section of a RHM to form a unit cell with a given phase shift.

The unit cell of the phase-shifting device is shown in FIG. **13** and has total dimension d_o. The unit cell includes a host transmission line medium TL periodically loaded with discrete lumped element components, L_o and C_o.

The dispersion characteristics of the unit cell can be determined by the following equations:

$$\cos(\beta_{Bloch}d) = \cos(\beta_{TL}d)\left(1 - \frac{1}{4\omega^2 L_o C_o}\right) + \sin(\beta_{TL}d)\left(\frac{1}{\omega C_o Z_o} + \frac{Z_o}{\omega L_o}\right) - \frac{1}{4\omega^2 L_o C_o} \tag{10}$$

where:

β_{Bloch} is the Bloch propagation constant; and

$\beta_{TL}=\omega\sqrt{LC}$ is the propagation constant of the host transmission line.

In order to consider a series of cascaded unit cells as an effective periodic medium, the physical length of the unit cell must be much smaller than a wavelength, therefore restricting the phase shift per unit cell and the length of the transmission lines to small values (βBlochd<<1 and βTLd<<1). Thus, an effective propagation constant, βeff, can be defined for the medium by the following equation:

$$\beta_{eff} = \pm\sqrt{\left(L - \frac{1}{\omega^2 C_o d}\right)\left(C - \frac{1}{\omega^2 L_o d}\right)} \tag{11}$$

FIG. **14** shows the dispersion relation for a unit cell with typical parameters C_o=15 pF, L_o=20 nH, Z_o=50Ω and $\beta_{TL}d$=0.28 rad. It can be observed that the corresponding dispersion diagram exhibits a band structure with two distinct passbands and stopbands. Expressions for the pertinent cut-off frequencies as indicated in FIG. **14** are as follows:

$$f_b = \frac{1}{4\pi\sqrt{L_o C_o}},\ f_{c1} = \frac{1}{2\pi\sqrt{L C_o}},\ f_{c2} = \frac{1}{2\pi\sqrt{L_o C}} \tag{12}$$

Equating f_{c1} and f_{c2} causes the stopband between these two cut-off frequencies to close and the band becomes continuous. This corresponds to the following matching condition for the LHM and RHM sections:

$$Z_o=\sqrt{L_o/C_o}=\sqrt{L/C} \tag{13}$$

Under this condition, it can be shown that the effective propagation constant of equation (11) simplifies to the expression:

$$\beta_{eff} \approx \omega\sqrt{LC} + \frac{-1}{\omega\sqrt{L_o C_o}} \tag{14}$$

Expression (14) can be interpreted as the sum of the propagation constants of the host transmission line and a uniform LHM line. For a transmission line medium with periodicity d and phase shift $\phi_{TL}=\beta_{TL}d$, the total phase shift per unit cell, $|\phi_o|$, given that the matching condition of equation (13) is satisfied, can be written as:

$$|\phi_o| = \beta_{eff} d = \phi_{TL} + \frac{-1}{\omega\sqrt{L_o C_o}} \tag{15}$$

Thus equations (13) and (15) can be used to determine unique values for L_o and C_o for any phase shift, ϕ_o, given a transmission line section with intrinsic phase shift β_{TL} and characteristic impedance Z_o.

The unit cell of FIG. **13** was implemented in coplanar-waveguide (CPW) technology at a design frequency of f_o=0.9 GHz using microwave-circuit simulator. The lumped element components and the transmission lines were assumed to be ideal. The transmission lines were designed with Z_o=50Ω, $\in_r$=2.2, dielectric height, h=20 mils and length d=7.4 mm. The dispersion relation for the unit cell with ϕ_o=0° is shown in FIG. **15**. The impedance matching condition of equation (13) has been satisfied and therefore, the stopband has been closed. It can also be observed that at f_o=0.9 GHz, the phase shift is 0°.

Based on the above parameters a +10-degree phase-shifting device was designed including 4 unit cells of total physical length 0.11λCPW=32 mm, with loading element values of C_o=21.66 pF and L_o=54.15 nH. The resulting phase characteristic of the phase-shifting device is shown in FIG. **16**. It can be observed that the phase shift is +10 degrees at fo=0.9 GHz with a phase slope of −95.7-degrees/GHz. In addition, the phase response around fo is linear with respect to frequency, implying a constant group delay for the phase-shifting device. Also shown in FIG. **16** is the phase characteristic of a conventional 50Ω transmission line of length 35/36 λCPW=275.7 mm for comparison. The transmission line exhibits as expected a −350-degree phase shift at fo, which is equivalent to +10 degrees, with a phase slope of −389-degrees/GHz. In addition, its physical length is approximately 9 times longer than the 4 unit cell phase-shifting line.

Referring back to FIG. **12**, an impression of the relative sizes of the two structures can be obtained. The proposed +10-degree phase-shifting device offers two main advantages over its equivalent transmission line implementation, namely a +10-degree phase shift while occupying significantly less physical space than a conventional transmission line, and a shorter group delay which is desirable for applications that require broadband operation.

Also shown in FIG. **16** is the phase characteristic of a 4-stage −10° phase-shifting device with loading element values of C_o=29.46 pF and L_o=73.65 nH. It can be observed that the −10° phase characteristic is similar to the +10° phase characteristic, with a phase slope of −72.9°/GHz. The −10° phase-shifting device occupies the same physical length as the +10° phase-shifting device and has loading element values that are at a ratio of approximately 1.4:1 compared to the ones used for the +10° phase shifting device. This implies that a reasonable variation of the loading element values using tuneable components can produce either a positive or a negative phase shift.

FIG. **16** also shows the phase characteristic of a 4-stage +10° uniform backward wave L-C line with loading element values of C_o=81.06 pF and L_o=202.64 nH. This corresponds to the unit cell of FIG. **13** with the transmission line sections removed. Setting therefore ϕTL to zero in equation (15), implies that ϕ_o will always remain negative, corresponding to a phase advance for a positively traveling plane wave of the form exp(−jβeffd). The absence of transmission lines implies that any combination of C_o and L_o values will produce a phase shift that will always be greater than 0 degrees for the unit cell. This is verified by the phase characteristic of the L-C line in FIG. **16** that stays well above 0 degrees at all times. Therefore, although the backward wave L-C line and the proposed unit cell of FIG. **13** can both provide positive phase shifts, the latter has the additional advantage of being able to provide a negative phase shift by simply varying the loading element values. More importantly, short, broadband 0-degree phase shift lines can be realized which can be used in numerous applications.

The 1-D phase-shifting structures were constructed using CPW technology on a Rogers™ RT5880 substrate with ∈r=2.2, and a dielectric height, h=20 mils. Standard size 0402 Panasonic ECJ-0EC capacitors were used for C_o with dimensions L=1 mm, W=0.5 mm, and H=0.5 mm, and standard size 0603 Panasonic ELJ-RE/FJ inductors were used for L_o with dimensions L=1.6 mm, W=0.8 mm, and H=0.8 mm. The 50Ω transmission lines were designed to have a length of d=7.4 mm, and a gap of s=0.6 mm was created to accommodate the series capacitors, resulting in a unit cell of total length d_o=8 mm. At 0.9 GHz the transmission lines were designed with a center conductor width of w=4 mm and a gap between the center conductor and the ground plane of g=0.106 mm. In addition, two 5 mm transmission lines were added to each end of the phase-shifting devices in order to accommodate SMA connectors. A Thru-Reflect-Line (TRL) calibration was carried out in order to remove the effects of the connectors and establish the reference planes at the input and output ports of the phase-shifting devices. All measurements were carried out on a HP8753C series vector network analyzer. The simulated results were obtained by replacing the ideal models for the capacitors and inductors with their corresponding S-parameter files provided by the vendors. Since the S-parameter files were extracted directly from device measurements, they therefore contain all the parasitic values associated with each component.

FIG. **17** shows the simulated (- - -) and measured (-) phase responses of 1-stage, 4-stage and 8-stage 0° phase-shifting devices with loading element values of C_o=12 pF and L_o=100 nH. It can be observed that the experimental results match very closely the simulated results, with phase slopes of −30.2°/GHz, −122.1°/GHz and −237.5°/GHz for the phase-shifting devices respectively. The measured insertion losses are 0.1 dB, 0.7 dB and 1.6 dB respectively at f_o=0.9 GHz. It is therefore evident that any number of stages can be used in order to produce the desired phase shift, however as the number of stages increases, so does the slope of the phase characteristic. For broadband applications therefore, it is desirable to keep the number of stages to a minimum.

FIG. **18** shows the simulated (- - -) and measured (-) phase responses of a 4-stage +10° phase-shifting device with loading element values of C_o=12 pF and L_o=68 nH and a 4-stage −10° phase-shifting device with loading element values of C_o=15 pF and L_o=120 nH. It can be observed that the experimental results correspond very closely to the simulated results with phase slopes of −136.6°/GHz and −108.4°/GHz for the +10° and −10° phase-shifting devices respectively. Also shown in FIG. **18** is the magnitude response of the −10° 4-stage phase-shifting device. The simulated (- - -) and measured (-) results correspond closely, however as the number of stages is increased the phase-shifting device begins to radiate, and resembles a leaky-wave structure. Therefore the phase-shifting devices should be kept short in length to avoid unnecessary energy leakage. The measured insertion losses at f_o=0.9 GHz are 0.4 dB and 0.5 dB respectively for the +10° and −10° phase-shifting devices.

The phase-shifting devices described above use cascaded sections of LHM lines and conventional transmission lines and offer significant advantages over conventional delay lines and uniform backward-wave L-C lines. The phase-shifting devices are compact in size, can be easily fabricated using standard etching techniques and exhibit a linear phase response around the design frequency. The phase-shifting devices can incur a negative or a positive phase, as well as a 0° phase depending on the values of the loading elements, while maintaining a short physical length. In addition, the phase incurred is independent of the length of the phase-shifting devices. Due to the compact, planar design, the phase-shifting devices are easily integrated with other microwave components and devices making them ideal for broadband applications requiring small, versatile, linear phase shifters.

Spectrum Analyser Using a Resonance Cone Metamaterial

The resonance cone metamaterial is useful in spectrum analysis. To illustrate this, a spectrum analyser incorporating resonance cone metamaterial was simulated. For the representation of metamaterial, analyser simulations employed a full-electromagnetic, thin-wire moment-method program, permitting the insertion of lumped circuit elements in finite-length wire segments. The basic network simulated is shown in FIG. **21**, and includes a 12×12 unit cell array, each unit cell being a 2.5 mm square, yielding a total grid size of 30 mm×30 mm. The grid is disposed above the ground plane a distance of 2.5 mm. The capacitance was 2 pF per segment and the inductance was 5.6 nH per segment. The chosen capacitance and inductance values were selected for availability as well as to achieve the desired resonance cone effects using existing X-Y scanning and network analysis equipment. The inductors used in the experiment had a manufacturer nominal quality factor Q of around 27. This was achieved in the simulation by adding a 1.6Ω resistor in series with each inductor. A resistor of the same value was added in series with each capacitor as well. Also shown in FIG. **21** is a 100 kΩ resistor from each grid intersection to ground to enable deduction of the grid-to-ground voltage from the computed resistor current. Along the edges, 50Ω resistors were connected to ground in place of the 100 kΩ resistors. For this value of edge resistance, resonance cone specular reflection from the edges is not significant.

Contour plots of grid-to-ground voltage magnitude and the Poynting vector real parts evaluated from the electric and magnetic fields on the ground-plane level at the center of each unit cell for a corner feed of 1 volt are shown in FIG. **22**. The simulated resonance cones and the way the cone orientation scans with frequency over the eight frequencies employed are illustrated. The arrows depict the Poynting vector real parts calculated on the ground plane. From the Poynting vector plot it can be seen that power flow is directed along the resonance cones. Apparent specular reflection from the grid edges is sometimes visible (for example at 0.7, 0.9, 1.9 and 2.3 GHz) but generally it is weak. As well, very weak parasitic resonance cones can be seen in FIG. **22** (for example at 0.9 and 2.3 GHz), probably caused by scattering from the inhomogeneities inherent in the grid structure. FIG. **23** shows the frequency response for the voltages across the 50Ω resistors to ground at all points along two edges of the grid.

Focusing Device Using Two Resonance Cone Metamaterials

Another focusing device making use of resonance cone metamaterial was created using a two media plane including metamaterial similar to that shown in FIG. **11**. By interchanging (transposing) the inductors and capacitors in the grid in part of the media, an interface is formed, separating two media. If the plasma negative and positive permittivities are interpreted by representing the media in terms of arrays of inductors and capacitors respectively, then interchanging the inductors and capacitors corresponds to moving the operating point from one hyperbolic region to the other in the parameter space as shown in FIG. **20**. The hyperbolic regions characterize the relevant partial differential equation. The vertical axis involves the electron cyclotron frequency ω_c and is proportional to the square of the ambient magnetic field, while the horizontal axis involves the plasma frequency ω_p and is proportional to the ambient electron density. The capacitor and inductor symbols characterize the impedance properties of the plasma medium as they effect the input of impedance of a small, spherical RF probe immersed in the plasma medium. These two hyperbolic regions suggest two quite different periodic physical media that, when interfaced, display unusual refraction phenomena. If the inductors and capacitors are regarded as physical entities and not just as aids to interpretation, a physical inductor-capacitor (L-C) network becomes a metamaterial that exhibits a resonance cone phenomena. This enables a two-medium configuration with a linear interface between the original medium and the transpose medium.

The component layout of a focusing device including multi-resonance cone materials is shown in FIG. **24**, and its physical grid realization is shown in FIG. **25** in which the grid elements are chip-type surface-mount inductors and capacitors soldered together. The grid has 12 by 12 unit cells. Medium **1** is constituted by the lower 5 rows of the grid and includes inductors on the horizontal segments and capacitors on the vertical segments. Medium **2** is constituted by the upper 6 rows of the grid and includes capacitors on the horizontal segments and inductors on the vertical segments. The sixth row of cells is the transition region, or interface, where the lower edges of the unit cell carry inductive loads, the upper edges carry capacitive loads, and the vertical segments carry a capacitor on the lower half and an inductor on the upper half. Therefore the boundary between the two media lies in the middle of the transition region.

To accommodate the physical size of the chip-type inductors and capacitors, the unit cells in the transition region are elongated i.e. the vertical dimension is doubled compared to unit cells elsewhere in the grid. The overall size of the grid is 30 mm×32.5 mm and the boundary is a horizontal line 15 mm from the bottom of the grid. The grid is aligned so that individual vertical transmission lines run directly through both media and the interface. In the refraction experiments, the vertical E-field is picked up by an open-ended coaxial probe positioned sequentially just above each conductor intersection, and the probe signal is fed to a network analyzer with output **S21** which, when plotted on a linear scale, is approximately proportional to the grid-to-ground voltage.

The two-medium result for corner feed with the feed point being below point A in FIG. **24**, is shown in FIG. **26**, which displays clearly the negative refraction of the resonance cone as it traverses the transposition interface. Experimental results at the top exhibit contours of normalized /**S21**/, and the simulation at the bottom exhibits contours of grid-to-ground voltage magnitude with the arrows depicting the Poynting vector real parts on the ground plane. The counter lines in the bottom plot are at the same levels as the three lowest contour lines in the top plot. Notice that specular reflection from the interface is negligible and there is no visible transmission into the second medium in the same direction as the incident cone. This can also be seen from the

Poynting vectors in the bottom plot of FIG. **26**, which displays the simulation results, where the power flow associated with specular reflection from the interface is negligible and no power along the original resonance cone is transmitted into the second medium. In order to extend the refraction results to include focusing, the physical set-up of FIG. **25** was used with coaxial feed from beneath the ground plane to point B on the grid in FIG. **24**, near the middle of one of the two planar metamaterials.

Simulation and experimental results are shown in FIGS. **27** and **28** at 1.3 GH_Z. The arrows display the Poynting vector real ports calculated on the ground plane. Heavy contour is at 0.707 of focal maximum of 0.867 units of normalized /**S21**/. Simulation and experimental results are also shown in FIGS. **29** and **30** at 1.7 GH_Z. In FIG. **29** heavy contour is at 0.707 of focal maximum of 0.708 and in FIG. **30** heavy contour is at 0.707 of focal maximum of 0.764. In all cases, one can see the cones emanating from the feed point (the "source"), the backward refraction at the y=15 mm interface, and the cones merging at the "focus". Note that in FIGS. **27** and **29** the power flow closely follows the directions of the resonance cones. Furthermore, FIG. **29** shows the power flow emanating from the source and following the resonance cones to converge at the focus, from where it diverges again into a new set of resonance cones. In FIGS. **29** and **30**, the focal region is taken to be bounded by the contour that is at 0.707 of the focal region maximum, so this contour may be termed the "half-power" contour. It is worth noting, that at 1.7 GHz, the experimental focal region boundary lies within a square measuring 1/25th free-space wavelength on a side, so the phenomenon properly may be termed "sub-wavelength focusing". In contrast, at 1.3 GHz the focal point is too close to the upper edge of the material to exhibit an equally well localized focal region, and thus this represents a useful low-frequency limiting case for the particular overall grid size in use. As well, with increasing frequency, it should be noted that the focal region moves gradually toward the boundary between the two media.

Couplers Using Linear Metamaterials

The present metamaterial is also useful in couplers by combining the metamaterial with conventional microchip (MS) transmission lines. In particular, it has been found that the metamaterial is useful in both a coupled-line coupler and a branch-line coupler. The MS/LHM coupled-line coupler is realized by means of one regular microstrip line that is edge-coupled to a LHM line. Such a coupler excites modes on the two lines whose propagation constants are co-directional but whose corresponding Poynting vectors are contra-directional, leading to backward power coupling. Moreover, by comparison to a conventional coupled-line microstrip coupler of the same length and line spacing, the present MS/LHM coupled-line coupler offers better performance in terms of coupled power, return loss and isolation without any bandwidth degradation. The MS/LHM coupled-line coupler can also be designed for arbitrary backward coupling.

The branch-line coupler uses LHM lines to split power equally between the output ports with interesting phase compensation characteristics. In conventional branch-line couplers, the through and coupled ports are −90° and −180° out of phase respectively with reference to the input port. The present branch-line coupler using an LHM line, results in the power emerging from the coupled port to be in phase with the input port and permits a choice of either positive or negative phase-quadrature (±90°) at the through port.

For ease of understanding, the theory behind the coupled-line coupler will be described. A MS/LHM coupled-line coupler can be realized by means of one regular microstrip line that is edge-coupled to a LHM line as shown in FIG. **31**. The LHM is synthesized by periodically loading a host microstrip transmission-line (TL) medium with series capacitors and shunt inductors. In such a backward-wave transmission line, the phase velocity and power flow are anti-parallel.

Coupled-line couplers rely on proximity interaction of fields between transmission lines to transfer power between them. In co-directional coupled-line couplers, the direction of the propagation vectors on the two lines is the same. When such couplers are constructed using either two regular or two LHM lines, the underlying propagation phenomenon is fundamentally similar to that on a corresponding isolated line of the same type. Indeed in such co-directional couplers including two lines of the same type, the power also couples co-directionally and appears at the far end of the coupler (i.e. away from the source). Peculiar features arise when two different types of lines are used. Herein it is shown that a MS/LHM coupler excites co-directional forward traveling waves but delivers power backwards as a result of the power flow being contra-directional.

The analysis presented herein assumes that port **1** in FIG. **31**(*a*) is the excited port. The operation of coupled line couplers can be understood in terms of constructive and destructive interference of the two modes that are allowed to propagate. Using a coupled-mode differential-equation formulation, it is possible to show that the general expressions for voltages and currents on the two lines take the form:

$$V_1=(V_c{}^+e^{-j\beta_c z}+V_c{}^-e^{j\beta_c z})+(V_\pi{}^+e^{-j\beta_\pi z}+V_\pi{}^-e^{j\beta_\pi z}) \tag{16}$$

$$V_2=R_c(V_c{}^+e^{-j\beta_c z}+V_c{}^-e^{j\beta_c z})+R_\pi(V_\pi{}^+e^{-j\beta_\pi z}+V_\pi{}^-e^{j\beta_\pi z}) \tag{17}$$

$$I_1=\frac{1}{Z_{c1}}(V_c^+e^{-j\beta_c z}-V_c^-e^{j\beta_c z})+\frac{1}{Z_{\pi 1}}(V_\pi^+e^{-j\beta_\pi z}-V_\pi^-e^{j\beta_\pi z}) \tag{18}$$

$$I_2=\frac{R_c}{Z_{c2}}(V_c^+e^{-j\beta_c z}-V_c^-e^{j\beta_c z})+\frac{R_\pi}{Z_{\pi 2}}(V_\pi^+e^{-j\beta_\pi z}-V_\pi^-e^{j\beta_\pi z}) \tag{19}$$

Subscripts **1** and **2** refer to line **1** (microstrip) and line **2** (LHM) of the coupler respectively; c and π are the two possible modes and the coefficients R_c and R_π are the ratios of the voltages on line **2** to line **1** for each mode.

To analyze the MS/LHM coupler of FIG. **31**, some simplifying assumptions are made. First, the lines of the coupler are assumed to the far removed from each other so that the coupled-mode propagation constants β_c and β_π correspond to the isolated propagation constants β_1 and β_2 respectively of each line. This implies that β_2 (LHM line) has a negative value. Hence all occurrences of β_c and β_π in equations (16) to (19) can be replaced by β_1 and β_2 respectively. Furthermore line **1** is assumed to be the excited line and therefore, it is anticipated that mode **1** will be the dominant mode for determining the voltage on line **1**. Second, to minimize reflections, only the case of properly terminated lines is considered. Thus each mode travels in one direction only. Mode **1**, being dominant on line **1**, has a forward phase velocity and carries power away from the source. The dilemma lies in the choice of the direction of the phase velocity for mode **2**. To resolve this issue, one can consider the limiting case when the magnitudes of the isolated propagation constants of the modes are close to each other. As mode **1** travels forward, due to phase matching constraints, it is expected that the coupled mode (mode **2** in this case) should also travel forward. The present situation is the reciprocal case and therefore, power from the excited regular MS line will couple backwards into port **2** of the LHM line (see FIG. **31**(*a*)). In the general case when the two mode propagation constants are not close to each other, maximum power can be coupled backwards into port **2** (see FIG. **32**(*a*)) only when the excited modes are co-directional. Hence in the following discussion, only the case of co-directional (phase-wise) coupled modes will be considered.

Lastly, it is assumed that the magnitudes of the impedances for each mode on line **1** are the same and are denoted by Z_1; similarly for line **2** they are denoted by Z_2. Caution must be exercised in choosing the signs of the impedances in the current equations 3 and 4. The impedance term in front of V_c^+ and V_c^- (equation 3) will be positive as usual. Since V_π^+ and V_π^- (equation 3) actually correspond to backward and forward traveling waves respectively on line **1** (due to the negative sign of β_2), and to ensure that the power carried by a forward traveling wave in a regular medium is also forward, $Z_{\pi 1}$ should be negative. Similar arguments applied to line **2** (LHM line) require that the sign of Z_{c2} be negative and that of $Z_{\pi 2}$ be positive to ensure that power flow and phase flow occur in opposite directions.

With these assumptions, equations (16) to (19) can be simplified to:

$$V_1=V_c^+e^{-j\beta_1 z}+V_\pi^-e^{j\beta_2 z} \tag{20}$$

$$V_2=R_cV_c^+e^{-j\beta_1 z}+R_\pi V_\pi^-e^{j\beta_2 z} \tag{21}$$

$$I_1=\frac{V_c^+}{Z_1}e^{-j\beta_1 z}+\frac{V_\pi^-}{Z_1}e^{j\beta_2 z} \tag{22}$$

$$I_2=-\frac{R_c}{Z_2}V_c^+e^{-j\beta_1 z}-\frac{R_\pi}{Z_2}V_\pi^-e^{j\beta_2 z} \tag{23}$$

Applying boundary conditions at port **1** (presence of source V_s and matching source impedance Z_1) and port **4** (termination to Z_2) respectively yields:

$$V_c^+e^{j\beta_1 d}+V_\pi^-e^{-j\beta_2 d}=\frac{V_s}{2} \tag{24}$$

$$R_cV_c^++R_\pi V_\pi^-=0 \tag{25}$$

Solving for V_c^+ and V_π^-, and plugging back into equations (20) and (21), the following expressions for the line voltages along the coupler can be obtained:

$$V_1(z)=\frac{V_s}{2}\frac{R_\pi e^{-j\beta_1 z}-R_ce^{j\beta_2 z}}{R_\pi e^{j\beta_1 d}-R_ce^{-j\beta_2 d}} \tag{26}$$

$$V_2(z)=\frac{V_s}{2}\frac{R_\pi R_c(e^{-j\beta_1 z}-e^{j\beta_2 z})}{R_\pi e^{j\beta_1 d}-R_ce^{-j\beta_2 d}} \tag{27}$$

Equations (26) and (27) confirm that port **4** (z=0) is the isolated port and port **2** (z=−d) is the coupled port. Expressions for power delivered to the through port (port **3**) and coupled port (port **2**) can be derived from equations (26) and (27):

$$P_2=\frac{1}{Z_2}\left(\frac{V_s}{2}\right)^2\frac{(R_CR_\pi)^2(1-\cos((\beta_1+\beta_2)d))}{(R_c^2+R_\pi^2-2R_cR_\pi\cos((\beta_1+\beta_2)d))} \tag{28}$$

-continued

$$P_3=\frac{1}{2Z_1}\left(\frac{V_s}{2}\right)^2\frac{(R_C-R_\pi)^2}{(R_c^2+R_\pi^2-2R_cR_\pi\cos((\beta_1+\beta_2)d))} \tag{29}$$

From equation (29) it can be seen that the power delivered to the through port (port **3**) is minimized (i.e. the power to the coupled port is maximized) when the argument of the cosine function in the denominator takes values which are odd integer multiples of π. This allows the definition of the coherence length of the coupler, which is the optimal length for maximum coupling, as:

$$d_{coherence}=\pi/|\beta_1+\beta_2| \tag{30}$$

Note that in the above equation (30), β_2 is negative. An interesting feature of such a coupler is that it is possible to achieve arbitrary backward coupling by making the magnitudes of the propagation constants of the two modes close to each other.

Two coupled-line couplers were constructed, namely a short MS/LHM coupled-line coupler and a long MS/LHM coupled-line coupler with arbitrary coupling. The substrate used to construct the couplers was a low-loss Rogers™ TMM4® having a relative dielectric constant $\in_r=4.6$ and 50 mils thickness (loss tangent=0.002). The regular MS lines were designed using a microwave-circuit simulator, whereas Matlab® simulations of the dispersion relation of the periodic LHM line were performed to estimate the corresponding propagation constant at the design frequency. Moreover, microwave circuit simulations of the couplers were performed using a microwave-circuit simulator.

The short MS/LHM coupled-line coupler was designed with a finite coherence length to demonstrate the validity of the theory. The long MS/LHM coupled-line coupler has magnitudes of the two modal propagation-constants close to each other. By taking a finite number of cells (the optimal length infinite as defined by equation 15), a comparison with a conventional (MS/MS) coupled-line microstrip coupler of the same length, line spacing, and propagation constant is permitted.

Using a microwave circuit simulator the widths of the lines of the short MS/LHM coupled-line coupler are designed so that both the MS and LHM lines have an impedance of 50Ω when placed 0.4 mm apart. The width of the MS and LHM lines used was 2.383 mm. Using a unit cell size of 5 mm for the LHM line, the choice of the loading elements are 2.7 nH shunt inductors and 0.9 pF series capacitors. Again, the microwave circuit simulator is used to determine the effective dielectric constant for the line segments. This is used in Matlab simulations of the periodically loaded transmission-line based LHM line to extract its dispersion characteristics (see FIG. **32**(*a*)). A design frequency of 2.2 GHz is selected for this configuration so that the magnitude of the isolated propagation constants differ enough (see FIG. **32**(*a*)) to give a short coherence length (see equation 15).

A spatial Fourier transform of the line voltages, extracted by exciting port **1** in ADS simulations, is shown for both lines of the coupled-line coupler at 2.2 GHz using 400 sample points (see FIGS. **33**(*a*) and (*b*)). These plots demonstrate the validity of the assumption that the propagation constants of the coupled modes ($|\beta_1|=85.3\ \text{m}^{-1}$ and $|\beta_2|=195\ \text{m}^{-1}$) are indeed close to the isolated propagation constants of the individual lines (compare to FIG. **32**(*a*)). Moreover the Fourier spectrum demonstrates that both modes propagate forward (co-directional) and that β_1 is dominant on line **1** (MS). Furthermore, the assumption of minimal reflections is evident from the small peaks in the negative region of the spectrum in FIGS. **33**(*a*) and **33**(*b*). In addition, using equation (30), the coherence length is calculated to be 28.6 mm and can be approximated by 6 unit cells of the MS/LHM coupled-line coupler. By comparing the coupled power as a function of frequency for various coupler sizes, this is verified to be true as shown in FIG. **34**. Moreover, the input power at port **1** is $V_s^2/8Z_0$ and the output power at the coupled port is given by equation (28). The implicated voltage ratios R_c and R_π can be estimated from FIGS. **33**(*a*) and **33**(*b*) and the corresponding theoretical value for the coupled power (S_{21}) amounts to −7.46 dB at the coherence length, which is close to the simulation results for a 6 unit cell long coupler (see FIG. **34**).

Fair comparison between a conventional coupled-line backward coupler (MS/MS) and a metamaterial long MS/LHM coupled-line coupler (see FIG. **32**(*b*)) can be made if their propagation constants and line impedances are similar. This is because, it is always possible to slow down the coupled waves and improve coupling in conventional couplers by loading them. For this purpose, a design frequency of 2.8 GHz (FIG. **32**(*a*)) is selected. The widths of the MS and LHM strips of the MS/LHM coupled-line coupler are taken to be 2.45 mm and 2.0 mm respectively so that at a line separation of 0.4 mm at the design frequency, the impedances of each line are close to 50Ω. The loading elements are the same as above with reference to the short MS/LHM coupled-line coupler.

A regular MS/MS backward coupler was also designed at the same frequency with a propagation constant of 108.2 m^{-1} and impedance close to 50Ω. The widths of the strips were taken to be 2.122 mm in this case and the length of the coupler was optimized to 14.5 mm (λ/4).

From MATLAB® simulations of the periodic LHM lines (see FIG. **32**(*a*)) it is found that at 2.8 GHz, the propagation constants for both lines of the MS/LHM coupled-line coupler are similar to those of the MS/MS coupler lines (MS and LHM propagation constants are 108.9 m^{-1} and 105 m^{-1} respectively). This makes the coherence length for the MS/LHM coupled-line coupler very long (equation 15). For comparison with the designed conventional MS/MS coupler, the former is constructed out of 3 unit cells corresponding to a total length of 15 mm (see FIG. **32**(*b*)).

From theory (equation (15)) it is anticipated that this MS/LHM coupled-line coupler will have a long coherence length if the propagation constants of its two modes are close to each other. This is true at 2.8 GHz as shown in the dispersion diagram for the LHM-line (see FIG. **31**(*a*)).

For comparison with a conventional MS/MS backward coupler, simulation and experimental results are presented in FIGS. **35**(*a*) to **35**(*c*), involving a λ/4 MS/MS coupler and a 3 unit cell long MS/LHM coupled-line coupler, both of approximately equal lengths, equal spacing and propagation constants (see FIG. **32**(*b*)). The return loss (S_{11}), isolation (S_{41}) and coupled power (S_{21}) for each coupler were measured using a HP8753D network analyzer and values obtained were compared to microwave-circuit simulation results. As shown in FIG. **35**(*a*), the coupled power of the MS/LHM coupled-line coupler is better than that of the MS/MS coupler. Furthermore, from FIG. **35**(*b*), it is seen that the amount of power leaking into the isolated port of the MS/LHM coupled-line coupler is lower than that of the corresponding MS/MS coupler. In addition, at the design frequency, the return loss of the MS/LHM coupled-line coupler is lower than its conventional counterpart (see FIG. **35**(*c*)). Hence, the metamaterial MS/LHM coupled-line coupler exhibits superior performance in terms of the coupled power, isolation and return loss when compared to a conventional MS/MS coupler of the same length, line separation and propagation constant.

As expected from the dispersion analysis and the theory presented earlier, the MS/LHM coupled-line coupler has a very large coherence length. Under such conditions, arbitrary coupling can be achieved as demonstrated in FIG. **35**(*d*). As the number of unit cells is increased, the coupled power improves and it is possible to achieve close to 0 dB coupled power.

Overall, there is good agreement between theory, simulation and experiment. The predictions of the coherence length as well as the level of the coupled power agree well with the experimentally observed results. The coupled-line coupler demonstrates better performance in terms of coupling, isolation and reflection when compared to a conventional microstrip coupler of the same length, line spacing and propagation constant.

Small discrepancies between simulation and experimental results can be attributed to a number of factors. The absence of accurate models for the inductors used and their self-resonance characteristics affect the design. Moreover chip components with finite dimensions add to the overall electrical length of the MS/LHM coupler which needs to be properly accounted for.

In conventional microstrip branch-line couplers, power splits equally between the through and coupled ports and the ports are −90° and −180° out of phase respectively with reference to the input port. Interesting phase compensation effects can be achieved when some of the branches of the branch-line couplers are replaced by LHM lines.

Two different branch-line coupler designs are presented below denoted as type **1** and type **2**. The type **1** branch-line coupler uses regular microstrip lines for the low-impedance branches and LHM lines for the high-impedance branches (see FIG. **36**(*a*)). The type **2** branch-line coupler is the dual of the type **1** branch-line coupler and utilizes LHM lines for the low-impedance branches and microstrip lines for the high-impedance branches (see FIG. **36**(*b*)). For both branch-line couplers, power splits equally between the two output ports with a 0°-phase shift with respect to the input at the coupled port. Furthermore, the type **1** branch-line coupler offers a negative phase quadrature (−90°) while the type **2** branch-line coupler provides a positive phase quadrature (+90°) at the through port with respect to the input port.

FIG. **36**(*a*) shows the structure of the type **1** branch-line coupler made of alternating MS and LHM lines. The scattering parameters of the coupler can be determined using even-odd mode analysis. It can be easily shown that the impedance looking into a segment of LHM transmission line of length d is given by:

$$Z_{in} = Z_0 \frac{Z_L + jZ_0 \tan\beta d}{Z_0 + jZ_L \tan\beta d} \tag{31}$$

where:

Z_L is the load impedance at the end of the line;

Z_0 is its characteristic impedance; and

β is a negative number corresponding to the propagation constant.

For the purpose of analysis, port **1** is excited by a source V_s of internal impedance Z_0 and all other ports are terminated at Z_0. In the even mode analysis, both port **1** and port **2** are excited by $V_s/2$ sources. In the odd mode, the source at port **1** is $V_s/2$ and the source at port **2** is $-V_s/2$ (see FIG. **37**).

In the even mode, looking down the LHM lines at port **1** and port **3** (see FIG. **36**(*a*)), one sees open λ/8 lines and hence βd is −π/4. From equation (31), this corresponds to shunt loading of the λ/4 MS line at its ends by a susceptance $-jY_0$ (see FIG. **37**). In the odd mode, one can observe shorted λ/8 segments of LHM lines instead, corresponding to shunt loading of the MS line by a susceptance jY_0.

By analyzing the equivalent circuits in FIG. **37**, it is seen that the impedance looking into each port is Z_0. Since the ports are always matched, S_{11}, S_{22}, S_{33} and S_{44} are zero. As the waves in each port travel in only one direction, by superposition of the even and odd modes, S_{21} is zero. In the even mode, the signal at port **1** can be related to the signal at port **3** by the following transfer matrices:

$$\begin{bmatrix} V_s/4 \\ I_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -jY_0 & 1 \end{bmatrix} \begin{bmatrix} 0 & j\frac{Z_0}{\sqrt{2}} \\ jY_0\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -jY_0 & 1 \end{bmatrix} \begin{bmatrix} V_3 \\ V_3 Y_0 \end{bmatrix} \tag{32}$$

$$\Longleftrightarrow V_3^{even} = \frac{1}{\sqrt{2}}\frac{V_s}{4}(1-j) = V_4^{even} \tag{33}$$

In the even mode, the voltage at port **3** and port **4** are identical (see FIG. **37**). In the odd mode, the transfer matrices are:

$$\begin{bmatrix} V_s/4 \\ I_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ jY_0 & 1 \end{bmatrix} \begin{bmatrix} 0 & j\frac{Z_0}{\sqrt{2}} \\ jY_0\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ jY_0 & 1 \end{bmatrix} \begin{bmatrix} V_3 \\ V_3 Y_0 \end{bmatrix} \tag{34}$$

$$\begin{bmatrix} -V_s/4 \\ I_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ jY_0 & 1 \end{bmatrix} \begin{bmatrix} 0 & j\frac{Z_0}{\sqrt{2}} \\ jY_0\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ jY_0 & 1 \end{bmatrix} \begin{bmatrix} V_4 \\ V_4 Y_0 \end{bmatrix} \tag{35}$$

$$\Leftrightarrow V_3^{odd} = -\frac{1}{\sqrt{2}}\frac{V_s}{4}(1+j);\ V_4^{odd} = \frac{1}{\sqrt{2}}\frac{V_s}{4}(1+j) \tag{36}$$

where equation (34) is the transfer matrix from port **1** to port **3** and equation (35) is the transfer matrix from port **2** to port **4** (see FIG. **37**). The amplitude of the total incident voltage (odd and even) at port **1** is $V_s/2$. Hence from superposition of the modes, S_{31} and S_{41} are equal to −j/√2 and 1/√2 respectively. The scattering matrix for the MS/LHM branch-line coupler is:

$$\begin{bmatrix} V_1^- \\ V_2^- \\ V_3^- \\ V_4^- \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & -j & 1 \\ 0 & 0 & 1 & -j \\ -j & 1 & 0 & 0 \\ 1 & -j & 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^+ \\ V_2^+ \\ V_3^+ \\ V_4^+ \end{bmatrix} \tag{37}$$

From equation (37), it is seen that in the MS/LHM branch-line coupler, the power divides equally between the through port (port **3**) and coupled port (port **4**) and they are 90° and 0° out of phase respectively with reference to the input (port **1**).

To construct the type **1** branch-line coupler a substrate similar to that of the coupled-line couplers was used. In the type **1** branch-line coupler, the MS strips have an impedance of 50/√2Ω whereas LHM strips have an impedance of 50Ω (using a microwave-circuit simulator). The design frequency is chosen to be 1.9 GHz so that each side of the coupler is λ/4 long (from Matlab® simulations of the dispersion characteristics of the line). The MS line used is 4.018 mm wide and 20.603 mm long. The LHM-line strip comprises 3 unit cells of width 2.343 mm and length 5.2 mm each. The latter is loaded with 3.3 nH shunt inductors and 1.6 pF series capacitors. The end capacitors of each LHM line segment are 2.4 pF. The terminating capacitors are chosen to be almost twice the value of the interior ones to preserve the symmetry of the unit cells making up the LHM lines (see FIG. **31**(*b*)).

During simulation, port **1** of the branch-line coupler shown in FIG. **36**(*a*) is excited in a microwave-circuit simulation and all other ports are terminated with 50Ω loads. The reflected power (S_{11}), isolation (S_{21}), through power (S_{31}) and coupled power (S_{41}) are recorded and compared to theoretical estimates from even-odd mode analysis. To apply the previously presented even-odd mode analysis, Matlab® simulations are performed for both lines to determine their dispersion characteristic. Subsequently, the corresponding phase shifts are used in the even-odd mode analysis. These theoretical results are compared with a microwave-circuit simulation and are presented in FIGS. **38**(*a*) to **38**(*c*).

In this branch-line coupler, the input power splits in half between the through port (port**3**) and coupled port (port **4**) as expected (see equation (22)). This is demonstrated in FIGS. **38**(*a*) and **38**(*b*). FIGS. **38***a* and **38***b* also show that at the design frequency, the through and coupled ports are matched (see S_{11} in FIG. **38**(*a*)) and port **2** is isolated (see FIG. **38**(*b*)). In FIGS. **38**(*a*) and **38**(*b*) the response of a commensurate, conventional, branch-line coupler is also included for comparison purposes. As shown, there is no degradation of the useful bandwidth between the conventional and the type **1** branch-line coupler. However, from theory the through port (port **3**) undergoes a phase shift of −90° and the coupled port (port **4**) undergoes a phase shift of 0° with respect to the input. This is clearly demonstrated from the simulation results in FIG. **38**(*c*).

The schematic for the type **2** MS/LHM branch-line coupler is shown in FIG. **36**(*b*). The scattering parameters for this branch-line coupler can be determined in the same way followed for the type **1** case. Since in this branch-line coupler, the MS and LHM lines are switched, every occurrence of j in equations (32) to (37) is replaced by −j. Hence the scattering matrix for the type **2** MS/LHM branch-line coupler is:

$$\begin{bmatrix} V_1^- \\ V_2^- \\ V_3^- \\ V_4^- \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & j & 1 \\ 0 & 0 & 1 & j \\ j & 1 & 0 & 0 \\ 1 & j & 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^+ \\ V_2^+ \\ V_3^+ \\ V_4^+ \end{bmatrix} \tag{38}$$

From equation (38), it is seen that in the type **2** MS/LHM branch-line coupler, the power divides equally between the through port (port **3**) and coupled port (port **4**) and the parts are +90° and 0° out of phase respectively with reference to the input (port **1**).

For the type **2** branch-line coupler, the LHM strips are designed with 50/√2Ω impedance and the MS strips with 50Ω impedance. The design frequency is chosen at 1.7 GHz so that again all sides are λ/4 long. In this case, the MS lines used are 2.342 mm wide and 23.645 mm long. The LHM-line strip comprises 3 unit cells of width 4.017 mm and length 5 mm each. The latter is loaded with 2.7 nH shunt inductors and 2.4 pF series capacitors, with terminating 4 pF capacitors for the LHM lines.

The type **2** branch-line coupler splits the input power equally between the two forward ports (through and coupled) and the backward port is isolated. As will be appreciated, the magnitude response is very similar to that of the type **1** branch-line coupler. From theory (equation (23)) and verified from the simulation results in FIG. **38**(*d*), the through port (port **3**) undergoes a phase shift of 90° and the coupled port (port **4**) undergoes a phase shift of 0° with respect to the input.

A conventional branch-line coupler allows for equal power split between the output ports but with −90° and −180° phase shifts with respect to the input (see FIGS. **38**(*a*) to **38**(*c*)). As will be appreciated, the present branch-line couplers offer more options in the design of circuits involving branch-line couplers. than available with conventional branch-line couplers. Specifically either a positive or a negative phase quadrature at the through port can be achieved, associated with a phase compensation at the coupled port.

The present coupled-line coupler demonstrates superior performance in terms of coupling, isolation and return loss when compared to a conventional microstrip coupled-line coupler of the same length, line spacing and propagation constant. Moreover, it is possible to realize such MS/LHM coupled-line couplers to achieve an arbitrary coupling coefficient. Furthermore, MS/LHM branch-line couplers possess similar functionality to their conventional counterparts when considering the magnitude of the power transfer between the ports. However, new functionality is added by the ability to obtain phase compensation at the coupled port (0° phase shift with respect to the input) along with a choice of either a positive or a negative phase quadrature (±90° phase shift) at the through port. On the other hand, the corresponding bandwidth remains comparable to that of conventional, commensurate, microstrip branch-line couplers.

Backward Leaky-Wave Antenna Using Metamaterial

Charged particles traveling at speeds greater than the phase velocity of light in a medium emit coherent radiation better known as Cherenkov radiation. The angle of the radiated conical wave front is given by the velocity of the particle (V) with respect to the phase velocity of EM waves (v) within the medium in the following manner:

$$\mathrm{Cos}(\theta) = v/V = \frac{c/n_0}{V} \tag{39}$$

where:

n_0 is the refractive index of the surrounding medium;

c is the speed of light in a vacuum; and

θ is the angle between the particle velocity and the radiated EM wave front.

This expression suggests that in a medium with negative no, the angle θ becomes obtuse. This implies that radiation is directed backward rather than forward, as is the case in a RHM.

Similarly, if a periodic guiding structure supports Bloch currents (moving charge) or equivalently EM waves with phase velocities greater than the speed of light, the angle of the radiated wave front is derived from equation (39) by letting V be the phase velocity of the EM wave along the guiding interface. This is due to the phase matching condition along the interface of the guide and surrounding medium. If the guiding structure is a medium with an effective refractive index n_1, and the surrounding medium has a refractive index n_0, then the angle of the radiated wave front is given by the following expression:

$$\mathrm{Cos}(\theta) = \frac{c/n_0}{c/n_1} = \frac{n_1}{n_0} \tag{40}$$

The above equation indicates that if the refractive index of the guiding medium (n_1) is negative, the radiation emitted into a RHM will be directed backward. It is in fact the phase advancement (backward-wave propagation) that causes backward radiation into the surrounding medium.

In order to excite backward-wave radiation from a dual transmission line LHM into free space, the L, C parameters need to be chosen such that the effective refractive index of the LHM, (n_1), is negative and smaller in magnitude than one, as indicated by equation (40). Equivalently, the magnitude of the propagation constant along the guiding LHM needs to be designed to be smaller in magnitude than the propagation constant in free-space k_0.

A radiating LHM antenna at 15 GHz based on dual transmission lines of the type shown in FIG. **13** was created. The LHM antenna includes 16 unit cells, with each unit cell having a dimension of 4.268 mm, approximately a factor of 5 smaller than the free-space wavelength of 2 cm at 15 GHz. Thus, the antenna can be treated as an effective medium. Commercial method of moments software was used to design the layout of the antenna. The LHM was fabricated by a mask/photo-etching technique on a 20-mil thick Rogers™ RO3203 circuit board with a relative permittivity of 3. The interconnecting transmission line sections were implemented in coplanar waveguide (CPW) technology. The transmission lines include a planar center conductor with two adjacent ground planes on either side. In the coplanar waveguide configuration, both ground planes and the center conductor lie in the same plane, which allows the simple integration of shunt inductors and series capacitors. The gaps in the CPW center conductor serve as the series capacitors and the narrow lines connecting the center conductor to the coplanar ground planes act as the shunt inductors. It is in fact these capacitive gaps that radiate in the structure and cause a backward emerging transverse magnetic wavefront. On the contrary, the inductive lines are non-radiating due to the antiparallel currents flowing on each pair of inductive lines. This odd symmetry causes cancellation in the far field and leads to low cross-polarization levels.

Although there are various radiating structures that exhibit phase and group velocities of opposite sign, the proposed structure, unlike conventional structures operates in the long wavelength regime and demonstrates backward-wave radiation in its lowest passband of operation. The proposed structure supports a backward-wave fundamental spatial harmonic that radiates. Early examples of backward-wave radiating structures include the helix antenna, corrugated dielectric, or metallic surfaces, and periodic arrays of radiating elements fed by slow wave transmission line sections of large periodic spacing (d>λ/2). These structures, however, radiate in higher-order passbands (exhibit higher-order radiating spatial harmonics) and therefore, effective material constants such as a refractive index cannot in general be defined. On the other hand, log-periodic dipole arrays and related uniform dipole arrays with a transposed feed have shown to produce backward waves even for small longitudinal periodic spacings. Nevertheless, the dipoles are resonant so element dimensions still remain electrically large (~λ/2) and therefore effective material parameters, e and m cannot be defined.

Super-Lens Formed From Metamaterials and RHM

The present metamaterial is also suitable for use in a focusing device that overcomes diffraction limits, hereinafter referred to as a "super-lens". The super-lens includes two-field focusing devices of the type previously discussed and is formed by sandwiching the LHM between the two RHM at two parallel interfaces. The LHM is a planar slab having a grid of printed metallic strips over a ground plane, loaded with series capacitors C and shunt inductors L. The two RHMs are two unloaded printed grids which act as homogeneous and isotropic media with positive indices of refraction.

FIG. **39** illustrates the planar transmission line left-handed super-lens. The unit cell of the left-handed grid is shown in the top insert while the unit cell of the positive refractive index grid is shown in the bottom insert. The super-lens is fabricated on a grounded microwave substrate (Rogers™ RO3003) of thickness 60 mil (1.52 mm) and dielectric constant $\in_r$=3.00. The left-handed slab has a 5×19 grid of printed metallic strips (microstrip lines) loaded with series chip capacitors C and shunt chip inductors L to the ground. Each unit cell has dimensions 8.40 mm×8.40 mm and therefore is much smaller than the nominal operating wavelength of 15.59 cm at 1 GHz. Hence, the loaded periodic grid acts as an effective medium exhibiting a negative index of refraction. The left-handed planar slab is sandwiched between two commensurate unloaded printed grids that act as effective homogenous media with positive indices of refraction. The unloaded grid and ground plane behave like a dielectrically loaded parallel-plate waveguide, but in addition allow simple measurement of the guided fields through proximity coupling. The left-handed grid is designed so that it is impedance matched and exhibits a refractive index of −1 relative to the unloaded grids at the design frequency of 1.00 GHz.

The first unloaded grid is excited at a point which is imaged by the left-handed slab to the second unloaded grid, with a vertical monopole fed by a coaxial cable through the ground plane. The monopole attaches the center conductor of the coaxial cable to a point on the unloaded grid, while the outer conductor of the coaxial cable attaches to the ground plane. The vertical monopole lies along the center row (row **0**), located 2.5 unit cells away from the first interface of the left-handed slab. According to geometrical optics, the source and image should be symmetrically positioned with respect to the left-handed slab. This is due to the fact that the distance from the source to the first interface (2.5 unit cells) is half the left-handed slab thickness (5 cells). The set-up shown in FIG. **39** is a practical implementation for imaging a point source from one homogeneous dielectric to another using a left-handed slab.

The vertical electric field over the entire structure is measured using a detecting probe. The field is detected 0.8 mm above the entire surface of the structure, using a short vertical probe connected to a Hewlett-Packard Vector Network Analyzer model 8753D. Port **1** of the network analyzer is connected to the coaxial cable that feeds the exciting monopole. Port **2** is connected by a separate coaxial cable to the detecting probe that is scanned above the surface of the structure using a computer-controlled stepper motor. The measured transmission coefficient is proportional to the voltage of the grid nodes with respect to the ground plane. The best focusing results were observed at 1.057 GHz, a frequency slightly higher than the design frequency of 1.00 GHz. This is primarily due to the variation of chip inductors and capacitors from their nominal values, as well as fabrication tolerances in printing the grid lines. The measured vertical electric field above each unit cell for the entire structure is shown in FIG. **40**, at a frequency of 1.057 GHz. The plot is normalized with respect to the source amplitude. The source is located at (column, row)=**(0,0)** and the image at **(10,0)**, whereas the first and second interfaces are located between columns **2** & **3** and **7** & **8**, respectively. As shown, the enhancement of evanescent waves is quite evident at the second interface of the left-handed slab near the center row (row **0**) and the results agree well with microwave-circuit simulations. FIG. **41** explicitly shows the measured electric field along row **0** to emphasize the growing evanescent fields within the left-handed lens. The vertical dashed lines identify the source at column **0** and the image at column **10** while the vertical solid lines identify the interfaces of the left-handed slab.

Geometrical optics establish that the source and its image should be separated by twice the thickness of the slab (2×5=10 unit cells). Thus, the measured vertical electric field along the image column (column **10**) is shown in FIG. **42**. Plotted in the same Figure are the measured transverse patterns at the source column (column **0**), as well as the theoretical diffraction-limited pattern. All patterns in FIG. **42** are normalized to unity for comparison purposes. Nevertheless, the measured source and image peaks lie within 7% of each other. In addition, it is important to explain how the theoretical diffraction-limited pattern was computed. Due to the close proximity (0.54 effective wavelengths) of the source and its image, evanescent waves that reach the image should be accounted for. For this reason, the diffraction-limited pattern assumes that the propagating components are focused whereas the evanescent components are not neglected, but rather assumed to exponentially decay from the source to the image, with attenuation factors corresponding to a refractive index of n=+1. Expressions for the voltage (V) at the image plane (column **10**) are given by the integral in equation 41 below:

$$V(x) = C\int_{-\infty}^{\infty} \frac{e^{ik_{z1}D} e^{ik_{z2}D} e^{ik_x x}}{k_{z1}} \, dk_x \tag{41}$$

$k_{z1}=-k_{z2}=\sqrt{k_o^2-k_x^2}$ *for* $k_{.x.}<k_o$ (propagating waves)

$k_{z1}=k_{z2}=i\sqrt{k_x^2-k_o^2}$ *for* $k_{.x.}>k_o$ (evanescent waves)

where:

C is a constant that normalizes V(0)=1;

D is the width of the lens; and

k_o is the intrinsic wavenumber in all grid media.

The half-power beamwidth of the diffraction-limited image shown in FIG. **42** is 129 degrees. It was found that the diffraction-limited image including attenuating evanescent waves and the image neglecting them altogether are almost identical.

With this clarification in mind, a comparison of the images shown in FIG. **42** can be made. Clearly, the measured image pattern is narrower than the theoretical diffraction-limited one. The measured half-power beamwidth is 75 degrees. This is significantly narrower than the 129 degrees for the diffraction-limited image. This establishes that it is possible to overcome the diffraction limit when imaging from one homogeneous dielectric to another through a left-handed isotropic lens. This narrowing of the beamwidth beyond the diffraction limit can be ascribed to the enhancement of the evanescent waves evident in FIGS. **39** and **40**. Nevertheless, the image is still imperfect since the source beamwidth is narrower than that of the image. This is not surprising considering that slight material losses and mismatches at the lens interfaces lead to departures from the n=−1 condition and thus, degrade the ability of a left-handed lens to achieve perfect imaging.

Resonance Cone Antenna

Resonance cone metamaterial may also be used in an antenna, by shunting the metamaterial with inductors, at the transmission line intersections. FIG. **43** shows a resonance cone antenna **4300** constructed using metamaterial. The antenna includes a 2×3 array of macrocells. Each macrocell includes a plurality of microcells of orthogonal pairs of transmission line segments joined in a substantially rectangular fashion. The segments are loaded periodically with lumped capacitors in one direction and lumped inductors in a perpendicular direction. The microcells of the macrocells have a shared orientation that is different from the orientation of microcells in adjacent macrocells by 90 degrees. Macrocells are joined by transition regions similar to those described previously with reference to the focusing device. The transmission line segments of the transition regions include both lumped capacitors and inductors arranged in a manner consistent with FIG. **24**. The device input, or feed point, is located at the transition region adjacent the center of the macrocell array.

This arrangement of metamaterials, and input location allows input waves of compatible wavelength, to form a resonance cone patern of two squares that are multi-refracted within the macrocell array. This gives radiation patterns that can be close to the ground plane or elevated, but do not suffer from unwanted edge interactions as shown in FIGS. **44** and **45**. Moreover, as the frequency departs from the design value, the resonance cone squares turn into expanding or contracting square spirals that become weak before reaching the grid edges due to radiation or ohmic losses. This phenomenon is inherently non-resonant which leads to broadened operating bandwidths as compares to patch antennas over ground.

As will be appreciated it has been shown that negative refraction and focusing of electromagnetic waves can be achieved in metamaterials that support backward waves without employing resonances or directly synthesizing the permittivity and permeability. Schemes for fabricating such media by appropriately loading a host transmission line medium have also been described. The resulting planar topology permits LHM structures to be readily integrated with conventional planar microwave circuits and devices and used in a variety of applications to guide and control electromagnetic radiation.

Although preferred embodiments of the present invention have been described, those of skill in the art will appreciate that variations and modifications may be made without departing from the spirit and scope thereof as defined by the appended claims.

What is claimed is:

1. A planar metamaterial comprising:

two substantially orthogonal, coplanar sets of transmission lines, said transmission lines being spaced with a periodicity, loaded with capacitors with said periodicity, and shunted with inductors with said periodicity such that for an electromagnetic wave having a wavelength greater than said periodicity and traveling along the plane of said transmission lines, an effective permeability and permittivity of said metamaterial are simultaneously negative.

2. A planar metamaterial according to claim **1** wherein the wavelength of said electromagnetic wave is greater than said periodicity by an order of magnitude.

3. A linear metamaterial comprising:

a transmission line, having a linear dimension, and being loaded with capacitors, and shunted with an inductor such that for an electromagnetic wave having a wavelength greater than said linear dimension and traveling along the axis of said transmission line, an effective permeability and permittivity of said metamaterial are simultaneously negative.

4. A planar metamaterial according to claim **3** wherein the wavelength of said electromagnetic wave is greater than said periodicity by an order of magnitude.

5. A metamaterial according to claim **4** wherein the values of said capacitors and inductor are proportioned to provide a continuous passband for said electromagnetic wave.

6. A metamaterial according to claim **4** wherein said transmission line includes two substantially identical capacitors on either side of said inductor.

7. A metamaterial according to claim **4** further comprising a plurality of substantially identical transmission lines arranged in series.

8. A metamaterial according to claim **7** wherein each transmission line includes two substantially identical capacitors on either side of said inductor.

9. A metamaterial according to claim **8** wherein the values of said capacitors and inductor are proportioned to provide a continuous passband for said electromagnetic wave.

10. A planar resonance cone metamaterial comprising:

a first set of transmission lines, spaced with a periodicity, and loaded with capacitors with said periodicity;

a second set of transmission lines, substantially orthogonal and coplanar with said first set of transmission lines, said second set of transmission lines being spaced with said periodicity, and loaded with inductors with said periodicity, said first and second sets of transmission lines exhibiting characteristics such that for an electromagnetic wave having a wavelength greater than said periodicity by an order of magnitude, and traveling along the linear axis of said first set of transmission lines, an effective permittivity of said metamaterial is positive such that for an electromagnetic wave having a wavelength greater than said periodicity by an order of magnitude and traveling along the linear axis of said second set of transmission lines, the effective permittivity of said metamaterial is negative.

11. A metamaterial according to claim **10** further comprising an input on the perimeter of said metamaterial coincident with an intersection of one transmission line of said first set and one transmission line of said second set and a plurality of outputs on the perimeter of said metamaterial coincident with intersections of transmission lines of said first and second sets.

12. A planar resonance cone metamaterial comprising:

a first set of transmission lines, spaced with a first periodicity, and loaded with capacitors with a second periodicity;

a second set of transmission lines, substantially orthogonal and coplanar with said first set of transmission lines, said second set of transmission lines being spaced with said second periodicity, and being loaded with inductors with said first periodicity whereby for an electromagnetic wave having a wavelength greater than said first and second periodicities by an order of magnitude and traveling along the longitudinal axis of said first set of transmission lines, an effective permittivity of said metamaterial is positive and for an electromagnetic wave having a wavelength greater than said first and second priodocities and traveling along the longitudinal axis of said second set of transmission lines, the effective permittivity of said metamaterial is negative.

13. A near field focusing device comprising:

a first set of transmission lines, said first set of transmission lines being spaced with a first periodicity and loaded with capacitors with a second periodicity;

a second set of transmission lines, substantially orthogonal to, and coplanar with said first set of transmission lines, said second set of transmission lines being spaced with said second periodicity, loaded with capacitors with said first periodicity, and shunted with inductors, said first set of transmission lines intersecting said second set of transmission lines such that for an electromagnetic wave having a wavelength greater than said first or second periodicity by an order of magnitude, and traveling along the plane of said transmission lines, an effective permeability and permittivity of said metamaterial are simultaneously negative; and

a planar waveguide having a flat extent, coupled to said transmission lines, such that said flat extent is parallel to one set of transmission lines.

14. A phase-shifting device comprising:

a transmission line, having a linear dimension and characteristic impedance;

capacitors loaded on said transmission line; and

an inductor shunting said transmission line, said transmission line exhibiting characteristics such that for an electromagnetic wave having a wavelength greater than said linear dimension by an order of magnitude and traveling along the axis of said transmission line, an

effective permeability and permittivity of transmission are simultaneously negative and such that said transmission line provides a continuous passband for said electromagnetic wave along said transmission line.

15. A phase-shifting device according to claim **14** further comprising a plurality of transmission lines coupled in series.

16. A phase-shifting-device according to claim **15** wherein each transmission line includes a pair of identical capacitors on either side of said inductor.

17. A near field focussing device comprising:

a parallel-plate waveguide; and

a negative refractive index metamaterial in line with said waveguide and forming an interface therewith.

18. A focussing device according to claim **17** wherein said waveguide serves as a right-handed medium, electromagnetic radiation passing through said interface from said waveguide being focussed in said metamaterial.

* * * * *

US008253639B2

(12) **United States Patent**
Cohen

(10) **Patent No.: US 8,253,639 B2**
(45) **Date of Patent: Aug. 28, 2012**

(54) **WIDEBAND ELECTROMAGNETIC CLOAKING SYSTEMS**

(76) Inventor: **Nathan Cohen**, Belmont, MA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 498 days.

(21) Appl. No.: **12/547,104**

(22) Filed: **Aug. 25, 2009**

(65) **Prior Publication Data**

US 2010/0156556 A1 Jun. 24, 2010

Related U.S. Application Data

(60) Provisional application No. 61/189,966, filed on Aug. 25, 2008.

(51) **Int. Cl.**
H01Q 15/02 (2006.01)
H01Q 19/06 (2006.01)

(52) **U.S. Cl.** **343/753**; 343/909

(58) **Field of Classification Search** 343/753, 343/909, 910; 333/135
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,471,224 A *	11/1995	Barkeshli	....	343/909
5,714,961 A *	2/1998	Kot et al.	....	343/769
6,473,048 B1 *	10/2002	Diaz	....	343/753

* cited by examiner

Primary Examiner — Michael C Wimer

(74) *Attorney, Agent, or Firm* — McDermott Will & Emery LLP

(57) **ABSTRACT**

Arrangement of resonators in an aperiodic configurations are described, which can be used for electromagnetic cloaking of objects. The overall assembly of resonators, as structures, do not all repeat periodically and at least some of the resonators are spaced such that their phase centers are separated by more than a wavelength. The arrangements can include resonators of several different sizes and/or geometries arranged so that each size or geometry corresponds to a moderate or high "Q" response that resonates within a specific frequency range, and that arrangement within that specific grouping of akin elements is periodic in the overall structure. The relative spacing and arrangement of groupings can be defined by self similarity and origin symmetry.

14 Claims, 4 Drawing Sheets

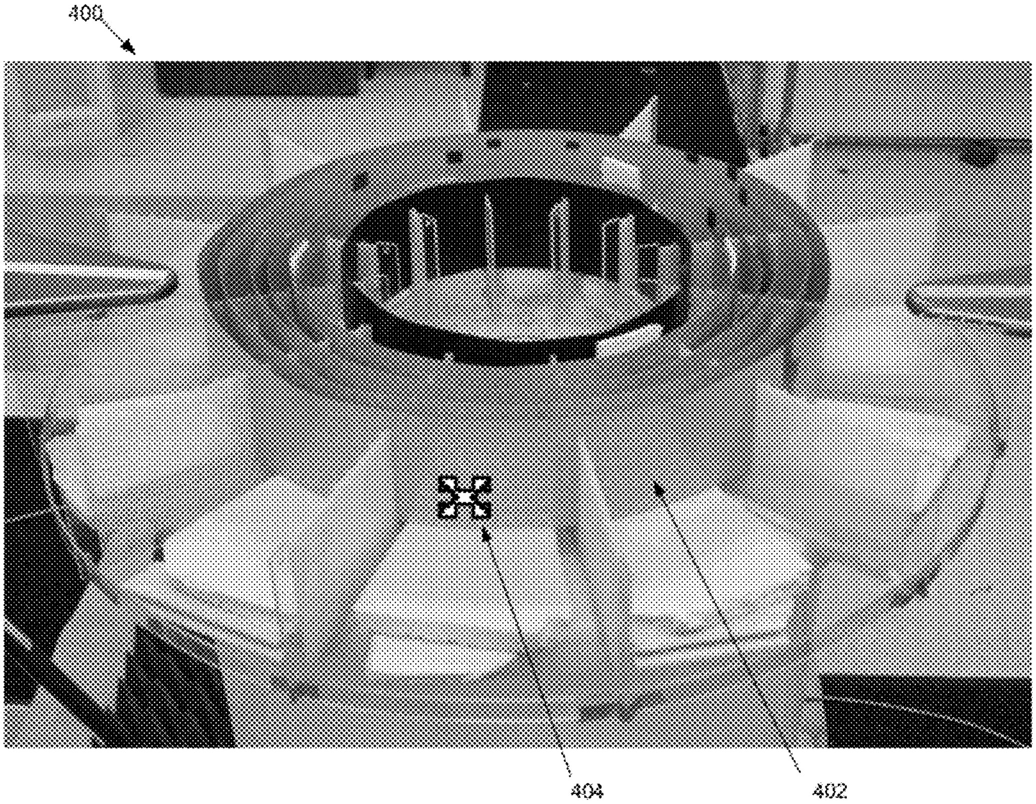

FIG. 1

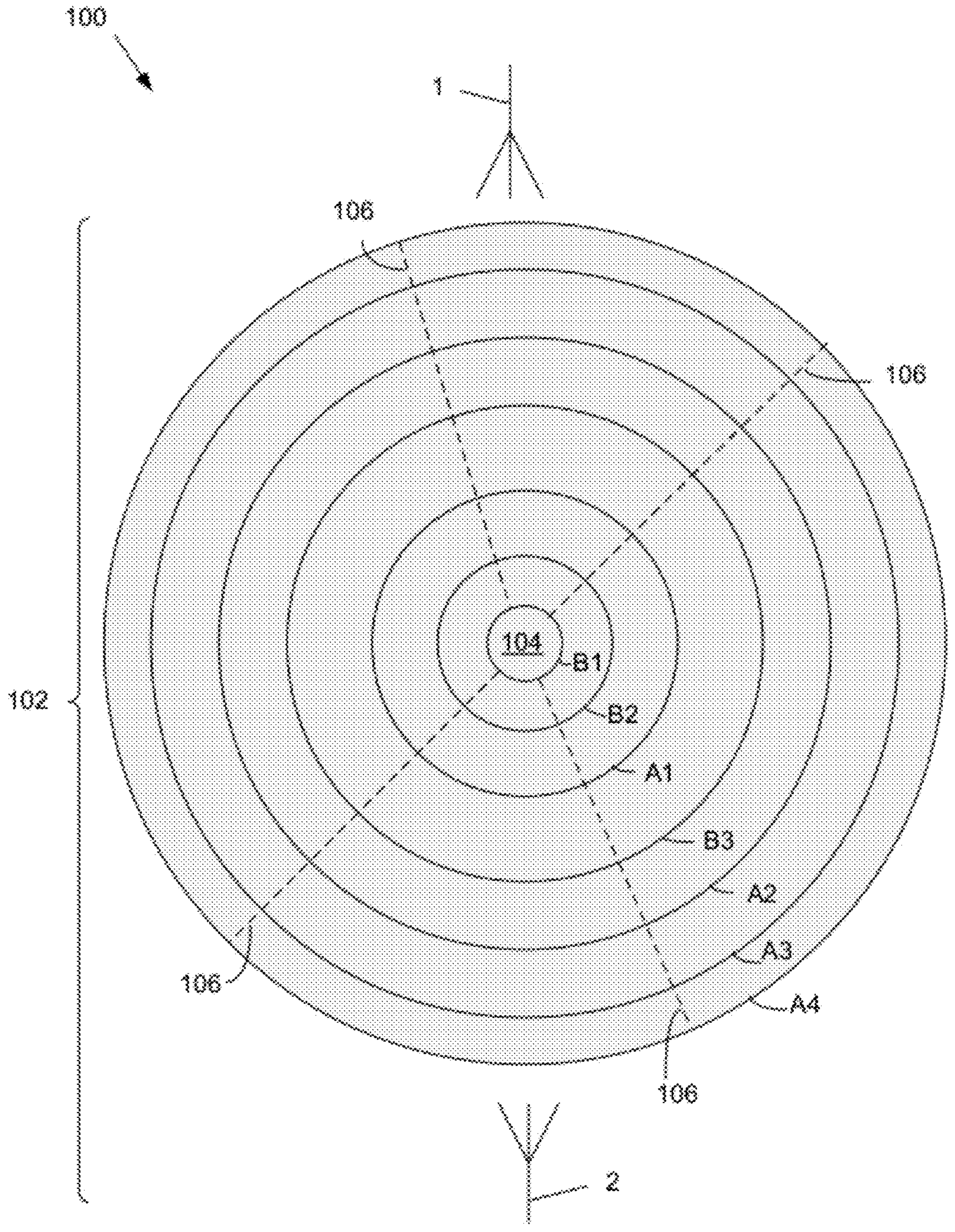
100
1
106
106
102
104
B1
B2
A1
B3
A2
A3
A4
106
106
2

FIG. 2

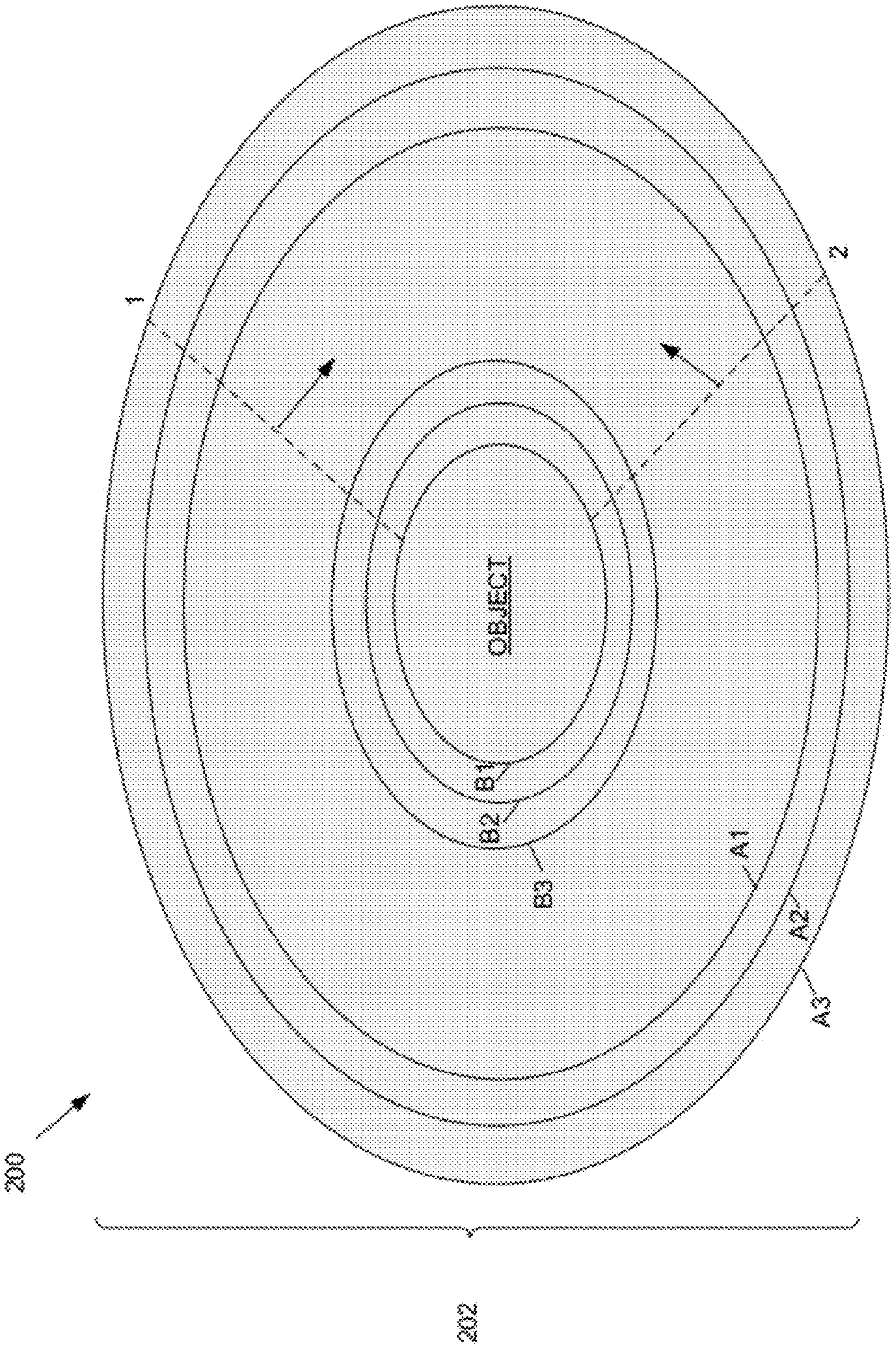
OBJECT
1
2
B1
B2
B3
A1
A2
A3
200
202

FIG. 3

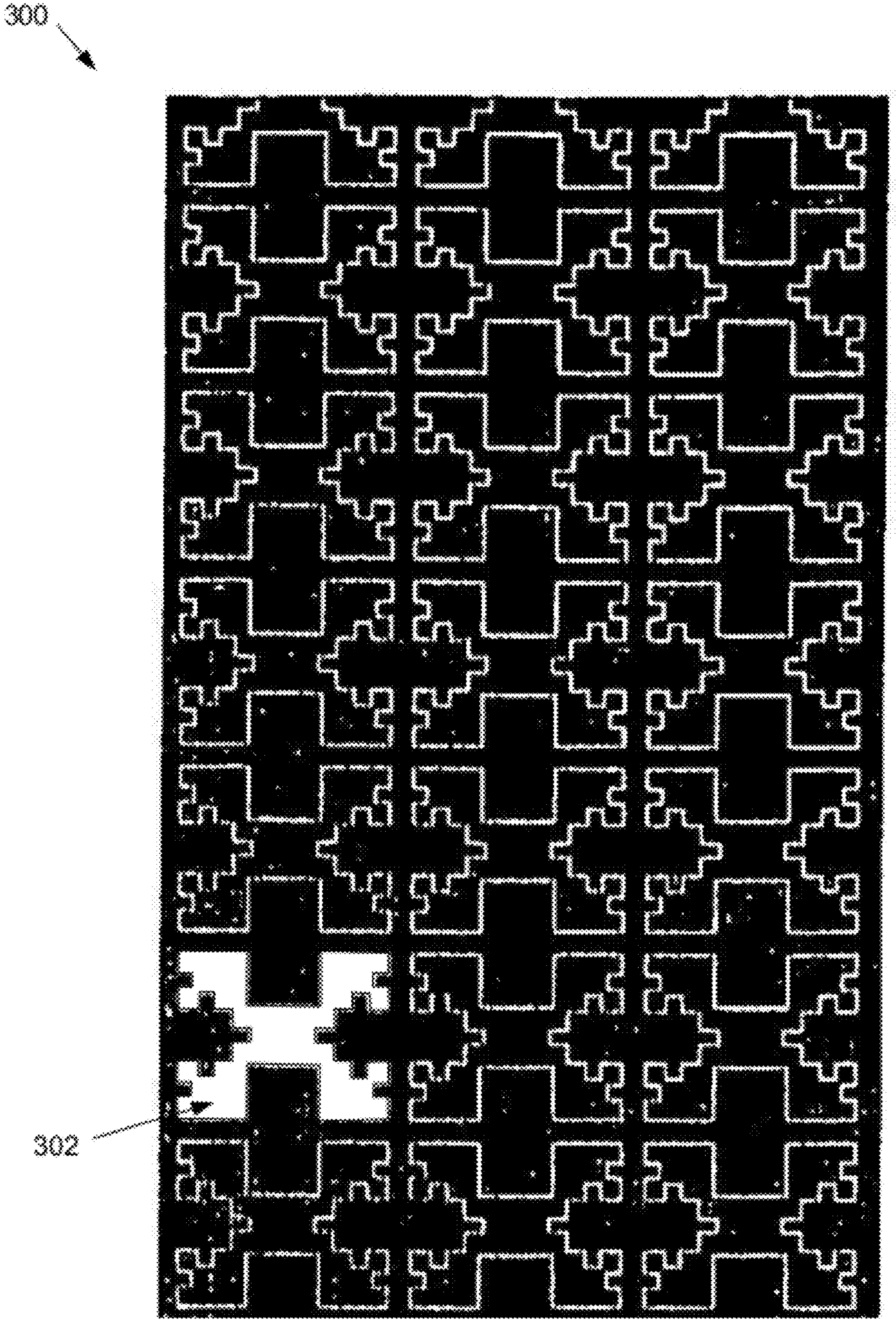

FIG. 4

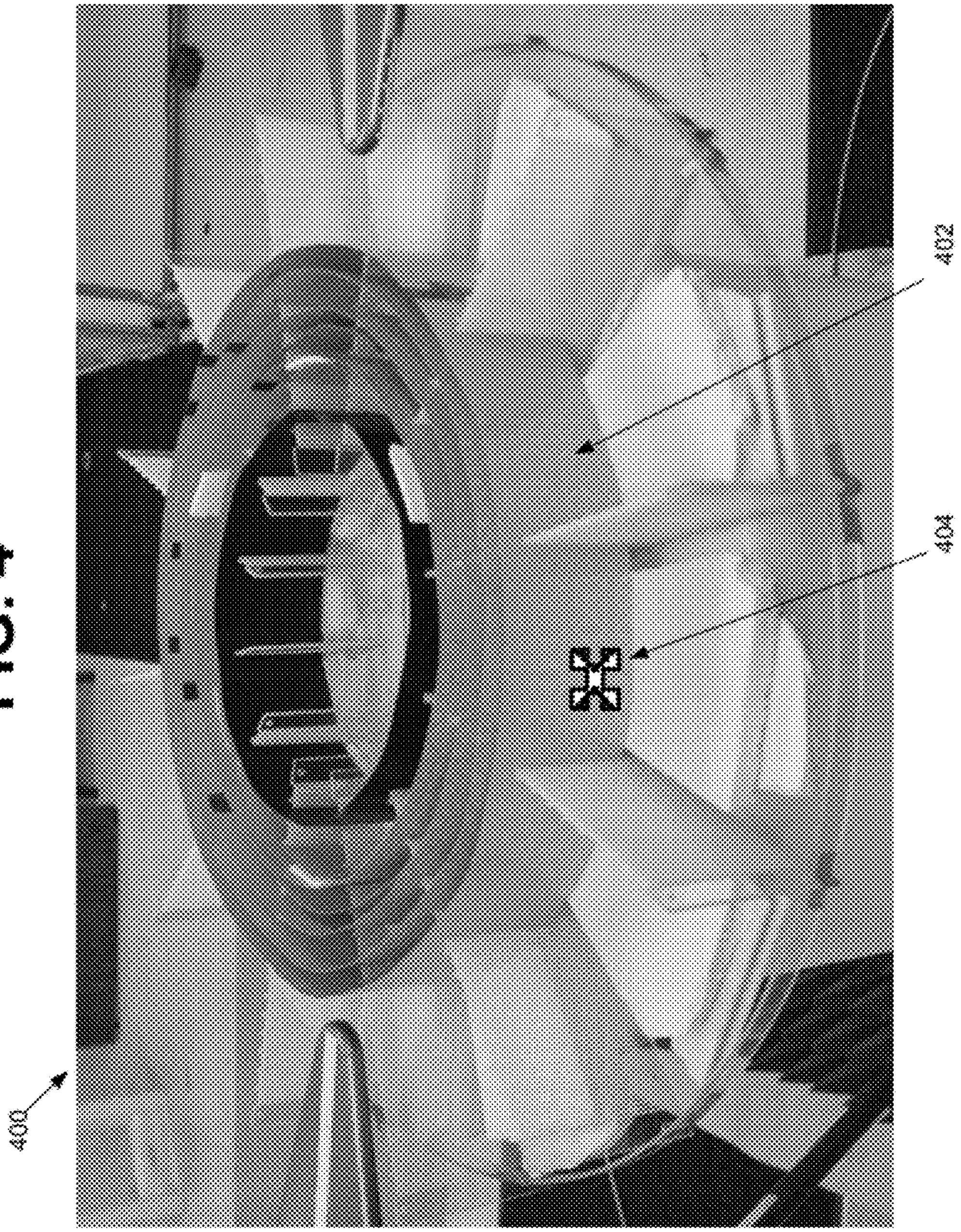

WIDEBAND ELECTROMAGNETIC CLOAKING SYSTEMS

RELATED APPLICATION

This application claims priority to U.S. Provisional Patent Application No. 61/189,966, filed 25 Aug. 2008 and entitled "Method and Apparatus for Wideband Electromagnetic Cloaking, Negative Refractive Index Lensing and Metamaterial Applications," the entire contents of which are incorporated herein by reference.

BACKGROUND

Much time and effort has been devoted to the quest for so-called invisibility machines. Beyond science fiction, however, there has been little if any real progress toward this goal.

Materials with negative permittivity and permeability leading to negative index of refraction were theorized by Russian noted physicist Victor Veselago in his seminal paper in *Soviet Physics* USPEKHI, 10, 509 (1968). Since that time, metamaterials have been developed that produce negative index of refraction, subject to various constraints. Such materials are artificially engineered micro/nanostructures that, at given frequencies, show negative permeability and permittivity. Metamaterials have been shown to produce narrow band, e.g., typically less than 5%, response such as bent-back lensing. Such metamaterials produce such a negative-index effect by utilizing a closely-spaced periodic lattice of resonators, such as split-ring resonators, that all resonate. Previous metamaterials provide a negative index of refraction when a sub-wavelength spacing is used for the resonators.

In the microwave regime, certain techniques have been developed to utilize radiation-absorbing materials or coatings to reduce the radar cross section of airborne missiles and vehicles. While such absorbing materials can provide an effective reduction in radar cross section, these results are largely limited to small ranges of electromagnetic radiation.

SUMMARY

Embodiments of the present disclosure can provide techniques, including systems and/or methods, for cloaking objects at certain wavelengths/frequencies or over certain wavelength/frequency ranges (bands). Such techniques can provide an effective electromagnetic lens and/or lensing effect for certain wavelengths/frequencies or over certain wavelength/frequency ranges (bands).

The effects produced by such techniques can include cloaking or so-called invisibility of the object(s) at the noted wavelengths or bands. Representative frequencies of operation can include, but are not limited to, those over a range of 500 MHz to 1.3 GHz, though others may of course be realized. Operation at other frequencies, including for example those of visible light, infrared, ultraviolet, and as well as microwave EM radiation, e.g., K, Ka, X-bands, etc. may be realized, e.g., by appropriate scaling of dimensions and selection of shape of the resonator elements.

Exemplary embodiments of the present disclosure can include a novel arrangement of resonators in an aperiodic configuration or lattice. The overall assembly of resonators, as structures, do not all repeat periodically and at least some of the resonators are spaced such that their phase centers are separated by more than a wavelength. The arrangements can include resonators of several different sizes and/or geometries arranged so that each size or geometry ("grouping") corresponds to a moderate or high "Q" (that is moderate or low bandwidth) response that resonates within a specific frequency range, and that arrangement within that specific grouping of akin elements is periodic in the overall structure—even though the structure as a whole is not an entirely periodic arrangement of resonators. The relative spacing and arrangement of groupings (at least one for each specific frequency range) can be defined by self similarity and origin symmetry, where the "origin" arises at the center of a structure (or part of the structure) individually designed to have the wideband metamaterial property.

For exemplary embodiments, fractal resonators can be used for the resonators in such structures because of their control of passbands, and smaller sizes compared to non-fractal based resonators. Their benefit arises from a size standpoint because they can be used to shrink the resonator (s), while control of passbands can reduce or eliminates issues of harmonic passbands that would resonate at frequencies not desired.

It should be understood that other embodiments of wideband electromagnetic resonator or cloaking systems and methods according to the present disclosure will become readily apparent to those skilled in the art from the following detailed description, wherein exemplary embodiments are shown and described by way of illustration. The systems and methods of the present disclosure are capable of other and different embodiments, and details of such are capable of modification in various other respects. Accordingly, the drawings and detailed description are to be regarded as illustrative in nature and not as restrictive.

BRIEF DESCRIPTION OF THE DRAWINGS

Aspects of the disclosure may be more fully understood from the following description when read together with the accompanying drawings, which are to be regarded as illustrative in nature, and not as limiting. The drawings are not necessarily to scale, emphasis instead being placed on the principles of the disclosure. In the drawings:

FIG. **1** depicts a diagrammatic plan view of a resonator cloaking system utilizing a number of cylindrical shells, in accordance with exemplary embodiments of the present disclosure;

FIG. **2** depicts a diagrammatic plan view of a resonator cloaking system utilizing a number of shells having an elliptical cross-section, in accordance with an alternate embodiment of the present disclosure;

FIG. **3** depicts an exemplary embodiment of a portion of shell that includes repeated conductive traces that are configured in a fractal-like shape; and

FIG. **4** depicts a perspective view (photograph) of an exemplary embodiment of the present disclosure.

While certain embodiments depicted in the drawings, one skilled in the art will appreciate that the embodiments depicted are illustrative and that variations of those shown, as well as other embodiments described herein, may be envisioned and practiced within the scope of the present disclosure.

DETAILED DESCRIPTION

The present disclosure is directed to novel arrangements of resonators useful for obscuring or hiding objects at given bands of electromagnetic radiation. Embodiments of the present disclosure can provide techniques, including systems and/or methods, for hiding or obscuring objects at certain wavelengths/frequencies or over certain wavelength/frequency ranges or bands. Such techniques can provide an

effective electromagnetic lens and/or lensing effect for certain wavelengths/frequencies or over certain wavelength/frequency ranges or bands. The effects produced by such techniques can include cloaking or so-called invisibility of the object(s) at the noted wavelengths or bands.

Representative frequencies of operation can include, but are not limited to, those over a range of about 500 MHz to about 1.3 GHz, though others may of course be realized. Operation at other frequencies, including for example those of visible light, infrared, ultraviolet, and as well as microwave EM radiation, e.g., K, Ka, X-bands, etc. may be realized, e.g., by appropriate scaling of dimensions and selection of shape of the resonator elements.

Embodiments of the present disclosure include arrangement of resonators or resonant structures in aperiodic configurations or lattices. The overall assembly of resonator structures can include nested or concentric shells, that each include repeated patterns of resonant structures. The resonant structures can be configured as a close-packed arrangement of electrically conductive material. The resonant structures can be located on the surface of a circuit board.

The overall assemblies, as structures, do not all repeat periodically and at least some of the resonators are spaced such that their phase centers are separated by more than a wavelength. The arrangements can include resonators of several different sizes and/or geometries arranged so that each size or geometry ("grouping") corresponds to a moderate or high quality-factor "Q" response (that is, one allowing for a moderate or low bandwidth) that resonates within a specific frequency range, and that arrangement within that specific grouping of like elements is periodic in the overall structure—even though the structure as a whole is not an entirely periodic arrangement of resonators. The relative spacing and arrangement of groupings (at least one for each specific frequency range) can be defined by self similarity and origin symmetry, where the "origin" arises at the center of a structure (or part of the structure) individually designed to have the wideband metamaterial property.

For exemplary embodiments, fractal resonators can be used for the resonators because of their control of passbands, and smaller sizes. A main benefit of such resonators arises from a size standpoint because they can be used to shrink the resonator(s), while control of passbands can reduce/mitigate or eliminate issues of harmonic passbands that would resonate at frequencies not desired.

Exemplary embodiments of a resonator system for use at microwave (or nearby) frequencies can be built from belts of circuit boards festooned with resonators. These belts can function to slip the microwaves around an object located within the belts, so the object is effectively invisible and "see thru" at the microwave frequencies. Belts, or shells, having similar closed-packed arrangements for operation at a first passband can be positioned within a wavelength of one another, e.g., $\frac{1}{10}\lambda$, $\frac{1}{8}\lambda$, $\frac{1}{4}\lambda$, $\frac{1}{2}\lambda$, etc.

An observer can observe an original image or signal, without it being blocked by the cloaked object. Using no power, the fractal cloak can replicates the original signal (that is, the signal before blocking) with great fidelity. Exemplary embodiments can function over a bandwidth from about 500 MHz to approximately 1500 MHz (1.5 GHz), providing 3:1 bandwidth; operation within or near such can frequencies can provide other bandwidths as well, such as 1:1 up to 2:1 and up to about 3:1.

FIG. **1** depicts a diagrammatic plan view of a cloaking system **100** and RF testing set up in accordance with exemplary embodiments of the present disclosure. As shown in FIG. **1**, a number of concentric shells (or bands) **102** are placed on a platform (parallel to the plane of the drawing). The shells include a flexible substrate (e.g., polyimide with or without composite reinforcement) with conductive traces (e.g., copper, silver, etc.) in fractal shapes or outlines. The shells **102** surround an object to be cloaked (shown as **104** in FIG. **1**). A transmitting antenna **1** and a receiving antenna **2** are configured at different sides of the system **100**, for verifying efficacy of the cloaking system **100** and recording results. The shells **102** can be held in place by radial supports **106** (while only four are shown, **12** were used in the exemplary embodiment indicated).

The shells indicated in FIG. **1** are of two types, one set (A**1**-A**4**) configured for optimal operation over a first wavelength/frequency range, and another set (B**1**-B**3**) configured for optimal operation over a second wavelength/frequency range. (The numbering of the shells is of course arbitrary and can be reordered, e.g., reversed.)

For an exemplary embodiment of system **100**, the outer set of shells (A**1**-A**4**, with A**1** being the innermost and A**4** the outmost) had a height of about 3 to 4 inches (e.g., 3.5 inches) and the inner set of shells had a height of about 1 inch less (e.g., about 2.5 to 3 inches). The spacing between the shells with a larger fractal shape (A**1**-A**4**) was about 2.4 cm while the spacing between shells of smaller fractal generator shapes (B**1**-B**3**) was about 2.15 cm (along a radial direction). In a preferred embodiment, shell A**4** was placed between shell B**2** and B**3** as shown. The resonators formed on each shell by the fractal shapes can be configured so as to be closely coupled (e.g., by capacitive coupling) and can serve to propagate a plasmonic wave.

It will be appreciated that while, two types of shells and a given number of shells per set are indicated in FIG. **1**, the number of shell types and number of shells for each set can be selected as desired, and may be optimized for different applications, e.g., wavelength/frequency bands.

FIG. **2** depicts a diagrammatic plan view of a cloaking system (or electrical resonator system) according to an alternate embodiment in which the individual shells have an elliptical cross section. As shown in FIG. **2**, a system **200** for cloaking can include a number of concentric shells (or bands) **202**. These shells can be held in place with respect to one another by suitable fixing means, e.g., they can be placed on a platform (parallel to the plane of the drawing) and/or held with a frame. The shells **202** can include a flexible substrate (e.g., polyimide with or without composite reinforcement) with a close-packed arrangement of electrically conductive material formed on the first surface. As stated previously for FIG. **1**, the closed-packed arrangement can include a number of self-similar electrical resonator shapes. The resonator shapes can be made from conductive traces (e.g., copper, silver, gold, silver-based ink, etc.) having a desired shape, e.g., fractal shape, split-ring shape, and the like. The shells **202** can surround an object to be cloaked, as indicated in FIG. **2**.

As indicated in FIG. **2** (by dashed lines **1** and **2** and arrows), the various shells themselves do not have to form closed surfaces. Rather, one or more shells can form open surfaces. This can allow for preferential cloaking of the object in one direction or over a given angle (solid angle). Moreover, while dashed lines **1** and **2** are shown intersecting shells B**1**-B**3** and A**1**-A**3** of system **200**, one or more shells of each group of shells (B**1**-B**3** and A**1**-A**3**) can be closed while others are open.

With further regard to FIGS. **1**-**2**, it should be appreciated that the cross-sections shown for each shell can represent closed geometric shapes, e.g., spherical and ellipsoidal shells.

As indicated previously, each shell of a cloaking system can include multiple resonators. The resonators can be repeated patterns of conductive traces. These conductive traces can be closed geometric shapes, e.g., rings, loops, closed fractals, etc. The resonator(s) can being self similar to at least second iteration. The resonators can include split-ring shapes, for some embodiments. The resonant structures are not required to be closed shapes, however, and open shapes can be used for such.

In exemplary embodiments, the closed loops can be configured as a fractals or fractal-based shapes, e.g., as depicted by **302** in FIG. **3** for an exemplary embodiment of a shell **300**, or **402** in FIG. **4**. The dimensions and type of fractal shape can be the same for each shell type but can vary between shell types. This variation (e.g., scaling of the same fractal shape) can afford increased bandwidth for the cloaking characteristics of the system (e.g., system **100** of FIG. **1**) This can lead to periodicity of the fractal shapes of common shell types but aperiodicity between the fractal shapes of different shell types.

Examples of suitable fractal shapes (for use for shells and/or a scatting object) can include, but are not limited to, fractal shapes described in one or more of the following patents, owned by the assignee of the present disclosure, the entire contents of all of which are incorporated herein by reference: U.S. Pat. Nos. 6,452,553; 6,104,349; 6,140,975; 7,145,513; 7,256,751; 6,127,977; 6,476,766; 7,019,695; 7,215,290; 6,445,352; 7,126,537; 7,190,318; 6,985,122; 7,345,642; and, U.S. Pat. No. 7,456,799.

Other suitable fractal shape for the resonant structures can include any of the following: a Koch fractal, a Minkowski fractal, a Cantor fractal, a torn square fractal, a Mandelbrot, a Caley tree fractal, a monkey's swing fractal, a Sierpinski gasket, and a Julia fractal, a contour set fractal, a Sierpinski triangle fractal, a Menger sponge fractal, a dragon curve fractal, a space-filling curve fractal, a Koch curve fractal, an Iypanov fractal, and a Kleinian group fractal.

FIG. **3** depicts an exemplary embodiment of a shell **300** (only a portion is shown) that includes repeated conductive traces that are configured in a fractal shape **302** (the individual closed traces). For the exemplary embodiment shown, each resonator shape **302** is about 1 cm on a side. Such resonator could, e.g., be used for the fractal shapes of shells B**1**-B**3** of FIG. **1**, in which case similar fractal shapes of larger size (e.g., about 1.5 cm on a side) could be used for shells A**1**-A**4**. The conductive trace is preferably made of copper. While exemplary fractal shapes are shown in FIG. **3**, the present disclosure is not limited to such and any other suitable fractal shapes (including generator motifs) may be used in accordance with the present disclosure.

It will be appreciated that the resonant structures of the shells may be formed or made by any suitable techniques and with any suitable materials. For example, semiconductors with desired doping levels and dopants may be used as conductive materials. Suitable metals or metal containing compounds may be used. Suitable techniques may be used to place conductors on/in a shell, including, but no limited to, printing techniques, photolithography techniques, etching techniques, and the like.

It will also be appreciated that the shells may be made of any suitable material(s). Printed circuit board materials may be used. Flexible circuit board materials are preferred. Other material may, however, be used for the shells and the shells themselves can be made of non-continuous elements, e.g., a frame or framework. For example, various plastics may be used.

FIG. **4** depicts a perspective view (photograph) of an exemplary embodiment of a cloak system **400** according to the present disclosure. As shown, the system includes a number of resonator shells **402** each having a close-packed arrangement of electrically conductive material (self-similar resonators **404**) formed on one surface. Two different shell configurations are shown, four larger shells and two smaller shells. The smaller shells included close-packed arrangements of resonator structures in which each resonator shape (as shown by **302** in FIG. **3**) was about 1 cm on a side. Similar fractal shapes of larger size (e.g., about 1.5 cm on a side) were used for the larger shells.

In FIG. **4**, a transmitting (source) antenna and a receiving antenna are shown as triangular shapes on the left and right, respectively (though functionally of each could of course be interchanged for the other). Twelve radially arrayed spacers are shown in FIG. **4**. The system **400** is shown supported on a Nalgene tank and Delrin platform and Delrin supports (radial supports) RF absorbers were placed in the immediate vicinity of the set up; further RF tripods (e.g., available from ETS) were used; all such materials were substantially transparent at the RF frequencies investigated/used. The cloak system **400** consists of six belts of fractal metamaterial (i.e., fractal-resonant structures shown in FIG. **3**) on flexible Taconic EF35 (low loss) circuit board. The belts are shown surround a scattering ring (object). The arrangement is supported by RF transparent plastics in a comb support. The entire system **400** was shown to be easily built up and broken down within a minute or two. The scale in FIG. **4** is about 0.7 meters across. The height of each shell can of course be selected as desired depending on the situation/application.

While embodiments are shown and described herein as having shells in the shape of concentric rings (circular cylinders), shells can take other shapes in other embodiments. For example, one or more shells could have a generally spherical shape (with minor deviations for structural support). In an exemplary embodiment, the shells could form a nested arrangement of such spherical shapes, around an object to be shielded (at the targeted/selected frequencies/wavelengths). Shell cross-sections of angular shapes, e.g., triangular, hexagonal, while not preferred, may be used.

One skilled in the art will appreciate that embodiments and/or portions of embodiments of the present disclosure can be implemented in/with computer-readable storage media (e.g., hardware, software, firmware, or any combinations of such), and can be distributed and/or practiced over one or more networks. Steps or operations (or portions of such) as described herein, including processing functions to derive, learn, or calculate formula and/or mathematical models utilized and/or produced by the embodiments of the present disclosure, can be processed by one or more suitable processors, e.g., central processing units ("CPUs") implementing suitable code/instructions in any suitable language (machine dependent on machine independent).

While certain embodiments and/or aspects have been described herein, it will be understood by one skilled in the art that the methods, systems, and apparatus of the present disclosure may be embodied in other specific forms without departing from the spirit thereof.

For example, while certain wavelengths/frequencies of operation have been described, these are merely representative and other wavelength/frequencies may be utilized or achieved within the scope of the present disclosure.

Furthermore, while certain preferred fractal generator shapes have been described others may be used within the scope of the present disclosure. Accordingly, the embodi-

ments described herein are to be considered in all respects as illustrative of the present disclosure and not restrictive.

What is claimed is:

1. An electrical resonator system, comprising:

a plurality of concentric electrical resonator shells, each shell including a substrate having first and second surfaces and a close-packed arrangement of electrically conductive material formed on the first surface, wherein the closed-packed arrangement comprises a plurality of self-similar electrical resonator shapes and is configured to operate at a desired passband of electromagnetic radiation;

wherein the close-packed arrangements of at least two of the electrical resonator shells are different in size and/or shape; and

wherein a resonator in the close-packed arrangement comprises a second order or higher fractal.

2. The system of claim **1**, wherein said passband is about 2:1.

3. The system of claim **2**, wherein said passband is about 3:1.

4. The system of claim **1**, wherein the electrical system is configured and arranged so that radiation incident on the system from a given direction has an intensity on a point-by-point basis such at each respective antipodal point, relative to an object placed at the center of the system, the radiation has the same or similar intensity.

5. The system of claim **1**, wherein the system is configured and arranged so that radiation incident on the system from a direction in cylindrical coordinates has the same or similar intensity at the antipodal point after having traversed the system.

6. The system of claim **1**, wherein the plurality of shells comprises a first pair of shells having similar closed-packed arrangements for operation at a first passband, wherein the two shells are positioned within ⅛λ of one another.

7. The system of claim **6**, wherein the plurality of shells comprises a second pair of shells having similar closed-packed arrangements for operation at a second frequency band, wherein the two shells are positioned within ⅛λ of one another.

8. The system of claim **1**, wherein the plurality of shells are hemispherical.

9. The system of claim **1**, wherein the plurality of shells are cylindrical.

10. The system of claim **1**, wherein the plurality of shells are spherical.

11. The system of claim **1**, wherein said fractal is selected from the group consisting of a Koch fractal, a Minkowski fractal, a Cantor fractal, a torn square fractal, a Mandelbrot, a Caley tree fractal, a monkey's swing fractal, a Sierpinski gasket, and a Julia fractal.

12. The system of claim **1**, wherein the fractal is selected from the group consisting of a contour set fractal, a Sierpinski triangle fractal, a Menger sponge fractal, a dragon curve fractal, a space-filling curve fractal, a Koch curve fractal, a Lypanov fractal, and a Kleinian group fractal.

13. The system of claim **1**, wherein the plurality of concentric electrical resonator shells are configured and arranged for operation at K band, Ka band, or X-band.

14. The system of claim **1**, wherein the system is operational over a bandwidth from about 500 MHz to about 1500 MHz.

* * * * *

US009847583B1

(12) **United States Patent**
Cohen

(10) **Patent No.: US 9,847,583 B1**
(45) **Date of Patent: *Dec. 19, 2017**

(54) **DEFLECTIVE ELECTROMAGNETIC SHIELDING**

(71) Applicant: **Fractal Antenna Systems, Inc.**, Bedford, MA (US)

(72) Inventor: **Nathan Cohen**, Belmont, MA (US)

(73) Assignee: **Nathan Cohen**, Belmont, MA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 167 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **14/853,298**

(22) Filed: **Sep. 14, 2015**

Related U.S. Application Data

(63) Continuation of application No. 14/071,235, filed on Nov. 4, 2013, now Pat. No. 9,134,465, which is a continuation-in-part of application No. 14/043,178, filed on Oct. 1, 2013.

(60) Provisional application No. 61/744,651, filed on Oct. 1, 2012, provisional application No. 61/796,119, filed on Nov. 3, 2012.

(51) **Int. Cl.**

H01Q 15/00	(2006.01)
H01Q 17/00	(2006.01)

(52) **U.S. Cl.**
CPC ***H01Q 15/0086*** (2013.01)

(58) **Field of Classification Search**
CPC G02B 5/008; H01Q 15/0086; H01Q 15/0093; H01Q 17/00; H01Q 17/002; H01Q 15/00; H01Q 21/06; H01Q 21/20; H01Q 21/205
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

1,434,167	A	10/1922	Thorner
5,250,950	A	10/1993	Scherrer et al.
6,642,881	B1	11/2003	Lawrence et al.
6,876,320	B2	4/2005	Puente Ballarda
7,102,814	B1	9/2006	Hughes
7,256,751	B2	8/2007	Cohen
7,301,493	B1	11/2007	Canales et al.

(Continued)

FOREIGN PATENT DOCUMENTS

KR	10-2010-0072382	7/2010
WO	WO2011/138670 A2	11/2011
WO	WO2014/05573 A1	4/2014

OTHER PUBLICATIONS

Korean International Searching Authority (ISA/KIPO). 2014. International Search Report and Written Opinion, dated Jan. 23, 2014, from PCT Application PCT/US2013/062940 (published as WO 2014/05573 A1).

Primary Examiner — Bernarr E Gregory
(74) *Attorney, Agent, or Firm* — McDermott Will & Emery LLP

(57) **ABSTRACT**

Systems according to the present disclosure provide one or more surfaces that function as power transferring surfaces for which at least a portion of the surface includes or is composed of "fractal cells" placed sufficiently closed close together to one another so that a surface (plasmonic) wave causes near replication of current present in one fractal cell in an adjacent fractal cell. A fractal of such a fractal cell can be of any suitable fractal shape and may have two or more iterations. The fractal cells may lie on a flat or curved sheet or layer and be composed in layers for wide bandwidth or multibandwidth transmission.

34 Claims, 7 Drawing Sheets

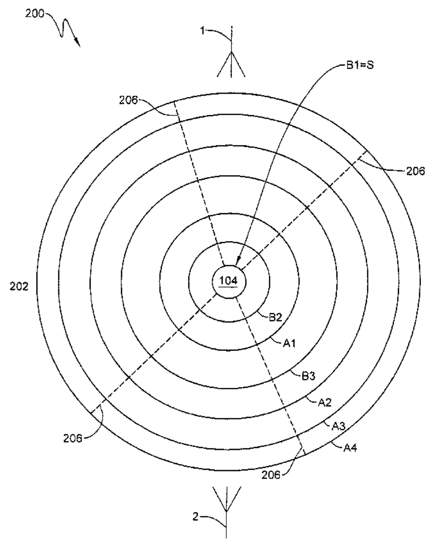

(56) References Cited

U.S. PATENT DOCUMENTS

7,994,962	B1	8/2011	Ben-Shmuel	
8,253,639	B2	8/2012	Cohen	
8,378,877	B2	2/2013	Tishin et al.	
8,610,617	B1	12/2013	Avouris et al.	
8,878,741	B2	11/2014	Mosallaei	
8,937,579	B2	1/2015	Cohen	
9,134,465	B1*	9/2015	Cohen	G02B 5/008
2007/0236406	A1*	10/2007	Wen	H01Q 15/0093 343/909
2008/0024792	A1*	1/2008	Pendry	H01Q 15/00 356/602
2009/0135042	A1	5/2009	Umishita et al.	
2009/0207087	A1	8/2009	Fang et al.	
2010/0126566	A1*	5/2010	Ji	G02B 5/008 136/252
2010/0288116	A1	11/2010	Cincotti et al.	
2011/0050360	A1*	3/2011	Cohen	H01Q 21/205 333/135
2011/0063189	A1	3/2011	Cohen et al.	

* cited by examiner

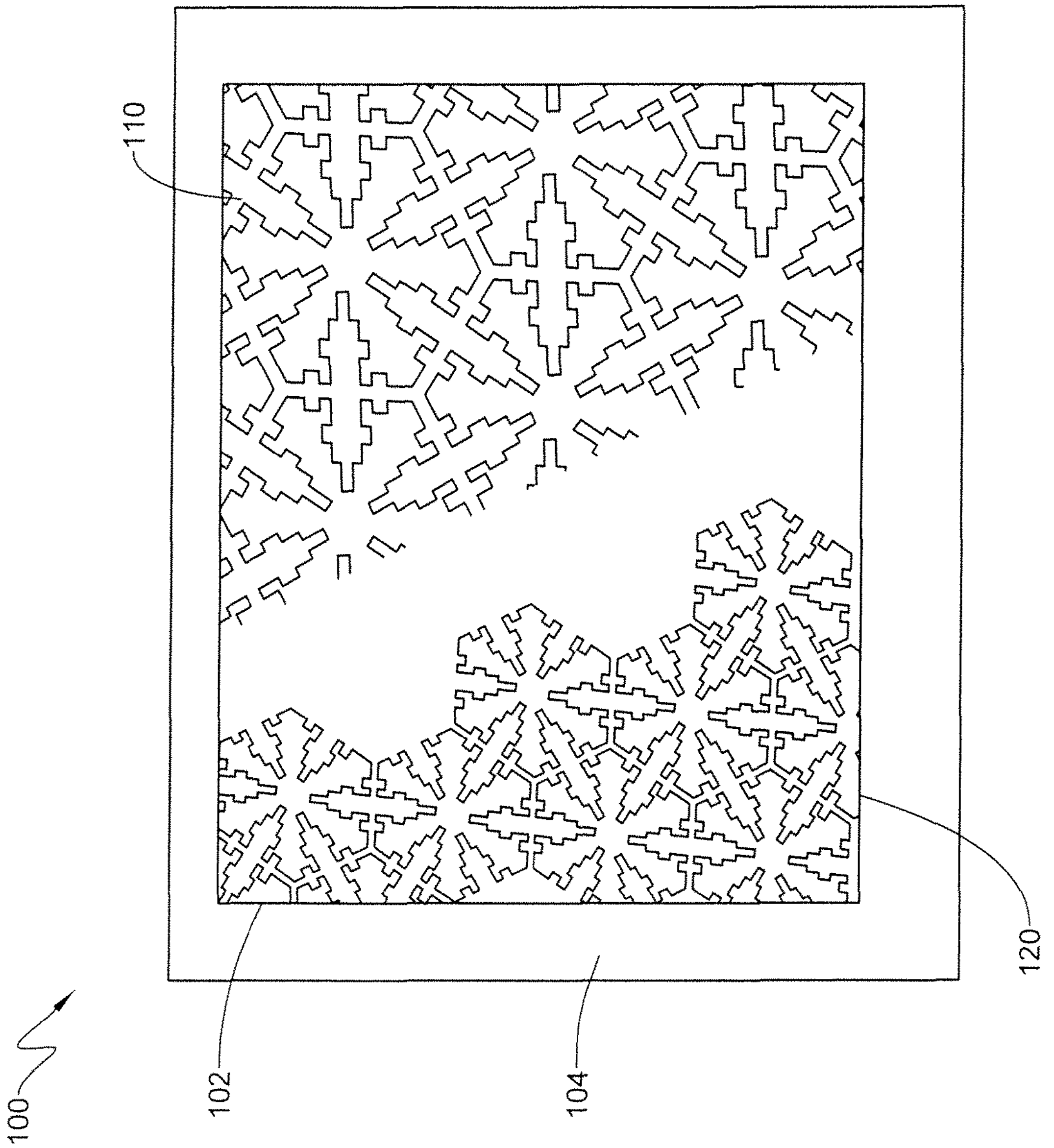
110
120
100
102
104

FIG. 1

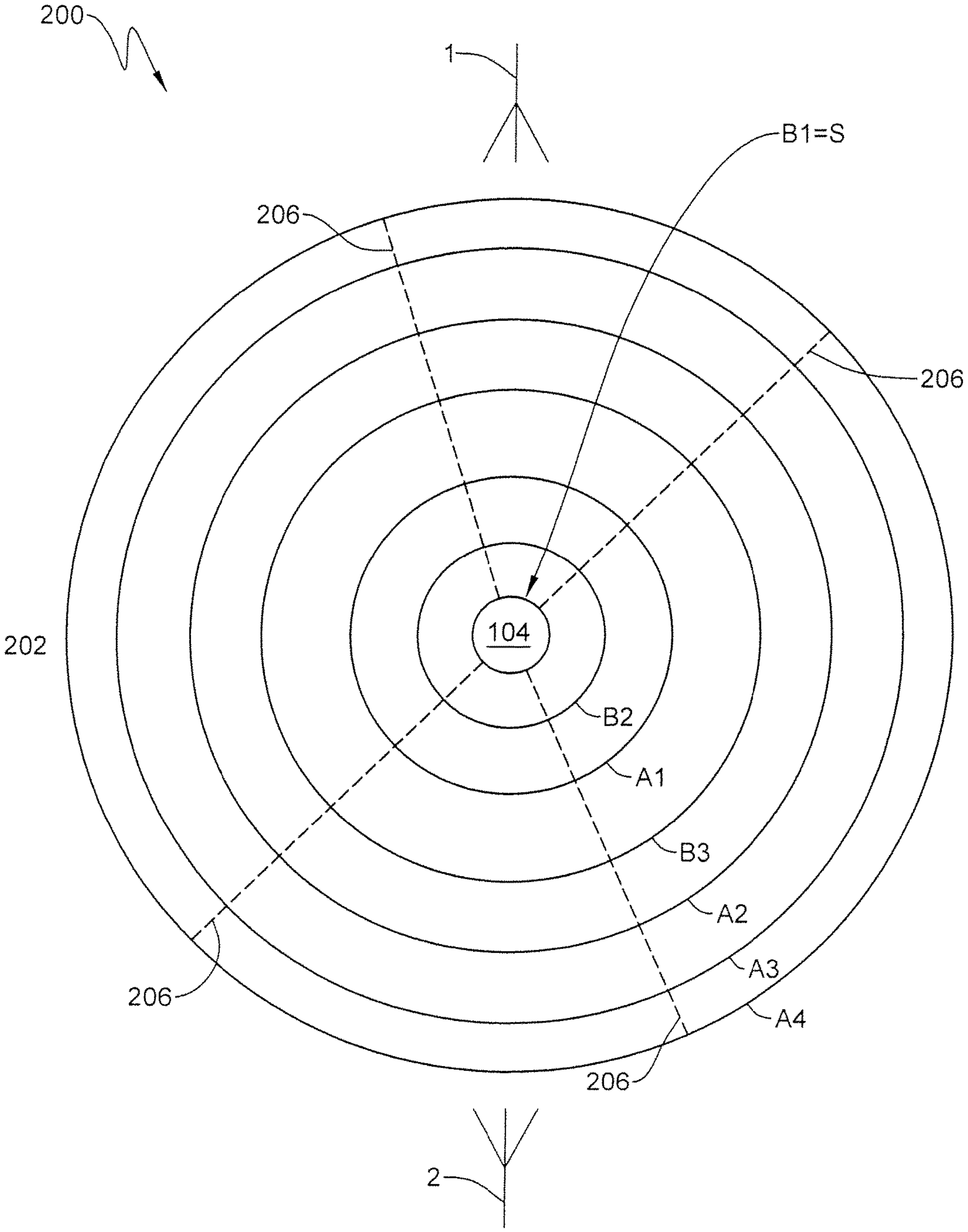
200
1
B1=S
206
206
202
104
B2
A1
B3
A2
A3
A4
206
206
2

FIG. 2

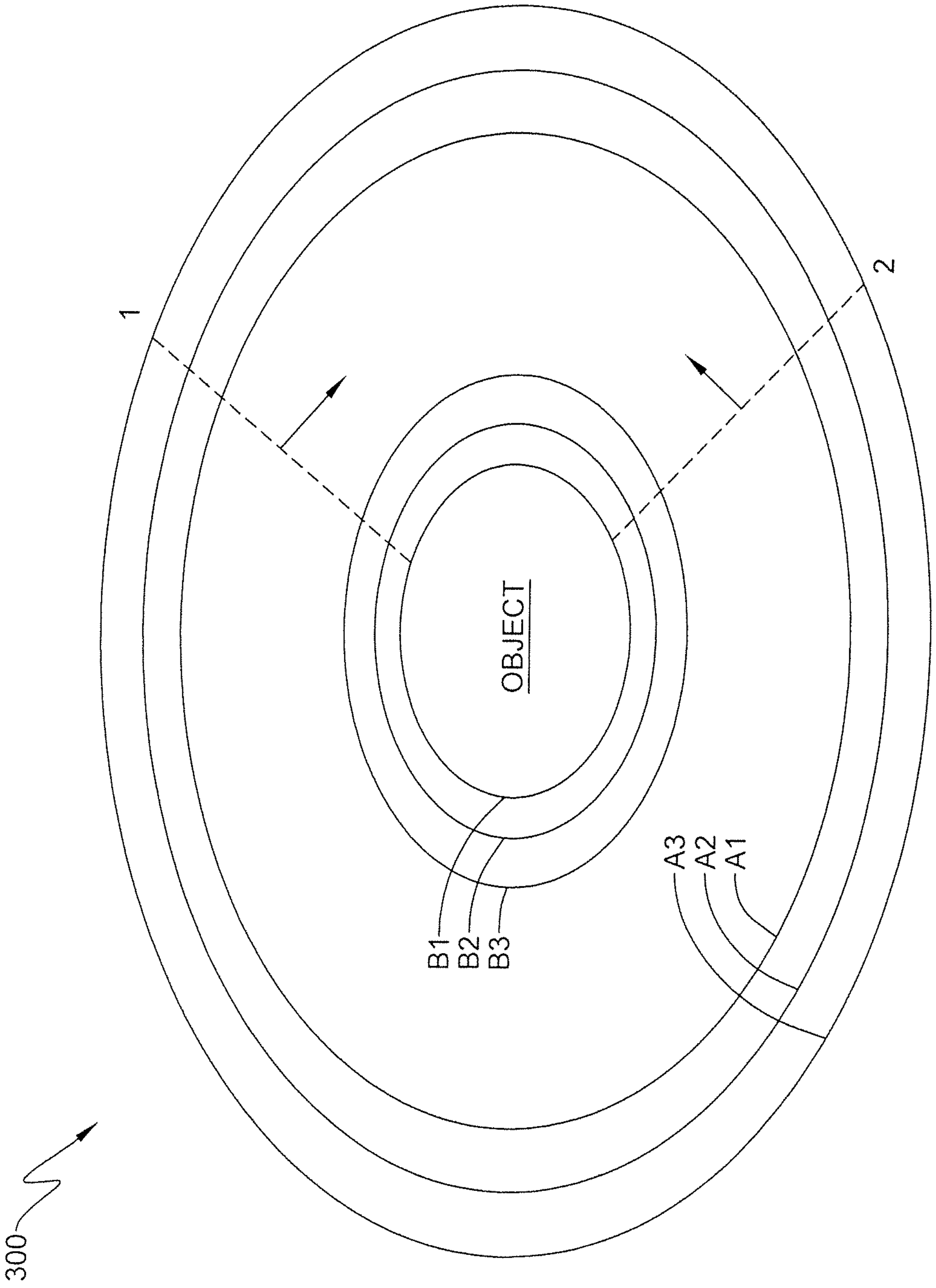
1
2
OBJECT
A3
A2
A1
B1
B2
B3
300

FIG. 3

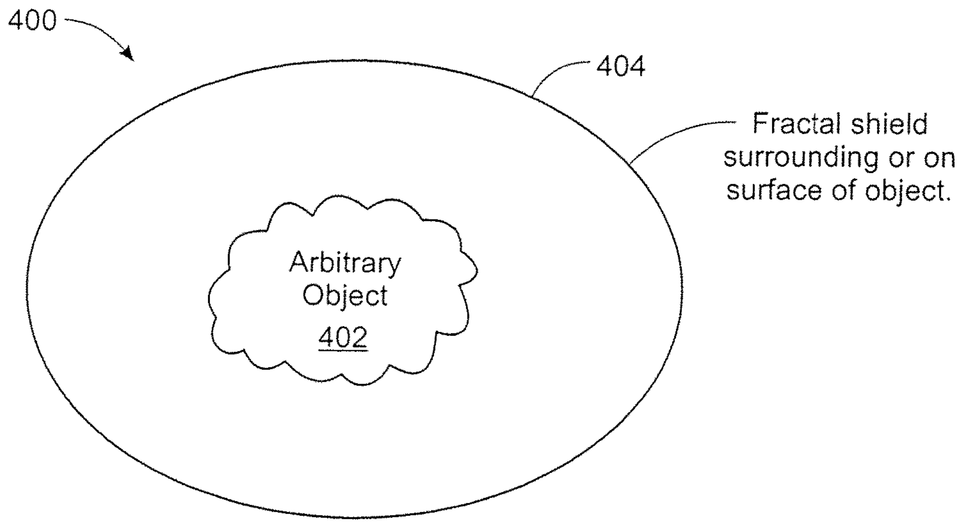
400
404
Fractal shield surrounding or on surface of object.
Arbitrary Object
402

FIG. 4A

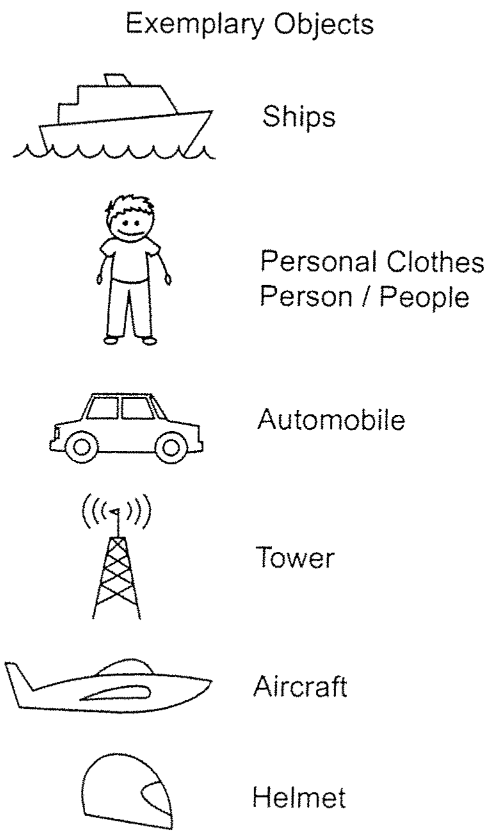
Exemplary Objects
Ships
Personal Clothes Person / People
Automobile
Tower
Aircraft
Helmet

FIG. 4B

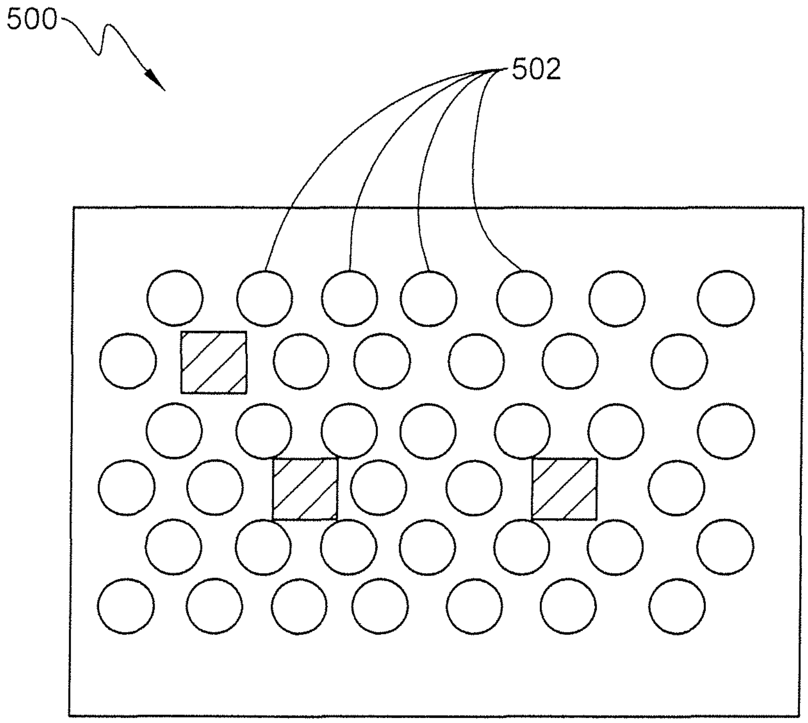

FIG. 5

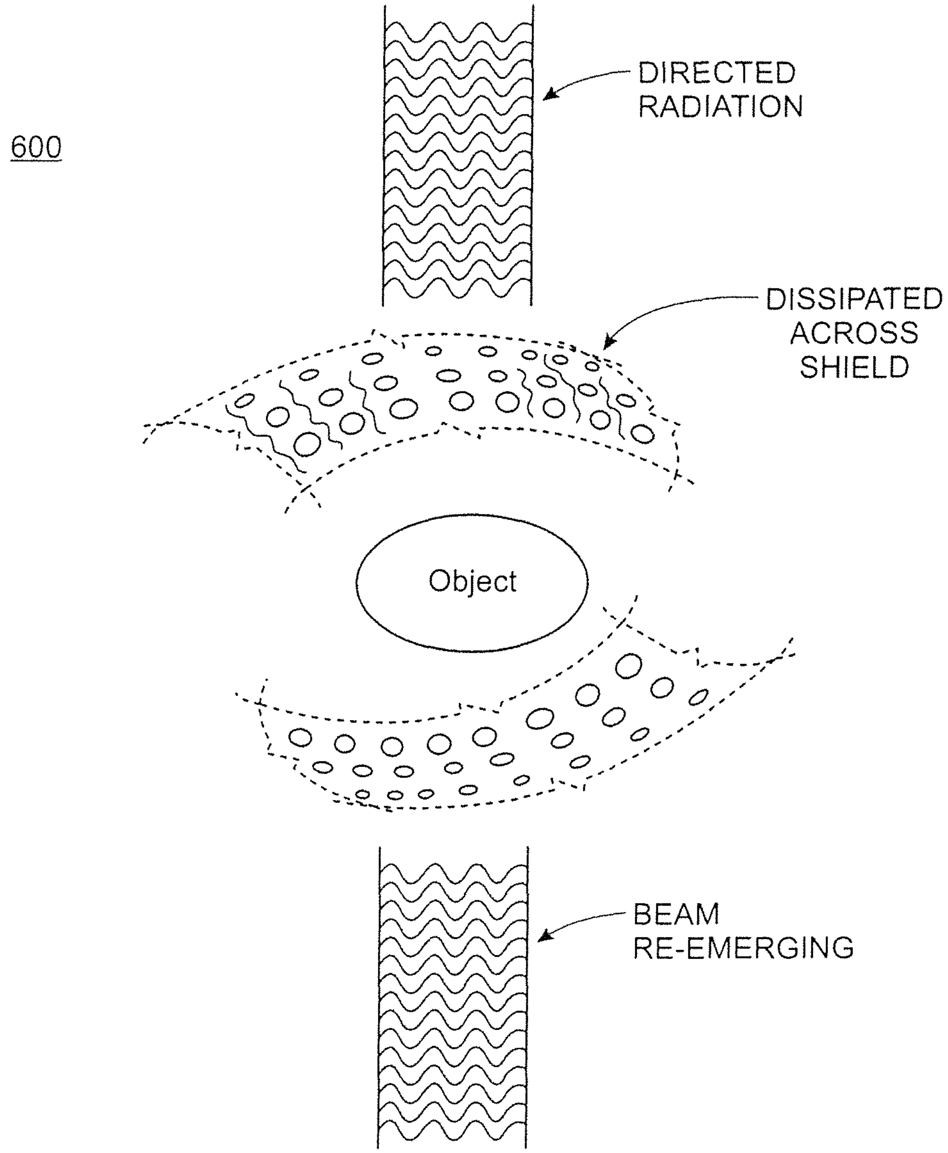
600
DIRECTED RADIATION
DISSIPATED ACROSS SHIELD
Object
BEAM RE-EMERGING

FIG. 6A

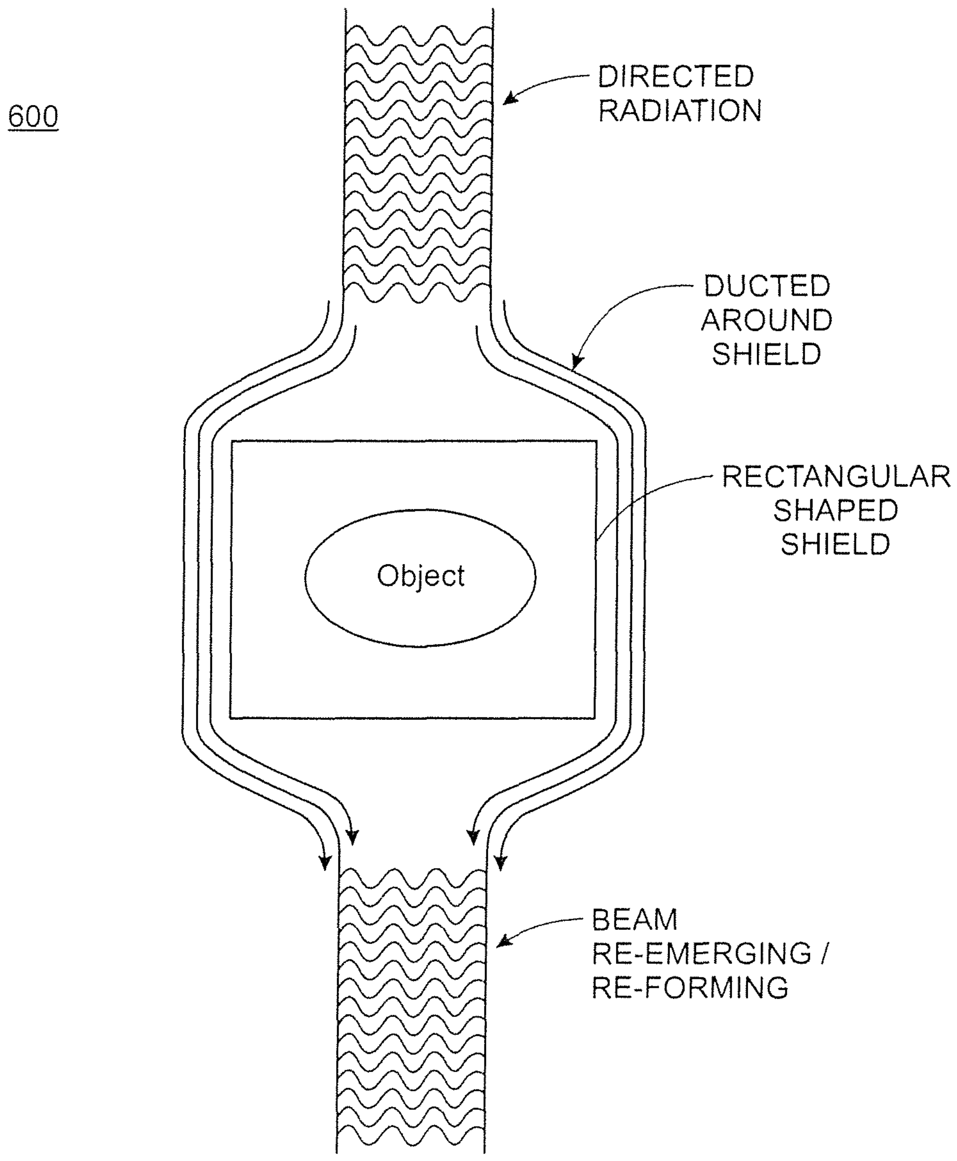
600
DIRECTED RADIATION
DUCTED AROUND SHIELD
RECTANGULAR SHAPED SHIELD
Object
BEAM RE-EMERGING / RE-FORMING

FIG. 6B

DEFLECTIVE ELECTROMAGNETIC SHIELDING

RELATED APPLICATION

This application is a continuation of application Ser. No. 14/071,235 filed 4 Nov. 2013, now U.S. Pat. No. 9,134,465, and entitled "Deflective Electromagnetic Shielding," which (i) claims priority to U.S. Provisional Patent Application No. 61/796,119, filed 3 Nov. 2012 and entitled "Deflective Shielding Via Electromagnetic Cloaking," and (ii) is also a continuation-in-part of U.S. patent application Ser. No. 14/043,781 filed 1 Oct. 2013, now abandoned, and entitled "Radiative Transfer and Power Control with Fractal Metamaterial and Plasmonics," which claims priority to U.S. Provisional Patent Application No. 61/744,651, filed 1 Oct. 2012 and entitled "Radiative Transfer and Power Control with Fractal Metamaterial and Plasmonics"; the entire contents of all of which previously-noted applications are incorporated herein by reference.

BACKGROUND

Much time and effort has been devoted to the quest for so-called invisibility machines. Beyond science fiction, however, there has been little if any real progress toward this goal.

Materials with negative permittivity and permeability leading to negative index of refraction were theorized by Russian noted physicist Victor Veselago in his seminal paper in *Soviet Physics* USPEKHI, 10, 509 (1968). Since that time, metamaterials have been developed that produce negative index of refraction, subject to various constraints. Such materials are artificially engineered micro/nanostructures that, at given frequencies, show negative permeability and permittivity. Metamaterials have been shown to produce narrow band, e.g., typically less than 5%, response such as bent-back lensing. Such metamaterials produce such a negative-index effect by utilizing a closely-spaced periodic lattice of resonators, such as split-ring resonators, that all resonate. Previous metamaterials provide a negative index of refraction when a sub-wavelength spacing is used for the resonators.

In the microwave regime, certain techniques have been developed to utilize radiation-absorbing materials or coatings to reduce the radar cross section of airborne missiles and vehicles. While such absorbing materials can provide an effective reduction in radar cross section, these results are largely limited to small ranges of electromagnetic radiation.

SUMMARY

The present disclosure is directed to systems and techniques that provide for desired transfer of radiation by using close packed arrangements of resonators having fractal shapes, i.e., "fractal cells." Systems and techniques according to the present disclosure provide one or more surfaces that act or function as power transfer surfaces or structures for which at least a portion of the surface or structure includes or is composed of "fractal cells" (fractal shapes) (as antennas or resonators) placed sufficiently closed close together to one another (e.g., less than 1/20 wavelength) to produce substantial replication of current present in one fractal cell in an adjacent fractal cell. The size of the fractal cells can be adjusted based on or tailored to a desired wavelength or wavelength range of operation.

A fractal of such a fractal cell can be of any suitable fractal shape and may have two or more iterations. The fractal cells may lie on a flat or curved sheet or layer and be composed in layers for wide bandwidth or multibandwidth transmission. The area of a surface (e.g., sheet) and its number of fractals determines the gain relative to a single fractal cell. As each cell is fed indirectly (e.g., plasmonicly) rather than directly, 'dead' or covered or out of resonance cells do not alter the ability of the other cells to work. The boundary edges of the surface may be terminated resistively so as to not degrade the cell performance at the edges. These surfaces may be referred to as fractal plasmonic surfaces (FPS's) and can provide benefits such as shielding for/against electromagnetic radiation.

Such a fractal plasmonic surface (FPS) may be used to transfer radiation, e.g., via evanescent wave transfer. In doing so such surfaces can be used to remove radiative power from one location to another, or divert it to another location. Such surfaces may be used to essentially diffuse power delivered to a part of the FPS and dissipate and or distribute it to other specific parts or globally. For example, a power "hotspot" supplied or hit with localized power may have the power spread out or diverted to other desired locations, e.g., different areas of a related machine or structure. In some embodiments, a FPS on a closed or curved surface may act as a cloaking device to divert power (in a wavelength or frequency of interest) from one side of the device to the other.

Exemplary embodiments of the present disclosure can provide techniques, including systems and/or methods, for cloaking objects at certain wavelengths/frequencies or over certain wavelength/frequency ranges (bands). Such techniques can provide an effective electromagnetic lens and/or lensing effect for certain wavelengths/frequencies or over certain wavelength/frequency ranges (bands).

In some embodiments, the effects produced by such techniques can include cloaking or so-called invisibility of the object(s) at the noted wavelengths or bands. Representative frequencies of operation can include, but are not limited to, those over a range of 500 MHz to 1.3 GHz, though others may of course be realized. Operation at other frequencies, including for example those of visible light, infrared, ultraviolet, and as well as microwave EM radiation, e.g., K, Ka, X-bands, etc. may be realized, e.g., by appropriate scaling of dimensions and selection of shape of the resonator elements.

Exemplary embodiments of the present disclosure can include a novel arrangement of resonators in a lattice-like configuration. The arrangements can include resonators of several different sizes and/or geometries arranged so that each size or geometry ("grouping") corresponds to a moderate or high quality factor, "Q," (that is moderate or low bandwidth) response that resonates within a specific frequency range.

For exemplary embodiments, fractal resonators can be used for the resonators in such structures because of their control of passbands, and smaller sizes compared to non-fractal based resonators. Their benefit arises from a size standpoint because they can be used to shrink the resonator(s), while control of passbands can reduce or eliminates issues of harmonic passbands that would resonate at frequencies not desired.

Further embodiments of the present disclosure are directed to scatterer or scattering structures. Additional embodiments of the present disclosure are directed to structures/techniques for activating and/or deactivating cloaking structures. Further embodiments of the present disclosure are directed to wideband absorbers.

It should be understood that other embodiments of FPS systems and methods according to the present disclosure will become readily apparent to those skilled in the art from the following detailed description, wherein exemplary embodiments are shown and described by way of illustration. The systems and methods of the present disclosure are capable of other and different embodiments, and details of such are capable of modification in various other respects. Accordingly, the drawings and detailed description are to be regarded as illustrative in nature and not as restrictive.

BRIEF DESCRIPTION OF THE DRAWINGS

Aspects of the disclosure may be more fully understood from the following description when read together with the accompanying drawings, which are to be regarded as illustrative in nature, and not as limiting. The drawings are not necessarily to scale, emphasis instead being placed on the principles of the disclosure. In the drawings:

FIG. **1** depicts a diagrammatic plan view of a fractal plasmonic surface (FPS), in accordance with exemplary embodiments of the present disclosure;

FIG. **2** depicts a diagrammatic plan view of a resonator cloaking system utilizing a number of cylindrical shells, in accordance with exemplary embodiments of the present disclosure; and

FIG. **3** depicts a diagrammatic plan view of a resonator cloaking system utilizing a number of shells having an elliptical cross-section, in accordance with an alternate embodiment of the present disclosure;

FIG. **4**A depicts an example of a fractal shield surrounding or on an arbitrary object; FIG. **4**B depicts exemplary objects for use with a fractal shield;

FIG. **5** illustrates the robustness or anti-fragility presented by a FPS under conditions where some fractal cells are damaged or otherwise inoperative; and

FIG. **6**A depicts an example of a fractal shield dissipating/transferring incident radiation; FIG. **6**B shows another example in which directed radiation is incident on a rectangular shield.

While certain embodiments depicted in the drawings, one skilled in the art will appreciate that the embodiments depicted are illustrative and that variations of those shown, as well as other embodiments described herein, may be envisioned and practiced within the scope of the present disclosure.

DETAILED DESCRIPTION

Systems and techniques according to the present disclosure provide for desired transfer of electromagnetic radiation by using close packed arrangements of resonators having fractal shapes, i.e., "fractal cells." Systems and techniques according to the present disclosure provide one or more surfaces that act or function as shielding and/or cloaking surfaces for which at least a portion of the surface includes or is composed of "fractal cells" (small fractal shapes, functioning as antennas or resonators) placed sufficiently close to one another so that current present in one fractal cell is replicated or reproduced to an extent in an adjacent fractal cell. Without being limited by any theoretical explanation, surface plasmonic waves are believed to cause such replication in conjunction with evanescent waves.

A fractal of such a fractal cell can be of any suitable fractal shape and may have two or more iterations. The fractal cells may lie on a flat or curved sheet or layer and be composed in layers for wide bandwidth or multibandwidth transmission. As each cell is fed plasmonicly rather than directly, "dead" or covered or out of resonance cells do not alter the ability of the other cells to work. The boundary edges of the surface may be terminated resistively so as to not degrade the cell performance at the edges. These surfaces may be referred to as fractal plasmonic surfaces (FPS's) and they can function to shield structures and/or objects by conveying or ducting EM radiation.

Such a fractal plasmonic surface (FPS) may be used to transfer radiation, e.g., via evanescent wave transfer. In doing so such surfaces can be used to remove or transfer radiative power and/or heat from one location (such as a location receiving incident radiation) to another location. Such surfaces may be used to essentially diffuse power delivered to a part of the FPS and dissipate and or distribute it to other specific parts or globally. For example, a power "hotspot" supplied or hit with localized electromagnetic power may have the power spread out or diverted to other desired locations, e.g., different areas of a related machine or structure. Thus, the structure may act as shielding for another structure or the structure itself. In some embodiments, a FPS on a closed or curved surface may act as a cloaking device to divert power (in a wavelength or frequency of interest) from one side of the device to the other. In other embodiments, a refrigeration or cooling effect may be induced by transferring heat away from a FPS.

In some embodiments, power control may also be achieved through wideband absorption by placing a resistive sheet at the edge of the fractal cells, creating a layer adjacent to the cells. Impinging electromagnetic radiation will be absorbed and not reflected or scattered. A FPS may be used to wirelessly couple a device for power transmission by placing the power as electromagnetic radiation on the FPS and then physically placing the (to be) powered device closely to the FPS. A FPS may also work even if some of the cells are damaged or missing as a variety of paths exist to convey the plasmonic transmission.

Exemplary embodiments of the present disclosure can provide techniques, including systems and/or methods, for cloaking objects at certain wavelengths/frequencies or over certain wavelength/frequency ranges (bands). Such techniques can provide an effective electromagnetic lens and/or lensing effect for certain wavelengths/frequencies or over certain wavelength/frequency ranges (bands). In some embodiments, the effects produced by such techniques can include cloaking or so-called invisibility of the object(s) at the noted wavelengths or bands. Representative frequencies of operation can include, but are not limited to, those over a range of 500 MHz to 1.3 GHz, though others may of course be realized. Operation at other frequencies, including for example those of visible light, infrared, ultraviolet, and as well as microwave EM radiation, e.g., K, Ka, X-bands, etc. may be realized, e.g., by appropriate scaling of dimensions and selection of shape of the resonator elements.

The fractal cells of a FPS can be designed or configured (e.g., sized) based on the wavelength or wavelength range of radiation that the FPS is anticipated as working with. In exemplary embodiments, a fractal cell may be sized such that its spatial extent in a given direction (e.g., a minor diameter, a major diameter, a major axis, a minor axis, etc.) is less than a wavelength of the radiation it is intended to work with. For example, a fractal cell may be constructed that has a spatial extent (in a particular direction) that is ⅛, or less, of the wavelength (lambda, λ) of radiation that the cell is intended to work with. For example, considering a roughly hexagonal shaped fractal cell (e.g., as shown in FIG. **1**), the "flat-to-flat" distance of the hexagonal cell may be ⅛ λ or less of the longest wavelength of radiation foreseen for use with the fractal cell. The spacing between the cell and each of the other adjacent cells (e.g., in this case, six other cells) is preferably 1/16 λ to 1/20 λ. When a wavelength range is anticipated for the cell or FPS having the cell, the wavelength used for design purposes can be the greatest wavelength of operation (corresponding to the lowest frequency of operation). Exemplary embodiments can be designed and built to operate to transfer radiation in the near-IR (NIR) wavelength range. Examples of such wavelengths include but are not limited to the edge of the visible red spectrum (~700 nm) through 1 mm. Other embodiments can operate at medium and/or long-wave infrared wavelengths. As stated above, the spacing between fractal cells in preferably sub-wavelength, e.g., 1/20 of a wavelength. For example, for a desired wavelength of operation of 1 mm (corresponding to 300 GHz), the spacing between adjacent fractal cells could be 50 microns (i.e., 1/20 λ).

Exemplary embodiments of the present disclosure can include an arrangement of resonators in a lattice-like configuration. The arrangements can include resonators of several different sizes and/or geometries arranged so that each size or geometry ("grouping") corresponds to a moderate or high quality factor, "Q," (that is moderate or low bandwidth) response that resonates within a specific frequency range.

For exemplary embodiments, fractal resonators can be used for the resonators in such structures because of their control of passbands, and smaller sizes compared to non-fractal based resonators. Their benefit arises from a size standpoint because they can be used to shrink the resonator(s), while control of passbands can reduce or eliminates issues of harmonic passbands that would resonate at frequencies not desired.

Further embodiments of the present disclosure are directed to scatterer or scattering structures. Additional embodiments of the present disclosure are directed to structures/techniques for activating and/or deactivating cloaking structures.

Related fractal technology is described in the following: (i) U.S. patent application Ser. No. 12/732,059 filed 25 Mar. 2010 and entitled "Wideband Electromagnetic Cloaking Systems," which is a continuation-in-part of (ii) U.S. patent application Ser. No. 12/547,104 filed 25 Aug. 2009, now U.S. Pat. No. 8,253,639, which claims benefit of (iii) U.S. Provisional Patent Application No. 61/189,966 filed 25 Aug. 2008 and (iv) U.S. Provisional Patent Application No. 61/163,824 filed 26 Mar. 2009 and entitled "Cloaking Techniques"; (v) U.S. Provisional Patent Application No. 61/163,837, filed 26 Mar. 2009 and entitled "Scatterer"; (vi) U.S. Provisional Patent Application No. 61/163,913, filed 27 Mar. 2009 and entitled "Cloaking Techniques"; and, (vii) U.S. Provisional Patent Application No. 61/237,360, filed 27 Aug. 2009 and entitled "Switching System for Cloak On Command"; the entire contents of all of which applications are incorporated herein by reference.

For exemplary embodiments, fractal resonators can be used for the resonators because of their control of passbands, and smaller sizes. A main benefit of such resonators arises from a size standpoint because they can be used to shrink the resonator(s), while control of passbands can reduce/mitigate or eliminate issues of harmonic passbands that would resonate at frequencies not desired.

Exemplary embodiments of a resonator system for use at infrared (or nearby) frequencies can be built from belts or loops having fractal cells on one or both sides. These belts or loops can function to slip or duct the infrared (heat) energy around an object located within the belts, so the object is effectively invisible and "see thru" at the infrared frequencies. Belts, or shells, having similar closed-packed arrangements for operation at a first passband can be positioned within a wavelength of one another, e.g., 1/10λ, ⅛λ, ¼λ, ½λ, etc.

In exemplary embodiments, as described in further detail below, an observer can observe an original image or signal, without it being blocked by the cloaked object. Using no power, the fractal cloak having FPS can replicate the original signal (that is, the signal before blocking) with great fidelity. Exemplary embodiments can function in the infrared region (e.g., ~700 nm to ~1 mm, corresponding to ~430 THz to ~300 GHz), providing 3:1 bandwidth; operation within or near such can frequencies can provide other bandwidths as well, such as 1:1 up to 2:1 and up to about 3:1.

FIG. **1** depicts a deflective shield system **100** having a fractal plasmonic surface (FPS) **102**, in accordance with the present disclosure. The FPS **102** includes a close packed arrangements of resonators having fractal shapes (e.g., "fractal cells") as denoted by **110** and **120**. The FPS **102** may be part of a larger surface or area **104**. The individual fractal cells are separated from the adjacent fractal cells but are sufficiently close to one another (e.g., preferably less than 1/20 wavelength) so that a surface (plasmonic) wave causes near replication of current present in one fractal cell in an adjacent fractal cell. While preferred fractal shapes are shown in FIG. **1** as being hexagonal or snowflake-like, any suitable fractal shape (e.g., deterministic) may be used and such a fractal may have two or more iterations. The fractal cells may lie on a flat or curved sheet or layer and be composed in layers for wide bandwidth or multibandwidth transmission; the fractal cells can also lie on more complex three-dimensional surfaces and structures of virtually any shape. For embodiments having layers, each layer holding a FPS can utilize fractal cells of different size and shape that those of another layer.

Examples of suitable fractal shapes (for use for shells and/or a scatting object) can include, but are not limited to, fractal shapes described in one or more of the following patents, owned by the assignee of the present disclosure, the entire contents of all of which are incorporated herein by reference: U.S. Pat. No. 6,452,553; U.S. Pat. No. 6,104,349; U.S. Pat. No. 6,140,975; U.S. Pat. No. 7,145,513; U.S. Pat. No. 7,256,751; U.S. Pat. No. 6,127,977; U.S. Pat. No. 6,476,766; U.S. Pat. No. 7,019,695; U.S. Pat. No. 7,215,290; U.S. Pat. No. 6,445,352; U.S. Pat. No. 7,126,537; U.S. Pat. No. 7,190,318; U.S. Pat. No. 6,985,122; U.S. Pat. No. 7,345,642; and, U.S. Pat. No. 7,456,799.

Other suitable fractal shape for the resonant structures can include any of the following: a Koch fractal, a Minkowski fractal, a Cantor fractal, a torn square fractal, a Mandelbrot, a Caley tree fractal, a monkey's swing fractal, a Sierpinski gasket, and a Julia fractal, a contour set fractal, a Sierpinski triangle fractal, a Menger sponge fractal, a dragon curve fractal, a space-filling curve fractal, a Koch curve fractal, an Lypanov fractal, and a Kleinian group fractal. While exemplary fractal shapes are shown in FIG. **1**, the present disclosure is not limited to such and any other suitable fractal shapes (including generator motifs) may be used in accordance with the present disclosure.

With continued reference to FIG. **1**, the fractal cells **110** and **120** can be formed on the FPS **102** by any suitable techniques. Of course, while a space devoid of cells is shown between cell groups **110** and **120**, such is not necessary (and is included in FIG. **1** merely to show the distinction between the individual cells of the two groups). Such techniques can include additive and/or subtractive techniques. Suitable lithography techniques may be used for some embodiments. In exemplary embodiments, the fractal shapes of the fractal cells **110** and **120** can be conductive traces that are deposited on the underlying surface, e.g., a suitable substrate. Any suitable deposition techniques can be utilized. In other embodiments, the fractal cells **110** and **120** can be etched or engraved onto a surface. Any suitable micromachining or nanomachining techniques may be used.

Exemplary embodiments of system **100** can utilize a material for a substrate that has low-loss characteristics in the infrared region to facilitate heat transfer by fractal cells disposed on, disposed in, or supported by the substrate, e.g., on a supporting surface provided by the substrate. Examples of suitable materials for such substrates can include, but are not limited to, the following: chalcogenide-glasses in general; high-purity As—S, As—Se, Ge—As—Se glasses; and, $Ge_{30}As_{10}Se_{30}Te_{30}$ glass, and the like.

FIG. **2** depicts a diagrammatic plan view of a cloaking system **200** and RF testing set up in accordance with exemplary embodiments of the present disclosure. As shown in FIG. **2**, a number of concentric shells (or bands) **202** are placed on a platform (parallel to the plane of the drawing). The shells include a flexible substrate (e.g., polyimide with or without composite reinforcement) with conductive traces (e.g., copper, silver, etc.) in fractal shapes or outlines, each separate shape representing a fractal cell (e.g., similar to cells **110** and **120** of FIG. **1**). The shells **202** surround an object to be cloaked (shown as **204** in FIG. **1**). A transmitting antenna **1** and a receiving antenna **2** are shown at different sides of the system **200**, for the purposes of illustration. The shells **202** can be held in place by suitable radial supports **206**.

The shells indicated in FIG. **2** are of two types, one set (A**1**-A**4**) configured for optimal operation over a first wavelength/frequency range, and another set (B**1**-B**3**) configured for optimal operation over a second wavelength/frequency range. (The numbering of the shells is of course arbitrary and can be reordered, e.g., reversed.)

For an exemplary embodiment of system **200**, the outer set of shells (A**1**-A**4**, with A**1** being the innermost and A**4** the outmost) had a height of about 3 to 4 inches (e.g., 3.5 inches) and the inner set of shells had a height of about 1 inch less (e.g., about 2.5 to 3 inches). The spacing between the shells with a larger fractal shape (A**1**-A**4**) was about 2.4 cm while the spacing between shells of smaller fractal generator shapes (B**1**-B**3**) was about 2.15 cm (along a radial direction). In a preferred embodiment, shell A**4** was placed between shell B**2** and B**3** as shown. The resonators formed on each shell by the fractal shapes can be configured so as to be closely coupled (e.g., by capacitive and/or evanescent-wave coupling) and function to propagate energy by a plasmonic wave.

It will be appreciated that while, two types of shells and a given number of shells per set are indicated in FIG. **2**, the number of shell types and number of shells for each set can be selected as desired, and may be optimized for different applications, e.g., wavelength/frequency bands, including the optical bands, i.e., infrared, visible, and ultraviolet, as well as x-ray.

FIG. **3** depicts a diagrammatic plan view of a cloaking system (or electrical resonator system) according to an alternate embodiment in which the individual shells have an elliptical cross section. As shown in FIG. **3**, a system **300** for cloaking can include a number of concentric shells (or bands) **302**. These shells can be held in place with respect to one another by suitable fixing means, e.g., they can be placed on a platform (parallel to the plane of the drawing) and/or held with a frame. The shells **302** can include a flexible substrate (e.g., polyimide with or without composite reinforcement) with a close-packed arrangement of electrically conductive material formed on the first surface. As stated previously for FIG. **2**, the closed-packed arrangement can include a number of self-similar electrical resonator shapes. The resonator shapes can be made from conductive traces (e.g., copper, silver, gold, silver-based ink, etc.) having a desired shape, e.g., fractal shape, split-ring shape, and the like. The shells **302** can surround an object to be cloaked, as indicated in FIG. **3**.

As indicated in FIG. **3** (by dashed lines **1** and **2** and arrows), the various shells themselves do not have to form closed surfaces. Rather, one or more shells can form open surfaces. This can allow for preferential cloaking of the object in one direction or over a given angle (solid angle). Moreover, while dashed lines **1** and **2** are shown intersecting shells B**1**-B**3** and A**1**-A**3** of system **300**, one or more shells of each group of shells (B**1**-B**3** and A**1**-A**3**) can be closed while others are open. Additionally, it should be appreciated that the cross-sections shown for each shell can represent closed geometric shapes, e.g., spherical and ellipsoidal shells.

As indicated previously, each shell of a cloaking system (e.g., system **300**) include multiple resonators in one or multiple close-packed configurations. The resonators can be repeated patterns of conductive traces. These conductive traces can be closed geometric shapes, e.g., rings, loops, closed fractals, etc. The resonator(s) can being self similar to at least second iteration. The resonators can include splitring shapes, for some embodiments. The resonant structures are not required to be closed shapes, however, and open shapes can be used for such.

In exemplary embodiments of shell **300**, the closed loops can be configured as a fractals or fractal-based shapes, e.g., as depicted by fractal cells **110** and **120** in FIG. **1**. The dimensions and type of a fractal shape for a fractal cell can be the same for each shell type but can vary between shell types. This variation (e.g., scaling of the same fractal shape) can afford increased bandwidth for the cloaking characteristics of the system. This can lead to periodicity of the fractal shapes of common shell types but aperiodicity between the fractal shapes of different shell types.

It will be appreciated that the resonant structures of the shells may be formed or made by any suitable techniques and with any suitable materials. For example, semiconductors with desired doping levels and dopants may be used as conductive materials. Suitable metals or metal containing compounds may be used. Suitable techniques may be used to place conductors on/in a shell, including, but no limited to, printing techniques, photolithography techniques, etching techniques, and the like.

It will also be appreciated that the shells may be made of any suitable material(s). Printed circuit board materials may be used. Flexible circuit board materials are preferred. Other material may, however, be used for the shells and the shells themselves can be made of non-continuous elements, e.g., a frame or framework. For example, various plastics may be used. In exemplary embodiments, the underlying surface or substrate on which a FPS is formed can have low loss with respect to the type of radiation that the FPS is designed for, so as to facilitate the intended heat and/or power transfer or dissipation.

Exemplary embodiments of the present disclosure can provide techniques, including systems and/or methods, for providing a radar cross section of different sizes than as would otherwise be dictated by the physical geometry of an object. Such techniques (objects/methods) can be useful for implementations such as radar decoys where a given object (decoy) is made to appear in radar cross section as like another object (e.g., missile). Representative frequencies of operation can include those over a range of 500 MHz to 1.3 GHz, though others may of course be realized. Other frequencies, include those of visible light may be realized, e.g., by appropriate scaling of dimensions and selection of shape of fractal elements.

FIG. **4**A depicts an example of a system **400** including an arbitrary object **402** surrounded by a fractal shield **404**. FIG. **4**B depicts exemplary objects **402** that may be shielded by or with a fractal shield **404**. Of course, such objects are merely representative and virtually any object or portion thereof may be shielded by EM radiation by a fractal shield designed (e.g., sized) for such radiation.

FIG. **5** illustrates the robustness or anti-fragility presented by a FPS **500** under conditions where some fractal cells are damaged or otherwise inoperative. As shown, FPS **500** has a close-packed arrangement of fractal cells, indicated by circles **502**. The close-packed arrangement provides many paths by which energy may be transferred from one area of the FPS to another, even in the presence of damaged or otherwise inoperative fractal cells (represented by the black squares shown).

FIG. **6**A depicts an example **600** of a FPS-based shield **602** used to dissipate incident radiation, which may be directed radiation of high intensity or fluence. As shown, radiation that is incident on a localized area of the shield **602** can be transferred from an area of incidence, around the object **604**, and dissipated and/or retransmitted via plasmonic coupling and radiative transfer between the fractal cells of the FPS-based shield **602**.

FIG. **6**B shows another example in which directed radiation is incident on a rectangular shield (exemplifying that a shield need not be circular/elliptical or have circular or elliptical symmetry). As shown, the incident radiation is ducted (indicated by flow lines) over the surface of the shield and around the object within the shield.

Accordingly the subject technology in accordance with the present disclosure can provide protection for a desired object (e.g., a person or thing) from incident radiation such as high energy density electromagnetic waves and/or impulses that would otherwise tend to damage or destroy the object(s). Such deflective shields, or deflectors, can be used on or for three-dimensional surfaces in addition to two-dimensional surfaces. Such a deflector allows the electromagnetic waves or impulses to emerge or re-form on the far side of the object, not passing through the object, wand with minimum (or reduced) absorption and reflection of the waves or impulses; moreover, the momentum of the incident radiation is not conveyed to the objected (or such conveyance is reduced compared to the case where no shield is used).

While embodiments are shown and described herein as having fractal-based EM shields in the shape of concentric rings (circular cylinders), EM shields in accordance with the present disclosure can take other shapes in other embodiments. For example, one or more shells could have a generally spherical shape (with minor deviations for structural support). In an exemplary embodiment, the shells could form a nested arrangement of such spherical shapes, around an object to be shielded (at the targeted/selected frequencies/wavelengths). Shell cross-sections of angular shapes, e.g., triangular, hexagonal, while not preferred, may be used. Further, shields may be placed on or integrated with one or more surfaces of a particular object to be shielded. Such shields and shield structures may be implemented in or configured as conformable surfaces which adapt or conform to any underlying surface or structure.

One skilled in the art will appreciate that embodiments and/or portions of embodiments of the present disclosure can be implemented in/with computer-readable storage media (e.g., hardware, software, firmware, or any combinations of such), and can be distributed and/or practiced over one or more networks. Steps or operations (or portions of such) as described herein, including processing functions to derive, learn, or calculate formula and/or mathematical models and geometric shapes utilized and/or produced by the embodiments of the present disclosure, can be processed by one or more suitable processors, e.g., central processing units ("CPU's") implementing suitable code/instructions in any suitable language (machine dependent on machine independent).

While certain embodiments and/or aspects have been described herein, it will be understood by one skilled in the art that the methods, systems, and apparatus of the present disclosure may be embodied in other specific forms without departing from the spirit thereof.

For example, while certain wavelengths/frequencies of operation have been described, these are merely representative and other wavelength/frequencies may be utilized or achieved within the scope of the present disclosure.

Furthermore, while certain preferred fractal generator shapes have been described others may be used within the scope of the present disclosure. Accordingly, the embodiments described herein are to be considered in all respects as illustrative of the present disclosure and not restrictive.

What is claimed is:

1. A fractal-cell deflective shield system comprising:

a plurality of close-packed fractal cells disposed on a supporting surface, wherein each fractal cell includes a fractal shape defining an electrical resonator, wherein each fractal cell has a spatial extent in a given direction of less than ⅛ of an operational wavelength of incident radiation, and wherein the plurality of fractal cells are positioned sufficiently close to one another to support plasmonic transfer of energy between the fractal cells;

wherein the supporting surface has a configuration adapted to surround an object.

2. The system of claim **1**, wherein the object, corresponding to the configuration of the supporting surface, is a vehicle.

3. The system of claim **1**, wherein the object, corresponding to the configuration of the supporting surface, is a satellite.

4. The system of claim **1**, wherein the object, corresponding to the configuration of the supporting surface, is an automobile.

5. The system of claim **1**, wherein the object, corresponding to the configuration of the supporting surface, is a tower.

6. The system of claim **1**, wherein the object, corresponding to the configuration of the supporting surface, is a person.

7. The system of claim **1**, wherein the object, corresponding to the configuration of the supporting surface, is a vessel.

8. The system of claim **1**, wherein the object, corresponding to the configuration of the supporting surface, is an antenna.

9. The system of claim **1**, wherein the object, corresponding to the configuration of the supporting surface, is a weapon.

10. The system of claim **1**, wherein the object, corresponding to the configuration of the supporting surface, is a ship.

11. The system of claim **1**, wherein the object, corresponding to the configuration of the supporting surface, is a rocket.

12. The system of claim **1**, wherein the object, corresponding to the configuration of the supporting surface, is a spacecraft.

13. The system of claim **1**, wherein the object, corresponding to the configuration of the supporting surface, is an aircraft.

14. The system of claim **1**, wherein the object, corresponding to the configuration of the supporting surface, is a piece of artillery.

15. The system of claim **1**, wherein the object, corresponding to the configuration of the supporting surface, is a radome.

16. The system of claim **1**, wherein the object, corresponding to the configuration of the supporting surface, is a body covering.

17. A fractal-cell shield system comprising:

a first plurality of concentric first shells, each first shell including a substrate having first and second surfaces and a plurality of close-packed fractal cells disposed on a supporting surface disposed on the first surface, wherein each fractal cell includes a fractal shape defining an electrical resonator, and wherein the plurality of fractal cells are positioned sufficiently close to one another to support plasmonic transfer of energy between the fractal cells;

wherein the first plurality of concentric first shells have a configuration adapted to surround an object.

18. The system of claim **17**, further comprising:

a second plurality of concentric second shells, each second shell including a substrate having first and second surfaces and a plurality of close-packed fractal cells disposed on a supporting surface disposed on the first surface, wherein each fractal cell includes a fractal shape defining an electrical resonator, and wherein the plurality of fractal cells are positioned sufficiently close to one another to suppoli plasmonic transfer of energy between the fractal cells;

wherein at least one of the second plurality of second shells is configured at a greater distance from the object than at least one of the first plurality of first shells.

19. The system of claim **18**, wherein each of the second shells are configured at a respective greater distance from the object than the first plurality of first shells.

20. The system of claim **17**, wherein the object, corresponding to the configuration of the first plurality of concentric first shells, is a vehicle.

21. The system of claim **17**, wherein the object, corresponding to the configuration of the first plurality of concentric first shells, is a satellite.

22. The system of claim **17**, wherein the object, corresponding to the configuration of the first plurality of concentric first shells, is an automobile.

23. The system of claim **17**, wherein the object, corresponding to the configuration of the first plurality of concentric first shells, is a tower.

24. The system of claim **17**, wherein the object, corresponding to the configuration of the first plurality of concentric first shells, is a person.

25. The system of claim **17**, wherein the object, corresponding to the configuration of the first plurality of concentric first shells, is a vessel.

26. The system of claim **17**, wherein the object, corresponding to the configuration of the first plurality of concentric first shells, is an antenna.

27. The system of claim **17**, wherein the object, corresponding to the configuration of the first plurality of concentric first shells, is a weapon.

28. The system of claim **17**, wherein the object, corresponding to the configuration of the first plurality of concentric first shells, is a ship.

29. The system of claim **17**, wherein the object, corresponding to the configuration of the first plurality of concentric first shells, is a rocket.

30. The system of claim **17**, wherein the object, corresponding to the configuration of the first plurality of concentric first shells, is a spacecraft.

31. The system of claim **17**, wherein the object, corresponding to the configuration of the first plurality of concentric first shells, is an aircraft.

32. The system of claim **17**, wherein the object, corresponding to the configuration of the first plurality of concentric first shells, is a piece of artillery.

33. The system of claim **17**, wherein the object, corresponding to the configuration of the first plurality of concentric first shells, is a radome.

34. The system of claim **17**, wherein the object, corresponding to the configuration of the first plurality of concentric first shells, is a body covering.

* * * * *

US 20080024792A1

(19) **United States**

(12) **Patent Application Publication**

Pendry et al.

(10) **Pub. No.: US 2008/0024792 A1**

(43) **Pub. Date: Jan. 31, 2008**

(54) **ELECTROMAGNETIC CLOAKING METHOD**

(76) Inventors: **John Pendry**, London (GB); **David Smith**, Durham, NC (US); **David Schurig**, Durham, NC (US)

Correspondence Address:
HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
530 VIRGINIA ROAD, P.O. BOX 9133
CONCORD, MA 01742-9133

(21) Appl. No.: **11/459,728**

(22) Filed: **Jul. 25, 2006**

Publication Classification

(51) **Int. Cl.**
G01B 11/24 (2006.01)

(52) **U.S. Cl.** **356/602**

(57) **ABSTRACT**

A method of constructing a concealing volume comprises constructing a plurality of concealing volume elements around a concealable volume. Each concealing volume element has a material parameter arranged to direct a propagating wave around the concealable volume.

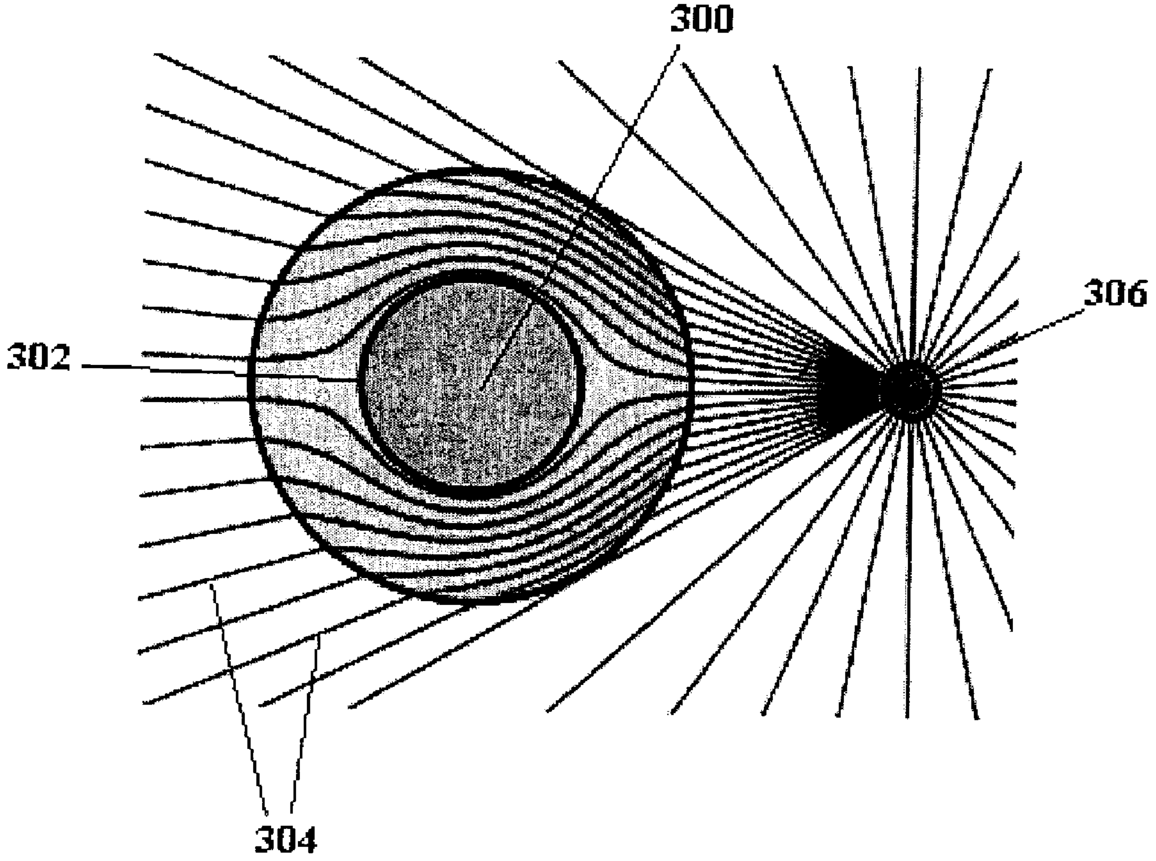

Fig 1

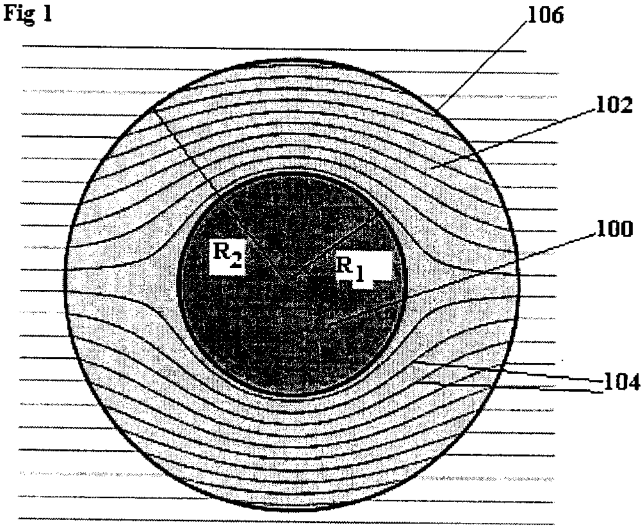
106
102
100
R2
R1
104

Fig 2a

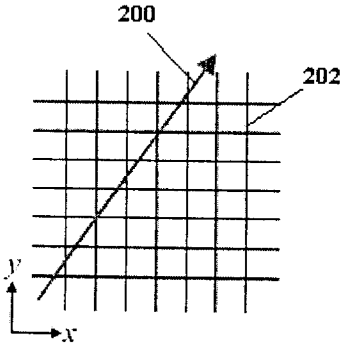

Fig 2b

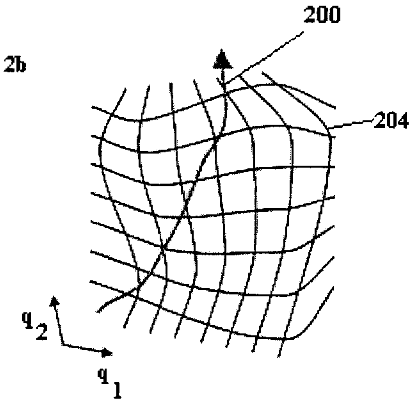
200
204
q2
q1

Fig 3

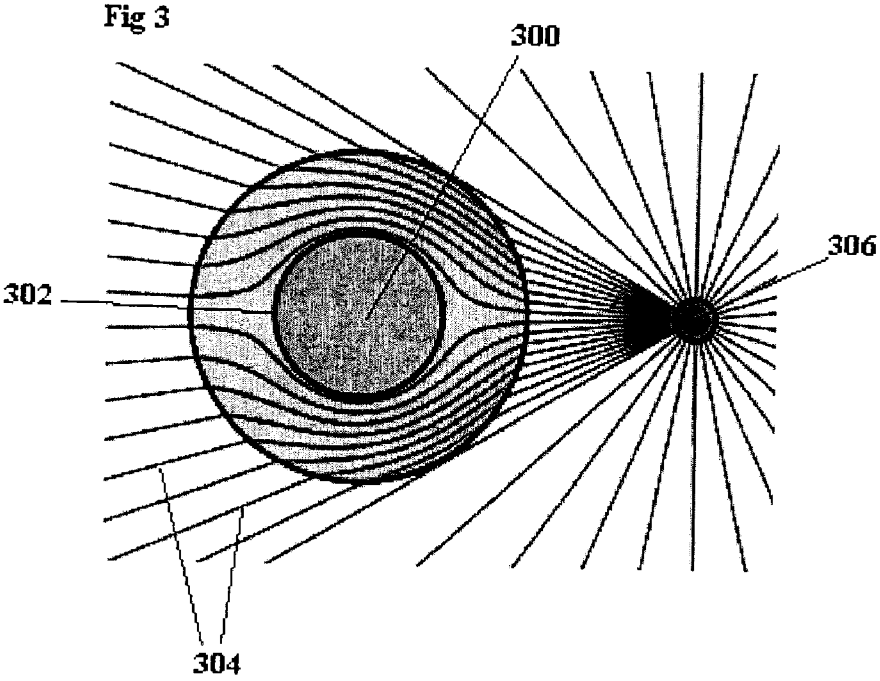
300
306
302
304

Fig 4a

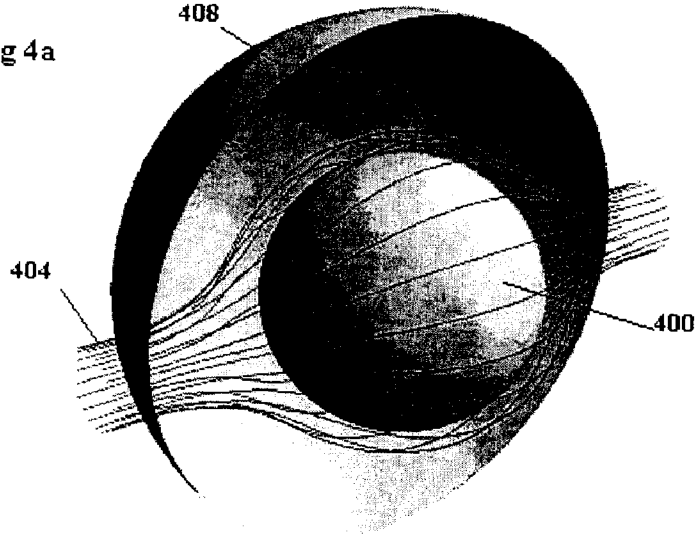

Fig 4b

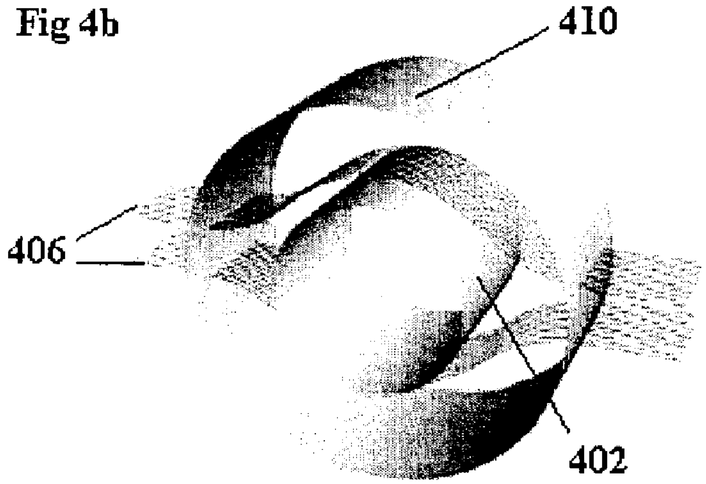

Fig 5

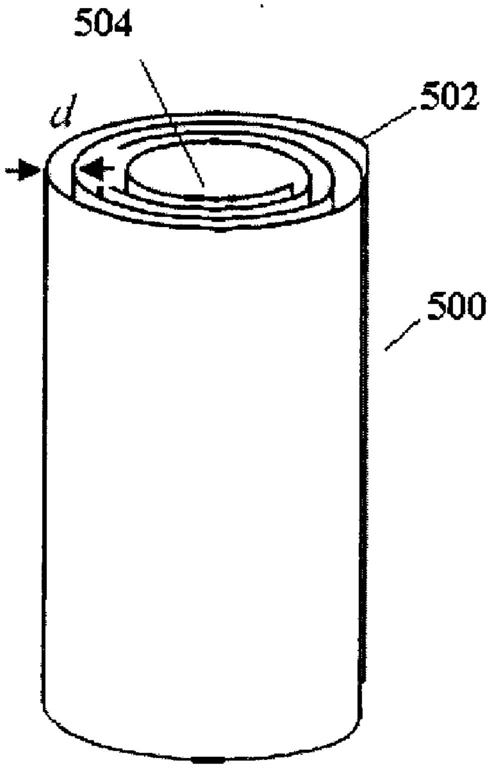
504
502
d
500

FIg 6

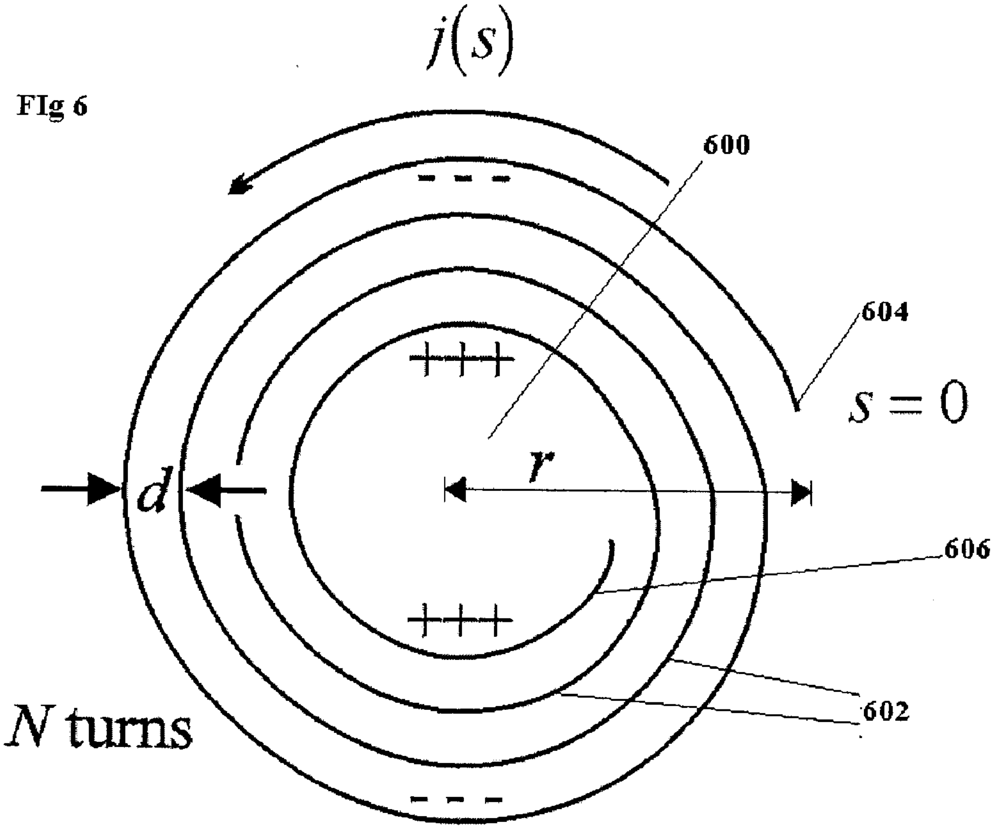

$j(s)$
600
604
$s = 0$
+++
r
d
606
+++
602
N turns
- - -
- - -

Fig 7

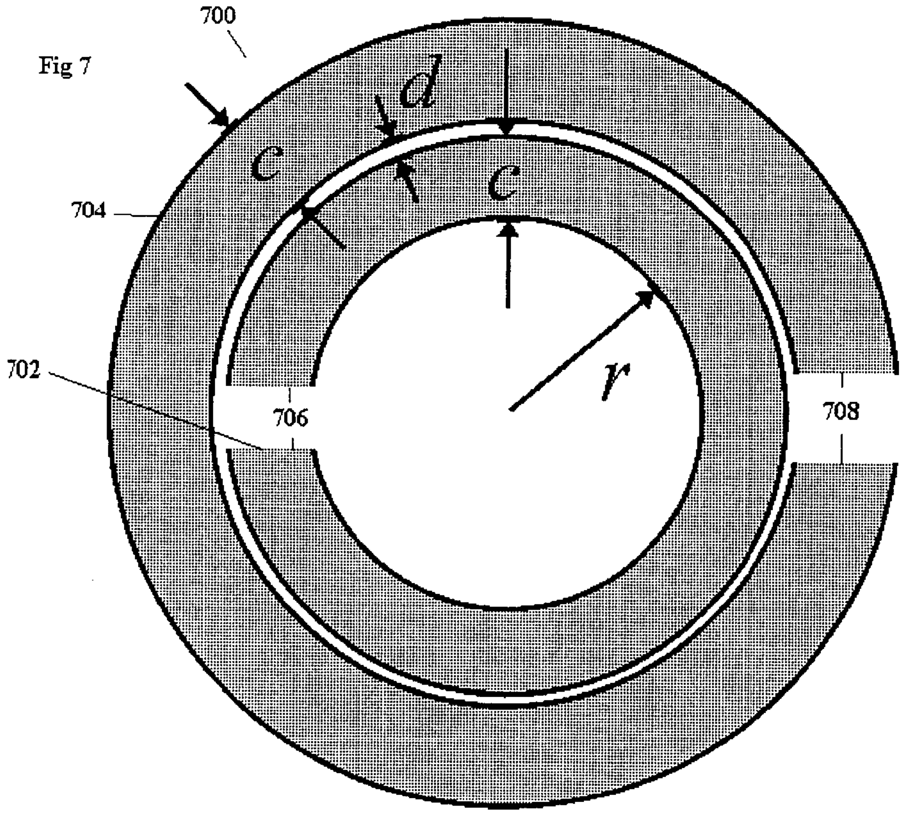
700
d
c
c
704
r
702
706
708

Fig 8

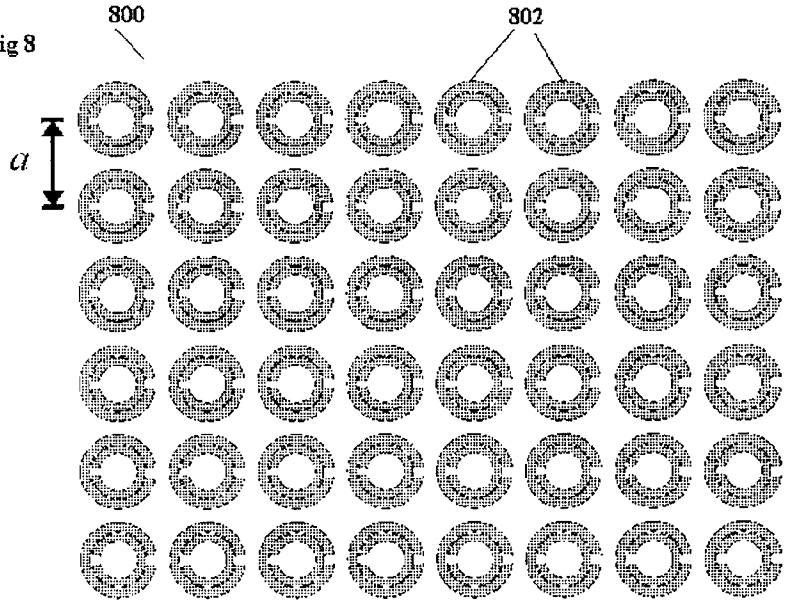

Fig 9

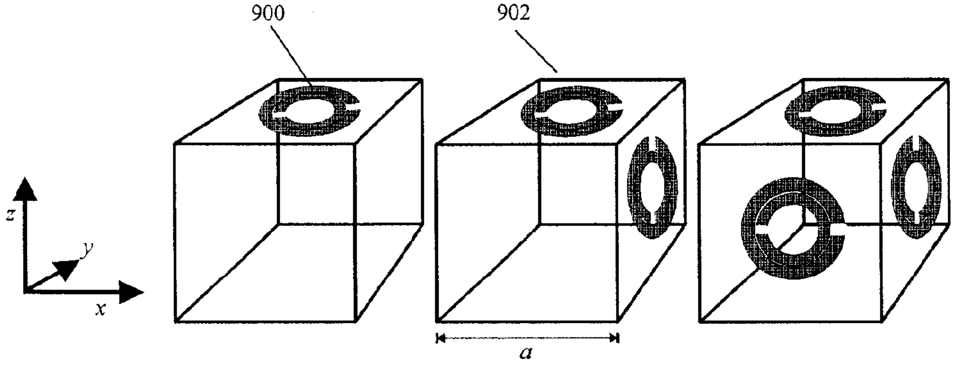
900
902
z
y
x
a

Fig 10

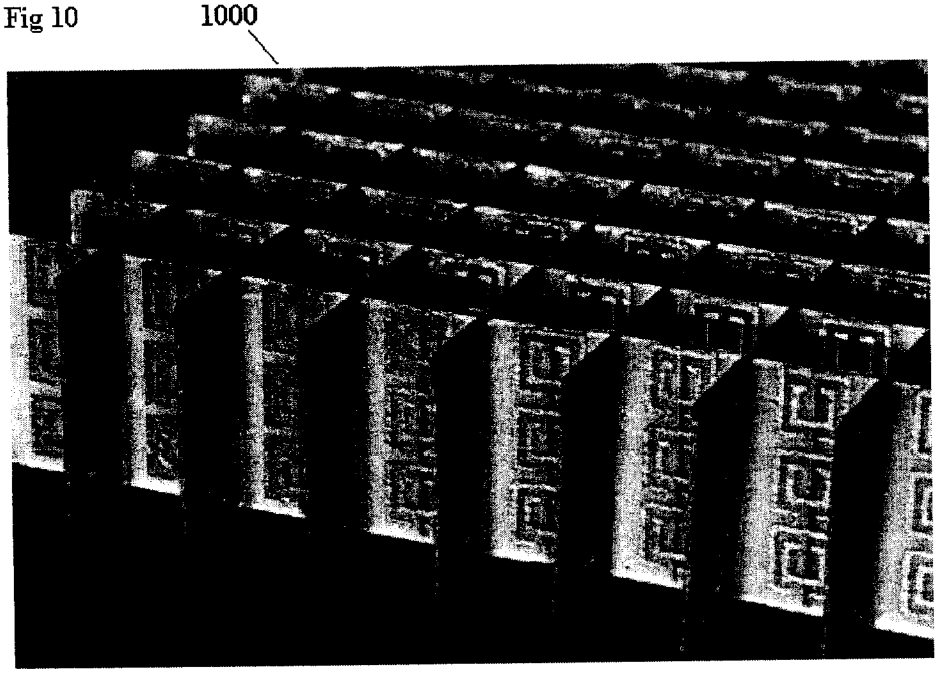

Fig 11

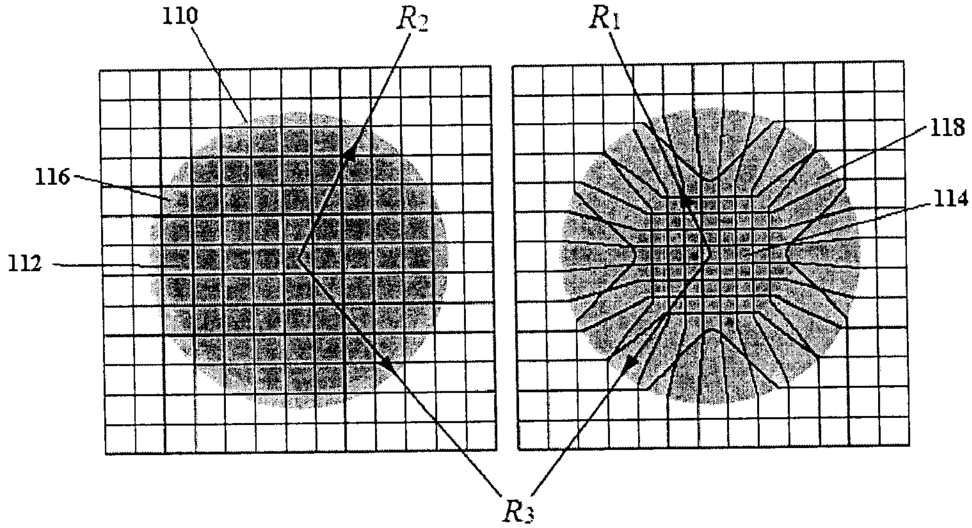
110
R_2
R_1
118
116
114
112
R_3

Fig 12

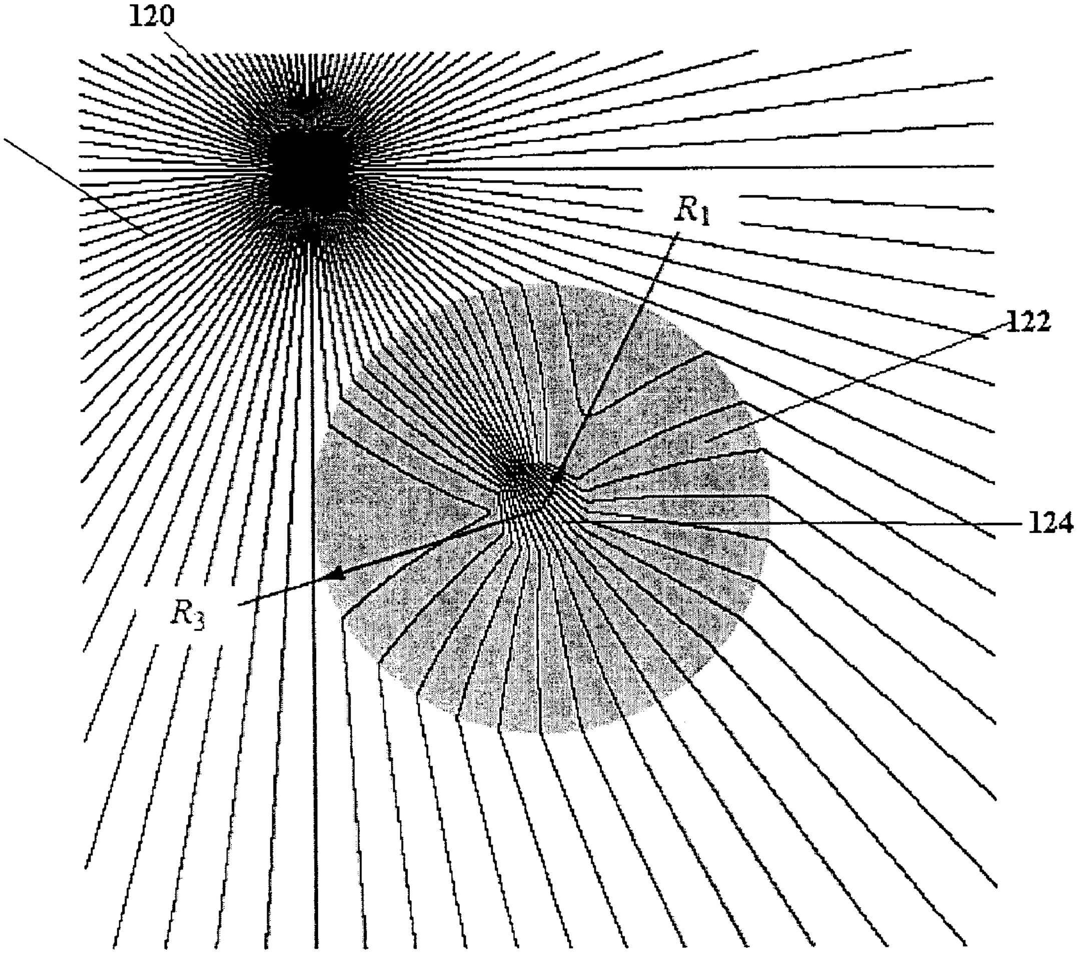
120
R_1
122
124
R_3

ELECTROMAGNETIC CLOAKING METHOD

BACKGROUND

[0001] In order to exploit electromagnetism, materials are commonly used to control and direct the electromagnetic fields. For example, a glass lens in a camera directs the rays of light to form an image, metal cages are used to screen sensitive equipment from electromagnetic radiation, and various forms of 'black bodies' are utilised to prevent unwanted reflections. One aspect of electromagnetism which is of particular interest, is the use of materials in the manipulation of electromagnetic waves such as to conceal or cloak objects or volumes of space from detection by an outside observer.

[0002] Several known methods exist which attempt to achieve electromagnetic concealment of objects. For example, it is possible to use a series of cameras to project an image to an observer of what he would see if an object in question were not blocking his view path. As a result, the observer does not realise that the object is present. This method, however, relies on the use of active components, and depends heavily on the relative positioning of the object, cameras and observer at all times.

[0003] Further known concealment methods include traditional "stealth technology" and the use of low radar cross section structures. These are designed to minimize back reflection of radar or other electromagnetic waves. Whilst these structures can provide a reduced or altered electromagnetic signature, because they involve either the scattering of incident waves away from the target or absorbing incident waves, the objects which they hide are still detectable in transmission.

[0004] In their paper, Physics Rev. E95, 016623 (2005), A Alu and N Engheta suggest a scheme for the concealment of spherical and cylindrical objects through the use of plasmonic and metamaterial 'cloaks' or covers. Whilst this paper provides a method of reducing the total scattering cross section of such objects, it relies on a specific knowledge of the shape and material properties of the object being hidden. In particular, the electromagnetic cloak and the concealed object form a composite, whose scattering properties can be reduced in the lowest order approximation. Therefore, if the shape of the object changes, the shape of the cloak must change accordingly. Furthermore, this method relies on a resonance effect, such that if the frequency drifts away from its resonant peak, the method is less effective. It is therefore a narrowband method, which cannot be implemented for broadband applications.

[0005] A further aspect of electromagnetism which is of interest is the use of materials in electromagnetic sensing and energy harvesting applications. Several known devices exist in this area, such as satellite dishes and solar energy panels. Whilst such prior art devices are operable to collect or detect electromagnetic radiation incident upon them from a number of different directions, and can be moveable to capture radiation incident from any desired direction, they do not have the capability to capture electromagnetic radiation incident from all directions at any given time. Problems therefore arise in applications when the direction of the electromagnetic source is initially unknown or constantly changing, such as in solar energy collection and microwave energy beaming on mobile platforms.

SUMMARY

[0006] The invention is set out in the claims. According to a first embodiment, because a method is provided in which those rays which would have passed through a particular volume of space are deflected around the volume and returned to their original trajectory, an observer would conclude that the rays had passed directly through that volume of space. This will be the case regardless of the relative positioning of the observer and the concealed volume. Furthermore, because no radiation can get into the concealed volume nor any radiation gets out, an object of any shape or material placed in the concealed volume will be invisible to the observer.

[0007] In one aspect, the invention utilises a co-ordinate transformation approach which is independent of the sizes of the concealed and concealing volume and does not suffer from any of the fundamental scaling issues which affect known concealment schemes. It is also possible to use the transformation method for any shape of concealed or concealing volume. The wavelength of the electromagnetic radiation does not appear in the solution, such that the invention can be applied for any size structure and for any wavelength, and for near and far fields.

[0008] In a further embodiment of the invention, all electromagnetic fields which are incident on an outer surface of a particular volume of space are concentrated into a inner core region, regardless of their direction of incidence on the outer surface. This enables a detector or collector to be placed at the inner core into which the electromagnetic fields have been concentrated, allowing it to interact with the intensified fields from all directions at that location. Because a method is described in which those rays which pass through the particular volume of space are returned to their original trajectory outwards of its outer surface, an observer would conclude that the rays had passed directly through that volume of space, regardless of the relative positioning of the observer and the volume in question. In addition, the present invention enables a detector placed within the inner core region to have its material properties matched to the surrounding layer such that it will also be invisible to an observer viewing the set up from any direction.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] FIG. **1** is a cross-sectional ray trajectory diagram for a sphere of radius $r<R_1$ cloaked within an annulus of radius $R_1<r<R_2$, assuming that $R_2>>\lambda$;

[0010] FIG. **2***a* shows a field line in free space plotted on a Cartesian mesh;

[0011] FIG. **2***b* shows the field line and background coordinates of FIG. **2***a* both distorted by the same transformation;

[0012] FIG. **3** is a cross-sectional view of the electrostatic displacement field lines for a sphere of radius $r<R_1$ cloaked within an annulus of radius $R_1<r<R_2$, located close to a point charge, assuming that $R_2<<\lambda$;

[0013] FIG. **4***a* is a three-dimensional depiction of the diversion of electromagnetic rays around a cloaked volume, according to the present invention;

[0014] FIG. **4***b* is a three-dimensional depiction of the diversion of electromagnetic rays around a cloaked surface, according to the present invention;

[0015] FIG. **5** shows a front perspective view of a 'Swiss-roll' metamaterial;

[0016] FIG. **6** shows a top view of the 'Swiss-roll' metamaterial of FIG. **5**;

[0017] FIG. **7** shows a single split ring metamaterial;

[0018] FIG. **8** shows a plan view of a split ring structure in a square array, lattice spacing a;

[0019] FIG. **9** depicts the building of three-dimensional split ring resonator (SRR) at unit cell level;

[0020] FIG. **10** is a prior-art negative index metamaterial comprising SRR's and wires;

[0021] FIG. **11** is a cross-sectional ray trajectory diagram of a concentration device of radius R_3 in which a spherical volume of space radius R_2 is compressed to a smaller volume of space radius R_1; and

[0022] FIG. **12** is a cross-sectional ray trajectory diagram in which rays from a point source incident on a sphere radius R_3 are concentrated into an inner sphere radius R_1.

DETAILED DESCRIPTION

[0023] In overview, according to a first embodiment the invention provides a method of concealing an object or volume of space by means of redirecting the electromagnetic field lines around it. Any field lines which normally would flow through the concealed volume are redirected through a surrounding concealing volume. Outside of the cloaking structure, the field lines remain unperturbed. The cloaking structure is formed from a metamaterial, that is a material designed to exhibit the specific material parameters required for cloaking the object or volume in question. In particular, the electric permittivity ϵ and magnetic permeability μ and hence the refractive index n of the cloaking metamaterial can be varied effectively continuously spatially in order to manipulate the electromagnetic field lines and hence effectively conceal the enclosed volume. As a result of the invention, the concealed volume is undetectable to an outside observer, and any electromagnetic radiation produced by an object within the concealed volume cannot escape from it.

[0024] In a further embodiment the invention provides a method of concentrating the electromagnetic fields incident upon the outer surface of a concentrator device into an inner core region. The concentrator structure comprises a concentrating volume and a collecting volume. Any field lines which would normally flow through the concentrating volume are redirected into an inner collecting volume. Outside of the concentrator annulus surrounding the concentrating volume, the field lines remain unperturbed. The concentrator structure is formed from a metamaterial in a similar manner to that described above for the cloaking structure above.

[0025] Whilst, as is outlined in more detail below, the transformation method described herein can be applied to conceal any desired volume using any desired cloaking structure, for simplicity it will first be described in relation to the simple spherical case as shown in FIG. **1**. According to FIG. **1**, the volume to be hidden is a sphere **100** of radius R_1 and the cloaking material is contained within the annulus **102** of radius $R_1<r<R_2$. In order to create a secure, radiation-free volume within the sphere **100** of radius R_1, all those electromagnetic field lines **104** which normally would flow through the sphere **106** of radius $r<R_2$ must therefore be compressed into the cloaking annulus **102** $R_1<r<R_2$. That is, in order to remove the sphere **100** of radius R_1 from view, the volume within that sphere **100** plus the volume within the annulus **102** must all be mapped onto the annulus **102** of radius $R_1<r<R_2$. In the medium outside of the sphere **106** $r<R_2$, the electromagnetic field lines **104** must remain unperturbed.

[0026] The theory of propagation of electromagnetic waves is well known and so only the relevant aspects are summarised here in the context of the invention. In free space, the magnetic field strength H is related to the magnetic field intensity B and the magnetic permeability of free space, μ_0, via the equation $B=\mu_0 H$. The electric field E in free space is related to the electric displacement field D and electric permittivity of free space ϵ_0 via the equation $D=\epsilon_0 E$. In a dielectric the value of ϵ and μ are modified by the relative permittivity and permeability value ϵ_r, μ_r and the refractive index of the invention is given by

$$n_r = \sqrt[\pm]{\varepsilon\mu}\,.$$

As will be well known to the skilled person in the art, the three vectors E, H, and S are mutually orthogonal, with the Poynting vector S being calculable via the vector cross product $S=E\times H$. The Poynting vector S gives the direction and magnitude of energy flow per unit area per unit time of an electromagnetic wave, and in a ray trajectory diagram such as that shown in FIG. **1**, the direction of the rays of light correspond to the electromagnetic field lines and follow the direction of the Poynting vector. In order to implement this invention therefore, it is necessary to focus or redirect any of the electric displacement field D, the magnetic field intensity B, and the Poynting vector S so that they avoid the sphere **100** of radius R_1, and return undisturbed to their original trajectories.

[0027] As described in more detail below the approach further requires solution of Maxwell's equations in relation to the redirected fields. For ease of depiction, redirection of the Poynting vector in particular will be addressed in the following description.

[0028] In order to calculate the electromagnetic properties which a cloaking structure must possess in order to effectively implement the desired cloaking effects, the first step is to calculate the distortion which must be applied to the Poynting vector S so that the rays do not pass through the sphere **100** of radius R_1. As shown in FIGS. **2***a* and **2***b*, this is achieved by recording the initial configuration of each ray, which would occur in the absence of a cloaked volume and cloaking material, on a known geometric mesh, such as a spherical or Cartesian mesh. For example in FIG. **2***a*, a single ray **200** is plotted on a Cartesian mesh **202**: Once the appropriate rays **200** have been plotted on it, the Cartesian mesh **202** is stretched or distorted in such a manner that the rays plotted thereon become distorted to follow their new desired path, avoiding the cloaked volume. The distortions which are imposed on the Cartesian mesh **202** in order to achieve this can then be recorded as a coordinate transformation between the original Cartesian mesh **202** and a second transformed mesh **204**, wherein the second mesh is defined by:

$$q_1(x,y,z),\ q_2(x,y,z),\ q_3(x,y,z) \qquad (1)$$

[0029] In this second mesh **204**, lines of constant q_2, q_3 define the generalized q_1 axis, and so on. It will be seen that conversely if the second mesh **204** comprised a set of points defined by equal increments along the q_1, q_2, q_3 axes, it would appear distorted in the x, y, z Cartesian frame, as a comparison of FIGS. **2***a* and **2***b* clearly demonstrates.

[0030] The benefit of performing this transformation is that, as described below, it can be shown that it is possible to solve Maxwell's equations in the same form in the transformed geometry as in the original geometry, with only a corresponding transform of the spatially distributed values of ϵ and μ being required. This can be understood intuitively from the recognition that variation of the permittivity ϵ and permeability μ values, and hence the refractive index n in a space, is equivalent to distorting the geometry of the space as seen by a light ray. It will be noted that the approach specifically addresses modification of the value of ϵ and μ individually rather than of the refractive index $n=\pm\sqrt{\epsilon\mu}$. In particular this allows reflection between interfaces of different media to be removed by ensuring impedance (Z) matching where

$$Z = \sqrt{\frac{\mu}{\varepsilon}}.$$

[0031] As will be known to the skilled person, Maxwell's equations in a system of Cartesian coordinates take the form:

$$\nabla\times E=-\mu\mu_0\partial H/\partial t$$

$$\nabla\times H=+\epsilon\epsilon_0\partial E/\partial t \tag{2}$$

where both electric permittivity ϵ and magnetic permeability μ depend on position. As described in Annex 1, when a coordinate transformation is applied to these equations, in the new co-ordinate system they take the form:

$$\nabla_q\times\hat{E}=-\mu_0\hat{\mu}\partial\hat{H}/\partial t,$$

$$\nabla_q\times\hat{H}=+\epsilon_0\hat{\epsilon}\partial\hat{E}/\partial t \tag{3}$$

where ϵ and μ are in general tensors, and Ê, Ĥ are renormalized electric and magnetic fields. All four quantities are simply related to originals. In other words the form of Maxwell's equation is preserved by a co-ordinate transformation: the co-ordinate transformation does not change the fact that we are still solving Maxwell's equations, it simply changes the definition of ϵ, μ.

[0032] In particular, as is shown in detail in Annex 1, the effect of the transformation on μ and ϵ is to scale them both by a common factor. A generalized solution to this is shown in Annex 1, however in the case of an orthogonal geometry (Cartesian, spherical or cylindrical) but the renormalized values of the permittivity and permeability are:

$$\varepsilon'_u = \varepsilon_u\frac{Q_uQ_vQ_w}{Q_u^2}, \tag{4}$$

$$\mu'_u = \mu_u\frac{Q_uQ_vQ_w}{Q_u^2}, \text{ etc.}$$

$$E'_u = Q_uE_u,\ H'_u = Q_uH_u, \text{ etc.} \tag{5}$$

where,

-continued

$$Q_u^2 = \left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2 + \left(\frac{\partial z}{\partial u}\right)^2 \tag{6}$$

$$Q_v^2 = \left(\frac{\partial x}{\partial v}\right)^2 + \left(\frac{\partial y}{\partial v}\right)^2 + \left(\frac{\partial z}{\partial v}\right)^2$$

$$Q_w^2 = \left(\frac{\partial x}{\partial w}\right)^2 + \left(\frac{\partial y}{\partial w}\right)^2 + \left(\frac{\partial z}{\partial w}\right)^2$$

And:

$$B'\mu_0\mu'H' \text{ and } D' = \varepsilon_0\varepsilon E' \tag{7}$$

[0033] Referring back to the spherical case as shown in FIG. **1**, the natural choice of mesh on which to plot the original ray configuration is a spherical one, which uses the co-ordinates r, θ and ϕ. The co-ordinate transformation operates by taking all the fields in the sphere **106** of radius $r<R_2$ and compressing them into the annulus **102** of radius $R_1<r<R_2$. This is achieved by any transformation of the form:

$$r' = R_1 + \frac{f(r)-f(0)}{f(R_2)-f(0)}(R_2-R_1), \tag{8}$$

$$\theta' = \theta,$$

$$\phi' = \phi$$

where *f* (r) is a monotonically increasing function of r, this will map the entire interior of a sphere **106** of radius R_2 into an annulus bounded by two spheres **100**, **106** radii R_1 and R_2 respectively. There is symmetry for ϕ and θ because the field lines are being compressed radially only, so their angular positioning is not distorted.

[0034] The simplest solution to equation (8) arises when f(r)=r, which simplifies to give:

$$r'=R_1+r(R_2-R_1)/R_2,$$

$$\theta'=\theta,$$

$$\phi'=\phi \tag{9}$$

[0035] In order to ascertain the electromagnetic properties which will be required of the cloaking annulus **102** material so that it achieves the desired result of cloaking any object in the sphere **100** R_1 the transformation which has been imposed on r, ϕ and θ must be applied in an analogous manner to ϵ and μ, as shown for the general case in Annex 1. At each point (x, y, z) in space, ϵ and μ will have component (ϵ_x, ϵ_y, ϵ_z), (μ_x, μ_y μ_z) which must be transformed to spherical coordinate valve (ϵ'_r, $\epsilon'_\ominus$, ϵ'ø), (μ_r, · μ⊖.. μ'ø) in the transformed geometry.

[0036] We do this in two steps:

First transform to spherical coordinates without any compression,

[0037]

$$x=r\cos\phi\sin\theta,\ y=r\sin\phi\sin\theta,\ Z=r\cos\theta \tag{A1}$$

so that, as shown in equation (4) above, in the new frame,

$$\tilde{\varepsilon}_i = \varepsilon_i \frac{Q_1 Q_2 Q_3}{Q_i^2}, \qquad \text{(A2)}$$

$$\tilde{\mu}_i = \mu_i \frac{Q_1 Q_2 Q_3}{Q_i^2}$$

We calculate,

[0038]

$$Q_r^2 = \left(\frac{\partial x}{\partial r} = \cos\phi\sin\theta\right)^2 + \left(\frac{\partial y}{\partial r} = \sin\phi\sin\theta\right)^2 + \left(\frac{\partial z}{\partial r} = \cos\theta\right)^2 = 1 \qquad \text{(A3)}$$

$$Q_\theta^2 = \left(\frac{\partial x}{\partial \theta} = r\cos\phi\cos\theta\right)^2 + \left(\frac{\partial y}{\partial \theta} r\sin\phi\cos\theta\right)^2 + \left(\frac{\partial z}{\partial \theta} = -r\sin\theta\right)^2 = r^2$$

$$Q_\phi^2 = \left(\frac{\partial x}{\partial \phi} = -r\sin\phi\sin\theta\right)^2 + \left(\frac{\partial y}{\partial Z} = +r\cos\phi\sin\theta\right)^2 + \left(\frac{\partial z}{\partial Z} = 0\right)^2 = r^2\sin^2\theta$$

and hence,

$$\begin{aligned} \tilde{\varepsilon}_r = \tilde{\mu}_r &= \frac{Q_r Q_\theta Q_\phi}{Q_r^2}\varepsilon_r \\ &= \frac{r^2\sin\theta}{1}\varepsilon_r \\ &= r^2\sin\theta\varepsilon_r \end{aligned} \qquad \text{(A4)}$$

$$\begin{aligned} \tilde{\varepsilon}_\theta = \tilde{\mu}_\theta &= \frac{Q_r Q_\theta Q_\phi}{Q_\theta^2}\varepsilon_\theta \\ &= \frac{r^2\sin\theta}{r^2}\varepsilon_\theta \\ &= \sin\theta\varepsilon_\phi \end{aligned}$$

$$\begin{aligned} \tilde{\varepsilon}_\phi = \tilde{\mu}_\phi &= \frac{Q_r Q_\theta Q_\phi}{Q_\phi^2}\varepsilon_\phi \\ &= \frac{r^2\sin\theta}{r^2\sin^2\theta}\varepsilon_\phi \\ &= \frac{1}{\sin\theta}\varepsilon_\phi \end{aligned}$$

where ϵ_r, ϵ_θ, ϵ_ϕ, the three components of the permittivity tensor in the original Cartesian frame, and we assume that the permittivity and permeability are equal. Note that we can easily extract the Cartesian view of ϵ_r, ϵ_θ, ϵ_ϕ from (A4) by removing the appropriate factors of r.

[0039] In order to create a protected space inside the sphere radius R_1 we make a further transformation to a new cylindrical system in which the radius is compressed taking any ray trajectories with it. It is noted that computer programs work in Cartesian coordinates and therefore that it is possible to re-express the compressed radial coordinate system in terms of a compressed Cartesian system, x' y' z', by removing factors of r'.

Consider the transformation from equation (9),

[0040]

$$r' = R_1 + r(R_2 - R_1)/R_2,\ \theta' = \theta,\ \phi' = \phi \qquad \text{(A5)}$$

which can be re-expressed as,

$$r = \frac{(r' - R_1)R_2}{R_2 - R_1},\ \theta' = \theta,\ \phi' = \phi \qquad \text{(A6)}$$

We calculate,

[0041]

$$Q_{r'}^2 = \left(\frac{\partial r}{\partial r'} = \frac{R_2}{R_2 - R_1}\right)^2 + \left(\frac{\partial \theta}{\partial r'} = 0\right)^2 + \left(\frac{\partial \phi}{\partial r'} = 0\right)^2 = \left[\frac{R_2}{R_2 - R_1}\right] \qquad \text{(A7)}$$

$$Q_{\theta'}^2 = \left(\frac{\partial r}{\partial \theta'} = 0\right)^2 + \left(\frac{\partial \theta}{\partial \theta'} = 1\right)^2 + \left(\frac{\partial \phi}{\partial \theta'} = 0\right)^2 = 1$$

$$Q_{\phi'}^2 = \left(\frac{\partial r}{\partial \phi'} = 0\right)^2 + \left(\frac{\partial \theta}{\partial \phi'} = 0\right)^2 + \left(\frac{\partial \phi}{\partial \phi'} = 1\right)^2 = 1$$

so that in the new frame,
for $R_1 < r' < R_2$

$$\begin{aligned} \tilde{\varepsilon}_{r'} = \tilde{\mu}_{r'} &= \tilde{\varepsilon}_r \frac{Q_{r'} Q_{\theta'} Q_{\phi'}}{Q_{r'}^2} \\ &= \frac{(R_2 - R_1)}{R_2} r^2 \sin\theta\varepsilon_r \\ &= \frac{(R_2 - R_1)}{R_2}\left[\frac{(r' - R_1)R_2}{R_2 - R_1}\right]^2 \sin\theta\varepsilon_r \\ &= \frac{R_2 (r' - R_1)^2}{R_2 - R_1}\sin\theta\varepsilon_r \end{aligned} \qquad \text{(A8a)}$$

$$\begin{aligned} \tilde{\varepsilon}_{\theta'} = \tilde{\mu}_{\theta'} &= \tilde{\varepsilon}_\theta \frac{Q_{r'} Q_{\theta'} Q_{\phi'}}{Q_{\theta'}^2} \\ &= \frac{R_2}{R_2 - R_1}\sin\theta\varepsilon_\theta \end{aligned}$$

$$\begin{aligned} \tilde{\varepsilon}_{\phi'} = \tilde{\mu}_{\phi'} &= \tilde{\varepsilon}_\phi \frac{Q_{r'} Q_{\theta'} Q_{\phi'}}{Q_{\phi'}^2} \\ &= \frac{R_2}{R_2 - R_1}\frac{1}{\sin\theta}\varepsilon_\phi \end{aligned}$$

[0042] Alternatively we can reinterpret these values in a Cartesian coordinate frame, x'y'z', which is easily done by comparison with (A4),

$$\begin{aligned} \varepsilon_{r'} = \mu_{r'} &= \frac{R_2}{R_2 - R_1}\frac{(r' - R_1)^2}{r'^2}\varepsilon_r \end{aligned} \qquad \text{(10)}$$

-continued

$$\begin{aligned}\varepsilon_{\theta'} &= \mu_{\theta'}\\ &= \frac{R_2}{R_2 - R_1}\varepsilon_\theta\\ \tilde{\varepsilon}_{\phi'} &= \tilde{\mu}_{\phi'}\\ &= \frac{R_2}{R_2 - R_1}\varepsilon_\phi\end{aligned}$$

where we have assumed that the starting material is vacuum. for $R_2<r'$

$$\begin{aligned}\tilde{\varepsilon}_{r'} &= \tilde{\mu}_{r'} = r'^2\sin\theta'\varepsilon_{r'}\\ \tilde{\varepsilon}_{\theta'} &= \tilde{\mu}_{\theta'} = \sin\theta'\varepsilon_{\theta'}\\ \tilde{\varepsilon}_{Z'} &= \tilde{\mu}_{Z'} = \frac{1}{\sin\theta'}\varepsilon_{\phi'}\end{aligned} \tag{A7b}$$

In free space, $\epsilon_r'=\mu_r'=\epsilon_\theta'=\mu_\theta'\epsilon_\varnothing'=\mu_\varnothing'=1$ (11)

for $r'<R_1$ we may choose any value we please for the dielectric functions since radiation never penetrates into this region. In this region ϵ', μ' are free to take any value without restriction whilst still remaining invisible, and not contributing to electromagnetic scattering. This will be the case for all cloaked volumes, regardless of their shape or size.

[0043] If a suitable cloaking material can be implemented for the annulus **102** of radius $R_1<r<R_2$, this will exclude all fields from entering the central region **100** of radius R_1 and, conversely, will prevent all fields from escaping from this region. The cloaking structure itself and any object placed in the concealed sphere **100** R_1 will therefore be undetectable in either reflection or transmission. It will be noted that at all points in the medium the impedance

$$z = \sqrt{\frac{\mu}{\varepsilon}} = 1$$

such that no unwanted reflection in the medium, caused by impedance mismatch, takes place.

[0044] It is noted that if the impendance $z=\sqrt{\mu/\epsilon}=1$ is achieved at all points in the metamaterial, the cloaking structure and concealed volume will appear to contain only free space, and so will be invisible if embedded in free space. If the cloaking structure is embedded in another medium the effect will differ, for example if embedded in water it would have the appearance of a bubble. In order to make the cloaking structure invisible in another medium it is necessary to match the ratio of μ and ε in the metamaterial to that of the surrounding medium at the interface between the two. As the skilled person will appreciate, whilst this changes the numerical values of ε and μ in the cloaking structure metamaterial, the general transformation theory outlined in Annex 1 remains the same. That is, because the present invention enables ε and μ to be controlled in a metamaterial independently of one another, it is possible to achieve any desired value of z, to match the impedance of the metamaterial to that of its surroundings.

[0045] Whilst in FIG. **1** the cloaking annulus **102** is depicted as being relatively thick, this does not have to be the case. The solution is not dependent on shell thickness or any other length scales, such that it does not suffer from any scaling issues. Additionally, the solution provided holds in both in the near and far fields, because the solution does not depend on the wavelength λ. This is demonstrated in FIGS. **1** and **3**. For purposes of illustration, suppose that $R_2>>\lambda$ where λ is the wavelength so that we can use the ray approximation to plot the Poynting vector. If our system is then exposed to a source of radiation at infinity we can perform the ray tracing exercise shown in FIG. **1**. Rays **104** in this figure result from numerical integration of a set of Hamilton's equations obtained by taking the geometric limit of Maxwell's equations with anisotropic, inhomogeneous media. This integration provides an independent confirmation that the configuration specified by (9) and (10) excludes rays from the interior region **100**. Alternatively if $R_2<<\lambda$ and we locate a point charge **306** nearby, the electrostatic (or magnetostatic) approximation applies. A plot of the local electrostatic displacement field which results is shown in FIG. **3**.

[0046] The theoretical solution outlined above can be demonstrated using a ray-tracing program to calculate the ray trajectories. According to such a program, a specified material is assumed to exist around a central volume that is to be 'cloaked', or rendered invisible to the user. Using the specified parameters as input, the ray-tracing code confirms the theoretical prediction. Ray tracing programs of this type will be well known to those skilled in the art, although some existing programs will not be able to deal with the complexity of the problem for this invention. In FIGS. **4*a*** and **4*b***, a ray tracing program has been used which was developed from a custom ray tracing code based on a Hamiltonian formulation to solve Maxwell's equations in the geometric limit (ie zero wavelength limit). The results are shown for a spherical structure **400** and a cylindrical structure **402** respectively, wherein the transformed parameters ε and μ in the cylindrical case are calculated in a similar manner to those for the sphere, as shown below:

[0047] First transform to cylindrical coordinates without any compression,

$x=r\cos\theta$, $y=r\sin\theta$, $Z=z$ (A8)

so that in the new frame,

$$\tilde{\varepsilon}_i = \varepsilon_i\frac{Q_1Q_2Q_3}{Q_i^2}, \qquad \tilde{\mu}_i = \mu_i\frac{Q_1Q_2Q_3}{Q_i^2} \tag{A9}$$

We calculate,

[0048]

$$\begin{aligned}Q_r^2 &= \left(\frac{\partial x}{\partial r} = \cos\theta\right)^2 + \left(\frac{\partial y}{\partial r} = \sin\theta\right)^2 + \left(\frac{\partial z}{\partial r} = 0\right)^2 = 1\\ Q_\theta^2 &= \left(\frac{\partial x}{\partial \theta} = -r\sin\theta\right)^2 + \left(\frac{\partial y}{\partial \theta} = +r\cos\theta\right)^2 + \left(\frac{\partial z}{\partial \theta} = 0\right)^2 = r^2\\ Q_Z^2 &= \left(\frac{\partial x}{\partial Z} = 0\right)^2 + \left(\frac{\partial y}{\partial Z} = 0\right)^2 + \left(\frac{\partial z}{\partial Z} = 1\right)^2 = 1\end{aligned} \tag{A10}$$

and hence,

$$\tilde{\varepsilon}_r = \tilde{\mu}_r = \frac{Q_rQ_\theta Q_Z}{Q_r^2}\varepsilon_r = r\varepsilon_r \tag{A11}$$

$$\tilde{\varepsilon}_\theta = \tilde{\mu}_\theta = \frac{Q_rQ_\theta Q_Z}{Q_\theta^2}\varepsilon_\theta = r^{-1}\varepsilon_\theta$$

$$\tilde{\varepsilon}_Z = \tilde{\mu}_Z = \frac{Q_rQ_\theta Q_Z}{Q_Z^2}\varepsilon_Z = r\varepsilon_Z$$

where $\epsilon_r, \epsilon_\theta, \epsilon_Z$ the three components of the permittivity tensor in the original Cartesian frame, and we assume that the permittivity and permeability are equal. Note that we can easily extract the Cartesian view of $\epsilon_r, \epsilon_\theta, \epsilon_Z$ from (A13) by removing the appropriate factors of r.

[0049] In order to create a protected space inside a cylinder radius R_1, we make a further transformation to a new cylindrical system in which the radius is compressed taking any ray trajectories with it. It is noted that computer programs work in Cartesian coordinates and therefore that it is possible to re-express the compressed radial coordinate system in terms of a compressed Cartesian system, x'y'z', by removing factors of r'.

Consider the transformation,

[0050]

$$r' = R_1 + r(R_2 - R_1)/R_2,\ \theta' = \theta,\ Z' = Z \tag{A12}$$

or,

$$r = \frac{(r' - R_1)R_2}{R_2 - R_1},\ \theta' = \theta,\ Z' = Z \tag{A13}$$

We calculate,

[0051]

$$Q_{r'}^2 = \begin{matrix}\left(\frac{\partial r}{\partial r'} = \frac{R_2}{R_2 - R_1}\right)^2 + \left(\frac{\partial \theta}{\partial r'} = 0\right)^2 + \\ \left(\frac{\partial Z}{\partial r'} = 0\right)^2 + \left[\frac{R_2}{R_2 - R_1}\right]^2\end{matrix} \tag{A14}$$

$$Q_{\theta'}^2 = \left(\frac{\partial r}{\partial \theta'} = 0\right)^2 + \left(\frac{\partial \theta}{\partial \theta'} = 1\right)^2 + \left(\frac{\partial Z}{\partial \theta'} = 0\right)^2 = 1$$

$$Q_{Z'}^2 = \left(\frac{\partial r}{\partial Z'} = 0\right)^2 + \left(\frac{\partial \theta}{\partial Z'} = 0\right)^2 + \left(\frac{\partial Z}{\partial Z'} = 1\right)^2 = 1$$

so that in the new frame,
for $R_1 < r' < R_2$

$$\begin{aligned}\tilde{\varepsilon}_{r'} &= \tilde{\mu}_{r'} \\ &= \tilde{\varepsilon}_r\frac{Q_{r'}Q_{\theta'}Q_{Z'}}{Q_{r'}^2} \\ &= \frac{(R_2 - R_1)}{R_2}r\varepsilon_r \\ &= \frac{(R_2 - R_1)}{R_2}\frac{(r' - R_1)R_2}{R_2 - R_1}\varepsilon_r \\ &= (r' - R_1)\varepsilon_r\end{aligned} \tag{A15a}$$

-continued

$$\begin{aligned}\tilde{\varepsilon}_{\theta'} &= \tilde{\mu}_{\theta'} \\ &= \tilde{\varepsilon}_\theta\frac{Q_{r'}Q_{\theta'}Q_{Z'}}{Q_{\theta'}^2} \\ &= \frac{R_2}{R_2 - R_1}r^{-1}\varepsilon_\theta \\ &= \frac{R_2}{R_2 - R_1}\frac{R_2 - R_1}{(r' - R_1)R_2}\varepsilon_\theta \\ &= \frac{1}{(r' - R_1)}\varepsilon_\theta\end{aligned}$$

$$\begin{aligned}\tilde{\varepsilon}_{Z'} &= \tilde{\mu}_{Z'} \\ &= \tilde{\varepsilon}_Z\frac{Q_{r'}Q_{\theta'}Q_{Z'}}{Q_{Z'}^2} \\ &= \frac{R_2}{R_2 - R_1}r\varepsilon_Z \\ &= \frac{R_2}{R_2 - R_1}\frac{(r' - R_1)R_2}{R_2 - R_1}\varepsilon_Z \\ &= \left[\frac{R_2}{R_2 - R_1}\right]^2(r' - R_1)\varepsilon_Z\end{aligned}$$

for $R_2 < r'$

$$\tilde{\epsilon}_{r'} = \tilde{\mu}_{r'} = r'\epsilon_{r'}$$

$$\tilde{\epsilon}_{\theta'} = \tilde{\Xi}_{\theta'} = r'^{-1}\epsilon_{\theta'}$$

$$\tilde{\epsilon}_{Z'} = \tilde{\mu}_{Z'} = r'\epsilon_{Z'} \tag{A15b}$$

for $r' < R_1$ we may choose any value we please for the dielectric functions since radiation never penetrates into this region.

Alternatively we can reinterpret these values in a Cartesian coordinate frame, x'y'z', which is easily done by comparison with (A4),

[0052]

$$\begin{aligned}\varepsilon_{r'} &= \mu_{r'} \\ &= \frac{r' - R_1}{r'}\end{aligned} \tag{A16}$$

$$\begin{aligned}\varepsilon_{\theta'} &= \mu_{\theta'} \\ &= \frac{r'}{(r' - R_1)}\end{aligned}$$

$$\begin{aligned}\varepsilon_{Z'} &= \mu_{Z'} \\ &= \left[\frac{R_2}{R_2 - R_1}\right]^2\frac{r' - R_1}{r'}\end{aligned}$$

where we have assumed that the starting material is vacuum.

[0053] FIGS. **4***a* and **4***b* show that the incident rays **404**, **406** are diverted around a central region **400**, **402** and emerge from the cloaking region **408**, **410** apparently unperturbed, as predicted by the theoretical solution.

[0054] Whilst the theoretical solution has been confirmed by ray-tracing methods, in order to put the invention into effect a suitable material to form the cloaking structure has been developed.

[0055] The material parameters of the cloaking structure require that the components of ϵ' equal those of μ', that ϵ' and μ' vary throughout space, and that the tensor components of $\epsilon'$and μ' vary independently. The anticipated material is thus inhomogeneous and anisotropic. Moreover, the material parameters ϵ' and μ' must assume values less than unity and approach zero at the interface between the concealed region and the cloaking shell. Furthermore, it is noted that equation (6) is singular at $r'=R_1$. This is unavoidable as can be seen by considering a ray headed directly towards the centre of the sphere **100** (FIG. **1**). This ray does not know whether to be deviated up or down, left or right. Neighbouring rays **104** are bent around tighter and tighter arcs closer to the critical ray they are. This in turn implies very steep gradients in ϵ' and μ'. Parameters ϵ' and μ' are necessarily anisotropic in the cloaking material because space has been compressed anisotopically. This set of constraints is not attainable by conventional materials for example because finding matching values for ϵ and μ in the same waveband does not occur. However, given the rapid progress that has occurred over the past several years in artificial materials, a metamaterial can be practically designed that satisfies the specifications.

[0056] Metamaterials are artificially constructed 'materials' which can exhibit electromagnetic characteristics that are difficult or impossible to achieve with conventional materials. From an electromagnetic point of view, the wavelength, λ, of a wave passing through a material determines whether a collection of atoms or other objects can be considered a material and properties at the atomic level determine ϵ and μ. However, the electromagnetic parameters ϵ and μ need not arise strictly from the response of atoms or molecules: Any collection of objects whose size and spacing are much smaller than λ can be described by an ϵ and μ. In that case, the values of ϵ and μ are determined by the scattering properties of the structured objects. Although such an inhomogeneous collection may not satisfy an intuitive definition of a material, an electromagnetic wave passing through the structure cannot tell the difference, and, from the electromagnetic point of view, we have created an artificial material, or metamaterial. This properties of metamaterials are explained further in "Metamaterials and Refractive Index", D R Smith, J B Pendry, MCK. Wiltshire, VOL 305, SCIENCE 061081 (2004).which is incorporated herein by reference.

[0057] One useful property exhibited by certain metamaterials is that of artificial magnetism. It has been found that structures consisting of non-magnetic arrays of wire loops in which an external field can induce a current consequently produce an effective magnetic response. One such design is the so-called 'Swiss-roll' structure, as shown in FIG. **5**. The roll **500** is manufactured by rolling an insulated metallic sheet around a cylinder **504**. A design with about 11 turns on a 1-cm-diameter cylinder gives a resonant response at 21 MHz. The metamaterial is formed by stacking together many of these cylinders. In this structure, the coiled copper sheets have a self-capacitance and self-inductance that create a resonance. The currents that flow when this resonance is activated couple strongly to an applied magnetic field, yielding an effective permeability that can reach quite high values. No current can flow around the coil except by self capacitance. As is shown in FIG. **6**, when a magnetic field parallel to the cylinder **600** is switched on it induces currents (j) in the coiled sheets **602**, which are spaced a distance d apart from one another. Capacitance between the first **604** and last **606** turns of the coil enables current to flow. This is described further in "Magnetism from Conductors and Enhanced Non-Linear Phenomena", J B Pendry, A J Holden, D J Robbins and W J Stewart IEEE Trans. Micr. Theory and Techniques, 47, 2075 (1999), which is incorporated herein by reference.

[0058] One problem with the Swiss-roll structures is that the continuous electrical path provided by metal cylinders can cause the metamaterial to respond like an effective metal when the electric field applied is not parallel to the cylinders, hence restricting its usefulness in certain applications. An adaptation described in the same paper which avoids this undesirable effect is the split ring resonator (SRR). The SRR is built up from a series of split rings, as shown in FIG. **7**. Each split ring **700** comprises at least two concentric thin metal rings **702**, **704** of width c, spaced apart by a distance d, each one with a gap, **706**, **708**. By eliminating the continuous conducting path which the cylinders provide, an SRR eliminates most of the electrical activity along this direction.

[0059] As shown in FIG. **8**, it is possible to form a planar split ring structure **800** from a series of split rings **802**, spaced a distance a apart. The two dimensional square array **800** shown in FIG. **8** can be made by printing with metallic inks. If each printed sheet is then fixed to a solid block of inert material, thickness a, the blocks can be stacked to give columns of rings. This would establish magnetic activity along the direction of stacking, z-axis.

[0060] The unit cell **900** of the stacked SRR structure is shown in FIG. **9** on the left. It is possible to form a symmetrical three dimensional structure starting from such a structure comprising successive layers of rings stacked along the z-axis. This is achievable by cutting up the structure into a series of slabs thickness a, making incisions in the y-z plane and being careful to avoid slicing through any of the rings. Each of the new slabs contains a layer of rings but now each ring is perpendicular to the plane of the slab and is embedded within. The next stage is to print onto the surface of each slab another layer of rings and stack the slabs back together again. The unit cell **902** of this second structure is shown in the middle of FIG. **9**. In the final step, a third set of slabs is produced by cutting in the x-z plane, printing on the surface of the slabs, and reassembling. This results in a structure with cubic symmetry whose unit cell **904** is shown on the right of FIG. **9**.

[0061] Similarly ϵ can be governed by an array of thin wires as described in "Extremely Low Frequency Plasmas in Metallic Mesostructures", J B Pendry, A J Holden, W J Stewart, I Youngs, Phys Rev Lett, 76, 4773 (1996) which is incorporated herein by reference.

[0062] Hence known techniques can be applied to construct the metamaterial required to put the current invention as described above into practice. The materials comprise many unit cells, each of which is designed to have a predetermined value for ϵ' and μ', such that they correspond to the values determined by equation (4) above (or for the non-orthogonal case, equation (23) of Annex 1). As will be apparent to the skilled reader, this may involve each unit cell having slightly different values of ϵ' and μ' to those of its radially neighbouring cells. Each unit cell in the metamaterial acts as a concealing volume element. The concealing volume elements are assembled in the correct respective spatial positions to form the cloaking structure, as calculated by the methods outlined above. The cloaking structure will, as a result, possess the spatially distributed material parameter values required for electromagnetic concealment of the enclosed volume.

[0063] For practical implementation of the theory outlined above, a three dimensional cloaking structure will be necessary in most cases. A solid composite metamaterial structure can be formed using fabrication techniques similar to those used to manufacture a gradient index lens. One such known structure **1000** is shown in FIG. **10**, in which SRR's and wires are deposited on opposite sides lithographically on a standard circuit board to provide a desired material response. This is described in more detail in "Reversing Light with Negative Refraction", J B Pendry, D R Smith, Physics Today, June 2004, which is incorporated herein by reference.

The unit cell structure of the metamaterial used ensures that anisotropy and continuous variation of the parameters, as required by the theoretical solution, can be achieved. Unrelated to cloaking, another application of the transformation theory described above is a device for concentrating fields. FIG. **11** shows a concentrator that concentrates electromagnetic fields incident upon its outer surface (**110**) onto an inner core region (**114**). The transformation that characterizes the concentrator, for the spherical geometry, is shown in FIG. **11**. The outer radius of the device (**110**) is R_3. A concentrating volume of space (**112**) with radius $R_2<R_3$ is compressed to a collecting volume (**114**) of radius R_1. The shell volume (**116**) that lies between R_2 and R_3 is expanded to fill the region (**118**) between R_1 and R_3, as FIG. **11** depicts. A simple example of this transformation that compresses the core (**112**) uniformly and stretches the shell region (**116**) with a linear radial function is given by

$$r = \begin{cases} \frac{R_2}{R_1} r' & 0 < r' < R_1 \\ \frac{R_2 - R_1}{R_3 - R_1} R_3 + \frac{R_3 - R_2}{R_3 - R_1} r' & R_1 < r' < R_3 \\ r' & R_3 < r' \end{cases} \tag{12}$$

$$\theta = \theta'$$

$$\phi = \phi' \tag{13}$$

The resulting material properties of permitting and permeability in the concentration device are then

[0064]

$$\begin{aligned} [\varepsilon'_{r'}, \varepsilon'_{\theta'}, \varepsilon'_{\phi'}] &= [\mu'_{r'}, \mu'_{\theta'}, \mu'_{\phi'}] \\ &= \begin{cases} \frac{R_2}{R_1}[1, 1, 1] & 0 < r' < R_1 \\ \frac{R_3 - R_2}{R_3 - R_1}\left[\left(1 + \frac{R_2 - R_1}{R_3 - R_2}\frac{R_3}{r'}\right)^2, 1, 1\right] & R_1 < r' < R_3 \\ [1, 1, 1] & R_3 < r' \end{cases} \end{aligned} \tag{15}$$

where the material properties are given relative to the external environment. As with the cloaking structure, the external environment may be free space or any other medium.

[0065] In order to concentrate fields into an arbitrarily small volume, we let $R_1 \rightarrow 0$ and obtain

$$\begin{aligned} [\varepsilon'_{r'}, \varepsilon'_{\theta'}, \varepsilon'_{\phi'}] &= [\mu'_{r'}, \mu'_{\theta'}, \mu'_{\phi'}] \\ &= \begin{cases} \frac{R_3 - R_2}{R_3}\left[\left(1 + \frac{R_3}{R_3 - R_2}\frac{R_2}{r'}\right)^2, 1, 1\right] & 0 < r' < R_3 \\ [1, 1, 1] & R_3 < r' \end{cases} \end{aligned} \tag{16}$$

[0066] If we wish the concentration cross section of the sphere to equal it's outer radius, then we let $R_2 \rightarrow R_3$.

$$\begin{aligned} \lim_{R_3 \to R_3} [\varepsilon'_{r'}, \varepsilon'_{\theta'}, \varepsilon'_{\phi'}] &= \lim_{R_3 \to R_3} [\mu'_{r'}, \mu'_{\theta'}, \mu'_{\phi'}] \\ &= \begin{cases} [+\infty, 0, 0] & 0 < r' < R_3 \\ [1, 1, 1] & R_3 < r' \end{cases} \end{aligned} \tag{17}$$

[0067] This is the perfect field concentrator. The material parameters which are required in the concentrator can be achieved through the use of metamaterials as described in relation to the cloaking structure. The permittivity ϵ and permeability μ can be controlled independently of one another in each of three dimensions within each unit cell of the metamaterial.

[0068] The concentrator device is unique, and different from focusing lenses and mirrors, in that it operates on fields incident from any direction. For electromagnetic sensing or energy harvesting applications, a detector or collector would be placed at the core (**114**) to interact with the intensified fields at that location. Part of the utility of this device is concentration of fields when the direction of their source is initially unknown or constantly changing e.g. solar energy collection or microwave energy beaming on mobile platforms.

[0069] A detector placed at the concentrator's core (**114**) could easily be configured to provide accurate directional information as well as collect the energy. This could be accomplished by, for example, using a spherical shaped detector patterned into eight quadrants. By comparing the energy collected from each quadrant, accurate directional information will be obtained. The detector need not be spherical. A transformation could be used that results in a core region that is a thin circular disk. This would make the concentrator compatible with common planar detectors.

[0070] The device can also provide impedance matching to the detector. This is achieved by applying the inverse transformation to the desired detector, expanding it from the core radius, R_1, to the radius R_2. It is then possible to design an impedance matching layer, (between R_1 and R_3), to match the transformed detector to its environment. This impedance matching layer is then transformed with the forward transformation expanding it to fill the region (**118**) between R_1 and R_3. The resulting shell layer between R_1 and R_3, provides both concentration and impedance matching functions.

[0071] Like the cloak structures the concentrator (**110**) can be perfectly matched to its environment, so that if the fields are not disturbed in the core (**114**) by a detector, the device (**110**) is undetectable by external electromagnetic probing. This device (**110**) can operate in both near and far field regimes and at any size scale and wavelength. The basic formalism of this transformation can be applied to other basic geometries, (e.g. cylindrical) or applied conformally to an arbitrarily shaped object.

FIG. 12 demonstrates the concentration of rays or field lines from a point source (**120**) incident upon a sphere (**122**) radius R_3, into an inner core region (**124**) radius R_1.

[0072] The co-ordinate transfer mapping theory outlined above has been described with respect to diverting the propagation of electromagnetic waves. However it is noted that there is a significant analogy between electromagnetic and other propagating waves such as acoustic waves. In particular, this theory may be applied to obtain specifications for the mechanical properties of a material such as density and stiffness that yield acoustic cloaking. Such a cloak may be used for example to conceal bodies from sonar detection.

[0073] Whilst the description of the cloaking theory above is directed mainly to concealing objects, it will be appreciated that the diversion of electromagnetic disturbances from the cloaked volume into a cloaking shell creates an electromagnetic quiet zone inside. Equation (9) dictates that any electromagnetic radiation produced by an object inside the quiet zone will not be able to emerge from it. This may be put to use in electromagnetic interference shielding applications, for example inside MRI machines. Analogously, acoustic quiet zones may also be achieved.

[0074] The application of the present invention is not limited to the cloaking and concentrator structures as shown in the Figures. As demonstrated clearly in Annex 1, the mathematical solution provided herein applies to cloak a structure of any shape. Furthermore, if a volume is cloaked, an object of any shape or material placed therein will not be detectable by an outside observer, either in reflection mode or transmission mode.

[0075] The methods described herein can be used to apply any appropriate coordinate transformation between any appropriate co-ordinate systems and using any appropriate function of the co-ordinates in the original frame. The cloaking structure used to conceal the material within it can be of any shape or thickness. In particular, the cloaking structure does not have to follow the same shape as the object within that is to be concealed. It will be apparent to the skilled reader that the methods described herein can be used to transform all frequencies of electromagnetic radiation, and for a given application can be applied to radiation in any specific frequency range of the electromagnetic spectrum, including infrared, radio, visible light, x-rays, microwaves and radar.

[0076] In the examples shown impedance matching has been achieved by the ratio

$$\frac{\mu}{\varepsilon} = 1.$$

As will be apparent to the skilled reader, impedance matching of the sort described herein can be achieved for

[0077]

$$\frac{\mu}{\varepsilon} = m$$

where m is any real constant value. In particular, the desired ratio of μ/ϵ will depend on the medium in which the cloaking structure is to be embedded.

[0078] While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

What is claimed is:

1. A method of constructing a concealing volume comprising constructing a plurality of concealing volume elements around a concealable volume, each concealing volume element having a material parameter arranged to direct a propagating wave around a concealable volume.

2. A method as claimed in claim **1** in which the material parameter comprises one of refractive index, electrical permittivity, magnetic permeability.

3. A method as claimed in claim **2** in which the propagating wave comprises an electromagnetic wave.

4. A method as claimed in claim **1** in which the material parameter comprises each of electrical permittivity and magnetic permeability.

5. A method as claimed in claim **2** in which the concealing volume is a metamaterial.

6. A method as claimed in claim **5** in which each element comprises at least one of a split ring and a conductive wire providing a respective permeability and permittivity value.

7. A method as claimed in claim **1** in which the concealing volume is designed to divert incoming propagating waves such that outgoing propagating waves appear to be unperturbed to an observer.

8. A method of designing a concealing volume for provision around a concealable volume comprising identifying a co-ordinate transform mapping a concealable volume space onto a concealing volume space and applying a corresponding transform to spatially distributed material parameter values in the concealable volume space.

9. A method as claimed in claim **8** in which the material parameter comprises at least one of the refractive index, electrical permittivity, magnetic permeability.

10. A method of constructing a concealing volume comprising designing a concealing volume as claimed in claim **8**, constructing a plurality of elements having said transformed material parameter values and assembling said elements at the respective transformed spatial positions.

11. A method as claimed in claim **10** in which the assembled elements form a metamaterial.

12. A concealing volume for concealing a concealable volume comprising a plurality of concealing volume elements each having a material parameter arranged to divert a propagating wave around a concealable volume.

13. A concealing volume as claimed in claim **12** in which the material parameter is at least one of refractive index, electrical permittivity and magnetic permeability.

14. A concealing volume as claimed in claim **12** in which the concealing volume elements form a metamaterial.

15. A concealing volume as claimed in claim **12** in which each concealing volume element comprises at least one of a split ring and an electrically conductive wire providing a respective magnetic permeability or electrical permittivity value.

16. A concealing volume as claimed in claim **12** further comprising a concealed volume within said concealing volume.

17. A concealing volume as claimed in claim **12** further comprising a plurality of bandwidth enhancing elements.

18. A computer readable medium comprising a set of instructions for implementing a method as claimed in claim **1**.

19. A computer configured to operate under the instructions of a computer readable medium as claimed in claim **18**.

20. A computer readable medium storing concealing volume element material parameter data derived according to the method of claim **8**.

21. A concealing volume as claimed in claim **12** configured for cloaking in at least one of the following electromagnetic spectrum bands: visible, infrared, ultraviolet, x-ray, radio, radar, microwave.

22. A computer readable medium as claimed in claim **18** in which the propagating wave is an acoustic wave and the material parameter comprises at least one of density and stiffness.

23. A method of constructing a concentrator comprising constructing a plurality of concentrator elements to form a concentrating volume and a collecting volume, each concentrator element having a material parameter arranged to direct a propagating wave from the concentrating volume to the collecting volume.

24. A method as claimed in claim **23** in which the material parameter comprises one of refractive index, electrical permittivity, magnetic permeability.

25. A method as claimed in claim **24** in which the propagating wave comprises an electromagnetic wave.

26. A method as claimed in claim **23** in which the material parameter comprises each of electrical permittivity and magnetic permeability.

27. A method as claimed in claim **24** in which the concentrator is a metamaterial.

28. A method as claimed in claim **27** in which each element comprises at least one of a split ring and a conductive wire providing a respective permeability and permittivity value.

29. A method as claimed in claim **23** in which the concentrator is designed to divert incoming propagating waves such that outgoing propagating waves appear to be unperturbed to an observer.

30. A method of designing a concentrator for provision around a concentrating volume comprising identifying a co-ordinate transform mapping the concentrating volume space onto a collecting volume space and applying a corresponding transform to spatially distributed material parameter values in the concentrator space.

31. A method as claimed in claim **30** in which the material parameter comprises at least one of the refractive index, electrical permittivity, magnetic permeability.

32. A method of constructing a concentrator comprising designing a concentrator as claimed in claim **36**, constructing a plurality of elements having said transformed material parameter values and assembling said elements at the respective transformed spatial positions.

33. A method as claimed in claim **32** in which the assembled elements form a metamaterial.

34. A concentrator for concentrating a concentrating volume comprising a plurality of concentrator elements each having a material parameter arranged to divert a propagating wave from the concentrating volume to a collecting volume.

35. A concentrator as claimed in claim **34** in which the material parameter is at least one of refractive index, electrical permittivity and magnetic permeability.

36. A concentrator as claimed in claim **34** in which the concentrator elements form a metamaterial.

37. A concentrator as claimed in claim **34** in which each concentrator element comprises at least one of a split ring and an electrically conductive wire providing a respective magnetic permeability or electrical permittivity value.

38. A concentrator as claimed in claim **34** further comprising a collecting volume within said concentrating volume.

39. A concentrator as claimed in claim **34** further comprising a plurality of bandwidth enhancing elements.

40. A computer readable medium comprising a set of instructions for implementing a method as claimed in claim **23**.

41. A computer configured to operate under the instructions of a computer readable medium as claimed in claim **40**.

42. A computer readable medium storing concentrator element material parameter data derived according to the method of claim **30**.

43. A concentrator as claimed in claim **34** configured for concentrating in at least one of the following electromagnetic spectrum bands: visible, infrared, ultraviolet, x-ray, radio, radar, microwave.

44. A method for controlling a propagating wave within a material comprising controlling the values of electrical permittivity and magnetic permeability exhibited by the material independently of one another.

45. A method as claimed in claim **44** in which the material is a metamaterial.

46. A method as claimed in claim **44** in which the propagating wave is an electromagnetic wave.

47. A method as claimed in claim **44** wherein the propagating wave is controlled to circumvent a concealable volume.

48. A method as claimed in claim **44** wherein the propagating wave is directed from a concentrating volume to a collecting volume.

49. A method of defining a spatial distribution of material parameter values of an object comprising performing a spatial coordinate transformation in relation to a volume having a pre-transformation material parameter value spatial distribution, and constructing an object from a metamaterial having a post-transformation material parameter value spatial distribution.

50. A method of defining a spatial distribution of material parameter values of an object comprising performing a spatial coordinate transformation in relation to a volume having a pre-transformation material parameter value spatial distribution having an effect on a volume surrounding the object, in which the spatial transformation is selected so that the post-transformation material parameter value spatial distribution achieves the same effect on the volume surrounding the object.

51. A method as claimed in claim **50** in which the effect on the surrounding volume is non-perturbation of electromagnetic radiation by virtue of presence of the object.

* * * * *

US 20080165442A1

(19) **United States**

(12) **Patent Application Publication**

Cai et al.

(10) **Pub. No.: US 2008/0165442 A1**

(43) **Pub. Date: Jul. 10, 2008**

---

(54) **SYSTEM, METHOD AND APPARATUS FOR CLOAKING**

(76) Inventors: **Wenshan Cai**, West Lafayette, IN (US); **Vladimir M. Shalaev**, West Lafayette, IN (US); **Uday K. Chettiar**, West Lafayette, IN (US); **Alexander V. Kildishev**, West Lafayette, IN (US)

Correspondence Address:
BRINKS HOFER GILSON & LIONE
P.O. BOX 10395
CHICAGO, IL 60610

(21) Appl. No.: **11/983,228**

(22) Filed: **Nov. 7, 2007**

Related U.S. Application Data

(60) Provisional application No. 60/857,526, filed on Nov. 8, 2006.

Publication Classification

(51) **Int. Cl.**
G02B 27/00 (2006.01)

(52) **U.S. Cl.** **359/896**

(57) **ABSTRACT**

An apparatus and method of cloaking is described. An object to be cloaked is disposed such that the cloaking apparatus is between the object and an observer. The appearance of the object is altered and, in the limit, the object cannot be observed, and the background appears unobstructed. The cloak is formed of a metamaterial where the properties of the metamaterial are varied as a function of distance from the cloak interfaces, and the permittivity is less than unity. The metamaterial may be fabricated as a composite material having a dielectric component and inclusions of particles of sub-wavelength size, so as to have a permeability substantially equal to unity.

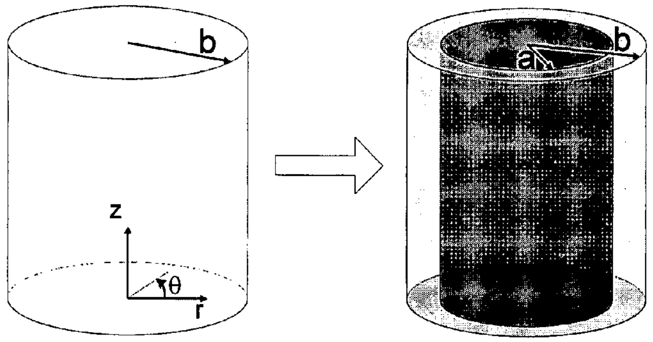

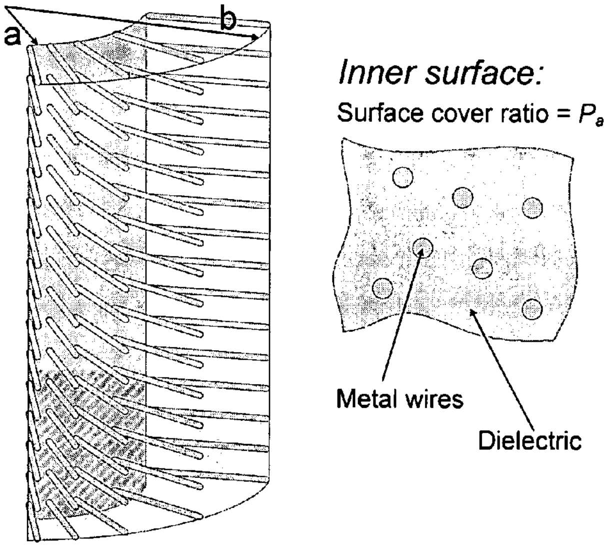

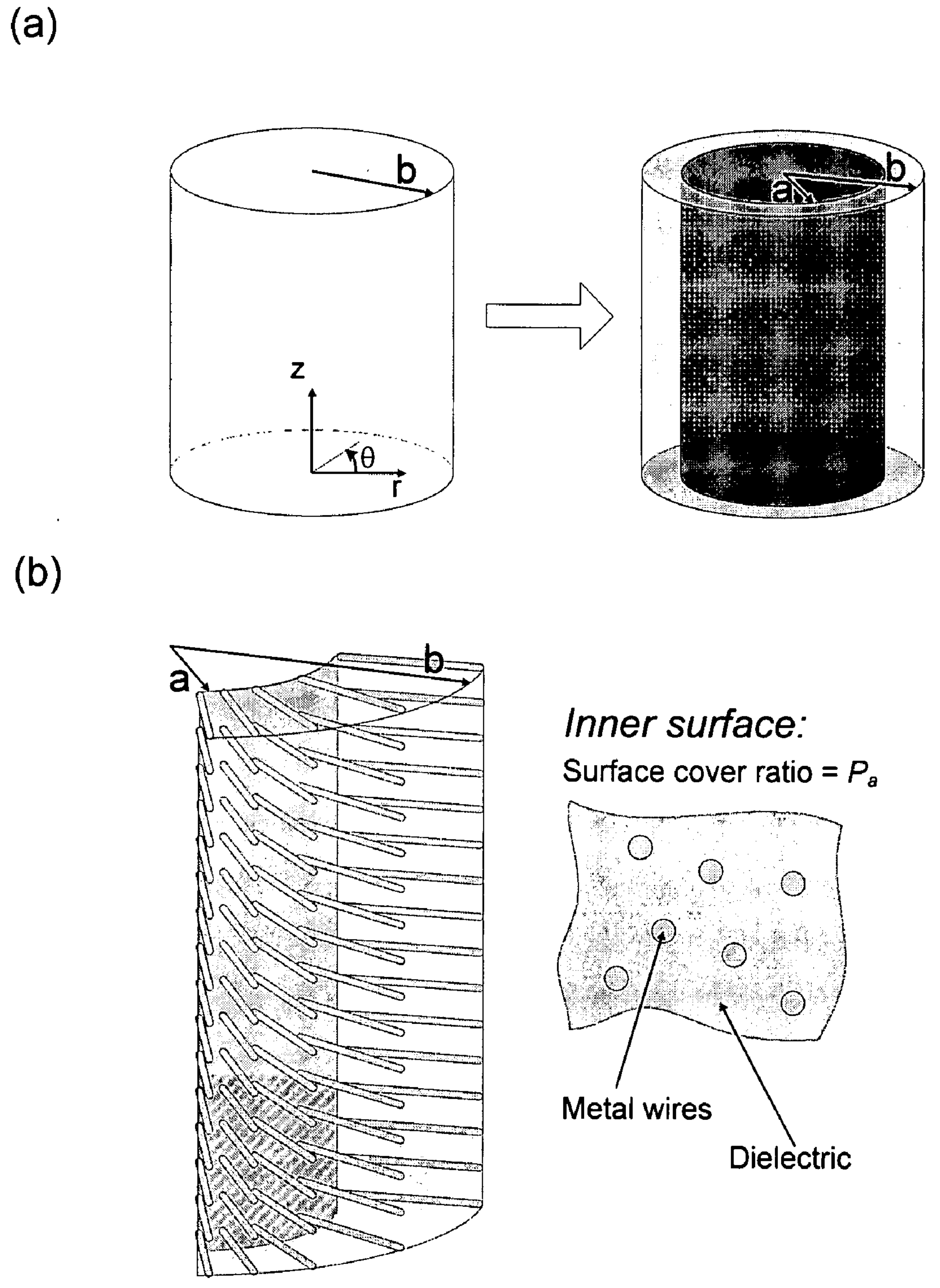
(a)
b
z
θ
r
a
b
(b)
a
b
Inner surface:
Surface cover ratio = P_a
Metal wires
Dielectric

FIG. 1

FIG. 2

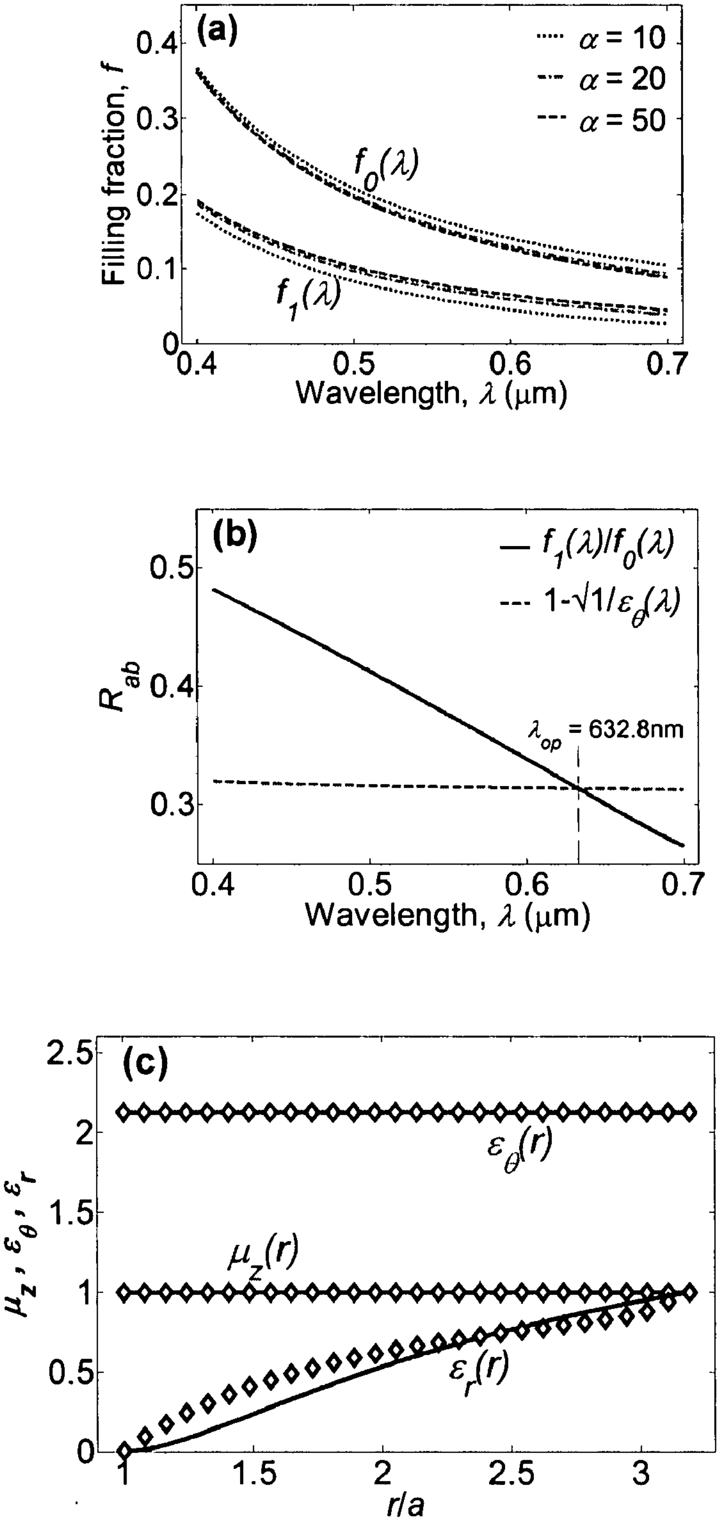
(a)
Filling fraction, f
0.4
0.3
0.2
0.1
0
0.4
0.5
0.6
0.7
Wavelength, λ (μm)
$\alpha = 10$
$\alpha = 20$
$\alpha = 50$
$f_0(\lambda)$
$f_1(\lambda)$
(b)
R_{ab}
0.5
0.4
0.3
0.4
0.5
0.6
0.7
Wavelength, λ (μm)
$f_1(\lambda)/f_0(\lambda)$
$1-\sqrt{1/\varepsilon_\theta(\lambda)}$
λ_{op} = 632.8nm
(c)
$\mu_z, \varepsilon_\theta, \varepsilon_r$
2.5
2
1.5
1
0.5
0
1
1.5
2
2.5
3
r/a
$\varepsilon_\theta(r)$
$\mu_z(r)$
$\varepsilon_r(r)$

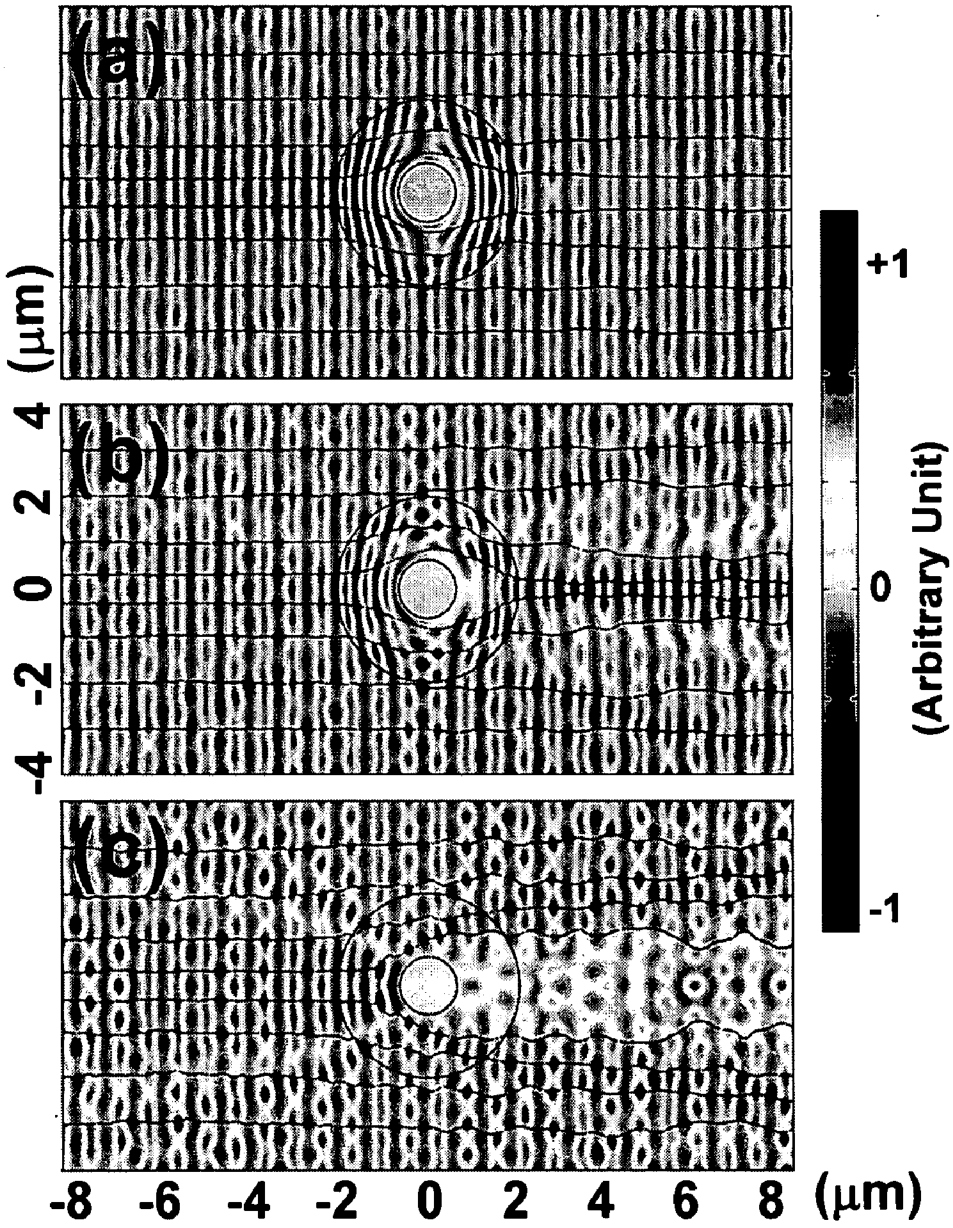

FIG. 3

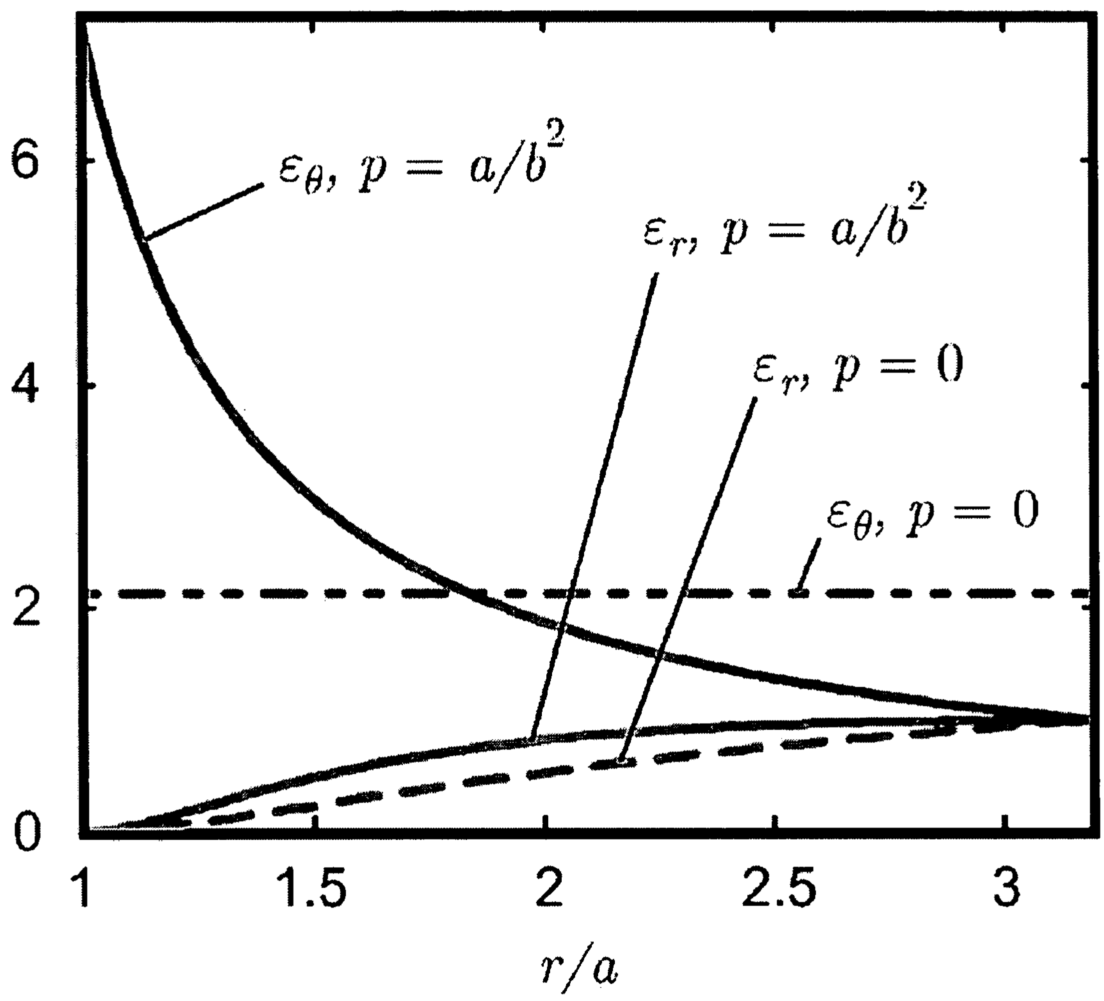
ε_θ, $p = a/b^2$
ε_r, $p = a/b^2$
ε_r, $p = 0$
ε_θ, $p = 0$
6
4
2
0
1
1.5
2
2.5
3
r/a

FIG. 4

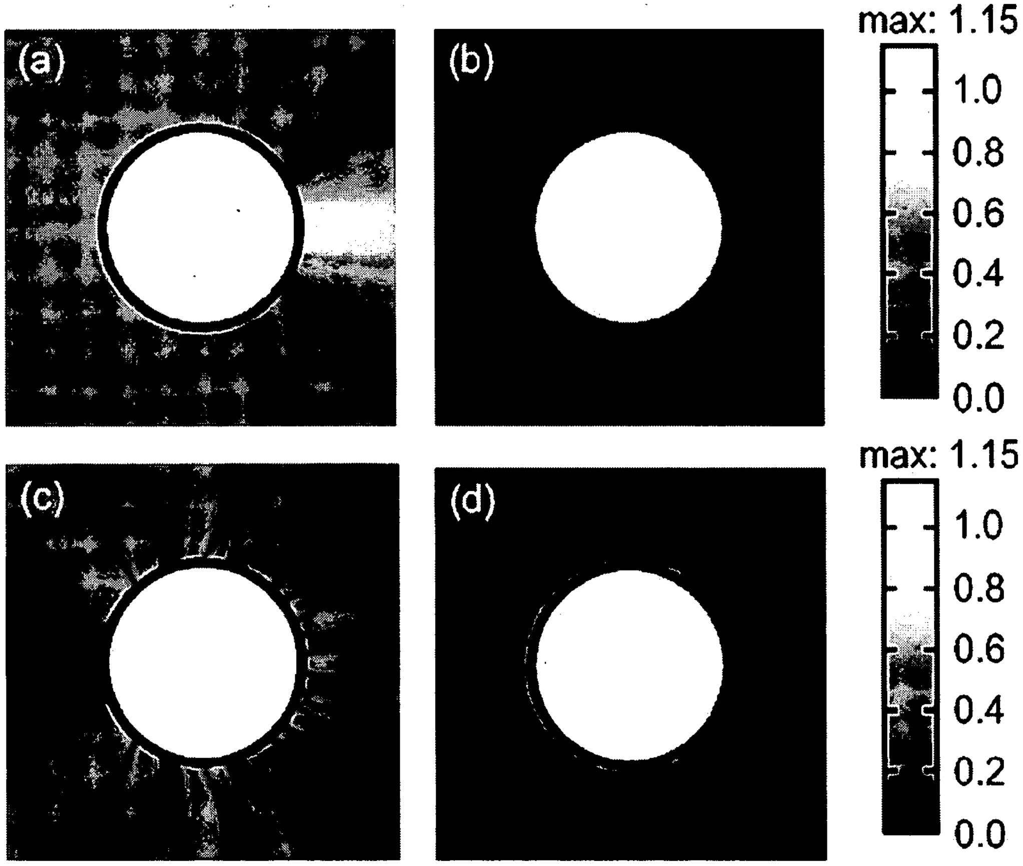

FIG. 5

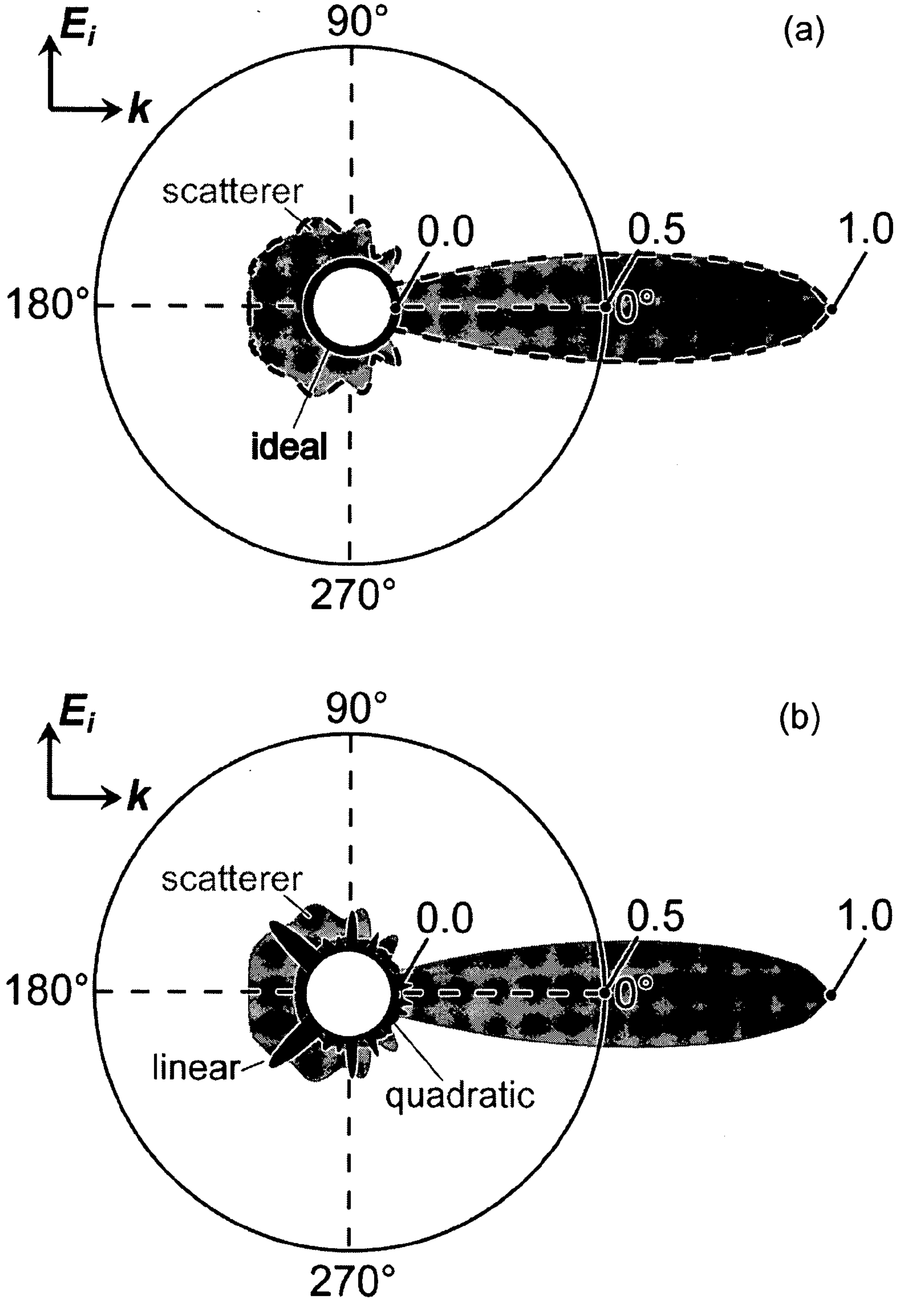
E_i
k
90°
(a)
scatterer
0.0
0.5
1.0
180°
0°
ideal
270°
E_i
k
90°
(b)
scatterer
0.0
0.5
1.0
180°
0°
linear
quadratic
270°

FIG. 6

SYSTEM, METHOD AND APPARATUS FOR CLOAKING

[0001] This application claims the benefit of U.S. provisional application Ser. No. 60/857,526, filed on Nov. 8, 2006, which is incorporated herein by reference.

STATEMENT OF GOVERNMENT SUPPORT

[0002] This application is based on research sponsored by the U.S. Army Research Office under 50342-PH-MUR.

TECHNICAL FIELD

[0003] This application relates to a system, method and apparatus for the modification of the scattering properties of an object. In particular, the effect may be achieved using non-magnetic materials.

BACKGROUND

[0004] An object may be sought to be made invisible at least over some frequency range. This has been termed a “cloak of invisibility”; the invisibility sought may be partial at a specific frequency, or over a band of frequencies, so the term “cloak of invisibility” or “cloak” may take on a variety of meanings. The cloak may be designed to decrease scattering (particularly “backscattering”) from an object contained within, while at the same time reducing the shadow (“forward scattering”), so that the combination of the cloak and the object contained therein have a resemblance to free space. When the phrase “cloaking”, “cloak of invisibility” or the like, is used herein, the effect is generally acknowledged to be imperfect, and the object may appear in a distorted or attenuated form, or the background obscured by the object may be distorted or partially obscured.

[0005] In an aspect the cloak has a similarity to “stealth” technology where the objective is to make the object as invisible as possible in the reflection or backscattering direction. One means of doing this is to match the impedance of the stealth material to that of the electromagnetic wave at the boundary, but where the material is strongly attenuating to the electromagnetic waves, so that the energy backscattered from the object within the stealth material is strongly attenuated on reflection, and there is minimal electromagnetic reflection at the boundary within the design frequency range. This is typically used in evading radar in military applications. Shadowing may not be a consideration in stealth technology.

[0006] The materials used for the cloak may have properties where, generally the permeability and permittivity tensors are anisotropic and where the magnitudes of the permeability and permittivity are less than one, so that the phase velocity of the electromagnetic energy being bent around the cloaking region is greater than that of the group velocity.

[0007] Materials having such properties have not been discovered as natural substances, but may be produced as artificial, man-made materials, where the permittivity and permeability are less than unity, and may be negative. Metamaterials, an extension of the concept of artificial dielectrics, were first designed in the 1940s for microwave frequencies. They typically consist of periodic geometric structures of a guest material embedded in a host material. Analogous to the circumstance where homogeneous dielectrics owe their properties to the nanometer-scale structure of atoms, metamaterials derive their properties from the sub-wavelength structure of its component materials. At wavelengths much longer than the unit-cell size, the structure can be assigned parameters that may be used describe homogeneous dielectrics, such as electric permittivity and refractive index.

[0008] A first experimental demonstration of a cloak operating over a narrow band of microwave frequencies was recently reported. Cloaking was achieved by varying the dimensions of a series of split ring resonators (SRRs) to yield a desired gradient of permeability in the radial direction. However, there appear to be limits to size scaling of SRRs so as to exhibit magnetic responses in the optical range. Replacing the SRRs with other optical magnetic structures like paired nano-rods or nano-strips may be difficult, primarily due to fabrication issues. Moreover, optical magnetism based on, for example, resonant plasmonic structures is usually associated with a high loss factor, which may be detrimental to the performance of cloaking devices.

SUMMARY

[0009] An apparatus is disclosed, the apparatus being a structure formed of a material having a permittivity less than unity and a permeability approximately equal to unity. The structural material may be a metamaterial: for example, a material having a permittivity greater than or equal to unity, partially filled with a material having a negative permittivity at a design wavelength. The permittivity of the material may be anisotropic.

[0010] In an aspect, the apparatus is sized and dimensioned such that the structure is disposable between an object and an observer.

[0011] A method of cloaking an object is described, the method including providing a structure formed of a metamaterial, and disposing the object to be cloaked such that the structure is positioned between the object and an observer. The metamaterial has a permittivity less than unity and a permittivity approximately equal to unity. The metamaterial may be fabricated from a dielectric with sub-wavelength inclusions of a material having a permittivity less than zero.

[0012] In another aspect, method of designing a cloaking structure includes, selecting a dielectric material; selecting a filler material having a permittivity less than zero; and, adjusting at least one of a distribution, a quantity or a dimension of the filler material within the dielectric material such that, at a design wavelength, a permittivity of a composite material comprised of the dielectric material and the filler material varies with a distance such that a visibility of an object disposed within the cloaking structure is reduced, and the region behind the object is visible.

[0013] In yet another aspect, a cloaked object, includes an object having a electromagnetic scattering property and a structure disposable between the object and an observer. The structure may be fabricated from a material having a permeability approximately equal to unity and a permittivity between approximately unity and approximately zero. The properties of the material may be selected so that a visibility of the object is reduced and a visibility of region behind the cloaked object may be substantially unaffected.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] FIG. **1**A illustrates the coordinate transformation that compresses a cylindrical region r<b into a concentric

cylindrical shell a<r<b; and, **1**B the structure of a non-magnetic optical cloak, together with a portion of the inner surface thereof;

[0015] FIG. **2**A is a graph of the functions f_0 and f_1 as defined by (8) for a silver-silica composite with aspect ratios of 10:1, 20:1 and 50:1, respectively; **2**B, the operational point determined by (11) for a cloak consisting of silver wires of $\alpha=10.7$ in silica; and, **2**C, the material parameters $\in_r$, $\in_\theta$ and μ_z of the cloak operating at $\lambda=632.8$ nm, where the solid line (—) represents the exact set of reduced parameters by Eq. 2, and the diamond (◇): parameters of the metal wire composite cloak; and

[0016] FIG. **3** is a cross-sectional view of finite-element simulations of the magnetic field mapping around the cloaked object with TM illumination at $\lambda=632.8$ nm when the object is surrounded by: A, a cloak with the exact set of reduced parameters; B, the metal wire composite cloak; and C, a vacuum (no cloak).

[0017] FIG. **4** shows a comparison of the anisotropic material properties for of the anisotropic material properties for the optimal quadratic design ($p=a/b^2$), solid lines and the linear transformation ($p=0$), dashed lines. The shape factor a/b is 0.31 and the diameter **2***b* is 4 microns.

[0018] FIG. **5** shows the results of full-wave field-mapping simulations of the magnitudes of the normalized scattered field for a metal cylinder with: A, no cloak; B, an ideal cloak; C, a linear transformation non-magnetic cloak; and D, an optimal quadratic transformation nonmagnetic cloak; and

[0019] FIG. **6** shows scattering patterns from the four cases shown in FIG. **5**; where A shows the metal cylinder scatterer with no cloak and the cylinder with the ideal cloak; and B shows the cylinder with a linear nonmagnetic cloak and the cylinder with an optimal quadratic nonmagnetic cloak.

DETAILED DESCRIPTION

[0020] Exemplary embodiments may be better understood with reference to the drawings, but these embodiments are not intended to be of a limiting nature.

[0021] When the phrase “cloaking”, “cloak of invisibility” or the like is used herein, the effect is generally acknowledged to be imperfect in practice, and the object may appear in a distorted or attenuated form, or the background obscured by the object may be distorted or partially obscured. Therefore, “cloak” should not be interpreted so as to require that the object within the cloak be “invisible” even at a design wavelength.

[0022] Reference may be made in this application to systems, apparatus, components, or techniques that are known, so as to enable a person of ordinary skill in the art to be able to comprehend the examples disclosed in the specification. The examples are intended to enable a person of ordinary skill in the art to practice the inventive concepts as claimed herein, using systems, apparatus, components, or techniques that may be known, disclosed herein, or hereafter developed, or combinations thereof. Where a comparison of performance is made between the examples disclosed herein and any known system, apparatus, component, or technique, such comparison is made solely to permit a person of skill in the art to more conveniently understand the present novel system, apparatus, component, or technique, and it should be understood that, in complex systems, various configurations may exist where the comparisons made may be better, worse, or substantially the same, without implying that such results are invariably obtained or constitute a limitation on the performance which may be obtained.

[0023] A non-magnetic cloak structure of cylindrical geometry which may be operable at visible wavelengths is described. The cloak may be designed to be effective at other wavelengths, and in other geometries.

[0024] In a first example, a cylindrical geometry is selected. A coordinate transformation may be used where a cylindrical region r<b is compressed into a concentric cylindrical shell a<r<b as shown in FIG. **1**A. There is no variation of permittivity or permeability along the z direction in the model geometry of this example.

[0025] Using a linear coordinate transformation results in the following properties for the anisotropic permittivity and permeability in a cloaking shell or structure:

$$\varepsilon_r = \mu_r = \frac{r-a}{r}, \qquad (1)$$

$$\varepsilon_\theta = \mu_\theta = \frac{r}{r-a},\ \varepsilon_z = \mu_z = \left(\frac{b}{b-a}\right)^2 \frac{r-a}{r}$$

[0026] Mathematically, there are a variety of coordinate transformations that may be used, and while a linear transformation is used in this example, the optimization of a quadratic coordinate transformation is described in a second example. Such mathematical examples are convenient for discussion, but more complex transformations may be used and, in conjunction with electromagnetic finite-element analysis, to design more complex structures having differing properties, and include the effects of lossy materials.

[0027] For transverse electromagnetic (TE) illumination of the cylindrical system with the incident electrical field polarized along the z axis, only ϵ_z, μ_r and μ_θ in (1) enter into Maxwell’s equations. The dispersion properties and wave trajectory in the cloaking shell remain the same as long as the values of the products $\epsilon_i\mu_j$ are maintained constant, where i and j represent any of the two distinct subscripts among r, θ and z.

[0028] For transverse magnetic (TM) illumination of the cylindrical system with the incident magnetic field polarized along the z axis is considered, only μ_z, ϵ_r and ϵ_θ may need to satisfy the requirements in (1), and the dispersion relations inside the cloak may remain unaffected as long as the product of $\mu_z\epsilon_r$ and $\mu_z\epsilon_\theta$ are maintained the same as the values determined by (1). Unlike the TE case, under TM illumination only one component of μ is of interest in the model. By multiplying ϵ_r and ϵ_θ by the value of μ_z, a reduced set of cloaking shell parameters is obtained:

$$\mu_z = 1 \qquad (2a)$$

$$\varepsilon_\theta = \left(\frac{b}{b-a}\right)^2 \qquad (2b)$$

$$\varepsilon_r = \left(\frac{b}{b-a}\right)^2 \left(\frac{r-a}{r}\right)^2 \qquad (2c)$$

[0029] By the process of normalizing the parameters to the permeability such that $\mu_z=1$, a material that does not exhibit magnetic properties within the effective frequency regime of the cloak may be used. The reduced set of parameters of (2) results in the same electromagnetic wave trajectories as for materials meeting the requirements of (1).

[0030] The ideal parameters in (1) result in a perfectly-matched impedance of $Z=\sqrt{\mu_z/\epsilon_\theta}=1$ at r=b, while the reduced set in (2) produces an impedance at the outer boundary of $Z=1-R_{ab}$, where R_{ab}=a/b denotes the ratio between the inner and outer radii. R_{ab} may be termed the "shape factor" of the cylindrical structure. The level of power reflection or back-scattering due to using reduced parameters in the design with a linear transformation can be estimated as $|(1-Z)/(1+Z)|^2=[R_{ab}/(2-R_{ab})]^2$.

[0031] The azimuthal permittivity ϵ_θ is a constant with a value larger than 1, which can be achieved with the usual dielectric materials, although specially designed materials are not intended to be excluded. The cylindrical shell may be constructed with a desired radial distribution of ϵ_r which may vary from 0 at the inner boundary (r=a) of the cloak to 1 at the outer surface (r=b).

[0032] The effect of using a material meeting the electromagnetic and spatial requirements of (2) is to guide incident electromagnetic waves such that they are excluded from the region interior to r=a, and which exit the cloaking region with minimal disturbance to the originally incident ray paths. As such, the background behind the object may appear to be substantially undisturbed by the presence of the object being cloaked.

[0033] Artificial dielectrics, such as metamaterials, with a positive permittivity ϵ_r less than unity are known. In this example, the characteristics of ϵ_r may be realized, for example, by using metal wires of sub-wavelength size disposed in a radial direction and embedded in a dielectric material, as shown FIG. **1**B. The wires may be disposed perpendicular to the cylinder inner and outer surfaces. The spatial positions may not be periodic and may be random. For large cloaks, the wires may be broken into smaller pieces. The aspect ratio of the metal wires, defined by the ratio of the length to the radius of the wire, is denoted by α. The whole structure of the cloaking system may conceptually resemble a round hair brush (except that the "bristles" of such a "hair brush" may consist of disconnected smaller pieces, with either random or periodic distribution within the cylinder).

[0034] The metal material may be chosen so as to have a negative permittivity in the wavelength regime chose for the design. Metals from the noble metals such as gold, silver, tantalum, platinum, palladium or rhodium may be used. Other materials are known to exhibit negative permittivity, such as silicon carbide. These materials may be combined with dielectric materials having a permittivity greater than unity, so as to result in a metamaterial with an effective permittivity between about zero and unity.

[0035] The shape-dependent electromagnetic response of a sub-wavelength particle can be characterized by the Lorentz depolarization factor q. For an ellipsoid of semiaxes a_i, a_j and a_k with electric field polarized along a_i, the depolarization factor may be expressed by:

$$q_i=\int_0^\infty \frac{a_i a_j a_k ds}{2(s+a_i^2)^{3/2}(s+a_j^2)^{1/2}(s+a_k^2)^{1/2}} \tag{3}$$

Another commonly used parameter, the screening factor κ of a particle, is related to q by κ=(1−q)/q. A long wire with large aspect ratio α results in a small depolarization factor and a large screening factor, which generally indicates strong interactions between the electromagnetic fields and the wire.

[0036] For a composite cloak with metal wires as inclusions in a dielectric, the electromagnetic properties may be described by "shape-dependent" effective-medium theory (EMT) that describes composites with particles of different shapes and thus different κ-factors. The effective permittivity ϵ_{eff} for a composite material comprising metal particles with permittivity ϵ_m a volume filling factor f and screening factor κ, along with a dielectric component with permittivity ϵ_d and a filling factor $1-f$ is given by:

$$f\frac{\varepsilon_m-\varepsilon_{eff}}{\varepsilon_m+\kappa\varepsilon_{eff}}+(1-f)\frac{\varepsilon_d-\varepsilon_{eff}}{\varepsilon_d+\kappa\varepsilon_{eff}}=0 \tag{4}$$

For spherical particles with q=⅓ and κ=2, (4) reduces to the common EMT expression which is a quadratic equation with the following solutions:

$$\varepsilon_{eff}=\frac{1}{2\kappa}\left\{\bar{\varepsilon}\pm\sqrt{\bar{\varepsilon}^2+4\kappa\varepsilon_m\varepsilon_d}\right\} \tag{5}$$

where $\bar{\epsilon}=[(\kappa+1)f-1]\epsilon_m+[\kappa-(\kappa+1)f]\epsilon_d$. The sign in (5) may be chosen such that $\epsilon_{eff}>0$.

[0037] When using metal wires in a composite cloak, the radial permittivity ϵ_r determined by (5) may exhibit a positive value less than 1 with a minimal imaginary part. Metamaterials having a metallic component may exhibit a permittivity that differs from unity, while having a permeability close to unity, and such material may be termed substantially non-magnetic so as to suggest that the permittivity characteristics are more important in achieving the cloaking.

[0038] For the structure in FIG. **1**B, the volume filling fraction is inversely proportional to r. The filling fraction in the EMT formula for calculating ϵ_r using (5) may be $f(r)=P_a\cdot(a/r)$, with P_a being the surface cover ratio of metal at the inner surface of the cloak (r=a). The filling fractions f at the inner and outer surface of the cloak are P_a and $P_a\cdot(a/b)$ respectively, and the overall metal filling fraction in the whole cloak layer is $P_a\cdot 2a/(a+b)$. The azimuthal permittivity ϵ_θ inside the cloak is substantially the same as that of the dielectric material because a response of wires to the angular electrical field E_θ oriented normally to the wires is small and, at low metal filling factors, it may generally be neglected.

[0039] The reduced set of cloak parameters in (2) is consistent with a smooth variation of the radial permittivity from 0 to 1 as r varies from a to b. That is,

$$\begin{cases}\varepsilon_{eff,r}(P_a)=0\\ \varepsilon_{eff,r}(P_a\cdot a/b)=1\end{cases} \tag{6}$$

[0040] The gradient in $\epsilon_{eff,r}$ may follow the function described in (2c) such that

$$\varepsilon_{eff,r}(P_a\cdot a/r)=\left(\frac{b}{b-a}\right)^2\left(\frac{r-a}{r}\right)^2 \tag{7}$$

[0041] In an actual design, $\epsilon_{eff,r}$ may have some deviation from the value given by (2c) inside the cloak. For example, in the reported microwave cloak with a layered structure, the desired permeability was fulfilled at only a few discrete positions along the radial direction, while in the majority of the cloak the material was air. At the inner and outer surfaces of the cloak, the conditions of (6) should be satisfied as closely as practical, although the conditions at the inner surface may be relaxed where there is a lossy component to the material. This may result in impedance index matching at r=b and minimal leaking energy at r=a.

[0042] To determine all the parameters of the design shown in FIG. **1**B, the general properties of a metal-dielectric composite with thin metal wires in the radial direction as the inclusion are modeled. Two filling fraction functions $f_0(\lambda)$ and $f_1(\lambda)$ are such that for given constituent composite materials and for a fixed aspect ratio α of the wires, the effective radial permittivity is.

$$\epsilon_{eff,r}(\lambda, f_0(\lambda))=0 \tag{8a}$$

and

$$\epsilon_{eff,r}(\lambda, f_1(\lambda))=1 \tag{8b}$$

The values of $f_0(\lambda)$ and $f_1(\lambda)$ calculated from (5) and (8) for a silver-silica composite with α=10, 20 and 50 at visible wavelengths are plotted in FIG. **2**A. In the calculation the metal permittivity is approximated by the Drude model, and the permittivity of the dielectric is calculated using Sellmeier equation.

[0043] Combining (6) and (8), at the design wavelength λ,

$$\begin{cases} f_0(\lambda) = P_a \\ f_1(\lambda) = P_a \cdot a/b \end{cases} \tag{9}$$

[0044] From the above equations R_{ab} may be expressed as:

$$R_{ab}=f_1(\lambda)/f_0(\lambda) \tag{10}$$

[0045] Using (10) with (2b), the operating condition of the cloak is obtained as:

$$f_1(\lambda)/f_0(\lambda)=1-\sqrt{1/\epsilon_\theta(\lambda)}, \tag{11}$$

where $\epsilon_\theta(\lambda)$ is the permittivity of the dielectric material surrounding the metal wires in the cloak. Thus, the geometrical factors of the cloak including R_{ab}, P_a and α are determined. The same design may work for similar cylindrical cloaks with the same shape factor R_{ab}. In FIG. **2**B we show the operational point obtained by (11) is shown for a cloak consisting of silver wires with α=10.7 in silica.

[0046] A cloaking device or structure may be designed for operating at an operational wavelength λ_{op}. A method of designing a cloaking device may include the steps of: choosing materials that are available for the metal wires and the surrounding dielectric, or other metamaterials; and, calculating the values of f_0 and f_1 as functions of the aspect ratio α at λ_{op} using, for example, the EMT model in (5). Other models may also be used. The desired aspect ratio for λ_{op} corresponds to the fulfillment of (12).

$$\varepsilon_\theta(\lambda_{op}) = \left(\frac{f_0(\lambda_{op}, \alpha)}{f_0(\lambda_{op}, \alpha) - f_1(\lambda_{op}, \alpha)}\right)^2 \tag{12}$$

Then, the structure of the cloak can be determined from (9) and (10).

[0047] The "round brush" design for a non-magnetic cloak may permit constructing an appropriate device operating at desired wavelength, which may be an optical wavelength, by choosing the proper materials and structures. As an example, the design of an optical cloak operating at the frequently-used wavelength of 632.8 nm (He—Ne laser), and consisting of silver and silica is described.

[0048] Expressions (5), (8), and (12) yield the aspect ratio α=10.7, and the volume filling fractions at the two boundaries are f_0=0.125 and f_1=0.039, respectively. From (9) and (10) the shape factor of the cylindrical cloak is Rab=0.314 while the surface cover ratio at the inner boundary is P_a=12.5%. The effective parameters of μ_z, ϵ_r and ϵ_θ from for this design and the set of reduced parameters from (2) are shown in FIG. **2**C. As seen in FIG. **2**C, μ_z and ϵ_θ match the theoretical requirements throughout the cylindrical cloak. In this example, the radial permittivity ϵ_r fits the values required by (2) exactly at the two boundaries of the cloak, and follows the overall tendency relatively well inside the cloak. The imaginary part of ϵr is almost zero at r=b where $\epsilon_r'(\lambda_{op}, b)=1$ and reaches around 0.6 at the inner surface where $\epsilon_r'(\lambda_{op}, a)=0$. These quantities are similar to reported low-index metamaterials with periodic-metal-wire arrays.

[0049] The effects of loss can be addressed in several ways. As an example, if the aspect ratio of the wires is varied along the radial direction, the imaginary part of ϵ_r may be smaller than 0.1 throughout the cloak. It may possible to compensate the loss by using a gain medium.

[0050] To illustrate the performance of the non-magnetic optical cloak with a design corresponding to FIG. **2**C, a finite element method simulation using the commercial finite element package COMSOL Multiphysics (available from COMSOL, Inc. Burlington, Mass.) was performed. An ideal metallic cylinder with radius r=a is disposed within the cloaked region. The simulated results of magnetic field distribution around the cloaked object together with the power flow lines are illustrated in FIG. **3** for three cases. As shown in FIG. **3**A, the cloak with the reduced set of material parameters represented by the solid curves in FIG. **2**C leads to a small perturbation of the external fields, which is limited by imperfect impedance matching. FIG. **3**B corresponds to a cloak with parameters given by the diamond shaped markers in FIG. **2**C. Comparing the field maps in FIGS. **3**A and **3**B, the simulations for the designed cloak are in good agreement with the exact case. Without the cloak (FIG. **3**C), the waves around the object are severely distorted and a clear shadow is cast behind the cylinder. These simulations show the capability of the structure described in reducing the scattering from the object inside the cloaked region.

[0051] The achievable invisibility with the example designed cloak is not perfect due to impedance mismatch associated with the reduced material specifications and the energy loss in a metal-dielectric structure.

[0052] In another aspect, a cloaking device can be based on vertical metal strips (instead of rods) placed in the radial directions within the cloaking structure. These strips can also be randomly or periodically disposed and may also consist of disconnected smaller strips. Chains of metal particles of various shapes may also be used.

[0053] In a second example, the coordinate transformation applied to (1) may be a high-order transformation such as a quadratic instead of the linear transformation. This example is one of a number of different analytic coordinate transformations which may be employed and illustrates a particular situation where the impedance is matched at the boundary between the cloaking cylinder and the assumed free space propagation medium of the incident electromagnetic wave.

[0054] By constraining the impedance of the outer boundary to be equal to that of the external propagation medium, an optimal quadratic coordinate transformation may be obtained as:

$$r=[(a/b((r'/b-2)+1]r'+a \tag{13}$$

[0055] The shape factor a/b should be less than 0.5 in order to have a monotonic transformation. When evaluating the material properties at the outer boundary, r=b, the material parameters ϵ_r, ϵ_θ, μ_r, are each equal to unity. As such, the impedance mismatch at the boundary has been obviated for the case of reduced parameters.

[0056] FIG. **4** compares the material properties of the nonmagnetic cloak structures of the first and the second examples, where the shape factor a/b is 0.31, and the diameter (**2***b*) is 4 micrometers and the wavelength λ=632.8 nanometers. The object disposed inside of the cloaks is an ideal metal cylinder with a radius that is the same as the inner surface; that is r=a. This may be a worst-case condition, and an arbitrarily shaped object of any configuration may be disposed inside of the inner surface with the same, similar, or better results.

[0057] The results were computed using the same finite-element software package as used for the first example, and the normalized magnitudes of the scattered fields are shown in FIG. **5**.

[0058] The scattered field from the cloaked metallic cylinder is itself is shown in FIG. **5**A. The strong forward scattering observed at the right-hand side of the diagram corresponds to a shadow cast behind the object. An idealized cloak is shown in FIG. **5**B, where the scattered field would be essentially zero in magnitude in all directions in the plane. These examples may be compared with the results obtained for the linear transformation (FIG. **5**C) and the quadratic case (FIG. **5**D). The linear case exhibits a scattering pattern from the outer boundary of the system, primarily due to the impedance mismatch. On the other hand, the quadratic transformation function results in substantially less scattering from the cloaking system. A figure of merit for cloaking, which may be defined as the ratio of the scattering cross sections with and without the cloak, is about 10 for quadratic cloak of the dimensions modeled, and it increases as the size of the cloaking system increases.

[0059] FIGS. **6** A and B. show the scattering radiation patterns corresponding to the four cases of FIG. **5**. The curves in FIG. **6** show the energy flow in the radial direction normalized by the maximum value in the noncloaked case at a boundary outside the outer surface of the cloak structures. In the ideal cloaking system, the scattering energy flow is zero, which is indicated by the solid inner circle in FIG. **6**A.

[0060] FIG. **6**B shows that the linear cloak structure with reduced parameters exhibits a noticeable although smaller scattering, and has and strongly directional scattering pattern. However, for the nonmagnetic quadratic cloak, the overall scattering is much less significant. The peak value of the radial Poynting vector in the quadratic cloak is more than six times smaller than that of the linear example. Moreover, the directivity in the scattering pattern is substantially suppressed.

[0061] In another aspect, a cloaking device or structure may be a spherical or other shaped cloaking structure. The specific geometrical shape, the size and other design parameters of the structure, such as the spatial variation of permittivity, may be chosen using the general approach described herein so as to be adaptable to the wavelength, the degree of cloaking, and the properties of the object to be cloaked. Loss and gain may be introduced in various portions of the structure.

[0062] The examples shown herein have used analytic profiles for the material properties so as to illustrate certain of the principles which may influence design of cloaking structures. However, since electromagnetic simulations using finite element methods are commonly used in design of complex shapes, and have been shown herein to yield plausible results, the use of such simulations are envisaged as useful in design.

[0063] Certain aspects, advantages, and novel features of the claimed invention have been described herein. It would be understood by a person of skill in the art that not all advantages may be achieved in practicing a specific embodiment. The claimed invention may be embodied or carried out in a manner that achieves or optimizes one advantage or group of advantages as taught herein without necessarily achieving other advantages as may have been taught or suggested.

[0064] It is therefore intended that the foregoing detailed description be regarded as illustrative rather than limiting, and that it be understood that it is the following claims, including all equivalents, that are intended to define the spirit and scope of this invention.

What is claimed is:

1. A cloaking apparatus, comprising:

a structure formed of a material having a permittivity less than unity and a permeability approximately equal to unity,

wherein the material is a metamaterial.

2. The apparatus of claim **1**, wherein the structure is disposable between an object and an observer.

3. The apparatus of claim **1**, wherein the structure has a first surface and a second surface such that an object is disposable so that the second surface lies between the object and the first surface.

4. The apparatus of claim **3**, wherein the permittivity of the metamaterial varies between approximately zero at an interface with the void and approximately unity at an outer radius of the cylinder.

5. The apparatus of claim **3**, wherein the properties of the metamaterial are selected so that the power flow crossing the second surface in the direction of an object to be cloaked is minimized.

6. The apparatus of claim **3**, wherein the properties of the metamaterial are selected so that the backscattering coefficient and the forward scattering coefficient of a combination of the structure and an object disposed such that the first and second surfaces are disposed between the object and a source of electromagnetic radiation are reduced.

7. The apparatus of claim **3**, wherein the properties of the metamaterial include a variation of permittivity with distance between the first and the second surface.

8. The apparatus of claim **3**, wherein the metamaterial has permittivity that is a function of the distance between the first surface and the second surface, and the variation of the permittivity is selected so that the impedance of the metamaterial at the first surface is the substantially the same as the impedance of a region outside of the structure and abutting the first surface.

9. The apparatus of claim **1**, wherein the structure has at least one axis of symmetry.

10. The apparatus of claim **1**, wherein the structure has a cylindrical symmetry, and an inner void.

11. The apparatus of claim **10**, wherein an object is disposable in the inner void.

12. The apparatus of claim **10**, wherein the permittivity of the metamaterial varies in a radial direction.

13. The apparatus of claim **1**, wherein the permittivity is anisotropic.

14. The apparatus of claim **1**, wherein the metamaterial comprises a dielectric material having a permittivity greater than or equal to unity partially filled with a material having a negative permittivity at a design wavelength.

15. The apparatus of claim **14**, wherein the filler material is a plurality of sub-wavelength sized structures.

16. The apparatus of claim **14**, wherein the material having a negative permittivity is a metal.

17. The apparatus of claim **16**, wherein the metal is selected from a noble metal group.

18. The apparatus of claim **1**, wherein the properties of the metamaterial are selected so that an impedance discontinuity between the metamaterial and an external environment is minimized.

19. The apparatus of claim **18**, wherein a refractive index of the external environment is unity.

20. The apparatus of claim **1**, wherein the properties of the metamaterial are selected so that the backscattering coefficient and the forward scattering coefficient of the structure are reduced.

21. The apparatus of claim **1**, wherein the rate of change of an outer geometrical shape of the structure is slow when compared with a value of a design wavelength.

22. The apparatus of claim **1**, wherein the metamaterial is lossless.

23. The apparatus of claim **1**, wherein the metamaterial has a permeability of unity.

24. The apparatus of claim **1**, wherein the properties of the metamaterial are selected so that a cloaking effect is a maximum at a visible wavelength of light.

25. The apparatus of claim **1**, wherein the metamaterial includes a gain medium material.

26. A method of cloaking an object, the method comprising:

providing a structure formed of a metamaterial; and

disposing the object to be cloaked such that the structure is positioned between the object and an observer,

wherein the metamaterial has a permittivity less than unity and a permittivity approximately equal to unity.

27. The method of claim **16**, wherein the metamaterial is a dielectric with sub-wavelength inclusions of a material having a permittivity less than zero.

28. The method of claim **26**, wherein the object is surrounded by the structure.

29. A method of designing a cloaking structure, the method comprising:

selecting a dielectric material;

selecting a filler material having a permittivity less than zero; and

adjusting at least one of a distribution, a quantity or a dimension of the filler material within the dielectric material such that, at a design wavelength, a permittivity of a composite material comprised of the dielectric material and the filler material varies with a distance such that a visibility of an object disposed within the cloaking structure is reduced, and the region behind the object is visible.

30. The method of claim **29**, where the energy reflected from a surface of the cloaking structure is substantially less than the energy that reflectable by the object without the cloaking structure.

31. The method of claim **29**, wherein an outer geometrical shape of the cloaking structure is slowly varying with respect to the design wavelength.

32. The method of claim **29**, wherein the composite material further includes a gain medium.

33. A cloaked object, the object comprising:

an object having a electromagnetic scattering property; and

a structure disposable between the object and an observer, wherein the structure further comprises:

a material having a permeability approximately equal to unity and a permittivity between approximately unity and approximately zero.

34. The cloaked object of claim **33**, wherein the properties of the material are selected so that a visibility of the object is reduced and a visibility of region behind the cloaked object is substantially unaffected.

35. The cloaked object of claim **33**, wherein the properties of the metamaterial include the variation of permittivity as a function of a location within the structure.

36. The cloaked object of claim **33**, wherein a design wavelength is within a visible light spectrum.

37. The cloaked object of claim **33**, wherein the material further includes a gain medium.

38. The cloaked object of claim **33**, where an outer geometrical shape of the structure varies slowly compared with a value of a design wavelength.

* * * * *

US 20100110559A1

(19) **United States**
(12) **Patent Application Publication**
Cai et al.

(10) **Pub. No.: US 2010/0110559 A1**
(43) **Pub. Date: May 6, 2010**

(54) **SYSTEM, METHOD AND APPARATUS FOR CLOAKING**

(76) Inventors: **Wenshan Cai**, Sunnyvale, CA (US); **Vladimir M. Shalaev**, West Lafayette, IN (US); **Uday K. Chettiar**, Philadelphia, PA (US); **Alexander V. Kildishev**, West Lafayette, IN (US)

Correspondence Address:
BRINKS HOFER GILSON & LIONE
P.O. BOX 10395
CHICAGO, IL 60610 (US)

(21) Appl. No.: **12/573,610**

(22) Filed: **Oct. 5, 2009**

Related U.S. Application Data

(60) Provisional application No. 61/103,025, filed on Oct. 6, 2008.

Publication Classification

(51) **Int. Cl.**

G02B 3/00	(2006.01)
G06F 17/50	(2006.01)

(52) **U.S. Cl.** **359/642**; 703/1

(57) **ABSTRACT**

An apparatus and method of cloaking is described. An object to be cloaked is disposed such that the cloaking apparatus is between the object and an observer. The appearance of the object is altered and, in the limit, the object cannot be observed, and the background appears unobstructed. The cloak is formed of a metamaterial where the properties of the metamaterial are varied as a function of distance from the cloak interfaces. The metamaterial may be fabricated as a composite material having a dielectric component and inclusions of particles of sub-wavelength size, and may also include a gain medium.

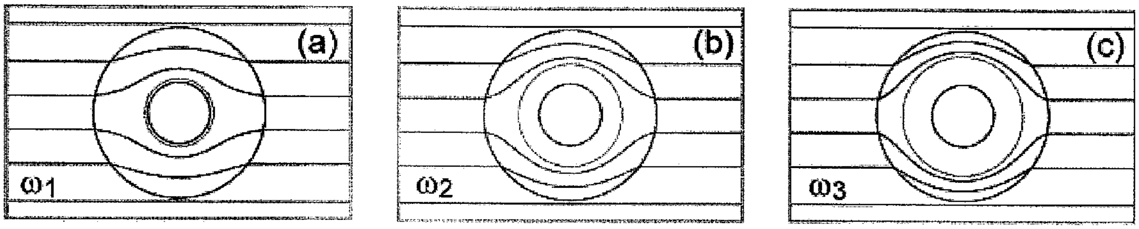

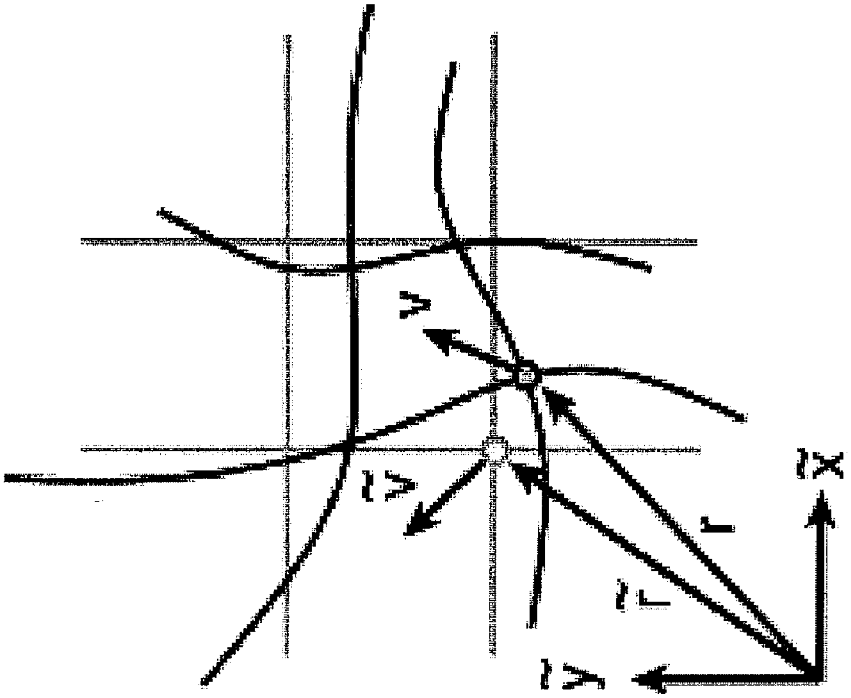
v
$\tilde{v}$
r
$\tilde{r}$
$\tilde{x}$
$\tilde{y}$

FIG. 1

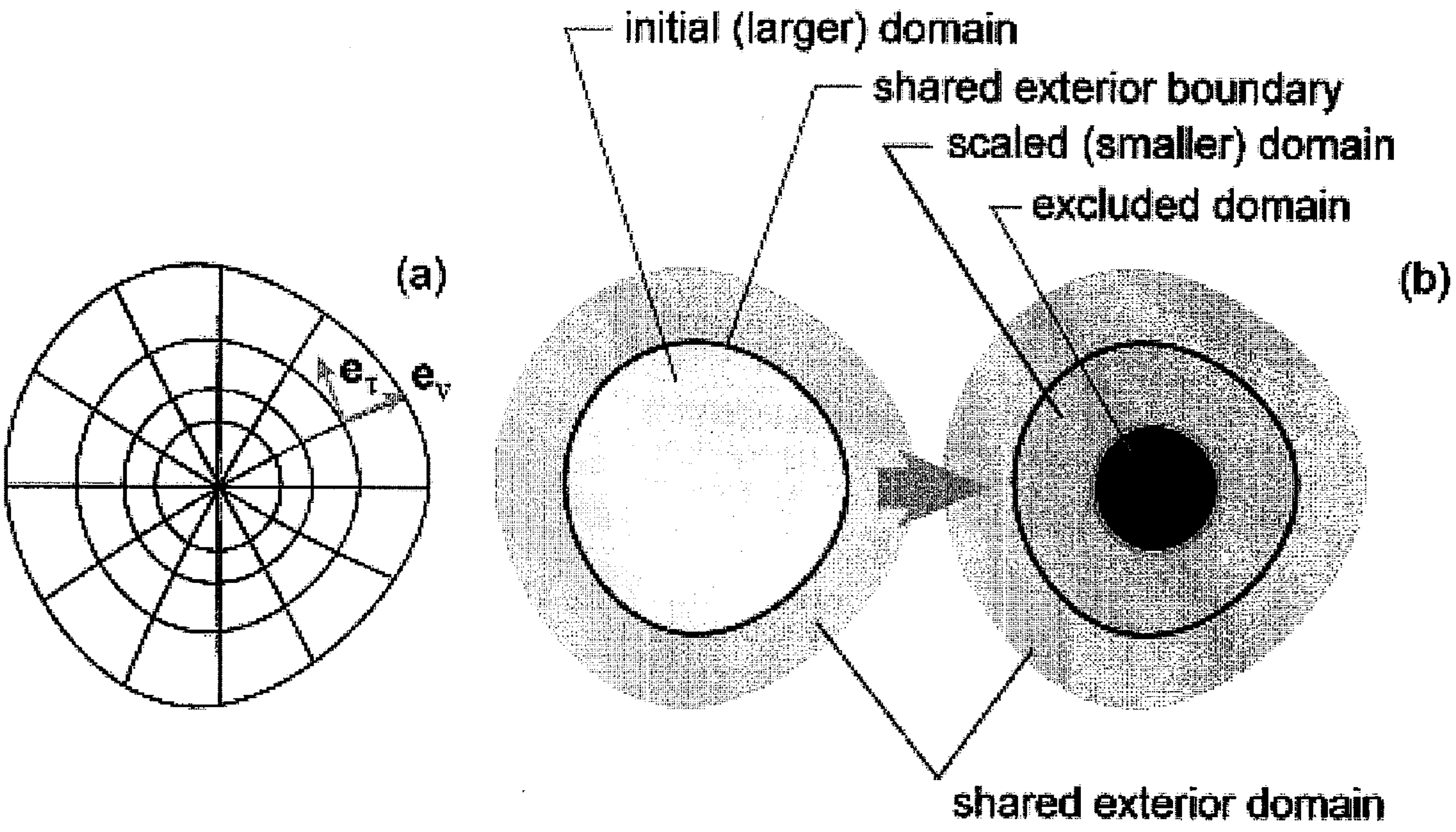
(a)
e_τ
e_v
initial (larger) domain
shared exterior boundary
scaled (smaller) domain
excluded domain
(b)
shared exterior domain

FIG. 2

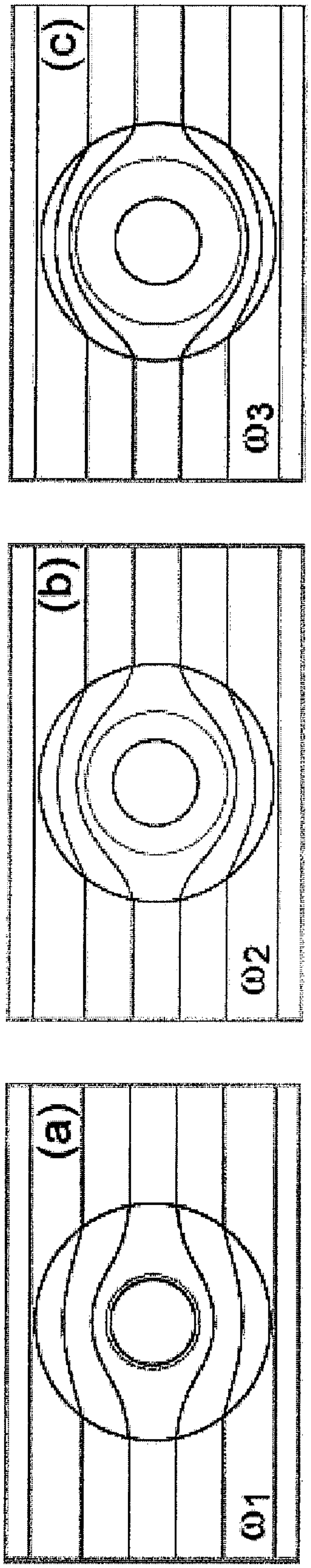

FIG. 3

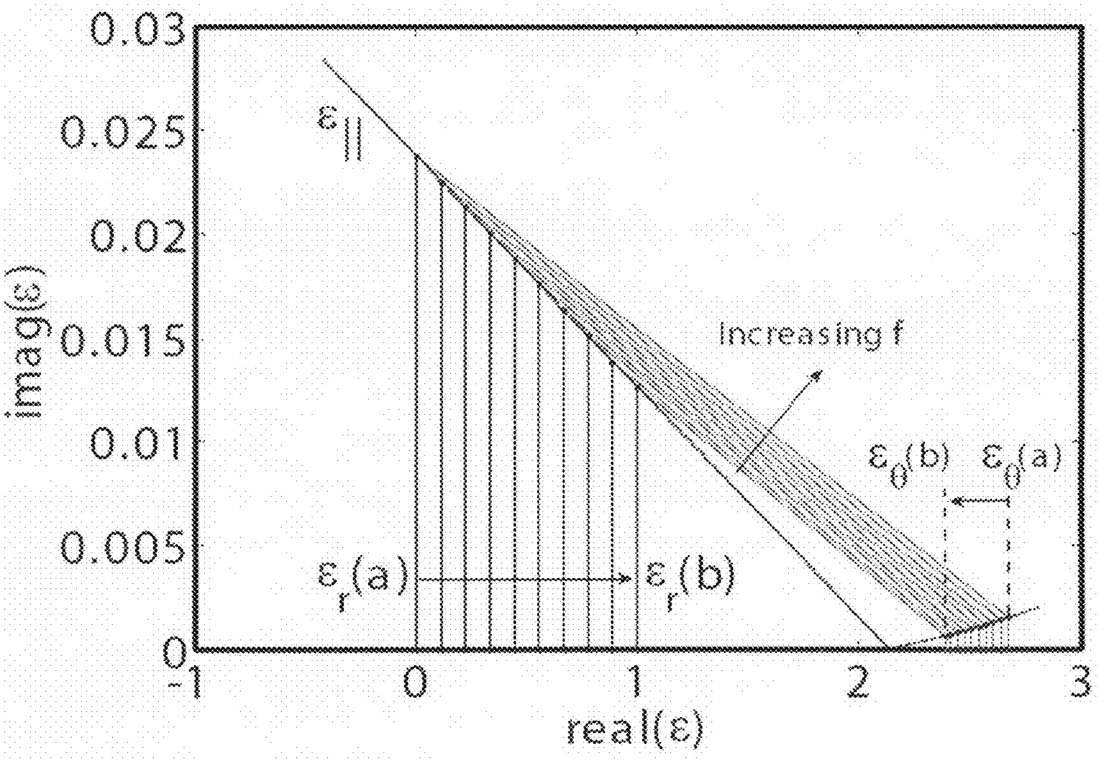
0.03
0.025
0.02
0.015
0.01
0.005
0
imag(ε)
-1
0
1
2
3
real(ε)
$\varepsilon_{||}$
Increasing f
$\varepsilon_0(b)$
$\varepsilon_0(a)$
$\varepsilon_r(a)$
$\varepsilon_r(b)$

FIG. 4

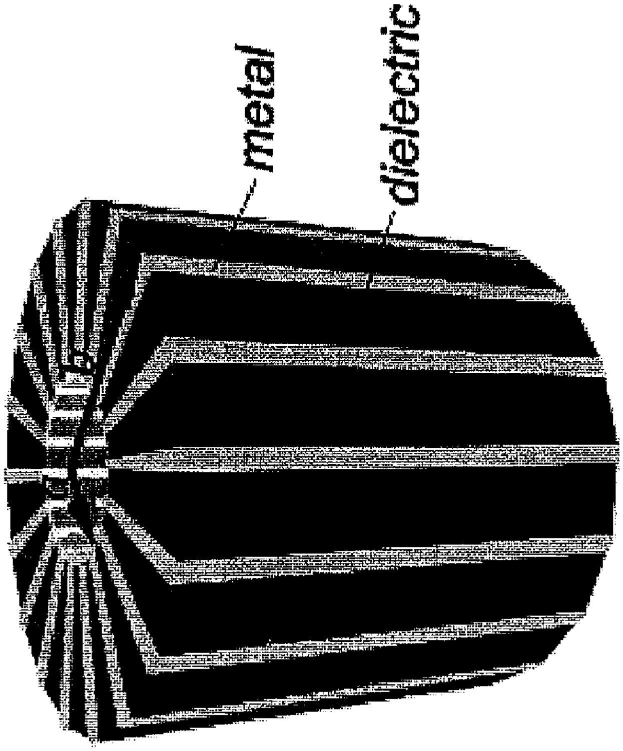
metal
dielectric

FIG. 5

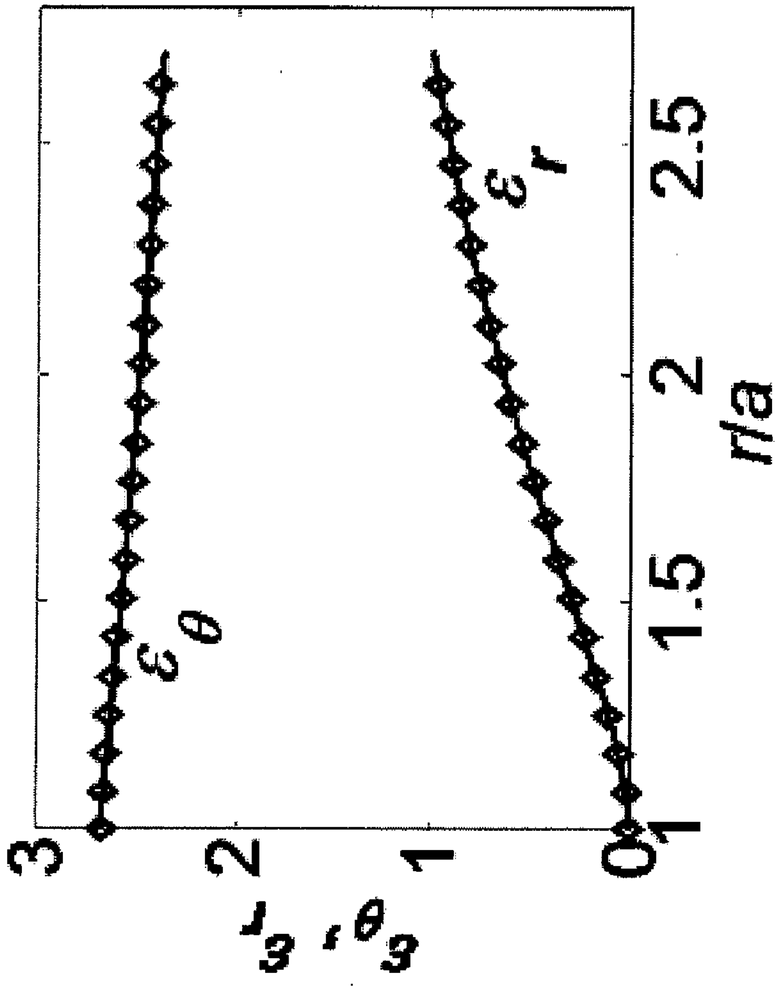
ε_θ, ε_r
3
2
1
0
1
1.5
2
2.5
r/a
ε_θ
ε_r

FIG. 6

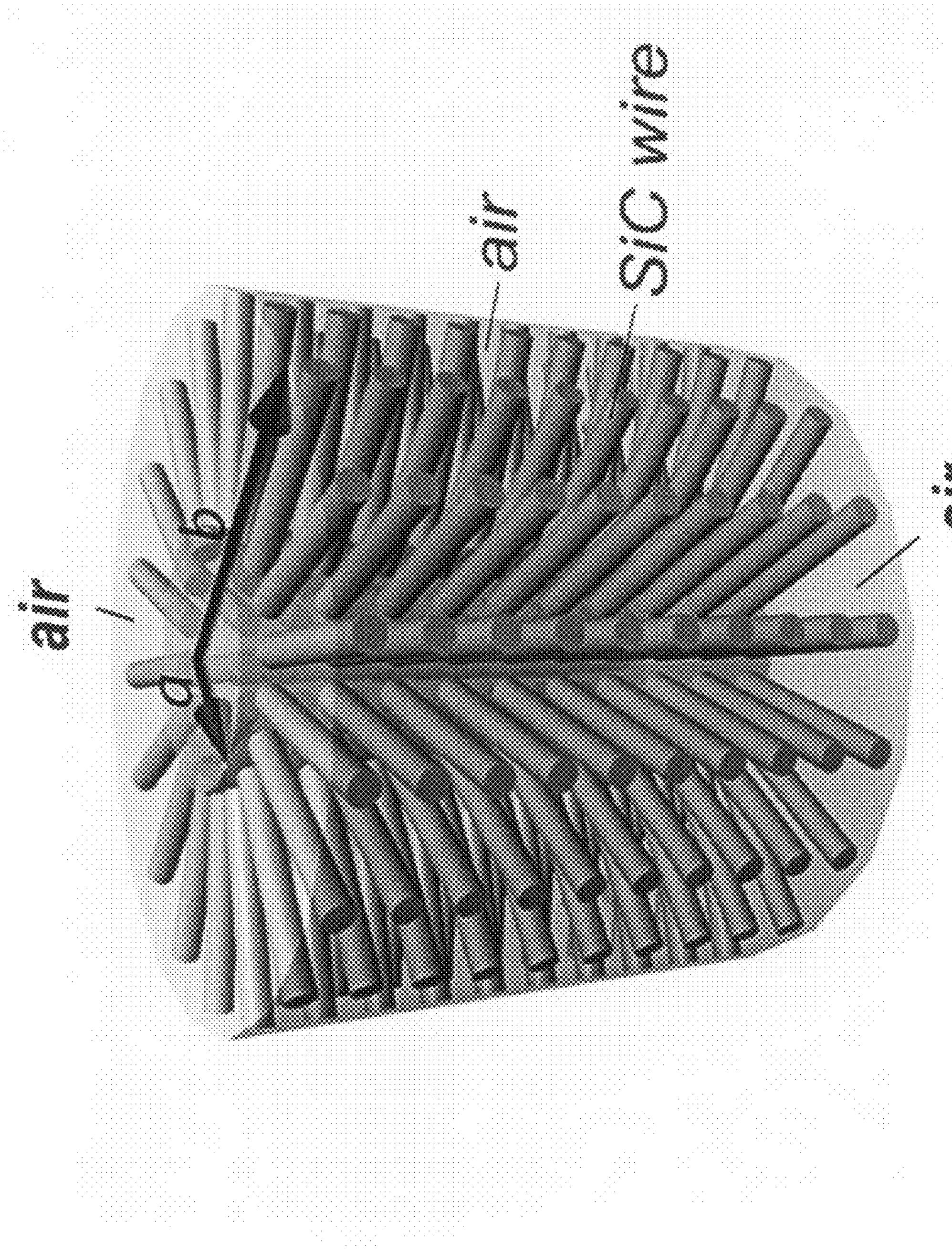
air
d
b
air
SiC wire
air

FIG. 7

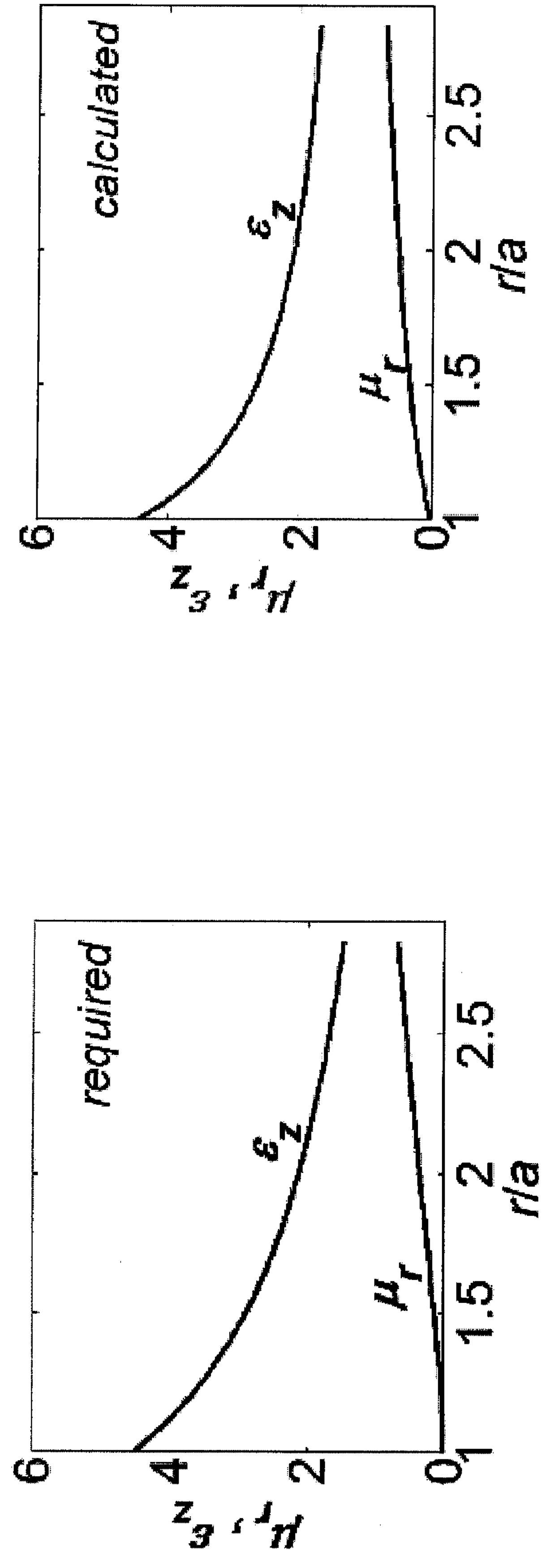
calculated
ε_z
μ_r
μ_r, ε_z
6
4
2
0
1
1.5
2
2.5
r/a
required
ε_z
μ_r
μ_r, ε_z
6
4
2
0
1
1.5
2
2.5
r/a

FIG. 8

SYSTEM, METHOD AND APPARATUS FOR CLOAKING

[0001] This application claims the benefit priority to U.S. provisional application Ser. No. 61/103,025, filed on Oct. 6, 2008, which is incorporated herein by reference.

STATEMENT OF GOVERNMENT SUPPORT

[0002] This work was supported in part Army Research Office grant W911NF-04-1-0350 and by ARO-MURI award 50342-PH-MUR.

TECHNICAL FIELD

[0003] This application relates to a system, method and apparatus for the modification of the observability properties of an object by a structure.

BACKGROUND

[0004] An object may be made effectively invisible at least over some frequency range. This has been termed a "cloak of invisibility"; the invisibility sought may be partial at a specific frequency, or over a band of frequencies, so the term "cloak of invisibility" or "cloak" may take on a variety of meanings. The cloak may be designed to decrease scattering (particularly "backscattering") from an object contained within, while at the same time reducing the shadow cast by the object, so that the combination of the cloak and the object contained therein have a resemblance to free space. When the phrase "cloaking," "cloak of invisibility," or the like, is used herein, the effect is generally acknowledged to be imperfect, and the object may appear in a distorted or attenuated form, or the background behind the object by the object may be distorted or partially obscured.

[0005] As will be understood by a person of skill in the art a "frequency" and a "wavelength" are inversely related by the speed of light in vacuo, and either term would be understood when describing an electromagnetic signal.

[0006] In some aspects, the cloak has a superficial similarity to "stealth" technology where the objective is to make the object as invisible as possible in the reflection or backscattering direction. One means of doing this is to match the impedance of the stealth material to that of the electromagnetic wave at the boundary, but where the material is strongly attenuating to the electromagnetic waves, so that the energy backscattered from the object within the stealth material is strongly attenuated on reflection, and there is minimal electromagnetic reflection at the boundary within the design frequency range. This is typically used in evading radar detection in military applications. Shadowing may not be a consideration in stealth technology. Shadowing may be understood as the effect of the object in blocking the observation of anything behind the object, for example the background, where the object is disposed between the observer and the background. A perfect cloak would result in no shadowing.

[0007] The materials used for the cloak may have properties where, generally, the permeability and permittivity tensors are anisotropic and where the magnitudes of the permeability and permittivity are less than one, so that the phase velocity of the electromagnetic energy being bent around the cloaking region is greater than that of the group velocity.

[0008] Materials having such properties have not been discovered as natural substances, but have been produced as artificial, man-made composite materials, where the permittivity and permeability of the bulk material are less than unity, and may be negative. They are often called "metamaterials" an extension of the concept of artificial dielectrics, that were first designed in the 1940s for microwave frequencies. Such materials typically consist of periodic geometric structures of a guest material embedded in a host material.

[0009] Analogous to the circumstance where homogeneous dielectrics owe their properties to the nanometer-scale structure of atoms, metamaterials may derive their properties from the sub-wavelength structure of its component materials. At wavelengths much longer than the unit-cell size of the material, the structure can be represented by effective electromagnetic parameters that are also used describe homogeneous dielectrics, such as an electric permittivity and a refractive index.

[0010] Cloaking has been experimentally demonstrated over a narrow band of microwave frequencies by achieved by varying the dimensions of a series of split ring resonators (SRRs) to yield a desired gradient of permeability in the radial direction.

SUMMARY

[0011] A apparatus for modifying the visibility properties of an object is disclosed, including a structure formed of a metamaterial. The metamaterial properties are selected so that an electromagnetic wave incident on the apparatus is guided around the object at plurality of wavelengths.

[0012] In an aspect, a method of designing a structure for use as a cloak effective at a plurality of wavelengths, includes the steps of: selecting a design wavelength; selecting a metamaterial having the property of having a low loss at the design wavelength and at least a permeability or a permittivity of less than unity; and determining, for a selected shape and size of structure, the variation of metamaterial properties as a function of position in the structure so as to guide electromagnetic waves of the design wavelength and polarization around a object disposed within the structure. A second design wavelength is selected and the design process is repeated for the second design wavelength.

[0013] In another aspect, a method of modifying the observability of an object, includes the steps of: providing a structure fabricated from a plurality of metamaterials, the metamaterials selected so as to guide electromagnetic waves around an object at a plurality of wavelengths; and disposing the structure between an observer and the object.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] FIG. **1** is a representation of a transformation of a vector field;

[0015] FIG. **2** is (a) an example of a general orthogonal cylindrical coordinate system; and, (b) a domain transformation for a cylindrical cloaking device where the larger initial domain in the left panel is mapped onto a scaled smaller annular domain shown in the right panel, leaving the central domain inaccessible to light; the initial and scaled domains share the same exterior boundary and the common space beyond;

[0016] FIG. **3** is a schematic representation of a cloaking system for multiple wavelengths or a finite bandwidth, with $w_1 > w_2 > w_3$, shown in (a), (b), and (c) respectively; the outer and inner circles represent the physical boundaries the cloaking device, and the circle between the two refers to an inner material boundary for each design wavelength;

[0017] FIG. **4** shows design constraints for constructing a non-magnetic cloak in the TM mode with high-order transformations; the thick solid and dashed lines represent the two Wiener bounds $\epsilon_{\|}(f)$ and $\epsilon_{\perp}(f)$, respectively: the basic material properties for this calculation are: $\epsilon_1=\epsilon_{Ag}=-10.6+0.14i$ and $\epsilon_2=\epsilon_{SiO2}=2.13$ at $\lambda=532$ nm;

[0018] FIG. **5** is a perspective view of a cylindrical non-magnetic cloak using the high-order transformations for TM polarization;

[0019] FIG. **6** is a graph of the anisotropic material parameters ϵ_r and ϵ_θ of a non-magnetic cloak made of silver-silica alternating slices corresponding to the third row ($\lambda=532$ nm) in Table 1; the solid lines represent the exact parameters determined by equation 35, and the diamond markers show the parameters on the Wiener's bounds given by equation (37);

[0020] FIG. **7** is a perspective view of a cylindrical non-magnetic cloak with high-order transformations for TE polarization; and

[0021] FIG. **8** shows a comparison of the theoretical and the calculated values of effective parameters μ_r and ϵ_z for a cylindrical TE cloak with SiC wire arrays at a design wavelength of $\lambda=13.5$ μm.

DETAILED DESCRIPTION

[0022] Exemplary embodiments of the apparatus and method may be better understood with reference to the drawings, but these embodiments are not intended to be of a limiting nature.

[0023] When the phrase "cloaking," "cloaking structure," "cloak of invisibility" or the like is used herein, the effect may be imperfect in practice, and the object may appear in a distorted or attenuated form, or the background obscured by the object may be distorted or partially obscured or attenuated, or the perceived color of the background may be modified. Therefore, "cloak" should not be interpreted so as to require that the object within the cloak be "invisible" even at a design wavelength, nor that the background be free of shadowing or distortion. Of course, a design objective may be to approach the ideal cloak at a wavelength or a range of wavelengths. A plurality of non-contiguous wavelength ranges may also be considered in a design for a structure.

[0024] The examples disclosed herein are intended to enable a person of ordinary skill in the art to practice the inventive concepts as claimed herein, using systems, apparatus, components, or techniques that may be known, disclosed herein, or hereafter developed, or combinations thereof. Where a comparison of performance is made between the examples disclosed herein and any known system, apparatus, component, or technique, such comparison is made solely to permit a person of skill in the art to more conveniently understand the present novel system, apparatus, component, or technique, and it should be understood that, in complex systems, various configurations may exist where the comparisons made may be better, worse, or substantially the same, without implying that such results are invariably obtained or constitute a limitation on the performance which may be obtained.

[0025] Broadband cloaking of electromagnetic waves can be understood by a person of skill in the art using a simplified example of a scaling transformation of a general cylindrical coordinate system. A generalized form of the transformation equations is presented so as to permit the application of this approach to other related designs.

[0026] The apparatus design may use metamaterials with specifically engineered dispersion. Constraints on the signs of gradients in the dispersion dependencies of dielectric permittivity and magnetic permeability for different operation wavelengths may result. Some constraints may be obviated by gain-assisted compensation for losses or electromagnetically induced transparency (EIT) are included in the design of cloaking system. So, when a structure, or a portion thereof, is described as "transparent," the transparency may be at a wavelength or a range of wavelengths, and should be understood to be achievable either by low loss materials, or materials with loss that has be compensated by a gain medium.

[0027] Electromagnetically induced transparency (EIT) is a coherent nonlinear process that may occur in some highly dispersive optical systems. EIT creates a narrow transparency window within an absorption peak. The anomalous dispersion along with a low optical loss available in an EIT system may be used for broadband optical cloaking. Similarly, in a gain medium the imaginary part of the refractive index has a negative value, and the dispersion curve exhibits an anti-Lorentz line shape. This property may result in anomalous dispersion with a low loss. Examples of EIT systems are three-state lead vapors. Examples of gain media include electrically or optically pumped semiconductors, dye modules, and quantum structures.

[0028] Examples of electromagnetic wave propagation in an isotropic bi-layer or for multilayer sub-wavelength inclusions of ellipsoidal (spheroidal or spherical) shapes in a dielectric host media are presented. Other geometrical shapes may be used. Such shapes may be known geometrical shapes, portions thereof, or shapes that are composites of geometrical shapes, including shapes that are arbitrary, but slowly varying with respect to the design wavelength.

[0029] In addition to numerical and theoretical studies of composite materials described herein, broadband transparency achieved by using multiphase spherical inclusions with appropriate layered geometries and materials is described. These examples are useful for estimating local electromagnetic fields and effective optical properties of heterogeneous media with binary or multi-phase inclusions, and as the starting point for more complex designs in accordance with the concepts described herein.

[0030] The basics of transformation optics (TO) approach to designing cloaking structures described herein may follow from the fundamental theoretical results of Dolin (Dolin. L. S., Izzv. Vyssh. Uchebn. Zaved., Radiofiz. 4, 694-7, 1961) which showed that Maxwell's equations can be considered to be form-invariant under a space-deforming transformation.

[0031] The underlying theoretical basis for the transformational optics (TO) approach is presented so as to enable a person of skill in the art to generalize the examples which follow. The transformation may be used at any wavelength, but the selection of materials and geometries may depend on the specific application of the design. As such, the terms "light," "optics," and the like, are understood to be interchangeable with "electromagnetic wave" at an appropriate frequency, and not to be limited to light visible to the human eye, infrared light, or the like. Specific examples are provided at visible (to the human eye) wavelengths, and in the mid infrared, so as to illustrate the concepts presented herein.

[0032] Consider an initial material space defined by its radius-vector $\tilde{r}(\tilde{x}, \tilde{y}, \tilde{z})$ and an inhomogeneous distribution of an anisotropic material property (e.g., either anisotropic permittivity, $\in$, or anisotropic permeability, μ), given by a tensor, $\tilde{m}=\tilde{m}(r)$ Suppose that the initial distribution of coupled vector fields, $\tilde{v}=\tilde{v}(\tilde{r})$ and $\tilde{u}=\tilde{u}(\tilde{r})$ is modified using a tensor j. The transformation may be formally achieved by mapping the initial space, using a coordinate transformation ($r=r(\tilde{r})$, i.e. $x=x(\tilde{x}, \tilde{y}, \tilde{z})$, $y=y(\tilde{x}, \tilde{y}, \tilde{z})$, $z=z(\tilde{x}, \tilde{y}, \tilde{z})$) with a non-singular Jacobian matrix j, ($|j|\neq 0$), so that it is a one-to-one transformation in a neighborhood of each point. The Jacobian matrix j is arranged from the columns of base vectors,

$$j=(r^{(x)}r^{(y)}r^{(z)}), \tag{1}$$

or its transposition can be arranged from the columns of gradients

$$j^T=(\tilde{\nabla}x\tilde{\nabla}y\tilde{\nabla}z). \tag{2}$$

In equation (1) and equation (2), $f^{(\cdot)}$ and $\tilde{\nabla}f=f^{(x)}\hat{x}+f^{(y)}\hat{y}+f^{(z)}\hat{z}$ denote a partial derivative and a gradient, respectively. The Jacobian determinant $|j|$ is equal to the triple vector product, $r^{(x)}r^{(y)}r^{(z)}$. Vectors $\tilde{v}$ and $\tilde{v}$ are have scalar components, as $\tilde{v}=\tilde{v}_x\hat{x}+\tilde{v}_y\hat{y}+\tilde{v}_z\tilde{z}$ and $v=v_x\hat{x}+v_y\hat{y}+v_z\hat{z}$, respectively.

[0033] Thus, a general invertible field-deforming transformation,

$$\tilde{v}=j^Tv, \tilde{u}=j^Tu, \tag{3}$$

links the vectors of the initial vector-space $\tilde{v}=\tilde{v}(\tilde{r})$ and $\tilde{u}=\tilde{u}(\tilde{r})$ with the new vectors of a deformed vector-space $v=v(r)$ and $\tilde{u}=\tilde{u}(\tilde{r})$ obtained at the corresponding points of the new material domain.

[0034] A solution may be sought so as to achieve a given transformation of the fields in equation (3). The initial material properties, $\tilde{m}=\tilde{m}(r)$, are modified in order to obtain the required transformation of the vector fields as determined by equation (3). A formal connection between the expressions for gradients before and after the change of variables may be expressed as:

$$x=x(\tilde{x},\tilde{y},\tilde{z}), y=y(\tilde{x},\tilde{y},\tilde{z}), z=z(\tilde{x},\tilde{y},\tilde{z})$$

where

$\tilde{\nabla}f(x, y, z)=f^{(x)}\tilde{\nabla}x+f^{(y)}\tilde{\nabla}y+f^{(x)}\tilde{\nabla}z$, which yields a general result that is analogous to equation (3)

$$\tilde{\nabla}=j^T\nabla. \tag{4}$$

[0035] The transformation identity for the curl can be derived first for pseudo-vectors $p=u\times v$ and $\tilde{p}=\tilde{u}\times\tilde{v}$. The standard vector algebra gives

$$\tilde{u}\times\tilde{v}=(j^Tu)\times(j^Tv)=|j|j^{-1}(u\times v), \tag{5}$$

connecting pseudo-vectors p and $\tilde{p}$ through

$$p=|j|^{-1}j\tilde{p}. \tag{6}$$

[0036] To obtain a formalism that is closer to Maxwell's curl equations, another product of the material tensor m and a vector u can be defined as $(mu)^{(t)}=\nabla\times v$, such that for time-independent material properties, $(mu)^{(t)}=mu^{(t)}$. This yields:

$$mu^{(t)}=\nabla\times v. \tag{7}$$

[0037] The right hand side of equation (7) is identical to $p=u\times v$, provided that vector u is replaced with ∇, following the result shown in equation (4). The use of the same sets of vector components, i.e. $\nabla\times\tilde{v}=(j^Tu)\times(j^Tv)=|j|j^{-1}(\nabla\times v)$ gives $mu^{(t)}=\nabla\times v$. Finally, using $\tilde{m}=|j|j^{-1}m(j^T)^{-1}$, equation (7) can be rewritten as,

$$\tilde{m}\tilde{u}^{(t)}=\tilde{\nabla}\times\tilde{v}. \tag{8}$$

[0038] Then, the required transform for tensors $\tilde{m}$ and m is given by

$$m=|j|^{-1}j\tilde{m}j^T. \tag{9}$$

[0039] As shown in FIG. **1**, the spatial transformation of the vector fields performed by tensor j through equation (3) can be considered as a spatial transformation $r=r(\tilde{x}, \tilde{y}, \tilde{z})$, with j being its Jacobian matrix, $j=(r^{(x)}\ r^{(y)}\ r^{(z)})$.

[0040] For the divergence relationships in Maxwell's equations, the derivation uses a scalar product of vector v and pseudo-vector p, which gives a scalar q (i.e., $v\cdot p=q$). Then, using equation (6) the scalar products yields $\tilde{v}\cdot\tilde{p}=(j^Tv)\cdot(|j|j^{-1}p)=|j|v\cdot p)=|j|v\cdot p$, and an equivalent divergence equation is obtained through substitution of v and $\tilde{v}$ with ∇ and $\tilde{\nabla}$, resulting in:

$$\tilde{\nabla}\cdot\tilde{p}=|j|\nabla\cdot p. \tag{10}$$

[0041] Equation (8) has cast the Maxwell curl equations $\nabla\times E=-\mu H^{(t)}$ and $\nabla\times H=\in E^{(t)}$ into a new set of similar equations, $\tilde{\nabla}\times\tilde{E}=-\tilde{\mu}\tilde{H}^{(t)}$ and $\tilde{v}\times\tilde{H}=\tilde{\in}\tilde{E}^{(t)}$, where

$$H=(j^T)^{-1}\tilde{H}, E=(j^T)^{-1}\tilde{E}, \tag{11}$$

and

$$\in=|j|^{-1}j\tilde{\in}j^T, \mu=|j|^{-1}j\tilde{\mu}j^T. \tag{12}$$

$(j^T)^{-1}$ in (11) is a matrix of the columns of reciprocal vectors $(j^T)^{-1}=(r^{(y)}\times r^{(z)}\ r^{(z)}\times r^{(x)}\ r^{(x)}\times r^{(y)})|j|^{-1}$.

[0042] Thus, provided that the electromagnetic properties of the new material space follow equation (12), the Poynting vector in the new space,

$$S=\frac{1}{2}(E\times H^*),$$

will obey equation (6), satisfying the following transformation of the initial Poynting vector

$$\tilde{S}=\frac{1}{2}(\tilde{E}\times\tilde{H}^*)$$

$$S=|j|^{-1}j\tilde{S}. \tag{13}$$

[0043] An analogous result would also be valid for other pseudo-vectors, e.g., the time derivatives of magnetic flux densities $B^{(t)}$ and $\tilde{B}^{(t)}$, and displacement currents, $D^{(t)}$ and $\tilde{D}^{(t)}$.

[0044] In a similar way, the divergence equations $\tilde{\nabla}\cdot\tilde{D}=\tilde{q}$ and $\nabla\cdot D=q$ would link the charge densities through equation (10) as

$$q=|j|^{-1}\overline{q}. \tag{14}$$

[0045] The above conversions provide a method of designing a continuous material space for a required spatial transformation of electromagnetic vectors and, therefore, achieving a desired functionality. That is, for the physical Poynting vector, S, to match the required transformation of the Poynting vector, $S=|j|^{-1}jS$, the material properties in the new space, $r=r(\tilde{r})$, should satisfy $\in=|j|^{-1}\,j\in j^{T}$ and $\mu=|j|^{-1}\,j\mu j^{T}$.

[0046] The result of Dolin is repeated here as equation (15), as the original work is in Russian and not readily available. The radially anisotropic permeability and permittivity of a spherical material inhomogeneity may be expressed as:

$$\|\varepsilon_{ik}\| = \|\mu_{ik}\| = \begin{vmatrix} \frac{R^2}{r^2(R)}\frac{dr(R)}{d(R)} & 0 & 0 \\ 0 & \frac{1}{\frac{dr(R)}{dR}} & 0 \\ 0 & 0 & \frac{1}{\frac{dr(R)}{dR}} \end{vmatrix} \tag{15}$$

corresponding to a spatial transformation from the spherical coordinates r, Q, j to the coordinates R (r), Q, j . A plane wave incident from infinity on an inhomogeneity with parameters in accordance with equation (15) would pass through the inhomogeneity without apparent distortion to the external observer.

[0047] A method is described herein for the design of broadband cloaking apparatus and systems comprising binary or multiphase metamaterials, where different optical paths are arranged for different wavelengths inside the macroscopic cloaking structures. The cloaking design requirements may be satisfied through appropriate dispersion engineering of metamaterials.

[0048] The concept of an electromagnetic cloak is to create a structure, whose permittivity and permeability distributions allow the incident waves to be directed around the inner region and be (at least ideally) emitted on the far side of the structure without distortion arising from propagating through the structure. From among simple geometries, including spherical, square and elliptical varieties, cloaking in a cylindrical system is may be the most straightforward to describe mathematically, and is used for the examples herein. However, solutions in other than cylindrical coordinate systems arise from the general transformational optics theory presented herein. A person of skill in the art would understand that such structures may not need to be solved analytically, as numerical analysis methods may be effectively used. Such numerical analysis techniques may also be used for more complex structures. For some scale sizes, ray tracing in an inhomogeneous anisotropic medium may be used. For numerical analysis of cloaking devices, there are a variety of numerical electromagnetic approaches that can be used, such as the finite-element methods (FEM), the finite-difference time-domain (FDTD) methods, the finite integration technique (FIT), and the method of moments (MoM). A number of commercial packages are widely, including COMSOL MULTIPHYSICS, CST MICROWAVE STUDIO, RSoft FULLWAVE, and others may be used to perform the numerical analysis and design.

[0049] A class of a general orthogonal cylindrical coordinate system (OCCS) can be arranged by translating an x-y-plane map ($x=x(\tilde{\nu},\tilde{\tau})$, $y=y(\tilde{\nu},\tilde{\tau})$) perpendicular to itself; the resulting physical coordinate system forms families of concentric cylindrical surfaces. Since the unit vectors are orthogonal, $\hat{e}_{\nu}\times\hat{e}_{\tilde{\tau}}=\hat{e}_{z}$, $\hat{e}_{\tilde{\tau}}\times\hat{e}_{z}=\hat{e}_{\nu}$, and $\hat{e}_{z}\times\hat{e}_{\nu}=\hat{e}_{\tilde{\tau}}$, the complexity of TO problems in TE or TM formulations can be significantly reduced.

[0050] Consider the initial OCCS, where a 2D radius-vector is defined by a parametric vector function $\tilde{r}(\tilde{\nu},\tilde{\tau})$, and a 2D vector $\tilde{\mu}$ is defined as $\tilde{u}=\tilde{\nu}_{\mu}\hat{e}_{\nu}+\tilde{\nu}_{\tilde{\tau}}\hat{e}_{\tilde{\tau}}$. The Jacobian matrix is the diagonal matrix, s%=diag(s$_{1\%}$s$_{1\%}$), with the metric coefficients $\tilde{s}=\mathrm{diag}(s_{\tilde{\nu}},s_{\tilde{\tau}})$ and $s_{\tilde{\nu}}=\sqrt{\tilde{r}^{(\tilde{\nu})}\cdot\tilde{r}^{(\tilde{\nu})}}$. Then, the following scalar wave equation may be obtained from the Maxwell curl equations in an orthogonal cylindrical basis for a general anisotropic media. Thus, from

$$\tilde{u}_{\tilde{\nu}}=\omega^{-1}\tilde{m}_{\tilde{\nu}}^{-1}\tilde{s}_{\tilde{\tau}}^{-1}\tilde{\nu}^{(\tilde{\tau})},\ \tilde{u}_{\tilde{\tau}}=-\omega^{-1}\tilde{m}_{\tilde{\tau}}^{-1}\tilde{s}_{\tilde{\nu}}^{-1}\tilde{\nu}^{(\tilde{\nu})},\ -\omega\tilde{m}_{z}\tilde{\nu}=|\tilde{s}|^{-1}[(\tilde{s}_{\tilde{\tau}}\tilde{u}_{\tilde{\tau}})^{(\tilde{\nu})}-(\tilde{s}_{\tilde{\nu}}\tilde{u}_{\tilde{\nu}})^{(\tilde{\tau})}], \tag{16}$$

we arrive at

$$(\tilde{s}_{\tilde{\tau}}\tilde{m}_{\tilde{\tau}}^{-1}\tilde{s}_{\tilde{\nu}}^{-1}\tilde{\nu}^{(\tilde{\nu})(\tilde{\nu})}+(\tilde{s}_{\tilde{\nu}}\tilde{m}_{\tilde{\nu}}^{-1}\tilde{s}_{\tilde{\tau}}^{-1}\tilde{\nu}^{(\tilde{\tau})})^{(\tilde{\tau})}-\omega^{2}\tilde{m}_{z}|s|\nu=0, \tag{17}$$

where $\tilde{m}_{\nu}$ and $\tilde{m}_{\tilde{\tau}}$ are the only components of a diagonal material property tensor, i.e., anisotropic permeability or anisotropic permittivity (for TM or TE polarization respectively); the scalar $\tilde{\nu}$ is the only component of the , transverse field: i.e., the magnetic field, $H=\hat{e}_{x}H_{z}$ (TM), or the electric field, $E=\hat{e}_{z}E_{z}$ (TE).

[0051] Similar to equation (17), another wave equation in a new physical OCCS, (ν, τ, z), can be written as

$$(s_{\tau}m_{r}^{-1}s_{\nu}^{-1}\nu^{(\nu)})^{(\nu)}+(s_{\nu}m_{\nu}^{-1}s_{\tau}^{-1}\nu^{(\tau)})^{(\tau)}-\omega^{2}m_{z}|s|\nu=0 \tag{18}$$

To mimic the behaviour of light waves obeying equation (16), a scaling transformation $\nu=\nu(\tilde{\nu})$ (with $\tau=\tilde{\tau}$, $z=\tilde{z}$, and $\nu^{1}=\nu^{(\tilde{\nu})}$) is introduced. Thus, to get closer to equations (16), equations (18) are expressed as

$$\left(\left[\frac{1}{v'}\frac{s_{T}\tilde{m}_{T}\tilde{s}_{\tilde{v}}}{\tilde{s}_{\tilde{T}}m_{T}s_{v}}\right]\frac{\tilde{s}_{\tilde{T}}}{\tilde{m}_{\tilde{T}}\tilde{s}_{\tilde{v}}}v^{(\tilde{v})}\right)^{(\tilde{v})}+\left(\left[v'\frac{s_{v}\tilde{m}_{\tilde{v}}\tilde{s}_{T}}{\tilde{s}_{v}m_{v}s_{T}}\right]\frac{\tilde{s}_{v}}{\tilde{m}_{v}\tilde{s}_{T}}v^{(\tilde{T})}\right)^{(\tilde{T})}-w^{2}\left(v'\frac{m_{z}|s|}{\tilde{m}_{z}|\tilde{s}|}\right)\tilde{m}_{z}|\tilde{s}|v=0.. \tag{19}$$

It follows that equation (19) is may be made to be the same as equation (16), provided that the ratios in the square brackets can be eliminated. Thus, the TO identities

$$\frac{1}{v'}\frac{s_{T}\tilde{m}_{\tilde{T}}\tilde{s}_{\tilde{v}}}{\tilde{s}_{T}m_{T}s_{v}}=1,\ v'\frac{s_{v}\tilde{m}_{\tilde{v}}\tilde{s}_{T}}{\tilde{s}_{v}m_{v}s_{T}}=1,\ v'=\frac{m_{z}|s|}{\tilde{m}_{z}|\tilde{s}|}=1, \tag{20}$$

should be valid in a new material space (m_{ν}, m_{τ}, and m_{z}) in order to mimic the behaviour of light in the initial material space ($\tilde{m}_{\nu}$, $\tilde{m}_{\tilde{\tau}}$, and $\tilde{m}_{z}$). The above identities define the material transformation requirements which may be used for cloaking design and other applications.

[0052] Equations (20) are a solution to the problem of designing an anisotropic continuous material space supporting a required electromagnetic wave behavior, which is equivalent to the behavior of the electromagnetic waves mapped back onto the initial space. Scaling transformations that expand the initially small domain onto a larger physical domain are pertinent to imaging or light concentration while a typical cloaking application uses scaling transforms that shrink the initially larger space to produce voids excluded from the initial domain. Such voids are therefore inaccessible to electromagnetic waves at least the design frequency. The initial virtual space shares a common exterior boundary with the rest of the transformed physical world. An example is shown in FIG. **2**.

[0053] In the circular cylindrical coordinates ($\nu=\rho$, $\tau=\phi$), and $s_\rho=1$, $s_\phi=\rho$, equations (20) give

$$m_\phi = \frac{\rho}{\rho'\tilde{\rho}}\tilde{m}_\phi,\ m_\rho = \frac{\rho'\tilde{\rho}}{\rho}\tilde{m}_\rho,\ m_z = \frac{\tilde{\rho}}{\rho\rho'}\tilde{m}_z, \tag{21}$$

which are the material space parameters for an exact cloak, which is analogous to a cylindrical free-space domain, and is defined by the following inhomogeneous and anisotropic material properties:

$$\in_\rho=\mu_\rho=\wp\rho^1/\rho;\ \in_\phi=\mu_\phi=\in_\rho^{-1};\ \in_z=\mu_z=\wp/(\rho^1\rho). \tag{22}$$

[0054] The constraints on the material properties may be relaxed in some circumstances. For example, for TM polarization with the magnetic field polarized along the z-axis, multiply $\in^\tau$ and $\in_\phi$ by μ_z in equation (22) to obtain the following reduced set of non-magnetic cloak parameters:

$$\in_\rho=(\wp/\rho)^2;\ \in_\phi=(\rho^1)^{-2};\ \mu_z=1. \tag{23}$$

[0055] Similarly, for the TE polarization, the required parameters for a general transformation are:

$$\mu_\rho=(\wp/\rho)^2(\rho^1)^2,\ \mu_\phi=1,\ \in_z=(\rho^1)^{-2}. \tag{24}$$

In equations (22)-(24), $\wp$ could be replaced by $\wp=\wp(\rho)$ to obtain closed-form expressions. Such closed form expressions are useful to verify numerical analysis results for a corresponding geometrical configuration. The numerical analysis may then be extended to situations where the geometry of the apparatus or the complexity of the material spatial variations may make a closed-form solution impractical as a design tool. A person of skill in the art would use the numerical analysis methods so as to extend the scope of the types of apparatus, materials and wavelength regimes which may be used in designs based on the theoretical analysis presented herein.

[0056] Consider the bandwidth of a cloaking structure when a design for a single specific central wavelength is used. A broadband cloak may be designed to function in a wavelength multiplexing manner. Since the anisotropic constituent materials of a cloak for one wavelength may not be transparent at other frequencies, cloaks for the wavelengths being considered should share the same outer boundary, may be is the physical outer boundary of the device. The inner boundary and the transformation for each operating wavelength is dependent on the wavelength. Thus, a number of different inner boundaries and different transformations may be used to provide a broadband cloaking capability.

[0057] In practice, the registration of the outer boundaries of the different material layers may have some variation without appreciable degeneration of the overall effectiveness of the broadband guidance. This follows from simulations which have suggested that variations from the ideal material parameter profile may be tolerated.

[0058] Moreover, as the theoretical results here and elsewhere in the description herein are obtained from analytic models, some adjustment of the results may be needed in practice to, for example, take account of the refraction of a signal of a wavelength that differs from the design wavelength, or which passes through a shell of another design wavelength prior to being refracted by a shell designed for the signal. In another aspect, while gain media may be needed in some cases for an exact cloaking result, some loss may be tolerated in the structure, depending on the application, and the sensitivity of the viewer or viewing device to changes in the strength of the background signal, the transmitted signal or the like.

[0059] FIG. **3** is a schematic representation of a cloaking system for multiple wavelengths or a finite bandwidth, with $w_1>w_2>w_3$, shown in (a), (b), and (c) respectively; the outer and inner circles represent the physical boundaries the cloaking device, and the circle between the two refers to an inner material boundary for each design wavelength;.

[0060] Since the wave components at different frequencies go through the system following different physical paths, the proposed system may permit the cloaking parameters to be appropriately realized over a finite bandwidth without violating basic physical laws or giving rise to a superluminal group velocity. As a result, a 'colorful' (multi-frequency) image would appear transparently through the cloaking device. At the central wavelength of each of the various designs, an image of the background region behind the cloaking structure in the design wavelength ("color") would be seen. This would be the situation for each of the design wavelengths of the structure.

[0061] The device may be constructed using multiple shells of material, where the material properties of each shell is appropriate for the wavelengths propagating therein. Further, it would be understood that each shell may also be comprised of a number of conformal shells with material properties that vary with a geometric dimension such as the radius. Such a construction may facilitate the manufacturing process. Further, although not shown, some shells may be a gain material, or dielectric materials or various types of materials may be fabricated as a composite material.

[0062] In order to better understand the limitations on cloaking over a contiguous band of frequencies, consider the TE propagation mode with material properties given in equation (24), which allows for flexible parameters at the outer boundary of $\rho=b$. Assume that at frequency ω_0, the material properties required by a TE cloak are exactly satisfied based on the transformation $\rho=\rho(\wp)$ within the range, $\alpha\leqq\rho\leqq b$;

$$\mu_\rho(\omega_0,\rho)=(\wp/\rho)^2(\rho^1)^2,\ \mu_\phi(\omega_0,\rho)=1,\ \in_z(\omega_0,\rho)=(\rho^1)^{-2}. \tag{25}$$

[0063] Dispersion needs to be considered for broadband performance of a cloaking system. Assuming that the cloaking materials exhibit a linear dispersion around the initial frequency ω_0 the dispersion function may be expressed in a Taylor series expansion:

$$\mu_\rho(\omega,\rho)=\mu_\rho(\omega_0,\rho)+\mu_\rho^{(\omega)}(\omega_0,\rho)(\omega-\omega_0), \tag{26}$$

and

$$\in_z(\omega,\rho)=\in_z(\omega_0,\rho)+\in_z^{(\omega)}(\omega_0,\rho)(\omega-\omega_0), \tag{27}$$

[0064] In equations (26) and (27) the two frequency derivatives $\mu_\rho^{(\omega)}$ and $\in_z^{(\omega)}$ are continuous functions of ρ. Since there is no magnetic response along the ϕ direction at ω_0, it may be reasonable to choose that $\mu_\phi(\omega,\tau)=\mu_\phi(\omega_0,\tau)=1$.

[0065] The initial formulation of the analysis is to determine, at a frequency $\omega_1=\omega_0+\delta\omega$, a combination of the transformation $\rho_1=\rho_1(\wp)$ along with yet another inner radius a_1 such that the function $\rho_1(\wp)$ maps [0, b] onto [a_1, b] with $a<a_1<b$, while satisfying the boundary conditions

$$\rho_1(0)=a_1\rho_1(b)=b \tag{28}$$

along with the monotonicity condition:

$$\rho_1{}^1>0 \tag{29}$$

and, the material transforms of the reduced TE cloak:

$$\mu_\rho(\omega_1,\rho_1)=(\varrho/\rho_1)^2(\rho_1{}^1)^2,\ \in_z(\omega_1,\rho_1)=(\rho_1{}^1)^{-2} \tag{30}$$

where $\varrho=g_1{}^{-1}(\rho)$, $a_1\leqq\rho\leqq b$.

[0066] The transformation $\rho_1(\varrho)$ for $\omega_1=\omega_0+\delta\omega$ is related to the original transformation at ω_0 and the dispersion functions by:

$$(\varrho(\rho_1)/\rho_1)^2(\rho_1{}^1)^2=(\varrho(\rho)/\rho)^2(\rho^1)^2+\mu_\rho{}^{(\omega)}(\omega_0,\rho)(\omega-\omega_0), \tag{31}$$

and

$$(\rho_1{}^1)^{-2}=(\rho^1)^{-2}+\in_z{}^{(\omega)}(\omega_0,\rho)(\omega-\omega_0), \tag{32}$$

within the range of $a_1\leqq\rho_1\leqq b$ with the boundary conditions mentioned above. It would appear that equations (31) and (32) may not be fulfilled exactly for arbitrary gradients of dispersion functions $\mu_\rho{}^{(\omega)}(\omega_0,\rho)$ and $\in_z{}^{(\omega)}(\omega_0,\rho)$.

[0067] Therefore, achieving complete cloaking over a bandwidth involves computational methods and materials for dispersion management. This requirement may be expressed as: What physically-possible functions $\mu_\rho{}^{(\omega)}(\omega_0,\rho)$ and $\mu_\rho{}^{(\omega)}(\omega_0,\rho)$ should be engineered to make the cloaking effect possible at a given frequency $\omega_1=\omega_0+\delta\omega$ in addition to cloaking at ω_0?

[0068] After some algebra, it may be seen that equations (28) to (32) can be satisfied by

$$\mu_\rho{}^{(\omega)}(\omega_0,\rho)\in_z{}^{(\omega)}(\omega_0,\rho)<0. \tag{33}$$

That is, equation (33) indicates that the dispersion of the radial permeability $\mu_\rho(\omega,\rho)$ and the axial permittivity $\in_z(\omega,\rho)$ should have opposite slopes as functions of the frequency.

[0069] The effective bandwidth of a transformation-based cloaking device is determined by the frequency range over which the material properties in equations (22)-(24) are substantially satisfied. The curved trajectory of the electromagnetic waves within the cloak implies a refractive index n of less than 1 in order to satisfy the minimal optical path requirement of the Fermat principle. However, a metamaterial with n<1 should be dispersive to fulfill causality.

[0070] In practice, the bandwidth of the apparatus may largely be determined by the performance tolerances. That is, how close to the performance of an ideal cloak over a bandwidth is achieved. The needed performance may be dependent on the application for which the structure is intended. So, while mathematically there may be a single wavelength value where the cloaking conditions are exactly fulfilled, the undesired scattering and distortion arising from the cloak structure may remain at a low level over a finite bandwidth. As such, cloaks share the property of many engineering solutions in that compromises in performance may be accepted as a trade-off with respect to cost, complexity, and the like.

[0071] Specifically engineered strong anomalous dispersion may be needed as equation (33) is not satisfied with normal dispersion, where $\partial\in(\omega)/\partial\omega>0$ and $\partial\mu(\omega)/\partial\omega>0$. However, anomalous dispersion characteristics are normally associated with substantial loss. In such designs, a broadband cloaking solution may need additional loss-compensation by incorporating gain media in the structure.

[0072] Passive materials exhibit normal dispersion away from the resonance band. Because anomalous dispersion usually occurs only around the absorption bands, a wavelength multiplexing cloak with broadband capability may be achievable when gain materials or electromagnetically induced transparency or chirality are introduced to make low-loss anomalous dispersion possible. For example, in an active medium, where the optical gain is represented by a negative imaginary part of permittivity over a finite bandwidth, the real part of permittivity around the active band will exhibit an anti-Lorentz line shape, as governed by the Kramers-Kronig relations. As a result, anomalous dispersion with relatively low loss can occur in the wings of the gain spectrum. Incorporating gain materials into plasmonics and metamaterials has been proposed and demonstrated in related applications such as, a near-field superlens, tunneling transmittance, enhanced surface plasmons, and lossless negative-index materials.

[0073] We present two structures for optical cloaking based on high-order transformations for TM and TE polarizations respectively. These designs are realizable for at least visible and infrared light wavelengths.

[0074] The constitutive dimensional and electromagnetic parameters of the cloak are determined by the specific form of the spatial transformation used. The parameters are usually anisotropic with gradient requirements that may be achieved using artificially engineered structures

[0075] Two design examples of optical cloaks based on high-order transformations are described. Specifically: i) a non-magnetic cylindrical cloaking system for TM incidence (magnetic field polarized along the cylindrical axis) which consists of a layered metal-dielectric without any variation in either material or structure along the vertical direction; and, ii) a magnetic cylindrical cloak for TE incidence (electric field polarized parallel to axis) utilizing Mie resonance in periodic rod-shaped high-permittivity materials.

[0076] For a cloak in the cylindrical geometry, a coordinate transformation function $r=g(r^1)$ from (r^1, θ^1, z^1) to (r, θ, z) is used to compress the region $r^1\leqq b$ into a concentric shell of $a\leqq r\leqq b$, and the permittivity and permeability tensors required for an exact cloak can be determined as:

$$\in_r=\mu_r=(r^1/r)\partial g(r^1)\partial r^1;\ \in_\theta=\mu_\theta=1/\in_r;\ \in_z=\mu_z=(r^1/r)[\partial g(r^1)/\partial r^1]^{-1} \tag{34}$$

For the standard states of incident polarization, the requirement of equation (34) can be relaxed such that only three of the six components are relevant. For example, for TE (TM) polarization, only μ_z, μ_r and μ_θ(μ_z, $\in_\tau$ and $\in_\theta$) enter into Maxwell's equations. As would be understood, the TM and the text in parenthesis are read in lieu of the TE and corresponding parameters so as to provide a compact presentation of the discussion

[0077] The parameters can be further simplified to form reduced parameters which are more realistic for practical applications. Since the trajectory of the waves is determined by the cross product components of the $\in$ and μ tensors instead of the two tensors individually, the cloaking performance is sustained as long as $n_\theta=\sqrt{\in_z\mu_r}$ and $n_r=\sqrt{\in_s\mu_\theta}$ ($n_0=\sqrt{\mu_z\in_r}$ and $n_r=\sqrt{\mu_z\in_\theta}$) meet equation (34). This technique results in a specific set of reduced parameters which allow for a permeability gradient along only the radial direction for the TE mode:

$$\mu_r=(r^1/r)^2[\partial g(r^1)/\partial r^1]^2;\ \mu_\theta=1;\ \in_z=[\partial g(r^1)/\partial r^1]^{-2} \tag{35}$$

and can be purely non-magnetic for the TM mode:

$$\in_r=(r^1/r)^2;\ \in_\theta=[\partial g(r^1)/\partial r^1]^{-2};\ \mu_z=1 \tag{36}$$

[0078] The designs of the example electromagnetic cloaks herein use known structures and materials to achieve the set of parameters corresponding to any of equations (34)-(36). Recently a demonstration of a microwave cloak satisfying equation (35) was reported and the previously described nonmagnetic optical cloak in U.S. patent application Ser. No. 11/983,228, filed on Nov. 7, 2007, and is incorporated herein by reference, corresponds to the case described by equation (36). One common aspect in the previous work is that the designs were based on a standard linear transformation $r=g(r^1)=(1-a/b)r^1+a$.

[0079] Designs based on more general high-order transformations are described. In particular, for the TM polarization, a non-magnetic cloak design which may compatible with mature fabrication techniques such as direct deposition and direct etching is described; for TE incidence, a structure that allows for a radial gradient in the magnetic permeability while avoiding the use of plasmonic metallic inclusions in the optical range is described.

[0080] Consider a non-magnetic cloak for the TM mode with parameters given in equation (36). In this case, the cloak material is designed to produce the required gradients in $\in_r$ and $\in_\theta$ using readily available materials. In an aspect, the design may employ the flexibility in realizing the effective permittivity of a general two-phase composite medium.

[0081] When an external field interacts with a composite material comprising two elements with permittivity of $\in_1$ and $\in_2$ respectively, minimal screening occurs when all internal boundaries between the two constituents are parallel to the electric field, and maximal screening occurs when all boundaries are aligned perpendicular to the field. These two extremes of orientation can be achieved by using an alternating layered structure, provided that the thickness of each layer is much less than the wavelength of the incident electromagnetic radiation. The two extreme values of the effective permittivity can be approximated as:

$$\in_{\|}=f\in_1+(1-f)\in_2;\ \in_{\perp}=\in_1\in_2/(f\in_2+(1-f)\in_1) \qquad (37a,\ b)$$

where f and $1-f$ denote the volume fractions of components **1** and **2**, and the subscripts $\|$ and $\perp$ indicate the cases with electric field polarized parallel and perpendicular to the interfaces of the layers, respectively. Such layered structures have been studied extensively in recent years for various purposes, especially in sub-diffraction imaging for both the near field and the far zone.

[0082] The alternating layers may be a plurality of layers, each layer having a bulk material property appropriate to a particular wavelength and the shape of the cloaking structure being designed, and some of these layers may be, for example gain media so as to compensate for the loss in passive layers.

[0083] The two extrema in equation (4) are termed the Wiener bounds on the permittivity, which set the bounds on the effective permittivity of a two-phase composite material. Other limits, for example those from the spectral representation developed by Bergman and Milton (see Bergman, D. J., Phys. Rev. Lett. 44, 1285-1287, 1980; Milton, G. W., Appl. Phys. Lett. 37 , 300-302, 1980) may also apply in addition to the Wiener bounds, but equation (37) nonetheless provides a straightforward way to evaluate the accessible permittivity range in a composite with specified constituent materials. The Wiener bounds can be illustrated on a complex $\in$-plane with the real and imaginary parts of $\in$ being the x and y axis, respectively. In this plane, the low-screening bound in equation (37a) corresponds to a straight line between $\in_1$ and $\in_2$, and the high-screening bound in equation (4b) defines an arc which is part of the circle determined by the three points: $\in_1$, $\in_2$ and the origin.

[0084] The material properties for the cloak design corresponding to equation (36) are such that, for a non-magnetic cylindrical cloak with any transformation function, $\in_r$ varies from 0 at the inner boundary of the cloak (r=a) to 1 at the outer surface (r=b), while $\in_\theta$ is a function of r with varying positive value, except for the linear transformation case where $\partial g(r^1)/\partial r^1$ is a constant.

[0085] Fulfilling the parameters in equation (36) may use, for example, alternating metal-dielectric slices whose properties may be estimated by equation (37). Phase **1** is a metal ($\in_1=\in_m<0$) and phase **2** is a dielectric ($\in_2=\in_d>0$), and the desired material properties of the cloak are achieved when the slices are within the r-z plane of the cylindrical coordinates. $\in_r$ and $\in_\theta$ correspond to $\in_{\|}$ and $\in_{\perp}$ in equation (37), respectively.

[0086] This situation is illustrated in FIG. **4**. The thick solid and dashed lines represent the two Wiener bounds $\in_{\|}(f)$ and $\in_{\perp}(f)$, respectively. The constituent materials used for the calculation presented in FIG. **4** are silver and silica at a "green" light wavelength of 532 nm. The pair of points on the bounds with the same filling fraction are connected with a straight line for clarity. When $\in_r$ varies between 0 and 1, the value of $\in_\theta$ varies accordingly as shown by the arrow between the two thin dashed lines. Therefore, the construction of a non-magnetic cloak establishes the relationship between the two quantities $\in_{\|}$ and $\in_{\perp}$ (as functions of f) within the range shown in FIG. **4** that fits the material properties given in equation 36 for a particular transformation function: $r=g(r^1)$.

[0087] The example design has a low loss factor. As shown in FIG. **4**, the loss factor described by the imaginary part of the effective permittivity is on the order of 0.01. This is considerably smaller than that of a pure metal or any resonant metal-dielectric structures. A schematic representation of the structure having interlaced metal and dielectric slices is illustrated in FIG. **5**.

[0088] For a selected design wavelength, a transformation together with the cylindrical shape factor a/b that fulfills the following equation may be suitable.

$$\varepsilon_m\varepsilon_d\left(\frac{\partial g(r')}{\partial r'}\right)^2+\left(\frac{r'}{g(r')}\right)^2-\varepsilon_m+\varepsilon_d=0 \qquad (38)$$

and

$$g(0)=a;\ g(b)=b;\ \partial g(r^1)/\partial r^{1>}0 \qquad (39)$$

[0089] An approximate solution to the equations may be found using a polynomial function such as:

$$r=g(r^1)=[1-a/b+p(r^1-b)]r^1+a \qquad (40)$$

with $|p|<(b-a)/b^2$

[0090] Such a quadratic transformation satisfies the boundary and monotonicity requirements in equation (39), and it is possible to fulfill equation (38) with minimal deviation from a theoretical profile when an appropriate shape factor is chosen. Table 1 sets forth transformations, materials and geometries for non-magnetic cloaks designed for several important central wavelengths across the visible wavelength regime including 488 nm (Ar-ion laser), 532 nm (Nd:YAG laser), 589.3 run (sodium D-line), and 632.8 nm (He—Ne laser). In the calculations, the permittivity of silver is taken from well accepted experimental data (see Johnson, P. B., and R. W. Christy, Phys. Rev. B **6 4370-4379** **,1972**), and the dielectric constant of silica is from tabulated data (see Palik, E. D., Handbook of Optical Constants of Solids, Academic Press, New York, 1997. The same design and transformation work for similar cylindrical cloaks with the same shape factor a/b. When the approximate quadratic function is fixed for a given design wavelength, the filling fraction function $f(r)$ is determined by:

$$f(r) = \frac{Re(\varepsilon_d) - (g^{-1}(r)/r)^2}{Re(\varepsilon_d - \varepsilon_m)} \tag{41}$$

TABLE 1

Approximate quadratic transformations and materials for constructing a cloak with alternating slices

λ	ϵ_1	ϵ_2	$p \times (b^2/a)$	a/b
488 nm	$\epsilon_{Ag} = -8.15 + 0.11i$	$\epsilon_{SiO2} = 2.14$	0.0662	0.389
532 nm	$\epsilon_{Ag} = -10.6 + 0.14i$	$\epsilon_{SiO2} = 2.13$	0.0517	0.370
589.3 nm	$\epsilon_{Ag} = -14.2 + 0.19i$	$\epsilon_{SiO2} = 2.13$	0.0397	0.354
632.8 nm	$\epsilon_{Ag} = -17.1 + 0.24i$	$\epsilon_{SiO2} = 2.12$	0.0333	0.347
11.3 nm	$\epsilon_{SiC} = -7.1 + 0.40i$	$\epsilon_{BaF2} = 1.93$	0.0869	0.356

[0091] FIG. **6** shows the calculated anisotropic material properties of a non-magnetic cloak corresponding to the λ=532 nm case. With the approximate quadratic transformation, the effective parameters $\in_r$ and $\in_\theta$ obtained with the Wiener bounds in equation (37) fit with the exact parameters required for this transformation by equation (35) quite well, with the average deviation of less than 0.5%.

[0092] Fabrication of the design is practical, as such vertical wall-like structures are compatible with mature fabrication techniques such as direct deposition and direct etching.

[0093] In another example, a cylindrical cloak for TE mode cloaking operable within the mid-infrared frequency range is described, with a gradient in the magnetic permeability, in accordance with equation (35). This frequency range is of interest as it corresponds to the thermal radiation band from human bodies.

[0094] Several different approaches involving silicon carbide as component of the metamaterial are described. SiC is a polaritonic material with a phonon resonance band falling into the spectral range centered at around 12.5 μm (800 cm^{-1}) This resonance band introduces a sharp Lorentz behavior in the electric permittivity. The dielectric function of SiC at mid-infrared may be described with the following model:

$$\in_{SiC} = \in_\infty[\omega^2 - \omega_L^2 + i\gamma\omega]/[\omega^2 - \omega_T^2 + i\gamma\omega] \tag{42}$$

where $\in_\infty$=6.5, ω_L=972 cm^{-1}, ω_T=796 cm^{-1} and γ=5 cm^{-1}. On the high-frequency side of the resonance frequency, the dielectric function is strongly negative, which makes the optical response similar to that of metals, and the material has been already been utilized in applications such as a mid-infrared superlens. At frequencies lower than the resonance frequency, the permittivity can be strongly positive, which makes SiC a candidate for producing high-permittivity Mie resonators at the mid-infrared wavelength range.

[0095] SiC structures may be used to build mid-infrared cloaking devices in a variety of physical configurations. For example, the needle-based structure may be used for the TM mode, where needles are made of a low-loss negative-∈ polaritonic material such as, for example, SiC or TiO_2, and are embedded in an infra-red-transparent dielectric such as, for example, ZnS.

[0096] In another aspect, a non-magnetic cloak using alternating slices of structure as previously described herein may be used. With SiC as the negative-s material and BaF_2 as the positive-∈ slices, the appropriate transformation function and shape factor that fulfills the material property requirements at a preset wavelength may be determined. The result for λ=11.3 μm (CO_2 laser range) is shown in the last row of Table 1.

[0097] In yet another example, a cylindrical cloak for the TE mode with the required material properties given in equation (35) is described, having a gradient in the magnetic permeability along the radial direction. μ_r may vary from 0 at the inner boundary (r=a) to $[\partial g(r^1)/\partial r^1]^2$: at the outer surface (r=b), while the $\in_z$ changes according to $[\partial g(r^1)/\partial r^1]^{-2}$. The magnetic requirement may be accomplished using metal elements like split-ring resonators, coupled nanostrips or nanowires. However, such plasmonic structures exhibit a high loss. A SiC based structure provides an all-dielectric design to a magnetic cloak for the TE mode due to the Mie resonance in subwavelength SiC inclusions.

[0098] Meta-magnetic responses and a negative index of refraction in structures made from high-permittivity materials have been studied extensively in recently years. Magnetic resonance in a rod-shaped high-permittivity particle can be excited by different polarizations of the external field with respect to the rod axis. When a strong magnetic resonance and an effective permeability substantially distinct from 1 are desired, the rod should be aligned parallel to the electric field to assure the maximum possible interaction between the rod and the external field. In the present example the radial permeability has values of less than (but close to) 1, and resonance behavior in the effective permittivity $\in_z$ should be avoided for a minimal loss. Therefore, with the electrical field polarized along the z axis of the cylindrical system, the SiC rods may be arranged along the r axis and form an array in the θ-z plane. The structure is depicted in FIG. **7**, where arrays of SiC wires along the radial direction are placed between the two surfaces of the cylindrical cloak.

[0099] The effective permeability of the system may be estimated as follows using the approach of O'Brien and Pendry (see O'Brien, S., and J. B. Pendry, J. Phys. Condens. Matter. 14, 4035-4044, 2002)

$$\mu_r = \frac{2}{kL_1^2}\frac{L_1J_1(kL_1) - tJ_1(kt) + a_0tH_1^{(1)}(kt) - a_0L_1H_1^{(1)}(kL_1) + c_0tJ_1(nkt)/n}{J_0(kL_2/2 - a_0H_0^{(1)}(kL_2/2)} \tag{43}$$

where h and ϕ represent the periodicities along the z and θ directions respectively, t denotes the radius of each wire, $n=\sqrt{e_{SiC}}$ is the refractive index, $k=2\pi/\lambda_0$ denotes the wave vector, $L_1=\sqrt{hr\phi/\pi}$ and $L_2=(h+r\phi)/2$ represent the two effective unit sizes based on area and perimeter estimations respectively. $a_0=[nJ_0(nkt)J_1(kt)-J_0(kt)J_1(nkt)]/[nJ_0(nkt)H_1^{(1)}(kt)-H_0^{(1)}(kt)J_1(nkt)]$ and $c_0=[J_0(kt)-a_0H_0^{(1)}(kt)]/J_0(nkt)$ are the scattering coefficients, and the Bessel functions in the equation follow the standard notations. The permittivity along the z direction may be approximated using Maxwell-Garnett method. In the design disclosed herein we choose the appropriate transformation geometry and operational wavelength such that the calculated effective parameters μ_r and $\in_z$ follow equation (35) with tolerable deviations. FIG. **8** shows the theoretically required and the calculated μ_r and $\in_z$ for a TE cloak at λ=13.5 μm. The parameters used for this calculation are a=15 μm, a/b=0.35, t=1.2 μm, h=2.8 μm, ϕ=10.6°, and the p coefficient in the quadratic transformation is $0.5a/b^2$. Good agreement between the required values and the calculated ones based on analytical formulae, and the imaginary part in the effective permeability is less than 0.06. This computation verifies the feasibility of the proposed cloaking system based on SiC wire arrays for the TE polarization. In FIG. **8** the magnetic parameter μ_r is calculated using equation (43), and the electric parameter $\in_z$ is obtained based on Maxwell-Garnett method.

[0100] In another aspect, a cloaking device structure may be a spherical or other shaped cloaking structure. The specific geometrical shape, the size and other design parameters of the structure, such as the spatial variation of material properties, may be chosen using the general approach described herein so as to be adaptable to the wavelength, the degree of cloaking, and the properties of the object to be cloaked. Loss and gain may be introduced in various portions of the structure.

[0101] The examples shown herein have used analytic profiles for the material properties so as to illustrate certain of the principles which may influence design of cloaking structures. However, since electromagnetic simulations using finite element methods, for example, are commonly used in design of complex shapes, and have been shown to yield plausible results, the use of such simulations are envisaged as useful in apparatus design. Ray tracing programs may be effectively used in situations where the spatial component of the material properties, and of the geometry, are slowly varying with respect to a wavelength at the operating frequencies. In optics, this is termed an adiabatic approximation.

[0102] Certain aspects, advantages, and novel features of the claimed invention have been described herein. It would be understood by a person of skill in the art that not all advantages may be achieved in practicing a specific embodiment. The claimed invention may be embodied or carried out in a manner that achieves or optimizes one advantage or group of advantages as taught herein without necessarily achieving other advantages as may have been taught or suggested.

[0103] It is therefore intended that the foregoing detailed description be regarded as illustrative rather than limiting, and that it be understood that it is the following claims, including all equivalents, that are intended to define the spirit and scope of this invention.

What is claimed is:

1. A apparatus for modifying the visibility properties of an object, comprising:

a structure formed of a metamaterial,

wherein the metamaterial properties are selected so that an electromagnetic wave incident on the apparatus is guided around the object at plurality of wavelengths.

2. The apparatus of claim **1**, wherein the structure is disposable between an object and an observer.

3. The apparatus of claim **1**, wherein the structure is comprised of a plurality of metamaterial layers, the layers having electromagnetic properties determined by one or more of the plurality of wavelengths.

4. The apparatus of claim **1**, wherein the structure includes a gain medium.

5. The apparatus of claim **4**, wherein the gain medium is a semiconductor capable of spontaneous emission at least one wavelength of the plurality of wavelengths.

6. The apparatus of claim **3**, where the material properties of each layer are selected so that the outermost boundary of a metamaterial layer of the plurality of metamaterial layers for each design wavelength is substantially coincident with the outer boundary of the structure.

7. The apparatus of claim **3**, where each layer is formed from a plurality of conformal layers.

8. The apparatus of claim **3**, wherein the location of an innermost boundary of a metamaterial layer of the structure is dependent on the design wavelength.

9. The apparatus of claim **3**, wherein a least a portion of the structure is formed of metamaterial layers wherein proximal layers have a effective refractive index of less than unity for an orthogonal polarization of an incident wave.

10. The apparatus of claim **3**, wherein the metamaterial properties of the layers at a first design wavelength are selected such that electromagnetic waves of a second design wavelength may penetrate into the structure so as to be guided by layers having metamaterial properties suitable for guiding the second design wavelength.

11. The apparatus of claim **10**, wherein gain medium layers are included in the layers of the first design wavelength so as to compensate for loss of the metamaterial at the second design wavelength.

12. The apparatus of claim **1**, wherein the structure is a sphere with an interior void containing an object to be cloaked.

13. The apparatus of claim **1**, wherein the structure is a cylinder of finite length having a symmetrical cylindrical void therein.

14. The apparatus of claim **1**, wherein the metamaterial comprises a dielectric having inclusions of a polaritonic material.

15. The apparatus of claim **14**, wherein the polaritonic material is silicon carbide (SiC).

16. The apparatus of claim **14**, wherein the polaritonic material is rod shaped, with the rods oriented in a radial direction.

17. The apparatus of claim **1**, wherein the metamaterial comprises a dielectric having inclusions of a metal.

18. The apparatus of claim **17**, wherein the metal and the dielectric are disposed in wedge shapes oriented in a radial direction.

19. A method of designing a structure for use as a cloak, comprising:

(a) selecting a design wavelength;

(b) selecting a metamaterial having the property of having a low loss at the design wavelength and at least a permeability or a permittivity of less than unity;

(c) determining, for a selected shape and size of structure, the variation of metamaterial properties as a function of position in the structure so as to guide electromagnetic waves of the design wavelength and polarization around a object disposed within the structure;

(d) selecting a second design wavelength and performing steps (a)-(c) for the second design wavelength.

20. The method of claim **19**, wherein the metamaterial includes a gain medium.

21. The method of claim **19**, wherein at least a portion of the metamaterial comprises a metamaterial effective at the first design wavelength, interspersed with a metamaterial effective at the second design wavelength.

22. The method of claim **19**, wherein the metamaterials for the first design wavelength and the second design wavelength are conformal alternating layers at least when proximal to an outer surface of the structure.

23. The method of claim **22**, wherein the step of determining, for a selected shape and size of structure, the variation of metamaterial properties as a function of position in the structure so as to guide electromagnetic waves of the design wavelength and polarization around a object disposed within the structure is performed iteratively to adjust the properties of the metamaterials at each design wavelength to account for the effect of the metamaterial selected for the other design wavelength.

24. The method of claim **22**, wherein the thickness of the metamaterial layers is small compared with either design wavelength.

25. A method of modifying the observability of an object, comprising:

providing a structure fabricated from a plurality of metamaterials, the metamaterials selected so as to guide electromagnetic waves around an object at a plurality of wavelengths; and

disposing the structure between an observer and the object.

* * * * *

US 20100277381A1

(19) **United States**
(12) **Patent Application Publication**
Smolyaninov et al.
(10) **Pub. No.: US 2010/0277381 A1**
(43) **Pub. Date: Nov. 4, 2010**

(54) **METAMATERIAL CLOAKED ANTENNA**

(75) Inventors: **Igor I. Smolyaninov**, Columbia, MD (US); **Nikolay K. Yushin**, Herndon, VA (US); **Louise C. Sengupta**, Ellicott City, MD (US)

Correspondence Address:
BAE SYSTEMS
PO BOX 868
NASHUA, NH 03061-0868 (US)

(73) Assignee: **BAE Systems Information and Electronic Systems Integration Inc.**, Nashua, NH (US)

(21) Appl. No.: **12/773,562**

(22) Filed: **May 4, 2010**

Related U.S. Application Data

(60) Provisional application No. 61/175,221, filed on May 4, 2009.

Publication Classification

(51) **Int. Cl.**
H01Q 19/06 (2006.01)
H01Q 1/32 (2006.01)

(52) **U.S. Cl.** **343/713**; 343/753

(57) **ABSTRACT**

In the method for operating a plurality antenna having near-field (NF) zone mounted on a vehicle, wherein the improvement comprises the step of cloaking in the near-field (NF) zone of each of said plurality of antennas.

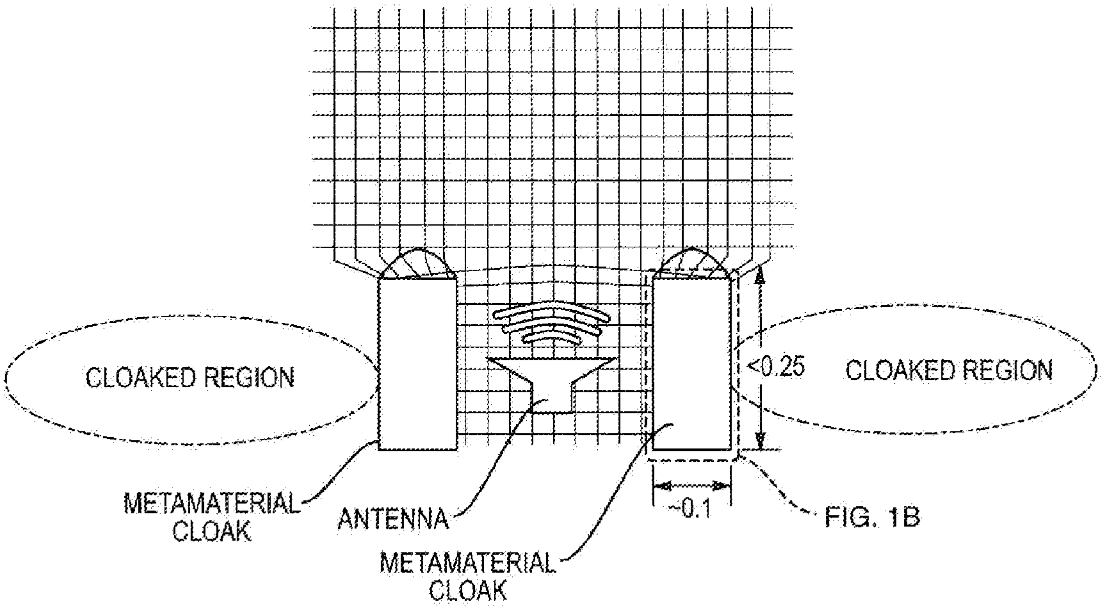

FIG. 1B

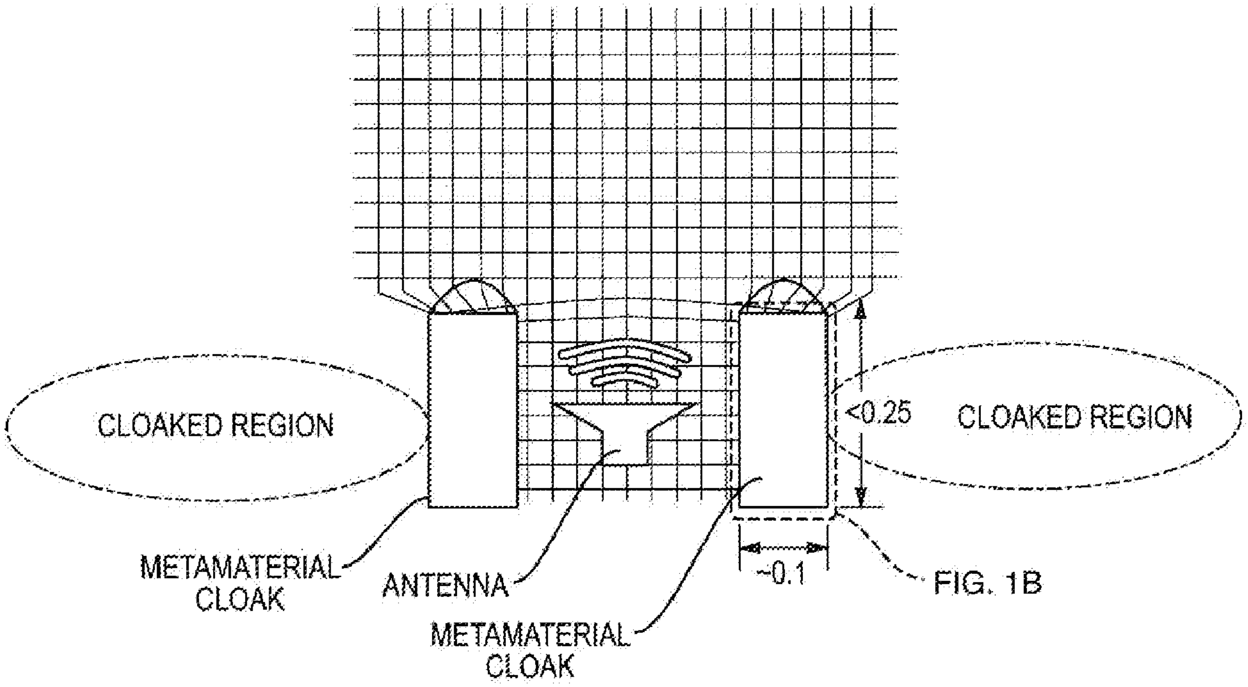
CLOAKED REGION
CLOAKED REGION
<0.25
~0.1
METAMATERIAL CLOAK
ANTENNA
METAMATERIAL CLOAK
FIG. 1B

FIG. 1A

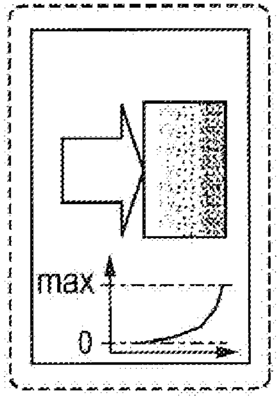
max
0

FIG. 1B

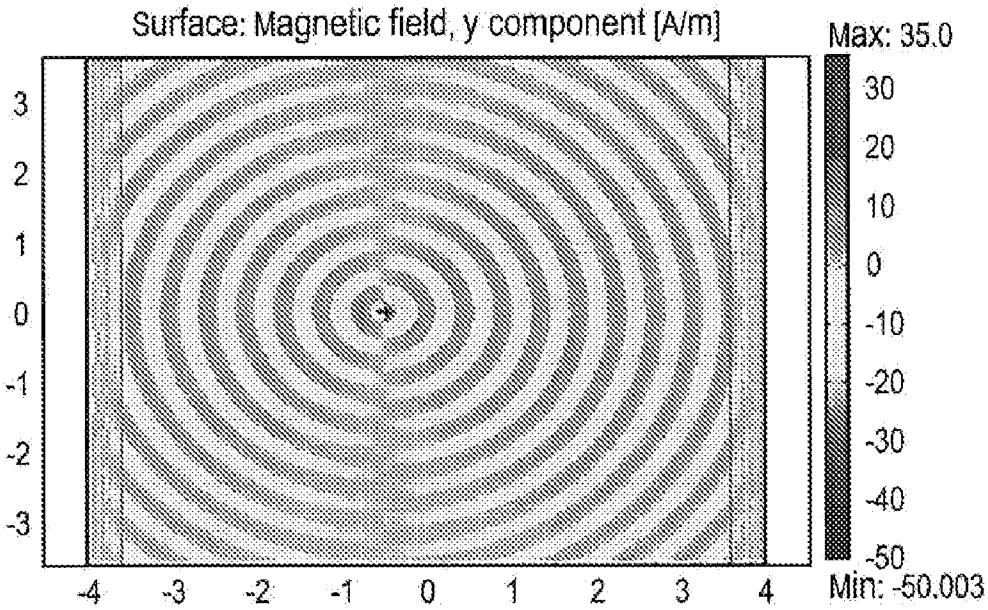
Surface: Magnetic field, y component [A/m]
Max: 35.0
30
20
10
0
-10
-20
-30
-40
-50
Min: -50.003
3
2
1
0
-1
-2
-3
-4
-3
-2
-1
0
1
2
3
4

FIG. 2A

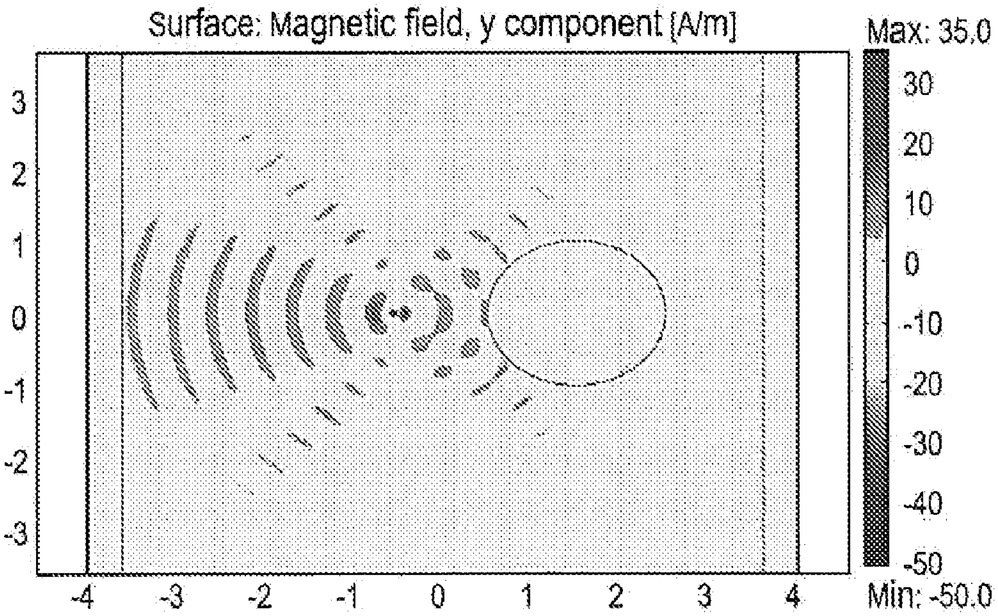
Surface: Magnetic field, y component [A/m]
Max: 35.0
30
20
10
0
-10
-20
-30
-40
-50
Min: -50.0
3
2
1
0
-1
-2
-3
-4
-3
-2
-1
0
1
2
3
4

FIG. 2B

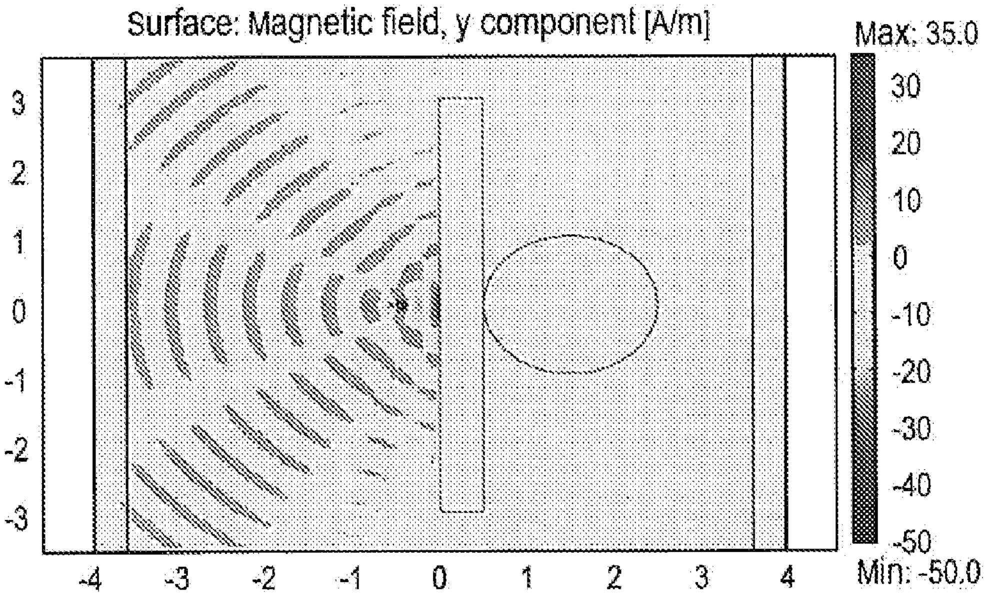
Surface: Magnetic field, y component [A/m]
Max: 35.0
30
20
10
0
-10
-20
-30
-40
-50
Min: -50.0
3
2
1
0
-1
-2
-3
-4
-3
-2
-1
0
1
2
3
4

FIG. 2C

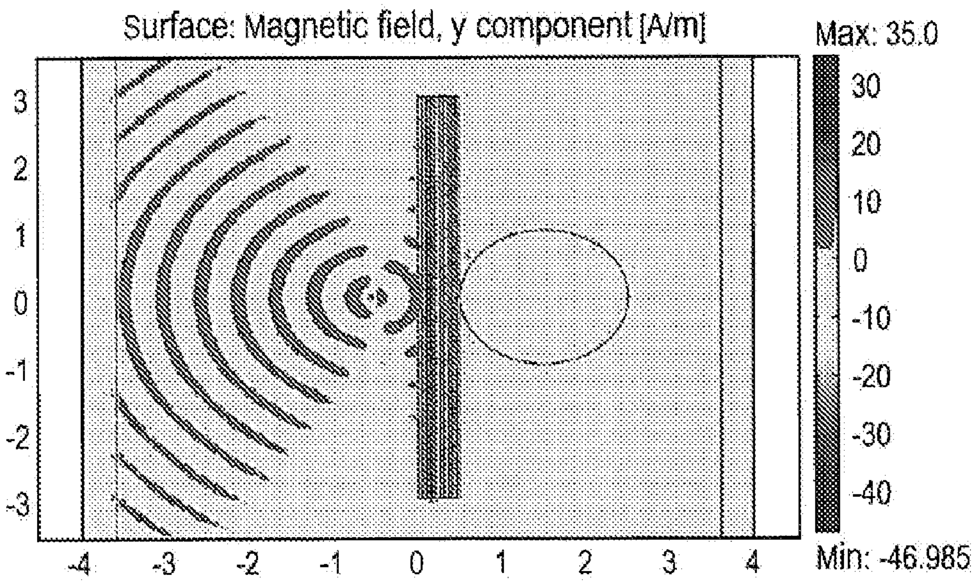
Surface: Magnetic field, y component [A/m]
Max: 35.0
30
20
10
0
-10
-20
-30
-40
Min: -46.985
3
2
1
0
-1
-2
-3
-4
-3
-2
-1
0
1
2
3
4

FIG. 2D

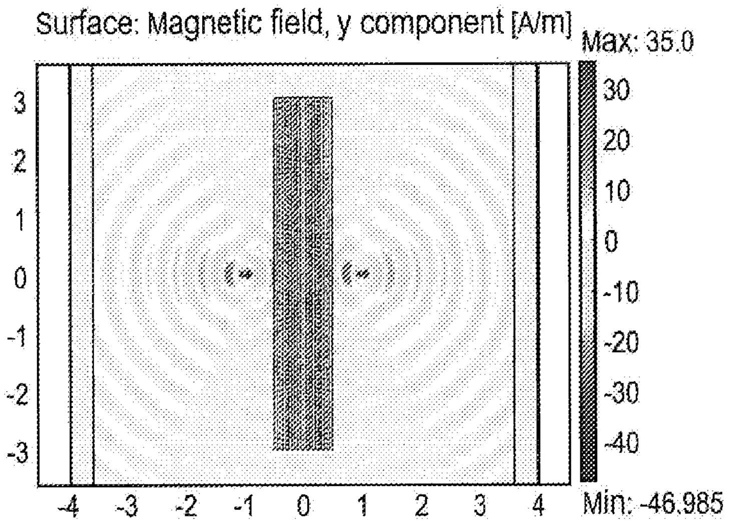
Surface: Magnetic field, y component [A/m]
Max: 35.0
30
20
10
0
-10
-20
-30
-40
Min: -46.985
3
2
1
0
-1
-2
-3
-4
-3
-2
-1
0
1
2
3
4

FIG. 3A

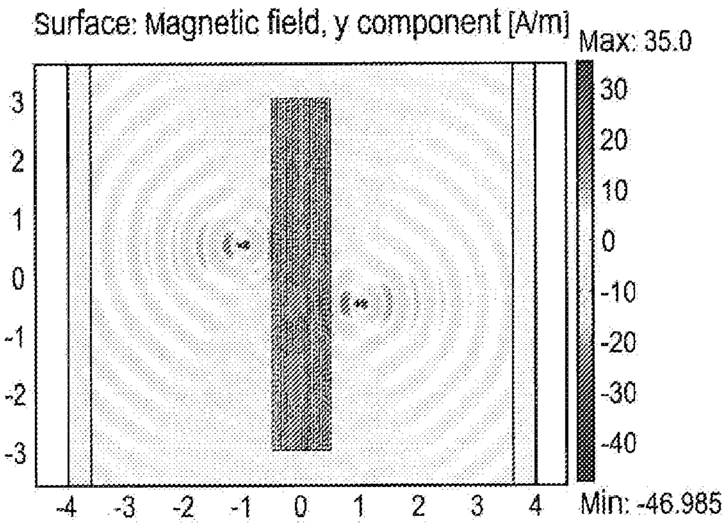
Surface: Magnetic field, y component [A/m]
Max: 35.0
30
20
10
0
-10
-20
-30
-40
Min: -46.985
3
2
1
0
-1
-2
-3
-4
-3
-2
-1
0
1
2
3
4

FIG. 3B

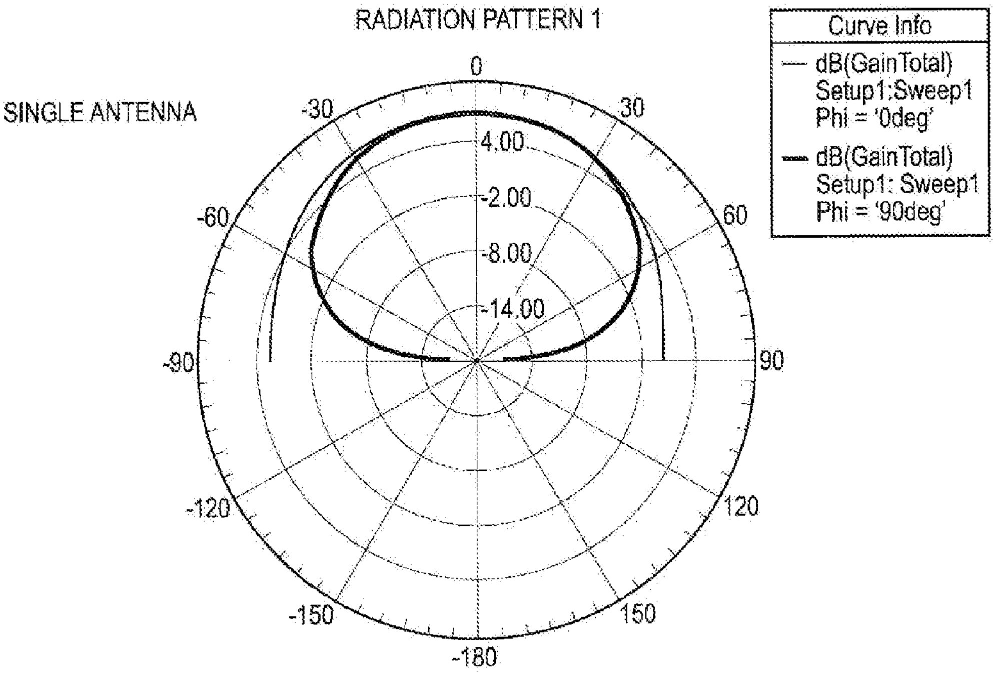
RADIATION PATTERN 1
SINGLE ANTENNA
Curve Info
dB(GainTotal)
Setup1:Sweep1
Phi = '0deg'
dB(GainTotal)
Setup1: Sweep1
Phi = '90deg'
0
30
60
90
120
150
-180
-150
-120
-90
-60
-30
4.00
-2.00
-8.00
-14.00

FIG. 4A

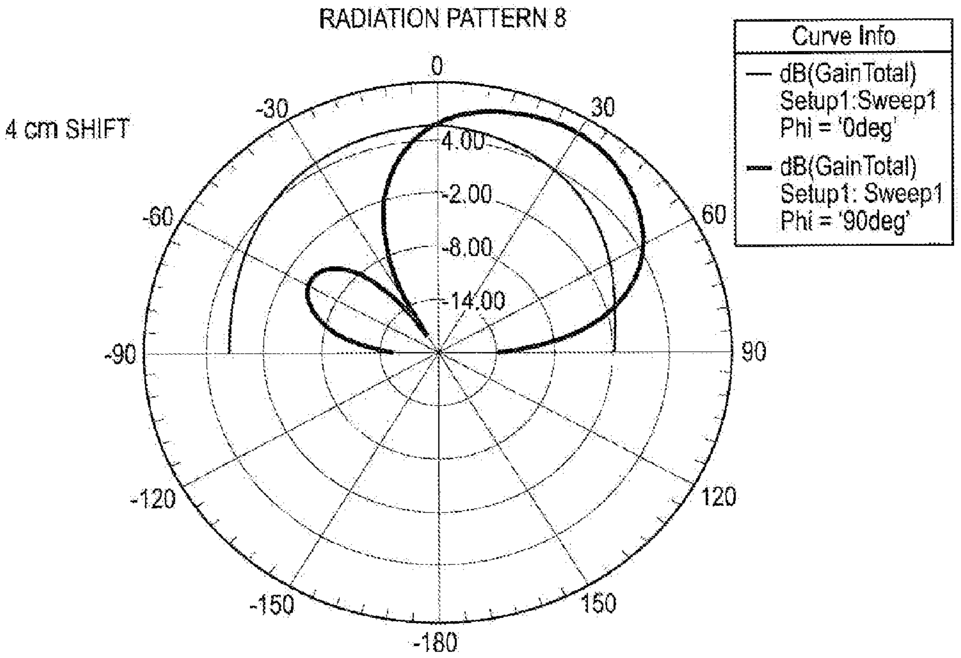
RADIATION PATTERN 8
4 cm SHIFT
Curve Info
dB(GainTotal)
Setup1:Sweep1
Phi = '0deg'
dB(GainTotal)
Setup1: Sweep1
Phi = '90deg'
0
30
60
90
120
150
-180
-150
-120
-90
-60
-30
4.00
-2.00
-8.00
-14.00

FIG. 4B

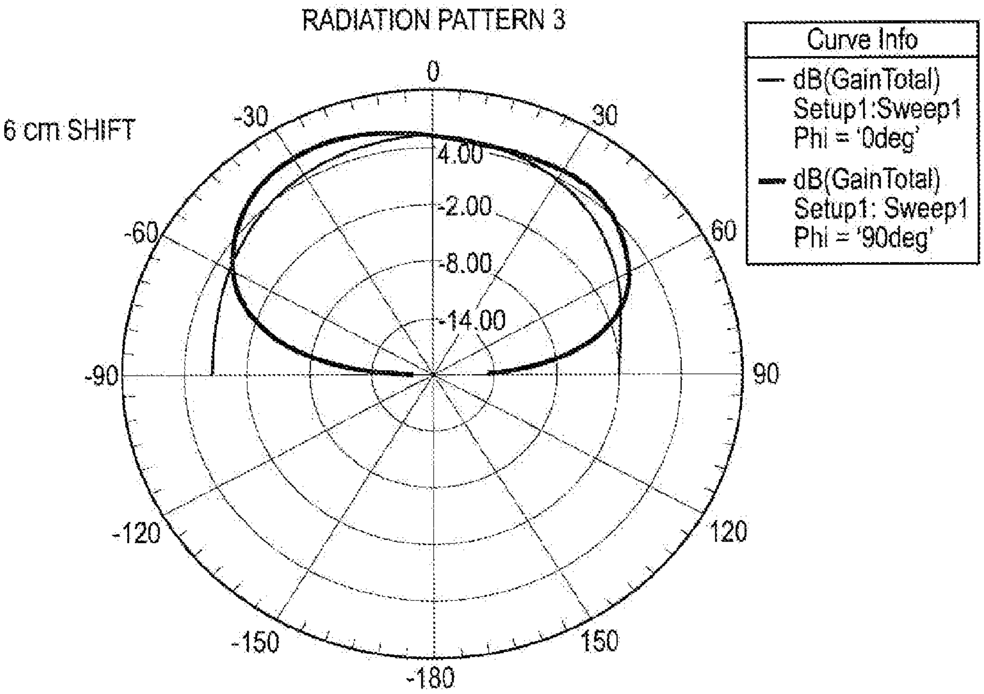
RADIATION PATTERN 3
6 cm SHIFT
Curve Info
dB(GainTotal)
Setup1:Sweep1
Phi = '0deg'
dB(GainTotal)
Setup1: Sweep1
Phi = '90deg'
0
30
60
90
120
150
-180
-150
-120
-90
-60
-30
4.00
-2.00
-8.00
-14.00

FIG. 4C

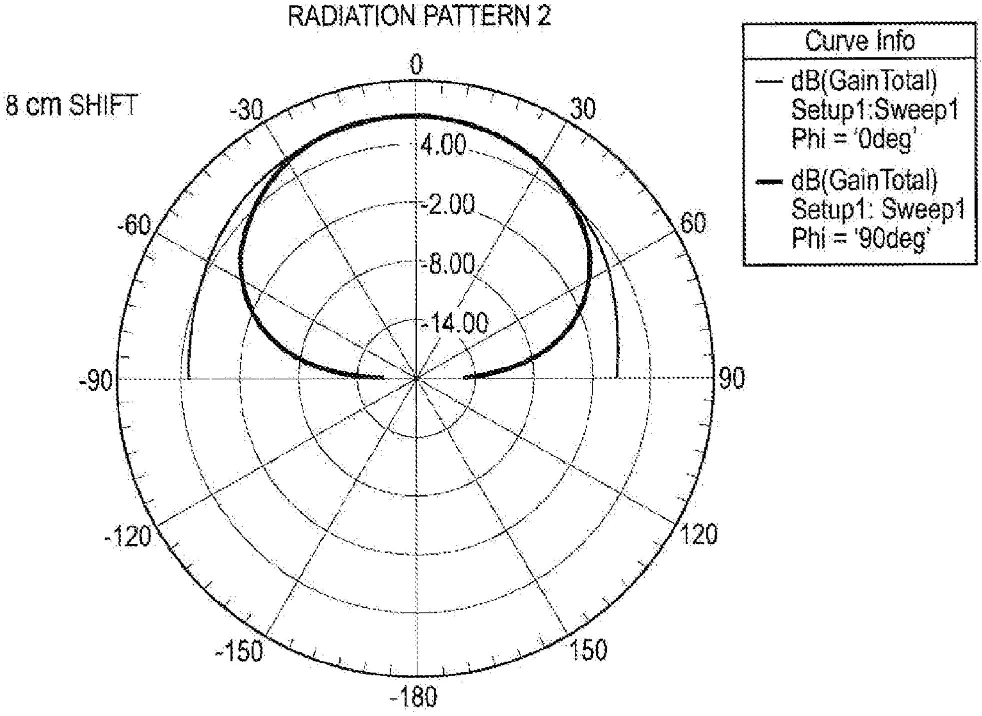
RADIATION PATTERN 2
8 cm SHIFT
Curve Info
dB(GainTotal)
Setup1:Sweep1
Phi = '0deg'
dB(GainTotal)
Setup1: Sweep1
Phi = '90deg'
0
30
60
90
120
150
-180
-150
-120
-90
-60
-30
4.00
-2.00
-8.00
-14.00

FIG. 4D

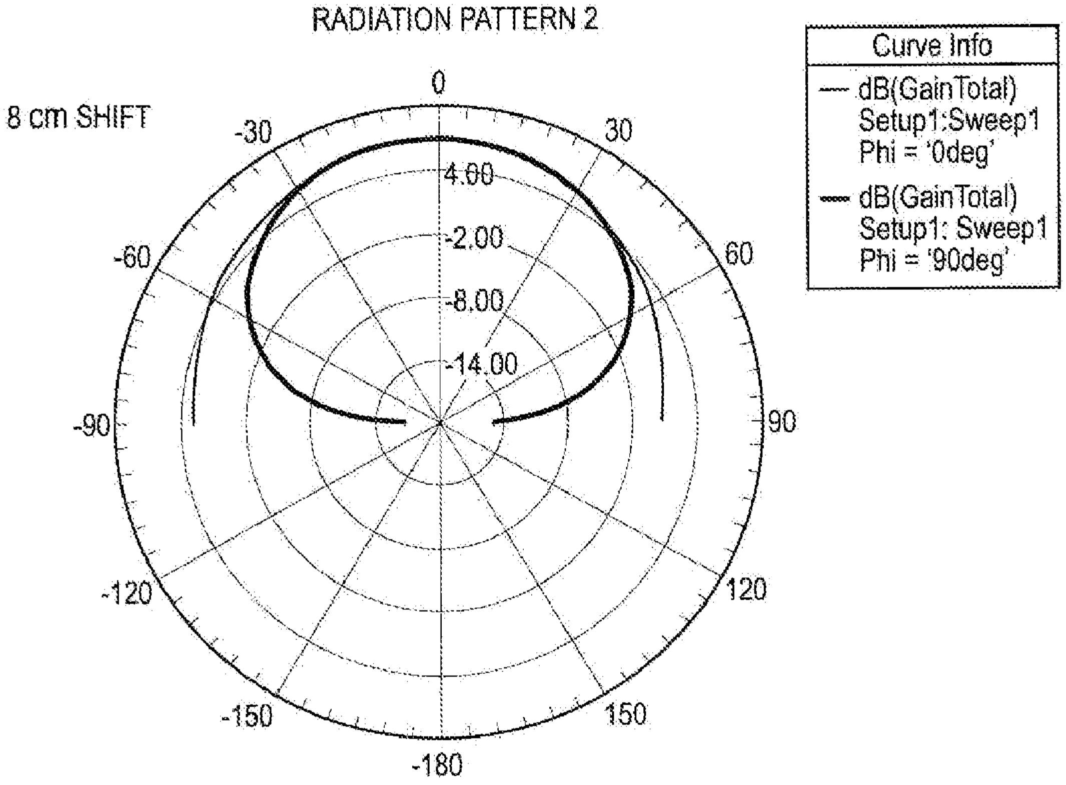
RADIATION PATTERN 2
8 cm SHIFT
Curve Info
— dB(GainTotal) Setup1:Sweep1 Phi = '0deg'
— dB(GainTotal) Setup1: Sweep1 Phi = '90deg'
0
30
60
90
120
150
-180
-150
-120
-90
-60
-30
4.00
-2.00
-8.00
-14.00

FIG. 4E

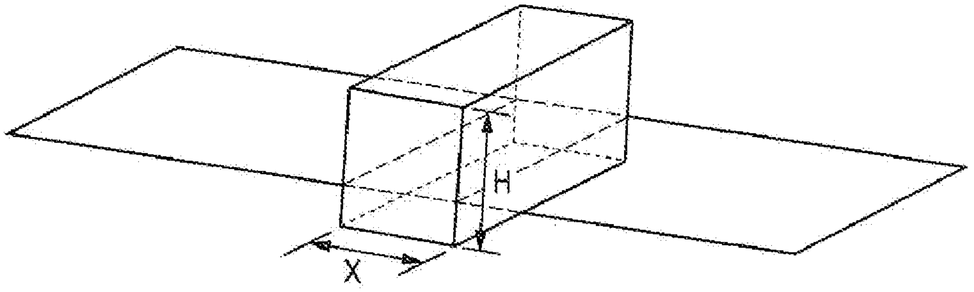
H
X

FIG. 4F

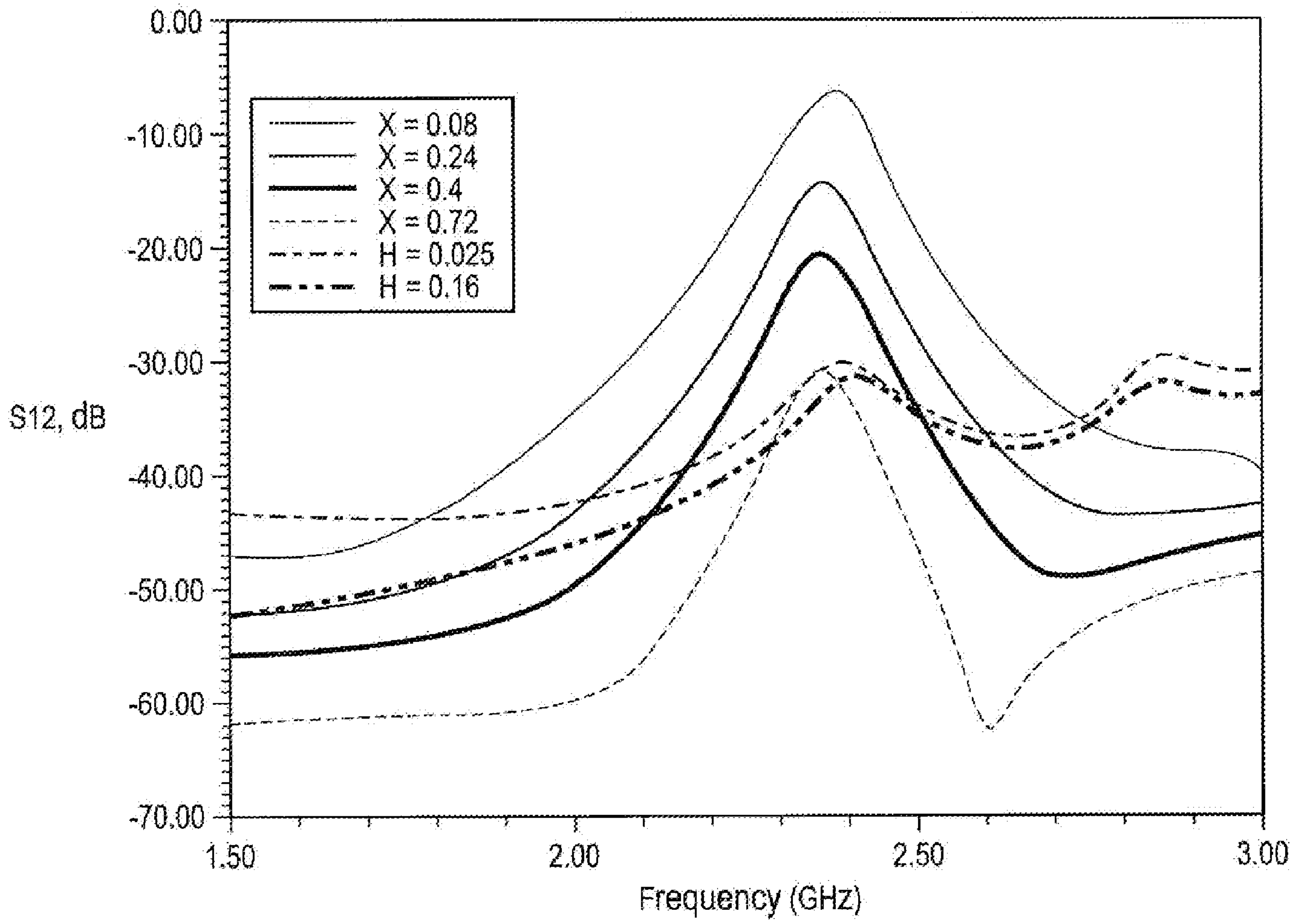
X = 0.08
X = 0.24
X = 0.4
X = 0.72
H = 0.025
H = 0.16
S12, dB
0.00
-10.00
-20.00
-30.00
-40.00
-50.00
-60.00
-70.00
1.50
2.00
2.50
3.00
Frequency (GHz)

FIG. 5

METAMATERIAL CLOAKED ANTENNA

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims rights under 35 U.S.C. §119 (e) from U.S. Application Ser. No. 61/175,221, filed. May 4, 2009, the contents of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

[0002] 1. Field of the Invention

[0003] The present invention relates to electromagnetic cloaking and more particularly to metamaterial cloaked antennas.

[0004] 2. Brief Description of Prior Developments

[0005] Prior art metamaterial-based cloaking techniques are described in the following publications, the contents of which are incorporated herein by reference:

[0006] 1. J. B. Pendry, D. Schurig, D. R. Smith, *Science* 312, 1780 (2006).

[0007] 2. U. Leonhardt, *Science* 312, 1777 (2006).

[0008] 3. D. Schurig, J. J. Mock, B. J. Justice, S. A. Cummer, J. B. Pendry, A. F. Starr, D. R. Smith, *Science* 314, 977 (2006).

[0009] 4. W. Cai, U. K. Chettiar, A. V. Kildishev, V. M. Shalaev, *Nature Photonics* 1, 224 (2007).

[0010] The performance of modern and future aircraft and satellites strongly depends on various antennas. Typically, a very large number of different types of antennas are mounted on board aircraft, satellites or other vehicles in order to cover as wide a radio-frequency range and bandwidth as is possible. Because space is at a premium on board, the airborne antennas are usually located in close proximity to each other. This arrangement may cause so-called estate issues with respect to antenna installation.

[0011] A need, therefore, exists for a way of reducing the detrimental effect of co-site interference between multiple antennas densely mounted on aircraft, satellites, ships or other vehicles.

SUMMARY OF INVENTION

[0012] According to the present invention, engineered metamaterial (MM) layers are applied to antenna surfaces to prevent the reflected/scattered radiation return on the active antenna surface. The cloaking effect is based on the ability of specially designed MM devices to prevent reflection of the electromagnetic radiation rays in such manner that the electromagnetic waves go around the object and make the object essentially "invisible". A transformational optics approach is used for the cloaked antenna design. An important feature of method of the present invention is cloaking in the near-field (NF) zone of an antenna. The radiation field distribution in the NF region of an antenna is much more complex than in the far-field (FF) zone. Therefore, the MM structure has to be designed in a different way than has been demonstrated in early cloaks for the FF zone.

[0013] This invention also encompasses the method of improving the performance of multiple antennas mounted on board of the aircraft, satellite, ship, or any vehicle by separating the near field (NF) radiation by the metamaterial cloaks positioned between the neighbor antennas (see FIGS. **1** and **2**). As a result the far field (FF) radiation patterns are significantly improved.

[0014] This invention also encompasses a cloaked antenna which directs the antenna radiation in such manner that it does not reach the neighbor objects and/or other antennas (see FIG. **2**). We would like to emphasize that the metamaterial cloak developed for the purposes of this project has a different configuration than to the conventional cloaks described by the Pendry et al. and Schurig et al. references cited above.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] The invention will be more closely understood hereinafter by reference to the drawings in which:

[0016] FIGS. **1**A and **1**B show drawings of an antenna cloaking by the coordinate transformation method. By selecting conformal coordinate transformation to curvilinear coordinate system schematically indicated by the black mesh, the donut-shaped region of space on the sides of the antenna is "eliminated": from the point of view of the radiating antenna, the other antennas and/or passive sections of the aircraft located in the donut-shaped cloaked region are located at infinity. As a result, the radiation pattern of the cloaked antenna is insensitive to its ambient. The inset shows gradual change of the metamaterial parameters inside the cloak wall.

[0017] FIGS. **2**A, **2**B, **2**C and **2**D show a COMSOL simulation of the dipole near-field distribution for a single dipole (a), the dipole and metallic cylinder (b), the metallic plate positioned between the dipole and the cylinder (c), and the metamaterial flat cloak located between the dipole and the cylinder (d). In real applications the edges of the cloak will be tapered in order to reduce scattering.

[0018] FIGS. **3**A and **3**B show a COMSOL simulation of the NF H-field of the two dipole antennas separated by the metamaterial cloak.

[0019] FIGS. **4**A, **4**B, **4**C, **4**D and **4**F show the FF radiation pattern calculated for a single patch antenna (a), the active antenna with a neighbor passive antenna at X=0.08λ (b), X=0.24λ (c), and X=0.72λ (d). The FF radiation pattern of the two antennas at X=0.08λ with the metamaterial cloak between them (e). The MM clock (with the width X and the height H) shown green is positioned between the two patch antennas shown pink (f).

[0020] FIG. **5** shows the transmission coefficient S_{12} of the two patch antennas separated by the distance X=0.08λ (a), X=0.24λ (b), X=0.4λ (c), and X=0.72λ (d), as well as with the metamaterial cloak of the height H between the antennas at X=0.08λ and H=0.025λ (e), X=0.08λ and H=0.16λ (f).

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

[0021] Based on the transformational optics approach we have developed a novel concept of the cloaked antenna, which is illustrated in FIG. **1**. The transformational optics approach is based on the conformal coordinate transformations to a specially chosen system of curvilinear coordinates (x_1,y_1,z_1,t_1), which distort the length elements but keep the angles intact. The Maxwell equations in such a curvilinear coordinate system may be written in the usual form by introducing effective anisotropic dielectric permittivity $\epsilon(x_1,y_1,z_1)$ and magnetic permeability $\mu(x_1,y_1,z_1)$ tensors of free space. On the other hand, if the so obtained ϵ and μ distributions will be engineered artificially using specially designed metamaterial medium, electromagnetic fields in such a medium will propagate as if in normal Minkowski space described by the (x_1,y_1,z_1,t_1) coordinates. Thus, transformational optics approach to antenna cloaking is applicable to both near and far fields of a cloaked antenna.

[0022] The schematic drawing in FIG. **1** represents our conceptual approach. In real-life applications the exact configuration of the cloak needs to be tailored in order to avoid undesirable blind spots. The approximate dimensions in FIG.

1 indicate that not much of the antenna field of view will be obscured. Those skilled in the art will appreciate that if any particular application would require an omnidirectional field of view, a combination of our approach with the approach disclosed in the Pendry et al. reference (cited above) may be used, so that antenna may see behind the obscuration.

[0023] The distinctive feature of the metamaterial for cloaking applications is that its properties (dielectric permittivity c and magnetic permeability μ) are gradually changed inside the cloak, while their ratio does not. It is matched to the impedance of free space. FIG. **1** illustrates our MM cloak that is essentially multilayered dielectric structure (with some metallic particles inclusions) each layer ϵ (and μ) being extrapolated by the step function close to desirable ϵ and μ distribution. Our simulations show that significant field distortion is observed when the radiation from the dipole hits metallic objects (e.g. metal cylinder or metal plate), while in the case of the metamaterial cloak positioned between the dipole and the metal object there is no such field distortion (compare FIGS. **2***b*, **2***c* and FIG. **2***d*). Thus, screening the objects by metallic plate or cylinder does not prevent the radiation pattern distortion. On the other hand, two antennas may work independently without interference with each other when each antenna is cloaked, even though they are situated next to each other (FIG. **3**).

[0024] The general algorithm for designing the MM cloak is the following:

Using the desired conformal coordinate transformation tensor we calculate the spatial variations of components of the permittivity $\epsilon(r)$ tensor and the permeability $\mu(r)$ tensor. From the so obtained $\epsilon(r)$ and $\mu(r)$ spatial distribution, the structure and composition of the metamaterial is determined. To simplify the cloak fabrication we will use the multilayered structure (as in the inset in FIG. **1**) where the $\epsilon(r)$ and $\mu(r)$ function are replaced by the step-like approximation.

[0025] An underlying feature of the proposed approach is the use of the metamaterial cloaks to prevent the electromagnetic radiation from the neighbor objects (active and/or passive ones) from hitting the emitting surface of the antenna(s). This returning radiation creates the phase distortion on the surface and deteriorates the FF radiation pattern of the antenna. In the conventional approach to antenna separation one has to position the antennas in such a way that the lateral distance between them is greater than the NF zone as is described in Handbook of Antennas in Wireless Communications, Ch. 5, Antenna Parameters, Various Generic Antennas and Feed Systems, and Available Softwares, CRC, 2002, the contents of which are incorporated herein by reference, size d_{NF} ($d_{NF}>>D$, $d_{NF}>>\lambda$, and $d_{NF}>>2D^2/\lambda$, here D is the antenna size and λ is its wavelength). Therefore, there is a room for $(D+d_{NF})^{-2}$ antennas per unit area. With the proposed metamaterial cloaks the distance d_{MM} between the antenna edges could be 0.08λ or less. Then we may put $(D+d_{MM})^2$ antennas per unit area. For $d_{MM}<<D$ the metamaterial clock permits the position of much more antennas than a conventional spatial separation.

[0026] To evaluate significance of our approach the antenna FF radiation pattern distortion and its correction with the metamaterial cloak have been simulated by HFSS software for a set of two antennas—one is an active patch antenna and second is a passive patch antenna located on a distance X from the edge of the active antenna.

[0027] The results presented in FIG. **4** show the radiation pattern of the active antenna is drastically deteriorated at low X, but the pattern coincides with a single antenna pattern when the distance X becomes larger. Meanwhile the metamaterial clock positioned between the antennas makes the pattern looks as a single antenna one even at $X=0.08\lambda$ (Compare FIGS. **4***a, d*, and *e*). Using HFSS we have also calculated the S-parameters of the set of the two identical patch antennas located at the distance X. One may see from the results presented in FIG. **5** that the closer antennas the higher the transmission S_{12}-parameter. Meanwhile the metamaterial cloak positioned between the antennas significantly decreases the S_{12} magnitude even at the distance as small as 0.08λ with a low value of the scattering S_{11}-parameter.

[0028] Thus, our HFSS simulations demonstrate that the proposed metamaterial cloak of the thickness of only 0.08λ works as the separation of the antennas by the distance of 0.72λ.

[0029] Those skilled in the art will appreciate that any combination of metal, dielectric, ferroelectric and/or ferromagnetic materials may be used in the antenna cloak design, as long as the resulting metamaterial parameters would satisfy the desired spatial distribution of ϵ and μ tensors, as described in FIG. **1**.

[0030] While the present invention has been described in connection with the preferred embodiments of the various figures, it is to be understood that other similar embodiments may be used or modifications or additions may be made to the described embodiment for performing the same function of the present invention without deviating therefrom. Therefore, the present invention should not be limited to any single embodiment, but rather construed in breadth and scope in accordance with the recitation of the appended claims.

What is claimed is:

1. In the method for operating a plurality antenna having near-field (NF) zone mounted on a vehicle, wherein the improvement comprises the step of cloaking in the near-field (NF) zone of each of said plurality of antennas.

2. The method of claim **1** wherein near-field (NF) radiation of neighboring radiation is separated by a plurality of metamaterial cloaks so that the far field (FF) radiator patterns are improved.

3. The method of claim **1** wherein a cloaked antenna is provided which directs radiation in such a way that said radiation is impeded from reaching neighboring antennas.

4. The method of claim **1** wherein a metamaterial cloak is used.

5. The method of claim **4** wherein using a desired conformal coordinate transformation tensor the spatial variations of components of the permittivity $\epsilon(r)$ tensor and the permeability $\mu(r)$ tensor is calculated.

6. The method of claim **5** wherein from $\epsilon(r)$ and $\mu(r)$ spatial distribution, the structure and composition of the metamaterial is determined.

7. The method of claim **6** wherein a multilayered cloak structure is used wherein $\epsilon(r)$ and $\mu(r)$ function are approximated by a step-like approximation.

8. The method of claim **4** wherein the metamaterial cloaks of the antennas prevent electromagnetic radiation from hitting emitting surfaces of the antenna.

* * * * *

June 9, 1925. **1,540,998**

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 1

Fig. 1.

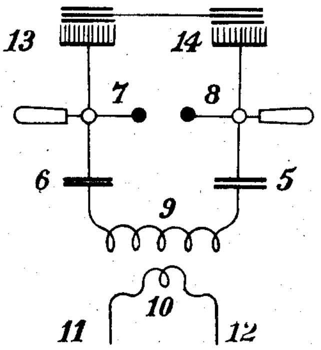
13
14
7
8
6
5
9
10
11
12

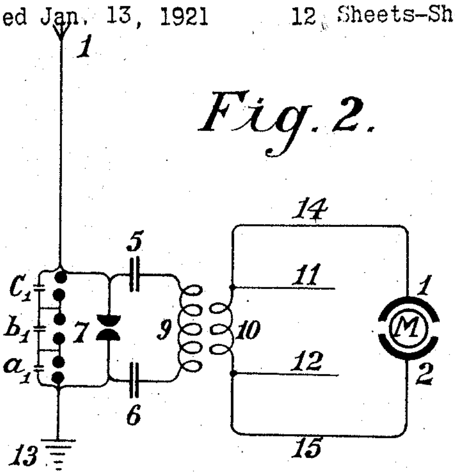
1
Fig. 2.
14
5
11
C_1
b_1
7
9
10
1
M
a_1
12
2
6
13
15

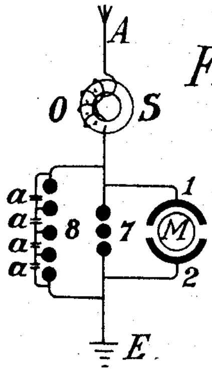
A
O
S
a
a
a
a
8
7
1
M
2
E.

Fig. 3.

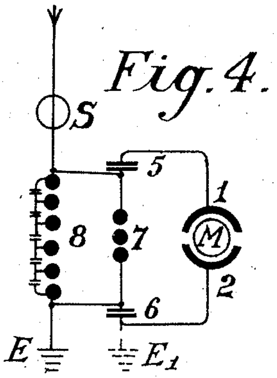
S
Fig. 4.
5
1
8
7
M
2
6
E
E_1

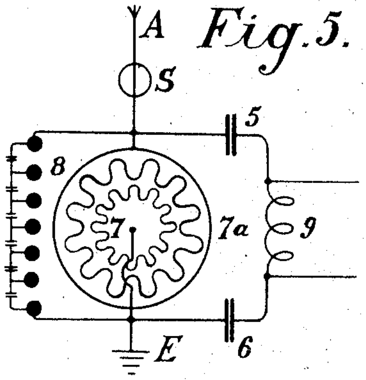
A
S
5
8
7
7a
9
E
6

Fig. 5.

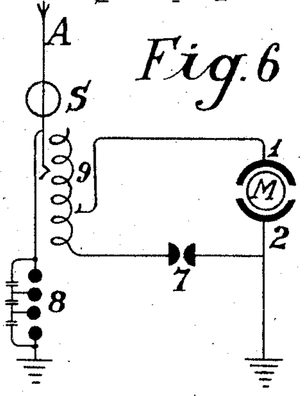
A
S
Fig. 6
9
1
M
2
7
8

Inventor
Hermann Plausen

by Knight Bros
attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 2

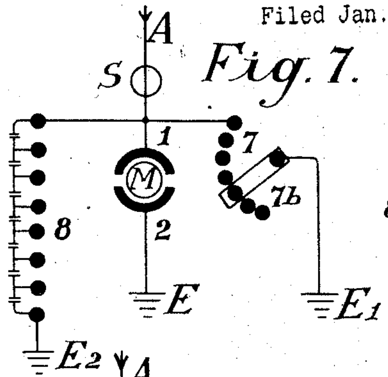
Fig. 7.
A
S
1
2
7
7b
8
E
E1
E2

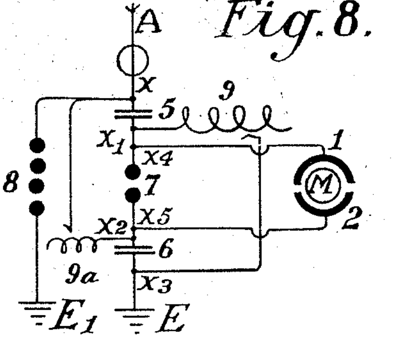
Fig. 8.
A
X
5
9
X1
X4
1
7
8
X5
2
X2
6
9a
X3
E1
E

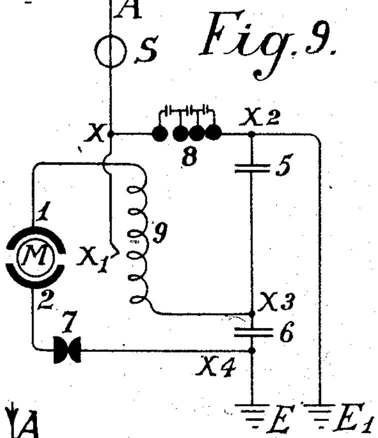
Fig. 9.
A
S
X
8
X2
5
1
9
X1
2
X3
6
7
X4
E
E1

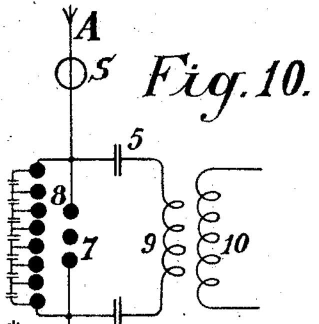
Fig. 10.

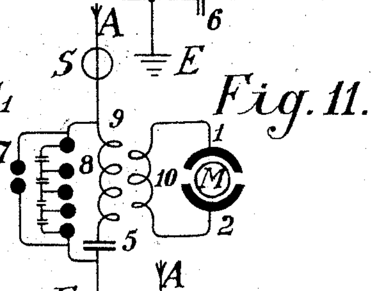
Fig. 11.
A
S
E
9
1
7
8
10
2
5

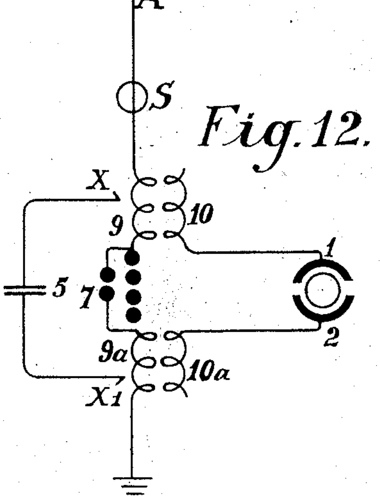
Fig. 12.
A
S
X
9
10
1
5
7
2
9a
10a
X1

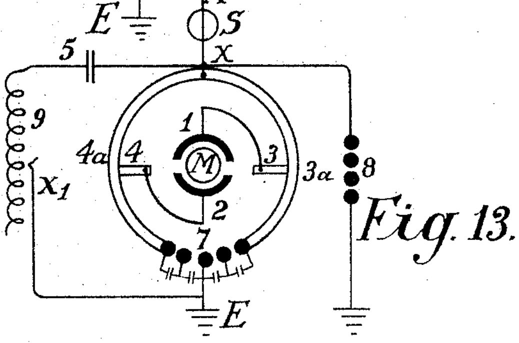
Fig. 13.
A
S
X
5
9
1
4a
4
3
3a
8
X1
2
7
E

Inventor

Normann Plauson

by Knight Bros
attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 3

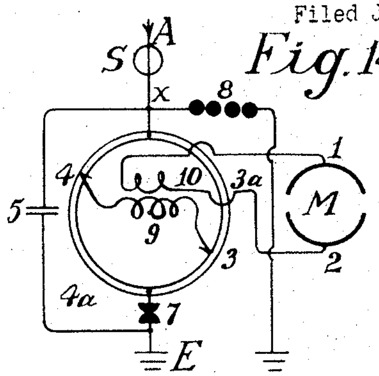
S
A
x
8
1
4
10
3a
M
5
9
3
2
4a
7
E

Fig. 14.

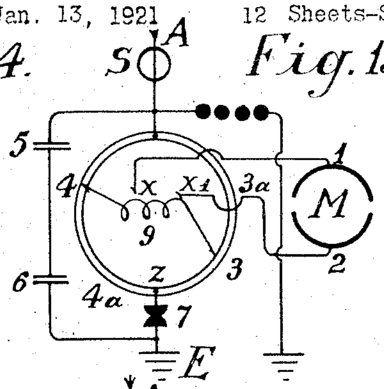
S
A
5
1
4
x
x1
3a
M
9
3
2
6
z
4a
7
E

Fig. 15.

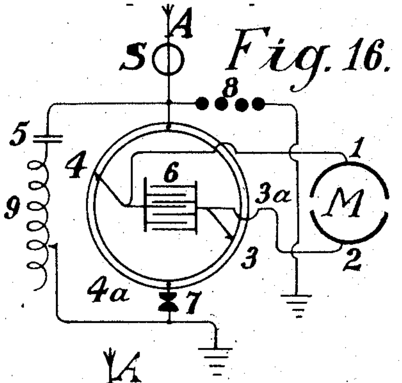
S
A
8
5
1
4
6
3a
M
9
3
2
4a
7

Fig. 16.

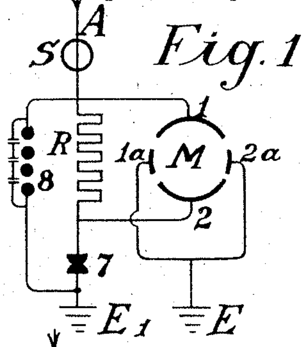
S
A
1
R
1a
M
2a
8
2
7
E1
E

Fig. 17.

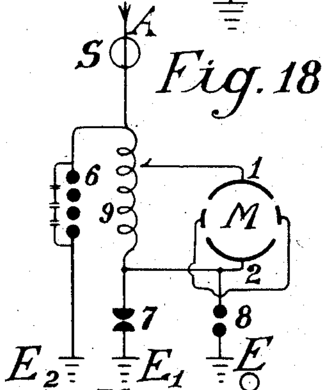
S
A
1
6
9
M
2
7
8
E2
E1
E

Fig. 18.

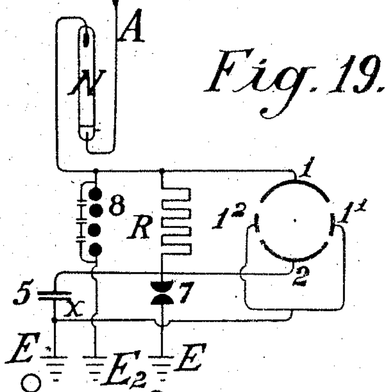
A
N
1
8
R
1²
1¹
2
5
x
7
E
E2
E

Fig. 19.

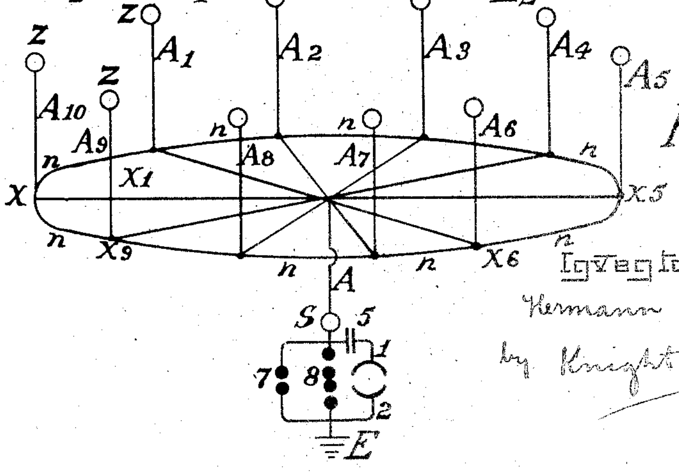
Z
Z
Z
A1
A2
A3
A4
A5
A10
A9
n
n
n
A8
A7
A6
n
x
x1
x5
n
x9
n
n
x6
A
S
5
1
7
8
2
E

Fig. 20.

Inventor
Hermann Plauson
by Knight Bros
Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 4

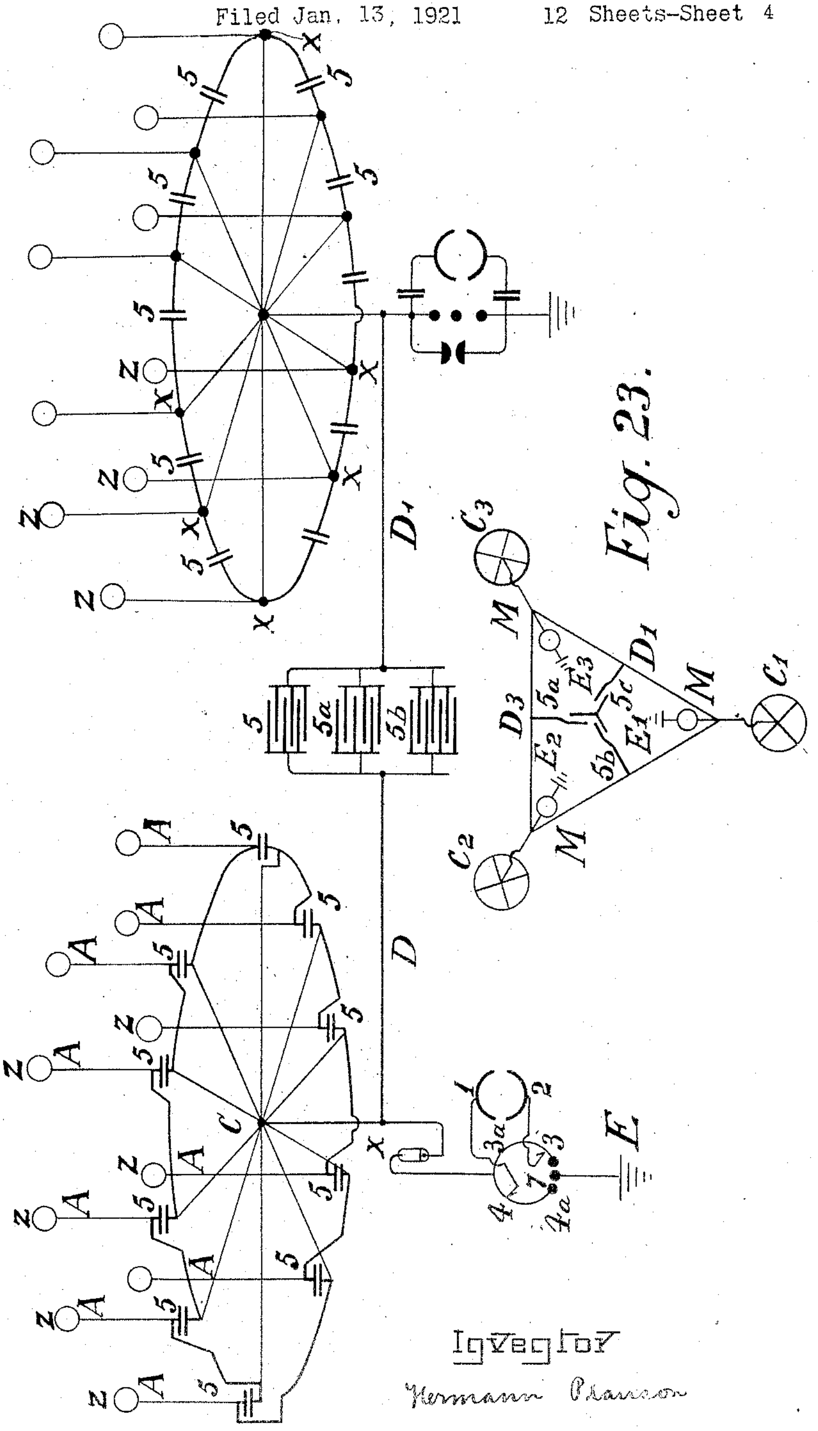
5
X
Z
D1
D
5a
5b
5c
C
A
1
2
3
3a
4
4a
7
E
C1
C2
C3
M
D1
D3
E1
E2
E3

Fig. 22.

Fig. 21.

Fig. 23.

Inventor
Hermann Plauson
by Knight Bros
Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 5

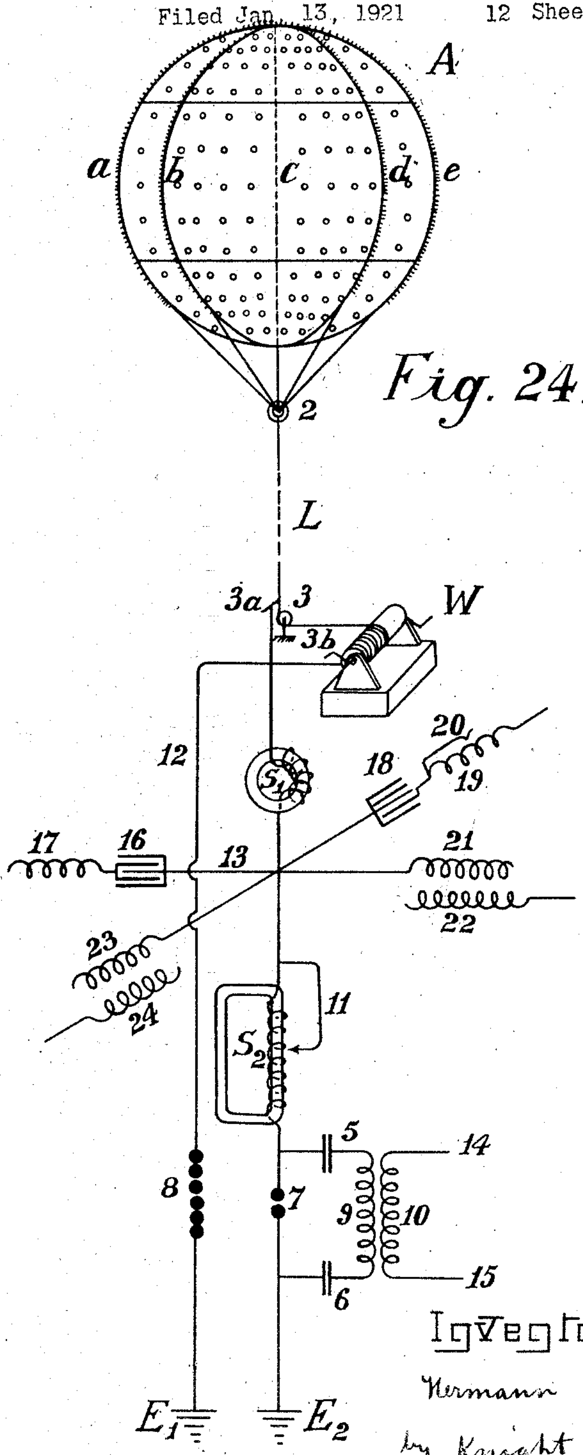
A
a
b
c
d
e
2
L
3a
3
3b
W
12
S1
18
20
19
17
16
13
21
22
23
24
11
S2
5
14
8
7
9
10
15
6
E1
E2

Fig. 24.

Inventor
Hermann Plauson
by Knight Bros
Attorneys

June 9, 1925. **1,540,998**

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 6

Fig. 25.

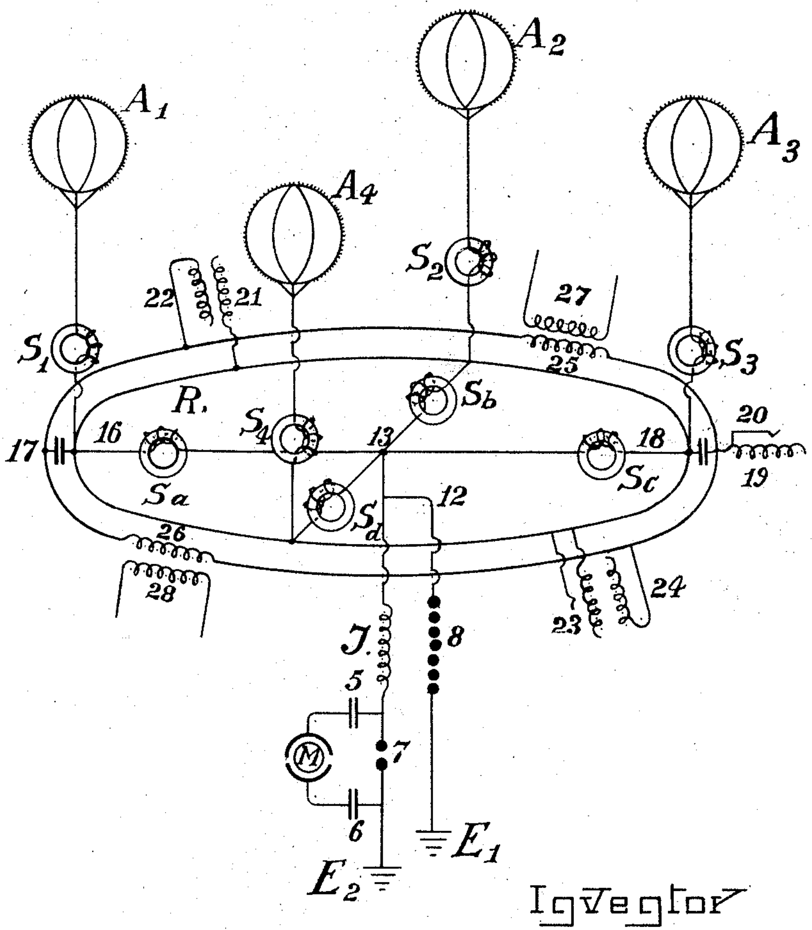

A1
A2
A3
A4
S1
S2
S3
S4
Sa
Sb
Sc
Sd
R.
22
21
27
25
16
13
18
17
20
19
12
26
28
23
24
J.
5
8
M
7
6
E1
E2

Inventor

Hermann Plauson

by

Knight Bros

attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 7

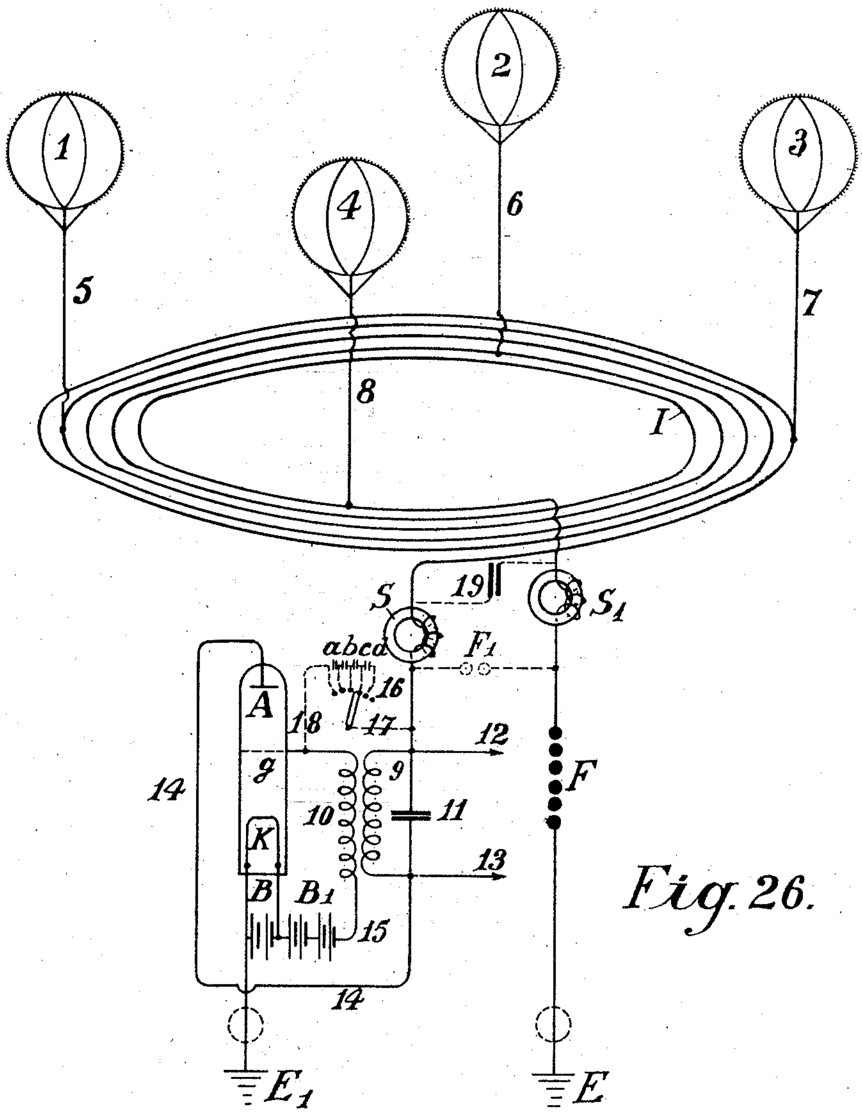
1
2
3
4
5
6
7
8
I
S
19
S1
abcd
F1
16
A
18
17
12
g
9
14
10
11
K
13
B
B1
15
14
F
E1
E

Fig. 26.

Inventor

Hermann Plauson

by Knight Bros

Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921. 12 Sheets-Sheet 8

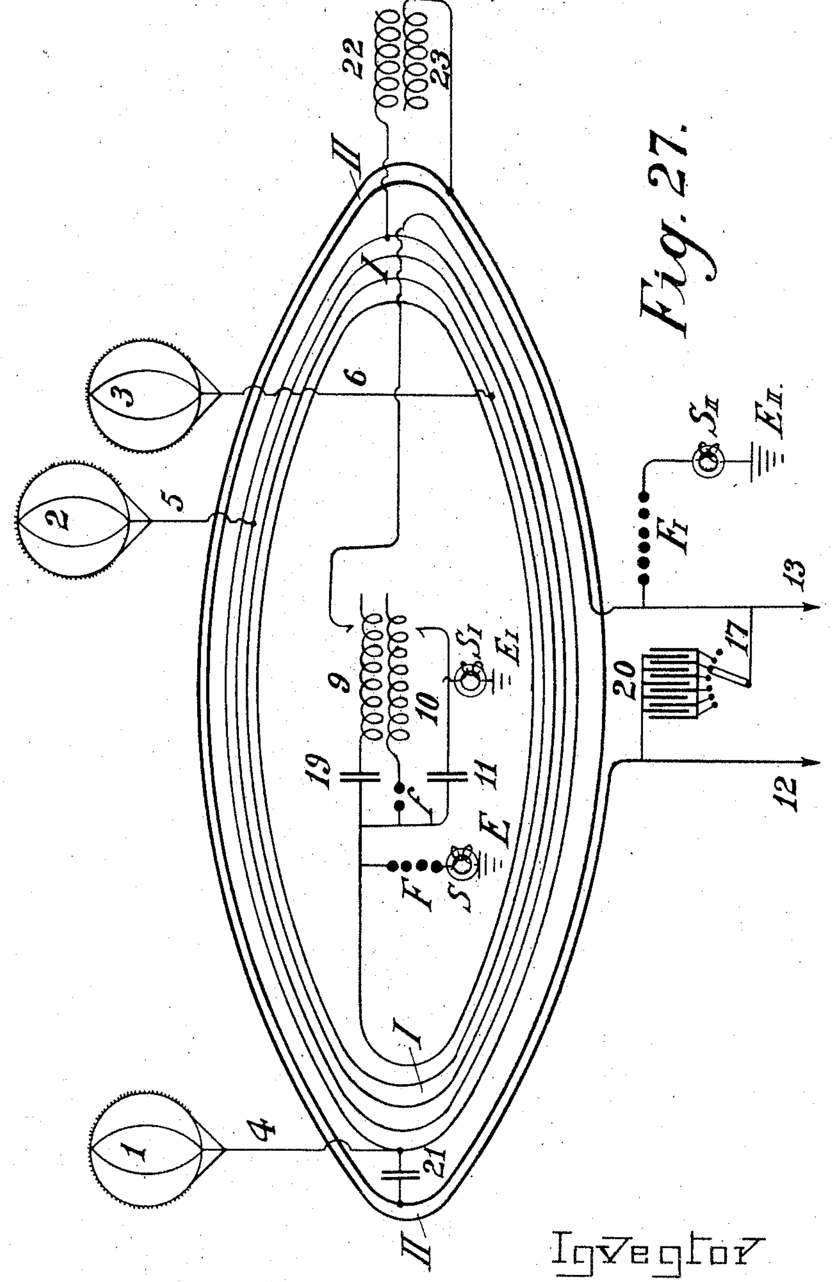
22
23
II
I
6
3
2
5
S_II
E_II
F_I
13
9
S_I
E_I
10
20
17
19
11
12
F
E
S
I
1
4
21
II

Fig. 27.

Inventor

Hermann Plauson

by Knight Bros

Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 9

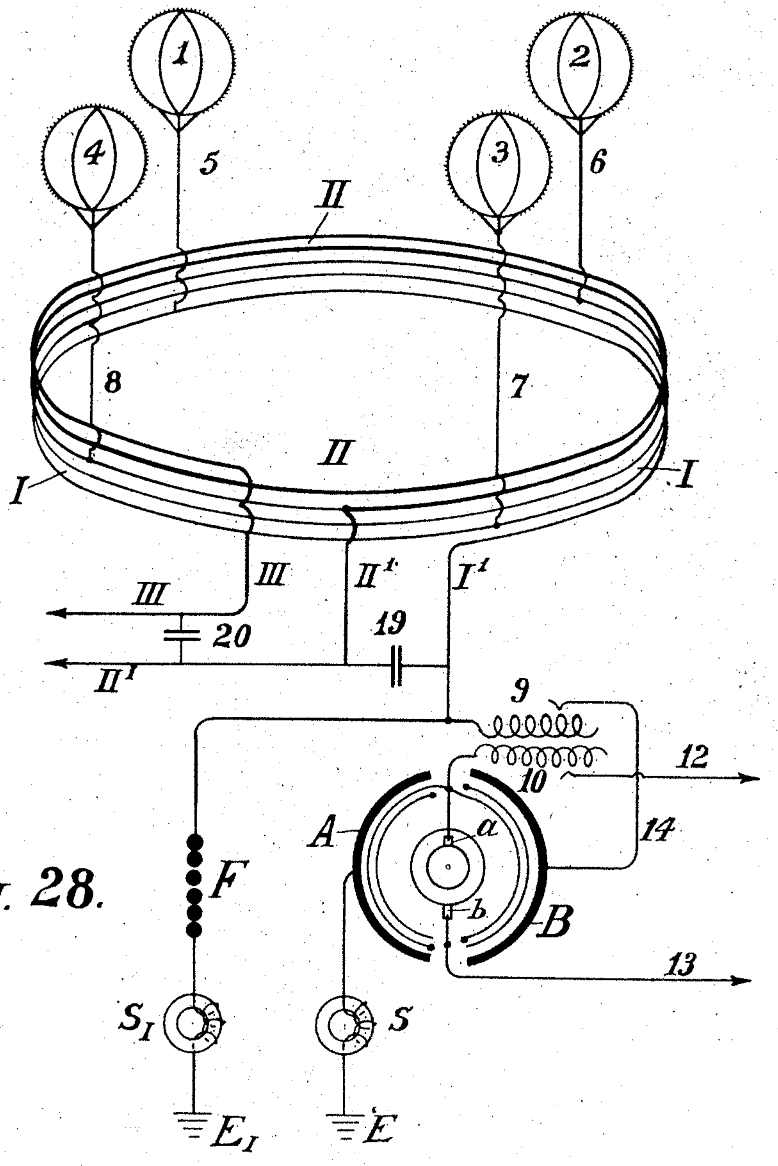
1
2
4
5
3
6
II
8
7
II
I
I
III
II1'
I1
III
20
19
II1
9
12
10
A
a
14
B
b
13
F
S1
S
E1
E

Fig. 28.

Inventor
Hermann Plauson
by
Knight Bros
attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921. 12 Sheets-Sheet 10

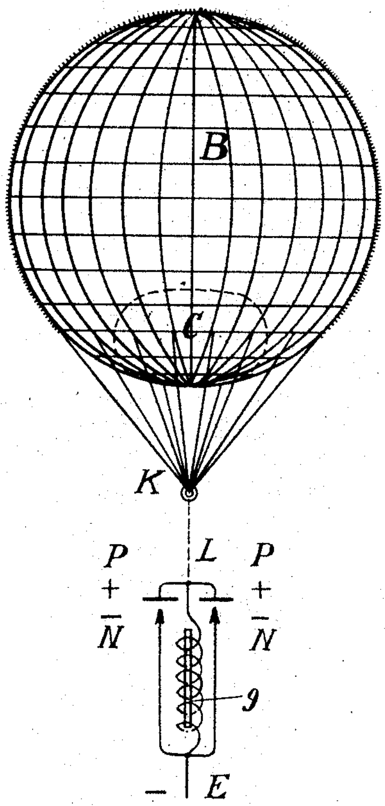
B
C
K
L
P
+
−
N
P
+
−
N
9
−
E

Fig. 29.

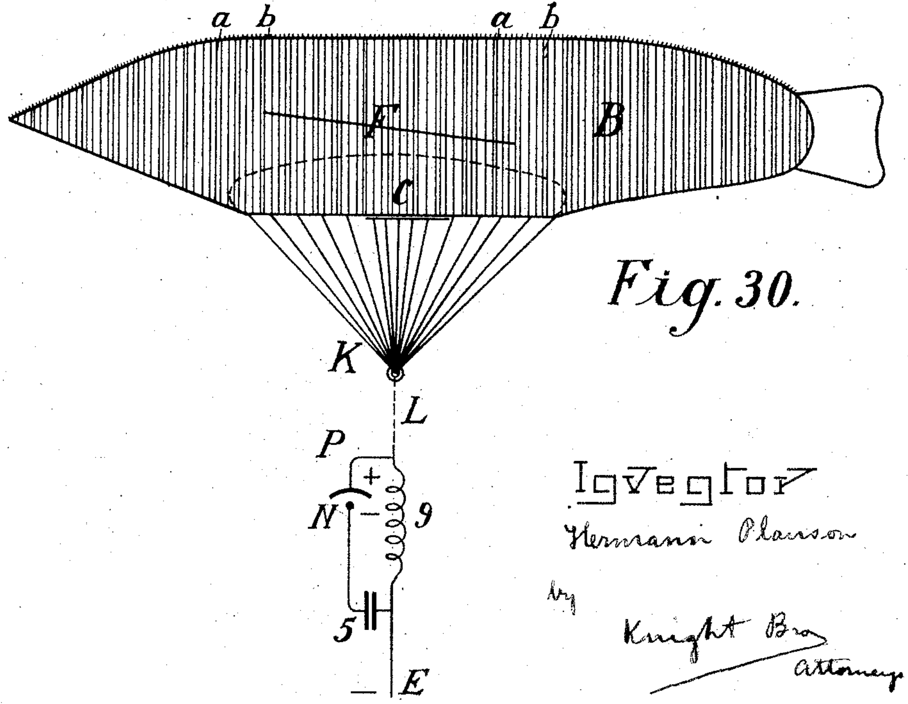
a
b
a
b
F
B
C
K
L
P
+
N
−
9
5
−
E

Fig. 30.

Inventor
Hermann Plauson
by
Knight Bros
Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 11

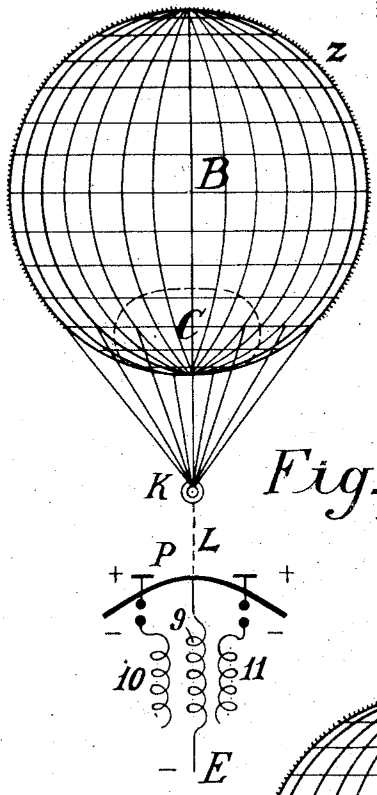
z
B
C
K
P
L
+
+
−
−
9
10
11
−
E

Fig. 31.

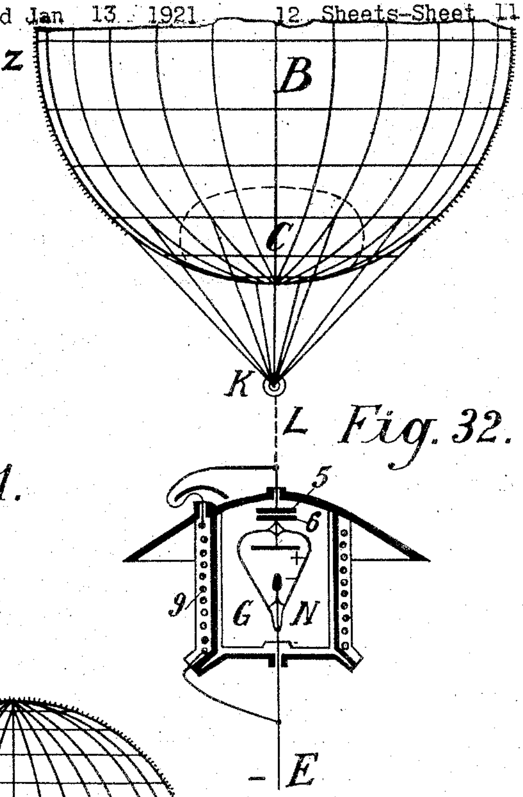
z
B
C
K
L
5
6
+
−
9
G
N
−
E

Fig. 32.

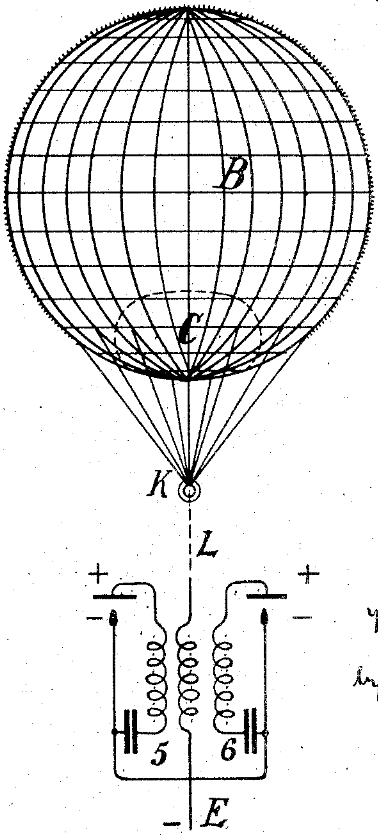

Fig. 33.

Inventor
Hermann Plauson
by Knight Bros
attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 12

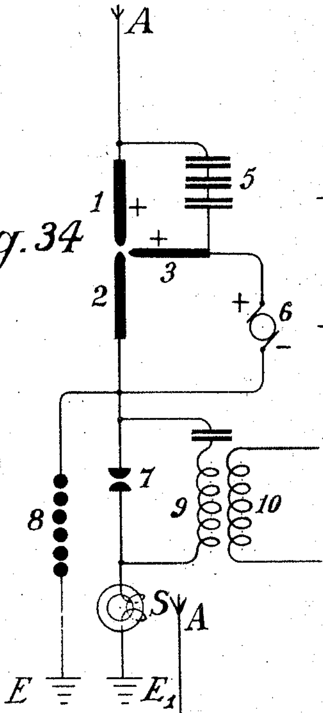
A
1
+
5
+
3
2
+
6
–
8
7
9
10
S
E
E1

Fig. 34

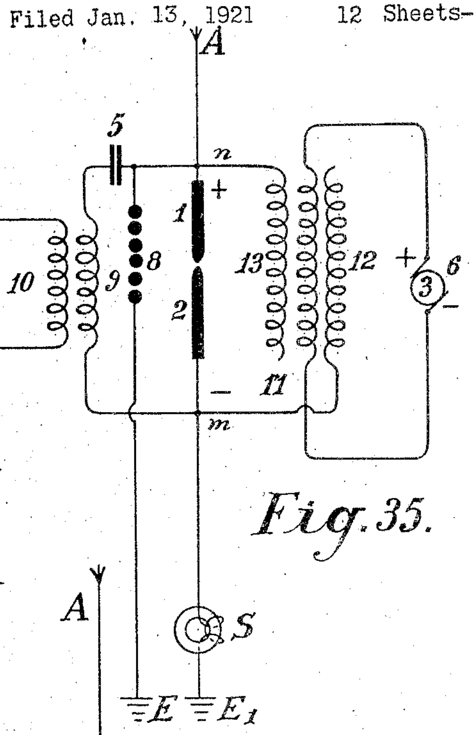
A
5
n
+
1
8
10
9
13
12
+
6
3
–
2
–
11
m
S
E
E1

Fig. 35.

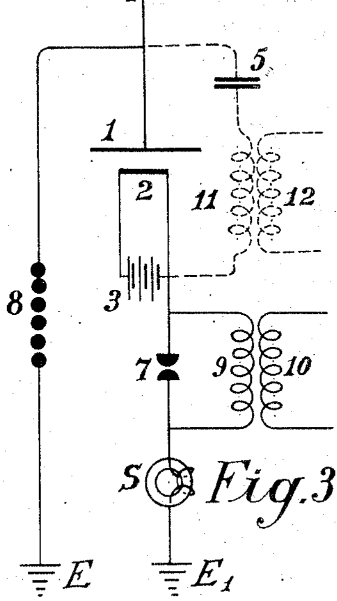
5
1
2
11
12
8
3
7
9
10
S
E
E1

Fig. 36.

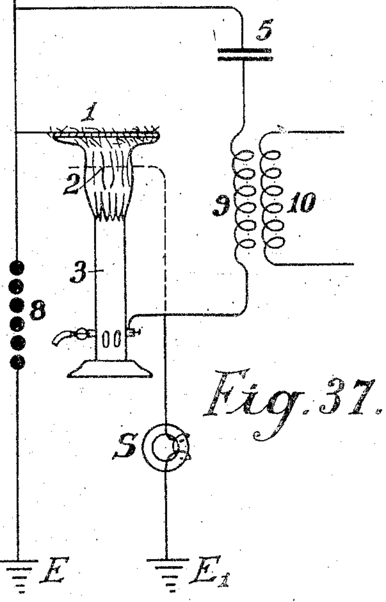
5
1
2
9
10
3
8
S
E
E1

Fig. 37.

Inventor

Hermann Plauson

by Knight Bros

Attorneys

Patented June 9, 1925. **1,540,998**

UNITED STATES PATENT OFFICE.

HERMANN PLAUSON, OF HAMBURG, GERMANY.

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY.

Application filed January 13, 1921. Serial No. 437,107.

To all whom it may concern:

Be it known that I, HERMANN PLAUSON, Esthonian subject, residing in Hamburg, Germany, have invented certain new and useful Improvements in the Conversion of Atmospheric Electric Energy, of which the following is a specification.

Methods of obtaining atmospheric electricity by means of metallic nettings set with spikes which are held by means of ordinary or anchored kite balloons made of fabrics and filled with hydrogen, are in theory already known. Atmospheric electricity obtained in this way has been suggested to be used in the form of direct current for the charging of accumulators. This knowledge however is at present only theoretical as the conversion in practice has hitherto been a failure. No means are known of protecting the apparatus from destruction by lightning. The balloons used for collecting the charge must also be made of very large size in order to be able to support the weight of the metallic netting and the heavy cable connections.

Instead of using heavy metallic netting as collectors attached to single air balloons of non-conducting materials which are liable to be torn and are permeable to the gas, it is proposed to use metallic balloon collectors which have the following important advantages—

(*a*) The metallic cases are impenetrable to helium and hydrogen; they also represent large metallic weather-proof collecting surfaces.

(*b*) Radio active means and the like may be easily applied internally or externally; whereby the ionization is considerably increased and therewith also the quantity of atmospheric electricity capable of being collected.

(*c*) Such balloon collectors of light metal do not require to be of large size as they have to carry only their own moderate weight, and that of the conducting cable or wire.

(*d*) The entire system therefore offers little surface for the action of storm and wind and is resistant and stable.

(*e*) Each balloon can be easily raised and lowered by means of a winch so that all repairs, recharging and the like can be carried out without danger during the operation.

It is further proposed to use a collecting aerial network of several separate collectors spread out in the air above the earth, which collectors are interconnected by electrical conductors.

According to this invention charges of atmospheric electricity are not directly converted into mechanical energy, and this forms the main difference from previous inventions, but the static electricity which runs to earth through aerial conductors in the form of direct current of very high voltage and low current strength is converted into electro-dynamic energy in the form of high frequency vibrations. Many advantages are thereby obtained and all disadvantages avoided.

The very high voltage of static electricity of a low current strength can be converted by this invention to voltages more suitable for technical purposes and of greater current strength. By the use of closed oscillatory circuits it is possible to obtain electromagnetic waves of various amplitude and thereby to increase the degree of resonance of such current. Such resonance allows various values of inductance to be chosen whereby again the governing of the starting and stopping of machines driven thereby by simply tuning the resonance between coils of the machine and the transformer circuit forming the resonance can easily be obtained. Further, such currents have the property of being directly available for various uses, even without employing them for driving motors, of which there may be particularly mentioned, lighting, production of heat and use in electro-chemistry.

Further, with such currents a series of apparatus may be fed without direct current supply through conductors and also the electro-magnetic high frequency currents may be converted by means of special motors adapted for electro-magnetic oscillations into mechanical energy, or finally converted by special machines into alternating current of low frequency or even into direct current of high potential.

The invention is more particularly described with reference to the accompanying diagrams in which:—

Figure 1 is an explanatory figure.

Figure 2 is a diagrammatic view of the simplest form.

Figure 3 shows a method of converting atmospheric electrical energy for use with motors.

Figure 4 is a diagram showing the use of protective means.

Figure 5 is a diagram of an arrangement for converting large current strengths.

Figure 6 is a diagram of an arrangement including controlling means.

Figure 7 shows means whereby the spark gap length can be adjusted.

Figure 8 shows a unipolar connection for the motor.

Figure 9 shows a weak coupled system suitable for use with small power motors.

Figures 10, 11 and 12 show modified arrangements.

Figure 13 shows a form of inductive coupling for the motor circuit.

Figure 14 is a modified form of Figure 13 with inductive coupling.

Figure 15 is an arrangement with non-inductive motor.

Figure 16 is an arrangement with coupling by condenser.

Figures 17, 18 and 19 are diagrams of further modifications.

Figure 20 shows a simple form in which the aerial network is combined with special collectors.

Figure 21 shows diagrammatically an arrangement suitable for collecting large quantities of energy.

Figure 22 is a modified arrangement having two rings of collectors.

Figure 23 shows the connections for three rings of collectors.

Figure 24 shows a collecting balloon and diagram of its connection of condenser batteries.

Figures 25 and 26 show modified collector balloon arrangements.

Figure 27 shows a second method of connecting conductor for the balloon aerials.

Figure 28 shows an auto-transformer method of connection.

Figure 29 shows the simplest form of construction with incandescent cathode.

Figure 30 shows a form with cigar shaped balloon.

Figure 31 is a modified arrangement.

Figure 32 shows a form with cathode and electrode enclosed in a vacuum chamber.

Figure 33 is a modified form of Figure 32.

Figure 34 shows an arc light collector.

Figure 35 shows such an arrangement for alternating current.

Figure 36 shows an incandescent collector with Nernst lamp.

Figure 37 shows a form with a gas flame.

Figure 1 illustrates a simple diagram for converting static electricity into dynamic energy of a high number of oscillations. For the sake of clearness in the drawings an influence machine is assumed to be employed and not an aerial antenna. 13 and 14 are combs for collecting the static electricity of the influence machine. 7 and 8 are spark discharging electrodes, 6 and 5 condensers, 9 an inductive primary coil, 10 secondary coil, 11 and 12 ends of conductors of the secondary coil 10. When the disc of the static influence machine is rotated by mechanical means, the combs collect the electric charges one the positive and the other the negative, and charge the condensers 5 and 6 until such a high potential is formed across the spark gap 7—8, that the spark gap is jumped. As the spark gap 7—8 forms a closed circuit with condensers 6 and 5, and inductive resistance 9, as is well known, waves of high frequency electromagnetic oscillations will pass in this circuit.

The high frequency of the oscillations produced in the primary circuit induces waves of the same periodicity in the secondary circuit. Thus in the primary circuit electromagnetic oscillations are formed by the passage of the spark over the spark gap and these waves are maintained by fresh charges of static electricity.

By suitably selecting the ratio between the number of the coils in the primary and secondary circuits with regard to a correct application of the co-efficients of resonance (capacity, inductance, and resistance) the high voltage of the primary circuit may be suitably converted into low voltage and high current strength.

When the oscillatory discharges in the primary circuit becomes weaker or entirely cease, the condensers are charged again by the static electricity until the accumulated charge again breaks down the spark gap. All this is repeated as long as electricity is produced by the static machine by employing mechanical energy.

An elementary form of the invention is shown in Figure 2 in which two spark gaps in parallel are used one of which may be termed the working gap 7 in Figure 2, whilst the second serves as a safety device for excess voltage and consists of a larger number of spark gaps than the working section, which gaps are arranged in series and are bridged by very small capacities as is illustrated in a_1, b_1, c_1, Figure 2 which allow of uniform sparking in the safety section.

In Figure 2 A is the aerial antenna for collecting charges of atmospheric electricity, 13 is the earth connection of the second part of the spark gap, 5 and 6 are condensers, 9 a primary coil. Now when through the aerial A the positive atmospheric electricity seeks to combine with the negative charge to earth, this is prevented by (the air gap between) the spark gaps. The resistance of the spark gap 7 is, as shown in the drawings, lower than that of the other safety section which consists of three spark gaps connected in

series, and consequently a three times greater air resistance is offered by the latter.

So long therefore, as the resistance of the spark gap 7 is not overloaded, so that the other spark gaps have an equal resistance with it the discharges take place only over spark gap 7. Should however the voltage be increased by any influences so that it might be dangerous for charging the condensers 5 and 6 or for the coil insulation 9 and 10 in consequence of break down, by a correct regulation of this spark gap the second spark gap can discharge free from inductive effects direct to earth without endangering the machine.

Without this second spark gap, arranged in parallel having a higher resistance than the working spark gap it is impossible to collect and render available large quantities of electrical energy.

The action of this closed oscillation circuit consisting of spark gap 7, two condensers 5 and 6, primary coil 9, and also secondary coil 10 is exactly the same as the one described in Figure 1 with the arrangement of the static induction machine with the only difference that here the second spark gap is provided. The electromagnetic high frequency alternating current obtained can be tapped off from the conductors 11 and 12 for lighting and heating purposes. Special kinds of motors adapted for working with these peculiar electrical charges may be connected at 14 and 15 which can work with static electricity charges or with high frequency oscillations.

In addition to the use of spark gaps in parallel a second measure of security is also necessary for taking off the current. This precaution consists according to this invention, in the introduction of and method of connecting certain protective electromagnets or choking coils in the aerial circuit as shown by S in Figure 3.

A single electromagnet only having a core of the thinnest possible separate laminations is connected with the aerial.

In the case of high voltages in the aerial network or at places where there are frequent thunder storms, several such magnets may however be connected in series.

In the case of large units or plants several electromagnets can be employed in parallel or in series parallel.

The windings of these electromagnets may be simply connected in series with the aerials. In this case the winding preferably consists of several thin parallel wires, which make up together, the necessary section.

The winding may be made of primary and secondary windings in the form of a transformer. The primary winding will be then connected in series with the aerial network, and the secondary winding more or less short-circuited over a regulating resistance or an induction coil. In the latter case it is possible to regulate to a certain extent the effect of the choking coils. In the further description of the connecting and constructional diagrams the aerial electromagnet choke coil is indicated by a simple ring S.

Figure 3 shows the simplest way of converting atmospheric electricity into electromagnetic wave energy by the use of special motors adapted for high oscillatory currents or static charges of electrical energy. Recent improvements in motors for working with static charges and motors working by resonance, that is to say, having groups of tuned electromagnetic cooperating circuits render this possible but such do not form part of the present invention.

A motor adapted to operate with static charges will for the sake of simplicity be diagrammatically indicated by two semicircles 1 and 2 and the rotor of the motor by a ring M. (Figure 3.) A is a vertical aerial or aerial network. S the safety choke or electromagnet with coil O as may be seen is connected with the aerial A. Adjacent the electromagnet S the aerial conductor is divided into three circuits, the circuit 8 giving the safety spark gap, the circuit 7 with the working spark gap, and then a circuit including the stator terminal 1, the rotor and stator terminal 2 at which a connection is made to the earth wire. The two spark gaps are also connected metallically with the earth wire. The method of working these diagrams is as follows:

The positive atmospheric electric charge collected tends to combine with the negative electricity (or earth electricity) connected with the earth wire. It travels along the aerial A through the electromagnet S without being checked as it flows in the same direction as the direct current. Further, its progress is arrested by two sparks gaps placed in the way and the stator condenser surfaces. The stator condenser surfaces are charged until the charge is greater than the resistance of the spark gap 7, whereupon a spark springs over the spark gap 7 and an oscillatory charge is obtained as by means of the motor M, stator surfaces 1 and 2, and spark gap 7, a closed oscillation circuit is obtained for producing the electromagnetic oscillations. The motor here forms the capacity and the necessary inductance and resistance, which, as is well known, are necessary for converting static electricity into electromagnetic wave energy.

The discharges formed are converted into mechanical energy in special motors and can not reach the aerial network by reason of the electromagnet or choke. If, however, when a spark springs over the spark gap 7 a greater quantity of atmospheric electricity tends to flow to earth, a counter voltage is

induced in the electromagnet, which is greater the more rapidly and strongly the flow of current direct to the earth is. By the formation of this opposing voltage a sufficiently high resistance is offered to the flow of atmospheric electricity direct to earth to prevent a short circuit with the earth.

The circuit containing spark gap 8 having a different wave length which is not in resonance with the natural frequency of the motor, does not endanger the motor and serves as security against excess voltage, which, as practical experiments have shown, may still arise in certain cases, but can be conducted direct to earth through this spark gap.

In the diagram illustrated in Figure 4 the spark gap 7 is shunted across condensers 5 and 6 from the motor M. This construction affords mainly a better insulation of the motor against excess voltage and a uniform excitation through the spark gap 7.

In Figure 5 a diagram is illustrated for transforming large current strengths which may be employed direct without motors, for example, for lighting or heating purposes. The main difference is that here the spark gap consists of a star shaped disc 7 which can rotate on its own axis and is rotated by a motor opposite similarly fitted electrodes 7^a. When separate points of stars face one another, discharges take place, thus forming an oscillation circuit over condensers 5 and 6 and inductance 9 for oscillatory discharges. It is evident that a motor may also be directly connected to the ends of the spiral 9.

The construction of the diagram shown in Figure 6 permits of the oscillation circuit of the motor being connected with an induction coil. Here a regulating inductive resistance is introduced for counter-acting excess voltages in the motor. By cutting the separate coils 9 (coupled inductively to the aerial) in or out the inductive action on the motor may be more or less increased or variable aerial action may be exerted on the oscillation circuit.

In Figure 7 the oscillation circuit is closed through the earth (E and E_1). The spark gap 7 may be prolonged or shortened by more or fewer spark gaps being successively connected by means of a contact arm 7^b.

Diagram 8 shows a unipolar connection of the motor with the aerial network. Here two oscillation circuits are closed through the same motor. The first oscillation circuit passes from aerial A through electromagnet S, point x, inductance 9^a to the earth condenser 6 and further, over spark gap 7 to the aerial condenser 5 and back to x. The second oscillation circuit starts from the aerial condenser 5 at the point x^1 over the inductance 9 to the earth condenser 6 at the point x^3 and through the condenser 6 over the spark gap 7 back to x^1. The motor itself is inserted between the two points of the spark gap 7. From this arrangement slightly damped oscillation wave currents are produced.

In the diagram illustrated in Figure 9 a loosely coupled system of connections is illustrated which is assumed to be for small motors for measuring purposes. A indicates the aerial conductor, S the electromagnet in the aerial conductor, 9 the inductance, 7 the spark gap, 5 and 6 condensers, E the earth, M the motor, and 1 and 2 stator connections of the motor. The motor is directly metallically connected with the oscillation circuit.

In Figure 10 a purely inductive coupling is employed for the motor circuit. The motor is connected with the secondary wire 10 as may be seen in Figure 11 in a somewhat modified diagram connection. The same applies to the diagram of Figure 12.

The diagrams hitherto described preferably allow of motors of small and medium strength to be operated. For large aggregates, however, they are too inconvenient as the construction of two or more oscillation circuits for large amounts of energy is difficult; the governing is still more difficult and the danger in switching on or off is greater.

A means of overcoming such difficulties is shown in Figure 13. The oscillation circuit here runs starting from the point x over condenser 5, variable inductance 9, spark gap 7 and the two segments (3^a and 4^a) forming arms of a Wheatstone bridge, back to x. If the motor is connected by brushes 3 and 4 transversely to the two arms of the bridge as shown in the drawings, electromagnetic oscillations of equal sign are induced in the stator surfaces 1 and 2 and the motor does not revolve. If however, the brushes 3 and 4 are moved in common with the conducting wires 1 and 2 which connect the brushes with the stator poles a certain alteration or displacement of the polarity is obtained and the motor commences to revolve.

The maximum action will result if one brush 3 comes on the central sparking contact 7 and the other brush 4 on the part x. They are however, usually in practice not brought on to the central contact 7 but only held in the path of the bridge segments 4^a and 3^a in order not to connect the spark gaps with the motor oscillation circuit.

As however, the entire oscillation energy can thereby not act on the motor it is better to carry out the same system according to the diagram 14. The diagram 14 differs from the foregoing only by the motor not being directly metallically connected with the segments of the commutator, but only a primary coil 9 which induces in a secondary coil 10, current which feeds the motor M and takes the place of the rotor. By this

arrangement a good transforming action is obtained, a loose coupling and also an oscillation circuit without a spark gap.

In Figure 15 the motor is not purely inductively as in 14, but directly metallically branched off from the primary coil (at x and x^1) after the principle of the auto-transformer.

In Figure 16 instead of an inductance a condenser 6 is in similar manner, and for the same object inserted between the segments 3^a and 4^a. This has the advantage that the segments 3^a and 4^a need not be made of solid metal but may consist of spiral coils whereby a more exact regulation is possible and further motors of high inductance may be employed.

The arrangements of Figures 17, 18 and 19 may be employed for use with resonance and particularly with induction condenser motors; between the large stator induction condenser surfaces, small reversing pole condensers are connected, which, as may be seen from Figures 17, 18 and 19 are led together to earth. Such reversing poles have the advantage that with large quantities of electrical energy the spark formation between the separate oscillation circuits ceases.

Figure 19 shows a further method which prevents electromagnetic oscillations of high number of alternations formed in the oscillation circuit striking back to the aerial conductor. It is based on the well known principle that a mercury lamp, one electrode of which is formed of mercury, the other of solid metal such as steel allows an electric charge to pass in only one direction from the mercury to the steel and not vice versa. The mercury electrode of the vacuum tube N is therefore connected with the aerial conductor and the steel electrode with the oscillation circuit. From this it results that charges can pass only from the aerial through the vacuum tube to the oscillation circuit, but not vice versa. Oscillations which are formed on being transformed in the oscillation circuit cannot pass to the aerial conductor.

In practice these vacuum tubes must be connected behind an electromagnet as the latter alone affords no protection against the danger of lightning.

As regards the use of spark gaps, all arrangements as used for wireless telegraphy may be used. Of course the spark gaps in large machines must have a sufficiently large surface. In very large stations they are cooled in liquid carbonic acid or better still in liquid nitrogen or hydrogen; in most cases the cooling may also take place by means of liquefied low homologues of the metal series or by means of hydrocarbons the freezing point of which lies at between −90° C. and −40° C. The spark gap casing must also be insulated and be of sufficient strength to be able to resist any pressure which may arise. Any undesirable excess super-pressure which may be formed must be automatically let off. I have employed with very good results mercury electrodes which were frozen in liquid carbonic acid, the cooling being maintained during the operation from the outside through the walls.

Figure 20 is one of the simplest forms of construction of an aerial network in combination with collectors, transformers and the like illustrated diagrammatically. E is here the earth wire, 8 the safety spark gap, 7 the working spark gap, 1 and 2 the stator surfaces of the motor, 5 a condenser battery, S the protective magnet which is connected with the coil in the aerial conductor, A^1 to A^{10} aerial antennæ with collecting balloons, N horizontal collecting or connecting wires from which, to the centre a number of connections run.

The actual collectors consist of metal sheaths preferably made of an aluminium magnesium alloy, and are filled with hydrogen or helium and are attached to copper plated steel wires. The size of the balloon is selected so that the actual weight of the balloon and the weight of the conducting wire is supported thereby. On the top of the balloon aluminium spikes, made and gilded in a special manner hereinafter described, are arranged in order to produce a conductor action. Small quantities of radium preparations, more particularly polonium-ionium or mesothorium preparations considerably increase the ionization, and therewith the action of these collectors.

In addition to metal balloons, fabric balloons which are superficially metal coated according to Schoop's metal spraying process, may however also be employed. A metallic surface may also be produced by lacquering with metallic bronzes, preferably according to Schoop's spraying process or lacquering with metallic bronze powders in two electrical series of widely different metals, because thereby the collecting effect is considerably increased.

Instead of the ordinary round balloons, elongated cigar shaped ones may be employed. In order also to utilize the frictional energy of the wind, patches or strips of non-conducting substances which produce electricity by friction, may be attached to the metallized balloon surfaces. The wind will impart a portion of its energy in the form of frictional electricity, to the balloon casing, and thereby the collecting effect is substantially increased.

In practice however, very high towers (up to 300 metres is fully admissible) may be employed as antennæ. In these towers copper tubes rise freely further above the top of the tower. A gas lamp secured

against the wind is then lit at the point of the copper tube and a netting is secured to the copper tube over the flame of this lamp to form a collector. The gas is conveyed through the interior of the tube up to the summit. The copper tube must be absolutely protected from moisture at the place at which it enters the tower and also rain must be prevented running down the walls of the tower which might lead to a bad catastrophe. This is done by bell shaped enlargements which expand downwards, being arranged in the tower in the form of high voltage insulators of Siamese pagodas.

Special attention must be devoted to the foundations of such towers. They must be well insulated from the ground, which may be obtained by first embedding a layer of concrete in a box form to a sufficient depth in the ground and inserting in this an asphalt lining and then glass bricks cast about 1 or 2 metres in thickness. Over this in turn there is a ferro-concrete layer in which alone the metal foot of the tube is secured. This concrete block must be at least 2 metres from the ground and be fully protected at the sides by a wooden covering, from moisture. In the lower part of the tower a wood or glass house for the large condenser batteries or for the motors may be constructed. In order to lead the earth connection to the ground water, a well insulated pit constructed of vitreous bricks, must be provided. Several such towers are erected at equal distances apart and connected with a horizontal conductor. The horizontal connecting wires may either run directly from tower to tower or be carried on bell shaped insulators similar to those in use for high voltage conductors. The width of the network may be of any suitable size and the connection of the motors can take place at any suitable places.

In order to collect large quantities of electricity with few aerials it is well to provide the aerial conductor with batteries of condensers as shown in two methods of construction in Figures 21 and 22. In Figure 21 the batteries of condensers 5 are connected on the one hand with the aerial electricity collectors Z by the aerial conductor A, and on the other hand interconnected in series with an annular conductor from which horizontal conductors run to the connecting points C to which the earth wire is connected.

Figure 22 shows a similar arrangement. Should two such series of antennæ rings be shown by a voltmeter to have a large difference of potential (for example, one in the mountains and one in the plain) or even of different polarity these differences may be compensated for by connecting sufficiently large condenser batteries (5, 5^a, 5^b) by means of Maji star conductors D and D^1. In Figure 23 a connection of three such rings of collectors to form a triangle with a central condenser battery is illustrated.

The condenser batteries of such large installations must be embedded in liquefied gases or in liquids freezing at very low temperatures. In such cases a portion of the atmospheric energy must be employed for liquefying these gases. It is also preferable to employ pressure. By this means the condenser surfaces may be diminished, and still allow for large quantities of energy to be stored, secure against breakdown. For smaller installations the immersing of the condensers in well insulated oil or the like, suffices. Solid substances on the other hand cannot be employed as insulators.

The arrangement in the diagrams hitherto described was always such that the condenser batteries were connected with both poles directly to the aerial conductors. An improved diagram of the connections for obtaining atmospheric electricity for the condenser batteries has however, been found to be very advantageous, this arrangement consists in that they are connected by only one pole (unipolar) to the collecting network. Such a method of arrangement is very important, as by means of it a constant current and an increase of the normal working pressure or voltage is obtained. If for example a collecting balloon aerial which is allowed to rise to a height of 300 metres, shows 40,000 volts above earth voltage, in practice it has been found that the working voltage (with a withdrawal of the power according to the method hereinbefore described by means of oscillating spark gaps and the like) is only about 400 volts. If however, the capacity of the condenser surfaces be increased, which capacity in the above mentioned case was equal to that of the collecting surface of the balloon aerials, to double the amount, by connecting the condenser batteries with only one pole, the voltage rises under an equal withdrawal of current up to and beyond 500 volts. This can only be ascribed to the favourable action of the connecting method.

In addition to this substantial improvement it has also been found preferable to insert double inductances with electromagnets and to place the capacities preferably between two such electromagnets. It has also been found that the useful action of such condensers can be further increased if an induction coil be connected as inductive resistance to the unconnected pole of the condenser, or still better if the condenser itself be made as an induction condenser. Such a condenser may be compared with a spring which when compressed carries in itself accumulated force, which it again gives off when released. In charging, a charge with reversed sign is formed at the

other free condenser pole, and if through the spark gap a short circuit results, the accumulated energy is again given back since now new quantities of energy are induced at the condenser pole connected with the conductor network, which in fact charges with opposite signs to that at the free condenser pole. The new induced charges have of course the same sign as the collector network. The whole voltage energy in the aerial is thereby however increased. In the same space of time larger quantities of energy are accumulated than is the case without such inserted condenser batteries.

In Figures 24 and 25 two different diagrams of connections are more exactly illustrated, Figure 24 shows a collecting balloon and the diagram of the connections to earth. Figure 25 four collecting balloons and the parallel connection of the condenser batteries belonging thereto.

A is the collecting balloon made of an aluminium magnesium alloy (electron metal, magnalium) of a specific gravity of 1.8 and a thickness of plate 0.1 to 0.2 mm. Inside there are eight strong vertical ribs of T shaped section about 10 to 20 mm. in height and about 3 mm. in thickness with the projecting part directed inwards (indicated by *a*, *b*, *c*, *d* and so forth); they are riveted together to form a firm skeleton and are stiffened in a horizontal direction by two cross ribs. The ribs are further connected with one another internally and transversely by means of thin steel wires, whereby the balloon obtains great power of resistance and elasticity. Rolled plates of 0.1 to 0.2 mm. in thickness made of magnalium alloy are then either soldered or riveted on this skeleton so that a fully metallic casing with smooth external surface is obtained. Well silvered or coppered aluminium plated steel wires run from each rib to the fastening ring 2. Further, the coppered steel hawser L preferably twisted out of separate thin wires (shown in dotted lines in Figure 24) and which must be long enough to allow the balloon to rise in the desired height, leads to a metal roller or pulley 3 and from thence to a winch W, well insulated from the earth. By means of this winch, the balloon, which is filled with hydrogen, or helium, can be allowed to rise to a suitable height (300 to 5,000 metres) and brought to the ground for recharging or repairs.

The actual current is taken directly through a friction contact from the metal roller 3 or from the wire, or even from the winch or simultaneously from all three by means of brushes (3, 3^a and 3^b). Beyond the brushes the conductor is divided, the paths being:—firstly over 12 to the safety spark gap 8, from thence to the earth conductor E^1, and secondly over electromagnet S^1, point 13, to a second loose electromagnet having an adjustable coil S^2, then to the spark gap 7 and to the second earth conductor E^2. The actual working circuit is formed through the spark gap 7, condensers 5 and 6, and through the primary coil 9; here the static electricity formed by oscillatory discharges is accumulated and converted into high frequency electromagnetic oscillations. Between the electromagnets S^1 and S^2 at the crossing point 13, four condenser batteries are introduced which are only indicated diagrammatically in the drawings each by one condenser. Two of these batteries (16 and 18) are made as plate condensers and prolonged by regulating induction coils or spirals 17 and 19 while the two others (21 and 23) are induction condensers. As may be seen from the drawings each of the four condenser batteries 16, 18, 21, 23 is connected only by one pole to the aerial or to the collector conductor. The second poles 17, 19, 22, 24 are open. In the case of plate condensers having no inductive resistance an induction coil is inserted. The object of such a spiral or coil is the displacement of phase of the induction current by ¼ periods, whilst the charging current of the condenser poles which lie free in the air, works back to the collector aerial. The consequence of this is that in discharges in the collector aerial the back inductive action of the free poles allows a higher voltage to be maintained in the aerial collecting conductor than would otherwise be the case. It has also been found that such a back action has an extremely favourable effect on the wear of the contacts. Of course the inductive effect may be regulated at will within the limits of the size of the induction coil, the length of the coil in action being adjustable by means of wire connection without induction (see Fig. 24, No. 20).

S^1 and S^2 may also be provided with such regulating devices in the case of S^2 (illustrated by 11). If excess voltage be formed it is conducted to earth through the wire 12 and spark gap 8 or through any other suitable apparatus, since this formation would be dangerous for the other apparatus.

The action of these condenser batteries has already been hereinbefore described.

The small circles on the collector balloon indicate places at which zinc amalgam or gold amalgam or other photoelectric acting metals in the form of small patches in extremely thin layers (.01 to .05 mm. in thickness) are applied to the balloon casing of light metal. Such metallic patches may also be applied to the entire balloon as well as in greater thickness to the conducting network. The capacity of the collector is thereby considerably strengthened at the surface. The greatest possible effect in collecting may be obtained by polonium amalgams and the like. On the surface of the

collector balloon metal points or spikes are also fixed along the ribs, which spikes serve particularly for collecting the collector charge. Since it is well known that the resistance of the spikes is less the sharper the spike is, for this purpose it is therefore extremely important to employ as sharp spikes as possible. Experiments made as regards these have shown that the formation of the body of the spike or point also plays a large part. for example, spikes made of bars or rollers with smooth surfaces, have a many times greater point resistance as collector accumulator spikes than those with rough surfaces. Various kinds of spike bodies have been experimented with for the collector balloons hereinbefore mentioned. The best results were given by spikes which were made in the following way. Fine points made of steel, copper, nickel, or copper and nickel alloys, were fastened together in bundles and then placed as anode with the points in a suitable electrolyte (preferably in hydrochloric acid or muriate of iron solutions) and so treated with weak current at 2 to 3 volts pressure. After 2 to 3 hours according to the thickness of the spikes or pins the points become extremely sharp and the bodies of the spikes have a rough surface. The bundle can then be removed and the acid washed off with water. The spikes are then placed as cathode in a bath consisting of solution of gold, platinum, iridium, paladium or wolfram salts or their compounds and coated at the cathode galvanically with a thin layer of precious metal, which must however be sufficiently firm to protect them from atmospheric oxidation.

Such spikes act at a 20 fold lower voltage almost as well as the best and finest points made by mechanical means. Still better results are obtained if polonium or radium salts are added to the galvanic bath when forming the protective layer or coating. Such pins have a low resistance at their points and even at one volt and still lower pressures have an excellent collector action.

In Figure 24 the three unconnected poles are not connected with one another in parallel. That is quite possible in practice without altering the principle of the free pole. It is also preferable to interconnect in parallel to a common collector network, a series of collecting aerials.

Figure 25 shows a diagram for such an installation. A^1, A^2, A^3, A^4 are four metal collector balloons with gold or platinum coated spikes which are electrolytically made in the presence of polonium emanations or radium salts, which spikes or needles are connected over four electro-magnets S^1, S^2, S^3, S^4, through an annular conductor R. From this annular conductor four wires run over four further electromagnets S^a, S^b, S^c, S^d, to the connecting point 13. There the conductor is divided, one branch passing over 12 and the safety spark gap 8 to the earth at E^1, the other over inductive resistance J and working spark gap 7 to the earth at E^2. The working circuit, consisting of the condenser 5 and 6 and a resonance motor or a condenser motor M, such as hereinbefore described, is connected in proximity round the sparking gap section 7.

Instead of directly connecting the condenser motor of course the primary circuit for high frequency oscillatory current may also be inserted.

The condenser batteries are connected by one pole to the annular conductor R and can be either inductionless (16 and 18) or made as induction condensers as shown by 21 and 23. The free poles of the inductionless condensers are indicated by 17 and 19, those of the induction condensers by 22 and 24. As may be seen from the drawings all these poles 17, 22, 19, 24 may be interconnected in parallel through a second annular conductor without any fear that thereby the principle of the free pole connection will be injured. In addition to the advantages already set forth the parallel connection also allows of an equalization of the working pressure in the entire collector network. Suitably constructed and calculated induction coils 25 and 26 may also be inserted in the annular conductor of the free poles, by means of which a circuit may be formed in the secondary coils 27 and 28 which allows current produced in this annular conductor by fluctuations of the charges or the like appearances to be measured or otherwise utilized.

According to what has been hereinbefore stated separate collector balloons may be connected at equidistant stations distributed over the entire country, either connected directly with one another metallically or by means of intermediate suitably connected condenser batteries through high voltage conductors insulated from earth. The static electricity is converted through a spark gap into dynamic energy of a high number of oscillations and may in such form be coupled as a source of energy by means of a suitable method of connecting, various precautions being observed and with special regulations. The wires leading from the collector balloons have hitherto been connected through an annular conductor without this endless connection, which can be regarded as an endless induction coil, being able to exert any action on the whole conductor system.

It has now been found that if the network conductor connecting the aerial collector balloons with one another is not made as a simple annular conductor, but preferably short circuited in the form of coils over a

condenser battery or spark gap or through thermionic tubes or valves or audions, then the total collecting network exhibits quite new properties. The collection of atmospheric electricity is thereby not only increased but an alternating field may be easily produced in the collector network. Further, the atmospheric electrical forces showing themselves in the higher regions may also be directly obtained by induction. In Figures 26 and 28 a form of construction is shown on the basis of which the further foundations of the method will be more particularly explained.

In Figure 26 1, 2, 3, 4 are metal collector balloons, 5, 6, 7, 8 their metallic aerial conductors and I the actual collector network. This consists of five coils and is mounted on high voltage insulators in the air, on high voltage masts (or with a suitable construction of cable embedded in the earth). One coil has a diameter of 1 to 100 km. or more. S and S^1 are two protective electromagnets, F the second safety section against excess voltage, E its earth conductor and E^1 the earth conductor of the working section. When an absorption of static atmospheric electricity is effected through the four balloon collectors, the current in order to reach the earth connection E^1 must flow spirally through the collector network over the electromagnet S, primary induction coil 9, conductor 14, anode A of the audion tube, incandescent cathode K, as the way over the electromagnet and safety spark gap F offers considerably greater resistance. Owing to the fact that the accumulated current flows in one direction, an electromagnetic alternating field is produced in the interior of the collector network coil, whereby the whole free electrons are directed more or less into the interior of the coil. An increased ionization of the atmosphere is therefore produced. In consequence of this the points mounted on the collector balloon show a considerably reduced resistance and therefore increased static charges between the points on the balloon and the surrounding atmosphere are produced. The result of this is a considerably increased collector effect.

A second effect which could not be obtained otherwise is obtained by the electromagnetic alternating field which running parallel to the earth surface, acts more or less with a diminishing or increasing effect on the earth magnetic field, whereby in the case of fluctuations in the current a return induction current of reversed sign is always produced in the collector coil by earth magnetism. Now if, however, a constantly pulsating continuous alternating field is produced as stated in the above collector network I, an alternating current of the same periodicity is produced also in the collecting network coil. As the same alternating field is further transmitted to the aerial balloon, the resistance of its points is thereby considerably reduced, whilst the collector action is considerably increased. A further advantage is that positive electrons which collect on the metal surfaces during the conversion into dynamic current produce a so-called drop of potential of the collector area. As an alternating field is present, the negative ions surrounding the collector surfaces, when discharge of the collector surfaces takes place produce by the law of induction, an induction of reversed sign on the collector surface and so forth (that is to say again a positive charge). In addition to the advantages hereinbefore set forth, the construction of connecting conductors in coil form when of sufficiently large diameter, allows of a utilization of energy arising in higher regions also in the simplest way. As is well known electric discharges frequently take place at very great elevations which may be observed, such as St. Elmo's fires or northern lights. These energy quantities have not been able to be utilized up to now. By this invention all these kinds of energy, as they are of an electromagnetic nature and the direction of the axis of the collector coils stands at right angles to the earth's surface, can be more or less absorbed in the same way as a receiver in wireless telegraphy absorbs waves coming from a far distance. With a large diameter of the spiral it is possible to connect large surfaces and thereby to take up also large quantities of energy.

It is well known that large wireless stations in the summer months, and also in the tropics are very frequently unable to receive the signals in consequence of interruptions which are caused by atmospheric electricity, and this takes place with vertical coils of only 40 to 100 m. diameter. If on the contrary horizontal coils of 1 to 100 km. diameter be employed very strong currents may be obtained through discharges which are constantly taking place in the atmosphere. Particularly in the tropics or still better in the polar regions where the northern lights are constantly present, large quantities of energy may probably be obtained in this way. A coil with several windings should act the best. In similar manner any alteration of the earth magnetism should act inductively on such a coil.

It is not at all unlikely that earthquakes and spots on the sun will also produce an induction in such collector coils of sufficient size. In similar manner this collector conductor will react on earth currents more particularly when they are near the surface of the earth or even embedded in the earth. By combining the previous kind of current collectors so far as they are adapted for

the improved system with the improved possibilities of obtaining current the quantities of free natural energy which are to be obtained in the form of electricity are considerably increased.

In order to produce in the improved collector coil uniform current oscillations of an undamped nature so-called audion high vacuum or thermionic tubes of suitable connection are employed instead of the previously known spark gaps (Fig. 26, Nos. 9–18). The main aerial current flows through electromagnet S (which in the case of a high number of alternations is not connected here but in the earth conductor E^1) and may be conveyed over the primary coils in the induction winding through wire 14 to the anode A of the high vacuum grid tube. Parallel with the induction resistance 9 a regulating capacity of suitable size, such as condenser 11 is inserted. In the lower part of the vacuum grid tube is arranged the incandescent filament or the cathode K which is fed through a battery B. From the battery B two branches run, one to the earth conductor E^1 and the other through battery B^1 and secondary coil 10 to the grid anode *g* in the vacuum tube. By the method of connections shown in dotted lines, a desired voltage at the grid electrode *g* may also be produced through the wire 17 which is branched off from the main current conductor through switches 16 and some small condensers (*a*, *b*, *c*, *d*) connected in series, and conductor 18, without the battery B^1 being required.

The action of the entire system is somewhat as follows:—

On the connecting conductor of the aerial collector network being short circuited to earth, the condenser pole 11 is charged and slightly damped oscillations are formed in the short circuited existing oscillation circuit formed of the condenser 11 and self inductance 9. In consequence of the coupling through coil 10, fluctuations of voltage take place in the grid circuit 15 with the same frequency, which, fluctuations in turn influence the strength of the electrode current passing through the high vacuum amplyfying tube and thus produce current fluctuations of the same frequency in the anode circuit. A permanent supply of energy to the oscillation circuits 9 and 10 consequently takes place, until a condition of balance is set up, in which the consumed oscillation energy is equal to that absorbed. Thereby constant undamped oscillations are now produced in the oscillation circuits 9–11.

For regular working of such oscillation producers high vacuum strengthening tubes are necessary and it is also necessary that the grid and anode voltages shall have a phase difference of 180° so that if the grid is negatively charged, then the anode is positively charged and vice versa. This necessary difference of phase may be obtained by most varied connections, for example, by placing the oscillation circuit in the grid circuit or by separating the oscillation circuit and inductive coupling from the anodes and the grid circuit and so forth.

A second important factor in this way of converting static atmospheric electricity into undamped oscillations is that care must be taken that the grid and anode voltages have a certain relation to one another; the latter may be obtained by altering the coupling and a suitable selection of the self induction in the grid circuit, or as shown by dotted lines 18, 17, 16 by means of a larger or smaller number of condensers of suitable size connected in series; in this case the battery B^1 may be omitted. With a suitable selection of the grid potential a glow discharge takes place between the grid *g* and the anode A, and accordingly at the grid there is a cathode drop and a dark space is formed. The size of this cathode drop is influenced by the ions which are emitted in the lower space in consequence of shock ionization of the incandescent cathodes K and pass through the grid in the upper space. On the other hand the number of the ions passing through the grid is dependent on the voltage between the grid and the cathode. Thus if the grid voltage undergos periodic fluctuations (as in the present case) the amount of the cathode drop at the grid fluctuates and consequently the internal resistance of the tube correspondingly fluctuates, so that when a back coupling of the feed circuit with the grid circuit takes place, the necessary means are afforded for producing undamped oscillations and of taking current, according to requirements from the collecting conductor.

The frequency of the undamped oscillations produced is with a suitably loose coupling equal to the self frequency of the oscillation circuits 9 and 10. By a suitable selection of the self induction of the coil 9 and capacity 11 it is possible to extend from frequencies which produce electromagnetic oscillations of only a few metres wave length down to the lowest practical alternating current frequency. For large installations a suitable number of frequency producing tubes in the form of the well known high vacuum transmission tubes of .5 to 2 kw. in size may be connected in parallel so that in this respect no difficulty exists.

The use of such tubes for producing undamped oscillations, and also the construction and method of inserting such transmission tubes in an accumulator or dynamo circuit is known and also that such oscillation producing tubes only work well at voltages of 1,000 up to 4,000 volts, so that on the contrary their use at lower voltages is

considerably more difficult. By the use of high voltage static electricity this method of producing undamped oscillations as compared with that through spark gaps must be regarded, as an ideal solution particularly for small installations of outputs of from 1 to 100 kw.

By the application of safety spark gaps, with interpolation of electro-magnets, not only is short circuiting avoided but also the taking up of current is regulated. Oscillation producers inserted in the above way form a constantly acting electromagnetic alternating field in the collector coil, whereby as already stated, a considerable accumulating effect takes place. The withdrawal wire or working wire is connected at 12 and 13, but current may be taken by means of a secondary coil which is firmly or movably mounted in any suitable way inside the large collector coil, i. e. in its electromagnetic alternating field, so long as the direction of its axis runs parallel with that of the main current collecting coil.

In producing undamped oscillations of a high frequency (50,000 per second and more) in the oscillation circuits 9 and 11, electromagnets S and S^1 must be inserted if the high frequency oscillations are not to penetrate the collector coil, between the oscillation producers and the collector coil. In all other cases they are connected shortly before the earthing (as in Figs. 27 and 28).

In Figure 27 a second method of construction of the connecting conductor of the balloon aerials is illustrated in the form of a coil. The main difference consists in that in addition to the connecting conductor I another annular conductor II is inserted parallel to the former on the high voltage masts in the air (or embedded as a cable in the earth) but both in the form of a coil. The connecting wire of the balloon aerials is indicated as a primary conductor and also as a current producing network; the other is the consumption network and is not in unipolar connection with the current producing network.

In Figure 27 the current producing network I is shown with three balloon collectors 1, 2, 3 and aerial conductors 4, 5, 6; it is short circuited through condenser 19 and inductance 9. The oscillation forming circuit consists in this diagram of spark gap *f*, inductance 10, and condenser 11; the earth wire E, is connected to earth over electromagnet S^1. F is the safety spark gap which is also connected to earth through a second electromagnet S at E. On connecting up the condenser circuit 11 this is charged over the spark gap *f* whereby an oscillatory discharge is formed. This discharging current acts through inductance 10 on the inductively coupled secondary 9, whereby in the producing network a modification of the potential of the condenser 19 is produced. The consequence of this is that oscillations arise in the coil shaped producer network. These oscillations induce a current in the secondary circuit II, which has a smaller number of windings and a less resistance, the voltage of which, according to the proportion of the number of windings and of the ohmic resistance, is considerably lower whilst the current strength is greater.

In order to convert the current thus obtained into current of an undamped character, and to tune its wave lengths, a sufficiently large regulatable capacity 20 is inserted between the ends 12 and 13 of the secondary conductor II. Here also current may be taken without an earth conductor, but it is advisable to insert a safety spark gap E^1 and to connect this with the earth over an electromagnet S^2.

The producer network may be connected with the working network II over an inductionless condenser 21 or over an induction condenser 22, 23. In this case the secondary conductor is unipolarly connected with the energy conductor.

In Figure 28 the connecting conductor between the separate accumulator balloons is carried out according to the autotransformer principle. The collecting coil connects four aerial balloons 1, 2, 3, 4, the windings of which are not made side by side but one above the other. In Figure 28 the collector coil I is shown with a thin line, the metallically connected prolongation coils II with a thick line. Between the ends I^1 and II^1 of the energy network I a regulating capacity 19 is inserted. The wire I^1 is connected with the output wire and with the spark gap F.

As transformer of the atmospheric electricity an arrangement is employed which consists in using rotary pairs of condensers in which the one stator surface B is connected with the main current, whilst the other A is connected with the earth pole. Between these pairs of short circuited condensers are caused to rotate from which the converted current can be taken by means of two collector rings and brushes, in the form of an alternating current, the frequency of which is dependent on the number of balloons and the revolutions of the rotor. As the alternating current formed in the rotor can act, in this improved method of connection described in this invention, through coils 10 on the inductance 9, an increase or diminution of the feed current in I can be obtained according to the direction of the current by back induction. Current oscillations of uniform rhythm thereby result in the coil shaped windings of the producer net-work.

As the ends of this conductor are short cir-

cuited through the regulatable condenser 19 these rhythms produce short circuited undamped oscillations in the energy conductor, the periodicity and wave lengths of which oscillations can be adjusted according to desire by altering the capacity 19 to a given wave length and therewith also to a given frequency. These currents may also be employed in this form directly as working current through the conductors II^1 and III. By inserting the condenser 20 a connection between these conductors may also be made, whereby harmonic oscillations of desired wave length are formed. By this means quite new effects as regards current distribution are obtained. The withdrawal of current can even take place without direct wire connection if, at a suitable point in the interior of the producing network (quite immaterially whether this has a diameter of 1 or 100 km.) a coil tuned to these wave lengths and of the desired capacity is firmly or movably mounted in the aerial conductor in such a way that its axial direction is in parallel with that of the collector coil. In this case a current is induced in the producing network, the size of which is dependent on the total capacity and resistance and also on the periodicity employed. A possibility is thereby afforded in future, of taking energy from the producer network by wireless means. As thereby in addition to atmospheric electricity also magnetic earth currents and the energy from the higher atmosphere (at least partially) may be simultaneously obtained, this last system for collecting the atmospheric energy is of particular importance for the future.

Of course everywhere instead of spark gaps suitable grid vacuum tubes may be employed as producers for undamped oscillations. The separate coils of the producer net-work with large diameters may be connected with one another through separate conductors all in parallel or all in series or in groups in series. By regulating the number of oscillations and also the extent of the voltage more or less large collector coils of this kind may be employed. The coils may also be divided spirally over the entire section. The coils may be carried out in annular form or also in triangular, quadrangular, hexagonal or octagonal form.

Of course wires may be carried from a suitable place to the centre or also laterally which serve the current waves as guides. This is necessary when the currents have to be conducted over mountains and valleys and so forth. In all these cases the current must be converted into a current of suitable periodicity.

As already hereinbefore mentioned separate collecting balloons may be directly metallically interconnected at equidistant stations distributed over the entire country or may be connected by interpolation of suitable condenser batteries by means of high voltage conductors. The static electricity is converted through a spark gap into dynamic energy of a high number of oscillations, and could then in such form, with a suitable arrangement of the connections, observing various measures of precaution, be employed as source of energy after separate or special regulation.

According to this invention in order to increase the collecting effect of the balloon in the aerial collector conductor or in the earth wire, radiating collectors are employed. These consist either of incandescent metal or oxide electrodes in the form of vacuum grid tubes, or electric arcs (mercury and the like electrodes) Nernst lamps, or finally flames of various kinds may be simply connected with the respective conductor.

It is well known that energy can be drawn off from a cathode consisting of an incandescent body opposite an anode charged with positive electricity (vacuum grid tube). Hitherto however, a cathode was always first directly placed opposite an anode, and secondly the system always consisted of a closed circuit.

Now if we dispense with the ordinary ideas in forming light or flame arcs in which a cathode must always stand directly opposite an anode, and if we place an incandescent cathode opposite an anode charged to a high potential or another body freely floating in the air, or regard the incandescent cathode only as a source of unipolar discharge (which represent group and point discharges in electro-static machines similar to unipolar discharges), it may be ascertained that incandescent cathodes and less perfectly all incandescent radiators, flames and the like admit of relatively large current densities and allow large quantities of electric energy to radiate into the open space in the form of electron streams as transmitters.

The object of this invention is as described below, if such incandescent oxide electrodes or other incandescent radiators or flames are not freely suspended in space but connected metallically with the earth so that they can be charged with negative terrestrial electricity, these radiators possess the property of absorbing the free positive electrical charges contained in the air space surrounding them (that is to say of collecting them and conducting them to earth). They can therefore, serve as collectors and have, in comparison to the action of the spikes, or points, a very large radius of action R; the effective capacity of these collector is much greater than the geometrical capacity (R_0-) calculated in an electro-static sense.

Now as our earth is surrounded as is well

known, with an electro-static field and the difference of potential

$$\frac{\delta V}{\delta h}$$

of the earth field according to the latest investigations, is in summer about 60 to 100 volts and in winter 300 to 500 volts per metre of difference in height (δh), a simple calculation gives the result that when such a radiation collector or flame collector is arranged for example on the ground, and a second one is mounted vertically over it at a distance of 2,000 metres and both are connected by a conducting cable, there is a difference of potential in summer of about 2,000,000 volts and in winter even of 6,000,000 volts and more.

According to Stefan Boltzmann's law of radiation, the quantity of energy which an incandescent surface (temperature T) of 1 sq. cm. radiates in a unit of time into the open air (temperature T_o) is expressed by the following formula:

$$S=\sigma\,(T^4-T^4{}_o)\ \text{watt/sq. cm.}$$

and the universal radiation constant σ is according to the latest researches of Ferry (Annales de Chimie et de physique 17 page 267 (1909)) equal to 6.30×10^{-12} watt/sq. cm.

Now if an incandescent surface of 1 sq. cm. shows, as compared with the surrounding space a periodic fall of potential δV it radiates (independent of the current direction, that is to say of the sign) in accordance with the above formula, for example at a temperature of 3725° C. an energy of 1.6 kw. per sq. cm. per second. As for the radiation the same value can be calculated for the collection of energy, but reversed. Now as carbon electrodes at the temperature of the electric arc support on the current basis a current density up to from 60 to 65 amperes per sq. cm. no difficulties will result in this direction in employing radiating collectors as accumulators.

If the earth be regarded as a cosmically insulated condenser in the sense of geometrical electro-statics *x* there results from the geometric (compare Ewald Rasch, "das elektrische Bogenlicht" (The electric arc light) page 169) capacity of the earth according to Chwolson:

For negative charging 1.3×10^6 Coulomb

For negative potential $V=10\times10^8$ volts.

From this there results however, $EJT\cong24.7\times10^{24}$ watt/Sec. Now if it is desired to make a theoretic short circuit through an earthed flame collector this would represent an electric total work of about 79,500 10^{10} kilowatt years. As the earth must be regarded as a rotating mechanism which is thermo-dynamically, electromagnetically, and also kinematically coupled with the sun and stars system by cosmic radiations and gravitation a diminution of the electric energy of the earth field is not to be feared. The energies which the incandescent collectors would withdraw from the earth field can only cause by the withdrawal of motor work a lowering of the earth temperature (temperature $T_E=300$) and reduce this to that of the world space ($T=O$) by using the entire energy. This is however not the case as the earth does not represent a cosmically entirely insulated system. On the contrary there is conveyed to the same according to the recent value corrected by Ferry for the solar constants through the radiation from the sun an energy of $18{,}500\times10^{10}$ kw. Accordingly any lowering of the earth temperature (T_E) without a simultaneous lowering of the sun's temperature (T_S) would contradict Stefan Boltzmann's law of radiation.

$$S=\sigma\,(T_S{}^4-T_E{}^4).$$

From this it must be concluded that if the earth temperature (T_E) sinks, the total radiation S absorbed by the earth increases, and further also that the secular speed of cooling of the earth is directly dependent on that of the sun and the other radiators cosmically coupled with the sun and is connected most closely with these.

The incandescent radiation collectors may, according to this invention, be employed for collecting atmospheric electricity if they (1) are charged with the negative earth electricity (that is to say when they are directly connected by means of a metallic conductor with the earth) and (2) if large capacities (metal surfaces) charged with electricity are mounted opposite them as positive poles in the air. This is regarded as the main feature of the present invention as without these inventive ideas it would not be possible to collect with an incandescent collector, sufficiently large quantities of the electrical charges contained in the atmosphere as technology requires; the radius of action of the flame collectors would also be too small, especially if it be considered that the very small surface density (energy density) (σ about$=2\times7\ .\ 10^9$ St. E. per sq. cm.) does not allow of large quantities of charge being absorbed from the atmosphere.

x) Calculated according to Poisson's calculation;

$\Delta V=-4\pi\delta$; as here the alteration of the potential or potential gradients only takes place in the direction of the normal, this calculation assumes the simple form

$$\delta=\frac{1}{4\pi}\times\frac{\delta^2V}{\delta n^2}$$

It has indeed already been proposed to employ flame collectors for collecting atmospheric electricity and it is known that

their collecting effect is substantially greater opposite the points. It is however, not known that the quantities of current which could hitherto be obtained are too small for technical purposes. According to my experiments the reason for this is to be found in the too small capacities of the collector conductor poles. If such flame or radiating collectors have no or only small positive surfaces, their radius of action for large technical purposes is too small. If the incandescent collectors be constantly kept in movement in the air they may collect more according to the speed of the movement, but this is again not capable of being carried out in practice.

By this invention the collector effect is considerably increased by a body charged with a positive potential and of the best possible capacity being also held floating (without direct earth connection) opposite such an incandescent collector which is held floating in the air at a desired height. If for example, a collecting balloon of sheet metal or of metalized balloon fabric be caused to mount to 300 up to 3,000 metres in the air and as positive pole it is brought opposite such a radiating collector connected by a conductor to earth, quite different results are obtained.

The metallic balloon shell (with a large surface) is charged to a high potential by the atmospheric electricity. This potential is greater the higher the collecting balloon is above the incandescent collector. The positive electricity acts concentratedly on the anode floating in the air as it is attracted through the radiation shock ionization, proceeding from the incandescent cathode. The consequence of this is that the radius of action of the incandescent cathode collector is considerably increased and thereby also the collecting effect of the collecting balloon surface. Further the large capacity of the anode floating in the air plays therefore an important part because it allows of the taking of large charges, and thereby a more uniform current is obtained even when there is a large consumption; this cannot be the case with small surfaces.

In the present case the metallic collecting balloon is a positive anode floating in the air and the end of the earth conductor of this balloon serves as positive pole surface opposite the surface of the radiating incandescent cathode, which in turn is charged with negative earth electricity being conductingly connected to earth.

The process may be carried out by two such contacts (negative incandescent cathode and anode end of a capacity floating in the air) a condenser and an inductive resistance being switched on in parallel, whereby simultaneously undamped oscillations may be formed.

In very large installations it is advisable to connect two such radiating collectors in series. Thus an arc light incandescent cathode may be placed below on the open ground and an incandescent cathode which is heated by special electro-magnetic currents be located high in the air. Of course for this the special vacuum Liebig tubes with or without grids may also be employed. An ordinary arc lamp with oxide electrodes may be introduced on the ground and the positive pole is not directly connected with the collecting balloon, but through the upper incandescent cathode or over a condenser. The method of connecting the incandescent cathode floating in the air may be seen in Figs. 29–33.

B is the air balloon, K a Cardan ring (connection with the hawser) C the balloon, L a good conducting cable, P a positive pole, N negative incandescent cathode, and E earth conductor.

Fig. 29 represents the simplest form of construction. If electric oscillations are produced below on the ground by means of a carbon arc lamp or in other suitable way a considerably greater electric resistance is opposed to that in the direct way by inserting an electrical inductive resistance 9. Consequently between P and N a voltage is formed, and as, over N and P only an inductionless ohmic resistance is present, a spark will spring over so long as the separate induction co-efficients and the like are correctly calculated. The consequence of this is that the oxide electrode (carbon or the like) is rendered incandescent and then shows as incandescent cathode an increased collecting effect. The positive poles must be substantially larger than the negative in order that they may not also become incandescent. As they are further connected with the large balloon area which has a large capacity and is charged at high voltage, an incandescent body which is held floating in the air and a positive pole which can collect large capacities is thereby obtained in the simplest way. The incandescent cathode is first caused to become incandescent by means of separate energy produced on the earth, and then maintained by the energy collected from the atmosphere.

Fig. 30 only shows the difference that instead of a round balloon a cigar shaped one (of metal or metalized fabric) may be employed and also a condenser 5 is inserted between the incandescent cathode and the earth conductor so that a short circuited oscillation circuit over P. N 5 and 9 is obtained. This has the advantage that quite small quantities of electricity cause the cathode to become incandescent and much larger cathode bodies may be rendered incandescent.

In this form of construction both the in-

candescent cathode and also the positive electrode may be enclosed in a vacuum chamber as may be seen in Fig. 32. A cable L is carried well insulated through the cover of a vessel and ends in a condenser disc 5. The cover is arched in order to keep off the rain. The vessel is entirely or partially made of magnetic metal and well insulated inside and outside. Opposite the disc 5 another disc 6 and on this again a metallic positive pole of the vacuum tube *g* with the incandescent cathode (oxide electrode) N is arranged. The negative electrode is on the one hand connected with the earth conductor E, and on the other hand with the inductive resistance 9 which is also connected with the cable L with the positive pole and wound round the vessel in coils. The action is exactly the same as that in Fig. 29 only instead of an open incandescent cathode one enclosed in vacuo is employed. As in such collectors only small bodies can be brought to incandescence in large installations a plurality of such vacuum tubes must be inserted in proximity to one another. According to the previous constructions Figs. 31 and 33 are quite self evident without further explanations.

Figs. 34–37 represent further diagrams of connections over radiating and flame collectors, and in fact, how they are to be arranged on the ground.

Fig. 34 shows an arc light collector with oxide electrodes for direct current and its connection; Fig. 35 a similar one for alternating current, Fig. 36 an incandescent collector with a Nernst lamp and Fig. 37 a similar one with a gas flame.

The positive pole 1 of the radiating collectors is always directly connected to the aerial collecting conductor A. In Fig. 34 this is further connected over the condenser battery 5 with a second positive electrode 3. The direct current dynamo *b* produces current which flows over between the electrodes 3 and 2 as an arc light. On the formation of an arc the negative incandescent electrode 2 absorbs electricity from the positive poles standing opposite it and highly charged with atmospheric electricity and conveys the same to the working circuit. The spark gap 7, inductive resistance 9 and induction coil 10 are like the ones previously described. The protective electromagnet S guards the installation against earth circuiting, the safety spark gap 8 from excess voltage or overcharging.

In Fig. 35 the connection is so far altered that the alternating current dynamo feeds the exciting coil 11 of the induction condenser. 12 is its negative and 13 its positive pole; if the coil 3 on the magnet core of the dynamo is correctly calculated and the periodicity of the alternating current is sufficiently high an arc light can be formed between the two poles 1 and 2. As the cathode 2 is connected with the negatively charged earth, and therefore always acts as a negative pole, a form of rectification of the alternating current produced by the dynamo 3 is obtained, the second half of the period is always suppressed. The working circuit may be carried out in the same way as in Fig. 34; the working spark gap 7 may however be dispensed with, and instead thereof between the points *n* and *m* a condenser 5 and an induction resistance 9 may be inserted from which the current is taken inductively.

Fig. 36 represents a form of construction similar to Fig. 34 only that here instead of an arc lamp a Nernst incandescent body is employed. The Nernst lamp is fed through the battery 3. The working section is connected with the negative pole, the safety spark gap with the + poles. The working spark gap 7 may also be dispensed with and the current for it taken at 12 over the oscillation circuit 5, 11 (shown in dotted lines).

Flame collectors (Fig. 37) may also be employed according to this invention. The wire network 1 is connected with the aerial collector conductor A and the burner with the earth. At the upper end of the latter, long points are provided which project into the flame. The positive electrode is connected with the negative over a condenser 5 and the induction coil 9 with the earth.

The novelty in this invention is firstly, the use of incandescent cathodes opposite positive poles which are connected with large metallic capacities as automatic collecting surfaces, (2) the connection of the incandescent cathodes with the earth whereby, in addition to the electricity conveyed to them from the battery or machine which causes the incandescing, also the negative charge of the earth potential is conveyed, and (3) the connection of the positive and negative poles of the radiating collectors over a condenser circuit alone or with the introduction of a suitable inductive resistance, whereby simultaneously an oscillatory oscillation circuit may be obtained. The collecting effect is by these methods quite considerably increased.

I declare that what I claim is:—

1. An electrical energy generating system, comprising a conducting surface for static charges, means to support same at a distance above the earth, a conductor leading to the earth level, a spark gap associated with said conductor to convert electrostatic charges into electromagnetic high frequency oscillations means to supply said electromagnetic energy to a net work, and a spark gap of greatly increased relative resistance in parallel therewith.

2. An electrical energy generating system

comprising a conductor, means to support same above the earth level, an inductance therein, a spark gap associated with said conductor, a second spark gap of much higher relative resistance in parallel therewith and an energy receiving circuit coupled with the spark gap of lesser resistance.

3. An electrical energy generating system comprising a collecting surface, means to support same above the earth level, a conductor connecting said collecting surface with the earth level, a choke in said conductor, an electromagnetic resistance converting electrostatic energy to electromagnetic energy, a safety higher resistance in parallel therewith and a net work coupled with the conversion resistance of lesser value.

4. An electrical energy generating system comprising electric conductors spaced above the earth to form electromagnetic oscillating circuits, conductors connecting to earth level, electrostatic to electromagnetic energy conversion means therein, a safety high electrostatic resistance in parallel therewith and means to alter the electromagnetic characteristic of the circuits.

5. An electrical energy generating system comprising in combination a static collecting surface arranged above the earth, conductors connecting to earth level, a pair of spark gaps in parallel of different electrostatic resistance, a utilization net work shunted across the spark gap of lesser resistance and an electromagnetic choke in said conductors.

6. An electrical energy generating system comprising an open circuit energy collecting aerial, a pair of sparking gaps in parallel of widely different resistance, connected thereto and a closed electric oscillation circuit in shunt across the gap of lesser resistance.

7. An electrical energy generating system comprising an open circuit energy collecting aerial, a pair of sparking gaps in parallel of widely different resistance connected thereto, a closed electric oscillation circuit in shunt across the gap of lesser resistance, a plurality of electrostatic collecting surfaces, means to connect said collecting surfaces in parallel in groups and means to connect said groups symmetrically with said aerial.

In witness whereof, I have hereunto signed my name this 30 day of Dec., 1920, in the presence of two subscribing witnesses.

HERMANN PLAUSON.

Witnesses:
H. F. Armstrong,
W. H. Beeston.

Nov. 12, 1957

L. R. CRUMP

2,813,242

POWERING ELECTRICAL DEVICES WITH ENERGY
ABSTRACTED FROM THE ATMOSPHERE
Filed March 12, 1954

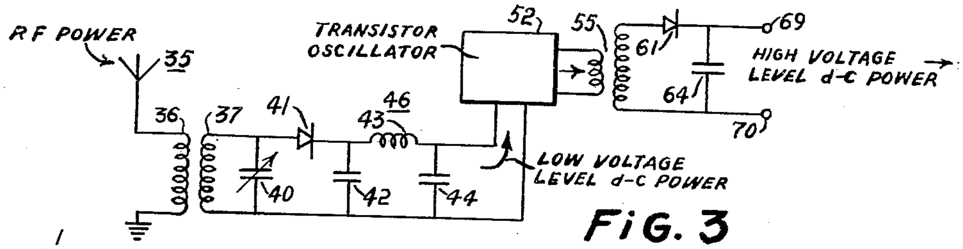
RF POWER
35
36
37
41
46
43
40
42
44
TRANSISTOR OSCILLATOR
52
55
61
64
69
70
HIGH VOLTAGE LEVEL d-c POWER
LOW VOLTAGE LEVEL d-c POWER

FiG. 3

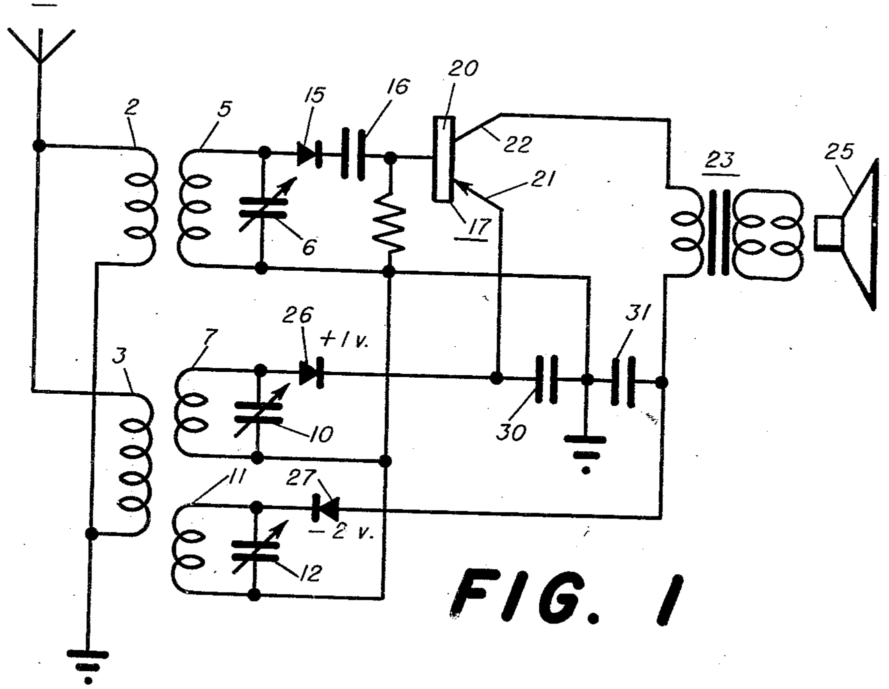
1
2
5
15
16
20
22
21
17
23
25
6
26
+1 v.
7
3
31
10
30
11
27
-2 v.
12

FIG. 1

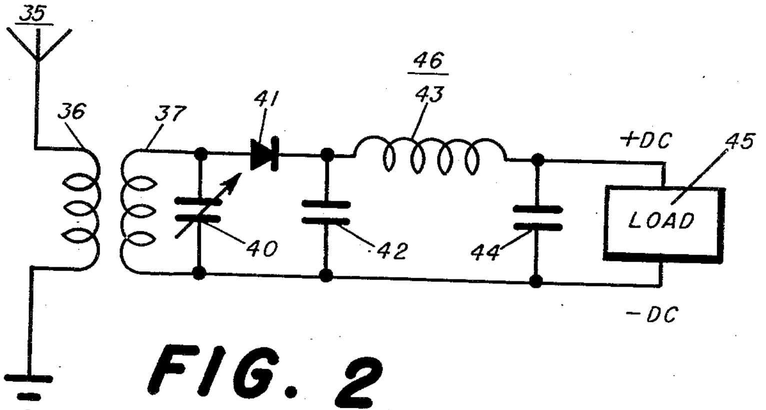
35
36
37
41
46
43
+DC
45
LOAD
40
42
44
-DC

FIG. 2

INVENTOR
Lloyd R. Crump

BY W.E. Thibodeau + A.W. Dew

ATTORNEYS

United States Patent Office

2,813,242

Patented Nov. 12, 1957

2,813,242

POWERING ELECTRICAL DEVICES WITH ENERGY ABSTRACTED FROM THE ATMOSPHERE

Lloyd R. Crump, Silver Spring, Md.

Application March 12, 1954, Serial No. 415,986

1 Claim. (Cl. 321—2)

(Granted under Title 35, U. S. Code (1952), sec. 266)

The invention described herein may be manufactured and used by or for the Government for governmental purposes without payment to me of any royalty thereon.

This invention relates to the convenient and economical provision of power for the operation of electronic circuits and devices using transistors, and of other electrical devices having modest power requirements.

A great advantage of transistors, and a major reason for their enthusiastic reception since their introduction a few years ago, is the fact that they will operate satisfactorily with very low supply voltages and currents. One milliwatt or even less is sufficient to power a transistor in many applications. Various batteries have been developed to provide, in a minimum of space, the relatively minute amounts of power needed by transistors.

My invention provides methods and means that permit transistor circuits, and also other low-powered electrical devices, to be economically and conveniently operated without any batteries whatever, and indeed without any power supply whatever as power supplies are ordinarily conceived.

The invention centers around my discovery that it is practicable to construct operative transistor circuits that are able to abstract from the atmosphere sufficient electromagnetic energy to provide all necessary supply voltages and currents for their own operation. Circuits and devices powered according to my invention will operate indefinitely without any local power source whatever.

I have successfully constructed and demonstrated such circuits. For example, I have constructed a batteryless transistor radio receiver on which I have listened to either nearby or distant broadcast stations as desired, using either headphones or a loudspeaker; this receiver has been powered entirely by electromagnetic energy abstracted from the atmosphere.

From the successful operation of this receiver, and from other experimental work, it becomes clear that, by the methods and means of the invention, a great variety of practical and useful transistor circuits can be powered entirely by energy abstracted from the atmosphere.

Furthermore, as will become apparent below, my invention is applicable to the powering of other electrical devices requiring relatively small amounts of power.

An object of the present invention is to provide methods and means for powering transistor circuits entirely from radiofrequency energy abstracted from the atmosphere.

Another object is to provide methods and means for powering remote radio receivers, low-powered radio transmitters, and other low-powered electrical devices, with energy received by radio from a master station, so that no local power supplies are needed by the devices and so that the powering or non-powering of the remote device is under the control of the master station.

A further object is to provide methods and means for powering transistor circuits and other low-powered electrical devices with radiofrequency energy received from one or more remote radio transmitters.

Other objects, aspects, uses, and advantages of the invention will become apparent from the following description and from the drawing.

Figure 1 is a schematic diagram of a transistor radio receiver in which all necessary power is supplied by energy abstracted from the atmosphere in accordance with the invention.

Figure 2 is a schematic diagram showing a general application of the invention to provide direct-current power to a load.

Figure 3 is a schematic diagram of a system for obtaining a high energy D.-C. source at a high voltage level using energy abstracted from the atmosphere.

Referring to Figure 1, a receiving antenna **1** is connected to antenna coupling coils **2** and **3**, the other ends of which are connected to ground. A parallel resonant circuit consisting of coil **5** and capacitor **6** is coupled to coil **2**. A second parallel resonant circuit consisting of coil **7** and capacitor **10** is coupled to coil **3**. A third parallel resonant circuit consisting of coil **11** and capacitor **12** is also coupled to coil **3**.

Coil **5** and capacitor **6** are tuned to the frequency of a radio transmitter from which it is desired to receive information—for instance, an amplitude-modulated standard broadcast station. The signal received from this transmitter need not be strong. The signal is detected by diode **15** to obtain an audio-frequency information signal. This audio signal is coupled through a capacitor **16** and is amplified by a circuit that includes a transistor **17** having a base **20**, an emitter **21**, and a collector **22**. The amplified audio output of the transistor is coupled through an audio transformer **23** to an electroacoustical transducer, preferably a permanent-magnet dynamic loudspeaker **25** as shown.

The novelty of the invention lies largely in the method and means by which the necessary direct-current power is supplied to the emitter and collector circuits of transistor **17**. This method and means will now be described.

Coil **7** and capacitor **10**, and also coil **11** and capacitor **12**, are tuned to receive radio signals of relatively high strength. It does not matter whether these signals contain information. These power signals are rectified by diodes **26** and **27** to provide direct-current power that is filtered by capacitors **30** and **31**. The D.-C. power thus obtained is utilized to power the transistor **17**.

In the circuit shown, two tuned circuits (coil **7** and capacitor **10**, and coil **11** and capacitor **12**) are tuned to power signals and the D.-C. voltages obtained from each are connected in series. The tuned power circuits may be tuned to the same or different power signals. Under certain circumstances it may be desirable to use more than two tuned power circuits and to tune them to more than two power signals; in this way power can be obtained from several signals and combined. On the other hand, if a strong power signal is available, a single tuned power circuit may suffice to give the needed D.-C. power.

Even weak information signals can be received successfully. A plurality of transistor amplifier stages can be used if desired, or other circuits such as superheterodyne circuits can be used. It is merely necessary that a sufficiently strong power signal or signals be available to provide the small amount of D.-C. needed to power the transistors.

If the information signal happens to be strong, it can be used as the power signal; all of the tuned circuits (coil **5** and capacitor **6**, coil **7** and capacitor **10**, coil **11** and capacitor **12**) are tuned to the information signal.

Engineers who have observed my invention in operation have been surprised at the unexpectedly good results obtained, even with readily available power signals of quite moderate strength. For instance, sufficient power for satisfactory operation of a loudspeaker at low volumes is

readily obtained from a 5-kilowatt standard broadcast station 5 miles away, using only an indoor antenna to pick up the power signal as well as information signals. In typical operation under these conditions a D.-C. voltage of about 2.5 to 3 volts is obtained between the emitter and the collector, at a current of about 250 microamperes; D.-C. power input to the transistor is thus of the order of 0.5 to 1 milliwatt. So far as I am aware, no one has ever before discovered and demonstrated the practicability of this method of powering a radio receiver.

Because existing broadcast stations within a radius of a number of miles provide adequate power signals, the invention is readily practicable with existing power signals in almost any location in or near any city in the United States.

Although I have described a transistor radio receiver powered by my invention, it will be readily apparent that the invention is applicable to the powering of any transistor circuit using one or a number of transistors, and to the powering of other devices requiring relatively small amounts of power. For instance, sensitive electromechanical, electrochemical, or electrothermal devices can be operated by the method of the invention.

Referring to Figure 2, which shows a more general embodiment of my invention, an antenna **35** picks up radiofrequency energy from the atmosphere. This energy flows through coil **36**, which is coupled to a tuned circuit consisting of coil **37** and capacitor **40**. The radiofrequency voltage across capacitor **40** is rectified by diode **41** and filtered by a low-pass filter **46** consisting of capacitors **42** and **44** and choke coil **43**. The resulting D.-C. voltage is applied to a load **45**.

In the practice of my invention, larger amounts of power can be obtained for short periods of time by storing received energy in a suitable energy storage device. Stored energy may then be withdrawn at intervals at a more rapid rate than that at which it was received and put into the storage device. In this way the invention can be used to provide short pulses of relatively very high electrical energy. This result can be readily obtained by charging a relatively large capacitor with direct current and then discharging the capacitor rapidly into a load when desired. This rapid discharge can be initiated automatically when the voltage across the capacitor reaches a certain level, or it can be initiated when a transistor radio receiver receives a certain information signal.

Higher voltages can be obtained with the invention by means of well known devices for raising D.-C. voltages as shown in Figure 3. The D.-C. voltage output from the capacitor **44** can be used to power a low frequency transistor oscillator **52** whose A.-C. output is raised to a higher voltage level by the transformer **55**. This relatively high A.-C. voltage can then be rectified by a diode **61** and fed to a capacitor **64** to provide a high energy D.-C. source at a relatively high voltage level at the terminals **69** and **70**. If desired, energy can now be withdrawn from the capacitor **64** at intervals in short pulses of high energy at a high voltage level. Pulsed radio transmission is one of the possible uses for this form of the invention. Other uses would be to provide a single relatively powerful pulse needed to actuate an electrothermal or electromechanical device.

As has been indicated above, in many locations and particularly anywhere in or near most American cities, power signals normally present in the atmosphere are readily available for the easy and convenient practice of the invention. However, the invention also has important applications in systems in which the necessary power signal is generated and transmitted specifically for the operation of the particular system. Such systems can, for example, comprise a master station transmitting all the power that is needed for hundreds or thousands of fixed or mobile transistor receivers or other remote devices over a range of many miles. This eliminates the need for hundreds or thousands, as the case may be, of local power supplies. At the same time, such a system has the advantage that all of the remote devices can be simultaneously activated or deactivated at the will of the master station, simply by starting or stopping the transmission of the power signal. In such systems it will often be advantageous to use power signals of frequencies sufficiently high to permit the use of resonant receiving antennas of small physical dimensions for signal pickup at the remote devices. In addition to the power signal, the master station may transmit an information signal on the same or a different carrier.

Certain types of devices powered entirely by received radio waves are of course well known. The well-known "crystal set" of the early days of radio, which used a diode rectifier to demodulate an amplitude-modulated radiofrequency signal, is an outstanding example of such a device. My invention is readily distinguishable from such prior devices, however. In typical prior devices a modulated radiofrequency signal is applied to a diode to obtain unidirectional half-wave pulses whose amplitudes vary with modulation. These pulses are integrated by means of a capacitor to obtain a unidirectional signal the amplitude of which follows the audiofrequency modulation envelope. If the radiofrequency signal is received with sufficient strength the audio signal may have sufficient power to operate headphones or similar utilization device without power amplification; but the signal is utilized for its information content, rather than to supply non-information-containing power.

My invention, on the other hand, entails the utilization of received radiofrequency energy to supply power to at least one pair of circuit points (across capacitor **31** in Fig. 1, for example), such circuit points requiring power solely for its power content and not for any information or modulation it may contain. In other words, my invention entails the utilization of radiofrequency energy to supply power that would otherwise have to be supplied by batteries, generator, or other local power source.

It will be apparent that the embodiments shown are only exemplary and that various modifications can be made in construction and arrangement within the scope of the invention as defined in the appended claim.

I claim:

An electrical device for obtaining a high energy D.-C. source at a high voltage level using energy abstracted from the atmosphere, said device comprising in combination: resonant means for receiving radio waves, first rectifier means for converting said radio waves into first direct current energy, first capacitor means for storing said first direct current energy, an oscillator powered by said direct current energy, said oscillator producing an A.-C. output, transformer means for raising said A.-C. output to an increased voltage level, second rectifier means for converting the A.-C. output of increased voltage level from said transformer into second direct current energy, and second capacitor means for storing said second direct current energy.

References Cited in the file of this patent

UNITED STATES PATENTS

1,540,998	Plauson	June 9, 1925
2,056,824	Cawley	Oct. 6, 1936
2,553,734	Adler	May 22, 1951
2,569,687	Pan	Oct. 2, 1951
2,647,957	Mallinckrodt	Aug. 4, 1953

OTHER REFERENCES

Turner: "A crystal receiver with transistor amplifier," Radio and Television News, January 1950.

Burn things out, mess things up—that's how you learn.

Make: Electronics

Learn by Discovery

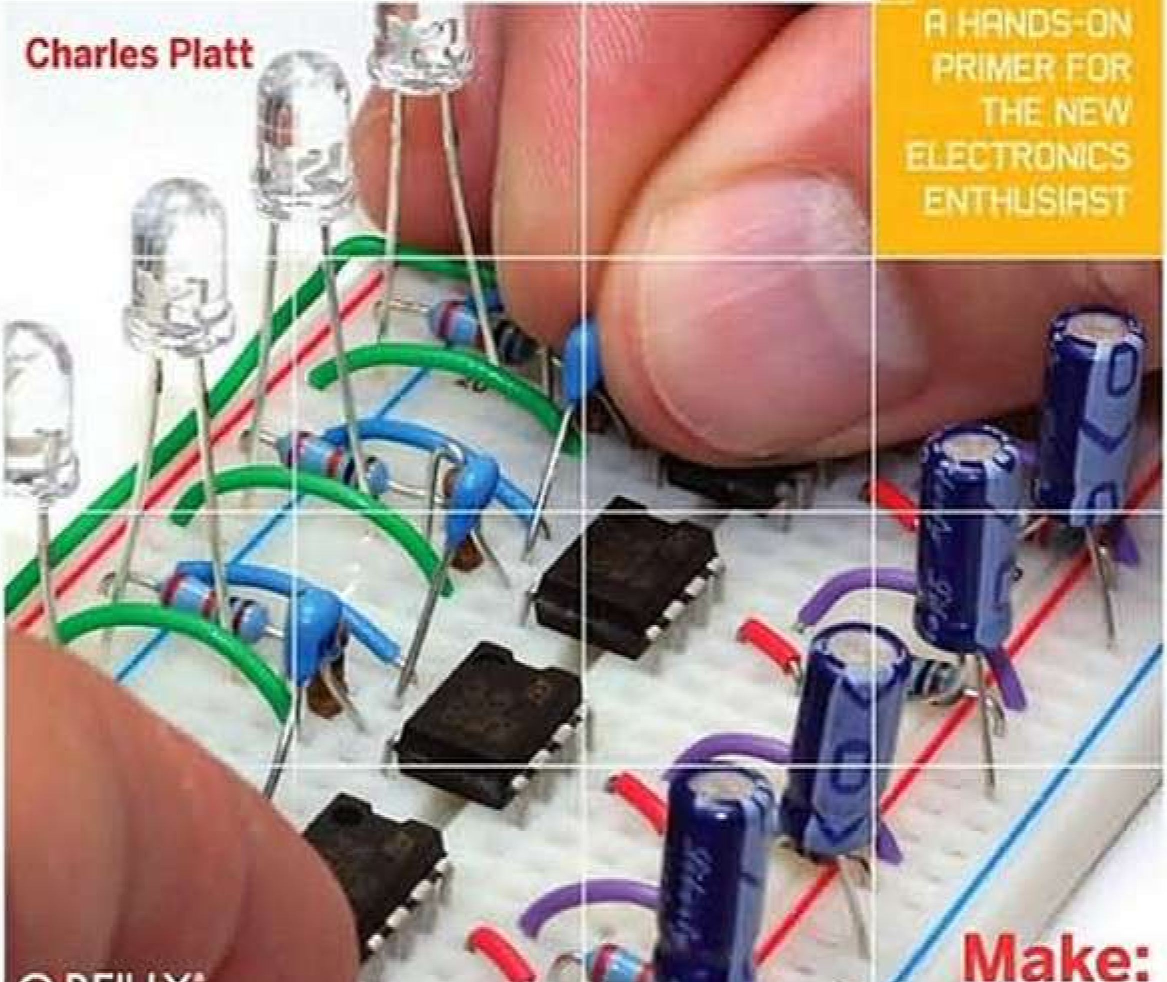

Charles Platt

A HANDS-ON PRIMER FOR THE NEW ELECTRONICS ENTHUSIAST

O'REILLY®

Make:
makezine.com

Contents

Preface ix

1. Experiencing Electricity 1

Shopping List: Experiments 1 Through 5 1

Experiment 1: Taste the Power! 5

Experiment 2: Let's Abuse a Battery! 9

Experiment 3: Your First Circuit 13

Experiment 4: Varying the Voltage 18

Experiment 5: Let's Make a Battery 32

2. Switching Basics and More 39

Shopping List: Experiments 6 Through 11 39

Experiment 6: Very Simple Switching 43

Experiment 7: Relay-Driven LEDs 55

Experiment 8: A Relay Oscillator 60

Experiment 9: Time and Capacitors 68

Experiment 10: Transistor Switching 73

Experiment 11: A Modular Project 82

3. Getting Somewhat More Serious 95

Shopping List: Experiments 12 Through 15 95

Experiment 12: Joining Two Wires Together 104

Experiment 13: Broil an LED 114

Experiment 14: A Pulsing Glow 117

Experiment 15: Intrusion Alarm Revisited 127

4. Chips, Ahoy! 147
Shopping List: Experiments 16 Through 24 147
Experiment 16: Emitting a Pulse 153
Experiment 17: Set Your Tone 162
Experiment 18: Reaction Timer 170
Experiment 19: Learning Logic 181
Experiment 20: A Powerful Combination 197
Experiment 21: Race to Place 205
Experiment 22: Flipping and Bouncing 211
Experiment 23: Nice Dice 214
Experiment 24: Intrusion Alarm Completed 223

5. What Next? 227
Shopping List: Experiments 25 Through 36 228
Customizing Your Work Area 228
Reference Sources 233
Experiment 25: Magnetism 236
Experiment 26: Tabletop Power Generation 239
Experiment 27: Loudspeaker Destruction 242
Experiment 28: Making a Coil React 246
Experiment 29: Filtering Frequencies 249
Experiment 30: Fuzz 257
Experiment 31: One Radio, No Solder, No Power 262
Experiment 32: A Little Robot Cart 268
Experiment 33: Moving in Steps 284
Experiment 34: Hardware Meets Software 293
Experiment 35: Checking the Real World 306
Experiment 36: The Lock, Revisited 311
In Closing 317

Appendix. Online Retail Sources and Manufacturers 319

Index 325

Preface

How to Have Fun with This Book

Everyone uses electronic devices, but most of us don't really know what goes on inside them.

Of course, you may feel that you don't need to know. If you can drive a car without understanding the workings of an internal combustion engine, presumably you can use an iPod without knowing anything about integrated circuits. However, understanding some basics about electricity and electronics can be worthwhile for three reasons:

- By learning how technology works, you become better able to control your world instead of being controlled by it. When you run into problems, you can solve them instead of feeling frustrated by them.
- Learning about electronics can be fun—so long as you approach the process in the right way. The tools are relatively cheap, you can do all the work on a tabletop, and it doesn't consume a lot of time (unless you want it to).
- Knowledge of electronics can enhance your value as an employee or perhaps even lead to a whole new career.

Learning by Discovery

Most introductory guides begin with definitions and facts, and gradually get to the point where you can follow instructions to build a simple circuit.

This book works the other way around. I want you to start putting components together right away. After you see what happens, you'll figure out what's going on. I believe this process of *learning by discovery* creates a more powerful and lasting experience.

Stay Within the Limits!

Although I believe that everything suggested in this book is safe, I'm assuming that you will stay within the limits that I suggest. Please always follow the instructions and pay attention to the warnings, denoted by the icon you see here. If you go beyond the limits, you will expose yourself to unnecessary risks.

Learning by discovery occurs in serious research, when scientists notice an unusual phenomenon that cannot be explained by current theory, and they start to investigate it in an effort to explain it. This may ultimately lead to a better understanding of the world.

We're going to be doing the same thing, although obviously on a much less ambitious level.

Along the way, you will make some mistakes. This is good. Mistakes are the best of all learning processes. I want you to burn things out and mess things up, because this is how you learn the limits of components and materials. Since we'll be using low voltages, there'll be no chance of electrocution, and so long as you limit the flow of current in the ways I'll describe, there will be no risk of burning your fingers or starting fires.

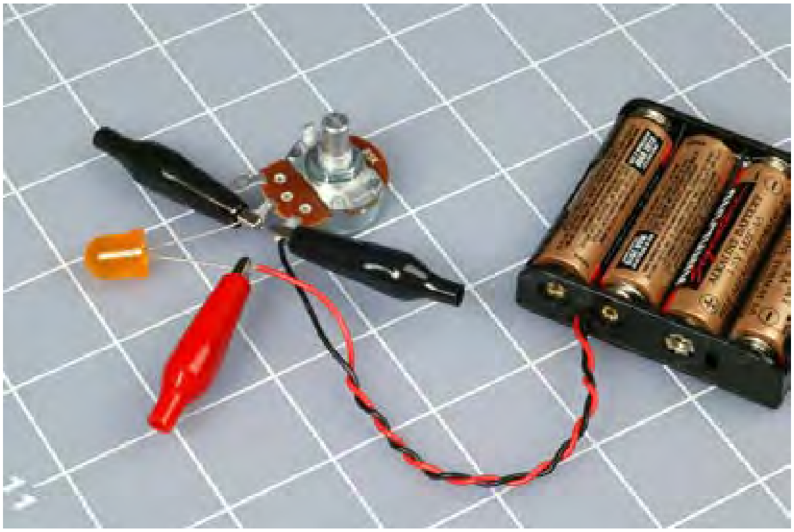

Figure P-1. *Learning by discovery allows you to start building simple circuits right away, using a handful of cheap components, a few batteries, and some alligator clips.*

How Hard Will It Be?

I assume that you're beginning with no prior knowledge of electronics. So, the first few experiments will be ultra-simple, and you won't even use solder or prototyping boards to build a circuit. You'll be holding wires together with alligator clips.

Very quickly, though, you'll be experimenting with transistors, and by the end of Chapter 2, you will have a working circuit that has useful applications.

I don't believe that hobby electronics has to be difficult to understand. Of course, if you want to study electronics more formally and do your own circuit design, this can be challenging. But in this book, the tools and supplies will be inexpensive, the objectives will be clearly defined, and the only math you'll need will be addition, subtraction, multiplication, division, and the ability to move decimal points from one position to another.

Moving Through This Book

Basically there are two ways to present information in a book of this kind: in tutorials and in reference sections. I'm going to use both of these methods. You'll find the tutorials in sections headed as follows:

- Shopping Lists
- Using Tools
- Experiments

You'll find reference sections under the following headings:

- Fundamentals
- Theory
- Background

How you use the sections is up to you. You can skip many of the reference sections and come back to them later. But if you skip many of the tutorials, this book won't be of much use to you. Learning by discovery means that you absolutely, positively have to do some hands-on work, and this in turn means that you have to buy some basic components and play with them. You will gain very little by merely imagining that you are doing this.

It's easy and inexpensive to buy what you need. In almost any urban or suburban area in the United States, chances are you live near a store that sells electronic components and some basic tools to work with them. I am referring, of course, to RadioShack franchises. Some Shacks have more components than others, but almost all of them have the basics that you'll need.

You can also visit auto supply stores such as AutoZone and Pep Boys for basics such as hookup wire, fuses, and switches, while stores such as Ace Hardware, Home Depot, and Lowe's will sell you tools.

If you prefer to buy via mail order, you can easily find everything you need by searching online. In each section of the book, I'll include the URLs of the most popular supply sources, and you'll find a complete list of URLs in the appendix.

Fundamentals

Mail-ordering components and tools

Here are the primary mail-order sources that I use myself online:

http://www.radioshack.com
RadioShack, a.k.a. The Shack. For tools and components. Not always the cheapest, but the site is easy and convenient, and some of the tools are exactly what you need.

http://www.mouser.com
Mouser Electronics.

http://www.digikey.com
Digi-Key Corporation.

http://www.newark.com
Newark.

Mouser, Digi-Key, and Newark are all good sources for components, usually requiring no minimum quantities.

http://www.allelectronics.com
All Electronics Corporation. A narrower range of components, but specifically aimed at the hobbyist, with kits available.

http://www.ebay.com
You can find surplus parts and bargains here, but you may have to try several eBay Stores to get what you want. Those based in Hong Kong are often very cheap, and I've found that they are reliable.

http://www.mcmaster.com
McMaster-Carr. Especially useful for high-quality tools.

Lowe's and Home Depot also allow you to shop online.

McMASTER-CARR
Soldering Irons & Guns
MOUSER ELECTRONICS
RobotShop
Robotics at your service!
Shop With Confidence
Contact the Experts
SERVO
Digi-Key
Mr. Clock Radio
LITTER-ROBOT
No more litter box cleaning!
ALL ELECTRONICS CORPORATION
3 1/2 DIGIT LCD MULTIMETER W/ BACK

Figure P-2. *You'll find no shortage of parts, tools, kits, and gadgets online.*

Companion Kits

Maker Shed (*www.makershed.com*) offers a number of *Make: Electronics* companion kits, both toolkits and bundles of the various components used in the book's experiments. This is a simple, convenient, and cost-effective way of getting all the tools and materials you need to do the projects in this book.

Comments and Questions

Please address comments and questions concerning this book to the publisher:

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472
800-998-9938 (in the United States or Canada)
707-829-0515 (international or local)
707-829-0104 (fax)

We have a web page for this book, where we list errata, examples, larger versions of the book's figures, and any additional information. You can access this page at:

http://oreilly.com/catalog/9780596153748

To comment or ask technical questions about this book, send email to:

bookquestions@oreilly.com

For more information about our books, conferences, Resource Centers, and the O'Reilly Network, see our website at:

http://oreilly.com

Safari® Books Online

Safari Books Online is an on-demand digital library that lets you easily search over 7,500 technology and creative reference books and videos to find the answers you need quickly.

With a subscription, you can read any page and watch any video from our library online. Read books on your cell phone and mobile devices. Access new titles before they are available for print, and get exclusive access to manuscripts in development and post feedback for the authors. Copy and paste code samples, organize your favorites, download chapters, bookmark key sections, create notes, print out pages, and benefit from tons of other time-saving features.

O'Reilly Media has uploaded this book to the Safari Books Online service. To have full digital access to this book and others on similar topics from O'Reilly and other publishers, sign up for free at *http://my.safaribooksonline.com*.

1

Experiencing Electricity

I want you to get a taste for electricity—literally!—in the first experiment. This first chapter of the book will show you:

- How to understand and measure electricity and resistance
- How to handle and connect components without overloading, damaging, or destroying them

Even if you have some prior knowledge of electronics, you should try these experiments before you venture on to the rest of the book.

IN THIS CHAPTER

Shopping List: Experiments 1 Through 5
Experiment 1: Taste the Power!
Experiment 2: Let's Abuse a Battery!
Experiment 3: Your First Circuit
Experiment 4: Varying the Voltage
Experiment 5: Let's Make a Battery

*Maker Shed (*www.makershed.com*) has put together a series of* Make: Electronics *companion kits. These include all of the tools and components used in book's experiments. This is a quick, simple, and cost-effective way of getting everything you need to complete the projects in this book.*

Shopping List: Experiments 1 Through 5

If you want to limit your number of shopping trips or online purchases, look ahead in the book for additional shopping lists, and combine them to make one bulk purchase.

In this first chapter, I will give you part numbers and sources for every tool and component that we'll be using. Subsequently, I won't expect you to need such specific information, because you will have gained experience searching for items on your own.

Tools

Small pliers

RadioShack Kronus 4.5-inch, part number 64-2953 or Xcelite 4-inch mini long-nose pliers, model L4G.

Or similar. See Figures 1-1 through 1-3. Look for these tools in hardware stores and the sources listed in the preface. The brand is unimportant. After you use them for a while, you'll develop your own preferences. In particular, you have to decide whether you like spring-loaded handles. If you decide you don't, you'll need a second pair of pliers to pull the springs out of the first.

Wire cutters

RadioShack Kronus 4.5-inch, part number 64-2951, or Stanley 7-inch model 84-108.

Or similar. Use them for cutting copper wire, not harder metals (Figure 1-4).

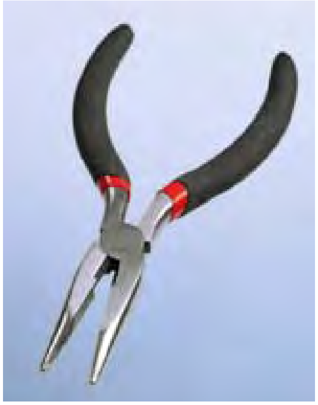

Figure 1-1. *Generic long-nosed pliers are your most fundamental tool for gripping, bending, and picking things up after you drop them.*

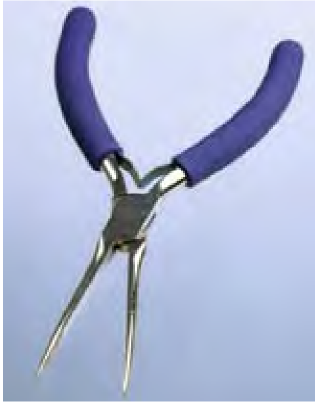

Figure 1-2. *Longer-nosed pliers: these are useful for reaching into tiny spaces.*

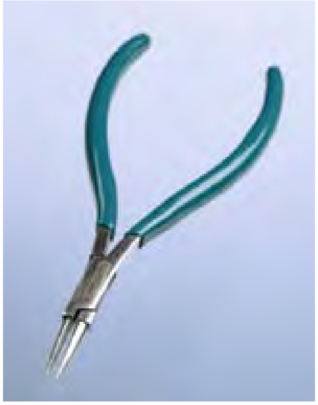

Figure 1-3. *Sharp-pointed pliers are designed for making jewelry, but are also useful for grabbing tiny components.*

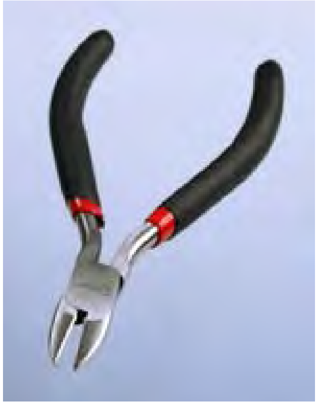

Figure 1-4. *Wire cutters, sometimes known as side cutters, are essential.*

Multimeter

Extech model EX410 or BK Precision model 2704-B or Amprobe model 5XP-A.

Or similar. Because electricity is invisible, we need a tool to visualize the pressure and flow, and a meter is the only way. A cheap meter will be sufficient for your initial experiments. If you buy online, try to check customer reviews, because reliability may be a problem for cheap meters. You can shop around for retailers offering the best price. Don't forget to search on eBay.

The meter must be digital—don't get the old-fashioned analog kind with a needle that moves across a set of printed scales. This book assumes that you are looking at a digital display.

I suggest that you do not buy an autoranging meter. "Autoranging" sounds useful—for example, when you want to check a 9-volt battery, the meter figures out for itself that you are not trying to measure hundreds of volts, nor fractions of a volt. The trouble is that this can trick you into making errors. What if the battery is almost dead? Then you may be measuring a fraction of a volt without realizing it. The only indication will be an easily overlooked "m" for "millivolts" beside the large numerals of the meter display.

On a manual-ranging meter, you select the range, and if the source that you are measuring is outside of that range, the meter tells you that you made an error. I prefer this. I also get impatient with the time it takes for the autoranging feature to figure out the appropriate range each time I make a measurement. But it's a matter of personal preference. See Figures 1-5 through 1-7 for some examples of multimeters.

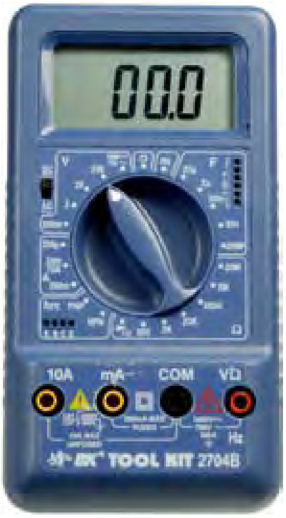

Figure 1-5. *You can see by the wear and tear that this is my own favorite meter. It has all the necessary basic features and can also measure capacitance (the F section, for Farads). It can also check transistors. You have to choose the ranges manually.*

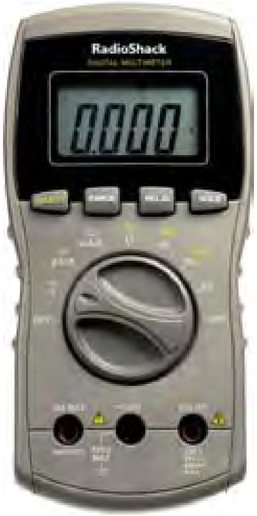

Figure 1-6. *Mid-priced RadioShack meter, which has the basic features; however, the dual purpose for each dial position, selected with the SELECT button, may be confusing. This is an autoranging meter.*

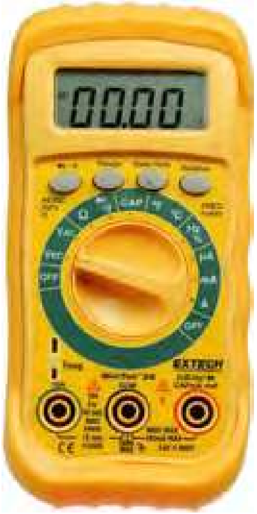

Figure 1-7. *An autoranging meter from Extech offers basic functions, plus a temperature probe, which may be useful to check whether components such as power supplies are running unduly hot.*

Supplies

Batteries

9-volt battery. Quantity: 1.

AA batteries, 1.5 volts each. Quantity: 6.

The batteries should be disposable alkaline, the cheapest available, because we may destroy some of them. You should *absolutely not* use rechargeable batteries in Experiments 1 and 2.

Battery holders and connectors

Snap connector for 9-volt battery, with wires attached (Figure 1-8). Quantity: 1. RadioShack part number 270-325 or similar. Any snap connector that has wires attached will do.

Battery holder for single AA cell, with wires attached (Figure 1-9). Quantity: 1. RadioShack part number 270-401 or Mouser.com catalog number 12BH410-GR, or similar; any single-battery holder that has thin wires attached will do.

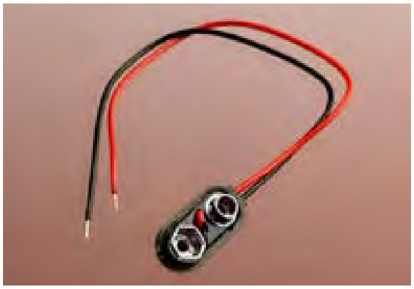

Figure 1-8. *Snap connector for a 9-volt battery.*

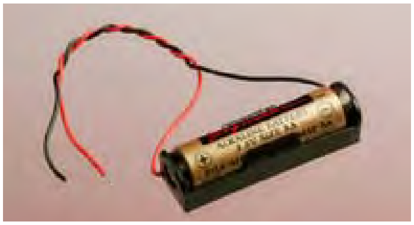

Figure 1-9. *Single AA-sized battery carrier with wires.*

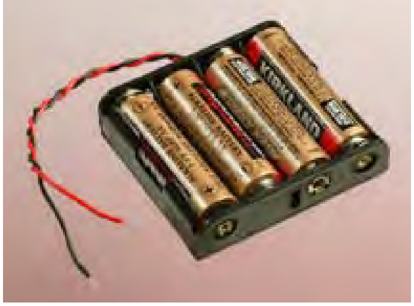

Figure 1-10. *Battery carrier for four AA cells, to be installed in series, delivering 6 volts.*

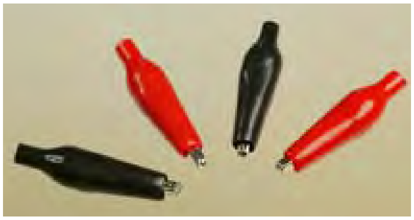

Figure 1-11. *Alligator clips inside vinyl sheaths, which reduce the chance of accidental short circuits.*

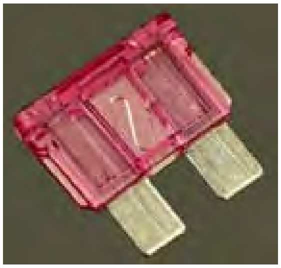

Figure 1-12. *A 3-amp fuse intended primarily for automotive use, shown here larger than actual size.*

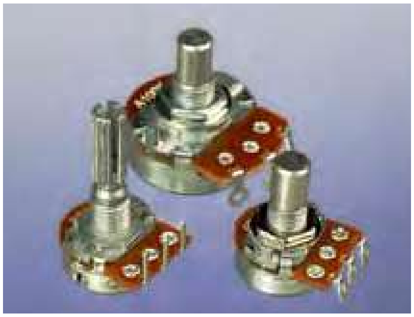

Figure 1-13. *Potentiometers come in many shapes and sizes, with different lengths of shafts intended for different types of knobs. For our purposes, any style will do, but the larger-sized ones are easier to play with.*

Battery holder for four AA cells, with wires attached (Figure 1-10). Quantity: 1. All Electronics catalog number BH-342 or RadioShack part 270-391 or similar.

Alligator clips

Vinyl-insulated. Quantity: at least 6. All Electronics catalog number ALG-28 or RadioShack part number 270-1545 or similar (Figure 1-11).

Components

You may not know what some of these items are, or what they do. Just look for the part numbers and descriptions, and match them with the photographs shown here. Very quickly, in the learning by discovery process, all will be revealed.

Fuses

Automotive-style, mini-blade type, 3 amps. Quantity: 3. RadioShack part number 270-1089, or Bussmann part ATM-3, available from automotive parts suppliers such as AutoZone (Figure 1-12).

Or similar. A blade-type fuse is easier to grip with alligator clips than a round cartridge fuse.

Potentiometers

Panel-mount, single-turn, 2K linear, 0.1 watt minimum. Quantity: 2. Alpha part RV170F-10-15R1-B23 or BI Technologies part P160KNPD-2QC25B2K, from Mouser.com or other component suppliers (Figure 1-13).

Or similar. The "watt" rating tells you how much power this component can handle. You don't need more than 0.5 watts.

Resistors

Assortment 1/4-watt minimum, various values but must include 470 ohms, 1K, and 2K or 2.2K. Quantity: at least 100. RadioShack part number 271-312.

Or search eBay for "resistor assorted."

Light-emitting diodes (LEDs)

Any size or color (Figures 1-14 and 1-15). Quantity: 10. RadioShack part number 276-1622 or All Spectrum Electronics part K/LED1 from Mouser.com.

Or similar. Just about any LEDs will do for these first experiments.

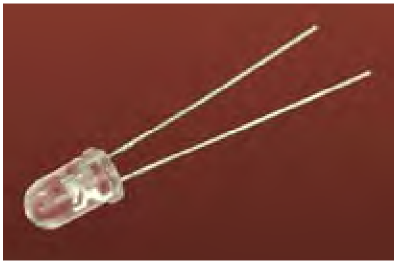

Figure 1-14. *Typical 5-mm diameter light-emitting diode (LED).*

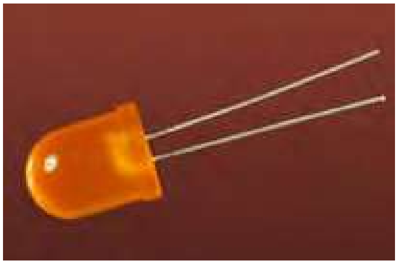

Figure 1-15. *Jumbo-sized LED (1 cm diameter) is not necessarily brighter or more expensive. For most of the experiments in this book, buy whatever LEDs you like the look of.*

Experiment 1: Taste the Power!

Can you taste electricity? Maybe not, but it feels as if you can.

You will need:

- 9-volt battery
- Snap connector for battery terminals
- Multimeter

No More Than 9 Volts

A 9-volt battery won't hurt you. But do not try this experiment with a higher-voltage battery or a larger battery that can deliver more current. Also, if you have metal braces on your teeth, be very careful not to touch them with the battery.

Procedure

Moisten your tongue and touch the tip of it to the metal terminals of a 9-volt battery. The sudden sharp tingle that you feel is caused by electricity flowing from one terminal of the battery (Figure 1-16), through the moisture on and in your tongue, to the other terminal. Because the skin of your tongue is very thin (it's actually a mucus membrane) and the nerves are close to the surface, you can feel the electricity very easily.

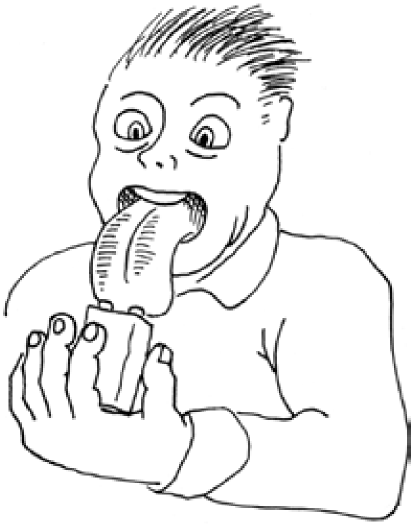

Figure 1-16. *Step 1 in the process of learning by discovery: the 9-volt tongue test.*

Now stick out your tongue, dry the tip of it very thoroughly with a tissue, and repeat the experiment without allowing your tongue to become moist again. You should feel less of a tingle.

What's happening here? We're going to need a meter to find out.

Tools

Setting up your meter

Check the instructions that came with the meter to find out whether you have to install a battery in it, or whether a battery is preinstalled.

Most meters have removable wires, known as *leads* (pronounced "leeds"). Most meters also have three sockets on the front, the leftmost one usually being reserved to measure high electrical currents (flows of electricity). We can ignore that one for now.

The leads will probably be black and red. The black wire plugs into a socket labeled "COM" or "Common." Plug the red one into the socket labeled "V" or "volts." See Figures 1-17 through 1-20.

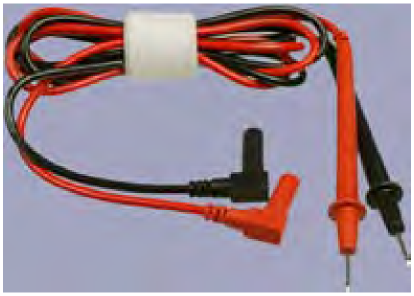

Figure 1-17. *The black lead plugs into the Common (COM) socket, and the red lead plugs into the red socket that's almost always on the righthand side of a multimeter.*

The other ends of the leads terminate in metal spikes known as *probes*, which you will be touching to components when you want to make electrical measurements. The probes detect electricity; they don't emit it in significant quantities. Therefore, they cannot hurt you unless you poke yourself with their sharp ends.

If your meter doesn't do autoranging, each position on the dial will have a number beside it. This number means "no higher than." For instance if you want to check a 6-volt battery, and one position on the voltage section of the dial is numbered 2 and the next position is numbered 20, position 2 means "no higher than 2 volts." You have to go to the next position, which means "no higher than 20 volts."

If you make a mistake and try to measure something inappropriate, the meter will show you an error message such as "E" or "L." Turn the dial and try again.

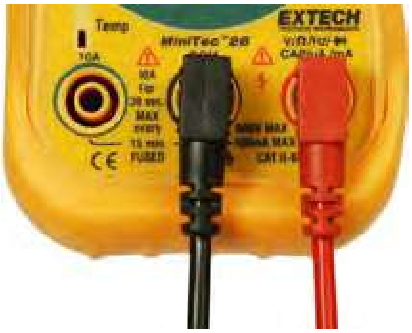

Figure 1-18

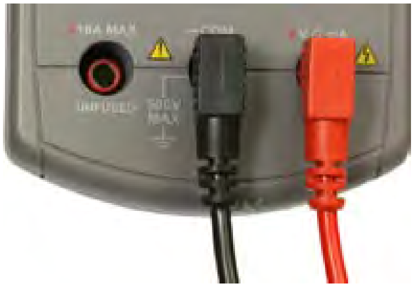

Figure 1-19

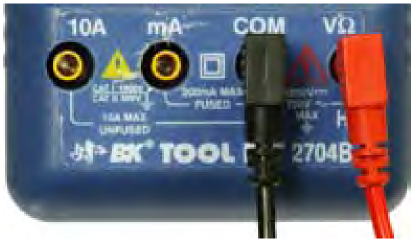

Figure 1-20. *To measure resistance and voltage, plug the black lead into the Common socket and the red lead into the Volts socket. Almost all meters have a separate socket where you must plug the red lead when you measure large currents in amps, but we'll be dealing with this later.*

FUNDAMENTALS

Ohms

We measure distance in miles or kilometers, weight in pounds or kilograms, temperature in Fahrenheit or Centigrade—and electrical resistance in ohms. The ohm is an international unit.

The Greek omega symbol (Ω) is used to indicate ohms, as shown in Figures 1-21 and 1-22. Letter K (or alternatively, KΩ) means a kilohm, which is 1,000 ohms. Letter M (or MΩ) means a megohm, which is 1,000,000 ohms.

Number of ohms	Usually expressed as	Abbreviated as
1,000 ohms	1 kilohm	1KΩ or 1K
10,000 ohms	10 kilohms	10KΩ or 10K
100,000 ohms	100 kilohms	100KΩ or 100K
1,000,000 ohms	1 megohm	1MΩ or 1M
10,000,000 ohms	10 megohms	10MΩ or 10M

A material that has very high resistance to electricity is known as an *insulator*. Most plastics, including the colored sheaths around wires, are insulators.

A material with very low resistance is a *conductor*. Metals such as copper, aluminum, silver, and gold are excellent conductors.

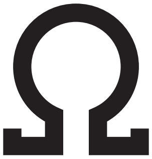

Figure 1-21. *The omega symbol is used internationally to indicate resistance on ohms.*

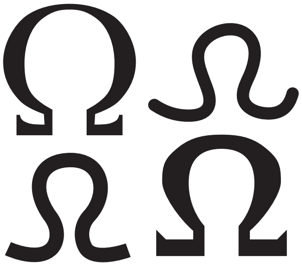

Figure 1-22. *You'll find it printed or written in a wide variety of styles.*

Procedure

We're going to use the meter to discover the electrical resistance of your tongue. First, set your meter to measure resistance. If it has autoranging, look to see whether it is displaying a K, meaning kilohms, or M, meaning megohms. If you have to set the range manually, begin with no less than 100,000 ohms (100K). See Figures 1-23 through 1-25.

Touch the probes to your tongue, about an inch apart. Note the reading, which should be around 50K. Now put aside the probes, stick out your tongue, and use a tissue to dry it very carefully and thoroughly. Without allowing your tongue to become moist again, repeat the test, and the reading should be higher. Finally, press the probes against the skin of your hand or arm: you may get no reading at all, until you moisten your skin.

When your skin is moist (for instance, if you perspire), its electrical resistance decreases. This principle is used in lie detectors, because someone who knowingly tells a lie, under conditions of stress, tends to perspire.

A 9-volt battery contains chemicals that liberate electrons (particles of electricity), which want to flow from one terminal to the other as a result of a chemical reaction inside it. Think of the cells inside a battery as being like two water tanks—one of them full, the other empty. If they are connected with a pipe, water flows between them until their levels are equal. Figure 1-26 may help you visualize this. Similarly, when you open up an electrical pathway between the two sides of a battery, electrons flow between them, even if the pathway consists only of the moisture on your tongue.

Electrons flow more easily through some substances (such as a moist tongue) than others (such as a dry tongue).

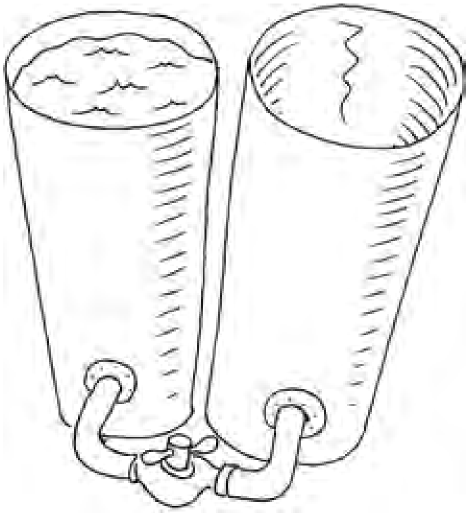

Figure 1-26. *Think of the cells in a battery as being like two cylinders: one full of water, the other empty. Open a connection between the cylinders, and the water will flow until the levels are equal on both sides. The less resistance in the connection, the faster the flow will be.*

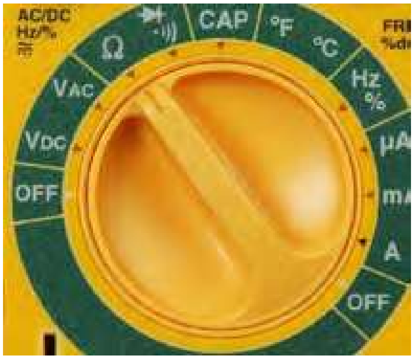

Figure 1-23

Figure 1-24

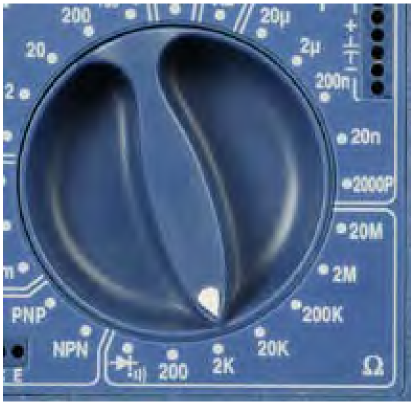

Figure 1-25. *To measure ohms, turn the dial to the ohm (omega) symbol. On an autoranging meter, you can then press the Range button repeatedly to display different ranges of resistance, or simply touch the probes to a resistance and wait for the meter to choose a range automatically. A manual meter requires you to select the range with the dial (you should set it to 100K or higher, to measure skin resistance). If you don't get a meaningful reading, try a different range.*

BACKGROUND

The man who discovered resistance

Georg Simon Ohm, pictured in Figure 1-27, was born in Bavaria in 1787 and worked in obscurity for much of his life, studying the nature of electricity using metal wire that he had to make for himself (you couldn't truck on down to Home Depot for a spool of hookup wire back in the early 1800s).

Despite his limited resources and inadequate mathematical abilities, Ohm was able to demonstrate in 1827 that the electrical resistance of a conductor such as copper varied in proportion with its area of cross-section, and the current flowing through it is proportional to the voltage applied to it, as long as temperature is held constant. Fourteen years later, the Royal Society in London finally recognized the significance of his contribution and awarded him the Copley Medal. Today, his discovery is known as Ohm's Law.

Figure 1-27. *Georg Simon Ohm, after being honored for his pioneering work, most of which he pursued in relative obscurity.*

Further Investigation

Attach the snap-on terminal cap (shown earlier in Figure 1-8) to the 9-volt battery. Take the two wires that are attached to the cap and hold them so that the bare ends are just a few millimeters apart. Touch them to your tongue. Now separate the ends of the wires by a couple of inches, and touch them to your tongue again. (See Figure 1-28.) Notice any difference?

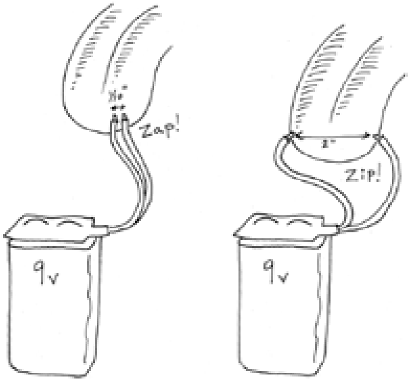
Zap!
9v
Zip!
9v

Figure 1-28. *Modifying the tongue test to show that a shorter distance, with lower resistance, allows greater flow of electricity, and a bigger zap.*

Use your meter to measure the electrical resistance of your tongue, this time varying the distance between the two probes. When electricity travels through a shorter distance, it encounters less total resistance. As a result, the current (the flow of electricity per second) increases. You can try a similar experiment on your arm, as shown in Figure 1-29.

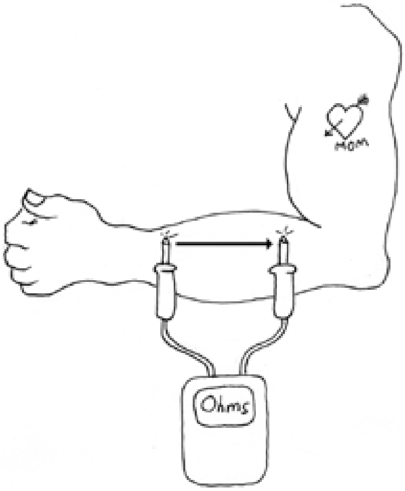

Figure 1-29. *Moisten your skin before trying to measure its resistance. You should find that the resistance goes up as you move the meter probes farther apart. The resistance is proportional to the distance.*

Use your meter to test the electrical resistance of water. Dissolve some salt in the water, and test it again. Now try measuring the resistance of distilled water (in a clean glass).

The world around you is full of materials that conduct electricity with varying amounts of resistance.

Cleanup and Recycling

Your battery should not have been damaged or significantly discharged by this experiment. You'll be able to use it again.

Remember to switch off your meter before putting it away.

Experiment 2: Let's Abuse a Battery!

To get a better feeling for electrical power, you're going to do what most books tell you not to do. You're going to short out a battery. A short circuit is a direct connection between the two sides of a power source.

Short Circuits

Short circuits can be dangerous. Do not short out a power outlet in your home: there'll be a loud bang, a bright flash, and the wire or tool that you use will be partially melted, while flying particles of melted metal can burn you or blind you.

If you short out a car battery, the flow of current is so huge that the battery might even explode, drenching you in acid (Figure 1-30).

Lithium batteries are also dangerous. Never short-circuit a lithium battery: it can catch fire and burn you (Figure 1-31).

Use only an alkaline battery in this experiment, and only a single AA cell (Figure 1-32). You should also wear safety glasses in case you happen to have a defective battery.

You will need:

- 1.5-volt AA battery
- Single-battery carrier
- 3-amp fuse
- Safety glasses (regular eyeglasses or sunglasses will do)

Procedure

Use an alkaline battery. Do not use any kind of rechargeable battery.

Put the battery into a battery holder that's designed for a single battery and has two thin insulated wires emerging from it, as shown in Figure 1-32. Do not use any other kind of battery holder.

Use an alligator clip to connect the bare ends of the wires, as shown in Figure 1-32. There will be no spark, because you are using only 1.5 volts. Wait one minute, and you'll find that the wires are getting hot. Wait another minute, and the battery, too, will be hot.

Figure 1-30. *Anyone who has dropped an adjustable wrench across the bare terminals of a car battery will tell you that short circuits can be dramatic at a "mere" 12 volts, if the battery is big enough.*

Figure 1-31. *The low internal resistance of lithium batteries (which are often used in laptop computers) allows high currents to flow, with unexpected results. Never fool around with lithium batteries!*

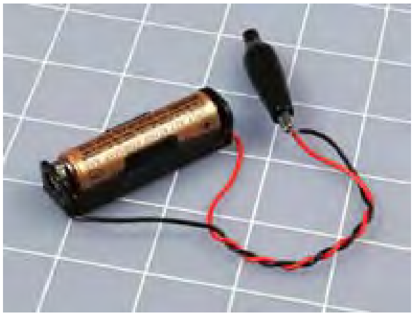

Figure 1-32. *Shorting out an alkaline battery can be safe if you follow the directions precisely. Even so, the battery is liable to become too hot to touch comfortably. Don't try this with any type of rechargeable battery.*

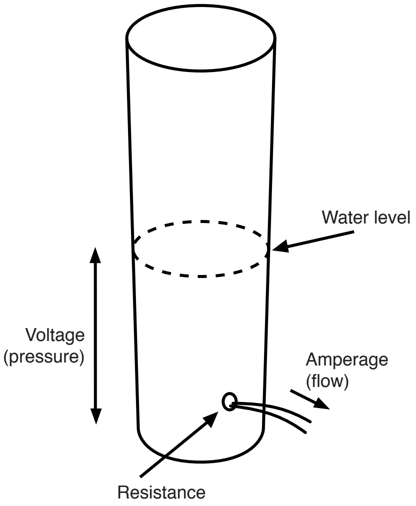

Figure 1-33. *Think of voltage as pressure, and amperes as flow.*

The heat is caused by electricity flowing through the wires and through the electrolyte (the conductive fluid) inside the battery. If you've ever used a hand pump to force air into a bicycle tire, you know that the pump gets warm. Electricity behaves in much the same way. You can imagine the electricity being composed of particles (electrons) that make the wire hot as they push through it. This isn't a perfect analogy, but it's close enough for our purposes.

Chemical reactions inside the battery create electrical pressure. The correct name for this pressure is *voltage*, which is measured in volts and is named after Alessandro Volta, an electrical pioneer.

Going back to the water analogy: the height of the water in a tank is proportionate to the pressure of the water, and comparable to voltage. Figure 1-33 may help you to visualize this.

But volts are only half of the story. When electrons flow through a wire, the flow is known as *amperage*, named after yet another electrical pioneer, André-Marie Ampère. The flow is also generally known as current. It's the current—the amperage—that generates the heat.

BACKGROUND

Why didn't your tongue get hot?

When you touched the 9-volt battery to your tongue, you felt a tingle, but no perceptible heat. When you shorted out a battery, you generated a noticeable amount of heat, even though you used a lower voltage. How can we explain this?

The electrical resistance of your tongue is very high, which reduces the flow of electrons. The resistance of a wire is very low, so if there's only a wire connecting the two terminals of the battery, more current will pass through it, creating more heat. If all other factors remain constant:

- Lower resistance allows more current to flow (Figure 1-34).
- The heat generated by electricity is proportional to the amount of electricity (the current) that flows.

Here are some other basic concepts:

- The flow of electricity per second is measured in amperes, or amps.
- The pressure of electricity causes the flow, measured in volts.
- The resistance to the flow is measured in ohms.
- A higher resistance restricts the current.
- A higher voltage overcomes resistance and increases the current.

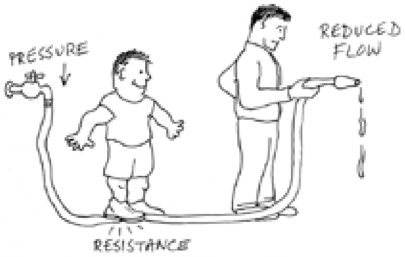

Figure 1-34. *Larger resistance results in smaller flow—but if you increase the pressure, it may overcome the resistance and increase the flow.*

If you're wondering exactly how much current flows between the terminals of a battery when you short it out, that's a difficult question to answer. If you try to use your multimeter to measure it, you're liable to blow the fuse inside the meter. Still, you can use your very own 3-amp fuse, which we can sacrifice because it didn't cost very much.

First inspect the fuse very carefully, using a magnifying glass if you have one. You should see a tiny S-shape in the transparent window at the center of the fuse. That S is a thin section of metal that melts easily.

Remove the battery that you short-circuited. It is no longer useful for anything, and should be recycled if possible. Put a fresh battery into the battery carrier, connect the fuse as shown in Figure 1-35, and take another look. You should see a break in the center of the S shape, where the metal melted almost instantly. Figure 1-36 shows the fuse before you connected it, and Figure 1-37 depicts a blown fuse. This is how a fuse works: it melts to protect the rest of the circuit. That tiny break inside the fuse stops any more current from flowing.

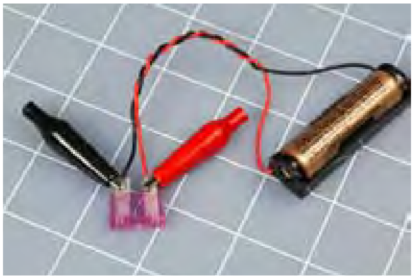

Figure 1-35. *When you attach both wires to the fuse, the little S-shaped element inside will melt almost instantly.*

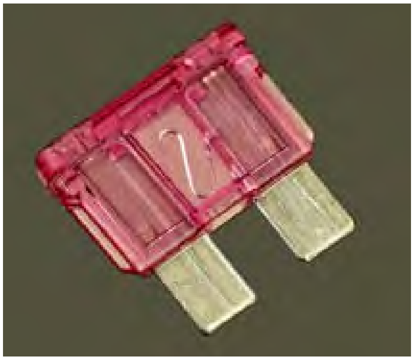

Figure 1-36. *A 3-amp fuse, before its element was melted by a single 1.5-volt battery.*

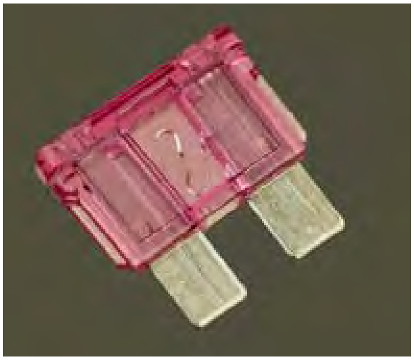

Figure 1-37. *The same fuse after being melted by electric current.*

FUNDAMENTALS

Volt basics

Electrical pressure is measured in volts. The volt is an international unit. A millivolt is 1/1,000 of a volt.

Number of volts	Usually expressed as	Abbreviated as
0.001 volts	1 millivolt	1 mV
0.01 volts	10 millivolts	10 mV
0.1 volts	100 millivolts	100 mV
1 volt	1,000 millivolts	1V

Ampere basics

We measure electrical flow in amperes, or amps. The ampere is an international unit, often referred to as an "amp." A milliamp is 1/1,000 of an ampere.

Number of amperes	Usually expressed as	Abbreviated as
0.001 amps	1 milliamp	1 mA
0.01 amps	10 milliamps	10 mA
0.1 amps	100 milliamps	100 mA
1 amp	1,000 milliamps	1A

BACKGROUND

Inventor of the battery

Alessandro Volta (Figure 1-38) was born in Italy in 1745, long before science was broken up into specialties. After studying chemistry (he discovered methane in 1776), he became a professor of physics and became interested in the so-called galvanic response, whereby a frog's leg will twitch in response to a jolt of static electricity.

Using a wine glass full of salt water, Volta demonstrated that the chemical reaction between two electrodes, one made of copper, the other of zinc, will generate a steady electric current. In 1800, he refined his apparatus by stacking plates of copper and zinc, separated by cardboard soaked in salt and water. This "voltaic pile" was the first electric battery.

Figure 1-38. *Alessandro Volta discovered that chemical reactions can create electricity.*

FUNDAMENTALS

Direct and alternating current

The flow of current that you get from a battery is known as *direct current*, or DC. Like the flow of water from a faucet, it is a steady stream, in one direction.

The flow of current that you get from the "hot" wire in a power outlet in your home is very different. It changes from positive to negative 50 times each second (in Great Britain and some other nations, 60 times per second). This is known as *alternating current*, or AC, which is more like the pulsatile flow you get from a power washer.

Alternating current is essential for some purposes, such as cranking up voltage so that electricity can be distributed over long distances. AC is also useful in motors and domestic appliances. The parts of an American power outlet are shown in Figure 1-39. A few other nations, such as Japan, also use American-style outlets.

For most of this book I'm going to be talking about DC, for two reasons: first, most simple electronic circuits are powered with DC, and second, the way it behaves is much easier to understand.

I won't bother to mention repeatedly that I'm dealing with DC. Just assume that everything is DC unless otherwise noted.

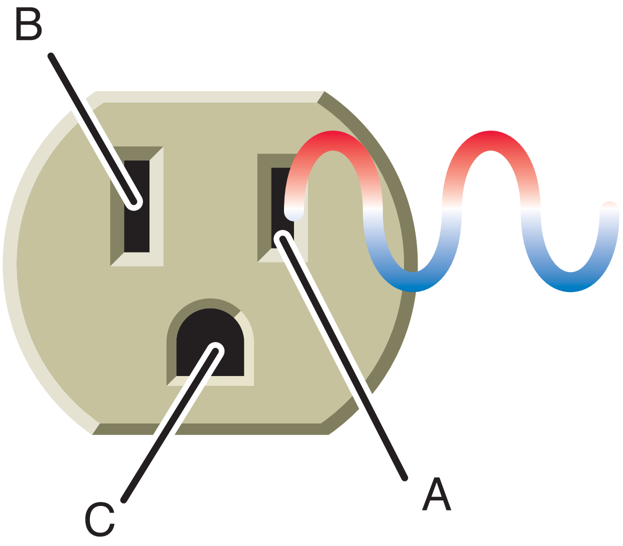

Figure 1-39. *This style of power outlet is found in North America, South America, Japan, and some other nations. European outlets look different, but the principle remains the same. Socket A is the "live" side of the outlet, supplying voltage that alternates between positive and negative, relative to socket B, which is called the "neutral" side. If an appliance develops a fault such as an internal loose wire, it should protect you by sinking the voltage through socket C, the ground.*

Cleanup and Recycling

The first AA battery that you shorted out is probably damaged beyond repair. You should dispose of it. Putting batteries in the trash is not a great idea, because they contain heavy metals that should be kept out of the ecosystem. Your state or town may include batteries in a local recycling scheme. (California requires that almost all batteries be recycled.) You'll have to check your local regulations for details.

The blown fuse is of no further use, and can be thrown away.

The second battery, which was protected by the fuse, should still be OK. The battery holder also can be reused later.

Experiment 3: Your First Circuit

Now it's time to make electricity do something that's at least slightly useful. For this purpose, you'll use components known as resistors, and a light-emitting diode, or LED.

You will need:

- 1.5-volt AA batteries. Quantity: 4.
- Four-battery holder. Quantity: 1.
- Resistors: 470Ω, 1K, and either 2K or 2.2K (the 2.2K value happens to be more common than 2K, but either will do in this experiment). Quantity: 1 of each resistor.
- An LED, any type. Quantity: 1.
- Alligator clips. Quantity: 3.

Setup

It's time to get acquainted with the most fundamental component we'll be using in electronic circuits: the humble resistor. As its name implies, it resists the flow of electricity. As you might expect, the value is measured in ohms.

If you bought a bargain-basement assortment package of resistors, you may find nothing that tells you their values. That's OK; we can find out easily enough. In fact, even if they are clearly labeled, I want you to check their values yourself. You can do it in two ways:

- Use your multimeter. This is excellent practice in learning to interpret the numbers that it displays.
- Learn the color codes that are printed on most resistors. See the following section, "Fundamentals: Decoding resistors," for instructions.

After you check them, it's a good idea to sort them into labeled compartments in a little plastic parts box. Personally, I like the boxes sold at the Michaels chain of crafts stores, but you can find them from many sources.

BACKGROUND

Father of electromagnetism

Born in 1775 in France, André-Marie Ampère (Figure 1-40) was a mathematical prodigy who became a science teacher, despite being largely self-educated in his father's library. His best-known work was to derive a theory of electromagnetism in 1820, describing the way that an electric current generates a magnetic field. He also built the first instrument to measure the flow of electricity (now known as a *galvanometer*), and discovered the element fluorine.

Figure 1-40. *Andre-Marie Ampere found that an electric current running through a wire creates a magnetic field around it. He used this principle to make the first reliable measurements of what came to be known as amperage.*

FUNDAMENTALS

Decoding resistors

Some resistors have their value clearly stated on them in microscopic print that you can read with a magnifying glass. Most, however, are color-coded with stripes. The code works like this: first, ignore the color of the body of the resistor. Second, look for a silver or gold stripe. If you find it, turn the resistor so that the stripe is on the righthand side. Silver means that the value of the resistor is accurate within 10%, while gold means that the value is accurate within 5%. If you don't find a silver or gold stripe, turn the resistor so that the stripes are clustered at the left end. You should now find yourself looking at three colored stripes on the left. Some resistors have more stripes, but we'll deal with those in a moment. See Figures 1-41 and 1-42.

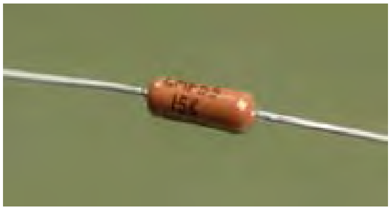

Figure 1-41. *Some modern resistors have their values printed on them, although you may need a magnifier to read them. This 15K resistor is less than half an inch long.*

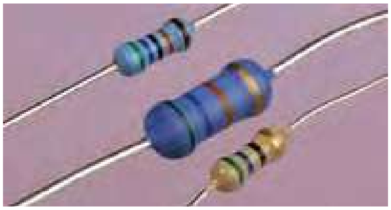

Figure 1-42. *From top to bottom, these resistor values are 56,000 ohms (56K), 5,600 ohms (5.6K), and 560 ohms. The size tells you how much power the resistor can handle; it has nothing to do with the resistance. The smaller components are rated at 1/4 watt; the larger one in the center can handle 1 watt of power.*

Starting from the left, the first and second stripes are coded according to this table:

Black	0
Brown	1
Red	2
Orange	3
Yellow	4
Green	5
Blue	6
Violet	7
Gray	8
White	9

The third stripe has a different meaning: It tells you how many zeros to add, like this:

Black	-	No zeros
Brown	0	1 zero
Red	00	2 zeros
Orange	000	3 zeros
Yellow	0000	4 zeros
Green	00000	5 zeros
Blue	000000	6 zeros
Violet	0000000	7 zeros
Gray	00000000	8 zeros
White	000000000	9 zeros

FUNDAMENTALS

Decoding resistors (continued)

Note that the color-coding is consistent, so that green, for instance, means either a value of 5 (for the first two stripes) or 5 zeros (for the third stripe). Also, the sequence of colors is the same as their sequence in a rainbow.

So, a resistor colored brown-red-green would have a value of 1-2 and five zeros, making 1,200,000 ohms, or 1.2MΩ. A resistor colored orange-orange-orange would have a value of 3-3 and three zeros, making 33,000 ohms, or 33KΩ. A resistor colored brown-black-red would have a value of 1-0 and two additional zeros, or 1KΩ. Figure 1-43 shows some other examples.

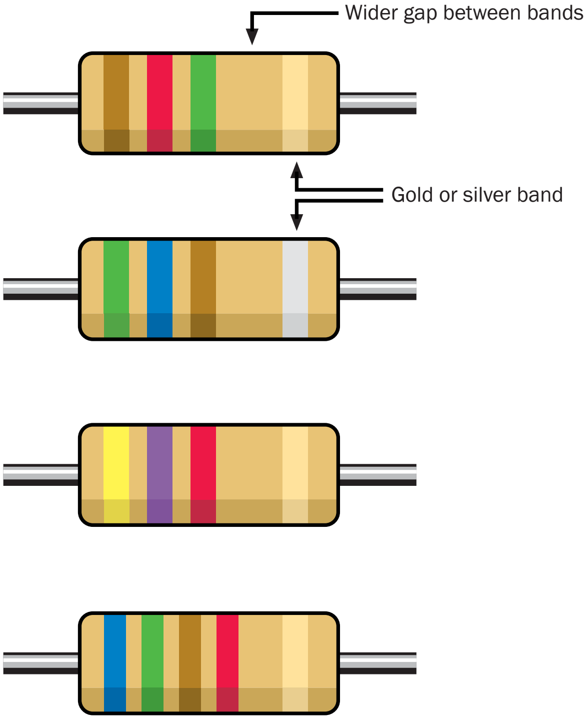

Figure 1-43. *To read the value of a resistor, first turn it so that the silver or gold stripe is on the right, or the other stripes are clustered on the left. From top to bottom: The first resistor has a value of 1-2 and five zeros, or 1,200,000, which is 1.2MΩ. The second is 5-6 and one zero, or 560Ω. The third is 4-7 and two zeros, or 4,700, which is 4.7KΩ. The last is 6-5-1 and two zeros, or 65,100Ω, which is 65.1KΩ.*

If you run across a resistor with four stripes instead of three, the first *three* stripes are digits and the *fourth* stripe is the number of zeros. The third numeric stripe allows the resistor to be calibrated to a finer tolerance.

Confusing? Absolutely. That's why it's easier to use your meter to check the values. Just be aware that the meter reading may be slightly different from the claimed value of the resistor. This can happen because your meter isn't absolutely accurate, or because the resistor is not absolutely accurate, or both. As long as you're within 5% of the claimed value, it doesn't matter for our purposes.

Lighting an LED

Now take a look at one of your LEDs. An old-fashioned lightbulb wastes a lot of power by converting it into heat. LEDs are much smarter: they convert almost all their power into light, and they last almost indefinitely—as long as you treat them right!

An LED is quite fussy about the amount of power it gets, and the way it gets it. Always follow these rules:

- The *longer* wire protruding from the LED must receive a *more positive* voltage than the shorter wire.
- The voltage difference between the long wire and the short wire must not exceed the limit stated by the manufacturer.
- The current passing through the LED must not exceed the limit stated by the manufacturer.

What happens if you break these rules? Well, we're going to find out!

Make sure you are using fresh batteries. You can check by setting your multimeter to measure volts DC, and touching the probes to the terminals of each battery. You should find that each of them generates a pressure of at least 1.5 volts. If they read slightly higher than this, it's normal. A battery starts out above its rated voltage, and delivers progressively less as you use it. Batteries also lose some voltage while they are sitting on the shelf doing nothing.

Load your battery holder (taking care that the batteries are the right way around, with the negative ends pressing against the springs in the carrier). Use your meter to check the voltage on the wires coming out of the battery carrier. You should have at least 6 volts.

Now select a 2KΩ resistor. Remember, "2KΩ" means "2,000 ohms." If it has colored stripes, they should be red-black-red, meaning 2-0 and two more zeros. Because 2.2K resistors are more common than 2K resistors, you can substitute one of them if necessary. It will be colored red-red-red.

Wire it into the circuit as shown in Figures 1-44 and 1-45, making the connections with alligator clips. You should see the LED glow very dimly.

Now swap out your 2K resistor and substitute a 1K resistor, which will have brown-black-red stripes, meaning 1-0 and two more zeros. The LED should glow more brightly.

Swap out the 1K resistor and substitute a 470Ω resistor, which will have yellow-violet-brown stripes, meaning 4-7 and one more zero. The LED should be brighter still.

This may seem very elementary, but it makes an important point. The resistor blocks a percentage of the voltage in the circuit. Think of it as being like a kink or constriction in a flexible hose. A higher-value resistor blocks more voltage, leaving less for the LED.

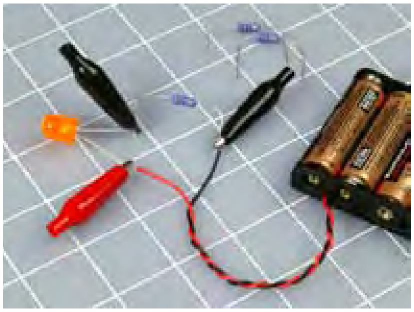

Figure 1-44. *The setup for Experiment 3, showing resistors of 470Ω, 1KΩ, and 2KΩ. Apply alligator clips where shown, to make a secure contact, and try each of the resistors one at a time at the same point in the circuit, while watching the LED.*

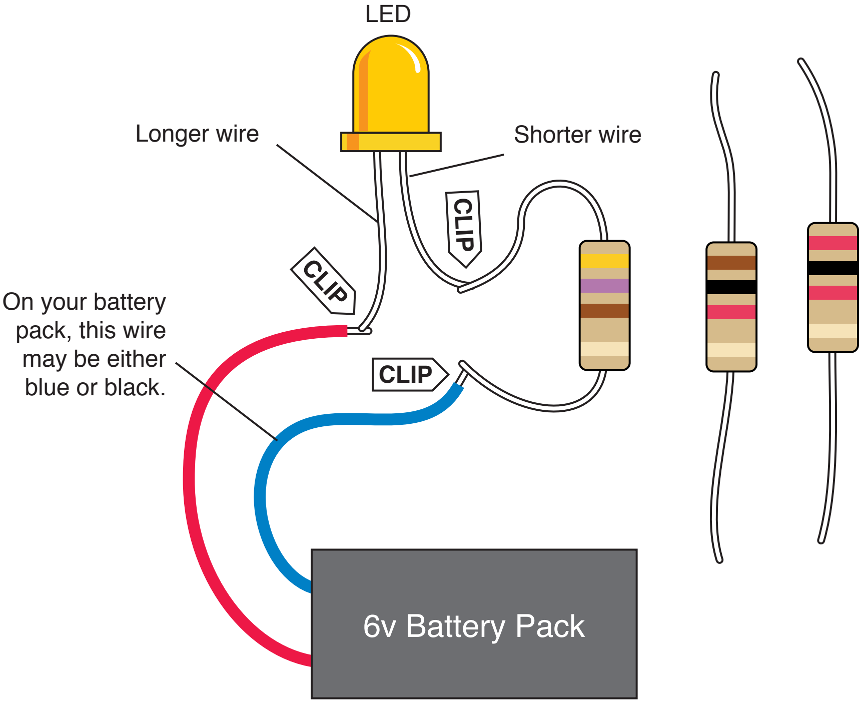
LED
Longer wire
Shorter wire
CLIP
CLIP
CLIP
On your battery pack, this wire may be either blue or black.
6v Battery Pack

Figure 1-45. *Here's how it actually looks, using a large LED. If you start with the highest value resistor, the LED will glow very dimly as you complete the circuit. The resistor drops most of the voltage, leaving the LED with insufficient current to make it shine brightly.*

Cleanup and Recycling

We'll use the batteries and the LED in the next experiment. The resistors can be reused in the future.

Experiment 4: Varying the Voltage

Potentiometers come in various shapes and sizes, but they all do the same thing: they allow you to vary voltage and current by varying resistance. This experiment will enable you to learn more about voltage, amperage, and the relationship between them. You'll also learn how to read a manufacturer's data sheet.

You will need the same batteries, battery carrier, alligator clips, and LED from the last experiment, plus:

- Potentiometer, 2KΩ linear. Quantity: 2. (See Figure 1-46.) Full-sized potentiometers that look like this are becoming less common, as miniature versions are taking their place. I'd like you to use a large one, though, because it's so much easier to work with.
- One extra LED.
- Multimeter.

Look Inside Your Potentiometer

The first thing I want you to do is find out how a potentiometer works. This means you'll have to open it, which is why your shopping list required you to buy two of them, in case you can't put the first one back together again.

Most potentiometers are held together with little metal tabs. You should be able to grab hold of the tabs with your wire cutters or pliers, and bend them up and outward. If you do this, the potentiometer should open up as shown in Figures 1-47 and 1-48.

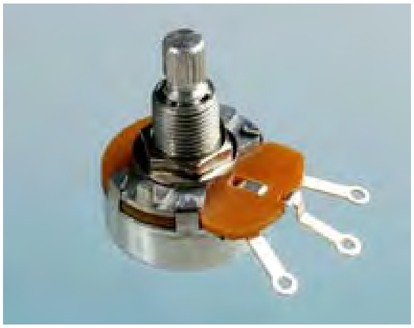

Figure 1-46

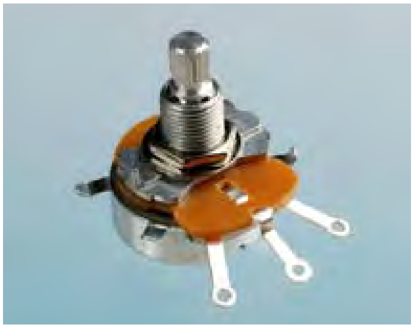

Figure 1-47

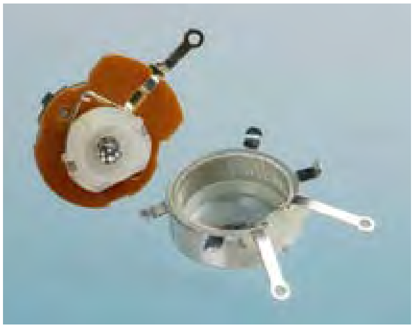

Figure 1-48. *To open the potentiometer, first pry up the four little metal tabs around the edge (you can see one sticking out at the left and another one sticking out at the right in Figure 1-47). Inside is a coil of wire around a flat plastic band, and a pair of springy contacts (the wiper), which conduct electricity to or from any point in the coil when you turn the shaft.*

Depending whether you have a really cheap potentiometer or a slightly more high-class version, you may find a circular track of conductive plastic or a loop of coiled wire. Either way, the principle is the same. The wire or the plastic possesses some resistance (a total of 2K in this instance), and as you turn the shaft of the potentiometer, a wiper rubs against the resistance, giving you a shortcut to any point from the center terminal.

You can try to put it back together, but if it doesn't work, use your backup potentiometer instead.

To test your potentiometer, set your meter to measure resistance (ohms) and touch the probes while turning the potentiometer shaft to and fro, as shown in Figure 1-49.

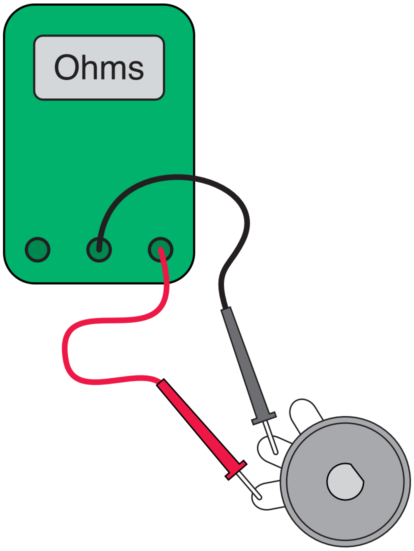

Figure 1-49. *Measure the resistance between these two terminals of the potentiometer while you turn its shaft to and fro.*

Dimming Your LED

Begin with the potentiometer turned all the way counterclockwise, otherwise you'll burn out the LED before we even get started. (A very, very small number of potentiometers increase and decrease resistance in the opposite way to which I'm describing here, but as long as your potentiometer looks like the one in Figure 1-48 after you open it up, my description should be accurate.)

Now connect everything as shown in Figures 1-50 and 1-51, taking care that you don't allow the metal parts of any of the alligator clips to touch each other. Now turn up the potentiometer *very* slowly. You'll notice the LED glowing brighter, and brighter, and brighter—until, oops, it goes dark. You see how easy it is to destroy modern electronics? Throw away that LED. It will never glow again. Substitute a new LED, and we'll be more careful this time.

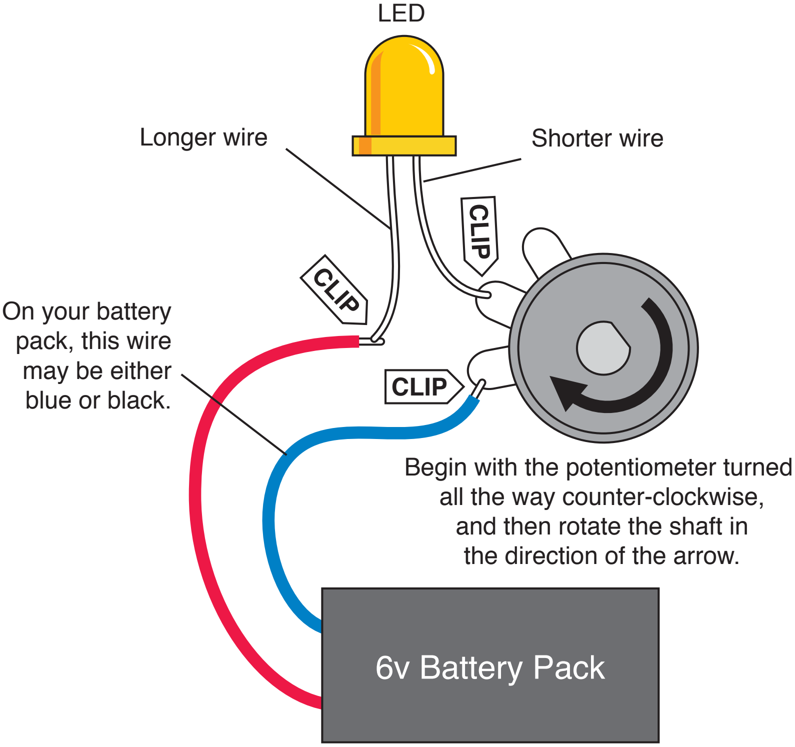
LED
Longer wire
Shorter wire
CLIP
CLIP
CLIP
On your battery pack, this wire may be either blue or black.
Begin with the potentiometer turned all the way counter-clockwise, and then rotate the shaft in the direction of the arrow.
6v Battery Pack

Figure 1-50. *The setup for Experiment 4. Rotating the shaft of the 2K potentiometer varies its resistance from 0 to 2,000Ω. This resistance protects the LED from the full 6 volts of the battery.*

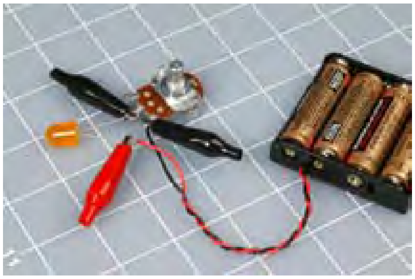

Figure 1-51. *The LED in this photo is dark because I turned the potentiometer up just a little bit too far.*

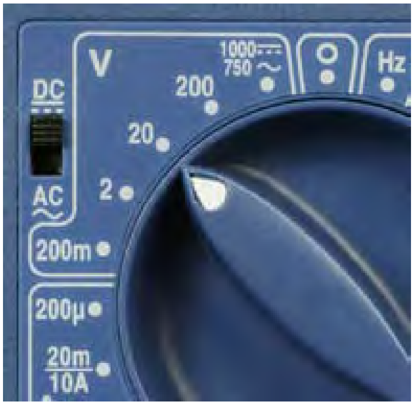
DC
AC
V
1000
750
200
20
2
200m
200μ
20m
10A
Hz

Figure 1-52

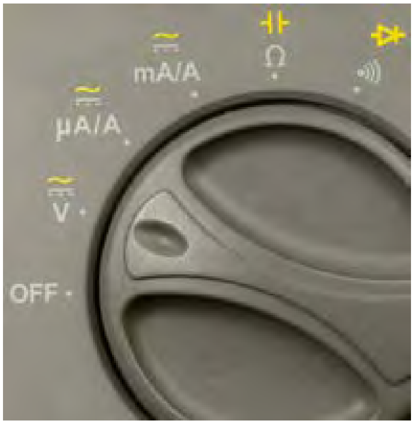
mA/A
Ω
μA/A
V
OFF

Figure 1-53

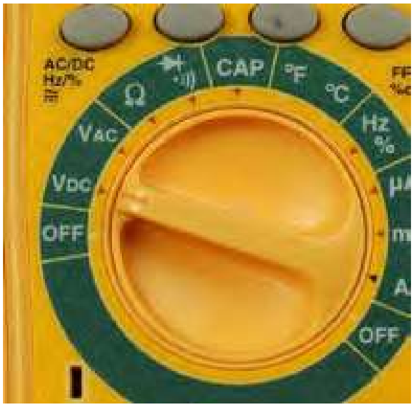
AC/DC
Hz/%
Ω
CAP
°F
°C
Hz
%
VAC
VDC
OFF
OFF

Figure 1-54. *Each meter has a different way to measure volts DC. The manually adjusted meter (top) requires you to move a slider switch to "DC" and then choose the highest voltage you want to measure: In this case, the selected voltage is 20 (because 2 would be too low). Using the autoranging RadioShack meter, you set it to "V" and the meter will figure out which range to use.*

While the batteries are connected to the circuit, set your meter to measure *volts DC* as shown in Figures 1-52 through 1-54. Now touch the probes either side of the LED. Try to hold the probes in place while you turn the potentiometer up a little, and down a little. You should see the voltage pressure around the LED changing accordingly. We call this the *potential difference* between the two wires of the LED.

If you were using a miniature old-fashioned lightbulb instead of an LED, you'd see the potential difference varying much more, because a lightbulb behaves like a "pure" resistor, whereas an LED self-adjusts to some extent, modifying its resistance as the voltage pressure changes.

Now touch the probes to the two terminals of the potentiometer that we're using, so that you can measure the potential difference between them. The potentiometer and the LED share the total available voltage, so when the potential difference (the voltage drop) around the potentiometer goes up, the potential difference around the LED goes down, and vice versa. See Figures 1-55 through 1-57. A few things to keep in mind:

- If you add the voltage drops across the devices in the circuit, the total is the same as the voltage supplied by the batteries.
- You measure voltage relatively, between two points in a circuit.
- Apply your meter like a stethoscope, without disturbing or breaking the connections in the circuit.

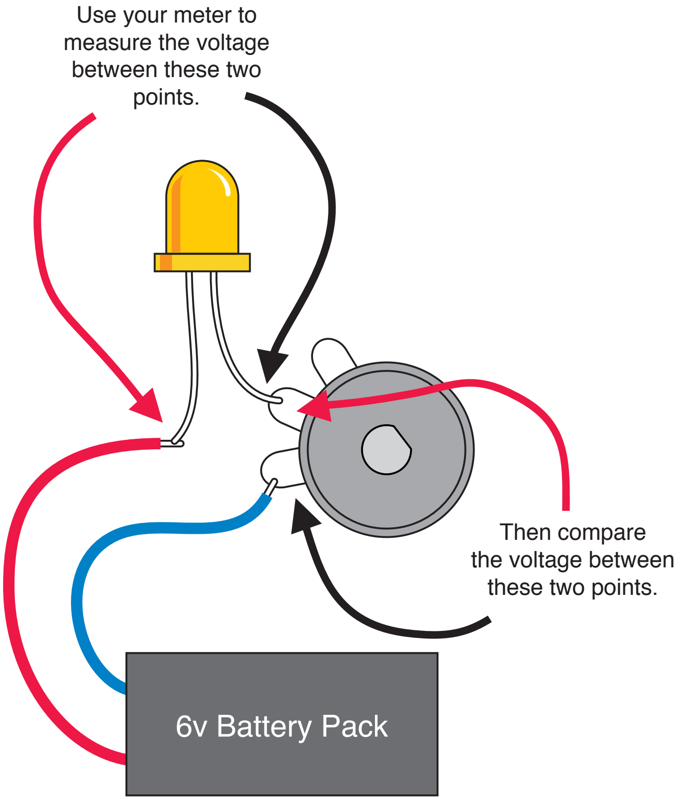
Use your meter to measure the voltage between these two points.
Then compare the voltage between these two points.
6v Battery Pack

Figure 1-55. *How to measure voltage in a simple circuit.*

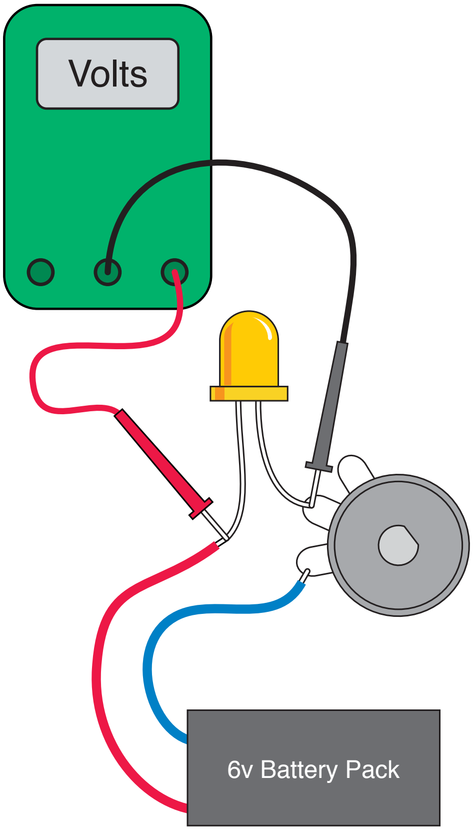

Figure 1-56. *The meter shows how much voltage the LED takes.*

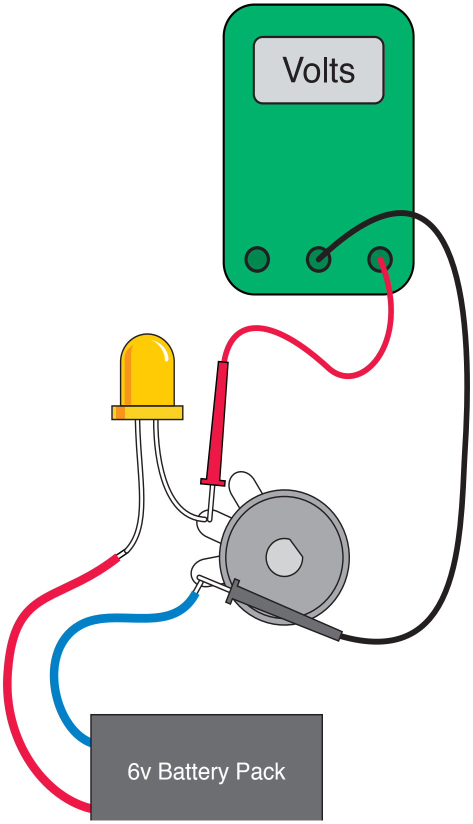

Figure 1-57. *The meter shows how much voltage the potentiometer takes.*

Checking the Flow

Now I want you to make a different measurement. I want you to measure the flow, or current, in the circuit, using your meter set to mA (milliamps). Remember, to measure current:

- You can only measure current when it passes *through* the meter.
- You have to insert your meter into the circuit.
- Too much current will blow the fuse inside your meter.

Make sure you set your meter to measure mA, not volts, before you try this. Some meters require you to move one of your leads to a different socket on the meter, to measure mA. See Figures 1-58 through 1-61.

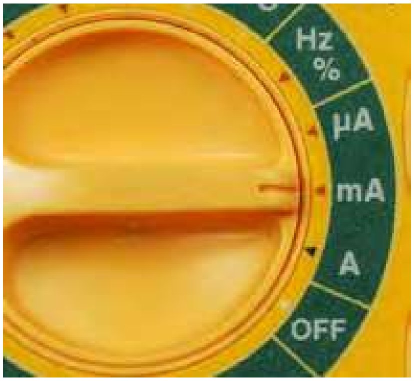

Figure 1-58. *Any meter will blow its internal fuse if you try to make it measure too high an amperage. In our circuit, this is not a risk as long as you keep the potentiometer in the middle of its range. Choose "mA" for milliamps and remember that the meter displays numbers that mean thousandths of an amp.*

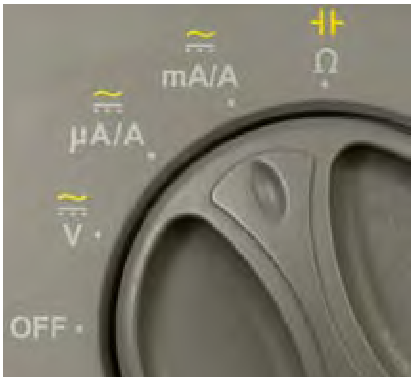

Figure 1-59

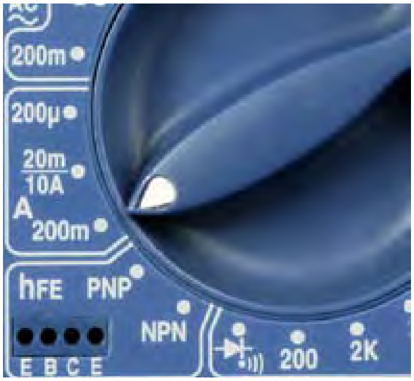

Figure 1-60

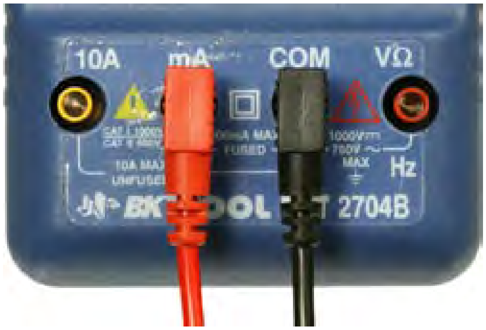

Figure 1-61. *A manual meter such as the one here may require you to shift the red lead to a different socket, to measure milliamps. Most modern meters don't require this until you are measuring higher currents.*

Insert your meter into the circuit, as shown in Figure 1-62. Don't turn the potentiometer more than halfway up. The resistance in the potentiometer will protect your meter, as well as the LED. If the meter gets too much current, you'll find yourself replacing its internal fuse.

As you adjust the potentiometer up and down a little, you should find that the varying resistance in the circuit changes the flow of current—the amperage. This is why the LED burned out in the previous experiment: too much current made it hot, and the heat melts it inside, just like the fuse in the previous experiment. *A higher resistance limits the flow of current, or amperage*.

Now insert the meter in another part of the circuit, as shown in Figure 1-63. As you turn the potentiometer up and down, you should get exactly the same results as with the configuration in Figure 1-64. This is because the current is the same at all points in a similar circuit. *It has to be, because the flow of electrons has no place else to go*.

It's time now to nail this down with some numbers. Here's one last thing to try. Set aside the LED and substitute a 1KΩ resistor, as shown in Figure 1-64. The total resistance in the circuit is now 1KΩ plus whatever the resistance the potentiometer provides, depending how you set it. (The meter also has some resistance, but it's so low, we can ignore it.)

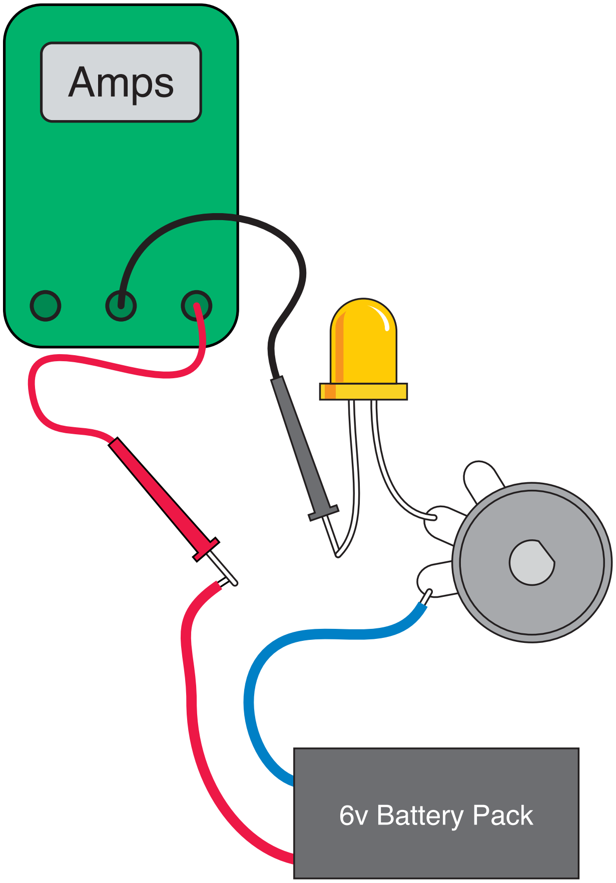

Figure 1-62. *To measure amps, as illustrated here and in Figure 1-63, the current has to pass through the meter. When you increase the resistance, you restrict the current flow, and the lower flow makes the LED glow less brightly.*

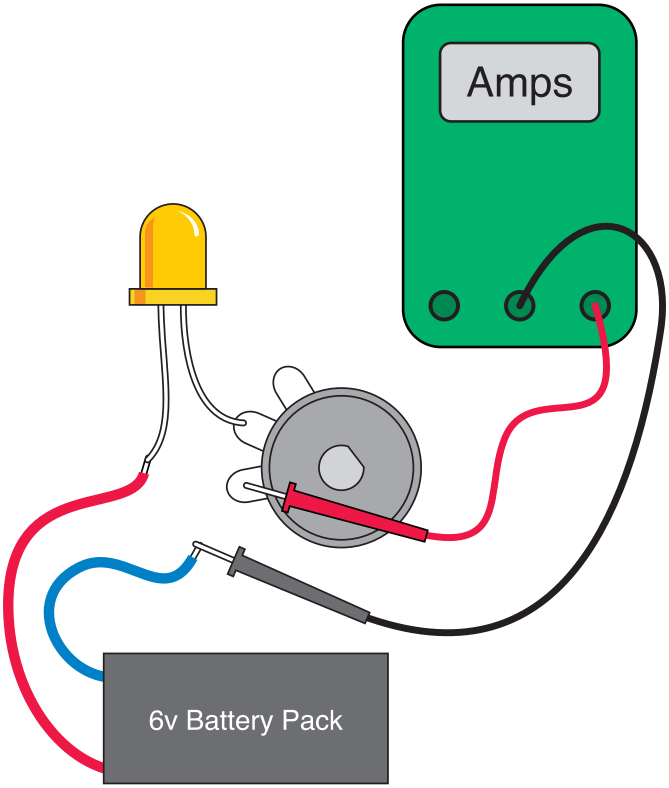

Figure 1-63

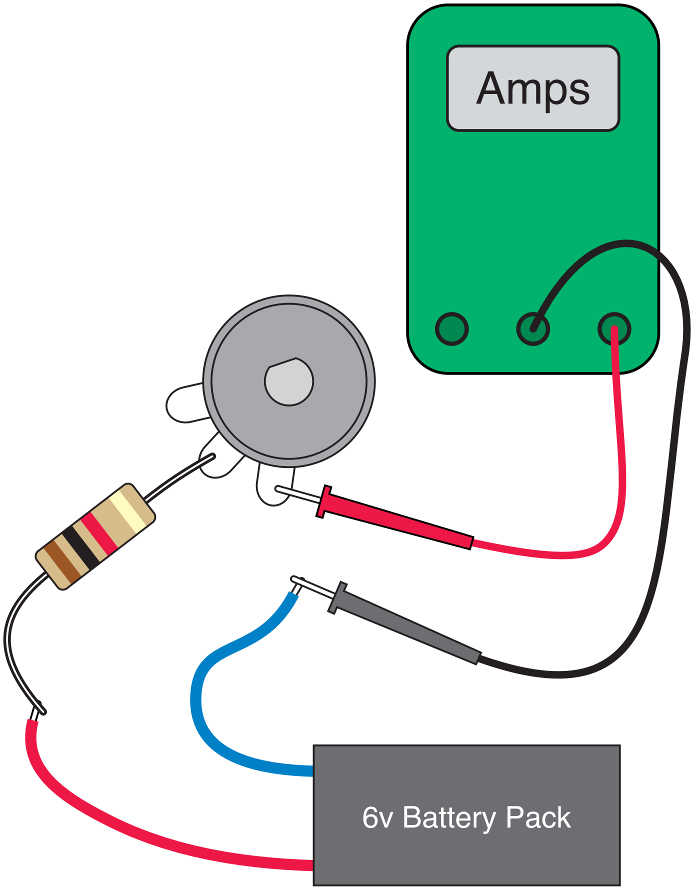

Figure 1-64. *If you substitute a resistor instead of the LED, you can confirm that the current flowing through the circuit varies with the total resistance in the circuit, if the voltage stays the same.*

Turn the potentiometer all the way counterclockwise, and you have a total of 3K resistance in the circuit. Your meter should show about 2 mA flowing. Now turn the potentiometer halfway, and you have about 2K total resistance. You should see about 3 mA flowing. Turn the potentiometer all the way clockwise, so there's a total of 1K, and you should see 6 mA flowing. You may notice that if we multiply the resistance by the amperage, we get 6 each time—which just happens to be the voltage being applied to the circuit. See the following table.

Total resistance	**Current**	**Voltage**
(KΩ)	(mA)	(Volts)
3	2	6
2	3	6
1	6	6

In fact, we could say:

voltage = kilohms × milliamps

But wait a minute: 1K is 1,000 ohms, and 1mA is 1/1,000 of an amp. Therefore, our formula should really look like this:

voltage = (ohms × 1,000) × (amps/1,000)

The two factors of 1,000 cancel out, so we get this:

volts = ohms × amps

This is known as Ohm's Law. See the section, "Fundamentals: Ohm's Law," on the following page.

FUNDAMENTALS

Series and parallel

Before we go any further, you should know how resistance in a circuit increases when you put resistors in series or in parallel. Figures 1-65 through 1-67 illustrate this. Remember:

- Resistors in series are oriented so that one follows the other.
- Resistors in parallel are oriented side by side.

When you put two equal-valued resistors in series, you double the total resistance, because electricity has to pass through two barriers in succession.

When you put two equal-valued resistors in parallel, you divide the total resistance by two, because you're giving the electricity two paths which it can take, instead of one.

In reality we don't normally need to put resistors in parallel, but we often put other types of components in parallel. Lightbulbs in your house, for instance, are all wired that way. So, it's useful to understand that resistance in a circuit goes down if you keep adding components in parallel.

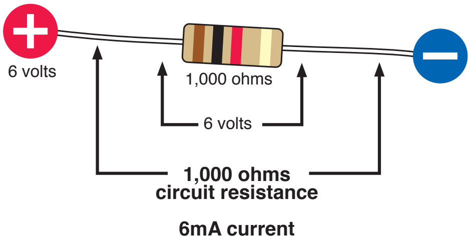
6 volts
1,000 ohms
6 volts
1,000 ohms
circuit resistance
6mA current

Figure 1-65. *One resistor takes the entire voltage, and according to Ohm's Law, it draws v/R = 6/1,000 = 0.006 amps = 6mA of current.*

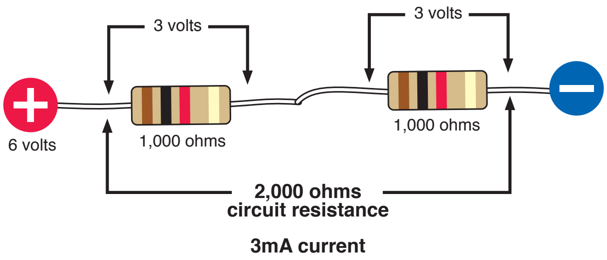
3 volts
3 volts
6 volts
1,000 ohms
1,000 ohms
2,000 ohms
circuit resistance
3mA current

Figure 1-66. *When two resistors are in series, the electricity has to pass through one to reach the other, and therefore each of them takes half the voltage. Total resistance is now 2,000 ohms, and according to Ohm's Law, the circuit draws v/R = 6/2,000 = 0.003 amps = 3mA of current.*

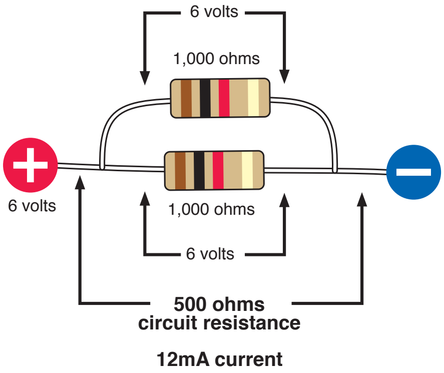
6 volts
1,000 ohms
6 volts
1,000 ohms
6 volts
500 ohms
circuit resistance
12mA current

Figure 1-67. *When two resistors are in parallel, each is exposed to the full voltage, so each of them takes 6 volts. The electricity can now flow through both at once, so the total resistance of the circuit is half as much as before. According to Ohm's Law, the circuit draws v/R = 6/500 = 0.012 amps = 12mA of current.*

FUNDAMENTALS

Ohm's Law

For reasons I'll explain in a moment, amps are normally abbreviated with the letter I. V stands for volts and R stands for resistance in ohms (because the omega symbol, Ω, is not easily generated from most keyboards). Using these symbols, you can write Ohm's Law in three different ways:

$V = I \times R$

$I = V/R$

$R = V/I$

Remember, V is a *difference* in voltage between two points in a simple circuit, R is the resistance in ohms *between* the same two points, and I is the current in amps flowing *through* the circuit between the two points.

Letter I is used because originally current was measured by its *inductance*, meaning the ability to induce magnetic effects. It would be much less confusing to use A for amps, but unfortunately it's too late for that to happen.

Using Ohm's Law

Ohm's Law is extremely useful. For example, it helps us to figure out whether a component can be used safely in a circuit. Instead of stressing the component until we burn it out, we can predict whether it will work.

For instance, the first time you turned the potentiometer, you didn't really know how far you could go until the LED burned out. Wouldn't it be useful to know precisely what resistance to put in series with an LED, to protect it adequately while providing as much light as possible?

How to Read a Data Sheet

Like most information, the answer to this question is available online.

Here's how you find a manufacturer's data sheet (Figure 1-68). First, find the component that you're interested in from a mail-order source. Next, Google the part number and manufacturer's name. Usually the data sheet will pop up as the first hit. A source such as *Mouser.com* makes it even easier by giving you a direct link to manufacturers' data sheets for many products.

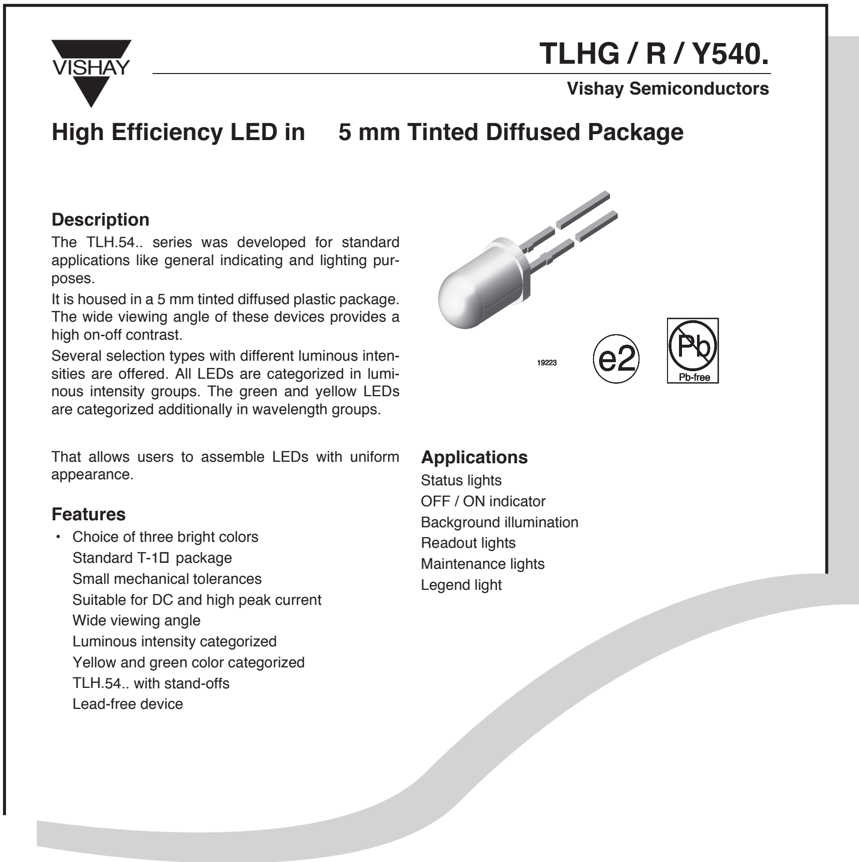

VISHAY

TLHG / R / Y540.

Vishay Semiconductors

High Efficiency LED in 5 mm Tinted Diffused Package

Description

The TLH.54.. series was developed for standard applications like general indicating and lighting purposes.

It is housed in a 5 mm tinted diffused plastic package. The wide viewing angle of these devices provides a high on-off contrast.

Several selection types with different luminous intensities are offered. All LEDs are categorized in luminous intensity groups. The green and yellow LEDs are categorized additionally in wavelength groups.

That allows users to assemble LEDs with uniform appearance.

Features

- Choice of three bright colors
- Standard T-1□ package
- Small mechanical tolerances
- Suitable for DC and high peak current
- Wide viewing angle
- Luminous intensity categorized
- Yellow and green color categorized
- TLH.54.. with stand-offs
- Lead-free device

Applications

Status lights
OFF / ON indicator
Background illumination
Readout lights
Maintenance lights
Legend light

Figure 1-68. *The beginning of a typical data sheet, which includes all relevant specifications for the product, freely available online.*

BACKGROUND

How much voltage does a wire consume?

Normally, we can ignore the resistance in electric wires, such as the little leads of wire that stick out of resistors, because it's trivial. However, if you try to force large amounts of current through long lengths of thin wire, the resistance of the wire can become important.

How important? Once again, we can use Ohm's Law to find out.

Suppose that a very long piece of wire has a resistance of 0.2Ω. And we want to run 15 amps through it. How much voltage will the wire steal from the circuit, because of its resistance?

Once again, you begin by writing down what you know:

$R = 0.2$

$I = 15$

We want to know V, the potential difference, for the wire, so we use the version of Ohm's Law that places V on the left side:

$V = I \times R$

Now plug in the values:

$V = 15 \times 0.2 = 3$ volts

Three volts is not a big deal if you have a high-voltage power supply, but if you are using a 12-volt car battery, this length of wire will take one-quarter of the available voltage.

Now you know why the wiring in automobiles is relatively thick—to reduce its resistance well below 0.2Ω. See Figure 1-69.

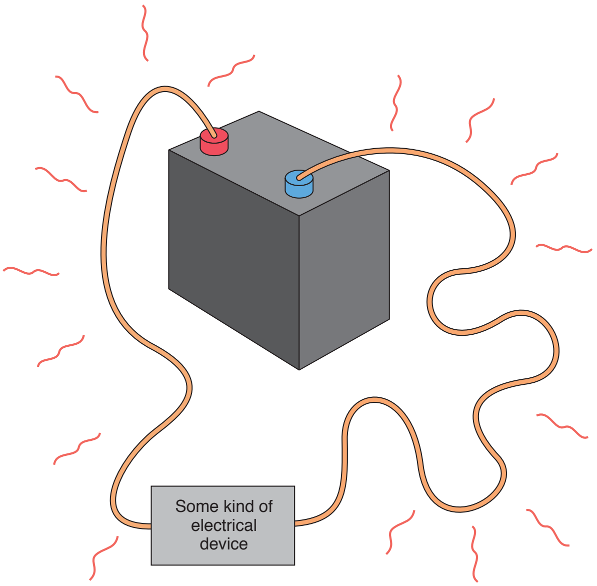
Some kind of electrical device

Figure 1-69. *When a 12-volt car battery runs some kind of electrical device through a long piece of thin wire, the resistance of the wire steals some of the voltage and dissipates it as heat.*

BACKGROUND

The origins of wattage

James Watt (Figure 1-70) is known as the inventor of the steam engine. Born in 1736 in Scotland, he set up a small workshop in the University of Glasgow, where he struggled to perfect an efficient design for using steam to move a piston in a cylinder. Financial problems and the primitive state of the art of metal working delayed practical applications until 1776.

Despite difficulties in obtaining patents (which could only be granted by an act of parliament in those times), Watt and his business partner eventually made a lot of money from his innovations. Although he predated the pioneers in electricity, in 1889 (70 years after his death), his name was assigned to the basic unit of electric power that can be defined by multiplying amperes by volts. See the Fundamentals section, "Watt Basics," on page 31.

Figure 1-70. *James Watt's development of steam power enabled the industrial revolution. After his death, he was honored by having his name applied to the basic unit of power in electricity.*

Here's an example. Suppose I want a red LED, such as the Vishay part TLHR5400, which has become such a common item that I can buy them individually for 9 cents apiece. I click the link to the data sheet maintained by the manufacturer, Vishay Semiconductor. Almost immediately I have a PDF page on my screen. This data sheet is for TLHR, TLHG, and TLHY types of LED, which are red, green, and yellow respectively, as suggested by the R, G, and Y in the product codes. I scroll down and look at the "Optical and Electrical Characteristics" section. It tells me that under conditions of drawing a current of 20 mA, the LED will enjoy a "Typ," meaning, typical, "forward voltage" of 2 volts. The "Max," meaning maximum, is 3 volts.

Let's look at one other data sheet, as not all of them are written the same way. I'll choose a different LED, the Kingbright part WP7113SGC. Click on the link to the manufacturer's site, and I find on the second page of the data sheet a typical forward voltage of 2.2, maximum 2.5, and a maximum forward current of 25 mA. I also find some additional information: a maximum reverse voltage of 5 and maximum reverse current of 10 uA (that's microamps, which are 1,000 times smaller than milliamps). This tells us that you should avoid applying excessive voltage to the LED the wrong way around. If you exceed the reverse voltage, you risk burning out the LED. Always observe polarity!

Kingbright also warns us how much heat the LED can stand: 260° C (500° F) for a few seconds. This is useful information, as we'll be putting aside our alligator clips and using hot molten solder to connect electrical parts in the near future. Because we have already destroyed a battery, a fuse, and an LED in just four experiments, maybe you won't be surprised when I tell you that we will destroy at least a couple more components as we test their limits with a soldering iron.

Anyway, now we know what an LED wants, we can figure out how to supply it. If you have any difficulties dealing with decimals, check the Fundamentals section "Decimals," on the next page, before continuing.

How Big a Resistor Does an LED Need?

Suppose that we're use the Vishay LED. Remember its requirements from the data sheet? Maximum of 3 volts, and a safe current of 20mA.

I'm going to limit it to 2.5 volts, to be on the safe side. We have 6 volts of battery power. Subtract 2.5 from 6 and we get 3.5. So we need a resistor that will take 3.5 volts from the circuit, leaving 2.5 for the LED.

The current flow is the same at all places in a simple circuit. If we want a maximum of 20mA to flow through the LED, the same amount of current will be flowing through the resistor.

Now we can write down what we know about the resistor in the circuit. Note that we have to convert all units to volts, amps, and ohms, so that 20mA should be written as 0.02 amps:

V = 3.5 (the potential drop across the resistor)
I = 0.02 (the current flowing through the resistor)

We want to know R, the resistance. So, we use the version of Ohm's Law that puts R on the left side:

$$R = V/I$$

Now plug in the values:

$$R = 3.5/0.02$$

Run this through your pocket calculator if you find decimals confusing. The answer is:

$$R = 175\Omega$$

It so happens that 175Ω isn't a standard value. You may have to settle for 180 or 220Ω, but that's close enough.

Evidently the 470Ω resistor that you used in Experiment 3 was a very conservative choice. I suggested it because I said originally that you could use any LED at all. I figured that no matter which one you picked, it should be safe with 470Ω to protect it.

Cleanup and Recycling

The dead LED can be thrown away. Everything else is reusable.

FUNDAMENTALS

Decimals

Legendary British politician Sir Winston Churchill is famous for complaining about "those damned dots." He was referring to decimal points. Because Churchill was Chancellor of the Exchequer at the time, and thus in charge of all government expenditures, his difficulty with decimals was a bit of a problem. Still, he muddled through in time-honored British fashion, and so can you.

You can also use a pocket calculator—or follow two basic rules.

Doing multiplication: combine the zeros

Suppose you need to multiply 0.04 by 0.005:

1. Count the total number of zeros following both of the decimal points. In this case, three zeros.
2. Multiply the numbers which follow the zeros. In this case, $4 \times 5 = 20$.
3. Write down the result as 0 followed by a decimal point, followed by the number of zeros, followed by the multiplication result. Like this: 0.00020, which is the same as 0.0002.

Doing division: cancel the zeros

Suppose you need to divide 0.006 by 0.0002:

1. Shift the decimal points to the right, in both the numbers, by the same number of steps, until both the numbers are greater than 1. In this case, shift the point four steps in each number, so you get 60 divided by 2.
2. Do the division. The result in this case is 30.

THEORY

Doing the math on your tongue

I'm going to go back to the question I asked in the previous experiment: why didn't your tongue get hot?

Now that you know Ohm's Law, you can figure out the answer in numbers. Let's suppose the battery delivered its rated 9 volts, and your tongue had a resistance of 50K, which is 50,000 ohms. Write down what you know:

$V = 9$

$R = 50{,}000$

We want to know the current, I, so we use the version of Ohm's Law that puts this on the left:

$I = V/R$

Plug in the numbers:

$I = 9/50{,}000 = 0.00018$ amps

Move the decimal point three places to convert to milliamps:

$I = 0.18$ mA

That's a tiny current that will not produce much heat at 9 volts.

What about when you shorted out the battery? How much current made the wires get hot? Well, suppose the wires had a resistance of 0.1 ohms (probably it's less, but I'll start with 0.1 as a guess). Write down what we know:

$V = 1.5$

$R = 0.1$

Once again we're trying to find I, the current, so we use:

$I = V/R$

Plug in the numbers:

$I = 1.5/0.1 = 15$ amps

That's 100,000 times the current that may have passed through your tongue, which would have generated much more heat, even though the voltage was lower.

Could that tiny little battery really pump out 15 amps? Remember that the battery got hot, as well as the wire. This tells us that the electrons may have met some resistance inside the battery, as well as in the wire. (Otherwise, where else did the heat come from?) Normally we can forget about the internal resistance of a battery, because it's so low. But at high currents, it becomes a factor.

I was reluctant to short-circuit the battery through a meter, to try to measure the current. My meter will fry if the current is greater than 10A. However I did try putting other fuses into the circuit, to see whether they would blow. When I tried a 10A fuse, it did not melt. Therefore, for the brand of battery I used, I'm fairly sure that the current in the short circuit was under 10A, but I know it was over 3A, because the 3A fuse blew right away.

The internal resistance of the 1.5-volt battery prevented the current in the short circuit from getting too high. This is why I cautioned against using a larger battery (especially a car battery). Larger batteries have a much lower internal resistance, allowing dangerously high currents which generate explosive amounts of heat. A car battery is designed to deliver literally hundreds of amps when it turns a starter motor. That's quite enough current to melt wires and cause nasty burns. In fact, you can weld metal using a car battery.

Lithium batteries also have low internal resistance, making them very dangerous when they're shorted out. High current can be just as dangerous as high voltage.

FUNDAMENTALS

Watt basics

So far I haven't mentioned a unit that everyone is familiar with: watts.

A watt is a unit of work. Engineers have their own definition of work—they say that work is done when a person, an animal, or a machine pushes something to overcome mechanical resistance. Examples would be a steam engine pulling a train on a level track (overcoming friction and air resistance) or a person walking upstairs (overcoming the force of gravity).

When electrons push their way through a circuit, they are overcoming a kind of resistance, and so they are doing work, which can be measured in watts. The definition is easy:

watts = volts × amps

Or, using the symbols customarily assigned, these three formulas all mean the same thing:

$W = V \times I$

$V = W/I$

$I = W/V$

Watts can be preceded with an "m," for "milli," just like volts:

Number of watts	Usually expressed as	Abbreviated as
0.001 watts	1 milliwatt	1mW
0.01 watts	10 milliwatts	10 mW
0.1 watts	100 milliwatts	100 mW
1 watt	1,000 milliwatts	1W

Because power stations, solar installations, and wind farms deal with much larger numbers, you may also see references to kilowatts (using letter K) and megawatts (with a capital M, not to be confused with the lowercase m used to define milliwatts):

Number of watts	Usually expressed as	Abbreviated as
1,000 watts	1 kilowatt	1 KW
1,000,000 watts	1 megawatt	1 MW

Lightbulbs are calibrated in watts. So are stereo systems. The watt is named after James Watt, inventor of the steam engine. Incidentally, watts can be converted to horsepower, and vice versa.

THEORY

Power assessments

I mentioned earlier than resistors are commonly rated as being capable of dealing with 1/4 watt, 1/2 watt, 1 watt, and so on. I suggested that you should buy resistors of 1/4 watt or higher. How did I know this?

Go back to the LED circuit. Remember we wanted the resistor to drop the voltage by 3.5 volts, at a current of 20 mA. How many watts of power would this impose on the resistor?

Write down what you know:

$V = 3.5$ (the voltage drop imposed by the resistor)

$I = 20mA = 0.02$ amps (the current flowing through the resistor)

We want to know W, so we use this version of the formula:

$W = V \times I$

Plug in the values:

$W = 3.5 \times 0.02 = 0.07$ watts (the power being dissipated by the resistor)

Because 1/4 watt is 0.25 watts, obviously a 1/4 watt resistor will have about four times the necessary capacity. In fact you could have used a 1/8 watt resistor, but in future experiments we may need resistors that can handle 1/4 watt, and there's no penalty for using a resistor that is rated for more watts than will actually pass through it.

Experiment 5: Let's Make a Battery

Long ago, before web surfing, file sharing, or cell phones, kids were so horribly deprived that they tried to amuse themselves with kitchen-table experiments such as making a primitive battery by pushing a nail and a penny into a lemon. Hard to believe, perhaps, but true!

This is seriously old-school—but I want you to try it anyway, because anyone who wants to get a feel for electricity should see how easy it is to extract it from everyday objects around us. Plus, if you use enough lemons, you just *might* generate enough voltage to power an LED.

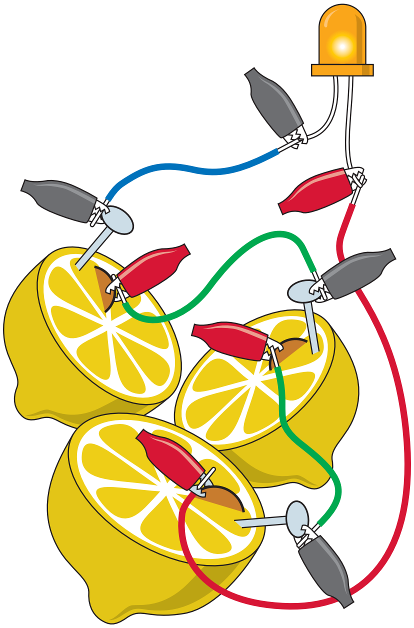

Figure 1-71. *A three-lemon battery. Don't be too disappointed if the LED fails to light up. The lemons have a high electrical resistance, so they can't deliver much current, especially through the relatively small surface area of the nails and the pennies. However, the lemon battery does generate voltage that you can measure with your meter.*

The basic components of a battery are two metal electrodes immersed in an electrolyte. I won't define these terms here (they're explained in the following section "Theory: The nature of electricity"). Right now all you need to know is that lemon juice will be your electrolyte, and copper and zinc will be your electrodes. A penny provides the necessary copper, as long as it is fairly new and shiny. Pennies aren't solid copper anymore, but they are still copper-plated, which is good enough.

To find some metallic zinc, you will have to make a trip to a hardware store, where you should ask for roofing nails. The nails are zinc-plated to prevent them from rusting. Small metal brackets or mending plates also are usually zinc-plated. They should have a slightly dull, silvery look. If they have a mirror-bright finish, they're more likely to be nickel-plated.

Cut a lemon in half, set your multimeter so that it can measure up to 2 volts DC, and hold one probe against a penny while you hold the other probe against a roofing nail (or other zinc-plated object). Now force the penny and the nail into the exposed juicy interior of the lemon, as close to each other as possible, but not actually touching. You should find that your meter detects between 0.8 volts and 1 volt.

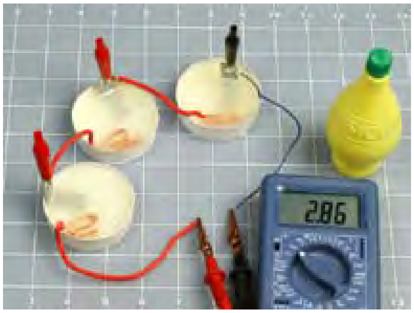

Figure 1-72. *Bottled lemon juice seems to work just as well as fresh lemon juice. I cut the bottoms off three paper cups, inserted a galvanized bracket into each, and used heavyweight stranded copper wire to make the positive electrodes*

You can experiment with different items and liquids to see which works best. Immersing your nail and penny in lemon juice that you have squeezed into a shot glass or egg cup may enhance the efficiency of your battery, although you'll have a harder time holding everything in place. Grapefruit juice and vinegar will work as substitutes for lemon juice.

To drive a typical LED, you need more than 1 volt. How to generate the extra electrical pressure? By putting batteries in series, of course. In other words, more lemons! (Or more shot glasses or egg cups.) You'll also need lengths of wire to connect multiple electrodes, and this may entail skipping ahead to Chapter 2, where I describe how to strip insulation from hookup wire. Figures 1-71 and 1-72 show the configuration.

If you set things up carefully, making sure than none of the electrodes are touching, you may be able to illuminate your LED with two or three lemon-juice batteries in series. (Some LEDs are more sensitive to very low currents than others. Later in the book I'll be talking about very-low-current LEDs. If you want your lemon-juice battery to have the best chance of working, you can search online for low-current LEDs and buy a couple.)

THEORY

The nature of electricity

To understand electricity, you have to start with some basic information about atoms. Each atom consists of a nucleus at the center, containing protons, which have a positive charge. The nucleus is surrounded by electrons, which carry a negative charge.

Breaking up the nucleus of an atom requires a lot of energy, and can also liberate a lot of energy—as happens in a nuclear explosion. But persuading a couple of electrons to leave an atom (or join an atom) takes very little energy. For instance, when zinc reacts chemically with an acid, it can liberate electrons. This is what happens at the zinc electrode of the chemical battery in Experiment 5.

The reaction soon stops, as electrons accumulate on the zinc electrode. They feel a mutual force of repulsion, yet they have nowhere to go. You can imagine them like a crowd of hostile people, each one wanting the others to leave, and refusing to allow new ones to join them, as shown in Figure 1-73.

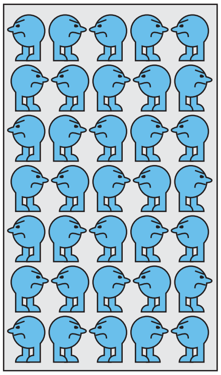

Figure 1-73. *Electrons on an electrode have a bad attitude known as mutual repulsion.*

Now consider what happens when a wire connects the zinc electrode, which has a surplus of electronics, to another electrode, made from a different material, that has a shortage of electrons. The electrons can pass through the wire very easily by jumping from one atom to the next, so they escape from the zinc electrode and run through the wire, propelled by their great desire to get away from each other. See Figure 1-74. This mutual force of propulsion is what creates an electrical current.

Now that the population of electrons on the zinc electrode has been reduced, the zinc-acid reaction can continue, replacing the missing electrodes with new ones—which promptly imitate their predecessors and try to get away from each other by running away down the wire. The process continues until the zinc-acid reaction grinds to a halt, usually because it creates a layer of a compound such as zinc oxide, which won't react with acid and prevents the acid from reacting with the zinc underneath. (This is why your zinc electrode may have looked sooty when you pulled it out of the acidic electrolyte.)

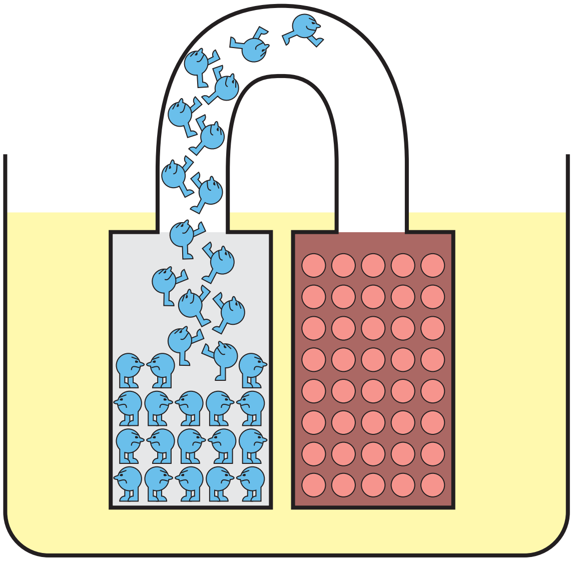

Figure 1-74. *As soon as we open up a pathway from a zinc electrode crowded with electrons to a copper electrode, which contains "holes" for the electrons, their mutual repulsion makes them try to escape from each other to their new home as quickly as possible.*

This description applies to a "primary battery," meaning one that is ready to generate electricity as soon as a connection between its terminals allows electrons to transfer from one electrode to the other. The amount of current that a primary battery can generate is determined by the speed at which chemical reactions inside the battery can liberate electrons. When the raw metal in the electrodes has all been used up in chemical reactions, the battery can't generate any more electricity and is dead. It cannot easily be recharged, because the chemical reactions are not easily reversible, and the electrodes may have oxidized.

In a rechargeable battery, also known as a secondary battery, a smarter choice of electrodes and electrolyte does allow the chemical reactions to be reversed.

How much current is being generated in your lemon battery? Set your meter to measure milliamps, and connect it between the nail and the penny. I measured about 2mA, but got 10mA when I used some #10 stranded copper wire instead of a penny and a large mending plate instead of a roofing nail, immersed in a cup of grapefruit juice. When a larger surface area of metal makes better contact with the electrolyte, you get a greater flow of current. (Don't ever connect your meter to measure amps directly between the terminals of a real battery. The current will be too high, and can blow the fuse inside your meter.)

What's the internal resistance of your lemon? Put aside the copper and zinc electrodes and insert your nickel-plated meter probes into the juice. I got a reading of around 30K when both probes were in the same segment of the lemon, but 40K or higher if the probes were in different segments. Is the resistance lower when you test liquid in a cup?

Here are a couple more questions that you may wish to investigate. For how long will your lemon battery generate electricity? And why do you think your zinc-plated electrode becomes discolored after it has been used for a while?

Electricity is generated in a battery by an exchange of ions, or free electrons, between metals. If you want to know more about this, check the section "Theory: The nature of electricity" on the previous page.

Cleanup and Recycling

The hardware that you immersed in lemons or lemon juice may be discolored, but it is reusable. Whether you eat the lemons is up to you.

BACKGROUND

Positive and negative

If electricity is a flow of electrons, which have a negative charge, why do people talk as if electricity flows from the positive terminal to the negative terminal of a battery?

The answer lies in a fundamental embarrassment in the history of research into electricity. For various reasons, when Benjamin Franklin was trying to understand the nature of electric current by studying phenomena such as lightning during thunderstorms, he believed he observed a flow of "electrical fluid" from positive to negative. He proposed this concept in 1747.

In fact, Franklin had made an unfortunate error that remained uncorrected until after physicist J. J. Thomson announced his discovery of the electron in 1897, 150 years later. Electricity actually flows from an area of greater negative charge, to some other location that is "less negative"—that is, "more positive." In other words, electricity is a flow of negatively charged particles. In a battery, they originate from the negative terminal and flow to the positive terminal.

You might think that when this fact was established, everyone should have discarded Franklin's idea of a flow from positive to negative. But when an electron moves through a wire, you can still think of an equal positive charge flowing in the opposite direction. When the electron leaves home, it takes a small negative charge with it; therefore, its home becomes a bit more positive. When the electron arrives at its destination, its negative charge makes the destination a bit less positive. This is pretty much what would happen if an imaginary positive particle traveled in the opposite direction. Moreover, all of the mathematics describing electrical behavior are still valid if you apply them to the imaginary flow of positive charges.

As a matter of tradition and convenience we still retain Ben Franklin's erroneous concept of flow from positive to negative, because it really makes no difference. In the symbols that represent components such as diodes and transistors, you will actually find arrows reminding you which way these components should be placed—and the arrows all point from positive to negative, even though that's not the way things really work at all! Ben Franklin would have been surprised to learn that although most lightning strikes occur when a negative charge in clouds discharges to neutralize a positive charge on the ground, some forms of lightning are actually a flow of electrons from the negatively charged surface of the earth, up to a positive charge in the clouds. That's right: someone who is "struck by lightning" may be hurt by *emitting* electrons rather than by receiving them, as shown in Figure 1-75.

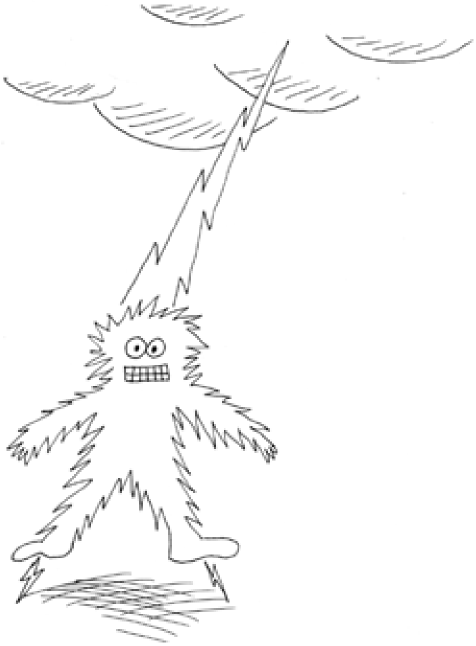

Figure 1-75. *In some weather conditions, the flow of electrons during a lightning strike can be from the ground, through your feet, out of the top of your head, and up to the clouds. Benjamin Franklin would have been surprised.*

THEORY

Basic measurements

Electrical potential is measured by adding up the charges on individual electrons. The basic unit is the *coulomb*, equal to the total charge on about 6,250,000,000,000,000,000 electrons.

If you know how many electrons pass through a piece of wire each second, this establishes the flow of electricity, which can be expressed in amperes. In fact 1 ampere can be defined as 1 coulomb per second. Thus:

1 ampere = 1 coulomb/second
= about 6.25 quintillion electrons/second

There's no way to "see" the number of electrons running through a conductor (Figure 1-76), but there are indirect ways of getting at this information. For instance, when an electron goes running through a wire, it creates a wave of electromagnetic force around it. This force can be measured, and we can calculate the amperage from that. The electric meter installed at your home by the utility company functions on this principle.

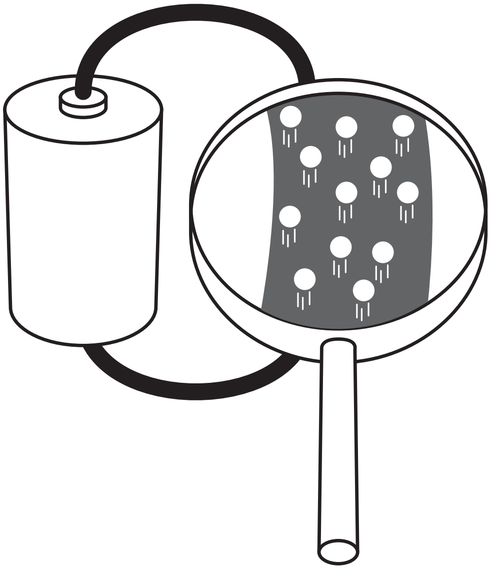

Figure 1-76. *If you could look inside an electric wire with a sufficiently powerful magnifying device, and the wire happened to be carrying 1 ampere of electron flow at the time, you might hope to see about 6.25 quintillion electrons speeding past each second.*

If electrons are just moving freely, they aren't doing any work. If you had a loop of wire of zero resistance, and you kick-started a flow of electrons somehow, they could just go buzzing around forever. (This is what happens inside a superconductor—almost.)

Under everyday conditions, even a copper wire has some resistance. The force that we need to push electrons through it is known as "voltage," and creates a flow that can create heat, as you saw when you shorted out a battery. (If the wire that you used had zero resistance, the electricity running through it would not have created any heat.) We can use the heat directly, as in an electric stove, or we can use the electrical energy in other ways—to run a motor, for instance. Either way, we are taking energy out of the electrons, to do some work.

One volt can be defined as the amount of pressure that you need to create a flow of 1 ampere, which does 1 watt of work. As previously defined, 1 watt = 1 volt × 1 ampere, but the definition actually originated the other way around:

1 volt = 1 watt/1 ampere

It's more meaningful this way, because a watt can be defined in nonelectrical terms. Just in case you're interested, we can work backward through the units of the metric system like this:

1 watt = 1 joule/second
1 joule = a force of 1 newton acting through 1 meter
1 newton = the force required to accelerate 1 kilogram by 1 meter per second, each second

On this basis, the electrical units can all be anchored with observations of mass, time, and the charge on electrons.

Practically Speaking

For practical purposes, an intuitive understanding of electricity can be more useful than the theory. Personally I like the water analogies that have been used for decades in guides to electricity. Figure 1-77 shows a tall tank half full of water, with a hole punched in it near the bottom. Think of the tank as being like a battery. The height of the water is comparable to voltage. The volume of flow through the hole, per second, is comparable to amperage. The smallness of the hole is comparable to resistance. See Figure 1-79 on the next page.

Where's the wattage in this picture? Suppose we place a little water wheel where it is hit by the flow from the hole. We can attach some machinery to the water wheel. Now the flow is doing some work. (Remember, wattage is a measurement of work.)

Maybe this looks as if we're getting something for nothing, extracting work from the water wheel without putting any energy back into the system. But remember, the water level in the tank is falling. As soon as I include some helpers hauling the waste water back up to the top of the tank (in Figure 1-78), you see that we have to put work in to get work out.

Similarly, a battery may seem to be giving power out without taking anything in, but the chemical reactions inside it are changing pure metals into metallic compounds, and the power we get out of a battery is enabled by this change of state. If it's a rechargeable battery, we have to push power back into it to reverse the chemical reactions.

Going back to the tank of water, suppose we can't get enough power out of it to turn the wheel. One answer could be to add more water. The height of the water will create more force. This would be the same as putting two batteries end to end, positive to negative, in series, to double the voltage. See Figure 1-80. As long as the resistance in the circuit remains the same, greater voltage will create more amperage, because amperage = voltage/resistance.

What if we want to run two wheels instead of one? We can punch a second hole in the tank, and the force (voltage) will be the same at each of them. However, the water level in the tank will drop twice as fast. Really, we'd do better to build a second tank, and here again the analogy with a battery is good. If you wire two batteries side by side, in parallel, you get the same voltage, but for twice as long. The two batteries may also be able to deliver more current than if you just used one. See Figure 1-81.

Summing up:

- Two batteries in series deliver twice the voltage.
- Two batteries in parallel can deliver twice the current.

All right, that's more than enough theory for now. In the next chapter, we'll continue with some experiments that will build on the foundations of knowledge about electricity, to take us gradually toward gadgets that can be fun and useful.

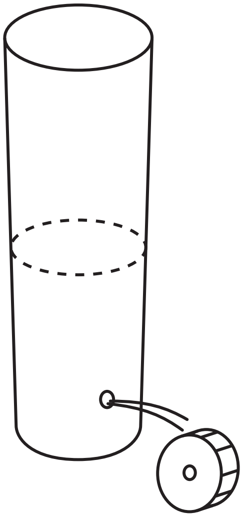

Figure 1-77. *If you want to get work out of a system...*

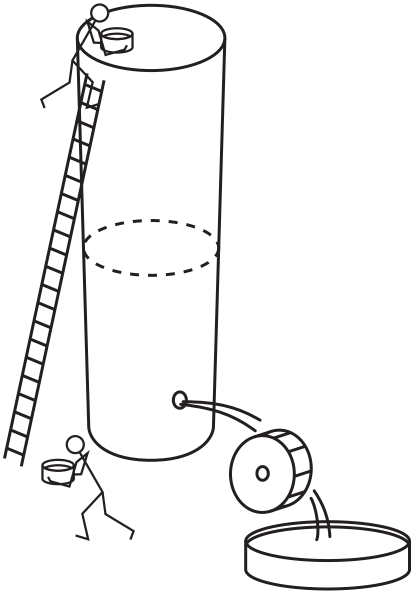

Figure 1-78. *. . . somehow or other you have to put work back into it.*

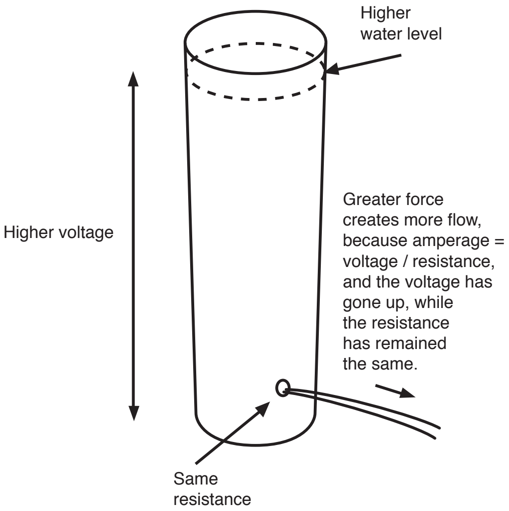
Higher water level
Higher voltage
Greater force creates more flow, because amperage = voltage / resistance, and the voltage has gone up, while the resistance has remained the same.
Same resistance

Figure 1-79. *Greater force generates more flow, as long as the resistance remains the same.*

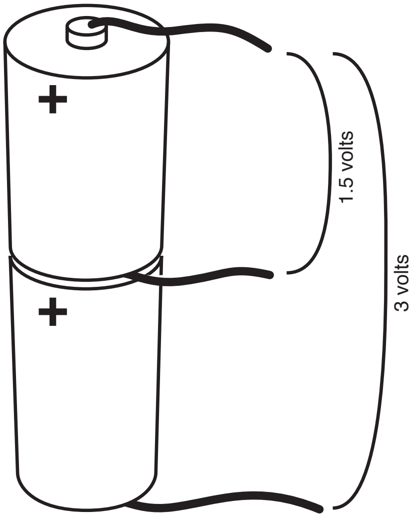
+
+
1.5 volts
3 volts

Figure 1-80. *When you place two equal batteries in series, you double the voltage.*

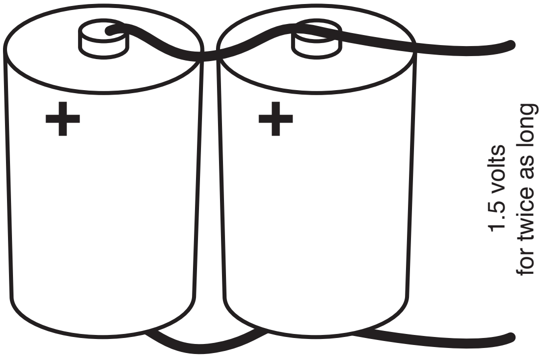
+
+
1.5 volts
for twice as long

Figure 1-81. *Two equal batteries that are wired in parallel will deliver the same voltage for twice as long as one.*

Switching Basics and More

2

The concept of switching is fundamental in electronics, and I'm not just talking about power switches. By "switching," I mean using one flow of electricity to switch, or control, another. This is such an important principle that no digital device can exist without it.

Today, switching is mostly done with semiconductors. Before I deal with them, I'll back up and illustrate the concept by introducing you to relays, which are easier to understand, because you can see what's going on inside them. And before I get to relays, I'll deal with everyday on/off switches, which may seem very simple—but we have to nail down the basics.

Also in this chapter, I'll deal with capacitance, because capacitance and resistance are fundamental to electronic circuits. By the end of the chapter, you should have a basic grounding in electronics and be able to build the noise-making section of a simple intrusion alarm. This will be your first circuit that does something genuinely useful!

IN THIS CHAPTER

Shopping List: Experiments 6 Through 11
Experiment 6: Very Simple Switching
Experiment 7: Relay-Driven LEDs
Experiment 8: A Relay Oscillator
Experiment 9: Time and Capacitors
Experiment 10: Transistor Switching
Experiment 11: A Modular Project

Shopping List: Experiments 6 Through 11

As in the previous shopping list, you should visit the various online suppliers for availability and pricing of components and devices. Manufacturers seldom sell small numbers of parts directly. Check the appendix for a complete list of URLs for all the companies mentioned here.

Devices

- Power supply/universal AC adapter, 3 to 12 volts at 1A (1,000 mA). See Figure 2-1. Part number 273-029 from RadioShack, part PH-62092 by Philips, or similar.
- Breadboard suitable for integrated circuits. Quantity: 1. See Figures 2-2 and 2-3. Part 276-002 from RadioShack, model 383-X1000 made by PSP, part 923252-I by 3M, or similar. A breadboard that has screw terminals mounted beside it will be a little easier to use but more expensive than one that does not have terminals.

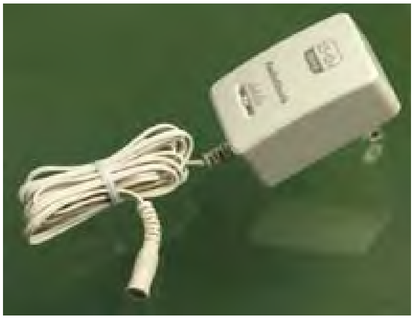

Figure 2-1. *This AC adapter delivers DC from 3 to 12 volts, and is ideal for electronics projects.*

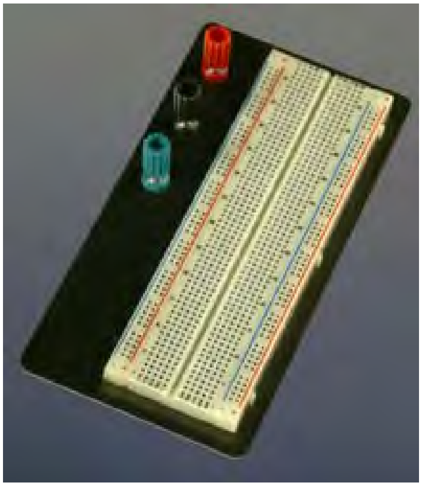

Figure 2-2. *This "breadboard" for quickly constructing electronic circuits has a metal base, and screw terminals for attaching wires from a power supply.*

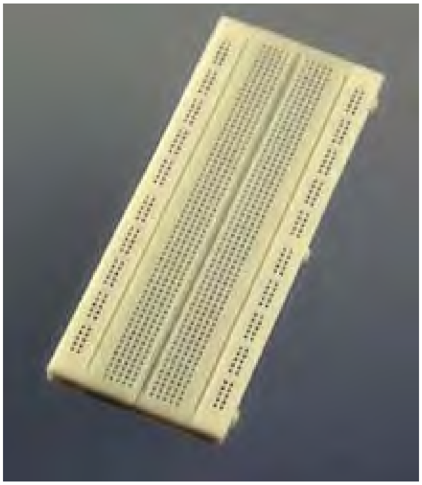

Figure 2-3. *A breadboard without screw terminals is almost as convenient, and is cheaper.*

Tools

Wire strippers

Ideal model 45-121 wire strippers for 16- to 26-gauge wire, or similar. See Figure 2-4. (The "gauge" of the wire tells you how thick it is. A higher gauge means a thinner wire. In this book, we will mainly be using thin wire of 20- to 24-gauge.)

You may also consider the Kronus Automatic Wire Strippers, part 64-2981 from RadioShack, or GB Automatic Wire Strippers, part SE-92 from Amazon.com. See Figure 2-5.

The Kronus and GB wire strippers are functionally identical. The advantage of their design is that it enables you to strip insulation from a wire with one hand. But they do not work well on really thin wire.

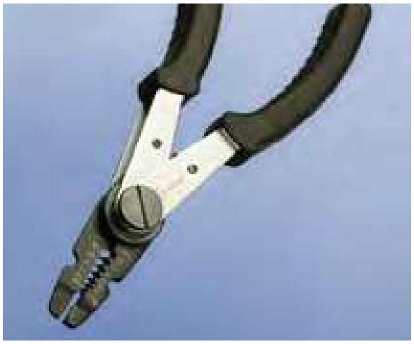

Figure 2-4. *To use these wire strippers, insert a piece of insulated wire in the appropriate-sized hole between the jaws, grip the handles, and pull a section of insulation away. See page 45.*

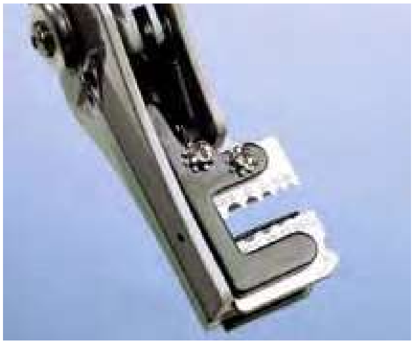

Figure 2-5. *These automatic wire strippers enable one-handed operation, but are not suitable for very small wire diameters. See page 44.*

Supplies

Hookup wire

Solid-conductor, 22-gauge, minimum 25 feet of each color. See Figure 2-6. Part 278-1222 from RadioShack, catalog item 9948T17 from McMaster-Carr, or check eBay for deals.

It's easy to buy the wrong kind of wire. You need solid-core wire, which has a single conductor inside the plastic insulation, not stranded, which has multiple, thinner conductors. See Figures 2-7 and 2-8. You're going to be pushing wires into little holes in a "breadboard," and stranded wire won't let you do this. You will also have problems if you buy wire thicker than 22-gauge. Remember: the lower the gauge number, the thicker the wire.

For a little extra money, you can buy an assortment of precut sections of wire, with ends stripped and ready for use. Try catalog item JW-140 (jumper wire assortment) from All Electronics or search eBay for "breadboard wire." See Figure 2-9.

Patch cords

Patch cords are not strictly necessary but very convenient. You don't want audio or video patch cords, which have a plug on each end; you want wires with alligator clips on each end, also sometimes referred to as "test leads." Try catalog item 461-1176-ND from Digi-Key or catalog item MTL-10 from All Electronics. See Figure 2-10.

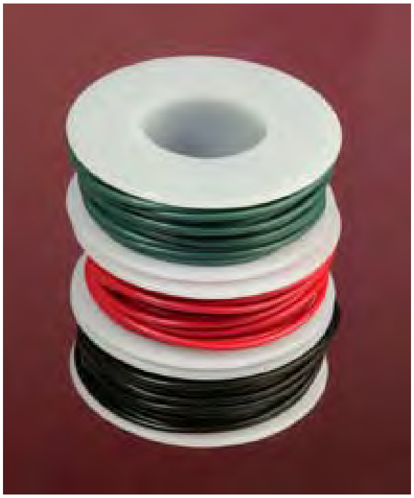

Figure 2-6. *Using hookup wire with different colors of plastic insulation will help you to distinguish one wire from another in your circuits.*

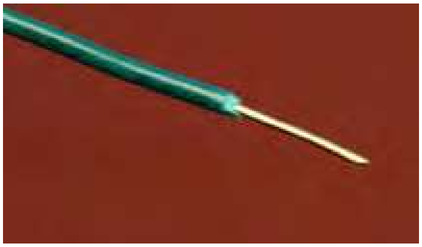

Figure 2-7. *Solid-conductor wire of 22 or 24 gauge is suitable for most of the experiments in this chapter.*

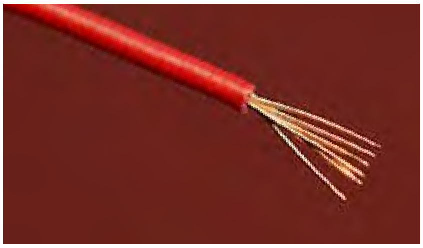

Figure 2-8. *Stranded is more flexible, but cannot be used easily with breadboards.*

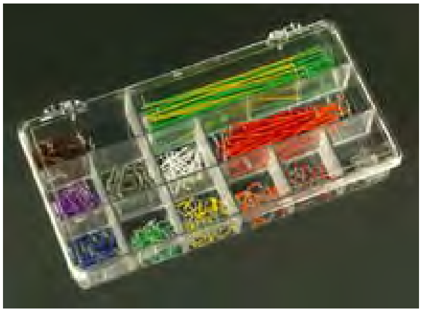

Figure 2-9. *Precut wires with stripped ends can save a lot of time and trouble—if you don't mind paying a little extra.*

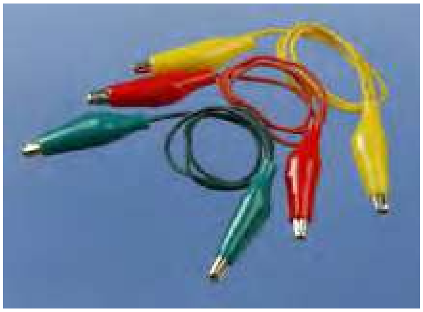

Figure 2-10. *Patch cords, sometimes known as test leads, consist of wires preattached to alligator clips. This is another of those little luxuries that reduces the hassle factor in hobby electronics.*

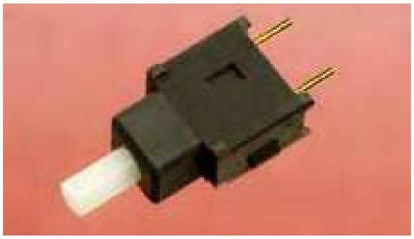

Figure 2-11. *The terminals protruding from this tiny pushbutton are spaced 0.2 inches apart, making it ideal for the "breadboard" that you'll be using.*

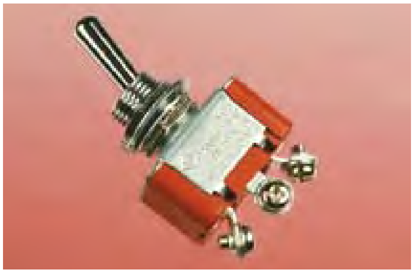

Figure 2-12. *This relatively large toggle switch made by NKK has screw terminals, which will reduce the inconvenience of attaching it to hookup wire.*

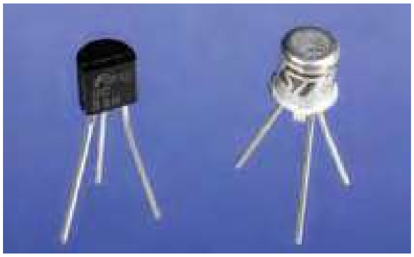

Figure 2-13. *Transistors are commonly sold either in little metal cans or sealed into little lumps of plastic. For our purposes, the packaging makes no difference.*

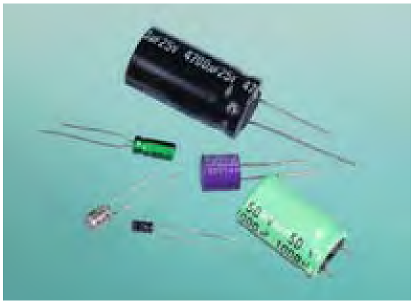

Figure 2-14. *An assortment of electrolytic capacitors.*

Components

Pushbutton

Momentary-on, SPST, sometimes referred to as OFF-(ON) or (ON)-OFF type. Must be PCB- or PC-mount, meaning is extremely small with thin spiky contacts on the bottom. Quantity: 1. See Figure 2-11.

Examples are part number AB11AP by NKK, part MPA103B04 by Alcoswitch, or part EP11SD1CBE by C&K. If you have a choice, buy the cheapest, as we're going to switch very low current.

Switches

Toggle switch, single-pole, double-throw (SPDT), sometimes referred to as ON-ON type. Quantity: 2. See Figure 2-12.

Model S302T-RO by NKK is ideal; it has screw terminals that will eliminate the need for alligator clips. Other options are catalog item MTS-4PC from All Electronics or part 275-603 from RadioShack.

We won't be switching large currents or high voltages, so the exact type of switch is unimportant. However, the terminals on larger-size switches are spaced wider apart, which makes them easier to deal with.

Relays

DPDT, nonlatching, 12v DC. Quantity: 2.

It's important to get the right kind of relay—one whose configuration matches the pictures I'll be using. Look for parts FTR-F1CA012V or FTR-F1CD012V by Fujitsu, G2RL-24-DC12 by Omron, or OMI-SH-212D by Tyco. Avoid substitutions.

Potentiometer

1 megohm linear potentiometer, Part number 271-211 from RadioShack, part number 24N-1M-15R-R from Jameco, or similar.

Transistors

NPN transistor, general-purpose, such as 2N2222 by STMicroelectronics, part PN2222 by Fairchild, or part 2N2222 from RadioShack. Quantity: 4. See Figure 2-13.

2N6027 programmable unijunction transistor manufactured by On Semiconductor or Motorola. Quantity: 4.

Capacitors

Electrolytic capacitors, assorted. Must be rated for a minimum of 25 volts and include at least one capacitor of 1,000 µF (microfarads). If you search on eBay, make sure you find *electrolytic* capacitors. If they're rated for higher voltages, that's OK, although they will be physically larger than you need. See Figure 2-14.

Ceramic capacitors, assorted. Make sure you get at least one rated at 0.0047 µF (which can also be written as 47 nF). See Figure 2-15.

Resistors

If you bought only a minimal selection for experiments 1 through 5, now's the time to buy a larger assortment, so that you won't be stuck needing the one value that you don't have. 1/4-watt minimum.

Loudspeaker

Any 8Ω, 1-inch miniature loudspeaker such as part 273-092 from RadioShack. See Figure 2-16.

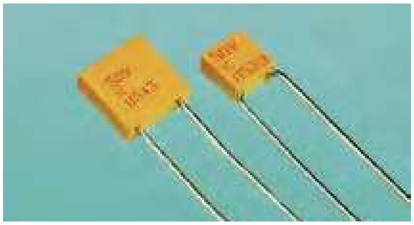

Figure 2-15. *Ceramic capacitors mostly look like this, although many of them are round or bead-shaped instead of square. The packaging shape is unimportant to us.*

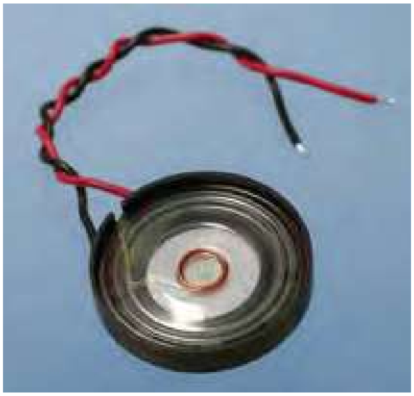

Figure 2-16. *This miniature loudspeaker, just over 1 inch in diameter, is useful for verifying audio output direct from transistor circuits.*

Experiment 6: Very Simple Switching

You will need:

- AA batteries. Quantity: 4.
- Battery carrier for 4 AA batteries. Quantity: 1.
- LED. Quantity: 1.
- Toggle switches, SPDT. Quantity: 2. See Figure 2-12.
- 220Ω or similar value resistor, 1/4-watt minimum. Quantity: 1.
- Alligator clips. Quantity: 8.
- Wire or patch cords. See Figure 2-10, shown previously.
- Wire cutters and wire strippers if you don't use patch cords. See Figure 2-4, shown previously.

In Experiment 3, you illuminated an LED by attaching a battery, and switched it off by removing the battery. For greater convenience our circuits should have proper switches to control power, and while I'm dealing with the general topic of switches, I'm going to explore all the varieties, using a circuit to suggest some possibilities.

Assemble the parts as shown in Figures 2-17 and 2-18. The long lead on the LED must connect with the resistor, because that is the more positive side of the circuit.

You'll notice that you have to include a couple lengths of wire. I suggest green wire to remind you that these sections are not connected directly to positive or to negative power. But you can use any color you like. You can also substitute patch cords, if you have them. However, learning to strip insulation from pieces of wire is a necessary skill, so let's deal with that now.

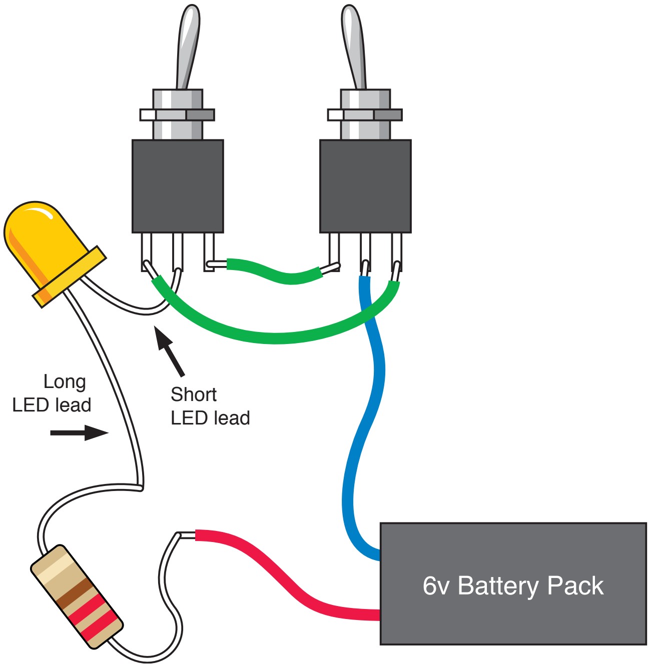

Figure 2-17. *If the LED is on, flipping either of the switches will turn it off. If the LED is off, either of the switches will turn it on. Use alligator clips to attach the wires to each other, and to the switches if your switches don't have screw terminals. Be careful that the clips don't touch each other.*

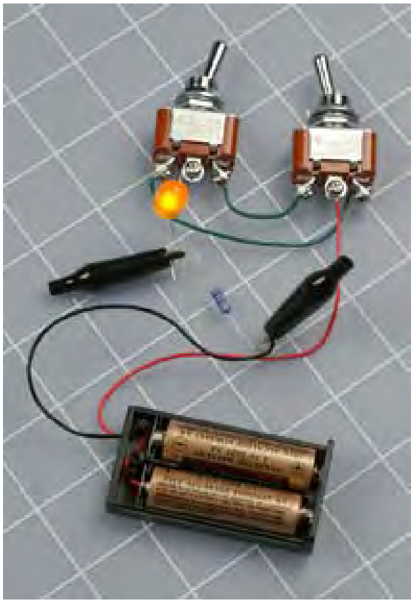

Figure 2-18. *Full-size toggle switches with screw terminals make it easy to hook up this simple circuit.*

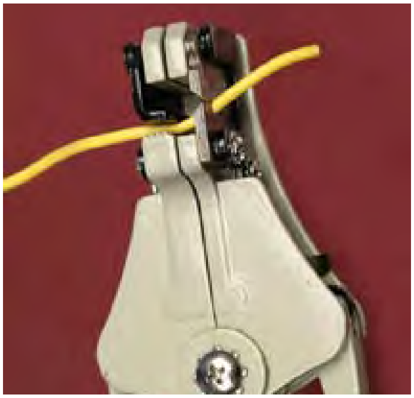

Figure 2-19. *Using automatic wire strippers, when you squeeze the handles the jaw on the left clamps the wire, the sharp grooves on the right bite into the insulation. Squeeze harder and the jaws pull away from each other, stripping the insulation from the wire.*

Tools

If automatic wire strippers (Figure 2-19) don't grip skinny 22-gauge wire very effectively, try the Ideal brand of wire strippers shown back in Figure 2-4, or use plain and simple wire cutters as shown in Figure 2-20. When using wire cutters, you hold the wire in one hand and apply the tool in your other hand, squeezing the handles with moderate pressure—just enough to bite into the insulation, but not so much that you chop the wire. Pull the wire down while you pull the cutters up, and with a little practice you can rip the insulation off to expose the end of the wire.

Macho hardware nerds may use their teeth to strip insulation from wires. When I was younger, I used to do this. I have two slightly chipped teeth to prove it. Really, it's better to use the right tool for the job.

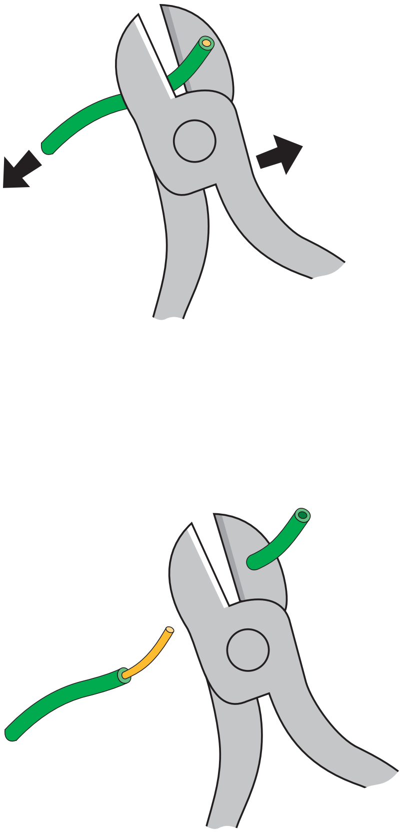

Figure 2-20. *To remove insulation from the end of a thin piece of wire, you can also use wire cutters. This takes a little practice.*

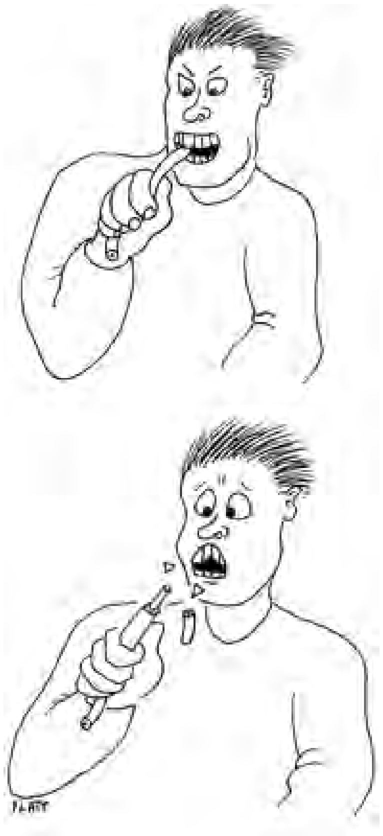

Figure 2-21. *Those who tend to misplace tools, and feel too impatient to search for them, may feel tempted to use their teeth to strip insulation from wire. This may not be such a good idea.*

Connection Problems

Depending on the size of toggle switches that you are using, you may have trouble fitting in all the alligator clips to hold the wires together. Miniature toggle switches, which are more common than the full-sized ones these days, can be especially troublesome (see Figure 2-22). Be patient: fairly soon we'll be using a breadboard, which will eliminate alligator clips almost completely.

Testing

Make sure that you connect the LED with its long wire toward the positive source of power (the resistor, in this case). Now flip either of the toggle switches. If the LED was on, it will go off, and if it was off, it will go on. Flip the other toggle switch, and it will have the same effect. If the LED does not go on at all, you've probably connected it the wrong way around. Another possibility is that two of your alligator clips may have shorted out the battery.

Assuming your two switches do work as I described them, what's going on here? It's time to nail down some basic facts.

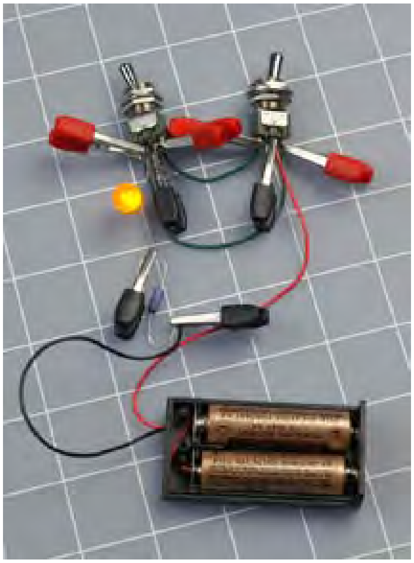

Figure 2-22. *Miniature toggle switches can be used—ideally, with miniature alligator clips—but watch out for short circuits.*

FUNDAMENTALS

All about switches

When you flip the type of toggle switch that you used in Experiment 6, it connects the center terminal with one of the outer terminals. Flip the switch back, and it connects the center terminal with the other outer terminal, as shown in Figure 2-23.

The center terminal is called the pole of the switch. Because you can flip, or throw, this switch to make two possible connections, it is called a *double-throw switch*. As mentioned earlier, a single-pole, double-throw switch is abbreviated *SPDT*.

Some switches are on/off, meaning that if you throw them in one direction they make a contact, but in the other direction, they make no contact at all. Most of the light switches in your house are like this. They are known as *single-throw switches*. A single-pole, single-throw switch is abbreviated *SPST*.

Some switches have two entirely separate poles, so you can make two separate connections simultaneously when you flip the switch. These are called *double-pole switches*. Check the photographs in Figures 2-24 through 2-26 of old-fashioned "knife" switches (which are still used to teach electronics to kids in school) and you'll see the simplest representation of single and double poles, and single and double throws. Various toggle switches that have contacts sealed inside them are shown in Figure 2-27.

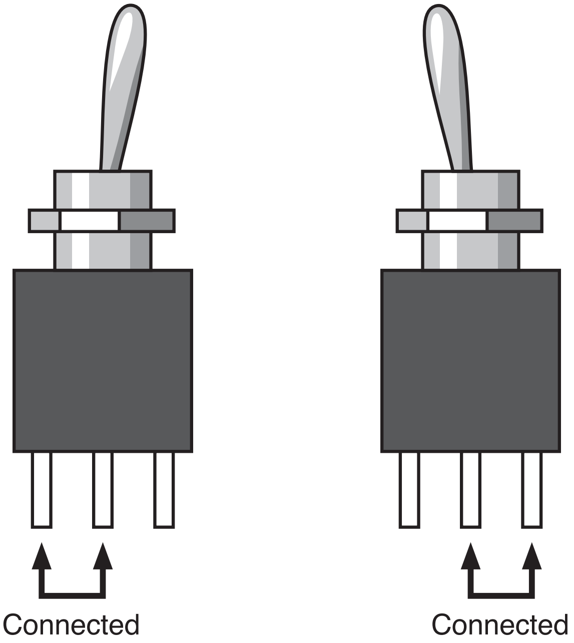

Figure 2-23. *The center terminal is the pole of the switch. When you flip the toggle, the pole changes its connection.*

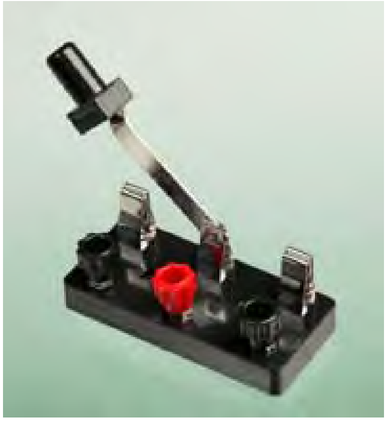

Figure 2-24. *This primitive-looking single-pole, double-throw switch does exactly the same thing as the toggle switches in Figures 2-23 and 2-27.*

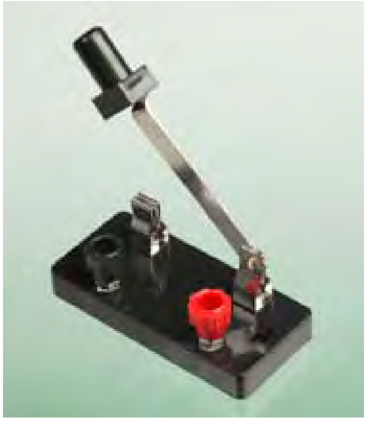

Figure 2-25. *A single-pole, single-throw switch makes only one connection with one pole. Its two states are simply open and closed, on and off.*

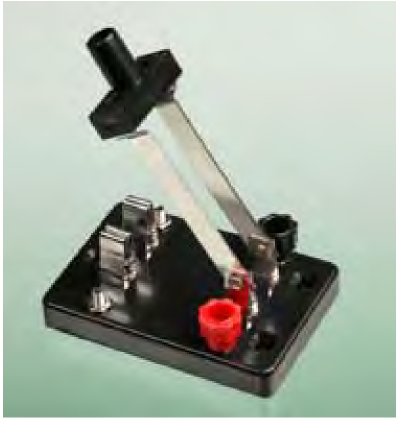

Figure 2-26. *A double-pole, single-throw switch makes two separate on/off connections.*

FUNDAMENTALS

All about switches (continued)

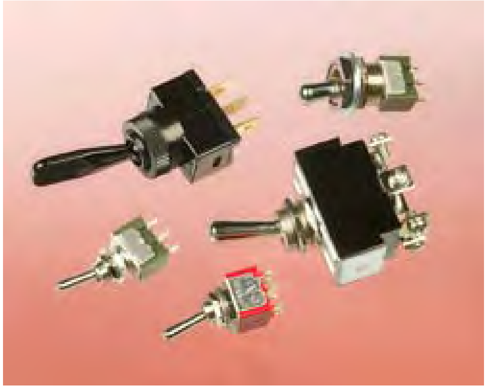

Figure 2-27. *These are all toggle switches. Generally, the larger the switch, the more current it can handle.*

To make things more interesting, you can also buy switches that have three or four poles. (Some rotary switches have even more, but we won't be using them.) Also, some double-throw switches have an additional "center off" position.

Putting all this together, I made a table of possible types of switches (Figure 2-28). When you're reading a parts catalog, you can check this table to remind yourself what the abbreviations mean.

	Single Pole	Double Pole	3-Pole	4-Pole
Single Throw	SPST ON-OFF	DPST ON-OFF	3PST ON-OFF	4PST ON-OFF
Double Throw	SPDT ON-ON	DPDT ON-ON	3PDT ON-ON	4PDT ON-ON
Double Throw with Center Off	SPDT ON-OFF-ON	DPDT ON-OFF-ON	3PDT ON-OFF-ON	4PDT ON-OFF-ON

Figure 2-28. *This table summarizes all the various options for toggle switches and pushbuttons.*

Now, what about pushbuttons? When you press a door bell, you're making an electrical contact, so this is a type of switch—and indeed the correct term for it is a momentary switch, because it makes only a momentary contact. Any spring-loaded switch or button that wants to jump back to its original position is known as a momentary switch. We indicate this by putting its momentary state in parentheses. Here are some examples:

- OFF-(ON): Because the ON state is in parentheses, it's the momentary state. Therefore, this is a single-pole switch that makes contact only when you push it, and flips back to make no contact when you let it go. It is also known as a "normally open" momentary switch, abbreviated "NO."
- ON-(OFF): The opposite kind of momentary single-pole switch. It's normally ON, but when you push it, you break the connection. So, the OFF state is momentary. It is known as a "normally closed" momentary switch, abbreviated "NC."
- (ON)-OFF-(ON): This switch has a center-off position. When you push it either way, it makes a momentary contact, and returns to the center when you let it go.

Other variations are possible, such as ON-OFF-(ON) or ON-(ON). As long as you remember that parentheses indicate the momentary state, you should be able to figure out what these switches are.

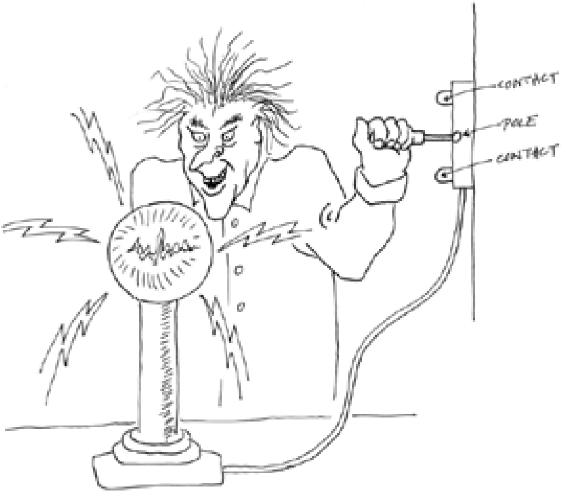

Figure 2-29. *This evil mad scientist is ready to apply power to his experiment. For this purpose, he is using a single-pole, double-throw knife switch, conveniently mounted on the wall of his basement laboratory.*

FUNDAMENTALS

All about switches (continued)

Sparking

When you make and break an electrical connection, it tends to create a spark. Sparking is bad for switch contacts. It eats them until the switch doesn't make a reliable connection anymore. For this reason, you must use a switch that is appropriate for the voltage and amperage that you are dealing with. Electronic circuits generally are low-current, and low-voltage, so you can use almost any switch, but if you are switching a motor, it will tend to suck an initial surge of current that is at least double the rating of the motor when it is running constantly. You should probably use a 4-amp switch to turn a 2-amp motor on and off.

Checking a switch

You can use your meter to check a switch. Doing this helps you find out which contacts are connected when you turn a switch one way or the other. It's also useful if you have a pushbutton and you can't remember whether it's the type that is normally open (you press it to make a connection) or normally closed (you press it to break the connection). Set your meter to measure ohms, and touch the probes to the switch terminals while you work the switch.

This is a hassle, though, because you have to wait while the meter makes an accurate measurement. When you just want to know whether there is a connection, your meter has a "continuity tester" setting. It beeps if it finds a connection, and stays silent if it doesn't. See Figures 2-30 through 2-32 for examples of meters set to test continuity. Figure 2-33 offers an example of a toggle switch being tested for continuity.

Use the continuity-testing feature on your meter only on circuits or components that have no power in them at the time.

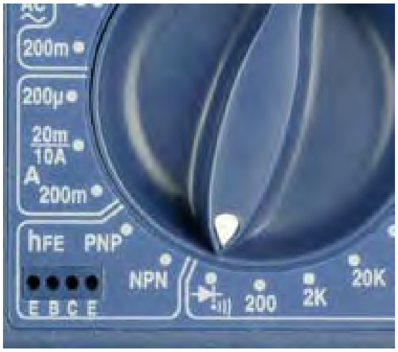

Figure 2-30

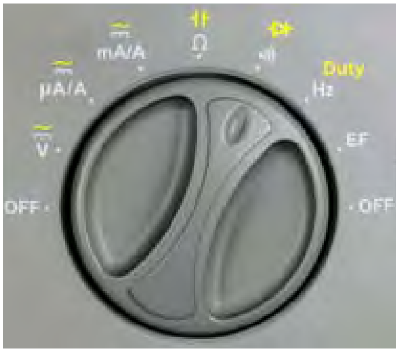

Figure 2-31

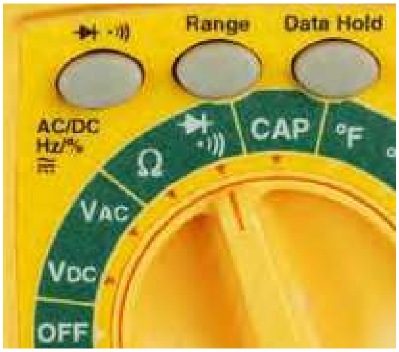

Figure 2-32. *To check a circuit for continuity, turn the dial of your meter to the symbol shown. Only use this feature when there is no power in the component or the circuit that you are testing.*

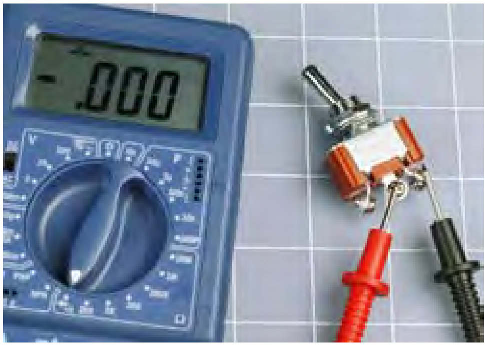

Figure 2-33. *When the switch connects two of its terminals, the meter shows zero resistance between them and will beep if you have set it to verify continuity.*

BACKGROUND

Early switching systems

Switches seem to be such a fundamental feature of our world, and their concept is so simple that it's easy to forget that they went through a gradual process of development and refinement. Primitive knife switches were quite adequate for pioneers of electricity who simply wanted to connect and disconnect electricity to some apparatus in a laboratory, but a more sophisticated approach was needed when telephone systems began to proliferate. Typically, an operator at a "switchboard" needed a way to connect any pair of 10,000 lines on the board. How could it be done?

In 1878, Charles E. Scribner (Figure 2-34) developed the "jack-knife switch," so called because the part of it that the operator held looked like the handle of a jackknife. Protruding from it was a plug, and when the plug was pushed into a socket, it made contact inside the socket. The socket, in fact, was the switch.

Figure 2-34. *Charles E. Scribner invented the "jack-knife switch" to satisfy the switching needs of telephone systems in the late 1800s. Today's audio jacks still work on the same basis.*[*]

Audio connectors on guitars and amplifiers still work on the same principle, and when we speak of them as being "jacks," the term dates back to Scribner's invention. Switch contacts still exist inside a jack socket.

Today, of course, telephone switchboards have become as rare as telephone operators. First they were replaced with relays—electrically operated switches, which I'll talk about later in this chapter. And then the relays were superceded by transistors, which made everything happen without any moving parts. Before the end of this chapter, you'll be switching current using transistors.

---

[*] The photo on which this drawing is based first appeared in *The History of the Telephone* by Herbert Newton Casson in 1910 (Chicago: A. C. McClurg & Co.).

Introducing Schematics

Larger versions of all schematics and breadboard photos are available online at this book's website: http://oreilly.com/catalog/9780596153748.

In Figure 2-35, I've redrawn the circuit from Experiment 6 in a simplified style known as a "schematic." From this point onward, I will be illustrating circuits with schematics, because they make circuits easier to understand. You just need to know a few symbols to interpret them.

Compare the schematic here with the drawing of the circuit in Figure 2-17. They both show exactly the same thing: Components, and connections between them. The gray rectangles are the switches, the zigzag thing is the resistor, and the symbol with two diagonal arrows is the LED.

The schematic LED symbol includes two arrows indicating that it emits light, because there are some kinds of diodes, which we'll get to later, that don't. The triangle inside the diode symbol always points from positive to negative.

Trace the path that electricity can take through the circuit and imagine the switches turning one way or the other. You should see clearly now why either switch will reverse the state of the LED from on to off or off to on.

This same circuit is used in houses where you have a switch at the bottom of a flight of stairs, and another one at the top, both controlling the same lightbulb. The wires in a house are much longer, and they snake around behind the walls, but because their connections are still the same, they could be represented with the same basic schematic. See Figure 2-36.

A schematic doesn't tell you exactly where to put the components. It just tells you how to join them together. One problem: Different people use slightly different schematic symbols to mean the same thing. Check the upcoming section, "Fundamentals: Basic schematic symbols," for the details.

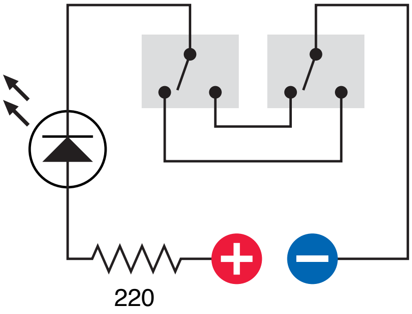

Figure 2-35. *This schematic shows the same circuit as in Figure 2-17 and makes it easier to see how the switches function.*

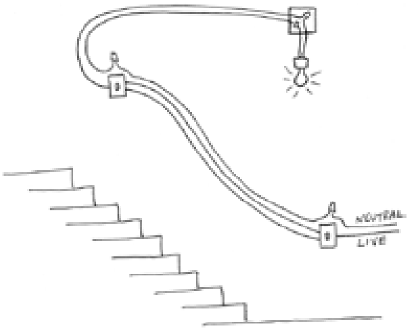

Figure 2-36. *The two-switch circuit shown in Figures 2-17 and 2-35 is often found in house wiring, especially where switches are located at the top and bottom of a flight of stairs. This sketch shows what you might find inside the walls. Wires are joined with "wire nuts" inside boxes that are hidden from everyday view.*

FUNDAMENTALS

Basic schematic symbols

Schematic symbols are like words in a language: they have mutated over the years into a confusing range of variations. A simple on/off (single-pole, single-throw, or SPST) switch, for instance, can be represented by any of the symbols shown in Figure 2-37. They all mean exactly the same thing.

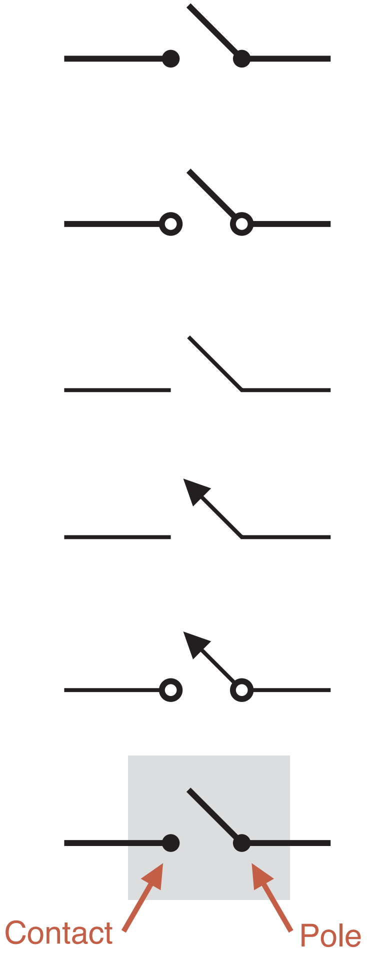

Figure 2-37. *Variations on a theme: Just some of the different styles used to depict a single-pole, single-throw switch in schematic diagrams. The bottom version is the style used in this book.*

Figure 2-38 shows double-pole, double-throw switches. A dotted line indicates a mechanical connection inside the switch, so that when you flip it, you affect both the poles simultaneously. Remember, the poles are electrically isolated from each other.

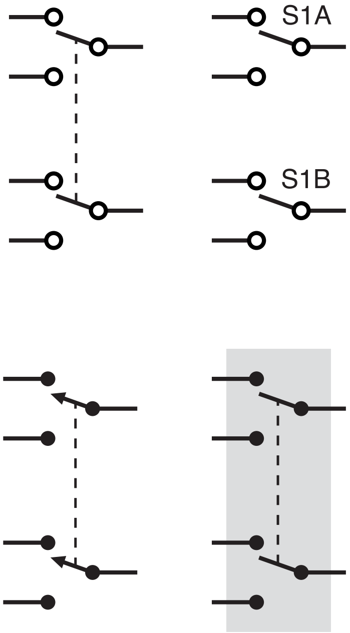

Figure 2-38. *More variations: some different styles for depicting a double-pole, double-throw switch. The style at bottom-right is used in this book.*

Once in a while, you may find a schematic in which switches seem to be scattered around, but the way they are identified (such as S1A, S1B, S1C, and so on) tells you that this is really all one switch with multiple poles.

In the schematics in this book, I'll place a gray rectangle behind each switch. This gray rectangle is not a standard symbol; you won't find it in other books. I'm just including it to remind you that the parts inside are all contained in one package.

A very important stylistic variation in schematics is the way they show whether wires make a connection with each other. Old schematics used to show a little semicircular bump in a wire if it crossed another wire without making a connection. Because modern circuit-drawing software doesn't create this style of schematic, it is no longer often used. The modern style, which you are likely to find if you browse through schematics online, can be summarized like this:

- A dot joining two wires indicates an electrical connection.
- No dot indicates no connection.

FUNDAMENTALS

Basic schematic symbols (continued)

The problem is that this is not very intuitive, especially when you're just beginning to use schematics. When you see two wires crossing, you can easily imagine that they are making a connection, even though there's no dot at the intersection. Therefore, in the interests of clarity, I've chosen to use the old "semicircular bump" style of schematic in this book (see Figure 2-39). It can be summarized like this:

- A dot joining two wires indicates an electrical connection.
- A bump in a wire that crosses another wire indicates no connection.

In this book, you won't find wires crossing each other without either a dot or a bump.

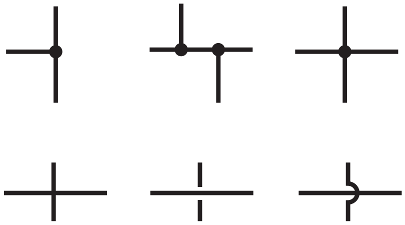

Figure 2-39. *In wiring schematics, a dot always indicates an electrical connection. However, the cross-shaped intersection of wires at top-right is considered bad style because if the dot is accidentally omitted or poorly printed, the intersection can be mistaken for the type shown at bottom left, in which the wires do not make a connection. All three of the configurations in the bottom row indicate no connection, the first example being the most common style, the center example being least common, and the third being the most old-fashioned—although for reasons of clarity, it is used in this book.*

In a battery-powered circuit, you may find a battery symbol, but more often you will find a little note indicating where positive voltage enters the system, while negative is indicated by a "ground" symbol. In fact there may be ground symbols all over the place. You have to remember that when you build a circuit, all the wires leading to grounds must actually be connected together, to the negative side of the voltage supply.

The idea of the ground symbol dates back to the time when electronic gadgets were mounted on a metal chassis, which was connected to the negative side of the power supply. The ground symbol really meant "connect to the chassis." Some variants in the ground symbol are shown in Figure 2-40.

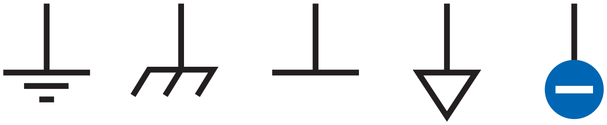

Figure 2-40. *All of these symbols are used to mean the same thing: connect the wire to "ground" or "chassis" or the negative side of the power source. The far-right symbol is used in this book.*

In this book, we have color throughout, so I'll show a red positive and blue negative to clarify where the power is connected, and I won't use ground symbols. Once again, my purpose is to minimize the risk of misunderstandings, because I know how frustrating it is to build a circuit that doesn't work.

A big inconsistency in schematics is the way in which they show resistors. The traditional zigzag symbol has been abandoned in Europe. Instead they use a rectangle with a number inside indicating the number of ohms. See Figure 2-41. The Europeans also changed the way in which decimal points are represented: they omit them as much as possible, because in badly printed schematics, the little dots tend to get lost (or can be confused with dust and dirt). So, a 4.7KΩ resistor will be listed as 4K7, and a 1.2MΩ resistor will be 1M2. I like this notation, so I'm going to be using it myself, but I'll be keeping the zigzag resistor symbol, which remains widely used in the United States.

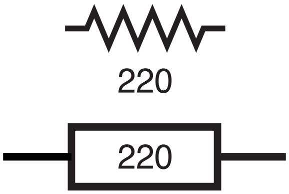

Figure 2-41. *Two styles for depicting a 220Ω resistor. The upper version is traditional, and still used in the United States. The lower version is European.*

FUNDAMENTALS

Basic schematic symbols (continued)

Potentiometers suffer from the same inconsistent style between the United States and Europe, but either way, you'll find an arrow showing where the wiper (usually, the center terminal) touches the resistance. See Figure 2-42. And sometimes LEDs are shown inside circles, and sometimes not. I prefer circles, myself. See Figure 2-46.

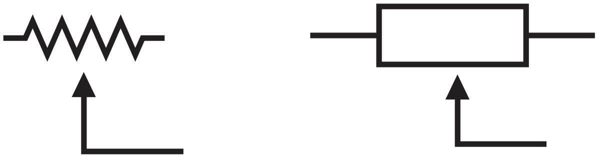

Figure 2-42. *Potentiometer symbols: the left is traditional and used in the United States, the right is European. In both cases the arrow indicates the wiper (usually the center terminal).*

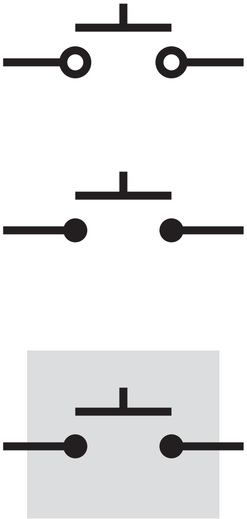

Figure 2-43. *Three ways of indicating a pushbutton switch.*

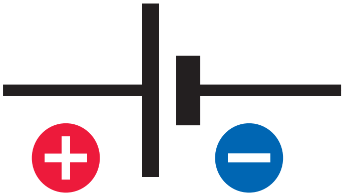

Figure 2-44. *The battery symbol is usually shown without + and – symbols. I've added them for clarity.*

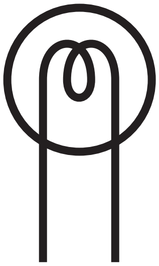

Figure 2-45. *Symbol for an incandescent lightbulb.*

I'll explore other symbol variants later in the book. Meanwhile, the most important things to remember are:

- The positions of components in a schematic are not important.
- The styles of symbols used in a schematic are not important.
- The connections between the components are extremely important.

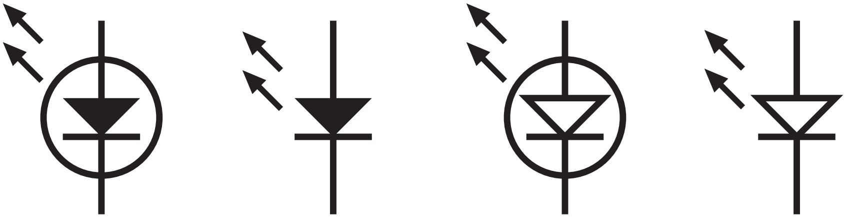

Figure 2-46. *Sometimes an LED is shown with a circle around it; sometimes not. In this book, I will include the circle. The arrows indicate emitted light.*

FUNDAMENTALS

Basic schematic symbols (continued)

For example, the three LED circuits that I have included in Figure 2-47 show components in different positions, using different symbols, but all three circuits function exactly the same way, because their connections are the same. In fact, they all depict the circuit that you built in Experiment 4, shown in Figure 1-50.

Often the symbols in a schematic are placed so that the circuit is most intuitively easy to understand, regardless of how you may build it with actual components. Compare the example in Figure 2-48, showing the two DPDT switches, with the version shown back in Figure 2-35. The previous one looked more like your bench-top version of it, but Figure 2-48 shows the flow of electricity more clearly.

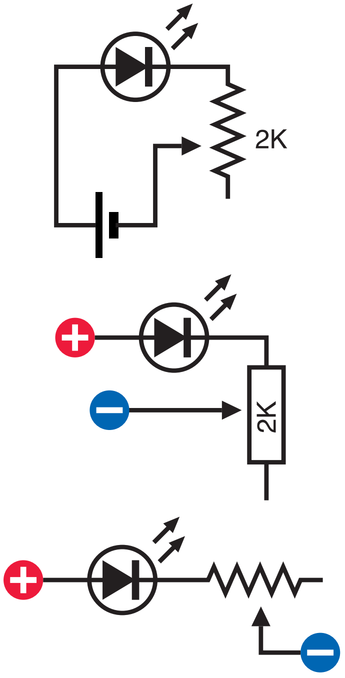

Figure 2-47. *These three schematics all depict the same basic circuit. It's the circuit that you built with the potentiometer in Experiment 4.*

In many schematics, the positive side of the power supply is shown at the top of the diagram, and negative or ground at the bottom. Many people also tend to draw a schematic with an input (such as an audio input, in an amplifier circuit) at the left side, and the output at the right. So, "positive voltage" flows from top to bottom while a signal tends to pass from left to right.

When I was planning this book, initially I drew the schematics to conform with this top-to-bottom, left-to-right convention, but as I started building and testing the circuits, I changed my mind. We use a device known as a "breadboard" to create circuits, and its internal connections require us to lay out components very differently from a typical schematic. When you're starting to learn electronics, it's very confusing to try to rearrange components from a schematic in the configuration that you need for a breadboard.

Therefore, throughout this book, you'll find that I have drawn the schematics to imitate the way you'll wire them on a breadboard. I believe the advantages of doing things this way outweigh the disadvantages of being a little different from the schematic styles that are used elsewhere.

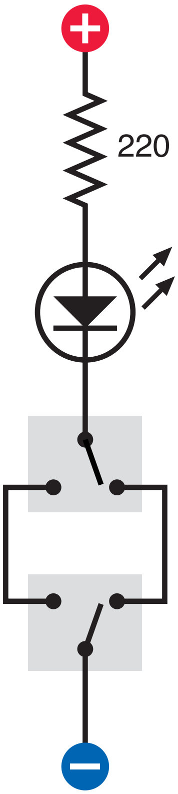

Figure 2-48. *This schematic is just another, clearer, simpler way of showing the circuit that appeared in Figure 2-35.*

Experiment 7: Relay-Driven LEDs

You will need:

- AC adapter, wire cutters and strippers.
- DPDT relay. Quantity: 2.
- LEDs. Quantity: 2.
- Resistor, 680Ω approx. Quantity: 1.
- Pushbutton, SPST. Quantity: 1.
- Hookup wire, 22 gauge, or patch cords.
- Alligator clips. Quantity: 8.
- Utility knife.

The next step in our exploration of switching is to use a remote-controlled switch. By "remote-controlled," I mean one to which you can send a signal to turn it on or off. This kind of switch is known as a *relay*, because it relays an instruction from one part of a circuit to another. Often a relay is controlled by a low voltage or small current, and switches a larger voltage or higher current.

This setup can be cost-effective. When you start your car, for instance, a relatively small, cheap switch sends a small signal down a relatively long, thin, inexpensive piece of wire, to a relay that is near the starter motor. The relay activates the motor through a shorter, thicker, more expensive piece of wire, capable of carrying as much as 100 amps.

Similarly, when you raise the lid on an a top-loading washing machine during its spin cycle, you close a lightweight switch that sends a small signal down a thin wire to a relay. The relay handles the bigger task of switching off the large motor spinning the drum full of wet clothes.

Before you begin this experiment, you need to upgrade your power supply. We're not going to use batteries anymore, because most relays require more than 6 volts, and in any case you should have a power supply that can give you a variety of voltages on demand. The simplest way to achieve this is by using an AC adapter.

First you'll set up the AC adapter. After you have it running, you'll use it to power the relay. Initially the relay will just switch between two LEDs, but then you'll modify the circuit to make the LEDs flash automatically. Finally you'll rebuild the circuit on a breadboard, and say goodbye to alligator clips, for most of the time at least.

Preparing Your AC Adapter

An AC adapter plugs into the wall and converts the high-voltage AC supply in your home into a safe, low DC voltage for electronic devices. Any charger that you use with your cell phone, or iPod, or laptop computer is a special-purpose AC adapter that delivers only one voltage via a specific type of plug. I've asked you to buy a general-purpose adapter that delivers many different voltages, and we're going to begin by getting rid of its plug.

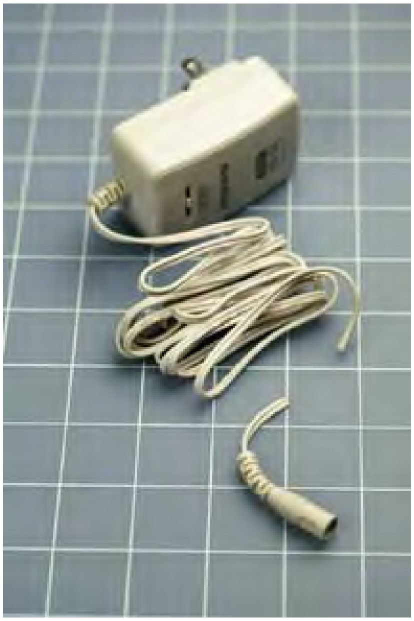

Figure 2-49. *Preparing an AC adapter. First, cut off the little low-voltage plug and throw it away.*

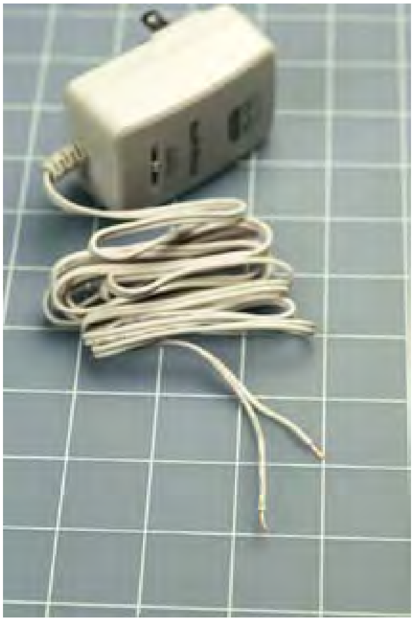

Figure 2-50. *Second, strip the wires, making one shorter than the other to reduce the risk of them touching. Color one of the adapter wires red with a marker, to identify it as the positive one.*

1. It's important to make sure that your AC adapter is not plugged into the wall!
2. Chop off the little plug at the end of its wire. See Figure 2-49.
3. Use a box cutter or utility knife or scissors to make a half-inch cut between the two conductors, and then pull the conductors apart a couple of inches.
4. Use wire cutters to trim one of the conductors shorter than the other, so that after you strip away a little of the insulation, the exposed copper wires cannot easily touch each other. This is a precaution against short-circuiting your AC adapter and burning it out.
5. Strip the two conductors using your wire strippers. Twist the copper strands between finger and thumb so that there are no loose strands sticking out. See Figure 2-50.
6. Make sure that the two wires are not touching each other, and plug your AC adapter into a wall outlet. Set your meter to DC volts and apply the meter probes to the wires from the adapter. If the voltage is preceded with a minus sign, you have the probes the wrong way around. Reverse them and the minus sign should go away. This tells you which wire is positive.
7. Mark the positive wire from the adapter. If the wire has white insulation, you can mark it with a red marker. If it has black insulation, you can tag it with a label. The positive wire will remain positive regardless of which way up you plug the AC adapter into a wall outlet.

The Relay

The type of relay that I want you to use has little spiky legs on the bottom, in a standardized layout. If you buy some other kind of relay, you will have to figure out for yourself which pins are connected to the coil inside, which pins go to the poles of the switch inside it, and which go to the normally closed and normally open contacts. You can check the manufacturer's data sheet for this purpose, but I strongly suggest you use one of the relays mentioned in the shopping list, so that you can follow the instructions here more easily.

I asked you to buy two relays so that you can use one for investigational purposes—meaning that you can break it open and take a look inside. If you do this very, very carefully, the relay should still be usable afterward. If not, well, you still have a spare.

The easiest way to open the relay is with a box cutter or utility knife. Figures 2-52, 2-53, and 2-54 show the technique. Shave the edges of the plastic shell containing the relay, beveling them until you see just a hair-thin opening. Don't go any farther; the parts inside are very, very close to its housing. Now pop the top off. You can use needle-nosed pliers to nibble the rest of the shell away. Read the following section, "Fundamentals: Inside a relay," and then apply power to the relay to see how it works.

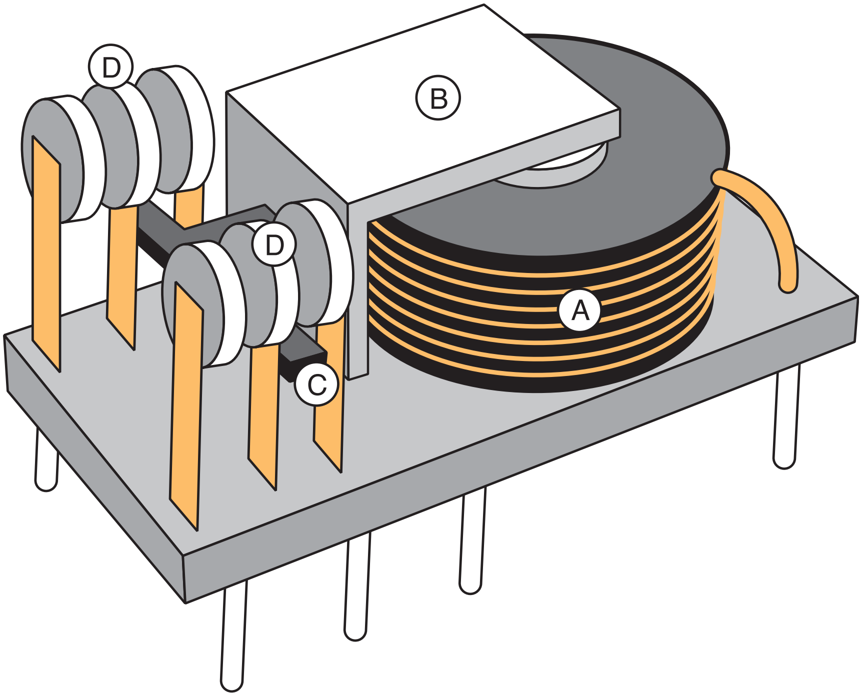

Figure 2-51. *This is one way that the parts inside a relay can be arranged. The coil, A, generates a magnetic attraction pulling lever B downward. A plastic extension, C, pushes outward against flexible metal strips and moves the poles of the relay, D, between the contacts.*

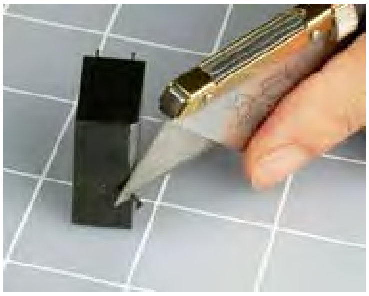

Figure 2-52. *To look inside a sealed relay, shave the top edges of the plastic package with a utility knife til you open a thin crack.*

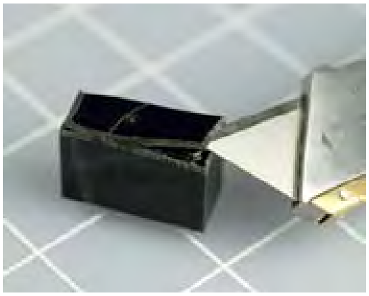

Figure 2-53. *Insert the blade of your knife to pry open the top, then repeat the procedure for the sides.*

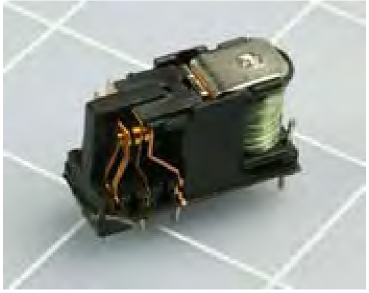

Figure 2-54. *If you are really, really careful, the relay should still work after you open it.*

Figure 2-55

Figure 2-56. *Patience is essential when carving the edges of a relay package in order to open it. Faster methods such as a tomahawk or a flamethrower will satisfy the emotional needs of those with a short attention span, but results may be unpredictable.*

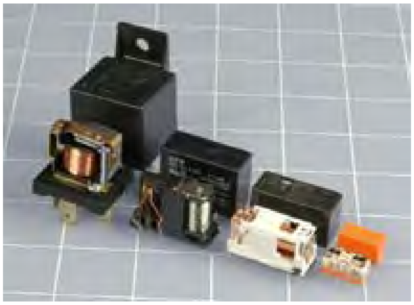

Figure 2-57. *Four assorted 12-volt relays, shown with and without their packages. The automotive relay (far left) is the simplest and easiest to understand, because it is designed without much concern for the size of the package. Smaller relays are more ingeniously designed, more complex, and more difficult to figure out. Usually, but not always, a smaller relay is designed to switch less current than a larger one.*

FUNDAMENTALS

Inside a relay

A relay contains a coil of wire wrapped around an iron core. When electricity runs through the coil, the iron core exerts a magnetic force, which pulls a lever, which pushes or pulls a springy strip of metal, closing two contacts. So as long as electricity runs through the coil, the relay is "energized" and its contacts remain closed.

When the power stops passing through the coil, the relay lets go and the springy strip of metal snaps back into its original position, opening the contacts. (The exception to this rule is a latching relay, which requires a second pulse through a separate coil to flip it back to its original position; but we won't be using latching relays until later in the book.)

Relays are categorized like switches. Thus, you have SPST relays, DPST, SPDT, and so on.

Compare the schematics in Figure 2-58 with the schematics of switches in Figure 2-38. The main difference is that the relay has a coil that activates the switch. The switch is shown in its "relaxed" mode, when no power flows through the coil.

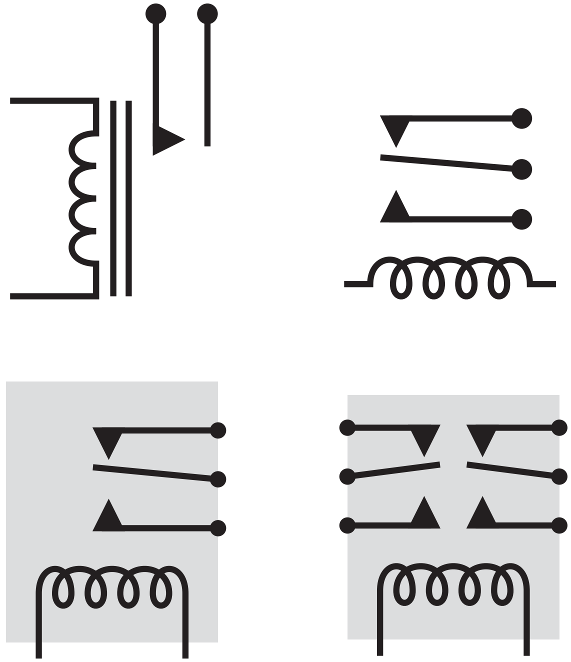

Figure 2-58. *Various ways to show a relay in a schematic. Top left: SPST. Top right and bottom left: SPDT. Bottom right: DPDT. The styles at bottom-left and bottom-right will be used in this book.*

The contacts are shown as little triangles. When there are two poles instead of one, the coil activates both switches simultaneously.

Most relays are nonpolarized, meaning that you can run electricity through the coil in either direction, and the relay doesn't care. You should check the data sheet to make sure, though. Some relay coils work on AC voltage, but almost all low-voltage relays use direct current—a steady flow of electricity, such as you would get from a battery. We'll be using DC relays in this book.

Relays suffer from the same limitations as switches: their contacts will be eroded by sparking if you try to switch too much voltage. It's not worth saving a few dollars by using a relay that is rated for less current or voltage than your application requires. The relay will fail you when you need it most, and may be inconvenient to replace.

Because there are so many different types of relays, read the specifications carefully before you buy one. Look for these basics:

Coil voltage

The voltage that the relay is supposed to receive when you energize it.

Set voltage

The minimum voltage that the relay needs to close its switch. This will be a bit less than the ideal coil voltage.

Operating current

The power consumption of the coil, usually in milliamps, when the relay is energized. Sometimes the power is expressed in milliwatts.

Switching capacity

The maximum amount of current that you can switch with contacts inside the relay. Usually this is for a "resistive load," meaning a passive device such as light bulb. When you use a relay to switch on a motor, the motor takes a big initial surge of current before it gets up to speed. In this case, you should choose a relay rated for double the current that the motor draws when it is running.

Procedure

Turn the relay with its legs in the air and attach wires and LEDs as shown in Figure 2-59, with a 680Ω resistor (a 1K resistor will be OK if you don't have the correct value). Also attach a pushbutton switch. (Your pushbutton switch may look different from the one shown, but as long as it is a SPST pushbutton with two contacts at the bottom, it will work the same way.) When you press the pushbutton, the relay will make the first LED go out and the second LED light up. When you release the pushbutton, the first LED lights up and the second one goes out.

How It Works

Check the schematic in Figure 2-60 and compare it with Figure 2-59. Also see Figure 2-62, which shows how the pins outside the relay make connections inside the relay when its coil is energized, and when it is not energized.

This is a DPDT relay, but we are only using one pole and ignoring the other. Why not buy a SPDT relay? Because I want the pins to be spaced the way they are when you will upgrade this circuit by transferring it onto a breadboard, which will happen very shortly.

On the schematic, I have shown the switch inside the relay in its relaxed state. When the coil is energized, the switch flips upward, which seems counterintuitive, but just happens to be the way that this particular relay is made.

When you're sure you understand how the circuit works, it's time to move on to the next step: making a small modification to get the relay to switch itself on and off, as we'll do in Experiment 8.

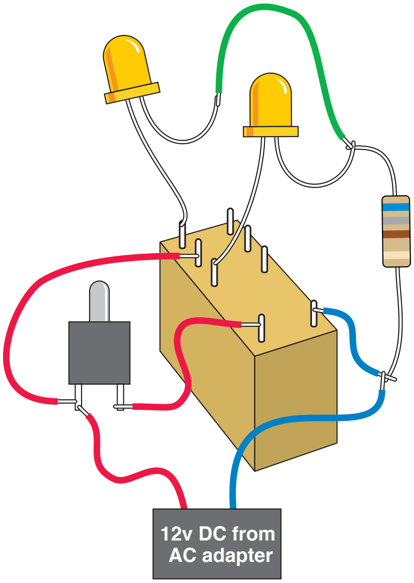

Figure 2-59. *As before, you can use patch cords, if you have them, instead of some of the wired connections shown here.*

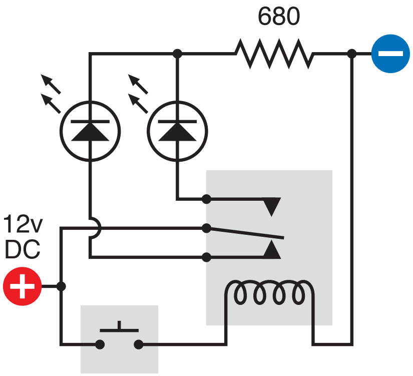

Figure 2-60. *Same circuit, shown in schematic form.*

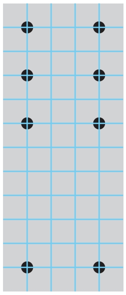

Figure 2-61. *The layout of the pins of the relay, superimposed on a grid of 1/10-inch squares. This is the type of relay that you will need in Experiment 8.*

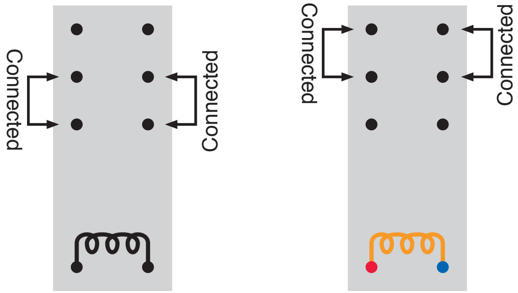

Figure 2-62. *How the relay connects the pins, when it is not energized (left) and when it is energized (right).*

Larger versions of all schematics and breadboard photos are available online at this book's website: http://oreilly.com/catalog/9780596153748.

Experiment 8: A Relay Oscillator

You will need:

- AC adapter, breadboard, wire, wire cutters and strippers.
- DPDT relay. Quantity: 1.
- LEDs. Quantity: 2.
- Pushbutton, SPST. Quantity: 1.
- Alligator clips. Quantity: 8.
- Resistor, approximately 680Ω. Quantity: 1.
- Capacitor, electrolytic, 1,000 µF. Quantity: 1.

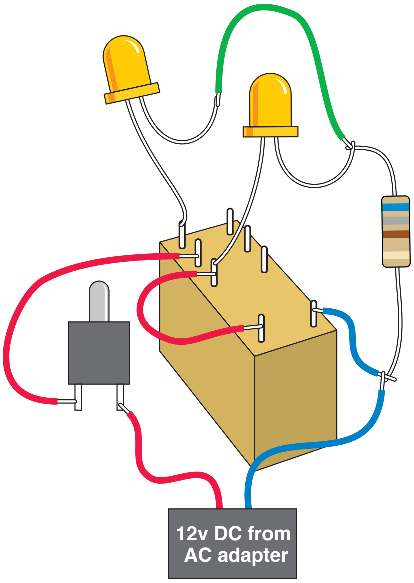

Figure 2-63. *A small revision to the previous circuit causes the relay to start oscillating when power is applied.*

Look at the revised drawing in Figure 2-63 and the revised schematic in Figure 2-64 and compare them with the previous ones. Originally, there was a direct connection from the pushbutton to the coil. In the new version, the power gets to the coil by going, first, through the contacts of the relay.

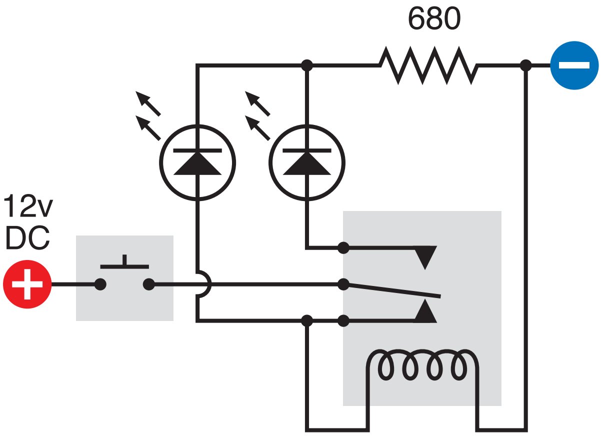

Figure 2-64. *The oscillator circuit shown in schematic form.*

Now, when you press the button, the contacts in their relaxed state feed power to the coil as well as to the lefthand LED. But as soon as the coil is energized, it opens the contacts. This interrupts the power to the coil—so the relay relaxes, and the contacts close again. They feed another pulse of power to the coil, which opens the contacts again, and the cycle repeats endlessly.

Because we're using a very small relay, it switches on and off extremely fast. In fact, it oscillates perhaps 50 times per second (too fast for the LEDs to show what's really happening). Make sure your circuit looks like the one in the diagram, and then press the pushbutton very briefly. You should hear the relay make a buzzing sound. If you have impaired hearing, touch the relay lightly with your finger, and you should feel the relay vibrating.

When you force a relay to oscillate like this, it's liable to burn itself out or destroy its contacts. That's why I asked you to press the pushbutton briefly. To make the circuit more practical, we need something to slow the relay down and prevent it from self-destructing. That necessary item is a capacitor.

Adding Capacitance

Add a 1,000 µF electrolytic capacitor in parallel with the coil of the relay as shown in the diagram in Figure 2-65 and the schematic in Figure 2-66. Check Figure 2-14 if you're not sure what a capacitor looks like. The 1,000 µF value will be printed on the side of it, and I'll explain what this means a little later.

Make sure the capacitor's *short* wire is connected to the *negative* side of the circuit; otherwise, it won't work. In addition to the short wire, you should find a minus sign on the body of the capacitor, which is there to remind you which side is negative. Electrolytic capacitors are fussy about this.

When you press the button now, the relay should click slowly instead of buzzing. What's happening here?

A capacitor is like a tiny rechargeable battery. It's so small that it charges in a fraction of a second, before the relay has time to open its lower pair of contacts. Then, when the contacts are open, the capacitor acts like a battery, providing power to the relay. It keeps the coil of the relay energized for about one second. After the capacitor exhausts its power reserve, the relay relaxes and the process repeats.

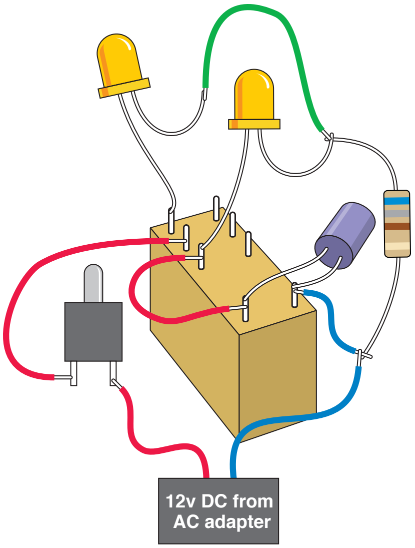
12v DC from AC adapter

Figure 2-65. *Adding a capacitor makes the relay oscillate more slowly.*

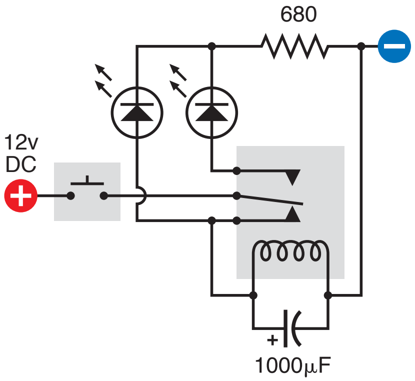

Figure 2-66. *The capacitor appears at the bottom of this schematic diagram.*

FUNDAMENTALS

Farad basics

The Farad is an international unit to measure capacitance. Modern circuits usually require small capacitors. Consequently it is common to find capacitors measured in microfarads (one-millionth of a farad) and even picofarads (one-trillionth of a farad). Nanofarads are also used, more often in Europe than in the United States. See the following conversion table.

0.001 nanofarad	1 picofarad	1 pF
0.01 nanofarad	10 picofarads	10 pF
0.1 nanofarad	100 picofarads	100 pF
1 nanofarad	1,000 picofarads	1,000 pF
0.001 microfarad	1 nanofarad	1 nF
0.01 microfarad	10 nanofarads	10 nF
0.1 microfarad	100 nanofarads	100 nF
1 microfarad	1,000 nanofarads	1,000 nF
0.000001 Farad	1 microfarad	1 µF
0.00001 Farad	10 microfarads	10 µF
0.0001 Farad	100 microfarads	100 µF
0.001 Farad	1,000 microfarads	1,000 µF

(You may encounter capacitances greater than 1,000 microfarads, but they are uncommon.)

FUNDAMENTALS

Capacitor basics

DC current does not flow through a capacitor, but voltage can accumulate very quickly inside it, and remains after the power supply is disconnected. Figures 2-67 and 2-68 may help to give you an idea of what happens inside a capacitor when it is fully charged.

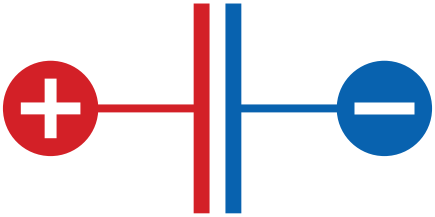

Figure 2-67. *When DC voltage reaches a capacitor, no current flows, but the capacitor charges itself like a little battery. The positive and negative charges are equal and opposite.*

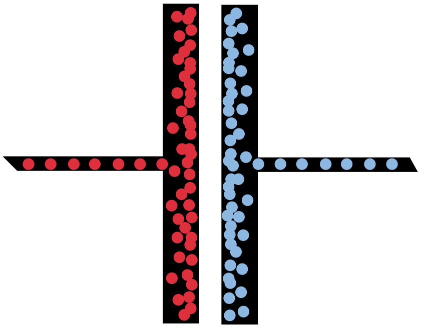

Figure 2-68. *You can imagine positive "charge particles" accumulating on one side of the capacitor and attracting negative "charge particles" to the opposite side.*

In most modern electrolytic capacitors, the plates have been reduced to two strips of very thin, flexible, metallic film, often wrapped around each other, separated by an equally thin insulator. Disc ceramic capacitors typically consist of just a single disc of nonconductive material with metal painted on both sides and leads soldered on.

The two most common varieties of capacitors are ceramic (capable of storing a relatively small charge) and electrolytic (which can be much larger). Ceramics are often disc-shaped and yellow in color; electrolytics are often shaped like miniature tin cans and may be just about any color. Refer back to Figures 2-13 and 2-14 for some examples.

Getting Zapped by Capacitors

If a large capacitor is charged with a high voltage, it can retain that voltage for a long time. Because the circuits in this book use low voltages, you don't have to be concerned about that danger here, but if you are reckless enough to open an old TV set and start digging around inside (which I do not recommend), you may have a nasty surprise. An undischarged capacitor can kill you as easily as if you stick your finger into an electrical outlet. Never touch a large capacitor unless you really know what you're doing.

FUNDAMENTALS

Capacitor basics (continued)

Ceramic capacitors have no polarity, meaning that you can apply negative voltage to either side of them. Electrolytics do have polarity, and won't work unless you connect them the right way around.

The schematic symbol for a capacitor has two significant variants: with two straight lines (symbolizing the plates inside a capacitor), or with one straight line and one curved line, as shown in Figure 2-69. When you see a curved line, that side of the capacitor should be more negative than the other. The schematic symbol may also include a + sign. Unfortunately, some people don't bother to draw a curved plate on a polarized capacitor, yet others draw a curved plate even on a nonpolarized capacitor.

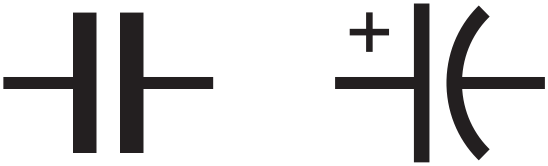

Figure 2-69. *The generic schematic for a capacitor is on the left. The version on the right indicates a polarized capacitor which requires its left plate to be "more positive" than its right plate. The plus sign is often omitted.*

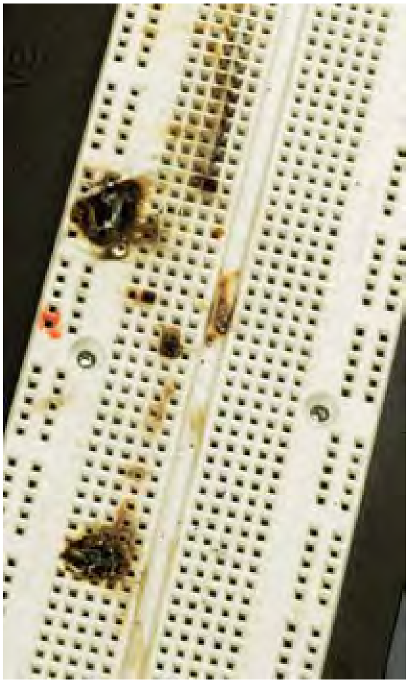

Figure 2-70. *A tantalum capacitor was plugged into this breadboard, accidentally connected the wrong way around to a power source capable of delivering a lot of current. After a minute or so of this abuse, the capacitor rebelled by popping open and scattering small flaming pieces, which burned their way into the plastic of the breadboard. Lesson learned: observe polarity!*

Capacitor Polarity

You must connect an electrolytic capacitor so that its longer wire is more positive than its shorter wire. The shell of the capacitor is usually marked with a negative sign near the shorter wire.

Some capacitors may behave badly if you don't observe their polarity. One time I connected a tantalum capacitor to a circuit, using a power supply able to deliver a lot of current, and was staring at the circuit and wondering why it wasn't working when the capacitor burst open and scattered little flaming fragments of itself in a 3-inch radius. I had forgotten that tantalum capacitors can be fussy about positive and negative connections. Figure 2-70 shows the aftermath.

BACKGROUND

Michael Faraday and capacitors

The earliest capacitors consisted of two metal plates with a very small gap between them. The principle of the thing was simple:

- If one plate was connected to a positive source, the positive charges attracted negative charges onto the other plate.
- If one plate was connected to a negative source, the negative charges attracted positive charges onto the other plate.

Figures 2-67 and 2-68, shown previously, convey the basic idea.

The electrical storage capacity of a capacitor is known as its *capacitance*, and is measured in farads, named after Michael Faraday (Figure 2-71), another of the pantheon of electrical pioneers. He was an English chemist and physicist who lived from 1791 to 1867.

Although Faraday was relatively uneducated and had little knowledge of mathematics, he had an opportunity to read a wide variety of books while working for seven years as a bookbinder's apprentice, and thus was able to educate himself. Also, he lived at a time when relatively simple experiments could reveal fundamental properties of electricity. Thus he made major discoveries including electromagnetic induction, which led to the development of electric motors. He also discovered that magnetism could affect rays of light.

His work earned him numerous honors, and his picture was printed on English 20-pound bank notes from 1991 through 2001.

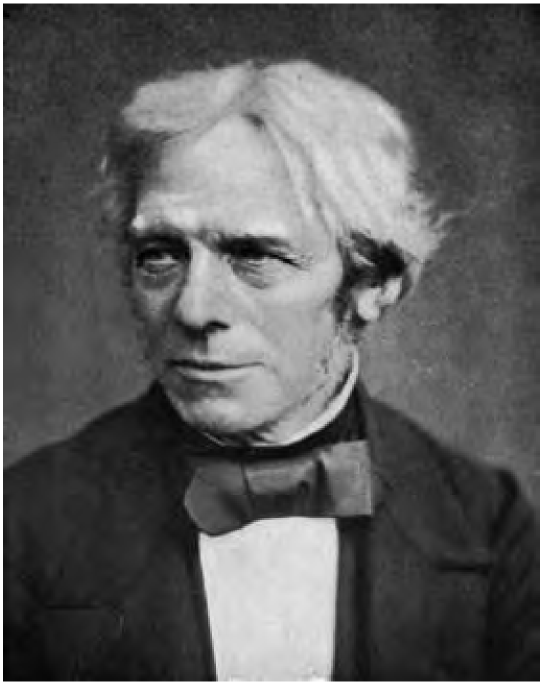

Figure 2-71. *Michael Faraday*

Breadboarding the Circuit

I promised to free you in time from the frustrations of alligator clips, and that time has come. Please turn your attention to the block of plastic with lots of little holes in it that I asked you to buy. For reasons that I do not know, this is called a *breadboard*. When you plug components into the holes, hidden metal strips inside the breadboard connect the components for you, allowing you to set up a circuit, test it, and modify it very easily. Afterward you can pull the components off the breadboard and put them away for future experiments.

Without a doubt, breadboarding is the most convenient way to test something before you decide whether you want to keep it.

Almost all breadboards are designed to be compatible with integrated circuit chips (which we will be using in Chapter 4 of this book). The chip straddles an empty channel in the center of the breadboard with rows of little holes either side—usually five holes per row. You insert other components into these holes.

In addition, the breadboard should have columns of holes running down each side. These are used to distribute positive and negative power.

Take a look at Figures 2-72 and 2-73, which show the upper part of a typical breadboard seen from above, and the same breadboard seen as if with X-ray vision, showing the metal strips that are embedded behind the holes.

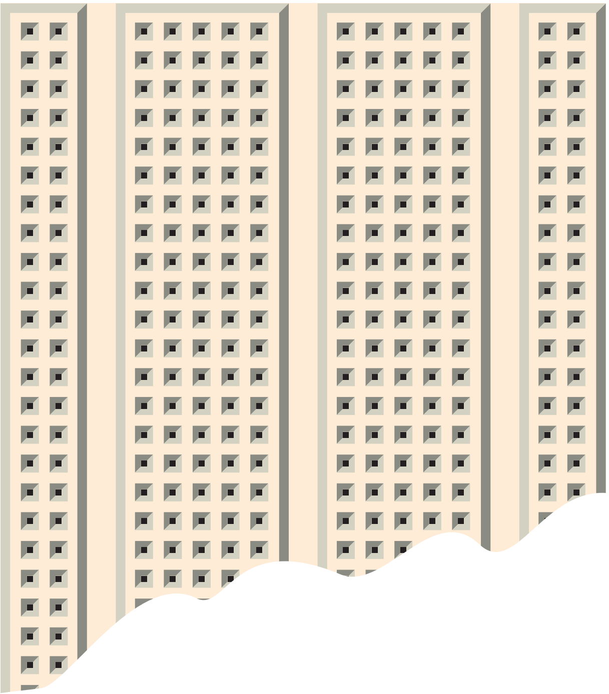

Figure 2-72. *A typical breadboard. You can plug components into the holes to test a circuit very quickly.*

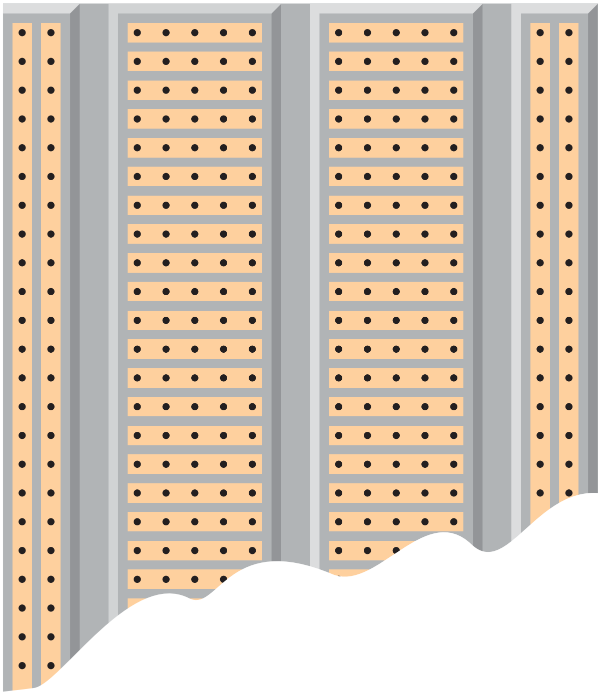

Figure 2-73. *This X-ray-vision view of the breadboard reveals the copper strips that are embedded in it. The strips conduct electricity from one component to another.*

Important note: some breadboards divide each vertical column of holes, on the left and the right, into two separate upper and lower sections. Use your meter's continuity testing feature to find out if your breadboard conducts power along its full length, and add jumper wires to link the upper and lower half of the breadboard if necessary.

Figure 2-74 shows how you can use the breadboard to replicate your oscillating relay circuit. To make this work, you need to apply the positive and negative power from your AC adapter. Because the wire from your AC adapter is almost certainly stranded, you'll have difficulty pushing it into the little holes. A way around this is to set up a couple of pieces of bare 22-gauge wire, and use them as terminals to which you clip the wire from the adapter, as in Figure 2-75. (Yes, you still need just a couple of alligator clips for this purpose.) Alternatively, you can use a breadboard with power terminals built into it, which is more convenient.

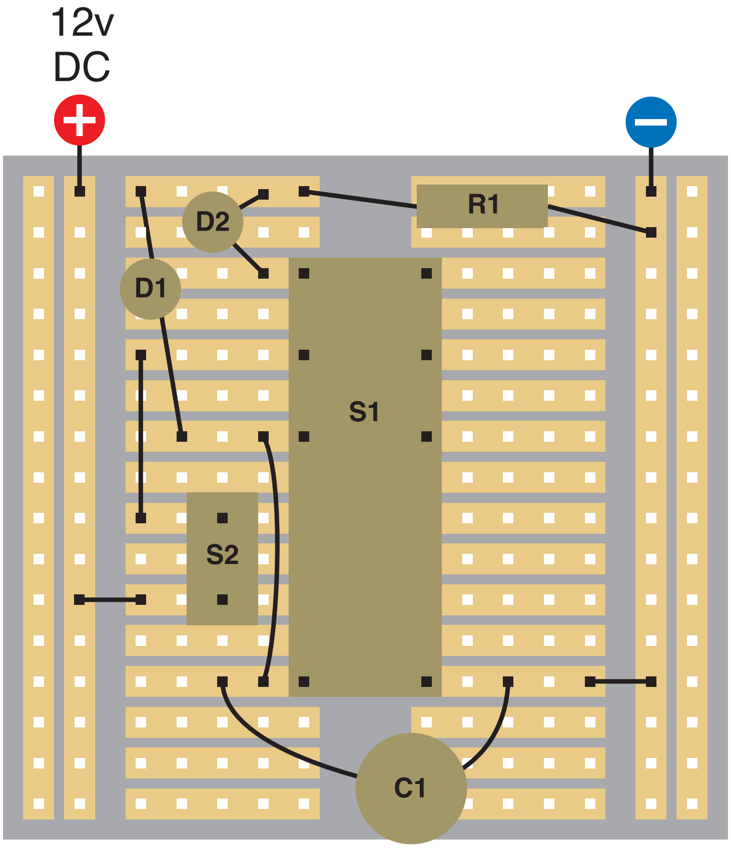

Figure 2-74. *If you place the components on your breadboard in the positions shown, they will create the same circuit that you built from wire and alligator clips in Experiment 8. Component values:*

D1, D2: Light-emitting diodes
S1: DPDT relay
S2: SPST momentary switch
C1: Electrolytic capacitor, 1,000 µF
R1: Resistor, 680Ω minimum

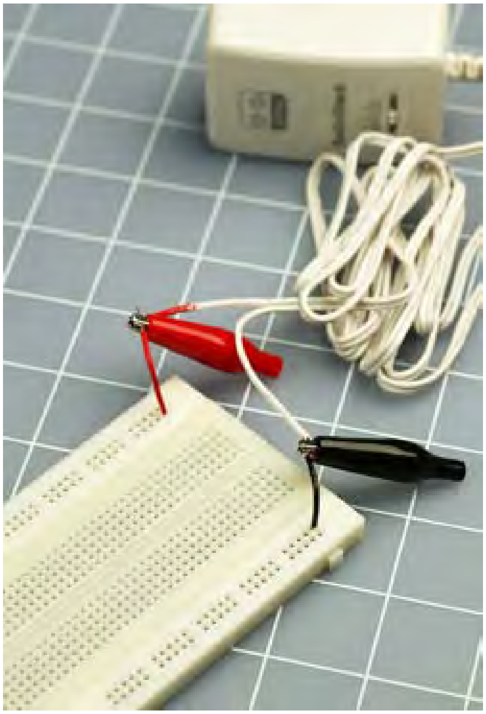

Figure 2-75. *If your breadboard doesn't have screw terminals, insert two short pieces of solid-core wire with stripped ends and then attach the stranded wires from the adapter using alligator clips.*

You'll need some more 22-gauge wire, or some precut hookup wire, to supply the power to your components, which are plugged into the breadboard as shown in Figures 2-76 and 2-77. If you get all the connections right, the circuit should function the same way as before.

The geometry of the metal connecting strips in the breadboard often forces you to connect components in a roundabout way. The pushbutton, for instance, supplies power to the pole of the relay but cannot be connected directly opposite, because there isn't room for it.

Remember that the strips inside the breadboard that don't have any wires or components plugged into them are irrelevant; they don't do anything.

I'll include some suggested breadboard layouts for circuits as you continue through this book, but eventually you'll have to start figuring out breadboard layouts for yourself, as this is an essential part of hobby electronics.

Larger versions of all schematics and breadboard photos are available online at this book's website: http://oreilly.com/catalog/9780596153748.

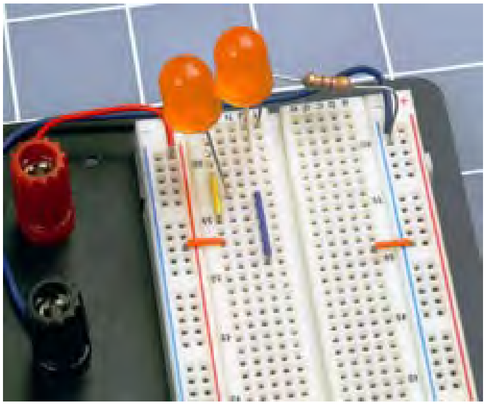

Figure 2-76. *Two oversized LEDs, one resistor, and the necessary jumper wires have been added to the breadboard.*

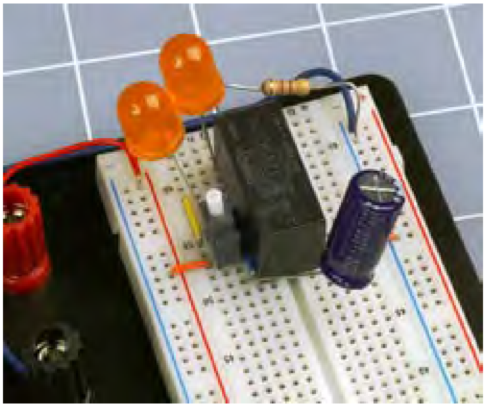

Figure 2-77. *Now the pushbutton, relay, and capacitor have been added to complete the circuit shown in the diagram and the schematic. When the pushbutton is pressed, the relay oscillates and the LEDs flash.*

Experiment 9: Time and Capacitors

You will need:

- AC adapter, breadboard, wire, wire cutters, and strippers.
- Multimeter.
- Pushbutton, SPST. Quantity: 1.
- Resistors and electrolytic capacitors, assorted.

In Experiment 8, when you put a capacitor in parallel with the coil of the relay, the capacitor charged almost instantly before discharging itself through the relay coil. If you add a resistor in series with a capacitor, the capacitor will take longer to charge. By making a capacitor take longer to charge, you can measure time, which is a very important concept.

Clean the components off your breadboard and use it to set up the very simple circuit shown in Figure 2-78, where C1 is a 1,000 μF capacitor, R1 is a 100K resistor, R2 is a 100Ω resistor, and S1 is the pushbutton that you used previously. Set your meter to measure volts DC, place the probes around the capacitor, and hold down the pushbutton. You should see the meter counting upward as the voltage accumulates on the capacitor. (This is easier with a meter that doesn't have autoranging, because you won't have to wait while the meter figures out which range to apply.) Resistor R1 slows the charging time for the capacitor.

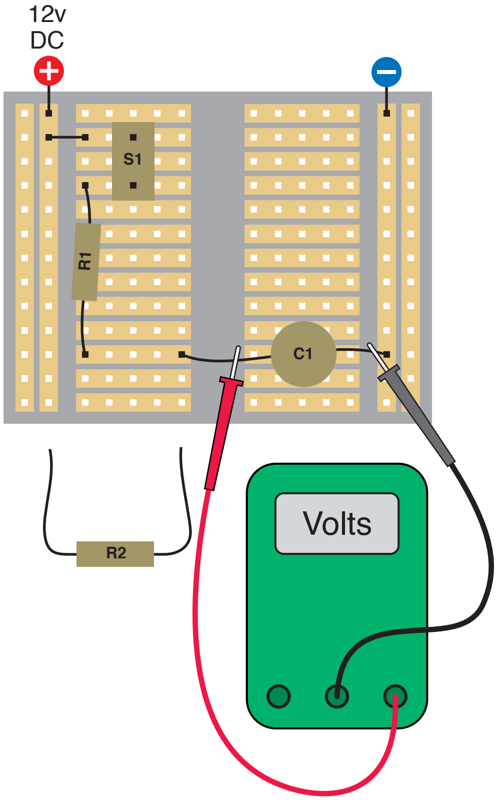

Figure 2-78. *Watch the voltage building up on the capacitor while you hold down the pushbutton. Substitute other values for R1, discharge the capacitor by touching R2 across it, and then repeat your measurement process.*

S1: Momentary pushbutton, OFF (ON)
R1: 100K initially
R2: 100Ω
C1: 1,000 μF

Release the pushbutton, set aside your meter, and discharge the capacitor by touching R2 across it for a second or two. Now substitute a 50K resistor for R1, and repeat the measurement. The meter should count upward almost twice as fast as before.

Voltage, Resistance, and Capacitance

Think of the resistor as a faucet, and the capacitor as a balloon that you are trying to fill with water. When you screw down the faucet until only a trickle comes through, the balloon will take longer to fill. But a slow flow of water will still fill the balloon completely if you wait long enough, and (assuming the balloon doesn't burst) the process ends when the pressure inside the balloon is equal to the water pressure in the pipe supplying the faucet. See Figure 2-79.

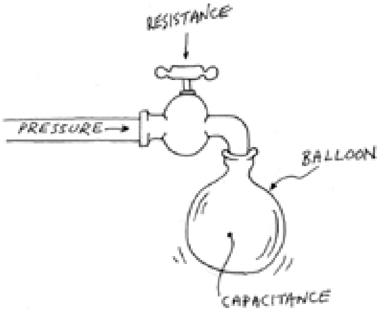

Figure 2-79. *When the faucet is closed half-way, the balloon will take longer to fill, but will still contain as much water and as much pressure in the end.*

Similarly, in your circuit, if you wait long enough, eventually the voltage across the capacitor should reach the same value as the voltage of the power supply. In a 12-volt circuit, the capacitor should eventually acquire 12 volts (although "eventually" may take longer than you think).

This may seem confusing, because earlier you learned that when you apply voltage at one end of a resistor, you get less voltage coming out than you have going in. Why should a resistor deliver the full voltage when it is paired with a capacitor?

Forget the capacitor for a moment, and remember how you tested just two 1K resistors. In that situation, each resistor contained half the total resistance of the circuit, so each resistor dropped half the voltage. If you held the negative probe of your meter against the negative side of your power supply and touched the positive probe to the center point between the two resistors, you would measure 6 volts. Figure 2-80 illustrates this.

Now, suppose you remove one of the 1K resistors and substitute a 9K resistor. The total resistance in the circuit is now 10K, and therefore the 9K resistor drops 90% of the 12 volts. That's 10.8 volts. You should try this and check it with your meter. (You are unlikely to find a 9K resistor, because this is not a standard value. Substitute the nearest value you can find.)

Now suppose you remove the 9K resistor and substitute a 99K resistor. Its voltage drop will be 99% of the available voltage, or 11.88 volts. You can see where this is heading: the larger the resistor, the larger its contribution to voltage drop.

However, I noted previously that a capacitor blocks DC voltage completely. It can accumulate an electrical charge, but no current passes through it. Therefore, a capacitor behaves like a resistor that has infinite resistance to DC current.

(Actually the insulation inside the capacitor allows a little bit of "leakage," but a perfect capacitor would have infinite resistance.)

The value of any resistor that you put in series with the capacitor is trivial by comparison. No matter how high the value of the resistor is, the capacitor still provides much more resistance in the circuit. This means that the capacitor steals almost the complete voltage drop in the circuit, and the voltage difference between one end of the resistor and the other will be zero (assuming that we ignore little imperfections in the components). Figure 2-80 may help to clarify this concept.

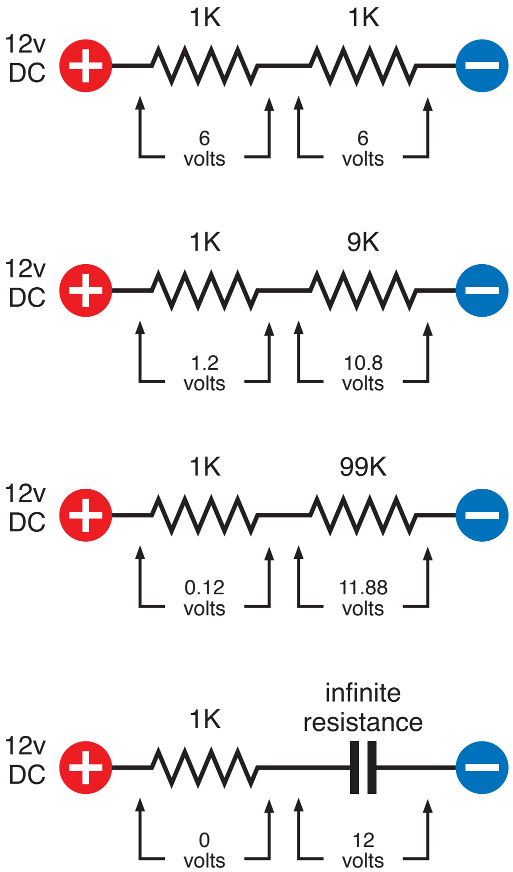

Figure 2-80. *When two resistances are in series, the larger one drops the voltage more than the smaller one. If the larger resistance becomes infinite (as in the case of a capacitor), the smaller one no longer has any measurable contribution to the voltage drop, and the voltage is almost exactly the same at both ends.*

You should try this using real resistors and capacitors—although if you do, you will run into a little problem. When you use your meter in its "DC volts" mode, it diverts a little of the current in the circuit—just a tiny taste—in the process of measuring it. The meter steals such a small amount, it doesn't affect the reading significantly when you are checking voltage across a resistor. The internal resistance of the meter is higher than the values of most resistors. However, remember that the internal resistance of a capacitor is almost infinite. Now the internal resistance of the meter becomes significant. Because you can never have an ideal meter, any more than you can have an ideal capacitor or resistor, your meter will always interfere with the circuit slightly, and you will get only an approximate indication.

If you try to measure the voltage on a capacitor that has been charged but is now not connected to anything else, you'll see the number slowly falling, as the capacitor discharges itself through the meter.

THEORY

The time constant

You may be wondering if there's a way to predict exactly how much time it takes for various capacitors to charge, when they are paired with various resistors. Is there a formula to calculate this?

Of course, the answer is yes, but the way we measure it is a bit tricky, because a capacitor doesn't charge at a constant rate. It accumulates the first volt very quickly, the second volt not quite as quickly, the third volt even less quickly—and so on. You can imagine the electrons accumulating on the plate of a capacitor like people walking into an auditorium and looking for a place to sit. The fewer seats that are left, the longer people take to find them.

The way we describe this is with something called a "time constant." The definition is very simple:

TC = R × C

where TC is the time constant, and a capacitor of C farads is being charged through a resistor of R ohms.

Going back to the circuit you just tested, try using it again, this time with a 1K resistor and the 1,000 µF capacitor. We have to change those numbers to farads and ohms before we can put them in the formula. Well, 1,000 µF is 0.001 farads, and 1K is 1,000 ohms, so the formula looks like this:

TC = 1,000 × 0.001

In other words, TC = 1—a lesson that could not be much easier to remember:

> ***A 1K resistor in series with a 1,000 µF capacitor has a time constant of 1.***

Does this mean that the capacitor will be fully charged in 1 second? No, it's not that simple. TC, the time constant, is the time it takes for a capacitor to acquire *63%* of the voltage being supplied to it, if it starts with zero volts.

(Why 63%? The answer to that question is too complicated for this book, and you'll have to read about time constants elsewhere if you want to know more. Be prepared for differential equations.) Here's a formal definition for future reference:

> TC, the time constant, is the time it takes for a capacitor to acquire 63% of the difference between its current charge and the voltage being applied to it. When TC=1, the capacitor acquires 63% of its full charge in 1 second. When TC=2, the capacitor acquires 63% of its full charge in 2 seconds. And so on.

What happens if you continue to apply the voltage? History repeats itself. The capacitor accumulates *another* 63% of the *remaining* difference between its current charge, and the voltage being applied to it.

Imagine someone eating a cake. In his first bite he's ravenously hungry, and eats 63% of the cake in one second. In his second bite, not wanting to seem too greedy, he takes just another 63% of the cake that is left—and because he's not feeling so hungry anymore, he requires the same time to eat it as he took to eat the first bite. In his third bite, he takes 63% of what still remains, and still takes the same amount of time. And so on. He is behaving like a capacitor eating electricity (Figure 2-81).

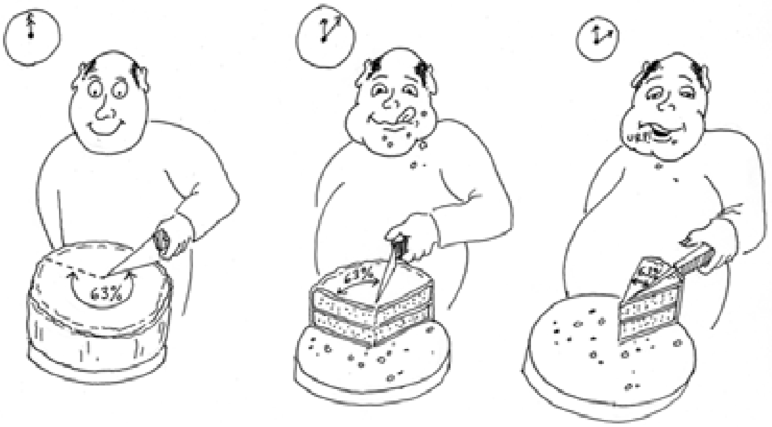

Figure 2-81. *If our gourmet always eats just 63% of the cake still on the plate, he "charges up" his stomach in the same way that a capacitor charges itself. No matter how long he keeps at it, his stomach is never completely filled.*

THEORY

The time constant (continued)

The cake eater will always have a few crumbs to eat, because he never takes 100% of the remainder. Likewise, the capacitor will never acquire a full charge. In a perfect world of perfect components, this process would continue for an infinite time.

In the real world, we say rather arbitrarily:

> After 5 × TC the capacitor will be so nearly fully charged, we won't care about the difference.

In the table is a calculation (rounded to two decimal places) showing the charge accumulating on a capacitor in a 12-volt circuit where the time constant is 1 second.

Here's how to understand the table. V1 is the current charge on the capacitor. Subtract this from the supply voltage (12 volts) to find the difference. Call the result V2. Now take 63% of V2, and add this to the current charge (V1) and call the result V4. This is the new charge that the capacitor will have after 1 second, so we copy it down to the next line in the table, and it becomes the new value for V1.

Now we repeat the same process all over again. Figure 2-82 shows this in graphical form. Note that after 5 seconds, the capacitor has acquired 11.92 volts, which is 99% of the power supply voltage. This should be close enough to satisfy anyone's real-world requirements.

If you try to verify these numbers by measuring the voltage across the capacitor as it charges, remember that because your meter steals a little current, there will be a small discrepancy that will increase as time passes. For practical purposes, the system works well enough.

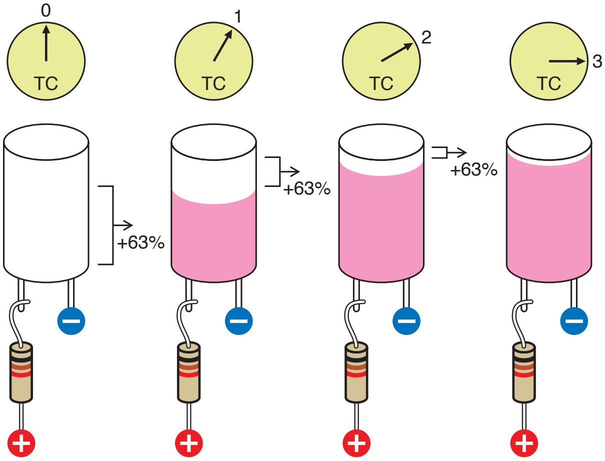

Figure 2-82. *A capacitor starts with 0 volts. After 1 time constant it adds 63% of the available voltage. After another time constant, it adds another 63% of the remaining voltage difference, and so on.*

Time in secs	V1: Charge on capacitor	V2: 12 – V1	V3: 63% of V2	V4: V1 + V3
0	0.00	12.00	7.56	7.56
1	7.56	4.44	2.80	10.36
2	10.36	1.64	1.03	11.39
3	11.39	0.61	0.38	11.77
4	11.77	0.23	0.15	11.92
5	11.92			

Experiment 10: Transistor Switching

You will need:

- AC adapter, breadboard, wire, and meter.
- LED. Quantity: 1.
- Resistors, various.
- Pushbutton, SPST. Quantity: 1.
- Transistor, 2N2222 or similar. Quantity: 1.

A transistor can switch a flow of electricity, just like a relay. But it's much more sensitive and versatile, as this first ultra-simple experiment will show.

We'll start with the 2N2222 transistor, which is the most widely used semiconductor of all time (it was introduced by Motorola in 1962 and has been in production ever since).

First, you should get acquainted with the transistor. Because Motorola's patents on the 2N2222 ran out long ago, any company can manufacture their own version of it. Some versions are packaged in a little piece of black plastic; others are enclosed in a little metal "can." (See Figure 2-83.) Either way, it contains a piece of silicon divided into three sections known as the collector, the base, and the emitter. I'll describe their function in more detail in a moment, but initially you just need to know that in this type of transistor, the collector receives current, the base controls it, and the emitter sends it out.

Use your breadboard to set up the circuit shown in Figure 2-85. Be careful to get the transistor the right way around! (See Figure 2-84.) For the three brands I have mentioned in the shopping list, the flat side should face right, if the transistor is packaged in black plastic, or the little tab should face toward the lower left, if the transistor is packaged in metal.

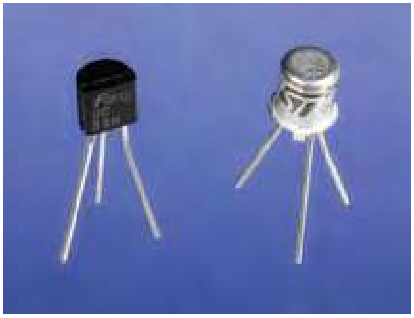

Figure 2-83. *A typical transistor is packaged either in a little metal can or a molded piece of black plastic. The manufacturer's data sheet tells you the identities of the three wire leads, relative to the flat side of a black plastic transistor or the tab that sticks out of a metal-can transistor.*

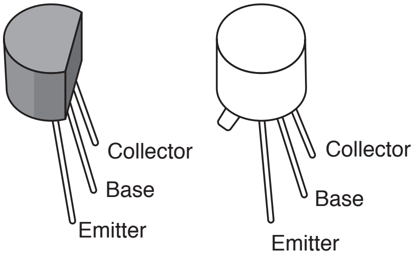

Figure 2-84. *The 2N2222 transistor may be packaged in either of these formats. Left: RadioShack or Fairchild. Right: STMicroelectronics (note the little tab sticking out at the lower-left side). If you use a different brand, you'll have to check the manufacturer's data sheet. Insert the transistor in your breadboard with the flat side facing right, as seen from above, or the tab pointing down and to the left, seen from above.*

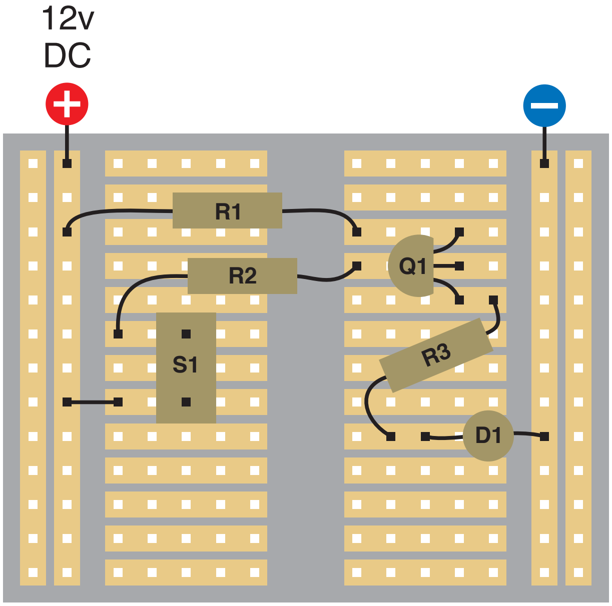

Figure 2-85. *The transistor blocks voltage that reaches it through R1. But when pushbutton S1 is pressed, this tells the transistor to allow current to pass through it. Note that transistors are always identified with letter Q in wiring diagrams and schematics.*

S1: Pushbutton, momentary, OFF (ON)
R1: 180Ω
R2: 10K
R3: 680Ω
Q1: 2N2222 or similar
D1: LED

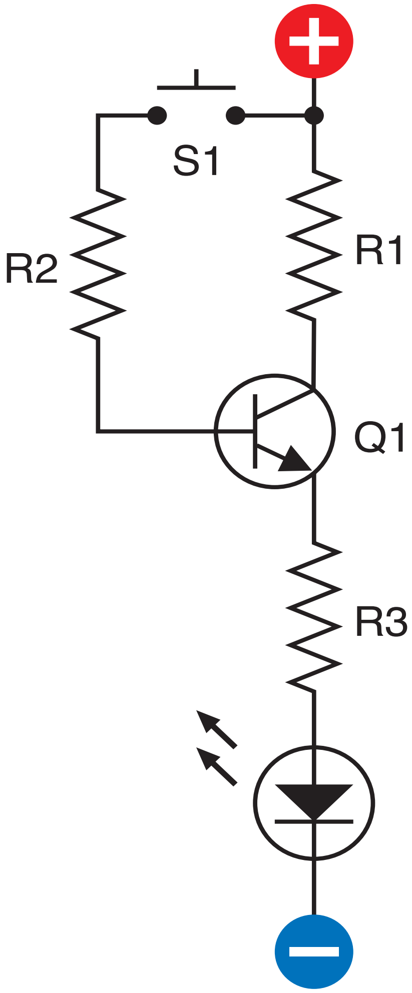

Figure 2-86. *This shows the same circuit as the breadboard diagram in Figure 2-85.*

Initially, the LED should be dark. Now press the pushbutton and the LED should glow brightly. Electricity is following two paths here. Look at the schematic in Figure 2-86, which shows the same circuit more clearly. I've shown positive at the top and negative at the bottom (the way most schematics do it) because it helps to clarify the function of this particular circuit. If you view the schematic from the side, the similarity with the breadboard layout is easier to see.

Through R1, voltage reaches the top pin (the collector) of the transistor. The transistor only lets a tiny trickle of it pass through, so the LED stays dark. When you press the button, voltage is also applied along a separate path, through R2 to the middle pin (the base) of the transistor. This tells the transistor to open its solid-state switch and allow current to flow out through its third pin (the emitter), and through R3, to the LED.

You can use your meter in volts DC mode to check the voltage at points in the circuit. Keep the negative probe from the meter touching the negative voltage source while you touch the positive probe to the top pin of the transistor, the middle pin, and the bottom pin. When you press the button, you should see the voltage change.

Fingertip Switching

Now here's something more remarkable. Remove R2 and the pushbutton, and insert two short pieces of of wire as shown in Figure 2-87. The upper piece of wire connects with the positive voltage supply; the lower piece connects with the middle pin of the transistor (its base). Now touch the tip of your finger to the two wires. Once again, the LED should glow, although not as brightly as before. Lick the tip of your finger, try again, and the LED should glow more brightly.

Never Use Two Hands

The fingertip switching demo is safe if the electricity passes just through your finger. You won't even feel it, because it's 12 volts DC from a power supply of 1 amp or less. But it's not a good idea to put the finger of one hand on one wire, and the finger of your other hand on the other wire. This would allow the electricity to pass through your body. Although the chance of hurting yourself this way is extremely small, you should never allow electricity to run through you from one hand to the other. Also, when touching the wires, don't allow them to penetrate your skin.

Your finger is conducting positive voltage to the base of the transistor. Even though your skin has a high resistance, the transistor still responds. It isn't just switching the LED on and off; it is *amplifying the current* applied to its base. This is an essential concept: *a transistor amplifies any changes in current that you apply to its base*.

Check Figure 2-88 to see more clearly what's happening.

If you studied the section "Background: Positive and negative" in Chapter 1, you learned that there is really no such thing as positive voltage. All we really have is negative voltage (created by the pressure of free electrons) and an absence of negative voltage (where there are fewer free electrons). But because the idea of a flow of electricity from positive to negative was so widely believed before the electron was discovered, and because the inner workings of a transistor involve "holes" which are an absence of electrons and can be thought of as positive, we can still pretend that electricity flows from positive to negative. See the following section, "Essentials: All about NPN and PNP transistors," for more details.

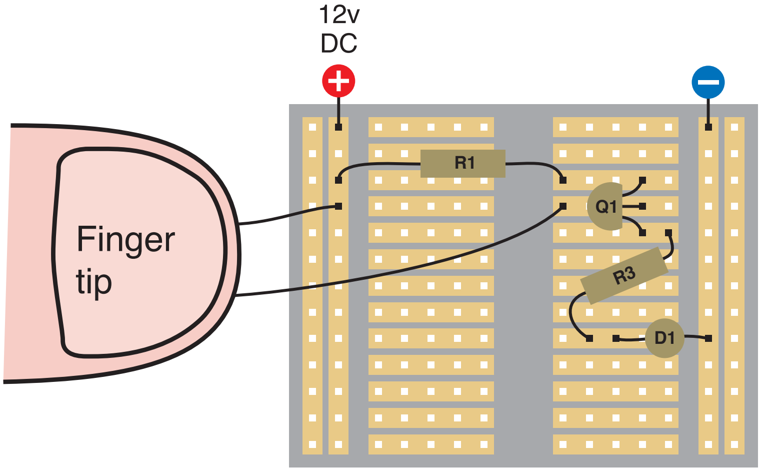

Figure 2-87

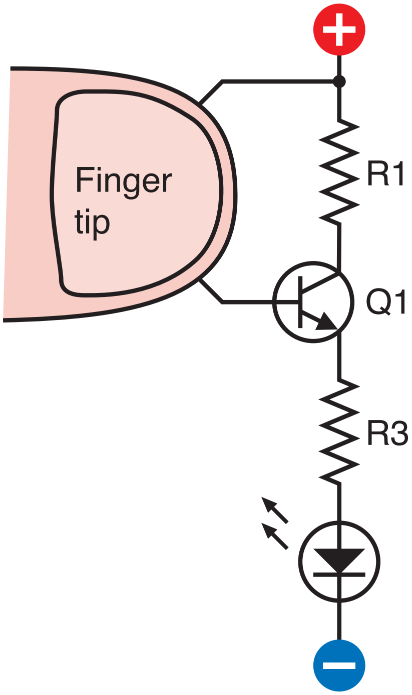

Figure 2-88. *These two diagrams show the same components as before, with a fingertip substituted for R2. Although only a trickle of voltage now reaches the base of the transistor, it's enough to make the transistor respond.*

ESSENTIALS

All about NPN and PNP transistors

A transistor is a semiconductor, meaning that sometimes it conducts electricity, and sometimes it doesn't. Its internal resistance varies, depending on the power that you apply to its base.

NPN and PNP transistors are bipolar semiconductors. They contain two slightly different variants of silicon, and conduct using both polarities of carriers—holes and electrons.

The NPN type is a sandwich with P-type silicon in the middle, and the PNP type is a sandwich with N-type silicon in the middle. If you want to know more about this terminology, and the behavior of electrons when they try to cross an NP junction or a PN junction, you'll have to read a separate source on this subject. It's too technical for this book. All you need to remember is:

- All bipolar transistors have three connections: Collector, Base, and Emitter, abbreviated as C, B, and E on the manufacturer's data sheet, which will identify the pins for you.
- NPN transistors are activated by *positive* voltage on the base relative to the emitter.
- PNP transistors are activated by *negative* voltage on the base relative to the emitter.

In their passive state, both types block the flow of electricity between the collector and emitter, just like an SPST relay in which the contacts are normally open. (Actually a transistor allows a tiny bit of current known as "leakage.")

You can think of a bipolar transistor as if it contains a little button inside, as shown in Figures 2-89 and 2-90. When the button is pressed, it allows a large current to flow. To press the button, you inject a much smaller current into the base by applying a small voltage to the base. In an NPN transistor, the control voltage is positive. In a PNP transistor, the control voltage is negative.

NPN transistor basics

- To start the flow of current from collector to emitter, apply a relatively positive voltage to the base.
- In the schematic symbol, the arrow points from base to emitter and shows the direction of positive current.
- The base must be at least 0.6 volts "more positive" than the emitter, to start the flow.
- The collector must be "more positive" than the emitter.

PNP transistor basics

- To start the flow of current from emitter to collector, apply a relatively negative voltage to the base.
- In the schematic symbol, the arrow points from emitter to base and shows the direction of positive current.
- The base must be at least 0.6 volts "more negative" than the emitter, to start the flow.
- The emitter must be "more positive" than the collector.

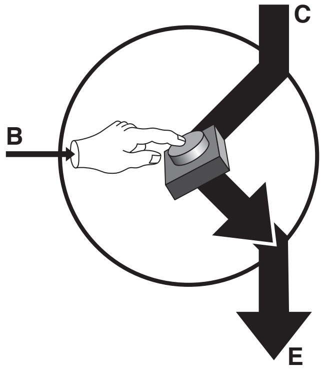

Figure 2-89. *You can think of a bipolar transistor as if it contains a button that can connect the collector and the emitter. In an NPN transistor, a small positive potential presses the button.*

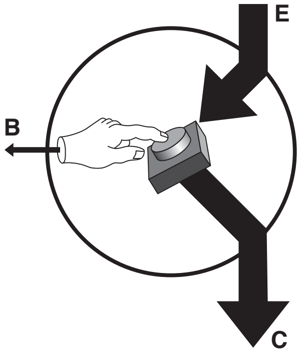

Figure 2-90. *In a PNP transistor, a small negative potential has the same effect. The arrows point in the direction of "positive current flow."*

ESSENTIALS

All about NPN and PNP transistors (continued)

All-transistor basics

- Never apply a power supply directly across a transistor. You can burn it out with too much current.
- Protect a transistor with a resistor, in the same way you would protect an LED.
- Avoid reversing the connection of a transistor between positive and negative voltages.
- Sometimes an NPN transistor is more convenient in a circuit; sometimes a PNP happens to fit more easily. They both function as switches and amplifiers, the only difference being that you apply a relatively positive voltage to the base of an NPN transistor, and a relatively negative voltage to the base of a PNP transistor.
- PNP transistors are used relatively seldom, mainly because they were more difficult to manufacture in the early days of semiconductors. People got into the habit of designing circuits around NPN transistors.
- Remember that bipolar transistors amplify current, not voltage. A small fluctuation of current through the base enables a large change in current between emitter and collector.
- Schematics sometimes show transistors with circles around them, and sometimes don't. In this book, I'll use circles to draw attention to them. See Figures 2-91 and 2-92.
- Schematics may show the emitter at the top and the collector at the bottom, or vice versa. The base may be on the left, or on the right, depending on what was most convenient for the person drawing the schematic. Be careful to look carefully at the arrow in the transistor to see which way up it is, and whether it is NPN or PNP. You can damage a transistor by connecting it incorrectly.
- Transistors come in various different sizes and configurations. In many of them, there is no way to tell which wires connect to the emitter, the collector, or the base, and some transistors have no part numbers on them. Before you throw away the packaging that came with a transistor, check to see whether it identifies the terminals.
- If you forget which wire is which, some multimeters have a function that will identify emitter, collector, and base for you. Check your multimeter instruction booklet for more details.

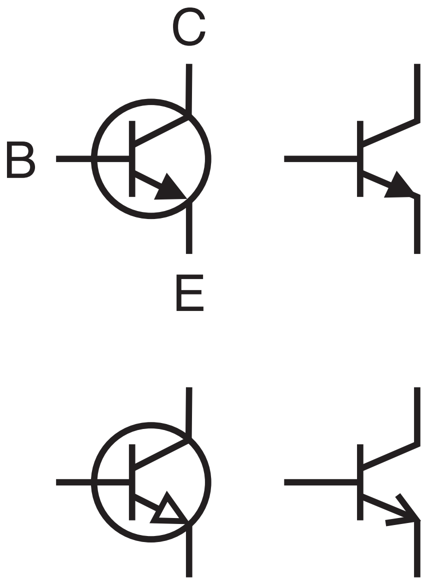

Figure 2-91. *The symbol for an NPN transistor always has an arrow pointing from its base to its emitter. Some people include a circle around the transistor; others don't bother. The style of the arrow may vary. But the meaning is always the same. The top-left version is the one I use in this book.*

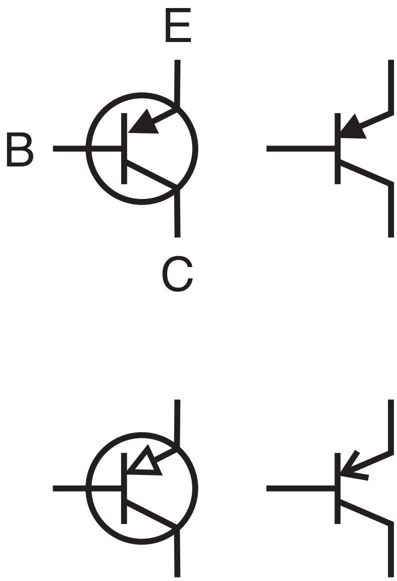

Figure 2-92. *The symbol for a PNP transistor always has an arrow pointing from its emitter to its base. Some people include a circle around the transistor; others don't bother. The style of the arrow may vary. But the meaning is always the same. The top-left version is the one I use in this book.*

BACKGROUND

Transistor origins

Though some historians trace the origins of the transistor back to the invention of diodes (which allow electricity to flow in one direction while preventing reversal of the flow), there's no dispute that the first working transistor was developed at Bell Laboratories in 1948 by John Bardeen, William Shockley, and Walter Brattain (Figure 2-93).

Shockley was the leader of the team, who had the foresight to see how potentially important a solid-state switch could be. Bardeen was the theorist, and Brattain actually made it work. This was a hugely productive collaboration—until it succeeded. At that point, Shockley started maneuvering to have the transistor patented exclusively under his own name. When he notified his collaborators, they were—naturally—unhappy about this idea.

A widely circulated publicity photograph didn't help, in that it showed Shockley sitting at the center in front of a microscope, as if he had done the hands-on work, while the other two stood behind him, implying that they had played a lesser role. In fact Shockley, as the supervisor, was seldom present in the laboratory where the real work was done.

The productive collaboration quickly disintegrated. Brattain asked to be transferred to a different lab at AT&T. Bardeen moved to the University of Illinois to pursue theoretical physics. Shockley eventually left Bell Labs and founded Shockley Semiconductor in what was later to become Silicon Valley, but his ambitions outstripped the capabilities of the technology of his time. His company never manufactured a profitable product.

Eight of Shockley's coworkers in his company eventually betrayed him by quitting and establishing their own business, Fairchild Semiconductor, which became hugely successful as a manufacturer of transistors and, later, integrated circuit chips.

Figure 2-93. *Photographs provided by the Nobel Foundation show, left to right, John Bardeen, William Shockley, and Walter Brattain. For their collaboration in development of the world's first working transistor in 1948, they shared a Nobel prize in 1956.*

ESSENTIALS

Transistors and relays

One limitation of NPN and PNP transistors is that they are naturally "off" until you turn them "on." They behave like a normally open pushbutton, which conducts electricity only for as long as you hold it down. They don't normally behave like a normal on switch, which stays on until you apply a signal to turn it off.

A relay offers more switching options. It can be normally open, normally closed, or it can contain a double-throw switch, which gives you a choice of two "on" positions. It can also contain a double-pole switch, which makes (or breaks) two entirely separate connections when you energize it. Single-transistor devices cannot provide the double-throw or double-pole features, although you can design more complex circuits that emulate this behavior.

Here's a list of transistor and relay characteristics.

	Transistor	Relay
Long-term reliability	Excellent	Limited
Configurable for DP and DT switching	No	Yes
Ability to switch large currents	Limited	Good
Able to switch alternating current (AC)	Usually not	Yes
Can be triggered by alternating current (AC)	Usually not	Optional
Suitability for miniaturization	Excellent	Very limited
Sensitive to heat	High	Moderate
Ability to switch at high speed	Excellent	Limited
Price advantage for low-voltage low-current	Yes	No
Price advantage for high-voltage high-current	No	Yes
Current leakage when "off"	Yes	No

The choice between relays or transistors will depend on each particular application.

THEORY

See the current

If you want to get a more precise understanding of how a transistor works, you should try this little test. It shows the precise behavior and limits of the 2N2222 transistor that you used in the previous experiment.

I've said that in an NPN transistor, the collector should always be more positive than the emitter and that the base should have a potential somewhere between those two voltages. Figure 2-94 shows this rather vague relationship. Now I want to substitute some numbers for these general statements.

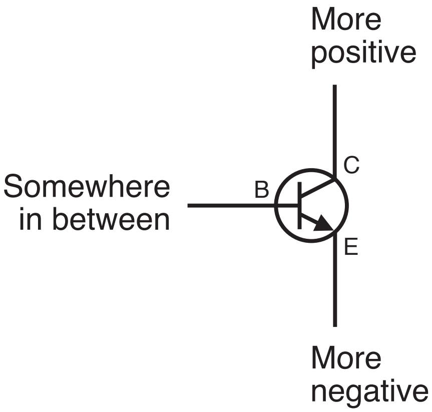

Figure 2-94. *The proper functioning of an NPN transistor requires you to maintain these voltage relationships.*

Take a look at the schematic in Figure 2-95, and check the component values. Notice that the total resistance above the transistor, from R1 + R2, is the same as the total resistance below it, from R3 + R4. Therefore the potential on the base of the transistor should be halfway between the two extremes—until you use potentiometer P1 to adjust the voltage of the base of the transistor up and down.

The two 180Ω resistors, R1 and R3, protect the transistor from passing excessive current. The two 10K resistors, R2 and R4, protect the base when the potentiometer is turned all the way up or all the way down.

I would like you to see what the transistor is doing by measuring the amperage flowing into the base at the position marked A1, and the total amperage flowing out through the emitter at the position marked A2. To do this, it would be really helpful if you had two meters. As that may be impractical, the breadboard diagrams in Figures 2-96 and 2-97 show how you can swap one meter between the two locations.

Remember that to measure milliamps, you have to pass electricity through the meter. This means that the meter must be inserted into the circuit, and whenever you remove the meter, you have to remake the connection where the meter was. The breadboard diagram shows how you can do this. Fortunately, it's very easy to remove and replace wires in a breadboard. Where wires are connected to the potentiometer, you may need to revert to using alligator clips.

Begin with the potentiometer turned about halfway through its range. Measure at A1 and A2. Turn the potentiometer up a bit, and measure current at the two locations again. Following is a table showing some actual readings I obtained at those two locations, using two digital meters simultaneously.

Milliamps passing through location A1	Milliamps passing through location A2
0.01	1.9
0.02	4.9
0.03	7.1
0.04	9.9
0.05	12.9
0.06	15.5
0.07	17.9
0.08	19.8
0.09	22.1
0.10	24.9
0.11	26.0
0.12	28.3

There's a very obvious relationship. The current emerging from the emitter of the transistor, through location A2, is about 24 times the current passing through location A1, into the base. The ratio of current coming out from the emitter of an NPN transistor to current going into the base is known as the *beta value* for a transistor. The beta value expresses the transistor's amplifying power.

It's a very constant ratio, until you push it a little too far. Above 0.12 mA, this particular transistor becomes "saturated," meaning that its internal resistance cannot go any lower.

THEORY

See the current (continued)

In my little experiment, I found that the maximum current at A2 was 33mA. A simple calculation using Ohm's Law showed me that this meant the transistor's internal resistance was near zero. This is why you should protect a transistor with some additional resistance in the circuit. If you don't, its low internal resistance would allow a huge current flow that would immediately burn it out.

What about the other end of its range? When it passes only 1.9 mA, the transistor has an internal resistance of around 6,000Ω. The conclusion is that depending how much current you apply to this transistor, its internal resistance varies between zero and 6,000Ω, approximately.

So much for the theory. Now what can we do with a transistor that's fun, or useful, or both? We can do Experiment 11!

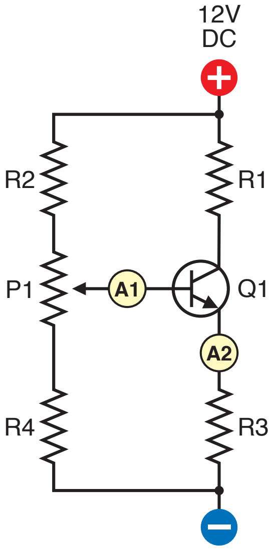

Figure 2-95. *This is basically the same as the previous circuit, with a potentiometer added and the LED removed. Component values:*

R1: 180Ω
R2: 10K
R3: 180Ω
R4: 10K
P1: 1M linear potentiometer
Q1: 2N2222 transistor

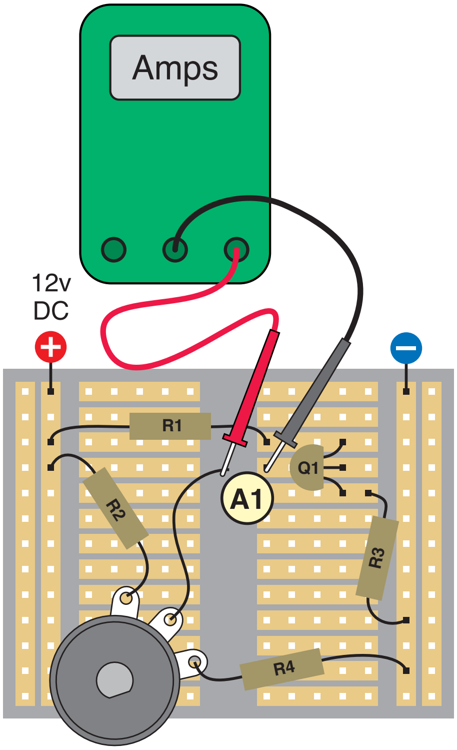

Figure 2-96. *The meter is measuring current flowing from the potentiometer into the base of the transistor at position A1 (see Figure 2-95).*

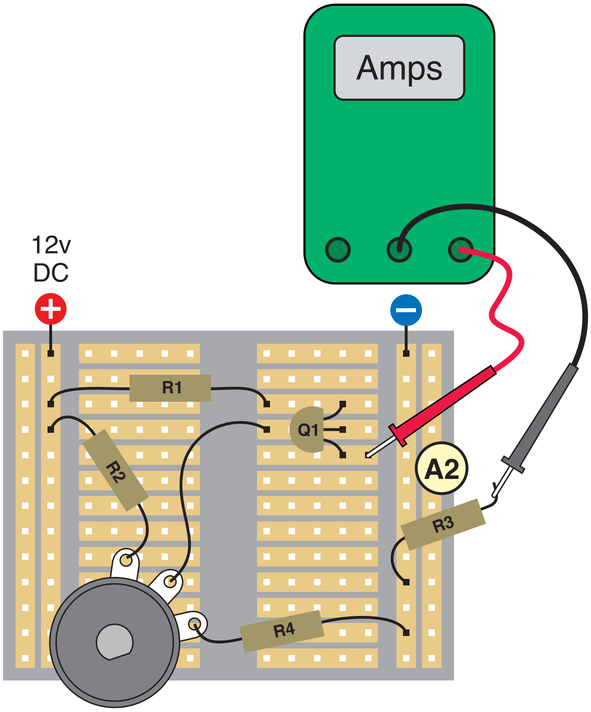

Figure 2-97. *One end of resistor R3 has been unplugged from the breadboard so that the meter now measures current flowing out through the emitter of the transistor, into R3, at position A2.*

Experiment 11: A Modular Project

You will need:

- AC adapter, breadboard, wire, and meter.
- LED. Quantity: 1.
- Resistors, various.
- Capacitors, various.
- Transistor, 2N2222 or similar. Quantity: 2.
- 2N6027 programmable unijunction transistor (PUT). Quantity: 2.
- Miniature 8Ω loudspeaker. Quantity: 1.

So far, I've described small circuits that perform very simple functions. Now it's time to show how modules can be combined to create a device that does a bit more.

The end product of this experiment will be a circuit that makes a noise like a small siren, which could be used in an intrusion alarm. You may or may not be interested in owning an alarm, but the four-step process of developing it is important, because it shows how individual clusters of components can be persuaded to communicate with each other.

I'll begin by showing how to use a transistor to make a solid-state version of the oscillating circuit that you built with a relay in Experiment 8. The relay, you may remember, was wired in such a way that the coil received power through the contacts of the relay. As soon as the coil was energized, it opened the contacts, thus cutting off its own power. As soon as the contacts relaxed they restored the power, and the process repeated itself.

There's no way to do this with a single bipolar transistor. You actually need two of them, switching each other on and off, and the way that this works is quite hard to understand. An easier option is to use a different thing known as a programmable unijunction transistor, or PUT.

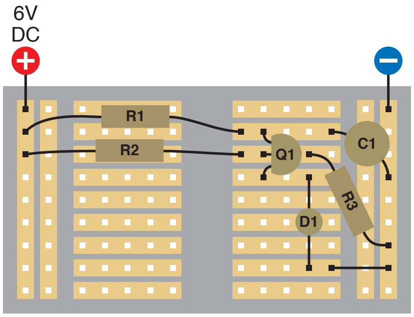

Figure 2-98. *Assemble these components, apply power, and the LED should start flashing.*

R1: 470K
R2: 15K
R3: 27K
C1: 2.2 μF electrolytic capacitor
D1: LED
Q1: 2N6027 programmable unijunction transistor

Unijunction transistors were developed during the 1950s, but fell into disuse when simple silicon chips acquired the ability to perform the same kinds of functions, more accurately and more cheaply. However, the so-called programmable unijunction transistor is still widely available, often used in applications such as lamp dimmers and motor controllers. Because its primary use is in generating a stream of pulses, it's ideal for our purposes.

If you put together the components shown in Figure 2-98, the LED should start flashing as soon as you apply power.

Note that this circuit will work on 6 volts. You won't damage anything if you run it with 12 volts, but as we continue adding pieces to it, you'll find that it actually performs better at 6 volts than at 12. If you read the next section, "Essentials: All about programmable unijunction transistors," you'll find out how the circuit works.

ESSENTIALS

All about programmable unijunction transistors

The schematic symbol for a programmable unijunction transistor, or PUT, looks very different from the symbol for a bipolar transistor, and its parts are named differently, too. Nevertheless, it does have a similar function as a solid-state switch. The symbol and the names of the three connections are shown in Figure 2-99.

Note that this is a rare case (maybe the only one in the whole of electronics!) in which you won't run into confusing variations of the basic schematic symbol. A PUT always seems to look the way I've drawn it here. Personally I think it would be clearer if we added a circle around it, but no one seems to do that, so I won't, either.

The 2N6027 is probably the most common PUT, and seems to be standardized in its packaging and pin-outs. I've only seen it in a plastic module rather than a little tin can. Figure 2-100 shows the functions of the leads if your 2N6027 is manufactured by Motorola or On Semiconductor. If you have one from another source, you should check the data sheet.

Note that the flat side of the plastic module faces the opposite way around compared with the 2N2222 bipolar transistor, when the two devices are functioning similarly.

The PUT blocks current until its internal resistance drops to allow flow from the "anode" to the "cathode." In this way, it seems very similar to an NPN transistor, but there's a big difference in the circumstances that cause the PUT to lower its resistance. The voltage at the anode determines when the PUT allows current to flow.

Suppose you start with, say, 1 volt at the anode. Slowly, you increase this voltage. The transistor blocks it until the anode gets close to 6 volts. Suddenly this pressure breaks down the resistance and current surges from the anode to the cathode. If the voltage goes back down again, the transistor reverts to its original state and blocks the flow.

I've included another version of the "finger on the button" drawing to convey this concept. The voltage on the anode is itself responsible for pushing the button that opens the pathway to the cathode. See Figure 2-101.

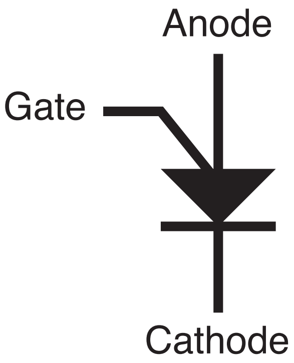

Figure 2-99. *The schematic symbol for a PUT.*

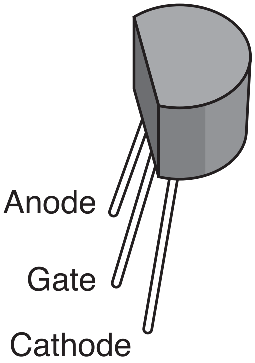

Figure 2-100. *In PUTs manufactured by On Semiconductor and Motorola, the leads have these functions.*

ESSENTIALS

All about programmable unijunction transistors (continued)

This may cause you to wonder what the function of the gate is. You can think of it as "assisting" the finger on the button. In fact, the gate is the "programmable" part of a PUT. By choosing a voltage for the gate, you establish the threshold point when current starts to flow.

Here's a simple take-home summary:

- The anode has to be more positive than the cathode, and the gate should be between those two extremes.
- If anode voltage increases above a threshold point, current bursts through and flows from the anode to the cathode.
- If anode voltage drops back down below the threshold, the transistor stops the flow.
- The voltage you apply to the gate determines how high the threshold is.
- The gate voltage is adjusted with two resistors, shown as R1 and R2 in the simple schematic in Figure 2-102. Typically, each resistor is around 20K. The PUT is protected from full positive voltage by R3, which can have a high value, 100K or greater, because very little current is needed to bias the transistor.
- You add your input signal in the form of positive voltage at the anode. When it exceeds the threshold, it flows out of the cathode and can work some kind of output device.

The only remaining question is how we make a PUT oscillate, to create a stream of on/off pulses. The answer is the capacitor that you included in the circuit that you breadboarded at the beginning of Experiment 11.

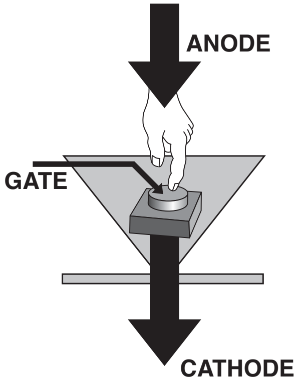

Figure 2-101. *When voltage at the anode of a PUT crosses a threshold (determined by a preset voltage at the gate), current breaks through and surges from the anode to the cathode. In this sense, the anode voltage acts as if it presses a button itself to open a connection inside the PUT, with some assistance from control voltage at the gate.*

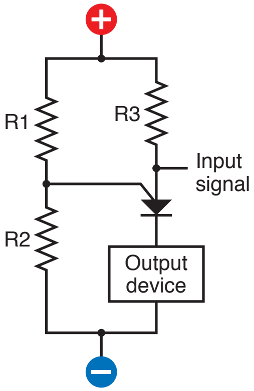

Figure 2-102. *This simple schematic shows how a PUT is used. R1 and R2 determine the voltage at the gate, which sets the threshold point for the input at the anode. Above the threshold, current flows from anode to cathode.*

Step 1: Slow-Speed Oscillation

Figure 2-103 is a schematic version of the previous PUT breadboard circuit shown in Figure 2-98, drawn so that the layout looks as much like the breadboard as possible.

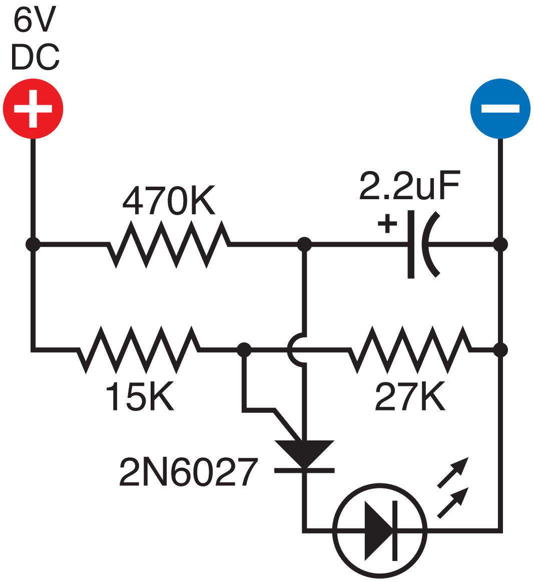

Figure 2-103. *This makes it easier to see what's happening in the breadboard version.*

The 15K resistor and 27K resistor establish the voltage at the gate. The 470K resistor supplies the anode of the PUT, but the PUT begins in its "off" condition, blocking the voltage. So the voltage starts to charge the 2.2 μF capacitor.

You may remember that a resistor slows the rate at which a capacitor accumulates voltage. The bigger the resistor and/or the larger the capacitor, the longer the capacitor takes to reach a full charge. In this circuit, the capacitor takes about half a second to get close to 6 volts.

But notice that the PUT is connected directly with the capacitor. Therefore, whatever voltage accumulates on the capacitor is also experienced by the PUT. As the voltage gradually increases, finally it reaches the threshold, which flips the PUT into its "on" state. The capacitor immediately discharges itself through the PUT, through the LED (which flashes), and from there to the negative side of the power supply.

The surge depletes the capacitor. The voltage drops back down, and the PUT returns to its original state. Now the capacitor has to recharge itself all over again, until the whole process repeats itself.

If you substitute a 22 μF capacitor, the charge/discharge cycle should take about 10 times as long, which will give you time to measure it. Set your meter to measure volts DC and place its probes on either side of the capacitor. You can actually watch the charge increasing until it reaches the threshold, at which point the capacitor discharges and the voltage drops back down again.

So now we have an oscillator. What's next?

Step 2: Beyond the Persistence of Vision

If you substitute a much smaller capacitor, it will charge much more quickly, and the LED will flash faster. Suppose you use a capacitor of 0.0047 µF (which can also be expressed as 47 nanofarads, or 47 nF). This seems like an odd number, but it's a standard value for a capacitor. This will reduce the capacitance by a factor of more than 500, and therefore the LED should flash about 500 times as fast, which should be about 1,000 times per second. The human eye cannot detect such rapid pulses. The human ear, however, can hear frequencies up to 10,000 per second and beyond. If we substitute a miniature loudspeaker for the LED, we should be able to hear the oscillations.

Figure 2-104 shows how I'd like you to make this happen. Please leave your original, slow-flashing circuit untouched, and make a duplicate of it farther down the breadboard, changing a couple of component values as indicated. In the schematic in Figure 2-105, the new part of the circuit is in solid black, while the previous section is in gray.

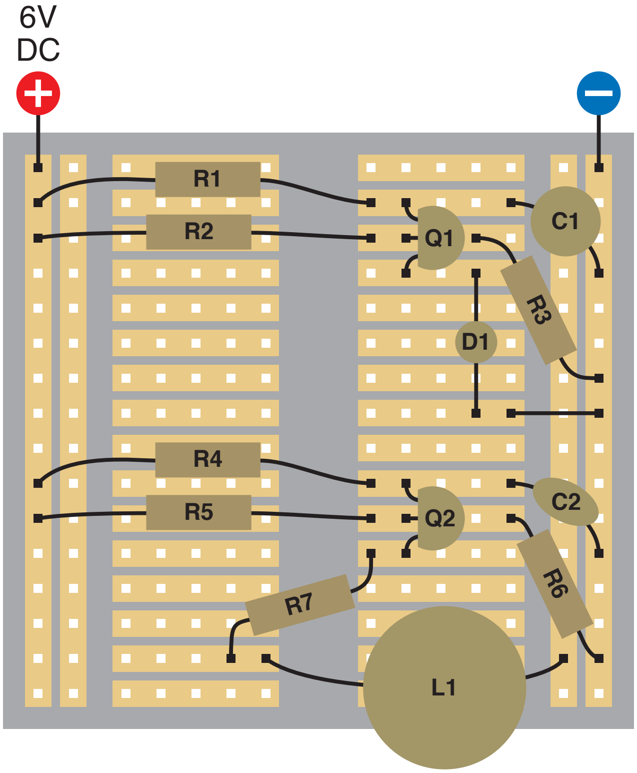

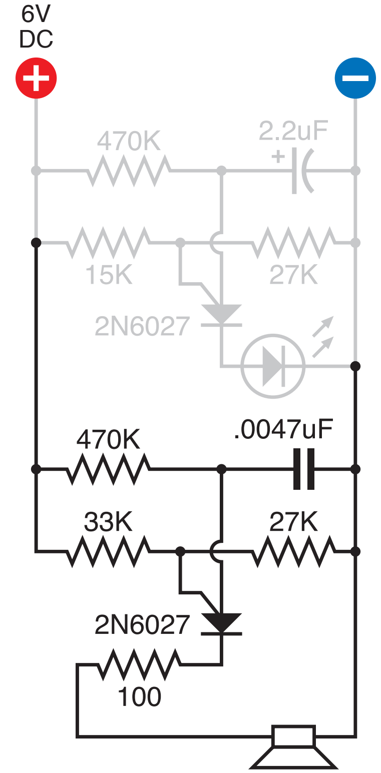

Figure 2-104. *The extra components which have been added at the lower half of the breadboard have the same functions as the components at the top, but some values are slightly different:*

R4: 470K
R5: 33K
R6: 27K
R7: 100Ω
C2: 0.0047 µF
Q2: 2N6027
L1: 8Ω 1-inch loudspeaker

Figure 2-105. *The previous section that you built is shown in gray. Just add the new section in black.*

I want you to keep the slow-flashing circuit separately, untouched, because I have an idea to make use of it a little later. You can leave the LED blinking.

The loudspeaker should be wired in series with a 100Ω resistor to limit the current that flows out of the PUT. The loudspeaker doesn't have any polarity, even though it is fitted with a red wire and a black wire. You can connect it either way around.

Initially, you may be disappointed, because the circuit will not seem to be doing anything. However, if you place your ear very, very close to the loudspeaker, and if you wired the circuit correctly, you should hear a faint buzz, like a mosquito. Obviously, this isn't loud enough to serve any practical purpose. We need to make it louder. In other words, we need to amplify it.

Maybe you remember that the 2N2222, which you played with previously, can function as an amplifier. So let's try using that.

Step 3: Amplification

Disconnect the loudspeaker and its 100Ω series resistor. Then add the 2N2222, which is linked with the output from the PUT via a 1K resistor to protect it from excessive current. See Figure 2-107.

The emitter of the 2N2222 is connected to ground, and the collector is supplied through the loudspeaker and its 100Ω series resistor. This way, small fluctuations in the output from the PUT are sensed by the base of the 2N2222 which converts them into bigger fluctuations between the collector and the emitter, which draw current through the loudspeaker. Check the schematic in Figure 2-108.

Now the sound should be louder than an insect buzz, but still not really loud enough to be useful. What to do?

Well—how about if we add another 2N2222? Bipolar transistors can be placed in series, so that the output from the first one goes to the base of the second one. The 24:1 amplification of the first one is multiplied by another 24:1, giving a total amplification of more than 500:1.

There are limits to this technique. The 2N2222 can only conduct so much current before getting overloaded, and excess amplification can cause distortion. But when I built this circuit, I used a meter to verify that we're still within the design limits of a 2N2222, and for this project, I don't care whether the sound is slightly distorted.

BACKGROUND

Mounting a loudspeaker

The diaphragm or cone of a loudspeaker is designed to radiate sound, but as it oscillates to and fro, it emits sound from its back side as well as its front side. Because the sounds are opposite in phase, they tend to cancel each other out.

The perceived output from a loudspeaker can increase dramatically if you add a horn around it in the form of a tube to separate the output from the front and back of the speaker. For a miniature 1-inch loudspeaker, you can bend and tape a file card around it. See Figure 2-106.

Better still, mount it in a box so that the box absorbs the sound from the rear of the loudspeaker. For purposes of these simple experiments, I won't bother to go into the details of vented enclosures and bass-reflex designs.

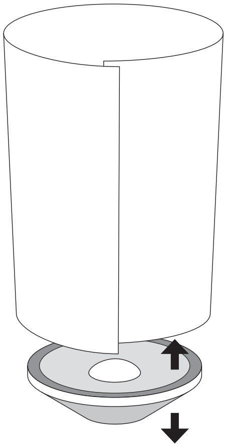

Figure 2-106. *A loudspeaker emits sound from its bottom surface as well as its top surface. To increase the perceived audio volume, use a cardboard tube to separate the two sound sources, or mount the speaker in a small box.*

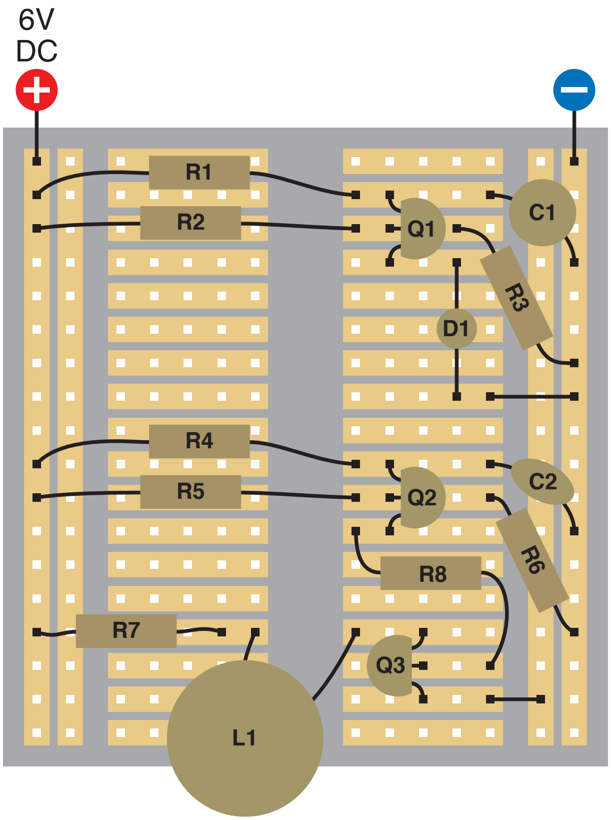

Figure 2-107. *By adding a 2N2222 general-purpose transistor, we amplify the signal from Q2:*

R8: 1K
Q3: 2N2222

Other components are the same as in the previous step in constructing this circuit.

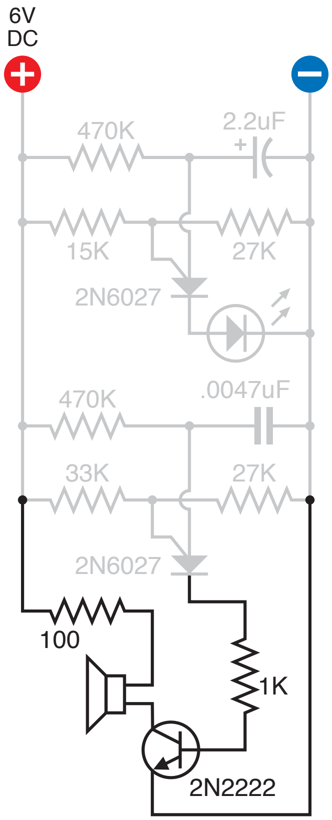

Figure 2-108

Add the second 2N2222 as shown in Figure 2-109. In Figure 2-110, once again the previously wired section is in gray.

If the accumulation of electrical components is beginning to seem confusing, remember that each cluster of parts has a separate defined function. We can draw a block diagram to illustrate this, as in Figure 2-112.

Using the second 2N2222, you should find that the output is more clearly audible, at least within the limits of your tiny 1-inch loudspeaker. Cup your hands around it to direct the sound, and you'll find that the volume seems to increase. You can also try using a 3-inch loudspeaker, which will create a generally better audio output while still remaining within the limits of the little 2N2222 transistor. See Figure 2-106, shown previously, and Figure 2-111.

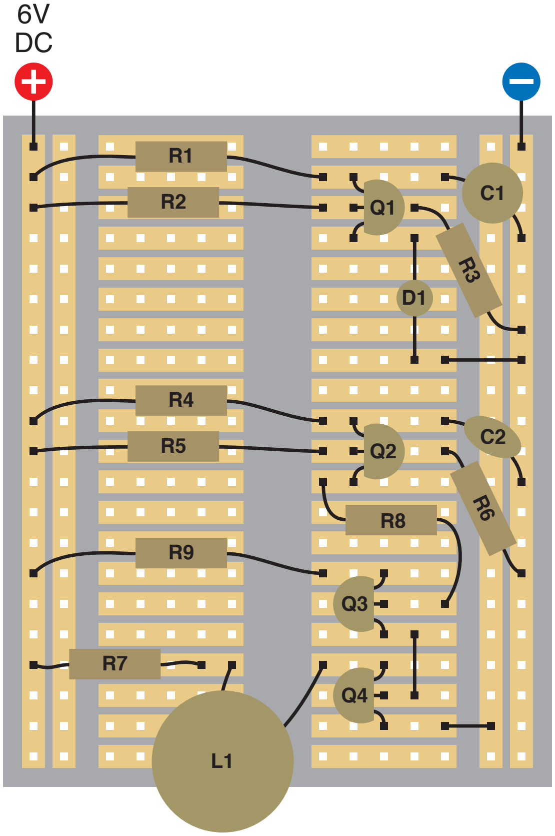
6V
DC
R1
R2
Q1
C1
R3
D1
R4
R5
Q2
C2
R6
R8
R9
Q3
R7
Q4
L1

Figure 2-109. *Q4 is another 2N2222 transistor that further amplifies the signal. It receives power through R9: 2.2K.*

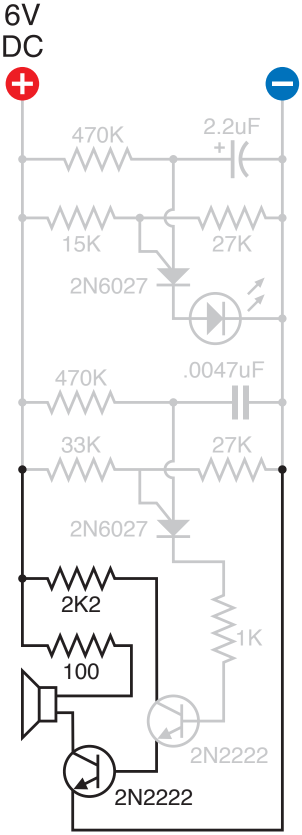
6V
DC
470K
2.2uF
+
15K
27K
2N6027
470K
.0047uF
33K
27K
2N6027
2K2
1K
100
2N2222
2N2222

Figure 2-110. *This schematic is comparable with the component layout in Figure 2-109.*

Figure 2-111. *The 2N2222 transistor is quite capable of driving a 3-inch loudspeaker, which will create much better sound than a 1-inch speaker.*

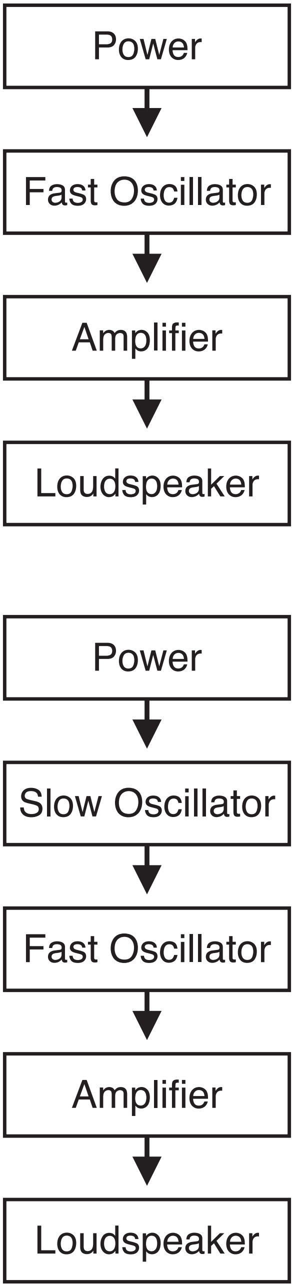

Figure 2-112. *Top: The basic functions of the noisemaking oscillator circuit shown as a block diagram. Bottom: The same functions with a slow oscillator added to control the fast oscillator.*

Step 4: Pulsed Output

If you wanted to use this audio signal as some kind of an alarm, a steady droning noise is not very satisfactory. A pulsing output would be a much better attention-getter.

Well, the first section of the circuit that you assembled created a pulsing signal about twice per second. You used it to flash an LED. Maybe we can get rid of the LED and feed the output from the first section to the second section. The lower block diagram in Figure 2-112 explains this concept.

Can it really be that simple? Well, yes and no. The trick is to make the output from the first section compatible with the input to the second section. If you simply connect a wire from the cathode of the first PUT to the anode of the second PUT, that's not going to work, because the second PUT is already oscillating nicely between low and high voltage, about 1,000 times each second. Add more voltage, and you will disrupt the balance that enables oscillation.

However, remember that the voltage on the gate of a PUT affects its threshold for conducting electricity. Maybe if we connect the output from Q1 to the gate of Q2, we'll be able to adjust that threshold automatically. The voltage still has to be in a range that the PUT finds acceptable, though. We can try various resistors to see which one works well.

This sounds like trial and error—and that's exactly what it is. Doing the math to predict the behavior of a circuit like this is far too complicated—for me, anyway. I just looked at the manufacturer's data sheet, saw the range of resistor values that the PUT would tolerate, and chose one that seemed as if it should work.

If you remove the LED and substitute R10 as shown in the breadboard diagram in Figure 2-113, you'll find that the fluctuating output from Q1 makes Q2 emit a two-tone signal. This is more interesting, but still not what I want. I'm thinking that if I make the pulses out of Q1 less abrupt, the result could be better, and the way to smooth a pulsing output is to hook up another capacitor that will charge at the beginning of each pulse and then release its charge at the end of each pulse. This is the function of C3 in Figure 2-114, and it completes the circuit so that it makes a whooping sound almost like a "real" alarm.

If you don't get any audio output, check your wiring very carefully. It's easy to make a wrong connection on the breadboard, especially between the three legs of each transistor. Use your meter, set to DC volts, to check that each section of the circuit has a positive voltage relative to the negative side of the power supply.

Figure 2-115 shows how your circuit should actually look on the breadboard.

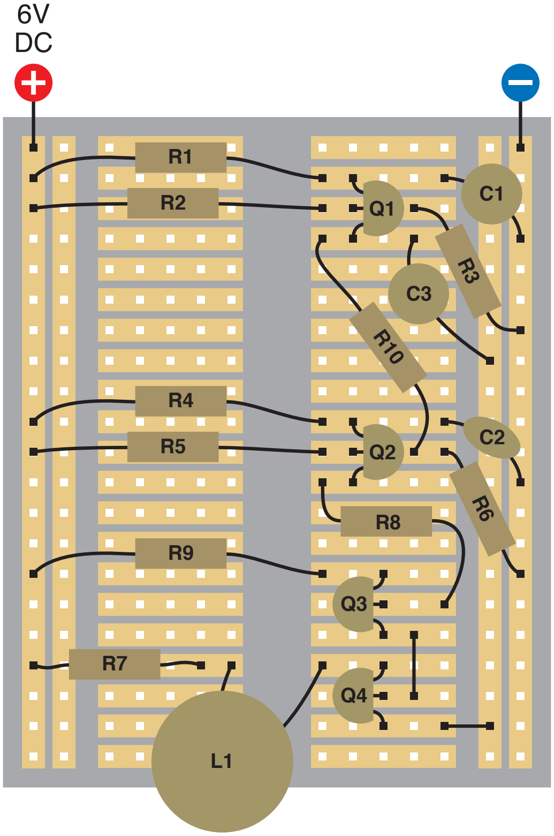
6V
DC
R1
R2
Q1
C1
R3
C3
R10
R4
R5
Q2
C2
R6
R8
R9
Q3
R7
Q4
L1

Figure 2-113. *R10 connects the slow-running oscillator at the top of the breadboard to the gate of Q2, the PUT in the middle of the breadboard. This modulates the audio oscillator, with addition of a smoothing capacitor.*

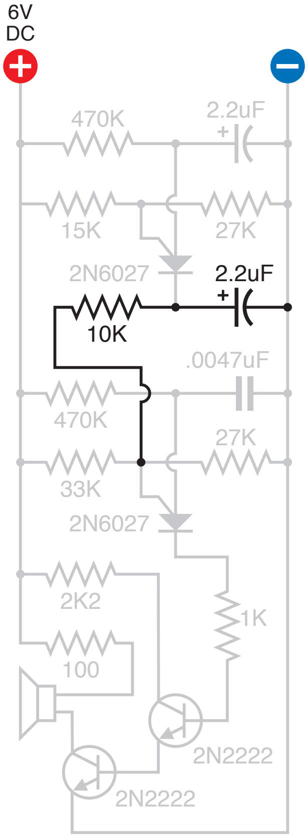
6V
DC
470K
2.2uF
+
15K
27K
2N6027
2.2uF
+
10K
.0047uF
470K
27K
33K
2N6027
2K2
1K
100
2N2222
2N2222

Figure 2-114. *This schematic shows the same circuit as in Figure 2-113:*

R10: 10K
C3: 2.2 μF

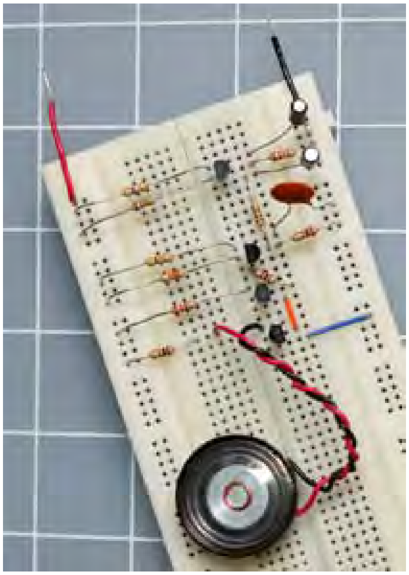

Figure 2-115. *This photograph shows the complete alarm-audio circuit on a breadboard.*

Tweaking it

There's still a lot of room for creativity here:

- Adjust the frequency of the sound: Use a smaller or larger capacitor instead of C2 (half or twice the current value). Use a smaller or larger value for R5.
- Adjust the pulsing feature: Use a smaller or larger capacitor instead of C1 (half or twice the current value). Use a smaller or larger value for R2.
- General performance adjustments: try a larger value for R1. Try smaller or larger values for C3.
- Try running the circuit at 7.5 volts, 10 volts, and 12 volts.

The circuits in this book are suggested as only a starting point. You should always try to tweak them to make them your own. As long as you follow the general rule of protecting transistors and LEDs with resistors, and respecting their requirements for positive and negative voltage, you're unlikely to burn them out. Of course, accidents will happen—I myself tend to be careless, and fried a couple of LEDs while working on this circuit, just because I connected them the wrong way around.

Step 5: Enhancements

A noisemaking circuit is just the output of an alarm. You would need several enhancements to make it useful:

1. Some kind of an intrusion sensor. Maybe magnetic switches for windows and doors?
2. A way to start the sound if any one of the sensors is triggered. The way this is usually done is to run a very small but constant current through all of the switches in series. If any one switch opens, or if the wire itself is broken, this interrupts the current, which starts the alarm. You could make this happen with a double-throw relay, keeping the relay energized all the time until the circuit is broken, at which point, the relay relaxes, opening one pair of contacts and closing the other pair, which can send power to the noisemaker.

   The trouble is that a relay draws significant power while it's energized, and it also tends to get hot. I want my alarm system to draw very little current while it's in "ready" mode, so that it can be powered by a battery. Alarm systems should never depend entirely on AC house current.

   If we don't use a relay, can we use a transistor to switch on the rest of the circuit when the power is interrupted? Absolutely; in fact, one transistor will do it.
3. But how do we arm the alarm in the first place? Really, we need a three-step procedure. First, check a little light that comes on when all the doors and windows are closed. Second, press a button that starts a 30-second countdown, giving you time to leave, if that's what you want to do. And third, after 30 seconds, the alarm arms itself.

4. If the alarm is triggered, what then? If someone forces open a window, should the alarm stop sounding as soon as the window is closed again? No, the alarm should lock itself on, until you turn it off.
5. How do you turn it off? Some kind of secret-code keypad would be good.
6. But to avoid driving everyone crazy if the alarm is triggered when you're not there, it should eventually stop itself, perhaps after about 10 minutes. At that point it should remain quiet but should light an LED to tell you what happened. You can then press a reset button to switch off the LED.

Implementing the Wish List

I've compiled a wish list that seems likely to make the project at least five times as complicated as it is already. Well, that's what tends to happen when you go beyond little demo circuits and try to design something that will be useful in everyday life. Suddenly you find yourself having to accommodate all kinds of circumstances and situations.

Actually, I can and will show you how to take care of all the enhancements on the wish list, but I'm thinking that they will require us to get a little more serious about electronics projects in general first. If you're going to build something ambitious, you'll want to make it more permanent, and probably more compact, than a breadboard with components pushed into it.

You will need to know how to connect everything permanently with solder, on a piece of perforated board that you can install in a neat little project box with switches and lights on the outside.

I'm going to deal with soldering and packaging in the next chapter. After that, we'll get back to the alarm project.

3

Getting Somewhat More Serious

IN THIS CHAPTER

Shopping List: Experiments 12 Through 15
Experiment 12: Joining Two Wires Together
Experiment 13: Broil an LED
Experiment 14: A Pulsing Glow
Experiment 15: Intrusion Alarm Revisited

I don't know how far you'll want to delve into electronics, but I do know that I've shown you about as much as I can with just a handful of components, some wires, a breadboard, and a few tools. To continue, you'll need:

- Some more tools and supplies
- Basic soldering skills
- Additional knowledge about:
    - Integrated circuits
    - Digital electronics
    - Microcontrollers
    - Motors

The tools are not particularly exotic or expensive, and the soldering skills are easily acquired. Learning to join wires with solder is far easier than mastering high-level crafts such as jewelry making or welding.

As for additional areas of knowledge about electronics, they are no more challenging than those that I have covered already.

By the end of this section, you should be able to transplant components from a breadboard onto perforated board, where you will solder everything together, and then mount the board in a little box with switches and warning lights on the front, for everyday use.

Shopping List: Experiments 12 Through 15

Tools

Each of the following tools is rated Essential, Recommended, or Optional. The Essential tools will take you through this chapter of the book. If you supplement them with the Recommended tools, they should be sufficient to get you to the end of the book. The Optional tools will make your work easier, but whether they're worth the money is for you to decide. Remember that URLs for manufacturers and sources of supply are all listed in the appendix.

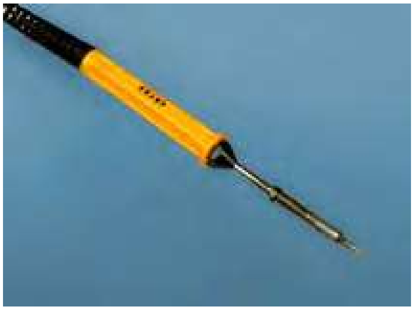

Figure 3-1. *The low wattage of this pencil-style soldering iron enables you to use it safely on sensitive components, and the sharp tip helps to apply heat selectively.*

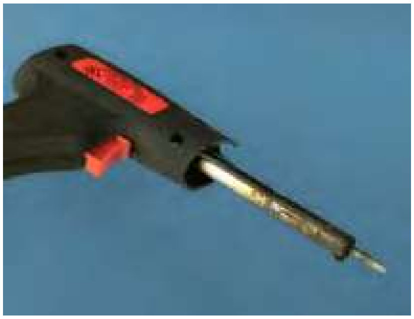

Figure 3-2. *This higher-wattage soldering iron delivers the additional heat necessary for thicker wire or larger components. The discoloration quickly occurs as a result of everyday use and has no effect on the capability of the iron, as long as the tip of it is clean.*

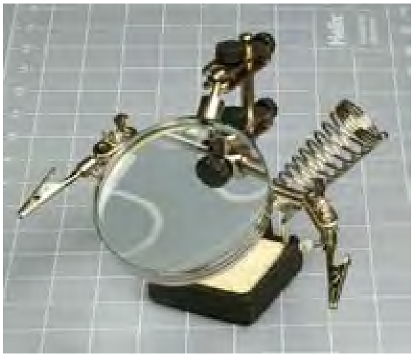

Figure 3-3. *The helping hand is fitted with two alligator clips to hold your work. The metal spiral is a safe place to holster a hot soldering iron, and you use the sponge to wipe its tip.*

I am assuming that you already have some commonly used workshop basics, such as an electric drill.

Essential: Pencil-type 15-watt soldering iron

Examples are RadioShack part 64-2051, McMaster-Carr catalog item 7016A34, or Xytronic model 252. See Figure 3-1. Soldering irons rated at 15 watts are less common than those that deliver 25 watts or more. Still, the 15-watt size is desirable for the small-scale work you'll be doing, and greatly reduces the risk of damaging components by inflicting excess heat.

When comparing prices, remember that a plated tip, which costs a little more, will last longer, will be easier to keep clean, and will conduct heat more reliably than a plain copper tip. If the manufacturer's specification doesn't mention a plated tip, the soldering iron probably doesn't have one.

Essential: General-duty soldering iron, 30 to 40 watts

Although most of the projects in this book entail small, heat-sensitive components and thin wire, at some point you're likely to want to make a solder joint with larger components and/or thicker wire. A 15-watt soldering iron will be unable to deliver enough heat. You should consider having a larger soldering iron in reserve, especially because they are relatively inexpensive.

Personally, I like the Weller Therma-Boost, shown in Figure 3-2, because it has an extra button that delivers more heat on demand. This is useful when you want the iron to get hot quickly, or if you are trying to solder very thick wire, which absorbs a lot of heat.

If you can't find or don't like the Weller, almost any 30-watt or 40-watt soldering iron will do. Check eBay or your local hardware store.

Essential: Helping hand

The so-called "helping hand" (or "third hand") has two alligator clips that hold components or pieces of wire precisely in position while you join them with solder. Some versions of the "helping hand" also feature a magnifying lens, a wire spiral in which you can rest your soldering iron, and a little sponge that you use to clean the tip of your iron when it becomes dirty. These additional features are desirable. Helping hands are available from all electronics hobby sources. Examples are the catalog item HH55 from Elenco or model 64-2991 from RadioShack. See Figure 3-3.

Essential: Magnifying lens

No matter how good your eyes are, a small, handheld, powerful magnifying lens is essential when you are checking solder joints on perforated board. The three-lens set in Figure 3-4 is designed to be held close to your eye, and is more powerful than the large lens on a "helping hand." The folding lens in Figure 3-5 stands on your workbench for hands-free operation. Both are available from RadioShack and similar items are stocked by art supply stores and hobby shops. Plastic lenses are quite acceptable if you treat them carefully.

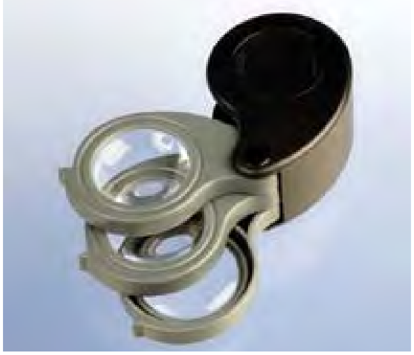

Figure 3-4. *As long as you treat it carefully, a cheap set of plastic magnifying lenses is perfectly acceptable. Handheld magnification is essential for inspecting the solder joints that you make on perforated board.*

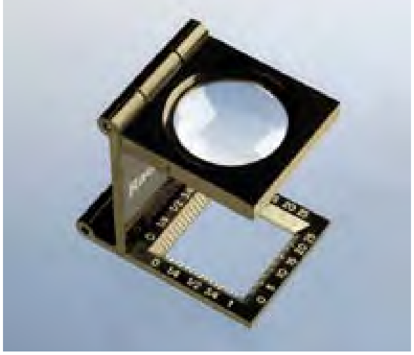

Figure 3-5. *This kind of folding magnifier can stand on your desktop and is useful for checking part numbers on tiny components.*

Essential: Clip-on meter test leads

The probes that came with your multimeter require you to hold them in contact while you make a reading. This requires both hands, preventing you from doing anything else at the same time.

When you use a pair of "minigrabber" probes with little spring-loaded clips at the end, you can attach the Common (negative) lead from your meter to the negative side of your circuit and leave it there, while you touch or attach the positive probe elsewhere.

The Pomona model 6244-48-0 (shown in Figure 3-6) from Meter Superstore and some other suppliers is what you need. If you have trouble finding it or you object to the cost, you may consider making your own by buying a couple of "banana plugs" (such as RadioShack part 274-721) that will fit the sockets on your meter, and then use 16-gauge or thicker stranded wire to connect the plugs with IC test clips, such as Kobiconn 13IC331 or RadioShack "mini hook clips," part number 270-372C. See Figures 3-7 and 3-8.

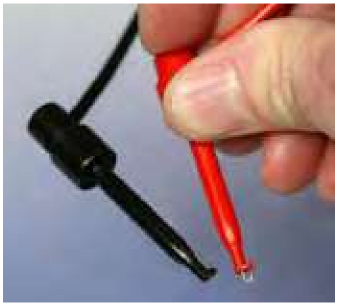

Figure 3-6. *These "minigrabber" add-ons for meter leads make it much easier to measure voltage or current. Push the spring-loaded button, and a little copper hook slides out. Attach it to a wire, release the button, and you have your hands free for other tasks. It's a mystery that meters are not supplied with these grabbers as standard equipment.*

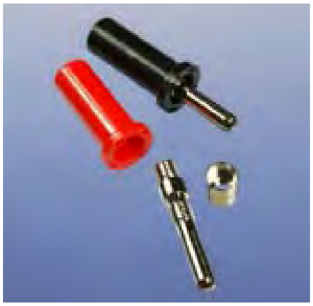

Figure 3-7. *To make your own minigrabber meter leads, first attach a banana plug to a wire by sliding the wire through the cap, into the plug, and out through a hole in the side.*

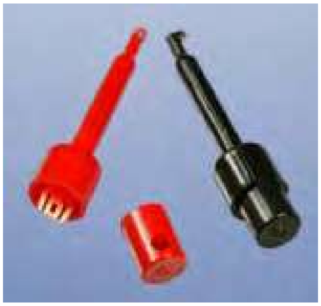

Figure 3-8. *Then screw a collar over the protruding piece of wire, and screw on the cap. The other end of the wire is soldered to a probe.*

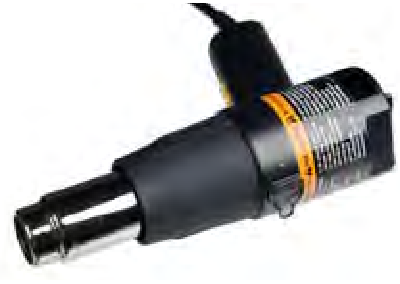

Figure 3-9. *Like an overpowered hair dryer, the heat gun is used with heat-shrink tubing to create a snug, insulated sheath around bare wire.*

Essential: Heat gun

After you join two wires with solder, you often need to insulate them. Electrical tape, sometimes called insulating tape, is messy and tends to come unstuck. You'll be using heat-shrink tube, which forms a safe, permanent sheath around a bare-metal joint. To make the tube shrink, use a heat gun, which is like a very powerful hair dryer. They're available from any hardware supply source, and I suggest you buy the cheapest one you can find. See Figure 3-9.

Essential: Solder pump

This little gadget sucks up hot, melted solder when you are trying to remove a solder joint that you made in the wrong place. Available from All Electronics (catalog item SSR-1) or RadioShack 64-2086. See Figure 3-10.

Essential: Desoldering wick

Also known as desoldering braid. See Figure 3-11. You use this to soak up solder, in conjunction with the Solder Pump. Available from All Electronics (catalog item SWK) or RadioShack (part 64-2090).

Essential: Miniature screwdriver set

Dinky little electronic parts often have dinky little screws in them, and if you try to use the wrong size of screwdriver, you'll tend to mash the heads of the screws. I like the Stanley precision set, part number 66-052, shown in Figure 3-12. But any set will do as long as it has both small Phillips and straight-blade screwdrivers.

Recommended: Soldering stand

Like a holster for a gun, you rest your soldering iron in this stand when the iron is hot but not on use. Examples are catalog item 50B-205 from All Electronics, RadioShack model 64-2078, or check eBay. See Figure 3-13. This item may be built into the helping hand, but you need an extra one for your second soldering iron.

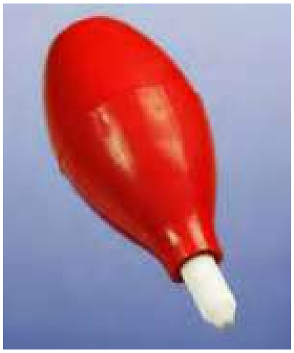

Figure 3-10. *To remove a solder joint, you can heat the solder until it's liquid, then suck it up into this squeezable rubber bulb.*

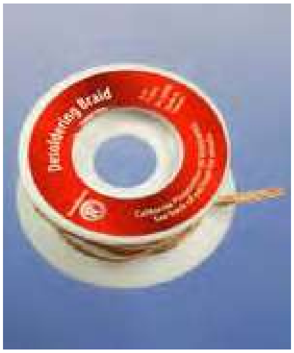

Figure 3-11. *An additional option for removing liquid solder is to soak it up in this copper braid.*

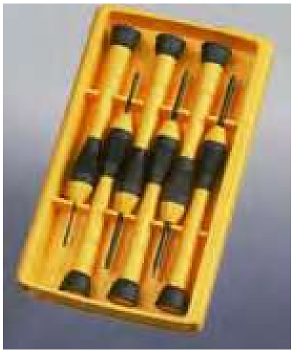

Figure 3-12. *A set of small screwdrivers is essential.*

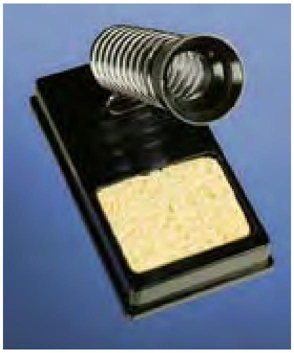

Figure 3-13. *A safe and simple additional stand for a hot soldering iron.*

Recommended: Miniature hand saw

I assume that you will want to mount a finished electronics project in a decent-looking enclosure. Consequently, you are likely to need tools to cut, shape, and trim thin plastic. For example, you may want to cut a square hole so that you can mount a square power switch in it.

Power tools are not suitable for this kind of delicate work. A miniature handsaw (a.k.a. a "hobby saw") is ideal for trimming things to fit. X-Acto makes a range of tiny saw blades. I suggest the #15 blade, plus the handle that it fits in, shown in Figure 3-14. Available online from Tower Hobbies, Hobbylinc, ArtCity, and many other arts/crafts sources. Also look for the larger X-Acto saw blade, #234 or #239, which you can use for cutting perforated board.

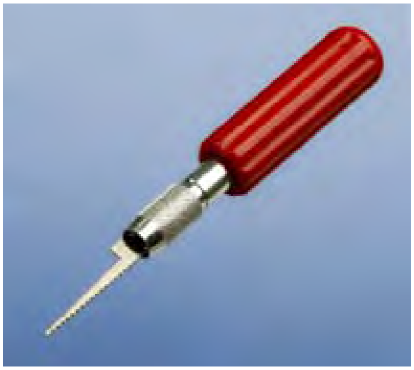

Figure 3-14. *X-Acto makes a range of small saw blades that are ideal for cutting square holes to mount components in plastic panels.*

Recommended: Miniature vise

A miniature vise can do things that the helping hand cannot. I use mine when I'm sawing small pieces of plastic and as a dead weight to anchor a piece of perforated board while I'm working on it. See Figure 3-15.

Look for a cast-iron vise that is listed as being 1 inch or slightly larger, available from Megahobby, eBay, and other arts/crafts sources. Also consider the PanaVise, which has a tilting head to allow you to turn your work to any angle.

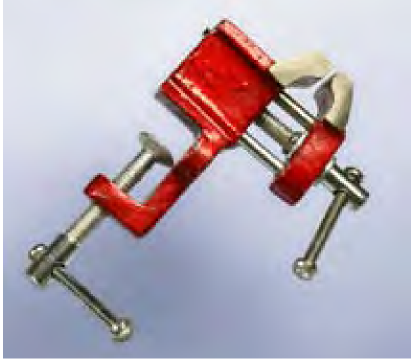

Figure 3-15. *This one-inch vise is available from the McMaster-Carr catalog.*

Recommended: Deburring tool

A deburring tool instantly smoothes and bevels any rough edge (when you have sawn or drilled a piece of plastic, for instance) and also can enlarge holes slightly. This may be necessary because some components are manufactured to metric sizes, which don't fit in the holes that you drill with American bits. Your small local hardware store may not stock deburring tools, but they are very inexpensively available from Sears, McMaster-Carr, KVM Tools, or Amazon. See Figure 3-16.

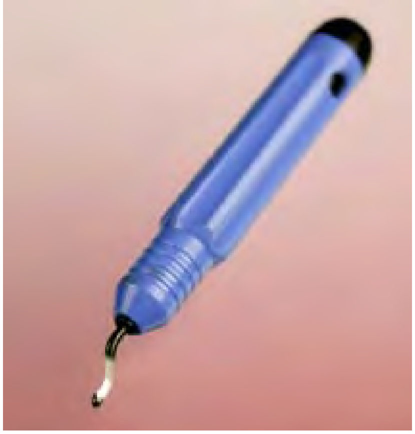

Figure 3-16. *This cunning little blade, safety-tipped with a round bump on the end, removes rough edges from saw cuts with a single stroke, and can enlarge holes that are almost big enough—but not quite.*

Optional: Hand-cranked countersink

You need a countersink to bevel the edges of screw holes to accept flat-headed screws. If you use a countersink bit in an electric drill, it won't give you precise control when you're working with thin, soft plastic.

Handheld countersinks that you grasp and turn like a screwdriver are easy to find, but McMaster-Carr (catalog item 28775A61) is the only source I've found for a hand-cranked tool that is much quicker to use. It comes with a set of bits, as shown in Figure 3-17.

Optional: Pick and hook set

Made by Stanley, part number 82-115, available from Amazon and hardware stores. You can find imported imitations for a few dollars less. See Figure 3-18.

Optional: Calipers

These may seem like a luxury, but are useful for measuring the external diameter of a round object (such as the screw thread on a switch or a potentiometer) or the internal diameter of a hole (into which you may want a switch or potentiometer to fit).

I like Mitutoyo calipers, and the low-end model 505-611 (shown in Figure 3-19) does everything I need. You can find cheaper brands, but economizing on precision measuring tools may not be a wise policy in the long term. The manufacturer's site will show you all their available models, after which you can Google "Mitutoyo" to find retail sources.

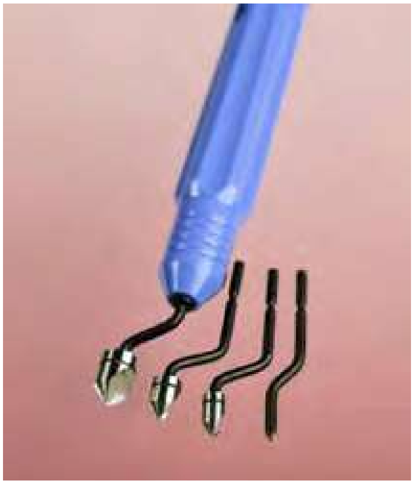

Figure 3-17. *You spin this countersink tool like a hand crank to add just the right amount of bevel to a hole, so that it will accommodate a flat-head screw.*

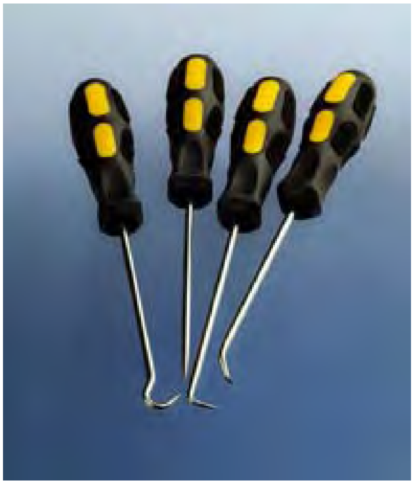

Figure 3-18. *This pick-and-hook set is useful in many unexpected ways.*

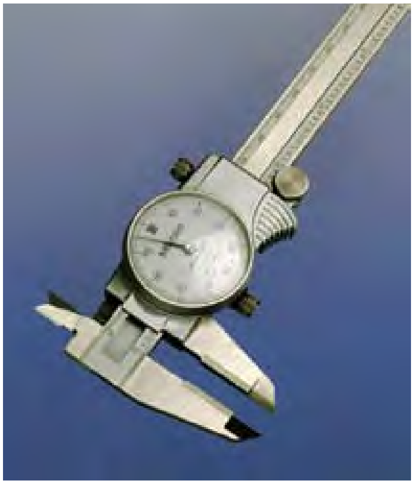

Figure 3-19. *Calipers can be digital (which automatically convert from millimeters to 1/64 inch to 1/1,000 inch), or analog like these (so you never need to worry about a dead battery).*

Supplies

Solder

This is the stuff that you will melt to join components together on a permanent (we hope) basis. You need some very thin solder, size 0.022 inches, for very small components, and thicker solder, 0.05 inches, for heavier items. Avoid buying solder that is intended for plumbers, or for craft purposes such as creating jewelry. A range of solder thicknesses is shown in Figure 3-20. You want to make sure to get lead-free solder.

Electronics solder has a nonacidic rosin core that is appropriate for electronic components. Rolls of solder are available from all hobby-electronics sources including All Electronics, RadioShack, and Jameco, or search for "electronic solder" on Amazon.

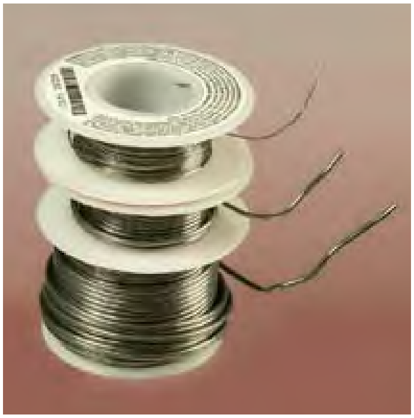

Figure 3-20. *Spools of solder in various thicknesses.*

Wire

You'll need some stranded wire to make flexible external connections with the circuit that you'll be building. Look for 22-gauge stranded hookup wire, in red, black, and green, 10 feet (minimum) of each.

If you want to install the intrusion alarm after completing that project in Experiment 15, you'll need white-insulated two-conductor wire of the type sold for doorbells or furnace controls. This is available by the foot from Lowe's, Home Depot, Ace Hardware, and similar stores. You'll decide how much to buy after you measure the distances between the magnetic sensor switches that you decide to install.

Heat-shrink tube

For use in conjunction with your heat gun, described previously. You'll need a range of sizes in any colors of your choice. See Figure 3-21. Check RadioShack part 278-1627, other electronics suppliers, or your local hardware store. Prices will vary widely. You can buy the cheapest.

Copper alligator clips

These absorb heat when you are soldering delicate components. The Mueller BU-30C is a full-size solid copper alligator clip for maximum heat absorption. RadioShack sells smaller clips (part number 270-373, shown in Figure 3-22) that are suitable for tiny components.

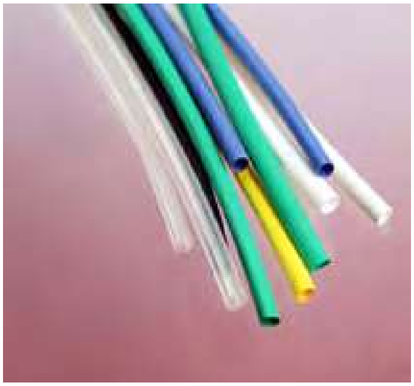

Figure 3-21. *Slide heat-shrink tubing over a bare joint and apply heat from a heat gun to make a tight insulating seal around the joint.*

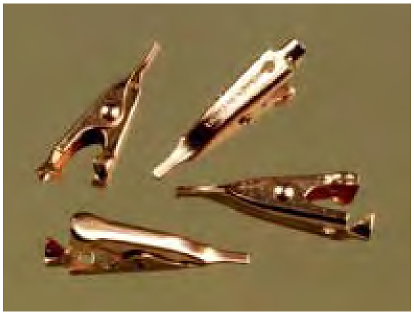

Figure 3-22. *These small clips absorb heat to protect components when you're soldering them.*

Perforated board

When you're ready to move your circuit from a breadboard to a more permanent location, you'll want to solder it to a piece of perforated board, often known as "prototyping board" but also called "perfboard."

You need the type that has copper strips etched onto the back, in exactly the same "breadboard layout" as the conductors hidden inside a breadboard, so that you can retain the same layout of your components when you solder them into place. Examples are RadioShack part 276-150 (shown in Figure 3-23) for small projects and part 276-170 (in Figure 3-24) for larger projects, such as Experiment 15.

For very small projects in which you will connect components using their wires alone, you need perfboard that isn't etched with copper strips connecting the holes. I like the Twin Industries 7100 range (available from Mouser.com) or Vectorboard from Newark Electronics, shown in Figure 3-25. You use a saw to cut out as small a piece as you need. Cheaper options are RadioShack part 276-147 (shown in Figure 3-26), or PC-1 from All Electronics. These have little copper circles around each hole that are not necessary for our purposes, but not a problem, either.

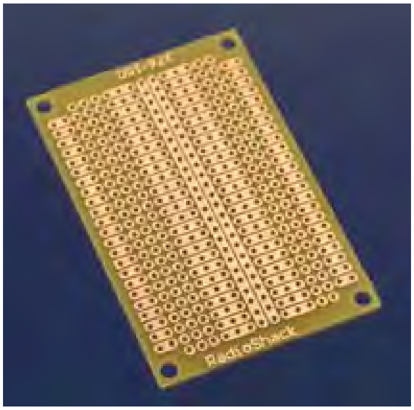

Figure 3-23. *This perforated board has a pattern of copper traces similar to the pattern inside a breadboard, so that you can lay out the components with minimal risk of wiring errors, when you're ready to create a permanently soldered version of your project.*

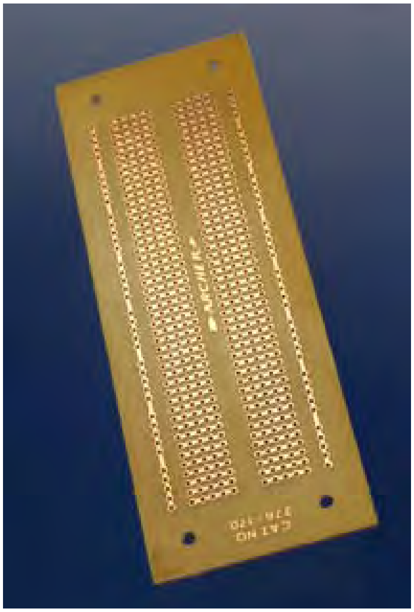

Figure 3-24. *A larger example of perforated board with breadboard geometry.*

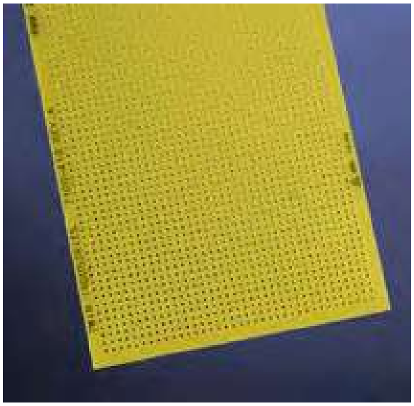

Figure 3-25. *Plain perforated board (with no copper traces) can be used for mounting components when you want to do point-to-point wiring.*

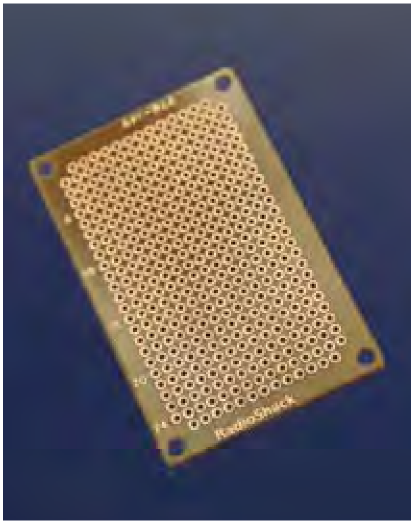

Figure 3-26. *A small piece of perforated board with individual copper solder pads to assist you in mounting components.*

Plywood

When you use a soldering iron, hot drops of solder tend to fall onto your table or workbench. The solder solidifies almost instantly, can be difficult to remove, and will leave a scar. Consider using a 2-foot square of half-inch plywood to provide disposable protection. You can buy it precut at Home Depot or Lowe's.

Machine screws

To mount components behind a panel, you need small machine screws (or "bolts"). They look nice if they have flat heads that fit flush against the panel. I suggest stainless-steel machine screws, #4 size, in 1/2-, 5/8-, 3/4-, and 1-inch lengths, 100 of each, plus 400 washers and 400 #4 locknuts of the type that have nylon inserts, so that they won't work loose. Check McMaster-Carr for a large and reasonably priced selection.

Project boxes

A project box is just a small box (usually plastic) with a removable lid. You mount your switches, potentiometers, and LEDs in holes that you drill through the box, and you attach your circuit on a perforated board that goes inside the box. Search All Electronics for "project box" or RadioShack for "project enclosure."

You need a box measuring approximately 6 inches long, 3 inches wide, and 2 inches high, such as RadioShack part 270-1805. Anything similar will do. I suggest you buy a couple other sizes as well, as they will be useful in the future.

Components

Power plugs, sockets, and binding posts

After you finish a project and put it in a box, you'll need a convenient way to supply it with power. Buy yourself a pair of insulated binding posts, such as RadioShack part 274-661, shown in Figure 3-27. Also obtain a panel-mounted power jack, size N, such as RadioShack part 274-1583, and DC power plug, size N, such as RadioShack 274-1573. The plug-and-socket pair is pictured in Figure 3-28.

Finally, you will need interconnects that are sized to fit a perforated board that is drilled at intervals of 1/10 inch. Sometimes known as "single inline sockets and headers," but also known as "boardmount sockets and pin-strip headers," they come in strips of 36 or more, and you can snip off as many as you need. Examples are Mill-Max part numbers 800-10-064-10-001000 and 801-93-050-10-001000, or 3M part numbers 929974-01-36-RK and 929834-01-36-RK. You can buy them from the usual electronics suppliers. Figure 3-29 shows headers before and after being snapped into small sections. Make sure that the interconnects have a terminal spacing of 0.1 inch.

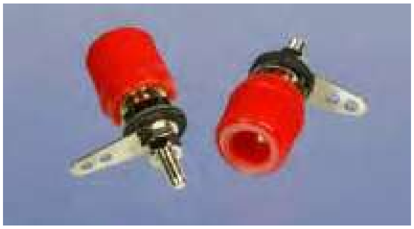

Figure 3-27. *These terminals, also known as binding posts, enable a solderless connection with wires that have stripped ends. Also available in black.*

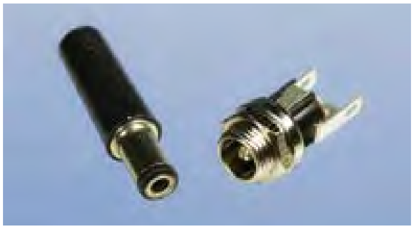

Figure 3-28. *The socket on the right can be mounted in a project box to receive power from the plug on the left.*

Battery

After you complete Experiment 15 at the end of this section of the book, if you want to use the project on a practical basis, you'll need a 12-volt battery. Search online for "12v battery" and you'll find many sealed, rechargeable lead-acid batteries that are designed for alarm systems, some measuring as small as 1×2×3-inch and costing under $10. You need a charger with it, which will probably cost you about $10.

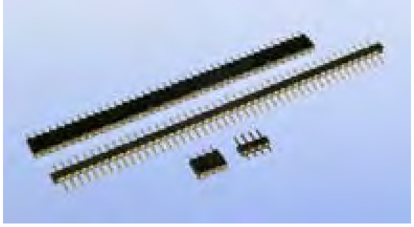

Figure 3-29. *Single inline sockets (top) and headers (middle) allow you to make very compact plug-and-socket connections to a PC board. They can be sawn, cut, or snapped into smaller sections (bottom). The terminals are 0.1 inch apart.*

Switches and relays

You will need the same DPDT relay and the same SPDT toggle switch that were mentioned in Chapter 2 shopping list.

For Experiment 15, you'll need magnetic switches that you can apply to doors or windows, such as the Directed model 8601, available from dozens of sources online.

Also you will need a DPDT pushbutton switch, ON-(ON) type, with solder terminals. Examples are model MPG206R04 by Tyco or model MB2061SS1W01-RO by NKK (with optional cap). Or search eBay for "DPDT pushbutton."

Diodes

Buy at least half-a-dozen red 5 mm LEDs rated for approximately 2 volts, such as the Optek part number OVLFR3C7, Lumex part number SSL-LX-5093IT, or Avago part HLMP-D155. Buy half-a-dozen similar green LEDs at the same time.

In addition, you'll need a signal diode, type 1N4001 (any brand will do). Figure 3-30 shows an example, highly magnified. They're cheap, and likely to be useful in the future, so buy 10 of them.

Figure 3-30. *This 1N4001 diode is about 1/4 inch long and can handle up to 50 volts.*

Loudspeaker

To complete the project in Experiment 15, you'll need a loudspeaker small enough to fit inside your project box but louder than the 1-inch speaker that you used previously. It should be 2 inches or 2.5 inches (50 to 60 mm) in diameter. If you can find a 100Ω speaker, it will give you more output, but an 8Ω speaker will be acceptable.

Experiment 12: Joining Two Wires Together

Your adventure into soldering begins with the prosaic task of joining one wire to another, but will lead quickly to creating a full electronic circuit on perforated board. So let's get started!

You will need:

- 30-watt or 40-watt soldering iron
- 15-watt pencil-type soldering iron
- Thin solder (0.022 inches or similar)
- Medium solder (0.05 inches or similar)
- Wire strippers and cutters
- "Helping hand" gadget to hold your work
- Shrink-wrap tubing, assorted
- Heat gun
- Something to protect your work area from drops of solder

Your First Solder Joint

We'll start with your general-duty soldering iron—the one rated for 30 or 40 watts. Plug it in, leave it safely in its holder, and find something else to do for five minutes. If you try to use a soldering iron without giving it time to get fully hot, you will *not* make good joints.

Strip the insulation from the ends of two pieces of 22-gauge solid wire and clamp them in your helping hand so that they cross each other and touch each other, as shown in Figure 3-31.

To make sure that the iron is ready, try to melt the end of a thin piece of solder on the tip of the iron. The solder should melt instantly. If it melts slowly, the iron isn't hot enough yet.

Soldering Irons Get Hot!

Please take these basic precautions:

Use a proper stand (such as the one incorporated in your helping hands) to hold your soldering iron. Don't leave it lying on a workbench.

If you have infants or pets, remember that they may play with, grab, or snag the wire to your soldering iron. They could injure themselves (or you).

Be careful never to rest the hot tip of the iron on the power cord that supplies electricity to the iron. It can melt the plastic in seconds and cause a dramatic short circuit.

If you drop a soldering iron, don't be a hero and try to catch it. Most likely you will grab the hot part, which hurts. (I speak from experience.) When you burn your hand, you will instinctively let go of the iron, so you may as well let it drop freely without the intermediate step of burning yourself while it's on its way to the floor. Naturally, you should pick it up quickly after it does hit the floor, but by then you will have gained the necessary time in which to make a sensible decision to grab it by the cool end.

Always bear in mind that others in your home are more at risk of hurting themselves on a soldering iron than you are, because they won't know that it's hot. Most soldering irons have no warning lights to tell you that they're plugged in. As a general rule, always assume that a soldering iron is hot, even if it's unplugged. It may retain sufficient heat to burn you for longer than you expect.

Now follow these steps (shown in Figures 3-32 through 3-36):

1. Make sure the tip of the soldering iron is clean (wipe it on the moistened sponge in the base of your helping hand if necessary), then touch it against the intersection of the wires steadily for three seconds to heat them. If you have hard tap water, use distilled water to wet the sponge to avoid a buildup of mineral deposits on the tip of your soldering iron.
2. While maintaining the iron in this position, feed a little solder onto the intersection of the wires, also touching the tip of the soldering iron. Thus, the two wires, the solder, and the tip of the iron should all come together at one point. The solder should spread over the wires within another two seconds.
3. Remove the iron and the solder. Blow on the joint to cool it. Within 10 seconds, it should be cool enough to touch.
4. Unclamp the wires and try to tug them apart. Tug hard! If they defeat your best attempts to separate them, the wires are electrically joined and should stay joined. If you didn't make a good joint, you will be able to separate the wires relatively easily, probably because you didn't apply enough heat or enough solder to connect them.

The reason I asked you to begin by using the higher-powered soldering iron is that it delivers more heat, which makes it easier to use.

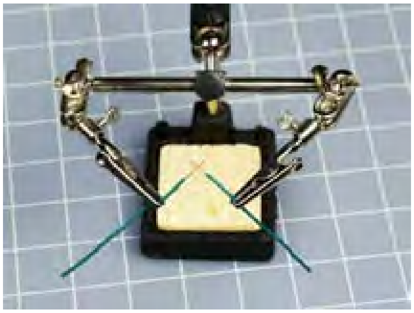

Figure 3-31. *A helping hand work aid is shown here holding two wires with their stripped ends touching. The magnifying glass has been hinged out of the way.*

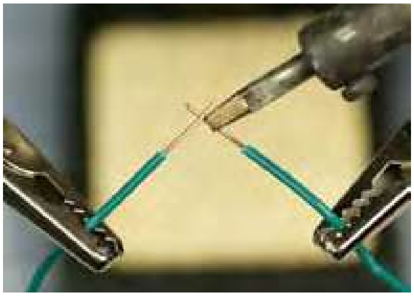

Figure 3-32

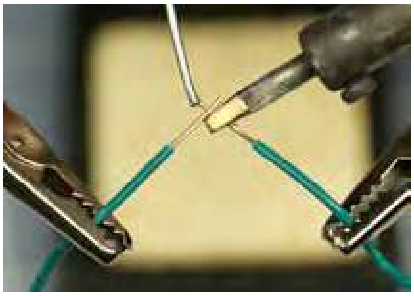

Figure 3-33

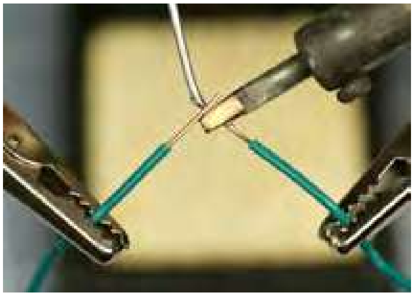

Figure 3-34

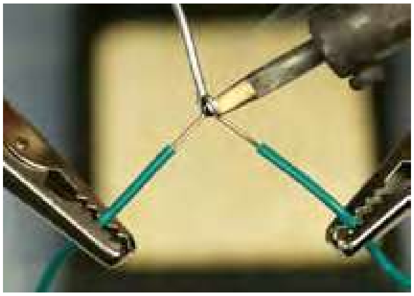

Figure 3-35. *This and the preceding three figures illustrate four steps to making a solder joint: apply heat to the wires, bring in the solder while maintaining the heat, wait for the solder to start to melt, and wait a moment longer for it to form a completely molten bead. The whole process should take between 4 and 6 seconds.*

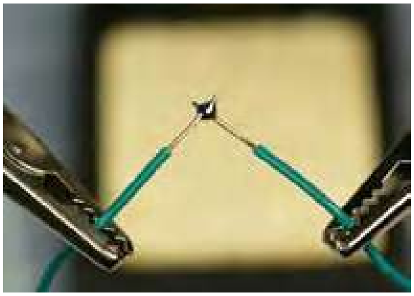

Figure 3-36. *The completed joint should be shiny, uniform, and rounded in shape.*

BACKGROUND

Soldering myths

Myth #1: Soldering is very difficult.

Millions of people have learned how to do it, and statistically, you are unlikely to be less coordinated than all of them. I have a lifelong problem with a tremor in my hands that makes it difficult for me to hold small things steadily. I also get impatient with repetitive detail work. If I can solder components, almost anyone should be able to.

Myth #2: Soldering involves poisonous chemicals.

Modern solder contains no lead. You should avoid inhaling the fumes for prolonged periods, but that also applies to everyday products such as bleach and paint. If soldering was a significant health hazard, we should have seen a high death rate among electronics hobbyists decades ago.

Myth #3: Soldering is hazardous.

A soldering iron is less hazardous than the kind of iron that you might use to iron a shirt, because it delivers less heat. In fact, in my experience, soldering is safer than most activities in a typical home or basement workshop. That doesn't mean you can be careless!

Soldering alternatives

As recently as the 1950s, connections inside electronic appliances such as radio sets were still being hand-soldered by workers on production lines. But the growth of telephone exchanges created a need for a faster way to make large numbers of rapid, reliable point-to-point wiring connections, and "wire wrap" became a viable alternative.

In a wire-wrapped electronics project, components are mounted on a circuit board that has long, gold-plated, sharp-cornered square pins sticking out of the rear. Special silver-plated wire is used, with an inch of insulation stripped from its ends. A manual or power-driven wire-wrap tool twirls the end of a wire around one of the pins, applying sufficient tension to "cold-weld" the soft silver plating of the wire to the pin. The wrapping process exerts sufficient pressure to make a very reliable joint, especially as 7 to 9 turns of wire are applied, each turn touching all four corners of the pin.

During the 1970s and 1980s, this system was adopted by hobbyists who built their own home computers. A wire-wrapped circuit board from a hand-built computer is shown in Figure 3-37. The technique was used by NASA to wire the computer in the Apollo spacecraft that went to the moon, but today, wire-wrapping has few commercial applications.

The widespread industrial use of "through-hole" components, such as the chips on early desktop computers, encouraged development of wave soldering, in which a wave or waterfall of molten solder is applied to the underside of a preheated circuit board where chips have been inserted. A masking technique prevents the solder from sticking where it isn't wanted.

Today, surface-mount components (which are significantly smaller than their through-hole counterparts) are glued to a circuit board with a solder paste, and the entire assembly is then heated, melting the paste to create a permanent connection.

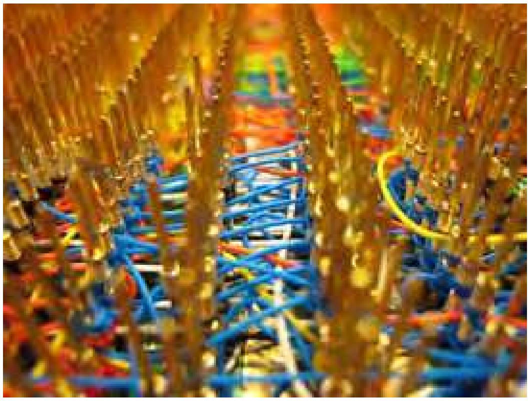

Figure 3-37. *This picture shows some of the wire-wrapping in Steve Chamberlin's custom-built, retro 8-bit CPU and computer. "Back in the day," connecting such a network of wires with solder joints would have been unduly time-consuming and prone to faults. Photo credit: Steve Chamberlin.*

TOOLS

Eight most common soldering errors

1. Not enough heat.

   The joint looks OK, but because you didn't apply quite enough heat, the solder didn't melt sufficiently to realign its internal molecular structure. It remained granular instead of becoming a solid, uniform blob, and you end up with a "dry joint," also known as a "cold joint," which will come apart when you pull the wires away from each other. Reheat the joint thoroughly and apply new solder.

   A leading cause of underheated solder is the temptation to use the soldering iron to carry solder to the joint. This results in the cold wires reducing the temperature of the solder. What you should do is touch the soldering iron to heat the wires first, and then apply the solder. This way, the wires are hot and help to melt the solder, which wants to stick to them.

   Because this is such a universal problem, I'll repeat myself: *Never melt solder on the tip of the iron and then use it to carry the solder to the joint.*

   You don't want to put hot solder on cold wires. You want to put cold solder on hot wires.

2. Too much heat.

   This may not hurt the joint, but can damage everything around it. Vinyl insulation will melt, exposing the wire and raising the risk of short circuits. You can easily damage semiconductors, and may even melt the internal plastic components of switches and connectors.

   Damaged components must be desoldered and replaced, which will take time and tends to be a big hassle (see "Tools: Desoldering" on page 109 for advice).

3. Not enough solder.

   A thin connection between two conductors may not be strong enough. When joining two wires, always check the underside of the joint to see whether the solder penetrated completely.

4. Moving the joint before the solder solidifies.

   You may create a fracture that you won't necessarily see. It may not stop your circuit from working, but at some point in the future, as a result of vibration or thermal stresses, the fracture can separate just enough to break electrical contact. Tracking it down will then be a chore. If you clamp components before you join them, or use perforated board to hold the components steady, you can avoid this problem.

5. Dirt or grease.

   Electrical solder contains rosin that cleans the metal that you're working with, but contaminants can still prevent solder from sticking. If any component looks dirty, clean it with fine sandpaper before joining it.

6. Carbon on the tip of your soldering iron.

   The iron gradually accumulates flecks of black carbon during use, and they can act as a barrier to heat transfer. Wipe the tip of the iron on the little sponge mounted in the base of your soldering iron stand or your helping hand.

7. Inappropriate materials.

   Electronic solder is designed for electronic components. It will not work with aluminum, stainless steel, or various other metals. You may be able to make it stick to chrome-plated items, but only with difficulty.

8. Failure to test the joint.

   Don't just assume that it's OK. Always test it, by applying manual force if you can (see Figures 3-38 and 3-39 for the ideal protocol) or, if you can't get a grip on the joint, slip a screwdriver blade under it and flex it just a little, or use small pliers to try to pull it apart. Don't be concerned about ruining your work. If your joint doesn't survive rough treatment, it wasn't a good joint.

Of the eight errors, dry/cold joints are by far the worst, because they are easy to make and can look OK.

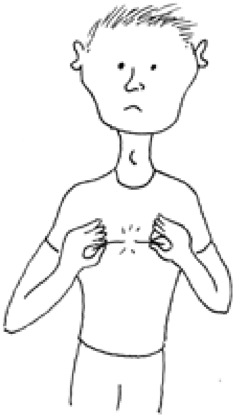

Figure 3-38. *Test result of a bad solder joint.*

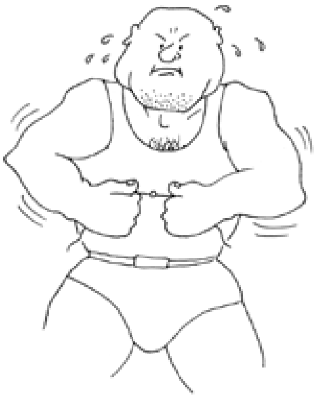

Figure 3-39. *Test result of a good solder joint.*

Your Second Solder Joint

Time now to try your pencil-style soldering iron. Once again, you must leave it plugged in for a good five minutes to make sure it's hot enough. In the meantime, don't forget to unplug your other soldering iron, and put it somewhere safe while it cools.

This time I'd like you to align the wires parallel with each other. Joining them this way is a little more difficult than joining them when they cross each other, but it's a necessary skill. Otherwise, you won't be able to slide heat-shrink tubing over the finished joint to insulate it.

Figures 3-40 through 3-44 show a successful joint of this type. The two wires do not have to make perfect contact with each other; the solder will fill any small gaps. But the wires must be hot enough for the solder to flow, and this can take an extra few seconds when you use the low-wattage pencil-style iron.

Be sure to feed the solder in as shown in the pictures. Remember: don't try to carry the solder to the joint on the tip of the iron. Heat the wires first, and then touch the solder to the wires and the tip of the iron, while keeping it in contact with the wires. Wait until the solder liquifies, and you will see it running eagerly into the joint. If this doesn't happen, be more patient and apply the heat for a little longer.

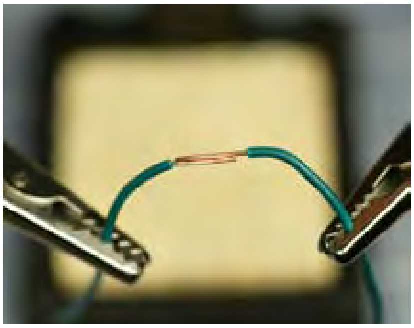

Figure 3-40

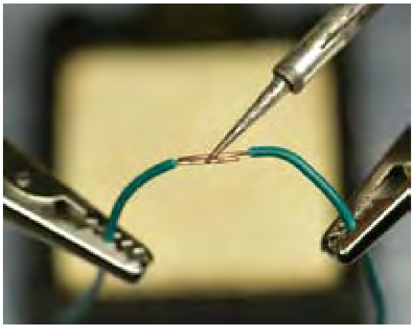

Figure 3-41

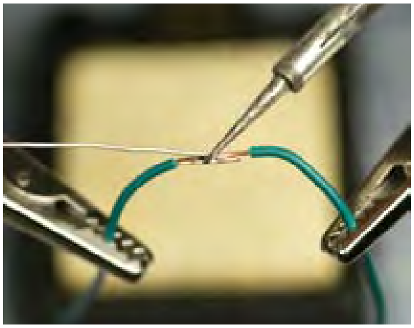

Figure 3-42

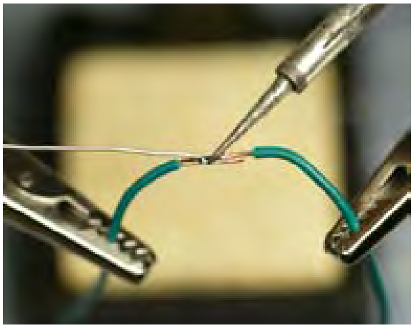

Figure 3-43. *This and the preceding three figures show how joining two wires that are parallel is more difficult, and the low-wattage, pencil-type soldering iron will require longer to heat them sufficiently for a good joint. Thinner solder can be used.*

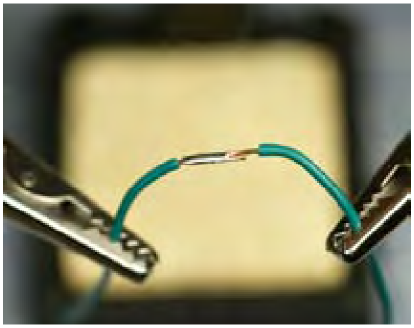

Figure 3-44. *The finished joint has enough solder for strength, but not so much solder that it will prevent heat-shrink tubing from sliding over it.*

THEORY

Soldering theory

The better you understand the process of soldering, the easier it should be for you to make good solder joints.

The tip of the soldering iron is hot, and you want to transfer that heat into the joint that you are trying to make. In this situation, you can think of the heat as being like a fluid. The larger the connection is between the soldering iron and the joint, the greater the quantity of heat, per second, that can flow through it.

For this reason, you should adjust the angle of the soldering iron so that it makes the widest possible contact. If it touches the wires only at a tiny point, you'll limit the amount of heat flow. Figures 3-45 and 3-46 illustrate this concept. Once the solder starts to melt, it broadens the area of contact, which helps to transfer more heat, so the process accelerates naturally. Initiating it is the tricky part.

The other aspect of heat flow that you should consider is that it can suck heat away from the places where you want it, and deliver it to places where you don't want it. If you're trying to solder a very heavy piece of copper wire, the joint may never get hot enough to melt the solder, because the heavy wire conducts heat away from the joint. You may find that even a 40-watt iron isn't powerful enough to overcome this problem, and if you are doing heavy work, you may need a more powerful iron.

As a general rule, if you can't complete a solder joint in 10 seconds, you aren't applying enough heat.

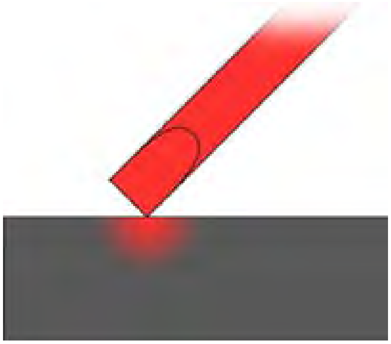

Figure 3-45. *With only a small surface area of contact between the iron and the working surface, an insufficient amount of heat is transferred.*

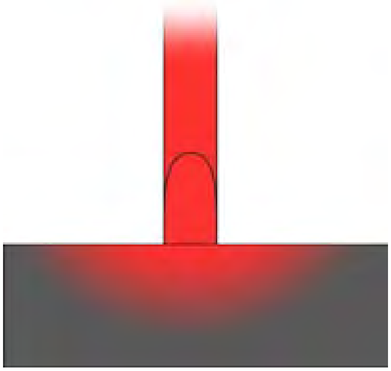

Figure 3-46. *A larger area of contact between the soldering iron and its target will greatly increase the heat transfer.*

TOOLS

Desoldering

Desoldering is much, much harder than soldering. Two simple tools are available:

- *Suction pump*. First, you apply the soldering iron to make the solder liquid. Then you use this simple gadget to try to suck up as much of the liquid as possible. Usually it won't remove enough metal to allow you to pull the joint apart, and you will have to try the next tool. Refer back to Figure 3-10.
- *Desoldering wick or braid*. Desoldering wick, also known as braid, is designed to soak up the solder from a joint, but again, it won't clean the joint entirely, and you will be in the awkward position of trying to use both hands to pull components apart while simultaneously applying heat to stop the solder from solidifying. Refer back to Figure 3-11.

I don't have much advice about desoldering. It's a frustrating experience (at least, I think so) and can damage components irrevocably.

Heat Guns Get Hot, Too!

Notice the chromed steel tube at the business end of your heat gun. Steel costs more than plastic, so the manufacturer must have put it there for a good reason—and the reason is that the air flowing through it becomes so hot that it would melt a plastic tube.

The metal tube stays hot enough to burn you for several minutes after you've used it. And, as in the case of soldering irons, other people (and pets) are vulnerable, because they won't necessarily know that the heat gun is hot. Most of all, make sure that no one in your home ever makes the mistake of using a heat gun as a hair dryer (Figure 3-47).

This tool is just a little more hazardous than it appears.

Adding Insulation

After you've succeeded in making a good inline solder connection between two wires, it's time for the easy part. Choose some heat-shrink tubing that is just big enough to slide over the joint with a little bit of room to spare.

Figure 3-47. *Other members of your family should understand that although a heat gun looks like a hair dryer, appearances may be deceptive.*

Slide the tubing along until the joint is centered under it, hold it in front of your heat gun, and switch on the gun (keeping your fingers away from the blast of superheated air). Turn the wire so that you heat both sides. The tubing should shrink tight around the joint within half a minute. If you overheat the tubing, it may shrink so much that it splits, at which point you must remove it and start over. As soon as the tubing is tight around the wire, your job is done, and there's no point in making it any hotter. Figures 3-48 through 3-50 show the desired result. I used white tubing because it shows up well in photographs. Different colors of heat-shrink tubing all perform the same way.

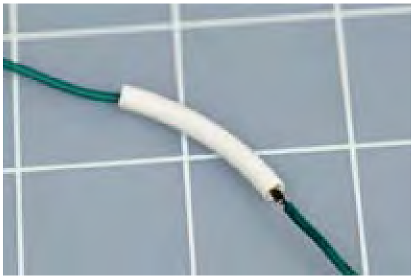

Figure 3-48. *Slip the tubing over your wire joint.*

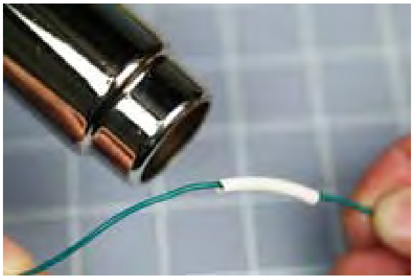

Figure 3-49. *Apply heat to the tubing.*

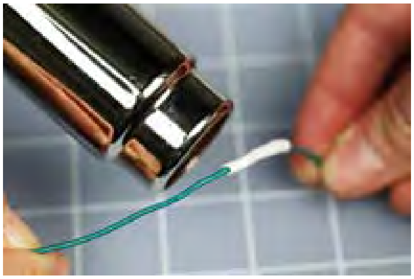

Figure 3-50. *Leave the heat on the tubing until it shrinks to firmly cover the joint.*

I suggest you next practice your soldering skills on a couple of practical projects. In the first one, you can add color-coded, solid-core wires to your AC adapter, and in the second one, you can shorten the power cord for a laptop power supply. You can use your larger soldering iron for both of these tasks, because neither of them involves any heat-sensitive components.

Modifying an AC Adapter

In the previous chapter, I mentioned the irritation of being unable to push the wires from your AC adapter into the holes of your breadboard. So, let's fix this right now:

1. Cut two pieces of solid-conductor 22-gauge wire—one of them red, the other black or blue. Each should be about 2 inches long. Strip a quarter-inch of insulation from both ends of each piece of wire.
2. Trim the wire from your AC adapter. You need to expose some fresh, clean copper to maximize your chance of getting the solder to stick.

   I suggest that you make one conductor longer than the other to minimize the chance of the bare ends touching and creating a short circuit. Use your meter, set to DC volts, if you have any doubt about which conductor is positive.

Solder the wires and add heat-shrink tubing as you did in the practice session. The result should look like Figure 3-51.

Choose the Right Tubing

If you use heat-shrink tubing on 110v AC cord, as is being done in this experiment, make sure you use tubing that's been rated for 110v use.

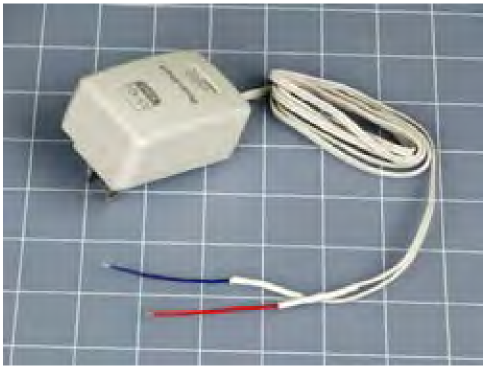

Figure 3-51. *Solid-core color-coded wires, soldered onto the wires from an AC adapter, provide a convenient way to feed power to a breadboard. Note that the wires are of differing lengths to reduce the risk of them touching each other.*

Shortening a Power Cord

When I travel, I like to minimize everything. It always annoys me that the power cord for the power supply of my laptop is 4 feet long. The thinner wire that connects the power supply to the computer is also 4 feet long, and I just don't need that much wire.

After searching exhaustively I couldn't find any laptop power cables shorter than 3 feet, so I decided to shorten one myself. If you feel no need to do this, you should try the following procedure on an old extension cord, just as an exercise. You do need to go through these steps to acquire some practice in soldering heavier, stranded wire and using heat-shrink tubing:

1. Use your wire cutters to chop the wire, and then a utility knife to split the two conductors, with one shorter than the other. When splicing a power cord or similar cable containing two or more conductors, it's good to avoid having the joints opposite each other. They fit more snugly if they are offset, and there's less risk of a short circuit if a joint fails.

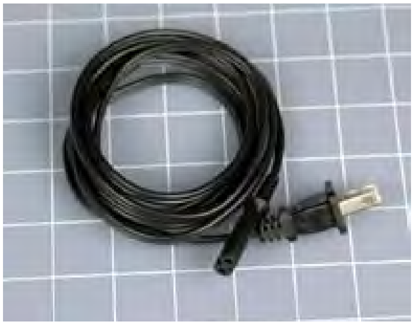

Figure 3-52

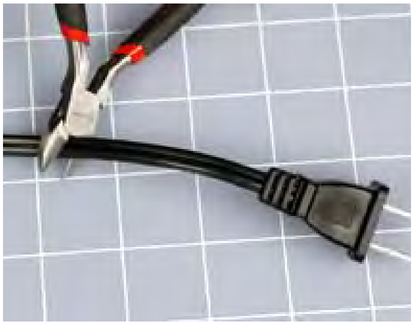

Figure 3-53

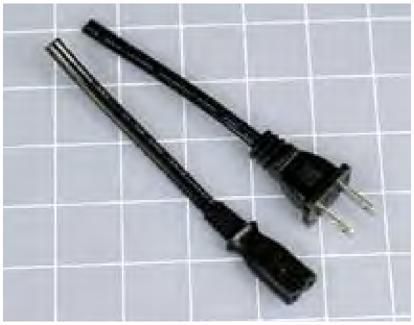

Figure 3-54

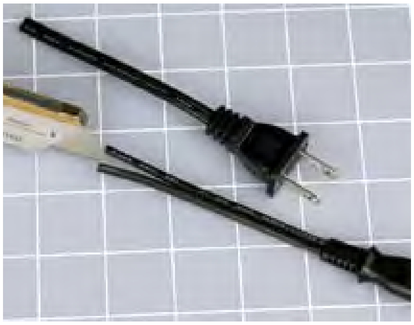

Figure 3-55

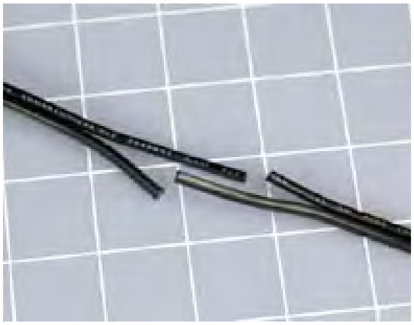

Figure 3-56

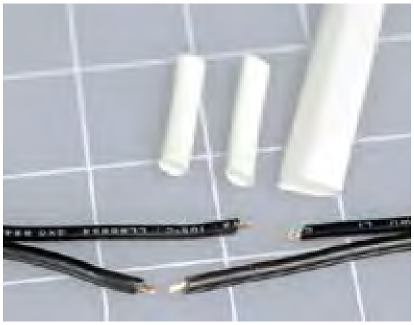

Figure 3-57

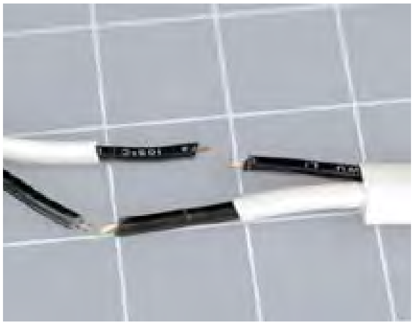

Figure 3-58. *Figures 3-52 through 3-58 illustrate the sequence of steps to prepare for making a shortened power cord for a laptop computer power supply.*

2. Strip off a minimal amount of insulation. One-eighth of an inch (3 mm) is sufficient. The automatic wire strippers that I mentioned in the shopping list in Chapter 1 are especially convenient, but regular wire strippers will do the job.
3. Cut two pieces of heat-shrink tubing, each 1 inch long, big enough to slide over the separate conductors in your cable. Cut a separate 2-inch piece of larger tubing that will slide over the entire joint when it's done. The steps described so far are illustrated in Figures 3-52 through 3-58.
4. Now for the most difficult part: activating your human memory. You have to remember to slide the tubing onto the wire *before* you make your solder joint, because the plugs on the ends of the wires will prevent you from adding any heat-shrink tubing later. If you're as impatient as I am, it's very difficult to remember to do this every time.

5. Use your helping hand to align the first joint. Push the two pieces of wire together so that the strands intermingle, and then squeeze them tight between finger and thumb, so that there are no little bits sticking out. A stray strand of wire can puncture heat-shrink tubing when the tubing is hot and soft and is shrinking around the joint.
6. The wire that you're joining is much heavier than the 22-gauge wire that you worked with previously, so it will suck up more heat, and you must touch the soldering iron to it for a longer time. Make sure that the solder flows all the way into the joint, and check the underside after the joint is cool. Most likely you'll find some bare copper strands there. The joint should become a nice solid, rounded, shiny blob. Keep the heat-shrink tubing as far away from the joint as possible while you're using the soldering iron, so that heat from the iron doesn't shrink the tubing prematurely, preventing you from sliding it over the joint later.

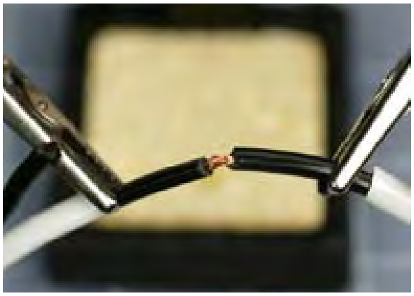

Figure 3-59

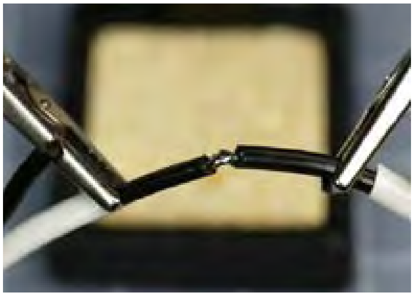

Figure 3-60

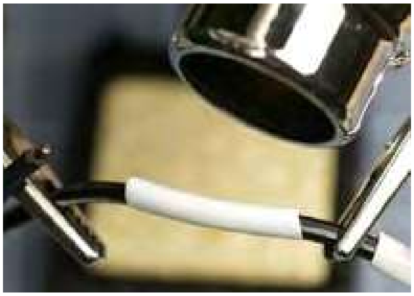

Figure 3-61

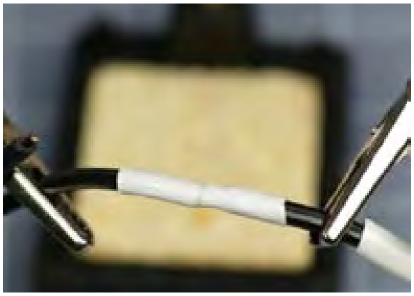

Figure 3-62

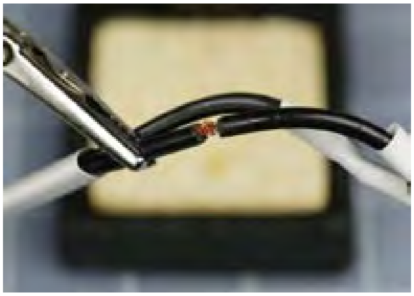

Figure 3-63

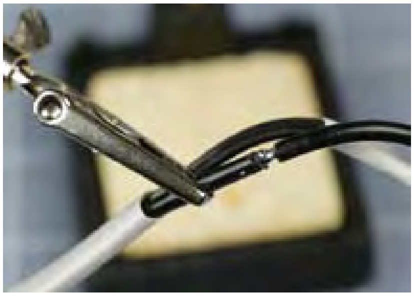

Figure 3-64

7. When the joint has cooled, slide the heat-shrink tubing over it, and apply the heat gun. Now repeat the process with the other conductor. Finally, slide the larger piece of tubing over the joint. You did remember to put the large tubing onto the wire at the beginning, didn't you?

Figures 3-59 through 3-65 show the steps all the way through to the end.

If you have completed the soldering exercises so far, you now have sufficient basic skills to solder your first electronic circuit. But first, I want you to verify the vulnerability of components to heat.

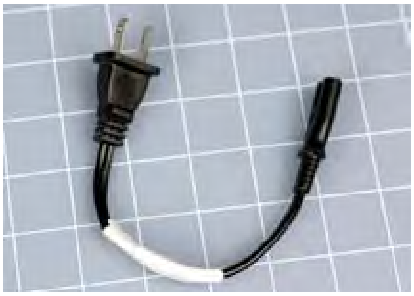

Figure 3-65. *Completion of the shortened power cord for a laptop power supply.*

Experiment 13: Broil an LED

In Chapter 1, you saw how an LED can be damaged if too much current flows through it. The electricity caused heat, which melted the LED. Unsurprisingly, you can just as easily melt it by applying too much heat to one of its leads with a soldering iron. The question is: how much heat is too much? Let's find out.

You will need:

- 30-watt or 40-watt soldering iron
- 15-watt pencil-type soldering iron
- A couple of LEDs (that are expendable)
- 680Ω resistor
- Wire cutters and sharp-nosed pliers
- "Helping hand" gadget to hold your work

I don't want you to use alligator clips to join the LED to a power supply, because the alligator clip will divert and absorb some of the heat from your soldering iron. Instead, please use some sharp-nosed pliers to bend each of the leads from an LED into little hooks, and do the same thing with the wires on a 680Ω load resistor. Finally bend the new wires on your AC adapter so that they, too, are tiny hooks. Now you can put the hooks together like links in a chain, as shown in Figure 3-66.

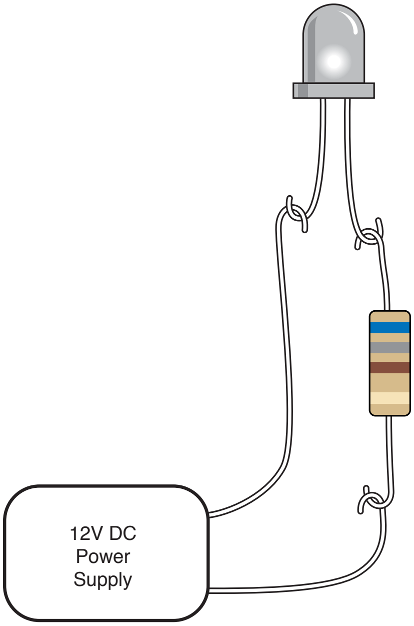

Figure 3-66. *By literally hooking together the leads from a resistor and a white-light LED, we minimize pathways for heat to escape during the subsequent test.*

Grip the plastic body of the LED in your helping hand. Plastic is not a good thermal conductor, so the helping hand shouldn't siphon too much heat away from our target. The resistor can dangle from one of the leads on the LED, and the wire from the AC adapter can hang from that, a little farther down. Gravity should be sufficient to make this work. Set your AC adapter to deliver 12 volts as before, plug it in, and your LED should be shining brightly. I used a white LED in this experiment, because it's easier to photograph.

Make sure your two soldering irons are really hot. They should have been plugged in for at least five minutes. Now take the pencil-style iron and hold its tip firmly against one of the leads on your glowing LED, while you check the time with a watch. Figure 3-67 shows the setup.

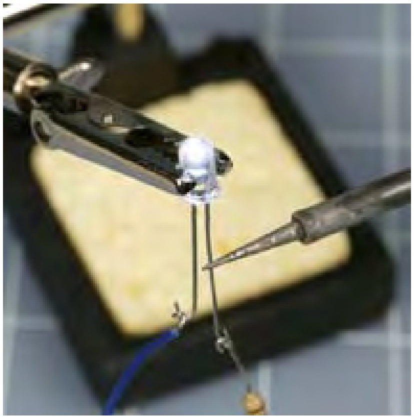

Figure 3-67. *Applying heat with a 15-watt soldering iron. A typical LED should withstand this treatment for two or three minutes, but if you substitute a 30-watt soldering iron, the LED is likely to burn out in under 15 seconds.*

I'm betting that you can sustain this contact for a full three minutes without burning out the LED. This is why you use a 15-watt soldering iron for delicate electronics work—it doesn't endanger the components.

Allow your LED wire to cool, and then apply your more powerful soldering iron to the same piece of wire as before. Again, make sure it is completely hot, and I think you'll find that the LED will go dark after as little as 10 seconds (note, some LEDs can survive higher temperatures than others). This is why you *don't* use a 30-watt soldering iron for delicate electronics work.

The large iron doesn't necessarily reach a higher temperature than the small one. It just has a larger heat capacity. In other words, a greater quantity of heat can flow out of it, at a faster rate.

Throw away your burned-out LED. Substitute a new one, connected as before, but add a full-size copper alligator clip to one of the leads up near the body of the LED, as shown in Figure 3-68. Press the tip of your 30-watt or 40-watt soldering iron against the lead just *below* the alligator clip. This time, you should be able to hold the powerful soldering iron in place for a full two minutes without burning out the LED.

Imagine the heat flowing out through the tip of your soldering iron, into the wire that leads to the LED—except that the heat meets the alligator clip along the way, as shown in Figure 3-69. The clip is like an empty vessel waiting to be filled. It offers much less resistance to heat than the remainder of the wire leading to the LED, so the heat prefers to flow into the copper clip, leaving the LED unharmed. At the end of your experiment, if you touch the clip, you'll find that it's hot, while the LED remains relatively cooler.

The alligator clip is known as a heat sink, and it should be made of copper, because copper is one of the best conductors of heat.

Because the 15-watt soldering iron failed to harm the LED, you may conclude that the 15-watt iron is completely safe, eliminating all need for a heat sink. Well, this may be true. The problem is, you don't really know whether some semiconductors may be more heat-sensitive than LEDs. Because the consequences of burning out a component are so exasperating, I suggest you should play it safe and use a heat sink in these circumstances:

- If you apply 15-watt iron extremely close to a semiconductor for 20 seconds or more.
- If you apply a 30-watt iron near resistors or capacitors for 10 seconds or more. (Never use it near semiconductors.)
- If you apply a 30-watt iron near anything meltable for 20 seconds or more. Meltable items include insulation on wires, plastic connectors, and plastic components inside switches.

Rules for Heat Sinking

1. Full-size copper alligator clips do work better.
2. Clamp the alligator clip as close as possible to the component and as far as possible from the joint. (You don't want to suck too much heat away from the joint.)
3. Make sure there is a metal-to-metal connection between the alligator clip and the wire to promote good heat transfer.

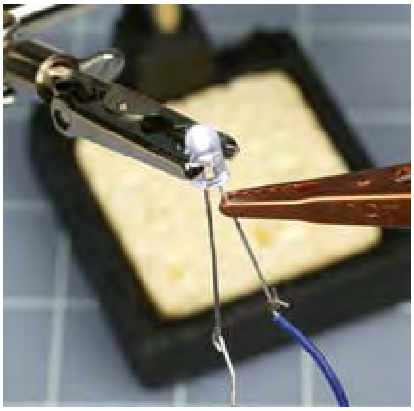

Figure 3-68. *When a copper alligator clip is used as a heat sink, you should be able to apply a 30-watt soldering iron (below the clip) without damaging the LED.*

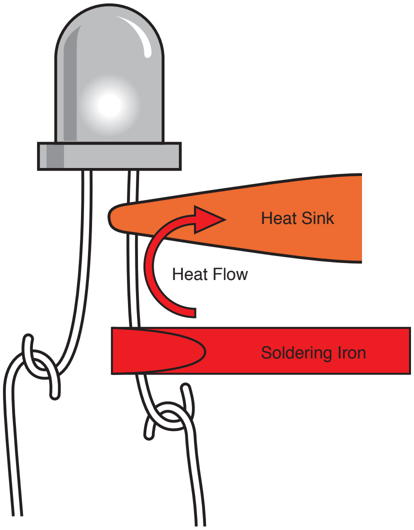

Figure 3-69. *The heat sink intercepts the heat, sucks it up, and protects the LED from damage.*

FUNDAMENTALS

All about perforated board

For the remainder of this book, you'll be using perforated board whenever you want to create permanent, soldered circuits. There are three ways to do this:

1. Point-to-point wiring. You use perforated board that has no connections behind the holes. Either the board has no copper traces on it at all, as in Figure 3-70, or you will find a little circular copper circle around each hole, as in Figure 3-71. These circles are not connected with each other and are used only to stabilize the components that you assemble.

   Point-to-point wiring allows you to place the components in a convenient, compact layout that can be very similar to a schematic. Under the board you bend the wires to link the components, and solder them together, adding extra lengths of wire if necessary. The advantage of this system is that it can be extremely compact. The disadvantage is that the layout can be confusing, leading to errors.

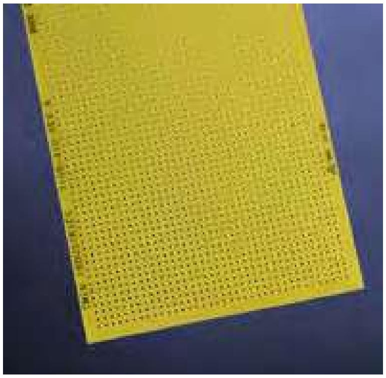

Figure 3-70

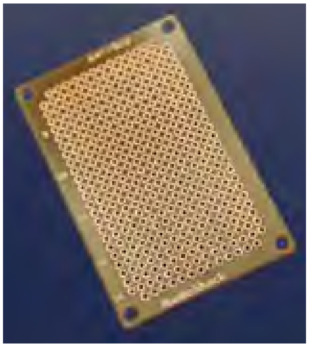

Figure 3-71. *Either this type of perforated board or the type in Figure 3-70 can be used for point-to-point wiring in Experiment 14.*

2. Breadboard-style wiring. Use perforated board that is printed with copper traces in exactly the same pattern as the conductors inside a breadboard. Once your circuit works on the breadboard, you move the components over to the perfboard one by one, maintaining their exact same positions relative to each other. You solder the "legs" of the components to the copper traces, which complete the circuit. Then you trim off the surplus wire. The advantage of this procedure is that it's quick, requires very little planning, and minimizes the possibility for errors. The disadvantage is that it tends to waste space. A cheap example is shown in Figure 3-72.
3. You can etch your own circuit board with customized copper traces that link your components in a point-to-point layout. This is the most professional way to complete a project, but it requires more time, trouble, and equipment than is practical in this book.

   Point-to-point wiring is like working with alligator clips, on a much smaller scale. The first soldered project will use this procedure.

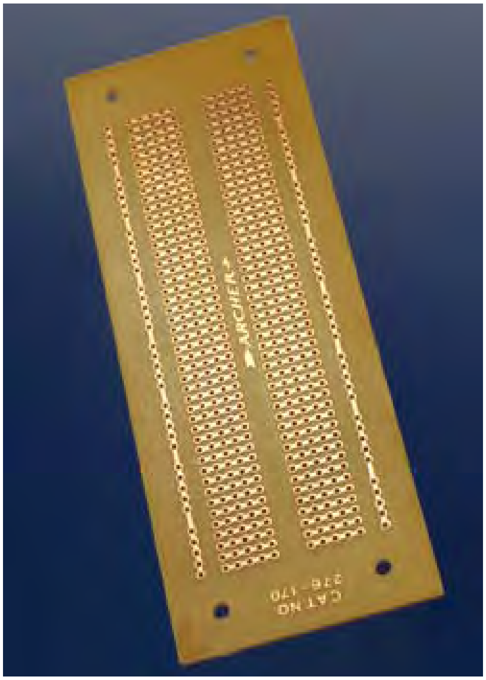

Figure 3-72. *Perforated board etched with copper in variants of a breadboard layout. This example is appropriate for Experiment 15.*

Experiment 14: A Pulsing Glow

You will need:

- Breadboard
- 15-watt pencil-type soldering iron
- Thin solder (0.022 inches or similar)
- Wire strippers and cutters
- Plain perforated board (no copper etching necessary)
- Small vise or clamp to hold your perforated board
- Resistors, various
- Capacitors, electrolytic, 100 µF and 220 µF, one of each
- Red LED, 5 mm, rated for 2 volts approximately
- 2N6027 programmable unijunction transistor

Your first circuit using a PUT was a slow-speed oscillator that made an LED flash about twice each second. The flashes looked very "electronic," by which I mean that the LED blinked on and off without a gradual transition between each state. I'm wondering if we can modify this circuit to make the LED pulse in a more gentle, interesting way, like the warning light on an Apple MacBook when it's in "sleep" mode. I'm thinking that something of this sort might be wearable as an ornament, if it's small enough and elegant enough.

I'm also thinking that this first soldering project will serve three purposes. It will test and refine your skill at joining wires together, will teach you point-to-point wiring with perfboard, and will give you some additional insight into the way that capacitors can be used to adjust timing.

Look back at the original schematic in Experiment 11, on page 82. Refresh your memory about the way it worked. The capacitor charges through a resistor until it has enough voltage to overcome the internal resistance in the PUT. Then the capacitor discharges through the PUT and flashes the LED.

If you drew a graph of the light coming out of the LED, it would be a thin, square-shaped pulse, as shown in Figure 3-73. How can we fill it out to make it more like the curve in Figure 3-74, so that the LED fades gently on and off, like a heartbeat?

One thing is obvious: the LED is going to be emitting a greater total amount of light in each cycle. Therefore, it's going to need more power. This means that C1, in Figure 3-75, must be a larger capacitor.

When we have a larger capacitor, it takes longer to charge. To keep the flashes reasonably frequent, we'll need a lower-value resistor for R1 to charge the capacitor quickly enough. In addition, reducing the values of R2 and R3 will program the PUT to allow a longer pulse.

Most important, I want to discharge the capacitor through a resistor to make the onset of the pulse gradual instead of sudden. Remember, when you have a resistor in series with a capacitor, the capacitor not only charges more slowly, but discharges more slowly.

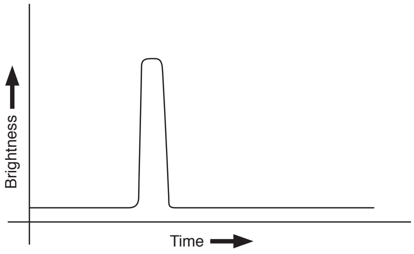

Figure 3-73

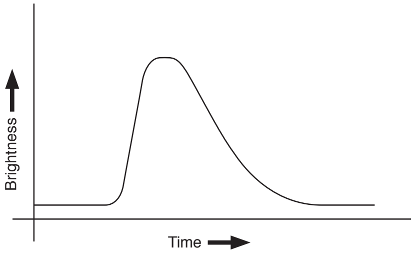

Figure 3-74. *The original PUT oscillator circuit in Experiment 11 made the LED emit sharp, short flashes. The upper graph shows what we might find if we measured light output over time. The second graph shows a gentler onset to each flash, followed by a slow fade-out. Capacitors can be used to create this effect.*

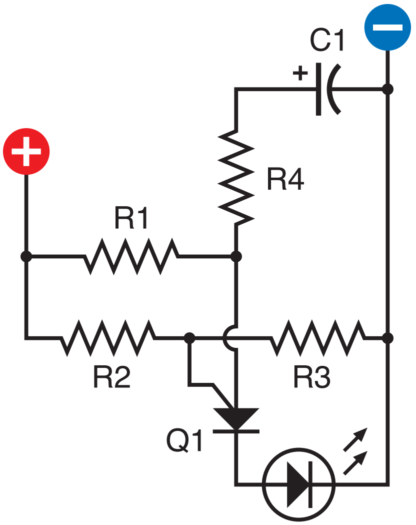

Figure 3-75. *The first step toward creating a gentler flashing effect is to use a larger capacitor for C1 and discharge it through a resistor, R4. Lower-value resistors are necessary to charge the capacitor rapidly enough.*

R1: 33K
R2: 1K
R3: 1K
R4: 1K
C1: 100 μF electrolytic
Q1: 2N6027

Figure 3-75 shows these features. Compare it with Figure 2-103 on page 85. R1 is now 33K instead of 470K. R2 and R3 are reduced to 1K. R4 also is 1K, to so that the capacitor takes longer to discharge through it. And C1 is now 100 µF instead of 2.2 µF.

Assemble this circuit on a breadboard, and compare the results when you include R4 or bypass it with a plain piece of jumper wire. It softens the pulse a bit, but we can work on it some more. On the output side of the PUT, we can add another capacitor. This will charge itself when the pulse comes out of the PUT, and then discharge itself gradually through another resistor, so that the light from the LED dies away more slowly.

Figure 3-76 shows the setup. C2 is large—220 µF—so it sucks up the pulse that comes out of the PUT, and then gradually releases it through 330Ω resistor R5 and the LED. You'll see that the LED behaves differently now, fading out inside of blinking off. But the resistances that I've added have dimmed the LED, and to brighten it, you should increase the power supply from 6 volts to 9 volts.

Remember that a capacitor imposes a smoothing effect only if one side of it is grounded to the negative side of the power supply. The presence of the negative charge on that side of the capacitor attracts the positive pulse to the other side.

I like the look of this heartbeat effect. I can imagine a piece of wearable electronic jewelry that pulses in this sensual way, very different from the hard-edged, sharp-on-and-off of a simple oscillator circuit. The only question is whether we can squeeze the components into a package that is small enough to wear.

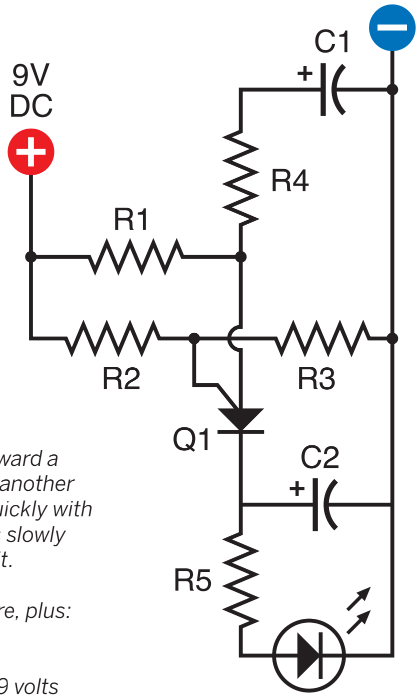

Figure 3-76. *The second step toward a gentler flashing effect is to add another capacitor, C2, which charges quickly with each pulse and then discharges slowly through R5 and the LED below it.*

Same components as before, plus:
R5: 330Ω
C2: 220 μF electrolytic
Power supply increased to 9 volts

Figure 3-77. *On a dark night in a rural area, the heartbeat flasher may be attractive in unexpected ways.*

Resizing the Circuit

The first step is to look at the physical components and imagine how to fit them into a small space. Figure 3-78 shows a 3D view of a compact arrangement. Check this carefully, tracing all the paths through the circuit, and you'll see that it's the same as the schematic. The trouble is that if we solder the components together like this, they won't have much strength. All the little wires can bend easily, and there's no easy way to mount the circuit in something or on something.

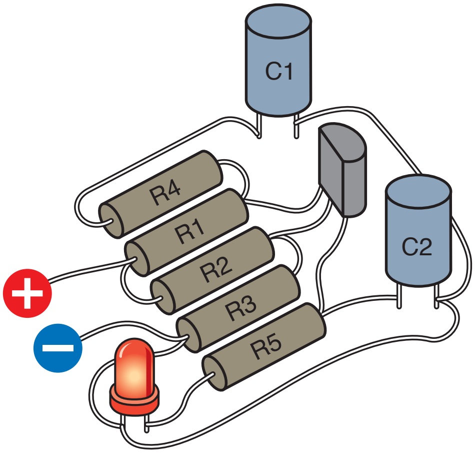

Figure 3-78. *This layout of components replicates their connections in the schematic diagram while squeezing them into a minimal amount of space.*

The answer is to put it on a substrate, which is one of those terms that people in the electronics field like to use, perhaps because it sounds more technical than "perfboard." But perforated board is what we need, and Figure 3-79 shows the components transferred onto a piece of board measuring just 1 inch by 0.8 inch.

The center version of this diagram uses dotted lines to show how the components will be connected with each other underneath the board. Mostly the leads that stick out from underneath the components will be long enough to make these connections.

Finally, the bottom version of the perfboard diagram shows the perfboard flipped left-to-right (notice the L and the R have been transposed to remind you, and I've used a darker color to indicate the underside of the board). Orange circles indicate where solder joints will be needed.

The LED should be unpluggable, because we may want to run it at some distance from the circuit. Likewise the power source should be unpluggable. Fortunately we can buy miniature connectors that fit right into the perforated board. You may have to go to large online retail suppliers such as Mouser.com for these. Some manufacturers call them "single inline sockets and headers," while others call them "boardmount sockets and pinstrip headers." Refer back to Figure 3-29 and check the shopping list for more details.

This is a very compact design that will require careful work with your pencil-style soldering iron. Because a piece of perforated board as small as this will tend to skitter around, I suggest that you apply your miniature vise to one end to anchor it with some weight while still allowing you to turn it easily.

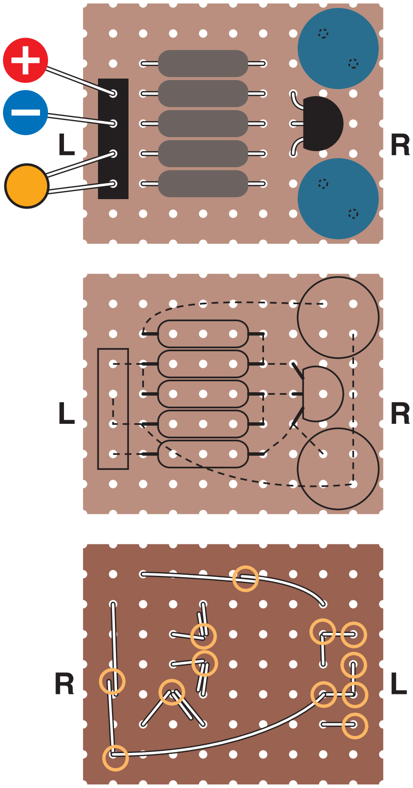

Figure 3-79. *Perforated board can be used to support the layout of components. Their leads are soldered together under the board to create the circuit. The middle diagram shows the wires under the board as dashed lines. The bottom diagram shows the board from underneath, flipped left to right. Orange circles indicate where solder joints will be necessary.*

When I'm working on this kind of project, I like to place it (with the vise attached) on a soft piece of polyurethane foam—the kind of slab that is normally used to make a chair cushion. The foam protects the components from damage when the board is upside-down, and again helps to prevent the work from sliding around unpredictably.

Step by Step

Here's the specific procedure for building this circuit:

1. Cut the small piece of perfboard out of a sheet that has no copper traces on it. You can cut the section using your miniature hobby saw, or you may be able to snap the board along its lines of holes, if you're careful. Alternatively, use a small ready-cut piece of perfboard with copper circles on it that are not connected to one another. You'll ignore the copper circles in this project. (In the next experiment, you'll deal with the additional challenge of making connections between components and copper traces on perforated board.)
2. Gather all the components and carefully insert them through holes in the board, counting the holes to make sure everything is in the right place. Flip the board over and bend the wires from the components to anchor them to the board and create connections as shown. If any of the wires isn't long enough, you'll have to supplement it with an extra piece of 22-gauge wire from your supply. You can remove all the insulation, as we'll be mounting the perfboard on a piece of insulating plastic.
3. Trim the wires approximately with your wire cutters.
4. Make the joints with your pencil soldering iron. Note that in this circuit, you are just joining wires to each other. The components are so close together that they'll prevent each other from wiggling around too much. If you are using board with copper pads (as I did), and some solder connects with them, that's OK—as long as it doesn't creep across to the neighboring component and create a short circuit.
5. Check each joint using a close-up magnifying glass, and wiggle it with pointed-nosed pliers. If there isn't enough solder for a really secure joint, reheat it and add more. If solder has created a connection that shouldn't be there, use a utility knife to make two parallel cuts in the solder, and scrape away the little section between them.

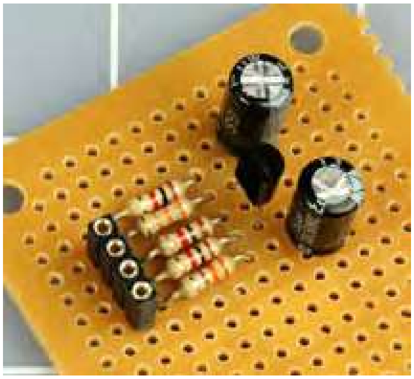

Figure 3-80. *Components mounted on a piece of perforated board.*

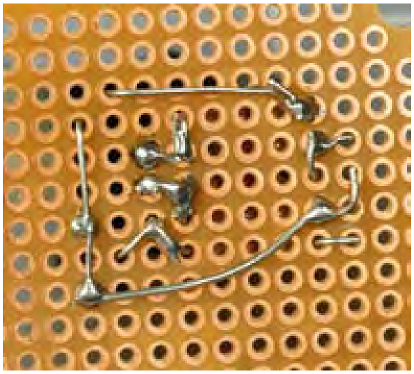

Figure 3-81. *The assembly seen from below. The copper circles around the holes are not necessary for this project. Some of them have picked up some solder, but this is irrelevant as long as no unintentional short circuits are created.*

Generally, I insert three or four components, trim the wires approximately, solder them, trim their wires finally, then pause to check the joints and the placement. If I solder too many components in succession, there's a greater risk of missing a bad joint, and if I make an error in placing a component, undoing it will be much more problematic if I have already added a whole lot more components around it.

Figures 3-80 and 3-81 show the version of this project that I constructed, before I trimmed the board to the minimum size.

Flying Wire Segments

The jaws of your wire cutters exert a powerful force that peaks and then is suddenly released when they cut through wire. This force can be translated into sudden motion of the snipped wire segment. Some wires are relatively soft, and don't pose a risk, but harder wires can fly in unpredictable directions at high speed, and may hit you in the eye. The leads of transistors are especially hazardous in this respect.

I think it's a good idea to wear safety glasses when trimming wires.

Finishing the Job

I always use bright illumination. This is not a luxury; it is a necessity. Buy a cheap desk lamp if you don't already have one. I use a daylight-spectrum fluorescent desk lamp, because it helps me identify the colored bands on resistors more reliably. Note that this type of fluorescent lamp emits quite a lot of ultraviolet light, which is not good for the lens in your eye. Avoid looking closely and directly at the tube in the lamp, and if you wear glasses, they will provide additional protection.

No matter how good your close-up vision is, you need to examine each joint with that close-up magnifier. You'll be surprised how imperfect some of them are. Hold the magnifier as close as possible to your eye, then pick up the thing that you want to examine and bring it closer until it comes into focus.

Finally, you should end up with a working circuit. You can insert the wires from your power supply into two of the tiny power sockets, and plug a red LED into the remaining two sockets. Remember that the two center sockets are negative, and the two outer sockets are positive, because it was easier to wire the circuit this way. You should color-code them to avoid mistakes.

So now you have a tiny circuit that pulses like a heartbeat. Or does it? If you have difficulty making it work, retrace every connection and compare it with the schematic. If you don't find an error, apply power to the circuit, attach the black lead from your meter to the negative side, and then go around the circuit with the red lead, checking the presence of voltage. Every part of this circuit should show at least some voltage while it's working. If you find a dead connection, you may have made a bad solder joint, or missed one entirely.

When you're done, now what? Well, now you can stop being an electronics hobbyist and become a crafts hobbyist. You can try to figure out a way to make this thing wearable.

First you have to consider the power supply. Because of the components that I used, we really need 9 volts to make this work well. How are you going to make this 9-volt circuit wearable, with a bulky 9-volt battery?

I can think of three answers:

1. You can put the battery in a pocket, and mount the flasher on the outside of the pocket, with a thin wire penetrating the fabric. Note that the tiny power connector on the perforated board will accept two 22-gauge wires if they are solid core, or if they are stranded (like the wires from a 9-volt battery connector) but have been thinly coated with solder.
2. You could mount the battery inside the crown of a baseball cap, with the flasher on the front.
3. You can put together three 3-volt button batteries in a stack, held in some kind of plastic clip. If you try this option, it may not be a good idea to try to solder wire to a battery. You will heat the liquid stuff inside the battery, which may not be good for it, and may not be good for you if the liquid starts boiling and the battery bursts open. Also, solder doesn't stick easily to the metallic finish on most battery terminals.

Most LEDs create a sharply defined beam of light, which you may want to diffuse to make it look nicer. One way to do this is to use a piece of transparent acrylic plastic, at least 1/4 inch thick, as shown in Figure 3-82. Sandpaper the front of the acrylic, ideally using an orbital sander that won't make an obvious pattern. Sanding will make the acrylic translucent rather than transparent.

Drill a hole slightly larger than the LED in the back of the acrylic. Don't drill all the way through the plastic. Remove all fragments and dust from the hole by blasting some compressed air into it, or by washing it if you don't have an air compressor. After the cavity is completely dry, get some transparent silicone caulking or mix some clear five-minute epoxy and put a drop in the bottom of the hole. Then insert the LED, pushing it in so that it forces the epoxy to ooze around it, making a tight seal. See Figure 3-82.

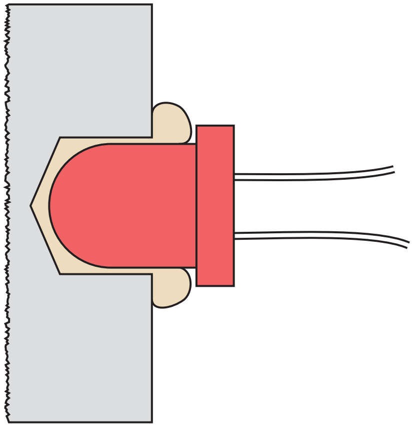

Figure 3-82. *This cross-sectional view shows a sheet of transparent acrylic in which a hole has been drilled part of the way from the back toward the front. Because a drill bit creates a hole with a conical shape at the bottom, and because the LED has rounded contours, transparent epoxy or silicone caulking can be injected into the hole before mounting the LED.*

Try illuminating the LED, and sand the acrylic some more if necessary. Finally, you can decide whether to mount the circuit on the back of the acrylic, or whether you want to run a wire to it elsewhere.

Because the LED will flash at about the speed of a human heart while the person is resting, it may look as if it's measuring your pulse, especially if you mount it on the center of your chest or in a strap around your wrist. If you enjoy hoaxing people, you can suggest that you're in such amazingly good shape, your pulse rate remains constant even when you're taking strenuous exercises.

To make a good-looking enclosure for the circuit, I can think of options ranging from embedding the whole thing in clear epoxy to finding a Victorian-style locket. I'll leave you to consider alternatives, because this is a book about electronics rather than handicrafts.

However, I will address one final issue: how long will this gadget continue flashing?

If you check the following section "Essentials: Battery life," you'll find that a regular alkaline 9-volt battery should keep the LED flashing for about 50 hours.

ESSENTIALS

Battery life

Any time you finish a circuit that you intend to run from a battery, you'll want to calculate the likely battery life. This is easily done, because manufacturers rate their batteries according to the "ampere hours" they can deliver. Keep the following in mind:

- The abbreviation for amp-hours is Ah, sometimes printed as AH. Milliampere-hours are abbreviated mAh.
- The rating of a battery in amp-hours is equal to the current, in amps, multiplied by the number of hours that the battery can deliver it.

Thus, in theory 1 amp-hour can mean 1 amp for 1 hour, or 0.1 amp for 10 hours, or 0.01 amp for 100 hours—and so on. In reality, it's not as simple as this, because the chemicals inside a battery become depleted more quickly when you draw a heavy current, especially if the battery gets hot. You have to stay within limits that are appropriate to the size of the battery.

For instance, if a small battery is rated for 0.5 amp-hours, you can't expect to draw 30 amperes from it for 1 minute. But you should be able to get 0.005 amps (i.e., 5 milliamps) for 100 hours without any trouble. Remember, though, that the voltage delivered by a battery will be greater than its rated voltage when the battery is fresh, and will diminish below its rated voltage while the battery is delivering power.

According to some test data that I trust (I think they are a little more realistic than the estimates supplied by battery manufacturers), here are some numbers for typical batteries:

- Typical 9 volt alkaline battery: 0.3 amp-hours, while delivering 100 mA.
- Typical AA size, 1.5-volt alkaline battery: 2.2 amp-hours, while delivering 100 mA.
- Rechargeable nickel-metal hydride battery: about twice the endurance of a comparably sized alkaline battery.
- Lithium battery: maybe three times the endurance of an alkaline battery.

BACKGROUND

Maddened by measurement

Throughout most of this book, I've mostly used measurements in inches, although sometimes I've digressed into the metric system, as when referring to "5-mm LEDs." This isn't inconsistency on my part; it reflects the conflicted state of the electronics industry, where you'll find inches and millimeters both in daily use, often in the very same data sheet.

The United States is the only major nation still using the old system of units that originated in England. (The other two holdouts are Liberia and Myanmar, according to the CIA's *World Factbook*.) Still, the United States has led many advances in electronics, especially the development of silicon chips, which have contacts spaced 1/10 inch apart. These standards became firmly established, and show no sign of disappearing.

To complicate matters further, even in the United States, you can encounter two incompatible systems for expressing fractions of an inch. Drill bits, for instance, are measured in multiples of 1/64 inch, while metal thicknesses may be measured in decimals such as 0.06 inch (which is approximately 1/16 inch).

The metric system is not necessarily more rational than the U.S. system. Originally, when the metric system was formally introduced in 1875, the meter was defined as being 1/10,000,000 of the distance between the North Pole and the equator, along a line passing through Paris—a quixotic, Francocentric conceit. Since then, the meter has been redefined three times, in a series of efforts to achieve greater accuracy in scientific applications.

As for the usefulness of a 10-based system, moving a decimal point is certainly simpler than doing calculations in 64ths of an inch, but the only reason we count in tens is because we happen to have evolved with that number of digits on our hands. A 12-based system would really be more convenient, as numbers would be evenly divisible by 2 and 3.

As we're stuck with the whimsical aspects of length measurement, I've created the charts in Figures 3-83 and 3-84 to assist you in going from one system to another. From these you will see that when you need to drill a hole for a 5 mm LED, a 3/16-inch drill bit is about right. (In fact, it results in a better, tighter fit than if you drill an actual 5 mm hole.)

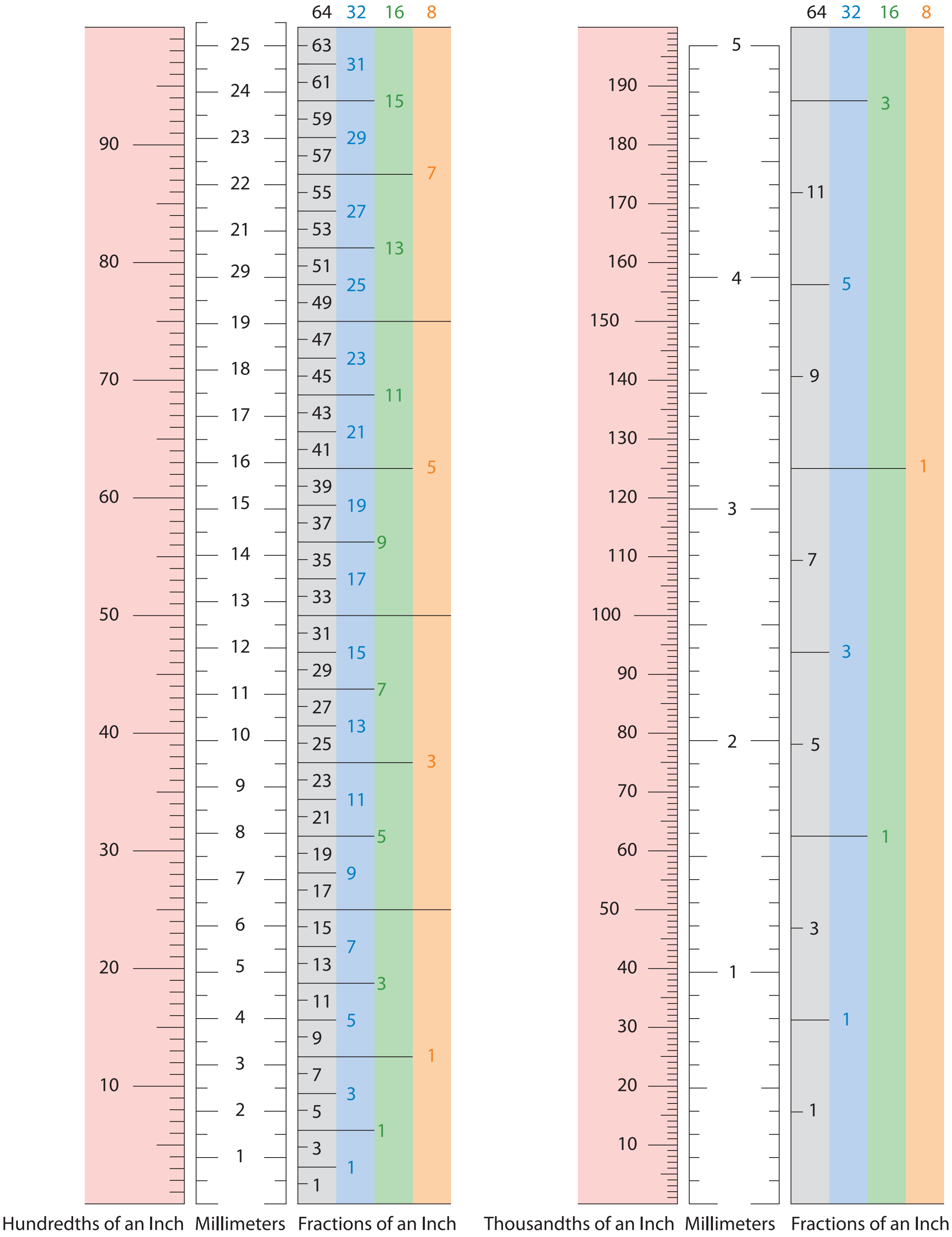

Figure 3-83. *Because units of measurement are not standardized in electronics, conversion is often necessary. The chart on the right is a 5x magnification of the bottom section of the chart on the left.*

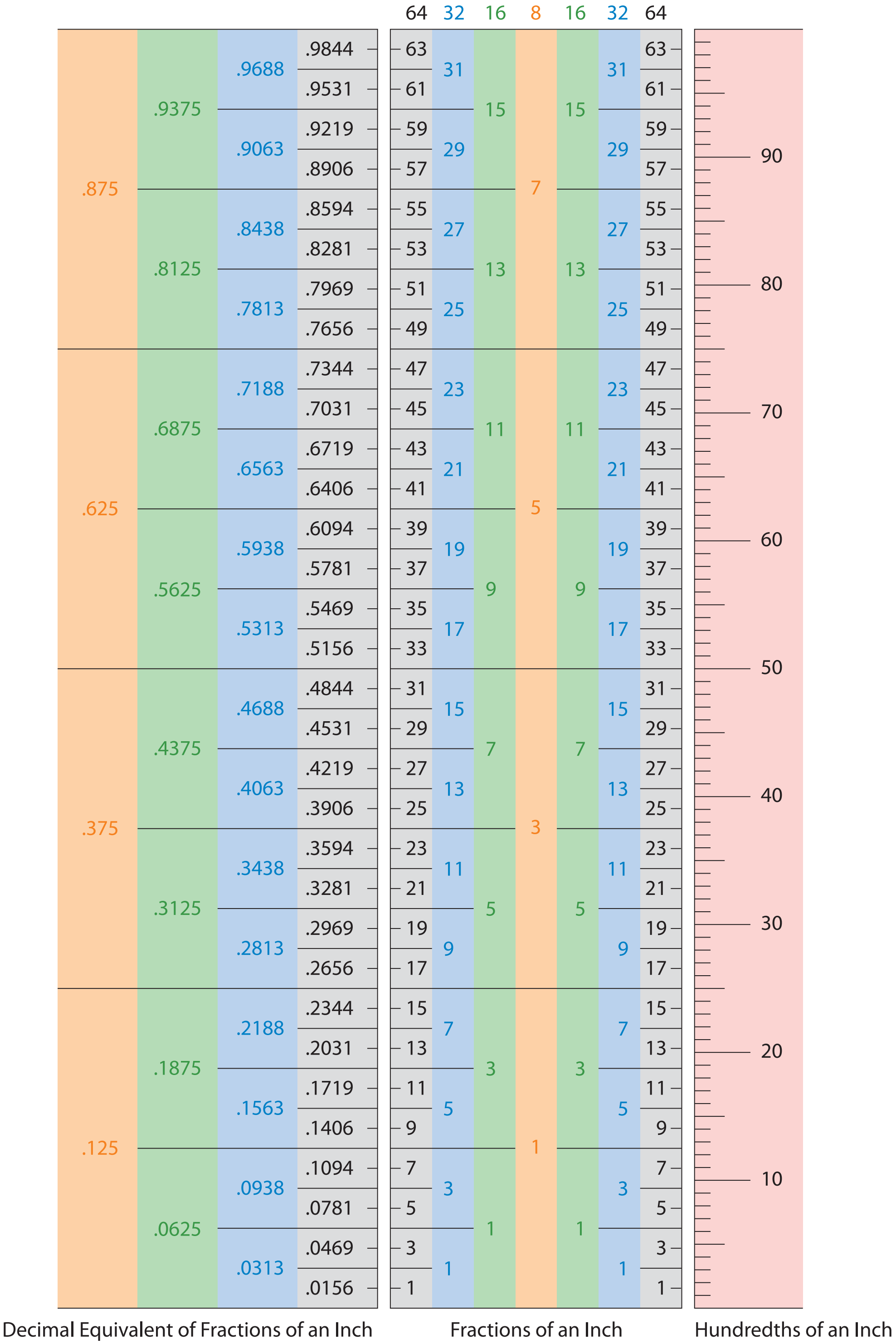

Figure 3-84. *This chart allows conversion between hundredths of an inch, conventional U. S. fractions of an inch, and fractions expressed in thousandths of an inch.*

Experiment 15: Intrusion Alarm Revisited

Time now to add some of the enhancements to the intrusion alarm that I discussed at the end of Experiment 11. I'm going to show you how the alarm can be triggered if you install various detectors on windows and doors in your home. I'll also show how the alarm can be wired so that it locks itself on and continues to make noise even after a door or window is reclosed.

This experiment will demonstrate the procedure for transferring a project from a breadboard to a piece of perforated board that has copper connections laid out identically to the ones inside the breadboard, as shown earlier in Figure 3-72. And you'll mount the finished circuit in a project box with switches and connectors on the front.

When all is said and done, you'll be ready for wholesale circuit building. The explanations in the rest of this book will get gradually briefer, and the pace will increase.

You will need:

- 15-watt pencil-type soldering iron
- Thin solder (0.022 inches or similar)
- Wire strippers and cutters
- Perforated board etched with copper in a breadboard layout
- Small vise or clamp to hold your perforated board
- The same components that you used in Experiment 11, plus:
  - 2N2222 NPN transistor. Quantity: 1.
  - DPDT relay. Quantity: 1.
  - SPDT toggle switch. Quantity: 1.
  - 1N4001 diode. Quantity: 1.
  - Red and green 5mm LEDs. Quantity: 1 each.
  - Project box, 6 × 3 × 2 inches.
  - Power jack, type N, and matching power socket, type N.
  - Binding posts.
  - Stranded 22-gauge wire, three different colors.
  - Magnetic sensor switches, sufficient for your home.
  - Alarm network wiring, sufficient for your home.

Magnetic Sensor Switches

A typical alarm sensor switch consists of two modules: the magnetic module and the switch module, as shown in Figures 3-85 and 3-86. The magnetic module contains a permanent magnet, and nothing else. The switch module contains a "reed switch," which makes or breaks a connection (like a contact inside a relay) under the influence of the magnet. When you bring the magnetic module close to the switch module, you may faintly hear the reed switch click as it flips from one state to the other.

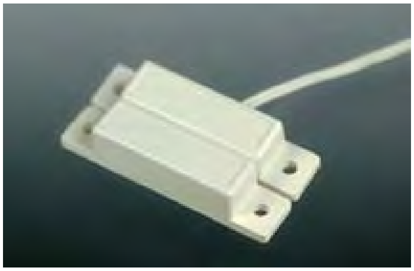

Figure 3-85. *In this simple alarm sensor switch, the lower module contains a magnet, which opens and closes a reed switch sealed into the upper module.*

Like all switches, reed switches can be normally open or normally closed. For this project, you want the kind of switch that is normally open, and closes when the magnetic module is close to it.

Attach the magnetic module to the moving part of a door or window, and attach the switch module to the window frame or door frame. When the window or door is closed, the magnetic module is almost touching the switch module. The magnet keeps the switch closed until the door or window is opened, at which point the switch opens.

The only question is: how do we use this component to trigger our alarm? As long as a small current flows through all our magnetic sensor switches, the alarm should be off, but if the flow of current stops, the alarm should switch on.

We could use a relay that is "always on" while the alarm is armed. When the circuit is interrupted, the relay relaxes and its other pair of contacts closes, which could power up the alarm noisemaker.

But I don't like this idea. Relays take significant power, and they can get hot. Most of them are not designed to be kept "always-on." I'd prefer to handle the task using a transistor.

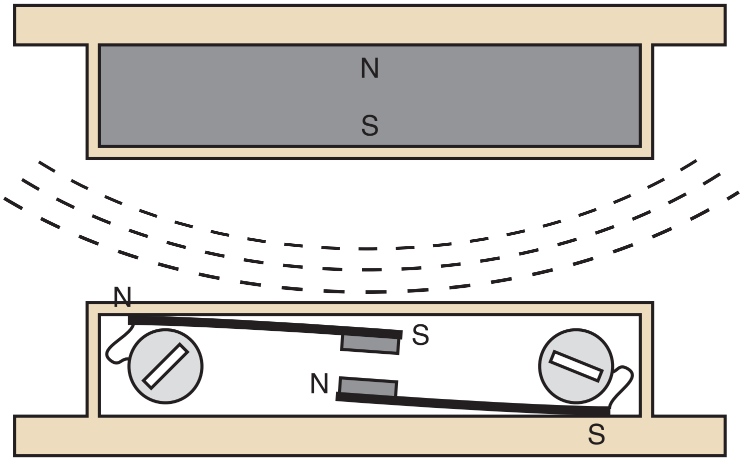

Figure 3-86. *This cutaway diagram shows a reed switch (bottom) and the magnet that activates it (top), inside an alarm sensor. The switch contains two flexible magnetized strips, the upper one with its south pole adjacent to an electrical contact, the lower one with its north pole adjacent to an electrical contact. When the south pole of the magnet approaches the switch, the magnetic force (shown as dashed lines) repels the south contact and attracts the north contact, causing them to snap together. Two screws on the outside of the casing are connected with the strips inside.*

A Break-to-Make Transistor Circuit

First, recall how an NPN transistor works. When the base is not sufficiently positive, the transistor blocks current between its collector and emitter, but when the base is relatively positive, the transistor passes current.

Take a look at the schematic in Figure 3-87, which is built around our old friend the 2N2222 NPN transistor. When the switch is closed, it connects the base of the transistor to the negative side of the power supply through a 1K resistor. At the same time, the base is connected with the positive side of the power supply through a 10K resistor. Because of the difference in resistances and the relatively high turn-on voltage for the LED, the base is forced below its turn-on threshold, and as a result, the transistor will not pass much current. The LED will glow dimly at best.

Now what happens when the switch is opened? The base of the transistor loses its negative power supply and has only its positive power supply. It becomes much more positive, above the turn-on threshold for the transistor, which tells the transistor to lower its resistance and pass more current. The LED now glows brightly. Thus, when the switch is turned off and breaks the connection, the LED is turned on.

This seems to be what we want. Imagine a whole series of switches instead of just one switch, as shown in Figure 3-88. The circuit will still work the same way, even if the switches are scattered all over your home, because the resistance in the wires connecting the switches will be trivial compared with the resistance of the 1K resistor.

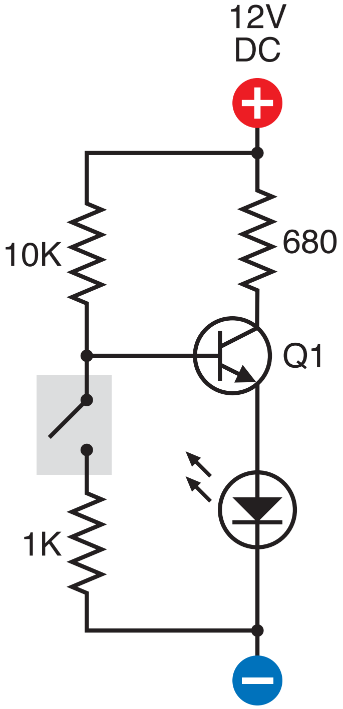

Figure 3-87. *In this demonstration circuit, when the switch is opened, it interrupts negative voltage to the base of the transistor, causing the transistor to lower its resistance, allowing current to reach the LED. Thus, when the switch is turned off, it turns on the LED.*

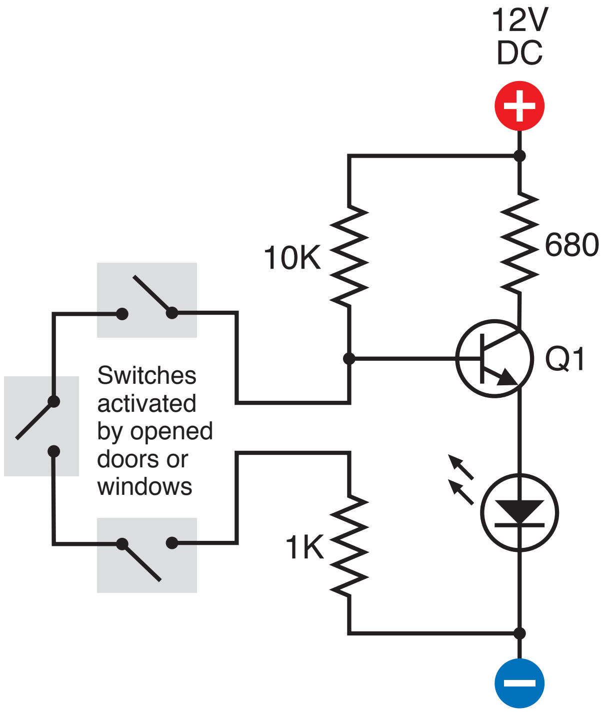

Figure 3-88. *A network of switches, wired in series, can be substituted for the single switch in Figure 3-87. Now any one switch will break continuity and trigger the transistor.*

I have shown the switches open, because that's the way the schematic for a switch is drawn, but imagine them all closed. The base of the transistor will now be supplied through the long piece of wire connecting all the closed switches, and the LED will stay dark. Now if just one switch is opened, or if anyone tampers with the wire linking them, the base of the transistor loses its connection to negative power, at which point the transistor conducts power and the LED lights up.

While all the switches remain closed, the circuit is drawing very little current—probably about 1.1 mA. So you could run it from a typical 12-volt alarm battery.

Now suppose we swap out the LED and put a relay in there instead, as shown in Figure 3-89. I don't mind using a relay in this location, because the relay will not be "always on." It will normally be off, and will draw power only when the alarm is triggered.

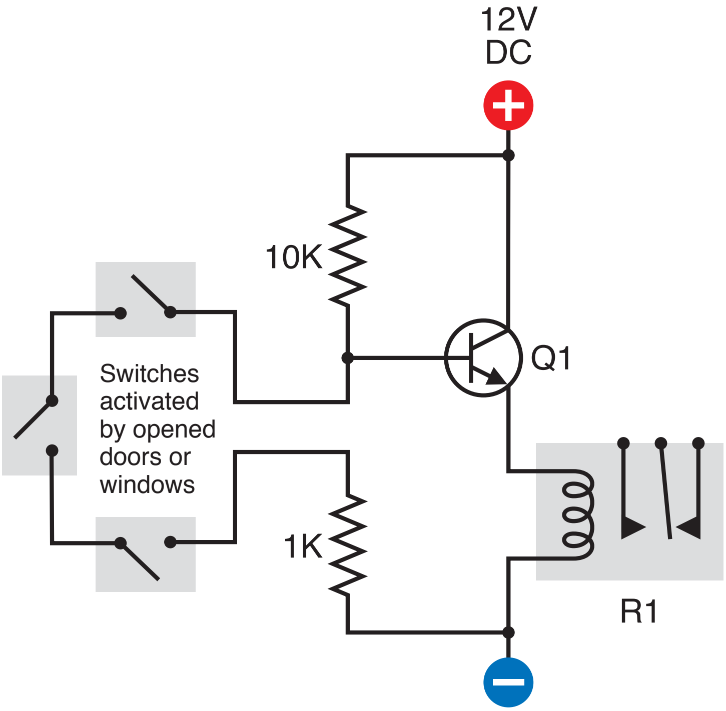

Figure 3-89. *If the LED and 680Ω resistor shown in are removed, and a relay takes their place, the relay will be activated when any switch in the sensor network is opened.*

Try one of the 12-volt relays that you used previously. You should find that when you open the switch, the relay is energized. When you close the switch, the relay goes back to sleep. Note that I eliminated the 680Ω resistor from the circuit, because the relay doesn't need any protection from the 12-volt power supply.

Self-Locking Relay

There's only one remaining problem: we want the alarm to continue making noise even after someone who has opened a door or window closes it again quickly. In other words, when the relay is activated, it must lock itself on.

One way to do this would be by using a latching relay. The only problem is that we would then need another piece of circuitry to unlatch it. I prefer to show you how you can make any relay keep itself switched on after it has received just one jolt of power. This idea will be useful to you later in the book as well.

The secret is to supply power to the relay coil through the two contacts inside the relay that are normally open. (Note that this is exactly opposite to the relay oscillator, which supplied power to its coil through the contacts that were normally closed. That setup caused the relay to switch itself off almost as soon as it switched itself on. This setup causes the relay to keep itself switched on, as soon as it has been activated.)

In Figure 3-90, the four schematics illustrate this. You can imagine them as being like frames in a movie, photographed microseconds apart. In the first picture, the switch is open, the relay is not energized, and nothing is happening. In the second, the switch has been closed to energize the coil. In the third, the coil has pulled the contact inside the relay, so that power now reaches the coil via two paths. In the fourth, the switch has been opened, but the relay is still powering its own coil through its contacts. It will remain locked in this state until the power is disconnected.

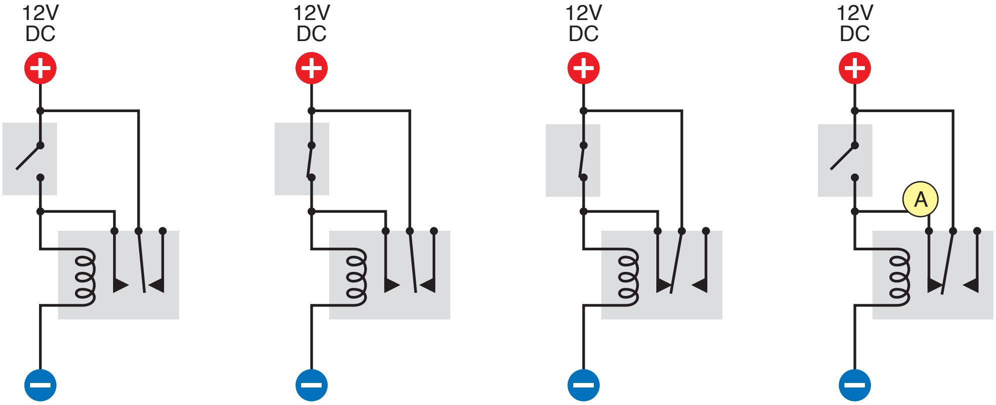

Figure 3-90. *This sequence of schematics shows the events that occur when a relay is energized. Initially, the switch is open. Then the switch is closed, activating the relay. The relay then powers itself through its own internal contacts. The relay remains energized even after the switch is opened again. Power switched by the relay can be taken from the circuit at point A.*

All we need to do, to make use of this idea, is to substitute the transistor for the on/off switch, and tap into the circuit at point A, running a wire from there to the noisemaking module.

Figure 3-91 shows how that would work. When the transistor is activated by any of the network of sensor switches, as previously explained, the transistor conducts power to the relay. The relay locks itself on, and the transistor becomes irrelevant.

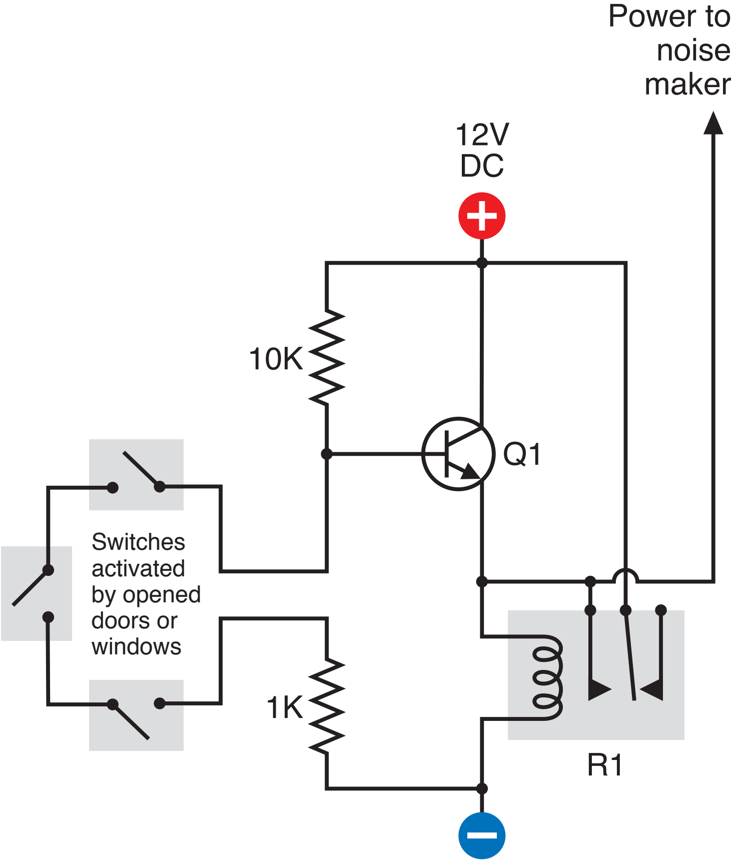
Power to
noise
maker
12V
DC
10K
Q1
Switches
activated
by opened
doors or
windows
1K
R1

Figure 3-91. *The self-locking relay depicted in has been incorporated in the alarm circuit, so that if any switch in the network is opened, the relay will continue to power the noise maker even if the switch is closed again.*

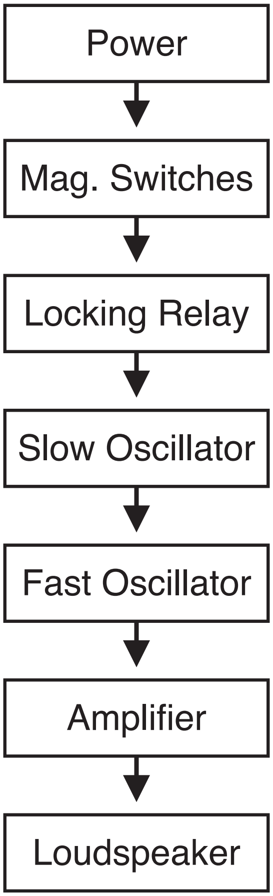
Power
Mag. Switches
Locking Relay
Slow Oscillator
Fast Oscillator
Amplifier
Loudspeaker

Figure 3-92. *This block diagram previously shown in Figure 2-112 on page 90 has been updated to include the magnetic-switch network and locking-relay control system.*

Because I've been adding pieces to the original alarm noisemaker circuit, I've updated the block diagram from Figure 2-112 to show that we can still break this down into modules with simple functions. The revised diagram is shown in Figure 3-92.

Blocking Bad Voltage

One little problem remains: in the new version of the circuit, if the transistor goes off while the relay is still on, current from the relay can flow back up the wire to the emitter of the transistor, where it will try to flow backward through the transistor to the base, which is "more negative," as it is linked through all the magnetic switches and the 1K resistor to the negative side of the power supply.

Applying power backward through a transistor is not a nice thing to do. Therefore the final schematic in this series shows one more component, which you have not seen before: a diode, labeled D1. See Figure 3-93. The diode looks like the heart of an LED, and indeed, that's pretty much what it is, although some diodes are much more robust. It allows electricity to flow in only one direction, from positive to negative, as shown by its arrow symbol. If current tries to flow in the opposite direction, the diode blocks it. The only price you pay for this service is that the diode imposes a small voltage drop on electricity flowing in the "OK" direction.

So now, positive flow can pass from the transistor, through the diode, to the relay coil, to get things started. The relay then supplies itself with power, but the diode prevents the positive voltage from getting back into the transistor the wrong way.

Perhaps a more elegant solution to the problem is to connect the NO leg of the relay via a 10k resistor to the base connection. When the relay is not energized, the NO leg is inert and simply behaves as a parasitic capacitance on the node. When the relay becomes energized, the NO leg shunts +12V through the common terminal via a 10k resistor into the base of the transistor. In this circuit configuration, the transistor is never exposed to a potentially harmful voltage and you are not depending on leakage currents of non-ideal elements to protect devices.

However, I needed an opportunity to introduce you to the concept of diodes. You can check the following section "Essentials: All about diodes" to learn more.

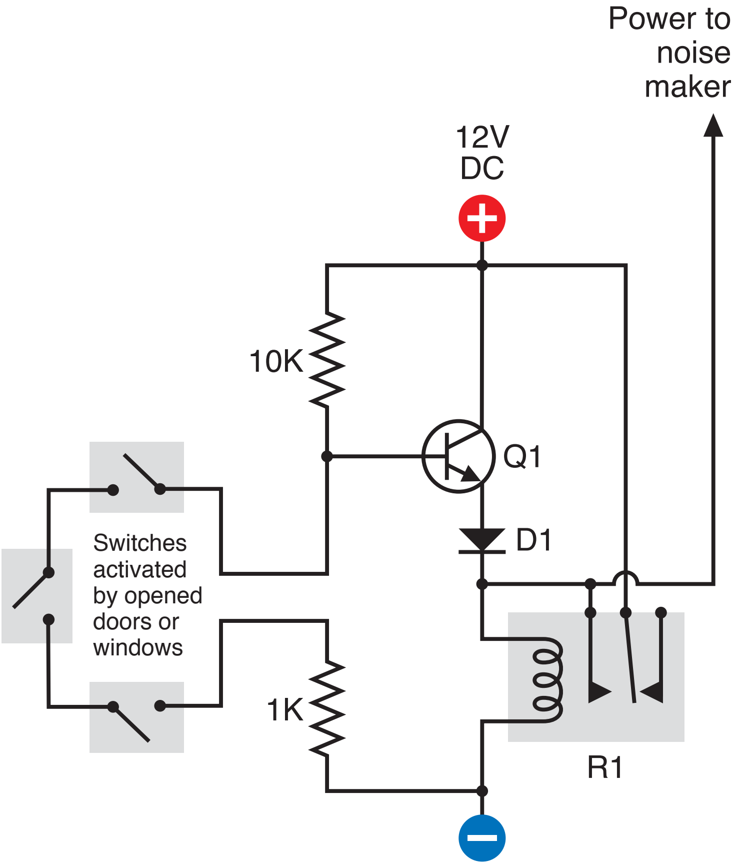

Figure 3-93. *Diode D1 has been added to protect the emitter of Q1 from positive voltage when the relay is energized.*

ESSENTIALS

All about diodes

A diode is a very early type of semiconductor. It allows electricity to flow in one direction, but blocks it in the opposite direction. (A light-emitting diode is a much more recent invention.) Like an LED, a diode can be damaged by reversing the voltage and applying excessive power, but most diodes generally have a much greater tolerance for this than LEDs. The end of the diode that blocks positive voltage is always marked, usually with a circular band, while the other end remains unmarked. Diodes are especially useful in logic circuits, and can also convert alternating current (AC) into direct current (DC).

A Zener diode is a special type that we won't be using in this book. It blocks current completely in one direction, and also blocks it in the other direction until a threshold voltage is reached—much like a PUT.

Signal diodes are available for various different voltages and wattages. The 1N4001 diode that I recommend for the alarm activation circuit is capable of handling a much greater load at a much higher voltage, but I used it because it has a low internal resistance. I wanted the diode to impose a minimal voltage drop, so that the relay would receive as much voltage as possible.

It's good practice to use diodes at less than their rated capacity. Like any semiconductor, they can overheat and burn out if they are subjected to mistreatment.

The schematic symbol for a diode has only one significant variant: sometimes the triangle is outlined instead of filled solid black (see Figure 3-94).

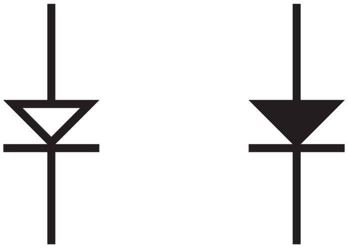

Figure 3-94. *Either of these schematic symbols may be used to represent a diode, but the one on the right is more common than the one on the left.*

Completing the Breadboard Alarm Circuit

It's time now to breadboard the control circuit for your alarm noisemaker. Figure 3-95 shows how this can be done. I am assuming that you still have the noisemaker, which functions as before. I'm assuming that you still have its relevant components mounted on the top half of the breadboard. To save space, I'm just going to show the additional components mounted on the bottom half of the same breadboard.

It's important to remember that you are not supplying power directly to the left and right "rails" on the breadboard anymore; you are supplying power to the relay-transistor section, and when the relay closes its contacts, the *relay* supplies power to the rails. These then feed the power up to the top half of the breadboard. So disconnect your power supply from the breadboard rails and reconnect it as shown in Figure 3-95.

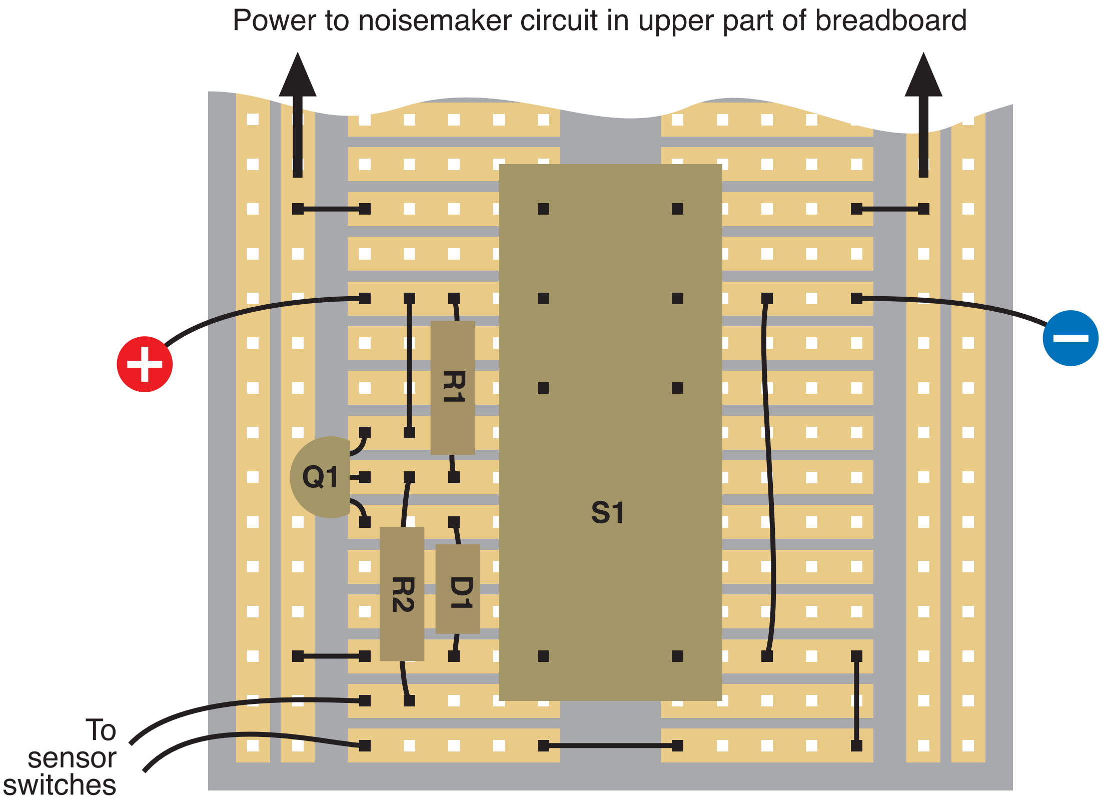

Figure 3-95. *The schematic that was developed in the previous pages can be emulated with components on a breadboard, as shown here. S1 is a DPDT relay. Wires to the sensor switch network and to the power supply must be added where shown.*

Because it's a double-pole relay, I am using it to switch negative as well as positive. This means that when the relay contacts are open, the noisemaking section of the circuit is completely isolated from the rest of the world.

The breadboarded relay circuit is exactly the same as the schematic in Figure 3-93. The components have just been rearranged and squeezed together so that they will fit alongside the relay. Two wires at the lower-left corner go to the network of magnetic sensor switches that will trip the alarm; for testing purposes, you can just hold the stripped ends of these two wires together to simulate all the switches being closed, and separate the wires to simulate a switch opening.

Two more wires bring power to the breadboard on either side of the relay. This is where you should connect your power supply during testing. The output from the relay, through its top pair of contacts, is connected with the rails of the breadboard by a little jumper wire at top left, and another at top right. Don't forget to include them! One more little wire at the lower-left corner (easily overlooked) connects the lefthand side rail to the lefthand coil terminal of the relay, so that when the relay is powering the noisemaker circuit, it powers itself as well.

When you mount the diode, remember that the end of it that is marked with a band around it is the end that blocks positive current. In this circuit, that's the lower end of the diode.

Try it to make sure that it works. Short the sensor wires together and then apply power. The alarm should remain silent. You can use your meter to check that no voltage exists between the side rails. Now separate the sensor wires, and the relay should click, supplying power to the side rails, which activates the noisemaker. Even if you bring the sensor wires back together, the relay should remain locked on. The only way to unlock it is to disconnect the power supply.

When the circuit is active, the transistor followed by the diode drops the voltage slightly, but the 12-volt relay should still work.

In my test circuit, trying three different relays, they drew between 27 and 40 milliamps at 9.6 volts. Some current still leaked through the transistor when it was in its "off" mode, but only a couple of milliamps at 0.5 volts. This low voltage was far below the threshold required to trip the relay.

Ready for Perfboarding

If the circuit works, the next step is to immortalize it on perforated board. Use the type of board that has a breadboard contact pattern etched on it in copper, as shown in Figure 3-72 on page 116. Check the following section, "Essentials: Perfboard procedure," for guidance on the best way to make this particular kind of solder joint—and the subsequent section for the most common problems.

ESSENTIALS

Perfboard soldering procedure

Carefully note the position of a component on your breadboard, and then move it to the same relative position on the perfboard, poking its wires through the little holes.

Turn the perfboard upside down, make sure that it's stable, and examine the hole where the wire is poking through, as shown in Figure 3-96. A copper trace surrounds this hole and links it with others. Your task is to melt solder so that it sticks to the copper and also to the wire, forming a solid, reliable connection between the two of them.

Take your pencil-style soldering iron in one hand and some solder in your other hand. Hold the tip of the iron against the wire and the copper, and feed some thin solder to their intersection. After two to four seconds, the solder should start flowing.

Allow enough solder to form a rounded bump sealing the wire and the copper, as shown in Figure 3-97. Wait for the solder to harden thoroughly, and then grab the wire with pointed-nosed pliers and wiggle it to make sure you have a strong connection. If all is well, snip the protruding wire with your cutters. See Figure 3-98.

Because solder joints are difficult to photograph, I'm using drawings to show the wire before and after making a reasonably good joint, which is shown in pure white, outlined with a black line.

Actual soldered perfboard is shown in the photographs in Figures 3-99 and 3-100.

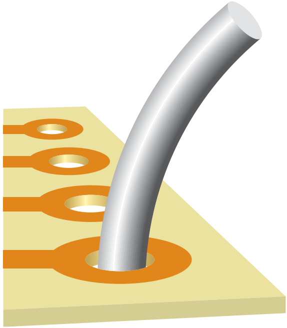

Figure 3-96

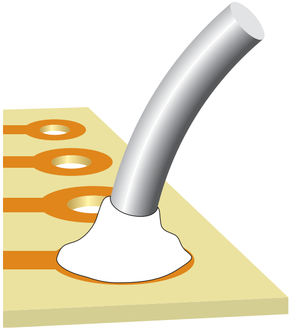

Figure 3-97

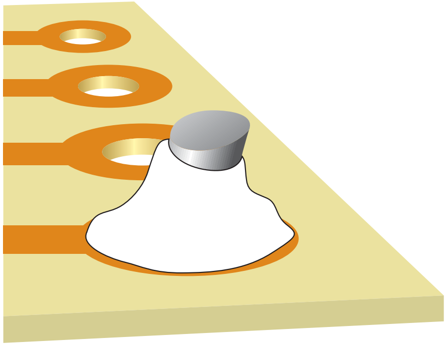

Figure 3-98. *To establish a connection between a section of wire and a copper trace on perforated board, the wire is pushed through the hole, and solder (shown in pure white for illustrative purposes) completes the connection. The wire can then be snipped short.*

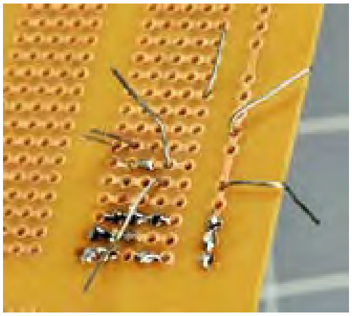

Figure 3-99. *This photograph was taken during the process of transferring components from breadboard to perforated board. Two or three components at a time are inserted from the other side of the board, and their leads are bent over to prevent them from falling out.*

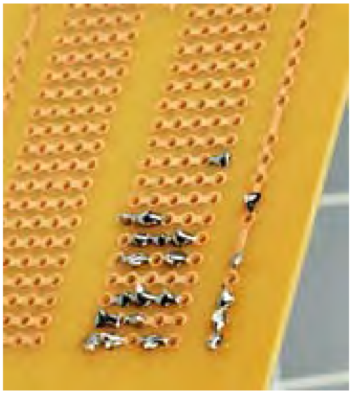

Figure 3-100. *After soldering, the leads are snipped short and the joints are inspected under a magnifying glass. Another two or three components can now be inserted, and the process can be repeated.*

TOOLS

Four most common perfboarding errors

1. Too much solder

Before you know it, solder creeps across the board, touches the next copper trace, and sticks to it, as depicted in Figure 3-101. When this happens, you have to wait for it to cool, and then cut it with a utility knife. You can also try to remove it with a rubber bulb and solder wick, but some of it will tend to remain.

Even a microscopic trace of solder is enough to create a short circuit. Check the wiring under a magnifying glass while turning the perfboard so that the light strikes it from different angles (or use your solder wick to suck it away).

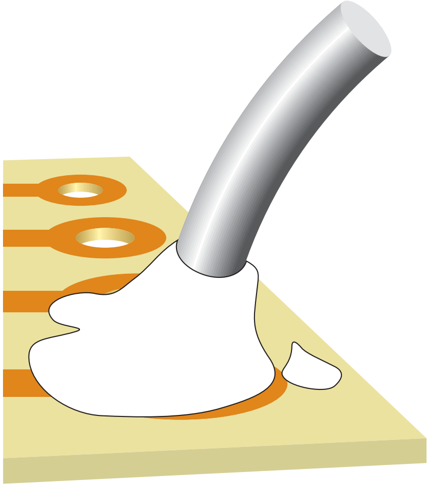

Figure 3-101. *If too much solder is used, it makes a mess and can create an unwanted connection with another conductor.*

2. Not enough solder

If the joint is thin, the wire can break free from the solder as it cools. Even a microscopic crack is sufficient to stop the circuit from working. In extreme cases, the solder sticks to the wire, and sticks to the copper trace around the wire, yet doesn't make a solid bridge connecting the two, leaving the wire encircled by solder yet untouched by it, as shown in Figure 3-102. You may find this undetectable unless you observe it with magnification.

You can add more solder to any joint that may have insufficient solder, but be sure to reheat the joint thoroughly.

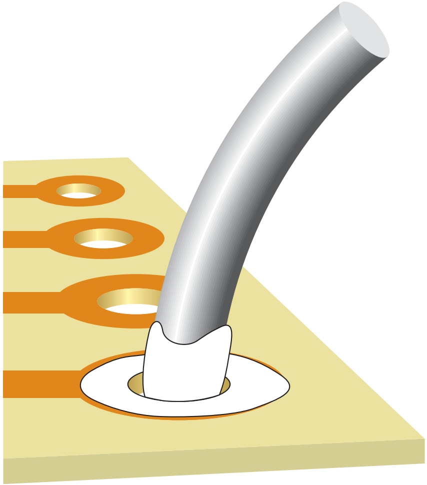

Figure 3-102. *Too little solder (or insufficient heat) can allow a soldered wire to remain separate from the soldered copper on the perforated board. Even a hair-thin gap is sufficient to prevent an electrical connection.*

3. Components incorrectly placed

It's very easy to put a component one hole away from the position where it should be. It's also easy to forget to make a connection.

I suggest that you print a copy of the schematic, and each time you make a connection on the perforated board, you eliminate that wire on your hardcopy, using a highlighter.

4. Debris

When you're trimming wires, the little fragments that you cut don't disappear. They start to clutter your work area, and one of them can easily get trapped under your perforated board, creating an electrical connection where you don't want it.

This is another reason for working with something soft, such as polyurethane foam, under your project. It tends to absorb or hold little pieces of debris, reducing the risk of you picking them up in your wiring.

Clean the underside of your board with an old (dry) toothbrush before you apply power to it, and keep your work area as neat as possible. The more meticulous you are, the fewer problems you'll have later.

Once again, be sure to check every joint with a magnifying glass.

Transferring components from the breadboard to the perforated board should be fairly simple, as long as you don't try to move too many at once. Follow the suggestions described in previous section "Essentials: Perfboard procedure," and pause frequently to check your connections. Impatience is almost always the cause of errors in this kind of work.

Figure 3-103 shows the noisemaker section of the circuit on perfboard, with the components positioned to minimize wasted space. Figure 3-104 shows the perfboard with the relay and its associated components added. The two black wires will go to the loudspeaker, the black-and-red pair of wires will bring power to the board, and the green wires will go to the magnetic sensor switches. Each wire penetrates the board, and its stripped end is soldered to the copper beneath.

Test it now, in the same way that you tested the same circuit on the breadboard. If it doesn't work, check the following section, "Essentials: Real-world fault tracing." If it does work, you're ready to trim the board and mount it in a project box.

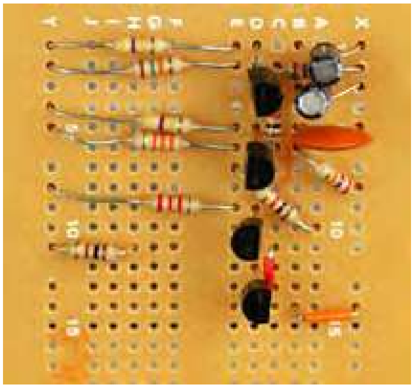

Figure 3-103. *The noisemaker circuit has been transplanted from breadboard to perforated board, with no additions or changes.*

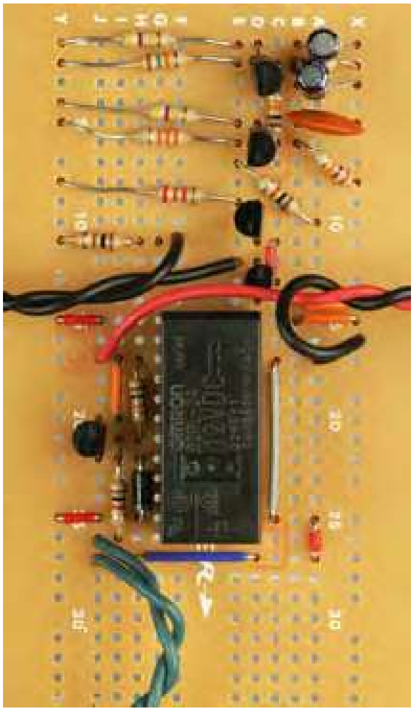

Figure 3-104. *The relay-transistor control circuit has been added. Wires to external devices have been stripped and poked into the perforated board, where they are soldered in place. The green wires connect with the sensor network, the black wires go to the loudspeaker, and the red-and-black wires supply power.*

ESSENTIALS

Real-world fault tracing

Here's a real-life description of the procedure for tracing a fault.

After I assembled the perfboard version of the combined noisemaker and relay circuit, I checked my work, applied power—and although the relay clicked, no sound came out of the loudspeaker. Of course, everything had worked fine on the breadboard.

First I looked at component placement, because this is the easiest thing to verify. I found no errors. Then I flexed the board gently while applying power—and the loudspeaker made a brief "beep." Any time this happens, you can be virtually certain that a solder joint has a tiny crack in it.

The next step was to anchor the black lead of my meter to the negative side of the power supply, and then switch on the power and go through the circuit point by point, from top to bottom, checking the voltage at each point with the red lead of the meter. In a simple circuit like this, every part should show at least some voltage.

But when I got to the second 2N2222 transistor, which powers the loudspeaker, its output was completely dead. Either I had melted the transistor while soldering it (unlikely), or there was a bad joint. I checked the perfboard beneath the transistor with the magnifying glass, and found that solder had flowed around one of the leads of the transistor without actually sticking to it. The gap must have been less than one-thousandth of an inch, but still, that was enough. Probably, the problem had been caused by dirt or grease.

This is the kind of patient inquiry you need to follow when a circuit doesn't work. Check whether your components are placed correctly, check your power supply, check the power on the board, check the voltage at each stage, and if you are persistent, you'll find the fault.

Switches and Inputs for the Alarm

Now you need to make the system easy to use. The block diagram in Figure 3-105 shows one additional box near the top of the sequence: User Controls. These will consist of switches, LEDs, and connections to the outside world. To plan this part of the job, first I have to summarize the way in which our alarm system works at this point in its evolution.

A full-featured home alarm system normally has two modes: in-home and away-from-home:

- Using the in-home mode, you switch on the alarm while you are at home so that it will alert you if an intruder opens a door or window.
- Using the away-from-home mode, typically you enter a code number, after which you have 30 seconds to leave and close the door behind you. When you return, you trigger the alarm by opening the door, but now you have 30 seconds to go to the control panel and enter your code number again to stop the alarm from sounding.

So far, the alarm system that you've been building has only an in-home mode. Still, many people find this function useful and reassuring. Later in the book I'll suggest a way in which you can modify it to incorporate an away-from-home mode, but for now, making it practical for in-home use is enough of a challenge.

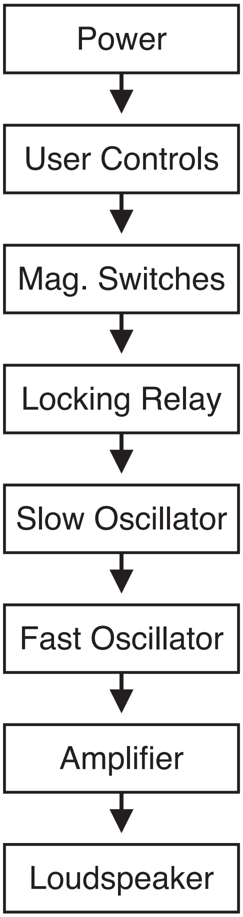

Figure 3-105. *The final block diagram for this phase of the project shows where user controls fit in the series of functions.*

Consider how it should be used on an everyday basis. It should have an on/off switch, naturally. When it's on, any of the magnetic sensor switches should trigger the alarm. But what if you switch it on without realizing that you've left a window open? At that time it won't be appropriate for the alarm to sound. What you really need is a circuit-test feature, to tell you if all the doors and windows are closed. Then you can switch on the alarm.

I think a pushbutton would be useful to test the alarm circuit. When you press it, a green LED should light up to indicate that the circuit is good. After you see the green light, you let go of the pushbutton and turn on the power switch, which illuminates a red LED, to remind you that the alarm is now armed and ready.

One additional feature would be useful: an alarm noisemaker test feature, so that you can be sure that the system is capable of sounding its alert when required to do so.

The circuit shown in Figure 3-106 incorporates all of these features. S1 is a SPDT switch; S2 is a DPDT momentary pushbutton of ON-(ON) type. The schematic shows it in its "relaxed" mode, when the button is not being pressed.

D1 is a red LED, D2 is a green LED, J1 is a power input jack (to be connected with an external 12-volt supply), and R1 is a 680Ω resistor to protect the LEDs. Note that J1 follows the usual practice of supplying positive voltage in its center contact, and negative in the circular shell around the center.

When S1 is in its Off position, it still supplies positive power through its upper contact to S2, the pushbutton. When the pushbutton is pressed, so that it goes into its "Test" position, the pole of S2 connects with the power and sends it out through the sensor switches on doors and windows. The wires to these switches will be attached via a couple of binding posts, shown here as two circles. If the sensor switches are all closed, power returns through the second binding post, passes through the lower set of contacts in S2, and lights D2, the green LED. Because S1 is not supplying power to the alarm circuit board, the alarm does not sound at this time.

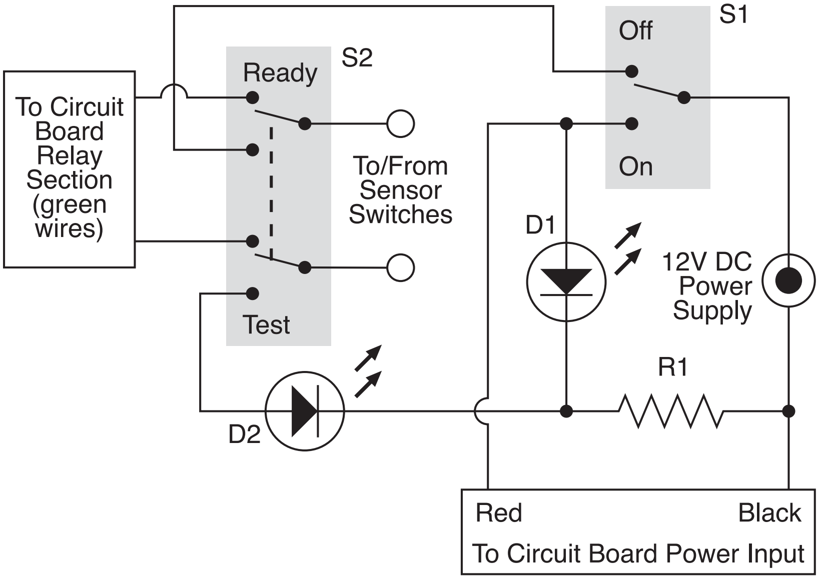

Figure 3-108. *This schematic suggests a convenient way to add an on/off switch, a continuity testing feature, and a noisemaker testing feature to the alarm.*

Now if S1 is turned to its On position, it sends power to the components on the perforated board. The transistor circuit sends power out along the green wires to S2, and as long as the button is not being pressed, the power goes out through the switch network and back through S2, to the transistor, keeping its base relatively negative. So the alarm remains silent. But as soon as a sensor switch is opened, the circuit is broken and the alarm sounds. The only way to stop it will be by switching S1 into its Off position.

Finally, if you press S2, the pushbutton, while S1 is in its On mode, you interrupt the network of switches and activate the noisemaker. In this way, S2 does dual duty: when S1 is off, pushbutton S2 tests the sensor switches for continuity. When S1 is on, S2 tests the noisemaker to make sure it creates a noise. I think this is the simplest possible way to implement these features.

Installing the Switches

If you bought a project box from RadioShack, it may have come with two optional top panels: one made of metal, the other made of plastic. I'll assume that you'll use the plastic one, as you'll have more trouble drilling holes in metal. The type of plastic used by RadioShack is ABS, which is very easy to shape with the tools that I have recommended.

You have to choose a layout for the switches and other components that will share the top panel of the project box. I like a layout to be neat, so I take the trouble to draw it using illustration software, but a full-size pencil sketch is almost as good. Just make sure there's room for the components to fit together, and try to place them similarly to the schematic, to minimize the risk of confusion.

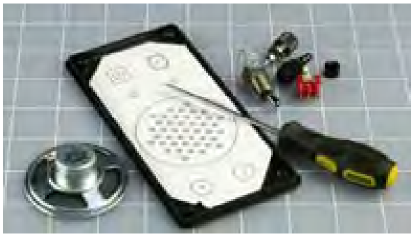

Figure 3-107. *A printed layout for the switches, LEDs, and other components has been taped to the underside of the lid of the project box. An awl is pressed through the paper to mark the center of each hole to be drilled in the lid.*

Tape your sketch to the inside of the top panel, as shown in Figure 3-107, and then use a sharp pointed tool, such as a pick, to press through and mark the plastic at the center of each hole. The indentations will help to center your bit when you drill the holes. Remember that you'll need to make multiple holes to vent the sound from the loudspeaker, which will be beneath the top panel of the box. The result is shown in Figure 3-108.

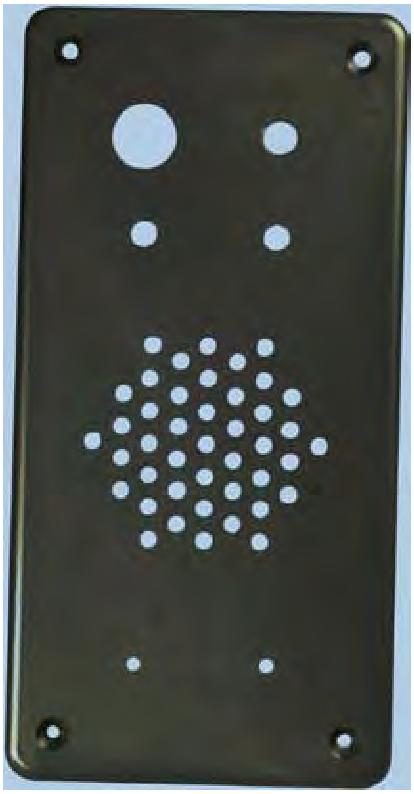

Figure 3-108. *The exterior of the panel after drilling. A small handheld cordless drill can create a neat result if the holes were marked carefully.*

I placed all the components on the top panel, with the exception of the power input jack, which I positioned at one end of the box. Naturally, each hole has to be sized to fit its component, and if you have calipers, they'll be very useful for taking measurements and selecting the right drill bit. Otherwise, make your best guess, too small being better than too large. A deburring tool is ideal for slightly enlarging a hole so that a component fits snugly. This may be necessary if you drill 3/16-inch holes for your 5-inch mm LEDs. Fractionally enlarge each hole, and the LEDs should push in very snugly.

If your loudspeaker lacks mounting holes, you'll have to glue it in place. I used five-minute epoxy to do this. Be careful not to use too much. You don't want any of the glue to touch the speaker cone.

Drilling large holes in the thin, soft plastic of a project box can be a problem. The drill bit tends to dig in and create a mess. You can approach this problem in one of three ways:

1. Use a Forstner drill bit if you have one. It creates a very clean hole.
2. Drill a series of holes of increasing size.
3. Drill a smaller hole than you need, and enlarge it with a deburring tool.

Regardless of which approach you use, you'll need to clamp or hold the top panel of the project box with its outside surface face-down on a piece of scrap wood. Then drill from the inside, so that your bit will pass through the plastic and into the wood.

Finally, mount the components in the panel, as shown in Figure 3-109, and turn your attention to the underneath part of the box.

The circuit board will sit on the bottom, held in place with four #4-size machine screws (bolts) with washers and nylon-insert locknuts. You need to use locknuts to eliminate the risk of a nut working loose and falling among components where it can cause a short circuit.

You'll have to cut the perfboard to fit, taking care not to damage any of the components on it. Also check the underside the board for loose fragments of copper traces after you finish cutting.

Drill bolt holes in the board, if necessary, taking care again not to damage any components. Then mark through the holes to the plastic bottom of the box, and drill the box. Countersink the holes (i.e., bevel the edges of a hole so that a flat-headed screw will fit into it flush with the surrounding surface), push the little bolts up from underneath, and install the circuit board. Be extremely careful not to attach the circuit board too tightly to the project box. This can impose bending stresses, which may break a joint or a copper trace on the board.

I like to include a soft piece of plastic under the board to absorb any stresses. Because you're using locknuts, which will not loosen, there's no need to make them especially tight.

Test the circuit again after mounting the circuit board, just in case.

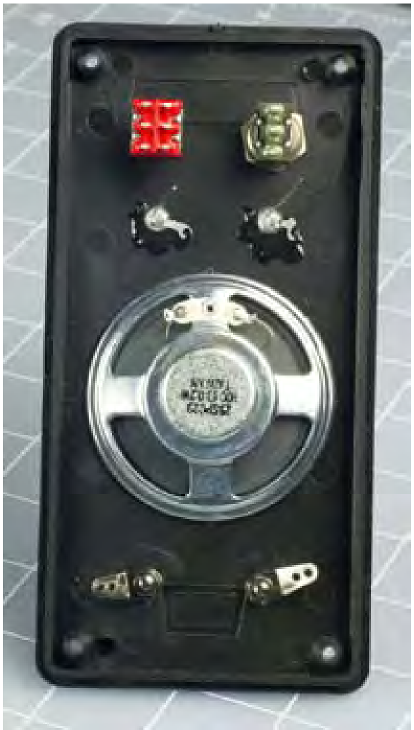

Figure 3-109. *Components have been added to the control panel of the project box (seen from the underside). The loudspeaker has been glued in place. Spare glue was dabbed onto the LEDs, just in case. The SPDT on/off switch is at the top right, the DPDT pushbutton is at top left, and the binding posts, which will connect with the network of magnetic sensor switches, are at the bottom.*

Soldering the Switches

Figure 3-110 shows how the physical switches should be wired together. Remember that S1 is a toggle switch and S2 is a DPDT pushbutton. Your first step is to decide which way up they should be. Use your meter to find out which terminals are connected when the switch is flipped, and when the button is pressed. You'll probably want the switch to be on when the toggle is flipped upward. Be especially careful with the orientation of the pushbutton, because if you wire it upside-down, it will constantly have the alarm in "test" mode, which is not what you want.

Remember, the center terminal of any double-throw switch is almost always the pole of the switch, connecting with the terminals immediately above it and below it.

Stranded wire is appropriate to connect the circuit board with the components in the top panel, because the strands flex easily and impose less stress on solder joints. Twisting each pair of wires together helps to minimize the mess.

Remember to install the LEDs with their short, negative wires connected with the resistor. This will entail some wire-to-wire soldering. You may want to protect some of these bare leads and joints with thin heat-shrink tubing, to minimize the risk of short circuits when you push all the parts into the box.

When you connect wires or components with the lugs on the switches, your pencil-style soldering iron probably won't deliver enough heat to make good joints. You can use your higher-powered soldering iron in these locations, but you absolutely must apply a good heat sink to protect the LEDs when you attach them, and don't allow the iron to remain in contact with anything for more than 10 seconds. It will quickly melt insulation, and may even damage the internal parts of the switches.

In projects that are more complex than this one, it would be good practice to link the top panel with the circuit board more neatly. Multicolored ribbon cable is ideal for this purpose, with plug-and-socket connectors that attach to the board. For this introductory project, I didn't bother. The wires just straggle around, as shown in Figure 3-111.

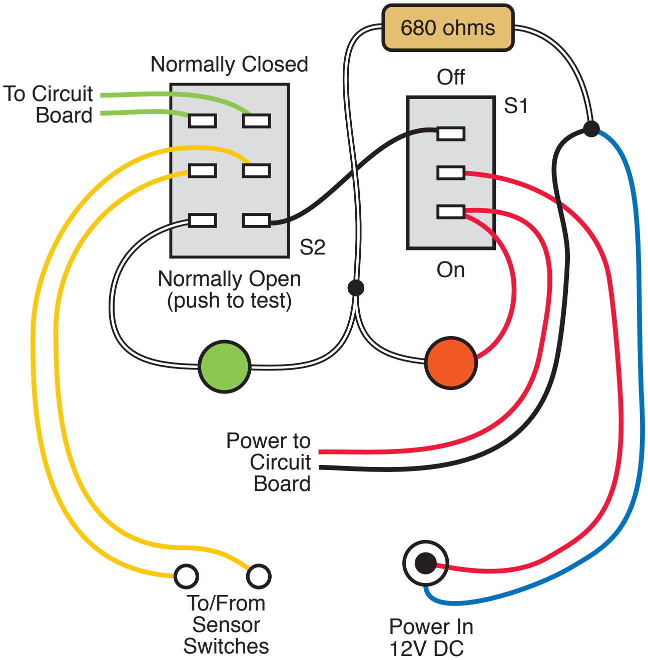

Figure 3-110. *The components can be wired together like this to replicate the circuit shown in The red and green circles are LEDs. Small, solid black circles indicate wire-to-wire solder joints.*

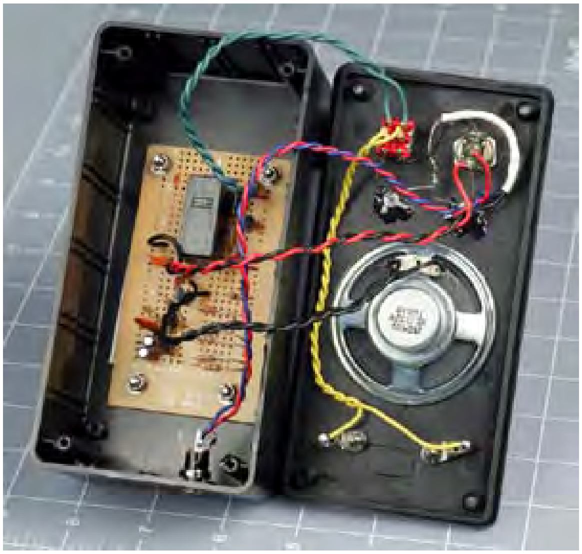

Figure 3-111. *The circuit board has been installed in the base of the project box, and the power input jack has been screwed into the end of the box. Twisted wire-pairs have been connected on a point-to-point basis, without much concern over neatness, as this is a relatively small project. The white insulation at the top-right corner of the front panel is heat-shrink tube that encloses a solder joint and the 680Ω load resistor. Soldering wires to the pushbutton switch requires care and precision, as the contacts are closely spaced.*

Final Test

When you've completed the circuit, test it! If you don't have your network of magnetic sensor switches set up yet, you can just use a piece of wire to connect the two binding posts. Make sure that S1 is in its Off position, then solder the appropriate plug to your 12-volt power source, and plug it into the power jack. When you press the button, the green LED should light up to show continuity between the two binding posts. Now disconnect the wire between the binding posts, press the button again, and the green LED should remain dark.

Reconnect the binding posts, flip S1 to its On position, and the red LED should light up. Press the button, and the alarm should start. Reset it by turning S1 off and then on again; then disconnect the wire between the binding posts. Again, the alarm should start, and it should continue even if you reconnect the wire.

If everything works the way it should, it's time to screw the top of the box in place, pushing the wires inside. Because you're using a large box, you should have no risk of metal parts touching each other accidentally, but still, proceed carefully.

Alarm Installation

Before you install your magnetic sensor switches, you should test each one by moving the magnetic module near the switch module and then away from it, while you use your meter to test continuity between the switch terminals. The switch should close when it's next to the magnet, and open when the magnet is removed.

Now draw a sketch of how you'll wire your switches together. Always remember that they have to be in series, not in parallel! Figure 3-112 shows the concept in theory. The two terminals are the binding posts on top of your control box (which is shown in green), and the dark red rectangles are the magnetic sensor switches on windows and doors. Because the wire for this kind of installation usually has two conductors, you can lay it as I've indicated but cut and solder it to create branches. The solder joints are shown as orange dots. Note how current flows through all the switches in series before it gets back to the control box.

Figure 3-113 shows the same network as you might actually install it in a situation where you have two windows and a door. The blue rectangles are the magnetic modules that activate the switch modules.

You'll need a large quantity of wire, obviously. The type of white, stranded wire that is sold for doorbells or furnace thermostats is good. Typically, it is 20-gauge or larger.

After you install all the switches, clip your meter leads to the wires that would normally attach to the alarm box. Set your meter to test continuity, and open each window or door, one at a time, to check whether you're breaking the continuity. If everything is OK, attach the alarm wires to the binding posts on your project box.

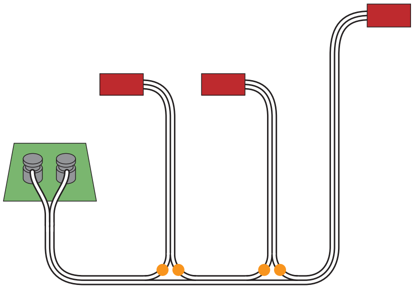

Figure 3-112. *Dual-conductor, white insulated wire can be used to connect the terminals on the alarm control box with magnetic sensors (shown in dark red). Because the sensors must be in series, the wire is cut and joined at positions marked with orange dots.*

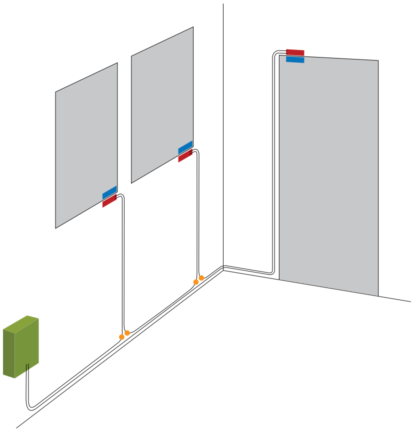

Figure 3-113. *In an installation involving two windows and a door, the magnetic components of the sensors (blue rectangles) could be placed as shown, while the switches (dark red) are located alongside them.*

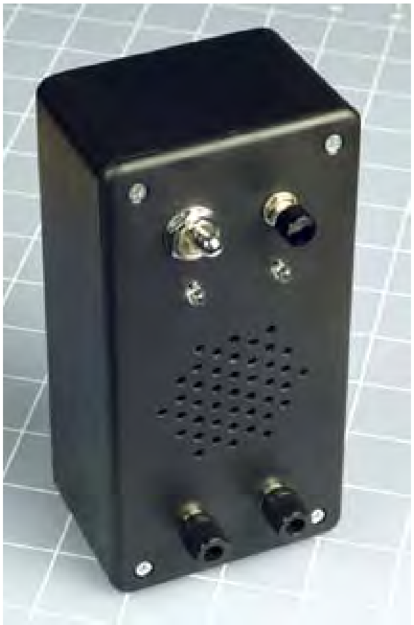

Figure 3-114. *The intrusion alarm completed and in its project box.*

Now deal with the power supply. Use your AC adapter, set to 12 volts, hooked up to your type N power plug, or attach the power plug to a 12-volt alarm battery.

If you use a battery, be especially careful that the wire leading to the center terminal of your power plug is positive! A 12-volt battery can deliver substantial current, which can fry your components if you connect it the wrong way around. It would be a shame to destroy your entire project at the very last step.

The only remaining task is to label the switch, button, power socket, and binding posts on the alarm box. You know that the switch turns the power on and off, and the button tests the circuit and the noisemaker, but no one else knows, and you might want to allow a guest to use your alarm while you're away. For that matter, months or years from now, you may forget some details. Will you remember that the power source for this unit should be 12 volts?

Labeling really is a good idea. But as you can see in Figure 3-114, I haven't quite gotten around to it for the box that I built.

Conclusion

The alarm project has taken you through the basic steps that you will usually follow any time you develop something:

1. Draw a schematic and make sure that you understand it.
2. Modify it to fit the pattern of conductors on a breadboard.
3. Install components on the breadboard and test the basic functions.
4. Modify or enhance the circuit, and retest.
5. Transfer to perforated board, test, and trace faults if necessary.
6. Add switches, buttons, power jack, and plugs or sockets to connect the circuit with the outside world.
7. Mount everything in a box (and add labeling).

While going through this sequence, I hope you've learned the basics of electricity, along with some simple electrical theory, and fundamentals about electronic components. This knowledge should enable you to move on to the much more powerful realm of integrated circuit—which I'll cover in Chapter 4.

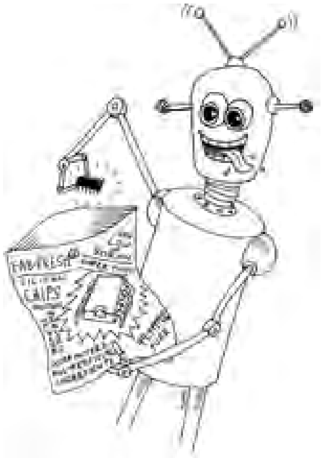

4

Chips, Ahoy!

IN THIS CHAPTER

Shopping List: Experiments 16 Through 24
Experiment 16: Emitting a Pulse
Experiment 17: Set Your Tone
Experiment 18: Reaction Timer
Experiment 19: Learning Logic
Experiment 20: A Powerful Combination
Experiment 21: Race to Place
Experiment 22: Flipping and Bouncing
Experiment 23: Nice Dice
Experiment 24: Intrusion Alarm Completed

Before I get into the fascinating topic of integrated circuit (IC) chips, I have to make a confession: some of the things I asked you to do in Chapter 3 could have been done a bit more simply. Does this mean you have been wasting your time? No, I firmly believe that by building circuits with old-fashioned components—capacitors, resistors, and transistors—you acquire the best possible understanding of the principles of electronics. Still, you are going to find that integrated circuit chips, containing dozens, hundreds, or even thousands of transistor junctions, will enable some shortcuts.

Shopping List: Experiments 16 Through 24

Tools

The only new tool that I recommend using in conjunction with chips is a logic probe. This tells you whether a single pin on a chip has a high or low voltage, which can be helpful in figuring out what your circuit is doing. The probe has a memory function so that it will light its LED, and keep it lit, in response to a pulse that may have been too quick for the eye to see.

Search online and buy the cheapest logic probe you can find. I don't have any specific brand recommendations. The one shown in Figure 4-1 is fairly typical.

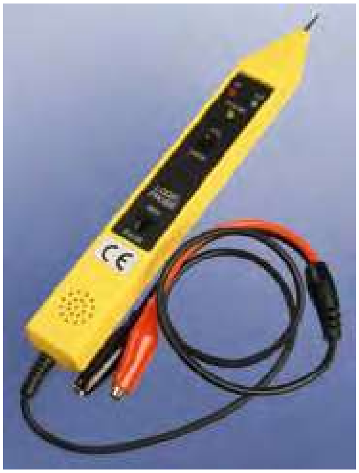

Figure 4-1. *A logic probe detects the high or low voltage on each pin of a chip, and reveals pulses that may occur too quickly for you to perceive them with the unaided eye.*

Supplies

Integrated circuit chips

If you buy everything on this shopping list, and you bought basic parts such as resistors and capacitors that were listed previously, you should have everything you need for all the projects in this chapter.

As chips are quite cheap (currently around 50 cents apiece), I suggest you buy extras. This way, if you damage one, you'll have some in reserve. You'll also have a stock for future projects.

Please read the next section, "Fundamentals: Choosing chips," before you begin chip-shopping. Chips should be easily obtainable from all the major electronics retail suppliers, and sometimes are found on eBay shops. Look in the appendix for a complete list of URLs.

FUNDAMENTALS

Choosing chips

Figure 4-2 shows what is often referred to as an integrated circuit (IC). The circuit is actually etched onto a tiny wafer or "chip" of silicon, embedded in a black plastic body, which is properly referred to as the "package." Tiny wires inside the package link the circuit with the two rows of pins on either side. Throughout this book, I will use the word "chip" to refer to the whole object, including its pins, as this is the most common usage.

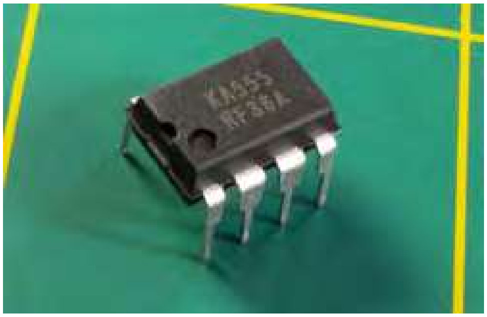

Figure 4-2. *An integrated circuit chip in Plastic Dual-Inline Pin package, abbreviated PDIP, or, more often, DIP.*

The pins are mounted at intervals of 1/10 inch in two rows spaced 3/10 inch apart. This format is known as a Plastic Dual Inline Package, abbreviated PDIP, or, more often, just DIP. The chip in the photograph has four pins in each row; others may have many more. The first thing you need to know, when shopping for chips, is that you'll only be using the DIP package. This book will not be featuring the more modern type, known as "surface-mount," because they're much smaller, more difficult to handle, and require special tools that are relatively expensive. Figure 4-3 shows a size comparison between a 14-pin DIP package and a 14-pin surface-mount package. Many surface-mount chips are even smaller than the one shown.

Just about every chip has a part number printed on it. In Figure 4-2, the part number is KA555. In Figure 4-3, the DIP chip's part number is M74HC00B1, and the surface-mount chip is a 74LVC07AD. You can ignore the second line of numbers and/or letters on each chip, as they are not part of the part number.

Notice in Figure 4-3 that even though the chips look quite different from each other, they both have "74" in their part numbers. This is because both of them are members of the "7400" family of logic chips, which originally had part numbers from 7400 and upward (7400, 7401, 7402, 7403, and so on). Often they are now referred to as "74xx" chips, where "xx" includes all the members of the family. I'll be using this family a lot, so you need to know how to buy them. I'll give you some advice on that without going into details yet about what the chips actually do.

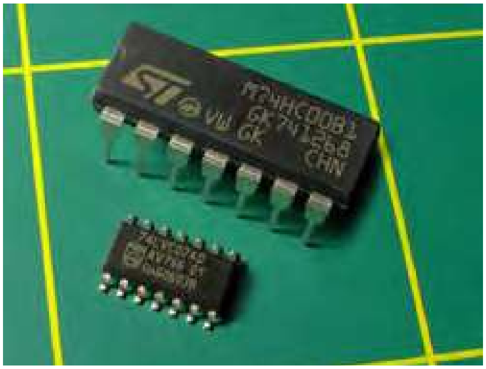

Figure 4-3. *The DIP chip, at the rear, has pins spaced 1/10 inch apart, suitable for insertion in a breadboard or perforated board. It can be soldered without special tools. The small-outline integrated circuit (SOIC) surface-mount chip (foreground) has solder tabs spaced at 1/20 inch. Other surface-mount chips have pins spaced at 1/40 inch or even less (these dimensions are often expressed in millimeters). Surface-mount chips are designed primarily for automated assembly and are difficult to work with manually. In this photo, the yellow lines are 1 inch apart to give you an idea of the scale.*

Take a look at Figure 4-4, which shows how to interpret a typical part number in a 74xx family member. The initial letters identify the manufacturer (which you can ignore, as it really makes no difference for our purposes). Skip the letters until you get to the "74." After that, you find two more letters, which are important. The 74xx family has evolved through many generations, and the letter(s) inserted after the "74" tell you which generation you're dealing with. Some generations have included:

- 74L
- 74LS
- 74C
- 74HC
- 74AHC

And there are more. Generally speaking, subsequent generations tend to be faster or more versatile than previous generations. In this book, for reasons I'll explain later, we are mostly using the HC generation.

FUNDAMENTALS

Choosing chips (continued)

After the letters identifying the generation, you'll find two (sometimes more) numerals. These identify the specific function of the chip. You can ignore any remaining letters and numerals. Looking back at Figure 4-3, the DIP chip part number, M74HC00B1 tells you that it is a chip in the 74xx family, HC generation, with its function identified by numerals 00. The surface-mount chip number, 74LVC07AD, tells you that it is in the 74xx family, LVC generation, with function identified by numerals 07. For convenience we could refer to the first chip as a "74HC00" and the second chip as a "74HC07" because, regardless of their different manufacturers and package sizes, the fundamental behavior of the circuit inside remains the same.

The purpose of this long explanation is to enable you to interpret catalog listings when you go chip shopping. You can search for "74HC00" and the online vendors are usually smart enough to show you appropriate chips from multiple manufacturers, even though there are letters preceding and following the term that you're searched for.

Suppose a circuit requires a 74HC04 chip. If you search for "74HC04" on the website of a parts supplier, you may find versions such as the CD74HC04M96 by Texas Instruments, the 74HC04N by NXP Semiconductors, or MM74HC04N by Fairchild Semiconductor. Because they all have "74HC04" in the middle, any of them will work.

Just be careful that you buy the larger DIP-style package, not the surface-mount package. If the part number has an "N" on the end, you can be sure that it's a DIP package. If there is no "N" on the end, it may or may not be a DIP package, and you will have to check a photo or additional description to make sure. If the part number begins with SS, SO, or TSS, it's absolutely definitely surface-mount, and you don't want it. Many catalogs show photographs of the chips to assist you in buying the right package style.

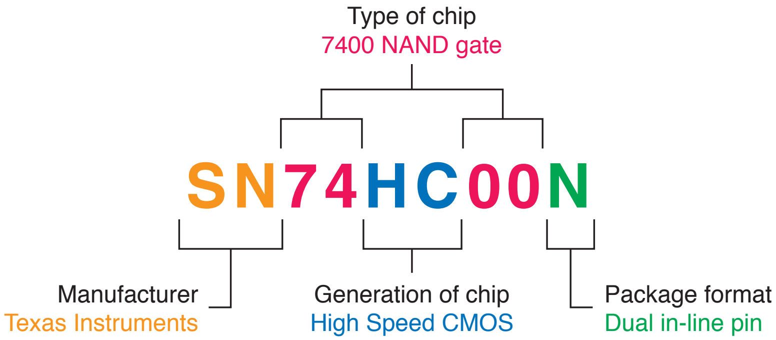

Figure 4-4. *Look for the chip family (74xx, in this case) with the correct generation (HC, in this case) embedded in the number. Make sure you buy the DIP version, not the surface-mount version. The manufacturer is unimportant.*

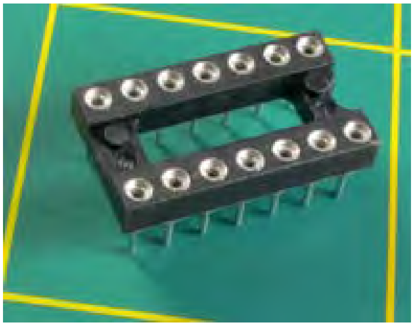

Figure 4-5. *When you're soldering a circuit onto perforated board, sockets eliminate the risk of overheating integrated circuit chips and reduce the risk of zapping them with static electricity, and enable easy replacement.*

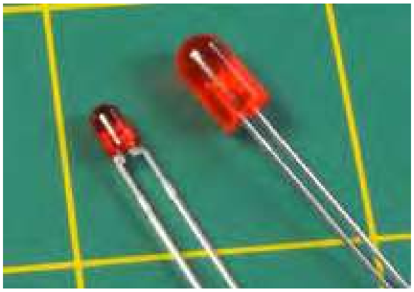

Figure 4-6. *An HC series logic chip is rated to deliver only 4mA at each pin. This is inadequate to drive a typical 5 mm LED (right), which is rated for 20mA forward current. Miniature, low-current LEDs (left) will use as little as 1mA in series with a suitable resistor, and are ideal for test circuits in which you want to see the output with a minimum of hassle.*

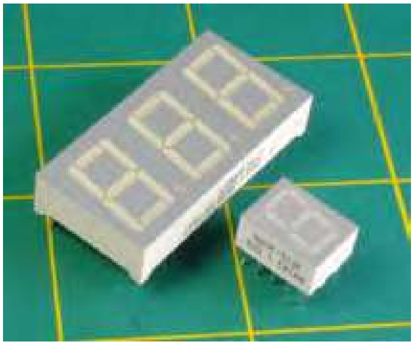

Figure 4-7. *Seven-segment displays are the simplest way to show a numeric output and can be driven directly by some CMOS chips. For finished projects, they are typically mounted behind transparent red acrylic plastic panels.*

Here's your chip list:

- 555 Timer. STMicroelectronics SA555N, Fairchild NE555D, RadioShack TLC555, or similar. Do not get the "CMOS" version of this chip, or any fancy versions such as those of high precision. Buy the cheapest you can find. Quantity: 10. The chip in Figure 4-2 is a 555 timer.
- Logic chips types 74HC00, 74HC02, 74HC04, 74HC08, 74HC32, and 74HC86. Actual part numbers could be M74HC00B1, M74HC02B1, M74HC04B1, and so on, by STMicroelectronics, or SN74HC00N, SN74HC02N, SN74HC04N, and so on, by Texas Instruments. Any other manufacturers are acceptable.
- Remember, each part number should have "HC" in the middle of it, and you want the DIP or PDIP package, not surface-mount. Quantity of each: At least 4.
- 4026 Decade Counter (a chip that counts in tens). Texas Instruments CD-4026BE or similar. Quantity: 4 (you'll need 3, but because this is a CMOS chip sensitive to static electricity, you should have at least one in reserve). Any chip with "4026" in its part number should be OK.
- 74LS92 counter chip, 74LS06 open-collector inverter chip, and 74LS27 triple-input NOR chip. Quantity: 2 of each. Note the "LS" in these part numbers! There will be one experiment in which I want you to use the LS generation instead of the HC generation.

IC sockets

I suggest that you avoid soldering chips directly onto perforated board. If you damage them, they're difficult to remove. Buy some DIP sockets, solder the sockets onto the board, and then plug the chips into the sockets. You can use the cheapest sockets you can find (you don't need gold-plated contacts for our purposes). You will need 8-pin, 14-pin, and 16-pin DIP sockets, such as parts 276-1995, 276-1999, and 276-1998 from RadioShack. See Figure 4-5. Quantity of each: 5 minimum.

Low-power LEDs

The logic chips that you'll be using are not designed to deliver significant power. You'll need to add transistors to amplify their output if you want to drive bright LEDs or relays. Because adding transistors is a hassle, I suggest an alternative: Special low-power LEDs that will draw as little as 1mA, such as the Everlight model T-100 Low Current Red, part number HLMPK150. Figure 4-6 shows a size comparison with a regular 5mm LED. Quantity: 10 (at a minimum).

LED numeric displays

In at least one of our projects, you'll want to illuminate some seven-segment LED numerals. You'll need either three individual numerals, or one package containing three numerals, such as the Kingbright High Efficiency Red Diffused, part number BC56-11EWA, which will be specifically referred to in schematics in this book. If you buy a different seven-segment display, it must be an LED with a "common cathode." (Don't buy liquid-crystal LCD numerals; they require different electronics to drive them.) If you have a choice of power consumption, buy whichever product consumes the least current. See Figure 4-7.

Latching relays

You're going to need a 5-volt latching relay that has two coils, instead of one. The first coil flips the relay one way; the second coil flips it back. The relay consumes no additional power while remaining passively in each state. I suggest the Panasonic DS2E-SL2-DC5V relay. If you buy a different relay, it must be dual-coil latching to run off 5 volts DC, switching at least 1 amp, in a "2 form C" package, to fit your breadboard.

Potentiometers

You'll need 5K, 10K, and 100K linear potentiometers (one of each). Also, a 10K trimmer potentiometer (which you may find described just as a "trimmer"). The manufacturer is unimportant.

Voltage regulators

Because many logic chips require precisely 5 volts DC, you need a voltage regulator to deliver this. The LM7805 does the job. Here again, the chip number will be preceded or followed with an abbreviation identifying the manufacturer and package style, as in the LM7805CT from Fairchild. Any manufacturer will do, but the package should look like the one in Figure 4-8, and if you have a choice, buy a regulator that can deliver at least 1 amp.

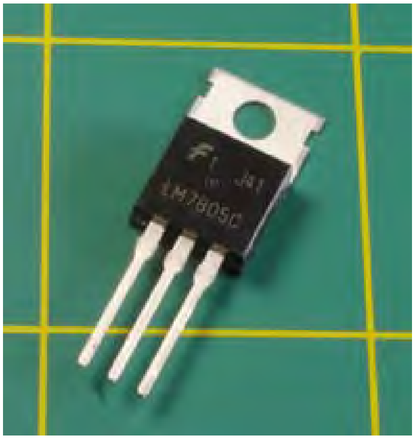

Figure 4-8. *Many integrated circuit chips require a controlled power supply of 5 volts, which can be delivered by this regulator when you apply 7.5 to 9 volts to it. The lefthand pin is for positive input, the center pin is a common ground, and the righthand pin is the 5V output. For currents exceeding 250mA, you should bolt the regulator to a metal heat sink using the hole at the top.*

Tactile switches

These are SPST pushbuttons (momentary switches), usually with four legs. Look for the ALPS part number SKHHAKA010 or any similar item that has pins to fit your breadboard or perforated board. See Figure 4-9.

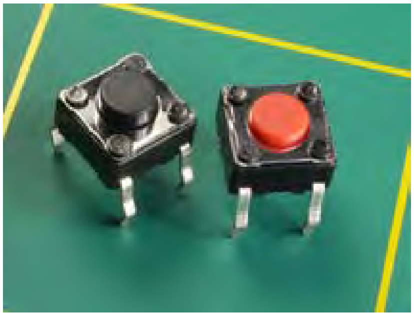

Figure 4-9. *A tactile switch delivers tactile feedback through your fingertip when you press it. They are almost always SPST pushbuttons designed for mounting in circuit boards with standard 1/10-inch hole spacing.*

12-key numeric keypad

Velleman "12 keys keyboard with common output" (no part number, but has been available through All Electronics under catalog code KP-12). Quantity: 1.

This type of keypad has the same layout as an old-fashioned touchtone phone. It should have at least 13 pins or contacts, 12 of which connect with individual pushbuttons, the thirteenth connecting with the other side of all the pushbuttons. In other words, the last pin is "common" to all of them, and this type of keypad is often described as having a "common output." The type of keypad that you don't want is "matrix-encoded," with fewer than 13 contacts, requiring additional external circuitry. See Figures 4-10 and 4-11. If you can't find the Velleman keypad that I suggest, look carefully at keypad descriptions and photographs to make sure that the one you buy is not matrix-encoded and has a common terminal.

Alternatively, you may substitute 12 cheap SPST NO pushbuttons and mount them in a small project box.

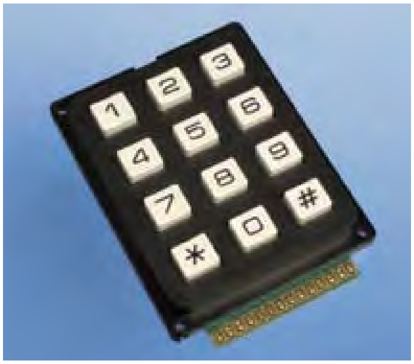

Figure 4-10. *When shopping for a numeric keypad, it should have 12 keys in "touch-tone phone" layout, and should have at least 13 contacts for input/output. The contacts are visible here along the front edge.*

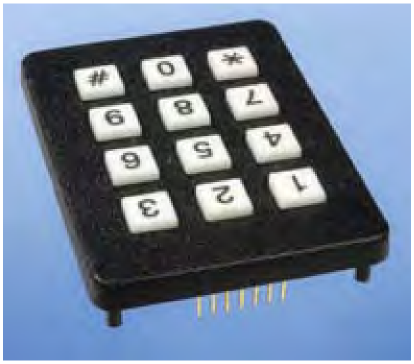

Figure 4-11. *This keypad has insufficient pins and will not work in the circuit in this book.*

BACKGROUND

How chips came to be

The concept of integrating solid-state components into one little package originated with British radar scientist Geoffrey W. A. Dummer, who talked about it for years before he attempted, unsuccessfully, to build one in 1956. The first true integrated circuit wasn't fabricated until 1958 by Jack Kilby, working at Texas Instruments. Kilby's version used germanium, as this element was already in use as a semiconductor. (You'll encounter a germanium diode when I deal with crystal radios in the next chapter of this book.) But Robert Noyce, pictured in Figure 4-12, had a better idea.

Born in 1927 in Iowa, in the 1950s Noyce moved to California, where he found a job working for William Shockley. This was shortly after Shockley had set up a business based around the transistor, which he had coinvented at Bell Labs.

Noyce was one of eight employees who became frustrated with Shockley's management and left to establish Fairchild Semiconductor. While he was the general manager of Fairchild, Noyce invented a silicon-based integrated circuit that avoided the manufacturing problems associated with germanium. He is generally credited as the man who made integrated circuits possible.

Early applications were for military use, as Minuteman missiles required small, light components in their guidance systems. These applications consumed almost all chips produced from 1960 through 1963, during which time the unit price fell from around $1,000 to $25 each, in 1963 dollars.

In the late 1960s, medium-scale integration chips emerged, each containing hundreds of transistors. Large-scale integration enabled tens of thousands of transistors on one chip by the mid-1970s, and today's chips can contain as many as several billion transistors.

Robert Noyce eventually cofounded Intel with Gordon Moore, but died unexpectedly of a heart attack in 1990. You can learn more about the fascinating early history of chip design and fabrication at *http://www.siliconvalleyhistorical.org*.

Figure 4-12. *This picture of Robert Noyce, late in his career, is from the Wikimedia Commons.*

Experiment 16: Emitting a Pulse

I'm going to introduce you to the most successful chip ever made: the 555 timer. As you can find numerous guides to it online, you might question the need to discuss it here, but I have three reasons for doing so:

1. It's unavoidable. You simply have to know about this chip. Some sources estimate that more than 1 billion are still being manufactured annually. It will be used in one way or another in most of the remaining circuits in this book.
2. It provides a perfect introduction to integrated circuits, because it's robust, versatile, and illustrates two functions that we'll be dealing with later: comparators and a flip-flop.
3. After reading all the guides to the 555 that I could find, beginning with the original Fairchild Semiconductor data sheet and making my way through various hobby texts, I concluded that its inner workings are seldom explained very clearly. I want to give you a graphic understanding of what's happening inside it, because if you don't have this, you won't be in a good position to use the chip creatively.

You will need:

- 9-volt power supply.
- Breadboard, jumper wires, and multimeter.
- 5K linear potentiometer. Quantity: 1.
- 555 timer chip. Quantity: 1.
- Assorted resistors and capacitors.
- SPST tactile switches. Quantity: 2.
- LED (any type). Quantity: 1.

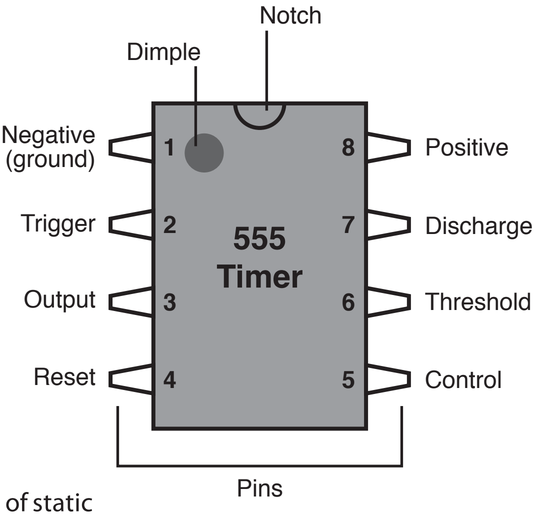

Figure 4-13. *The 555 timer chip, seen from above. Pins on chips are always numbered counterclockwise, from the top-left corner, with a notch in the body of the chip uppermost, or a circular indentation at top-left, to remind you which end is up.*

Procedure

The 555 chip is very robust, but still, in theory, you can zap it with a jolt of static electricity and kill it. Therefore, to be on the safe side, you should ground yourself before handling it. See the "Grounding yourself" warning on page 172 for details. Although this warning primarily refers to the type of chips known as CMOS, which are especially vulnerable, grounding yourself is always a sensible precaution.

Look for a small circular indentation, called the *dimple*, molded into the body of the chip, and turn the chip so that the indentation is at the top-left corner with the pins pointing down. Alternatively, if your chip is of the type with a notch at one end, turn the chip so that the notch is at the top.

The pins on chips are always numbered counterclockwise, starting from the top-left pin (next to the dimple). See Figure 4-13, which also shows the names of the pins on the 555 timer, although you don't need to know most of them just yet.

Insert the chip in your breadboard so that its pins straddle the channel down the center. Now you can easily feed voltages to the pins on either side, and read signals out of them. See Figure 4-14 for a precise guide to placement, in the first project. The timer is identified as "IC1," because "IC" is the customary abbreviation for "Integrated Circuit."

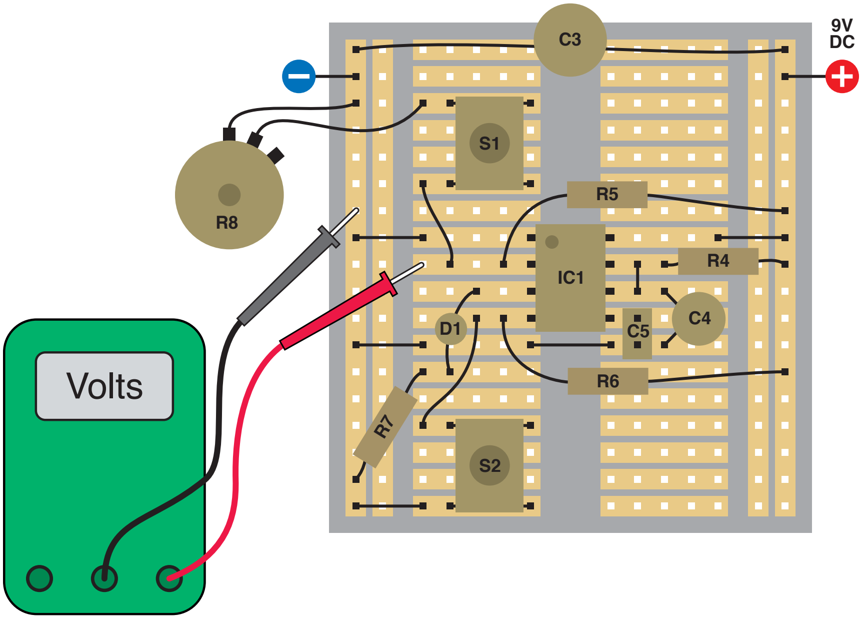

Figure 4-14. *This circuit allows you to explore the behavior of the 555 timer chip. Use your meter to monitor the voltage on pin 2 as shown. There are no resistors labeled R1, R2, or R3 and no capacitors labeled C1 or C2, because they'll be added in a later schematic. Component values in this schematic:*

R4: 100K
R5: 2K2
R6: 10K
R7: 1K
R8: 5K linear potentiometer
C3: 100 μF electrolytic
C4: 47 μF electrolytic
C5: 0.1 μF ceramic
IC1: 555 timer
S1, S2: SPST tactile switches (pushbuttons)
D1: Generic LED

R5 holds the trigger (pin 2) positive until S1 is pressed, which lowers the voltage depending on the setting of potentiometer R8. When the trigger voltage falls below 1/3 of the power supply, the chip's output (pin 3) goes high for a period determined by the values of R4 and C4. S2 resets (zeros) the timer, by reducing the voltage to pin 4, the Reset. C3 smoothes the power supply, and C5 isolates pin 5, the control, so that it won't interfere with the functioning of this test circuit. (We'll use the control pin in a future experiment.)

All integrated circuit chips require a power supply. The 555 is powered with negative voltage applied to pin 1 and positive to pin 8. If you reverse the voltage accidentally, this can permanently damage the chip, so place your jumper wires carefully.

Set your power supply to deliver 9 volts. It will be convenient for this experiment if you supply positive down the righthand side and negative down the lefthand side of the breadboard, as suggested in Figure 4-14. C3 is a large capacitor, at least 100 µF, which is placed across the power supply to smooth it out and provide a local store of charge to fuel fast-switching circuits, as well as to guard against other transient dips in voltage. Although the 555 isn't especially fast-switching, other chips are, and you should get into the habit of protecting them.

Begin with the potentiometer turned all the way counterclockwise to maximize the resistance between the two terminals that we're using, and when you apply the probe from your meter to pin 2, you should measure about 6 volts when you press S1.

Now rotate the potentiometer clockwise and press S1 again. If the LED doesn't light up, keep turning the potentiometer and pressing and releasing the button. When you've turned the potentiometer about two-thirds of the way, you should see the LED light up for just over 5 seconds when you press and release the button. Here are some facts that you should check for yourself:

- The LED will keep glowing after you release the button.
- You can press the button for any length of time (less than the timer's cycle time) and the LED always emits the same length of pulse.
- The timer is triggered by a fall in voltage on pin 2. You can verify this with your meter.
- The LED is either fully on or fully off. You can't see a faint glow when it's off, and the transition from off to on and on to off is very clean and precise.

Check Figure 4-16 to see how the components should look on your breadboard, and then look at the schematic in Figure 4-15 to understand what's happening. I will be adding more components later, which I will be labeling R1, R2, C1, and C2 to be consistent with data sheets that you may see for the 555 timer. Therefore, in this initial circuit the resistors are labeled R4 and up, and capacitors C3 and up.

When S1 (the tactile switch) is open, pin 2 of the 555 timer receives positive power through R5, which is 2K2. Because the input resistance of the timer is very high, the voltage on pin 2 is almost the full 9 volts.

When you press the button, it connects negative voltage through R8, the 5K potentiometer to pin 2. Thus, R8 and R5 form a voltage divider with pin 2 in the middle. You may remember this concept from when you were testing transistors. The voltage between the resistances will change, depending on the values of the resistances.

If R8 is turned up about halfway, it is approximately equal to R5, so the midpoint, connected to pin 2, has about half the 9-volt power supply. But when you turn the potentiometer so that its resistance falls farther, the negative voltage outweighs the positive voltage, so the voltage on pin 2 gradually drops.

If you have clips on your meter leads, you can hook them onto the nearest jumper wires and then watch the meter while you turn the potentiometer up and down and press the button.

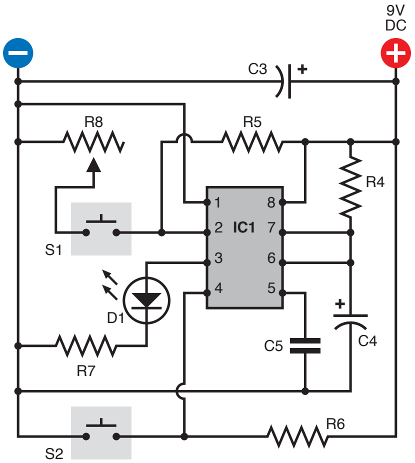

Figure 4-15. *A schematic view of the circuit shown in Figure 4-14. Throughout this chapter, the schematics will be laid out to emulate the most likely placement of components on a breadboard. This is not always the simplest layout, but will be easiest for you to build. Refer to Figure 4-14 for the values of the components.*

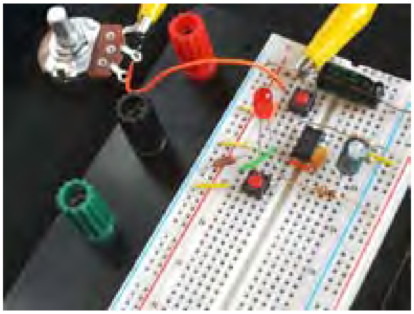

Figure 4-16. *This is how the components look when installed on the breadboard. The alligator clips are attached to a patch cord that links the 100 µF capacitor to the potentiometer. The power supply input is not shown.*

The graphs in Figure 4-17 illustrate what is happening. The upper graph shows the voltage applied to pin 2 by random button-presses, with the potentiometer turned to various values. The lower graph shows that the 555 is triggered if, and only if, the voltage on pin 2 actively drops from above 3 volts to below 3 volts. What's so special about 3 volts? It's one-third of our 9-volt power supply.

Here's the take-home message:

- The output of the 555 (pin 3) emits a *positive* pulse when the trigger (pin 2) drops *below* one-third of the supply voltage.
- The 555 delivers the *same duration* of positive pulse every time (so long as you don't supply a prolonged low voltage on pin 2).
- A *larger* value for R4 or for C4 will *lengthen* the pulse.
- When the output (pin 3) is high, the voltage is almost equal to the supply voltage. When the output goes low, it's almost zero.

The 555 converts the imperfect world around it into a precise and dependable output. It doesn't switch on and off absolutely instantly, but is fast enough to *appear* instant.

Now here's another thing to try. Trigger the timer so that the LED lights up. While it is illuminated, press S2, the second button, which grounds pin 4, the reset. The LED should go out immediately.

When the reset voltage is pulled *low*, the output goes *low*, regardless of what voltage you apply to the trigger.

There's one other thing I want you to notice before we start using the timer for more interesting purposes. I included R5 and R6 so that when you first switch on the timer, it is not emitting a pulse—but is ready to do so. These resistors apply a positive voltage to the trigger and the reset pin, to make sure that the 555 timer is ready to run when you first apply power to it.

As long as the *trigger* voltage is *high*, the timer *will not* emit a pulse. (It emits a pulse when the trigger voltage drops.)

As long as the *reset* voltage is *high*, the timer is *able* to emit a pulse. (It shuts down when the reset voltage drops.)

R5 and R6 are known as "pull-up resistors" because they pull the voltage up. You can easily overwhelm them by adding a direct connection to the negative side of the power supply. A typical pull-up resistor for the 555 timer is 10K. With a 9-volt power supply, it only passes 0.9mA (by Ohm's Law).

Finally, you may be wondering about the purpose of C5, attached to pin 5. This pin is known the "control" pin, which means that if you apply a voltage to it, you can control the sensitivity of the timer. I'll get to this in more detail a little later. Because we are not using this function right now, it's good practice to put a capacitor on pin 5 to protect it from voltage fluctuations and prevent it from interfering with normal functioning.

Make sure you become familiar with the basic functioning of the 555 timer before you continue.

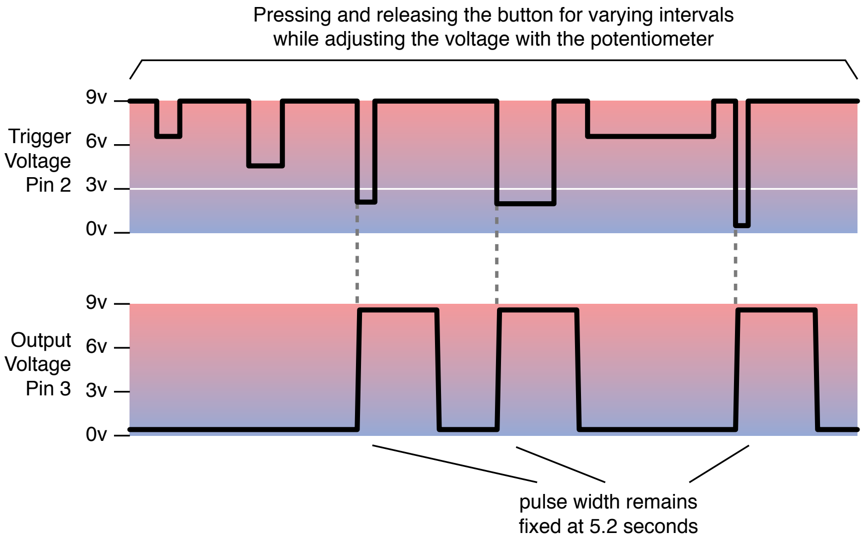

Figure 4-17. *The top graph shows voltage on the trigger (pin 2) when the pushbutton is pressed, for different intervals, at different settings of the potentiometer. The lower graph shows the output (pin 3), which rises until it is almost equal to the power supply, when the voltage on pin 2 drops below 1/3 the full supply voltage.*

FUNDAMENTALS

The following table shows 555 pulse duration in monostable mode:

- Duration is in seconds, rounded to two figures.
- The horizontal scale shows common resistor values between pin 7 and positive supply voltage.
- The vertical scale shows common capacitor values between pin 6 and negative supply voltage.

To calculate a different pulse duration, multiply resistance × capacitance × 0.0011 where resistance is in kilohms, capacitance is in microfarads, and duration is in seconds.

47 µF	0.05	0.11	0.24	0.52	1.1	2.4	5.2	11	24	52
22 µF	0.02	0.05	0.11	0.24	0.53	1.1	2.4	5.3	11	24
10 µF	0.01	0.02	0.05	0.11	0.24	0.52	1.1	2.4	5.2	11
4.7 µF		0.01	0.02	0.05	0.11	0.24	0.52	1.1	2.4	5.2
2.2 µF			0.01	0.02	0.05	0.11	0.24	0.53	1.1	2.4
1.0 µF				0.01	0.02	0.05	0.11	0.24	0.52	1.1
0.47 µF					0.01	0.02	0.05	0.11	0.24	0.52
0.22 µF						0.01	0.02	0.05	0.11	0.24
0.1 µF							0.01	0.02	0.05	0.11
0.047 µF								0.01	0.02	0.05
0.022 µF									0.01	0.02
0.01 µF										0.01
	1K	2K2	4K7	10K	22K	47K	100K	220K	470K	1M

THEORY

Inside the 555 timer: monostable mode

The plastic body of the 555 timer contains a wafer of silicon on which are etched dozens of transistor junctions in a pattern that is far too complex to be explained here. However, I can summarize their function by dividing them into groups, as shown in Figure 4-18. An external resistor and two external capacitors are also shown, labeled the same way as in Figure 4-15.

The negative and positive symbols inside the chip are power sources which actually come from pins 1 and 8, respectively. I omitted the internal connections to those pins for the sake of clarity.

The two yellow triangles are "comparators." Each comparator compares two inputs (at the base of the triangle) and delivers an output (from the apex of the triangle) depending on whether the inputs are similar or different. We'll be using comparators for other purposes later in this book.

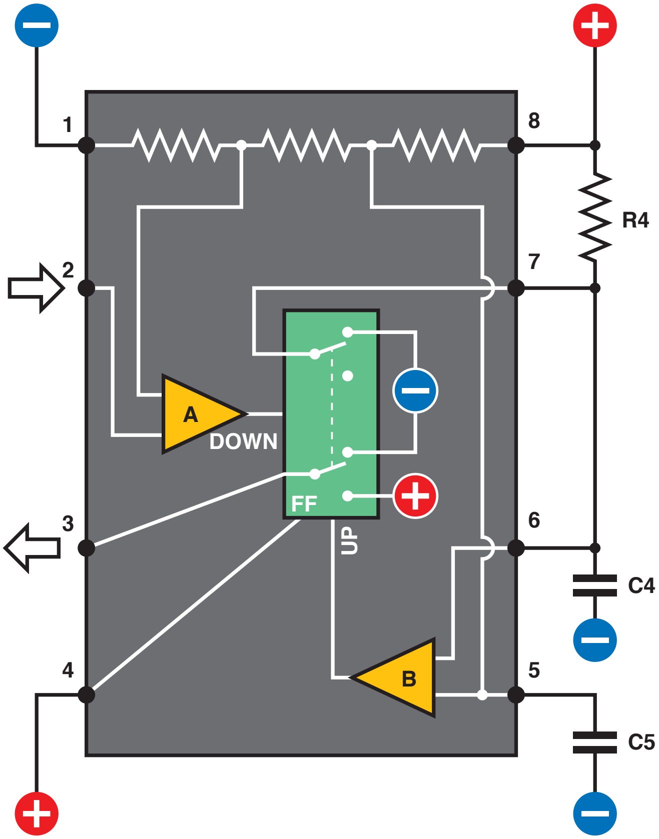

Figure 4-18. *Inside the 555 timer. White lines indicate connections inside the chip. A and B are comparators. FF is a flip-flop which can rest in one state or the other, like a double-throw switch. A drop in voltage on pin 2 is detected by comparator A, which triggers the flip-flop into its "down" position and sends a positive pulse out of pin 3. When C4 charges to 2/3 of supply voltage, this is detected by comparator B, which resets the flip-flop to its "up" position. This discharges C4 through pin 7.*

THEORY

Inside the 555 timer: monostable mode (continued)

The green rectangle, identified as "FF," is a "flip-flop." I have depicted it as a DPDT switch, because that's how it functions here, although of course it is really solid-state.

Initially when you power up the chip, the flip-flop is in its "up" position which delivers low voltage through the output, pin 3. If the flip-flop receives a signal from comparator A, it flips to its "down" state, and flops there. When it receives a signal from comparator B, it flips back to its "up" state, and flops there. The "UP" and "DOWN" labels on the comparators will remind you what each one does when it is activated.

Flip-flops are a fundamental concept in digital electronics. Computers couldn't function without them.

Notice the external wire that connects pin 7 with capacitor C4. As long as the flip-flop is "up," it sinks the positive voltage coming through R4 and prevents the capacitor from charging positively.

If the voltage on pin 2 drops to 1/3 of the supply, comparator A notices this, and flips the flip-flop. This sends a positive pulse out of pin 3, and also disconnects the negative power through pin 7. So now C4 can start charging through R4. While this is happening, the positive output from the timer continues.

As the voltage increases on the capacitor, comparator B monitors it through pin 6, known as the threshold. When the capacitor accumulates 2/3 of the supply voltage, comparator B sends a pulse to the flip-flop, flipping it back into its original state. This discharges the capacitor through pin 7, appropriately known as the discharge pin. Also, the flip-flop ends the positive output through pin 3 and replaces it with a negative voltage. This way, the 555 returns to its original state.

I'll sum up this sequence of events very simply:

1. Initially, the flip-flop grounds the capacitor and grounds the output (pin 3).
2. A drop in voltage on pin 2 to 1/3 the supply voltage or less makes the output (pin 3) positive and allows capacitor C4 to start charging through R4.
3. When the capacitor reaches 2/3 of supply voltage, the chip discharges the capacitor, and the output at pin 3 goes low again.

In this mode, the 555 timer is "monostable," meaning that it just gives one pulse, and you have to trigger it again to get another.

You adjust the length of each pulse by changing the values of R4 and C4. How do you know which values to choose? Check the table on page 157, which gives an approximate idea and also includes a formula so that you can calculate values of your own.

I didn't bother to include pulses shorter than 0.01 second in the table, because a single pulse of this length is usually not very useful. Also I rounded the numbers in the table to 2 significant figures, because capacitor values are seldom more accurate than that.

BACKGROUND

How the timer was born

Back in 1970, when barely a half-dozen corporate seedlings had taken root in the fertile ground of Silicon Valley, a company named Signetics bought an idea from an engineer named Hans Camenzind. It wasn't a huge breakthrough concept—just 23 transistors and a bunch of resistors that would function as a programmable timer. The timer would be versatile, stable, and simple, but these virtues paled in comparison to its primary selling point. Using the emerging technology of integrated circuits, Signetics could reproduce the whole thing on a silicon chip.

This entailed some trial and error. Camenzind worked alone, building the whole thing initially on a large scale, using off-the-shelf transistors, resistors, and diodes on a breadboard. It worked, so then he started substituting slightly different values for the various components to see whether the circuit would tolerate variations during production and other factors such as changes in temperature when the chip was in use. He made at least 10 different versions of the circuit. It took months.

Next came the crafts work. Camenzind sat at a drafting table and used a specially mounted X-Acto knife to scribe his circuit into a large sheet of plastic. Signetics then reduced this image photographically by a ratio of about 300:1. They etched it into tiny wafers, and embedded each of them in a half-inch rectangle of black plastic with the product number printed on top. Thus, the 555 timer was born.

It turned out to be the most successful chip in history, both in the number of units sold (tens of billions and counting) and the longevity of its design (unchanged in almost 40 years). The 555 has been used in everything from toys to spacecraft. It can make lights flash, activate alarm systems, put spaces between beeps, and create the beeps themselves.

Today, chips are designed by large teams and tested by simulating their behavior using computer software. Thus, chips inside a computer enable the design of more chips. The heyday of solo designers such as Hans Camenzind is long gone, but his genius lives inside every 555 timer that emerges from a fabrication facility. (If you'd like to know more about chip history, see *http://www.semiconductormuseum.com/Museum_Index.htm*.)

Figure 4-13. *Hans Camenzind, inventor and developer of the 555 timer chip for Signetics.*

FUNDAMENTALS

Why the 555 is useful

In its monostable mode (which is what you just saw), the 555 will emit a single pulse of fixed (but programmable) length. Can you imagine some applications? Think in terms of the pulse from the 555 controlling some other component. A motion sensor on an outdoor light, perhaps. When an infra-red detector "sees" something moving, the light comes on for a specific period—which can be controlled by a 555.

Another application could be a toaster. When someone lowers a slice of bread, a switch will close that triggers the toasting cycle. To change the length of the cycle, you could use a potentiometer instead of R4 and attach it to the external lever that determines how dark you want your toast. At the end of the toasting cycle, the output from the 555 would pass through a power transistor, to activate a solenoid (which is like a relay, except that it has no switch contacts) to release the toast.

Intermittent windshield wipers could be controlled by a 555 timer—and on earlier models of cars, they actually were.

And what about the burglar alarm that was described at the end of Chapter 3? One of the features that I listed, which has not been implemented yet, is that it should shut itself off after a fixed interval. We can use the change of output from a 555 timer to do that.

The experiment that you just performed seemed trivial, but really it implies all kinds of possibilities.

555 timer limits

1. The timer can run from a stable voltage source ranging from 5 to 15 volts.
2. Most manufacturers recommend a range from 1K to 1M for the resistor attached to pin 7.
3. The capacitor value can go as high as you like, if you want to time really long intervals, but the accuracy of the timer will diminish.
4. The output can deliver as much as 100mA at 9 volts. This is sufficient for a small relay or miniature loudspeaker, as you'll see in the next experiment.

Beware of Pin-Shuffling!

In all of the schematics in this book, I'll show chips as you'd see them from above, with pin 1 at top left. Other schematics that you may see, on websites or in other books, may do things differently. For convenience in drawing circuits, people shuffle the pin numbers on a chip so that pin 1 isn't necessarily shown adjacent to pin 2.

Look at the schematic in Figure 4-20 and compare it with the one in Figure 4-15. The connections are the same, but the one in Figure 4-20 groups pins to reduce the apparent complexity of the wiring.

"Pin shuffling" is common because circuit-drawing software tends to do it, and on larger chips, it is necessary for functional clarity of the schematic (i.e., logical groupings of pin names versus physical groupings on memory chips, for example). When you're first learning to use chips, I think it's easier to understand a schematic that shows the pins in their actual positions. So that's the practice I will be using here.

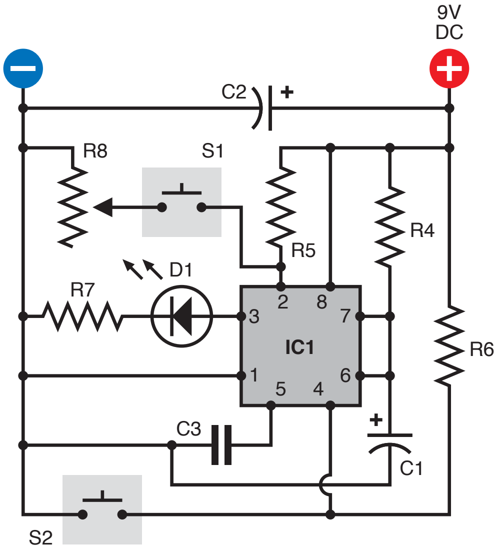

Figure 4-20. *Many people draw schematics in which the pin numbers on a chip are shuffled around to make the schematic smaller or simpler. This is not helpful when you try to build the circuit. The schematic here is for the same circuit as in Figure 4-15. This version would be harder to recreate on a breadboard.*

Experiment 17: Set Your Tone

I'm going to show you two other ways in which the 555 timer can be used.

You will need the same items as in Experiment 16, plus:

- Additional 555 timer chip. Quantity: 1.
- Miniature loudspeaker. Quantity: 1.
- 100K linear potentiometer. Quantity: 1.

Procedure

Leave the components from Experiment 16 where they are on the breadboard, and add the next section below them, as shown in Figures 4-21 and 4-22. Resistor R2 is inserted between pins 6 and 7, instead of the jumper wire that shorted the pins together in the previous circuit, and there's no external input to pin 2 anymore. Instead, pin 2 is connected via a jumper wire to pin 6. The easiest way to do this is by running the wire across the top of the chip.

I have omitted the smoothing capacitor from the schematic in Figure 4-22, because I'm assuming that you're running this circuit on the same breadboard as the first, where the previous smoothing capacitor is still active.

A loudspeaker in series with a 100Ω resistor (R3) has been substituted for the LED to show the output from the chip. Pin 4, the reset, is disabled by connecting it to the positive voltage supply, as I'm not expecting to use the reset function in this circuit.

Now what happens when you apply power? Immediately, you should hear noise through the loudspeaker. If you don't hear anything, you almost certainly made a wiring error.

Notice that you don't have to trigger the chip with a pushbutton anymore. The reason is that when C1 charges and discharges, its fluctuating voltage is connected via a jumper wire across the top of the chip to pin 2, the trigger. In this way, the 555 timer now triggers itself. I'll describe this in more detail in the next section "Theory: Inside the 555 timer: astable mode," if you want to see exactly what is going on.

In this mode, the chip is "astable," meaning that it is not stable, because it flips to and fro endlessly, sending a stream of pulses for as long as the power is connected. The pulses are so rapid that the loudspeaker reproduces them as noise.

In fact, with the component values that I specified for R1, R2, and C1, the 555 chip is emitting about 1,500 pulses per second. In other words, it creates a 1.5 KHz tone.

Check the table on page 166 to see how different values for R2 and C1 can create different pulse frequencies with the chip in this astable mode. Note that the table assumes a fixed value of 1K for R1!

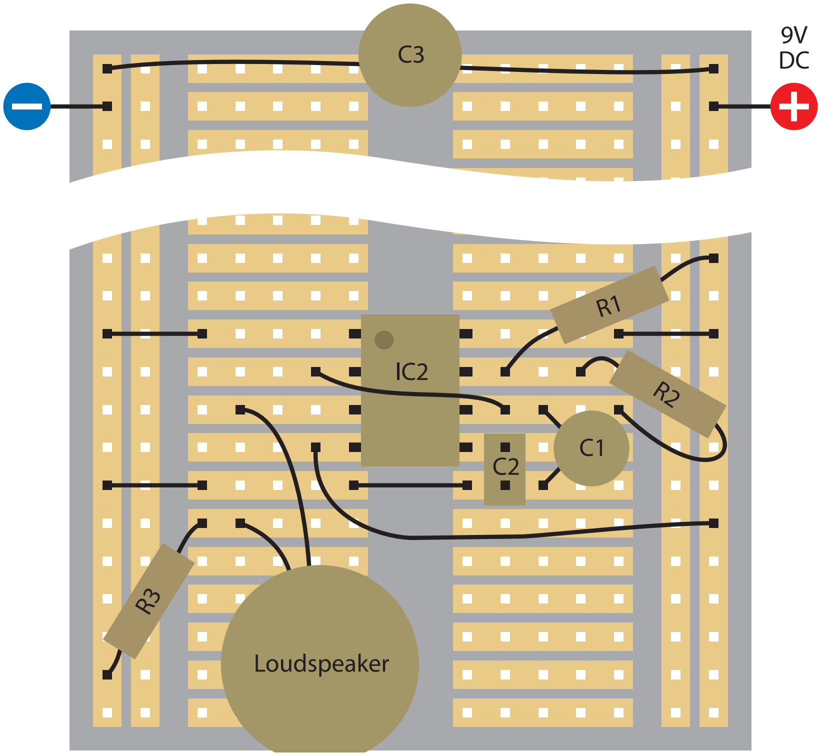
C3
9V
DC
R1
IC2
R2
C1
C2
R3
Loudspeaker

Figure 4-21. *These components should be added on the same breadboard below the components shown in Figure 4-14. Use the following values to test the 555 timer in its astable mode:*

R1: 1K
R2: 10K
R3: 100Ω
C1: 0.047 µF ceramic or electrolytic
C2: 0.1 µF ceramic
IC2: 555 timer

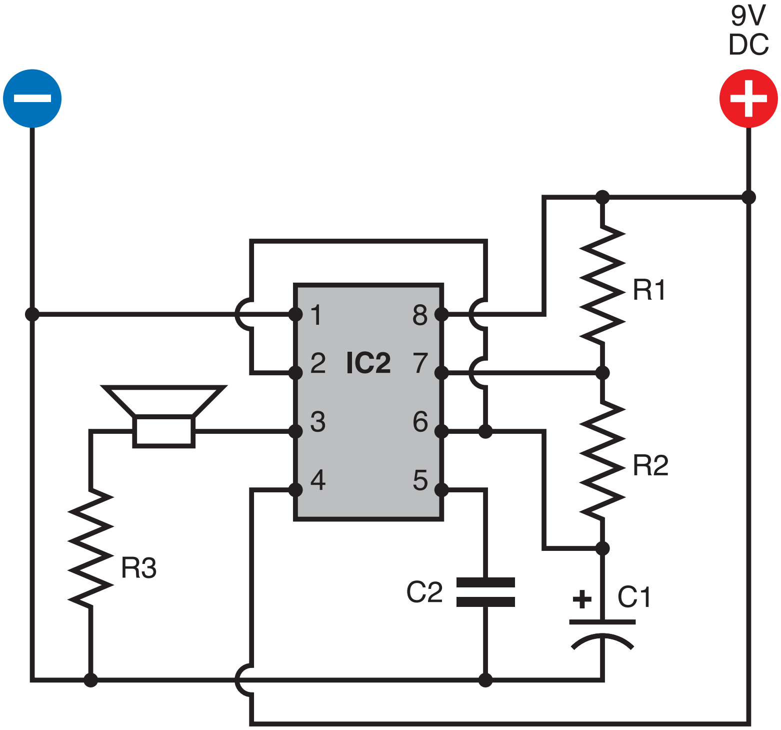
9V
DC
1
2
3
4
8
7
6
5
IC2
R1
R2
R3
C2
C1
+

Figure 4-22. *This is the schematic version of the circuit shown in Figure 4-21. The component values are the same.*

THEORY

Inside the 555 timer: astable mode

Here's what is happening now, illustrated in Figure 4-23. Initially, the flip-flop grounds C1 as before. But now the low voltage on the capacitor is connected from pin 7 to pin 2 through an external wire. The low voltage tells the chip to trigger itself. The flip-flop obediently flips to its "on" position and sends a positive pulse to the loudspeaker, while removing the negative voltage from pin 7.

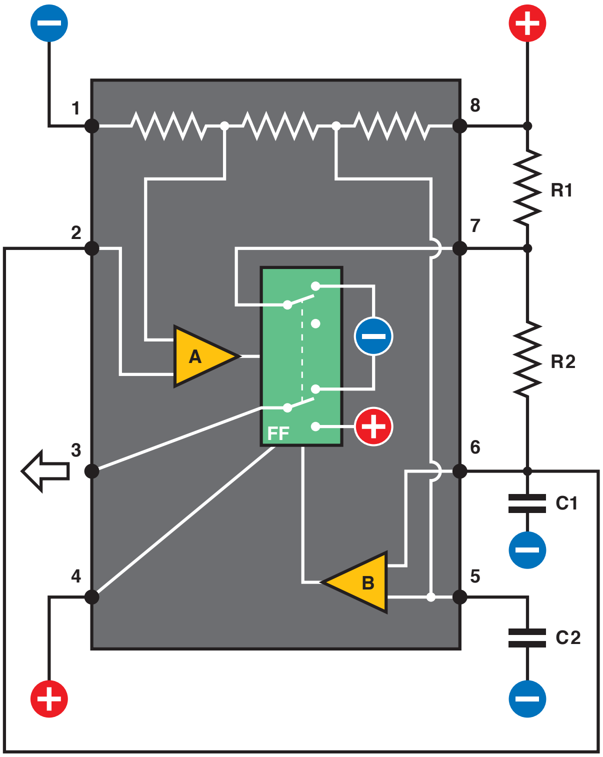

Figure 4-23. *When the 555 timer is used in astable mode, resistor R2 is placed between pin 6 and pin 7, and pin 6 is connected via an external wire to pin 2, so that the timer triggers itself.*

Now C1 starts charging, as it did when the timer was in monostable mode, except that it is being charged through R1 + R2 in series. Because the resistors have low values, and C1 is also small, C1 charges quickly. When it reaches 2/3 full voltage, comparator B takes action as before, discharging the capacitor and ending the output pulse from pin 3.

The capacitor takes longer to discharge than before, because R2 has been inserted between it and pin 7, the discharge pin. While the capacitor is discharging, its voltage diminishes, and is still linked to pin 2. When the voltage drops to 1/3 of full power or less, comparator A kicks in and sends another pulse to the flip-flop, starting the process all over again.

Summing up:

1. In astable mode, as soon as power is connected, the flip-flop pulls down the voltage on pin 2, triggering comparator A, which flips the flip-flop to its "down" position.
2. Pin 3, the output, goes high. The capacitor charges through R1 and R2 in series.
3. When the capacitor reaches 2/3 of supply voltage, the flip-flop goes "up" and the output at pin 3 goes low. The capacitor starts to discharge through R2.
4. When the charge on the capacitor diminishes to 1/3 of full voltage, the pull-down on pin 2 flips the flip-flop again and the cycle repeats.

Unequal on/off cycles

When the timer is running in astable mode, C1 charges through R1 and R2 in series. But when C1 discharges, it dumps its voltage through R2 only. This means that the capacitor charges more slowly than it discharges. While it is charging, the output on pin 3 is high; while it is discharging, the output on pin 3 is low. Consequently the "on" cycle is always longer than the "off" cycle. Figure 4-24 shows this as a simple graph.

If you want the on and off cycles to be equal, or if you want to adjust the on and off cycles independently (for example, because you want to send a very brief pulse to another chip, followed by a longer gap until the next pulse), all you need to do is add a diode, as shown in Figure 4-25.

Now when C1 charges, the electricity flows through R1 as before but takes a shortcut around R2, through diode D1. When C1 discharges, the diode blocks the flow of electricity in that direction, and so the discharge goes back through R2.

THEORY

Inside the 555 timer: astable mode (continued)

R1 now controls the charge time on its own, while R2 controls the discharge time. The formula for calculating the frequency is now:

Frequency = 1440 / ((R1 + R2) × C1)

If you set R1 = R2, you should get almost equal on/off cycles ("almost" because the diode itself imposes a small internal voltage drop of about 0.6V). The exact value depends primarily on the manufacturing process used to make the diode.

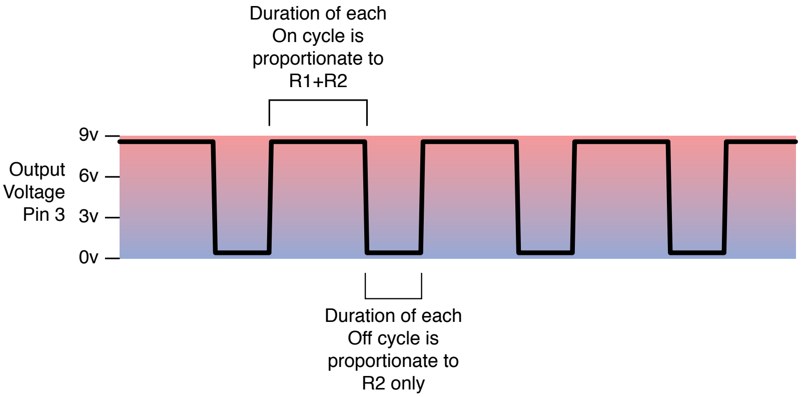

Figure 4-24. *In its usual astable configuration, the timer charges a capacitor through R1+R2 and discharges the capacitor through R2 only. Therefore its output on cycles are longer than its output off cycles.*

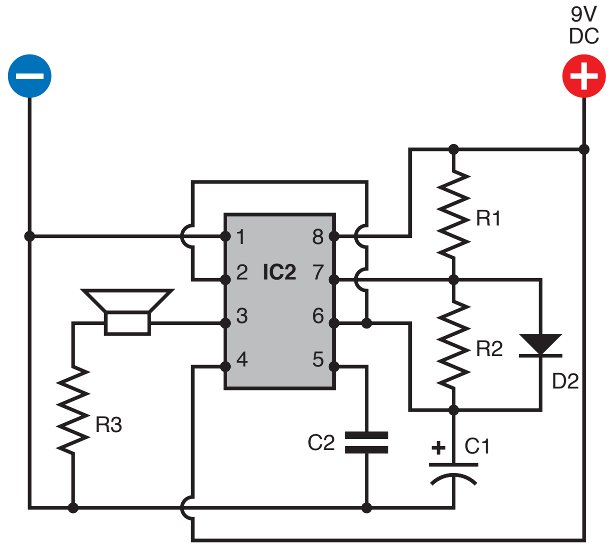

Figure 4-25. *This is a modification of the schematic shown in Figure 4-22. By adding a diode to a 555 timer running in astable mode, we eliminate R2 from the charging cycle of capacitor C1. Now we can adjust the output on cycle with the value of R1, and the output off cycle with the value of R2, so that the on and off durations are independent of each other.*

FUNDAMENTALS

The following table shows 555 timer frequency in astable mode:

- Frequency is in pulses per second, rounded to two figures.
- The horizontal scale shows common resistor values for R2.
- The vertical scale shows common capacitor values for C1. Resistor R1 is assumed to be 1K.
- Resistor R1 is assumed to be 1K.

To calculate a different frequency: double R2, add the product to R1, multiply the sum by C1, and divide the result into 1440. Like this:

Frequency = 1440 / ((R1 + 2R2) × C1) cycles per second

In this formula, R1 and R2 are in kilohms, C1 is in microfarads, and the frequency is in hertz (cycles per second). Note that the frequency is measured from the start of one pulse to the start of the next. The duration of each pulse is not the same as the length of time between each pulse. This issue is discussed in the previous section.

47 µF	10	5.7	3.0	1.5	0.7	0.3	0.2	0.1		
22 µF	22	12	6.3	3.1	1.5	0.7	0.3	0.2	0.1	
10 µF	48	27	14	6.9	3.2	1.5	0.7	0.3	0.2	0.1
4.7 µF	100	57	30	15	6.8	3.2	1.5	0.7	0.3	0.2
2.2 µF	220	120	63	31	15	6.9	3.3	1.5	0.7	0.3
1.0 µF	480	270	140	69	32	15	7.2	3.3	1.5	0.7
0.47 µF	1,000	570	300	150	68	32	15	7	3.3	1.5
0.22 µF	2,200	1,200	630	310	150	69	33	15	7	3.3
0.1 µF	4,800	2,700	1,400	690	320	150	72	33	15	7.2
0.047 µF	10,000	5,700	3,000	1,500	680	320	150	70	33	15
0.022 µF	22,000	12,000	6,300	3,100	1,500	690	330	150	70	33
0.01 µF	48,000	27,000	14,000	6,900	3,200	1,500	720	330	150	72
	1K	2K2	4K7	10K	22K	47K	100K	220K	470K	1M

Astable Modifications

In the circuits shown in Figures 4-22 or 4-25, if you substitute a 100K potentiometer for R2, you can adjust the frequency up and down by turning the shaft.

Another option is to “tune” the timer by using pin 5, the control, as shown in the Figure 4-26. Disconnect the capacitor that was attached to that pin and substitute the series of resistors shown. R9 and R11 are both 1K resistors, either side of R10, which is a 100K potentiometer. They ensure that pin 5 always has at least 1K between it and the positive and negative sides of the power supply. Connecting it directly to the power supply won’t damage the timer, but will prevent it from generating audible tones. As you turn the potentiometer to and fro, the frequency will vary over a wide range. If you want to generate a very specific frequency, a trimmer potentiometer can be used instead.

A primary advantage of using pin 5 to adjust frequency is that you can control it remotely. Take the output from pin 3 of another 555 timer running slowly in astable mode, and pipe it through a 2K2 resistor to pin 5. Now you get a two-tone siren effect, as one timer controls the other. If, in addition, you add a 100 µF capacitor between pin 5 and ground, the charging and discharging of the capacitor will make the tone slide up and down instead of switching abruptly. I'll describe this in more detail shortly. This leads me to the whole topic of one chip controlling another chip, which will be our last variation on this experiment.

Chaining Chips

Generally speaking, chips are designed so that they can talk to each other. The 555 couldn't be easier in this respect:

- Pin 3, the output, from one 555 can be connected directly to pin 2, the trigger, of a second 555.
- Alternatively, the output can be sufficient to provide power to pin 8 of a second 555.
- The output is appropriate to control or power other types of chips too.

Figure 4-27 shows these options.

When the output from the first 555 goes high, it is about 70 to 80% of its supply voltage. In other words, when you're using a 9V supply, the high output voltage is at least 6 volts. This is still above the minimum of 5V that the second chip needs to trigger its comparator, so there's no problem.

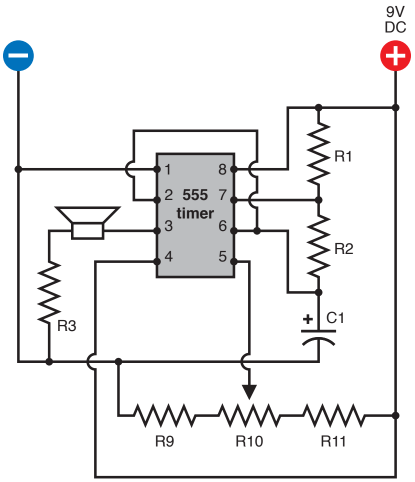
9V DC
1
2
3
4
555 timer
8
7
6
5
R1
R2
R3
C1
R9
R10
R11

Figure 4-28. *The control (pin 5) is seldom used but can be useful. Varying the voltage on it will adjust the speed of the timer. This circuit enables you to test the behavior of it. Component values:*

R1: 1K
R2: 10K
R3: 100 ohms
R9, R11: 1K
R10: 100K linear potentiometer
C1: 0.047 µF

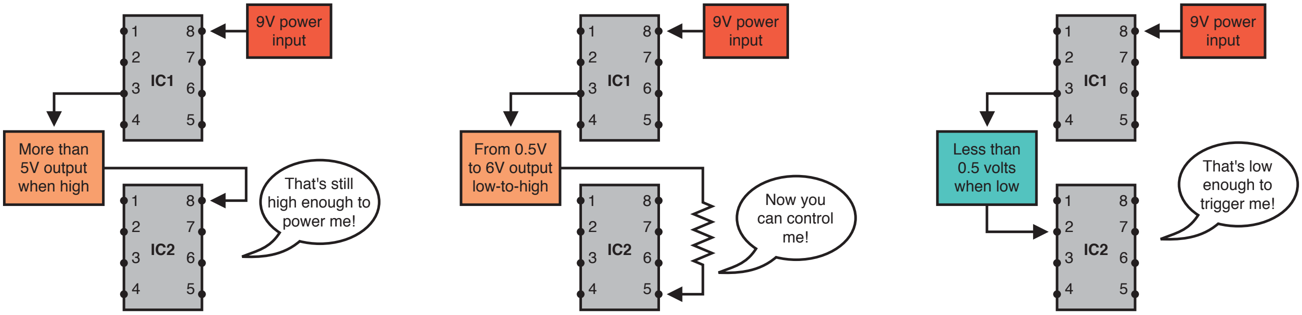
9V power input
IC1
1
2
3
4
8
7
6
5
More than 5V output when high
IC2
That's still high enough to power me!
9V power input
IC1
From 0.5V to 6V output low-to-high
IC2
Now you can control me!
9V power input
IC1
Less than 0.5 volts when low
IC2
That's low enough to trigger me!

Figure 4-27. *Three ways to chain 555 timers together. The output of IC1 can power a second timer, or adjust its control voltage, or activate its trigger pin.*

You can chain together the two 555 timers that you already have on your breadboard. Figure 4-28 shows how to connect the two circuits that were shown previously in Figures 4-15 and 4-22. Run a wire from pin 3 (the output) of the first chip to pin 8 (the positive power supply) of the second chip, and disconnect the existing wire connecting pin 8 to your power supply. The new wire is shown in red. Now when you press the button to activate the first chip, its output powers the second chip.

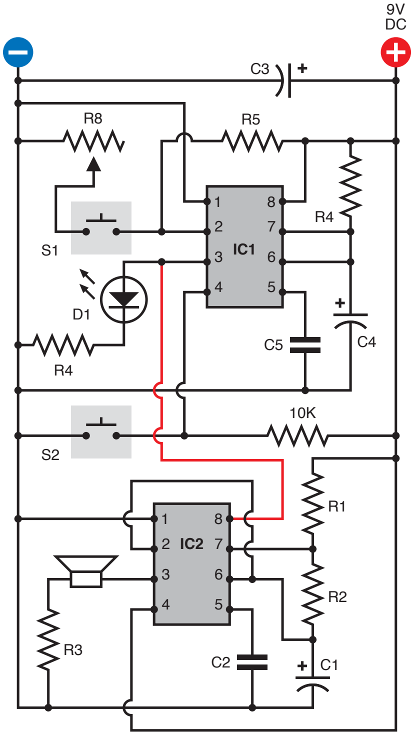

Figure 4-28. *You can combine the two circuits shown in Figures 4-15 and 4-22 simply by disconnecting the wire that provides power to pin 8 of the second timer, and running a substitute wire (shown in red).*

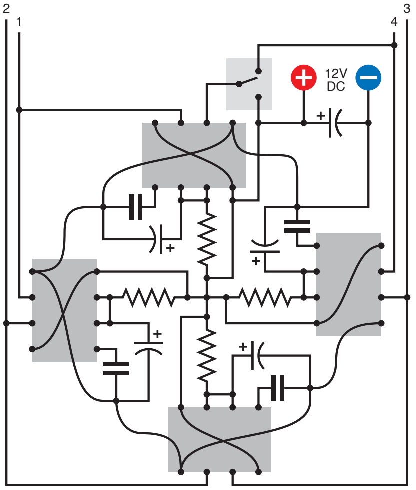

Figure 4-29. *Four 555 timers, chained together in a circle, can flash a series of four sets of LEDs in sequence, like Christmas lights or a movie marquee.*

You can also use the output from one chip to trigger another (i.e., you can connect pin 3 from the first chip to pin 2 of the second). When the output from the first chip is low, it's less than half a volt. This is well below the threshold that the second chip requires to be activated. Why would you want to do this? Well, you might want to have both timers running in monostable mode, so that the end of a high pulse from the first one triggers the start of a high pulse in the second one. In fact, you could chain together as many timers as you like in this way, with the last one feeding back and triggering the first one, and they could flash a series of LEDs in sequence, like Christmas lights. Figure 4-29 shows how four timers could be linked this way, in a configuration that would occupy minimal space (and would be wired point-to-point on perforated board, not on breadboard-format board). Each of the outputs numbered 1 through 4 would have about enough power to run maybe 10 LEDs, if you used relatively high load resistors to limit their current.

Incidentally, you can reduce the chip count (the number of chips) by using two 556 timers instead of four 555 timers. The 556 contains a pair of 555 timers in one package. But because you have to make the same number of external connections (other than the power supply), I haven't bothered to use this variant.

You can even get a 558 timer that contains four 555 circuits, all preset to function in astable mode. I decided not to use this chip, because its output behaves differently from a normal 555 timer. But you can buy a 558 timer and play with it if you wish. It is ideal for doing the "chain of four timers" that I suggested previously. The data sheet even suggests this.

Lastly, going back to the idea of modifying the frequency of a 555 timer in astable mode, you can chain two timers, as shown in Figure 4-30. The red wire shows the connection from the output of the first timer to the control pin of the second. The first timer has now been rewired in astable mode, so that it creates an oscillating on/off output around four times per second. This output flashes the LED (to give you a visual check of what's going on) and feeds through R7 to the control pin of the second timer.

But C2 is a large capacitor, which takes time to charge through R7. While this happens, the voltage detected by pin 5 slowly rises, so that the tone generated by IC2 gradually rises in pitch. Then IC1 reaches the end of its on cycle and switches itself off, at which point C2 discharges and the pitch of the sound generated by IC2 falls again.

You can tweak this circuit to create all kinds of sounds, much more controllably then when you were using PUT transistors to do the same kind of thing. Here are some options to try:

- Double or halve the value of C2.
- Omit C2 completely, and experiment with the value of R7.
- Substitute a 10K potentiometer for R7.
- Change C4 to increase or decrease the cycle time of IC1.

- Halve the value of R5 while doubling the value of C4, so that the cycle time of IC1 stays about the same, but the On time becomes significantly longer than the Off time.
- Change the supply voltage in the circuit from 9 volts to 6 volts or 12 volts.

Remember, you can't damage a 555 timer by making changes of this kind. Just make sure that the negative side of your power supply goes to pin 1 and the positive side to pin 8.

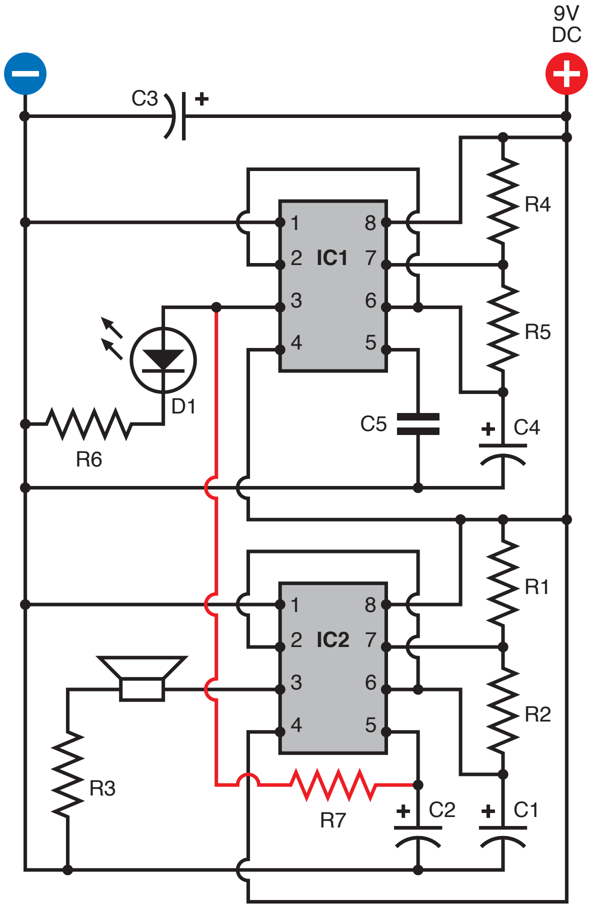

Figure 4-30. *When both timers are astable, but IC1 runs much more slowly than IC2, the output from IC1 can be used to modulate the tone generated by IC2. Note that as this is a substantial modification to the previous schematics, several components have been relabeled. To avoid errors, you may need to remove the old circuit from your breadboard and build this version from scratch. Try these values initially:*

R1, R4, R6, R7: 1K
R2, R5: 10K
R3: 100 ohms
C1: 0.047 µF
C2, C3: 100 µF
C4: 68 µF
C5: 0.1 µF

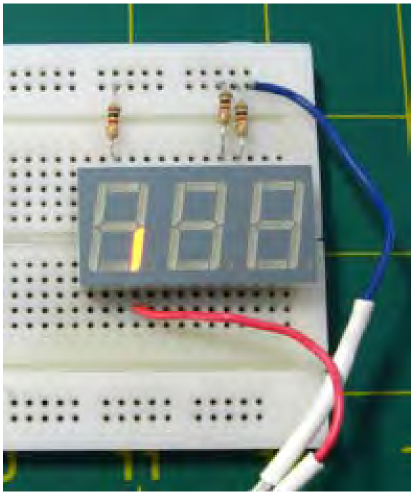

Figure 4-31. *After putting a 1K resistor between the common cathode of the display and the negative supply voltage, you can use the positive supply voltage to illuminate each segment in turn.*

Experiment 18: Reaction Timer

Because the 555 can easily run at thousands of cycles per second, we can use it to measure human reactions. You can compete with friends to see who has the fastest response—and note how your response changes depending on your mood, the time of day, or how much sleep you got last night.

Before going any further, I have to warn you that this circuit will have more connections than others you've tackled so far. It's not conceptually difficult, but requires a lot of wiring, and will only just fit on a breadboard that has 63 rows of holes. Still, we can build it in a series of phases, which should help you to detect any wiring errors as you go.

You will need:

- 4026 chip. Quantity: 4 (really you need only 3, but get another one in case you damage the others).
- 555 timers. Quantity: 3.
- Tactile switches (SPST momentary switches). Quantity: 3.
- Three numeric LEDs, or one 3-digit LED display (see the shopping list at the beginning of this chapter). Quantity: 1.
- Breadboard, resistors, capacitors, and meter, as usual.

Step 1: Display

You can use three separate LED numerals for this project, but I suggest that you buy the Kingbright BC56-11EWA on the shopping list at the beginning of this chapter. It contains three numerals in one big package.

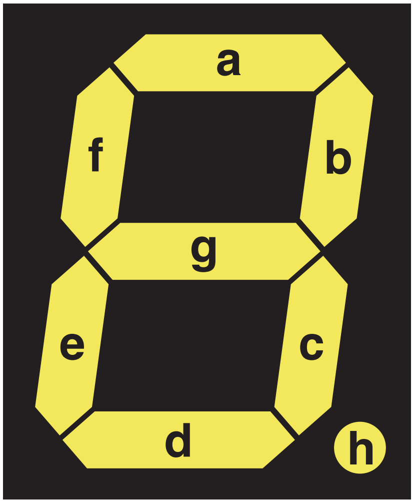

Figure 4-32. *The most basic and common digital numeral consists of seven LED segments identified by letters, as shown here, plus an optional decimal point.*

You should be able to plug it into your breadboard, straddling the center channel. Put it all the way down at the bottom of the breadboard, as shown in Figure 4-31. Don't put any other components on the breadboard yet.

Now set your power supply to 9 volts, and apply the *negative* side of it to the row of holes running up the breadboard on the *righthand* side. Insert a 1K resistor between that negative supply and each of pins 18, 19, and 26 of the Kingbright display, which are the "common cathode," meaning the negative connection shared by each set of LED segments in the display. (The pin numbers of the chip are shown in Figure 4-33. If you're using another model of display, you'll have to consult a data sheet to find which pin(s) are designed to receive negative voltage.)

Switch on the power supply and touch the free end of the positive wire to each row of holes serving the display on its left and right sides. You should see each segment light up, as shown in Figure 4-31.

Each numeral from 0 to 9 is represented by a group of these segments. The segments are always identified with lowercase letters *a* through *g*, as shown in Figure 4-32. In addition, there is often a decimal point, and although we won't be using it, I've identified it with the letter *h*.

Check Figure 4-33 showing the Kingbright display, and you'll see I have annotated each pin with its function. You can step down the display with the positive wire from your power supply, making sure that each pin lights an appropriate segment.

Incidentally, this display has two pins, numbered 3 and 26, both labeled to receive negative voltage for the first of the digits. Why two pins instead of one? I don't know. You need to use only one, and as this is a passive chip, it doesn't matter if you leave the unused one unconnected. Just take care not to apply positive voltage to it, which would create a short circuit.

A numeric display has no power or intelligence of its own. It's just a bunch of light-emitting diodes. It's not much use, really, until we can figure out a way to illuminate the LEDs in appropriate groups—which will be the next step.

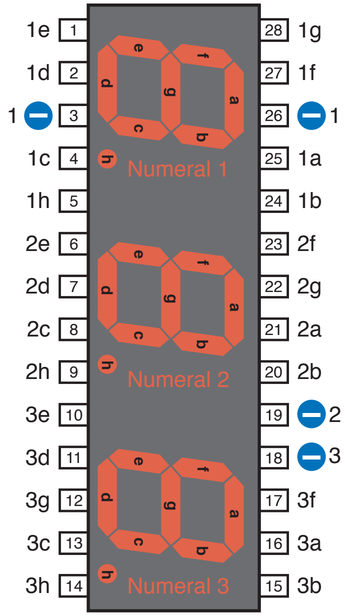

Figure 4-33. *This Kingbright unit incorporates three seven-segment numeric displays in one package, and can be driven by three chained 4026 decade counters. The pin numbers are shown close to the chip. Segments a through g of numeral 1 are identified as 1a through 1g. Segments a through g of numeral 2 are identified as 2a through 2g. Segments a through g of numeral 3 are identified as 3a through 3g.*

Step 2: Counting

Fortunately, we have a chip known as the 4026, which receives pulses, counts them, and creates an output designed to work with a seven-segment display so that it shows numbers 0–9. The only problem is that this is a rather old-fashioned CMOS chip (meaning, Complementary Metal Oxide Semiconductor) and is thus sensitive to static electricity. Check the caution on page 172 before continuing.

Switch off your power supply and connect its wires to the top of the breadboard, noting that for this experiment, we're going to need positive and negative power on both sides. See Figure 4-34 for details. If your breadboard does not already have the columns of holes color-coded, I suggest you use Sharpie markers to identify them, to avoid polarity errors that can fry your components.

The 4026 counter chip is barely powerful enough to drive the LEDs in our display when powered by 9 volts. Make sure you have the chip the right way up, and insert it into the breadboard immediately above your three-digit display, leaving just one row of holes between them empty.

The schematic in Figure 4-35 shows how the pins of the 4026 chip should be connected. The arrows tell you which pins on the display should be connected with pins on the counter.

Figure 4-36 shows the "pinouts" (i.e., the functions of each pin) of a 4026 counter chip. You should compare this with the schematic in Figure 4-35.

Include a tactile switch between the positive supply and pin 1 of the 4026 counter, with a 10K resistor to keep the input to the 4026 counter negative until the button is pressed. Make sure all your positives and negatives are correct, and turn on the power. You should find that when you tap the tactile switch lightly, the counter advances the numeric display from 0 through 9 and then begins all over again from 0. You may also find that the chip sometimes misinterprets your button-presses, and counts two or even three digits at a time. I'll deal with this problem a little further on.

The LED segments will not be glowing very brightly, because the 1K series resistors deprive of them of the power they would really like to receive. Those resistors are necessary to avoid overloading the outputs from the counter.

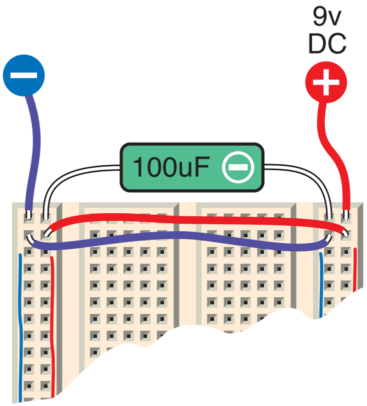

Figure 4-34. *When building circuits around chips, it's convenient to have a positive and negative power supply down each side of your breadboard. For the reaction timer circuit, a 9V supply with a 100 μF smoothing capacitor can be set up like this. If your breadboard doesn't color-code the columns of holes on the left and right sides, I suggest you do that yourself with a permanent marker.*

Grounding Yourself

To avoid the frustration that occurs when you power up a circuit and nothing happens, be sure to take these precautions when you use the older generation of CMOS chips (which often have part numbers from 4000 upward, such as 4002, 4020, and so on):

Chips are often shipped with their legs embedded in black foam. This is electrically conductive foam, and you should keep the chips embedded in it until you are ready to use them.

If the chips are supplied to you in plastic tubes, you can take them out and poke their legs into pieces of conductive foam or, if you don't have any, use aluminum foil. The idea is to avoid one pin on a chip acquiring an electric potential that is much higher than another pin.

While handling CMOS components, grounding yourself is important. I find that in dry weather, I accumulate a static charge merely by walking across a plastic floor-protecting mat in socks that contain some synthetic fibers. You can buy a wrist strap to keep yourself grounded, or simply touch a large metal object, such as a file cabinet, before you touch your circuit board. I am in the habit of working with my socked foot touching a file cabinet, which takes care of the problem.

Never solder a CMOS chip while there is power applied to it.

Grounding the tip of your soldering iron is a good idea.

Better still, don't solder CMOS chips at all. When you're ready to immortalize a project by moving it from a breadboard into perforated board, solder a socket into your perforated board, then push the chip into the socket. If there's a problem in the future, you can unplug the chip and plug in another.

Use a grounded, conductive surface on your workbench. The cheapest way to do this is to unroll some aluminum foil and ground it (with an alligator clip and a length of wire) to a radiator, a water pipe, or a large steel object. I like to use an area of conductive foam to cover my workbench—the same type of foam that is used for packaging chips. However, this foam is quite expensive.

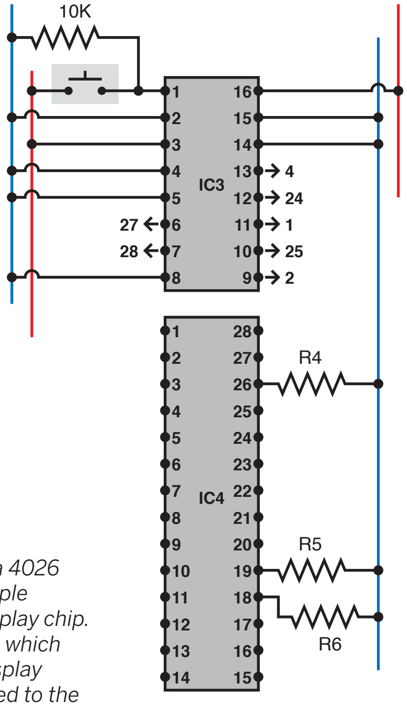
10K
IC3
1
2
3
4
5
6
7
8
16
15
14
13
12
11
10
9
27
28
4
24
1
25
2
IC4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
28
27
26
25
24
23
22
21
20
19
18
17
16
15
R4
R5
R6

Figure 4-35. *IC3 is a 4026 counter. IC4 is a triple seven-segment display chip. The arrows tell you which pins on the LED display should be connected to the pins on the counter.*

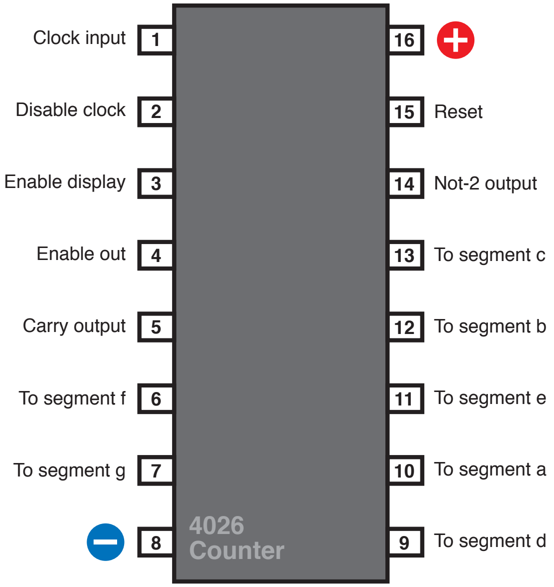
Clock input
1
Disable clock
2
Enable display
3
Enable out
4
Carry output
5
To segment f
6
To segment g
7
8
16
15
Reset
14
Not-2 output
13
To segment c
12
To segment b
11
To segment e
10
To segment a
9
To segment d
4026
Counter

Figure 4-36. *The 4026 decade counter is a CMOS chip that accepts clock pulses on pin 1, maintains a running total from 0 to 9, and outputs this total via pins designed to interface with a seven-segment LED numeric display.*

FUNDAMENTALS

Counters and seven-segment displays

Most counters accept a stream of pulses and distribute them to a series of pins in sequence. The 4026 decade counter is unusual in that it applies power to its output pins in a pattern that is just right to illuminate the segments of a 7-segment numeric display.

Some counters create positive outputs (they "source" current) while others create negative outputs (they "sink" current). Some seven-segment displays require positive input to light up the numbers. These are known as "common cathode" displays. Others require negative input and are known as "common anode" displays. The 4026 delivers positive outputs and requires a common cathode display.

Check the data sheet for any counter chip to find out how much power it requires, and how much it can deliver. CMOS chips are becoming dated, but they are very useful to hobbyists, because they will tolerate a wide range of supply voltage—from 5 to 15 volts in the case of the 4026. Other types of chips are much more limited.

Most counters can source or sink only a few milliamps of output power. When the 4026 is running on a 9-volt power supply, it can source about 4mA of power from each pin. This is barely enough to drive a seven-segment display.

You can insert a series resistor between each output pin of the counter and each input pin of the numeric display, but a simpler, quicker option is to use just one series resistor for each numeral, between the negative-power pin and ground. The experiment that I'm describing uses this shortcut. Its disadvantage is that digits that require only a couple of segments (such as numeral 1) will appear brighter than those that use many segments (such as numeral 8).

If you want your display to look bright and professional, you really need a transistor to drive each segment of each numeral. An alternative is to use a chip containing multiple "op amps" to amplify the current.

When a decade counter reaches 9 and rolls over to 0, it emits a pulse from its "carry" pin. This can drive another counter that will keep track of tens. The carry pin on that counter can be chained to a third counter that keeps track of hundreds, and so on. In addition to decade counters, there are hexadecimal counters (which count in 16s), octal counters (in 8s), and so on.

Why would you need to count in anything other than tens? Consider that the four numerals on a digital clock each count differently. The rightmost digit rolls over when it reaches 10. The next digit to the left counts in sixes. The first hours digit counts to 10, gives a carry signal, counts to 2, and gives another carry signal. The leftmost hours digit is either blank or 1, when displaying time in 12-hour format. Naturally there are counters specifically designed to do all this.

Counters have control pins such as "clock disable," which tells the counter to ignore its input pulses and freeze the display, "enable display," which enables the output from the chip, and "reset," which resets the count to zero.

The 4026 requires a positive input to activate each control pin. When the pins are grounded, their features are suppressed.

To make the 4026 count and display its running total you must ground the "clock disable" and "reset" pins (to suppress their function) and apply positive voltage to the "enable display" pin (to activate the output). See Figure 4-36 to see these pins identified.

Assuming that you succeed in getting your counter to drive the numeric display, you're ready to add two more counters, which will control the remaining two numerals. The first counter will count in ones, the second in tens, and the third in hundreds.

In Figure 4-37, I've continued to use arrows and numbers to tell you which pins of the counters should be connected to which pins of the numeric display. Otherwise, the schematic would be a confusing tangle of wires crossing each other.

At this point, you can give up in dismay at the number of connections—but really, using a breadboard, it shouldn't take you more than half an hour to complete this phase of the project. I suggest you give it a try, because there's

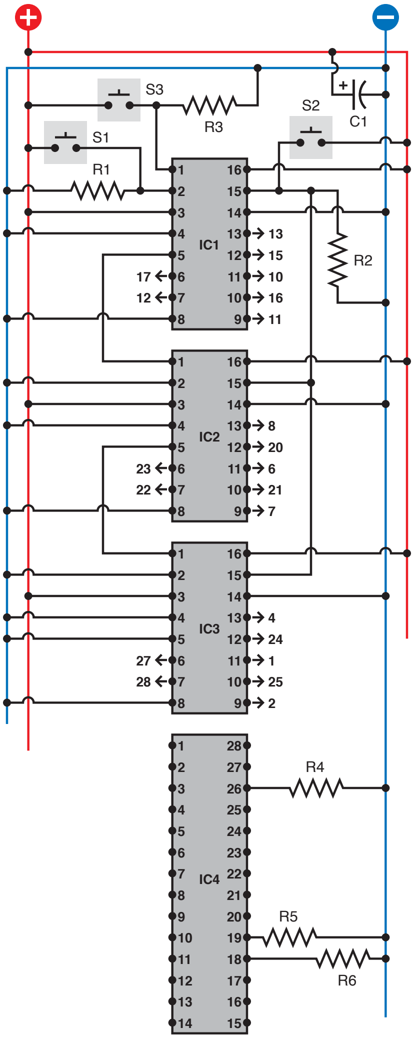

Figure 4-37. *This test circuit, laid out as you would be likely to place it on a breadboard, allows you to trigger a counter manually to verify that the display increments from 000 upward to 999.*

Component values:

All resistors are 1K.
S1, S2, S3: SPST tactile switches, normally open
IC1, IC2, IC3: 4026 decade counter chips
IC4: Kingbright 3-digit common-cathode display
C1: 100 µF (minimum) smoothing capacitor

Wire the output pins on IC1, IC2, and IC3 to the pins on IC4, according to the numbers preceded by arrows. The actual wires have been omitted for clarity. Check for the pinouts of IC4.

something magical about seeing a display count from 000 through 999 "all by itself," and I chose this project because it also has a lot of instructional value.

S1 is attached to the "clock disable" pin of IC1, so that when you hold down this button, it should stop that counter from counting. Because IC1 controls IC2, and IC2 controls IC3, if you freeze IC1, the other two will have to wait for it to resume. Therefore you won't need to make use of their "clock disable" features.

S2 is connected to the "reset" pins of all three counters, so that when you hold down this button, it should set them all to zero.

S3 sends positive pulses manually to the "clock input" pin of the first counter.

S1, S2, and S3 are all wired in parallel with 1K resistors connected to the negative side of the power supply. The idea is that when the buttons are not being pressed, the "pull-down" resistors keep the pins near ground (zero) voltage. When you press one of the buttons, it connects positive voltage directly to the chip, and easily overwhelms the negative voltage. This way, the pins remain either in a definitely positive or definitely negative state. If you disconnect one of these pull-down resistors you are likely to see the numeric display "flutter" erratically. (The numeric display chip has some unconnected pins, but this won't cause any problem, because it is a passive chip that is just a collection of LED segments.)

Always connect input pins of a CMOS chip so that they are either positive or negative. See the "No Floating Pins" warning on the next page.

I suggest that you connect all the wires shown in the schematic first. Then cut lengths of 22-gauge wire to join the remaining pins of the sockets from IC1, IC2, and IC3 to IC4.

Switch on the power and press S2. You'll see three zeros in your numeric display.

Each time you press S3, the count should advance by 1. If you press S2, the count should reset to three zeros. If you hold down S1 while you press S3 repeatedly, the counters should remain frozen, ignoring the pulses from S3.

FUNDAMENTALS

Switch bounce

When you hit S3, I think you'll find that the count sometimes increases by more than 1. This does not mean that there's something wrong with your circuit or your components; you are just observing a phenomenon known as "switch bounce."

On a microscopic level, the contacts inside a pushbutton switch do not close smoothly, firmly, and decisively. They vibrate for a few microseconds before settling; the counter chip detects this vibration as a series of pulses, not just one.

Various circuits are available to "debounce" a switch. The simplest option is to put a small capacitor in parallel with the switch, to absorb the fluctuations; but this is less than ideal. I'll come back to the topic of debouncing later in the book. Switch bounce is not a concern in this circuit, because we're about to get rid of S3 and substitute a 555 timer that generates nice clean bounceless pulses.

Pulse Generation

A 555 timer is ideal for driving a counter chip. You've already seen how to wire a 555 to create a stream of pulses that made noise through a loudspeaker. I'm reproducing the same circuit in Figure 4-38 in simplified form, using the positive and negative supply configuration in the current project. Also I'm showing the connection between pins 2 and 6 in the way that you're most likely to make it, via a wire that loops over the top of the chip.

For the current experiment, I'm suggesting initial component values that will generate only four pulses per second. Any faster than that, and you won't be able to verify that your counters are counting properly.

Install IC5 and its associated components on your breadboard immediately above IC1. Don't leave any gap between the chips. Disconnect S3 and R3 and connect a wire directly between pin 3 (output) of IC5 and pin 1 (clock) of IC1, the topmost counter. Power up again, and you should see the digits advancing rapidly in a smooth, regular fashion. Press S1, and while you hold it, the count should freeze. Release S1 and the count will resume. Press S2 and the counter should reset, even if you are pressing S1 at the same time.

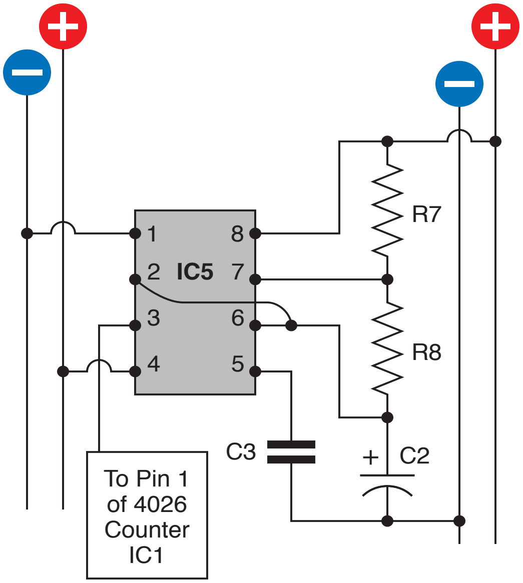

Figure 4-38. *A basic astable circuit to drive the decade counter in the previous schematic. Output is approximately 4 pulses per second.*

R7: 1K
R8: 2K2
C2: 68 µF
C3: 0.1 µF
IC5: 555 timer

No Floating Pins!

A CMOS chip is hypersensitive. Any pin that is not wired either to the supply voltage or to ground is said to be "floating" and may act like an antenna, sensitive to the smallest fluctuations in the world around it.

The 4026 counter chip has a pin labeled "clock disable." The manufacturer's data sheet helpfully tells you that if you give this pin a positive voltage, the chip stops counting and freezes its display. As you don't want to do that, you may just ignore that pin and leave it unconnected, at least while you test the chip. This is a very bad idea!

What the data sheet doesn't bother to tell you (presumably because "everyone knows" such things) is that if you want the clock to run normally, the clock-disable feature itself has to be disabled, by wiring it to negative (ground) voltage. If you leave the pin floating (and I speak from experience), the chip will behave erratically and uselessly.

All input pins must be either positively or negatively wired, unless otherwise specified.

Refinements

Now it's time to remember that what we really want this circuit to do is test a person's reflexes. When the user starts it, we want an initial delay, followed by a signal—probably an LED that comes on. The user responds to the signal by pressing a button as quickly as possible. During the time it takes for the person to respond, the counter will count milliseconds. When the person presses the button, the counter will stop. The display then remains frozen indefinitely, displaying the number of pulses that were counted before the person was able to react.

How to arrange this? I think we need a flip-flop. When the flip-flop gets a signal, it starts the counter running—and keeps it running. When the flip-flop gets another signal (from the user pressing a button), it stops the counter running, and keeps it stopped.

How do we build this flip-flop? Believe it or not, we can use yet another 555 timer, in a new manner known as bistable mode.

FUNDAMENTALS

The bistable 555 timer

Figure 4-39 shows the internal layout of a 555 timer, as before, but the external components on the righthand side have been eliminated. Instead, I'm applying a constant negative voltage to pin 6. Can you see the consequences? Suppose you apply a negative pulse to the trigger (pin 2). Normally when you do this and the 555 starts running, it generates a positive output while charging a capacitor attached to pin 6. When the capacitor reaches 2/3 of the full supply voltage, this tells the 555 to ends its positive output, and it flips back to negative.

Well, if there's no capacitor, there's nothing to stop the timer. Its positive output will just continue indefinitely. However, pin 4 (the reset) can still override everything, so if you apply negative voltage to pin 4, it flips the output to negative. After that, the output will stay negative indefinitely, as it usually does, until you trigger the timer by dropping the voltage to pin 2 again. This will flip the timer back to generating its positive output.

Here's a quick summary of the bistable configuration:

- A negative pulse to pin 2 turns the output positive.
- A negative pulse to pin 4 turns the output negative.
- The timer is stable in each of these states. Its run-time has become infinite.

It's OK to leave pins 5 and 7 of the timer unconnected, because we're pushing it into extreme states where any random signals from those pins will be ignored.

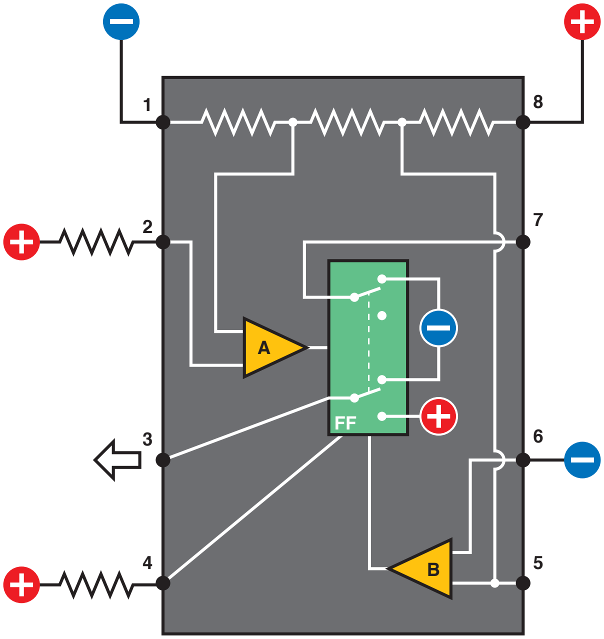

Figure 4-39. *In the bistable configuration, pin 6 of a 555 timer is perpetually negative, so the timer cycle never ends, unless you force it to do so by applying a negative pulse to pin 4 (the reset).*

In bistable mode, the 555 has turned into one big flip-flop. To avoid any uncertainty, we keep pins 2 and 4 normally positive via pull-up resistors, but negative pulses on those pins can overwhelm them when we want to flip the 555 into its opposite state. The schematic for running a 555 timer in bistable mode, controlled by two pushbuttons, is shown in Figure 4-40. You can add this above your existing circuit. Because you're going to attach the output from IC6 to pin 2 of IC1, the topmost counter, you can disconnect S1 and R1 from that pin. See Figure 4-41.

Now, power up the circuit again. You should find that it counts in the same way as before, but when you press S4, it freezes. This is because your bistable 555 timer is sending its positive output to the "clock disable" pin on the counter. The counter is still receiving a stream of pulses from the astable 555 timer, but as long as pin 2 is positive on the counter, the counter simply ignores the pulses.

Now press S5, which flips your bistable 555 back to delivering a negative output, at which point the count resumes.

We're getting close to a final working circuit here. We can reset the count to zero (with S3), start the count (with S5), and wait for the user to stop the count (with S4). The only thing missing is a way to start the count unexpectedly.

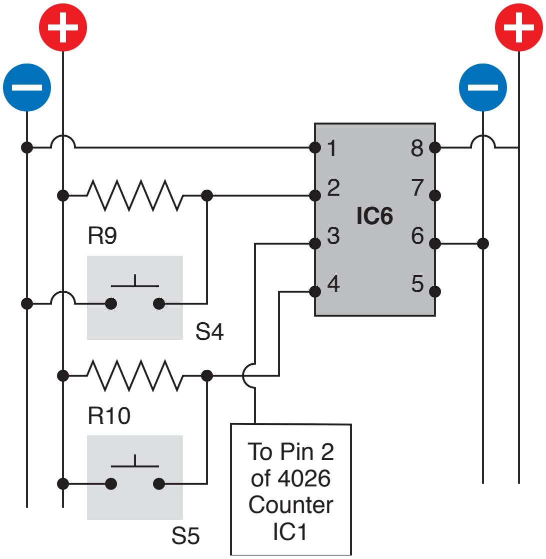

Figure 4-40. *Adding a bistable 555 timer to the reflex tester will stop the counter with a touch of a button, and keep it stopped.*

R9, R10: 1K
IC6: 555 timer

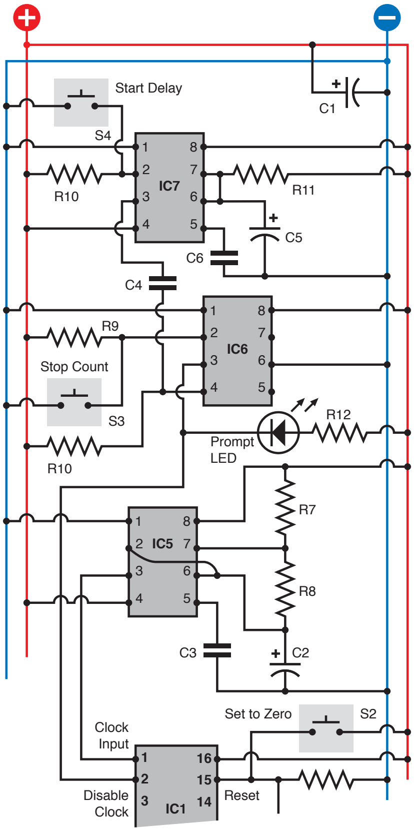

Figure 4-41. *The completed control section of the circuit, to be added above these timers.*

R7, R9, R10, R12: 1K
R8: 2K2
R11: 330K
C1: 100 µF
C2: 68 µF
C3, C4, C6: 0.1 µF
C5: 10 µF
S1, S2, S3: tactile switches
IC5, IC6, IC7: 555 timers

The Delay

Suppose we set up yet another 555 in monostable mode. Trigger its pin 2 with a negative pulse, and the timer delivers a positive output that lasts for, say, 4 seconds. At the end of that time, its output goes back to being negative. Maybe we can hook that positive-to-negative transition to pin 4 of IC6. We can use this instead of switch S5, which you were pressing previously to start the count.

Check the new schematic in Figure 4-41 which adds another 555 timer, IC7 above IC6. When the output from IC7 goes from positive to negative, it will trigger the reset of IC6, flipping its output negative, which allows the count to begin. So IC7 has taken the place of the start switch, S4. You can get rid of S4, but keep the pull-up resistor, R9, so that the reset of IC6 remains positive the rest of the time.

This arrangement works because I have used a capacitor, C4, to connect the output of IC7 to the reset of IC6. The capacitor communicates the sudden change from positive to negative, but the rest of the time it blocks the steady voltage from IC7 so that it won't interfere with IC6.

The final schematic in Figure 4-41 shows the three 555 timers all linked together, as you should insert them above the topmost counter, IC1. I also added an LED to signal the user. Figure 4-42 is a photograph of my working model of the circuit.

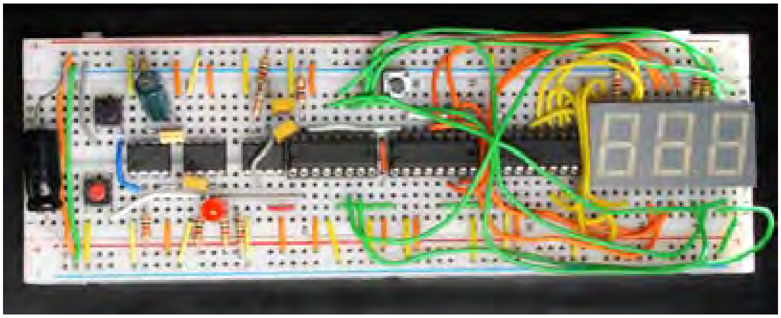

Figure 4-42. *The complete reaction-timer circuit barely fits on a 63-row breadboard.*

Because this circuit is complicated, I'll summarize the sequence of events when it's working. Refer to Figure 4-41 while following these steps:

1. User presses Start Delay button S4, which triggers IC7.
2. IC7 output goes high for a few seconds while C5 charges.
3. IC7 output drops back low.
4. IC7 communicates a pulse of low voltage through C4 to IC6, pin 4.
5. IC 6 output flips to low and flops there.
6. Low output from IC6 sinks current through LED and lights it.
7. Low output from IC6 also goes to pin 2 of IC1.

8. Low voltage on pin 2 of IC1 allows IC1 to start counting.
9. User presses S3, the "stop" button.
10. S3 connects pin 2 of IC6 to ground.
11. IC6 output flips to high and flops there.
12. High output from IC6 turns off the LED.
13. High output from IC6 also goes to pin 2 of IC1.
14. High voltage on pin 2 of IC1 stops it from counting.
15. After assessing the result, user presses S2.
16. S2 applies positive voltage to pin 15 of IC1, IC2, IC3.
17. Positive voltage resets counters to zero.
18. The user can now try again.
19. Meanwhile, IC5 is running continuously throughout.

In case you find a block diagram easier to understand, I've included that, too, in Figure 4-43.

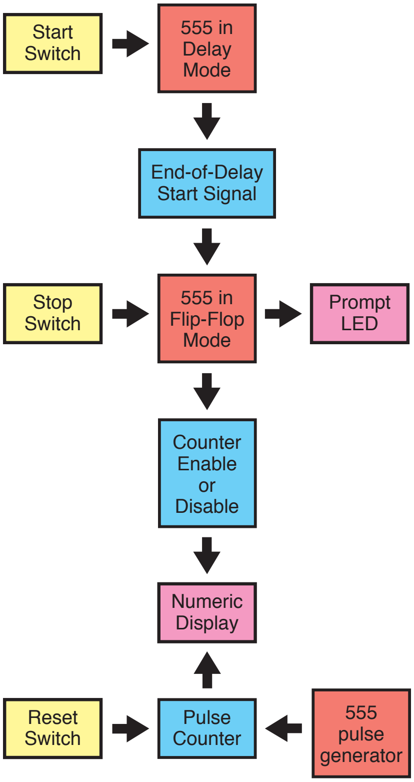

Figure 4-43. *The functions of the reflex tester, summarized as a block diagram.*

Using the Reflex Tester

At this point, you should be able to test the circuit fully. When you first switch it on, it will start counting, which is slightly annoying, but easily fixed. Press S3 to stop the count. Press S2 to reset to zero.

Now press S4. Nothing seems to happen—but that's the whole idea. The delay cycle has begun in stealth mode. After a few seconds, the delay cycle ends, and the LED lights up. Simultaneously, the count begins. As quickly as possible, the user presses S3 to stop the count. The numerals freeze, showing how much time elapsed.

There's only one problem—the system has not yet been calibrated. It is still running in slow-motion mode. You need to change the resistor and capacitor attached to IC5 to make it generate 1,000 pulses per second instead of just three or four.

Substitute a 10K trimmer potentiometer for R8 and a 1 μF capacitor for C2. This combination will generate about 690 pulses per second when the trimmer is presenting maximum resistance. When you turn the trimmer down to decrease its resistance, somewhere around its halfway mark the timer will be running at 1,000 pulses per second.

How will you know exactly where this point is? Ideally, you would attach an oscilloscope probe to the output from IC5. But, most likely, you don't have an oscilloscope, so here are a couple other suggestions.

First remove the 1 μF capacitor at C2 and substitute a 10 μF capacitor. Because you are multiplying the capacitance by 10, you will reduce the speed by 10. The leftmost digit in your display should now count in seconds, reaching 9 and rolling over to 0 every 10 seconds. You can adjust your trimmer potentiometer while timing the display with a stopwatch. When you have it right, remove the 10 μF capacitor and replace the 1 μF capacitor at C2.

The only problem is, the values of capacitors may be off by as much as 10%. If you want to fine-tune your reflex timer, you can proceed as follows.

Disconnect the wire from pin 5 of IC3, and substitute an LED with a 1K series resistor between pin 5 and ground. Pin 5 is the "carry" pin, which will emit a positive pulse whenever IC3 counts up to 9 and rolls over to start at 0 again. Because IC3 is counting tenths of a second, you want its carry output to occur once per second.

Now run the circuit for a full minute, using your stopwatch to see if the flashing LED drifts gradually faster or slower than once per second. If you have a camcorder that has a time display in its viewfinder, you can use that to observe the LED.

If the LED flashes too briefly to be easily visible, you can run a wire from pin 5 to another 555 timer that is set up in monostable mode to create an output lasting for around 1/10 of a second. The output from that timer can drive an LED.

Enhancements

It goes without saying that anytime you finish a project, you see some opportunities to improve it. Here are some suggestions:

1. No counting at power-up. It would be nice if the circuit begins in its "ready" state, rather than already counting. To achieve this you need to send a negative pulse to pin 2 of IC6, and maybe a positive pulse to pin 15 of IC1. Maybe an extra 555 timer could do this. I'm going to leave you to experiment with it.
2. Audible feedback when pressing the Start button. Currently, there's no confirmation that the Start button has done anything. All you need to do is buy a piezoelectric beeper and wire it between the righthand side of the Start button and the positive side of the power supply.
3. A random delay interval before the count begins. Making electronic components behave randomly is very difficult, but one way to do it would be to require the user to hold his finger on a couple of metal contacts. The skin resistance of the finger would substitute for R11. Because the finger pressure would not be exactly the same each time, the delay would vary. You'd have to adjust the value of C5.

Summing Up

This project demonstrated how a counter chip can be controlled, how counter chips can be chained together, and three different functions for 555 timers. It also showed you how chips can communicate with each other, and introduced you to the business of calibrating a circuit after you've finished building it.

Naturally, if you want to get some practical use from the circuit, you should build it into an enclosure with heavier-duty pushbuttons—especially the button that stops the count. You'll find that when people's reflexes are being tested, they are liable to hit the stop button quite hard.

Because this was a major project, I'll follow it up here with some quicker, easier ones as we move into the fascinating world of another kind of integrated circuit: logic chips.

Experiment 19: Learning Logic

You will need:

- Assorted resistors and capacitors.
- 74HC00 quad 2-input NAND chip, 74HC08 quad 2-input AND chip, and LM7805 voltage regulator. Quantity: 1 of each.
- Signal diode, 1N4148 or similar. Quantity: 1.
- Low-current LED. Quantity: 1.
- SPST tactile switches. Quantity: 2.

You're going to be entering the realm of pure digital electronics, using "logic gates" that are fundamental in every electronic computing device. When you deal with them individually, they're extremely easy to understand. When you start chaining them together, they can seem intimidatingly complex. So let's start with them one at a time.

Logic gates are much fussier than the 555 timer or the 4026 counter that you used previously. They demand an absolutely precise 5 volts DC, with no fluctuations or "spikes" in the flow of current. Fortunately, this is easy to achieve: just set up your breadboard with an LM7805 voltage regulator, as shown in the schematic in Figure 4-44 and the photograph in Figure 4-45. The regulator receives 9 volts from your usual voltage supply, and reduces it to 5 volts, with the help of a couple of capacitors. You apply the 9 volts to the regulator, and distribute the 5 volts down the sides of your breadboard instead of the unregulated voltage that you used previously. Use your meter to verify the voltage, and make sure you have the polarity clearly marked.

After installing your regulator, take a couple of tactile switches, two 10K resistors, a low-current LED, and a 1K resistor, and set them around a 74HC00 logic chip as shown in Figure 4-46. You may notice that many of the pins of the chip are shorted together and connected to the negative side of the power supply. I'll explain that in a moment.

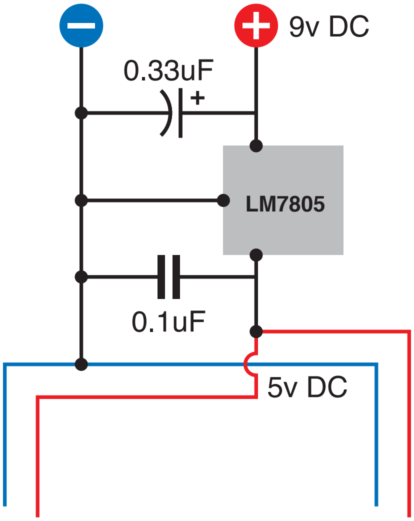

Figure 4-44. *This simple circuit is essential to provide a regulated 5V DC supply for logic chips.*

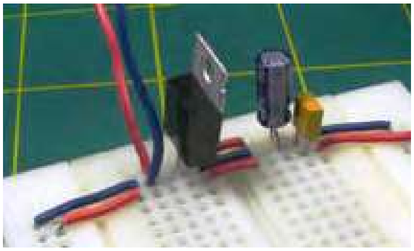

Figure 4-45. *The voltage regulator and its two capacitors can fit snugly at the top of a breadboard. Remember to apply the 9V input voltage at the left pin of the regulator, and distribute the 5V output down the sides of the breadboard.*

FUNDAMENTALS

Voltage regulators

The simplest versions of these little semiconductors accept a higher DC voltage on one pin and deliver a lower DC voltage on another pin, with a third pin (usually in the middle) serving as a common negative, or ground. You should also attach a couple of capacitors to smooth the current, as shown in Figure 4-46.

Typically you can put a 7.5-volt or 9-volt supply on the "input" side of a 5-volt regular, and draw a precise 5 volts from the "output" side. If you're wondering where the extra voltage goes, the answer is, the regulator turns the electricity into heat. For this reason, small regulators (such as the one in Figure 4-8) often have a metal back with a hole in the top. Its purpose is to radiate heat, which it will do more effectively if you bolt it to a piece of aluminum, since aluminum conducts heat very effectively. The aluminum is known as a heat sink, and you can buy fancy ones that have multiple cooling fins.

For our purposes, we won't be drawing enough current to require a heat sink.

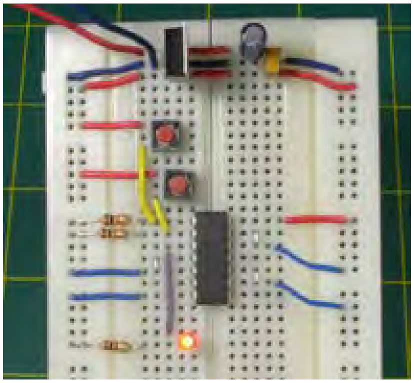

Figure 4-47. *This breadboard layout is exactly equivalent to the schematic in Figure 4-46.*

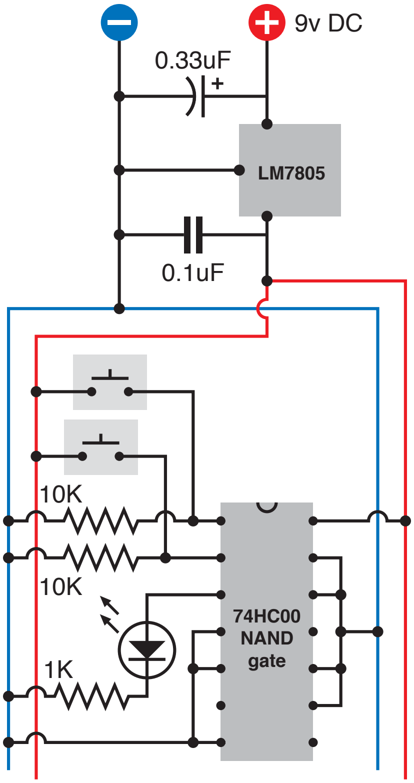

Figure 4-46. *By observing the LED when you press either, both, or neither of the buttons, you can easily figure out the logical function of the NAND gate.*

When you connect power, the LED should light up. Press one of the tactile switches, and the LED remains illuminated. Press the other tactile switch, and again the LED stays on. Now press both switches, and the light should go out.

Pins 1 and 2 are logic inputs for the 74HC00 chip. Initially they were held at negative voltage, being connected to the negative side of the power supply through 10K resistors. But each pushbutton overrides its pull-down resistor and forces the input pin to go positive.

The logic output from the chip, as you saw, is normally positive—but *not* if the first input *and* the second input are positive. Because the chip does a "Not AND" operation, it's known as a NAND logic gate. You can see the breadboard layout in Figure 4-47. Figure 4-48 is a simplified version of the circuit. The U-shaped thing with a circle at the bottom is the logic symbol for a NAND gate. No power supply is shown for it, but in fact all logic chips require a power supply, which enables them to put out more current than they take in. Anytime you see a symbol for a logic chip, try to remember that it has to have power to function.

The 74HC00 actually contains four NAND gates, each with two logical inputs and one output. They are arrayed as shown in Figure 4-49. Because only one gate was needed for the simple test, the input pins of the unused gates were shorted to the negative side of the power supply.

Pin 14 supplies positive power for the chip; pin 7 is its ground pin. Almost all the 7400 family of logic chips use the same pins for positive and negative power, so you can swap them easily.

In fact, let's do that right now. First, disconnect the power. Carefully pull out the 74HC00 and put it away with its legs embedded in conductive foam. Substitute a 74HC08 chip, which is an AND chip. Make sure you have it the right way up, with its notch at the top. Reconnect the power and use the pushbuttons as you did before. This time, you should find that the LED comes on if the first input AND the second input are both positive, but it remains dark otherwise. Thus, the AND chip functions exactly opposite to the NAND chip. Its pinouts are shown in Figure 4-50.

You may be wondering why these things are useful. Soon you'll see that we can put logic gates together to do things such as create an electronic combination lock, or a pair of electronic dice, or a computerized version of a TV quiz show where users compete to answer a question. And if you were really insanely ambitious, you could build an entire computer out of logic gates.

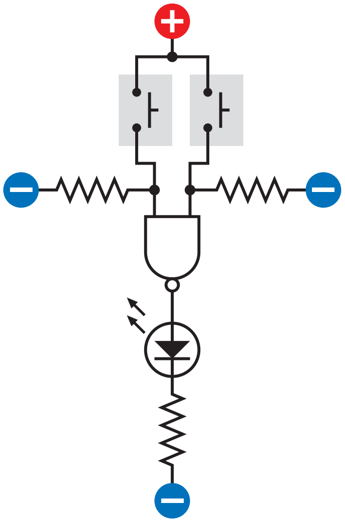

Figure 4-48. *The structure and function of the NAND gate is easier to visualize with this simplified schematic that omits the power supply for the chip and doesn't attempt to place the wires to fit a breadboard layout.*

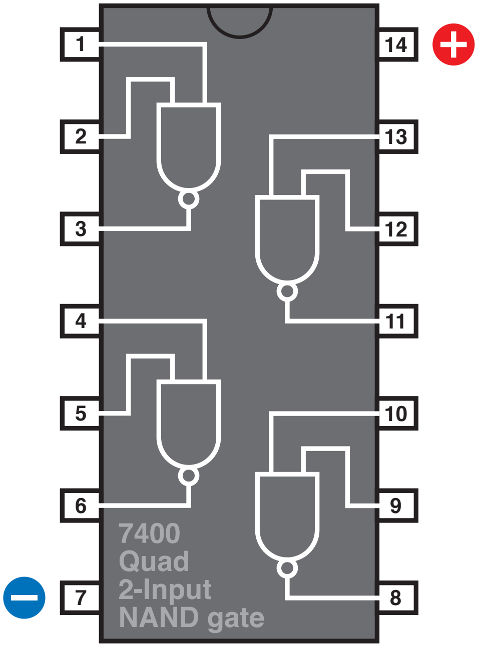

Figure 4-49. *The pinouts of the logic gates in a 74HC00chip.*

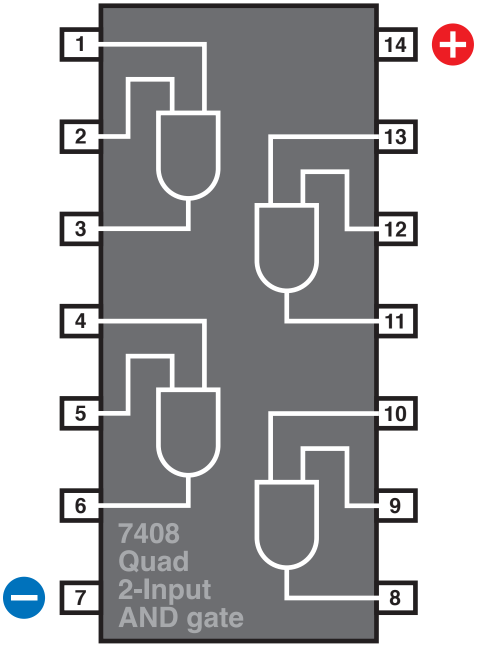

Figure 4-50. *The pinouts of the logic gates in a 74HC08chip.*

BACKGROUND

From Boole to Shannon

George Boole was a British mathematician, born in 1815, who did something that few people are ever lucky enough or smart enough to do: he invented an entirely new branch of mathematics.

Interestingly, it was not based on numbers. Boole had a relentlessly logical mind, and he wanted to reduce the world to a series of true-or-false statements which could overlap in interesting ways. For instance, suppose there is a couple named Ann and Bob who have so little money, they only own one hat. Clearly, if you happen to run into Ann and Bob walking down the street, there are four possibilities: neither of them may be wearing a hat, Ann may be wearing it, or Bob may be wearing it, but they cannot both be wearing it.

The diagram in Figure 4-51 illustrates this. All the states are possible except the one where the circles overlap. (This is known as a Venn diagram. I leave it to you to search for this term if it interests you and you'd like to learn more.) Boole took this concept much further, and showed how to create and simplify extremely complex arrays of logic.

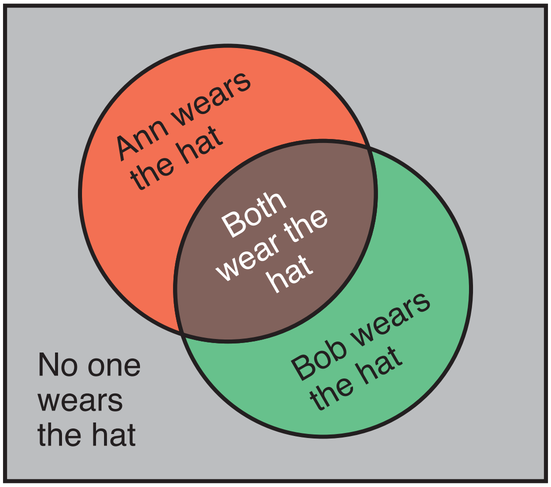

Figure 4-51. *This slightly frivolous Venn diagram illustrates the various possibilities for two people, Ann and Bob, who own only one hat.*

Another way to summarize the hat-wearing situation is to make the "truth table" shown in Figure 4-52. The rightmost column shows whether each combination of propositions can be true. Now check the table in Figure 4-53. It's the same table but uses different labels, which describe the pattern you have seen while using the NAND gate.

Boole published his treatise on logic in 1854, long before it could be applied to electrical or electronic devices. In fact, during his lifetime, his work seemed to have no practical applications at all. But a man named Claude Shannon encountered Boolean logic while studying at MIT in the 1930s, and in 1938 he published a paper describing how Boolean analysis could be applied to circuits using relays. This had immediate practical applications, as telephone networks were growing rapidly, creating complicated switching problems.

BACKGROUND

From Boole to Shannon (continued)

Ann is wearing the hat	Bob is wearing the hat	This combination can be
NO	NO	TRUE
NO	YES	TRUE
YES	NO	TRUE
YES	YES	FALSE

Figure 4-52. *The hat-wearing possibilities can be expressed in a "truth table."*

Input A	Input B	Output
OFF	OFF	ON
OFF	ON	ON
ON	OFF	ON
ON	ON	OFF

Figure 4-53. *The truth-table from can be relabeled to describe the inputs and outputs of a NAND gate.*

A very simple telephone problem could be expressed like this. Suppose two customers in a rural area share one telephone line. If one of them wants to use the line, or the other wants to use it, or neither of them wants to use it, there's no problem. But they cannot both use it at once. You may notice that this is exactly the same as the hat-wearing situation for Ann and Bob.

We can easily draw a circuit using two normally closed relays that creates the desired outcome (see Figure 4-54), but if you imagine a telephone exchange serving many thousands of customers, the situation becomes very complicated indeed. In fact, in Shannon's time, no logical process existed to find the best solution and verify that it used fewer components than some other solution.

Shannon saw that Boolean analysis could be used for this purpose. Also, if you used an "on" condition to represent numeral 1 and an "off" condition to represent numeral 0, you could build a system of relays that could count. And if it could count, it could do arithmetic.

When vacuum tubes were substituted for relays, the first practical digital computers were built. Transistors took the place of vacuum tubes, and integrated circuit chips replaced transistors, leading to the desktop computers that we now take for granted today. But deep down, at the lowest levels of these incredibly complex devices, they still use the laws of logic discovered by George Boole. Today, when you use a search engine online, if you use the words AND and OR to refine your search, you're actually using Boolean operators.

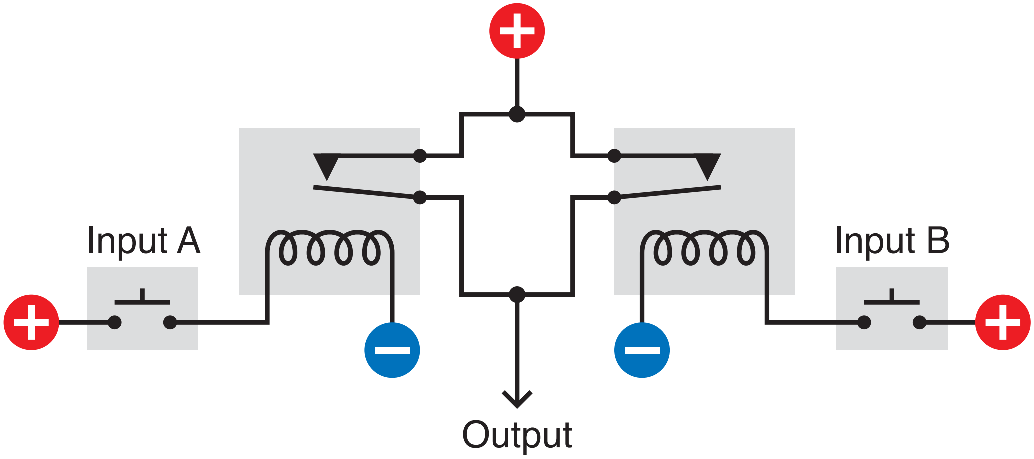

Figure 4-54. *This relay circuit could illustrate the desired logic for two telephone customers wanting to share one line, and its behavior is almost identical to that of the NAND schematic shown in Figure 4-48.*

Figure 4-55. *Ann and Bob attempt to overcome the limitations of Boolean logic.*

ESSENTIALS

Logic gate basics

The NAND gate is the most fundamental building block of digital computers, because (for reasons which I don't have space to explain here) it enables digital addition. If you want to explore more try searching online for topics such as "binary arithmetic" and "half-adder."

Generally, there are seven types of logic gates:

- AND
- NAND
- OR
- NOR
- XOR
- XNOR
- NOT

Of the six two-input gates, the XNOR is hardly ever used. The NOT gate has a single input, and simply gives a negative output when the input is positive or a positive output when the input is negative. The NOT is more often referred to as an "Inverter." The symbols for all seven gates are shown in Figure 4-56.

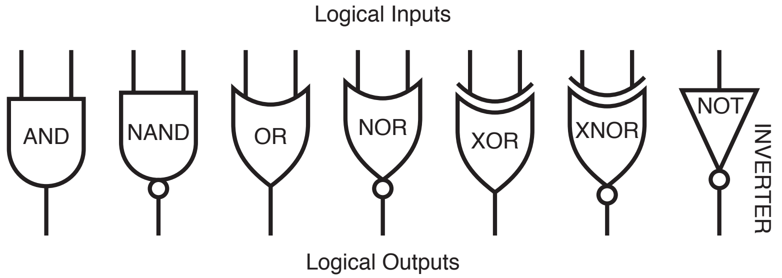

Figure 4-56. *American symbols for the six types of two-input logic gates, and the single-input inverter.*

I've shown the American symbols. Other symbols have been adopted in Europe, but the traditional symbols shown here are the ones that you will usually find, even being used by Europeans. I also show the truth tables, in Figure 4-57, illustrating the logical output (high or low) for each pair of inputs of each type of gate.

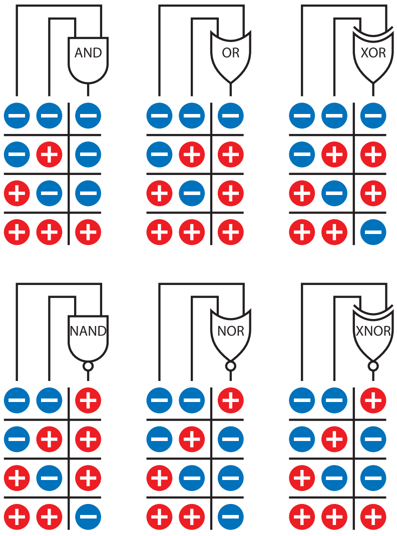

AND		
−	−	−
−	+	−
+	−	−
+	+	+

OR		
−	−	−
−	+	+
+	−	+
+	+	+

XOR		
−	−	−
−	+	+
+	−	+
+	+	−

NAND		
−	−	+
−	+	+
+	−	+
+	+	−

NOR		
−	−	+
−	+	−
+	−	−
+	+	−

XNOR		
−	−	+
−	+	−
+	−	−
+	+	+

Figure 4-57. *Inputs and corresponding outputs for the six types of logic gates (note that the XNOR gate is seldom used). The minus signs indicate low voltage, close to ground potential. The plus signs indicate higher voltage, close to the positive potential of the power supply in the circuit. The exact voltages will vary depending on other components that may be actively connected.*

ESSENTIALS

Logic gate basics (continued)

If you have difficulty visualizing logic gates, a mechanical comparison may help. You can think of them as being like sliding plates with holes in them, in a bubblegum machine. Two people, A and B, can push the plates. These people are the two inputs, which are considered positive when they are doing something. (Negative logic systems also exist, but are uncommon, so I'm only going to talk about positive systems here.)

The flow of bubblegum represents a flow of positive current. The full set of possibilities is shown in Figures 4-58 through 4-63.

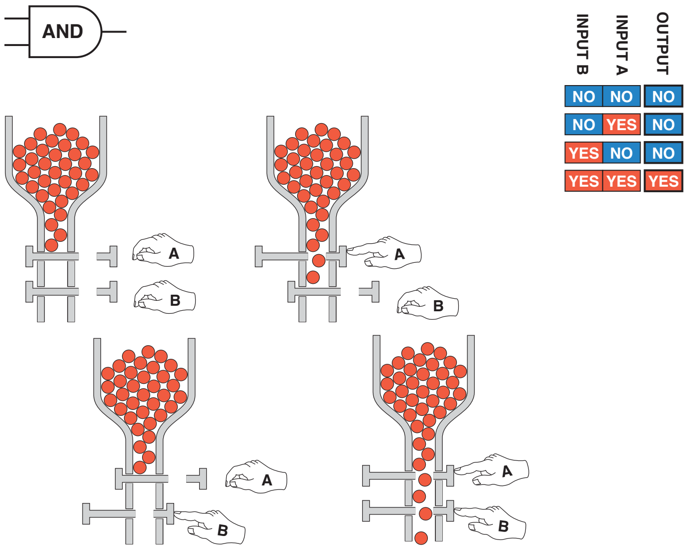

INPUT B	INPUT A	OUTPUT
NO	NO	NO
NO	YES	NO
YES	NO	NO
YES	YES	YES

Figure 4-58

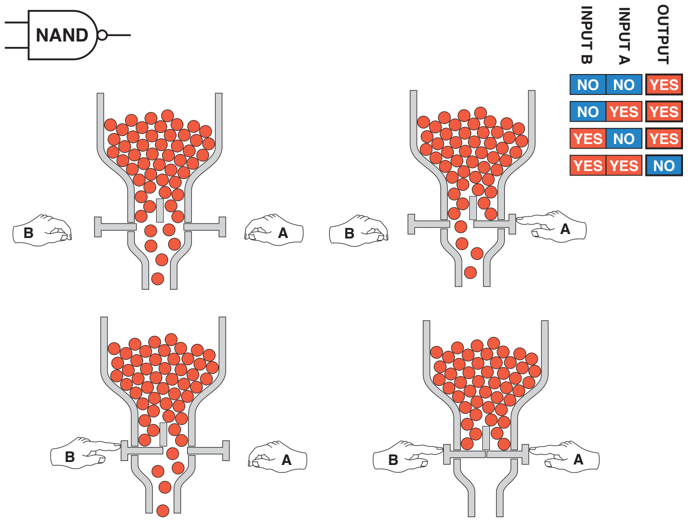

INPUT B	INPUT A	OUTPUT
NO	NO	YES
NO	YES	YES
YES	NO	YES
YES	YES	NO

Figure 4-59

ESSENTIALS

Logic gate basics (continued)

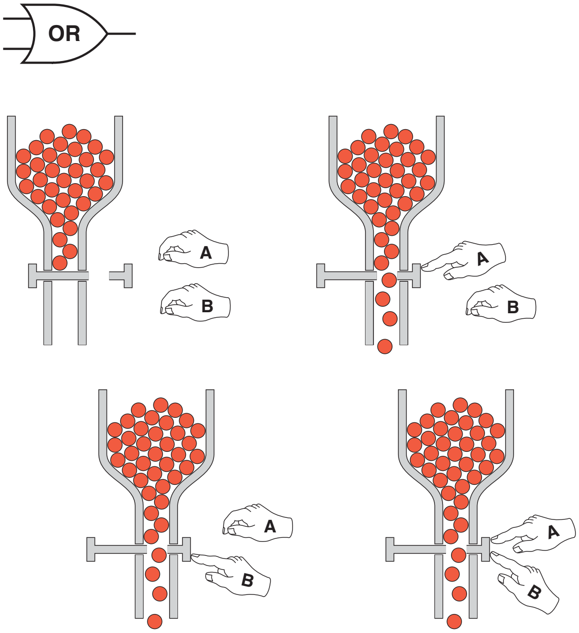
OR
A
B
A
B
A
B
A
B

INPUT B	INPUT A	OUTPUT
NO	NO	NO
NO	YES	YES
YES	NO	YES
YES	YES	YES

Figure 4-80

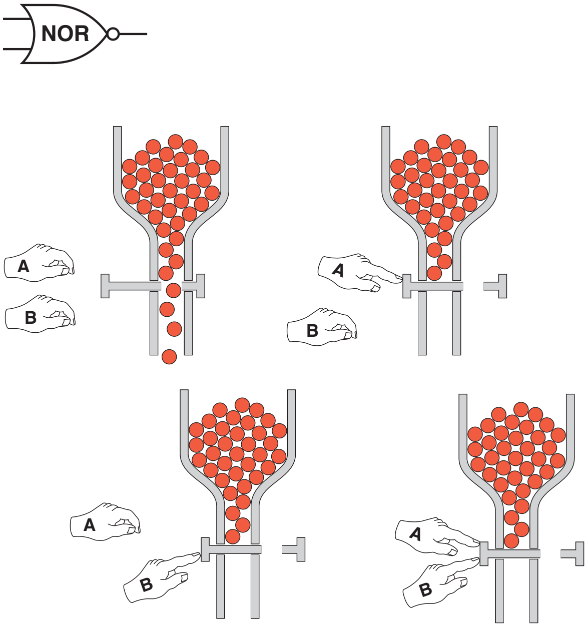
NOR
A
B
A
B
A
B
A
B

INPUT B	INPUT A	OUTPUT
NO	NO	YES
NO	YES	NO
YES	NO	NO
YES	YES	NO

Figure 4-81

ESSENTIALS

Logic gate basics (continued)

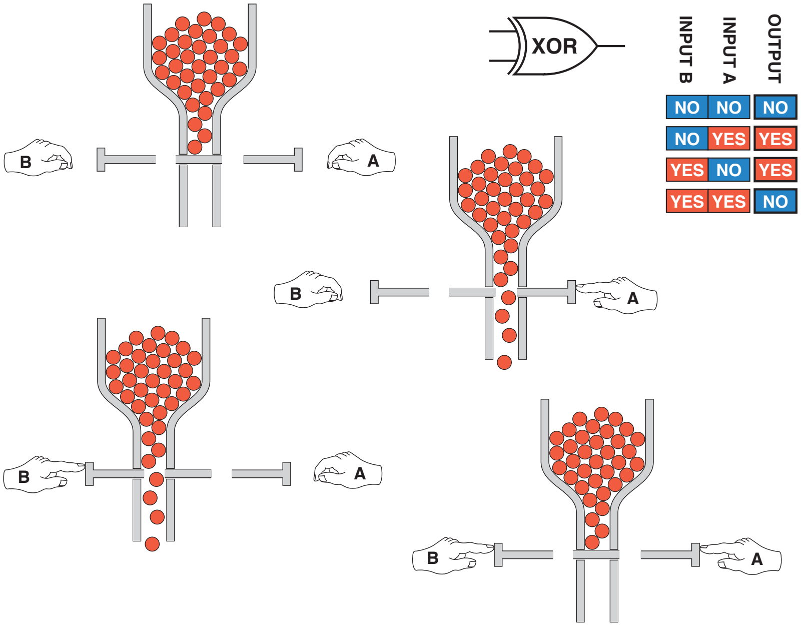

INPUT B	INPUT A	OUTPUT
NO	NO	NO
NO	YES	YES
YES	NO	YES
YES	YES	NO

Figure 4-82

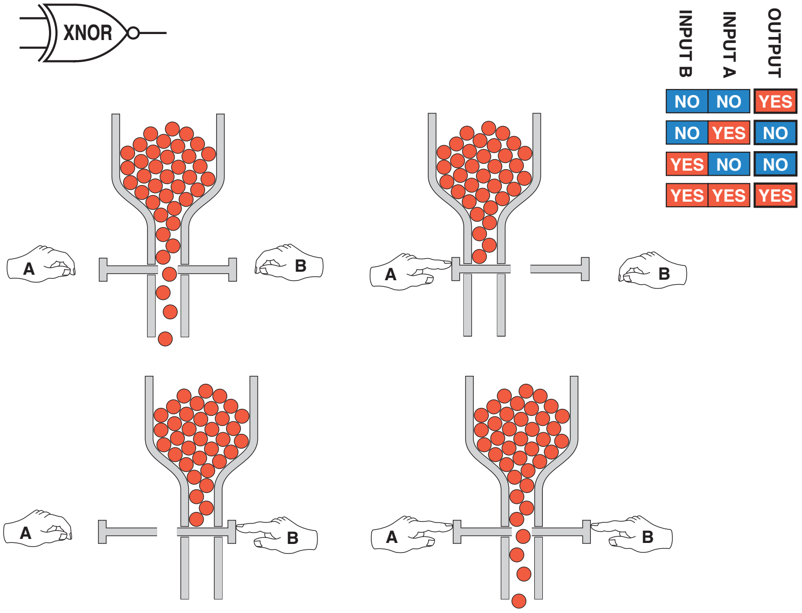

INPUT B	INPUT A	OUTPUT
NO	NO	YES
NO	YES	NO
YES	NO	NO
YES	YES	YES

Figure 4-83

BACKGROUND

The confusing world of TTL and CMOS

Back in the 1960s, the first logic gates were built with Transistor-Transistor Logic, abbreviated TTL, meaning that tiny bipolar transistors were etched into a single wafer of silicon. Soon, these were followed by Complementary Metal Oxide Semiconductors, abbreviated CMOS. Each of these chips was a collection of metal-oxide field-effect transistors (known as MOSFETs). The 4026 chip that you used earlier is an old CMOS.

You may remember that bipolar transistors amplify current. TTL circuits are similar: they are sensitive to current, rather than voltage. Thus they require a significant flow of electricity, to function. But CMOS chips are like the programmable unijunction transistor that I featured previously. They are voltage-sensitive, enabling them to draw hardly any current while they are waiting for a signal, or pausing after emitting a signal.

The two families named TTL and CMOS still exist today. The table in Figure 4-64 summarizes their basic advantages and disadvantages. The CMOS series, with part numbers from 4000 upward, were easily damaged by static electricity but were valuable because of their meager power consumption. The TTL series, with part numbers from 7400 upward, used much more power but were less sensitive and very fast. So, if you wanted to build a computer, you used the TTL family, but if you wanted to build a little gizmo that would run for weeks on a small battery, you used the CMOS family.

From this point on everything became extremely confusing, because CMOS manufacturers wanted to grab market share by emulating the advantages of TTL chips. Newer generations of CMOS chips even changed their part numbers to begin with "74" to emphasize their compatibility, and the functions of pins on CMOS chips were swapped around to match the functions of pins on TTL chips. Consequently, the pinouts of CMOS and TTL chips are usually now identical, but the meaning of "high" and "low" states changed in each new generation, and the maximum supply voltages for CMOS chips were revised downward. Note I have included question marks beside two categories in the CMOS column, as modern CMOS chips have overcome those disadvantages—at least to some extent.

Here's a quick summary, which will be useful to you if you look at a circuit that you find online, and you wonder about the chips that have been specified.

Where you see a letter "x," it means that various numbers may appear in that location. Thus "74xx" includes the 7400 NAND gate, the 7402 NOR gate, the 74150 16-bit data selector, and so on. A combination of letters preceding the "74" identifies the chip manufacturer, while letters following the part number may identify the style of package, may indicate whether it contains heavy metals that are environmentally toxic, and other details.

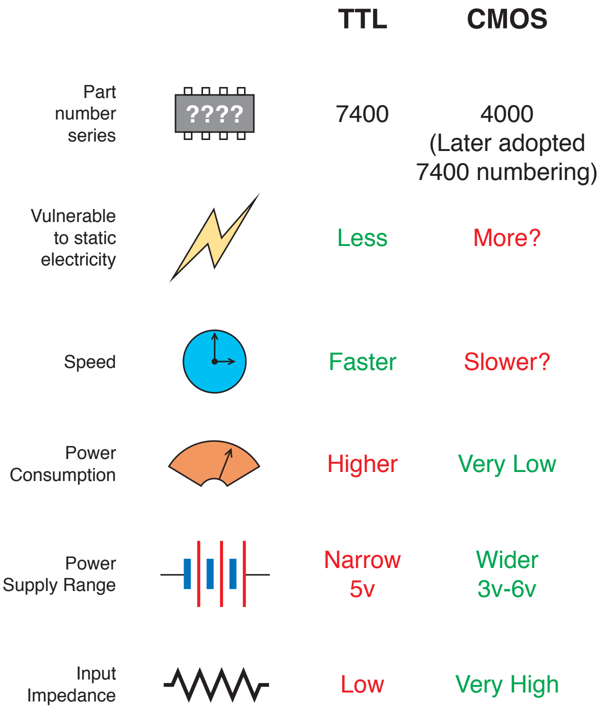

	TTL	CMOS
Part number series	7400	4000 (Later adopted 7400 numbering)
Vulnerable to static electricity	Less	More?
Speed	Faster	Slower?
Power Consumption	Higher	Very Low
Power Supply Range	Narrow 5v	Wider 3v-6v
Input Impedance	Low	Very High

Figure 4-64. *The basic differences between the two families of logic chips. In successive generations, these differences have gradually diminished.*

TTL family:

74xx
The old original generation, now obsolete.

74Sxx
Higher speed "Schottky" series, now obsolete.

74LSxx
Lower power Schottky series, still used occasionally.

74ALSxx
Advanced low-power Schottky.

74Fxx
Faster than the ALS series.

BACKGROUND

The confusing world of TTL and CMOS (continued)

CMOS family:

40xx

The old original generation, now obsolete.

40xxB

The 4000B series was improved but still susceptible to damage from static electricity. Many hobby circuits still use these chips because they will run from relatively high voltages and can power LEDs and even small relays directly.

74HCxx

Higher-speed CMOS, with part numbers matching the TTL family, and pinouts matching the TTL family, but input and output voltages not quite the same as the TTL family. I've used this generation extensively in this book, because it's widely available, and the circuits here have no need for greater speed or power.

74HCTxx

Like the HC series but matching the TTL voltages.

74ACxx

Advanced version of HC series. Faster, with higher output capacity.

74ACTxx

Like the AC series but with the same pin functions and voltages as TTL.

74AHCxx

Advanced higher-speed CMOS.

74AHCTxx

Like the AHC series but with the same pin functions and voltages as TTL.

74LVxx

Lower voltage (3.3v) versions, including LV, LVC, LVT, and ALVC series.

As you can see, these days you have to interpret part numbers very carefully. But which family and generation of chips should you use? Well, that depends! Following are some guidelines.

What you don't need:

1. Speed differences are irrelevant from our point of view, as we're not going to be building circuits running at millions of cycles per second.
2. Price differences are so small as to be inconsequential.
3. Lower-voltage (LV) CMOS chips are not very interesting for our small experimental circuits.
4. Try to avoid mixing different families, and different generations of the same family, in the same circuit. They may not be compatible.
5. Some modern chip varieties may be only available in the surface-mount package format. Because they're so much more difficult to deal with, and their only major advantage is miniaturization, I don't recommend them.
6. In the TTL family, the LS and ALS series cannot handle as much output current as the S series and the F series. You don't need them.

What you should use:

1. The old 74LSxx series of TTL chips was so popular, you'll still find schematics that specify these chips. You should still be able to buy them from sources online, but if not, you can substitute the 74HCTxx chips, which are designed to function identically.
2. The old 4000B series of CMOS chips are still used by hobbyists because their willingness to tolerate high voltages is convenient. While TTL or TTL-compatible chips require a carefully regulated 5 volts, the 4000B chips would handle 15 volts—and also delivered enough power to energize LEDs or even very small relays. Some hobbyists also have a nostalgic affection for the 74Cxx series of chips, which had the same pin connections as the TTL chips but could still tolerate higher voltages and higher output current. The trouble is, some of the 74Cxx chips are almost extinct, and while the 4000B chips are still available, they are considered almost obsolete.

BACKGROUND

The confusing world of TTL and CMOS (continued)

Bottom line: I suggest you use the 4000B chips only if you want to replicate an old circuit, or if a modern equivalent is unavailable (which is why I specified the 4026B chip for the reaction timer—I could not find a modern equivalent that will drive seven-segment numeric displays directly, and I didn't want you to have to deal with more parts than necessary).

If you check online suppliers such as *Mouser.com* you'll find that the HC family is by far the most popular right now. They are all available in through-hole format (to fit your breadboard and perforated board). They have the high input impedance of CMOS (which is useful) and they have the same pin identities as the old 74LSxx series.

Abbreviations

When looking at data sheets, you are likely to encounter some or all of these abbreviations:

- VOH min: Minimum output voltage in high state
- VOL max: Maximum output voltage in low state
- VIH min: Minimum input voltage to be recognized as high
- VIL max: Maximum input voltage recognized as low

BACKGROUND

Logic gate origins

The 7400 family of integrated circuits was introduced by Texas Instruments, beginning with the 7400 NAND gate in 1962. Other companies had sold logic chips previously, but the 7400 series came to dominate the market. The Apollo lunar missions used a computer built with 7400 chips, and they were a mainstay of minicomputers during the 1970s.

RCA introduced its 4000 series of logic chips in 1968, built around CMOS transistors; Texas Instruments had chosen TTL. The CMOS chips used less power, thus generating much less heat and enabling flexible circuit design, as each chip could power many others. CMOS was also tolerant of wide voltage ranges (from 3 to 15 volts) but prohibited switching speeds faster than around 1MHz. TTL was 10 times faster.

Design tweaks gradually eradicated the speed penalty for CMOS, and TTL chips have become relatively rare. Still, some people retain a special nostalgic loyalty to "the logic gates that went to the moon." A hardcore enthusiast named Bill Buzbee has built an entire web server from TTL-type 7400 chips, currently online at *http://magic-1.org*. Figure 4-65 shows just one of the handmade circuit boards that Bill assembled to run his computer.

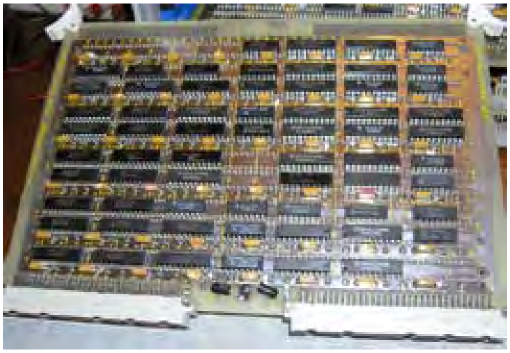

Figure 4-65. *Hobbyist Bill Buzbee built himself a web server entirely from 7400 series logic chips, the oldest of which was fabricated back in 1969. The web server can be found online at http://magic-1.org, displaying pictures of itself and details of its construction. The picture here that Bill took shows just one of the circuit boards of this remarkable project.*

FUNDAMENTALS

Common part numbers

Each 14-pin chip can contain four 2-input gates, three 3-input gates, two 4-input gates, one 8-input gate, or six single-input inverters, as shown in the following table.

	2 input	3 input	4 input	8 input
AND	7408	7411	7421	
NAND	7400	7410	7420	7430
OR	7432			744078*
NOR	7402	7427		744078*
XOR	7486			
XNOR	747266			
Inverter	(1 input) 7404			

*The 744078 has an OR output and a NOR output on the same chip.

Figures 4-66 through 4-74 show the internal connections of the logic chips that you are most likely to use. Note that the 7402 NOR gate has its logical inputs and outputs arranged differently from all the other chips.

Figure 4-66. *Figures 4-66 through 4-74 show pinouts for some of the most widely used logic chips. Note that the inputs of the 7402 are reversed compared with the other chips.*

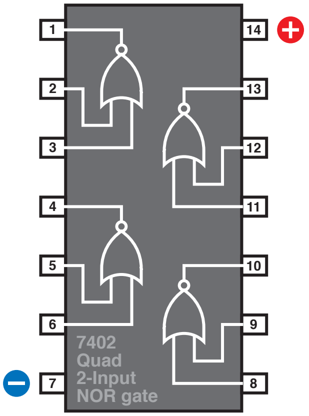
1
2
3
4
5
6
7
14
13
12
11
10
9
8
+
−
7402 Quad 2-Input NOR gate

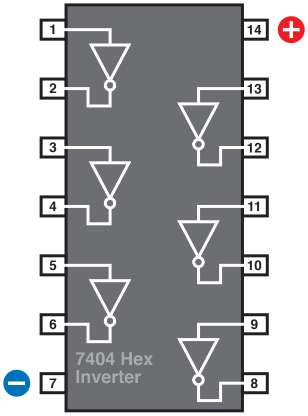
1
2
3
4
5
6
7
14
13
12
11
10
9
8
+
−
7404 Hex Inverter

Figure 4-67

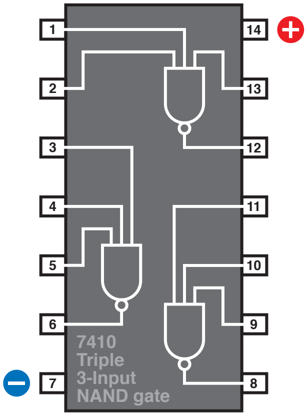
1
2
3
4
5
6
7
14
13
12
11
10
9
8
+
−
7410 Triple 3-Input NAND gate

Figure 4-68

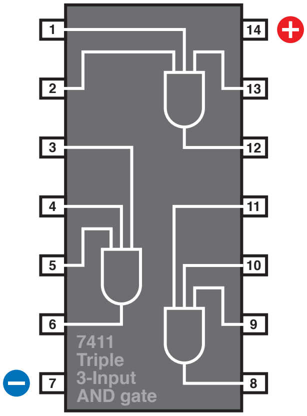
1
2
3
4
5
6
7
14
13
12
11
10
9
8
+
−
7411 Triple 3-Input AND gate

Figure 4-69

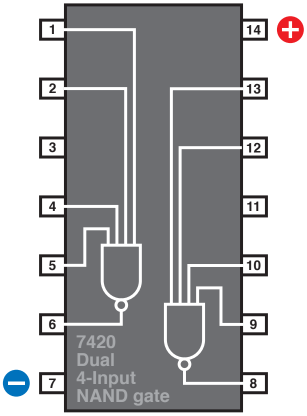
1
2
3
4
5
6
7
14
13
12
11
10
9
8
+
−
7420 Dual 4-Input NAND gate

Figure 4-70

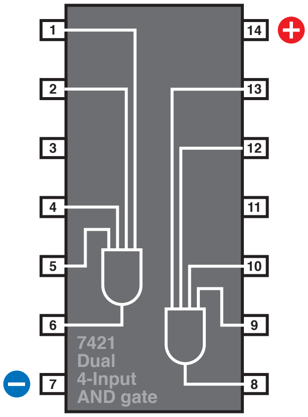
1
2
3
4
5
6
7
14
13
12
11
10
9
8
+
−
7421 Dual 4-Input AND gate

Figure 4-71

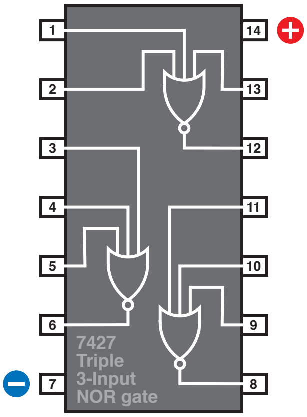
1
2
3
4
5
6
7
14
13
12
11
10
9
8
+
−
7427 Triple 3-Input NOR gate

Figure 4-72

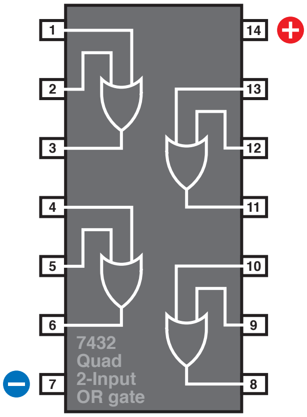
1
2
3
4
5
6
7
14
13
12
11
10
9
8
+
−
7432 Quad 2-Input OR gate

Figure 4-73

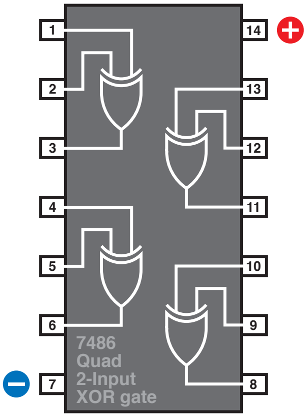
1
2
3
4
5
6
7
14
13
12
11
10
9
8
+
−
7486 Quad 2-Input XOR gate

Figure 4-74

FUNDAMENTALS

Rules for connecting logic gates

Permitted:

- You can connect the input of a gate directly to your regulated power supply, either positive side or negative side.
- You can connect the output from one gate directly to the input of another gate.
- The output from one gate can power the inputs of many other gates (this is known as "fanout"). The exact ratio depends on the chip, but you can always power at least ten inputs with one logic output. The output from a logic chip can drive the trigger (pin 3) of a 555 timer. The output from the timer can then deliver 100mA, easily enough for half-a-dozen LEDs or a small relay.
- Low input doesn't have to be zero. A 74HCxx logic gate will recognize any voltage up to 1 volt as "low."
- High input doesn't have to be 5 volts. A 74HCxx logic gate will recognize any voltage above 3.5 volts as "high."

See Figures 4-75 and 4-76 for a comparison of permitted voltages on the input and output side of 74HCxx and 74LSxx chips.

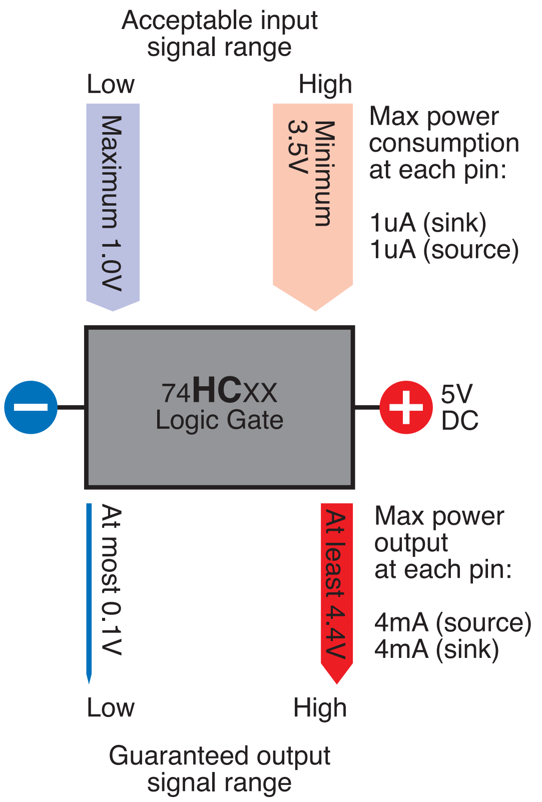
Acceptable input signal range
Low
High
Maximum 1.0V
Minimum 3.5V
Max power consumption at each pin:
1uA (sink)
1uA (source)
74HCxx Logic Gate
5V DC
At most 0.1V
At least 4.4V
Max power output at each pin:
4mA (source)
4mA (sink)
Low
High
Guaranteed output signal range

Figure 4-75. *Each family of logic chips, and each generation in each family, has different standards for input and output minimum and maximum voltages. This diagram shows the standards used by the HC generation of the CMOS family, which was chosen for most of the projects in this book. Note that the current required for input is minimal compared with the current available for output. The power supply to the chip makes up the difference.*

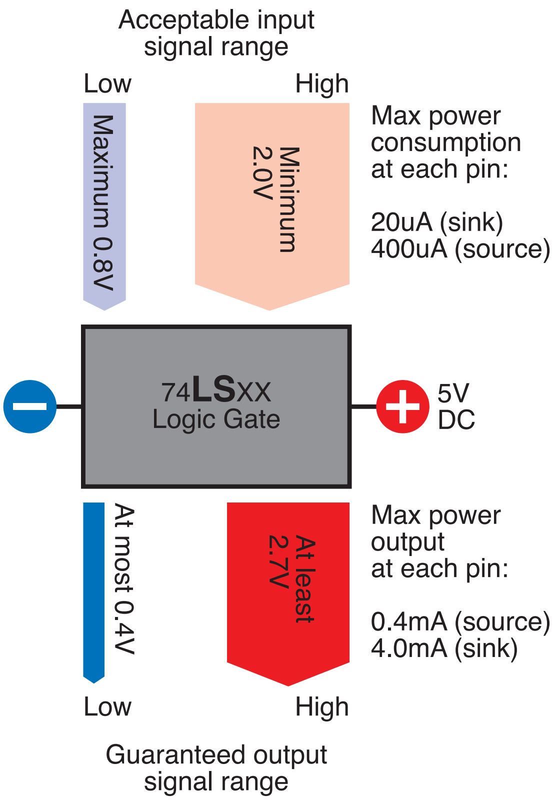
Acceptable input signal range
Low
High
Maximum 0.8V
Minimum 2.0V
Max power consumption at each pin:
20uA (sink)
400uA (source)
74LSxx Logic Gate
5V DC
At most 0.4V
At least 2.7V
Max power output at each pin:
0.4mA (source)
4.0mA (sink)
Low
High
Guaranteed output signal range

Figure 4-76. *Because the LS generation of the TTL family has such different tolerances for input voltages and different standards for output voltages, the LS generation of TTL chips should not be mixed in the same circuit as the HC generation of CMOS chips, unless pull-up resistors are used to bring the LS chips into conformance with standards expected by the HC chips. See Experiment 21 for a case study in using LS chips.*

FUNDAMENTALS

Rules for connecting logic gates (continued)

Not permitted:

- No floating-input pins! On CMOS chips such as the HC family, you must always connect all input pins with a known voltage, even if they supply a gate on the chip that you're not using. When you use a SPST switch to control an input, remember that in its "off" position, it leaves the input unconnected. Use a pull-up or pull-down resistor to prevent this situation. See Figure 4-77.
- Don't use an unregulated power supply, or more than 5 volts, to power 74HCxx or 74LSxx logic gates.
- Be careful when using the output from a logic gate to power even a low-current LED. Check how many milliamps are being drawn. Also be careful when "sharing" the output from a logic gate with the input of another gate, at the same time that it is driving an LED. The LED may pull down the output voltage, to a point where the other gate won't recognize it. Always check currents and voltages when modifying a circuit or designing a new one.
- Never apply a significant voltage or current to the output pin of a logic gate. In other words, don't force an input into an output.
- Never link the outputs from two or more logic gates. If they must share a common output wire, use diodes to protect them from each other. See Figure 4-78.

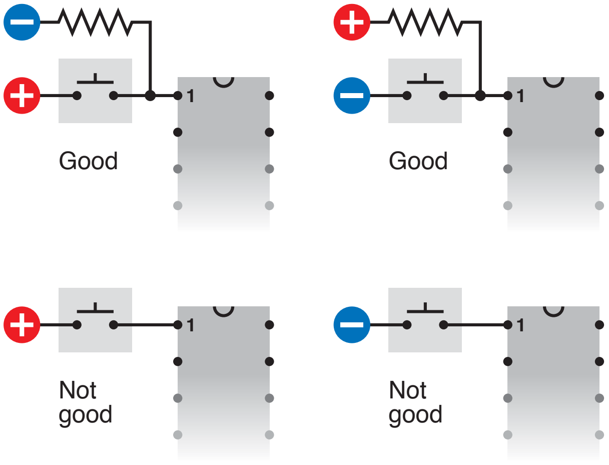

Figure 4-77. *Because a CMOS chip is so sensitive to input fluctuations, a logical input should never be left "floating," or unattached to a defined voltage source. This means that any single-throw switch or pushbutton should be used with a pull-up or pull-down resistor, so that when the contacts are open, the input is still defined.*

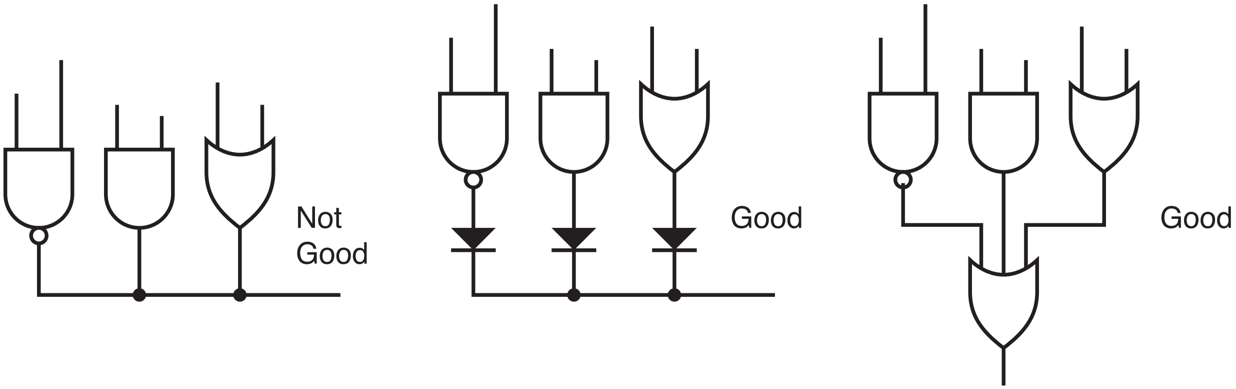

Figure 4-78. *The output from one logic gate must not be allowed to feed back into the output from another logic gate. Diodes can be used to isolate them, or they can be linked via another gate.*

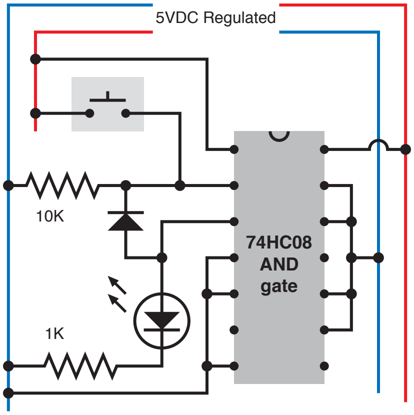

Figure 4-79. *Using a diode, the logical output from a gate can be allowed to feed back to one of its inputs, so that the gate latches after receiving a brief logical input pulse.*

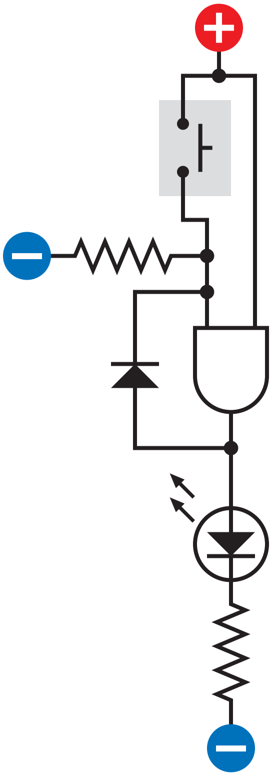

Figure 4-80. *The breadboard-format schematic in is simplified here to show more clearly the way in which a gate can latch itself after receiving an input pulse.*

In the 74HCxx logic family, each input of a logic gate consumes just a microamp, while the output can source 4 milliamps. This seems paradoxical: how can the chip give out more than it takes in? The answer is that it also consumes power from the power supply attached to pins 7 and 14. That's where the additional electricity comes from.

Because the logical output from a chip can be greater than the logical input, we can put the chip in a state where it keeps itself "switched on" in a way which is similar to the way the relay in the alarm project was wired to lock itself on. The simplest way to do this in a logic chip is by feeding some of the output back to one of the inputs.

Figure 4-79 shows an AND gate with one of its inputs wired to positive and its other input held low by a pull-down resistor, with a pushbutton that can make the input high. A signal diode connects the output of the chip back to the pushbutton-controlled input. Remember that the diode has a mark on it indicating the end which should be connected to the *negative* side of the power supply, which in this case will be the end of the 10K resistor.

The schematic in Figure 4-79 shows how the circuit should look in breadboard format. Figure 4-80 shows it in a simpler format.

> *From this point on, I won't bother to show the power regulator and the capacitors associated with it. Just remember to include them every time you see the power supply labeled as "5V DC Regulated."*

When you switch on the power, the LED is dark, as before. The AND gate needs a positive voltage on both of its logical inputs, to create a positive output, but it now has positive voltage only on one of its inputs, while the other input is pulled down by the 10K resistor. Now touch the pushbutton, and the LED comes on. Let go of the pushbutton, and the LED stays on, because the positive output from the AND gate circulates back through the diode and is high enough to overcome the negative voltage coming through the pull-down resistor.

The output from the AND gate is powering its own input, so the LED will stay on until we disconnect it. This arrangement is a simple kind of "latch," and can be very useful when we want an output that continues after the user presses and releases a button.

You can't just connect the output from the gate to one of its inputs using an ordinary piece of wire, because this would allow positive voltage from the tactile switch to flow around and interfere with the output signal. Remember, you must never apply voltage to the output pin of a logic gate. The diode prevents this from happening.

If you've grasped the basics of logic gates, you're ready now to continue to our first real project, which will use all the information that I've set out so far.

Experiment 20: A Powerful Combination

Suppose you want to prevent other people from using your computer. I can think of two ways to do this: using software, or using hardware. The software would be some kind of startup program that intercepts the normal boot sequence and requests a password. You could certainly do it that way, but I think it would be more fun (and more relevant to this book) to do it with hardware. What I'm imagining is a numeric keypad requiring the user to enter a secret combination before the computer can be switched on.

You will need:

- Numeric keypad. As specified in the shopping list at the beginning of this chapter, it must have a "common terminal" or "common output." The schematic in Figure 4-82 shows what I mean. Inside the keypad, one conductor (which I have colored red to distinguish it from the others) connects with one side of every pushbutton. This conductor is "common" to all of them. It emerges from the keypad on an edge connector or set of pins at the bottom, which I've colored yellow.

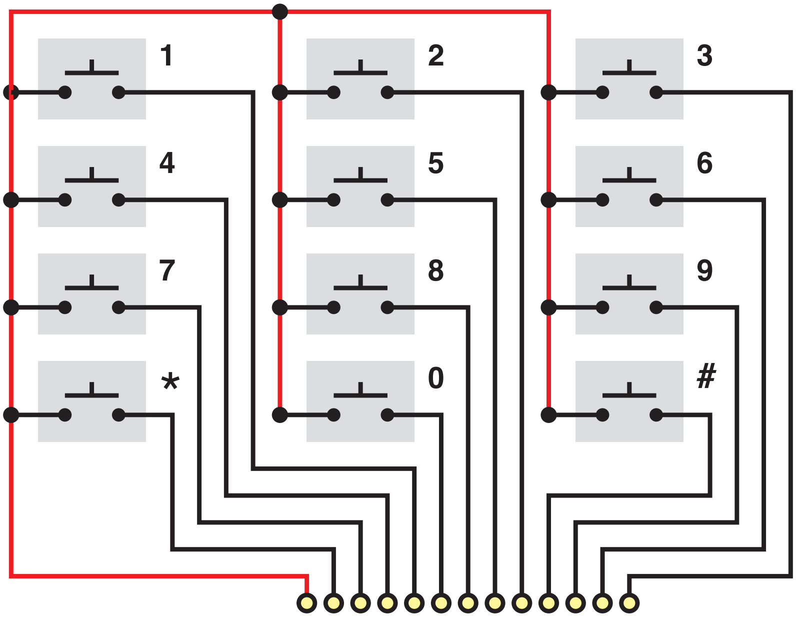

Figure 4-82. *A keypad of the type required for Experiment 20 incorporates a common terminal connected to one side of each of the 12 pushbuttons. The wire from the common terminal is shown red, here, to make it more easily identifiable.*

- Keypads that use "matrix encoding" won't work with the circuit that I'm going to describe. If the Velleman keypad, which I recommend, is unavailable, and you can't find another like it, you can use 12 separate SPST pushbuttons. Of course, that will cost a little more.
- 74HC08 logic chip containing four AND gates. Quantity: 1.
- 74HC04 logic chip containing six inverters. Quantity: 1.
- 555 timer chip. Quantity: 1.

The Warranty Issue

If you follow this project all the way to its conclusion, you'll open your desktop computer, cut a wire, and saw a hole in the cabinet. Without a doubt, this will void your warranty. If this makes you nervous, here are three options:

1. Breadboard the circuit for fun, and leave it at that.

2. Use the numeric keypad on some other device.

3. Use it on an old computer.

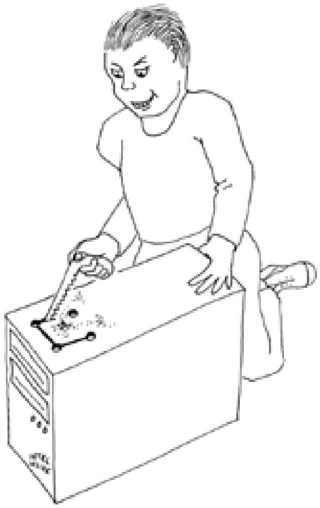

Figure 4-81. *Caution: This just might void your warranty.*

- Latching relay, 5 volt, DPST or DPDT, "2 form C" package, Panasonic DS2E-SL2-DC5V or similar. Must have two separate coils (one to latch, one to unlatch) with separate inputs. Quantity: 1.
- LEDs, 5mm generic, your choice of colors. Quantity: 3.
- Ribbon cable, with six conductors minimum, if you want to do a really neat job. You can use a cable of the type sold for hard drives, and split off the six conductors that you need, or shop around on eBay.
- Tools to open your computer, drill four holes, and make saw cuts between the holes, to create a rectangular opening for the keypad (if you want to take this project to its conclusion). Also, four small bolts to attach the keypad to the computer cabinet after you create the opening for it.

The Schematic

This time I'd like you to study the schematic before building anything. Let's start with the simplified version, shown in Figure 4-83.

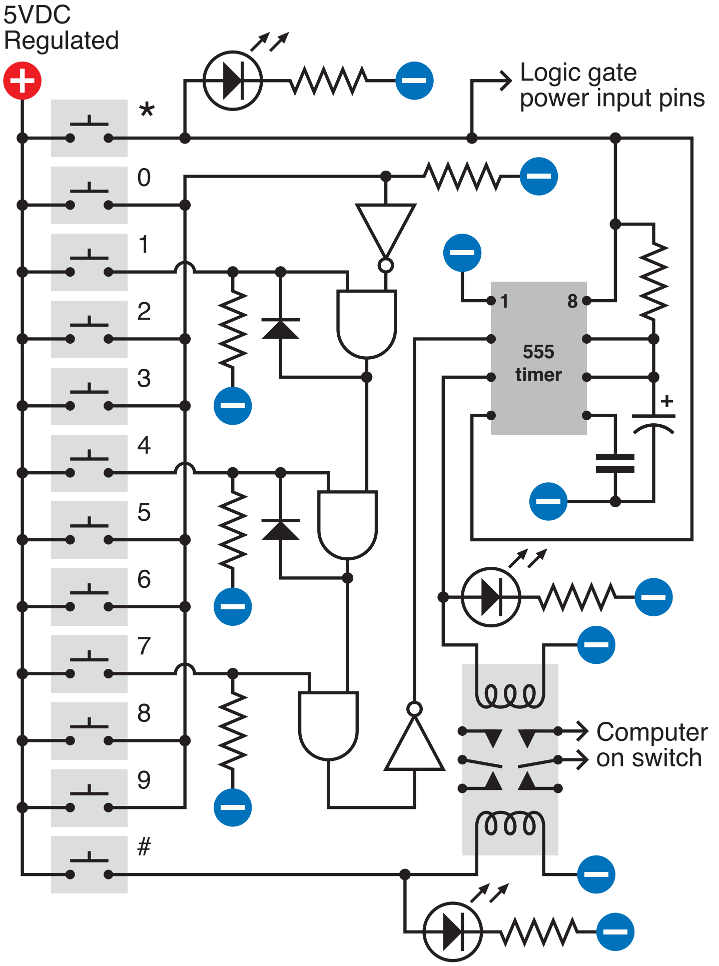

Figure 4-83. *A simplified schematic showing the basic structure of the combination lock circuit.*

I want this to be a battery-powered circuit, so that you don't have to run a separate power supply to it or (worse) try to tap into your computer's 5-volt bus. Battery power means that the circuit has to be "off" most of the time, to prevent the battery from running down. Because the keypad has two spare buttons (the asterisk and the pound sign), I'm going to use the asterisk as the "power on" button. When you press it, the LED at the top of the schematic lights up to confirm that everything's working, and the button sends power to the two logic chips and the 555 timer. You have to hold down the asterisk button while you punch in a three-digit code to unlock the computer.

Arbitrarily, I've chosen 1-4-7 as the three-digit code. Let's track what happens when you enter this sequence. (Naturally, if you build the circuit, you can wire it to choose any three digits you prefer.)

Pressing the 1 button sends positive power to one logical input of the first AND gate. The other logical input of this gate is also positive, because an inverter is supplying it, and the input of the inverter is being held negative by a pull-down resistor. When an inverter has a negative input, it gives a positive output, so pressing the 1 button activates the AND gate, and makes its output positive. The AND gate locks itself on, as its output cycles back to its switched input via a diode. So the gate output remains high even after you let go of the 1 button.

The output from the first AND gate also supplies one logical input of the second AND gate. When you press the 4 button, you send positive voltage to the other logical input of this AND gate, so its output goes high, and it locks itself on, just as the first gate did.

The second AND gate feeds the third AND gate, so when you press the 7 button, the third AND gate changes its output from low to high. This passes through an inverter, so the output from the inverter goes from high to low. This in turn goes to the trigger of a 555 timer wired in monostable mode.

When the trigger of a 555 timer goes from high to low, the timer emits a positive pulse through its output, pin 3. This runs down to the upper coil of the latching relay, and also flashes an LED to confirm that the code has been accepted and the relay has been activated.

Two of the contacts in the relay are wired into the power-up button of your computer. A little later in this description I'll explain why this should be safe with any modern computer.

Because we're using a latching relay, it flips into its "on" state and remains there, even when the power pulse from the timer ends. So now you can let go of the asterisk button to disconnect the battery power to your combination lock, and press the power-up button that switches on your computer.

At the end of your work session, you shut down your computer as usual, then press the pound button on your keypad, which flips the relay into its other position, reactivating the combination lock.

Incorrect Inputs

What happens if you enter the wrong code? If you press any button other than 1, 4, or 7, it sends positive voltage to the inverter near the top of the schematic. The positive voltage overwhelms the negative voltage being applied to the inverter through a pull-down resistor, and causes the inverter to output a negative voltage, which it applies to one of the logical inputs of the first AND gate. If the AND gate was locked on, the negative input will switch it off. If it was supplying the second AND gate, it'll switch that one off too.

Thus, any error when entering the first, second, or third digit of the secret code will reset the AND gates, forcing you to begin the sequence all over again.

What if you enter 1, 4, and 7 out of their correct sequence? The circuit won't respond. The third AND gate needs a high input supplied by the second AND gate, and the second AND gate needs a high input supplied by the first AND gate. So you have to activate the AND gates in the correct sequence.

Questions

Why did I use a 555 timer to deliver the pulse to the relay? Because the logical output from an AND gate cannot deliver sufficient power. I could have passed it through a transistor, but I liked the idea of a pulse of a fixed length to flip the relay and illuminate an LED for about 1 second, regardless of how briefly the user presses the 7 button.

Why do I need three LEDs? Because when you're punching buttons to unlock your computer, you need to know what's going on. The Power On LED reassures you that your battery isn't dead. The Relay Active LED tells you that the system is now unlocked, in case you are unable to hear the relay click. The System Relocked LED reassures you that you have secured your computer.

Because all the LEDs are driven either directly from the 5-volt supply or from the output of the 555 timer, they don't have to be low-current LEDs and can be used with 330Ω series resistors, so they'll be nice and bright.

How do you connect the keypad with the circuit? That's where your ribbon cable comes in. You carefully strip insulation from each of the conductors, and solder them to the contact strip or edge connector on your keypad. Push the conductors on the other end of the cable into your breadboard (when you're test-building the circuit) or solder them into perforated board (when you're building it permanently). Find a convenient spot inside your computer case where you can attach the perforated board, with double-sided adhesive or small bolts or whatever is convenient. Include a 9-volt battery carrier, and don't forget your power regulator to step the voltage down to 5 volts.

Breadboarding

No doubt you have realized by now that breadboards are very convenient as a quick way to push in some components and create connections, but the layout of their conductors forces you to put components in unintuitive configurations. Still, if you carefully compare the breadboard schematic in Figure 4-83 with the simplified schematic in Figure 4-84, you'll find that the connections are the same.

To help it make sense, I've shown the logic gates that exist inside the chips. I've also colored the power supply wires, as before, to reduce the risk of confusion. The positive side of the supply goes only to the common terminal on your keypad, and you have to press the asterisk key to send the power back down the ribbon cable, to supply the chips.

If you build the circuit and you can't understand why everything's dead, it's most likely because you forgot to hold down the asterisk button.

Note that the "wrong" numbers on the keypad are all shorted together. This will create some inconvenience if you want to change the combination in the future. I'll suggest a different option in the "enhancements" section that follows. For now, ideally, you should run a wire from every contact on your keypad, down to your circuit on its breadboard, and short the "wrong" keypad numbers together with jumper wires on the breadboard.

Also note that if you use a meter to test the inputs to the AND gates, and you touch your finger against the meter probe while doing so, this can be sufficient to trigger the sensitive CMOS inputs and give a false positive.

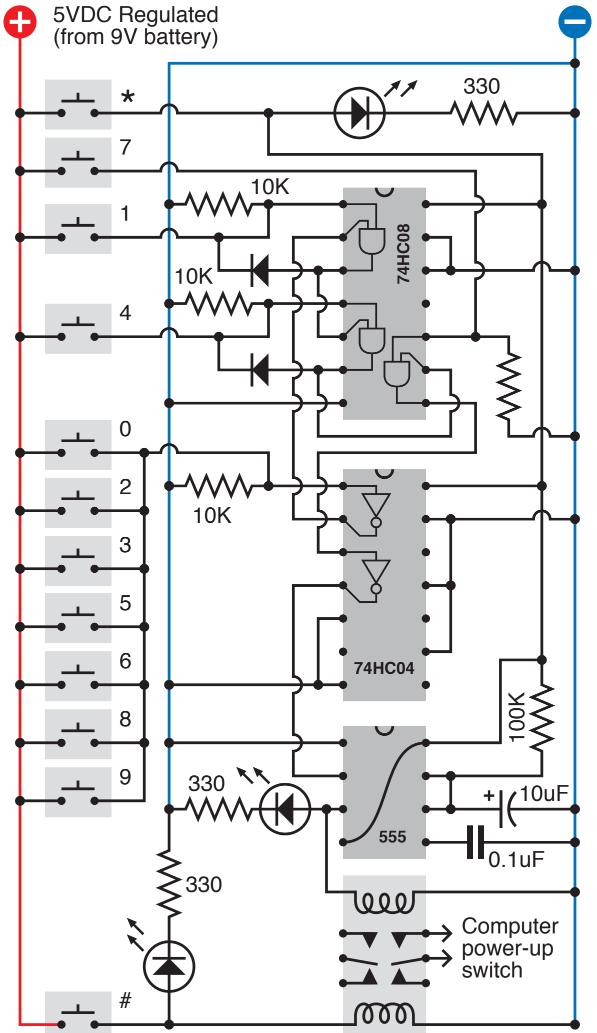

Figure 4-84. *The combination lock schematic redrawn to show how the components can be laid out on a breadboard.*

One Little Detail: The Computer Interface

Old computers used to have a big switch at the back, attached to the heavy metal box inside the computer, that transformed house current to regulated voltages that the computer needs. Most modern computers are not designed this way; you leave the computer plugged in, and you touch a little button on the box (if it's a Windows machine) or the keyboard (if it's a Mac), which sends a low-voltage pulse to the motherboard.

This is ideal from our point of view, because we don't have to mess with high voltages. Don't even think of opening that metal box with the fan mounted in it, containing the computer power supply. Just look for the wire (usually containing two conductors, on a Windows machine) that runs from the "power up" button to the motherboard.

To check that you found the right one, *make sure that your computer is unplugged*, ground yourself (because computers contain CMOS chips that are sensitive to static electricity) and very carefully snip just one of the two conductors in the wire. Now plug in your computer and try to use the "power up" button. If nothing happens, you've probably cut the right wire. Even if you cut the wrong wire, it still prevented your computer from booting, which is what you want, so you can use it anyway. Remember, we are not going to introduce any voltage to this wire. We're just going to use the relay as a switch to reconnect the conductor that you cut. You should have no problem if you maintain a cool and calm demeanor, and look for that single wire that starts everything. Check online for the maintenance manual for your computer if you're really concerned about making an error.

After you find the wire and cut just one of its conductors, unplug your computer again, and keep it unplugged during the next steps.

Find where the wire attaches to the motherboard. Usually there's a small unpluggable connector. First, mark it so that you know how to plug it back in the right way around, and then disconnect it while you follow the next couple of steps.

Strip insulation from the two ends of the wire that you cut, and solder an additional piece of two-conductor wire, as shown in Figure 4-85, with heat-shrink tube to protect the solder joints. (This is very important!)

Run your new piece of wire to the latching relay, making sure you attach it to the pair of contacts which close, inside the relay, when it is energized by the unlocking operation. You don't want to make the mistake of unlocking your computer when you think you're locking it, and vice versa.

Reconnect the connector that you disconnected from your motherboard, plug in your computer, and try to power it up. If nothing happens, this is probably good! Now enter the secret combination on your keypad (while holding down the asterisk button to provide battery power) and listen for the click as the relay latches. Now try the "power up" button again, and everything should work.

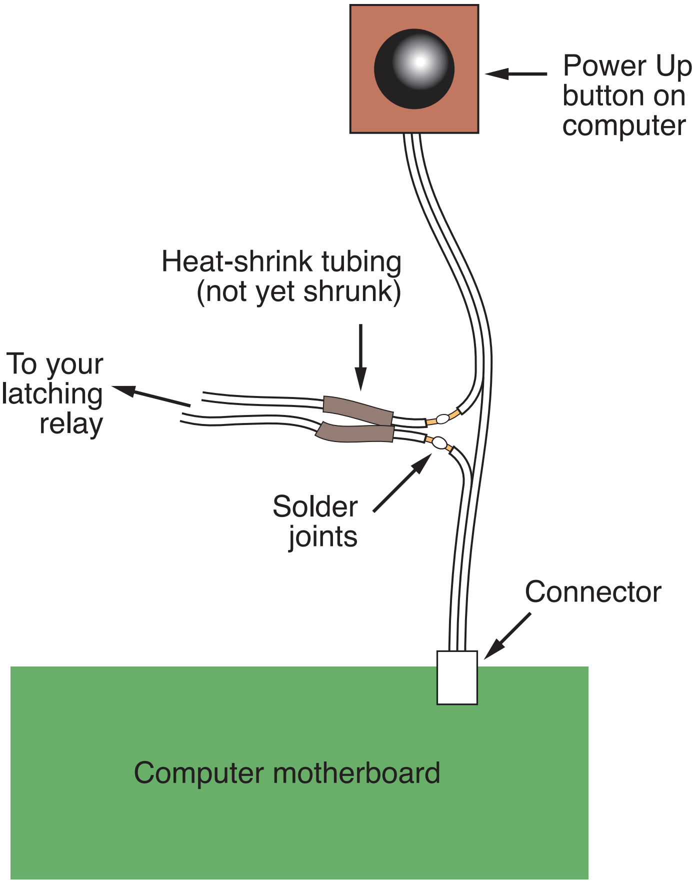
Power Up
button on
computer
Heat-shrink tubing
(not yet shrunk)
To your
latching
relay
Solder
joints
Connector
Computer motherboard

Figure 4-85. *The combination lock project can be interfaced with a typical desktop computer by cutting one conductor in the wire from the "power up" pushbutton, soldering an extension, and covering the joints with heat-shrink tube.*

Enhancements

At the end of any project, there's always more you can do.

To make this setup more secure, you could remove the usual screws that secure the case of the computer, and replace them with tamper-proof screws. Check any online source for "tamper-proof screws," such as *http://www.mcmaster.com*. Naturally, you will also need the special tool that fits the screws, so that you can install them (or remove them, if your security system malfunctions for any reason).

Another enhancement could be an additional 555 timer that is activated by the asterisk button, and delivers power to the other chips for, say, a limited period of 30 seconds, allowing you that much time to unlock the system. This would eliminate the need to hold down the asterisk button while you enter the unlocking code. A 555 timer can supply power to all the other chips, because they don't use very much. I omitted this feature for the sake of simplicity.

Yet another enhancement, if you are security-crazed, is to go for a four-button code. After all, the 74HC08 chip still has one unused AND gate. You could insert that into the chain of the existing AND gates and wire it to another keypad button of your choice.

Still another enhancement would be a way to change the code without unsoldering and resoldering wires. You can use the miniature sockets that I suggested in the heartbeat flasher project. These should enable you to swap around the ends of your wires from the keypad.

And for those who are absolutely, positively, totally paranoid, you could fix things so that entering a wrong code flips a second high-amperage relay which supplies a massive power overload, melting your CPU and sending a big pulse through a magnetic coil clamped to your hard drive, instantly turning the data to garbage (Figure 4-86). Really, if you want to protect information, messing up the hardware has major advantages compared with trying to erase data using software. It's faster, difficult to stop, and tends to be permanent. So, when the Record Industry Association of America comes to your home and asks to switch on your computer so that they can search for illegal file sharing, just accidentally give them an incorrect unlocking code, sit back, and wait for the pungent smell of melting insulation.

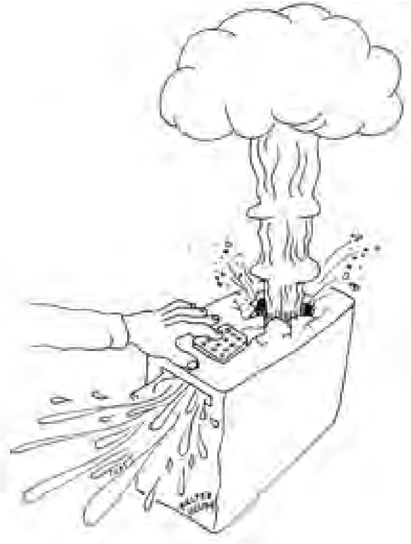

Figure 4-86. *For those who are absolutely, positively, totally paranoid: a meltdown/self-destruct system controlled by a secret key combination provides enhanced protection against data theft or intrusions by RIAA investigators asking annoying questions about file sharing.*

If you pursue this option, I *definitely* take no responsibility for the consequences.

On a more realistic level, no system is totally secure. The value of a hardware locking device is that if someone does defeat it (for instance, by figuring out how to unscrew your tamper-proof screws, or simply ripping your keypad out of the computer case with metal shears), at least you'll know that something happened—especially if you put little dabs of paint over the screws to reveal whether they've been messed with. By comparison, if you use password-protection software and someone defeats it, you may never know that your system has been compromised.

Experiment 21: Race to Place

The next project is going to get us deeper into the concept of feedback, where the output is piped back to affect the input—in this case, blocking it. It's a small project, but quite subtle, and the concepts will be useful to you in the future.

You will need:

- 74HC32 chip containing four OR gates. Quantity: 1.
- 555 timers. Quantity: 2.
- SPDT switch. Quantity: 1.
- SPST tactile switches. Quantity: 2.
- Various resistors.
- 5V supply using power regulator as before.

The Goal

On quiz shows such as *Jeopardy*, contestants race to answer each question. The first person who hits his answer button automatically locks out the other contestants, so that their buttons become inactive. How can we make a circuit that will do the same thing?

If you search online, you'll find several hobby sites where other people have suggested circuits to work this way, but they lack some features that I think are necessary. The approach I'm going to use here is both simpler and more elaborate. It's simpler because it has a very low chip count, but it's more elaborate in that it incorporates "quizmaster control" to make a more realistic game.

I'll suggest some initial ideas for a two-player version. After I develop that idea, I'll show how it could be expanded to four or even more players.

A Conceptual Experiment

I want to show how this kind of project grows from an idea to the finished version. By going through the steps of developing a circuit, I'm hoping I may inspire you to develop ideas of your own in the future, which is much more valuable than just replicating someone else's work. So join me in a conceptual experiment, thinking our way from a problem to a solution.

First consider the basic concept: two people have two buttons, and whoever goes first locks out the other person. I always find it helps me to visualize this kind of thing if I draw a sketch, so that's where I'll begin. In Figure 4-87, the signal from each button passes through a component that I'll call a "button blocker," activated by the other person's button. I'm not exactly sure what the button blocker will be or how it will work, yet.

Now that I'm looking at it, I see a problem here. If I want to expand this to three players, it will get complicated, because each player must activate the "button blockers" of *two* opponents. Figure 4-88 shows this. And if I have four players, it's going to get even more complicated.

Anytime I see this kind of complexity, I think there has to be a better way.

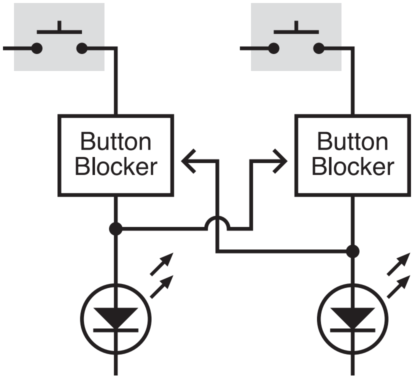

Figure 4-87. *The basic concept of the quiz project is that the output from one button should feed back to intercept the output of another button. At this point, the way in which the "button blocker" circuit works has not been figured out.*

Also, there's another problem. After a player lets his finger off the button, the other players' buttons will be unblocked again. I need a latch to hold the signal from the first player's button and continue to block the other players.

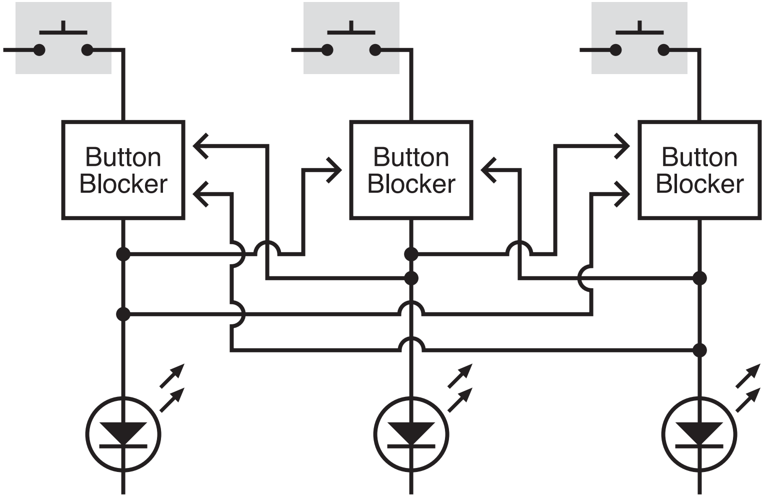

Figure 4-88. *The quiz concept becomes more complicated when an additional pushbutton is added. Now each button must block two other buttons. If a fourth button is added, the circuit will become unmanageably complex. There has to be a better way.*

This now sounds even more complicated. But wait a minute, if I have a latch which allows the winning player to take his finger off his button, I don't care if any of the buttons are being pressed anymore—including the button of the winning player. As soon as his signal is latched, *all* the buttons can be blocked. This makes things much simpler. I can summarize it as a sequence of events:

1. First player presses his button.
2. The signal is latched.
3. The latched signal feeds back and blocks all the buttons.

The new sketch in Figure 4-89 shows this. Now the configuration is modular, and can be expanded to almost any number of players, just by adding more modules.

There's something important missing, though: a reset switch, to put the system back to its starting mode after the players have had time to press their buttons and see who won. Also, I need a way to prevent players from pressing their buttons too soon, before the quizmaster has finished asking the question. Maybe I can combine this function in just one switch, which will be under the quizmaster's control. In its Reset position, the switch can reset the system and remove power to the buttons. In its Play position, the switch stops holding the system in reset mode, and provides power to the buttons. Figure 4-90 shows this. I've gone back to showing just two players, to minimize the clutter of lines and boxes, but the concept is still easily expandable.

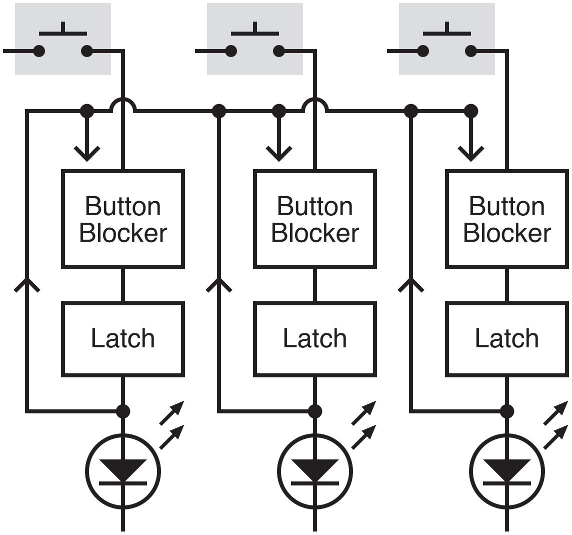

Figure 4-89. *If a latch is added below each button, it can retain one input and then block all inputs from all buttons. This simplifies the concept.*

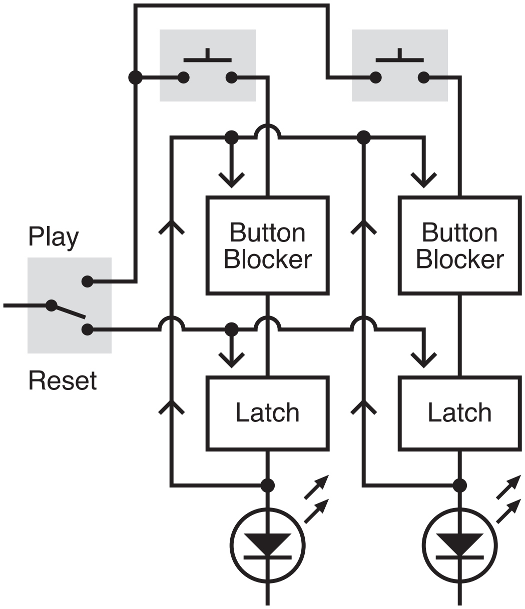

Figure 4-90. *A quizmaster switch will be needed to activate the buttons initially and then reset the circuit after a winning input has been recorded.*

Now I have to deal with a logic problem in the diagram. The way I've drawn it, after the output from the lefthand latch goes up to the "button blockers," it can also run down the wire to the other half of the circuit (against the direction to the arrows), because everything is joined together. In other words, if the lefthand LED lights up, the righthand LED will light up, too. How can I stop this from happening?

Well, I could put diodes in the "up" wires to block current from running down them. But I have a more elegant idea: I'll add an OR gate, because the inputs to an OR gate are separated from each other electrically. Figure 4-91 shows this.

Usually an OR gate has only two logical inputs. Will this prevent me from adding more players? No, because you can actually buy an OR that has eight inputs. If any one of them is high, the output is high. For fewer than eight players, I can short the unused inputs to ground, and ignore them.

Looking again at Figure 4-91, I'm getting a clearer idea of what the thing I've called a "button blocker" should actually be. I think it should be another logic gate. It should say, "If there's only one input, from a button, I'll let it through. But if there is a second input from the OR gate, I won't let it through."

That sounds like a NAND gate, but before I start choosing chips, I have to decide what the latch will be. I can buy an off-the-shelf flip-flop, which flips "on" if it gets one signal and "off" if it gets another, but the trouble is, chips containing flip-flops tend to have more features than I need for a simple circuit like this. Therefore I'm going to use 555 timers again, in flip-flop mode. They require very few connections, work very simply, and can deliver a good amount of current. The only problem with them is that they require a negative input at the trigger pin to create a positive output. But I think I can work with that.

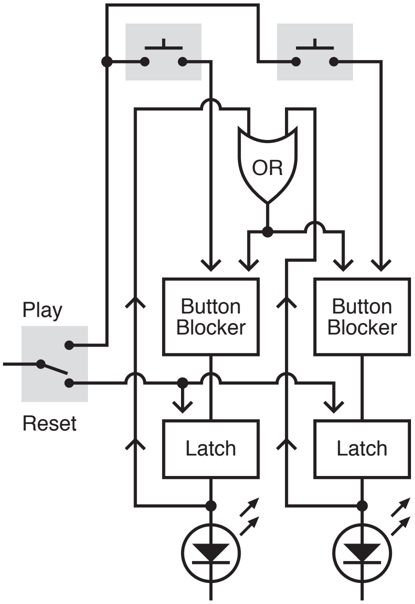

Figure 4-91. *To prevent the output from one latch feeding back around the circuit to the output from another latch, the outputs can be combined in an OR gate.*

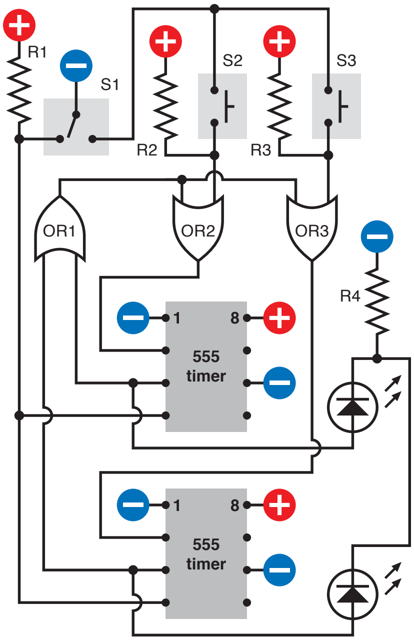

Figure 4-92. *Now that the basic concept of the quiz circuit has been roughed out, specific components can be inserted, with compatible inputs and outputs.*

So now, finally, here's a simplified schematic, in Figure 4-92. I like to show the pins of the 555 timers in their correct positions, so I had to move the components around a little to minimize wire crossovers, but you can see that logically, it's the same basic idea.

Before you try to build it, just run through the theory of it, because that's the final step, to make sure there are no mistakes. The important thing to bear in mind is that because the 555 needs a negative input on its trigger pin to create its output, when any of the players presses a button, the button has to create a negative "flow" through the circuit. This is a bit counterintuitive, so I'm including a three-step visualization in Figure 4-93, showing how it will work.

In Step 1, the quizmaster has asked a question and flipped his switch to the right, to supply (negative) power to the players' buttons. So long as no one presses a button, the pull-up resistors supply positive voltage to OR2 and OR3. An OR gate has a positive output if it has any positive input, so OR2 and OR3 keep the trigger inputs of the 555 timers positive. Their outputs remain low, and nothing is happening yet.

In Step 2, the lefthand player has pressed his button. Now OR2 has two negative inputs, so its output has gone low. But IC1 hasn't reacted yet.

In Step 3, just a microsecond later, IC1 has sensed the low voltage on its trigger, so its output from pin 3 has gone high, lighting the LED. Remember, this 555 timer is in flip-flop mode, so it locks itself into this state immediately. Meanwhile its high output also feeds back to OR1. Because OR1 is an OR gate, just one high input is enough to make a high output, so it feeds this back to OR2 and OR3. And now that they have high inputs, their outputs also go high, and will stay high, regardless of any future button-presses.

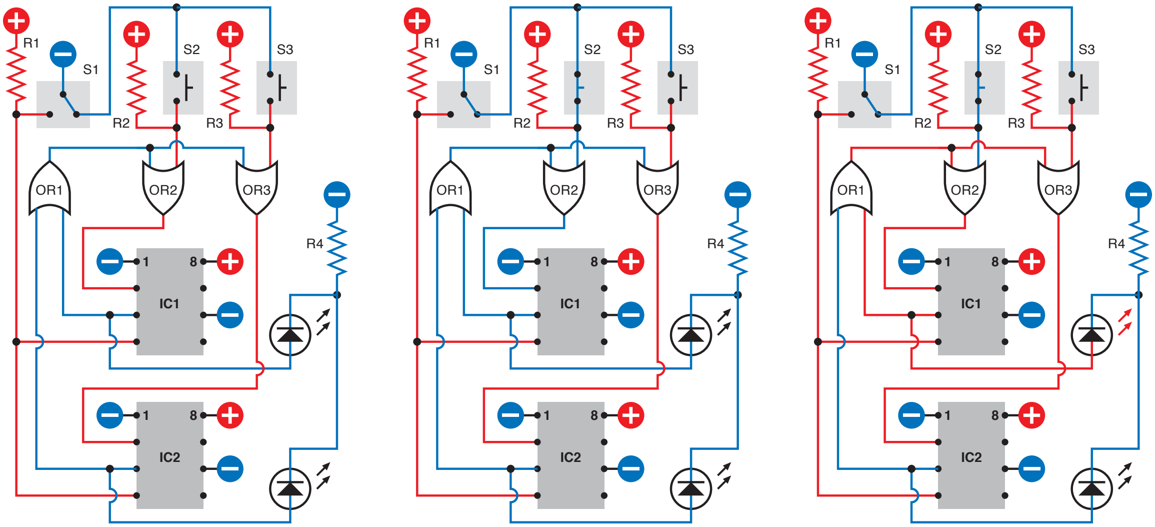

Figure 4-93. *These three schematics show the prevalence of higher and lower voltages (red and blue lines) through the quiz circuit when a pushbutton is pressed.*

Because OR2 and OR3 now have high inputs and outputs, IC1 and IC2 cannot be triggered. But IC1 is still locked into its "on" state, keeping the LED illuminated.

The only way to change IC1 is if the quizmaster flips his switch back to the left. That applies negative power to the reset pins of both the timers. Consequently their outputs go low, the LED goes out, and the circuit goes back into the same state as where it started. Having reset it, the quizmaster can ask another question, but the players' buttons are not activated until the quizmaster flips the switch back to the right again.

There's only one situation that I haven't addressed: what if both players press their buttons absolutely simultaneously? In the world of digital electronics, this is highly unlikely. Even a difference of a microsecond should be enough time for the circuit to react and block the second button. But if somehow both buttons are pressed at the same instant, both of the timers should react, and both of the LEDs will light up, showing that there has been a tie.

In case you feel a little uncertain about the way in which a two-player circuit can be upgraded to handle extra players, I've included a simplified three-player schematic in Figure 4-94.

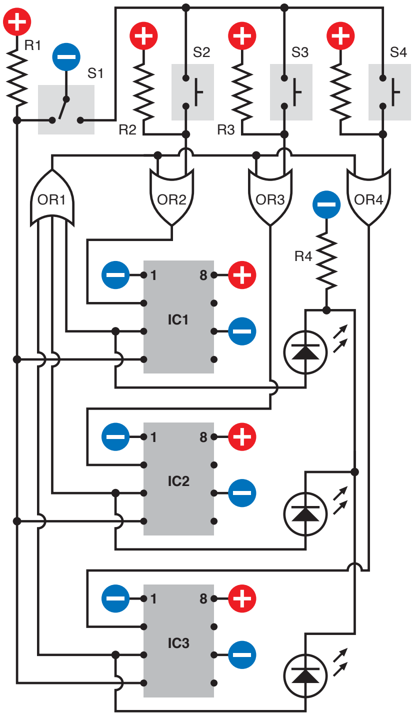

Figure 4-94. *The two-player schematic in can be easily upgraded to a three-player version, as shown here, provided the first OR gate can handle three inputs.*

Breadboarding It

Now it's time to create a schematic that's as close to the breadboard layout as possible, so that you can build this thing easily. The schematic is shown in Figure 4-95 and the actual components on a breadboard are in Figure 4-96. Because the only logic gates that I've used are OR gates, and there are only three of them, I just need one logic chip: the 74HC32, which contains four 2-input OR gates. (I've grounded the inputs to the fourth). The two OR gates on the left side of the chip have the same functions as OR2 and OR3 in my simplified schematic, and the OR gate at the bottom-right side of the chip works as OR1, receiving input from pin 3 of each 555 timer. If you have all the components, you should be able to put this together and test it quite quickly.

You may notice that I've made one modification of the previous schematic. A 0.01 µF capacitor has been added between pin 2 of each 555 timer (the Input) and negative ground. Why? Because when I tested the circuit without the capacitors, sometimes I found that one or both of the 555 timers would be triggered simply by flipping S1, the quizmaster switch, without anyone pressing a button.

At first this puzzled me. How were the timers getting triggered, without anyone doing anything? Maybe they were responding to "bounce" in the quizmaster switch. Sure enough, the small capacitors solved the problem. They may also slow the response of the 555 timers fractionally, but not enough to interfere with slow human reflexes.

As for the buttons, it doesn't matter if they "bounce," because each timer locks itself on at the very first impulse and ignores any hesitations that follow.

You can experiment building the circuit, disconnecting the 0.01 µF capacitors, and flipping S1 to and fro a dozen times. If you have a high-quality switch, you may not experience any problem. If you have a lower-quality switch, you may see a number of "false positives." I'm going to explain more about "bounce," and how to get rid of it, in the next experiment.

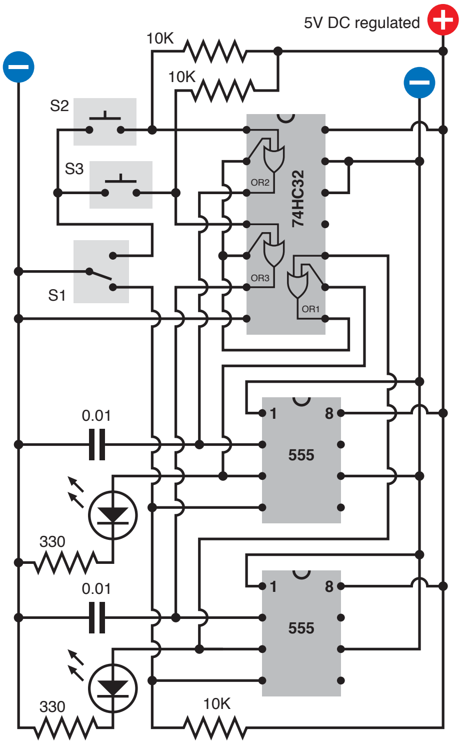

Figure 4-95. *Applying the simplified schematic to a breadboard inevitably entails a wiring layout that is less intuitively obvious and appears more complex. The connections are the same, though.*

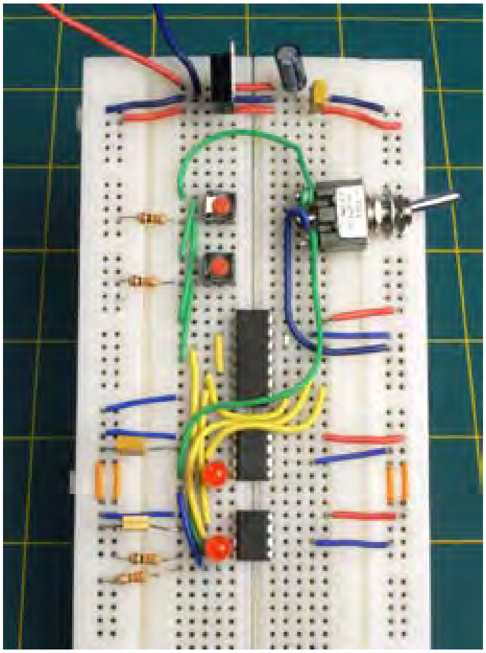

Figure 4-96. *The quiz schematic applied to a breadboard, to test the concept prior to full-scale implementation.*

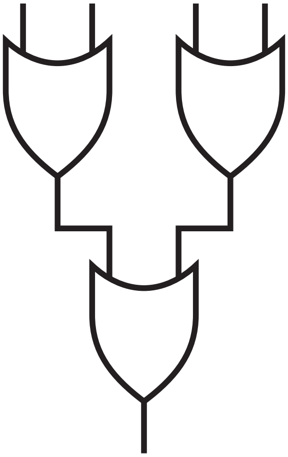

Figure 4-97. *Although a four-input OR gate is not manufactured, its functionality can be achieved easily by linking three 2-input OR gates together.*

Enhancements

After you breadboard the circuit, if you proceed to build a permanent version, I suggest that you expand it so that at least four players can participate. This will require an OR gate capable of receiving four inputs. The 74HC4078 is the obvious choice, as it allows up to eight. Just connect any unused inputs to negative ground.

Alternatively, if you already have a couple of 74HC32 chips and you don't want to bother ordering a 74HC4078, you can gang together three of the gates inside a single 74HC32 so that they function like a four-input OR. Look at the simple logic diagram in Figure 4-97 showing three ORs, and remember that the output from each OR will go high if at least one input is high.

And while you're thinking about this, see if you can figure out the inputs and output of three ANDs in the same configuration.

For a four-player game, you'll also need two additional 555 timers, of course, and two more LEDs, and two more pushbuttons.

As for creating a schematic for the four-player game—I'm going to leave that to you. Begin by sketching a simplified version, just showing the logic symbols. Then convert that to a breadboard layout. And here's a suggestion: pencil, paper, and an eraser can still be quicker, initially, than circuit-design software or graphic-design software, in my opinion.

Experiment 22: Flipping and Bouncing

I mentioned in the previous experiment that "bounce" from the buttons in the circuit wouldn't be a problem, because the buttons were activating 555 timers that were wired in bistable, flip-flop mode. As soon as the timer receives the very first pulse, it flips into its new state and flops there, ignoring any additional noise in the circuit. So can we debounce a switch or a button using a flip-flop? And as some 74HCxx chips are available containing flip-flops, can we use them?

The answers are yes, and yes, although it's not quite as simple as it sounds.

You will need:

- 74HC02 logic chip containing 4 NOR gates. 74HC00 logic chip containing 4 NAND gates. Quantity: 1 of each.
- SPDT switch. Quantity: 1.
- LEDs, low-current. Quantity: 2.
- 10K resistors and 1K resistors. Quantity: 2 of each.

Assemble the components on your breadboard, following the schematic shown in Figure 4-98. When you apply power (through your regulated 5-volt supply), one of the LEDs should be lit.

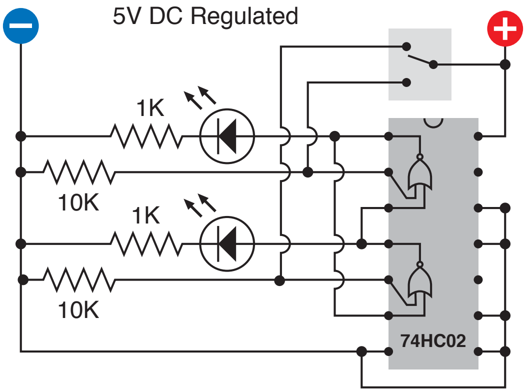

Figure 4-98. *A simple circuit to test the behavior of two NOR gates wired as a simple flip-flop that retains its state after an input pulse ceases.*

Now I want you to do something odd. Please disconnect the SPDT switch by taking hold of the wire that connects the positive power supply to the pole of the switch, and pulling the end of the wire out of the breadboard. When you do this, you may be surprised to find that the LED remains lit.

Push the wire back into the breadboard, flip the switch, and the first LED should go out, while the other LED should become lit. Once again, pull out the wire, and once again, the LED should remain lit.

Here's the take-home message:

- A flip-flop requires only an initial pulse.
- After that, it ignores its input.

How It Works

Two NOR gates or two NAND gates can function as a flip-flop:

- Use NOR gates when you have a positive input from a double-throw switch.
- Use NAND gates when you have a negative input from a double-throw switch.

Either way, you have to use a double-throw switch.

I've mentioned the double-throw switch three times (actually, four times if you count this sentence!) because for some strange reason, most introductory books fail to emphasize this point. When I first started learning electronics, I went crazy trying to understand how two NORs or two NANDs could debounce a simple SPST pushbutton—until finally I realized that they can't. The reason is that when you power up the circuit, the NOR gates (or NAND gates) need to be told in which state they should begin. They need an initial orientation, which comes from the switch being in one state or the other. So it has to be a double-throw switch. (Now I've mentioned it five times.)

I'm using another simplified multiple-step schematic, Figure 4-99, to show the changes that occur as the switch flips to and fro with two NOR gates. To refresh your memory, I've also included a truth table showing the logical outputs from NOR gates for each combination of inputs.

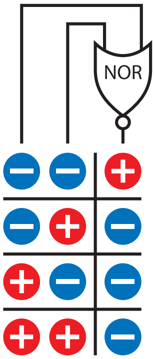

−	−	+
−	+	−
+	−	−
+	+	−

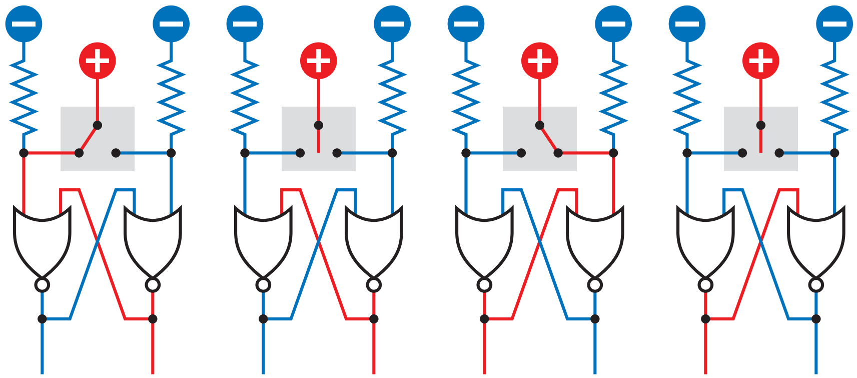

Figure 4-99. *Using two NOR gates in conjunction with a positive input through a SPDT switch, this sequence of four diagrams shows how a flip-flop circuit responds.*

Suppose that the switch is turned to the left. It sends positive current to the lefthand side of the circuit, overwhelming the negative supply from the pull-down resistor, so we can be sure that the NOR gate on the left has one positive logical input. Because any positive logical input will make the NOR give a negative output (as shown in the truth table), the negative output crosses over to the righthand NOR, so that it now has two negative inputs, which make it give a positive output. This crosses back to the lefthand NOR gate. So, in this configuration everything is stable.

Now comes the clever part. Suppose that you move the switch so that it doesn't touch either of its contacts. (Or suppose that the switch contacts are bouncing, and failing to make a good contact. Or suppose you disconnect the switch entirely.) Without a positive supply from the switch, the lefthand input of the left NOR gate goes from positive to negative, as a result of the pull-down resistor. But the righthand input of this gate is still positive, and one positive is all it takes to make the NOR maintain its negative output, so nothing changes. In other words, the circuit has "flopped" in this state.

Now if the switch turns fully to the right and supplies positive power to the righthand pin of the right NOR gate, quick as a flash, that NOR recognizes that it now has a positive logical input, so it changes its logical output to negative. That goes across to the other NOR gate, which now has two negative inputs, so its output goes positive, and runs back to the right NOR.

In this way, the output states of the two NOR gates change places. They flip, and then flop there, even if the switch breaks contact or is disconnected again. The second set of drawings in Figure 4-100 shows exactly the same logic, using a negatively powered switch and two NAND gates. You can use your 74HC00 chip, specified in the parts list for this experiment, to test this yourself.

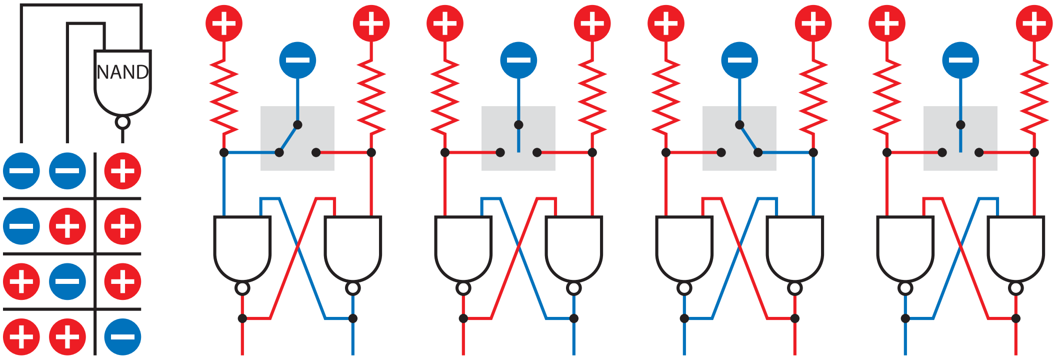

Figure 4-100. *The schematic from Figure 4-99 can be rewired with NAND gates and a negative switched input.*

Both of these configurations are examples of a jam-type flip-flop, so called because the switch forces it to respond immediately, and jams it into that state. You can use this circuit anytime you need to debounce a switch (as long as it's a double-throw switch).

A more sophisticated version is a clocked flip-flop, which requires you to set the state of each input first and then supply a clock pulse to make the flip-flop respond. The pulse has to be clean and precise, which means that if you supply it from a switch, the switch must be debounced—probably by using another jam-type flip-flop! Considerations of this type have made me reluctant to use clocked flip-flops in this book. They add a layer of complexity, which I prefer to avoid in an introductory text.

What if you want to debounce a single-throw button or switch? Well, you have a problem! One solution is to buy a special-purpose chip such as the 4490 "bounce eliminator," which contains digital delay circuitry. A specific part number is the MC14490PG from On Semiconductor. This contains six circuits for six separate inputs, each with an internal pull-up resistor. It's relatively expensive, however—more than 10 times the price of a 74HC02 containing NOR gates. Really, it may be simpler to use double-throw switches that are easily debounced as described previously.

Experiment 23: Nice Dice

This is the one experiment where I want you to use the 74LSxx generation of the TTL logic family, instead of the 74HCxx family of CMOS. Two reasons: first, I need to use the 7492 counter, which is unavailable in the HC family. And second, you should know the basic facts about the LS series of TTL chips, as they still crop up in circuits that you'll find in electronics books and online.

In addition, you'll learn about "open collector" TTL chips such as the 74LS06 inverter, which can be a convenient substitute for transistors when you want to deliver as much as 40mA of current.

The idea of this circuit is simple enough: run a 555 timer sending very fast pulses to a counter that counts in sixes, driving LEDs that are placed to imitate the spots on a die. (Note that the word "die" is the singular of "dice.") The counter runs so fast, the die-spots become a blur. When the user presses a button, the counter stops arbitrarily, displaying an unpredictable spot pattern.

Dice simulations have been around for many, many years, and you can even buy kits online. But this one will do something more: it will also demonstrate the principles of binary code.

So, if you're ready for the triple threat of TTL chips, open collectors, and binary, let's begin.

You will need:

- 74LS92 counter such as SN74LS92N by Texas Instruments. Quantity: 1 if you want to create one die, 2 to make two dice.
- 74LS27 three-input NOR gate such as SN74LS27N by Texas Instruments. Quantity: 1.
- 555 timers. Quantity: 1 if you want to make one die, 2 to make two dice.
- Signal diodes, 1N4148 or similar. Quantity: 4, or 8 to make two dice.

Seeing Binary

The counter that we dealt with before was unusual, in that its outputs were designed to drive seven-segment numerals. A more common type has outputs that count in binary code.

The 74LS92 pinouts are shown in Figure 4-101. Plug the chip into your breadboard and make connections as shown in Figure 4-102. Initially, the 555 timer will drive the counter in slow-motion, at around 1 step per second. Figure 4-103 shows the actual components on a breadboard.

Note that the counter has unusual power inputs, on pins 5 and 10 instead of at the corners. Also four of its pins are completely unused, and do not connect with anything inside the chip. Therefore, you don't need to attach any wire to them on the outside.

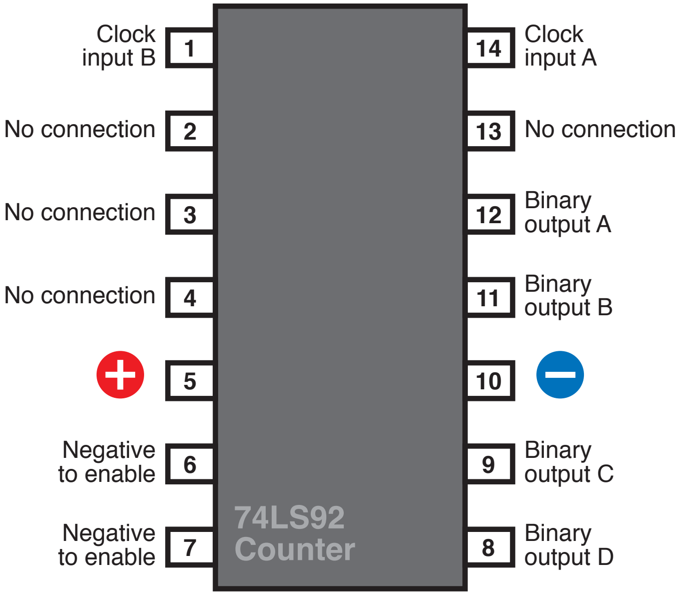
Clock input B
1
No connection
2
No connection
3
No connection
4
5
Negative to enable
6
Negative to enable
7
74LS92 Counter
14
Clock input A
13
No connection
12
Binary output A
11
Binary output B
10
9
Binary output C
8
Binary output D

Figure 4-101. *The unusual pin assignments include four that have no connection of any kind inside the chip, and can be left unattached.*

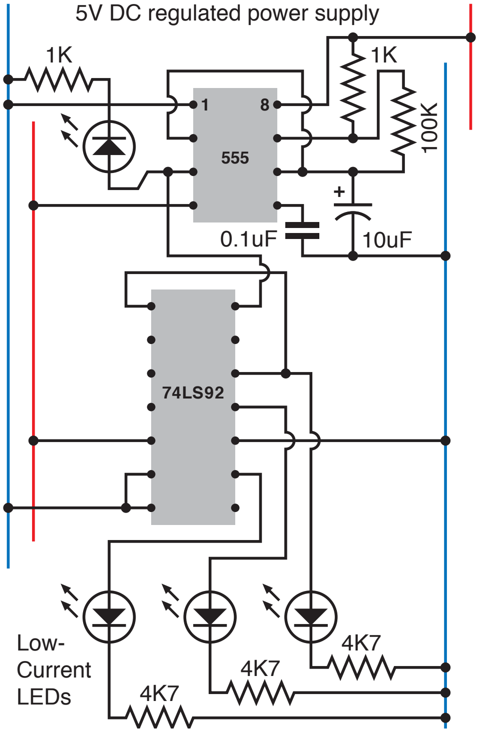
5V DC regulated power supply
1K
1K
1
8
555
100K
+
0.1uF
10uF
74LS92
Low-Current LEDs
4K7
4K7
4K7

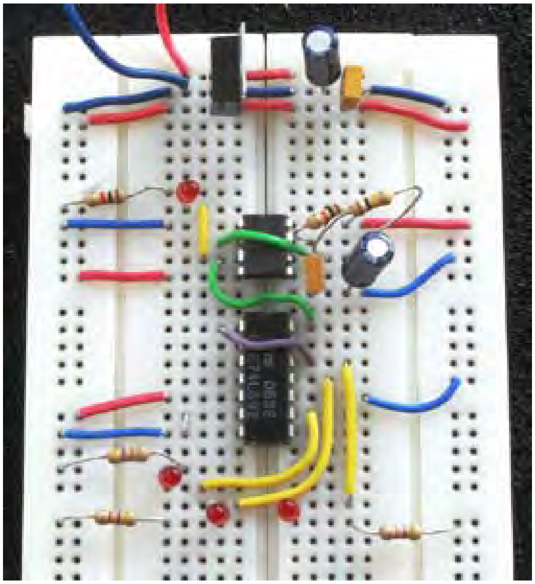

Figure 4-102. *This simple circuit uses a 555 timer running slowly to control the 74LS92 binary counter and display the succession of high states from its outputs.*

Figure 4-103. *The breadboard version of the schematic in Figure 4-102 to display the outputs from a 74LS92 counter.*

Now we come to the first new and difficult fact about the 74LSxx generation of TTL chips that makes them less desirable, for our purposes, than the 74HCxx generation of CMOS chips that I have recommended in previous projects. The modern and civilized HC chips will source 4mA or sink 4mA at each logical output, but the older LS generation is fussier. It will sink around 8mA into each output pin from a positive source, but when its output is high, it hardly gives you anything at all. This is a very basic principle:

- Outputs from TTL logic chips are designed to sink current.
- They are not designed to source significant current.

In fact, the 74LS92 is rated to deliver less than half a milliamp. This is quite acceptable when you're just connecting it with another logic chip, but if you want to drive an external device, it doesn't provide much to work with.

The proper solution is to say to the chip, "All right, we'll do it your way," and set things up with a positive source that flows through a load resistor to the LED that you want to use, and from there into the output from the chip. This is the "better" option shown in Figure 4-104.

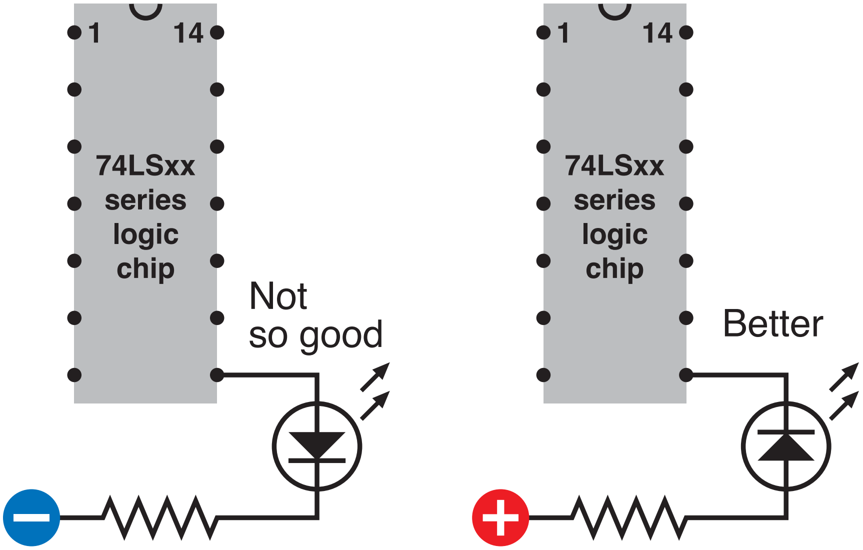

Figure 4-104. *Most TTL chips, including those in the LS generation, are unable to source much current from their logical output pins (left) and should usually be wired to sink current from a positive source (right).*

The only problem is that now the LED lights up when the counter's output is low. But the counter is designed to display its output in high pulses. So your LED is now off when it should be on, and on when it should be off.

You can fix this by passing the signal through an inverter, but already I'm getting impatient with this inconvenience. My way around the problem, at least for demo purposes, is to use the "Not so good" option in Figure 4-104 and make it work by connecting a very-low-current LED with a large 4K7 load resistor. This will enable us to "see" the output from the counter without asking it to give more than its rated limit, and if you want to create a more visibly powerful display for a finished version of the dice circuit, I'll deal with that later. According to my meter, the 4K7 resistor holds the current between 0.3mA and 0.4mA, which is the counter's rated maximum.

Set up your initial version of the circuit as shown in Figures 4-102 and 4-103. Be careful when you wire the positive and negative power supply to the counter chip, with its nonstandard pin assignments.

The 555 will run in astable mode, at about 1 pulse per second. This becomes the clock signal for the counter. The first three binary outputs from the counter then drive the three LEDs.

The counter advances when the input signal goes from *high* to *low*. So when the LED beside the 555 timer goes out, that's when the counter advances.

If you stare at the pattern generated by the outputs for long enough, you may be able to see the logic to it, bearing in mind that its zero state is when they are all off, and it counts up through five more steps before it repeats. The diagram in Figure 4-105 shows this sequence. If you want to know why the pattern works this way, check the following section, "Theory: Binary arithmetic."

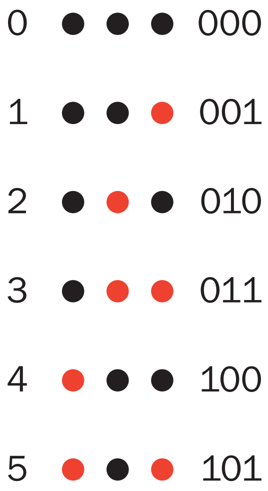

Figure 4-105. *The three output pins of the 74LS92 counter have high states shown by the red circles as the counter steps from 000 to 101 in binary notation.*

THEORY

Binary arithmetic

The rule for binary counting is just a variation of the rule that we normally use for everyday counting, probably without thinking much about it. In a 10-based system, we count from 0 to 9, then carry 1 over to the next position on the left, and go from 0 to 9 again in the right-most position. We repeat this procedure until we get to 99, then carry a 1 over to a new position to make 100, and continue counting.

In binary we do the same thing, except that we restrict ourselves to digits 0 and 1 only. So begin with 0 in the rightmost position, and count up to 1. As 1 is our limit, to continue counting we carry 1 over to the next place on the left, and start again from 0 in the right-most position. Count up to 1, then add 1 to the next place on the left—but, it already has a 1 in it, so it can't count any higher. So, carry 1 from there one space further, to the next place beside that—and so on.

If a glowing LED represents a 1, and a dark LED represents a 0, the diagram in Figure 4-105 shows how the 74LS92 counts up from 0 to (decimal) 6 or (binary) 101 in its inimitable fashion. I've also included a diagram in Figure 4-106 showing how a counter with four binary outputs would display decimal numbers from 0 through 15, again using the LEDs to represent 1s and 0s.

Here's a question for you: how many LEDs would you need to represent the decimal number 1024 in binary? And how many for 1023?

Obviously binary code is ideally suited to a machine full of logic components that either have a high or a low state. So it is that all digital computers use binary arithmetic (which they convert to decimal, just to please us).

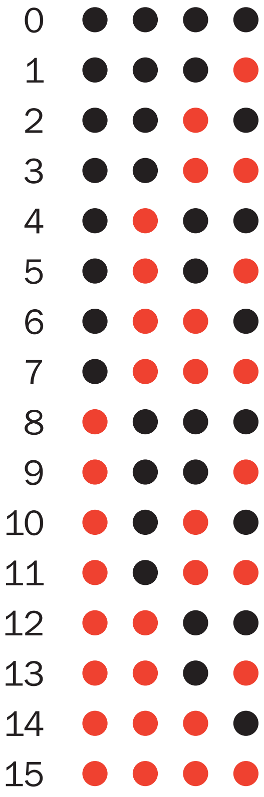

Figure 4-106. *A hexadecimal (16-based) binary counter would generate this succession of high states from its four output pins as it counts from 0 through 15 in decimal notation.*

Getting back to our project: I want to take the three binary outputs and make them create patterns like the spots on a die. How can I do this? Quite easily, as it turns out.

I'm assuming that I'll use seven LEDs to simulate the patterns of spots on a die. These patterns can be broken down into groups, which I have assigned to the three outputs from the counter in Figure 4-108. The first output (farthest to the right) can drive an LED representing the dot at the center of the die face. The second (middle) output can drive two more diagonal LEDs. The third output must switch on all four corner LEDs.

This will work for patterns 1 through 5, but won't display the die pattern for a 6. Suppose I tap into all three outputs from the counter with a three-input NOR gate. It has an output that goes high only when all three of its inputs are low, so it will only give a high output when the counter is beginning with all-low outputs. I can take advantage of this to make a 6 pattern.

Note that it's bad practice to mix the LS generation of TTL chips with the HC generation of CMOS chips, as their input and output ranges are different; so, the NOR chip has to be a 74LS27, not a 74HC27.

We're ready now for a simple schematic. In Figure 4-107 I've colored some of the wires just to make it easier for you to distinguish them. The colors have no other significance.

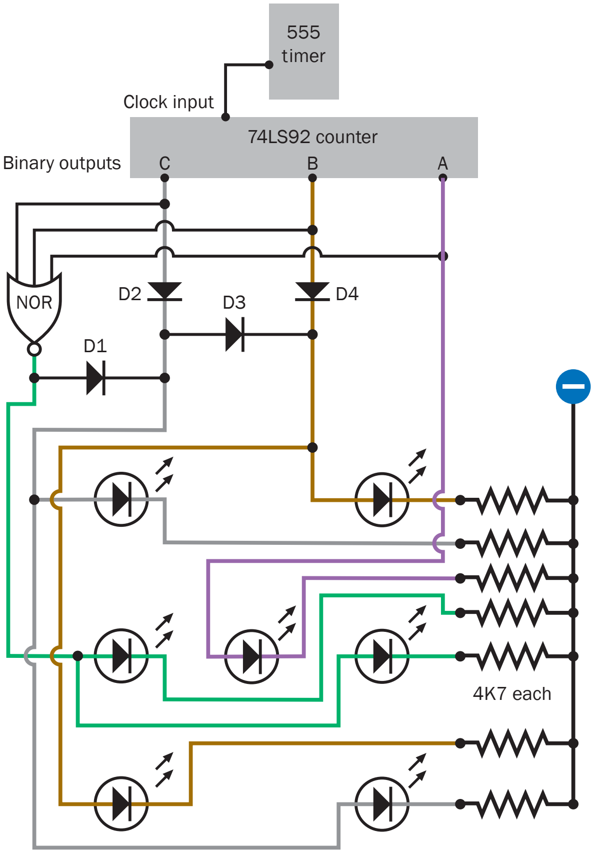
555
timer
Clock input
74LS92 counter
Binary outputs
C
B
A
NOR
D2
D3
D4
D1
4K7 each

Figure 4-107. *A simplified schematic shows how outputs from the 74LS92 counter can be combined, with signal diodes and a single three-input NOR gate, to generate the spot patterns on a die. The wire colors have no special meaning and are used just to distinguish them from each other.*

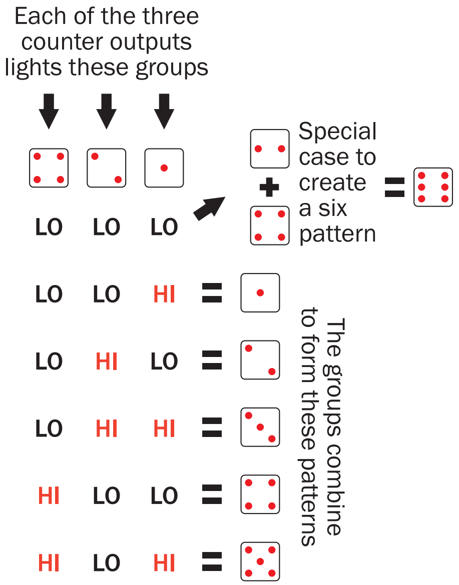
Each of the three
counter outputs
lights these groups
LO LO LO
Special
case to
create
a six
pattern
LO LO HI
LO HI LO
LO HI HI
HI LO LO
HI LO HI
The groups combine
to form these patterns

Figure 4-108. *Binary outputs from the 74LS92 counter can be used to power LEDs arrayed in groups to simulate the pattern of spots on a die.*

Each of the LEDs is grounded through a separate 4K7 load resistor. Unfortunately, this means that when they are displaying the pattern for a 6, all of them are running in parallel from the output of the NOR gate, which overloads it. As long as you don't leave the display in this mode for very long periods, it shouldn't cause a problem. You could compensate by increasing the load resistors, or by running pairs of the LEDs through one resistor, but this will make them so dim that they'll be difficult to see, as they're so close to their lower limit for current already.

Notice how I have added four signal diodes, D1 through D4. When Output C goes high, it has to illuminate all four corner LEDs, and so its power goes into the brown wire as well as the gray wire. But we must never allow one output to feed back into another, so D4 is needed to protect Output B when Output C is high.

Because there is now a connection between B and C, we need D2 to protect Output C when Output B is high. And because Output B must only feed two of the corner LEDs, we also need D3 to stop it from illuminating the other two. And, we have to protect the output from the NOR gate when either Output C or Output B is high. This requires D1.

Figure 4-109 shows everything that I've described so far assembled in breadboard format, while Figure 4-110 shows the test version that I built. Note that the unused logical inputs on the 74LS27 chip are shorted together and connected to the *positive* side of the power supply. Here's the rule:

- When using CMOS chips (such as the HC series), connect unused logical inputs to the negative side of the power supply.
- When using TTL chips (such as the LS series), connect unused logical inputs to the positive side of the power supply.

I assume that you have had enough fun watching the LEDs count slowly, so I've changed the capacitor and resistor values for the 555 to increase its speed from approximately 1 pulse per second to about 50,000 pulses per second. The counter could run much faster than this, but I just want it to cycle fast enough so that when the user presses and releases a button, the count will stop at an unforeseeable number.

The button starts and stops the 555 timer by applying and releasing power to the timing circuit only. This is the equivalent of shaking and then throwing the die.

While the counter is running fast, the LEDs are flashing so fast that all of them will seem to be on at once. At the same time, the circuit charges a new 68 µF capacitor, which I have added between the pushbutton and ground. When you release the button, this capacitor discharges itself through the 1K timing resistor. As the charge dissipates, the timing capacitor will take longer and longer to charge, and discharge, and the frequency of the 555 will gradually diminish. Consequently the LED display will also flash slower, like the reel on a Las Vegas slot machine gradually coming to a stop. This increases the tension as players can see the die display counting to the number that they're hoping for—and maybe going one step beyond it.

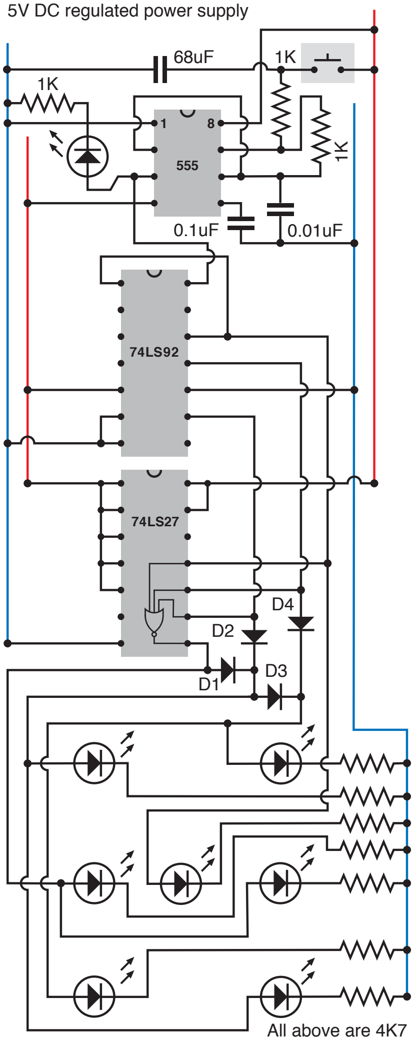

Figure 4-109. *With some extra components, the schematics from Figures 4-102 and 4-107 can be combined to make the working dice simulation.*

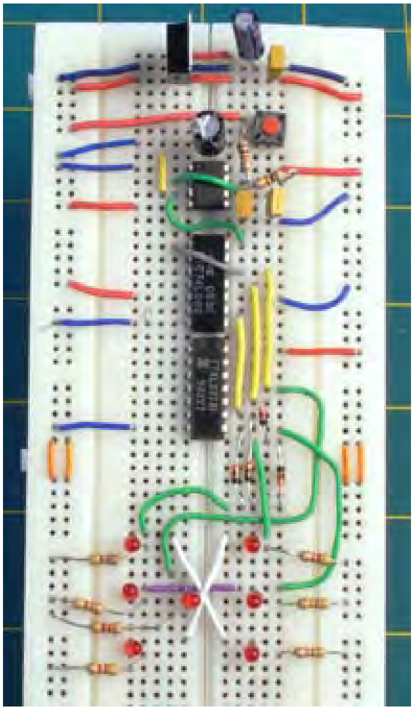

Figure 4-110. *The electronic dice schematic applied to a breadboard, with a pushbutton at the top to start and stop the counter, and 7 LEDs at the bottom to display the output.*

Note that to maximize this effect, the button has to be held down for a full second or more, so that the 68 µF capacitor becomes fully charged before the button is released.

So, this circuit now fulfills the original goal. But can it be better? Of course it can.

Enhancements

The main thing I want to improve is the brightness of the LEDs. I could add a transistor to amplify the current to each one, but there's a simpler alternative: a TTL "open collector" inverter.

I want to use an inverter because in the world of TTL, as I mentioned earlier, we can sink much more power into the output pin of a chip than we can source from it. So, I'm going to turn each LED the other way around and connect their load resistors to the positive side of the power supply. This way, they'll sink their power into the outputs of the inverter.

And the great advantage of an "open collector" version of the inverter chip is that it is designed to sink much more current than a normal TTL logic chip. It is rated for 40mA per pin. The only disadvantage is that it cannot source any current at all; instead of its output going high, it just behaves like an open switch. But that's OK for this circuit.

So the next and final schematic, in Figure 4-111, includes the 74LS06 inverter, which has also been added to the breadboarded version shown in Figure 4-112. I suggest that you put aside the little low-current LEDs and substitute some normal-size ones. Using Kingbright "standard" WP15031D 5mm LEDs, I find that each draws almost exactly 20mA with a voltage drop of about 2V with a 120 ohm series resistor. Because each output pin from the 74LS06 inverter powers no more than two LEDs at a time, this is exactly within its specification. I suggest that if you build this circuit, you check the consumption of your particular choice of LEDs and adjust the resistors if necessary.

Remember: to measure the voltage drop across an LED, simply touch the probes of your meter across it while it is illuminated. To measure the current, disconnect one side of the LED and insert the meter, in milliamp mode, between the leg of the LED and the contact that it normally makes in the circuit.

For a really dramatic display, you can get some 1 cm diameter LEDs (Figure 4-113). Check the specification, and you should find that many of these don't use more power than the usual 5 mm type. But whatever kind you use, don't forget to turn them around so that their negative sides face toward the inverter, and their positive sides face the resistors, which are connected to the positive side of the power supply.

One last detail: I had to add two 10K resistors to this version of the circuit. Can you see why? Diodes D1 through D4 are designed to transmit positive voltage through to the inverter when appropriate, but they prevent the inputs of the inverter from "seeing" the negative side of the power supply when the counter outputs are low. These inverter inputs require pull-down resistors to prevent them from "floating" and producing erroneous results.

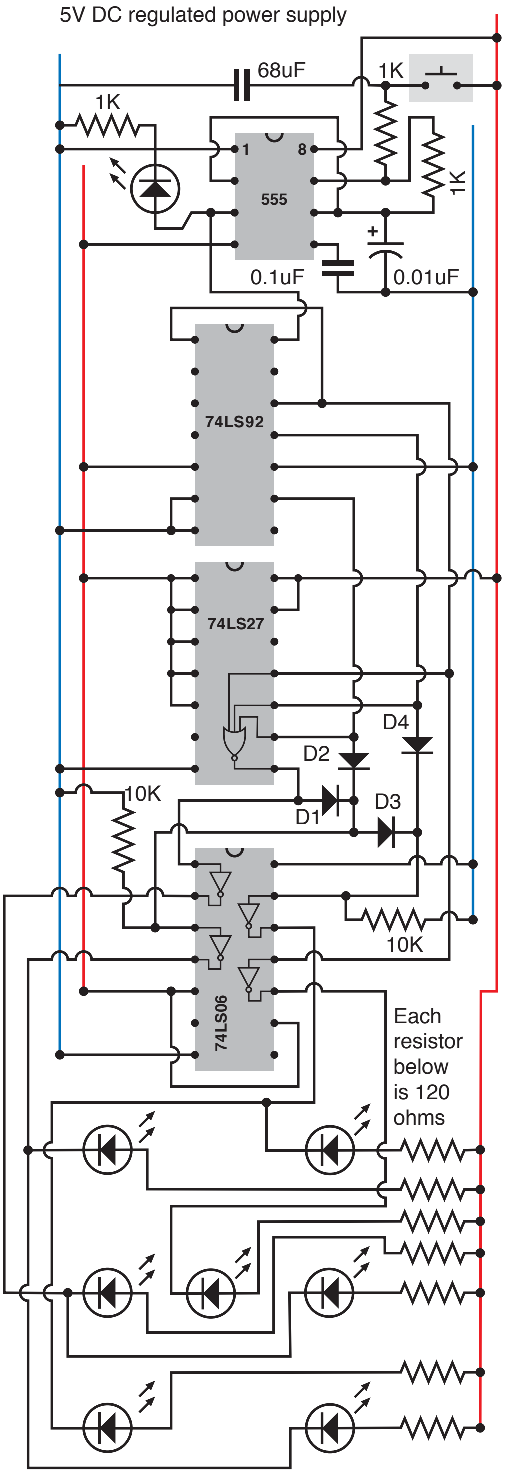
5V DC regulated power supply
68uF
1K
1K
1
8
555
1K
+
0.1uF
0.01uF
74LS92
74LS27
D4
D2
10K
D1
D3
74LS06
10K
Each resistor below is 120 ohms

Figure 4-111. *If open-collector inverters are added to the dice schematic, it can drive full-size LEDs with up to 40mA, as long as the LEDs are turned around to sink current into the TTL output stage instead of trying to source current from it.*

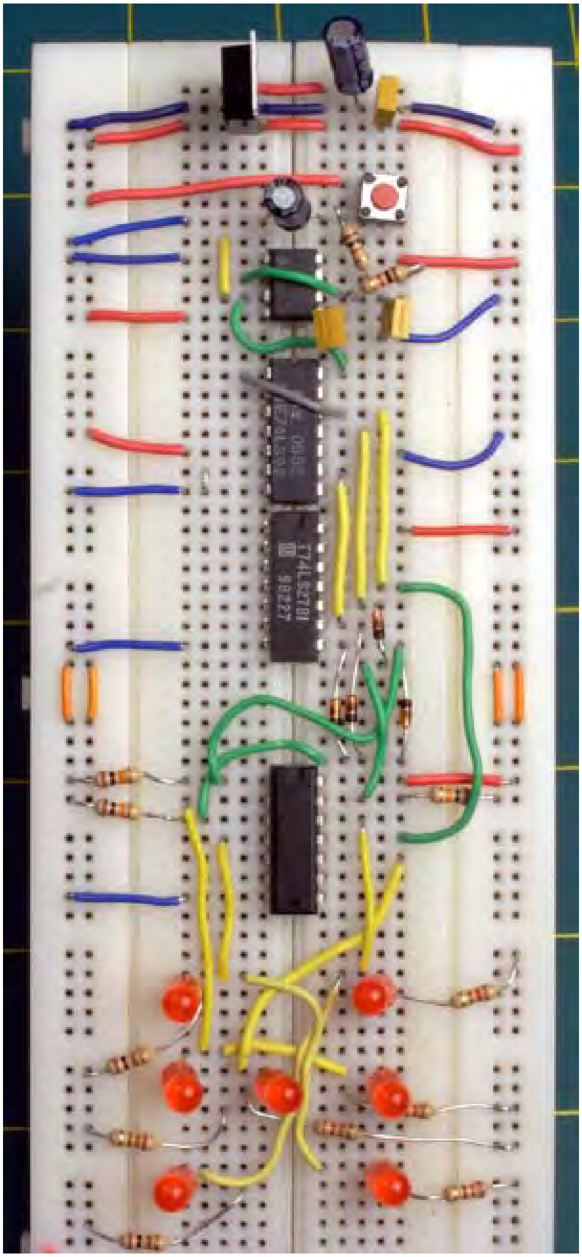

Figure 4-112. *The completed circuit using an open-collector inverter to drive full-size LEDs.*

The final enhancements are up to you. Most obviously, you can add a second die, as many games require two dice. The 74LS27 chip still has a couple of spare NOR gates in it, one of which you can make use of, but you will need an additional 555 timer, running at a significantly different speed to ensure randomness, and it will have to drive a second counter.

After you get your dice up and running, you may want to test them for randomness. Because the pulses from a 555 timer are of equal length, every number has an equal chance of coming up; but the longer you hold down the Start button, the better your odds are of interrupting the counting process at a truly random moment. Anyone using your electronic dice should be told that "shaking" them for a full second is mandatory.

Of course, I could have simulated dice more easily by writing a few lines of software to generate random numbers on a screen, but even a fancy screen image cannot have the same appeal as a well-made piece of hardware. Figure 4-113 shows white 1 cm LEDs mounted in a sanded polycarbonate enclosure for dramatic effect.

Most of all, I derived satisfaction from using simple, dedicated chips that demonstrate the binary arithmetic that is fundamental in every computer.

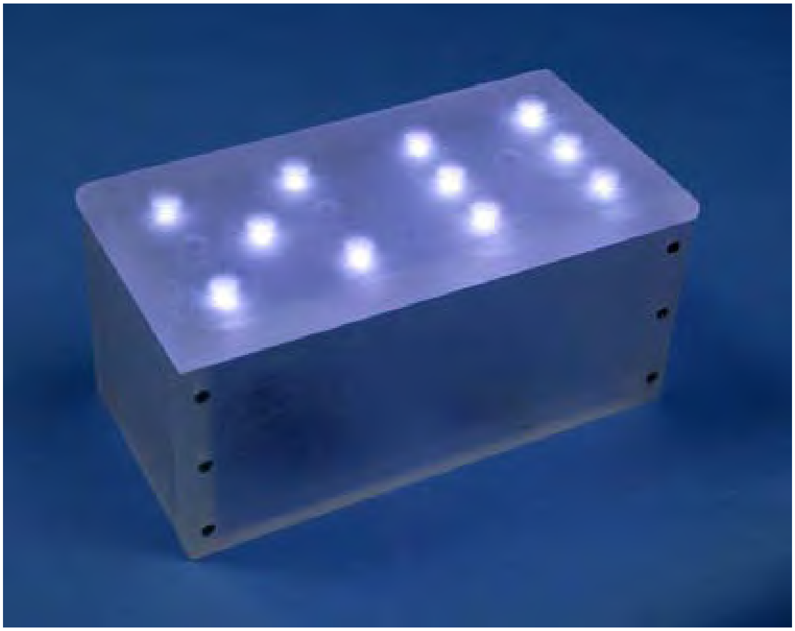

Figure 4-113. *The open-collector inverter chip in the dice circuit is sufficiently powerful to drive 1-cm white LEDs that draw about 20mA each, using a potential of 2V. In this finished version, the LEDs were embedded in cavities drilled from the underside of half-inch polycarbonate, which has been treated with an orbital sander to create a translucent finish.*

Experiment 24: Intrusion Alarm Completed

Now let me suggest how you can apply the knowledge from this chapter of the book to upgrade the burglar alarm project that was last modified in Experiment 15. You'll probably need to check Chapters 2 and 3 to refamiliarize yourself with some features of the alarm.

Upgrade 1: Delayed Activation

The biggest flaw in the alarm was that as soon as it was activated, it would immediately respond to any signal from the door and window sensors. It needed a feature to delay activation to give you a chance to exit from the building before the alarm armed itself. A 555 timer can provide this functionality, probably in conjunction with a relay. The power to the alarm should pass through the contacts of the relay, which are normally closed. When you press a button on the timer, it sends a positive pulse to the relay lasting for around 30 seconds, holding the relay open for that period. You could mount the timer in its own little box with a button on it, which you press when you're ready to leave the building. The 12-volt power supply to the burglar alarm passes through the box containing the delay circuit. For 30 seconds, the 555 interrupts power to the alarm, and then restores it, ready for action.

Upgrade 2: Keypad Deactivation

This is now really simple. You can substitute a latching relay instead of the switch, S1, on the alarm box (shown in Figure 3-110), and use a keypad to set and reset the relay in exactly the same way as in the combination lock in Experiment 20. You'll have to run an additional three wires from the relay, out of the alarm box, to the keypad (one supplying power to the "on" relay coil, another supplying power to the "off" coil, the third being a common ground). You can either use a 9-volt battery to power the electronics in association with the keypad, or run an additional fourth wire from the alarm box, to carry positive power to the logic chips, bearing in mind that you have to insert a voltage regulator at some point, to drop the 12 volts that the alarm uses to the 5 volts that the logic gates require. As the gates consume so little power, the 12-to-5 drop should be OK for the voltage regulator; it won't have to dissipate too much heat.

With this additional feature, you can use the alarm like this:

- Press the pound key on the keypad to flip the latching relay into its "on" mode, so that it passes power to the alarm box, which is now armed.
- If you want to leave the house, push the button on the delay unit to give you 30 seconds in which to do so.
- If the alarm is triggered, enter your secret code on the keypad to deactivate it by flipping the latching relay to its "off" position and cutting power to the alarm box.

These modifications are so simple that I think the block diagram in Figure 4-114 should be all you need. I don't think I need to give you any schematics. The only change you have to make to the existing alarm is to substitute the latching relay for the on/off switch.

But, there is still one obvious necessary enhancement needed: how can you get back into the house without instantly triggering the alarm?

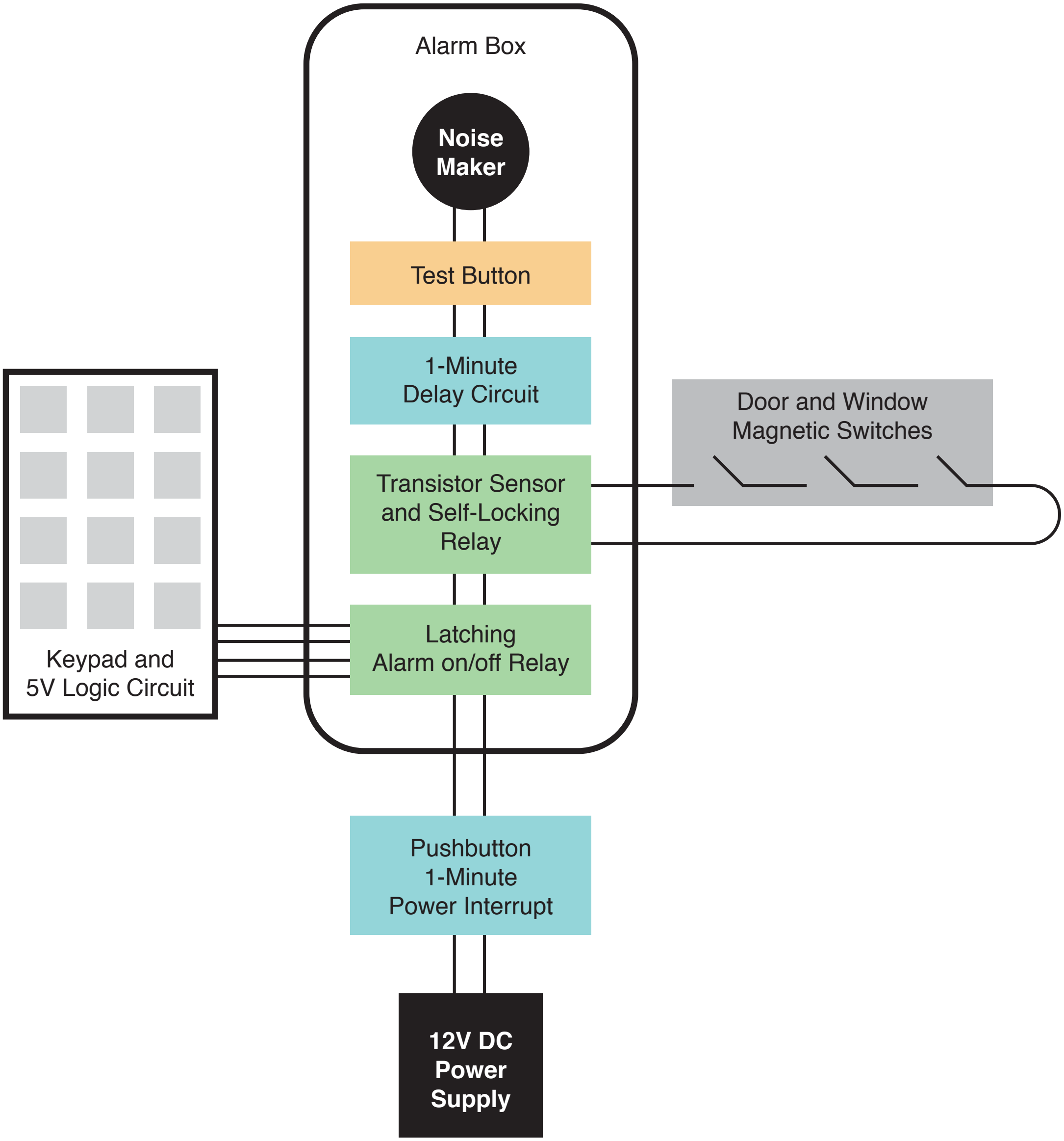
Alarm Box
Noise Maker
Test Button
1-Minute Delay Circuit
Transistor Sensor and Self-Locking Relay
Latching Alarm on/off Relay
Door and Window Magnetic Switches
Keypad and 5V Logic Circuit
Pushbutton 1-Minute Power Interrupt
12V DC Power Supply

Figure 4-114. *This block diagram shows the relative placement of the old and new components. The pushbutton power interrupt (which allows you to leave the house before the alarm switches itself back on) goes between the power supply and everything else. The latching relay substitutes for the DPDT switch on the previous version of the alarm. The transistor and self-locking relay, connected with the door and window magnetic switches, remain unchanged. The new delay circuit is inserted between the self-locking relay and the noisemaker. The test button is wired with the latching relay in the same way that it was wired previously with the DPDT switch.*

Upgrade 3: Delay Before Deactivation

Typically, alarms include another delay feature. When you open a door on your way into the building and it triggers the alarm, you have 30 seconds to deactivate it, before it starts making a noise.

How can we implement this delay feature? If I try to use another 555 timer to generate a pulse to inhibit the noise, that won't work, because the output from either the transistor or the relay can continue indefinitely. The relay locks itself on, and the transistor can continue passing voltage for as long as someone leaves a door open. If either of these signals activates a timer in monostable mode, the pulse from the timer will never end, and it will suppress the alarm indefinitely.

I think what I have to do is use a resistor and a capacitor to create a delay. I'll power them through the existing relay, so that I can be sure that they'll receive the full voltage of the power supply, after beginning from zero. Gradually the capacitor will acquire voltage—but I can't connect this directly to the noisemaker, because the noisemaker will gradually get louder as the voltage increases.

I have to insert a device that will be triggered to give full voltage when the input rises past a certain point. To do this, I'll use a 555 timer that's wired in bistable mode. This kind of modification is generally known as a "kludge," because it's not elegant, uses too many components, and does not use them appropriately. What I really need is a comparator, but I don't have space to get into that topic. So, using the knowledge that you have so far, the schematic in Figure 4-115 shows how a delay can be added to the alarm—not elegantly, but reliably.

The only problem is that if you power up a 555 timer in bistable mode, there's a 50-50 chance that the timer starts itself with its output high or low. So I need to pull the voltage low on the reset pin (to start the timer with its output inhibited) and gradually let it become positive (to permit the output). At the same time, I want to start with the voltage high on the trigger pin and gradually lower it, until it falls below 1/3 of the power supply and triggers the output.

So there are two timing circuits. The one for the reset pin works faster than the one on the trigger pin, so that at the point when the timer is triggered, the reset won't stop it.

The schematic shows component values that will do this. The 10 µF capacitor starts low but is charged through the 10K resistor in a couple of seconds. The timer is now ready to be triggered. But the 68 µF capacitor starts high (being connected with the positive side of the power supply) and takes a full minute to be pulled down to 1/3 of supply voltage through the 1M resistor. At that point, its voltage is low enough to trigger the 555. The timer output goes high and supplies the noisemaker.

You should be able to insert this little delay module in your alarm box, between the output from the relay and the input to the noisemaker without too much trouble. And if you want to adjust the delay, just use a higher or lower value resistor than 1M.

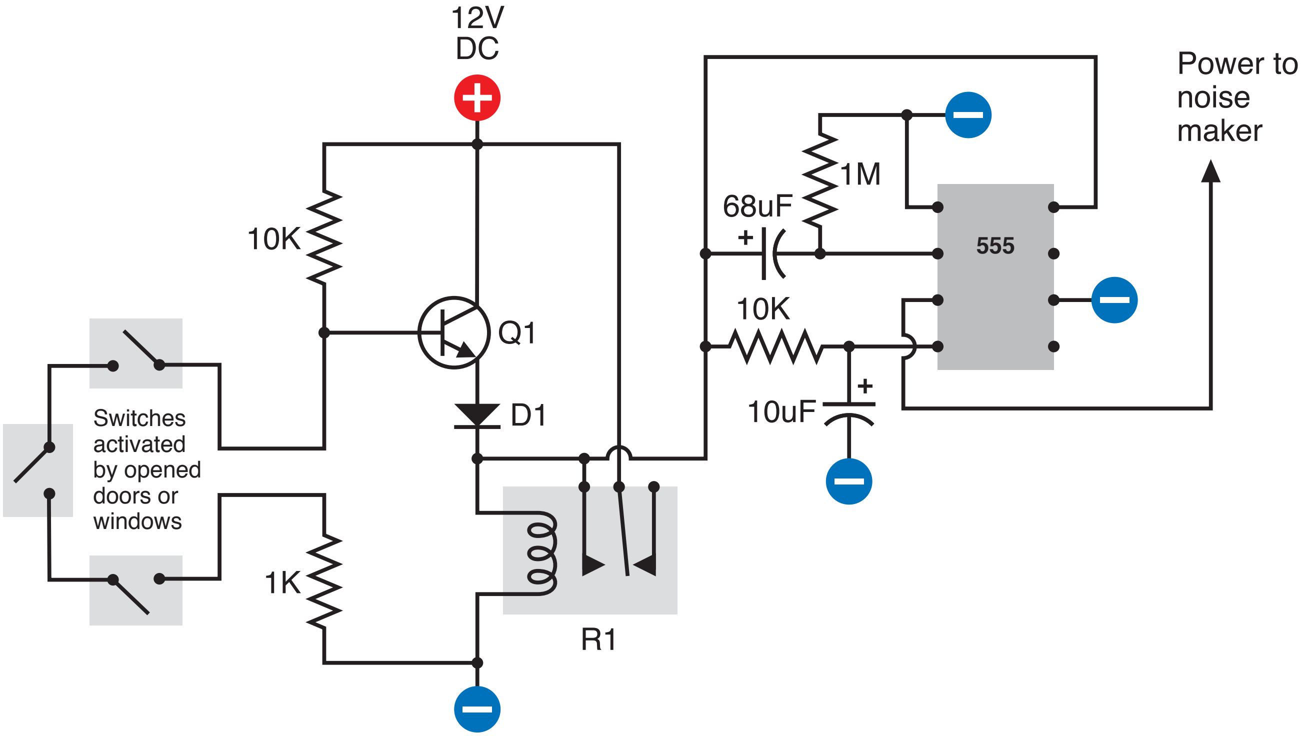

Figure 4-115. *This addition to the original alarm circuit imposes a one-minute delay before the alarm sounds. The 555 timer (wired in bistable mode) receives power through relay R1. The lower timing circuit initially applies negative voltage to the reset, ensuring that the 555 powers up with its output suppressed. This voltage quickly rises. Meanwhile the upper timing circuit applies a voltage to the trigger that gradually diminishes as the 68 µF capacitor equalizes its charge through the 1M resistor. When the voltage diminishes to 1/3 of the supply, the timer's output goes high and starts the noisemaker. If the power to the circuit is interrupted at any time before this, the relay relaxes, the capacitors gradually discharge, and the alarm does not sound.*

The Wrap-Up

If you add these three enhancements, your alarm will have all the features on my original wish list. Of course, if I were designing it from scratch, with all the information that has been added in this chapter of the book, it could be more elegant. But the modifications have not entailed making destructive changes to our original project, and all the design goals have been met.

5

What Next?

IN THIS CHAPTER

Shopping List: Experiments 25 Through 36
Customizing Your Work Area
Reference Sources
Experiment 25: Magnetism
Experiment 26: Tabletop Power Generation
Experiment 27: Loudspeaker Destruction
Experiment 28: Making a Coil React
Experiment 29: Filtering Frequencies
Experiment 30: Fuzz
Experiment 31: One Radio, No Solder, No Power
Experiment 32: A Little Robot Cart
Experiment 33: Moving in Steps
Experiment 34: Hardware Meets Software
Experiment 35: Checking the Real World
Experiment 36: The Lock, Revisited
In Closing

At this point, we can branch out in numerous directions. Here are some possibilities:

Audio electronics

This is a field in itself, including hobby projects, such as simple amplifiers and "stomp boxes," to modify guitar sound.

Radio-frequency devices

Anything that receives or transmits radio waves, from an ultra-simple AM radio to remote controllers.

Motors

The field of robotics has encouraged the growth of many online sites selling stepper motors, gear motors, synchronous motors, servo motors, and more.

Programmable microcontrollers

These are tiny computers on a single chip. You write a little program on your desktop computer, which will tell the chip to follow a series of procedures, such as receiving input from a sensor, waiting for a fixed period, and sending output to a motor. Then you download your program onto the chip, which stores it in nonvolatile memory. Popular controllers include the PICAXE, BASIC Stamp, Arduino, and many more. The cheapest ones retail for a mere $5 each.

Obviously, I don't have space to develop all of these topics fully, so what I'm going to do is introduce you to them by describing just one or two projects in each category. You can decide which interests you the most, and then proceed beyond this book by reading other guides that specialize in that interest.

I'm also going to make some suggestions about setting up a productive work area, reading relevant books, catalogs, and other printed sources, and generally proceeding further into hobby electronics.

Shopping List: Experiments 25 Through 36

Tools

You won't need any new tools for this section of the book.

Supplies and Components

As we have progressed to the point where you may want to pick and choose which projects you attempt, I will list the supplies and components at the beginning of each experiment.

Customizing Your Work Area

At this point, if you're getting hooked on the fun of creating hardware but haven't allocated a permanent corner to your new hobby, I have some suggestions. Having tried many different options over the years, my main piece of advice is this: don't build a workbench!

Many hobby electronics books want you to go shopping for 2×4s and plywood, as if a workbench has to be custom-fabricated to satisfy strict criteria about size and shape. I find this puzzling. To me, the exact size and shape of a bench is not very important. I think the most important issue is storage.

I want tools and parts to be easily accessible, whether they're tiny transistors or big spools of wire. I certainly don't want to go digging around on shelves that require me to get up and walk across the room.

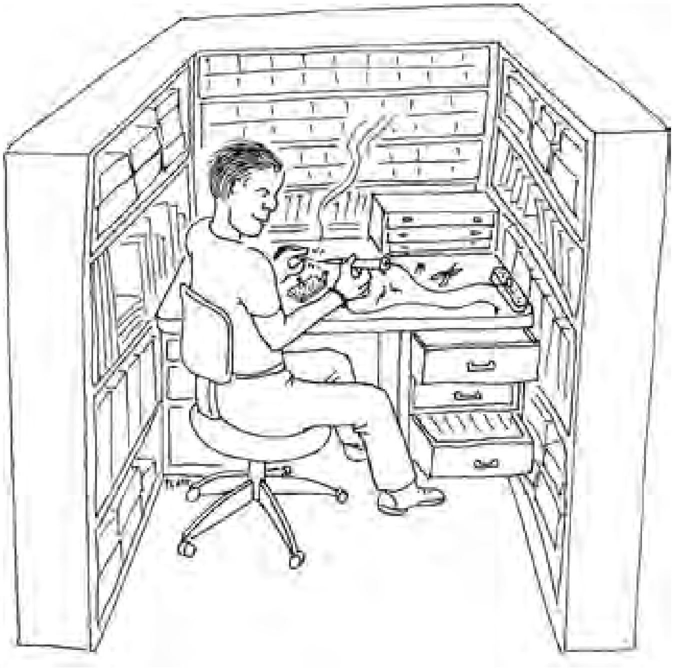

Figure 5-1. *The ideal work area: surrounded by storage. Never again will you need to get out of your chair!*

This leads me to two conclusions:

1. You need storage above the workbench.
2. You need storage below the workbench.

Many DIY workbench projects allow little or no storage underneath. Or, they suggest open shelves, which will be vulnerable to dust. My minimum configuration would be a pair of two-drawer file cabinets with a slab of 3/4-inch plywood or a Formica-clad kitchen countertop placed across them. File cabinets are ideal for storing all kinds of objects, not just files.

Of all the workbenches I've used, the one I liked best was an old-fashioned steel office desk—the kind of monster that dates back to the 1950s. They're difficult to move (because of their weight) and don't look beautiful, but you can buy them cheaply from used office furniture dealers, they're generous in size, they withstand abuse, and they last forever. The drawers are deep and usually slide in and out smoothly, like good file-cabinet drawers. Best of all, the desk has so much steel in it that you can use it to ground yourself before touching components that are sensitive to static electricity. If you use an antistatic wrist strap, you can simply attach it to a sheet-metal screw that you drive into one corner of the desk.

What will you put in the deep drawers of your desk or file cabinets? Some paperwork may be useful, perhaps including the following documents:

- Product data sheets
- Parts catalogs
- Sketches and plans that you draw yourself

The remaining capacity of each drawer can be filled with plastic storage boxes. The boxes can contain tools that you don't use so often (such as a heat gun or a high-capacity soldering iron), and larger-sized components (such as loudspeakers, AC adapters, project boxes, and circuit boards). You should look for storage boxes that measure around 11 inches long, 8 inches wide, and 5 inches deep, with straight sides. Boxes that you can buy at Wal-Mart will be cheaper, but they often have tapering sides (which are not space-efficient).

The boxes that I like best are Akro-Grids, made by Akro-Mils (see Figures 5-2 and 5-3). These are very rugged, straight-sided, with optional transparent snap-on lids. You can download the full Akro-Mills catalog from *http://www.akro-mils.com* and then search online for retail suppliers. You'll find that Akro-Mils also sells an incredible variety of parts bins, but I don't like open bins because their contents are vulnerable to dust and dirt.

For medium-size components, such as potentiometers, power connectors, control knobs, and toggle switches, I like storage containers measuring about 11 inches long, 8 inches wide, and 2 inches deep , divided into four to six sections. You can buy these from Michaels (the craft store), but I prefer to shop online for the Plano brand, as they seem more durably constructed. The Plano products that are most suitable for medium-size electronic parts are classified as fishing-tackle boxes, and you'll see them at *http://www.planomolding.com/tackle/products.asp*.

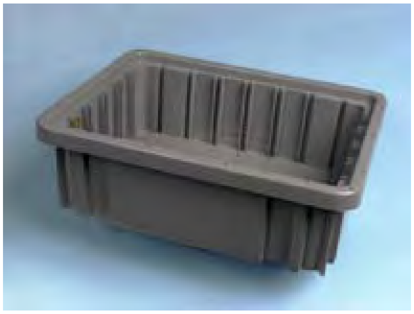

Figure 5-2. *Akro-Grid boxes contain grooves allowing them to be partitioned into numerous compartments for convenient parts storage.*

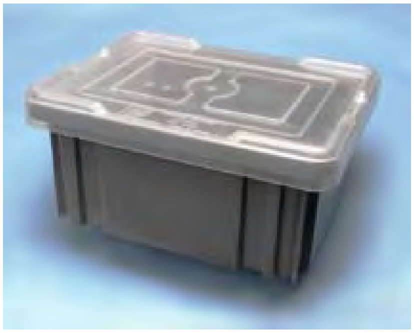

Figure 5-3. *Lids are sold separately for Akro-Grid boxes to keep the contents dust-free. The height of the box in Figure 5-2 allows three to be stacked in a typical file-cabinet drawer. The box shown here allows two to be stacked.*

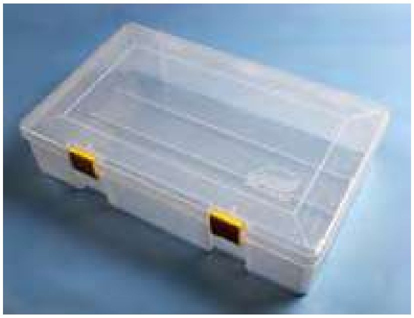

Figure 5-4. *This Plano brand box is undivided, making it useful for storing spools of wire or medium-size tools. When stacked upright on its long edge, three will fit precisely in a file-cabinet drawer.*

For undivided, flat-format storage boxes, the Prolatch 23600-00 is ideally sized to fit a file-cabinet drawer, and the latches are sufficiently secure that you could stack a series of them on their long edges. See Figure 5-4.

Plano also sells some really nicely designed toolboxes, one of which you can place on your desktop. It will have small drawers for easy access to screwdrivers, pliers, and other basics. Because you need a work area that's only about three feet square for most electronics projects, surrendering some desk space to a toolbox is not a big sacrifice.

If you have a steel desk with relatively shallow drawers, one of them can be allocated for printed catalogs. Don't underrate the usefulness of hard copy, just because you can buy everything online. The Mouser catalog, for instance, has an index, which is better in some ways than their online search feature, and the catalog is divided into helpful categories. Many times I've found useful parts that I never knew existed, just by browsing, which is much quicker than flipping through PDF pages online, even with a broadband connection. Currently, Mouser is still quite generous about sending out their catalogs, which contain over 2,000 pages. McMaster-Carr will also send you a catalog, but only after you've ordered from them, and only once a year.

Now, the big question: how to store all those dinky little parts, such as resistors, capacitors, and chips? I've tried various solutions to this problem. The most obvious is to buy a case of small drawers, each of which is removable, so you can place it on your desk while you access its contents. But I don't like this system, for two reasons. First, for very small components, you need to subdivide the drawers, and the dividers are never secure. And second, the removability of the drawers creates the risk of accidentally emptying the contents on the floor. Maybe you're too careful to allow this to happen, but I'm not!

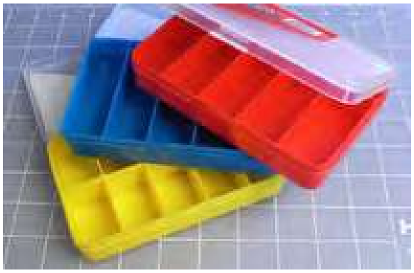

Figure 5-5. *Darice Mini-Storage boxes are ideal for components such as resistors, capacitors, and semiconductors. The boxes can be stacked stably or stored on shelves, with their ends labeled. The brand sticker is easily removed after being warmed with a heat gun.*

My personal preference is to use Darice Mini-Storage boxes, shown in Figure 5-5. You can find these at Michaels in small quantities, or buy them more economically in bulk online from suppliers such as *http://www.craftamerica.com*. The blue boxes are subdivided into five compartments that are exactly the right size and shape for resistors. The yellow boxes are subdivided into ten compartments, which are ideal for semiconductors. The purple boxes aren't divided at all, and the red boxes have a mix of divisions.

The dividers are molded into the boxes, so you don't have the annoyance associated with removable dividers that slip out of position, allowing components to mix together. The box lids fit tightly, so that even if you drop one of the boxes, it probably won't open. The lids have metal hinges, and a ridge around the edge that makes the boxes securely stackable.

I keep my little storage boxes on a set of shelves above the desk, with a gap of 3 inches between one shelf and the next, allowing two boxes to be stacked on each shelf. If I want to work with a particular subset of boxes, I shift them onto the desktop and stack them there.

Labeling

No matter which way you choose to store your parts, labeling them is essential. Any ink-jet printer will produce neat-looking labels, and if you use peelable (nonpermanent) labels, you'll be able to reorganize your parts in the future, as always seems to become necessary. I use color-coded labels for my collection of resistors, so that I can compare the stripes on a resistor with the code on the label, and see immediately if the resistor has been put in the wrong place. See Figure 5-6.

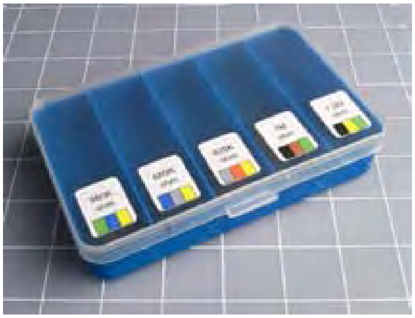

Figure 5-6. *To check that resistors are not placed in the wrong compartments, print the color code on each label.*

Even more important: you need to place a second (non-adhesive) label inside each compartment with the components. This label tells you the manufacturer's part number and the source, so that reordering is easy. I buy a lot of items from Mouser, and whenever I open their little plastic bags of parts, I snip out the section of the bag that has the identifying label on it, and slide it into the compartment of my parts box before I put the parts on top of it. This saves frustration later.

If I were *really* well organized, I would also keep a database on my computer listing everything that I buy, including the date, the source, the type of component, and the quantity. But I'm not that well organized.

On the Bench

Some items are so essential that they should sit on the bench or desktop on a permanent basis. These include your soldering iron(s), helping hands with magnifier, desk lamp, breadboard, power strip, and power supply. For a desk lamp, I prefer the type that has a daylight-spectrum fluorescent bulb, because it spreads a uniform light and helps me to distinguish colors of adjacent stripes on resistors.

The power supply is a matter of personal preference. If you're serious about electronics, you can buy a unit that delivers properly smoothed current at a variety of properly regulated and calibrated voltages. Your little wall-plug unit from RadioShack cannot do any of these things, and its output may vary depending on how heavily you load it. Still, as you've seen, it is sufficient for basic experiments, and when you're working with logic chips, you need to mount a 5-volt regulator on your breadboard anyway. Overall, I consider a good power supply optional.

Another optional item is an oscilloscope. This will show you, graphically, the electrical fluctuations inside your wires and components, and by applying probes at different points, you can track down errors in your circuit. It's a neat gadget to own, but it will cost a few hundred dollars, and for our tasks so far, it has not been necessary. If you plan to get seriously into audio circuits, an oscilloscope becomes far more important, because you'll want to see the shapes of the waveforms that you generate.

You can try to economize on an oscilloscope by buying a unit that plugs into the USB port of your computer and uses your computer monitor to display the signal. I have tried one of these, and was not entirely happy with the results. It worked, but did not seem accurate or reliable for low-frequency signals. Maybe I was unlucky; I decided not to try any other brands.

The surface of your desk or workbench will undoubtedly become scarred by random scuffs, cut marks, and drops of molten solder. I use a piece of half-inch plywood, two feet square, to protect my primary work area, and I clamp a miniature vise to its edge. To reduce the risk of static electricity when working with sensitive components, I cover the plywood with a square of conductive foam. This is not cheap, but offers advantages in addition to protecting chips from being zapped. Instead of scattering stray components, I can stick them into the foam, like plants growing in a garden. And like a garden, I can divide it into sections, with resistors on one side, capacitors on the other, and chips straight ahead.

Inevitably, during your work you'll create a mess. Little pieces of bent wire, stray screws, fasteners, and fragments of stripped insulation tend to accumulate, and can be a liability. If metal parts or fragments get into a project that you're building, they can cause short circuits. So you need a trash container. But it has to be easy to use. I use a full-size garbage pail, because it's so big that I can't miss it when I throw something toward it, and I can never forget that it's there.

Last, but most essential: a computer. Now that all data sheets are available online, and all components can be ordered online, and many sample circuits are placed online by hobbyists and educators, I don't think anyone can work efficiently without quick Internet access. To avoid wasting space, I suggest you use a small, cheap laptop that has a minimal footprint. A possible workbench configuration, using a steel desk, is shown in Figure 5-7.

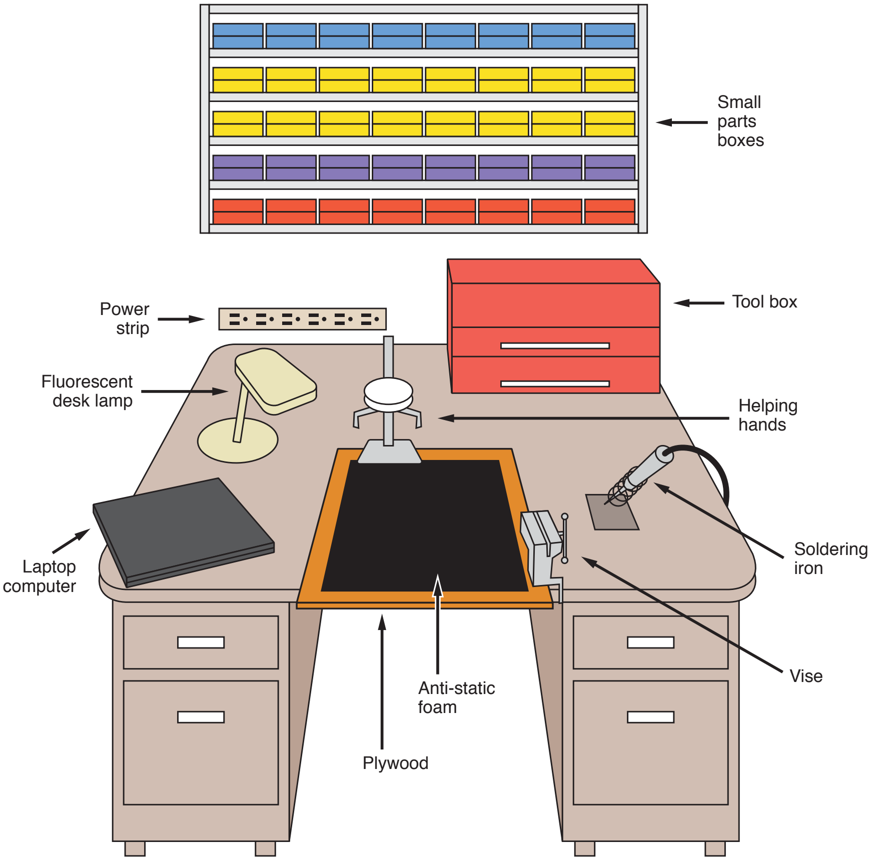

Figure 5-7. *An old steel office desk can be as good as, if not better than, a conventional workbench when building small electronics projects. It provides a large work area and ample storage, and has sufficient mass for you to ground yourself when dealing with components that are sensitive to static electricity.*

Reference Sources

Online

My favorite educational and reference site is Doctronics (*http://www.doctronics.co.uk)*. I like the way they draw their schematics, and I like the way they include many illustrations of circuits on breadboards (which most sites don't bother to do). They also sell kits, if you're willing to pay and wait for shipping from the UK. Part of a page from the doctronics website is reproduced in Figure 5-8.

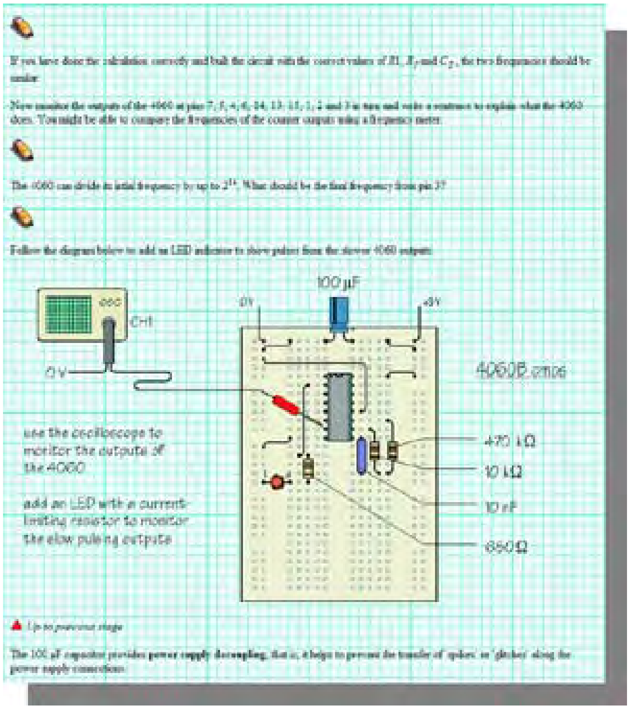
100 µF
0V
+9V
OSC
CH1
0V
4060B cmos
use the oscilloscope to monitor the outputs of the 4060
add an LED with a current-limiting resistor to monitor the slow pulsing outputs
470 kΩ
10 kΩ
10 nF
680Ω
Up to previous stage

Figure 5-8. *A sample page from http://www.doctronics.co.uk shows their detailed instructional approach. This is a valuable free online resource.*

My next favorite hobby site is also British-based: the Electronics Club (*http://www.kpsec.freeuk.com)*. It's not as comprehensive as Doctronics, but very friendly and easy to understand.

For a more theory-based approach, try *http://www.electronics-tutorials.ws*. This will go a little farther than the theory sections I've included here.

For an idiosyncratic selection of electronics topics, try Don Lancaster's Guru's Lair (*http://www.tinaja.com*). Lancaster wrote *The TTL Cookbook* more than 30 years ago, which opened up electronics to at least two generations of hobbyists and experimenters. He knows what he's talking about, and isn't afraid of getting into some fairly ambitious areas such as writing his own PostScript drivers and creating his own serial-port connections. You'll find a lot of ideas there.

Books

Yes, you do need books. As you're already reading this one, I won't recommend other beginners' guides. Instead, in keeping with the orientation of this chapter, I'll suggest some titles that will take you farther in various directions, and can be used for reference. I own all of these myself, and find them valuable:

Practical Electronics for Inventors, by Paul Scherz (McGraw-Hill, Second Edition, 2007)

> This is a massive, comprehensive book, well worth the $40 cover price. Despite its title, you won't need to invent anything to find it useful. It's my primary reference source, covering a wide range of concepts, from the basic properties of resistors and capacitors all the way to some fairly high-end math. If you buy only one book (in addition to this one, of course!), this would be my recommendation.

Getting Started with Arduino, by Massimo Banzi (Make: Books, 2009)

> If you enjoy the simplicity and convenience of the PICAXE programmable microcontroller that I describe later in this chapter, you'll find that the Arduino can do a lot more. *Getting Started* is the simplest introduction around, and will help to familiarize you with the Processing language used in Arduino (similar to the C language, not much like the version of BASIC used by the PICAXE).

Making Things Talk, by Tom Igoe (Make: Books, 2007)

> This ambitious and comprehensive volume shows how to make the most of the Arduino's ability to communicate with its environment, even getting it to access sites on the Internet.

TTL Cookbook, by Don Lancaster (Howard W. Sams & Co, 1974)

> The 1974 copyright date is not a misprint! You may be able to find some later editions, but whichever one you buy, it will be secondhand and possibly expensive, as this title now has collectible value. Lancaster wrote his guide before the 7400 series of chips was emulated on a pin-for-pin basis by CMOS versions, but it's still a good reference, because the concepts and part numbers haven't changed, and his writing is so accurate and concise.

CMOS Sourcebook, by Newton C. Braga (Sams Technical Publishing, 2001)

> This book is entirely devoted to the 4000 series of CMOS chips, not the 74HC00 series that I've dealt with primarily here. The 4000 series is older and must be handled more carefully, because it's more vulnerable to static electricity than the generations that came later. Still, the chips remain widely available, and their great advantage is their willingness to tolerate a wide voltage range, typically from 5 to 15 volts. This means you can set up a 12-volt circuit that drives a 555 timer, and use output from the timer to go straight into CMOS chips (for example). The book is well organized in three sections: CMOS basics, functional diagrams (showing pinouts for all the main chips), and simple circuits showing how to make the chips perform basic functions.

The Encyclopedia of Electronic Circuits, by Rudolf F. Graf (Tab Books, 1985)

A totally miscellaneous collection of schematics, with minimal explanations. This is a useful book to have around if you have an idea and want to see how someone else approached the problem. Examples are often more valuable than general explanations, and this book is a massive compendium of examples. Many additional volumes in the series have been published, but start with this one, and you may find it has everything you need.

The Circuit Designer's Companion, by Tim Williams (Newnes, Second Edition, 2005)

Much useful information about making things work in practical applications, but the style is dry and fairly technical. May be useful if you're interested in moving your electronics projects into the real world.

The Art of Electronics, by Paul Horowitz and Winfield Hill (Cambridge University Press, Second Edition, 1989)

The fact that this book has been through 20 printings tells you two things: (1) Many people regard it as a fundamental resource; (2) Secondhand copies should be widely available, which is an important consideration, as the list price is over $100. It's written by two academics, and has a more technical approach than *Practical Electronics for Inventors*, but I find it useful when I'm looking for backup information.

Getting Started in Electronics, by Forrest M. Mims III (Master Publishing, Fourth Edition, 2007)

Although the original dates back to 1983, this is still a fun book to have. I think I have covered many of its topics here, but you may benefit by reading explanations and advice from a completely different source, and it goes a little farther than I have into some electrical theory, on an easy-to-understand basis, with cute drawings. Be warned that it's a brief book with eclectic coverage. Don't expect it to have all the answers.

Making Things
Getting Started with Arduino
Make:

Figure 5-9. *These books from MAKE provide guidance if you want to go beyond basic microcontrollers into the more exotic realms of the Arduino chip.*

TTL Cookbook
HOROWITZ AND HILL
THE ART OF ELECTRONICS
PRACTICAL ELECTRONICS FOR INVENTORS

Figure 5-10. *A sun-damaged copy of the Don Lancaster's classic TTL Cookbook, a 2,000-page catalog from the Mouser Electronics supply company, and two comprehensive reference books that can provide years of additional guidance in all areas of electronics.*

Experiment 25: Magnetism

This experiment should be a part of any school science class, but even if you remember doing it, I suggest that you do it again, because setting it up takes only a matter of moments, and it's going to be our entry point to a whole new topic: the relationship between electricity and magnetism. Quickly this will lead us into audio reproduction and radio, and I'll describe the fundamentals of self-inductance, which is the third and final basic property of passive components (resistance and capacitance being the other two). I left self-inductance until last because it's not very relevant to the experiments that you've done so far. But as soon as we start dealing with analog signals that fluctuate, it becomes essential.

You will need:

- Large screwdriver.
- 22-gauge wire (or thinner). Quantity: 6 feet.
- AA battery.

Procedure

This couldn't be simpler. Wind the wire around the shaft of the screwdriver, near its tip. The turns should be neat and tight and closely spaced, and you'll need to make 100 of them, within a distance of no more than 2 inches. To fit them into this space, you'll have to make turns on top of previous turns. If the final turn tends to unwind itself (which will happen if you're using stranded wire), secure it with a piece of tape. See Figure 5-11.

Now apply a battery, as shown in Figure 5-12. At first sight, this looks like a very bad idea, because you're going to short out your battery just as you did in Experiment 2. But by passing the current through a wire that's coiled instead of straight, we'll get some work out of it before the battery expires.

Put a small paper clip near the screwdriver blade, on a soft, smooth surface that will not present much friction. A tissue works well. Because many screwdrivers are already magnetic, you may find that the paper clip is naturally attracted to the tip of the blade. If this happens, move the clip just outside the range of attraction. Now apply the 1.5 volts to the circuit, and the clip should jump to the tip of the screwdriver. Congratulations: you just made an electromagnet.

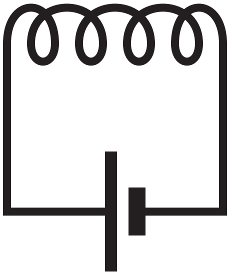

Figure 5-12. *A schematic can't get much simpler than this.*

FUNDAMENTALS

A two-way relationship

Every electric motor that was ever made uses some aspect of the relationship between electricity and magnetism. It's absolutely fundamental in the world around us. Remember that electricity can create magnetism:

When electricity flows through a wire, it creates a magnetic force around the wire.

The principle works in reverse: magnetism can create electricity.

When a wire moves through a magnetic field, it creates a flow of electricity in the wire.

This second principle is used in power generation. A diesel engine, or a water-powered turbine, or a windmill, or some other source of energy either turns coils of wire through a powerful magnetic field, or turns magnets amid some massive coils of wire. Electricity is induced in the coils. In the next experiment, you'll see a dramatic mini-demo of this effect.

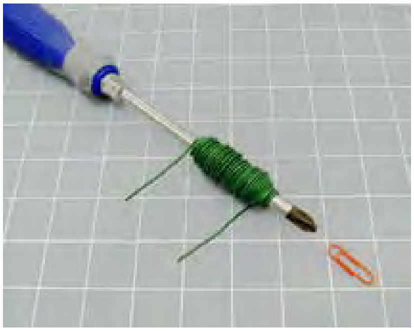

Figure 5-11. *Anyone who somehow missed this most basic childhood demo of electromagnetism should try it just for the fun of proving that a single AA battery can move a paper clip.*

THEORY

Inductance

When electricity flows through a wire, it creates a magnetic field around the wire. Because the electricity "induces" this effect, it is known as *inductance*. The effect is illustrated in Figure 5-13.

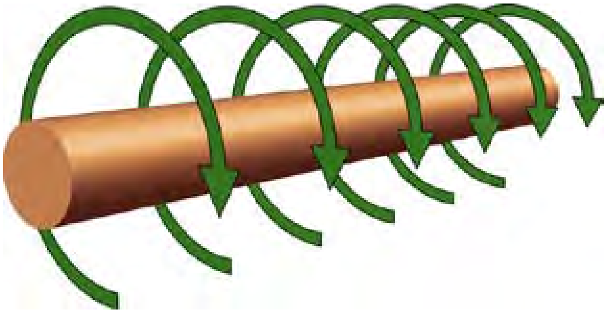

Figure 5-13. *When the flow of electricity is from left to right along this conductor, it induces a magnetic force shown by the green arrows.*

The field around a straight wire is very weak, but if we bend the wire into a circle, the magnetic force starts to accumulate, pointing through the center of the circle, as shown in Figure 5-14. If we add more circles, to form a coil, the force accumulates even more. And if we put a magnetic object (such as a screwdriver) in the center of the coil, the effectiveness increases further.

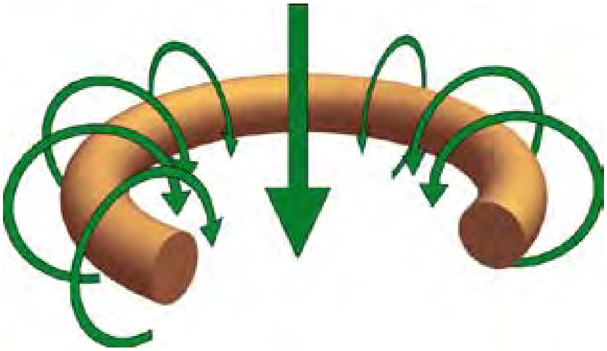

Figure 5-14. *When the conductor is bent to form a circle, the cumulative magnetic force acts through the center of the circle, as shown by the large arrow.*

Here's an approximated formula showing the relationship between the diameter of the coil, the width of the coil from end to end, the number of turns, and its inductance. The letter L is the symbol for inductance, even though the unit is the Henry, named after an American electrical pioneer named Joseph Henry:

> L (in microHenrys) =
> [(D × D) × (N x N)] / [(18 × D) + (40 × W)]
> (Approximately)

In this formula, D is the diameter of the coil, N is the number of turns, and W is the width of the coil from end to end. See Figure 5-15. Here are three simple conclusions from this formula:

- Inductance increases with the diameter of the coil.
- Inductance increases with the square of the number of turns. (In other words, three times as many turns create nine times the inductance.)
- If the number of turns remains the same, inductance is lower if you wind the coil so that it's slender and long, but is higher if you wind it so that it's fat and short.

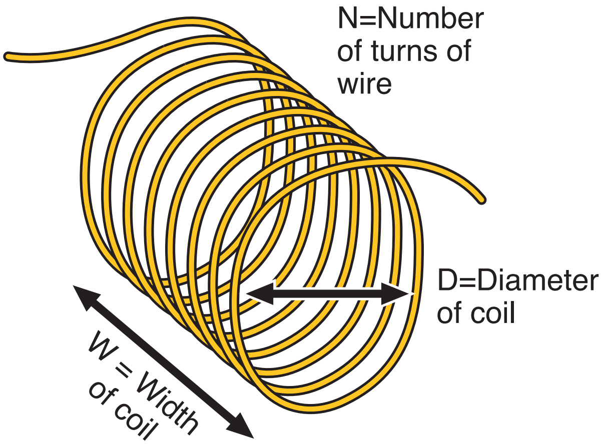

Figure 5-15. *The inductance of a coil increases with its diameter and with the square of its number of turns. If all other parameters remain the same, reducing the width (the distance from end to end) by packing the turns more tightly will increase the inductance.*

FUNDAMENTALS

Coil schematics and basics

Check the schematic symbols for coils in Figure 5-16. Note that if a coil has an iron core, this is shown with an extra couple of lines (sometimes only one line). If it has a ferrite core, the line is sometimes shown with dashes.

An iron core will add to the inductance of a coil, because it increases the magnetic effect.

A coil in isolation does not generally have any polarity. You can connect it either way around, but the magnetic force will be reversed accordingly (coils that interact with stuff—such as in transformers and solenoids—do have polarity).

Perhaps the most widespread application of coils is in transformers, where alternating current in one coil induces alternating current in another, often sharing the same iron core. If the primary (input) coil has half as many turns as the secondary (output) coil, the voltage will be doubled, at half the current—assuming hypothetically that the transformer is 100% efficient.

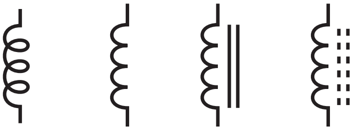

Figure 5-16. *Schematic diagrams represent coils. At far right is the older style. The third and fourth symbols indicate that the coil is wound around a solid or powdered magnetic core, respectively.*

BACKGROUND

Joseph Henry

Born in 1797, Joseph Henry was the first to develop and demonstrate powerful electromagnets. He also originated the concept of "self-inductance," meaning the "electrical inertia" that is a property of a coil of wire.

Henry started out as the son of a day laborer in Albany, New York. He worked in a general store before being apprenticed to a watchmaker, and was interested in becoming an actor. Friends persuaded him to enroll at the Albany Academy, where he turned out to have an aptitude for science. In 1826, he was appointed Professor of Mathematics and Natural Philosophy at the Academy, even though he was not a college graduate and described himself as being "principally self-educated." Michael Faraday was doing similar work in England, but Henry was unaware of it.

Henry was appointed to Princeton in 1832, where he received $1,000 per year and a free house. When Morse attempted to patent the telegraph, Henry testified that he was already aware of its concept, and indeed had rigged a system on similar principles to signal his wife, at home, when he was working in his laboratory at the Philosophical Hall.

Henry taught chemistry, astronomy, and architecture, in addition to physical science, and because science was not divided into strict specialties as it is now, he investigated phenomena such as phosphorescence, sound, capillary action, and ballistics. In 1846, he headed the newly founded Smithsonian Institution as its secretary.

Figure 5-17. *Joseph Henry was an American experimenter who pioneered the investigation of electromagnetism. This photograph is archived in Wikimedia Commons.*

Experiment 26: Tabletop Power Generation

If you have just three components, you can see magnetism generating electricity right in front of you, right now.

You will need:

- Cylindrical neodymium magnet, 3/4-inch diameter, axially magnetized. Quantity: 1. (Obtainable online at sites such as *http://www.kjmagnetics.com*.)
- Spool of hookup wire, 26-gauge, 100 feet. Quantity: 1.
- Spool of magnet wire, quarter-pound, 26-gauge, about 350 feet. Quantity: 1. (Search online for sources for "magnet wire.")
- Generic LED. Quantity: 1.
- 100 µF electrolytic capacitor. Quantity: 1.
- Signal diode, 2N4001 or similar. Quantity: 1.
- Jumper wires with alligator clips on the ends. Quantity: 2.

Procedure

You may be able to make this experiment work with the spool of hookup wire, depending on the size of the spool relative to the size of your magnet, but as the results are more likely to be better with the magnet wire, I'll assume that you're using that—initially, at least. The advantage of the magnet wire is that its very thin insulation allows the coils to be closely packed, increasing their inductance.

First peek into the hollow center of the spool to see if the inner end of the wire has been left accessible, as is visible in Figures 5-18 and 5-19. If it hasn't, you have to unwind the wire onto any large-diameter cylindrical object, then rewind it back onto the spool, this time taking care to leave the inner end sticking out.

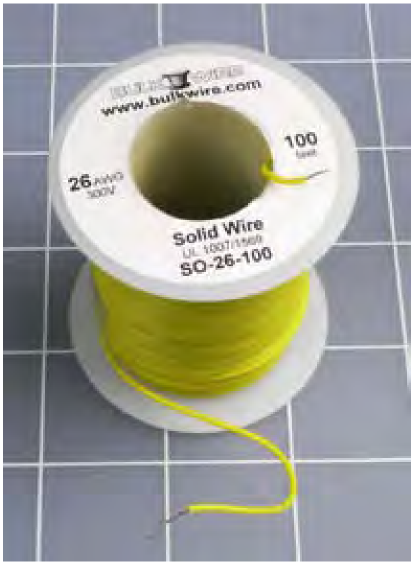
Figure 5-18. *An everyday 100-foot spool of hookup wire is capable of demonstrating the inductive power of a coil.*

Scrape the transparent insulation off each end of the magnet wire with a utility knife or sandpaper, until bare copper is revealed. To check, attach your meter, set to measure ohms, to the free ends of the wire. If you make a good contact, you should measure a resistance of 30 ohms or less.

Place the spool on a nonmagnetic, nonconductive surface such as a wooden, plastic, or glass-topped table. Attach the LED between the ends of the wire using jumper wires. The polarity is not important. Now take a cylindrical neodymium magnet of the type shown in Figure 5-20 and push it quickly down into the hollow core, then pull it quickly back out. See Figure 5-21. You should see the LED blink, either on the down stroke or the up stroke.

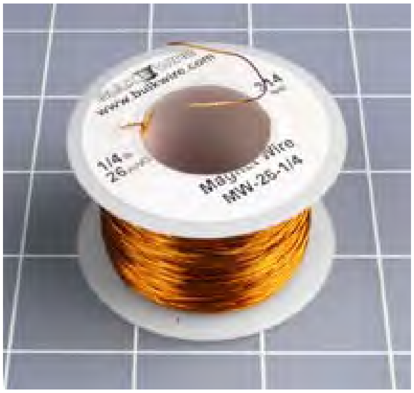
Figure 5-19. *Magnet wire has thinner insulation than hookup wire, allowing the turns to be more densely packed, inducing a more powerful magnetic field.*

The same thing may or may not happen if you use 100 feet of 26-gauge hookup wire. Ideally, your cylindrical magnet should fit fairly closely in the hollow center of the spool. If there's a big air gap, this will greatly reduce the effect of the magnet. Note that if you use a weaker, old-fashioned iron magnet instead of a neodymium magnet, you may get no result at all.

Blood Blisters and Dead Media

Neodymium magnets can be hazardous. They're brittle and can shatter if they slam against a piece of magnetic metal (or another magnet). For this reason, many manufacturers advise you to wear eye protection.

Because a magnet pulls with increasing force as the distance between it and another object gets smaller, it closes the final gap very suddenly and powerfully. You can easily pinch your skin and get blood blisters.

If there's an object made of iron or steel anywhere near a neodymium magnet, the magnet will find it and grab it, with results that may be unpleasant, especially if the object has sharp edges and your hands are in the vicinity. When using a magnet, create a clear area on a nonmagnetic surface, and watch out for magnetic objects underneath the surface. My magnet sensed a steel screw embedded in the underside of a kitchen countertop, and slammed itself into contact with the countertop unexpectedly.

Be aware that magnets create magnets. When a magnetic field passes across an iron or steel object, the object picks up some magnetism of its own. Be careful not to magnetize your watch!

Don't use magnets anywhere near a computer, a disk drive, credit cards with magnetic stripes, cassettes of any type, and other media. Also keep magnets well away from TV screens and video monitors (especially cathode-ray tubes). Last but not least, powerful magnets can interfere with the normal operation of cardiac pacemakers.

Figure 5-20. *Three neodymium magnets, 1/4-, 1/2-, and 3/4-inch in diameter. I would have preferred to photograph them standing half-an-inch apart, but they refused to permit it.*

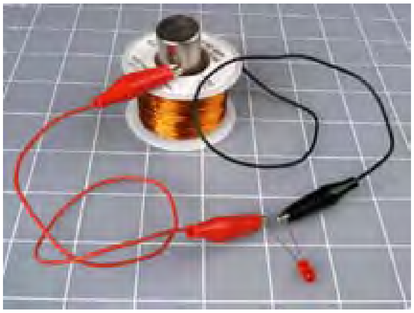

Figure 5-21. *By moving a magnet vigorously up and down through the center of a coil, you generate enough power to make the LED flash brightly.*

Here's another thing to try. Disconnect the LED and connect a 100 μF electrolytic capacitor in series with signal diode, as shown in Figure 5-23. Attach your meter, measuring volts, across the capacitor. If your meter has a manual setting for its range, set it to 20V DC. Make sure the positive (unmarked) side of the diode is attached to the negative (marked) side of the capacitor, so that positive voltage will pass through the capacitor and then through the diode.

Now move the magnet vigorously up and down in the coil. The meter should show that the capacitor is accumulating charge, up to about 10 volts. When you stop moving the magnet, the voltage reading will gradually decline, mostly because the capacitor discharges itself through the internal resistance of your meter.

This experiment is more important than it looks. Bear in mind that when you push the magnet into the coil, it induces current in one direction, and when you pull it back out again, it induces current in the opposite direction. You are actually generating alternating current.

The diode only allows current to flow one way through the circuit. It blocks the opposite flow, which is how the capacitor accumulates its charge. If you jump to the conclusion that diodes can be used to change alternating current to direct current, you're absolutely correct. We say that the diode is "rectifying" the AC power.

Experiment 24 showed that voltage can create a magnet. Experiment 25 has shown that a magnet can create voltage. We're now ready to apply these concepts to the detection and reproduction of sound.

Figure 5-22. *Because inductance increases with the diameter of a coil and with the square of the number of turns, your power output from moving a magnet through the coil can increase dramatically with scale. Those wishing to live "off the grid" may consider this steam-powered configuration, suitable for powering a three-bedroom home.*

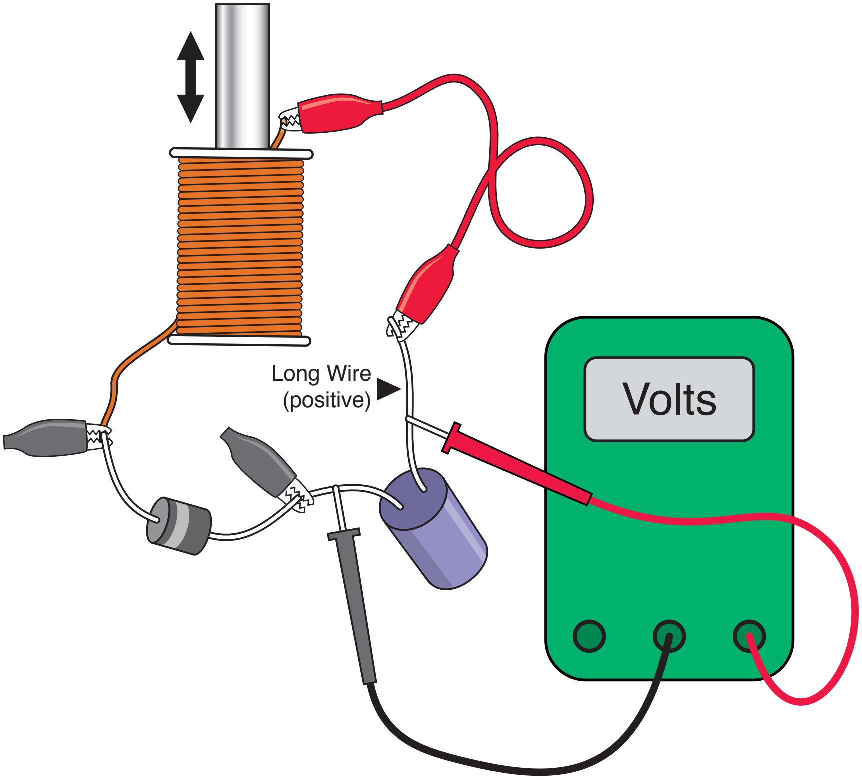

Figure 5-23. *Using a diode in series with a capacitor, you can charge the capacitor with the pulses of current that you generate by moving the magnet through the center of the coil. This demo illustrates the principle of rectifying alternating current.*

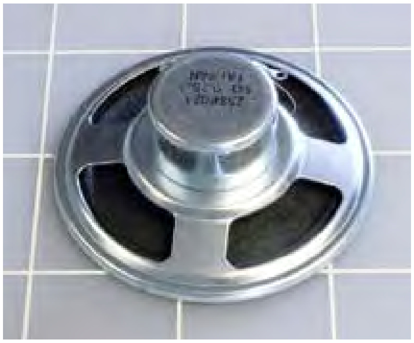

Figure 5-24. *A 2-inch loudspeaker can be instructively destroyed with a utility knife or X-Acto blade.*

Experiment 27: Loudspeaker Destruction

I'd like you to sacrifice a 2-inch loudspeaker, even though it means wasting the $5 or so that it probably costs. Actually, I don't consider this a waste, because if you want to learn how a component works, there's no substitute for actually seeing inside it. You might also have such a speaker already, part of a piece of cast-off personal electronics or toy you have in your basement.

You will need:

- Cheapest possible 2-inch loudspeaker. Quantity: 1. Figure 5-24 shows a typical example.

Procedure

Turn the loudspeaker face-up (as shown in Figure 5-25) and cut around the edge of its cone with a sharp utility knife or X-Acto blade. Then cut around the circular center and remove the ring of black paper that you've created. The result should look like Figure 5-26: you should see the flexible neck of the loudspeaker, which is usually made from a yellow weave. If you cut around its edge, you should be able to pull up the hidden paper cylinder, which has the copper coil of the loudspeaker wound around it. In Figure 5-27, I've turned it over so that it is easily visible. The two ends of this copper coil normally receive power through two terminals at the back of the speaker. When it sits in the groove visible between the inner magnet and the outer magnet, the coil reacts to voltage fluctuations by exerting an up-and-down force in reaction to the magnetic field. This vibrates the cone of the loudspeaker and creates sound waves.

Large loudspeakers in your stereo system work exactly the same way. They just have bigger magnets and coils that can handle more power (typically, as much as 100 watts).

Whenever I open up a small component like this, I'm impressed by the precision and delicacy of its parts, and the way it can be mass-produced for such a low cost. I imagine how astonished the pioneers of electrical theory (such as Faraday and Henry) would be, if they could see the components that we take for granted today. Henry spent days and weeks winding coils by hand to create electromagnets that were far less efficient than this cheap little loudspeaker.

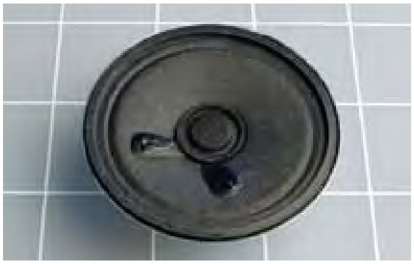

Figure 5-25. *Loudspeaker ready for creative destruction.*

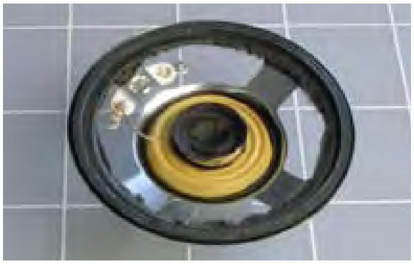

Figure 5-26. *The cone has been removed.*

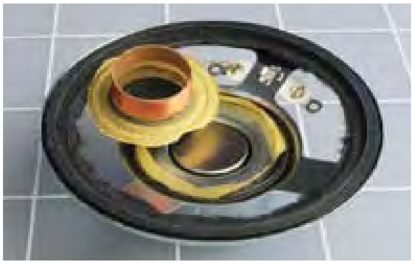

Figure 5-27. *The neck of the cone has been pulled out. Note the coil of copper wire, which fits precisely in the groove between two magnets in the base of the speaker.*

BACKGROUND

Origins of loudspeakers

Loudspeakers utilize the fact that if you run a varying electrical current through a coil situated in a magnetic field, the coil will move in response to the current. This idea was introduced in 1874 by Ernst Siemens, a prolific German inventor. (He also built the world's first electrically powered elevator in 1880.) Today, Siemens AG is one of the largest electronics companies in the world.

When Alexander Graham Bell patented the telephone in 1876, he used Siemen's concept to create audible frequencies in the earpiece. From that point on, sound-reproduction devices gradually increased in quality and power, until Chester Rice and Edward Kellogg at General Electric published a paper in 1925 establishing basic principles that are still used in loudspeaker design today.

At *http://www.radiolaguy.com/Showcase/Gallery-HornSpkr.htm* you'll find photographs of very beautiful early loudspeakers, which used a horn design to maximize efficiency. As sound amplifiers became more powerful, speaker efficiency became less important compared with quality reproduction and low manufacturing costs. Today's loudspeakers convert only about 1% of electrical energy into acoustical energy.

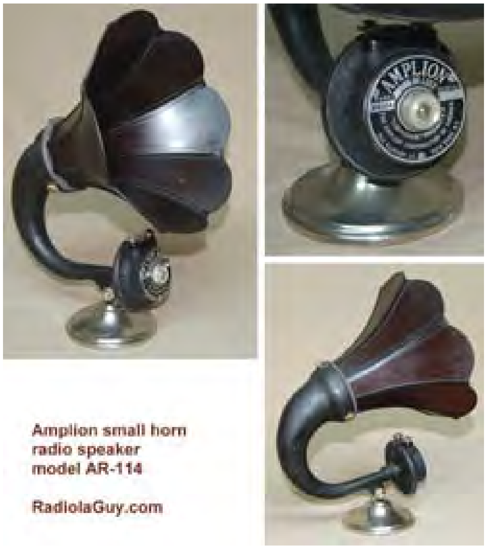

Figure 5-28. *This beautiful Amplion AR-114x illustrates the efforts of early designers to maximize efficiency in an era when the power of audio amplifiers was very limited. Photos by "Sonny, the RadiolaGuy." Many early speakers are illustrated at www.radiolaguy.com. Some are for sale.*

THEORY

Sound, electricity, and sound

Time now to establish a clear idea of how sound is transformed into electricity and back into sound again.

Suppose someone bangs a gong with a stick. The flat metal face of the gong vibrates in and out, creating sound waves. A sound wave is a peak of higher air pressure, followed by a trough of lower air pressure.

The wavelength of the sound is the distance (usually ranging from meters to millimeters) between one peak of pressure and the next peak.

The frequency of the sound is the number of waves per second, usually expressed as hertz.

Suppose we put a very sensitive little membrane of thin plastic in the path of the pressure waves. The plastic will flutter in response to the waves, like a leaf fluttering in the wind. Suppose we attach a tiny coil of very thin wire to the back of the membrane so that it moves with the membrane, and let's position a stationary magnet inside the coil of wire. This configuration is like a tiny, ultra-sensitive loudspeaker, except that instead of electricity producing sound, it is configured so that sound produces electricity. Sound pressure waves make the membrane move to and fro along the axis of the magnet, and the magnetic field creates a fluctuating voltage in the wire.

This is known as a *moving-coil* microphone. There are other ways to build a microphone, but this is the configuration that is easiest to understand. Of course, the voltage that it generates is very small, but we can amplify it using a transistor, or a series of transistors. Then we can feed the output through the coil around the neck of a loudspeaker, and the loudspeaker will recreate the pressure waves in the air. Figures 5-29 through 5-32 illustrate this sequence.

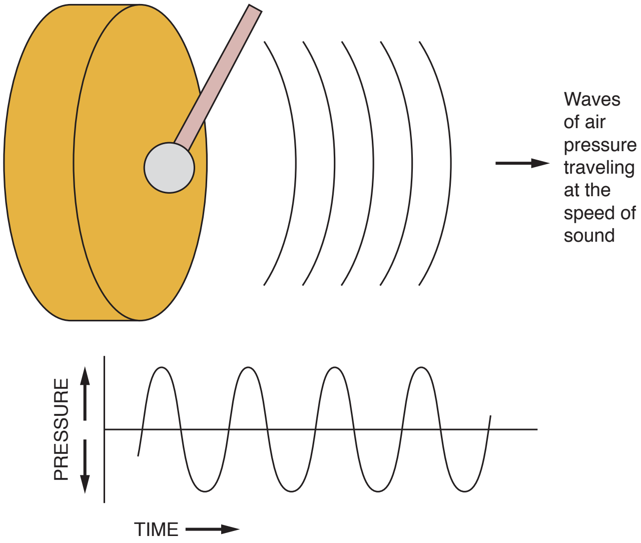

Figure 5-29. *Step 1 in the process of converting sound to electricity, and back again. When the hammer hits the gong, the face of the gong vibrates, creating pressure waves that travel through the air.*

THEORY

Sound, electricity, and sound (continued)

Somewhere along the way, we may want to record the sound and then replay it. But the principle remains the same. The hard part is designing the microphone, the amplifier, and the loudspeaker so that they reproduce the waveforms *accurately* at each step. It's a significant challenge, which is why accurate sound reproduction can be elusive.

Time now to think about what happens inside the wire when it generates a magnetic field. Obviously, some of the power in the wire is being transformed into magnetic force. But just what exactly is going on?

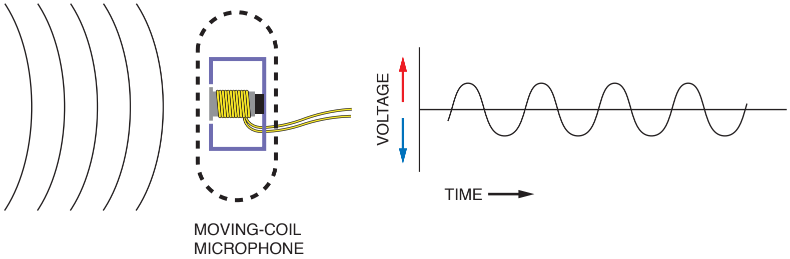

Figure 5-30. *Step 2: the pressure waves penetrate the perforated shell of a microphone and cause a diaphragm to vibrate in sympathy. The diaphragm has a coil attached to it. When the coil vibrates to and fro, a magnet at its center induces alternating current.*

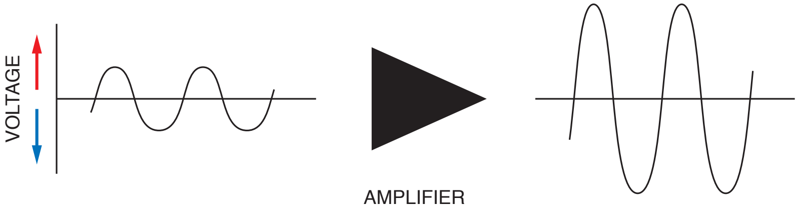

Figure 5-31. *Step 3: the tiny signals from the microphone pass through an amplifier, which enlarges their amplitude while retaining their frequency and the shape of their waveform.*

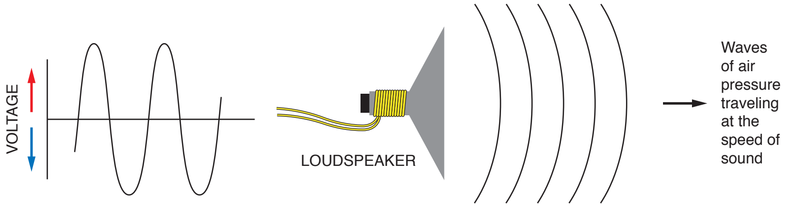

Figure 5-32. *Step 4: the amplified electrical signal is passed through a coil around the neck of a loudspeaker cone. The magnetic field induced by the current causes the cone to vibrate, reproducing the original sound.*

Experiment 28: Making a Coil React

A capacitor will absorb some DC current until it is fully charged, at which point it blocks the flow. There's another phenomenon that I haven't mentioned so far, which is the exact opposite of capacitance. It's known as *self-inductance*, and you find it in any coil of wire. Initially it blocks DC current (it reacts against it), but then its opposition gradually disappears. Here are a few definitions:

Resistance

Constrains current flow and drops voltage.

Capacitance

Allows current to flow initially and then blocks it. This behavior is properly known as capacitive reactance.

Self-Inductance

Blocks the flow of current initially and then allows it. This is also often referred to as *inductive reactance*. In fact, you may find the term "reactance" used as if it means the same thing, but since self-inductance is the correct term, I'll be using it here.

In this experiment, you'll see self-inductance in action.

You will need:

- LEDs, low-current type. Quantity: 2.
- Spool of hookup wire, 26-gauge, 100 feet. Quantity: 1.
- Resistor, 220Ω, rated 1/4 watt or higher. Quantity: 1.
- Capacitor, electrolytic, 2,000 µF or larger. Quantity: 1.
- SPST tactile switch. Quantity: 1.

Procedure

Take a look at the schematic in Figure 5-33. At first it may not make much sense. The curly symbol is a coil of wire—nothing more than that. So apparently the voltage will pass through the 220Ω resistor, and then through the coil, ignoring the two LEDs, because the coil obviously has a much lower resistance than either of them (and one of them is upside-down anyway).

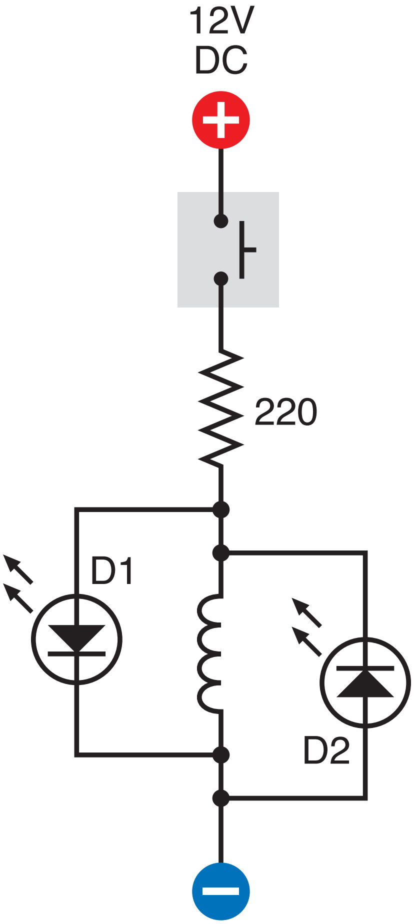

Figure 5-33. *In this demonstration of self-inductance, D1 and D2 are light-emitting diodes. When the switch is closed, D1 flashes briefly because the coil obstructs the initial flow of electricity. When the switch is opened, D2 flashes as the collapsing magnetic field induced by the coil releases another short burst of current.*

Is that what will happen? Let's find out. The coil can be a spool of 100 feet of 26-gauge (or smaller) hookup wire, although the spool of magnet wire listed in Experiment 25 will work better, if you have that. Once again, you will need access to both ends of the wire, and if the inner end is inaccessible, you'll need to rewind the coil, leaving the end sticking out.

Now that you have a coil, you can hook it up on your breadboard as shown in Figure 5-34, where the green circle is a tactile switch and the two circular red objects are LEDs. Make sure that you use low-current LEDs (otherwise, you may not see anything) and make sure that one of them is negative-side-up, positive-side-down and the other is positive-side-up, negative-side-down. Also, the 220Ω resistor should be rated at 1/4 watt or higher, if possible (see the following caution).

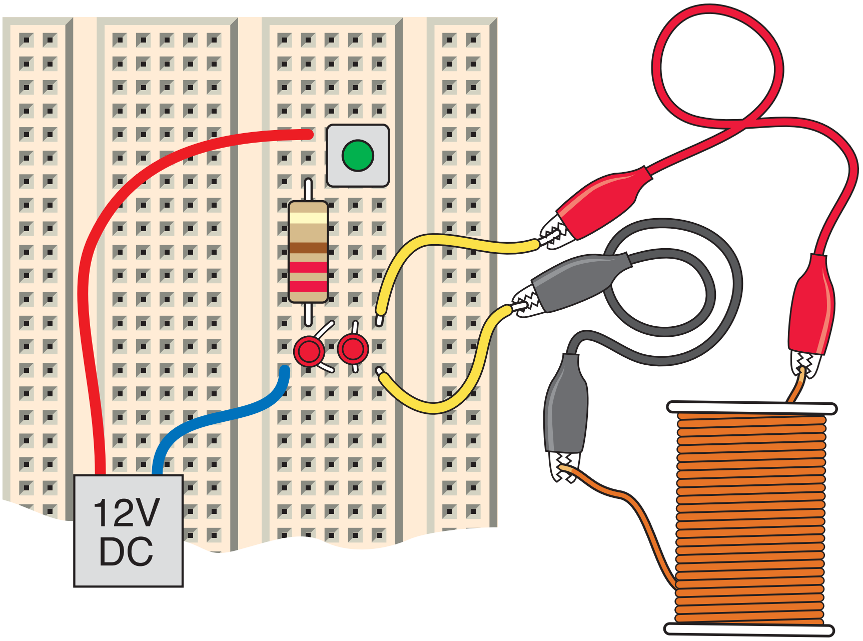

Figure 5-34. *The breadboarded version of the schematic in Figure 5-33 shows a simple way to set it up for a quick demo. The green button is a tactile switch. The two red LEDs should be placed so that the polarity of one is opposite to the polarity of the other.*

When you press the button, one LED should flash briefly. When you release the button, the other LED should flash.

What's happening here? The coil possesses self-inductance, which means that it reacts against any sudden change in the flow of electricity. First it fights it, and during that brief moment, it blocks most of the current. Consequently, the current looks for an alternative path and flows through D1, the lefthand LED in the schematic. (D2 doesn't respond, because it can pass current only in the opposite direction.)

Meanwhile, the voltage pressure overcomes the coil's self-inductance. When the self-inductance disappears, the resistance of the coil is no more than 10 ohms—so now the electricity flows mostly through the coil, and because the LED receives so little, it goes dark.

When you disconnect the power, the coil reacts again. It fights *any* sudden changes. After the flow of electricity stops, the coil stubbornly sustains it for a moment, because as the magnetic field collapses, it is turned back into electricity. This residual flow of current depletes itself through D2, the LED on the right.

In other words, the coil stores some energy in its magnetic field. This is similar to the way a capacitor stores energy between two metal plates, except that the coil blocks the current initially and then lets it build up, whereas the capacitor sucks up current initially, and then blocks it.

The more turns of wire you have in your coil, the more self-inductance the coil will have, causing your LEDs to flash more brightly.

Hot Resistors

You'll be passing about 50mA through the 220Ω resistor, while the current is flowing. At 12 volts, this works out at 0.6 watts. If you use a 1/8-watt resistor, you will be overloading it, and it will get quite hot and may burn out. If you use a 1/4-watt resistor, it will still get hot, but is unlikely to burn out, as long as you don't press the button for more than a second or two.

Don't run the circuit without the coil of wire; you'll be trying to pass more than 50mA through the LEDs.

Here's one last variation on this experiment to test your understanding of electrical fundamentals. Remove the 220Ω resistor, and substitute a 1K resistor (to protect your LED from sustained current). Remove the coil, and substitute a very large capacitor—ideally, about 4,700 µF. (Be careful to get its polarity the right way around.) What will you see when you press the button? Note that you will have to hold it down for a couple of seconds to get a result. And what will you see when you release the button? Remember: the behavior of capacitance is opposite to the behavior of self-inductance.

THEORY

Alternating current concepts

Here's a simple thought experiment. Suppose you set up a 555 timer to send a stream of pulses through a coil. This is a primitive form of alternating current.

We might imagine that the self-inductance of the coil will interfere with the stream of pulses, depending how long each pulse is, and how much inductance the coil has. If the pulses are too short, the self-inductance of the coil will tend to block them. Maybe if we can time the pulses exactly right, they'll synchronize with the time constant of the coil. In this way, we can "tune" a coil to allow a "frequency" to pass through it.

What happens if we substitute a capacitor? If the pulses are too long, compared with the time constant of the capacitor, it will tend to block them, because it will have enough time to become fully charged. But if the pulses are shorter, the capacitor can charge and discharge in rhythm with the pulses, and will seem to allow them through.

I don't have space in this book to get deeply into alternating current. It's a vast and complicated field where electricity behaves in strange and wonderful ways, and the mathematics that describe it can become quite challenging, involving differential equations and imaginary numbers. However, we can easily demonstrate the audio filtering effects of a loudspeaker and a coil.

Experiment 29: Filtering Frequencies

In this experiment, you'll see how self-inductance and capacitance can be used to filter audio frequencies. You're going to build a crossover network: a simple circuit that sends low frequencies to one place and high frequencies to another.

You will need:

- Loudspeaker, 8Ω, 5 inches in diameter. Quantity: 1. Figure 5-35 shows a typical example.
- Audio amplifier, STMicroelectronics TEA2025B or similar. Quantity: 1. See Figure 5-36.

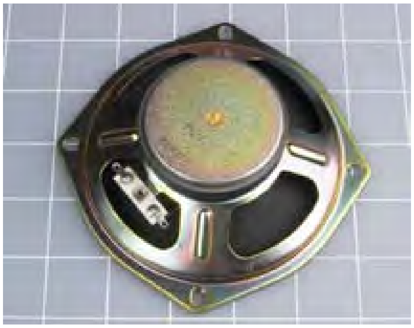

Figure 5-35. *To hear the effects of audio filters using coils and capacitors, you'll need a loudspeaker capable of reproducing lower frequencies. This 5-inch model is the minimum required.*

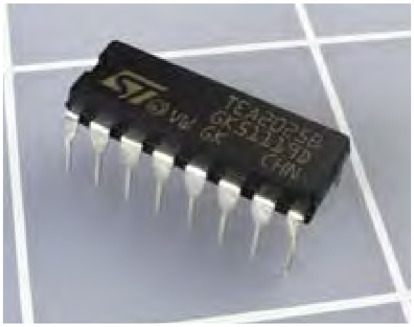

Figure 5-36. *This single chip contains a stereo amplifier capable of delivering a total of 5 watts into an 8Ω speaker when the two channels are combined.*

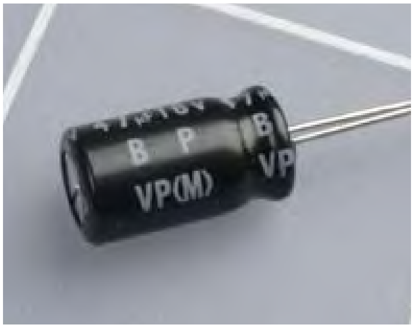

Figure 5-37. *A nonpolarized electrolytic capacitor, also known as a bipolar capacitor, looks just like an electrolytic capacitor, except that it will have "NP" or "BP" printed on it.*

- Nonpolarized electrolytic capacitors (also known as bipolar). 47 µF. Quantity: 2. A sample is shown in Figure 5-37. They should have "NP" or "BP" printed on them to indicate "nonpolarized" or "bipolar."
- Nonpolarized electrolytic capacitors (also known as bipolar). 100 µF. Quantity: 5. (Because you'll be working with audio signals that alternate between positive and negative, you can't use the usual polarized electrolytic capacitors. If you want to avoid the trouble and expense of ordering nonpolarized capacitors, you can substitute two regular electrolytics in series, facing in opposite directions, with their negative sides joined in the middle. Just remember that when you put capacitors in series, their total capacitance is half that of each individual component. Therefore, you would need two 220 µF electrolytics in series to create 110 µF of capacitance. See Figure 5-38.)
- Potentiometer, with audio taper if possible. 100K. Quantity: 1.
- Coil, for crossover network. Quantity: 1. You can search a source such as eBay for keywords "crossover" and "coil," but if you can't find one at a reasonable price, you can make do with a spool of 100 feet of 20-gauge hookup wire.
- Plastic shoebox. Quantity: 1.

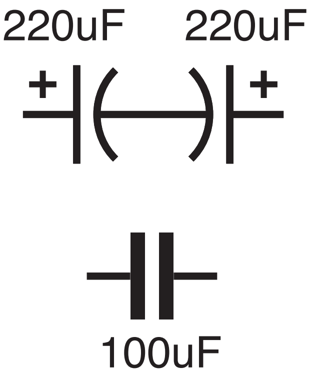

Figure 5-38. *You can make a nonpolarized electrolytic capacitor by putting two regular electrolytics in series. (In fact, that's what you'd find if you opened a real nonpolarized capacitor.) The symbol at the bottom is roughly equivalent to the pair of symbols at the top; bear in mind that two capacitors in series have a total capacitance that is half that of each of them.*

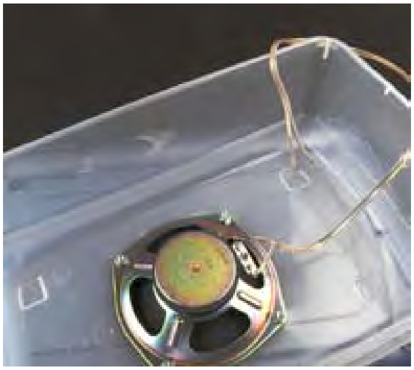

Figure 5-39. *A resonant enclosure is necessary if you want to hear some bass (lower frequencies) from your speaker. A cheap plastic shoebox is sufficient for demo purposes.*

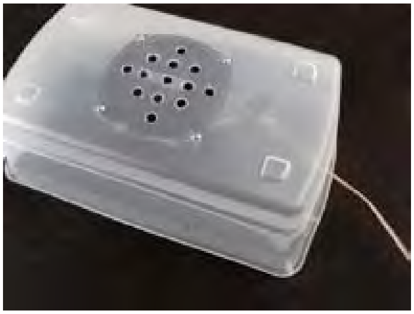

Figure 5-40. *Drill some half-inch holes in the bottom of the box, then bolt the speaker in place, running a wire out through a hole in one end. Snap on the lid, and you're ready for not-quite-high-fidelity audio.*

Procedure

The purpose of the audio amplifier chip is to provide enough power to get a decent amount of sound out of your loudspeaker. The purpose of using a 5-inch speaker is to enable you to hear lower-frequency sounds than the baby speakers that we have used previously. Bass notes have long wavelengths that small speakers are not able to generate effectively.

Maybe you remember from building the intrusion alarm that a speaker makes much more noise if you prevent the sound waves from the back of the cone from cancelling the sound waves from the front of the cone. The obvious way to achieve this is by enclosing the speaker in a box. I suggest a plastic box, because they're cheap, and we don't care too much about sound quality as long as we can hear at least some of the low frequencies. Figure 5-39 shows the speaker bolted into the bottom of a plastic box, and Figure 5-40 shows the box turned upside-down after snapping its lid into place.

Normally, a speaker should be mounted in a cabinet of heavy, thick material that has a very low resonant frequency—below the limits of human hearing. To minimize the resonance of the shoebox, you can put some soft, heavy fabric inside it before you snap the lid on. A hand towel or some socks should be sufficient to absorb some of the vibration.

Adding an Amplifier

Back in the 1950s, you needed vacuum tubes, transformers, and other power-hungry heavyweight components to build an audio amplifier. Today, you can buy a chip for about $1 that will do the job, if you add a few capacitors around it, and a volume control. The TEA2025B that I'm recommending is intended for use in cheap portable cassette players and CD players, and can work in stereo or mono mode, from a power supply ranging from 3 to 9 volts. With 9 volts and the two sides of the chip bridged together to drive one 8Ω speaker, it can generate 5 watts of audio power. That doesn't sound much compared with a typical home theater system rated at 100 watts per channel, but because loudness is a logarithmic scale, 5 watts will be quite enough to irritate any family members in the same room—and possibly even in other rooms.

If you can't find the TEA2025B chip, you can use any alternative listed as an audio amplifier. Try to find one that is designed to drive an 8Ω speaker with up to 5 watts in mono mode. Check the manufacturer's data sheet to see where you attach capacitors around it. Note carefully whether some of the capacitors have no polarity marked, even though they have fairly high values, such as 100 µF. These capacitors must function regardless of which way the alternating current is flowing, and I've marked them "NP" in my schematic in Figure 5-41, meaning "nonpolarized." (You may find them identified as "bipolar" or "BP" in parts catalogs.) As noted in the shopping list, you can put two 220 µF capacitors in series, negative-to-negative, to get the same effect as a single 100 µF nonpolarized capacitor.

For this project, it's essential to include the regular 100 µF electrolytic smoothing capacitor across the power supply. Otherwise, the amplifier will pick up and—yes, amplify—small voltage spikes in the circuit.

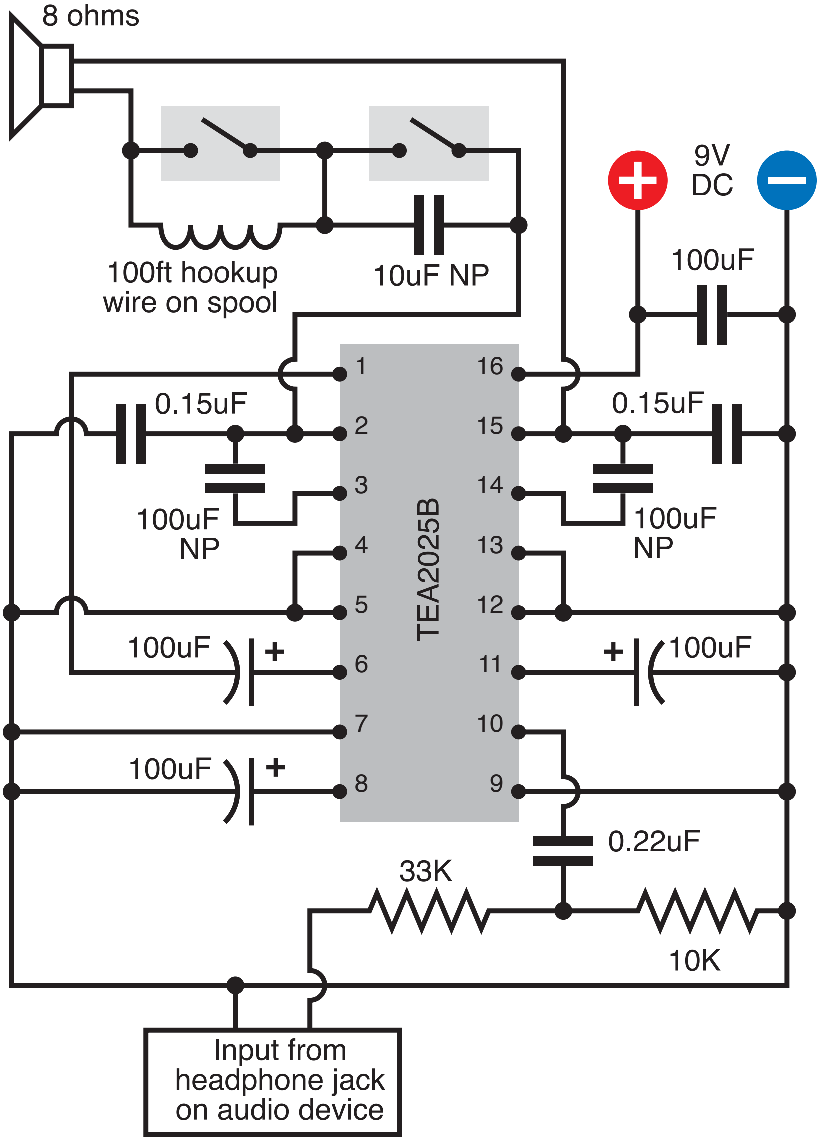
8 ohms
9V DC
100ft hookup wire on spool
10uF NP
100uF
0.15uF
0.15uF
100uF NP
100uF NP
TEA2025B
1
2
3
4
5
6
7
8
16
15
14
13
12
11
10
9
100uF
100uF
100uF
0.22uF
33K
10K
Input from headphone jack on audio device

Figure 5-41. *The audio amplifier chip should be wired with capacitors around it as shown, "NP" denoting the ones that are not polarized. The acronym "BP," meaning bipolar, is also often used to mean the same thing. The output from pins 2 and 15 of the chip can be passed through a coil or a 10 μF capacitor to demonstrate audio filtering.*

The input shown in the schematic can receive a signal from a typical media player, such as a portable MP3 player, CD player, or cassette player. To connect its headphone jack to the breadboard, you can use an adapter that converts it to a pair of RCA-type audio jacks, and then stick a wire into one of them as shown in Figure 5-42. The wire will connect to the 33K resistor on the breadboard circuit. The chromed neck of the RCA jack (which is sometimes gold-plated, or at least gold-colored) *must* be connected with the negative side of your power supply on the breadboard; otherwise, you won't hear anything. You can ignore the second output on the adapter, because we're working in mono, here, not stereo.

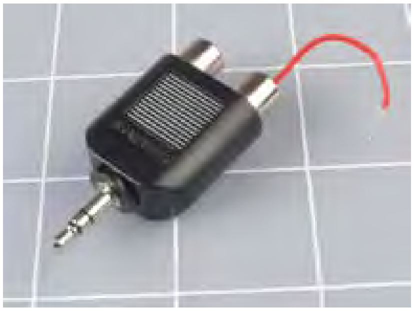

Figure 5-42. *To sample the output from the headphone socket of a music player, you can use this adapter and push the stripped end of a piece of hookup wire into one of the sockets. Then use alligator clips on a jumper wire to connect the audio to your breadboarded circuit. Don't forget to use an additional jumper wire to connect the outside of the socket to the negative side of the power supply on the breadboard. Because we're only using one speaker, the amplifier is connected to only one of the stereo outputs from your player. The other is ignored.*

The 33K resistor is necessary to protect the amplifier from being overdriven. If you don't get enough volume using your music player, decrease the 33K value. If the music is too loud and distorted, increase the value. You can also try omitting or increasing the 10K resistor next to it, which is included in an effort to reduce background hum noise.

I've shown two switches at the top of the schematic: one to bypass a coil, the other to bypass a capacitor. You can use alligator clips instead, as long as you can easily compare the sound when each of the components is inserted into the circuit.

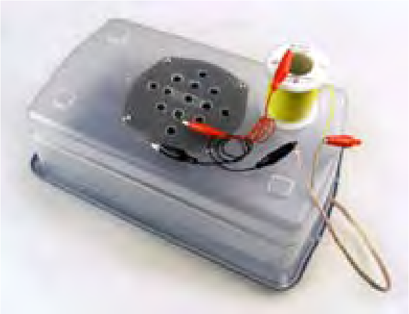

Figure 5-43. *The red and black alligator clips, lying on top of the shoebox, should connect with the output from your amplifier chip. The red jumper wire passes the signal through a coil of hookup wire on its way to the speaker. Note the change in sound when you short out the coil.*

Figure 5-43 shows a coil consisting of a spool of hookup wire being used. The red and black alligator clips resting loose on top of the shoebox will go to the output from the chip (on pins 2 and 15). There is no polarity; it doesn't matter which clip goes to which pin.

Begin by turning the volume control on your music source *all the way down* before you apply power. Don't be surprised if you hear humming or crackling noises when you activate the amplifier; it will pick up any stray voltages, because in this simple experiment, I haven't suggested that you should shield the input, and the amplifier circuit can pick up noise, as its wires can act like antennas.

Note that you may also get additional unwanted sound if you use the amplifier on a conductive desktop surface. Remove any aluminum foil or conductive foam for this project.

Make sure that your player is playing music, and slowly turn up its volume control until you hear it. If you don't hear anything, you'll have to check for circuit errors.

Now comes the interesting part. Insert the 100-foot spool of hookup wire between one output from the amplifier, and one input of the speaker (it doesn't matter which one), or if you used switches, open the switch that bypasses the coil. You should find that the music loses all its high-end response. By comparison, if you disconnect the coil and substitute a 10 µF capacitor, you should find that the music sounds "tinny," meaning that it loses all its low range, leaving only the high frequencies.

You've just tested two very simple filters. Here's what they are doing:

- The coil is a low-pass filter. It passes low frequencies but blocks high frequencies, because brief audio cycles don't have time to overcome the coil's self-inductance. A bigger coil eliminates a wider range of frequencies.
- The capacitor is a high-pass filter. It passes high frequencies and blocks low frequencies because longer audio cycles can fill the capacitance, at which point the capacitor stops passing current. A smaller capacitor eliminates a wider ranger of frequencies.

You can go a lot farther into filter design, using complex combinations of coils and capacitors to block frequencies at any point in the audible spectrum. Search online for audio filter schematics—you'll find hundreds of them.

Crossover Networks

In a traditional audio system, each speaker cabinet contains two drivers—one of them a small speaker called a *tweeter*, which reproduces high frequencies, the other a large speaker known as a *woofer*, which reproduces low frequencies. (Modern systems often remove the woofer and place it in a separate box of its own that can be positioned almost anywhere, because the human ear has difficulty sensing the direction of low-frequency sounds.)

The schematic that you just looked at and may have constructed is known as a "crossover network," and truly hardcore audiophiles have been known make their own (especially for use in car systems) to go with speakers of their choice in cabinets that they design and build themselves.

If you want to make a crossover network, you should use high-quality polyester capacitors (which have no polarity, last longer than electrolytics, and are better made) and a coil that has the right number of turns of wire and is the right size to cut high frequencies at the appropriate point. Figure 5-44 shows a polyester capacitor.

Figure 5-45 shows an audio crossover coil that I bought on eBay for $6. I was curious to find out what was inside it, so I bought two of them, and took one apart.

First I peeled away the black vinyl tape that enclosed the coil. Inside was some typical magnet wire—copper wire thinly coated with shellac or semitransparent plastic, as shown in Figure 5-46. I unwound the wire and counted the number of turns. Then I measured the length of the wire, and finally used a micrometer to measure the diameter of the wire, after which I checked online to find a conversion from the diameter in mils (1/1,000 of an inch) to American wire gauge.

As for the spool, it was plain plastic with an air core—no iron or ferrite rod in the center. Figure 5-47 shows the spool and the wire.

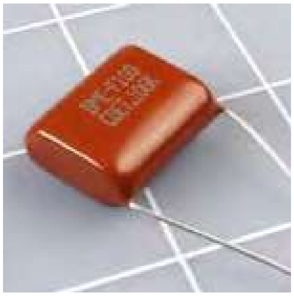

Figure 5-44. *Some nonelectrolytic capacitors have no polarity, such as this high-quality polyester film capacitor. However, they tend to be much more expensive, and are hard to find in values higher than 10 μF.*

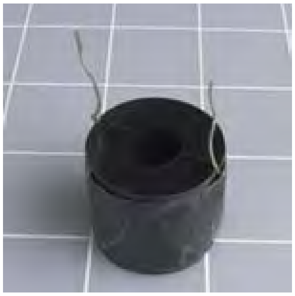

Figure 5-45. *What exotic components may we find inside this high-end audio component that's used with a subwoofer to block high frequencies?*

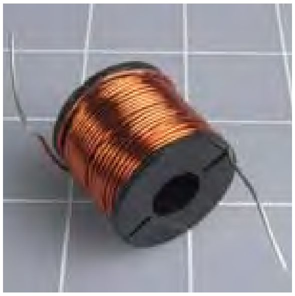

Figure 5-46. *The black tape is removed, revealing a coil of magnet wire.*

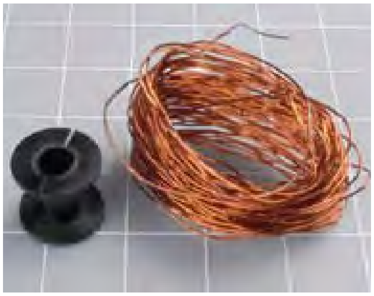

Figure 5-47. *The audio crossover coil consists of a plastic spool and some wire. Nothing more.*

So here's the specification for this particular coil in an audio crossover network. Forty feet of 20-gauge copper magnet wire, wrapped in 200 turns around a spool of 1/16–inch-thick plastic with a hub measuring 7/8 inch in length between the flanges and 1/2-inch external diameter. Total retail cost of materials if purchased separately: probably about $1, assuming you can find or make a spool of the appropriate size.

Conclusion: there's a lot of mystique attached to audio components. They are frequently overpriced, and you can make your own coil if you start with these parameters and adjust them to suit yourself.

Suppose you want to put some thumping bass speakers into your car. Could you build your own filter so that they only reproduce the low frequencies? Absolutely—you just need to wind a coil, adding more turns until it cuts as much of the high frequencies as you choose. Just make sure the wire is heavy enough so that it won't overheat when you push 100 or more audio watts through it.

Here's another project to think about: a color organ. You can tap into the output from your stereo and use filters to divide audio frequencies into three sections, each of which drives a separate set of colored LEDs. The red LEDs will flash in response to bass tones, yellow LEDs in response to the mid-range, and green LEDs in response to high frequencies (or whatever colors you prefer). You can put signal diodes in series with the LEDs to rectify the alternating current, and series resistors to limit the voltage across the LEDs to, say, 2.5 volts (when the music volume is turned all the way up). You'll use your meter to check the current passing through each resistor, and multiply that number by the voltage drop across the resistor, to find the wattage that it's handling, to make sure the resistor is capable of dissipating that much power without burning out.

Audio is a field offering all kinds of possibilities if you enjoy designing and building your own electronics.

THEORY

Waveforms

If you blow across the top of a bottle, the mellow sound that you hear is caused by the air vibrating inside the bottle, and if you could see the pressure waves, they would have a distinctive profile.

If you could slow down time and draw a graph of the alternating voltage in any power outlet in your house, it would have the same profile.

If you could measure the speed of a pendulum swinging slowly to and fro in a vacuum, and draw a graph of the speed relative to time, once again it would have the same profile.

That profile is a *sine wave*, so called because you can derive it from basic trigonometry. In a right-angled triangle, the sine of an angle is found by dividing the length of the side opposite the angle by the length of the hypoteneuse (the sloping side of the triangle).

To make this simpler, imagine a ball on a string rotating around a center point, as shown in Figure 5-48. Ignore the force of gravity, the resistance of air, and other annoying variables. Just measure the vertical height of the ball and divide it by the length of the string, at regular instants of time, as the ball moves around the circular path at a constant speed. Plot the result as a graph, and there's your sine wave, shown in Figure 5-49. Note that when the ball circles below its horizontal starting line, we consider its distance negative, so the sine wave becomes negative, too.

Why should this particular curve turn up in so many places and so many ways in nature? There are reasons for this rooted in physics, but I'll leave you to dig into that topic if it interests you. Getting back to the subject of audio reproduction, what matters is this:

- Any sound can be broken down into a mixture of sine waves of varying frequency and amplitude.

Or, conversely:

- If you put together the right mix of audio sine waves, you can create *any sound at all*.

Suppose that there are two sounds playing simultaneously. Figure 5-50 shows one sound as a red curve, and the other as pale blue. When the two sounds travel either as pressure waves through air or as alternating electric currents through a wire, the amplitudes of the waves are added together to make the more complex curve, which is shown in black. Now try to imagine dozens or even hundreds of different frequencies being added together, and you have an idea of the complex waveform of a piece of music.

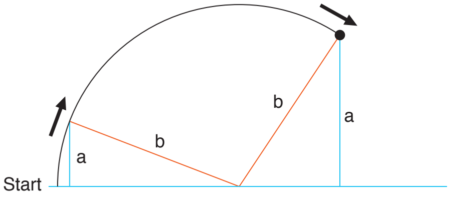

Figure 5-48. *If a weight on the end of a string (length b, in the diagram) follows a circular path at a steady speed, the distance of the weight from a horizontal center line (length a, in the diagram) can be plotted as a graph relative to time. The graph will be a sine wave, so called because in basic trigonometry, the ratio of a/b is the sine of the angle between line b and the horizontal baseline, measured at the center of rotation. Sinewaves occur naturally in the world around us, especially in audio reproduction and alternating current.*

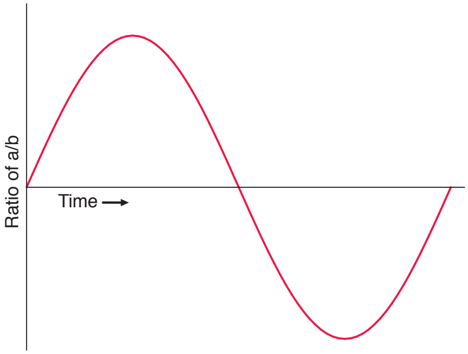

Figure 5-49. *This is what a "pure" sinewave looks like.*

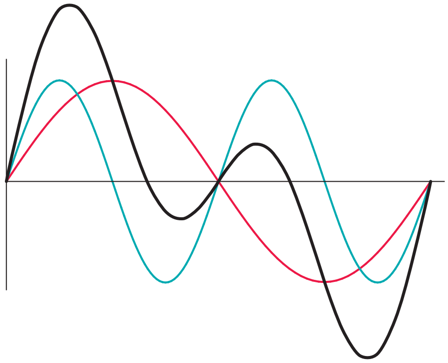

Figure 5-50. *When two sinewaves are generated at the same time (for instance, by two musicians, each playing a flute), the combined sound creates a compound curve. The blue sinewave is twice the frequency of the red sinewave. The compound curve (black line) is the sum of the distances of the sinewaves from the baseline of the graph.*

THEORY

Waveforms (continued)

You can create your own waveform as an input for your audio amplifier with the basic astable 555 timer circuit shown in Figure 5-51. You have to be careful, though, not to overload the amplifier input. Note the 680K series resistor on the output pin of the timer. Also note the 500Ω potentiometer.

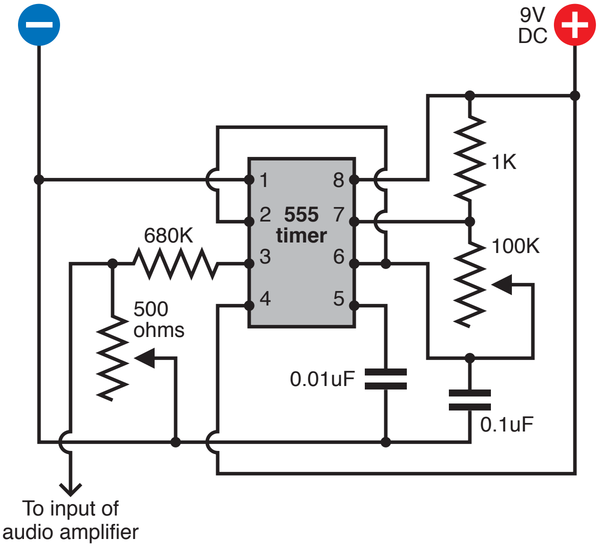

Figure 5-51. *A 555 timer is wired in astable mode using the component values shown here to generate a wide range of audible frequencies when the 100K potentiometer is adjusted. After the output is reduced in power, it can feed into the amplifier chip that was used previously.*

Disconnect your music player and connect the output from the 555 circuit to the input point (the 33K resistor) in the amplifier circuit shown earlier in Figure 5-41. You don't have to worry about a separate connection on the negative side as long as the 555 timer shares the same breadboard and the negative side of its power supply.

Make sure that the 500Ω potentiometer is turned all the way to short the output from the timer to the negative side of the power supply. This functions as your volume control. Also make sure the 100K potentiometer is in the middle of its range. Switch on the power and slowly turn up the 500Ω potentiometer until you hear a tone.

Now adjust the 100K potentiometer to create a low-pitched note. You'll find that it doesn't have a "pure" sound. There are some buzzing overtones. This is because the 555 timer is generating square waves such as those shown in Figure 5-52, not sine waves, and a square wave is actually a sum of many different sine waves, some of which have a high frequency. Your ear hears these harmonics, even though they are not obvious when you look at a square-shaped waveform.

Route one of the connections to your loudspeaker through your spool of hookup wire, and now you should hear a much purer tone, as the buzzing high frequencies are blocked by the self-inductance of the coil. Remove the coil and substitute the 10 µF capacitor, and now you hear more buzzing and less bass.

You've just taken a small step toward sound synthesis. If this subject interests you, you can go online and search for oscillator circuits. For a thorough understanding of the relationship between waveforms and the sounds you hear, you'll really need an oscilloscope, which will show you the shape of each waveform that you generate and modify.

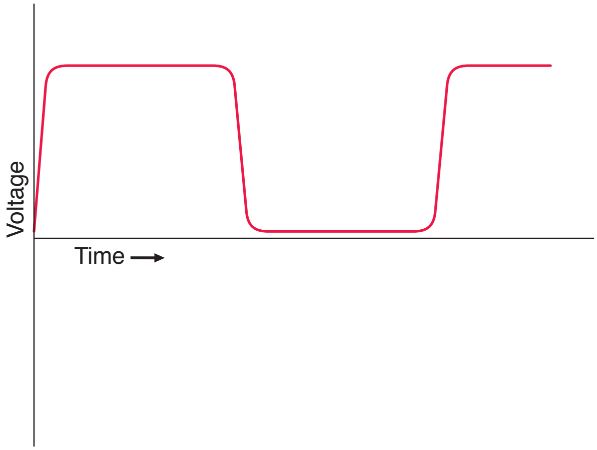

Figure 5-52. *The output from a 555 timer is either "on" or "off," with a very fast transition between those two states. The result is an almost perfect square wave. Theoretically, this can be disassembled into a complex set of sine waves that have many different frequencies. The human ear hears the high frequencies as harsh overtones.*

Experiment 30: Fuzz

Let's try one more variation on the circuit in Experiment 28. This will demonstrate another fundamental audio attribute: distortion.

You will need:

- One more 100K potentiometer.
- Generic NPN transistors: 2N2222 or similar. Quantity: 2.
- Various resistors and capacitors.

BACKGROUND

Clipping

In the early days of "hi-fi" sound, engineers labored mightily to perfect the process of sound reproduction. They wanted the waveform at the output end of the amplifier to look identical with the waveform at the input end, the only difference being that it should be bigger, so that it would be powerful enough to drive loudspeakers. Even a very slight distortion of the waveform was unacceptable.

Little did they realize that their beautifully designed tube amplifiers would be abused by a new generation of rock guitarists whose intention was to create as much distortion as possible.

The most common form of waveform abuse is technically known as "clipping." If you push a vacuum tube or a transistor to amplify a sine wave beyond the component's capabilities, it "clips" the top and bottom of the curve. This makes it look more like a square wave, and as I explained in the section on waveforms, a square wave has a harsh, buzzing quality. For rock guitarists trying to add an edge to their music, the harshness is actually a desirable feature.

Figure 5-53. *This Vox Wow-Fuzz pedal was one of the early stomp boxes, which deliberately induced the kind of distortion that audio engineers had been trying to get rid of for decades.*

The first gadget to offer this on a commercial basis was known as a "fuzz box," which deliberately clipped the input signal. An early fuzz box is shown in Figure 5-53. The clipping of a sine wave is shown in Figure 5-54.

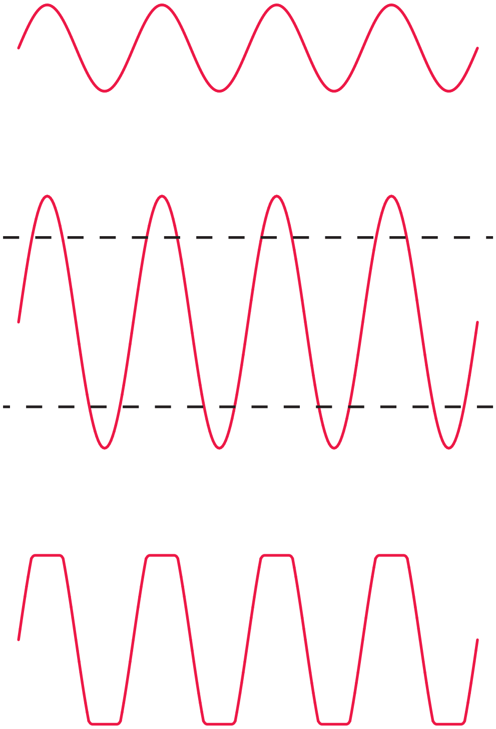

Figure 5-54. *When a sinewave (top) is passed through an amplifier which is turned up beyond the limit of its components (shown as dashed lines, center), the amplifier chops the wave (bottom) in a process known as "clipping." The result is close to a square wave and is the basic principle of a fuzz box commonly used to create a harsh guitar sound.*

Schematic

The output from the 555 timer is a square wave, so it already sounds quite "fuzzy," but we can make it more intense to demonstrate the clipping principle. I've redrawn the whole circuit in Figure 5-55, as several components have changed. The principal alteration is the addition of two NPN transistors.

If you assemble this circuit on your breadboard, note carefully that the 33K and 10K resistors at the bottom of the amplifier chip have been removed, and there's now just an 820Ω resistor in that location. The bottom of the adjacent 0.22 µF capacitor is still the input point for the amplifier, and if you follow the connection around to the middle of the schematic, you'll find it leading to a 100K potentiometer. This is your "fuzz adjuster."

The two NPN transistors are arranged so that the one on the left receives output from the 555 timer. This signal controls the flow of electricity through the transistor from a 33K resistor. This flow, in turn, controls the base of the righthand transistor, and the flow of current through that is what ultimately controls the amplifier.

When you power up the circuit, use the 100K potentiometer attached to the 555 timer to adjust the frequency (as before) and crank the "fuzz adjuster" potentiometer to hear how it adds increasing "bite" to the sound until ultimately it becomes pure noise.

The two transistors act as amplifiers. Of course, we didn't need them for that purpose—the input level for the amplifier chip was already more than adequate. The purpose of the lefthand transistor is simply to overload the righthand transistor, to create the "fuzz" effect. And when you turn up the output from the transistors with the "fuzz adjuster," eventually they overload the input of the amplifier chip, creating even more distortion.

If you want to tweak the output, try substituting different values for the 1K resistor and the 1 µF capacitor that are positioned between the emitter of the righthand transistor and the negative side of the power supply. A larger resistor should overload the transistor less. Different capacitor values should make the sound more or less harsh.

You can find literally thousands of schematics online for gadgets to modify guitar sound. The circuit I've included here is one of the most primitive. If you want something more versatile, you should search for "stomp box schematics" and see what you can find.

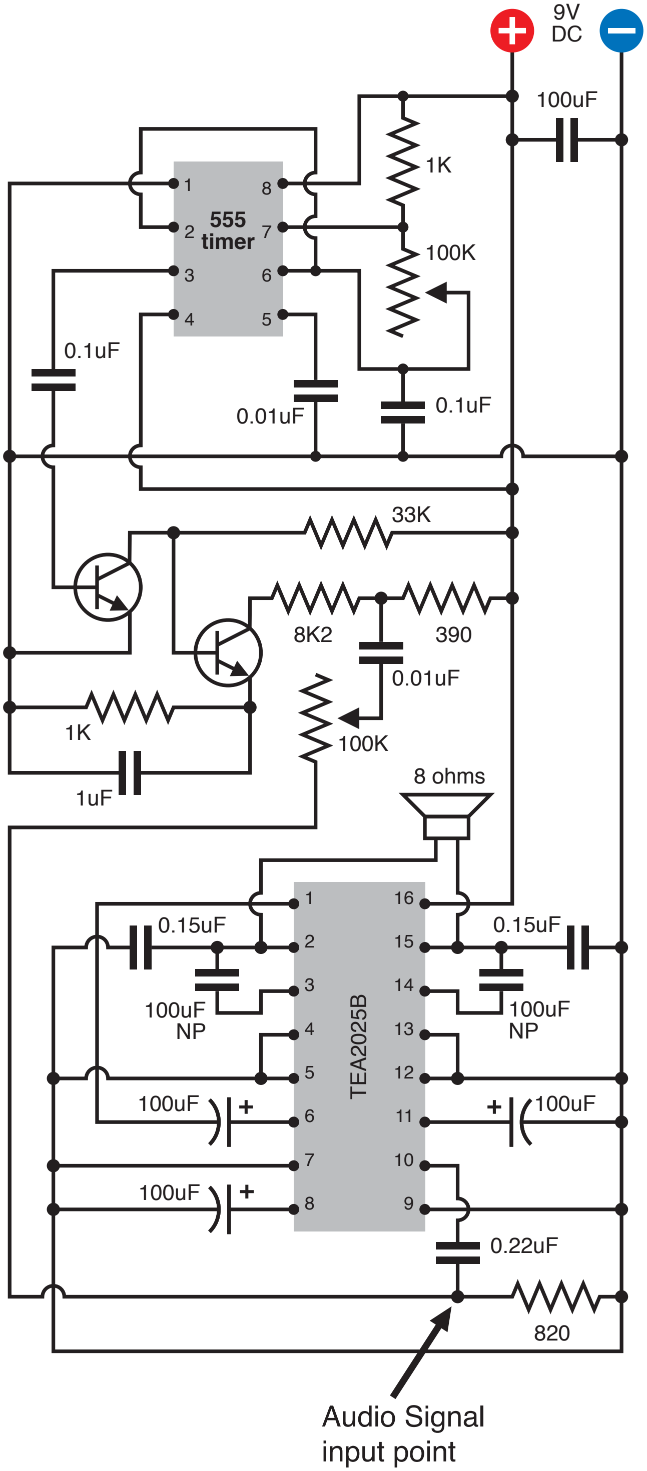
9V
DC
100uF
1K
100K
555
timer
1
2
3
4
8
7
6
5
0.1uF
0.01uF
0.1uF
33K
8K2
390
0.01uF
1K
1uF
100K
8 ohms
TEA2025B
1
2
3
4
5
6
7
8
16
15
14
13
12
11
10
9
0.15uF
100uF
NP
100uF
100uF
0.15uF
100uF
NP
100uF
0.22uF
820
Audio Signal
input point

Figure 5-55. *For a quick demo of clipping, insert a couple of transistors between the output from the 555 timer and the input of the amplifier chip. One transistor overdrives the other, so that when you adjust the potentiometer at the center of the circuit, you hear an increasingly harsh, distorted sound.*

BACKGROUND

Stomp-box origins

The Ventures recorded the first single to use a fuzz box, titled "The 2,000 Pound Bee," in 1962. Truly one of the most awful instrumentals ever made, it used distortion merely as a gimmick and must have discouraged other musicians from taking the concept seriously.

Ray Davies of the Kinks was the first to embody distortion as an integral part of his music. Davies did it initially by plugging the output from one amp into the input of another, supposedly during the recording of his hit "You Really Got Me." This overloaded the input and created clipping—the basic fuzz concept. From there it was a short step to Keith Richards using a Gibson Maestro Fuzz-Tone when the Rolling Stones recorded "(I Can't Get No) Satisfaction" in 1965.

Today, you can find thousands of advocates promoting as many different mythologies about "ideal" distortion. In Figure 5-56, I've included a schematic from Flavio Dellepiane, a circuit designer in Italy who gives away his work (with a little help from Google AdSense) at *http://www.redcircuits.com*. Dellepiane is self-taught, having gained much of his knowledge from electronics magazines such as the British *Wireless World*. In his fuzz circuit, he uses a very high-gain amplifier consisting of three field-effect transistors (FETs), which closely imitate the rounded square-wave typical of an overdriven tube amp.

Dellepiane offers dozens more schematics on his site, developed and tested with a dual-trace oscilloscope, low-distortion sinewave oscillator (so that he can give audio devices a "clean" input), distortion meter, and precision audio volt meter. This last item, and the oscillator, were built from his own designs, and he gives away their schematics, too. Thus his site provides one-stop shopping for home-audio electronics hobbyists in search of a self-administered education.

Before fuzz, there was tremolo. A lot of people confuse this with vibrato, so let's clarify that distinction right now:

- Vibrato applied to a note makes the frequency waver up and down, as if a guitarist is bending the strings.
- Tremolo applied to a note makes its volume fluctuate, as if someone is turning the volume control of a guitar up and down very quickly.

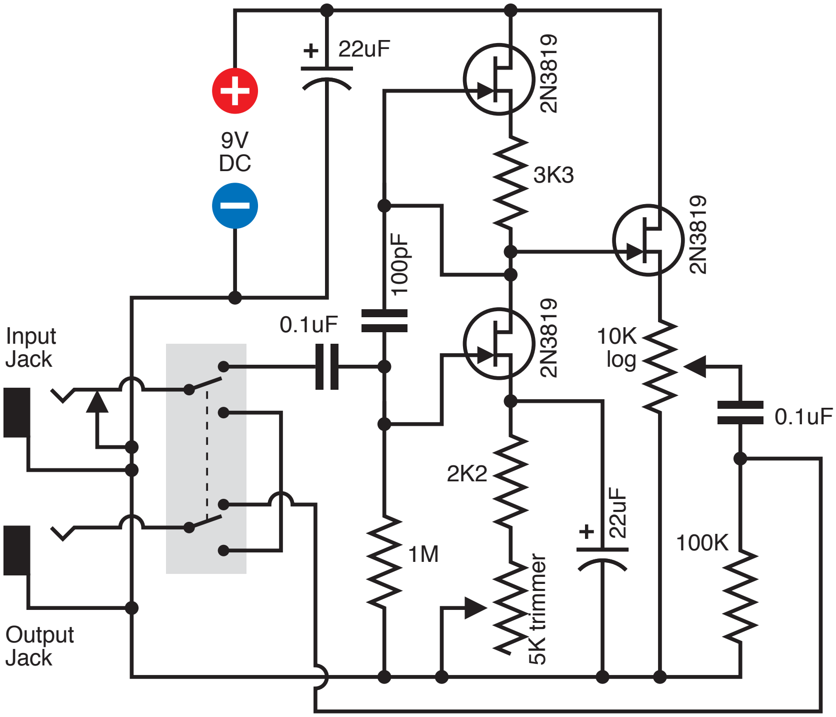

Figure 5-56. *This circuit designed by Flavio Dellepiane uses three transistors to simulate the kind of distortion that used to be created by overloading the input of a tube amplifier.*

BACKGROUND

Stomp-box origins (continued)

Harry DeArmond sold the first tremolo box, which he named the Trem-Trol. It looked like an antique portable radio, with two dials on the front and a carrying handle on top. Perhaps in an effort to cut costs, DeArmond didn't use any electronic components. His steam-punkish Trem-Trol contained a motor fitted with a tapered shaft, with a rubber wheel pressing against it. The speed of the wheel varied when you turned a knob to reposition the wheel up and down the shaft. The wheel, in turn, cranked a little capsule of "hydro-fluid," in which two wires were immersed, carrying the audio signal. As the capsule rocked to and fro, the fluid sloshed from side to side, and the resistance between the electrodes fluctuated. This modulated the audio output.

Today, Trem-Trols are an antique collectible. When industrial designer Dan Formosa acquired one, he put pictures online at *http://www.danformosa.com/dearmond.html*. And Johann Burkard has posted an MP3 of his DeArmond Trem-Trol so you can actually hear it: *http://johannburkard.de/blog/music/effects/DeArmond-Tremolo-Control-clip.html*.

The idea of a mechanical source for electronic sound mods didn't end there. The original Hammond organs derived their unique, rich sound from a set of toothed wheels turned by a motor. Each wheel created a fluctuating inductance in a sensor like the record head from a cassette player.

It's easy to think of other possibilities for motor-driven stomp boxes. Going back to tremolo: imagine a transparent disc masked with black paint, except for a circular stripe that tapers at each end. While the disc rotates, if you shine a bright LED through the transparent stripe toward a light-dependent resistor, you would have the basis for a tremolo device. You could even create never-before-heard tremolo effects by swapping discs with different stripe patterns. Figures 5-57 and 5-58 show the kind of thing I have in mind. For a real fabrication challenge, how about an automatic disc changer?

Today's guitarists can choose from a smorgasbord of effects, all of which can be home-built using plans available online. For reference, try these special-interest books:

- *Analog Man's Guide to Vintage Effects* by Tom Hughes (For Musicians Only Publishing, 2004). This is a guide to every vintage stomp box and pedal you can imagine.
- *How to Modify Effect Pedals for Guitar and Bass* by Brian Wampler (Custom Books Publishing, 2007). This is an extremely detailed guide for beginners with little or no prior knowledge. Currently it is available only by download, from sites such as *http://www.openlibrary.org*, but you may be able to find the previous printed edition from secondhand sellers, if you search for the title and the author.

Of course, you can always take a shortcut by laying down a couple hundred dollars for an off-the-shelf item such as a Boss ME-20, which uses digital processing to emulate distortion, metal, fuzz, chorus, phaser, flanger, tremolo, delay, reverb, and several more, all in one convenient multi-pedal package. Purists, of course, will claim that it "doesn't sound the same," but maybe that's not the point. Some of us simply can't get no satisfaction until we build our own stomp box and then tweak it, in search of a sound that doesn't come off-the-shelf and is wholly our own.

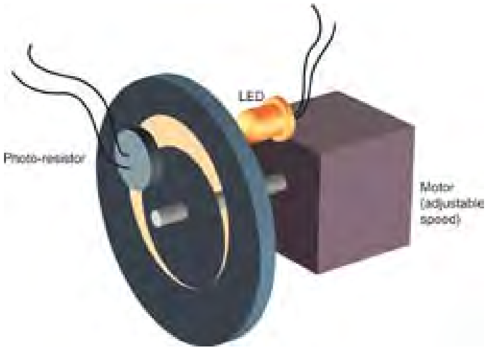

Figure 5-57. *Although electromechanical audio devices are obsolete now, some unexplored possibilities still exist. This design could create various tremolo effects, if anyone had the patience to build it.*

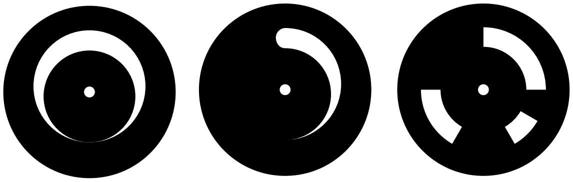

Figure 5-58. *Different stripe patterns could be used in conjunction with the imaginary electromechanical device in Figure 5-57 to create various tremolo effects.*

Experiment 31: One Radio, No Solder, No Power

Time now to go back one more time to inductance and capacitance, and demonstrate an application which also makes use of the way that waveforms can be added to each other. I want to show you how a simple circuit with no power supply at all can receive AM radio signals and make them audible. This is often known as a crystal radio, because the circuit includes a germanium diode, which has a crystal inside it. The idea dates back to the dawn of radio, but if you've never tried it, you've missed an experience that is truly magical.

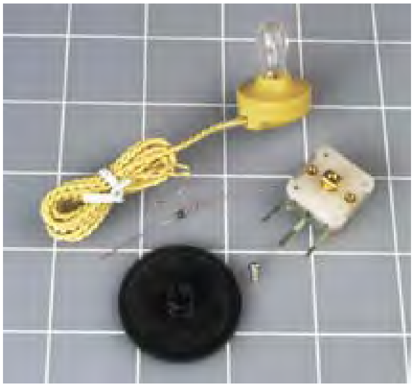

Figure 5-59. *Just add wire and a coil, and this is all you need to receive AM radio signals. The black disc becomes the tuning dial, after it is screwed onto the variable capacitor (right). This is actually an optional extra. A germanium diode (center) rectifies the radio signal. The high-impedance earphone (top) creates a barely audible sound.*

You will need:

- Rigid cylindrical object, such as a vitamin bottle. Quantity: 1.
- 22-gauge hookup wire, solid-core. Quantity: 60 feet.
- 16-gauge wire, stranded. Quantity: 100 feet.
- Polypropylene rope ("poly rope") or nylon rope. Quantity: 10 feet.
- Germanium diode. Quantity: 1.
- High-impedance headphone. Quantity: 1.

The diode and headphone can be ordered from *http://www.scitoyscatalog.com*. You cannot use a modern headphone of the type you wear with an MP3 player.

Some of these items are shown in Figure 5-59.

First, you need to make a coil. It should be about 3 inches in diameter, and you can wind it around any empty glass or plastic container of that size, so long as it's rigid. A soda bottle or water bottle isn't suitable, because the cumulative squeezing force of the turns of wire can deform the bottle so that it isn't circular anymore.

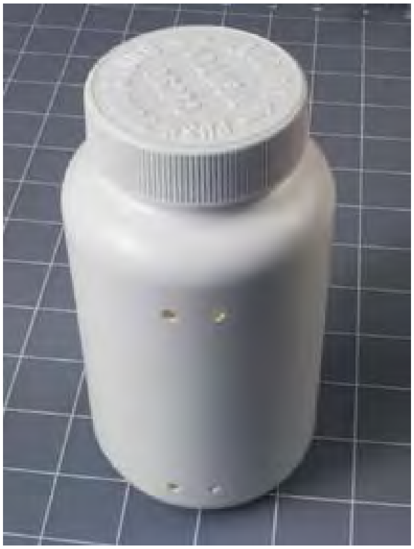

Figure 5-60. *A large, 3-inch diameter empty vitamin bottle makes an ideal core for a crystal radio coil. The drilled holes will anchor wire wrapped around the bottle.*

I chose a vitamin bottle that just happened to be exactly the right size. To remove the label, I softened its adhesive with a heat gun (lightly, to avoid melting the bottle) and then just peeled it off. The adhesive left a residue, which I removed with Xylol (also known as Xylene). This is a handy solvent to have around, as it can remove "permanent" marker stains as well as sticky residues, but you should always use latex or nitrile gloves to avoid getting it on your skin, and minimize your exposure the fumes. Because Xylol will dissolve some plastics, clearly it's not good for your lungs.

After you prepare a clean, rigid bottle, drill two pairs of holes in it, as shown in Figure 5-60. You'll use them to anchor the ends of the coil.

Now you need about 60 feet of 22-gauge solid-core wire. If you use magnet wire, its thin insulation will allow the turns of the coil to be more closely spaced, and the coil may be slightly more efficient. But everyday vinyl-insulated wire will do the job, and is much easier to work with.

Begin by stripping the first 6 inches of insulation from the end of the wire. Now measure 50 inches along the insulated remainder and apply your wire strippers at that point, just enough to cut the insulation without cutting the wire. Use your two thumb nails to pull the insulation apart, revealing about a half-inch of bare wire, as shown in Figure 5-61. Bend it at the center point and twist it into a loop, as shown in Figure 5-62.

You just made a "tap," meaning a point where you will be able to tap into the coil after you wind it. You'll need another 11 of these taps, all of them spaced 50 inches apart. (If the diameter of the bottle that you'll be using as the core of your coil is not 3 inches, multiply its diameter by 16 to get the approximate desired spacing of taps.)

After you have made 12 taps, cut the wire and strip 6 inches off that end. Now bend the end into a U shape about a half-inch in diameter, so that you can hook it through the pair of holes that you drilled at one end of the bottle. Pull the wire through, then loop it around again to make a secure anchor point.

Now wind the rest of the wire around the bottle, pulling it tightly so that the coils stay close together. When you get to the end of the wire, thread it through the remaining pair of holes to anchor it as shown in Figure 5-63. The completed coil is shown in Figure 5-64.

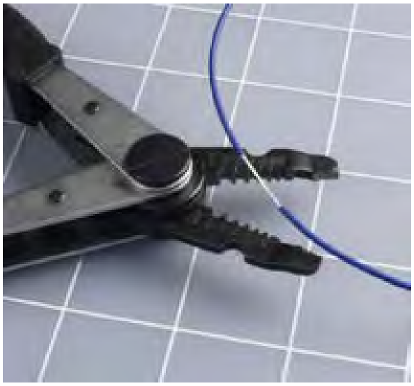

Figure 5-61. *Wire strippers expose the solid conductor at intervals along a 22-gauge wire.*

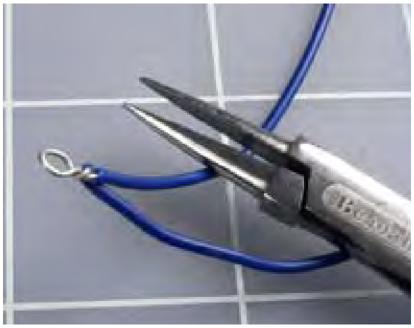

Figure 5-62. *Each exposed section of wire is twisted into a loop using sharp-nosed pliers.*

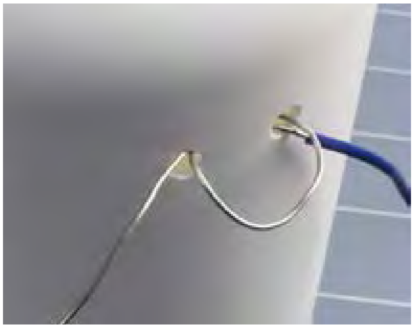

Figure 5-63. *The stripped end of the wire is secured through the holes drilled in the bottle.*

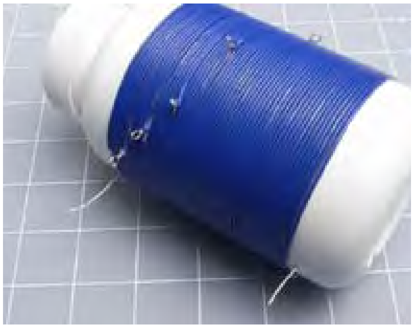

Figure 5-64. *The completed coil, wrapped tightly around the bottle.*

Your next step is to set up an antenna. If you live in a house with a yard outside, this is easy: just open a window, toss out a reel of 16-gauge wire while holding the free end, then go outside and string up your antenna by using polypropylene rope ("poly rope") or nylon rope, available from any hardware store, to hang the wire from any available trees, gutters, or poles. The total length of the wire should be about 100 feet. Where it comes in through the window, suspend it on another length of poly rope. The idea is to keep your antenna wire as far away from the ground or from any grounded objects as possible.

If you live in an apartment where you don't have access to a yard outside, you can try stringing your antenna around the room, hanging the wire from more pieces of poly rope. The antenna should still be about 100 feet long, but obviously it won't be in a straight line.

Hook the free end of your antenna to one end of your coil. At this point, you also need to add a germanium diode, which functions like a silicon-based diode but is better suited to the tiny voltages and currents that we'll be dealing with. The other end of the diode attaches to one of the wires leading to a high-impedance

High Voltage!

The world around us is full of electricity. Normally we're unaware of it, but a thunderstorm is a sudden reminder that there's a huge electrical potential between the ground below and the clouds above.

If you put up an outdoor antenna, never use it if there is any chance of a lightning strike. This can be extremely dangerous. Disconnect the indoor end of your antenna, drag it outside, and push the end of the wire into the ground to make it safe.

earphone. *A normal modern earphone or headphone will not work in this circuit.* The return wire from the headphone is connected to a jumper wire, the other end of which can be clipped to any of the taps in your coil.

One last modification, and you'll be ready to tune in. You have to ground the jumper wire. By this I mean connect it to something that literally goes into the ground. A cold-water pipe is the most commonly mentioned option, but (duh!) this will work only if the pipe is made of metal. Because a lot of plumbing these days is plastic, check under the sink to see if you have copper pipes before you try using a faucet for your ground.

Another option is to attach the wire to the screw in the cover plate of an electrical outlet, as the electrical system in your house is ultimately grounded. But the sure-fire way to get a good ground connection is to go outside and hammer a 4-foot copper-clad grounding stake into reasonably moist earth. Any wholesale electrical supply house should be able to sell you a stake. They're commonly used to ground welding equipment.

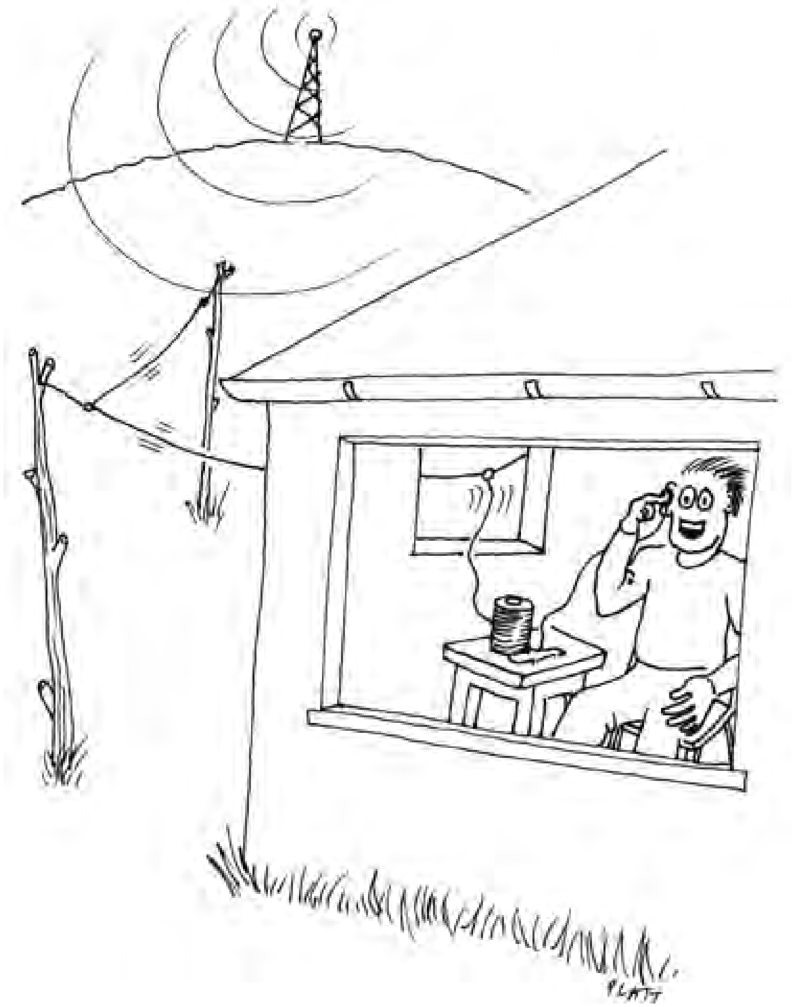

Figure 5-85. *The simple pleasure of picking up a radio signal with ultra-simple components and no additional power.*

Figures 5-66 and 5-67 show the completed radio.

If you've managed to follow these instructions (one way or another), it's time to tune your radio to the nearest station. Move the alligator clip at the end of your patch cord from one tap to another on your coil. Depending on where you live, you may pick up just one station, or several, some of them playing simultaneously.

It may seem that you're getting something for nothing here, as the earphone is making noise without any source of power. Really, though, there is a source of power: the transmitter located at the radio station. A large amplifier pumps power into the broadcasting tower, modulating a fixed frequency. When the combination of your coil and antenna resonates with that frequency, you're sucking in just enough voltage and current to energize a high-impedance headphone.

The reason you had to make a good ground connection is that the radio station broadcasts its signal at a voltage relative to ground. The earth completes the circuit between you and the transmitter. For more information on this and other concepts relating to radio, see the upcoming section "Theory: How radio works."

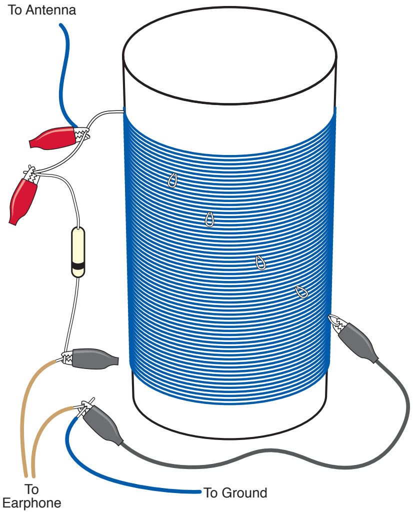

Figure 5-66. *A signal from the antenna can pass through the coil to ground. If the jumper wire is attached to an appropriate tap on the coil, it resonates with the radio signal, just powerfully enough to energize the earphone which is wired in series with a diode.*

Enhancements

The higher your antenna is, the better it should work. In my location, this is a major problem, as I live in a desert area without any trees. Still, just stringing the wire out of my window and tethering it (with rope) to the front bumper of my car enabled me to pick up a faint radio signal.

To improve the selectivity of your radio, you can add a variable capacitor, as shown in the following section. This allows you to "tune" the resonance of your circuit more precisely. Variable capacitors are uncommon today, but you can find one at the same specialty source that I recommended for the earphone and the germanium diode: the Scitoys Catalog (*http://www.scitoyscatalog.com*).

This source is affiliated with a smart man named Simon Quellan Field, whose site suggests many fun projects that you can pursue at home. One of his clever ideas is to remove the germanium diode from your radio circuit and substitute a low-power LED in series with a 1.5-volt battery. This didn't work for me, because I live 40 miles from the nearest AM broadcaster. If you're closer to a transmitter, you may be able to see the LED varying in intensity as the broadcast power runs through it.

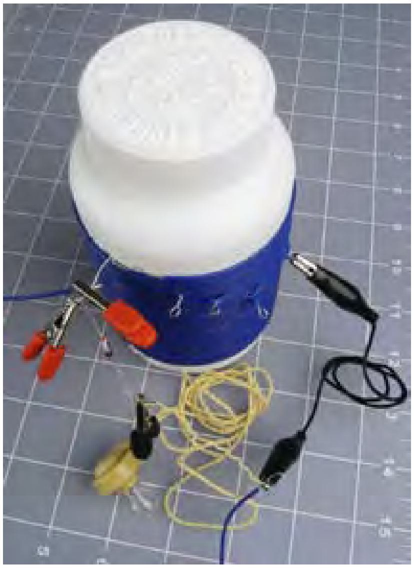

Figure 5-67. *The real-life version of Figure 5-66.*

THEORY

How radio works

When electrical frequencies are very high, the radiation they create has enough energy to travel for miles. This is the principle of radio transmission: A high-frequency voltage is applied to a broadcasting antenna, relative to the ground.

When I say "ground" in this instance, I literally mean the planet beneath us. If you set up a receiving antenna, it can pick up a faint trace of the transmission relative to the ground—as if the earth is one huge conductor. Actually the earth is so large and contains so many electrons, it can function as a common sink, like a gigantic version of the file cabinet that I suggested you should touch to get rid of static electricity in your body before touching a CMOS logic chip.

To make a radio transmitter, I could use a 555 timer chip running at, say, 850 kHz (850,000 cycles per second), and pass this stream of pulses through an amplifier to a transmission tower; if you had some way to block out all the other electromagnetic activity in the air, you could detect my signal and reamplify it.

This was more or less what Marconi (shown in Figure 5-68) was doing in 1901, after he had purchased rights to Edison's wireless telegraphy patent, although he had to use a primitive spark gap, rather than a 555 timer, to create the oscillations. His transmissions were of limited use, because they had only two states: on or off. You could send Morse code messages, and that was all.

Figure 5-68. *Marconi, the great pioneer of radio (photograph from Wikimedia Commons).*

Five years later, the first true audio signal was transmitted by imposing lower audio frequencies on the high-frequency carrier wave. In other words, the audio signal was "added" to the carrier frequency, so that the power of the carrier varied with the peaks and valleys of the audio.

At the receiving end, a very simple combination of a capacitor and a coil detected the carrier frequency out of all the other noise in the electromagnetic spectrum. The values of the capacitor and the coil were chosen so that their circuit would "resonate" at the same frequency as the carrier wave. Figures 5-69 and 5-70 illustrate these concepts.

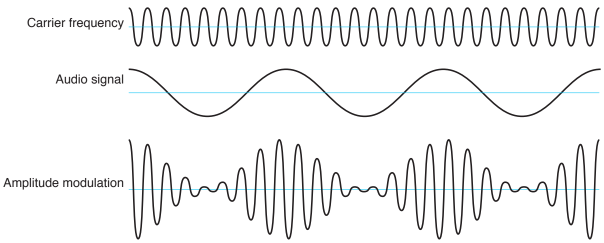

Figure 5-69. *When an audio signal (middle) is combined electronically with a high carrer frequency (top), the result looks something like the compound signal at the bottom. In actuality, the carrier frequency would be much higher compared with the audio frequency, by a radio of perhaps 1,000:1.*

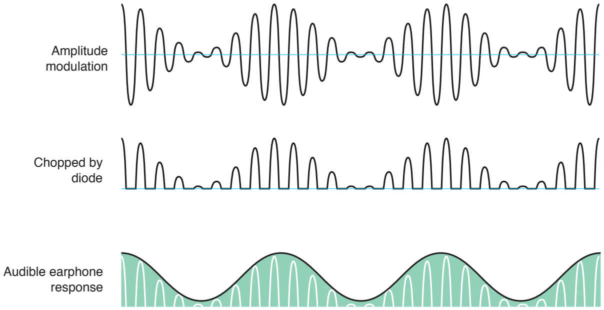

Figure 5-70. *When the compound signal is passed through a diode, only the upper half remains. An earphone cannot react fast enough to reproduce the high carrier frequency, so it "rides" the peaks and thus reproduces the audio frequency.*

The schematic in Figure 5-71 shows the simple circuit that you built by wrapping a coil around an empty vitamin bottle. When a positive pulse was received by the antenna, it resonated with the antenna and the coil, provided that the antenna was long enough and the coil was tapped at the appropriate number of turns.

THEORY

How radio works (continued)

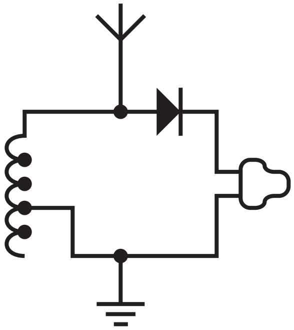

Figure 5-71. *An antenna at the top of the schematic picks up faint electromagnetic radiation from a distant transmitter. The coil at the left side is tapped at intervals so that its resonance can be adjusted to match the carrier frequency of the radio signal. Other frequencies are grounded (at the bottom of the schematic). The diode passes the "top half" of the signal to the earphone at the right, which is incapable of responding fast enough to reproduce the carrier frequency, and thus filters it out, leaving only the audio frequencies that were superimposed on it.*

By adding a capacitor, you can tune the circuit. Now an incoming pulse from the transmitter is initially blocked by the self-inductance of the coil, while it charges the capacitor. If an equally negative pulse is received after an interval that is properly synchronized with the values of the coil and the capacitor, it coincides with the capacitor discharging and the coil conducting. In this way, the right frequency of carrier wave makes the circuit resonate in sympathy. At the same time, audio-frequency fluctuations in the strength of the signal are translated into fluctuations in voltage in the circuit.

What happens to other frequencies pulled in by the antenna? The lower ones pass through the coil to ground; the higher ones pass through the capacitor to ground. They are just "thrown away."

The righthand half of the circuit samples the signal by passing it through a germanium diode and an earphone. The power from the transmitter is just sufficient to vibrate the diaphragm in the earphone, after the diode has subtracted the negative half of the signal.

Look back at the diagram of the amplitude-modulated signal. You'll see that it fluctuates up and down so rapidly, the earphone cannot possibly keep up with the positive-negative variations—hence the need for the diode. It will remain hesitating at the midpoint between the highs and lows, producing no sound at all. The diode solves this problem by subtracting the lower half of the "audio envelope," leaving just the positive spikes of voltage. Although these are still very small and rapid, they are now all pushing the diaphragm of the earphone in the same direction, so that it averages them out, approximately reconstructing the original sound wave.

Figure 5-72 shows how the circuit can be enhanced with a variable capacitor, to tune it without needing to tap the coil at intervals.

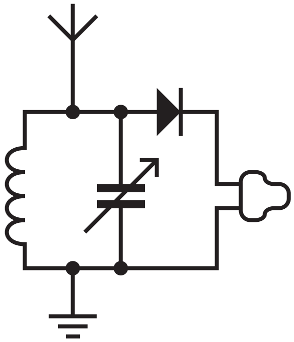

Figure 5-72. *By adding a capacitor to the circuit, its resonance can be tuned more precisely. The diagonal arrow indicates that a variable capacitor is used.*

The radio can pull in the stations on the AM (amplitude-modulated) waveband that happen to be most powerful in your area. The waveband ranges from 300 kHz to 3 MHz. If you find yourself interested in radio, your next step could be to build a powered radio using a couple of transistors. Alternatively you could build your own (legal) low-power AM transmitter. There's an ultra-simple kit available from *http://www.scitoys.com* consisting of just two principal components: a crystal oscillator, and a transformer, shown in Figure 5-73. That's all it takes.

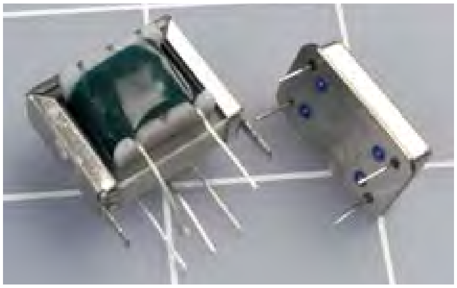

Figure 5-73. *An AM radio transmitter can be made from just two components: a transformer (left) and a crystal oscillator (right), available from http://www.scitoys.com.*

Experiment 32: A Little Robot Cart

Robotics is another application of electronics that deserves a book in itself—or several books. So, once again, I'm going to give you an introduction followed by some points that you can follow if you want to go further. As always, I will start with the simplest possible device, which in the world of robotics is a cart that finds its way around your living room.

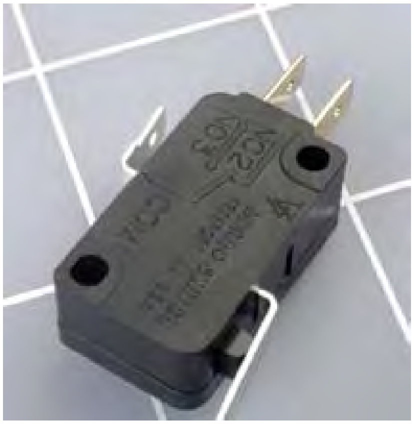

Figure 5-74. *A microswitch has a small button (at the front, righthand side in this picture) that is often actuated by a pivoted metal lever. The switch can respond to a very light pressure, but can handle relatively high currents.*

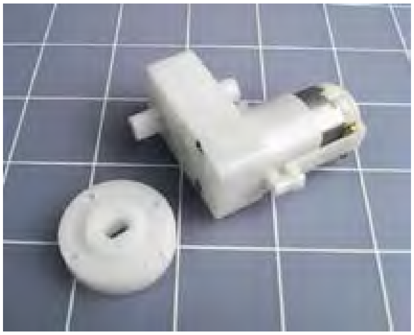

Figure 5-75. *For the Little Robot Cart, I found this 5-volt motor, which is supplied with a disc that fits its output shaft. The combination cost less than $10.*

You will need:

- SPST or SPDT microswitches requiring minimal pressure to activate them. A force between 0.02 and 0.1 newtons would be ideal. Quantity: 2. See Figure 5-74.
- DC gear-motor, rated for any voltage between 5 and 12, drawing a maximum of 100mA in its free-running state, output shaft turning between 30 and 60 RPM. Quantity: 1. A motor is shown in Figure 5-75.
- Disc or arm that fits securely onto your motor shaft. Quantity: 1.
- 555 timer. Quantity: 1.
- DPDT nonlatching relay rated for the same voltage as your motor. Quantity: 1.
- 1/4-inch plywood or plastic, one piece about 2 feet square.
- #4 sheet-metal screws, 5/8 inch or 3/4 inch long. Quantity: 2 dozen.
- #6 bolts, 3/4 inch long, with nylon-insert lock nuts. Quantity: 2 dozen.
- 1/4-inch bolts, 1 inch long, with nuts, to mount the wheels. Quantity: 4.

I'm not specifying one particular motor, because if I did, it might not be available by the time you read this. Motors aren't like logic chips, which have retained their basic function throughout various improvements over a period of several decades. Motors come and go, and many that you may run across will be surplus parts that will never been seen again. Search online for "gearmotor" or "gearhead motor" and find one as close as possible to the specification that I have provided. The mechanical power output of the motor shouldn't be important, because we won't be requiring it to do much work.

The important consideration when you buy your motor is that you should also obtain something that fits onto its output shaft. Typically, this will be a disk or arm that can be screwed into place. To this you can then add a larger wheel of your own, which you can cut with a hole saw or make from the screw-on lid of a jar, or anything else circular that you may find in the house.

A larger wheel will make your cart move faster than a smaller wheel, but will reduce its torque, thus limiting its power to overcome obstacles.

This brings me to my next topic: fabrication. Although this is an electronics book, motors are electromechanical devices, and you have to be able to install them in some kind of a machine to get any interesting results. You can use plywood to complete the two little robotics projects here (ideally, the kind of thin, high-quality plywood sold by hobby stores) but I recommend something that looks better and is easier to work with: ABS plastic. Before you start on the robotic cart, you may want to check the section "Fundamentals: All about ABS."

FUNDAMENTALS

All about ABS

Unless you think the steampunk movement isn't going back far enough, you probably don't want your autonomous robot cart to resemble a relic from before the 1800s. Therefore, wood may not be the best construction material. Metal can look nice, but is not easy to work with. For quick results that have a twentieth-century look (maybe even a 21st-century look), plastic is the obvious choice, and I feel that ABS is the best type of plastic to use, because it provides such quick, easy results. ABS stands for "acrylonitrile butadiene styrene." Lego® blocks are made of ABS. Car-stereo installers and model-railroad buffs use it. You can use it, too. You can saw it, drill it, sand it, whittle it, and drive screws into it, and it won't warp, split, or splinter. It's washable, doesn't need to be painted, and will last almost forever.

Delrin is another type of plastic, but tends to cost more and is a little tougher to drill and cut. It's a matter of personal preference. ABS machines fairly well, but when you drill it, for example, it can "catch" on the bit and the piece will spin with the bit due to the way that plastic chips off with the bit. Delrin is self-lubricating and has better melting properties under the heat of machining, so it drills and cuts much more cleanly and easily than ABS.

Where to find ABS

Pieces of ABS a couple of feet square are available from online sources such as *http://hobbylinc.com* or estreetplastics (an eBay store), but you'll save money if you can truck on down to your nearest plastic supply house and buy it like plywood, in sheets measuring 4 by 8 feet. To discover whether you have a nearby plastic supply house, search for "plastic supply" in your yellow pages or Google Local.

Piedmont Regal Plastics has many supply centers around the nation, but you'll have to collect it yourself, and they may not be willing to cut small pieces. You can check online at *http://www.piedmontplastics.com* for their locations.

Stock colors of ABS include black, white, and "natural," which is beige. Sheets usually are textured on one side, which is the side that should face outward, as it is more scratch-resistant than the smooth side.

Because you won't be adding paint or other finishes, you'll have to be careful not to scuff the plastic or scratch it while working. Clean your bench thoroughly before you begin, taking special care to remove any metal particles, which tend to become embedded in the plastic. Use wooden shims in the jaws of your vise, and avoid resting the plastic accidentally on any sharp tools or screws. Working with ABS requires a clean environment and a very gentle touch.

Cut with Care

You can saw ABS, but if you use a table saw, the plastic will tend to melt and stick to the blade. These smears will get warm and sticky when you feed the next piece of plastic into the saw, and the result will be extremely unpleasant. The whirling blade will grab the plastic and hurl it at you powerfully enough to break bones. This is known as "kickback" and is a very serious risk when sawing plastic.

If you have extensive experience using a table saw, you are actually more vulnerable, because the reflexes and cautions you have developed while dealing with wood will not be adequate for working with plastic. Please take this warning seriously!

Your first and most obvious precaution is to use a plastic-cutting blade, which has a larger number of thicker teeth to absorb the heat. The blade I have used is a Freud 80T, but there are others. If you use a blade that is not suitable, you will see it starting to accumulate sticky smears. This is the only warning you will get. Clean that blade with a solvent such as acetone, and never use it for ABS again.

Regardless of other precautions, always wear gloves and eye protection when using a table saw, and stand to one side when feeding materials into it. Personally, after one episode of kickback that I thought had broken my arm, I prefer not to use a table saw on plastic at all.

For long, straight cuts, the alternatives include:

- *Panel saw (big and expensive, but safe and accurate).*
- *Miniature handheld circular saw with a blade around 4 inches in diameter, guided with a straight edge clamped to the sheet.*
- *Hand saw. This is my old-school preference. My favorite is a Japanese pull-to-cut saw, which makes very clean cuts: the Vaughan Extra-Fine Cross-Cut Bear Saw, 9-1/2 inches, 17 tpi (teeth per inch). If you use one of these, be careful to keep your free hand out of the way, as the saw can easily jump out of the cut. Because it is designed to cut hard materials such as wood, it has no difficulty cutting soft flesh. Gloves are strongly recommended.*

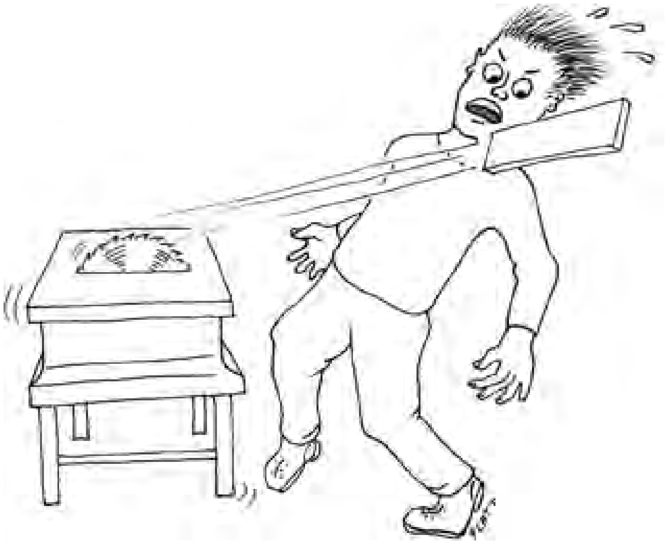

Figure 5-76. *The perils of kickback. Plastic easily sticks to the blade of a table saw, which will hurl it at you unexpectedly. Use other tools to cut plastic.*

Curving Cuts

Curving cuts involve relatively little danger, although eye protection and gloves are still advisable. My preferred tools:

- Band saw with a 3/8- or 1/4-inch blade designed for thin wood or plywood.
- Jigsaw. I have a special liking for the DeWalt XRP using Bosch blades that are designed for hardwood or plastic. This will cut complex curves in ABS as easily as scissors cutting paper.

No matter what type of saw you use, you'll have to clean ragged bits of plastic off the cut afterward, and the absolutely necessary item for this purpose is a deburring tool, available from *http://www.mcmaster.com* and most other online hardware sources. A belt sander or disc sander is ideal for rounding corners, and a metal file can be used to remove bumps from edges that are supposed to be straight.

Figures 5-77 through 5-80 show various cutting tools. Figure 5-81 shows a deburring tool, and Figure 5-82 shows a disc sander.

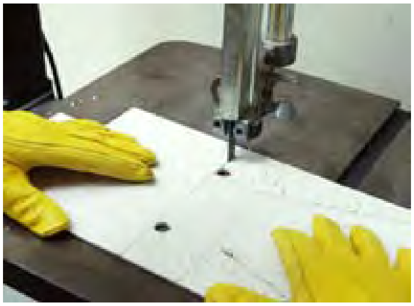

Figure 5-77. *A band saw is an ideal tool for cutting complex shapes out of ABS plastic. You can often find them secondhand for under $200.*

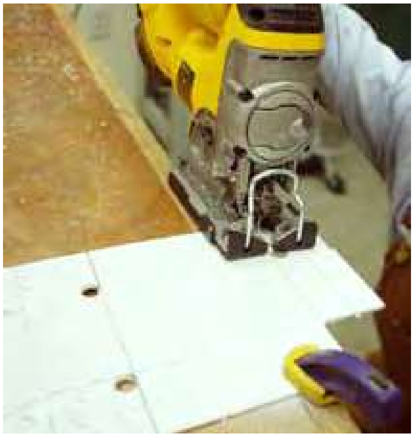

Figure 5-80. *This DeWalt jigsaw can run at very slow speeds, enabling precise and careful work with plastic.*

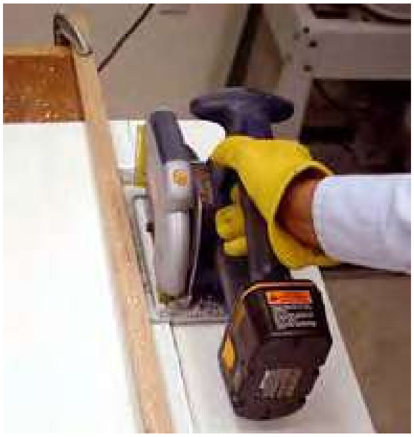

Figure 5-78. *A handheld circular saw, running along a straight-edge, is much safer than a table saw for cutting plastic, and can produce comparable results.*

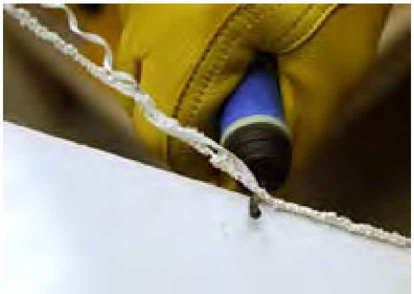

Figure 5-81. *A deburring tool will clean and bevel the sawn edge of a piece of plastic in just a couple of quick strokes.*

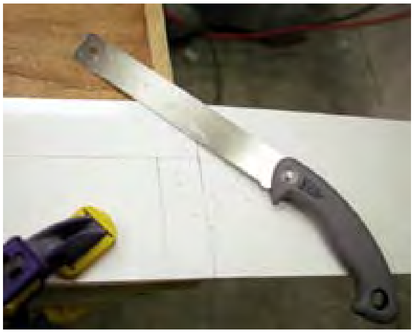

Figure 5-79. *This Japanese-style saw cuts when you pull it, rather than when you push it. After some practice, you can use it to make very accurate cuts. Because ABS is so soft, minimal muscle-power is required.*

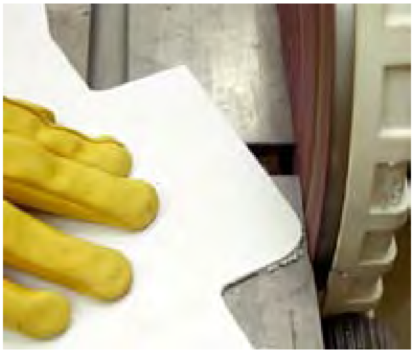

Figure 5-82. *A belt sander or disc sander is the ideal tool for rounding corners when working with ABS plastic.*

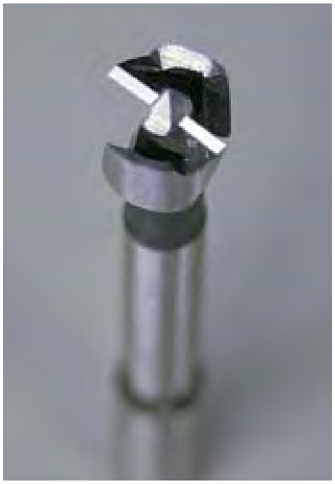

Figure 5-83. *A Forstner drill bit makes clean, precise holes; a large regular drill bit will chew up ABS plastic and make a mess.*

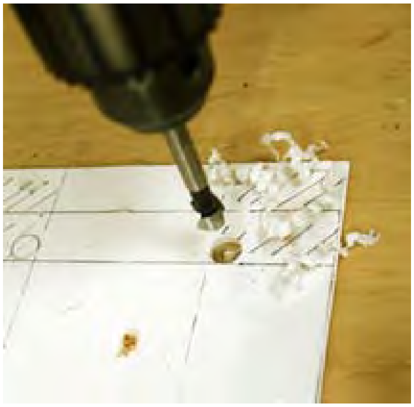

Figure 5-84. *By drilling holes at any location where two bends intersect, you reduce the risk of the plastic fissuring.*

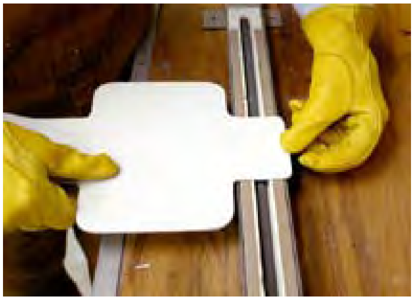

Figure 5-85. *Making clean, precise bends in ABS is simply a matter of resting the plastic over a bender that consists of an electric heating element.*

Making Plans

I like to use drawing software to create plans, and I try to print them at actual scale. I tape them to the smooth side of a piece of white or natural-color ABS, then use an awl to prick through the plan into the soft surface beneath. I remove the paper and connect the awl marks by drawing onto the plastic using a pencil or a fine-point water-soluble pen. Its lines can be wiped away later with a damp cloth. Don't use a permanent marker, as the solvents that you will need to clean it may dissolve the plastic.

Because ABS will tend to open a fissure when you bend it at any inside corner where you don't have a smooth radius, you need to drill holes at these locations, as shown in the cart plans in Figure 5-92 on page 275.

A regular half-inch drill bit is too aggressive; it will tend to jam itself into the plastic within one turn of the drill. Use Forstner bits (shown in Figures 5-83 and 5-84) to cut nice smooth circles.

Note that the heat from bending will tend to make any marks on the plastic permanent.

Bending It

A big advantage of plastic over wood is that you can make complex shapes by bending them, instead of cutting separate pieces and joining them with nails, screws, or glue. Unfortunately, bending does require an appropriate bender: an electric heating element mounted in a long, thin metal enclosure that you place on your workbench. The bender I use is made by FTM, a company that offers all kinds of neat gadgets for working with plastic. Their cheapest bender, shown in Figure 5-85, is just over $200 with a 2-foot element. You can get a 4-foot model for about $50 more. Check them out at *http://thefabricatorssource.com*.

Avoid Burns While Bending

A plastic bender will inflict serious burns if you happen to rest your hand on it accidentally, and because it has no warning light, you can easily forget that you have left it plugged in. Use gloves!

To bend plastic, lay it over the hot element of a plastic bender for a brief time (25 to 30 seconds for 1/8-inch ABS, 40 to 45 seconds for 3/16-inch, and up to a minute for 1/4-inch). If you overheat the plastic, you'll smell it, and when you turn it over you'll find it looks like brown melted cheese. Naturally you should learn to intervene before the plastic reaches that point.

ABS is ready to bend when it yields to gentle pressure. Take it off the bender and bend it *away* from the side that you heated. If you bend it toward the hot side, the softened plastic will bunch up inside the bend, which doesn't look nice.

You can work with it for about half a minute, and when you have it the way you want it, spray or sponge water onto it to make it set quickly. Alternatively, if you need more time, you can reheat it. The amount of force necessary to bend the sheet increases in proportion with the length of the bend, so a long bend can be difficult, and I usually insert it into a loose vise, push it a bit, move it along to the next spot, and push it again.

Because plastic bending is very similar to making shapes in origami, it's a good idea to model your projects in paper before you commit yourself to ABS.

If you decide that you don't want to spend money on a bender, don't abandon plastic just yet—you can use screws to assemble separate sections with greater ease and convenience than if you were working with wood.

Making 90-Degree Joints

Driving screws into the edge of a piece of plywood will almost always separate its layers, but ABS has no layers (or grain, either), and never splits or shatters. This means that you can easily join two pieces at 90° using small screws (#4 size, 5/8-inch long).

Figures 5-86 through 5-90 show the procedure for joining 1/8-inch (or thicker) ABS to 1/4-inch ABS, which I regard as the minimum thickness when you're inserting screws into its edge:

1. Mark a guideline on the thinner piece of plastic, 1/8 inch from its edge. For #4 screws, drill holes using a 7/64-inch bit. If you're using flat-headed screws, countersink the holes very gently.

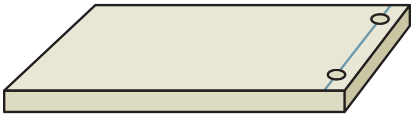

Figure 5-86

2. Hold or clamp the pieces in place and poke a pen or pencil through the holes to mark the edge of the 1/4-inch plastic beneath.
3. Remove the thin plastic, clamp the 1/4-inch plastic in a vise, and drill guide holes for the screws at each mark, centered within the thickness of the plastic. Because ABS does not compress like wood, the holes must be larger than you may expect; otherwise, the plastic will swell around the screw. A 3/32-inch bit is just right for a #4 screw.
4. Assemble the parts. Be careful not to overtighten the screws; it's easy to strip the threads that they cut in the soft plastic.

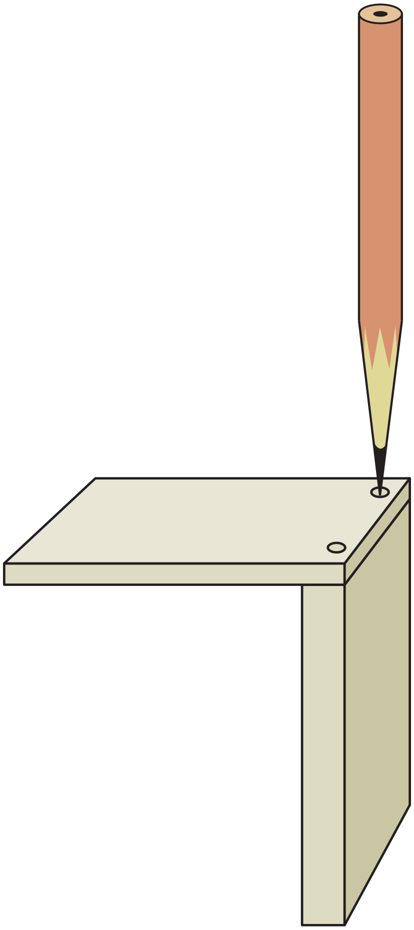

Figure 5-87

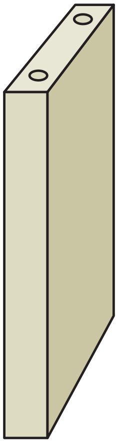

Figure 5-88

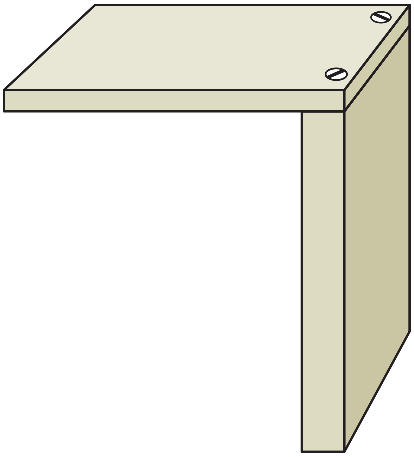

Figure 5-89. *Figures 5-86 through 5-89 illustrate four steps to join two pieces of ABS using #4 sheet-metal screws. Cut 7/64-inch holes on a line 1/8 inch from the edge of the first piece, then mark through the holes to the edge of the second piece. Drill 3/32-inch holes that are precisely centered in the edge, then screw the pieces together.*

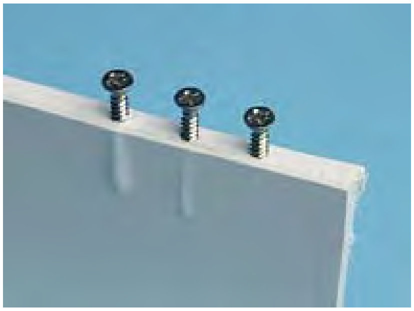

Figure 5-90. *Three #4 screws driven into the edge of ABS, using a 1/16-inch guide hole, a 5/64-inch guide hole, and a 3/32-inch guide hole. respectively Because the first two guides holes were too small, the plastic swelled around the screw (but did not break).*

Framing Your Cart

For reasons that will soon be apparent, I've chosen an unusual diamond-shaped configuration of wheels. In the rendering shown in Figure 5-91, the front wheel (at the far end of the cart) applies power, the rear wheel (at the near end of the cart) steers the cart when backing up, and the side wheels prevent it from falling over.

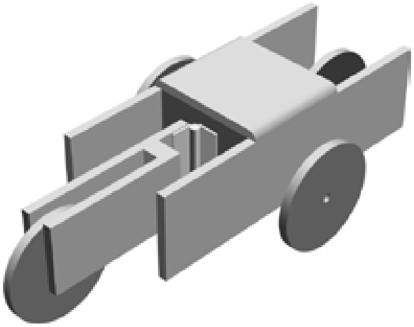

Figure 5-91. *If you have 3D rendering software, it can be a great way to test the feasibility of a construction project before you start cutting materials and trying to fit pieces together. This rendering was a proof-of-concept for the Little Robot Cart.*

Depending on the type of motor that you buy, you'll have to improvise a way to mount it in the front section of the cart. Don't be afraid to use kludges such as cable ties, duct tape, or even rubber bands to attach the motor to the frame. We're making a rough prototype, here, not a thing of beauty (although if you decide you like the cart, you can always rebuild it beautifully later).

The plan in Figure 5-92 shows the pieces that you will need. Part A is the body of the cart. If you're going to bend it from ABS, you should drill half-inch holes, with a forstner bit, at the four inside corners, so that these corners have rounded edges. If you simply saw the plastic to make sharp 90° corners, the plastic may develop fissures at the corners when you bend it. If you don't have a plastic bender and don't feel inclined to buy one, you can make Part A from three separate rectangles and then screw them together.

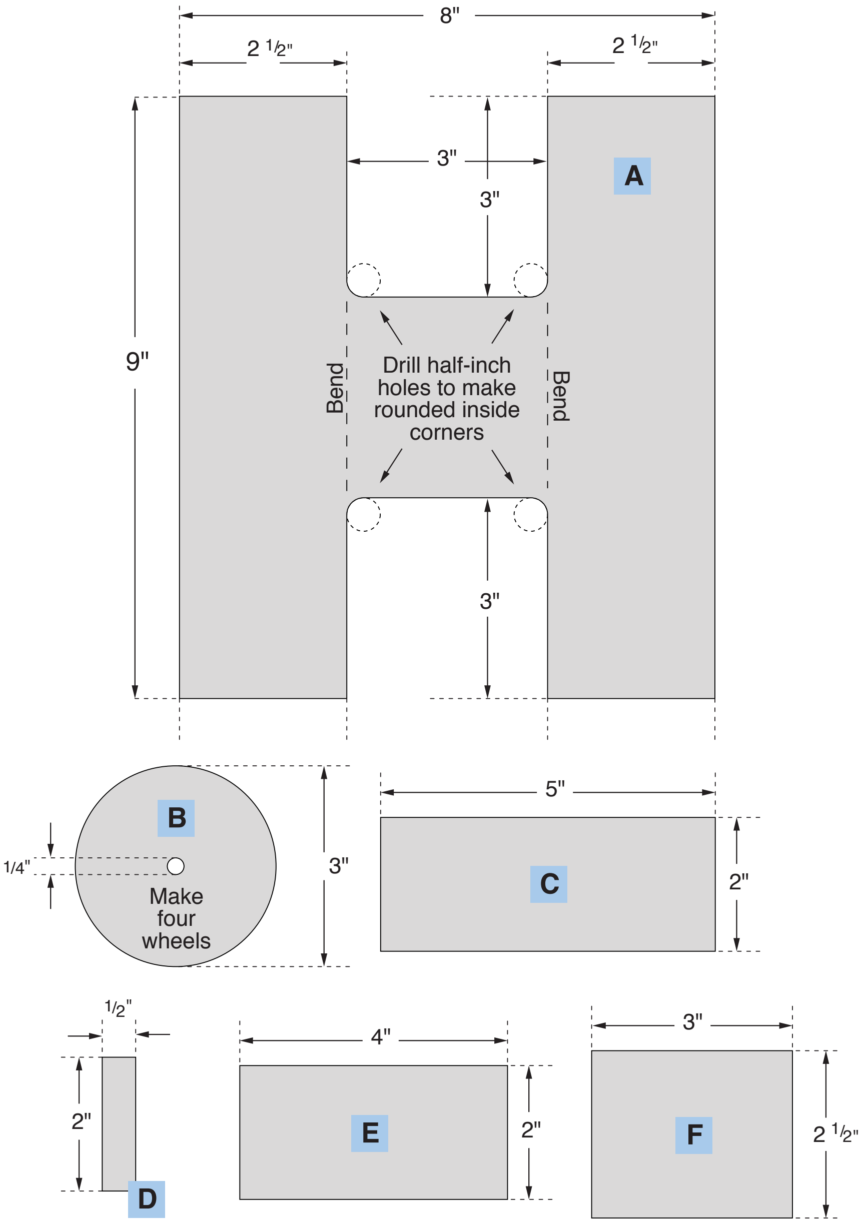
8"
2 1/2"
2 1/2"
A
3"
3"
9"
Bend
Bend
Drill half-inch holes to make rounded inside corners
3"
B
1/4"
Make four wheels
3"
5"
C
2"
1/2"
2"
D
4"
E
2"
3"
F
2 1/2"

Figure 5-92. *These sections of 1/4-inch plastic can be assembled to create the simple cart described in Experiment 31.*

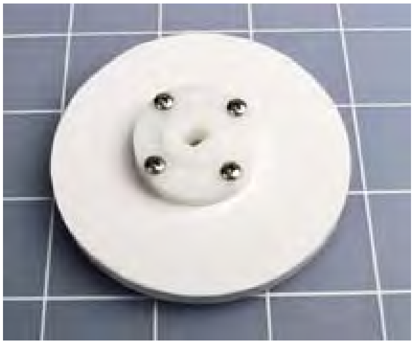

Figure 5-93. *A 3-inch wheel is screwed to the disc that mates with the drive shaft of the motor.*

Part B is a wheel, of which you will need four. I cut them using a 3-inch hole saw. The front wheel is screwed to whatever disc or arm you obtained to mate with the shaft of your motor. See Figure 5-93.

Parts C, D, and E assemble to form a yoke in which the rear wheel is mounted. I used a 2-inch hinge to pivot the yoke. The hinge is mounted on Part F, which is a partition located midway in the frame of the cart. The photographs in Figures

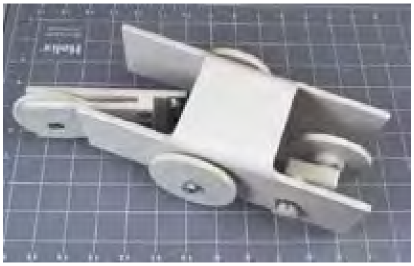

Figure 5-94. *The assembled body of the cart, before adding control electronics. The wheel at the righthand end will pull the cart from left to right. The hinged trailing wheel will allow the cart to move in a relatively straight line when it moves forward, but will tend to turn it when it backs up.*

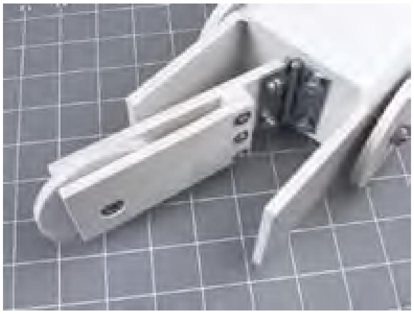

Figure 5-95. *A closeup of the hinged trailing wheel, which rotates freely and can flip from side to side with minimal friction.*

5-94 and 5-95 will help to make this clear. Initially, when you install Part F, use only two screws, one each side, so that you can adjust its angle a little. This will be necessary to optimize the contact of the wheels with the floor.

The side wheels and rear wheel must spin freely, but on the other hand, they shouldn't wobble. I simply tightened the nuts on the bolts that serve as axles for the wheels, until there was maybe half a millimeter of clearance. I added a drop of Loctite to stop the nuts from getting loose.

The plans don't show precisely where to drill holes for the axle bolts, because the location will depend on the size of your wheels. You can figure this out as you go along. Just make sure that the side wheels aren't mounted too low. We don't want them to lift the front wheel or the rear wheel off the floor. If the side wheels are a fraction higher off the ground than the front and rear wheels, that's good.

If you have tile or wood floors, your cart may acquire better traction if you wrap a thick rubber band around each disc that you use for the drive wheel and the steering wheel.

The most important aspect of the construction is to place microswitches where they'll be triggered when the cart runs into something. I placed mine at the front corners, as shown in Figures 5-96 and 5-97. And that brings me to the electronics.

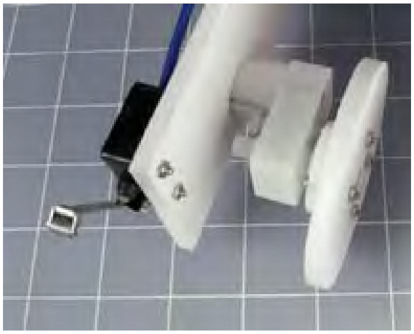

Figure 5-96

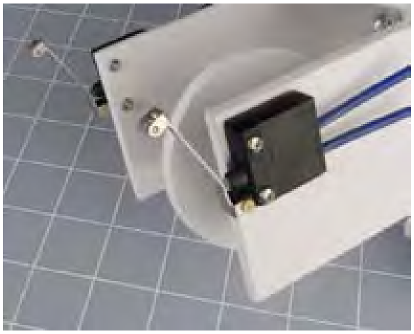

Figure 5-97. *Two microswitches with metal arms are mounted on each side of the cart, where they will sense any obstacle.*

The Circuit

The schematic is very, very simple, with only four principal components: two microswitches that sense obstacles in front of the cart, one relay, and one 555 timer. You will also need a small power switch, a battery or battery pack, and a resistor, and capacitors to go with the timer. A trimmer potentiometer will allow you to adjust the "on" time of the 555 timer, which will determine how long the cart takes to back up. See Figure 5-98.

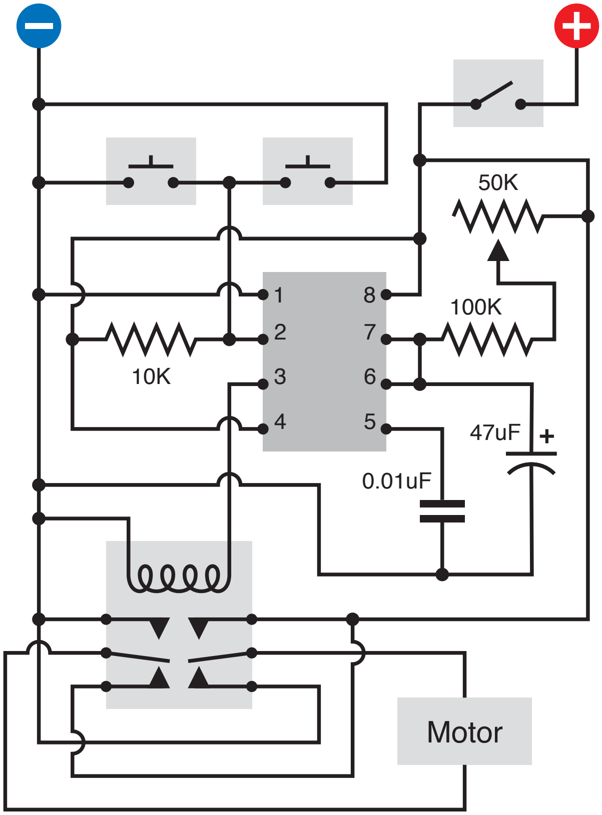

Figure 5-98. *This ultrasimple schematic is all the cart needs to enable it to back up when it hits an obstacle.*

The motor I chose requires 5 volts, so I had to use a voltage regulator with a 9-volt battery. If your motor uses 6 volts, you can wire four AA batteries to it directly. If you have a 12-volt motor, you can use two 9-volt batteries in series, supplying power through a 12-volt voltage regulator.

Assemble the components, mount them on the cart, and switch it on, and it should move forward slowly in a more-or-less straight line. If it moves backward, reverse your connection to the terminals on the motor.

When the cart bumps into something, either of the microswitches will connect negative voltage to the input pin of the 555 timer. This triggers the timer, which runs in monostable mode, generating a single pulse lasting about 5 seconds, which closes the relay, which is wired so that it reverses the voltage to the motor.

When the voltage is reversed to a simple DC motor, it runs backward. So the cart backs up. Because the rear wheel is mounted in a yoke that pivots, the yoke will tend to flip one way or the other, causing the cart to describe an arc as it moves backward. At the end of the timer cycle, the relay relaxes and the cart starts moving forward again. In forward mode, the rear wheel just follows along without applying any steering force, so the cart tends to follow a straight line—until it hits another obstacle, at which point it backs up, and tries another path.

FUNDAMENTALS

All about limit switches

The most obvious enhancement for your cart would be a better steering mechanism. You could use another motor to take care of this, with a pair of limit switches. Because limit switches are a basic, important idea in conjunction with motors, I'll explain them in detail.

Figure 5-99 shows three successive views of a motor with an arm attached to it, which can press either a lower pushbutton or an upper pushbutton. Both of the pushbuttons are normally closed, but will open when pressed by the motor arm. These buttons are the limit switches. Typically you would use microswitches for this purpose, just like the ones that I suggested as barrier-sensors at the front of the cart.

In addition, there's a DPDT relay that is activated by a simple on/off switch at the righthand side. On the cart, the 555 timer takes the place of the on/off switch, by feeding power to the relay.

Suppose that the motor begins with the arm pointing downward, as shown in the top view in Figure 5-99, and the motor is wired so that when it receives negative voltage at its lower terminal and positive at its upper terminal, it rotates counter-clockwise. This is what happens when the on/off switch closes and sends power to the DPDT relay. Positive voltage from the relay contacts cannot pass through the upper diode, but can pass through the upper limit switch, which is closed. Negative voltage cannot pass through the lower limit switch, because it's open, but can pass through the lower diode. So, the motor starts to turn counterclockwise. During the midpoint of its arc, it receives power through both of the limit switches.

Finally, the motor arm reaches the upper switch, and opens it. This prevents positive voltage from reaching the motor through that switch, and the positive voltage is also blocked by the upper diode. So, at this time, the motor stops.

Now suppose that the on/off switch is opened, as in the top view in Figure 5-100. The relay loses its power, so its contacts relax. The voltage to the motor is now reversed. Negative voltage passes through the upper diode, while positive voltage reaches the motor through the lower limit switch. The motor starts running clockwise, until its arm hits the lower switch, opening it and cutting off power to the motor.

Limit switches are necessary, because if you continue to apply voltage to a simple DC motor that is unable to turn, the motor sucks more current, gets hot, and may burn out.

You can easily see how this kind of system could be used to control the cart's steering. Even though the motor has only two positions, these are sufficient to make the cart turn when going backward, and proceed straight ahead when going forward.

To reduce power consumption, the DPDT relay could be replaced with a two-coil latching relay. The circuit would then have to be revised so that the relay is flipped to and fro by a pulse to each of its coils.

FUNDAMENTALS

All about limit switches (continued)

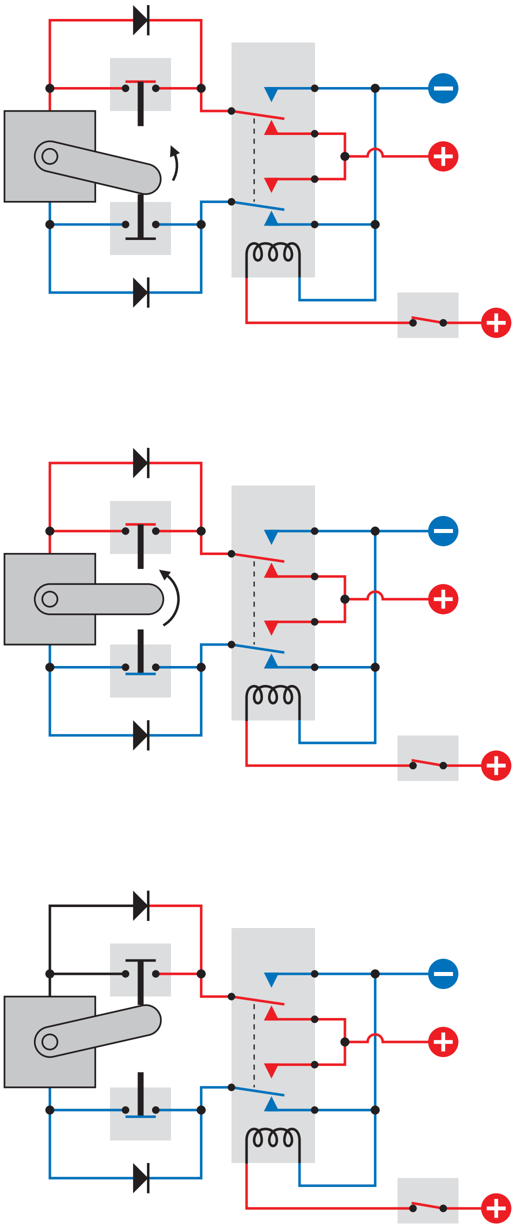

Figure 5-99. *The three diagrams, from top to bottom, show three snapshots of a motor controlled by a DPDT relay and two limit switches. When the on/off switch at bottom-right sends power to the relay, the lower relay contacts cause the motor to run counterclockwise until it stops itself as its arm opens the upper limit switch.*

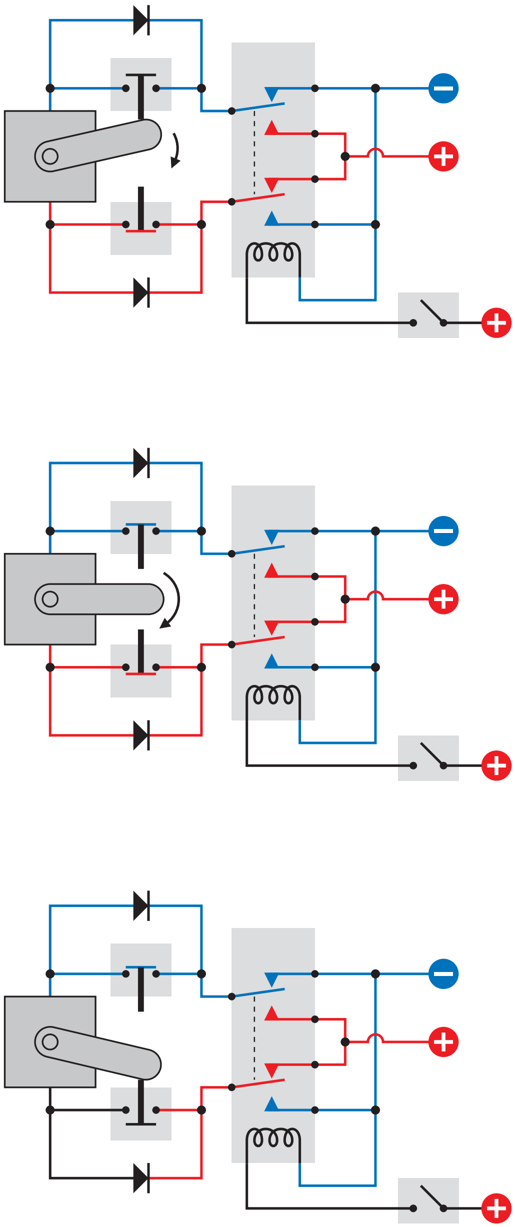

Figure 5-100. *When the on/off switch at bottom-right opens, the relay connects its upper contacts. This causes the motor to run clockwise until its arm opens the lower limit switch. Limit switches avoid the overheating and possible damage that are likely when power is delivered to a motor that is prevented from turning.*

FUNDAMENTALS

All about motors

Brushed DC motor

This is the oldest, simplest design for an electric motor, shown in very simplified form in Figure 5-101. Coils are attached to a shaft where they can interact with stationary magnets around them. The magnetic attraction turns the shaft a little, at which point the next coil on the shaft is energized to turn the shaft a little more, and then the next coil—and so on. To make this happen, electricity has to be fed into the coils by "brushes," often consisting of soft carbon pads that conduct power to a hub, known as a commutator, divided into sections, each of which is connected to a separate coil.

This basic design has several advantages if we want to build a small motorized gadget, such as a miniature robot or even a model airplane:

- Widely available
- Low cost
- Simple
- Reliable
- Will run in reverse when voltage reverses

In addition, brushed motors are often sold with reduction gearing built in. Such units are known as *gearhead motors* or *gear motors*. They free you from the need to use your own gears or belts to adjust the output speed yourself. You simply choose the motor that fits your specification.

DC stepper motor

This requires a controller, consisting of some electronics to tell the motor to rotate its shaft in small, discrete steps. The advantages of a stepper motor are:

- Precise positioning of the shaft
- Precise speed adjustment

Stepper motors are ideal for devices such as computer printers, where the paper has to roll up by a precise distance and the print head has to move laterally by an equally precise distance, but they are also useful in robots. If the motor is small enough to draw less than 200mA and will run on 12 volts or less, you can control it with pulses from a 555 timer. I'll describe stepper motors in more detail in Experiment 33.

Servo motor

This is generally used in conjunction with a programmable microcontroller, which sends instructors to rotate the motor shaft to a specific position and then hold it there. I'll mention servo motors when I introduce you to microcontrollers, but we won't be dealing with them in detail.

Other types of motors exist, including brushless DC motors (which require a different type of controller and are found in computer disk drives and CD players), and AC motors (including synchronous motors, which synchronize their rotation with the frequency of AC voltage, and were used extensively in clocks, before clocks mostly became digital).

In this book, I'll be talking mostly about brushed DC motors and DC stepper motors.

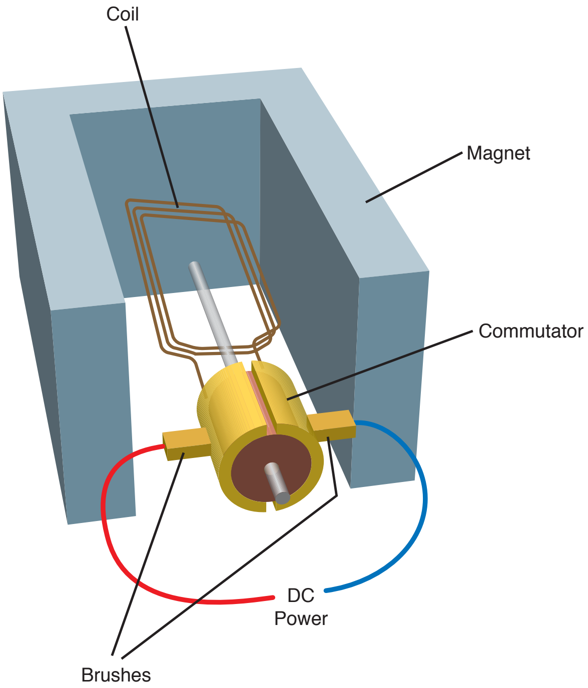

Figure 5-101. *The basic principle of a simple DC motor. The commutator passes electricity through a coil, creating a magnetic field that interacts with a magnet around the motor. The coil turns, and the commutator turns with it, until the electric field through the coil is reversed. This causes the process to repeat. In reality, a motor is likely to have a commutator formed from multiple segments, connected with multiple coils. The principle, however, remains the same.*

Take-home messages from this experiment include the following:

- You can buy simple DC motors with reduction gearing built in, providing your choice of RPM. Literally hundreds of websites will sell you small motors for robotics projects.
- When you reverse the voltage to a DC motor, the motor runs in reverse.
- A DPDT relay can be wired so that when it closes its contacts, it reverses a power supply to a motor.
- You can use two limit switches and a pair of diodes to stop a motor at two positions. In each of its stopping positions, the motor consumes no power and you won't have the risk of it burning out.

What other projects can you imagine using this simple set of techniques?

Mechanical Power

In the United States, the turning force, or torque, of a motor is usually measured in pound-feet or ounce-inches. In Europe, the metric system is used to measure torque in dynes.

A pound-foot is easy to understand. Imagine a lever pivoted at one end, as shown in Figure 5-102. If the lever is one foot long, and you hang a one-pound weight at the end of it, the turning force is one pound-foot.

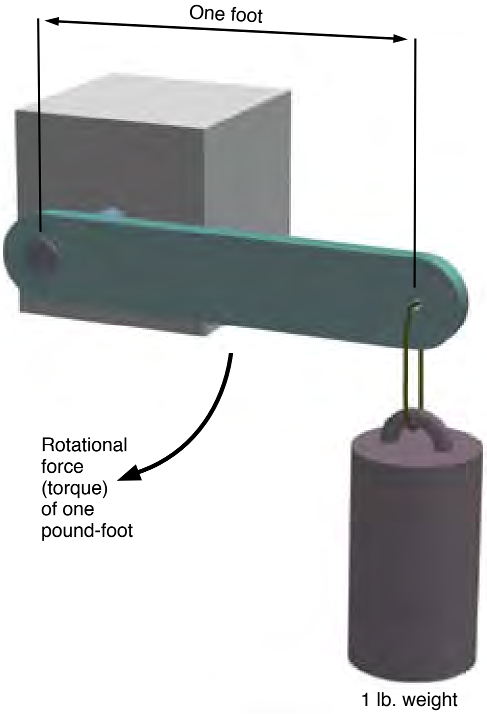

Figure 5-102. *The rotational force created by a motor is known as "torque," and in the United States it is measured in pound-feet (or ounce-inches, for small motors). In the metric system, torque is measured in dynes. Note that the torque created by a motor will vary according to the speed at which the motor is running.*

FUNDAMENTALS

Wire gauges

If you're going to power larger motors, or other components that take more current than LEDs or small relays, you really need to know about wire gauges. In particular, what's the relationship between wire thickness and AWG (American Wire Gauge)? And what gauge of wire should you use for any given current?

You can find numerous charts and tables if you go online, but many of these sources contradict each other, especially on the topic of how much current is safe to run through each gauge of wire.

After making several comparisons (and testing some wire samples myself), I've compiled the table in Figure 5-103, which I recommend as a compromise. Note the following:

- This table applies to solid-core copper wire.
- For stranded wire, or copper that has been tinned (giving it a silver appearance), the number of ohms per foot will *increase*, the number of feet per ohm will *decrease*, and the maximum amperage will *decrease*, probably by around 20%.

The maximum amperage assumes that the wire is insulated, preventing it from radiating heat as effectively as a bare conductor. I am also assuming that the wire is likely to be at least partially enclosed, inside a box or cabinet. At the amperages listed for each gauge of wire, you should expect the wire to become noticeably warm, and personally I would tend to use thicker wire instead of the maximums indicated in the table.

Most tables of this type only tell you the resistance of each gauge of wire in ohms per 1,000 feet. I have included that number but have also expressed the function the other way around, as the number of feet per ohm, as this doesn't require you to do so much arithmetic with decimals.

AWG		Diameter in inches	Ohms per 1,000 ft	Feet per ohm	Maximum amps (insulated)
0000	●	0.46	0.049	20,400	225
000	●	0.41	0.062	16,200	200
00	●	0.365	0.078	12,800	175
0	●	0.325	0.098	10,200	150
1	●	0.289	0.124	8,070	125
2	●	0.258	0.156	6,400	100
3	●	0.229	0.197	5,080	90
4	●	0.204	0.249	4,020	80
5	●	0.182	0.313	3,190	70
6	●	0.162	0.395	2,510	60
7	●	0.144	0.498	2,010	51
8	●	0.128	0.628	1,590	44
10	●	0.102	0.999	1,000	32
12	●	0.081	1.59	630	23
14	●	0.064	2.53	396	17
16	●	0.051	4.02	249	13
18	●	0.04	6.39	157	10
20	●	0.032	10.2	99	8
22	●	0.025	16.1	62	5
24	●	0.02	25.7	39	2.5
26	●	0.016	40.1	25	1.5
28	●	0.013	64.9	15	1.0
30	●	0.010	103.2	10	0.5

Figure 5-103. *American wire gauges (AWG) and their properties.*

THEORY

Calculating voltage drop

Another fact that you often need to know is how much of a voltage drop a particular length of wire will introduce in a circuit. If you want to get maximum power from a motor, you don't want to lose too much voltage in the wires that go to and from the motor.

Voltage drop is tricky, because it depends not only on the wire, but also on how heavily the circuit is loaded. Suppose that you are using 100 feet of 22-gauge wire, which has a resistance of about 1.5 ohms. If you attach it to a 12-volt battery and drive an LED and a series resistor offering a total effective resistance of about 1,200 ohms, the resistance of the wire is trivial by comparison. According to Ohm's Law:

amps = volts / ohms

so the current through the circuit is only about 10mA.

Again, by Ohm's Law:

volts = ohms × amps

so the wire with resistance of 1.5 ohms imposes a voltage drop of 1.5 × 0.01 = 0.015 volts.

Now suppose you're running a motor. The coils in the motor create impedance, rather than resistance, but still if we measure how much current is going through the circuit, we can establish its effective resistance. Suppose the current is 1 amp. Repeating the second calculation:

volts = ohms × amps

So the voltage drop in the wire is now 1.5 × 1 = 1.5 volts! This is illustrated in Figure 5-104.

Bearing these factors in mind, I have compiled a table for you. I've rounded the numbers to just two digits, as variations in the wire that you use make any pretense of greater accuracy unrealistic.

To use this table, you need to know how much current is passing through your circuit. You can calculate it (by adding up all the resistances and dividing it into the voltage that you are applying) or you can simply measure the current with a meter. Just make sure that your units are consistent (all in ohms, amps, and volts, or milliohms, milliamps, and millivolts).

In the table, I have arbitrarily assumed a length of 10 feet of wire. Naturally you will have to make allowances for the actual length of wire in your circuit. The shorter the wire, the less the loss will be. A circuit with only 5 feet of wire, and the same amperage and voltage, will suffer half of the percentage loss shown in the table. A circuit with 15 feet of wire, and the same amperage and voltage, will suffer 1.5 times the percentage loss. So, to use the table:

1. Divide your length of wire by 10. (Make sure that you measure the length in feet.)
2. Use the result to multiply the number in the table.

The table also arbitrarily assumes that you have a 12-volt supply. Again, you will have to make allowances if you are using a different voltage. So, to use the table:

1. Divide 12 by the actual voltage of your power supply.
2. Use the result to multiply the number in the table.

I can summarize those two steps like this:

Percent voltage lost = P × (12 / V) × (L / 10)

where P is the number from the table, V is your power-supply voltage, and L is the length of your wire.

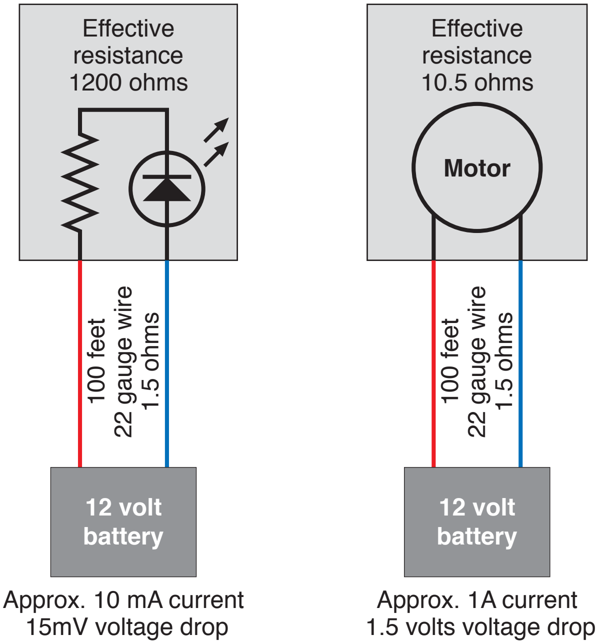
Effective resistance 1200 ohms
100 feet 22 gauge wire 1.5 ohms
12 volt battery
Approx. 10 mA current 15mV voltage drop
Effective resistance 10.5 ohms
Motor
100 feet 22 gauge wire 1.5 ohms
12 volt battery
Approx. 1A current 1.5 volts voltage drop

Figure 5-104. *The voltage drop imposed by wiring will depend on the current and the resistance in the circuit. The drop will be greatest when the resistance of the circuit is low and the amperage is high.*

THEORY

Calculating voltage drop (continued)

This table shows the percent voltage lost in a circuit with 10-foot wire at 12 volts.

Wire Gauge	Amperes									
	1	2	3	4	5	6	7	8	9	10
10	0.08	0.17	0.25	0.33	0.42	0.50	0.58	0.67	0.75	0.83
12	0.13	0.27	0.40	0.53	0.66	0.80	0.93	1.1	1.2	1.3
14	0.21	0.42	0.63	0.84	1.1	1.3	1.4	1.5	1.9	2.1
16	0.33	0.67	1.0	1.3	1.7	2.0	2.3	2.7	3.0	3.4
18	0.53	1.1	1.6	2.1	2.7	3.2	3.7	4.3	4.8	5.3
20	0.85	1.7	2.6	3.4	4.3	5.1	6.0	6.8	7.7	8.5
22	1.3	2.7	4.0	5.4	6.7	8.1	9.4	11	12	13
24	2.1	4.3	6.4	8.6	11	13	15	17	19	21
26	3.4	6.8	10	14	17	20	24	27	31	34
28	5.4	11	16	22	27	32	38	43	49	54
30	8.6	17	26	34	43	52	60	69	77	86

Remember, though, that the wire resistance will be higher if you are using stranded copper wire or tinned copper wire, and this will increase the percentage of voltage lost.

Experiment 33: Moving in Steps

Time now to build something more sophisticated: a cart that orients itself toward a light source. I'm going to tell you all you need to get started on this project, but this time I won't go all the way to the end in exhaustive detail. I want you to get into the habit of figuring out the details, improving on plans, and eventually inventing things for yourself.

You will need:

- 555 timers. Quantity: 8.
- Trimmer potentiometer, 2K linear. Quantity: 2.
- LEDs. Quantity: 4. If you get tired of using series resistors to protect LEDs in a 12-volt circuit, consider buying 12-volt LEDs such as Chicago Miniature 606-4302H1-12V, which contain their own resistors built in. However, the schematic in Figure 5-108 assumes that you will use regular 2V or 2.5V LEDs.
- Stepper motor: Unipolar, four-phase, 12-volt. Parallax 27964 or similar, consuming 100mA maximum. Quantity: 2.
- Photoresistors, ideally 500 to 3,000Ω range. Quantity: 2.

- ULN2001A or ULN2003A Darlington arrays by STMicroelectronics. Quantity: 2.
- CMOS octal or decade counter. Quantity: 2.
- Various resistors and capacitors.

Exploring Your Motor

I've specified a unipolar, four-phase, 12-volt motor because this is a very common type. A typical sample is shown in Figure 5-105. If you can't easily find the one that I've listed, you should feel safe in buying any other that has the same generic description. "Unipolar" means that you don't have to switch the power supply from positive to negative and back to positive again, to run the motor. Four-phase means that the pulses that run the motor must be applied in sequence to four separate wires. Because you will be running your motor directly from 555 timers, the lower its power consumption, the better.

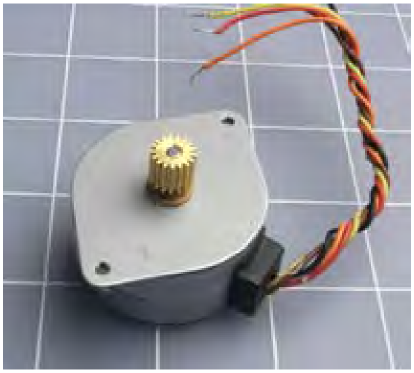

Figure 5-105. *A typical stepper motor. The shaft rotates in steps when negative pulses are applied to four of the wires in sequence, the fifth wire being common-positive.*

First, though, we can apply voltage to the motor without using any other components at all. Most likely it will have five wires already attached, with the ends stripped and tinned, so that you can easily insert them into holes in a breadboard, as shown in Figure 5-106. Check the data sheet for your motor; you should find that four of the wires are used to energize the motor and turn it in steps, while the fifth is the common connection. In many cases, the common connection should be hooked to the positive side of your power supply, while you apply negative voltage to the other four wires in sequence, one step at a time.

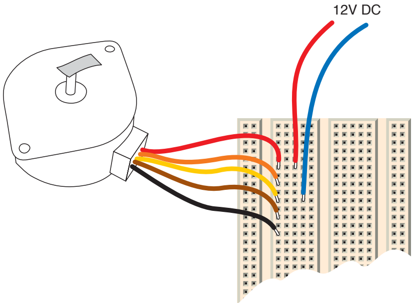

Figure 5-106. *The simplest test of a stepper motor is to apply voltage manually to each of its four control wires, while a piece of duct tape, attached to the output shaft, makes it easy to see how the motor responds.*

The data sheet will tell you in what sequence to apply power to the wires. You can figure this out by trial and error if necessary. One thing to bear in mind: a stepper motor is very tolerant. As long as you apply the correct voltage to it, you can't burn it out.

To see exactly what the motor is doing, stick a piece of duct tape to the end of the shaft. Then apply voltage to wires, one at a time, by moving your negative power connection from one to the next. You should see the shaft turning in little steps.

Inside the motor are coils and magnets, but they function differently from those in a DC motor. You can begin by imagining the configuration as being like the diagram in Figure 5-107. Each time you apply voltage to a different coil, the black quadrant of the shaft turns to face that coil. In reality, of course, the motor turns less than 90° from one coil to the next, but this simplified model is a good way to get a rough idea of what's happening. For a more precise explanation, see the upcoming section "Theory: Inside a stepper motor."

Bear in mind that as long as any of the wires of the motor are connected, it is constantly drawing power, even while sitting and doing nothing. Unlike a regular DC motor, a stepper motor is designed to do nothing for much of the time. When you apply voltage to a different wire, it steps to that position and then resumes doing nothing.

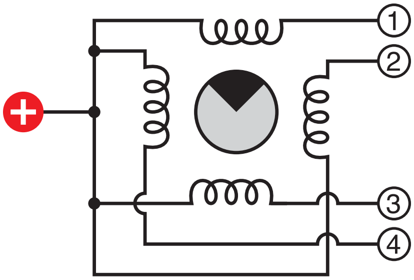

Figure 5-107. *This greatly simplified diagram helps in visualizing the way in which a stepper motor works. In reality, almost all motors rotate by less than 90° in response to each pulse.*

The coil inside the motor is holding the shaft in position, and the power that the motor draws will be dissipated as heat. It's quite normal for the motor to get warm while you're using it. The trouble is, if you use a battery to power it, and you forget that you have it connected, the battery will not hold its charge for long.

Quick Demo

Now that you've proved that your motor is functional, how can you actually run it? You need to send a pulse to each of the four wires in turn, in a rapidly repeating sequence. If you can also adjust the speed of the pulses, so much the better. I'm thinking that for a quick and simple demo, you can handle the challenge simply by using four 555 timers, all of them in monostable mode, with each one triggering the next.

The schematic in Figure 5-108 shows what I have in mind. It looks more complex than it really is. Each timer has the same pattern of components around it, so after you create the first module, you just make three copies of it.

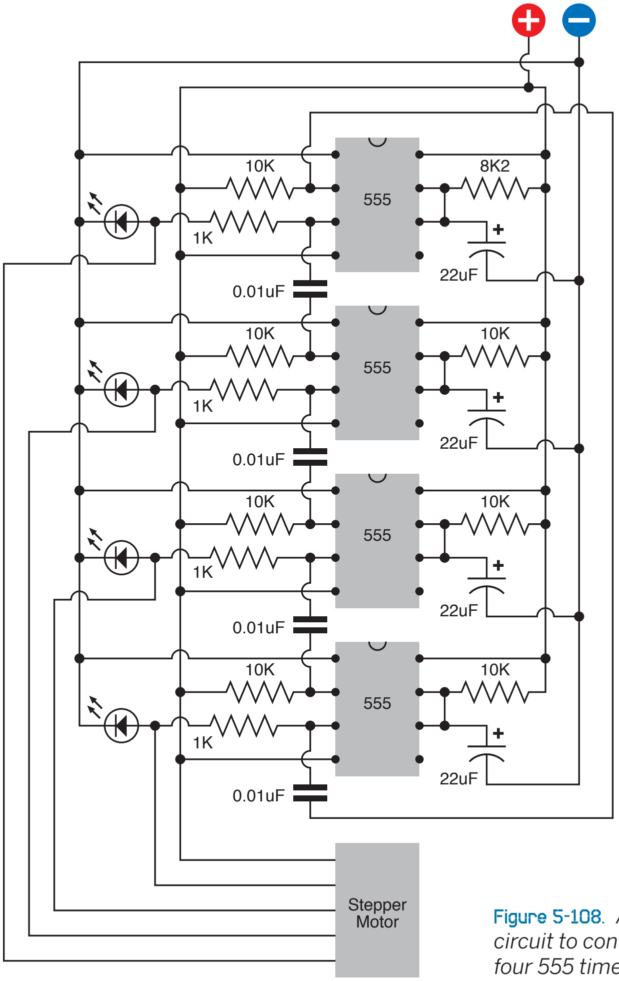

Figure 5-108. *A very quick and simple circuit to control a stepper motor uses four 555 timers, each in monostable mode, triggering each other in a repeating sequence.*

I've used a 10K resistor to pull up the input to each 555, so that the timers are naturally in their quiescent state. A 0.01 µF capacitor links the output from one timer to the input of the next so that they are electrically isolated from each other, and the capacitor just conveys a "spike" of voltage when one timer finishes its "on" cycle, and its output goes low, which triggers the next.

On the righthand side, I've used 10K resistors and 22 µF capacitors to generate a cycle of about a quarter of a second—except that the topmost timer has a 8K2 timing resistor. The reason for this is that when power is first applied, the timers will all be waiting for each other to begin, and timers 2 and 4 or 1 and 3 may fire together. By giving one timer a shorter cycle than the others, I minimize this problem.

The LEDs are included just to give you some visual verification of what's happening. Without them, if you make a wiring error, the motor may turn to and fro erratically, and you won't know why. Initially you can run your circuit with only the LEDs connected, just to make sure it works. Figure 5-109 shows the breadboarded circuit before the motor is plugged in. Then add the motor by plugging its wires into the breadboard, where you'll make connection with the outputs (pins 3) of the timers. See Figure 5-110.

Apply power, and you should see the motor turning in steps, in sequence with the LEDs. If the LED sequence isn't stable:

1. Connect a wire directly from the input (pin 2) of the topmost timer to the positive side of the voltage supply, and wait for the timers to calm down.
2. Restart the sequence by disconnecting the free end of this wire, or (if necessary) touch the free end of it briefly to the negative side of the supply, to trigger the first timer.

One thing you may have noticed, if you're paying very close attention: the common terminal of the motor is connected to positive. Therefore, when each timer flashes positive, that positive signal isn't actually powering the motor. The *low* outputs from the three timers that are *not* firing at any given moment are sinking current from the motor. It seems quite happy with this arrangement. You'll need some theory to understand why.

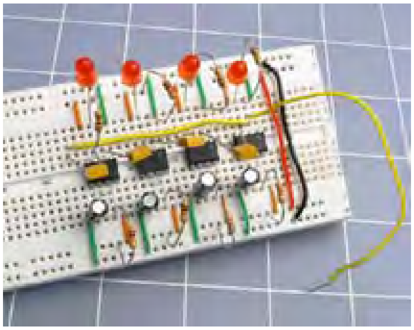

Figure 5-109. *To test the control circuit for errors, four LEDs show the outputs from the four 555 timers. The loose yellow wire at the righthand side connects to pin 2 of the first timer. Touch the free end of this wire to the positive side of the power supply to reset the timers, and then, if necessary, make a brief negative connection with the free end of the wire to restart their sequence.*

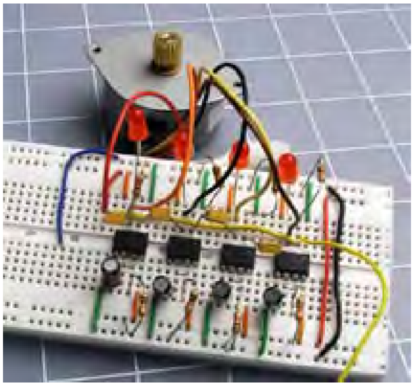

Figure 5-110. *After the circuit has been tested, the motor can be added by hooking its control wires to the outputs of the four 555 timers.*

THEORY

Inside a stepper motor

If you check the Wikipedia entry for stepper motors, you may see a very nice 3D rendering showing a toothed rotor and four coils arrayed around it. Maybe stepper motors used to be manufactured like this once upon a time, but not anymore.

Imagine two horizontal rows of coils. In the space between them is a series of little magnets, like a freight train, that can move left or right, as shown in Figures 5-111 and 5-112. Each coil has two windings, in opposite directions, so that current through one winding will create an upward magnetic force while current through the other will create a downward force. Each row of windings is connected in parallel, so that they switch on and off simultaneously.

In Step 1, the negative connection energizes the upper windings of the upper coils, which creates an upward magnetic force. I've shown this force using blue-green arrows so that you won't mistake it for a flow of electricity. It so happens that this force attracts the north poles of the magnets and repels the south poles, so if the magnets begin in the position shown in Step 1, they will want to move one step to the right.

This brings them to the position shown in Step 2. Now the upper windings of the lower coils are energized, and again, this produces an upward force, which again attracts the north poles and repels the south poles.

This advances the magnets to their location in Step 3. Now the lower windings of the upper coils are energized, producing a downward force. This repels the north poles of the magnets and attracts their south poles. So the magnets keep moving.

They reach the position shown in Step 4. The lower windings of the lower coils are energized, producing a downward force which continues to attract the south poles while repelling the north poles. So the magnets move a final step to the right—which leaves them in the same orientation shown in Step 1. And the process can repeat all over again.

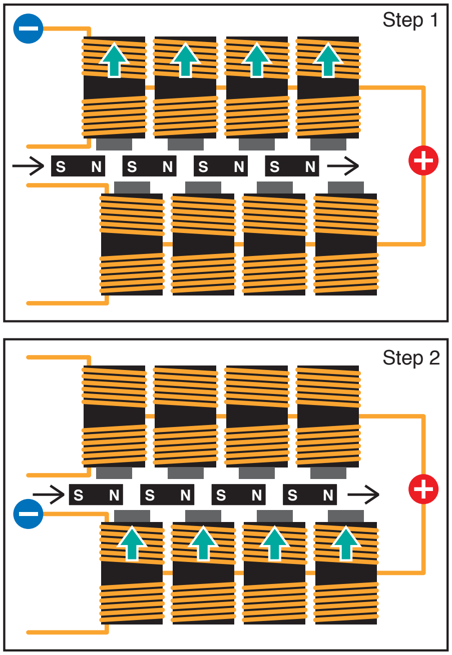

Figure 5-111. *This sequence shows the first two steps as the rotor of a stepper motor (shown as a series of north-south magnets) moves in response to pulses through electromagnets.*

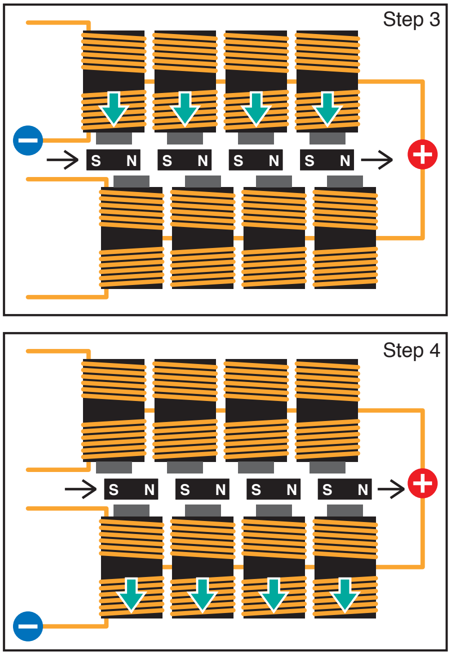

Figure 5-112. *After taking another two steps, the motor will be back where it started at Step 1 in Figure 5-111.*

THEORY

Inside a stepper motor (continued)

In reality, the magnets are not separate from each other. The edge of a rotor is magnetized in zones that alternate between south and north polarity. And instead of multiple coils, there are just four windings that go around all the magnetic cores. But the principle is exactly the same. The 3D rendering gives a general idea, and the photograph shows what I found when I cracked open a typical stepper motor.

Now bear in mind that when this device is driven by a set of 555 timers, we don't just connect negative to one wire at a time on the left, leaving the others floating. In reality, at any given moment, three of the timers have a negative output and the fourth has a positive output. The last diagram in Figure 5-112 shows this situation.

Suppose the top wire is positive while the other three are negative, as shown in Figure 5-113. The positive output does nothing, because it is balanced by the positive power on the other end of the coils. The two negatives attached to the bottom set of coils create equal and opposite forces that cancel each other out (while wasting some power). So the net result is the same as in Step 3.

In fact, you should find that you can disconnect the common wire completely while using the stepper motor with 555 timers, and the motor will still turn, because one of the timers is providing positive power while the others are negative. In fact, you'll be running them more efficiently this way.

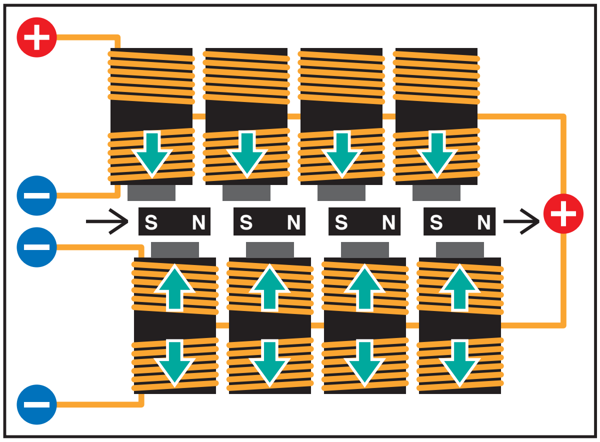

Figure 5-113. *When the motor is driven by four 555 timers, they are activating it by sinking positive voltage from it. The interior workings of the motor look something like this. It's not the most efficient way to do the job.*

Figures 5-114 and 5-115 may help to give you a clearer idea of what the motor actually looks like inside.

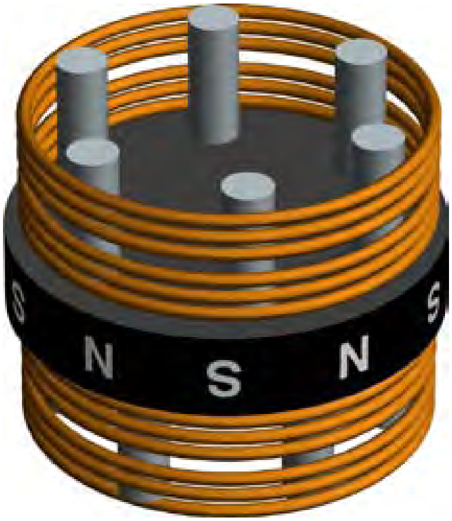

Figure 5-114. *This 3D rendering gives a better idea of what a typical stepper motor looks like inside. The copper coils and gray cylinders are stationary, while the black disc rotates between them.*

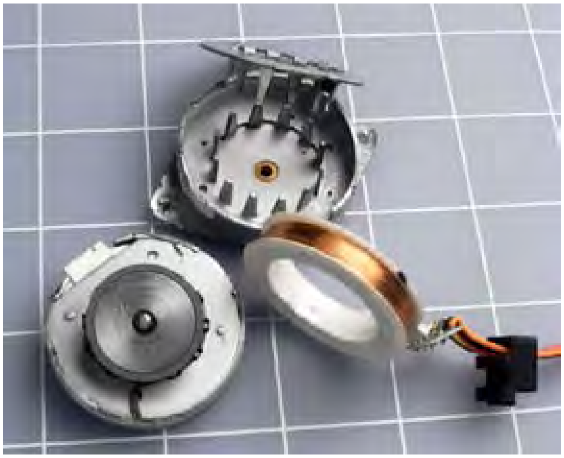

Figure 5-115. *When a stepper motor is broken open, this is what you're likely to find. On the left, the rotor of the motor, which has a magnetized band around its circumference, is still attached to the lower half of the casing. On the right, the upper half of the casing has been opened, and the coil has been removed (actually the winding you can see consists of two coils, wound in opposite directions). The spikes are the magnetic cores that exert force on the rotor.*

Speed Control

If you are a truly exceptionally observant, you may have noticed that I left pin 5 of each of the timers unconnected in the schematic for driving the stepper motor in Figure 5-108. Normally, pin 5 should be grounded through a capacitor to prevent it from picking up stray voltages which can affect the accuracy of the chip.

I left the pins unconnected because I had a plan for them. In fact, changing the timing of the chip is exactly what we want to do now, as a way to change the speed of the stepper motor.

If you tie pin 5 of all four timers together, as shown in Figure 5-116, and put a 2K trimmer potentiometer (shown in Figure 5-117) between them and the negative side of the power supply, you'll find that as you turn the trimmer to reduce its resistance, the timers start to run faster. Figure 5-118 shows the breadboard layout. Eventually, when the resistance goes below around 150 ohms, everything stops. The LEDs go dark, because you've reduced the voltage on pin 5 below the threshold level that the 555 timer finds acceptable.

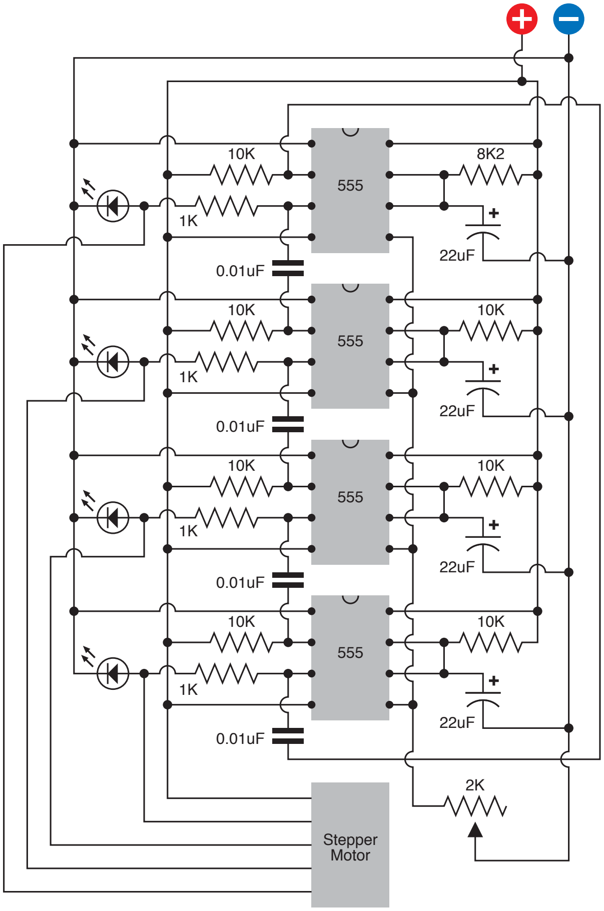

Figure 5-116. *To adjust the speed of the sequence of 555 timers, their control pins (pin 5 on each timer) are linked together and attached to a trimmer potentiometer that adjusts the resistance between the pins and the negative side of the power supply.*

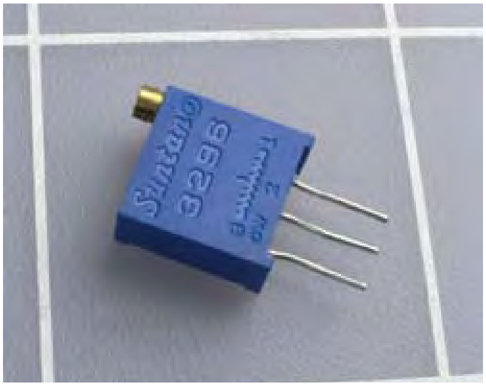

Figure 5-117. *Close-up of a trimmer potentiometer with pins spaced at 1/10 inch for insertion in a breadboard or perforated board. The brass screw, at top-left, turns a worm gear inside the unit, allowing precise adjustment of internal resistance.*

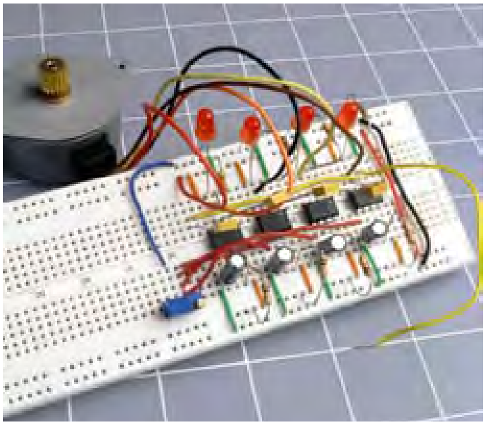

Figure 5-118. *The trimmer potentiometer has been added to the circuit, allowing motor speed control.*

Initially I suggested a step time of 1/4 second just so that you could see what was happening. When you're actually using this circuit, you'll never need it to run as slowly as that. So you can increase the entire range of speeds. Remove the 22 µF timing capacitors and substitute, say, 4.7 µF capacitors, or smaller. Now when you adjust the potentiometer, you'll get a useful range of speed.

Adding Autonomy

Currently, the circuit simply does what you tell it to do. The next step is to make it autonomous—in other words, give it the illusion of making up its own mind. I'm thinking that instead of a trimmer potentiometer, we could substitute a photocell, properly known as a photoresistor. Typically, the resistance of a cadmium sulfide photo resistor is highest in the dark, and lowest when light shines on it.

One problem with photoresistors is that they're not as widely available as many other types of electronic components. If you search Mouser.com, for instance, you'll find virtually nothing. Partly this is because the online search function at Mouser is the weakest feature of the site, and partly it's because Mouser is not oriented toward hobbyists. What you need to do is conduct a "product search." Go to *http://www.google.com/products*, enter the search terms "CdS" and "photocell," and you'll find a bunch of cheap cadmium sulfide components from places you may never have heard of.

Because photoresistors seem to come and go as erratically as DC motors, I am not offering any part numbers. You can buy any product that has an appropriate minimum resistance (in bright light) and maximum resistance (in the dark). If you find a component that ranges from 500 to 3,000Ω, that would be a good choice. If the only ones you can find have a higher minimum than 500Ω, you could consider putting a couple of them in parallel.

Setting Up Your Light Seeking Robot

Why would you want to control the speed of a stepper motor by using a photo resistor? Because the original objective was to build a robot that is attracted to light.

The idea is simple enough: use two stepper motors, each controlling the speed of one wheel of the cart. Use two photoresistors, each controlling the speed of the opposite stepper motor. When the righthand photoresistor picks up more light, its resistance lowers, causing the lefthand set of timers to run faster, which will make the lefthand wheel run faster. Thus, the cart will turn toward the light. Figure 5-119 illustrates the concept.

Before you start wiring more 555 timers, though, you might consider doing the job with a more appropriate component. The ULN2001A and ULN2003A are chips containing Darlington amplifiers specifically designed to deliver current to inductive loads such as solenoids, relays, and (you guessed it) motors. Each chip has seven inputs that require very little current, and seven outputs that can deliver 500mA each. The inputs are TTL and CMOS compatible (the 2001A has a wider tolerance for voltages than the 2003A) and each channel of

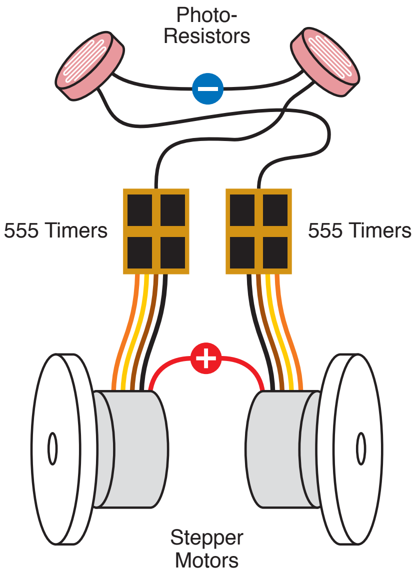

Figure 5-119. *If two photoresistors control the speed of two 555-timer arrays, the difference in speed between one wheel and the other can turn the cart toward a light source.*

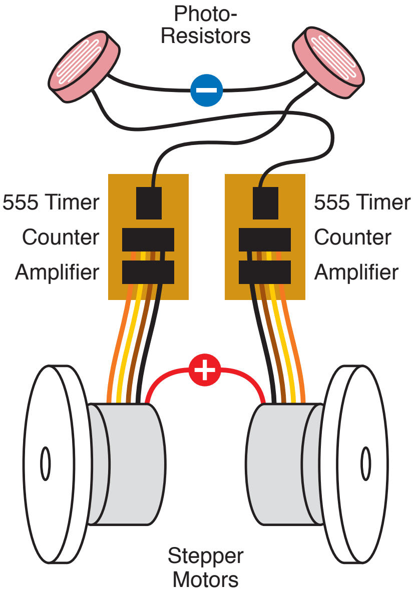

Figure 5-120. *A more efficient way to drive the motors is to use just one timer to set the speed of each, with a counter and amplifier (such as a Darlington array chip) sending the pulses down the wires. The principle is still the same, though.*

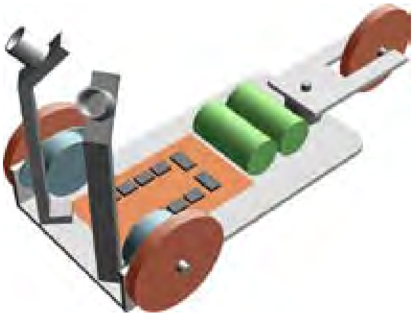

Figure 5-121. *This 3D rendering shows a possible configuration of the light-seeking cart, with two photoresistors enclosed in small tubes to restrict their response to light.*

the chip functions as an inverter, so that when the input goes high, the output goes low and sinks current. This is of course just what we need for our stepper motor that has a common positive connection.

The ULN2001A is only an amplification device, so you have to precede it with a counter that runs from 1 to 4 and then repeats. You can stick with your 555 timers, as you've already assembled them, or substitute almost any CMOS octal or decade counter that sends its output pulses to a series of pins. Just use the output from the fifth pin as the "carry" output to restart the counting sequence. I suggest a CMOS counter simply because it will run on 12 volts, so you can use the same power supply that suits your stepper motors.

If you switch to CMOS counters, you will still need a pair of 555 timers sending pulses to the counters. The timers will be free-running in astable mode, and your photoresistors will control their speed. Figure 5-120 shows the configuration.

One last item: you'll need a 12-volt battery. You can of course put eight AA cells together, but I think you should consider a rechargeable pack from a source such as *http://www.all-battery.com*, which has a section entirely devoted to "robot batteries."

If you put it all together, you should find that when you place your robot cart in a very dimly lit room, it will turn toward the beam from a bright, well-focused flashlight. To get reliable results, you may have to recess each of the photoresistors in little tubes, so that they receive much more light when they face your flashlight than when they face away from it. Figure 5-121 is a 3D rendering of the concept.

Another idea is to rewire your cart so that it actually runs away from the light. Can you imagine how this might be done?

Just one more thought: if you use infrared photoresistors, you can control your cart with beams from infrared LEDs, in normal room lighting. If you and a couple of friends all have infrared transmitters, you can get your cart to run from one of you to the next, like an obedient dog.

This takes us about as far as I'm going to go into robotics. I urge you to check out the sites online if you want to pursue the topic further. You can also buy a wide variety of robot kits, although of course I feel that it's more fun to invent or develop things for yourself.

All that's left now is to perform one last introduction: to a device that should make your life much easier, even though the device is much more complicated than anything we have dealt with so far.

Experiment 34: Hardware Meets Software

Throughout this book, in accordance with the goal of *learning by discovery*, I have asked you to do an experiment first, after which I've suggested the general principles and ideas that we can learn from it. I now have to change that policy, because the next experiment involves so much setup that it's only fair to tell you what to expect before you begin the preparations.

We are about to enter the realm of controller chips, often known as MCUs, which is an acronym for micro controller unit. An MCU contains some flash memory, which stores a program that you write yourself. The flash memory is like the memory in a portable media player, or the memory card that you use in a digital camera. It needs no electricity to power it. In addition, the chip has a processor which carries out the instructions in your program. It has RAM to store the values of variables on a temporary basis, and ROM, which tells it how to perform tasks such as sensing a varying voltage input and converting it into digital form for internal use. It also contains an accurate oscillator, so that it can keep track of time. Put it all together, and it's a tiny computer that you can buy for under $5.

Let's suppose that you have a greenhouse where the temperature must never fall below freezing. You set up a temperature sensor, and you have two different heaters. You want to switch on the first heater if the temperature falls below 38° Fahrenheit. But if, for some reason, that heater is broken, you want to switch on the second, backup heater when the temperature goes below 36° Fahrenheit.

Programming a MCU to take care of this can be very simple indeed. You could even add extra features, such as a second temperature sensor, just in case the first one fails, and you could tell the chip to use whichever sensor gives a lower reading.

Another application for a MCU would be in a fairly elaborate security system. The chip could monitor the status of various intrusion sensors, and can take various preprogrammed steps, depending on the sensors' status. You could include delay intervals, too.

Many MCUs have additional useful features built in, such as the ability to control servo motors that turn to a specific angle in response to a stream of pulses. Servos are widely used in radio-controlled model boats, airplanes, and hobby robotics.

Perhaps you are now wondering why, if MCUs can do all this, haven't we been using them all along? Why did I spend so much time describing the development of an alarm system using discrete components, if one chip could have done everything?

There are three answers:

1. MCUs cannot do everything. They need other components to help them interact with the world, such as transistors, relays, sensors, and amplifiers. You need to know how those things work, so that you can make intelligent use of them.
2. MCUs can introduce their own kinds of problems and errors, associated with using software in addition to hardware. I'll have more to say about this later.
3. MCUs have limits and restrictions, most obviously their requirement for a 5-volt regulated power supply, and their inability to source or sink much current from each pin. They also demand that you learn a programming language (which differs from one brand of MCU to the next). And to get the program into the chip, you have to be able to plug it into a computer and do a download, which is not always convenient.

In this experiment, you'll learn how to write a program for a small and simple MCU, and you'll transfer the program into it and see how it works.

BACKGROUND

Origins of programmable chips

In factories and laboratories, many procedures are repetitive. A flow sensor may have to control a heating element. A motion sensor may have to adjust the speed of a motor. Microcontrollers are perfect for this kind of routine task.

A company named General Instrument introduced an early line of MCUs in 1976, and called them PICs, meaning Programmable Intelligent Computer—or Programmable Interface Controller, depending which source you believe. General Instrument sold the brand to another company named Microchip Technology, which owns it today.

"PIC" is trademarked, but is sometimes used as if it's a generic term, like Scotch tape. In this book, I've chosen a range of controllers based on the PIC architecture. They are licensed by a British company named Revolution Education Ltd., which calls its range of chips the PICAXE, for no apparent logical reason other than that it sounds cool.

I like these microcontrollers because they were developed originally as an educational tool and because they are very easy to use. They're cheap, and some of them are quite powerful. Despite their odd name, I think they're the best way to get acquainted with the core concepts of MCUs.

After you play with the PICAXE, if you want to go farther into MCUs, I suggest the BASIC Stamp (which uses a very similar language, but with additional powerful commands) and the very popular Arduino (which is a more recent design, packed with powerful features, but requires you to learn a variant of C language to program it). I'll say more about these chips later.

If you search for "picaxe" on Wikipedia, you'll find an excellent introduction to all the various features. In fact, I think it's a clearer overview than you'll get from the PICAXE website.

Supplies

Figure 5-122 shows some of the chips in the PICAXE range. I'll be telling you how to use the smallest—the 08M—which costs less than $5 and is cheaper than any other MCU that I've found. It has only 256 bytes of memory to store a program (not gigabytes, megabytes, or kilobytes, just 256 bytes!), but you'll be surprised how many possibilities this amount allows for. Figure 5-123 shows a closeup of the 08M with its legs safely embedded in a piece of conductive foam.

In the United States, there are three distributors of this chip:

- *http://www.advancedmicrocircuits.com*
- *http://www.phanderson.com/picaxe*
- *http://www.sparkfun.com*

I like P. H. Anderson for its grass-roots hobbyist approach, and they have good prices if you want to buy multiple chips. But SparkFun Electronics offers other associated products that you may find interesting.

All the distributors will want to sell you "starter kits," such as the one in Figure 5-124, perhaps because the PICAXE itself is so cheap that it doesn't offer much of a profit margin. Still, for our purposes, you should buy the chip as a standalone item. And buy two of them, just in case you damage one (for example, by connecting voltage to it incorrectly).

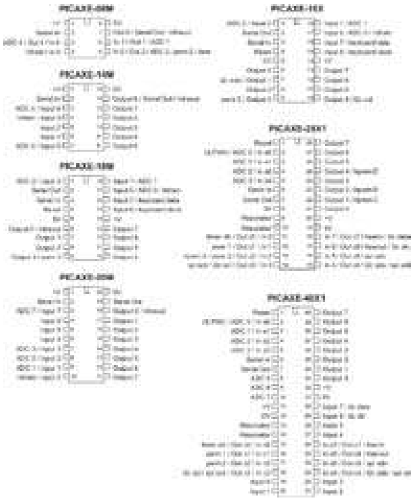

Figure 5-122. *A page from the PICAXE catalog lists only some of the chips that are available. What began as an educational aid has become a useful prototyping tool.*

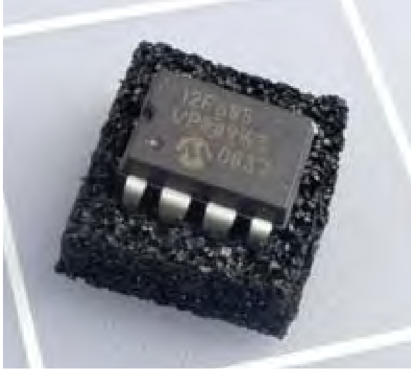

Figure 5-123. *When supplied by one of its American distributors, a PICAXE 08M arrives embedded in a little square of conductive foam. The chip is the same size as a 555 timer but has the power of a tiny computer.*

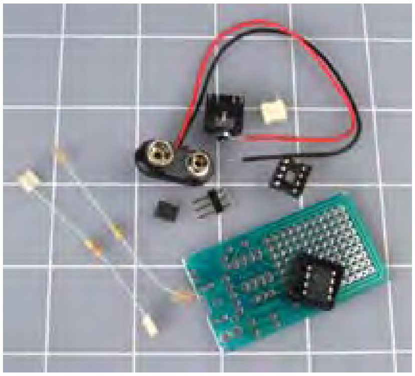

Figure 5-124. *A typical PICAXE kit includes a printed circuit board, which you may not really need, and some other not-entirely-essential items. But the 3.5-mm stereo jack socket (top, center) is absolutely necessary.*

To download your programming instructions into the chip, first you'll type the instructions on a computer, and then you'll feed them through a cable into the PICAXE memory. So you'll need to buy a cable, and you'll need software to help you to write the program.

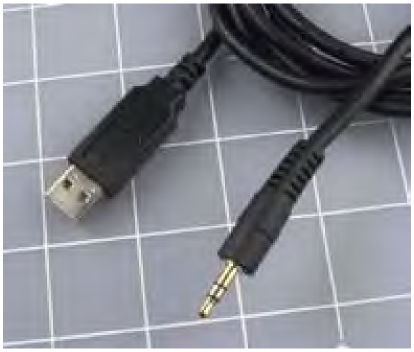

Figure 5-125. *The USB download cable made for use with the PICAXE terminates in a 3.5-mm audio plug. This should not be inserted in any audio device. It establishes a serial connection with a computer, allowing program code to be downloaded into the chip.*

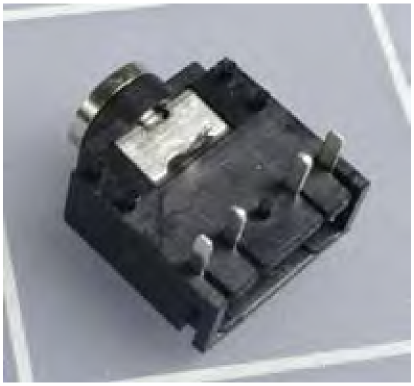

Figure 5-126. *Closeup of the 3.5-mm stereo socket that is used with the USB download cable.*

You can use the PICAXE with a serial cable, but I don't recommend it. The old RS-232 serial communications standard on PCs is pretty much obsolete, and PICAXE has recognized this by offering a USB cable (which contains a serial converter inside its plug). The USB cable is a little more expensive, but is also simpler and compatible with Apple computers. From any of the U.S. distributors, buy USB cable part AXE027, also sold as part PGM-08312 by *http://www.sparkfun.com* (quantity: 1). The cable is shown in Figure 5-125.

To write your software and send it down the wire to the chip, the PICAXE Programming Editor is the tool of choice. It comes in only a Windows version. For those who prefer Mac OS or Linux, you can get a free download of another piece of software known as AXEpad, which has fewer features, but will do the job. All the downloadable software is freely available from *http://www.rev-ed.co.uk/picaxe/software.htm*.

Finally, you need a 3.5-mm stereo audio socket with solder connections. The reason for this is that the manufacturers of the PICAXE have used a stereo audio plug on the free end of their USB cable, and you have to be able to plug it into something. The PICAXE breadboard adapter, SparkFun stock number DEV-08331, contains the necessary stereo socket in addition to a few other little items. Quantity: 1. See Figure 5-126.

Oddly enough, the USB cable is the most expensive item on the list, because of the electronics hidden inside it.

Software Installation and Setup

Now you have to go through a setup procedure. There is no way around this. Here is what you will be doing:

1. Install a driver so that your computer will recognize the special USB cable.
2. Install the Programming Editor software (or AXEpad for Mac/Linux) so that your computer will help you to write programs and then download them into the chip.
3. Mount the PICAXE on your breadboard and add the socket to receive downloads.

These steps are explained in the following sections.

The USB driver

Fair warning: If you go to the PICAXE website and try to use its search function, it probably won't find what you want. Search for "USB Driver," for instance, and it will pretend it has never heard of such a thing.

The PICAXE home page also has irritating drop-down menus that tend to disappear just when you're about to click on them, but at the time of writing, you can bypass these issues by going straight to the Software Downloads section at *http://www.rev-ed.co.uk/picaxe/software.htm*.

Scroll down past all the software until you get to Additional Resources. Look for the AXE027 PICAXE USB Download Cable. At first glance, it looks as if they want to sell you a cable, but in fact this is the list of drivers. Double-click the one appropriate to your computer, and choose a destination on your computer for the download—a place where you will find it easily, such as your desktop.

Be careful not to download the driver for the USB010 USB-Serial adapter by mistake. The USB-Serial adapter is something else entirely.

The download will leave you with a zipped file folder. You will have to unzip it. On Windows XP, right-click the folder and choose "Extract all." View the extracted files and you will find a PDF installation guide. Linux and Mac users can find instructions currently stashed at *http://www.rev-ed.co.uk/docs/AXE027.pdf*.

When installing the driver on a Windows platform, here are a few tips to minimize your exasperation level:

1. Remember, the special USB cable contains some electronics. It is not just a cable, but a device designed for interacting with a PICAXE chip. *Don't try to use it for anything else!*
2. You have to plug the cable into a USB port *before* you install the driver, because your computer will need to verify that the driver matches the cable.
3. You must not attach the PICAXE to the other end of the cable until *after* you have installed the driver.
4. Every USB port on your computer has a separate identity. Whichever one you choose when you first plug in the cable, you should use that port *every time in the future*. Otherwise, you will have to repeat the process of telling your computer what the cable is.
5. Bearing in mind Tip #4, you should avoid using the cable in a standalone USB hub.
6. The cable is fooling the PICAXE into thinking that it's talking to a serial port on your computer. Those "communication" ports are known as COM1, COM2, COM3, or COM4. When you install the driver, the installer will choose one of those COM ports for you, and later you will have to know which one it is. The PDF guide should help you through this procedure. Unfortunately, you cannot skip it.

The Programming Editor software

If you have come this far, you're ready for the next big step, which is much easier. You need the PICAXE Programming Editor, available for free on the Software Downloads web page where you found the USB driver. (If you are using a Mac or Linux, you will need AXEpad, which is on the same web page.)

Downloading and installing the Programming Editor should be simple and painless. Once you have done that, you should find that it has placed a shortcut on your desktop. Double-click it, go to View→Options, and in the window that opens (shown in Figure 5-127), click the Serial Port tab. You should see a dialog box like the one in Figure 5-128. Now make sure that the Programming Editor is looking at the same COM port that was chosen by the USB driver. Otherwise, the Programming Editor won't know where to find your PICAXE chip.

In the Programming Editor, go to View→Options and click the Mode tab, then click the button to select the 08M chip.

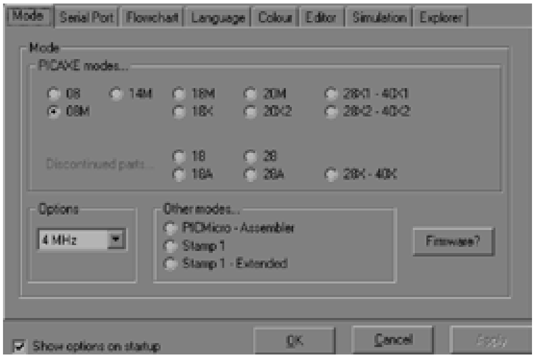

Mode

PICAXE modes...

○ 08 ○ 14M ○ 18M ○ 20M ○ 28X1 - 40X1

● 08M ○ 18X ○ 20X2 ○ 28X2 - 40X2

Discontinued parts... ○ 18 ○ 28

○ 18A ○ 28A ○ 28X - 40X

Options: 4 MHz

Other modes...

○ PICMicro - Assembler

○ Stamp 1

○ Stamp 1 - Extended

☑ Show options on startup

Figure 5-127. *This screenshot shows the options window of the PICAXE Program Editor, which you must use to select the type of chip that you intend to program (in our case, the 08M).*

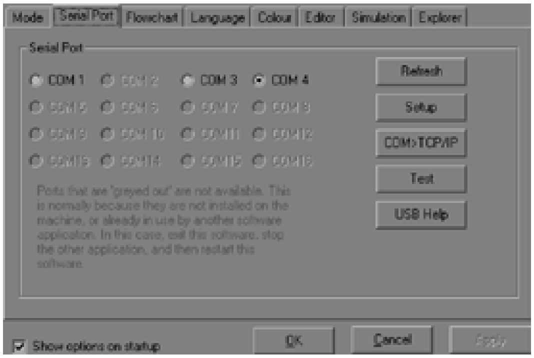

Serial Port

○ COM 1 ○ COM 2 ○ COM 3 ● COM 4

○ COM 5 ○ COM 6 ○ COM 7 ○ COM 8

○ COM 9 ○ COM 10 ○ COM11 ○ COM12

○ COM13 ○ COM14 ○ COM15 ○ COM16

Ports that are "greyed out" are not available. This is normally because they are not installed on the machine, or already in use by another software application. In this case, exit this software, stop the other application, and then restart this software.

☑ Show options on startup

Figure 5-128. *Another screenshot of the options window shows the second essential choice that you must make: selection of the COM port that the installer chose on your computer.*

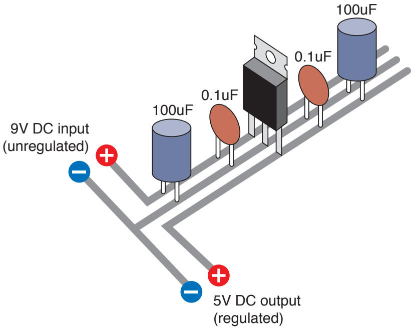

Figure 5-129. *PICAXE documentation specifies a 100 µF and 0.1 µF capacitor on the input side of a 5-volt regulator, and a similar pair of capacitors on its output side. On a breadboard, they can be arrayed like this.*

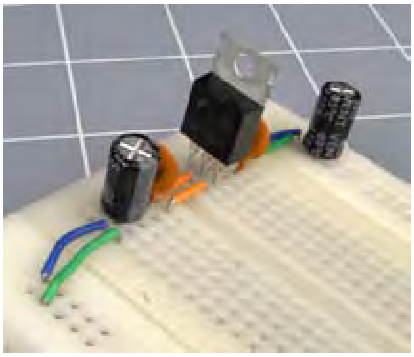

Figure 5-130. *The actual components for power regulation, applied to a breadboard, delivering 5 volts (positive and negative) down each side.*

Are we having fun yet? Obviously not, but you're through with software hassles for the time being. The last step before you're ready to use the PICAXE is to mount it, and its socket, on your breadboard.

Setting up the hardware

The PICAXE 08M looks like a 555 timer. (Other chips in the PICAXE range have more pins and more features.) It requires a properly regulated 5 volts, just like the logic chips you dealt with previously. In fact, the PICAXE people are rather emphatic about protecting it from voltage spikes. They want you to use two capacitors (one 100 µF, one 0.1 µF) on either side of an LM7805 regulator. This seems like overkill, but the PICAXE is more inconvenient to replace than a 555 timer. You certainly can't run down to RadioShack to buy one. So let's do what the manufacturer says, just in case, and set up a breadboard as shown in Figures 5-129 and 5-130.

Now for the chip itself. Note that the pins for positive and negative power are *exactly opposite* to those for the 555 timer, so be careful!

Set up your breadboard following the schematic shown in Figure 5-131. Note that I am showing the stereo socket on its underside, because I think that's how you'll have to use it with the breadboard. If you try to stick its pins into the holes in the board, they will fit, but when you insert the plug into the socket, the thickness of the plug will tend to raise the socket up so that it loses contact. I really think the way to go is to solder wires to the pins on the socket and push the wires into the breadboard. See Figure 5-133.

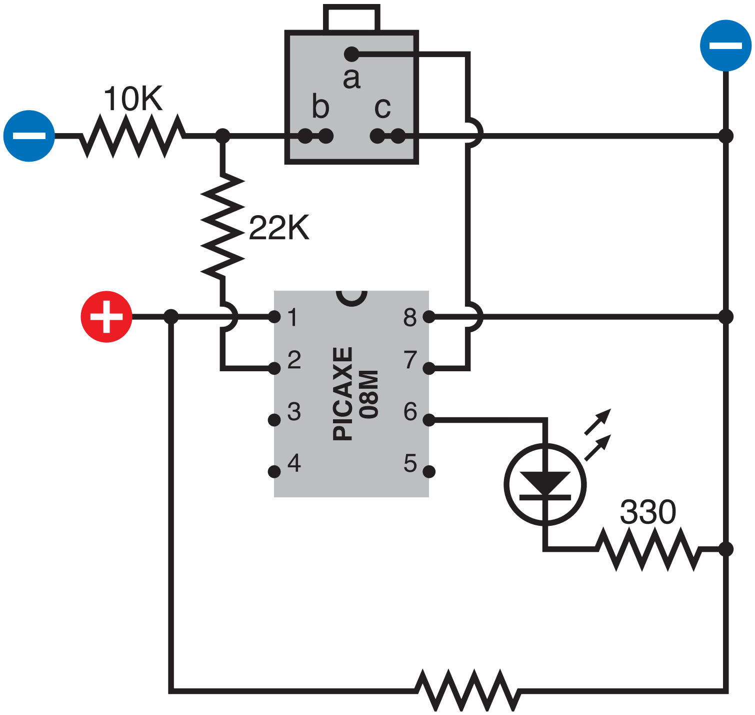

Figure 5-131. *The schematic of a test circuit for the PICAXE 08M shows the underside of the stereo socket, the essential 10K and 22K resistors on the input pin, and an LED to show an output from the chip.*

Be aware that the PICAXE manual shows things differently (although I have retained their labeling convention for the parts of the socket and the parts of the plug, identified as a, b, and c).

One little detail about the socket that is commonly supplied for use with the PICAXE: typically it has two *pairs* of contacts for the connections labeled b and c in the manual, and in my diagram. When you solder a connection, your solder joint should include *both* of the contacts in each pair, as shown in Figure 5-132.

Remember that the PICAXE must have 5 volts DC, and remember that your voltage regular will deliver this voltage reliably only if you give it extra voltage on its input side. If you provide it with 9 volts, that will provide a good amount of headroom.

The 22K and 10K resistors are essential for using the chip; see the following warning note for an explanation. My schematic also includes an LED and a 330Ω resistor, but they are needed only for the test that we'll be making momentarily.

Pin 2 Pull-Down

Always include the 22K resistor and the 10K resistor in the configuration shown in Figure 5-131. These resistors apply correct voltage to the serial connection, and when you're using the PICAXE on its own, they pull down the voltage on pin 2.

If pin 2 is left unconnected (floating), it may pick up random voltages, which the chip can misinterpret as a new program or other instructions, with unpredictable and undesirable results.

The 22K and 10K resistors should be regarded as permanent items accompanying your PICAXE regardless of whether you have it attached to your computer.

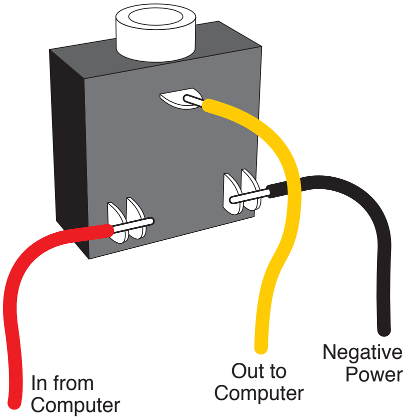

Figure 5-132. *Correct wiring of the socket is essential. When soldering wires to the lower terminals, make sure that you attach the wires to both of the terminals in each pair.*

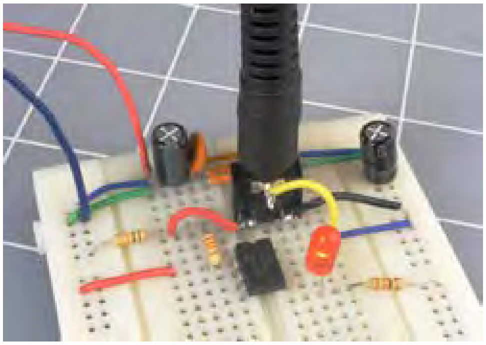

Figure 5-133. *The breadboard version of the test schematic, with the plug of the USB download cable inserted in the socket on the board. The PICAXE chip can now receive a downloaded program, and will immediately start to execute it.*

Verifying the Connection

Follow these steps carefully every time you want to program or reprogram your PICAXE chip:

1. Insert the USB plug of your PICAXE cable into the same USB port that you used before.
2. Start the Programming Editor (or AXEpad if you are using a Mac OS or Linux).
3. In the Programming Editor, select View→Options to verify that the editor is using the right COM port and is expecting the 08M PICAXE chip.
4. Plug the stereo plug on the free end of the USB cable into the stereo socket that is now wired into your breadboard. See Figures 5-133 and 5-134.
5. Check your wiring, and then connect your power supply to the breadboard.
6. Click the button labeled "program" in the Program Editor window to tell the software to look for the PICAXE.

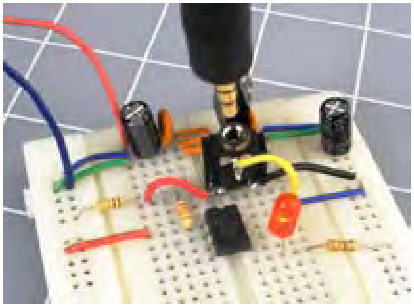

Figure 5-134. *After the program has been downloaded, the plug can be removed, and the program will continue to run, causing the LED to continue flashing.*

What If It Doesn't Work?

The first thing to do is pull out the plug of the USB cable from the PICAXE breadboard, leaving the other end of the cable attached to your computer. Set your multimeter to measure DC volts, and attach its probes to sections b and c of the plug. See Figure 5-135. Now click the "program" button again, and your meter should show 5 volts briefly coming out of your computer to the plug on the end of the cable.

If you detect the voltage, the software is installed and working properly. In that case, there's a problem on your breadboard, either in the chip or in the wiring around it.

If you cannot detect any voltage, the software probably wasn't installed properly, or is looking for the wrong serial port. Try uninstalling it and reinstalling it.

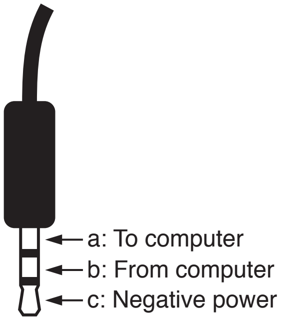
a: To computer
b: From computer
c: Negative power

Figure 5-135. *The stereo plug on the end of the USB download cable can be used for fault tracing. A multimeter set to measure DC volts can be attached to sections b and c of the plug to establish whether the Programming Editor is sending data through the serial connection.*

Your First Program

Finally you're ready to create your first program. Type the following code into the Programming Editor window:

```
main:
    high 1
    pause 1000
    low 1
    pause 1000
    goto main
```

Be sure to include the colon after the word "main" on the first line. See Figure 5-136 for a screenshot. The indents are created by pressing the Tab key. Their only purpose is to make program listings more legible. The software ignores them.

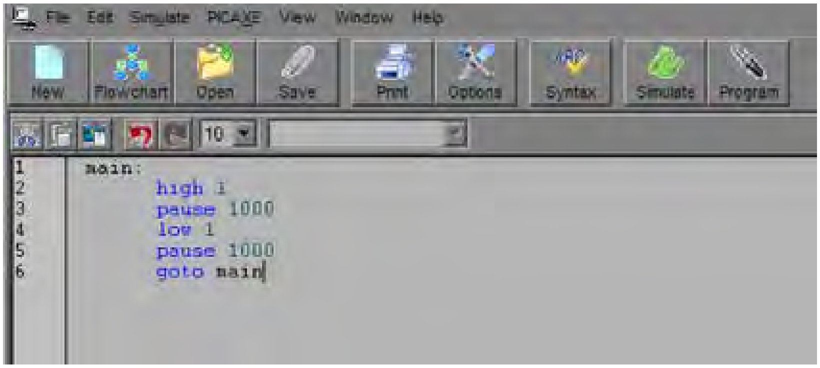
File
Edit
Simulate
PICAXE
View
Window
Help
New
Flowchart
Open
Save
Print
Options
Syntax
Simulate
Program
10

```
main:
        high 1
        pause 1000
        low 1
        pause 1000
        goto main
```

Figure 5-136. *This screenshot shows the first test program as it should be displayed by the Programming Editor (on a Windows computer).*

Click the Program button in the Programming Editor to download this program into the chip. As soon as the download is complete, the chip should start flashing the LED, lighting it for 1 second and then switching it off for 1 second. Figure 5-137 shows the steps that you should have followed to program the chip.

Now for the interesting part: disconnect the USB cable from the breadboard. The chip should continue flashing.

Disconnect the power supply from the breadboard and wait a minute or two for the capacitors to lose their charge. Reconnect the power, and the chip will start flashing again.

The program that you downloaded to the chip will remain in the memory inside the chip and will begin running every time power is applied to the chip.

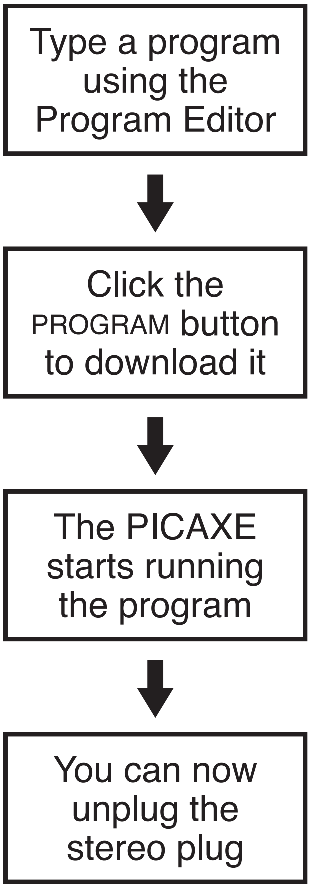
Type a program using the Program Editor
Click the PROGRAM button to download it
The PICAXE starts running the program
You can now unplug the stereo plug

Figure 5-137. *Four steps to create and run a program on the PICAXE controller chip.*

Decoding the Code

The program doesn't use the usual pin numbers on the chip. It uses what I will call "logic pin numbers." Figure 5-138 shows how they are numbered. Figure 5-139 shows their multiple functions. I have put Logic Pin 0 in parentheses because its main purpose is to send data to the computer through the USB cable. It can do double duty as a digital output, but you have to disconnect it from the USB cable first. It's easy to forget to do this. It's a hassle that I prefer to avoid.

Let's take a look at the little program that you typed in. The first line identifies a section of the program. This program only has one section, and we're calling it "main." Any word with a colon after it is the name of a section of a program:

```
main:
```

The second line tells the chip to send a high output from Logic Pin 1:

```
high 1
```

The third line of the program tells the chip to wait for 1,000 milliseconds. This of course is the same as one second:

```
pause 1000
```

The fourth line tells the chip to change Logic Pin 1 back to its low state:

```
low 1
```

The fifth line tells the chip to wait for another 1,000 milliseconds:

```
pause 1000
```

The last line tells the chip to go back to the beginning of the "main" section:

```
goto main
```

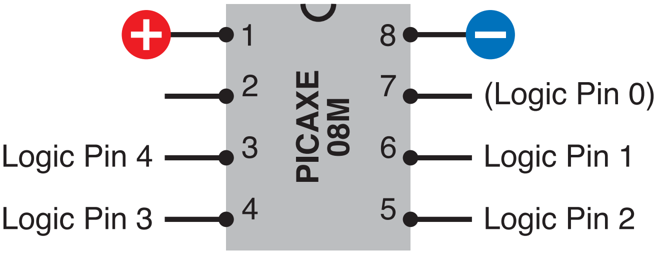
1
2
Logic Pin 4
3
Logic Pin 3
4
PICAXE
08M
8
7
(Logic Pin 0)
6
Logic Pin 1
5
Logic Pin 2

Figure 5-138. *The conventional pin numbers of the PICAXE chip are incompatible with the numbering system that is used in the PICAXE programming language. To minimize confusion, this guide refers to "Logic Pins" when using the numbering system that is required for programming the chip.*

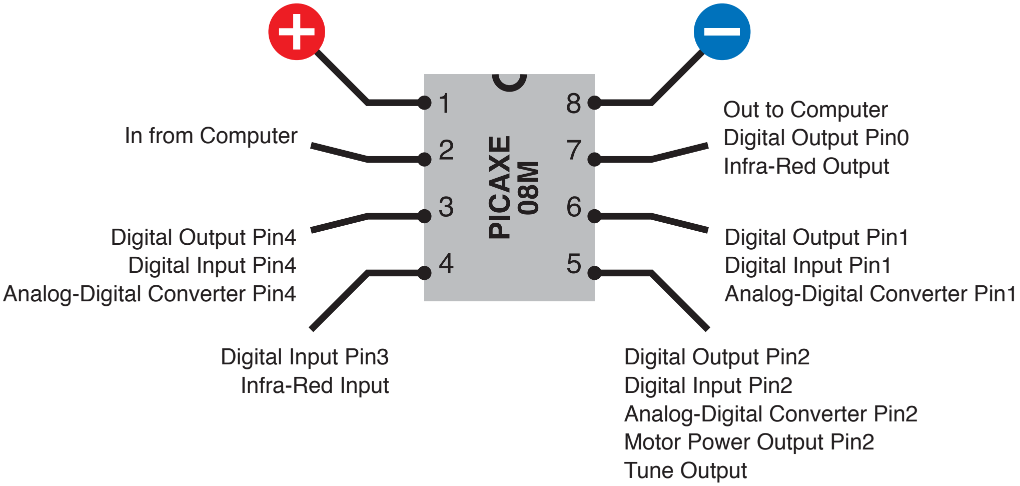
1
In from Computer
2
Digital Output Pin4
Digital Input Pin4
Analog-Digital Converter Pin4
3
4
Digital Input Pin3
Infra-Red Input
PICAXE
08M
8
Out to Computer
Digital Output Pin0
Infra-Red Output
7
6
Digital Output Pin1
Digital Input Pin1
Analog-Digital Converter Pin1
5
Digital Output Pin2
Digital Input Pin2
Analog-Digital Converter Pin2
Motor Power Output Pin2
Tune Output

Figure 5-139. *Many of the pins on the PICAXE 08M have multiple functions, which can be selected by appropriate program instructions.*

Editing

What if you want to change the program? No problem! Use the Programming Editor to change one of the lines in the program. Substitute 100 instead of 1000 milliseconds, for instance. (The `pause` command can be followed by any number up to 65535.) In your program, don't use a thousands comma in any of the numbers that you specify.

Plug the USB cable into the breadboard again, hit the Program button on the screen, and the new version of the program will be automatically downloaded to the chip, overwriting the old version.

What if you want to save the program for future use? Just go to the File menu in the Programming Editor and save the program onto your computer's hard drive. Because the PICAXE uses a variant of the BASIC computer language, it adds a *.bas* filename extension.

Simulation

If you make a simple typing error, the Programming Editor will find it and stop you from downloading your program. It will leave you to figure out how to fix the line that contains the error.

Even if all the statements in your program are correctly typed, it's a good idea to run a simulation of what they'll do, before you download them. This is easily done: click the "simulate" button on the menu bar of the Programming Editor. A new window will open, displaying a diagrammatic view of the PICAXE chip and showing you the states of its pins. (Note that if you use very short `pause` commands, the simulation won't run fast enough to display the time accurately.) A simulation screenshot is shown in Figure 5-140.

The >> button at the bottom-right corner of the simulation window will open up a list of all the variables in your program. So far, it doesn't have any variables, but it soon will. All the zeros on the righthand side are binary numbers, which you can ignore for now.

You'll need to check the second part of the PICAXE documentation, which contains all the programming statements and their correct syntax. At the time of writing, this is stored at http://www.rev-ed.co.uk/docs/picaxe_manual2.pdf.

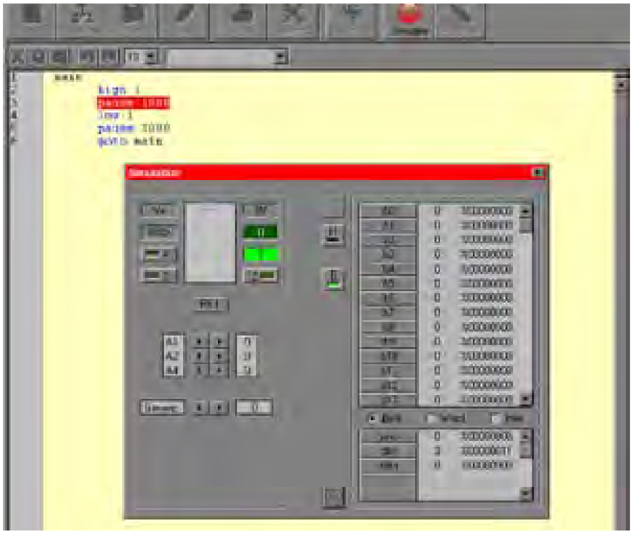

Figure 5-140. *This screenshot shows the simulation window that can be opened in the Program Editor to test program code before it is downloaded to the chip. The values of variables are shown in the section on the right. The pin states are shown on the left.*

Loops

Here's another thing I'd like you to try. Rewrite your program code as shown here and download it onto the PICAXE:

```
main:
    for b0 = 1 to 5
        high 1
        pause 200
        low 1
        pause 200
    next
    wait 2
    goto main
```

Note that `b0` is letter b followed by a zero, not letter b followed by letter O. The extra indents once again are added to make the listing easier to understand. The four lines beginning "high 1" and ending "pause 200" will be executed repeatedly. It's helpful to see them as a block.

Watch the light and see what happens. It should flash five times quickly, then wait for two seconds, and then repeat. You just added a *loop* to your program. You can use a loop if you want something to happen more than once.

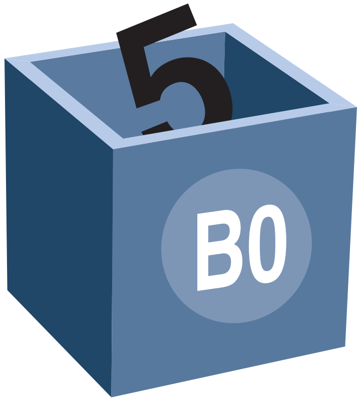

Figure 5-141. *To understand how a program works, visualize a variable as being like a "memory box" with its name on the outside and a number stored on the inside.*

b0 is known as a *variable*. Think of it as being like a little "memory box" with its name, b0, on a label on the outside. Figure 5-141 illustrates this concept. This particular memory box can contain any number from 0 through 255. The loop begins by telling the computer to put number 1 in the box, then process the remaining statements, until the word "next" sends the processor back to the first line, at which point it adds 1 to the contents of b0. If the value of b0 is 5 or less, the loop repeats. If the value is 6, the loop has run five times, so it's over, and the PICAXE skips down to the "wait 2" statement after "next." See Figure 5-142 for an annotated version of the program listing.

"Wait" is a PICAXE command that is measured in whole seconds, so "wait 2" waits for 2 seconds. Then "goto main" begins the procedure all over again.

If your flashing-light demo worked out as planned, it's time to take the next step and make the chip do something more useful.

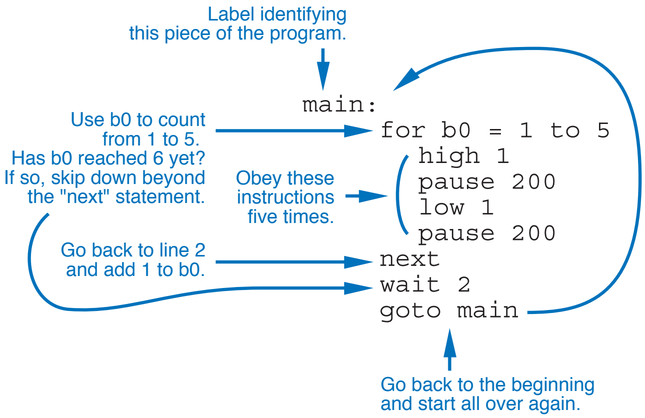
Label identifying
this piece of the program.
Use b0 to count
from 1 to 5.
Has b0 reached 6 yet?
If so, skip down beyond
the "next" statement.
Obey these
instructions
five times.
Go back to line 2
and add 1 to b0.
Go back to the beginning
and start all over again.

```
main:
for b0 = 1 to 5
   high 1
   pause 200
   low 1
   pause 200
next
wait 2
goto main
```

Figure 5-142. *The blue annotations explain what the program, on the right, is telling the PICAXE to do.*

FUNDAMENTALS

Basic PICAXE parameters

Here are some of the most useful parameters of the PICAXE:

- The PICAXE requires 5 volts DC, regulated.
- The inputs and outputs of the PICAXE are compatible with 5-volt logic chips. You can attach them directly.
- Each PICAXE pin can sink or source up to 20mA. The whole chip can deliver up to 90mA. This means that you can run LEDs directly from the pins, or a piezo noisemaker (which draws very little current), or a transistor.
- You can use a chip such as the ULN2001A Darlington array (mentioned in the previous experiment) to amplify the output from the PICAXE and drive something such as a relay or a motor.
- The chip executes each line of your program in about 0.1 milliseconds.
- The 08M chip has enough flash memory for about 80 lines of program code. Other PICAXE chips have more memory.
- The PICAXE provides 14 variables named b0 through b13. The "b" stands for "byte," as each variable occupies a single byte. Each can hold a value ranging from 0 through 255.
- No negative or fractional values are allowed in variables.
- You also have 7 double-byte variables, named w0 through w6. The "w" stands for "word." Each can hold a value ranging from 0 through 65535.
- The "b" variables share the same memory space as the "w" variables. Thus:
  - b0 and b1 use the same bytes as w0.
  - b2 and b3 use the same bytes as w1.
  - b3 and b4 use the same bytes as w2.
  - b5 and b6 use the same bytes as w3.
  - b7 and b8 use the same bytes as w4.
  - b9 and b10 use the same bytes as w5.
  - b11 and b12 use the same bytes as w6.
  - b13 and b14 use the same bytes as w7.

  Therefore, if you use w0 as a variable, do not use b0 or b1. If you use b6 as a variable, do not use w3, and so on.
- Variable values are stored in RAM, and disappear when the power is switched off.
- The program is stored in nonvolatile memory, and remains intact when the power is off.
- The manufacturer's specification claims that the nonvolatile memory is rewritable up to about 100,000 times.
- If you want to attach a switch or pushbutton to a pin and use it as an input, you should add a 10K pull-down resistor between the pin and the negative side of the power supply to hold the pin in a low state when the switch is open. Figure 5-143 shows how pull-down resistors should be used in conjunction with a SPST switch or a pushbutton.
- On the 08M chip, if you apply a varying resistance between Logic Pins 1, 2, or 4, and the negative side of the power supply, the chip can measure it and "decide" what to do. This is the "Analog-Digital Conversion" feature—which leads to our next experiment.

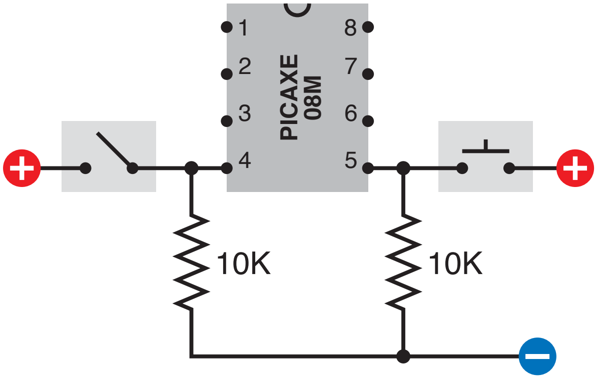

Figure 5-143. *The PICAXE can respond to the state of a switch or button attached to any of its input-capable pins. A 10K resistor must be used to pull down the state of the pin when the contact of the switch or button is open. Otherwise, you may get unpredictable results.*

Experiment 35: Checking the Real World

Often we want a microcontroller to measure something and respond in an appropriate way. For instance, it can measure a low temperature and sound an alarm, as I suggested in the example that I gave earlier.

The PICAXE has three analog-to-digital converters (ADCs) built in, accessible via logic pins 1, 2, and 4, as shown in Figure 5-139. The best way to use them is by applying a potential somewhere between 0 and 5 volts. In this experiment, I'll show you how to calibrate the response of the chip.

You will need:

- Trimmer potentiometer, 2K. Quantity: 1.
- PICAXE 08M chip and associated USB cable and socket. Quantity: 1 of each.

Procedure

Take the same trimmer potentiometer that you used in Experiment 32 and wire its center terminal to Logic Pin 2 of the PICAXE (which is hardware pin 5). The other two terminals of the 2K trimmer go to positive and to negative, respectively. So depending how you set the trimmer, the pin of the PICAXE is directly connected to positive (at one end of the scale), or directly connected to negative (at the other end of the scale), or somewhere in between. See Figure 5-144 for the revised schematic, and Figure 5-145 for a photograph of the breadboarded circuit.

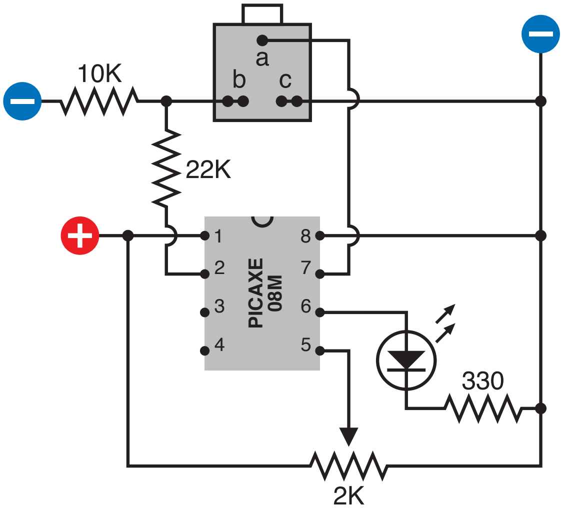

Figure 5-144. *This schematic, drawn in a layout suitable for breadboarding, shows how a 2K potentiometer can be used to apply a varying voltage to one of the pins of the PICAXE that is capable of converting an analog signal to a digital value.*

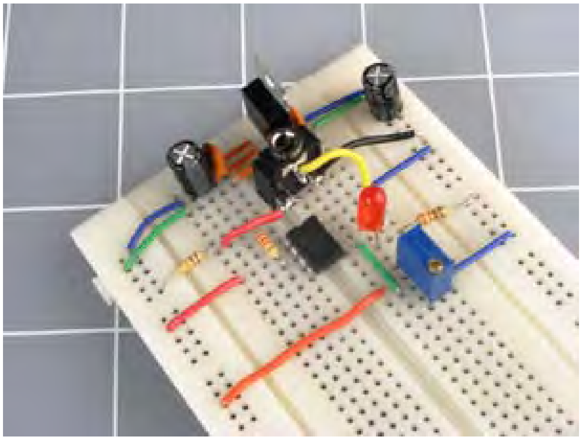

Figure 5-145. *The trimmer potentiometer added to the previously breadboarded circuit.*

Now we need a program to tell the chip what to do. Using the Programming Editor, start a new document. The code should look like this:

```
main:
    readadc 2,b0
    debug b0
    goto main
```

The command "readadc 2,b0" means "read the analog input on Logic Pin 2, convert from analog to digital, and store the result in b0."

The command "debug b0" tells the chip to go into program debugging mode, in which it uses its USB cable to tell the Programming Editor the values of all the variables while the program is running. The variables are displayed in a debugging window.

Download the program, and as the program starts to execute, the debugging window should open. Start adjusting the trimmer while looking at the value of b0, and you'll see b0 change its value.

You can make a table and draw a graph showing the relationship between the resistance between Logic Pin 2 and ground, and the value of b0. Just pull the trimmer off the breadboard, measure its resistance with a meter, then increase its resistance by, say, 200Ω, put it back into the breadboard, and look at the value of b0 again.

This is laborious, but calibrating equipment is always laborious—and in any case, I decided to do it for you. The graph is shown in Figure 5-146. You can also see the raw data numbers in the following table. I was pleased to find that the PICAXE gives a very precise, linear response to the input voltage. In other words, the graph is a straight line.

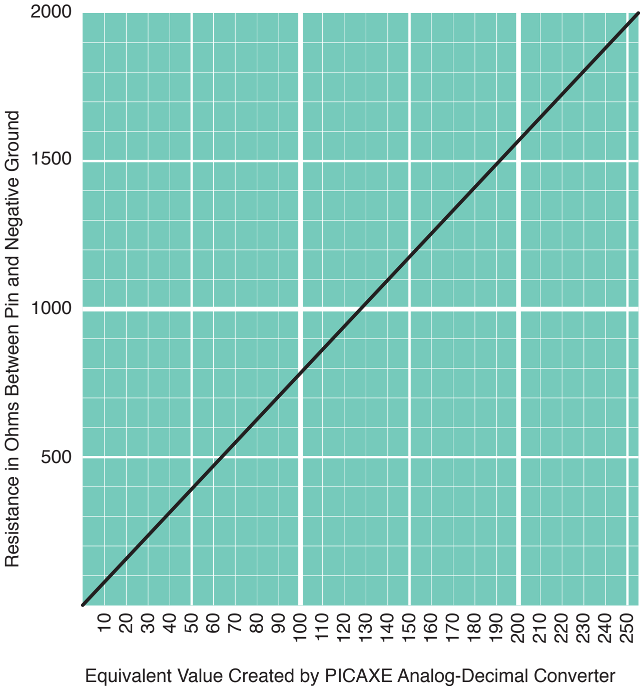

Figure 5-146. *When an ADC input pin is hooked up to a 2K potentiometer, which is connected across the same voltage that powers the chip, you should find that the resistance between the input pin and the negative side of the power supply generates the series of digital values shown on the graph. Note that the potentiometer must have a 2K value, and the power supply is assumed to be precisely 5 volts.*

This table shows measurements made with PICAXE 08M controller.

Resistance (in ohms) between the ADC pin and the negative supply	Equivalent digital value
2000	255
1900	243
1800	230
1700	218
1600	205
1500	192
1400	179
1300	166
1200	154
1100	141
1000	128
900	115
800	102
700	90
600	77
500	64
400	51
300	38
200	26
100	13
0	0

Now we can modify the program to make it do something with the information that it's taking in:

```
main:
    readadc 2,b0
    let w1 = 5 * b0
    high 1
    pause w1
    low 1
    pause w1
    goto main
```

Notice what's happening here. First we get a value in b0, and then on the next line, we do some arithmetic with it. The asterisk means "multiply." So the statement says, "Take whatever value is in b0, multiply by 5, and transfer it to another variable, w1." We have to use a w variable, because when we multiply the value of b0 by 5, we may get a number that is bigger than 255—too big to fit into a byte variable.

Finally, we take variable w1 and use it with a "pause" statement instead of a fixed number value. We're saying to the PICAXE, "pause for whatever number of microseconds you get by checking the value of w1."

So the software checks a variable resistance, turns it into a number, and applies that number to adjust the flashing speed of the LED.

Think back to the need of the cart powered by stepper motors. It was supposed to check two photoresistors, and adjust the speed of each motor accordingly. Well, this PICAXE program is a step in that direction. It can measure voltage on a pin and change the output frequency on another pin. If you had two PICAXE chips, you could wire each of them to a photoresistor and a motor. Then you could adjust the behavior of your cart by editing the second line in the program, where it converts the value of b0 to the value of w1 which will be used in the "pause" command to determine the number of pulses per second. Instead of multiplying by 5 you could multiply by 7 or whatever number gives you the result you need. This leads to an important conclusion: *a big advantage of a programmable chip is that you can make adjustments in software*.

Because the PICAXE 08M actually has more than one ADC input, and has three pins that can be used for output, you might wonder whether you could use just the one chip to control both motors in response to inputs from two sensors. The problem is that the three output pins on the 08M also function as the three ADC input pins. You'd do better to buy one of the more advanced PICAXE chips, such as the 18M, which has more pins to choose from. It uses the same basic set of programming instructions, and doesn't cost much more money.

Also, you should read the PICAXE documentation and look up the "pwmout" command, which is short for "pulse-width modulation output," but you can think of as meaning "power motor output." This is specifically intended to run stepper motors. It establishes an output frequency of pulses that will continue while the chip obeys other instructions in its program.

FUNDAMENTALS

Extra features

A complete guide to the 08M would fill a book of its own, and of course such books already exist (just search the books section of Amazon.com for keyword "picaxe"). But I'll finish my introduction to the controller by listing some of its extra capabilities, leaving you to look them up and explore them. Then I'm going to suggest one last experiment.

Interrupts

The PICAXE 08M allows you to set one "interrupt." This feature tells the chip to make a mental note that if a particular event occurs—such as a switch applying voltage to one pin—it should stop doing whatever else it was doing, and respond to the interruption.

Infrared

One pin on the PICAXE 08M can be used to receive infrared signals from a TV-style remote that you can buy from the same suppliers that sell the PICAXE itself. With an infrared sensor attached to the chip, you can issue commands remotely. If you want to build a remote-controlled robot, the chip is specifically designed with this in mind.

Servo motors

Every PICAXE chip has at least one pin that can send a stream of pulses to control a typical servo motor. On the 08M chip, it's Logic Pin 2. The width of each pulse tells the motor how far to rotate from its center position before stopping. A 555 timer can send this stream, but the PICAXE makes it easier. You can search online for more information about servo motors, which are especially useful for applications such as steering model vehicles, adjusting the flaps on model airplanes, and actuating robots.

Music

The PICAXE has an onboard tone generator that can be programmed with a "tune" command to play tunes that you write using a simple code.

Alphanumeric input/output

The "kbin" programming command is available in the PICAXE models 20X2, 28X1 and 28X2, and 40X1 and 40X2. You can plug a standard computer keyboard into the chip, and it will read the keypresses. You can also attach alphanumeric displays, but these procedures are nontrivial. For instance, when you're trying to figure out which key someone has pressed on a keyboard, your program has to contain a list of the special hexadecimal codes that the keyboard creates.

Pseudorandom number generation

All PICAXE models can generate pseudorandom numbers using a built-in algorithm. If you initialize the number generator by asking the user to press a button, and you measure the arbitrary time that this takes, you can seed the pseudorandom number generator with the result, and the pseudorandom number generator will have a less repeatable sequence.

Visit *http://www.rev-ed.co.uk/docs/picaxe_manual1.pdf* to learn more.

Experiment 36: The Lock, Revisited

The combination lock that I described in Experiment 20 is especially appropriate for a microcontroller, because it requires a series of operations that resemble a computer program. I'm going to show how this project can be redesigned using a PICAXE 08M, and then leave it to you to consider how some of the other projects in this book could be converted.

You will need:

- The same type of keypad and relay recommended in Experiment 20.
- A transistor or Darlington array to amplify the output from the PICAXE so that it can drive the relay.

Getting the User Input

Any of the input pins on the PICAXE can sense a switch closing. The trouble is that we only have three pins capable of doing this, and even the most advanced PICAXE chip has fewer than 10 such pins. So how can we attach a 10-key keypad to the 08M?

I have a suggestion: attach various resistors to the keypad, so that each key applies a different voltage to one of the ADC pins. Then use the ADC feature to convert the voltage to a number, and use a table of possible numbers to figure out which key is being pressed. This may not be the most elegant solution, but it works!

The keypad can be wired as shown in Figure 5-147. The asterisk key is still being used to supply power, as in the original experiment, while the pound key resets the relay at the end of your computing session, as before.

Current flows through a series of resistors, beginning with one that has a value of 500Ω. Because this is not a standard value, you will either have to make it by combining other resistors in series, or by presetting a trimmer potentiometer. After that, each button is separated from the next button by a 100Ω resistor. Finally, at the end of the chain, a 600Ω resistor separates the last button from the negative side of the power supply. Again, this is not a standard value, and you may have to use a trimmer.

Add up all the resistances and you have 2K, which is the range that the PICAXE wants us to use. When you press a button, you tap into the chain of resistances. Button 9 puts 600Ω between the PICAXE ADC pin and ground. Button 6 is 700Ω, button 3 is 800Ω, and so on. (You may prefer to lay out the buttons so that the resistance progresses in a more logical fashion. That's up to you. I chose to lay them out in the way that would be easiest to visualize on a keypad.)

Now look back at the ADC values that I supplied in the table on page 308. These are the values that you should get when you press various keypad buttons—but you cannot count on them being absolutely precise, because they may vary if your resistor values are not quite accurate, or if your power supply isn't exactly 5 volts. It's not safe to say, for example, that the PICAXE will deliver an ADC conversion value of precisely 77 when the resistance is 600Ω. It's safer to say that the value will be between 71 and 83. If we specify a range as shown in the following table, we have a much better chance of interpreting each button correctly.

Button number		Range
9	600	71–83
6	700	84–96
3	800	97–108
0	900	109–121
8	1000	122–134
5	1100	135–147
2	1200	148–160
7	1300	161–172
4	1400	173–185
1	1500	191–198

Suppose you attach the common pin of your keypad to ADC Logic Pin 2 of the PICAXE. You can now use the Program Editor to write a program that looks like this:

```
getkey:
    readadc 2,b0
    let b1 = 9
    if b0 < 84 then finish
    let b1 = 6
    if b0 < 97 then finish
    let b1 = 3
    if b0 < 109 then finish
    let b1 = 0
    if b0 < 122 then finish
    let b1 = 8
    if b0 < 135 then finish
    let b1 = 5
    if b0 < 148 then finish
    let b1 = 2
    if b0 < 161 then finish
    let b1 = 7
    if b0 < 173 then finish
    let b1 = 4
    if b0 < 186 then finish
    let b1 = 2

finish:
    return
```

What does the word "return" mean at the end? I'll get to that in a second. I want to explain the rest of the routine first.

b0 receives the value supplied by the analog-digital converter when it looks at the keypad. After storing the number in b0, the routine has to figure out which keypad key it matches. The key identity (0 through 9) will be stored in another variable, b1.

The program starts by assigning value 9 to b1. Then it checks to see whether b0 < 84. This means "if b0 is less than 84." If it is, then the routine tells the PICAXE to "finish," which means "jump to the finish: label." But if b0 is *not* less than 84, by default the PICAXE continues on to the next line, which makes a second attempt at guessing which key has been pressed. It assigns number 6 to b1. Now there's another if-then test—and so on. This process of reassigning values to b1 stops only when it gets to the point where b0 is greater than a number in the table.

If you're familiar with other dialects of BASIC, this may seem a bit laborious to you. You may wonder why we can't use a statement such as this:

```
    if b0 > 70 and b0 < 84 then b1 = 9
```

The answer is that PICAXE BASIC isn't sufficiently sophisticated to allow this. An if-then statement has to result in a jump to another section of the program. That's the only permitted outcome.

If you don't have any prior programming experience, the routine may still seem laborious to you, and perhaps a bit puzzling, too. This is understandable, because you're getting a crash course in software design without any formal preparation. Still, the PICAXE Programming Editor can be a big help, because it has its simulation feature. Before you can use this, though, you have to precede the routine that I just supplied with a control routine that you must type above it. The screenshot in Figure 5-148 shows you how it should look.

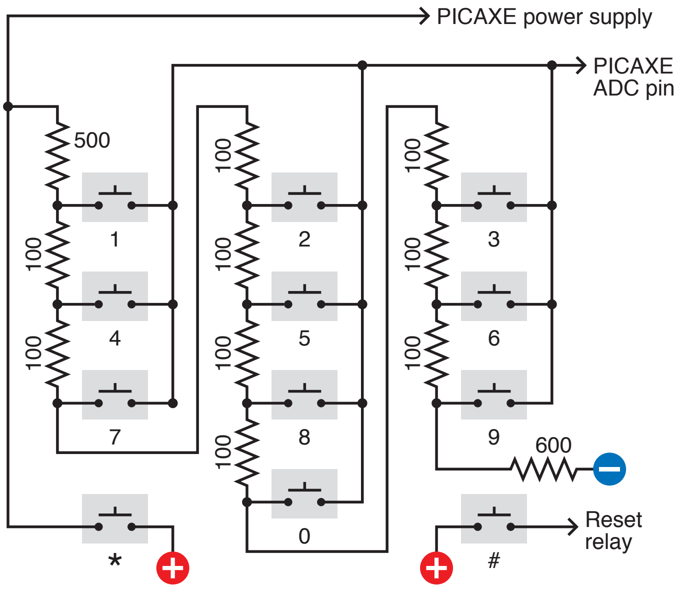

Figure 5-147. *A quick and simple way of attaching a keypad to provide numeric input to the PICAXE uses a chain of resistors totalling 2,000Ω. When a button is pressed, it connects the ADC input pin to a point in the chain. The resistance detected by the input pin can then be converted by the program in the chip to determine which key has been pressed.*

I have chosen an arbitrary combination of 7-4-1 for our combination lock. Using these numbers, the main section of the program looks like this:

```
main:
    low 1
    gosub getkey
    if b1<>7 then main
    gosub getkey
    if b1<>4 then main
    gosub getkey
    if b1<>1 then main
    high 1
    end
```

I should explain that the <> pair of symbols mean "is not equal to." So the fourth line of the program means, "if b1 is not equal to 7."

The value of b1 is supposed to be 7 if the user is putting in the correct combination. So if it's not 7, the user has entered the wrong value, and the if-then statement sends the PICAXE back to the beginning. In fact anytime the user inputs a number that is not in the correct 7-4-1 sequence, the program sends the PICAXE back to the beginning. This is the way the pure-hardware version of this experiment was set up.

But what is this word "gosub"? It means "go to a subroutine." A subroutine is any sequence of program statements that ends with the instruction to "return." So "gosub getkey" tells the PICAXE to mark its current place in the program while it skips to the `getkey:` section of code, which it obeys, until it finds the word "return," which returns it to the place from where it came.

The PICAXE continues in this fashion until it reaches the word "end." I had to insert the word "end" because otherwise the PICAXE will continue executing the program and will fall into the subroutine. "End" stops it from doing so. Figure 5-148 shows a screenshot of the complete listing.

So—is that all? Yes, that's it. If you enter the code into the Programming Editor exactly as I have supplied it, you should be able to run it in simulation mode, and in the simulation window, click the right-arrow beside Logical Pin A2 to increase its value in steps. Each time you pass one of the values in the `getkey:` subroutine, you should see the value for variable b1 change in the display.

This is really all you need to perform the functions of the combination lock. When the PICAXE runs this program, it waits for the correct combination. If it receives the combination, it sends the output from logical pin 1 high; otherwise, logical pin 1 stays low.

The only additional item you need is a transistor or CMOS gate between logical pin 1 and the relay that unlocks the computer, because the PICAXE cannot deliver enough current to operate the relay by itself.

Putting this procedure into a controller chip not only simplifies the circuit, but offers another advantage: you can change the combination simply by rewriting the program and downloading the new version into the chip.

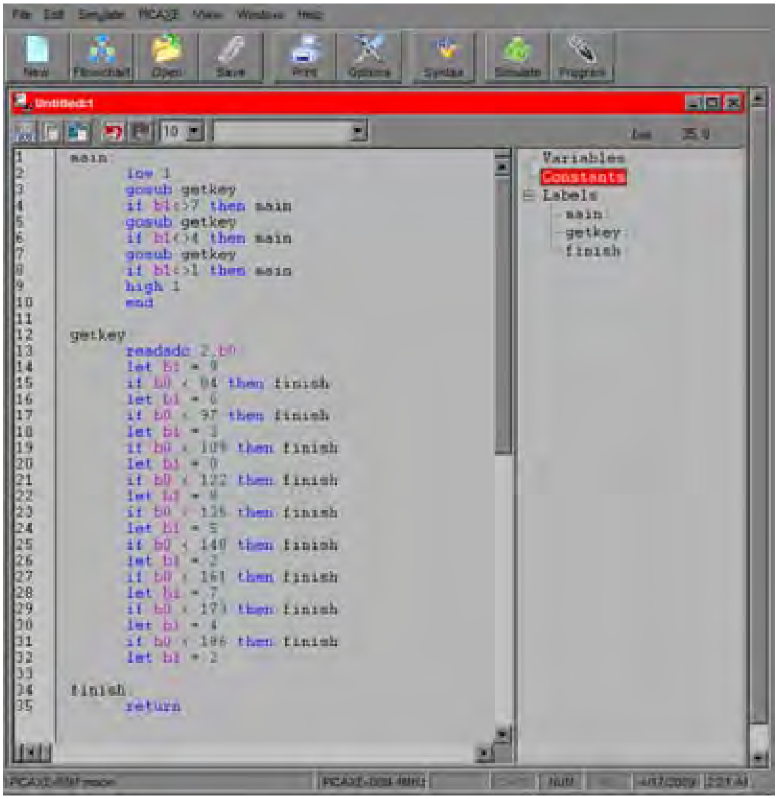

```
main:
        low 1
        gosub getkey
        if b1<>7 then main
        gosub getkey
        if b1<>4 then main
        gosub getkey
        if b1<>1 then main
        high 1
        end

getkey:
        readadc 2,b0
        let b1 = 9
        if b0 < 84 then finish
        let b1 = 6
        if b0 < 97 then finish
        let b1 = 3
        if b0 < 109 then finish
        let b1 = 0
        if b0 < 122 then finish
        let b1 = 8
        if b0 < 135 then finish
        let b1 = 5
        if b0 < 140 then finish
        let b1 = 2
        if b0 < 161 then finish
        let b1 = 7
        if b0 < 171 then finish
        let b1 = 4
        if b0 < 186 then finish
        let b1 = 2

finish:
        return
```

Figure 5-148. *This screenshot shows the complete listing of a program to read a sequence of three keypresses in conjunction with a combination lock. If the sequence is correct, the PICAXE sends a high output from one of its pins. If the sequence is incorrect, the program loops back to the beginning.*

FUNDAMENTALS

Limitations of MCUs

The PICAXE does have some disadvantages. Its voltage requirements alone restrict you from using it with the kind of freedom of a 555 timer.

Also, although I can get an instant result by plugging a 555 timer into a breadboard and adding a couple of resistors and a couple of capacitors, the PICAXE requires me to add a download socket, hook it up to my computer, write a program in the Programming Editor, and download the program.

Some people don't like writing software, or they have difficulty thinking in the relentlessly left-brain way that computer programming requires. They may prefer the hands-on process of assembling hardware.

Other people may have the opposite preference. This of course is a matter of taste, but one thing we know beyond all doubt is that computer programs often contain errors that may not reveal themselves until weeks or months later. The PICAXE, for instance, doesn't protect you if a number is assigned to a variable that exceeds the limit for that type of variable. Suppose b1=200 and b2=60 and your program tells the PICAXE:

```
let b3 = b1 + b2
```

The result should be 260, but byte-size variables can only count up to 255. What happens? You will find that b3 acquires a value of 4, without any warning or explanation. This is known as an "overflow error," which can be very difficult to predict, because it happens at runtime, when external factors are in control. The code looks perfectly good; the Programming Editor doesn't find any syntax errors; the simulation behaves properly. But in the real world, days or even months later, an unexpected set of circumstances results in an input that causes the overflow, and because the code is residing inside the chip at this point, you may have a hard time figuring out what on earth went wrong.

Software has its problems. Hardware has its advantages.

FUNDAMENTALS

Unexplored territory

If you've taken the time to complete most of the projects in this book with your own hands, you have gained a very rapid introduction to the most fundamental areas of electronics.

What have you missed along the way? Here are some topics that remain wide open for you to explore. Naturally you should search online if they interest you.

The informal, learning-by-discovery approach that I have used in this book tends to be light on theory. I've avoided most of the math that you'd be expected to learn in a more rigorous course on the subject. If you have mathematical aptitude, you can use it to gain a much deeper insight into the way in which circuits work.

I didn't deal much with computer architecture, either. We didn't go very far into binary code, and you didn't build a half-adder, which is a great way to learn how computers function on the most fundamental level. Perhaps you should think about assembling one.

I avoided going deeply into the fascinating and mysterious properties of alternating current. Here again, some math is involved, but just the behavior of current at high frequencies is an interesting topic in itself.

For reasons already stated, I avoided surface-mount components—but you can still go into this area yourself for a relatively small investment, if you like the idea of creating fascinatingly tiny devices. This may be the future of hobby electronics, so if you stick with it, you'll probably end up in the world of surface-mount.

Vacuum tubes were not mentioned, because at this point, they are mainly of historical interest. But there's something very special and beautiful about tubes, especially if you can enclose them in fancy cabinetwork. In the hands of a skilled craftsperson, tube amplifiers and radios become art objects.

I didn't show you how to etch your own printed circuit boards. This is a task that appeals to only certain people, and the preparation for it requires you to make very neat drawings or use computer software for that purpose. If you happen to have those resources, you might want to do your own etching. It could be a first step toward mass-producing your own devices.

I didn't cover static electricity at all. High-voltage sparks don't have any practical applications, and they entail some safety issues—but they are stunningly impressive, and you can easily obtain the necessary information to build the equipment. Maybe you should try.

Other Controllers

If you want something more powerful, a BASIC Stamp is the logical next step after a PICAXE. the BASIC Stamp is so called because it originally looked like a postage stamp. The BASIC Stamp has a larger vocabulary of commands and a bigger range of add-on devices (including displays with graphical capability, and a little keyboard that is specifically designed for use with the controller). The BASIC Stamp is shown in Figure 5-149.

On the downside, you'll find that everything associated with the BASIC Stamp is a bit more expensive than in the PICAXE world, and the download procedure isn't quite as simple.

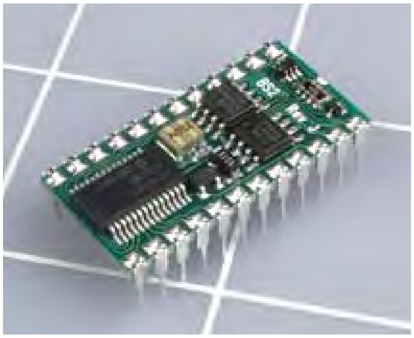

Figure 5-149. *The BASIC Stamp controller consists of surface-mounted components on a platform that has pins spaced at 1/10-inch intervals, for insertion in a breadboard or perforated board. This component uses a version of BASIC that is similar to the programming language of the PICAXE, but has many more extensions. The BASIC Stamp is available for use with a wide range of peripheral devices, including many alphanumeric dot-matrix displays.*

One of the more recent developments in the world of MCUs is the Arduino, which is both sophisticated and powerful. It does require programming in the C language. This language is a little more difficult to understand, and has only the vaguest similarity to the syntax that is used in the PICAXE and BASIC Stamp. On the other hand, because C dominates the larger world of computing, learning it might not be such a bad idea—and the Arduino offers some truly amazing capabilities. Because it is so popular, there are also many software tools, documentation, user forums, and many enthusatic hobbyists to help you. Two other Make: Books titles that I mentioned previously, *Getting Started with Arduino* and *Making Things Talk*, provide a great introduction.

In Closing

I believe that the purpose of an introductory book is to give you a taste of a wide range of possibilities, leaving you to decide for yourself what you want to explore next. Electronics is ideal for those of us who like to do things ourselves, because almost any application—from robotics, to radio-controlled aircraft, to telecommunications, to computing hardware—allows opportunities that we can explore at home, with limited resources.

As you delve deeper into the areas of electronics that interest you most, I trust you'll have a satisfying learning experience. But most of all, I hope you have lots of fun along the way.

Online Retail Sources and Manufacturers

A

This appendix contains URLs for companies mentioned as retail sources or manufacturers, along with the commonly used name of the source and the company name.

Colloquially used name	Actual corporate or company name	URL
3M	Minnesota Mining and Manufacturing Co.	*http://solutions.3m.com/en_US/*
Ace Hardware	Ace Hardware Corporation	*http://www.acehardware.com*
Advanced Micro Circuits	Advanced Micro Circuits Corp	*http://www.advancedmicrocircuits.com*
Akro-Mils	Myers Industries, Inc.	*http://www.akro-mils.com*
Alcoswitch	Division of Tyco Electronics Corporation	*http://www.tycoelectronics.com/catalog/menu/en/18025*
All Electronics	All Electronics Corporation	*http://www.allelectronics.com*
All Spectrum Electronics	All Spectrum Electronics	*http://www.allspectrum.com*
All-Battery.com	Tenergy Corporation	*http://www.all-battery.com*
Alpha potentiometers	Alpha Products Inc.	*http://www.alphapotentiometers.net*
ALPS pushbutton	ALPS Electric Co., Ltd.	*http://www.alps.com*
Amazon	Amazon.com, Inc.	*http://www.amazon.com*
Amprobe	Amprobe Test Tools	*http://www.amprobe.com*
Arduino	No corporate identity	*http://www.arduino.cc*
ArtCity	ArtCity	*http://www.artcity.com*
AutoZone	AutoZone, Inc.	*http://www.autozone.com*
Avago	Avago Technologies	*http://www.avagotech.com*
BASIC Stamp	Brand owned by Parallax, Inc.	*http://www.parallax.com*
BI Technologies	BI Technologies Corporation	*http://www.bitechnologies.com*
BK Precision	B&K Precision Corp.	*http://www.bkprecision.com*
Bussmann fuses	Cooper Bussman, Inc.	*http://www.cooperbussmann.com*
C&K switch	CoActive Technologies, Inc.	*http://www.ck-components.com*
Chicago lighting	CML Innovative Technologies	*http://www.cml-it.com*
CraftAmerica	Cardinal Enterprises	*http://www.craftamerica.com*
Darice	Darice Inc.	*http://www.darice.com*

Colloquially used name	Actual corporate or company name	URL
DeWalt	DeWalt Industrial Tool Company	*http://www.dewalt.com*
Digi-Key	Digi-Key Corporation	*http://www.digikey.com*
Directed switches	Directed Electronics Inc.	*http://www.directed.com*
Doctronics	Doctronics Educational Publishing	*http://www.doctronics.co.uk*
eBay	eBay Inc.	*http://www.ebay.com*
Elenco	Elenco Electronics Inc.	*http://www.elenco.com*
Everlight	Everlight Electronic Co. Ltd.	*http://www.everlight.com*
Extech	Extech Instruments Corporation	*http://www.extech.com*
Fairchild	Fairchild Semiconductor Incorporated	*http://www.fairchildsemi.com*
FTM	FTM Incorporated	*http://thefabricatorssource.com*
Fujitsu	Fujitsu America, Inc.	*http://www.fujitsu.com/us/*
GB wire strippers	Gardner Bender Inc.	*http://www.gardnerbender.com*
Hobbylinc	Hobbylinc Hobbies	*http://www.hobbylinc.com*
Home Depot	Homer TLC, Inc.	*http://www.homedepot.com*
Ideal wire strippers	Ideal Industries Inc.	*http://www.idealindustries.com*
Jameco	Jameco Electronics	*http://www.jameco.com*
K&J Magnetics	K&J Magnetics Inc.	*http://www.kjmagnetics.com*
Kingbright	Kingbright Corporation	*http://www.kingbrightusa.com*
Kobiconn	No web page found; use mouser.com	
KVM Tools	KVM Tools Inc.	*http://www.kvmtools.com*
Lowe's hardware	LF, LLC	*http://www.lowes.com*
Lumex	Lumex Inc.	*http://www.lumex.com*
McMaster-Carr	McMaster-Carr Supply Company	*http://www.mcmaster.com*
Megahobby	Megahobby.com	*http://www.megahobby.com*
Meter Superstore	Division of SRS Market Solutions Inc.	*http://www.metersuperstore.com*
Michaels craft stores	Michaels Stores, Inc.	*http://www.michaelscrafts.com*
Mill-Max	Mill-Max Manufacturing Corp.	*http://www.mill-max.com*
Mitutoyo	Mitutoyo America Corporation	*http://www.mitutoyo.com*
Motorola	Motorola, Inc.	*http://www.motorola.com/us*
Mouser electronics	Mouser Electronics, Inc.	*http://www.mouser.com*
Mueller alligator clip	Mueller Electric Company	*http://www.muellerelectric.com*
Newark	Subsidiary of Premier Farnell plc	*http://www.newark.com*
NKK switches	Nihon Kaiheiki Industry Co. Ltd.	*http://www.nkkswitches.com*
NXP semiconductors	NXP Semiconductors	*http://www.nxp.com*
Omron	Omron Corporation	*http://www.omron.com*
On Semiconductor	Semiconductor Components Industries, LLC	*http://www.onsemi.com*
Optek	Subsidiary of TT Electronics plc	*http://www.optekinc.com*
Panasonic	Panasonic Electric Works Corporation	*http://pewa.panasonic.com*

Colloquially used name	Actual corporate or company name	URL
PanaVise	Panavise Products, Inc.	*http://www.panavise.com*
Parallax	Parallax, Inc.	*http://www.parallax.com*
Pep Boys	Pep Boys-Manny, Moe and Jack	*http://www.pepboys.com*
Philips	Koninklijke Philips Electronics N.V.	*http://www.usa.philips.com*
PICAXE	Revolution Education Ltd.	*http://www.rev-ed.co.uk*
Piedmont Plastics	Piedmont Plastics, Inc.	*http://www.piedmontplastics.com*
Plano storages boxes	Plano Molding Company	*http://www.planomolding.com*
Pomona test equipment	Pomona Electronics Inc.	*http://www.pomonaelectronics.com*
RadioShack	RadioShack Corporation	*http://www.radioshack.com*
Sears	Sears Brands, LLC	*http://www.sears.com*
SparkFun Electronics	Sparkfun Electronics	*http://www.sparkfun.com*
Stanley tools	The Stanley Works	*http://www.stanleytools.com*
STMicroelectronics	STMicroelectronics Group	*http://www.st.com*
Texas Instruments	Texas Instruments Incorporated	*http://www.ti.com*
Tower Hobbies	Tower Hobbies	*http://www.towerhobbies.com*
Twin Industries	Twin Industries	*http://www.twinind.com*
Tyco	Tyco Electronics Corporation	*http://www.tycoelectronics.com*
Vaughan	Vaughan & Bushnell Mfg.	*http://hammernet.com/vaughan/*
Velleman keyboards	Velleman nv	*http://www.velleman.eu*
Vishay	Vishay Intertechnology Inc.	*http://www.vishay.com*
Wal-Mart	Wal-Mart Stores, Inc.	*http://www.walmart.com*
Weller	Division of Cooper Industries, LLC	*http://www.cooperhandtools.com/brands/weller/*
X-Acto	Division of Elmer's Products, Inc.	*http://www.xacto.com*
Xcelite	Division of Cooper Industries, LLC	*http://www.cooperhandtools.com/brands/xcelite/*
Xytronic	Xytronic Industries Ltd.	*http://www.xytronic-usa.com*

Secrets of Building Electrostatic
LIGHTNING BOLT GENERATORS

including high voltage test equipment, experiments, motors and more!

by Walt Noon

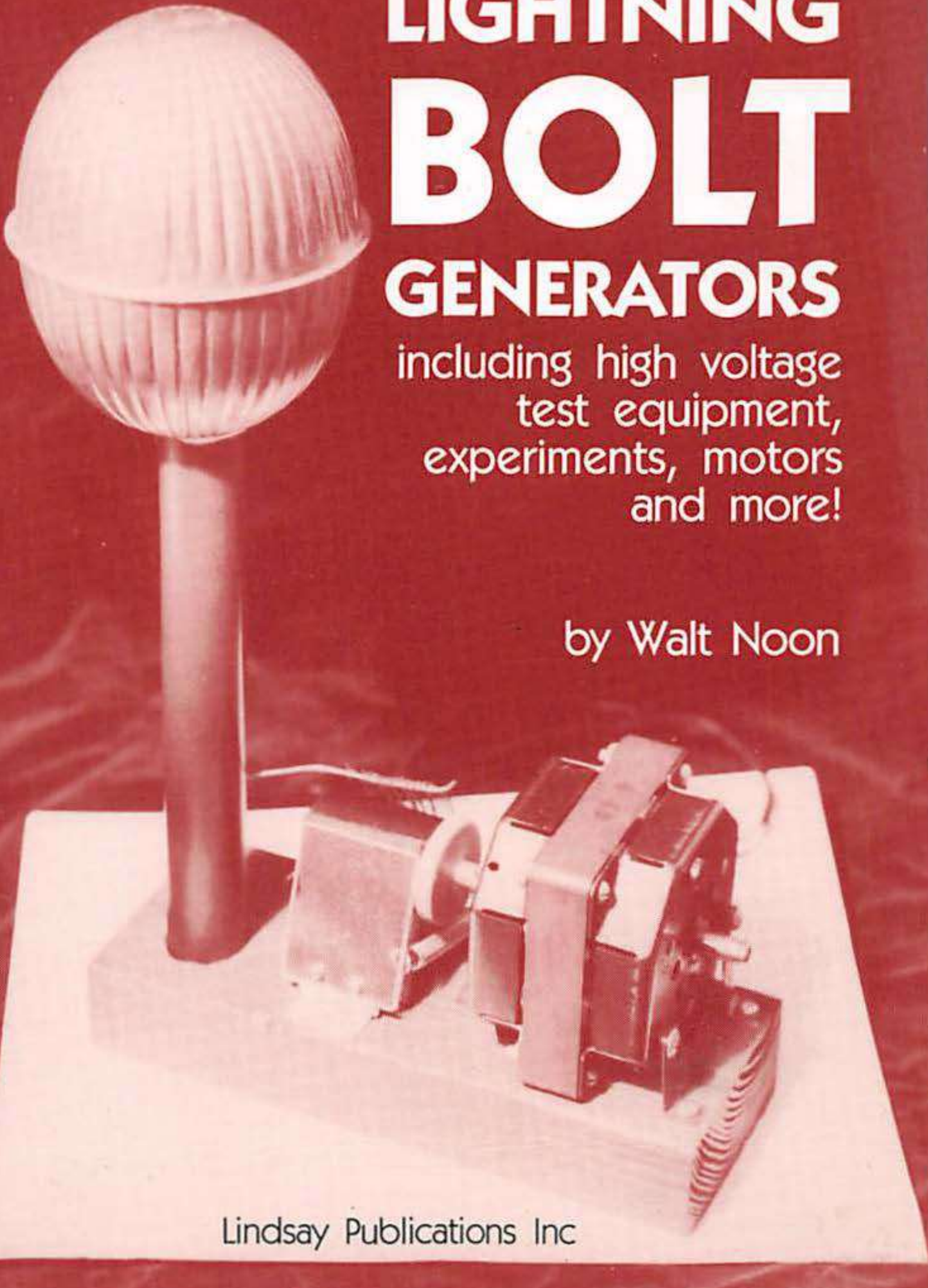

Lindsay Publications Inc

Secrets of Building Electrostatic

LIGHTNING BOLT GENERATORS

including high voltage test equipment, experiments, motors and more!

by Walt Noon

Lindsay Publications Inc

Copyright ©1992 by Lindsay Publications Inc

All rights reserved.

Lindsay Publications Inc
PO Box 12
Bradley IL 60915-0012

ISBN 1-55918-090-0

2 3 4 5 6 7 8 9 0

To my Mom and Dad who
always took the time to answer
my questions

Original CAD drawings
and photographs by
Matthew J. Roddy

Cartoon drawings by
Mike Noon

Special thanks to
Kevin E. Curtin
Matthew J. Roddy
Mike Noon

WARNING

The author of this book is not a professional engineer nor has he had formal training in the design or operation of high voltage devices. The author is an amateur but has been successful in building and operating the devices discussed herein.

The methods that he used and describes are presented merely as guidelines for other amateurs in developing similar devices. These devices can be dangerous, possibly even lethal, and dangers have been pointed out wherever possible. Since the author is not a professional in this field, there may well be other dangers involved in the building and operation of the devices described.

The author hereby disclaims any liability for injury to persons or property that may result while using this project book, and the author does not intend by this publication to explain all dangers known or unknown that may exist in the building and operation of this project.

Lindsay Publications Inc has not built this project nor does it endorse the methods described. Lindsay Publications Inc assumes no liability for injury to person or property that may result form used of this information.

Contents

Electrophorus 4
Rotostatic Generator 9
"Cat-O-Static" Generator 17
Motor Control 18
External Van de Graaff Generator 23
Classic Van de Graaff Generator 31
Extremely High Power Van de Graaff Generator 39
Inductive Electrostatic Generator 42
High Voltage Measurements 46
FET Electroscope 47
Mechanical Electroscope 50
Neon Lamp Blank 51
Spark Gap 53
Leyden Jar 58
Water Capacitor 61
Plate Capacitor 62
Experiments 67
Ion Blower 76
Franklin Motor 78
Corona Motor 80
Atmospheric Electricity 81
Bibliography 86
Electroplating Terminals 87

Every text I've seen on the subject of electrostatics seems to begin with a history of the subject. A. D. Moore, one writer whose works I particularly enjoy, said that if cavemen had cats that rubbed up against them then they must certainly have been aware of the presence of static electricity. There are theories now going around the scientific community that claim lightning triggered the formation of the first chains of DNA. I'll take that one step further and guess that lightning began the evolutionary process that created our species.

Static electricity is one of the most fascinating and possibly most overlooked fields of modern electronics. We all take static for granted. Before I began experimenting with static electricity it would have been very difficult to convince me that it was easy to get 3500 volts from an old plastic bottle or 10,000 volts and more from a rubber band! Even now as I watch one of these tiny generators in action it is difficult to imagine that it is producing these voltages.

As you learn about electrostatics you will find many strange and wonderful experiments to perform. For example, it is possible to mysteriously levitate objects, blow a candle flame to one side with an invisible wind, and to light neon bulbs without wires.

There are motors that can be built to run on static electricity! Well-made electrostatic motors are so sensitive that they have been powered by the earth's natural electrical charge alone! Connect your motor to an antenna, and off it goes!

In addition to the many curiosities static electricity makes possible, there is also the potential for serious research with many astounding possibilities.

Consider the work of Professor John G. Trump, formerly of the Massachusetts Institute of Technology. Professor Trump points out that electrostatic forces are the most powerful forces known in nature – stronger by far than gravity or magnetism. Professor Trump demonstrates the practical application of electrostatic generators with the following illustration:

Consider two conductive plates, 100 square inches in area, facing each other and separated by an insulator. If the plates are each charged oppositely to 300 volts per centimeter the force of attraction between them will be one 2,000th of a pound. Increase this charge to 30,000 volts per centimeter and the attraction escalates to half a pound. Now increase this voltage to the order of three million volts per centimeter and the force of attraction jumps to 5,700 pounds! As Trump has said "Force of this order has more than a passing interest for power engineers." Trump and his colleagues had been pursuing the creation of electrostatic generators they hoped would one day rival modern electromagnetic generators.

Just how much electrostatic energy is theoretically possible? Consider an aluminum cube only one centimeter square. Imagine that all the electrons in that cube were removed and kept one meter away. The cube has only positive charges, and the electrons one meter away only negative charges. Since positive and negative charges attract, what would be the force of attraction generated between them? The answer turns out to be an astounding thirty-two million million million million pounds! Professor A.D. Moore, an expert in electrostatics, points out that this force would be equal to the weight of a steel cube 76 miles high!

Of course, as much fun and as inspiring as it is to ponder the theoretical world of electrostatics, back in the real world we find that electrons don't leave home so easily. The most useful electrostatic inventions to date make use of the 'little' forces of electrostatics. But keep that cube in mind and let everyone wonder why you have that curious smirk on your face whenever you run your generator.

What voltages are possible for the experimenter? The four easily assembled generators shown in the following text will give you an idea. Each one intentionally operates on a slightly different principal of gathering an electrostatic charge. If you build them, or just simply become familiar with them, you will develop a understanding of the most popular ways of generating static electricity.

I am confident that if you become acquainted with these machines you will have no difficulty at all in constructing many others of your own design. The reason that I am so confident is that I designed and constructed all of these machines to teach myself about electrostatics! Each machine represents the best of my thoughts at the time I constructed it.

You will soon find that it is easily possible to construct a static generator capable of producing 50 to 100 thousand volts! If you are a craftsman, 200 thousand volts and more can be obtained by optimizing the designs of the higher power generators such as the Van De Graaff shown later in the text.

Electrostatic generators are also remarkably inexpensive. With a little foraging through the surplus bins, no generator in this text cost me more than $30 to build! I know of no other way to approach these voltages for such a low cost.

A Little History of the First Practical Generators

The first electrostatic generators produced their voltage with friction. Towards the end of the 17th century Otto von Guericke, an amateur physicist from Magdeburg, developed the first practical electrical machine – an electrostatic generator. What follows are his own instructions for building one:

"Secure one of the glass globes which are called phials, about the size of a youngster's head; fill it with sulfur, ground in a mortar and melted by the application of flame. After it freezes, break the phial, take out the sulfur globe and keep it in a dry place, not a moist one. Perforate it with a hole so that it can spin upon an iron axle. Thus the globe is prepared.

"To demonstrate the power developed by this globe, place it with its axis on two supports in the machine – a hand's breadth above the baseboard – and spread under it various sorts of fragments such as bits of leaves, gold dust, silver filings, snips of paper, hairs, shavings, etc. Apply a dry hand to the globe so that it is stroked or grazed two or three times or more. Now it attracts the fragments and, as it turns on its axis, carries them around with it.

"When a feather is in contact with the globe, and afterwards in the air, it puffs itself out and displays a sort of vivacity ... and if someone places a lighted candle on the table and brings the feather to within a hand's breadth of the flame, the feather regularly darts back suddenly to the globe and, as it were, seeks sanctuary there."

Otto Von Guericke goes on to describe experiments by which he produced both light and sound from his amazing globe!

The instructions for von Guericke's globe are simple enough for any experimenter today to follow. The only note that I would add is that in casting my own globe I found that large quantities of sulfur would become much smaller pools when melted, and that if the globe were not filled on the first pouring it would not appear nearly as attractive when finished. So, remember to begin with a large amount of sulfur.

THE ELECTROPHORUS

Many generations passed and many ingenious machines were developed without great change using materials and processes similar to von Guericke's globe. The most important historically was Alessandro Volta's *electro perpetuo* known now as the electrophorus. With it we can dramatically demonstrate that electrostatic charges are much like magnetic forces in that like charges repel one another while opposite charges attract.

An electrophorus is very easily assembled. It consists simply of a rectangular block of insulating material such Lucite or, better yet, polyethylene and a metal disk which is attached to an insulating handle (Figure 1).

When the Lucite is rubbed with a woolen cloth, electrons are removed from the cloth and held on the Lucite. This happens because the cloth will hold its electrons less tightly than the Lucite. An insulating material, in this case Lucite, by its nature, will not allow the electrons to move around freely. The charges are therefore 'static' or held in place by the insulating material. The Lucite will now have a negative charge, and the wool will pick up a positive charge, respectively.

Next, the metal disk with the handle is set on top of the Lucite. The negative charges on the Lucite are attracted to positive charges on the metal disk, so, the positive charges in the metal disk move to the side of the metal plate facing the Lucite. The negative charges are similarly repelled to the back of the metal disk. Now, your hand or, better yet, a ground wire are touched briefly to the top (handle side) of the metal disk and then removed. In this way the negative charge is transferred to your hand (or ground) leaving only a positive charge. The metal plate will now have a positive charge, and the plastic negative.

If you have achieved a large enough charge, the Lucite can be picked up by the metal disk!

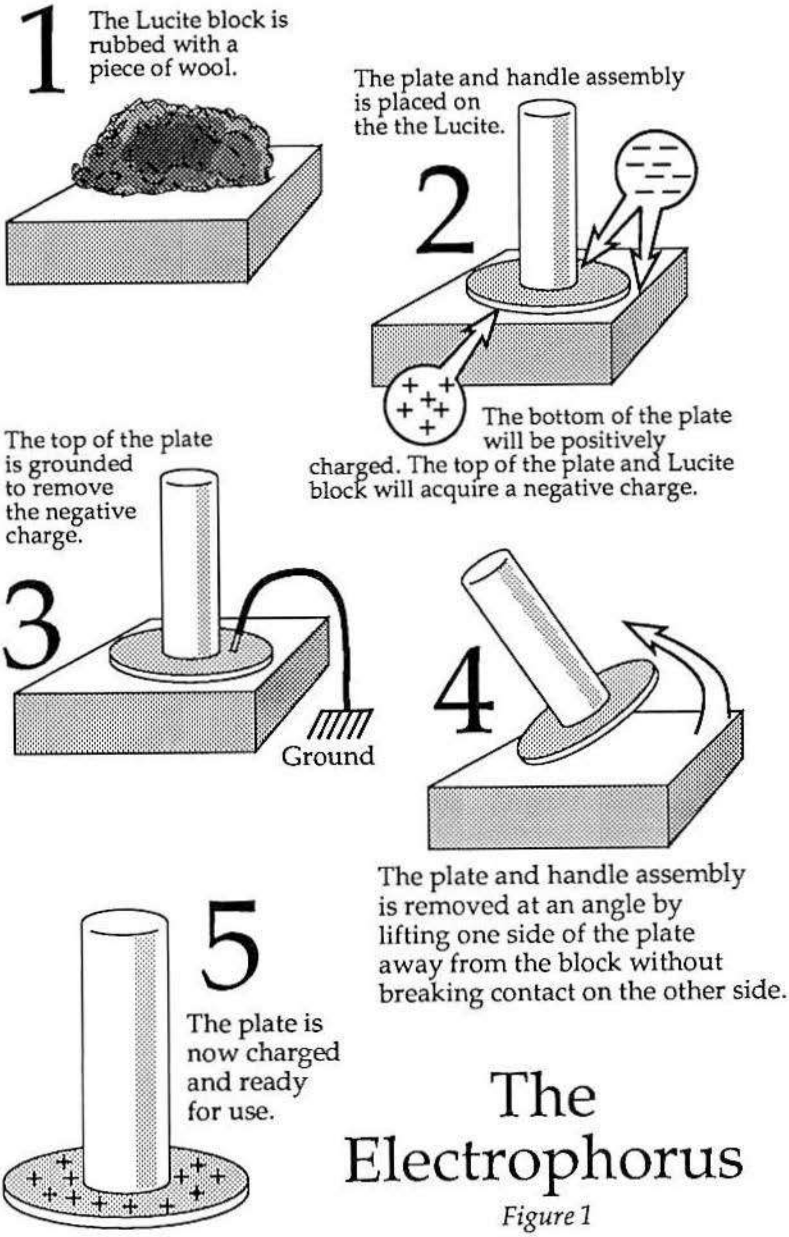

The Lucite block is rubbed with a piece of wool.

The plate and handle assembly is placed on the the Lucite.

The bottom of the plate will be positively charged. The top of the plate and Lucite block will acquire a negative charge.

The top of the plate is grounded to remove the negative charge.

The plate and handle assembly is removed at an angle by lifting one side of the plate away from the block without breaking contact on the other side.

The plate is now charged and ready for use.

The Electrophorus

Figure 1

Slowly remove the metal plate by the handle, drawing it away from the Lucite at an angle. The metal plate now holds the charge you seek. Having charged your electrophorus, you can perform the experiments with Von Guericke's globe mentioned above including attracting hair, styrofoam and other small objects.

Because the charged surface of the electrophorus is flat, a startling demonstration can be performed wherein the electrophorus disk is actually floated above the Lucite by electrostatic forces. Performing this trick takes some practice. Make a second metal plate and handle. Remember that if the plate is to float above the other it must be very light weight. (Figure 2)

Now, we have the challenge of charging both the disk and the Lucite with a like charge so that they will repel each other. Charge your first disk with a positive charge as described above, and then bring it close to, but not touching, the second disk. This will charge the second disk by attracting a negative charge to the front of the disk. Next, ground the handle side of the second disk to bleed off the positive charge. The second disk is now negatively charged. The Lucite is also negatively charged, so, provided everything has been accomplished and the weight is not too great, the second disk will now float above the Lucite block!

Electrophoruses act a great deal like capacitors discussed later on and can be used for many surprising experiments including charging a Leyden jar and flickering a neon bulb in a darkened room!

Special Notes on Friction Generators

Before we move on to the more powerful friction generators, there are a number of things you should be aware of that will help you in both troubleshooting and designing generators.

My first bit of advice is to avoid humidity. It makes air more conductive. Moisture in the air can have a devastating effect on the charges stored by electrostatic devices.

I was given a first rate example of this shortly after building my first two generators. Both machines, only days old in their construction, were whirling away in my shop. I was doing "careful scientific study" – in other words, I was zapping everything in sight to see what would happen. Suddenly, and most mysteriously, both generators

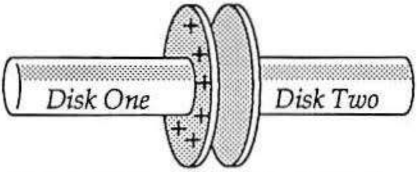
Disk One
Disk Two

Charge disk one on the Lucite block.
Then bring it close to a second disk.

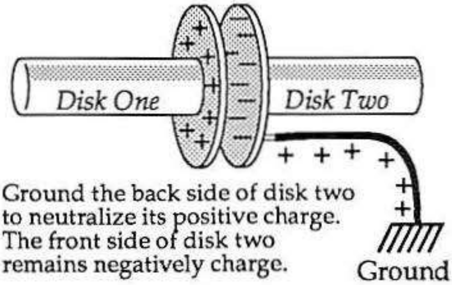
Disk One
Disk Two
Ground

Ground the back side of disk two to neutralize its positive charge. The front side of disk two remains negatively charge.

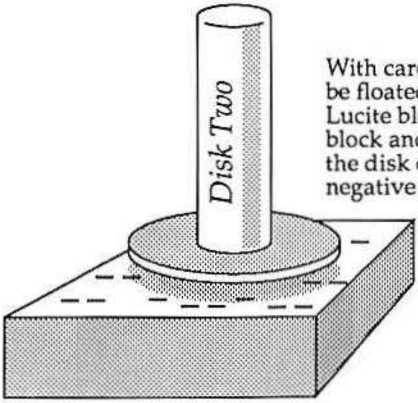

With care disk two can be floated above the Lucite block. Both the block and bottom of the disk carry repelling negative charges.

Floating the Electrophorus

Figure 2

stopped producing a charge. I couldn't imagine why both, each being very different in construction, might fail at exactly the same time. Then the culprit made itself known! Just outside the door the automatic lawn sprinklers had come on and extremely humid air was coming through my door!

Another good example of this problem is demonstrated by walking across a padded carpet. When the relative humidity is 20% or less, charges of up to 35,000 volts can be obtained by shuffling one's feet along the carpet. However, raise the humidity to 60% or more, and the maximum charge will plummet to a mere 1,500 volts! If you live in a humid region, you may not be able to get the same performance from a generator that someone living in the desert might. Hopefully, humidity will only be a seasonal problem for you. Dry winter days are best.

There are fascinating static generators which are more or less immune to humidity. These rely on generating a charge by induction rather than by friction. These machines will also be discussed in detail later on.

Be careful in selecting paint for your generator. I have found that many spray paints contain ingredients that will conduct static electricity. For the most part I have had few problems with this, although in a darkened room I have seen the telltale blue corona glow of escaping electrons coming from painted plastic surfaces on a generator. To correct the problem I simply coated the nearby metal surfaces that were losing their charges to the paint with high resistance "corona dope" sold in T.V. and better electronic stores.

Avoid using wood to build generators. Although we think of wood as an insulator, in reality it becomes an excellent conductor of very high voltages. In the machines that follow, you'll see polyvinyl chloride (pvc) pipe used extensively. If you can get acrylic or phenolic tubing, use it. It offers much better dielectric strength.

Dust can bleed away a charge. Dust on a high potential terminal can reduce your generator's efficiency by a whopping 40%. Because your high potential terminal carries a charge, dust will be attracted to it in much the same way that dust accumulates on the front of a T.V. screen. I have actually seen a tiny particle of dust prevent an eighth inch spark from arcing. Keeping your terminal clean is important.

On the other hand, the attraction of dust to your generator can be quite entertaining. One way that I enjoy cleaning my generator is to ground myself and then pass my hand over the high potential terminal. The dust which has been sitting there has acquired a high voltage charge and jumps to your hand which it sees as ground. The effectiveness of this method for really getting your generator clean is doubtful since after a few seconds the dust in your hand looses its charge and heads back for the generator, but it's lots of fun.

Once during a rare Southern California thunderstorm I observed a particularly confused particle of dust. Apparently the storm had charged the air, and my generator had charged a large particle of fluff. The fluff couldn't decide where it wanted to be and would repeatedly fly around my bench only to return to the generator and then again off into the air. It was an astounding sight.

SUPER FRICTION GENERATOR – ROTOSTATIC

Now that you understand the basics of how a charge is picked up by friction, doesn't it seem that it would be possible to design a mechanical means to make and store a charge automatically and continuously? One solution to the problem is what I call a "rotostatic" generator, so named for an old magazine article that called a similar device a "rotostat." I said earlier that it was possible to get 3500 volts from an old plastic bottle, and I wasn't kidding. All you have to do is build the "rotostat".

These generators are extremely simple to build and may be hand cranked or operated by a small electric motor. They consist simply of a plastic drum – an old plastic jar works great – and a rubbing block – wool, nylon, asbestos, leather. When the bottle is rotated against the rubbing block a charge is developed on the plastic and on the block.

We need to attach some metallic brushes to the side of the plastic drum opposite the block to pick up the charges as they are generated. The brushes are attached to a metal sphere where the charges can be stored. A complete schematic of a rotostatic generator is shown in figure 3.

There is an infinite variety of possible configurations for such a generator, but keep in mind in designing yours that for static voltages even wood is a good conductor. It is most important to remember to keep your high potential components well insulated from ground.

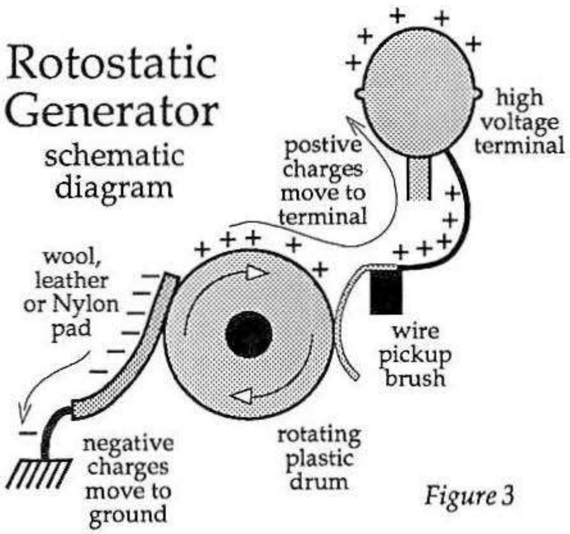
Rotostatic
Generator
schematic
diagram
postive
charges
move to
terminal
high
voltage
terminal
wool,
leather
or Nylon
pad
wire
pickup
brush
negative
charges
move to
ground
rotating
plastic
drum

Figure 3

If your generator is well made it will generate enough power to light a neon bulb (Ne2 type), levitate styrofoam, and arc a spark perhaps as long as 1/16 to 1/8 inch – all of this while still being perfectly safe to touch. When touched, your generator will give your finger only a very light "pop" much like a very small static discharge from a doorknob. As an interesting side note, if you can feel a static charge at all, it will be at least 3,500 volts!

Though static machines deliver these high voltages, their rapid, low current discharges make them unusually safe. Even generators producing 100,000 volts or more rarely present a danger, though I for one would not intentionally take a shock from one. A later chapter on capacitors (energy storage devices) will give you an equation for finding out when a charge has become dangerous.

Building a Film Case Rotostatic Generator

In my rotostatic generator design I have attempted to maximize the output and storage capacity of this type of machine. I am sure that there are still many improvements that can be made. This device works extremely well, is reliable, and very easy to construct. A schematic and a photograph are shown in figure 4.

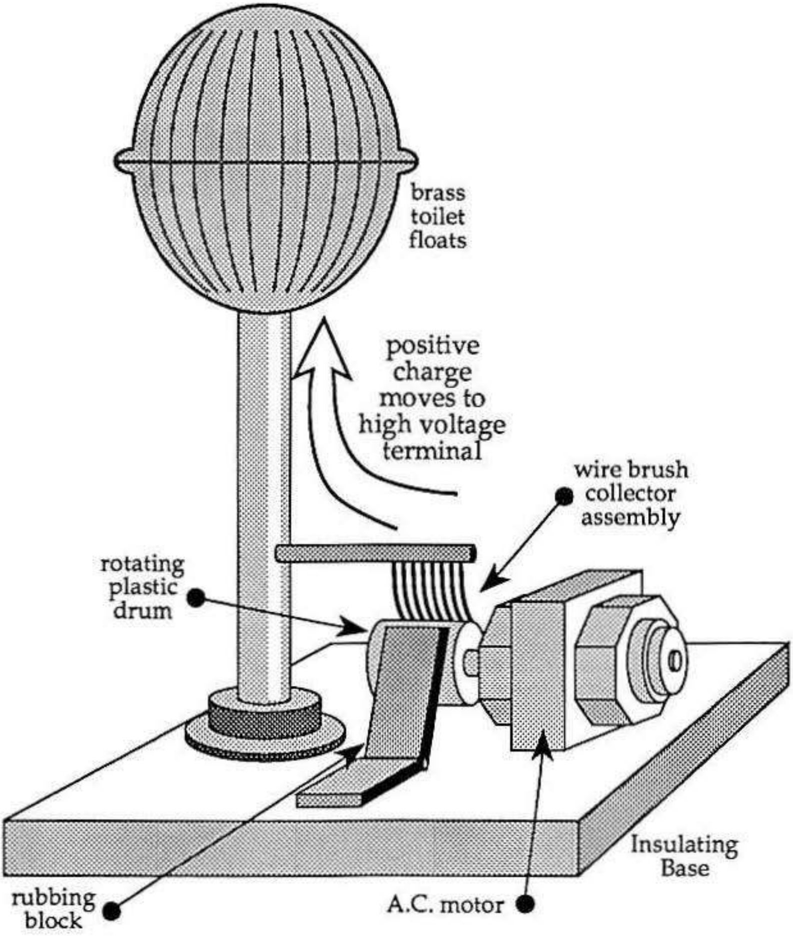
brass toilet floats
positive charge moves to high voltage terminal
wire brush collector assembly
rotating plastic drum
Insulating Base
rubbing block
A.C. motor

Rotostatic Generator

Figure 4

I used a surplus AC tape recorder motor to drive the unit. I like to use tape recorder motors in many projects because they are quiet and often have threaded axles that are very easy to mount components onto. In this case, I wanted to be able to switch plastic drums easily so I could experiment with a wide variety of materials. If you are unable to find such a motor, something similar can be used. A breakdown of

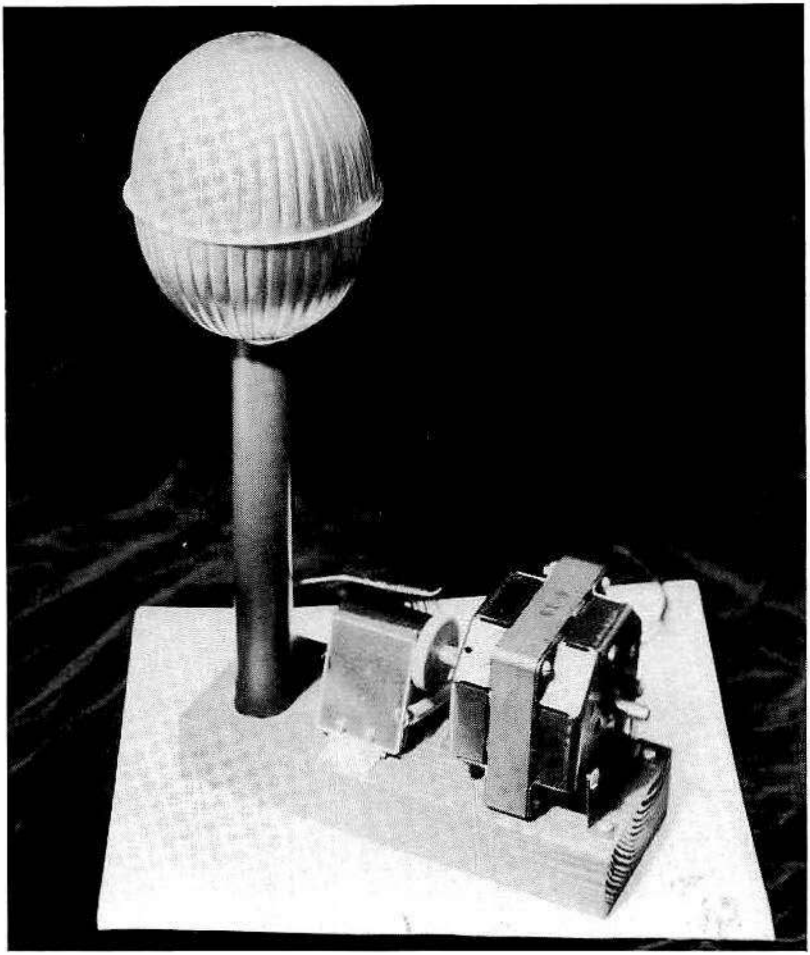

Fig. 5 – The complete Rotostatic Generator uses only a small plastic drum and a few common parts, and yet will light a hand-held 90 volt neon lamp!

AC and DC motors and their suggested hookups are shown in detail in the next chapter.

Assembly is straightforward and simple. The motor is mounted securely with wood screws to the base – in this case a section of 2x4. Make sure this connection is secure since in experimenting with different drums you will undoubtedly have some which will wobble due to imbalances in the drum or in the mounting. Also, remember to select a motor with plenty of power. Tiny D.C. toy motors can work, but they have little torque and can stall, and that can be frustrating.

The drum can be mounted to the motor in a variety of ways. If the motor axle is hollow as the tape recorder axles are, you can simply use a screw to secure the bottle cap directly to the motor axle. If the axle is solid, you will probably want to extend the axle all the way through the drum.

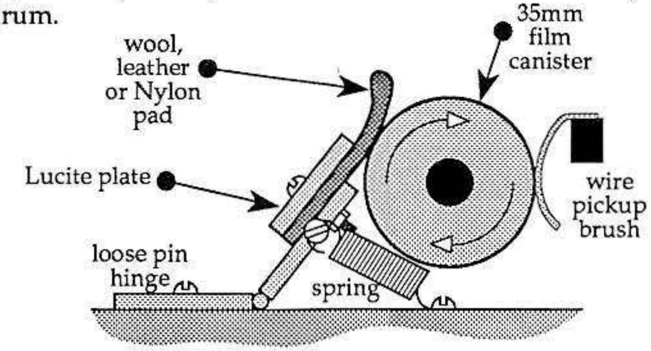

Wiper Assembly
for Roto Static Generator

Figure 6

To mate the motor axle to the axle running through your drum, glue progressively larger pieces of plastic tubing to the smaller axle until the tubing size matches that of the larger axle, then slip the tubing over both axles. Tubing used as fuel line on model airplanes, aquarium tubing, compressor hose, and the insulation stripped from wire can be used to make excellent adapters. Heat shrink tubing can also be used, however, after it has been shrunk it tends to become somewhat brittle.

As long as the pieces of tubing fit tightly over one another, any glue can be used to hold them together. I like cyanoacrylate (Super Glue) because it seems to hold rubber tubing nearly as well as it does skin. If you have used "Super Glue" for any length of time will know what I mean by this.

This method of connecting a motor to an axle also has the beneficial effect of acting as a "universal" joint. That is, your motor can actually power your axle from a variety of angles since the tubing will flex without difficulty. This is useful when using a variety of motors to drive a variety of unusual drums.

The drum must be very well centered on the motor axle, but fortunately, this is surprisingly easy to do. Most plastic pill bottles, film cases etc. already have a little nub right on the center of the lid. I assume this is some natural artifact of the process by which they are formed. Just hunt around a bit and use a bottle with this built-in center mark, and you'll have no difficulty here.

Personally, I like to mount the lid from the bottle to the axle so that the bottle can be easily removed and replaced.

The rubbing block shown in figure 6 turned out to be a very handy design. By looping a piece of wool against the Lucite as shown, the block makes excellent contact with the drum. A variety of rubbing materials can be tried by simply slipping them in between the block and the drum. I think that leather or nylon probably provided the best results when used with the film case. To make these experiments easier, use "loose pin" hinges for mounting the rubbing block.

Rather than use steel brushes to collect the charge from the drum, I came up with what I call "riders." These are formed by wrapping a short length of wire around the copper wire conductor coming from the globe. These wires should be loose enough to dangle down onto the drum and "ride" along its surface. Riders have many advantages. They are very tolerant of wobble in the drum. You may be able to change drum types and sizes without adjusting them in any way. You can improve the efficiency of "riders" by coating unusued surfaces with corona dope.

The PVC tube and collecting ball are easily mounted by drilling a tight hole in the 2x4 to accommodate the PVC tubing, by applying a little Silicone glue to the base, and by tapping it into place with a hammer. A large copper wire is run through the tubing as shown and is connected to the collecting ball with a brass screw. The ball, itself, is then mounted with a bead of Silicone glue.

Testing and Operation

Turn on the static generator's motor and allow it to run for 20 to 30 seconds. It takes some time to build up a charge. Next, hold one lead of a Ne2 neon bulb, available at electronics stores for about 49 cents, and bring the other lead near the brass bulb. When the free lead is nearly touching the bulb, the lamp should flicker or flash. Though the shock from this tiny generator can barely be felt, keep in mind that it takes nearly 90 volts to light the bulb!

Next spread some small, dry, pieces of styrofoam on the top of the ball. By bringing your hand close to the styrofoam you can actually cause them to move mysteriously toward your hands.

Touching this generator is perfectly safe. Under good conditions you should feel and hear a light static "pop" when you touch the fully charged top terminal. If you have trouble feeling the "pop" with your fingers try using your nose. I don't know why, but bringing your nose near the ball brings on a very subtle and strange electric sensation.

With this generator you can do many of the experiments mentioned later including charging up capacitors and energizing spark gaps. So experiment and enjoy.

REMEMBER!

If you are using yourself as ground, as described in the above experiments, you might find that generator performance seems to decrease after several experiments have been performed. What has actually happened is that by repeatedly touching the ball, you have charged yourself up to the same degree of charge as that of the ball, giving the electricity no desire to jump to you! This is simply remedied. Just touch a cold water pipe or other ground from time to time while you experiment.

One minor problem that can pop up is motor speed. Your motor may run too fast. We'll discuss a motor speed control a little later on. Don't worry if the riders hop around a bit. This is usually OK and won't cause problems.

Although a plastic film canister appears in the photos, jars with semi-transparent finishes perform. PVC might also make an exellent drum. Experiment!

A Useful List of Materials for Friction Generators

What follows is a useful list of materials that have a predisposition for taking on a negative or positive charge. By choosing materials that are far apart on this list you can increase the efficiency of a generator.

This list is excerpted from *Nature's Electricity* TAB books No. 2769 and is called "The TriboElectric Series."

POSITIVE CHARGE

Air
Human hands
Asbestos
Glass
Mica
Human hair
Nylon
Wool
Fur
Lead
Silk
Aluminum
Paper

NEUTRAL

Cotton
Steel
Wood
Amber
Sealing Wax
Hard rubber
Nickel, Copper
Brass, Silver
Gold, Platinum
Sulfur
Acetate
Polyester
Styrofoam
Orlon
Saran
Polyurethane
Polyethylene
Polypropylene
Vinyl (PVC)
Silicon
Teflon

NEGATIVE CHARGE

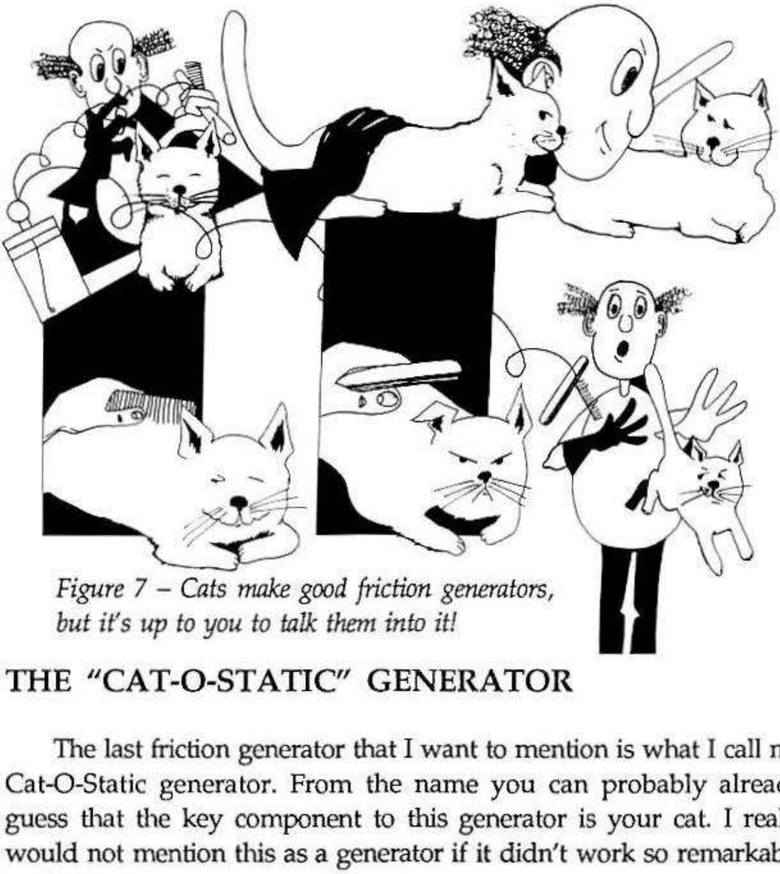

Figure 7 – Cats make good friction generators, but it's up to you to talk them into it!

THE "CAT-O-STATIC" GENERATOR

The last friction generator that I want to mention is what I call my Cat-O-Static generator. From the name you can probably already guess that the key component to this generator is your cat. I really would not mention this as a generator if it didn't work so remarkably well!

For the most part cats enjoy being electrostatic generators, the trick is to make them think they are being petted, and not used as a science experiment. (The cat commandments state somewhere: "Thou shalt allow thyself to be petted only when it suits your purposes and not those of another.")

First, you will need a rubbing material. The very best material I found for this purpose was the artificial leather used in cheap motorcycle gloves. It's actually a type of vinyl. You can secure a pair of these gloves for 5 or 10 dollars at any motorcycle shop. Some are called "man-made leather."

The next thing you will need is a metal flea comb and a length of PVC tubing. You will also need a Leyden jar to store the charge built up by your Cat-o-stat. Leyden jar construction is described in detail later on. The metal flea comb is held at the end of the PVC tubing by a slit cut in the tubing and some glue. A wire is attached from the flea comb to the top of the Leyden jar.

Put a glove on one hand and hold the PVC tubing with the other. Coax you cat into your lap, and tell him wonderful things about himself. This never fails to distract them. Pet him with the glove, and follow the glove with the metal comb to pick up the charge.

With my cat's fur and my particular gloves I am able to build up a charge so large that it will penetrate my vinyl gloves with a strong prickling! However, keep in mind that this type of generator will usually run off and hide before a large Leyden jar can be fully charged. Of course, if you have several cats, use them all to the charge the jar!

AC vs DC Motors and Controls

As we move on to the higher power generators, a little discussion is in order with regards to AC (alternating current) and DC (direct current) motors, and some good ways to control them.

In the generators I've built and discuss here, I have used both AC and DC motors. AC motors have advantages in that usually no external transformers or circuitry are required. They develop plenty of torque – so much so in fact, that you rarely have to worry about stalling the motor. Most AC motors have no brushes or other components to short or wear. They will generally run for years and years, often until the bearings themselves wear out! In general, I prefer using AC motors over DC motors whenever possible.

The major shortcoming of an AC motor is low rpm. The high performance Van-De-Graaff-type generators describe below work best at around 5,000 rpm. If the motor runs faster than this, electrons can literally be "blown" off of the carrying belt, and performance suffers. At lower rpm, leakage can impede the build up of high voltage. Since AC motors do not usually develop high rpm, DC motors should be considered for these high performance generators.

Unlike AC motors, DC motors are portable when battery powered and achieve high rpm easily. Keep in mind that a DC motor rated at

6500 rpm may not achieve that speed when driving your device. The rpm rating is usually the maximum no load speed of the motor.

Although not always necessary, you will probably find that you will want a motor with speed control as part of your DC static generator design. A number of dependable methods are shown in figures 8 through 12.

Figure 8 shows a piece of high resistance wire soldered or tied between two copper posts. One side of the wire is tied to a 12v (or similar) power supply. A wire coming from your motor is attached to an alligator clip which can be moved along the wire to increase or decrease the motor speed. This method works extremely well. However, be aware that the wire can become red hot when supplying current to the motor! Make sure to house it safely.

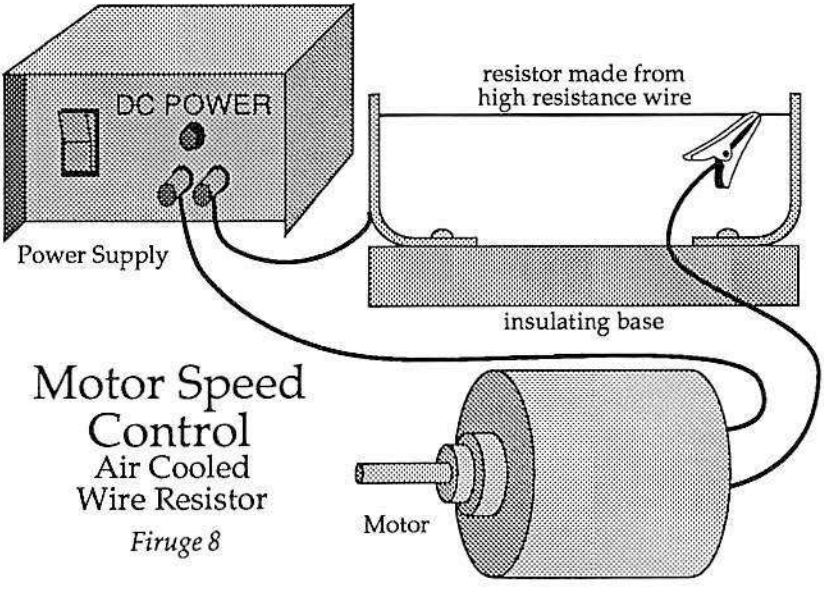

Motor Speed
Control
Air Cooled
Wire Resistor
Firuge 8

High resistance wire is easily obtained from a variety of sources. You can use the NiChrome wire that comprises the heating coils in an old toaster, hair dryer, or similar heating device. You can also purchase NiChrome wire from any well-stocked hardware store. Keep in mind that a fairly long piece may be required. Another variation that can be tried is to use a high wattage rheostat, or even the speed control handle from a "slot car" set.

Figure 9 shows the use of power resistors to slow a DC motor.

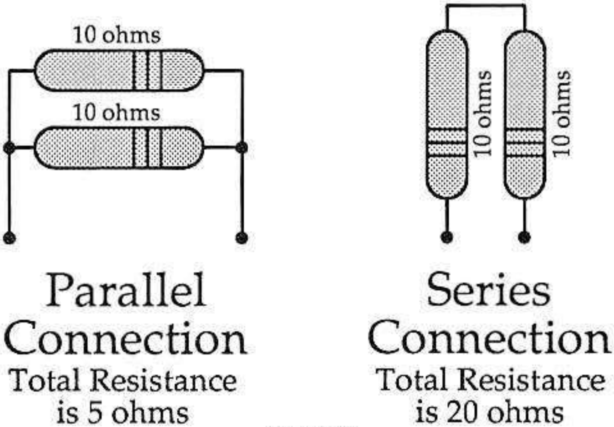

Figure 9

Power resistors are inexpensive, and if you buy a handful of different values you can connect them in parallel or series to get a wide variety of different values. Keep in mind that resistors have two ratings, the power handling capacity measured in Watts, and the resistance measured in Ohms.

An Ohm is a measure of the opposition to current flow. A resistor of 10 Ohms will oppose current flowing through it twice as much as a resistor of five Ohms.

Wattage is simply the amount of power your resistor can handle without being destroyed by overheating. In general, I'd say the higher the wattage rating the better. The Winter's Ring generator shown next required a 25 watt, 10 Ohm resistor, but had to be water cooled to increase its power handling capacity above the 25 watt air cooled rating.

Resistors oppose current flow and in doing so, convert electrical energy into heat. It's normal for them to become very warm. If your resistor gets too hot, its resistance can increase well above its stated value. A change in resistance can cause your motor to run at uneven speeds. You will either need to buy a resistor of a larger wattage, or use some means to cool it such as water (shown and discussed later). As a rule-of-thumb, don't use a resistor of less than 10 watts unless you have a very small or efficient motor.

Figure 10

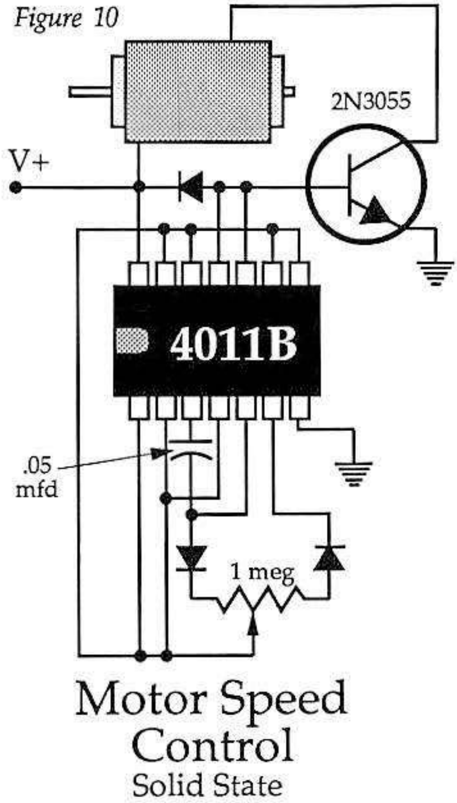

Figure 10 shows a very accurate and stable means of controlling a DC motor. An integrated circuit and power transistor are used to send a stream of pulses to the motor. As the pulse rate increases, the motor speed also increases. By simply turning the potentiometer, you can adjust motor speed at any time. For mid-sized DC motors a very large heat sink or water cooling is required for the transistor.

I have found that DC motors operating at high speeds can behave strangely. Slight variances in the motor's speed can set up an oscillation within the belt being driven. Because one side of the belt is being "pulled" and the other

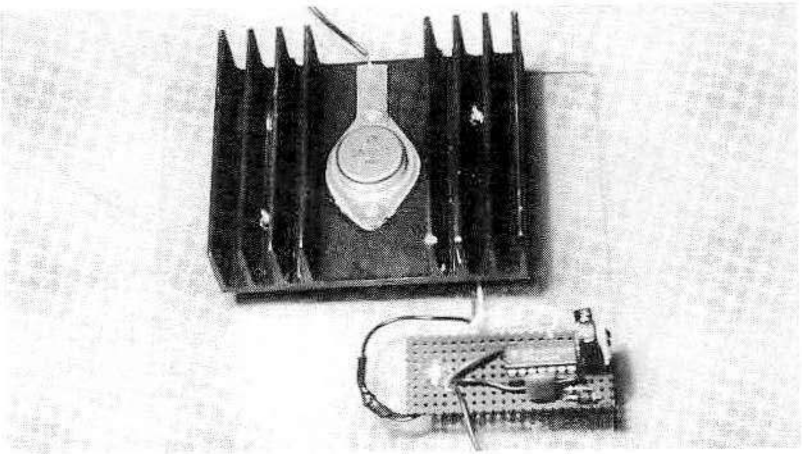

Figure 11 – Solid State Motor Control

is being "pushed", uneven tensions react with the belt's flexibility to produce a "flying" belt that makes only intermittent contact with the rollers or pulleys. This can cause a belt to fly off, bend brushes, or just make a lot of noise. Also, if ball bearings are not used in the rollers, an oscillation can occur among the bushings of the upper roller, the axle, and the belt. All of these problems can be easily solved by using a versatile speed control.

Component Cooling

You may to have to improvise methods of keeping electrical components cool. If you are using a fairly large, high RPM, DC motor which will need sizeable amounts of electrical power, components in the speed control are going to get hot. It may seem a little unorthodox, but water makes a fantastic heat sink for electrical components! Power resistors and even power transistors can be run well beyond ordinary heat sink tolerances when submerged in water. (Figure 12)

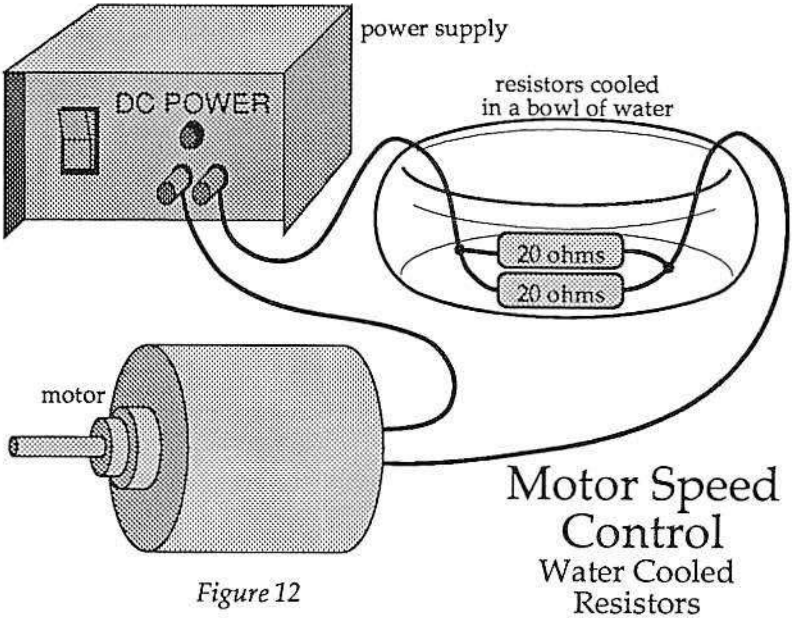
power supply
DC POWER
resistors cooled
in a bowl of water
20 ohms
20 ohms
motor
Motor Speed
Control
Water Cooled
Resistors

Figure 12

Distilled water is a very poor conductor of electricity and will not short out most components. Tap water is usually OK for resistors, but minerals and other gunk will cloud the water after a short time.

As long as the circuit is submerged in water, you know that the temperature of the water must be below 212 degrees Fahrenheit (100 Celsius) at normal atmospheric pressures. Above this temperature the water will boil away. Many electronic components work quite well up to 212 degrees. Just remember to add a little water now and then to compensate for evaporation.

AN EXTERNAL VAN DE GRAAFF USING "WINTER'S RING"

If you built the last generator described, you know that I wasn't exaggerating when I said you could get 3500 volts (or more) from an old film canister. Although interesting, this simple rotostatic generator won't keep you satisfied for long. There are higher voltages to conquer! So let's move on to the next type of generator and try for 10,000 volts (and possibly several times that amount) from a rubber band.

I call the following generator an "external" generator because all of the components of the generator have been mounted externally in such a way as to make construction and experimental modifications fast and easy. Since all of the workings are exposed, this generator is great for the experimenter who wants to quickly try different materials and configurations in developing his own generator design, or for someone who wants to whip together a simple generator to demonstrate.

The high voltage terminal of this generator consists of a metal ring, or toroid, of exceptional efficiency. I call it a "Winter's Ring" although such rings have been traditionally made of wood. They're rarely used anymore. As with other parts of this generator, the Winter's Ring may be quickly replaced by a more traditional spherical-shaped terminal or many others you might wish to try. Why not keep several on hand for demonstration?

Despite the fact that, as we've discussed, electrostatic principles have been well know for thousands of years, it was not until as recently as 1928 that a truly practical machine for generating high static voltages was perfected. Robert J. Van de Graaff, a young Rhodes scholar working at Princeton, designed his first machine to generate the high potential needed to power a particle accelerator to bombard atomic nuclei as part of his experiments in nuclear research.

Although it can be difficult, it is not impossible for the home experimenter to make his own atom smasher! A Van de Graaff generator such as described here is an integral part of such an apparatus. If you are interested in taking on a project of epic proportions, find a copy *"The Amateur Scientist"* by C. L. Stong which is now out of print and very difficult to find, but worth the effort. Or write for similar information from Information Unlimited, P. O. Box 716, Amherst NH 03031.

The Van de Graaff is so simple and obvious in design that it amazes me that is wasn't developed sooner. I have always thought that this is a good example of just how inventions don't have to be extremely complex to have a significant impact on society. How many other simple devices are there waiting to be designed and built? It could be you that dramatically changes the world!

The easiest way to understand the Van de Graaff is to refer to the schematic diagram in figure 13. A Van de Graaff generator consists of a belt, two pulleys, a motor, brushes, a high and a low potential terminals.

In newer machines the bottom pulley is plastic while older machines used an insulator covered by wool or similar rubbing material. The top pulley must be of a conductive material such as metal or even wood which is not a good insulator at high voltages.

When the machine is in operation, a small motor attached to the bottom pulley drives the belt that is lightly stretched between the pulleys. Friction occurs between the surfaces of the belt and the plastic wheel, putting an electrostatic charge on the belt. The charge is carried by the belt toward the upper terminal. A grounded wire brush makes contact with the lower portion of the belt and bleeds off some of the negative charge. The brush at the top of the belt accepts the positive charge remaining and distributes it on the high potential terminal. (Note: It is possible to make a Van de Graaff generator that develops a negative charge on the high potential terminal.)

As the belt moves, charge builds up on the high potential terminal. Theoretically, this charge could increase forever. However, leakage into the air from the upper terminal limits the maximum potential to

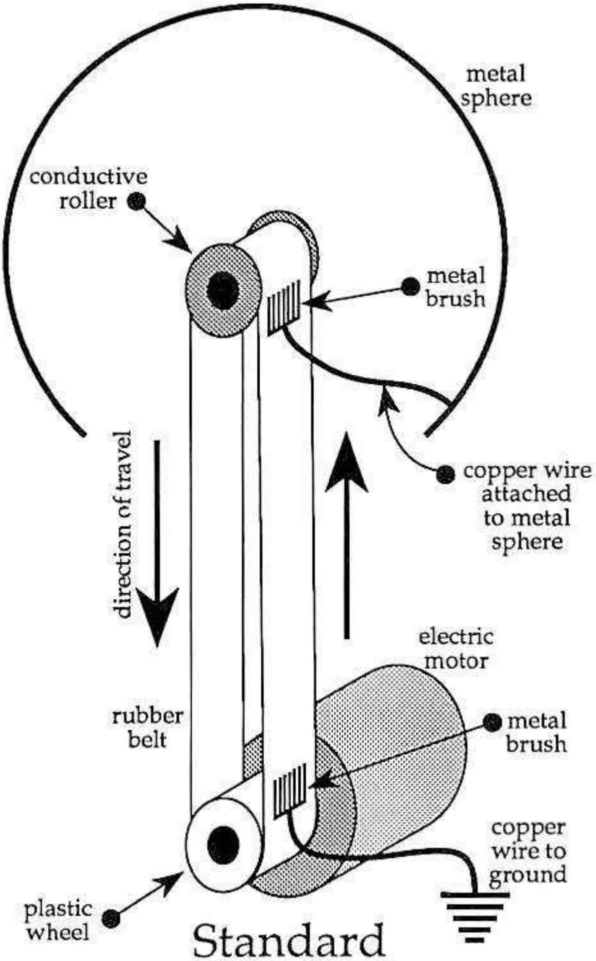
metal sphere
conductive roller
metal brush
direction of travel
copper wire attached to metal sphere
electric motor
rubber belt
metal brush
copper wire to ground
plastic wheel

Standard
Van de Graaff Generator
Schematic Diagram
Figure 13

about 50 to 100 thousand volts typically.

A simple way to calculate the theoretical maximum for a "perfect" spherical top potential is to figure a voltage 70,000 times the smallest radius of the curvature of the high potential terminal. For example–most spheres are sold with reference to their diameter. For a 12 inch diameter sphere, we divide 12 by 2 to get the radius, or 6 inches. We multiply 6 by 70,000 to get a theoretical voltage of 420,000 volts.

Because your terminal will never be a perfect sphere, and because there will be leakage into the atmosphere, theoretical voltages won't be reached. You might find these ball-park calculations interesting, nonetheless.

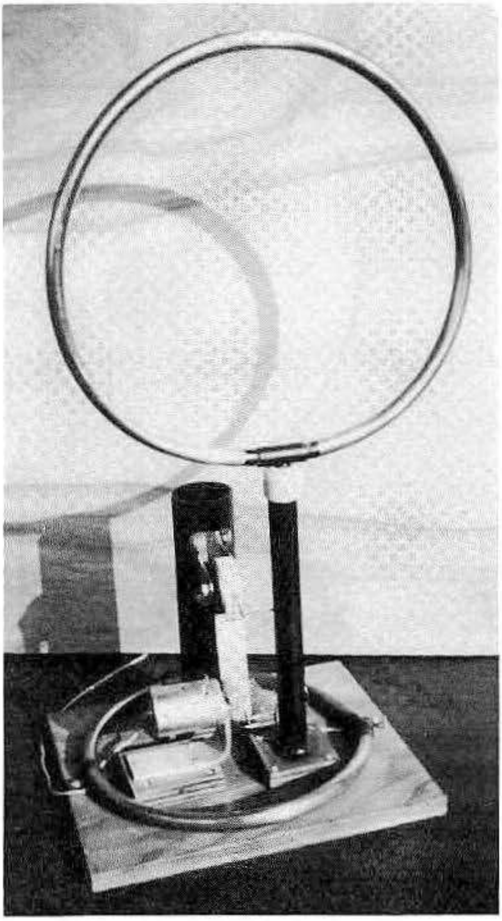
Figure 14 – External Van de Graaff Generator

A photograph of my external Van De Graaff and its schematic are shown in figures 14 and 15. I found this to be an extremely simple and inexpensive design. In operation it will produce a reliable voltage in the range of 10 to 30 thousand volts using an ordinary rubber band. It will produce a solid and exhilarating "pop" when it discharges to your hand. It can throw sparks to ground more than an inch in length. Neon bulbs such as a Ne2 will glow brightly and continuously when near the generator. A blue corona glow can be readily observed from the collecting brushes in a darkened room once your eyes have adjusted to the dark. Negative and positive corona can also be identified by their appearance using this device as discussed later.

The external Van De Graaff shown has some minor performance limitations due to the relatively high friction of the top wheel (no ball bearings) and excessive flexing of the rubber band. A rubber band should be used as a belt only for demonstration purposes. For more

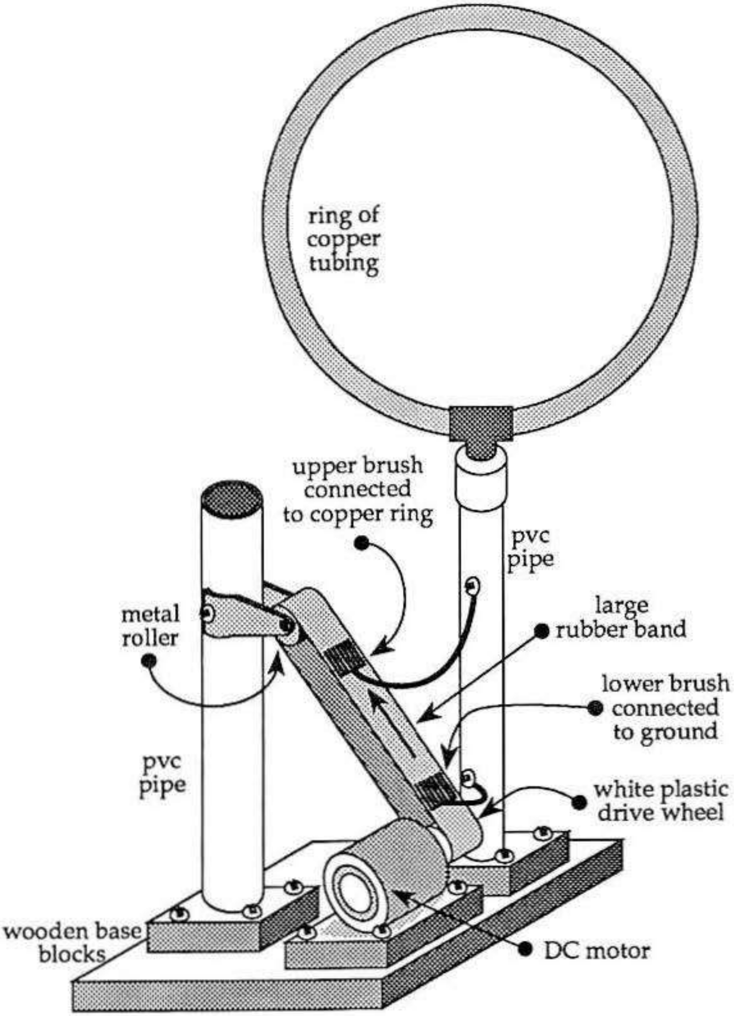
ring of
copper
tubing
upper brush
connected
to copper ring
pvc
pipe
metal
roller
large
rubber band
lower brush
connected
to ground
pvc
pipe
white plastic
drive wheel
wooden base
blocks
DC motor

External
Van de Graaff Generator

Figure 15

serious experimenting, replace the belt with a less flexible belt such as made from a bicycle inner tube or, ideally, neoprene. Also, without ball bearings, the top wheel required occasional lubrication that frequently brought about the risk of oil contamination. Except for these minor problems, this generator has proven to be an excellent and reliable performer. It still remains the most easily modified and observed of the Van de Graaff designs I'll describe.

The construction of the external Van De Graaff is so simple and straightforward that little description is required outside of the drawing. A few tricks can simplify and improve construction.

An easy way to get your motor, roller and belt in alignment is to run your generator while holding the motor with your hand, experimenting until you find the right position for it. If your belt slips off, shift the motor slightly to change its attack or grip on the belt.

When connecting a belt between two rollers in any Van De Graaff type generator, it is natural to assume that pulleys should have flanges on their edges to provide a groove in which the belt will run preventing it from slipping off the driving wheel. Surprisingly enough, the belt will have a tendency to stay centered on the pulleys even when they are somewhat out of alignment! I assume this has something to do with the way the pressures change in the belt as it approaches the edge of a pulley. You may find that a small "hump" in the drive wheel can even be helpful at high rpm. Once aligned, you should have no problems at all with belts slipping off.

Use the largest diameter of copper tubing that you conveniently can for the top terminal. The problem is that when larger diameter copper tubing is bent into small diameter circles, undesireable ripples form. Larger gauge tubing must be bent into larger diameter circles forcing you to compromise between size and efficiency.

When soldering the two ends of the copper tube into the "T" connector, first clean them with extra fine steel wool, then coat (tin) them with a little solder before inserting them. This will give an exceptionally strong and smooth bond. You can use any ordinary propane torch for heating and soldering the tube.

If you use the same type of brushes as shown in the drawing, it might be tempting to shorten the copper wires leading to them from the PVC tubing or otherwise alter this mounting. Don't do it! It is the flexibility in these long copper wires that provides the "spring" needed to keep the brushes flat against the belt.

The plastic bottom pulley is simply one of the white plastic wheels used on cheap furniture casters. It is attached to the motor shaft by carefully drilling out the wheel's center hole to a size slightly smaller than the diameter of the motor axle. You then apply a few drops of super glue to the axle and press or tap the wheel into place.

If the hole in the wheel is larger than the diameter of the axle, glue progressively larger diameter pieces of rubber tubing to the motor axle until the wheel fits snugly in place.

One last note: Since your top terminal is copper which quickly tarnishes, you might want to apply a little Silicone cleaner/protectant such as that sold for cleaning chrome automotive parts. This will keep your top terminal bright and shiny at all times. It also seems to limit leakage and produce an interesting double "pop" when touched.

The Evils of Oil

The only major problem that I encountered in building and operating Van De Graaff generators regardless of design is oil contamination. Since a Van De Graaff generator relies on friction between the plastic wheel and rubber belt to produce a charge, you can imagine the devastating effect oil or any lubricant can have on belt performance. Even the oil from your fingers can reduce the charge carried by the belt. Handle your belts sparingly, and hold them only from the outside edges as you might a record album.

The worst lubricant is a petroleum-based oil such as used to lubricate motor axles and bearings. Angular momentum from the motor axle can actually pull a drop of oil all the way along the axle, along the side of a wheel, and out onto a belt. *Therefore,* ***Do not oil your motor or bearings excessively.***

If you want a graphic demonstration of this effect, place a drop of fast drying super glue on your motor's axle. Before it has a chance to dry, run your motor for several seconds. If your motor is like mine, you will have a little trail of super glue from the axle to the outside of the wheel where it will form a little blob.

Once your belt and wheels are contaminated, they can be very difficult to clean. The best way to clean the belt is not to clean it all, but to throw it in the trash and get another belt.

The conductive metal top roller can be cleaned with a residue-free cleaner like ammonia, followed by a light sanding with extra fine sand paper and then by another ammonia cleaning.

Cleaning the plastic wheel is the most difficult. The plastic wheel usually builds up a static charge of its own during the time the generator is run. This charge seems to act like a sponge for oil! The only way I have found to remove it is it by attaching a wire to ground and running the motor while dragging the wire along the outside of the wheel. After discharging the pulley in this way, ground a piece of metallic sand paper and again run the motor while sanding away some of the outside of the wheel. This is accompanied with or followed by ammonia cleanings. Other solvents may work, but watch out for those that may dissolve the plastic. I realize this procedure may sound extreme, but I have even tried soaking the plastic wheels in industrial degreasers not available to the general public and have had no luck at removing this oil in any other fashion. If you are careful in your use of oil you will never encounter this problem.

A Few Words On Brushes

In both of the Van De Graaff generator designs described in this book, I used a simple tuft of wire soldered to the end of a thicker solid copper wire spring which provides the force necessary to keep the brush riding firmly against the belt. By far the biggest problem you are likely to encounter will be in keeping the brush riding solidly on the belt. If your brushes are not well placed, or your belt has undue flexibility, the brushes may have to be readjusted frequently.

Many configurations, some very elaborate, have been tried in an attempt to perfect a fool-proof brush assembly. I haven't found one yet. You may want to try a hinge, a spring, a piece of metal screen, or even a conductive rubber or foam such as those use to package integrated circuits. In designing a brush assembly, keep in mind that, ideally, it should be comprised of many fine points like the ends of wires because the density of electrons in a fine point is higher and electron transfer is more likely. Nevertheless I have had very good luck with non-pointed objects such as conductive foams. Most texts, and my own experience has shown that a good old tuft of wire usually wins out, but, as mentioned, this is not a perfect solution to the problem. So, experiment. And have fun!

A Few Words On Belts

A good belt can be made from virtually any insulating material. Rayon, Dacron, rubber, paper, silk, plastic and even ordinary cloth have been used with success. Wear is probably the most important factor to consider in choosing a material. Rubber, for instance, rots in my home town due to the ozone contained in the smog here. Cloth will fray at high speed, although I have coated cloth with an extremely thin coating of Silicone rubber and have made good belts. Neoprene, joined with a diagonal splice is a long-time favorite.

In my opinion, the easiest all-around belt is made from an old bicycle inner tube, or even better, the rubber "liner" that covers the heads of spokes under the rim. One of these belts can be carefully spliced with inner tube cement from a tire patch kit. Even a little "Super Glue" will work if the splice is clamped.

In fabricating a belt also remember that the wider the belt, the faster a charge accumulates (within reason). And, you may also want to give a little thought to the noise generated by certain materials. At 5000 rpm a poorly made belt will make considerable noise.

One last belt option is to purchase a ready made belt from a scientific supply house such as Analytical Scientific, Post Box 675, Helotes, TX 78023. Such companies can often supply many different sizes and types of belts. I would suggest ordering the thickest, and strongest belt offered for a home-made generator. I find the price of commercial belts a bit too expensive for my bank account, but you might find their prices more acceptable depending on your project.

THE HIGH POWER "CLASSIC" VAN DE GRAAFF

Although these voltages are impressive, keep in mind that the classic Van De Graaff will easily produce 60 to 100 thousand volts or more and is only slightly more difficult to build.

The classic design for the Van De Graaff is shown in photograph 16. The one minor concession I made in my design was to use two stainless steel bowls for the high potential collector instead of the traditional sphere. I did this simply because I could not readily find a

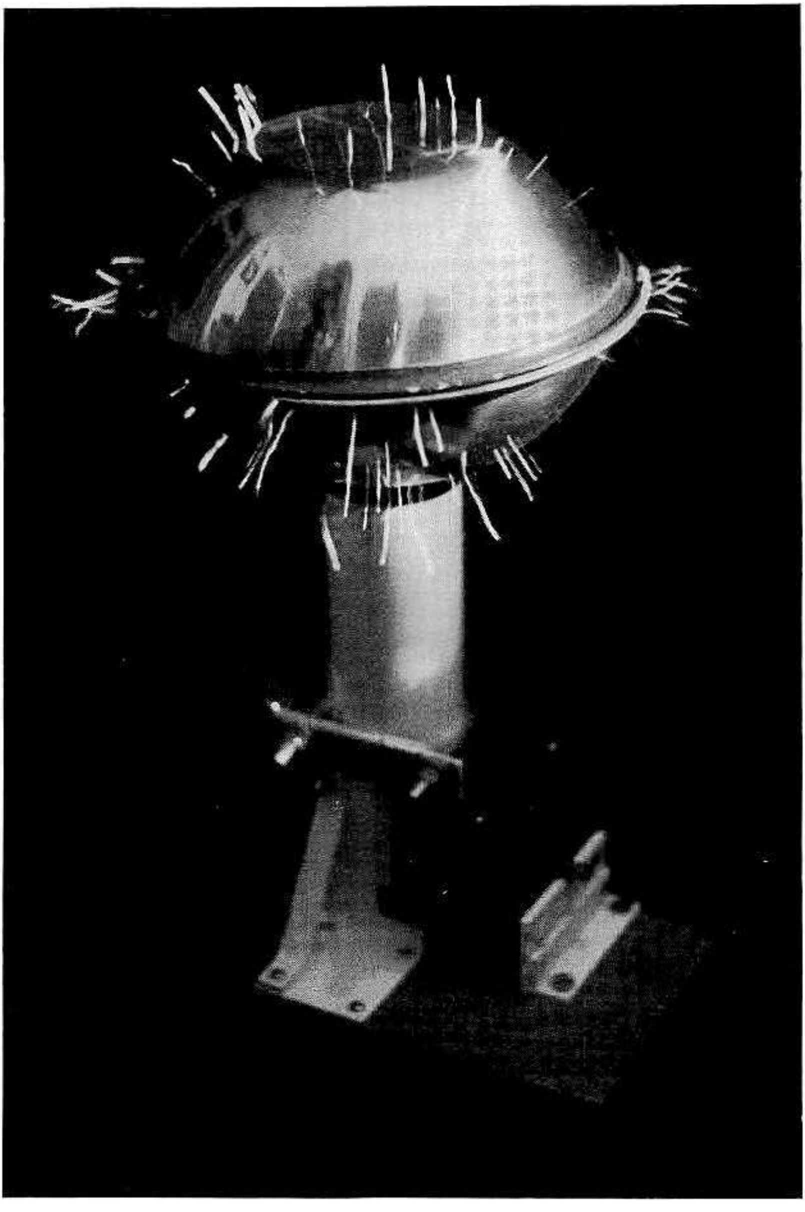

Figure 16 – Classic Van de Graaff Generator in Operation – This long-exposure photograph clearly shows the remarkable energy generated by this simple device.

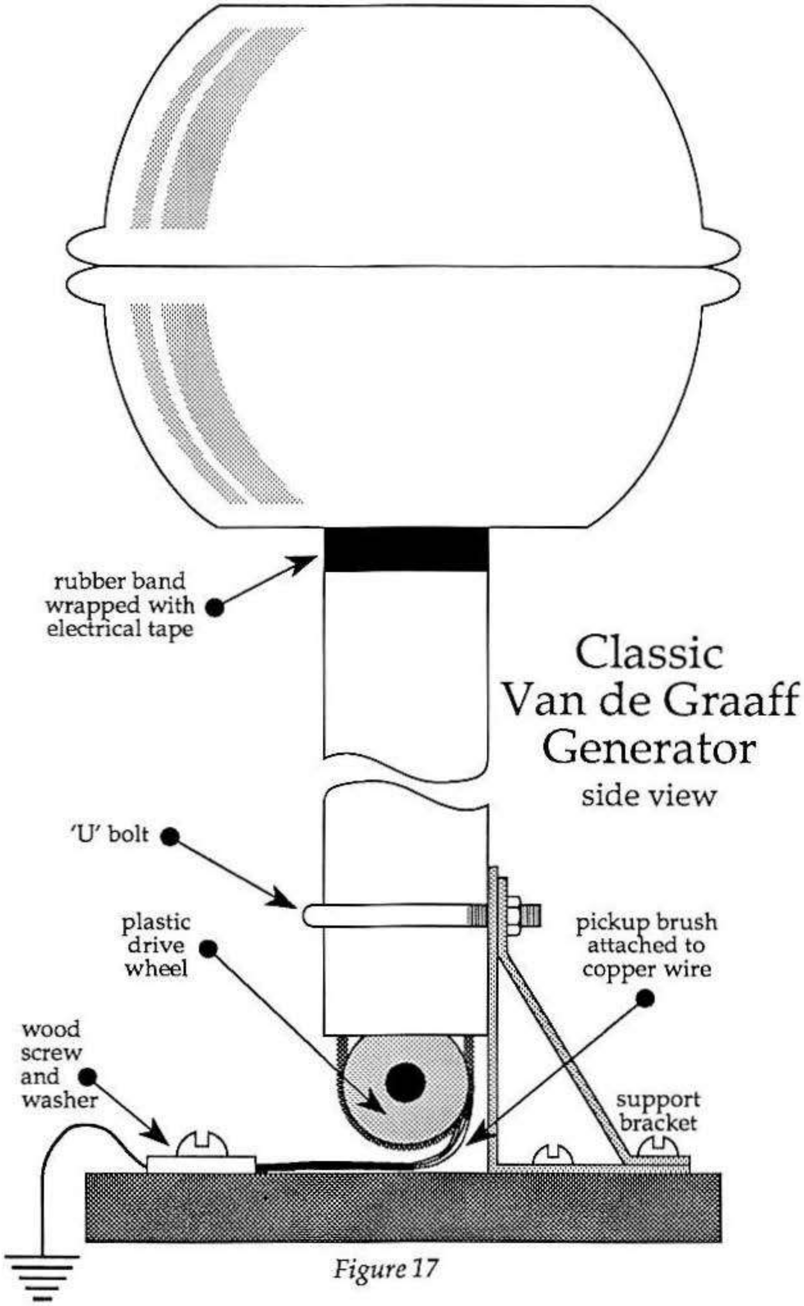
Classic Van de Graaff Generator side view

rubber band wrapped with electrical tape
'U' bolt
plastic drive wheel
pickup brush attached to copper wire
wood screw and washer
support bracket

Figure 17

sphere of the proper dimensions. Besides, two bowls allow easy access to the top roller and brushes, and I wanted to test my hunch that almost as much power could be obtained from the bowls if the edges were heavily doped and insulated as could be obtained from a traditional spherical terminal.

All of the experiments to be described can be performed with this unit. It is an all-around solid machine and is fun to operate. If you plan to build only one generator from this book, and are wondering which generator to select, this machine is the best all-around performer. If carefully built, this design will exceed a voltage of 100,000 volts. It will arc thick, blue sparks well over four inches to an outstretched hand and deliver an invigorating, if not slightly uncomfortable, shock.

This generator is capable of charging up a person's body to a high voltage. It makes possible many experiments such as causing a person's hair to stand on end, or the rapid charging of large capacitors. The device will even produce an ominous "crackling" as your hand approaches the high potential terminal.

By standing on a large plastic insulator such as an overturned detergent bucket, a friend of mine was able to charge her body to such a high potential that her hands would "crackle" any time she pointed to a person nearby in the room. I think a physics or chemistry teacher would have little trouble keeping his/her students attention if the mere pointing of their hands produced this ominous cracking!

Though I am certainly not recommending the following approach, this generator was also inadvertently useful in training a dog in a single lesson not to jump on people. My dog, Merfi, had the bad habit of greeting people by jumping all over them. Nothing, it seemed, could discourage this behavior. One afternoon some friends and I were standing around a completed Van De Graaff and commenting on what an unusually large charge we had managed to bestow upon a person. Suddenly Merfi bounded into the shop. Despite our pleas, Merfi shot across the shop with his usual enthusiasm and leapt nose first into my highly charged friend, no doubt, intending to bestow upon him his traditional greeting.

We could do nothing but watch as a large blue spark jumped to Merfi's moist, waiting nose. Merfi stumbled back, temporarily lost control of his bladder, and then slunk out the door. A few moments later everyone was petting and reassuring Merfi, yet, to this day we have had virtually no problems with Merfi's traditional greetings. He still eyes the friend from whom he received the shock with suspicion. And with good reason!

I suppose it is needless to say that it is wise to keep any and all pets a safe distance from your experiments. As with the external Van De Graaff generator, very little description outside of the drawings is required for constructing the classic Van De Graaff design.

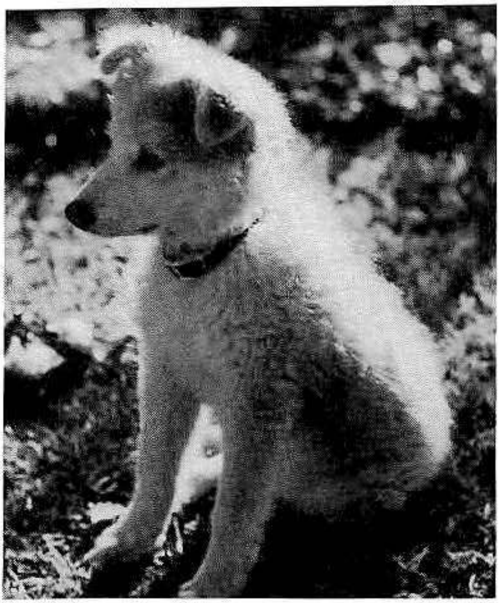

Figure 18 - "Merfi" – (An accidental electrostatic participant)

Things to keep in mind

The edges of the bowls are relatively sharp, even though the edges have been folded over. To prevent the charge from leaking through these sharp edges, they are coated with a special paint called "corona dope" which can be purchased in most T.V. and well-stocked electronic stores. It is available in both spray and brush-on forms. Corona dope has a high dielectric strength, and will prevent a charge from leaking off into the air. It can be used on any spots suspected of leakage on any of your high voltage projects. Keep in mind that the thicker the coating of corona dope, the better it will shield. I used around eight thick coats on my Van De Graaff and should have used more. Sparks fly right through it!

As with the last generator, the lower pulley is a white plastic wheel taken from an inexpensive furniture caster. It was mounted in the same way as was done in the external Van de Graaff.

The upper pulley is the front hub of a 10 speed bicycle wheel. I tried several different methods to reduce the friction of the hub bearings to make it turn more freely. I tried a variety of lighter oils and even

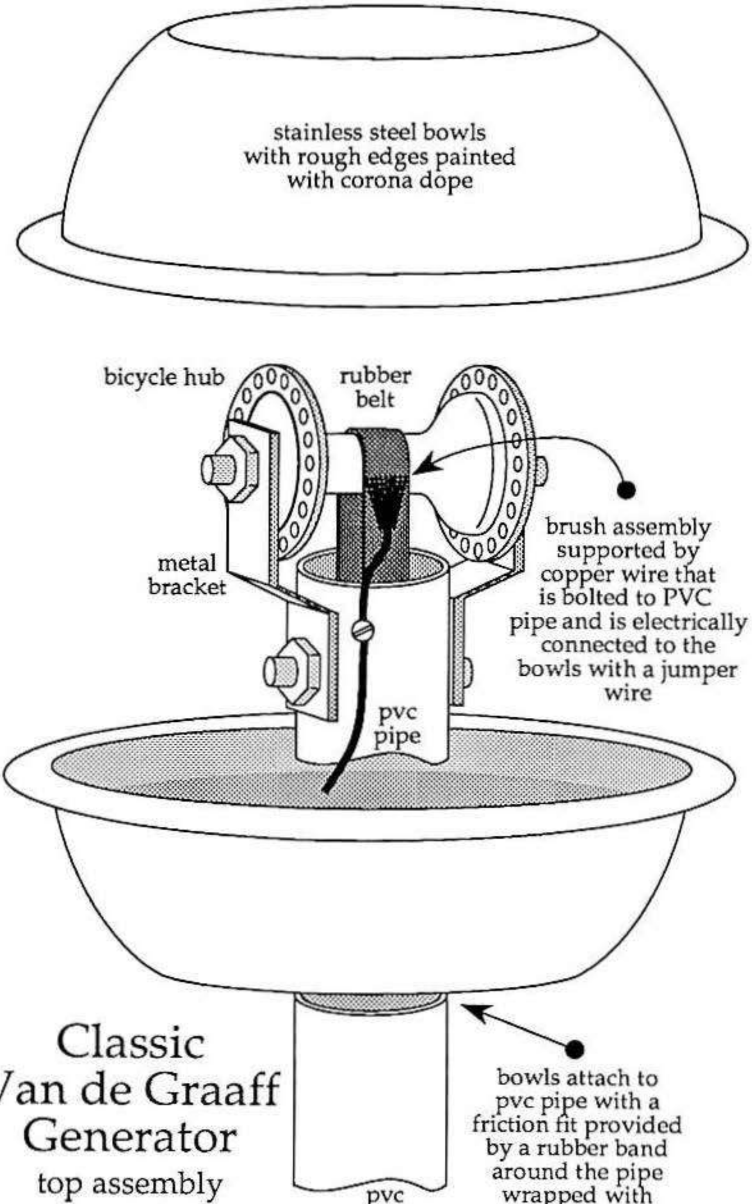
stainless steel bowls
with rough edges painted
with corona dope
bicycle hub
rubber belt
metal bracket
brush assembly supported by copper wire that is bolted to PVC pipe and is electrically connected to the bowls with a jumper wire
pvc pipe
bowls attach to pvc pipe with a friction fit provided by a rubber band around the pipe wrapped with electrical tape
pvc pipe

Classic
Van de Graaff
Generator
top assembly

Figure 19

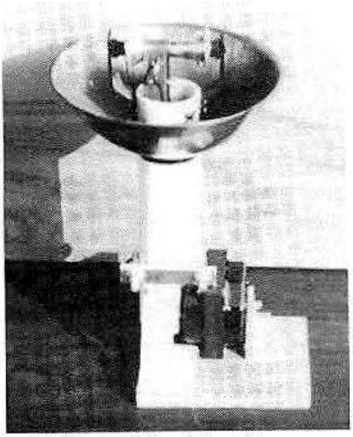

Figure 20 – Top Open

removed the rubber covers that protect the ball bearings from dirt contamination. I found that lighter oil works well, but because of the risk of oil contamination as discussed earlier, I ended up using grease. Removing the covers also helps, but when the generator is running and the roller is spinning, a sizable charge builds up on its surface, and every dust particle in the area is attracted to it! Although you may want to experiment, I have come to the conclusion that the hub should be left pretty much as it is when it comes off the bike.

The two top bowls can be hinged together from the inside with a wide strip of black tape, or can be bolted together by some nylon screws through the lip. Keep in mind that you want the surface of these bowls to remain as smooth as possible to prevent electrons from leaking off. Don't use metal hinges or any sharp objects in attaching them.

The main body tube is fabricated from three-inch diameter PVC drain pipe. (Remember that acrylic or phenolic pipe might be superior to

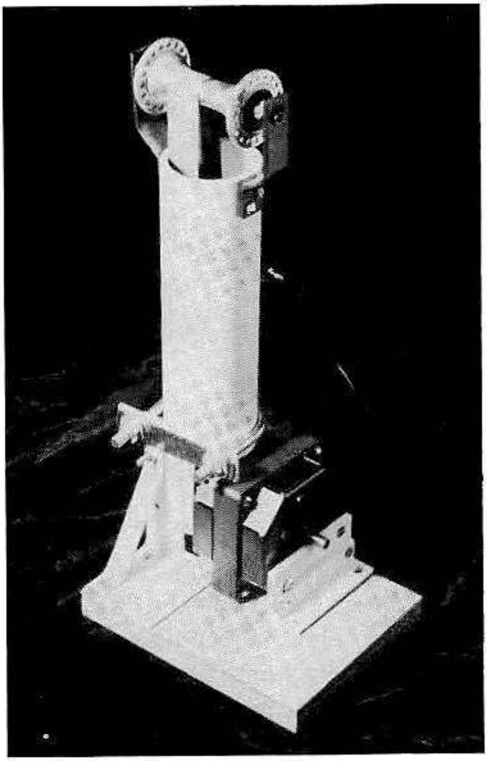

Figure 21 – Terminal Removed

Figure 22

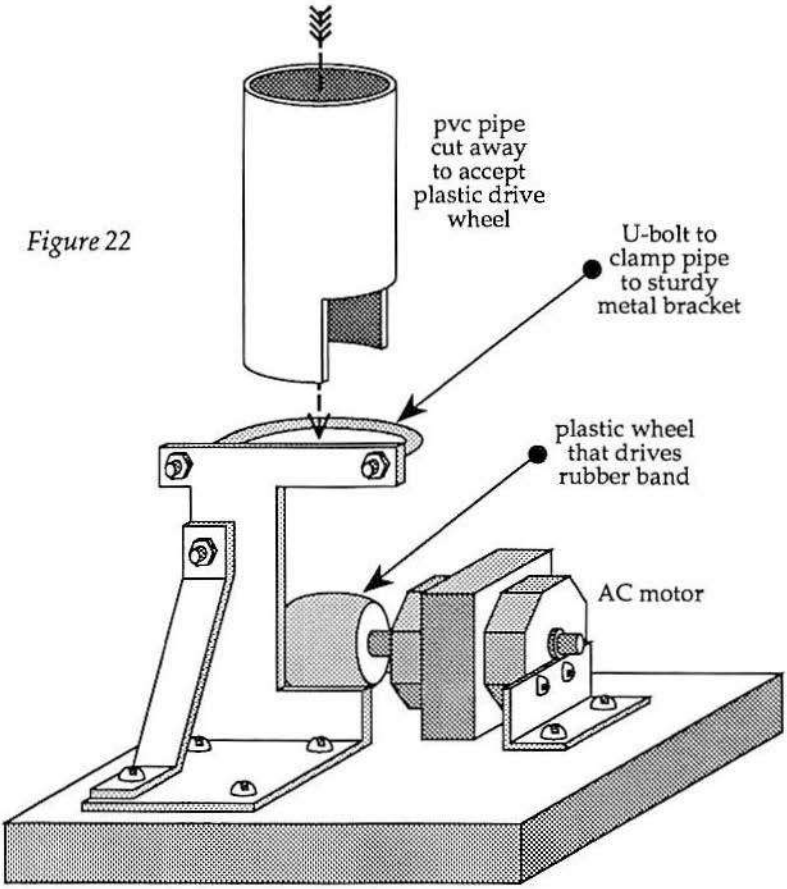
pvc pipe cut away to accept plastic drive wheel
U-bolt to clamp pipe to sturdy metal bracket
plastic wheel that drives rubber band
AC motor

Classic Van de Graaff Generator
bottom assembly

PVC.) A slot is cut into the lower portion of the tube to allow the motor and plastic drive wheel to be inserted through wall. The PVC pipe is supported by a metal L-bracket and a U-bolt. Depending on the size of your motor, you may want to cut a notch into one side of the L-bracket to provide enough clearance for the motor to fully enter the bottom of the tube. Since this "L" bracket will support your whole generator, make sure it is sturdy and of a heavy gauge. You may even wish to add a metal brace as I did to increase rigidity.

When you have completed this generator, you will be able to perform any of the experiments that follow. This easy-to-build, high performance generator is a very exciting machine to own!

AN EXTREMELY HIGH POWER VAN DE GRAAFF

Before we move on to other types of static generators, I'd like to briefly describe one of the highest performance Van De Graaff generators I have ever seen. Shown in figure 24 is a design for an extremely high power Van De Graaff generator. I have not constructed this unit. I am providing details on this machine should you want to explore machines beyond the limits of my humble designs. This schematic will provide the essential design details.

At maximum efficiency this design could generate voltages exceeding 1,000,000 volts without using an auxiliary storage capacitor!

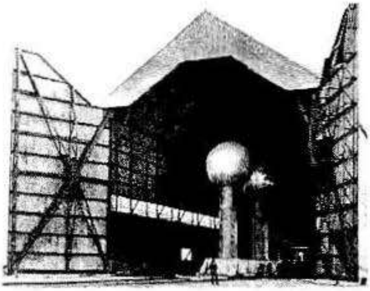

Figure 23 – Ten million volt Van de Graaff generators were built in the 1930's. One generator here is negatively charged, and the other is positive. Generators of this size are not recommended as a first project. (Perhaps not even as a second! ...but call me if you build one!)

Voltages of this order can be dangerous, possibly lethal. This is not a toy and should only be constructed by a serious and cautious experimenter. Because of its high output, I think this unit would make an outstanding laboratory power source for a homemade atom smasher.

The high power unit uses an external high voltage power supply to "spray" an initial charge on the rubber belt. Notice in the schematic that the ends of the spray wires are pointed. A rectified high voltage source between 5,000 to 10,000 volts is ideal. For low cost and ease of assembly you may want to try using an automotive coil as a transformer. For details see Lindsay Publications *"Build a 40,000 Volt Induction Coil."* (Written by an author who is generally a good sort.)

In this design, both top and bottom rollers are metal. The top roller itself is insulated from the high potential sphere. Charge is sprayed

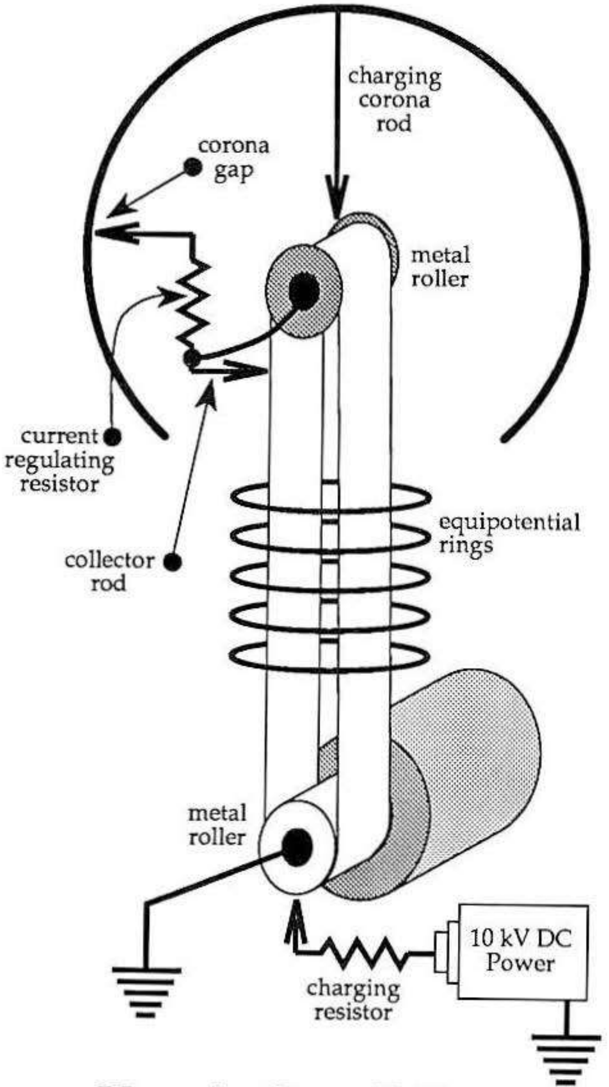
charging
corona
rod
corona
gap
metal
roller
current
regulating
resistor
collector
rod
equipotential
rings
metal
roller
10 kV DC
Power
charging
resistor

Van de Graaff Generator
high power configuration
using external excitation
(schematic diagram)

Figure 24

Figure 25

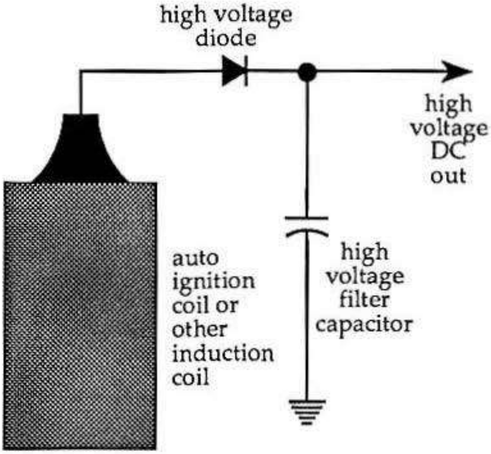
high voltage diode
high voltage DC out
auto ignition coil or other induction coil
high voltage filter capacitor

Induction Coil Exciter for Van de Graaf Generator
output must be recitified and filtered

onto the carrying belt as it passes through the corona discharge between the high voltage power supply and the lower roller which is grounded. A second set of points located just inside the upper high potential terminal picks up the charge and conducts it to the upper pulley. Once highly charged, current flows to the high potential terminal through the current regulating resistor. This configuration can also include a corona gap as shown near the inside surface of the high potential terminal.

An additional set of spray points, labeled "charging corona rod" on drawing, connect directly to the high potential terminal and are situated directly above the pulley.

According to *The Amateur Scientist*, *"The difference in potential between the upper pulley, made 'live' by the voltage drop across the current regulating resistor and corona gap, and the high-voltage terminal causes these points to 'spray a charge of opposite sign onto the downward run of the belt.' "*

The value of the current-regulating resistor can be computed roughly by Ohm's law. In small machines it is usually on the order of

40 megOhms.

Optimum belt speed for a high power unit is between 4,000 and 6,000 feet per minute. Most high power units use belts made from rubberized fabric.

If your design is to exceed 200,000 volts, charge distribution along the insulating column becomes important. Machines using belts more than four inches wide should be equipped with equipotential rings as shown spaced every two inches.

A large Van De Graaff generator intended to reach the million volt range is often sealed in a steel tank containing inert gases such as carbon dioxide and kept at many atmospheres of pressure. Keeping the gases at high pressure serves to increase their insulating properties and improves both the voltage and current output of the generator.

Owning a generator of this size would be incredible!

INDUCTIVE ELECTROSTATIC GENERATORS

Wouldn't it be wonderful to have a generator that could produce almost 90,000 volts while being virtually immune to changes in humidity? What if that same generator could be run extensively without concern for wear or maintenance? What if the same generator could be "daisy chained" to other generators to increase their output? What if it could be run as a motor, operating on electrostatic charges, just as a common D.C. generator can also by run as a motor? Well, what we would certainly have is a Dirod generator.

The Dirod generator is another invention of Mr. A. D. Moore. The Dirod gets its name from the main components of the generator: a high grade Plexiglas disk and a series of metal rods. "Dirod" is short for disk and rods. Its operation is most easily understood by observing a schematic representation of its workings.

Figure 26 shows a schematic of a basic Dirod generator. A number of metal rods are attached to an insulating disk (Lucite, glass, plastic, etc.) which is mounted on a bearing that allows it to rotate. In the center of the disk is an axle which is stationary, and attached to this axle is a wire with two brush ends as shown. As the disk rotates, each pair of metal rods is momentarily electrically connected to the stationary brushes as the rods pass by them.

Outside of the wheel we have two smooth metal plates each with

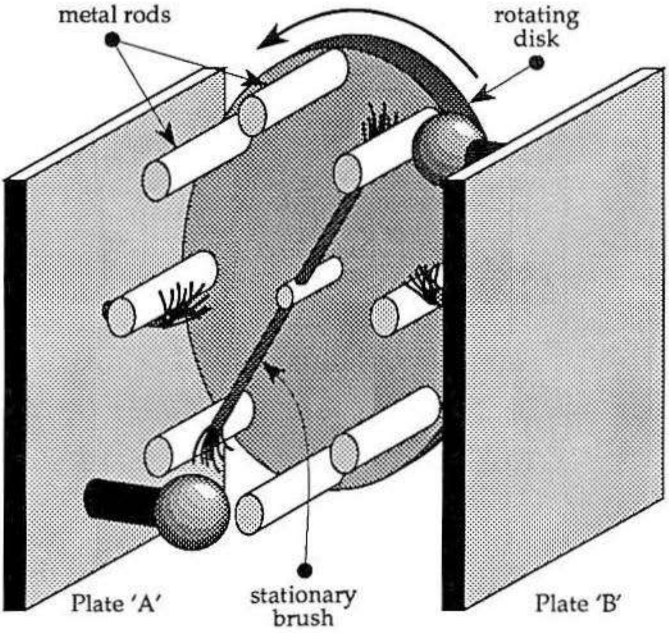

Induction Generator
schematic diagram

Figure 26

an electrode attached as shown and each having a pair of metal brushes as well. Positive and negative charges will form on these plates in the following manner. When the wheel begins its rotation, plate A and plate B will have similar amounts of negative and positive charges – similar, but of course, not exactly the same! Let's say that plate B has a slightly negative charge, and plate A has a slightly positive charge. Now, as a metal rod moves by the electrode from plate A, it is momentarily connected to the rod moving past the electrode from plate B, connected by the stationary brushes from the center axle as shown in figure 27a. At this point, since plate B has a slightly negative charge, the positive charges in the two bars rush into the bar closest to plate B (negative and positive charges attract), and the negative charges rush into the bar closest to plate A. This is comparable to one electrophorus charging the other without having them

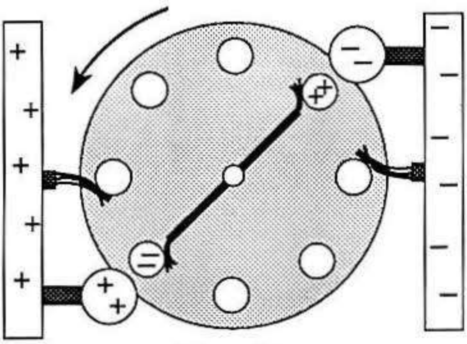

Figure 27a

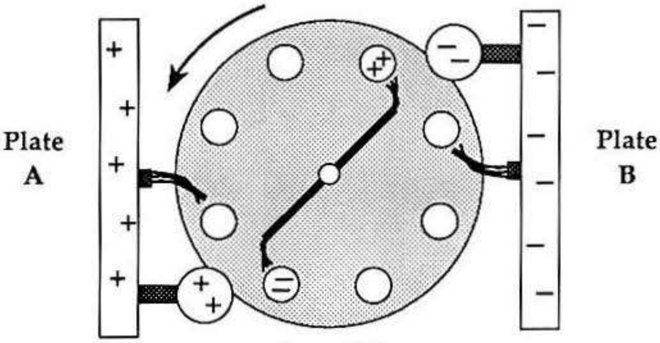
Plate
A
Plate
B

Figure 27b

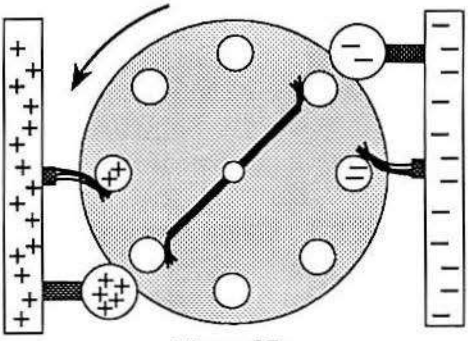

Figure 27c

Induction Generator Operation

actually touch.

As the disk continues to rotate, the two bars move past the brushes and break electrical contact. Now we have one bar with a mostly positive charge and one with a mostly negative charge. Figure 27b

As the bars continue to revolve around to the opposite plate, they briefly make a connection with the brushes from that opposite plate and bestow upon that plate their charge. Figure 27c. Now the plate has an even higher charge than it did originally. It charges the next rod passing by it even more forcefully. In this fashion the plates rapidly build up a larger and larger charge.

When you consider how fast these rods can be spun, and how fast these charges double, you realize how quickly this generator can produce high voltage!

To be a little more technical in describing the difference between our previous friction generators and an induction generator of this type, remember that friction generators charge by contact between two bodies and by the sharing of electrons from the resulting contact. Inductive generators make no demand on free electrons, but on the field they produce. This field causes electrons in the charging material to veer from their orbits. This displacement sets up or induces a charge in the formerly uncharged material. The series of figures 27a through 27c show not only how the machine moves but also how charges are built up during each step of the charging process described.

It is important to understand these simple principles of induction since they are critical to understanding and building the electrostatic motors that follow.

It is interesting to note that the Dirod generator is virtually the same as the world's first electric motor known as the "Franklin" motor for its inventor, Benjamin Franklin, and discussed later on.

Construction of the Dirod Generator

Construction of the Dirod generator is, again, straightforward and simple when viewing the drawings. Keep in mind that the metal rods can be mounted horizontally as shown or vertically, radiating out from the center. Vertical mounting is much trickier. However it is much less stressful on the Lucite disk allowing higher rotational speed to be achieved.

Also very important in achieving maximum voltages is a corona shield which is not shown in the drawing. A corona shield can be fabricated simply by placing a small sheet of Plexiglas between the metal spheres and the rotating rods. I would suggest that you first try your generator without a corona shield, and then with. If your design is optimum, your voltages will increase with the shield.

Because the brushes in this type of design can really take a beating I strongly suggest using a conductive foam such as the foam integrated circuits are packaged in or a conductive rubber brush. If forced to use a metal brush, use a metal that is very springy.

I built an induction generator by modifying the Franklin motor shown later on. By adding brushes and two stainless steel collectors to the Franklin motor, I had both an electrostatic motor and generator! Although I plan to build a much larger, more traditional Dirod machine, this small motor/generator still holds my fascination and continues to teach me the principles I will need to employ in devising the larger unit.

MEASURING VOLTAGES AND POLARITY

WARNING: As mentioned before, the electrostatic voltages produced by the generators in this text are generally safe for the experimenter even should he be accidentally shocked. The operation of the following devices is described for use with electrostatic devices. Should you choose to measure the voltage of other devices, more stringent safety precautions should be taken. Use extreme caution at all times.

Information in this chapter is useful not only in the field of electrostatics, but can also be valuable to those who work with Tesla coils, Oudin coils, induction coils, or other high voltage power generators. One of the most surprisingly difficult things to do with a high voltage is to accurately measure it. I was frustrated to find that text after text said simply that measurement of high voltages was impossible for the amateur. As a result of experimentation and having studied the writings of extraordinary experimenters such as Professor A.D. Moore, I can show you several methods of measuring high voltage that seem quite dependable and accurate. You can also use one of several means to determine the polarity of a high DC voltage – something which is rarely mentioned in any text.

Finding Polarity and Simple Voltage

There are several very simple means for determining the polarity of a high potential. The first, and by far the most sensitive is an FET based "electroscope." Figure 45 shows an electroscope of my own design. The device is simple to assemble and is shown in both schematic and component layout forms. The device works extremely well.

I highly recommend that you build one of these. It can show you the polarity of both very high voltages as well as low electrostatic voltages. In fact, this electroscope will respond to the static build up found in clothes taken from a dryer, or the charge that builds up on a hard rubber comb.

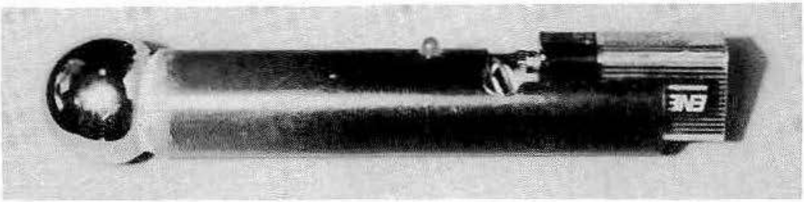

Figure 45 – The FET Electroscope

You may want to modify my design. You should consider replacing the 10 megohm resistor with one of a smaller or larger value. It controls the sensitivity of the circuit. The larger the resistor used, the more sensitive the device becomes, the smaller the resistor, the less

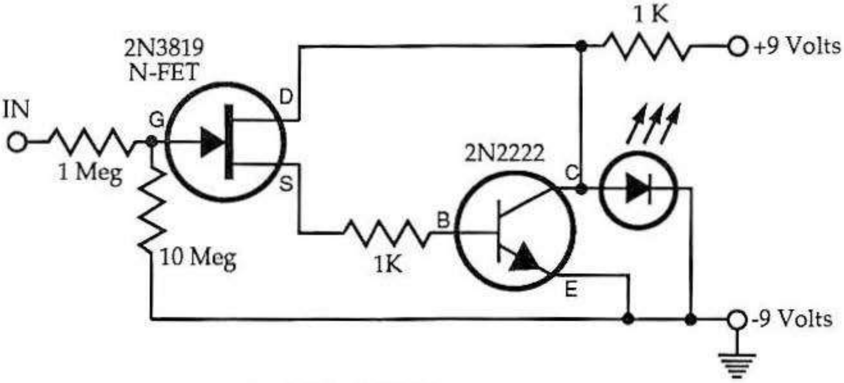
2N3819
N-FET
IN
G
D
S
1 Meg
10 Meg
1K
B
2N2222
C
E
1 K
+9 Volts
-9 Volts

FET Electroscope
Schematic Diagram
Figure 29

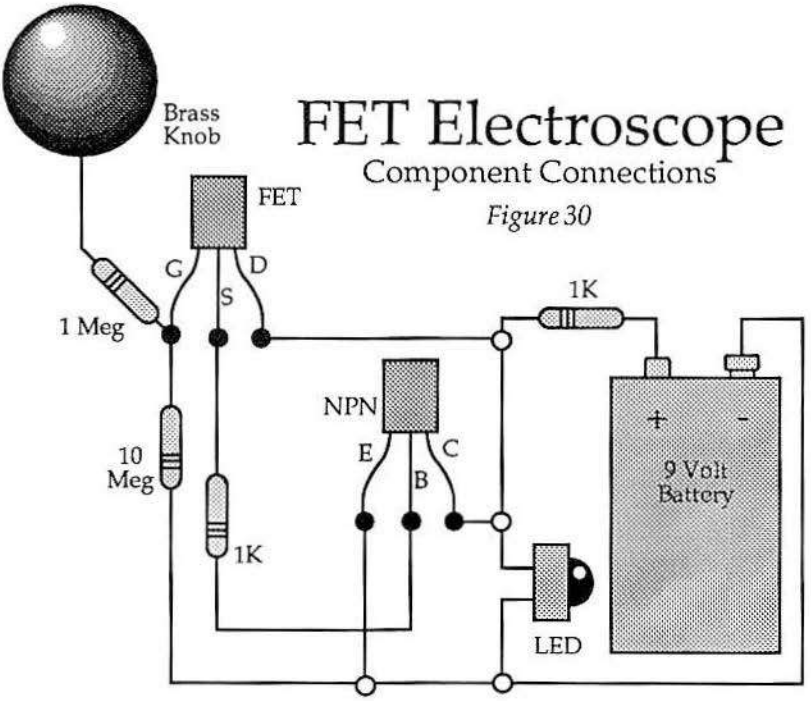
Brass
Knob
FET
G
S
D
1 Meg
10
Meg
1K
NPN
E
B
C
1K
+
-
9 Volt
Battery
LED

FET Electroscope
Component Connections
Figure 30

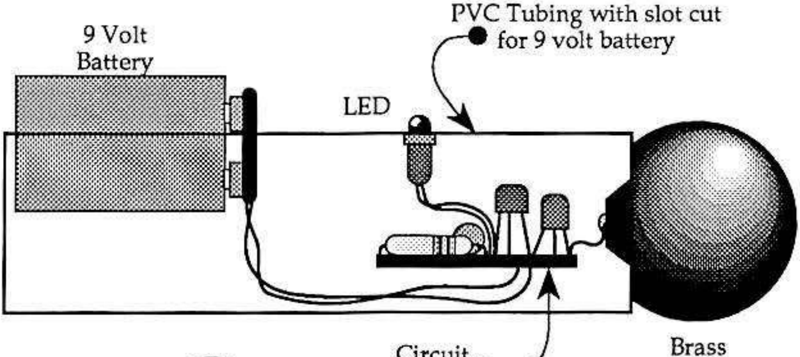
PVC Tubing with slot cut for 9 volt battery
9 Volt Battery
LED
Circuit Board
Brass Knob

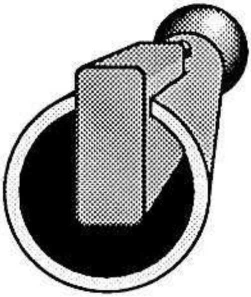
View From Battery End

FET-Based Electroscope and Polarity Indicator

Figure 31

sensitive – within reason, of course. The resistor can also be removed all together for maximum sensitivity, but the device becomes unreliable because even stray radio emissions will be picked up.

When properly adjusted, the FET scope will light when placed in contact with a positive static charge. Keep in mind that the FET transistor, which is the heart of the circuit, could be damaged by an extremely high voltage, however, I have used my scope without difficulty to more than 30 thousand volts (30 Kv). This voltage was measured by placing the scope directly against the Van de Graaff with an appropriate spark gap.

A MECHANICAL ELECTROSCOPE

A mechanical version of the electroscope can be made quite easily from a glass jar, a little wire, and some aluminum foil. A mechanical electroscope is shown in figure 32.

When the top electrode on this scope is touched to a static voltage source, the aluminum leaves in the jar will separate. This happens because the foil in the jar takes on a uniform charge. Since like charges repel each other, the two leaves of foil are forced apart.

The demonstration of an electroscope can be quite impressive. Although it would be difficult, if not impossible, to calibrate, a

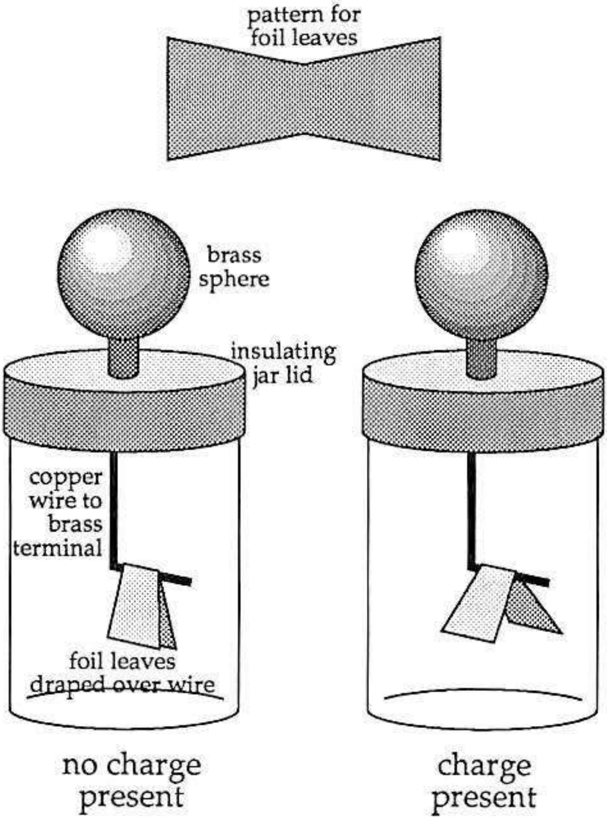
pattern for foil leaves
brass sphere
insulating jar lid
copper wire to brass terminal
foil leaves draped over wire
no charge present
charge present

Mechanical Electroscope

Figure 32

mechanical electroscope of this type can also be used to measure voltage. Heavier leaves require greater voltages to separate them. Therefore, the distance the leaves move away from one another indicates a relative voltage.

The Neon Lamp Bank

Another less sensitive but much simpler way to measure low electrostatic voltages is with the use of a neon lamp bank. Figure 33 shows the assembly of such a bank. Ne2 neon lamps are wired in series along a Plexiglas panel as shown in the drawing. Since each lamp fires at about 70 volts, five bulbs wired in series will read a voltage of up to 350 volts. You can quickly see how larger or smaller banks can be easily built. The cost of these bulbs is quite low, so even fairly large banks can be built without great cost.

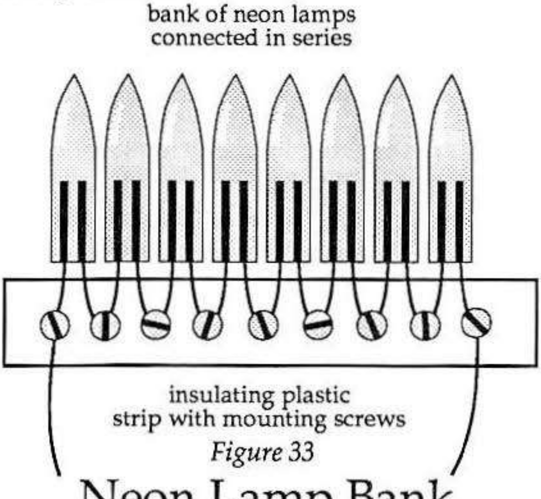

Figure 33

Neon Lamp Bank

When connected to a source of high voltage, it is quite easy to see which bulbs are lit. Sometimes, however, you may have to darken the room to get a more accurate reading. Neon lamps are not always easy to see.

You can increase visibility by leaving a small gap between the lamp bank and your high voltage source. Letting a spark periodically jump the gap will trigger the lamp bank. Attaching the bank directly to generator can draw enough current from the generator (load it down) such that it never reaches peak voltage. With a small Van de Graaff a steady stream of bright flashes will occur every two or three seconds.

Though this means of voltage measurement is crude, you could improve it somewhat by using bulbs with very close firing voltages.

The best way to find such bulbs is with the circuit such as shown in figure 34. Note that a volt-ohm meter can also be connected across the leads shown in order to gauge the exact voltage at which the bulbs will fire. By slowly bringing up the input voltage to a set of bulbs wired in parallel you can find which bulbs fire together and at what voltage.

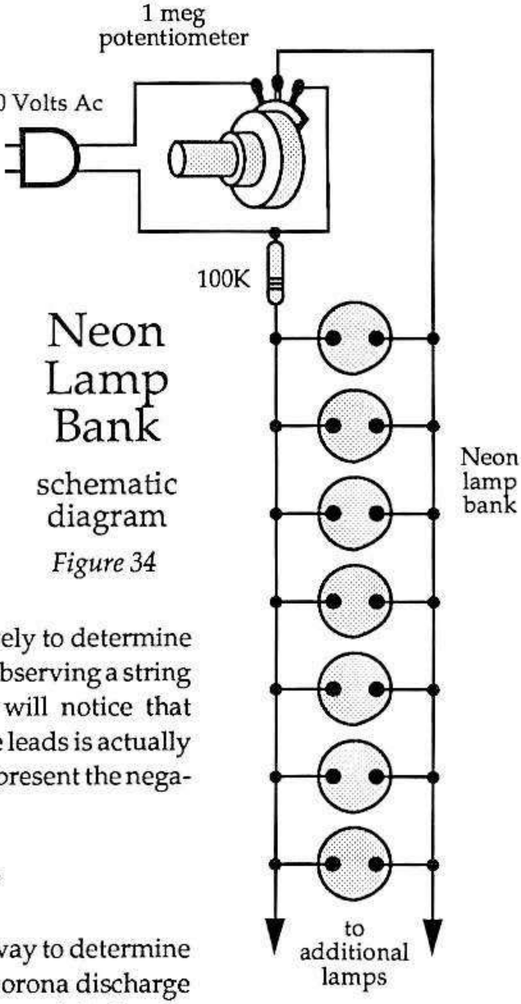

Neon Lamp Bank
schematic diagram
Figure 34

A neon bank can also be used very effectively to determine polarity. After carefully observing a string of flickering bulbs you will notice that only one of the two inside leads is actually lighting. This lead will represent the negative electrode.

Corona and Polarity

Another interesting way to determine polarity is by observing corona discharge from your high voltage potential. Corona discharge is that strange very faint blue glow that surrounds portions of your high potential terminal. To observe it, you will need to be in a darkened room with your eyes fully adjusted to the dark and your generator running.

One might assume that positive corona and negative corona look the same, but this is not true! Positive corona is even and uniform in appearance while negative corona is bushy, often with streaks or streamers that move around.

I first noticed this while using my external Van de Graaff, the same one described earlier. Not having had much experience with this generator, I immediately assumed that the light blue glow I was seeing in the darkness was due to arcing in the D.C. motor. It was too dark to see where the corona was coming from. Fearing the contacts in the motor would eventually be damaged, I tried every method I could think of to stop motor arcing, but the pesky blue glow went on unabated. Finally, I reached out in the darkness thinking I was going to touch the motor, but much to my surprise ended up touching the top of the generator! I had not expected to observe a corona discharge on this machine.

Another interesting fact is that negative corona can be heard! If you find that a hissing sound is coming from your generator, this will be generated by the negative corona discharges!

Using Spark Gaps To Determine Voltage

A spark gap is the best way I know of to accurately measure high voltage without having to resort to very expensive laboratory equipment.

Figure 35 shows the basic configuration of a "sphere gap" capable of accurately measuring up to 90,000 volts. A sphere gap consists simply of two hollow stainless steel spheres 1/2 inches in diameter which are attached to a length of threaded rod with epoxy glue that has

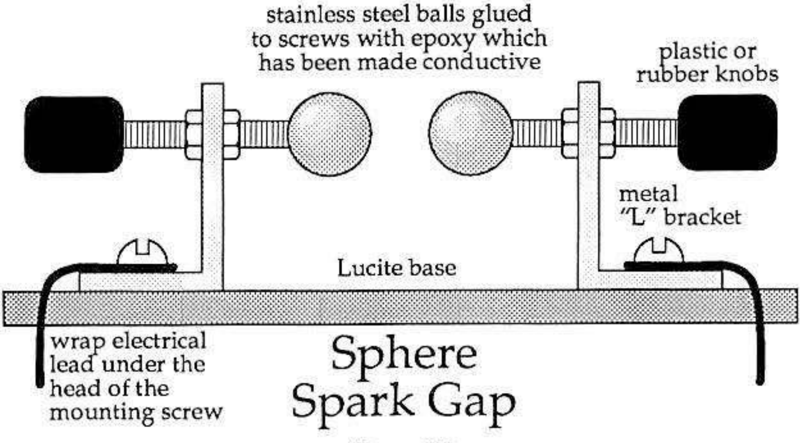
stainless steel balls glued to screws with epoxy which has been made conductive
plastic or rubber knobs
metal "L" bracket
Lucite base
wrap electrical lead under the head of the mounting screw
Sphere Spark Gap

Figure 35

been made conductive by some additive or has been coated with a little silver paint in order to make electrical contact between the sphere and the rod. The rod is threaded through a metal stand, and plastic handles are added to allow the spheres to be turned while the device is operating. The spheres must be at least 3 sphere diameters above the base to prevent distortion of the electrical field which can give inaccurate results.

Another type of gap you can build is a rod gap consisting of two aluminum rods 3/4 inch in diameter. The rods are lined up with each other just like the sphere gap. The ends of the rods are rounded by turning on a lathe. Precision is important. The ends of the finished rods must be as close to hemispherical in shape as possible and must be highly polished. A rod gap of this type will give fairly accurate measurements up to 60,000 volts.

Unfortunately, there are too many variables to give a simple equation for determining at what voltage a spark will bridge a given gap. The only practical solution is to either calibrate your gap by using devices that generate known voltages, or to use a set of measurements which have already been taken.

What follows is a chart of known values for the apparatus described. These charts are excerpts from *Electrostatics* by A.D. Moore.

kV	Rod gap	Sphere gap
5	0.056	
10	0.112	
15	0.170	
20	0.228	0.234
25	0.288	
30	0.355	0.370
35	0.430	
40	0.520	0.531
45	0.625	
50	0.750	0.710
55	0.903	
60	1.122	0.950
70		1.102
80		1.380
90		1.700

Note: Gaps listed are in inches.

One limitation of measuring high voltages with a sphere gap is that you will loose accuracy dramatically if the gap between the spheres becomes larger than the diameter of one sphere. Because of this, you might need a general way to calculate voltages for spheres of different sizes. Also, you may not be able to find a sphere of the exact size mentioned and will need to have a reference for adjusting your figures.

What follows is a listing of spark gaps for spheres of several different sizes. By using these numbers and plotting along a curve, you can remain fairly accurate in determining the characteristics of spheres whose sizes fall between those shown.

Gap Distances in Inches
for a variety of voltages and sphere sizes

	sphere sizes in cm				
kV	2.5	3	4	5	10
10	0.30	0.30	0.30	0.30	0.30
20	0.61	0.61	0.61	0.61	0.61
30	0.95	0.95	0.95	0.95	0.95
40	1.40	1.32	1.30	1.30	1.30
50	2.00	1.82	1.73	1.71	1.65
60	2.81	2.40	2.21	2.16	2.02
70	4.05	3.16	2.80	2.68	2.41
80		4.40	3.50	3.26	2.82
90			4.40	3.93	3.28
100				4.76	3.75

THE "NOON" ION HIGH VOLTAGE METER

OK, I have to admit that it is terribly arrogant for me to name one of my devices after myself, but what the heck. For the record, I won't complain if you affix your name to one of your inventions! I think that you will find this meter which has solved so many of my voltage measuring problems to be an extremely useful tool in experimenting with high voltage.

As we have discussed, the surfaces of a high potential terminal should be smooth and round, ideally spherical, in order to prevent electrons from flying off into space. If we intentionally direct these electrons onto a sharply pointed terminal, the density of electrons will increase to the point where they begin to jump out into the atmosphere that surrounds the terminal. As these electrons stream away, the pointed terminal moves in an opposite direction by virtue of Newton's second law of physics (ie. for every action there is an equal and opposite reaction.)

This reaction is extremely small. However, it is measurable, and can even drive a small motor (described later) or, as I have discovered, push the pointer on a standard panel meter! Since the strength of the reaction is directly related to the voltage accelerating the electrons, it is possible to construct a meter capable of measuring high voltage.

One difficulty with this meter is that, if constructed from off-the-shelf parts, no two meters will read exactly the same due to variations in the exact dimensions of the pointed terminal and the meter itself. It is necessary to use other methods to initially calibrate your meter such as comparing meter readings to voltages ascertained by the spark gap method.

When using your meter remember to keep it well clear of ground or any object that might attract the ions streaming off your high voltage generator. Anything that upsets the stream of electrons can affect the meter reading.

HIGH VOLTAGE AND SAFETY

Because storage devices can accumulate enough electrical charge to be lethal, we must discuss safety.

Most texts simply warn the experimenter to stay away from a high voltage discharge, and, of course, this is a good idea. No shock is completely safe, and people have been injured and even killed by remarkably small voltages.

Even low voltage can deliver high damaging current. A capacitor carrying just a few volts has been known to melt the end of a screwdriver that accidentally bridged its terminals!

If your experiences with electricity are anything like mine, no matter how careful you try to be, electricity will occasionally find some clever way to zap you when you aren't looking. Because of this, you should know how to determine when any source of high voltage you are experimenting with is safe.

The U.S. Atomic Energy Commission helps us by telling us that a discharge to the human body of 10 joules is known to be hazardous to life. A discharge of a mere 1/4 joules will deliver a heavy shock.

Use the following equation for figuring the charge in Joules:

$$\text{Joules} = \frac{1}{2}CV^2$$

where:

C = Capacitance in Farads
V = Voltage in Volts

What follows (again from A. D. Moore) is a short list of capacitors and the number of kilovolts needed to store up a dangerous 10 joule charge.

microfarads (uF)	kilovolts (kV)
0.002	100
0.20	10
20	1
80	.5
320	.25
2000	.1

From this you'll see that even small capacitors can store lethal amounts of electrical charge. Just because you may be dealing with less than 10 Joules does NOT mean you are safe! Again, remember that no shock is completely safe. If you are careful and use good judgement, you will have little to fear from electricity.

Most of the generators in this book will produce more than enough power to run any of the experiments outlined directly. High voltage static electricity is safe because there is so little current available. A Van de Graaff machine moving 50 square inches of belt per second generates only one microampere. Although the voltage (the pressure behind current) can be very high, there is very little current to be accelerated. Power is the product of voltage times current. One microampere would have to be accelerated by one million volts to be able to deliver one watt of power. With such low current levels, power levels are necessarily low and relatively safe.

If, however, small currents accelerated by high voltage are stored in a capacitor or bank of capacitors, it is possible to accumulate enough charge at high voltage that when it quickly discharges it can produce deadly levels of current. Some commercial high power static generators go as far as to present the warning "Do not attach capacitors." If you are experimenting with larger capacitors than those described, remember that it is no longer safe to verify static charges by touching terminals! Capacitors can store charges from *minutes to months* depending on the dielectric! Careless or thoughtless use can prove fatal!

Electrostatic generators are unusually safe. Most generators will produce no worse a shock than a good jolt resulting from shuffling across the carpet. However, in coupling your generators to energy storage capacitors you're moving into dangerously high voltages. Be very careful!

ENERGY STORAGE WITH THE LEYDEN JAR

A Leyden jar is simply a capacitor, that is, a device that stores electrical energy. Like all capacitors it consists of two conducting metal plates separated by an insulator. Figure 36

A brass knob such as those used on the handles of bathroom drawers is attached to an insulator, in this case, a flat piece of Lucite. You could also use cork, plastic, wax, or almost any insulator. The Lucite is cemented to the lid of a jar with a little Silicone glue. A large hole cut in the lid of the jar allows a wire to pass from the bottom of the knob to the bottom of the jar. The inside and outside of the jar are coated with a layer of aluminum foil. The foil should be worked with your fingers until it smoothly conforms to the sides of the jar. A wire is attached to the outside foil, and the entire outside is then wrapped with black tape.

To use the Leyden jar, the wire lead coming from the outside foil is attached to ground. A cold water pipe makes an excellent ground. The brass knob is connected to the output terminal of your generator.

Leyden jars can be assembled in any size, larger jars having larger capacitance. My personal favorite Leyden jar was made from a 5

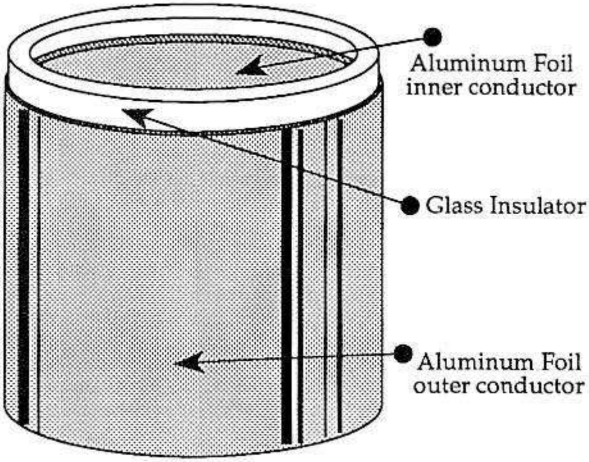

Leyden Jar & Other Capacitors
showing the
conductor-insulator-conductor layers

Figure 36

gallon white plastic pail that had contained detergent. This huge capacitor has been charged by the "Classic Van de Graaff" generator. Under good conditions it yields a very heavy, thick, multi-colored, spark of more than 4 inches accompanied by a "crack" that could be heard for some distance.

A word of warning: I made the error of touching this capacitor when I had thought it fully discharged, and though it was partially discharged at the time, I received a remarkably painful shock. This capacitor should be treated with respect. You can never be too careful.

Leyden Jar Tool

One tool you should build for your Leyden jar is a discharge rod. After experimenting with your Leyden jar, you will always want to be sure that it is fully discharged before handling or storing it. This is a very good habit to get into to avoid getting zapped when you are in a forgetful mood.

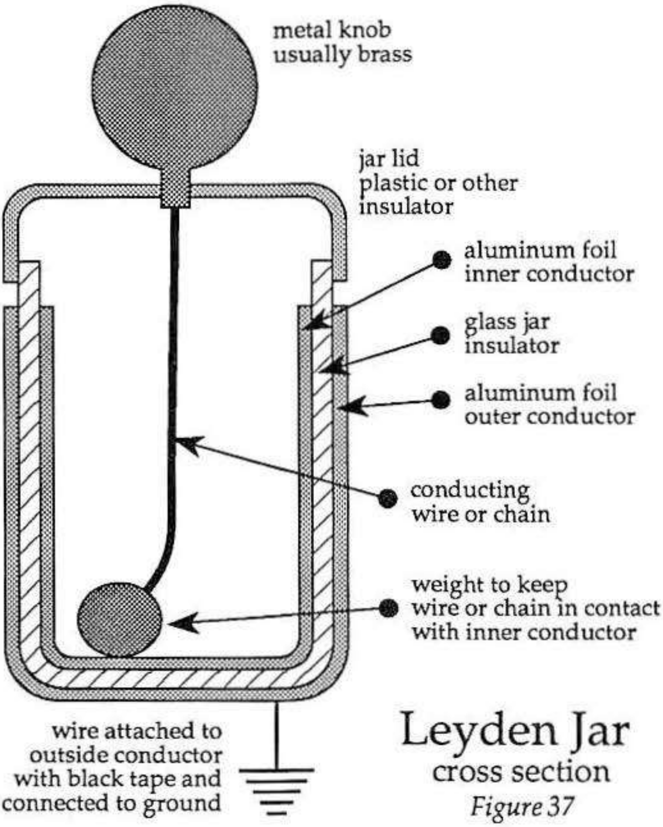
metal knob
usually brass
jar lid
plastic or other
insulator
aluminum foil
inner conductor
glass jar
insulator
aluminum foil
outer conductor
conducting
wire or chain
weight to keep
wire or chain in contact
with inner conductor
wire attached to
outside conductor
with black tape and
connected to ground

Leyden Jar
cross section
Figure 37

Electricity often seems to find some way to get into the jar without telling you. I was once shocked by a Leyden jar that was standing a few feet from a generator. It was not until I observed the generator and Leyden jar in a completely darkened room that I noticed a blue glow around a length of wire which had been acting as an aerial above the Leyden jar and had partially charged it although it was not even directly connected to the generator!

To make a discharge rod, simply take a length of heavy copper wire and bend it in a "U" shape. This "U" should cover the distance between the brass ball and the aluminum side of your Leyden jar. Attach an insulating handle to the center of the "U". Make sure this handle is long enough that a spark won't find its way along the handle to your hand. Holding it by the plastic handle, you can bridge the gap between the brass ball and side of the Leyden jar, and safely discharge it.

Discharging a large Leyden jar with this tool can be lots of fun since a large noisy spark can be safely produced. This provides a terrific demonstration for those who doubt the power of static electricity!

THE WATER CAPACITOR

Believe it or not, ordinary water and plastic bowls can be used to form an extraordinarily high dielectric, and make a good capacitor. Figure 38. A water capacitor is nothing more than two plastic bowls set inside one another each being partially filled with tap water. Any tap water will be somewhat conductive like the metal plates of the Leyden jar because of impurities. Dissolved salts in the water form positive and negative ions which move readily. The plastic wall of the inside bowl forms the dielectric while the water in each bowl forms the electrodes. A small capacitor of this type having 53 inches of useful surface area has a capacitance of approximately 100 pF.

This type of capacitor was supposed to have been a favorite of inventor, Nikola Tesla. Tesla, it is said, would wire together banks of

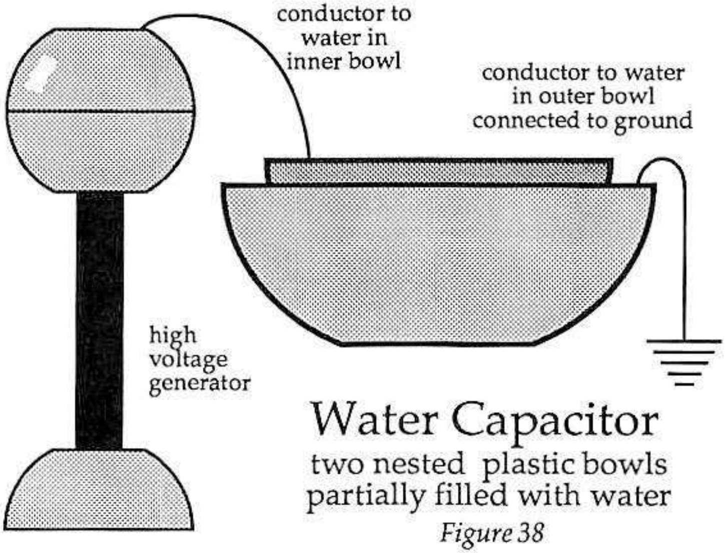

Water Capacitor
two nested plastic bowls
partially filled with water
Figure 38

insulated metal trays which were filled partially with water. By connecting these water capacitors in parallel, Tesla could reach very large storage potentials with a minimum of effort.

Plate Capacitors

As you have undoubtedly gathered, all that is required for a simple capacitor to hold a charge are two conductive plates with an insulator (dielectric) in between. There are an infinite number of configurations to be experimented with. Figure 39 shows what is probably the simplest useful capacitor possible.

The largest plate capacitor I attempted during the writing of this text was one made from aluminum foil and my sliding glass door. As soon as I became interested in electrostatics, and making high voltage capacitors, I started seeing the makings of capacitors everywhere. Every day I looked at my sliding glass door, and every day it looked less like a door and more like a capacitor. Finally I broke down and

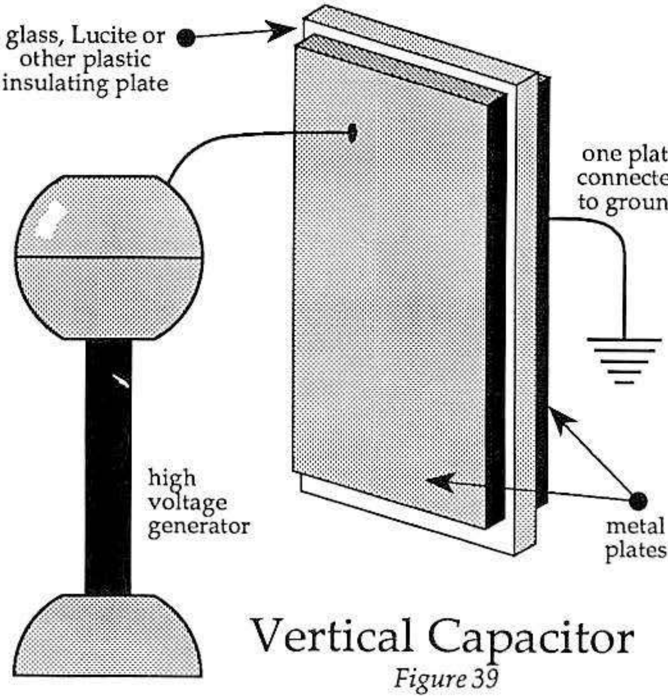
glass, Lucite or other plastic insulating plate
one plate connected to ground
high voltage generator
metal plates

Vertical Capacitor
Figure 39

bought a large roll of aluminum foil and some tape. After a thorough cleaning, I transformed my door into a giant capacitor, by taping sheets of foil to each side of the glass. But alas! The results were disappointing. Temperature differences between the inside and outside of the glass caused condensation and all the problems that moisture brings, and the large sharp edges of the foil which I attempted to insulate bled off charge rapidly into the air. The foil did odd things such as crinkling and making sounds when charged, but failed to yield an interesting spark or hold a charge for any length of time at all.

I'm sure you'll come up with as many ideas as I do. Most recently, a television special showed the huge metal domes on buildings in Moscow square giving me visions of tremendous high potentials. As I watched, the buildings kept looking more and more like high potential tops of Van de Graaff generators and less and less like buildings. I think this high voltage stuff is affecting my mind! I think it's a good thing that I don't live in Moscow.

Calculating Capacitance

Once you begin designing and building capacitors, you will undoubtedly want to know what charges they are capable of storing. Capacitance is fairly easy to calculate, although homemade capacitors will be much less than perfect and probably perform more poorly than calculations might indicate. The following equations will give you a good idea of what can be expected from your designs.

Air is a good dielectric, and is accounted for in calculations. If we have two metal plates standing parallel to one another with air between, the equation for finding capacitance in picofarad is:

$$C = .08854 \frac{A}{S}$$

Where

A is the area of one side of one plate in square centimeters

S is the distance between plates in centimeters.

The same equation can also be revised for inches:

$$C = .2249 \frac{A}{S}$$

Note that these equations remain fairly accurate as long as the smallest dimension of the plate is ten or more times "S". In other words, the area of the plate should be much, much larger than the distance between them.

Different materials have different insulating capabilities which are reflected in their dielectric constant. Since you will want to use materials other than air as a dielectric, you will need to look up the dielectric constant from standard tables. You'll find that Plexiglas provides about four times the dielectric strength of air. A capacitor built from two conducting plates would have four times the capacitance of an air capacitor.

Another example: We might calculate from the equation above that a capacitor with a desired plate size has a capacitance of 100 pF using air as the dielectric. Using the chart below we see that if dry paper were substituted for air, the capacitance will be raised to 200 pF.

Shown below are some typical dielectrics relative to air borrowed from Attwood's book *Electric and Magnetic Fields:*

Dielectric Constant	Material
1	Air
25	Ethyl alcohol
2.5	Transformer oil
4.5 - 7	Various types of glass
6.8-8.4	Plate Glass
8.0	Window Glass
4.1-6.1	Pyrex
5.4-5.8	Bakelite
2	Dry paper
2.3	Paraffin
4	Plexiglas
2.5	Pure rubber
4	Wood

Capacitors can be connected in parallel or in series. Parallel connection increases total capacity. The total capacity is the sum of the individual capacities.

$$C_{Total} = C_1 + C_2 + C_3 + \ldots$$

For example, three capacitors having capacity of 10, 20 and 30 pF each would act as one capacitor having 60 pF when connected in parallel. The breakdown voltage of the combined capacitors is only as great as the lowest breakdown voltage of any single capacitor.

Calculating the capacitance of capacitors connected in series is more complicated, being the reciprocal of the sum of the reciprocals of the individual capacitances.

In series you would figure:

$$C_{Total} = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots}$$

For example, three capacitors have capacitance of of 2, 4 and 4 pF. When connected in series, the total capacitance would be 1 pF. Although capacitance has decreased, the maximum working voltage has increased. If each capacitor had a maximum working voltage of 10,000 volts, the series connection would have a maximum working limit of 30,000 volts.

Power Capacitors

You can build a very powerful capacitor by creating a stack of alternating sheets of dielectric material and metal plates. On a large sheet of dielectric place a metal plate, then another sheet of dielectric, another metal plate and so on. The plates must be absolutely clean and free of even the tiniest contaminant or the capacitor will not work. To complete the capacitor connect alternating metal plates to one lead and the remaining metal plates to another.

For your first attempt, build a power capacitor with 5 sheets of dielectric and 3 metal plates. Connect the top and bottom metal plate to one lead, and the middle metal plate to another lead.

Benjamin Franklin — The Turkey Killer

Benjamin Franklin, an incurable prankster, was an ardent experimenter with electrostatic machines. Amazing as it may seem, according to his biography, the major reason Ben Franklin tried his dangerous kite experiment was that he was tired of charging his Leyden jar by hand cranking his static electricity generator. He wanted to charge it with lightning which he was able to do without getting himself killed! (Note: this method of charging a Leyden jar is not recommended for those other than "founding fathers" and other experimenters who seem to have a supernatural gift for avoiding harm. In other words, don't you try it.)

Franklin's hand-cranked static machine was large enough when coupled to Leyden jars to produce a lethal voltage. In fact, Franklin often used his static machine to kill his Thanksgiving turkey. In his own words "*... to the amazement of many... A turkey is to be killed for our dinner by electrical shock, and roasted by the electric jack, before a fire kindled by the electric bottle...*"

Franklin's pranks were not without risk. Just a short time before his famous kite experiment Franklin accidentally took the full force of his turkey killing device. He wrote, "*Being about to kill a turkey by the shock from two large glass jars, containing as much electrified fire as forty common phials, I inadvertently took the whole through my own arms and body, by receiving the fire from the united top wires with one hand while the other held a chain connected with the outside of both jars. The company present (whose talking to me, and to one another, I suppose occasioned my inattention to what I was about) say that the flash was very great and the crack as loud as a pistol; yet, my senses being instantly gone, I neither saw the one nor heard the other; nor did I feel the stroke of my hand.... I then felt what I know not well how to describe: a universal blow throughout my body from head to foot, which seemed within as well as without; after which the first thing I took notice of was a violent quick shaking of my body, which gradually remitting, my sense gradually returned.... That part of my hand and fingers that held the chain was left white, as though the blood had been driven out, and remained so eight to ten minutes after, feeling like dead flesh; and I had a numbness in my arms and the back of my neck that continued till the next morning...*"

Franklin's account of his own mishap should certainly give thought to those who do not take the dangers of static electricity seriously!

ELECTROSTATIC EXPERIMENTS

The experiments here are divided into two categories. The first group describes electrostatic experiments that can be performed for an audience. In other words they're merely for entertainment.

The second section features experiments which possible useful applications of electrostatics which are not often discussed. Many of these "useful" experiments could be or are being developed into very real, marketable industries and should be particularly interesting to the serious experimenter.

Hair Raising Experiences

Using electrostatic voltages to make a person's hair stand on end is a classic experiment every audience should see. It can be done in a number of ways. Find a subject with hair that is light and clean. With some experience you will be able to recognize the best subjects immediately. Next, convince the subject that the ominous apparatus that they have just seen throwing four inch sparks will not harm them.

You may want to explain why the experiment is perfectly safe by describing the "skin effect" of high voltages. This effect refers to the strange fact that voltages of this order travel only on the outside or "skin" of the object they pass over and will not travel through the inside of a body where they could do any harm. For alternating currents this is a true skin effect, but for direct currents there is a similar effect that occurs probably because of the repulsion of like charges rather than due to a true skin effect.

The subject should stand within reach of the generator and should be well insulated from any ground. This is accomplished by having the subject stand on a large insulator such as an overturned plastic bucket, or, as a joke, a stack of college text books since they are usually very dry. Instruct the subject to place one hand on the high potential terminal, and then turn the generator on. Remember to remind a nervous subject not to remove his hand from the terminal until the experiment is completed since removing his hand could provide a

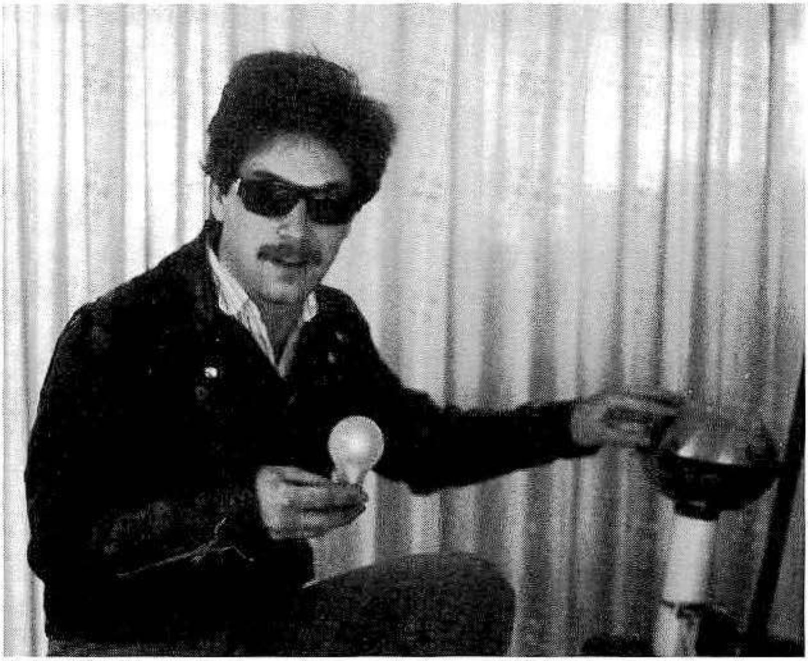

Figure 40 – The author's hair is usually flat and combed somewhat back. After a few minutes connected to a running Van de Graaff generator, and he's ready for a punk-rock concert!

shock. Depending on the output of your generator you may need to allow it to run for several minutes before your subjects hair will begin to rise. Having your subject shake his head occasionally with speed up the process.

During this whole procedure the subject will not feel the electricity passing over him other than perhaps a light tickling as his hairs stand on end. What is happening is that the subject's hair and whole body is taking on a charge from the generator. Since all of the hairs have the same charge whether it be positive or negative, and because like charges repel each other, every hair is repelled from every other hair and any nearby parts of the body. Warnings: **1) Don't use capacitors. 2) Hairspray is flammable. 3) Voltages over 500,000 are unsafe.**

Strangely enough, your generator will actually have a higher potential with someone is touching it than when it is operated alone since it can now use the person's body as a high potential collector. The success of this experiment depends largely on the subject's hair. Some women with short frizzy hair will have their hair stand out nearly straight from their head!

After the experiment is complete, have your subject remain standing on the insulator. This will allow the charge accumulated in their body to bleed off and will reduce the possibility of a large spark occurring when they eventually touch something that is grounded.

There are a few simpler although perhaps less dramatic ways to perform the same experiment. One is to make a wig out of fuzzy yarn and to place it over the generator's high potential terminal. You can place it on the spherical terminal of a Van de Graaff generator to simulate a person's head. Some Mr. Wizard impersonators even like to put a pair of plastic sun glasses under the wig to give the generator a little character. After the generator has run for a time, the hairs of the wig will stand on end.

A very quick way to demonstrate this effect is to simply take a few hairs from an old hair brush and place them on the generator's surface. They will immediately stand on end, and will often fly off the generator all together! Many different light weight materials can be used for these type of experiments. Bits of styrofoam work extraordinarily well.

The Magic Wand

As discussed, smooth, round objects tend to hold their charges while sharp pointed objects will allow electrons to fly off in a stream sometimes referred to as an "electric wind." The effects of this wind can easily be seen. Connect a short length of wire to your high potential terminal and hold the free end of the wire with a handle made of an insulating materials such as plastic. Point the end of the wire toward a lighted candle, and observe the flame. The flame will bend mysteriously away from the invisible "electric wind." In a well darkened room with ones eyes well adjusted the "wind" may actually be visible.

Another interesting experiment is to use the wand to chase a soap bubble. Because there are two forces acting on the bubble, the outcome of the experiment may be unpredictable. One force is the charge of the generator itself which draws the bubble toward the wand. The other force is the electric wind from the wand that pushes the bubble away. How the bubble will act is anyone's guess! It depends on which force is dominant.

A little later on we'll use this surprisingly powerful wind to create a small pinwheel motor.

Levitation

There are a number of ways that you can levitate lightweight materials using your generator. The easiest ways to demonstrate this is to simply spread very small shavings of styrofoam along the top of your generator. When the power is on, the pieces will take on the same charge as the surface of your generator and be repelled by it since like charges repel one another. The bits of styrofoam will levitate off into the air after the generator has been on a few seconds.

An entertaining experiment can be done by placing a number of tiny styrofoam balls in a short section of clear glass or plastic tubing about 2 1/2 inches in diameter and a few inches in length setting on your high potential terminal. Cover the top of the tubing with a metal plate, and connect it to ground. When your generator is turned on, the balls become charged and levitate to the top of the plastic tube. When they touch the metal plate at the top of the tube they lose their charge to ground and fall back to the generator again. The balls will oscillate between the ends of the tube until the generator is shut off. You should coat these balls with a conductive paint for a more complete charge transfer.

If you enjoy reproducing electrostatic experiments as they were originally done, you might want to try making your own old-fashioned pith balls instead of using styrofoam. Pith balls were used for these experiment before modern products like styrofoam were available. Pith balls are made by removing, shaping, and drying the insides of certain plants such as dandelions. Details can be found in old physics textbooks before 1900. I'll stick with good old styrofoam.

Place larger bits of styrofoam on your generator, and turn your generator on while touching the bits with your hand. The styrofoam will assume that you are ground and that your generator is a high voltage source. One side of the styrofoam will be charged one way, and the other side of the styrofoam the other. After removing your hand, a number of things can happen. The styrofoam may flip over and be stuck to your generator like a magnet. The styrofoam may stick to your hand like a magnet. Or you may find the styrofoam floating on end as the hair did in the last experiment.

An interesting trick can be performed by passing your hand over the styrofoam remaining on your generator terminal. As you do so, the electrical charges in the foam and in your hand will cause the styrofoam to jump about.

I have meant for some time to place a metal plate on a table underneath a table cloth and to energize that plate from a hidden electrostatic machine. I could then sit at the table and "mysteriously" move light weight objects such as bits of hair, sawdust, styrofoam or other similar materials around the table simply by passing my hands over them. I think this would make a very convincing illusion of mental telekinesis. Try it.

Lighting Without Wires

If you or your friends doubt the power of your generator, have on hand a supply of Ne2 neon bulbs which are available for less than a dollar at any electronics store. Remind those onlooking that it takes at least 70 volts to light a neon bulb. Hold one lead of the neon bulb in your hand and move the bulb toward your generator. Within inches, or even feet, the bulb will flicker or light. I have seen some experimenters hold the bulb in their mouth while reaching toward a generator with their hand. Seeing a 70 volt bulb light in a persons mouth with no wires attached is an impressive demonstration.

You may want to try large florescent bulbs as well. When my generator has been running I have seen flickers in the florescent lights in the ceiling of my home. These bulbs are many feet above the top of my generator yet respond well in a darkened room.

Perpetual Motion

Anything that people assume can't be done interests me. Many of the greatest inventions of all time would never have come into being if their inventors hadn't at one point been told that it couldn't be done. Currently, perpetual motion qualifies as such.

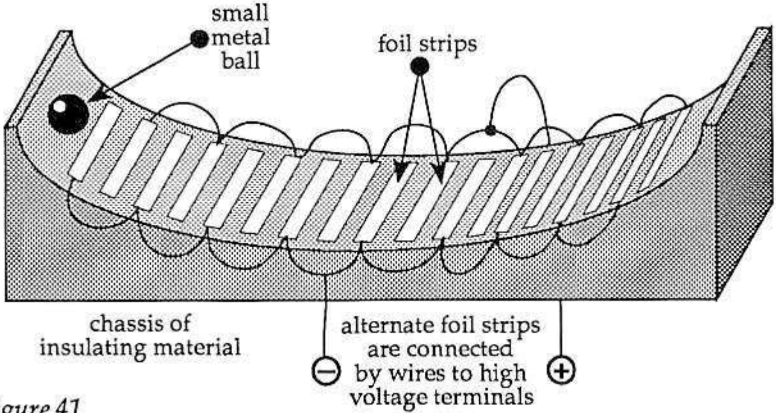

Figure 41

"Perpetual Motion"

The following device doesn't really qualify as perpetual motion, but it appears to be! Figure Pl shows the device. The idea is simple. A few light-weight conductive balls are released onto a U-shaped track on which the balls rock back and forth. It is normal to assume that the balls will slowly lose their energy and come to a stop in the center of the track. Instead, because of the electrodes in the track,the balls continue back and forth indefinitely and even pick up some speed along the way.

What has happened is that you have constructed simple electrostatic motor. As the balls pass over the foil strips they pick up the charge from that strip. Having the same charge as the strip, they are repelled by that strip and are simultaneously attracted to the next strip which has a dissimilar charge. Inertia will cause the balls to roll over the dissimilarly charged strip and on to the next. The electrostatic forces replace the energy lost in rolling friction. In this manner the balls will remain in motion for as long as power is supplied.

If you wish to really complete the illusion given by this device you will naturally want to hide the foil strips from view. To do this you could either use a conductive material other than foil that blends into the background, or, if your generator is powerful enough, you could probably bury the foil strips under the ramp material and still transfer a sufficient charge.

Trick Gun

We know that like charges repel one another. This has been demonstrated with the 'hair' experiments and others. What would happen if we introduced a charge into a liquid, and then shot that liquid out in a stream? We can find out with the use of a squirt gun. Purchase an inexpensive water gun that stores its water in the handle. Drill a small hole into the handle and twist a brass sheet metal screw into the hole. You may want to put a little Silicone glue around the top of the screw to make sure that this point will not leak when the gun is later filled with water. The brass screw provides electrical contact between your hand and the water.

Fill the gun with water, and pump out a few squirts. The water will come out in its customary straight stream probably covering the length of the room. Then stand on an insulator and charge yourself up as described earlier. When you squeeze the trigger, a stream of scattered droplets will be ejected only a few feet. What has happened is that all water droplets have a similar charge so each droplet is repelled by every other droplet. Other than being just an interesting way to get an unsuspecting friend wet, this mechanism can be applied anywhere that a liquid must be atomized.

I've wondered about making an electrostatic chamber on the carburetor of my car to see if better atomization of the fuel would result in better combustion. But first I need to check into the dangers of a spark occurring! I don't really want to blow myself up!

An Ion Motor

There are several different electrostatic motors that can be run by your generator, the simplest being an ion motor. You can make one by simply bending a length of light weight wire into a "Z". At the center of the "Z" shape, place a conductive bearing. Nothing fancy is required. In fact, just looping the center of the Z around an axle may work. Make sure however, that the Z will rotate very easily. A simple schematic of the ion motor can be seen in figure 42.

When power is applied to ion motor electrons stream off the pointed ends of the wire creating an electric wind. The force of the wind pushes the wire in the opposite direction. Again, for every action

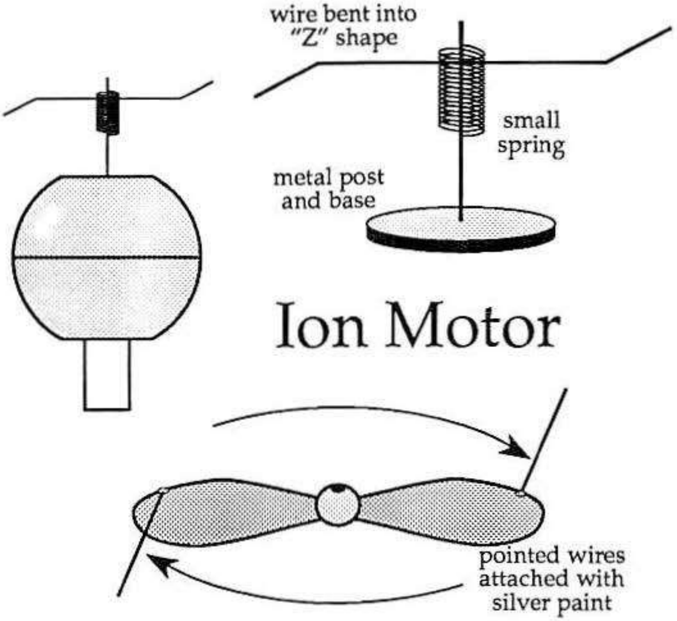
wire bent into "Z" shape
small spring
metal post and base
Ion Motor
pointed wires attached with silver paint

Plastic Toy Airplane Propellor
Figure 42

there is an equal and opposite reaction. If your rotor is well balanced and the bearing friction is very low, your motor can and will achieve speeds of several thousands revolutions per minute! Of course, the torque generated by this motor will be very, very small. Nevertheless, it is fascinating to watch a motor propelled by a stream of electrons!

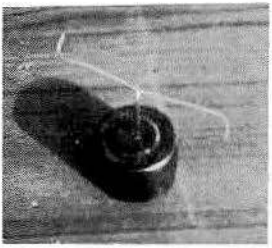

Figure 43 –Simple Ion Motor

Although the thrust of this motor is very small, I am certainly not the type of experimenter to let this hard-earned stream of electrons get away without doing some useful work along the way. I have attached one of my pinwheel motors to a tiny plastic airplane propeller as shown to blow a little air cool air my way. I think this is probably one of the few fans that I know of that must be connected to 100,000 volts or so to operate!

Another thought that I had for this fan was to drive it with a negative charge. Negative ions are thought to have many health benefits, and to cut down on pollutants in the air. Negative ions apparently attach themselves to specks of airborne dust and cause them to fall from the air. There are a number of negative ion generators being marketed for this purpose. I thought that a spinning propeller of this sort would be an absolutely outstanding way of both distributing and producing negative electrons in a room environment.

Ion motors also have serious applications in space travel. The Soviets have used ion motors to propel some of their long range probes. The problem with these motors is that the thrust is very small. However, ion motors offer at least two enormous advantages. First, they can be powered by a nuclear or solar electrical source which can provide power far longer than conventional chemical rockets. Second, an ion motor can provide higher velocity thrust than current chemical rockets because the nozzle velocity of a rocket is limited by the internal pressures that can be safely generated inside the rocket motor. An efficient ion motor can accelerate ions to speeds approaching the speed of light! Although the thrust is small, over time the space craft could itself approach the speed of light, at least, in theory.

A few tips for getting the best operation out of your ion motor—Make sure that your rotor is well balanced especially if you want your motor to be self-starting. Your rotor should spin easily for several seconds when tapped lightly by your finger. The more freely your rotor can spin, the faster it will revolve. The bearing used in a radiometer would make an excellent ion motor bearing.

Run your generator/motor in an area free of objects that might provide sources of ground. My first ion motor was well built and carefully balanced, yet it refused to spin. When I reached down to give it a tap to get it started, it refused to move. In fact it felt as though it were being held in place by a magnet. It was then that I realized that one of the arms of the rotor must have been attracted to an object with an opposite charge. That object turned out to be me! When I moved a short distance away, the ion motor sprang to life and immediately whirled up to several hundred RPM.

For this motor, voltage is more important than current. The motor will run and do useful work on currents under a microampere. Electrostatic motors have been run with currents of less than .001

microamperes, and have even from the earth's electrical charge itself.

Another interesting use that I have found for an ion motor is that of a primitive high voltage meter. The higher the voltage, the greater the thrust from the motor. This may be used to move a pointer such as in a panel meter or may be used simply to a relative indication of voltage.

I like to place a small ion motor on top of my Van de Graaff generator when I am charging capacitors. When the capacitors are charging, the motor will not move since the charge is being bled off from the generator to the capacitors. When the capacitors are fully charged, they stop drawing current from the generator. The terminal voltage rises, and the motor begins to spin. This is a very useful application of the ion motor since it is usually impossible to tell when a capacitor is fully charged any other way.

I am told that ion motors have been used in scientific toys such as those that can influence a balsa airplane in flight. Gustave LeBon's electrified rocket was a small aluminum foil rocket that could actually fly on electrostatic forces (Patent No. 2,018,585). I have not seen these devices, but they sound like something worth investigating.

Experimenting with "micro" sources of power can bring out the engineer in everyone.

An Ion Blower

The last experiment uses a stream of ions to propel a small rotor. Since in essence what we really have is a little jet engine, why not put our blower into a tube and see if we can pull a little air along with it. Figure 44 shows one possible configuration.

One end of a small, clear tube such of glass, Plexiglas, or other insulating material is fitted with a sharp metal point. A metal screen covers the other end. Electrons reach a high density at the metal point, break free into the air, and stream through the tube to the screen at the far end. As they move rapidly through the tube, they pull some of the air inside the tube with them. Obviously, a blower of this type has very low efficiency probably in the neighborhood of 1% under ideal conditions, Nevertheless, a puff of smoke at the intake side of the tube will easily be drawn through.

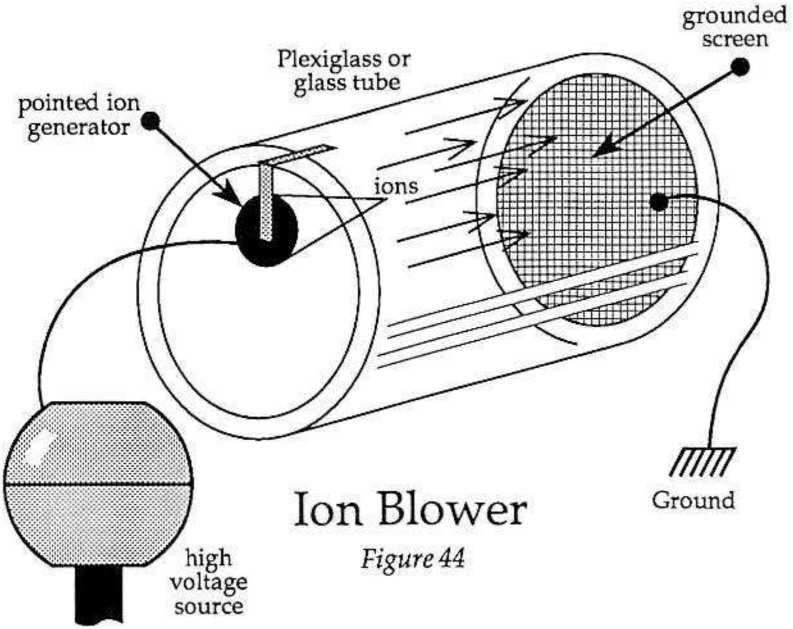

Ion Blower

Figure 44

Clearing The Air

Before moving on, here's another interesting smoke experiment you can try. Get a clear plastic or glass jar having no metal parts. Drive a pointed nail through the center of the lid. Blow a little smoke into the jar, and seal the jar with the lid. Next, connect the head of the nail to the output terminal of one of your high voltage generators.

As charge is applied to the smoke, it will begin to swirl. Soon the smoke will seem to disappear, although in reality it has been redistributed along the sides of the jar.

Variations of this effect are used in industry to place very fine films on surfaces. Similar electrostatic methods are being used to make sandpaper. When the tiny metallic abrasive particle are distributed along the surface of the sand paper they are given a charge that causes them to stand on end creating sandpaper with a better abrasive action.

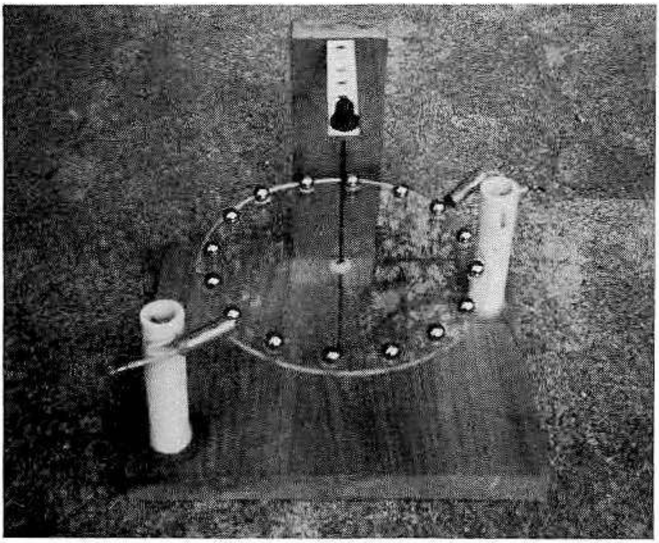

Figure 45 - Franklin Motor – Using large sling-shot balls and a simple Lucite disk, this little motor quickly speeds to over 600 rpm on less than 1 microampere of current!

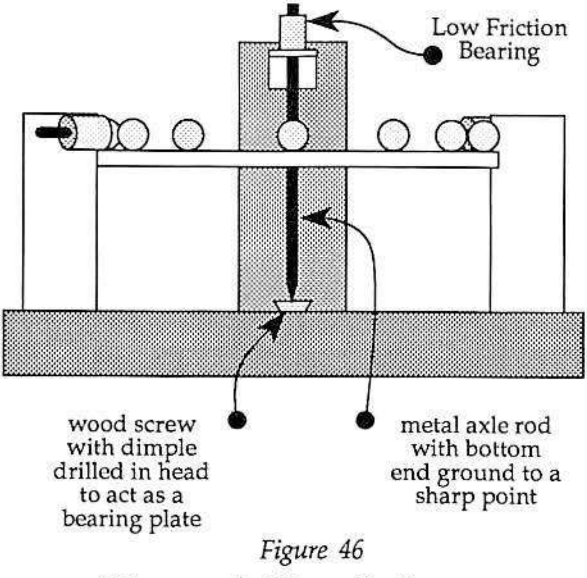
Low Friction Bearing
wood screw with dimple drilled in head to act as a bearing plate
metal axle rod with bottom end ground to a sharp point

Figure 46

Franklin Motor

Front View

Figure 47

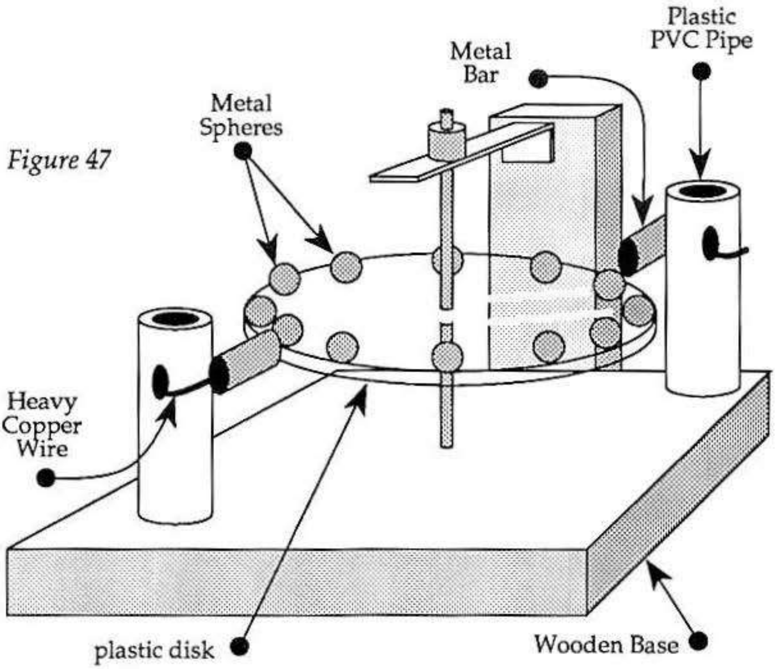

Franklin Motor

THE FRANKLIN ELECTROSTATIC MOTOR

The Franklin motor, named after its inventor Benjamin Franklin, is often said to be the first electric motor ever made. Its construction is simple, and, in fact, it is very similar in operation to the Dirod generator described earlier which can be run as a motor although inefficiently. Figure 47 shows the basic Franklin motor.

A series of lightweight spheres is glued in place along the outside of an insulating disk of Plexiglas, glass, etc. As these spheres rotate, they pass through the corona emanating from a highly charged high potential terminal at each side of the disk. One terminal is charged negatively and the other positively. As a sphere passes a terminal, it picks up a charge that will be attracted to other terminal causing the whole assembly to turn until the high voltage is removed.

One modification that you might like to experiment with is to make the Franklin motor self-starting. This can be accomplished by adding points on the high potential sources to create an electric wind that gives the spheres just enough of a nudge to start the rotor revolving.

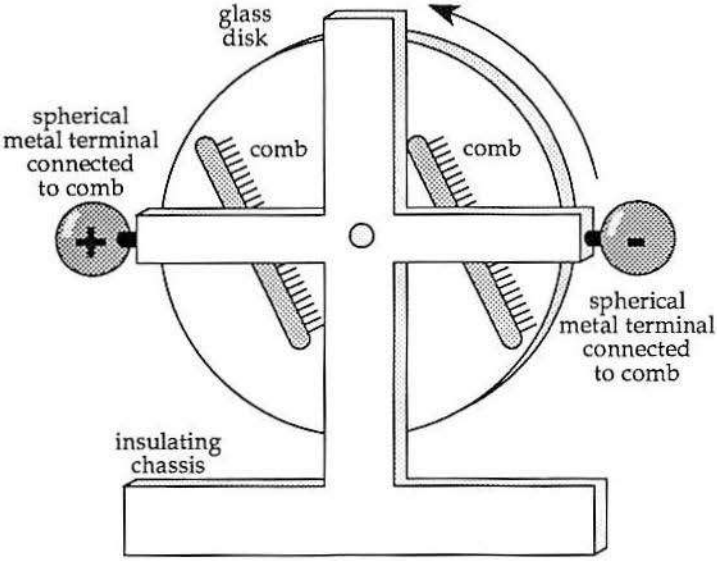
glass disk
spherical metal terminal connected to comb
comb
comb
+
-
spherical metal terminal connected to comb
insulating chassis

Corona Motor
schematic diagram
Figure 48

Poggendorff's Corona Motor

Franklin's motor works by charging a series of spheres as they pass by alternating high potentials. It is also possible to alternately charge an insulating surface as it passes near high potential terminals. Plexiglas, glass, and probably other materials will work. A drawing of Poggendorff's motor is shown in figure 48.

Here is how it works. Notice that there are two sets of brushes or combs with sharply pointed ends mounted at an angle. One brush sprays a negative charge onto the disk and the other a positive charge. As the disk continues to revolve, the charges on its surface find themselves attracted to the entire oppositely charged comb – not just the tiny spray points. It is important that the brushes be mounted at an angle, so that charges on the disk may be attracted to the comb of opposite polarity before they are replaced with an opposite charge as the disk rotates under the points of the comb. Understanding how a motor like this works is simple. Getting one balanced well enough to be self-starting and smoothly running takes quite a bit of patience. It is nonetheless an exciting and interesting project.

Atmospheric Electricity

The earth itself is a huge electrostatic generator. In the air at any time there are some 200 volts of electricity encircling you all of the time. The reason that we aren't being zapped all the time is simply that we are all at roughly the same potential. That is we all have the same level of charge, so, the electricity has no reason to move from one point to another.

Figure 49

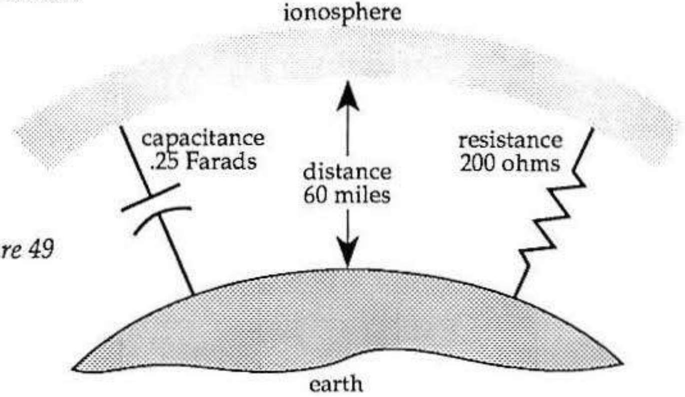

Earth's Atmosphere
electrical parameters

Between the earth and the electrically conductive ionosphere, both of which store electrical charge, exists an electrical potential. The same sort of buildup exists between clouds and the earth. When the potential builds to a level great enough, the air ionizes, loses its insulating capacity and becomes an excellent conductor. The result is a huge arc of static electricity we call lightning. As current flows between the cloud and the earth, the potential is reduced until the arc can no longer be sustained, and the lightning bolt is extinguished. It fascinates me to think that I live on one of two highly charged electrodes.

As an electrostatic generator the earth is fairly impressive. Between two and six thousand thunderstorms are in progress at any one time. Worldwide, a hundred bolts of lightning strike the earth every second. It is estimated that this gives the earth a high voltage potential of approximately 360,000 volts between ground and ionosphere with an average lightning bolt conducting 1800 amperes. The density of the electric field decreases as altitude increases. Near the earth's surface, however, the field is estimated to be approximately 100 volts per yard of height.

I say that the earth is fairly impressive because for all its size it is limited to these voltages by the fact that the atmosphere is relatively conductive having a resistance of approximately 200 Ohms. This is a relatively low resistance which quickly bleeds away charges.

Although it can take a significant amount of time depending on weather conditions, any conductive material insulated from ground can acquire charge from the surrounding air, and become charged to a high potential. Although the charge accumulated is usually negative in polarity, thunderstorms and other atmospheric conditions may create a positive potential with respect to ground. Radio antennas and flag poles insulated from ground have been known to throw hefty sparks and even start fires.

Plugging In To The Earth's Charge

If you're as curious as I am, you're asking yourself, "Can I tap into this charge?" The answer is yes! The first order of business is to measure and verify that a charge is present. If you are a skeptic like me, you've "gotta see that needle move" before you are going to believe there is something there!

The best antenna for capturing atmospheric charge is a metal structure, insulated from ground, pointing straight up into the air. The simplest, and probably easiest way to accomplish this is to simply attach a length of very light, bare, copper wire to a large helium balloon. This can be dangerous! Remember to avoid power lines and airports if you plan to launch such an energy-gathering balloon!

Having your bare copper wire come in contact with a high-tension power line is not a pleasant way to die! Thunderstorms are extremely dangerous. (Ben Franklin was lucky to have survived his experiment.)

The bottom of the wire should be attached to a high quality insulator well above ground. Attach one lead of a extremely sensitive meter to ground. Occasionally attach the other lead to the vertical metal conductor. Be careful! Do not take for granted that the potential will be harmless! If left unchecked it could easily reach a level that will bite you. You may find that it can take an hour or more to build up a significant charge, however.

At one time I tried using a kite flown with bare copper wire but had mixed results. I experienced significant difficulty with the wire kinking as I let it out and reeled it in, in response to changing winds.

Another time I hooked a diode into the line and got surprising voltages, but I had been fooled. The diode and coil in the meter had made a primitive crystal radio. I was extracting power from the atmosphere that had been placed there by radio broadcasting stations. Although this is not static electricity, it is nonetheless interesting because it is a demonstration of wireless transmission of power that Tesla was trying to perfect.

An antenna of this type has been used quite adequately to run a corona motor from the natural potential that exists between the earth and its atmosphere! A motor of this type must be extremely well built having great sensitivity and very low friction bearings. I have seen such a motor run, but I have not run one of my motors from this charge.

Because of weather, it can take a long time to build the potential up to a level capable of doing useful work such as moving a sensitive meter or a small corona motor. Several methods have been devised to speed up the charging process. Early experiments devised by Lord Kelvin used an insulated metal tank filled with water suspended above the ground constructed in such a way so as to allow a steady stream of water droplets to reach the ground. Appropriately, he called the setup a "water dropper." He reasoned that the surface of each water drop would be an extension of the surface of the water in the tank at the moment it dropped away. As it did, the drop would carry charges away with it. This continues until equilibrium was reached between the tank and the air.

Using this same idea, charging time can also be reduced by spraying a fine mist of water along a bare copper wire a hundred feet or more in length which is insulated from and suspended above ground. The falling water droplets can help to charge this wire to a measurable level in less than a minute's time.

Another method of speeding up the charging process is to increase the number of ions in the vicinity of the wire itself. A flame applied to the wire will increase the number of ions in the area of the flame and can bring a wire up to the same potential as the air around it in a matter of seconds. To accomplish this easily, wrap the wire with a flammable substance such as string or paper soaked in lead nitrate, and ignite it. It is also possible to get good results by simply waving a torch lightly across the wire.

In Conclusion

I hope that you will enjoy experimenting with electrostatics as much as I have! I find few subjects more fascinating than high voltage and its generation. As a technically minded person, you will undoubtedly see many improvements that can be made in my designs. I know that this is true because at times it seemed I might never complete this text because there was always one more improvement to make and one more experiment to try. No doubt, you will conceive of dozens of new ways to apply the basic principles of electrostatics. Like all experimentation, every change, every test, every improvement will suggest many more questions to be answered and new improvements to be made. You'll probably devise many new inventions!

Electrostatics is also a subject that everyone is absolutely fascinated by – even those who can't change flashlight batteries,. When I was experimenting with different friction generator materials, I purchased quite a number of little swatches of material from a local fabric store. The young ladies working in the store demanded to know what the fabric was for, and, were so fascinated by the idea of generating a spark that I emerged from the store piled high with swatches of material which they insisted be tried on the generator.

My own mother, who despises anything technical, shocked me to no end by demanding that I explain to her the workings of the atom immediately after seeing the classic Van De Graaff in operation. I do not think anyone can watch a powerful Van De Graaff generator at work and not be curious about the forces at work. (And you and I will always want control of those forces!)

So dig in, get to work! Mysteries await!

SELECTED BIBLIOGRAPHY

B.I. Bleaney & B. Bleaney, *Electricity and Magnetism.* 1957 p.762
Tricker, R.A.R., *The Contributions of Faraday & Maxwell to Electrical Science.* 1966
Moore, Arthur Dearth, *Electrostatics.* 1968 p.240
Simon and Schuster, *The Way Things Work.* 1967 p.590
R.A. Ford, *Tesla Coil Secrets.* 1985 p.74
Lodge, Oliver, Sir, *The Nature and Properties of Negative Electricity.* 1906
Adams, Charles K., *Nature's Electricity.* 1987 p.147
Sootin, Harry, *Experiments With Static Electricity.* 1969 p.85
Van Doren, Carl, *Benjamin Franklin.* 1938 p.845
Ammen, C. W., *The Electroplater's Handbook,* 1986, p. 213

ADDENDUM

Electroplating High Potential Terminals

As you have certainly realized by now, a tremendous amount of your generator's efficiency rests in the quality of your high potential terminal and in your ability to effectively store the charges produced. I think you will be very happy with the performance of any of the classic generators and storage devices described earlier. However, for those of you seeking to go beyond these boundaries of performance, I'd like to let you in on new area of experimentation: electroplating.

Electroplating is simply a process wherein metal salts such as copper sulfate or nickel ammonium sulfate dissolved in an acidic aqueous solution can be deposited as elemental metal on conductor by applying an electrical current. This means simply that any conductive object can be covered through the process of electrodeposition with a layer of metal!

Since electrostatic charges can be distributed quite nicely on even very thin conductive surfaces, this is a perfect system for making a flawless Leyden jar or a huge high potential terminal. I am currently working on creating very large, near-perfect high potential terminals by plating large plastic globes sold in many lighting stores for under $15.00!

Anything can be plated – plastic, glass, pvc, etc. The only trick is in making their surface conductive initially so that a current can be applied.

The basic procedure for producing a perfect Leyden jar is as follows.

Begin by making the glass jar extremely clean. No fingerprints, water, or oil of any kind can remain. Use masking tape to cover the top 20% of the jar to prevent plating in this area and to preserve a well-proportioned jar. Paint the exposed glass with a resin such as white shellac thinned with denatured alcohol, if necessary. I've even used a spray lacquer.

From here there are several routes that you can take. The easiest method is to take a finely powered metal such as iron or aluminum (something around 400 mesh) and sprinkle it on the sticky, wet shellac. Since these finely powdered metals will float in the air much like talcum powder, I usually put the metal powder and jar in a box and shake the two around around. Remember! These metals SHOULD NOT BE INHALED! Wear a safety mask, and if working with heavy metals, wearing protective clothing is a good idea. The metal powder should create a beautiful, even, light coating over the painted surface of the jar. Allow the paint to fully dry, and repeat the procedure on the inside of the jar. This is a little trickier, but not impossible.

At this point you can actually stop, and use the jar quite successfully. The tiny sandpaper edges of this powder metal coating will bleed off a stored charge when the jar is not in use. This makes a good natural safety feature.

The only drawback to the powdered-metal approach is that if you choose to copper plate it, the copper plating will be just as bumpy as the powdered metal surface itself. As a result, this method will always allow some leakage.

Another simple method of making the jar conductive is to cover the shellac with powdered graphite. This can be rubbed smoothly onto the surface when the shellac is partially dry and will provide a smoother surface which in turn will produce smoother plating. On the other hand, graphite surfaces often require considerably longer plating times and the process is less reliable.

Although both of the above methods work well and have merits, the very best way to make the jar or any non-conductor conductive is to allow the shellac to dry completely, then paint or dip it in a solution of

1/2	ounce	silver nitrate
4	ounces	distilled water
6	ounces	alcohol

While this is still wet, suspend the jar inside a larger jar or other acid-proof non-metallic vessel, and place a few lumps of potassium sulfite or iron pyrite in the bottom of the larger jar. Cover this with a solution made up of one part sulfuric acid and eight parts distilled water, and close the lid. You now have a jar covered with silver nitrate solution hanging above the potassium sulfite-acid solution in a sealed vessel. This will produce hydrogen sulfide fumes that will convert the nitrate to a sulfide which is a conductive coating for you jar. Within a

A glass jar for the Leyden jar capacitor is selected and thoroughly cleaned.

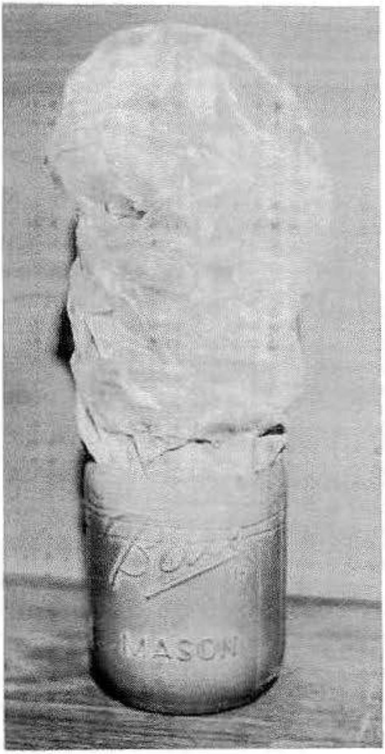

The jar is masked off and coated with conductive graphite.

few minutes the chemical reaction will stop, and your jar will be ready to be plated.

Now that the outside of the jar has been made conductive, it can easily be plated with copper. In an appropriate acid-proof vessel, make a solution of

25	ounces copper sulfate
6	ounces sulfuric acid
1	gallon distilled water

Use warm water when mixing to dissolve the copper sulfate more quickly. Do not allow this solution to touch your skin or clothes.

After a few hours, the deposit of copper is easily seen.

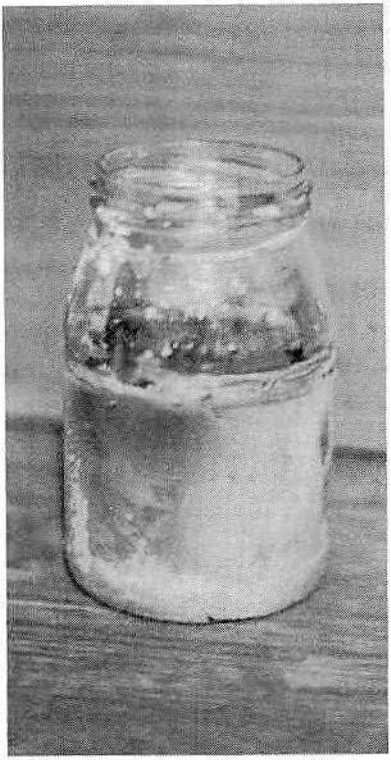

The plated jar is ready to be cleaned up and put into use.

Attach a wire from the negative lead of a 2 to 5 volt battery or transformer to the conductive part of your jar. A rubber band works well for this. Immerse the portion of your jar to be plated in the copper sulfate solution. Finally, attach a wire from the positive lead of your power source to a copper electrode having a large surface area such as a thin copper sheet. Set this electrode in your solution.

All that you have to do now is watch the jar as the plating takes place. Be sure you keep the solution warm, or the plating will proceed very slowly. I like a thick copper coat and usually allow my plating to take place overnight. However in a fairly short period of time you should see a reddish copper coating forming on the outside of your jar! Simply repeat this process for the inside of your jar.

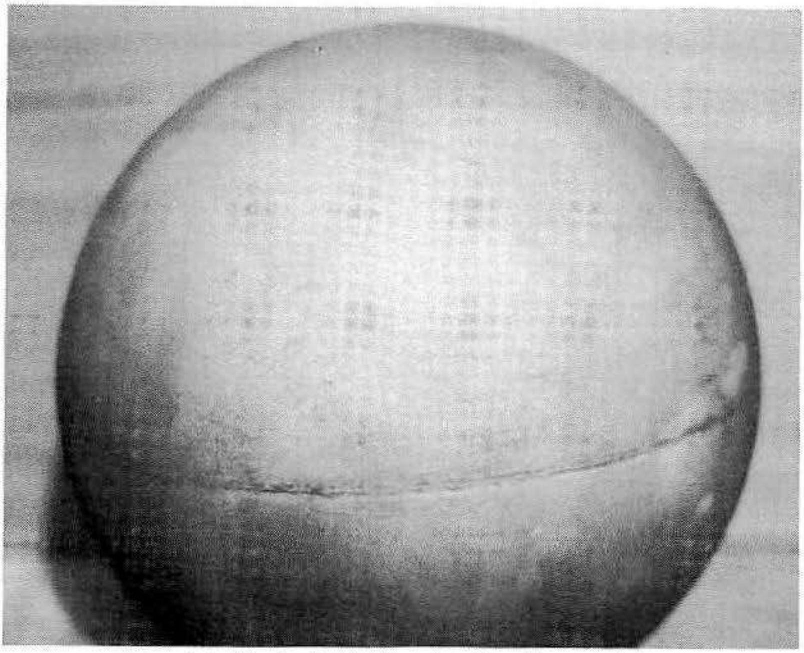

A twelve inch plastic sphere has been coated with graphite and is ready to be plated.

Very large spheres and other shapes can also be plated by scaling up this process. I have seen many inexpensive large glass and plastic globes since I began experimenting with this process that would lend themselves to this technique. For more information, refer to books on electroplating and electroforming that can give you advanced details on plating superior finishes.

As you can see, simple electroplating can make possible a vast number of new shapes and designs for high voltage terminals!

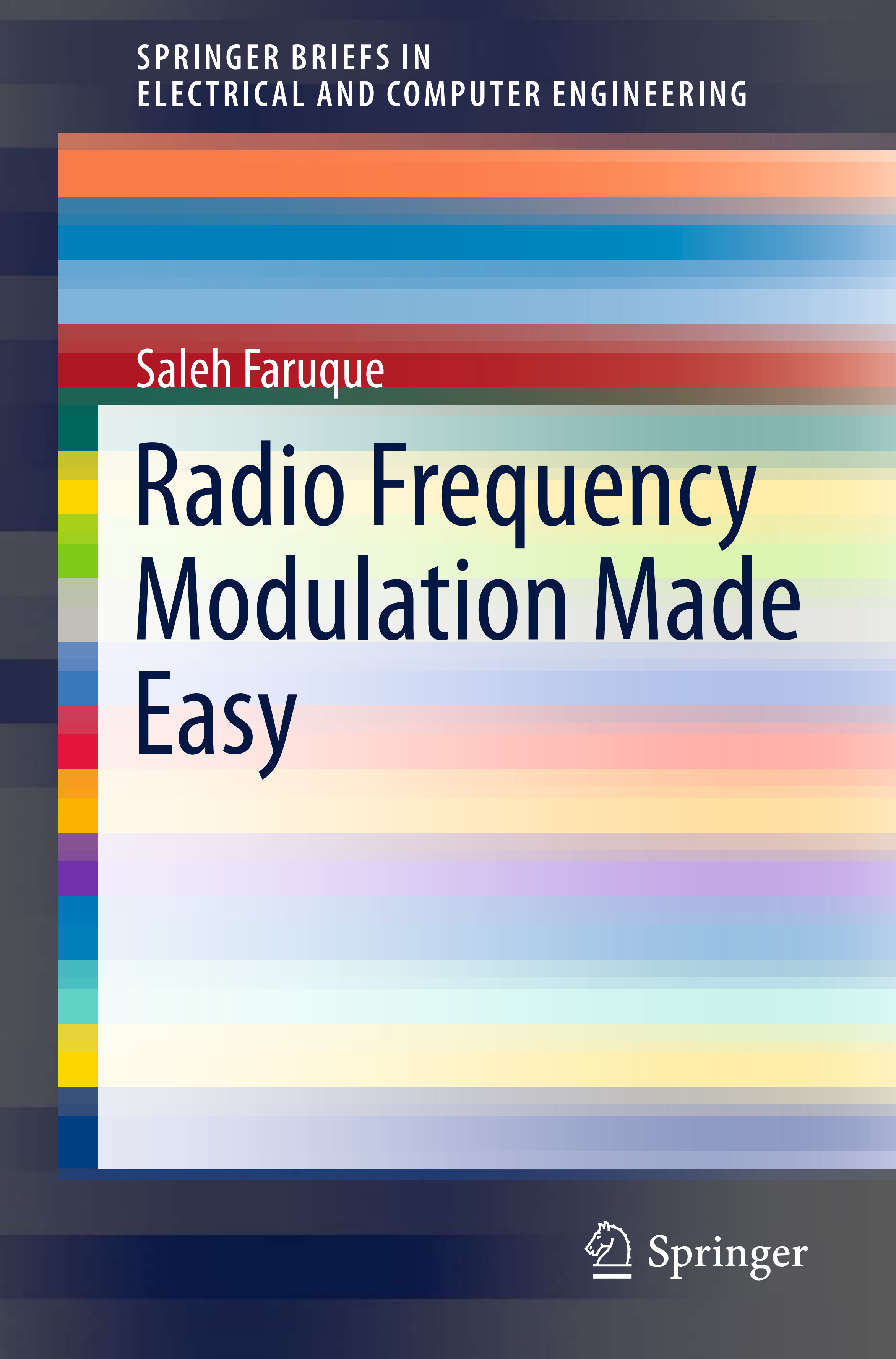

SPRINGER BRIEFS IN
ELECTRICAL AND COMPUTER ENGINEERING

Saleh Faruque

Radio Frequency Modulation Made Easy

Springer

SpringerBriefs in Electrical and Computer Engineering

More information about this series at http://www.springer.com/series/10059

Saleh Faruque

Radio Frequency Modulation Made Easy

Springer

Saleh Faruque
Department of Electrical Engineering
University of North Dakota
Grand Forks, ND
USA

ISSN 2191-8112 ISSN 2191-8120 (electronic)
SpringerBriefs in Electrical and Computer Engineering
ISBN 978-3-319-41200-9 ISBN 978-3-319-41202-3 (eBook)
DOI 10.1007/978-3-319-41202-3

Library of Congress Control Number: 2016945147

© The Author(s) 2017
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

By inventing the wireless transmitter or radio in 1897, the Italian physicist Tomaso Guglielmo Marconi added a new dimension to the world of communications. This enabled the transmission of the human voice through space without wires. For this epoch-making invention, this illustrious scientist was honored with the Nobel Prize for Physics in 1909. Even today, students of wireless or radio technology remember this distinguished physicist with reverence. A new era began in Radio Communications.

The classical Marconi radio used a modulation technique known today as "Amplitude Modulation" or just AM. This led to the development of Frequency Modulation (FM), amplitude shift keying (ASK), phase shift keying (PSK), etc. Today, these technologies are extensively used in various wireless communication systems. These modulation techniques form an integral part of academic curricula today.

This book presents a comprehensive overview of the various modulation techniques mentioned above. Numerous illustrations are used to bring students up-to-date in key concepts and underlying principles of various analog and digital modulation techniques. In particular, the following topics will be presented in this book:

- Amplitude Modulation (AM)
- Frequency Modulation (FM)
- Bandwidth occupancy in AM and FM
- Amplitude shift keying (ASK)
- Frequency shift keying (FSK)
- Phase shift keying (PSK)
- N-ary coding and M-ary modulation
- Bandwidth occupancy in ASK, FSK, and PSK

This text has been primarily designed for electrical engineering students in the area of telecommunications. However, engineers and designers working in the area of wireless communications would also find this text useful. It is assumed that the student is familiar with the general theory of telecommunications.

In closing, I would like to say a few words about how this book was conceived. It came out of my long industrial and academic career. During my teaching tenure at the University of North Dakota, I developed a number of graduate-level elective courses in the area of telecommunications that combine theory and practice. This book is a collection of my courseware and research activities in wireless communications.

I am grateful to UND and the School for the Blind, North Dakota, for affording me this opportunity. This book would never have seen the light of day had UND and the State of North Dakota not provided me with the technology to do so. My heartfelt salute goes out to the dedicated developers of these technologies, who have enabled me and others visually impaired to work comfortably.

I would like to thank my beloved wife, Yasmin, an English Literature buff and a writer herself, for being by my side throughout the writing of this book and for patiently proofreading it. My darling son, Shams, an electrical engineer himself, provided technical support in formulation and experimentation when I needed it. For this, he deserves my heartfelt thanks.

Finally, thanks are also to my doctoral student Md. Maruf Ahamed who found time in his busy schedule to assist me with the simulations, illustrations, and the verification of equations.

In spite of all this support, there may still be some errors in this book. I hope that my readers forgive me for them. I shall be amply rewarded if they still find this book useful.

Grand Forks, USA Saleh Faruque
May 2016

Contents

1 **Introduction to Modulation** 1
  1.1 Background 1
  1.2 Modulation by Analog Signals 3
    1.2.1 AM, FM, and PM 3
    1.2.2 AM and FM Bandwidth at a Glance 4
  1.3 Modulation by Digital Signal 5
    1.3.1 Amplitude Shift Keying (ASK) Modulation 5
    1.3.2 Frequency Shift Keying (FSK) Modulation. 6
    1.3.3 Phase Shift Keying (PSK) Modulation 7
  1.4 Bandwidth Occupancy in Digital Modulation 7
    1.4.1 Spectral Response of the Encoded Data 8
    1.4.2 Spectral Response of the Carrier Frequency Before Modulation. 9
    1.4.3 ASK Bandwidth at a Glance. 10
    1.4.4 FSK Bandwidth at a Glance 11
    1.4.5 BPSK Bandwidth at a Glance. 12
  1.5 Conclusions 14
  References 15

2 **Amplitude Modulation (AM)** 17
  2.1 Introduction 17
  2.2 Amplitude Modulation 19
  2.3 AM Spectrum and Bandwidth 20
    2.3.1 Spectral Response of the Input Modulating Signal. 20
    2.3.2 Spectral Response of the Carrier Frequency 21
    2.3.3 AM Spectrum and Bandwidth. 21
    2.3.4 AM Response Due to Low and High Modulating Signals 23
    2.3.5 AM Demodulation 24
    2.3.6 Drawbacks in AM. 24

2.4 Double Sideband-Suppressed Carrier (DSBSC) 25
  2.4.1 DSBSC Modulation. 25
  2.4.2 Generation of DSBSC Signal 26
  2.4.3 DSBSC Spectrum and Bandwidth 27
  2.4.4 DSBSC Drawbacks 28
2.5 Single Sideband (SSB) Modulation 29
  2.5.1 Why SSB Modulation? 29
  2.5.2 Generation of SSB-Modulated Signal. 29
  2.5.3 SSB Spectrum and Bandwidth 30
2.6 Conclusions 32
References 32

3 Frequency Modulation (FM) 33
3.1 Introduction 33
3.2 Frequency Modulation (FM). 34
  3.2.1 Background 34
  3.2.2 The Basic FM 35
3.3 FM Spectrum and Bandwidth 37
  3.3.1 Spectral Response of the Input Modulating Signal. 37
  3.3.2 Spectral Response of the Carrier Frequency 38
  3.3.3 FM Spectrum 39
  3.3.4 Carson's Rule and FM Bandwidth. 40
  3.3.5 Bessel Function and FM Bandwidth 41
  3.3.6 FM Bandwidth Dilemma 42
3.4 Conclusions 44
References 44

4 Amplitude Shift Keying (ASK). 45
4.1 Introduction 45
4.2 ASK Modulation. 46
4.3 ASK Demodulation 48
4.4 ASK Bandwidth 49
  4.4.1 Spectral Response of the Encoded Data 49
  4.4.2 Spectral Response of the Carrier Frequency Before Modulation. 51
  4.4.3 ASK Bandwidth at a Glance. 51
4.5 BER Performance 53
4.6 Conclusions 54
References 55

5 Frequency Shift Keying (FSK) 57
5.1 Introduction 57
5.2 Frequency Shift Keying (FSK) Modulation. 58
5.3 Frequency Shift Keying (FSK) Demodulation 60

- 5.4 FSK Bandwidth 61
  - 5.4.1 Spectral Response of the Encoded Data 61
  - 5.4.2 Spectral Response of the Carrier Frequency Before Modulation 63
  - 5.4.3 FSK Bandwidth at a Glance 63
- 5.5 BER Performance 65
- 5.6 Conclusions 67
- References 67

6 Phase Shift Keying (PSK) 69
- 6.1 Introduction 69
- 6.2 Binary Phase Shift Keying (BPSK) 70
  - 6.2.1 BPSK Modulation 70
  - 6.2.2 BPSK Demodulation 73
- 6.3 QPSK Modulation 74
- 6.4 8PSK Modulation 75
- 6.5 16PSK Modulation 75
- 6.6 PSK Spectrum and Bandwidth 77
  - 6.6.1 Spectral Response of the Encoded Data 77
  - 6.6.2 Spectral Response of the Carrier Before Modulation 79
  - 6.6.3 BPSK Spectrum 79
- 6.7 Conclusions 82
- References 82

7 N-Ary Coded Modulation 85
- 7.1 Introduction 85
- 7.2 N-Ary Convolutional Coding and M-Ary Modulation 86
  - 7.2.1 Background 86
  - 7.2.2 Generation of Complementary Convolutional Codes 86
  - 7.2.3 2-Ary Convolutional Coding with QPSK Modulation 88
  - 7.2.4 4-Ary Convolutional Coding with 16PSK Modulation 89
- 7.3 N-Ary Convolutional Decoder 91
  - 7.3.1 Correlation Receiver 91
  - 7.3.2 Error Correction Capabilities of N-Ary Convolutional Codes 93
- 7.4 N-Ary Orthogonal Coding and M-Ary Modulation 94
  - 7.4.1 Background 94
  - 7.4.2 Orthogonal Codes 95
  - 7.4.3 2-Ary Orthogonal Coding with QPSK Modulation 95
  - 7.4.4 4-Ary Orthogonal Coding with 16PSK Modulation 97
  - 7.4.5 2-Ary Orthogonal Decoding 97
  - 7.4.6 4-Ary Orthogonal Decoding 99
  - 7.4.7 Error Correction Capabilities of N-Ary Orthogonal Codes 99
- 7.5 Conclusions 103
- References 104

Chapter 1
Introduction to Modulation

Topics

- Background
- Modulation by Analog Signal
- AM and FM Bandwidth at a Glance
- Modulation by Digital Signal
- ASK, FSK and PSK Bandwidth at a Glance

1.1 Background

Modulation is a technique that changes the characteristics of the carrier frequency in accordance with the input signal. Figure 1.1 shows the conceptual block diagram of a modern wireless communication system, where the modulation block is shown in the inset of the dotted block. As shown in the figure, modulation is performed at the transmit side and demodulation is performed at the receive side. This is the final stage of any radio communication system. The preceding two stages have been discussed elaborately in my previous book in this series [1, 2].

The output signal of the modulator, referred to as the modulated signal, is fed into the antenna for propagation. Antenna is a reciprocal device that transmits and receives the modulated carrier frequency. The size of the antenna depends on the wavelength (λ) of the sinusoidal wave where

$\lambda = c/f$ m
c = Velocity of light = 3×10^8 m/s
f = Frequency of the sinusoidal wave, also known as "carrier frequency."

Therefore, a carrier frequency much higher than the input signal is required to keep the size of the antenna at an acceptable limit. For these reasons, a high-frequency carrier signal is used in the modulation process. In this process, the

© The Author(s) 2017

S. Faruque, *Radio Frequency Modulation Made Easy*,
SpringerBriefs in Electrical and Computer Engineering,
DOI 10.1007/978-3-319-41202-3_1

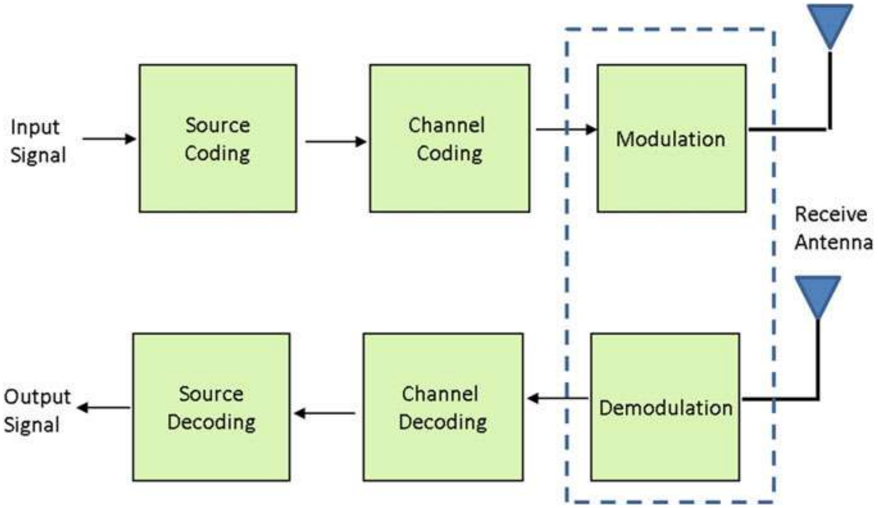

Fig. 1.1 Block diagram of a modern full-duplex communication system. The modulation stage is shown as a *dotted* block

low-frequency input signal changes the characteristics of the high-frequency carrier in a certain manner, depending on the modulation technique. Furthermore, as the size and speed of digital data networks continue to expand, bandwidth efficiency becomes increasingly important. This is especially true for broadband communication, where the digital signal processing is done keeping in mind the available bandwidth resources.

Hence, modulation is a very important step in the transmission of information. The information can be either analog or digital, where the carrier is a high-frequency sinusoidal waveform. As stated earlier, the input signal (analog or digital) changes the characteristics of the carrier waveform. Therefore, there are two basic modulation schemes as listed below:

- Modulation by analog signals and
- Modulation by digital signals.

This book presents a comprehensive overview of these modulation techniques in use today. Numerous illustrations are used to bring students up-to-date in key concepts and underlying principles of various analog and digital modulation techniques. For a head start, brief descriptions of each of these modulation techniques are presented below.

1.2 Modulation by Analog Signals

1.2.1 AM, FM, and PM

For analog signals, there are three well-known modulation techniques as listed below:

- Amplitude Modulation (AM),
- Frequency Modulation (FM),
- Phase Modulation (PM).

By inventing the wireless transmitter or radio in 1897, the Italian physicist Tomaso Guglielmo Marconi added a new dimension to the world of communications [3, 4]. This enabled the transmission of the human voice through space without wires. For this epoch-making invention, this illustrious scientist was honored with the Nobel Prize for Physics in 1909. Even today, students of wireless or radio technology remember this distinguished physicist with reverence. A new era began in Radio Communications. The classical Marconi radio used a modulation technique known today as "Amplitude Modulation" or just AM. In AM, the amplitude of the carrier changes in accordance with the input analog signal, while the frequency of the carrier remains the same. This is shown in Fig. 1.2 where

- *m*(*t*) is the input modulating audio signal,
- *C*(*t*) is the carrier frequency, and
- *S*(*t*) is the AM-modulated carrier frequency.

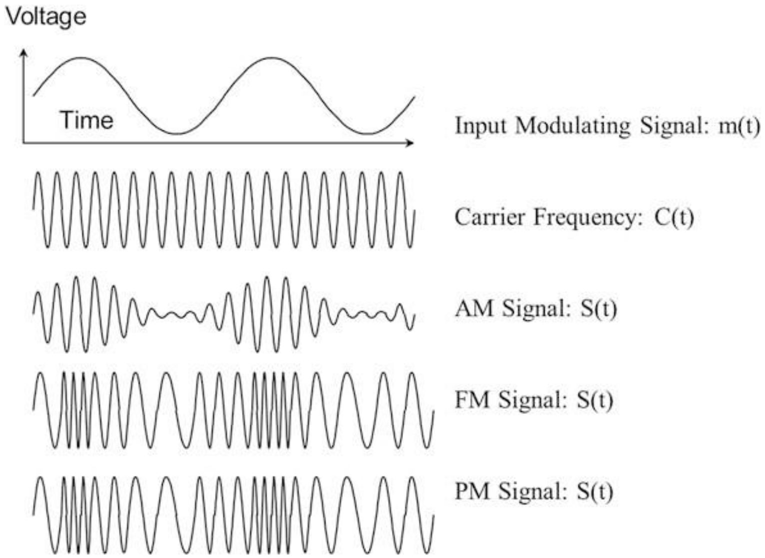

Fig. 1.2 Modulation by analog signal

As shown in the figure, the audio waveform changes the amplitude of the carrier to determine the envelope of the modulated carrier. This enables the receiver to extract the audio signal by demodulation. Notice that the amplitude of the carrier changes in accordance with the input signal, while the frequency of the carrier does not change after modulation. It can be shown that the modulated carrier $S(t)$ contains several spectral components, requiring frequency-domain analysis, which will be addressed in Chap. 2. It may be noted that AM is vulnerable to signal amplitude fading.

In Frequency Modulation (FM), the frequency of the carrier changes in accordance with the input modulation signal as shown in Fig. 1.2 [5]. Notice that in FM, only the frequency changes while the amplitude remains the same. Unlike AM, FM is more robust against signal amplitude fading. For this reason, FM is more attractive in commercial FM radio. In Chap. 3, it will be shown that in FM, the modulated carrier contains an infinite number of sideband due to modulation. For this reason, FM is also bandwidth inefficient.

Similarly, in Phase Modulation (PM), the phase of the carrier changes in accordance with the phase of the carrier, while the amplitude of the carrier does not change. PM is closely related to FM. In fact, FM is derived from the rate of change of phase of the carrier frequency. Both FM and PM belong to the same mathematical family. We will discuss this more elaborately in Chap. 3.

1.2.2 AM and FM Bandwidth at a Glance

The bandwidth occupied by the modulated signal depends on bandwidth of the input signal and the modulation method as shown in Fig. 1.3. Note that the unmodulated carrier itself has zero bandwidth.

In AM:

- The modulated carrier has two sidebands (upper and lower) and
- Total bandwidth = 2 × base band.

In FM:

- The carrier frequency shifts back and forth from the nominal frequency by Δf, where Δf is the frequency deviation.
- During this process, the modulated carrier creates an infinite number of spectral components, where higher-order spectral components are negligible.
- The approximate FM bandwidth is given by the Carson's rule:
  - FM BW = $2f\,(1 + \beta)$
  - f = Base band frequency
  - β = Modulation index
  - $\beta = \Delta f/f$
  - Δf = Frequency deviation.

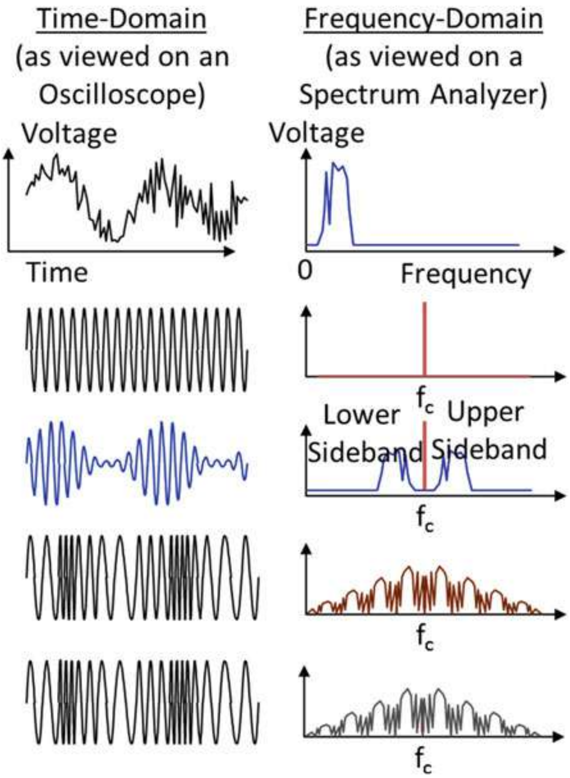
Time-Domain
(as viewed on an
Oscilloscope)
Voltage
Time
Frequency-Domain
(as viewed on a
Spectrum Analyzer)
Voltage
0
Frequency
f_c
Lower
Sideband
Upper
Sideband
f_c
f_c
f_c

Fig. 1.3 Bandwidth occupancy in AM, FM, and PM signals

1.3 Modulation by Digital Signal

For digital signals, there are several modulation techniques available. The three main digital modulation techniques are as follows:

- Amplitude shift keying (ASK),
- Frequency shift keying (FSK), and
- Phase shift keying (PSK).

Figure 1.4 illustrates the modulated waveforms for an input modulating digital signal. Brief descriptions of each of these digital modulation techniques along with the respective spectral responses and bandwidth are presented below.

1.3.1 Amplitude Shift Keying (ASK) Modulation

Amplitude shift keying (ASK), also known as on–off keying (OOK), is a method of digital modulation that utilizes amplitude shifting of the relative amplitude of the

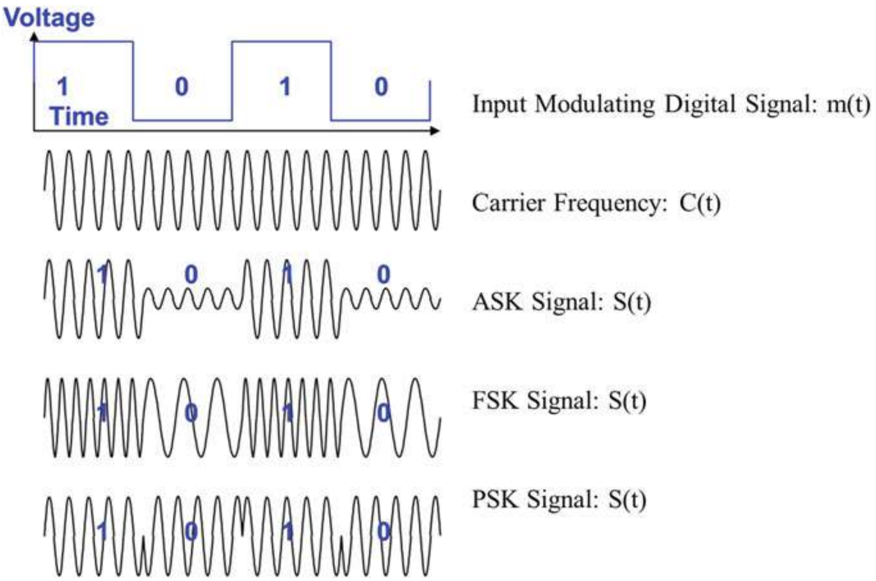
Voltage
1
0
1
0
Time
Input Modulating Digital Signal: m(t)
Carrier Frequency: C(t)
1
0
1
0
ASK Signal: S(t)
1
0
1
0
FSK Signal: S(t)
1
0
1
0
PSK Signal: S(t)

Fig. 1.4 Modulation by digital signal

career frequency [6–8]. The signal to be modulated and transmitted is binary; this is referred to as ASK, where the amplitude of the carrier changes in discrete levels, in accordance with the input signal, as shown.

- Binary 0 (bit 0): Amplitude = Low and
- Binary 1 (bit 1): Amplitude = High.

Figure 1.4 shows the ASK-modulated waveform where

- Input digital signal is the information we want to transmit.
- Carrier is the radio frequency without modulation.
- Output is the ASK-modulated carrier, which has two amplitudes corresponding to the binary input signal. For binary signal 1, the carrier is ON. For the binary signal 0, the carrier is OFF. However, a small residual signal may remain due to noise, interference, etc.

1.3.2 Frequency Shift Keying (FSK) Modulation

Frequency shift keying (FSK) is a method of digital modulation that utilizes frequency shifting of the relative frequency content of the signal [6–8]. The signal to be modulated and transmitted is binary; this is referred to as binary FSK (BFSK), where the carrier frequency changes in discrete levels, in accordance with the input signal as shown below:

- Binary 0 (bit 0): Frequency = $f + \Delta f$.
- Binary 1 (bit 1): Frequency = $f - \Delta f$.

Figure 1.4 shows the FSK-modulated waveform where

- Input digital signal is the information we want to transmit.
- Carrier is the radio frequency without modulation.
- Output is the FSK-modulated carrier, which has two frequencies ω_1 and ω_2, corresponding to the binary input signal.
- These frequencies correspond to the messages binary 0 and 1, respectively.

1.3.3 Phase Shift Keying (PSK) Modulation

Phase shift keying (PSK) is a method of digital modulation that utilizes phase of the carrier to represent digital signal [6–8]. The signal to be modulated and transmitted is binary; this is referred to as binary PSK (BPSK), where the phase of the carrier changes in discrete levels, in accordance with the input signal as shown below:

- Binary 0 (bit 0): $\text{Phase}_1 = 0°$.
- Binary 1 (bit 1): $\text{Phase}_2 = 180°$.

Figure 1.4 shows the modulated waveform where

- Input digital signal is the information we want to transmit.
- Carrier is the radio frequency without modulation.
- Output is the BPSK-modulated carrier, which has two phases φ_1 and φ_2 corresponding to the two information bits.

1.4 Bandwidth Occupancy in Digital Modulation

In wireless communications, the scarcity of RF spectrum is well known. For this reason, we have to be vigilant about using transmission bandwidth. The transmission bandwidth depends on the following:

- Spectral response of the encoded data,
- Spectral response of the carrier frequency, and
- Modulation type (ASK, FSK, PSK), etc.

Let us take a closer look!

1.4.1 Spectral Response of the Encoded Data

In digital communications, data is generally referred to as a non-periodic digital signal. It has two values:

- Binary-1 = High, Period = T.
- Binary-0 = Low, Period = T.

Also, data can be represented in two ways:

- Time-domain representation and
- Frequency-domain representation.

The time-domain representation (Fig. 1.5a), known as non-return-to-zero (NRZ), is given by

$$\begin{aligned} V(t) &= V \qquad &< 0 < t < T \\ &= 0 \qquad &\text{elsewhere} \end{aligned} \tag{1.1}$$

The frequency-domain representation is given by "Fourier transform" [9]:

$$V(\omega) = \int_0^T V \cdot e^{-j\omega t} dt \tag{1.2}$$

$$|V(\omega)| = VT\left[\frac{\sin(\omega T/2)}{\omega T/2}\right]$$

$$P(\omega) = \left(\frac{1}{T}\right)|V(\omega)|^2 = V^2 T\left[\frac{\sin(\omega T/2)}{\omega T/2}\right]^2 \tag{1.3}$$

Here, $P(\omega)$ is the power spectral density. This is plotted in (Fig. 1.5b). The main lobe corresponds to the fundamental frequency and side lobes correspond to harmonic components. The bandwidth of the power spectrum is proportional to the frequency. In practice, the side lobes are filtered out since they are relatively insignificant with respect to the main lobe. Therefore, the one-sided bandwidth is given by the ratio $f/fb = 1$. In other words, the one-sided bandwidth $= f = f_b$, where $f_b = R_b = 1/T$, T being the bit duration.

The general equation for two-sided response is given by

$$V(\omega) = \int_{-\alpha}^{\alpha} V(t) \cdot e^{-j\omega t} \mathrm{d}t$$

In this case, $V(\omega)$ is called two-sided spectrum of $V(t)$. This is due to both positive and negative frequencies used in the integral. The function can be a voltage

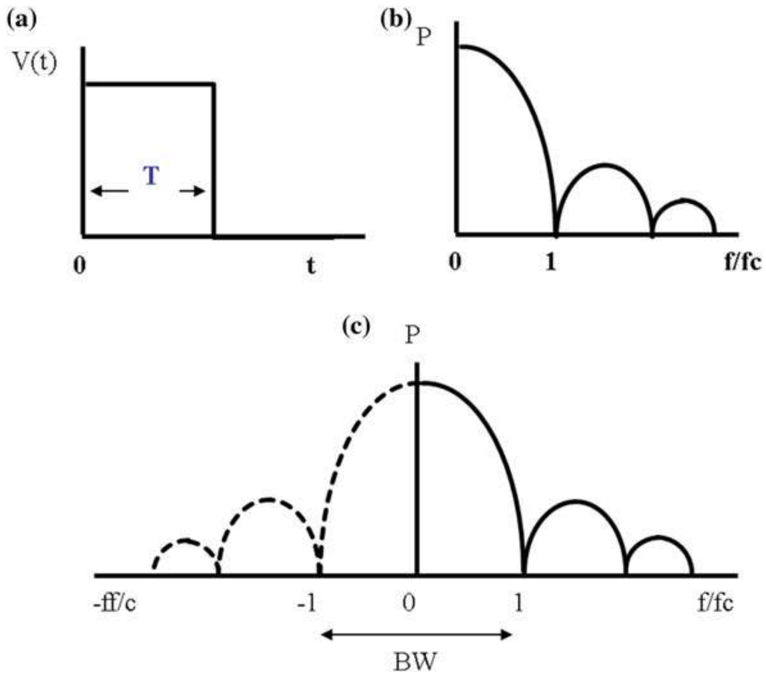

Fig. 1.5 **a** Discrete time digital signal, **b** it is one-sided power spectral density, and **c** two-sided power spectral density. The bandwidth associated with the non-return-to-zero (NRz) data is $2R_b$, where R_b is the bit rate

or a current (Fig. 1.5c) shows the two-sided response, where the bandwidth is determined by the main lobe as shown below:

$$\text{Two sided bandwidth (BW)} = 2R_b \; (R_b = \text{Bit rate before coding}) \tag{1.4}$$

1.4.2 *Spectral Response of the Carrier Frequency Before Modulation*

A carrier frequency is essentially a sinusoidal waveform, which is periodic and continuous with respect to time. It has one frequency component. For example, the sine wave is described by the following time-domain equation:

$$V(t) = V_p \sin(\omega t_c) \tag{1.5}$$

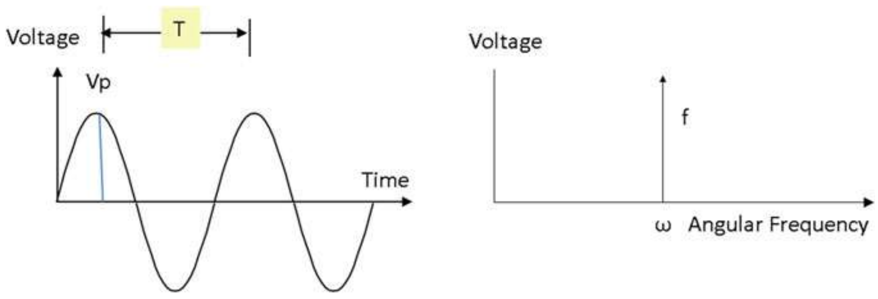

Fig. 1.6 A sine wave and its frequency response

where

$$V_p = \text{Peak voltage}$$

- $\omega_c = 2\pi f_c$
- f_c = Carrier frequency in Hz

Figure 1.6 shows the characteristics of a sine wave and its spectral response. Since the frequency is constant, its spectral response is located in the horizontal axis and the peak voltage is shown in the vertical axis. The corresponding bandwidth is zero.

1.4.3 ASK Bandwidth at a Glance

In ASK, the amplitude of the carriers changes in discrete levels, in accordance with the input signal where

- Input data: $m(t) = 0$ or 1.
- Carrier frequency: $C(t) = A_c \cos(\omega_c t)$.
- Modulated carrier: $S(t) = m(t)C(t) = m(t)A_c \cos(\omega_c t)$.

Since $m(t)$ is the input digital signal and it contains an infinite number of harmonically related sinusoidal waveforms and that we keep the fundamental and filter out the higher-order components, we write:

$$m(t) = A_m \sin(\omega_m t)$$

The ASK-modulated signal then becomes:

$$\begin{aligned} S(t) &= m(t)S(t) = A_m A_c \sin(\omega_m t)\cos(\omega_c) \\ &= A_m A_c \cos\left(\omega_c \pm \omega_m\right) \end{aligned}$$

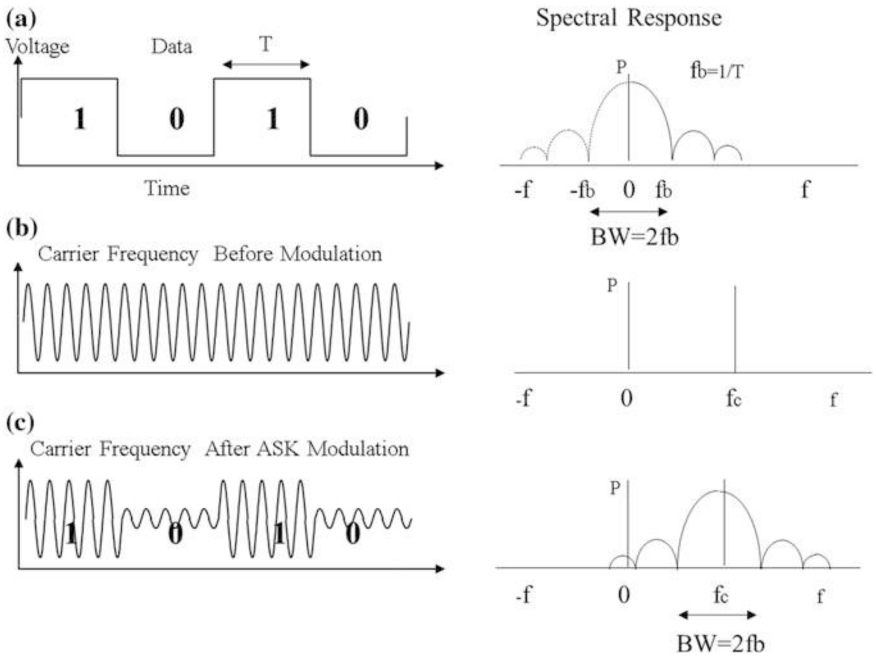
(a)
Voltage
Data
T
1
0
1
0
Time
Spectral Response
P
fb=1/T
-f
-fb
0
fb
f
BW=2fb
(b)
Carrier Frequency Before Modulation
P
-f
0
fc
f
(c)
Carrier Frequency After ASK Modulation
1
0
1
0
P
-f
0
fc
f
BW=2fb

Fig. 1.7 ASK bandwidth at a glance. **a** Spectral response of NRZ data before modulation. **b** Spectral response of the carrier before modulation. **c** Spectral response of the carrier after modulation. The transmission bandwidth is $2f_b$, where f_b is the bit rate and $T = 1/f_b$ is the bit duration for NRZ data

The spectral response is depicted in Fig. 1.7. Notice that the spectral response after ASK modulation is the shifted version of the NRZ data. Bandwidth is given by,

BW = $2R_b$ (coded), where R_b is the coded bit rate.

1.4.4 FSK Bandwidth at a Glance

In FSK, the frequency of the carrier changes in two discrete levels, in accordance with the input signals. We have:

- Input data: $m(t) = 0$ or 1
- Carrier frequency: $C\ (t) = A\ \cos\ (\omega t)$
- Modulated carrier: $S(t) = A\ \cos(\omega - \Delta\ \omega)t$, For $m(t) = 1$
  $S(t) = A\ \cos(\omega + \Delta\omega)t$, For $m(t) = 0$

where

- $S(t)$ = The modulated carrier,
- A = Amplitude of the carrier,

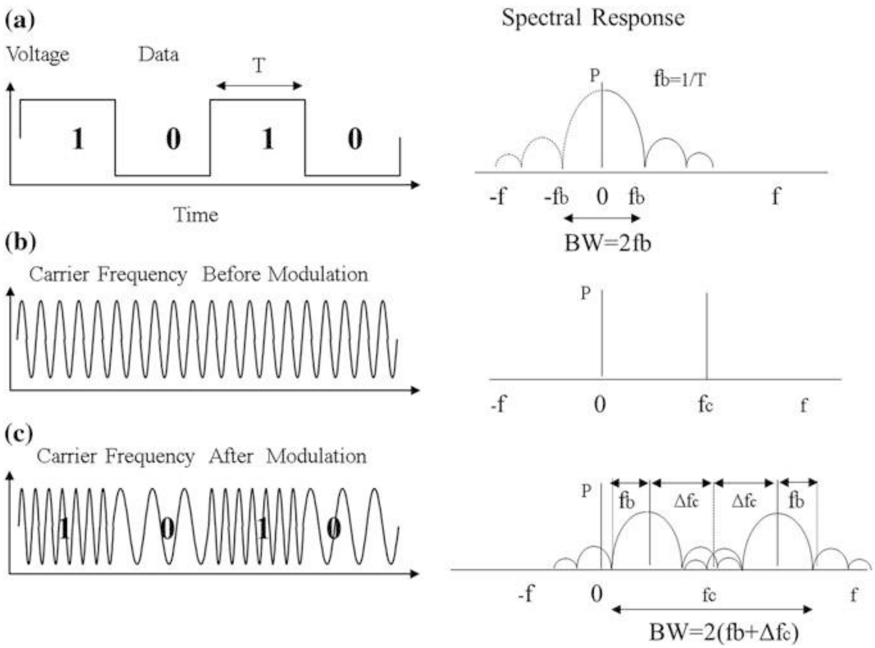
(a)
Voltage
Data
T
1
0
1
0
Time
Spectral Response
P
fb=1/T
-f
-fb
0
fb
f
BW=2fb
(b)
Carrier Frequency Before Modulation
P
-f
0
fc
f
(c)
Carrier Frequency After Modulation
1
0
1
0
P
fb
Δfc
Δfc
fb
-f
0
fc
f
BW=2(fb+Δfc)

Fig. 1.8 FSK bandwidth at a glance. **a** Spectral response of NRZ data before modulation. **b** Spectral response of the carrier before modulation. **c** Spectral response of the carrier after modulation. The transmission bandwidth is $2(f_b + \Delta f_c)$. f_b is the bit rate and Δf_c is the frequency deviation = $1/f_b$ is the bit duration for NRZ data

- ω = Nominal frequency of the carrier frequency, and
- $\Delta\omega$ = Frequency deviation.

The spectral response is depicted in Fig. 1.8. Notice that the carrier frequency after FSK modulation varies back and forth from the nominal frequency f_c by $\pm\,\Delta f_c$, where Δf_c is the frequency deviation. The FSK bandwidth is given by

$$\mathrm{BW} = 2(f_b + \Delta f_c) = 2f_b(1 + \Delta f_c/f_b) = 2f_b(1 + \beta),$$

where $\beta = \Delta f/f_b$ is known as the modulation index and f_b is the coded bit frequency (bit rate R_b).

1.4.5 BPSK Bandwidth at a Glance

In BPSK, the phase of the carrier changes in two discrete levels, in accordance with the input signal. Here, we have:

- Input data: $m(t) = 0$ or 1
- Carrier frequency: $C(t) = A \cos(\omega t)$
- Modulated carrier: $S(t) = A \cos(\omega + \varphi)t$

where

- A = Amplitude of the carrier frequency,
- ω = Angular frequency of the carrier, and
- φ = Phase of the carrier frequency.

Table below shows the number of phases and the corresponding bits per phase for MPSK modulation schemes for $M = 2, 4, 8, 16, 32, 64$, etc. It will be shown that higher-order MPSK modulation schemes ($M > 2$) are spectrally efficient.

Modulation	Number of phases φ	Number of bits per phase
BPSK	2	1
QPSK	4	2
8PSK	8	3
16	16	4
32	32	5
64	64	6
:	:	:

Figure 1.9 shows the spectral response of the BPSK modulator. Since there are two phases, the carrier frequency changes in two discrete levels, one bit per phase, as follows:

$\varphi = 0°$ for bit 0 and
$\varphi = 180°$ for bit 1.

Notice that the spectral response after BPSK modulation is the shifted version of the NRZ data, centered on the carrier frequency f_c. The transmission bandwidth is given by

$$\text{BW(BPSK)} = 2R_b/\text{Bit per Phase} = 2R_b/1 = 2R_b$$

where

- R_b is the coded bit rate (bit frequency).
- For BPSK, $\varphi = 2$, one bit per phase.

Also, notice that the BPSK bandwidth is the same as the one in ASK modulation. This is due to the fact that the phase of the carrier changes in two discrete levels, while the frequency remains the same.

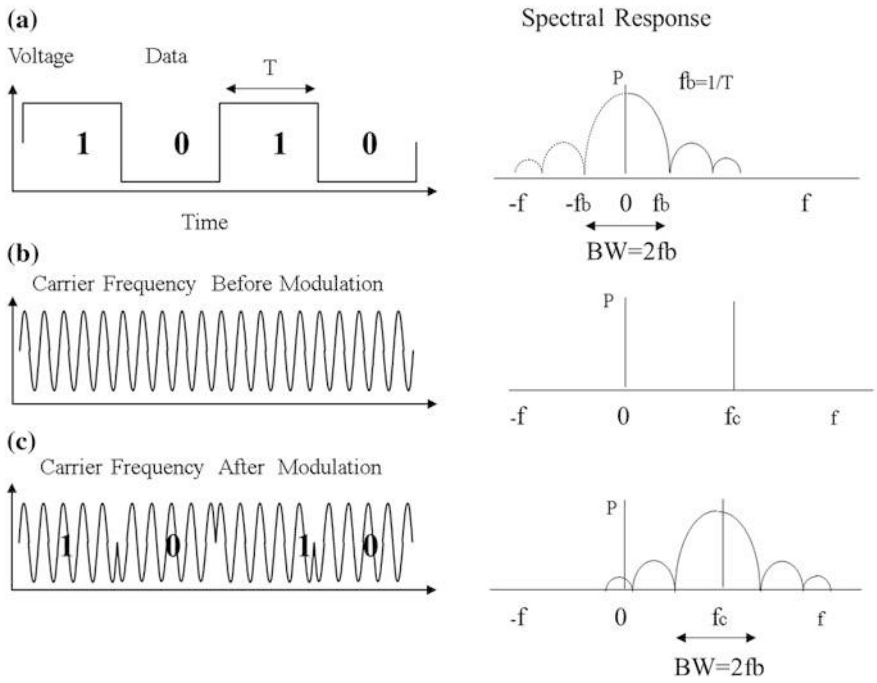
(a)
Voltage
Data
T
1
0
1
0
Time
Spectral Response
P
fb=1/T
-f
-fb
0
fb
f
BW=2fb
(b)
Carrier Frequency Before Modulation
P
-f
0
fc
f
(c)
Carrier Frequency After Modulation
1
0
1
0
P
-f
0
fc
f
BW=2fb

Fig. 1.9 PSK bandwidth at a glance. **a** Spectral response of NRZ data before modulation. **b** Spectral response of the carrier before modulation. **c** Spectral response of the carrier after modulation

1.5 Conclusions

This chapter presents a brief overview of modulation techniques covered in this book. Numerous illustrations are used to bring students up-to-date in key concepts and underlying principles of various analog and digital modulation techniques. In particular, following topics will be presented in this book:

- Amplitude Modulation (AM),
- Frequency Modulation (FM),
- Bandwidth occupancy in AM and FM,
- Amplitude shift keying (ASK),
- Frequency shift keying (FSK),
- Phase shift keying (PSK), and
- Bandwidth occupancy in ASK, FSK, and PSK.

References

1. Faruque S (2014) Radio frequency source coding made easy. Springer, New York
2. Faruque S (2014) Radio frequency channel coding made easy. Springer, New York
3. Marconi G (1987) Improvements in transmitting electrical impulses and signals, and in apparatus therefor. British patent No. 12,039 . Date of Application 2 June 1896; Complete Specification Left, 2 Mar 1897; Accepted, 2 July 1897 (later claimed by Oliver Lodge to contain his own ideas which he failed to patent)
4. Marconi G (1900) Improvements in apparatus for wireless telegraphy. British patent No. 7,777. Date of Application 26 Apr 1900; Complete Specification Left, 25 Feb 1901; Accepted, 13 Apr 1901
5. Armstrong EH (1936) A Method of reducing disturbances in radio signaling by a system of frequency modulation. Proc IRE (IRE) 24(5):689–740. doi:10.1109/JRPROC.1936.227383
6. Smith DR (1985) Digital transmission system. Van Nostrand Reinhold Co. ISBN: 0442009178
7. Leon W, Couch II (2001) Digital and analog communication systems, 7th edn. Prentice-Hall Inc, Englewood Cliffs. ISBN 0-13-142492-0
8. Sklar B (1988) Digital communications fundamentals and applications. Prentice Hall, Upper Saddle River
9. Joseph Fourier JB (1878) The analytical theory of heat (trans: Freeman A). The University Press, London

Chapter 2
Amplitude Modulation (AM)

Topics

- Introduction
- Amplitude Modulation (AM)
- AM Spectrum and Bandwidth
- Double Side Band Suppressed Carrier (DSBSC)
- DSBSC Spectrum and Bandwidth
- Single Side Band (SSB) Carrier
- SSB Spectrum and Bandwidth

2.1 Introduction

By inventing the wireless transmitter or radio in 1897, the Italian physicist Tomaso Guglielmo Marconi added a new dimension to the world of communications [1, 2]. This enabled the transmission of the human voice through space without wires. For this epoch-making invention, this illustrious scientist was honored with the Nobel Prize for Physics in 1909. Even today, students of wireless or radio technologies remember this distinguished physicist with reverence. A new era began in Radio Communications.

The classical Marconi radio used a modulation technique known today as "Amplitude Modulation" or just AM, which is the main topic of this chapter. In AM, the amplitude of the carrier changes in accordance with the input analog signal, while the frequency of the carrier remains the same. This is shown in Fig. 2.1, where

- $m(t)$ is the input modulating audio signal,
- $C(t)$ is the carrier frequency, and
- $S(t)$ is the AM-modulated carrier frequency.

© The Author(s) 2017

S. Faruque, *Radio Frequency Modulation Made Easy*,
SpringerBriefs in Electrical and Computer Engineering,
DOI 10.1007/978-3-319-41202-3_2

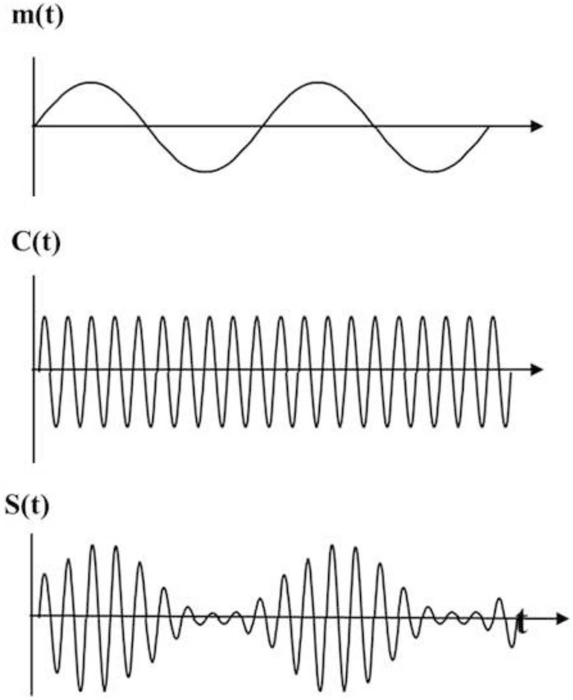

Fig. 2.1 AM waveforms. The amplitude of the carrier changes in accordance with the input analog signal. The frequency of the carrier remains the same

As shown in the figure, the audio waveform changes the amplitude of the carrier to determine the envelope of the modulated carrier. This enables the receiver to extract the audio signal by demodulation. Notice that the amplitude of the carrier changes in accordance with the input signal, while the frequency of the carrier does not change after modulation. However, it can be shown that the modulated carrier $S(t)$ contains several spectral components, requiring frequency domain analysis. In an effort to examine this, this chapter will present the following topics:

- Amplitude Modulation (AM) and AM spectrum
- Double Sideband-suppressed carrier (DSBSC) and DSBSC spectrum
- Single sideband (SSB) carrier and SSB spectrum

In the following sections, the above disciplines in AM modulation will be presented along with the respective spectrum and bandwidth. These materials have been augmented by diagrams and associated waveforms to make them easier for readers to grasp.

2.2 Amplitude Modulation

Amplitude Modulation (AM) is a method of analog modulation that utilizes amplitude variations of the relative amplitude of the career frequency [3–5]. The signal to be modulated and transmitted is analog. This is referred to as AM, where the amplitude of the carrier changes in accordance with the input signal.

Figure 2.2 shows a functional diagram of a typical AM modulator for a single tone. Here, $m(t)$ is the input analog signal we want to transmit, $C(t)$ is the carrier frequency without modulation, and $S(t)$ is the output AM-modulated carrier frequency. These parameters are described below:

$$\begin{aligned} &\bullet\ m(t) = A_m \cos(2\pi f_m t) \\ &\bullet\ C(t) = A_c \cos(2\pi f_c t) f_c \gg f_m \\ &\bullet\ \begin{aligned} S(t) &= [1 + m(t)]C(t) \\ &= C(t) + m(t)C(t) \end{aligned} \end{aligned} \tag{2.1}$$

Therefore,
When $m(t) = 0$:

$$\bullet\ S(t) = A_c \cos(2\pi f_c t) \tag{2.2}$$

When $m(t) = A_m \cos(2\pi f_m t)$:

$$\bullet\ S(t) = A_c \cos(2\pi f_c t) + A_c A_m \cos(2\pi f_m t) \cos(2\pi f_c t) \tag{2.3}$$

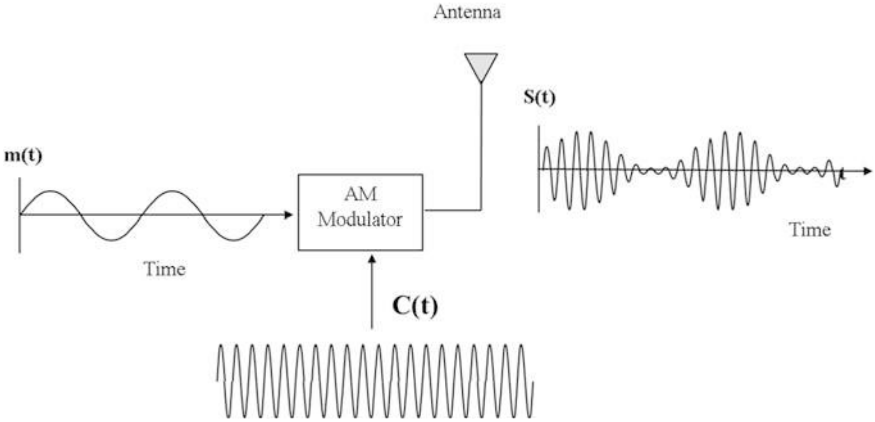

Fig. 2.2 Illustration of amplitude modulation. The amplitude of the carrier $C(t)$ changes in accordance with the input modulating signal $m(t)$. $S(t)$ is the modulated waveform which is transmitted by the antenna

In the above equation, we see that:

- The first term is the carrier only, which does not have information
- The second **term** contain**s the** information, which has several spectral components, requiring further analysis to quantify them.

2.3 AM Spectrum and Bandwidth

In wireless communications, the scarcity of RF spectrum is well known. For this reason, we have to be vigilant about using transmission bandwidth and modulation. The transmission bandwidth depends on the following:

- Spectral response of the input modulating signal
- Spectral response of the carrier frequency and
- Modulation type (AM, FM ASK, FSK, PSK, etc.)

Let us take a closer look!

2.3.1 Spectral Response of the Input Modulating Signal

In AM, the input modulating signal is a continuous time low-frequency analog signal. For simplicity, we use a sinusoidal waveform, which is periodic and continuous with respect to time. It has one frequency component. For example, the sine wave is described by the following time domain equation:

$$V(t) = V_p \sin(\omega_m t) \tag{2.4}$$

where

- Vp = Peak voltage
- $\omega_m = 2\pi f_m$
- f_m = Input modulating frequency in Hz

Figure 2.3 shows the characteristics of a sine wave and its spectral response. Since the frequency is constant, its spectral response is located in the horizontal axis at f_m, and the peak voltage is shown in the vertical axis. The corresponding bandwidth is zero.

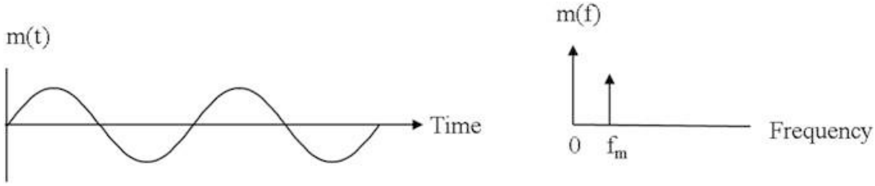

Fig. 2.3 A low-frequency sine wave and its frequency response

2.3.2 Spectral Response of the Carrier Frequency

A carrier frequency (f_c) is essentially a sinusoidal waveform, which is periodic and continuous with respect to time. It has one frequency component, which is much higher than the input modulating frequency ($f_c \gg f_m$). For example, the sine wave is described by the following time domain equation:

$$V(t) = V_p \sin(\omega_c t) \tag{2.5}$$

where

$$V_p = \text{Peak voltage}$$

- $\omega_c = 2\pi f_c$
- f_c = Carrier frequency in Hz

Figure 2.4 shows the characteristics of a sine wave and its spectral response. Since the frequency is constant, its spectral response is located in the horizontal axis at f_c and the peak voltage is shown in the vertical axis. The corresponding bandwidth is zero.

2.3.3 AM Spectrum and Bandwidth

Let us consider the AM signal again, which was derived earlier:

$$S(t) = A_c \cos(2\pi f_c t) + A_c A_m \cos(2\pi f_m t) \cos(2\pi f_c t) \tag{2.6}$$

Using the following trigonometric identity:

$$\cos A \cos B = 1/2 \cos(A + B) + 1/2 \cos(A - B) \tag{2.7}$$

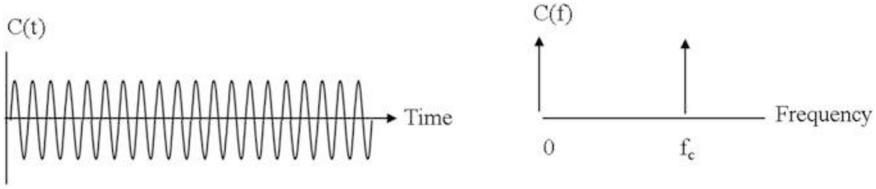

Fig. 2.4 A high-frequency sine wave and its frequency response

where

- $A = 2\pi\ (f_c + f_m)t$
- $B = 2\pi\ (f_c - f_m)t$

we get

$$S(t) = A_c \cos(2\pi f_c t) + (1/2)A_c A_m \cos[2\pi(f_c + f_m)t] + (1/2)A_c A_m \cos[2\pi(f_c - f_m)t] \tag{2.8}$$

This is the spectral response of the AM-modulated signal. It has three spectral components:

- The carrier: f_c
- Upper sideband: $f_c + f_m$
- Lower sideband: $f_c - f_m$

where f_c is the carrier frequency and f_m is the input modulating frequency. This is shown in Fig. 2.5. The AM bandwidth (BW) is given by

$$\mathrm{BW} = 2f_m \tag{2.9}$$

Notice that the power is distributed among the sidebands and the carrier, where the carrier does not contain any information. Only the sidebands contain the information. Therefore, AM is inefficient in power usage.

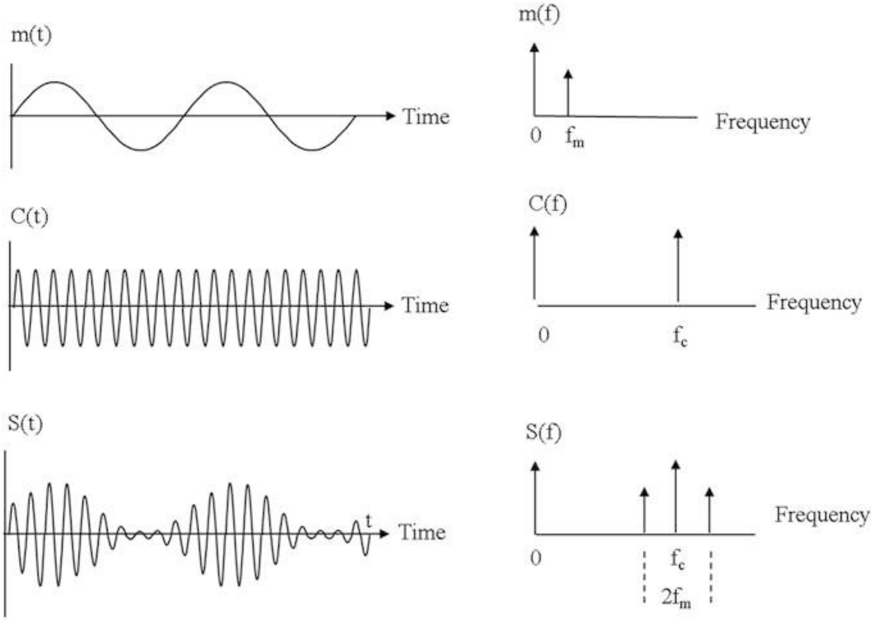
m(t)
Time
m(f)
0
f_m
Frequency
C(t)
Time
C(f)
0
f_c
Frequency
S(t)
t
Time
S(f)
0
f_c
$2f_m$
Frequency

Fig. 2.5 AM spectrum. The bandwidth is given by $2\,f_m$

2.3.4 AM Response Due to Low and High Modulating Signals

If $m(t)$ has a peak positive value of less than +1 and a peak negative value of higher than −1, then the modulation is less than 100 %. This is shown in Fig. 2.7.

On the other hand, if $m(t)$ has a peak positive value of +1 and a peak negative value of −1, then the modulation is 100 % [3–5]. Therefore,

- For $m(t) = -1: \quad S(t) = A_c[1-1]\cos(2\pi_c t) = 0 \qquad (2.10)$
- For $m(t) = +1: S(t) = A_c[1+1]\cos(2\pi f_c t) = 2A_c\cos(2\pi f_c t) \qquad (2.11)$

This is called 100 % modulation, as shown in Fig. 2.6. The percent modulation is described by the following equation:

$$\begin{aligned}&\text{The overall modulation percentage is:}\\ &\%\,\text{Overall Modulation} = \frac{A_{\max} - A_{\min}}{A_c} \times 100\\ &= \frac{\max[m(t)] - \min[m(t)]}{2A_c} \times 100.\end{aligned} \tag{2.12}$$

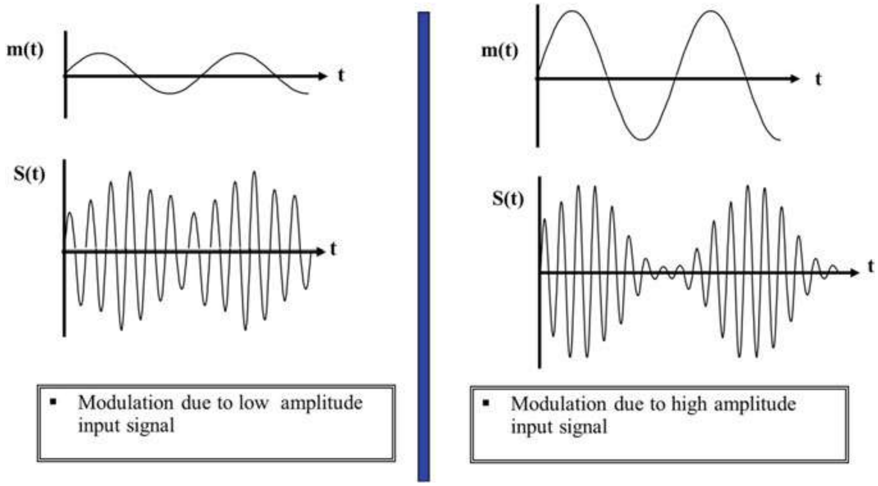

Fig. 2.6 Amplitude modulation due to low and high modulating signals

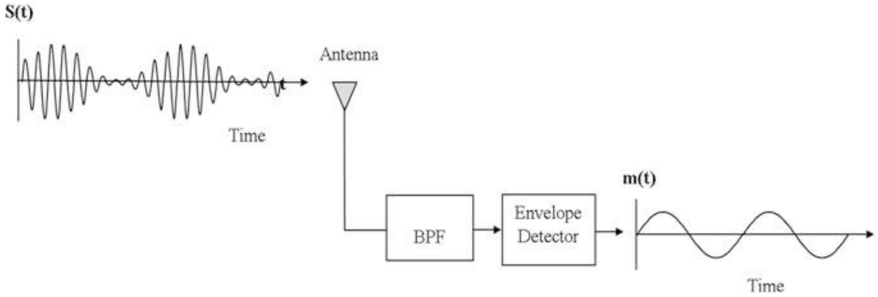

Fig. 2.7 AM demodulation technique. As the signal enters the receiver, it passes through the band-pass filter, which is tuned to the carrier frequency f_0. Next, the recovered signal is passed through an envelope detector to recover the original signal that was transmitted

2.3.5 AM Demodulation

Once the modulated analog signal has been transmitted, it needs to be received and demodulated. This is accomplished by the use of a band-pass filter that is tuned to the appropriate carrier frequency. Figure 2.7 shows the conceptual model of the AM receiver. As the signal enters the receiver, it passes through the band-pass filter, which is tuned to the carrier frequency f_0. Next, the recovered signal is passed through an envelope detector to recover the original signal that was transmitted.

2.3.6 Drawbacks in AM

- The modulated signal contains the carrier; carrier takes power and it does not have the information
- Therefore, AM is inefficient in power usage
- Moreover, there are two sidebands, containing the same information
- It is bandwidth inefficient
- AM is also susceptible to interference, since it affects the amplitude of the carrier.

Therefore, a solution is needed to improve bandwidth and power efficiency.

Problem 2.1 Given:

- Input modulating frequency $f_m = 10$ kHz
- Carrier frequency $f_c = 400$ kHz

Find

- Spectral components
- Bandwidth

Solution
Spectral Components:

- $f_c = 400$ kHz
- $f_c + f_m = 400$ kHz + 10 kHz = 410 kHz
- $f_c - f_m = 400$ kHz − 10 kHz = 390 kHz

Bandwidth

- BW = $2\,f_m = 2 \times 10$ kHz = 20 kHz.

2.4 Double Sideband-Suppressed Carrier (DSBSC)

2.4.1 DSBSC Modulation

Double sideband-suppressed carrier (DSBSC), also known as product modulator, is an AM signal that has a suppressed carrier [3–5]. Let us take the original AM signal once again, as given below:

$$S(t) = A_c\cos(2\pi f_c t) + (1/2)A_cA_m\cos[2\pi(f_c + f_m)t] + (1/2)A_cA_m\cos[2\pi(f_c - f_m)t] \tag{2.13}$$

Notice that there are three spectral components:

- The first term is the carrier only, which does not have any information
- The second and third terms contain information.

In DSBSC, we suppress the carrier, which is the first term that does not have any information. Therefore, by suppressing the first term we obtain the following:

$$S(t) = (1/2)A_cA_m\cos[2\pi(f_c + f_m)t] + (1/2)A_cA_m\cos[2\pi(f_c - f_m)t] \tag{2.14}$$

Next, we use the following trigonometric identities:

- $\cos(A + B) = \cos A \cos B - \sin A \sin B$
- $\cos(A - B) = \cos A \cos B + \sin A \sin B$

With $A = 2\pi f_m t = \omega_m t$ and $B = 2\pi f_c t = \omega_c t$, we obtain:

$$S(t) = A_cA_m \cos(\omega_m t)\ \cos(\omega_c t) \tag{2.15}$$

Now, define

- $m(t) = A_m \cos(\omega_m t)$
- $C(t) = A_c \cos(\omega_c t)$

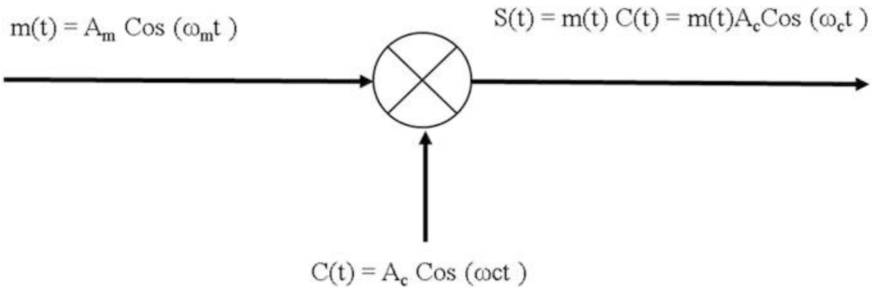
m(t) = A_m Cos (ω_m t)
S(t) = m(t) C(t) = m(t)A_c Cos (ω_c t)
C(t) = A_c Cos (ωct)

Fig. 2.8 Symbolic representation of DSBSC, also known as product modulator

Then, we can write the above equations as:

$$S(t) = m(t)\,C(t) \tag{2.16}$$

This is the DSBSC waveform. Since the output is the product of two signals, it is also known as product modulator. The symbolic representation is given in Fig. 2.8, where *m*(*t*) is the input modulating signal and *C*(*t*) is the carrier frequency.

2.4.2 Generation of DSBSC Signal

A DSBSC signal can be generated using two AM modulators arranged in a balanced configuration as shown in Fig. 2.9 [3–5]. The outcome is a cancellation of the discrete carrier. Also, the output is the product of two inputs: $S(t) = m(t)\ C(t)$. This is why it is called "product modulator."

Proof of DSBSC

Consider the DSBSC modulator as shown in Fig. 2.9. Here, the AM modulators generate $S_1(t)$ and $S_2(t)$, which are given by:

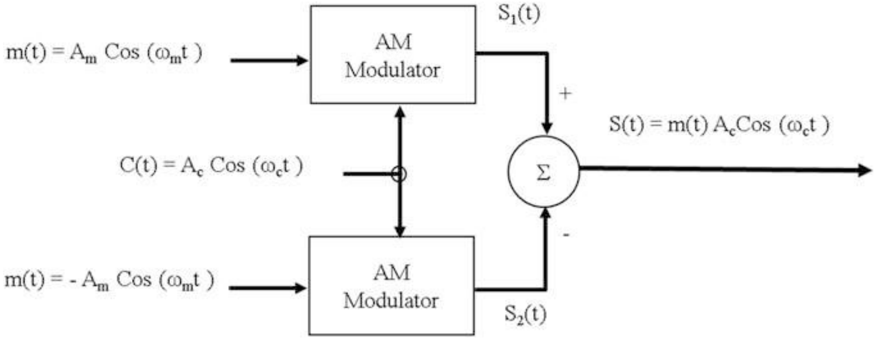
m(t) = A_m Cos (ω_m t)
AM Modulator
S_1(t)
+
S(t) = m(t) A_c Cos (ω_c t)
C(t) = A_c Cos (ω_c t)
Σ
-
m(t) = - A_m Cos (ω_m t)
AM Modulator
S_2(t)

Fig. 2.9 Construction of DSBSC modulator. The output is the product of two signals

$$S_1(t) = A_c[1 + m(t)]\cos(\omega_c t) \tag{2.17}$$

$$S_2(t) = A_c[1 - m(t)]\cos(\omega_c t)$$

Subtracting $S_2(t)$ from $S_1(t)$, we essentially cancel the carrier to obtain:

$$\begin{aligned} S(t) &= S_1(t) - S_2(t) \\ &= 2\,m(t)\,A_c\,\cos(\omega_c t) \end{aligned} \tag{2.18}$$

Therefore, except for the scaling factor 2, the above equation is exactly the same as the desired DSBSC waveform shown earlier, which does not have the carrier. In other words, the carrier has been suppressed, hence the name double sideband-suppressed carrier (DSBSC).

2.4.3 DSBSC Spectrum and Bandwidth

We begin with the DSBSC-modulated signal:

$$S(t) = 2\,m(t)\,A_c\cos(\omega_c t) \tag{2.19}$$

where

$$m(t) = A_m\cos(\omega_m t)$$

Therefore,

$$S(t) = 2\,A_c A_m\cos(\omega_m t)\,\cos(\omega_c t) \tag{2.20}$$

This is the desired DSBSC waveform for spectral analysis.

Now, use the trigonometric identity:

- $\cos A\,\cos B = 1/2\,\cos(A+B) + 1/2\,\cos(A-B)$ (2.21)

where

$$A = \omega_m t = 2\pi f_m t \text{ and } B = \omega_c t = 2\pi f_c t$$

Therefore,

$$\begin{aligned} S(t) &= A_c[\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t] \\ &= A_c[\cos 2\pi(f_c + f_m)t + \cos 2\pi(f_c - f_m)t] \end{aligned} \tag{2.22}$$

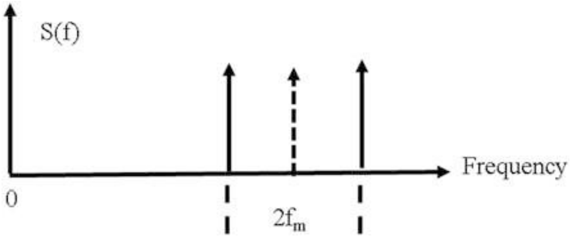

Fig. 2.10 DSBSC spectrum where the carrier frequency is suppressed. The bandwidth is given by 2 f_m

Notice that the carrier power is distributed among the sidebands (Fig. 2.10). Therefore, it is more efficient. The bandwidth is given by:

$$BW = 2f_m \tag{2.23}$$

2.4.4 DSBSC Drawbacks

- There are two identical sidebands.
- Each sideband contains the same information
- Bandwidth is 2 f_m
- Unnecessary power usage

Therefore, a solution is needed to improve bandwidth efficiency.

Problem 2.2 Given:Two product modulators using identical carriers are connected in a series, as shown below:

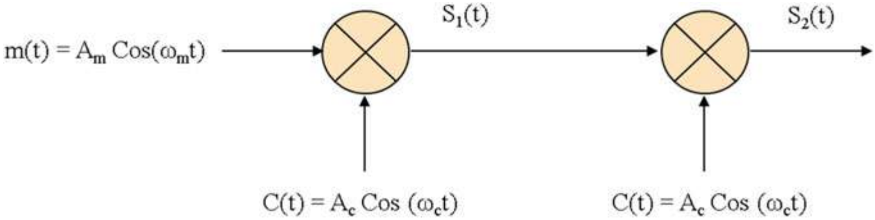

Find

(a) The output waveform $S_2(t)$
(b) What is the function of this circuit?

Solution

(a)
$$S_1(t) = m(t)\,C(t)$$
$$S_2(t) = S_1(t)\,C(t) = m(t)C^2(t) = A_m\cos(\omega_m t)\,Ac^2\cos\,2(\omega_c t)$$

(b) The function of the circuit is to demodulate DSBSC signals, where the carrier frequency is filtered out.

2.5 Single Sideband (SSB) Modulation

2.5.1 *Why SSB Modulation?*

- The basic AM has a carrier which does not carry information—*Inefficient power usage*
- The basic AM has two sidebands contain the same information—*Additional loss of power*
- DSBSC has two sidebands, containing the same information—*Loss of power*
- Therefore, the basic AM and DSBSC are bandwidth and power inefficient
- SSB is bandwidth and power efficient.

2.5.2 *Generation of SSB-Modulated Signal*

Single sideband (SSB) modulation uses two product modulators as shown in Fig. 2.11 [3–5], where

$$\bullet \quad m(t) = A_m \cos(\omega_m t) \tag{2.24}$$

$$\bullet \quad m(t)^* = A_m \sin(\omega_m t) - (\text{Hilbert Transform}) \tag{2.25}$$

$$\bullet \quad C(t) = A_m \cos\,(\omega_c t) \tag{2.26}$$

$$\bullet \quad C(t)^* = A_c \sin(\omega_c t) - (\text{Hilbert Transform}) \tag{2.27}$$

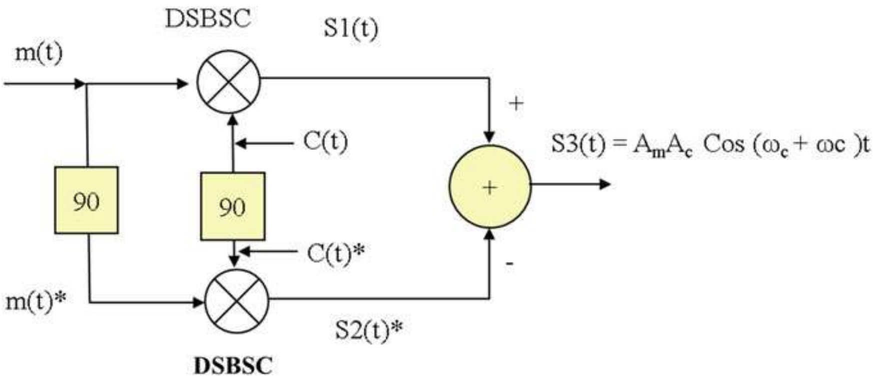

Fig. 2.11 Generation of SSB signal

Solving for S_1, S_2, and S_3, we obtain:

- $S_1(t) = A_c A_m \cos(\omega_m t)\cos(\omega_c t)$ (2.28)
- $S_2(t) = A_c A_m \sin(\omega_m t)\sin(\omega_c t)$ (2.29)

$$\begin{aligned} S_3(t) &= S_1(t) - S_2(t) \\ &= A_c A_m \cos(\omega_m t)\cos(\omega_c t) - A_c A_m \sin(\omega_m t)\sin(\omega_c t) \end{aligned} \quad (2.30)$$

Using the following formula:

- cos$A$ cos$B$ = 1/2cos(A + B) + 1/2cos(A − B)
- sin$A$ sin$B$ =1/2cos(A − B) − 1/2cos(A + B)

Solving for S_3, we get:

$$S_3 = A_c A_m \cos(\omega_c + \omega_m)t \quad (2.31)$$

In the above equation, $S_3(t)$ is the desired SSB signal, which is the upper sideband only.

2.5.3 SSB Spectrum and Bandwidth

Let us consider the SSB signal again:

$$\begin{aligned} S_3 &= A_c A_m \cos(\omega_c + \omega_m)t \\ &= A_c A_m \cos\, 2\pi(f_c + f_m)t \end{aligned} \quad (2.32)$$

Here, we see that the SSB spectrum contains only one sideband. Therefore, it is more efficient. The SSB bandwidth is given by:

$$\text{SSB BW} = f_m \quad (2.33)$$

Figure 2.12 displays the SSB spectrum.

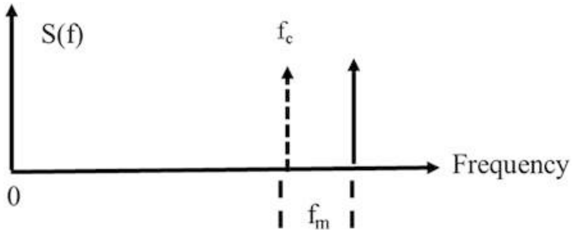

Fig. 2.12 SSB spectrum showing the upper sideband. The SSB bandwidth is f_m

Problem 2.3

Given:

- $m(t) = A_m \cos(\omega_m t)$
- $m(t)^* = A_m \sin(\omega_m t)$ − (Hilbert Transform)
- $C(t) = A_c \cos(\omega_c t)$
- $C(t)^* = A_c \sin(\omega_c t)$ − (Hilbert Transform)

Design an SSB modulator to realize the lower sideband. Sketch the spectral response.

Solution

Solving for S_1 and S_2, we obtain:

- $S_1(t) = A_c A_m \cos(\omega_m t) \cos(\omega_c t)$
- $S_2(t) = A_c A_m \sin(\omega_m t) \sin(\omega_c t)$

Obtain S_3 as:

$$\begin{aligned} S_3(t) &= S_1(t) + S_2(t) \\ &= A_c A_m \cos(\omega_m t) \cos(\omega_c t) + A_c A_m \sin(\omega_m t) \sin(\omega_c t) \end{aligned}$$

Using the following formula:

- $\cos A \cos B = 1/2\cos(A + B) + 1/2\cos(A - B)$
- $\sin A \sin B = 1/2\cos(A - B) - 1/2\cos(A + B)$

Solving for S_3, we get:

$$\begin{aligned} S_3 &= A_c A_m \cos(\omega_m - \omega_c)t \\ &= A_c A_m \cos 2\pi(f_c - f_m)t \end{aligned}$$

In the above equation, S_3(t) is the desired SSB signal, which is the lower sideband only. The spectral response, showing the lower sideband, is presented below.

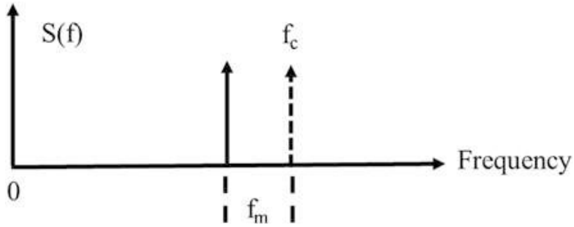

2.6 Conclusions

This chapter presents the key concepts and underlying principles of Amplitude Modulation. It was shown how the audio waveform changes the amplitude of the carrier to determine the envelope of the modulated carrier. It was also shown that the modulated carrier contains several spectral components that lead to DSBSC and SSB modulation techniques. In particular, the following topics were presented in this chapter:

- Amplitude Modulation (AM)
- AM spectrum and bandwidth
- Double sideband-suppressed carrier (DSBSC)
- DSBSC spectrum and bandwidth
- Single sideband (SSB)
- SSB spectrum and bandwidth

These materials have been augmented by diagrams and associated waveforms to make them easier for readers to grasp.

References

1. Marconi G (1897) Improvements in transmitting electrical impulses and signals, and in apparatus therefor. British patent No. 12,039. Date of Application 2 June 1896; Complete Specification Left, 2 Mar 1897; Accepted, 2 July 1897 (later claimed by Oliver Lodge to contain his own ideas which he failed to patent)
2. Marconi G (1900) Improvements in apparatus for wireless telegraphy. British patent No. 7,777. Date of Application 26 Apr 1900; Complete Specification Left, 25 Feb 1901; Accepted, 13 Apr 1901
3. Leon W, Couch II (2001) Digital and analog communication systems, 7th edn. Prentice-Hall Inc, Englewood Cliffs. ISBN 0-13-142492-0
4. Godse AP, Bakshi UA (2009) Communication engineering. Technical Publications, p 36. ISBN 978-81-8431-089-4
5. Silver W (ed) (2011) Chapter 14 transceivers. The ARRL handbook for radio communications, 88th edn. American Radio Relay League. ISBN 978-0-87259-096-0

Chapter 3
Frequency Modulation (FM)

Topics

- Introduction
- Frequency Modulation (FM)
- FM Spectrum
- Carson's Rule & FM Bandwidth
- Bessel Function & FM Bandwidth
- FM bandwidth Dilemma

3.1 Introduction

In Frequency Modulation (FM), the frequency of the carrier changes in accordance with the input analog signal, while the amplitude of the carrier remains the same [1–5]. This is shown in Fig. 3.1, where

- $m(t)$ is the input modulating audio signal,
- $C(t)$ is the carrier frequency, and
- $S(t)$ is the FM-modulated carrier frequency.

As shown in the figure, the audio waveform changes the frequency of the carrier. This enables the receiver to extract the audio signal by demodulation. Notice that the frequency of the carrier changes in accordance with the input signal, while the amplitude of the carrier does not change after modulation. However, it can be shown that the modulated carrier $S(t)$ contains an infinite number of spectral components, requiring frequency domain analysis [3]. In an effort to examine this, this chapter will present the following topics:

- The basic Frequency Modulation (FM),
- FM spectrum, and
- FM bandwidth.

© The Author(s) 2017

S. Faruque, *Radio Frequency Modulation Made Easy*,
SpringerBriefs in Electrical and Computer Engineering,
DOI 10.1007/978-3-319-41202-3_3

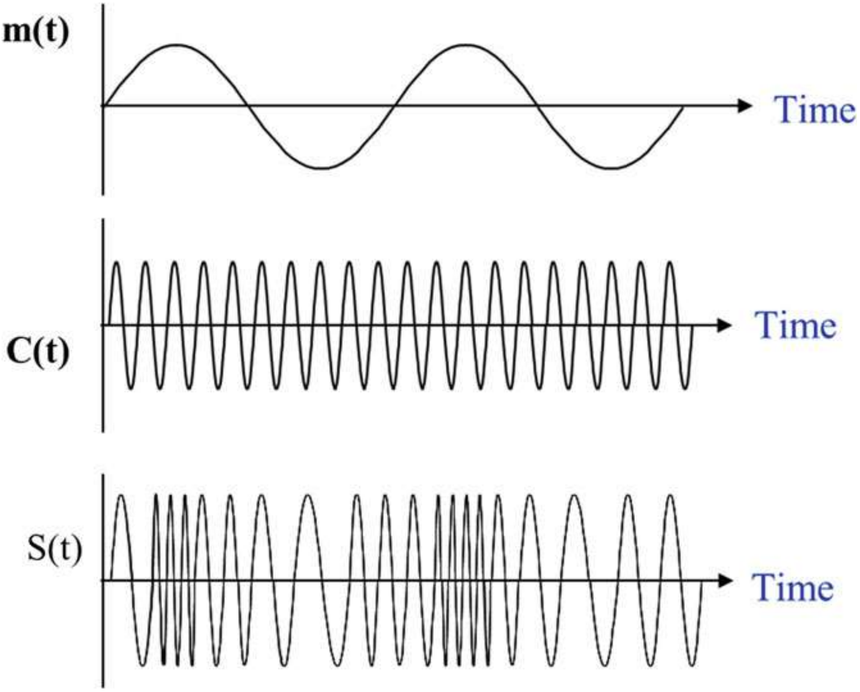

Fig. 3.1 Illustration of FM

3.2 Frequency Modulation (FM)

3.2.1 Background

FM is a form of angle modulation, where the frequency of the carrier varies in accordance with the input signal. Here, the angle refers to the angular frequency (ω). The angular frequency ω is also recognized as angular speed or circular frequency. It is a measure of rotation rate or the rate of change of the phase of a sinusoidal waveform as illustrated in Fig. 3.2.

Magnitude of the angular frequency ω is defined by one revolution or 2π radians:

$$\omega = 2\pi f = \mathrm{d}\theta/\mathrm{d}t \quad \text{Radians per second} \tag{3.1}$$

where,

- ω = Angular frequency in radians per seconds,
- f = Frequency in Hertz (Hz) or cycles per second, and
- θ = Phase angle.

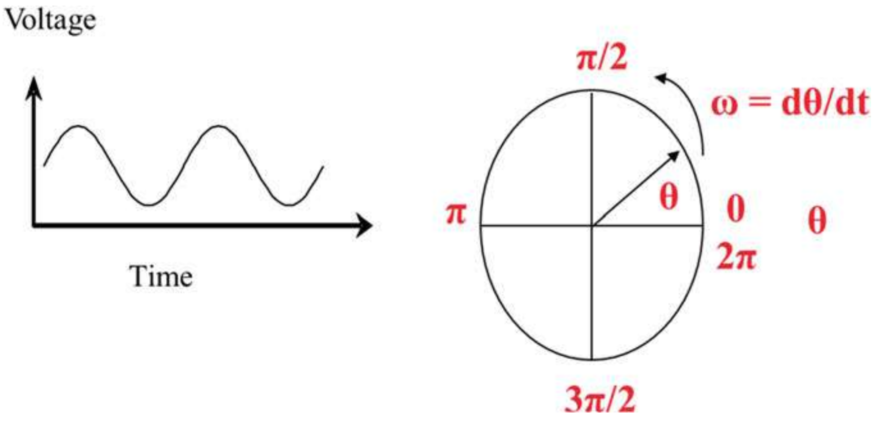

Fig. 3.2 A sinusoidal waveform in the time domain and its representation in the phase domain

Notice that in Eq. 3.1, the angular frequency ω is greater than the frequency f by a factor of 2π. Now solving for the phase angle, we obtain,

$$\theta_i(t) = 2\pi \int_0^t f_i dt \tag{3.2}$$

- θ_i = Instantaneous phase angle and
- f_i = Instantaneous frequency.

This forms the basis of our derivation of FM as presented in the following section.

3.2.2 *The Basic FM*

Frequency Modulation (FM) is a method of analog modulation that utilizes frequency variation of the relative frequency of the career [1]. The signal to be modulated and transmitted is analog. This is referred to as FM, where the frequency of the carrier changes in accordance with the input signal. The modulated carrier frequency f_c varies back and forth and depends on amplitude A_m and frequency f_m of the input signal.

Figure 3.3 shows the functional diagram of a typical FM, using a single-tone modulating signal. Here, $m(t)$ is the input analog signal we want to transmit, $C(t)$ is the carrier frequency without modulation, and $S(t)$ is the output FM-modulated carrier frequency. These parameters are described below.

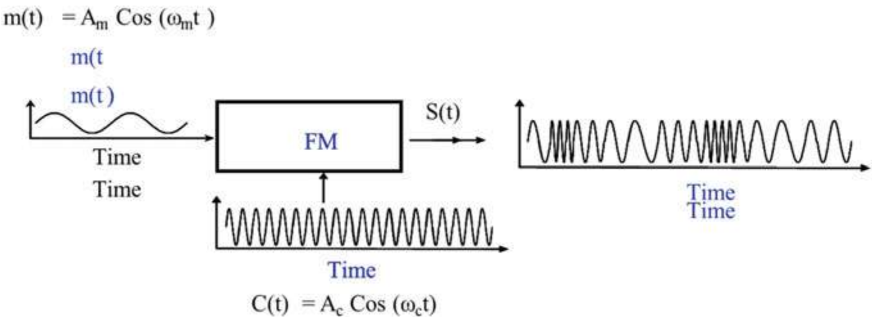

Fig. 3.3 A functional diagram of a typical FM modulator using a single-tone input modulating signal. Here, $m(t)$ is the input analog signal we want to transmit, $C(t)$ is the carrier frequency without modulation, and $S(t)$ is the output FM-modulated carrier frequency

We examine this by means of a single-tone input modulating signal and its angular frequency as shown in Fig. 3.2. Since the frequency of the carrier varies in accordance with the input signal, the instantaneous frequency of the carrier is given by

$$f_i(t) = f_c(t) + D_f m(t) \tag{3.3}$$

where,

- f_i = Instantaneous frequency,
- f_c = Carrier frequency,
- D_f = Constant, and
- $m(t) = A_m \cos(wmt)$.

The FM-modulated signal is given by

$$S(t) = A_c \cos(\theta_i) \tag{3.4}$$

where Ac is the amplitude of the carrier frequency and θ_i is the instantaneous angle. Substituting Eq. (3.2) into Eq. (3.4) for θi, we get

$$S(t) = A_c \cos[2\pi \int_0^t f_i dt] \tag{3.5}$$

where f_i is the instantaneous frequency. Substituting Eq. (3.3) into Eq. (3.5) for fi, we obtain

$$S(t) = A_c \cos[2\pi \int_0^t \{fc(t) + Df\, m(t)\} dt] \tag{3.6}$$

where $m(t) = A_m \cos(\omega_m t)$ is the input modulating signal. Integrating the above equation, we obtain the desired FM signal as follows:

$$\begin{aligned} S(t) &= A_c \cos\left[2\pi f_c t + \left(\frac{\Delta f}{f_m}\right)\sin(2\pi f_m t)\right] \\ &= A_c \cos[2\pi f_c t + \beta \sin(2\pi f_m t)] \qquad (3.7) \\ \beta &= \left(\frac{\Delta f}{f_m}\right) = \text{Modulation Index} \quad \Delta f = DfAm = \text{Freq. Deviation} \end{aligned}$$

where,

- $S(t)$ = FM-modulated carrier signal,
- f_c = Frequency of the carrier,
- A_c = Amplitude of the carrier frequency,
- $\Delta f = D_f A_m$ = Frequency deviation,
- f_m = Input modulating frequency,
- Am = Amplitude of the input modulating signal,
- D_f = A constant parameter, and
- $\beta = \Delta f / f_m$ = Modulation index.

Note that the modulation index β is an important design parameter in FM. It is directly related to FM bandwidth. It may also be noted that FM bandwidth depends on both frequency and amplitude of the input modulating signal. Let us take a closer look.

3.3 FM Spectrum and Bandwidth

In wireless communications, the scarcity of RF spectrum is well known. For this reason, we have to be vigilant about using transmission bandwidth and modulation. The transmission bandwidth depends on the following:

- Spectral response of the input modulating signal,
- Spectral response of the carrier frequency, and
- Modulation type.

Let's take a closer look!

3.3.1 Spectral Response of the Input Modulating Signal

In FM, the input modulating signal is a continuous time low-frequency analog signal. For simplicity, we use a sinusoidal waveform, which is periodic and

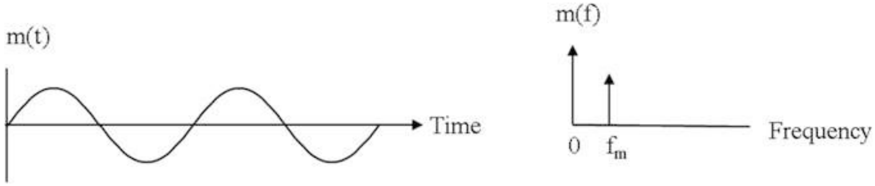

Fig. 3.4 A low-frequency sine wave and its frequency response

continuous with respect to time. It has one frequency component. For example, the sine wave is described by the following time domain equation:

$$V(t) = V_p \sin(\omega_m t) \tag{3.8}$$

where,

- Vp = Peak voltage,
- $\omega_m = 2\pi f_m$, and
- f_m = Input modulating frequency in Hz.

Figure 3.4 shows the characteristics of a sine wave and its spectral response. Since the frequency is constant, its spectral response is located in the horizontal axis at f_m and the peak voltage is shown in the vertical axis. The corresponding bandwidth is zero.

3.3.2 Spectral Response of the Carrier Frequency

A carrier frequency (f_c) is essentially a sinusoidal waveform, which is periodic and continuous with respect to time. It has one frequency component, which is much higher than the input modulating frequency ($f_c \gg f_m$). For example, the sine wave is described by the following time domain equation:

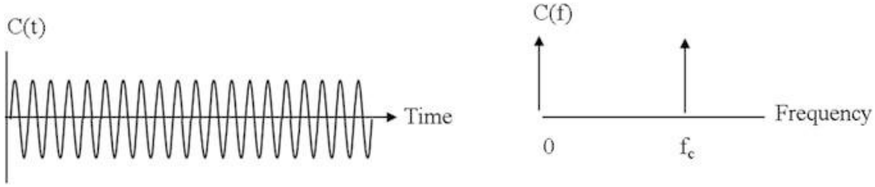

Fig. 3.5 A high-frequency sine wave and its frequency response before modulation

$$V(t) = V_p \sin(\omega_c t) \tag{3.9}$$

where

$$V_p = \text{Peak voltage}$$

- $\omega_c = 2\pi f_c$ *and*
- f_c = Carrier frequency in Hz.

Figure 3.5 shows the characteristics of a high-frequency sine wave and its spectral response. Since the frequency is constant, its spectral response is located in the horizontal axis at f_c and the peak voltage is shown in the vertical axis. The corresponding bandwidth is zero.

3.3.3 FM Spectrum

In FM, the frequency of the carrier changes in accordance with the input signal. Here, we have:

$$\begin{aligned} &\bullet\ \text{Input Signal :} && m(t) = A_m \cos(\omega_m t) \\ &\bullet\ \text{Carrier Frequency :} && C(t) = A_c \cos(\omega_c t) \\ &\bullet\ \text{Modulated Carrier :} && S(t) = A_c \cos[(\omega_c t) + \beta \sin(\omega_m t)] \end{aligned} \tag{3.10}$$

where

- $S(t)$ = The modulated carrier,
- Ac = Frequency of the carrier,
- ω_c = Nominal frequency of the carrier frequency,
- $\beta = \Delta f/fm$ = Modulation index,
- $\Delta f = D_f A_m$ = Frequency deviation,
- D_f = Constant, and
- A_m = Amplitude of the input modulating signal.

By inspecting the modulated carrier frequency, we observe that $S(t)$ depends on both frequency and amplitude of the input signal. The spectral response is given in Fig. 3.6. Notice that the carrier frequency after modulation varies back and forth from the nominal frequency f_c as depicted in the figure. In Fig. 3.6, we see that:

- As time passes, the carrier moves back and forth in frequency in exact step with the input signal.
- Frequency deviation is proportional to the input signal voltage.
- A group of many sidebands is created, spaced from carrier by amounts $N \times f_i$.
- Relative strength of each sideband depends on Bessel function.
- Strength of individual sidebands far away from the carrier is proportional to (freq. deviation × input frequency).
- Higher order spectral components are negligible.

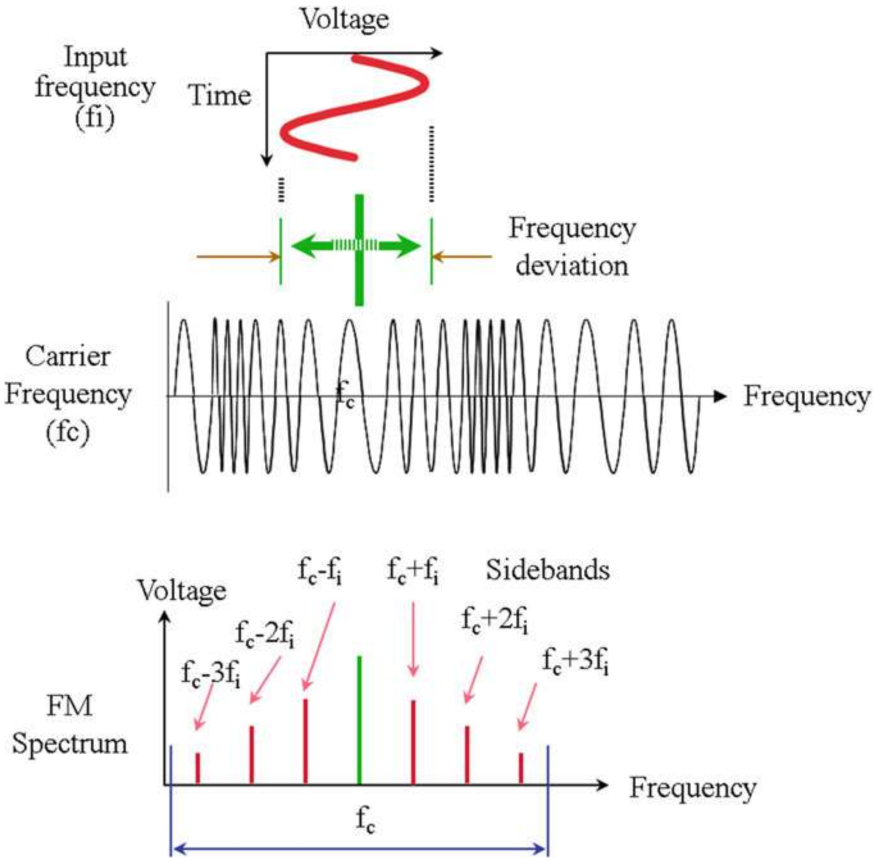
Voltage
Input frequency (fi)
Time
Frequency deviation
Carrier Frequency (fc)
f_c
Frequency
Voltage
f_c-f_i
f_c+f_i
Sidebands
f_c-2f_i
f_c+2f_i
f_c-3f_i
f_c+3f_i
FM Spectrum
Frequency
f_c

Fig. 3.6 FM spectrum. As time passes, the carrier moves back and forth in frequency in exact step with the input signal and generates an infinite number of sidebands

- Carson's rule can be used to determine the approximate bandwidth: bandwidth required = 2 × (highest input frequency + frequency deviation).

3.3.4 Carson's Rule and FM Bandwidth

The Carson's rule, a rule of thumb, states that more than 98 % of the power of FM signal lies within a bandwidth given by the following approximation:

$$\text{FM Bandwidth (BW)} = 2f_m(1+\beta) \tag{3.11}$$

where

Table 3.1 Bessel function table

β	J0	J1	J2	J3	J4	J5
0	1					
0.25	0.98	0.12				
0.5	0.94	0.24	0.03			
1	0.77	0.44	0.11	0.02		
1.5	0.51	0.56	0.23	0.06	0.01	
2	0.22	0.58	0.35	0.13	0.03	

- $\beta = \Delta f/fm$ = Modulation index,
- Δf = Peak deviation of the instantaneous frequency from the center of the carrier frequency, and
- fm = Highest frequency of the modulating signal.

3.3.5 Bessel Function and FM Bandwidth

FM bandwidth can be estimated by means of Bessel function of the first kind. For a single-tone modulation, it can be obtained as a function of the sideband number and the modulation index.

For a given β, the solution for $S(t)$ is given as follows:

$$\begin{aligned} S(t) &= A_c\cos[(\omega_c t) + \beta\sin(\omega_m t)] \\ &= J_0\,(\beta)\cos(\omega_c t) \\ &\quad + J_1(\beta)\cos(\omega_c t + \omega_m t) + J_2(\beta)\cos(\omega_c t + 2\omega_m t) + J_3(\beta)\cos(\omega_c t + 3\omega_m t) + \ldots \\ &\quad - J_1(\beta)\cos(\omega_c t - \omega_m t) - J_2(\beta)\cos(\omega_c t - 2\omega_m t) - J_3(\beta)\cos(\omega_c t - 3\omega_m t) + \ldots \end{aligned} \tag{3.12}$$

Here, J's are the Bessel functions, representing the amplitude of the sidebands. J_0 (β) is the amplitude of the fundamental spectral component, and the remaining spectral components are the sidebands. Each sideband is separated by the input modulating frequency. These values are also available as a standard Bessel function table.

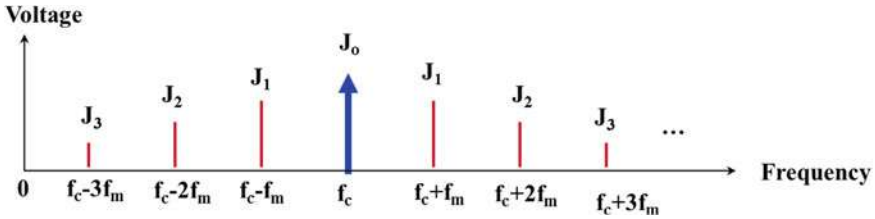

Fig. 3.7 FM sidebands. Power is taken from the carrier $J0$ and distributed among the sidebands $J1$, $J2$, $J3$,…, etc. Each sideband is separated by the modulating frequency f_m

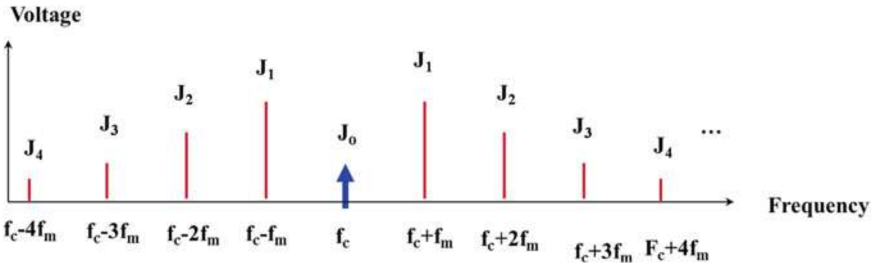

Fig. 3.8 FM sidebands for large β. More power is taken from the carrier $J0$ and distributed among the sidebands $J1$, $J2$, $J3$, $J4$,… Each sideband is separated by the modulating frequency f_m

As an example, Table 3.1 provides a few Bessel parameters to illustrate the concept. In this table, the carrier and sideband amplitude powers, $J0$, $J1$, $J2$,…, etc., are presented for different values of β. Here, $J0$ is carrier power before modulation. After modulation, with a given value of β, power is taken from the carrier and distributed among the sidebands. Also, there is a unique value of β, for which the carrier amplitude becomes zero and all the signal power is in the sidebands.

Note that, in FM, the sidebands are on both sides of the carrier. Therefore, the total bandwidth includes spectral components from both sides of the carrier. For a given low β, this is shown in Fig. 3.7. Here, J0 is the spectral component of the carrier. $J1$, $J2$, $J3$, etc., are the sidebands. Each sideband is separated by the input modulating frequency f_m. After modulation, power is taken from the carrier J0 and distributed among the sidebands, depending on the modulation index β.

Figure 3.8 shows another scenario where β is large. In this case, more power is taken from the carrier and distributed among the sideband, while creating more significant sidebands, requiring more bandwidth.

3.3.6 FM Bandwidth Dilemma

In FM, we notice that if the modulation index is low, the occupied bandwidth is low and the sidebands take less power from the carrier, making the modulation less efficient. On the other hand, if the modulation index is high, the occupied bandwidth is also high, while the sidebands retain most of the power, making the modulation more efficient at the expense of bandwidth. This is a dilemma in FM. However, FM is more popular because it is less sensitive to noise.

Problem 3.1
Given:

$$f_m = 1\,\text{kHz}$$
$$f_c = 1\,\text{MHz}$$
$$\beta = 1$$

Find:

(a) The FM bandwidth using Carson's rule and
(b) The FM bandwidth using Bessel function

Solution:

(a) FM BW (Carson's rule) = $2\,f_m\,(1 + \beta) = 2 \times 1$ kHz $(1 + 1) = 4$ kHz.
(b) For $\beta = 1$, See Table 3.1:

- $J0 = 0.77$, $J1 = 0.44$, $J2 = 0.11$, $J3 = 0.02$

where $J0$ is the carrier and $J1$, $J2$, and $J3$ are the sidebands. Here, we can neglect $J3$ since it has a negligible power. Therefore, there are two significant upper and lower sidebands, and each sideband is 1 kHz apart. Therefore, FM bandwidth is 2 × Number of significant sidebands = 2 × 2 = 4 kHz. See figure below.

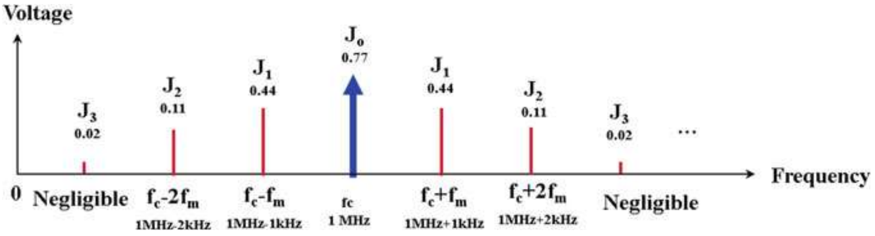

Problem 3.2
Given:

$$fm = 1\,\text{kHz}$$
$$fc = 1\,\text{MHz}$$
$$\beta = 2$$

Find:

(a) The FM bandwidth using Carson's rule and
(b) The FM bandwidth using Bessel function

Solution:

(a) FM BW (Carson's rule) = $2\,f_m\,(1 + \beta) = 2 \times 1$ kHz $(1 + 2) = 6$ kHz.
(b) For $\beta = 2$, See Table 3.1:

- $J0 = 0.22$, $J1 = 0.58$, $J2 = 0.35$, $J3 = 0.13$, $J4 = 0.03$

Here, we have 4 upper and 4 lower sidebands, and each sideband is separated by 1 kHz. Since the J4 is negligible, we take 3 upper sidebands and 3 lower sidebands. Therefore, the FM bandwidth is 2 × 3 = 6 kHz. See figure below.

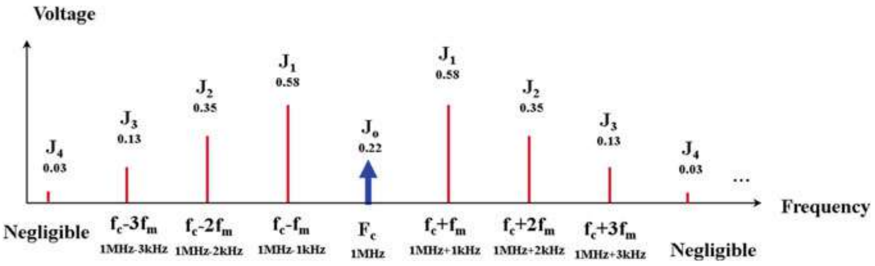

3.4 Conclusions

- This chapter presents a brief overview of Frequency Modulation (FM) and its attributes.
- The key concept and the underlying principle of FM are presented with numerous illustrations to make them easier for readers to grasp.
- FM spectrum is addressed with illustrations.
- FM bandwidth is estimated by means of Carson’s rule and Bessel function.
- FM bandwidth dilemma is explained, and problems are given to illustrate the concept.

References

1. Armstrong EH (1936) A method of reducing disturbances in radio signaling by a system of frequency modulation. Proc IRE (IRE) 24(5):689–740. doi:10.1109/JRPROC.1936.227383
2. Lathi BP (1968) Communication systems. Wiley, New York, pp 214–217. ISBN 0-471-51832-8
3. Leon W, Couch II (2001) Digital and analog communication systems, 7th edn. Prentice-Hall Inc, Englewood Cliffs. ISBN 0-13-142492-0
4. Haykin S (2001) Communication systems, 4th edn. Wiley, New York
5. Felix MO (1965) FM systems of exceptional bandwidth. Proc IEEE 112(9):1664

Chapter 4
Amplitude Shift Keying (ASK)

Topics

- Introduction
- Amplitude Shift Keying (ASK)
- ASK spectrum
- ASK Bandwidth
- Performance Analysis

4.1 Introduction

In amplitude shift keying (ASK), the amplitude of the carrier changes in discrete levels in accordance with the input digital signal, while the frequency of the carrier remains the same. This is shown in Fig. 4.1, where

- $m(t)$ is the input modulating digital signal,
- $C(t)$ is the carrier frequency, and
- $S(t)$ is the ASK-modulated carrier frequency.

As shown in the figure, the digital binary signal changes the amplitude of the carrier in two discrete levels. This enables the receiver to extract the digital signal by demodulation. Notice that the amplitude of the carrier changes in accordance with the input signal, while the frequency of the carrier does not change after modulation. However, it can be shown that the modulated carrier $S(t)$ contains several spectral components, requiring frequency domain analysis.

In the following sections, the above disciplines in ASK modulation will be presented along with the respective spectrum and bandwidth. These materials have been augmented by diagrams and associated waveforms to make them easier for readers to grasp.

© The Author(s) 2017

S. Faruque, *Radio Frequency Modulation Made Easy*,
SpringerBriefs in Electrical and Computer Engineering,
DOI 10.1007/978-3-319-41202-3_4

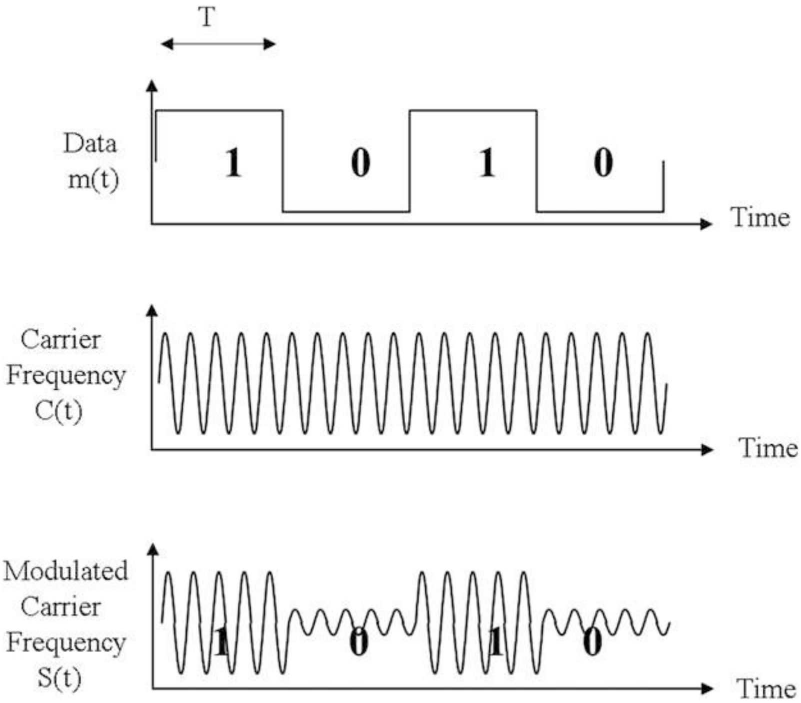

Fig. 4.1 ASK waveforms. The amplitude of the carrier changes in accordance with the input digital signal. The frequency of the carrier remains the same

4.2 ASK Modulation

Amplitude shift keying (ASK), also known as on–off keying (OOK), is a method of digital modulation that utilizes amplitude shifting of the relative amplitude of the career frequency [1–3]. The signal to be modulated and transmitted is binary, which is encoded before modulation. This is an indispensable task in digital communications, where redundant bits are added with the raw data that enables the receiver to detect and correct bit errors, if they occur during transmission [4–16].

While there are many error-coding schemes available, we will use a simple coding technique, known as "Block Coding" to illustrate the concept.

Figure 4.2 shows an encoded ASK modulation scheme using (15, 8) block code where an 8-bit data block is formed as M-rows and N-columns ($M = 4, N = 2$). The product $MN = k = 8$ is the dimension of the information bits before coding. Next, a horizontal parity P_H is appended to each row and a vertical parity P_V is appended to each column. The resulting augmented dimension is given by the product $(M + 1)(N + 1) = n = 15$, which is then ASK modulated and transmitted row by row. The rate of this coding scheme is given by

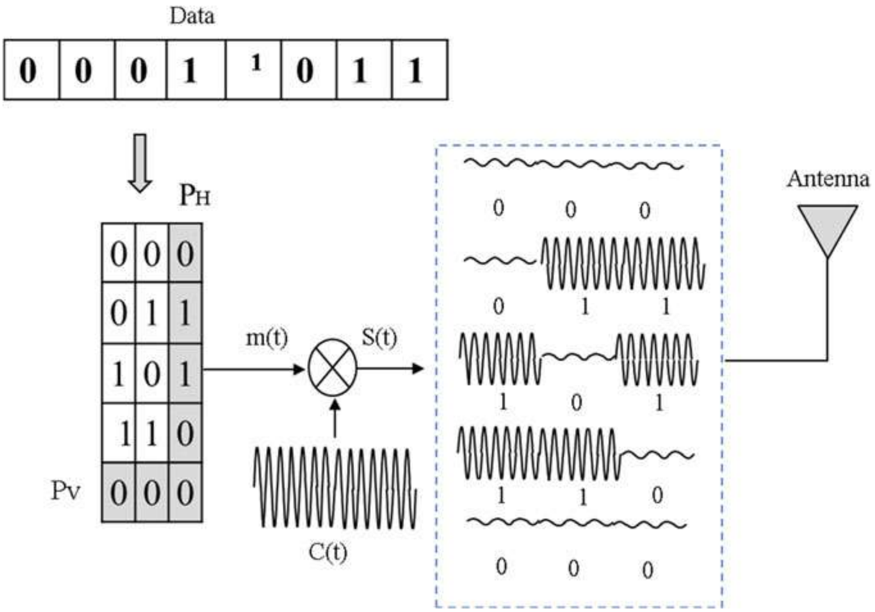

Fig. 4.2 Amplitude shift keying (ASK) is also known as on–off keying (OOK). The input encoded data block is transmitted row by row. The amplitude of the carrier frequency changes in accordance with the input digital signal

$$\text{Code Rate: } r = (MN)/[(M+1)(N+1)] = (4 \times 2)/(5 \times 3) = 8/15 \tag{4.1}$$

The coded bit rate R_{b2} is given by

$$R_{b2} = \text{Uncoded Bit Rate}/\text{Code Rate} = R_{b1}/r = R_{b1}\,(15/8) \tag{4.2}$$

Next, the coded bits are modulated by means of the ASK modulator as shown in the figure. Here,

- The Input digital signal is the encoded bit sequence we want to transmit
- Carrier is the radio frequency without modulation
- Output is the ASK-modulated carrier, which has two amplitudes corresponding to the binary input signal. For binary signal 1, the carrier is ON. For the binary signal 0, the carrier is OFF; however, a small residual signal may remain due to noise, interference, etc., as indicated in the figure.

As shown in the figure, the amplitude of the carrier changes in discrete levels, in accordance with the input signal, where

$$\begin{aligned} &\bullet\ \text{Input Data :} && m(t) = 0 \text{ or } 1 \text{ (coded data)} \\ &\bullet\ \text{Carrier Frequency :} && C(t) = A\cos(\omega t) \\ &\bullet\ \text{Modulated Carrier :} && S(t) = m(t)C(t) = m(t)A\cos(\omega t) \end{aligned} \tag{4.3}$$

Therefore,

$$\begin{aligned} &\text{For } m(t) = 1 : S(t) = A\cos(\omega t) \text{ i.e. the carrier is ON} \\ &\text{For } m(t) = 1 : S(t) = 0, \text{ i.e. the carrier is OFF} \end{aligned} \tag{4.4}$$

where A is the amplitude and ω is the frequency of the carrier.

4.3 ASK Demodulation

Once the modulated binary data has been transmitted, it needs to be received and demodulated. This is often accomplished with the use of a band-pass filter. In the case of ASK, the receiver needs to utilize one band-pass filter that is tuned to the appropriate carrier frequency. As the signal enters the receiver, it passes through the filter and a decision as to the value of each bit is made to recover the encoded data block, along with horizontal and vertical parities. Next, the receiver appends horizontal and vertical parities P_H* and P_V* to check parity failures and recovers the data block. This is shown in Fig. 4.3 having no errors. If there is an error, there will be a parity failure in P_H* and P_V* to pinpoint the error.

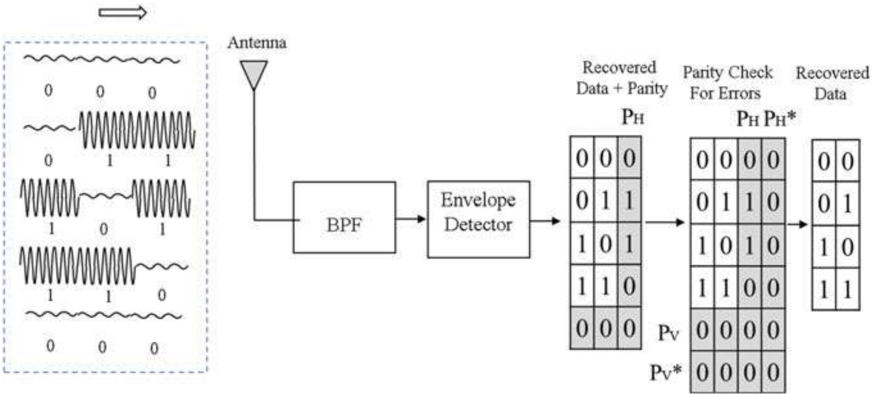

Recovered Data + Parity

		PH
0	0	0
0	1	1
1	0	1
1	1	0
0	0	0

Parity Check For Errors

			PH	PH*
	0	0	0	0
	0	1	1	0
	1	0	1	0
	1	1	0	0
Pv	0	0	0	0
Pv*	0	0	0	0

Recovered Data

0	0
0	1
1	0
1	1

Fig. 4.3 Data recovery process in ASK, showing no errors. If there is an error, there will be a parity failure in P_H* and P_V* to pinpoint the error

4.4 ASK Bandwidth

In wireless communications, the scarcity of RF spectrum is well known. For this reason, we have to be vigilant about using transmission bandwidth in error control coding and modulation. The transmission bandwidth depends on:

- Spectral response of the encoded data
- Spectral response of the carrier frequency, and
- Modulation type.

4.4.1 Spectral Response of the Encoded Data

In digital communications, data is generally referred to as a non-periodic digital signal. It has two values:

- Binary-1 = High, Period = T
- Binary-0 = Low, Period = T

Also, data can be represented in two ways:

- Time domain representation and
- Frequency domain representation

The time domain representation (Fig. 4.4a), known as non-return-to-zero (NRZ), is given by:

$$\begin{aligned} V(t) &= V \qquad <0<t<T \\ &= 0 \qquad \text{elsewhere} \end{aligned} \tag{4.5}$$

The frequency domain representation is given by "Fourier transform":

$$V(\omega) = \int_0^T V \cdot e^{-j\omega t} dt \tag{4.6}$$

$$\begin{aligned} |V(\omega)| &= VT\left[\frac{\sin(\omega T/2)}{\omega T/2}\right] \\ P(\omega) &= \left(\frac{1}{T}\right)|V(\omega)|^2 = V^2 T\left[\frac{\sin(\omega T/2)}{\omega T/2}\right]^2 \end{aligned} \tag{4.7}$$

Here, $P(\omega)$ is the power spectral density. This is plotted in Fig. 4.4b. The main lobe corresponds to the fundamental frequency and side lobes correspond to harmonic components. The bandwidth of the power spectrum is proportional to the frequency. In practice, the side lobes are filtered out since they are relatively insignificant with respect to the main lobe. Therefore, the one-sided bandwidth is given by the ratio $f/f_b = 1$. In other words, the one-sided bandwidth $= f = f_b$, where $f_b = R_b = 1/T$, T being the bit duration.

The general equation for two-sided response is given by:

$$V(\omega) = \int_{-\propto}^{\propto} V(t) \cdot e^{-j\omega t} dt$$

In this case, $V(\omega)$ is called two-sided spectrum of $V(t)$. This is due to both positive and negative frequencies used in the integral. The function can be a voltage or a current. Figure 4.4c shows the two-sided response, where the bandwidth is determined by the main lobe as shown below:

$$\text{Two-sided bandwidth (BW)} = 2R_b \; (R_b = \text{Bit rate before coding}) \tag{4.8}$$

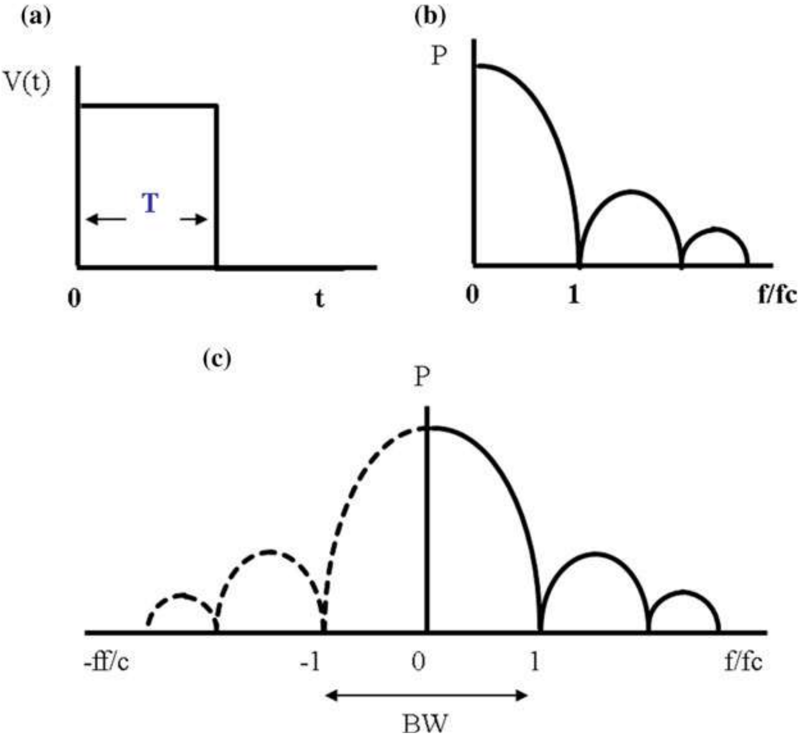

Fig. 4.4 **a** Discrete time digital signal **b** it is one-sided power spectral density and **c** two-sided power spectral density. The bandwidth associated with the non-return-to-zero (NRZ) data is $2R_b$, where R_b is the bit rate

Important Notes

1. If R_b is the bit rate before coding, and if the data is NRZ, then the bandwidth associated with the raw data will be $2R_b$. For example, if the bit rate before coding is 10 kb/s, then the bandwidth associated with the raw data will be 2 × 10 kb/s = 20 kHz.
2. If R_b is the bit rate before coding, code rate is r, and if the data is NRZ, then the bit rate after coding will be R_b (coded) = R_b (uncoded)r. The corresponding bandwidth associated with the coded data will be $2R_b$ (coded) = $2R_b$ (uncoded)/r. For example, if the bit rate before coding is 10 kb/s and the code rate $r = 1/2$, then the coded bit rate will be R_b (coded) = R_b (uncoded)/r = 10/0.5 = 20 kb/s. The corresponding bandwidth associated with the coded data will be 2 × 20 = 40 kHz.

4.4.2 Spectral Response of the Carrier Frequency Before Modulation

A carrier frequency is essentially a sinusoidal waveform, which is periodic and continuous with respect to time. It has one frequency component. For example, the sine wave is described by the following time domain equation:

$$V(t) = V_p \sin(\omega t_c) \tag{4.9}$$

where

$$V_P = \text{Peak voltage}$$

- $\omega_c = 2\pi f_c$
- f_c = Carrier frequency in Hz

Figure 4.5 shows the characteristics of a sine wave and its spectral response. Since the frequency is constant, its spectral response is located in the horizontal axis and the peak voltage is shown in the vertical axis. The corresponding bandwidth is zero.

4.4.3 ASK Bandwidth at a Glance

In ASK, the amplitude of the carriers changes in discrete levels, in accordance with the input signal, where

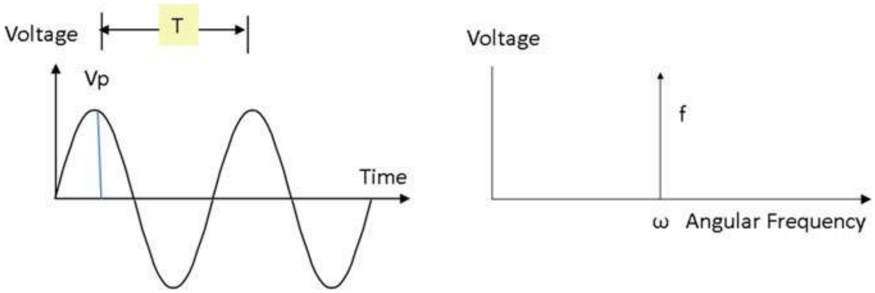

Fig. 4.5 High-frequency carrier frequency response

- Input Data : $m(t) = 0 \text{ or } 1$
- Carrier Frequency : $C(t) = Ac\cos(\omega_c t)$
- Modulated Carrier : $S(t) = m(t)C(t) = m(t)A_c\cos(\omega_c t)$

Since $m(t)$ is the input digital signal and it contains an infinite number of harmonically related sinusoidal waveforms and that we keep the fundamental and filter out the higher-order components, we write:

$$m(t) = A_m \sin(\omega_m t)$$

The ASK-modulated signal then becomes:

$$\begin{aligned} S(t) = m(t)S(t) &= A_m A_c \sin(\omega_m t)\cos(\omega_m t) \\ &= 1/2\,A_m A_c\,[\sin(\omega_c - \omega_m)t + \sin(\omega_c + \omega_m)t] \end{aligned} \tag{4.10}$$

The spectral response is depicted in Fig. 3.14. Notice that the spectral response after ASK modulation is the shifted version of the NRZ data. Bandwidth is given by:

BW = $2R_b$ (coded), where R_b is the coded bit rate (Fig. 4.6).

Problem 1

Given:

- Bit rate before coding: R_{B1} = 10 kb/s
- Code rate: r = 8/15
- Modulation: ASK

Find:

(a) the bit rate after coding: R_{B2}
(b) Transmission bandwidth: BW

Solution

(a) Bit rate after coding (R_{B2})
$R_{B2} = R_{B1}/r$ = 10 kb/s (15/8) = 18.75 kb/s
(b) Transmission BW = 2 × 18.75 kb/s = 37.5 kHz.

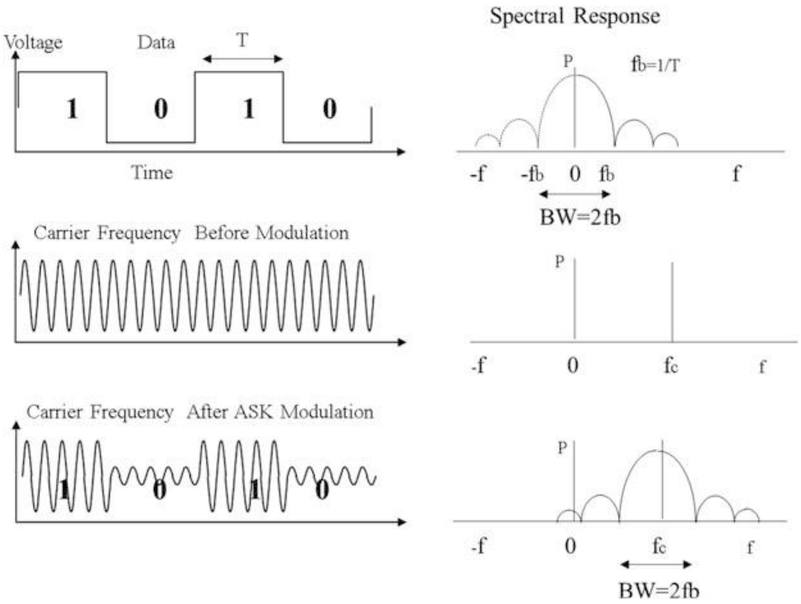
Voltage
Data
T
1
0
1
0
Time
Spectral Response
P
fb=1/T
-f
-fb
0
fb
f
BW=2fb
Carrier Frequency Before Modulation
P
-f
0
fc
f
Carrier Frequency After ASK Modulation
1
0
1
0
P
-f
0
fc
f
BW=2fb

Fig. 4.6 ASK bandwidth at a glance. **a** Spectral response of NRZ data before modulation. **b** Spectral response of the carrier before modulation. **c** Spectral response of the carrier after modulation. The transmission bandwidth is $2f_b$, where f_b is the bit rate and $T = 1/f_b$ is the bit duration for NRZ data

4.5 BER Performance

It is well known that an (n, k) bock code, where k = number of information bits, n = number of coded bits, can correct t errors [4, 5]. A measure of coding gain is then obtained by comparing the uncoded word error WER_U with the coded word WER_C. We examine this by means of the following analytical means.

Let the uncoded word error be defined as (WER_U). Then, with ASK modulation, the uncoded BER will be given by:

$$\mathrm{BER}_U = 0.5\,\mathrm{EXP}(-E_B/2N_0) \tag{4.11}$$

The probability that the uncoded word (WER_U) will be received in error is 1 minus the product of the probabilities that each bit will be received correctly. Thus, we write:

$$\mathrm{WER}_U = 1 - (1 - \mathrm{BER}_U)^k \tag{4.12}$$

Let the coded word error be defined as (WER_C). Since $n > k$, the coded bit energy to noise ratio will be modified to E_c/N_O, where $E_c/N_O = E_b/N_O + 10\log(k/n)$. Therefore, the coded BERc will be:

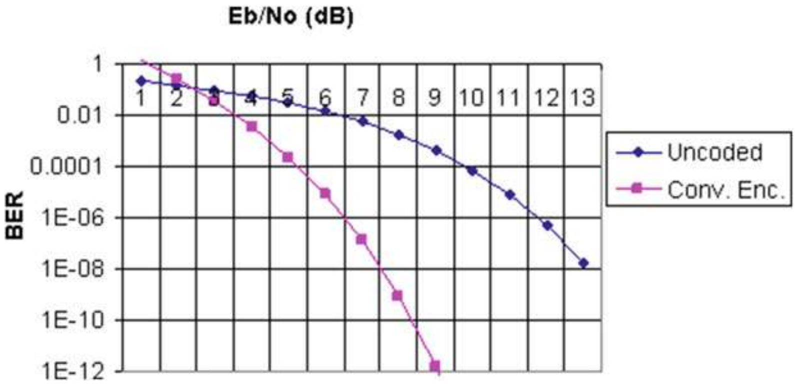

Fig. 4.7 Typical error performance in AWGN

$$\mathrm{BER}_C = 0.5\,\mathrm{EXP}(-E_C/2N_0) \tag{4.13}$$

The corresponding coded word error rate is given as:

$$\mathrm{WER}_C = \sum_{k=t+1}^{n} \binom{n}{k} \mathrm{BER}_C^K (1 - \mathrm{BER}_C)^{n-k} \tag{4.14}$$

When $\mathrm{BER}_c < 0.5$, the first term in the summation is the dominant one; therefore, Eq. (4.14) can be simplified as

$$\mathrm{WER}_C \approx \binom{n}{k} \mathrm{BER}_c^k (1 - \mathrm{BER}_c)^{n-k} \tag{4.15}$$

Using (15, 8) block code ($n = 15$, $k = 8$, $t = 1$), we obtain the coded and the uncoded WER as shown in Fig. 4.7. Coding gain is the difference in E_b/N_0 between the two curves. Notice that at least 3–4 db coding gain is available in this example where $r = k/n = 8/15$.

4.6 Conclusions

- This chapter presents ASK modulation and its attributes.
- Numerous illustrations are provided to show how amplitude of the carrier changes in discrete levels in accordance with the input digital signal, while the frequency of the carrier remains the same.
- The Fourier transform is used to derive the spectral components and ASK bandwidth is calculated.

- Bit error rate (BER) performance is presented.
- These materials have been augmented by diagrams and associated waveforms to make them easier for readers to grasp.

References

1. Smith DR (1985) Digital transmission systems, Van Nostrand Reinhold Co. ISBN: 0442009178
2. Leon W, Couch II (2001) Digital and analog communication systems, 7th edn. Prentice-Hall, Inc., Englewood Cliffs. ISBN: 0-13-142492-0
3. Sklar B (1988) Digital communications fundamentals and applications. Prentice Hall, Englewood Cliffs
4. Faruque S (2016) Radio frequency channel coding made easy. Springer. ISBN: 978-3-319-21169-5
5. Clark GC et al (1981) Error correction coding for digital communications. Plenum press, New York
6. Ungerboeck G (1982) Channel coding with multilevel/multiphase signals. IEEE Trans Inf Theory IT28:55–67
7. Lin S, Costello DJ Jr (1983) Error control coding: fundamentals and applications. Prentice-Hall, Inc., Englewood Cliffs
8. Blahut RE (1983) Theory and practice of error control codes. Addison-Wesley, Reading
9. van Lint JH (1992) Introduction to coding theory. GTM 86, 2nd edn. Springer, p 31. ISBN 3-540-54894-7
10. Mac Williams FJ, Sloane NJA (1977) The theory of error-correcting codes. North-Holland, p 35. ISBN 0-444-85193-3
11. Huffman W, Pless V (2003) Fundamentals of error-correcting codes. Cambridge University Press. ISBN 978-0-521-78280-7
12. Ebert PM, Tong SY (1968) Convolutional Reed-Soloman codes. Bell Syst Tech 48:729–742
13. Gallager RG (1968) Information theory and reliable communications. Wiley, New York
14. Kohlenbero A, Forney GD Jr (1968) Convolutional coding for channels with memory. IEEE Trans Inf Theory IT-14:618–626
15. Reddy SM (1968) Further results on convolutional codes derived from block codes. Inf Control 13:357–362
16. Reddy SM (1968) A class of linear convolutional codes for compound channels. Technical report, Bell Telephone Laboratories, Holmdel, New Jersey

Chapter 5
Frequency Shift Keying (FSK)

Topics

- Introduction
- Frequency Shift Keying (FSK)
- FSK spectrum
- FSK Bandwidth
- Performance Analysis

5.1 Introduction

In frequency shift keying (FSK), the frequency of the carrier changes in discrete levels in accordance with the input digital signal, while the amplitude of the carrier remains the same. This is shown in Fig. 5.1 where

- $m(t)$ is the input modulating digital signal,
- $C(t)$ is the carrier frequency, and
- $S(t)$ is the FSK-modulated carrier frequency.

As shown in the figure, the digital binary signal changes the frequency of the carrier on two discrete levels. This enables the receiver to extract the digital signal by demodulation. Notice that the frequency of the carrier changes in accordance with the input signal, while the amplitude of the carrier does not change after modulation. However, it can be shown that the modulated carrier $S(t)$ contains several spectral components, requiring frequency-domain analysis.

In the following sections, the above disciplines in FSK modulation will be presented, along with the respective spectrum and bandwidth. These materials have been augmented by diagrams and associated waveforms to make them easier for readers to grasp.

© The Author(s) 2017

S. Faruque, *Radio Frequency Modulation Made Easy*,
SpringerBriefs in Electrical and Computer Engineering,
DOI 10.1007/978-3-319-41202-3_5

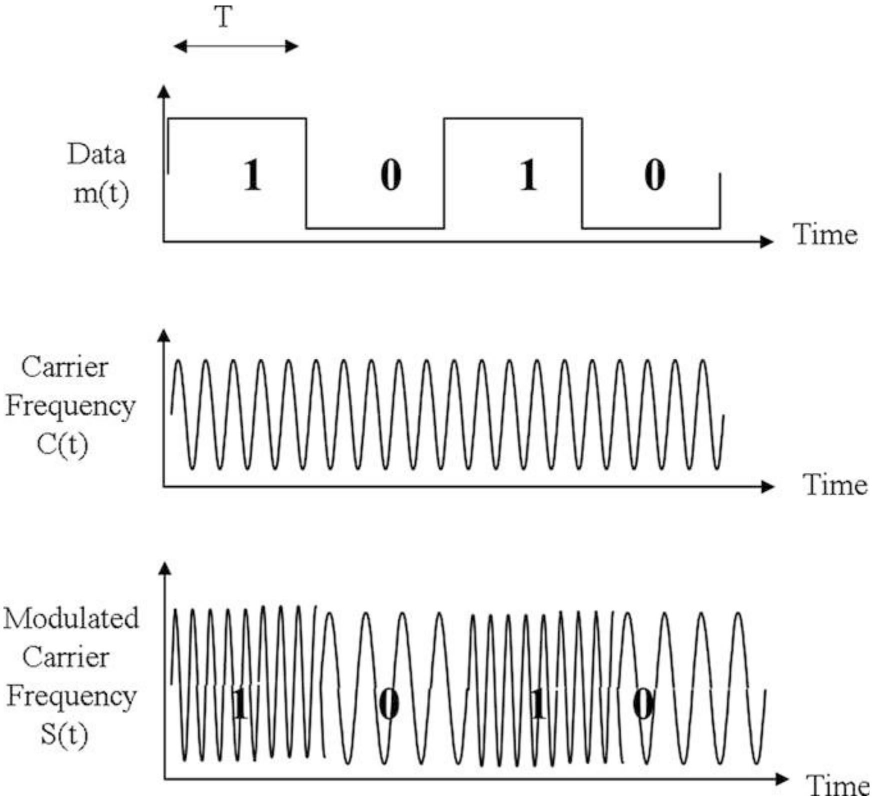

Fig. 5.1 FSK waveforms. The frequency of the carrier changes in accordance with the input digital signal. The amplitude of the carrier remains the same

5.2 Frequency Shift Keying (FSK) Modulation

Frequency shift keying (FSK) is a method of digital modulation that utilizes frequency shifting of the relative frequency content of the signal [1–3]. The signal to be modulated and transmitted is binary, which is encoded before modulation. This is an indispensable task in digital communications, where redundant bits are added with the raw data that enables the receiver to detect and correct bit errors, if they occur during transmission [4–16]. While there are many error-coding schemes available, we will use a simple coding technique, known as "block coding" to illustrate the concept.

Figure 5.2 shows an encoded FSK modulation scheme using (15, 8) block code where an 8-bit data block is formed as M rows and N columns ($M = 4$, $N = 2$). The product $MN = k = 8$ is the dimension of the information bits before coding. Next, a horizontal parity P_H is appended to each row and a vertical parity P_V is appended to each column. The resulting augmented dimension is given by the product $(M + 1)$

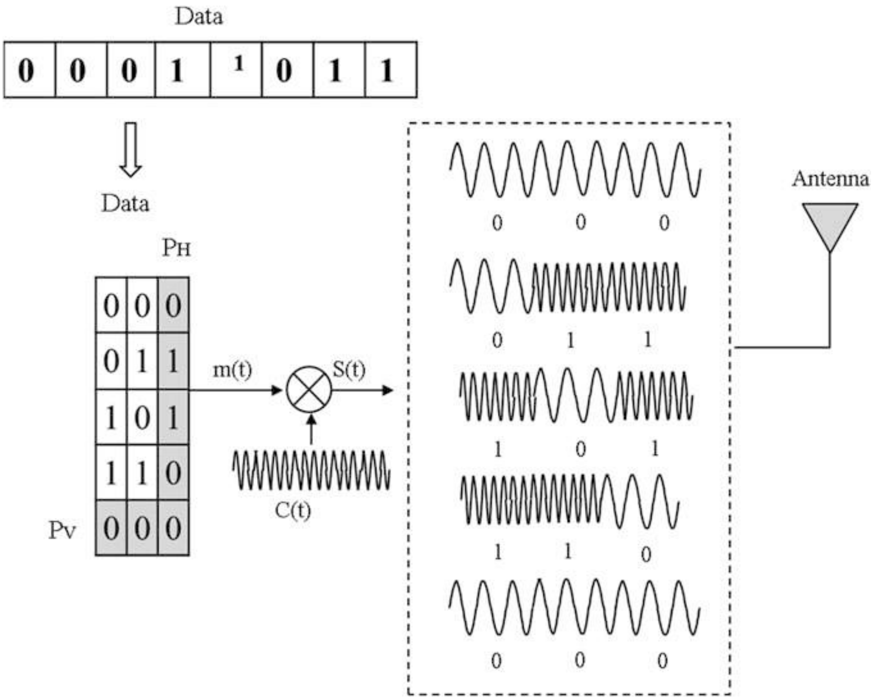

Fig. 5.2 Binary frequency shift keying (BFSK) modulation. The input encoded data block is transmitted row by row. The frequency of the carrier changes in accordance with the input digital signal

$(N + 1) = n = 15$, which is then FSK-modulated and transmitted row by row. The rate of this coding scheme is given by:

$$\text{Code Rate} : r = (MN)/[(M+1)(N+1)] = (4 \times 2)/5 \times 3) = 8/15 \qquad (5.1)$$

The coded bit rate R_{b2} is given by:

$$R_{b2} = \text{Uncoded Bit Rate}/\text{Code Rate} = R_{b1}/r = R_{b1}(15/8) \qquad (5.2)$$

Next, the coded bits are modulated by means of the FSK modulator as shown in the figure. Here,

- The input digital signal is the encoded bit sequence we want to transmit.
- Carrier is the radio frequency without modulation.
- Output is the FSK-modulated carrier, which has two frequencies corresponding to the binary input signal.
- For binary signal 1, the carrier changes to $f_c - \Delta f$.
- For binary signal 0, the carrier changes to $f_c + \Delta f$.
- The total frequency deviation is $2\Delta f$.

As shown in Fig. 3.8, the frequency of the carrier changes in discrete levels, in accordance with the input signals. We have:

$$\begin{array}{lll} \bullet & \text{Input Data:} & m(t) = 0 \text{ or } 1 \\ \bullet & \text{Carrier Frequency:} & C(t) = A\cos(\omega t) \\ \bullet & \text{Modulated Carrier:} & S(t) = A\cos(\omega - \Delta\omega)t, \text{For } m(t) = 1 \\ & & S(t) = A\cos(\omega + \Delta\omega)t, \text{For } m(t) = 0 \end{array} \tag{5.3}$$

where

$$\begin{array}{ll} \bullet & A = \text{Frequency of the carrier} \\ \bullet & \omega = \text{Nominal frequency of the carrier frequency} \\ \bullet & \Delta\omega = \text{Frequency deviation.} \end{array} \tag{5.4}$$

5.3 Frequency Shift Keying (FSK) Demodulation

Once the modulated binary data has been transmitted, it needs to be received and demodulated. This is often accomplished by the use of band-pass filters. In the case of binary FSK, the receiver needs to utilize two band-pass filters that are tuned to the appropriate frequencies. Since the nominal carrier frequency and the frequency deviation are known, this is relatively straightforward. One band-pass filter will be centered at the frequency ω_1 and the other at ω_2. As the signal enters into the receiver, it passes through the respective filter and the corresponding bit value is made. This is shown in Fig. 5.3. In order to assure that the bits are decoded correctly, the frequency deviation needs to be chosen with the limitations of the filters in mind to eliminate crossover.

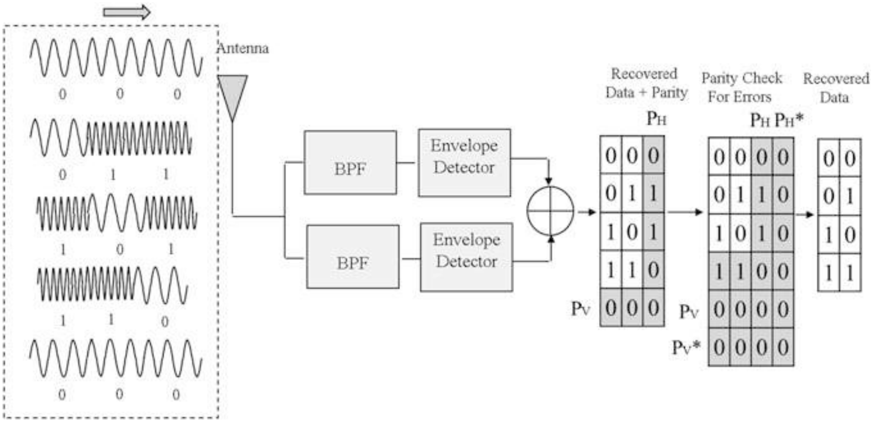

Fig. 5.3 Binary FSK detector utilizing two matched band-pass filters

5.4 FSK Bandwidth

In wireless communications, the scarcity of RF spectrum is well known. For this reason, we have to be vigilant about using transmission bandwidth in error control coding and modulation. The transmission bandwidth depends on the following:

- Spectral response of the encoded data
- Spectral response of the carrier frequency and
- Modulation type.

5.4.1 Spectral Response of the Encoded Data

In digital communications, data is generally referred to as a non-periodic digital signal. It has two values:

- Binary 1 = high, period = T
- Binary 0 = low, period = T

Also, data can be represented in two ways:

- Time-domain representation and
- Frequency-domain representation

The time-domain representation (Fig. 5.4a), known as non-return-to-zero (NRZ), is given by:

$$\begin{aligned} V(t) &= V \quad < 0 < t\ < T \\ &= 0 \quad ;\text{elsewhere} \end{aligned} \tag{5.5}$$

The frequency-domain representation is given by “Fourier transform”:

$$V(\omega) = \int_0^T V \cdot e^{-j\omega t} dt \tag{5.6}$$

$$|V(\omega)| = VT\left[\frac{\sin(\omega T/2)}{\omega T/2}\right] \tag{5.7}$$

$$P(\omega) = \left(\frac{1}{T}\right)|V(\omega)|^2 = V^2 T\left[\frac{\sin(\omega T/2)}{\omega T/2}\right]^2$$

Here, $P(\omega)$ is the power spectral density. This is plotted in Fig. 5.4b. The main lobe corresponds to the fundamental frequency, while side lobes correspond to harmonic components. The bandwidth of the power spectrum is proportional to the

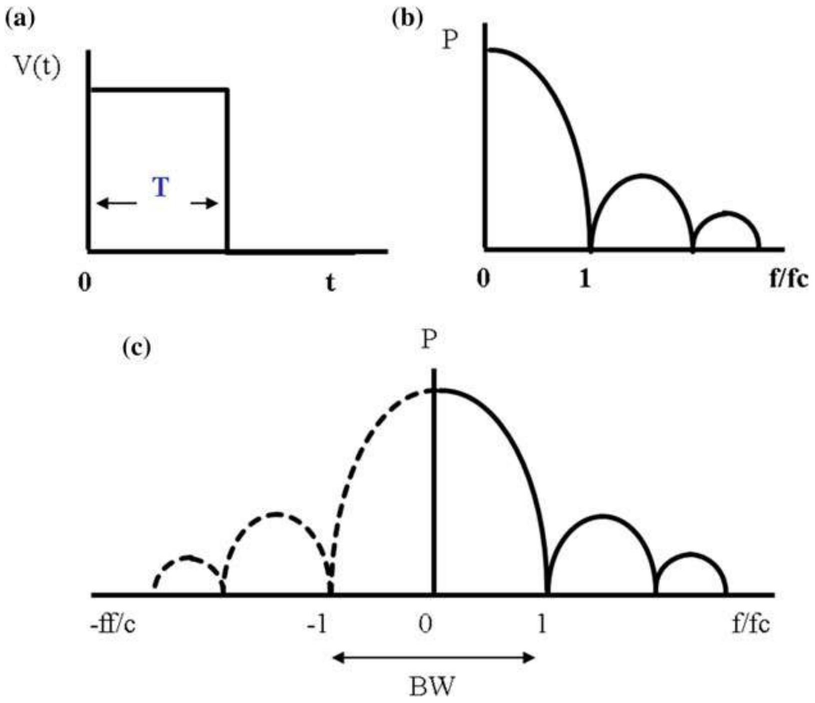

Fig. 5.4 **a** Discrete-time digital signal, **b** its one-sided power spectral density, and **c** two-sided power spectral density. The bandwidth associated with the non-return-to-zero (NRz) data is $2R_b$, where R_b is the bit rate

frequency. In practice, the side lobes are filtered out, since they are relatively insignificant with respect to the main lobe. Therefore, the one-sided bandwidth is given by the ratio $f/f_b = 1$. In other words, the one-sided bandwidth $= f = f_b$, where $f_b = R_b = 1/T$, T being the bit duration.

The general equation for two-sided response is given by:

$$V(\omega) = \int_{-\propto}^{\propto} V(t) \cdot e^{-j\omega t} dt$$

In this case, $V(\omega)$ is called the two-sided spectrum of $V(t)$. This is due to both positive and negative frequencies used in the integral. The function can be either a voltage or a current. Figure 5.4c shows the two-sided response, where the bandwidth is determined by the main lobe as shown below:

$$\text{Two sided bandwidth (BW)} = 2R_b(R_b = \text{Bit rate before coding}) \tag{5.8}$$

Important Notes:

1. If R_b is the bit rate before coding, and if the data is NRZ, then the bandwidth associated with the raw data will be $2R_b$. For example, if the bit rate before coding is 10 kb/s, then the bandwidth associated with the raw data will be 2×10 kb/s = 20 kHz.
2. If R_b is the bit rate before coding, code rate is r, and if the data is NRZ, then the bit rate after coding will be R_b (coded) = R_b(uncoded)r. The corresponding bandwidth associated with the coded data will be $2R_b$ (coded) = $2R_b$ (uncoded)/r. For example, if the bit rate before coding is 10 kb/s and the code rate $r = 1/2$, the coded bit rate will be R_b (coded) = R_b (uncoded)/r = 10/0.5 = 20 kb/s. The corresponding bandwidth associated with the coded data will be $2 \times 20 = 40$ kHz.

5.4.2 Spectral Response of the Carrier Frequency Before Modulation

A carrier frequency is essentially a sinusoidal waveform, which is periodic and continuous with respect to time. It has one frequency component. For example, the sine wave is described by the following time-domain equation:

$$V(t) = V_p \sin(\omega t_c) \tag{5.9}$$

where

$$V_P = \text{Peak voltage}$$

- $\omega_c = 2\pi f_c$
- f_c = Carrier frequency in Hz

Figure 5.5 shows the characteristics of a sine wave and its spectral response. Since the frequency is constant, its spectral response is located in the horizontal axis and the peak voltage is shown in the vertical axis. The corresponding bandwidth is zero.

5.4.3 FSK Bandwidth at a Glance

In FSK, the frequency of the carrier changes in two discrete levels, in accordance with the input signals. We have:

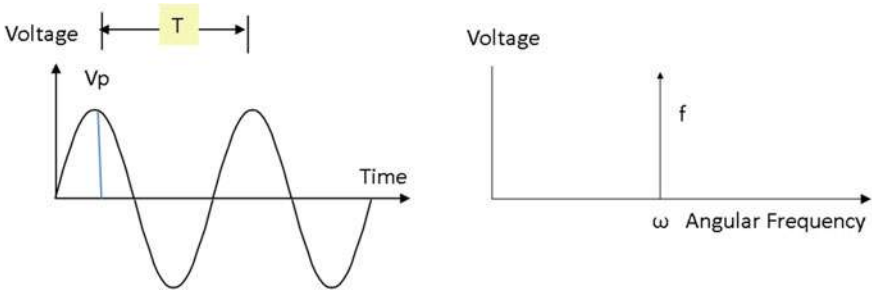

Fig. 5.5 High-frequency carrier response

- Input Data: $m(t) = 0 \text{ or } 1$
- Carrier Frequency: $C(t) = A\cos(\omega t)$
- Modulated Carrier: $S(t) = A\cos(\omega - \Delta\omega)t, \text{For } m(t) = 1$
  $S(t) = A\cos(\omega + \Delta\omega)t, \text{For } m(t) = 0$

where

- $S(t)$ = The modulated carrier
- A = Amplitude of the carrier
- ω = Nominal frequency of the carrier
- $\Delta\omega$ = Frequency deviation

The spectral response is depicted in Fig. 5.6. Notice that the carrier frequency after FSK modulation varies back and forth from the nominal frequency f_c by $\pm\Delta f_c$, where Δf_c is the frequency deviation. The FSK bandwidth is given by:

$$\begin{aligned} \mathrm{BW} &= 2(f_b + \Delta f_c) \\ &= 2f_b(1 + \Delta f_c/f_b) \\ &= 2f_b(1 + \beta) \end{aligned} \tag{5.10}$$

where $\beta = \Delta f/f_b$ is known as the modulation index and f_b is the coded bit frequency (bit rate R_b). The above equation is also known as "Carson's rule."

Problem

Given:

- Bit rate before coding: $R_{B1} = 10$ kb/s
- Code rate: $r = 8/15$
- Modulation: FSK
- Modulation index $\beta = 1$

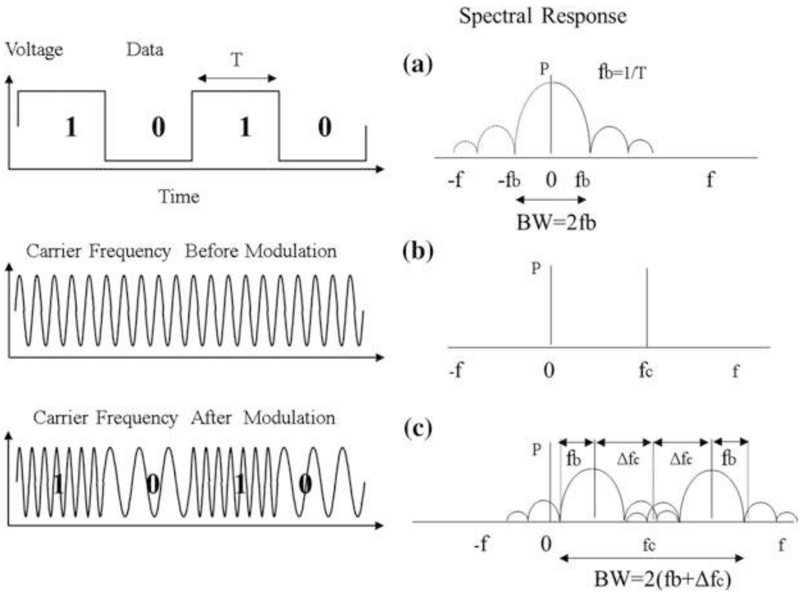
Spectral Response
Voltage
Data
T
1
0
1
0
Time
(a)
P
fb=1/T
-f
-fb
0
fb
f
BW=2fb
Carrier Frequency Before Modulation
(b)
P
-f
0
fc
f
Carrier Frequency After Modulation
1
0
1
0
(c)
P
fb
Δfc
Δfc
fb
-f
0
fc
f
BW=2(fb+Δfc)

Fig. 5.6 FSK bandwidth at a glance. **a** Spectral response of NRZ data before modulation. **b** Spectral response of the carrier before modulation. **c** Spectral response of the carrier after modulation. The transmission bandwidth is $2(f_b + \Delta f_c)$. f_b where f_b is the bit rate and Δf_c is the frequency deviation $=1/f_b$ is the bit duration for NRZ data

Find:

(a) the bit rate after coding: R_{B2}
(b) Transmission bandwidth: BW

Solution

(a) Bit rate after coding (R_{B2}):
$R_{B2} = R_{B1}/r = 10$ kb/s $(15/8) = 18.75$ kb/s
(b) Transmission BW $= 2\ R_{B2}(1+\beta)$
$= 2 \times 18.75$ kb/s $(1 + 1) = 75$ kHz.

5.5 BER Performance

It is well known that an (n, k) bock code, where k = number of information bits and n = number of coded bits, can correct t errors [4, 5]. A measure of coding gain is then obtained by comparing the uncoded word error WER_U, to the coded word WER_C. We examine this by means of the following analytical means.

Let the uncoded word error be defined as (WER_U). Then, with FSK modulation, the uncoded BER will be given by:

$$\mathrm{BER}_U = 0.5\,\mathrm{EXP}(-E_B/2N_0) \tag{5.11}$$

The probability that the uncoded word (WER_U) will be received in error is 1 minus the product of the probabilities that each bit will be received correctly. Thus, we write:

$$\mathrm{WER}_U = 1-(1-\mathrm{BER}_U)^k \tag{5.12}$$

Let the coded word error be defined as (WER_C). Since $n > k$, the ratio of coded bit energy to noise will be modified to E_c/N_0, where $E_c/N_0 = E_b/N_0 + 10\log(k/n)$. Therefore, the coded BER_c will be:

$$\mathrm{BER}_C = 0.5\,\mathrm{EXP}(-E_c/2N_0) \tag{5.13}$$

The corresponding coded word error rate is:

$$\mathrm{WER}_C = \sum_{k=t+1}^{n} \binom{n}{k} \mathrm{BER}_C^K (1 - \mathrm{BER}_C)^{n-k} \tag{5.14}$$

When $\mathrm{BER}_c < 0.5$, the first term in the summation is the dominant one; therefore, equation can be simplified as

$$\mathrm{WER}_C \approx \binom{n}{k} \mathrm{BER}_c^k (1 - \mathrm{BER}_c)^{n-k} \tag{5.15}$$

Using (15, 8) block code ($n = 15$, $k = 8$, $t = 1$), we obtain the coded and the uncoded WER as shown in Fig. 5.7. Coding gain is the difference in E_b/N_0 between the two curves. Notice that at least 3–4-db coding gain is available in this example where $r = k/n = 8/15$.

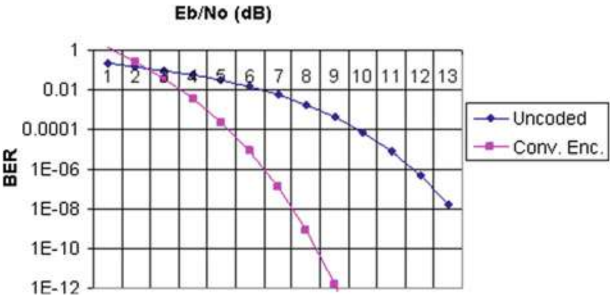

Fig. 5.7 Typical error performance in AWGN

5.6 Conclusions

- This chapter presents FSK modulation and its attributes.
- Numerous illustrations are provided to show how frequency of the carrier changes in discrete levels in accordance with the input digital signal, while the amplitude of the carrier remains the same.
- The Fourier transform is used to derive the spectral components, and FSK bandwidth is calculated.
- Bit error rate (BER) performance is presented.
- These materials have been augmented by diagrams and associated waveforms to make them easier for readers to grasp.

References

1. Smith DR (1985) Digital transmission systems, Van Nostrand Reinhold Co. ISBN: 0442009178
2. Leon W, Couch II (2001) Digital and analog communication systems, 7th edn. Prentice-Hall Inc., Englewood Cliffs. ISBN 0-13-142492-0
3. Sklar B (1988) Digital communications fundamentals and applications. Prentice Hall, Englewood Cliffs
4. Faruque S (2016) Radio frequency channel coding made easy. Springer. ISBN: 978-3-319-21169-5
5. Clark GC et al (1981) Error correction coding for digital communications. Plenum press, New York
6. Ungerboeck G (1982) Channel coding with multilevel/multiphase signals. IEEE Trans Inf Theory IT28:55–67
7. Lin S, Costello DJ Jr (1983) Error control coding: fundamentals and applications. Prentice-Hall, Inc., Englewood Cliffs
8. Blahut RE (1983) Theory and practice of error control codes. Addison-Wesley, Reading
9. van Lint JH (1992) Introduction to coding theory. GTM 86, 2nd edn. Springer, p 31. ISBN 3-540-54894-7
10. Mac Williams FJ, Sloane NJA (1977) The theory of error-correcting codes. North-Holland, p 35. ISBN 0-444-85193-3
11. Huffman W, Pless V (2003) Fundamentals of error-correcting codes. Cambridge University Press. ISBN 978-0-521-78280-7
12. Ebert PM, Tong SY (1968) Convolutional Reed-Soloman codes. Bell Syst Tech 48:729–742
13. Gallager RG (1968) Information theory and reliable communications. Wiley, New York
14. Kohlenbero A, Forney GD Jr (1968) Convolutional coding for channels with memory. IEEE Trans Inf Theory IT-14:618–626
15. Reddy SM (1968) Further results on convolutional codes derived from block codes. Inf Control 13:357–362
16. Reddy SM (1968) A class of linear convolutional codes for compound channels. Technical report, Bell Telephone Laboratories, Holmdel, New Jersey

Chapter 6
Phase Shift Keying (PSK)

Topics

- Introduction
- Binary Phase Shift Keying (BPSK) Modulation
- QPSK Modulation
- 8PSK Modulation
- 16PSK Modulation
- PSK spectrum and Bandwidth
- SSB Spectrum and Bandwidth

6.1 Introduction

In phase shift keying (PSK), the phase of the carrier changes in discrete levels in accordance with the input digital signal, while the amplitude of the carrier remains the same. This is shown in Fig. 6.1, where

- $m(t)$ is the input modulating digital signal,
- $C(t)$ is the carrier frequency, and
- $S(t)$ is the PSK-modulated carrier frequency.

As shown in the figure, the digital binary signal changes the phase of the carrier on two discrete levels. This enables the receiver to extract the digital signal by demodulation. Notice that the phase of the carrier changes in accordance with the input signal, while the amplitude of the carrier does not change after modulation. However, it can be shown that the modulated carrier $S(t)$ contains several spectral components, requiring frequency domain analysis.

In the following sections, the above disciplines in PSK modulation will be presented, along with the respective spectrum and bandwidth. These materials have been augmented by diagrams and associated waveforms to make them easier for readers to grasp.

© The Author(s) 2017

S. Faruque, *Radio Frequency Modulation Made Easy*,
SpringerBriefs in Electrical and Computer Engineering,
DOI 10.1007/978-3-319-41202-3_6

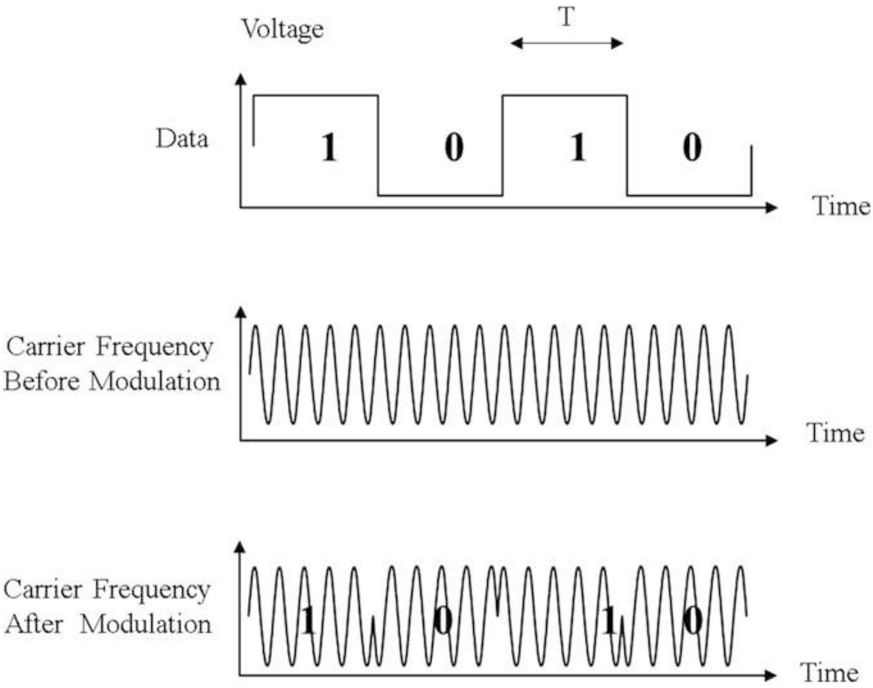
Voltage
T
Data
1
0
1
0
Time
Carrier Frequency Before Modulation
Time
Carrier Frequency After Modulation
1
0
1
0
Time

Fig. 6.1 Binary PSK (BPSK) waveforms. The phase of the carrier changes in accordance with the input digital signal. The amplitude of the carrier remains the same

6.2 Binary Phase Shift Keying (BPSK)

6.2.1 BPSK Modulation

Phase shift keying (PSK) is a method of digital modulation that utilizes phase shifting of the relative phase content of the signal [1–3]. The signal to be modulated and transmitted is binary, which is encoded before modulation. This is an indispensable task in digital communications, where redundant bits are added with the raw data that enable the receiver to detect and correct bit errors, if they occur during transmission [4–16]. While there are many error-coding schemes available, we will use a simple coding technique, known as “Block Coding” to illustrate the concept.

Figure 6.2 shows an encoded BPSK modulation scheme using (15, 8) block code where an 8-bit data block is formed as M-rows and N-columns ($M = 4$, $N = 2$). The product $MN = k = 8$ is the dimension of the information bits before coding. Next, a horizontal parity P_H is appended to each row and a vertical parity P_V is appended to each column. The resulting augmented dimension is given by the product $(M + 1)\ (N + 1) = n = 15$, which is then PSK modulated and transmitted row by row. The rate of this coding scheme is given by

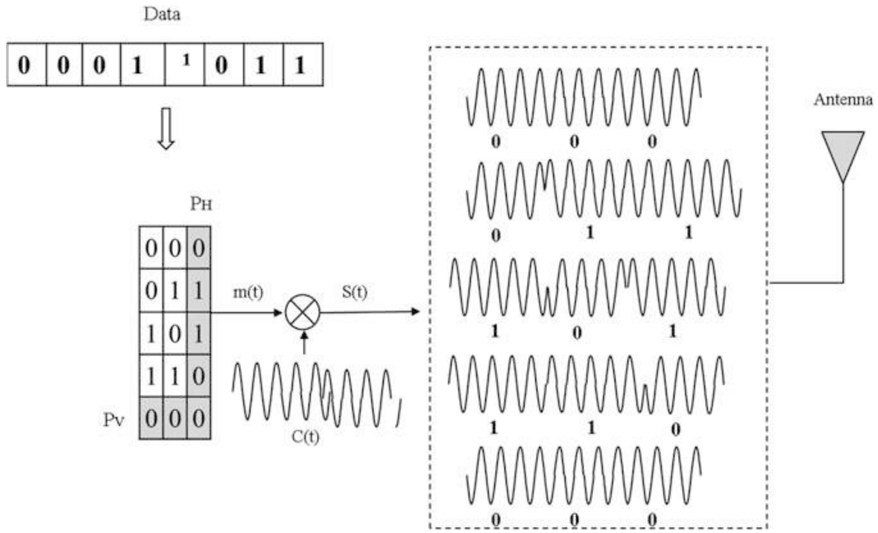

Fig. 6.2 Binary phase shift keying (BPSK) modulation. The input encoded data block is transmitted row by row. The phase of the carrier changes in accordance with the input digital signal

$$\text{Code Rate} : r = (MN)/[(M+1)(N+1)] = (4 \times 2)/5 \times 3) = 8/15 \tag{6.1}$$

The coded bit rate R_{b2} is given by:

$$R_{b2} = \text{Uncoded Bit Rate}/\text{Code Rate} = R_{b1}/r = R_{b1}(15/8) \tag{6.2}$$

Next, the coded bits are modulated by means of the PSK modulator as shown in the figure. Here,

- The Input digital signal is the encoded bit sequence we want to transmit;
- Carrier is the radio frequency without modulation;
- Output is the PSK-modulated carrier, which has two phases corresponding to the binary input signals;
- For binary signal 0, $\varphi = 0^\circ$; and
- For binary signal 1, $\varphi = 180^\circ$.

As shown in Fig. 6.2, the phase of the carrier changes in two discrete levels, in accordance with the input signals. Here, we have the following:

- Input Data : $m(t) = 0 \text{ or } 1$
- Carrier Frequency : $C(t) = A\cos(\omega_{ct})$
- Modulated Carrier : $S(t) = A_c \cos[\omega_c t + 2\pi/M)m(t)] \quad m(t) = 0,\ 1,\ 2,\ 3, \ldots M-1$ (6.3)

where,

- A_c = Amplitude of the carrier frequency
- ω_c = Angular frequency of the carrier
- $M = 2, 4, 8, 16, \ldots$
- In BPSK, there are two phases 1 bit/phase($M = 2$) (6.4)
- In QPSK, there are four phases, 2-bits/phase, $M = 4$
- In 8PSK, there are 8 phases, 3-bits/phase, $M = 8$
- In 16PSK, there are 16 phases, 4-bits/phase, $M = 16$

We can also represent the BPSK modulator as a signal constellation diagram with $M = 2$, 1 bit per phase. This is shown in Fig. 6.3, where the input raw data, having a bit rate R_{b1}, is encoded by means of a rate r encoder. The encoded data, having a bit rate R_{b2}, = $R_{b1}r$ ($r < 1$), is modulated by the BPSK modulator as shown in Fig. 6.3.

The BPSK modulator takes one bit at a time to construct the phase constellation having two phases, also known as "Symbols," where each symbol represents one bit. The symbol rate is therefore the same as the encoded bit rate R_{b2}.

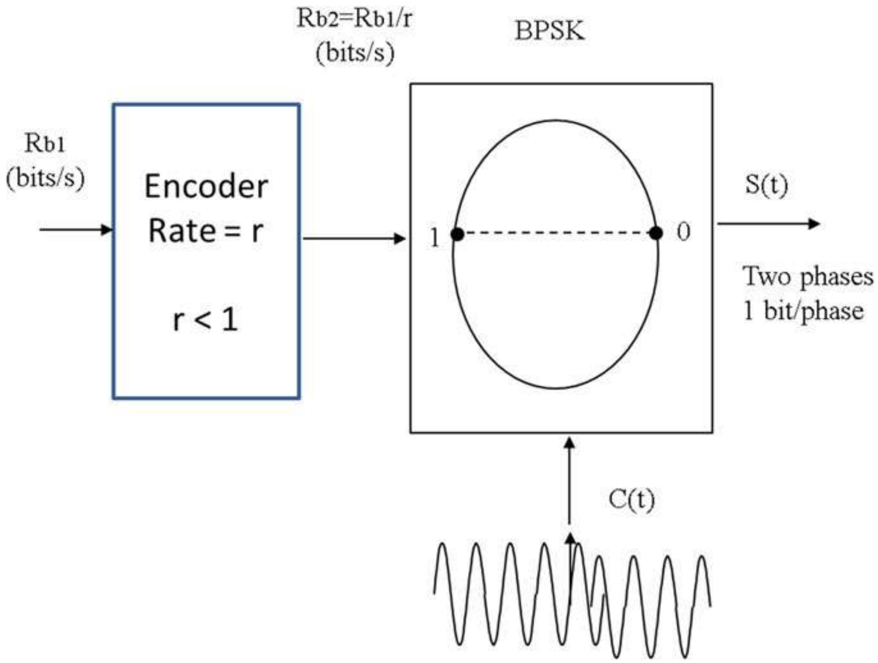

Fig. 6.3 BPSK signal constellation having 2 symbols, 1-bit per symbol

Therefore, the BPSK modulator has the following specifications:

- 2 phases or 2 symbols;
- 1-bit/symbol,

The above specifications govern the transmission bandwidth, as we shall see later.

6.2.2 BPSK Demodulation

Once the modulated binary data has been transmitted, it needs to be received and demodulated. This is often accomplished with the use of a phase detector, typically known as phase-locked loop (PLL). As the signal enters the receiver, it passes through the PLL. The PLL locks the incoming carrier frequency and tracks the variations in frequency and phase. This is known as coherent detection technique, where the knowledge of the carrier frequency and phase must be known to the receiver.

Figure 6.4 shows a simplified diagram of a BPSK demodulator along with the data recovery process. In order to assure that the bits are decoded correctly, the phase deviation needs to be chosen with the limitations of the PLL in mind to eliminate crossover.

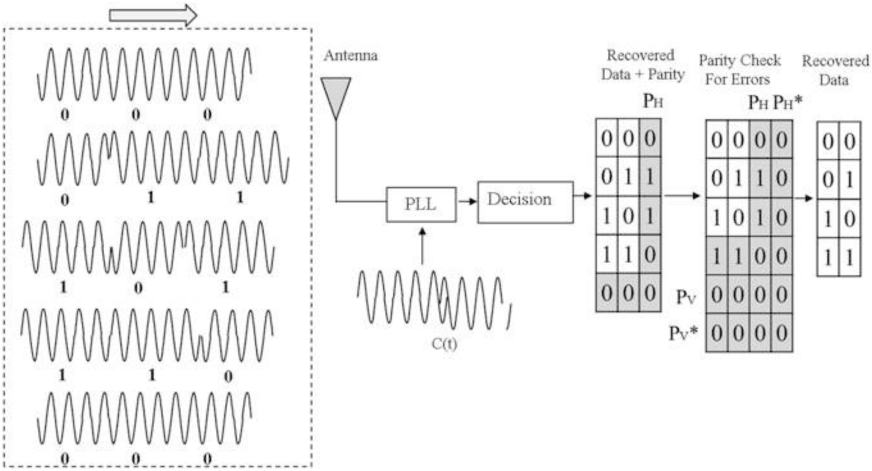
0 0 0
0 1 1
1 0 1
1 1 0
0 0 0
Antenna
PLL
Decision
C(t)
Recovered Data + Parity
PH
0 0 0
0 1 1
1 0 1
1 1 0
0 0 0
Parity Check For Errors
PH PH*
0 0 0 0
0 1 1 0
1 0 1 0
1 1 0 0
PV 0 0 0 0
PV* 0 0 0 0
Recovered Data
0 0
0 1
1 0
1 1

Fig. 6.4 Binary PSK detector showing data recovery process

6.3 QPSK Modulation

In QPSK, the input raw data, having a bit rate R_{b1}, is encoded by a rate r ($r < 1$) encoder. The encoded data, having a bit rate $R_{b2} = R_{b1}/r$, is serial to parallel converted into two parallel streams. The encoded bit rate, now reduced in speed by a factor of two, is modulated by the QPSK modulator as shown in Fig. 6.5.

The QPSK modulator takes one bit from each stream to construct the phase constellation having four phases, also known as "Symbols," where each symbol represents two bits. The symbol rate is therefore reduced by a factor of two. The QPSK modulator has four phases or 4 symbols, 2-bits/symbol as shown in the figure.

Therefore, the QPSK modulator has the following specifications:

- 4 phases or 4 symbols
- 2-bits/symbol

The above specifications govern the transmission bandwidth, as we shall see later.

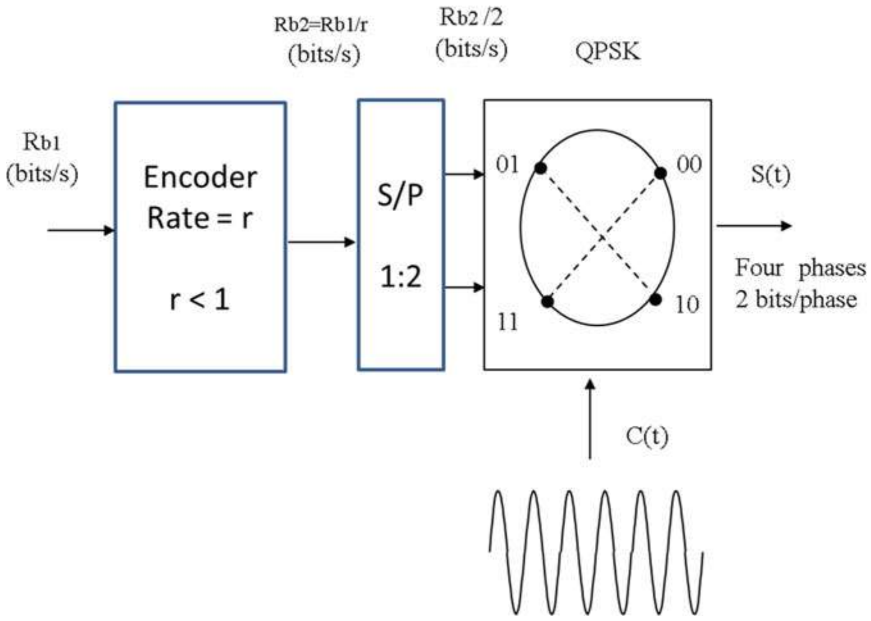
Rb2=Rb1/r
(bits/s)
Rb2 /2
(bits/s)
QPSK
Rb1
(bits/s)
Encoder
Rate = r
r < 1
S/P
1:2
01
00
11
10
S(t)
Four phases
2 bits/phase
C(t)

Fig. 6.5 QPSK signal constellation having 4 symbols, 2-bits per symbol

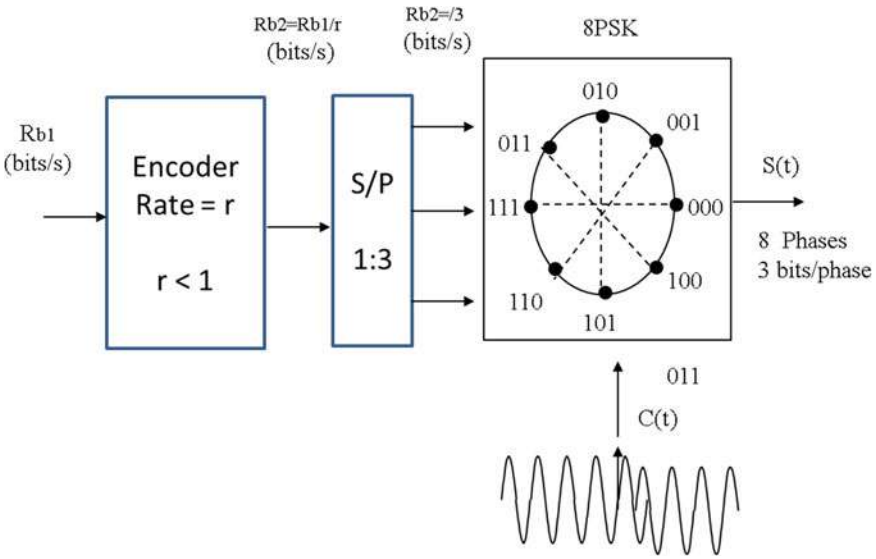

Fig. 6.6 8PSK signal constellation having 8 symbols, 3-bits per symbol

6.4 8PSK Modulation

In 8PSK, the input raw data, having a bit rate R_{b1}, is encoded by a rate r ($r < 1$) encoder. The encoded data, having a bit rate $R_{b2} = R_{b1}/r$, is serial to parallel converted into three parallel streams. The encoded bit rate, now reduced in speed by a factor of three, is modulated by the 8PSK modulator as shown in Fig. 6.6.

The 8PSK modulator takes one bit from each stream to construct the phase constellation having 8 phases, also known as "Symbols," where each symbol represents 3-bits. The symbol rate is therefore reduced by a factor of 3. The 8PSK modulator has 8 phases or 8 symbols, 3-bits/symbol as shown in the figure.

Therefore, the 8PSK modulator has the following specifications:

- 8 phases or 8 symbols;
- 3-bits/symbol.

The above specifications govern the transmission bandwidth, as we shall see later.

6.5 16PSK Modulation

In 16PSK, the input raw data, having a bit rate R_{b1}, is encoded by means of a rate r ($r < 1$) encoder. The encoded data, having a bit rate $R_{b2} = R_{b1}/r$, is serial to parallel converted into four parallel streams. The encoded bit rate, now reduced in speed by a factor of four, is modulated by the 16PSK modulator as shown in Fig. 6.7.

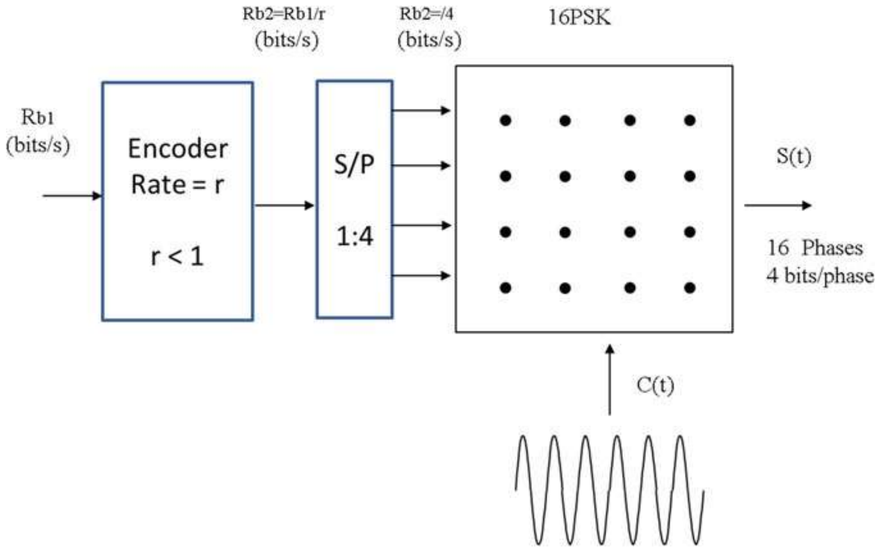

Fig. 6.7 16PSK signal constellation having 16 symbols, 4-bits per symbol. Here, each symbol is represented by a *dot*, where each dot represents 4-bits

The 16PSK modulator takes one bit from each stream to construct the phase constellation having 16 phases, also known as “Symbols,” where each symbol represents four bits. The symbol rate is therefore reduced by a factor of four. Therefore, the 16PSK modulator has 16 phases or 16 symbols, 4-bits/symbol as shown in the figure.

Therefore, the 16PSK modulator has the following specifications:

- 16 phases or 16 symbols;
- 4-bits/symbol.

The above specifications govern the transmission bandwidth, as we shall see later.

Table 6.1 shows the number of phases and the corresponding bits per phase for MPSK modulation schemes for $M = 2, 4, 8, 16, 32, 64$, etc.

Table 6.1 MPSK modulation parameters. $M = 2, 4, 8, 16$, and 32

Modulation	Number of phases φ	Number of bits per phase
BPSK	2	1
QPSK	4	2
8PSK	8	3
16	16	4
32	32	5
64	64	6
:	:	:

6.6 PSK Spectrum and Bandwidth

In wireless communications, the scarcity of RF spectrum is well known. For this reason, we have to be vigilant about using transmission bandwidth in error control coding and modulation. The transmission bandwidth depends on:

- Spectral response of the encoded data;
- Spectral response of the carrier phase; and
- Modulation type.

Let us take a closer look:

6.6.1 Spectral Response of the Encoded Data

In digital communications, data is generally referred to as a non-periodic digital signal. It has two values:

- Binary-1 = High, Period = T
- Binary-0 = Low, Period = T

Also, data can be represented in two ways:

- Time domain representation and
- Frequency domain representation

The time domain representation (Fig. 6.8a), known as non-return-to-zero (NRZ), is given by:

$$\begin{aligned} V(t) &= V \quad < 0 < t < T \\ &= 0 \quad \text{elsewhere} \end{aligned} \tag{6.5}$$

The frequency domain representation is given by “Fourier Transform”:

$$V(\omega) = \int_{0}^{T} V \cdot e^{-j\omega t} dt \tag{6.6}$$

$$\begin{aligned} |V(\omega)| &= 2\left(\frac{V}{\omega}\right)\sin\left(\frac{\omega T}{2}\right) = VT\left[\frac{\sin(\omega T/2)}{\omega T/2}\right] \\ P(\omega) &= \left(\frac{1}{T}\right)|V(\omega)|^2 = V^2 T\left[\frac{\sin(\omega T/2)}{\omega T/2}\right]^2 \end{aligned} \tag{6.7}$$

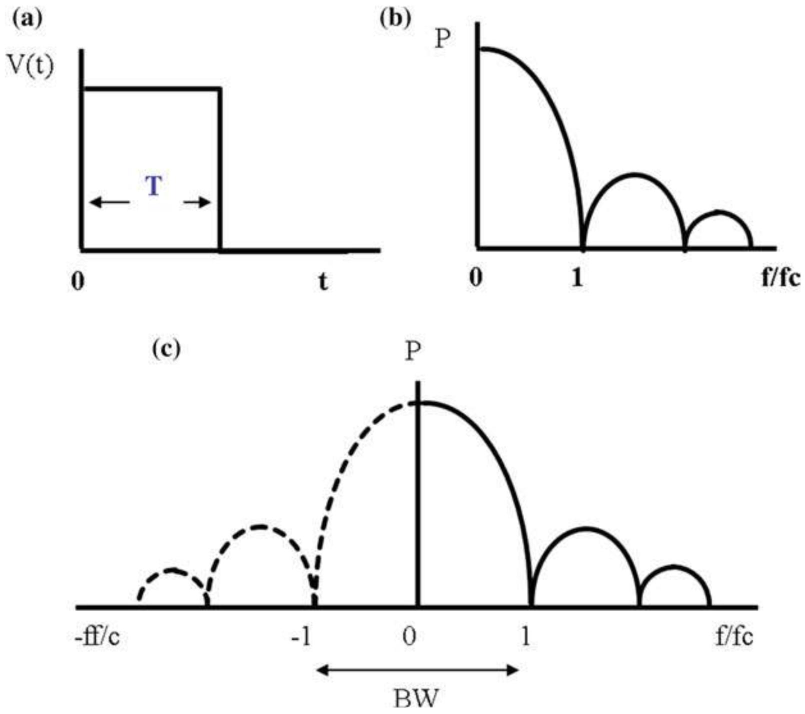

Fig. 6.8 **a** Discrete time digital signal **b** it is one-sided power spectral density and **c** two-sided power spectral density. The bandwidth associated with the non-return-to-zero (NRz) data is $2R_b$, where R_b is the bit rate

Here, $P(\omega)$ is the power spectral density. This is plotted in Fig. 6.8b. The main lobe corresponds to the fundamental frequency, while the side lobes correspond to harmonic components. The bandwidth of the power spectrum is proportional to the frequency. In practice, the side lobes are filtered out, since they are relatively insignificant with respect to the main lobe. Therefore, the one-sided bandwidth is given by the ratio $f/fb = 1$. In other words, the one-sided bandwidth $= f = f_b$, where $f_b = R_b = 1/T$, T being the bit duration.

The general equation for two-sided response is given by:

$$V(\omega) = \int_{-\propto}^{\propto} V(t) \cdot e^{-j\omega t} dt$$

In this case, $V(\omega)$ is called the two-sided spectrum of $V(t)$. This is due to both positive and negative frequencies used in the integral. The function can be either a voltage or a current. Figure 6.8c shows the two-sided response, where the bandwidth is determined by the main lobe as shown below:

$$\text{Two-sided bandwidth (BW)} = 2R_b\ (R_b = \text{Bit rate before coding}) \tag{6.8}$$

6.6.2 Spectral Response of the Carrier Before Modulation

A carrier frequency is a sinusoidal waveform, which is periodic and continuous with respect to time. It has one phase component. For example, the sine wave is described by the following time domain equation:

$$V(t) = V_p \sin(\omega_c t) \tag{6.9}$$

Where

$$Vp = \text{Peak voltage}$$

- $\omega_c = 2\pi f_c$
- f_c = Carrier phase in Hz

Figure 6.9 shows the characteristics of a sine wave and its spectral response. Since the phase is constant, its spectral response is located in the horizontal axis and the peak voltage is shown in the vertical axis. The corresponding bandwidth is zero.

6.6.3 BPSK Spectrum

In BPSK, the input is a digital signal and it contains an infinite number of harmonically related sinusoidal waveforms. This is given by (see Sect. 6.6.1):

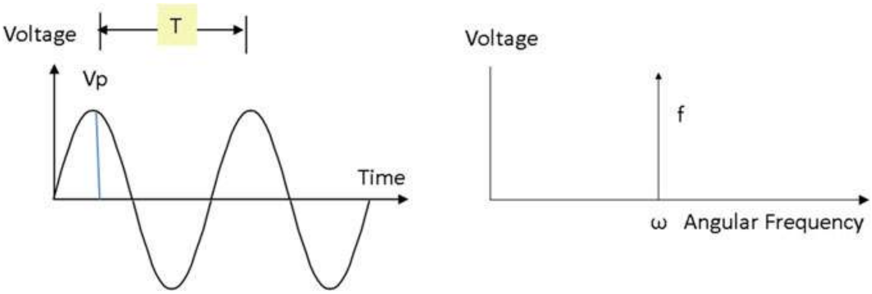

Fig. 6.9 High-frequency carrier response

$$|V(\omega)| = 2\left(\frac{V}{\omega}\right)\sin\left(\frac{\omega T}{2}\right) = VT\left[\frac{\sin(\omega T/2)}{\omega T/2}\right] \tag{6.10}$$

Here, $V(\omega)$ is the frequency domain representation of the input digital signal, which has a $\sin(x)/x$ response that governs the phase of the carrier frequency.

With $V(t) = m(t)$, we write the following as:

$$S(t) = A_c \cos[\omega_c t + \beta m(t)] \tag{6.11}$$

where β is the phase deviation index of the carrier and $m(t)$ has a $\sin(x)/x$ response, which is given by

$$m(t) = VT\left[\frac{\sin(\omega m T/2)}{\omega m T/2}\right] \tag{6.12}$$

Therefore, the spectral response after BPSK modulation also has a $\sin(x)/x$ response, which is the shifted version of the NRZ data, centered on the carrier frequency f_c, as shown in Fig. 6.10. The transmission bandwidth associated with the main lobe is given by

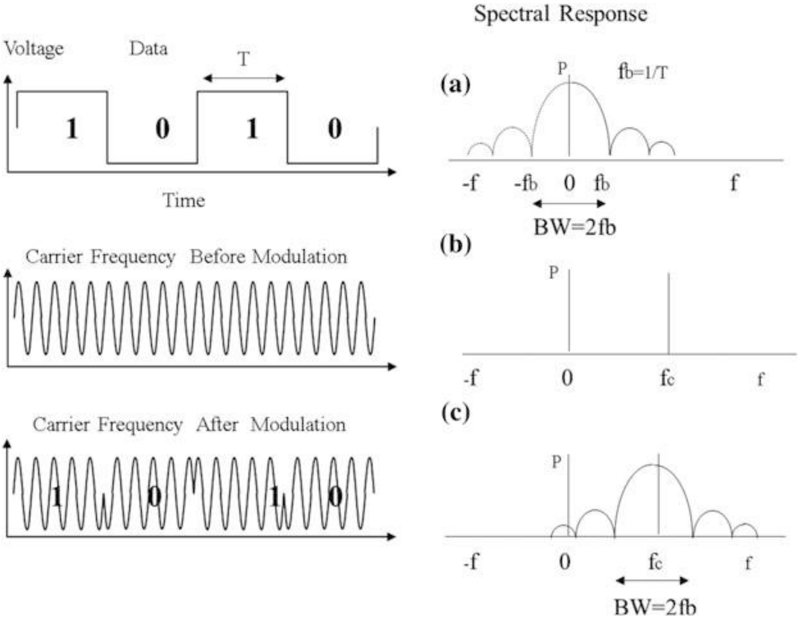
Spectral Response
Voltage
Data
T
1
0
1
0
Time
Carrier Frequency Before Modulation
Carrier Frequency After Modulation
1
0
1
0
(a)
P
fb=1/T
-f
-fb
0
fb
f
BW=2fb
(b)
P
-f
0
fc
f
(c)
P
-f
0
fc
f
BW=2fb

Fig. 6.10 BPSK bandwidth. **a** Spectral response of NRZ data before modulation. **b** Spectral response of the carrier before modulation. **c** Spectral response of the carrier after modulation

$$\begin{aligned} \text{BW (BPSK)} &\approx 2R_{b2}/\text{Bit per Phase} \\ &\approx 2R_{b2}/1 \approx 2R_{b2} \end{aligned} \tag{6.13}$$

where R_{b2} is the coded bit rate (bit frequency). Notice that the BPSK bandwidth is the same the ASK bandwidth. It may be noted that the higher-order spectral components are filtered out.

Problem 6.1
Given:

- Uncoded input bit rate: R_{b1} = 10 kb/s
- Code rate: r = 8/15
- Carrier frequency fc = 1 MHz
- Modulation: BPSK, QPSK, 8PSK, and 16PSK

Find:

(a) Coded bit rate R_{b2}
(b) BPSK bandwidth (BW)
(c) QPSK bandwidth (BW)
(d) 8PSK bandwidth (BW)
(e) 16PSK bandwidth (BW)

Solution:

(a) Coded bit rate: $R_{B2} = R_{B1}/r$ = 10 kb/(15/8) = 18.75 kb/s
(b) BPSK bandwidth: $BW = 2R_{B2}/1 = 2 \times 18.75 = 37.5$ kHz
(c) QPSK bandwidth: $BW = 2R_{B2}/2 = 2 \times 18.75/2 = 18.75$ kHz
(d) 8PSK bandwidth: $BW = 2R_{B2}/3 = 2 \times 18.75/3 = 12.5$ kHz
(e) 16PSK bandwidth: $BW = 2R_{B2}/4 = 2 \times 18.755/4 = 9.375$ kHz

NOTE: Higher-order PSK modulation is bandwidth efficient.

Problem 6.2
Given:

- Input bit rate R_{b1} = 10 kb/s
- Code rate r = 1/2
- Modulation: BPSK, QPSK, 8PSK, 16PSK

Find:

(a) Bit rate after coding R_{b2}
(b) Transmission bandwidth BW

Solution:

(a) $R_{b2} = R_{b1}/r$ = 10 kb/s × 2 = 20 kb/s
(b) Transmission bandwidth:

- BPSK $BW = 2R_{b2}$/Bits per Symbol = 2 × 20 kb/s/1 = 40 kHz
- QPSK $BW = 2R_{b2}$/Bits per symbol = 2 × 20 kb/s/2 = 20 kHz
- 8PSK $BW = 2R_{b2}$/Bits per symbol = 2 × 20 kb/s/3 = 13.33 kHz
- 16PSK $BW = 2R_{b2}$/Bits per symbol = 2 × 20 kb/s/4 = 10 kHz

NOTE: Higher-order PSK modulation is bandwidth efficient.

6.7 Conclusions

- This chapter presents PSK modulation and its attributes.
- Numerous illustrations are provided to show how phase of the carrier changes in discrete levels in accordance with the input digital signal, while the amplitude of the carrier remains the same.
- The Fourier transform is used to derive the spectral components and PSK bandwidth is calculated.
- These materials have been augmented by diagrams and associated waveforms to make them easier for readers to grasp.

References

1. Smith DR (1986) Digital transmission systems. Van Nostrand Reinhold Co. ISBN: 0442009178
2. Couch II LW (2001) Digital and analog communication systems, 7th edn. Prentice-Hall, Inc.: Englewood Cliffs. ISBN: 0-13-142492-0
3. Sklar B (1988) Digital communications fundamentals and applications. Prentice Hall
4. Faruque S (2016) Radio frequency channel coding made easy. Springer, ISBN: 978-3-319-21169-5
5. Clark GC et al (1981) Error correction coding for digital communications. Plenum press
6. Ungerboeck G (1982) Channel coding with multilevel/multiphase signals. IEEE Trans Inf Theory IT28:55–67
7. Lin S, Costello DJ Jr (1983) Error control coding: fundamentals and applications. Prentice-Hall, Inc., Englewood Cliffs
8. Blahut RE (1983) Theory and practice of error control codes. reading. Addison-Wesley, Massachusetts
9. van Lint JH (1992) Introduction to coding theory. GTM 86. 2nd ed. Springer. p 31. ISBN 3-540-54894-7
10. Mac Williams FJ, Sloane NJA (1977) The theory of error correcting codes. North-Holland. p 36. ISBN 0-444-85193-3
11. Huffman W, Pless V (2003) Fundamentals of error-correcting codes. Cambridge University Press. ISBN 978-0-521-78280-7
12. Ebert PM, Tong SY (1968) Convolutional Reed-Soloman codes, Bell Syst Tech, pp 729–742
13. Gallager RG (1968) Information theory and reliable communications. John Wiley, New York

14. Kohlenbero A, Forney GD Jr (1968) Convolutional coding for channels with memory, IEEE Trans. Information Theory IT-14, 618–626. 1968
15. M.A. Reddy, S. M, "Further results on convolutional codes derived from block codes. Inf Control 13:357–362
16. Reddy SM (1968) A class of linear convolutional codes for compound channels. Technical Report, Bell Telephone Laboratories, Holmdel, New Jersey

Chapter 7
N-Ary Coded Modulation

Topics

- Introduction
- N-ary Complementary Convolutional Coding and M-ary Modulation
- N-ary Complementary Orthogonal Coding and M-ary Modulation
- Coding Gain and Bandwidth
- Conclusions

7.1 Introduction

N-ary coded modulation is a multi-level channel coding and multi-level modulation technique, where instead of coding one bit at a time two or more bits are encoded simultaneously and then modulated by means of M-ary modulation. In its most basic construction, the input serial data is converted into several parallel streams. These parallel bit streams, now reduced in speed, are mapped into a bank of *N* unique convolutional or orthogonal codes. The coded information bits are then modulated by an M-ary PSK modulator and transmitted through a channel. At the receiver, the decoder recovers the data by means of code correlation. A lookup table at the receiver contains the input/output bit sequences. Upon receiving an encoded data pattern, the receiver validates the received data pattern by means of code correlation.

In this chapter, we will present the key concept, underlying principles and practical application of N-ary coded modulation schemes, offering spectrum efficiency with improved error correction capabilities. Construction of N-ary coded modulation schemes based on convolutional and orthogonal codes will be presented to illustrate the concept.

© The Author(s) 2017

S. Faruque, *Radio Frequency Modulation Made Easy*,
SpringerBriefs in Electrical and Computer Engineering,
DOI 10.1007/978-3-319-41202-3_7

7.2 N-Ary Convolutional Coding and M-Ary Modulation

7.2.1 Background

In convolutional coding, a sequence of data signals is transformed into a longer sequence that contains enough redundancy to protect the data [1–6]. This type of error control is also classified as forward error control coding (FECC), because these methods are often used to correct errors that are caused by channel noise. In a typical convolutional encoder, k information bits enter into the encoder sequentially. The convolutional encoder generates n parity bits as encoded bits ($n > k$). The code rate is defined as $r = k/n$. The coded information bits are modulated and transmitted through a channel.

Another class of convolutional codes, known as parallel concatenated codes or turbo codes [7–9], is popular in wireless communications because of its superior error control capabilities. Yet, turbo codes are limited to two convolutional codes connected in parallel, where delays are introduced due to concatenation and interleaving, thereby limiting high-speed data communications in wireless communications. In addition, it is also too complex to realize.

In this section, a modified technique, defined as N-ary complementary convolutional coding, along with M-ary modulation, is presented to overcome these problems. N-ary complementary convolutional coding is a multi-level convolutional code, where more than two convolutional codes can be used simultaneously to further enhance coding gain without concatenation and interleaving.

7.2.2 Generation of Complementary Convolutional Codes

In a typical convolutional encoder, k information bits enter into the encoder sequentially. The encoder generates n parity bits as encoded bits ($n > k$). The code rate is defined as $r = k/n$. The proposed N-ary convolutional codes are a block of n-convolutional codes and their complements.

In its most basic construction, as shown in Fig. 7.1, a typical convolutional encoder is taken as the basis to generate a block of n-convolutional codes. Next, each convolutional code is inverted to construct a block of n-antipodal codes. Therefore, a block of N-ary convolutional code has n-convolutional codes and n-antipodal codes, for a total of $2n = N$ complementary convolutional codes. In short, we shall define this as N-ary complementary convolutional codes.

Let us consider Fig. 7.1 once again and assume that the input bit sequence to the encoder is:

$$m(t) = 1\ 0\ 1 \tag{7.1}$$

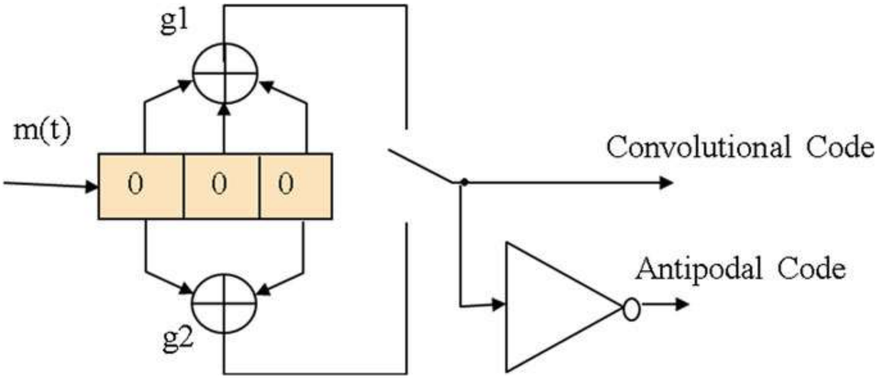

Fig. 7.1 Generation of N-ary complimentary convolutional codes

Which is described by the following polynomial?

$$\begin{aligned} m(X) &= 1 + 0X + 1X^2 \\ &= 1 + X^2 \end{aligned} \tag{7.2}$$

The encoder is described by the following generator polynomials:

$$\begin{aligned} g1(X) &= 1 + X + X^2 \\ g2(X) &= 1 + X^2 \end{aligned} \tag{7.3}$$

Then, the product of polynomials for the encoder can be described as follows:

$$\begin{aligned} m(X)g1(X) &= (1 + X^2)(1 + X + X^2) = 1 + X + X^3 + X^4 \\ m(X)g2(X) &= (1 + X^2)(1 + X^2) = 1 + X^4 \end{aligned} \tag{7.4}$$

with $X^2 + X^2 = 0$, the output bit sequence can be found as $U(X) = m(X)g1(\mathrm{X})$ multiplexed with $m(X)g2(X)$, where $m(X)$ is the input bit sequence [7]. We write the above two equations as follows:

$$\begin{array}{l} \mathrm{m(X)g1(X)} = 1 + 1\mathrm{X} + 0\mathrm{X}^2 + 1\mathrm{X}^3 + 1\mathrm{X}^4 \\ \mathrm{m(X)g2(X)} = 1 + 0\mathrm{X} + 0\mathrm{X}^2 + 0\mathrm{X}^3 + 1\mathrm{X}^4 \\ \hline \mathrm{U1U2} = (1,1) + (1,0) + (0,0) + (1,0) + (1,1) \end{array} \tag{7.5}$$

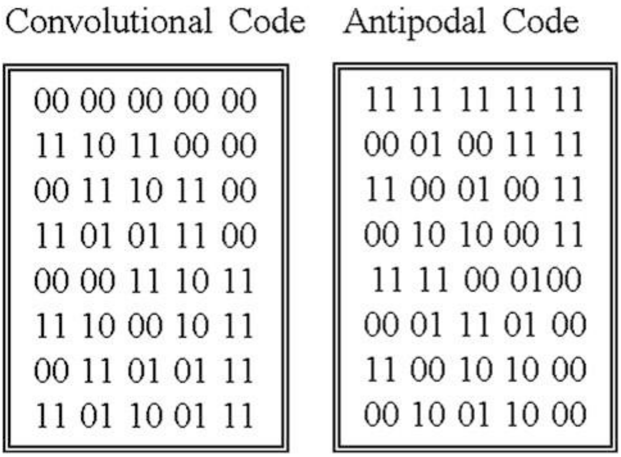

Convolutional Code	Antipodal Code
00 00 00 00 00	11 11 11 11 11
11 10 11 00 00	00 01 00 11 11
00 11 10 11 00	11 00 01 00 11
11 01 01 11 00	00 10 10 00 11
00 00 11 10 11	11 11 00 0100
11 10 00 10 11	00 01 11 01 00
00 11 01 01 11	11 00 10 10 00
11 01 10 01 11	00 10 01 10 00

Fig. 7.2 N-ary complimentary convolutional code blocks

Taking only the coefficients, we obtain the desired convolutional code sequence corresponding to the input bit sequence $m(T) = 101$ as follows:

$$\text{Convolutional Code} = 1\,1 \quad 1\,0 \quad 0\,0 \quad 1\,0 \quad 1\,1 \tag{7.6}$$

The corresponding antipodal code is obtained simply by inverting the convolutional code as shown in the figure. This is given by:

$$\text{Antipodal Code} = 0\,0 \quad 0\,1 \quad 1\,1 \quad 0\,1 \quad 0\,0 \tag{7.7}$$

As can be seen, each 3-bit data corresponds to a unique 10-bit convolutional code. Antipodal code is just the inverse of the convolutional codes. Since a 3-bit sequence has 8 combinations, the above convolutional encoder generates 8 unique convolutional codes and 8 unique antipodal codes, for a total of 16 complementary convolutional codes. These code blocks are displayed in Fig. 7.2.

7.2.3 2-Ary Convolutional Coding with QPSK Modulation

A 2-ary ($N = 2$) coded QPSK modulator can be constructed by inverse multiplexing the incoming traffic into 6-parallel streams as shown in Fig. 7.3. These bit streams, now reduced in speed by a factor of 6, are partitioned into two 8 × 3 data blocks. The first 8 × 3 data block maps the 8 × 10 convolutional code block, and the next 8 × 3 data block maps the next 8 × 10 antipodal code block. These code blocks are stored in two 8 × 10 ROMs. The output of each ROM is a unique 10-bit

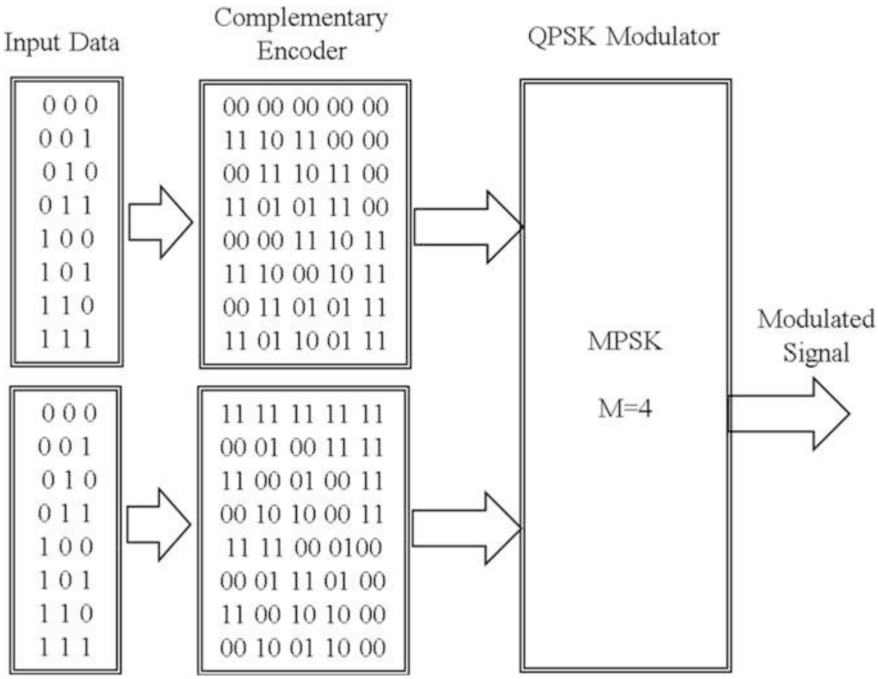
Input Data
Complementary Encoder
QPSK Modulator
0 0 0
0 0 1
0 1 0
0 1 1
1 0 0
1 0 1
1 1 0
1 1 1
00 00 00 00 00
11 10 11 00 00
00 11 10 11 00
11 01 01 11 00
00 00 11 10 11
11 10 00 10 11
00 11 01 01 11
11 01 10 01 11
0 0 0
0 0 1
0 1 0
0 1 1
1 0 0
1 0 1
1 1 0
1 1 1
11 11 11 11 11
00 01 00 11 11
11 00 01 00 11
00 10 10 00 11
11 11 00 0100
00 01 11 01 00
11 00 10 10 00
00 10 01 10 00
MPSK
M=4
Modulated Signal

Fig. 7.3 2-ary convolutional coding with QPSK modulation

convolutional/antipodal code, which is modulated by a QPSK modulator using the same carrier frequency. The code rate is given by:

$$\text{Code Rate: } r = 6/10 = 3/5 \tag{7.8}$$

Since there are two convolutional waveforms (one convolutional and one antipodal), the number of errors that can be corrected is doubled. Moreover, the transmission bandwidth is also reduced, which is given by:

Transmission bandwidth: $BW = 2R_{b2}/\text{bits per symbol Hz}$

$$B = 2R_{b2}/2 = R_{b2} \tag{7.9}$$

where $R_{b2} = R_{b1}/r$. R_{b2} is the coded bit rate, and R_{b1} is the uncoded bit rate.

7.2.4 4-Ary Convolutional Coding with 16PSK Modulation

A 4-ary ($N = 4$) coded 16PSK modulator can be constructed by inverse multiplexing the incoming traffic into 8-parallel streams as shown in Fig. 7.4. These bit streams, now reduced in speed by a factor of 8, are partitioned into four 4 × 2 data

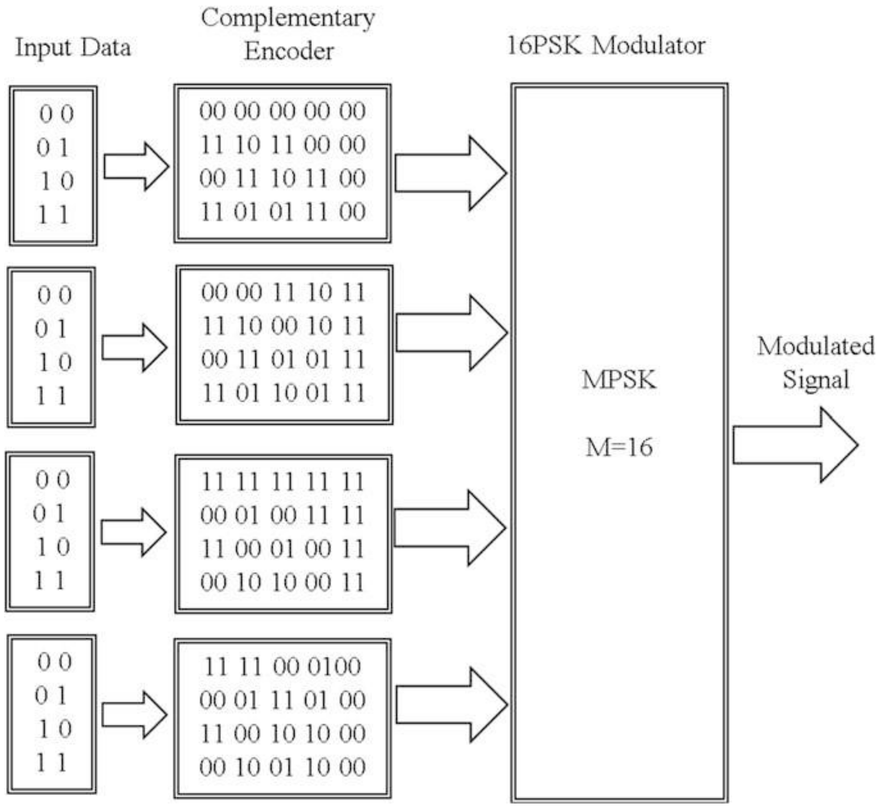
Input Data
Complementary Encoder
16PSK Modulator
0 0
0 1
1 0
1 1
00 00 00 00 00
11 10 11 00 00
00 11 10 11 00
11 01 01 11 00
0 0
0 1
1 0
1 1
00 00 11 10 11
11 10 00 10 11
00 11 01 01 11
11 01 10 01 11
0 0
0 1
1 0
1 1
11 11 11 11 11
00 01 00 11 11
11 00 01 00 11
00 10 10 00 11
0 0
0 1
1 0
1 1
11 11 00 0100
00 01 11 01 00
11 00 10 10 00
00 10 01 10 00
MPSK
M=16
Modulated Signal

Fig. 7.4 4-ary convolutional coding with 16PSK modulation

blocks. Each 4 × 2 data block maps corresponding 4 × 10 convolutional code blocks. These code blocks are stored in four 4 × 10 ROMs. The output of each ROM is a unique 10-bit convolutional/antipodal code, which is modulated by a 16PSK modulator using the same carrier frequency. The code rate is given by:

$$\text{Code Rate:} r = 8/10 = 4/5 \tag{7.10}$$

Since there are four convolutional waveforms (two convolutional and two antipodal), the number of errors that can be corrected is quadrupled. Moreover, the transmission bandwidth is further reduced, which is given by:

Transmission bandwidth: $BW = 2R_{b2}$/bits per symbol Hz

$$= 2R_{b2}/4 = R_{b2}/2 \tag{7.11}$$

where $R_{b2} = R_{b1}/r$. R_{b2} is the coded bit rate, and R_{b1} is the uncoded bit rate.

7.3 N-Ary Convolutional Decoder

7.3.1 Correlation Receiver

Decoding is a process of code correlation. In this process, the receiver compares the received data with the expected data set to recover the actual data. The expected data is stored into a lookup table [see Table 7.1]:

- The lookup table at the receiver contains the input/output bit sequences.
- For $m = 3$, there are 8 possible output combinations of 3-bit data.
- For each combination of 3-bit data, there is a unique encoded 10-bit data (see table).
- The receiver receives one of 8 output sequences.
- Upon receiving an encoded data pattern, the receiver validates the received data pattern by means of code correlation.

Let us examine the correlation process using the following example:

- The input bit pattern $m = 0\ 1\ 1$
- Encoded transmit data: $U = 00\ 11\ 01\ 01\ 11$
- Received data with errors $U^* = 00\ 11\ 01\ 01$ **00**

Notice that the last two bits are in error, identified in bold. Now, let us determine how the receiver recovers the correct data, where the actual input data is $m = 0\ 1\ 1$. This is a correlation process, requiring several tests to validate the actual data. The correlation process is described below:

Test 0:

This test compares the received data with the 1st row of data stored in the lookup table and counts the number of positions it does not match. This is accomplished by MOD2 operation (EXOR operation). The result is presented below:

- Received data: 00 11 01 01 **00**
- 1st row of data in the lookup table: 00 00 00 00 00
- Mod-2 Add: 00 11 01 01 00

Table 7.1 Lookup table a

Input (m)	Output (U)
0. 0 0 0	00 00 00 00 00
1. 0 0 1	00 00 11 10 11
2. 0 1 0	00 11 10 11 00
3. 0 1 1	00 11 01 01 11
4. 1 0 0	11 10 11 00 00
5. 1 0 1	11 10 00 10 11
6. 1 1 0	11 01 01 11 00
7. 1 1 1	11 01 10 01 11

- Correlation value = 4 (count the number of 1’s in MOD2 Add)
- Verdict: No match, continue search.

Test-0: Look Up Table

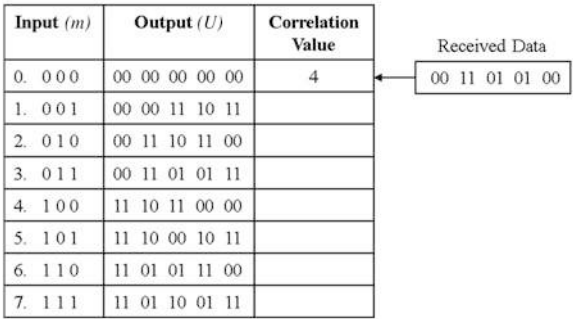

Input (*m*)	Output (*U*)	Correlation Value
0. 0 0 0	00 00 00 00 00	4
1. 0 0 1	00 00 11 10 11	
2. 0 1 0	00 11 10 11 00	
3. 0 1 1	00 11 01 01 11	
4. 1 0 0	11 10 11 00 00	
5. 1 0 1	11 10 00 10 11	
6. 1 1 0	11 01 01 11 00	
7. 1 1 1	11 01 10 01 11	

Received Data

00 11 01 01 00

Test 1:

This test compares the received data with the 2nd row of data stored in the lookup table and counts the number of positions it does not match. This is accomplished by MOD2 operation (EXOR operation). The result is presented below:

- Received data: 00 11 01 01 00
- 2nd row of data in the lookup table: 00 00 11 10 11
- Mod-2 Add: 00 11 10 111 11
- Correlation value = 7
- Verdict: No match, continue search

Test 1: Look Up Table

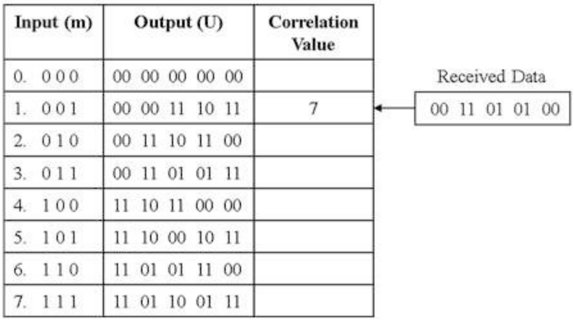

Input (m)	Output (U)	Correlation Value
0. 0 0 0	00 00 00 00 00	
1. 0 0 1	00 00 11 10 11	7
2. 0 1 0	00 11 10 11 00	
3. 0 1 1	00 11 01 01 11	
4. 1 0 0	11 10 11 00 00	
5. 1 0 1	11 10 00 10 11	
6. 1 1 0	11 01 01 11 00	
7. 1 1 1	11 01 10 01 11	

Received Data

00 11 01 01 00

The remaining tests may be conducted by using the same method to reach the final verdict as given below:

Final Verdict:

Collect the correlation values and validate the data that indicates the lowest correlation value. This is presented in the following lookup table.

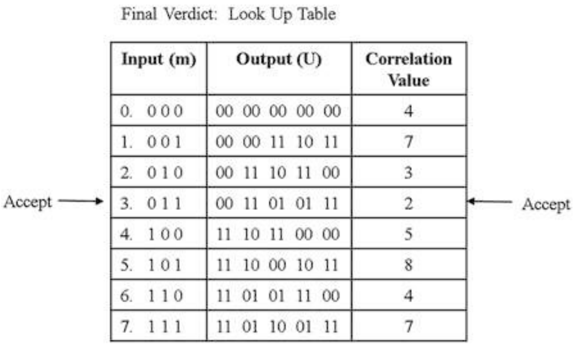

Final Verdict: Look Up Table

Input (m)	Output (U)	Correlation Value
0. 0 0 0	00 00 00 00 00	4
1. 0 0 1	00 00 11 10 11	7
2. 0 1 0	00 11 10 11 00	3
3. 0 1 1	00 11 01 01 11	2
4. 1 0 0	11 10 11 00 00	5
5. 1 0 1	11 10 00 10 11	8
6. 1 1 0	11 01 01 11 00	4
7. 1 1 1	11 01 10 01 11	7

Accept
Accept

In examining the above table, we find that:

- The lowest correlation value is 2.
- The corresponding data is $m = 0\ 1\ 1$.
- This is the data which has been transmitted to the receiver.

7.3.2 *Error Correction Capabilities of N-Ary Convolutional Codes*

An n-bit convolutional code is generated by a rate ½ encoder of constraint length K where

$$n = 4K - 2 \tag{7.12}$$

In the above equation, K is the length of the shift register and the factor −2 is due to the initial content of the shift register, which is 000. For example, a rate ½ convolutional encoder with $K = 3$ generates 00 when the initial content of the shift register is 000. Therefore, the value of n is reduced by 2. Therefore, each 3-bit data sequence corresponds to a unique 10-bit convolutional code. The antipodal code is just the inverse of the convolutional codes. Since a 3-bit sequence has 8 combinations, the above convolutional encoder generates 8 unique convolutional codes and 8 unique antipodal codes, for a total of 16 complementary convolutional codes.

These code blocks are displayed in Fig. 7.2. As can be seen, each code is unique and there is a minimum distance $d_{\min}$ between any two convolutional codes where

$$d_{\min} \leq n/2 \tag{7.13}$$

This distance property can be used to detect an impaired received code by setting a threshold midway between two convolutional codes. It can be shown that an n-bit convolutional code can correct the errors, where

Table 7.2 Error correction capabilities of N-ary convolutional codes

Constraint length (K)	Number of errors corrected if the modulation scheme is (BPSK)	Number of errors corrected if the modulation scheme is (QPSK)	Number of errors corrected if the modulation scheme is (16PSK)
3	2	4	8
5	4	8	16
7	6	12	24
9	8	16	32

$$t \leq d_{\min}/2 \leq n/4 \leq (4K - 2)/4 \tag{7.14}$$

where

- t = Number of errors that can be corrected by means of a single convolutional code
- n = Code length
- K = Constraint length

The proposed N-ary coded modulation is a multi-level channel coding technique, where instead of coding one bit at a time two or more bits are encoded simultaneously. Therefore, we modify the above equation to obtain

$$N\,t \leq N\,d_{\min}/2 \leq Nn/4 \leq N(4K - 2)/4 \tag{7.15}$$

Table 7.2 displays error correction capabilities of N-ary convolutional codes.

7.4 N-Ary Orthogonal Coding and M-Ary Modulation

7.4.1 Background

Orthogonal codes are used in CDMA cellular communications for spectrum spreading and user ID [10–13]. The use of orthogonal codes for forward error control coding has also been investigated by a limited number of authors [14–16]. This chapter presents a method of channel coding based on N-ary orthogonal codes [17–20]. N-ary orthogonal coding is a multi-level channel coding technique, where instead of coding one bit at a time two or more bits are encoded simultaneously. This type of channel coding is bandwidth efficient and further enhances the coding gain with bandwidth efficiency. In the proposed scheme, the input serial data is converted into several parallel streams. These parallel bit streams, now reduced in speed, are mapped into N-ary orthogonal codes. The coded information bits are then modulated by means of MPSK modulator and transmitted through a channel. At the receiver, the decoder recovers the data by means of code correlation. The proposed coded modulation scheme is bandwidth efficient and offers a coding gain near the Shannon’s limit.

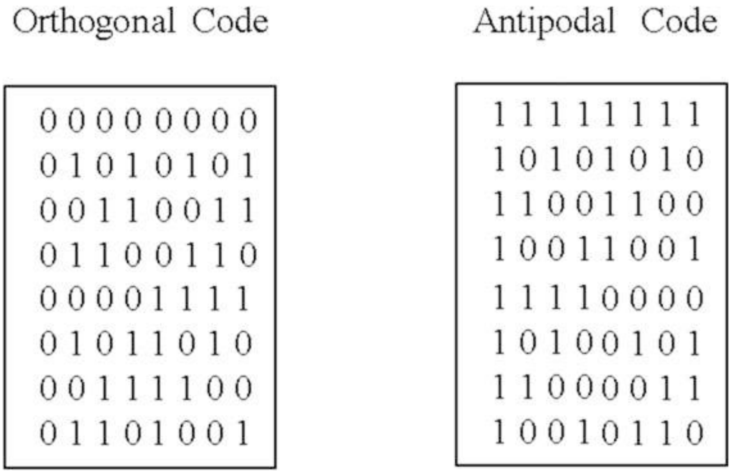

Orthogonal Code	Antipodal Code
0 0 0 0 0 0 0 0	1 1 1 1 1 1 1 1
0 1 0 1 0 1 0 1	1 0 1 0 1 0 1 0
0 0 1 1 0 0 1 1	1 1 0 0 1 1 0 0
0 1 1 0 0 1 1 0	1 0 0 1 1 0 0 1
0 0 0 0 1 1 1 1	1 1 1 1 0 0 0 0
0 1 0 1 1 0 1 0	1 0 1 0 0 1 0 1
0 0 1 1 1 1 0 0	1 1 0 0 0 0 1 1
0 1 1 0 1 0 0 1	1 0 0 1 0 1 1 0

Fig. 7.5 Bi-orthogonal code set for $n = 8$. An 8-bit orthogonal code has 8 orthogonal codes and 8 antipodal codes for a total of 16 bi-orthogonal codes

7.4.2 *Orthogonal Codes*

Orthogonal codes are binary-valued and have equal numbers of 1's and 0's. Antipodal codes, on the other hand, are just the inverse of orthogonal codes. Antipodal codes are also orthogonal among them. Therefore, an n-bit orthogonal code has n-orthogonal codes and n-antipodal codes, for a total of 2n bi-orthogonal codes. For example, an 8-bit orthogonal code has 8 orthogonal codes and 8 antipodal codes, for a total of 16 bi-orthogonal codes, as shown in Fig. 7.5 [12]. Similarly, a 16-bit orthogonal code has 16 orthogonal codes and 16 antipodal codes for a total of 32 bi-orthogonal codes, as shown in Fig. 7.6.

We note that orthogonal codes are essentially (n, k) block codes, where a k-bit data set is represented by a unique n-bit orthogonal code ($k < n$). We now show that code rates such as rate ½, rate ¾, and rate 1 are indeed available out of orthogonal codes. The principle is presented below.

7.4.3 *2-Ary Orthogonal Coding with QPSK Modulation*

A 2-ary orthogonal coded QPSK modulator with $n = 8$, having 16 complementary orthogonal codes, can be constructed by inverse multiplexing the incoming traffic into 6-parallel streams, as shown in Fig. 7.7. These bit streams, now reduced in speed by a factor of 6, are partitioned into two data blocks. Each data block is mapped into an 8 × 8 code block, as depicted in the figure.

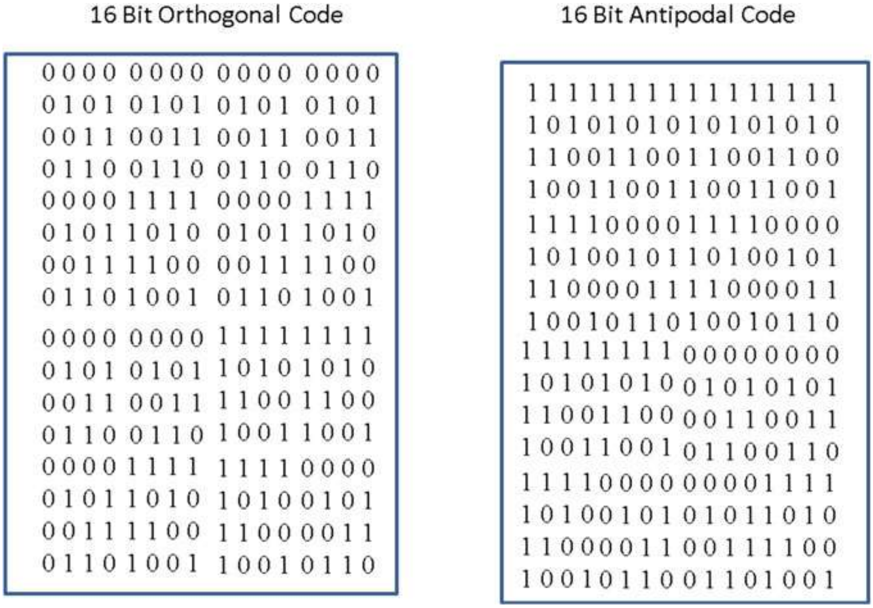

16 Bit Orthogonal Code

0000 0000 0000 0000
0101 0101 0101 0101
0011 0011 0011 0011
0110 0110 0110 0110
00001111 00001111
0101 1010 0101 1010
0011 1100 0011 1100
0110 1001 0110 1001
0000 0000 11111111
0101 0101 10101010
0011 0011 11001100
0110 0110 10011001
00001111 11110000
0101 1010 10100101
0011 1100 11000011
0110 1001 10010110

16 Bit Antipodal Code

1111111111111111
1010101010101010
1100110011001100
1001100110011001
1111000011110000
1010010110100101
1100001111000011
1001011010010110
11111111 00000000
10101010 01010101
11001100 00110011
10011001 01100110
11110000 00001111
10100101 01011010
11000011 00111100
10010110 01101001

Fig. 7.6 Bi-orthogonal code set for $n = 16$. A 16-bit orthogonal code has 16 orthogonal codes and 16 antipodal codes for a total of 32 bi-orthogonal codes

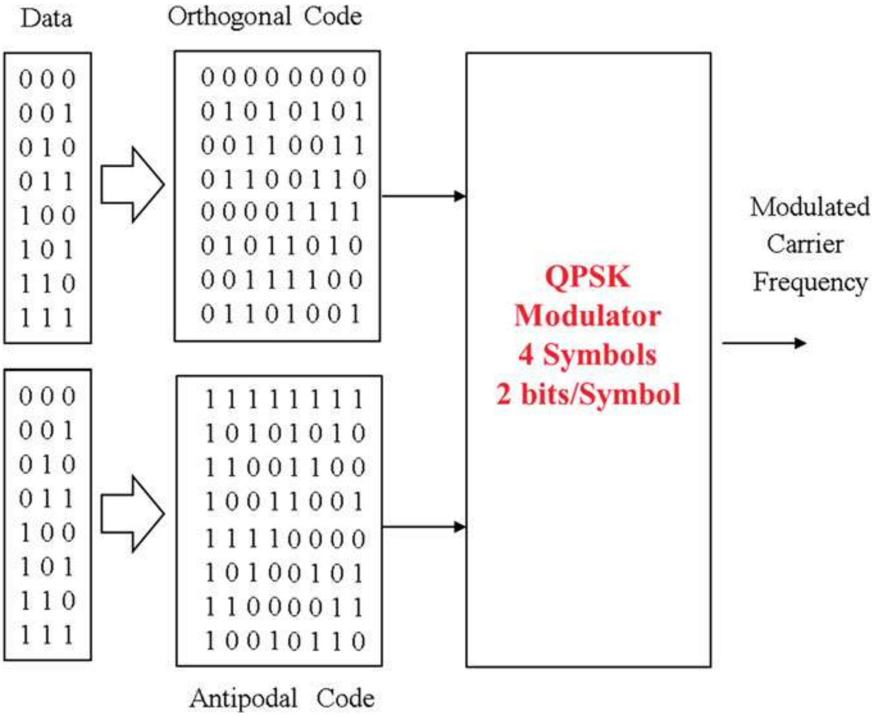
Data
Orthogonal Code
000
001
010
011
100
101
110
111
00000000
01010101
00110011
01100110
00001111
01011010
00111100
01101001
QPSK
Modulator
4 Symbols
2 bits/Symbol
Modulated
Carrier
Frequency
000
001
010
011
100
101
110
111
11111111
10101010
11001100
10011001
11110000
10100101
11000011
10010110
Antipodal Code

Fig. 7.7 2-ary orthogonal coding with $N = 2$

According to the communication scheme, when a 3-bit data pattern needs to be transmitted, the corresponding orthogonal/antipodal code is transmitted instead, requiring a QPSK modulator. The modulated waveforms are in orthogonal space and have fewer errors. The code rate is given by:

$$\text{Code Rate: } r = 6/8 = 3/4$$

Since there are two orthogonal waveforms (one orthogonal and one antipodal), the number of errors that can be corrected is doubled. Moreover, the transmission bandwidth is also reduced, which is given by:

Transmission Bandwidth: $BW = 2R_{b2}$/bits per symbol Hz

$$= 2R_{b2}/2 = R_{b2}\,\text{Hz} \tag{7.16}$$

where $R_{b2} = R_{b1}/r$. R_{b2} is the coded bit rate and R_{b1} is the uncoded bit rate.

7.4.4 4-Ary Orthogonal Coding with 16PSK Modulation

A 4-ary orthogonal coded 16PSK modulator, with $n = 8$, having 16 complementary orthogonal codes, can be constructed by inverse multiplexing the incoming traffic into 8-parallel streams, as shown in Fig. 7.8. The bit streams, now reduced in speed by a factor of 8, are partitioned into four data blocks. Each data block is mapped into a 4×8 code block as depicted in the figure. According to the communication scheme, when a 2-bit data pattern needs to be transmitted, the corresponding orthogonal/antipodal code is transmitted instead, requiring a 16PSK modulator. The modulated waveforms are in orthogonal space and have fewer errors. The code rate is given by:

$$\text{Code Rate: } r = 8/8 = 1 \tag{7.17}$$

This is achieved without bandwidth expansion.

Since there are four orthogonal waveforms (two orthogonal and two antipodal), the number of errors that can be corrected is quadrupled. Moreover, the transmission bandwidth is further reduced, which is given by:

Transmission bandwidth: $BW = 2R_{b2}$/bits per symbol Hz

$$= 2R_{b2}/4 = R_{b2}/2\ \text{Hz} \tag{7.18}$$

where $R_{b2} = R_{b1}/r$. R_{b2} is the coded bit rate, and R_{b1} is the uncoded bit rate.

7.4.5 2-Ary Orthogonal Decoding

Decoding is a correlation process. Notice that the entire biorthogonal code block is partitioned into two code blocks. Each code block represents a data block as shown in Fig. 7.9. Upon receiving an impaired code, the receiver compares it with each

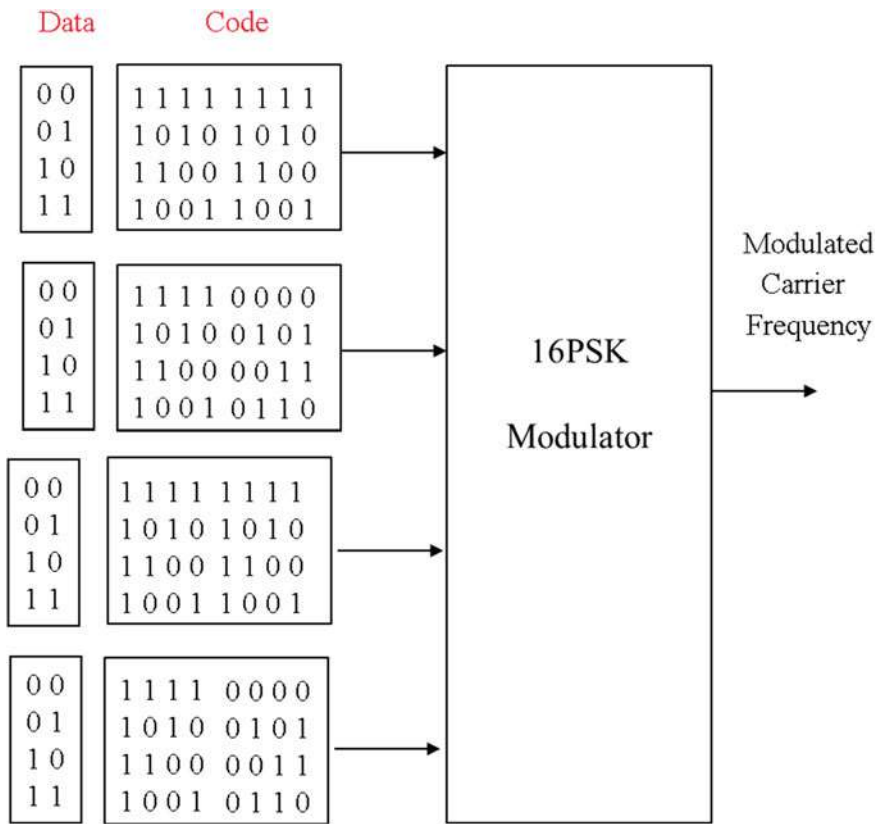
Data
Code
00
01
10
11
1111 1111
1010 1010
1100 1100
1001 1001
00
01
10
11
1111 0000
1010 0101
1100 0011
1001 0110
00
01
10
11
1111 1111
1010 1010
1100 1100
1001 1001
00
01
10
11
1111 0000
1010 0101
1100 0011
1001 0110
16PSK
Modulator
Modulated
Carrier
Frequency

Fig. 7.8 N-ary orthogonal coding with $N = 4$

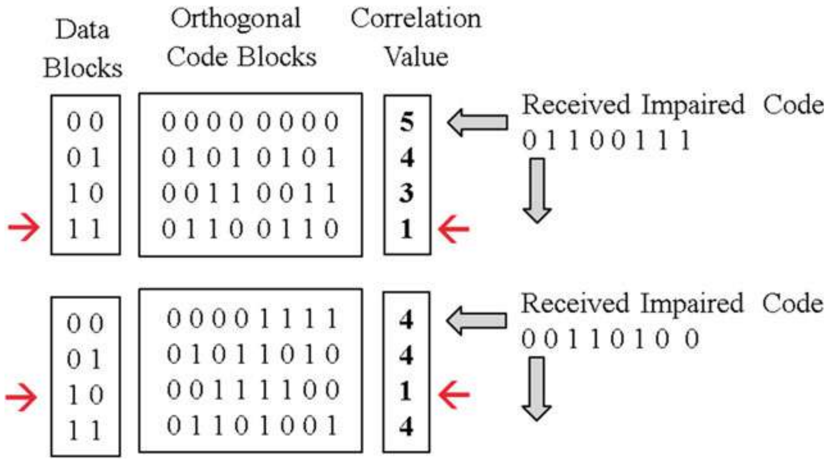
Data
Blocks
Orthogonal
Code Blocks
Correlation
Value
00
01
10
11
0000 0000
0101 0101
0011 0011
0110 0110
5
4
3
1
Received Impaired Code
01100111
00
01
10
11
00001111
01011010
00111100
01101001
4
4
1
4
Received Impaired Code
00110100

Fig. 7.9 2-ary orthogonal decoding

entry in the code block and appends a correlation value for each comparison. A valid code is declared when the closest approximation is achieved. As can be seen, the minimum correlation value in each block is 1 as depicted in the figure. The corresponding data is 1 1, 1 0, 1 1, 1 0, respectively.

Since there are two orthogonal waveforms (one orthogonal and one antipodal), the number of errors that can be corrected is given by 2 × 1 = 2. Moreover, the bandwidth is also reduced, as presented below:

$$BW = 2Rb2/\text{bits per symbol} = 2_{Rb2}/2 = \mathrm{R}_{b2} \tag{7.19}$$

7.4.6 4-Ary Orthogonal Decoding

Decoding is a correlation process. Notice that the entire biorthogonal code block is partitioned into four code blocks. Each code block represents a data block, as shown in Fig. 7.10. Upon receiving an impaired code, the receiver compares it with each entry in the code block and appends a correlation value for each comparison. A valid code is declared when the closest approximation is achieved. As can be seen, the minimum correlation value in each block is 1 as depicted in the figure. The corresponding data is 1 1, 1 0, 1 1, 1 0, respectively.

Since there are four orthogonal waveforms (two orthogonal and two antipodal), the number of errors that can be corrected is given by 4 × 1 = 4. Moreover, the bandwidth is further reduced as presented below:

$$BW = 2Rb2/\text{bits per symbol} = 2_{Rb2}/4 = R_{b2}/2 \tag{7.20}$$

7.4.7 Error Correction Capabilities of N-Ary Orthogonal Codes

Error correction capabilities of orthogonal codes have been discussed earlier [16]. We present it again for convenience. An n-bit orthogonal code has $n/2$ 1's and $n/2$ 0's; i.e., there are $n/2$ positions where 1's and 0's differ. Therefore, the distance between two orthogonal codes is $d = n/2$. This distance property can be used to detect an impaired received code by setting a threshold midway between two orthogonal codes as shown in Fig. 7.11, where the received code is shown as a dotted line. This is given by:

$$d_{th} = \frac{n}{4} \tag{7.21}$$

where n is the code length and d_{th} is the threshold, which is midway between two valid orthogonal codes. Therefore, for the given 8-bit orthogonal code, we have

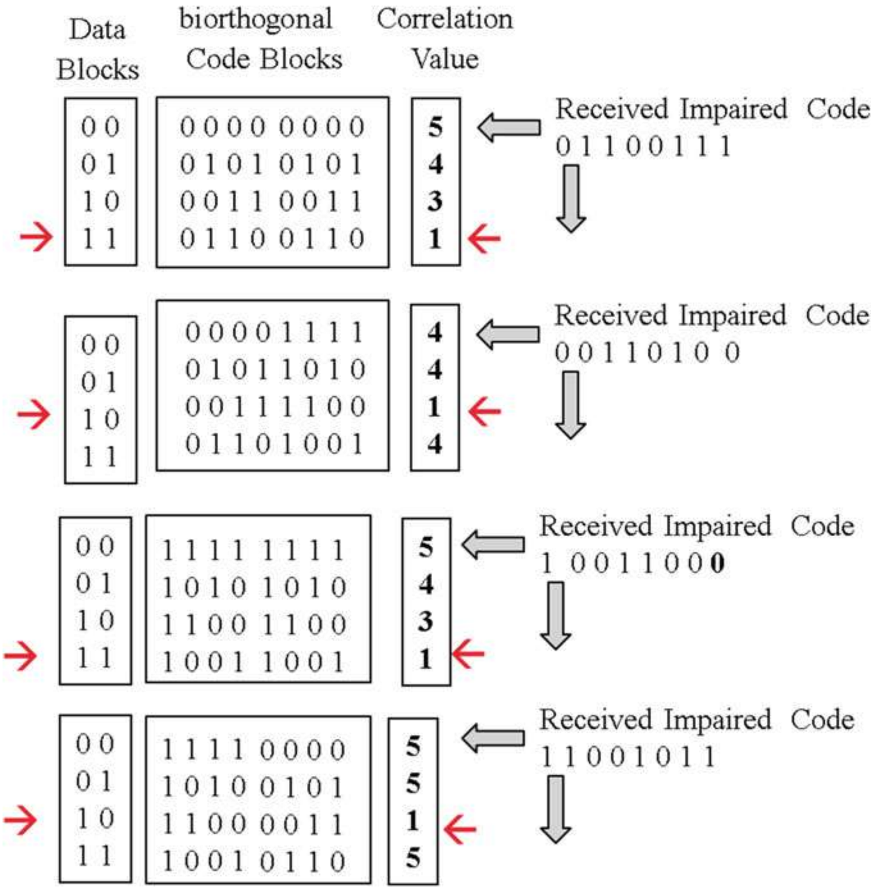
Data Blocks
biorthogonal Code Blocks
Correlation Value
00
01
10
11
0000 0000
0101 0101
0011 0011
0110 0110
5
4
3
1
Received Impaired Code
01100111
00
01
10
11
00001111
01011010
00111100
01101001
4
4
1
4
Received Impaired Code
0011010 0
00
01
10
11
1111 1111
1010 1010
1100 1100
1001 1001
5
4
3
1
Received Impaired Code
1 0011000
00
01
10
11
1111 0000
1010 0101
1100 0011
1001 0110
5
5
1
5
Received Impaired Code
11001011

Fig. 7.10 4-ary orthogonal decoding

Fig. 7.11 Distance properties of orthogonal codes

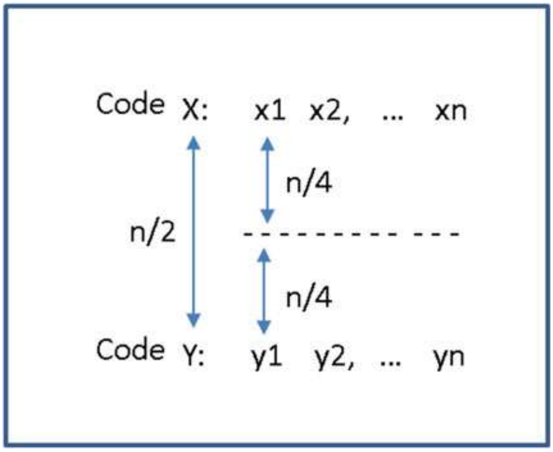
Code X: x1 x2, ... xn
n/4
n/2
n/4
Code Y: y1 y2, ... yn

$d_{th} = 8/4 = 2$. This mechanism offers a decision process, where the incoming impaired orthogonal code is examined for correlation with the neighboring codes for a possible match.

The received code is examined for correlation with the neighboring codes for a possible match. The acceptance criterion for a valid code is that an n-bit comparison must yield a good autocorrelation value; otherwise, a false detection will occur. The following correlation process governs this, where an impaired orthogonal code *i*s compared with a pair of n-bit orthogonal codes to yield:

$$R(x, y) = \sum_{i=1}^{n} x_i y_i \geq (n - d_{th}) + 1 \tag{7.22}$$

where *x* and *y* are two *n*-bit orthogonal codes, *R*(*x*, *y*) is the autocorrelation function, *n* is the code length, and d_{th} is the threshold as defined earlier. Since the threshold is in the midway between two valid codes, an additional 1-bit offset is added to Eq. 7.21 for reliable detection. The average number of errors that can be corrected by means of this process can be estimated by combining Eq. 7.21 and Eq. 7.22, yielding:

$$t = n - R(x, y) = \frac{n}{4} - 1 \tag{7.23}$$

In the above equation, *t* is the number of errors that can be corrected by means of an *n*-bit orthogonal code. For example, a single-error-correcting orthogonal code can be constructed by means of an 8-bit orthogonal code ($n = 8$). Similarly, a three-error-correcting orthogonal code can be constructed by means of a 16-bit orthogonal code ($n = 16$) and so on. Table 7.3 shows a few orthogonal codes and the corresponding error-correcting capabilities.

Problem 7.1 This problem relates to N-ary convolutional codes and M-ary PSK modulation.

Given:

- Input bit rate before coding: $R_{b1} = 10$ kb/s
- $K = 3$, Rate ½ complementary convolutional codes
- 4-ary convolutional coding ($N = 4$)

Table 7.3 Error correction capabilities of N-ary orthogonal codes

Code length *n*	Number of errors corrected per code *t*	Number of errors corrected by N-ary code *Nt* $N = 1, 2, 4, 8, \ldots$
8	1	N
16	3	3 N
32	7	7 N
64	15	15 N

Find:

(a) Modulation level M
(b) Code rate r
(c) Bit rate after coding R_{b2}
(d) Transmission bandwidth BW
(e) Number of errors corrected

Solution:

(a) A 4-ary ($N = 4$) coded modulator can be constructed by inverse multiplexing the incoming traffic into 8-parallel stream. These bit streams, now reduced in speed by a factor of 8, are partitioned into four 4 × 2 data blocks. Each 4 × 2 data block maps a corresponding 4 × 10 convolutional code blocks. These code blocks are stored in four 4 × 10 ROMs. The output of each ROM is a unique 10-bit convolutional/antipodal code, which is modulated by a $2^4 = 16$ PSK modulator using the same carrier frequency. Therefore, $M = 16$.

(b) The code rate is given by:

$$\text{Code Rate: } r = 8/10 = 4/5$$

(c) Bit rate after coding $R_{b2} = R_{b1}/r = (5/4)\ 10\ \text{kb/s} = 12.5\ \text{kb/s}$

(d) Transmission bandwidth: $BW = 2R_{b2}/\text{bits per symbol Hz}$

$$= 2R_{b2}/4 = R_{b2}/2 = (12.50\,\text{kb/s})/2 = 6.25\,\text{kHz}$$

(e) Since there are four convolutional waveforms (two convolutional and two antipodal), the number of errors that can be corrected is quadruple, $4t = 4 \times 2 = 8$ [Each code corrects 2 errors.]

Problem 7.2 This example relates to N-ary orthogonal codes and M-ary PSK modulation.

Given:

- Input bit rate before coding: $R_{b1} = 10$ kb/s
- $n = 8$ orthogonal codes having 16 complementary orthogonal codes
- 4-ary orthogonal Coding

Find:

(a) Modulation level M
(b) Code rate r
(c) Bit rate after coding R_{b2}
(d) Transmission bandwidth BW
(e) Number of errors corrected

Solution:

(a) A 4-ary orthogonal coded modulator, with $n = 8$, having 16 complementary orthogonal codes, can be constructed by inverse multiplexing the incoming traffic into 8-parallel streams. The bit streams, now reduced in speed by a factor of 8, are partitioned into four data blocks. Each data block is mapped into a 4×8 code block. According to the communication scheme, when a 2-bit data pattern needs to be transmitted, the corresponding orthogonal/antipodal code is transmitted instead, requiring a $2^4 = 16$ PSK modulator. Therefore, $M = 416$.

(b) The code rate is given by:

$$\text{Code Rate:}\, r = 8/8 = 1$$

(c) The bit rate after coding is given by:

$$R_{b2} = R_{b1} r = 10\,\text{kb/s}\ (r = 1)$$

(d) The transmission bandwidth is given by:

$$\begin{aligned} BW &= 2Rb2/\text{bits per symbol Hz} \\ &= 2R_{b2}/4 = 2 \times 10\,\text{kb/s}/4 = 5\ \text{k Hz} \end{aligned}$$

(e) Since there are four orthogonal waveforms (two orthogonal and two antipodal), the number of errors that can be corrected is quadrupled, i.e.,

$$4t = 4 \times 1 = 4$$

7.5 Conclusions

A method of coded modulation, based on N-ary complementary codes and M-ary PSK (MPSK) modulation, is presented. N-ary coded modulation is a multi-level channel coding and multi-level modulation technique, where instead of coding one bit at a time two or more bits are encoded simultaneously and then modulated by means of M-ary modulation. At the receiver, the decoder recovers the data by means of code correlation. A lookup table at the receiver contains the input/output bit sequences. Upon receiving an encoded data pattern, the receiver validates the received data pattern by means of code correlation. Construction of N-ary coded modulation schemes based on convolutional and orthogonal codes is presented to illustrate the concept. The proposed coded modulation schemes are bandwidth efficient and offer a coding gain near the Shannon's limit.

References

1. Ebert PM, Tong SY (1968) Convolutional Reed-Soloman codes, Bell Syst Tech, pp 729–742
2. Gallager RG (1968) Information theory and reliable communications. Wiley, New York
3. Kohlenbero A,.Forney GD Jr (1968) Convolutional coding for channels with memory, IEEE Trans Inf Theory IT-14:618–626
4. Reddy SM (1968) Further results on convolutional codes derived from block codes. Inf Control 13:357–362
5. Reddy SM (1968) A class of linear convolutional codes for compound channels. Technical Report, Bell Telephone Laboratories, Holmdel, New Jersey
6. Reddy SM, Robinson JP (1968) A construction for convolutional codes using block codes. Inf Control 12:55–70
7. Berrou C, Glavieux A, Thitimajshima P (2010) Near Shannon limit error—correcting (PDF). Retrieved 11 February 2010
8. Berrou C (2010) The ten-year-old turbo codes are entering into service, Bretagne, France. Retrieved11 February 2010
9. McEliece RJ, MacKay DJC, Cheng J-F (1998) Turbo decoding as an instance of Pearl's "belief propagation" algorithm. IEEE J Sel Areas Commun 16(2):140–152, doi:10.1109/49.661103, ISSN 0733-8716
10. IS-95 Mobile station—base station compatibility standard for dual mode wide band spread
11. Spectrum Cellular Systems, TR 55, PN-3115, March 15 (1993)
12. Samuel CY (1998) CDMA RF system engineering. Artech House Inc
13. Vijay KG (1999) IS-95 CDMA and cdma 2000. Prentice Hall
14. Bernard S (1988) Digital communications fundamentals and applications. Prentice Hall
15. Faruque S, Faruque S, Semke W (2013) Orthogonal on-off keying for free-space laser communications with ambient light cancellation. SPIE Optical Eng J 52(9), September 26
16. Faruque S (2016) Radio frequency channel coding made easy, Springer, ISBN: 978-3-319-21169-5
17. Faruque S (1999) Battlefield wideband transceivers based on combined N-ary orthogonal signaling and M-ary PSK modulation, SPIE proceedings, vol. 3709 digitization of the battle space 1 V, pp 123–128
18. Faruque S et al (2001) Broadband wireless access based on code division parallel access, US Patent No. 6208615, March 27
19. Faruque S et al (2001) Bi-orthogonal code division multiple access system, US Patent No. 6198719, March 6
20. Faruque S (2003) Code division multiple access cable modem, US Patent No. 6647059, November

S. Gernsback's
Radio Encyclopedia

S. GERNSBACK'S

RADIO ENCYCLOPEDIA

SIDNEY GERNSBACK
PUBLISHER
NEW YORK

All rights reserved including that of translating
into foreign languages including the Scandinavian

Copyright 1927
by
SIDNEY GERNSBACK

Printed in the United States of America

A

“A” BATTERY—An appliance for furnishing electrical current to light the *filaments* of a *vacuum tube*, and cause emission of *electrons* (see *Theory of Vacuum Tube Operation*). This battery may be in any of several forms, the two in popular use being the storage type (see *Storage Battery*), and the dry cell (see *Dry Cell*). Fig. 1 shows method of connecting the “A” Battery in a vacuum tube circuit.

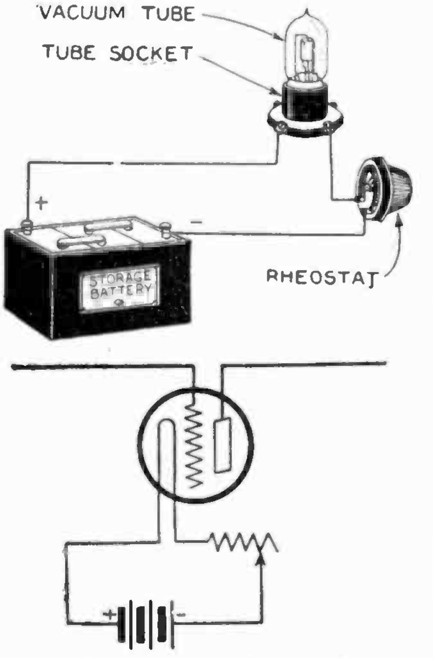
VACUUM TUBE
TUBE SOCKET
+
−
STORAGE BATTERY
RHEOSTAT

Fig. 1. Method of connecting the “A” battery to a vacuum tube in radio receiving circuits. The lower diagram shows the “A” battery as represented in schematic drawings

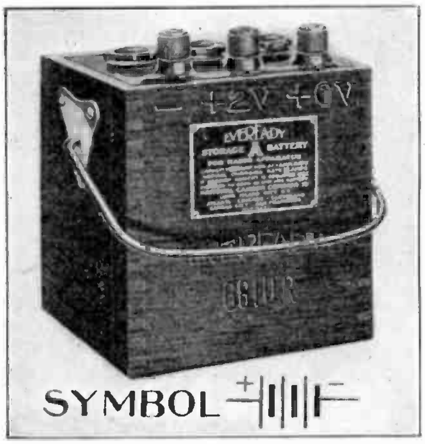
EVEREADY
STORAGE A BATTERY
SYMBOL
+
−

Photos by courtesy of National Carbon Co., Inc.

Fig. 2. A storage battery such as used for “A” battery current supply

In Fig. 2 a popular type of storage battery is shown. Batteries of this type come in many individual forms, the usual voltage being standardized at 6 volts. However, for use with certain tubes, 2 and 4 volt storage batteries are available. For general radio work, using the low-consumption tubes, the battery, as shown in Fig. 2, can hold a charge of from 60-80 ampere hours. (*See Storage Battery.*)

The other common form of “A” Battery is the dry cell, usually having a difference of potential (voltage) of 1.5 volts. Fig. 3 illustrates a popular type of dry cell. In connecting a number of cells for use with a vacuum tube, the method of connecting will depend on the voltage rating of the particular tube. If the tube is rated at 1.5 volts, a single dry cell may be used as in Fig. 4, or several may be connected in *parallel* (*q.v.*) shown in Fig. 5, the total voltage however, being only the voltage of one cell. Three dry cells connected in *series* (*q.v.*) will give 4½ volts, Fig. 6. A connection for dry cells, known as *series-parallel* or *multiple* (*q.v.*), is shown in Fig. 7. This connection will also produce 4½ volts and at the same time increase the life of the cells. Four dry cells connected in series will give six volts for use with a standard six-volt tube, Fig. 8. Dry cells are only used with six-volt tubes where the tube has a current rating of ¼ ampere, and seldom more than one tube is used in this manner due to the relatively high discharge and consequent short life of the battery.

A storage battery can be used for the “A” Battery current with a 4-volt tube if the connection is made as in Fig. 9. Here a standard 6-volt storage battery is tapped at the four volt terminal.

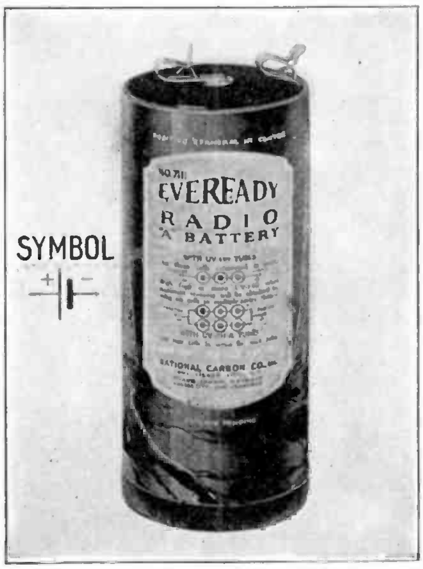
SYMBOL
+
−
No. 711
EVEREADY
RADIO
"A" BATTERY
NATIONAL CARBON CO.

Fig. 3. A dry cell “A” battery

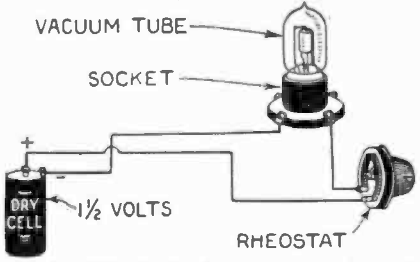
VACUUM TUBE
SOCKET
+
−
DRY CELL
1½ VOLTS
RHEOSTAT

Fig. 4. This circuit shows how one dry cell is used as the “A” battery for types of tubes requiring 1½ volts filament current

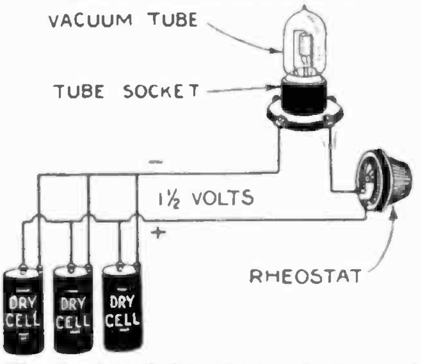
VACUUM TUBE
TUBE SOCKET
−
1½ VOLTS
+
DRY CELL
DRY CELL
DRY CELL
RHEOSTAT

Fig. 5. Several dry cells may be connected in parallel for increased amperage or longer life of the “A” battery

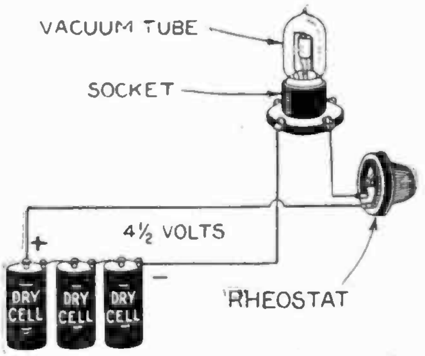
VACUUM TUBE
SOCKET
4½ VOLTS
+
−
DRY CELL
DRY CELL
DRY CELL
RHEOSTAT

Fig. 6. Three dry cells connected in series give 4½ volts, serving as the “A” battery for the U.V.199 or C299 type of tube

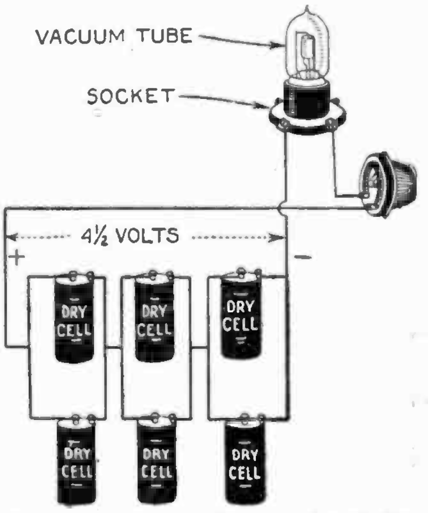
VACUUM TUBE
SOCKET
4½ VOLTS
+
−
DRY CELL
DRY CELL
DRY CELL
DRY CELL
DRY CELL
DRY CELL

Fig. 7. A series—parallel connection of three dry cells for longer life of the “A” battery with 4½ volt tubes

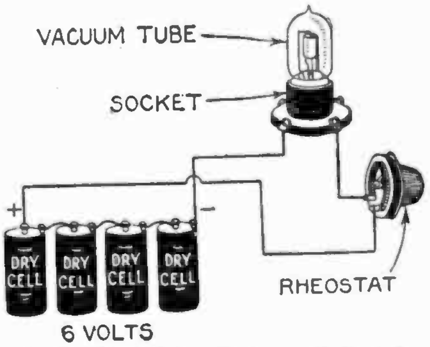
VACUUM TUBE
SOCKET
+
−
DRY CELL
DRY CELL
DRY CELL
DRY CELL
RHEOSTAT
6 VOLTS

Fig. 8. Four dry cells connected in series give 5 volts for low consumption tubes

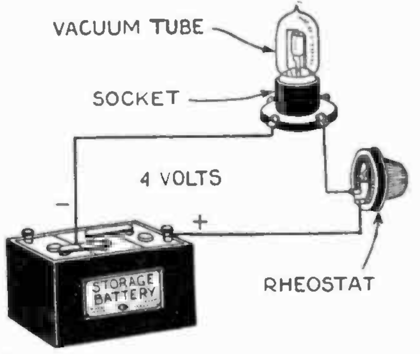
VACUUM TUBE
SOCKET
4 VOLTS
−
+
STORAGE BATTERY
RHEOSTAT

Fig. 9. Showing method of tapping at 4-volt terminal of a 6-volt storage battery for type U.V.199 or C299 tubes or other 4-volt tubes

Another connection for low voltage tubes can be made as illustrated in Fig. 10. Here only one cell of the storage battery is used. This cell can be used until exhausted when the other cells can be used in the same manner. When the three cells are exhausted, the storage battery may be charged by means of a *charger* (*q.v.*).

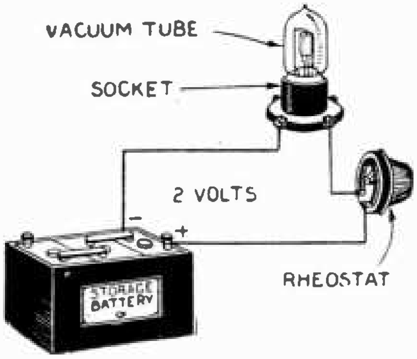

Fig. 10. How a single cell of a 6-volt storage battery may be employed to furnish "A" battery current for 1½ volt tubes

ABAMPERE—A unit of current used in special theoretical work. It has a value of 10 *amperes* (*q.v.*) and is not generally used in electrical work.

ABBREVIATIONS, AMATEUR—The transmitting amateurs of the United States use their own peculiar abbreviations in their inter-communication work. These are used to lessen the interference between transmitting stations as it permits conversation in a much quicker manner. The abbreviations generally used are listed below. Simplified spelling is always resorted to, words being spelled according to their sounds. Some such words are given, but others are so obvious that they are not listed. There are also the standard "Q" signals adopted by the International Radiotelegraphic Convention, to be found in official publications and given under the heading "Abbreviations, International Radiotelegraphic Convention."

ACCW—A transmitter using rectified alternating current as the plate supply.

AFTRN or P.M.—Afternoon.

AMMTR—Ammeter.

ARRL—American Radio Relay League.

BCL—Broadcast listener, one who does not have a transmitting station.

BC STATION—Broadcast Station.

BFRE—or B4—Before.

BLO—Blow.

BOTTLE or V.T.—Vacuum Tube.

C—See or Call.

CGE—Cage.

CM—City Manager of A. R. R. L.

CTPSE—Counterpoise.

CUL—See you later or call you later.

CW—Continuous wave transmitter.

DCCW—A transmitter using direct current as the supply for the plate.

DM—District Manager of A. R. R. L.

DOPE—Information, or in another sense, a varnish for coating inductance coils.

DS—District Superintendent of the A. R. R. L.

DX—Distance, receiving or transmitting.

FB—Fine Business.

FONE—Phone.

GA—Good afternoon.

GLD—Glad.

GM—Good morning.

GN—Good night.

FRQ—Frequency.

HAM—Being a term for the transmitting amateur.

HI—High or I am laughing—it's funny, according to use in sentence.

HI LOS—High loss.

HR—Here.

HV—Have.

KCK BCK—Kick back.

KONGRATS—Congratulations.

KY—Key.

LW LOS—Low loss.

LW WVE—Low waves.

MGHT—Might.

MIKE—Microphone.

MSG—Message.

NITE—Night.

OM—Old man.

OP—Operator.

OPS—Operators.

ORS—Official Radio Station.

OT—Oscillation transformer.

OW—Female operator.

PEEP-PEEP—Continuous wave transmitter.

QCW—Listen for my CW transmitter.

QCW?—Shall I listen to your CW transmitter?

QRMER—An operator who persists in interfering with other stations after being warned to stop.

QSR—I will relay the message.

QSR?—Will you relay the message?

QSS—You are fading.

QSS?—Is my signal fading?

RDO—Radio.

RECVR—Receiver.

RELY—Relay.

ROCKCRUSHER—Spark transmitter.

SA—Say.

SIGS—Signals.

SPIKE—Message file.

SPK—Speak.

SUREFIRE—Something positive in action.

THG—Thing.

THT—That.

TLL—Till.

TM—Traffic manager of A. R. R. L.

TMR—Tomorrow.

TRAFFIC—The handling of messages among the stations.

U—You.

UNLIS—Unlicensed.

UR—Your.

VLTMTR—Voltmeter.

WL—Will.

WRLS—Wireless.

WVL—Wavelength.

"X" LICENSE—Experimental License.

YL—Young lady.

"Z"—Special license.

73—Best regards.

ABBREVIATIONS — RADIO — ELECTRICAL—A letter or group of letters of the alphabet arranged so that there is a definite electrical meaning when used in radio telegraphy or telephony. These abbreviations are the means by which technical writings are simplified, *i.e.*, Alternating Current is simply A.C. when used in general telegraphy. As most of the latest radio text books use these abbreviations, the following are given to aid the reader in understanding the various abbreviations:

A.C.—Alternating Current.

A.F.—Audio Frequency.

B.E.M.F.—Back Electromotive Force.

B. & S.—Brown & Sharpe Wire Gauge.

B.S.W.G.—Birmingham Standard Wire Gauge.

C.—Capacity.

C.G.S. — Centimetre-gramme-second (*q.v.*).

C.W.—Continuous Waves.

D.C.—Direct Current.

D.C.C. — Double Cotton Covered (Wire).

D.S.C.—Double Silk Covered.

E.—Induced Pressure (See Abvolt).

E.M.F.—Electromotive Force.

H.F.—High Frequency.

H.F.C.—High Frequency Current.

I.R.E.—Institute of Radio Engineers.

L.—Inductance (in Formulae).

L.F.—Low Frequency.

L.F.C.—Low Frequency Current.

M.A.—Milliampere.

Mfd.—Microfarad.

P.D.—Potential Difference.

R.—Electrical Resistance.

R.F.—Radio Frequency.

R.P.M.—Revolutions per Minute.

R.P.S.—Revolutions per Second.

S.G.—Specific Gravity.

S.H.M.—Simple Harmonic Motion.

S.S.C.—Single Silk Covered.

S.W.G.—Standard Wire Gauge.

W/L—Wavelength.

X's—Atmospherics.

ABBREVIATIONS, INTERNATIONAL RADIO TELEGRAPHIC CONVENTION—All messages and conversations are not transmitted in regular form, various forms of abbreviations being widely used to increase the speed of transmission and to reduce cost. It also has the tendency to decrease the interference as messages are sent swiftly in half the time. In wireless telegraphy, abbreviations have been adopted as international, and today there are very few ships or shore stations that do not use at least part of them daily. Below we give the International Radiotelegraphic Abbreviations as adopted at The International Radiotelegraphic Convention. An example of the use of the abbreviations is as follows:

A station wishing to learn the name of a ship which is seen off the coast, sends in code the letters QRA?, which means, what is the name of your ship? The station or ship answers with the letters, QRA S.S. Lapland, thus in a few seconds the question had been answered.

Abbreviation	*Question*	*Answer or Notice*
CQ	—.—.—.—.—	Signal of inquiry made by a station desiring to communicate.
TR	— .—.	Signal announcing the sending of particulars concerning a station on shipboard.
	— —. .— —	Signal indicating that a station is about to send a high power.
PRB	Do you wish to communicate by means of the International Signal Code?	I wish to communicate by means of the International Signal Code.
QRA	What ship or coast station is that?	This is ....
QRB	What is your distance?	My distance is ....
QRC	What is your true bearing?	My true bearing is.........degrees.
QRD	Where are you bound for?	I am bound for.............
QRF	Where are you bound from?	I am bound from............
QRG	What line do you belong to?	I belong to the.............Line.
QRH	What is your wave length in meters?	My wave length is..........meters.
QRJ	How many words have you to send?	I have .... words to send.
QRK	How do you receive me?	I am receiving well.
QRL	Are you receiving badly? Shall I send 20? ...—. for adjustment?	I am receiving badly. Please send 20. ...—. for adjustment.
QRM	Are you being interfered with?	I am being interfered with.
QRN	Are the atmospherics strong?	Atmospherics are very strong.
QRO	Shall I increase power?	Increase power.
QRP	Shall I decrease power?	Decrease power.
QRQ	Shall I send faster?	Send faster.
QRS	Shall I send slower?	Send slower.
QRT	Shall I stop sending?	Stop sending.
QRU	....	I have nothing to transmit. I have nothing for you.
QRV	Are you ready?	I am ready. All right now.
QRW	Are you busy?	I am busy (or, I am busy with.... Please do not interfere).
QRX	Shall I stand by?	Stand by. I will call you when required.
QRY	When will be my turn?	Your turn will be No. ....
QRZ	Are my signals weak?	Your signals are weak.
QSA	Are my signals strong?	Your signals are strong.
QSB	Is my tone bad? Is my spark bad?	The tone is bad. The spark is bad.
QSC	Is my spacing bad?	Your spacing is bad.

(When an abbreviation is followed by a mark of interrogation, it refers to the question indicated for that abbreviation)

ABSCISSA—A term in geometry adapted to radio use in making *curves* to show various values. The abscissae are the horizontal lines, and the ordinates are the vertical lines. In the diagram, the line AY is the axis of *ordinates* (*q.v.*) and the line AX the axis of abscissae. Fig. 2 shows a simple curve indicating the relation between voltage and current (amperes) in an *arc* (*q.v.*). The volts are given as ordinates and the amperes are shown as abscissae.

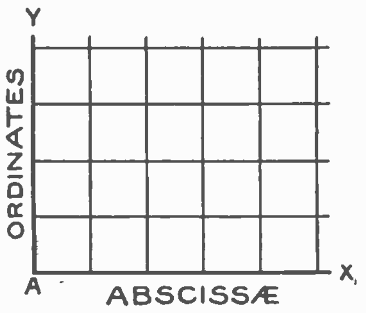

Fig. 1. Indicating axis of abscissae by the line A to X

The distance of any point from the axis of ordinates, measured on a line parallel to the other axis represents the value of the quantitv and locates a point on the curve. Where abscissae and ordinates are known, it is a simple matter to plot a series of points and then draw the curve through these points. (See *curve.*) Thus, in Figure 2, the curve shows the operation of the arc to be such that when the voltage is zero, the current is 20 amperes; when the voltage is 60, the current is 6 amperes, etc.

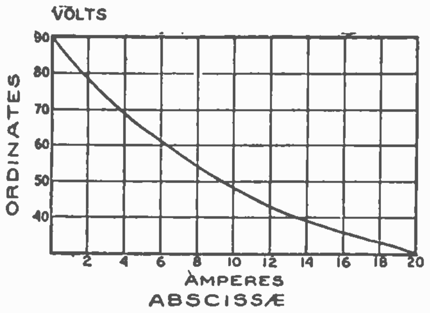

Fig. 2. A simple curve to illustrate the relation between volts and amperes

ABSORPTION, ELECTRIC—(See *Soaking-in.*)

ABSORPTION MODULATION—In the production of modulated (see *modulation*) waves, used to transmit speech in radio, it is necessary first to produce *undamped* (*q.v.*) waves of high or *radio frequency* (*q.v.*). Then other waves of low or *audio frequency* (*q.v.*) are superposed or run together with them. Now, if we insert a *microphone* (*q.v.*) in the *aerial circuit* (*q.v.*) of the transmitting station, providing that the *resistance* (*q.v.*) of that instrument is such as to reproduce accurately the vibrations of the voice or music, it is possible to vary the current output of the system producing the high frequency currents, so as to transmit the voice vibrations. If the microphone has resistance comparably low with the antenna or *antenna resistance* (*q.v.*), then the high frequency power output is handled (apportioned) equally between the aerial and microphone. The result is the production of fairly well modulated waves. There are, however, better methods of modulating the waves for which see *Modulation.*

ABVOLT—The *CGS unit* (*q.v.*) indicating the pressure or induced E.M.F. (*electromotive force*) (*q.v.*) set up between the ends of a wire one centimeter long, moving with a speed of one centimeter per second across the magnetic lines of force (*flux lines*) (*q.v.*) of a uniform magnetic field of unit intensity or flux density. An abvolt is therefore a very small unit and 100,000,000 (10^8) are required to equal one volt. In all formulae for determining the output or E.M.F. that can be expected to be produced by a certain *alternator* (*q.v.*), the result is in abvolts, and the complete formula usually provides for obtaining the actual volts. This is done by dividing the result in abvolts by 100,000,000 or 10^8 as it is generally expressed. (See *Alternator* also *abampere.*)

A. C.—Abbreviation for Alternating Current. (See *Alternating Current.*)

ACCELERATION—The rate at which the speed of a body in motion increases within a certain period of time. If we assume a train running at a certain rate of speed when the brakes are applied, the speed will decrease, and if measured over a period of one second, there will be a definite amount of decrease or a rate of decrease. This is known as *negative* acceleration or *deceleration.* When the train is speeded up there will be a definite change (increase) in velocity. The rate of change of velocity per unit of time is the acceleration. (*See Velocity* also *Frequency.*)

ACCEPTOR—A supplementary combination of *inductance* (*q.v.*) and *capacity* (*q.v.*) tuned to the frequency of the desired signal and connected in series in the receiving antenna. Generally, a *coil* (*q.v.*) and *condenser* (*q.v.*) of a value to correspond with a certain given wavelength—thus literally to "accept" the desired signal. (See also *Rejector.*)

A series combination of *inductance* (*q.v.*) and *capacity* (*q.v.*) connected in series with a high-frequency C.M.F. When tuned to resonance with the impressed E.M.F., the impedance drops to a low value, limited by the resistance of the system

ACCUMULATION OF ELECTRICITY—The process of storing electrical energy as (1)—a *storage battery* (*q.v.*) in which electric charges are placed to be later withdrawn and expended at will for various purposes, or (2)—the collection of energy by a *condenser* (*q.v.*) wherein the *electrons* (*q.v.*)

or minute charges of electricity are stored up momentarily on the plates or electrodes of the *condenser* (*q.v.*) and then released into the circuit.

ACCUMULATOR—(See *Storage Battery.*)

ACID—An active chemical compound formed by combining hydrogen with various acid radicals. Acids are much used in radio, particularly in *storage batteries* (*q.v.*). The most commonly used acid in this connection is sulphuric acid—a compound or union of hydrogen, sulphur and oxygen in certain definite proportions as indicated by the chemical symbol H_2SO_4. (See *Flux, Soldering.*)

ACIDIMETER or ACIDOMETER—An instrument used to determine the *specific gravity* (*q.v.*) of an acid solution. It is similar to a hydrometer and used in the same manner. Much used for testing strength or purity of acids or solutions. Operates on the principle that the strength of an acid will be directly proportional to the quantity of carbonic acid gas which it will liberate from a carbonate of soda or potash. (See *Hydrometer.*)

ACLINIC LINE—An imaginary line on the earth's surface assumed as passing through points where there is no magnetic inclination or dip of the needle of a compass. (See *Agonic Line.*)

ACOUSTICS—The science of sound or the study of the cause and effect of vibrations which effect the hearing. Generally speaking, the production and transmission of sound.

ACOUSTIC WAVE—The term occasionally used to denote a *sound wave.*

ACTINIC RAY—A light ray or beam of invisible *radiant energy* (*q.v.*). Rays of light, generally considered as lying at the blue or extreme left of the *spectrum* (*q.v.*) and being of such short *wavelength* as to be invisible to the eye. These rays have the power to induce or bring about chemical action. The most powerful actinic rays are the violet (X Ray) and the *ultraviolet.*

ACTIVE CONDUCTOR—(See *Conductor.*)

ACTIVE MATERIAL—The spongy part of a plate of a storage battery. The part which changes in nature and appearance due to the flow of electric current and having the ability to redevelop the current by a secondary chemical change.

ACTIVE PRESSURE—The active *electromotive force* (*q.v.*) or the pressure which produces a current. The term is used to distinguish this pressure from one impressed on the circuit.

A component in phase with the current in an *alternating current* (*q.v.*)

ACTIVE SPARK—A spark produced from energy contained in a charged condenser, which produces active *oscillations* (*q.v.*).

Thus, the spark discharge from a coil without any method of storing energy will be "inactive." When a condenser is inserted in such manner as to allow it to store up energy and then discharge, the resultant spark will be "active."

Fig. 1 shows a simple spark coil circuit with the secondary discharging across a gap. This is an Inactive Spark. Fig. 2 represents the same coil with a condenser in the circuit to store up the energy before discharging across the *gap* (*q.v.*). This is an Active Spark.

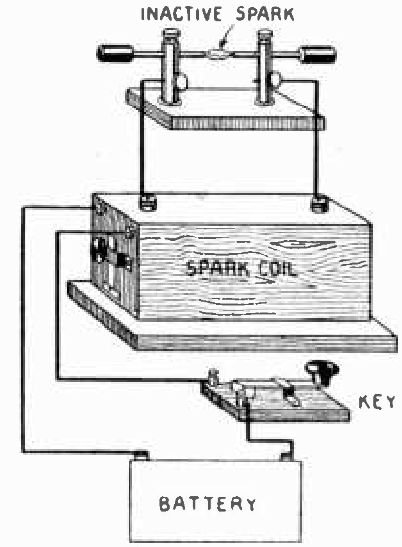

Fig. 1. Showing how an inactive spark is produced

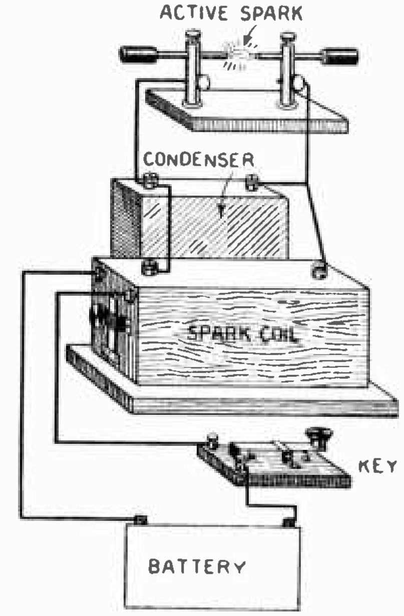

Fig. 2. The active spark is produced by means of a condenser in the circuit as shown

ADAPTER, AERIAL—A device for utilizing an electric light line as an aerial for the reception of radio signals. Essentially, two or more fixed mica condensers of comparatively high capacity, in series with two metal legs attached rigidly to a standard plug for insertion in socket of electric light fixture. Terminals are provided for connection to receiving apparatus. The illustration shows one of the popular types of Aerial Adapters which consists of merely a condenser to prevent the voltage of the line from passing directly to, or through, the instruments.

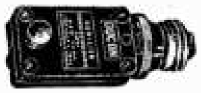

Illustration by courtesy of Dubilier Condenser and Radio Corp.

ADAPTER, TUBE—A device similar in form to a vacuum tube socket. Usually comprises a cylindrical form or shell of either metal or insulating material with legs identical with the tips or legs of a vacuum tube. A receptacle is allowed for the insertion of a tube, and the lead from each tip recess is so connected to a leg or tip at its base, as to permit the use of a non-standard base tube in a standard *socket* (*q.v.*). The illustration shows adapter for U.V. 199 or C-299 type of tube. When it is desired to employ a tube having a non-standard base in a set already equipped with standard sockets, without the necessity of changing the sockets, an adapter may be used. (See *Socket.*)

Illustsation by courtesy of Pacent Elec. Co., Inc.

ADHESION, ELECTRIC—Affinity of one body for another due to dissimilar charges of electricity passing through or being carried by them. The electrical equivalent of magnetic attraction.

ADJUSTABLE CONDENSER—(See *Variable Condenser.*)

ADJUSTABLE GRID LEAK—A form of *grid leak* (*q.v.*) that can be easily adjusted over a wide range of values without removal from the circuit. The illustration shows one type of Adjustable Grid Leak.

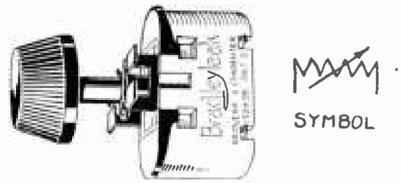

Illustration by courtesy of Allen-Bradley Co.

ADJUSTABLE RESISTANCE—(See *Variable Resistance.*)

ADMITTANCE—The inverse of *impedance* (*q.v.*). In an *alternating current* (*q.v.*) the Admittance acts in the opposite manner to impedance. A circuit having low impedance is said to have relatively high Admittance.

AERIAL—A system of wires suspended in the air or in any form in which they may be insulated or kept free from surrounding objects, used for the purpose of receiving or transmitting impulses as in radio transmission and reception. The term Aerial is actually identical with *Antenna*, but due to the general English definition of Antenna, authorities are gradually coming to distinction between transmission and reception Aerial systems by referring to a receiving device of this nature as an Antenna, and when used to transmit, as an Aerial.

There are different types and forms of Aerials, varying mainly according to purpose and facilities. The standard and perhaps more commonly used receiving Aerial consists of a single wire as in Fig. 1 or several wires as in Fig. 2. The wire used is usually of copper, as this metal combines low resistance with economy, strength, and durability. The theory of the action of the waves either received by or transmitted from an Aerial system will be found under the heading "Electromagnetic Waves."

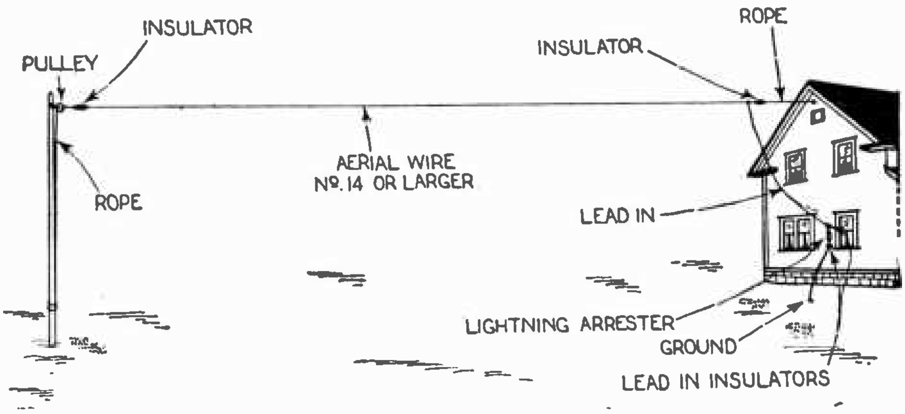

Fig. 1. A single wire aerial in the form of an inverted L. If a long wire is used, the lead-is may be connected in the center, forming a T type

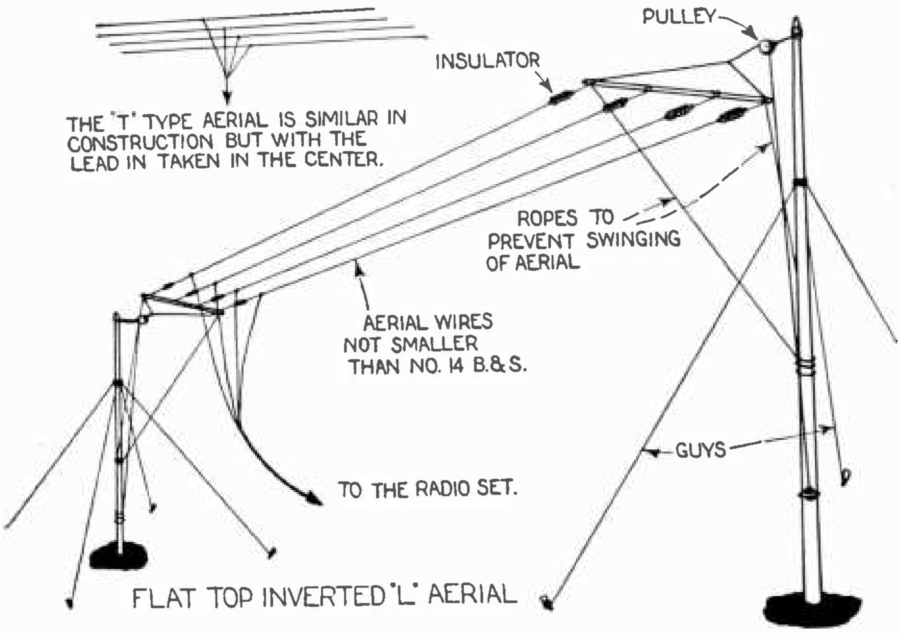
THE 'T' TYPE AERIAL IS SIMILAR IN CONSTRUCTION BUT WITH THE LEAD IN TAKEN IN THE CENTER.
INSULATOR
PULLEY
ROPES TO PREVENT SWINGING OF AERIAL
AERIAL WIRES NOT SMALLER THAN NO. 14 B.&S.
GUYS
TO THE RADIO SET.
FLAT TOP INVERTED 'L' AERIAL

Fig. 2. Method of erecting a multi-wire aerial. This type of aerial may be made in either the inverted L or T form as shown

The exact type and extent of an Aerial will depend entirely on the available facilities and the particular purpose for which it is intended. These various forms will be taken up under their respective headings. In the receiving Aerials, Figs. 1 and 2, a certain definite *wavelength* (*q.v.*) is obtained by arranging a certain length to the wires of the Aerial system including the *ground* and *lead-in.* (See *Fundamental Wave-Length of Aerials.*) In the single wire Aerial, the lead-in is usually taken from the end nearest the receiving apparatus, but where a particularly long Aerial is to be used, it is customary to attach the lead-in to the centre of the suspended wire. The former is known as an *inverted L* type and the latter as a *T* type Aerial. These designations will apply as well where a number of wires are used as in Fig. 2.

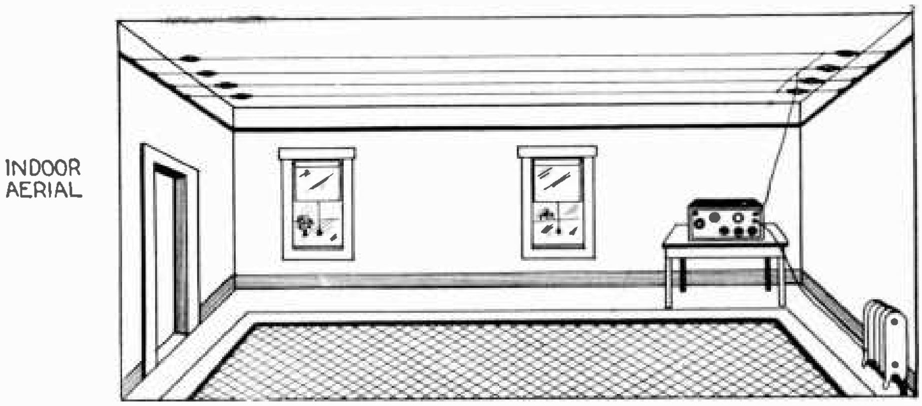

Fig. 3. Showing one method of installing an indoor aerial. While this illustrates a multi-wire type, a single wire run around the moulding of the room may give excellent results

Generally speaking, an Aerial should be as high above the ground as possible, of course within practical limits. From 25 to 100 feet is the most practical height to use. Where the Aerial is suspended at a substantial height the incoming waves will be less likely to be obstructed by surrounding trees or buildings, and in the case of transmission the same factor will apply, inasmuch as the waves may be obstructed in their transmission, if buildings or trees closely surround the Aerial.

The insulation of an Aerial is of particular importance, and especially so in the case of a transmitting Aerial. The reason for insulating an Aerial is to prevent the currents from escaping to the ground by way of trees, buildings, or the masts to which the Aerial is attached.

Fig. 3 shows a method of arranging an indoor Antenna system where it is not possible to use the out-door type. The construction may be essentially the same as the usual form of outdoor Aerial and can be arranged in a number of different ways. Very often a wire is merely suspended around two sides of a room and insulated from the moulding. One end will be connected as a lead-in to the receiving apparatus. In constructing an indoor Antenna, it is well to remember that the effective length of the Antenna is not much greater than the direct distance from the receiving set to the farthest point of the Antenna.

Fig. 4 shows a different type of Aerial known as the *Cage* type. The illustration shows an inverted L, and the T arrangement. Fig. 5 shows a *Fan Aerial* and Fig. 6 the *umbrella* type. The illustration, Fig. 7, shows one type of *loop Aerial* (*q.v.*) and Fig. 8 the same general type in somewhat different form. Owing to the fact that the theory of operation of a *loop Aerial* is somewhat different than the ordinary types, it will be described more fully under the heading "Loop Aerial." Fig. 9 illustrates a form of *Aerial Adapter* which is arranged to allow the electric light wires to be used as an Aerial. (See *Adapter Aerial.*) Fig. 10 shows an *underground Aerial* system designed by Dr. Rogers and which is not appreciably affected by atmospheric disturbances. (See *Strays.*) This form of Aerial, however, is generally impractical due to the difficulties in construction.

Aerials for transmitting may be of the same general types as receiving Antennas arranged outdoors, with the exception as explained previously, that heavy insulators must be used owing to the high voltage employed. Aerial for transmitting will depend for size on the wave-length desired.

Fig. 11 illustrates several types and arrangements of *counterpoise* (*q.v.*).

In the case of the receiving Aerial a single wire may be from 75 to 150 feet in length, or if a number of wires are used, the over-all length may be considerably less. (See *Fundamental Wave-length.*)

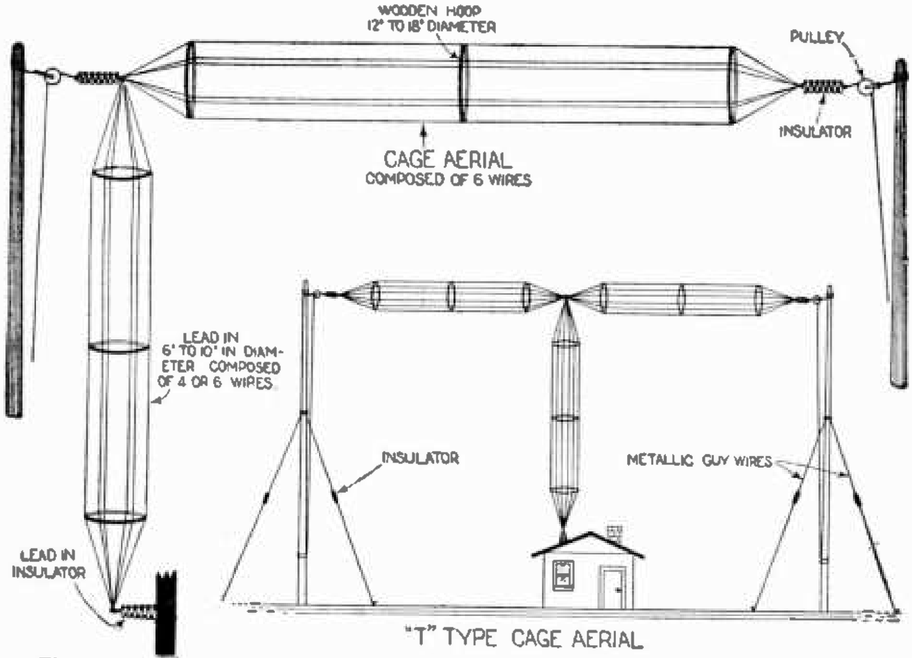
WOODEN HOOP
12" TO 18" DIAMETER
PULLEY
INSULATOR
CAGE AERIAL
COMPOSED OF 6 WIRES
LEAD IN
6' TO 10' IN DIAMETER COMPOSED OF 4 OR 6 WIRES
LEAD IN
INSULATOR
INSULATOR
METALLIC GUY WIRES
"T" TYPE CAGE AERIAL

Fig. 4. Cage aerials are usually employed for transmitting purposes. The inverted L and T types are erected as shown

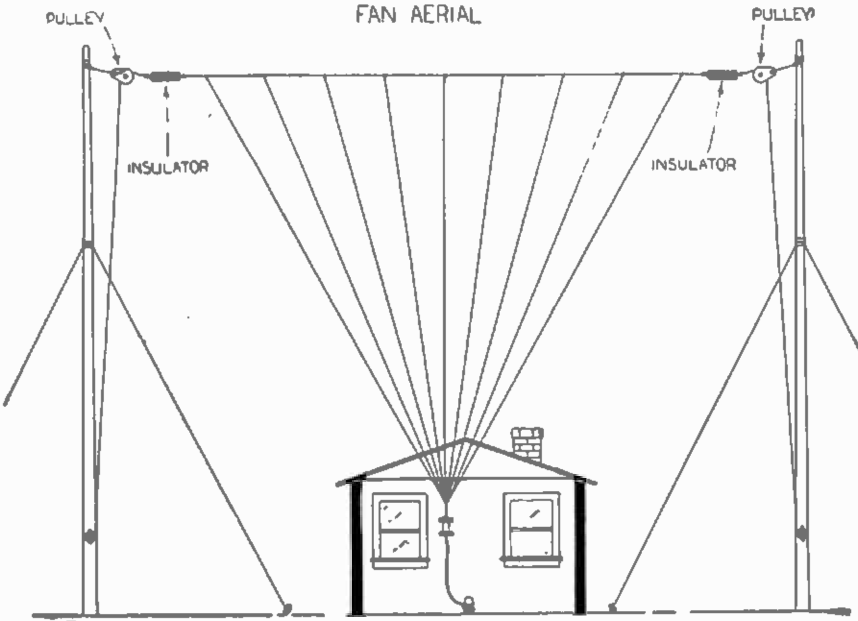
PULLEY
FAN AERIAL
PULLEY
INSULATOR
INSULATOR

Fig. 5. The fan aerial has its wires spread apart at the farthest point, forming a fan

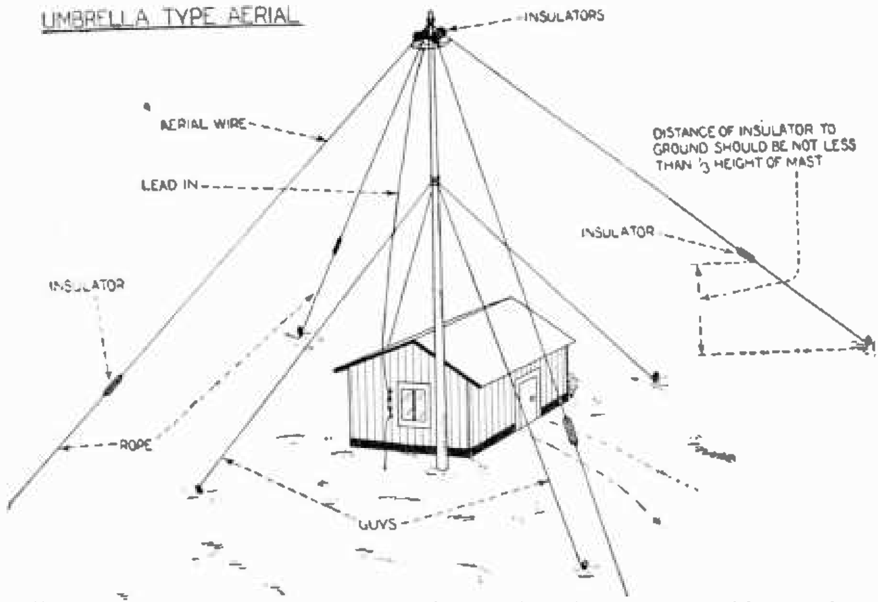
UMBRELLA TYPE AERIAL
INSULATORS
AERIAL WIRE
LEAD IN
DISTANCE OF INSULATOR TO GROUND SHOULD BE NOT LESS THAN ⅓ HEIGHT OF MAST
INSULATOR
INSULATOR
ROPE
GUYS

Fig. 6. The umbrella type of aerial derived its name from the arrangement of its wires in the form of an umbrella

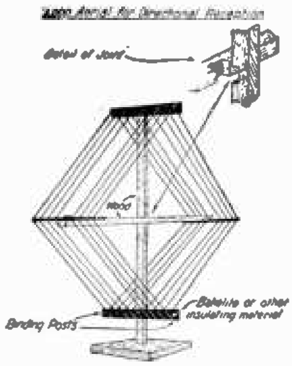
Detail of Joint
Wood
Bakelite or other insulating material
Binding Posts

Fig. 7. The selenoid type loop aerial

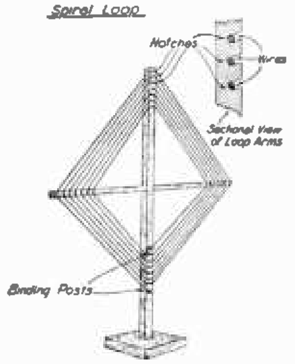
Spiral Loop
Notches
Wires
Sectional View of Loop Arms
Binding Posts

Fig. 8. The spiral type loop aerial

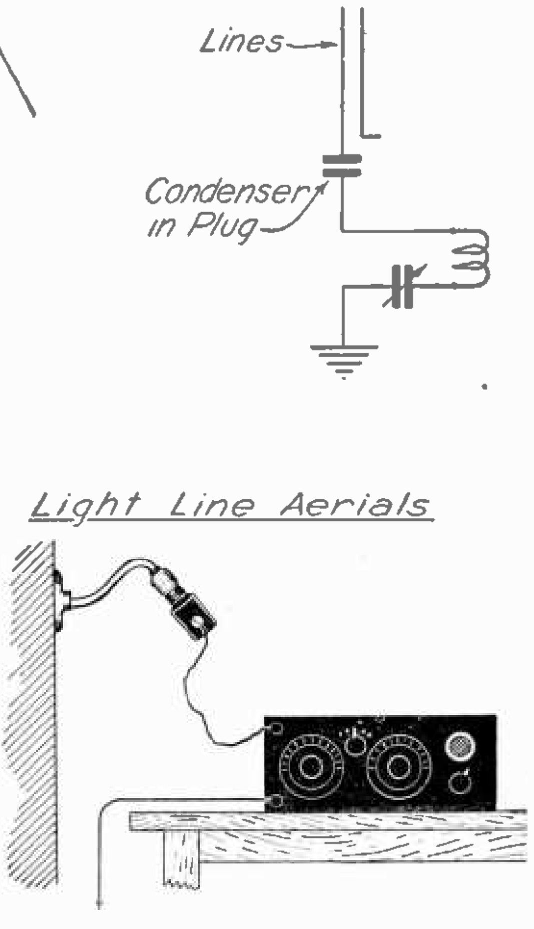
Lines
Condenser in Plug
Light Line Aerials

Fig. 9. Method of employing the electric light line as an aerial with an aerial adapter. Above is shown the diagram of the condenser in connection with the aerial circuit of a radio set

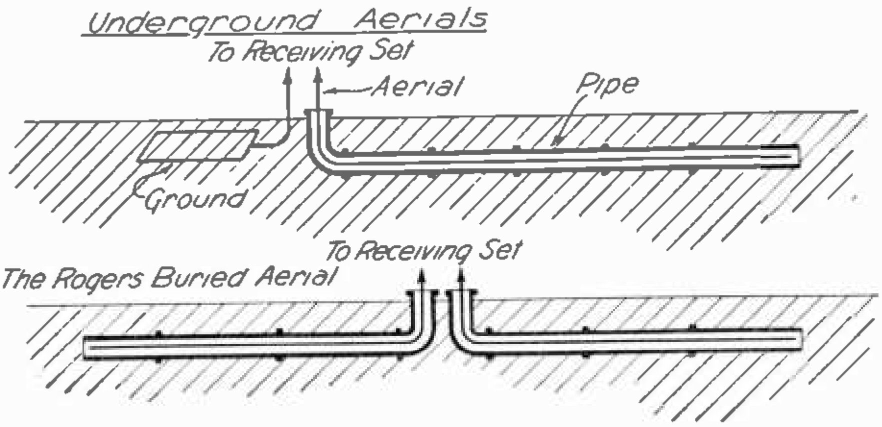

Fig. 10. Showing how wires are buried in the ground for aerials

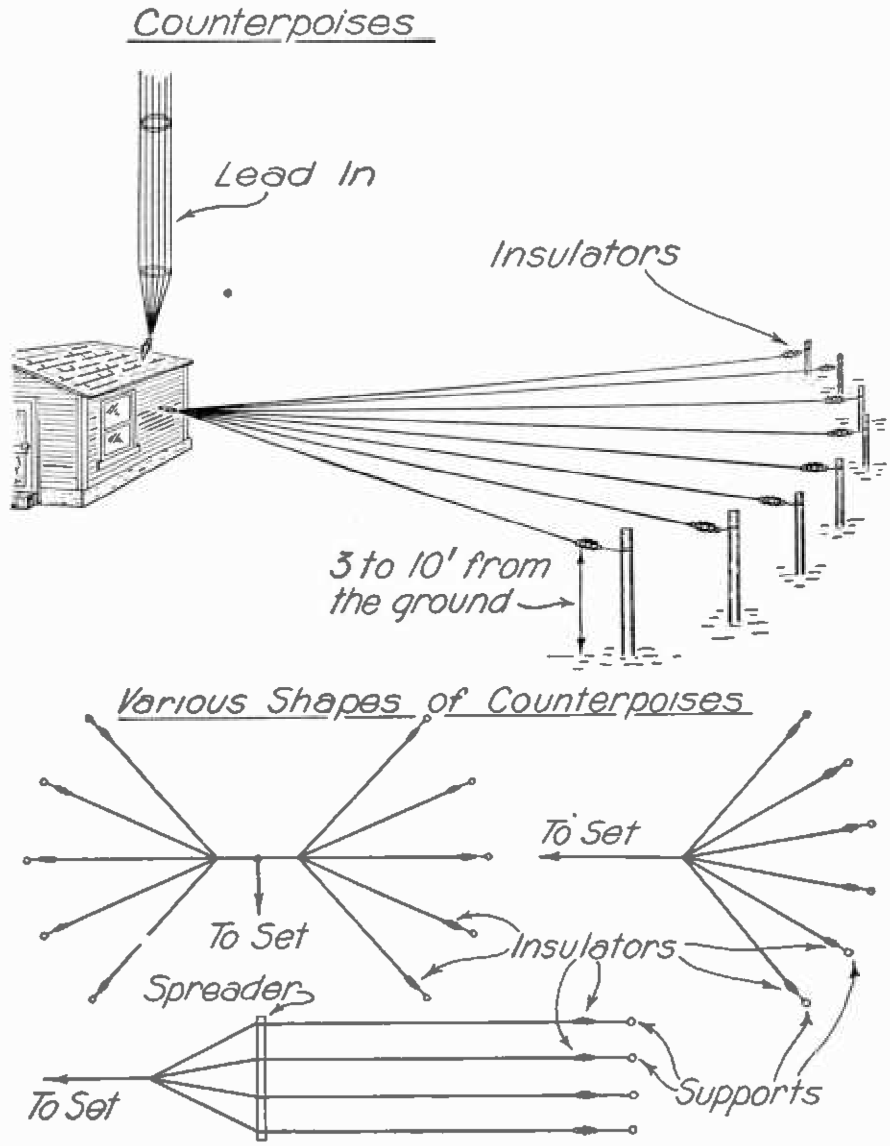

Fig. 11. Different methods of arranging counterpoises for either transmission or reception

AERIAL CAPACITY—(See *Capacity of Antenna.*)

AERIAL CIRCUIT—Consists of *aerial* and *earth* or *ground*, including all *coils* or *inductances*, *condensers*, etc., which may be connected with the aerial and earth and forming a direct path between those points.

AERIAL SWITCH—A device to transfer the *aerial* and *ground* connections from the transmitting circuit to the receiving circuit or vice versa. Its object is to separate the high tension current of the transmitter from the low tension of the receiver. The switch is usually installed in a convenient position so that the operator may change from transmitting to receiving by one motion of the hands. A simple form of aerial switch is illustrated in Fig. 1. The circuit for same is shown in Fig. 2. A more efficient circuit is the one shown in Fig. 3. This circuit requires another blade to control the primary current to the transformer. The latter circuit is most satisfactory as it entirely disconnects the current to the transformer, doing away with the possibility of dangerous shocks to the operator. The switch commonly used in such circuits is shown in Fig. 4.

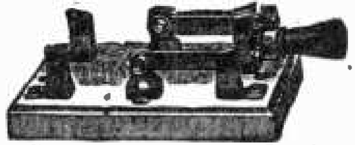

Fig. 1. A simple Aerial Switch

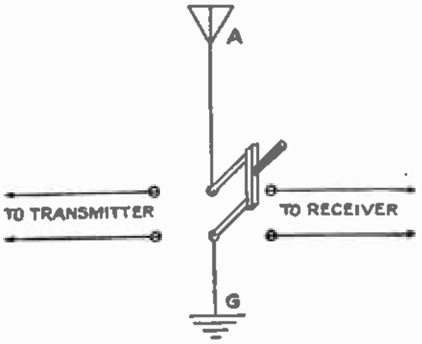

Fig. 2. Circuit for a simple Aerial Switch

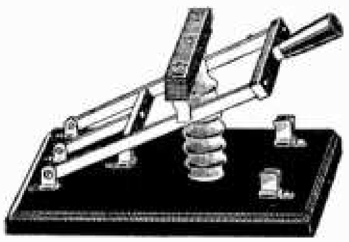

Fig. 4. An Aerial Switch with an extra blade

The switches used in the various commercial wireless stations are of course more complicated. There is a need to control many circuits with the simplest possible method. In order to do this it was found more convenient to only change the aerial connections, and allow the ground connection to both the receiver and transmitter to remain permanently. This of course necessitates a small air gap in the aerial for the transmitter. This is called an *anchor gap* (*q.v.*). This small appliance allows the transmitter to be connected continuously to the aerial as the high voltage from the transformer discharges across the small air gap, as provided for in the *anchor gap* (*q.v.*). Fig. 5 shows the circuit when this type of aerial transfer is to be used. When the switch handle is pushed down it makes connections with the receiver, when it is raised, the transmitter is ready to be used as the primary current is connected also.

AERIAL TUNING CONDENSER—A *condenser* (*q.v.*), usually variable, connected in the *aerial circuit* (*q.v.*) for the purpose of adjusting the natural period of the receiving circuit to the period of the incoming waves (see *Wavelength*) or signals coming from the *aerial*. In the illustration is shown an Aerial Tuning Condenser.

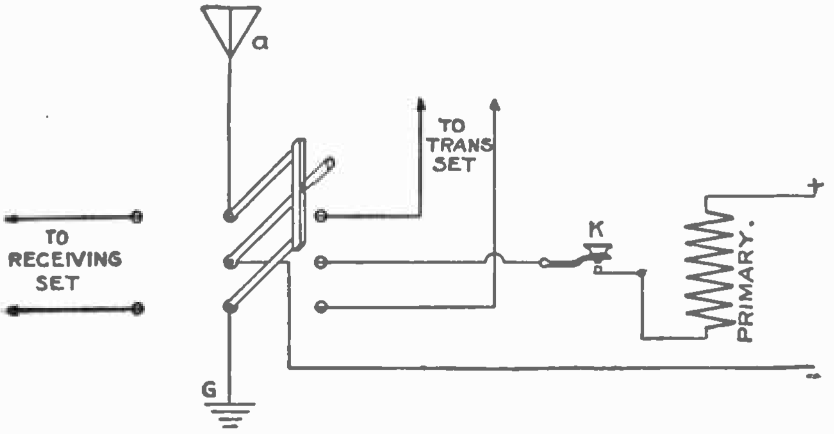

Fig. 3. Circuit for the Aerial Switch shown in Fig. 4 on page 7

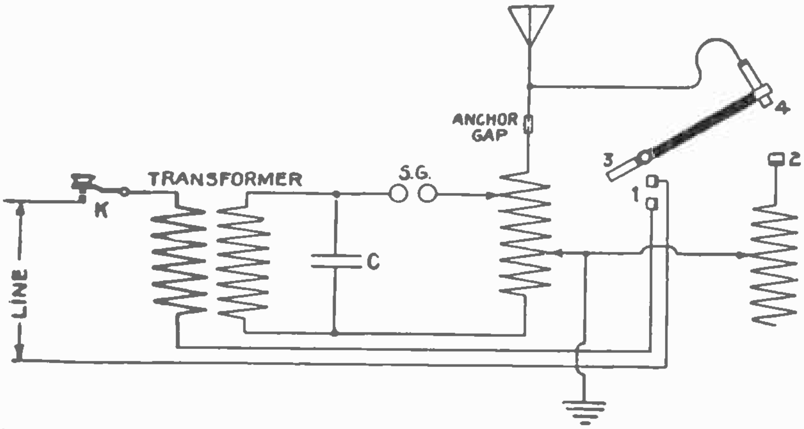

Fig. 5. An Aerial Switch employed in conjunction with an Anchor Gap as explained on page 7

Any *capacitive* (*q.v.*) means of establishing or varying the *oscillation constant* (*q.v.*) of the receiver.

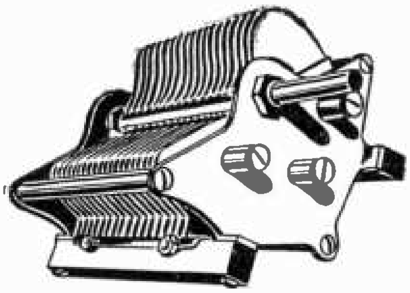

A variable condenser such as commonly used for an Aerial Tuning Condenser (see page 7).

AERIAL WIRE—The wire forming the *aerial* (*q.v.*). Copper is generally used for this purpose as it offers very little *resistance* (*q.v.*) to the feeble currents in receiving.

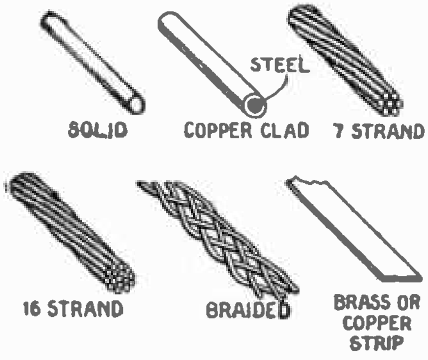

The illustrations show various types of wire as used for aerials. Among them there are the solid copper wire, copper clad steel wire, stranded wire, braided wire and brass or copper strip.

AERO-FERRIC INDUCTANCE—The *inductance* (*q.v.*) of coils wherein the *magnetic circuit* (*q.v.*) is completed through both air and iron.

AEROPLANE ANTENNA—Receiving or transmitting *antenna* (*q.v.*) used on an aeroplane. The commonest form is a suspended wire shown as A in the illustration. The *ground* or *counterpoise* (*q.v.*) in this case may be another wire suspended on and insulated from the fuselage as indicated at B. When the plane is in flight, the hanging wire trails up and backward, depending on the speed of the plane and the size of the weight used. This type of antenna will be slightly *directional* (*q.v.*) in the direction the plane is traveling. If it is impractical to use a trailing wire antenna for any reason, a loop or coil of wire may be used, the results of course being less satisfactory. The metal frame of the plane is often used as a *counterpoise* (*q.v.*)

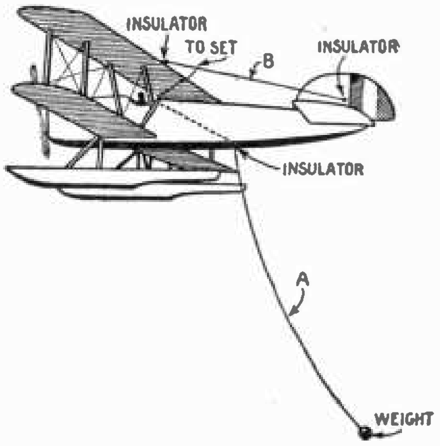

A. F.—Abbreviation for *Audio-Frequency* (*q.v.*)

AFTER-GLOW—Fluorescent phenomena in a *vacuum tube* after current has been withdrawn.

AGONIC LINES—Lines imagined as passing through points on the earth's surface where the magnetic inclination or dip of a magnetic needle (compass) is zero. (See *Aclinic Lines.*)

AGING, VACUUM TUBE—Gradual diminishing of the brilliancy of a tube due to the deterioration of the *filament* (*q.v.*) and the coating or deposit on the bulb. A tube is said to be Aging when it gradually loses its power to emit *electrons* (*q.v.*)

AIR CONDENSER—*Condenser* utilizing air as the *dielectric*. The majority of variable condensers used in radio work employ the air dielectric, that is, the space between plates is merely an air gap. Most *fixed condensers* (*q.v.*) use some substance such as mica, or sheets of waxed paper as a dielectric.

AIR CORE TRANSFORMER—A *transformer* used in radio, having no metal core, *the lines of force* having their path through air. (See *Amplifier, Radio Frequency.*)

ALBIZ, Count—Managing Director of the Compañia National de Telegrafia sin Hiols in 1910, of Scotch origin.

He was born in Madrid in 1858 and received a schooling at the Madrid University and London University College. He was a Tory member of Parliament.

ALEXANDERSON, Ernst Fredrick Werner (1878)—Radio engineer and inventor. Born at Upsala, Sweden, January 25th, 1878, he was educated at the High School and University of Lund, Sweden, at the Royal Institute of Technology, Stockholm, and at Berlin. In 1902 he joined the General Electric Company, and has been for some years their consulting engineer. He holds the post of chief engineer to the Radio Corporation of America, and

Photo by courtesy of General Elec. Co.

E. F. W. Alexanderson

is a member of the American Institute of Radio Engineers. Alexanderson has read many papers on electrical subjects before the chief technical societies of America.

Alexanderson is famous for his work on *high-frequency alternators* used in Radio telegraphy. The Alexanderson alternator is connected directly or *inductively* to the aerial and earth, and constitutes the simplest possible connection for producing *continuous* or *undamped waves.* It is a machine of great speed and many *field poles*, and frequencies as high as 200,000 *cycles* are obtained. For long distance work, employing very *long wavelengths*, this high-frequency alternator is largely used. Alexanderson is also responsible for the *magnetic amplifier*, patented in 1913, and has carried out successful experiments on *duplex wireless telephony.* In the Alexanderson *microphone transmitter* the *modulation* is mainly effected by variations in the tuning of the *aerial circuit.*

ALEXANDERSON ALTERNATOR—A type of *high-frequency alternator* (*q.v.*) which produces *continuous oscillations* (*q.v.*)

ALLOY—A compound of two or more metals. (See *Woods Metal.*)

ALTERNATING CURRENT—An electric current that does not flow steadily in one direction—as from *positive* to *negative*—but completely reverses its direction or *polarity* (*q.v.*) at certain definite intervals. In other words, the current flows first in one direction and then reverses and flows in the opposite direction—these changes being referred to as changes of *phase* (*q.v.*). An electric current is either *direct* (*q.v.*), alternating, or *pulsating*. The latter is an alternating current that has been rectified. (See *Rectified Current.*) Practically the only direct current that is employed or appears in a radio receiver is that supplied by the *"A"* and *"B" Batteries* (*q.v.*). In an alternating current, the change in direction or polarity takes place in a steady, even manner.

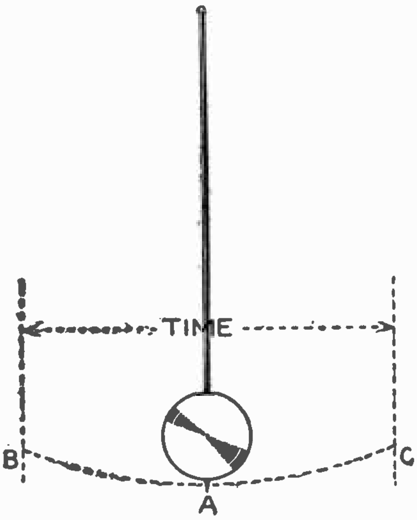

Fig. 1

An understandable analogy for alternating current is its comparison with a pendulum. In Fig. 1 the pendulum is supposed to be the alternating current or impulse. Now when the bob is swung from its natural (neutral) position A, which may be considered as zero value of voltage or current, it will move to a maximum point, either B or C. If we assume B to be the *positive* side, then the bob will move to B which will be the maximum positive *amplitude* (*q.v.*). It will then fall back to zero again, and with the power behind it, the momentum in this case, will rise on the other side to C, which is the maximum *negative* amplitude. The bob will then fall back to A once more and will have undergone a complete change of direction and values; it will have gone through a *cycle* (*q.v.*). That is to say, it will have risen from zero A to positive maximum B, then back to A, rising on the other side to negative maximum C and back to A. Now the length of time required for this complete cycle (see *Alternation*), is known as the *time period* of vibration or *phase change*. If, for example, it required one-sixtieth of a second for this change to take place, then we say that the *frequency* (*q.v.*) is sixty cycles, because obviously, a change requiring one-sixtieth of a second will occur sixty times in one second.

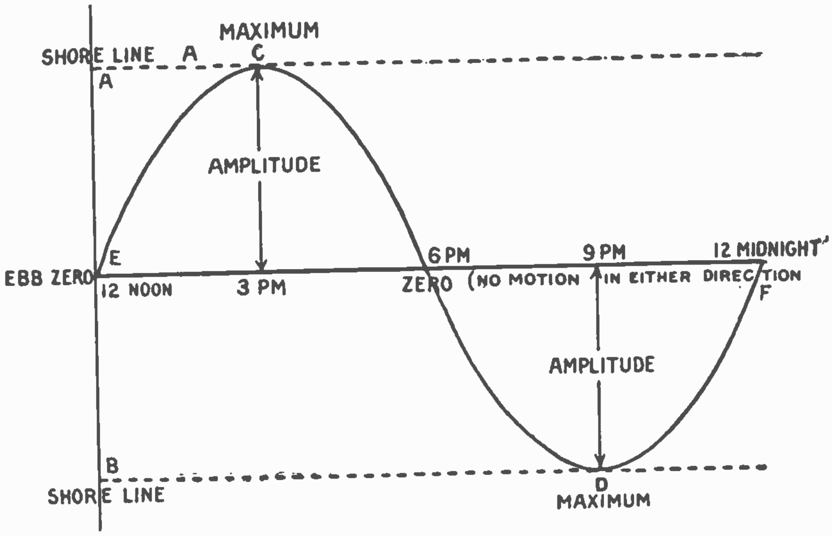

Fig. 2

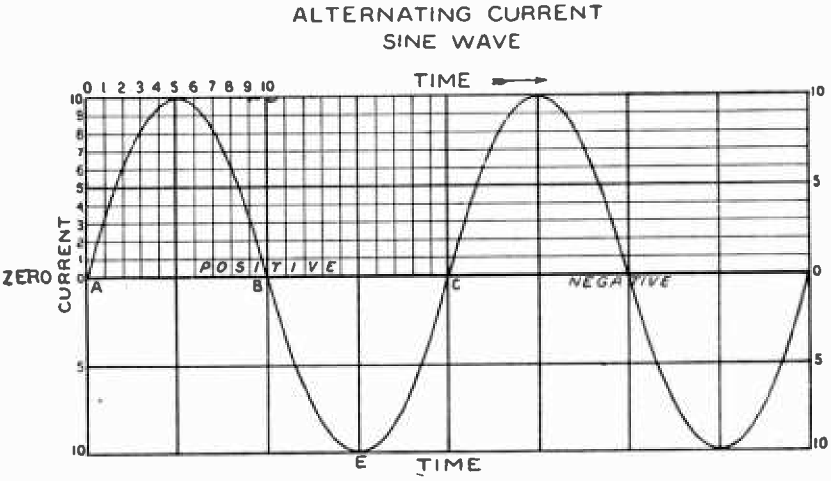

Fig. 3

The action of alternating currents may also be likened to that of ocean tides. In Fig. 2 the curved line represents the water in motion and is to be compared with current in motion in an alternating circuit. For the sake of clarity, we can assume that the tide rises and falls between two shore lines, A and B. At 12 noon the tide is at zero or "ebb" and there is no motion in either direction. Now the flow starts toward shore line A and gradually rises toward that side until at 3 P. M. it has reached its maximum height. In the case of an alternating current, we may consider that at 3 P. M. the current has reached maximum value on the positive side or in a positive direction. It then reverses and falls toward zero until at 6 P. M. it has reached the zero point. In the case of the current, it will have become zero. Motion again starts in the other direction and the tide rises once more, but in the opposite direction, toward shore line B. By 9 P. M. it will have reached maximum height in this direction, or in the case of the current it has reached maximum value or *amplitude* (*q.v.*) at D on the negative side. It then falls back again to zero, and at 12 midnight has come back to its original starting point, having gone through all stages of value or height from zero to maximum at one shore and then back to zero; then from there to maximum at the other shore and back to zero. The action of the alternating current then will have been to flow from zero to positive maximum, back to zero and reversing, from zero to maximum on the negative side, then back to zero. It will have completed one cycle. The time required for the tide to go through these changes E to F is 12 hours and is known as the *time period* or *period* (*q.v.*). Therefore, the time required for a complete change in the case of alternating current will be referred to as the period. The number of times that these changes take place within a certain time limit is known as the *frequency* (*q.v.*)

The current value in Fig. 3 is assumed to be 10 amperes, and the current values at any moment are shown as *ordinates* (*q.v.*), (vertical) while the time is shown in the *Abscissas* (*q.v.*) (horizontals). If the above curve is to be assumed as representing the Sine wave of an Alternating current having a frequency of 60 cycles per second, the total time in the chart would be one-thirtieth of a second. Thus, Time A to C would be one-sixtieth second, the Time of one complete cycle. A B is a positive Alternation and B C a negative Alternation, D and E represent maximum current values in each direction.

If the above Sine Wave is to be used as an example of a high frequency oscillation as in radio, the Time of one cycle can be assumed as about $\frac{1}{1,000,000}$ second. (See *Frequency-Oscillation - Sine - Phase - Characteristic of Alternating Current.*)

The frequency of an alternating current depends on the number of poles (field magnets) and the speed of rotation when the A.C. is generated by an alternator. (See *Alternator.*)

ALTERNATING CURRENT, THEORY OF PRODUCTION OF—Electric current in motion along a conductor always has a magnetic field (*magnetic flux*) surrounding it. If this current is passed near a number of turns of wire, the lines of flux will link with the wire turns in the circuit. The number of flux lines through a coil will depend upon the current. A change of current will change the number of linkages. If two turns of wire are used, the circuit will link twice with the same magnetic flux. Thus, a change in the current or a change in the number of turns of wire will alter the linkages.

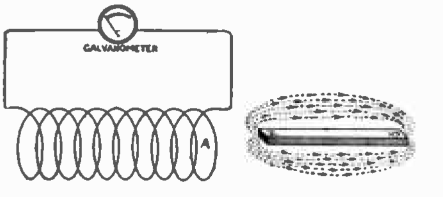

Fig. 1

Now in the case of Fig. 1, the magnet being placed near the coil causes lines of force to pass through the coil.

Each turn of wire will link the flux lines. Now any *change* in these linkages between the flux lines and wire turns will produce an E.M.F. (*electro motive force*) in the circuit of the wire turns. This is called an "induced current" because it is induced by the change in linkages. In Fig. 1, we have a coil of wire in a circuit with some sensitive measuring instrument for minute currents, in this case a *galvanometer*. Near it we have a bar magnet. When this magnet is moved toward or within the coil, a current is induced in the circuit by reason of the change in linkages. This current is *always* in a direction to *oppose* the change that causes it. (See *Lenz Law.*) "Whenever an induced current arises by reason of some change in the linkages, the magnetic field about the induced current is in such direction as to oppose the change." When the magnet is moved toward the coil, the change in linkages creates or induces a current that flows in the circuit in such manner as to make A act as a North pole. As the North pole of the magnet is toward this point A, the similar poles will oppose each other according to the simple law of magnetic attraction and repulsion. Now if we attempt to withdraw the magnet, the change will alter the linkages again, this time creating an induced current that makes A act as a South pole, thus *attracting* the magnetic North pole and opposing in the opposite manner, the change in linkages that have created the current.

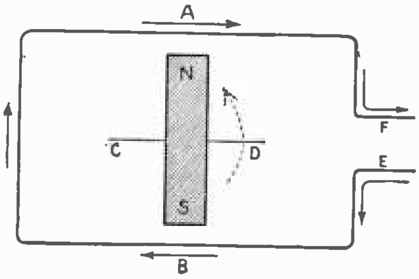

Fig. 2

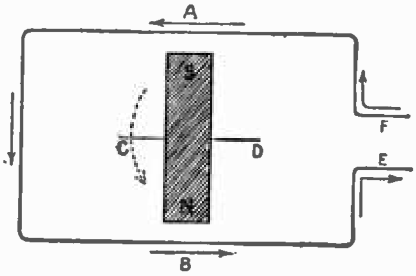

Fig. 3

We already know that if a conductor is moved across a magnetic field, an E.M.F. is induced in the conductor. The action is the same whether the conductor is moved across the magnetic field or vice versa. In Fig. 2, a magnet is shown in position on an axis—C D for rotation within the loop A B. As the magnet is rotated, the North pole passes across A, an E.M.F. is created toward F in the same direction. As the position of the magnet reverses in rotation and S passes A while N passes B, a current is induced in the *opposite* direction, both times in accordance with the law previously stated. Thus, if a meter is connected between the ends of the loop at E F and the magnet rotated, a current will be registered. With each half turn of the magnet, this current will reverse its direction. Therefore, we have produced an alternating current that changes direction with each reversal of the position of the magnets. This is the simple *AC generator* or *alternator*. The illustration, Fig. 4, shows a standard type of low frequency alternator.

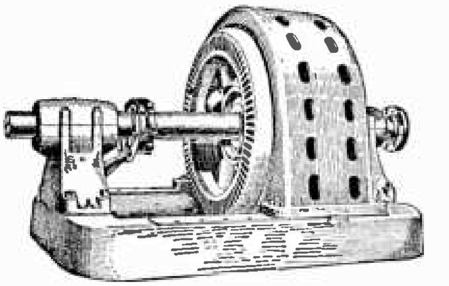

Fig. 4

ALTERNATION—In an alternating current, is the rise from zero to maximum *amplitude* and back to zero on either side, i.e., *positive* or *negative* directions. Two alternations—one positive and one negative—make a complete *cycle*. (See *Alternating Current.*)

ALTERNATOR—A machine for the purpose of changing mechanical energy into electrical energy. Fig. 1 shows an alternator in its elementary or simplest form—merely a conductor, B C moved across a *magnetic flux* (*q.v.*) (Note: For a complete explanation of the production of alternating currents, see *"Alternating Current, Theory of Production of."*) Here the conductor B C is moved to the right *across* the magnetic flux shown as dotted lines between N and S, the poles of the magnets. As the conductor moves across the flux lines of the magnet, an *electro-motive force* (*q.v.*) *or electric* current is set up or induced in the conductor, this force being in such a direction that the current produced will create a flux or magnetic field, in the same direction as the flux on the side of the conductor that comes first into contact with the flux lines N S on moving the conductor. In this case E.M.F. (electromotive force) is in the direction B to C in the conductor. In order to establish a circuit and utilize the current pro-

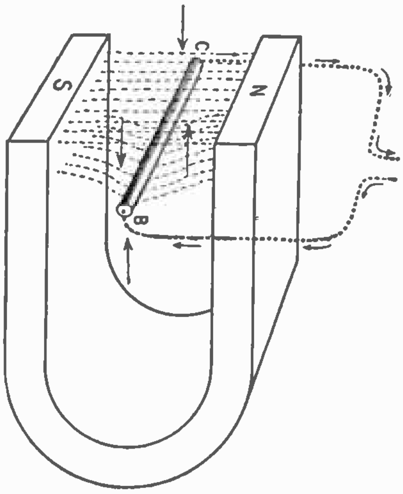

Fig. 1

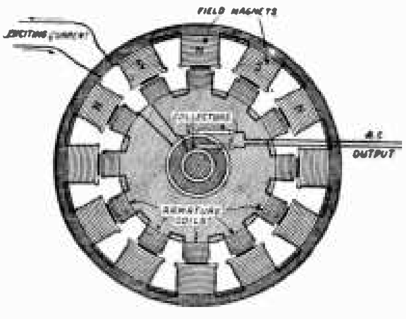

Fig. 2

duced in this manner, it is only necessary to attach wires to both ends of the conductor as shown by the dotted loop. Now at this moment the conductor is being traversed by a current having a definite polarity, that is to say, it is flowing in one direction only. When the direction of the flux lines N S are reversed, as by reversal of the magnet, or the conductor is turned so that the other end comes into contact with the flux lines, then the direction of the current in the conductor will reverse and it will flow in the opposite direction. In other words, with every reversal of the position of the conductor in its relation to the magnetic flux N S, or the reversal of the direction of the flux lines, with respect to the conductor, the E.M.F. that is created will change its direction of flow. (See *Alternating Current.*)

In an alternator it has been shown that the result is approximately the same whether the magnet is rotated or the conductor is rotated through the magnetic flux. In either case the moving member is known as the *rotor* (*q.v.*) and the stationary member is called the *stator* (*q.v.*). The magnets producing the flux lines may be stationary or they may revolve; in either case they are referred to as the *"field"* (*q.v.*).

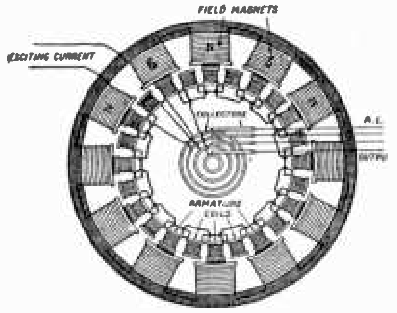

Fig. 3

It has been shown that with every change in the relative position of the field magnets and the *armature* (*q.v.*), irrespective of which member is the rotating one, a change of direction of the induced E.M.F. takes place. Then it is apparent that in order to increase the number of changes, that is, to increase the *frequency* (*q.v.*), it is only necessary to use (1) a greater number of field magnets, or (2) have the armature or conductor pass the magnets more frequently by means of greater speed in the rotating member, or by means of both. Now each of these changes is an alternation (see *Alternating current*) and two such changes will be a *cycle* (*q.v.*). Let us suppose that the moving member of a certain alternator has a rotary speed of 720 revolutions per minute and there are 10 poles or magnets. As frequency is figured by the changes per second, it will be necessary to reduce this number to the revolutions per second. Thus, there will be 12 revolutions per second. This figure multiplied by the number of *poles* (*q.v.*), which we assume as 10, indicates 120 alternations per second. As there are two alternations to each *cycle* (*q.v.*), the frequency in this case will be 60 cycles. The equation is as follows:

$$F = \frac{P}{2} \times \frac{N}{60}$$

where F is the frequency, P, the number of poles and N the speed of rotation per minute. The exact voltage or E.M.F. obtained will depend on the number of poles and the speed of rotation, and also on the distribution of the field windings and total flux from all the poles or magnets. In designing alternators the effective E.M.F. (electromotive force) is expressed by the formula:

$$E = \frac{K P \phi N Z}{10^8}$$

E is the effective electromotive force of the alternator.
K represents the E.M.F. factor of the alternator. This factor depends upon certain charcteristics of the poles and also the distribution of the windings on the armature.
P is the number of poles of the field magnates.
ϕ is the number of lines of magnetic flux that flow from one pole across the gap to the armature—known as the useful magnetic flux per pole; (see *Maxwells*).
N is the speed of rotation of the rotor in revolutions per second.
Z is the total number of conductors on the surface of the armature.
10^8 represents the number 100,000,000 —the calculation given would be determined in *abvolts* (*q.v.*), $\frac{1}{100,000,000}$ volt and it is therefore necessary to reduce it to volts by division.

In the "revolving field" alternator, as explained above, the field magnets are rotated and the conductor remains stationary. The windings in which the E.M.F. is produced, are placed in slots in the inner surface of the stator, the stator and the windings composing it being the *armature.*

In the alternators of the revolving armature type, the conductors or armatures are the rotating member and the field magnets remain stationary. Fig. 2 shows the revolving armature type of alternator in diagrammatic form. A small direct current *generator* (*q.v.*) may be used to excite or build up the field. That is to say, a current is sent through the field magnets (windings in this case) for the purpose of increasing or building up the *flux density* (*q.v.*). This is known as an *"exciter"* (*q.v.*). The exciter may be mounted directly on the end of the rotating shaft of the alternator, or mounted separately and connected to the alternator rotor shaft by means of a belt. This permits a heavy field to be built up by the electro-magnets used in place of the ordinary magnet shown in Fig. 1. If the machine is to be *single phase* (*q.v.*) there will be two *collector rings* (*q.v.*), the two phase type will have four rings, and the three phase six rings. The illustration Fig. 2 shows the general design of a single phase alternator. In this case there will be of course only two rings to collect the E.M.F. set up by the machine. Fig. 3 illustrates the general scheme of armature winding for a two phase machine. Here there are actually two separate E.M.F.'s set up by the separate windings and the ends of each of these windings are connected to collector rings as shown. In the three phase arrangement there may be three separate windings and six collector rings.

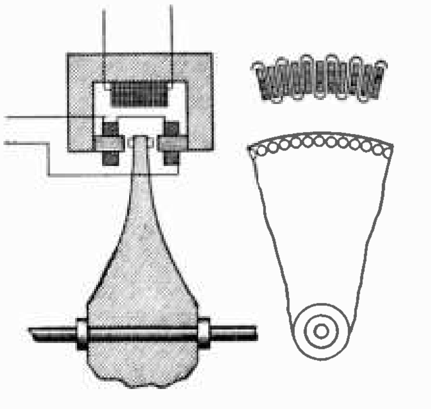

Fig. 4

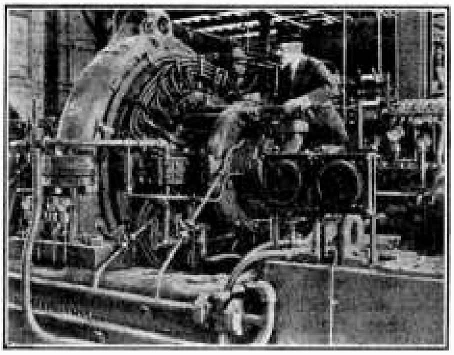

Fig. 5

The use of the revolving armature type for comparatively small voltages is due, as mentioned previously, to the fact that they are easily constructed, and also the advantage that the exciter need not be an external source, but may be a separate winding with a *commutator* (*q.v.*) mounted on the main shaft. The revolving field type of course allows much greater space for the armature winding, inasmuch as the stationary member is not limited to any particular space, whereas the moving member must be arranged as small as possible to do away with mechanical troubles. Thus, the stationary armature can be wound and properly insulated for great voltages. For this reason the revolving field is generally used in high power work. Another point of importance is that where the armature revolves, the windings must be made very carefully as the centrifugal force due to rapid rotation of the members places a great strain on them.

One of the widely used forms of alternator for extremely high frequencies, is the Alexanderson, named after its originator, *Ernst Alexanderson* (*q.v.*), a noted engineer. This machine is of the inductor type. It is a well-known fact that magnetism is more readily established through iron than through air, and therefore when iron is placed in the air gap of a magnetic circuit, the lines of force or flux lines are greatly increased. Then when the iron is withdrawn, the flux lines fall to their original value. We have seen that as changes of flux take place throughout a magnetic circuit an electromotive force is induced in a coil or conductor (armature) surrounding the magnetic circuit. Now, if iron is periodically inserted into and withdrawn from the magnetic field, an alternating current will be set up in the coil surrounding the field. This principle is made use of in the inductor alternator and the field and armature thus remain stationary, contrary to the usual form of an alternator. Now if we arrange to have pieces of iron moved rapidly in the gap between field and armature, changes of flux will take place. Obviously, it is easy in this manner to create many changes of flux within a certain time, and therefore the alternating current set up (E.M.F.) will have a high frequency as seen by the explanation of this phenomena given previously. These moving pieces are so arranged that they practically close the gap between field and armature with only enough clearance to prevent the rotating member striking the stationary members. This is briefly the principle of the Alexanderson alternator. Fig. 4 shows the general plan of an inductor alternator. On the right shows the scheme of winding in slots around the entire form, and as there are necessarily many windings in order to obtain the high frequency, they are generally single turns of wire, each turn having a separate slot. On the left is given a general idea of the arrangement of the rotor, which in this case, as explained previously, is neither field nor armature, but iron teeth arranged on the periphery of the moving disk. The field is increased or excited in the usual manner by means of a source of direct current. In the assembly, the relation of the moving disk to the field and armature is shown. Fig. 5 is a commercial type of Alexanderson Alternator having a frequency (*q.v.*) of 100,000 cycles used for high power radio transmission (*q.v.*).

ALTERNATOR, SINGLE PHASE—(See *Alternator.*)

ALTERNATOR, POLYPHASE—(See *Alternator.*)

AMMETER OR AMPERE METER—An instrument for measuring current in amperes in a circuit. Is connected in series with a circuit. Exists in a variety of forms. The operation of the most commonly used type depends upon the fact that the force a magnet exerts depends upon the number of *ampere turns* (*q.v.*). Therefore, the greater the current sent through its coils, the greater will be its attraction for a balanced armature. Fig. 1 shows a common form of ammeter.

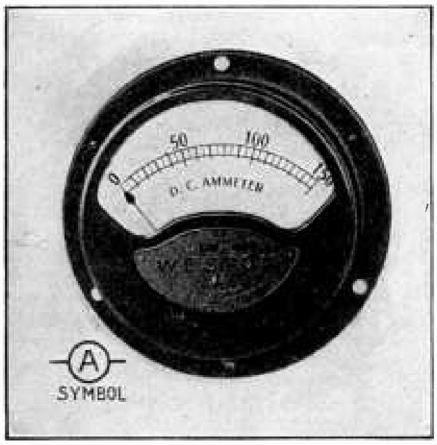

Photo by courtesy of Weston Elec. Inst. Co.

Fig. 1

AMMETER, THERMAL — Commonly known as "*Hotwire Meter.*" Also called "Aerial Ammeter" and "Radio Frequency Ammeter." Operates on the principle of thermal expansion, i.e., the tendency of a wire to expand with heat of current passing through it. Is used to measure current in amperes at extremely high voltages as in radio transmission. Figure 2 shows a Thermal Ammeter which is generally employed in radio work.

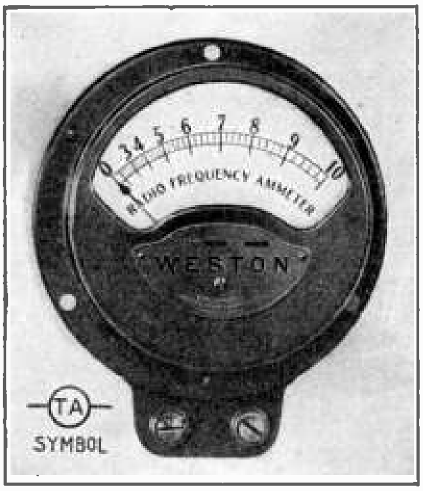

Photo by courtesy of Weston Elec. Inst. Co.

Fig. 2

AMP.—Abbreviation for *Ampere.*

AMPERE—The practical unit of the rate of flow of electricity. It is the unit current, or the rate of flow when one *coulomb* (*q.v.*) flows each second. The value or strength of electric current is almost invariably given in amperes. This unit is used to express the flow of a definite quantity of electricity in a second. If, for example, ten coulombs pass a given point in a circuit in one second, then we say that the current in that circuit is ten amperes. (See *Coulomb, Milli-ampere, Micro-ampere.*)

AMPERE, Andre Marie—A distinguished French physicist, born 1775, died 1836. Noted for his researches in the field of *electro-dynamics.*

AMPERE-HOUR—Commercial unit of quantity. Is that quantity which flows in one hour through a circuit carrying a current of one *ampere.* It is equal to 3,600 coulombs.

AMPERE, INTERNATIONAL—Symbol a—is defined as the unvarying electric current which deposits silver at the rate of 0.00111800 gram per second from a specified solution of nitrate of silver in water.

AMPERE'S RULE FOR DEFLECTION OF NEEDLE—"If one swims with the current and looks at the plus (+) or *north seeking pole* of a magnetic needle it will be deflected to the left, while the negative (—) or *south seeking pole* will be urged to the right."

AMPERE'S THEORY OF MAGNETISM—The theory advanced by *Ampere* that an *electric current continually circulates around each molecule of a magnetic substance and that the process of magnetization is one of arranging these currents in such manner that they all take the same direction.*

AMPERE TURNS—Expressed by the product of number of turns of, and the number of amperes flowing through, the coils of an electro-magnet. Thus, one ampere turn would be one ampere flowing through one turn. The ampere turn is very frequently used as a unit of *magneto-motive force* (*q.v.*) due to the ease with which it can be obtained.

AMPERE VOLT OR VOLT AMPERE—The expression sometimes used to denote *watts* (*q.v.*). For example 1 volt ampere will be one watt, as watts are equal to the product of the volts and amperes in D.C. circuits. (See *Efficiency,* also *Watts.*)

AMPLIFICATION CONSTANT — The factor expressing the maximum voltage *amplification* (*q.v.*) that it is possible to obtain with a given vacuum tube. Thus, if a certain tube will permit amplification of eight times the original voltage, the amplification constant is said to be 8. (See *Vacuum Tube.*)

AMPLIFICATION FACTOR—The ratio of the change of instantaneous voltage, between filament and plate of a *vacuum tube* to a small change of instantaneous voltage between filament and grid for a given constant plate current:

$$y = \frac{\delta V_p}{\delta V_g}$$

The ratio of power (radio signals), voltage or current output of an amplifying device to the power, voltage or current delivered to the input terminals. Generally speaking the degree of increase in amplitude or volume by insertion of an amplifying device in the circuit. (See *Amplification Voltage.*)

AMPLIFICATION, VOLTAGE — In a *vacuum tube* the ratio of the voltage apparent between the terminals of an impedance in series with the *plate,* to the voltage impressed on the *grid.* (See *Amplifier.*)

AMPLIFIER—One of the most important features of the *vacuum tube* (*q.v.*) is its ability under certain conditions to amplify or increase the intensity of electric currents. Such tubes are generally referred to as amplifier tubes and an amplifier is understood to be one or more such tubes with the necessary associated circuits to accomplish amplification.

X—In the accompanying diagram a circuit of a typical one stage *audio frequency* (*q.v.*) amplifier is shown in conjunction with the *detector circuit* (*q.v.*). In this case the *radio frequency signals* (*q.v.*) are impressed at the points A-B from the aerial and ground circuit. The resulting signals at points C-D are audio frequency or voice frequency because they have undergone a change or *rectification* (*q.v.*). Now, if we consider only the solid lines, the circuit will be an ordinary detector circuit and the signals at C-D can be heard by merely connecting the phones at that point in place of the primary of the transformer—P-B. However, by sending these currents through the transformer and impressing them on the second tube it is possible to obtain signals at points E-F of much greater volume and yet having the same general characteristics or sound. The signals or currents will have undergone amplification—the effect of the amplified tube on a local source of power increase the amplitude, but the form or nature of the signals or waves will be the same. (This statement is not entirely true as amplification is generally accomplished only at the expense of some slight changes in the nature of the waves). They will, however, sound about the same in the phones at E-F as they will when the phones are placed at C-D, the volume at E-F being greater due to the amplification. (See *Distortion*). Now it must not be assumed that an amplifier in itself actually has the ability to increase power without some assistance. It is no more possible to obtain something for nothing in electricity than it is in mechanics. What actually takes place is that the amplifier tube and associated circuits control local power—in this case the voltage of "B" battery—in such a manner as to boost or amplify the power of the signals without materially changing their nature. Therefore, an amplifier may be considered as a device which modifies the effect of a local source of power in accordance with the variations of input power, and produces an increased output power or a means of stepping-up or increasing the amplitude or volume of a signal or series of signals. (See *Amplifier Action or Vacuum Tube.*) (See *Amplifier, Audio Frequency, Amplifier Radio Frequency,* also *Transformer.*)

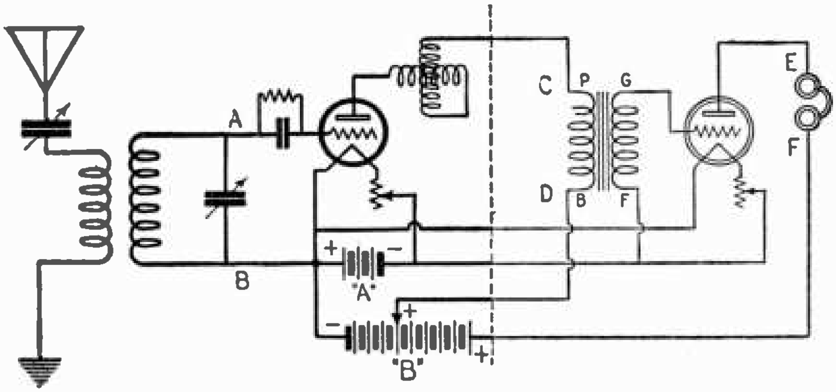

A single stage radio frequency amplifier as used in conjunction with a tuner and detector
Showing the method of amplifying rectified signals at audio frequency

AMPLIFIER ACTION OF VACUUM TUBE—The tendency of a vacuum tube to increase the power or voltage of incoming signals under proper conditions. In the illustration is shown a simple vacuum tube circuit. The incoming signals are impressed on the tube at points A and B. Now it has been noted that small changes of voltage at the *grid* (*q.v.*) will produce much greater changes at the *plate* (*q.v.*). (See *Vacuum Tube, Theory of Operation.*) In other words, the constant changes in value of the incoming signals that are impressed at A-B will produce comparatively larger changes in the current at the plate and therefore larger changes of voltage in the output circuit at C-D. These changes will of course manifest themselves in the form of signals in the phones placed at C-D. Now if these voltages are impressed on a *resistance* (*q.v.*) of suitable value—such as the primary of an amplifying transformer—they can be sent to another tube and again amplified. (See *Amplifier, Audio Frequency.*) In other words the amplifying action of a vacuum tube means the ability of that tube to use input power—the power of the signals delivered to the tube—to control a local source of power such as the "B" battery voltage used on the plate of the tube, and deliver increased power. The signals to be increased or amplified in this manner may be *Radio Frequency* (*q.v.*) or *Audio Frequency* (*q.v.*), each instance requiring a different type of *amplifier* (*q.v.*). The radio frequency signals are amplified by means of a tube and suitable apparatus (transformer or resistance) before they have been changed to audio frequency. The signals that have been *rectified* (*q.v.*) or changed to audio frequency may be amplified by means of a similar arrangement of somewhat different type. The latter would be known as audio frequencv amplification. The increase in power is accomplished without any material change in the nature of the signals or wave form (*q.v.*)

AMPLIFIER, INTERMEDIATE FREQUENCY—An arrangment of a *radio frequency amplifier* to increase the strength of signals at the *intermediate frequency* (*q.v.*) as used in a *Super-Heterodyne receiver.* This is usually a radio frequency amplifier using *transformer coupling* (*q.v.*) the transformers being designed to cover extremely high wave-lengths. (See *Super-Heterodyne,* also *Transformer.*)

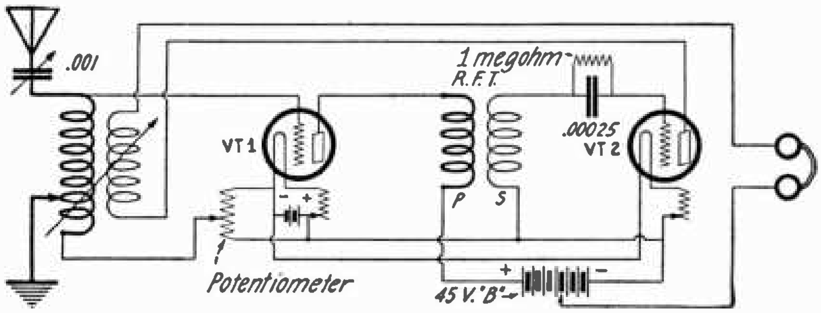

A single stage radio frequency amplifier as used in conjunction with a tuner and detector

AMPLIFIER, RADIO FREQUENCY — An amplifier used to increase the volume of the incoming signals at *radio frequency* (*q.v.*) before they have been *rectified* by the detector. In the illustration is shown a single stage of radio frequency amplification. This is the transformer coupled type. In this case the incoming signals of radio frequency are increased in strength and then passed on to the detector tube VT2. If further amplification is desired, more stages may be used, the number being held within certain limits due to *distortion* (*q.v.*). If more than three stages are used, it is often difficult to control the *feed-back effect* (q.v.) and efficiency is cut down.

AMPLITUDE—The current or voltage on an A.C. or H.F. circuit at any instant.

The greatest vertical distance above or below zero of the crest or trough of a wave is the maximum amplitude of the wave. In the accompanying illustration the numerals at the left (*ordinates q.v.*) may be assumed to be volts, current (amperes), or voice vibrations. A is the zero point; B the maximum *positive,* C the maximum *negative,* and D the length of the wave. Thus, A. B will be the maximum *positive* amplitude, A C the maximum *negative* amplitude, and D the wave length. B is known as the crest or Node—C is the trough or antinode.

X—Now if in a certain train or group of waves, the amplitude of one wave is less than the one immediately preceding it, there is said to be *damping* (*q.v.*), the degree of this damping or diminishing in amplitude being referred to as *decrement* (*q.v.*). Thus, if we assume that in Fig. 1 the current or amplitude of the first wave is 10 amperes, and the next wave has an amplitude of only 5 amperes, then the damping or decrement is 50 per cent, because the amplitude has decreased that much. Similarly, if the first wave amplitude were 10, and the second were 9 amperes, then the decrease

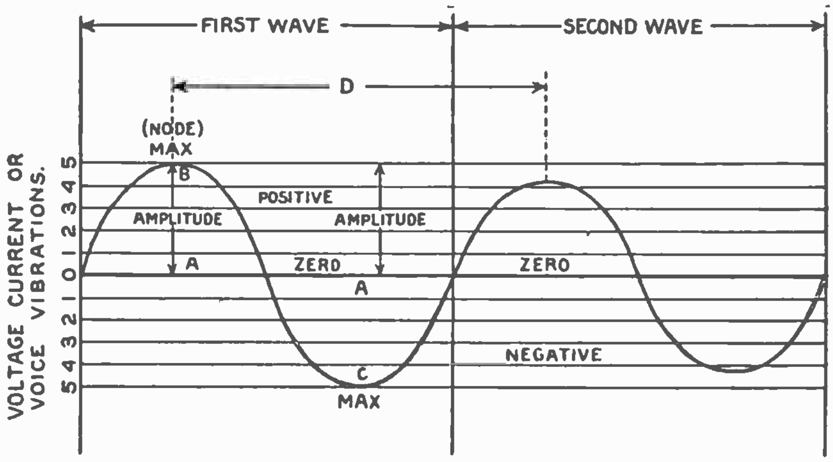

Line wave showing two cycles of an alternating current, indicating the amplitude and damping

Super Regenerative and *Super Heterodyne*. The principle of regeneration is one that has revolutionized the art of radio reception and broadcasting. The vast majority of present day receivers use regeneration in one form or another. Basically, regeneration is the process of returning a certain portion of the unrectified or radio frequency (*q.v.*) current from the plate circuit. of the detector tube to the grid circuit. This has the effect of reducing the resistance of the grid circuit and thus permitting more ready passage of the weak impulses. The result is not only to greatly increase the sensitivity of the detector to weak signals, but actually to step up or amplify the volume of the received signals. In Fig. 1, coils L1 and L2 are the aerial tuning units in conjunction with the variable condenser, C1; C2 and C3 are variable condensers of low capacity; L3 is the plate control for regenerative action; T is the telephone receiver, G, P and F are respectively the grid, plate and filament elements of the tube; R the filament rheostat, and A and B the batteries for filament and plate. In the illustration Fig. 2, C1 is the aerial tuning condenser, L1 the primary coil of a standard vario coupler; V1 the grid variometer and V2 the plate or regeneration variometer. L2 is the secondary of the vario coupler, and the balance of the elements are as in Fig. 1, GL being the customary grid leak and condenser. Regenerative receivers have come into some disfavor owing to their tendency to radiate annoying signals into the antenna, and also the howling within the set when not properly adjusted. While regeneration in some form or other is used in the majority of receivers, it is not widely used now in its original state. (See *Neutrodyne*, *Tuned Radio Frequency*, etc.)

There are many types of regenerative set, some extremely simple and others more complicated. The original Armstrong regenerative circuit is shown in Fig. 1. Fig. 2 shows a more modern application of the regenerative principle and at the same time a more complicated one. (See *Regeneration*, *Super Regeneration* and *Super Heterodyne*.)

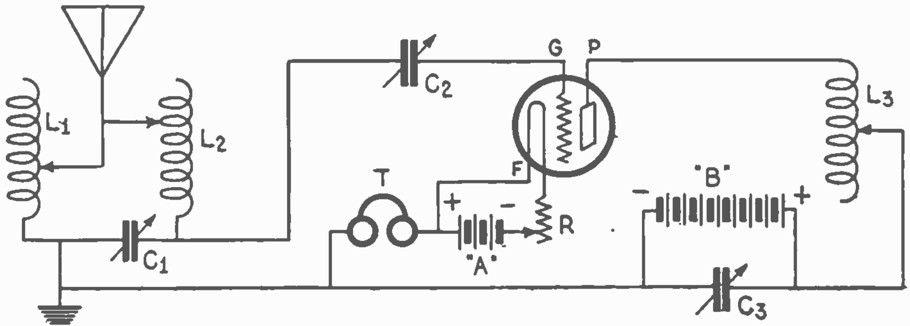

Fig. 1. The Original Armstrong Circuit.

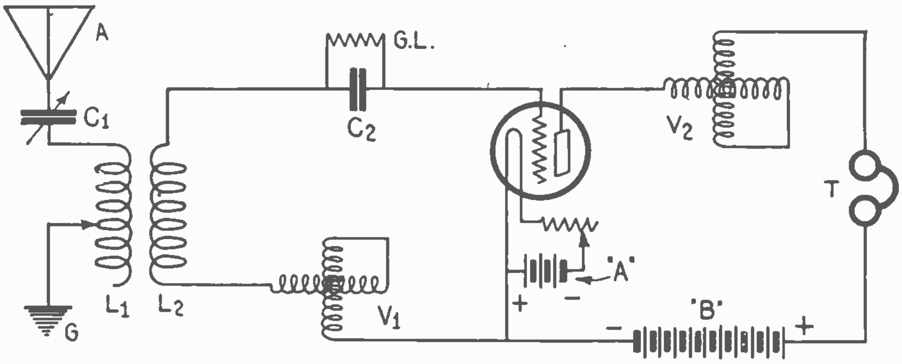

Fig. 2. The Armstrong 3-Circuit Regenerative Receiver.

ARRESTER—A device used for by-passing heavy electrical discharges in the atmosphere in the vicinity of the antenna when they strike the antenna. It ordinarily consists of a small gap between two metallic points connected in series with the antenna and ground in order to protect the apparatus and property against danger. (See *Lightning Arrester*.)

A. R. R. L.—Abbreviation for American Radio Relay League.

ARTIFICIAL ANTENNA—See *Mute Antenna*.

ARTIFICIAL MAGNET—A magnet produced by magnetizing a piece of steel that previously had no magnetic attraction. May be produced by rubbing a *natural magnet* or another artificial magnet across a piece of hard steel in the same direction a number of times. (See *Electro-magnet*.)

ASSUMED DIRECTION OF CURRENT FLOW—The direction which an electric current is assumed to take in its flow. Current is generally considered as leaving the *positive* terminal of its source, flowing through the circuits external to the source, and thence to the *negative* terminal of the source. The direction of flow within the source itself is from the negative to the positive terminal. (See *Theory of Current Flow*.)

ASTATIC COILS—Coils wound in such a manner that, when connected together, they neutralize each other's effect and produce no external *magnetic field*. They are used in the measurement of *inductance*.

ASYMMETRIC CONDUCTOR — One that permits the flow of greater current in one direction than in another. Non-symmetrical as to conductivity.

ASYMMETRICAL EFFECT—The lack of symmetry in the directional effect in a loop or frame aerial due to lack of symmetry in construction. (See *Loop Aerial*.)

ASYNCHRONOUS—A term used in referring to AC (Alternating Current) motors or generators and also to *dischargers* used in radio transmission. As applied to alternating current machines it is signified that the speed of rotation of the machine does not have any definite relation to the *frequency* of the currents produced—thus, out of *synchronism*. (See *Synchronous*.)

An asynchronous radio circuit is one which is not tuned to, or in syntony with, the frequency of the *oscillations* impressed on it. (See *Aperiodic*.)

ASYNCHRONOUS ROTARY DISCHARGE—A spark discharge produced at electrodes of a *rotary discharger* (*q. v.*) for transmitting radio signals, in which the spark rate or number of sparks per second has no relation to the *frequency* of the *alternating current*. (See *Spark Discharge*.)

ASYNCHRONOUS SPARK GAP—See *Spark Gap*.

ATMOSPHERIC ABSORPTION — The portion of the total reduction of radiated power due to atmospheric conductivity, reflection and refraction. The amount of energy loss in radio transmission due to atmospheric factors. This applies only to *electromagnetic waves* (*q.v.*) in *motion in the ether* and should not be confused with the reception losses due to atmospheric disturbances.

ATMOSPHERICS — See *X's*, *Strays*, *Static*.

ATTENUATION—of the electric or magnetic intensity or of the average energy density in electric waves refers to the reduction in strength with increase of distance traversed due to absorption or equivalent losses, as distinguished from the reduction in strength due to geometrical divergence. In free space, the geometrical divergence of waves from a small source involves a diminution of the average energy density in accordance with the inverse square law; plane waves (e. g., along wires) suffer no reduction by divergence, but undergo attenuation due to resistance and other line faults. (See *Attenuation, Geometrical Divergence*).

ATTENUATION, GEOMETRICAL DIVERGENCE—The gradual decrease in strength of electromagnetic waves (radio signals) as the distance which they have traversed increases. The effect on the waves as they travel through the air is essentially the same as in the case of light waves. Now if a beam of light is viewed one mile from its source it will have a certain intensity, but if it is viewed again from a distance of two miles the intensity will be much less. In the case of radio waves, the intensity is said to decrease as the square of the increased distance traveled; that is to say, if they have a certain power at a distance of five miles from the transmitter, the same waves at twice the distance or ten miles will be found to have a strength only one-quarter as great. Similarly, if the waves travel fifteen miles or three times the distance, the signal strength will be only one-ninth that of the same waves at a distance of five miles from the transmitter.

AUDIBILITY—The measure of strength of received radio signals as heard in a telephone receiver connected to the radio receiving apparatus. A radio telegraph signal which is at all audible can be read, i.e., the dots and dashes distinguished from one another; radio broadcast programs heard with at least some degree of clarity.

AUDIBILITY (Radio Telegraph)—A measure of the ratio of the telephone current producing a signal in a telephone receiver to that producing a barely audible signal. (A barely audible signal is one which just permits the differentiation of the dot and dash elements of the letters sent in code.) In the simple shunted telephone method of measuring signals the audibility is defined as:

$$\frac{s+t}{s}$$

where s and t are the *impedances* of shunt and telephone respectively.

AUDIBILITY FACTOR—A measure of signal strength obtained by observing the *resistance* to be placed in shunt with the telephones of a receiving station to reduce the signals to unit audibility. Approximately, the audibility factor is proportional to the power absorbed from the signal waves by the receiving apparatus. Unit audibility is the strength of signal at which dots and dashes can just be discriminated.

AUDIBILITY METER—A device for comparison of the strength of received signals, either from different stations or with different circuits by means of a variable *resistance* shunted (placed) across the phones. This resistance is decreased until the signals are just audible in the phones. Now, if another station is tuned in, or the circuit changed in some manner, the signals may either be entirely inaudible or they may be much louder, this giving an idea of the *comparative* strength without actual measurement. A popular type of Audibility Meter is shown in the above illustration.

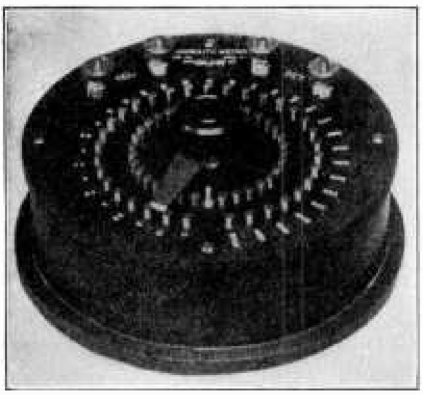

Photo by courtesy of General Radio Co.

An Audibility Meter.

For measurements with an audibility meter, the following shunt formula is used:

$$K = \frac{s+t}{s}$$

Where s is the *impedance* of the *shunt* resistance and t the impedance of the phones, then K is the audibility constant. In general practice a *series* resistance is devised to compensate for the reduction in resistance of the shunt circuit, this keeping the impedance of the plate circuit constant.

AUDIO AMPLIFIER—An *amplifier* used to increase the volume of signals of *audio frequency* (*q. v.*). Audio amplifiers are connected in the circuit of the detector in place of the phones (input) as shown in Fig. 1. This is the standard two stage audio amplifier. The separate stages each consist of a *vacuum tube, a transformer, socket* for the tube and a control *rheostat*. If desired one rheostat may be used for both amplifier tubes as the *filament* current (*q. v.*) can be the same if the tubes are of a similar design. This is known as "transformer coupled amplification." Another much used method is known as "resistance coupled amplification."

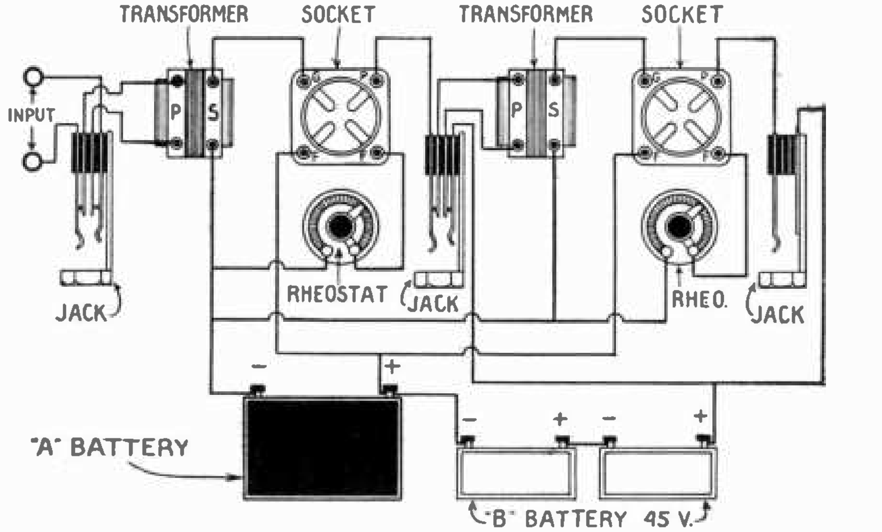

Fig. 1. A standard two stage Audio Amplifier employing phone jacks for detector, first stage, and second stage Audio Amplifiers.

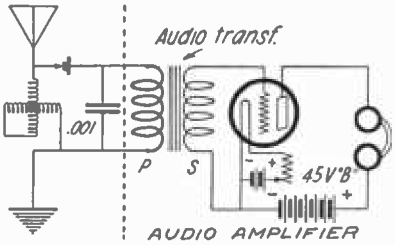

Fig. 2. A one stage Audio Amplifier connected to a crystal receiving set.

Fig. 2 shows one stage of audio frequency amplification connected to a crystal receiving set. There are numerous variations of audio amplification. (See *Push-Pull Amplification*, also *Resistance Coupled Amplifier*, and *Power Amplifier*.)

AUDIO FREQUENCY—A frequency corresponding to normally audible sound waves, i.e., between 40 and 10,000 per second. The limits of audibility are given variously from 16 to 20,000 cycles, but for ordinary usage may be taken as from 40 to 10,000 cycles per second. In audio frequency amplification as used in conjunction with the tuning and detector apparatus of the radio receiving set, the volume of the received sound is amplified or increased over that of the originally rectified sound waves. (See *Radio Frequency*; also *Amplifier*.)

AUDION—The name given by Dr. L. De Forest to a *vacuum tube detector* which possesses a glowing *cathode* and has a steady E. M. F. (*Electromotive force*) permanently applied between *anode* and hot *cathode* and two cold *anodes*. (See *Vacuum Tube*.)

AUSTIN, Louis Winslo, Ph.D.—Born in Ordwell, October 30th, 1867. Graduated at Middlebury College and studied at Clarke University and the Universities of Strasburg and Berlin. Noted for the fact that he is head of the U. S. Naval Telegraphic Laboratory, Washington, D. C., and for his work in measuring high frequency currents.

AUTO-COHERER—A coherer having automatic action to release the particles that cohere when circuit is established. Now obsolete. (See *Coherer*.)

AUTODYNE RECEPTION—See *Self-Heterodyne Reception*.

AUTO INDUCTIVE COUPLING—See *Coupling*.

AUTOMATIC INTERRUPTER—A vibrating electro-mechanical device to make and break an electric circuit by using the energy passing through it. The most common example is that of a *spark coil*. An interrupter may be used for several purposes. In the case of a spark coil it is used to produce sparks for transmission of signals. In another form it may appear in a mechanical battery *charger*, in this instance being used to interrupt the *alternating current* and permit passage of the current in one direction only. A *buzzer* uses a type of interrupter—a vibrating device producing a buzzing sound having a pitch dependent on the rapidity with which the vibrations take place.

AUTOMATIC TRANSMITTER—An apparatus for operating a sending *key* mechanically. (See *Wheatstone Transmitter*.)

AUTOPLEX—A type of radio receiver employing one vacuum tube, operating along the lines of the *Super-regenerative* principle and designed to operate a Loud-speaker. The circuit comprises two standard variometers, a 1500 turn duo-lateral coil, and the usual vacuum

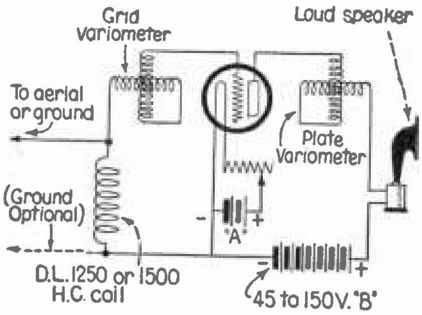

Circuit of the Autoplex radio receiver.

tube, socket, batteries, etc. It is extremely critical in operation, but if carefully assembled will give excellent volume under good conditions. Illustration shows the circuit diagram of the Autoplex receiver.

AUTO-RECEIVER—A device for the automatic reception and recording of radio signals. It is essential in the reception of signals transmitted by an Automatic transmitter (see *Wheatstone Transmitter*) as the speed is too high to permit reception by ear. There are three general classes of automatic receiver: a *resonance intensifier* in conjunction with an inking arrangement; a *phonographic recorder* and a recorder using *photographic means*. In the latter method, a powerful beam of light is projected on to a mirror which is arranged to respond to the signals in approximately the same manner as a mirror *galvanometer*. In the phonographic type, the signals are reproduced on a drum. (See *Resonance Intensifier, Photographic Recorder,* also *Phonographic Recorder.*)

AUTO TRANSFORMER—A device for changing voltage in a circuit carrying alternating currents. A form of *transformer* having but one winding, any part of which may be used as the *primary* and any part as the *secondary*. In Fig. 1 the two windings P and S are continuous, although different sizes of wire may be used according to the purpose for which intended. In this case the voltage taken from the secondary D E will be less than the voltage impressed at A B, because the secondary has fewer turns of wire than the primary. Now in Fig. 2, this conditions is reversed, and as the primary has only a few turns of wire and the secondary has a comparatively large number, the voltage obtained from D E will be greater than that impressed at A B. The voltage ratio is approximately proportional to the ratio of wire turns. Thus, if the primary has 100 turns of wire and a pressure of 100 volts is placed at A B, then the voltage at D E would be within certain limits about equal to the number of turns of the secondary. In other words, if the number of turns on the secondary is one-third the number of turns on the primary, then the voltage at D E will be approximately one-third that impressed at A B. The main advantage of the autotransformer is that a saving is accomplished by having the primary and secondary windings combined. The chief disadvantage is that where extremely high voltages are used there is always danger of the two windings becoming crossed or shorted. (See *Short.*) For this reason they are used mainly for comparatively low voltages. One of the most common forms is the step-down type, known as a "bell-ringing transformer." These are used to cut down the voltage of an electric light line to suitable value for use with small toys or bells. If the transformer is used to increase the voltage, it is known as a "step-up" transformer. (See *Transformer—Radio, Audio.*)

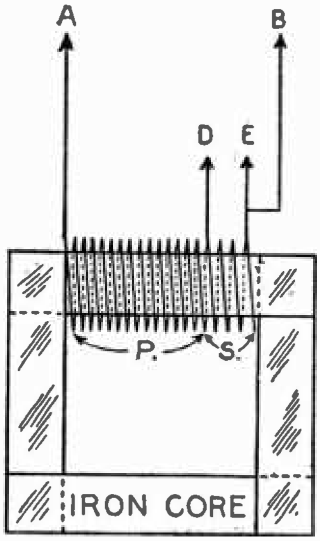

Fig. 1.

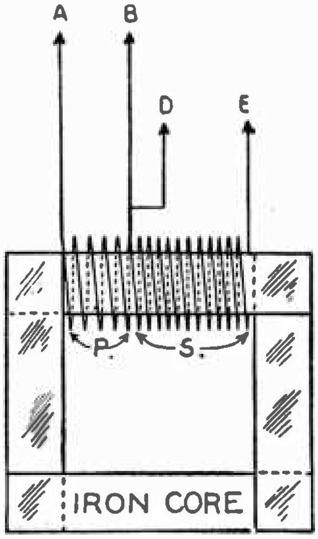

Fig. 2.

AUTO TRANSFORMER COUPLING—See *Coupling.*

B

B—Abbreviation for Baumé, the scale used in a *hydrometer* (*q.v.*)

B. A. AMPERE—The standard ampere fixed by the British Association for the Advancement of Science.

BACK COUPLING—See *Reactance Coils.*

BACK ELECTRO-MOTIVE FORCE—An *Electromotive force* (E.M.F.) which opposes the original electromotive force. The more common term is "*Counter E. M. F.*" In many arrangements of electrical circuits there will be an electromotive force set up in such manner as to oppose the impressed voltage or force. This will be true in the case of a *choke coil* in which any variation in the pressure or electromotive force applied to the coil will result, under proper conditions, in the production of a back or counter-electromotive force that will oppose the change in the original force.

This phenomena is frequently found in electrical circuits; at times introduced for a purpose and at others self-introduced, and therefore demanding consideration in design of apparatus. The rotation of the *armature* of a motor creates a counter E. M. F. opposed to the impressed E. M. F. and upon which the driving power of the motor depends. When an external E. M. F. is impressed on a circuit in which there exists a local E. M. F., the flow of the current through that portion of the circuit containing the local E. M. F. will result in an increase or decrease of energy depending on whether the local E. M. F. is opposed to, or in the same direction as, the external E. M. F. (See *Counter Electromotive force.*)

BACK OSCILLATION—When the *condenser* of a transmitting circuit of the spark type discharges across the *spark gap* some of the *high frequency* current may flow back through the secondary of the transformer instead of taking its normal path across the gap. This back flow is usually referred to as *Back oscillation* or kick back, and is apt to break down (see *Breaking Down of Insulation*) the insulation of the transformer if not prevented or controlled. Back oscillation is retarded ordinarily by means of coils of wire having a certain arbitrary *inductance* value. When the high frequency discharge from the condenser has any tendency to break through, the *reactance* of these coils becomes so great that they act in effect as insulators. As the capacity and inductance of these coils might create an *oscillatory circuit* of the same *natural period* as the primary oscillator, they are shunted with non-inductive carbon resistances. These coils are termed "*choke*" coils.

BAKELITE—An insulating material having widespread use in radio. It possesses great strength and has a high *dielectric* (insulating) value together with the ability to withstand high temperatures. This material is furnished in many forms and under a variety of trade names, being considered one of the best of the insulating compounds.

Chemically known as *Oxybenzylmethylenglycol anhydride.* It is produced by the union under heat and pressure of phenol and formaldehyde together with a small percentage of some alkaline agent. Is heat resisting up to about 500 degrees Fahrenheit—disintegration setting in somewhere above that point.

BALANCED CIRCUIT—An electric circuit arranged in such manner with relation to adjacent or neighboring circuits as to do away with the effect of *mutual induction* (*q.v.*).

BALANCED DETECTORS—A term applied to an arrangement of opposed rectifying *detectors* for reducing the effects of strong *strays* in receiving apparatus. The detectors must be balanced for violent *electro-motive forces,* but not balanced for the moderate E. M. F.'s due to the signals. Actually, a method of eliminating stray impulses in reception.

BALANCED METALLIC CIRCUIT—A circuit through conductors (wire or any metal) in which the total resistance is the same on each side of the circuit. (See *Circuit.*)

BALANCED ROTARY CONDENSER—A variable condenser having some arrangement for keeping the plates accurately balanced in rotation at all points. Fig. 1 shows a form known as the "balanced plate" type. In this case the plates are so arranged that half the movable plates are on one side and half on the other side of the shaft.

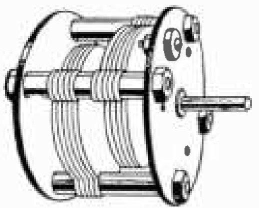

Fig. 1. A "Balanced-Plate" type Rotary Condenser.

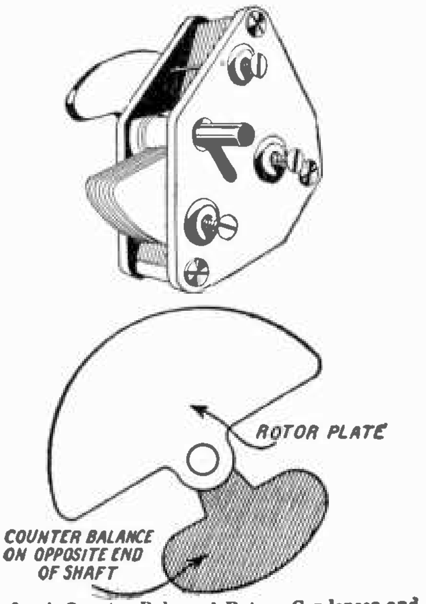

Fig. 2. A Counter-Balanced Rotary Condenser and illustration showing how the counter balance is attached to shaft.

This maintains the balance of the part. Fig. 2 shows another method known as "counter-balanced." The weight of the balance is the same as that of the movable plates and its position on the other side of the shaft serves to maintain balance. The idea is to prevent the plates slipping due to lack of balance and thus keep the setting at the desired point after adjustment.

BALANCING—Any method of eliminating interference by reducing the *amplification* of the undesired signals while permitting full amplification of the desired signals. (See *Interference Eliminator.*)

BALANCING AERIAL—An aerial system designed to permit reception of signals from distant points without interference from a powerful nearby transmitter. The system involves the use of two aerials, one for balancing, the other for receiving. The receiving aerial is tuned to the wavelength of the desired signals, while the balancing aerial is tuned to the wavelength of the nearby station (the interfering signals).

In operation the receiving circuit is coupled to each aerial *electro-magnetically,* and is acted on by both series of *oscillations* (distant and nearby waves). By careful adjustment of *coupling* and phase relation of the *induced currents,* the interfering wave may be balanced out in effect and the desired oscillations received.

BALDWIN RECEIVER—Perhaps the most sensitive type of telephone receiver thus far developed. It employs what is known as a "balanced armature," the signals being produced by the vibrations of a *diaphragm* which is not acted on directly by the magnets. The illustration shows the general plan of a receiver of this type. The *armature* (A) is of soft iron and is pivoted between two U-shaped soft iron pieces (M1, M2) mounted on the ring-shaped horseshoe magnet (M) as indicated. The armature is acted on by the magnets in response to incoming signals and the movement of this element in turn acts on the mica diaphragm (D) by means of a fine brass wire (C). The usual windings (R) are placed between the two pole pieces, the armature being mounted in a central slot. These receivers are much used both for headphones and for loud-speaker units. In the latter case the mica diaphragm is generally made heavier to handle the more powerful vibrations due to the *amplification* of the signals.

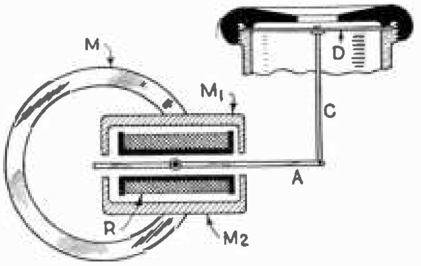

Principle plan of the Baldwin Receiver unit.

When no fluctuating impulses (signals) flow through the windings (R), there is no *magnetic stress* on the armature (A), because this member is suspended centrally between the pieces (M1) and (M2), the magnetic attraction of which are equal. Now when a fluctuating impulse flows through the windings (R) it produces a magnetic flux which combines with the flux of the permanent magnet (M) and the total flux is distributed asymmetrically or unevenly on both sides of the armature. The result is a rocking vibratory movement to this member which is in turn communicated by the juncture wire (C) to the diaphragm (D), producing the audible signals. (For a comparison of this action with the conventional telephone receiver action, see *Telephone Receiver.*)

BALKITE BATTERY CHARGER—An electrolytic battery charger which makes use of a sulphuric acid solution and tantalum as the valve metal. Used for charging *storage batteries* from *alternating current.* See illustration.

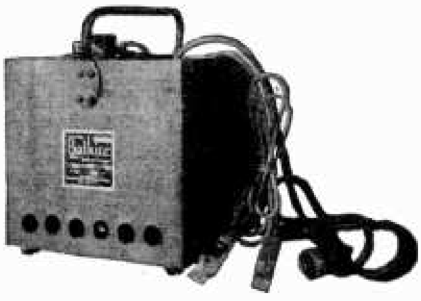

Photo by courtesy of Fansteel Products Co.

A Balkite Battery Charger.

Note: The name Balkite is derived from the name of the man who perfected the process of extraction of tantalum from its ores.

BALLAST TUBE—An automatic means of regulating the *filament current* of vacuum tubes. The device is enclosed in a glass tube and is arranged to permit passage of a certain fixed amount of current. The ballast element is a wire whose *resistance* changes with the current through it, due to the change of resistance with the temperature of the wire. When the *electromotive force* increases beyond normal, the increased current raises the temperature and hence the resistance, so that the current is held to the normal value.

B. and S.—Abbreviation for Brown and Sharpe Gauge (*q.v.*).

BAND OF FREQUENCIES—A continuous range of *frequencies* extending between two frequencies. (See *Band of Wavelengths;* also *Band Pass Filter.*)

BAND PASS FILTER—A *filter* arrangement designed to permit passage of all *frequencies* extending between two definite frequencies but excluding all other frequencies. (See *Band of Frequencies.*)

BAND OF WAVE-LENGTHS—A continuous range of *wave-lengths* extending between two definite wave-lengths. For example, the usual broadcast band of wave-lengths is approximately 250 meters to 550 meters.

BANKED BATTERY—A *battery* that has its *cells* connected in *parallel.*

BANK WINDING—A form of winding for coils in which the turns are staggered so that two or more layers may be used without having large distributed capacity losses. The illustration shows method of bank winding for two layers. Note that in this form the *electrostatic capacity* between any two turns is only that between adjacent turns, representing only a small part of the total winding, whereas on the usual form of winding the adjacent turns would represent a good portion of the total *inductance* and hence the effect of distributed capacity would then be much greater. (See *Distributed Capacity.*)

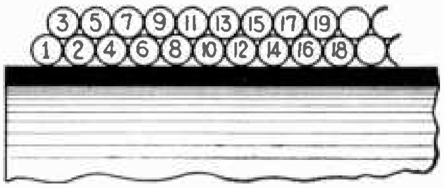

Method of bank winding wire on a coil form.

BARRETER—A receiving instrument consisting essentially of a small mass of conducting material that is heated by the passage of an *oscillatory* current, and arranged so that the consequent alteration of electrical *conductivity* affects an indicating instrument, such as a telephone receiver or *galvanometer.*

BASKET COIL—Variometers—couplers, etc. (See *Basket Wound.*)

BASKET WOUND—A method of winding coils, such as variometers or variocouplers. The winding is a latticework of wires in skeleton form, the object being to save weight and eliminate large losses such as occur in the types using a great deal of solid insulating material. The illustration shows a popular form of basket variometer.

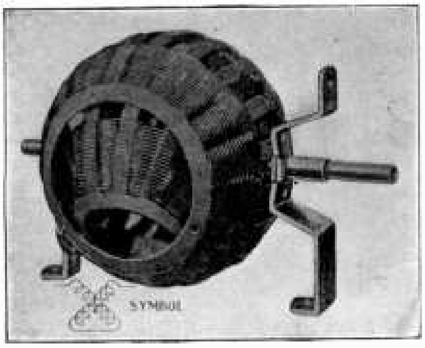

Photo by courtesy of Amer. Radio & Research Corp.

A Basket Wound Variometer.

BATTERY — In the electrical field, a group of *primary or storage cells, dynamos*, or *condensers* grouped together to form a single unit. The term is often used incorrectly to refer to a single cell—as "dry battery," the correct name for which is "dry cell." Thus a group of condensers joined together as one unit is termed a battery, also a number of dynamos arranged as a single unit for the production of electrical energy. (See *Dry Cell*, also *Battery of Alternators, Storage Battery, "A," "B" Battery*, etc.)

"B" BATTERY—An appliance for obtaining electricity of sufficient voltage to supply the plate of a *vacuum tube*. There are two types of "B" batteries, viz., the *dry cell* type and the storage types. In order to obtain the high voltage required for vacuum tube plates, many small cells are connected in series. The usual method is to connect them together and seal in a square package. Often taps are brought out so that various voltages may be obtained for *critical* vacuum tubes. The most popular voltages which dry cell "B" batteries give are 22½ and 45 volts. The storage "B" battery is similar to the ordinary storage battery with the exception that the cells are much smaller, but contain more cells in series. These can be charged in the same manner as the "A" or *storage battery*. In Fig. 1, we have a popular type of "B" battery. This is tapped in order that different voltages may be obtained.

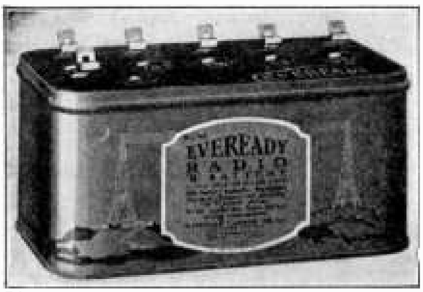
EVEREADY
RADIO

Photo by courtesy of National Carbon Co., Inc.

Fig. 1-A. A popular type of "B" Battery of 22½ volts.

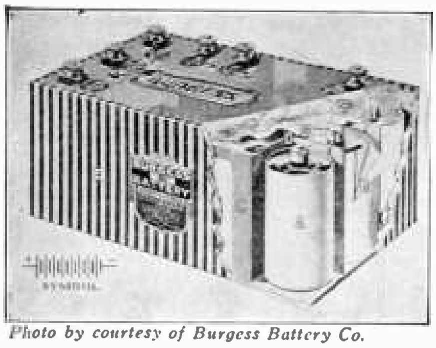

Photo by courtesy of Burgess Battery Co.

Fig. 1-B. A "B" Battery showing the contents of small cells.

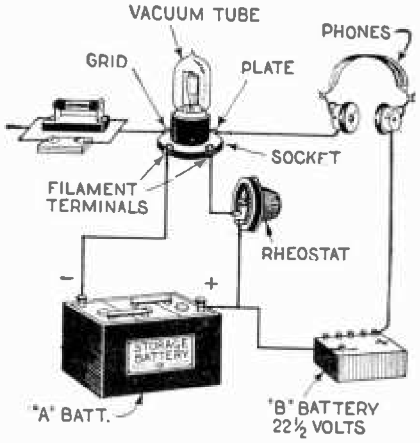
VACUUM TUBE
PHONES
GRID
PLATE
SOCKET
FILAMENT TERMINALS
RHEOSTAT
-
+
STORAGE BATTERY
"A" BATT.
"B" BATTERY 22½ VOLTS

Fig. 2. Manner of connecting a "B" Battery in the circuit of a vacuum tube detector.

The manner of connecting up the "B" batteries in radio receiving is shown in Fig. 2. Here one battery of 22½ volts is used. In Fig. 3, three batteries are connected in series so that 67½ volts are obtained.

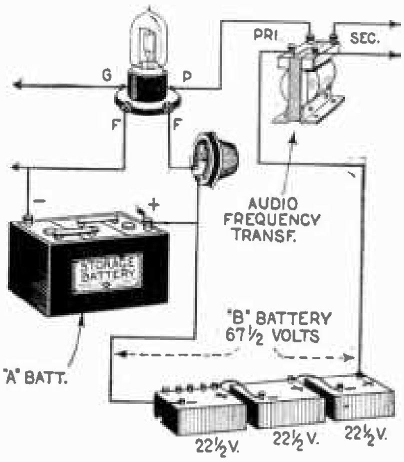
PRI
SEC.
G
P
F
F
-
+
AUDIO FREQUENCY TRANSF.
STORAGE BATTERY
"B" BATTERY 67½ VOLTS
"A" BATT.
22½V.
22½V.
22½V.

Fig. 3. "B" Battery of 67½ volts connected in an amplifier circuit.

A common "B" battery may be used for the detector tube and the amplifier as shown in Fig. 4.

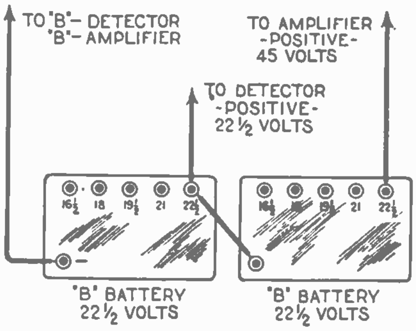
TO "B"– DETECTOR
"B"– AMPLIFIER
TO AMPLIFIER -POSITIVE- 45 VOLTS
TO DETECTOR -POSITIVE- 22½ VOLTS
16½
18
19½
21
22½
–
16½
18
19½
21
22½
"B" BATTERY 22½ VOLTS
"B" BATTERY 22½ VOLTS

Fig. 4. How a common "B" Battery may be used both for detector and amplifier.

If separate batteries for each tube are desired, the circuit should be connected as in Fig. 5. This lengthens the life of the "B" battery as the drain on the battery is much less, i.e., only one vacuum tube is operated on each battery.

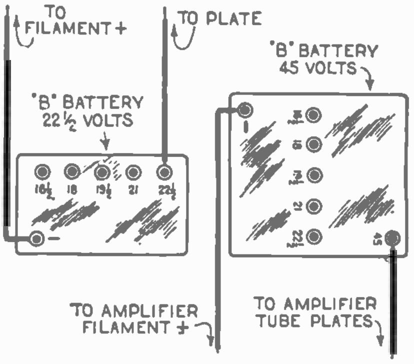
TO FILAMENT +
TO PLATE
"B" BATTERY 45 VOLTS
"B" BATTERY 22½ VOLTS
16½
18
19½
21
22½
–
16½
18
19½
21
22½
45
TO AMPLIFIER FILAMENT +
TO AMPLIFIER TUBE PLATES

Fig. 5. Method of employing separate "B" Batteries for the detector and amplifier.

BATTERY OF ALTERNATORS — In ordinary electrical work such as house lighting, several *alternators* for the production of alternating current may be used together as one source of power. Also in high-power radio transmission, several *high-frequency alternators* may be used together to produce added power. (See *High-Frequency Alternator.*)

BATTERY TESTING INSTRUMENTS —The various devices used for determining the condition of batteries used as radio apparatus. (See *Ammeter, Voltmeter*, also *Hydrometer.*)

BEAT FREQUENCY—The *frequency* or number of vibrations per second of a series of *oscillations* representing the difference between two other series having different frequencies. It is numerically equal to the difference between the original two series of oscillations. Thus, if we have a series of oscillations having a frequency of 100,000 *cycles* per second, and combine it with another series having a frequency of 90,000 or 110,000 cycles, the *beat frequency* will be 10,000 cycles or the difference between the other two frequencies.

X. The frequency of recurrence per second of either maxima of addition or minima of opposition of two superposed periodic phenomena having the same nature but of different frequencies. The time measure of vibration of a beat or heterodyne. (See *Heterodyne*, also *Beats.*)

BEATS—When two sets of *oscillations* of different *frequencies* occur in the same system, the difference in the frequency or rapidity of vibration causes a new set of oscillations having a frequency equal to the difference between the original two sets. This may be accidental as in the case of certain types of receivers such as the *regenerative* type, in which case the beat oscillations are undesired and cause interference or disturbances in the set. Then again a beat frequency may be purposely introduced as in the case of the *Super-Heterodyne receiver*. Here we have the natural incoming oscillations from the antenna combined with a series of oscillations produced locally in the system. The result is a new set of oscillations where the incoming series and the locally produced series are not of the same frequency. (For more complete explanation see *Super-Heterodyne Receiver.*)

X. Where two *periodic phenomena* are *super-posed* or run together and the frequencies differ, the gradual change in *phase difference* produces a condition wherein the *amplitudes* are in opposition at one instant and in concurrence at a later instant with the various intermediate stages during the interval. (See *Heterodyne*, also *Beat Frequency*.)

BEG OHM—A *resistance* of one billion *ohms*. (See *Megohm*.)

BELIN, EDOUARD—French scientist, the inventor of what is considered the first practical system of radio transmission of photographs. The device can be used for sending photos by cable or telephone lines as well as radio. (See *Transmission of Photographs by Radio*.)

Brown photo.

Edouard Belin.

BELL, Alexander Graham (1847-1922)—Scottish scientist. Born in Edinburgh, March 3rd, 1847, he was educated at the High School and University and graduated as a doctor of medicine. In 1870 he went to Canada, and in 1872 became professor of vocal physiology in the University of Boston. In 1876 he exhibited his apparatus for the transmission of sound, afterwards developed into the telephone. Bell was experimenting with an electric invention by means of which he hoped to make speech visible to the deaf. A delicate metal reed was caused to vibrate by spoken speech and to transmit an electric current to the opposite end of a wire, where the vibrations of the first reed were reproduced in a second reed by a magnet. He found that it was possible to transmit not merely vibrations of the original reed, but to reproduce the sound itself in the vibrations of the second reed. In 1878 he invented the photophone, to enable sound to be transmitted by variations on a beam of light, and later a phonograph. Bell was the author of many scientific papers, was awarded the Albert Medal of the Royal Society of Arts in 1902 and the Hughes Medal of the Royal Society in 1918. He died Aug. 2nd, 1922.

Alexander Graham Bell.

BELLINI, Dr. Ettore—Born in Foligno, Italy, April 13th, 1876. Educated at the University of Naples. Noted as Electrical Engineer to the Royal Italian Navy, and chief of the Naval Electrical Laboratory at Venice. Joint inventor with Captain Tosi of the "Radiogoniometer" (*q. v.*), a device for finding the direction of transmitted radio signals.

BELLINI-TOSI AERIAL — See *Goniometer*.

BELLINI-TOSI DIRECTION FINDER —See *Goniometer*.

BEZEL — In mechanics, generally a groove and flange made to receive a beveled edge. In its adaptation to radio some liberty was taken with its real meaning. Its significance in this sense is a small metal ring having a wire mesh or glass center. It is fitted into a circular hole in the panel of a radio receiver and used as a peep hole to enable the operator to know at all times the conditions of the tubes, whether or not they are properly lighted.

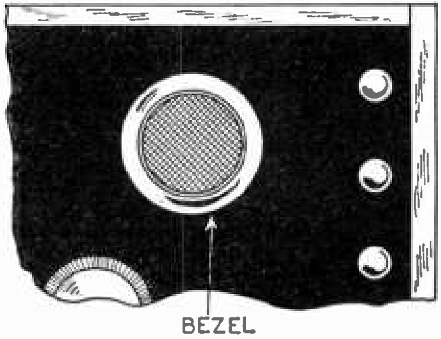

A Bezel on the panel of a radio receiving set.

BILLI CONDENSER—A variable *condenser* of low *capacity*, consisting of two brass tubes, one of which is arranged to slide in and out of the other for the purpose of varying the capacity. Such condensers were much used before the days of broadcasting, in conjunction with any circuit using a crystal and battery for detector. This type of condenser has long been obsolete in the United States.

BINDING POST—A screw and nut arrangement used on electrical units and radio apparatus to make convenient connections from an external source to the desired point within the apparatus. Binding posts are provided, for instance, to make handy connections from batteries to the set and from the aerial and ground leads. The illustration shows a popular type of binding post.

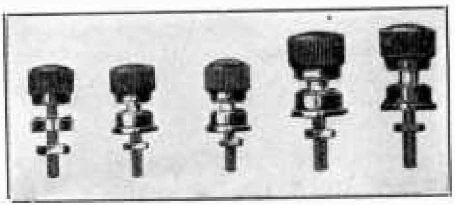

Photo by courtesy of The H. H. Eby Mfg. Co.

A group of Binding Posts as used on electrical and radio apparatus.

BI-POLAR—Having two *poles*. Usually a *dynamo* or motor whose *armature* rotates between a *field magnet* having only two poles. The most modern types are of the Multi-polar variety. (See *Bi-polar Magnetic Field*.)

BI-POLAR ARMATURE WINDING—An *armature* wound in a manner to permit its use with a dynamo having a bi-polar magnetic field.

BI-POLAR MAGNETIC FIELD — The magnetic field (*q.v.*) created between two magnetic poles. The illustration shows a common form of bi-polar magnet for a dynamo. Bi-polar magnetic fields are now used as a rule only for direct current machines of low power, such as those under 5 *Kilowatts*. (See *Dynamo*, also *Generator*.)

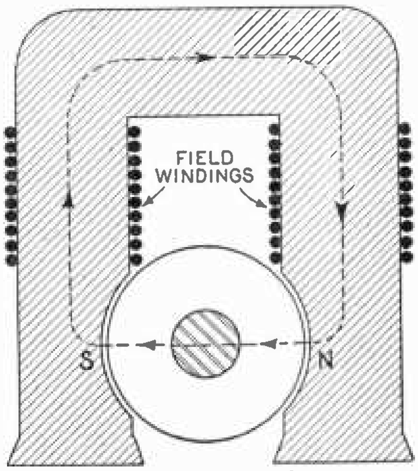

A Bi-Polar Magnet of a dynamo.

BI-TELEPHONE RECEIVER—A headset with two receivers or phones as generally used in radio. In the early days of radio transmission it was the custom to use only one phone.

BLONDEL, Andree—French electrical expert. Born at Chaumont, France, in 1863, he graduated at Paris University, and studied electric waves, on which subject he early contributed a number of papers to various scientific journals. In 1893 he invented the *oscillograph*, an instrument somewhat similar to a mirror *galvanometer*, for showing curves of *oscillating* or *alternating currents*. This invention opened up a fresh field in the study of alternating currents. In the same year Blondel explained for the first time mathematically the effect of inertia in the shunting of *alternators*. He is responsible for a system of acoustically syntonic wireless telegraphy, and for directed waves produced by a double aerial. In 1902 he patented a method for producing electric oscillations for wireless telephony, and has written many papers on microphonic control for transmitters, wireless telephony, the singing arc, etc.

BLONDLOT, Professor Prosper Rene—French wireless expert. Born at Nancy, France, in 1849, studied at

Paris, and became professor at the faculty of sciences, Nancy, and afterwards Honorary Professor and Correspondent of the Institute of France. Professor Blondlot is famous for his studies of electro-magnetic waves, particularly with regard to their speed, and the laws of *propagation* of wireless waves in various media.

BLOWER MOTOR—A motordriven fan used to deliver a high pressure blast at the *spark gap* to prevent *arcing*. (See *Spark Gap*, also *Spark Discharger*.)

BLOWOUT—See *Magnetic Blowout*.

BLUE GLOW—A condition within a *vacuum tube* when the vacuum has become poor. After continued use a tube will often hold a small percentage of gas, which causes a blue glow when current is passed through it. A tube in this condition is a poor *detector* or *amplifier* and should be replaced. The condition should not be confused with the glow often caused by an excessive voltage. In this latter case decreasing the voltage to the normal value will generally permit the tube to be operated efficiently. The blue glow in this case is due to *ionization* of the residual gas by the excessively high *potential* impressed across the elements of the tube.

BOARD OF TRADE UNIT—B.O.T. 1,000 Watt Hours. One and a third Horse Power.

BODY CAPACITY—The effect of the human body when tuning a radio receiving set. The hand when placed on or near the controls very often throws the receiver out of balance with the incoming signals. In the case of very *sharp tuning* this effect is more in evidence and is likely to cause howls due to *self-oscillation*. The howling is due to the production of an audible *beat frequency* in the system, caused by the combination of the local *oscillations* with the incoming signal oscillations. The remedy may be to shield the panel with metal foil or in some cases to merely alter the direction of the leads from the *tuning condenser*. The experienced operator seldom pays much attention to the phenomena, as a little practice enables the listener-in to compensate for the effect. To make this more clear, if, when tuning the set, the withdrawal of the hand from the dial *detunes* the set or throws it out of balance, it is due to the fact that the hand has acted as a certain amount of *capacity* in the circuit, and naturally its withdrawal is equivalent to a change in the condenser setting. A simple system is to turn the control beyond the point of maximum volume, then when the hand is withdrawn the capacity will drop to its proper value. This can be mastered very readily. If the tuning condenser is connected across the secondary, i.e., one side to the grid, and the other to the grid return, the stator plates of the condenser should be connected to the grid of the tube. If metal foil or thin copper sheeting is used for shielding it should be connected to the ground post of the set.

The tendency of the human body to insert an arbitrary capacity in the circuit. The phenomena is especially noted where maximum volume is obtained only by very critical control close to the oscillating point, in which case the circuit may be thrown out of resonance.

BOLITHO CIRCUIT—A *super-regenerative circuit* patented in England in 1919 by Captain J. B. Bolitho. The circuit was originally intended to operate a relay device. The circuit has been adapted to use as an *amplifier* for radio reception.

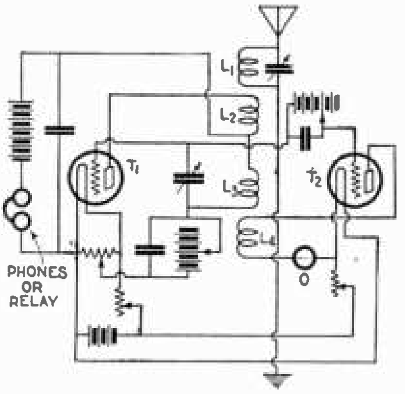

Diagram of the Bolitho Circuit.

In the illustration the phones may be replaced by a loud speaker or relay. In operation, T1 is held at a point just below that where *oscillation* sets in by means of T2 which is excited by an oscillator O in the plate circuit and coupled by means of coil L4 to the tuned grid circuit through L2 and L3, the coupling, as explained, being so arranged that the tube is always just below the oscillating point. A *reaction coil* L4 is placed in the plate circuit of T2 in such manner as to oppose the magnetic linkage between coils L2 and L3. The *frequency* of the generator O is lower than that of the received signals. As indicated, the grids of both tubes T1 and T2 are joined together and the tube T2 maintains the circuit in a receptive condition when tube T1 is tuned just below the oscillating point.

The oscillator O makes the *plate* of T2 alternately *positive* and *negative* with the following effect: When the plate of T2 is negatively charged there is no current in coil L4, this coil therefore not affecting coils L2 and L3, and permitting tube T1 to build up self oscillation. When the plate is made positive by the action of the oscillator, current flows through coil L4, and as it is coupled in opposition to coils L2 and L3 it neutralizes the coupling between L2 and L3 and prevents transfer of energy between the grid and plate circuits of the tube T1. This tends to stop the self oscillation and makes the circuit receptive to the energy (signal) produced by outside signals in the aerial coil L1. The frequency, of course, is determined by the frequency of the oscillator O. The system is especially adapted to the reception of continuous wave signals. (See *Super-regenerative circuit*, also *Feed-back* and *Regenerative circuits*.)

BOLOMETER—Type of *Wheatstone Bridge* having an easily heated *resistance*, such as a very fine wire in one arm. (See *Wheatstone Bridge*.)

BOOSTER—An expression signifying a small *dynamo* used in conjunction with main dynamo to temporarily raise, when necessary, its normal *pressure*. It is generally driven by a motor supplied with energy from the main generator and thus becomes in effect a *continuous current transformer*. Frequently used for charging *accumulators* of a generating plant. The term "Boost" is also used to denote increase or a stepping-up of any electrical quantity. (See *Amplification*.)

BORNITE—A *crystal rectifier* much used in radio reception. It is a natural sulphide of iron and copper, having a metallic blue lustre. This mineral is used in combination with zincite or copper pyrites as a crystal detector. Such combinations are generally known as "crystal to crystal" detectors to distinguish them from the ordinary variety using one mineral and a wire contact.

BORON—A non-metallic chemical element used in radio as one of the electrodes for the *T. Y. K. Arc*. Chemical symbol B, atomic weight 11.0, specific gravity 2.6.

B. O. T.—The customary abbreviation for *Board of Trade Unit*. The unit represents 1,000 *watt hours*. (See *Board of Trade Unit*.)

BOX AERIAL—A term occasionally applied to loop aerials. (See *Loop aerial*, also *Frame Aerial*.)

BRADFIELD INSULATOR—A particular form of *Lead-in Insulator*, consisting of an ebonite tube provided with zinc cone and ebonite spark discs, for breaking up continuous streams of rain running down outside which might cause the aerial to become grounded. The whole is held in position, half way through roof of operating room, by means of a stuffing box. The aerial is led in by means of a conducting rod through center of tube. (See *Petticoat insulator*.)

BRANLY COHERER—Early form of *Marconi Coherer*. (See *Coherer*.)

BRANLY, EDOUARD—French radio expert. Born at Amiens, France, Oct. 23rd, 1844. He was educated at Paris and afterwards became Fellow of the University, doctor of physical science, and doctor of medicine. Branley early made a study of electro-magnetic waves, and in 1890 and 1891 patented methods of operating a local relay circuit from a distance by means of wireless waves. In 1900 he was awarded the Grand Prix by the International Jury of Superior Precept Instruction for his exhibition of radio-conductors.

In 1890 Branly published an account of his very extensive series of observations on the electrical conductivity of loosely packed metal filings, and he made the extremely important observation that an electric spark at a dis-

Edouard Branly.

tance had the power of suddenly changing the electric *conductivity* of loose masses of powdered *conductors*. To Branley is due the *coherer* named after him.

BRAUN, FERDINAND—Professor at the University of Strasburg, and one of the leading world authorities on Radio transmission. As early as 1899 Braun was granted a patent for *closed oscillating* systems with an *inductively coupled* antenna. The system was claimed by Braun to possess a much greater efficiency than the directly coupled systems. The Braun transmitting set, as manufactured by Siemens and Halske, consisted of a large coil worked into an *electrolytic interrupter*, a set of *Leyden jars*, enclosed *spark gap*, *oscillation transformer* wound with insulated wire placed in oil, and for the receiving set the standard type of *coherer relay* and Morse register. In 1899 Braun established communication between Cuxhaven and Heligoland, using aerial wires 90 feet high, and the inductive coupled aerial connection for transmitting. In 1903 Braun joined with Slaby, von Arco, and Siemens to form the Telefunken system of transmission. Professor Braun was awarded the Nobel prize with Marconi in 1900, for his work in wireless. Braun has devised a method of directional wireless which depends upon the interference of *electric waves* travelling in the same direction but different in *phase*. Three simple vertical wire aerials are set up in positions corresponding to the angular points of an equilateral triangle, and *oscillations* are created in these which differ from one another in phase. In 1897 Professon Braun published a description of his cathode ray tube, and afterwards pointed out, in 1902, how such a tube could be used to trace the forms of alternating current waves.

BRITISH STANDARD WIRE GAUGE—The standard wire gauge of Great Britain. The table shows the various diameters in thousandths of an inch (mils.).

WIRE GAUGES IN MILS.

GAUGE	New British Standard
0 000 000	500
000 000	464
00 000	432
0 000	400
000	372
00	378
0	324
1	300
2	276
3	252
4	232
5	212
6	192
7	176
8	160
9	144
10	128
11	116
12	104
13	92
14	80
15	72
16	64
17	56
18	48
19	40
20	36
21	32
22	28
23	24
24	22
25	20
26	18
27	16.4
28	14.8
29	13.6
30	12.4
31	11.6
32	10.8
33	10.0
34	9.2
35	8.4
36	7.6
37	6.8
38	6.0
39	5.2
40	4.8

BREAK or BREAKER—See *Circuit Breaker.*

BREAKDOWN POTENTIAL or BREAKDOWN VOLTAGE—The voltage necessary to *break down* or puncture a *dielectric*. The ability of any *insulating* material to resist breakdown is an important matter. An insulator or dielectric material is usually given a certain rating or point above which it is likely to fracture. Thus, a condenser may have a breakdown rating of 500 volts. That is to say that voltages up to 500 and slightly over may be handled with safety, but any large increase above that figure may result in puncture of the material by the excessive potential or voltage. The relation between dielectric strength and breakdown voltage is shown in the following formula:

$$V = cd\ \tfrac{2}{3}$$

Where d is the thickness of dielectric in millimeters
V the potential difference in volts
c-a constant representing the potential difference required to cause a dielectric breakdown of a sample of the material 1 mm. thick. (See *Dielectric Strength.*)

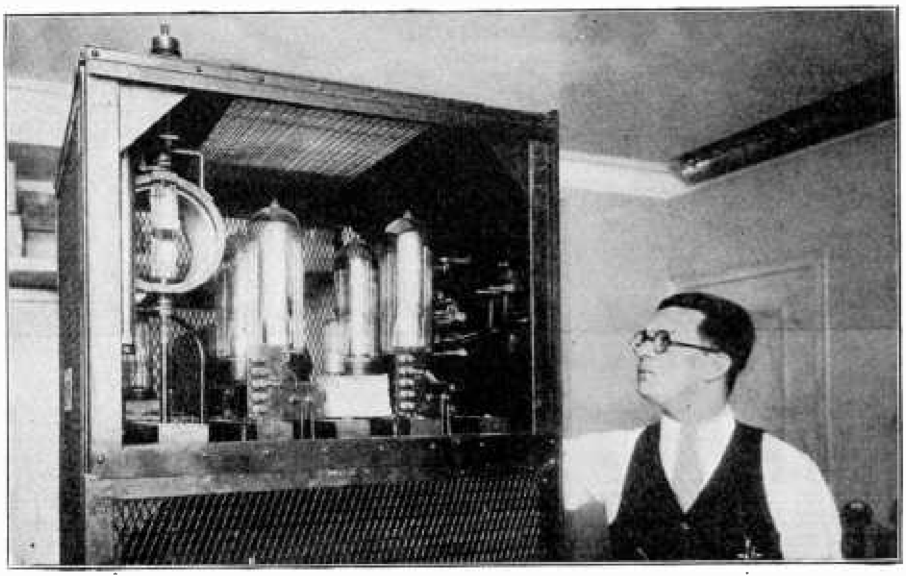

Fig. 1. A bank of 250-watt transmitting tubes in a Radio Broadcast Station. Note the small receiving tube for comparison of size.

BREAKING DOWN OF DIELECTRIC—In a *condenser*, the gradual weakening of the *dielectric* or insulating sheets which eventually causes them to permit direct passage of the current. For example, if a force of 1000 volts is applied for a considerable period to a condenser that has been designed for a maximum of 500 volts, the insulating material between the plates would probably be punctured by the current. (See *Break-down Potential.*)

BREAKING DOWN OF INSULATION—The same effect as referred to under *Breaking Down of Dielectric*. The insulation weakens and permits the ready passage of electric currents. (See *Break-down Potential.*)

BRIDGES—See *Capacity Bridge*, also *Wheatstone Bridge.*

BRIDGING—A term occasionally used for *shunt* (q.v.) connections. Thus, a grid condenser is generally shunted or bridged by a *grid leak.*

BROADCAST—Transmission of music, news and other matters of general interest and entertainment by means of *radio telephony*. (See *Broadcasting, General Treatise on Methods.*)

BROADCASTING, GENERAL TREATISE ON METHODS—The application of radio telephony to the transmission through the ether of musical programs, speech or any form of news or entertainment, in such manner that it can be readily received by anyone possessing a radio receiver sufficiently sensitive and capable of being tuned to resonance with the particular waves being sent out. While broadcasting in its present world-wide scope is a comparatively new enterprise, it has actually been used to a more or less degree since the practical development of the wireless telephone. As near as records can be determined, the first actual broadcasting was done in 1912. It was, however, not until after the world war that it began to develop into an industry of prominence. Now it occupies a place of very definite importance in everday life; carrying news and entertainment into the remote corners of the earth, far removed from telephone or telegraph communication.

The fundamental principles of broadcasting are essentially those of radio transmission, with the addition of means of combining the speech or music vibrations with the regular radio waves. First we have to consider the radio waves of high-frequency—alternating current that changes in direction many thousands of times each second. These high frequency waves are known as the carrier waves. The frequency of these waves varies according to the length of the waves (wavelength), in the case of American broadcasting stations, being between 250 and 550 meters in most cases, which means a frequency range of approximately 550,000 to 1,200,000 cycles.

Now by means of a microphone, a device similar in principle to the mouthpiece of an ordinary telephone, but designed to handle large currents, the air pressure waves, i.e., speech or music, are converted into variations of electric current. By use of suitable amplifying apparatus, the strength of the electric waves radiated from the antenna is varied to correspond electrically to the acoustic or sound variations due to the speech or music being impressed on the microphone. The speech or music waves are low frequency waves, generally ranging between 100 and 4000 per second. Thus, while the carrier wave oscillations are of high frequency, the variations in it are low frequency, corresponding to the ordinary speech or music vibrations.

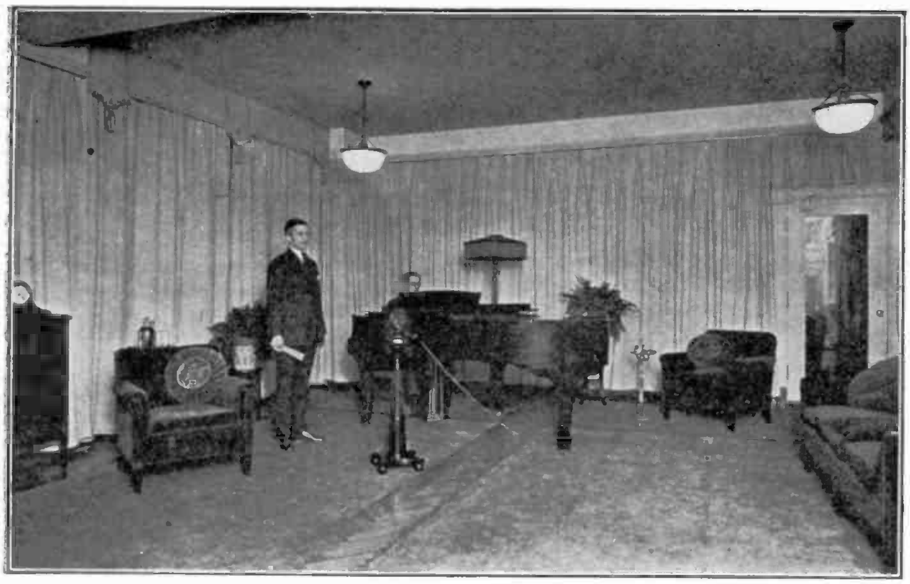

Fig. 2. Studio of a modern Radio Broadcast Station.

The process of varying the radiated wave in accordance with the sound waves is known as *modulation*. This adjustment is of the utmost importance, as defects here result in distorted or improperly modulated sound waves and the reproduction cannot be perfect. The reception of broadcast programs is essentially the same as that of ordinary radio signals, with the exception that care must be used at the receiving end to preserve the form of the original vibrations. When dot and dash signals are received, the tone is of comparatively little importance, whereas with voice or music it is essential that the tone quality be retained.

In broadcasting the high frequency oscillations are produced by means of a tube or group of tubes, practically identical with those used for reception, but designed to handle large currents and voltages. The majority of American stations use vacuum tubes of 250 watts power rating, one or more of them being used to obtain the total power required. These tubes are shown in the illustration, Fig. 1.

When a program is being broadcast, the action is somewhat as follows: The artists are disposed before the microphone with due regard to the effect of certain instruments. That is to say, one type of instrument may be placed near the microphone, while another must be placed farther away, owing to the differences in tone and volume. This is done to prevent any one tone or range of tones from predominating. These speech or music currents are converted into electrical variations by the microphone and then increased many times in power or amplitude by means of a speech amplifier system, comprising vacuum tubes and the necessary associated circuits.

These amplified currents are applied to the grid of a control tube or bank of tubes and result in large variations in the plate current of the tubes. (See *Theory of Vacuum Tube.*) The control tubes are supplied with filament current from storage batteries or special low voltage D. C. generators, and the high voltage applied to the plates of the tubes is furnished by generators. This power will vary in most cases from about 1,000 to 5,000 volts. In some broadcasting stations, the power is obtained through high voltage direct current generators, while in others, the alternating current supply is stepped up or transformed through high voltage transformers and then rectified by means of special vacuum tubes.

The high frequency waves are applied to the antenna system tuned circuits and carefully modulated in accordance with the variations of the speech or music being transmitted. The action of the transmitter and the effect of the modulation must be carefully observed at all times during the transmission of programs. (See *Harmonic Suppressor*, also *Oscillograph.*) Fig. 2 shows the studio of a modern broadcasting station. It is interesting to note that these studios are very carefully arranged and the walls cushioned to prevent troublesome acoustic effects. There must be no reverberation—echoes—or the transmitted program would be distorted by the additional or repeated tones. Fig. 3 shows the main control panels. Fig. 4 illustrates the power generating system of a typical broadcasting station.

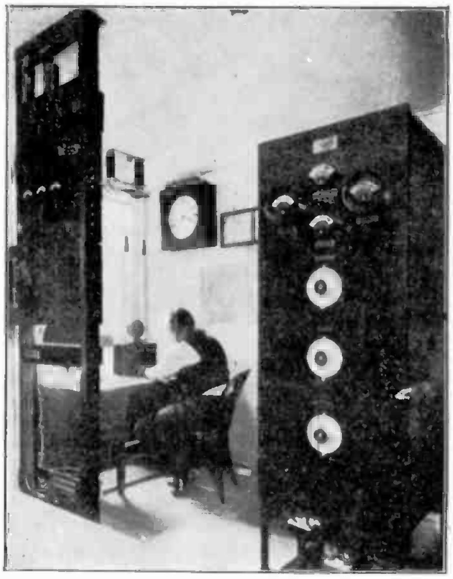

Fig. 3. Main control panels of a Radio Broadcast Station.

The radio waves generated and radiated from the broadcasting station travel away from the antenna at tremendous speed—that of light waves, or about 186,000 miles per second. These waves radiate off into space in

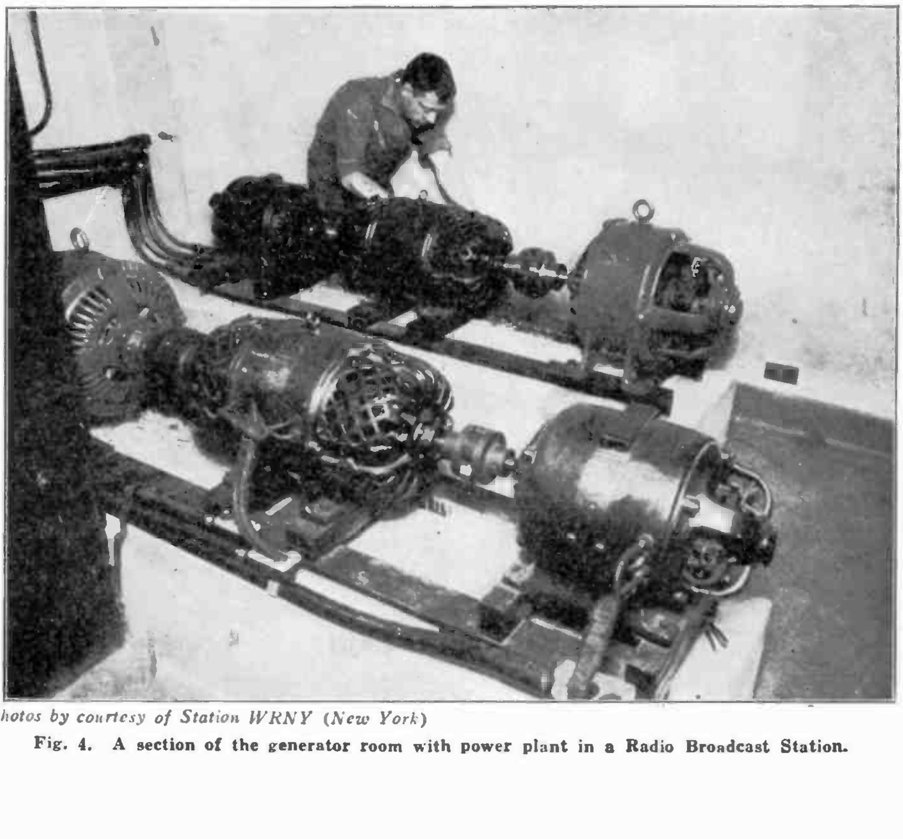

Photos by courtesy of Station WRNY (New York)

Fig. 4. A section of the generator room with power plant in a Radio Broadcast Station.

much the same manner as light waves, but a certain portion apparently follow the curvature of the earth. When these waves impinge on a receiving aerial they follow the path of least resistance, and being then in the nature of electrical impulses, they follow the aerial and lead-in wire to the receiving set, and thence to the ground. The receiver is arranged in such fashion that it corresponds electrically to the antenna circuit of the transmitting station. That is to say, the receiver is placed in resonance with the particular signals desired. The usual system of detecting, either with a simple crystal or by means of vacuum tubes, may be followed by special amplification which will make the signals audible on a loud speaker. In many cases the signals are amplified in their original form (radio frequency amplification), then rectified by the detector and further increased in volume at audio frequency. (See *Modulation*, *Speech Amplifier*, *Electromagnetic Waves, etc.*)

BROAD TUNING—A term used to designate the lack of *selectivity* in reception or in transmission, the use of several *waves* rather than a sharp, pure wave. If a receiving set is of such a nature that it is difficult or impossible to receive the desired signals without hearing any other, the set is said to be "broadly tuned." (See *Tuning*, also *Selective Tuning.*)

BROKEN CIRCUIT—See *Open Circuit.*

BROWN AMPLIFYING RELAY—An arrangement whereby the comparatively weak signals received by a *crystal detector*, usually of the *carborundum type*, may be *amplified* or increased in intensity, either to allow a recording machine to be used or merely to make the signals more readily audible in the ear phones. The device consists essentially of an arrangement for controlling and stepping up the vibrations in conjunction with a small *battery* or *dry cell.*

One example of such a relay is shown in the illustration. In this case the soft iron *cores* cc are *magnetized* by the *permanent magnets* mm, the poles of which are marked n and s. R1 is a fine wire winding such as is ordinarily used in ear phones and R2 is a larger winding placed over the main magnets mm. Directly above the soft core cc with the windings R1 is placed a steel vibrating tongue V having a contact D which touches lightly contact G. Contact G is of a special alloy of osmium and iridium and the contact D is a carbon button, which in contact with G forms a *microphone.* The winding R2 is joined with the telephone P and the battery B, all in series with the microphonic contact. When current (signal current) enters the leads A—E through the windings R1 the result is a change of *flux* in the core of the magnet which vibrates the tongue V. This in turn varies the current of the battery B and causes correspondingly greater sounds in the ear phones P. Several of these relays may be connected together to permit a much greater *amplification* of the original signals.

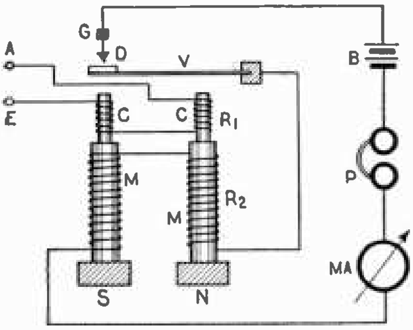

The Brown Amplifying Relay Circuit.

BROWN & SHARPE GAUGE—The American wire gauge adopted as standard for wires for electrical purposes. Commonly known as B. & S. The various sizes run from 40 to 0000, 40 having a diameter of .00314 inch, and 0000 being .46 inches in diameter. We refer to wire having a diameter of .00314 inch as being 40 B. & S. gauge. (*See page 26 for table.*)

BROWN, Sidney George—Born in Chicago, U. S. A., in 1873, of English parents, he was educated at Harrogate and London University. An early study was made by him of the subject of submarine telegraphy. During these investigations he invented the magnifying relay for cables. Another important invention was the cable drum relay. In the year 1898 he invented the magnetic shunt. Since that time many of his activities have been directed to the solution of problems in telephone and wireless telegraphy.

In particular the radio experimenter will know the telephone receivers bearing his name and the important developments of the *microphone relay.* In the field of land telegraphy and telephony, his activities have resulted in the invention of such items as the carbon telephone relay system, which is largely used on land trunk lines for the transmission and reception of telephony.

He is a Fellow of the Royal Society and Vice-President of the Radio Society of Great Britain. His writings on technical subjects are extensive, and numerous valuable patents have been taken out by him, including the vacuum tube *oscillation generator* in 1916 and ionic electric relays in 1918.

BRUSH—A device for collecting current from the *commutator* (*q.v.*) of a *dynamo* or supplying current to the commutator of a motor. Brushes are made in a great variety of forms, the most common being carbon blocks or brass or copper gauze. The illustration shows two different types.

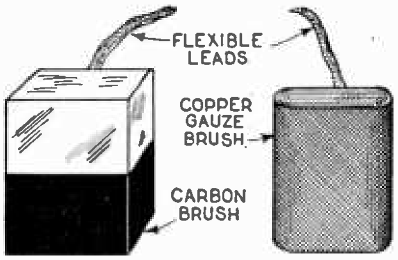

Two different types of Brushes. At the left is shown a carbon brush, and at the right a copper gauze brush.

BRUSH ANGLE—The angle formed by a brush in its contact with the *commutator.* The angle is usually 45°. If the brushes do not bed properly, sparking is apt to occur.

BRUSH DISCHARGE—A faintly luminous discharge which takes place on the surface of *conductors* charged to *high potential*, due to *ionization.* This effect can be noted at times on the *antenna* of a powerful transmitting station. (See *Corona.*)

BRUSH HOLDER—Metal clamp capable of adjustment which holds the *brush* in position on the *commutator* of a dynamo or motor. (See *Brush.*)

BRUSH LOSS—The loss in *watts* or power due to the friction of the brushes against the *commutator.* This loss is at a minimum when the brushes make perfect contact at the proper angle.

BRUSH PRESSURE—A term used to designate the pressure with which the brushes bear on the *commutator* of a dynamo machine and also referring to the voltage or *electrical pressure* delivered at the brushes.

B. S. G.—Abbreviation for British Standard Gauge of Wire, commonly known as B. S. (See *British Standard Wire Gauge.*)

B. T. U.—Abbreviation for "British Thermal Unit," being the commonly used heat unit. The amount of heat required to raise the temperature of a pound of water one degree Fahrenheit at ordinary atmospheric pressure.

BUCHER, Elmer E.—Born Akron, Ohio, Nov. 11, 1885. Educated at high school and private tutors. An American pioneer, Experimental Engineer with De-Forest Wireless Telegraph Company in 1903; constructed several high power stations in Middle West and Gulf Coast and later for United Wireless, which absorbed the Deforest Company. Organized training school for United Wireless and instituted first radio schools for Y. M. C. A. in New York City. Supervised commercial operations and conducted research work for United Wireless; later instructing engineer Marconi Wireless Tel. Co. Technical Editor "Wireless Age" 1913 to 1917. Appointed Commercial Engineer for newly formed Radio Corporation of America in 1920, and since 1922 has been in charge of general sales for that corporation. Author of numerous standard works on radio, including "Practical Wireless Telegraphy," "Wireless Experimenters' Manual," "Vacuum Tubes in Wireless Communication," and many others. Holds numerous U. S. patents on radio systems and devices.

BUCKLED DIAPHRAGM—Warping of the *diaphragm* of a phone as used in radio. Excessive voltages if applied direct to the head-phones will often bend the diaphragms. Any such defect may seriously impair the operation of the part and when it is noted that a diaphragm is bent or warped through rough usage or the application of too great voltages, it is best to replace it immediately.

BUCKLING OF PLATES—During discharge of a *storage battery* the plates gradually expand, owing to the fact that lead sulphate has about twice the volume of the same quantity of lead peroxide. Should this expansion or discharge take place too quickly, the plates will bend or buckle. (See *Storage Battery.*)

BUNSEN BURNER—A form of gas burner frequently used in radio construction for heating the soldering iron where electricity is not available. By special arrangement of the parts, the gas is combined with the proper amount of air before it reaches the burning part. This is done by driving the gas through a tube with small holes drilled to allow the air to be sucked in. This method promotes complete combustion and gives a non-luminous flame that heats the soldering iron without

THE BROWN & SHARPE (B. & S.) COPPER WIRE GAUGE—Below is a comprehensive copper wire table, which is the standard one in use in the United States. The resistance and cross-sectional areas of various sizes of conductors are given, so that considerable calculating can be done with the data at hand. A mil means 1-1000 (one thousandth) of an inch. Hence a wire with a diameter of 9 thousandths of an inch could also be stated as having a diameter of 9 mils. The area of cross-section of the wire in circular mils is found by squaring the diameter in mils; thus the 9 mil conductor would have 9 times 9, or 81 circular mils area. This is the end area of the wire, of course, and is used in all electrical figuring. If the resistance per 1,000 feet of a certain size wire is, say, 50 ohms, then 500 feet of the same size wire would have a proportionately less amount of resistance, or ½ of 50 ohms, that is, 25 ohms. The resistance per foot is found by dividing the resistance in ohms per 1,000 feet by 1,000. This wire table is for bare copper wire. Conductors, according to the B. & S. standard gauge, halve their sectional area in circular mils (Cir. Mils) for every 3 gauge sizes smaller wire; and double their sectional area in Cir. Mils for every 3 gauge sizes increase.

Gauge Number.	Size.		Weight and Length.			Resistance.			Carrying Capacity, 2,000 Amperes p. sq. in. section.
	Diameter in Mils.	Square of Diameter or circular Mils.	Grains per Foot.	Pounds per 1000 Feet.	Feet per Pound.	Ohms per 1000~~0~~ Feet.	Feet per Ohm.	Ohms per Pound.	Amperes.
0000	460.000	211600.0	4477.2	639.60	1.564	.051	19929.7	.0000785	430
000	409.640	167804.9	3550.5	507.22	1.971	.063	15804.9	.000125	262
00	364.800	133079.0	2815.8	402.25	2.486	.080	12534.2	.000198	208
0	324.950	105592.5	2236.2	319.17	3.133	.101	9945.3	.000815	165
1	289.300	83694.49	1770.9	252.98	3.952	.127	7882.8	.000501	130
2	257.630	66373.22	1404.4	200.63	4.994	.160	6251.4	.000799	103
3	229.420	52633.53	1113.6	159.09	6.285	.202	4957.3	.001268	81
4	204.310	41742.57	883.2	126.17	7.925	.254	3931.6	.002016	65
5	181.940	33102.16	700.4	100.05	9.995	.321	3117.8	.003206	52
6	162.020	26250.48	555.4	79.34	12.604	.404	2472.4	.005098	41
7	144.280	20816.72	440.4	62.92	15.893	.509	1960.6	.008106	32
8	128.490	16509.68	349.3	49.90	20.040	.643	1555.0	.01289	26
9	114.430	13094.22	277.1	39.58	25.265	.811	1233.3	.02048	20
10	101.390	10381.57	219.7	31.38	31.867	1.023	977.8	.03259	16
11	90.742	8234.11	174.2	24.89	40.176	1.289	775.5	.05181	13
12	80.808	6529.93	138.2	19.74	50.659	1.626	615.02	.08237	10.2
13	71.961	5178.39	109.6	15.65	63.898	2.048	488.25	.13087	8.1
14	64.084	4106.75	86.87	12.41	80.580	2.585	386.80	.20830	6.4
15	57.068	3256.76	68.88	9.84	101.626	3.177	306.74	.33133	5.1
16	50.820	2582.67	54.67	7.81	128.041	4.582	243.25	.52638	4.0
17	45.257	2048.19	43.33	6.19	161.551	5.183	192.91	.83744	3.2
18	40.303	1624.33	34.37	4.91	203.666	6.536	152.99	1.3312	2.5
19	35.390	1252.45	26.50	3.786	264.136	8.477	117.96	2.2392	1.96
20	31.961	1021.51	21.60	3.086	324.045	10.394	96.21	3.3438	1.60
21	28.462	810.09	17.14	2.448	408.497	13.106	76.30	5.3539	1.28
22	25.347	642.47	13.59	1.942	514.933	16.525	60.51	8.5099	1.08
23	22.571	509.45	10.77	1.539	649.773	20.842	47.98	13.334	.80
24	20.100	404.01	8.55	1.221	819.001	26.284	38.05	21.524	.63
25	17.900	320.41	6.77	.967	1044.126	33.135	30.18	34.298	.50
26	15.940	254.08	5.38	.768	1302.083	41.789	23.93	54.410	.40
27	14.195	201.49	4.26	.608	1644.737	52.687	18.98	86.657	.31
28	12.641	159.79	3.39	.484	2066.116	66.445	15.05	137.283	.25
29	11.257	126.72	2.69	.384	2604.167	83.752	11.94	218.104	.20
30	10.025	100.50	2.11	.302	3311.258	105.641	9.466	349.805	.16
31	8.928	79.71	1.67	.239	4184.100	133.191	7.508	557.286	.13
32	7.950	63.20	1.33	.190	5263.158	168.011	5.952	884.267	.098
33	7.080	50.13	1.06	.151	6622.517	211.820	4.721	1402.78	.078
34	6.304	39.74	.847	.121	8264.463	267.165	3.743	2207.98	.062
35	5.614	31.52	.658	.094	10638.30	336.81	2.969	3583.12	.049
36	5.000	25.00	.525	.075	13333.33	424.65	2.355	5661.71	.039
37	4.453	19.83	.420	.060	16666.66	535.33	1.868	8922.20	.031
38	3.965	15.72	.315	.045	22222.22	675.22	1.481	15000.5	.025
39	3.531	12.47	.266	.038	26315.79	851.789	1.174	22415.5	.020
40	3.144	9.88	.210	.030	33333.33	1074.11	.931	35803.8	.015

depositing soot as in the case of an ordinary luminous flame.

BURSTYN, Dr. W.—Born in Austria, 1877. Educated at Vienna University. Noted as: Developer of the *quenched spark system* in radio transmission, upon which he worked in conjunction with Baron Lepel in 1907 to 1912.

BUS BAR—A single bar, usually copper, which serves as a common connector for a number of pieces of apparatus. Also referred to as Omnibus Bar.

BUS BAR WIRE—Square copper wire, usually tinned, much used in making radio connections. This wire is generally furnished in lengths of two or two and a half feet and is particularly efficient for radio work, as it offers a low *resistance* connection, makes a neat appearance, and is suitable for soldering at joints.

BUSHING—Pieces of insulating material, usually fibre or rubber composition, used to *insulate* parts of radio apparatus or in mounting machine screws or shafts of *condensers* and other devices. The term is also applied to brass or other metal bearings through which the shaft of a piece of apparatus is run.

BUTT JOINT—A method joining two lengths of wire together by placing them end to end or at right angles and soldering the joint. The illustration shows a common butt joint.

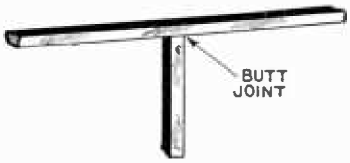

Showing how a Butt Joint is made.

BUZZER—An electric call signal device that makes a buzzing noise caused by the rapid vibrations of a contact breaker in a circuit using a small *battery* or *cell*. (See *Testing Circuit.*)

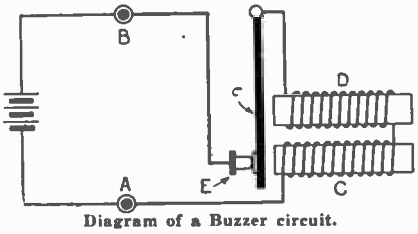

Diagram of a Buzzer circuit.

As the *current* flows from the posts A B the *electro-magnets* C and D are energized and draw down the vibrator spring F. This breaks the contact between E and F and as the current no longer flows the magnets release the spring F, which returns to its normal position and establishes the circuit once more. This making and breaking of the circuit takes place so rapidly that the vibrator makes a buzzing or humming sound, the speed with which it vibrates determining the pitch of the note. (See *Frequency*, also *vibration.*)

BUZZER EXCITER—A method of using a *buzzer* to produce local signals in a crystal receiving circuit to permit sensitive adjustment of the *detector*. The illustration shows a common method of producing signals to test the receiving set in this manner. In this case a small coil having a few turns of wire is placed in the buzzer circuit as indicated and a similar one is placed in the ground lead of the receiving set. When the button is depressed the buzzer is set in vibration and a transfer of energy takes place between L1 and L2. This causes an *oscillating* impulse (*q.v.*) to flow through the receiver circuit. The detector can then be adjusted to the most sensitive point at which it can be left until outside signals are heard. Another but less efficient method of exciting the circuit is to connect a short length of wire from the contact point of the buzzer to the ground connection of the receiving set. The former system is much more practical and furnishes better test signals.

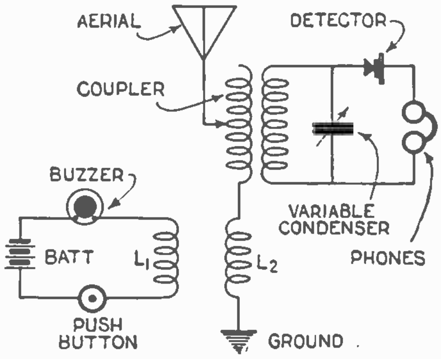

Buzzer Exciter.

Coil L1 being placed in *inductive relation* to coil L2, which is in the antenna circuit, undergoes a change of *flux* when the buzzer is *excited.* This sets up a *difference of potential* across coil L2 which charges the antenna ground system and produces oscillation at a definite *frequency* according to the adjustment. Now if the secondary circuit is tuned to resonance with the primary circuit, the signals will be audible in the phones when the detector is properly adjusted. Also called *Buzzer Wave Generator.*

BUZZER MODULATION—See *Modulation.*

BUZZER PRACTICE SET—A combination of a *buzzer* and signalling key arranged on a baseboard used for the purpose of practicing signalling.

B. W. G.—The abbreviation for Birmingham Wire Gauge—a system of gauging wire used in England. (See *B & S Gauge*, also *British Standard Wire Gauge.*)

BY-PASS CONDENSER — See *Condenser*, also *Filter.*

C

C—The chemical symbol for carbon. When printed in *italics*, it is also the symbol for *coulomb* in electrical practice. It is the international symbol for *capacity.*

CABLE—A number of wires stranded together, used to carry electrical currents, heavily insulated and often covered by a sheath or outer coating of lead, rubber, gutta percha, or braided silk or cotton. The term is used very often in a broad sense to signify any heavy wire or strand of wires that may be insulated more thoroughly than is ordinarily the case. (See *Conductor*, *Stranded Wire*, etc.)

CADMIUM—A chemical metallic element somewhat resembling zinc. In color, white with a bluish tinge. It is useful in fusible alloys because of its property for uniting readily with other metals, having a low melting point. Its electrical conductivity is about 25% of that of silver. It is very useful in the construction of a crystal detector, the molten metal holding the crystal firmly and as it melts at a low temperature the heat will not affect the sensitivity of the crystal. Cadmium is also used in making cells, notably the *Weston Cell.* (See *Weston Cell.*)

CAGE ANTENNA—An aerial having a number of wires arranged in cylindrical form, suspended either vertically or as a horizontal top. So named because of its resemblance to a cage. (See *Aerial.*)

CALCULATION OF CAPACITY—The mathematical or comparative determination of *capacity.* There are various formulas in use for calculating capacity under various conditions, such as for capacity of various types of condensers, or the capacity of a circuit. Rough approximations of capacity of condensers might be obtained by direct comparison with a known value. Other formulas are used for the calculation of antenna capacity. For determining the capacity of flat condensers, such as fixed condensers used for transmitting or receiving circuits, the following formula is generally used:

$$C = 0.0885\,K\,\frac{S}{r}$$

In this case, C is the capacity in micro-micro farads; to obtain result in the usual units of micro-farads, divide the result by one million. S is the surface area of one plate in square centimeters; r is the thickness of the dielectric in centimeters and K a constant depending on the material between the plates of the condenser. (See *Dielectric Constants.*)

CAL-ELECTRICITY — Electricity produced in the secondary of a transformer due to changes in temperature of the core. This current is in addition to the current normally induced in the secondary. (See *Transformer*, *Core*, also *Core Losses.*)

CALIBRATION—A process whereby instruments such as *galvanometers* ammeters, *wavemeters*, etc., are made to indicate certain values. Before such

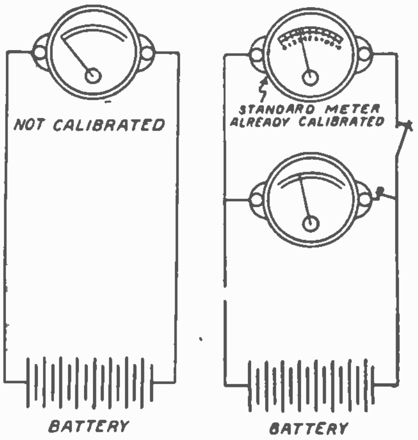

Fig. 1. Fig. 2.

Calibrating a Meter

instruments can be made use of, it is necessary to mark the dials in such a manner that the indicator showing on the dial or scale will have a definite meaning in terms of electrical values. Thus, a voltmeter before calibration is shown in Fig. 1. When it is placed across a battery, the pointer will move to some position on the white space, but without indicating any definite value. Now by testing with another meter which is already marked or calibrated in volts, we find, for example, that when the pointer is at the middle point in its swing, it indicates 10 volts. Similarly various other positions are found to indicate more or less voltage by comparison with the standard. This is shown in Fig. 2. It will be understood that one of the small switches must be open at all times. The reading of the calibrated meter is taken with the other meter out of the circuit (switch open), then the reading is marked on the other meter at the point indicated with the calibrated meter out of the circuit. Such a method would, of course, be inefficient and it is shown only to give a rough idea of the purpose and meaning of calibration. Another example of calibration is in the case of a dial on a receiver. Fig. 3 shows the usual form of dial used in tuning a receiving set. The dial is marked off to scale, from 1 to 100, the numbers having no particular significance. Now if we find that when the tuning dial is set at 20 it is adjusted for a wavelength of 260 meters, we can put that figure in place of the 20. The same procedure can be followed throughout the entire range of the dial, and when thus marked in terms of a definite quantity—wavelength in meters in this case—the dial is said to be calibrated.

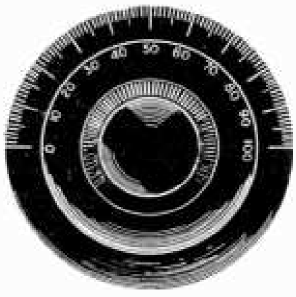

Fig. 3.

A dial as used on radio receiving sets may be Calibrated according to different wavelengths.

The actual method to be followed out in calibrating will depend on the equipment available and the accuracy desired. If a standard of comparison is available, it saves a great deal of work. Thus, a voltmeter can easily be calibrated by means of a standard meter, already calibrated, and a source of steady voltage controlled by a rheostat. An ammeter can be calibrated in the same manner by using a standard ammeter, and a wavemeter can be calibrated from another wavemeter of known values.

Calibration may be done in such a manner that the readings require mathematical solution to determine their value in terms of actual volts, amperes, meters and so on, or it may be *absolute*, in which case the readings are directly in terms of the quantity and require no solution. (See *Voltmeter Calibration, Wavemeter Calibration*, also *Rough Calibration.*)

CALIDO—An alloy containing nickel and chromium with a small percentage of iron. The melting point is very high, about 1550 degrees centigrade. Calido wire is much used as a *resistance*, particularly for heating devices. The chief characteristics are as follows: maximum working temperature 1100 degrees centigrade; microhms per cubic centimeter at 20 degrees centigrade, 100; temperature coefficient per degree centigrade, 0.00034; specific gravity, 8.2.

CALLAND CELL—A primary cell used in French telegraph work. It is a form of gravity cell having a negative electrode of copper and a positive electrode of amalgamated zinc, the electrolyte being zinc sulphate. Crystals of copper sulphate are used as a depolarizer. (See *Gravity Battery.*)

CALORIE—The unit of heat in the C. G. S. (Centimeter-Gram-Second) (q.v.) system. It is equivalent to the amount of heat necessary to raise the temperature of one gram of water from 0° to 1° centigrade. (See *Thermal*, also *C. G. S.*)

CALORIMETER—In electrical practice, an instrument used to measure the heat generated in a conductor carrying an electric current.

CAMBRIC, VARNISHED—Varnished muslin used as an insulating material. Also known as Empire Cloth. (See *Spaghetti.*)

CANAL RAYS—When an electric discharge takes place between the *anode* and a perforated *cathode* in a vacuum tube, fine pencils of light are seen to pass through the perforations in the cathode. These rays are called *canal rays*, and consist of positively charged particles. They produce phosphorescence on the wall of the tube.

Canal rays travel at a lesser velocity and in the opposite direction to the cathode rays. This fact and also the fact that the rays are deflected by powerful electric or magnetic fields in the opposite direction to cathode rays, is taken as proof that they consist of positively charged particles. (See *Cathode Rays.*)

CAPACITANCE—A term very often used as synonymous with *capacity*. Due to the fact that capacity may refer to the current carrying ability of a conductor and also to electrostatic capacity, it has been suggested that capacitance be used to refer only to *electrostatic capacity* of a body or device. Capacity would then refer to *current carrying ability*. (See *Capacity*, also *Electrostatic.*)

CAPACITATIVE REACTANCE—That part of the *reactance* of a circuit carrying alternating current which is due to the *capacity* in the circuit. (See *Reactance* also *Impedance.*)

CAPACITATIVE COUPLING — See *Coupling*.

CAPACITY—Generally speaking, the quantity of electricity in any form which a body is able to store or contain. The term is usually qualified to denote the particular case, such as *electrostatic capacity*, which is a measure of the ability of a condenser to store up energy in the form of electrostatic charges; also *current-carrying capacity* of a conductor, which is considered as the ability of a wire to carry a certain amount of electric current without overheating. The tendency in electrical literature is to use the term capacitance as applying to the electrostatic capacity of a condenser. Thus, a certain condenser may have a capacitance of 2 micro-farads. The unit of electrostatic capacity is the *farad*. This is understood as the capacity of a condenser that will store one *coulomb* of electricty under an electromotive force of one *volt*. When V is expressed in volts, C in farads, and Q in coulombs, $C = Q/V$.

CAPACITY CIRCUIT—An electrical circuit in which the *capacity* is very large compared with the *inductance*. The inductance may be considered negligible.

CAPACITY CONVERSION FACTORS—In the calculation of *capacity*, as the capacity of a condenser, there are several different units. For instance, a *farad* is a large unit of capacity and is equal to one million *microfarads*, or conversely, a microfarad is equal to one millionth of a farad. In the same way a farad is equal to one billionth of an *abfarad*. The four units used in capacity measurement are the *farad*, *micro-farad*, *micromicrofarad*, *statfarad* and *abfarad*. The following table gives the conversion factors for these various units:

$$1 \text{ abfarad} = \begin{cases} 10^{9} \text{ farads} \\ 10^{15} \text{ mfds.} \\ 9 \times 10^{20} \text{ stfds.} \end{cases}$$

$$1 \text{ farad} = \begin{cases} 10^{-9} \text{ abfd.} \\ 10^{6} \text{ mfds.} \\ 9 \times 10^{11} \text{ stfds.} \end{cases}$$

$$1 \text{ mfd.} = \begin{cases} 10^{-15} \text{ abfd.} \\ 10^{-6} \text{ farads} \\ 9 \times 10^{5} \text{ stfds.} \end{cases}$$

$$1 \text{ stfd.} = \begin{cases} 1/9 \times 10^{-20} \text{ abfd.} \\ 1/9 \times 10^{-11} \text{ farad} \\ 1/9 \times 10^{-5} \text{ mfd.} \end{cases}$$

$$1 \text{ mmfd.} = 10^{-6} \text{ mfd.}$$

In the table above, farad is abbreviated Fd.; abfarad is shown as abfd.; microfarad is mfd., micromicrofarad is mmfd. and statfarad is stfd. In dealing with large figures containing many ciphers the amount is shown as 1,000,000 is 10^6 and $\frac{1}{1{,}000{,}000}$ is 10^{-6}, etc. (See *Capacity, Farad*, also *Unit.*)

CAPACITY EARTH or CAPACITY GROUND—A substitute for the usual ground connection where the wires or plates are buried beneath the ground or attached to a water or steam pipe. It is in effect a second aerial, and is placed either under the regular aerial or to one side. It is more used in radio transmission than in reception. (See *Counterpoise.*)

CAPACITY ELECTROSTATIC — See *Capacity* also *Electrostatic Capacity.*

CAPACITY FREQUENCY FACTOR—The relation between the apparent capacity of a condenser and the *electrostatic capacity*. (See *Capacity, Condenser* also *Electrostatic.*)

CAPACITY OF CONDENSERS IN PARALLEL—When several condensers are connected in parallel the resultant capacity is the sum of the individual capacities. It is written:

$$\text{Capacity} = C_1 + C_2 + C_3, \text{etc.}$$

It will be evident that the resultant capacity of condensers connected in parallel is just the reverse of case for resistances, where the total resistance is less than that of the smallest individual resistance in the parallel connection. (See *Resistance Measurement* also *Capacity of Condenser in Series*, also *Condensers in Series* and *Condensers in Parallel.*)

CAPACITY OF CONDENSERS IN SERIES—When several condensers are connected in series with each other the resultant capacity is always less than the capacity of the smallest condenser in the group. The capacity of condensers in series is given as follows:

$$\text{Capacity} = \cfrac{1}{\cfrac{1}{C_1} + \cfrac{1}{C_2} + \cfrac{1}{C_3}, \text{ etc.}}$$

The resultant capacity is thus the reciprocal of the sum of the reciprocals of the several capacities. (See *Condenser, In Series.*)

CAPACITY MEASUREMENT OF ANTENNA—A simple method for the measurement of the capacity of an antenna as given by Austin follows: The illustrations show the general arrangement. The inductance L in Fig. A is a coil that will increase the wavelength (natural) of the antenna system 4 or 5 times; C is a calibrated variable condenser and W a standard wavemeter. In operation the coil L is connected in the antenna circuit, the buzzer is used to excite the antenna circuit and the wavelength is measured by w. The coil L is then taken out of the antenna circuit and shunted by the condenser C. The condenser C is then varied until the wavelength of the combined coil L and condenser C is the same as that obtained before with the coil in the antenna circuit but without the variable condenser. The capacity of the condenser now will be approximately that of the antenna, ignoring the distributed inductance of the antenna system. Fig. A shows the first operation, the measurement of the antenna with the coil in the circuit, and Fig. B shows the measurement of the wavelength of the coil and condenser. (See *Mutual Capacity, Wavemeter, Natural Wavelength,* also *Distributed Capacity.*)

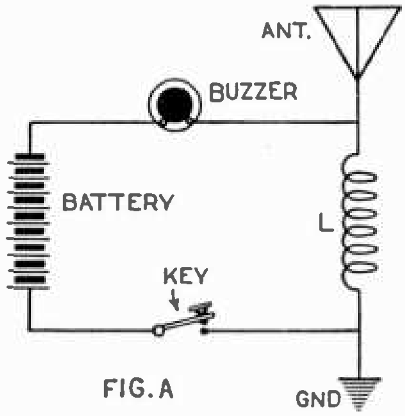

FIG. A

The first operation in the measurement of the antenna. The wavemeter is shown below.

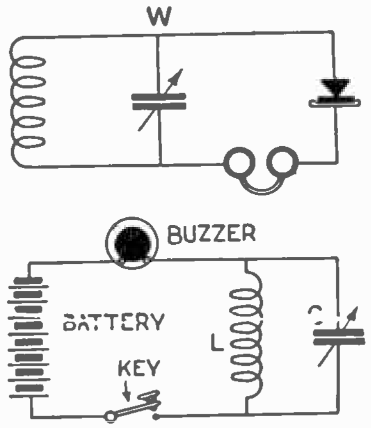

FIG. B

Circuit for the measurement of wavelength of coil and condenser.

CAPACITY MUTUAL—See *Mutual Capacity.*

CAPACITY RESISTANCE—The resistance to alternating currents offered by a body possessing electric capacity. (See *Capacity,* also *Resistance.*)

CAPACITY SWITCH—Any switch used in a circuit to introduce or cut out capacity (condenser).

CAPACITY (UNIT OF)—The unit of capacity is the farad. A condenser has a capacity of 1 farad when 1 coulomb is required to raise its potential from zero to 1 volt. Since the farad is very large, its millionth part, or the microfarad is generally used.

CARBON—A non-metallic chemical element used for many purposes in electricity and radio. It exists in three forms, two crystalline and one non-crystalline. The first two are diamond and graphite, the third charcoal. Carbon is used as an electrode in electric arcs, as one of the poles in certain types of primary cells, as a material in place of wire for various forms of resistance, and as brushes in motors and generators. (See *Diaphragm, Carbon.*)

CARBON RHEOSTAT—A rheostat or variable resistance using carbon in place of the customary wire. In its simplest form a carbon rheostat may be merely a piece of carbon having a sliding contact. This is shown in Fig. 1. Owing to the relatively high resistance of carbon used in this manner, such a controlling resistance will generally be termed a *potentiometer* (q.v.) and used as such.

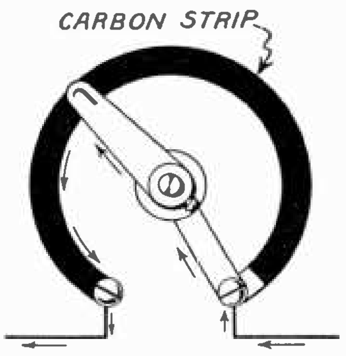

Fig. 1.—A carbon rheostat with a rotary sliding contact.

The most practical form of a carbon rheostat is known as a carbon pile rheostat. This type uses a number of flat pieces of carbon of relatively low resistance placed together and arranged with a knob and screw in such manner that the pressure can be varied. This change in pressure alters the resistance of the mass and allows control of the current for vacuum tubes. The same principle is used for very high resistances as well as for low values. The principle will be seen from the illustration Fig. 2.

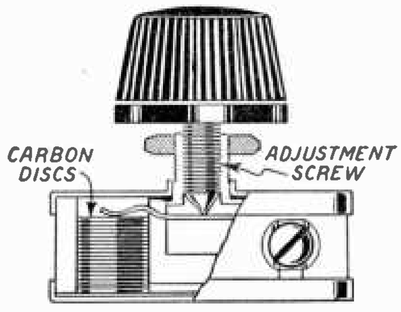

Fig. 2.—A carbon pressure type rheostat.

CARBORUNDUM—An abrasive material, a product of the electric furnace, composed of silica and carbon. It is used in radio as a crystal rectifier, although the strides of radio have practically rendered it obsolete, many other better crystals being in use. The chief difficulty with its use as a rectifier or crystal detector was the fact that for efficient operation it required a cell or battery and potentiometer to regulate the current flowing through it. The simple crystal circuit using a carborundum detector is shown in the illustration. C is a tuning coil, D the carborundum crystal and contact, P a potentiometer and T the head phones. (See *Rectifier,* also *Detector.*)

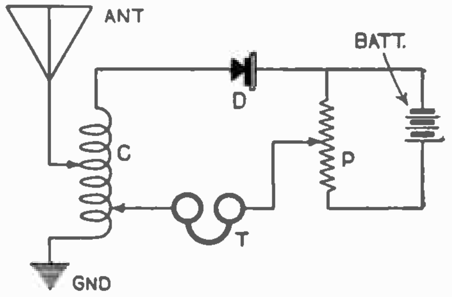

Diagram of a radio circuit employing a carborundum crystal detector.

CARDBOARD TUBING—Tubing made of laminations of paper pasted together. They are used for home made coils of all descriptions, and while not as efficient as the composition tubing, can be used for many purposes without any great loss. As a general rule it is better to use some of the numerous compositions, such as bakelite or rubber compounds.

CARPENTIER, JULES—French radio expert. Born in Paris, in 1854, he joined the Ecole Polytechnique in 1871, and in 1876 was appointed principal stores engineer of the Lyons Railway Company, making a special study of electricity. For his work in electricity he obtained in 1881 the cross of Chevalier of the Legion of Honor. One of the early pioneers of radio in France, he founded the Compagnie Generale Radiotelegraphique, afterwards absorbed in the Compagnie Generale de Telegraphie sans Fil. Carpentier is a member of the Academie des Sciences, Commander of the Legion of Honor, and President of many scientific societies.

CARRIER WAVE—The *continuous wave* (q.v.) of radio frequency generally thought of as carrying the voice or music waves from the radio broadcasting station. Actually it is a radio wave of high frequency, which is altered in amplitude by the music or speech transmitted. In action, this continuous wave has another wave super-imposed on it. This other wave having wave form and amplitudes determined by the voice or other sounds being transmitted. The illustration shows the carrier wave, the waves representing speech or music and the combination of the two. This effect of varying the amplitude or strength of a continuous wave by means of some sort of program is known as *modulation.* (q.v.) The wave representing the speech or music is imposed on the carrier wave resulting in the combined wave which reaches the listener and is translated by suitable apparatus into approximately the same sounds that are transmitted. (See *Modulation, Broadcasting*, also *Speech Vibrations.*)

CARRYING CAPACITY—See *Current Carrying Capacity.*

CARTRIDGE FUSE—A fuse in which the fuse wire is surrounded by some non-inflammable substance, enclosed in a cartridge-like tube and having brass lugs soldered to caps at ends. Used to prevent the hot wire from "flying" when fused. (See *Fuse.*)

CASCADE—A term applied to pieces of apparatus connected together in series, particularly vacuum tubes. In this case, the arrangement would be such that each vacuum tube would amplify the signal output of the preceding tube. The tubes will be so connected that the output of one is introduced as input to the next and so on the total result being greatly increased signal strength from the output of the last tube. This term may apply to either radio frequency or audio frequency amplification. (See *Amplifier, Radio Frequency*, also *Audio Frequency.*) A simple cascade arrangement is shown in the illustration. Fig. 1 is a circuit with two stages of cascade radio frequency amplification, and Fig. 2 a circuit using two stages of audio frequency amplification with the usual detector tube in each case. A combination of both radio frequency and audio frequency stages of amplification might also be used, the term cascade amplification also applying here.

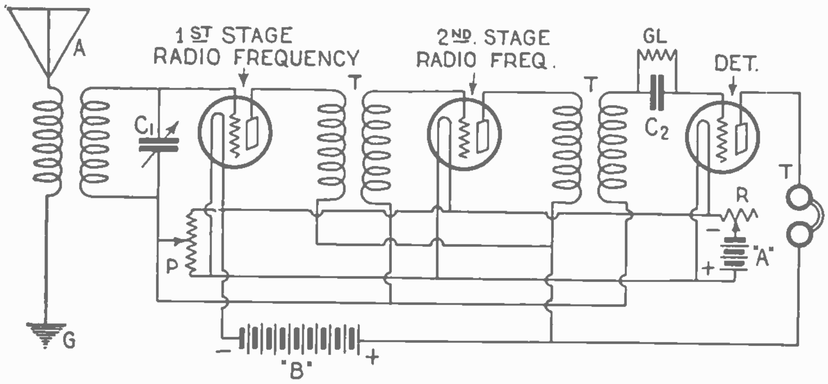

Fig. 1.—Circuit diagram showing two stages of cascade radio frequency amplification and detector.

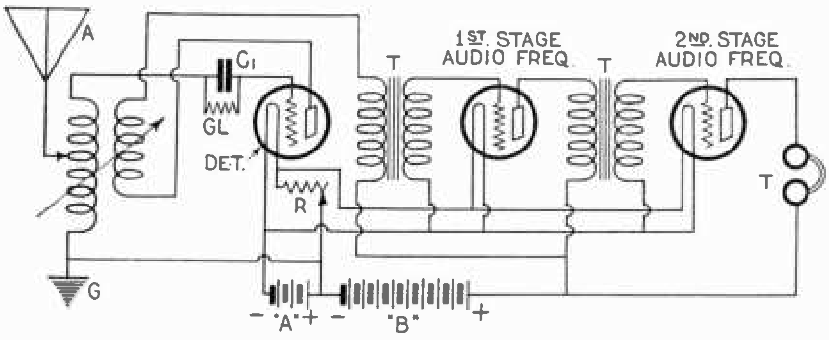

Fig. 2.—Diagram of detector and two stage cascade audio frequency amplifier.

CATHION or CATION—The charged particles which appear at the cathode or move toward it through the electrolyte of an electrolytic cell. (See *Anode.*)

CATHODE or KATHODE—A *negative electrode*. The term is often applied to the filament of a vacuum tube to distinguish it from the anode or plate. (See *Cell, Electrolysis*, also *Anode.*)

CATHODE RAY TUBE—A vacuum tube having a high vacuum, that is from which practically all air has been exhausted, used in the production of cathode rays. These tubes in their modern form are used for many purposes in studying the nature and form of electric waves. The illustration shows one of the latest types of tube, known as the *Cathode Ray Oscillograph*. This tube is used for the following purposes: examination of the wave forms of various rectifiers, study of wave forms of different types of generators, study of vacuum tube characteristics, examination of the characteristics of X-ray and other types of tubes, examination of radio frequency waves modulated as in broadcasting, measurement of hysteresis and dielectric loss in various materials, study of phenomena in arcs and sparks, together with numerous other associated effects. The development of this tube is one of the marvels of science, enabling the research engineer to virtually see the electric wave, and to study exhaustively, electric phenomena otherwise invisible.

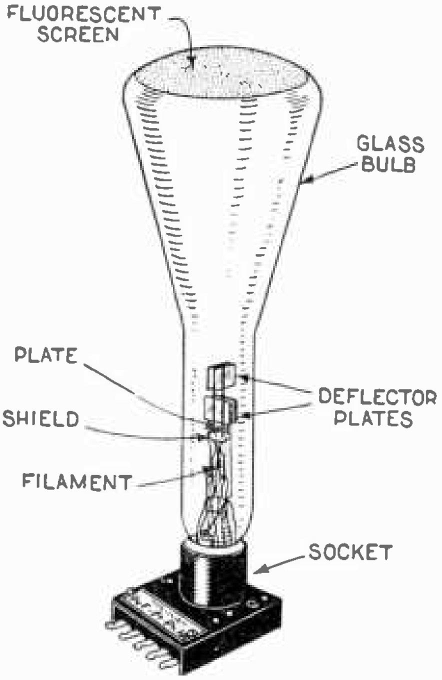

A cathode ray tube with socket.

The Western Electric modification of the Braun Cathode ray tube or oscillograph, is a three electrode glass tube about 30 centimeters (about twelve inches) long, in the form of a cylinder an inch and three-quarters wide, spreading out in conical shape, the tube being about three and a half inches wide at the top. The *cathode* is a filament coated with active oxides and arranged to emit the number of electrons required for the cathode rays, at a dull red heat. The *anode* or plate is a small platinum tube set very near the filament. Between the filament and the anodes is a small circular screen having a hole just a little smaller than the circular filament. A battery of several hundred volts is connected between the filament and plates with the positive terminal connected to the plate. The electrons emitted from the hot filament are controlled by this field in the same manner as in an ordinary vacuum tube. A small portion of these electrons pass completely through the tubular anode and constitute the *cathode* rays. These rays are passed between two pairs of deflecting plates set at right angles to each other, and fly against the large end of the tube. This end is covered with a fluorescent mixture which renders the rays visible. These rays are deflected in various ways. For example, if an alternating current is applied the visible spot on the screen will be drawn out into a line. (See *Electron, Vacuum Tube*, also *Wave Analysis.*)

CATHODE RAYS—The stream of electrons or electrical particles sent out from the *cathode* or filament of a vacuum tube. These rays are negatively charged. (See *Cathode, Electrons, Vacuum Tube* also *Oscillograph.*)

CAT-WHISKER—The fine wire used with certain crystal detectors to make contact with the crystal. Usually a springy metal such as phosphor bronze wire.

"C" BATTERY—One or more small cells having a voltage generally between 2 and 10 volts, used in the grid

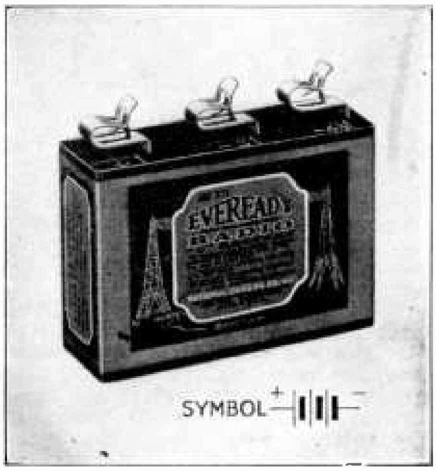

Fig. 1.—A 4½ volt "C" battery. This type of battery has a common positive terminal and is tapped at negative terminals for 3 and 4½ volts.

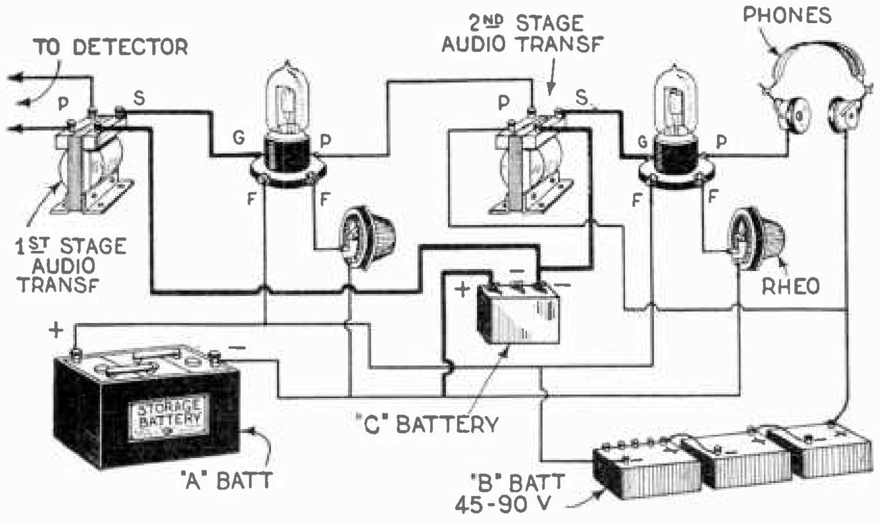

Fig. 2.—A diagram showing how the "C" battery is connected in the audio frequency amplifier.

circuit of an amplifier tube for the purpose of impressing a negative potential on the grid of the tube. The method of connecting it in the circuit is shown in Fig. 2. Fig. 1 shows a typical 4½ volt "C" battery. (See also *Grid Bias* and *"C" Battery Curve.*)

"C" BATTERY CURVE—In order to show the effect of a "C" battery on the consumption of current from the "B" battery in an audio amplifier, a *curve* (q.v.) may be drawn as in the illustration. In this case the test was made with a type UV 201A tube used as a one stage *audio frequency amplifier* in a conventional circuit. It is readily apparent on referring to the lines A and B that when the "C" battery is in use the consumption of current by the plate is much less than without it.

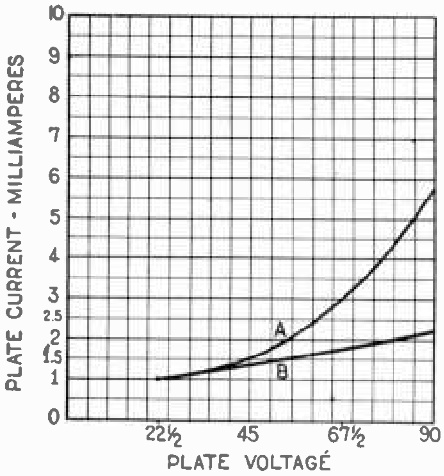

Test conducted on UV201A tube using single stage audio amplifier with conventional circuit.
Plate Voltage—Abscissae.
Plate Current in milliamperes—ordinates.
A—Consumption curve without "C" Battery.
B—Consumption curve with "C" Battery.
(Note: 1.5 ma. at 45v plate rising to 5.75 ma. at 90v plate without "C". With "C" Battery consumption maximum is 2¼ ma. at 90v.)

CELL—A cell is one of the four chief sources of electric energy. In a cell the energy is created by an *electro-chemical* process, using in its simple form two unlike metals immersed in a dilute acid or alkaline solution. In the illustration Fig. 1 the common form of electric cell is shown. C is a strip of *carbon* and Z a strip of *zinc*. These are placed in the jar in a conducting solution of sal ammoniac. If the exposed terminals from the carbon and zinc strips are connected together by a piece of wire, current will flow from one side of the cell to the other through the wire. The current flows from the carbon or *positive* pole to the *negative* or zinc pole through the wire, completing the circuit through the solution from zinc to carbon. (See *Current, Direction of Flow.*) The cell described is known as a *primary cell.*

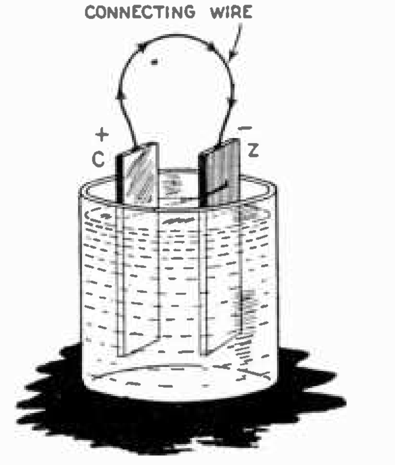

Fig. 1.—Common form of electric cell.

In Fig. 2 a simple *storage cell* or *secondary cell* is shown. In this case there are two lead plates immersed in a dilute solution of sulphuric acid. A source of primary current (primary cell) may be connected to the secondary cell in the manner shown in Fig. 3 and after being connected for a certain length of time it will be found that the energy of the primary cell has been transferred to the secondary cell and stored in it. This process of placing electric charges in a secondary cell is known as *charging*, and the current stored up in it is known as the *charge*. Now, if the wire is connected from one terminal to the other it will be found that electric energy can be withdrawn in a similar manner to that of the primary cell. A secondary cell may contain several plates and several of these cells may be connected together to form a storage battery. (See *Storage Battery, Current, Production of,* also *Anode* and *Cathode.*)

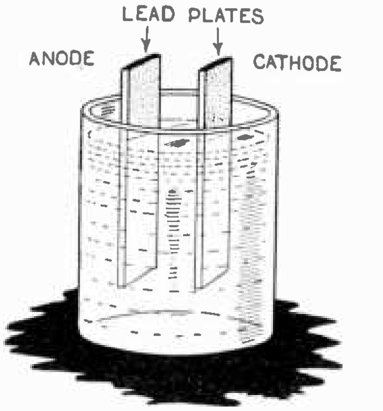

Fig. 2.—Storage or secondary cell.

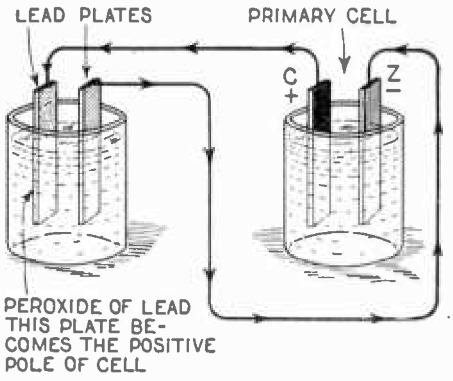
LEAD PLATES
PRIMARY CELL
C
+
Z
−
PEROXIDE OF LEAD THIS PLATE BECOMES THE POSITIVE POLE OF CELL

Fig. 3.—Primary cell connected to secondary or storage cell.

CELL, SECONDARY—When a *primary cell* is connected to a *storage* or *secondary cell*, the current from the primary cell flows into the secondary cell and in effect deposits energy in that cell. Then when the charging source (the primary cell) is withdrawn, the secondary cell can be used in practically the same manner as the primary cell. A wire connected from one terminal to the other will permit passage of current. The action of storing up energy in a storage cell is approximately as follows: When the charging current flows from plate to plate through the solution in the secondary or storage cell, the plate connected to the *positive* (+) *pole* (carbon or copper) of the primary cell, receives a brown coating of *peroxide of lead.* At the same time the other plate becomes spongy or porous. Now as one plate has received a coating while the other remains unchanged as far as its surface is concerned—it acts as a primary cell if the charging source is taken away and the two terminals connected. It now has all the essentials of an ordinary chemical cell. As long as the coating of peroxide of lead remains on the surface of one of the plates, the cell will be capable of delivering current.

When the coating has been worn off or eaten away the cell is said to be in a discharged condition. It will then be necessary to go through the same process of connecting a primary source to the terminals in order to again deposit peroxide of lead on one plate. It will therefore be evident that when the *storage cell* is charged, *current* is not actually stored in it, although in effect it has received potential energy. What has actually occurred is that the current supplied to the cell during the charging process has produced electro-chemical changes, making the plates dissimilar and thus producing a difference of potential. (See *Current, Production of*). The *voltage* or *electromotive force* of primary cells varies from .06 to 1.5 volts according to the nature of the elements used and the grade of the *electrolyte* (q.v.). Secondary or *storage cells* of the lead plate type produce from 2.1 to 2.6 volts. (For more complete explanation of the theory and action of a secondary cell, see *Storage Battery.* See also *Electro Motive Force, Current* and *Voltage.*)

CELL, THEORY OF PRIMARY—It has been explained under the heading *Cell,* that a primary cell produces electrical energy by chemical action. More specifically the action of a simple cell is as follows: When the copper and zinc, or carbon and zinc placed in a cell are connected together by a conductor, (wire or other conducting object) the acid in the case of a sulphuric acid solution, attacks the surface of the zinc plate and forms a compound known as zinc sulphate. While this sulphate is being formed some of the hydrogen contained in the sulphuric acid is liberated in the form of bubbles which appear on the copper plate. Some of these bubbles escape into the surrounding air by rising to the surface of the solution, but others cling to the surface of the copper plate and gradually cover it with a film of hydrogen. It may be said here that the decomposition of the zinc plate in the acid solution furnishes the *electro-motive force* required to cause the flow of current between the plates, through the solution and through the conductor. Now, hydrogen is a non-conductor, and as the hydrogen gradually covers the copper plate, the surface of the plate in contact with the solution gradually decreases. The flow of *current* from zinc plate to copper plate in the solution is thus gradually diminished, until eventally we can imagine the copper plate as being entirely covered with hydrogen and no longer in contact with the solution. In addition to this insulating action of the hydrogen, it tends to set up a current within the cell in the opposite direction to the normal flow of current produced by the chemical action. A cell in this condition is said to be "polarized," and various means have been used to cut down or eliminate the *polarization effect.* The purpose of these various means is to prevent the hydrogen bubbles clinging to the surface of the copper plate, thus allowing them to escape to the top of the solution and thence into the surrounding air. (*See Polarization, Current, Voltage, also Storage Cells.*)

CENTIMETER—A measure of length in the metric system. It is one hundredth part of a meter, or approximately .3937 inch. It is used as the unit of length in the Centimeter Gram Second System (q.v.).

CENTIMETER GRAM SECOND or C. G. S. A system of units of measure employed in practically every branch of engineering, particularly in electrical practice. It is a decimal system, the centimeter being the unit of length, the gram the unit of weight and the second the unit of time. All other units are derived from these three. When the magnitude of a quantity is spoken of, it refers to the relative magnitude of the quantity when compared with some other quantity of the same nature. The magnitude used as a standard of comparison is termed a unit. Thus, if the length of a body is measured in feet, as for example, ten feet, it means that the magnitude of the body is ten times unity or ten times one foot. Similarly a *current* of ten *amperes,* is a current of magnitude ten times as great as the unit quantity—one ampere. Now an ampere is the practical unit of rate of flow of electric current. The unit has a certain definite value. (See Ampere.) The magnitude of any other similar quantity measured in amperes will mean a value equal to a certain number of units. In addition to the C. G. S. system of units, there is a practical system. This is commonly used for ordinary calculations. Under this system, the following are the main units: The *volt* is the practical unit of *electro-motive force;*—the force required to maintain a flow of current of one ampere through a resistance of one ohm. The unit of *current strength* or rate of flow of current is the *ampere;*—the strength of current maintained by a force of one volt through a resistance of one *ohm.* The unit of *resistance* is one *ohm,* and is the resistance of a *conductor* or *circuit* that will permit the passage of a current of one ampere under the force of one volt. The unit of *quantity* is the *coulomb,* which is the quantity of current flowing in a circuit when one ampere passes a given point during one second of time. The *watt* is the practical unit of *electrical power,* being the power of a current of one ampere flowing in a circuit under the pressure of one volt.

It will thus appear that the practical system is actually based on the C. G. S. system. Calculations of an involved nature necessarily make use of the C. G. S. system, and when multiples or sub-multiples are used they are expressed in Greek and Latin prefixes. For example, a million ohms is 10^6 ohms and is expressed as a *meg-ohm.* "Meg" being derived from *mega millian.* Similarly a millionth of an ampere is expressed as a *micro-ampere* or 10^{-6} AMPERE. The various units under the C. G. S. and practical systems are taken up in detail under the respective headings, as *JOULE. ERG, DYNE* and so forth. (See also *Inductance (Unit of); Capacity (Unit of); Electrostatic Units and Electromagnetic Units.*)

C. E. M. F.—The abbreviation for *counter electromotive force* (q.v.).

C. G. S. UNIT OF CURRENT—The absolute *electromagnetic unit of current.* It is the current which when passed through a conductor bent into a circle of one centimeter radius will attract or repel a unit magnetic pole placed at its center, with a force of one *dyne* (q.v.).

CERUSITE—A kind of mineral used as a rectifier in some types of *crystal detectors.* It is a carbonate of lead having a whitish-grey color. This mineral is not as good a rectifier as some of the others, but has the advantage of more uniform sensitivity over practically its entire surface. (See *Crystal Detectors.*)

CHALCOPYRITE—A crystal used in a *pericon detector,* which is composed of zincite and chalcopyrite in contact with each other. It is a combination detector suitable for use where it is subject to jarring, since it has nearly uniform characteristics over its entire surface. It works without the aid of an electric battery.

CHARACTERISTICS—A term broadly applied in electricity to refer to any distinguishing feature in electrical apparatus. In radio usage the term is generally applied to *vacuum tubes.* The characteristic of a tube is understood as the relation between the potential (voltage) at the *grid* (q.v.) and the resulting current obtained from the *plate.* (q.v.) It may also occasionally refer to other features of a tube, such as its current requirements or amplifying quality as distinct from the ratio of *grid voltage* to *plate current.* (See *Characteristic Curve, Vacuum Tube Characteristics;* also *Plate Current.*)

CHARACTERISTIC CURVE—A diagram showing graphically in the form of a curved line the relation of changing values. Thus, in the case of a *vacuum tube,* a change in the *grid* voltage produces certain changes in the *plate* current (q.v.) and these changes may be plotted as curves. The temperature of a wire carrying current increases with increase in the *current* passing through it, and this increase may be shown by a curve. (See *Vacuum Tube Characteristics.*)

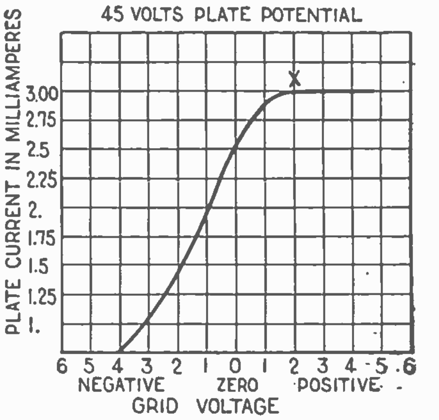

Characteristic chart of vacuum tube grid and plate voltage.

CHARGE—The presence of electricity in a body. Is usually defined as an excess or deficit of *electrons* (q.v.) on an insulated body. In the case of a *condenser* (q.v.) after it has accumulated electrical energy it is said to be in a charged condition, or to have an *electro-static charge.* The charge in a *condenser* is measured in *coulombs.* (q.v.) A *storage battery* is said to contain a charge when it has been connected to a source of direct current for a length of time, and the hydrometer test shows the electrolyte at proper specific gravity.

CHARGE, SPACE—See *Vacuum Tubes,* also *Space Charge.*

CHARGED BODY—Any body that has accumulated electrical energy. (See *Charge.*)

CHARGER, STORAGE BATTERY—Any means of controlling the current delivered from a primary source for the purpose of charging *storage batteries* or other *secondary cells.* Chargers are made in many forms and are designed for either direct or alternating current. Direct current chargers consist of certain resistances, with the necessary meters and switches. A simple form of direct current charger is shown in the illustration. In constructing a direct current charger there are several important details that must be

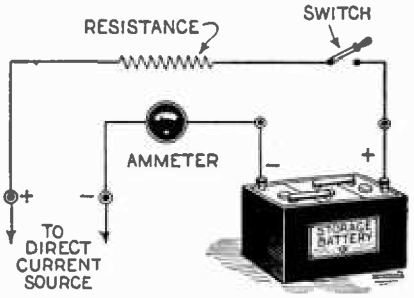

Charging storage battery from direct current source.

taken into consideration. The plates of a storage battery require certain charging rates in amperes, depending on the size and number of the plates. That is to say, the manufacturer specifies that when charging, a certain number of amperes must be passed to the battery from the charging source for a definite number of hours. This figure may be decreased but not increased. As an example, some storage batteries will take a charge of five amperes without injury to the plates, which rate would be ruinous for other batteries. Now the question of obtaining the proper amount of charging current arises. The majority of house lighting systems have a voltage of from 110 to 125 volts. (Reference is now made only to direct current systems, alternating current chargers being taken up later on.) According to *Ohm's Law* (q.v.) the current flowing in a circuit is directly proportional to the *voltage* and inversely proportional to the *resistance*. Voltage is the product of resistance times the current, the resistance in ohms and the current in amperes. Therefore if we desire a current of five amperes from a source of 110 volts, it is necessary to have a resistance of a value in ohms which, when multiplied by the current in amperes,—five in this case—will produce 110. Conversely, we can divide the voltage—110 by the current—5 to obtain the correct resistance. This would indicate a resistance of 22 ohms. There is another factor which enters into the problem however, the back pressure, or Counter Electromotive Force exerted on the charging source by the battery. In the case of a six volt storage battery, this will be six volts. This must be subtracted from the voltage of the charging source to obtain the effective value. Thus, the effective value in this case would be 110 — 6 or 104 volts. Then as explained before, dividing the voltage by the current required, we find that it is 104 ÷ 5 or approximately 21 ohms. Then by placing a resistance of 21 ohms in the circuit, as shown in the illustration, and connecting the battery as indicated, a charge of five amperes can be applied to the battery. The same result can be obtained by placing an ordinary electric lamp in series in place of the resistance. (See *Lamp Bank.*) The *internal resistance* of the battery is very small compared to that of the line resistance, and can be ignored. A point to remember is that the voltage of the charging source must always be higher than that of the storage battery, because, as explained above, the back or *counter electromotive force* of the battery, must be subtracted from the voltage of the charging source to indicate the effective voltage. Thus, if the battery exerts a counter electromotive force of six volts and the charging source is only six volts the counter force would be equal to the charging force and there would be no charging current. When *direct current* is not available and *alternating current* is used to charge the battery, it will be necessary to change it to direct current. For this purpose it is essential that some form of *rectifier* be used. This rectifier may be of a chemical, mechanical or electrical type. These various forms will be taken up under their separate headings. (See *Rectifier, Balkite, Tungar, Rectigon, Electrolytic Rectifier, Tube Rectifier.*)

CHARGING—The act of expending electrical energy to place a charge in or on a battery or body. The term is used generally to indicate the process of impressing electrical energy on the *storage battery* elements for the purpose of making the battery a secondary source of electricty. (See *Charger, Chemical Rectifier* also *Tube Rectifier.*)

CHARGING BOARD—A term used synonymously with *charging panel* to denote an arrangement for controlling the current used for charging *storage batteries.* It generally consists of an insulated panel with various *switches, fuses* and *resistances* thereon to limit the amount of current delivered, and at the same time furnish a ready means of connecting and disconnecting the various circuits connected through it. The fuses act as safety devices and break the circuit when too great a current is delivered. (See *Fuse, Switch, Charger.*)

CHEMICAL RECTIFIER—Device for changing alternating current to direct current. (*See Electrolytic Rectifier.*)

CHLORIDE BATTERY A type of storage battery for heavy duty operation. The positive plate is made of lead with a number of holes pierced in it and each filled with a coil of pure lead. This coil is forced into the holes during manufacture and is said to enable the battery to withstand heavy discharge without damage to the plates. (See *Storage Battery.*)

CHLORIDE CELL—A cell (q.v.) which uses in most cases, a positive plate having lead as the active material, the negative plate being made of metallic zinc. A solution of chloride of zinc is used, from which the name is derived. (See *Chloride Battery.*)

CHOKE OR CHOKING—The action of a device in a circuit wherein it tends to choke, or hold back, certain forms of current or certain *frequencies* while permitting others to pass freely, or to oppose fluctuations in the strength of the current passing through it in the circuit.

CHOKE COIL—A coil of wire possessing considerable *self-inductance* and relatively little *resistance.* These coils are used in various ways and for a variety of purposes. When connected in a direct current circuit, and possessing the proper value, they have a tendency to prevent any fluctuations in the current and thus keep the current smooth. An example of its use is with a *loudspeaker.* A *condenser* will permit passage of *high frequency (alternating) currents* such as we find in radio reception, but will effectively block direct currents. On the other hand a choke coil will pass direct current without offering any appreciable resistance, but will retard the *modulated currents* (q.v.) It will be apparent then that by a proper combination of the two (Choke Coil and Condenser) as in the illustration, the direct currents can be shunted around the loud speaker and thus prevent injury to it, while the necessary modulated currents are allowed to pass through without difficulty. Another use for the Choke is in a circuit carrying alternating currents. In this case the choke will limit the current in the circuit without loss, whereas an ordinary resistance, while it would limit the current, would dissipate a certain amount of energy. Chokes may be made with iron cores of the open or closed type, for *low frequency currents*, or may only have air for the core when used in *high frequency* circuits. The closed core type is the more common form. The formula most commonly used in determining the inductance of a choke coil in *henries* is as follows:

$$L = \frac{1.257 \times \mu \times A \times N^2 \times 10^{-8}}{l}$$

Where L is the inductance in henries.
μ is the permeability. (This varies from about 1000 to 2000 in most cases).
A is the effective area of iron cross section (square centimeters).
N is the number of turns of wire.
l is the length of iron path (in centimeters).

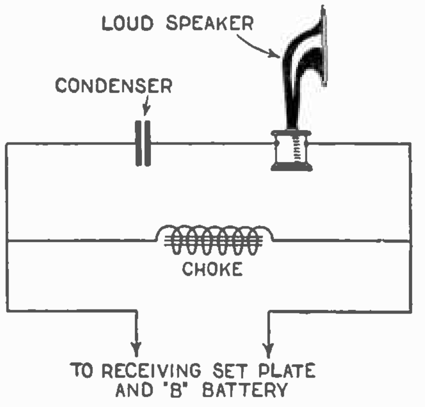

Method of using choke coil in connection with loud speaker.

CHOKE, HIGH FREQUENCY — A choke coil of high *self-inductance* (q.v.) used in transmitting circuits to prevent puncture of the secondary windings of the *transformer* (q.v.). When the condensers in a transmitting circuit are charged or discharged, a back pressure is exerted on the windings of the transformer supplying the high voltage. The illustration shows the position of these choke coils in the transmitting circuit. (See *Back Oscillation.*)

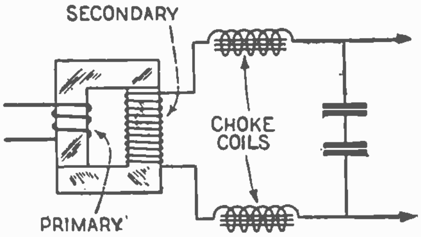

Position of choke coils in transmitting circuit.

CHOPPER—A device used in the aerial circuit of a *continuous wave* (q.v.) transmitter to break up the continuous trains of waves into separate groups and thus permit them to become audible. In the case of reception of ordinary *damped waves* (q.v.) the head phone produces signals that have a frequency of vibration equal to the number of wave trains per second. When the transmitted waves are con-

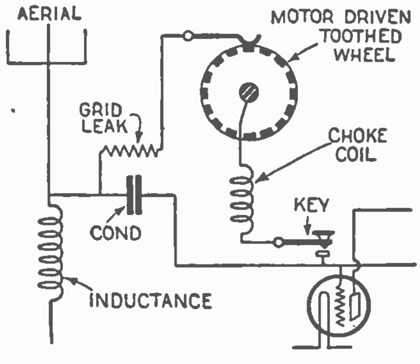

Showing how chopper is used to interrupt signals from continuous wave transmitter.

tinuous there is obviously no interval, that is, there are no groups of waves and they are therefore inaudible in the usual receiving apparatus. The *chopper* is only necessary when *crystal detectors* are used for reception. (See *Ticker, Beat Reception and Continuous Wave Transmission.*)

CIRCUIT—The path followed by an electric current passing from its source through a wire conductor or series of such *conductors*, and back again to its starting point. A *closed circuit* is one that is continuous and will permit the passage of current. An *open circuit* is one that is not continuous and, being broken or open at some point, will not permit passage of the current. The term is used very commonly, and broadly covers any system for the conducting of electric current from its source, through some electrical instrument or appliance and back to the starting point.

CIRCUIT BREAKER—Generally speaking, any device for automatically breaking or opening a circuit and thus preventing flow of current through the circuit. More specifically, it may be a device similar to a *switch*, used to automatically open the circuit when the current falls below a certain point, or rises above a certain fixed value. Thus an *overload circuit breaker* is an arrangement which will automatically open the circuit if the current becomes too great. This is shown in Fig. 1. It is a simple form and consists of an *electro magnet* which requires a certain force to operate it. When the current in the circuit reaches a certain point, the magnet operates, pulling down the bar "A" against the action of the spring "P," and thus breaking the circuit. As long as the current is not more than that fixed by the requirements of the circuit—the magnet being arranged in accordance with the particular requirements — the current flows, the breaker operating to open the circuit only when more than the desired current is introduced into the circuit. In the illustration we can consider "S1" and "S2" as the source of current, "M" as an electro-magnet in series with the motor "D," and "A-B" a switch, in which "A" has attached to it a piece of soft iron. Now if the current required for the motor is ten amperes, the magnet can be so designed that it will not act on the bar "A" until 15 amperes flow through it. Thus, as long as the current is less than this, the circuit is continuous, but if the current is increased to 15 amperes, the magnet operates and attracts the bar "A," breaking its contact with "B" and thus opening the circuit. In Fig. 2 a simple *underload breaker* is shown. In this case the operation is the reverse of that of the *overload breaker*. Here the bar "A" is held down in contact with "B" by the magnet "M" thus keeping the circuit closed through the motor "D". Now assume that the motor requires ten amperes and the magnet requires the same amount to hold the bar "A" in place against "B". If then the current in the circuit falls below ten amperes, the magnet can no longer hold the bar "A" against the action of the spring "P" and it rises, breaking the contact between "A" and "B" and thus opening the circuit. These circuits are of course only assumed for the purpose of illustrating the theory of circuit breakers. Actually the system is generally much more elaborate. Circuit breakers are used for many purposes—chiefly to prevent any sudden increase or decrease of current in a circuit, and to control the flow of current for charging batteries. (See *Charger-Storage Battery.*)

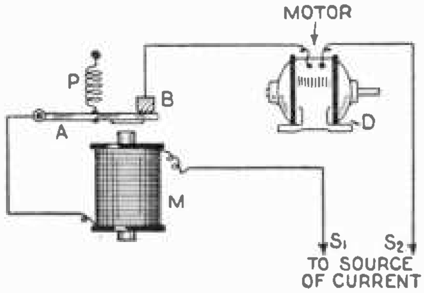

Fig. 1. Overload Circuit Breaker.

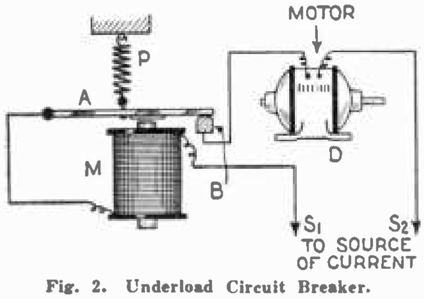

Fig. 2. Underload Circuit Breaker.

CIRCUIT, BROKEN—A circuit that is not continuous, having been opened either purposely, or due to some defect in the circuit. The term is usually applied where the circuit is opened by the breaking of a wire, or through a faulty contact.

CIRCUIT, OPEN — A circuit which is open and through which current cannot flow, or a circuit which is normally open but which can be closed at will by pressing a key or button. A common example of open circuit is in the case of an electric bell circuit. The illustration shows such a circuit. Normally it is open and current does not flow to the bell, but on pressing the button B, the circuit is closed and current passes to the bell. (See *Circuit*, also *Closed Circuit.*)

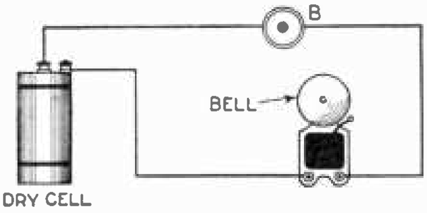

Open circuit which can be closed by pressing button "B".

CIRCUITS, PARALLEL — Circuits starting at a common point and ending at a common point. The illustration shows three *parallel circuits*. Actually each one is a part of the whole circuit. (See *Parallel Resistances.*)

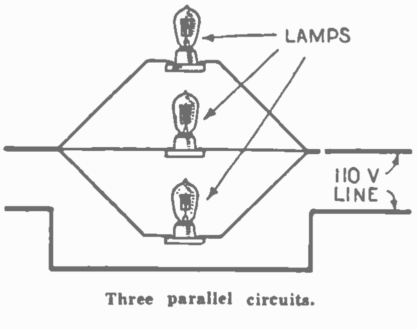

Three parallel circuits.

CIRCULAR MIL—The cross-sectional area of conductors (wires) is usually designated in terms of circular mils. A circular mil is the area of a circle having a diameter of one thousandth of an inch. Thus a wire having a diameter of one quarter of an inch is said to have an area of 250 circular mils. (See *Square Mil.*)

CLEAT—A form of porcelain *insulator* used in house wiring. They are generally arranged in two parts, with corresponding grooves in which the wire is held, and are nailed or otherwise fastened to the wall. Cleats are about the cheapest form of insulator and where the voltages used are low, are quite as effective as the more elaborate types. Cleats are much used in radio work, both for insulating lead in wires coming from the aerial, and for insulating the aerial itself.

CLERK-MAXWELL, JAMES—Born at Edinburgh, Scotland, in 1831, and educated at Edinburgh and Cambridge University, he became Professor of Natural Philosophy at Aberdeen, 1856-69, and Professor of Physics and Astronomy at King's College, London, 1860-65. In 1871 he became the first holder of the new chair of experimental physics at Cambridge, where he died, November 15, 1879. Clerk-Maxwell was recognized in his later years as one of the greatest authorities on physics of his time, and his fame has been steadily increasing since his death. Radio owes Clerk-Maxwell a deep debt. Electricity was the chief study of his lifetime, and his first important paper on the theory of *electromagnetism* was communicated to the Royal Society in 1867. In 1873 he published his "Electricity and Magnetism," a work on the subject which has never been surpassed. In it he formulated his famous *electro-magnetic theory* of light and his theories on electric waves, which developed into the modern system of wireless telegraphy and telephony through the experiments of Hertz.

James Clerk-Maxwell.

CLIMAX—A high resistance nickel steel alloy, used extensively in *rheostats* (q.v.). It is one of the cheapest and most practical alloys for low-temperature resistance wire. Its chief properties are as follows: Maximum working temperature 540 degrees centigrade, resistance in microhms per cubic centimeter at 20 degrees centigrade, 87.2; temperature coefficient per degree centigrade, .00054, and having a specific gravity of 8.14.

CLIPS—A clamp connector used for various purposes in radio receiving and transmitting circuits. (See *illustration.*) In a receiving set a clip may be used at the end of a flexible wire for the purpose of varying the connection to the various "B" battery taps; clips are very frequently used in making the connections to the storage "A" battery, and for many other purposes about the receiving set. One of their principal uses in the transmitter is in connection with the leads to the oscillation transformer. Flexible leads with clips attached come from the other units of the transmitter and are clamped at will along the various turns of the oscillation transformer until the point giving the desired wave length or highest radiation reading is found.

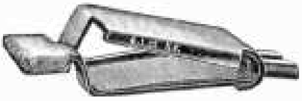

Type of clip in general use.

CLOCKWISE—A term used to indicate that the rotating part of an electrical machine or instrument moves from left to right (when facing it), or as the hands of a clock. In the illustration the needle or pointer of the meter moves in a clock-wise direction. In an electrical measuring instrument this is always true where the zero point is at the left part of the dial. The opposite is known as *counter-clockwise.* When looking at the *armature* of a motor or generator, if the rotating part moves from left to right it is said to have a clock-wise motion.

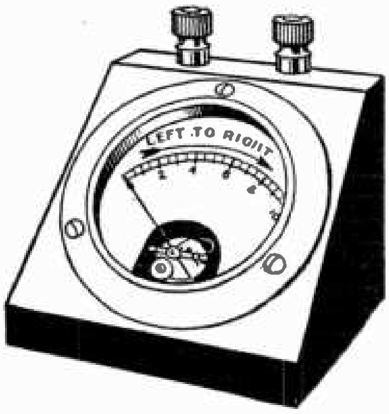

Indicator needle moves from left to right in clockwise motion.

CLOSE COUPLING The arrangement of two coils acting as *primary* and *secondary,* placed close together in such manner that the coupling or electrical relation between them is close. In the illustration the secondary coil is shown entirely within the primary. In other words the transfer of energy will be large in this case as the coupling is close. If the secondary is withdrawn a certain distance from the primary, the effect is referred to as loosening the coupling, and the coils are then said to be *loose coupled,* the transfer of energy from one to the other being considerably less than in the case of close or *tight coupling.* (See *Coupling.*)

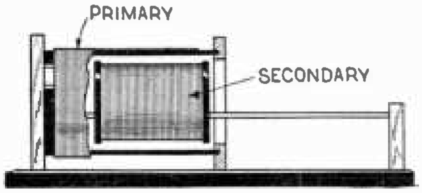

Loose coupler with secondary placed to produce close coupling.

CLOSED CIRCUIT—A circuit which permits the continuous flow of electricity. The illustration shows an ordinary dry cell connected in circuit with a small lamp. In Fig. 1A the circuit is closed and current is permitted to flow to the lamp. In Fig. 1B the circuit is open and no current can flow. A closed circuit is therefore an electrical circuit that is continuous. (See *Circuit,* also *Open Circuit.*)

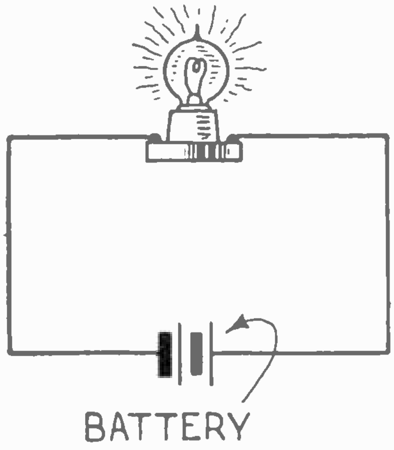

-1A-

Circuit closed.

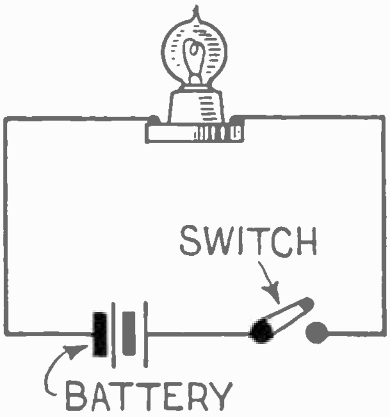

-1B-

Circuit open by means of switch.

CLOSED CIRCUIT CELL—A cell that can be used in a closed circuit for a considerable length of time without the impairing effects of *polarization* (q.v.) An *open circuit* cell is one that can only be used intermittently, as for example, with an electric bell where the circuit is only closed for a brief period. If an open circuit cell is used in a circuit that remains closed for some time, polarization sets in and ruins the cell in a very short time. An open circuit cell might be used with the *test buzzer* (q.v.) because in this case the circuit is closed intermittently as the key is depressed. (See *Cell.*)

CLOSED CORE—A core (q.v.) used in *transformers, chokes,* etc., which is continuous, forming a *closed magnetic path.* The illustration shows the difference between a closed core and an open core. In the open core type, one end is the north pole, the other the south pole, whereas in the closed core, there are obviously no poles. The closed core type is occasionally referred to as a *non-polar* core. (See *Core, Transformer,* also *Choke.*)

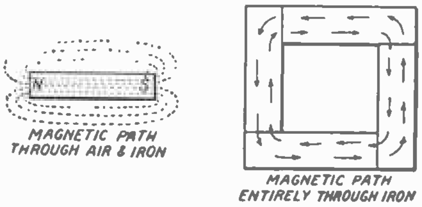
N
S
MAGNETIC PATH THROUGH AIR & IRON
MAGNETIC PATH ENTIRELY THROUGH IRON

Open core above on left. closed on right.

CLOSED CORE TRANSFORMER—A *transformer* (q.v.) having a *closed core.* In the illustration the magnetic path, or core, is continuous, being made of strips of uniform size. This form of core is known as *"laminated."* An efficient transformer is generally made with a closed core. A transformer may be for a variety of purposes, such as *step-up, step-down, audio frequency* or *radio frequency.* (See *Transformer.*)

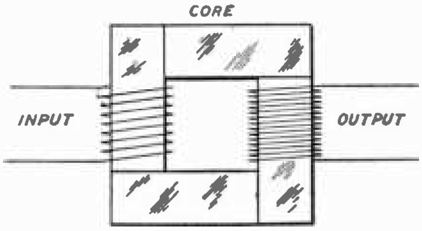

Closed core transformer.

C. M.—Abbreviation for circular mil. (q.v.)

Cm.—The abbreviation for *centimeter,* the unit of length in the C. G. S. system (q.v.).

COCKADAY CIRCUIT—A popular radio receiving circuit devised by the American experimenter, Lawrence M.

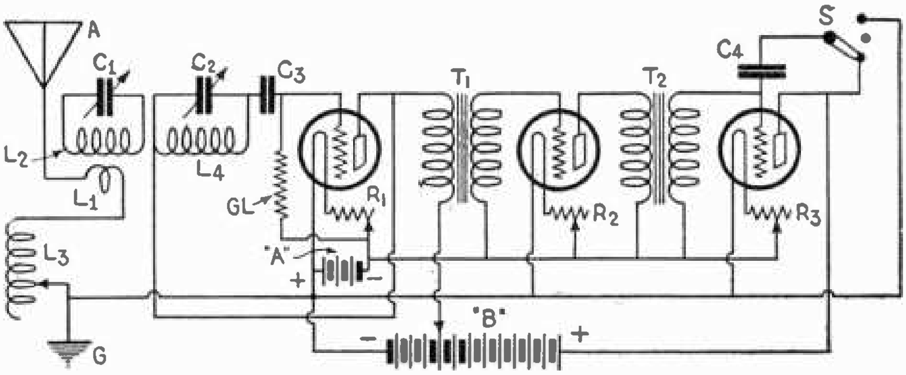

Fig. 1. Original 3 tube Cockaday Circuit.

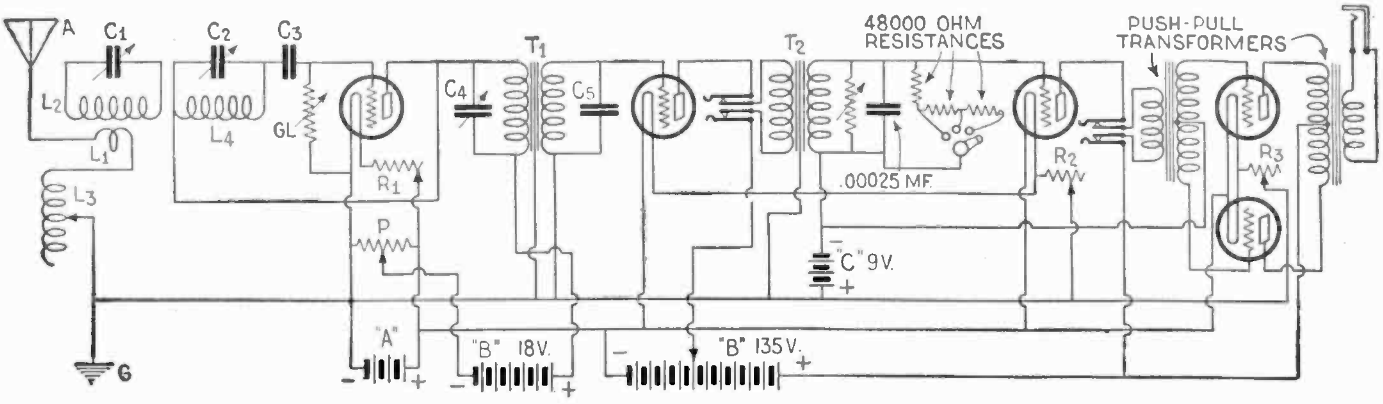

Fig. 2. Cockaday Circuit with push-pull amplifier.

Cockaday. It includes several radical features not found in the usual systems. Receivers using this circuit are generally made in three, four or five tube types. The chief value of this circuit lies in its extreme selectivity. It is usually referred to as a four circuit receiver, owing to the addition of a sensitizing circuit to the regular three circuit regenerative system. The diagram for the three tube receiver is given in Fig. 1. Here the primary or aerial tuning circuit is seen to have a single turn of wire, L1, wound around another coil L2, which is shunted by a variable condenser C1, thus forming a separate closed, tunable circuit, coupled very loosely with the antenna circuit. The primary, or antenna circuit, has in addition to the single turn, a bank wound coil L3 connected in the ground side of the open circuit and tapped to permit coarse tuning. Thus, the primary is coupled inductively to the separate closed circuit, which in turn is inductively coupled to the grid or secondary circuit containing a coil L4 shunted by the condenser C2, allowing tuning.

The tone of the music is varied by means of the novel arrangement using a condenser C4 connected to a three point switch. When the switch is placed in the center position the condenser is eliminated from the circuit and the wiring is standard. With the switch on the top tap, the grid of the tube is connected to the ground, thus by-passing a portion of the oscillations and giving a smooth soft tone with a certain loss of volume. With the switch arm on the bottom tap, the grid of the last tube is connected to the plate circuit, thus producing a different note due to the added grid-plate capacity. Coil L2 in the grid circuit is tuned by means of C1, which is a variable condenser of .0005 mfd. capacity.

A peculiarity of the circuit is that the length of the antenna has no bearing on the tuning. This is due to the *semi-aperiodic primary* and the fact that the sensitizing or pick-up coil is only one turn of wire, the inductance of which, compared to the whole inductance of the aerial circuit, is so small that any change in the over-all inductance has little, if any, effect.

This circuit, while originally having three tubes, has been arranged in conjunction with a standard *push-pull audio amplifier*, as shown in Fig. 2, making a five tube set with great power and possessing the characteristic of the original circuit with increased amplification. (See *Push-pull, Cockaday Coil* also *Regenerative Receiver*.)

COCKADAY COIL—The coil used in the *Cockaday circuits*. It comprises four windings, i.e. tapped primary, separate closed circuit coil, grid or secondary winding and a single turn looped around the closed circuit coil. (See *Cockaday Circuit*.)

CODE—Telegraph codes consist of characters formed by combinations of dots, dashes and spaces which represent letters, numerals, and punctuation marks. These characters are sent out by the radio operator with the aid of electrical impulses, and by using suitable receiving apparatus the receiving operator hears the incoming signal and is thus able to interpret the various combinations of dots and dashes and reproduce the original message. The Morse Code of characters came into general use in wire telegraphy shortly after the establishment of that means of communication. In this system, which is also called the "American Morse Code," some of the characters are made up with so-called spaces which are part of the group signal, and are necessary in distinguishing those characters. The International Morse code is a modified form of the American Morse code, and no spaces are used in the characters of the International Code. The International Morse Code is used all over the world for radio telegraphy, and for wire telegraphy in almost every country except the United States, Canada and parts of Australia. The American Morse, owing to the fact that there are fewer dashes in the characters, is about

LETTERS	MORSE	CONTINENTAL
A	• ▬	• ▬
B	▬ • • •	▬ • • •
C	• • •	▬ • ▬ •
D	▬ • •	▬ • •
E	•	•
F	• ▬ •	• • ▬ •
G	▬ ▬ •	▬ ▬ •
H	• • • •	• • • •
I	• •	• •
J	▬ • ▬ •	• ▬ ▬ ▬
K	▬ • ▬	▬ • ▬
L	▬▬	• ▬ • •
M	▬ ▬	▬ ▬
N	▬ •	▬ •
O	• •	▬ ▬ ▬
P	• • • • •	• ▬ ▬ •
Q	• • ▬ •	▬ ▬ • ▬
R	• • •	• ▬ •
S	• • •	• • •
T	▬	▬
U	• • ▬	• • ▬
V	• • • ▬	• • • ▬
W	• ▬ ▬	• ▬ ▬
X	• ▬ • •	▬ • • ▬
Y	• • • •	▬ • ▬ ▬
Z	• • • •	▬ ▬ • •
&	• • • •	
1	• ▬ ▬ •	• ▬ ▬ ▬ ▬
2	• • ▬ • •	• • ▬ ▬ ▬
3	• • • ▬ •	• • • ▬ ▬
4	• • • • ▬	• • • • ▬
5	▬ ▬ ▬	• • • • •
6	• • • • • •	▬ • • • •
7	▬ ▬ • •	▬ ▬ • • •
8	▬ • • • •	▬ ▬ ▬ • •
9	▬ • • ▬	▬ ▬ ▬ ▬ •
0	▬▬▬	▬ ▬ ▬ ▬ ▬
.	• • ▬ ▬ • •	• • • • • •
,	• ▬ • ▬	• ▬ • ▬ • ▬
:	▬ • ▬ • •	▬ ▬ ▬ • • •
;	• • • • •	▬ • ▬ • ▬ •
?	▬ • • ▬ •	• • ▬ ▬ • •

Morse and Continental Codes.

5% more rapid than the International Morse. The American Morse and International (Continental) Codes are given herewith.

COEFFICIENT—In algebra, a number or multiplier of a *symbol.* Thus in the expression 5Y, the coefficient is 5. In electricity, and in its radio application, it is a number expressing the relation or ratio between quantities, one of the quantities generally being unity (1) as a basis or standard for the coefficients. For example, let us assume a copper wire is carrying an electric current, the *resistance* of the wire at 32 degrees Fahrenheit being one *ohm.* If we consider the resistance of the wire at 32 degrees to be unity, or one, then its resistance at any other temperature will be some number or coefficient times one.

To make this clearer, let us take the ratio of an inch to a foot. An inch is one-twelfth, or approximately .0833, of a foot. Any number of inches multiplied by this fraction will give another number in feet. Therefore we say that the coefficient, or multiplier, of an inch as compared to a foot is .0833. A coefficient is thus any number by which a quantity must be multiplied to give another quantity.

In radio the most common cases of coefficients are the following. The *coefficient of coupling* is the ratio between the *mutual inductance* (q.v.) (between the two coils, *primary* and *secondary*) and the square root of the product of the total *self-inductances* (q.v.) in the two circuits containing the coils. Similarly, the coefficient of amplification is the ratio of the effect produced with an *amplifier* (q.v.) to the effect produced without an amplifier. (See *Mutual Induction Coefficient,* also *Self-induction Coefficient.*)

COHEN, LOUIS, Ph. D.—Born in Russia, Dec. 16, 1876. Educated at Armour Institute of Technology, University of Chicago and Columbia University. Noted for his formulae and tables for the calculation of alternating current problems. He has written a book on the subject. Has evolved many formulas for the measurement of inductance and capacity. Has secured a number of patents on radio apparatus.

COHERER—A device formerly employed for detecting *radio waves.* In its original form it consisted of two metal rods separated by a narrow space in a glass tube and the gap between the rods filled with nickel and silver filings. It was found that when the *electromagnetic waves* or signals were passed through the coherer the particles of filings clung together and became *conductors.* Then by using some device such as the tapper hammer of an electric bell to strike the glass tube, again separating the particles after each signal, it was possible to keep the coherer constantly in a state of change between a *conductive* and a *non-conductive* condition, thus making the signals audible. There were many variations of this type of *detector,* which, however, is now obsolete. (See *Marconi Coherer.*)

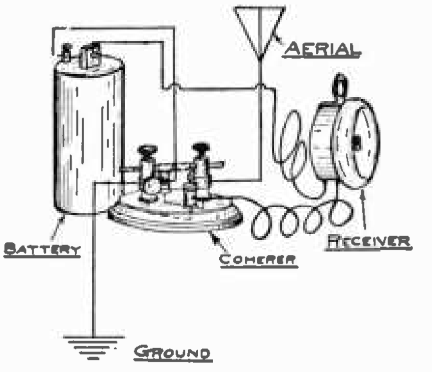

Coherer used in radio receiving circuit.

COIL—A term having very broad application in electricity and radio. It may be used to designate any arrangement of a number of turns of wire, usually copper, and used for almost any purpose in the production of electric and magnetic effects or phenomena. The most common significance of a "coil" in radio is in reference to a number of turns of insulated wire wound on various forms and used for *tuning* (q.v.). The term *inductance coil* is generally used, although actually, any coil of wire will have inductance. Coils come in countless numbers and variety for many purposes. A coil may be used not alone for tuning—that is to adjust a receiver or transmitter to resonance with incoming or outgoing signals; it may be used to connect two circuits, in which case it is known as a *coupling coil;* it may be used with a *core* (q.v.) of iron wire, or laminated pieces, as a *choke coil* (q.v.) or it may be used in connection with *high frequency* apparatus, such as used by physicians. Coils are generally referred to by individual names indicating their construction or use. (See *Inductance, Tuning Coil, Vario-coupler, Choke Coil, Honey Comb Coil, Pancake Coil, Spider Web Coil, Lorenz Coil, Duo Lateral Coils.*)

COIL ANTENNA—Or Loop Aerial. An antenna (aerial) constructed by winding the wire around a square form. The coil (or loop) antenna, because of its comparatively small size, is often mounted on a frame work which may be rotated. If one end of the coil is pointed toward the broadcasting station, the antenna (aerial) will pick up its signals with maximum intensity, but if the coil is rotated from that position the signals from that station grow weaker, until when the plane of the loop is at right angles with the intercepted signals no sound will be obtained. It is a part of the radio compass used at sea. (See *Loop-Aerial.*)

COLLECTOR RINGS—Insulated metal rings attached to the *armature* (q.v.) of an *alternator* (generator of alternating current) in such fashion that the alternating currents can be collected and communicated to the *brushes* without change, these rings are used to collect alternating current from that type of generator and brushes are used to collect direct current from a direct current generator.

COLLODION—A solution used by surgeons to cover slight abrasions or wounds, forming a protective coating similar to skin. It has been adapted to radio use as an insulating coating and is used as a binder on coils of various forms. It has been found more efficient than shellac for this purpose. A solution of pyroxylin or soluble guncotton in ether, with a certain amount of alcohol, depending on the purpose for which intended. When used as a coating for coils it serves as an insulator and also to hold the coils in place, but it should be sparingly applied.

COMBINATION DETECTOR—A detector composed of two different crystals in contact, instead of one crystal with a fine wire to search out the sensitive spots. Some combination detectors are composed of silicon and antimony; zincite and bornite; lenzite and cerusite, and zincite and chalcopyrite. These detectors all function without a battery and are simple to operate.

COMMON CONNECTIONS—A term often applied to a connection that joins several points. For example, in most radio receivers the binding posts for battery connections are arranged so that the "B—" and the "A+" or "A—" are connected to the same point. That is to say the post is a common connection for these leads. The term is also applied to a battery, as a *"B" battery,* which is used for both *detector* and *amplifier.* It is then known as a common "B" battery. The other method would, of course, be to use separate batteries for the detector and amplifier.

COMMUTATOR—In general electrical use, a device for reversing the direction of electric currents in a circuit. A *commutator* as used with a *generator* usually consists of a number of pieces called *segments,* mounted on a circular form but carefully insulated from each other. The illustration shows a common form. In the case of an ordinary motor, the commutator is employed to distribute the current to the windings and in a direct current generator it serves to furnish a direct current to the brushes although the current collected from the windings of the *armature* is alternating.

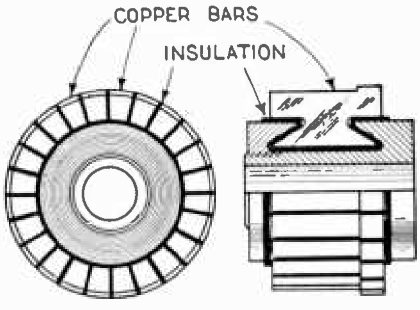

Front and side view of Commutator.

COMPASS, RADIO — See *Direction Finder, Goniometer,* also *Bellin-Tosi Direction Finder.*

COMPONENT—A part of the whole. In mechanics it is one of the parts of a stress or strain (mechanical force). In electricity, *component currents* are the several currents into which a single current may be assumed to be separated, so that if assumed to be acting together they would produce the equivalent effect of the actual current. Thus also, component *E. M. F.'s* (*electro-motive forces*) are understood as the several components into which any electro motive force may be divided. *Impedance* (q.v.) is said to consist of two components, the actual resistance and the apparent resistance present in the opposition offered to the flow of current termed the *reactance* (q.v.) by a circuit. For instance, the total current in the plate circuit of a vacuum tube responding to high *frequency grid voltages* may be considered as made up of two components, viz., a steady or constant current and a high frequency component superimposed on it.

COMPRESSED AIR CONDENSER—A condenser which uses compressed air as the *dielectric* or insulating material between the plates. This type of condenser is used chiefly in transmitting circuits. (See *Condenser.*)

COMPOUND MAGNETS—A combination of several *permanent magnets.* (See *Magnets, Compound.*)

COMPOUND WOUND—A *dynamo* or *motor field magnet* wound with two *field windings,* one of which is connected in series, the other in parallel with the *armature* (q.v.).

CONDENSER—In electricity or radio practice a *condenser* consists of two or more conducting surfaces placed in relation to each other and separated by an insulating medium such as air, impregnated paper, mica, etc. The conducting surfaces are generally referred to as *plates* or *electrodes* (q.v.) and the insulating material is known as the *dielectric* (q.v.). Condensers are used in radio for a variety of purposes, their broad use being to allow tuning. In action, a condenser permits charges of electricity known as *electrostatic charges* (q.v.) to be stored up and released periodically, according to the frequency desired. They may also be used to exclude certain types of energy. (See *Condenser, By Pass,* also *Filter.*) The action of a condenser may be compared to that of a spring when placed under a mechanical strain. In the illustration Fig. 1A a flat spring S is shown in a normal position, NB. In 1B a weight W is placed on the surface of the spring. This weight will force the spring down from its normal position to one of the positions (a), (b), (c) or (d), according to the force exerted and the strength of the spring. If a heavier weight is then placed on the spring it will depress still further until we can imagine a point P in Fig. 1C at which the elasticity of the spring is used up and the opposition is such that it will no longer depress and comes to rest. When the weight is removed the spring will fly back into normal position, Fig. 1D, and in doing so will vibrate back and forth a number of times, thus producing, or rather, returning the energy that was stored in it during the moment of stress.

Considered as analagous to a *condenser,* in the above case the weight can be likened to the electrical charge applied to the condenser, the *resistance* of the spring may be considered somewhat similar to the *capacity* of the condenser, or the power returned on release of the spring may be regarded as the electrical power released by the condenser when the charging force is removed. The capacity of the condenser then will be seen as the ability to store up energy placed in it.

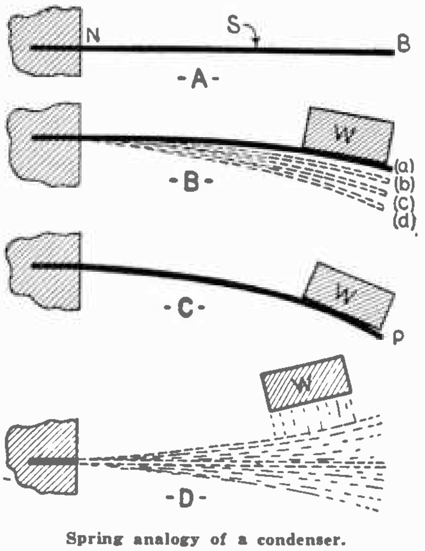

Spring analogy of a condenser.

In the case of the spring, if too great a force is exerted, it may break down or as would actually happen, the elasticity would be lost. The same applies in the case of a condenser. If the charging force is too great, it may exceed the dielectric strength of the condenser and result in puncturing the insulating material between the plates. (See *Breaking Down of Dielectric.*) The energy obtainable from a charged condenser will be equal to the charge placed in it providing there is no leakage or other loss. In a well-designed condenser the leakage will be so negligible that it need not be taken into consideration.

The relation between the charge in a *condenser,* the *voltage* applied and the *capacity of the condenser* is expressed as $Q = C \times E$, where Q is the quantity of the charge in *coulombs,* E the potential in *volts* and C the capacity in *farads.* (See *Capacity, Condenser, Variable; Condenser Fixed; Condenser Transmitting,* also *Condenser Curves.*)

CONDENSER, AIR—See *Air Condenser.*

CONDENSER, ANTENNA—See *Aerial Tuning Condenser.*

CONDENSER, BY-PASS—A condenser, usually of the fixed (q.v.) type arranged in a circuit in such manner that currents of a certain frequency will pass freely around any obstructing object, as, for instance, a high resistance element necessary to the circuit. An example of the use of a condenser in this manner is shown in the case of a "B" battery where the battery has a high internal resistance. A fixed condenser is placed across the terminals of the battery without injury to it and offers a path of low resistance for the high frequency radio currents.

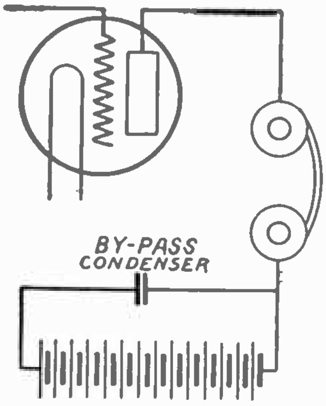

By-pass condenser across "B" battery.

CONDENSER CURVES—Curves showing the relation of variations in wavelength, frequency, and capacity with rotation of the dial of a variable condenser. Three types of condensers are shown here, the *compression,* Fig. 1, the *straight line capacity,* Fig. 2, and the *square law,* Fig. 3. In Fig. 1A is shown the relation of capacity and wavelength changes when the dial of a sample compression condenser was rotated. The condenser was shunted across a fixed coil. It will appear that the increase in capacity is gradual at the lower part of the range but increases very rapidly at the upper ranges. This is due mainly to the relation of the curvature of the upper plate to the lower plate as will be seen from the illustration, Fig. 1. These condensers are very satisfactory on about the lower two-thirds of their range, but through the remainder of their range a very slight adjustment of the knob causes an extremely large change in capacity, making them very difficult to handle near their maximum capacity. Fig. 2 shows the shape of the variable plates in the straight line capacity condenser and Fig. 2A illustrates the relation of wave-length and capacity changes with rotation of the dial. Here the capacity depends upon the overlapping area and is therefore proportional to the angle of rotation of the movable plates. The capacity curve appears as a practically straight line running diagonally across the chart. The wave-length, however, changes rapidly at the lower range and very gradually at the upper range as indicated by the wave-length curve. Fig.

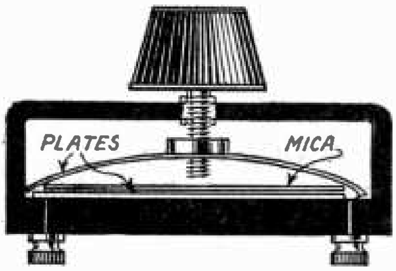

Fig. 1. Compression type condenser.

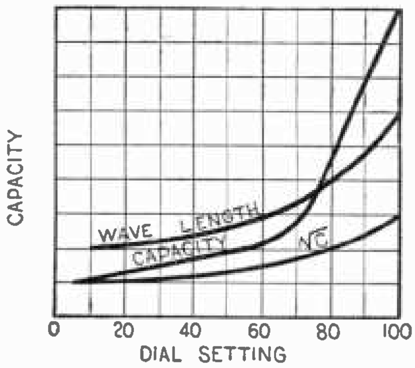

Fig. 1A. Curve for compression type condenser.

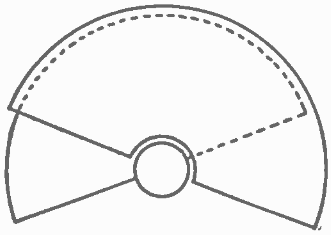

Fig. 2. Straight line condenser.

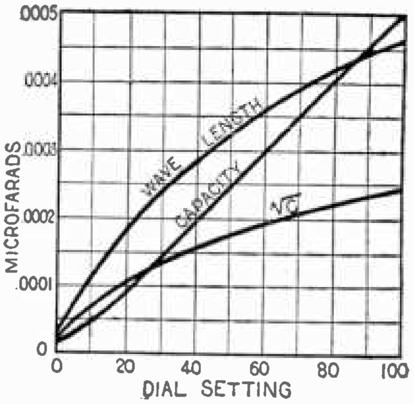

Fig. 2A. Curve for straight line condenser.

3 shows the shape and disposition of rotor and stator plates of a condenser so designed as to read directly in wavelength from the settings of the dial.

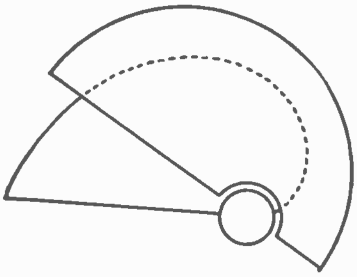

Fig. 3. Square law condenser.

This is known as a square law condenser, and the relational curves are shown in Fig. 3A. Here we find the wave-length curve a straight line and directly proportional to the angle through which the dial is turned. The line indicating the square root of the capacity √C is a straight line, and as the wavelength is the product of the square root of the capacity multiplied by a constant (varying with the individual condenser) the wavelength curve must also be a straight line.

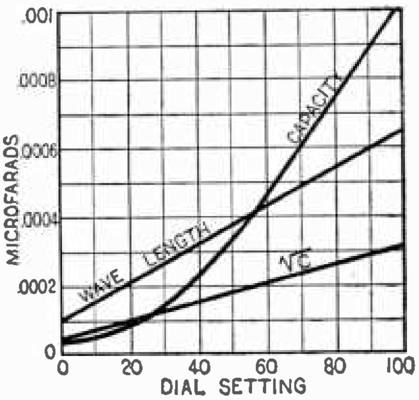

Fig. 3A. Curve for square law condenser.

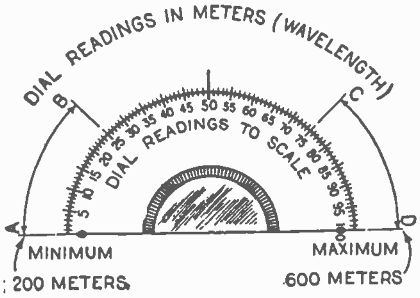

Fig. 3B. Showing effect of straight line condenser on dial reading.

Such condensers are much used in wave meters (q.v.). Condensers are also made so that the frequency curve is a nearly straight line, varying more or less directly as the angle of rotation of the plates. Fig. 3B illustrates the result when an ordinary straight line capacity condenser is arranged with a dial calibrated directly in meters. It appears that the readings are bunched at the lower range (A to B) and spread out at the upper ranges (C to D). The reason for this can readily be seen by glancing at the curve Fig. 2A.

CONDENSER, FIXED—A *condenser* which has a fixed or constant capacity, allowing no variation of the value. (See *Condenser*, also *Condenser, By-Pass.*)

CONDENSER, GRID—A condenser, generally of low capacity, of the fixed or variable type, used in series with the grid of the *detector tube* to prevent excessive charges from affecting the *grid.* These condensers generally have a capacity between .00025 and .0005 *micro farads.* Due to the blocking action of these condensers, a high resistance, known as a *grid leak* (q.v.) is placed across the condenser or from the grid to filament, thus allowing the accumulated charges to leak off the grid of the tube in time for the grid to be free for the succeeding electric charge. The grid leak may be considered as a valve controlling the amount of energy that the detector can efficiently take care of without overload. (See *Grid Leak.*)

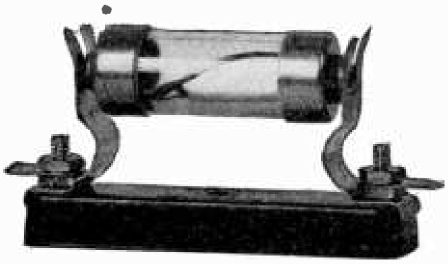

Grid leak and condenser.

CONDENSER, IN PARALLEL—A condenser arranged in a circuit in such manner that the current is divided between it and the other condenser or coil with which it is connected. The illustration shows a *variable condenser* connected in parallel with a coil, both placed in series between aerial and ground and forming a tuned primary or aerial circuit. (See *Parallel Connections.*)

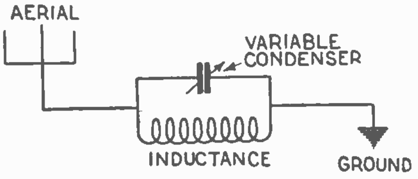

Coil and condenser in parallel.

CONDENSER, IN SERIES—A condenser connected in a circuit with other condensers or coils in such manner as to form a continuous path, allowing the current to pass through each unit without dividing. The illustration shows a condenser connected in series with a *tuning coil,* aerial and ground, thus forming a tuned primary or aerial circuit. (See *Condenser, in Parallel,* also *Series Connections.*)

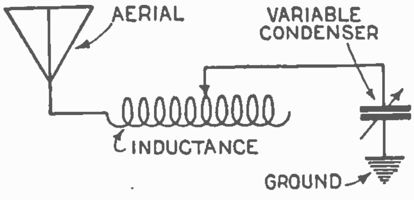

Coil and condenser in series.

CONDENSER, MICA—A condenser, usually of the fixed type, employing sheets of mica as the dielectric or insulating material between the plates. The condensers of small capacity used in receiving circuits are generally made with mica as a dielectric, but owing to its cost and thickness it is seldom used in low-power, high-capacity condensers. Mica is also used in condensers for high-power work such as transmitting, when the capacity is low. (See *Condenser.*)

CONDENSER, PHONE—A fixed condenser of comparatively low capacity used across the *telephone receivers* to allow a by-pass for a certain portion of the currents and thus prevent undue strain on the phones. (See *Condenser, By-Pass.*)

CONDENSER, TRANSMITTING—See *Condenser, Mica.*

CONDENSER, VARIABLE—Any condenser so arranged as to allow variation of its capacity. This may be accomplished in many ways, the usual method being to have the plates divided so that one series is movable with respect to the others, the capacity changing with the rotation of the variable plates. (See *Condenser.*)

CONDUCTANCE — Sometimes called the conductivity. The quality of a given conductor in virtue of which it facilitates the flow of an electric current. A *conductor* (q.v.) is a body having large *conductance* and an *insulator* (q.v.) is one having negligible conductance. *Conductance* is thus the opposite of *resistance.* The unit of conductance is the *mho,* the reciprocal of *ohm.* The symbol is obviously derived from ohm, reversed. It is the conductance of a column of mercury 106.3 cm. (centimeters) long, 1 square mm. (milimeter) in cross-sectional area and of a weight of 14.4521 grams.

CONDUCTIVE COUPLING—The association of one circuit with another by means of inductances (coils) mutual to both circuits.

CONDUCTIVITY, SPECIFIC—The reciprocal of Specific Resistance; being a standard of reference for comparing the conductances of different substances. (See *Specific Resistance, Conductance.*)

CONDUCTOR—Any substance that offers small *resistance* to the passage of electric currents may be termed a *conductor.* This term is the opposite of *insulator,* but there is actually no line of demarcation between the two, since any insulator will permit the passage of current when sufficiently high voltages are used. For all practical purposes, however, a conductor is considered as material that presents little difficulty to the passage of current. The most common conductor is copper, usually in the form of wire. An idea of the relative conductivity (power of carrying electric current possessed by various substances, pure copper being taken as standard) of various substances can be gained from the table following. The materials on the left are comparatively good conductors given in the order of their conductivity, while those on the right are poor conductors and generally used for resistances. (Note: A *resistance* is used to retard the flow of electricity without necessarily stopping it entirely, whereas an insulator is used to stop the passage of current.)

Silver
Copper
Brass
Gold
Aluminum
Zinc
Platinum
Iron
Nickel
Tin

German Silver
Platinum Silver
Manganin
Mercury
Graphite or carbon

There are, of course, many other substances that can be classed as good conductors, or as poor conductors, but the above are more generally known and used. It is a peculiar fact that while good conductors become less conductive as their temperature is raised, insulators lose a certain amount of their insulating properties as their temperature is raised. As too great current in a conductor will cause heat, and as explained above, cause lessening of conducting ability, it will be obvious that for the fullest efficiency a conductor will necessarily be arranged to carry the required current without undue heating. It is understood also that conductors in any part of a radio circuit should be carefully gauged as to size and the effective cross-section of the conductor determined, in order to offset the losses due to heating. It must not be assumed from the foregoing that the larger the wire the greater the efficiency—in fact a large wire is often less satisfactory than a small one in high frequency circuits.

When the conductor is to carry direct current the only consideration is to have the wire large enough to prevent heat losses and at the same time not be too bulky. Where the conductor is to carry *high-frequency* (q.v.) currents there are other factors to consider. High frequency currents are said to travel only on the outside of the wire (see *Skin Effect*) and therefore it is possible to use very fine wire for this purpose when the very low current values obtaining in reception are in question. A phenomenon that must be given attention, however, is that of *distributed capacity* (q.v.). (For more complete explanation of this subject see *Low-loss Coils*. See also *Bus bar*.)

CONNECTING UP or HOOKING UP—These general terms are used in radio to refer to the process of wiring a receiver or any particular part of a receiver or other circuit.

CONSTANT—A quantity or magnitude, which does not vary, derived from actual experiment, which is included as a factor in most formulas for the purpose of bringing theoretical calculations into agreement with experience. Constants are much used in radio calculations. As an illustration, different dielectric or insulating materials used in a condenser have different constants which must be taken into consideration when calculating the capacity of the *condenser*. Thus, as a dielectric, air at ordinary pressure is taken as the standard, the constant being stated as 1. Flint glass has a high dielectric value, its constant being stated as 10.10, other materials varying according to their comparison with air as a standard. (See *Dielectric*.)

CONSTANT CURRENT—An unvarying *current*. The term "constant" is applied to *voltage* when it is steady and unchanging, to *resistance* when it is fixed and unvarying, or to almost any value that does not change.

CONSUMPTION, CURRENT — The term usually employed to denote the current required or consumed by any part of an electrical circuit or by the circuit itself. A resistance in a circuit consumes a certain amount of energy because it dissipates it as heat energy. A filament in the case of a vacuum tube, requires a certain amount of current to heat it to brilliancy. If the filament thus requires one-quarter of an ampere we say its current consumption is one-quarter ampere. (See *Plate Consumption*.)

CONSUMPTION, PLATE—The current required to energize the plate of a vacuum tube. (See *Plate Consumption*.)

CONTACT BREAKER—A contrivance for quickly and automatically making or breaking an electrical circuit normally existing between two contacts. (See *Circuit Breaker*.)

CONTACT DETECTOR—A form of crystal detector in which a small wire, or "catwhisker," as it is called, makes contact with the crystal in the detector and acts as a *rectifier* of the *high frequency waves* received from the broadcasting station. (See *Crystal Detector—Theory of Operation*.)

CONTACT POINTS, or SWITCH POINTS—Generally refers to a small flat-headed machine screw used as a contact point for a *switch*. The illustration shows the types mostly used.

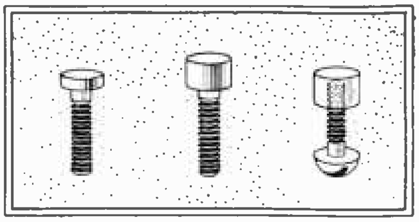

Various contact points.

Leads from various points of a tapped coil may be connected to the several contacts in order to permit variation of the *inductance*. A "contact point" may also mean any point in an electrical circuit at which contact is made to close the circuit.

CONTACT RECTIFIER—See *Contact Detector*.

CONTACT RESISTANCE—When two connecting surfaces in a circuit carrying electrical energy do not make absolute or perfect contact, a certain amount of *resistance* (q.v.) results. In any part of a radio circuit where two wires or *conductors* are joined to form a circuit, it is essential that they make a positive contact. The wires used for connections in a radio receiver are generally soldered or fastened tightly by nuts or bolts, or machine screws in order to form a good contact. If this is not done, energy may be wasted, due to the resistance at the point of contact.

CONTINUOUS OSCILLATIONS—Each oscillation or vibration being of the same amplitude as the preceding one. (See *Sine wave*.)

C. W.—Abbreviation for continuous wave (q.v.).

CONTINUOUS DIRECT CURRENT—A *direct current* which flows steadily without interruption or reversal. The current supplied from a battery for radio use is continuous direct current, while that portion of the *high frequency alternating current* which is rectified by a *vacuum tube* detector is called *pulsating direct current*, as there is an interval of time between the rectified half of each incoming wave, due to the other half not being rectified. (See *Current, Pulsating Direct, Pulsating Current*, also *Rectification*.)

CONTINUOUS WAVES—Also called *Undamped* or *Sustained waves* and usually referred to as *C. W. Electromagnetic waves*, the separate oscillations (each vibration) of which are of constant or unvarying amplitude. In Fig. A a series of *oscillations* (q.v.) are shown. It will appear that one oscillation or *cycle* is of the same amplitude as the preceding one—the amplitude (q.v.) being considered as the height of the wave above the straight line representing time or its depth below that line. Each loop above the line or below it represents an *alteration* (q.v.) (the increase in current to maximum and back to minimum in either direction)—the loops above the line being considered as *positive alternations* and those below the line as *negative alternations*. Now it is evident from the sketch that each succeeding alternation has the same strength (amplitude) as that preceding it, and the same is true of the complete cycle, consisting of one alternation in each direction. Thus the waves are continuous, or constant, as far as amplitude is concerned. Now in Fig. B the amplitude is not constant—it decreases with each succeeding alternation and hence with each succeeding cycle (two alternations). After a comparatively few oscillations the current has decreased to zero, that is, the oscillations have died out completely. In this case the percentage of decrease in amplitude from one cycle to another is known as *decrement* or *damping*. Thus continuous waves may be considered as waves in which, with any train or group of oscillations, each succeeding oscillation has the same amplitude as that preceding it.

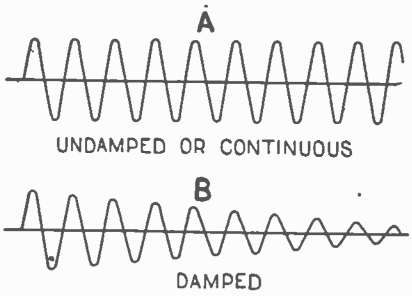

Series of undamped and damped waves.

CONTINUOUS WAVES—KEY MODULATED—Continuous waves (q.v.) which have been broken into dots and dashes of the telegraphic code by means of a *telegraphic key*, these signals being made audible at the receiving end by the use of an oscillating receiver.

CONTINUOUS WAVES—MODULATED AT AUDIO FREQUENCY—The process of variation of the frequency (vibration) of waves radiated by a broadcasting station in accordance with the sound waves is known as modulation. Frequencies from 10,000 down to about 32 are audible to the human ear and are known as audio frequencies. Hence continuous waves which are modulated at *audio frequencies* are waves which have been modulated (varied) by having the voice or music impressed on the *microphone* at the broadcasting station, which in turn causes the strength of the radiated electric wave to vary so as to correspond electrically to the sound vibrations at the microphone.

CONTINUOUS WAVE SIGNALS—See *Continuous Waves*.

CONTINUOUS WAVE TRANSMISSION—System of radio telegraphy using waves as described under "Continuous Waves."

CONVECTION CURRENTS—When a hot and a cold body are separated by a fluid which is free to circulate, heat is transferred from the hot to the cold body by currents of the fluid itself flowing from one to the other; similarly, all parts of a fluid which is being heated quickly come to approximately the same temperature. This transfer of heat by currents of the fluid itself is called "convection." In electricity "Convection Currents" are currents arising from the motion of charged particles thrown off in electrified streams, i.e., the flowing of charged air particles in streams from the pointed end of a highly electrified conductor, sometimes called "electric wind."

CONVENTIONAL—This term is much used in radio to denote a form established by custom. A conventional circuit for some particular purpose is understood as a circuit which has been generally used and which is generally considered as standard. Broadly, the most common form or use of anything. The conventional method of amplifying signals of *audio frequency* is by means of *audio frequency transformers*. Any other means is unconventional, although it may produce similar results.

CONVERSION FACTORS—A numerical factor which gives relation between the magnitude of a quantity expressed in terms of one system of units, and the same quantity expressed in terms of any other system of units. For example, the yard is regarded in the United States as 3600/3937 meter (a meter is approximately 39.37 inches). We say, therefore, that the conversion factor between the yard and the meter is 3600/3937 because any number of yards multiplied by this figure or factor will give the result in meters. (See *Capacity Conversion Factors.*)

CONVERTER—A term applied generally to any mechanical, electrical or electro-chemical device which will convert or change one form of energy to another. More specifically in electricity the term is used in reference to (1) a dynamotor designed with two *armatures*, mounted on one shaft, one being a motor armature driven by the original current, the other armature generates new current; this is also called a "motor-dynamo" and it can transform continuous (direct) current to higher or lower voltages, or, (2) a rotary electrical device used to change direct to alternating current or vice versa. (See *Rotary Converter.*)

COORDINATES—A thing of the same rank with another thing may be termed a coordinate. For example, lines by means of which the position of any point as of a curve may be defined with respect to certain fixed lines. In the illustration Fig. 1 lines known as axes of coordinates are shown. Such a system as this will be the basis for curves (q.v.) showing the relation between varying values. If we assume the horizontal line X as the axis of abscissas, all distances measured parallel to this line will be known as abscissas; distances measured parallel to the vertical line Y being referred to as ordinates. The line A-B is parallel to the axis of abscissas X and is therefore an abscissa. The line B-C is parallel to the axis of ordinates and is referred to as an ordinate. Now if we place a number of lines parallel to each axis (ordinates and abscissas) and assign to each of them a certain value, such as amperes, to each abscissa, and volts to each ordinate, a line can be drawn through a number of points to represent the various values of each quantity as related to the other. Thus, in Fig. 2 we have a purely hypothetical curve representing the assumed value of current in amperes under assumed voltages. In the example, the curve at point A shows approximately three amperes under a pressure of ten volts. This imaginary curve is, of course, shown merely for the purpose of illustrating the meaning of coordinates, ordinates, abscissa, etc. (See *Abscissa, Characteristic curve*, also *Graph.*)

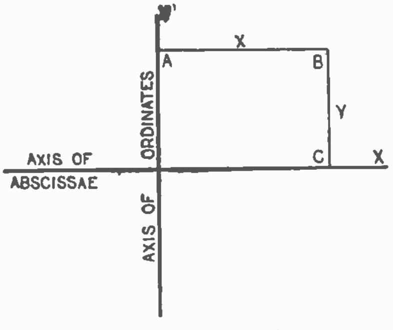

Fig. 1. Axis of coordinates.

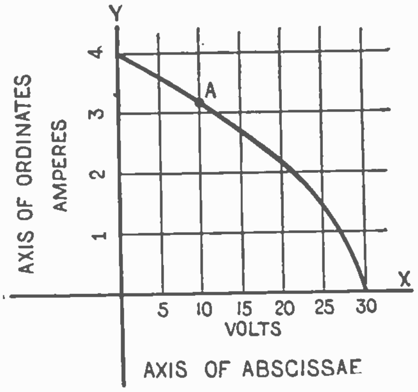

Fig. 2. Curve showing assumed value of current in amperes under assumed voltages.

COPPER PYRITES—Sometimes referred to as *Chalcopyrite*. It is used in radio reception as one of the elements of a *crystal rectifier* (detector) generally in conjunction with zincite or tellurium.

COPPER SULPHATE—A salt of copper. Also known as blue vitriol or copperas. A copper sulphate solution is sometimes used as a preservative for wooden aerial masts. It is used in certain forms of primary cells (devices for transforming chemical action into electric current) and also produces sulphuric acid when subjected to *electrolysis* (q.v.).

CORD, PHONE—The cord or flexible wire used to connect a *headset* (phone) or *loudspeaker* to the terminals of a receiver. Phone cord is usually copper tinsel wrapped around a cotton cord, the whole covered by braided silk or cotton threads. It cannot be used to carry high currents, but is efficient for the comparatively low currents flowing through the phone circuit of a radio receiver. Some phone cords are also formed of a number of fine copper wires covered with a suitable insulating material.

CORE—The central portion of a piece of certain electrical apparatus. Thus the center of a coil will be the *core*, irrespective of its nature, whether air or iron. The core of a *choke coil* (q.v.) is usually made of a bundle of soft iron wires or strips of iron. The core of the closed type is called a *laminated core*, being made up of laminations or layers of the core metal. (See *Choke or Choking Transformer, Closed Core.*)

CORE LOSSES—The losses due to the *stray currents* (q.v.) set up in an iron core while under the stress (strain) of a *magnetic field* (q.v.). Also losses due to *hysteresis* and *eddy-currents*. (See *Magnetic flux, Eddy current*, also *Hysteresis.*)

CORE TRANSFORMER (Iron)—A transformer having the wire wound over an *iron core*, the iron forming a path for the *magnetic lines of force*, as distinguished from a *transformer* in which there is no iron core—called *air core*. (See *Transformer.*)

CORONA—The Latin word meaning crown, originally applied to the crown shaped lights occurring in the *aurora borealis*. In electrical parlance it refers to the luminous discharge which occurs along high tension transmission lines or the aerial wires of a high power radio transmitting station. (See *Brush Discharge;* also *Corona Loss.*)

CORONA LOSS—The energy loss due to corona discharge. (See *Corona;* also *Brush Discharge.*)

CORRECTION FACTORS—A formula for the measurement of certain values is generally made by using certain conditions or types of apparatus as a standard. For example a formula for calculating the inductance of a coil is made by using a particular type and shape of coil as the standard. Therefore the formula will be essentially correct for this type of coil, assuming that its shape and other characteristics are the same. If, however, the nature of the coil to be tested differs from that of the standard coil in some way not covered by the formula, a system of correction must be used to adjust the formula to the particular case. Thus, the formula for inductance of coils takes into consideration the size of the wire or the number of turns, the length of the coil, its diameter and so on. However, it does not take into consideration the ratio between the length of the coil and its diameter. For this purpose a correction factor (necessary consideration) is needed to correct the result of the formula to fit the individual occasion. This correction factor is generally referred to as a *constant* (q.v.), its value being determined generally from a chart prepared to cover various ratios of length to diameter.

CORROSION—Chemical action which causes destruction of the surface of a metal. In the case of iron it is generally rust, corrosion of copper being due to oxidation, or perhaps the action of stray currents causing *electrolysis* (q.v.). This effect is very often noted in the case of an aerial wire where the wire is necessarily exposed to the action of the atmosphere, rain, etc. The effect is particularly important and undesirable where the lead in wire is joined to the aerial.

If the two wires are not soldered securely or at all, corrosion may set in, which will produce in time a high resistance connection and waste the valuable energy of the incoming signals. This is the main point of danger in an aerial system and the best protection is to solder the joint carefully. As a matter of fact the remainder of the wire may corrode to a considerable extent without disastrous effect. In the case of copper, the wire very quickly corrodes when exposed. For this reason insulated wire is sometimes used for the aerial. As the current received by an aerial travels only on the outer surface of the wire (see *Skin Effect*), and a corroded surface is not a particularly efficient conductor, there seems to be some basis for the use of insulated wire for this purpose. Another instance of corrosion is in a soldered joint within a receiving set. All soldered joints should be carefully wiped to remove any trace of *acid flux* (q.v.), which may remain to produce a destroying chemical effect on the wire and cause loss of energy. (See *Electrolysis*, also *High Resistance Joint.*)

COULOMB—The practical unit for electrical quantity. It is defined as the quantity of electricity delivered by a current of one *ampere* flowing for one second. Therefore a *coulomb* of electricity will pass each second in a circuit having one *ohm* of resistance and a pressure of one *volt.* It is also used to express the quantity of electricity which a condenser having one *farad* of capacity will absorb when placed under a pressure of one volt. (See *Capacity, Electro-static, Volt, Ampere Hour, Practical Units, CGS Units.*)

COUNTER ELECTROMOTIVE FORCE—Commonly known as *Counter E. M. F.* It is an opposing electromotive force which tends to resist or act against the original electromotive force. It may be introduced in a circuit to prevent any fluctuation of the original force. (See *Back Electromotive Force*, also *Choke Coil.*)

COUNTERPOISE—A *counterpoise* is generally a second aerial suspended either below the main aerial or to one side of it. It is used more often in transmitting stations in place of the usual ground connection, but is occasionally also used in reception where it is impossible to obtain an efficient ground. (See *Aerial.*)

COUPLE, THERMO ELECTRIC—See *Thermo-Couple.*

COUPLED CIRCUITS—The term generally used to refer to circuits that are joined by *coupling.* (q.v.) In the illustration the *primary circuit* (the aerial ground and tuning devices) is coupled or joined to the *secondary circuit* by *inductive* coupling. That is to say, the signals impressed on the *primary circuit* are transferred to the *secondary circuit* by means of *inductive coupling;* when the lines of magnetic force around the primary coil cut the secondary coil, a voltage is set up in this coil which causes a current to flow in the secondary circuit, and it is then said the coils are "inductively coupled", there being no connecting wires between them. Signals might be transferred to the secondary circuit by any of several means, but in each case the method of joining the two for the purpose of effecting a transfer of energy is referred to as "coupling the circuits." (See *Coupling.*)

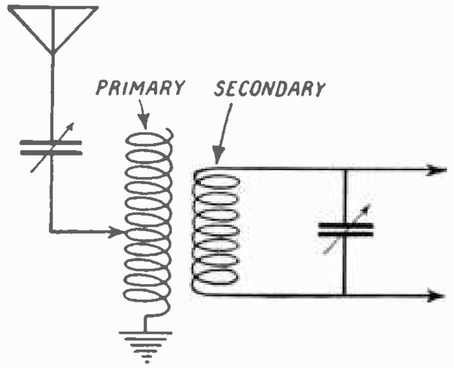

Circuit in which primary and secondary are joined by inductive coupling.

COUPLER—See "Coupler—(Loose)" and "Coupler—(Vario)".

COUPLER—(LOOSE)—A tuner used in a radio receiver, consisting of a primary and secondary coil so constructed that the secondary coil slides on rods in and out of the primary coil. When the secondary is entirely within the primary the coupling is said to be "tight," and as the secondary is withdrawn from the primary the coupling becomes "loose," depending upon the distance which the coils are separated.

COUPLER—(VARIO)—Form of tuner used in a radio receiver, having primary and secondary coils arranged in such a manner that the secondary rotates on a shaft within the primary. The degree of coupling is governed by the variation of the angle between the axes of the two coils.

COUPLING—The relation of one coil of wire carrying alternating currents to another in which energy is set up by the first coil. It is used to define the intimacy between the primary and secondary windings of various types of *oscillation transformers, couplers* and *tuning devices.* This transfer of energy may be produced in several ways, as by *inductive coupling, conductive coupling* or *electro-static coupling.* The more common method is to place the primary and secondary coils in close proximity and produce the energy in the secondary through induction. The illustration Fig. 1 shows a simple two slide tuning coil which is divided into a primary and secondary by *conductive* coupling. Here the transfer of energy from the aerial and ground, or primary circuit, to the secondary circuit (detector and phones) is accomplished with little loss of energy as the circuit is through wire conductors. This is also known as *direct coupling.* Now in Fig. 2 a simple form of *loose-coupler* is shown. This device is made in numerous forms, the principle of operation however, being the same in all cases. Here one coil, the secondary in this case, is made to fit into the larger coil which is the primary. The energy flowing in the primary coil induces or sets up a similar current in the secondary by reason of its proximity. The energy in the secondary, however, will not be as great as in the case of Fig. 1 where it is transferred directly through a wire conductor. The coupling shown in Fig. 2 will on the other hand, permit much closer adjustment to the desired signals.

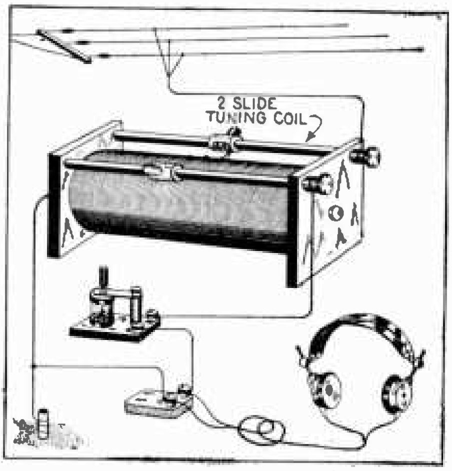

Fig. 1.—Two slide tuner divided into primary and secondary circuits by conductive coupling.

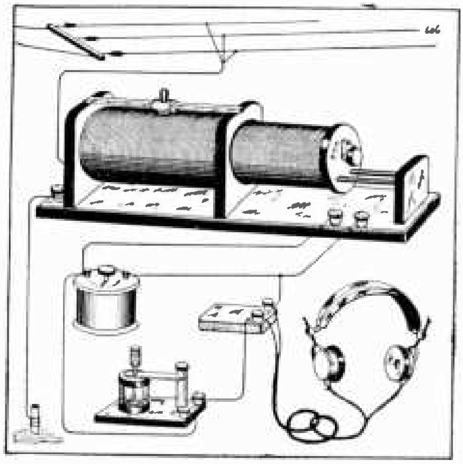

Fig. 2.—Circuit in which primary and secondary are inductively coupled by loose coupler.

The other method is known as *capacitive* or *electrostatic coupling.* This is shown in Fig. 3. Here the secondary coil is not in direct inductive relation to the primary, in fact it might well be almost any distance away were it not for the loss in efficiency due to long connections. The connection between the primary and secondary coils is by means of condensers. The transfer takes place through these and is determined by the setting of the condensers. It will appear that from a standpoint of maximum transfer of energy, the direct or *conductive coupling* system is best. The factor of tuning enters here however, and must be considered. In order to eliminate undesired signals the various systems of coupling are resorted to. It is known that *selectivity*—the ability to tune out undesired signals while receiving the desired signals—is often obtained only at the expense of a certain amount of energy or volume. It is a safe assumption therefore, that if there is any such thing as maximum efficiency in tuning apparatus, it must be a general term covering both selectivity and volume.

Coupling in its various forms is intended to refer not only to receiving circuits but to transmitting circuits as well. Its application in transmitting is covered under the heading *Spark transmission* (q.v.).

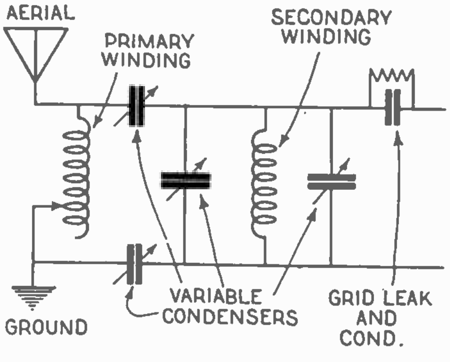

Fig. 3.—Electrostatic coupling between primary and secondary by means of variable condensers.

COUPLING COEFFICIENT — The closeness of *coupling* is specified by a quantity representing the ratio of the *mutual inductance* between two coils to the square root of the product of the two *self inductances.* In its practical sense coupling coefficient refers to the inductive relation of two coils. Coupling coefficient is defined as the ratio

$$\frac{Xm}{X.\ X_2}$$

Where Xm is the mutual or common reactance X is the total inductive or capacitive reactance in the primary circuit and X_2 the similar reactance in the secondary circuit. (See *Coupling, Mutual Inductance.*)

COUPLING, DEGREE OF—The extent of relationship between coupled circuits. In a great number of circuits the coupling is arranged so that it may be adjusted either close or loose, as the requirements of the reception may be. Variable coupling may be obtained either by varying the distance between the two coils of a *loose coupler*, or by changing the angle between the coils of a *variocoupler*. (See *Coupling.*)

COUPLING TRANSFORMER—(OSCILLATION TRANSFORMER) — This is the "tuner" of the transmitting circuit. It consists of primary and secondary windings usually of flat strips of copper, or copper ribbon, instead of wire, in order to reduce losses due to *skin effect* (q.v.). The primary coil receives the energy in the form of alternating currents from the generating device; these currents set up lines of force which interlink with the windings of the secondary coil connected in series with the antenna system, and in this manner the energy of the oscillation (alternating) current is transferred to the radiating system (aerial and ground) and thence into space. The purpose of the oscillation transformer is to produce a sharp wave which is readily tuned at the receiving end and causes a minimum of interference. The greater the distance between the primary and secondary coils the more sharp the wave becomes.

A device serving the same purpose in a receiving set, i.e. to enable the receiver to separate sharply the received signals, is sometimes called a receiving transformer, although commonly known as a tuner, coupler, vario-coupler, etc.

C. Q.—Abbreviation of inquiry made in code by a radio telegraph station desiring to communicate with another station. (See *Abbreviations.*)

C. Q. D.—The original Distress Signal (q.v.) established in 1904 by the Marconi Company. The origin of this signal is said to have developed from the use by telegraphers on land lines sending the C. Q. call when they wished everyone on their circuit to listen to the message being transmitted.

The signal C. Q. D. is now obsolete, having been replaced by S. O. S. at the International Radio Telegraph Convention held at Berlin, Germany, in July, 1908.

CRITICAL—A term often used in reference to the tuning qualities of a radio set, and, less frequently, to indicate a *vacuum tube* that *oscillates* too freely. If a receiving set is not critical in adjustment it is said to tune broadly, that is to say, not sharp in tuning. A critical set has a very narrow margin of control in tuning—the particular signals being receivable within only a very few points on either side of their true wave setting on the tuning dials. If a set is difficult to tune it is termed "too critical". This means that it tunes too sharply and it is difficult to get proper results, due to the possibility of passing by signals unnoticed when turning the dials. Obviously there is a middle point wherein a set may be selective and yet not too difficult to tune. If a vacuum tube goes into *self-oscillation* (q.v.) too readily it is said to be critical. This effect is not desired in a tube except in rare instances. (See *Resonance, Tuning, Broad, etc.*)

CRITICAL CURRENT—The current required to bring about some particular effect in an electrical circuit. In radio use, a common example is the exact amount of current needed by the filament of a vacuum tube to bring it close to the *oscillating point* (q.v.) without actually going into oscillation. (See *Critical.*)

CRITICAL POINT OF CURVE—The peak or point of maximum value in a curve showing the relation of any varying values. The illustration shows the curve indicating relation between voltage and signal strength in two types of vacuum tubes. The points marked "Peak" in each curve represent the critical points of the curves at which maximum audibility is obtained, any further increase in voltage not increasing the audibility. (See *Characteristic Curve*, also *Curve.*)

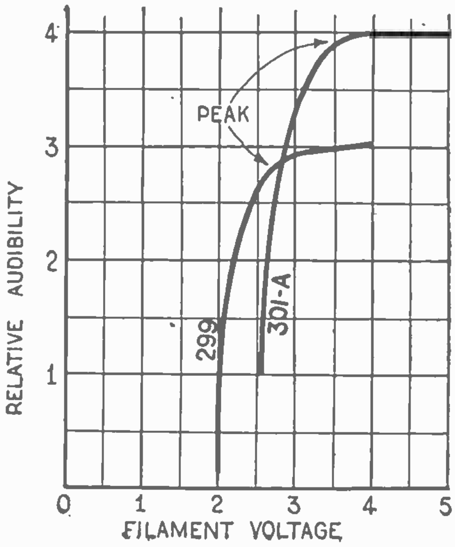

Illustration showing peaks or critical points of curve.

CROOKES, SIR WILLIAM (1832-1919) —English chemist and physicist. Born in London June 17th, 1832, Crookes was educated at the Royal College of Chemistry and afterwards became assistant in the meteorological department of the Radcliffe Observatory, Oxford. In 1855 he obtained a chemical post at Chester.

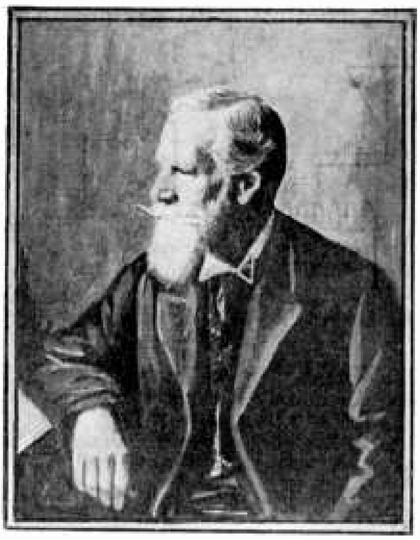

Sir William Crookes.

Crookes early began that series of brilliant researches which has left its mark on the scientific progress of the nineteenth century. In 1861 he isolated the new element *thallium*, during the investigation of the atomic weight of which he made the discovery of his *radiometer*. This led to his researches on the phenomena of electric discharges through highly exhausted tubes, or *Crookes tubes*. The illustration shows such a tube. When a current is passed through it. cathode particles travel along it. His discoveries on these phenomena led directly to the development by Sir J. J. Thomson of the now generally accepted *electron theory*. (q.v.) In 1883 he began his study of the rare earths, and in 1892 he forecast Radio telegraphy on the strength of Lodge's and Hughes' experiments.

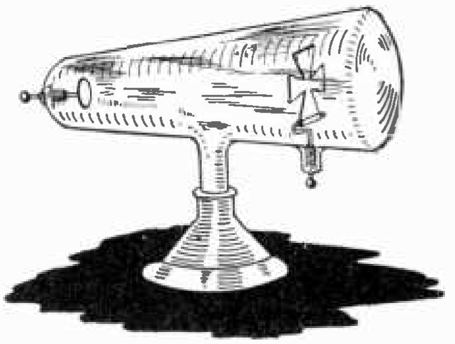

Crookes' tube.

Knighted in 1897, Crookes was awarded the Royal medal in 1875, the Davy medal in 1880, and the Copley medal of the Royal Society in 1904, and the O.M., in 1910. He died in London, April 4th, 1919.

CRYSTAL—A general radio term for a mineral used as a *detector* (q.v.) of radio signals. The Physicist F. Braun noted in 1874 that certain pairs of crystals when arranged so that only a small area of surface was in contact, offered high resistance to the passage of currents in one direction while permitting them to pass readily in the other. This was really the birth of the crystal detector, although radio was not then developed. There are a great many different minerals that will act as detectors for radio signals, some of which operate with a wire, or other metallic contact, others which require a combination of minerals, one in contact with the other. There are also numerous *synthetic crystals* such as *carborundum*. Then again some of the more common natural crystals are subjected to treatment by heat, or other means, to change their nature or to make them more sensitive.

The following crystals are most used in radio reception: *galena, silicon, molybdenum, carborundum, zincite, bornite, chalcopyrite*, and *cerusite*. All of these are taken up under their respective headings. (See *Crystal Detector*, also *Combination Detector.*)

CRYSTAL DETECTOR—A combination of a *crystal rectifier* with contact and supports, used to detect *radio frequency oscillations*. Detectors of this nature are made in a great variety of styles, and use various kinds of crystals. In the more widespread form

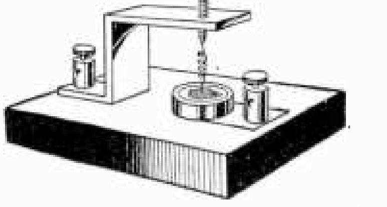

Fig. 1. Crystal detector using contact wire, or "cat-whisker."

the crystal detector consists of a crystal with rectifying properties, held in a cup or spring grip and arranged with a contact wire as shown in the illustration Fig. 1. This may be elaborated upon by supplying a series of flat springs to obtain very fine adjustment of the contact. Fig. 2. shows a combination detector, using two crystals, in contact with each other. This form has several advantages over the wire or *cat-whisker* (as the contact wire is called) type, chief among them the fact that it will withstand severe shocks without losing its adjustment. Fig. 3 illustrates a popular form of fixed crystal detector, in which the contact seldom requires adjustment. This type is of advantage providing it can be made sufficiently sensitive. (See *Carborundum, Vacuum Tube,* also *Crystal Receivers.*)

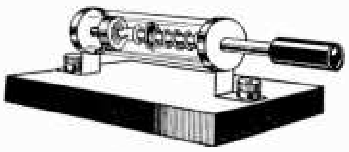

Fig. 2. Combination crystal detector.

Fig. 3. Fixed crystal detector.

CRYSTAL DETECTOR—THEORY OF OPERATION—Theory of operation of a device used to detect radio waves, usually comprising a crystal in combination with a metallic contact, or, in some cases, two crystals arranged to rest against each other. In action this device serves to rectify the incoming signals, that is to say, it changes their nature from an alternating current to pulsating direct current. The incoming waves or signals consist of a series of alternations, the current flowing first in one direction, then reversing and flowing in the opposite direction along the circuit of the receiver. The changes of direction take place many times each second and would not be audible in the head phones in their original state. The crystal, as used in radio reception, has the peculiar property of furnishing a ready path for the flow of current in one direction while offering high resistance to its passage in the other. Now an *alternation* (q.v.) is a rise in current from minimum to maximum in one direction and back to minimum or zero, repeating the process in the other direction. Two of these alternations, one in a positive direction and the other in the negative direction constitute a *cycle.* (q.v.) If the crystal will permit the passage of current in one direction only, it is apparent that only half of each cycle can pass through; the other half of each cycle being lost or blocked out by the action of the crystal.

The illustration "A" gives a graphic representation of the *radio frequency oscillations* as they move over the aerial circuit. The curves above the line show the current in a positive direction while those below the line show it in the opposite or negative direction. If the signals or waves in this state were sent through a telephone receiver or head phone in a radio receiving set no sounds would be reproduced. Then again if the incoming oscillations were continuous, of unvarying strength, the rectified current would merely act on the diaphragm of the phone and hold it in a certain position as long as the signal persisted. This of course would not create any audible or intelligible sounds. If, however, the incoming oscillations were in the forms of groups or trains of waves, in any one of which the successive oscillations were less than the previous ones, the current would be rectified as shown in B and the result would be a series of pulses in the phone as indicated in C. These pulses would occur as often as the trains of oscillations occur (for more complete explanation of the difference between continuous and damped waves, see Damped Waves, also Continuous Waves). It will therefore appear that a *crystal detector* can only be used to receive *damped waves.* When the rectified portion of a damped wave train reaches the phones, the *diaphragm* is depressed and held until the train has died out, at which point the diaphragm is released until another train reaches it. This creates the audible signals and the number of trains occcurring each second will determine the number of depressions of the diaphragm and hence the pitch of the audible signals. (See *Crystal Receiver, Crystal Detectors,* also *Tickler.*)

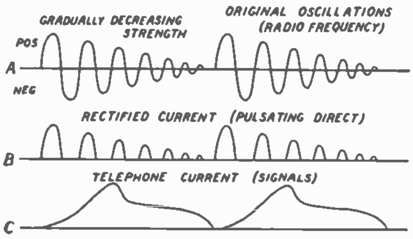

Rectifying process of radio signals.

CRYSTAL RECEIVERS—A crystal receiver is the complete set of apparatus for receiving broadcast programs or *spark transmission* (signals), employing a crystal as the *detector* or *rectifier.* The set comprises a crystal detector, tuning device and phones, together with the usual aerial and ground arrangement. In its simplest form it may be merely a crystal detector, coil of wire and phones as in Fig. 1.

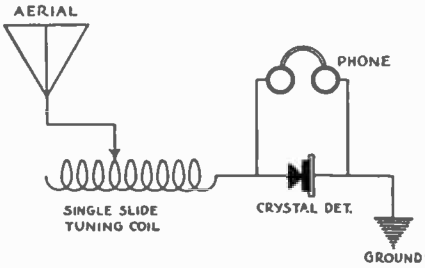

Fig. 1. Simple crystal set with single slide tuner.

Fig. 2 shows a more efficient form, using a two slide tuning coil. This method permits closer adjustment of the receiver to the particular signals desired. A still more efficient form is the *inductively coupled* type shown in Fig. 3. This method allows a very close adjustment and considerably reduces interference. The kind of detector to be used in any of the above types of receiver is immaterial as long as it is sufficiently sensitive. Such an arrangement as shown in Fig. 3 is very inexpensive to construct and when used within ten to twenty miles of the broadcasting station gives excellent results. A very simple arrangement for use with a crystal detector is shown in Fig. 4. This circuit uses a standard *variometer* and eliminates the necessity for a *condenser,* thus having only one control in addition to the adjustment of the crystal.

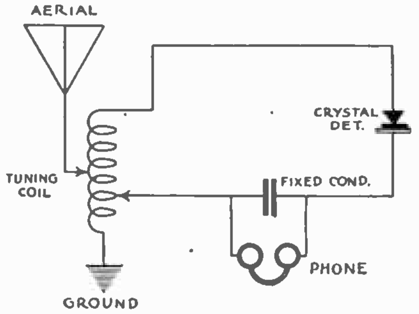

Fig. 2. Crystal receiver using double slide tuner.

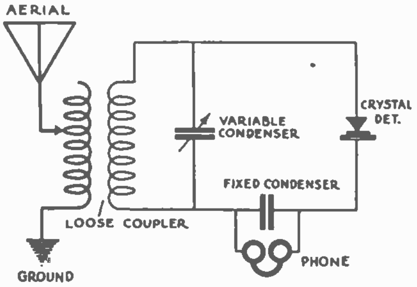

Fig. 3. Selective crystal receiver with loose coupler for tuning.

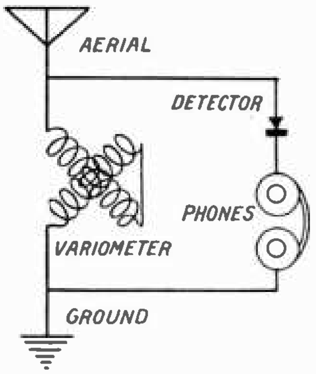

Fig. 4. A variometer is used as the tuning element in this crystal set.

CRYSTAL RECTIFIER—A device for changing an alternating current in a radio receiver into a pulsating direct current. This action is fully explained in "*Crystal Detectors (Theory of Operation).*"

CURRENT—A term used in electrical practise to signify the rate at which electricity flows from point to point in a circuit. The current in a circuit may be likened to the water flowing in a pipe. In this case the water would represent the electricity, the force or pressure represents the voltage, the pipe can be compared to the conductor or wire through which the electricity flows, and the current can be likened to the discharge from the pipe over a given period of time, that is the cubic feet of water per second. For practical purposes the unit of current is considered as the *Ampere.* This is the rate of flow of electricity when one *coulomb* (q.v.) passes a given point in the circuit each second. It will thus be apparent that while the ampere is often considered as the amount of current in a circuit, in reality it is the rate of flow of the electricity, the *coulomb* being the definite unit of quantity. Electric current is of course invisible. The only means of determining when there is current in a circuit is by observing its effect and it is always apparent by one or more of a number of effects. For example, if current is flowing in a wire or other conductor and a compass is placed near the conductor in such a position that the axis on which the needle revolves is parallel to the axis of the wire, the needle will be deflected. If the current reverses and flows in the opposite direction in the wire, the needle will be deflected in the opposite direction. The effect of current is also apparent by reason of the heat which it generates in passing through a wire conductor. If the wire is comparatively small and the current correspondingly large, the heat may be sufficient to be noticeable by placing the hand on the conductor. When the current is small, it can be detected by means of delicate instruments. This heating effect is made use of in various types of *ammeters*. The effect of current may also be noted in an incandescent lamp. In this case the current is made to flow through a fine wire inside a globe from which the air has been extracted, and the heat generated by the passing current causes the filament, or fine wire, to glow.

It is a well-known fact that heat causes expansion and this expansion of a fine wire is the principle on which one type of current measuring device operates.

Current is divided broadly into two divisions, *direct* and *alternating*. These two classes may again be separated into several further groups such as *high* and *low frequency alternating*, *continuous* and *pulsating*. Each phase or characteristic of current is taken up more in detail under its particular heading in this encyclopedia. (See *Current, Direction of Flow*, also *Current, Production of*.)

CURRENT, ALTERNATING—A current which does not flow steadily in one direction but changes its direction in the circuit periodically. (See *Alternating Current*.)

CURRENT CARRYING CAPACITY OF WIRE—The amount of current which a wire or other metallic conductor will carry without over-heating. Whenever current flows through a conductor, heat is generated. The amount of heat will be directly proportional to the resistance of the conductor and the square of the current in it. Now the resistance of a conductor depends on its cross-sectional area, length, and also its nature. It is therefore correct to state that the heating will depend on the amount of current flowing and the cross section and length of the wire. Obviously, a large wire will carry more current without over-heating, than will a small wire. Thus, it is necessary when choosing wire for a conductor to carry a certain amount of current, to select a wire of a cross section sufficient to carry the desired amount of current without undue heating.

The following formula for determining the current carrying capacity of various sizes and kinds of wire uses a certain factor T, which is the permissible rise in temperature above surrounding medium—such as air, earth or water. Where d is the diameter of the conductor in inches, T—the permissible temperature rise in degrees centigrade, N the resistance of conductor in ohms per mil-foot at final temperature, and I the current in amperes, then for solid conductors:

$I = \frac{K\sqrt{Td^3}}{r}$ and for stranded conductors: $I = 0.85\,K\frac{\sqrt{Td^3}}{r}$

K is a *constant* (q.v.) depending upon the condition of the surface of the wire and upon the amount of heat convection due to air currents. Values of the constant K for air differ according to different authorities from 800 to 1000, the former referring to still air and the latter to open air.

Table of current carrying capacities follows:

Wires and Cables, Insulated: Carrying Capacities, in Amperes, allowed by The Regulations of The National Board of Fire Underwriters for Interior Copper Conductors.

(For Aluminum 84 Per Cent of These Currents is allowed.)

Single Conductor Cables or Each Conductor of Multiple Conductor Cables.

A. W. G.	Area in Circular Mils	Table A. Rubber Insulation	Table B Varnished cloth	Table C. other insulation
18	1,624	3	....	5
16	2,583	6	....	10
14	4,107	15	(18)	20
12	6,530	20	(25)	25
10	10,380	25	(30)	30
8	16,510	35	(40)	50
6	26,250	50	60	70
5	33,100	55	65	80
4	41,740	70	85	90
3	52,630	80	95	100
2	66,370	90	110	125
1	83,690	100	120	150
0	105,500	125	150	200
00	133,100	150	180	225
000	167,800	175	210	275
	200,000	200	240	300
0000	211,600	225	270	325
	250,000	250	300	350
	300,000	275	330	400
	400,000	325	390	500
	500,000	400	480	600
	600,000	450	540	680
	700,000	500	600	760
	800,000	550	660	840
	900,000	600	720	920
	1,000,000	650	780	1000
	1,100,000	690	830	1080
	1,200,000	730	880	1150
	1,300,000	770	920	1220
	1,400,000	810	970	1290
	1,500,000	850	1020	1360
	1,600,000	890	1070	1430
	1,700,000	930	1120	1490
	1,800,000	970	1160	1550
	1,900,000	1010	1210	1610
	2,000,000	1050	1260	1670

Varnished cloth smaller than No. 6 may be used by special permission only.

CURRENT, CONVECTION—See *Convection Current*.

CURRENT, CRITICAL—See *Critical Current*.

CURRENT DENSITY—The number of amperes passing through a wire conductor per square inch of area of the conductor. For example, if a conductor is a wire or bar of copper having a cross section of one square inch and ten amperes are passed through it, the current density will be ten amperes per square inch.

CURRENT, DIRECT—Current that flows steadily, and only in one direction. (See *Direct Current*.)

CURRENT, DIRECTION OF FLOW (ASSUMED)—In electrical practice the *assumed direction of flow* of current is often referred to. This refers to the fact that direct current is assumed to flow from the *positive pole* of a battery, or other source of electrical energy, to the *negative pole* and thence back through the battery from negative to positive. This assumed action is shown in the illustration Fig. 1, where the arrows indicate the assumed direction of flow. Now while the current is stated as flowing from positive to negative outside the source, it is the generally accepted theory that the electron stream moves in the opposite direction, or from negative to positive. In other words, the current as we use the term, flows backward along the path of electrons. (See *Electron Flow*.) This reverse action of current and electrons is shown by the illustration Fig. 2 This apparent anomaly

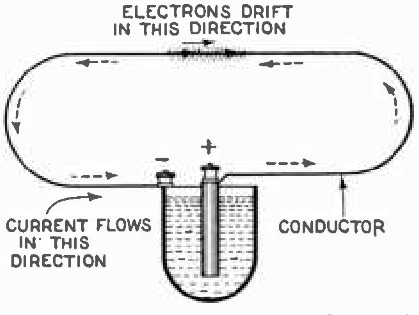
ELECTRONS DRIFT IN THIS DIRECTION
–
+
CURRENT FLOWS IN THIS DIRECTION
CONDUCTOR

Fig. 2. Showing reverse action of current and electrons.

can be likened to a salmon swimming up a rapid. The movement of electrons can be compared to that of the the water in the rapids and while the fish actually swim against the stream, and the electric current flows against the electron flow, the water and the electron stream nevertheless furnish the path of motion in each instance. The reason for the contradiction lies in the unfortunate arbitrary choice of positive and negative poles long before electrons were known. (See *Electron.*)

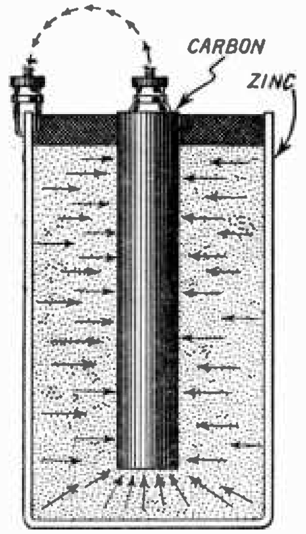

Fig. 1. Illustration showing the assumed direction of current flow in a dry battery.

CURRENT, EDDY—See *Eddy Currents.*

CURRENT, HIGH FREQUENCY—A current that changes direction many times each second. In radio usage, currents having a frequency over ten thousand cycles per second are considered as *high frequency* or *radio frequency currents.* (See *Frequency, Radio Frequency* also *Low Frequency.*)

CURRENT, HIGH TENSION—Alternate term *High Pressure.* When the voltage, tension or pressure, in a circuit is comparatively low, the current is said to be a low pressure or *low tension current.* When the voltage is high, the current is said to be of high tension or pressure. Thus, as applied to radio usage, a dry cell or storage battery giving, respectively, one and a half and six volts for a certain type, are said to produce low tension or low pressure current. A "B" battery (q. v.) giving from 22½ to 45 volts (the usual value) furnishes *high tension* or *high pressure current.* While there is not a very great difference in voltage between a small "B" battery and an "A" type storage battery, the ratio of voltage to current in a "B" battery is far greater than in the ordinary case of the "A" battery and this must be taken into consideration. In general electrical practice the terms are not so arbitrary, and high tension currents might be taken as those above a pressure of several thousand volts.

CURRENT, LOW FREQUENCY—An alternating current that changes its direction of flow a comparatively few times per second. In commercial electrical work any frequency up to about 500 cycles per second is considered as low frequency current. In radio phraseology *audio frequency* (q.v.) currents are termed low frequency. For ordinary radio practice low frequency may be considered as any frequency below about 10,000 cycles. (See *Oscillation* also *Oscillatory current.*)

CURRENT, OSCILLATORY—See *Oscillatory Current.*

CURRENT, PRIMARY—A current which flows directly from its source, such as a cell or generator. The current obtained from any source through a direct conductive circuit is termed a primary current. (See *Secondary Current,* also *Induced Current.*)

CURRENT, NATURE OF ELECTRIC—As explained under the heading *Current,* electric current is manifest or understood by its effects. It may be apparent through the production of heat or light, by producing mechanical action such as required to operate a bell or move an electric motor, or, again, by producing chemical changes (See *Electrolysis*), and still further in its disastrous effects on the human body when the current has sufficient intensity. Without any actual or visible form of manifestation it can best be understood from the *electron* basis. It will be understood that an *electron* (q.v.) small as it is, contains a charge of electricity. When a sufficient number of electrons are in motion in a conductor (Note: current consists of electrons in motion) the current becomes large enough to be measured. As an instance of this fact a current of one ampere intensity (one coulomb per second) requires that about 10^{19} electrons flow past a given point in the circuit each second. When the electrons are not progressing along a conductor, no current flows in the circuit. If no current is being carried by the conductor, the electrons may move about, but they do not progress along the conductor.

CURRENT, PRODUCTION OF ELECTRIC—An electric current is produced in a conductor by applying an *electro-motive force* (voltage) at points of the conductor or by maintaining a difference of *potential* (q.v.) between two points. In the first example we can assume a metal ring or closed coil of wire through which a magnet is pushed or perhaps in the core of which a magnet is being excited, or in the general sense any complete circuit through which the number of magnetic lines of force are being varied. In the second instance the difference of potential may be due to a *primary cell* (q.v.). Thus, the first case can be understood as referring to any mechanical means of producing current, such as *generators.* The other general method is by use of a *voltaic cell,* that is, by chemical action. There are four important methods of producing an *electro-motive force.* The first is by friction, the apparatus used in this connection being termed a static machine; the second is by chemical action, as a primary or secondary cell; third, by mechanical motion, i.e., *dynamos* and *generators,* and fourth, by thermal action—known as a *thermo-junction* which is the wiring of two unlike metals in a conductory circuit so that by heating one of the points of joining an electric current will be set up in the circuit. (See *Static Electricity, Generator, Thermo Couple,* also *Current and Electro-Motive Force.*)

CURRENT, PULSATING DIRECT—When an alternating current is rectified by means of any device which permits current to pass only in one direction, the resultant current will move only in one direction, hence it is known as a *direct current·* but unless both negative and positive halves of the *cycle* (q.v.) are rectified or changed to direct current, the resultant current will appear as a series of pulses in one direction. The most common application of this is found in the case of a *crystal rectifier,* or *detector* as used for receiving radio signals or broadcasting.

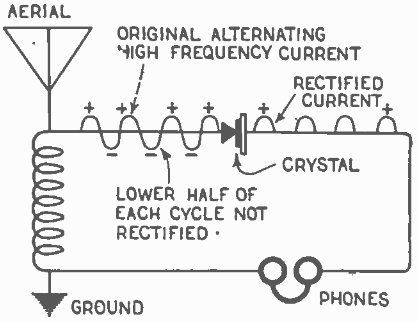
AERIAL
ORIGINAL ALTERNATING HIGH FREQUENCY CURRENT
RECTIFIED CURRENT
CRYSTAL
LOWER HALF OF EACH CYCLE NOT RECTIFIED
GROUND
PHONES

Diagram showing how crystal detector converts alternating high frequency current into pulsating direct current.

When the incoming waves, or high frequency *oscillations* (Q.v.), are impressed on the detector, due to its peculiar properties they can only pass through it in one direction. In other words, granting that the incoming signal, or current, reverses its direction a great many times each second, it will be obvious that at any instant in the *cycle* (q.v.) current is moving in a certain definite direction, as from positive to negative, or vice versa, as stated, due to the peculiar properties of the detector the incoming signal current can only pass through the detector when the current is flowing in the correct direction. This means that only half of the *wave* (q.v.) is passed through and rectified. The result is a series of direct current surges through the balance of the circuit, each surge of current being followed by a space or time interval while the opposite half of the *cycle* or opposite *alternation* (q.v.) is blocked by the uni-directional nature of the crystal. These surges are then audible in the *head phones,* as they will actuate the *diaphragm* (q.v.) whereas a high frequency alternating current such as the original impulse from the broadcasting station would be inaudible. For a more complete explanation of this action see Rectifying action of crystal detector. (See also, *Rectifier, Full Wave Rectification and Rectifying Tube.*)

CURRENT, SECONDARY—A current produced by a *primary current* (q.v.) acting on a *secondary curcuit* which does not receive its energy by direct connection with the primary source. The illustration shows both primary and secondary currents and the manner of production. It will be understood that the primary current in this case must be a varying current. (See *Induced Current,* also *Inductive Coupling.*)

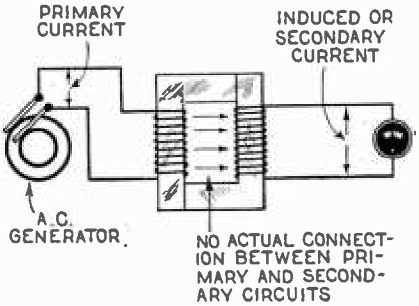
PRIMARY CURRENT
INDUCED OR SECONDARY CURRENT
A.C. GENERATOR
NO ACTUAL CONNECTION BETWEEN PRIMARY AND SECONDARY CIRCUITS

Illustration showing primary and secondary currents and method of production.

CURRENT, THEORY OF ELECTRIC—Under Electromotive force it will be explained that voltage or pressure is the result of a difference of potential (the power possessed by a charge of electricity for doing work) between two points in a conductor. Electric current will flow from the point of higher pressure or *voltage* to the point of lower pressure, the flow continuing as long as the difference is maintained. This action is the same as with water, which will always seek to move from a high point to a lower point. The flow of current is thus an effort to equalize the two *potentials*. (See *Current, Assumed Direction of Flow*, also *Pressure and Voltage*.)

CURVE—A straight or curved line showing the relation between various electrical phenomena or the characteristics (relation of changing values) of an electrical instrument. (See *Characteristic Curve*, also *Sine Wave*.)

CURVE OF SINES—A curved line representing the vibration of a body that oscillates like a pendulum. The illustration shows a *curve of sines* which may be used to represent an alternating current. (See *Alternating Current*, also *Sine Wave* and *Simple Harmonic Vibration*.)

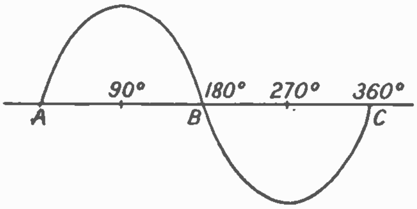

Sine curve.

CUT OUT—An electrical device to interrupt the flow of current through any particular piece of apparatus or instrument, either automatically or by hand; a switch.

CYCLE—In electricity, a period of time during which certain changes take place in an *alternating current*, (q.v.) the same changes occurring again in each successive cycle. We can imagine an alternating current as travelling along a wire in the form of a wave, one half of the wave being in one direction and the other half being in the opposite direction. That is to say, the current flows from positive to negative for one half of the cycle and then in effect, flows from negative to positive the other half of the cycle. (It will be understood that actually current always flows in the same direction, but the action of alternating current is due to a reversal of polarity at the source.) The difference between direct and alternating current is that, in the case of direct, the polarity always remains the same at the source and hence the direction of flow of the current is always in one direction, whereas with alternating current, the polarity changes twice each cycle due to changes at the source, and thus it flows in one direction for a certain fixed period, then reverses and flows in the opposite direction. In the illustration the current is shown as a *sine wave*. (q.v.) We assume that this represents an alternating current of 110 volts pressure such as in the average house lighting system. Now we consider the wave as a circle. The current flows in a positive direction toward its maximum value and at the 90 degrees it has reached the positive maximum. It then falls back to zero at 180 degrees and reverses, rising to maximum on the negative side, which it reaches at 270 degrees. It then falls back to zero at 360 degrees having completed a cycle or two alternations. The operation is repeated over and over a certain number of times each second. Thus, the rise and fall from A to B or 180 degrees is one *alternation*, the rise and fall from B to C is another alternation, the two alternations, one positive and one negative making a complete cycle. If sixty of these cycles occur each second we say that the frequency (q.v.) is sixty cycles. (See *Alternating Current, Phase, Sine Wave*.)

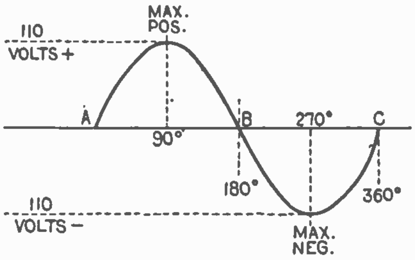

Curve showing action of 110 volts alternating current during one cycle.

CYMOMETER—The name given a type of *wave-meter* designed by Dr. J. A. Fleming. It is used to measure the wave-length of oscillatory (vibrating) circuits. (See *Wave-Meter*.)

CYMOSCOPE—A term used to designate any instrument which enables one to see the effect of electrical waves or to detect their presence. The original form was merely a loop of wire not completely closed as shown in Fig. 1. When held close to a transmitter of short waves a small spark will pass between the ends of the loop providing the *wave-length* is properly adjusted. This was the form used by Hertz (the discoverer of electro-magnetic waves which form the basis of radio telegraphy and telephony) in his early experiments. The more modern form is shown in Fig. 2 where a small lamp is made to light by holding the loop close to the transmitter. A complicated arrangement used to acually see the waves and study their form is called the *Oscillograph*. (See *Oscillograph*, also *Cathode Ray Tube*.)

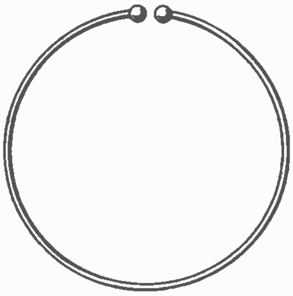

Fig. 1. Early type of cymoscope used to indicate resonance in a circuit.

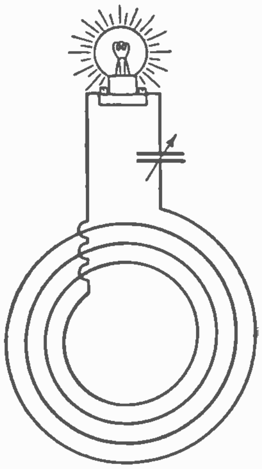

Fig. 2. More modern form of cymoscope.

D

D—The symbol of *electric displacement* (q.v.); also occasionally used as a symbol for diameter in electrical calculations.

DAMPED OSCILLATIONS—Electrical oscillations which die away, each succeeding oscillation in a group having lesser *amplitude* (q.v.) or strength than the preceding one. (See *Damped Waves*.)

DAMPED WAVES—Radio waves in the form of successive trains or groups, in each of which the amplitude or strength decreases with each successive wave. When electrical oscillations are caused by a single impulse, they do not continue indefinitely, but decrease in amplitude or die away. Thus, if oscillations are created in an antenna or other circuit by a discharge from a *condenser* (see spark discharge), each electric spark creates a train of oscillations which die away more or less gradually. Fig. 1 shows a pendulum which is given a momentary impulse. Now as this source of power is only momentary the pendulum will not swing from A to D indefinitely. If the first swing under the impulse carries the pendulum the entire distance A to D, the next swing will cover a lesser arc and so on until the movement has died out altogether.

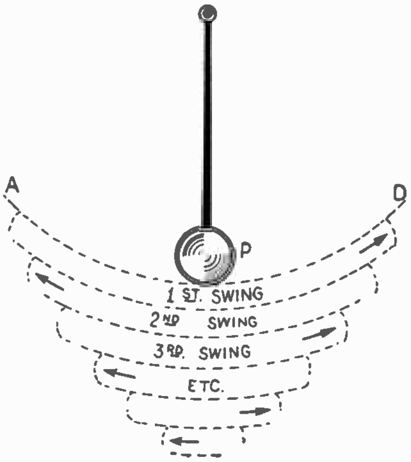

Fig. 1. Pendulum coming to rest after being given a momentary impulse.

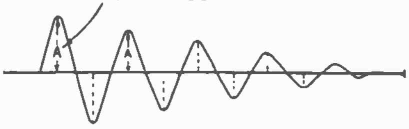

Fig. 2. Train of damped waves.

Now in Fig. 2, the graphic illustration of a train of *damped waves* shows that each crest is a little lower than the one preceding it and after a short period the waves have died out altogether. These curved lines may be considered as an illustration of the path taken by the pendulum in its motion as shown in Fig. 1. The decrease in amplitude of one crest over that preceding is known as the damping, and the naperian logarithm of the ratio of amplitude of any oscillation to the one preceding it is known as the *logarithmic decrement.*

If the impulse of force causing the oscillations is steady—not intermittent—the result will be a sustained series of oscillations that will continue at constant strength or amplitude until the exciting force is withdrawn. Thus, in the case of the pendulum, we can very easily keep exciting it by an evenly exerted force and so insure that the swings will continue without diminishing, as in the case of a momentary impulse. Oscillations or waves that have constant amplitude are known as *Continuous Waves, Undamped Waves* or *Sustained Waves.* (q.v.) (See also *Decrement, Logarithmic;* also *Naperian Base.*)

DAMPING—A progressive decrease in amplitude or strength. (See *Damped Waves, Damping in Instruments.*)

DAMPING COILS—*Short circuited* (q.v.) coils used in electrical apparatus for the purpose of causing damping (i. e., decrease in the intensity of electrical oscillations). The reaction of induced currents in the shorted coil creates the damping effect. (See *Damping.*)

DAMPING DECREMENT—See *Decrement, Logarithmic.*

DAMPING FACTOR—The product of the *logarithmic decrement* (q.v.) and the frequency (rate of alternation) of a damped alternating current. (See *Decrement,* also *Amplitude.*)

DAMPING IN INSTRUMENTS — The term damping is applied in a mechanical sense as in the case of the needle of a measuring instrument such as a *galvanometer.* (q.v.)

In a delicate instrument for measuring current the needle is controlled by the current through the instrument, and when this current is withdrawn the needle is brought back to the neutral position by a fine spring. Now, unless some precaution is taken the needle will have a tendency to swing or vibrate when measurement is being taken. Various devices are used to bring the needle quickly to rest without undue vibration. If the needle comes to rest almost immediately it is said to be a highly damped or *dead beat instrument,* because the oscillations or vibrations die out rapidly the same as in the case of damped waves.

There are two principal methods of damping instruments. When the instrument is of the moving coil type, the coil is generally wound on a light aluminum frame which is mounted between the poles of a permanent magnet. A soft iron core fills the gap between the two poles, allowing space for free movement of the coil. The result is that the metal frame carrying the coil moves through a very strong *magnetic field* and *Eddy currents* (q.v.) are set up in the frame and serve to damp its motion. The other common method is to have a light metallic vane carried by the pointer in its motion and moving in an air chamber to stop vibration. While damping is highly essential in a measuring instrument, too high a degree of damping is not advisable. If there is a very slight tendency on the part of the needle to vibrate, it permits the operator to determine whether it is swinging freely. (See *Voltmeter, Wattmeter.*)

DAMPING MEASUREMENT—See *Decrement, Logarithmic.*

DAMPING WAVES—Term commonly used to denote the gradual decrease in amplitude of a train of oscillations as in radio transmission. If the oscillations persist at constant amplitude or strength, each oscillation being equal in amplitude to the one preceding, there is no damping. When the amplitude of each successive oscillation is less than that of the preceding one there is said to be damping and the oscillations will die out entirely, at a rate dependent on the degree of damping. If the oscillations die out very rapidly they are said to be highly damped; if they die out slowly the damping is said to be feeble.

DANIELL'S CELL—A two fluid *voltaic cell* containing a zinc plate immersed in dilute sulphuric acid, and a copper plate in a saturated solution of copper sulphate; the two solutions being separated by a porous cup. This cell has a constant voltage and shows only slight polarization. Was formerly much used in telegraph work.

D'ARSONVAL GALVANOMETER—A type of galvanometer in which a coil carrying the current to be measured is suspended in a *fixed magnetic field.* (See *Galvanometer.*)

DASH—The long stroke used in *Morse* and *Continental Code.* It is considered as equivalent in length to three dots. (See *Code.*)

D. C. C.—The customary abbreviation for double cotton covered wire.

D. C.—The abbreviation commonly used for *direct current.* (q.v.)

"D" COIL—Occasionally known as Figure Eight coil. A form of inductance, used particularly in tuned radio frequency circuits. The windings are in the shape of a figure eight, either in pancake form (flat wound) or wound on a slotted tube. Owing to the peculiar arrangement of the windings these coils have very little external field and thus can be used in radio frequency amplifiers without the usual oscillating due to conflicting fields. (See *Field, Inductance, Radio Frequency Coils.*)

DEAD BEAT INSTRUMENT—A measuring instrument arranged so that the needle comes quickly to rest with little or no vibration is said to be *dead beat* or highly *damped.* (See *Damping in Instruments.*)

DEAD CELL—A *cell* which has been discharged. (See *Cell.*)

DEAD CRYSTAL—A crystal used in a *crystal detector* (q.v.) that has lost its sensitivity. Very often a crystal becomes dirty or oily from contact with the fingers. In some cases they can be brought back to usefulness by cleaning with alcohol, or if the crystal is dead through causes other than dirt on its surface, it may be broken to present a new surface and thus find a new sensitive spot.

DEAD END EFFECT—The effect on a receiving circuit (oscillatory circuit) resulting from unused turns of wire in a coil. In the customary tuning coil or tapped coil a certain portion of the wire is not in use and this idle section, being in metallic connection with the main portion, but not actually in use, acts as a miniature oscillatory circuit. This tends to reduce the efficiency of the tuning unit as a whole. Dead end effect may be eliminated by using the entire coil at all times, doing the tuning solely by the variable condenser, or it may be reduced by the use of a *dead end switch* (q.v.). (See *Tuning Coil.*)

DEAD END SWITCH—A switch used to cut out unwanted turns of wire in a coil in order to eliminate or reduce the *dead end effect* (q.v.). Such a switch is generally arranged so that instead of taps being taken from various points on the coil, the winding is actually in sections and the switch blade connects into the circuit the desired number of sections.

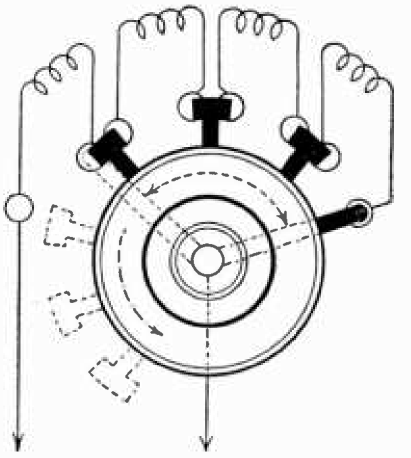

Dead end switch.

DEAD SPACE—In the case of *beat heterodyne* reception if the two frequencies are brought together or the difference is very slight there will be no beat. This space where the two frequencies are identical, or very nearly so, is known as the *dead space.* (See *Beats,* also *Heterodyne.*) The beat resulting from the superposition of two waves differing in frequency is the difference or the sum of the two frequencies. For instance, if two waves are present in the same circuit, of 1,000 and 1,005 kilocycles respectively, the two beat frequencies will be 5 and 2,005 kilocycles. If, however, the two frequencies are made identical, the difference between the two will be zero, and the sum of the two will be twice the original frequency. In the one case, where we take the difference of the frequencies, the beat note is zero, or there is no beat note at all; in the other case, where we add the two frequencies, the resulting frequency is the *harmonic* (q.v.) of the original frequencies, which is the same as an octave in music. In the present instance, this note will not be heard at all as it is above the range of audibility.

DECADENT WAVE—A damped wave. A wave, or oscillation, in a train of *damped waves* (q.v) which is of lesser amplitude or strength than the one preceding it. (See *Damped Waves,* also *Undamped Waves.*)

DECAY OF CURRENT—The gradual dying out of current in an *inductive circuit* (q.v.) after the *impressed potential* has been removed. The presence of inductance in a circuit causes the current to lag in time behind the voltage producing it. As a result, when a highly inductive circuit is broken by opening a switch or by other means, the current still tends to flow for a short interval, and the *induced voltage* in the circuit may become very high. This is the reason why a spark often occurs when the switch in a highly inductive circuit is opened. As a result the current decays gradually after the impressed voltage is removed.

It will be noted that the *decrement* (q.v.) or the rate of decay of the oscillations directly affects the number of oscillations in the wave train that will be useful in transmitting signals. For instance, if a certain oscillation in the figure should represent the smallest voltage that will operate a detector, it is evident that all oscillations in the wave train that have smaller amplitudes will be useless. Consequently, the lower the decrement of the wave, the greater will be the number of oscillations before the amplitude becomes too small for useful purposes. This likewise means that a greater amount of the energy in the wave can be utilized.

DECOHERER—In the type of detector, now obsolete, known as the *coherer*, the incoming signals caused the particles of nickel and silver filings in the glass tube of the device to cohere or cling together. (See *Coherer.*) In order to place the coherer in its original state after each impulse, it was necessary to have some means of separating the particles. These were known as *decoherers* and were in many different forms, some as separate attachments, others as a part of the detector or coherer. (See *Detector.*)

DECOMPOSITION, ELECTROLYTIC—The decomposition of chemicals taking place in a cell due to the action of the electrolyte. (See *Cell*, also *Electrolyte.*)

DECREMENT—A term used to indicate the rate of decay or dying out of an electrical oscillation that is subject to *damping.* (q.v.) In the illustration is shown a series of damped waves. It will be noted that the distance from zero to the crest of wave (C-2) is less than the one preceding it (C-1). This distance represents the amplitude and the ratio of any amplitude to the one preceding it is constant, that is the decay or decrease in amplitude is constant. The *Naperian Logarithm* (q.v.) of the ratio—of one wave to the one preceding it—is called the *logarithmic decrement.* The damping of a train of waves is an important consideration as it affects the tuning qualities to a great extent. This is covered by a U. S. Government Statute Concerning Transmitting Stations, decreeing that the logarithmic decrement per complete *oscillation* must not exceed 0.2, which means that for each single spark discharge from the transmitter there must be not less than 24 complete oscillations in the *antenna system.* (See *Oscillations, Damping, Naperian Logarithm, Sharp Wave.*)

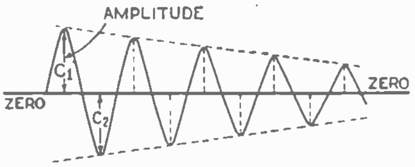

Damped waves gradually dying out.

DECREMETER—An instrument used to measure the *decrement* or degree of damping in an oscillatory current. Such meters are made in several different forms. Some types are direct reading in terms of the logarithmic decrement, while others require mathematical reduction from the readings of a dial. All decremeters operate on the basis of comparison of the resonance current in a tuned circuit with that in a circuit out of tune by a known or unknown percentage. (See *Logarithmic Decrement*, also *Damping.*)

DE FOREST, DR. LEE—Electrical engineer; b. Council Bluffs, Ia., Aug. 26, 1873. Grad. Yale (Sheffield Scientific), Ph. B., 1896; Ph. D., 1899. Inventor De Forest system wireless telegraphy and founder Am. De Forest Wireless Telegraph Co.; system originally adopted largely by U. S. Govt.; demonstrated to the British, Danish, German, Russian and Indian Govts., first used by the U. S. Signal Corps in the war manoeuvres and inst. in Alaska. 1905 invented the Audion or three-electrode vacuum tube. 1906-09 was the pioneer in Radio Telephony. First to broadcast music and opera. 1908 equipped the 24 battleships and destroyers of Admiral Evans for their historic round-the-world cruise. 1913 incorporated De Forest Radio Tel. & Tel. Co., which developed the audion amplifier and oscillator for commercial purposes. These De Forest devices completely revolutionized the radio art and made possible the Bell Tel. Transcontinental System. He was the father of radio broadcasting and all radio broadcasting apparatus is built upon his audion inventions. Since 1919 he has devoted his attention to the phonofilm, photographing sound waves directly on motion picture film. Pres. of the De Forest Phonofilms, Inc.; Dir. of De Forest Phonofilms, Ltd., London. 1921, Awarded Cross of Legion of Honor by the French Govt. in appreciation rendered by audion to the Allies during the war. 1922 awarded Elliott Cresson medal by Franklin Inst. 1922 awarded medal by Inst. of Radio Engineers. Fellow American Inst. of Elec. Engineers. Fellow Inst. of Radio Engineers. Mem. of N. Y. Elec. Society; mem. of Nat. Geog. Soc.

Dr. Lee De Forest.

DE FOREST COILS—A type of coil named after *Lee De Forest* (q.v.). The wires are wound in a diamond shaped pattern in such manner that the turns of one layer cross the turns of the preceding layer at an angle, thus making the *distributed capacity* (q.v.) very small compared with the capacity of other types of coils. (See *Honeycomb Coils.*)

DEGREES, ELECTRICAL—In an alternating current curve each complete oscillation or cycle is shown as two loops, one above the line representing time, and the other below it, the entire cycle being considered as 360 *electrical degrees.* It is possible in this manner to show relations between current and voltage phases, amplitude at any point, etc. (See *Alternating Current*, also *Sine Wave.*)

The idea of electrical degrees comes originally from the *elementary electric generator*, with two-slip-rings taking off the alternating current, and having only two poles. During one-half revolution of the armature turns, the current increases from zero to a maximum and then back to zero again. During the other half revolution the current increases in an opposite direction from a zero value to maximum and then back to zero again. The curve shown in the figure represents the current variation during the complete revolution, a crosssection of the rotating armature wire being shown at A. The angles indicated on the line BC are laid off to any convenient scale, and the intersection of the ordinates from these points with horizontals drawn from the circle as shown, give the points on the curve.

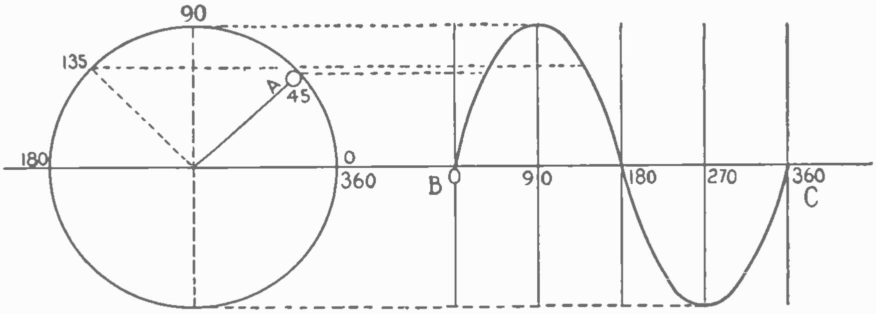

Curve representing current variation during one revolution of the armature of a generator.

The same ideas hold when the generator has more than two poles; in this case a full 360 electrical degrees is taken to represent the armature motion from across two adjacent pairs of poles, so that a complete cycle of voltage or current is generated. In a four pole machine, therefore, a full revolution of the armature could be represented by 2 x 360 or 720 electrical degrees. This corresponds to 2 complete cycles. Or, looking at it the other way, 360 electrical degrees would represent a half revolution of the armature wire.

DEIONIZATION—The action of returning a mass of gas that has become *ionized* (breaking up of the atoms of gas into positive and negative ions, being the elements to which and to whose motion, under the action of electric forces, is supposed to be due their electric conductivity) to its original state. In the case of a *spark gap* (q.v.) the ionization creates a conductive path which remains after the actual spark has passed, thus preventing sharply defined oscillations. There are various methods of deionization, including cooling, absorption, air-blast or magnetic field, all tending to deionize or place the spark gap in its previous condition to prevent an arc from following the discharge. This action is also known as *quenching.* (See *Quenched Gap.*)

DELTA—The fourth letter in the Greek alphabet. The capital letter delta (Δ) is used as a symbol for delta connection of *three phase alternating current generators or transformers.* The small or lower case delta (δ) is used as a symbol for *Logarithmic decrement.* (See *Decrement, also Delta Connection.*)

DELTA CONNECTION—Δ In a three phase *alternating current generator,* where three coils are mounted symmetrically around a shaft rotating in a *magnetic field,* the connection or grouping of the windings is called *delta* after the Greek letter which the grouping somewhat resembles. The illustration shows delta grouping in a three phase system. The sum of the instantaneous Electromotive Forces of the three coils is zero, or in other words the sum of the Electromotive Forces of any two coils is equal and opposite to that of the other coil, the *line voltage* thus being equal to the *phase voltage.* (See *Alternator, Polyphase.*)

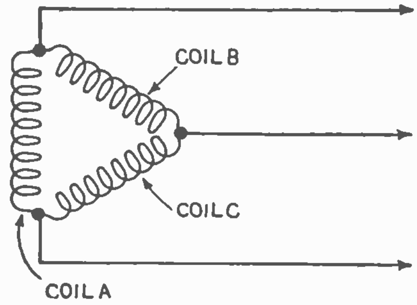

Delta grouping in a three phase system.

DEMAGNETIZATION—The act of returning a body to an unmagnetized state or of reducing the degree of *magnetization.* In the case of phones used in radio reception the permanent magnets are often weakened or *demagnetized* by connecting them in the *electron-tube circuits* in the wrong way. (See *Magnetization,* also *Telephone Receiver.*)

DENSITY, CURRENT—See *Current Density.*

DENSITY, FLUX—See *Flux Density.*

DENSITY OF ELECTROSTATIC CHARGE—The *electrostatic charge* per unit area. (See *Electrostic Charge,* also *Condenser.*)

DEPOLARIZATION—In an electric cell hydrogen bubbles form on the surface of the positive electrode and unless some arrangement is made to counteract this effect, the usefulness of the cell will soon be impaired due to the insulating action caused by this film of gas, known as *polarization.* Usually some means of oxidizing is employed to act upon the hydrogen as fast as it is produced on the positive electrode, thus preventing or reducing the effect of *polarization.* (See *Cell,* also *Polarization,* and *Depolarizer.*)

DEPOLARIZER—The oxidizing agent or other means used in a cell to counteract the effect of polarization. Bichromate of potash or peroxide of manganese is commonly used for this purpose. (See *Dry Battery,* also *Polarization.*)

DETECTOR—A device for converting oscillating currents of high frequency (radio waves) into a form suitable for operating a telephone receiver or sensitive measuring instrument. It is often referred to as a *rectifier,* because it serves to change the incoming currents from *alternating* to *pulsating direct currents. Detectors* vary in type and efficiency, ranging from the now obsolete *coherer,* to the modern sensitive *regenerative vacuum tube.* The different types of detectors are taken up under their various headings. (See *Crystal Detector, Vacuum Tubes, Electrolytic Detector.*)

DETECTOR CIRCUIT—That part of the circuit in a radio receiver which contains the *detector* (q.v.). The *detector circuit* may be closely coupled to the balance of the receiving circuit by means of a tuning coil or it may be inductively coupled by using a variocoupler or loose coupler. (See *Coupling*).

DETECTOR, DAMPED WAVE—Any special type of detector, such as a crystal, etc., used for reception of *damped waves.* (q.v.). (See *Tikker,* also *Heterodyne,* and *Vacuum Tube.*)

DETECTOR, VACUUM TUBE—The *vacuum tube* used as a *detector* of *radio frequency* (*high frequency*) *oscillations.* A vacuum tube may be used to change or rectify the high frequency alternating currents received at the aerial to *pulsating direct currents* (q.v.) capable of operating a telephone receiver or recording device. Formerly vacuum tubes were made for an individual purpose, such as *radio frequency amplifier, audio frequency amplifier, or detector,* the amplifier tubes being highly exhausted and the detector tube having a low vacuum and requiring critical filament adjustment. Of late tubes have been developed to such an extent that no special tube is necessary for the detector circuit, all being adaptable for almost any purpose. Fig. 1 shows the ordinary circuit for a vacuum tube detector. This method is not much used now since the advent of regeneration, except where stages of radio frequency amplification are used ahead of the detector to amplify the incoming signals. In this case incoming signals are impressed on the grid "G" and the slight variations of potential at this point control comparatively large currents at the plate "P," actuating the telephone "T." (See *Vacuum Tube, Theory of Detector Action.*)

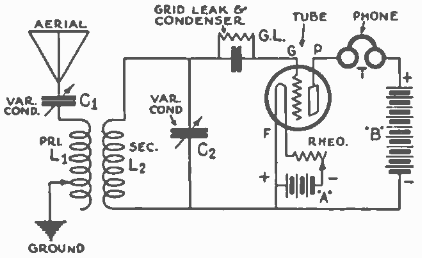

Fig 1. Non-regenerative vacuum tube circuit.

Used in this manner the tube is a considerable improvement over the crystal detector in point of sensitivity. However, by means of regeneration, the sensitivity may be increased many times. Fig. 2 shows a simple *regenerative circuit.* Here the unrectified currents passed through to the *plate* of the vacuum tube are fed back by means of the coil L2 and by being impressed again on the *grid,* further amplification can be attained. (For complete explanation see *Regenerative, Theory of Operation of Vacuum Tube, Detector, Pulsating Direct Current,* also *Crystal Detector.*)

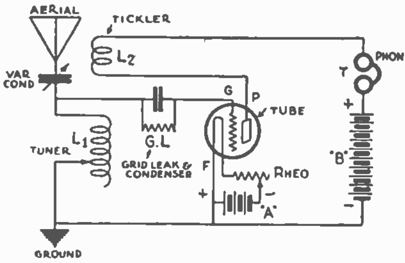

Fig. 2. Simple regenerative circuit.

DETUNING—The opposite of *tuning* (q.v.) The process of varying the effective inductance or capacity or both, to throw the radio receiver out of resonance with the particular signals to which it is tuned. This may be done to decrease the volume of the signals or it may be employed to eliminate or reduce interference from some undesired signals. For example, a station operating with a wavelength of 400 meters might interfere with reception of a station operating on 360 meters, providing it has sufficient power, or the receiver is not particularly selective. If the radio receiver is detuned slightly below 360 meters, the desired signals will still come in, though at lesser volume, but the interfering signals will be lost entirely. Tuning is the act of *producing* resonance and detuning is the act of *destroying* resonance. In the case of a *heterodyne receiver* (q.v.) the heterodyne circuit is slightly detuned from the incoming oscillations, thus producing a difference in frequency or *beat frequency* (q.v.). (See *Resonance,* also *Dead-space.*)

DIAGRAM—A system of lines drawn to represent the circuit or connections for radio receivers or transmitters or associated apparatus, etc. Diagrams may be in any of several forms, the two common methods being known as schematic and perspective. In the schematic form, symbols are generally used to represent the various pieces of apparatus, whereas in the perspective form, drawings of the apparatus are used. Fig. 1 illustrates a typical schematic diagram of a simple crystal receiver. Here the tuning coils, de-

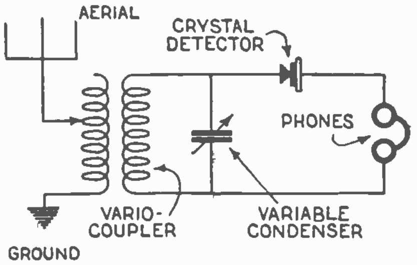

Fig. 1. Schematic diagram of a simple crystal receiver.

tector, phones and so on, are shown in the form of symbols. Fig. 2 shows the same circuit in perspective form, with the tuning coil and other parts actually pictured. Diagrams may be used to show connections for any electrical or radio circuits or for individual parts of instruments. (See *Hookup.*)

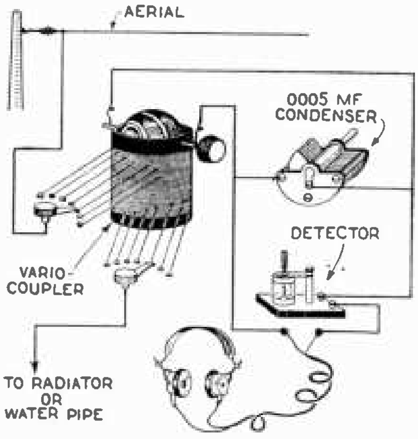

Fig. 2. Showing in perspective form the circuit given in Fig. 1.

DIAL—Devices used to control the movement of condensers or other moving parts of radio apparatus. Most dials are made of a composition such as Bakelite or hard rubber; some are made of metal. The periphery is generally spaced off in numbers from 0 to 100 or 180, for purposes of keeping a record of the positions at which the various stations are received. Dials may be arranged to have the readings refer to definite values by means of *calibration* (q.v.). (See *Dial Vernier.* See *Log.*)

DIAL, VERNIER—A term incorrectly applied to a dial arranged with slow-motion linkage in order to permit very fine adjustment. The illustrations show two types of vernier dials. Vernier dials are generally used on *variable condensers* and occasionally on other instruments, where close adjustment is necessary. They are particularly useful in the case of a *Super-Heterodyne receiver*, the oscillator control being critical, and the wavelength control as well, in certain cases. (See *Dial*, also *Critical.*)

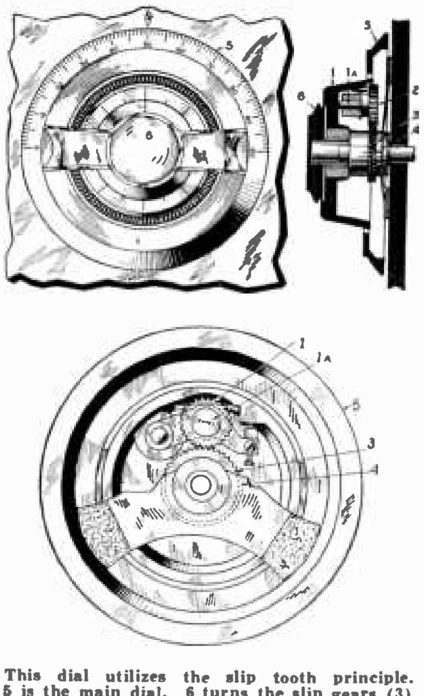

This dial utilizes the slip tooth principle. 5 is the main dial. 6 turns the slip gears (3) and (4), which cause 2 to turn. 2 is mounted on the shaft (1A) fastened in the main dial (1). 5 is the scale, fastened to the condenser shaft.

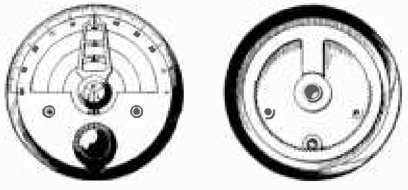

A simple vernier dial using a pinion and internal rack.

DIAMAGNETIC MATERIAL—Substances not readily susceptible to *magnetism.* If a substance does not pass magnetic lines of force, but rather is repelled by a magnetic pole, it is said to be *diamagnetic;* its permeability, or ability to permit magnetism to pass through it, is considered negative or less than unity (air). Diamagnetic substances are apparently repelled from the poles of a magnet. If placed in a magnetic field they have a tendency to diminish the *magnetic induction* (q.v.). (See *Paramagnetic,* also *Permeability.*)

DIAPHRAGM—A thin disk, generally of soft iron used in *telephone receivers* and *microphones* (q.v.) to either produce or detect electrical pulsations by means of vibrations caused by an *electromagnet.* Variations in the current through the electromagnet cause corresponding vibrations in the diaphragm. While the common type of diaphragm is made of soft iron, there are numerous other materials used. Certain alloys have higher permeability, and therefore are more responsive to the attraction of the magnet. In one type of telephone receiver that is particularly sensitive, there are two diaphragms—one that is actuated directly by the electromagnet under the influence of a varying current and another of thin mica, coupled to it by a fine wire. Another type employs a small iron armature suspended on a pivot, and attached to a mica diaphragm by means of a delicate lever, at right angles to the armature. The vibrations of the armature in the magnetic field of the magnets are communicated to the diaphragm by means of this lever. (See *Baldwin Receiver.*) Another type is of metal with fine corrugations and still another is a mica or parchment disk with a piece of soft iron in the center. (See *Microphone, Telephone Receivers.*)

DIAPHRAGM, CARBON—A phone diaphragm in the form of a thin carbon disc. (See *Diaphragm.*)

DIELECTRIC—This term is rather broadly used in electrical work to indicate a non-conductor or *insulator.* A dielectric may be a solid, liquid or gas and is generally employed to separate two conducting surfaces. The most common instance of the application of a *dielectric* is in the case of a condenser. Here the plates are separated by a dielectric material, the nature of this material depending on the particular type and purpose of condenser. A *variable condenser* for receiving purposes generally employs air as the dielectric. That is, the two sets of plates are kept from actual contact while one is being rotated, by an air space. In *fixed condensers* the dielectric is usually waxed paper, mica, glass, oil or compressed air, depending on the purpose for which the condenser is to be used and the dielectric strength necessary. While we refer to a dielectric as a non-conductor or insulator, there is actually no dividing line between conductors and insulators, practically any dielectric permitting passage of a certain amount of current under the proper conditions. If a substance is a very good insulator it is said to have good dielectric properties, while one that is a poor insulator and permits leakage of current is a correspondingly poor dielectric. (See *Insulator, Conductor, Resistance,* also *Dielectric Coefficient and Constant.*)

DIELECTRIC ABSORPTION—The tendency of the dielectric in a condenser to apparently absorb a certain amount of the power applied to it. When a condenser is charged by a direct current source such as a battery, the instantaneous charge is often followed by a small charge or flow of current that steadily decreases as it flows into the condenser. This additional charge is in effect absorbed by the dielectric of the condenser. The reverse effect is obtained when the condenser is discharged, the instantaneous discharge being followed by a steadily decreasing additional discharge. It will be apparent, then, that if the condenser is charged from an *alternating current* source, such as by *radio frequency currents,* the dielectric will show a tendency to withhold a certain part of the charge, and as the charging and discharging goes on rapidly due to the rapid reversals of the current (see *High Frequency Alternating Current*), the dielectric will continue to hold back this portion of the power as long as the charging and discharging process persists.

This loss, due to the absorption in the dielectric, must not be confused with the losses due to faulty dielectric materials, wherein a certain portion of the current is allowed to leak away by conductance through or on the surface of the insulator. *Dielectric absorption* is often referred to as *dielectric viscosity* or *hysteresis* because of its similarity to viscosity in liquids. Viscosity is the property possessed by liquids to resist deformation. The usual method of measuring viscosities is by measuring the time taken by a known volume of the liquid, at a known temperature, in flowing through an aperture of known form and dimensions under a known pressure. Thus tested, water will flow rapidly, while cylinder oil is very sluggish, and hence is said to possess great viscosity. (See *Condenser,* also *Dielectric.*)

DIELECTRIC COEFFICIENT and CONSTANT—The specific *inductive capacity* of a dielectric. Generally speaking, its properties to act as a *dielectric.* In the *Centimeter Gram Second system* (q.v.) the *inductive capacity* and the *dielectric constant* are numerically equal, the constant being the dielectric value of the material as compared with air at ordinary pressure taken as the standard (1). The table of constants for the more important dielectric materials follows.

It will be seen that glass, oils and mica have the highest values, for which reason they are widely used. (See *Specific Inductive Capacity, Dielectric,* also *C. G. S.*)

TABLE OF DIELECTRIC CONSTANTS

Dielectric	*Constant*
Air at ordinary Pressure (Taken as the Standard)	1.000
Manila Paper	1.50
Celluloid	1.555
Parrafine (clear)	1.68 to 2.32
Beeswax	1.86
Parrafine Wax	1.9936 to 2.32
Parrafined Paper	3.65
Hard Rubber (Enonite)	2.05 to 3.15
India Rubber (pure)	2.22 to 2.50
Gutta Percha	2.46 to 4.20
Shellac	2.74 to 3.60
Olive Oil	3.00 to 3.16
Glass (Low Frequency value)	3.25 to 4.00
Glass (High Frequency value)	4.21
Mica (Pure Sheet)	4.00 to 8.00
Porcelain	4.38
Castor Oil	4.80
Flint Glass, very light	6.57
Flint Glass, light	6.85
Flint Glass, very dense	7.40
Flint Glass, double extra dense	10.10

DIELECTRIC CONSTANTS—See *Table of Constants* under *"Dielectric Coefficient* and *Constants."*

DIELECTRIC HYSTERESIS—When the *electric field* in a dielectric material is varied rapidly, as when the condenser is charged by *high frequency currents*, heat may be generated. This is due to *dielectric hysteresis which is* synonymous with *dielectric absorption* (q.v.)

DIELECTRIC STRENGTH—When an electric field is established in a dielectric, that is to say, a charge is applied to a condenser, and the field attains a certain intensity, the dielectric ceases to be an insulator and becomes, in effect, a conductor. This condition generally is accompanied by a puncturing of the dielectric material. When the charge applied to a condenser is too high the spark will burn through the dielectric and the condenser is then said to have broken down. In some cases the condenser may be permanently ruined by this action, and in others, the dielectric may be conductive at a certain voltage, while remaining effective as long as the voltage is held below the safety point. The particular voltage (critical voltage) or field intensity at which the breakdown of the dielectric occurs is called the dielectric strength of the material. (See *Dielectric*, also *Table of Dielectric Strength.*)

DIFFERENCE FREQUENCY—The frequency of oscillations produced by superposing oscillations of one frequency on oscillations of a different frequency. In Super-heterodyne receivers the action is based on the production of a difference or beat frequency. If the incoming oscillations from the antenna are combined with oscillations of a different frequency produced locally in the receiver, a new series of oscillations will be produced, these oscillations having a frequency numerically equal to the difference in frequency between the other two sets of oscillations. (See *Beat Frequency*, also *Super-Heterodyne.)*

DIODE—A thermionic vacuum tube having only two *electrodes*, namely *filament* and *plate*. The original vacuum tubes were known as *Fleming Valves* and contained a hot filament and cold plate but no grid as found in the three element vacuum tube now in general use. Several types of *diode* are still in use as detectors, their chief value lying in the fact that they require no careful adjustments as in the case of a *crystal detector.* (See *Triode, Vacuum Tube,* also *Fleming Valve.)*

DIPLEX RECEPTION OR TRANSMISSION — The simultaneous reception or transmission of two series of signals by or from a single operating station. The systems are so arranged that two messages may be sent or received at the same time without interfering with each other. (See *Duplex Signalling.)*

DIRECT CONDUCTIVE CIRCUIT—Any circuit having a direct conductive path and not depending on capacity or electromagnetic coupling. A metallic, conducting circuit.

DIRECT COUPLING—The coupling or relation between two or more coils or circuits, wherein the connection is *metallic*. In the illustration Fig. 1 is shown a two slide tuning coil. In this case A-B is the primary or aerial circuit and C-D the secondary or detector circuit. The two circuits are joined together by metallic connection.

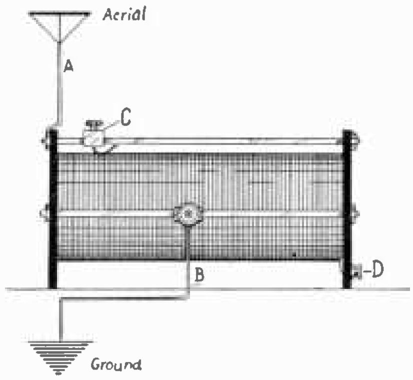

Fig. 1. Double slide tuner which affords direct coupling between the primary and secondary circuits.

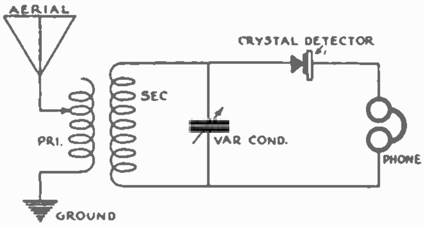

Fig. 2. Circuit in which primary and secondary are inductively coupled.

The illustration Fig. 2 shows essentially the same circuit, except that the primary and secondary are two separate coils. Here the two coils or circuits are joined by *inductive* coupling.

DIRECT COUPLED—Circuit in which primary and secondary circuits are metallically connected. Generally permits little selectivity. For this reason, *inductive coupling* is more generally used to permit close tuning or selectivity (q.v.). (See *Coupling, Inductive Coupling.)*

DIRECT CURRENT—See *Current, Direct.*

DIRECT READING—A term applied to various measuring or recording instruments which are so arranged as to show values directly without requiring mathematical reduction. *Voltmeters, Ammeters* and such instruments, used widely for showing the *voltage, current*, etc., of *batteries* or *cells*, are direct reading, showing the values directly in *volts* or *amperes*. A *decremeter* or *wavemeter* very often shows a reading that requires application of a formula to produce the desired quantity. (See *Calibration.)*

DIRECTION FINDER—See *Radio Compass.*

DIRECTIONAL—The effect of an aerial wherein waves are transmitted better, or entirely, in a certain direction, or received from a certain direction, depending upon the type and direction of aerial. (See *Transmitting Aerial*, also *Receiving Aerial.)*

DIRECTIONAL EFFECT OF ANTENNA—An effect of an antenna (aerial) for reception or transmission wherein signals from a certain direction are more readily received than from another direction, or in the case of transmission, the range being greater in a certain direction.

DISCHARGE—Generally speaking, a comparatively sudden passage of electricity. The term however, as applied to a *storage battery* will mean merely the effect of releasing the *electrical charge* stored up in it, in this case the discharge not being necessarily rapid. A *condenser* has the ability to hold charges of electrical energy and to release them suddenly when the proper contact is made. The term "discharge" as applied to a *transmitter* indicates the passage, usually in the form of a spark or succession of sparks, of the electrical energy across a gap between *electrodes*. The illustration shows spark discharge from spark coil. (See *Spark Discharge, Storage Battery,* also *Discharger.)*

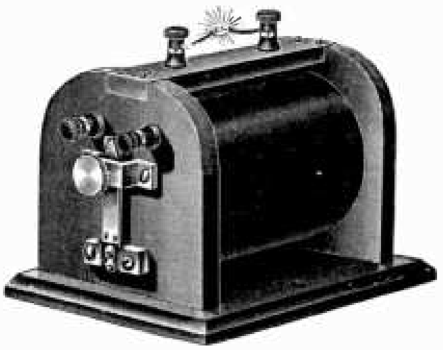

Discharge from a spark coil.

DISCHARGER—Any device allowing a path for an electrical discharge. The term may be considered roughly a synonym for *spark gap*. Usually two or more *electrodes*, either stationary or rotary, spaced a short distance apart to permit the released electrical energy to bridge the space and thus complete an *oscillatory circuit*. (See *Disc Discharger Rotary Gap.)*

DISC DISCHARGER—A form of *discharger* employing one or two rotating discs carrying the sparking surfaces. With the advent of modern undamped transmission methods, these dischargers are gradually passing out. (See *Synchronous Discharger.)*

DISPLACEMENT CURRENT—A current which flows for a short interval in an insulating material or *dielectric* when an *electromotive force* is impressed across it, or when the intensity of the electromotive force impressed across it is increased or decreased. This *displacement current* will flow only when the impressed electromotive force is altered in intensity. After the initial current due to the sudden change in the electromotive force being impressed across the material, the dielectric material or insulating material will remain in a state of strain as long as the *charging force* persists without further change in intensity, and no further displacement current will flow. (See *Current*, also *Dielectric.)*

DISPLACEMENT, PHASE—See *Phase Displacement*, also *Phase Angle*.

DISRUPTIVE VOLTAGE—The voltage sufficient to disrupt or break down a sample of *dielectric* material under given conditions. Known also as *breakdown potential*. For example, if a *condenser* will stand an impressed potential or voltage of 500 volts without injury, but will be punctured by a potential of 1000 volts minimum, the disruptive voltage is said to be 1000. (See *Break Down Potential.*)

DISSOCIATION THEORY—A theory advanced by Arrhenius in 1887, explaining *electrolytic conduction* by the assumption that a substance in solution is dissociated or separated into *positive and negative ions*, these ions carrying their respective charges in opposite directions. (See *Electrolytic Action.*)

DISSONANCE—"Discord; disagreement"—Webster. The antonym of *resonance*. A term broadly applied in radio to indicate lack of resonance, or coordination of signals or impulses. When an alternating current is superimposed on another alternating current of different frequency, the resulting lack of resonance—*dissonance*—is known as a *beat*. (See *Distortion.*)

DISTORTION—Lack of purity or faithfulness in the reproduction of a vibration or series of vibrations. The most common application of the term in broadcast reception is in the case of reproduction by a loud-speaker. It will very often be found that music or speech is not perfectly reproduced, due to any of a variety of causes. Vacuum tubes may themselves cause distortion; too high potential applied to the plate of a tube may result in distortion; many transformers used in audio frequency amplifier circuits may distort notes of certain frequencies, or it may be due to *self-oscillation* or *regeneration* in the receiving set. This effect is also caused by inefficient transmitting (broadcasting) apparatus, or by lack of careful adjustment. The transformation of speech or music into electrical impulses and its subsequent transmission and reception are attended by many difficulties.

The control of wave form in broadcasting is known as *modulation*. (q.v.). If the transformation of speech or music into electrical impulses and its propagation into space is accomplished without materially changing or distorting the wave form from the original voice or music vibrations, the problem of accurate reproduction is entirely dependent on the receiving apparatus. Distortion may often be traced to the diaphragm of the loud-speaker, certain types being more efficient in this respect than others. (See *Loud-speakers.*) If the ordinary type of disc diaphragm is used, it may become bent and thus cause inequalities in the vibrations, producing distortion of the music or speech being reproduced. (See *Wave Form, Modulation*, also *Amplifier.*)

DISTRESS SIGNAL—At the International Radio Telegraph convention held at Berlin, Germany, in July, 1908, the call letters C Q D, established by the Marconi Co. in February, 1904, as the official distress call, were superseded by the letters S O S as the Marine distress signal.

S O S is the International distress call for ships and airships requiring assistance. The letters have no particular significance, being chosen mainly for their distinctive sound. In the International Morse code, S O S is composed of three dots, three dashes and three dots, thus: . . . — — — . . . an unusual combination which permits easy recognition among other messages and calls. When distress signals are heard, the nearest government transmitting station generally sends out immediate notification to all broadcasting stations in the vicinity to suspend operations until the ship has been located and assistance rendered.

DISTRIBUTED CAPACITY—The condenser effect in a coil of wire. Any coil of wire possesses *inherent capacity* to a certain extent, depending on its particular shape, size, etc. In the case of a coil of wire wound on a cylindrical form as indicated in the illustration Fig. 1, the passage of current through the coil sets up a *local magnetic field*, but also an *electrostatic field*, the latter being in the form of lines of force perpendicular to the conductor. Thus in effect, each two adjacent turns of wire act as a miniature condenser (Fig. 2).

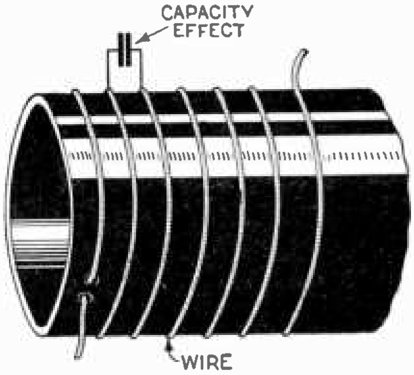

Fig. 1. Coil showing effect of distributed capacity between winding.

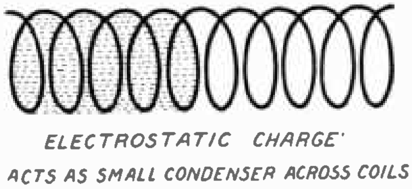

Fig. 2.

There are numerous methods and forms for winding coils in order to do away as much as possible with this effect. (See *Low-Loss Coils.*)

DISTRIBUTED INDUCTANCE—In a long cable or any great length of wire used for electrical purposes, additional inductance (wire) distributed throughout the entire length to compensate for the inherent capacity of the line. (See *Distributed Capacity.*)

DOT—The short signal in the International Morse Code. (See *Code.*)

DOUBLE AMPLIFICATION CIRCUIT—Also known as *dual amplification circuit*. The arrangement whereby a vacuum tube or several tubes may be made to do double duty, acting as both *radio frequency* and *audio frequency amplifiers*. (See *Reflex.*)

DOUBLE COTTON COVERED WIRE—Abbreviation D. C. C.—Cotton covered copper wire widely used in radio work for coils and all forms of inductances. The insulating covering is composed of two distinct layers of cotton, wound on in opposite directions to prevent loosening. Such wire is obtainable in a variety of sizes according to the standard gauges. (See *Wire Gauge.*)

DOUBLE FREQUENCY OSCILLATIONS—Sometimes referred to as a *double humped wave*. An irregular wave resulting from two frequencies, generally due to too close coupling of the *oscillation transformer*. When the coupling between the open and *closed oscillatory circuits* is too tight, the *open circuit* very often oscillates at two frequencies, resulting in the radiation of an irregular or *impure wave*. The illustration Fig. 1 shows in graphic form, such an irregular wave. The two peaks of resonance explain the term double humped wave. Fig. 2 shows another curve illustrating two frequencies.

In this case however, the amplitude of the second wave is considerably less than that of the first or main wave. According to U. S. Government regulations, if the amplitude of the lesser wave is not more than 0.1 of the amplitude of the main wave, it is said to be pure. In other words, where there are two frequencies, but the strength of one less than one tenth that of the other, it is considered negligible as it cannot be expected to cause any appreciable interference. (See *Resonance, Amplitude, Decrement*, etc.)

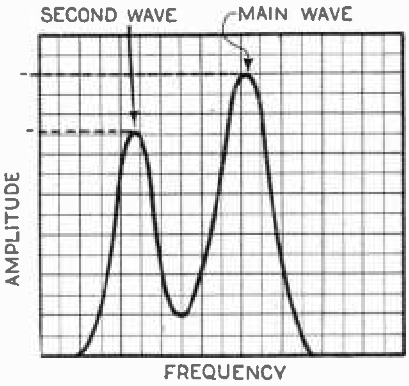

Fig. 1. Resonance curve showing second wave of interfering nature.

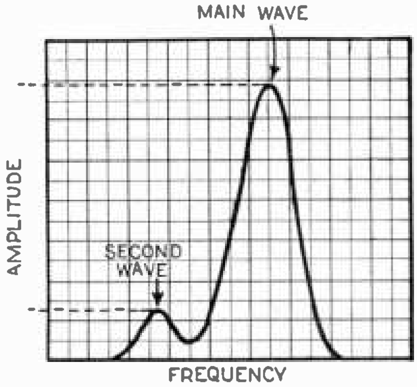

Fig. 2. Resonance curve showing second wave of negligible proportion.

DOUBLE GRID TUBE—A vacuum tube having two distinct *grid* members in addition to the usual *filament* and *plate* elements. Such a tube is used in circuits where the customary "B" battery is not employed, the extra grid acting as a booster for the electron flow, the "A" battery furnishing a small positive potential to the plate. (See *Solodyne*, also *Theory of Vacuum Tubes.*)

DOUBLE MODULATION—Successive modulation of a *radio frequency alternating current* at two lower frequencies. The *intermediate frequency* is usually above *audio frequency* and the lowest frequency is customarily within audio limits or a combination of audio frequencies, as in the case of *radio* *telephony* (broadcast). (See *Modulation.*)

DOUBLE POLE SWITCH—A switch used in electrical practice and in radio installations having two poles or connections, thus permitting both sides of a circuit to be opened or closed simultaneously.

DOUBLE RANGE METERS—Meters used for electrical measurements, arranged to read to two scales. As an example, the voltmeter shown has two scale readings, one in fine degrees for a maximum of 7.5 volts, the other a coarser scale to a maximum of 150 volts. Three binding post connectors are furnished, one being a common positive for both scales. A meter for almost any measurement might be made to have double range, although such instruments are generally confined to measurement of current in amperes and pressure in volts—i. e., ammeters and voltmeters. (See *Meter.*)

Photo by Courtesy of Weston Electrical Instrument Co.

Double range meter, one range to 7½ volts, higher range to 140 volts maximum.

DOUBLE SLIDE TUNER—A tuning coil provided with two sliding contacts, generally used in crystal receivers.

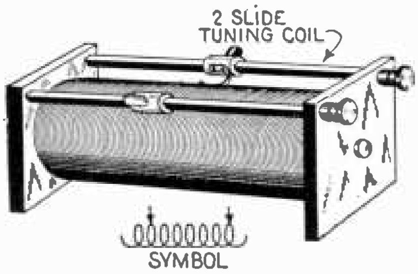

Two-slide tuner.

The illustration shows a common type of two slide tuner. (See *Coupling Crystal Receiver*, also *Tuning.*)

DOUBLE THROW SWITCH—A switch so arranged that a circuit or a certain instrument is connected in either of two different positions by throwing the switch lever. The illustration shows a single pole double throw switch. (See *Switch.*)

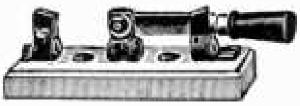

Double throw single pole switch.

DOWN LEAD—More commonly known as *lead-in*. The wire running from the aerial to the receiving or transmitting apparatus, whereby the signal energy is either collected from, or fed to, the aerial. The illustration shows a typical down lead or lead-in for a receiving set. Obviously the chief difference between such a lead for receiving purposes and for connecting a transmitter to the antenna is the matter of insulation. The comparatively feeble impulses received or collected by the aerial when receiving require only ordinary care, whereas the high voltages used in transmitting require careful insulation and much heavier apparatus all around. (*For more complete details regarding insulation, etc., see Aerial.*)

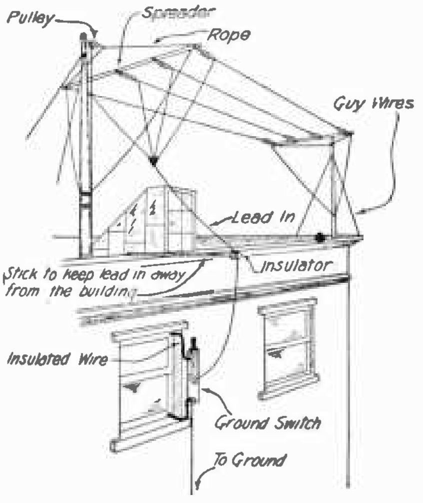

A typical 4-wire aerial showing lead in and method of installing lightning switch.

DRIFT, AVERAGE ELECTRON—The assumed rate of flow or drift of *electrons* under average or specific conditions. (See *Electronic Flow.* Also *Current, Assumed Direction of Flow.*)

DRIVER—A term broadly used to denote any system used to produce *oscillations* (vibrations) of a local nature. More specifically any means of producing oscillatory currents as used to test or make measurements in radio circuits. A buzzer connected inductively to a circuit to produce oscillations in that circuit for the purpose of making measurements of *capacity, wavelength*, etc. Another means might be a circuit involving a vacuum tube and the necessary apparatus to make it oscillate at a given or variable frequency. The term may be used to signify the tube circuit used to produce local oscillations in the case of a *super-heterodyne* (q.v.). Here the driver produces a series of oscillations of a frequency different from the incoming oscillations, the difference between the two being known as the beat frequency. (See *Heterodyne, Local Oscillations, Buzzer, Excitor.*)

DRIVER CIRCUIT—The circuit of the apparatus used to produce oscillations for purposes of test or measurement, or, in the case of a super-heterodyne, to produce a beat effect. (See *Driver.*)

DRUM ARMATURE—A form of armature winding in the approximate shape of a drum. (See *Armature.*)

DRY BATTERY—A battery or group of cells not employing a liquid electrolyte, the cell being filled with a mixture of carbon, manganese dioxide and sawdust (or other absorbent) saturated with a solution of sal ammoniac, these forming a paste, as distinguished from Wet Battery. (See *Battery*, also *Cell* and *B Battery.*)

DRY CELL—See *Dry Battery.*

D. S. C.—The abbreviation for "Double Silk Covered" as applied to copper wire having two distinct layers of silk, wound generally in opposite directions. This wire is furnished in standard sizes according to the various wire gauges. (See *Brown & Sharpe Gauge.*)

DUAL AMPLIFICATION—The process of obtaining both *radio frequency* and *audio frequency amplification* from the same tube instead of using two separate tubes. (See *Double Amplification Circuit*, also *Reflex.*)

DUBILIER, WILLIAM—President and technical director; born, New York, July 25, 1888. Educated New York Schools, Technical Inst. and Cooper Union; Chief Engineer of Continental Wireless Tel. & Tel. Co.; 1910 President of Com. Wireless Tel. Co.; at present Technical Director of Dubilier Condenser and Radio Corp. of New York, The Dubilier Condenser Company Ltd. of London, The Deutsche Dubilier Kondensator Gesellschaft in Berlin, La Protection Electrique Capart-Dubilier in Paris. Has obtained over three hundred patents and applications of electrical devices, which have been purchased or licensed by many companies. Member American Institute Electrical Engineers, Inst. of Radio Engineers, Honorary Member of Societe Academique D'Historie Internationale. Member of Royal Society of Arts.

DUCON—Trade name of a device which can be fitted to any electric light socket and made to serve in place of the usual outdoor aerial. This attachment consists essentially of two condensers, so arranged as to prevent the passage of any direct current or low frequency alternating current from the lighting main to the receiving set, but at the same time to permit passage of the incoming radio signals, the electric light line acting, therefore, as an aerial. (See *Adapter, Aerial.*)

DUDDELL SINGING ARC—Also called *Musical arc*. An *arc* actuated by a source of direct current through a *resistance* and shunted by a *condenser* and *inductance* in series. An *oscillating current* is thus produced in the condenser circuit, the result being a singing note corresponding in pitch to the frequency of the oscillations in the condenser circuit. The schematic arrangement is shown by the illustration. (See *Arc Generator.*)

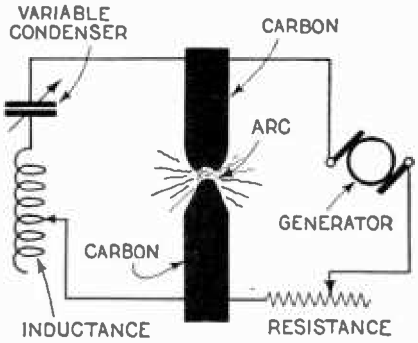

Schematic arrangement for operation of Duddell singing arc.

DULL EMITTER—The English term for vacuum tubes having *thoriated filaments* and operating with low current consumption. While the term is generally used to denote the tubes requiring only one or two dry cells to operate, it may apply as well to storage battery tubes where the current requirements are low. (See *Filament, Thoriated.*)

DUOLATERAL COILS—An alternate term for *honey-comb* coils; inductances wound in diamond-shaped layers to reduce the distributed capacity effect. (See *Honey-Comb Coils.*)

DUPLEX SIGNALLING or DUPLEX TELEGRAPHY OR TELEPHONY—The simultaneous transmission of signals or telephony in both directions between two stations. (See *Diplex Reception or Transmission*, also *Telephony.*)

DUST CORE—A form of core for certain types of transformers, employing iron filings or dust in place of the customary iron wire or laminations. Iron dust is used occasionally where simplicity of construction is desired, the assembly of such a core being obviously much easier than with wire or laminated types. (See *Core.*)

DX—The popular term for long distance; referring to the transmission to or reception from distant points of radio signals or broadcasting.

DYNAMIC CHARACTERISTIC—The curves obtained by impressing an alternating E. M. F. (electromotive force) on the grid of a vacuum tube, as distinguished from the curves obtained by application of a steady direct potential. When alternating potentials such as incoming signals are impressed on the grid of a tube, the curve resulting is apt to take an entirely different form from that of the curve showing *static characteristics* (q.v.). (See *Vacuum Tube characteristics.*)

DYNAMIC CONDENSER—A term occasionally applied to a *synchronous motor* used to improve power factor (in alternating current circuits, the ratio of the electric power in *watts* to the apparent power in *volt amperes* is known as the power factor). Used in this manner the motor has the effect of advancing the current phase in a manner similar to that of a condenser. (See *Angle of Lead or Lag.*)

DYNAMIC ELECTRICITY—A term sometimes used for electric currents to distinguish them from *static electricity* (q.v.).

DYNAMO—A machine for converting mechanical energy into electrical energy. Is also known as *generator*, but usually confined to machines for generating *direct* current. The action of a dynamo is based on the production of an *electromotive force* in a conductor moving in a magnetic field. Fundamentally a dynamo is a machine that generates alternating current, but instead of collecting this current by means of rings, which would cause alternating current to flow in the external circuit, a device known as an *armature* is employed in such a manner that the current collected by brushes and delivered to the external circuit is direct—that is, it flows more or less steadily in one direction. Thus the kind of current delivered for use depends on the method used to collect the currents generated. (For more complete explanation of the production of alternating current see *Alternating Current, Theory of Production.*) In the case of a dynamo, the essential parts are the *field magnets*, the *armature* and the *commutator*. Now in the case of an *alternator*, each end of the loop or armature is connected by a brush bearing against a collector ring and the end of each loop is always connected to the same brush. This *results in an external current that changes its direction with each reversal of direction of the induced electromotive force.* If the external current is to be direct it is necessary to have some means of collecting the current through one brush at the positive instant and the other at the negative instant. This is done by means of an armature, which is in effect a switching arrangement, so designed that it will reverse the connections of the external circuit at the instant of each reversal of current in the armature. The illustration shows a *shunt* wound dynamo. MM are the field magnets, FF the field windings, BB the brushes, A the armature and C the commutator. (See *Alternator, Armature, Generator.*)

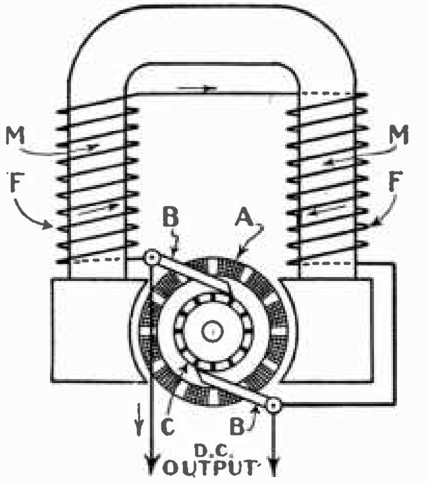

Shunt wound dynamo.

DYNAMOMETER—A delicate and accurate instrument for the measurement of currents and voltages or both. Such instruments depend upon the action of a circuit carrying current upon another carrying the same current. Essentially it comprises two coils, one fixed and the other movable. This system of measurement is used in a *wattmeter*, where it is necessary to measure the instantaneous current and voltage, i.e., power. (See *Wattmeter*, also *Electro Dynamometer.*)

DYNAMOTOR—A direct current machine which acts either as a *motor* or *dynamo*. It has an armature with two separate windings and two separate commutators, one at each end of the armature. Either winding may be used as a motor and the other as the generator winding. Such a machine is used to convert high voltage direct current into low voltage direct current or vice versa, thus performing the same function with direct current as a power transformer performs for alternating current. (See *Converter.*)

DYNATRON—A form of vacuum tube generally used for producing oscillations as in radio telephony, wherein the phenomena of "*secondary electron emission*" is used. In this form of tube the electrons traveling at high speed from the hot filament or cathode are made to collide with a metallic surface. The collision of these electrons on the surface of the interposed element has the effect, under proper conditions, of jarring other electrons out of the metal. The secondary emission thus obtained depends upon the speed of the original electrons which collide with the metal surface. Normally, these secondary electrons will immediately re-enter the surface from which they were emitted, but if another electrode of higher potential is in the vicinity they will travel toward it in the same manner as the electrons are attracted to the plate of an ordinary three element tube. A standard tube can be connected as shown in the illustration to act as a dynatron. Here the filament F emits electrons, some of which pass through the grid G and collide with the plate P. This collision with the plate P may jar loose from the surface additional electrons, which normally would immediately re-enter P. However, as the grid G is held at a higher potential (higher positive voltage) than P, the electrons will be drawn toward it. A tube arranged in this manner may be used for practically any of the purposes of the standard tube, such as regenerative detector, detector of continuous waves or as a generator of high frequency oscillations, but it has not been shown to have any advantages in this respect, being used mainly as an *oscillator* (q.v.). (See *Electron Emission, Vacuum Tube*, etc.)

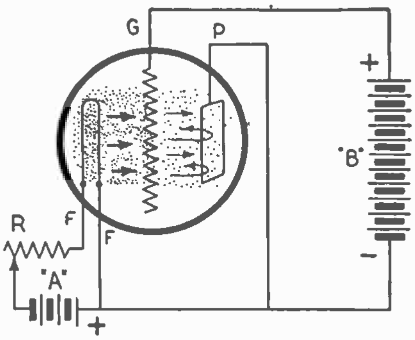

Showing how a standard vacuum tube can be connected to act as a dynatron.

DYNE—The unit of force in the absolute or CGS system of units. It is defined as the force which, acting on a mass of one gram for one second, will impart to the mass a velocity of one centimeter per second. (See *CGS System*, also *Force.*)

E

E—Common symbol for *Electromotive force* (q.v.).

EAR CUSHION—Pads or cushions of soft rubber used in conjunction with head-phones to prevent unpleasant pressure against the ears and also to exclude outside sounds when listening in.

EARTH—An alternate term for *ground*, where the earth or any metallic connection thereto is used as a return in transmission or reception of *electromagnetic waves.* (See *Ground.*)

EARTH CURRENTS—See *Ground Currents.*

EARTH, DEAD—See *Ground.*

EARTH DETECTOR—See *Ground Detector.*

EBONITE—See *Insulating Materials.*

EBURIN—An insulating compound used for *strain insulators* (q.v.). (See *Insulating Materials.*)

EDDY CURRENT LOSS—The portion

of the total loss in electrical apparatus due to *Eddy Currents* (q.v.).

EDDY CURRENTS—Currents induced in the mass of a solid *conductor* due to the action of a varying *magnetic field.* The most common example of eddy currents will be found in the case of a *dynamo* or *generator.* Here the useful current generated in the *armature* is produced by the motion of the armature in the magnetic field. At the same time currents are generated in the iron core due to its motion in the magnetic field. As all currents generated in an electrical machine are produced at the expense of a certain amount of energy, and since the Eddy currents cannot be gathered or put to useful account in this case, they represent a loss. This loss usually is apparent in the form of heat. If the core is made up of solid metal it will be obvious that a good closed path is offered for eddy currents produced. Now if the core is composed of a number of sheets or *laminations,* insulated by thin layers of paper or by an insulating scale (see *Transformer Steel*) the path for the eddy currents is partially broken and the effect is reduced. Eddy currents are produced in the core of a transformer due to the same action—the variation of the magnetic field in which the core is located. In order to reduce the effect, laminated cores are employed as in the case of many generating machines. Eddy current loss is often given as the loss in watts per pound of core material at 10,000 *gausses* (q.v.) and 60 cycles for a sheet 0.0141 inch or 0.0358 centimeter thick. While eddy currents are usually productive of losses they are put to useful account in certain types of instruments. For example, the eddy currents produced in the metal frame of a moving coil meter may be employed to *damp* or retard the free *oscillations* of the needle. (See *Damping of Instruments.*) Eddy currents are also used in a form of speed indicator where the reaction of eddy currents created in a moving disk are made to deflect a pivoted, spring controlled, magnetic needle. (See *Core Loss, Hysteresis,* also *Foucault Currents.*)

EDDY CURRENT COEFFICIENT—The coefficient or numerical multiplier, generally termed K_1, used in calculations of eddy current loss. Its numerical value depends upon the *specific resistance* (q.v.) of the iron used in the core of the machine in question, the character or *wave shape* of the induced voltage, the distribution of *magnetic flux* (q.v.) in the core material, the degree of insulation between sheets, where *laminations* are used, and upon the *flux distribution* (q.v.) due to the shape of the magnetic circuit. Its value will also vary according to the units used in the computation.

EDDY CURRENT LOSS, FORMULA FOR CALCULATING—A common formula for calculation of eddy current loss is the following: $P_\epsilon = \epsilon V(xfB)^2$. Here the loss is expressed in watts as P_ϵ, V is the volume of core metal in cubic centimeters, x the thickness of the sheets in centimeters, f the *frequency* in cycles per second, B the *flux density* (q.v.) in *gausses* and ϵ the eddy current coefficient.

EDGE EFFECT—The effect on the capacity of a condenser due to the curving of the lines of stress at the edge of the plates. This effect is not very pronounced where there is considerable *dielectric* surface (insulating material) extending beyond the edge of the plates. (See *Distributed Capacity.*)

EDISON BATTERY—A number of Edison cells grouped together in one case to supply various currents and voltages as in radio. The illustration shows an Edison "B" battery, used to supply potential to the plates of vacuum tubes. (See *"A" Battery, "B" Battery, Plate, Filament,* also *Storage Battery.*)

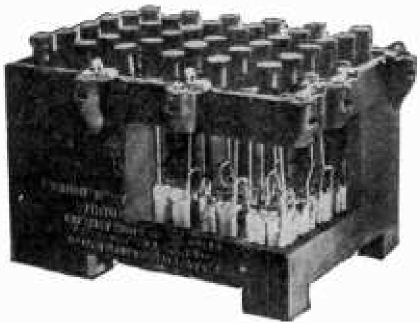

Photo by Courtesy of Edison Storage Battery Co.

Edison "B" battery with portion of case removed to show interior construction.

EDISON CELL—A storage cell employing electrodes of nickeled steel and a solution of potassium hydrate for the electrolyte as distinguished from the usual type of storage cell using lead plates and dilute sulphuric acid as the electrolyte. The chief point of superiority of this type of cell is its comparative ruggedness, due partially to steel construction and also its freedom from the usual ills of cells employing acids and lead plates. The positive electrode in the Edison cell is made by tamping nickel hydrate and alternate layers of pure flake nickel into perforated steel tubes under pressure. The negative electrode is made by pressing iron oxide into flat perforated steel pockets. A number of these pockets are forced into a steel grid to form the complete negative plate. (Note: The *negative* electrode connection on the outside of the cell is referred to as the *positive* pole because it is positive in its relation to the external circuit. Similarly the *positive* electrode connection will be known as the *negative* pole.) The potassium hydrate or hydroxide which takes the place of the acid solution used in the lead plate type of cell acts as a preservative of the steel elements. This naturally means greater life; in fact, it is not subject to the chemical deterioration of other cells. The illustration shows a cell with part of the sides cut away to show the interior construction. (See *Edison Battery.*)

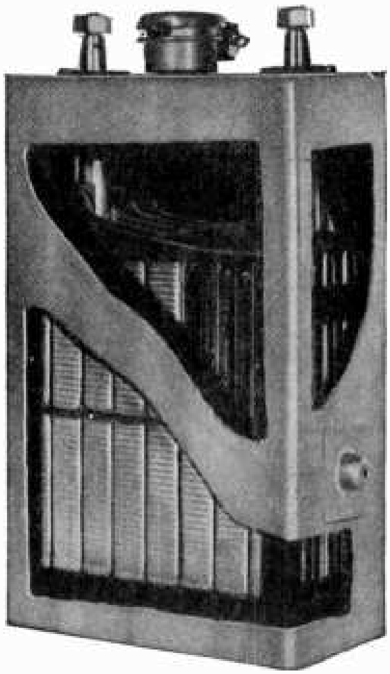

Photo by Courtesy of Edison Storage Battery Co.

Edison cell with part of case cut away to show how constructed.

EDISON, THOMAS ALVA—Born 1847. An inventor famous for his experiments in applied electricity. He began life with newspaper work which he soon abandoned for telegraphy, making many original inventions in duplex systems of operation. After a varied experience in that line he came to New York in 1871, where his talents were recognized and he had opportunity to profitably develop his ideas. The duplex telegraph was made a success the following year, and two years later the quadruplex; and thereupon he began manufacturing on a large scale for the Western Union Telegraph Co. In 1876 he gave up his factory, and established his experimental station at Menlo Park, N. J., where for several years he worked upon the problem of the incandescent electric light, exhibiting a successful bamboo filament lamp in Paris in 1881. He invented the phonograph in 1878. He superintended the construction of the first incandescent lighting station in New York in 1882. Moving his laboratory to Orange, N. J., he established there a large plant for electrical experiment and invention and as a result of his labors there he has taken out 400 patents. Among his inventions may be further named: a type of dynamo, a microphone, the chemical electrical meter, an electric pen, the mimeograph, the magnetic ore separator, dead beat galvanometer, the electric torpedo, a telephone transmitter, and a storage battery. His chief fame rests with his development of the telegraph, his invention of the incandescent lamp and the phonograph.

Thomas A. Edison

EDISON EFFECT—The blackening of the inner surface of an electric light bulb during use. Edison noticed that after the lamp had been burned for some time a coating of black formed on the inner surface and increased in density, becoming finally almost opaque. Edison's experiments, followed by those of *Professor Fleming* (q.v.) resulted in the development of the *Fleming valve* (q.v.)., it having been determined that the black coating was due to the discharge of *electrons* from the hot *filament*. (See *Vacuum Tube, Electron Emission*, etc.)

EFFICIENCY—A very flexible term in electrical practice. Generally speaking, the ratio of useful output of any piece of apparatus to total input. Efficiency is customarily expressed in terms of percentage. Thus, broadly, a piece of apparatus may produce as useful output 80 per cent of the total input. (See *Power Factor.*)

EFFECTIVE ELECTROMOTIVE FORCE—The square root of the mean square of the full alternating current wave. Usually abbreviated R. M. S. (root—mean—square). The effective value of the E. M. F. (electromotive force) is thus taken as the square root of the mean of the squares of the *instantaneous values* over a complete period. (See *Instantaneous Values.*)

EGG INSULATOR—A name applied to a certain type of *strain insulator* due to its egg-like form. The illustration shows such an insulator. The two connections to the insulators are so arranged that the wires will still be looped together in the event that the insulator breaks. (See *Insulator*, also *Insulation of Aerial.*)

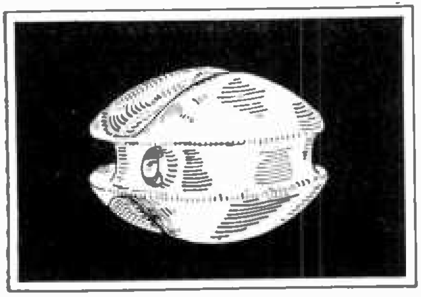

An egg insulator.

ELASTICITY—A property of matter which permits it to resist any change in shape or bulk and permits it after such a change to return to its original state. This property of materials by virtue of which they are enabled to return to their original state after the force causing the distortion has been removed is a very important one in all branches of engineering. Its chief application in radio is in the matter of aerials and aerial guy wires. While the average aerial installation does not represent any serious problems in stress, the large aerial systems of commercial stations must be carefully designed in this respect. For example, a guy wire of one material may have more elasticity than one of similar size in another material. The amount which the wire will stretch under a certain load may be determined by use of a table giving the elasticity of various metals. (See *Electric Elasticity.*)

ELECTRIC ABSORPTION—An effect in *condensers*. That quality of a condenser by means of which it absorbs or "soaks up" a charge of electricity, and conversely, retains part of the charge when the condenser is momentarily discharged. This effect is more apparent in cases where a solid *dielectric* material is used. In fact, *air condensers* (q.v.) are said to have little electric absorption. The reason for this is that mica or other solid dielectrics are strained to a greater extent than air dielectric after the condenser has received a charge. (See *Soaking In*, also *Dielectric Absorption.*)

ELECTRIC DISPLACEMENT—A term proposed by Maxwell to denote a quantity expressing the state or condition of a dielectric in an electric field in accordance with the supposition that such a field is created by the transfer of positive electricity to one end of the field of force and negative electricity to the other end of the so-called *tubes of force* (q.v.). (See *Displacement Current*, also *Electrostatic Induction*, and *Flux Density, Electrostatic.*)

ELECTRIC ELASTICITY—In a dielectric material, the property which permits it to arrest passage of a *displacement current* (q.v.) due to electric stress. It is equivalent to the electric stress divided by the electric strain. (See *Dielectric*, also *Displacement Current.*)

ELECTRIC FIELD—The area surrounding an electrified body in which the electrical influence of that body can be measured or noticed. Also any area or region in which there is an electric force either steady or varying in intensity. For example, the *dielectric* or insulating surface of a *condenser* (material separating the plates) contains an electric field. Similarly, an electric field always surrounds a wire or other conductor carrying electric current. (See *Field, Magnetic; Field, Electromagnetic; Field, Electrostatic;* also *Flux.*)

ELECTRIC INDUCTION—The transfer of an electric state from a charged or electrified body to a non-electrified body without electrical contact. (See *Induction.*)

ELECTRICITY—The term given to an invisible form of energy. Electricity is not tangible to the human senses, but its effects can readily be detected. The phenomenon of electricity is regarded as due to the separation and independent movement of constituent parts of atoms—known as *electrons*. While electricity is really an intangible thing so far as the human perceptions are concerned, it is known to science not only through its effects or manifestations of energy, but has actually been reduced to quantities. The mass of electricity has been determined, as has also the *rate of flow*, etc. We have units which are used for the basis of all calculations in electricity. Thus, *voltage* is regarded as the propelling force; the quantity of electricity is given in *coulombs*, the *current* or rate of flow of electricity is given in *amperes*, etc. Each term for the various phases of electricity is based on a certain unit which represents definite values under given conditions. The universally accepted theory in chemistry is that all matter is made up of molecules, which in turn are composed of one or more atoms. These atoms now are regarded as comprising minute solar systems made up of positive and negative electrical particles. (For more complete explanation of the *electron theory* see *Electron Theory;* see also *Magnetism, Current, Voltage.*)

ELECTRO-CHEMICAL CONDENSER—See *Electrolytic Condenser.*

ELECTRO-CHEMICAL EQUIVALENT—The amount of a substance liberated by *electrolytic action* (electrolysis) by the passage of one *coulomb* (q.v.) of electric current. It represents the weight in grams of each element of an *electrolyte* (q.v.) which is deposited by one coulomb of electricity. It has been determined that one coulomb will liberate .00001035 grams of hydrogen, which is therefore used as the electrochemical equivalent of hydrogen. Electro-chemical equivalents are much used in applying the law of electrolysis. This law states that the amount (weight) of an *ion* liberated at an *electrode* each second during *electrolysis* is equal to the strength of the *current* (amperes) multiplied by the electro-chemical equivalent. The corresponding equivalents of other elements compared to hydrogen may be determined by multiplying the figure .00001035 by the atomic weight of the other element in grams and dividing the result by the *valency* (q.v.) of the element in question. (See *Electrolysis*, also *Electrolyte.*)

ELECTRO-CHEMICAL SERIES—A table of metals arranged in accordance with the *potential* they produce in a particular *electrolytic* solution.

ELECTRO-CHEMISTRY—The science dealing with chemical changes due to electricity or effected by means of electricity. (See *Electrolysis*, also *Cell*, and *Storage Batteries.*)

ELECTRODE—The conductor through which electric current enters or leaves an electrical device. That is to say the pole or terminal of the current carrying conductors separated by a medium through which electric current can flow from one to the other. The most common usage of the term is in referring to the *positive* and *negative* poles (q.v.) of a *primary* or *storage cell*. In the case of a *dry cell*, the two electrodes are the carbon rod or block in the center and the zinc container in most cases. In radio it is used to designate the *filament* and *plate* elements of a vacuum tube. The filament is termed the *cathode*, or *negative electrode* and the plate is the *positive electrode* or *anode*. The illustration shows an ordinary *dry cell* or *primary cell* with the two electrodes designated as positive and negative. (See *Anode, Cathode, Battery, Electrolysis, Positive Electrode* and *Negative Electrode.*)

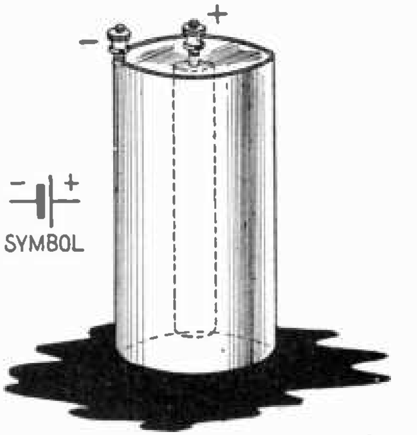

Dry cell showing positive and negative electrodes.

ELECTRODYNAMIC—Term used to denote electric currents or the forces exerted by one current upon another.

ELECTRO-DYNAMICS—A branch of the science of electricity dealing with electricity in motion; the study of the force exerted by electric currents upon each other. This branch of electricity was first accorded serious attention when the French scientist, *A. M. Ampere* (q.v.) announced the result of his investigations in collaboration with D. F. J. Arago. Ampere stated in 1820 that *parallel conductors* through which electric currents were flowing in the same direction had a tendency to be attracted to each other and to be repelled when the currents were flowing in opposite directions. The subject is discussed under its various separate headings. (See *Ampere, Electricity, Magnetism.*)

ELECTRO-DYNAMOMETER—A device for measuring the *current* and *voltage* in a circuit; that is, the actual power in a circuit. It is a form of *wattmeter* (q.v.) designed to indicate the power in *watts*. It is of particular importance in measuring the power in *alternating current* circuits. In the case of *direct current*, the watts in the circuit are the product of the volts multiplied by the amperes, the result being the actual power. In the case of alternating current, however, the product of the volts and amperes is not necessarily, in fact, seldom, the true power or watts. Here we have the problem of true watts and apparent watts, or volt-amperes. If a voltmeter and an ammeter are connected in a circuit and read separately, the readings indicate the maximum voltage and the maximum current or amperage. The product of these two values is the power in watts in a direct current circuit. If the current is an alternating one, the true watts or actual power will be the product of the *instantaneous values* (q.v.) of voltage and current. Under certain conditions in a circuit carrying alternating currents, the current may lag behind or precede the voltage. In other words, the maximum value of current is not reached at the same moment as the maximum value of the voltage. In this case the true watts can only be determined by obtaining the values of volts and amperes at the same instant, i.e., instantaneous values. The electro-dynamometer shown in the illustration gives the instantaneous values, or rather their product, and thus indicates the power or true watts. Here the heavy lines represent the current or moving coil and the light lines the stationary or voltage coil. (See *Dynamometer*, also *Wattmeter.*)

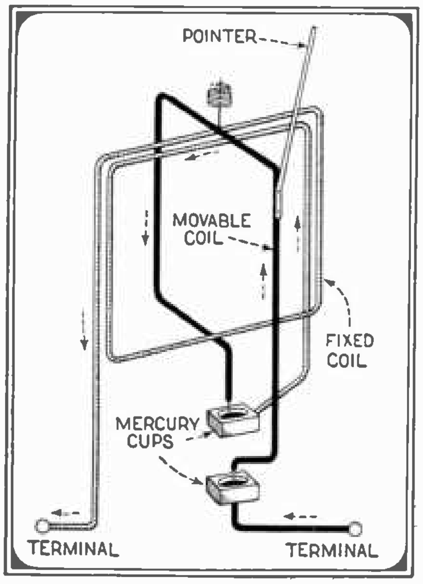

Electro-dynamometer with the current or moving coil shown by heavy lines and the stationary or voltage coil by light lines.

ELECTROLINES—A term advanced by *Professor J. A. Fleming* (q.v.) to designate the lines of electric force radiating from an electron. These lines are pictured as analogous to long straight wires extending in all directions from the center of a small sphere. (See *Electron*, also *Electromagnetic Waves.*)

ELECTROLYSIS—The decomposition of a comopund substance, generally a liquid, into its component parts by the action of an electric current passing through it. While electrolysis is an undesired effect in certain phases of electrical power work, causing composition of grounded structures, its application to radio is mainly in the case of cells for producing an *electromotive force* or to furnish current for lighting tubes. The constituent elements of a chemical compound such as used in a wet cell or any device using an electrolyte for the production of electric current, are known as *ions* (q.v.). These ions, which are separated during electrolysis, are of two kinds. The electro-positive ions are known as *cations*, and the electro-negative ions are known as *anions*. The electro-positive ions appear at the cathode or negative electrode during electrolysis and therefore are of positive origin. The electro-negative ions appear at the positive electrode and are therefore of negative origin. This is easily understood when the ions are considered as moving from the positive electrode or to the negative. Thus the ions appearing at the negative electrode must have come from the positive electrode and are therefore considered electro-positive in nature. The illustration shows the action of the ions in a typical *primary cell* (q.v.). The ions are considered as carrying the current through the electrolyte—each ion carrying a fixed charge of electricity of positive or negative nature. An ion is capable of carrying only a certain fixed charge of electricity and therefore any increase in the current will necessarily be accomplished by an increase in the number of ions. (See *Primary Cell, Electrolyte, Electrolytic Detector, Interrupter* and *Condenser.*)

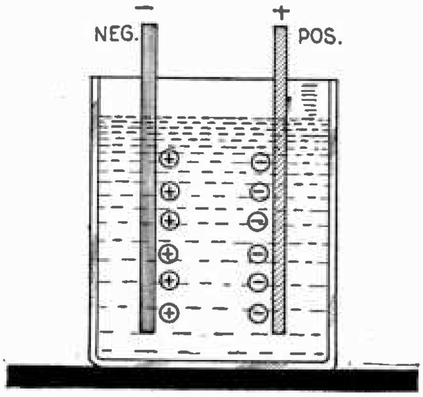

Illustration showing the action of ions in a primary cell.

ELECTROLYTE—The liquid decomposed during electrolysis, as the exciting fluid or solution in a wet cell or primary cell. The liquid used in any primary, secondary or electrolytic cell. (See *Cell*, also *Storage Battery.*)

ELECTROLYTIC ACTION—The decomposition of a chemical compound (electrolyte) into its constituent elements—ions, by the passage of an electric current through it. (See *Electrolysis, Ions*, also *Cell.*)

ELECTROLYTIC BATTERY CHARGER—A device for converting alternating current such as supplied for house lighting purposes, to direct current for the purpose of charging storage batteries. Essentially one or more cells containing an electrolyte and two metal electrodes, generally lead and aluminum, so arranged that current can only pass in one direction, thus making it uni-directional and suitable for charging batteries. Such chargers are arranged to rectify either one-half of the alternating current cycle, or both halves of the wave. (See *Charger, Storage Battery, Electrolytic Rectifier*, also *Electrolytic Cell.*).

ELECTROLYTIC CELL—An arrangement of two *electrodes* in an electrolyte through which electric current can be passed to produce electrolysis. The electrolyte is decomposed or separated into ions by the action of the current, and at each of the electrodes one of the substances of which the electrolyte is composed accumulates. Such a cell is often used to measure electric current, as the amount of a substance deposited on the electrode depends directly on the amount of current passed through the electrolyte. (See *Electrolysis.*)

ELECTROLYTIC CONDENSER—A form of condenser making use of two electrodes placed in an electrolyte and used for alternating current circuits of high voltage and low frequency, low voltage and high frequency or high voltage and high frequency. The action of such an arrangement as a condenser depends on the fact that upon the passage of certain forms of alternating currents through it, a gas is formed around the electrodes, causing *polarization* (q.v.) and opposing the current. This polarization creates a very good insulating medium and a condenser of such nature can be made to have very high capacity. An electrolytic condenser may be used to prevent any sudden surge of pressure due to lightning or other causes. The polarization is instantaneous when any surge of voltage takes place and thus forms an automatic safety device which operates instantly and returns to its normal characteristics as a condenser as soon as the surge has passed. (See *Capacity*, also *Condenser.*)

ELECTROLYTIC CONDUCTION—The action in an electrolyte wherein the ions that have been separated carry positive and negative electric charges in opposite directions to the electrodes. (See *Electrolysis.*)

ELECTROLYTIC DETECTOR A device which converts high frequency currents into direct current pulsations capable of operating a telephone receiver. It is an extremely sensitive detector of electromagnetic waves (radio signals). The illustration shows a simple form as originally used in place of the customary crystal detector. A fine platinum wire dips into a

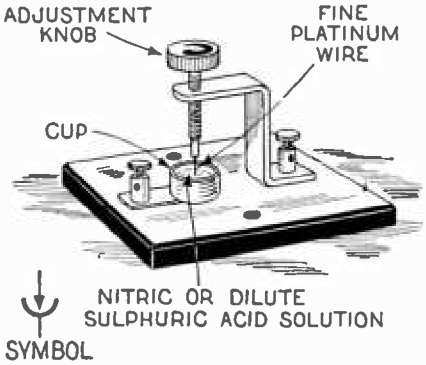

Electrolytic detector originally used in place of a crystal detector for radio reception.

cup of acid solution, generally either nitric or dilute sulphuric. The fine wire is lowered until it just touches the acid solution and must be very carefully adjusted as it has a tendency to curl up, its diameter being usually but one five thousandth of an inch or less. The high frequency currents can pass in only one direction through the detector which gives the rectifying action and makes the signals audible in the head phones. There are several theories regarding the action of an electrolytic detector, the most probable one being that the contact of the wire on the surface of the acid solution gives the effect of a very small *electrolytic condenser.* (q.v.) When the high frequency currents are passed through the detector there is a polarization action which causes high resistance to current in one direction but not in the other. (See *Detector,* also *Polarization.*)

ELECTROLYTIC INTERRUPTER—A device using electrodes immersed in an electrolyte for the purpose of interrupting or breaking up the flow of current. It is used mainly for induction coils for producing high voltages. The common form comprises a lead plate as a cathode or negative element and a platinum wire extending beyond a glass or porcelain tube in which it is tightly imbedded, both immersed in an electrolyte of dilute sulphuric acid. When current passes through, bubbles form on the platinum electrode causing temporary stoppage of the current. The bubbles are automatically dissipated and the current flows again, this operation occurring at rapid intervals and giving as many as 1000 interruptions per second. (See *Interrupter,* also *Induction Coil.*)

ELECTROLYTIC RECTIFIER—A device for changing alternating current into direct current by use of electrodes placed in an electrolytic solution. The action of such electrolytic cells or rectifiers is to permit current to pass only in one direction, thus suppressing one half of each cycle and producing a direct current. The illustration shows one of the customary arrangements for electrolytic rectification, being an electrolytic cell with two electrodes, one being aluminum and the other a lead electrode. One or more lamps are placed in series or shunt in the circuit to permit the proper current flow. The positive pole of a storage battery is connected to the aluminum plate as shown and the lead electrode connects to the other side of the alternating current line and back to the negative side of the battery. At one instant current flows in the normal direction through the battery, but when the current reverses and flows in the opposite direction through the electrolytic cell, oxygen is given off at the aluminum electrode and forms a thin film of aluminum oxide which acts as an insulator and prevents the passage of the current in that direction. Thus with one cell one half of each alternating current cycle is stopped or cut off when the aluminum electrode is made positive and the result is an intermittent direct current suitable for charging batteries. Such rectifiers are also used for supplying current to the filaments of vacuum tubes, or voltage to the plate members of the tubes, particularly for transmitting. (See *Electrolytic Rectifier Full Wave,* also *Balkite Battery Charger.*)

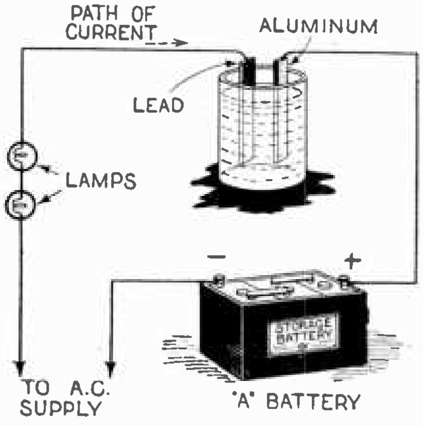

Customary method of electrolytic rectification.

ELECTROLYTIC RECTIFIER FULL WAVE—This system is much used for charging storage batteries from an alternating current house lighting supply. Four cells are used in this instance. The electrolyte may be a solution of hydrolchloric or sulphuric acid, or a saturated solution of sodium phosphate or bicarbonate of soda. Each cell has two electrodes, one iron or aluminum and the other lead. The positive pole of the storage battery is connected to the lead electrodes of two cells and the negative pole is connected to the aluminum electrodes of two cells. By this means both halves of the alternating current cycle are rectified to direct, pulsating current. This is known as *full wave rectification* (q.v.) (See *Charger, Storage Battery, Balkite Battery Charger,* also *Rectifier.*)

ELECTRO-MAGNET—A magnet using a curved or straight piece of soft iron or other magnetic material so arranged as to become magnetized by the passage of electric current through a coil of wire surrounding it. The illustration shows one of the commonest forms of electro-magnet. This type is used in a bell or buzzer to draw down the vibrator bar and break the circuit. (For explanation of action see *Buzzer.*) When the current is withdrawn from an electro-magnet, the magnetic force or attraction ceases. Electro-magnets are used for a variety of purposes in radio, such as for test buzzers, head phones, relays, etc. (See *Magnet,* also *Magnetic Lines of Force.*)

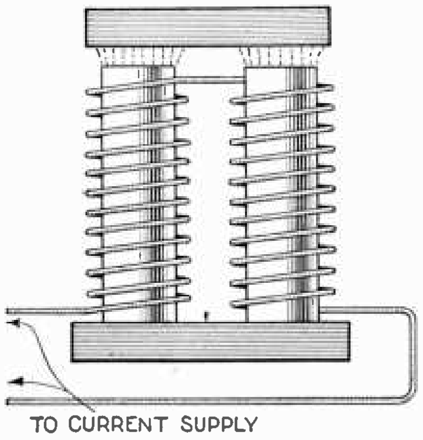

One of the common forms of electro-magnets.

ELECTROMAGNETIC CONTROL—The control of various switches and other apparatus by means of *electromagnets,* usually through an auxiliary contact at a distant point. (See *Remote Control,* also *Electromagnet.*)

ELECTROMAGNETIC FIELD—The space understood to be filled with *electromagnetic lines of force.* (See *Flux,* also *Lines of Force.*)

ELECTROMAGNETIC FLUX—The distribution of the lines of force through an electromagnetic field. (See *Flux,* also *Lines of Force.*)

ELECTROMAGNETIC IMPULSE—An impulse or electromagnetic disturbance conveyed to the ether from a conductor carrying high frequency currents. Another term for the electromagnetic waves involved in radio. (See *Electromagnetic Waves.*)

ELECTROMAGNETIC INDUCTION — The production of electric currents through a change in a *magnetic field* (q.v.). Its production is due to a change in the number of lines of force which are linked with a conducting circuit. The simplest example of this means of producing current is shown in the illustration. Here a conductor is arranged to be moved in the immediate vicinity of a magnet, causing a change in the number of lines of force that link the conductor and creating an induced current. This current will persist only so long as the conductor continues to be moved in the magnetic field. That is to say when there is no longer any change in the magnetic lines of force linking the conductor, there will be no current induced. The discovery of electromagnetic induction and formulation of its laws was primarily due to the scientist *Faraday* (q.v.). Faraday's law governing electromagnetic induction follows: "The induced electromotive force around any circuit is the rate of the decrease of the total flux of magnetic induction through the circuit." (See *Induction, Flux,* and *Lines of Force.*)

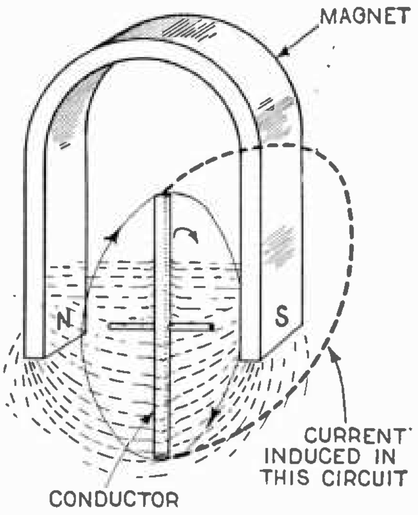

Simple method of producing electric currents by causing changes in the magnetic field.

ELECTROMAGNETIC INSTRUMENTS —Electrical measuring instruments such as voltmeters, ammeters, etc., which depend for their action on the exertion of electromotive forces upon iron armatures. (See *Ammeter, Thermal;* also *Voltmeter.*)

ELECTROMAGNETIC MICROPHONE —A type of microphone in which the sound waves due to speech or music cause vibrations of a suspended coil in a magnetic field, inducing in the coil minute waves of electromotive force corresponding in form to the original sound waves. A form of microphone used in radiophone (broadcasting) transmission when purity of sound is an important consideration. Such a microphone requires that the induced currents be heavily amplified. (See *Microphone*, also *Speech Amplifier.*)

ELECTROMAGNETIC RADIATION—The propagation or radiation of electromagnetic waves through the ether. (See *Electromagnetic Waves*, also *Ether.*)

ELECTROMAGNETIC ATTRACTION—The attraction existing between unlike poles of electromagnets. This attraction is identical with that of two unlike poles of ordinary magnets. (See *Magnetic Attraction and Repulsion.*)

ELECTROMAGNETIC REPULSION—The tendency of two like poles of electromagnets to act against or repel each other. (See *Magnetic Attraction and Repulsion.*)

ELECTROMAGNETIC SWITCH—A switch for breaking electrical circuits by the action of an electromagnet. (See *Electromagnet*, also *Remote Control.*)

ELECTROMAGNETIC THEORY OF LIGHT—The theory and consequent scientific proof that radio waves and light waves are practically identical in nature. Light waves travel at a tremendous speed through the ether, approximately 186,000 miles per second, which is also understood to be the velocity of electromagnetic or radio waves. (See *Electromagnetic Waves.*)

ELECTROMAGNETIC UNITS—The fundamental or absolute units employed in electricity as a basis for calculations and on which the practical units are based. The effects of electricity, and hence electricity itself, can be measured in two ways. The first is the measurement of the quantity of electricity by the force it exerts upon another stationary quantity of electricity. This is known as the *electrostatic system*. The second is by measuring the force which a given quantity of electricity exerts upon a *magnetic pole* when flowing through a nearby conductor. This is referred to as the electromagnetic system. Fundamental or absolute units are derived from the fundamental units of *length, mass* and *time*. These units are the *centimeter of length, the gramme of mass* and the *second of time*. The term centimeter-gramme-second, referring to this system, is abbreviated C.G.S. system. The practical units of electricity are known as *volts, amperes*, etc., and are based on the absolute units. (See *Electrostatic, Units, C.G.S. System*, also *Ampere, Erg, Dyne* and *Coulomb.*)

ELECTROMAGNETIC WAVES—The term applied to the energy radiated from an *antenna* in radio transmission. So called because it is understood to travel in the form of waves. One-half the energy of such a wave is electrical and the other half electromagnetic. *Electromagnetic waves* travel with the speed of light waves or approximately 186,000 miles per second. These waves are of extremely high frequency—that is, their direction changes many thousand times each second. This will be better understood by means of the illustration showing the arrangement of the electric and *magnetic fields* around a radiating antenna. Here the lines represent the electric field and the shaded portion the magnetic field. The lower loops or "feet" of these waves are assumed to pass over the surface of the earth which is a conductor. Thus when electromagnetic waves are being radiated from an antenna, circular distributions of electric charges are in the earth's surface around the antenna, these charges being alternately negative and positive, the bands of charge spreading out or radiating from the antenna in the manner shown in the illustration. These waves are subject to various influences which have a tendency to refract or reflect the waves much in the manner of light waves.

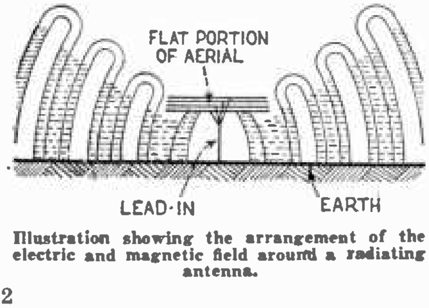

Illustration showing the arrangement of the electric and magnetic field around a radiating antenna.

As these waves move out from the antenna they gradually increase in height, although the distance between the "feet," i.e., from one wave to another, remains the same. This distance is known as the wave-length. (See *Reception of Electromagnetic Waves*, also *Wave-length, Radiation Transmission.*)

ELECTROMETER—An instrument used for measuring potential differences (voltages) and, under certain conditions, power in *watts*. The operation of the instrument depends upon the attraction or repulsion of electrostatic charges and it may be used for either direct or alternating current measurements. (See *Voltmeter, Electrostatic.*)

ELECTROMETRY—The practice or science of electrical measurement. (See *Meter, Voltmeter.*)

ELECTROMOTIVE FORCE—Symbol E or e—abbreviation E.M.F. The force or electrical pressure which starts and maintains a current of electricity through a conductor. Electromotive force is commonly referred to in terms of volts and is abbreviated E.M.F. The electromotive force in electricity is analogous to the difference of level which produces pressure in pipes carrying water and forces the water to flow. Electromotive force is produced by a difference of potential, and this in turn sets up the flow of current. The illustration *Fig. 1A* shows a simple primary cell or electrolytic cell, and *Fig. 1B* is the analogous action of two tanks holding water, connected by a pipe and so arranged that the water level in one tank is higher than in the other, thus having a difference of level and creating pressure which makes the water flow. The water in tank A being at a higher level than in the other, the pressure causes the water to seek its common level and it flowed into tank B through the pipe. This flow continues as long as there is difference of level. In the case of the cell *Fig. 1A* there are two electrodes, one of which is at a higher potential than the other. That is to say, one is in a state of higher electrification than the other, causing a difference of potential and producing an electromotive force. This force is in the direction from the point of greater potential to point of lesser potential as in the case of the water, and causes a flow of current in that direction. Here the flow is from B to A in the electrolytic solution and from A to B externally through the conductor C. Thus as long as there exists a difference of potential, current will flow in an attempt to equalize the potentials. (This statement holds true only under ideal conditions. See *Polarization.*) This flow of current in the illustration *Fig. 1A* is according to *Ohm's law* (q.v.), i.e., a difference of potential of one volt will cause one ampere to flow through the external circuit when the resistance of the circuit is one ohm. (See *Ampere, Ohm, Volt, Resistance.*)

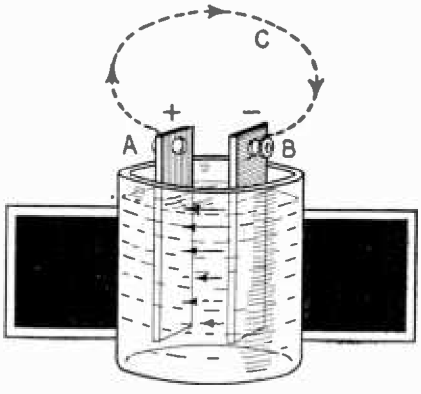

Fig. 1-A. A simple primary or electrolytic cell.

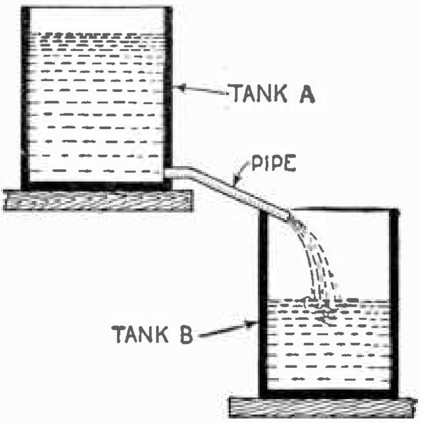

Fig. 1-B. Water in tank "A" seeks its level by discharging into tank "B" in a manner analogous to the flow of electric current where a potential difference exists.

ELECTRON—The smallest negative charge of electricity known. An English investigator, Dr. Johnstone Stoney, is credited with having named the electron, although it was so named by him in anticipation of its being isolated and measured. The actual proof of the earler assumptions was due to Sir J. J. Thomson and others, who actually measured, and in all effect weighed, the minute particles of negative electricity. (See *Electron Theory.*)

ELECTRON FLOW—The theory of movement of free electrons along a conductor through which an electric current is flowing. (See *Electron*, also *Current, Assumed Direction of Flow.*)

ELECTRON EMISSION—The emission or discharge of electrons or minute particles of negative electricity from a heated body, particularly in a vacuum. The emission of electrons in a vacuum tube is the basis of operation of all vacuum tubes as used in radio. (See *Electron Theory*, also *Vacuum Tube, Theory of Operation.*)

ELECTRON THEORY—The assumption that the atoms of all substances contain one or more particles of negative electricity, which is of course paramount to saying that all matter is electricity. These particles are known as electrons, and while the assumption is referred to as a theory, it has been rather definitely proven by numerous experiments, and electrons have actually been measured. Under the electron theory it is assumed that the charge carried by a single electron is the smallest possible charge which can exist in nature and that no charge of electricity exists or can be produced which is not an integral multiple of this charge. (Note: This is merely a mathematical way of stating that any charge of electricity is equivalent to the combined charges of a certain number of electrons.) Probably the most accurate conclusion regarding the charge carried by an electron is due to experiments carried on by Prof. A. E. Millikan of the University of California. His measurements state this charge of a single electron as being 4.77×10^{-10} electrostatic unit, which is 1.59×10^{-19} of a *coulomb* (q.v.). While the mass or weight of an electron is said to vary with its velocity, assuming a velocity of less than one-tenth that of light, the mass will be 8.9×10^{-28} grams. This is approximately one seventeen hundredth of the weight or mass of an atom of hydrogen. For purposes of calculation the electron is assumed to be in the shape of a sphere. The radius has been estimated at approximately 10^{-13} centimeter. When it is considered that a centimeter is somewhat less than half an inch it is readily seen that an electron is an infinitesimal affair.

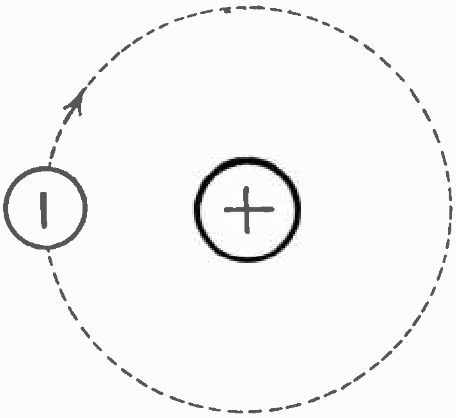

Illustration showing the electron, or negative particle, revolving around the proton or positive nucleus.

The figures as to velocity of electrons vary somewhat with different authorities, but it is probably about 100 kilometers (approximately 62½ miles) per second at a temperature of zero degrees centigrade, the speed increasing as the temperature rises. Electrons may be understood as comprising minute solar systems of their own. In the illustration the *electron* or negative particle is shown revolving around the *proton* (q.v.) or positive nucleus in the manner of a planet revolving around the sun. This represents a normal atom of hydrogen, the simplest of all atoms, which, as shown, is composed of a single positive nucleus and a single electron revolving around it. In the case of an atom of helium there is a doubly charged positive nucleus, or proton, and two negative electrons revolving around it. Other atoms have as many as 92 electrons.

All vacuum tubes used in radio operate on the basis of the emission of electrons from a heated filament within the tube. These electrons are being emitted from the hot filament at all times during operation of the tube. When the incoming signals are such as to charge the grid negatively they are repelled and held back to the filament, thus offering no path for the signals. When the incoming signals have reversed their polarity, as they do thousands of times each second, the grid will be positively charged and the electrons will be aided in their travel to the plate, thus forming a conductive path for the currents. This, of course, results in the elimination of one-half of each cycle of the incoming currents or oscillations. That is to say, they are rectified to direct pulsating currents. (See *Amplifier, Vacuum Tube, Theory of Operation* and *Space Charge.*)

ELECTRON RELAY—A term occasionally applied to a three element tube (vacuum tube). This alternate term is obviously derived from the fact that the vacuum tube makes use of the path of electrons and in effect acts as a relay. (See *Electron.*)

ELECTRON TUBE—Any tube using a high vacuum and depending for operation on the emission of electrons from a heated *cathode.* (See *Electron,* also *Vacuum Tube.*)

ELECTRONEGATIVE—Having the property of being attracted to the positive pole of an electrolytic cell. The electronegative component. (See *Anion,* also *Cathion.*)

ELECTROPOSITIVE—Having the property of being attracted to the negative pole of an electrolytic cell during electrolysis. The electropositive component. (See *Anion,* also *Cathion.*)

ELECTROSCOPE—A device for detecting minute differences of potential (voltages). Such a machine will not measure the voltage but merely indicate its presence. The simplest form of electroscope is shown in the illustration. Here a ball B, composed of the pith of some wood, preferably elder due to its extreme lightness, is suspended by a thin silk thread from an insulated stand. The object under investigation is moved near the pith ball, which will be repelled or attracted by the object if the object is in a state of electrification; but will remain motionless if no difference of potential exists. (See *Voltmeter, Potential, Difference of.*)

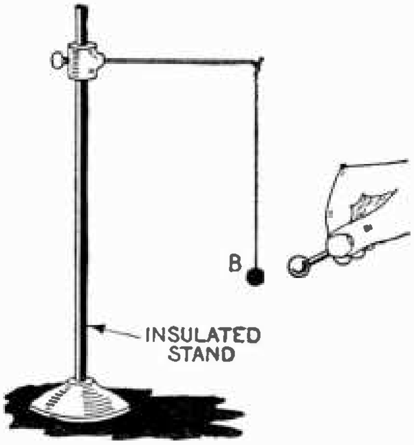

Simplest form of electroscope.

ELECTROSE—An insulating composition much used in electrical power work and for radio purposes. It has very high compressive strength, is hard and tough without being brittle, and is moisture, water and oil proof. Electrose is used extensively for antenna insulation. (See *Insulator*).

ELECTROSTATIC—The term used to designate EMF's (electromotive forces) or potentials, charges, etc., and the resultant forces. That is to say, electricity at rest or stationary in a body in order to distinguish it from the forces due to *electro-dynamics* (q.v.).

ELECTROSTATIC CAPACITY—The amount of electricity which a condenser or similar body is capable of storing. If a condenser is capable of receiving and storing one *coulomb* of electricity under a potential of one *volt* it is said to have an electrostatic capacity of one *farad,* which is the unit of electrostatic capacity. The electrostatic capacity of a condenser depends on several factors, mainly the size of the plates and the nature and thickness of the dielectric material (insulating sheets). (See *Capacity,* also *Condenser.)*

ELECTROSTATIC CHARGE—The presence of electricity in the form of electrostatic lines of force in a body such as a charged conductor. (See *Charge.)*

ELECTROSTATIC COUPLING—The association or coupling of circuits one to another by means of electrostatic action, i.e., condensers. Any capacitive association of one circuit with another. (See *Coupling.)*

ELECTROSTATIC FIELD—A region wherein forces are present and exerted due to the presence of electric charges. It is the area immediately surrounding a charged body. (See *Field, Electromagnetic,* also *Field, Electrostatic*).

ELECTROSTATIC FLUX—Lines of force making up an electrostatic field. (See *Field, Flux,* also *Flux, Electrostatic*).

ELECTROSTATIC FLUX DENSITY—The density or number of electrostatic lines of force in a unit area of an electrostatic field. (See *Electrostatic Field,* also *Flux Density.)*

ELECTROSTATIC FORCE—The force due to interaction of electric charges. (See *Electrostatic Field,* also *Electrostatic Flux Density.)*

ELECTROSTATIC INDUCTION—The production of a charge in a body due to the presence of an opposite charge in a nearby conductor, or synonymous with displacement in the case of a condenser. Electrostatic influence or induction is the basis on which condensers operate. In the illustration *Fig.* 1 is a simple circuit containing a source of potential, which in this case is a battery. As the circuit is closed through the key a momentary flow of current takes place. This flow of current is accompanied by a condition of electric strain in the insulating material of the condenser. The electrically strained dielectric or insulating material exerts a back pressure against the pressure or force of the battery, and when this back pressure is equal to the original battery pressure the flow ceases. The strain in this dielectric is known as *electric displacement* and this displacement can be transmitted or passed on to neighboring parts of the medium and thence into space. Under the proper conditions this displacement and the magnetic field which accompanies it may be detached from the circuit and move independently in space. This is the fundamental process in the radiation of *electromagnetic waves* (q.v.). Now, in *Fig.* 2, A and B represent two conductors, both insulated. If conduc-

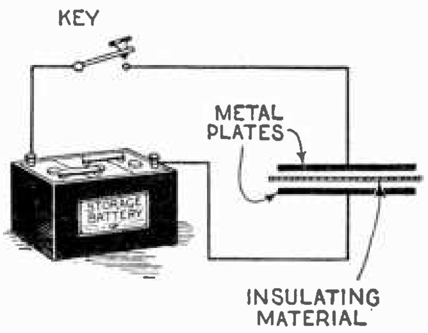

Fig. 1. A simple circuit for illustrating electrostatic induction.

tor A has a positive charge and is brought close to conductor B, although not in actual contact with it, the result is a displacement of electric charge which results in the near side of conductor B becoming negative, while the opposite side assumes the same sign (positive) as conductor A. Actually B has received no charge, but has been placed in a state of strain due to electrostatic induction between it and the conductor A. Due to the negative charge induced on B, the positive potential difference of A has been decreased and in order to maintain the constant potential it must be given a further supply of electricity. In this manner it is actually possible to store the electricity by electrostatic induction between two conductors or plates, by reason of the strain in the insulating medium—air in the case of the two conductors—which in turn passes the displacement on to conductor B. (See *Conductor, Capacity, Electrostatic,* also *Induction.*)

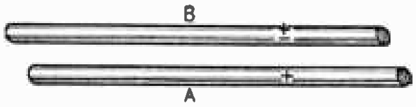

Fig. 2. Illustration of displacement of electric charge due to proximity of charged conductors.

ELECTROSTATIC LEAKAGE—The gradual loss through leakage, of the electrostatic charge in a body, due to the fact that perfect insulation is impossible to obtain. (See *Electrostatic Induction.*)

ELECTROSTATIC LINES OF FORCE—The lines of force traversing an electrostatic field. That is to say the *electrostatic flux* (q.v.).

ELECTROSTATIC STRAIN—The strain or state of stress in which a body is placed when in an electrostatic field. In a condenser, the strain experienced by the *dielectric* (q.v.) when the condenser is in a state of charge. (See *Condenser,* also *Capacity, Electrostatic*).

ELECTROSTATIC SYSTEM—The electrostatic system of C.G.S. units. One of the absolute unit systems in electrical practice. (See *C.G.S.*, also *Electrostatic Units* and *Electromagnetic Units.*)

ELECTROSTATIC UNITS—A system of measurement used in electrical engineering based on the unit of electric quantity under the *C.G.S. system* (q.v.). This unit quantity of electricity is the quantity of electricity which, placed on a small sphere will repel an equal quantity of electricity of the same sign (both negative or both positive) on a similar sphere with a force of one *dyne.* (q.v.) when the centers of the respective spheres are one centimeter apart. With this definition it is assumed that the spheres are in a vacuum, otherwise the distance between centers would vary according to the dielectric or insulating medium. (See *C.G.S., Electromagnetic Units,* also *Conversion Factors.*)

ELECTROSTATIC VOLTMETER — A meter for measuring voltage, depending for its action on the electrostatic action between a fixed and a moving plate connected respectively to the two poles or terminals between which the source of potential is connected. (See *Electrostatic Force,* also *Voltmeter.*)

ELECTROSTATIC WATTMETER — A type of wattmeter depending on the electrostatic action between two sets of fixed and moving plates. (See *Electrostatic Force,* also *Wattmeter.*)

ELECTRO-THERMAL METER — The term sometimes applied to a *hot-wire ammeter* (q.v.) or voltmeter operating on the principle that the heat generated by the passage of an electric current through a constant resistance will be equal to the square of the current. (See *Thermal Ammeter.*)

ELEMENT—A term primarily applied in chemistry to the substances which are not subject to decomposition by electric analysis (see *Electrolysis*). These elements are hydrogen, helium, etc. The term is also used quite generally to denote components such as the electrodes in a primary cell or the plate, grid and filament members of a vacuum tube. (See *Cell,* also *Vacuum Tube.*)

EMANATE—"To flow forth or proceed, as from a source."—Webster. Thus, emanating, issuing from, as the emanation of electrons from a hot filament in a vacuum tube. The throwing off of electrons. (See *Electron Emission.*)

E.M.F. — The abbreviation for electromotive force. The practical unit of E.M.F. is the volt. (See *Electromotive Force,* also *Volt.*)

EMISSION—The act of sending forth. Used chiefly to designate the emission of electrons from the filament of a vacuum tube as used in radio. The flow of these electrons between filament and plate is controlled by the grid. (See *Electron Emission,* also *Vacuum Tube, Theory of Operation.*)

EMISSION CURRENT—In a vacuum tube for transmitting it is the product of the tube current multiplied by the plate voltage and is generally given in watts. Generally speaking, the power for which the tube is designed. (See *Transmitting Tube, Power Rating,* also *Vacuum Tubes.*)

EMISSIVITY OF FILAMENT — The ability of a filament to emit electrons when heated by an electric current. (See *Electron,* also *Vacuum Tube.*)

EMPIRE CLOTH OR PAPER—A closely woven cambric coated with two or more films of an oxidized oil, or a tough paper treated with some oxidized oil. Empire products serve as insulating materials for transformers, condensers and various other devices requiring thin insulating sheets. In a transformer for relatively high voltages, insulated paper is used between the layers of windings and in certain types of fixed condensers it is used as a dielectric material. The cloth has a higher breakdown strength; that is, it will stand higher voltages than paper. (See *Breakdown Potential,* also *Dielectric.*)

ENAMELED WIRE — Wire, generally copper, covered with a coating of enamel as an insulating covering and to protect it from moisture and corrosion. (See *Magnet Wire.*)

ENDODYNE—A term occasionally used to describe reception of radio signals by means of locally generated oscillations such as in the case of the *heterodyne* (q.v.), but wherein these oscillations are created by a circuit or system that is part of the receiving circuit and not a separate arrangement. (See *Self Heterodyne.*)

ENERGY—The capacity for doing work in an electrical sense. Energy is something furnished to a body by means of work done on it. Thus, if a ball is thrown, the operation of throwing is the work, and the ball itself in traveling performs work, the means of doing this work being the energy. There are two kinds of energy, *potential,* as in the case of electricity, and *kinetic,* the energy of a body due to motion. Kinetic energy belongs properly in the realms of mechanics and so will be passed over here with only this brief mention.

When we speak of a difference of potential in a primary cell, it means in effect that the body or cell has potential energy. That is to say, the energy is present although not active. If, now, a wire is connected across the two terminals of the cell, the difference of potential will cause electric current to flow from one pole to the other. Actually, in a cell of this type, we have chemical energy stored up. When a circuit is completed, the stored chemical energy is converted into electrical energy or creates the difference of potential, which in turn causes current to flow and electrical *work* is thus done. (See *Potential,* also *Voltage* and *Generator.*)

EPSTEIN HYSTERESIS TESTER — A device for measuring *hysteresis* and *Eddy current loss* in a sample of sheet iron by using an alternating current wattmeter. The device is used for determining the properties of samples of iron, particularly those to be used for transformers. (See *Eddy Currents,* also *Transformer, Hysteresis* and *Wattmeter.*)

EQUIVALENT RESISTANCE — The value which a resistance would be required to have in order to permit the same quantity of current to flow with the same applied electromotive force (voltage) as in the case of a piece of apparatus wherein there are other factors besides pure resistance to determine the amount of current. (See *Impedance,* also *Reactance.*)

ERG—The fundamental unit of mechanical work under the *C.G.S. system* (q.v.). This is defined as the work done by a force of one *dyne* (q.v.) when the body producing the force and the body on which the force acts are one centimeter apart. (See *Joule,* also *Practical Units.*)

E.S.U. — Abbreviation sometimes used for *electrostatic units* (q.v.). (See *Electromagnetic Units.*)

ETHER—or **AETHER**—The term used to denote the supposed or hypothetical medium by means of which electromagnetic and light waves are propagated. (See *Ether Waves.*)

ETHER WAVES—Term sometimes applied to electromagnetic or radio waves on account of the assumption that they travel through the supposed medium ether. The assumption of an all-pervading medium such as ether has been the basis of the major part of the extensive research work carried on with light and electromagnetic waves for many years. Various experiments have shown beyond reasonable doubt that there is some such medium, capable of conducting or rather permitting passage of light waves and radio waves. So accurate has this work been that it has been proven beyond question that light waves travel through the ether with the tremendous velocity of 186,000 miles per second (approximately). It has also been determined that radio waves travel at this same speed through the ether. It has further been proven that radio waves all travel at the same identical velocity, regardless of their length, form and other factors. (See *Electromagnetic Waves,* also *Theory of Propagation of Electromagnetic Waves.*)

EXCITER—A small direct current dynamo used for the purpose of exciting the field magnets of an alternating current generator. Such a device is generally a part of the main machine and mounted on the main shaft, but it may also be a separate unit. The purpose is to furnish current to the field magnets (electromagnets) to create a magnetic flux. (See *Alternator*, also *Generator.*)

EXPLORING COIL—A coil of insulated wire wound on a rotating form as an integral part of a direction finder or *goniometer*. (See *Direction Finder*, also *Goniometer.*)

F

F—The abbreviation for *dielectric field intensity* (see *Field*, also *Dielectric*), also used as the symbol for the unit of *magneto-motive force*, the *gilbert* (q.v.). The small or lower case f is sometimes used to denote the *farad*, the practical unit of *electrostatic capacity* (q.v.) and also as a symbol for *frequency* (q.v.).

FADING—The tendency of radio signals to decrease in volume or amplitude, or to fade. The strength of a signal is seldom the same at different hours of the day or night, generally being stronger during the dark hours of the night. This has been attributed to a variety of causes, the most plausible being that the air is more free of disturbances during the night and thus permits more ready passage of the waves through the ether. The difference between the strength of signals in the daytime and at night is sometimes very pronounced, but the phenomenon of fading is usually noticed only when distant signals are being received. It has been fairly definitely established that fading of signals is due to the reflection of the transmitted waves. (See *Wave Distortion*, also *Reflection and Refraction.*)

FALL OF POTENTIAL—See *Potential Drop.*

FAN AERIAL—An aerial, usually for transmission, constructed in the shape of a fan. (See *Aerial.*)

FAN CONNECTOR—A small triangular or fan-shaped piece of metal arranged with several holes along one edge, with set screws for fastening in each the several wires of the lead-in of a multi-wire aerial. The lower apex is fitted with some sort of connecting device to permit a single wire to be led in to the receiver. (See *Lead-In.*)

Illustration by
Courtesy of Radio Specialty Co., Inc.

Fan Connector—The leads from each wire of a multi-wire aerial go to the wide portion and are held by the set screws: the lead to the set goes to the small end.

FARAD — The unit of *electrostatic capacity*. It is the practical unit of capacity and is defined as the capacity of a condenser that is capable of being charged to a potential difference of one volt by a charge of one ampere for one second; i.e., one *coulomb* (q.v.). "Farad" is a contraction of the name of the distinguished English scientist *Michael Faraday* (q.v.). Inasmuch as the farad is too large a unit for practical purposes, the *micro-farad*, or one millionth of a farad, is commonly used. (See *Electrostatic Capacity*, also *Condenser.*)

FARADAY, MICHAEL. Born 1791, died 1867. An English scientist famous for his discoveries in chemistry, electricity and magnetism. He first produced the rotation of the magnetic needle around the electric current in 1821, based upon Oersted's discovery of electromagnetism in 1820; he discovered electromagnetic induction (1831), a principle upon which is founded the development of dynamo machinery; specific inductive capacity (1838); magnetic polarization of light (1845); diamagnetism (1846). He was a brilliant experimenter, and contributed greatly to the knowledge upon which is based present-day practice of electricity.

Michael Faraday.

FEEBLY DAMPED—In spark transmission, signals that are only slightly damped. Where each succeeding oscillation in a train of oscillations is of slightly lesser *amplitude* (q.v.) than the one preceding it, the signals are said to be slightly damped to distinguish them from signals that are highly damped and thus die out quickly. (See *Damping*, also *Damped Waves.*)

FEED—In electrical parlance, to furnish with a current of electricity. (See *Feed-Back.*)

FEED-BACK—A term applied in radio to the coupling of one circuit to another, whereby a portion of the current present in one circuit is fed back or returned to the other circuit. The common use of the term is in reference to regeneration, where a certain portion of the radio frequency current present in the plate circuit of a detector is returned to the grid circuit of the tube by some form of coupling between the two circuits. It may also apply to the tendency of high-frequency currents to become transferred from one circuit to another due to undesired coupling between two or more circuits. The illustration shows a simple feed-back or regenerative circuit. The incoming signal energy is transferred from the primary of the coupler to the secondary by means of inductive coupling and tuned by the condenser C. A grid condenser and leak GC are placed in the grid circuit in the customary manner. A feed-back or reaction coil T is placed in the plate circuit and inductively coupled to the main coil. A portion of the radio frequency current in the plate circuit or output circuit is transferred back to the grid circuit by the coupling effect between the coil T and the secondary of the coupler. This serves in effect to reduce the resistance of the grid circuit and produces more current in the output or plate circuit, thus acting as a form of amplifier. (See *Regeneration, Tickler.*)

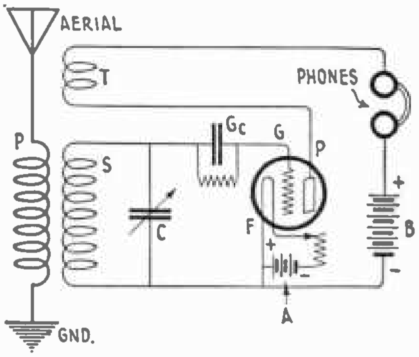

Regenerative circuit in which the coil "T" returns, or "feeds back" to the grid circuit a portion of the radio frequency current in the plate circuit, resulting in increased amplification.

FEED-BACK COIL—An inductance coil placed in the plate circuit of a regenerative receiver and arranged in inductive relation to the grid coil (secondary of tuner) for the purpose of effecting feed-back or regeneration. (See *Feed-Back*, also *Regeneration.*)

FEED-BACK COUPLING—Any process by which the plate circuit of a detector is coupled to the grid circuit to allow a portion of the output current to be transferred back to the grid circuit to produce regeneration. (See *Feed-Back*, also *Regeneration.*)

FEED-BACK EFFECT — The effect of feeding back of currents from one circuit to another as in the case of a *regenerative circuit* (q.v.). Generally the expression "feed-back effect" refers to an undesired feed-back. This may be the action of radio frequency components of the currents present in one part of a circuit being fed back to another part of the circuit due to the coupling effect of various coils or other devices used in the circuit. (See *Feed-Back*, also *Regeneration.*)

FEED-VOLTAGE MODULATION — A process of modulating or varying the amplitude of a radio frequency alternating current to correspond to any wave form of speech vibrations as in radio broadcasting. The system involves the introduction of additional power into the circuit of the radio frequency generator until the desired wave form variations are obtained. This is done particularly in the case of a vacuum tube transmitter, in which case the voltage supplied to the plate of the tube is altered (increased or decreased) to accomplish the purpose. (See *Modulation.*)

FERRO-MAGNETIC MODULATOR—A device for modulating or varying the amplitude of radio frequency current to correspond in wave form to the speech or music vibrations in broadcasting. The system makes use of what is known as the *hysteretic energy absorption* of iron, or in some cases utilizes the variation of inductance of iron-core coils. (See *Modulation,* also *Ferro-magnetic Substances.*)

FERRO-MAGNETIC SUBSTANCES — Elements or compounds such as iron, nickel, cobalt and others which are strongly attracted by magnetic fields of force. Such substances vary widely in their magnetic properties when subjected to temperature changes. Soft iron loses its magnetization almost immediately on removal of the magnetizing source, and for this reason it is used as a pole piece in an electromagnet. (See *Electromagnet,* also *Magnetism* and *Magnetic Properties.*)

FESSENDEN, REGINALD AUBREY. —Canadian-American Radio expert. Born at Milton, Canada, October 6, 1866, and educated at New York and Port Hope, Ontario. Fessenden became inspecting engineer for the Edison Company, New York, and afterwards professor of physics and electrical engineering at Western University, 1892. Professor Fessenden is the author of a well-known system of wireless, and below are briefly described some of the patents bearing his name.

In 1906 and 1907 Fessenden invented a number of microphone transmitters which carried heavy currents for long periods, and also a heavy current telephone relay which allowed the controlling of heavy currents by means of small currents originating in an ordinary microphone circuit or coming from a telephone line. One of these transmitters was called by Fessenden a trough transmitter. It consisted of a soapstone annulus to which were clamped two plates having platinum-iridium electrodes. Through a hole in the center of one plate passed a rod attached at one end to a diaphragm, and at the other to a platinum-iridium spade. The two outside electrodes were water-jacketed. This form of transmitter required no adjusting, all that was necessary being to place about a teaspoonful of carbon granules in the center space. It was able to carry as much as 15 amperes continuously without articulation falling off, and had the advantage that it never packed. By a combination of the trough transmitter and a differential magnetic relay, Fessenden produced a transmitting relay for magnifying very feeble currents. An amplification of fifteen times is possible without any loss of distinctness. Fessenden is also responsible for a duplex system of Radio telephony, and the *heterodyne* method of reception is due to him. Fessenden has written largely on radio subjects, and is one of the leading authorities on both transmission and reception.

Reginald Aubrey Fessenden.

FIBER—or **FIBRE**—A term used for a variety of substances, composed of fine slender thread-like materials, particularly a composition of vulcanized or compressed paper. It is much used in radio as an insulating material for panels and connection blocks. It is not as efficient as insulating medium as bakelite or some forms of rubber composition, but is durable, cheap and comparatively easy to drill. (See *Bakelite.*)

FIELD—The space occupied by electric or magnetic lines of force. (See *Electric Field.*)

FIELD DENSITY—The strength of an electric or magnetic field. It is measured by the number of electric or magnetic lines of force contained in a given cross-sectional area. (See *Density. Flux,* also *Electric Field* and *Field, Magnetic.*)

FIELD, ELECTRIC—The space traversed or occupied by electric lines of force. (See *Electric Field.*)

FIELD, ELECTROMAGNETIC — The total area traversed by electromagnetic lines of force. The region in which electromagnetic lines of force are exerted. (See *Field,* also *Electrostatic Force.*)

FIELD, ELECTROSTATIC—The region in which electrostatic forces are present or exerted. The total area traversed by electrostatic lines of force. (See *Field,* also *Electrostatic Flux.*)

FIELD MAGNET—An *electromagnet,* usually a core of soft iron surrounded by a winding of insulated wire, employed in generating machines to produce a strong magnetic field. As the *armature* of a *generator* revolves the conducting surfaces pass through the *magnetic field* created by the *field magnets.* This field may be merely the ordinary magnetic force due to permanent magnets, but is usually due to the powerful magnetic lines of force resulting from passage of electric current through electromagnets. (See *Alternator,* also *Alternating Current, Theory of Production* and *Electromagnet.*)

FIELD, MAGNETIC—The region through which magnetic flux passes. Usually refers to the space (air) through which the magnetic flux lines pass as distinguished from the activity in the iron path tself. (See *Field,* also *Magnetic Flux.*)

FIELD REGULATOR or RHEOSTAT—Any device for varying the magnetic strength of the field magnets in a dynamo or electric motor, by which variations certain changes may be accomplished in the operation or performance of the machine. A field rheostat is often used in an electric motor to vary the strength of the field and thus either increase or decrease the speed of the machine without the necessity of changing the applied voltage. If the strength of the field (magnetic) is reduced by insertion of resistance, the armature or rotating part speeds up in an attempt to generate the same *counter-electromotive force* (q.v.), which means greater speed. On the other hand, if the resistance is removed—that is, the strength of the field increased—the fact that the armature is running through a stronger field reduces the speed of rotation, the same counter-electromotive force being present. The use of a field regulator or rheostat in connection with a generator is for the purpose of varying the output voltage without the necessity of altering the speed of the machine. The illustration shows the connection of a field rheostat or regulator in the circuit of the *shunt wound generator* (q.v.). (See *Generator, Rheostat, Resistance.*)

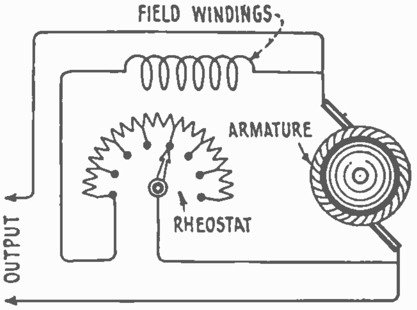

Illustration showing method of connecting a field rheostat or regulator in the circuit of a shunt wound generator.

FILAMENT—A fine wire, one of the elements of a vacuum tube, which is heated to throw off electrons. The term is also used to refer to the fine wire used in an incandescent lamp. The filament of a vacuum tube is generally made of tungsten and is supported by glass pillars within the tube in such manner as to be insulated from the other elements. Filaments are made to be heated with various voltages from 1.1 to 6 volts in most receiving tubes, but much higher in the case of high power transmitting tubes. The filament may consume considerable current as in the old styles of six-volt tubes which consumed one ampere. The modern tubes have thoriated filaments, tungsten treated with thorium, which permits maximum efficiency with a minimum of current consumption. When the filament is heated in the vacuum tube, by having the proper amount of voltage applied, electrons are thrown off and fill the space between filament and grid and grid and plate as shown in the illustration. (See *Electron Theory,* also *Vacuum Tube, Theory of Operation.*)

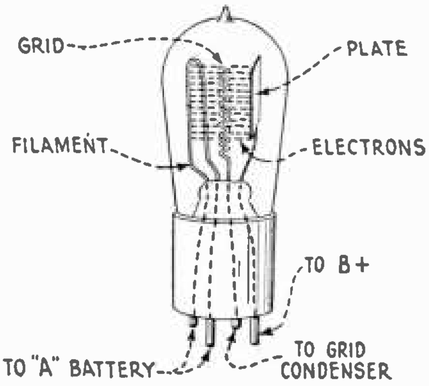

When an "A" battery is connected to the filament terminals, heating the filament, electrons are thrown off and fill the space between filament, grid and plate.

FILAMENT BATTERY—The battery or cell used in lighting the filament of a vacuum tube. These range from a *dry battery* (q.v.) having a voltage of one and a half volts, to storage batteries of six volts. Filament batteries are connected across the filament of a vacuum tube in the manner shown in the illustration. The rheostat is inserted to permit the supply of current to the filament to be varied and thus change the operation of the tube. (See *Filament*, also *Vacuum Tube, Theory of Operation*.)

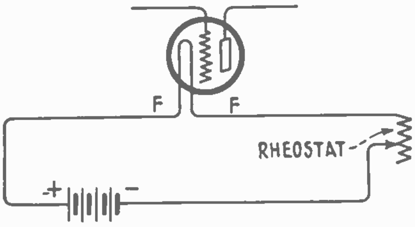

Method of connecting "A" battery in series with rheostat to regulate filament of vacuum tube.

FILAMENT CONSUMPTION — The amount of current consumed by the filament of a vacuum tube. The current requirements of vacuum tubes are always stated by the manufacturer, either in amperes or fractional amperes, or in the case of extremely low consumption, in thousandths of an ampere (milliamperes). Some typical examples are the UV201A and C201A types, which require one-quarter ampere at five volts; the WD11 and 12 and C11 and 12 types, which require one-quarter ampere at approximately 1.1 volts, and the UV200 and C200 tubes, which require one ampere at six volts. Of the so-called standard types, the UV199 and C199 tubes consume the least current, requiring but sixty milliamperes or .06 ampere, with from three to four and a half volts. The statement that a certain tube consumes one-quarter ampere indicates that in operation one-quarter ampere is consumed each hour, this current being supplied by dry cells or storage batteries known as the *"A" battery* (q.v.).

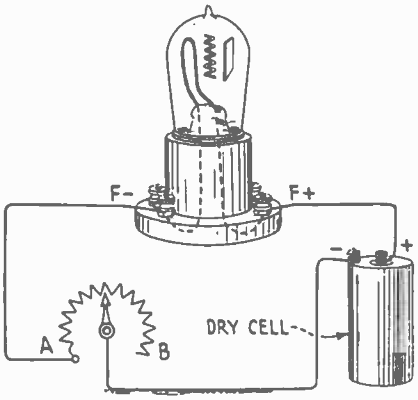

The current produced by the dry cell battery, flowing through the rheostat and filament, is consumed by the filament, at a rate depending upon the type of tube.

It will be apparent from the above that the more current consumed by the filament the greater will be the drain on the dry cells or storage battery. For this reason, and also to lengthen the life of the filaments, it is well to operate the tubes at the lowest efficient brilliancy. The brilliancy of the filament may be controlled by adjustment of the filament rheostat. In the illustration the usual "A" battery, which in this case is a dry cell, is shown connected to the filament of a vacuum tube with a rheostat in series between one filament connection and the negative terminal of the dry cell. Now as the arm of the rheostat is moved from A toward B resistance is introduced into the circuit between the filament and dry cell. As this resistance increases the current in the circuit of the filament will decrease according to *Ohm's Law* (q.v.); i.e., Current = Voltage ÷ Resistance, the current in amperes or fractional amperes and the resistance in ohms. (See *Ampere, Current, Filament Resistance or Rheostat* and *Vacuum Tube*.)

FILAMENT CURRENT — The current used to light the filament of a vacuum tube. Different tubes require different amounts of current; that is, one tube may consume a small fraction of an ampere and another may require a full ampere. Of the tubes most commonly used, the UV199 type uses the least current, this tube requiring only .06 of an ampere each hour. The WD11 and 12 types require one-quarter ampere per hour and also the UV201A, while the now obsolete UV200 and UV201 required a full ampere. The 199 type operates with filament voltage of 3 volts, although it is often used with three dry cells in series which deliver 4½ volts; the WD11 and 12 require 1.1 volts each, which is usually furnished by a dry cell, while the UV201A type is operated with a six-volt storage battery. (See *Vacuum Tubes, Types of*, also *Vacuum Tube, Theory of Operation*.)

FILAMENT, FORMULA FOR ELECTRON FLOW—The extent of electron emission—that is, the number of electrons—emitted from a hot filament in a vacuum tube is subject to certain definite laws. In ordinary practice the actual number of electrons are not referred to, the ability of the filament to emit electrons being measured in fractional amperes of output current. Thus a tungsten filament having a diameter of .0125 centimeters operating with a power dissipation of 3.1 watts per centimeter of length has been found to emit a current of .03 amperes. This result was obtained with a filament having a clean surface and with no absorbed gas on its surface. The formula for electron flow most commonly used is as follows:

$$I_{B} = AT^{1/2}\,\epsilon - \frac{b}{\tau}$$

I_B is the electron current in milliamperes per square centimeter of filament surface, T is the absolute temperature and A and b depend on the metal used for the filament.

FILAMENT RESISTANCE or RHEOSTAT—A device placed in the filament circuit of a vacuum tube, connected between the filament battery and either filament lead on the tube socket for the purpose of reducing or controlling the filament voltage. The illustration Fig. 1 shows one type of filament rheostat which is used to permit variation of the filament current. Fig. 2 shows the manner of connecting such resistance in the filament circuit of a vacuum tube. Fig. 3 illustrates a type of filament control known as an automatic rheostat. In this form of filament resistance a fine wire is placed in a glass tube filled with an inert gas. The wire has the property of increasing its resistance rapidly with any increase in the current passing through it. It is thus possible to arrange such a device to pass only a fixed amount of current for any type of vacuum tube. The device is thus not only an automatic rheostat, but also a safety valve which prevents excessive current reaching the filament and injuring it. These controls are coming into widespread use due to these features and the fact that they reduce the number of controls necessary to operate a receiver. Fig. 4 shows a resistance which is used in series with a rheostat having small resistance in order to enable its use with tubes requiring less current to operate. (See *Filament*, also *Vacuum Tube*.)

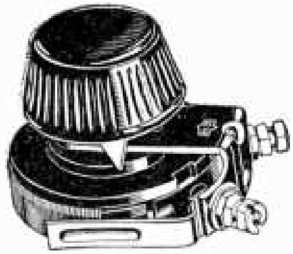

Illustration by Courtesy of The Cutler-Hammer Mfg. Co.

Fig. 1. A type of rheostat used to control the flow of current to the filament of a vacuum tube.

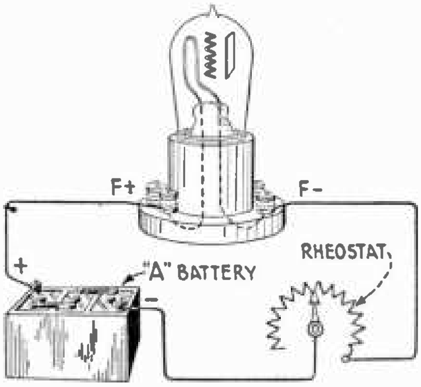

Fig. 2. Showing method of connecting a rheostat in "A" battery circuit of a vacuum tube.

Illustration by Courtesy of the Radiall Co.

Fig. 3. A type of automatic rheostat.

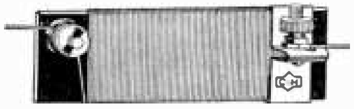

Illustration by Courtesy of The Cutler-Hammer Mfg. Co.

Fig. 4. Resistance used in series with a rheostat having a small resistance in order to enable its use with tubes requiring less current.

FILAMENT, THORIATED—A filament or heating element of a vacuum tube in which a certain amount of *thorium* (q.v.) has been incorporated. The introduction of this rare metal into the filament in small quantities gives it the property of emitting electrons at a relatively low temperature. From a practical standpoint this permits a vacuum tube to be operated with a low *filament consumption* (q.v.). A common example of the use of thorium in filaments is that of the UV201A or C201A vacuum tubes, which operate with a filament voltage of 5 volts and consume but one-quarter ampere of current, giving essentially the same results as the former type UV201 and C201 tubes which consumed one ampere. The value of thorium used in this manner is readily appreciated, the current consumption being materially lessened and the life of the tube conceivably lengthened. (See *Thorium Treatment of Filaments*, also *Vacuum Tubes, Type UV201A and C201A.*)

FILTER—In radio, a device arranged to permit passage through an electrical circuit of currents of certain frequency or *frequencies* (q.v.) while excluding others of different frequencies, or any device for producing smooth direct current from a generating source such as a *direct curent dynamo*, or for smoothing out *rectified direct current.* Filters of one sort or another are much used where rectified direct curent or direct current from a generator is to be used for operation of tubes in radio transmission or reception. In the case of rectified direct current, that is the current derived from any device for changing alternating to direct current, there is always present a tendency on the part of the current to surge or ripple. In other words, it is not smooth and of constant voltage, or there may be present a generator hum which will seriously interfere with proper operation of a tube. The illustration shows the conventional method of filtering rectified direct current for use as plate supply with vacuum tubes in receiving sets. This system may also be used in connection with the output of a direct current generator or dynamo. A filter circuit may be arranged to pass a band of frequencies (see *Band Pass Filter*) or it may be made to pass either high or low frequency current. (*See Filter, High Pass;* also *Filter, Low Pass.*)

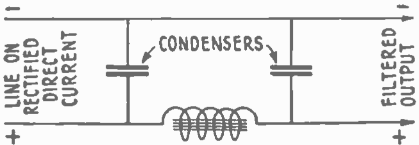

Conventional method of filtering rectified direct current for use as plate supply of a vacuum tube.

FILTER, BAND PASS—See *Band Pass Filter.*

FILTER CONDENSER—The condenser, either fixed or variable, forming a part of a circuit for filtering. (See *Filter*, also *Acceptor*, *Wave Trap* and *Rejector.*)

FILTER, HIGH PASS—A filter circuit arranged to permit current above a certain frequency to pass freely while blocking out or attenuating all frequencies below it. (See *Filter*, also *Filter, Low Pass.*)

FILTER, LOW PASS—A filter circuit arranged to permit currents below a certain frequency to pass freely while blocking currents of higher frequency. (See *Band of Frequencies*, also *Filter, High Pass.*)

FIXED COILS—Inductance coils as used in radio circuits, having no mechanical means of variation, such as in a *Variocoupler or movable coil.* (q.v.) The majority of tuned radio frequency receivers employ fixed coils, shunted by variable condensers. These are known as *tuned radio frequency transformers.* (q.v.) (See *Coupler, Loose*, also *Coupler, Vario.*)

FIXED CONDENSER — A condenser having a certain definite or fixed value of capacity. Any condenser that is not variable. (See *Condenser*, also *Condenser, Variable.*)

FIXED RESISTANCE—Any resistance device having a fixed or non-variable value. (See *Rheostat*, also *Potentiometer.*)

FIXED DISCHARGER—A fixed *spark gap.* (q.v.) A spark gap used in transmission of radiotelegraphic signals in which the electrodes or sparking surfaces are not moved during operation. (See *Disc Discharger.*)

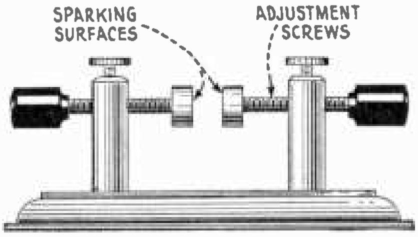

A type of fixed spark gap.

FLAME MICROPHONE—A type of microphone for radio telephone transmission, wherein the height of a coal gas flame is varied by and according to the speech fluctuations and the resulting pulsations made audible by means of a selenium cell and phones. (See *Microphone*, also *Selenium Cell.*)

FLASH SIGNALLING— See *Code*, also *Morse Light.*

FLAT-TOP AERIAL or ANTENNA—An aerial, generally composed of a number of wires placed adjacent to each other equally spaced on a spreader, or hoop, and suspended parallel with the earth. A flat-top aerial may be of the kind known as the *T type aerial* (q.v.) or in the form of an inverted L. (See *Aerial.*)

FLAT TUNING—A term sometimes applied where the *resonance curve* (q.v.) has a flat top, that is where the peak of resonance is not sharply defined. The effect is known as *broad tuning.* (q.v.) (See *Resonance*, also *Tuning.*)

FLEMING, JOHN AMBROSE. British radio expert. Born in Lancaster, November 29, 1849, he was educated at University College, London, the Royal School of Mines, and St. John's College, Cambridge.

From 1873-74 Fleming was demonstrator at the Royal College of Chemistry, and in 1877 he began work under Clerk-Maxwell at the Cavendish Laboratory, Cambridge. There he carried out a series of experimental researches on the British Association Standards of Electrical Resistance. In 1881 Dr. Fleming was appointed the first professor of mathematics and physics at University College, Nottingham, but the following year he joined the Edison Electric Light Company, and on the amalgamation of the Edison and Swan Companies, he was appointed advising electrician, a post he held for twenty years. In 1885 Fleming was appointed to the newly founded professorship of electrical engineering at University College, London, and he was entirely responsible for the design and equipment of the new electrical and engineering laboratories which were opened in 1893. In 1892 he was elected Fellow of the Royal Society.

Dr. Fleming has been very closely associated for many years with radio telegraphy and telephony, and has since 1899 acted as the scientific advisor to Marconi's Wireless Telegraph Company. He is well known as the inventor of the thermionic two electrode vacuum tube, or glow-lamp detector, which was one of the greatest steps forward in radio telephony.

Dr. Fleming has written a very large number of papers and books on wireless telegraphy and telephony, and his "Principles of Electric Wave Telegraphy" is the standard treatise on the subject. He has also published an "Elementary Manual of Radio Telegraphy and Radiotelephony"; "The Wonders of Wireless Telegraphy"; "The Thermionic Valve and Its Development"; "A Pocket Book for Wireless Telegraphists," and other books. As a lecturer at the Royal Society of Arts he has obtained a world-wide reputation for his lectures on electric oscillations and electric waves, Hertzian wave telegraphy, high-frequency measurements, etc. He has given many lectures on wireless at the Royal Institution.

John Ambrose Fleming.

Fleming was awarded the Hughes gold medal of the Royal Society in 1910 as an acknowledgement of the value of his work in electrical science and engineering, and he has twice been awarded the Institution Premium of the Institution of Electrical Engineers, the highest award for communicated papers, and he is an Albert Gold Medallist of the Royal Society of Arts awarded to him for the pioneer invention of the thermionic valve.

FLEMING'S RULE—A rule for determining the direction of the *induced current* (q.v.) in a circuit. Hold the right

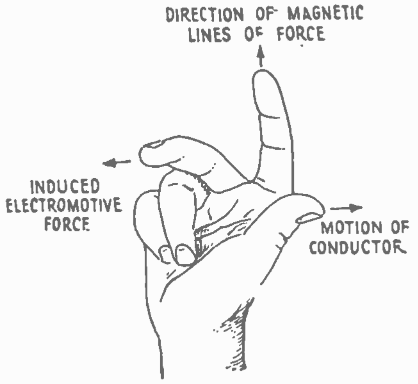
DIRECTION OF MAGNETIC LINES OF FORCE
INDUCED ELECTROMOTIVE FORCE
MOTION OF CONDUCTOR

Method of using the right hand in determining the direction of the induced current.

hand as shown in the illustration with the thumb, forefinger and middle finger as nearly as possible at right angles to each other. If the thumb is then pointing in the direction of motion of the conductor and the forefinger points along the direction of the magnetic lines of force, the middle finger will then point in the direction of the induced electromotive force. (See *Induced Current*, also *Electromotive Force.*)

FLEMING VALVE—The forerunner of the modern vacuum tube. It was invented in 1904 by Professor J. A. Fleming and was for a short time used as a detector in place of the customary crystal or coherer types. The value of the discovery of the Fleming valve lies chiefly in its influence on the course of radio development, as in itself it was not nearly as efficient as a good crystal detector. As a result of the discovery of the Fleming valve, however, many leading scientists bent their efforts toward the discovery of better methods of detecting radio waves, culminating in the three element vacuum tube designed by Dr. Lee De Forest.

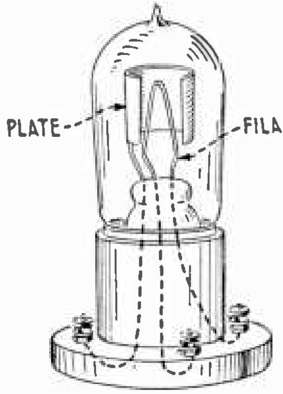

Fig. 1. The Fleming valve, which contained but two elements; i. e., filament and plate. It was the forerunner of the modern vacuum tube.

The Fleming valve consists essentially of an incandescent lamp using a filament of carbon, tungsten or tantalum which can be rendered incandescent by an electric current. The filament is sealed in a glass tube which is highly evacuated, that is from which practically all air has been excluded. Surrounding the filament, but separated from it is a metal cylinder, both ends of the filament and the plate being attached to outside connections by means of leads sealed in the tube. The illustration, Fig. 1, shows the general arrangement of such a tube. The operation of the tube is due to the emission of electrons from the heated filament. It was demonstrated by several investigators that a filament made of certain metals has the property of throwing off minute particles of negative electricity or electrons when heated to a relatively high temperature. It was also found (see *Electron*) that these electrons are thrown off in greater profusion when the filament is placed in high vacuum. It was also found that the escape of negative electrons from the filament would in time leave it positively charged, this charge having an attraction for the negative electrons and preventing their escape from the filament. Hence it was necessary to apply a charge of negative electricity to the filament. The electrons were now found to escape in great number, but had the effect of forming a *space charge* (q.v.) of negative electricity around the filament, which in time served to repel the escape of any more electrons. To counteract this tendency a strong positive charge was applied to the metal plate surrounding the filament. This made the plate positive in respect to the electrons and served to attract them away from the filament. The stream of electrons thus affords a path for current to flow from filament to plate, and the action of the tube as a detector of high frequency waves depends on this principle. If the tube is connected as indicated in the illustration, Fig. 2, it will serve to rectify the incoming signals, that is change them from high frequency alternating currents, to pulsating direct current. The action is as follows: The aerial is connected to the secondary circuit by means of a standard coupler. The coils, marked L2, are the primary and secondary of this coupler; C1 and C2 are variable condensers, F the filament of the Fleming valve, B the battery for lighting the filament of the valve and also to supply voltage to the plate, T a pair of headphones, P a 400 ohm potentiometer and R a standard filament rheostat. The incoming signals in the form of high frequency currents, change direction many times each second. Thus, at one instant the plate will have a positive charge while the next instant it will have a negative charge as a result of the changes in direction of the currents impressed across the tube from the antenna circuit. While the plate is positive, electrons will be attracted away from the filament, which at that instant will be negative with respect to the plate. The current thus flows for an instant, stopping when the impressed current changes its direction and makes the plate negative. In this manner we have a flow of current from filament to plate in one direction in the form of surges appearing at every other alternation or change of current. The diaphragms of the phones will respond to these uni-directional surges and will give out sounds almost identical with the original signals transmitted. The tube thus serves as a rectifier of the incoming signals and is known as a detector of high frequency waves, or oscillations. The Fleming valve principle is used in rectifying tubes for changing alternating current to direct current for charging batteries, its use as a detector for radio reception having died out almost entirely. (See *Charger, Storage Battery, Detector, Tungar, Rectigon,* also *Vacuum Tube.*)

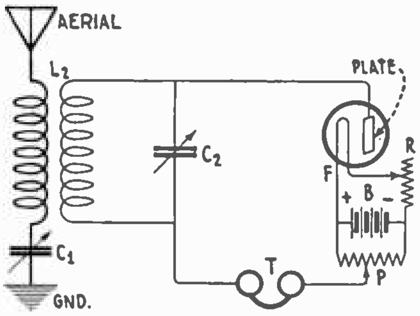

Fig. 2. Method of connecting a Fleming valve in a radio receiving circuit.

FLEWELLING CIRCUIT—A modification of the Armstrong Super-regenerative circuit, designed by the American experimenter, Flewelling. In this system, the tendency toward *self-oscillation* (q.v.) is retarded by the use of fixed condensers which are arranged to have a *damping* (q.v.) effect on the grid of the detector tube, automatically controlling the oscillations of the tube. The circuit controls the self-oscillation of the tube without the necessity for the separate control tube used in the Armstrong circuit. The illustration shows the original Flewelling circuit. It is essentially a standard regenerative circuit with certain necessary modifications.

The coils L1 and L2 give magnetic coupling between the plate circuit and the grid circuit. C1 is a variable condenser for tuning and condensers C2, C3 and C4 are of the fixed type and afford a path of relatively low *impedance* (q.v) for the high frequency currents. C5 is the customary grid blocking condenser across which a grid leak is shunted. The usual telephone receivers T, the plate battery B and the filament battery A are included in the circuit in the customary manner. The coils L1 and L2 may be two honey-comb coils having respectively 50 and 75 turns. The tuning condenser C1 should be .0005 mfd. for the broadcast band of wave-lengths and the loop may be almost any form.

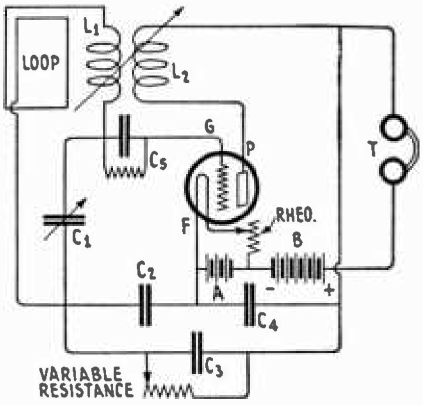

The Flewelling circuit.

In operation three forms of current pass through the plate circuit including the coil L2, namely, the direct current from the plate battery B, the unidirectional surges of rectified current due to the action of the tube as a rectifier and the high frequency currents transferred through the medium of the magnetic coupling of the two coils L1 and L2. The battery current is present only in the plate circuit, the condensers C2, C3 and C4 blocking the direct current from the grid circuit. These condensers also block the rectified currents but afford two paths for the high frequency currents or oscillations. The condensers C2, C3 and C4 have definite values, respectively .005, .005 and .006 micro-farads, to allow a by-pass through the grid to the filament, thus causing accumulation of electrons on the grid and stopping oscillation of the tube. The charge then leaks off by means of the grid leak connected across C5, the tube then building up oscillations again and continuing the process during operation. The high resistance connected across C3 is variable and permits control of the by-pass of high frequency current of the grid. Careful adjustment is necessary to obtain maximum signal strength without distortion. (See *Armstrong Circuits, Regenerative Circuit,* also *Super-Regenerative Circuit.*)

FLEX—Term used in Great Britain and occasionally in this country to refer to a number of small wires twisted or stranded together to form a single flexible conductor. (See *Flexible Wire.*)

FLEXIBLE LEAD—Any conductor, particularly the connections from rotary parts of couplers or condensers, in which the conductor is made up of a number of fine wires in such a way that the necessary bending in operation of the unit will not snap the wire as would be the case with a solid conductor. (*See Pig-tail*, also *Flexible Wire.*)

FLEXIBLE WIRE—Any conductor composed of a number of fine wires to permit constant handling and bending without breaking. (See *Flexible Lead.*)

FLOAT, BATTERY — Term sometimes applied to the bulb or float in a hydrometer for testing the condition of storage batteries. (See *Hydrometer.*)

FLOW OF CURRENT—An expression used to denote the passage of electric current along a conductor. (See *Current, Direction of Flow* (*Assumed*), also *Current.*)

FLOW OF MAGNETIC FLUX—The passage of lines of magnetic force through any magnetic circuit. (See *Flux*, also *Field, Magnetic.*)

FLUCTUATING CURRENT—An electric current that does not have steady value, rising and falling in strength or pressure. A current that is not constant. (See *Constant Current*, also *Voltage, Constant.*)

FLUORESCENT SCREEN—A screen or surface coated with fluorescent matter, used to exhibit shadows cast by rays from various types of tubes. Its application in radio is chiefly in the *Cathode Ray Tube.* (q.v.)

FLUX—A very broad term in physics. The three more common examples of flux are: (a) to denote *electrostatic or magnetic flux* (q.v.); (b) a substance used in soldering for the purpose of making the solder flow readily and making it adhere to the parts being united, and (c) in metal smelting, a mineral—customarily chalk or limestone—added to the charge in the furnace to absorb mineral impurities and dispose of them in the form of slag. *Flux* is also used to refer to the amount of heat or light flowing or passing through a given distance or area in a given time. (See *Electrostatic Flux, Flux Density*, also *Flux, Magnetic.*)

FLUX DENSITY—The number of lines of force per square centimeter of sectional area of a magnetic path. The unit of magnetic flux in the C.G.S. (Centimeter Gram Second) or electromagnetic system of units is the Maxwell. It is generally referred to as a "line." As a magnetic flux is merely the effect of magneto-motive force, we can take the rule that the magnetomotive force is 1.257 x ampere turns as the rule for determining the intensity or density of magnetic flux. (See *Flux; Field, Magnetic*, also *Ampere Turns* and *Maxwells.*)

FLUX DISTRIBUTION—The distribution of lines of force in a magnetic or electrostatic field of force. (See *Field*, also *Flux.*)

FLUX, ELECTROSTATIC — Another term for electric displacement as in the case of a condenser. The density of electrostatic flux refers to the intensity of an *electrostatic field.* (q.v.) (See *Flux Density, Field, Electrostatic*, also *Electric Displacement.*)

FLUX LEAKAGE—Any loss or dissipation of magnetic or electrostatic flux. (See *Flux, Electrostatic*, also *Flux, Magnetic.*)

FLUX LINES—The lines of force in a field of force in which electrostatic or electromagnetic force is exerted. An alternate term for *lines of force.* (q.v.)

FLUX, MAGNETIC—The term broadly used to denote the grouping of magnetic lines of force in a *magnetic field.* (q.v.). The most common example of magnetic flux is in the case of an ordinary bar magnet as shown in the illustration Fig. 1. Here the lines of force or magnetic flux create a magnetic field indicated by the dotted lines. The number of lines of force in a field of unit area is used as a basis for determining the density of the flux. (See *Flux Density.*) This effect or magnetic flux field is also present when a current of electricity is allowed to circulate around a coil of wire, the larger the current and the greater the number of turns, the more powerful the flux created. The illustration Fig. 2 shows a coil of wire through which current is assumed to be passing, the dotted lines indicating the flux field. (See *Choke Coil, Transformer, Magnet*, also *Flux, Electrostatic* and *Magneto-motive Force.*)

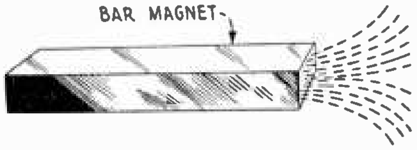

Fig. 1. Bar magnet, showing how the lines of force, or magnetic flux, create a magnetic field as indicated by dotted lines.

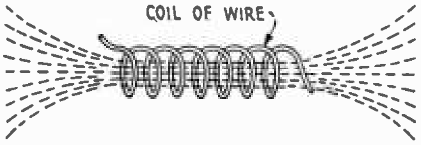

Fig. 2. Coil of wire through which current is assumed to be passing, the dotted lines indicating the flux field.

FLUX MAGNETISM — The magnetic flux, or total number of lines of force passing through a magnetic circuit. It is equal to the product of the magnetic density (see *Flux Density*) and the cross sectional area of the magnetic path. It is also referred to as magnetic flow. (See *Magnetic Flow.*)

FLUXMETER—A device for measuring flux density. The Grassot fluxmeter uses an *exploring coil* (q.v.) in conjunction with a special form of moving coil galvanometer. The coil is suspended on a silk fiber and its movements actuate a pointer across a scale. (See *Flux, Magnetic.*)

FLUX, SOLDERING — See *Soldering Flux, Flux.*

FOCUS or FOCI—The point or points at which converging lines meet. The magnetic foci are the points on the earth's surface in the vicinity of the magnetic poles, where the magnetic force of the earth is the greatest. Essentially a mathematical term.

FORCE—An influence exerted upon a body so as to produce a change or a tendency to change its state of rest or motion. This is mechanical force. In electricity we have its electrical analogy—*electromotive force.* This is the force which causes motion of electricity and sustains the electric currents against the resistance of a circuit. (See *Electromotive Force.*)

FORCE, MAGNETO-MOTIVE—The force which produces magnetic flux as in the case of a magnet moving through a magnetic circuit. The abbreviation is M. M. F. and its unit is the *gilbert.* (q.v.) (See *Magneto-motive Force.*)

FORCED OSCILLATIONS — Electrical vibrations or high frequency oscillations which are imposed upon an *oscillatory circuit* without being in tune or resonance with the natural period of vibration of the circuit. The opposite of a free oscillation in which case the circuit oscillates freely on its own account at the same frequency as the imposed or incoming oscillations or one of the *harmonics* (q.v.) of this frequency. Free and forced oscillations may be present in the same oscillatory circuit, the forced oscillations being set up during the first stages of induction of free oscillations. (See *Free Oscillations*, also *Oscillatory Circuit.*)

FORM FACTOR—The ratio of the *root-mean-square value* (q.v.) to the algebraic mean ordinate taken over one alternation (half cycle), starting with zero value of current. It is the ratio of the r.m.s. value of an alternating electromotive force, current, etc., to its true mean value. (See *Wave Analysis.*)

FORMULA — A principle or rule expressed in mathematical terms or chemical symbols. Formulae are much used in radio to state a problem as, for example, a formula for calculation of wave-length, which is written: $\text{Wave-length}=1884\sqrt{LC}$. This states the problem or rule for the determination of wave-length. Here we have certain values represented by symbols. The capacity is represented by C, the inductance by L, the wave-length in meters being the product of the square root of the product of these two multiplied by the factor 1884. (See *Measurement of Wave-length.*)

FOSTER BRIDGE—A form of *Wheatstone Bridge* (q.v) of the slide wire variety designed to permit comparison of two low value resistances that are nearly equal. It was devised by Professor Carey Foster and is used to determine the difference between two resistances as distinguished from the usual bridge for measuring the resistance of a single unit. (See *Wheatstone Bridge.*)

FOUCAULT CURRENT—Another name for *eddy currents*, the *electro-magnetic currents* due to the circulation of induced current in the mass of a conductor. Foucault first proved that these currents were productive of losses of energy as they cannot be utilized except in certain forms of meters. (See *Damping of Instruments.*) Alternating currents tend to set up eddy currents in metal masses near them. Thus in the case of a transformer, eddy currents are produced in the iron core due to the induced currents circulating through its mass. These currents are wasteful and means are therefore used for reducing them. The core of a transformer may be made of numerous flat pieces called laminations, which serve to prevent the free movement of eddy currents in the core. (See *Eddy Currents*, also *Transformer* and *Core Losses.*)

FOUR ELEMENT or FOUR ELECTRODE TUBE—A vacuum tube having four elements instead of the usual three. The standard tube has three elements, namely, grid, filament and

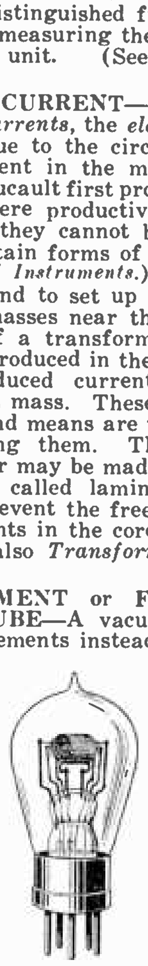

A four element vacuum tube.

plate. The four element tube has an additional grid element. The illustration shows one form of tube having a double grid arrangement. Such a tube can be used for high and low frequency amplification and as a detector simultaneously or it may be used in conjunction with the Solodyne system, which does not use the customary B battery to supply potential to the plate element. (For more complete explanation of the action of a four element tube in the latter manner see *Solodyne*). (See also *Vacuum Tube.*)

FOURTH CIRCUIT—A method of controlling oscillation of a regenerative detector used in the *Cockaday Circuit.* (q.v.)

FRAME AERIAL—An alternate name for *loop aerial,* a type of aerial consisting of a number of turns of wire wound on an insulated frame and used in place of the regular outdoor aerial with sensitive receiving sets. (See *Loop Aerial.*)

FREE ALTERNATING CURRENT—Term applied to alternating current assumed to be produced by the application of a single impulse of electricity to an *oscillatory circuit.* (q.v.) See *Free Oscillations.*)

FREE ELECTRONS—Name given to the outer electrons of an atom, which become detached under various chemical transformations and are assumed to participate in electrical effects or phenomena. Free electrons are so named to distinguish them from the bound electrons of the proton or positive nucleus of an atom. (See *Electron,* also *Proton.*)

FREE MAGNETISM—The portion of magnetism which does not follow the magnetic path through the metal of a magnetized body, but leaves the surface of the magnet. Sometimes referred to as *surface magnetism.* (See *Magnetism.*)

FREE OSCILLATIONS—Oscillations having the natural frequency of the circuit in which they occur. (See *Forced Oscillations.*)

FREQUENCIES—The plural of the term denoting the number of complete cycles per second in an alternating current. In referring to vibrations, particularly audible or so-called *voice frequencies,* the term is used to imply all frequencies that are normally considered audible. (See *Frequency.*)

FREQUENCY—A term adapted from academic English to use in electrical practice. Thus, the number of times a certain action is repeated per unit of time. In the case of an *alternating current* (q.v.) the current reverses its direction of flow at certain definite intervals. If the reversing process takes place 120 times per second—that is 60 reversals from negative to positive and vice versa—the frequency of the current is said to be 60 cycles. In ordinary commercial practice the majority of alternating current generators produce an alternating current or E.M.F. having a frequency from 25 cycles per second to 60 cycles per second. In radio transmission the alternating current or E.M.F. (Electro Motive Force) has an infinitely higher frequency—the reversals taking place many thousands of times each second. (See *Alternating Current, Theory of Production,* also *Alternation, Cycle* and *Periodicity.*)

FREQUENCY, AUDIO—In radio, a frequency within the limits at which audible sounds are produced. It is generally considered as a frequency less than 10,000 cycles per second, although some authorities refer to frequencies up to 16,000 or 20,000 cycles as being within the limits of audibility. For ordinary practice, the human ear is capable of hearing sounds only up to about 2,500 cycles per second. In receiving sets the audio frequency currents appear only in the circuits carrying the rectified currents, although some types of super-heterodyne receiver use an intermediate frequency as low as 30,000 cycles per second, these signals then being rectified and made audible in the telephone receivers. (See *Frequency Circuit, Audio,* also *Frequency, High.*)

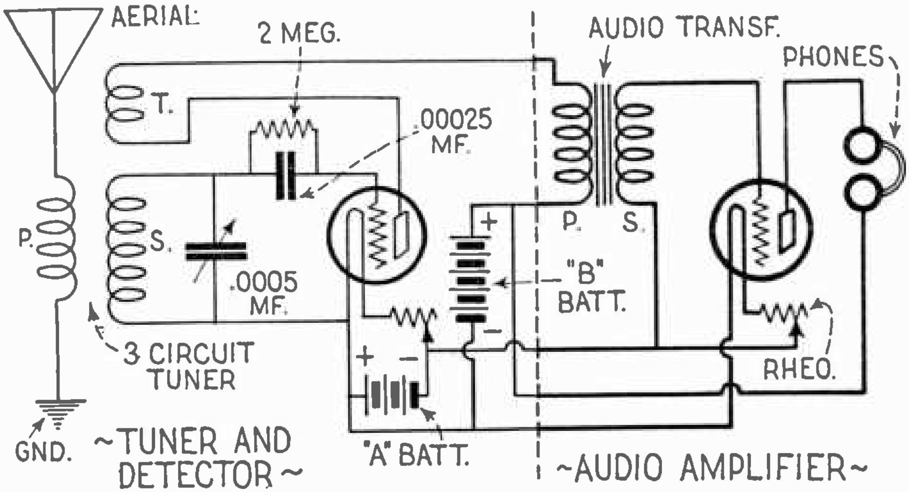

In the diagram above, the heavy lines to the right of the dotted line indicate the audio-frequency circuit of a radio receiver.

FREQUENCY, ALTERNATOR OF—The frequency of an alternating current dynamo or alternator is expressed in cycles per second. With each complete revolution of the *armature* a complete *cycle* or double *alternation* of current is generated. (See *Alternator.*) It is therefore a simple matter to determine the frequency of alternating current generated when the number of revolutions per second and the number of field poles or magnets is known. The frequency will be $f=\frac{N \times S}{2}$, N being the number of field poles and S the number of revolutions per second. The divisor 2 is used to convert the result in *alternations* per second to cycles per second, one cycle representing two alternations. (See *Alternating Current, Theory of Production of,* also *Frequency.*)

FREQUENCY BAND—A term used to denote all the frequencies lying within a certain minimum and maximum limit. Thus, in broadcasting, the frequency band will be the frequencies corresponding to the wave-lengths within the broadcast band. This will be generally between 550,000 and 1,500,000 cycles, or a wave-length band of approximately 200 to 550 meters. (See *Band of Frequencies.*)

FREQUENCY CARRIER—In radio broadcast transmission, the radio frequency (high frequency) wave with which the audio frequency speech variations are superposed. (See *Broadcasting, General Treatise on Methods,* also *Carrier Wave.*)

FREQUENCY CHANGER—A device for converting currents at a certain frequency into currents of a higher or lower frequency, without change of phase or voltage. A frequency converter. Generally an *alternator* driven by a *synchronous motor.* The term is also applied erroneously to certain forms of super-heterodynes. (See *Ultradyne.*)

FREQUENCY CIRCUIT, AUDIO—Any circuit in a radio receiver or transmitter which carries currents within the arbitrary limits of audibility. In the diagram shown herewith, the heavy lines are the connections of the audio frequency circuit. In this part of the circuit only currents of audio frequencies are present. (See *High Frequency Component of Plate.*)

FREQUENCY CIRCUIT, RADIO—The circuits in a receiver in which only high or radio frequency currents are supposed to be present. (See *Feed-Back, Audio Frequency,* also illustration in connection with *Frequency Circuit, Audio.*)

FREQUENCY, FUNDAMENTAL—The frequency of the fundamental wave. (See *Fundamental Frequency.*)

FREQUENCY, GROUP—The frequency of the separate trains of waves in radio communication. (See *Group Frequency,* also *Wave Train.*)

FREQUENCY, HIGH—Frequencies over certain definite limits, generally from 10,000 to 20,000 cycles per second in radio usage. Commonly used as synonymous with radio frequency. It is the opposite of low or audio frequency. (See *Frequency,* also *Audio Frequency, Limits of Audibility.*)

FREQUENCY, LOW—Frequencies below certain limits, generally 10,000 or 20,000 cycles per second. (See *Frequency; Limits of Audibility.*)

FREQUENCY METER—A device for indicating the frequency of an alternating current in cycles per second. The most simple form uses a system of tuned, vibrating reeds. The illustration shows a frequency meter of the

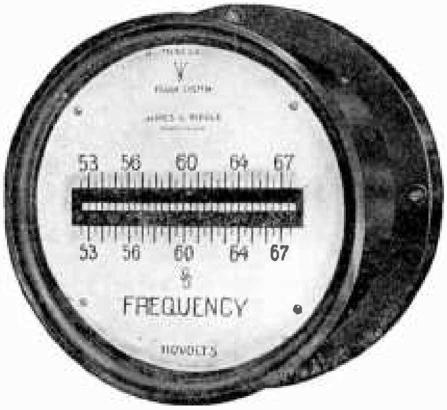

Illustration by Courtesy of Jas. G. Biddle.
A vibrating reed type of frequency meter.

reed type. These reeds may be arranged to respond to various frequencies in a similar manner to the reeds in various musical instruments. These reeds are acted upon by an alternating current magnet. When the magnet is acted upon by an alternating current sent through the windings, the reed nearest to that frequency will vibrate in sympathy with the vibration due to the alternating magnetic effect on the magnet. Such meters and also those of the *induction type* are used for comparatively low frequencies such as occur in commercial lighting systems. For measurement of the high frequencies used in radio communication somewhat different systems are used.

FREQUENCY, RADIO—An alternating current or electromotive force above certain definite limits, generally above 10,000 or 20,000 cycles per second. The line of demarcation between *audio frequency* and *radio frequency* is not very sharply defined, as, in radio usage frequencies that are actually inaudible to the human ear may yet be referred to as audio frequencies. The term is used to differentiate between the rectified or audio currents and the high frequency oscillations such as are impressed on the antenna in receiving or transmitting. (See *Audio Frequency,* also *Frequency, Low.*)

FREQUENCY, UNIT OF—Frequency of alternating currents is expressed in cycles per second and is generally written f. (See *Frequency,* also *Cycle.*)

FREQUENCY, VARIATION OF POWER FACTOR WITH—When a condenser is charged and *dielectric absorption* (q.v.) takes place, it is always accompanied by a loss of power which is apparent in the form of heat produced in the condenser. This loss of power varies with the frequency of the current impressed on the condenser and it is therefore said that the *power factor* (q.v.) varies with the frequency. (See *Condenser, also Dielectric Absorption.*)

FULL LOAD CURRENT—The current delivered from any electrical source, such as a dynamo or generator operating at its maximum load. The maximum current obtainable from an electric source.

FULL WAVE RECTIFICATION — In rectifying alternating current, the process of rectifying both the positive and negative alternations or half cycles. The majority of rectifiers for charging storage batteries rectify only a half wave, or half cycle, and thus do not utilize the full alternating current cycle. (See *Rectifier, Full Wave,* also *Charger, Storage Battery.*)

FULLER CELL—A type of primary cell formerly much used in telephone, telegraph or any intermittent work. (See *Cell, Open Circuit.*) A zinc electrode cast in the shape of a cone is at the end of a soft copper wire and rests at the bottom of a porous cup immersed in a dilute solution of sulphuric acid. The carbon electrode is in an outer jar in a bichromate solution. (See *Cell.*)

FUNDAMENTAL FREQUENCY — The frequency or number of cycles per second of the fundamental wave. (See *Frequency,* also *Fundamental Wave.*)

FUNDAMENTAL UNITS—Units based on the fundamental quantities of length, mass and time. Under the *C. G. S. system* the unit of length is the centimeter; of mass, the gram; and time, the second. The absolute system of units. (See *Centimeter-Gram-Second.*)

FUNDAMENTAL WAVE—In an alternating current, the component of the wave form that is a pure sine wave of the principal frequency. Upon this principal frequency the various higher frequencies (harmonics) may be superposed. A fundamental wave when combined with other frequencies which are a whole multiple of it produce the harmonious note. These terms are derived from acoustics in which the fundamental is the pure sound wave or pure note. (See *Sine Wave of Alternating Current or EMF.*)

FUNDAMENTAL WAVELENGTH — Term used to denote the wave-length at which the inductance and capacity of a coil or antenna would be in resonance. In referring to antennas the fundamental wave-length is the wave-length corresponding to the frequency at which free oscillations will occur. This is often referred to as the natural wave-length and the corresponding frequency is referred to as the natural frequency of the antenna.

For practical purposes the fundamental wave-length of an aerial (antenna) may be computed from the dimensions. Where no loading inductance is used, a single wire flat top antenna (T or inverted L type) will have a fundamental wave-length approximately four times the distance from the ground connection to the extreme end of flat or horizontal portion as measured in meters. Thus, if the antenna is of the L type with a top one hundred feet long and a lead-in and ground connection sixty feet long, the fundamental wave-length will be about 4×(100+60)×1.1, the result being given in feet. Now as wave-length is always stated in meters it will be necessary to divide this result by 3.28 which is the rough conversion factor from feet to meters. The multiplier 1.1 adds ten per cent to the figures for necessary correction of length. If the antenna is of the T type, the length of the flat portion would be considered as half of the total horizontal length or 50 feet instead of 100 feet. The illustration shows method of measuring length. (See *Wave-length Calculation for Antennae.*)

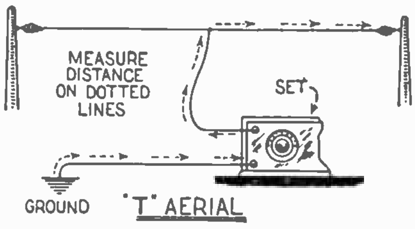

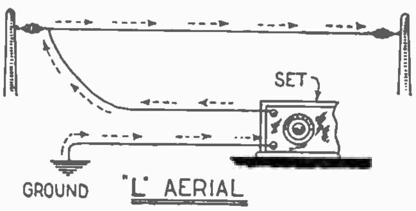

FUSE—A device used in electrical circuits to prevent the passage of excessive currents. Generally a strip of fusible metal, commonly lead with a small percentage of tin, enclosed in a fiber or insulated and fireproof container and inserted in the circuit. When the temperature, due to excessive current, reaches a certain limit, the fuse "blows" causing a break in the circuit. It is thus possible to limit the current flowing in a circuit to a definite maximum in order to protect instruments from excessive charges. The most common use is in commercial lighting and heating systems where the current on any line is generally limited to ten or twenty amperes for the protection of the lines leading in to the building.

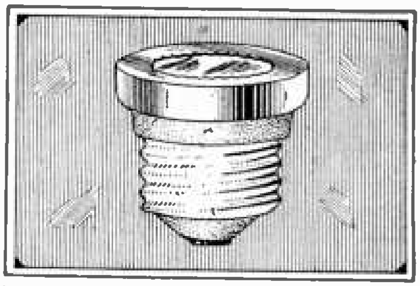

Fig. 1. A common type of fuse used in lighting, heating and power circuits.

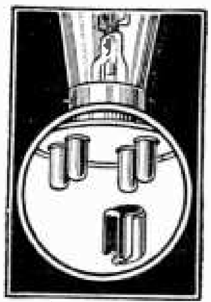

Illustration by Courtesy of Radio Equipment Co.

Fig. 2. A fuse used to protect the filament of a vacuum tube from damage by excessive voltage.

G

GALENA—A natural crystal sulphide of lead. Chemical symbol PbS. It is an ore of lead, grey in color and showing brilliant metallic lustre when fractured in cubical form. Galena crystals are used either alone or in conjunction with other minerals as *crystal rectifiers (detectors)* in radio reception. When used in combination it is generally in conjunction with *graphite* or *zincite.* Used alone, the contact may be a fine wire or *cat whisker.* (q.v.) (See *Crystal Detector.*)

GALVANIC BATTERY—The term, now obsolete, formerly applied to a primary battery or group of primary cells.

GALVANI, LUIGI. Born 1737, died 1798.—An Italian physician and physiologist noted as the discoverer of galvanic or current electricity. While dissecting the legs of a frog, they came by accident into contact with dissimilar metals which caused muscular action in them. Galvani thought that he had discovered electricity in animal matter. Volta attributed the action to the metallic contact and thereupon constructed his voltaic pile, the forerunner of the primary cell.

GALVANOMETER—An extremely sensitive instrument used for detecting and measuring electric currents and also for indicating the direction of the current in a circuit. In its simplest form it consists of a small iron needle pivoted in the center of a hollow coil of wire. The needle is controlled by the earth's field and points north and south when at rest. When an electric current is passed through the galvanometer winding the needle is deflected over a graduated scale the extent of the deflection depending on the force of the current. The instrument is arranged in such a manner that the needle will be deflected in one direction when the current flows through the windings in a certain direction, and when the current is reversed the needle will be deflected in the opposite direction. Thus when the connections are known the direction of the current is accurately indicated as well as its strength. The galvanometer is one of the most sensitive current measuring devices, the deflection of the needle being due to the electro-magnetic field created by the passage of electric current through the coil of wire. (See *Ammeter* or *Ampere Meter,* also *Galvanometer, D'Arsonval* and *Galvanometer, Mirror.*)

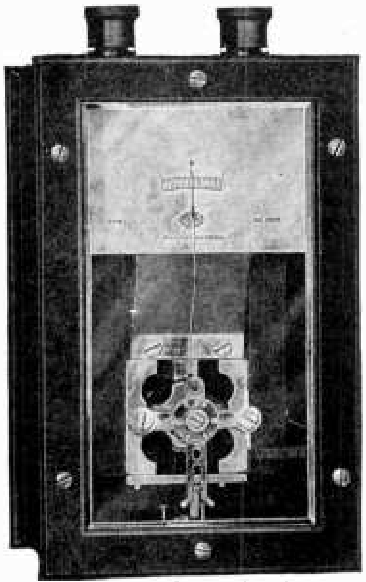

Illustration by Courtesy of the Roller-Smith Co.

A galvanometer which is used for measuring electric currents, and also for indicating the direction of flow of current in a circuit.

GALVANOMETER CONSTANT—Generally speaking, a number used to change the galvanometer reading to ordinary units. The instrument measures extremely small currents and this number times a certain function of the galvanometer reading gives the value in customary units. Technically, the strength of the field produced at the center of the coil of a galvanometer by the unit of current to be used. (See *Constant.*)

GALVANOMETER, D'ARSONVAL—A form of galvanometer having a dead beat indicator (see *Dead Beat Instrument*) in which the indicating coil is suspended in the field of a powerful horseshoe magnet. It is one of the most sensitive forms of galvanometer. The name was derived from that of the designer, A. D'Arsonval. (See *Galvanometer.*)

GALVANOMETER, MIRROR—A type of galvanometer, the moving coil or indicator of which carries a mirror which reflects a beam of light from a fixed lamp. It is so arranged that the lamp is provided with a lens and the beam is so reflected that a visible spot of light is seen on a scale and deflected when the moving element (carrying the mirror) swings with the passage of current. In the beam of light is a sharp image of fine wire stretched in front of the lamp. This type of galvanometer may also be arranged so that the reflection of a fixed scale may be observed in the mirror by means of a small telescope arrangement. (See *Galvanometer.*)

GAMMA—Γ—The third letter in the Greek alphabet. (See *Gamma Type Antenna.*) The lower case *Gamma* γ is used as a symbol for *conductivity.* (q.v.)

GAMMA TYPE ANTENNA—An alternate name for the inverted L type of antenna, so called because the third letter in the Greek alphabet, Γ, resembles an inverted "L," and a single wire antenna with lead-in from one end is in the shape of an inverted L. (See *T Type Aerial.*)

GAP—See *Spark Gap,* also *Lightning Gap.*

GAP, AIR—Term applied to the gap between sections of an iron core as in a transformer for the purpose of preventing *saturation.* (q.v.) Also applied to speak dischargers where a space is provided between the electrodes across which the spark jumps in operation. The term is used as well in referring to the radial depth of the space between surfaces of the rotor and stator in generating machines or motors. (See *Spark Gap,* also *Rotary Gap.*)

GAP ARRESTER—A form of *lightning arrester* (q.v.) in which small air gaps are provided between non-arcing metal electrodes. (See *Arrester.*)

GAP, MICROMETER—A small gap designed to protect apparatus from excessive potentials. (See *Safety Gap.*)

GASEOUS—Anything of the nature of a gas, having no natural physical boundaries but accommodating itself to the shape of the container, expanding to the dimensions of any space in which it may be confined.

GASEOUS TUBE—A term sometimes applied to a vacuum tube in which a certain amount of gas is present. Such tubes under ordinary conditions are not efficient as amplifiers in vacuum tube circuits, the minute gas atoms presenting a more or less effective obstacle to the free passage of the electrons. (See *Vacuum Tube, Theory of Operation.*)

GAS RECTIFIER—The term applied to a vacuum tube which is only partially exhausted of its residual gas. They are also known as *soft* tubes and are generally used as detectors although the tendency now is toward *hard* or highly evacuated tubes due to their adaptability to all purposes in radio receivers whereas the soft tubes do not function well as *amplifiers.* (See *Vacuum Tube.*)

GASSING—When the charging current is continued after a storage battery has taken a normal charge the free liberation of oxygen and hydrogen takes place. This is manifest in the form of gaseous emanation or liberation of bubbles from the electrolyte. The same effect may be noted when a battery is charged at higher rate than the normal charging rate stated by the manufacturer, and under certain conditions may occur at the later stages of charging, even when the battery is not fully charged. In a storage battery that is in good condition it is a fairly accurate indication of a full charge. (See *Charging,* also *Rate of Charging.*)

GAUGE, WIRE—The various systems of measurement used for denoting the physical dimensions of wire. The standard gauge in the United States is known as the B & S or Brown and Sharpe gauge. (See *Brown & Sharpe Guage.*)

GAUSS—The term adopted by the American Institute of Electrical Engineers as applying to the *magnetic flux density.* (q.v.) It is written B or ℬ and is also used in conjunction with the *Gilbert* to denote magnetizing force, one gauss being one gilbert per Centimeter. It is incorrectly used to apply to *magneto motive force.* (See *Centimeter Gram Second or C. G. S.*)

GENERATOR—Any machine for producing electric energy. Usually a device for converting mechanical power into electrical energy. It may apply to a vacuum tube used to generate oscillations as in transmission or for *heterodyne action.* (See *Heterodyne,* also *Vacuum Tube Generator.*)

GERNSBACK, HUGO—Born 1884. Educated Ecole Industrielle, Luxembourg; Technikum, Bingen, Germany. President, The Experimenter Publishing Company, Publishers of Radio News, Science and Invention, Amazing Stories and Radio Internacional. President of The Consrad Company, Publishers of Radio Review, Radio Listeners' Guide and Call Book, and Money Making. Editor of Radio News, Science and Invention, Amazing Stories. Author of "Radio for All," "Wireless Telephone," "Ralph 124C-41+."

Hugo Gernsback.

GERNSBACK, SIDNEY—Born 1876. Educated College of Luxembourg, Lycee de Nancy, France. Vice-President and Treasurer of The Experimenter Publishing Company, Publishers of Radio News, Science and Invention, Amazing Stories, and Radio Internacional. Vice-President and Treasurer of the Consrad Company, Publishers of Radio Review, Radio Listeners' Guide and Call Book, and Money Making. Editor of Radio Review, Radio Listeners' Guide and Call Book, and Money Making. Author of "A Thousand and One Formulas," "Wireless Course in Twenty Lessons," "Experimental Electricity Course," "Radio Encyclopedia."

Sidney Gernsback.

GILBERT—Symbol *F*, is the unit of magneto-motive force adopted by the American Institute of Electrical Engineers. (See *Magneto-motive Force.*) It is equivalent to $4\pi/10$ or .7958 ampere turns. (See *Ampere Turns.*)

GLASS PLATE CONDENSER—A condenser, generally of the heavy duty fixed type, consisting of alternate layers of metal or foil and thin glass sheets as the *dielectric*. Glass plate condensers are mainly used for transmitting, glass having a very high dielectric constant. Such condensers may be immersed in oil to stand higher potentials without breaking down. (See *Condenser*, also *Dielectric Coefficient and Constant.*)

GLASS SILENCER—A glass tube having wooden ends with holes just permitting discharge electrodes or rods to pass, used in conjunction with spark coils and gaps to enclose the sparking surfaces and thus dampen the noise of the sparking. (See *Spark Gap.*)

GLOW DISCHARGE—See *Corona.*

GLIDING THEORY—A theory regarding the propagation of electromagnetic (radio) waves, which assumes the waves as following the curvature of the earth and not being subject to reflection due to objects in their path or by any reflecting surface above the earth. This theory has been discarded; due to observations of phenomena of transmission it now is generally assumed that there must be a reflecting surface high in the atmosphere. However, it is considered probable that the waves do not extend very high as a general rule, the exceptions being the cases of extreme distance transmission in certain directions. In the upper atmosphere, due to the rarification of the atmosphere *ionization* (q.v.) is very pronounced and the air may become a fairly good conductor. This conducting region or layer may impose severe obstacles to the electromagnetic waves and thus limit them to a certain restricted zone. In the majority of cases this ionized region may thus be regarded as damping or absorbing the waves, and in others it may conceivably intensify or concentrate them in one direction, producing instances of transmission over extreme distances with low power. For more complete discussion along these lines see *Heaviside Layer*, also *Refraction and Reflection of Electromagnetic Waves.*

GOLD LEAF ELECTROSCOPE—A very sensitive device for determining the presence and nature of an electric charge. It comprises, essentially, two strips of gold foil joined at one end by a conductor and suspended in a glass jar. When an electric charge is present the two strips of foil have a tendency to repel one another and thus, by diverging, indicate the presence of the charge. (See *Galvanometer.*)

GOLDSCHMIDT ALTERNATOR—One of the first practical *alternators* for generating *high frequency currents* suitable for *continuous wave* radio transmission. It was designed by *Rudolph Goldschmidt* in 1912 and was used mainly in Europe. This machine operated on a frequency step-up principle. A single phase alternating current was first produced, the frequency being about ten thousand cycles. This was about the highest possible frequency obtainable with an ordinary type of alternator. By means of a *cascade* effect, obtained by an arrangement of tuned circuits connected to stator and rotor of the alternator, a much higher frequency could be impressed on the antenna to radiate continuous waves. The machine had a number of disadvantages which limited its commercial value, chiefly the fact that the air gap between rotor and stator was necessarily very small—a difficult matter with a bulky machine, the rotor of which weighed several tons. (See *Alternator, Continuous Waves*, also *Goldschmidt, Rudolf.*)

GOLDSCHMIDT, RUDOLF—Born March 19, 1876, at Neu-Buckow, Mecklenburg, Germany. After finishing his education at Wiemar Municipal School, he studied engineering at Charlottenburg and Darmstadt Technical High School. In Darmstadt he obtained his degree as electrical engineer in January, 1898, and then became assistant to Professor Kittler. In 1900 he obtained the college and travelling scholarship, which enabled him to visit engineering works in Belgium, England and France. Later in the same year he was appointed engineer in the laboratory of the Allgemeine Elektrizitaets Gesellschaft in Berlin. In 1901-2 he occupied the position of chief laboratory engineer and designer to Kolben & Co., Ltd. in Prague. He came to England in connection with the Willesden Electricity Supply Station and was later appointed chief engineer to Messrs. Crompton & Co., of Chelmsford. In 1905 he joined the Westinghouse Company in Manchester. After private preparation he passed the German abitur-examination and obtained the degree of Dr. Eng. In 1907 he returned to Germany as lecturer at Darmstadt Technical College. Here he practiced as a consulting engineer, and also pursued the development of several inventions, chiefly occupying himself with the invention and design of high-frequency alternators for wireless telegraphy. In 1911 he became manager of the "Hochfrequenz-Machinen Aktiengesellschaft fur drahtlose Telegraphie" in Berlin, a company formed for the utilization of his inventions in wireless telegraphy. In this position he established two large wireless stations at Eilvesen, Province of Hanover, and Tuckerton, New Jersey, for wireless communication between Germany and America.

GOLDSMITH, ALFRED NORTON—Elec. and Radio Engr.; b. New York, N. Y., Sept. 15, 1887; B.S., Coll. of the City of New York, 1907; Ph.D., Columbia Univ., 1911; cons. radio expert, U. S. Dept. of Justice, 1912; cons. radio engr., Atlantic Communication Co., 1914; cons. engr., Gen. Elec. Co., 1915-17; dir. of research, Marconi Wireless Telegraph Co. of America, 1917-19; associated professor of electrical engineering, College of City of New York, since 1924; dir. of research dept., Radio Corp'n of America, 1919-1922; Chief Broadcast Engineer, Radio Corp'n of America, since 1922; editor, Proceedings of the Institute of Radio Engrs., since 1912. Made investigations in simplex and duplex radio telegraphy and telephony, transmission of canal rays, precision measurements in radio eng. Author: "Radio Telephony" (Wireless Press), 1918; "Radio Measurements," "Radio Frequency Changers" (Proc. Inst. Radio Engrs.), 1915; "World Communication" (Jour. A. I. E. E.), 1921. Technical director, U. S. Signal Corps School of Communication, 1917-18; U. S. Naval Radio School, 1917-18. Fellow, A. I. E. E., I. R. E., American Association for Advancement of Science. Hon. mem., Radio Club of America, Radio Society of Great Britain, Am. Physical Soc.

Dr. Alfred N. Goldsmith.

GONIOMETER—A device used in radio reception to determine the direction from which the electromagnetic waves (signals) are coming. The term is also used at times to designate any system of directional transmission or reception. The device was originally designed as an aid to navigation.

The original goniometer was designed by Messrs. Bellini and Tosi (see *Bellini, Dr. Ettore*, also *Tosi, A.*), but numerous changes have since been made, notably by Marconi and his associates, to adapt it to the requirements of ships at sea. The form most used is based on the original Bellini-Tosi machine and is essentially a special form of *oscillation transformer* used in conjunction with a special aerial or loop. As an example of the use of a goniometer in connection with ships at sea, suppose a ship is in distress at a point about two hundred miles off the eastern coast. Perhaps the ship's radio apparatus is working, but for some reason the mariner does not know the ship's position. The radio signals are then received, let us say, by three different land stations at widely separated points. The receiving stations use their goniometers and determine more or less accurately the direction from which the signals are coming. The three stations compare notes on the respective directions and any one of them can readily plot out lines running on a chart from each station in the directions indicated by the goniometers. Then by simple triangulation the location of the ship can be figured, enabling other ships in the vicinity to be instructed accordingly and go to their rescue. There are, of course, many other uses that occur in everyday practice. Goniometers are also much used by radio inspectors to locate stations that are interfering, using an unlicensed transmitter, or are otherwise not complying with the regulations.

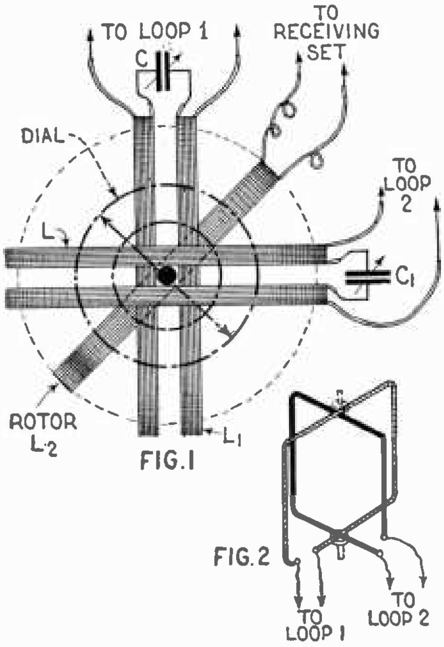

Fig. 1 shows the general arrangement of a goniometer and Fig. 2 shows the loops to which coils L and L-1 are connected.

The illustration Fig. 1 shows the general arrangement of the instrument. Here, coils L and L1 are the main windings of the goniometer. They are wound in two sections with condensers C and C1 inserted between the sections of the two windings for purposes of tuning. Coil L2 is the *exploring coil.* This is connected to the receiving apparatus, which is the customary detector circuit. Coils L and L1 are connected to the two loop aerials shown in Fig. 2. Here it will appear that the loops are placed at right angles, their relation in this respect being variable. In operation the coils L and L1 have minimum *mutual induction,* as they are placed at right angles. The exploring coil L2 can be rotated through a complete circle and the pointer indicates the angle with relation to the fixed coils. When one of the two loops points in the direction of a transmitting station it will pick up the signals, the other being at that moment inoperative, as it is at right angles to the first, as previously explained. If the exploring coil L2 is now turned until maximum signal strength is obtained, the pointer will be in the direction of the plane of the loop. If the advancing wave strikes the two loops simultaneously at any angle, a resulting magnetic field will be set up in the coils L and L1 of the goniometer, the field having a direction at right angles to the direction from which the waves are advancing. Now when the exploring coil L2 is turned to allow maximum signal strength due to this field (permitting maximum inductive transfer), the pointer will be in the direction of the transmitting station. The direction indicated by the pointer is, of course, not the geographical direction, but this can readily be determined by means of a compass. It also appears from the foregoing description that the pointer will indicate two directions, which, however, will be on the same geographic plane. In other words, the pointer may be, for example, exactly north and south.

GRAMME ARMATURE—A form of ring type *armature* (q.v.) used in generators, named for its originator. (See *Generator.*)

GRANULAR CARBON—The carbon grains used in a telephone transmitter to vary the resistance between the electrodes. The highest grade is made by carbonizing pure anthracite coal in an electric furnace, after which it is screened and graded. (See *Microphone, Carbon.*)

GRANULAR COHERER—A form of *coherer* (q.v.) used in the early days of wireless. It employed carbon granules between two electrodes, the whole enclosed in a glass tube. (See *Detector.*)

GRAPH—The presentation in diagrammatic form of facts, formulae, etc., particularly in the electrical and mechanical fields. Graphs are much used in radio to give graphic illustrations of various functions, such as the relation of currents and voltages. For this purpose ruled and squared paper is used. One of the most common and useful forms of graph is known as a characteristic curve of a vacuum tube. These curves may show the relation of changing values, such, for instance, as the relation of grid voltage to plate current, indicating certain of the operating characteristics of the vacuum tube. (See *Characteristic Curve,* also *Sine Curve.*)

GRAPHITE RESISTANCE—A form of resistance composed of graphite, used in any of several ways in radio circuits. The most common usage is in the form of a *grid leak* (q.v.). Here it may take the form of a thin sheet of paper coated with graphite or a glass tube filled with graphite and some binding medium. Such resistances are also used in *resistance coupled amplifiers* (q.v.), in which case they are interposed between the plate and "B" battery and couple the successive stages of amplification. (See *Resistance,* also *Resistance Coupled Amplifier.*)

GRASSOT FLUXMETER—An instrument used to measure the strength of magnetic fields and the regions around magnetic poles. It is of particular value in measuring the relative strengths of various magnets. The instrument comprises a very sensitive galvanometer of the moving coil type with a special exploring coil. This coil is introduced into the magnetic field to be measured and readings of the proportionate strength of the field are obtained directly from the scale of the galvanometer, which is appropriately divided. The field strength is obtained by dividing the deflection value in *Maxwell turns* (q.v.) by the area of the search or exploring coil in square centimeters and the number of turns of the coil. (See *Galvanometer,* also *Flux Density.*)

GRAVITY BATTERY—A combination of gravity cells. A group of cells, each cell having two different electrolytes which remain separated without use of a diaphragm due to their difference in specific gravity. (See *Gravity Cell.*)

GRAVITY CELL—A type of *primary cell* (q.v.) in which the specific gravity of one electrolyte is less than the other, which causes this lighter electrolyte to float on top of the heavier fluid. One of the typical examples of gravity cells is the *Daniell's Cell* (q.v.). (See *Secondary Cell.*)

GRID—Generally, any system of wires placed parallel to each other and in the same plane, as a mesh. Specifically in radio usage, the fine network of wires interposed between the filament and plate elements in a vacuum tube. It may be in the form of a flat network of wires or in the form of an oval cylindrical mesh completely surrounding the filament and in turn surrounded by a similarly shaped plate of metal. The illustration shows the form taken by a typical grid member in a modern type of vacuum tube. Here the plate is not shown in order to permit the grid member to show more clearly.

In receiving circuits the grid acts as a valve to control the flow of electrons from hot filament to cold plate. When signals are impressed on the aerial they are tuned by the customary agencies and passed to the *grid circuit* (q.v.). As the incoming signals are in the form of *high frequency oscillations* (q.v.) changing their polarity (direction) many thousands of times each second, it will be apparent that the grid is altered rapidly from positive to negative and back. As the electrons flowing from the filament are negative, the grid will thus have a tendency to retard or increase the flow. That is to say, at the instant the grid is positive it will attract more electrons, but when it is momentarily negative the like charges will repel each other and less electrons will flow, some of them being held back to the filament. (See *Grid Battery, Grid Bias,* also *Vacuum Tube, Theory of Operation of.*)

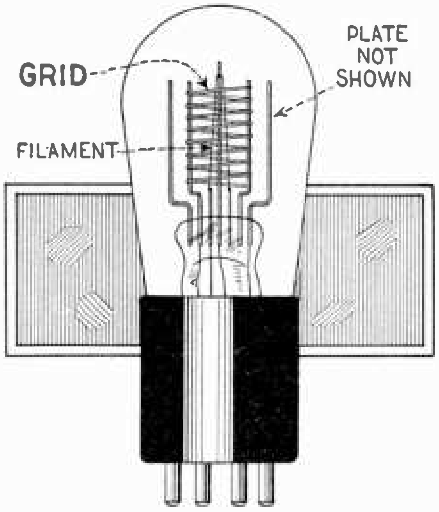

Showing the form taken by a typical grid in a modern vacuum tube.

GRID BATTERY—A battery of small cells connected in the grid circuit of a vacuum tube to place a negative or positive bias or charge on that element. The illustration Fig. 1 shows a small four and one-half volt battery inserted in such a manner that a negative potential is impressed on the grid. In this case the *characteristic curve* (q.v.) of the tube is assumed to indicate that more difference in current occurs between points B and C than between A and B in Fig. 2. For this reason a negative potential is applied as shown in Fig. I. (See *Grid Bias.*)

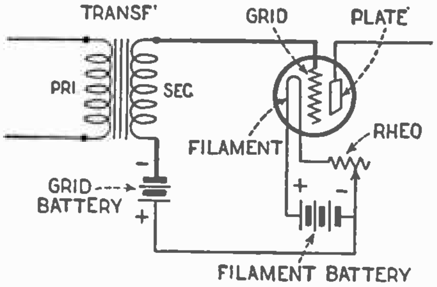

Fig. 1. Method of a 4½ volt battery to impress a negative potential on the grid of a vacuum tube.

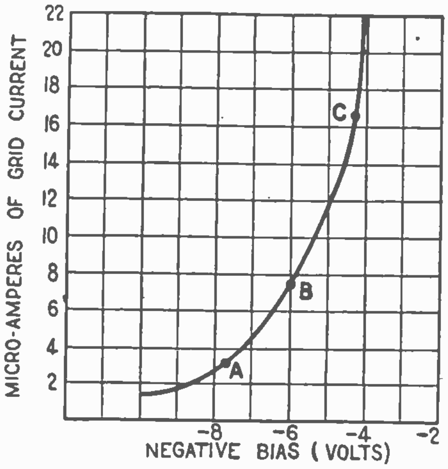

Fig. 2. Characteristic curve of a vacuum tube indicating more difference in current occurs between "B" and "C," than between "A" and "B."

GRID BIAS—A potential of a few volts, generally from four to six, applied to the grid of a vacuum tube to influence its operation by making it more or less negative. The grid bias is usually negative and determines the point of the *characteristic curve* at which the tube will operate. In a sensitive receiver, and particularly where a tube is used as an *amplifier* (q.v.), it is essential to obtain as great a change of grid current as possible. (Note: The greater the change of grid current, the greater the change in plate current and hence the more powerful will be the output.) By applying a negative potential on the grid it is possible to hold it at the point of maximum response. (See *Grid Battery,* also *Vacuum Tube, Theory of Operation of.*)

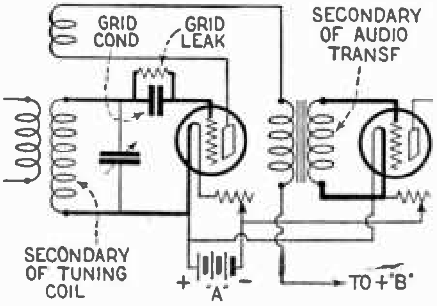

Diagram in which heavy lines at left show the grid circuit of the detector, while those on right indicate audio amplifier grid circuit.

GRID CIRCUIT—In a receiver employing vacuum tubes, the part of the circuit containing the grid of the tube or tubes, but generally referring to the tuning circuit or that part of the system which contains the tuning elements. In the illustration is shown a standard arrangement of a regenerative detector and one stage of audio-frequency amplification. The heavy lines at the left of the diagram represent the tuning circuit or detector *grid circuit,* while the section with heavy lines at the right is the grid circuit of the amplifier tube. The grid circuit of a detector tube may generally be distinguished by means of the grid leak and condenser (so marked in the illustration), which controls the incoming signals and permits the tube to function as a rectifier. (See *Detector, Grid Control,* also *Grid Condenser* and *Grid Leak.*)

GRID CONDENSER—A small fixed condenser, generally from .00025 to .0005 microfarad capacity, inserted in the circuit of a detector tube between the tuning coil and the grid member of the tube. The illustration shows the manner of connecting the grid condenser in a conventional detector circuit. This condenser insulates the grid from the filament by breaking the path F to G and permits the tube to act as a rectifier or detector. A resistance (called *grid leak*) is usually placed across this condenser to allow the accumulated charges on the grid to leak off. (See *Grid Leak,* also *Grid Control* and *Vacuum Tube, Theory of Operation of.*)

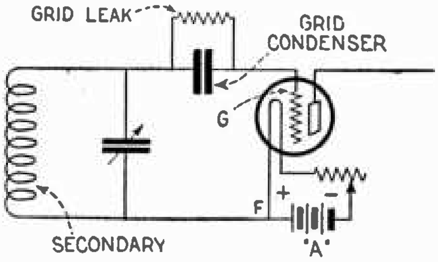

Showing how grid condenser is connected in a detector circuit.

GRID CONTROL—The general term used to designate the various devices and associated connections placed in or allied with the grid circuit of a vacuum tube for the purpose of controlling to a greater or lesser extent the potential of the grid, or to tune or control the incoming oscillations. In the case of a vacuum tube used as a rectifier (detector) the grid condenser and grid leak are placed in series with it. (Note: the grid leak is often connected from the grid directly to the filament instead of across the grid condenser, the operation, however, being approximately the same in either case.) Here the controls permit the tube to function as a detector of the incoming oscillations. Where the tube is used as a radio-frequency (high-frequency) amplifier, the radio-frequency transformer or other amplifying unit controls the grid of the following tube by influencing the fluctuating potential differences impressed on it. (See *Amplifier, Radio-Frequency.*) The most important example of a grid control is by using some means to control the potential applied to the grid, such, for example, as a grid bias or grid battery. (See *Grid Bias, Grid Battery,* also *Vacuum Tube, Theory of Operation of.*)

GRID CURRENT—The current present in the grid circuit of a vacuum tube. This is generally very small, perhaps of the order of fifteen to twenty *micro-amperes* (q.v.). (See *Plate Current,* also *Grid Bias.*)

GRID LEAK—A high resistance placed in the grid circuit of a vacuum tube to permit the electrons forming the grid current to leak off after each charge, thus preventing their accumulating on the grid in such numbers as to stop the flow from the filament. The electrons being negative, a surplus of them on the grid member of the tube would act as a barrier to the free flow of other electrons from the filament through the grid to the plate, thus diminishing the plate current toward zero. It will be understood that a heavy negative charge on the grid will repel a like or negative charge, in this case the negative electrons being thrown off by the hot filament.

The grid leak may have a value ranging from about 250,000 ohms to several million ohms. The value of a grid leak is generally stated in *megohms*—one megohm being one million ohms. Thus a grid leak of 250,000 ohms resistance would be referred to as a ¼-megohm leak and, similarly, one having a resistance of two million ohms would be referred to as a two-megohm grid leak.

Such resistances are furnished in many forms, the most common being a strip of paper impregnated with graphite or some similar high resistance preparation, placed in a glass tube and sealed to prevent moisture from affecting the value. The illustration, Fig. 1, shows a tubular type of grid leak.

FRESHMAN
NOISELESS
TESTED GRID LEAK

Courtesy of Chas. Freshman Co.

Fig. 1.—A tubular type of grid leak.

For experimental work a fairly satisfactory grid leak can be made by the simple expedient of drawing a line on a piece of cardboard with a soft pencil—the graphite mark acting as the resistance. At each end of the cardboard is placed a binding post in contact with the graphite, each binding post being connected to opposite sides of the grid condensers. Variable grid leaks, the resistance of which can be changed at will, are also furnished in an infinite variety of types, two of which are shown in Fig. 2. (See *Variable Grid Leaks,* also *Grid Control* and *Resistance.*)

CHAS. FRESHMAN CO.
NEW YORK PAT'S PEND.
CHAS. FRESHMAN CO.

Courtesy of Chas. Freshman Co.

Fig. 2.—Two types of variable grid leaks with condenser incorporated.

GRID POTENTIAL, MODULATION OF—In the transmission of *undamped waves* (q.v.) generated by a vacuum tube, the process of varying or modulating the potential of the grid element of the tube with respect to the filament. In this method an alternating potential may be applied to the grid circuit, the frequency of this varying potential being within the audio band —possibly about 500 cycles—supplied by a generator of that frequency. (See *Modulation.*)

GRID POTENTIOMETER—A potentiometer—a variable resistance unit—Fig. 1, used in the grid circuit of a vacuum tube for the purpose of controlling the potential applied to the grid. The conventional method is to place the potentiometer across the "A" battery with the center or variable contact arm connected to the grid of the tube in place of the usual connection to either positive or negative filament lead. This is shown in the illustration, Fig. 2, Here the customary circuit is shown for the "A" battery to light the filament of the tube. Pri. and Sec. are respectively primary and secondary of the transformer (either radio-frequency or audio-frequency) and C the center arm of the potentiometer. The center arm may be moved either to the right or left, permitting either negative or positive potential to be applied to the grid. This system is much used in amplifying tube circuits to control the action of the tube as an amplifier. Another method would be to connect one of the ends of the potentiometer to the negative lead of the "A" battery and the center arm to the grid leak. In this way the potentiometer would be acting as a pure variable resistance, the potential applied to the grid being always negative, but susceptible of variation within the limits of the ressitance and the voltage of the "A" battery. (See *Grid Bias, Grid Battery,* also *Grid Control.*)

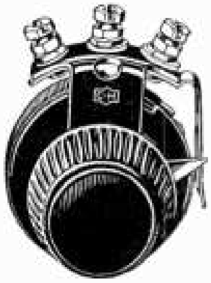

Courtesy of the Cutler-Hammer Co.

Fig. 1.—A type of potentiometer.

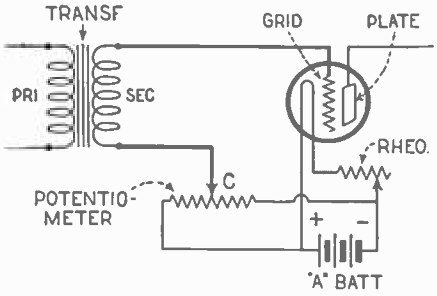

Fig. 2.—Method of connecting potentiometer in grid circuit.

GRINDER—Term occasionally applied to one of the various forms of atmospheric disturbance known under the general heading of *static.* This form of static is more prevalent in warm weather and does not appear to have its origin in local electric storms. Its effect is to produce rattling sounds of no definite pitch. (See *Strays,* also *Static.*)

GROUND—The earth considered as an electric conductor. In radio it is considered as the return circuit for electromagnetic waves. (See *Ground Connection,* also *Theory of Propagation of Electromagnetic Waves.*)

GROUND CLAMP—A strip of metal, generally copper, having some sort of arrangement for fastening rigidly to a water or steam pipe to form the ground connection for a radio receiver. The illustration shows a conventional form of clamp. (See *Ground Connection.*)

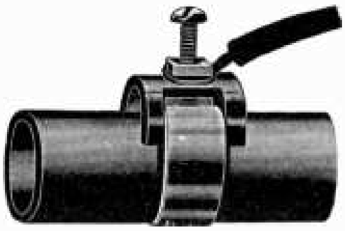

A ground clamp for fastening the ground lead to a water or steam pipe.

GROUND CONNECTION—In radio reception or transmission, the connection between the earth and the apparatus whereby the currents set up by the electromagnetic waves are returned to the earth as one side of the assumed circuit. The ground or earth connection from a radio receiver or transmitter should be as short as possible and preferably a heavy wire as shown in the illustration. Such a connection for receiving purposes may be one or more heavy metal plates sunk in the ground and connected by a heavy wire to the apparatus. The more common method is to connect the set to a convenient water or steam pipe, such pipes being almost invariably well grounded, especially the former. (See *Earth,* also *Ground* and *Transmission.*)

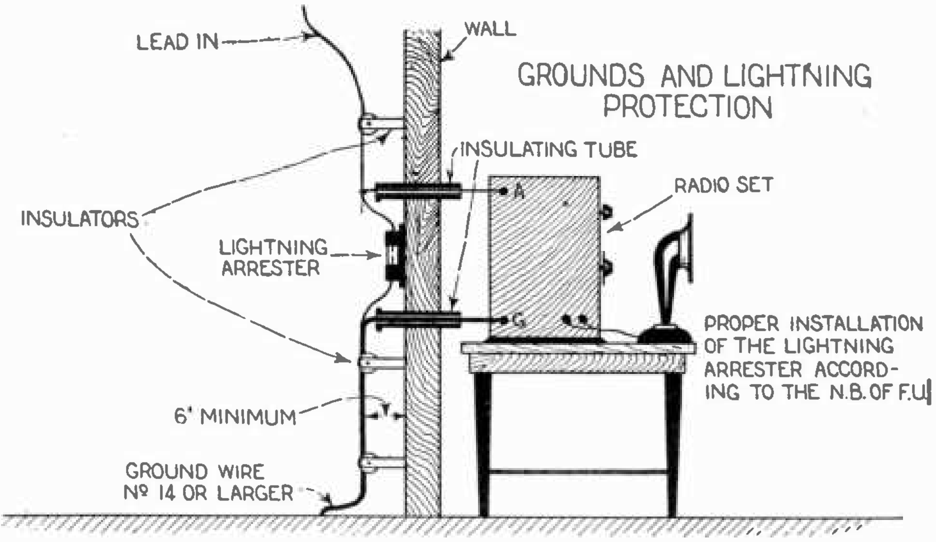
LEAD IN
WALL
GROUNDS AND LIGHTNING PROTECTION
INSULATING TUBE
RADIO SET
A
INSULATORS
LIGHTNING ARRESTER
G
PROPER INSTALLATION OF THE LIGHTNING ARRESTER ACCORDING TO THE N.B.OF F.U
6' MINIMUM
GROUND WIRE Nº 14 OR LARGER

Illustration showing by means of a heavy line, the ground connection of a radio receiver.

GROUND CURRENTS—Electric currents present in the ground due to the grounding of commercial electric machines. These currents very often prove a disturbing factor in radio reception. (See *Ground.*)

GROUND DETECTOR—A device for indicating an accidental ground on a power transmission line or for any conducting systems for electric currents. (See *Detector,* also *Ground.*)

GROUND RESISTANCE—See *Resistance of Ground Connection.*

GROUND SWITCH—A switch, generally of rugged construction, connected to the aerial, ground and apparatus in such manner that at one position the aerial is connected to the set, which in turn is connected to the ground, and in the other position the aerial being connected directly to the ground to protect the set from excessive static when not in use. (See *Lightning Switch,* also *Aerial Switch, Down Lead.*)

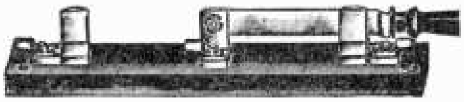

Switch having a break distance of not less than 4 inches.

GROUND WIRE—The wire connecting a receiving or transmitting set to the earth. The underwriters specify not less than No. 14 B & S gauge copper wire. (See *Ground Connection.*)

GROUNDED ROTOR—See *Rotor, Grounded Connection.*

GROUP FREQUENCY—The number of separate trains or groups of waves per second in a damped or undamped system of radio transmission. This frequency is to be distinguished from the frequency of the individual waves. Thus in a spark transmitter, for example, the condenser in the transmitting circuit will be momentarily charged and discharged. The discharges will occur in the form of trains of oscillations, each oscillation having a certain definite frequency, but the trains or groups occurring at definite intervals, the frequency of these intervals being called the group frequency. (See *Frequency,* also *Spark Transmission* and *Undamped Waves.*)

GUARD LAMP—An incandescent lamp having a straight filament, sometimes shunted across the armature and field leads of a rotary converter used in radio transmission. Its purpose is to protect the windings from injury due to the inducing of oscillatory surges from the high-frequency circuits. (See *Rotary Converter.*)

GUTTA PERCHA—An insulating material derived from the milky secretion of certain trees, used as an insulation for submarine cables on account of its ability to resist the action of salt water. It is used to a lesser extent as a general insulating medium for wires, but owing to its scarcity and limited field of usefulness, rubber or rubber compounds are more widely used. (See *Insulating Materials.*)

GUY WIRE—Wires used to brace aerial masts for receiving or transmitting stations. Such guys are generally heavy galvanized wires attached to the mast at various points and extending radially downward and anchored to surrounding rigid objects or by means of some form of anchor in the ground. (See *Aerial.*)

H

H—The chemical symbol for *hydrogen.* In electrical and radio usage it is the symbol for *magnetizing force* or *magnetic field intensity*, adopted as such by the International Electrotechnical Commission and by the American Institute of Electrical Engineers. It has also been adopted by the I. E. C. as the symbol for *Henry* (q.v.). (See *Magnetic Field.*)

HALF WAVE RECTIFICATION—The process of changing alternating currents to pulsating direct currents wherein only one-half of the full alternating current wave or cycle is rectified, the other portion being blocked by the action of the rectifier. In the illustration is shown a typical half wave rectification process. The upper graph shows the sine wave representing an alternating current, such as supplied from an ordinary alternating current house lighting system. Here A-B represents one complete cycle or two alternations, one positive and the other negative. In the lower graph is shown the form taken by the current after passing through a half wave rectifier. It now appears as a pulsating or intermittent direct current, only the upper or positive alternations having passed through the rectifier. The shaded portions below the line represent the negative or unused alternations. (See *Full Wave Rectification, Sine Wave of Alternating Current or E M F*, also *Charger, Storage Battery.*)

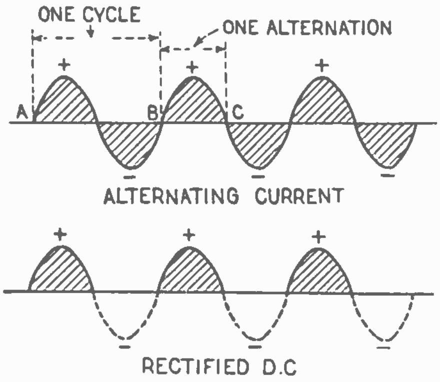

The upper graph represents an alternating current, and the lower graph shows the form taken by the current after passing through a half wave rectifier.

HALF WAVE RECTIFIER—Any system for rectifying or changing alternating currents to pulsating direct currents, wherein only half of the full alternating current wave is used. Essentially such a device is a uni-directional conductor, permitting current to pass in one direction only and submerging or blotting out the surges in the opposite direction. A typical example of a half wave rectifier is the crystal detector used in simple receiving circuits. Here the nature of the crystal is such that the currents are permitted to pass only in one direction. In half wave rectifiers used for charging storage batteries the effect is much the same. Where tubes such as the Tungar or Rectigon are employed, one tube is more commonly used, allowing the current to flow in one direction but submerging half of each cycle. (See *Rectifier, Full Wave*, also *Rectifier and Charger, Storage Battery.*)

HAMMER BREAK—A magnetically operated device for making and breaking contact in an electrical circuit. Hammer breaks are essentially the same as a vibrator or buzzer break but are made to withstand heavy currents without undue arcing, and hence are much slower in action. (See *Buzzer*, also *Spark Coil.*)

HAMMOND, JOHN HAYS, JR.—Inventor. Born San Francisco, Calif., April 13, 1888. B.S., Sheffield Scientific School (Yale), 1910; Sc.D., George Washington University. Inventor of type of torpedo for coast defense, controlled by wireless energy from coast fortifications, which was recommended to Congress for exclusive purchase by U. S. by Board of Ordnance and Fortifications, U. S. A.; invented system of automobile torpedo firing; aluminothermic incendiary projectiles employed by Allied armies in World War; radio system of control of ships, employed on U. S. S. Iowa for target practice; a system of coastal patrol by aeroplane; a system of selective radio telegraphy, adopted by the U. S. Navy, U. S. Signal Corps, and U. S Army; system of aerial coast surveying adopted by the Bartlett Expedition for Polar exploration. President Radio Engineering Company of New York; director Hammond Radio Research Labs.; director Radio Corp. of America; consulting engineer to Radio Corp. of America and General Electric Co. Has licensed the American Tel. & Tel. Co., and the Radio Corp. of America under his electrical patents in U. S. and S. A. for commercial purposes only, U. S. Gov't retaining option on military uses. Has applied for over 224 patents in U. S. and Europe relating to Radio Telegraphy and Telephony and Wirelessly controlled torpedoes and various improvements in pipe organ mechanisms. Formerly member Advisory Board, U. S. Naval Board of Inventors, member Advisory Com. Langley Aerodynamic Lab. of Smithsonian Inst., co-operating with Third Naval Dist.; member Conf. Com. on National Preparedness (Sub.-Com. on Finance); Gov. Aero Club America (mem. Aerodynamic, Tech., Pub. Safety and Map and Landing Places Coms.); v.-p. Am. Soc. Aeronautic Engineers; mem. Inst. Radio Engineers (ex-treas., etc.), Am. Inst. E. E., Royal Soc. Arts (London); asso. mem. Am. Soc. M. E.; hon. mem. Nat. Inst. of Inventors, Harvard Aeronaut. Soc.; Fellow Am. Geog. Soc.; U. S. del. Radio-Telegraphic Com., London, 1912.

John Hays Hammond, Jr.

HAND CAPACITY—See *Body Capacity.*

HARD DRAWN WIRE—In radio usage generally confined to hard drawn copper wire, i. e., copper wire that has been hardened by repeated drawing without being annealed. It is used chiefly for aerials. (See *Aerial.*)

HARD TUBE—A vacuum tube is said to be hard when it has been evacuated of all gases to a high degree. The opposite is a soft tube, one in which the vacuum is not of a very high order and more or less gas particles are present. Hard tubes are used for amplifiers, both audio and radio frequency, while the soft tubes are used mainly as detectors or oscillators. Hard tubes may be used with high plate voltages, whereas soft tubes require comparatively low plate voltages, the tubes having a tendency to turn blue (see *Blue Glow*) under the influence of high plate potentials. (See *Detector*, also *Amplifier.*)

HARMONIC CURRENT—An alternating current which can be represented by a simple harmonic or sine curve. It may be considered as an alternating current, the waves of which are sinusoidal or uniform. (See *Simple Harmonic*, also *Sine Curve.*)

HARMONIC MOTION—A term signifying any periodic oscillatory motion, as, for example, the vibration or beating of a pendulum. The illustration shows the manner of indicating simple harmonic motion. The distance A-B represents 360 degrees of a circle, or, if it is assumed to represent a sine wave of alternating current, it will represent one complete cycle. Curves representing harmonic motion are used extensively in radio engineering to illustrate, in graphic form, the oscillations of an alternating current. (See *Sine Wave*, also *Simple Harmonic Vibration* and *Oscillations.*)

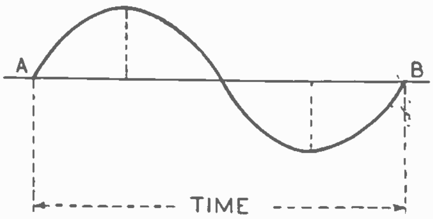

Illustration showing method of indicating simple harmonic motion.

HARMONICS—Oscillations to which a circuit will respond in addition to the basic or fundamental oscillations. The effect is similar to that in sound, where, for instance, the primary or fundamental tone struck by an instrument may be accompanied by notes or tones of a higher pitch. In sound these harmonics are principally the third, fifth, seventh and the octave. In radio, a circuit may respond to oscillations of a frequency either higher or lower than the fundamental frequency, as, for example, one third, one fifth or one seventh, or, again, double or treble the original frequency. As a concrete example, with a fundamental wave-length of 240 meters, the harmonics might be

40 meters or 80 meters. The corresponding frequencies would be approximately 1,250,000 cycles, 7,500,000 cycles and 3,750,000 cycles. (See *Harmonic Suppressor*, *Side Bands*, also *Harmonic Motion*, *Oscillations*, *Wave-length*.)

HARMONIC SUPPRESSOR—A device used in radio transmission and in laboratory work, for the purpose of eliminating undesired frequencies (wave-lengths). Essentially such a device is a *filter* (q.v.) so arranged as to suppress or eliminate certain frequencies or *harmonics* (q.v.). In radio broadcasting it is particularly essential that all undesired frequencies or *harmonics* be suppressed. Basically the system involves a special coil shunted by a condenser and tuned to the frequency of the interfering wave, the main tuning circuit at the same time being adjusted to the desired frequency. (See *Broadcasting*, *General Treatise on Methods*, also *Transmission* and *Radio Telephony*.)

HAZELTINE NEUTRODYNE RECEIVER—The radio receiver designed by Professor Hazeltine, employing the Neutrodyne circuit. Essentially these are five-tube receivers employing two stages of tuned and neutralized radio-frequency amplification with the conventional detector and two audio-frequency stages. In addition to the five-tube Neutrodyne receiver there are numerous types involving the Neutrodyne principle in conjunction with *reflex systems*. (See *Neutrodyne*.)

HEAD TELEPHONE OR HEADPHONES—Telephone receivers arranged with a band to fit over the operator's head, thus leaving him free use of both hands. Head-phones are used in commercial radio work, the operator being free to handle a key or the controls of the receiver or transmitter. They are also used in broadcast reception where the strength of signals is not sufficient to permit the use of a loud speaker, and also in some cases to tune in distant signals before attaching a loud speaker. (See *Watch Case Receivers*.)

Illustration by courtesy of C. Brandes, Inc.
A type of headphone used extensively in radio reception.

HEAVISIDE, DR. OLIVER. — Born in London on May 13, 1850, Dr. Heaviside was engaged in telegraph work for a few years, but after 1874 he lived in retirement studying Clerk-Maxwell's theory and applying it to telegraph and wireless problems. Elected a Fellow of the Royal Society in 1891, he was also Faraday Medallist of the Institution of Electrical Engineers and Hon. Ph.D. of Gottingen University. The name of Dr. Heaviside is primarily associated with the theory, propounded by himself, of the existence of a permanently ionized layer in the upper atmosphere capable of deflecting electro-magnetic waves and thus permitting wireless communication round the earth.

HEAVISIDE LAYER—An assumed layer or stratum of the atmosphere supposed to exist from fifty to one hundred miles above the earth's surface according to the Heaviside Theory. Heaviside's theory supposes this stratum to consist of heavily ionized gas which is conductive and acts as a gigantic reflector of electromagnetic waves (radio waves). It is this layer which is supposed to reflect the electric waves and force them to follow the curvature of the earth. This theory partially explains—if true—why radio signals are transmitted over great distances instead of being radiated off into space and lost. It has been accorded the support of many eminent scientists. Elihu Thompson has propounded an alternative theory which assumes the waves as attached to the earth and gliding along over the surface. (See *Gliding Theory*.)

HEAVISIDE LAYER THEORY—The theory of an upper ionized stratum from fifty to one hundred miles above the earth's surface, which is assumed to act as a reflector of radio waves and thus force them to follow the curvature of the earth. None of the theories thus far advanced has been proven, but it has been definitely established that these waves do follow the curvature of the earth. (See *Heaviside Layer*, also *Gliding Theory*.)

HEDGEHOG TRANSFORMER—A type of audio transformer so called because of its peculiar construction. The core is cylindrical in shape and consists of a great number of soft iron wires around which is a bobbin holding the primary and secondary windings. After the windings are completed the wires forming the core are bent back over the form and held in place, completely surrounding the windings. (See *Audio-Frequency*, also *Transformer*.)

HEIGHT OF AERIAL—The elevation of an aerial system as used in radio transmission or reception. The effective height is roughly considered as the height above the earth or surrounding objects. The range of a station having constant power is dependent upon the height of the aerial, other factors of radiation being unchanged. In other words, increase of height increases the range within certain limits. (See *Radiation from Antenna*, also *Transmission*.)

HEISING, RAYMOND A.—Electrical Engineer, born Albert Lea, Minn., August 10, 1888. Educated University of N. D., Elec. Engineer, 1912; University of Wisc., M.S., 1914 Grad. Work in Physics, Research and Development of Radio Tel. since 1914; Trans-Atlantic test in 1915; for Army and Navy use during War, aeroplane, chaser, intership communication; since War continued research and development ship to shore use, trans-Atlantic test, 1923. Author of several technical papers; numerous inventions widely used in present broadcasting stations.

HEISING MODULATION—A system of modulating radio currents for radio transmission, such as broadcasting, developed by *R. A. Heising*. The fundamental principle is shown by the illustration. Here the customary antenna and ground are connected with a high-frequency alternator, A; the field of the alternator, FW, is connected to the grid return and filament of the vacuum tube circuit, and through the "B" battery BB to the plate P of the tube. The microphone circuit comprises a microphone M connected to the primary of a transformer T through the battery MB. The secondary of the transformer is connected to the grid G of the vacuum tube and the filament F. R is the usual rheostat to control the filament of the tube and AB is the filament lighting battery. In operation the speech or sound vibrations are impressed on the microphone M, thus varying the resistance of the current flowing from the battery MB and through the transformer T. This serves to vary the potential of the grid G, and as this potential changes, the current in the plate circuit will also be varied, the same characteristics being maintained. As the field winding FW is in the plate circuit, it is evident that the antenna current will be modulated in accordance with the variations of the speech or sounds introduced at M. (See *Grid Control*, also *Modulation* and *Broadcasting*, *General Treatise on Methods*.)

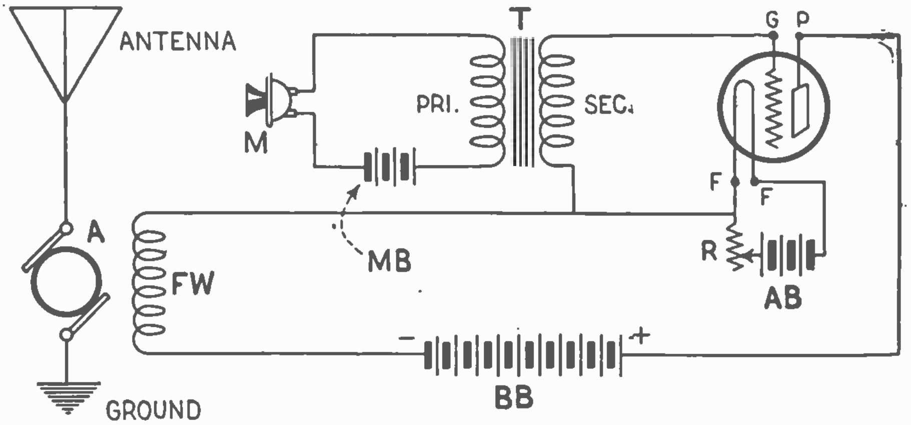

Circuit illustrating the fundamental principle of the Heising system of modulation.

HELIOGRAPH—An instrument involving a movable mirror, by means of which the sun's rays are reflected and made use of as a signalling device. In practice it is usually arranged with a blind or shade to cover the mirror at rapid intervals by pressing on a key device. In this manner the flashes may be made at definite intervals, as in the case of the Morse light. (See *Code*, also *Morse Light*.)

HELIX—The shape assumed by a line or wire wound in a single layer around a cylinder. Generally speaking, a *solenoid* (q.v.). A device in the form of a helix was formally used in transmitting circuits to vary the inductance, or, practically speaking, the wave-length. It usually was a number of turns of heavy copper or aluminum wire wound on an insulating form. Clips were provided to permit a variation in the contact point and hence any necessary change in the value of the inductance. (See *Oscillation Transformer.*)

HENRY, JOSEPH—Born in 1797, died in 1878. An American Physicist, noted for his researches in electromagnetism. He developed the electromagnet which had been invented by Sturgeon in England, so that it became an instrument of far greater power than before. In 1831 he employed a mile of fine copper wire with an electromagnet, causing the current to attract the armature and strike a bell, thereby establishing the principle employed in modern telegraph practice. He was made a professor at Princeton in 1832 and during his experimenting there he devised an arrangement of batteries and electromagnets embodying the principle of the telegraph relay which made possible long distance transmission. He was the first to observe magnetic self-induction, and performed important investigations in oscillating electric discharges (1842), and other electrical phenomena. In 1846 he was chosen secretary of the Smithsonian Inst. at Washington, an office which he held until his death. As chairman of the U. S. Lighthouse Board he made important tests in marine signals and lights.

Joseph Henry.

In meteorology, terrestrial magnetism and acoustics he carried on important researches. Henry enjoyed an international reputation and is acknowledged to be one of America's greatest scientists.

HENRY—The unit of inductance, named after the American physicist, Joseph Henry, who first produced high-frequency oscillations in 1832. A coil of wire is said to have an inductance value of one henry when a current passing through it changes at the rate of one ampere per second and an electromotive force of one volt is produced. As the henry is too large a unit for practical purposes, the thousandth part of a henry, known as the *milli-henry*, or the millionth part, known as the *micro-henry*, are more commonly used. Another much-used term is the centimeter, which is the thousandth part of a micro-henry. The abbreviation for henry is H, and for micro-henry μH. (See *Centimeter Gram Second, Centimeter*, also *Inductance.*)

HERTZ, HEINRICH RUDOLF—Born 1857, died 1894. A German physicist noted as the discoverer of electromagnetic waves (1888) which form the basis of wireless telegraphy. He found that waves produced by the spark of a simple device called the oscillator could be detected by a loop or square of wire known as the resonator, and he was able to show the reflection, refraction, diffraction and poralization of the waves. These remarkable discoveries demonstrated the practical possibilities of radio-telegraphy. He also made valuable experiments with electric phenomena in vacuum tubes.

Heinrich Rudolf Hertz.

HERTZIAN WAVES—Term occasionally applied to radio or electromagnetic waves, they being named after Dr. Heinrich Rudolf Hertz owing to his extensive researches into electromagnetic waves. (See *Hertz*, also *Clerk-Maxwell*, and *Ether Theory.*)

HETERODYNE—The term originally applied by Fessenden to a method of reception of electromagnetic waves, wherein the principle of *Beats* (q.v.) was employed. It is derived from the words "hetero," meaning other or different, and "dyne," meaning power.

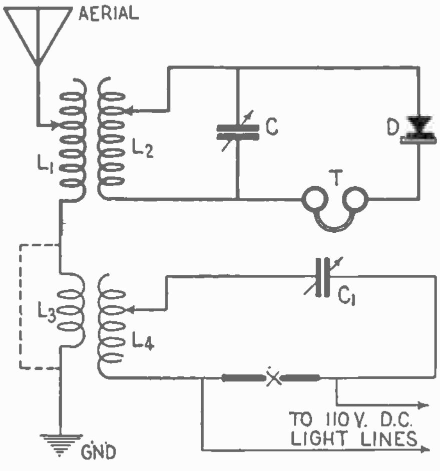

Fig. 1.—The original form of heterodyne receiver.

In its original conception by Fessenden, before the widespread use of vacuum tubes, it was used in conjunction with crystal receivers for the reception of *undamped waves* (q.v.). The heterodyne principle involves production of a local series of oscillations which are tuned to a frequency differing from the frequency of the incoming oscillations from the antenna. The difference between the two is known as the *beat frequency*. (See *Beats*, also *Heterodyne Receiver.*)

HETERODYNE RECEIVER—A receiver for electromagnetitc (radio) waves employing the *heterodyne* principle. The illustration Fig. 1 shows the original form of heterodyne receiver, wherein the local oscillations were produced by means of a direct current arc gap and associated circuits, and combined through a coupling medium, with the oscillations in the antenna circuit. Here the heavy lines represent a conventional crystal receiver circuit. L1 and L2 are respectively primary and secondary of a vario-coupler, C is a variable tuning condenser, D the crystal detector and T the head phones. If the dotted line is assumed as a connection to the ground, this will constitute a complete receiver. For the heterodyne purpose, however, a special coil L3 is employed in the ground circuit. Coupled to this is the secondary L4. C1 is another tuning condenser and A is an arc gap fed by 110 volts direct current. In operation the circuit represented by the light lines acts as a generator of undamped oscillations, these oscillations being supplied to the antenna circuit and the frequency varied by means of the condenser C1 and the coil L4. If the receiving tuner is tuned to a certain wave-length corresponding to incoming signals, and the generator system tuned to a different frequency, the difference between the two frequencies will be a *beat* frequency which will be present in the detector circuit. This new series of oscillations will be rectified by the detector D and will be audible in the head phones, T. The action of this system is shown by the series of curves given in Fig. 2. Here the incoming oscillations (signals) from the antenna are shown in graph 1 at the top of the illustration. These oscillations are of continuous amplitude or *undamped* (q.c.). Graph 2 shows the local oscillations. If we assume that the distance from A to B represents the time, it will be apparent that there are more oscillations per unit of time in the case of the incoming signals—

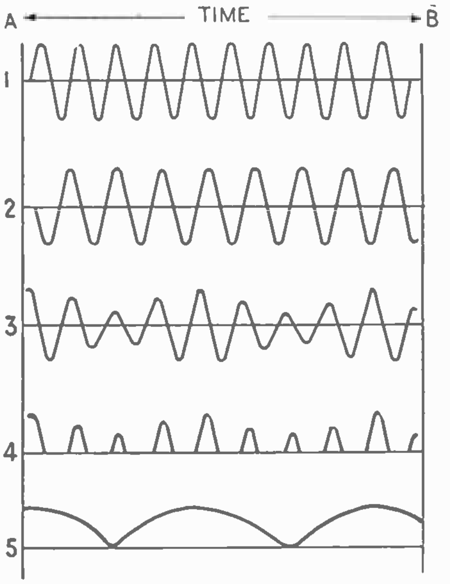

Series of graphs showing the action of a heterodyne receiver.

hence the second graph indicates that the local oscillations have a lower frequency. Graph 3 represents the *beat* resulting from interaction of the two upper series of oscillations and the frequency corresponds to the numerical difference. When the two upper groups of oscillations are opposed, the beat current is zero, and when they are in phase, or assisting each other, the beat current is maximum. When rectified by the crystal detector action the beats take the form shown in graph 4, the lower half of each cycle having been submerged. The telephone current of a periodic nature is indicated by the bottom graph, the original signals being represented in the phones by a series of contractions of the diaphragm. (See *Heterodyne*, also *Beats* and *Super Heterodyne.*)

HETERODYNE, SELF—See *Self-Heterodyne.*

HETERODYNE WAVE-METER—A wave-meter—device for measuring the wave-length of coils, circuits or incoming and outgoing signals—which utilized the heterodyne principle. Briefly the action is as follows: It has already been explained under Heterodyne Receiver that when the local oscillations and the incoming oscillations are opposed, the beat current is zero. It follows, then, that with a pair of phones connected in the circuit, the beat signals will be audible only when the two frequencies are different. Now if a suitable means is present for varying the frequency of the local oscillations, in practice a combination of coil and condenser, the local circuit can be tuned to a point where the two frequencies concur or are similar and no beat is audible. As the frequency is proportional to the capacity and inductance in a circuit, and as wave-length can easily be figured when the frequency of the local generating system is known, the wave-length of the incoming signals must be that of the local circuit when the beat is zero. The condenser dial on the heterodyne wavemeter can thus be calibrated directly in meters and a nearly true reading of the wave-length of incoming signals obtained. (See *Beats*, also *Heterodyne* and *Wave-meter.*)

HIGH FREQUENCY—A very general term applied to alternating electric currents. As a rule, alternating currents having frequencies above one thousand *cycles* (q.v.) per second are referred to as high frequency currents. When the frequency is extremely high, of the order of hundreds of thousands of cycles, the current is usually referred to as high frequency, oscillatory current. There is no definite level or line of distinction between high and low frequency currents. In radio practice, the audio frequency currents are generally considered below twenty thousand cycles per second and known as low frequency currents. The radio frequency or frequencies above twenty thousand are considered as the high frequency currents, although here again there is no definite level or line of demarcation between the two areas. (See *Audio Frequency*, also *Radio Frequency* and *Frequency.*)

HIGH FREQUENCY ALTERNATING CURRENT—See *High Frequency Current.*

HIGH FREQUENCY ALTERNATOR—An alternating current machine or alternator, designed to produce currents having high frequencies, generally above one thousand cycles per second. Such machines are used in high power radio telegraph transmission (not broadcasting) producing undamped or continuous waves. (See *Alternator*, also *Alexanderson Alternator.*)

HIGH FREQUENCY AMPLIFICATION—The process of stepping up or amplifying currents of high frequency. In radio circuits, the high frequency currents are known as radio frequency oscillations. In radio circuits the audio frequency (low frequency) currents and the high or radio frequency currents are usually in different parts of a circuit. When we consider alternating currents having a frequency above, say, twenty thousand cycles per second as radio frequency, and those below that limit as audio frequency, it is easy to consider the radio frequency currents as existing in all stages between the aerial and the detector. Beyond the detector, the alternating currents having become uni-directional surges, we can consider these currents as being audio frequency currents.

Thus we have a definite dividing line in any radio receiver using but one detector, the detector or rectifier marking the transition from radio frequency to audio frequency. Thus, any means of amplifying the signals before they reach the detector will be known as high frequency amplification. (See *Audio Frequency*, also *Amplifier.*)

HIGH FREQUENCY BUZZER—A *buzzer* (q.v.) designed to produce a very high note. The term "high frequency" in this case merely indicates that the pitch or frequency (vibration) is high in comparison with the ordinary buzzer, the sounds of course being actually of audible frequency. The general scheme of such a buzzer is about that of the ordinary type as used in conjunction with electric bells or signal systems, the armature, however, being very thin and thus permitting rapid make and break for the current. Such buzzers are used extensively in connection with crystal receivers for the purpose of testing and also for code practice. (See *Buzzer* also *High Frequency* and *Armature.*)

HIGH FREQUENCY COMPONENT OF PLATE—The high frequency currents present in the plate circuit of a vacuum tube receiver, as distinguished from the audio or low frequency currents. (See *Feed-Back*, also *Plate Component, High Frequency*, and *Regeneration.*)

H.F.C.—Abbreviation used abroad and occasionally in the United States for high frequency current. The more common term is radio frequency current and no abbreviation is used as a rule. (See *Radio Frequency Current*, also *High Frequency.*)

HIGH FREQUENCY CURRENT—Alternating currents having frequencies above certain limits, as, for example, the radio frequency currents employed in radio transmission, which are usually of a frequency above one hundred thousand cycles per second. (See *High Frequency*, also *Low Frequency* and *Radio Frequency.*)

HIGH FREQUENCY OSCILLATION—The term oscillation is generally applied to an alternating current of very high frequency. Thus radio frequency currents, which may run into hundreds of thousands of cycles per second, are called high frequency oscillations. (See *Oscillation.*)

HIGH FREQUENCY RESISTANCE—Resistance to the passage of alternating currents of high frequency. Said of wires, coils, circuits and apparati. The resistance of a certain conductor or part of an electrical circuit, as in radio, may be comparatively higher in the case of high frequency currents than to alternating currents of a low frequency. This is due to the "*skin effect*" (q.v.) it being a phenomenon of radio frequency (high frequency) currents that they have a tendency to concentrate near the surface of the conductor. For this reason, small wires are more apt to have high radio frequency resistance, as obviously the surface on which the currents can travel is restricted. (See *High Frequency.* also *Skin Effect* and *Resistance.*)

HIGH PASS FILTER—An arrangement of electric circuits designed to permit ready passage of high frequency alternating currents while at the same time presenting high resistance or impedance to alternating currents of low frequency, generally within certain definite limits. (See *Filter* also *Filter, High Pass.*)

HIGH PRESSURE—A term occasionally used to denote high potentials or voltages. (See *High Tension*, also *High Potential.*)

HIGH POTENTIAL—One of the several terms used to indicate high voltage. (See *High Potential Battery*, also *High Voltage.*)

HIGH POTENTIAL BATTERY—Another term for "B" battery, the battery used to supply potential to the plate of a vacuum tube. The term "high potential" is used here, only in a comparative sense, the voltage of the battery being high in proportion to that of the filament or "A" battery. High potential batteries for radio usage are generally furnished in blocks of 22½ or 45 volts. (See *High Voltage, High Tension* and *"B" Battery.*)

HIGH RESISTANCE—The property of any circuit whereby it offers considerable opposition to the flow of electric currents. There is no very distinct line of demarcation between high and low resistance. One of the common applications of the term is in the case of a *grid leak* or grid resistance. Here the resistance may be upward of a half million ohms, running into several million ohms in certain cases. An example of low resistance elements is that of a *rheostat* for controlling the filament current. Here the resistance is relatively low, rarely more than thirty ohms. The average resistance of the magnet windings in headphones (as used in radio) is one thousand ohms. The term "high resistance" is also used in a comparative sense in respect to various instruments or parts of radio circuits. Here the resistance in most cases should be as low as possible, and therefore a resistance of even a few ohms may be referred to as *high.* While resistance to direct currents is generally the feature of resistance coils and units, the opposition to currents of an oscillatory nature (high frequency alternating currents) may be of the utmost importance, and it is often the case that a circuit has little direct current resistance but very high resistance or *impedance* (q.v.) to high frequency currents. (See *Resistance, Impedance*, also *Skin Effect.*)

HIGH RESISTANCE JOINT—Any connection between two wires or conductors which offers resistance to the passage of current due to defective contact or imperfect connection through solder or any other medium. When bare wires are joined they should be carefully scraped and soldered. Too much solder is a bad feature, as the resistance of solder is naturally far higher than that of the copper wire. (See *Corrosion*, also *Soldered Joints.*)

HIGH RESISTANCE TELEPHONES—Headphones or telephones used in radio reception, having high resistance windings. In the ordinary telephone receiver used with the commercial telephone systems, the windings of the telephone magnets have a relatively low resistance, generally less than one hundred ohms. Such phones would not be sufficiently sensitive for radio usage, the average resistance in this case being about one thousand ohms. The resistance here refers to the ohmic resistance of the many turns of wire wound around the magnets. This resistance is not a required feature, the factor of importance being the number of turns. Inasmuch as space is a factor to be considered, it is necessary to use very fine wire in order to permit use of the thousands of turns essential. The resistance of this fine wire is of course high. Where very high resistance windings are used—and this indicates small wire—it is sometimes advisable to use a by-pass for the heavy plate potentials to protect the windings. (See *Ampere Turns*, also *Output Transformer* and *High Resistance*.)

HIGH SPEED RECEPTION AND TRANSMISSION—The reception or transmission of radio messages by automatic means in order to obtain much greater speed than would be possible under the manual system or hand opereating and recording. The general practice is to record the messages to be transmitted on a tape by means of a punching machine. They are then sent automatically through a high speed automatic transmitter and received automatically at the same high speed by a machine which punches the messages or records them on some form of tape or sheet. The more common of these methods is the *Wheatstone Transmitter*. (q.v.).

HIGH TENSION—One of the numerous terms used to denote high voltage. "B" batteries as used in radio reception are generally known as high tension batteries to distinguish them from the low voltage storage battery or dry cells used to light the filaments. Abbreviation H. T. mostly used in magazines and books printed in England.

HIGH TENSION BATTERY—Term applied to "B" batteries used in radio reception to supply potential to the plates of the vacuum tubes. These batteries usually are in blocks of twenty-two and a half or forty-five volts. (See *"B" Battery*.)

HIGH TENSION CIRCUIT—Term occasionally applied to the part of a radio circuit in which high potentials are present, as in the plate circuit of a vacuum tube. Here the "B" battery is in the circuit. (See *High Tension*, also *"B" Battery*.)

HIGH TENSION INSULATOR—Insulators of large dimension and rugged design, possessing unusual insulating properties in order to withstand high voltages. As a rule such insulators are made of high grade porcelain and arranged with annular ridges or "petticoats" to give greater surface and reduce the possibility of surface leakage. The illustration shows a form of high tension or high potential insulator. (See *Insulator*, *High Tension*, also *Petticoat Insulators*.)

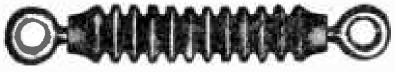

A high potential insulator.

HIGH VACUUM—A term applied to vacuum tubes when they have been highly exhausted of air. That is, an almost perfect vacuum. When a vacuum tube has a very good vacuum—has been highly evacuated of gases—it is generally called a "hard" tube. A high vacuum is essential in tubes to be used as amplifiers or for transmission. (See *Vacuum Tubes, Theory of Operation*.)

HIGH VOLTAGE—A very indefinite term, used in ordinary electrical practice to denote voltages above 600, but applied in radio practice to much lower potentials, as, for example, a "B" battery, which is generally referred to as a high voltage battery in spite of being generally below 100 volts. The term used in this sense is one of comparison with the low voltage of a storage or other type of "A" battery. (See *Voltagee*, also *Potential*.)

HIGHLY DAMPED WAVES—An expression applied to waves or trains of waves wherein the amplitude or strength of each succeeding oscillation or train of oscillations is greatly decreased. If the progressive decrease in amplitude or strength is very gradual, that is, if the waves die out slowly, they are said to be feebly damped, but if they die out rapidly they are referred to as being highly damped. (See *Damped Waves*, also *Decrement* and *Amplitude*.)

HOGAN, JOHN V. L.—American radio expert. Born at Philadelphia, Pa. He was educated at Sheffield Scientific School, Yale University, where he made a special study of physics and mathematics. In 1906 he became assistant to Dr. Lee De Forest, and in 1909 he joined the National Electric Signalling Co. In 1914 he was appointed chief research engineer to the International Radiotelegraphic Company. Hogan has written many articles on radio and is a past president of the Institute of Radio Engineers, member of the American Institute of Electrical Engineers, of the American Association for the Advancement of Science, of the Radio Club of America and other societies.

John V. L. Hogan.

HOMOPOLAR—Synonymous with unipolar. Having but one pole. Said of a dynamo having its conductor or armature rotating around a single pole of a field magnet. (See *Multi-Polar*.)

HONEYCOMB COIL—A coil used in radio, having a peculiar construction, somewhat similar to a honeycomb. The illustration shows a typical honeycomb coil used as an inductance in receiving circuits. The object of this staggered winding system is to reduce the self-capacity between wires to a minimum without reducing the inductive value. The chief value of such a winding method is that it permits the retaining of high inductance value in a limited space, without undue *high frequency resistancee* (q.v.). (See *Inductance*, *Distributed Capacity* and *High Frequency Resistance*.)

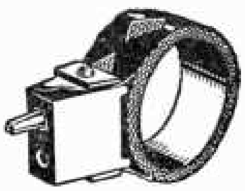

Honeycomb coil used in radio reception.

HOOK UP—The general term in the United States to designate the diagram for any complete radio receiver or transmitter, or for any specific part of such a circuit. Thus, when we refer to a hook-up for Radio receiver, we mean a diagram, either schematic (using symbols for the parts) or perspective (showing pictures of the parts) which gives the various parts necessary and the manner of connecting them to each other to form the complete circuit. Similarly we may use the term "hook-up" to describe a sketch or circuit for connecting various instruments such as *ammeters*, or *voltmeters*, *wavemeters* and the like. (See *Circuit*, also *Diagram*.)

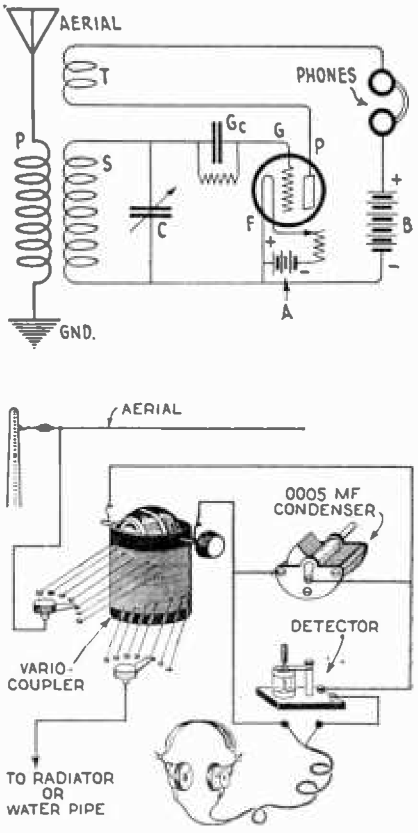

The upper diagram is in schematic form and shows a single tube regenerative receiver, and the lower diagram is in perspective form and shows a crystal receiving set.

HOOPER, STANFORD C.—American Naval wireless expert. Born August 16, 1884, at Colton, California, and educated at San Bernardino, California. He joined the Southern Pacific Company as a telegraph operator. In 1901 he entered the Naval Academy at Annapolis, and entering the Navy was advanced to Commander in 1918. From 1910-11 he was instructor in electrical engineering, physics and chemistry at the U. S. Naval Academy, and in 1912-13 Fleet Radio Officer of the United States Atlantic Fleet. During the great war he was responsible for the supply of wireless instruments, etc., for the American Navy, and he was also responsible for the construction of many of the larger American wireless stations including those at Annapolis, San Diego and Pearl Harbor. Hooper was one of the chief men concerned with the radio-compass system now used in the American Navy.

HORIZONTAL AERIALS—Aerials—overhead wire systems—for the reception or transmission of radio signals, wherein the major portion of the aerial is horizontal—parallel to the earth's surface. Typical examples are the "T" and inverted "L" aerials. In the former type, the flat portion suspended horizontally between two supporting structures and a lead-in wire is attached to the center and thence to the apparatus. In the latter type, the lead-in is attached to either end, giving a shape somewhat similar to an inverted L, from which the name is derived. The more common term is "flat top." This term is now somewhat ambiguous, as single wire aerials are used more or less extensively for receiving purposes, whereas formerly a system of several horizontal wires was employed. (See *Flat Top Aerial or Antenna*, also *Aerial T Type* and *Inverted Γ Aerial.*)

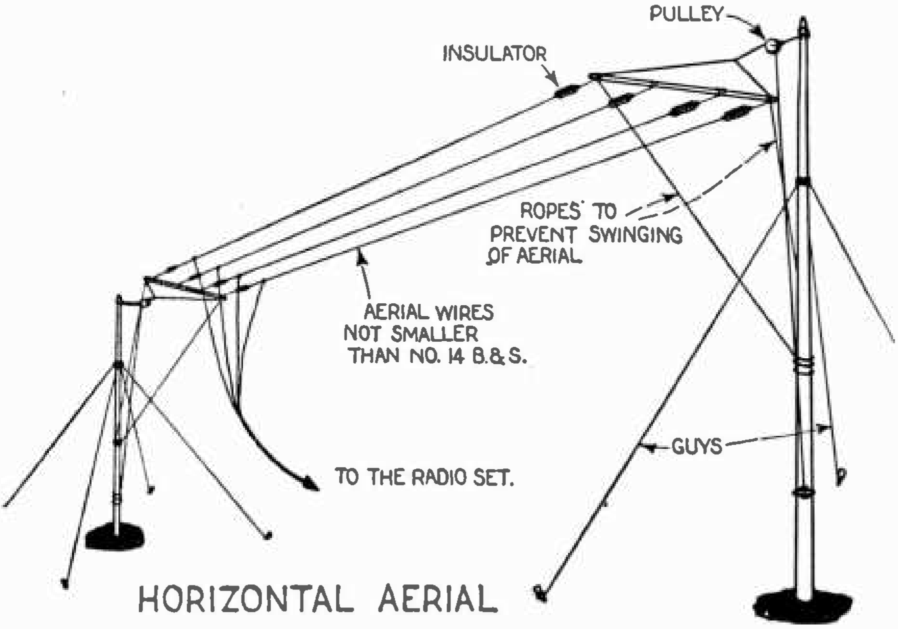
PULLEY
INSULATOR
ROPES TO PREVENT SWINGING OF AERIAL
AERIAL WIRES NOT SMALLER THAN NO. 14 B.&S.
GUYS
TO THE RADIO SET.
HORIZONTAL AERIAL

A horizontal aerial of the inverted "L" type.

HORN, LOUDSPEAKER—A device, generally in the shape of a horn, used in conjunction with a heavy duty telephone receiver or loudspeaker unit, to reproduce loudly the signals or music from a radio receiver. In operation, the horn has attached to it at its small end a unit or telephone receiver. When the diaphragm of the unit is actuated by the signals, speech or music, the column of air in the open tube is displaced, creating audible sounds which can be heard at a distance or throughout a room. Assuming other conditions equal, the volume will be more or less in proportion to the size of the norn, or more specifically to the amount of air displaced by the signals. When a telephone receiver is used without horn, the opening above the diaphragm being unconfined, only a small column of air is displaced and the volume will not be great under ordinary circumstances. When a horn is attached, however, the sound being confined will displace a relatively greater amount of air—depending mainly on the shape and size of the horn—and the volume will be measurably greater. (See *Loudspeaker*, also *Acoustics* and *Loudspeaker Units.*)

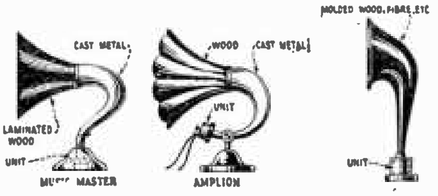

Three types of horn loudspeaker.

HORSE POWER—The unit of power used in the United States and Great Britain. It is defined as the power required to lift 550 pounds to a height of one foot in one second, or similarly, 33,000 pounds to that height in a minute. The term "foot-pounds" is derived from this value. Thus, one horsepower is the power required to perform 550 foot-pounds of work per second of time. (See *Horse Power, Electrical.*)

HORSE POWER, ELECTRICAL—The unit of electrical work similar to the mechanical horse power. It is actually mechanical horse power expressed in *watts* (q.v.). One electrical horse power is equivalent to 746 watts. In the *centimeter gram second* (q.v.) system of units, the unit of work is the *erg.* This unit is too small for ordinary usage and the customary unit is the watt, which is equal to 10^7 ergs per second. The unit used in electrical power work is still larger; it is the *kilowatt* or one thousand watts. (See *Units*, also *Watts.*)

HORSESHOE MAGNET—A *magnet* (q.v.) in the shape of a horseshoe or letter U, both poles being brought nearer together so that they may act on the armature or "keeper." Generally a magnetized steel bar which retains its magnetism for a long period. The illustration, Fig. 1, shows a permanent horseshoe magnet. Such magnets are also used with external power, then being termed *electromagnets.* In this case the magnet is made of soft iron. The illustration, Fig. 2, shows a horseshoe *electromagnet.* Here the magnetic attraction is supplied by application of an external electric force to the windings WW, and the magnet being of soft iron, loses its magnetic attraction almost immediately upon withdrawal of the electric current, magnetizing force. (See *Telephone Receiver*, also *Electromagnet.*)

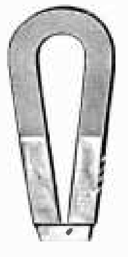

Fig. 1. A permanent horseshoe magnet.

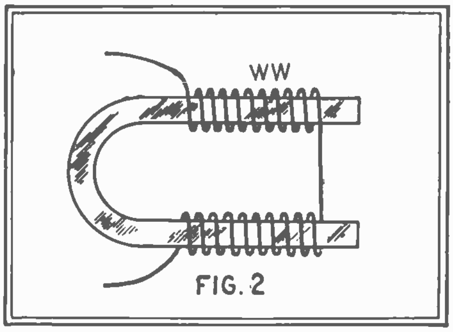

FIG. 2

Fig. 2. A horseshoe electromagnet.

HOT-WIRE AMMETER—An ammeter, or instrument for measuring electric current amperes, which depends for its action on the expansion of a fine wire under the influences of the heat produced in it, by the passage of the electric current to be measured. In the illustration the current to be measured is passed through the fine resistance wire A-B. P is a pointer moving over a graduated scale, and attached at its other end to a coiled spring S, which in turn is attached to the wire A-B by the silk thread T. The tendency of the spring is to move the pointer to the right, but owing to the pull of the thread in the normal position of the wire A-B, the pointer rests at zero at the left of the scale. As current is passed through the wire A-B, it heats and expands, thus permitting more or less slack to occur in the thread T and allowing the spring S to move the pointer over the scale toward the right, As soon as the current is withdrawn, the wire cools and contracts to its former size, thus again restoring the tension to the thread and returning the pointer to zero. Hot Wire ammeters are used extensively for measurement of currents at *radio frequency*, that is, extremely high frequency as in radio transmission. The average ammeter using a coil of fine wire would not be adaptable to this use, as the high voltage, high frequency current would burn out the coil. In addition to this. the length of these windings would affect the circuit and probably throw it out of *resonance* (q.v.). (See *Ammeter* or *Ampere Meter, Frequency, Resonance,* also *Instrument Shunts.*)

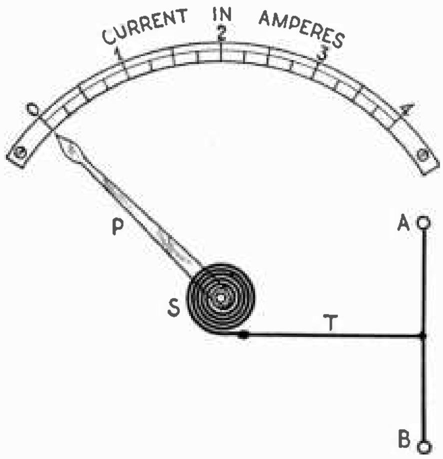

Illustration showing operation of a hot wire ammeter.

HOT WIRE INSTRUMENTS—Current measuring instruments which utilize the tendency of a fine wire to sag or expand under the heat caused by passage of an electric current through it. Such instruments may be used for direct or alternating current. (See *Hot Wire Ammeter.*)

HOT WIRE OSCILLOGRAPH — See *Oscillograph.*

HOT WIRE TELEPHONE—See *Thermal Telephone.*

HOT WIRE WATTMETER—An instrument used to measure combined volts and amperes, i. e., watts in a circuit. It comprises essentially a mirror, which is deflected according to the difference in expansion of two fine wires. (See *Hot Wire Instruments,* also *Wattmeter.*)

HOUR METER, AMPERE—An instrument used extensively in connection with storage batteries, to measure the input or output in amperes per hour. (*Ammeter* or *Ampere Meter.*)

HOUSE MAINS—A term much used in connection with storage battery charger and battery elimination devices. The term implies the main circuit of any standard house lighting system, either direct or alternating current. Thus in the case of a rectifier for charging storage batteries from alternating current, a lead is provided for connecting to the house mains, offering a means for introducing the required current into the device. (See *Current Direct,* also *Alternating Current* and *Chargers, Storage Battery.*)

HOWLING—The general term used to imply any undesired sounds produced in or by a radio receiver. Howling is usually caused by *feed-back effects* (q.v.). The most common example of howling is in the case of a regenerative receiver. Here the detector tube may have a tendency to generate its own oscillations As long as these oscillations are equal in frequency to any incoming signals there will be no howling, but should these local oscillations differ in frequency from the incoming series of signals, the result will be a howl resulting from the *beats* (q.v.) established owing to the difference between the two sets of oscillations. This is essentially the principle of *heterodyne* or "beat" reception, but in the case of a heterodyne, the effect is utilized, whereas in the case of a regenerative receiver it is undesirable. Interaction between various parts or different parts of a circuit may often lead to howling. (See *Feed-Back,* also *Regeneration, Reflex, Beats* and *Re-Radiation.*)

H. P.—Abbreviation sometimes used in radio for *high potential* or *high pressure.* In mechanics it is the standard abbreviation for *horse power.*

HOZIER-BROWN DETECTOR—A form of electrolytic detector, which, however, does not use a liquid electrolyte. It consists of a small pellet of lead oxide held more or less rigidly between two electrodes, one platinum and the other lead. (See *Detector,* also *Electrolytic Detector.*)

HYDROGEN—A gaseous chemical element. The symbol is H and its specific gravity is taken as 1, that of air being 14.4. It is colorless, tasteless and odorless and is the lightest known body. Hydrogen is a constituent of water, which is composed of one atom of oxygen and two atoms of hydrogen according to the chemical formula for water, H_2O. Hydrogen is used in radio transmission in connection with *arc generators.* (q.v.) (See *Poulsen Arc.*)

HYDROMETER—An instrument used to measure the *specific gravity* (q.v.) of a liquid. In radio it has an extensive use in connection with storage batteries. Various types of hydrometers are used to indicate the specific gravity of the electrolyte of a storage battery and hence in a manner, to indicate the state of charge of the battery. The illustration shows the standard form of hydrometer using a weighted float. In the illustration the float is shown inside the barrel of the instrument. In operation the bulb B is compressed and the tube T inserted in the open vent of a storage battery. When pressure on the bulb is released some of the electrolyte is drawn up inside the barrel and the float rises more or less within the barrel in proportion to the specific gravity or density of the electrolyte. With the hydrometer held in the manner shown, that is, upright, the numbers just at the level of the liquid in the barrel indicate the specific gravity. Thus, if the reading is 1250° or higher the battery is considered in fairly good operating condition —full charge being 1280° to 1300°. If the reading is below 1200° it indicates the need of charging, these levels not being arbitrary, a discharged condition being indicated by a reading of 1150° or lower. The use of a hydrometer is generally advisable and the reading accurate under ordinary conditions. If, however, water has just been placed in the battery, the reading may not be accurate for several hours due to the fact that the water and acid may not mix immediately. (See *Storage Battery Tests,* also *Electrolyte.*)

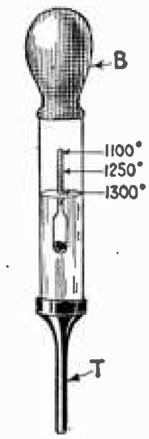

Hydrometer used to test the specific gravity of liquids.

HYGROSCOPIC EFFECT — The effect in any substance which permit it to absorb moisture. In the case of forms or tubes for radio coils and inductances this feature is particularly undesirable, as obviously, moisture lowers the dielectric value of the form and may cause it to warp. (See *Bakelite,* also *Tubing.*)

HYSTERESIS—The tendency of magnetization to lag behind the magnetizing force as, for instance, in an iron core in a *transformer* (q.v.). When the core of such a transformer is undergoing rapid reversals of magnetism, there is often an expenditure of energy which is not useful and which is converted into heat, thus representing a loss in power. This effect is more noticeable with certain qualities of iron. In other words, when rapid reversals of magnetism take place in a poor quality of iron core it may display a certain sluggishness. This is actually the lagging of the magnetic flux behind the magnetizing force producing it. This loss of energy is assumed due to the work required to alter the position of the molecules of the iron composing the core and the less energy expended in this manner the less the hysteretic losses. In designing transformers or armature cores, hysteresis must be taken into consideration and the iron must be of high quality to hold these losses at a minimum. (See *Eddy Currents,* also *Core, Transformer.*)

HYSTERESIS COEFFICIENT — The amount of energy wasted or lost in the process of magnetizing and de-magnetizing a unit volume of magnetic material is referred to as the hysteresis coefficient of that material. (See *Hysteresis,* also *Coefficient.*)

HYSTERESIS LOOP—Graphic illustration of losses due to hysteresis in any sample of core material. Loops are plotted in graphic curves of cycles of magnetization. (See *Curve,* also *Hysteresis.*)

HYSTERESIS LOSSES — The energy wasted or lost through hysteresis. (See *Hysteresis.*)

HYSTERETIC LAG — The lagging of magnetization due to the effect of hysteresis. (See *Hysteresis,* also *Magnetization.*)

I

I—The abbreviation adopted by the *International Electrotechnical Commission* (q.v.) and American Institute of Electrical Engineers, to denote current in amperes. The small or lower case (i) may also be used where the capital letter is not advisable. (See *Current.*)

I ARMATURE—An armature used in generating machines, having a core resembling the letter I in shape. (See *Armature,* also *Core.*)

I.C.W.—The customary abbreviation for "interrupted continuous waves." This is a system of radio transmission wherein the waves are modulated at a constant low frequency. (See *Interrupted Continuous Waves,* also *Modulation.*)

IDLE CURRENT—A name often applied to the component of an alternating current which contributes nothing to the power. This is also known as the *wattless component* (q.v.) or the *wattless current* (q.v.). It represents the component which, being in quadrature with the electromotive force in the circuit, thus has no active value—adds nothing to the total power. (See *Current,* also *Electromotive Force* and *Alternating Current.*)

I.E.C.—The abbreviation for *International Electrotechnical Commission,* the body which suggested the international electrical and magnetic abbreviations and symbols. (See *Symbols,* also *Units.*)

IMPACT OR SHOCK EXCITATION—The method of producing *free oscillations* (q.v.) in a circuit by means of an exciting current of short duration compared to that of the resultant oscillations. The term is a general one and broadly covers the various methods of generating oscillations in which an oscillatory circuit is thrown into electrical vibration at its natural or fundamental frequency by means of an electromotive shock or exciting force of short duration. Also covers the term *Impulse Excitation.* (See *Oscillations,* also *Fundamental Frequency.*)

IMPACT TRANSMITTER—Apparatus for transmitting radio waves wherein the transfer of energy from the primary or exciting circuit to the oscillatory circuit takes place during one pulse or vibration of the exciting current. (See *Impact or Shock Excitation,* also *Transmitter.*)

IMPEDANCE—The total resistance to the flow of alternating current in a circuit. Impedance combines the ohmic resistance and the apparent resistance due to self-induction. When a direct electromotive force is impressed on a circuit, the current flowing in that circuit depends directly on the ohmic resistance of the circuit. The formula is according to Ohm's law $I = \frac{E}{R}$, I representing the current in amperes, R the resistance in ohms and E the electromotive force in volts. Now if this direct electromotive force is replaced by an alternating EMF (electromotive force) and the ohmic resistance remains the same, the total resistance of the circuit to the alternating EMF will nevertheless be greater than in the case of direct current. Now the ohmic resistance is still the same but we have added to it the apparent resistance due to the self-induction of the circuit. In other words the inductance of the circuit has a marked effect on the current flowing in the circuit. What actually occurs is that the self-induction sets up a *counter-electromotive force* (q.v.) which continues to act against the impressed EMF as long as the flow of current persists. The result is a reduction in the current flowing in the circuit, the EMF having been, in effect, reduced, and hence according to the law above, $I = \frac{E}{R}$, the current is less. This added resistance, due to self-induction, is referred to as *reactance* and the flow of current in an alternating current circuit is thus governed by the ohmic resistance and the reactance. The application of the term "apparent resistance" to describe the added resistance due to self-induction, is a concession to the fact that it is not actually resistance in the proper sense of the word. In a direct current circuit, the resistance actually wastes or dissipates a certain amount of energy. In the alternating current circuit, the ohmic resistance has the same effect. The reactance, however, does not dissipate energy. It merely reduces the effective EMF by means of the counter EMF and thus necessitates the application of a higher impressed voltage in order to maintain the same flow of current.

This is one of the major advantages of alternating over direct current. Alternating current can be transformed readily and can also be readily controlled by means of a reactance or choke coil, which by creating a counter EMF limits the current flowing in the circuit without dissipating any great amount of energy. (See *Alternating Current, Current Direct, Choke Coil,* and *Reactance.*)

IMPEDANCE COIL—A coil of wire or inductance coil, generally wound on an iron core, used in alternating current circuits to limit the current flowing in the circuit. In action such a coil serves through the medium of its self-induction, to create a counter electromotive force which acts against the impressed EMF (electromotive force) and thus reduces the current in the circuit according to ohm's law that the current is equivalent to the EMF in volts, divided by the resistance in ohms. An impedance coil is also known as a choke coil although the two may have somewhat different meanings. For example, a choke coil might be used in either direct or alternating current circuits, whereas an impedance coil as its name indicates, offers impedance, an effect only present in alternating current circuits. The simplest formula for impedance of a coil is the following: Impedance $= \sqrt{R^2+P^2L^2}$. Here it is assumed that the current takes the form of a *sine wave* (q.v.), then R is the resistance of the coil, L the coefficient of self-induction in henries and p is 2 π times the number of alternations per second. (See *Choke Coil, Alternating Current, Current Direct, Reactance, Impedance,* also *Impedance Coupled Amplifier.*)

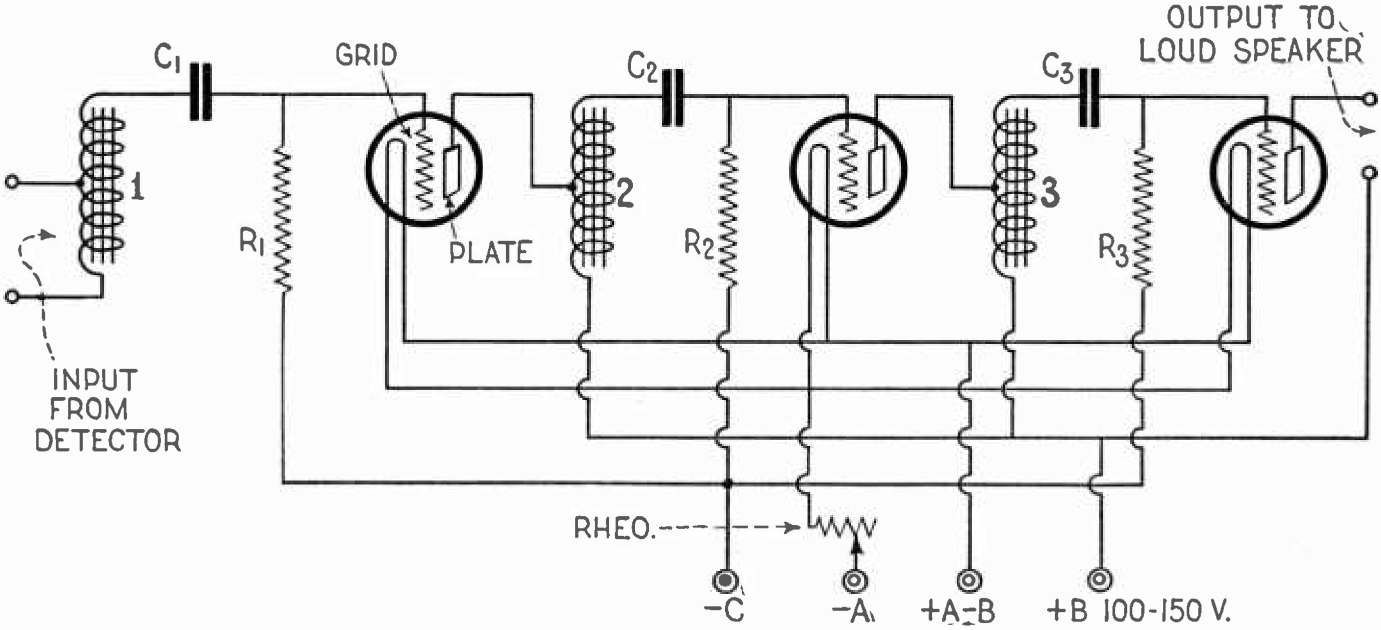

Schematic diagram of an impedance coupled audio frequency amplifier. *(See Impedance Coupled Amplifier.)*

IMPEDANCE COUPLED AMPLIFIER—A system of amplifying radio signals, voice or music, whereby the successive stages are coupled or joined by means of impedance coils. This system is much used in audio amplifiers, the amplification being much more uniform over the necessary range of frequencies than in the case of the ordinary audio frequency transformer. This is also known as choke coil coupling or choke coil amplification. The illustration on the preceding page shows a typical impedance coupled amplifier for audio frequency currents. It will be noted that three stages are used. This is necessary owing to the lower amplification per stage. Using this system, then, three stages are necessary in place of the customary two stages of transformer coupled amplification. In the illustration, which is a schematic diagram of the circuit for such an amplifier, the impedances 1, 2 and 3 are in the form of *auto transformers.* (q.v.) R1, R2 and R3 are grid leaks ranging in value between one quarter megohm and one and one-half megohms. C1, C2 and C3 are stopping condensers of about 1 micro-farad capacity each. The variations in potential are transferred successively from stage to stage, the amplification factor being roughly the sum of the individual factors of the tubes. In other words the ratio of the windings of the impedances is one to one, or unity, and therefore no voltage amplification takes place. The resistances are the usual grid leaks and the condensers C1, C2 and C3 are to block the high plate voltages from the grids of the succeeding tubes. (See *Tuned Impedance*, also *Amplifier* and *Vacuum Tube Amplifiers, Theory of.*)

IMPEDANCE FACTOR—The ratio of the impedance to the ohmic resistance in an alternating current circuit. (See *Impedance*, also *Resistance.*)

IMPEDANCES, MATCHED—Arrangement of an amplifying circuit and the choice of an audio transformer or loudspeaker with due regard for the impedance of the circuit. Thus if the impedance of a certain amplifier tube is 10,000 ohms, and the resistance of a loudspeaker unit being used is 2,500 ohms, then for maximum response when used with an *output transformer* (q.v.) the impedance of the primary of the transformer should be equivalent to the total or 12,500 ohms. (See *Matched Impedances.*)

IMPREGNATED INSULATING MATERIALS—Cloth or paper impregnated with an oily or resinous substance to render it moisture proof and improve the insulating properties. Cardboard is sometimes treated in this manner and used for tubing to wind inductance coils. (See *Empire Cloth or Paper*, also *Insulating Property.*)

IMPRESSED ELECTROMOTIVE FORCE—The term used to distinguish the EMF (electromotive force) impressed on or applied to a certain part of an electrical circuit from the counter or back-electromotive force due to constants of the circuit. The term is sometimes used where a direct current is applied in a circuit, but it should be confined to use where a counter EMF (electromotive force) is present. (See *Counter Electromotive Force*, also *Choke or Choking* and *Reactance.*)

IMPRESSED FIELD—An electric or magnetic field of force impressed on or applied to a body or region (area) for the purpose of producing other fields of force. (See *Field.*)

IMPULSE—Electrically, an electromotive force which produces an impulsive rush or discharge of electricity. The term is used to distinguish between this form and an electromotive force producing a steady flow of current. (See *Current*, also *Electromotive Force.*)

IMPULSE E.M.F.—The more specific term designating an EMF (electromotive force) which produces an impulsive rush or discharge of electricity. It is defiined technically, as having a maximum value which is large compared with its average value, the average value being taken over a time equivalent to the *time-constant* (q.v.) of the circuit on which the EMF is impressed. (See *Current, also Electromotive Force.*)

IMPULSE EXCITATION—Method of exciting or starting oscillations in a circuit, such as the aerial circuit of a transmitter, by a sudden surge of impulse EMF (electromotive force) rather than by application of oscillations of the frequency of the circuit.

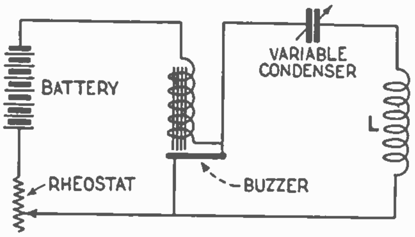

Connections of buzzer for generating damped oscillations suitable for testing by impulse excitation.

This system of impulse or shock excitation may be used for transmission and also has application in connection with testing or comparing of wavemeters. The illustration shows a buzzer connected with the necessary components to form a source of high frequency oscillations. This system is capable of producing through coupling from the coil L, currents by shock excitation. These currents can be used in a variety of ways for testing purposes. In operation the condenser is charged to the voltage of the battery at the instant the buzzer contact is open and when the contact closes, discharges through the coil L thus furnishing a series of oscilations to surge through the coil and by inductive coupling to any other coil placed in proximity to it. (See *Buzzer Exciter.*)

INCANDESCENCE—The glowing of any substance when brought to a white heat. The term is used in radio to signify the condition of the heated cathode or filament in a vacuum tube. In order to produce *electrons* (q.v.) it is necessary to cause the filament to become white hot, the condition being known as incandescence. (See *Filament.*)

INCONDUCTIVITY—A little used term signifying the condition of a substance by virtue of which it will not act as a conductor for electric currents. (See *Conductivity, Specific.*)

INDEPENDENT UNIT—A unit of a quantity, such as length, mass or time, chosen arbitrarily and without having any relation to any other unit. (*Centimeter Gram Second, or C. G. S.*, also *Fundamental Units.*)

INDICATING INSTRUMENTS—Any instrument or device which indicates values, as for example, a *voltmeter* or *ammeter.* (q.v.) (See *Instruments, Measuring.*)

INDIRECT CURRENT—Term for alternating current—now obsolete. (See *Alternating Current.*)

INDUCED CURRENT—The current in a circuit due to an induced EMF (electromotive force). The current produced by an EMF induced in a circuit due to a change in the magnetic flux linking the circuit with another. (See *Induced EMF*, also *Induction.*)

INDUCED E.M.F.—The electromotive force induced in a conductor or body. The phenomena of induced EMF (electromotive force) is the basis of operation of generating machines. Here a conductor is rotated or moved in a field of force, the action causing a change in the flux lines linking the circuits or the conductor. An EMF may also be induced by placing a coil in inductive relation to another coil through which alternating current is passing. Here the transfer is caused by induction. (See *Induction.*)

INDUCTANCE—The broad term used to designate any wire wound coil as used in radio practice for tuning units, inductive couplers and any variety of purpose. Actually the term is derived from self-induction or self-inductance, a property of a coil of wire by virtue of which it sets up a counter EMF (electromotive force) which serves to oppose the original EMF passing through it. When a certain circuit has inductance, it is to say that a number of turns of wire possessing self-induction are present in the circuit. The greater the inductance in a circuit the greater its opposition to the passage of an alternating current. The unit of inductance is the Henry. This is defined as the self-induction of a circuit or coil when the induced EMF is one volt with the inducing current varying at the rate of one ampere per second. (See *Alternating Current, Henry*, also *Inductance Coils.*)

A type of antenna inductance used in radio transmission.

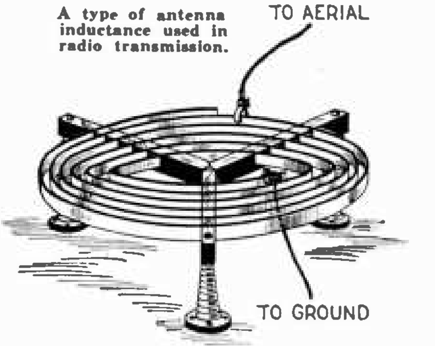

INDUCTANCE, ANTENNA—Inductance (coil) placed in antenna circuit of transmitting set for purpose of increasing wavelength of the open or antenna circuit to any desired level. It is generally referred to as an antenna loading inductance or loading coil. In high power transmitting stations it is very often the case that the natural period or fundamental wavelength of the antenna is not as great as the intended operating wavelength. In this case a loading inductance is used to bring it up to the desired value. In some instances the fundamental wavelength is too high, in which case a condenser is placed in series to reduce the wavelength and then a few turns of inductance placed in series with the condenser and antenna for the purpose of tuning. Such inductances must be of a very rugged nature when high power is used, generally consisting of a comparatively few turns of heavy copper ribbon mounted on high tension insulators. The illustration shows a common form of antenna inductance used in radio transmission. (See *Loading Inductance*, also *Tuned Antenna.*)

INDUCTANCE COILS—Any coil of wire wound in such a manner as to possess the property of self-induction. Such a coil may be in almost any conceivable form, providing the turns are separated or insulated from each other. Inductance coils are a very necessary part of radio circuits, being used in conjunction with condensers to obtain resonance or desired wave-length. Such coils can also be arranged as separate circuits and placed in inductive relation to each other, thus permitting energy present in one to be transferred through induction to the other. The most common example of inductance in a receiving set is that of a tuning coil or vario-coupler. The illustration shows several forms of inductances known as tuning inductances. (See under separate headings as follows: *Inductance, Loading, Tuning, Tapped,* also *Inductance, Units of.*

Honeycomb type inductance coils.

Low-loss air-wound inductance coil made especially for tuned radio frequency work.

INDUCTANCE, DISTRIBUTED—Added inductance distributed throughout the length of a cable or telephone line to compensate for the capacity of the line. (See *Distributed Inductance.*)

INDUCTANCE, LOADING—Literally an inductance to load a circuit. The term is used to designate any inductance coil or unit used in series with the antenna for the purpose of increasing the wave-length of the circuit. (See *Inductance, Antenna.*)

INDUCTANCE, MEASUREMENT OF—Any process or means of determining the inductance value of a coil or circuit. The usual method is by formula. (Fully described under *Measurement of Inductance.*)

INDUCTANCE, MUTUAL—Symbol M. The number of linkages of flux lines or the flux that is common or mutual to two circuits which are adjacent to each other or inductively coupled. It is generally defined as the amount of the *flux* (q.v.) which is common to two linked circuits per unit of current flowing in the inducing circuit. In the illustration circuit A contains an inductance L_1 which is being traversed by an alternating electromotive force. Circuit B contains merely an inductance L_2 and a measuring device, M. Now, as the alternating E.M.F. passes through coil L_1 it sets up a field of force, or rather flux lines as shown. These lines link with the coil L_2 and an alternating E.M.F. is set up. Now, suppose that the inducing current flowing in circuit A is changing at the rate of one ampere per second and an E.M.F. of one volt is induced in the circuit B, or rather in coil L_2. We then say that the mutual inductance of the two circuits will be one henry. This is the same unit used for self-inductance and in fact, mutual inductance and self-inductance are virtually the same thing.

The effect of mutual inductance must be taken into consideration when measuring the fundamental wave-length, capacity or inductance of an antenna. That is to say, where a number of parallel wires are used, the mutual inductance between them must be taken into consideration as well as other values or constants. (See *Self-Inductance, Mutual Inductance, Induction, Coupling* and *Induced E.M.F.*)

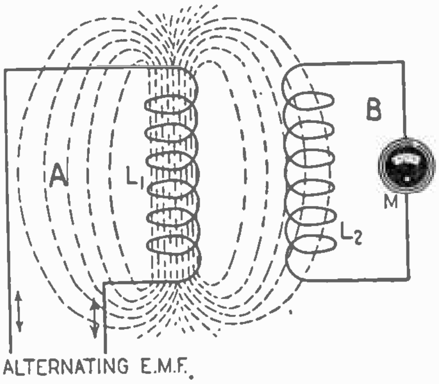

Diagram which illustrates the mutual inductance between coil L_1 in circuit A and coil L_2 in circuit B.

INDUCTANCE, REACTANCE—Symbol x—The *reactance* present in a coil possessing self-inductance. The self-inductance in such a coil opposes the flow of alternating current due to the back *electromotive force* set up as a result of the changing magnetic field of the current. *Inductance reactance* (or *inductive reactance*) is the term used to define this opposition. *Inductance reactance* acts similarly to *resistance* (q.v.) in that it opposes the flow of an electric current. It is measured in *ohms* (q.v.), just as *resistance* is measured. It should be noted, however, that *inductance reactance* only opposes the flow of alternating currents, whereas resistance opposes the flow of both direct and alternating currents. Furthermore, resistance in offering opposition to current flow consumes power, while *inductance reactance* merely takes up energy from the circuit momentarily while current is increasing and thereafter gives it back while the current is decreasing.

The combined opposition offered to the flow of alternating current by the resistance and the inductance reactance of a circuit is called the *impedance* (q.v.) of that circuit. The voltage drop due to resistance is in *phase* (q.v.) with the current. The voltage drop due to *inductance reactance* is 90 degrees in advance of the current.

Expressed as a formula, $x = \sqrt{Z^2 - R^2}$, where x is the inductance reactance, Z is the impedance and R is the resistance, all measured in ohms.

Knowing the *frequency* (q.v.) of the alternating current and the inductance of the circuit through which this current flows, it is possible to calculate the *inductance reactance* by using the formula $x = 2\pi fL$, where π is a constant equal to 3.1416, f is the frequency in *cycles* per second, and L is the *inductance* of the circuit in *henry's* (q.v.). This formula is sometimes abbreviated by substituting the symbol ω (omega) for $2\pi f$. From the formula it can be seen that the *inductance reactance* increases directly both with the inductance of the circuit and with the frequency of the current flowing. (See *Current,* also *Electromotive Force, Resistance, Reactance.*)

INDUCTANCE, RECEIVING—Broadly, any form of inductance used in receiving circuits. The term thus includes all types of coils for tuning purposes, such as couplers, tuned transformers, variometers, etc. The illustration shows a common form of tuning inductance. Here the coil L is an inductance in the aerial circuit and L_1 the coil in the secondary circuit. Here the energy from the aerial traverses coil L and by induction is transferred to coil L_1. (See *Tuning Coil, Vario-Coupler, Receiving Circuit.*)

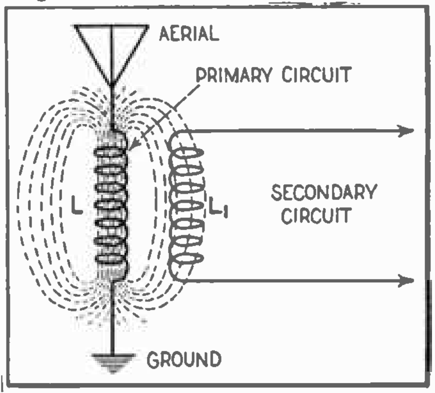

A form of tuning inductance sometimes used in radio receiving sets.

INDUCTANCE, SELF—Symbol L—An alternate term with self-induction for the *inductance* (q.v.) due to the field of force produced by the circuit itself. It is virtually the same as mutual inductance. For example, in a coil of wire being traversed by alternating currents the adjacent turns may act upon each other on the principle of mutual induction between two separate adjacent circuits. The induced currents are the result of self-induction. (See *Self-Induction,* also *Mutual Inductance* and *Induction.*)

INDUCTANCE, TAPPED—Any type of inductance having taps for connecting at different points on the coil in order to vary the inductance value. The illustration shows a form of tapped inductance. Here the taps at different portions of the coil are connected to switch points and the switch lever permits instant change to any of the available values. (See *Inductance Coils,* also *Tuning Coil.*)

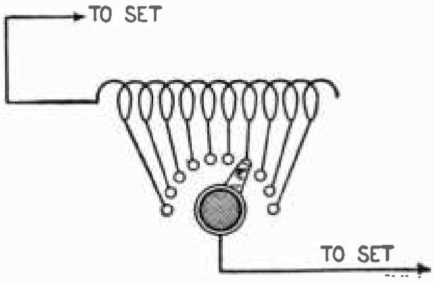

A tapped inductance which permits inductance to be added or taken out of the circuit by rotating the lever switch.

INDUCTANCE, UNITS OF—The measuring units for inductance. The practical unit is the henry (H), the value of inductance which, if current varies at the rate of one ampere each second, results in an *induced electromotive force* (q.v.) of one volt. When we speak of an induced E.M.F. of one volt, the volt is used as a practical unit. Under the electromagnetic system of units, the absolute unit of induced E.M.F. is known as the abvolt. This unit is very small, 100,000,000 of them being required to equal one volt. We thus say that the volt of induced E.M.F. is equal to $1\text{x}10^8$ abvolts. The electromagnetic unit of E.M.F., or the abvolt, is defined as the E.M.F. induced in a conductor cutting one line of magnetic force each second. As we have shown that one volt is equal to $1\text{x}10^8$ abvolts of induced E.M.F., and since one ampere is equal to $1\text{x}10^{-1}$ abamps of current (see *Units, Electromagnetic*) then the unit of inductance or one henry $= \dfrac{1\text{x}10^8}{1\text{x}10^{-1}}$ or) $1\text{x}10^9$ abhenrys. The unit of inductance (the henry) being too large for practical purposes, the *milli-henry* (one-thousandth henry) or the *micro-henry* (one-millionth henry) are commonly used. Occasionally the *centimeter* is used as a measure of inductance, being the thousandth part of a micro-henry or a *milli-micro-henry*. The milli-henry is written as $1\text{x}10^{-3}$ henry, the micro-henry as $1\text{x}10^{-6}$ henry, and the centimeter as $1\text{x}10^{-9}$ henry. (See *Induced Current, Induced E.M.F., Induction, Electromagnetic,* also *Units.*)

INDUCTANCE, VARIABLE—Any coil possessing self-inductance and being arranged in such a manner that the value of the inductance can be changed at will. The simplest form of variable inductance is the sliding contact tuner. This is merely a coil of wire of any predetermined value (maximum) mounted and arranged with a sliding contact moving over the wires, which are bared of insulation at a certain point to permit contact. By using the sliding contact as one connection and either end of the coil as the other lead, the effective value of the inductance in the circuit may be varied within the limits of the coil itself. Other forms of variable inductance include the variometer, vario-coupler and various modifications. The illustration shows the two more common forms of variable inductance. (See *Inductance,* and *Inductance Coils.*)

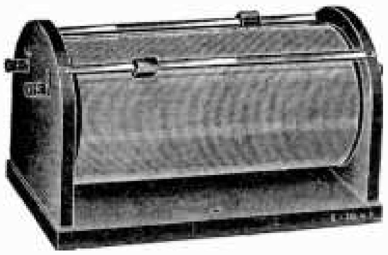

A sliding contact tuner formerly used in radio receiving sets.

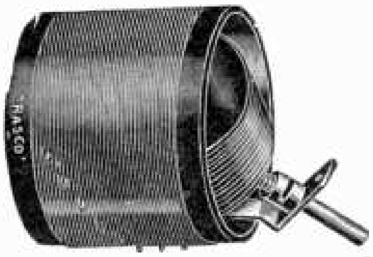

Illustration by courtesy of Radio Specialty Co.
Variable inductance known as a variocoupler.

Illustration by courtesy of Radio Specialty Co.
Another type of variocoupler.

INDUCTANCES IN PARALLEL—Coils possessing the property of self-induction or self-inductance, connected in parallel or shunt to each other. The illustration shows two standard coils connected in a circuit in this manner. The effective inductance is less than that of either coil alone. (See *Inductance Coils.*)

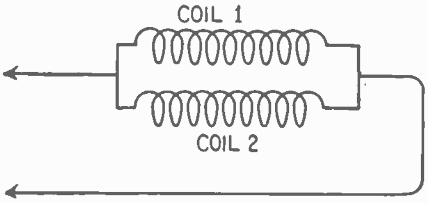

Diagram showing connection of two inductance coils in parallel in a circuit.

INDUCTANCES IN SERIES—Coils possessing the property of self-induction, connected in series in a circuit. The illustration shows two coils connected in series in a circuit. Here the effective inductance value is the sum of the separate inductances. (See *Inductance Coils.*)

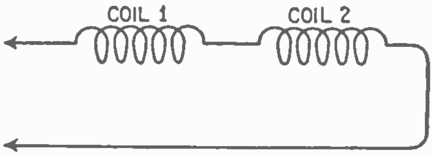

Diagram showing two inductance coils in series in a circuit.

INDUCTION—The influence exerted by one field of force (electromagnetic or electrostatic) on another or on a conductor. Whenever a magnetic field interlinked with an electric circuit is changed, an *electromotive force* (q.v.) is induced in that circuit. The magnetic field is considered to originate from the heart of the conductor, spreading out concentrically from the conductor as the current increases and shrinking back into the center of the conductor as the current decreases. In either case, the conductor is cut by the magnetic field, and in accordance with the *law of electromagnetic induction* (q.v.) an electromotive force is induced in the circuit proportional in amount to the rate of change of the magnetic field and acting in a direction which would, by producing a current, tend to oppose the change.

A bothersome type of electromagnetic induction sometimes occurring in radio takes place when a radio aerial is run parallel to a high-tension line or even a trolley line. The varying field of force about the high-tension line sets up, by induction, a corresponding field in the parallel aerial, thus causing a disturbing hum in the loud speaker. Such a condition can usually be eliminated by running the aerial at right angles to the power line. By doing this, induction between the two conductors is reduced to a minimum.

Electrostatic induction is a rearrangement of an electrostatic field whereby a body in the neighborhood of a charged body will have an opposite charge induced on it. An *induction machine* (q.v.) known as the Wimshurst utilizes the principle of electrostatic induction.

INDUCTION BRIDGE—Also called inductance bridge. A balance arranged similarly to a *Wheatstone Bridge* (q.v.), used in induction measurement. (See *Induction, Measurement of Inductance,* also *Induction Density.*)

INDUCTION COIL—A form of transformer, having a primary winding arranged with a means of making the direct current applied to it, pulsate.

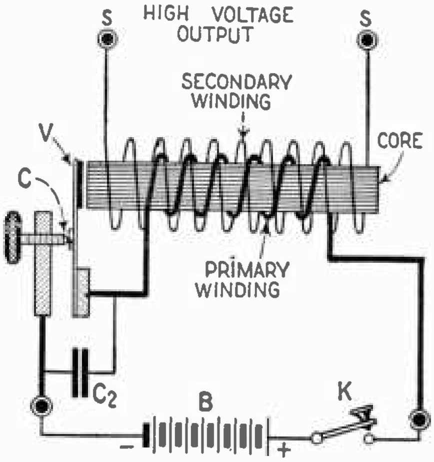

A typical form of induction coil showing diagram of connections, vibrator, core, battery, condenser and key.

Essentially a transformer with open magnetic circuit, carrying a pulsating direct current which induces a high voltage, alternating current in the secondary by means of the step-up effect of the windings. The illustration shows a typical form of induction coil and the associated circuits for producing high voltages. Here, there are two windings, a primary and a secondary, over and insulated from a core, composed of a number of soft iron wires. The primary is arranged with a key, K, and a battery, B, to supply direct current to the circuit. When the key is depressed the current flows through the primary circuit, including the contact C and the vibrator V. The core becomes *saturated* (q.v.) or magnetized and attracts the armature or spring V. As this spring is drawn toward the core it breaks the contact between the spring and the contact post. This, of course, breaks the circuit from the bat-

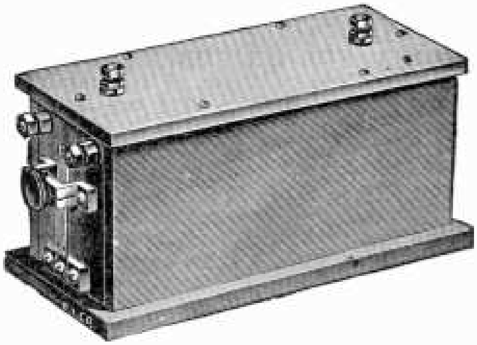

A form of spark coil.

tery and the flow of current stops, resulting in the collapse of the flux through the core. When the magnetic attraction ceases the spring goes back into normal position and the contact is once more complete, allowing current to flow through the windings. This make-and-break process is more or less rapid and is repeated continuously. By induction, an alternating E.M.F. (electromotive force) is created in the secondary windings, and as these windings have many more turns than the primary winding the induced E.M.F. will be proportionately greater. If the breaks occur at a rate of from thirty to one hundred per second, the key can be used to interrupt the primary current and produce at the secondary terminals a high voltage suitable for spark transmission. (See *Induction, Transformer, Spark Coil.*)

INDUCTION COIL, PRIMARY — The simplest form of induction coil. It consists of an iron core wound with a relatively large amount of low resistance copper wire, which is well insulated. The iron core increases the effect of self-induction and the size or length of the core and the number of turns of wire determine the efficiency of the coil. Such coils are used extensively for low tension ignition. (See *Induction Coil.*)

INDUCTION COIL, SECONDARY — A type of induction coil employing a primary and secondary winding. The primary is wound over an iron core and the secondary is generally wound over the primary. (See *Induction Coil.*)

INDUCTION DENSITY—A term occasionally used to refer to the density of flux in a magnetic or electromagnetic field of force. It is synonymous with flux density, under which heading it will be found more fully defined. (See *Induction, Electromagnetic*, also *Gauss*, and *Induction, Magnetic.*)

INDUCTION, ELECTROMAGNETIC — The production or *inducing* of electric currents in a conductor when it is moved in a magnetic field of force in such manner as to cut magnetic lines of force. The simplest example of this phenomena is the Faraday experiment in which he discovered that a wire with its ends joined, when moved rapidly in the field of a magnet, would result in current being induced in the wire or conductor. All dynamo electric machinery is based upon this principle of electromagnetic induction, as are also alternating current transformers and various other electrical devices.

In the illustration is shown the theory of production of current by electromagnetic induction. Fig. 1 shows a coil of wire placed in a uniform magnetic field of force between two magnets. Here the coil is assumed to have been moved from A to B in the same plane. This is known as a *motion of translation*, and the field being uniform at all points the same number of lines pass through the coil at both points. Therefore no current is induced in the coil and a sensitive instrument placed across the leads, C, C, would show no current present. In Fig. 2 the same coil is shown at A in its original position. Here, however, it is rotated, and when in position B it is apparent that less lines of force pass through it. We have thus fulfilled the conditions for inducing current, and if the instrument is placed across the leads, C, C, the presence of electric current will be indicated. If the motion of rotation is continued the number of lines of force passing through the coil will continue to increase and decrease and current will be produced in the coil as long as the motion persists. From the above it will appear that current can be induced either by rotating the coil around any axis in its plane, or by tilting it in its motion across the field. The coil of wire in this case is known as an *inductor* because the current is induced in it. It is also referred to as a *conductor* when describing the production of alternating currents.

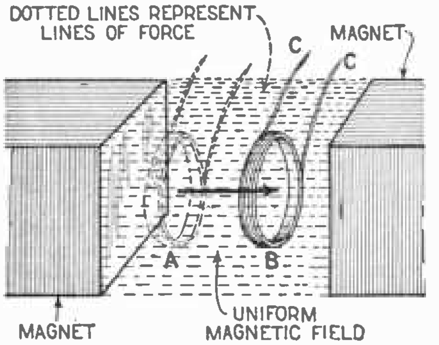

Fig. 1. Coil of wire in a uniform field moving parallel to direction of the field.

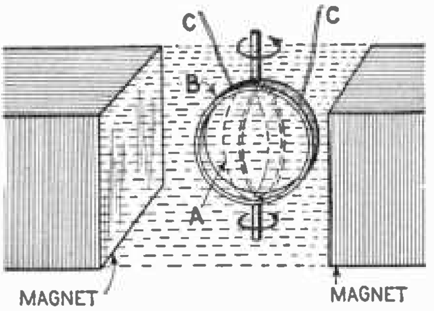

Fig. 2. Same coil, showing cutting of field due to the rotation of the coil.

Now it was stated that current is induced in the inductor whether the lines of force through it are *increased* or *decreased* by its motion through the field. We can reduce the action to two simple rules. First, the relative motion between an inductor and a magnetic field must be such that the number of lines passing through the inductor are altered, i.e., increased or decreased. Second, the electromotive force induced in the inductor circuit will be proportional to the rate of increase or decrease in the number of magnetic lines embraced by the inductor circuit. We know from the *Lenz Law* (see *Alternating Current, Theory of Production of*) that current will flow in the inductor in a direction dependent upon whether the lines of force through it are being increased or decreased, and as the inductor is being rotated continuously the current will change direction with each change in the number of lines. (See *Alternating Current, Alternator, Electromotive Force, Fleming's Rule, Laws of Electromagnetic Induction, Lenz's Law.*)

INDUCTION, ELECTROSTATIC — The production of an electric charge in a body due to the presence of a nearby conductor having an opposite charge. The effect is known as a displacement of electric charge. (See *Electrostatic Induction*, also *Induction.*)

INDUCTION INSTRUMENTS—Measuring instruments used in electrical and radio practice, which make use of the principle of electromagnetic induction. Such instruments are used for measuring alternating currents, the torque or actuating force for the indicator being produced by the effect of a rotating field on a metal disc or drum which is not connected to the circuit carrying the currents. The instruments operate on the same basic principle as the *induction motor* (q.v.).

INDUCTION MACHINE — A machine wherein the primary and secondary windings rotate with respect to each other. Examples of induction machines are induction motors, certain types of frequency converters and induction generators. An entirely different kind of induction machine is one which generates low potentials, utilizing the principle of *electrostatic induction* (q.v.). The most common form of this type is known as the Wimshurst (q.v.).

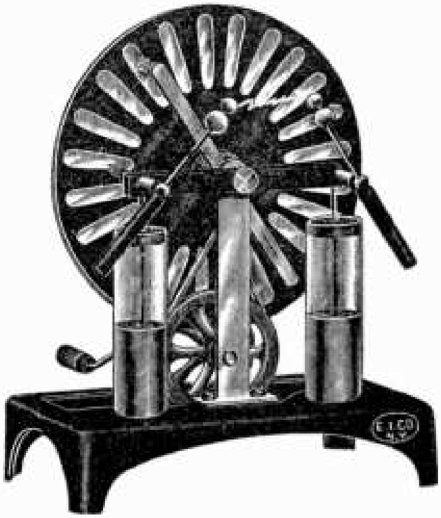

A common type of induction machine known as the Wimshurst.

INDUCTION, MAGNETIC — The communication of magnetism to a metal by the presence of a magnet without any actual contact. The effect of the presence of a magnet in the neighborhood of a piece of magnetic material is to induce magnetic poles in it. (See *Magnetism, Magnetic Poles, Induction*, also *Induction, Electromagnetic.*)

INDUCTION MOTOR — An alternating current motor in which the input current is passed through the field coils only, the armature not being connected to the external circuit. Here the armature is rotated by currents induced by the varying field set up through the field coils. The induction motor is the most widely used of the alternating current types, maintaining a practically constant speed from no load to full load and being simple and rugged in construction. This type of motor has two windings, the primary and secondary, and two members called stator and rotor. The primary winding is usually on the stator and the secondary winding, which is therefore customarily on the rotating member, is not connected electrically to the primary. Current to operate the rotor is induced in the secondary windings by the magnetic action of the current circulating in the primary windings. (See *Alternating Current*, also *Induction, Electromagnetic.*)

INDUCTION, MUTUAL—The interference or mutual effect between two electric or magnetic fields, due entirely to their proximity and without electrical contact. The mutual induction or electromagnetic influence of one circuit upon another is measured by the coefficient of mutual inductance or induction. (See *Mutual Induction*, also *Mutual Induction Coefficient*, and *Mutual Inductance.*)

INDUCTION REACTANCE—The value in ohms of the inductance in an electrical circuit as distinguished from the capacity or capacitive reactance. It is the part of the total impedance in an alternating current circuit that is due to the presence of inductance in the circuit. (See *Inductance Reactance, Ohm, Impedance.*)

INDUCTION REGULATOR—A system of regulating the voltage in alternating current circuits, employing a choke coil with removable core, or some form of variable ratio transformer which will assist or oppose the current by readjustment of the ratio. (See *Choke Coil, Reactance,* and *Variable Ratio Transformer.*)

INDUCTION SCREEN—A metal screen or shield placed between two electrified or magnetic bodies to reduce the effect of induction between them. A common example of this in radio usage is the placing of a sheet of metal between two coils, the sheet being generally grounded, or the provision of a metal *shield* (q.v.) entirely around each coil. This reduces the interaction of the fields of the coils, which leads to losses or howling, especially in *tuned radio frequency amplifiers* (q.v.). (See *Shielding.*)

INDUCTION UNIT—The *henry,* defined as the induction in a circuit when the electromotive force induced in that circuit is one *international volt* (see *volt*) while the inducing current varies at the rate of one ampere each second. (See *Henry,* also *Self-Induction.*)

INDUCTION VOLTMETER—See *Induction Wattmeter,* also *Induction Instruments.*

INDUCTION WATTMETER—An instrument for measuring the power delivered to a circuit. The induction type wattmeter operates on a principle used in the induction motor, a magnetic body attached to a pointer being rotated by a changing magnetic field. (See *Induction Motor,* also *Wattmeter.*)

INDUCTIVE CAPACITY—Generally referred to as the specific inductive capacity in rating the dielectric qualities of a substance. It is the quality of any dielectric substance by reason of which it permits electrostatic induction through it. The term is more or less synonymous with *Dielectric Constant,* the latter, however, being generally given in terms of a numerical value representing its comparative specific inductive capacity with some material taken as a standard. (See *Dielectric Coefficient and Constant,* also *Specific Inductive Capacity.*)

INDUCTIVE CIRCUIT—An electrical circuit possessing inductance but not capacity. (See *Inductance,* also *Capacity.*)

INDUCTIVE CONNECTION—A connection between two circuits entirely through the property of induction and without any metallic contact. The two coils, primary and secondary, of a vario-coupler as used in radio receiving circuits, are said to have inductive connection where the windings are separate. Thus, the primary coil will be in one electrical circuit and the secondary coil will receive energy from it by means of induction. (See *Induction, Coupling, Vario-Coupler.*)

INDUCTIVE COUPLER—The inductances or coils placed in inductive relation or inductively connected with each other for the purpose of transferring energy by electromagnetic coupling. Any device consisting of a primary and secondary inductance and used to transfer energy from one circuit to another by means of induction. (See *Coupler, Induction, Vario-Coupler.*)

INDUCTIVE COUPLING—The coupling or connection between two circuits by means of electromagnetic induction.

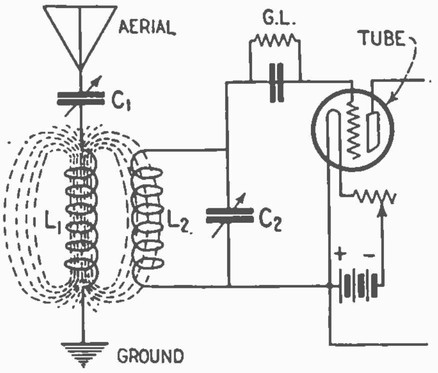

The utilization of a vario-coupler to inductively couple the aerial system with the rest of a receiving circuit.

The most common example of inductive coupling is in radio receiving circuits where coupling coils are used to associate the primary or aerial circuit with the secondary circuit. In the illustration L$_1$ is the primary of a vario-coupler and L$_2$ the secondary. The primary is inserted between the ground and aerial and the secondary is connected to the grid and filament of the detector tube. The variable condenser C$_1$ tunes the primary circuit, the primary coil also being variable by means of taps if desired. The variable condensed C$_2$ tunes the secondary inductance and the condenser and grid leak GL are the customary elements in the detector grid circuit. As the high frequency currents traverse the primary circuit, passing through coil L$_1$, lines of force are set up as shown and some of these lines embrace the secondary coil L$_2$, inducing similar currents in the secondary or vacuum tube circuit. The balance of the circuit is not shown for obvious reasons. This type of coupling between circuits is much more selective than the straight coupling or conductive coupling. (See *Coupling,* also *Conductive Coupling.*)

INDUCTIVE DISTURBANCE—The disturbing effects sometimes experienced in radio broadcast or telegraph reception due to induction from nearby power or telephone lines. This sometimes occurs when an aerial is strung close to and parallel with a line carrying high voltages, and occasionally is experienced with telephone wires near the receiver. (See *Induction.*)

INDUCTIVE DROP IN VOLTAGE—The drop or diminution of voltage in an alternating current circuit due to the presence of inductance. Inductance in a circuit carrying alternating current tends by reason of the self-induction to set up a counter-electromotive force which opposes the original E.M.F. (electromotive force), thus reducing the effective voltage in the line. (See *Inductance,* also *Inductance Reactance.*)

INDUCTIVE EMF—The electromotive force due to induction. That is to say, the electromotive force induced in a circuit. (See *Induction, Induced EMF.*)

INDUCTIVE REACTANCE—The reactance or portion of the total impedance in a circuit, due to the presence of self-inductance. (See *Inductance Reactance,* also *Impedance.*)

INDUCTIVE RESISTANCE—A resistance element possessing inductance, as for example a resistance coil, which also has the property of self-inductance. (See *Resistance,* also *Inductance.*)

INDUCTIVE RISE—The rise in voltage noted in transformers or other alternating current apparatus due to the presence of a *leading current* (q.v.). (See *Inductance,* also *Transformer* and *Angle of Lead or Lag.*)

INDUCTIVITY—Another term used alternately with dielectric constant and conductive capacity, referring to the dielectric properties of materials. (See *Dielectric Coefficient or Constant,* also *Inductive Capacity.*)

INDUCTOR—The coils or conductors in which current is induced through electromagnetic induction, as in an induction motor. Any conductor in which current is induced due to a change in magnetic flux may be called an inductor. (See *Induction, Inductor Alternator,* and *Induction Motor.*)

INDUCTOR ALTERNATOR—An alternator for producing high frequency currents. It employs the *armature* and *field windings* on projections inside the *stator,* the rotor consisting of a drum carrying a magnetic material. By using a *rotor* of solid steel, very high speeds and consequently high frequency currents are obtainable. The *Alexanderson Alternator* (q.v.) is an inductor type alternator. (See *Alternator,* also *Inductor.*)

INDUCTOMETER—A device or instrument used to measure self or mutual inductance. The basic principle depends on the relation of two coils, a primary and secondary, one of which can be moved in its relation to the other, the inductance, either self or mutual, being registered on a scale calibrated in units of inductance. (See *Inductance, Mutual,* also *Inductance, Self.*)

INERTIA—In physics, the property of a body which tends to keep it in a state of rest or to resist any change of motion. (See *Inertia, Electric.*)

INERTIA, ELECTRIC—Term occasionally applied in place of *self-inductance* (q.v.). When a current flows in a circuit possessing inductance and the circuit is broken, current continues to flow for a short time. It thus appears that in this instance the resistance alone will not stop the flow of current on the instant of breaking the circuit, an interval of time being required as in the act of bringing to rest a moving material. (See *Inertia, Inductance, Self,* also *Resistance.*)

INERTIA, ELECTROMAGNETIC—The same as electric inertia, being, in effect, the energy required to start or stop a current in a circuit possessing self-inductance. (See *Inertia.*)

INFERRED ZERO—A term used in connection with certain instruments having extremely high sensitivity, where the zero is removed from the scale. In a galvanometer, for example, the zero position on the scale may be merely assumed, an electrical or mechanical force being applied to bring the zero off the scale entirely, only a part of the full range of the full scale being utilized. (See *Galvanometer.*)

INFLUENCE—A broad term signifying action at a distance, as by electrostatic induction, without any actual physical contact. It is considered as the effect of a charged body on a neutral body or conductor coming within its field. (See *Field*, also *Induction.*)

INFRA-RED RAYS—The heat waves or rays which lie between light waves and electromagnetic (radio) waves. The infra-red lies at the extreme low frequency end of the invisible spectrum, the wave-length being extremely high and the frequency conversely low. (See *Wave*, also *Ether Waves.*)

INFRA-SATURATION PART OF CHARACTERISTIC—In a characteristic curve of a vacuum tube, this is the portion over which the tube is operable, or the operating portion of the curve. (See *Characteristic Curve*, also *Vacuum Tube, Theory of Operation.*)

INITIAL MAGNETIZATION—The initial stages of magnetization as distinguished from the effect after saturation has been approached or reached. (See *Saturation, Magnetic.*)

INPUT—Generally speaking, the energy absorbed by a machine as distinguished from the output energy. (See *Input Voltage*, etc.)

INPUT CIRCUIT OF VACUUM TUBE—Vacuum tube circuits are generally divided into three separate classes or circuits. The first is the circuit including the filament, i.e., the "A" battery circuit through the filament. The second is the circuit between filament and grid and the third is the circuit between filament and plate. The filament grid circuit is called the input circuit because the incoming oscillations are impressed on this circuit. The filament plate circuit is known as the output because the rectified or amplified currents are taken from across the plate and filament. The illustration shows both input and output circuits, the former in heavy lines and the latter in light lines. The input circuit of a tube is also known as the *grid circuit.* (See *Grid, Filament, Plate*, also *Vacuum Tube, Theory of Operation.*)

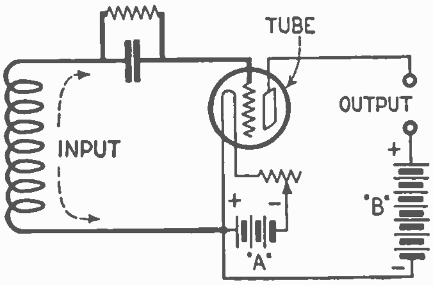

Input and output circuits of a vacuum tube receiver. Heavy lines show the former, light lines the latter.

INPUT VOLTAGE—The voltage impressed on a circuit or machine as distinguished from the output voltage. (See *Input.*)

INSTANTANEOUS CURRENT—The value of current in an alternating current circuit measured at any instant, as distinguished from the average current value. (See *Instantaneous Values.*)

INSTANTANEOUS VALUES—The actual value of an alternating current, E.M.F. (electromotive force), etc., measured at any instant. This value may be anything from zero to maximum or peak It must be remembered that in an alternating current the current and voltage are both variable factors, rising to a maximum in one direction, then falling to a minimum, and then rising to a maximum in the other direction. Thus, a wattmeter measures the actual value of current and voltage at each instant and the instantaneous torque at each instant is proportional to the instantaneous power. The average torque acting on the moving element during each cycle of the current and voltage is then proportional to the average power. If we measure an alternating current with an ammeter, and then measure the voltage with a volt-meter, the product of the two will be the apparent power, (apparent watts). It should be noted that these instruments measure effective and not instantaneous values. In order to ascertain the true *watts* (q.v.), it is necessary to multiply the apparent watts by the power factor. A wattmeter measures true watts. (See *Effective Electromotive Force, Power Factor*, also *Wattmeter.*)

INSTANTANEOUS VOLTAGE OR PRESSURE—The voltage or pressure in an alternating current circuit measured at any instant, as distinguished from the average value of E.M.F. (electromotive force). (See *Instantaneous Values.*)

INSTRUMENTS, MEASURING—In electricity or radio, any device or instrument for recording or indicating values, such as an ammeter for indicating the current in amperes, or a voltmeter to indicate the electromotive force in volts. (See *Ammeter* or *Ampere Meter, Voltmeter, Wattmeter.*)

INSTRUMENT SHUNTS—In electrical measuring instruments, a branch conductor, generally of relatively heavy wire or metal strip, joining the meter circuit at two points and forming a parallel circuit, thus dividing the current and protecting the delicate windings of the instrument. (See *Shunt*, also *Ammeter*, or *Ampere Meter, Hot Wire Ammeter.*)

INSULATE—To provide a protective covering or shield for a part or conductor, preventing flow of electricity from it or the passage of electricity through it. A conductor of electric currents is generally insulated to prevent the flow of the electricity through contact with other parts or conductors except where contact is intended to be made. (See *Insulated Wire.*)

INSULATED WIRE—Wire, generally copper, covered with insulating material to prevent leakage of the electric current. Standard insulated wire of the heavier gauges is usually covered with a layer of rubber and an outer covering of braided cotton. (See *Lead-In*, also *Insulation.*)

INSULATING COMPOUND—Liquid or easily liquifiable mixture of some insulating material, used in radio to coat cardboard tubes and to cover any open point of contact in an otherwise insulated conductor or part. Coils are very often coated with shellac or some other liquid and quick-drying substance. This serves to hold the windings in place and also acts as a protective covering. A very valuable application of an insulating liquid is shown in the illustration. Here an inductance is wound on a form of stiff paper or cardboard tube and after being completed is coated with one or two treatments of collodion or similar liquid. (Shellac introduces losses through the medium of *distributed capacity.*) The form is then cut away and the coil remains self-supporting as shown. This method produces a coil having the least possible dielectric and distributed capacity loss for a cylindrical shape. (See *Dielectric, Inductance Coil, Low Loss Coils, Collodion.*)

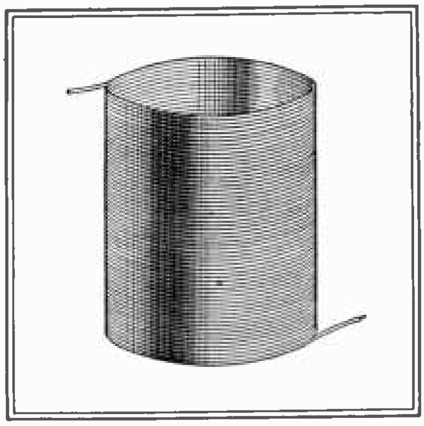

Air core coil made rigid by use of insulating compound.

INSULATING MATERIALS—Non-conducting substances, for example, glass, hard rubber or porcelain, which do not conduct electric currents. Strictly speaking, there are no perfect insulating materials. The most perfect insulators are certain gases, although at low pressure these may act as very good conductors. The more common insulating materials in radio are the following: mica, glass, bakelite and hard rubber, the insulating values being in that order. When we speak of the insulating property of a material we are actually referring to its resistance, a good insulator offering infinite resistance to the passage of electric currents and poor ones being relatively good conductors. The term *specific resistance* is used extensively in connection with the insulating properties of materials. This term is defined as the resistance of a piece of material of unit length and unit cross-sectional area at a given temperature, the value being given in ohms. (See *Dielectric Coefficient and Constant, Table of*, also *Ohms, Resistance* and *Insulate.*)

INSULATING PROPERTY—The ability of any material to *insulate* electric currents. Accepting mica as an insulator, we say that its insulating properties are superior to glass or hard rubber. (See *Insulating Materials*, also *Dielectric Coefficient and Constant* and *Insulation of Aerial.*)

INSULATION—Broadly any material or gas used to insulate electric currents. In radio, insulation plays a most important part, providing protective coverings or separators between conductors of the electric current and between the conductors and the ground. Insulation is really a material or gas having extremely high specific resistance; the better the insulating property, the higher the resistance to passage of electricity. There is no clearly defined line of demarcation between an insulator and a conductor, a poor insulator being to some extent a conductor. (See *Bakelite, Conductor, Dielectric Coefficient and Constant, Insulations of Aerial, Panel.*)

INSULATION OF AERIAL—The means used in connection with an aerial for reception or transmission of radio waves, to prevent leakage or ground-

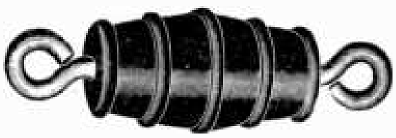

Aerial Insulator.

ing of the currents. The insulation of an aerial for reception purposes should be given very careful consideration. The wires may be suspended by means of small glass or porcelain insulators and the lead-in rubber covered wire, entering the building through a *lead-in insulator* (q.v.). The aerial for transmission must be accorded specially careful attention as far as insulation is concerned. The bare wires must be suspended with heavy duty insulators, capable of withstanding high voltages without *"breaking down."* The lead wire must be conducted into the building through a high-tension insulating tube or other lead-in device to prevent leakage of the current to the ground. (See *Aerial, Insulation of Aerial* and *High Tension Insulators,* also *Lead-In Insulator.*)

INSULATION RESISTANCE—The actual resistance in ohms, generally megohms, of the insulation of any conductor or piece of apparatus. If we assume two conductors, carrying separate currents and crossed so that the only protection would be the actual insulation of the wires, the insulation resistance would be the actual resistance in megohms from one path to the other. Conceivably a small current might leak from one conductor to the other, this current, of course, depending on the resistance of the insulation. The term "insulation resistance" is more easily understood when it is considered that an insulator or insulating material is merely a very poor conductor, or one having extremely high resistance to electric currents, there being actually no perfect insulator. (See *Insulation, Insulated Wire,* also *Conductor* and *Resistance.*)

INSULATION TESTER—An instrument or set of meters for testing the resistance of an insulator in fractional ohms, generally megohms. (See *Insulation Resistance.*)

INSULATOR—Any substance having insulating properties. *Insulators* are used to keep electric currents flowing in predetermined paths. It is necessary to use insulators on all electric circuits in order to prevent *grounds, leakage* and *short circuits.* Insulators are of the utmost importance in radio. At the transmitting end, where high voltages as well as high frequencies are used, the insulators must be specially designed to prevent leakage. A typical installation of insulators in use on the aerial system of a modern broadcasting station utilizes a 12-inch pyrex insulator at each end of each wire between the spreaders. A standard porcelain strain insulator is used every 14 feet on each guy wire. At each end of each guy wire a standard 24-inch porcelain insulator is used.

Insulators used on receiving antennas are made of porcelain, pyrex, rubber and composition. These are usually about 4 inches in length. Sometimes porcelain tubes are used as insulators to bring a *lead-in* (q.v.) into a building, but these have been generally superseded by flat *window lead-ins.*

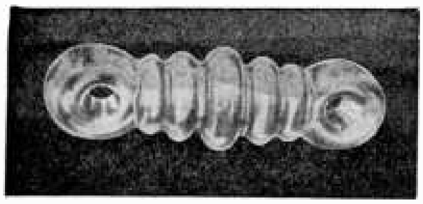

Illustration by courtesy of Corning Glass Works.

A typical small antenna insulator, showing corrugated surface.

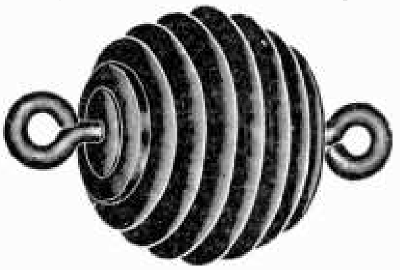

A corrugated ball insulator.

INSULATOR, LEAD IN—The insulating tube through which a lead wire from a radio aerial enters a building, used to prevent grounding or leakage of the current by contact with another conducting surface. (See *Lead-In Insulator.*)

A flat window lead-in which permits entrance of lead-in into building without the necessity of boring holes.

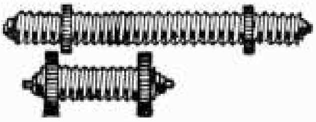

Lead-in insulators of conventional type.

INTEGRATING DETECTOR—The name sometimes given to any detector which yields a response proportional to the total energy received from a spark train rather than to the maximum value of the current or voltage in the train.

INTEGRATING INSTRUMENTS—Meters so designed as to record the total quantity of electricity (ampere hours) or the total energy (watt-hours) passing in a circuit in a given time. (See *Ammeter* or *Ampere Meter,* also *Integrating Wattmeter.*)

INTEGRATING WATTMETER—An instrument which measures and records the total amount of electrical energy being consumed in a circuit. These instruments are also known as *recording watthour meters* (q.v.), *integrating watthour meters,* and sometimes are simply referred to as *watthour meters* (q.v.).

Broadly speaking the *integrating wattmeter* is a type of small motor whose rotations are counted by means of a worm on the armature shaft engaging a set of cogs working a counter. The construction is such that the average *torque* (q.v.) exerted by the motor is proportional to the average power taken by the load. In other words, the motor rotates at a speed directly proportional to the energy being expended. For direct current work, *integrating wattmeters* are either of the commutator type, such as the Thompson, Westinghouse and Duncan meters or of the no-commutator type, such as the Sangamo. The commutator type instruments operate on the *dynamometer* (qv.) principle. Meters of the no-commutator type are essentially motors whose armatures always cut the same direction flux. That is to say, they are essentially *homopolar* (q.v.) motors. The principle of operation of these meters is the law of motor action which states that a conductor, free to move and carrying a current whose direction of flow is at right angles to a fixed magnetic field, will be moved out of this field, and in a direction at right angles both to the direction of current flow and to the field. For alternating current work, *integrating wattmeters* are of the *induction* (q.v.) type. The principle of operation of these meters is identical to that of *induction motors* (q.v) having shunt and series windings stationary and so related and located as to produce a rotating field acting upon a closed rotable secondary. In the induction meter the secondary consists of a light aluminum disc. These meters may be either of *single phase* (q.v.) or *polyphase* (q.v.) construction. (See *Electro-dynamometer.*)

INTEGRATOR—A device which automatically, by means of clockwork, adds up and records on a dial items of calculation or measurement. The system used in a recording or integrating wattmeter. (See *Integrating Wattmeter.*)

INTENSIFICATION—Broadly an intensifying or increasing of the density of electric current, but specifically in radio it refers to the tendency of radio signals to increase in intensity under certain conditions. The phenomena of intensification is naturally closely related to fading, as obviously the intensification process must necessarily follow or be followed by a fading process. The tendency toward intensification is more pronounced toward night and generally on the shorter wavelengths, it being noted that quite often signals from a distant station will be increased in volume or intensity as much as fifty per cent or more between the late afternoon and late evening. While the periodic *fading* of signals and the subsequent increase in volume over regular periods, usually of a few minutes' duration, may be considered as *fading* and *intensification,* it is preferable to think of fading as one operation together with the subsequent increase in volume, and the term *intensification* as referring strictly to the change to greater volume that takes place toward night. (See *Fading,* also *Phenomena of Electric Wave Propagation.*)

INTENSITY OF ELECTRIC CURRENT—A term more or less in use to imply the strength of an electric current in amperes, the symbol I for current being derived from it. (See *Ampere, Current Density,* also *Intensity of Field.*)

INTENSITY OF FIELD—The intensity or strength of the force exerted in a magnetic field. It is measured by its action or effect on a unit pole placed at any point in the field, the intensity of the field, of course, varying at different points. (See *Flux Density, Intensity of Magnetic Flux.*)

INTENSITY OF MAGNETIC FLUX—The strength of the force exerted in a magnetic field. (See *Intensity of Field,* also *Flux Density.*)

INTENSITY OF MAGNETIZATION—The density of force per unit cross-sectional area of a magnetized material. The extent of the magnetizing effect on a material placed in a magnetic field of force. This applies only to material in a magnetic circuit and should not be confused with the lines of force due to a field arising from the presence of neighboring currents or magnets. (See *Magnetization,* also *Intensity of Magnetic Flux.*)

INTERFERENCE—The interruption or interference with desired electromagnetic waves by undesired or extraneous waves. The interference of undesired broadcast or radio-telegraph signals with the desired signals. Also, broadly, the detrimental effect of power lines and other electric circuits on radio communication. Interference is encountered chiefly in receiving sets having poor *selectivity* (q.v.). In some instances interference may be due to the fact that another transmitter is operating on the desired wave-length, and being sufficiently close as to make it practically impossible to obtain clear signals from the desired station. Ordinarily, and where the effect is due to lack of sharp resonance or tuning, certain changes can be made in the receiver to sharpen the tuning, or if the disturbance is due to the transmitting stations and the receiver is normally sharp, various circuits can be employed to filter out the undesired signals. (See *Interference Preventer*, also *Wave Trap.*)

INTERFERENCE ELIMINATOR—A device or system as shown in the illustration for reducing or preventing reception of undesired signals in radio work. Practically all *interference eliminators* make use of a separately tuned inductance across the aerial and ground and in parallel with the receiving set. The inductance is usually tuned by means of a variable condenser which allows it to be tuned to the wave-length of the interfering signal. The interference can thus be by-passed to the ground. *Interference eliminators* of the kind described are generally known as *wave traps* (q.v.). A special form of interference due to atmospheric and electrical disturbances is known as *static* (q.v.). (See *Static Eliminator.*)

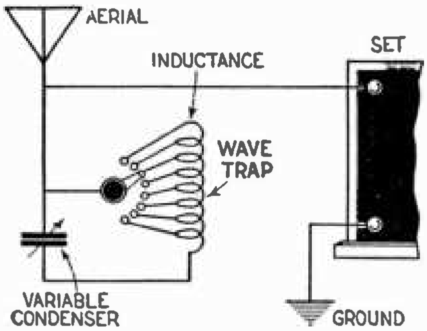

A wave trap or interference eliminator which dispenses with the ground connection.

INTERFERENCE PREVENTERS—Any arrangements either within the receiving set, or external to the set, for limiting or preventing interference. In addition to *interference eliminators* (q.v.) and static *eliminators* (q.v.) various other methods are utilized in radio to prevent interference. Thus the arrangement of a receiving set to give sharp tuning and great selectivity, results in the prevention of much interference. A method of preventing interference which has been used, is the utilization of filter circuits which can be tuned to pass any desired frequency and exclude all others. The use of the *loop aerial* (q.v.) results in great selectivity. The loop has a very pronounced directional effect and this in itself acts to narrow the field of stations offering possible interference. For example, if stations of very nearly the same wave length are in opposite or different directions and are sending at the same time, the loop permits selection of the station wanted, with no interference from the others.

INTERMEDIATE CIRCUIT—A closed circuit consisting of two inductance coils shunted by a variable condenser. One coil acts as primary to the secondary coil of the detector circuit and the other coil acts as secondary to the primary of the aerial circuit. A circuit such as this is sometimes referred to as a *multiple tuner intermediate* circuit and may be used as an interference eliminator.

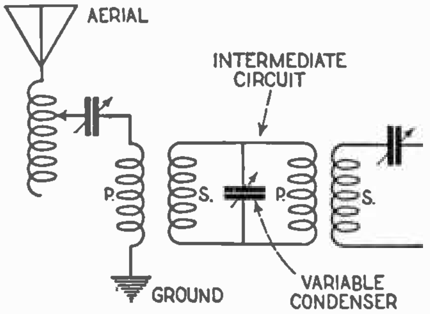

A multiple tuner intermediate circuit.

INTERMEDIATE FREQUENCY—This is a frequency higher in number of oscillations than *audio frequency* (q.v.) but lower than *radio frequency* (q.v.). It is generally around 30,000 cycles or 10,000 meters wave-length. In the *super-heterodyne* (q.v.) the incoming signal is converted from the broadcast frequency to the so-called *intermediate frequency* at which it can be amplified by means of fixed winding transformers to almost any extent desired. The method of converting the incoming signal in a *super-heterodyne* from radio frequency to *intermediate frequency* is by the *beat* method. In this case an oscillating circuit controlled by means of a variable condenser so as to obtain any desired frequency, is brought into interference with the first detector and produces a *beat note* (q.v.) for the *intermediate frequency* which has all the characteristics of the original signal. In this way, no matter what the frequency of the intercepted signal may be, the intermediate frequency amplification is always carried on at a fixed wave-length, thus making for high efficiency. (See *Beat Frequency; Beats.*)

INTERMEMEDIATE FREQUENCY AMPLIFIER—An arrangement of tubes and transformers for stepping up the current, used in the *super-heterodyne* (q.v.) circuit. The intermediate frequency amplifier is located between the first detector and the second detector. (See *Intermediate Frequency.*)

INTERMEDIATE TRANSFORMER—The name given to the transformers used in the *intermediate frequency amplifier* (q.v.) of the *super-heterodyne* circuit. *Intermediate transformers* are often referred to as *intermediate frequency transformers.* They are *radio frequency transformers* (q.v.) but have an iron core and instead of tuning to a maximum of 550 or 600 meters they are designed to cover wave-lengths of 10,000 meters or higher. (See *Intermediate Frequency.*)

INTERMITTENT CURRENT—A current which flows irregularly or which is interrupted at intervals, or without continuity. The current obtained from a *magneto generator* (q.v.) is an interrupted or *intermittent current.* Such a current may be either alternating or direct. A magneto generator constructed to give direct current is equipped with a two segment commutator having a stationary brush in contact with it. (See *Pulsating Current.*)

INTERNAL RESISTANCE—Resistance within an electric source. In the case of a primary cell such as a dry battery, the internal resistance causes an interference with the flow of current. This varies in amount with the construction and materials of the battery.

INTERNATIONAL MORSE CODE—This is also known as the *Continental code.* (See *Code.*)

INTERPOLE—A small magnetic pole placed between the main field poles of an electric generator for the purpose of obtaining better commutation. Interpole windings are connected in series with the armature winding and the load. The action of the interpole is exactly analogous to the shifting of the brushes, but when interpoles are used, brush shifting is dispensed with and the magnetic flux is shifted instead.

INTERRUPTED CONTINUOUS WAVES—abbreviation I.C.W.—These waves are obtained by the modulation at audio frequency, during signalling, of an otherwise *continuous wave* (q.v.). (See *Continuous Waves Key Modulated*, also *Continuous Waves Modulated at Audio Frequency.*)

INTERRUPTED WAVES—abbreviation I.W.—Interrupted waves are waves produced by modulation at audio frequency, of otherwise continuous waves. (See *Interrupted Continuous Waves.*)

INTERRUPTER—A combination of an electro-magnet with a vibrating armature which carries a contact point on a piece of spring steel fastened to the armature. When the electro-magnet attracts the armature the contact points are drawn apart thus interrupting the circuit. This action is only momentary for the spring again brings the circuit back to its original condition. The most important use of the interrupter in radio is as a *rectifying valve* (q.v.) in in mechanical chargers. A typical vibrator panel is shown in the illustration. The armature, A, is the moving element. This carries at its outermost end a heavy Tungsten contact, B, which is caused to vibrate in synchronism with the supply current, by the actuating coil, C. When the current from the transformer is of the correct polarity and value the actuating coil attracts the armature, closing the charging circuit to the battery through the fixed contact, D. The contact end of the armature, A, vibrates within the air gap of the pole shoe, G.

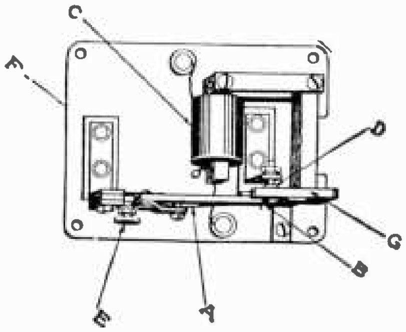

The vibrator panel of an efficient type of mechanical charger.

A common use of the *interrupter* is in the ordinary electric bell. In this apparatus, the electromagnet pulls the armature carrying the spring and contact point, breaking contact and opening the circuit. This immediately de-energizes the magnet since its coils are in series with the circuit. The spring tension pulls the armature back to its original position thus again closing the contact. As soon as the contact is closed, the electromagnet is again energized and it again attracts the armature, opening the circuit. This action keeps on indefinitely as long as a difference of potential is maintained at the terminals of the bell.

The *interrupter* is used in the *induction coil* (q.v.) for breaking up the direct current into a series of impulses, thus producing an intermittent current. (See *Circuit Breaker.*)

INTERRUPTER, ELECTROLYTIC—A jar containing diluted sulphuric acid as the electrolyte, a large sheet of lead as one electrode and a platinum needle point introduced through a glass tube as the other electrode. If these electrodes are connected through an inductance to a source of current supply, the current in the circuit will be rapidly interrupted. *Electrolytic interrupters* are suitable for radio transmitters using small spark coils and are also used in X-ray work. (See *Electrolytic Interrupter.*)

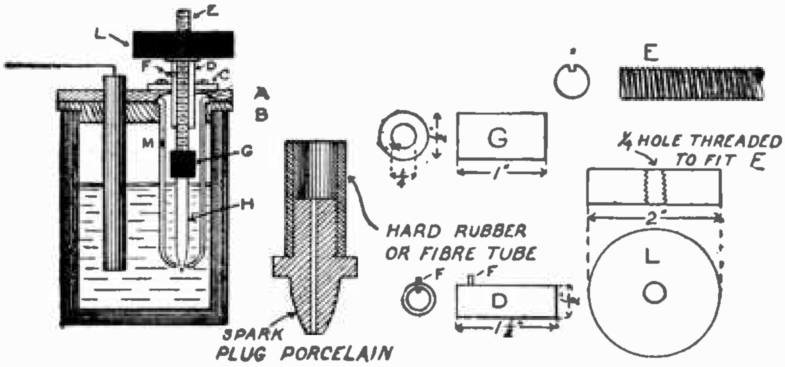

A simple type of Wehnelt interrupter. The positive pole of the direct current supply is connected to the brass part C, while the negative pole is attached to the lead rod. L is a fibre handle; E is a threaded rod, which is constructed to fit the threaded handle, but slotted so as to engage the pin F and prevent it from turning; H is a copper or platinum rod; D is a brass tube; M. is the overflow hole in the fibre tube, into the lower end of which a portion of an old spark plug porcelain part has been driven; finally C is a small brass block as shown. This is the adjustable type of Wehnelt interrupter, since the fibre handle can be operated to allow more or less surface to be exposed to the acidulated water.

INVERTED "L" AERIAL—A flat top aerial (antenna) in which the down lead is tapped off one end of its horizontal span. (See *Length of Aerials, Horizontal Aerial.*)

INVERSE DUPLEX CIRCUIT—A circuit in which the vacuum tubes are utilized both for radio frequency and for audio frequency amplification as in the *reflex* (q.v.) circuit with this important differentiation that the first radio frequency tube is utilized as the second audio frequency tube and the second radio frequency tube is used as the first audio frequency tube. In the ordinary reflex circuit the first radio frequency tube is also the first audio tube and the second radio frequency tube is the second audio tube. The *inverse duplex* circuit has the advantage over the reflex in that the tubes which are reflexed operate more efficiently since the relatively weak radio currents pass through the same tube as the stronger audio currents and vice-versa. The *inverse duplex circuit* is the invention of David Grimes.

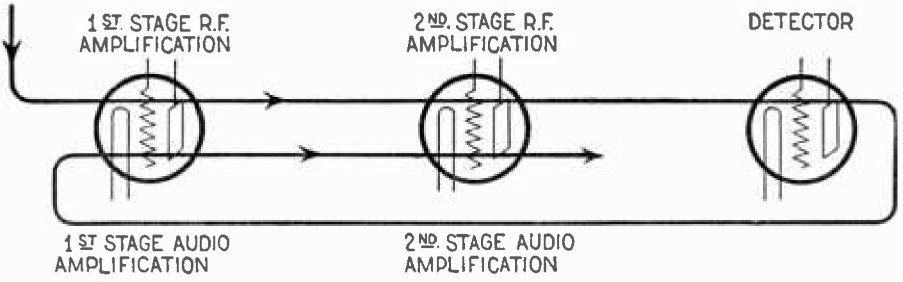
1ST. STAGE R.F. AMPLIFICATION
2ND. STAGE R.F. AMPLIFICATION
DETECTOR
1ST STAGE AUDIO AMPLIFICATION
2ND. STAGE AUDIO AMPLIFICATION

Diagram showing the current path and reflexing in an ordinary reflex circuit.

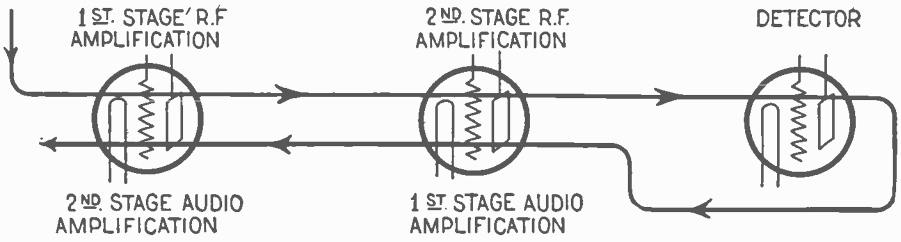
1ST. STAGE R.F AMPLIFICATION
2ND. STAGE R.F. AMPLIFICATION
DETECTOR
2ND. STAGE AUDIO AMPLIFICATION
1ST. STAGE AUDIO AMPLIFICATION

Diagram showing method of reflexing in an inverse duplex circuit.

ION—An atom of matter carrying an *electron* or an *atom* deprived of electrons. *Monad-ion* refers to a unit charge, *dyad-ion, divalent* carries two units, the *triad-ion* carries three unit charges. The positive *ion* is an *atom* minus an *electron.* A negative *ion* is an *atom* plus an *electron.* (See *Anion,* also *Cathion, Electron, Electron Theory.*)

IONIZATION—sometimes spelled **IONISATION**—The splitting up of *molecules* into *ions* (q.v.). *Ionization* may apply either to a chemical compound or to a gas. When a electric current is passed through an *electrolyte* (q.v.) the constituents of the electrolyte which are liberated or deposited at the *electrodes* (q.v.) are called *ions* and this process is called *ionization.* Gases at or near atmospheric pressure are normally good insulators. They are said to undergo *ionization* when they are made conductive, that is when they allow the flow of electricity. A gas may be made conductive by passing *X-rays* (q.v.) through it, by the action of ultra-violet light, by subjecting it to radium rays, by passing it close to a flame or hot piece of metal or by forming in it a brush discharge or *corona* (q.v.). Gases which have been made conductive carry current because of their dissociation into *ions,* which carry positive or negative charges of electricity. (See *Electrolysis, Electrolytic Action,* also *Valve Detectors.*)

I.R. DROP.—The voltage drop due to the resistance in a current-carrying conductor. It is directly proportional to the current flowing in the circuit multiplied by the resistance of that circuit.

I^2R LOSS—The power loss in any circuit due to the resistance offered to current flow. It is proportional to the square of the current flowing in the circuit multiplied by the resistance of the circuit. Whenever current flows there is an I^2R loss. This is measured in *watts* (q.v.) or *kilowatts* (q.v.).

IRON PYRITES—Formula FeS_2—This is a mineral, a disulphide of iron, found in large quantities. It was formerly used in radio as a crystal detector mineral but has been almost entirely superseded by vacuum tube detectors. *Iron pyrites* is flaked with shiny spots which have the appearance of gold, and for this reason it is sometimes called "fools gold." *Iron pyrites* is very hard and brittle. The present commercial use of *iron pyrites* is in the production of sulphuric acid. Ferrous Sulphide, FeS, has also been employed in the past as a radio detector. This substance is chemically formed by fusing equivalent quantities of sulphur and iron together. The resultant substance has a smooth surface. (See *Crystal Detector.*)

ISOCHRONOUS—A term applied to two or more oscillatory circuits, meaning that they have the same *natural frequency* (q.v.). Stated in another way, two or more radio frequency circuits are isochronous when they are in electrical *resonance* (q.v.). Circuits are said to be in resonance when they have the same *oscillation constant* (q.v.). The *oscillation constant* of a circuit is equal to the square root of the product of the inductance multiplied by the capacity of that circuit. Circuits having equal oscillation constants will have the same discharge frequency.

ISODYNAMIC LINES—Lines on a magnetic map connecting all parts of the earth where the magnetic intensity is the same. In other words, these lines pass through points of equal horizontal component of the earths magnetic field.

ISOGONIC LINES—Lines on a magnetic map passing through points of equal *declination.* Lines connecting all points of equal variation of the magnetic meridian from the geographic meridian.

ISOTHERMS—Lines on a meteorological chart joining all points having the same temperature.

ISOTROPIC CONDUCTIVITY—Equal conductivity in every direction. A conductor is said to be *isotropic* when it offers the same resistance to the flow of an electric current in every direction through its mass.

J

JACK—An arrangement of spring terminals as shown in the illustrations which can be connected together or separated thus closing or opening circuits through the insertion of a *plug* which also forms a part of the circuit. The *plug* (q.v.) has a tip and a sleeve at one end (that is at the end inserted into the jack) and at the other end it has the two terminals for connecting the apparatus to be cut into the circuit. As an example, the terminals of the plug may be connected to a *loud speaker* (q.v.). The jack in this case is of the *open circuit* variety. That is to say, the springs of the jack are separated when not in use and the circuit is normally open. When the plug is inserted, the tip of the plug makes contact with one spring and the sleeve makes contact with the other thus closing the circuit through the *loud speaker*. In cases where the plug is used to open a circuit, that is to make the *jack* springs break contact with each other, the *jack* is called a *closed circuit jack* since its spring terminals are normally making contact. In cases where the jack is used not only to put the loud speaker in the circuit, but also to act as a filament switch, it is known as an *automatic filament control jack*. Jacks which are used to control a single circuit only, are referred to as *single circuit jacks*. Those controlling two or more circuits simultaneously are known as *multiple circuit jacks*. *Jacks* are usually mounted on a panel so that the plug can be inserted from the outside of the panel while the spring terminals are concealed behind the panel.

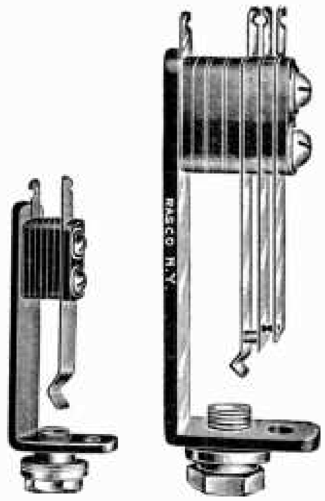

Illustration by courtesy of Radio Specialty Co.

Illustration at left shows one-spring open circuit jack. At the right is a three-spring jack.

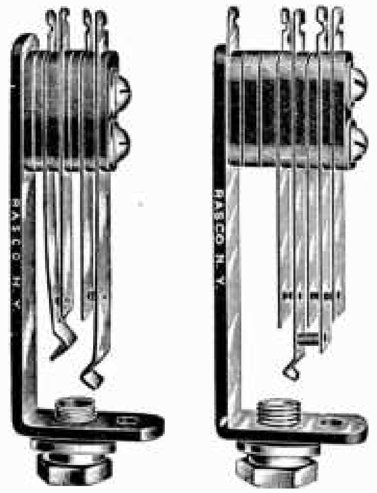

Illustration by courtesy of Radio Specialty Co.

The jack shown at the left is a four-spring double circuit jack. At the right is a five-spring automatic jack.

JAMMING—An expression denoting interference at a receiving station caused by a transmitting station not desired to be heard. More specifically, such interference, prevents the reception of signals, *jamming* the receiving station by drowning out with stronger signals or noises the station desired.

JANET, PAUL—Born in Paris, January 10, 1863. Educated at Lycee Louis-le-Grand and at the high school. He was Professor of Physics at the University of Paris and University of Grenoble. Professor Janet was the author of several important works, and first to make the experiments in electrical resonance successful. The results of these experiments are applied in present day wave meters.

JAR—A unit of electrostatic capacity. It is mainly used in the British Naval service. One *jar* equals 1000 *centimeters* which equals .0011 *microfarads* (q.v.).

JIGGER—A term generally accepted to denote an *oscillation transformer* (q.v.). Used for transforming trains of oscillations from one circuit to another. Dr. J. Erskine Murray refers to high frequency oscillations as *Jigs*. *Jiggers* are also known as *magnetic coupling transformers* (q.v.). They are also sometimes referred to as *auto-jiggers*.

JOULE—symbol J—The practical unit of electrical energy. It is equal to 10^7 *ergs* (q.v.). The *joule* is also the unit of electrical work. It is equal to the work done or the heat generated by a *watt second* (q.v.); or an *ampere* (q.v.) flowing for a second through a resistance of one *ohm* (q.v.). Expressed another way the *joule* is the work done by one *coulomb* (q.v.) flowing under the pressure of one *volt* (q.v.). One *joule* is equal to .73732 foot pounds, or .24 calories (q.v.).

JOULE EFFECT, or JOULEAN HEATTING—This occurs when an electric current traverses a resisting conductor. The rate of generation of heat in the conductor by a steady current equals the square of the current multiplied by the resistance.

JOULES EQUIVALENT—The amount of energy equal to a heat unit.

JOULES LAW—The heat produced in an electric circuit is directly proportional to the square of the current, to the resistance of the conductor and to the time of current flow.

K

KANALSTRAHLEN—A German word meaning Rays of Canal. It has been found that by perforating the cathode of a tube producing *cathode rays* (q.v.), faint luminous streaks will come through these perforations in a direction opposite to that of the *cathode* rays. These are called *kanalstrahlen*. They communicate a positive charge to an insulated conductor. In terms of the *electron theory* (q.v.) they are positively charged *ions* (q.v.) The cathode particle is an electron traveling in one direction, and the kanalstrahlen particle traveling in the opposite direction is what remains of the atom which has lost an electron.

KATHODE—This is a negative *electrode* (q.v.). (See *Cathode*, also *Electrolysis*.)

KATHODE OF CELL—The positive pole of a cell. This is indicated by the plus (+) sign. (See *cell*.)

KATHODE RAYS—The stream of *electrons* (q.v.) thrown off by the *kathode* of a vacuum tube. These produce a glow when they strike the walls forming the tube. They can be deflected by a magnet. (See *Cathode Rays*.)

KATION—The charged particles or *ions* (q.v.) moving in the direction of the *cathode* (q.v.). (See *Cathion*.)

KEEPER—The piece of iron used to close a magnetic circuit to protect it from external disturbances. The term *keeper* is generally used to refer to such a piece of iron when used with a permanent magnet such as a horseshoe magnet. In the case of the piece of iron which closes the magnetic circuit of an electromagnet, this is more often referred to as an *armature* (q.v.). The *keeper* may be used with a straight bar magnet also. In this case, two bar magnets are laid side by side with unlike poles adjacent and two keepers are used, one at each end. The use of a keeper avoids the demagnetizing effect of *leakage lines* (q.v.). (See *Horseshoe Magnet*.)

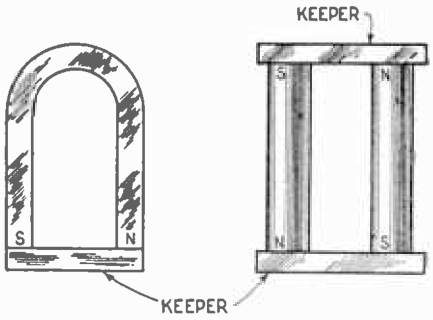

At left is shown keeper used with a horseshoe magnet. At right two keepers are shown used with two bar magnets.

KELVIN, LORD (William Thomson) was born at Belfast in 1824; he entered Glasgow University as a student at 10 years of age; graduated at Cambridge in 1845. When only 22 years of age he was called to occupy the chair of natural philosophy in the University of Glasgow, a chair which he adorned by his genius for fifty years. He was for 40 years or more regarded as the acknowledged leader of British science on its physical and mathematical side. His great inventions in telegraphy, his magnetic compass and sea sounding apparatus brought him fame and fortune. He was knighted in 1866; and in 1892 was raised to the peerage as Baron Kelvin of Largs; he died on December 17, 1907.

Lord Kelvin (William Thomson)

KELVIN'S BRIDGE—A method of measuring low resistance in which the voltage drops, produced by the same current in the resistance under test and in a standard low resistance slide wire are balanced against each other. This *bridge* is a modification of the *Wheatstone Bridge* (q.v.) and was designed by Lord Kelvin to eliminate the errors introduced when measuring resistances much less than one *ohm* (q.v.). In the case of the *Wheatstone Bridge*, such resistances could not be measured accurately on account of the errors produced by the terminal and contact resistances.

KELVIN'S ELECTROSTATIC VOLTMETER—A voltmeter used for measuring high, and in some cases low, alternating current voltages. It is constructed on the principle of an air condenser. One type of electrostatic voltmeter is a high potential instrument having the needle made of a thin aluminum plate suspended vertically on delicate knife edges, with a pointer extending from the upper part to a scale. Two quadrant plates, metallically connected together, are placed on either side of the needle and parallel to its face. These serve as one terminal of the circuit to be measured. The needle acts as the opposite terminal. When there is a difference of potential between the needle and the plates, the needle is deflected out of its neutral position. The value of the scale indications can be changed by hanging calibrated weights on the bottom of the needle.

KENOTRON—A type of *vacuum tube rectifier* (q.v.) in which the vacuum is extremely high and the discharge is carried almost entirely by electrons, not by gas ions. In these tubes, the plate current is always less than that actually emitted by the hot filament. *Kenotrons* are made sufficiently large to rectify several kilowatts.

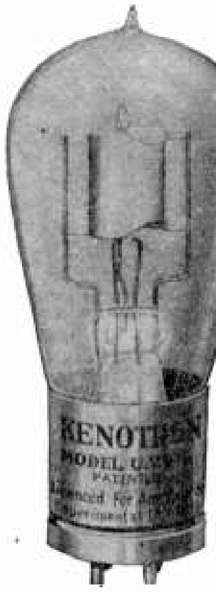

Illustration by courtesy of Radio Corp. of America
A kenotron tube.

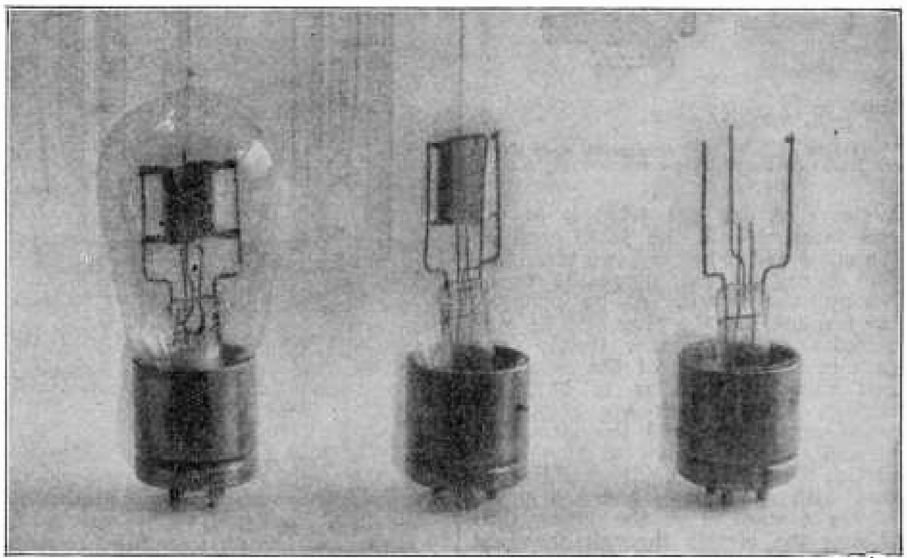
Illustration by courtesy of Radio Corporation of America.
Stages in the construction of a kenotron.

KENNELLY, A. E.—Anglo-American radio expert. Born at Colaba, Bombay, December 17, 1861, he was educated in England, Belgium, France and Italy. In 1875 he became a telegraph operator in the employ of the Eastern Telegraph Company, and in 1886 became the principal electrical assistant to Thomas Alva Edison in the laboratories at Orange, N. J., a post he held until 1892.

He was engineer-in-chief with E. J. Houston, of the Thomson-Houston Company, for the laying of the cables from Vera Cruz to Campeche, 1902. Since 1902 he has been Professor of Electrical Engineering at Harvard University, and since 1914 professor of the same subject at Massachusetts Institute of Technology.

Kennelly has written a large number of books on electricity and radio, and is the author of one of the standard elementary text-books on same. He has written a very large number of papers on radio, and he is an authority on alternating currents.

He is past president of the American Institute of Electrical Engineers, was president in 1916 of the Institute of Radio Engineers, and vice-president of the International Electrical Congresses, held in Paris and Turin. He is a member of many scientific societies, and has received many honorary degrees. In 1921 he was appointed a delegate to the Interallied Radio Technical Committee in Paris.

KEY—A form of switch for conveniently and quickly opening and closing a transmitting circuit, in the act of transmitting signals. An *operating*, or *telegraph*, so-called *Morse key* is a form of *tapping key* designed especially for the rapid sending of signals by the dot and dash code. A tapping key has one contact carried by a springy strip of metal. In signalling, this key is depressed by the fingers, thus closing the circuit by bringing the contact together with a fixed contact. A standard type of *transmitting key* (q.v.) such as used in sending wireless messages is shown in the illustration. These keys must necessarily be made with larger contacts and of more solid construction than those used to interrupt the smaller currents used in wire telegraphy.

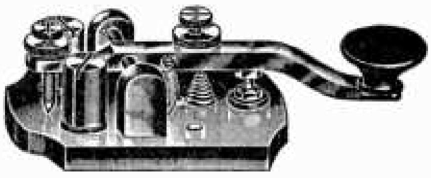
A standard transmitting key.

KEY, ANTI-CAPACITY—This is in reality a form of switch. (See *Switch.*)

KEY, HIGH-SPEED—A transmitting key usually operated from side to side instead of downwards, which makes

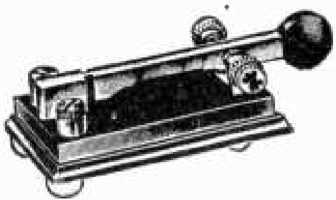
High speed key.

contact both on the right and on the left side. Such a key requires only half the travel of the ordinary transmitting key and thus facilitates high-speed manual transmission of code signals.

KEY SWITCH—A type of switch adapted for operation by a removable portion in the form of a handle or key. In radio work, *key switches* are sometimes used to lock a receiving set so that it cannot be operated unless the filament circuit is closed by the insertion and turning of the key. Key switches are generally operated by a hollow slotted key, although there are other types on the market. The slot in the key fits a projection or pin on the spindle or shaft of the switch which carries the movable contact. When the key is inserted in the socket and turned, it turns the shaft and brings the movable contact against a fixed contact. *Key switches* are also used for turning on or off electric lights and for locking or unlocking the ignition systems of automobiles.

KILO—Symbol K—A prefix placed before the name of a unit to indicate the multiple 1000 or 10^3. In cases where standard units are too small for convenient use, it is customary to utilize this prefix as, for example, in stating the voltage of a high tension transmission line. This would often be given as 10 kilovolts (K.V.) rather than as 10,000 volts. In like manner, in referring to large amounts of power, it would be preferable to give the amounts in kilowatts rather than in watts. Thus the better way to express 10,000,000 watts would be to divide by 1,000, thus converting to 10,000 kilowatts.

KILOAMPERE-TURN—One kiloampere-turn is equivalent to 1,000 ampere-turns. In referring to the magnetic circuits of electrical machinery it is usual to specify the magnetomotive-forces in kiloampere-turns rather than in ampere-turns. (See *ampere-turns*, also *magnetomotive-force.*)

KILOCYCLE—1,000 *cycles*—Two immediately succeeding half waves of an alternating current constitute a *cycle* (q.v.). The number of *cycles* per second is called the *frequency* (q.v.). In house-lighting and similar circuits, where a comparatively low frequency is used, it is usual to refer to the *frequency* in cycles per second. However, in radio circuits, where currents as high as 300,000,000 cycles per second are used, it is preferable to divide by 1,000, thus converting to *kilocycles.* If the wave-length in meters is known, it is a simple matter to determine the frequency in *kilocycles.* This is calculated as follows: The speed of light is 186,000 miles per second and experiments have shown that this is practically the same as the velocity of electromagnetic or radio waves. Expressed in meters this is equal to 300,000,000 (appx.) meters per second. Assuming that a transmitting station is using a wave-length of 500 meters, the frequency will be obtained in cycles per second by dividing the velocity of the waves by the

KILOCYCLES TO METERS, OR METERS TO KILOCYCLES

Meters	Kilocycles	Meters	Kilocycles	Meters	Kilocycles	Meters	Kilocycles	Meters	Kilocycles	Meters	Kilocycles	Meters	Kilocycles
10	29980	720	416.4	1430	209.7	2140	140.1	2850	105.2	4120	72.77	6400	46.85
20	14990	730	410.7	1440	208.2	2150	139.5	2860	104.8	4140	72.42	6450	46.48
30	9994	740	405.2	1450	206.8	2160	138.8	2870	104.5	4160	72.07	6500	46.13
40	7496	750	399.8	1460	205.4	2170	138.1	2880	104.1	4180	71.73	6550	45.77
50	5996	760	394.5	1470	204.0	2180	137.5	2890	103.7	4200	71.39	6600	45.43
60	4997	770	389.4	1480	202.6	2190	136.9	2900	103.4	4220	71.05	6650	45.09
70	4283	780	384.4	1490	201.2	2200	136.3	2910	103.0	4240	70.71	6700	44.75
80	3748	790	379.5	1500	199.9	2210	135.7	2920	102.7	4260	70.38	6750	44.42
90	3331	800	374.8	1510	198.6	2220	135.1	2930	102.3	4280	70.05	6800	55.09
100	2998	810	370.2	1520	197.2	2230	134.4	2940	102.0	4300	69.73	6850	43.77
110	2726	820	365.6	1530	196.0	2240	133.8	2950	101.6	4320	69.40	6900	43.45
120	2499	830	361.2	1540	194.7	2250	133.3	2960	101.3	4340	69.08	6950	43.14
130	2306	840	356.9	1550	193.4	2260	132.7	2970	100.9	4360	68.77	7000	42.83
140	2142	850	362.7	1560	192.2	2270	132.1	2980	100.6	4380	68.45	7050	42.53
150	1999	860	348.6	1570	191.0	2280	131.5	2990	100.3	4400	68.14	7100	42.23
160	1874	870	344.6	1580	189.9	2290	130.9	3000	99.94	4420	67.83	7150	41.93
170	1764	880	340.7	1590	188.6	2300	130.4	3020	99.28	4440	67.53	7200	41.64
180	1666	890	336.9	1600	187.4	2310	129.8	3040	98.62	4460	67.22	7250	41.35
190	1578	900	333.1	1610	186.2	2320	129.2	3060	97.98	4480	66.91	7300	41.07
200	1499	910	329.5	1620	185.1	2330	128.7	3080	97.34	4500	66.63	7350	40.79
210	1428	920	325.9	1630	183.9	2340	128.1	3100	96.72	4520	66.33	7400	40.52
220	1363	930	322.4	1640	182.8	2350	127.6	3120	96.10	4540	66.04	7450	40.24
230	1304	940	319.0	1650	181.7	2360	127.0	3140	95.48	4560	65.75	7500	39.98
240	1249	950	315.6	1660	180.6	2370	126.5	3160	94.88	4580	65.46	7550	39.71
250	1199	960	312.3	1670	179.5	2380	126.0	3180	94.28	4600	65.18	7600	39.45
260	1153	970	309.1	1680	178.5	2390	125.4	3200	93.69	4620	64.90	7650	39.19
270	1110	980	305.9	1690	177.4	2400	124.9	3220	93.11	4640	64.62	7700	38.94
280	1071	990	302.8	1700	176.4	2410	124.4	3240	92.54	4660	64.34	7750	38.69
290	1034	1000	299.8	1710	175.3	2420	123.9	3260	91.97	4680	64.06	7800	38.44
300	999.4	1010	296.9	1720	174.3	2430	123.4	3280	91.41	4700	63.79	7850	38.19
310	967.2	1020	293.9	1730	173.3	2440	122.9	3300	90.86	4720	63.52	7900	37.95
320	936.9	1030	291.1	1740	172.3	2450	122.4	3320	90.31	4740	63.25	7950	37.71
330	908.6	1040	288.3	1750	171.3	2460	121.9	3340	89.77	4760	62.99	8000	37.48
340	881.8	1050	285.5	1760	170.4	2470	121.4	3360	89.23	4780	62.72	8050	37.25
350	856.6	1060	282.8	1770	169.4	2480	120.9	3380	88.70	4800	62.46	8100	37.02
360	832.8	1070	280.2	1780	168.4	2490	120.4	3400	88 18	4820	62.20	8150	36.79
370	810.3	1080	277.6	1790	167.5	2500	119.9	3420	87.67	4840	61.95	8200	36.56
380	789.0	1090	275.1	1800	166.6	2510	119.5	3440	87.16	4860	61.69	8250	36.34
390	768.8	1100	272.6	1810	165.6	2520	119.0	3460	86.65	4880	61.44	8300	36.12
400	749.6	1110	270.1	1820	164.7	2530	118.5	3480	86.16	4900	61.19	8350	35.91
410	731.3	1120	267.7	1830	163.8	2540	118.0	3500	85.66	4920	60.94	8400	35.69
420	713.9	1130	265.3	1840	162.9	2550	117.6	3520	85.18	4940	60.69	8450	35.48
430	697.3	1140	263.0	1850	162.1	2560	117.1	3540	84.70	4960	60.45	8500	35.27
440	681.4	1150	260.7	1860	161.2	2570	116.7	3560	84.22	4980	60.20	8550	35.07
450	666.3	1160	258.5	1870	160.3	2580	116.2	3580	83.75	5000	59.96	8600	34.86
460	651.8	1170	256.3	1880	159.5	2590	115.8	3600	83.28	5050	59.37	8650	34.66
470	637.9	1180	254.1	1890	158.6	2600	115.3	3620	82.82	5100	58.79	8700	34.46
480	624.6	1190	252.0	1900	157.8	2610	114.9	3640	82.37	5150	59.22	8750	34.27
490	611.9	1200	249.9	1910	157.0	2620	114.4	3660	81.92	5200	57.66	8800	34.07
500	599.6	1210	247.8	1920	156.2	2630	114.0	3680	81.47	5250	47.11	8850	33.88
510	587.9	1220	245.8	1930	155.3	2640	113.6	3700	81.03	5300	56.75	8900	33.69
520	576.6	1230	243.8	1940	154.5	2650	113.1	3720	80.60	5350	57.11	8950	33.50
530	565.7	1240	241.8	1950	153.8	2660	112.7	3740	80.17	5400	55.52	9000	33.31
540	555.2	1250	239.9	1960	153.0	2670	112.3	3760	79.74	5450	55.01	9050	33.13
550	545.1	1260	238.0	1970	152.2	2680	111.9	3780	79.32	5500	54.51	9100	32.95
560	535.4	1270	236.1	1980	151.4	2690	111.5	3800	78.90	5550	54.02	9150	32.77
570	526.0	1280	234.2	1990	150.7	2700	111.0	3820	78.49	5600	53.54	9200	32.59
580	516.9	1290	232.4	2000	149.9	2710	110.6	3840	78.08	5650	53.07	9250	32.41
590	508.2	1300	230.6	2010	149.2	2720	110.2	3860	77.67	5700	52.60	9300	32.24
600	499.7	1310	228.9	2020	148.4	2730	109.8	3880	77.27	5750	52.14	9350	32.07
610	491.5	1320	227.1	2030	147.7	2740	109.4	3900	76.88	5800	51.69	9400	31.90
620	483.6	1330	225.4	2040	147.0	2750	109.0	3920	76.49	5850	51.25	9450	31.73
630	475.9	1340	223.7	2050	146.3	2760	108.6	3940	76.10	5900	50.82	9500	31.56
640	468.5	1350	222.1	2060	145.5	2770	108.2	3960	75.71	5950	50.39	9550	31.39
650	461.3	1360	220.4	2070	144.8	2780	107.8	3980	75.33	6000	49.97	9600	31.23
660	454.3	1370	218.8	2080	144.1	2790	107.5	4000	74.96	6050	49.56	9650	31.07
670	447.5	1380	217.3	2090	143.5	2800	107.1	4020	74.58	6100	49.15	9700	30.91
680	440.9	1390	215.7	2100	142.3	2810	106.7	4040	74.21	6150	48.75	9750	30.75
690	434.5	1400	214.2	2110	142.1	2820	106.3	4060	73.85	6200	48.36	9800	30.59
700	428.3	1410	212.6	2120	141.4	2830	105.9	4080	73.49	6250	47.97	9850	30.44
710	422.3	1420	211.1	2130	140.8	2840	105.6	4100	73.13	6300	47.59	9900	30.28
										6350	47.22	9950	30.13

length of a single wave, which in this case would be equal to 300,000,000 (appx.) divided by 500, equalling 600,000 cycles per second or (reduced to kilocycles by dividing by 1,000) 600 kilocycles. On the preceding page is given a chart for converting kilocycles to meters, or meters to kilocycles.

KILOLINE—This is equal to 1,000 *lines of force*, or 1,000 *maxwells* (q.v.). The *maxwell* or the *line of force* is the unit used to measure *magnetic flux.* (See *Line of Force, Maxwell, Magnetic Flux.*)

KILOVOLT—abbreviated K.V.—The kilovolt is equal to 1,000 volts.

KILOVOLT-AMPERE — Abbreviated K.V.A. or kv-a.—The product of the effective volts across the terminals of a circuit by the effective amperes flowing in that circuit, divided by 1,000. Alternators are rated in *kilovolt-amperes* because their load determines the power factor. Thus with a zero *power factor* (q.v.) the *true power* (q.v.) in kilowatts would equal zero even if the alternator were delivering maximum amperes and maximum volts. At 50% power factor the kilowatt capacity would be half that at 100% power factor. On the other hand the kilovolt-ampere capacity remains the same no matter what the power factor.

KILOWATT—abbreviation K.W.—The kilowatt is a unit of electrical power having the value of 1,000 watts. One kilowatt equals 1,000 watts. To reduce the power expressed in kilowatts to watts, it is necessary to multiply the number of kilowatts by 1,000. Most electrical formulas use watts rather than kilowatts. Therefore, in solving such formulas, where the power is given in kilowatts, it is necessary to multiply this by 1,000 before using it in the formula. For example, suppose the power expended in an electrical circuit carrying 100 amperes is 20 kilowatts. The voltage will equal the power in watts divided by the current in amperes. This is $\frac{20 \times 1{,}000}{100} = 200$ volts. Motors are often rated in kilowatts instead of horsepower. For practical calculations, the horsepower rating of a motor is equivalent to four-thirds of its kilowatt rating. (See *watts.*)

KILOWATT-HOUR—The energy expended when work is done for one hour at the rate of one kilowatt. It is possible to calculate the energy expended in an electrical circuit over a certain period of time, knowing the current in amperes and the resistance of the circuit. Using the following formula the energy expended will be obtained in kilowatt-hours:

$$\text{Work performed or energy expended} = \frac{\text{Current}^2 \times \text{Resistance}^* \times \text{Time}^\dagger}{3{,}600{,}000} = \text{kilowatt-hours.}$$

KIMURA SHUNKICHI—Japanese radio authority. Kimura was born in 1866, and educated at the Scientific College of the Tokyo Imperial University, his special study being physics, which he continued to study at Harvard and Yale Universities, United States.

In 1901 Kimura entered the Japanese Navy and first began to study wireless telegraphy, particularly for naval use. In 1906 he was appointed the Japanese delegate to the International Wireless Telegraph Conference, Berlin, and after his retirement from the Navy, 1912, he became the director of the Nippon Radio Telegraph and Telephone Company. Kimura has written largely on radio subjects for various scientific journals.

KINETICS—That branch of dynamics dealing with forces that produce or change motion. Energy is usually classified either as *potential* or *kinetic.* When energy is available for the production of work it is referred to as potential energy. Energy at work is *kinetic energy* (q.v.). The term *kinetic* refers to motion.

KINETIC ENERGY—The work a body is able to do by reason of its motion. The unit of energy is the *erg.* This unit is also used to measure work. (See *Erg.*) It should be noted that *kinetic energy* is stored in the magnetic field. It requires more energy to start a current in a circuit having great inductance than in one having small inductance. This greater amount of energy is stored in the magnetic field and is returned to the circuit when the current decreases to zero.

KINRAIDY SPARK GAP—A form of *quenched spark gap* (q.v.) consisting of two water-cooled flat terminals close together and between which discharge occurs.

KIRCHOFF'S LAWS—These are two in number. First Law: The sum of the currents flowing to a point, in any electrical circuit, is equal to the sum of the currents flowing away from that point. Second Law: In any closed electrical circuit, the sum of the impressed *electromotive forces* (q.v.) will equal the sum of the *voltage drops* (q.v.). This statement requires modification, in so far as "addition" of voltages is concerned. Voltages are added provided that they are in the same direction, but must be subtracted if in opposite directions. These two laws depend on *Ohm's Law* (q.v.) and constitute a further application of that law to more complicated electrical circuits. In addition to simple electrical circuits, conductors may be connected in various complicated networks, all of which come under the heading of *divided circuits.* By means of *Kirchoff's Laws*, the current in any part of a divided circuit may be found, if the resistances of the various parts, and the electromotive forces are given.

KITES AS AERIAL SUPPORTS—During the early experiments with wireless, kites were successfully employed by Marconi as aerial supports.

KLEIN, RENE HENRI—Anglo-French radio expert. Born at Soults-sous-Foret, France, in 1880, he first became interested in wireless telegraphy in 1908. He was the founder and first honorary secretary of the Wireless Society of London, 1913-20, and became a vice-president of the society. He was one of the inventors with H. L. McMichael of the synthetic galena crystal Radiocite, and has written many articles on wireless subjects.

KNIFE SWITCH—A switch in which the movable arm wedges in between two parallel spring clips. Knife switches may be either of the *single pole* (s.p.) or *double pole* (d.p.) types and also may either be *single throw* (s.t.) or *double throw* (d.t.). Knife switches are often referred to as *Lever* switches They are used on direct current circuits up to 250 volts and on alternating current circuits up to 500 volts (See *Ground Switch; Double Throw Switch.*)

KLEMENCIC THERMAL JUNCTION—A device consisting of two thin sheets of brass, designated in the illustration as A, A. These are 4 inches by 11.8 inches and are placed about 1 inch apart. Soldered to one plate is a fine platinum wire and to the other plate is a fine platinum-nickel wire. These two wires are connected together at J to form a *thermal junction* (q.v.) and are then led off at right angles and are soldered at their other ends to the leads R, R, of a sensitive *galvanometer* (q.v.). This resonance system is fixed at the focal line of a suitable cylindrical metallic reflector. When electric waves, with the electric force parallel to A A fall on this receiver, electric oscillations between A and A produce heating at the junction J which is the point of contact between the two dissimilar metals. In consequence, the heat developed gives rise to a thermoelectromotive force at the junction and consequently to a current in the galvanometer. This instrument can be used to detect and measure electric waves.

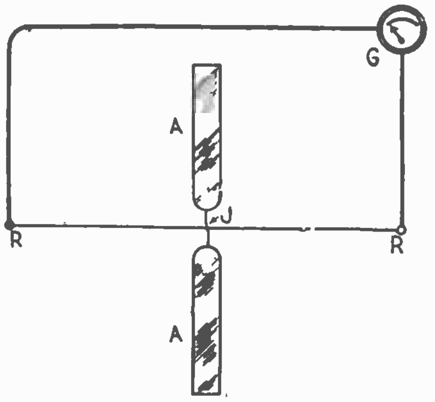

Schematic diagram showing Klemencic thermal junction.

KOEPSEL PERMEAMETER—An instrument used for measuring the *permeability* (q.v.) of an iron or steel bar. In its principle of operation, this device is a movable coil milli-voltmeter in which the permanent magnet is replaced by an electromagnet. The sample to be tested forms a part of this electromagnet. The flux produced in the sample and in the pole pieces of the instrument is measured by a movable coil similar to those used in direct current voltmeters. The coil is supplied with energizing current from a small dry battery. Deflections are indicated by a pointer on a dial, the scale being calibrated directly in magnetic densities.

KOHLRAUSCH'S LAW—This law states that the rate of motion of each atom in a solution undergoing electrolysis for a given liquid is independent of the element with which it may have been in combination.

KOLSTER, FREDERICK A.—American radio authority. Born at Geneva, Switzerland, January 13, 1883, he was educated at Cambridge, Mass., and Harvard University. From 1902-08 he was assistant to J. S. Stone, the radio expert, and from 1909-12 to Dr. Lee de Forest. He was appointed chief of the Radio Section of the Bureau of Standards, a post he held until 1921, when he became research engineer to the Federal Telegraph Company.

Kolster has written a large number of articles on radio subjects, including those on the "Effects of Distributed Capacity in Coils," "Reinforced Harmonics in Radio Transmission," etc. He is the inventor of the Kolster decremeter, of a radio compass, of directional radio systems, etc., and is a member of the American Institute of Radio-Engineers.

---

*Ohms. †Seconds.

KOOMANS, NICHOLAS—Dutch radio authority. Born at Delft, 1879, he was educated there as a mechanical and electrical engineer. In 1908 he received his doctorate for his dissertation "Regarding the Influence of Self-Induction in Telephone Conducting Wires." He was one of the founders of and editor of the "Monthly Review of Telephony and Telegraphy," and was also one of the founders of the Dutch Society for Radio Telegraphy. Koomans is a member of the International Electrotechnical Commission and professor in physics and theoretical electrical engineering at the school of the Dutch Post and Telegraph Administration.

KORDA AIR CONDENSER—A variable *condenser* (q.v.) using air as the *dielectric* (q.v.) consisting of two sets of semi-circular plates, one set being connected together and fixed in position and the other set being also connected together but capable of being rotated about a central axis. Rotating the movable plates brings a greater or less area of the two sets of plates into interlapping positions, thus permitting the capacity of the condenser to be varied. This condenser was patented by Korda in Germany in 1893.

KORN, ARTHUR—German radio expert. Born at Breslau, Germany, 1870, Korn was educated at Leipzig and Paris and became professor of physics at the University of Munich, 1903-8, and afterwards professor at the Polytechnical High School, Charlottenburg, Berlin.

Korn is well known for his experiments on and the invention of a method of transmitting photographs by telegraphy, the first photograph being telegraphed from Munich to Berlin in 1907 by his methods. He is the inventor of a system of telautography and wireless phototelegraphy, and the author of a number of standard books on electricity.

L

L—The symbol of *inductance* (q.v.) or the *coefficient* of *self-induction.* In general, the co-efficient *L* by which the rate of change of current in any circuit must be multiplied to obtain the value of the self-induced *electromotive force* (q.v.) is called the *coefficient* of *self-induction* or simply the *inductance.* (See *Self-induction, coefficient.*)

LAG—One *alternating current* quantity reaching its maximum at a later time than a second alternating current quantity. As an example,—the current is said to *lag* behind the impressed voltage when maximum current occurs after maximum voltage. In other words, the current and voltage in such an alternating current circuit are out of *phase* (q.v.). In the above example, since the current *lags* behind the voltage, the voltage is said to lead the current. *Inductance* (q.v.) in an alternating current circuit will cause the current to *lag* behind the impressed voltage. *Capacity* (q.v.) in such a circuit will cause the current to *lead* (q.v.) the impressed voltage. The word *lag* may be applied to magnetic quantities as well as electrical quantities. Thus, one magnetic flux may *lag* behind another as in the case of an induction motor where the induced *magnetic flux* (q.v.) *lags* behind the impressed magnetic flux. Another type of magnetic *lag* is the retardation of magnetic effects behind their causes, due to *hysteresis* (q.v.) in the magnetized substance. (See *Lead,* also *Leading Current.*)

LAMINATED—Composed of a number of thin plates, one on top of another. Examples of laminated construction are *laminated brushes, laminated cores laminated conductors,* etc.

LAMINATED CORE—An arrangement of soft iron plates or stampings forming the core of a transformer, or of an armature of a motor or generator. Cores are laminated in this way, principally to reduce *eddy current* (q.v.) losses. In the case of an armature core, this must be laminated parallel to the direction of rotation (in other words, perpendicular to the axis of rotation). Laminations are insulated from each other by means of japan, varnish, shellac or simply rust.

LAMP BANK—A bank of incandescent lamps arranged so that they can be connected in series or parallel. *Lamp banks* are used in laboratories in conjunction with electrical machinery under test. In this case, the lamps act as the *load,* the more lamps being connected in *parallel,* the greater the amount of current being drawn. *Lamp banks* are sometimes connected in *series* in a circuit for the purpose of reducing voltage. An example in this connection is the use of a lamp or bank of lamps in charging a 6-volt storage battery from a 110-volt direct current source. Lamp banks are sometimes called *lamp panels* or *lamp batteries.*

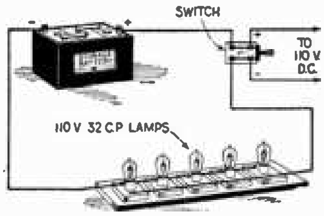

Use of lamp bank in charging storage battery.

LANGMUIR, IRVING—American radio authority. Born at Brooklyn, New York, January 31st, 1881, he was educated at the School of Mines, Columbia University, and was for some time assistant to Professor Nernst at the University of Gottingen. In 1906-9 he was instructor in chemistry at the Stevens Institute of Technology, and in the latter year he became a research assistant to the General Electric Company, at Schenectady. He has carried out a series of brilliant researches on apparatus used in wireless telegraphy and telephony, on electron discharge apparatus, atomic and molecular structure, etc.

Langmuir has paid particular attention to high-vacuum valves, and is the inventor of the Langmuir valve. To him is due the discovery that by treating the tungstic oxide used in the construction of tungsten filaments with certain compounds of thorium the filament becomes thoriated tungsten and the electronic emission is enormously increased. These filaments are the ones used in dull emitter valves, the electrons being given off at comparatively low temperatures Langmuir is the author of many papers to scientific and technical journals, including those on pure electron discharge, thermionic currents in high vacua, etc. He is also well known for his mercury condensation pump for high vacua.

LAP WINDING—An armature winding in which the opposite ends of each coil are connected to adjoining commutator segments. As a result the windings lap back and upon themselves forming loops. *Lap winding* is also called *loop winding, parallel winding* and *multiple winding.* In the elementary form of *lap winding* the method of connection to the commutator bars is as follows: One side of a coil is connected to a commutator bar, and the other side of the coil, which is located 180 electrical degrees (the distance from the center of a north pole to the center of the adjacent south pole) away, has its end connection brought back and soldered to the next adjacent commutator segment. This method of overlapping the ends is continued all the way around the armature until all the slots have been filled and the circuit has been closed on itself. The coils must all be symmetrical. This means that if there is a forward throw of a certain num-

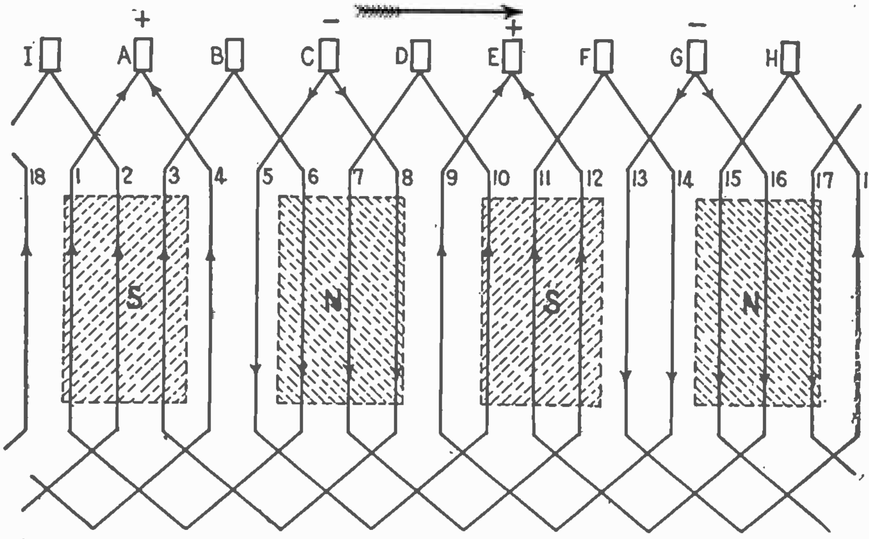

A typical lap winding shown in developed view. In the winding shown each inductor is connected at the rear of the armature to one five slots away. Thus 3 is connected to 8, 5 to 10, etc. After having made one complete "element"—as, for example, A-1-6-B—the winding forms a second "element," B-3-8-C, which laps over the first. This is continued around the armature until the winding closes on itself.

ber of slots on the first coil, all the other coils must have the same forward throw. In like manner all the return throws must be the same. In the *lap wound* generator there are as many paths for the current as there are poles. Hence the name *parallel* or *multiple winding*. This necessitates the use of as many sets of brushes as there are poles. An increased number of paths in parallel means increased current capacity. In the case of a motor, the greater the number of armature paths between brushes, the greater will be the speed, other conditions remaining equal.

LAW OF ELECTROMAGNETIC INDUCTION—Also known as the "*Law of Generator Action*" and as "*Faraday's Law*." If there is relative motion between an electrical conductor and a magnetic field, such that the conductor cuts *flux*, an *electromotive force* (q.v.) will be induced in the conductor. The amount of the induced electromotive force will depend upon the strength of the magnetic field, upon the speed of cutting, upon the length of the conductor cutting the field and upon the direction of motion between the conductor and the field. Thus the stronger the magnetic field, the greater the induced electromotive force. The induced electromotive force will also increase with increased speed of cutting and with increased conductor length. When the conductor cuts the flux at right angles, the induced electromotive force is maximum. If the flux is cut in any other direction, the induced electromotive force will be less. The induced electromotive force decreases from maximum to zero as the angle of cutting decreases from 90 degrees to zero degrees. If a conductor is of such length and moves in such a direction, and at such a speed, as to cut 100,000,000 lines of force in one second, one *volt* (q.v.) will be induced in it. Expressed as an equation the induced electromotive force (in volts) will equal the flux cut (in *lines* or *maxwells*) divided by the product of the time (in seconds) times 100,000,000. This equation refers to the average value of the electromotive force induced in a single conductor. If a number of conductors are used the same electromotive force will be induced in each. If the conductors are so connected that their electromotive forces add up, the total average electromotive force can be found by multiplying the above equation by the number of conductors.

LATOUR, MARIUS—French radio expert. Born in October, 1875, he was educated at the University of Paris and the Ecole Superieure d'Eléctricité, Paris. He was for many years consulting engineer to the General Electric Company, and he specialized in the construction of high-frequency machines. Latour is the designer of the so-called S. F. R. alternator, in which there are a smaller number of stator slots than usual.

During the World War Latour was engaged in research work under General Ferrie, and he invented a system of elimination of the interference produced in telephone lines by neighboring high-tension power lines which has been installed throughout the greater part of France. Latour is the inventor of the now widely used system of high-frequency multiplex telegraphy and telephony, using the three-electrode valve for generation and reception. He is a member of many scientific societies, including the Institute of Radio Engineers, and of the American Institute of Electrical Engineers. He has written many articles on radio for scientific and technical journals, and is the inventor of a great many important radio developments covering practically the whole range of the art.

LEAD—Pronounced LEED—One alternating current quantity reaching its maximum at an earlier time than a second alternating current quantity. Thus in an alternating current circuit, the current is said to *lead* the voltage if maximum current occurs before maximum voltage. If the current reaches its maximum 90 *electrical degrees* ahead of the voltage, current and voltage are said to be in *quadrature* (q.v.) with current *leading* and voltage *lagging*. (See *Degrees, Electrical*, also *Leading Current* and *Lag*.)

LEAD SULPHATE—Formula $PbSO_4$—A white insoluble salt which is found in native form in the ground and which can be formed artificially by adding sulphuric acid to a soluble lead salt. The active elements of a lead *storage battery* (q.v.) consists of (PbO_2) *lead peroxide* on the positive plate and (Pb) *spongy lead* on the negative plate, with dilute *sulphuric acid* (H_2SO_4) as the *electrolyte* (q.v.). The final result, on discharge of such a battery is the formation of *lead sulphate* on both the positive and negative plates. This is a part of the process of producing current. After a normal discharge, the sulphate is finely crystalline and of such nature that it is easily reduced by the current flowing through the battery on charge. If charging is neglected and the battery is allowed to stand in a discharged state, the sulphate gradually ceases to be crystalline. It fills the pores of the plates and finally makes the active material hard and dense, thus lengthening the time required for a charge. A battery in this condition is said to be *sulphated*. If sulphating proceeds too far, it will result in *buckling* of *plates* (q.v.), loss of active material, greatly decreased capacity, lower efficiency and an increased internal resistance. A moderately sulphated battery can be restored by means of a prolonged charging at a moderate rate. In a badly sulphated battery, a number of successive charges and discharges may be necessary.

LEAD-IN—The wire leading from the aerial to the radio apparatus. *Lead-in* wire, as used for receiving sets, is usually rubber covered wire although bare antenna wire can be used provided it is properly insulated from the walls of the building. A poorly insulated lead-in will greatly reduce the efficiency of the finest antenna. (See *Down Lead.*)

LEAD-IN INSULATOR—Any form of insulator used for passing aerial lead-in through window, wall or roof. These insulators are usually made of porcelain, electrose or insulated copper strips. (See *Insulator, lead-in*, also *Window Lead-in.*)

LEADING CURRENT—An alternating current which reaches its maximum value at an earlier time than the voltage. When an inductance is inserted in an alternating current circuit, the self-induced *electromotive force* (q.v.) will combine with the *impressed* electromotive force and the resultant of the two will be the active electromotive force which causes the current flow. The current will always be exactly in *phase* (q.v.) with and proportional to the resultant electromotive force. The induced electromotive force is 90 degrees behind the current and therefore behind the resultant electromotive force, which is in phase with the current. The impressed electromotive force will lead the current and will be higher than the resultant electromotive force by an amount which at each instant is equal to the counter-electromotive force of the inductance. When a condenser is inserted in an alternating current circuit the current will charge the condenser. As the amount of current stored in the condenser increases, the electromotive force of the condenser increases also until the impressed and the condenser electromotive forces are equal. The condenser electromotive force, being a counter pressure, the current flow will cease when the two electromotive forces balance. The current being zero at this point and the condenser electromotive force a maximum, it follows that the condenser electromotive force is 90 degrees in advance of the current and hence of the resultant electromotive force which is in phase with the current. The impressed electromotive force lags behind the current which in this case is a *leading current* and moreover the impressed electromotive force is greater than the resultant electromotive force. It follows that if either capacity or inductance be placed in an alternating current circuit, the phase of the current with respect to the impressed electromotive force will change and the current flow will be reduced. Capacity sets up an electromotive force 90 degrees in advance of current flow, while inductance sets up an electromotive force of 90 degrees behind current flow. These two effects tend to neutralize each other when the proper amount of inductance and capacity are connected in series in a circuit and when they are just equal, no electromotive force other than the impressed is left to act on the circuit, the resultant and impressed electromotive forces are identical and there is no phase displacement. This is said to be a *resonant* (q.v.) circuit. (See *Lead, Lag*, also *Inductive Rise.*)

LEADS—Wires conveying or "leading" current from one point to another in a circuit. The conductors in any system of electrical distribution are referred to as *leads*. Sometimes the term *lead* is applied to the flexible wire or cable conveying current to or from a piece of apparatus.

LEAK—See *Grid Leak*.

LEAKAGE LINES—Also termed *leakage flux*—That part of the total magnetic flux in a magnet, transformer, generator, etc., which fails to complete the closed magnetic circuit and hence is dissipated or wasted. (See *Keeper*.)

LEAKAGE, SURFACE—The flow of electricity across the surface of insulators. Its amount depends on the material and preparation of the surface and the humidity of the atmosphere and is recorded by measuring the surface *resistivity* (q.v.). Several methods are employed to reduce *surface leakage*. Porcelain *insulators* for outdoor use are often of the *petticoat* type so that a portion of the surface of the insulator remains dry during a rain or snowstorm. Another method very generally used in radio work to reduce surface leakage, is the use of corrugated surfaces on insulators. This method increases the length and hence the resistance of the surface across which leakage is likely to take place.

LENGTH OF AERIAL—The distance from the aerial binding post at the instrument to the farthest point of the antenna system, measured to any of its extremities. Only one component wire of a system is taken into consideration even if two, three or four are used. In a *T aerial* only half of the horizontal span must be added to the length of the down lead. In the *umbrella* type, the length of one radial lead is added to the down lead.

LENZ'S LAW—An induced current always tends to stop the motion which produces it. Stated in another way, whenever current flows in a conductor, due to electromagnetic induction, the action between the magnetic field which is being cut and the field around the conductor is always such as to tend to oppose further motion of the conductor. The following is still another way of stating *Lenz's Law*: The field due to the induced current always acts to prevent a change of the original field.

LEYDEN JAR—A condenser consisting of a glass jar with an outer and an inner coating of tinfoil covering the bottom and the sides nearly to the neck.

A small leyden jar.

A brass rod terminating in an external knob passes through a wooden stopper and is connected to the inner coating by a loose chain. This condenser received its name from Leyden, the city in Holland in which it was invented.

L. F. C.—Abbreviation for *Low Frequency Current.* (See *Frequency, Low,* also *Frequency.*)

LIGHT—Radiant energy or visible *radiation* which affects the eye so that objects become visible. Light consists of *ether* vibrations varying between 400 billions per second producing waves 271 ten-millionths of an inch in length and giving red light, up to 750 billions per second producing waves of 165 ten-millionths in length, giving violet light. Faster or slower vibrations produce waves not visible to the human eye. Shorter waves are called *ultraviolet rays,* while those immediately above the red rays are known as *infra-red rays.* The speed of light waves is identical with that of other electromagnetic waves and electricity, and is 186,000 miles or 300,000,000 meters per second.

LIGHT LINE AERIAL—A device which acts as an aerial for a radio receiving set and which utilizes the electric lighting system by means of a plug to which a condenser is attached. The plug is inserted in any lighting socket and the aerial connection of the receiving set it attached by means of a flexible lead to one terminal of the condenser. (See *Adapter, Aerial.*)

LIGHTNING ARRESTER—(Also called *lightning protector.*) A device for protecting radio or other electrical apparatus from lightning or from atmospheric discharges. Radio *lightning arresters* are provided with two terminals, one being connected with the aerial and the other directly to a ground. Between these terminals is a medium which offers practically no resistance (or technically speaking, which offers low *impedance* (q.v.)) to the flow of very high frequency currents such as lightning discharges. However, this medium acts as an insulator to the normal operating currents. Hence, during the ordinary reception of radio signals, the currents flow past the *arrester* to the receiving instrument but they are unable to pass through the insulating medium of the *ligtning arrester* to the ground. When there is a lightning discharge, the arrester offers the shortest and most direct path to the ground and hence the dangerous current never reaches the apparatus but proceeds directly to the ground through the *lightning arrester.* The more popular designs of lightning arresters are of three types: vacuum, non-air-gap and air-gap. The vacuum type is the more sensitive of the three, due to its having a low functioning voltage. In a usual form, it consists of two terminals in a sealed glass chamber from which air has been partially exhausted. This type of arrester can be discharged repeatedly without impairing its efficiency. The non-air-gap type of arrester involves the use of a carborundum stone, with a special clay binder. This type of arrester has a rectifying action similar to that used in a carborundum detector. The air-gap type of *lightning arrester* depends for its efficiency on a closely spaced air-gap which offers a low break-down voltage. The Underwriters requirements call for the use of an approved lightning protective device on all radio receiving sets where outside aerials are used. Further requirements are that such arresters shall operate on 500 volts or less. (See *Arrester.*)

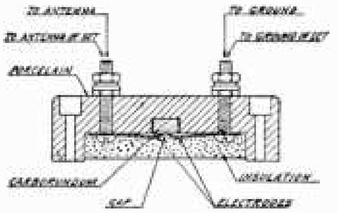

A carborundum type lightning arrester consisting of a block of carborundum held between two pieces of high resistance carbon.

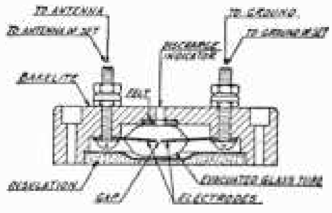

Vacuum type of lightning arrester for indoor work.

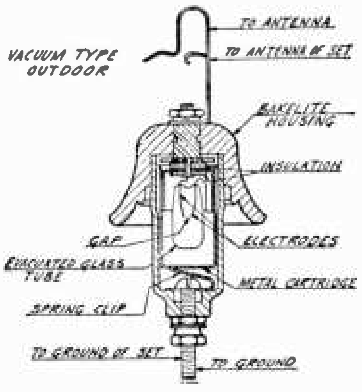

Sectional view of vacuum type lightning arrester used for outdoor work.

Outdoor vacuum type lightning arrester shown above in sectional view.

LIGHTNING GAP—This refers to the *air gap* or *vacuum gap* in a *lightning arrester* (q.v.). This gap is connected across the aerial and ground leads to the receiving set. When the potential in the aerial rises above a certain value (usually 500 volts or less) a spark jumps the gap, discharging the excessive energy to the ground without damage to the radio apparatus. (See *Gap Arrester.*)

LIGHTNING SWITCH—A single pole, double throw knife *switch* (q.v.) used for grounding the antenna during a storm, or in some cases at all times when the apparatus is not in use. The upper terminal of the switch is usually connected to the aerial, the central terminal to which the switch arm is connected, leads to the radio set and the lower terminal is connected to the ground. In most installations, the *lightning switch* is mounted outside the building in which the radio apparatus is located. (See *Ground Switch.*)

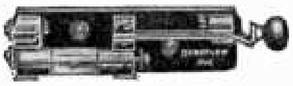

Lightning switch combined with vacuum type lightning arrester.

LIKE CHARGES—Bodies which carry like *electrostatic charge* (q.v.) exert a repulsion against each other.

LIKE POLES—Like magnetic poles repel one another. This is one of the *laws of magnetism.* (See *Magnetism, Laws of.*)

LIMITING VOLTAGE OF BRUSH DISCHARGE—The potential at which a brush discharge passes into an arc or spark. Brush discharges and arcs may be reduced in radio transmitting apparatus by eliminating sharp points or edges on conductors and by coating the edges of metal plates with paraffin. (See *Brush Discharge,* also *Corona.*)

LIMITING VOLTAGE OF GLOW DISCHARGE—The potential at which the glow gives place to the *brush discharge* (q.v.) in a high tension line or in a radio transmitting aerial (q.v.). (See *Corona.*)

LIMITS OF AUDIBILITY—The limiting *frequencies* (q.v.) at which audible sounds are produced. These range from 500 to 1000 cycles per second up to as high as 10,000 cycles per second. *Audio frequencies* (q.v.) are sometimes defined as those conveniently heard in the telephone. (See *Frequency, High,* also *Frequency, Low.*)

LINE—A conducting wire between terminals in a system of electric communication or distribution. The positive wire is sometimes specifically referred to as the *line* and the return wire as *earth.* In many cases the earth or ground is used in the place of a return wire.

LINE OF FORCE—A theoretical conception used to map out a magnetic field and to show the direction along which the magnetism acts. *Lines of force* are conceived as forming closed loops and the path they traverse is called the *magnetic circuit* (q.v.) The positive direction of the *lines of force* is arbitrarily taken as that in which the north-seeking end of a compass moves. *Lines of force* are represented in diagrams by curves so drawn that the direction of the tangent at any point is the direction of the *magnetic force* (q.v.) at that point. The closeness of the lines is used to indicate the intensity of the force. The *line of force* is also used to indicate and map out an *electrostatic* (q.v) field in the same way as it is used for the magnetic field. The *line of force* is also called the *line of flux* or the *line of induction*. This latter term is used especially where the line of force is being considered in respect to its electrical effect. In some cases it is simply referred to as a *line*. Thus a line (*of force*, or *of flux*, or *of induction*) is an electromagnetic unit of magnetic flux and is equal to 10^{-8} *webers* (q.v.). In the practical system of measurements, the *maxwell* (q.v.) equal to one *line*, is the unit of *magnetic flux* (q.v.). When 100,000,000 *maxwells* are cut by a single conductor, in one second, one *volt* (q.v.) will be induced. (See *Law of Electromagnetic Induction*, also *Electromagnetic field.*)

LINEAR OSCILLATOR—See *Oscillator, Linear.*

LINE RADIO—Also known as *Wired Wireless*—A system of communciation devised by Major General George O. Squier which permits the transmission of electromagnetic waves along a wire or conductor. Instead of the waves spreading out in all directions, as in ordinary broadcasting, the waves travel along a predetermined path, viz: the conductor. It is possible to transmit either code or speech by *wired wireless.* Vacuum tubes are used as oscillators at the transmitting end and a specially designed vacuum tube set is used as the receiver. When code transmission is desired, a telegraph key is inserted in the circuit to intermittently stop and start the oscillating waves. This produces dots and dashes and in this way it is possible to send code using undamped radio frequency waves. When it is desired to telephone, the key is omitted and the radio frequency or *carrier waves* (q.v.) are produced continuously except when they are varied or modulated in amplitude corresponding to the voice of the speaker at the transmitting end. The wave which passes from the transmitting end of the wire to the receiving end is of radio frequency and is called the *carrier wave.* In a way, this wave is analogous to the wire which carries the voice current in wire telephony. *Line radio* can be transmitted over ordinary telephone or telegraph lines at the same time as these lines are being used for their ordinary purposes and without any interference whatsoever. High-tension lines are also used for transmission of *wired wireless.*

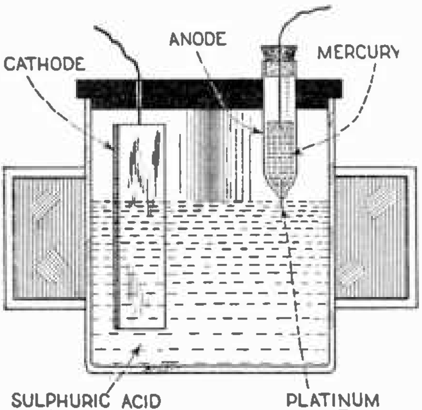

A liquid barreter.

LIQUID BARRETER—An electrolytic detector which utilizes a very fine platinum wire dipped to a small depth in nitric or sulphuric acid. These detectors were formerly used to some extent in receiving wireless code messages. The electrolytic detector of the *barreter* type was invented and named by *Fessenden* (q.v.). The illustration shows a liquid barreter. In the glass jar is a platinum glass-anode, lead cathode and sulphuric acid electrolyte.

LITZENDRAHT—abbreviated Litz—A German word, meaning braided wire. A high frequency conductor built up of a number of very fine conducting strands. The individual strands are enameled and the conductor is usually so constructed that each strand comes to the surface or very near the surface at regular intervals. In the most effective form of conductor the strands are braided so as to form a woven tube. The strands are usually covered by a single braided silk wrapper. The use of *litz* tends to reduce *skin effect* (q.v.). This type of wire has been found to be especially effective when used in the construction of *loop antennas* (q.v.).

LOADSTONE—also spelled *Lode-stone*—Iron ore which shows magnetic properties. The name loadstone is derived from a word meaning leading stone and refers to the property of a compass of pointing to the north magnetic pole. The loadstone is also referred to as a *natural magnet.* (See *Magnetite*, also *Magnet.*)

LOADING—Adding capacity or inductance to a circuit. *Inductance coils* (q.v.) called *loading coils* are used to increase the inductance and condensers are used to increase the capacity. A loading coil may be used to increase the *wave length* (q.v.) of an antenna by placing it in series between the *lead-in* (q.v.) and the receiving set. In the case of a transmitting antenna, the loading coil may be used to tune the aerial for *resonance* (q.v.) at a desired frequency. An example of the use of loading, is that of the addition of a loading coil to a receiving set for the purpose of enabling the set to receive higher wave lengths.

LOADING INDUCTANCE—An inductance coil used to increase the wave length of a circuit. This term usually refers to an inductance coil used in connection with a radio receiving set for increasing its receptive wave length range. Sometimes the term *loading coil* is applied to a unit consisting of an inductance coil, mounted beneath a panel in a small cabinet. On the outside of the panel is a contact arm and a number of studs. These latter lead to taps taken from the inductance coil. By moving the contact arm from one stud to another it is possible to vary the loading of the circuit as desired. A loading coil connected in series with the aerial will increase the natural wave length of the aerial. (See *Loading*, also *Inductance, Antenna.*)

LOCAL ACTION—Electrochemical action within a cell, usually at the positive plate, which does not add to the useful output. In primary cells, zinc is the material usually employed for the positive plate. This zinc practically always contains impurities such as carbon, iron, etc. When the zinc is immersed in the electrolyte, there is a difference of potential between the zinc and the impurities and small local currents flow from the zinc through the electrolyte to the impurities and back to the zinc again. These currents serve no useful purpose and moreover use up the zinc and the electrolyte. *Local action* can be practically eliminated by coating the surface of the zinc with mercury. This is called *amalgamation.* In secondary cells or *accumulators* (q.v.) local action is due to the action which takes place between the lead of the positive and the coating of lead sulphate.

LOCAL OSCILLATIONS—Oscillations set up in the vicinity of a receiving set for test purposes or within the set, as in the *super-heterodyne*, to produce oscillations differing in frequency from the incoming oscillations so as to establish a beat frequency. *Local oscillations* may be produced by means of a buzzer. This is a method frequently used in testing. The use of such oscillations enables a manufacturer of radio receiving sets to test these sets for selectivity, etc., without waiting for actual broadcasting. In the *super-heterodyne* (q.v.) local oscillations are produced by a suitably arranged vacuum tube. (See *Driver.*)

LOCUS—The path of a movement graphically represented by a curve. In mathematical language, the *locus* of an equation is the curve that contains all the points whose co-ordinates satisfy the equation. The circle is defined as the *locus* of all points in a plane equidistant from a fixed point in the plane called the center.

LODGE, SIR OLIVER—British Physicist and radio pioneer. Born at Penkhull, near Stoke-upon-Trent, June 12th, 1851, Oliver Lodge went to the Newport Grammar School at the age of eight, but at fourteen was taken into business to help his father. He worked in the evenings for the intermediate examination in science at the University of London, and eventually took first-class honors in physics. In 1872 he went to University College, London, and in 1877 took his D.Sc. in electricity and became demonstrator, and afterwards assistant professor in physics at University College.

Sir Oliver Lodge.

In 1881 he was appointed first professor of physics at Liverpool, and in 1887 he was made a Fellow at the Royal Society. In 1888 he received the honorary L.L.D. of the University of St. Andrews, the first of a long series of such distinctions. In 1900 he was appointed the first principal of the University of Birmingham, a post from which he retired in 1920. In 1902 he was knighted. Sir Oliver Lodge is one of the foremost physicists of his time, and his brilliant series of researches into electrical phenomena, which he has made his special study have had a profound effect in all branches of that subject. To him, indeed radio owes as much as to any man, for he dealt with fundamentals when the whole subject of electromagnetic radiation was in its infancy. He was led to study the surging and oscillating character of an electrical discharge along wires as a result of his investigations as to the best means of protection against lightning.

One very remarkable experiment of his is now well known under the name of Lodge's resonating jar, now used as a measurer of wave length. In these experiments Lodge was really dealing with the electromagnetic waves discovered by Hertz in 1888. Lodge early recognized the vast importance of the discovery by Hertz, which so brilliantly confirmed the theories of Clerk-Maxwell, and he threw himself into work of investigating these ways with an energy which made him the foremost British physicist on the subject at the time. All this was pioneer work before Marconi and others came along and built upon the solid foundations which Lodge had prepared.

It was during this early period that Lodge discovered his coherer, and with this detector devised the first practical wireless telegraph, sending signals over a distance of several hundred yeards. In May, 1897, Lodge took out his fundamental patent for tuning—a patent that was successfully upheld in the law courts, and extended for seven years by Lord Parker. The patent was acquired by the Marconi Company in 1911. Sir Oliver Lodge has been a pioneer in many other directions. His theories and experiments in connection with the ether, of which he is the leading exponent, are well known, and he has carried out a number of interesting researches on the passage of electricity through liquids. Sir Oliver was the first to devise a simple and direct experiment to show the speed and direction of travel of the constituents of a salt in solution when broken up by a current of electricity. For his researches on electric waves and the passage of light through a moving medium, Sir Oliver Lodge received in 1898 the Rumford Medal of the Royal Society, one of the highest honors the society bestows.

Sir Oliver Lodge has won a well-deserved reputation for the clearness and simplicity with which he is able to explain many of the difficult and fundamental truths of physics, and this clarity of expression is well shown in the many articles he has written.

LODGE DISC DETECTOR—An imperfect contact detector consisting of a steel disc rotating with its sharp knife edge just touching some mercury having a film of oil on it. When an oscillation passes through the circuit, momentary cohesion sets in between the disc and the mercury, but this is immediately broken by the revolving motion of the disc. In this way, oil re-insulates the disc at the completion of each passage of oscillations. Using this type of detector, code signals can be recorded with great precision on a *Morse* tape using a syphon recorder. The *disc detector* is in reality a specialized form of *coherer* (q.v.).

LOG—A list of dial readings of a radio receiving set with corresponding call letters of the broadcasting stations received at these readings. A radio set is said to *log* when the same stations always come in at exactly the same dial readings. Thus on a three dial set, if WEAF comes in at 61, 60 and 60 and always comes in at these identical readings using the same aerial, the set *logs*. All present day sets are constructed so that stations can be accurately logged. Sometimes the term log is used to refer to a graph or chart on which dial settings are plotted against wave lengths of stations recorded. A graph is a diagrammatic representation by means of lines, of the relation between several quantities. Figure 1, gives a simple illustration of this. Suppose it is desired to obtain a general idea of how the radio business varies from year to year or from season to season. On a piece of paper, ruled both vertically and horizontally, as in Figure 1, the months and years are marked at the bottom horizontally. The vertical lines then represent the months and the bottom horizontal line is called the *axis of time*. These vertical lines are known technically as *ordinates*. The vertical line to the extreme left is called the axis of ordinates and after it has been marked as shown in Figure 1, to represent dollars, it may be referred to as the *axis of dollars*. The horizontal lines are known as *abscissas*, the bottom horizontal line being the *axis of abscissas*. Assume that a certain radio concern did $10,000 worth of business in the month of January, 1924. Looking upward along the vertical line representing this month, until the horizontal line representing $10,000 is reached, a point is put at the intersection of these two lines. Suppose that in the following month, February, 1924, the firm does $12,000 worth of business. The point is found as before by following the line February, 1924, up to the line indicating $12,000. Each horizontal line in the figure represents $1,000, so that this point should be found two lines higher than the one marked for $10,000. This can then be continued indefinitely, plotting the amount of business for each month. Assume that in the summer there is a lull in business. The decline starts, say, in the month of May. The business done amounted to $8,000. By glancing at the chart it is possible to ascertain a number of facts which stand out much clearer than where columns of figures are given. Thus the seasonal character of the business is apparent at a glance. It can also be seen readily that the business was better in 1925 than in 1924, etc. Radio logs are plotted in exactly the same way. Each point must be obtained individually in order for the curve to be absolutely accurate, although in many cases a few well chosen points are sufficiently satisfactory. In Figure 2, the dial setting is marked along the bottom of the chart. The scale is here chosen at one division on the chart equal to 5 divisions of the dial. If preferred the scale could be chosen so that one division on the chart would equal one division on the dial as shown on chart on the following page, which is a typical log of an actual set. Proceeding, it is first necessary to find out the limits of the chart. The limits of the horizontal scale are already known, i.e., zero to 100 on the dial. Looking over the list of broadcasting stations, it is found that they all have wave lengths which fall, roughly between 200 and 600 meters. Therefore it is not necessary to make the vertical scale include wave lengths outside of this range. Consequently the vertical scale is started at 200. Counting upward, two divisions on the chart are taken, for convenience, as equal to 50 meters. Having done this, actual readings are next taken on the radio set. Suppose that the first station tuned in is WEAF and that this is tuned in at 85 on the dial. This station has a wave length of 492 meters. Looking along the bottom of the chart for 85, the vertical line is then followed until the position of 492 meters is reached. This is slightly below the 500 line on the chart, Figure 2, and is indicated by the letter A. Next suppose that WOO, having a wave length of 509 meters is tuned in at 90 on the dial. This, then gives the point marked B. Then another station is tuned in, indicated by the point C.

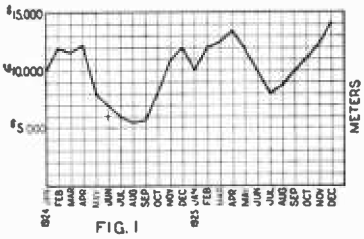

FIG. 1

Log showing variation of a business.

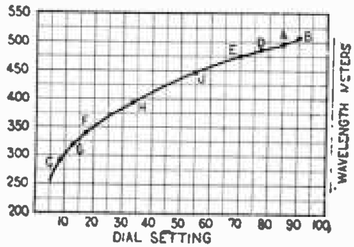

FIG. 2.

A radio log.

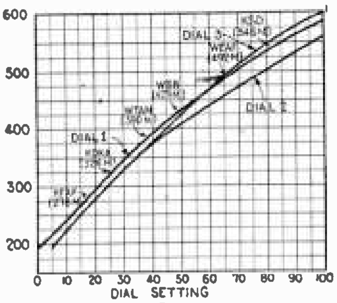

FIG. 3

A three-dial radio log using only one chart.

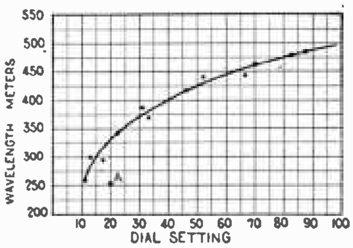

FIG. 4

Radio log showing curve drawn so as to touch as many points as possible.

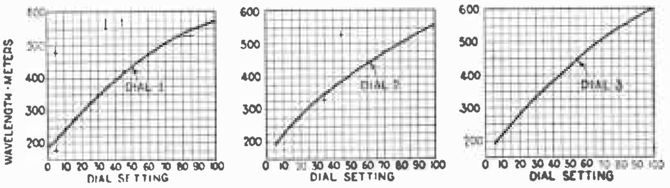

Fig. 5—A three-dial radio log using three separate charts.

Various stations are thus located by points on the chart until enough points are obtained to draw the curve. If the points do not give a smooth curve, it is necessary to approximate a curve, that is to draw it so as to pass through as many of the points as possible. This has been done in Figure 4.

Sometimes points are obtained that do not seem to have any meaning, such as the point A in Fig. 4. This point may have been subject to inaccuracies in reading the dial, or perhaps the operator did not read the wave length correctly from the list of stations. There is always the possibility, also, that he did not plot the point properly. He has plotted the point A as 251 meters at a setting of 20 when it might have been 335 meters at 20 on the dial, in which case it would have fallen directly upon the curve. If none of these difficulties can be found with the point A it can be discarded, however, because it is an absurd point, as the 12 other points, which determine the curve, will testify. Generally it is not good practise to discard points, but the experimenter must learn to use his good judgment.

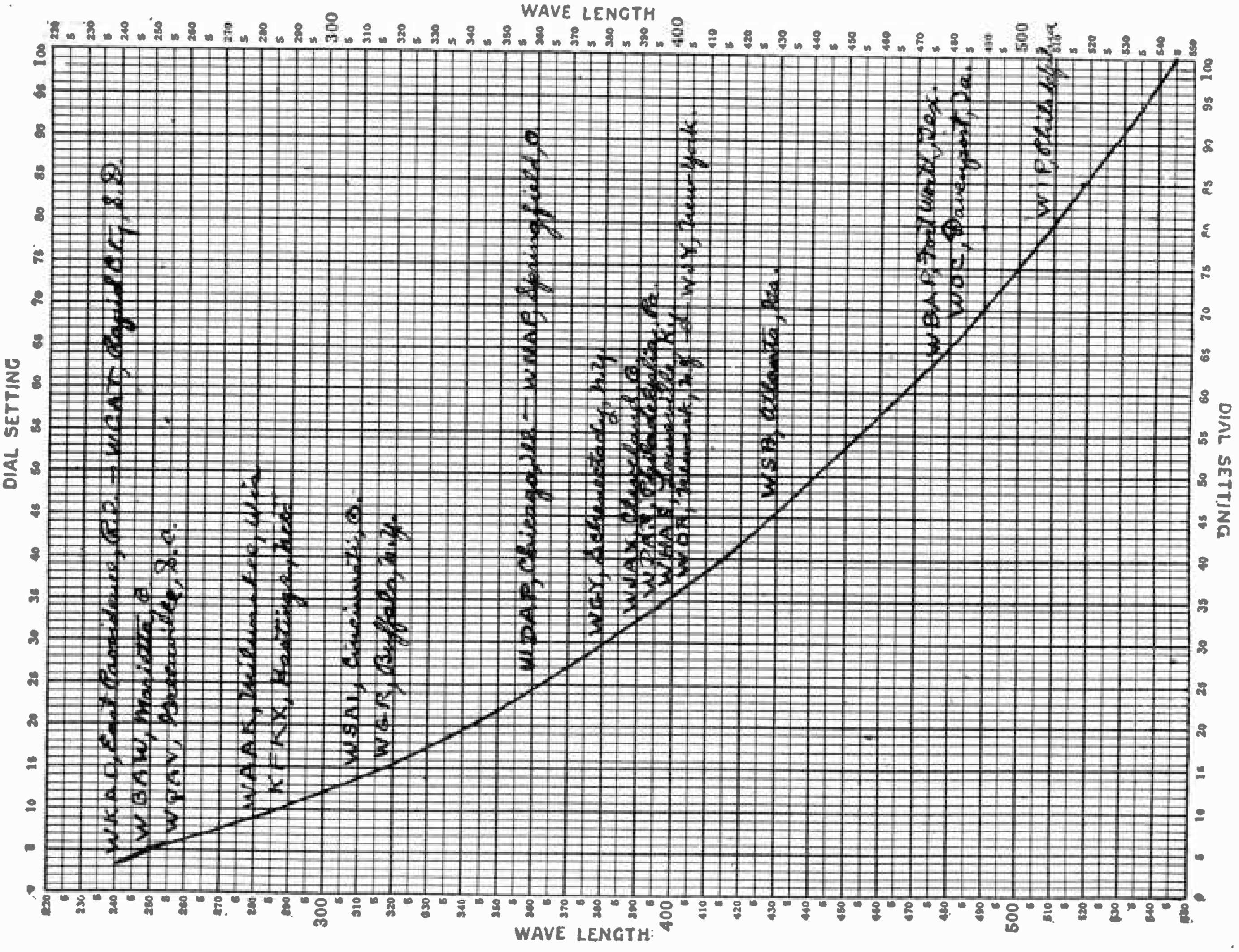

An example of a station log chart.

In single-control sets, where only one dial is employed it is only necessary to have one of these radio log curves. In receivers, such as the Neutrodyne or any two-stage radio frequency amplifier, where three dials are employed, it is generally necessary to log each dial separately. The reason for this is, that although these sets are supposed to tune to stations with identical dial settings, many of them do not do so, a difference as great as 10 divisions on the dial often being found. A three dial log is shown in Fig. 5. Separate charts may be used for each dial, or, for convenience, as has been done in Fig. 3, the three charts may be combined into one. If the difference between the dial settings is very small, say only a division or two, one curve will be sufficient.

Many things can be learned from studying the curves obtained. Take for instance, Fig. 2. This is practically a straight line at the higher wave lengths, changing from about 500 meters to about 400 in 55 divisions of the dial. In the remaining 35 divisions of the dial the wave length changes from 400 to 250. In the first case there was a range of 100 meters for 55 divisions and in the latter case a range of 150 meters in 35 divisions. This is because the curve bends downward. At the low dial settings a small change in the dial reading means a relatively large change in wave length, while at the upper end of the dial the same change in dial reading will mean a relatively small change in wave length. The reason why the stations are so crowded at present, at the small end of the dial, is then very apparent, i.e.:

To illustrate, a station operating on a 200 meter wave corresponds to a frequency of 1500 kilocycles (1 kilocycle equals 1000 cycles), and a station on 200 meters corresponds to a frequency of 1,363.5 kilocycles. In other words the difference in wave length is only 20 meters whereas the difference in kilocycles in 136.5. These two wave bands are at the low end of the present broadcast wave band. Now at the high end of the band we find a wave length of 600 meters corresponds to 500 kilocycles, and a wave of 580 meters corresponds to 517.2 kilocycles, or a difference of but 17.2 kilocycles for the 20 meter difference in wave length. At the lower end of the broadcast band between 200 and 220 meters we found a difference of 20 meters equalled 136.5 kilocycles. This indicates that as the wave length becomes lower the difference in frequency becomes very high.

To bring this most strongly to the mind, go down to the extremely low wave of 5 meters. Here the frequency is found to be 60,000 kilocycles. Going up to 25 meters, this corresponds to 12,000 kilocycles, so that for this difference of 20 meters there is the tremendous difference of 48,000 kilocycles.

The accompanying illustration, Fig. 6 will explain the matter very clearly. In changing the dial setting from 100 to, let us say, 90, the wave length is changed roughly from 600 to 580 meters. This is a change from 500 to 517 kilocycles. On the other end of the dial from zero to about 10 we have a change of wave length of roughly 200 to 220 meters, an approx-mate change of from 1500 to 1363 kilocycles. In the first case on the upper part of the dial we have only made a change of 17 kilocycles in the frequency. On the lower end of the dial we have made a change of 137 kilocycles.

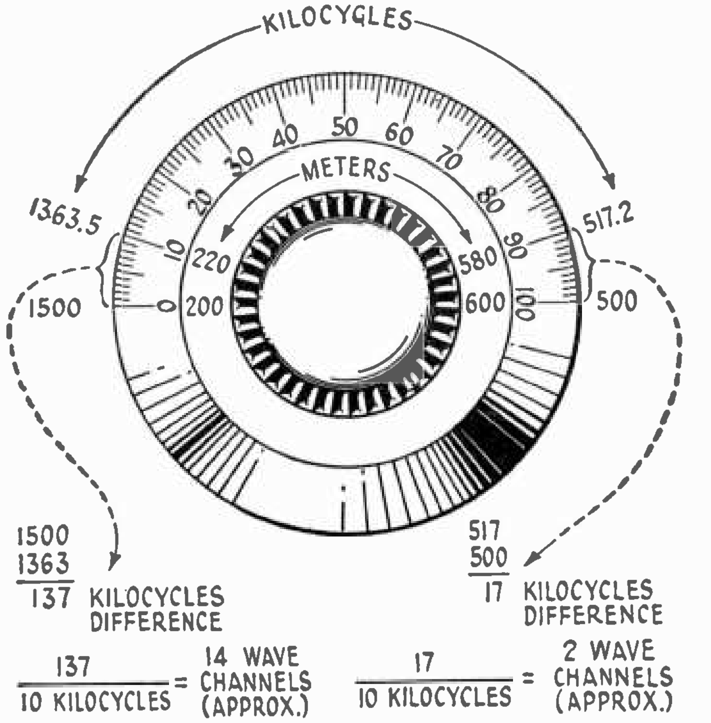
KILOCYCLES
METERS
1363.5
517.2
1500
500
0
10
20
30
40
50
60
70
80
90
100
200
220
580
600
1500
1363
137 KILOCYCLES DIFFERENCE
517
500
17 KILOCYCLES DIFFERENCE
137 / 10 KILOCYCLES = 14 WAVE CHANNELS (APPROX.)
17 / 10 KILOCYCLES = 2 WAVE CHANNELS (APPROX.)

Fig. 6—This illustration shows clearly why stations are crowded on the dial at the lower wave lengths.

According to the frequency allocation as made by the Government, which places broadcast frequencies 10 kilocycles apart, on the upper 10 divisions of the dial there is room for approximately 2 frequency channels whereas on the lower 10 divisions of the dial there is room for roughly 14 frequency channels. This shows very clearly that we would have to be able to separate 14 stations on the lower 10 divisions whereas we have only 2 to separate on the upper 10 divisions of the dial.

So it will be understood why turning the dials but a degree or two at the low setting causes a relatively large change in frequency, tuning in perhaps several stations within one degree of the dial, while at the higher settings, due to the small difference in frequency, as stated, on the higher waves, it may be necessary to turn the dials two or three degrees on either side of the peak of the broadcast wave before it is tuned out. (See *Dial.*)

LOGARITHM—abbreviation LOG—The power to which a number (defined as a base) must be raised in order to equal a given number. As an example, given the number 100, assume that it is desired to determine the *logarithm* of this number to the base 10. (When 10 is used as a base, the *logarithms* obtained are said to belong to the *common* or *Briggs system.*) Then the exponent or power which will raise 10 to the value of 100 is known as the *logarithm* of 100. Obviously, 10^2 equals 100, therefore 2 is the logarithm of 100. *Logarithms* are used in the processes of multiplication, division, raising to powers and taking roots. The whole number, or integer part of a logarithm is called its *characteristic* and the fractional part, is referred to as its *mantissa.* The characteristic of a number greater than unity is positive and is one less than the number of figures to the left of the decimal. The characteristic of a number less than unity is negative and is one greater than the number of ciphers between the decimal and the first significant figure. A *table of logarithms* gives the mantissas only, the characteristic being determined as indicated above.

Logarithms are also used which have as their base the number 2.718 +. These are called *natural* or *Naperian logs.* The base number is usually represented by the symbol ε (epsilon). The natural logarithm of any number may be found from a table of common logarithms by dividing the common log by the logarithm of 2.718 or simply by multiplying the common log by 2.30259. The symbol l_n is often used for the natural logarithm and the abbreviation *log* without a subscript is usually used for the logarithm to the base 10. (See *Naperian Log,* also *Mantissa,* and *Log Table.*)

LOGARITHMIC CURVE—A curve which has no definite minimum in a decreasing curve, or no definite maximum in an increasing curve. A *logarithmic curve* may be further defined as a curve whose ordinates increase arithmetically while its abscissas increase geometrically. A series of damped electrical oscillations diminish according to a logarithmic law and hence a curve drawn tangent to the successive maximum displacements will be a *logarithmic curve.*

LOGARITHMIC DAMPING—See *Damping.*

LOGARITHMIC DECREMENT—The *Naperian logarithm* (q.v.) of the ratio of any maximum to the next following maximum, with the current in the same direction in a decaying wave train. In other words the logarithm of the ratio of two maxima, one cycle apart. The rapidity with which the oscillations decay depends not only on the resistance of the circuit, but on the inductance as well. The greater the resistance and the smaller the inductance, the more rapid is the damping and the rate at which the oscillations increase. If the resistance, inductance and capacity of the circuit have fixed values, each successive maximum of current is the same fraction of the preceding maximum as the latter is of the maximum immediately preceding it. Thus, if the second maximum is 75 percent of the first maximum, the third will be 75 percent of the second, etc. The numerical value of the rate of decrease in this case is .75 and the natural logarithm of .75 is the *logarithmic decrement* or simply the *decrement.* (See *Decrement,* also *Damped Waves.*)

LOG TABLE—A tabulated arrangement of logarithms which gives the *mantissas* (q.v.) only. The numbers whose logarithms are to be looked up are arranged for convenience in consecutive order. (See *Logarithm.*)

LOMBARDI, DR. LUIGI—Italian radio authority. Dr. Lombardi was born at Dronero, August 21, 1876. He graduated from the Royal Engineering School at Turin, Italy. He was Professor of Electricity at the Zurich Polytechnical School, Zurich, Switzerland, and also at the Royal Polytechnic School, Naples, Italy. Dr. Lombardi published many scientific papers including a study of condensers for trans mission and he also invented a special high-tension electrical condenser.

LONGITUDE—The distance east or west of a meridian passing through Greenwich, England, measured in degrees, minutes and seconds.

LOOP ANTENNA—*Loop Aerial*—Also called *Coil Aerial* or simply *Loop*—A receiving or transmitting *aerial* (q.v.) consisting essentially of one or more turns of wire connected to the radio set so as to form a closed circuit. The *loop* acts as a simple inductance coil. The ordinary radio aerial acts as a huge condenser with the aerial system as one plate of the condenser and the *ground* or *counterpoise* (q.v.) as the other. Electromagnetic waves reaching the aerial set up an alternating electromotive force between the wires forming the upper plate of the condenser and the ground or lower plate of the condenser. This action takes place through *electrostatic induction* (q.v.). In the *loop aerial, electromagnetic induction* (q.v.) sets up an induced electromotive force thus causing alternating current to flow to the detecting apparatus. While coil antennas may be used for transmitting, their use in this connection is rather limited. On the other hand, the use of the loop for radio receiving is widespread. The greater the area enclosed

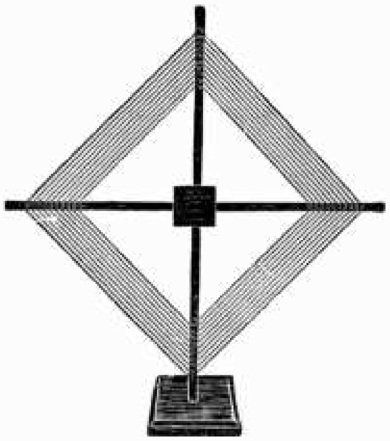

A typical loop antenna.

by a loop the greater the amount of energy it will receive. Loops more than two feet in diameter are usually constructed so as to be collapsible. A common type of loop shown in the illustration is 36 inches high and 28 inches wide. It consists of twelve turns of wire wound in the same vertical plane and held in place by polished hard rubber strips. It has a heavy cast base permitting rotation of the loop as desired. This loop folds up so as to fit into a four inch tube. In another type of loop, sometimes called the *box type*, the wires are wound so as to lie in the same horizontal plane instead of the vertical plane. The use of *Litzendraht* (q.v.) for winding loops has been found to add to their efficiency. Loops are commonly used on *super-heterodyne* (q.v.) sets, *reflex* (q.v.) sets and in some *tuned radio frequency* (q.v.) sets. These latter generally require an extra stage of amplification, although tuned radio frequency sets having only five tubes (detector, two stages of radio frequency and two stages of audio frequency) will operate on a loop and give especially good results when used with the new power tube. As a general rule, it may be stated that any set which will give good results with a loop, will give better results, insofar as distance is concerned, with an outside aerial. For this reason loop operating sets are quite generally constructed with extra binding posts for connecting aerial and ground. The loop is connected directly into the grid circuit, while the aerial and ground are connected to the circuit through an inductance such as an *oscillo-coupler* or antenna coupler. Loops may be used with satisfactory results inside of buildings. Where a compact, portable and selective aerial is desired, the loop is especially useful. The use of the loop, results in a marked directional effect and also has a tendency to reduce *static* (q.v.) and other forms of *interference* (q.v.). Loops are employed in wireless direction finders or *radio compasses* (q.v.) and are also used as aerials in submarine and aeroplane work. Where the loop is used as an aeroplane antenna it may be wound on the wings of the plane or it may be wound on a small frame and placed in the rear of the plane. (See *Coil Antenna*, also *Frame Aerial.*)

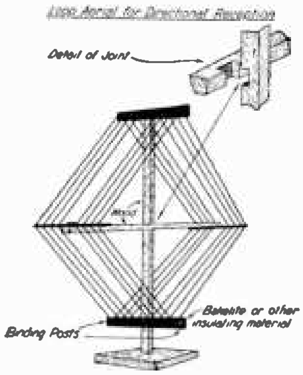
Loop Aerial for Directional Reception
Detail of Joint
Bakelite or other insulating material
Binding Posts

A box-type loop.

LOOPS—The points of greatest amplitude in a wave train. Sometimes called *antinodes*. The term *loop* is also used in describing the electrostatic field spreading out from a charged aerial. Each time an aerial is charged (for example by an induction coil) and discharged across a spark gap a group of gradually decaying oscillations is produced in the aerial called a *train* of oscillations. The interval of time which elapses between a flow of the electricity in the aerial in one direction, and a succeeding similar movement is called the *periodic time* and the reciprocal of this is called the frequency. For example, if the periodic time is one-millionth of a second, the frequency is one million. The diagram shows the manner in which *semi-loops* of electric force represented by the dotted lines move away from the aerial which is represented by the short black line.

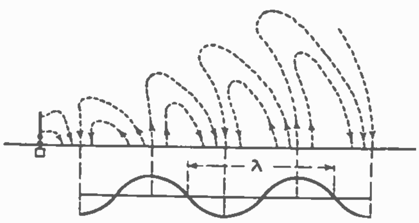

Semi-loops of electric force represented by the dotted lines moving away from the aerial represented by the short black line.

LORENZ COIL—A form of low loss inductance coil, having an air core and of cylindrical shape, being wound in a single layer, with turns slightly kinked to make the coil self-supporting. (See *Coil, Inductance Coils*, also *Low Loss Coils.*)

LOUD SPEAKER—A sound producing device for converting audio frequency currents into sounds loud enough to be heard by an assemblage of people. Nearly all *loud speakers* utilize an electromagnet in order to directly or indirectly cause a diaphragm to vibrate.

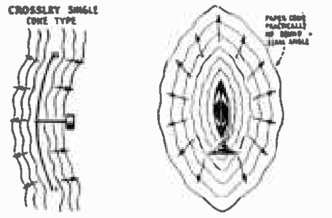

At the left is shown a cone-type speaker having a single diaphragm; at the right one with a double diaphragm.

The audio frequency currents flow through the windings of the electromagnet and the vibrations of the diaphragm correspond to the minute changes and fluctuations of the current thus accurately reproducing sounds. In loud speakers which utilize a metallic diaphragm, the electromagnet directly actuates the diaphragm. In other types, the diaphragm may be of mica, as in the Baldwin unit type, or of parchment or wood as in the cone type speakers. In these speakers, the electromagnet actuates an armature which in turn actuates the diaphragm through a lever action. In certain loud speakers, the electromagnet is of very powerful type and has a very small coil of fine wire attached to a diaphragm but not touching the poles of the magnet. The small coil is connected to the radio set or amplifier and the fluctuating audio frequency currents pass through it. These currents force the coil out of the strong magnetic field in the air gap. In moving, it moves the diaphragm to which it is attached. Speakers of this nature are sometimes called electrodynamic speakers. They are also referred to as power speakers. Such speakers give plenty of volume and in certain makes are reasonably good in tonal quality but they have the disadvantage that the electromagnet must be energized from a source of current such as the storage battery. A loud speaker has been developed in England called the Crystavox for use with crystal sets. This consists essentially of an electromagnetically operated speaker connected with a microphonic relay. The relay is actuated by a storage battery. Its function is to magnify the incoming signals from the crystal set, passing them along with increased strength to the electromagnet of the loud speaker. Still another type of speaker which has been used to a limited extent abroad, dispenses entirely with the electromagnet as the actuating source of diaphragm vibration. In this type, electrostatic principles are involved. An agate cylindrical drum, over which a band of copper or similar metal is wound, is revolved at a uniform rate by a small motor or by clockwork. To one end of the band a diaphragm is fastened mechanically. The other end of the band is secured by a fairly stiff spring. Current from the receiving set is led to the cylinder by means of a brush contact and as the current fluctuates the attraction of the band to the cylinder fluctuates with it and the drag on the diaphragm due to the rotation is increased and decreased. The use of the motor constitutes the most serious disadvantage in this type of loud speaker.

Having discussed the various *types* of loud speakers, from the standpoint of principles of operation, the more important of these will now be considered from the constructional standpoint. The early forms of directly actuated metallic diaphragm, electromagnetic loud speakers consisted essentially of a watch-case telephone receiver, or a pair of receivers attached to a horn. The purpose of the horn was to concentrate and redistribute the sounds caused by the vibration of the diaphragm. One of the first improvements was to modify the shape of the horn so as to make it conform more nearly with correct acoustical principles. Another improvement was to introduce an adjustment by means of which the air gap could be varied. In some cases this was accomplished by moving the magnet closer to the diaphragm and in others by moving the diaphragm itself. This adjustment provided a means of getting greater sensitivity on distant stations (by drawing diaphragm closer to the magnet) and of preventing sticking of diaphragm or rattling, on local stations. The ordinary metallic diaphragm has *one* frequency to which it will respond most strongly. This is called its *resonant* frequency. Special speakers

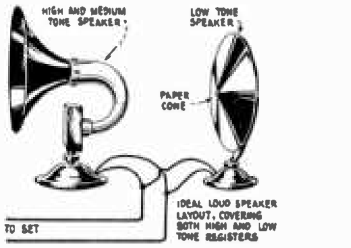
HIGH AND MEDIUM TONE SPEAKER
LOW TONE SPEAKER
PAPER CONE
IDEAL LOUD SPEAKER LAYOUT, COVERING BOTH HIGH AND LOW TONE REGISTERS
TO SET

A high and a low pitched speaker connected in parallel for simultaneous operation.

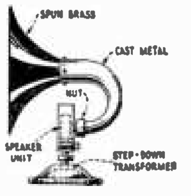

A loud speaker of the horn type having curved cast metal neck and spun brass bell.

called *tuned receivers* have been designed in England which permit the resonant frequency to be varied. In one type, the variations of the magnetic field operate a vibrating reed attached to a non-magnetic diaphragm. The resonant frequency can be varied by adjusting the position of the vibrating reed with a set screw. Another method of getting rid of resonant points in diaphragms has been to use corrugated instead of flat diaphragms. In some of these diaphragms the center is left flat, the remainder of the diaphragm being corrugated with concentric circles either evenly spaced or spaced in section at radii bearing a ratio to each other corresponding to prime numbers. In some cases conical diaphragms with corrugated edges are used. The material as well as the shape of the horn have been found to materially affect the sound reproduction. Horns are constructed of metal, fibre, composition, hard rubber, wood, etc. In general the horn should be of a material which will not give out metallic sounds. The tone arm which carries the speaker unit at one end and the horn at the other is usually constructed of cast metal such as a lead alloy or aluminum. The requirement here is that it must be of an acoustically "dead" material. In some cases the speaker unit and sound chamber are concealed in an artistically designed cabinet thus forming a *cabinet speaker*. Loud speaker units are also made especially for use in conjunction with the sound chamber of a phonograph. These are referred to as *phonograph attachments*. They are ordinarily attached to the phonograph tone arm being put in place of the phonograph reproducer. Some phonographs are now made with provision for attachment of the speaker unit without removal of the reproducer. There is also an attachment on the market for use in connection with a phonograph which has provision for conveniently changing over from the reproducer to the speaker unit or vice versa. Cone speakers may be of the single cone variety or of the double cone style. In these speakers a metal rod extends from the magnet armature and is rigidly fastened to the center of the cone. The most common type of cone speaker has a parchment diaphragm. This seems to give excellent reproduction especially for tones in the lower register. Its points of disadvantage are that it is affected to some extent by moisture and temperature change and furthermore seems to bring out static interference more clearly than some other types. Another type of cone speaker uses a teak wood diaphragm. Some of these speakers have a small fastening screw at the point of attachment between the rod and the diaphragm. This is provided so that tension can be removed from the diaphragm during shipment and also if the speaker is to remain in a room where change in temperature may take place. A cylindrical parchment speaker which utilizes the same principle as the cone speaker, has recently been placed on the market. Its chief point of difference is that the parchment is not under tension as in the case of the cone speaker. Large volume loud speakers are provided with power amplifiers which are often an integral part of the speaker itself. One of the most recent loud speakers is contained in a large cabinet and contains apparatus for utilizing ordinary lighting circuits for operating the set and speaker. The power amplifier unit is also contained within the speaker cabinet and the design of the speaker is such that it can handle any volume without distortion. The rectifier is arranged to supply not only the necessary voltages for the power amplifier but also plate voltage for practically any type of receiving set. (See *Horn, Loudspeaker,* also *Distortion.*)

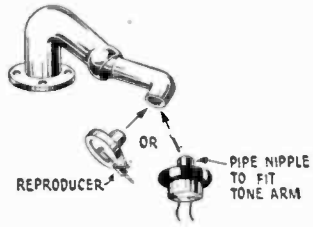

A loud speaker unit used as a phonograph attachment.

LOUDSPEAKER UNIT—The sound reproducing unit attached to the tone arm of the horn type speaker or to the diaphragm of the cone type speaker. Various types of units are described under the heading *Loudspeakers.*

LOW FREQUENCY—See *Low Frequency Current.*

LOW FREQUENCY CURRENT—An alternating current of less than 10,000 cycles. In radio work the alternating currents used are often classified as *low* frequency currents and *high* frequency currents. While there is no fixed value separating high and low frequencies, 10,000 cycles is generally taken as the dividing line. Thus audio frequency currents are considered as low frequencies and radio frequency currents are high frequencies. The term *low frequency* current as applied to power and lighting circuits refers to circuits having frequencies of 25, 60, etc., cycles per second. (See *Current, High Frequency.*)

LOW FREQUENCY IRON CORE INDUCTANCE—A variable inductance having an open-ended iron wire core. This inductance is used with spark transmitters to place the primary transformer circuit in resonance with the alternating current frequency.

LOW LOSS COILS—*Inductance coils* (q.v.) having low resistance and specially wound so as to reduce to a minimum condenser effect and hence dielectric losses. The term low loss has been used rather loosely but for practical purposes coil losses should be taken as referring to self-capacity, often called *distributed capacity* (q.v.) ohmic resistance, *skin effect* (q.v.) in the wire due to the effects of frequency and leakage and absorption in or through the insulating materials upon which the coil is built or with which the wire is covered. While it is impossible to obtain a coil having pure inductance only, it is feasable to design the coil so that the various losses will be reduced to a minimum. Often, it has been found that a design which will greatly reduce one type of loss will result in an increase of another. The illustrations show various types of low loss coils together with curves showing comparative efficiency tests. Figure 1 shows the results of an investigation regarding the distributed

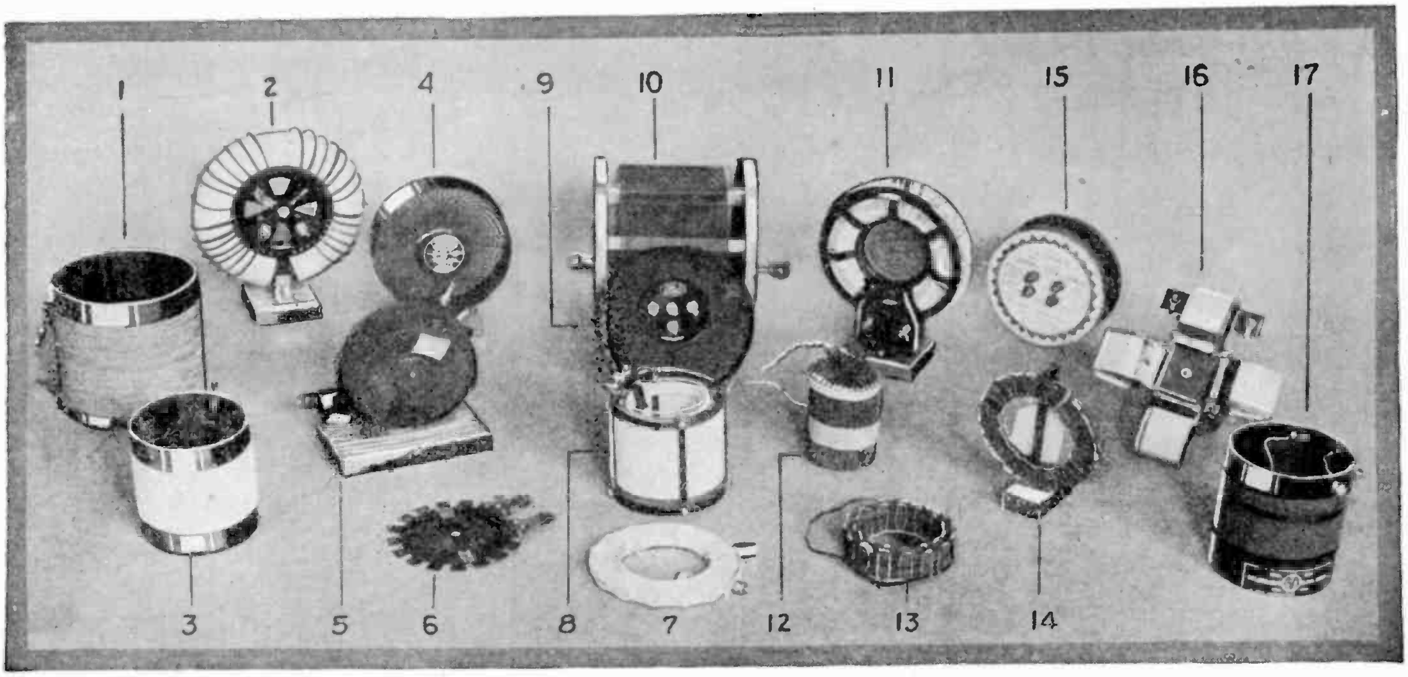

Various types of low loss coils including: 1. Bell wire coil; 2. Naxon; 3. Marco; 4. All-American; 5. Orbit; 6. Turney; 7. Sickles; 8. Aerocoil; 9. Erla; 10. Wavemeter coil; 11. Thorola; 12. Pathe; 13. Freshman; 14. Coast Coil; 15. Summitt; 16. Quadroformer; 17. Bruno.

capacity of various coils. This was obtained by making measurements of the inductance of the coils. If the self-capacity of the coil is very small, amounts to anything, the curve showing the relation between inductance and frequency will bend upward as the frequency increases, and the the self-inductance will not change perceptibly as the frequency changes. On the other hand, if the self-capacity greater the capacity of the coil, the more sharply will the curve bend. The curves shown in Figure 1 do not give any idea of what value the self-capacity may have, but simply show whether the inductance varies or not, which is one method of determining how harmful is the effect of the distributed capacity. The more sharply the curve bends upward the less desirable is the coil. In general, the smaller the size of the wire (for sizes smaller than about No. 16 B & S) the higher the resistance. The larger the wire and the thinner the insulation, the greater is the skin effect. The greatest inductance is obtained for a given amount of wire when the coil has a true cylindrical shape. The skin-effect in multi-layer coils is much greater than in single-layer coils. The effect of coil capacity and absorption or leakage in insulation is small compared with the skin-effect except in multi-layer coils, where the distributed capacity may become very great. To keep the physical size of toroidal coils within practical limits the diameter of the turns must be relatively small, so that many more turns are required to obtain a given inductance.

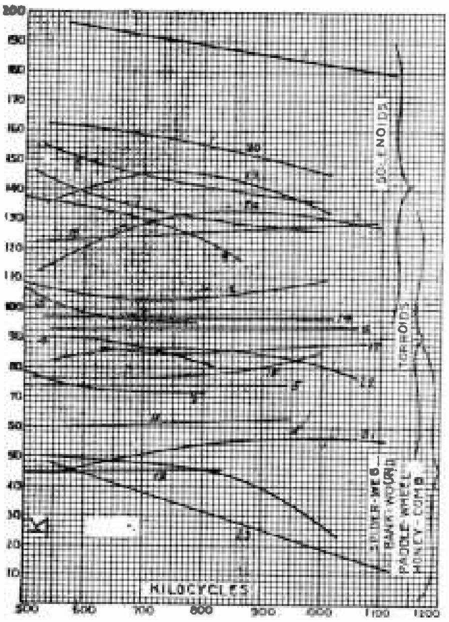

The efficiency of a coil is indicated by the height of its curves on this graph.

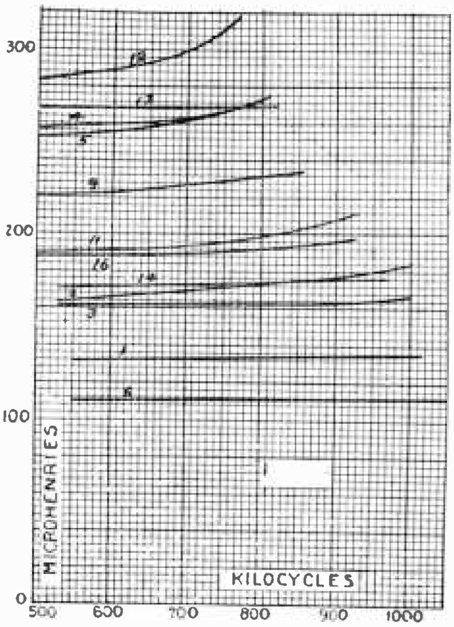

Fig. 1.—The curvature of the graphs indicate the presence of distributed capacity in the coils.

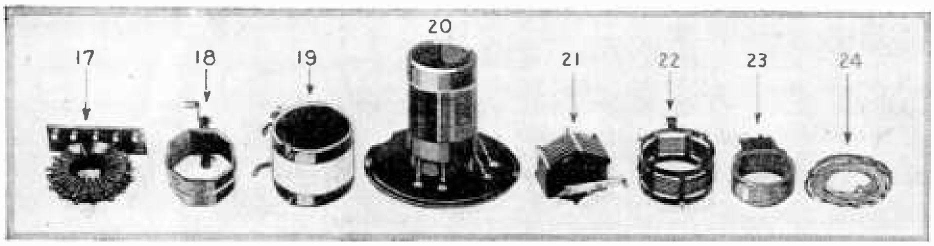

Additional examples of low loss coils: 17, Marwol; 18, Eastern Coil; 19, Workrite; 20, Walbert; 21, Andrews; 22, Bremer Tully; 23, Cotocoil; 24, Kresge Lorenz coil.

LOW TENSION — abbreviation L. T. — A comparatively low voltage. A circuit supplied by a low voltage battery such as the filament circuit of a vacuum tube is called a low tension circuit. In electrical engineering work, any circuit of 600 volts or under is considered to be a low tension circuit.

LUGS—Metal projections used for permanently connecting conductors with electrical apparatus. In one type of lug, the metal is bent over at one end to form a tube into which the conductor can be soldered, while the other end is flat and has a hole drilled to receive any desired terminal.

Typical lugs.

M

MAGNET—A piece of iron or steel having the power to attract other pieces of iron and to attract or repel other magnets. Natural magnets are known as *Loadstones* (q.v.). (See *Electromagnet*, also *Permanent Magnet* and *Temporary Magnets*.)

MAGNET WIRE — Insulated wire for the construction of magnet coils. In radio extremely fine wire is used to wind the coils of the electromagnets used in headsets and loud speaker units. The most common insulation used in radio work for magnet wire is enamel. Other types of magnet wire utilize a single cotton covering, a double cotton covering or a silk covering. (See *Enameled Wire*.)

MAGNETIC ATTRACTION AND REPULSION—The mechanical force tending to draw together two magnetic poles of unlike polarity, or to force apart two magnetic poles of like polarity.

MAGNETIC BLOWOUT—An electromagnet with its pole pieces placed so that the direction of the magnetic field will be at right angles to the movement of the ions in an electric arc. The coil of this magnet is energized in some cases in series with the line and acts to expel or blow out the arc. In the series arrangement, the blow-out action is stronger, the larger the current to be interrupted. For small shunted arcs, permanent magnets are often used. The operating principle of the blowout coil depends upon the fact that in a magnetic field, an electric current tends to move across the lines of force or in other words to cut them. In the illustration, the pole pieces are shown so placed that the general direction of the lines of force is at a right angle to the flow of the ions in the arc. The electromagnetic action causes the arc to move outside the field. This increases the length of the arc, thus rapidly breaking it. Magnetic arcs are used in connection with *keys* (q.v.) used for interrupting a large current, in certain types of *lightning arresters* (q.v.) and also in motor controllers..

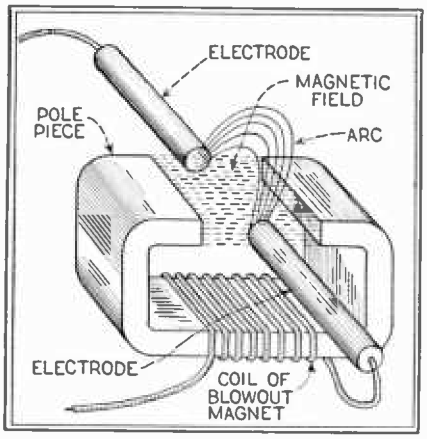

A magnetic blowout.

MAGNETIC CIRCUIT — The path traversed by magnetic lines of force. A magnetic circuit may be entirely through iron, as in the case of a closed core transformer or partly through iron and partly through air as in the case of a motor or generator. (See *Line of Force*.)

MAGNETIC COUPLING—See *Coupling*.

MAGNETIC COUPLING TRANSFORMERS—See *Coupling Transformer*, also *Jigger*.

MAGNETIC FIELD — The whole space over which a magnet exerts influence. The space traversed by the magnetic flux. Wherever magnetism is present the space in the immediate vicinity is called a magnetic field. (See *Electromagnetic Induction.*)

MAGNETIC FLOW—The passage of magnetic lines through a magnetic circuit. This term is sometimes used in the place of magnetic flux. (See *Flux Magnetism.*)

MAGNETIC FLUX—Symbol ϕ—The total magnetism present at any cross-section perpendicular to the lines of force. Magnetic flux is analagous to current flow. That is to say, it is convenient to consider a flow of magnetism in the same way as electricity is considered to flow. The unit of magnetic flux in the c.g.s. (centimeter gram second system) electromagnetic system is the *maxwell* (q.v.) or simply the *line*. The unit of magnetic flux in the practical system is also the *maxwell.* (See *Line of Force, Alternator,* also *Flux, Magnetism.*)

MAGNETIC FLUX DENSITY—The number of lines of force per unit cross-sectional area of a magnetic path. The *gauss* (q.v.) is the unit of magnetic flux density. It is equal to one *line* (q.v.) per square centimeter.

MAGNETIC FORCE—The force with which a magnet attracts or repels any piece of iron or steel. The intensity of a magnetic field at any point as measured by the force which it would exert upon a unit magnetic pole if placed at that point.

MAGNETIC FORCE DUE TO CURRENT—A magnetizing force due to current flow which can be determined mathematically for any point in the surrounding medium. Consider any closed stream line of electric current surrounded by a medium having uniform magnetic properties. It can be demonstrated that each elementary length of the current stream line, dl, contributes to the magnetizing force H, at any point T an amount dH =

$$\frac{(I \sin \theta)\ dl}{r^2}$$

where I is the current flowing along this stream line, r is the distance from T to dl, and θ (theta) is the angle between r and dl. The direction of dH is perpendicular to the plane of r and dl. The total magnetizing force at T is the vector sum of dH for all the elementary lengths into which the current stream line is divided.

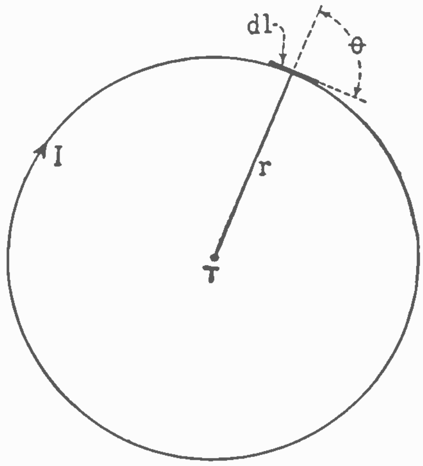

Diagram illustrating method of determining magnetic force due to current.

MAGNETIC HYSTERESIS—The opposition which minute particles of a magnetic material offer to being magnetized or to having their magnetization changed. It is sometimes referred to as molecular friction or as *magnetic lag.* (See *Hysteresis.*)

MAGNETIC INDUCTION—See *Induction,* also *Diamagnetic Material.*

MAGNETIC INDUCTIVE CAPACITY—symbol μ (mu)—Also termed Magnetic Inductivity—Names sometimes used instead of *Permeability* (q.v.). The ratio between the number of lines of force per unit area passing through a magnetizable substance and the magnetizing force which produces them. Stated another way, it is the ratio of flux density to magnetizing force. It can be considered as a measure of the ease with which magnetism passes through any substance. The magnetic inductive capacity of air is taken as unity. The term magnetic inductive capacity is analogous to *specific inductive capacity* (q.v.) which is the ratio of the dielectric constant of a material to that of a vacuum.

MAGNETIC INTERRUPTER—A device energized by an electromagnet for interrupting a circuit. (See *Interrupter.*)

MAGNETIC KEY—An electromagnetic adjunct to a wireless transmitting key for reducing or eliminating sparking due to the self-induction of the heavy currents which are interrupted. A form of magnetic key is shown in the illustration. Underneath the lever arm of the key is placed a spring to which an iron armature C and a platinum contact A_1 are attached. If the key and hence the spring are pressed down contact A_1 touches contact A_2 and the path of the current is completed. The current passes not only through the key, but also through the winding of the electromagnet. If the key is released so as to move upward, the contact A_1 still remains touching A_2, as the magnetic action due to the current continues to hold down the armature C and hence the spring with the contact A_1. Not until the alternating current reaches a zero value is the armature released. This releases A_1 from A_2, at an instant of zero current and hence there is no spark.

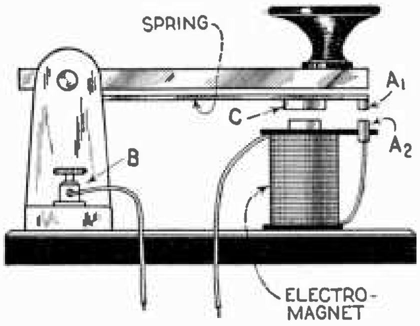

A magnetic key.

MAGNETIC LEAKAGE—Magnetic flux which serves no useful purpose. There are no substances which will prevent the passage of magnetic flux. This fact introduces a certain amount of difficulty into practical magnetic circuit calculations, since it is impossible to confine all the flux into a predetermined magnetic path. Since there is no magnetic insulator there will always be a certain amount of *leakage flux.* However this can be reduced to a minimum by proper design of the magnetic circuit. Magnetic Leakage is also referred to as *Stray Flux.*

MAGNETIC LINES OF FORCE—See *Line of Force,* also *Electromagnet.*

MAGNETIC MOMENT—The product of the strength of pole of a magnet and its virtual length. The torque exerted by a magnetic field upon a magnet depends upon the magnetic moment.

MAGNETIC NEEDLE—A light, thin steel magnet mounted on a pivot or suspended so as to be free to move. It takes a position pointing to the magnetic north.

MAGNETIC POLES—Points on a magnet where the lines of force leaving or entering the iron are concentrated. In a magnetic needle, that end which tends to point north is called a north or positive pole; that end which tends to point south is called a south or negative pole. A unit magnetic pole is defined as one which, when placed at a distance of one centimeter from a like pole of equal strength, will repel it with the force of one dyne. The north magnetic pole of the earth is situated in latitude 70 North and longitude 97 West. The south pole is located at latitude 70 South and longitude 102 East. It should be noted that the magnetic poles do not coincide with the geographical poles. (See *Induction, Magnetic.*)

MAGNETIC PROPERTIES—Materials are classified according to whether or not they offer a good path for the magnetic lines of force. Substances of comparatively low *reluctance* (q.v.) are termed *magnetic.* They are also often referred to as *ferro-magnetic.* They include iron, steel, nickel, cobalt, manganese, chromium, magnetite, certain alloys of copper, manganese and aluminum (called Heusler Alloys), and certain other oxides. The magnetic laws governing these substances can best be set forth by means of magnetization curves. Substances having high reluctance are termed *non-magnetic.* For commercial purposes, air is the most important non-magnetic substance. However, practically all substances, with the exception of the ferro-magnetic ones, follow the same magnetic laws as air. While the reluctance of a non-magnetic substance is constant, the reluctance of magnetic substances varies, as the amount of *flux* (q.v.) in the substance changes. Although air and all other substances, with the exception of the ferro-magnetic ones are classed as non-magnetic substances, and are poor conductors of magnetic flux, when compared to iron, etc., nevertheless they are by no means magnetic insulators. There are no substances which will absolutely prevent the passage of magnetic flux. (See *Ferro-Magnetic Substances.*)

MAGNETIC PYRITES—See *Pyrites.*

MAGNETIC STORM—Sudden and irregular variations of the earths magnetic field. Magnetic storms are sometimes coincident with the appearance of sun spots. These storms often disrupt telegraph service and also cause much interference with radio communication.

MAGNETISM—A phenomena exhibited by certain materials which is characterized by the attraction or repulsion of soft iron or of conductors carrying electricity. Magnetism is one of the manifestations of electricity, since a current flow is always associated or linked with a *magnetic field* (q.v.). (See *Free Magnetism, Magnet,* also *Magnetic Properties.*)

MAGNETISM, LAWS OF—Like magnetic poles repel one another; unlike magnetic poles attract one another. The force exerted between two magnetic poles is proportional to the product of their strengths and is inversely proportional to the square of the distance between them. (See *Like Poles.*)

MAGNETITE — symbol Fe_3O_4 — A mineral composed mainly of magnetic oxide of iron. When found in magnetized condition it is known as a *loadstone* (q.v.) or a natural magnet. It is a black brittle substance having a specific gravity of 5.2.

MAGNETIZATION—(also spelled *Magnetisation*)—The process of communicating magnetism. The act of rendering iron, steel, etc., magnetic. According to Ewing's theory, a substance which is magnetized has its molecules rearranged so that they assume symmetrical positions, each molecule lying in line with or parallel with its neighbor. In this rearranged position each molecule adds its separate magnetic force to every other one and the cumulative effect of this is that the substance exhibits the property of a magnet. (See *Demagnetization.*)

MAGNETO GENERATOR—A small generator consisting essentially of an armature rotating between permanent magnets, usually of the horse-shoe type. Magnetos are mainly used for motor ignition, medical coils, bell ringing and for firing high explosives. (See *Intermittent Current.*)

MAGNETOMOTIVE FORCE OR DIFFERENCE OF MAGNETIC POTENTIAL—abbreviation m.m.f.—The force causing magnetic flow. *Magnetomotive force* is the magnetic cause, while *flux* (q.v.) is the magnetic effect, just as in electrical terms, *electromotive force* (q.v.) is the cause and current is the effect. The c.g.s. (centimeter gram second) unit of magnetomotive force is the gilbert (q.v.). The unit ordinarily employed in practical calculations of the magnetic circuit is the *ampere-turn* (q.v.). (One ampere-turn is equal to 1.2566 gilberts.) An ampere-turn consists of a current of one ampere flowing through one complete turn of a conductor. The expression "turn" means that the conductor is wound in a solenoid. That is to say, it is wrapped around a core which may consist of air or of any other substance. The same magnetic field will result from ten ampere-turns consisting of ten amperes flowing through one turn, as from ten ampere-turns consisting of one ampere flowing through ten turns. (See *Force, Magnetomotive.*)

MAGNIFIER, OR RELAY—An apparatus for intensifying the variations of a current. The term *magnifier* has been incorrectly used for *amplifier* (q.v.). One type of *magnifier* used in radio is the microphonic relay (see loud speaker) by means of which the minute currents in a crystal set can be magnified so as to actuate a loud speaker. In the ordinary telephone repeater, a microphone is acted upon by the receiver of the first line circuit so as to introduce a local circuit which acts inductiveley on the second line wire. Relays are also used in wire telegraphy to boost weak line currents by means of strong local batteries. (See *Relay.*)

MANTISSA — The decimal part of a *logarithm* as differentiated from the integral part or characteristic. For example in the logarithm 1.23578, 1 is the characteristic and the decimal .23578 is the mantissa. (See *Logarithms.*)

MARBLE—The name given to any limestone sufficiently compact to permit being polished. Pure marble is white, the presence of oxides giving the different colors. Marble is used extensively for switchboard work and when so used must be free from metallic veins. Where circuits carrying 1000 volts or more are used, the switchboard should be saturated with an insulating varnish and baked. Inasmuch as the marble shows oil spots, it is sometimes stained black and given a finish called a *marine finish.*

MARCHANT, EDGAR WALFORD. British radio expert. Born in 1876, he was educated at University School, Hastings, and Central Technical College and London University. In 1897 he was appointed superintendent of the laboratory and workshops of Lord Blythswood at Renfrew, where he carried out many experiments in wireless telegraphy. He became chief assistant at Finsbury Technical College, 1900, under Professor Silvanus P. Thompson, and the following year was appointed lecturer in electro-technic at University College, Liverpool. In 1903 he was appointed the first professor of electrical engineering at the university.

Marchant was for some time closely associated with W. Duddell in developing the oscillograph, and the two read a joint paper on the study of the electric arc by the aid of oscillographs to the Institution of Electrical Engineers, of which institution Marchant became vice-president. He is also a vice-president of the Radio Society of Great Britain. He has written many papers on radio, and contributed an article to the proceedings of the Royal Society on the magnetic behaviour of iron under oscillatory discharge of a condenser.

MARCONI FILINGS COHERER—Also called Marconi Coherer—A detector of electric waves consisting of fine nickel and silver filings contained in a suitable glass tube between two electrodes consisting of two silver plugs. These two plugs are connected to platinum wires brought out through the sealed ends of the tube, which is carefully exhausted of its air. The metallic filings in a loose condition have feeble conductivity for the current from a single dry cell. If a very weak electric oscillation is passed through the coherer, the filings cling together or cohere and hence conduct electricity better and they will then pass sufficient current from a single dry cell to operate a telegraphic relay. To bring these filings back to a non-conductive condition, the tube is tapped by an electromagnetic tapper, similar to an electric bell with the gong removed. The Marconi coherer is now obsolete. (See *Coherer.*)

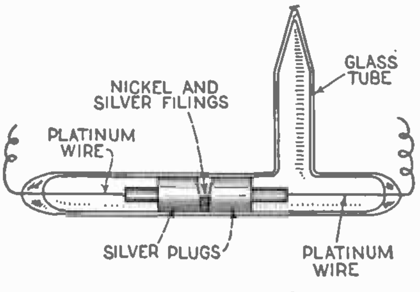

The Marconi filings coherer.

MARCONI, GUGLIELMO—Italian-Irish radio pioneer. Guglielmo Marconi was born at Bologna, Italy, April 25, 1874. and was educated at the Leghorn Technical School, under Professor Rosa. From his earliest years he was keenly interested in communication by means of Hertzian waves, and began his brilliantly successful series of experiments in June, 1895. These first experiments were carried out on his father's estate near Bologna. Marconi soon found that the Hertzian form of resonator gave only feeble signals at a distance, and he substituted a vertical wire, with the result that in 1895 he was able to transmit signals to a distance of one and a half miles. About the same time he improved the Branly coherer, which he was using as a detector, and invented an electric tapping device to decohere the filings.

This early apparatus of Marconi, from which has sprung the far-flung wireless chain of to-day, consisted of a coherer, a relay, a decoherer, and a Morse printing instrument, all working from storage batteries. Between the coherer and the relay Marconi interposed choke coils, and this had a very marked effect on the receptivity of his set. By close attention to the details of his system he was able to carry on his signals at greater ranges than had up to this time been accomplished by other experimenters.

The transmitting apparatus used by Marconi in these early efforts consisted of a large spark gap to which the aerial and earth wires were connected. The high-tension current for the spark was provided through an induction coil from batteries. The spark gap consisted of a ball discharger comprising four brass balls. The two middle balls were separated by a small space filled with vaseline oil, the actual spark jumping from the two end balls to the middle ones and through the vaseline, producing a high-frequency spark.

Guglielmo Marconi.

Marconi came to England in 1896, where he took out the first patent ever granted for a practical system of wireless telegraphy. He made a number of experiments at Westbourne Park, and demonstrated his system before Sir William Preece and other high officials at the Post Office. In the following year he increased the range of his set to nine miles using a 29 in. spark coil and kites to raise the vertical aerials. In July, 1897, in demonstrating before the Italian Government, he covered 12 miles between warships, and he began to instal a number of his sets for lighthouses, and the success of his experiments led to stations being erected for the corporation of Trinity House.

In 1899, the first proof of the advantages of wireless over other forms of communications came with the saving of the lives on board the ship R. F. Matthews, which ran into the East

Goodwin lightship. The latter was equipped with one of Marconi's transmitting sets and was able to communicate with the South Foreland lighthouse and summon assistance.

The genius of Marconi lies not so much in his inventions, as in his far-sightedness in being the first man to realize the immense commercial possibilities of radio, and to make the best use of all the scientific effort of his time to further the object he had in view.

Marconi has received innumerable honors from all countries. He holds the honorary degrees of universities all over the world, and the freedom of a large number of cities. In 1909 with Professor Braun, he was awarded the Nobel Prize for Physics, and in 1914 he was given the Honorary Knighthood of the Grand Cross of the Victorian Order. He has been awarded the Albert Medal of the Royal Society of Arts, the Gold Medal of the Institute of Radio Engineers of New York, the John Fritz Gold Medal for the invention of wireless telegraphy, and the Franklin Gold Medal of the Franklin Institute. He was the Italian delegate to the Peace Conference, and signed on behalf of Italy. He is chairman of the board of directors of the Marconi Company.

MARINE TYPE CHARGING PANEL—A switchboard equipped with resistance coils, carbon filament pilot lights, voltmeter and necessary switches. Manipulation of the switches allows the battery to be placed on charge or on the line. By throwing a switch on this board, it is also possible to work the wireless transmitter either from the ship's generator or from the storage battery.

MASS—The quantity of matter which a body contains. Physical quantities such as force, velocity, etc., are expressed in terms of length, *mass*, and time. The unit of length is the centimeter; of mass, the gram; and of time, the second. Hence this system of measurement is known as the C.G.S. (centimeter gram second) system. The gram is equal to 15.432 grains and represents the mass or quantity of a cubic centimeter of water at 4 degrees centigrade.

MASTER-OSCILLATOR SYSTEM—A radio telephone transmitting circuit in which a small extra *audion* (q.v.) is used to supply the power to excite the grid circuit. This extra oscillator or exciter is referred to as the master-oscillator. The master-oscillator system has several advantages over the self-excited system. Changes in antenna constants, such as might be caused by a swaying aerial or *lead-in* (q.v.) do not affect the transmission where the former system is used. In addition the master-oscillator system is more convenient to work with and the adjustment for maximum output for different wave lengths and antenna resistances is more easily made.

MASTER-OSCILLATOR SYSTEM AMPLIFIER—In the master-oscillator system, the power tube acts as an amplifier of the power supplied by the small extra exciting, or master-oscillator. The exciter must develop sufficient power to supply the losses in its own oscillating circuit and those of the grid circuit of the power tube.

MASTS, ERECTION OF—High powered stations generally use latticed steel masts. In some cases tubular metal masts in telescoping sections are used. Wooden masts have been used to quite an extent. For portable outfits, these are made in sections which can be fitted together like a fishing rod. Guy ropes or wires are necessary in practically every instance. In erecting a mast, it is usual not to make the installation too rigid, but rather to allow some freedom of movement so that the mast may sway slightly in the wind. Where a one-wire aerial is being erected this may be fastened to any support available. In marine construction, the ship's masts are used to support the aerial, the wires being stretched between the two booms or spreaders, from which halyards run to the masts. (See *Aerial.*)

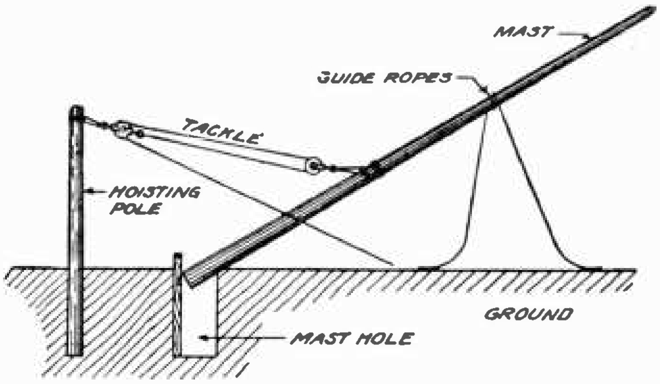

A method of erecting a tall mast using tackle and hoisting pole.

MATCHED IMPEDANCES—Circuits balanced or matched so that the total impedance of one circuit will equal the impedance of another circuit with which it is coupled. (See *Impedances, Matched.*)

MAXWELL—symbol φ—The unit of *magnetic flux* (q.v.). It is also known as the *line* (q.v.). One maxwell is equal to one line of force. (See *Flux Density*, also *Kiloline.*)

MAXWELL'S CORKSCREW RULE—If the direction of travel of a right-handed corkscrew represents the direction of current in a straight conductor, the direction of rotation of the corkscrew will represent the direction of the magnetic lines of force.

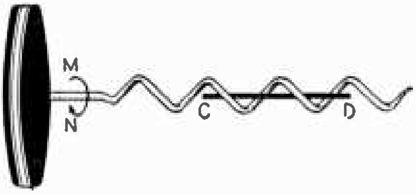

Illustration of Maxwell's corkscrew rule. When current flows from C to D, magnetic lines encircle the wire in the direction of the curved arrow MN.

MAXWELL'S LAWS—The magnetic fields surrounding parallel currents in the same direction will react upon each other in such a way that the conductors will tend to move together. In the case of parallel currents in opposite directions, the fields between the two conductors will be in the same direction and will merge together, thus tending to push the conductors apart. The magnetic fields around two conductors at an angle will react upon each other in such a way as to tend to bring the conductors parallel, and with the currents flowing in the same direction.

MAXWELL'S LAW OF MOTOR ACTION—A conductor carrying a current whose direction of flow is at right angles to a magnetic field, will tend to move out of this field.

MAXWELL'S RULE—Every portion of a circuit is acted upon by a force urging it in such a direction as to make it enclose the greatest possible number of lines of force.

MAXWELL TURNS—The deflection value in a ballistic galvanometer or *fluxmeter* (q.v.). This depends upon the number of turns in the exploring coil and the instantaneous value of the flux in maxwells linked with this coil. (See *Grassot Fluxmeter.*)

McLACHLAN, NORMAN W.—British radio authority. Born at Longtown, Cumberland in 1888, and educated at Carlisle Grammar School and the George Watson and the Heriot Watt Colleges, Edinburgh, and Liverpool University. He was appointed lecturer in Engineering and Mathematics at Newcastle-on-Tyne in 1909. In 1913 McLachlan was appointed supervisor of classes in engineering subjects in the Liverpool Technical Institutes, and after the Great War, during which he was engaged in aeronautical research and the study of anti-submarine devices, he made a special study of magnetos at the National Physical Laboratory, Teddington. Appointed research engineer to the Marconi Company, he is the author of many papers on wireless and electrical subjects in the journal of the Institution of Electrical Engineers and other scientific journals, including those on Characteristic Curves of a Poulsen Arc., the Magnetic Behaviour of Iron at very High Frequencies; and Theory of Iron-cored High Frequency Transformers. He has taken out a number of radio patents.

MEASUREMENT OF INDUCTANCE—Some methods of measuring inductance (also applicable to the measurement of *capacity*) are various bridge methods; by measuring the voltage, current, power, and frequency; and by the wave meter method. Connections in one bridge method for measuring inductance are similar to those in the ordinary *Wheatstone Bridge* (q.v.). A coil having a known inductance is balanced against the inductance to be measured. A device called a *Secohmmeter* is in use to increase the sensibility of bridge measurement of inductance. The Secohmmeter serves the purpose of making an alternating current to use in measurements of self-induction and of commutating such portion of this current as flows in the galvanometer circuit to a direct current. When a source of alternating current of known frequency is available the following method is convenient. Place the inductance across the alternating current mains, measure the power absorbed, the current flowing, the voltage across the unknown inductance and the frequency of the circuit. Then the inductance L is equal to the reactance, x, divided by

2 π times the frequency. Expressed as a formula

$$L = \frac{X}{2\pi f} = \frac{\sqrt{\left(\frac{E}{I}\right)^2 - \left(\frac{\text{Power}}{I}\right)^2}}{2\pi f}$$

The inductance L should be expressed in henries. This method gives results of fair accuracy but is not adapted to measurements of low values of inductance. Small inductances are more usually measured by bridge methods. In the wave meter method, a condenser is discharged through a spark gap in a circuit containing a known capacity and by means of a wave meter measuring the wave length. The expression for the number of oscillations in a closed circuit containing both inductance and capacity is given by the formula

$$n = \frac{1}{2\pi\sqrt{CL}}$$

and the wave length is equal to the velocity of propogation divided by the number of waves and calling λ the wave length we have

$$\lambda = V \text{ times } 2\pi\sqrt{CL}$$

In this last equation everything is known but the inductance which can be calculated. (See *Inductance*, Measurement of.)

MEASUREMENT OF WAVE LENGTH—The usual method of measuring an unknown wave length is by means of a wavemeter. Other apparatus required in this connection is a hot wire ammeter indicator in a coupled circuit, or crystal detector and telephones. Also an antenna excited oscillating circuit supplied by electron tube. A wavemeter placed in inductive coupling with a coil or antenna carrying radio current will show decided increase of current in its own coil and condenser when it is tuned to resonance with the source. The wave length is read directly from the wavemeter setting for resonance or from a calibration curve. A receiving set can be used to measure the wave lengths of received waves if it is first standardized in terms of wave lengths. The operator listens in through the head set attached to the receiving set. As the wavemeter condenser knob is turned the loudest sound is heard when the wavemeter circuit is tuned to the same wave length as that for which the receiving set is adjusted. The wave length is then read from the wavemeter scale or calibration curve. Continuing in this manner, the receiving circuit can be calibrated as a wavemeter by setting it at many different adjustments and reading the wave lengths at resonance each time. The wave length of the waves emitted by an antenna, when no added inductance or capacity is inserted in the antenna circuit, is known as its fundamental wave length. For a single vertical wire grounded antenna the fundamental wave length is slightly greater than four times the length of the wire. The approximate constant often used is 4 2 and this applies also to "L" or "T" type aerials with vertical lead in wire, the total length being measured from the transmitting apparatus up the lead-in wire and over to the end of the flat top. It is of course, more accurate to measure the wave length radiated from an antenna directly by use of a wavemeter. The wavemeter coil needs merely to be brought near the antenna or lead-in wire and the condenser of the wavemeter adjusted to give maximum current in the wavemeter indicator. The wave length corresponding to the wavemeter setting is then the length of the waves radiated by the antenna. The fundamental wave length of the antenna may be determined by gradually decreasing the number of turns in the *loading inductance* (q.v.), measuring the wave length for each setting of the loading coil and plotting a curve showing the wave length coresponding to the various numbers of turns of the loading coil. The fundamental, then, is the wave length corresponding to zero turns and is found at the point where the extension of the curve cuts the wave length axis. (See *Measurement of Inductance*, also *Formula.)*

MECHANICAL MIXTURE— A mixture of two or more elements which leaves them essentially the same. This is sometimes incorrectly called a mechanical compound. Sugar and sand mixed together give an example of a mechanical mixture. A *chemical compound* is formed when two or more substances combine chemically to form another substance having different characteristics. Thus iron filings and sulphur can be combined to form a third substance unlike either of the original two.

MEG OR MEGA—symbol M—The prefix placed before the name of a unit to denote one million times (10^6) that unit, as for example *megohm* (q.v.) *mega-volt, megerg*, etc.

MEGGER—An instrument for measuring insulation resistance or in general any resistance by the direct deflection of a pointer on a scale. The "Evershed" type is made in two forms (1) the *megger* insulation testing set and (2) the bridge-megger testing set which combines the functions of type (1) with those of a *Wheatstone Bridge* (q.v.). Both types consist of an ohmmeter of the moving coil type combined in one box with a hand driven generator for providing the necessary testing voltage and current. When the handle of the generator is not being turned, the pointer is entirely free and will rest anywhere on the scale. The general principles underlying the construction and operation of the megger are shown in the illustration. As can be seen, the megger is a combination of a *magneto generator* (q.v.) shown at the right, with an ohmmeter shown at the left, in a magnetic circuit common to both and consisting of two pairs of field poles braced by strong bar magnets, NS, NS, and forming two bi-polar field magnets in series. In the right hand one, and rotated by a folding handle and spur gearing D, is the armature of the generator, with its brush gear B, and terminal bars marked + and —. In the left hand field is the current coil A, pressure coil P, and compensating coil C of the ohmmeter, connected to resistances Q and R, a *guard plate* G, and the only two external terminals L and E (marked Line and Earth on the actual instrument). The bridge-megger testing set is available for use both as an insulation testing set and as a specialized Wheatstone Bridge. It differs from the ordinary megger in outward appearance only by the addition of two pairs of terminals and of two switches, one of which is the *ratio* switch of the Wheatstone Bridge and the other one is a two-way change over-switch. This latter when set to "Megger" prepares the instrument for measuring insulation resistance or high metallic resistance by coupling the two windings of the constant voltage generator in series and making the two front terminals, marked Line and Earth, the only two available for connection. When the change-over switch is set to "Bridge" the instrument is ready for Wheatstone Bridge work, the two windings of the generator now being connected in parallel in order to increase the current obtainable from it. In this case the ohmmeter part is changed into a galvanometer for the bridge and the arms of the bridge are switched into their appropriate places in the circuit and connected to the terminals marked R and X at the end.

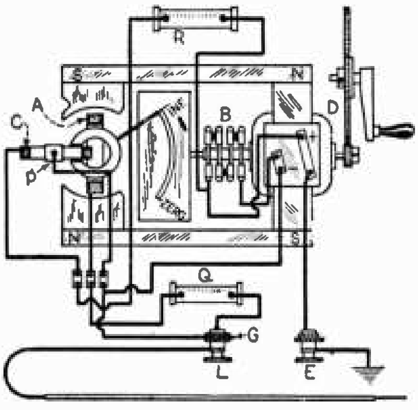

A diagrammatic illustration of a megger.

MEGOHM—One million ohms. This unit is usually employed in the measurement of insulation resistance. It is also used in measuring *grid leaks* (q.v.), etc. See *Ohm*, also *Begohm.)*

MEGOHMITE, also spelled Megohmit— An insulating material similar to micanite, obtainable in several forms, one of which is known as *hard megohmite* and is made of thin sheets of mica built up of shellac. *Flexible megohmite* consists of sheets of mica stuck together by vegetable adhesives. Flexible megohmite covered with Japanese paper is called mica paper while flexible megohmite covered with linen, is called *mica linen.*

MEISSNER, ALEXANDER.— Austrian radio expert. Born at Vienna, 1883, he was educated at the Technical High School and University of Vienna. He joined the Telefunken Company in Berlin, 1907, and is one of the leading authorities in Germany on radio. To him is due a large number of wireless inventions, including the Telefunken compass, musical quenched sparks, interference preventers, the direct current cathode valve relay for the reception of Morse Code, etc. In 1913 he invented the generation of oscillations by means of the three-electrode valve. He holds many patents, including spark gap for impulse excitation, quenched-spark signalling in conjunction with G. Von Arco, etc., and he is the author of many papers for scientific journals and societies on the subject.

MERCURY ARC RECTIFIER (MERCURY VAPOR CONVERTER)—A device for changing alternating current to direct current, which depends for its action upon the unidirectional conductivity of the mercury arc in an exhausted vessel. An idea of the principle of operation of the mercury arc rectifier may be obtained from the illustration. Assume an instant when the terminal H of the supply transformer is positive, the anode A is then positive and the arc is free to flow between A and B, B being the mercury cathode. Following the direction of the arrows still further, the current passes through the lead J, through the reactance coil E and back to the negative terminal G on the transformer. A little later when the impressed electromotive force falls below a value sufficient to maintain the arc against the counter electromotive force of the arc and the lead, the reactance E which heretofore has been charging, now discharges, the discharge current being in the same direction as formerly. This serves to maintain the arc in the rectifier until the electromotive force of the supply has passed through zero, reverses and builds up to such a value as to cause A' to have a sufficiently positive value to start an arc between it and the mercury cathode B. The discharge circuit of the reactance coil E is now through the arc A'B, instead of through its former circuit. Consequently the arc A'B is now supplied with current, partly from the transformer and partly from the reactance coil E. The new circuit from the transformer is indicated by the arrows enclosed in circles. The amount of reactance inserted in the circuit reduces the pulsations of the direct current sufficiently for all ordinary commercial purposes. The mercury rectifier is used in small sizes for charging storage batteries from alternating current circuits and in the larger sizes for operating direct current arc lamps from alternating circuits and also in electric traction work in the place of rotary converters in the substations or even on the train itself.

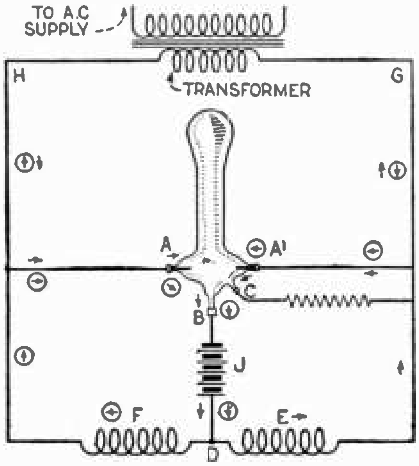

Illustration showing action of mercury arc rectifier.

MERCURY INTERRUPTER (MERCURY JET INTERRUPTER)—A contact breaker in which contact is made between moving metallic contacts and a jet of mercury provided by a small centrifugal pump driven by the same motor as the contacts. When the primary winding of an induction coil is supplied with a continuous current it is necessary to interrupt the current in order that a secondary voltage shall be induced in the secondary winding. The mercury interrupter offers a satisfactory method of accomplishing this. In one type of mercury interrupter a stream of mercury is forced from a jet against a metal plate. The stream of mercury is interrupted by a toothed wheel of insulating material so that the electrical circuit is periodically made and broken between the mercury and the metal plate. In some forms of mercury interrupters, the jet itself revolves, the mercury impinging against a fixed metal plate. The mercury interrupter is used in operating X-ray bulbs from an induction coil where the source is direct current. A form of mercury interrupter is shown in the illustration. A is a tooth-shaped wheel made of insulating material. This wheel is driven by a small motor and the number of revolutions can be varied within wide limits. A small centrifugal pump forces the mercury upon the jet B, while the revolving wheel A interrupts the contact of the mercury with the plate C. The wheel A can be raised or lowered, thereby enabling the duration of contact of the mercury with the plate to be varied without altering the number of interruptions in a given time.

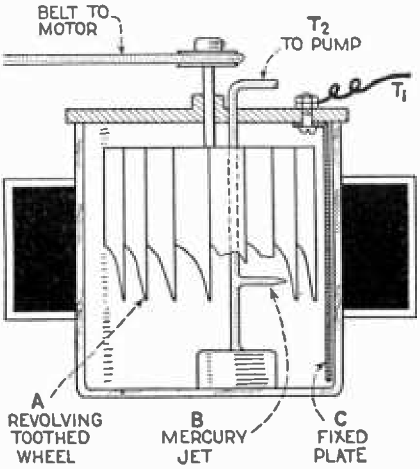

A type of mercury interrupter.

MESH GROUPING—A method of connecting coils in a polyphase circuit whereby they form a closed circuit and have the line wires attached to the points of junction between the coils. The term mesh is used interchangeably with delta. (See *Delta Connection.*)

METER—An instrument used to make measurements. (See *ammeter* or *amperemeter, double range meters, integrating wattmeter, voltmeter, wavemeter, wattmeter.*)

METER (*Measure*)—A unit of length in the *metric system* (q.v.). One meter is equal to 100 centimeters or 39.37 inches. It is equal to the length of a standard platinum bar, kept in Paris, and representing approximately a ten-millionth part of a quadrant of the earth's meridian measured from the equator to the pole through Paris. Wireless waves are measured in meters. The length of a wave is the distance, usually measured in meters between two points in the successive waves where the disturbance is at a maximum or at a minimum or between any two points of equal disturbance. There is a fixed relationship between the length of a single wave, the frequency of oscillation and the velocity of the wave. Thus if 3,000,000 waves pass a given point in one second, each wave must be 100 meters long, since the velocity of radio waves is equal to 300,000,000 meters per second.

METER-AMPERES—The product of the antenna current in amperes at the point of maximum current and the antenna effective height in meters, for any radio transmitting station. It constitutes a factor for indicating the radiating strength of radio transmitting stations.

METRIC SYSTEM—A system of measurement in which the meter is the fundamental unit of length and in which all the units both fundamental and derived are divided decimally and higher units are formed in multiples of ten. (See *Centimeter Gram Second.*)

MFD—Abbreviation for *Microfarad* (q.v.)

MICA—An anhydrous silicate of aluminum and potash or sodium. Mica has very high insulating qualities and can withstand high temperatures. Mica obtained in the natural state is separated into laminations which are sorted and graded. These are then cemented together to form plate or flexible mica of any desired thickness or purity. Mica is used in radio work as the dielectric of small fixed condensers. It is also used as the sound producing diaphragm in certain types of head sets and *loud speakers* (q.v.).

MICABOND CLOTH—India sheet mica faced on one side with muslin and on the other side with Japanese insulating paper bonded together with a special binder. It is used for insulating field coils and transformer coils.

MICABOND PAPER—India sheet mica faced on one side with Japanese insulating paper of one grade and on the other side with paper of a different grade.

MICABOND PLATE—Mica sheets bonded together with orange shellac. Various grades of plate are made some of which are used for magnet spools, others for insulating of commutator segments, while still others are used in electric irons.

MICANITE—A form of reconstructed mica. Canadian amber mica and white India mica are used.

MICRO—A prefix placed before the name of a unit to denote one-millionth part of that unit, as for example *microfarad* (q.v.).

MICRO-AMPERE—One-millionth of an ampere. A unit used in measuring extremely small currents. (See *Ampere.*)

MICROFARAD—abbreviation Mfd.—A unit of electrical capacity equal to one-millionth of a farad. Since the farad is too large for most practical measurements, the microfarad is generally used. (See *Farad,* also *Jar.*)

MICROHM—A unit of resistance equal to one-millionth of an *ohm* (q.v.).

MICROMETER—An instrument for making minute measurements. Usually controlled by an accurate screw of fine pitch.

MICROMETER SPARK GAP—A minute adjustable spark gap of about four-hundredths of an inch placed in the aerial circuit of a multiple tuner to allow heavy charges to pass readily to earth by sparking across its points.

Courtesy of Western Electric Co.

An example of the microphone used in modern broadcasting.

MICROPHONE—A sound magnifier. The ordinary form of microphone consists essentially of a diaphragm set in vibration by the sound waves and causing by its motion variation in the resistance of a mass of loosely packed carbon granules. This variation in resistance causes corresponding variations in the current through the instrument so that at every instant the form of the current corresponds to the variations in sound. This type of microphone is often referred to as a *Carbon Microphone*. The telephone transmitter is a form of microphone, and the microphone (nicknamed *Mike*) used in radio broadcasting is merely a larger and more sensitive variation of the same instrument. (See *Electromagnetic Microphone.*)

MICROPHONE, CARBON—See *Microphone*, also *Granular Carbon*.

MICROPHONE TRANSMITTER—See *Microphone*.

MICROPHONIC JOINT—A loose contact between two solid conductors, so mobile that feeble vibrations vary its resistance.

MICROPHONIC RELAY—An instrument for magnifying the variations of a given current by passing it through an electromagnet that acts on the diaphragm or reed of a microphone. The resistance variations are repeated in the current produced by a constant voltage in a circuit in series with the *microphone* (q.v.). In general a microphonic relay can be described as a microphone combined with a telephone so that a message transmitted over the telephone is repeated by the microphone over another line.

MIL—A unit of length equal to one-thousandth part of an inch.

MIL-FOOT—A wire one foot long having a diameter of one mil. Unit resistance may be measured in *ohms* (q.v.) per mil-foot. This refers to a conductor having a cross-sectional area of a *circular mil* (q.v.) and a length of one foot.

MILLI—The prefix placed before the name of a unit to indicate one-thousandth part of that unit, as for example *milli-ampere* (q.v.).

MILLI-AMPERE—One-thousandth of an *ampere* (q.v.). Sometimes abbreviated to *milli-amp*.

MILLI-MICRO—A prefix denoting one thousand millions. (10^{-9}). The term *Billi* is an alternate prefix.

MILLI-VOLT—One-thousandth of a volt.

MILS, CIRCULAR—See *Circular Mils*.

MIRROR GALVANOMETER—A galvanometer for measuring very small currents which utilizes a reflected beam of light for a pointer. In one type of mirror galvanometer a fixed nearly circular magnet is used and a coil of wire is suspended in its field. When the current flows through the coil it tends to turn the coil so as to set its axis in the direction of the magnetic force. This movement can be observed by means of a beam of light reflected from a mirror fixed to the coil. (See *Galvanometer.*)

MINERALLAC—An insulating compound made in several different grades all of which are impervious to water and show no alkali or acid reactions.

M.M.F.—Abbreviation for *Magnetomotive force* (q.v.).

MODULATED CURRENTS—Currents which have their amplitude periodically varied as in the case of the carrier current used in radio telephony which is varied or modulated in accordance with the vibrations of a microphone. If a sound wave actuates a microphone, its inward and outward displacement, varying the resistance in the aerial circuit, results in a high frequency current in the aerial of variable amplitude, called a modulated high frequency current. The illustration shows an extremely simplified diagram of the apparatus in a broadcasting station. The initial impulse that starts a radio signal towards the receiving set is furnished by the voice of the speaker or singer or by some other sound originating in the studio. This sound strikes the microphone, the operating element of which is a thin metal disk or diaphragm. The sound waves cause this diaphragm to vibrate. In doing so, the disk causes changes to take place in the electrical circuit. Pulsations of current are set up, the pulsations corresponding in strength to the variations of the spoken word, music or other sound as the case may be. These fluctuations are quite weak when they first come from the microphone and therefore they must be strengthened. They go through what is known as a voice amplifier which utilizes vacuum tubes to make the current stronger. The next piece of apparatus to consider is the *oscillator* (q.v.). This consists of one or more vacuum tubes, connected in a circuit of such a type that the tubes, when lighted and properly furnished with a high voltage direct current, will generate another current, alternating in character, which is said to be oscillating at radio frequency. Upon this current so generated, the voice current is impressed or super-imposed. This process is called *modulation* (q.v.). The current from the oscillator is known as the *carrier wave* (q.v.) current and when voice currents are impressed upon it, it is known as modulated radio frequency current. This current flows in the aerial and there sets up radio waves.

MODULATION—Variation of current or wave form to conform to sound waves or to other predetermined forms. Various forms of modulation are dot and dash modulation, chopper modulation, buzzer modulation and *speech modulation* (q.v.). (See *Distortion*, *Double Modulation*, also *Modulation Frequency Ratio.*)

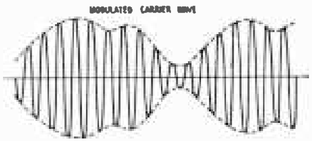

Graph of a modulated carrier wave.

MODULATION, DOUBLE—See *Double Modulation*.

MODULATION FREQUENCY RATIO—The ratio of modulation frequency to wave frequency. An alternating current is said to be *modulated* when the amplitude of its oscillations is varied periodically. The frequency at which the variations occur, that is to say the *modulation frequency* is necessarily less than the frequency of the alternating current which is being modulated. The nature of the variations may assume practically any form. As examples, there are dot and dash modulation, chopper modulation, buzzer modulation and *speech modulation* (q.v.).

MOLECULE—The smallest group of *atoms* of an element or compound which can exist by themselves. A familiar analogy states that if a drop of water could be magnified to the size of the earth, its component molecules would be the size of base-balls.

MOLYBDENITE—symbol MoS_2—A lead gray sulphide of Molybdenum. It is used as a rectifier detector in contact with copper.

MOLYBDENUM—Symbol Mo—A silver white metal found in nature as *molybdenite* (q.v.), a sulphide of molybdenum. When pure this metal is ductile and can be forged. It is used chiefly as an alloy to produce extremely hard steel. Molybdenite is used as a *crystal* (q.v.) in the detection of radio waves.

MORSE LIGHT—A search light used to signal code messages by means of intermittent flashes. (See *Code.*)

MORSE INKER—An instrument for recording Morse code signals in the form of dots and dashes on a moving paper band. In one form of Morse recorder, the attraction of the armature raises an inked wheel against the paper ribbon and thereby prints upon it the dots and dashes of the message.

MOSCICKI CONDENSER—A condenser in which the dielectric thickens out at the edges. It is essentially a form of *Leyden Jar* (q.v.) composed of a glass tube, especially thickened at the neck where the dielectric stress is greatest, coated inside and out with metal foil. The design of this condenser is such as to minimize *brush discharge* (q.v.) or *corona* (q.v.) effect at its edges.

MOTOR—A device for transforming electrical energy into mechanical energy. Motors may be designed for operation on direct current, on alternating current or on either direct or alternating current. These latter are known as *universal* motors. Direct current motors are classified as shunt, series and compound. This classification is arrived at according to the manner in which the field windings are connected. In the shunt motor the field winding is connected in parallel with the brushes (that is with the armature winding). In the series motor the field winding is connected in series with the armature and the line. Compound motors have a shunt winding in parallel with the brushes and a series winding in series with the armature and the line. Compound motors may be further classified as differentially compounded and cumulatively compounded. A differentially compounded motor is one in which the series field winding is so connected as to oppose the shunt field. A cumulatively compounded motor is one in which the series field aids the shunt field. The speed of the shunt motor is nearly constant, falling slightly as the load increases. The speed of the series motor decreases as the load increases. The differential motor may be constructed so that an absolutely constant speed will be maintained at all loads. The speed of the cumulative motor decreases rapidly as the load is increased. Alternating current motors are of the series type, the induction type, commutator type, synchronous type and repulsion type. There are a number of special classifications of alternating current motors such as single phase induction, polyphase induction, repulsion induction, etc. An ordinary direct current series motor used on alternating current will operate but at a low power factor, with a large eddy current loss and with violent sparking at the commutator. In order to adapt the series motor for universal use, it must be specially designed, the field structure being *laminated* (q.v.) in order to avoid *eddy currents* (q.v.) and the field coils having only a few turns to avoid too great *self-inductance* (q.v.) and consequently low *power factor* (q.v.). Sparking at the brushes can be avoided to a certain extent by designing the motor for a small field flux and with but a few turns in series in each armature coil.

MOTOR-ALTERNATOR—A *motor generator* (q.v.) set consisting of an *alternator* (q.v.) driven by a direct or alternating current motor. Such a set may be used for converting from direct to alternating current, or for frequency changing.

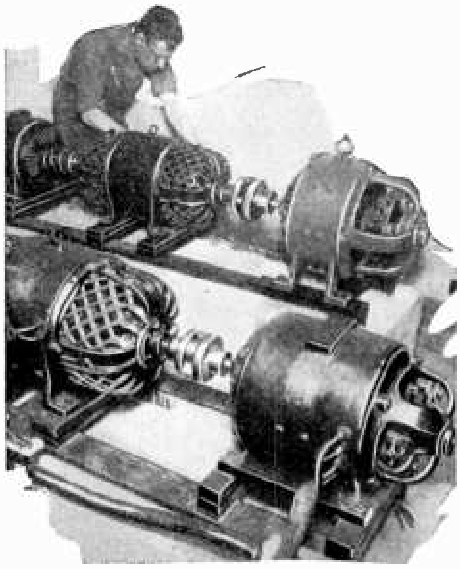

Motor Alternator set installed at Station WRNY.

MOTOR-GENERATOR—A generator or generators driven by an electric motor. Motor generators are used for converting alternating current to direct current or vice versa, for changing from one voltage or from one frequency to another or to obtain a variable from a fixed voltage.

MOVABLE COIL—Inductance coils which can be moved or rotated so as to vary the inductance. Examples of movable coils used in radio are vario couplers, loose couplers, etc. (See *Coupler, Loose; Coupler, Vario,* also *Fixed Coils.*)

MULLARD, S. R.—British radio authority. Born in 1884, he was educated as an electrical engineer and was ultimately appointed head of the research laboratory of Edison & Swan, Ltd., where he developed the Pointolite arc lamp. During the World War, he served in the Royal Air Force and became head of the wireless section laboratory at the Imperial College of Science for the Air Ministry. He carried out a series of researches on vacuum tubes and in 1920 founded the Mullard Radio Valve Company.

MULTIPLE—As applied to an electric circuit, a divided circuit. Two conductors are said to be connected in multiple or *parallel* (q.v.) when their two ends are joined together.

MULTIPLE SERIES—See *Series Parallel.*

MULTIPLIER—A resistance placed in series with a voltmeter for limiting and controlling the amount of current flowing through the voltmeter windings. This resistance may be either internal (enclosed by the voltmeter case) or external.

MULTI-POLAR—Having more than two poles. Usually a generator or motor whose field magnet has more than two poles. Small machines may be made with two poles, *bi-polar* (q.v.), but all others have a number of poles, that is to say they are *multi-polar.* The use of multi-poles results in a lighter machine and incidentally there is a saving of field copper and a reduction in sparking. If there are too many poles, however, there will be excessive *leakage flux* between the poles. (*See Homo-polar.*)

MUSICAL SPARK—A spark giving a regular distinct musical note. This may be produced by a high speed rotary discharger, a *quenched gap* (q.v.) or an arc.

MUSICAL SPARK SIGNALS—Signals in which the sparks occur at regular intervals of time and fast enough to give a musical note. Usually the spark rate is between 100 to 1200 discharges per second.

MUTE ANTENNA—A local circuit or resistor used in testing transmitting apparatus. The mute antenna in this case is substituted in the place of the actual antenna. It is also referred to as a *dummy aerial, phantom antenna, mock antenna,* or *artificial antenna.*

MUTUAL CAPACITY—The capacity effect of one conductor upon another one in the same electrostatic field. The mutual capacity is not the same as the capacity of the two wires regarded as the two plates of a condenser, one positively charged while the other is charged negatively. It really represents a decrease in the capacity of one of the wires with respect to earth caused by the presence of the electrostatic field of the other. The total capacity of the two wires is diminished by the overlapping of the two individual fields. (See *Capacity Measurement of Antenna.*)

MUTUAL INDUCTANCE—Symbol M—The magnetic flux that is common or mutual to two inductively coupled circuits. The mutual inductance of two circuits is the change in the interlinkage of flux that takes place in one circuit for a change of unit current in the other. This is also called the *coefficient of mutual induction* or the *mutual induction coefficient.* The units of mutual inductance are the same as those of *self-inductance.* (See *Inductance, Mutual; Self-Inductance, Induction, Coupling, Induced E.M.F., Induction, Mutual; Coefficient,* and *Mutual Induction Coefficient.*)

MUTUAL INDUCTION—The interaction between two current-carrying coils, not having direct metallic connection. The reactance of a coil such as shown in the illustration at A_1 is reduced by the proximity of another coil, A_2 if the circuit of the latter contains no external source of electromotive force, such as an alternator. The reason for this lessening in the primary reactance is that the flux excited by the first coil induces a current in the secondary coil and this current opposes the action of the primary current. As a result the total *magnetomotive force* (q.v.) the *flux* (q.v.) and consequently the *counter-electromotive force* (q.v) induced in the first coil are reduced. The effect of the secondary coil depends upon its distance from the primary coil and upon the amounts of the resistance R_2 and reactance X_2. If both coils are mounted on the same iron core, the action is greatly increased. In some instances mutual induction is desirable, as in the case of coupled radio circuits, or even in the ordinary transformer. In other cases, mutual inductance may be undesired as in the case where a radio antenna parallels a power line. (See *Induction, Mutual; Mutual Inductance,* also *Mutual Induction Coefficient.*)

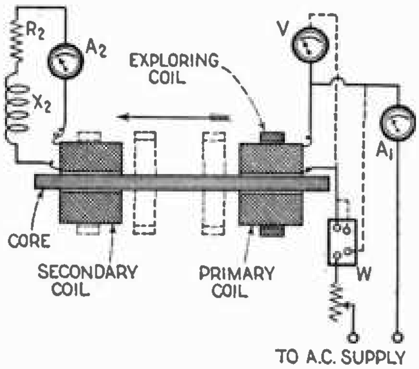

Diagram illustrating mutual inductance. The reactance of the coil A_1 is reduced by the proximity of coil A_2, since the circuit of A_2 contains no source of electromotive force. The effect of the secondary coil depends not only upon its distance from the primary coil but also upon R_2 and X_2. The presence of the iron core greatly increases the action.

MUTUAL INDUCTION COEFFICIENT—Another name for *mutual induction* (q.v.) That coefficient by which the rate of change of the current in a circuit must be multiplied to give the electromotive force induced in an adjacent circuit. From the principle of conservation of energy, it can be shown that the mutual inductance of a circuit A with respect to a second circuit B equals the mutual inductance of B with respect to A. (See *Mutual Inductance, Inductance, Mutual; Self-Inductance, Induction, Coefficient,* also *Coupling.*)

N

NAGYAGITE—A mineral containing lead, gold, antimony, sulphur, and tellurium. It receives its name from the location of its discovery near Nagyag in Transylvania. It is also known as black or leaf tellurium, and has been used as a crystal rectifier in conjunction with zincite.

NALLY, Edward Julian—American radio pioneer. Born April 11th, 1859, in Philadelphia, he joined the Western Union Telegraph Company in St. Louis and worked his way up in the telegraph service until in 1913, he was appointed vice-president and general manager of the Marconi Wireless Telegraph Company of America. Nally was one of the first to see the possibilities of wireless, and under his control the first commercial wireless communication was established between the United States and Japan, in 1914, and in 1920, he founded the first commercial wireless service between the United States and Great Britain, and afterwards to other countries. He was appointed the first president of the

Radio Corporation of America, and is a member of many scientific and other societies.

Edward Julian Nally.

NAME PLATE—A metal plate affixed to a radio set or other apparatus, giving the maker's name, trade name of apparatus, serial number or other details concerning the machine.

NAPERIAN BASE—The base of the so-called natural system of *logarithms* (q.v.). It is usually represented by the symbol ϵ (epsilon) and is numerically equal to 2.718 appx. (See *Damped Waves.*)

NAPERIAN LOG—See *Naperian Logarithms.*

NAPERIAN LOGARITHMS—Also called *Natural Logarithms.* Logarithms to the base ϵ. Logarithms constitute a tabular system of numbers, by which the operation of multiplication can be performed by addition, division by subtraction, involution by a single multiplication, and evolution by a single division. John Napier, Laird of Merchiston, Scotland, is generally regarded as the inventor of logarithms. The logarithms set forth by Napier were those of trigonometric functions. Later on Napier's logarithms were adapted to positive integers, using as a base the number 2.718 appx. and these logarithms are now called natural or Naperian logarithms. (See *Decrement*, also *Logarithms.*)

NATIONAL ELECTRIC CODE—A uniform code of rules, based upon the requirements of fire underwriters, for the electric wiring and electric installations in buildings. Unless these rules are compiled with fire insurance will not be issued or if issued previously will be voided. The National Electric Code contains a section covering special requirements for radio installations. The code is revised annually to meet new conditions and requirements. Copies of the code can be obtained from local insurance agents or from the National Board of Fire Underwriters at New York or Chicago.

NATURAL ELECTRIC WAVES—Wireless waves due to natural causes such as lightning discharges. When received in the radio apparatus, they are called *strays* (q.v.), *static* (q.v.), atmospherics, *X's* (q.v.), and various other names. There are many different causes for these stray waves. Some produce a grinding noise in the loud speaker, others a hissing noise often associated with snow or rain. Nearby lightning produces a sharp snap. Another type of strays cause loud crashes in the speaker.

NATURAL MAGNET—A substance which possesses in a natural state the properties of a magnet. Such a substance is the *lodestone*, a magnetic oxide of iron also called *magnetite* (q.v.). (See *Loadstone.*)

NATURAL FREQUENCY—The frequency at which free oscillations occur in a circuit. If L, C, and R are respectively the inductance, capacity and resistance of a circuit, then free oscillations in the circuit will have the frequency

$$f = \frac{1}{2\pi}\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}$$

This is the natural or *fundamental frequency* of the circuit. The natural frequency of an aerial is the frequency corresponding to the natural wave length. (See *Natural Wave Length*, also *Measurement of Wave Length.*)

NATURAL RECTIFIER—A mineral which possesses the property of conducting an electric current in one direction only. (See *Crystal*, also *Crystal Detector.*)

NATURAL WAVE-LENGTH—Length of the wave emitted by an aerial when no added inductance or capacity is inserted in the aerial circuit. This is also known as the *fundamental wave length* (q.v.). By putting *inductance coils* (q.v.) (*loading inductance*) (q.v.) in the aerial circuit, longer waves may be radiated, while condensers put in series with the antenna enable it to produce shorter waves than the aerial would ordinarily radiate. The use of a series condenser is avoided wherever possible since it has the effect of decreasing the total capacity of the aerial circuit and thereby decreasing the amount of power which can be given to the antenna. The addition of some inductance has a beneficial effect, since the decrement of the aerial is thereby lessened and a sharper wave results. (See *Measurement of Wave Length; Wave Length.*)

NAUTICAL MILE—abbreviation *Naut* —A marine unit of distance. Equivalent to one minute of longitude at the equator. This unit is also sometimes referred to as a *Telegraph Naut.* It is equal to 1.1528 statute miles (the statute mile being equal to 5,280 feet). The term *nautical mile* is used especially in submarine cable work, and also in matters pertaining to navigation. The term *naut* is differentiated from *knot* in that the former refers to a distance whereas the latter refers to a rate or unit of speed.

NEGATIVE CARRIER*—An electron combined with neutral gas molecules. When the velocity of the electrons in a vacuum tube is less than the value necessary to cause ionization by collision, the electrons attract the neutral gas molecules and so form heavy negative carriers. The ease with which this formation of negative carriers takes place depends on the nature of the gas. Such gases as argon and mercury vapor do not readily form negative carriers, while hydrogen and oxygen combine with electrons more easily. The effect of this negative carrier formation is to counteract the reduction of the negative space charge occasioned by the heavy positive ions formed by collision ionization. The positive ions are ions of the gas from which one or more electrons have been removed. The ions therefore have very nearly the same weight as the gas atoms. The negative carriers, on the other hand, may consist of an atom or molecule to which has been attached an electron. It is also possible that the attraction between an electron and the neutral gas molecules can result in the formation of clusters consisting of more than one molecule held together by the electron. These negative carriers, therefore, move as slowly as, and often more slowly than, the positive ions and consequently have a relatively great effect in counteracting the tendency of the positive ions to reduce the negative space charge of the electrons. (See *Electron Theory*, also *Ion.*)

NEGATIVE CORPUSCLE—The natural unit of negative electricity. The *electron* (q.v.). It is more usual to refer to the negative corpuscle simply as the *corpuscle.* This was the name given by Sir J. J. Thomson to the carriers of electricity shot off from the cathodes in vacuum tubes. The corpuscle was found to have a charge equal to the electron and a mass which is $\frac{1}{1845}$ of that of the hydrogen atom.

NEGATIVE ELECTRODE—In a *primary cell* (q.v.) the *cathode* (q.v.) which is the carbon, copper, etc., electrode, is the *negative electrode*, while the *pole* of this electrode is the positive pole, since it is positive in relation to the external circuit. In a *secondary* or *storage cell*, the spongy lead plate, which is the anode during discharge, is called the *negative electrode* and its pole the *negative pole.*

NEGATIVE ION—An atom, which is the smallest particle of an element capable of existing, plus an *electron* (q.v.) (See *Anion, Cathion, Electron Theory*, also *Ion.*)

NEGATIVE POLE—The south-seeking end of a magnet. In a generator, the terminal into which the current returns from the external circuit. In a storage cell, the terminal of the negative plate. In a primary cell, the pole of the positive plate.

NEGATIVE RESISTANCE—A current path within a vacuum tube or an arc in which current decreases as voltage increases. The emission of electrons from cold electrodes under the impact of electrons (known as *secondary electron emission* (q.v.) or *delta rays*) results in a negative resistance or *falling characteristic.* The presence of secondary electrons can be shown by referring to the illustration. The plate P is kept at a constant positive potential with respect to the filament F by means of the battery B. When there is no difference of potential between the filament and the grid G, the cur-

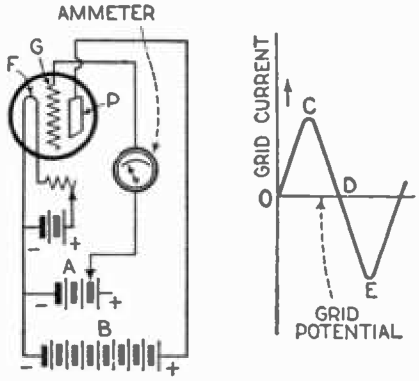

At the left is a hook-up for demonstrating negative resistance. The curve at the right shows the decrease of grid current although grid potential is increased.

(*Van der Bijl—The Thermionic Vacuum Tube.)

rent in the circuit FGA is very small, since practically all the electrons emitted from the filament are drawn through the openings of the grid and thrown on to the plate. If the potential of the grid (positive with respect to the filament) is increased, the current to the grid at first increases, as shown by the part OC in the accompanying curve. When the grid potential reaches a certain value, the current as indicated by the ammeter begins to decrease, drops down to zero at D and then reverses its direction of flow. The reason for this is that while the difference of potential between the filament and the grid is small, the electrons that strike the grid enter it, but as the positive grid potential is increased the electrons on striking the grid emit *secondary electrons* from it and these are drawn to the plate which is maintained by the battery B at a positive potential, with respect to the grid. The net current as shown by the ammeter is the sum of the electrons entering the grid and those leaving it. When the velocity with which the electrons strike the grid increases beyond a certain value, one primary electron can knock out more than one secondary electron from the grid and the current in the filament-grid circuit reverses. When the positive grid potential is increased to such an extent that the grid becomes positive with respect to the plate, the secondary electrons are no longer drawn away to the plate, but are driven back to the grid so that the reversed current in the grid circuit again decreases as shown at E on the curve, and finally assumes the original direction. Hence, considering the current as indicated by the ammeter and the voltage between filament and grid, it can be seen that over the portion of the curve CDE, the current decreases as the voltage increases. Therefore CDE represents a *negative resistance* characteristic. It has been found that a device which possesses a negative resistance can function as an amplifier and also as a generator of continuous oscillations.

NEGATRON—An English vacuum tube containing a filament, a grid and two plates. This tube was devised by J. Scott-Taggart, who used it for obtaining *negative resistance* (q.v.) characteristics. The two plates are fixed one on each side of the filament. The accompanying illustration shows a typical negatron. The bulb is tubular in shape and the four electrodes are plainly shown. The grid, in this tube, is a metal rod.

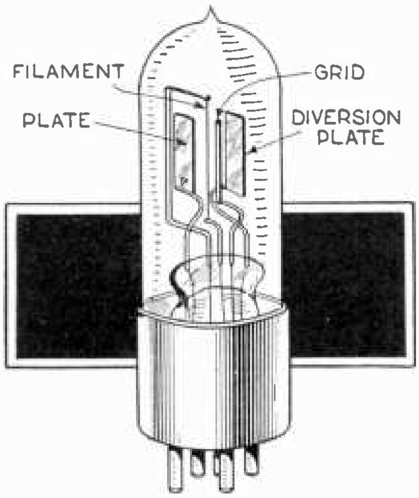

A negatron tube, showing the two plate construction.

NEON LAMP—An incandescent lamp in which a reddish light is produced by the incandescence of neon gas at low pressure. This lamp has very low current consumption and has been applied to pilot lights, signs, etc.

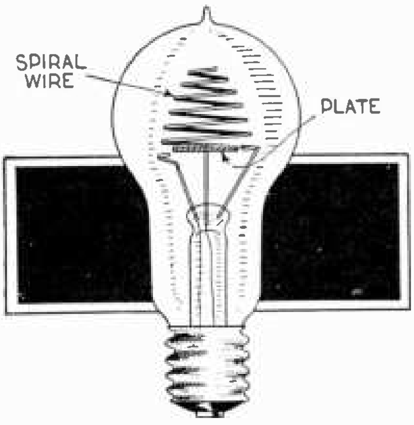

Neon Lamp

NEON TUBE—A vacuum tube containing the gas neon at low pressure used in the Fleming *Cynometer* (q.v.) etc. In this instrument, which is used to measure the frequency of electric waves, wave lengths, etc., a capacity formed by one brass tube sliding over another can be varied simultaneously with an inductance consisting of an air core solenoid and arranged so that the point at which resonance takes place is indicated by the glowing of the *neon tube* and can be read from a calibrated scale.

NERNST LAMP—An incandescent lamp in which the incandescent body consists of a strip of material composed of a mixture of oxides of metal such as zirconium, magnesium and other refractory oxides. The incandescent portion of the lamp is called the *glower* and is conductive only at a high temperature. Consequently it has to be heated by an auxiliary heating coil. The heating coil is automatically cut out of the circuit when the current starts to flow through the glower.

NEUTRAL BODY—A body void of electrification. A charged body which loses its electrification is said to be *discharged* or *neutralized*. All matter is capable of electrification.

NEUTRAL LINE OF MAGNET—The middle portion between the two magnetic poles of a bar magnet, where there is no manifestation of magnetism.

NEUTRAL WIRE—A middle wire, in a *three wire system* (q.v.) of power distribution, which is kept at a potential mid-way between the positive and the negative mains. In the three-phase system of alternating current distribution, the conductor connected to the neutral point of a "star" or "Y" connected alternator is also known as the neutral wire. The neutral wire is here also known as the *common return*.

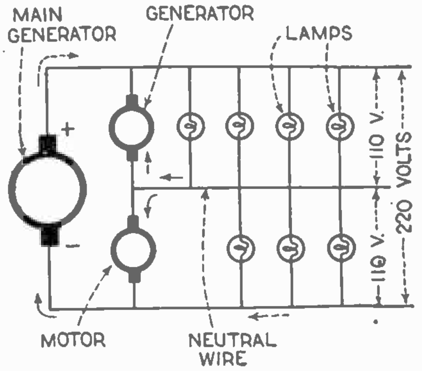

A three wire system with unbalanced load, showing neutral wire.

NEUTRODYNE—See *Neutrodyne Circuit*, also *Hazeltine Neutrodyne Receiver*.

NEUTRODYNE CIRCUIT—A circuit used in radio reception in which radio frequency amplification is used for neutralizing the effects of capacity of the tube and its socket. Figure 1 shows a portion of a receiving circuit. In this diagram the tube capacity is represented by a condenser shown in dotted lines. By studying this circuit, it can be seen that a closed oscillatory circuit exists, made up of the inductances and capacities in the plate and grid circuits. Just as radio frequency currents can pass through condensers in a receiving set, so they can pass through or across the vacuum tube from the plate to the grid, which forms a condenser. This condenser effect is the reason why radio frequency amplifiers oscillate. A difference of potential occurs in the plate side of the coil in the plate circuit. Instead of this being handed on to the next tube for additional amplification, a portion of this potential leaks through the capacity of the tube and its socket and affects the grid in such a way that trouble is experienced with self-oscillation. In the neutrodyne circuit, the

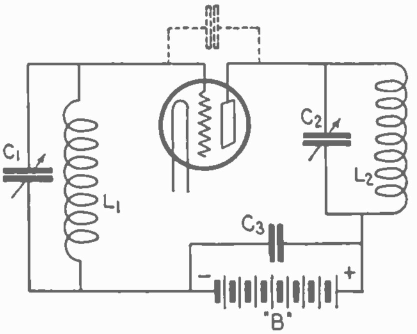

Fig. 1. A portion of a receiving circuit. Dotted line shows tube capacity represented by an equivalent condenser.

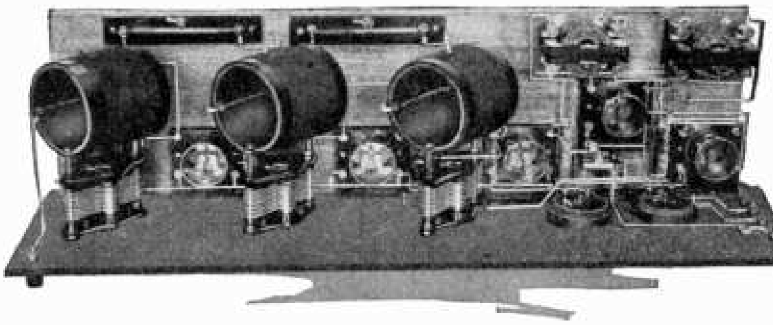

Fig. 2. A typical neutrodyne set.

inter-element capacity of the amplifier vacuum tubes is neutralized by means of special condensers called *neutrodons*. The capacity of these is very low, being approximately equal to the internal capacity of a vacuum tube. By reason of this equality, any tendency of a large amount of radio frequency current to pass back through the tube through the grid is defeated and instead is neutralized by the combination of the neutralizing capacities, the inter-element capacity of the vacuum tubes and the secondary windings of the tuned radio frequency transformers. This effect is in reality a bucking one, since the current is made to take two paths. Each neutralizing condenser must be adjusted so that its capacity will equal that of the vacuum tube it is connected with. Since it is impossible for a neutrodyne circuit to oscillate, there is no radiation of energy from the set, with consequent disturbance of other receiving sets in the vicinity.

The neutrodyne circuit was invented by Professor L. A. Hazeltine. The typical neutrodyne receiving set uses five tubes, employing two stages of tuned and neutralized radio frequency amplification, with detector and two audio frequency stages. (See *Hazeltine Neutrodyne Receiver.*)

NICHROME WIRE—A nickel-chromium-steel alloy wire, which is able to withstand a bright red heat in the atmosphere without oxidizing. Nichrome is practically non-corrosive. It has a very high melting point, about 1550 degrees Centigrade and is used extensively in electrical heating appliances and resistance elements.

NIGHT EFFECT—Changes in the strength of radio signals noticeable especially at night. This effect is also known as *fading* (q.v.). Fading does not generally occur in the immediate vicinity of the transmitting station. It seems to happen more often on wave lengths below 400 meters. A certain station is being received with normal intensity, when suddenly the sounds become very faint and in a few minutes they may again become normal or may even become stronger than usual. In some cases the variations from strong signals to weak take place every few minutes while in other cases there may be an interval of several hours. It has been observed that fading takes place over land more often than in radio transmission at sea.

NOBLE, Sir William—British engineer. Born in 1861 and educated at Gordon's College, Aberdeen, he thereafter entered the Aberdeen telegraph office. In 1893 he was appointed engineer for the northeast area of Scotland, rising rapidly in the service until he became chief engineer in 1919. He retired from the Post Office in 1922, and became a director of the General Electric Company and of the British Broadcasting Company. Sir William Noble has written many articles on telegraphy and telephony and is considered to be an authority on these subjects.

NODES—In a wave form, such as shown in the illustration which represents an alternating current sine wave, the zero points. Thus in the alternating current sine wave, the points of zero current or potential, at A, C, E, and G, are nodes, and the points B, D, and F are the anti-nodes or *loops* (q.v.).

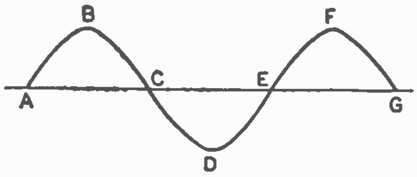

Sine wave illustrating nodes and anti-nodes.

NODEN VALVE—An *electrolytic rectifier* (q.v.) which allows only current in one direction to pass through it. This rectifier utilizes an aluminum rod or cylinder as a cathode and the anode may be of iron, lead or of carbon. Ammonium phosphate is used as the electrolyte. The principal of operation of this rectifier depends upon the fact that the current in the positive direction is stopped at the aluminum plate through the formation of an insulating crust of aluminum phosphate and aluminum oxide. These two substances present an extremely high resistance to the passage of the current in the positive direction, but the compounds dissolve again upon reversal of the current. By suitably combining two or more cells, both the half waves of an alternating current can be rectified. In Fig. 1 is shown the principle upon which the Noden Valve acts, and Fig. 2 shows a hook-up for connecting four cells to obtain full wave rectification. (See *Electrolytic Rectifier Full Wave,* also *Rectifier.*)

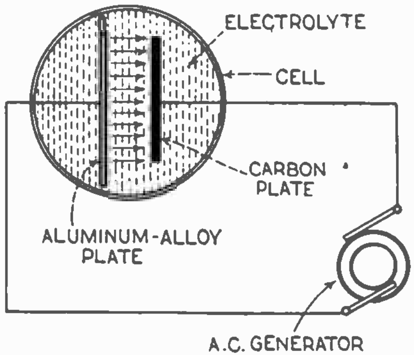

Fig. 1. Diagrammatic illustration of the principle of the Noden valve.

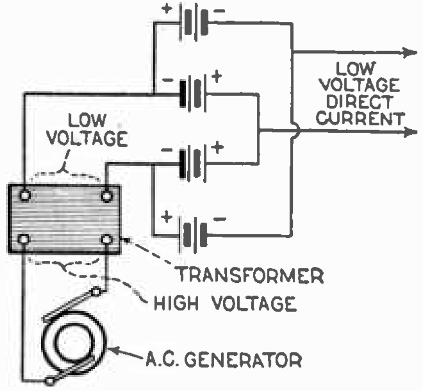

Fig. 2. A hook-up for connecting four Noden cells to obtain full wave rectification.

NON-CONDUCTOR—See *Insulator.*

NON-INDUCTIVE COIL—A coil having negligible inductance. Such a coil may be wound by doubling the wire upon itself and then winding the two parallel halves side by side.

NON-OSCILLATORY—Free from oscillations or vibrations. Without periodicity, that is to say, *aperiodic* (q.v.). A current which is uniform in direction and flow and free from oscillations is said to be a non-oscillatory current. A non-oscillatory circuit is one in which an impressed potential will produce current that gradually diminishes in amplitude without reversing its direction of flow. This is also called an *aperiodic circuit* (q.v.). Such a circuit can have no natural period of oscillation.

NON-PERIODIC—See *Aperiodic,* also *non-oscillatory.*

NON-RADIATIVE ANTENNA—A combination of capacity and inductance equivalent to that of an aerial, used to test transmitting apparatus without radiating waves. (See *Mute Antenna.*)

NORTH MAGNETIC POLE—A point on the earth at a latitude 70 degrees North, longitude 97 West. The north magnetic pole does not coincide with the north (geographic) pole. (See *Magnetic Poles.*)

NORTH POLE—See *Magnetic Poles,* also *North Magnetic Pole.*

NO-VOLT RELEASE—An electro-magnetically controlled device, inserted in the field circuit of a motor so that it holds the handle of the motor starter in place on the last stud, but if the supply current fails, the magnet becomes de-magnetized and allows the spring attached to the starter handle to pull the handle back to the "off" position. This prevents damage to the windings of the motor, in case the current comes on again, since it is then necessary to start up the motor in the usual way, using the starter and gradually increasing the voltage by cutting out the limiting resistances provided in the starter. (See *Overload Release.*)

NULL METHOD—An electrical method of testing in which adjustments are made so that zero deflection is obtained in a galvanometer, as for example when using a *Wheatstone Bridge* (q.v.). This is sometimes referred to as the *zero method.* (Null is also a German word meaning zero.)

O

OCTAHEDRITE—Chemical Symbol TiO_2—Titanium dioxide, an eight sided crystal which has been used as a crystal detector for detecting and rectifying radio currents. (See *Crystal.*)

OERSTED—The unit of *reluctance* (q.v.). The reluctance or in other words, the magnetic resistance offered by a cubic centimeter of vacuum. Unit *magneto-motive-force* (q.v.) will generate a unit of *magnetic flux* (q.v.) through unit reluctance. The oersted was named after the Danish physicist, H. C. Oersted.

OHM—Symbol Ω (Omega)—The unit of electrical resistance. It is the ratio of unit *electromotive force* (q.v.) to unit current. The ohm is the basis of the entire electrical system of units and the volt and the ampere are defined in relation to it. Very elaborate apparatus is required in order to determine the ohm absolutely. However, resistances can be compared easily and once the standard is determined it is a simple matter to make duplicates. In accordance with an act of Congress, the ohm was adopted as a legal unit and defined as follows: The unit of resistance shall be what is known as the international ohm, which is substantially equal to one thousand million units of resistance of the centimeter-second-gram system of electromagnetic units and is represented by the resistance offered to an unvarying electric current by a column of mercury at the temperature of melting ice, 14.4521 grams in mass, of a constant cross-sectional area and of the length of 106.3 centimeters. Stated in English units, the ohm is specified as the resistance offered to an unvarying electric current by a column of mercury at the temperature of melting ice (32 degrees Fahrenheit), 0.5050 of an ounce in mass (approximately 3/100ths of a pound), of a constant cross-sectional area and of a length of 3.487 feet.

OHMMETER—An instrument used to measure resistance directly in ohms. It is particularly adapted for measuring high resistances, although certain types of ohmmeters are made for measuring lower resistances. In one form of ohmmeter, the moving system is deflected by forces due to currents in two coils at right angles to one another. These carry currents in one case proportional to the current through the conductor under test, and in the other case proportional to the potential drop across it. This is in effect, a special form of *galvanometer* (q.v.) and when combined with a hand driven generator gives the conventional *megger* (q.v.) The bridge type ohmmeter depends on the *Wheatstone Bridge* (q.v.) principle. In this instrument, however, the rheostat resistance remains constant and the bridge arms are formed by a wire, called a *slide wire*, resistances being varied by means of a moving contact as shown in the illustration. This type of instrument is often referred to as a *slide wire ohmmeter*. While a straight slide wire is used in some types, in others the wire is wound on an insulating cylinder. In the Weston direct-reading ohmmeter, there is a permanent magnet and a moving coil having two windings and an adjustable magnetic shunt. Provisions are made for connecting an unknown resistance and an auxiliary battery in the circuit of the instrument. A checking plug is provided for checking the accuracy of the instrument before use. When the key is closed, the battery current is divided between two windings. The magnetic force exerted on the one winding, which is in series with a fixed resistance, tends to deflect the pointer up the scale, while the force exerted on the other winding tends to return it to zero. With the correct battery voltage the magnetic shunt may be adjusted to bring the pointer to full scale position. When the plug is shifted to the low range position, a resistance equal to that of the low range is removed from the circuit whose magnetic force tends to deflect the pointer towards the zero position. By connecting a resistance equal to that of the low range across the "unknown" binding posts, the previous condition is restored and the pointer will go back to the full scale position. For values of resistance lower than this the opposing force will be greater and hence the coil will be brought back to a lower position on the scale. At zero resistance across the "unknown" binding posts, the magnetic forces of the two windings will balance each other, thus causing the coil to stand at zero on the scale. The Vawter indicating ohmmeter is of the galvanometer type and utilizes two coils carried by a shaft and moving in the field of a permanent magnet. A specially shaped core is used to give the desired scale characteristics. By making connections to the coils through spirals of negligible torque, the position of the pointer does not depend upon the value of the current nor on the strength of the permanent magnet.

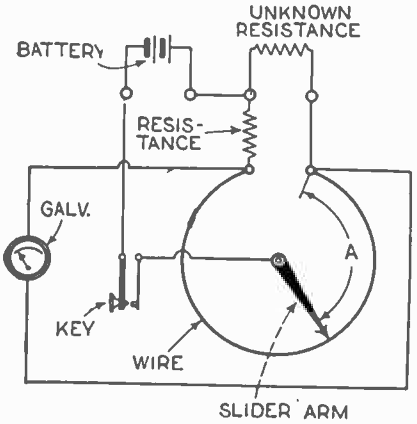

Diagram of connections of a slide wire ohmmeter.

OHMIC COUPLING—See *Resistance Coupling*.

OHMIC DROP—The fall in potential which occurs when an electric current flows through a resistance. (See *IR Drop*.)

OHMIC LOSS—Power or energy loss due to the resistance which an electrical circuit offers to the flow of current. (See *I²R Loss*.)

OHM'S LAW—Current flowing in a conductor will increase directly with increase of voltage and will decrease directly with increase of resistance. In other words, voltage is the cause, while current is the effect. The effect is directly dependent upon the cause and inversely dependent upon the opposition offered to the cause. Ohm's law is the fundamental law of electrical engineering. Ohm's law can be stated as an equation as follows:

$$\text{Current (in amperes)} = \frac{\text{Voltage (in volts)}}{\text{Resistance (in ohms)}}$$

or

$$\text{Voltage} = \text{Current} \times \text{Resistance}$$

$$\text{Resistance} = \frac{\text{Voltage}}{\text{Current}}$$

The above three equations all mean the same thing and serve to express in various forms, the idea set forth above. Ohm's law can be applied to every conductor, provided that the conductor stays at a constant temperature, that there is no internal electromotive force in the conductor and that the distribution of the stream lines in the conductor remains unchanged. Ohm's law is easy to understand and easy to apply. Thus in a particular circuit, a certain voltage causes a certain current to flow. Double the voltage will cause double the current to flow, if the resistance is the same. If the resistance is doubled, however, the same voltage will result in only half the current flow. Double voltage and double resistance results in unchanged current flow. Ohm's law applies to a portion of an electrical circuit as well as to the entire circuit. Care must be taken when applying the law to only part of the circuit, to consider only the resistance, current and voltage of that part.

OMNIBUS BAR—A conducting bar of copper, mounted in back of a switchboard, or panel and serving as a common connector for two or more pieces of apparatus. This term is more usually shortened to *bus bar* (q.v.)

OMNIGRAPH—An intrument which sends code (dots and dashes) mechanically. It is usually operated by clockwork and is connected with a buzzer. It employs metal disc records each of which has a series of notches which correspond to the dots and dashes of the telegraph code. The omnigraph is used by beginners in learning the code and is also used to test the speed of reception of applicants for radio operators' licenses.

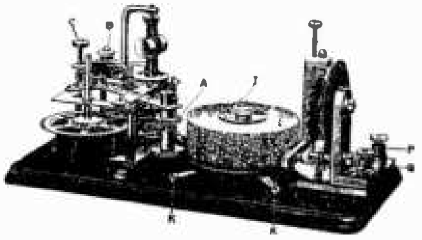

The omnigraph—an instrument which sends code mechanically.

ONDOGRAPH—An instrument for autographically recording the wave forms of varying currents, especially currents which are alternating rapidly. This form of curve tracer was invented by Hospitalier.

ONDOMETER—Another name for a *wave meter* (q.v.)

OPEN ANTENNA—According to the report of the Institute of Radio Engineers standardization committee, this is a condenser antenna, that is to say an antenna consisting of two capacity areas. Antennas may be subdivided into two general classes, those which act principally as condensers, usually simply called antennas, and those which act principally as inductances. These latter are *loop antennas* (q.v.), *coil antennas* (q.v.), etc. When electric waves reach a condenser antenna, they set up an alternating electromotive force between the wires forming the upper plate of the condenser and the ground or the other lower plate of the condenser. As a result of this alternating electromotive force, alternating current will flow in the antenna system. (See *Aerial*.)

OPEN CIRCUIT—An electrical circuit which does not offer a complete path for the flow of electricity. There may be one or more breaks in the path. An open circuit is often referred to as a *broken circuit*. Open circuits often cause trouble in radio apparatus. They may be due to the burning out of a winding, to a connection coming off a terminal, to a soldered connection breaking loose or to similar causes. The method of testing for an open circuit is to connect a battery and a headset in series with the circuit in question. If there is no click in the headset, there is evidently a break in the circuit.

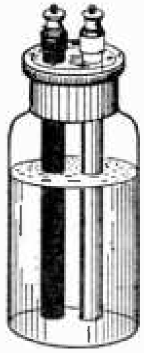

An open circuit primary cell.

OPEN CIRCUIT PRIMARY CELL—A *primary cell* (q.v.) designed for intermittent use and normally kept on open circuit. Such cells usually have a *depolarizer* (q.v.) which acts slowly. In certain cases, no depolarizer whatsoever is employed. Open circuit cells are designed for use during short intervals of time only and must stand for long periods on open circuit, during which the hydrogen is gradually taken away from the negative plate. The operation of the open circuit cell will not be satisfactory if the rest periods are not long enough or are not at frequent intervals.

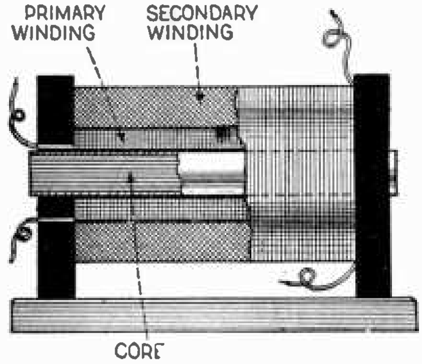

An open-core transformer.

OPEN-CORE TRANSFORMER—A transformer in which the (useful) magnetic circuit is partly through iron and partly through air, instead of entirely through iron as in the *closed-core transformer* (q.v.). It is sometimes referred to as a *polar* transformer. In one type of open-core transformer the arrangement is similar to that of an *induction coil* (q.v.), a hollow iron core, filled with soft iron wires being employed to concentrate the magnetic field.

OPPOSING ELECTROMOTIVE-FORCE — See *Back Electromotive-force.*

OPTOPHONE—An instrument utilizing a telephone receiver controlled by the variations of resistance of a *selenium cell* (q.v.) by means of which the blind are enabled to read ordinary type, through the sense of hearing.

ORDINATE—A mathematical term used in radio in making *logs* (q.v.) or other curves. Defined in mathematical language, the ordinate is that one of the coordinates of a point which is drawn parallel to a line (called the axis of ordinates) to the point from the other axis (called the axis of abscissas) or from the plane of the other axis of coordinates, assumed as a basis of reference. This definition will be clearer if reference is made to the accompanying illustration. The dotted line drawn from the point A to the axis of abscissas is the ordinate of the point A.

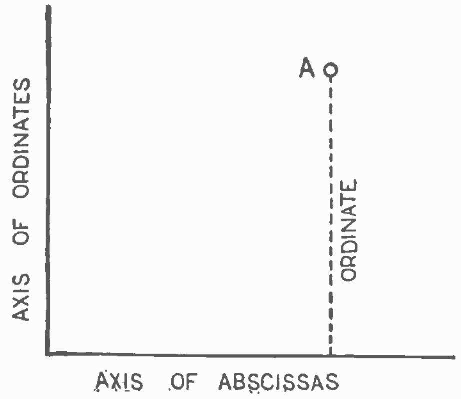

The dotted line is an ordinate.

OSCILLATING CURRENT — An alternating current of high frequency, consisting of a succession of waves of constant length. In some cases, the term oscillating current refers particularly to a current where the amplitude is decreasing in constant proportion due to damping. (See *Oscillation, Oscillation Frequency,* also *Logarithmic Decrement.*)

OSCILLATING IMPULSE — *An oscillating current* set up for the purpose of producing electric waves.

OSCILLATING PERIOD—The time of one complete *oscillation* (q.v.).

OSCILLATION—Periodical surging first in one direction and then in the other. Electrical oscillations are the surgings backward and forward of current periodically, as, for example, in the case of a condenser being discharged in an oscillating circuit. The oscillating action in a discharging condenser is as follows: In the case of a condenser, positively charged, upon being discharged through an inductance, the charge rushes out of the condenser, the current increasing in strength until it reaches a maximum. Then, due to self-induction the current prolongs itself, and it continues until its energy is changed back into an electric charge of an opposite sign, current flowing into the condenser which is then charged negatively. The condenser again discharges giving a maximum current in the opposite direction. the current thereafter flowing back into the condenser with the same direction charge as at first. This action constitutes one complete swing or oscillation and the current continues to oscillate, each time becoming weaker and weaker as the energy is dissipated into heat and also partially into electric waves. It is important to note that the propagation of electric waves could not be accomplished without electrical oscillations. Thus at the transmitting station, the electrical oscillations are converted into electric waves and at the receiving station, the waves are re-converted back into oscillations.

OSCILLATION CONSTANT — The square root of the product of the capacity and the inductance of a circuit. The *natural period* of a circuit is determined by its inductance and capacity in the same manner as the natural period of a steel spring clamped at one end in a vise is determined by its mass and elasticity. Where a simple aerial is used it is an easy matter to calculate the length of a wave radiated from it, provided the oscillation constant has been found. The velocity of electric waves is approximately 3 x 10^{10}cm. per second, and hence the length of the wave radiated is about 60 times the oscillation constant where the wave length is measured in meters. (See *Aerial Tuning Condenser,* also *Isochronous.*)

OSCILLATION FREQUENCY—The number of complete oscillations or cycles flowing in a circuit per second. The oscillation frequency of a circuit varies inversely as the square root of the product of the inductance and capacity. The frequency is inversely proportional to the wave length, hence the addition of either *capacity* (q.v.) or *inductance* (q.v.) to the oscillatory circuit will result in an increase of *wave length* (q.v.). (See *Frequency.*)

OSCILLATION POINT — That current value, known as the *critical current* (q.v.), at which the vacuum tube or tubes in a radio receiving set start to oscillate.

OSCILLATION TRANSFORMER—A high-frequency air-core transformer used for transferring electrical oscillations from one circuit to another. Either one or both the windings may be tuned to the frequency of the oscillations. The transformer used for coupling the oscillator of a transmitting set to the aerial is an example of an oscillation transformer. (See *Coupling, Inductive Coupler,* also *Transformer.*)

OSCILLATION VALVE—See *Oscillator,* also *Vacuum Tube.*

OSCILLATOR—A device or apparatus for producing electrical *oscillations.* The *vacuum tube* (q.v.) under correct conditions may be used to produce sustained oscillations of a definite amplitude and frequency. In order for a vacuum tube to act as an oscillation generator it must be capable of amplifying, part of its output energy must be returned to the input, an oscillation circuit must be combined with the tube, which possesses inductance, capacity and resistances of values such that the tube will oscillate with the desired frequency and finally the tube must have certain characteristics which when combined with the constants of the oscillatory circuit, will determine the amplitude of the oscillations. Oscillations may also be produced by means of a high frequency alternator and by a direct current generator and electric arc. (See *Alternator,* also *Alexanderson.*)

OSCILLATOR, HERTZIAN — Two insulated rods, with axes in line, with their nearer ends forming a spark gap and their outer ends bearing plates or balls to give the system the required capacity. The overall length is usually less than 6½ ft. and the whole constitutes an open oscillator. A Hertzian oscillator is sometimes called a *dipole.* When the rods are brought near each other but not in actual contact and the conductors are connected to the terminals of an *induction coil* (q.v.) a succession of sparks can be caused to jump the gap and these set up electric waves in the *ether* (q.v.).

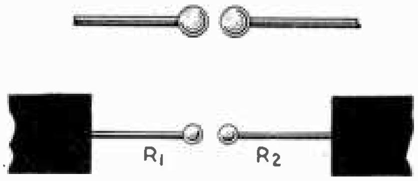

Typical Hertzian Oscillators.

OSCILLATOR, LINEAR—The straight wires connected on either side of the spark gap of an induction coil, traversed by oscillations. Linear oscillators are of the open type. As used by Hertz, they consisted of two rods, having a spark gap between their adjusted ends, while their outer ends terminated in plates or balls in order to increase the capacity of the system.

OSCILLATOR, RIGHI—In this type of oscillator, the spark gap is established between two metal spheres placed between two smaller spheres. (See *Oscillator,* also *Oscillator, Hertzian.*)

OSCILLATOR, THERMIONIC—A vacuum tube used to generate continuous oscillations. (See *Oscillator.*)

OSCILLATORY CIRCUIT — A circuit possessing *inductance* (q.v.) and *capacity* (q.v.), the electrical constants being such that an oscillatory current can be set up. The accompanying figure shows a simple oscillatory circuit, consisting of a condenser, an inductance and a spark gap. When a source of high potential is connected across the spark gap and the latter is properly adjusted, the condenser is charged and there is a spark across the gap. Provided that there is not too much resistance in this circuit, there will be set up an *oscillating current* (q.v.) of high frequency, decreasing in amplitude according to a fixed ratio. Such a current can be used to set up electric waves. Oscillatory circuits are of two general classes, closed as shown in the illustration and open, as in the case of an aerial system. An aerial which has no tuning condenser or inductance in its circuit still has distributed inductance and capacity, whereas an aerial having a tuning inductance and a condenser in its circuit, has a combination of concentrated and distributed inductance and capacity.

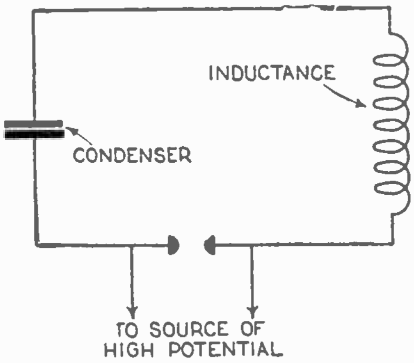

A simple oscillatory circuit.

OSCILLATORY CURRENT—See *Oscillating current.*

OSCILLION—Name which has been applied to the *vacuum tube* (q.v.) especially where used as an oscillation generator.

OSCILLOGRAM — A picture, graph or record of a wave form usually obtained by means of an *oscillograph* (q.v.).

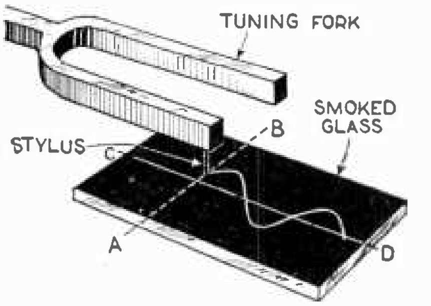

An oscillogram obtained by means of a tuning fork.

The illustration shows an oscillogram obtained on smoked glass by means of a tuning fork. (See *Oscillagraph.)*

OSCILLOGRAPH — An instrument for recording photographically or showing visually the wave form of alternating currents or of other rapidly changing currents or voltages. The essential parts of such an instrument are a moving coil mirror galvanometer, a rotating or vibrating mirror and a moving photographic plate or film. Figure 1 illustrates the principle of the oscillograph and shows the curves it produces. The oscillograph was invented by M. A. Blondel in 1893. By means of the oscillograph the wave form of a current may be shown as a curve and its characteristics may thus be examined and studied. In order to form the curve, it is necessary to have simultaneous motion in two directions. The vibrator of the oscillograph which consists of a delicate coil of wire, with mirror attached, is made extremely light so that the moving system will have as little inertia as possible. The coil is placed in a powerful magnetic field and when an oscillating current is passed through it, it oscillates in synchronism with the current. A spot of light focused on the mirror is made to fall on a screen at the predetermined distance from the vibrator. This spot of light traces out a straight line on the screen. Substituting a photographic film for the screen and arranging the film so that it will move in a vertical path across the moving beam of light (which is assumed to be moving from right to left across the screen) a curve is traced on the film. The commercial form of oscillographs are provided with apparatus whereby the curve may be examined without the necessity of photographing it. In these the beam of light is reflected from a rotating or vibrating mirror in such a way that it receives a horizontal and an up and down motion and is focused onto a curved plate of glass. Tracing paper may be stretched over the glass so that the image of the spot of light traces a visible curve on it. The curve being repeated very rapidly appears to be standing still and hence it is possible to trace over it with a pencil or even to place a piece of photographic paper on the curved glass and make a record of the curve in this way. The motion picture camera has been applied to the oscillograph and is useful where longer records are required. It should be noted that oscillographs which utilize a moving coil, such as the Duddell or Blondin types. will work over a band of frequencies up to a maximum of about 300 cycles per second. Inasmuch as the oscillations common in radio work are of much higher frequency than 300 cycles, it has been found necessary to use an oscillograph whose moving parts are without appreciable weight. The use of electrons as moving parts has solved this problem. A Frenchman, Alexander Dufour has applied the Braun tube to the construction of an oscillograph which has no appreciable inertia and which is capable of operating at a frequency up to a million cycles per second. This device uses only minute amounts of energy in its operation and therefore does not disturb the original circuit. It is able to register both voltage and current simultaneously and is so arranged that a single impulse is sufficient for a photographic impression. The Dufour oscillograph consists essentially of two glass tubes fitted by means of a ground joint into a bronze chamber. The upper glass tube carries a cathode and an anode and the other tube has one pair of deflecting plates for electrostatic deflection of the electron stream. Two sets of coils perpendicular to each other are used for magnetic deflection and these are placed outside the tube and slightly below the deflecting plates. In order to photograph the oscillations, a drum is used which is provided with a film magazine allowing six films to be taken in succession. When visual examinations of the oscillations are to be made, a fluorescent screen is turned up into position covering the opening into the interior of the drum thus preventing the films from being exposed when using the screen. After the films are put into the drum, this is placed within the bronze chamber and locked into position. An air-tight door is used to close the opening and the film changing mechanism is operated by means of external controls. Glass windows on either side of the bronze chamber allow the fluorescent screen to be seen. The accompanying sketch shown at Figure 2, gives a diagrammatic representation of the operation of the Dufour oscillograph.

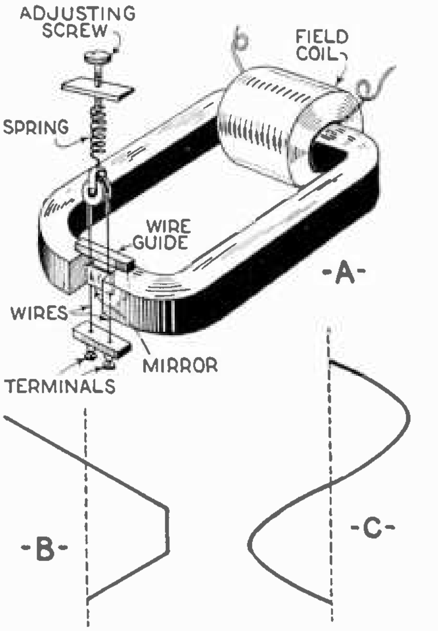

Fig. 1. The principle of operation of the oscillograph is shown at A. At B are shown diagonal and straight paths traced by the beam of light falling on a moving photographic plate. The type of curve which an oscillograph traces for alternating currents is shown at C.

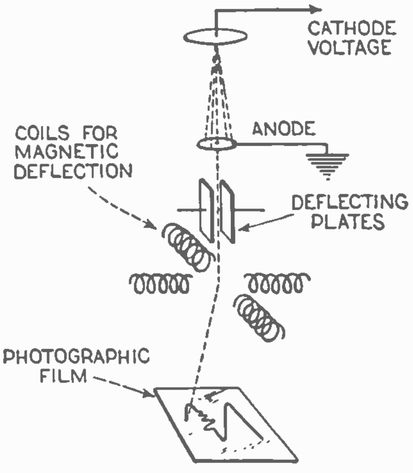

Fig. 2. A diagrammatic representation of the Dufour oscillograph.

OSMOSIS, ELECTRIC—When an electric current is passed through an electrolyte having an anode on one side of a porous diaphragm and a cathode on the other side, there is a tendency for the liquid to pass through the diaphragm towards the cathode raising the level of the liquid on the cathode side. This phenomenon is known as electric osmosis.

OTOPHONE—A device for the hard of hearing, which utilizes an extremely sensitive microphone connected to a two-stage vacuum tube amplifier. The instrument uses dry cell tubes and the batteries are contained in the case with the microphone. A small telephone receiver completes the outfit.

OUDIN RESONATOR — A device used for obtaining high-frequency brush discharges. Discharges of this nature are used in the investigation of resonance effects and also in medical work. The accompanying diagram explains the principle of the Oudin resonator.

B D is an uninsulated coil of copper wire wound on an insulating core with each turn separated from the next adjacent one. At B, the helix is connected to an induction coil and it is tapped by means of a sliding contact at C. The sliding contact is connected through a condenser to the induction coil H, with a spark gap G in parallel in the circuit as shown. The circuit BGFC is a closed oscillatory circuit receiving its excitation from the induction coil H. The circuit C D E is an open circuit in contact with the oscillatory circuit. By moving the sliding contact, C, a point is reached where a brush discharge appears at E, thus indicating that electric oscillations are being set up in the open circuit. (See *Brush Discharge*, *Corona*, also *Oscillator.*)

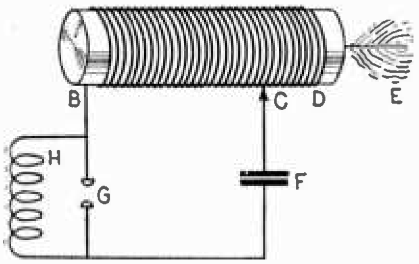

Connections of the Oudin resonator.

OUTPUT—The useful energy given out by a machine, circuit, vacuum tube or other device. (See *Input; Output, Transformer; Output Circuit of Vacuum Tube.*)

OUTPUT CIRCUIT OF VACUUM TUBE—The filament-plate circuit as differentiated from the filament-grid or input circuit. (See *Input Circuit of Vacuum Tube.*)

OUTPUT, TRANSFORMER—The product of the voltage at the secondary terminals of a transformer by the current flowing in the secondary winding. The output of any transformer is always less than the input by an amount equal to the losses. The *efficiency* of a transformer is the ratio of its useful output to the total *input.* (q.v.)

OVERLOAD RELEASE—An electromagnetic device which is used to cause a motor starter handle to return to the off position when the current exceeds a pre-determined overload value.

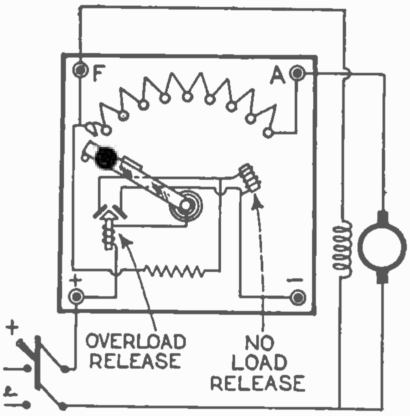

Overload Release.

The overload magnet is usually arranged so that too strong a current will result in short-circuiting the no-volt release thus switching off the driving current. One form of overload release consists of a magnet in series with the line and the armature. This magnet is arranged to attract a pivoted iron keeper whenever excessive current flows. The pivoted arm, when attracted by the magnet, makes connection between two contacts, themselves connected so as to short circuit the holding magnet when the circuit is closed. Therefore, when more than the allowable current is drawn from the line, the overload magnet attracts its pivoted keeper which closes the circuit short-circuiting the holding magnet. The contact arm is then released by the holding magnet and the arm is pulled back to the off position, thus cutting the motor off the line. Another form of overload release utilizes an electromagnet to attract a pivoted arm, arranged so that this will break the circuit between the motor and the line. This results in operating the no-voltage release. (See *No-Volt Release.*)

OVERTONE CURRENTS—Alternating currents which have harmonic frequencies of a higher value than the fundamental or first harmonics.

OVERTONES—Higher harmonics associated with the first harmonic. This term particularly refers to sound waves. In various musical instruments, different overtones are brought out and this results in each instrument having its own tone characteristic.

P

PACKING—(referring to a *microphone*). The tendency of carbon granules in a microphone to settle or pack. In this position they are no longer sensitive and do not readily respond to the vibrations of the diaphragm. Packing may be caused by irregularity in the sizes of the carbon granules. It may be remedied by gently tapping the microphone.

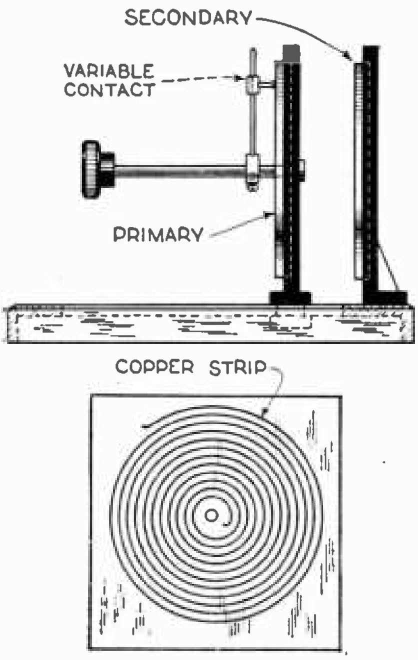

Front and side view of a pancake coil.

PANCAKE COIL—A flat inductance coil. Coils of this type are used in radio receiving sets as radio frequency transformers, also as antenna couplers and oscillator couplers. They may be *basket wound* (q.v.), spider wound, etc. In wireless transmitting apparatus, aerial tuning inductances, loading inductances, and oscillation transformers (illustrated) are frequently of the pancake coil type. (See *Inductance, Antenna.*)

PANEL—A board of insulating material carrying the controlling or measuring devices of an electric circuit. The panel for controlling the charging of storage batteries is called a *charging panel.* A panel carrying generator controls is referred to as a *generator panel.* Where only a single panel is used this is usually called the *switchboard.* The usual switchboard comprises a number of panels. Switchboard panels are commonly made of marble. The usual varieties available are white Italian, pink or grey Tennessee and blue Vermont marbles. Plain slate is sometimes used for panels where the voltage is not too high. When the panels do not require a finish, soapstone is sometimes used. The front of a radio receiving set on which the dials and rheostat controls, etc., are mounted is also called a panel. These panels are usually made of Bakelite, hard rubber, composition or similar insulating substances, although a number of sets are being made with metal panels. The horizontal panel on which the sockets and transformers are usually mounted is known as the *sub-panel* (q.v.).

PANNILL, Charles Jackson—American radio pioneer. Born in Petersburg, Virginia, May 13, 1879. He entered the American Navy, 1898. In 1902 he took a post under Professor R. A. Fessenden and carried out a series of wireless experiments at Hampton Roads, later inaugurating wireless communication between New York and Philadelphia. Pannill was the first man to install wireless on the battleships of the U. S. Navy. Afterwards he carried out a series of communication experiments between various parts of the United States and erected a number of wireless stations. He joined the Marconi Wireless Telegraph Company in 1912 and became a radio adviser to the United States Government, 1914, and assistant director of Naval Communications, 1916. Pannill is a fellow of the Institute of Radio Engineers and a member of the Washington Society of Engineers.

PAPER CONDENSER—A fixed condenser, usually made of tin foil and utilizing a dielectric of paraffin-waxed paper. The 1 microfarad by-pass condenser is a typical example of a paper condenser. (See *Condenser, By-pass.*)

PARALLEL—See *Parallel Connection.*

PARALLEL CONNECTION—Two or more parts of an electrical circuit so connected that the current divides between them. This is also known as a *multiple* (q.v.) connection, *divided circuit*, or *shunt* circuit. Where a number of parts of a circuit are connected in parallel, the same potential is impressed on each part. The current flowing in each branch will depend upon the impressed voltage and the resistance of each branch. (See *Circuit, Parallel.*)

PARALLEL RESISTANCES—Resistances connected so that their terminals have the same difference of potential between them. The greater the number of resistances in parallel, the less will be the total resistance of the circuit. If there are three resistances in parallel of ½, ⅓ and ⅕ ohm, respectively, then the total resistance is equal to the reciprocal (one divided by a number) of the sum of the separate reciprocals. In this case the sum of the reciprocals is 2 plus 3 plus 5 equal to 10 and the reciprocal of this is equal to one-tenth, so that the total resistance of this parallel circuit is one-tenth of an ohm. It can be seen that the total resistance of any circuit having a number of resistances in parallel is less than the resistance of any *one* of the branches. (See *Circuits Parallel*, also *Parallel Connection.*)

PARALLEL RESONANCE—When a concentrated capacity and a concentrated inductance are connected in parallel between terminals to which an alternating electromotive force is applied, and the inductance or capacity or frequency is varied, the condition of parallel resonance exists when the current supplied by the source is a minimum. It should be noted that in series resonance, the total current supplied by the source flows through both the inductance and the capacity whereas in parallel resonance the current supplied by the source is the vectorial sum of the two currents, one flowing through the capacity and one through the inductance.

PARAMAGNETIC—Having a permeability greater than unity. Magnetic as opposed to diamagnetic. The term *paramagnetic* is used in some cases to apply to substances which have only a slightly greater permeability than air such as liquid oxygen, the rare metal erbium, etc. If used in this sense it does not include *ferro-magnetic substances* (q.v.). (See *Diamagnetic material.*)

PARTIAL—An acoustical term denoting any one of the natural vibrations of which a body is capable. An electrical *oscillator* (q.v.) possessing distributed capacity and inductance has, theoretically, an infinite number of possible frequencies, each of which is called a *partial.* The lowest frequency is called *the fundamental.* (See *Harmonics, Harmonic Current, Overtones,* also *Overtone Currents.*)

PARTITION INSULATOR—An ebonite tube having a metallic rod running through its center with wing nuts at each end. It is used for continuing a circuit through a wall or partition.

P.D.—Abbreviation for *potential difference.* (See *Potential, Difference of.*)

PELTIER EFFECT—The change in temperature, either heating or cooling, at the junction of two unlike metals, depending upon whether an electric current originating from an external electromotive force, is sent through the junction in one direction or the other. The Peltier effect is the opposite of the *thermo-electric* effect.

PERCENTAGE COUPLING—The coefficient of coupling between the primary and the secondary of an oscillation transformer in a wireless transmitting system. When the primary winding is placed close to the secondary, the coupling is said to be *close* or *tight.* When the two windings are apart the coupling is said to be *loose.* With the ordinary spark gap in the closed circuit, if the primary and the secondary windings are closely coupled a broad wave will be radiated from the aerial system. If they are loosely coupled, a *sharp* wave will be radiated.

PERIKON DETECTOR—A crystal rectifier utilizing a piece of zincite in firm contact with a piece of chalcopyrite. (See *Combination Detector*, also *Crystal Detector.*)

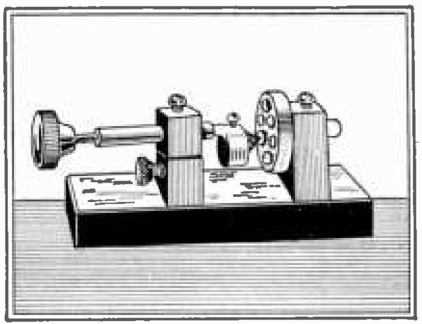

A Perikon detector.

PERIOD—The time required for one complete *cycle.* (q.v.) Referring to an *alternating current* (q.v.) it is the time required for the current or electromotive force to pass through the various values from zero to a positive maximum back to zero again, then to a negative maximum and finally to zero.

PERIODIC—A vibration is said to be periodic if all the phenomena are repeated regularly at equal intervals of time. This time is called *periodic time* or the *period.*

PERIODICITY—A synonym for frequency, or for angular velocity.

PERMANENT MAGNET—Hardened steel which when magnetized retains its magnetism after the removal of the magnetizing force. The usual method of making permanent magnets is to place the steel in a powerful electromagnetic field. They can also be made by stroking with another magnet. Permanent magnets are used in all electromagnetically operated headsets and *loud speakers.* (q.v.) (See *Magnet, Electromagnet,* also *Temporary Magnets.*)

PERMEABILITY—The *permeance* (q.v.) existing between the opposite faces of a cube of the substance each side of which is one centimeter in length. Since the permeability of air is assumed to be unity, the permeability of any substance is the ratio of the flux that passes through it to the flux that would exist in air if the magnetomotive and flux path remained unchanged.

PERMEANCE—That property of a magnetic circuit which allows the flow of magnetic flux. Its reciprocal is *reluctance* (q.v.) which is the property of a magnetic circuit by which it resists the flow of magnetic flux.

PERMITTANCE—A synonym for *capacity* (q.v.) or *capacitance* (q.v.). This property is possessed by every electrical circuit, but does not manifest itself unless there is a change in voltage. The greater the permittance of a circuit, the greater will be the opposition offered to a change in voltage.

PERMITTIVITY—See *Dielectric Coefficient and Constant, Inductive Capacity,* also *Inductivity.*

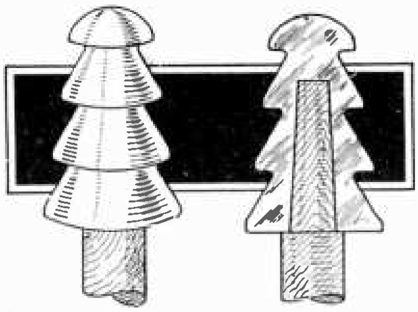

A petticoat insulator. Shown in cross-section at the right.

PETTICOAT INSULATOR—An insulator having two or more flanges or petticoats arranged one on top of the other for the purpose of reducing leakage and preventing accumulation of moisture. Petticoat insulators are usually made of porcelain. Insulators used on low tension lines are also of the petticoat type, in such cases being constructed of glass. (See *High Tension Insulator,* also *Insulator.*)

PHANTOM AERIAL—See *Mute antenna.*

PHASE—A particular state in a regularly recurring cycle of changes. In simple harmonic motion, uniform circular motion, or in periodic changes of any magnitude varying according to a simple harmonic law (as for example an alternating current) the point or stage in the period to which the rotation, oscillation or variation has advanced, considered in its relation to the position or instant of starting or to some other standard position. This relationship is usually expressed in angular measure. Defined in electrical terms, phase is the distance, usually measured as an angle, of the base of any *ordinate* (q.v.) of an alternating wave from any chosen point on the time axis. This distance represents the phase of the ordinate with respect to the point. In any sinusoidal alternating current quantity, the phase at any instant may be shown by the corresponding position of a *vector* (q.v.) revolving about a point with such an angular velocity that its projection at each instant upon a chosen reference line is proportional to the value of the quantity at that instant. As an example, consider a conductor rotating in a magnetic field. The induced electromotive force, in one complete revolution, passes through the following values. It starts at zero and rises to a maximum positive value, then falls to zero and rises to a maximum positive value in the reverse direction, finally falling to zero again. The transitions through these various values constitute a *cycle* (q.v.). The time of this transition is called a *period,* the number of cycles per second being the *frequency* (q.v.). The angle at any instant through which the conductor has rotated since the electromotive force changed its sign from negative to positive is the *phase angle.* This is ordinarily designated by the Greek letter Φ (phi). If ω (omega) is the angular velocity of the rotating conductor and f the frequency, then $\omega = 2\pi f$ (since there are 2π radians in a complete revolution). If the time in seconds which has elapsed since the conductor last passed the point of zero electromotive force be represented by t, then the phase Φ at that particular instant is equal to $\omega t = 2\pi f t$. Periodic variations are conveniently expressed by sine curves and the relative phases are indicated by the relative positions of the *nodes* (q.v.) and *loops* (q.v.).

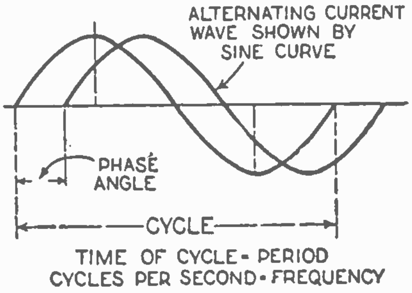
ALTERNATING CURRENT WAVE SHOWN BY SINE CURVE
PHASE ANGLE
CYCLE
TIME OF CYCLE = PERIOD
CYCLES PER SECOND = FREQUENCY

Phase displacement. cycle, etc., illustrated by means of sine curves.

Magnitudes having maximum and minimum values occurring simultaneously are said to be *in phase.* If they do not occur simultaneously, they are *out of phase.* When the corresponding maximum or minimum values of two sinusoidal alternating quantities of the same frequency occur at different instances, the two quantities are said to differ in phase by the angle between their nearest negative maxima or their nearest positive maxima or any other nearest corresponding values. The quantity whose maximum value occurs first is said to *lead* (q.v.) the other. The quantity whose maximum value occurs later is said to *lag* (q.v.). Where one quantity lags behind another by 180 degrees, the values are said to be of *opposite phase.* An alternating current may be *single phase* or *polyphase.* Another term for polyphase is *multiphase.* In a three-phase alternating current system the three separate currents differ in phase from each other by 120 degrees. In a polyphase system, if the various phases each carry or supply equal current, the system is said to be *balanced.*

PHASE ANGLE—See *Phase.*

PHASE DISPLACEMENT—The difference in phase or the phase angle between two alternating quantities, such as voltage and current, of the same frequency. Such quantities do not pass through their respective maximum positive quantities or through their zero values at the same time. The instantaneous values of any sine-shaped wave can be represented as being proportional to the sine of an angle, a complete cycle being equal to 360 degrees. When two quantities are out of phase, the respective angles corresponding to each quantity at any given instant are unequal, the difference between the two angles being called the *phase displacement* or the *phase difference.* In the accompanying illustration, the current and voltage relationship in an alternating current circuit are shown by the two sine curves. As can be seen, the current and voltage are out of phase, the phase displacement being equal to the angle Φ, in this case 90 degrees ($\frac{\pi}{2}$ radians). The current reaches its maximum value at a later time than the voltage. In this case the current *lags* behind the voltage by the angle Φ, the voltage leading the current by this angle. Since the angle of lag of the current is 90 degrees, the current and voltage are said to be in *quadrature.* In an alternating current circuit having pure inductance and no resistance (a theoretical circuit, since all electrical conductors possess resistance), the current would lag by 90 degrees as in the example given. (See *Lag, Lead, Leading Current,* also *Phase.*)

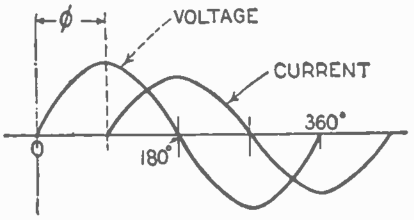

Sine curves of current and voltage, showing difference in phase.

PHENOMENA OF ELECTRIC WAVE PROPAGATION—Electric waves are created by electric *oscillations* (q.v.). They are often referred to as *radio waves.* Electric waves, light waves and radiated heat waves are included in the general term *electromagnetic waves* (q.v.). The velocity of electric waves is the same as that of light waves, 300,000,000 meters per second or 186,000 miles per second (appx.). Electric waves used for radio work have frequencies from about 10,000 to 3,000,000 cycles per second. Since the velocity of electric waves is known, it is possible to calculate the length of a wave if the frequency is known. Thus if the electric wave has a frequency of 10,000 cycles per second, the wavelength is found by dividing the velocity in meters per second by the frequency in cycles per second, which gives in this case a wavelength of 30,000 meters. Electric waves are in reality a combination of electromagnetic and electrostatic disturbances in space. The displacements are at right angles to the motion of the wave train. The electric field and the magnetic field are at right angles and travel together. They spread out or expand from the point of disturbance just as water waves spread out from a point where a stone strikes the surface of the water. Electric waves are propagated from an aerial when certain forms of alternating current flow in the aerial. Since the frequency of the propagated wave corresponds to the frequency of the current in the aerial, high frequency current must flow in the aerial to produce the high frequency electric waves used in radio. Such current may be produced by a high-frequency generator, by a direct current generator and an electric arc, or by means of vacuum tubes. The latter offer the simplest means of producing the rapidly reversing currents. Electric waves may be classified as *continuous* and *discontinuous.* Examples of the former are the waves produced by a high frequency alternator, an oscillating vacuum tube or a Poulsen arc. Examples of the latter are the waves produced by condenser discharges in a spark circuit. In these the amplitude of the waves diminishes in each wave train and the waves are referred to as *damped* waves. The high frequency current used to produce continuous waves used in radio telephony is *modulated* in accordance with the vibrations of a *microphone* (q.v.). As a result modulated waves are sent out from the aerial and these correspond in wave form to the variations of the sound waves. The sensitive apparatus at the distant receiving set picks up or intercepts the modulated electric waves, reconverts them to high frequency electric currents which are rectified into direct currents capable of actuating a 'phone and usually amplified so that they have energy enough to operate a loud speaker. (See *Oscillator, Carrier Wave,* also *Modulated Currents.*)

PHILLIPS, Raymond—British radio expert. Born October 6, 1879, he was educated at Edgebaston, Birmingham Schorne College, Buckinghamshire, and Windsor High School. For three years he was engaged on railway constructions and repairs, with special application to electric railways. In 1902 he invented a system of automatic train control and a number of appliances for electric railways which were widely adopted. In 1905 he took up the study of wireless and specialized in the control of mechanisms at a distance by electric waves. In 1910 he patented a system for controlling airships by radio and by means of a working model demonstrated how an airship could be thus controlled. During the World War he was inspector of ordnance machinery and in 1921 was appointed a member of the Inter-Allied Commission of Control in Germany. He is the author of many articles on the control of machines by electric waves.

PHONOGRAPHIC RECORDING—This refers to the amplifying of code signal currents by means of relays in tandem until the sounds produced in a telephone receiver are loud enough to cut a distinct record on the wax of a phonograph. The signals can be received at high speed and thereafter read at low speed. Instead of a phonograph a telegraphone may be used.

PHOTO-ELECTRIC EFFECT—A change in the electrical conductivity of a gaseous or solid substance when radiations of certain wavelengths come in contact with them. This refers particularly to rays of the visible spectrum and ultra-violet and infra-red rays.

PHOTOGRAPHIC RECORDER—A device for recording high-speed wireless messages by recording the deflections of a sensitive galvanometer on a moving photographic film. (See *Auto-Receiver.*)

PICKARD, Greenleaf Whittier—American radio expert. He was born at Portland, Maine, on February 14, 1877, and was educated at Westbrook Seminary, Lawrence Scientific School, Harvard and Massachusetts Institute of Technology. He made a special study of wireless telegraphy and telephony and has taken out many United States and foreign patents for radio inventions. He is noted for his pioneer work in radio telephony. Pickard began radio work in 1899 at Blue Hill observatory, Milton, Massachusetts, under a grant from the Smithsonian Institute. He was on the engineering staff of the American Telephone and Telegraph Company from 1902 to 1906. Later he was connected with the Wireless Specialty Apparatus Company as consulting engineer. He has an extensive practice as a patent expert in radio patent litigation. Pickard is a fellow of the American Institute of Electrical Engineers, a Member of the American Chemical Society and also is a Member of the Society of Chemical Industry.

Greenleaf Whittier Pickard.

PIG-TAIL—A flexible braided or stranded copper conductor used to carry the current to the rotor of a condenser or a coupler. Pig-tails are also used on the carbon brushes of motors and generators. (See *Flexible Lead.*)

PITCH—Frequency of vibration of a sound. A shrill note is said to be of high pitch, a bass note of low pitch. The pitch of an armature winding refers to the distance from the center of a winding to the center of the next winding.

PLAIN AERIAL—An aerial having the transmitting or receiving circuit connected directly to it without the utilization of inductive coupling or any intermediate tuned circuit. An example of a plain aerial is one used to transmit code, having a spark gap in series with the aerial and the ground.

PLATE—The anode or positive electrode of a vacuum tube. The vacuum tube plate is usually made of nickel. The shape of the plate varies according to the type and construction of the tube. In some tubes the plate is tubular; in others flat. When the cathode or filament is heated, it emits electrons which pass to the grid or through the grid to the plate, depending upon the grid charge. In power tubes (transmitting) it is necessary to use some cooling device to conduct away the heat dissipated at the plate. In some cases a liquid bath or a blast of air is used or the plate may be constructed so that it can be cooled by water circulation. Other methods of preventing overheating of these tubes are increasing the area of the plate as for example increasing the diameter of a tubular plate, or blackening the surface of the plate to increase its heat emissivity.

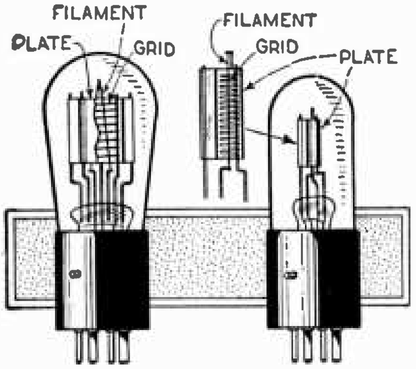

Vacuum tubes, showing construction of plates.

PLATE BATTERY—A battery, usually referred to as the *"B" Battery* having its positive terminal connected to the plate of the three-electrode vacuum tube and its negative terminal leading to the filament. The purpose of this battery is to keep the plate voltage positive with respect to the filament so that the electrons will be attracted sufficiently to the plate. Plate batteries are usually dry cells although there are a number of storage plate or "B" batteries on the market, both of the acid and the alkaline types. It is possible to dispense with the plate battery by using a *"B" eliminator*, which is connected to the ordinary house-lighting circuit. (See *"B" Battery.*)

PLATE CIRCUIT—The circuit connected to the *plate* (q.v.) or anode of a three-electrode vacuum tube.

PLATE COMPONENT, HIGH FREQUENCY—The radio frequency current flowing in the *plate circuit* (q.v.) of a vacuum tube. (See *High Frequency Component of Plate.*)

PLATE CONSUMPTION—This refers to the current consumed in energizing the plate of a vacuum tube. Increasing the grid voltage tends to increase the plate consumption. An increasing plate voltage also increases the plate current, but tends to decrease the relatively small grid current. (See *Consumption, Current;* also *Consumption, Plate.*)

PLATE CONTROL—The vacuum tube starts oscillating at a plate voltage which will depend on the adjustments of the circuit. When the plate voltage is raised, the oscillation current becomes greater, increasing in direct proportion.

PLATE CURRENT—The current in the *plate circuit* (q.v.) of the three-electrode vacuum tube. The current passing between the plate and the heated cathode or filament.

PLATE IMPEDANCE—The total resistance offered to alternating current between the filament and the plate in a vacuum tube. It is often referred to as the *tube impedance.* The impedance is independent of the alternating current input voltage.

PLATE VOLTAGE—The voltage of the plate of a vacuum tube above the filament. The potential difference between the plate and the filament. This term is usually used to refer to the voltage of the *"B" battery* (q.v.) connected to the plate, as for example, if the 22½-volt terminal of the "B" battery is connected to the plate of the detector tube, the plate voltage is said to be 22½ volts. Soft tubes, formerly used as detectors, operated best at this voltage. Most of the amplifier tubes now used as detectors operate best with 45 volts on the plate, although these tubes vary individually, some operating better with slightly increased voltage and others with a voltage less than 45. The plate voltage generally used on the amplifiers is 90 volts. The new power reception tubes use plate voltages up to 135 volts.

PLIODYNATRON—A special vacuum tube, designed by A. W. Hull, having two grids instead of one. This tube utilizes the normal amplifying property of the vacuum tube to obtain a very high voltage amplification, as high as 1,000 fold. When such high amplification is obtained, however, the operation is unstable and extremely careful adjustment is necessary.

PLIOTRON—A name given to a highly evacuated vacuum tube used for transmitting purposes. Large pliotrons are capable of developing as high as 500 watts output since they are so highly exhausted that several thousand volts may be applied to the plate and also since they can carry a plate current of nearly one-half an ampere.

PLUG—A connecting device used extensively in radio and telephone work for making connection or "plugging in" *loud speakers* (q.v.), head sets, power supply, microphone transmitters, etc. The plug has an insulating handle and the metal contact tip is designed to fit into a standard *jack* (q.v.). A flexible wire fits into terminals in the plug, the other end of the wire being connected to the apparatus such as the loud speaker, etc. The plug connection terminals may be of the conventional small binding post type or they may consist of internal spring clips.

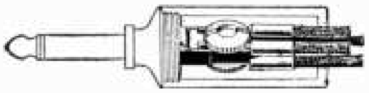

Cross-section of plug.

PLUG-IN COIL—An interchangeable coil, often of the honey-comb type, which can be plugged into a socket, thus permitting it to be quickly and conveniently put into or taken out of a circuit. Plug-in coils are used in some cases in receiving sets designed for operation on widely varying wave bands.

PLUG-IN TRANSFORMER—*A plug-in coil* (q.v.) specifically used as a radio frequency transformer.

POLARITY—In magnets the differentiation between *North Pole* and *South Pole.* In electrical devices or circuits, the difference between *positive* and *negative poles.* It is often necessary to determine which is the positive pole or terminal of a direct current circuit, and which the negative. There are various methods of indicating polarity. A voltmeter or an ammeter will show the direction of current flow. An extremely simple method of determining polarity is to place the two terminals in a glass jar containing salt water or some other electrolyte. A much greater number of bubbles will be observed to collect around the terminal of negative polarity. A piece of used blue print paper can be utilized to indicate polarity. It is simply necessary to moisten the paper and place the terminals on it, at a distance of about ¼ of an inch apart. The paper will be whitened at the negative terminal.

POLARIZATION—In a primary cell, the current liberates hydrogen from the electrolyte in passing through it. The hydrogen is carried to the negative plate with the current and collects there. It thus increases the internal resistance of the cell. This makes it harder for the current to pass through it and as a result the electromotive force available at the terminals is materially lowered. In addition there is a slight difference of potential between the negative plate and the hydrogen and this acts in opposition to the electromotive force of the cell. This further reduces the effective voltage of the cell. The entire effect is known as polarization. Polarization may be carried to such an extent that current can no longer flow from the cell. It may be counteracted or prevented by the use of *depolarizers* (q.v.). There are three different kinds of depolarizers. Chemical depolarizers are placed in the cell near the cathode. Substances such as manganese dioxide, potassium bichromate, nitric acid or cuprous chloride are used. When the hydrogen reaches the depolarizer it combines with it to form a new compound. Electrochemical depolarizers depend for their action upon the use of a substance which will liberate a metal such as copper at the cathode instead of hydrogen. Thus with no hydrogen present there is no polarization. Of course it is necessary to use the proper electrodes and the correct electrolyte to obtain this result. Mechanical depolarizers merely blow or sweep the hydrogen away from the cathode. Being the least effective of the three methods, the latter is seldom used. (See *Local Action.*)

POLARIZATION, ELECTRIC—The state of a dielectric when subjected to electrostatic forces. Electric polarization is synonymous with electric displacement.

POLARIZED RELAY—A relay in which a magnetized swinging arm is placed between poles of two electromagnets. When current passes, one pole must change, so that the arm is attracted by one and repelled by the other. A polarized relay is much more sensitive to weak currents than the ordinary relay.

POLES—See *Magnetic Poles.*

POLYPHASE—Referring to an alternating current system in which the circuits are divided up into two or more branches having their currents displaced in phase from each other. The term *multiphase* has the same meaning as polyphase. (See *Phase.*)

POLYPHASE ALTERNATOR—An alternating current generator wound in such a way that it supplies two or more currents which differ in *phase* (q.v.) from each other by definite phase angles. The usual form of polyphase alternator is the three-phase alternator.

PORCELAIN CLEAT—See *Cleat.*

POSITIVE ELECTRODE—In a primary cell, the zinc plate or *anode* (q.v.) is the *positive electrode,* while the *pole* of this electrode is the *negative pole* (q.v.) because it is negative in relation to the external circuit. In a storage cell, the lead peroxide plate, which is the cathode during discharge, is called the *positive electrode* and its pole the *positive pole.*

POSITIVE TERMINAL POLE—The pole or terminal out of which current is considered to flow. This is more or less of a handy convention, but is contrary to the present day knowledge of electrical flow inasmuch as electrons move from negative to positive.

POTENTIAL—See *Potential, Electric.*

POTENTIAL, DIFFERENCE OF—abbreviation P.D.—Difference of electrical level. Theoretically the potential difference between two points in an electric field is measured by the work done by the electric field on a unit point charge of positive electricity moved from one point to another without disturbing the field. Potential difference is measured in volts. It may refer to any two points in a circuit, even though there is no source of electromotive force between these points. When applied to circuit between two points not possessing a source of electromotive force, the potential difference between the two points under consideration is equal to the product of the current flowing times the resistance of the circuit between the two points. In the case of an open circuit, such as the terminals of a cell, the potential difference is equal to the electromotive force. (See *Potential, Electromotive force,* also *Volt.*)

POTENTIAL DROP—The fall in voltage in a circuit due to the resistance of the conductor. (See *Potential, Electric;* also *Potential, Difference of.*)

POTENTIAL, ELECTRIC—Electrical pressure, or the degree of *electromotive force* (q.v.). Potential is electrical level. Electrical potential can be compared with the "head" of water. Thus it is customary to measure the height or head of water above sea level, which is taken as an arbitrary base. In a somewhat similar manner, the potential of the earth is taken as zero and all potential is measured from this as an arbitrary base. With this understood, it is no longer necessary to refer to the degree of electromotive force, but simply to the electromotive force or the potential. See *Potential, Difference of,* also *Volt.*)

POTENTIAL ENERGY—Energy of position. When energy is available for the production of work, it is referred to as potential energy. Energy at work is *kinetic energy* (q.v.).

POTENTIAL RECTIFIER—A *crystal* (q.v.) or other form of rectifier which requires an initial current to pass through it to become sensitive.

POTENTIOMETER (Potential Divider) —A resistance connected across a battery or other source of electromotive force and equipped with a slider so that the voltage on another circuit may be varied from zero to the full potential of the battery. Where a gas content detector vacuum tube is used, the purpose of the potentiometer is to permit a close adjustment of the *plate battery* (q.v.) potential. The potentiometer is shunted around the filament heating or "A" battery, its resistance being so high that its power consumption is practically negligible. In tuned radio frequency sets, a potentiometer is used to control the tube oscillation. The advantage of being able to make a system oscillate is obvious, as the greatest amplification is obtained when the circuits are operated just at the point before self-oscillation starts. This condition is easily found in operation, since the instant self-oscillation starts, a whistle is heard in the phone or loud speaker when tuning to any station it is desired to receive. The use of a potentiometer to control oscillations in a tuned radio frequency circuit is illustrated in the diagram shown at Fig. 1. Potentiometers used in radio apparatus are usually made either from high resistance wire or from carbon granules. In electrical engineering work, the term potentiometer means, in general, an arrangement of circuits for measuring potential difference.

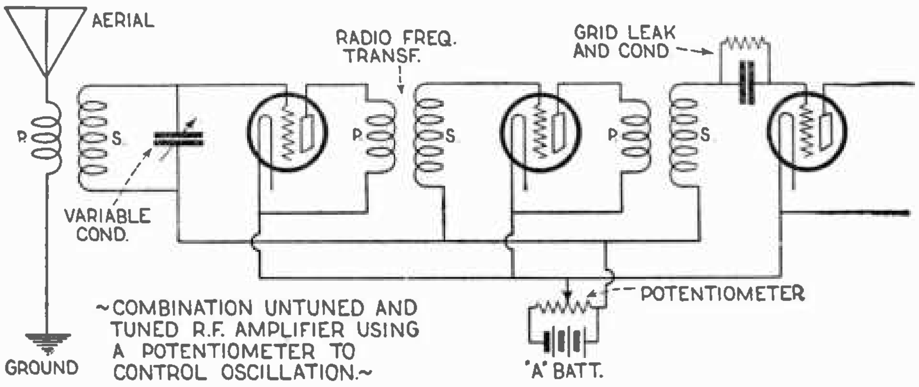
AERIAL
RADIO FREQ. TRANSF.
GRID LEAK AND COND
P
S
P
S
P
S
VARIABLE COND.
~COMBINATION UNTUNED AND TUNED R.F. AMPLIFIER USING A POTENTIOMETER TO CONTROL OSCILLATION.~
POTENTIOMETER
'A' BATT.
GROUND

Fig. 1. Tuned radio frequency circuit showing use of potentiometer to control oscillations.

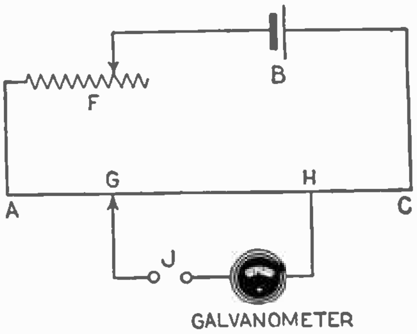

Fig. 2. A simple form of potentiometer used for electrical measurements.

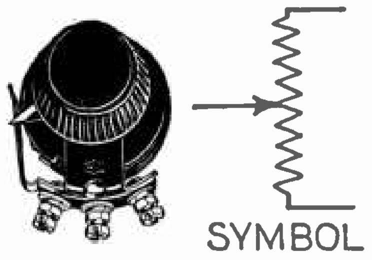

Illustration of a radio potentiometer, showing symbol at the right.

Using certain accessories, laboratory potentiometers can be used to measure voltages over all ranges, and by applying's *Ohm's law* (q.v.) the measurements obtained can be used to determine a wide range of current values.

The simplest form of potentiometer used for electrical measurements is shown in Fig. 2. AC is a resistance in which a constant current from the battery B is flowing. F is a regulating resistance used to compensate for the variations of the electromotive force of the battery. The current in AC is checked for constancy by noting that the drop in potential between the two points selected in it is equal to the electromotive force of a standard cell. The standard cell is connected in the circuit GHJ at J and the regulating resistance is adjusted until the sensitive galvanometer shows no deflection. With the assumption that AC has a uniform resistance throughout its length, and the current in it remaining constant, it is obvious that any other voltage not greater than the drop between A and C can be measured by connecting it in the circuit at J and shifting the points G and H until the galvanometer again comes to a balance.

A direct reading scale may be placed between A and C. For ordinary potentiometer measurements, the drop between A and C is made about 1½ volts as this is approximately the electromotive force of a standard Clark cell.

POULSEN ARC—An arc usually burned in hydrogen between a rotating carbon cathode and a water-cooled copper anode in a magnetic field. This arc is connected to an oscillating circuit, and due to its instability it maintains continuous oscillations which are utilized for the production of continuous waves for radio telegraphy and telephony.

POULSEN, Valdemar—Danish radio pioneer. Born in Copenhagen, November 23, 1869; he was educated at Copenhagen University and in 1893 he joined the Copenhagen Telephone Company. In 1903 he invented his famous system of arc transmission. Poulsen is a fellow of the Danish Society of Sciences. In 1900 he was awarded the Grand Prix at Paris.

POUNDAL—A unit of force. That force which acting on a mass of one pound gives it a velocity of one foot per second. It is equal to the weight of a pound mass divided by the acceleration of gravity; equal to 13,825 dynes or approximately half an ounce.

POWER—The rate of doing work. A certain amount of work may be done in a day or in a week. In the former case, more power is expended than in the latter case. Mechanical power is measured in *horse power* (q.v.). One horse power is equivalent to 33,000 foot-pounds per minute. Since power is the rate of doing work, the expression for power must contain a time element and a work element. Work is a measure of a force exerted through a distance. Hence the foot-pound measures work. Electrical power is the product of voltage times amperage. It is measured in *watts* (q.v.). (See *Kilowatt.*)

POWER AMPLIFIER—See *Power Amplification.*

POWER AMPLIFICATION—Amplification by means of apparatus consisting of a combination of power vacuum tubes, designed to carry larger currents than the ordinary tubes and specially designed transformers. Power amplifiers provide additional audio amplification and are used in connection with radio receiving sets where great volume is required as in the case of apparatus used for public address systems and for use with power operated loud speakers. (See *Amplifier.*)

POWER AMPLIFICATION COEFFICIENT—The ratio of the plate output power to the grid input power of a vacuum tube. This value may be as high as 10,000. The output power of a vacuum tube comes from the plate battery. When the grid is operated at low or negative voltage, the input power needed to produce the changes in grid voltage may be made very small compared with the output power of the tube.

POWER, APPARENT—The product of the voltage and the current in an alternating current circuit expressed in volt-amperes. This expression does not take into account any difference in *phase* (q.v.) between the voltage and the current. Thus in an extreme case (theoretical) where an alternating current flows in a purely inductive circuit, the current will lag behind the voltage by a phase angle of 90 degrees. In such a circuit, the *true power* (q.v.) would be zero, since the cosine of the phase angle between the current and the voltage would be zero, whereas the apparent power might be quite high and a heavy current might be flowing in the circuit. Apparent power is ordinarily expressed in *kilovolt-amperes* (q.v.). (See *Power Factor.*)

POWER FACTOR—The ratio of the *true power* (watts) in an alternating current circuit to the *apparent power* (volt-amperes). Where the wave form of the alternating current is sinusoidal, the power factor is equal to the cosine of the angle between the voltage and the current. Expressed as an equation, power P is equal to the product of the current I and the component of the impressed electromotive force which is in phase with the current, $E \cos \Phi$, where Φ is the angle of phase displacement between the current and the voltage ($P = IE \cos \Phi$). Expressed in another way, the power factor is the fractional amount which, if multiplied by the effective values of the current and the voltage, will give the value of the power expended. (See *Effective Electromotive force, Power, Apparent,* also *Power, True; Kilovolt-ampere, Kilowatt.*)

POWER, TRUE—The product of the *apparent power* and the *power factor* (q.v.) in an alternating current circuit. A wattmeter will measure true power, inasmuch as it takes into account the *phase displacement* (q.v.) between the current and the voltage. A voltmeter and an ammeter cannot be used to measure power, however, since the product of the two readings only gives the apparent power.

POWER TUBE—See *Vacuum Tube, Power.*

PRACTICAL UNITS—In the practical system of electrical units, the ohm, the ampere and the volt are the fundamental units. While the practical system is based originally on the Centimeter Gram Second systems, it is in no way dependent upon these systems for practical use. When used with the fundamental units of length and time, all other electrical as well as magnetic units can be derived from the fundamental units of the practical system. At the recommendation of the Chicago International Electrical Congress in 1893, the fundamental electrical units of the practical system were adopted as legal units in the United States by an act of Congress, July 12, 1894.

P R B—International Radiotelegraphic Convention abbreviation. Followed by question mark it means, "Do you wish to communicate by means of the International Signal Code?" Without the question mark it means, "I wish to communicate by means of the International Signal Code." (See *Abbreviations, International Radio Telegraphic Conventional.*)

PRESSURE—See *Potential, Electric.*

PRESSURE, ELECTRICAL—The term is synonymous with *voltage* (q.v.) or *electromotive* force (q.v.). Electrical pressure is the cause of current flow. It is measured in *volts* (q.v.). The flow of electricity along a conductor is analogous to the flow of water through a pipe. In order to make water flow, a pump is often used to force the water or in other words to apply pressure. Similarly, electrical pressure must be applied in order for an electric current to flow.

PRIMARY CELL—A device for transforming chemical energy into electrical energy. It consists of two different conductors placed in a liquid which acts chemically on one conductor, or more on one conductor than on the other, thus keeping a difference of potential between them. While the cell is being discharged either one or both of the conductors is used up and the liquid is also consumed. In many types of primary cells these can be replaced. The liquid is called an *electrolyte* (q.v.). In some cases an acid is used while in others a salt solution is the *electrolyte.* (q.v.) (See *Cell, Cell Secondary; Cell, Theory of Primary; Electrolysis, Local Action, Polarization,* also *Storage Cells.*)

PRIMARY CIRCUIT—A circuit supplying current to another which is called the secondary circuit. The two circuits are usually coupled by means of a transformer, the primary winding being in the primary circuit and the secondary winding being in the secondary circuit.

PRIMARY CURRENT—See *Current, Primary.*

PRIMARY ELECTRONS—The main stream of electrons given off by the hot filament of a vacuum tube, as differentiated from *secondary electrons* (q.v.) which may be detached from the plate by the impact of the primary electrons. (See *Negative Resistance.*)

PRIMARY TUNING INDUCTANCE—A variable inductance in the primary circuit of a radio transmitter. It is also referred to as the *aerial tuning inductance.* The aerial inductance coil may be continuously variable utilizing a sliding contact which bears on an edgewise copper strip or it may be of the drum type, in this case being made of stranded copper cable. (See *Inductance Coils,* also *Pancake Coils.*)

PRINCE, Charles Edmond—British radio expert. Born in 1874 at Cape Town and educated at Clifton and Faraday House, he joined the Marconi Wireless Telegraph Company in 1907, where he was concerned chiefly in research work on wireless telephony. In 1909 he carried out experiments in Italy and Switzerland with the first Marconi Field Station and made a number of valuable improvements in the Bellini-Tosi direction finder. In the World war, he joined the Royal Flying Corps and in 1915 he installed the first radio telephone on aircraft.

PRINTER, MORSE—See *Morse Inker.*

PROTON—A unit of positive electricity. The smallest quantity of positive electricity capable of existing in a free state. A minute particle or quantity of positive electricity carried by the nucleus of a hydrogen atom. See *Electron Theory,* also *Free Electrons.*

PULSATING CURRENT—A periodic current, that is to say a current passing through the successive equal cycles of values, the average value of which is not zero. A pulsating current is the sum of an alternating and a direct current. A current which varies regularly in magnitude is said to be pulsating while if it remains the same in magnitude it is called continuous. (See *Continuous Direct Current,* also *Current, Pulsating Direct.*)

PULSATING DIRECT CURRENT—See *Pulsating Current.*

PUPIN, Michael—American radio authority. Born in Hungary, October 4th, 1858, he came to the United States at the age of 10, and was educated at Columbia University. He also studied at Cambridge University, England. In 1891 he was appointed professor of mathematical physics at Columbia University, where he carried out an important series of researches in radio telegraphy and telephony. He is the holder of a large number of patents relating to telephony and radio, has done valuable research work on reso-

Michael Pupin.

nance and has developed radio circuits of extreme selectivity.

PUSH-PULL AMPLIFICATION—A method of connection in a vacuum tube receiving set, for supplying more power to the loud speaker than that ordinarily obtainable from a one or two stage audio amplifier. In the push-pull amplifier two tubes are used in the final stage of the amplifier, connected in such a way that they are used alternately on the two halves of each audio frequency cycle. A push-pull amplification circuit is shown in Figure 1. The first tube at the left gives the first stage of amplification and the other two in combination give the second stage. An ordinary audio frequency amplifier transformer is connected to the grid and filament of the first tube. Special transformers are usually used for the other two transformers. These have connections brought out from the center point of one of the windings as shown. The transformer between the two stages has a split secondary while the last transformer has a split primary. In cases where a tapped wound transformer is not available, two transformers can be used instead. In this case the two primary windings are connected in series and the two secondary windings are also connected in series. Figure 2 shows the method of connecting two transformers for this purpose, a connection being made to the mid-point between the secondary windings. It is essential that the two windings in series be connected so that they are continuous in the same direction around the transformer cores. In theory, the push-pull amplifier uses one tube during one half of the cycle and the other tube during the other half of the cycle. While the terminal of the secondary transformer winding leading to the grid of the tube is at a positive potential with respect to the middle of the winding, the tube is effective. During the remaining half of the cycle when the terminal is negative with respect to the tap, the tube is not operating. The action however is such that one tube is effective at all times. Although ordinary amplifier tubes can be used in the push-pull amplifier circuit it is desirable to use special power tubes which have a high amplification coefficient.

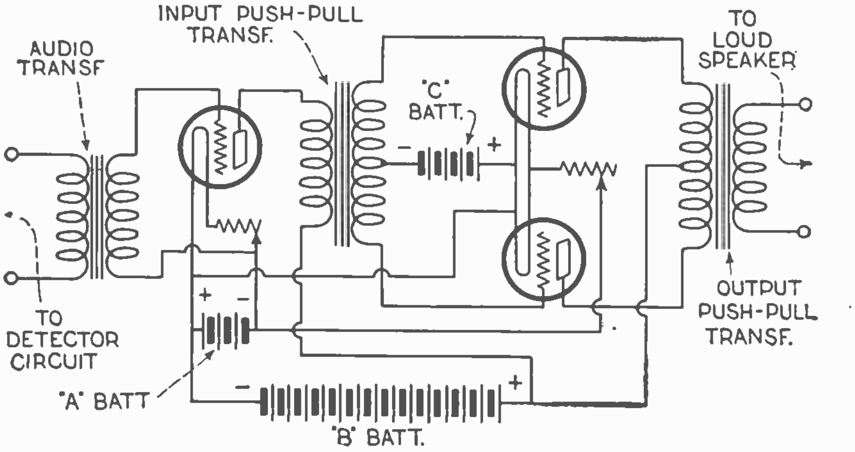

Fig. 1. Diagram of a push-pull amplification circuit.

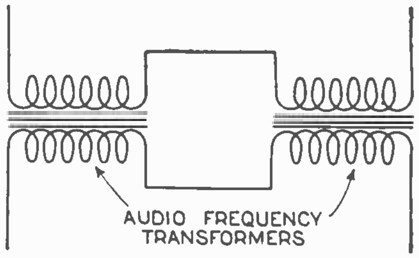

Fig. 2. Method of connecting two transformers where a tapped wound tranformer is not available.

PYRITES—Compounds of iron or copper and sulphur. Three of these are *Ferrous Sulphide* (FeS), *Bisulphide* (FeS_2), and an intermediate known as *Magnetic Pyrites* (Fe_3S_4). Pyrites crystals have been used in radio as detector crystals. (See *Crystal,* also *Crystal Detector.*)

PYRON DETECTOR—A *crystal detector* (q.v.) having an iron pyrites crystal and a copper or other metallic contact point.

Q

Q SIGNALS—International Radiotelegraphic Convention abbreviations. When the signal is followed by a question mark it asks the question. Otherwise the answer is indicated. A list of these abbreviations and their meanings follow:

Abbrev'n	*Question*	*Answer*
QRA	What ship or coast station is that?	This is . . . .
QRB	What is your distance?	My distance is . . . .
QRC	What is your true bearing?	My true bearing is . . . degrees.
QRD	Where are you bound for?	I am bound for . . .
QRF	Where are you bound from?	I am bound from . . . . .
QRG	What line do you belong to?	I belong to the . . . Line.
QRH	What is your wave length in meters?	My wave length is . . . . meters.
QRHH	What tune shall I adjust for?	Adjust to receive on tune . . . .
QRJ	How many words have you to send?	I have . . . . words to send.
QRK	How do you receive me?	I am receiving well.
QRL	Are you receiving badly? Shall I send 20 . . . — . for adjustment?	I am receiving badly. Please send 20 . . . — for adjustment.
QRLL	Request permission to test . . . minutes.	Permission to test granted.
QRM	Are you being interfered with?	I am being interfered with.
QRN	Are the atmospherics strong?	Atmospherics are very strong.
QRO	Shall I increase power?	Increase power.
QRP	Shall I decrease power?	Decrease power.
QRQ	Shall I send faster?	Send faster.
QRS	Shall I send slower?	Send slower.
QRT	Shall I stop sending?	Stop sending.
QRU		I have nothing to transmit. I have nothing for you.
QRV	Are you ready?	I am ready. All right now.
QRW	Are you busy?	I am busy (or, I am busy with . . . . Please do not interfere).
QRX	Shall I stand by?	Stand by. I will call you when required.
QRY	When will be my turn?	Your turn will be No. . . . .
QRZ	Are my signals weak?	Your signals are weak.
QSA	Are my signals strong?	Your signals are strong.
QSB	Is my tone bad? Is my spark bad?	The tone is bad. The spark is bad.
QSC	Is my spacing bad?	Your spacing is bad.
QSD	What is your time?	My time is . . . .
QSF	Is transmission to be in alternate order or in series?	Transmission will be in alternate order.
QSG		Transmission will be in a series of 5 messages.
QSH		Transmission will be in a series of 10 messages.
QSJ	What rate shall I collect for?	Collect . . . .

Abbrev'n	*Question*	*Answer*
QSK	Is the last radiogram cancelled?	The last radiogram is cancelled.
QSL	Did you get my receipt?	Please acknowledge.
QSM	What is your true course?	My true course is . . . . . degrees.
QSN	Are you in communication with land?	I am not in communication with land.
QSO	Are you in communication with any ship or station (or: with . . . .)?	I am in communication with . . . . (through . . . . .
QSP	Shall I inform . . . . that you are calling him?	Inform . . . . that I am calling him.
QSQ	Is . . . . . calling me?	You are being called by . . . .
QSR	Will you forward the radiogram?	I will forward the radiogram.
QSS	Are my signals fading?	Your signals are fading.
QST	Have you received the general call?	General call to all stations.
QSU	Please call me when you have finished (or: at . . o'clock).	Will call when I have finished.
QSV	Is public correspondence being handled?	Public correspondence is being handled. Please do not interfere.
QSW	Shall I increase my spark frequency?	Increase your spark frequency.
QSX	Shall I decrease my spark frequency?	Decrease your spark frequency.
QSY	Shall I send on a wave length of . . . . meters.	Let us change to a wave length of . . . . meters.
QSZ		Send each word twice. I have difficulty in receiving you.
QTA		Repeat the last radiogram.
QTC	Have you anything to transmit?	I have something to transmit. I have one or more radiograms for . . . . . .
QTE	What is my true bearing?	Your true bearing is . . . . degrees from . . . . . .
QTF	What is my true position?	Your true position is . . latitude . . . . . .longitude.

QUADRANT—A term formerly in use as the unit of *inductance* (q.v.), now called the *henry* (q.v.). Inductance has the same dimensions as length and this unit is equivalent to 10^9 cm., this being approximately the length of one quadrant of the earth.

QUADRANT ELECTROMETER—A sensitive *electrometer* (q.v.) which in its calibrated form is known as an *electrostatic voltmeter* (q.v.). In the usual type, the "needle" consists of a thin aluminum plate suspended on delicate knife edges, with a pointer extending from the upper part to a scale. Two pairs of quadrant plates, having opposite faces metallically connected together, are placed on either side of the movable plate and parallel to its face. Each pair is insulated from the other pair. When the pairs of quadrants are subjected to a difference of potential, the needle, if highly charged, will be attracted into one pair or repelled out of the other. (See *Kelvin's Electrostatic Voltmeter.*)

QUADRATURE—The differing in *phase* (q.v.) of two alternating current quantities by 90 degrees. For example, when the current lags behind the impressed voltage by 90 degrees, as in the case of pure inductance in a circuit, the current and the voltage are said to be in quadrature. (See *Lag, Lead,* also *Leading Current.*)

QUANTITY OF ELECTRICITY—When referring to static electricity, it is the amount of the charge. Referring to electricity in motion, or an electric current, it is the amount of current flowing in a given time. Quantity of electricity is measured in *coulombs* (q.v.). The coulomb is the quantity of electricity which has passed when a current of one ampere flows for one second. The symbol which is usually used to signify quantity of electricity is Q.

QUANTOMETER—A type of ballistic galvanometer in which the swing of the movable system is proportional to the quantity of electricity passing through the instrument. A *galvanometer* (q.v.) of this type is used for magnetic testing as in the case of the *fluxmeter* (q.v.).

QUARTER PHASE—A two-phase alternating current system. An alternating current system having two currents differing in phase by 90 degrees or in other words by one-quarter of a cycle.

QUARTZ LAMP—A mercury vapor lamp which utilizes a tube of quartz instead of a glass container. Increased current densities can then be employed, since higher temperatures can be used with quartz than with glass.

QUARTZ OSCILLATOR—The piezo-electric quartz plate may be used as a *resonator* (q.v.) or as an *oscillator* (q.v.). When used as an oscillator it may be used either as a *master oscillator* controlling the frequency of a radio transmitting station's output or as a frequency indicator. All three of these uses are applicable to the work of maintaining station frequencies constant. The applicability of the piezo oscillator as a frequency indicator in broadcasting stations has been tested carefully by the United States Bureau of Standards, and found to be useful and satisfactory. This method gives a means of maintaining an extremely accurate check of the transmitting station frequency. (See *Master-Oscillator System,* also *Frequency Meter.*)

QUENCHED GAP—See *Quenched Spark Gap.*

QUENCHED SPARK—See *Quenched Spark Gap.*

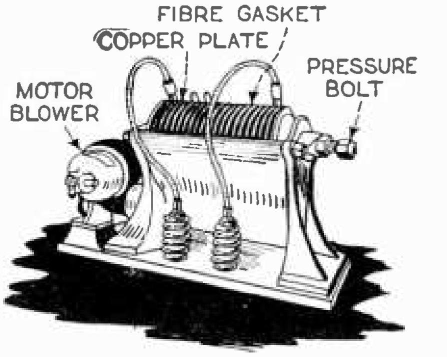

A typical quenched spark plug.

QUENCHED SPARK GAP—A spark discharger in a wireless telegraph transmitting circuit in which the spark is deionized to prevent the gap remaining conductive after the spark has passed. In this way only a few sharply defined oscillations are permitted to occur. The particular advantages of the quenched spark gap is that it enables the aerial to oscillate at its own frequency regardless of the period of the spark system. In one form of construction, a number of heavy copper plates are used, separated by micanite, fiber insulating washers or other insulating material. These are placed in an iron rack and compressed by means of a pressure bolt. The inside edge of the washer rests upon a groove which is cut in each plate. This makes it impossible for the spark to discharge over the edge of the washer. In order to make the discharge surface airtight, the washers are specially treated and as a result this helps in quenching the primary oscillations and makes the discharge noiseless. A small motor driven blower is mounted at the base of the spark discharger for the purpose of forcing air through the base of the rack against the cooling flanges of the plates. (See *Deionization, Musical Spark, Kinraidy Spark Gap,* also *Quenched Spark Transmitter.*

QUENCHED SPARK TRANSMITTER—A radio telegraph transmitter which employs a quenched spark gap. The quenched spark transmitter is not used for very high powers as it is not satisfactory for such use. In this case, a form of disc discharger is usually used. The quenched spark transmitter, however, has no moving parts and is noiseless in operation. In addition it gives large values of antenna current because closer coupling is possible. (See *Quenched Spark Gap.*)

QUICK BREAK SWITCH—A switch provided with auxiliary contacts which are released after the main contacts. The auxiliary contacts are separated rapidly, regardless of the speed with which the switch handle is pulled. This action is obtained by the tension of a spring connected between the main switch blade and the auxiliary blade, as shown in the illustration.

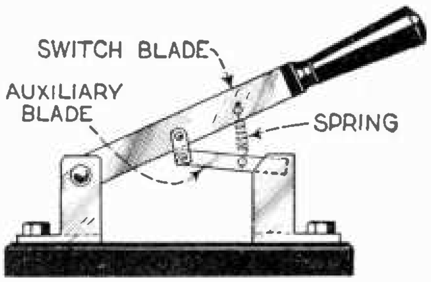

Quick Break Switch. Note the auxiliary spring-controlled blade.

QUIESCENT AERIAL—In a radio telephone transmitter, an aerial operated in such a manner that the carrier wave is suppressed when transmission is not actually taking place.

R

"R"—Symbol for *resistance* (q.v.)

RADIANT ENERGY—Energy transmitted through space by means of vibrations of the ether, such as electromagnetic energy, light or heat. (See *Actinic Ray.*)

RADIANT HEAT—Infra-red *radiation* (q.v.) as for example radiation from substances at temperatures insufficient to cause visible light rays.

RADIATING CIRCUIT—An electrical circuit which emits energy in the form of ether waves. In a transmitting system, the aerial circuit. Any circuit which radiates energy is a radiating circuit.

RADIATION—The transference of energy in waves through space which is not necessarily occupied by matter. Transference of energy through *ether* (q.v.). Electric or electromagnetic radiation refers to the radiation by means of electromagnetic induction. The radiation field is transmitted by wave motion. Various types of electromagnetic waves may be sent out by a radiating system. In one system the waves are known as *damped waves* (q.v.), in another *undamped waves* (q.v.) are used.

RADIATION EFFICIENCY—The radiation efficiency of an antenna at a given wave length is the ratio of the power radiated to the total power delivered to the antenna (resistance).

RADIATION FROM ANTENNA—In the case of a simple vertical antenna, electromagnetic waves are radiated in semi-loops since the lower end of the aerial is grounded. Fig. 1 gives an idea of the way in which the intertwined electric and magnetic fields radiate from the antenna. It should be noted that the electric and the magnetic fields are in planes which are at right angles to each other. The electric lines are shown in elevation while the magnetic lines appear in plan, with the wave form common to both being shown between. At any single point in space, the lines separate and come together like a bellows. After the wave has travelled a certain distance, the lines can be regarded as sections of planes and the electric lines can be represented by the diagram shown in Fig. 2. The magnetic lines can be shown by the same figure, provided this is considered to have been rotated about its horizontal axis through an angle of 90 degrees. Regardless of whether the aerial is of the open or closed coil type, the radiation will take place in the same general way, *loops* (q.v.) being formed and detached as described. (See *Electromagnetic Waves,* also *Height of Aerial.*)

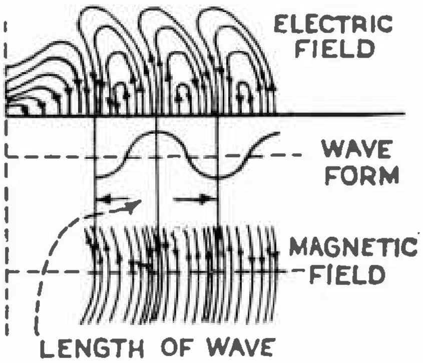

Fig. 1. Electric and magnetic fields about a vertical wire aerial.

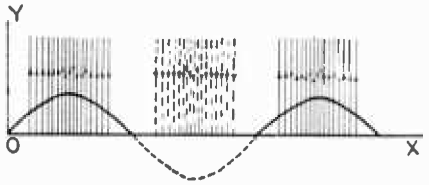

Fig. 2. Electric lines of force shown after the wave has travelled some distance from the aerial.

RADIATION RESISTANCE—The ratio of the total power radiated by an antenna to the square of the effective current at the point of maximum current. The property of an antenna by means of which radiation of energy takes place. Some of the power supplied to a circuit, carrying radio currents, is radiated in the form of electric waves. The radiation is a measure of the useful work obtained from the circuit. The greater the power radiated in comparison with the power dissipated in the circuit itself, the better will be the transmission. At any frequency, the power radiated is proportional to the square of the current flowing, hence the radiative effect may be regarded as causing artificially, an increase in the effective resistance of the circuit. This resistance increase is known as "radiation resistance" and is directly proportional to the square of the frequency or inversely proportional to the square of the wave length. (See *Antenna Resistance.*)

RADIATION TRANSMISSION—See *Radiation from Antenna.*

RADIO—The transmission of intelligence by means of electromagnetic waves through the ether. Radio falls under two general classifications; *radio telephony* (q.v.) and *radio telegraphy* (q.v.). (See *Radio Channel, Radio Communication, Radio Control,* also *Radio Frequencies.*)

RADIO-ACTIVITY—The emanation of energy by certain substances, such as radium, caused by a continuous detachment of electrons from the substance.

RADIOCAST—A term sometimes used instead of *Broadcast* (q.v.). To transmit speech, music or other sounds by *radio telephony* (q.v.).

RADIO CHANNEL—A band of wave lengths or frequencies of a width sufficient to permit of its use for radio communication without the radiation of subsidiary waves of more than a certain intensity at wave lengths or frequencies outside of such a band.

RADIO COMMUNICATION—The transmission of intelligible signals by means of electromagnetic waves. The signals may be in the form of code as in radio telegraphy or in the form of speech or music as in radio telephony. They may be transmitted through the *ether* (q.v.) without the use of intervening conductors as in ordinary radio, or along conductors as in *wired wireless* (q.v.).

RADIO COMPASS—A simple form of radio *direction finder* especially used for marine work. The radio compass depends for its principle of operation upon the directional effect of a coil antenna. Hence a coil antenna mounted on a vertical axis so that it can be rotated freely, is an essential part of the radio direction finder. By rotating the coil while a particular station is transmitting, it is possible to fix the line of direction of the station. The coil may be mounted so that the maximum signal will be received when the transmitting station lies in the plane of the coil. The device may also be set for minimum signal reception in which case the transmitting station lies in a direction perpendicular to the face of the coil. It is possible to obtain a much sharper determination of direction by setting the coil on the minimum position and hence this method is generally used for direction finder work. (See *Goniometer.*)

RADIO CONTROL—The control of motor driven devices such as marine vessels, trains, torpedoes, automobiles, etc., by means of special radio receiving apparatus designed to respond to radio impulses. E. P. Glavin's radio controlled automobile is an example of the radio controlled vehicle. *John Hays Hammond, Jr.* (q.v.) has invented radio controlled torpedoes, radio controlled ships and also radio controlled air craft.

R.F.—Abbreviation for radio frequency.

RADIO FREQUENCIES—Frequencies higher than those corresponding to normally audible sound waves. (See *Frequency, Radio; Frequency, High; Audio Frequency;* also *Frequency, Low.*)

RADIO FREQUENCY COILS—*Inductance coils* (q.v.) used to carry radio frequency currents or to couple radio frequency circuits. Radio frequency *transformers* (q.v.) may be of either iron core or air core types although the air core transformers are more generally used. (See *Low Loss Coils,* also *"D" Coils.*)

RADIO FREQUENCY CURRENT—An alternating current having a frequency of from 20,000 to 2,000,000 cycles per second. There is no hard and fast rule regarding the upper and lower limits, since much higher or much lower frequencies may still be considered as radio frequency currents. (*See High Frequency Current.*)

RADIO FREQUENCY SELECTIVITY—The radio frequency selectivity of a simple element of a receiving system is the ratio of resonant response (in terms of effective voltage or current measured at the indicator) to the nonresonant response when the radio frequency portions of the elements of that system are detuned by one percent of the resonant frequency. In this case, a simple element refers to a combination of an inductance, a capacitance and optionally a resistance.

RADIO FREQUENCY SIGNALS—Currents flowing in a radio receiving circuit up to the point of rectification. The currents flowing in the antenna circuit are radio frequency currents corresponding to the currents flowing in the transmitting aerial. (See *Amplifier, Radio Frequency.*)

RADIO GONIOMETER—See *Goniometer.*

RADIOGRAM—A message sent in code by radio. A radio telegraphic message.

RADIOPHARE—A radiotelegraphic "lighthouse" intended to aid navigation by emitting characteristic signals. By estimating the bearings of two charted radiophares, the mariner may determine the position of his ship.

RADIO TELEGRAPHY—The transmission of messages by means of electromagnetic waves using dot and dash code signals. A transmitter furnishes a source of high frequency alternating current which oscillates rapidly in the antenna system thus sending out electromagnetic waves. These are intercepted at the receiving station by the receiving aerial. At the receiving aerial, the electromagnetic waves are changed back into alternating current which is rectified by a detector in the receiving apparatus, then amplified and made audible in a *head telephone* (q.v.) or *loud speaker* (q.v.). Of course, a means of interrupting the current to produce dots and dashes must be present in the transmitter and suitable tuning devices are used in both the transmitter and the receiver.

RADIO TELEPHONY—The transmission of speech, music or other sounds by means of electromagnetic waves. In the standard systems of radio telephony continuous waves are used of a much higher frequency than those of audible sounds. The amplitude of these is modulated by the connection of a microphone in the transmitting circuit. The receiving apparatus is similar to that employed in radio telegraphy and is arranged so that the head set or loud speaker responds to the modulations of the waves and reproduces the sounds as originally produced at the transmitter.

RADIO TRANSMISSION—The sending of signals, speech, music, or other sounds by means of electromagnetic waves. The transmission of intelligence through the ether. (See *Radio Communication, Radio Control, Radio Telegraphy*, also *Radio Telephony.*)

RAT-TAIL—A single wire, or a group of wires used to connect the main part of an aerial with the *lead-in* (q.v.) wire.

RATIO—Relative values of quantities of the same kind, or number of times one quantity is contained in the other.

RATIO OF TRANSFORMATION—The ratio between the primary and secondary windings of a transformer. This determines the ratio of the primary to the secondary voltages. Thus if a transformer has a ratio of transformation of 3 to 1, then the secondary will have three times as many turns as the primary and the secondary voltage will be, approximately triple that of the primary voltage.

RATIO OF UNITS—The ratio of the unit pole in the electromagnetic system to the unit pole in the electrostatic system, or of the unit charge in the electromagnetic system to the unit charge in the electrostatic system. Its numerical value is approximately 3×10^{10}.

RAYLEIGH, John William Strutt—British physicist. Born at Langford Green, Essex, November 12th, 1842. He was educated at Trinity College, Cambridge. In 1879 he was appointed Cavendish professor of experimental physics in succession to Clerk-Maxwell, a post he held until 1884. In 1887 he became professor of natural philosophy at the Royal Institution and was appointed president of the Royal Society in 1905. In 1908 he was appointed Chancellor of Cambridge University, and died June 30th, 1919. Rayleigh was considered by many to have been one of the most brilliant experimental physicists of the nineteenth century. He not only had a remarkable power of mathematical analysis, but was equally skillful in carrying out the experimental proofs of his theoretical researches. Many of his experiments were made with home-built apparatus. He threw a fresh light on nearly every branch of physics, from the theory of gases to wave theories and electric and magnetic problems. Due to the care with which he carried out his experiments, he was the first to detect the presence of neon in the atmosphere, this being the forerunner of the discovery of a number of inert gases in the atmosphere. The Rayleigh balance for absolute current measurement is as well known as it is important. Lord Rayleigh was one of the original members of the Order of Merit. In 1882 he was awarded the Royal Medal, in 1899 the Copley Medal of the Royal Society and in 1904 the Nobel Prize for physics.

REACTANCE—That property of an electric circuit, aside from its ohmic resistance, which tends to oppose the flow of an alternating current. Mathematically, reactance is defined as the square root of the difference between the square of the *impedance* (q.v.) and the square of the effective resistance of a given portion of an electric circuit. Expressed as an equation reactance $x = \sqrt{z^2 - r^2}$ where z is the impedance and r the resistance of the part of the circuit under consideration. Reactance due to capacity in a circuit is capacity or condensive reactance. The reactance of a condenser of capacity C to an alternating current (of sine wave form) of frequency f is expressed by the equation $X_c = \dfrac{1}{2\pi fC}$. Reactance due to inductance in a circuit is *inductive reactance* (q.v.). The reactance of an inductance coil of inductance L to an alternating current (of sine wave form) of frequency f is expressed by the equation $X_L = 2\pi fL$. Reactance is measured in *ohms*.

REACTANCE COIL—REACTION COIL—Terms used in England to describe the *tickler coil* (q.v.) used in regenerative sets to couple back part of the energy in the plate or output circuit to the input circuit.

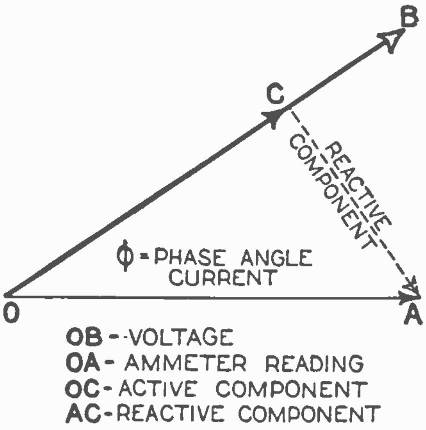

The current represented by vector OA may be considered as divided into two components, OC in phase with the voltage, and CA the reactive component.

REACTIVE COMPONENT—The component of an alternating current used to overcome the reactance of the circuit. This component acts at an angle of 90 degrees from the voltage. It is also known as *wattless current* (q.v.).

REACTIVE CURRENT—See *Reactive Component.*

REACTIVE DROP—The portion of the voltage drop in an alternating current circuit, due to the reactance.

REAL POWER—See *True Power.*

RECEIVER—A term rather loosely applied to cover the entire radio receiving apparatus within a station. Also sometimes used as a synonym for *Detector* (q.v.)

RECEIVING AERIAL—A wire or group of wires suspended at a suitable height and connected to a radio receiving set. The purpose of the receiving aerial is to intercept the electromagnetic waves. It should be placed well above surrounding buildings if possible, and should be well insulated from the ground. (See *Aerial*, also *Directional.*)

RECEIVING CIRCUIT—Any electrical circuit used in connection with the reception of radio. The more important vacuum tube receiving circuits are the *super-heterodyne* (q.v.), the *neutrodyne* (q.v.), the *reflex circuit* (q.v.), the *regenerative circuit* (q.v.) and the *tuned frequency circuit* (q.v.). There are various modifications and combinations of these circuits.

RECEPTACLE—A form of socket into which a plug is inserted for making electrical connections.

RECONSTRUCTED MICA—See *Micanite.*

RECORDER—A telegraphic instrument which automatically records messages sent over a telegraph system.

RECORDING WATTHOUR METER—An instrument for measuring and recording the total amount of electrical energy being consumed in a circuit. (See *Integrating Wattmeter.*)

RECTIFICATION—The changing of alternating current to direct current. In the usual radio receiving circuit rectification is accomplished by means of the vacuum tube, although the ordinary crystal detector rectifies the alternating current also. Before a storage battery can be charged from an alternating current source, the current must first be changed to direct current. Hence the battery charger designed to work on alternating current must be equipped with a rectification device, or rectifier (q.v.).

RECTIFIED CURRENT—An *alternating current* (q.v.) which has been changed to uni-directional or direct current either by preventing alternate waves from flowing or by reversing their direction. (See *Rectified Signals, Rectification*, also *Rectifier.*)

RECTIFIED SIGNALS—Radio signals which have passed through a detector, such as a vacuum tube or a crystal. Rectification of the oscillations results in audible or audio frequency currents which can be heard in a telephone receiver. (See *Rectification, Rectifier, Rectifying Tube*, also *Rectifying Action of Crystal Detector.*)

RECTIFIER—An apparatus for converting alternating current to unidirectional or direct current. Examples of rectifiers are certain crystals used as crystal detectors, the commutators of direct current motors and generators, vacuum tubes, electrolytic devices used for charging batteries, etc. (See *Rectified Signal, Rectification, Rectifying Tube, Rectifying Action of Crystal Detector, Rectified Current, Rectifier, Full Wave, Full Wave Rectification*, also *Half Wave Rectifier, Electrolytic Rectifier, Mercury Arc Rectifier, Charger, Storage Battery; Chemical Rectifier; Tungar; Rectigon.*)

RECTIFIER, FULL WAVE—A device for converting alternating current to direct current, in which both the positive and negative alternations are utilized. This differs from the half-wave rectifier which rectifies by suppressing alternate waves. (See *Full*

Wave Rectification, also *Half Wave Rectifier.*)

RECTIFYING ACTION OF CRYSTAL DETECTOR—The crystal detector has the property of allowing radio frequency alternating currents to flow through it in one direction only. In other words the alternating current is rectified. (See *Crystal Detector-Theory of Operation, Crystal Detector, Crystal Rectifier, Rectified Signals, also Rectified Current.*)

RECTIFYING TUBE—A vacuum tube used to convert alternating current into direct current. Rectification can be accomplished by a two-electrode tube, having filament and plate. In this case the electron flow is from the hot filament to the plate. A two-electrode vacuum tube made especially for rectifying purposes is known as the *kenotron* (q.v.). Another recent type of rectifying tube, known as the *helium tube* (trade name, *Raytheon*) utilizes a different operating principle, known as the "short path" principle, whereby the rarefied gas acts as an insulator between points which are in close proximity. This is in apparent contradiction of the observed phenomenon that the smaller the distance between two points the more readily a spark will jump between them due to the ionization of the gas. However, if the distance is small enough and a suitable gas is used at a low enough pressure, an electron may encounter no gas molecules in its path between the points and there will be no ionization by collision. Consequently, the inert gas helium, may be made to act as a perfect insulator at low pressures. Furthermore, it has been found that when a larger electrode of a gas conduction tube is negative, there is a greater current flow than when the smaller electrode is negative and that the smaller the positive relative to the negative electrode, the smaller is the back current and hence the more complete is the rectification. Figure 1 shows a typical helium tube, while Figure 2 illustrates its construction. The two small positive electrodes *A*, *A*, are carried through two small glass tubes imbedded in a lava insulating block *L* so as to project very slightly into a lava cup *C* whose wells constitute the negative electrode, being connected to the negative terminal through the base. The diameters of the small wires and the diameters and positions of the holes whereby they enter the cup are so proportioned as to give the necessary short path to the negative electrode. The cup *C* contains helium gas at such low pressure as to prevent gaseous conduction. This gives an insulation of great reliability and long life and makes possible extremely small anode surfaces which reduce the back current to a negligible quantity.

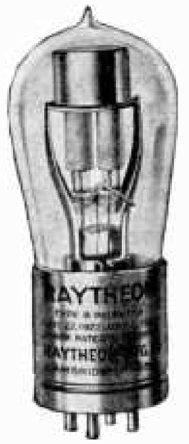

Fig. 1. A Raytheon tube.

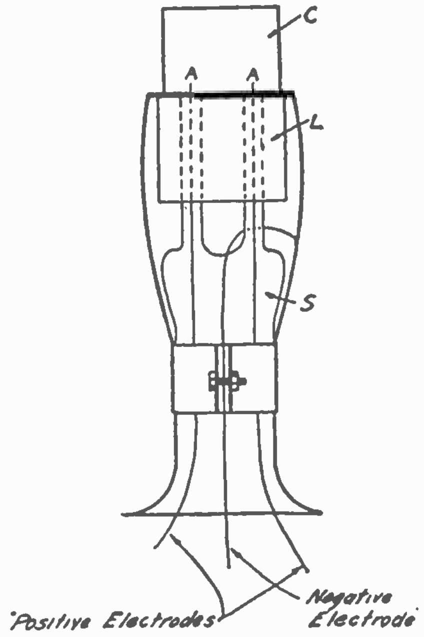

Fig. 2. Construction of the Raytheon tube.

RECTIFYING VALVE—This usually refers to a *rectifying tube* (q.v.) although it may refer to an *electrolytic rectifier* (q.v.)

RECTIGON—The trade name of a storage battery charger which utilizes the vacuum tube rectifier principle. (See *Charger, Storage Battery*, also *Rectifying Tube.*)

RED MAGNETISM—An obsolete term for the lines of force coming from the north pole of a magnet.

REDUCE WAVE-LENGTH OF AERIAL—The wave-length of an antenna may be reduced by the use of series condensers. However, this method should be avoided where possible, since it results in a loss of power. Shortening the aerial is a preferable method of reducing wave-length.

REFLECTING GALVANOMETER—See *Mirror Galvanometer.*

REFLECTION—Radio waves are reflected and also refracted when they pass into a region of different dielectric constant. A perfect conducting sheet would reflect all of the wave energy striking it. A conductor parallel to the direction of motion of a wave serves to guide the direction of the wave. The reflection and other modifications of radio waves serves to explain *fading* (q.v.) and other irregularities of broadcast signals. It has been found that broadcast signals will have most uniform intensity when long waves are used and during the daytime. Moreover, transmission will be better over a uniform conducting surface, such as the ocean. It is assumed that the space through which a radio wave travels is bounded by the surface of the earth and by an upper region which during the day is a fairly good conductor. The transmitted waves are guided by the two conducting mediums but in addition are assisted materially by reflection especially during the time that the air is partially ionized by the radiation of the sun.

REFLECTION COEFFICIENT—(Of a surface of discontinuity between two media.) The ratio of the reflected field intensity near the surface to the incident field intensity near the surface.

REFLEX—See *Reflex Circuit.*

REFLEX CIRCUIT—A radio receiving circuit invented by Marius Latour in which the vacuum tubes are made to perform double duties as both radio and audio frequency amplifiers. The incoming radio frequency current is amplified at radio frequency, rectified by a detector (which may be a crystal detector) and then amplified at audio frequency, using the same tube. A typical four tube reflex set uses the second and third tubes as amplifiers simultaneously for both the radio and the audio frequency amplification. As a result the four tubes are made to produce the effect of six. Reflexing may be accomplished in a number of different ways. In some cases all the tubes are made to work twice. In other cases, only a part of the tubes are used for dual amplification. A variation of the reflex circuit, is the *Inverse Duplex Circuit* (q.v.). In this system the incoming signal reaches the first tube where it is amplified at radio frequencies and passes to the second tube, where it is again amplified at radio frequencies. From the second radio frequency tube, it is passed to the detector tube, where it is converted to audio frequency. The current then flows back to the second radio frequency tube which is now used as an audio frequency amplifier. Thence it flows to the first radio frequency tube which acts as the second

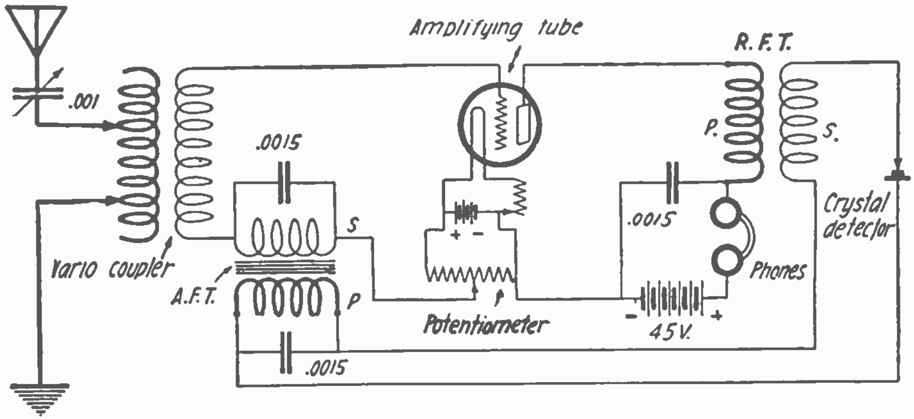

A typical reflex circuit using one tube and a crystal detector.

stage of audio frequency amplification.

REFRACTION—The refraction of electric waves is analogous to the refraction of light. The electric (radio) waves are refracted when they pass into a region of different dielectric constant. That is to say, their direction is slightly altered. (See *Fading, Gliding Theory,* also *Reflection.*)

REFRACTION AND REFLECTION OF ELECTROMAGNETIC WAVES—See *Reflection,* also *Refraction.*

REGENERATION—The amplifying properties of the three-electrode vacuum tube can be used to obtain what is termed *regeneration.* Since it is possible to have greater output energy than input energy, part of the output may be returned to the input side, thus resulting in amplification of energy or in regeneration. By *feeding back* the correct amount of energy and in the right phase relation a constant reamplification may be obtained, and the tube can be made to act as a generator of sustained oscillations. (See *Armstrong Circuits, Feed-Back, Feed-Back Coil, Feed-Back Coupling, Feed-Back Effect,* also *Regenerative Circuit.*)

REGENERATIVE—See *Regenerative Circuit.*

REGENERATIVE CIRCUIT—A method of connecting up vaccum tubes for radio reception in which the amplified variations of the plate circuit are superimposed inductively on the grid or imput circuit thus producing a reinforcing effect which results in increased sensitivity. (See *Armstrong Circuits.*)

REGENERATIVE COUPLING—A radio receiving system designed to increase amplification in a vacuum tube, by coupling the plate or output circuit back to the grid or input circuit thus producing regeneration. (See *Feed-Back, Feed-Back Coil, Feed-Back Coupling, Feed-Back Effect,* also *Regenerative Circuit.*)

REINARTZ CIRCUIT—A selective regenerative circuit devised by John L. Reinartz in which a *spider-web* inductance coil is used, the coil acting as primary, secondary and tickler. Capacity *feed-back* (q.v.) is accomplished by means of a variable condenser. In the improved Reinartz a variocoupler is used, connected as a variometer, in addition to the condenser feed-back.

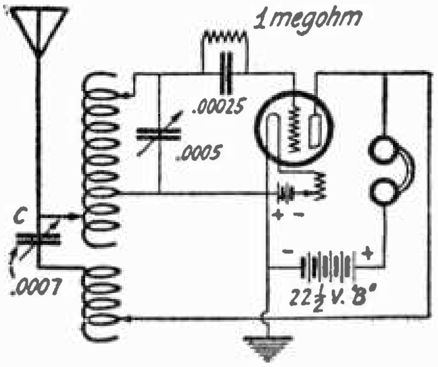

Diagram showing Reinartz circuit

REJECTOR—See *Rejector Circuit.*

REJECTOR CIRCUIT—A circuit consisting of a combination of inductance and capacity for filtering out or preventing the passage of currents of certain frequencies. This circuit is used in connection with a radio receiving set to obtain greater selectivity. The inductance and the capacity are arranged in parallel to form a resonant circuit. The constants of this circuit are chosen so that at the particular frequencies to be filtered, the capacity current will exactly equal the inductive current, thus giving zero resultant current. The primary application of the rejector circuit is as an *interference eliminator* (q.v.) or *wave trap* (q.v.). Figure 1 shows the application of the rejector circuit applied to a radio receiving set. The circuit shown is a combination *acceptor-rejector* circuit. The rejector circuit is shown in heavy lines. The inductance and the capacity are connected in parallel with each other and with the acceptor circuit and this latter is inductively coupled to the receiving set circuit. Both the acceptor and the rejector circuits are tuned to the wave length of the receiving aerial. By using more than one rejector and acceptor circuit increased selectivity can be obtained. A rejector circuit used to filter out undesired spark signals is shown in Figure 2. A small inductance of about ten turns is employed in this circuit and this is inductively coupled to a closed tuned circuit and wired in series with the aerial, and the aerial terminal of the set. If a spider-web coil is used in the closed circuit, a suitable value for broadcast wave lengths is a 50 turn coil, placing the aerial coil in such a position as to get closest coupling. The variable condenser should have a capacity of .0005 mfd. (See *Acceptor.*)

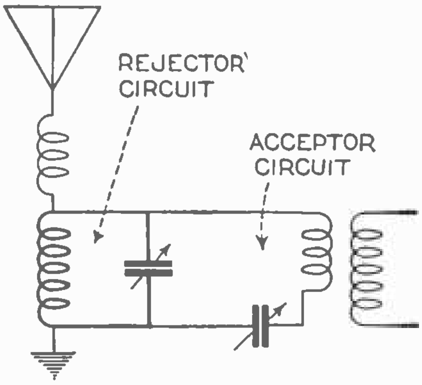

Fig. 1. An acceptor-rejector circuit used as a wave trap.

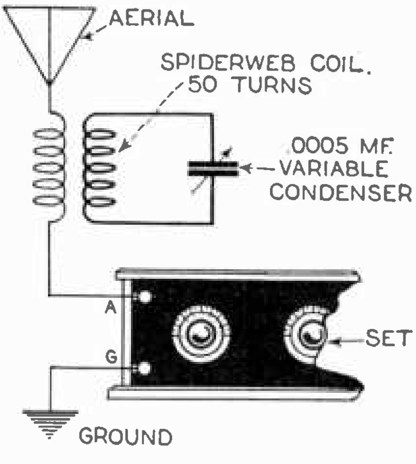

Fig. 2. A rejector circuit designed to filter out undesired spark signals.

RELAY—An electromagnetic device by means of which contacts in one circuit are operated by a change in conditions in the same circuit or in one or more associated circuits. (See *Magnifier.*)

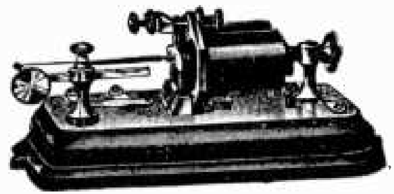

A telegraphic relay.

RELUCTANCE—Magnetic resistance. The reluctance of a magnetic circuit is analogous to electrical resistance. All magnetic circuits offer reluctance to flux. The name *rel* is widely used as the unit of reluctance. The rel is the reluctance of a magnetic circuit which will have one line of force in it, upon application of one *ampere-turn.* The rel can also be specified as the reluctance of a cylinder of air, or of any other non-magnetic material, one square inch in cross-sectional area, and 3.19 inches long. The flux through a magnetic circuit is equal to the *magnetomotive force* (q.v.) divided by the reluctance. (See *Magnetic Properties.*)

RELUCTIVITY—The *reluctance* (q.v.) per unit length and per unit cross-section. The reluctance per inch^3 or per cm^3. Reluctivity may be defined as specific magnetic resistance. It is the reciprocal of *permeability.* (q.v.).

REMANENCE—The term applied to the magnetism remaining in a magnet after the magnetizing force is removed.

REMOTE CONTROL—A control system whereby a radio transmitting or receiving apparatus may be operated at any desired distance from the apparatus. The chief advantage of remote control lies in the fact that the high tension apparatus and the tall aerial masts may be located in outlying districts whereas the studio can be located at any desired, convenient point. In addition it is possible to broadcast concerts, theatrical entertainments, prize fights, etc., direct from the scenes of action by utilizing remote control.

REPEATING COIL—A transformer of unity ratio used in telephone practice.

RE-RADIATION—The relaying of a radio wave through the action of a receiver employing regeneration. This phenomenon is most commonly observed in receivers of the *super-heterodyne* type.

RESIDUAL CHARGE—The residuary electricity retained by a condenser after discharge. This is also known as *electric residue* or *soakage.* After the condenser is discharged in the ordinary way, the residual charge permits a second discharge, smaller than the first. (See *Soaking in.*)

RESIDUAL MAGNETISM—The magnetism retained by a piece of iron, after the magnetizing force is withdrawn. After a mass of iron has become magnetized, some of the magnetism will remain, even though the magnetizing force has been taken away. The amount of residual magnetism depends on the quality of the iron. Pure wrought iron, in general, retains very little residual magnetism, whereas wrought iron which has gone through a hardening process, or which contains a large percentage of impurities and also cast iron possess a much larger amount of residual magnetism.

RESINOUS ELECTRICITY—An obsolete term for negative charges of electricity, adopted because of the fact that resinous bodies become negatively charged by friction. (See *Vitreous.*)

RESISTANCE—Symbol Ω (Omega)—Every conductor of electricity offers opposition to the flow of electricity. Resistance is a property of the conductor itself. It increases with increased length of conductor, decreases with increased cross-sectional area, and is greater or less depending upon the material of the conductor. Resistance is measured in *ohms.* (q.v.). The ohm is equal to 10^9 C.G.S. electromagnetic units.

RESISTANCE BOX—A box containing a number of wire coils of known resistances connected in various combinations, by means of switches or plugs, to produce any desired total resistance.

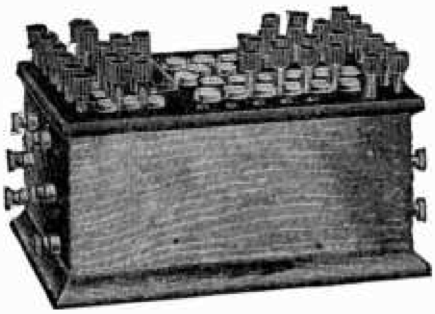

A resistance box.

RESISTANCE COUPLED AMPLIFIER—A type of *amplifier* (q.v.) which utilizes high resistances for interstage coupling. Resistance coupled amplification is generally used only for audio frequency amplification although it may be used for radio frequencies above 1000 meters. The chief advantage of resistance coupling consists in the purity of tone obtainable and the absence of distortion. In addition radio receivers of this type are comparatively inexpensive to build. It should be noted, however, that with resistance coupled amplification, the amplification per tube is much less than with transformer coupled amplification and furthermore the drain on the "B" batteries is also much greater. Of course, where a "B" eliminator is used "B" battery consumption is not a consideration. The resistances used for resistance coupling are non-inductive and may have resistances of from .1 megohm to 2 megohms. Figure 1 shows an elementary circuit utilizing a non-inductive resistance in the plate circuit of the tube. When a signal is impressed upon the grid circuit of this amplifying tube, there is a corresponding voltage set up in the plate circuit. The resistance R in the plate circuit causes a voltage drop to take place in two places in the tube. The total resistance of the plate circuit is equal to the internal resistance of the tube added to the resistance R, and if the "B" battery has 90 volts and if R is equivalent to the tube resistance at that moment the voltage drop across the tube will be 45 volts with the same drop across R. Theoretically it is possible to vary the voltage drop across R from zero to almost 90 degrees. Discounting the small resistances of the "A" and "B" batteries, a changing voltage will occur across the filament and plate when the grid potential varies, and also across the resistance R. Naturally these last variations depend upon the tube resistance and upon the value of the resistance R. These variations will be largest when R has a greater resistance than that of the tube. The resistance or impedance of the tube is constantly in a state of change and can assume very high values. For this reason the value of R should be equal to the higher values of the tube impedance in order to maintain maximum effectiveness. As used with ordinary amplifier tubes, a value of 100,000 ohms has been found to give satisfactory results. The next step is to apply the voltage variations across R to the grid circuit of the next tube. In Figure 2, the resistance R lies in the grid circuit of tube 2 as well as in the plate circuit of tube 1. Hence the magnified changes in tube 1 are transferred from R to the grid and filament of tube 2, thus the latter tube will be affected directly by the output circuit of the first tube and thus there will be amplification in both tubes thereby causing a resultant magnified signal to be heard. However, in such a hook-up, the same "A" and "B" battery cannot be used since this would mean that there would be a positive charge on the grid of tube No. 2, which would then be unable to function. Naturally, it would be impracticable to use a circuit having so many batteries, so a means must be found whereby it is possible to have a power supply consisting of but one filament battery and one plate battery for any given number of tubes in an amplifier. In order to do this the grids of the tubes must be separated electrically from the high voltage of the plates of the preceding tubes, but only as far as the steady voltage is concerned, while the fluctuating voltage of these plates must be impressed on the grids of the following tubes. This is done by introducing condensers between the two resistances. These effectively block the direct currents while they afford easy passage to alternating currents. Figure 3 shows the introduction of the condenser between the plate resistance and the grid resistance. Inspection of this diagram shows that although tube No. 1 remains the same, the plate instead of being connected directly to the grid of tube No. 2 is separated from the latter by means of the condenser C. If voltage fluctuations now occur across the resistance R, they will also be effective across the grid and filament of tube No. 2, because R exerts its influence between plate and filament of tube No. 1. As the condenser, C, conducts the variations very easily, the voltage fluctuations against R are then effective across the filament and grid of tube No. 2. It is necessary that tube No. 2 be provided with a grid leak, R_1 for the accumulated negative charges to leak off the grid to the filament. This grid leak at the same time serves the purpose of a by-pass for voltages, occurring due to possible leaks in the condenser C, which would tend to make the grid positive, thus allowing a grid current to flow and stopping the perfect action of the tube as an amplifier. Since the grid leak R_1 is a direct shunt across the grid and filament of tube No. 2, it must have a relatively high value in order to retain the highest possible amplification. R_1 should have a value lying between 500,000 and 2,000,000 ohms. If this resistance is made variable, it will be possible to control the resulting amplification completely. The value of the condenser C, must be chosen so that its resistance to the flow of alternating current will be as low as possible. Figure 4 shows three stages of resistance coupled amplification with a "C" battery used to give proper grid bias. The value of the resistance R, in the plate circuit of the tubes is 100,000 ohms and the value of the resistance in the grid circuits is 1,000,000 ohms (1 megohm). The condenser C has in each case a capacity of 0.05 mfd. The value of the "C" battery is about 4½ volts, but this should be determined by experiment for best results.

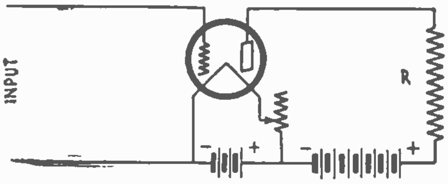

Fig. 1. The introduction of the non-inductive resistance in the plate circuit of the tube is the first step for explaining the principles involved.

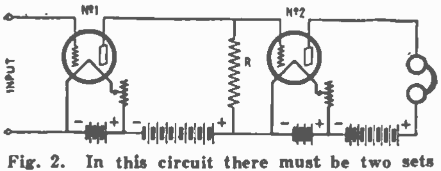

Fig. 2. In this circuit there must be two sets of batteries, but even then the grid of tube No. 2 will have a positive bias, which is undesirable.

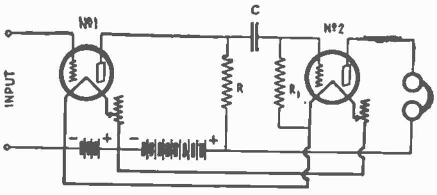

Fig. 3. The resistance coupled amplifier circuit with the introduction of a grid leak and condenser. The reasons for this are explained in the accompanying text.

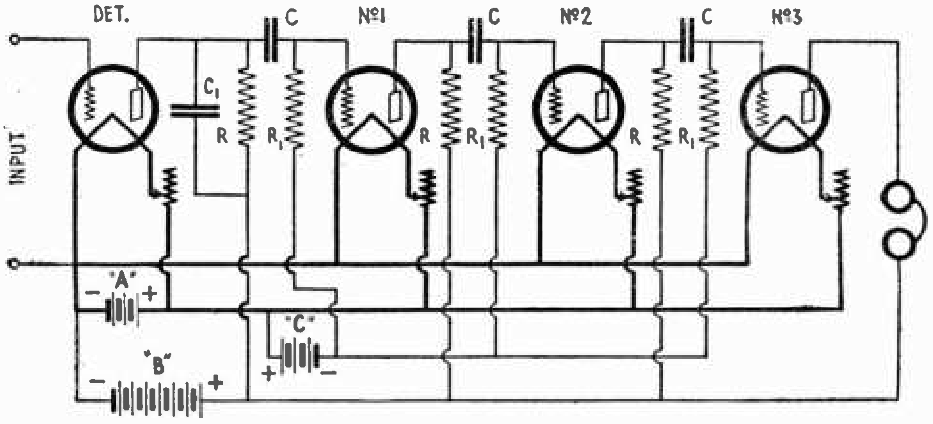

Fig. 4. Diagram showing three stages of resistance coupled amplification.

RESISTANCE COUPLING—RESISTIVE COUPLING—The transference of electrical energy from one circuit to another, by means of resistance common to both. (See *Resistance Coupled Amplifier.*)

RESISTANCE, HIGH FREQUENCY—The increased resistance of a conductor to high frequency currents, due to concentration of current at the surface of the conductor. This phenomenon is known as *skin effect* (q.v.) Skin effect becomes more pronounced, the greater the frequency of the impressed electromotive force. Skin effect also increases with increased cross-section of the conductor, with increased conductivity of the conductor and with increased magnetic *permeability* (q.v.). As a result of skin effect, the actual resistance of a conductor to alternating currents is greater than to direct currents.

RESISTANCE OF GROUND CONNECTION—The connecting wire leading from the radio apparatus to the ground should be of low resistance. If possible, the ground should be made by burying a copper plate in moist earth. The usual way to obtain a ground is to fasten a copper strap or *ground clamp* (q.v.) around a water or a steam pipe.

RESISTANCE (RADIATION)—This is the ratio of the total energy radiated (per second) by the antenna to the square of the root-mean-square current at a potential node (generally the ground connection).

RESISTANCE, RADIO FREQUENCY—The ratio of the heat produced per second, in watts, to the square of the root-mean-square current (radio frequency), in amperes, in a conductor.

RESISTIVITY—See *Resistivity of a Material*, also *Resistivity, Surface*.

RESISTIVITY OF A MATERIAL—SPECIFIC RESISTANCE—Symbol ρ (rho)—The resistance in ohms of unit length and unit cross-section of the material under consideration. Resistivity may be measured in ohms per inch cube or per centimeter cube, but more commonly it is measured in ohms per *mil-foot* (q.v.). The material of a conductor has an important bearing upon its resistance. Thus a unit length and unit cross-section of aluminum has about one and one-half times the resistance of copper having the same dimensions. Platinum has about six times the resistance.

RESISTIVITY, SURFACE—Referring to an insulating substance, surface resistivity is the resistance between two opposite edges of a surface film, 1 centimeter square. Surface resistivity depends to a considerable extent on humidity.

RESONANCE—A condition which exists in an electrical circuit where the frequency of the applied electromotive force is the same as the natural frequency of the circuit. Resonance is the cumulative effect produced by a periodic force in a circuit, so adjusted in frequency that the effect can attain the highest value. The effect measured is usually the current and the circuit when adjusted for maximum current, is said to be in resonance with the force, or with the circuit from which the force comes. The effect of resonance is to bring the current into phase with the electromotive force. Resonance in a circuit has the effect of bringing the reactance due to the inductance (inductive reactance) in exact opposition to the reactance due to capacity (capacity reactance). The total reactance is therefore zero and therefore the impedance is equal to the resistance. When two or more circuits have the same natural frequency they are said to be in resonance. Resonance of a circuit to a given exciting alternating electromotive force is that condition due to variation of the inductance or capacity in which the resultant effective current (or voltage) in that circuit is maximum. Instead of varying the inductance and the capacity, the frequency of the exciting field may be varied. The condition of resonance is determined by the frequency at which the current or voltage is maximum.

RESONANCE CURVE—A curve showing the relationship between the current induced in an oscillatory circuit and the frequency of the inducing current. In the case of two oscillating circuits placed near each other so that oscillations in one circuit will induce oscillations in the other, the induced oscillations will be weak or strong depending whether or not the second circuit is in tune with the first one. The capacity and the inductance of the second circuit can be varied until it is exactly in tune with the first at which time maximum current will flow in it. The resonance curve is obtained by measuring the current in the second circuit for various noted oscillation constants, and plotting the curve with the ordinates representing the current and the abscissae the corresponding natural frequency of the second circuit. The diagram shows a typical resonance curve. The peak of the curve is the point where the two circuits are in tune, or in other words, the point of resonance.

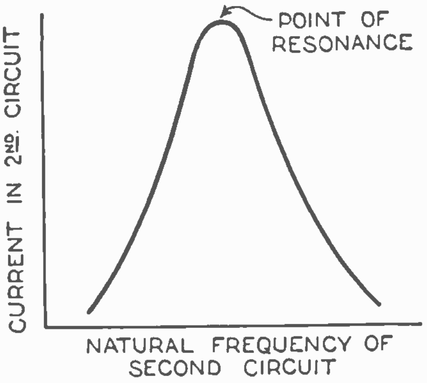

A typical resonance curve.

RESONANCE INTENSIFIER—An automatic receiver used for the reception of high-speed radio telegraph signals.

RESONANCE TRANSFORMER—Any loose-coupled tuning inductance having a primary and secondary each with a variable condenser in the circuit. Tuning the secondary circuit brings it in resonance with the primary, thus enabling signals to be heard with greatest volume.

RESONANT—See *Resonance*.

RESONATOR—A device for detecting by resonance, oscillations produced by an oscillator. This term is also used to refer to the sound box employed with the telegraph sounder.

RESONATOR, ACOUSTIC—A vessel containing a volume of air which can easily be set into vibration by sound waves or by the oscillation of a solid object. Typical forms are cylindrical or spherical and the ear is applied to a small tube opening into the air chamber.

RETARDATION COIL—A reactance coil used in a circuit for the purpose of selectively reacting on currents which vary at different rates.

RHEOSTAT—A variable resistance used to control current flow. Rheostats may be made of resistance wire or of carbon or for special purposes liquids may be used. In radio work, rheostats are connected in series with the filaments of vacuum tubes and the "A" battery in order to control the amount of current flowing in the filaments.

A rheostat constructed of resistance wire.

R. M. S.—Abbreviation for *Root Mean Square* (q.v.)

ROBERTS, Joseph Harrison Thomson—Born in 1890, he was educated at Liverpool and Cambridge Universities. He carried out many physical researches at the Cavendish Laboratory, Cambridge, and did a great deal of work on thermionics and other radio problems. He is a doctor of science and a Fellow of the Institute of Physics. He has several radio patents to his credit and is an expert in electrical, optical and acoustical sciences.

ROBINSON, James—British radio inventor. Born in 1884, he was educated at Durham and Gottingen Universities. He was appointed lecturer in physics at Durham University, 1906-07, lecturer in physics Sheffield University, 1910-12, and lecturer in physics at the East London College, London University, 1912-15. Robinson is known chiefly for his direction finding apparatus. He is a member of the Radio Society of Great Britain.

ROBINSON, Samuel S.—Rear-Admiral, U.S.N. He was born May 10th, 1867, and was educated at the United States Naval Academy, from which he graduated. He wrote the Manual of Wireless Telegraphy for Naval Electricians, and was in charge of the Bureau of Equipment, Navy Department.

ROOT MEAN SQUARE—The square root of the mean value of the square of an alternating current or voltage. An alternating current whose root-mean-square value is one ampere has the same heating effect as a direct current of one ampere. Root mean square is also referred to as *effective* or *virtual value*. The root mean square is derived as follows. In electrical measurements, it is necessary to compare direct and alternating currents. Moreover, the same units of measurement, the *ampere* (q.v.) and the *volt* (q.v.) are used in referring to both direct and alternating currents. However, in a direct current, a certain voltage will produce a steady current, whereas in an alternating current, both the voltage and the current are constantly varying both in magnitude and direction. The common basis of measurement is found in the value of the alternating current which will produce the same amount of heat as a certain direct current. Heating effect, according to *Joules Law* (q.v.) is directly proportional to the square of the current. Hence, in an alternating current varying according to a sine law, it is necessary to square each ordinate, find the mean or average value of these squares, and then extract the square root of this mean. Stated in concise terms, it is necessary to find the root mean square. An alternating current having a *maximum* value of one ampere would have a root mean square value of .707 amp. (See *Form Factor.*)

ROOT-MEAN-SQUARE VALUE—See *Root-Mean-Square.*

ROTARY CONDENSER—A synchronous motor used on a power line to advance the *phase* (q.v.) of the current. As a result there is an improvement in the *power factor* (q.v.). The action of the synchronous motor is similar to that of a condenser, whence the name rotary condenser.

ROTARY CONVERTER—A device for converting alternating current into direct current or vice versa. The rotary converter is also known as a *synchronous converter.* It is built like a direct current generator, except that it has certain commutator segments or conductors in the armature, connected to slip rings. Two, three, four, or six rings may be used. When the armature is rotated there is an alternating voltage between the rings. If the converter is driven by a motor or by an engine, it may be used as a *double current generator,* direct current being obtained from the brushes and alternating current from the slip rings (collector rings). When the collector rings are supplied with alternating current, the converter runs as a synchronous motor and direct current can be obtained from the brushes on the commutator. It should be understood that there is only one set of windings on the rotary converter which act simultaneously as alternating current motor and direct current generator. A converter which is used to convert direct current to alternating current is called an *inverted converter.* In the *single-phase* converter, there are two slip rings connected to the windings by as many equally spaced taps as there are poles. In the three-phase converter, three rings are used, with three equally spaced taps for each pair of poles. The single phase rotary converter must be brought up to synchronous speed by external power as it will not start itself. The polyphase converter will start itself from the alternating current side, but since it draws a very great lagging current, it should be started whenever possible from the direct current side. (See *Converter.*)

ROTARY CURRENT—A term infrequently used to denote *polyphase* (q.v.) current.

ROTARY DISCHARGER—A rotating disc discharger used in wireless transmitters in which the sparks occur between metal studs mounted on a motor driven disc. (See *Disc Discharger.*)

ROTARY FIELD—A magnetic field rotating in space. It may be due to rotating windings or stationary windings properly spaced through which polyphase currents are flowing.

ROTARY GAP—A form of rotating spark gap utilizing two or more rotating discs carrying studs between which the sparks pass. Spark gaps of this type were formerly used in wireless telegraph transmitters. (See *Discharger,* also *Disc Discharger.*)

ROTARY HYSTERESIS—Hysteresis (q.v.) produced by a rotating field as distinguished from that produced by a changing field.

ROTARY SPARK GAP—See *Rotary Discharger.*

ROTARY TICKER—A *ticker* (q.v.) in which the interruptions are produced by the chattering of a springy brush on a rapidly rotating metal wheel.

ROTOR—The rotating part of any electrical apparatus. It may refer to the revolving portion of a *motor* (q.v.) or *generator* (q.v.) or to the rotable part of a variable *condenser* (q.v.). The rotor in a direct current motor or generator is the armature. In an alternator or an alternating current motor, the rotor may be either the armature or the field.

ROTOR, GROUNDED CONNECTION—This refers to the grounding of the rotors of variable condensers in a radio receiving set. This is often done to eliminate hand capacity. Sometimes the rotors are connected to the shielding, which is grounded.

ROUGH CALIBRATION—Meters such as voltmeters, ammeters, etc., are often calibrated by comparison with standard instruments. In case such instruments are not available, it may be necessary to use less accurate methods. As an example of rough calibration, an experimental wave meter could be calibrated roughly by tuning in on various broadcasting stations whose wave lengths are known. (See *Calibration.*)

RUHMKORFF COIL—See *Induction Coil.*

RUTHERFORD, Sir Ernest—British physicist. Born at Nelson, New Zealand, he was educated at Canterbury College, Christchurch, New Zealand University, and Trinity College, Cambridge. From 1898 to 1907 he held the position of MacDonald professor of physics at Manchester University. In the latter year he was appointed Cavendish professor of physics and director of the Cavendish laboratory at Cambridge. Rutherford is an extremely brilliant physicist and has carried out a series of researches on the ultimate composition of matter. Many of these experiments have confirmed the researches of J. J. Thomson and others. Some of Sir Rutherford's experiments have revolutionized the theories of matter. He has received many honors for his brilliant researches including the Nobel prize for chemistry awarded to him in 1908. He has written extensively on radio activity, his books including "Radioactivity," 1904; "Radioactive Transformations," 1906; "Radioactive Substances and their Radiations," 1913; as well as many papers for technical journals. Sir Ernest was knighted in 1914.

S

SAFETY GAP—A spark gap which is connected in parallel with a condenser or other apparatus, and which protects the apparatus from excessive voltage. When the voltage exceeds a predetermined limit, the gap breaks down, thus passing the current around the apparatus. (See *Gap, Micrometer.*)

SAL AMMONIAC—A term commonly used for ammonium chloride. This substance is a white crystal, very soluble in water, and is used as an electrolyte in primary cells and also in many electroplating processes. (See *Cell.*)

SATURATED—The magnetic condition of a substance when increase of magnetizing force produces no further increase in flux density. A point is reached in every magnetic material, where further increase of magnetic intensity will not produce any increase of flux density. At this point, the material is said to be saturated. Cast iron becomes saturated at a lower flux density than steel or wrought iron.

SATURATION—See *Saturated.*

SATURATION, MAGNETIC—See *Saturated.*

SECONDARY CELL—A unit consisting of a positive and a negative electrode immersed in an *electrolyte* (q.v.) which becomes active only after the passage of an electric current. The electrodes are of metal or of a metallic compound. An electric current, flowing into a secondary cell causes certain chemical actions to take place whereby electrical energy is transformed into chemical energy. This process is termed "charging." After the cell is charged, the reverse action takes place and chemical energy is transformed into electrical energy. In this case, the cell is able to deliver an electric current to an outside circuit. This process is termed "discharging." The secondary cell is also known as an *accumulator* or *storage cell.* However, the cell does not store electricity. It merely holds or stores the chemical energy for immediate transformation into electrical energy. It is usual practice for several positive plates to be connected in multiple to form one positive electrode and for several negative plates to be connected in parallel to form a single negative electrode. The *storage battery* consists of two or more storage cells. The operation of the storage cell depends upon the fact that certain metals immersed in an electrolyte may differ in potential. In the secondary cell, no electrical action takes place until after charging. Thereafter, chemical action maintains a difference of potential until the stored energy has been retransformed. While the negative electrode is used up to a certain extent during the process of discharging, it is replaced electrochemically during the charging process, due to the flow of current through the electrolyte in the opposite direction to that of discharge. The two principal types of storage cells are the lead cell and the nickle-iron-alkaline (or Edison Cell). In lead cells, the negative electrode is made of spongy lead and the positive electrode is of lead peroxide. The electrolyte is dilute sulphuric acid. The following chemical action takes place within the lead cell. When the cell is charged, hydrogen is released at the negative plate, and oxygen at the positive plate. The oxygen combines with the lead of the positive plate to form lead peroxide while the hydrogen reduces the oxide in the negative plate. When the cell is discharged, both electrodes are changed to lead sulphate and the liberated hydrogen combines with the oxygen to form water. Care must be taken not to allow the cell to discharge too far as in this case, too great an amount of lead sulphate may form and this may cause buckling of the plates or may make it very difficult to recharge the cell. The indications of a completely charged cell are increased gassing of electrolyte together with a stop in the rising of the voltage and the specific gravity. In the Edison-type secondary cell, the active materials are of nickel and iron. Nickel hydrate forms the active material of the positive plate, while iron oxide forms the active material of the negative plate. The electrolyte is a 21 per cent. solution of potassium hydrate combined with a small amount of lithium hydrate. The chief advantage of the *Edison cell* (q.v.) is that it can be left idle indefinitely in a fully charged, semi-charged, or fully discharged state without any harm to the cell. In addition it is lighter and smaller than the same capacity lead cell. The unit used for measuring the capacity of storage cells is the *ampere-hour* (q.v.). This is based on the rate of discharge normally used in measurement, the eight-hour rate. As an example, an 80 ampere-hour storage battery will furnish a continuous current of 10 amperes for 8 hours. If the battery is discharged at a more rapid rate, it will not have the rated capacity.

SECONDARY CURRENT—The current flowing in the secondary windings of inductive apparatus. Current induced in a secondary winding due to the action of *primary current.* (See *Current, Primary.*)

SECONDARY ELECTRON EMISSION—See *Secondary Electrons,* also *Secondary Emission.*

SECONDARY ELECTRONS—Electrons emitted by the plate of a vacuum tube when struck by the electron stream from the filament. (See *Negative Resistance,* also *Secondary Emission.*)

SECONDARY EMISSION—Electron emission in which the exciting agency is bombardment of the emitting material by electrons. The emission of electrons from cold electrodes under the impact of other electrons. (See *Negative Resistance.*)

SELECTIVE—See *Selectivity.*

SELECTIVITY—The sharpness with which a radio receiving set can be tuned to a particular wavelength. A selective set is one which tunes sharply, whereas a non-selective set tunes broadly. Too long an aerial will often result in making a set tune broadly. A short aerial makes the receiving set more selective but reduces the volume.

SELECTOR—A tuning device connected in a radio receiving circuit for the purpose of eliminating undesired stations. (See *Wave Trap.*)

SELENIUM CELL—A cell constructed of a thin layer of the metal selenium. This metal is extremely sensitive to light. In its crystalline form, it is a conductor of electricity, its conductivity varying with the amount of light allowed to fall on it. As a result it is possible to regulate the flow of current in a circuit by varying the light. The selenium cell thus offers a means of transmitting pictures by wire or radio. Selenium cells are also being used in experimental devices for the transmission of motion pictures and for obtaining *television* (q.v.).

SELF-CAPACITY—The capacity of a circuit or a portion of a circuit. Every electrical circuit possesses inherent capacity, which varies in degree, depending upon the shape, size, etc., of the circuit. Self-capacity is also referred to as *distributed capacity* (q.v.). In general self-capacity is detrimental, especially in radio circuits, since it introduces losses. By using special forms of windings, the self-capacity of coils can be greatly reduced. (See *Low Loss Coils.*)

SELF-EXCITED GENERATOR—A generator which depends upon its own current for its field excitation. Such a generator must build up its magnetic field and hence its voltage, gradually. In order to get a start, it depends upon the *residual magnetism* (q.v.) present in the iron of the poles. There are three types of self-excited generators, the *shunt generator* (q.v.), the *series generator* (q.v.) and the *compound generator.* (See *Compound Wound.*)

SELF-HETERODYNE—A system of reception of continuous wave signals by the production of audio frequency beats, through the use of a device which is both a radio frequency generator and a detector of the audio frequency beat currents produced.

SELF-INDUCTANCES—See *Inductance, Self.*

SELF-INDUCTION—Inductance due to the field produced by the circuit itself. Changes in current are always accompanied by self-induction. When the current is decreasing, the electromotive force of self-induction is in the same direction, thus tending to prevent the decrease. When the current is increasing, the electromotive force of self-induction tends to prevent the increase. (*See Inductance, Self; Mutual Inductance,* also *Induction.*)

SELF-INDUCTION COEFFICIENT—The ratio of the number of lines of induction produced by a current flowing in the circuit divided by the current in the circuit. The coefficient of self-induction of a conductor of finite cross-section may be defined as the ratio of twice the energy of the magnetic field produced by the current flowing to the square of the current. The following factors affect the coefficient of self-induction of a circuit: the cross-section and shape of the conductor, the size and shape of the circuit, the permeabilities of the surrounding medium and of the conductor itself. Where the skin effect is large, the frequency of the current and the specific resistance of the conductor will also affect the coefficient of self-induction.

SELF-OSCILLATION—The tendency for the vacuum tubes in a radio receiving set to act as generators of radio frequency oscillations. (See *Body Capacity,* also *Bolitho Circuit.*)

SENSITIVE SPOTS—Points on the mineral of a crystal detector which show a greater response to incoming oscillations than others. Dust, grease, or dampness reduce the sensitivity of these spots and therefore the crystal should be properly protected.

SEPARATELY EXCITED GENERATOR—A generator in which current for producing the magnetic field is obtained from some source external to the generator itself. This source is either another generator or else a storage battery. Separately excited generators are generally used only for special work such as battery charging, electroplating and testing.

SERIES—See Series Connection.

SERIES CONNECTION—When two or more parts of an electrical circuit are so connected that the same current flows through them they are said to be in *series* and the circuit is called a *series circuit,* the connection being a *series connection.* In order to connect primary or other cells in series, it is necessary to connect the positive terminal of one to the negative terminal of another, etc. In this case, the total voltage of the battery equals the sum of the separate voltages.

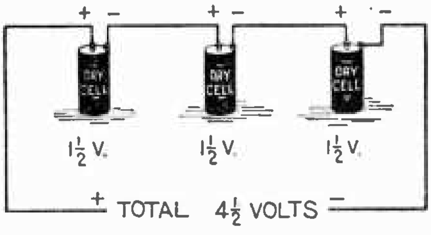

The three dry cells are here shown connected in series.

SERIES GENERATOR—A generator in which the exciting winding is connected in series with the armature and the load. This type of generator is of small commercial importance being used mainly for certain constant current systems.

SERIES MOTOR—A commutator-type motor having the field and armature windings connected in series. Ordinarily, the series motor is a direct current motor, but with minor modifications, such as lamination of the field as well as armature, this motor can be used for alternating current also. With the applied voltage constant, the greater the load, the less will be the speed of a series motor. As the load is lessened, the motor will speed up, and if the load is entirely removed, it will "run away." The *torque* (q.v.) characteristic is the most important feature of a series motor. Instead of varying directly with the current as in the shunt motor, the torque increases with the square of the current.

Hence, if the current is doubled, the torque will be increased four times, while if the current is quadrupled, the torque will be increased sixteen times. For the same current, the torque of the series motor is higher than that of the shunt and in addition, the series motor has a much better starting torque.

SERIES-PARALLEL—This refers to a circuit which is a combination of a *series circuit* and a *parallel circuit.* The term *series-parallel* is generally used to indicate a series circuit having certain minor branches in parallel; the term *parallel-series* denotes a parallel circuit having certain minor branches in series. The illustration shows a series parallel circuit having two dry cells in parallel and three dry cells in series. The push button and the bell are also in series. (See *Parallel Connection, Series Connection,* also *"A" Battery.*)

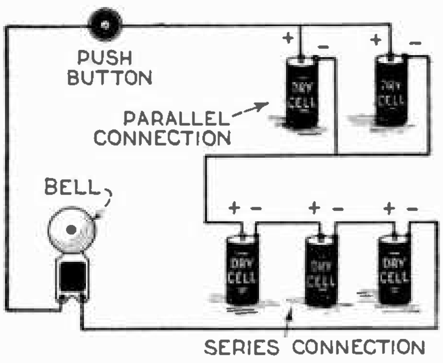

A series-parallel circuit. The three dry cells are connected in series with each other and with the bell and the push button. They are also in series with the other two dry cells, these latter being in parallel with each other.

SERIES RESONANCE—When a single lumped capacity and a single lumped inductance are connected in series between terminals to which an alternating electromotive force is applied, and the inductance or capacity or frequency is varied, the condition of series resonance exists, when the current is a maximum. (See *Resonance,* also *Resonance Curve.*)

SHARPNESS OF RESONANCE—A quantity expressing the fractional change of current in a simple series circuit for a given fractional change in either capacitative or inductive reactance.

SHARPNESS OF TUNING—See *Selectivity.*

SHARP WAVE—A wave in which the energy radiated is confined to a single frequency of oscillation. Tight coupling of the oscillation transformer tends to give a broad wave, whereas loose coupling gives a sharp wave. (See *Decrément.*)

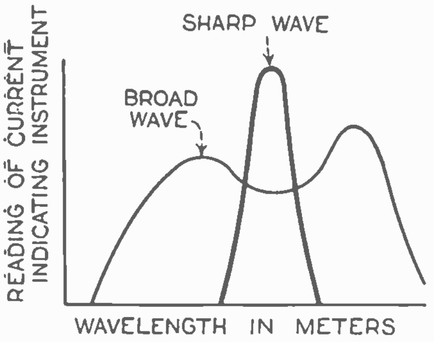

Graph showing difference between sharp and broad waves.

SHIELDED TRANSFORMER—An audio or a radio frequency transformer surrounded by a metal casing. Shielding of radio frequency transformers prevents interference from external fields and also keeps body capacity from affecting the tuning. Audio frequency transformers are shielded mainly to prevent interaction between the flux from the transformer and other parts of the radio set.

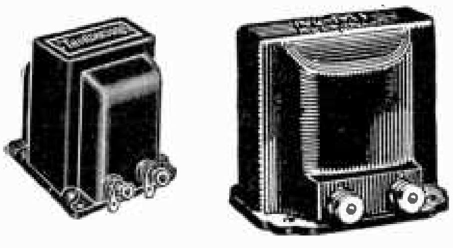

Typical shielded transformers.

SHIELDING—Metallic screening placed around radio apparatus to prevent interference from stray electromagnetic or electrostatic fields. In radio sets, the inside surface of the panel is often shielded by means of a metallic lining, thus eliminating body capacity. It is usual to ground the shielding. In sets having more than two stages of radio frequency amplification complete shielding is essential. (*See Induction Screen.*)

S.H.M.—Abbreviation for *Simple Harmonic Motion* (q.v.).

SHOCK EXCITATION—See *Impact Excitation.*

SHORT—See *Short Circuit.*

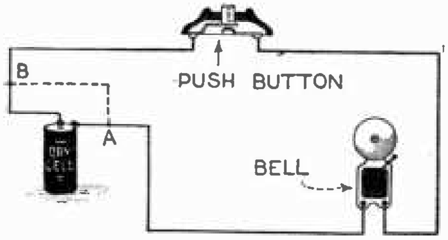

The dotted line between A and B represents a conductor short-circuiting the dry cell.

SHORT CIRCUIT—A short circuit refers to a condition of an electrical circuit whereby an external path of low resistance is connected across the source of electrical energy. A short circuit is accompanied by an increase in current flow, although this current is no longer usefully employed. Apparatus may be protected from short circuits by means of *circuit breakers* (q.v.) or *fuses* (q.v.).

SHORTENING CONDENSER, or ANTENNA TUNING CONDENSER—A condenser connected in series with the antenna and the ground, for the purpose of diminishing the natural wave length.

SHUNT WOUND GENERATOR—A generator in which the field winding is shunted across the armature terminals. The resistance of the shunt winding is made high enough so that only a small percentage of the load current flows through it.

SHUNT WOUND MOTOR—A motor in which the field winding is connected in parallel with the brushes (and the armature winding). The speed of a shunt motor is practically constant, but falls to a slight extent as the load is increased. The drop in speed may be only a few per cent between no load and full load, for large motors, while for small ones it may be as high as 15 per cent. As the load on a shunt motor is increased, the torque which the motor develops will increase and the armature current will also increase in almost direct proportion.

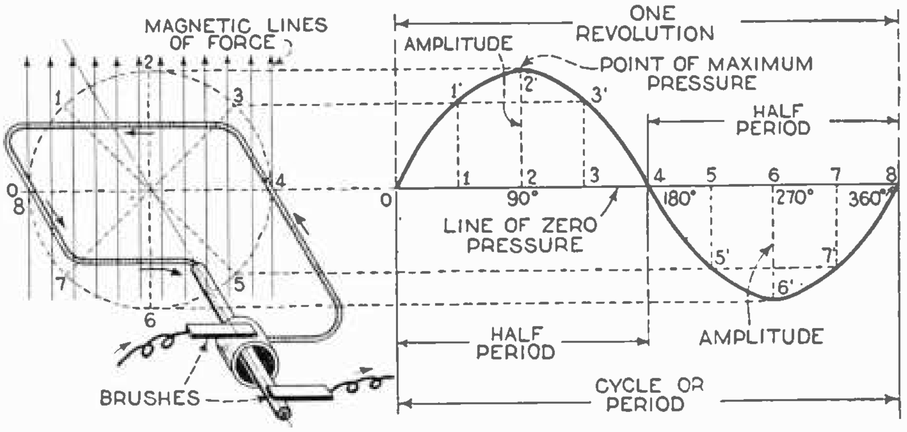

A sine curve is shown at the right. At the left a coil is shown rotating in a magnetic field. As the coil cuts the magnetic lines it generates the varying electromotive force represented by the sine curve.

S.I.C.—Abbreviation for *Specific Inductive Capacity* (q.v.).

SIDE BANDS—A group or band of frequencies formed by the interaction of the *Carrier Wave* (q.v.) and the modulations above and below the main carrier wave frequency. (See *Harmonics.*)

SIGNAL, STRAY RADIO—See *Strays.*

SILICON—A metalloid occurring widely in nature. It is a component of most of the rocks forming the earth's crust. Silicon crystals are used as detectors in radio work. (See *Crystal*, also *Silicon Detector.*)

SILICON DETECTOR—A *crystal detector* (q.v.) using silicon as its crystal. The silicon is used in a crystalline state and a *cat-whisker* (q.v.) is generally used to locate the *sensitive spots* (q.v.) and make contact.

SIMPLE HARMONIC—Referring to a *simple harmonic motion* (q.v.). If a point moves at a uniform rate around a circle and the point be projected in a straight line in the plane of the circle, the back and forth motion of the projected point is referred to as *simple harmonic motion.* True simple harmonic motions are the motions of a pendulum, a vibrating tuning fork, particles of water in a wave, etc. Other motions which are not precisely simple harmonic motions may be treated as resulting from several such motions. Examples of these are motions occurring in alternating currents in light waves, in electric waves and in sound waves.

SIMPLE HARMONIC MOTION—Abbreviation S.H.M.—The motion executed by the foot of a perpendicular let fall on the diameter of a circle from a particle moving with uniform velocity in that circle. The piston rod of a steam engine, turning a crank uniformly approximates a simple harmonic motion.

SIMPLE HARMONIC VIBRATIONS—The vibrations of a body such as a tuning fork which follow the law of *simple harmonic motion* (q.v.). (See *Angular Velocity*, also *Simple Harmonic.*)

SINE CHARACTERISTIC OF ALTERNATING CURRENT—See *Sine Curve*, also *Sine Wave.*

SINE CURVE—A curve whose perpendicular at any point is proportional to the sine of the angle corresponding to that point. The sine curve is used to represent the changes in direction and strength of an alternating current or voltage. The coil illustrated at the left in the accompanying diagram generates a sine wave current as it revolves through the magnetic field shown by the arrows. When the coil is moving parallel to the magnetic flux, no electromotive force is induced and hence current flow is zero. As the coil is rotated in a clockwise direction, it cuts the flux at a greater and greater angle, until the maximum is reached when the flux is cut at a right angle. This gives a gradually increasing electromotive force (and hence current) until maximum is reached, then when the angle at which flux is cut decreases, the electromotive force decreases until the coil again moves parallel to the flux, at which point electromotive force is again zero. Then when the coil starts to cut the flux in the opposite direction, electromotive force is reversed, increasing to a maximum and then decreasing to zero thus completing the *cycle* (q.v.). Current or electromotive force in one direction (positive) is represented by the height of the ordinates of the sine curve above the line, while current in the reverse direction is represented by the height of the ordinates below the line. The horizontal distances (abscissae) represent time generally measured in phase angles (position of coil at any particular instant). (See *Alternating Current*, also *Curve of Sines.*)

SINE WAVE—As referring to an alternating current, a sine wave or a simple alternating current, is a current whose wave shape is sinusoidal. The wave shape is the shape of the curve obtained when the instantaneous values of the current are plotted against time in rectangular co-ordinates. The wave shape is independent of the frequency and of the scale to which the curve is plotted.

SINE WAVE OF ALTERNATING CURRENT OR E.M.F.—See *Sine Wave.*

SINGING ARC—See *Duddell Singing Arc.*

SINGING SPARK—A spark used in a wireless telegraph transmitter which produces a singing note in the receiving head set. A spark which occurs at regular intervals and at an audible frequency such as that obtained in a quenched spark system of wireless telegraphy.

SINGLE CIRCUIT TUNER—A regenerative receiving circuit in which the antenna and grid circuits are conductively coupled. *Feedback* (q.v.) is obtained by placing the plate circuit in inductive relation with the primary.

SINGLE-PHASE—This term is used to refer to an alternating current, an alternating current system of distribution, or to alternating current apparatus. A single phase system or circuit is one energized by a single alternating electromotive force. A single phase circuit is usually supplied through two wires. The currents in these two wires counted outwardly from the source, differ in phase by one-half a cycle (180 degrees). In the single phase generator the voltage per phase is the same as the voltage between the lines and the current per phase is the same as the current per line. Since there is a great saving of copper in the use of three phase transmission of power, three phase generators are more frequently used than single phase machines. (See *Alternator.*)

SINOIDAL VIBRATION—See *Simple Harmonic Motion.*

SIPHON RECORDER—A moving coil galvanometer originally used for recording messages sent over long submarine cables. The received currents are passed to a coil suspended in a strong magnetic field. This coil carries a fine glass syphon which discharges ink onto a moving paper tape. In cable signaling, dots and dashes are usually indicated by opposite deflections which result in a wavy line being produced on the tape.

SKINDERVIKEN TRANSMITTER BUTTON—A small carbon grain microphone (q.v.) button less than an inch in diameter and about half an inch high containing a polished metallic button affixed to a mica diaphragm and a surrounding case of brass, the space between the case and the polished button being partially filled with a good grade of carbon granules.

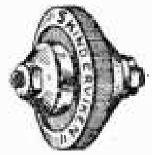

Fig. 1. **Skinderviken Transmitter Button.**

The button is very sensitive to sound waves when it is attached to any form of a diaphragm, thus making it applicable wherever a sensitive telephone transmitter should be used.

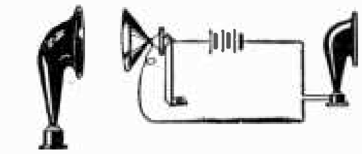

Fig. 2. **Retransmitting radio programs.**

Fig. 1 shows the external appearance, full size, of the Skinderviken button. This transmitter button has a number of applications in radio work and can be used in connection with speech amplifiers, telephone transmitters, detectaphones, etc.

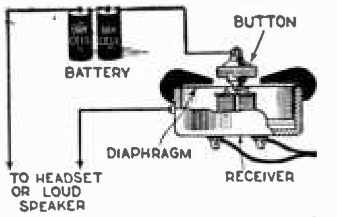

Fig. 3. **Retransmission from ear phone.**

It can be made to transmit in any position, either horizontal or vertical. It is small enough to be readily concealed. Fig. 2 shows how radio music can be re-transmitted to distant points and incidentally amplified at the same time. The microphone button is shown fastened to a cone and mounted in front of the loud speaker.

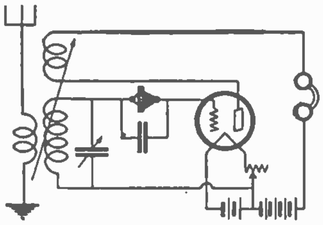

Fig. 4. **Button used as grid leak.**

The leads from the button are connected in series with several dry cells and another loud speaker at a distant point. Fig. 3 shows how the transmitter button can be attached to an ordinary receiver to produce the same effect.

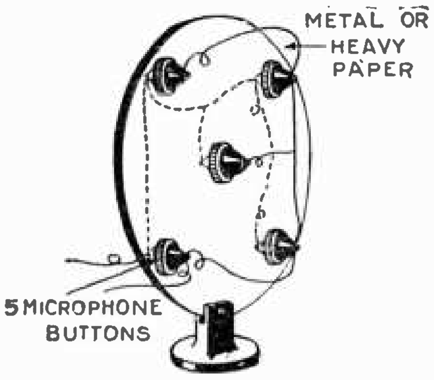

Fig. 5. **Combining buttons to form sensitive microphone.**

In this case also, the leads are connected in series with several dry cells and with a low resistance phone at the distant end. In Fig. 4 the transmitter button is shown in use as a grid leak in a regenerative circuit. Where it is desired to get greater sensitivity and volume for either radio telephone transmitting or for ordinary phone transmission the system illustrated in Fig. 5 will be found highly suitable.

Fig. 6 shows the construction of a hand microphone for transmitting sets or for inter-departmental telephone circuits.

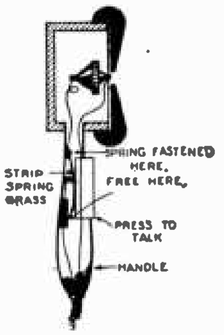

Fig. 6. **Construction of hand microphone.**

A very simple circuit closer is mounted in the handle of the microphone which in this particular case is simply an ordinary piece of wood hollowed out to receive the switch and the wires. This wooden handle is then fastened to the rubber case of a telephone receiver. The coils and magnets in the telephone receiver are removed, a small hole is drilled or punched in the diaphragm, and the transmitter button is fastened as indicated in the diagram.

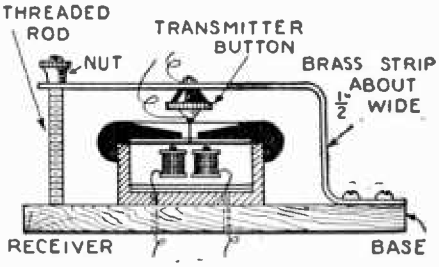

Fig. 7. **Method of fastening a button to ear 'phone.**

Another method of fastening a transmitter button to a telephone receiver is illustrated in the diagram in Fig. 7. Here the transmitter button is mounted on a brass strip which is bent so that the center of the button will rest upon the diaphragm of the telephone receiver fixed to the base. A threaded rod and nut regulate the pressure of the button against the diaphragm. In the event that it is not desired to make any changes to a pair of telephone receivers, the transmitter button can be fastened in the shell of an old receiver, B in Fig. 8, and the head phone may be rested on the shell.

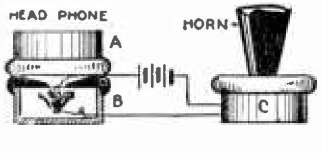

Fig. 8. **A simple form of amplifier.**

For

use with a Baldwin receiver, a microphone transmitter button is fastened to it in the following manner: The diaphragm of the Baldwin receiver is first removed, and a thin rod is soldered to the lever, as indicated in Fig. 9. The transmitter button is then soldered to the rod as shown. Another way of making a loud speaker set from an ordinary one-tube receiving set is illustrated in Fig. 10. Where a phonograph is available, the hook-up illustrated may be employed, placing the reproducer of the phonograph on the receiver, which should be placed in a suitable position on the turntable. Inasmuch as the movement of the diaphragm of the seventy-five ohm receiver is an up and down one, the reproducer of the phonograph should be turned as it would be for "hill and dale" records. A tiny drop of solder should be dropped upon the diaphragm of the seventy-five ohm receiver and a small nick placed in the center of the solder for receiving the needle of the phonograph reproducer.

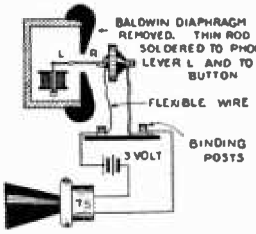
BALDWIN DIAPHRAGM REMOVED. THIN ROD R SOLDERED TO PHONE LEVER L AND TO BUTTON
L
R
FLEXIBLE WIRE
3 VOLT
BINDING POSTS
75

Fig. 9. Button attached to Baldwin 'phone.

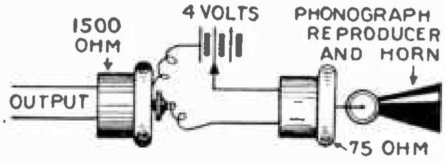
1500 OHM
4 VOLTS
PHONOGRAPH REPRODUCER AND HORN
OUTPUT
75 OHM

Fig. 10. Loud speaker operation from 1-tube set.

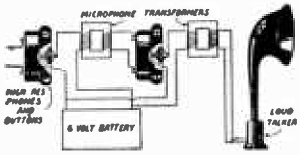
MICROPHONE TRANSFORMERS
HIGH RES PHONES AND BUTTONS
6 VOLT BATTERY
LOUD TALKER

Fig. 11. An effective radio amplifier.

Very often it is desired to further amplify the music or sounds picked up by a transmitter button. For this purpose, microphone transformers, or as they are better known, modulation or telephone induction transformers, are employed. In Fig. 11 this instrument is illustrated, together with the circuit for its use. Here the incoming radio signals are made to operate the microphone fastened to the receiver. The current which the microphone passes then goes through the primary of the microphone transformer, the secondary of which is connected to another receiver similarly fitted with another transmitter button. The output of the second microphone transformer is then connected to a loud speaker. A diagram for using but one stage of amplification through transformers is illustrated in Fig. 12. For the better class of work, however, it is desirable to use a standard vacuum tube amplifier as illustrated in Fig. 13. Here the microphone button is placed in series with the primary of a modulation transformer, and in series with several dry cells or flash light batteries. The secondary of the modulation transformer then connects to the grid and filament, the same as it would in a standard vacuum tube amplifying circuit. The music from the output is, of course, very loud.

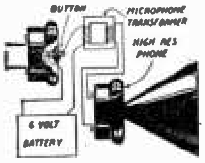
BUTTON
MICROPHONE TRANSFORMER
HIGH RES PHONE
6 VOLT BATTERY

Fig. 12. A single stage amplifier.

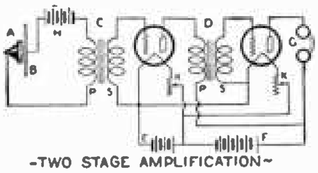
A
H
C
D
G
B
P S
P S
K
E
F
-TWO STAGE AMPLIFICATION-

Fig. 13. A 2-stage vacuum tube amplifier.

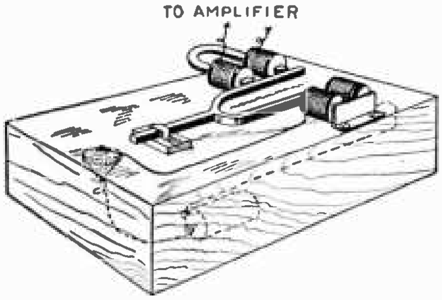
TO AMPLIFIER

Fig. 14. A tuning fork oscillator.

A very splendid oscillator is illustrated in Fig. 14. Here a tuning fork of standard frequency is mounted on a suitable holder in a cigar box. Two low resistance coils, about five to seventy-five ohms, are then placed under one prong of the tuning fork. These are connected in series with the transmitter button fastened in the cigar box, which serves as a base for mounting the tuning fork. The batteries may be concealed in the cigar box. Thus when the tuning fork is struck, it will continue to vibrate as long as the current is applied to the microphone. To use this as an oscillator, a pair of magnets from a telephone receiver and the permanent magnets are mounted near the other vibrating prong. The current constantly being changed by means of the vibrating bar of steel produces a standard frequency, which when amplified can be used for making radio measurements of different kinds. Fig. 15 shows the transmitter button connected in a radio circuit for use as a detector of radio waves. Fig. 16 shows a method of making a code practice device. An audio frequency howl is produced which will enable an instructor to teach an entire classroom the radio code. Sometimes the key, instead of interrupting the circuit, merely shunts a small portion of the current passing through the transmitter button, and for this purpose the key would have to be placed directly across the transmitter button and in series with a small resistance. Fig. 17 shows how the transmitter button may be used in Heising modulation. Fig. 18 indicates its adaptation to the variation of grid voltage of the oscillator tube and Fig. 19 shows how the transmitter button may be hooked up to modulate the output current. These last three figures indicate the use of transmitter buttons in vacuum tube transmitters.

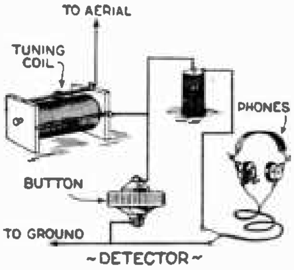
TO AERIAL
TUNING COIL
PHONES
BUTTON
TO GROUND
~DETECTOR~

Fig. 15. Transmitter button used as radio detector.

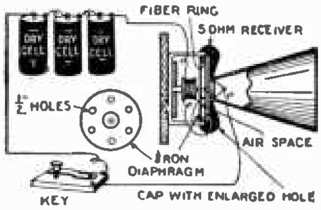
FIBER RING
5 OHM RECEIVER
DRY CELL
DRY CELL
DRY CELL
1/2" HOLES
IRON DIAPHRAGM
AIR SPACE
KEY
CAP WITH ENLARGED HOLE

Fig. 16. A code practice machine.

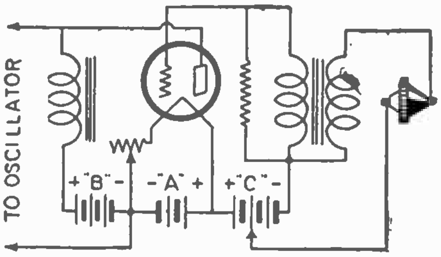
TO OSCILLATOR
+"B"-
-"A"+
+"C"-

Fig. 17. Use of button in Heising modulation system.

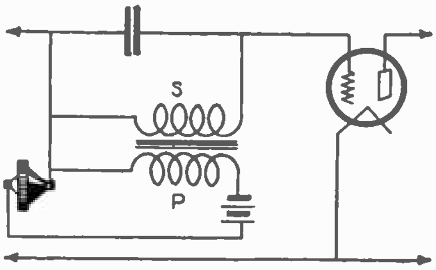

Fig. 18. Method of varying grid voltage of oscillator tube.

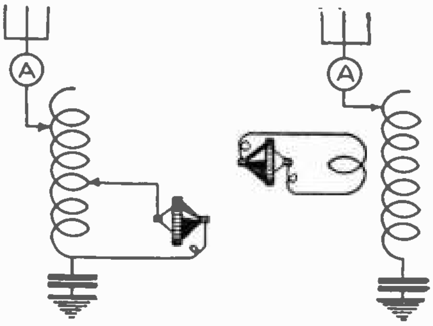

Fig. 19. Method of modulating output current.

SKIN EFFECT — The name given to the crowding of alternating or oscillatory currents into the surface layers of a solid conductor. This phenomenon increases with increased frequency. As a result the resistance of solid conductors is much higher at high than at low frequencies. (See *High Frequency*, also *High Frequency Resistance.*)

SLABY-ARCO-SYSTEM — A system of wireless telegraphy which used a quenched spark gap. This system is now obsolete.

SLABY-ARCO-VACUUM COHERER — A filings coherer enclosed in a sealed glass tube from which the air has been extracted.

A Slaby-Arco vacuum coherer.

SLIP RINGS — Metallic rings, separated by insulators, from which collecting brushes pick up current generated by an alternator. Slip rings take the place of the commutator used on the direct current generator.

SLOTS—Channels cut in the armature core discs and into which the windings fit.

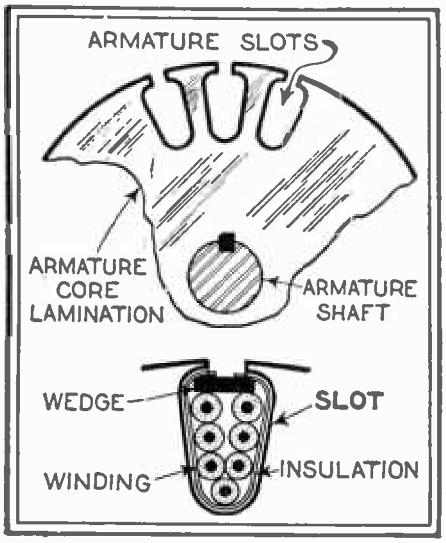

Details of armature slot.

SLUDGING OF TRANSFORMER OIL —The thickening of the oil used for cooling in large transformers. A muddy deposit is formed due to oxidation. The presence of metallic copper in contact with the oil is considered to hasten this action catalytically.

SOAKING-IN—An increase in a condenser charge above its initial value, which takes place gradually if the potential is kept applied. Soaking-in takes place due to a change in the *dielectric* (q.v.).

SOAKING-OUT — A gradual continuation of the discharge of a condenser after the first rapid discharge. In this case, the dielectric is returning to its original state.

SOCKET—A receptacle for holding and making electrical connections with a vacuum tube, an incandescent lamp, or a connection plug. Vacuum tube sockets are made of bakelite, hard rubber, porcelain, pyrex and other suitable *dielectrics* (q.v.). They are usually molded. One type of socket is composed of a shell of insulating material, or in some cases of metal. Four contact blades extend within the base of the socket, for making contact with the prongs of the tube. These blades may be of strong phosphor bronze. Their outer ends are attached to binding posts. It is important that the base supporting the binding posts and contacts be of good insulating material such as bakelite or hard rubber. Modern sockets are of the *spring* or *shock absorbing* type. This construction protects the tube and does away with microphonic sounds. One type of spring socket consists of two molded bakelite parts, the outer one square with a binding post at each corner and the inner one circular with holes for the tube prongs and steel spring contacts fastened underneath the holes.

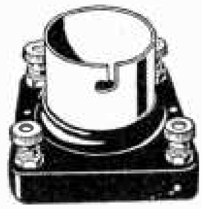

A metal shell socket.

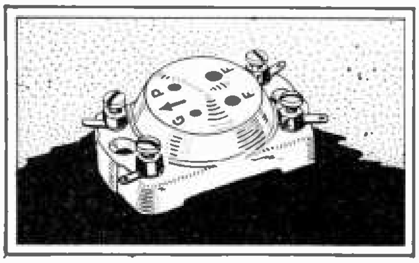

Shock absorbing or spring-type socket.

The circular molding is suspended within the square one by means of two rubber pads. Other types of shock absorbing sockets use spring suspension instead of rubber.

SOFT IRON INSTRUMENTS — Ammeters, voltmeters and other measuring instruments which depend upon the magnetic force between a current carrying coil and a movable core of soft iron to produce the deflection of the needle or pointer. The movement of the iron may be controlled by a spring or by gravity. These instruments may be used for either direct or alternating current measurements. They are not as accurate as instruments of the dynamometer type.

SOFT TUBE — A vacuum tube which contains a gas content instead of being highly evacuated. Soft tubes were formerly recommended for use as detector tubes but at present have been superseded for this use by the ordinary amplifying tube.

SOLDER—An alloy of lead and tin used for making good electrical connections.

SOLDERED JOINT — A connection between two metallic surfaces, cemented together by means of solder. Where the joint is between two electrical conductors, it must be made both mechanically and electrically secure.

SOLDERING — The process of joining together two or more conductors or other metallic surfaces by means of solder, so as to give a good electrical or mechanical connection, or both. After the surfaces to be united are carefully cleaned, a flux is applied to dissolve the oxides which occur on the surface of the parts to be joined with solder.

SOLDERING FLUX—A substance used to dissolve the oxides on the surfaces of the metals to be soldered. When these oxides are dissolved, the flux enables the solder to enter the minute pores of the metal surface, effectually sealing it against the penetration of oxygen. Fluxes range in character from very strong acids to very mild acid bearing substances. For radio use, it is necessary to use a flux which is non-corrosive and which in its use will leave a residual matter that will have no tendency to collect moisture, dust, or other foreign material. Rosin has been found to be very suitable as a flux in radio soldering. (See *Flux, Soldering.*)

SOLENOID — A coil of wire of helical form used as an electromagnet, generally to attract a movable iron plunger, which is drawn within the solenoid when the current flows through the windings. The solenoid never has a fixed core. A coil wound on a cylinder so as to give uniform magnetic field within it when traversed by a current is often called a *straight solenoid.*

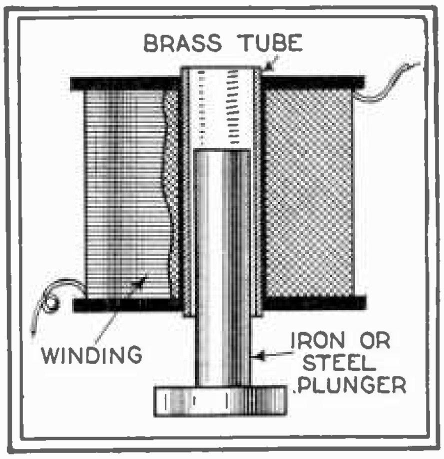

A typical solenoid.

SOLODYNE—A radio circuit which dispenses with high-tension or "B" batteries and which utilizes a double grid vacuum tube. The solodyne principle is also known as the *Unidyne.* The theory of the solodyne is as follows: The small plate current due to the electron emission from the lighted filament passes through the tickler coil which feeds back to the main grid circuit in the usual way, the primary of a step-up transformer, through the telephone receivers, and then back again to the filament of the tube. The electron stream passing from the filament to the plate inside the tube must pass the two grids. The first of these, which may be referred to as the additional grid, is primarily made positive by connecting it directly to the positive terminal of the "A" battery. This in itself tends to assist the electron stream, to reduce the resistance of the vacuum of the tube. But the additional grid is assisted in its work by having impressed upon it the stepped up voltage from the plate circuit, due to the transformer, the secondary of which is in direct connection with the additional grid. Thus a building up process is introduced, every possible electron primarily due to the electron emission of the filament of the tube being made use of and ultimately passing through the telephone receivers to be reproduced in the form of audible signals. The main grid functions in the usual manner, except that this too, can be made to help the additional grid as well, by giving it a strong positive bias. A typical Solodyne circuit is shown in Figure 1. In this circuit L and L^1 may be honeycomb coils or a variocoupler. In Figure 2 is shown the Cowper Solodyne circuit which utilizes standard tubes. In this diagram, the numbers printed near each coil indicate the size of honeycomb or duo-lateral coils to use in each circuit. (See *Four Element Tube.*)

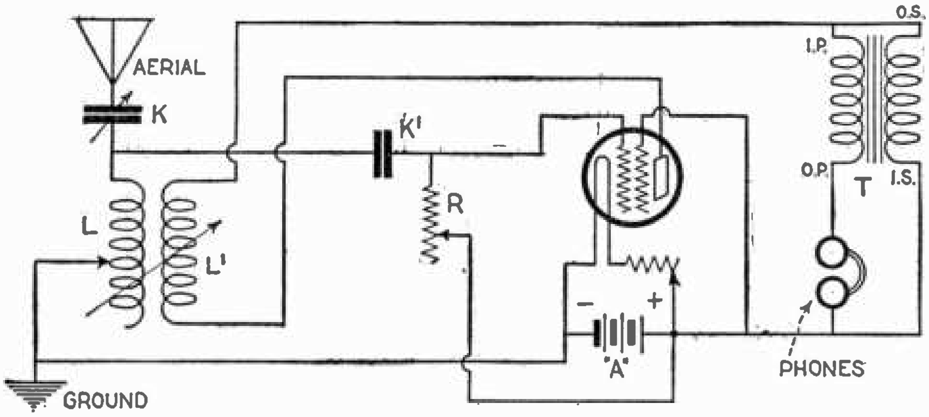

A Solodyne circuit. L and L^1 may be honeycomb coils or a vario-coupler.

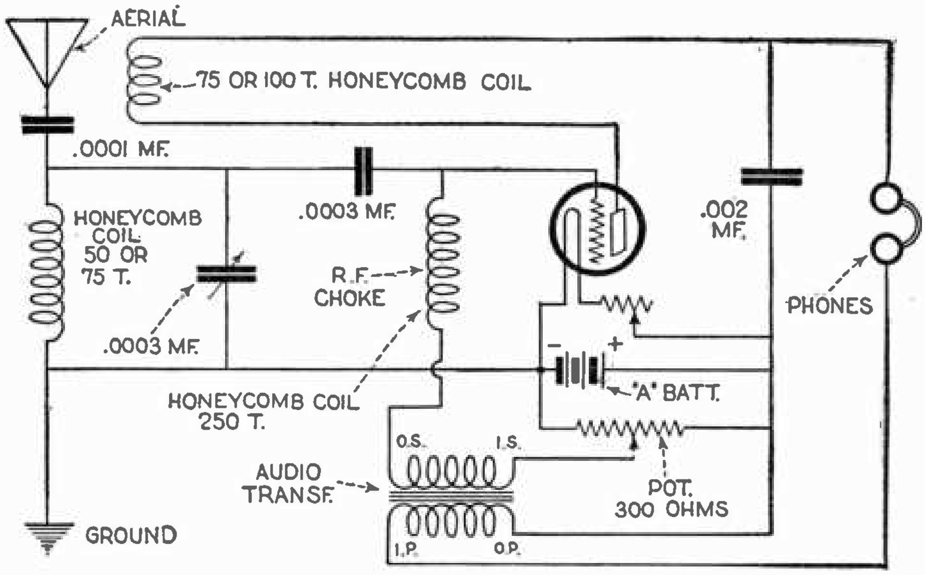

The Cowper circuit uses no "B" battery and is quite stable in operation.

SOUNDER—An instrument used for receiving telegraph signals which utilizes the attraction of an armature by an electromagnet to make sounds as the armature hits against stops at the beginning and end of each current impulse. The sounds thus produced form the characteristic dots and dashes of the telegraph code and can be easily read by the experienced telegraph operator.

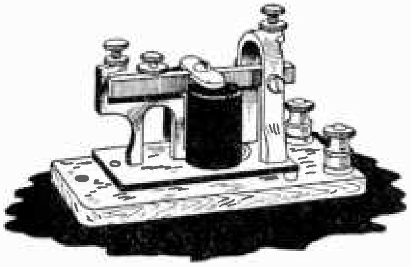

Telegraph sounder.

SOUND WAVE—A wave of alternate condensation and rarefication through an elastic body such as air, water, etc. Sound waves travel at the rate of 1090 feet per second in air.

SOUTH MAGNETIC POLE—Situated in Latitude 70 South and Longitude 102 East. The South Magnetic Pole does not coincide with the geographic South Pole.

SOUTH POLE—When referring to a magnet, the pole which tends to point to the south when the magnet is freely suspended.

SOUTH SEEKING POLE—See *South Pole.*

SPACE CHARGE—The difference between the number of electrons and positive ions in unit volume, multiplied by the charge per ion. This is also called the volume density of electrification. Let us consider the distribution of electrons in a vacuum tube, between the hot filament and the plate. An electron close to the surface of the plate is attracted to the plate by two forces, attraction from the plate and repulsion from the electrons located between it and the filament. An electron close to the surface of the filament, however, is repelled by the electrons between the plate and itself, although it receives some attraction from the plate. It may either move towards the plate or go back to the filament. This will depend upon whether the plate voltage is high enough to result in a force of attraction sufficiently great to overcome the repulsive force of the space charge.

SPACE CURRENT—Current which flows between the cathode and the anode in a vacuum tube. Space current is the result of the motion of electrons through space. It does not follow the same laws which govern the flow of current in a metallic circuit, deviating from Ohm's law to a certain extent, depending upon the relation between the impedances of the tube and the metallic portion of the circuit and also upon the nature of the discharge.

SPAGHETTI—A varnish-impregnated cloth tubing used to insulate bare conductors. (See *Cambric, Varnished.*)

SPARK—See *Spark Discharge.*

SPARK COIL—An *induction coil* (q.v.) used to produce *spark discharges* (q.v.).

SPARK DISCHARGE—A passage of electrical current between two conductors across a previously non-conducting space. The discharge is accompanied by light, heat and sound. It may take place through a liquid, a solid or a gas but whenever it occurs, it means that there has been a breaking down of the dielectric stress. A spark passing through a solid will puncture it. The passage of a spark through a liquid does not affect the dielectric properties of the liquid. On the other hand, a spark passing through a gas, usually ionizes the gas for a period of time after the passage of the discharge and as a result a continuous arc may follow the first spark. An *oscillating current* (q.v.) may produce a rapid train of sparks which persists until the oscillations are damped out. (See *Sparking.*)

SPARK DISCHARGER—A form of *spark gap* (q.v.) such as a *rotary discharger* (q.v.) or a fixed gap, such as a *quenched spark gap* (q.v.). (See *Discharger*, also *Disc Discharger.*)

SPARK GAP—A break in an oscillating circuit which acts as an automatic safety valve to the condenser. Since the air between the gap has a high resistance, the condenser cannot discharge until the potential is sufficient to break down its insulation, thus permitting only heavy discharges to take place. (See *Quenched Spark Gap, Fixed Discharger,* also *Safety Gap.*)

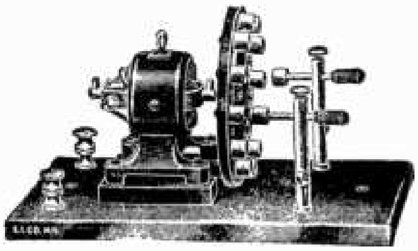

A rotary spark gap.

SPARKING—The production of spark discharges, especially when an inductive circuit is broken. Sparking at the brushes of a motor or generator is caused by armature reaction and self-induction. Sparkless commutation can be obtained by shifting brushes or by the use of *interpoles* (q.v.). Sparking is injurious to contact surfaces and should be avoided if possible. Sparking between contacts of induction coil vibrators is usually reduced by placing a condenser across the gap.

SPARK MICROMETER—A graduated adjustable spark gap for determining sparking distances for various voltages, etc. This device also permits the approximate measurement of high voltages provided the sparking distances for the particular type of terminals, are known.

SPARK RATE, SPARK FREQUENCY, or GROUP FREQUENCY—The number of sparks per second occurring in a spark system of wireless telegraph transmission. It refers to the *Group Frequency* of the wave train rather than to the waves themselves.

SPARK RECORDER—An instrument for recording telegraph signals, in which sparks from an induction coil pass through and mark a paper tape carried on a drum which turns under a metallic pointer. The use of a spark recorder dispenses with the use of ink, thus doing away with the friction of the pen on the paper.

SPARK TRANSMISSION—A wireless telegraph transmitting system which uses a succession of spark discharges in an oscillating circuit, to produce oscillations. These traverse an aerial system and a series of short trains of damped waves are emitted.

SPECIFIC GRAVITY—abbreviation S.G.—The weight of a body compared with that of another, having equal bulk, considered as a standard. The standard for liquids and solids is water, while hydrogen or air is the standard for gases. The specific gravity test of the radio storage battery is an important means of testing the condition of the battery. A *hydrometer* (q.v.) is used for this purpose.

SPECIFIC INDUCTIVE CAPACITY—symbol k—abbreviation S. I. C.—A measure of the degree to which a body permits electrostatic induction through it. The ratio between the capacities of two condensers, one employing the material under consideration as the dielectric and the other using an air or vacuum dielectric. The Specific Inductive Capacity of a material is the *inductivity* (q.v.) of that material relative to that of air. (See *Inductive Capacity* also *Dielectric Coefficient or Constant.*)

SPECIFIC RESISTANCE—The resistance in ohms, of unit length and unit cross-section of a conductor. (See *Resistivity of a Material*, also *Resistivity, Surface.*)

SPECTRUM—An image formed by rays of light or other radiant energy in which the parts are arranged according to their refrangibility or wave length. The image may be visible or invisible. The arrangement is such that all parts of the same wave length fall together, while those of different wave lengths are separated from each other forming a regular series. The spectrum produced by the light of the sun passing through a triangular glass prism and falling on a screen is one of the most common forms. The various colors, since they are unequally refracted, are spread out into a band, showing the seven rainbow colors. The red is at one end (that of the least deviation), while the violet is at the other end.

SPEECH AMPLIFIER—An audio frequency amplifier designed especially for public address systems, where addresses are to be heard over a comparatively wide area. By means of speech amplifiers, it is possible for one speaker to address an unlimited number of listeners.

SPEECH MODULATION—The modulation of radio frequency currents, as utilized in radio telephony. The production of speech modulated waves calls for a source of undamped waves and a method of causing variations in the current output of this source which will accurately correspond to the vibrations of the voice. The source of undamped waves may be a high-frequency alternator or an electron tube generator, the latter being in most common use. The radio frequency antenna current may be varied by inserting a speech-controlled variable resistance, such as a microphone, in the antenna circuit at the transmitter. The microphone may be put in the direct current power supply of the generating system, so that the radio frequency output of the system will be varied as the power input is varied. (See *Modulation*, also *Modulation Frequency Ratio.*)

SPEECH VIBRATIONS—The wave corresponding to a given sound. There are various methods of obtaining graphically, the picture of the wave form of any particular sound. One method is to make a phonograph record of the sound and then as the sound is reproduced from the record, to magnify the movement of the needle, using a lever to trace the form of the waves.

SPIDER WEB COIL—A form of inductance coil in which the wires are wound on a frame consisting of radiating arms similar to the spokes of a wheel. The wire is wound in successive turns in and out around the arms, starting from the center, until it reaches the ends of the arms. An odd number of turns are used so that alternate turns will follow the same wave; that is to say, adjacent turns will be on opposite sides and separated by an arm. Coils of this type have comparatively low distributed capacity and in addition have the advantage of being extremely compact. Figure 1 illustrates the frame on which the wire is wound. This is known as the *spider.* A spider cut from 1/16th inch bakelite, as shown in Figure 2, can also be used, although, owing to a smaller number of divisions, the latter does not have such a large inductance as the former. In winding, the wire is fastened around one of the pegs and is then taken in and out around the pegs, as shown in Figure 3. When starting the second turn, the wire will be around the pegs the opposite way to the first turn, and all successive odd-number turns will be on the same side as No. 1, while all successive even number turns will be on the same side as No. 2. The finished coil is shown in Figure 4. Spider web coils are also known as *basket wound* (q.v.), *basket woven*, or *stagger wound* coils.

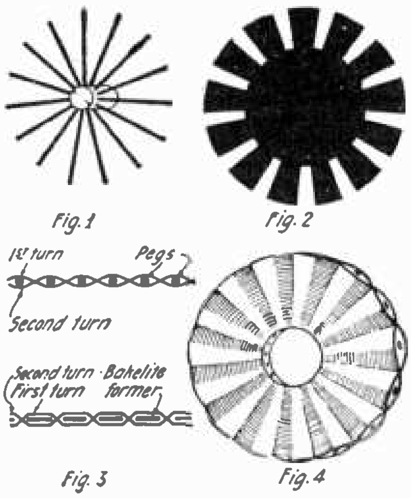

Details of spider web coil construction.

SPLICE—A method of joining two or more conductors by interweaving or entwining the strands, in a similar manner to that of splicing a rope.

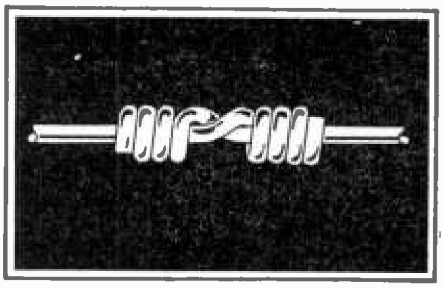

A strong splice.

SPREADER—A spar or pole used on an aerial, where two or more wires are used, to keep these wires properly spread out and parallel to each other.

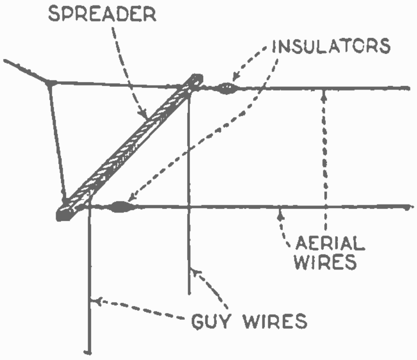

Aerial showing use of spreader.

SQUARE MIL—A square mil is the area of a square, one mil, or 0.001 inch on each side. (See *Circular, Mil.*)

SQUIER, MAJOR-GENERAL SIR GEORGE OWEN—American radio authority. He was educated at the Johns Hopkins University and became a research student under Professor Rowland and Sir William Preece at the British General Post Office. In 1904 he published his famous paper on the absorption of electromagnetic waves by living vegetable organisms and showed how trees could be used for the reception of radio messages. In 1911 he read a paper on multiple telephony before the American Institute of Electrical Engineers. In 1912 he was awarded the Elliot Cresson Gold Medal for his researches in multiplex telephony and in 1919 the Franklin Medal of the Franklin Institute. Major-General Squier was awarded

Major-General George Owen Squier.

the K. C. M. G. for distinguished services during the World War. He is a member of the National Academy of Sciences and the International Electrotechnical Commission. He is the inventor of *Line Radio* (q.v.) which is also known as *Wired Wireless, Wire-Radio Telephony, Guided Wave Telephony*, and *Carrier Current Telephony.*

SQUIRREL CAGE INDUCTION MOTOR—A type of induction motor having a rotor consisting of copper bars connected to rings at each end so as to form a short-circuited system. There are no windings in this type of rotor nor are there any external connections through slip rings. These motors are used for constant speed work, where starting is necessary only at infrequent intervals. The squirrel cage motor draws a large starting current, but has a relatively small starting torque. It is possible, however, by properly designing the rotor so as to have enough resistance, to use small motors of this type for loads requiring frequent starting, rapid acceleration and high starting torque.

SQUIRREL CAGE ROTOR—See *Squirrel Cage Induction Motor.*

STAR GROUPING—A method of connecting up polyphase apparatus or circuits. One end of each phase is connected to a common point, usually called the neutral point. This method of connection is called a "Y" grouping, in the case of a three-phase system.

STATIC—An irregular disturbing noise, heard in the radio loud speaker or head set due to atmospheric discharges, lightning and similar phenomena. A common form of static produces an intermittent crashing sound. Other types cause grinding noises or hissing sounds. Nearby lightning is usually accompanied by a sharp snap in the loud speaker. Various methods have been devised for eliminating or reducing static, although it cannot be said that any of these has been entirely successful. The combination of coil and ground aerials results in the reduction of static; also the use of *beat reception*, using continuous waves as in certain methods of wireless telegraphy. Static is also referred to under various other names, such as *strays, atmospherics, X's*, etc.

STATIC CHARACTERISTIC—The curves obtained by the use of a steady direct-current potential for showing the performance, efficiency, etc., of a vacuum tube. The static characteristics of the tube itself are obtained when the resistance of the external circuit is very small in comparison to the plate resistance of the tube. In cases where the external circuit contains reactance, the true characteristic could not be obtained using direct current. In this case, an alternating electromotive force is impressed on the grid and the curves obtained are known as *dynamic characteristics* (q.v.).

STATIC ELECTRICITY—Electricity which is stored in a circuit, manifesting itself in the form of charges at high potential. Electricity, as produced by frictional or influence machines.

STATIC ELIMINATORS—A device for eliminating or reducing the effect of *static* (q.v.). Various circuits have been proposed for accomplishing static elimination. Some of the most recent and successful of these are the McCaa anti-static circuits described in the July, 1925, issue of RADIO REVIEW, on page 63. There are two types of McCaa circuits, one applicable to radio telegraphy and the other to radio telephony. Both have for their object the reduction of signal static ratio. The wiring diagram shown is that of a standard five-tube neutrodyne receiver plus a McCaa anti-static device of the receiver type. In this diagram Ta is the repeater tube, Tb and Tc are the two radio-frequency amplifier tubes, Td is the detector tube and Te and Tf are the two audio-frequency amplifier tubes. The circuit from tube Tb on is unchanged and is the same as any standard neutrodyne receiver. The input circuit of tube Tb, however, and also the antenna circuit, have been altered to make possible the installation of the anti-static device.

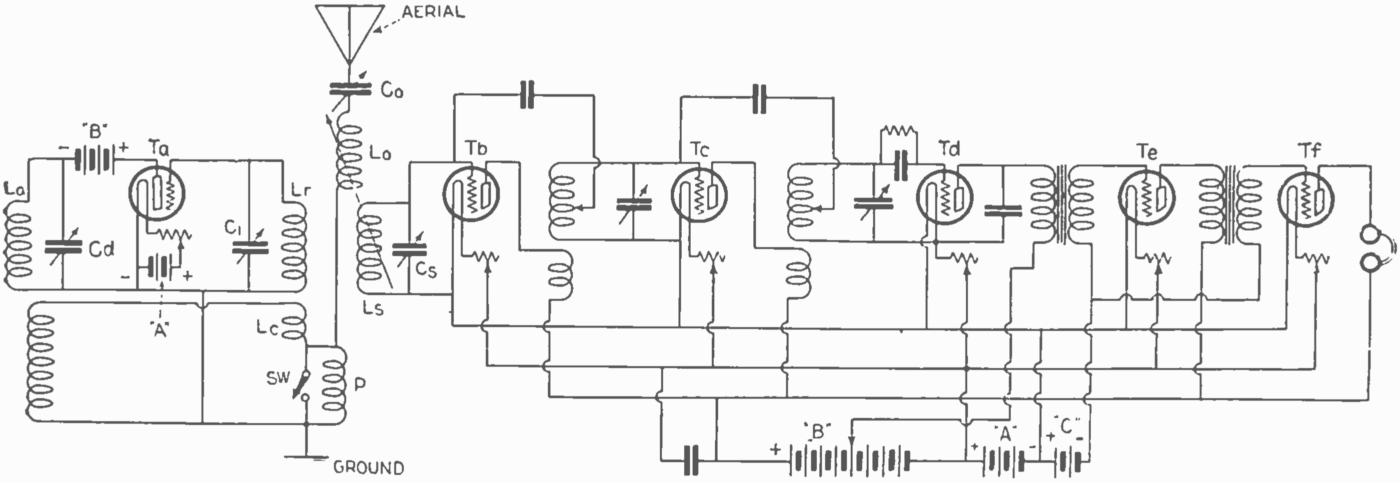

Arrangements of a static reducing device employing repeater circuits.

For the neutroformer which usually couples the first tube of a neutrodyne to the antenna circuit, the coils Lo, Ls and Lp have been substituted. These coils may either be homemade coils or honeycomb coils. If honeycomb coils are selected each may have fifty turns. The coupling between Ls and Lp should be fixed so that the coils are about an inch apart and some arrangement should be made for changing the coupling between Lo and Ls. If homemade coils are used each coil should have approximately sixty turns of No. 22 d. c. c. wire on a three-inch diameter tube and they should be coupled the same as the honeycomb coils.

The coils Lr and Lc are also coupled and may be either 50 and 10 turn honeycomb coils, respectively, or 60 and 10 turn homemade coils wound on a three-inch diameter tube with No. 22 d. c. c. wire. The coupling between these coils is not variable. After the proper value of coupling has been found by experiment it may be made permanent.

Coils Ld and Lq are also shown coupled in the diagram. They may be 50 and 75 turn honeycomb coils, respectively, or 60 and 90 turn homemade coils wound on a three-inch diameter tube with No. 22 d. c. c. wire. The coupling between these coils is also fixed after the proper value has been determined by experiment.

The proper values for the various capacities used in this circuit are as follows: Co., .001 mfd. (43 plate); Cs, .000035 mfd. (17 plate); Cr., .00035 mfd. (17 plate); Cd, .00035 mfd. (17 plate).

In building this set it is advisable to place both the anti-static device and the receiver in the same cabinet. This cabinet must be shielded and the shield should be connected to the ground. It is also advisable to shield the anti-static device from the receiver proper. The batteries should also be shielded either by placing them inside the cabinet or in metal boxes which are connected to the ground. The wires from the batteries to the set may be effectively shielded by using BX cable, the outside covering of which should also be connected to the ground. For the phone wires special phone cords which are shielded with a flexible copper braid are desirable.

STATOR—The stationary plates of a rotary condenser. The fixed portion of a motor or a generator, which carries a winding. In induction motors, that portion of the motor which is fixed in position and which carries the winding connected to the external circuit. (See *Rotor*, also *Alternator.*)

STATOR PLATES—The plates of a condenser which remain stationary in position as differentiated from the plates which rotate. (See *Shielding*, also *Stator.*)

STOPPING CONDENSER—A condenser which opposes or prevents the flow of direct current, but which allows alternating current to flow in a circuit.

STORAGE BATTERY—See *Secondary Cell.*

STORAGE BATTERY TESTS—The testing instrument most commonly used in storage battery work is the *hydrometer* (q.v.). The purpose of the hydrometer is to determine the state of the charge of the battery, as indicated by the *relative* specific gravity of the electrolyte. The specific gravity of a fully charged battery is not a fixed value but depends entirely upon the standard specific gravity employed by its manufacturer which may be anywhere from 1.250 to 1.300. The user should always ascertain the correct specific gravity for the make of battery before attempting to test it with a hydrometer. A voltmeter may be employed as a means of testing a storage battery provided it is of the proper type and is used correctly. It must be accurate with sufficient length of scale to read tenths of a volt clearly. Voltmeters are obtainable which may be used for both the "A" and "B" batteries whether dry or storage. It is not possible to test the storage "A" battery with cheap types of so-called pocket voltmeters. It must be kept in mind that a voltmeter reading on a storage "A" battery is of value only when the battery is either charging or discharging, except when the battery is practically dead. A voltmeter reading taken when the battery is not connected to the circuit, either charging or discharging, may be misleading. If a reading is taken with the filaments lighted and the voltmeter indicates a full six volts, the battery is charged. Under the same conditions, if the voltmeter lags appreciably below six volts the battery should be charged in order to keep it in the best condition. When it falls to 5.4 volts the battery is discharged and cannot possibly operate the set with any degree of satisfaction. When the battery is connected to the charger and the voltmeter shows from 7.5 to 7.8 volts the battery is fully charged. It is possible to test roughly without instruments. If neither a hydrometer nor a voltmeter is handy it is possible to recognize the necessity for charging by the fact that the amplifying rheostats have to be moved forward of the normal working position, indicating low voltage and that the battery must be charged. When the battery is on the charging line, being charged at the ordinary rate of the commercial charger, and all of the cells of the battery are bubbling, this is an indication that the battery is charged. An interesting meter has recently been placed on the market which should prove of considerable use to owners of storage batteries. It is an ammeter of special design which is to be connected in series with the storage battery on charge. There is a third terminal on back of the instrument by means of which the same instrument can be used to measure the current consumption of the filaments. By pressing a small button in the center of the meter, the needle swings over to the right-hand portion of the scale and indicates whether the battery is fully charged, half charged or low.

STORAGE CELL—See *Secondary Cell.*

STRAIN INSULATOR—An insulator used under tension, as for example the insulators to which the guy wires of an aerial are attached. The simplest form of strain insulator consists of a cylinder or ball having an eye-bolt at each end. Where the aerial guy is fastened to a strain insulator, the purpose of the insulator is to prevent the guy from having a natural wave length nearly the same as the wave length of the antenna. A common form of strain insulator is of nearly spherical form and is so grooved as to carry the two wires firmly, without permitting them to come into contact.

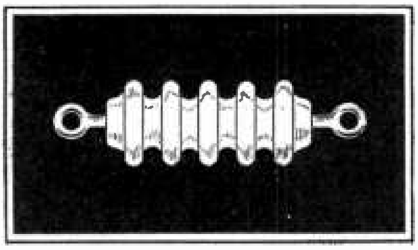

A strain insulator.

STRANDED WIRE—A wire made up of a number of smaller wires, twisted or braided together. An uninsulated, large stranded wire is generally referred to as a bare cable. Stranded wires are of four general classes; (1) wire braid, (2) bunched wire, (3) rope-lay cables, (4) concentric-lay cables. Wire braid is used to afford protection to the insulation of various types of cable. In flat form, it is used as a flexible lead. Bunched wire is a type of cable also referred to as a cord. This form of cable is very flexible, since the individual wires are small. Rope-lay cables are made up of a central core of stranded wire, having one or more layers of stranded wire wound helically about it. Concentric-lay cables are formed in the same way as the rope-lay type, except that the core and layers are of individual solid wires instead of stranded. Stranded wire is very often used in radio work as aerial wire. It has the advantage of being more flexible than solid wire and it also has a lower resistance at high frequencies because of *skin effect* (q.v.). A greater cross-sectional area is available in the stranded conductor than in the solid conductor of the same weight, for carrying current. If used for radio frequency currents, however it is essential that the individual strands be enameled if the lower resistance is to be attained.

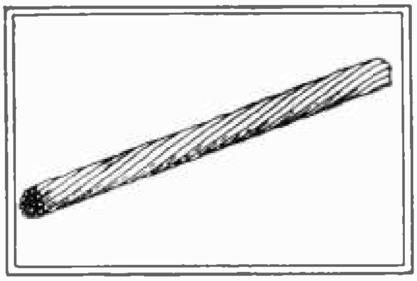

Stranded wire.

STRAY CURRENTS—Currents induced by stray magnetic fields such as eddy currents. Such currents always result in an energy loss. In electric railway systems, the currents returning through the earth, through piping, etc., rather than through the path provided are known as stray currents.

STRAY FLUX—Magnetic flux which is not usefully employed. The *leakage flux* or *leakage lines* (q.v.) which stray from the closed magnetic path provided in a transformer or other electromagnetic or magnetic apparatus. (See *Magnetic Leakage.*)

STRAYS—Atmospheric disturbances which manifest themselves as noises in the radio loud speaker or head set. (See *Static.*)

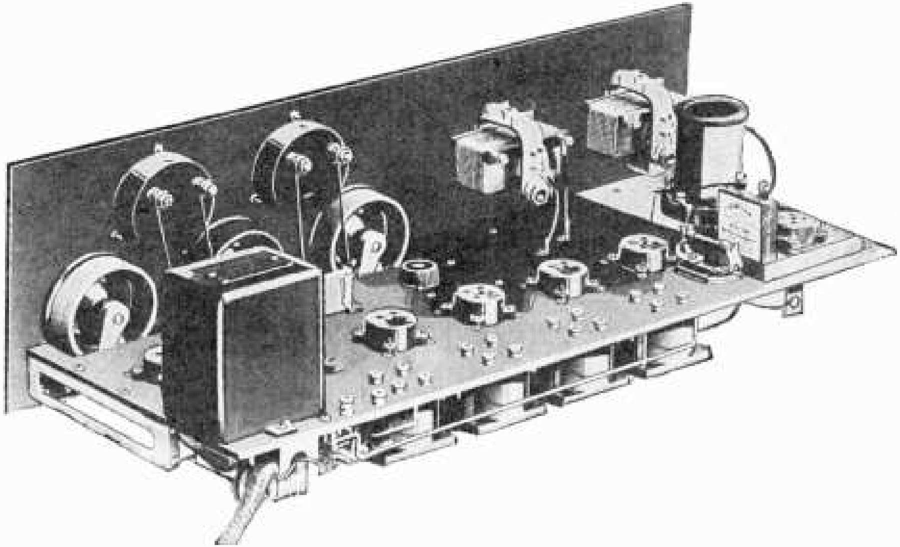

This is the rear view of the Callies Super. Note particularly the long wave transformers at the rear of the set and how they support the sub-panel.

SUB-PANEL—A secondary panel in a radio set, mounted at an angle with the main panel. The sub-panel usually carries the transformers, sockets, grid condenser and grid leak. Sub-panels are made of bakelite or composition, hard rubber, etc. (See *Panel.*)

SUPER-HETERODYNE—See *Super-Heterodyne Receiver.*

SUPER-HETERODYNE RECEIVER—A circuit used in radio reception in which the wave lengths of the incoming signals are increased to several thousand meters, by the aid of a local source of oscillations. It is a recognized fact that radio signals at "radio frequencies," that is, the original signals as they are impressed on the receiving set from the aerial, must be amplified or built up in some way to operate the detector. However, signals at low wave lengths such as in use for broadcasting, cannot be amplified very efficiently at radio frequencies. By changing the waves from the ordinary broadcast band between about 200 and 550 meters to wave lengths of from 4,500 to 10,000, it is possible to obtain more complete amplification. As a result the receiver is more sensitive and greater distance can be received. In the super-heterodyne receiver, a *difference frequency*, termed a *heterodyne note*, is created. This is done by an arrangement for generating oscillations locally. When the incoming oscillations are combined or superimposed with the local oscillations in such a manner that there is a difference in frequency, this difference will be in the form of a new set of oscillations. If the receiver is arranged in such a form that the new set of oscillations has a comparatively low frequency, the wave length will be correspondingly high. In a standard super-heterodyne receiver the action can be imagined as follows: the signal energy from a certain transmitting station is gathered on an antenna or loop aerial. These signals are tuned in by means of a condenser in the case of a loop aerial, or by a regular tuning device such as a coupler and condenser in the case of an outside antenna. The energy thus obtained is passed through to the grid member of a vacuum tube, which is known as the first detector tube. (In some cases the energy is first amplified by a radio frequency amplifier, consisting essentially of a vacuum tube and a radio frequency transformer). At this point in the circuit, between the tuning device and the first detector, the apparatus for producing the heterodyne action is located. This part of the outfit consists of a vacuum tube with coupled coils in the grid and plate circuits, with a variable condenser (the heterodyne condenser) connected across them, having a suitable value for obtaining any desired local frequency. In this way, the difference between the incoming oscillations and the locally generated oscillations may be adjusted to a prearranged value or frequency. In inductive relation to the grid and plate coils, another coil is arranged, referred to as a *pick-up* coil. The energy obtained from the first detector tube signals having a frequency equal to the difference between that of the incoming signals and the local oscillations, is sent to a *filter transformer* or *tuned filter*. The construction of this filter and the number of windings used is of the utmost importance, as it determines the frequency of the new set of oscillations and will permit the passage of no other frequency than the desired band. This is a very essential feature, as it is necessary to pass only the desired frequency in order to permit maximum transfer of energy. The energy at this predetermined frequency is then passed to the *intermediate frequency amplifier* (q.v.). After being amplified or built up by the intermediate frequency amplifier consisting of tubes and transformers, the energy is then passed to the second detector tube and rectified or changed from radio frequency. The audio frequency signals are now amplified by one or two audio frequency amplifiers consisting of amplifying tubes and transformers and the resulting signals, speech or music

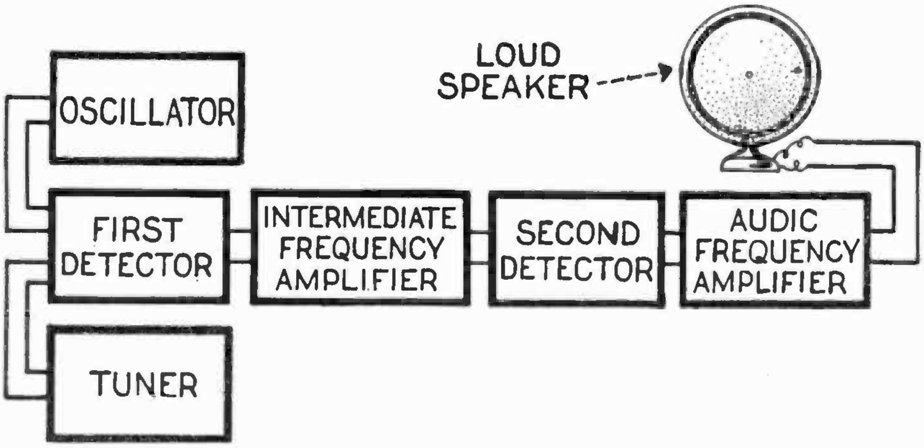

LOUD SPEAKER
OSCILLATOR
FIRST DETECTOR
INTERMEDIATE FREQUENCY AMPLIFIER
SECOND DETECTOR
AUDIC FREQUENCY AMPLIFIER
TUNER

Fig. 1.—This shows the general arrangement of the units of a typical super-heterodyne receiver.

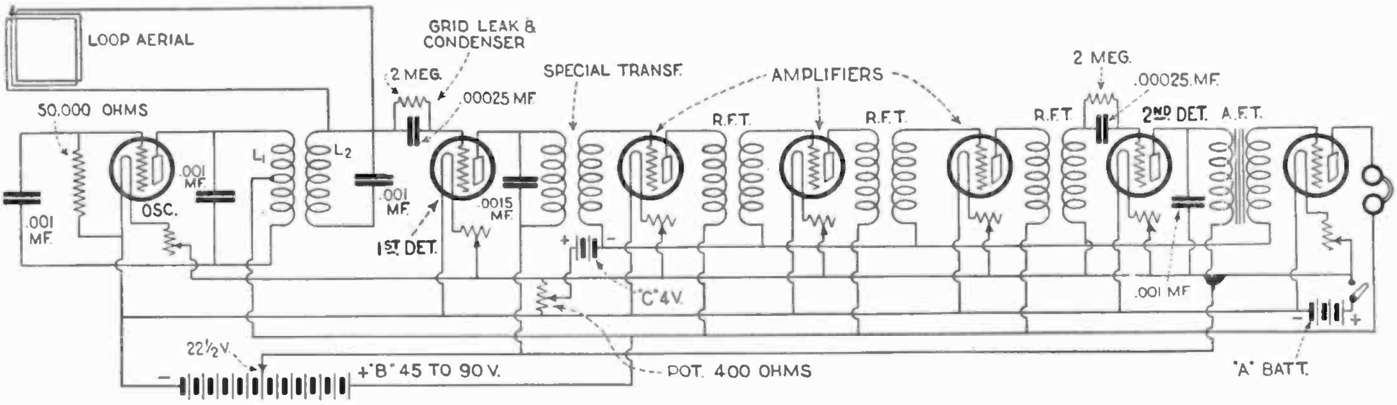

LOOP AERIAL
GRID LEAK & CONDENSER
2 MEG.
.00025 MF
SPECIAL TRANSF.
AMPLIFIERS
2 MEG.
.00025 MF
50,000 OHMS
R.F.T.
R.F.T.
R.F.T.
2ND DET.
A.F.T.
.001 MF
OSC.
.001 MF
L1
L2
.001 MF
.0015 MF
1ST DET.
"C" 4V.
.001 MF
22½V.
+"B" 45 TO 90 V.
POT. 400 OHMS
"A" BATT.

Fig. 2.—Wiring diagram of a standard 7-tube super-heterodyne.

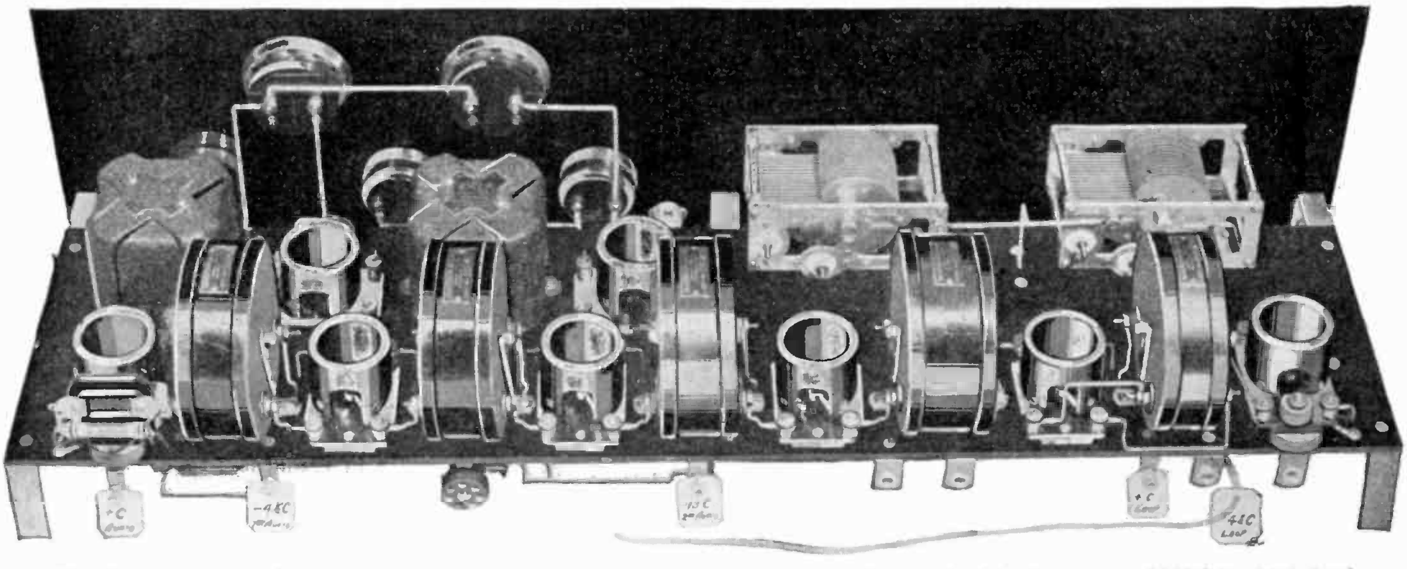

The photograph above shows an eight-tube super-heterodyne set of extra fine quality. Each of the heteroformers are shielded in nickel-plated metal shields. The two variable condensers for tuning the aerial and oscillator circuits should be of the straight line frequency type. The two audio frequency transformers are extra large concert type, the first stage of audio being all that is necessary for average reception on a loud speaker.

are used to operate the loud speaker. Figure 1 shows the general arrangement of the various sections or units of a typical Super-Heterodyne receiver. To go a little more fully into the action of the receiver, let us suppose that a certain station is broadcasting on a wave length of 400 meters. Every wave length is equal to a certain frequency.

That is to say, every wave has a certain number of vibrations per second. The frequency of a wave 400 meters in length will be about 750,000 cycles or 750 kilocycles. By adjusting the wave length condenser, the aerial or loop circuit is placed in resonance with this particular wave length or frequency. This energy is passed through the oscillator pick-up coil and the local oscillations super-imposed on it.

The oscillator condenser, let us say, has been adjusted to permit the oscillator circuit to produce oscillations of a frequency such that the difference between the incoming oscillations and the local ones will be in resonance with the windings of the tuned filter. For instance let us suppose that the tuned filter is arranged for tuning to 10,000 meters, which corresponds to 30,000 cycles.

Now the oscillator circuit must be tuned either 30,000 cycles above or below the incoming wave frequency. In other words, if the incoming wave frequency is as stated—750,000 meters, the oscillator circuit will be tuned to either 780,000 or 720,000 cycles so that the difference between the two frequencies will be 30,000 cycles. No matter what the frequency of the incoming oscillations, the local oscillating circuit will be tuned in a manner to produce the definite difference of 30,000 cycles. These signals will then be passed to the intermediate frequency amplifier and built up before being passed to the second detector tube for rectification.

After this, the signals that are now of audio frequency can be further amplified or built up by the audio frequency amplifier. The two operations that are of the utmost importance in receiving with the super-heterodyne are first, the tuning of the incoming signals and second, the adjustment of the oscillator circuit to produce the difference between two frequencies according to a definite plan. Figure 2 shows the hook-up of a typical Super-Heterodyne receiver.

SUPER-REGENERATION—See *Super-Regenerative Circuit.*

SUPER-REGENERATIVE CIRCUIT— A radio receiving circuit which permits reception under conditions of operation, which in an ordinary regenerative set would result in howling. In the regenerative set, which is tightly coupled, the signal strength increases to a maximum, and thereafter a still further increase in regeneration will result in greatly increased signal strength for a minute fraction of time, followed by violent oscillations. In the super-regenerative circuit, the signals are received at the point of very high amplification by suitably controlling the set. The great advantage of the super-regenerative circuit is the enormous amplification obtainable. Another advantage is the fact that radio code signals are not amplified to the same extent as radio telephony. The super-regenerative circuit was devised by Major Armstrong. The *Flewelling Circuit* (q.v.) is a modification of the Armstrong circuit. (See *Armstrong Circuits.*)

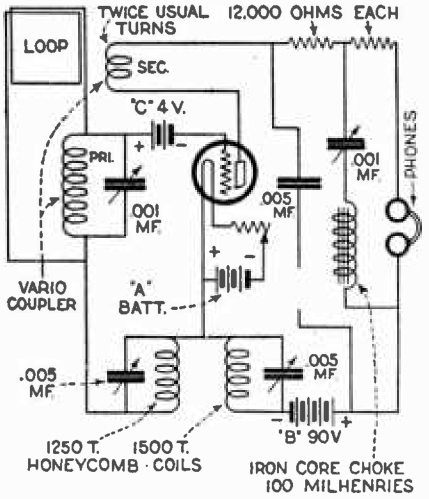

Flivver Armstrong Super-Regenerative circuit.

SURFACE MAGNETISM—A synonym for *Free Magnetism* (q.v.).

SUSTAINED WAVES—Continuous or undamped waves. Waves in which similar current cycles follow each other continuously instead of being broken up into groups like *damped waves* (q.v.). The sustained waves may be interrupted by means of a sending key and thus used for sending radio telegraph messages or they may be modulated and thus utilized in radio telephony. Sustained waves are obtained by the use of high frequency alternators, arc condensers, or thermionic electron tubes (vacuum tubes). The use of sustained waves has made radio telephony practicable. Sustained waves permit sharper tuning, require less energy at the transmitter than in the case of undamped waves, and permit the use of "beat" reception and other sensitive methods of reception. (See *Continuous Waves.*)

S.W.G.—Abbreviation for Standard Wire Gauge. The full name of this wire gage is the *British Standard Wire Gauge*. It is also referred to as the *New British Standard*, the *Imperial Wire Gauge*, and the *English Legal Standard.* It is the legally adopted standard of Great Britain. The following table shows a comparison between the American Wire Gauge (B. & S. Gauge and the Standard Wire Gauge:

Gauge No.	*American Wire Gauge Dia. in Mils*	*Standard Wire Gauge Dia. in Mils*
10	102	128
11	91	116
12	81	104
13	72	92
14	64	80
15	57	72
16	51	64
17	45	56
18	40	48
19	36	40
20	32	36
21	28.5	32
22	25.3	28
23	22.6	24
24	20.1	22
25	17.9	20
26	15.9	18
27	14.2	16.4
28	12.6	14.8
29	11.3	13.6
30	10.0	12.4
31	8.9	11.6
32	8.0	10.8
33	7.1	10
34	6.3	9.2
35	5.6	8.4
36	5.0	7.6
37	4.5	6.8
38	4.0	6.0
39	3.5	5.2
40	3.1	4.8

SWITCH—A device for conveniently making or breaking an electrical circuit. Switches must be designed to carry their rated current without overheating or undue voltage drop, to handle overloads, to prevent arcs on being opened and to properly insulate live parts when switch is open. The most common form of switch is the *knife switch* (q.v.). A switch which opens or closes but one circuit, that is operating in only a single position, is called a *single throw switch.* One operating when thrown in either of two positions is called a *double throw switch* (q.v.). A switch which controls only one side of a circuit is a *single pole switch;* both sides of a circuit, a *double pole switch.* The abbreviation for a single pole, single throw switch is S. P. S. T. The abbreviation for a double pole, double throw switch is D. P. D. T., etc. There are various types of switches ranging from *oil switches* used to handle enormous currents to the small push button or snap switches for turning on or off electric lights. Among the switches used in radio work may be mentioned the *ground switch* (q.v.), *aerial switch* (q.v.), *quick break switch* (q.v.) and *anti-capacity switch.* This latter usually consists of a small handle with a cam attached to its other end which serves to press together or release spring contacts. The electrostatic capacity between the springs is low, due to the construction used. The *plug* (q.v.) and *jack* (q.v.) constitute a form of switch extensively used in radio equipment. The *push-pull switch,* operating as its name implies, is often used to control the filament lighting circuits of radio receiving sets.

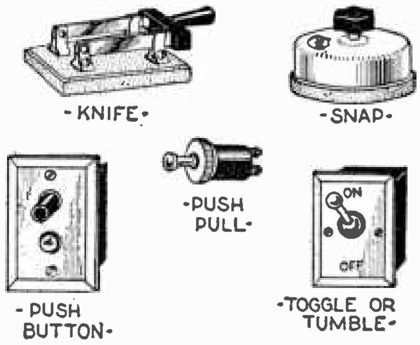

Various types of switches.

SWITCHBOARD—In its broadest sense, this term is applied to any collection of control, operating and measuring apparatus mounted on a panel or panels for the purpose of starting, stopping or otherwise controlling an electrical installation. In a small electrical plant, all control and switch gear may be mounted on a single structure, which is referred to as the switchboard. Switchboards are used at the radio broadcasting station to carry the various meters, control rheostats, switches, etc. (See *Panel.*)

SWITCHGEAR—Apparatus used in connection with the control of an electrical installation, such as switches, rheostats, starters, circuit breakers, etc.

SYMBOLS—Conventional signs used in radio hook-ups to represent various electrical and radio apparatus. The use of symbols has been standardized and results in clearer and more easily drawn diagrams. A table of symbols is given below:

Table of Radio Symbols

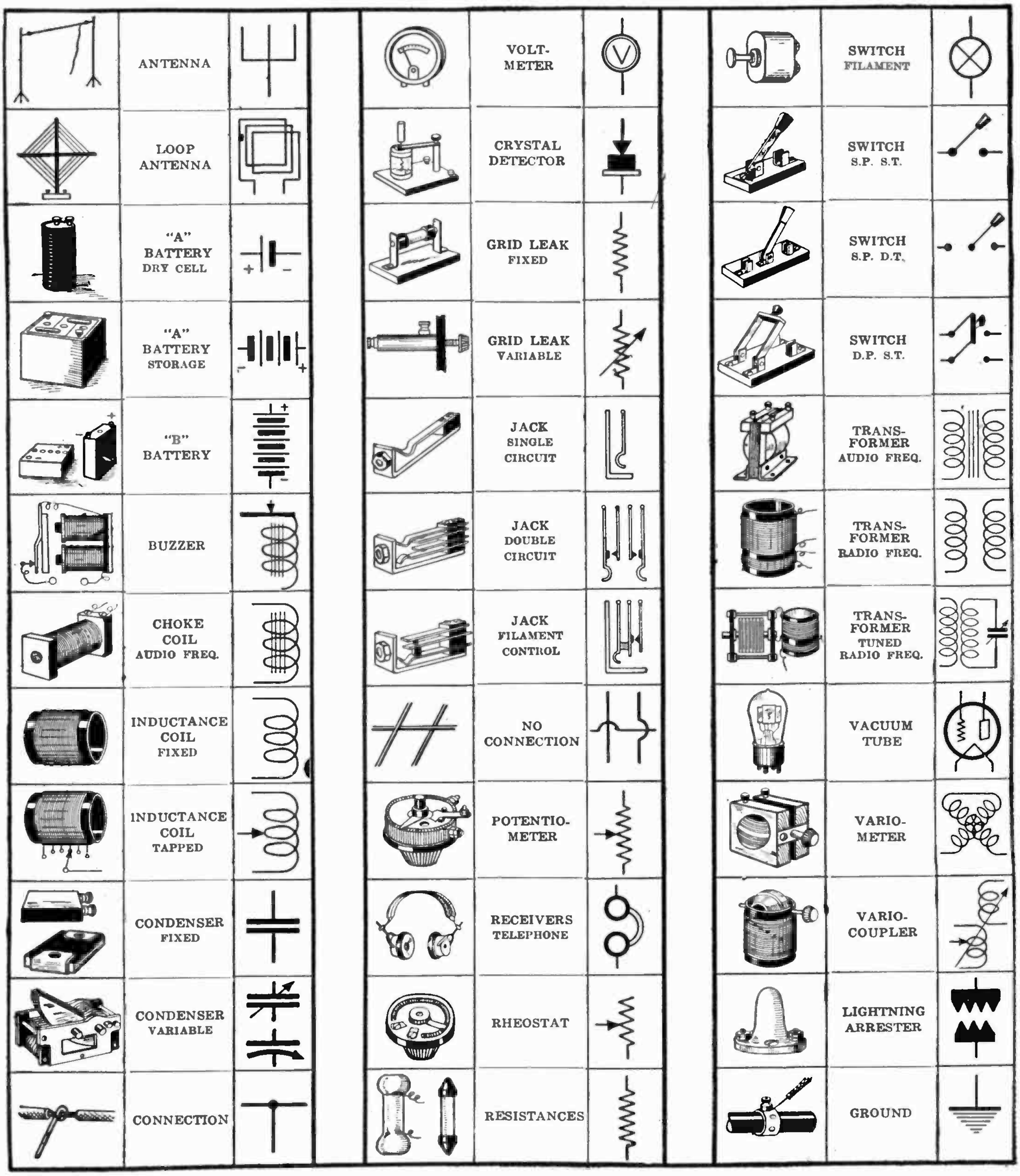
ANTENNA
LOOP ANTENNA
"A" BATTERY DRY CELL
+
–
"A" BATTERY STORAGE
–
+
"B" BATTERY
+
–
BUZZER
CHOKE COIL AUDIO FREQ.
INDUCTANCE COIL FIXED
INDUCTANCE COIL TAPPED
CONDENSER FIXED
CONDENSER VARIABLE
CONNECTION
VOLT-METER
V
CRYSTAL DETECTOR
GRID LEAK FIXED
GRID LEAK VARIABLE
JACK SINGLE CIRCUIT
JACK DOUBLE CIRCUIT
JACK FILAMENT CONTROL
NO CONNECTION
POTENTIO-METER
RECEIVERS TELEPHONE
RHEOSTAT
RESISTANCES
SWITCH FILAMENT
SWITCH S.P. S.T.
SWITCH S.P. D.T.
SWITCH D.P. S.T.
TRANS-FORMER AUDIO FREQ.
TRANS-FORMER RADIO FREQ.
TRANS-FORMER TUNED RADIO FREQ.
VACUUM TUBE
VARIO-METER
VARIO-COUPLER
LIGHTNING ARRESTER
GROUND

SYNCHRONOUS—As applied to alternating currents, these are said to be synchronous when of the same *frequency* (q.v.) and exactly in *phase* (q.v.).

SYNCHRONOUS DISCHARGER—A rotary gap discharger driven by a synchronous motor from the same line supplying the transformer in a radio telegraph transmitting system or a rotary gap mounted on the shaft of the alternator supplying the 500 cycle or higher frequency current. (See *Disc Discharger.*)

SYNTONY—The relationship between two oscillating currents when one resonates to the waves produced by the oscillations of the other.

T

TABLE OF DIELECTRIC STRENGTH—A table showing the voltages at which certain thicknesses of various materials will break down or puncture. The following table has been obtained from Pender's Handbook for Electrical Engineers:

MATERIAL	DIELECTRIC STRENGTH Specimen thickness mm.	DIELECTRIC STRENGTH Kv. per mm. (a)
Ambroin	0.84	6.0
Asbestos paper	1.2	4.2
Asphalt (Byerlyte)	3.6	14.0
Bakelite, C-1		up to 27.5
Bakelite, wood molding mixture		17.7 to 21.6
Bakelite, asbestos molding mixture		up to 9.8
Bakelite, Continental	3.2	15.7
Bakelite, Micarta-213		up to 31.4
Bakelite, Micarta-21D		5.9
Bakelite, Micarta-21H		15.7
Bakelite-Dilecto-X	3.2	25.6
Bakelite-Dilecto-XX	3.2	25.6
Celluloid (clear)	0.25	12 to 28
Celluloid (colored)	0.25	10.2 to 18.9
Condensite (molded)	5.7	19.7
Condensite (celoron)	5.7	29.5
Conite	0.13	15.7
Copal	3.0	3.2
Empire cloth, canvas	.41	28.9
Empire cloth, linen	.15	54.0
Empire cloth, muslin	.38	48.0
Empire cloth, silk	.15	48.0
Faturan	3.	10
Fiber, vulcanized, including hard fiber all colors	0.79	8.9 to 16.7
	3.2	4.9 to 10.8
	6.4	3.9 to 8.9
	12.7	3 to 5.9
Galalith (white)		6 to 8.5
Glass (ordinary)		8 to 9
Hermit		2 to 3
Jute (impregnated)	6	1.2
Lava		3 to 10
Litholite	4.5	4.5
Marble		6.5
Mica	6	21 to 28
Micabond, plate	1.6	37.5
Micabond, flexible	1.6	23.1
Micanite, plate	1.6	37.5
Micanite, flexible	1.6	23.1
Minerallac		39.4
Paper	0.13	8.7
Paraffin (parawax)		11.5
Porcelain	20	8
Pressboard (oiled)	0.25	39.3
Pressboard (oiled)	1.58	29.2
Pressboard (oiled)	3.17	21.1
Pressboard (varnished)	0.25	26.3
Pressboard (varnished)	1.58	15.5
Pressboard (varnished)	3.17	9.5
Presspahn		5.2 to 9.3
Redmanol (molded)	5.1	11.8 to 18.5
Redmanol (laminated)	0.8	41 to 51
Rubber (hard)	0.5	70
Slate	10.3	1.3
Vulcabeston	1.9	31.5 to 7.1
Wood (maple), paraffined	15.2	4.6

(a) To obtain volts per mil, multiply kilovolts per millimeter by 25.4.

"T" ANTENNA—A flat top antenna in which the down lead is taken from the center of the horizontal portion. (See *Aerial*, also *Flat Top Aerial.*)

TAILS—Name given to small iron wires forming core of *induction coils* (q.v.).

TANGENT GALVANOMETER—A type of galvanometer in which the strength of the currents which pass through the coil is proportional to the tangent of the angle of deflection of the magnetic needle in the center of the coil. In construction, the tangent galvanometer consists essentially of a small magnetic needle suspended in the center of a large circular coil made of the few turns of insulated wire. Attached to the needle is a light aluminum pointer which allows the deflections to be read on a horizontal scale. The coil is placed in the magnetic meridian with the needle and the coil in the same plane. When the current passes through the coil, the pointer is deflected and the angle of deflection recorded. Knowing the constant of the galvanometer, it is possible to calculate the current in absolute units or in amperes by multiplying the constant by the tangent of the angle recorded.

TANGENT OF AN ANGLE—In a right angle triangle having the given angle as one of its angles, the tangent of the angle under consideration is equal to the ratio of the side opposite the angle to the base of the triangle.

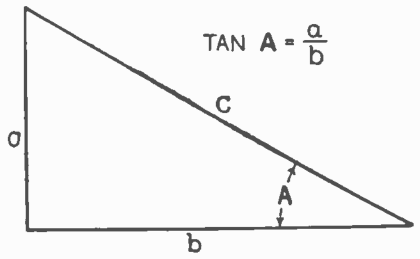

The tangent of angle A is the ratio of side a to side b.

TAP-TAPPING—The connection made to an intermediate point in a winding thus permitting the number of turns in the circuit to be varied at will. Variometers, variocouplers and other coils are made with taps, also transformers, armatures of rotary converters, etc.

TAPPED—See *Tap*, also *Tapped Inductance.*

TAPPER—An electromagnetic device, similar to an electric bell, but with gong removed, arranged so that a hammer gently taps a coherer thus decohering filings after the passage of incoming oscillations. This device was formerly used as a detector of wireless telegraph signals.

TAPPING-BACK—The application of a light blow to a filings coherer to decohere the filings. (See *Tapper.*)

TELEFUNKEN SYSTEM—A wireless telegraph system employing a *quenched spark gap* (q.v.) and giving a singing spark of rather high note. This system has been developed and used by the Germans.

TELEGRAPH—Any system of transmitting intelligence from one point to another over a distance. The word was at one time applied to visual systems such as wig-wagging and semaphore, but at the present is used only to refer to electrical systems. The one exception to this rule is the use of the term to refer to signals sent either mechanically or electrically from the bridge of a ship to the engine room.

TELEGRAPHONE—A magnetic phonograph in which speech currents or signal currents are recorded as permanent magnetic impressions on a moving steel wire. Instruments such as this have been used for recording telephone conversations and also in high speed automatic wireless telegraph transmission. (See *High Speed Reception and Transmission.*)

TELEGRAPH NAUT—A nautical mile of 2,029 yards or 1.1528 statute miles. (See *Nautical Mile.*)

TELEGRAPHY, RADIO—See *Radio Telegraphy.*

TELEPHONE CONDENSER—In radio work, this is a fixed condenser shunted across the head set or loud speaker. Used in this connection, the condenser is also referred to as a by-pass condenser since it offers a path of low resistance to the radio frequency currents. By-pass condensers are usually made of tin foil with a mica or paper dielectric. Condensers used for ordinary telephone work are also referred to as telephone condensers.

TELEPHONE RECEIVER—An electromagnetic instrument by means of which variations of current are caused to reproduce sound waves corresponding to words spoken into a transmitter at a distant point. The essential features of the telephone receiver are an electromagnet, a permanent magnet and a diaphragm held in a suitable case. The diaphragm is supported by the rim of the case, at a short distance from the face of the magnet and it is attracted by the magnet when it is energized by the current flowing in the windings. The steady magnetic flux of the permanent magnet draws the diaphragm to the pole pieces, leaving a small gap between, so that the diaphragm is held under tension. The currents passing through the electromagnet are modulated in accordance with the transmitted speech. These change the tension of the diaphragm due to the change in magnetic pull and as a result the diaphragm vibrates giving forth speech or music. The telephone receivers used in radio work are of the *watch case* type. (See *Head Telephone or Head-Phone.*)

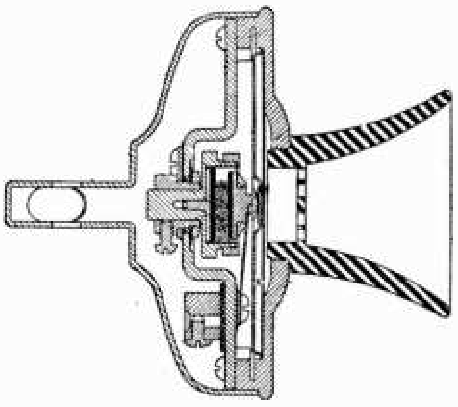

A cross sectional view of a telephone transmitter showing details of construction.

TELEPHONE TRANSMITTER—A sound-wave operated or vibration-operated device designed to produce electromagnetic waves or vibrations which correspond to the sound waves or vibrations actuating it. The standard transmitter used in ordinary telephone work is known as the "solid-back" transmitter. A gong-shaped back supports the transmitter and carries all the parts. A mouth piece of hard rubber is screwed into the front. Just to the rear of the front is placed an aluminum diaphragm which fits into a receptacle cut out for it. An insulated cushion seat of rubber is placed above and below the diaphragm. The diaphragm is held securely by damping springs having soft rubber cushions at their tips, so that it can only assume forced vibrations. An auxiliary diaphragm of mica is used which is fastened to the front electrode of the transmitter button. Both the front and the rear electrodes are made of carbon discs, between which carbon granules are placed in a cylindrical chamber lined with varnished paper. A rigidly connected pin communicates the vibrations of the diaphragm to the front electrode. The movement of the electrode varies the pressure on the carbon granules, thus giving the required variation of resistance which in turn varies the current. The *microphone* (q.v.) used in radio broadcasting is a modification of the telephone transmitter.

TELEPHONE, TUNED—See *Tuned Telephone.*

TELEPHONY, RADIO—See *Radio Telephony.*

TELEVISION—Electric vision at a distance. The transmission by wire or by radio of vision—that is to say the ability to see objects at any distance by changing the light waves to electric currents or to electromagnetic waves, transmitting these to the distant receiving point, and then retransforming them back to the original light waves. Television is still in the experimental stage. Of course, the sending of pictures by radio has been accomplished by Capt. Ranger, in this country, Belin in France, Karolus in Germany, Baird in England, and several others. Both Belin and Baird have been working on the problem of television and are reported to be nearing success. In this country, Mr. C. Francis Jenkins, of Washington, D. C., has constructed a non-commercial (that is in its present stage) machine which will send motion pictures by radio. This is practically the last word in the present development of television. It is possible to broadcast a film picture with this device at the rate of sixteen pictures per second showing the moving outlines as silhouettes, but without the intermediate tones. The method by which a photograph is transmitted by radio in the Jenkins system may be illustrated by the example of a coin covered by a thin sheet of paper on which serried pencil lines are drawn. The design of the coin appears in lines varying in intensity. The variations in the lines may be transformed into pulsing electric current by means of a light sensitive cell passing over the lines in question. Place the lines end to end and the current can modulate a carrier wave. The forming of the continuous line presented a difficulty which was solved in the following way. The inventor proceeded along the line that a prism bends a ray of light and that an image will appear upon a photographic lens placed before a picture. From this it follows that a prism placed near a lens will cause the image to be displaced sideways, the extent of the displacement being dependent on the angle of the prism. If the prism could change its angle, the result would be motion on the part of the image. The prism used is circular, with rapid change of form across the prismatic section. Around the circumference of the disc, the prism changes in thickness according to a graduated scale. Revolving the disc causes the image to move in a straight line. When the end of the prism passes the lens, there is instantaneous snapping back of the picture and the journey is repeated. If a small hole be made in the screen, the light from a given point on the picture will pass through and can be made to operate a light sensitive cell at the back of the screen. When the image is made to move down over the screen by the prismatic ring, the aperture will admit light of varying intensity, according to the light value along a line crossing the picture. This means that the light falling on the photo cell as the image travels will represent a line across the picture from top to bottom. This leads to the creation of the successive lines that make up the picture as a whole. The image is moved along from side to side by means of a second prismatic ring working along at right angles to the first, but at much lower speed. This second prism makes but one revolution during the time that the first makes one hundred, producing the hundred lines assumed to be necessary for making the picture. The cell used may be selenium or thalium oxide. The former requires twenty minutes to cover the picture, the latter six. It is obvious that to reproduce the picture at the other end the process must be reversed, the lines being replaced side by side by means of a pair of prismatic discs similar to those at the transmitting end. As noted above, the main difference between sending photographs and motion pictures is in the necessary speeding up. For this purpose, Mr. Jenkins uses a potassium light-sensitive bulb, the action of which is much more rapid than that of either selenium or thalium oxide. The cell is composed of a bulb with its inner surface coated with metallic potassium. The bulb has a high vacuum. The device has two electrodes, one being a wire through the stem, ending in a loop in the center of the bulb; the other passing through the side of the bulb and connecting with the potassium deposit. Absence of the potassium from a small part of the surface forms a "window" opposite the last named contact. Application of a positive potential to the central electrode, attended by connection with the potassium coating, permits no current to flow as long as no light enters the bulb. The admission of light frees electrons from the potassium, in numbers proportionate to the strength of the light. In this way a current flows through the bulb when light enters the "window." In this cell, there is no appreciable lag due to the amazing speed of the electrons.

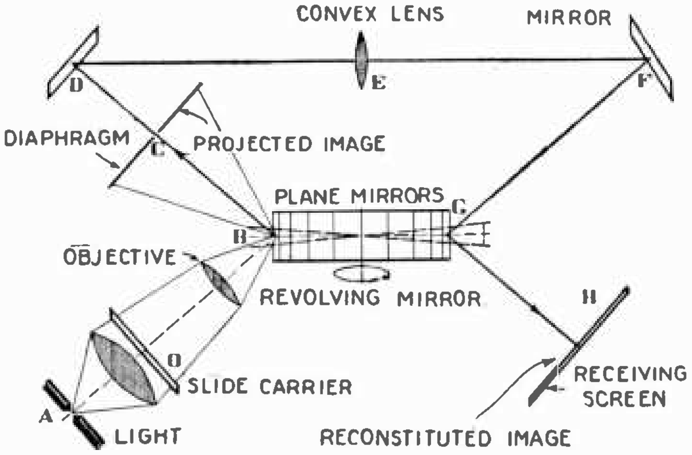

The cylinder is covered with plane mirrors, revolving downward on the side toward the lantern. One ray at a time from 1/25,000 of the area of the image, passes through the opening in C. The fixed mirrors D and F send it back to the mirror G, opposite B on the cylinder, and it is finally reflected against H in a position corresponding exactly to the portion of the image from which it was first taken. The effect of continuous vision is produced.

The sending station has a motion picture projector, a small machine used to cut up the image, a light-sensitive cell, and a transmitter. The receiving station has a radio receiving set, a machine for reassembling the picture, and a screen for the reproduction of the image.

The lamp required for the receiving station must be able to pass swiftly from darkness to extreme brilliancy, and the ordinary filament lamp could not attain the tremendous speed involved. The lamp used by Mr. Jenkins is the creation of Professor McFarlane Moore, and is capable of handling complete cycles at the rate of 75,000 to the second. This lamp has a glass bulb containing two concentric cylinders, the larger almost filling the bulb, and the smaller almost filling the larger. The inner cylinder has a small axial hole, drilled almost to the bottom and forming a deep cup. The principle of this lamp is much the same as that of the Amrad "S" valve.

It is a truism that the fundamental principle of cinema pictures is that the human eye continues to see an image

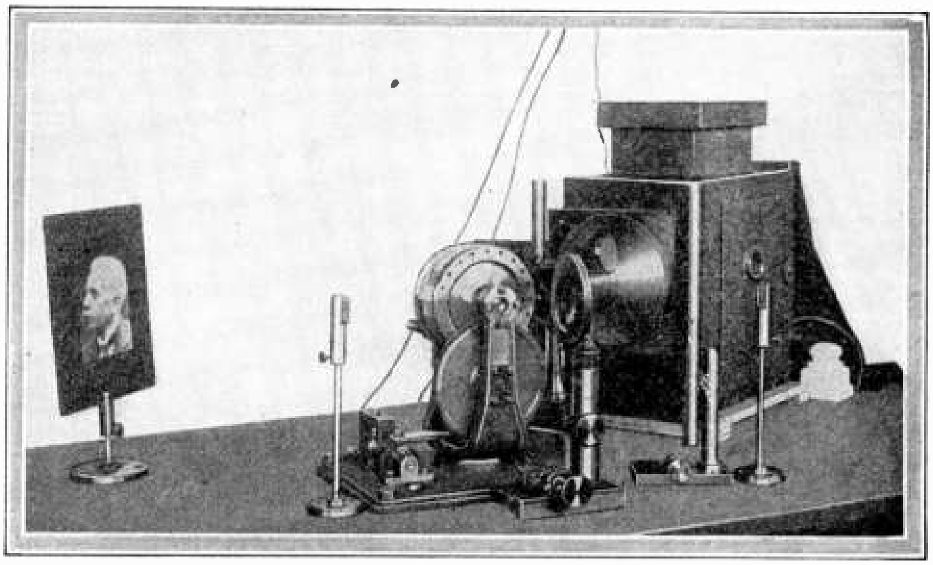

This apparatus is diagrammed above: the projecting lantern at the right; the drum of mirrors, center; and the adjusting stands for diaphragm and fixed mirrors in front. At the left a continuous image appears on the screen, although only 1/25,000 of it is actually projected at any instant.

after that image itself has disappeared. In the case of an electric spark lasting one ten thousandth of a second, the image lasts for at least one-sixteenth of a second as visualized by the eye. Mr. Jenkins assumes that if 10,000 dots of light and shade flash successively on a screen with sufficient speed, the eye will see the picture to which the dots belong even though but one dot is on the screen at a time. He, therefore, aimed at flashing the 10,000 dots on the screen within the space of one-sixteenth of a second.

Theoretically, this feat might be accomplished by halving the slow disc revolving rate to 960 revolutions per minute, keeping the faster disc at a rate exactly one hundred times greater. But the glass disc would be shattered by the centrifugal force before the attainment of a speed of 96,000 revolutions per minute.

Mr. Jenkins gets over this difficulty by the use of forty-eight lenses attached to a large aluminum disc rotating in front of the prismatic disc, thus giving the effect of a slow-moving lens. Each of these forty-eight lenses makes a line across the plate, which thus becomes the screen of the motion picture apparatus. This makes it possible to reduce the speed to one forty-eighth of that which would otherwise be required, that is, to one of 2,000 revolutions per minute.

At the receiving station, the modulated wave is received by a set but slightly different from many in common use. The last amplifier tube is one of 5 watts. By the use of the McFarlane Moore light already described, the modulations in the plate circuit of the last valve are transformed into light, and the rays are distributed over the screen by prismatic and lens discs similar to those used at the sending station. In actual practice, the picture appears on a screen about 6 in. by 8 in. More powerful lamps would allow of larger pictures.

The process described is for sending photographic pictures. But it can be adapted for the sending of direct views without the intervention of photography.

All that is done in this case, is to remove the projector and focus the lens so that it will throw the view as an image.

So far as radio features of the transmission are concerned, there is little to be explained. At the receiving station the circuit resembles the superheterodyne. The picture is at 75 kilocycles frequency and does not interfere with broadcasting.

The carrier wave may be modulated with voice frequency, so that the voice can travel with the picture, which enables vocal explanations to be given simultaneously with the flashing of the film or actual scene being transmitted.

TEMPORARY MAGNETS—Magnets which lose their magnetic properties as soon as the magnetizing force is removed. Soft iron displays the properties of a temporary magnet and hence is ideal for use as the core of electromagnets. (See *Magnet*, *Electromagnet*, also *Permanent Magnet*.)

TERMINAL—A binding post or other fitting attached to the electrodes, ends of windings or other parts of electrical apparatus so as to permit the external circuit to be connected to the apparatus.

TESLA COIL—An oscillation transformer for producing high potential discharges from oscillations of low potential. The *Tesla Coil* or *Tesla Transformer*, as it is sometimes called, is somewhat similar to the ordinary transformer, although it is much more heavily insulated and has the ends of the secondary connected to a condenser which discharges across a spark gap, thus increasing the rapidity of the oscillations, which then pass into a secondary induction coil. This second coil has no iron core. The Tesla Coil consists essentially of a primary winding having a relatively small number of turns of heavy wire and a secondary having a large number of turns of fine wire. An air, glass, or ebonite dielectric separates the primary and the secondary. Due to the sufficient thickness of the dielectric, there is no possibility of a direct discharge between the two windings. Both the primary winding and the secondary winding consist of but one layer each. The primary winding forms part of an ordinary oscillation discharge circuit. A high frequency electromotive force is induced in the secondary winding, due to the high frequency oscillating currents which flow in the primary winding thus setting up an oscillating magnetic field. Due to the high ratio between the windings of the primary and the secondary, very high voltages are set up in the secondary, in some cases as high as a million volts, and hence a very powerful brush discharge can be obtained.

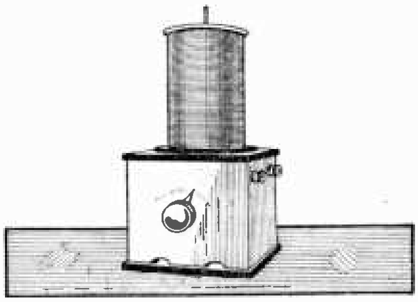

Tesla coil.

TESLA, NIKOLA—American electrical and radio expert. Tesla was born at Smiljan, in Jugo-Slavia, 1857, and was educated at Graz and Prague Universities. He entered the Austrian Telegraph Service and in 1884 he came to the United States where he became an assistant to Edison. He specialized in the study of high frequency and high potential alternating currents. Tesla is noted for his many revolutionary inventions which include polyphase alternating current systems, the rotating field alternating current motor, arc lighting, the *Tesla transformer* (q.v.), and radio apparatus of many kinds including *radio control* (q.v.) devices.

Nikola Tesla.

TESTING CIRCUIT (BUZZER)—A circuit in which a buzzer is used for the purpose of testing or of locating faults in radio apparatus. A buzzer connected in series with a dry cell can be used for locating open circuits and also for testing for short circuits. A buzzer connected in series with a battery and an inductance coil, so as to form a closed circuit and having the inductance coil placed so as to be coupled with the antenna coil of a crystal detector receiving set, can be used to determine the most sensitive point of contact of the crystal. A *tuned buzzer tester* is sometimes used in connection with radio direction finder systems.

THEORY OF CURRENT FLOW—While it is assumed for convenience, that an electric current flows along a conductor from the positive terminal of the source and back to the negative terminal, modern theory is that the electrons actually travel in the opposite direction. Whenever there is a flow of electricity in a conductor, it is believed that extremely small particles of electricity, called *electrons* (q.v.), pass along this conductor. Matter is made up of atoms, which in turn are composed of electrons. The electrons are minute particles of negative electricity revolving around a positive particle. All electric current flow is based upon the motion of electrons within a conductor.

THEORY OF DETECTOR ACTION (OF VACUUM TUBE)—The vacuum tube used as a detector performs a two-fold purpose—it rectifies the current so that it is capable of actuating a telephone receiver and in addition it amplifies or strengthens the current. The rectification takes place due to the fact that the electron flow is only possible from the filament to the plate, but not in the reverse direction. The radio frequency currents applied to the grid regulate the number of electrons passing from the filament to the plate, the variation being exactly in accordance with the variations of the radio frequency current. (See *Detector*, *Vacuum Tube*.)

THEORY OF OPERATION OF VACUUM TUBES—The operation of the three-electrode vacuum tube depends upon the emission of electrons from the hot filament to the plate and the control of the flow of these electrons by means of the potential variations applied to the grid. The vacuum tube consists of an evacuated bulb containing a tungsten filament treated with thorium, an anode or plate, which may be in the form of a cylinder or plate, and a grid consisting of a wire grating inserted between the filament and the plate. When the filament is heated it gives forth electrons which are attracted to the plate under the influence of a difference of potential by which the plate is maintained positive with respect to the filament. The grid is an automatic control, regulating the flow of electrons from filament to plate. Potential variations applied to the grid serve to vary the electron flow between the filament and the plate. The circuit in which the vacuum tube is used may be divided into two parts, the input circuit connecting the filament to the grid through the secondary of a transformer or other method or applying potential changes to the grid, and the output circuit between filament and plate. The vacuum tube may be used as a detector or rectifier, as an oscillator or generator of high frequency oscillations, as an amplifier, as an electrostatic voltmeter, as a voltage and current regulator, a power limiting device, etc.

THEORY OF PROPAGATION OF ELECTROMAGNETIC WAVES—Waves are due to vibrations transmitted through a continuous elastic medium. Electromagnetic waves are due to vibrations caused by an oscillating electric charge in a circuit (the transmitting aerial circuit). The current flowing in the aerial is oscillating (surging back and forth) at an enormously high frequency and in addition to its induction field, the aerial has associated with it a *radiation field* which spreads out in all directions from the aerial. The strength of the radiation field decreases in inverse proportion to the distance from the aerial. At any given point the strength is directly proportional to the frequency. The transmission of electromagnetic waves is considered as due to the motion of the lines of force. There is a displacement of electricity along these lines against the elastic force of the medium, termed for convenience the *ether* (q.v.). Displaced electricity continuously tends to go back to its position of rest under the action of the elastic forces. There is pressure at right angles to the lines of force in addition to the tension along them. The pressures may be considered as due to the repulsion between the displaced charges of the same sign in neighboring lines of force.

THEORY OF RECEPTION OF ELECTROMAGNETIC WAVES—As electromagnetic waves cut across a receiving aerial, the electric field intensity along the aerial alternates in value. There is an alternating potential between the aerial and the ground which gives rise to the flow of an alternating current. Another explanation is based on the principle of induction. The magnetic and electric fields, in cutting the aerial, induce electromotive forces which cause current to flow. This is radio frequency current and it is passed to the grid of the detector tube and rectified as explained under *Theory of Detector Action* (*of Vacuum Tube*). (See *Electromagnetic Waves.*)

THEORY OF VACUUM TUBE OPERATION—See *Theory of Operation of Vacuum Tubes.*

THERMAL—Pertaining to heat. There are various electrical and radio devices which depend upon heating or thermal properties. The *thermal ammeter* (q.v.) is described under the heading *Hot-Wire Ammeter.* The *thermal detector* is taken up under *Barreter* and *Liquid Barreter.*

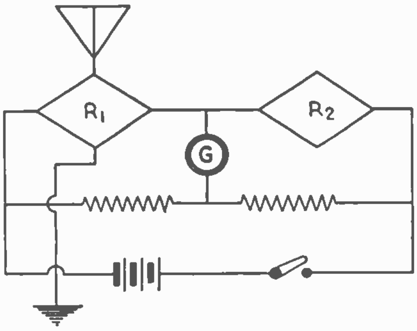

A thermal detector. By means of the change of resistance of a conductor, due to heating, electro-magnetic waves are detected. R_1 and R_2 are two rectangles of very fine iron wire.

THERMAL AMMETER—A current meter, in which the current or a fixed proportion of the current in question, passes along a fine wire. This heats the wire, causing it to expand or sag. This deflects a pointer or mirror across a calibrated scale. The thermal ammeter can be used for measuring direct or alternating currents. (See *Electrothermal Meter,* also *Hot-Wire Ammeter.*)

THERMAL DETECTOR—A radio *detector* (q.v.) which depends for its operation upon the heating of a fine wire by the passage of electrical oscillations. (See *Barreter.*)

THERMAL JUNCTION—THERMOELECTRIC JUNCTION—The contact *point* or joining point of the two dissimilar metals of a *thermo-electric couple* (q.v.). (See *Klemencic Thermal Junction.*)

THERMAL TELEPHONE—A telephone receiver in which the movements of the diaphragm are regulated by the variations of expansion of a wire heated by the telephone currents thus reproducing sound waves. In another type of thermal telephone, the diaphragm is dispensed with, sound waves being reproduced directly through the expansion and contraction of air in contact with the heated wire. This device is also known as a *thermal receiver* or a *hot-wire telephone.*

THERMIONIC EMISSION—The emission of a stream of negative electrons from a heated filament (cathode) in a vacuum tube.

THERMO-ELECTRIC COUPLE—THERMO-COUPLE—A pair of dissimilar metal pieces placed in contact and connected by electrical conductors in a closed circuit. When the junction of the two dissimilar metals is heated, an electric current will flow.

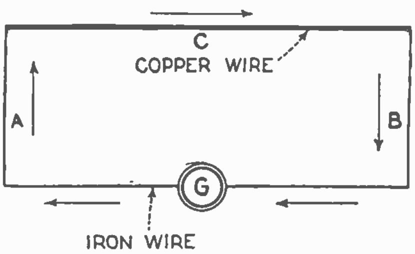

A current is caused to flow by heating the junction of two dissimilar metals. If the end at B is heated the direction of flow will be as shown by the arrows.

THERMO-ELECTRIC CURRENT—A current which is caused to flow due to an electromotive force set up at the junction of two dissimilar metals when these metals are at different temperatures. (See *Peltier Effect.*)

THERMOPILE—A device for magnifying the effect of the thermo-couple by connecting a number of these in series and exposing one set of alternate junctions to the heat, thus causing the electromotive forces to add up.

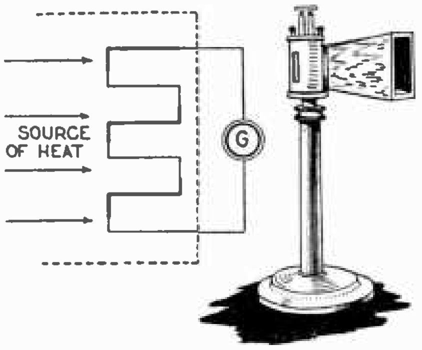

At the left is shown method of joining thermopile junctions to multiply the effect. At the right is a sensitive thermopile.

THOMSON EFFECT—The phenomenon of the appearance or disappearance of heat when a current flows from a cold towards a hot part of a conductor. If copper is unequally heated, heat is liberated at a point when the current and the heat flow in the same direction and is absorbed when they flow in opposite directions. The Thomson effect is measured by a quantity termed "the specific heat of electricity."

THORIUM—A rare metal which in certain compounds in radio-active. This metal has a specific gravity of eleven and an atomic weight of 232.5. Thorium is used for coating Welsbach mantles and is also used in connection with the coating of the tungsten filaments used in vacuum tubes. (See *Filament, Thoriated,* also *Thorium Treatment of Filaments.*)

THORIUM TREATMENT OF FILAMENTS—A compound of thorium, thoria (ThO_2) is used extensively in the manufacture of filaments for vacuum tubes. Sometimes thorium nitrate is used for this purpose. The filament so treated is termed thoriated tungsten. Since the rate of electron emission from a thoriated tungsten filament at a given temperature is several thousand times greater than in the case of the ordinary tungsten filament, it is possible to use the former at a much lower temperature. (See *Filament, Thoriated.*)

THREE CIRCUIT TUNER—A *regenerative circuit* (q.v.) in which the primary, the secondary or grid circuit, and the plate circuit are capable of being tuned in resonance with each other.

THREE-ELECTRODE THERMIONIC TUBE—An evacuated bulb or vessel containing a combination of a heated cathode, a relatively cold anode and a third electrode for controlling the current flowing between the other two electrodes. (See *Triode,* also *Vacuum Tube.*)

THREE-ELECTRODE TUBE—The three element thermionic vacuum tube having a filament or cathode, a plate or anode, and a grid. This tube can be used with minor modifications as a *rectifier* (q.v.) or *detector* (q.v.) of radio signals, an *amplifier* (q.v.), an *oscillation generator* (*oscillator* (q.v.)), an *electrostatic voltmeter* (q.v.), as a *power limiting device,* etc.

THREE WIRE SYSTEM—In direct current work, a system of transmitting electrical energy, which utilizes three wires, one of which is maintained at a potential, midway between the other two. A part of the load is connected between one wire and the neutral and part between the other wire and the neutral. As a result the neutral wire only carries current equal to the difference in the loads. The main advantages of this system are a saving in copper and availability of double voltage. Thus 110 volts may be required for lights, whereas 220 may be wanted for electric motors. In alternating current work reference is sometimes made to a two-phase three wire system. In this system there are two single phase currents which differ in phase by 90 degrees. Each current has a separate outgoing wire, but unites in a common return wire. (See *Neutral Wire.*)

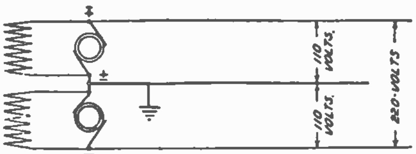

THE EDISON THREE WIRE SYSTEM

TICKER or TIKKER—A form of interrupter or an equivalent apparatus, used in sustained wave telegraphy. It is really a commutator interrupter. In a commercial type illustrated, a disk is mounted on a motor shaft. The disk has a number of teeth filled in between with fibre or other insulating material. The radio frequency currents flow from one brush to the other through the disk which interrupts them at the rate of from 300 to 1000 times per second. A charge is built up in the condenser C_a by resonance with the aerial system and this discharges into the telephone condenser C_b at regular intervals. The telephone condenser then discharges through the head set, thus producing a single sound corresponding to the charge accumulated. Since the tikker discharges the condenser, C_a at different places on the cycle of the incoming oscillations, the note produced is not uniform and is hard to read through atmospheric interference. (See *Chopper*.)

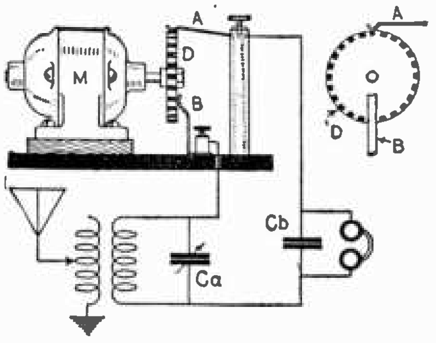

A simple Poulsen ticker. This is used to interrupt circuits of a receiving tuner at a uniform rate per second.

TICKLER—TICKLER COIL—A coil placed in the plate circuit of a regenerative receiver and coupled inductively in the grid coil in order to obtain regeneration. (See *Feed-Back, Feed-Back Coil, Feed-Back Coupling, Regenerative Circuit, Regenerative Coupling,* also *Regeneration.*)

TIME CONSTANT—The ratio, in an electric circuit, of the inductance in henries to the resistance in ohms. It is the time, in seconds, required for a current to attain 63.2 per cent of its ultimate value as given by Ohm's Law. (See *Impulse E.M.F.*)

TONE WHEEL—An apparatus used in the reception of wireless telegraph signals in continuous wave systems, utilizing a wheel in which the conducting teeth are separated by insulating material. A brush is used to make contact with the wheel. The tone wheel is similar in principle to the *tikker* (q.v.) except that it converts the incoming oscillations to audio frequency currents, thus doing away with the non-uniform note of the tikker. In addition to this the tone wheel utilizes more of the incoming energy than does the tikker. The tone wheel is run at such a speed that the frequency of the interruptions differs slightly from that of the incoming oscillations thus giving *beats* (q.v.) which produce audio frequency currents which can be heard in the telephone receiver. Since the tone wheel mechanically converts radio frequency currents into audio frequency currents it is sometimes called a *frequency transformer.*

TORQUE—The moment of a system of forces tending to produce rotation. It is usually measured in foot-pounds. Torque may also be defined as that which produces or tends to produce torsion. Thus if a force acts at a tangent to the periphery of a circle, thus producing a turning moment about the center of the circle, the product of the force and the radius of the circle is called the torque.

TRAIN RADIOPHONE—Communication by means of radio between moving trains and ordinary land-line telephone systems. Such a system is in operation commercially in Germany on trains operating between Berlin and Hamburg. The voice frequency currents from the telephone are impressed on a carrier wave by a method similar to the *Heising Modulation* (q.v.) system. Each train has a special wave length assigned to it. The modulated wave is emitted through a four-wire antenna, strung along the roofs of two coaches. The nearby telegraph wires pick it up and carry it to the receiving sets, located in stations along the line. The ground for the system is obtained through contact of the car wheels with the rails. In reverse order, the voice frequencies of messages from the telephone at the other end are used to modulate a carrier wave (using a different frequency) at the nearest train-telephone switchboard. They are then transmitted over the wires and inductively to the train. An arrangement is provided to prevent interference between the sending and the receiving sets.

TRANSFORMER—A type of induction apparatus, usually stationary, having primary and secondary windings, ordinarily insulated from each other. These are wound about an iron core. The function of the transformer is to change the electrical energy either increasing or decreasing the voltage, current, frequency or phase. Transformers can only be used in connection with alternating currents. Transformers for changing voltage have a different number of turns in the primary than in the secondary. In the step-up transformer, the primary usually consists of a small number of turns of comparatively heavy wire, while the secondary consists of a larger number of turns of fine wire. The ratio of the primary and secondary windings determines the ratio between the input and the output voltages; thus a 10 to 1 step-up transformer could be used to step a primary voltage of ten to a secondary voltage of 100. There are various methods of classifying transformers, namely, according to their operating characteristics, according to their method of cooling, or to their construction. Transformers which are meant to give a constant voltage on the secondary side are known as *constant potential transformers.* Those meant to give constant current are called *constant current transformers.* Transformers may be air cooled, water cooled or oil cooled. In some cases a natural air draft is used as in transformers of the smaller types. Larger transformers may be designed with special ducts through which blowers force a current of air. Oil cooled transformers may have the cores and windings immersed in oil, or the oil may be circulated through external coils. Water cooled transformers are immersed in oil, but have pipes carrying running water submerged in the oil. According to construction, transformers are of the core type and the shell type. The core type transformer is used more for high voltage, low capacity work, while the shell type is used for low voltages and high capacities. Transformers find a wide application in radio apparatus. *Audio transformers* are used to couple audio frequency circuits as for example in the case of an audio frequency amplifier. Audio transformers have an iron core and may be *shielded* or *unshielded. Radio frequency* transformers are used to couple radio frequency circuits. They are usually of air core type, although in special instances iron cores are used. Transformers are also used in alternating current chargers, in "A" and "B" battery eliminators and in power sets which operate directly from alternating current houselighting circuits. (See *Charger, Storage Battery, Audio Amplifier, Coupling Transformer, Jigger, Intermediate Transformer, Inductive Coupler, Oscillation Transformer, Air Core Transformer, Core Transfomer, Ratio of Transformation,* also *Shielded Transformer.*)

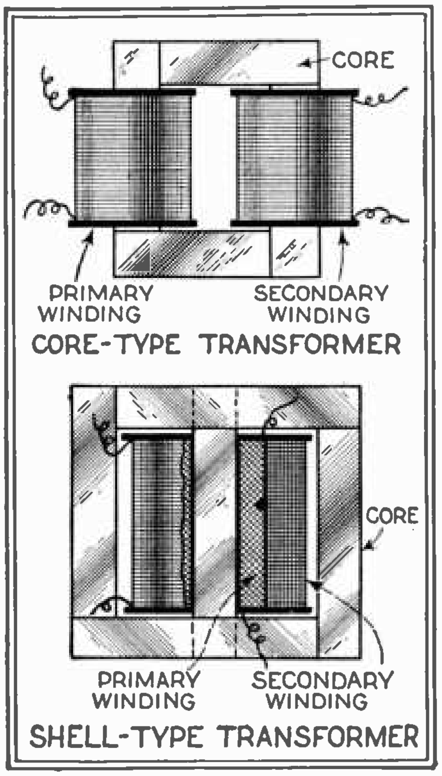

Two standard types of transformers.

TRANSFORMER COUPLING—A method of transferring the electrical energy of an alternating current from one circuit to another inductively by means of a transformer, one of whose windings is connected in one circuit, the other winding being connected in the second circuit. Radio frequency circuits are usually coupled by *radio frequency transformers, audio frequency transformers* being used to couple audio circuits. (See *Coupling,* also *Amplifier, Intermediate Frequency.*)

TRANSFORMER STEEL—A special steel usually containing a high silicon content used for stamping out transformer laminations. (See *Eddy Currents.*)

TRANSFORMER, TUNED—See *Resonance Transformer.*

TRANSMISSION—As applied to radio communication, this is the sending of signals, speech or music from one point to another by means of electromagnetic waves. (See *Radio Telephony, Radio Telegraphy, Theory of Propagation of Electromagnetic Waves.*)

TRANSMISSION OF PHOTOGRAPHS BY RADIO—Various methods have been devised and are now in use for the transmission of photographs by radio. Among these may be mentioned the systems of *Belin* (q.v.), *Baird*, and *Jenkins*. The principles underlying the Jenkins system are explained under the heading of *Television*. Using the system developed by Capt. R. H. Ranger, photographs were transmitted by radio from Honolulu to New York, a distance of 5,136 miles. Recently commercial picture transmission service has been inaugurated between New York and London using the Ranger apparatus. Two distinct methods have been applied for analyzing the picture in the process of transmission. One arrangement consists of producing an image as a non-conducting deposit upon a metal foil which is traversed by a stylus, while the other method makes use of an opaque image deposited upon a transparent film which is traversed by a beam of light, the light interruptions being recorded by a light sensitive cell. The Ranger system makes use of this latter method. The image is photographically recorded upon a celluloid sheet large enough to accommodate easily a picture of half-plate size. In the case of sketches and written messages the image for transmission is made directly by writing upon a piece of transparent film with a dense black ink. The image is then secured to the face of a glass cylinder, and by means of a lamp, focussing lens and reflecting prism, a narrow pencil of light is passed through the film. The cylinder is mounted on a carrier which is caused to be moved backward and forward so that the beam of light is concentrated in turn on all parts of the picture. A rotary motion is applied, as well as the transverse motion, the cylinder being given a slight rotation when it completes each transverse motion. The beam of light is passed through a special photo-electric cell. This consists of a spherical globe, coated on the inside with potassium hydroxide, which is very sensitive to light. The coating is connected to the grid of a vacuum tube, while an "electron collector" near the center of the tube is joined to the plate of the first amplifier. When no light is falling on the deposit on the inner surface of the globe, the grid acquires a negative charge, stopping the flow of electrons between filament and plate, and hence no current flows in the external circuit. The ray of light, however, causes an electron stream to flow between the coating and the collector, and since the coating is connected to the grid, the electron flow constitutes a discharged circuit, so that the grid becomes less negative. The first amplifying tube is a direct current potential amplifier, and is resistance coupled. The grid and plate connections of the amplifier are connected across a condenser which becomes discharged with the fall in the grid to plate resistance of the valve brought about by the grid potential fluctuations. A charging circuit is connected to the condenser and is controlled by a valve, the grid circuit of which operates by variations of the potential across the condenser. The charging current is fed through the plate circuit of this valve, in which a relay is connected, which working through other mechanical relays in cascades, controls the radio transmitter. Wave trains from the transmitting station after detection and amplification, are applied to the picture recorder. The recording mechanism, in order that it may be sensitive to exceedingly small currents, comprises, a small moving coil, in a magnetic field created by three electromagnets. The coil of wire, in moving in the field, as the received fluctuations are applied through its windings, operates a stylus while travelling across the surface of the paper. The stylus traverses the paper in perfect synchrony with the carriage of the transmitter, the paper being lifted each time the stylus completes a forward and backward movement across the paper. A small flashing neon lamp is used to indicate the correct speed adjustment of the driving motor.

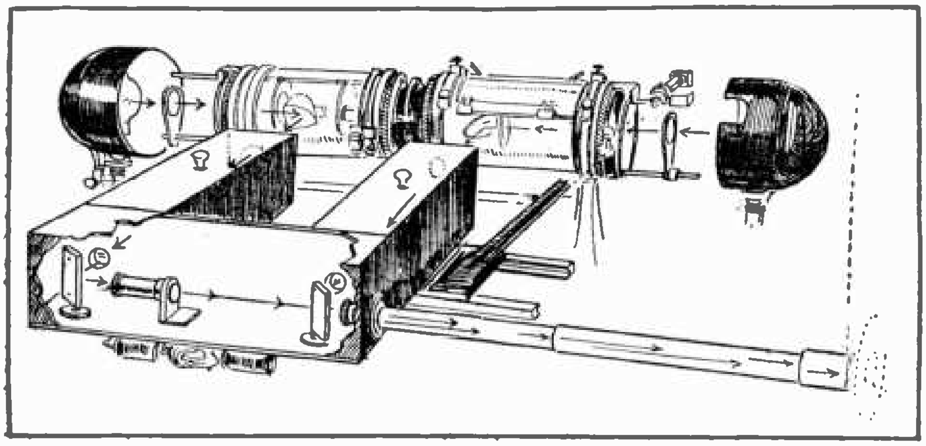

A pencil of light traverses the picture which is attached to the glass drums and is analyzed by a slow rotating action as well as a backwards and forwards movement of the carrier.

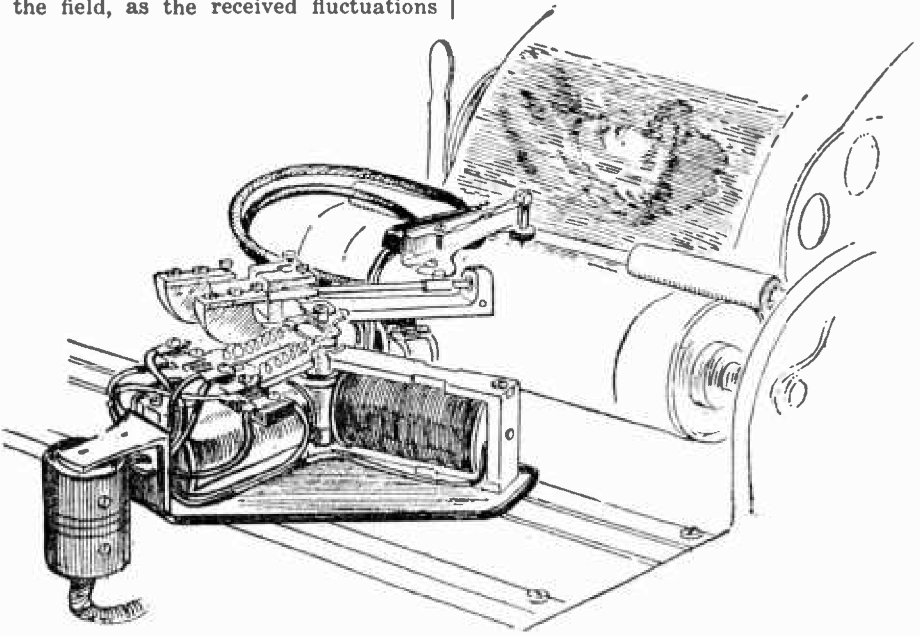

The recording mechanism of the receiver. Three electromagnets produce the magnetic field in which a moving coil controls the stylus.

TRANSMITTER—An apparatus for sending out electrical messages. As applied to radio telegraphy or radio telephony, the transmitter refers to the entire sending apparatus. The term transmitter is often used to refer to a *telephone transmitter* (q.v.). (See *Automatic Transmitter*, also *Wheatstone Transmitter*.)

TRANSMITTING AERIAL—A wire, or more usually a group of wires, suspended at a suitable height and connected to a radio transmitting set. The purpose of the aerial is to facilitate the radiation of the electromagnetic waves generated by the high frequency oscillating current which flows in the aerial. (See *Directional*, also *Receiving Aerial*.)

TRANSMITTING JIGGER—An *oscillation transformer* (q.v.) having a variable secondary, permitting of various degrees of coupling, by adjustment, between the two circuits. (See *Jigger*).

TRANSMITTING KEY—A telegraph key, used in the sending of radio code messages. This key must be of rather heavy construction since it handles larger currents than those used in ordinary wire telegraphy. Pressing the key, closes the circuit, and by holding the key down for a longer or a shorter period, the dots and dashes of the continental code are reproduced. (See *Code, Key, Key High Speed*.)

TRANSMITTING TUBE, POWER RATING—The useful power output from an oscillating tube is the power expended by the oscillating current in the resistance of the output circuit. The power input to the tube, not counting that expended in heating the filament, is the product of the plate supply voltage and the average plate current during an oscillation.

TREMBLER—A springy metallic blade carrying a soft iron armature. The spring makes contact with a fixed contact point, but when current passes through the electromagnet, the armature is attracted and the contact is broken. The spring then resumes its normal position, re-establishing the contact and the same process is repeated again and again. (See *Interrupter.*)

TRIGGER BATTERY—A term (seldom used) denoting a small battery inserted in the grid circuit to give the grid its initial charge when a tube is being used for radio transmission. In this connection, the battery is used to replace the potentiometer. (See *Grid Battery,* also *Grid Bias.*)

TRIODE—A name used to designate the three-electrode type of *thermionic tube.* Other names sometimes applied to the same device are *audion* (q.v.), *audiotron, aerotron, electron relay* (q.v.), *electron tube* (q.v.), *pliotron* (q.v.), *thermionic valve, oscillion* (q.v.), *vacuum tube* (q.v.), etc. (See *Three-Electrode Thermionic Tube.*)

TRUE POWER—The product of the *apparent power* and the *power factor* (q.v.) in an alternating current circuit. (See *Power, True,* also *Power, Apparent.*)

"T" TYPE AERIAL—See *"T" Antenna.*

TUBE OF FORCE—A theoretical conception, similar to *Line of Force* (q.v.) used in mathematical considerations of electrostatic or electromagnetic fields.

TUBE RECTIFIER—A vacuum tube or gas filled tube used to change or rectify alternating current to direct current. The principle of the vacuum tube rectifier is the fact that the electrons will flow from the hot filament to the plate, but not in the reverse direction. Tubes are made for half-wave or full-wave rectification. (See *Rectifying Tube.*)

TUBE, THREE ELEMENT—See *Triode,* also *Three-Electrode Thermionic Tube.*

TUBE, TWO ELEMENT—See *Fleming Valve.*

TUBING—In general, this refers to the bakelite or hard rubber forms used to wind inductance coils on. Another type of radio tubing is defined under *Spaghetti.*

TUDOR ACCUMULATOR—A storage battery having special positive plates formed by the *Plante* process and negative plates of the *Faure* or pasted type. The Tudor positive is of the cast-lead type. This plate has no central web and is made by casting pure soft lead in a mold. Casting is advantageous since it permits the distribution of the metal in the plate without limitations in the manufacturing process. After the plate is removed from the mold, its surface consists of a large number of short vertical ribs which run entirely through the plate. These are bound together by short transverse ribs for the purpose of giving strength to the plate. This gives a large surface and since there is no central web, the electrolyte is enabled to circulate through the plate thus working the active material uniformly.

TUNED ANTENNA—A transmitting aerial, having a loading inductance or a variable condenser or both in series with it, for the purpose of changing the antenna wave length.

TUNED IMPEDANCE—This refers especially to a circuit, or to a method of radio frequency amplification in which a tuned inductance connected in parallel with a variable condenser is used as the method of coupling rather than transformer or resistance coupling. (See *Impedance Coupled Amplifier.*)

TUNED RADIO FREQUENCY—See *Tuned Radio Frequency Circuit.*

TUNED RADIO FREQUENCY AMPLIFIER—A combination of a *radio frequency transformer,* a variable condenser and an amplifying vacuum tube in a radio receiving circuit for the purpose of increasing the intensity or amplifying the radio frequency signals.

TUNED RADIO FREQUENCY CIRCUIT—A radio receiving circuit employing radio frequency amplification in which the radio frequency amplifier circuits may be tuned to the desired wave lengths by varying the inductance or the capacity, or both although the usual method of tuning is by means of a variable condenser in parallel with the secondary of the radio frequency transformer. A typical five-tube tuned radio frequency circuit is shown in the accompanying illustration. This consists of two stages of tuned radio frequency amplification, a detector and two stages of audio frequency amplification.

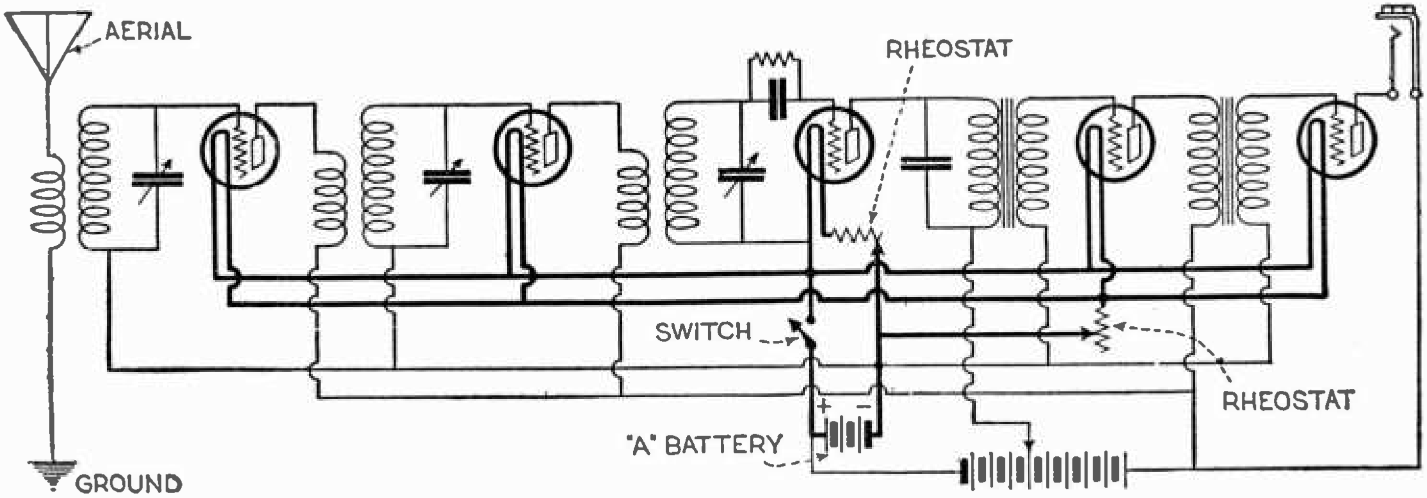

A standard five tube tuned radio frequency circuit.

TUNED RADIO FREQUENCY TRANSFORMERS—An air core *transformer* (q.v.) shunted by a variable condenser used in a *tuned radio frequency circuit* (q.v.). (See *Fixed Coils,* also *Low Loss Coils.*)

TUNED TELEPHONE—One in which the diaphragm is adjusted to vibrate at the same frequency as the current impulses to be indicated. In addition the telephone circuit may have its inductance and capacitance chosen to have the same electrical frequency, and a tuned acoustic resonator may intervene between the diaphragm and the ear.

TUNED TRANSFORMER—See *Resonance Transformer.*

TUNGAR—A trade name for a two-electrode vacuum tube especially designed for rectifying purposes. Tungar tubes are made in two ampere and five ampere capacities. They are used in *chargers* (q.v.) and also in radio sets operating directly from the lighting socket without the use of batteries. (See *Rectifying Tube.*)

A typical rectifier tube of the Tungar type. The screw-base allows it to be inserted in a standard lamp socket from which current is obtained for the filament. Connection is made to the graphite disc by means of a Fahnestock connector which clips on to the wire post projecting from the other end of the tube.

TUNING—The process of regulating the inductance and the capacity of a radio circuit in order to be in unison with a desired wave length.

TUNING, BROAD—Reception of a signal over a wide range, that is on a number of wave lengths, rather than on a single one. A receiving set which tunes broadly cannot be used where the broadcasting stations are transmitting on wave lengths near to one another. Too long an aerial may result in a receiving set tuning broadly. (See *Selectivity.*)

TUNING COIL—A coil of wire so arranged that each turn is electrically insulated from its neighbor and with a device by which contact may be made to bring any desired number of turns into the circuit so as to increase or decrease the inductance of a radio circuit so as to tune it to the desired wave length. A variable *inductance* (q.v.) used for tuning an oscillatory circuit. (See *Lengthening Coil.*)

T.Y.K. ARC—A transmitting arc used for radio telephony invented by W. Torikata, E. Yokoyama and M. Kitamura of Japan. This arc uses magnetite and brass electrodes instead of copper-carbon electrodes. A circuit diagram of the T.Y.K. system is shown in the illustration. The materials used for the electrodes are such that a high resistance film forms on their surfaces. This requires a temporary high voltage to start the discharge. This is accomplished as follows: An armature, A, is attached to one of the electrodes. The two electrodes are in contact, and a steady current is caused to flow in the circuit, through the spark induction coil B. This attracts the armature A, drawing the electrodes apart. The break of the spark-coil current at its interrupter induces a high electromotive force which, acting through the coil and the spark gap in series, breaks down the film on the gap electrodes. The power supplied to the gap is 500 volts and 0.2 ampere. A condenser of approximately .05 mfd. is used in the primary oscillating circuit.

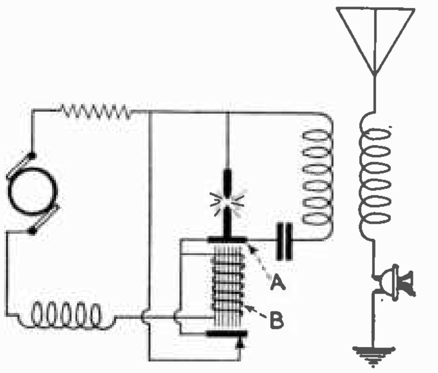

Circuit diagram of the T. Y. K. arc.

T.Y.K. sets have been used only in small units, having a range of 30 or 40 miles.

Several unusual features appear in the circuit shown in the accompanying illustration, which gives the circuit diagram of the T.Y.K. arc.

U

ULTRA-AUDION—A name sometimes applied to the vacuum tube used in connection with *beat* reception for supplying local oscillations. In general, however, the name ultra-audion applies to a circuit for radio reception. (See *Ultra-Audion Circuit.*)

ULTRA-AUDION CIRCUIT—A type of circuit used for long wave radio reception which uses a form of regeneration, without calling for the introduction of auxiliary equipment in the circuit. A typical ultra-audion circuit is illustrated. It will be noticed that while one terminal of the secondary goes to the grid, the other terminal, which usually goes to the filament, in this case goes to the plate. Thus there is a direct plate-filament circuit which does not involve the plate and filament batteries in the usual manner.

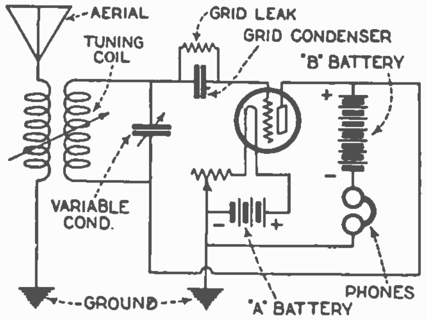

Diagram of an ultra-audion regenerative circuit.

ULTRADYNE—A modification of the *super-heterodyne circuit* (q.v.), which uses the modulation method to produce beats. In this method the incoming signal is caused to modulate the oscillations produced locally in a similar way to that in which the speech or music modulate the carrier wave of a broadcast station. It is claimed that this method is simpler than the ordinary super-heterodyne and also more sensitive to weak signals.

The circuit of the improved model L-2 Ultradyne, incorporating regeneration is shown below. The long wave intermediate transformers are accurately tuned to a wave length of 3000 meters. Referring to the diagram, it will be noted that a special form of oscillator, comprising a grid and plate coil, are used together with an oscillator tuning condenser connected across the grid coil. The first tube, which is usually the detector in the standard super-heterodyne circuit, is known as the modulator tube. (See *Supcr-Heterodyne Receiver.*)

UMBRELLA ANTENNA—An antenna, the conductors of which form elements of a cone with the apex at the top to which the *down lead* (q.v.) is connected.

UNDAMPED—See *Undamped Alternating Current,* also *Undamped Waves.*

UNDAMPED ALTERNATING CURRENT—Periodic current (i.e., current passing through successive equal cycles of values) whose average value is zero.

UNDAMPED OSCILLATIONS—Oscillations which are sustained. *Undamped alternating current* (q.v.). Oscillations such as are generated by a *vacuum tube oscillator* or by an *arc generator.*

UNDAMPED WAVES—Continuous waves. Waves in which similar current cycles follow each other continuously instead of being broken up into groups. (See *Sustained Waves, Decadent Wave,* also *Damped Waves.*)

UNITS—Specified amounts of physical quantities used as a basis of measurement. (See *Practical Units.*)

UNITS, ELECTROMAGNETIC—See *Electromagnetic Units.*

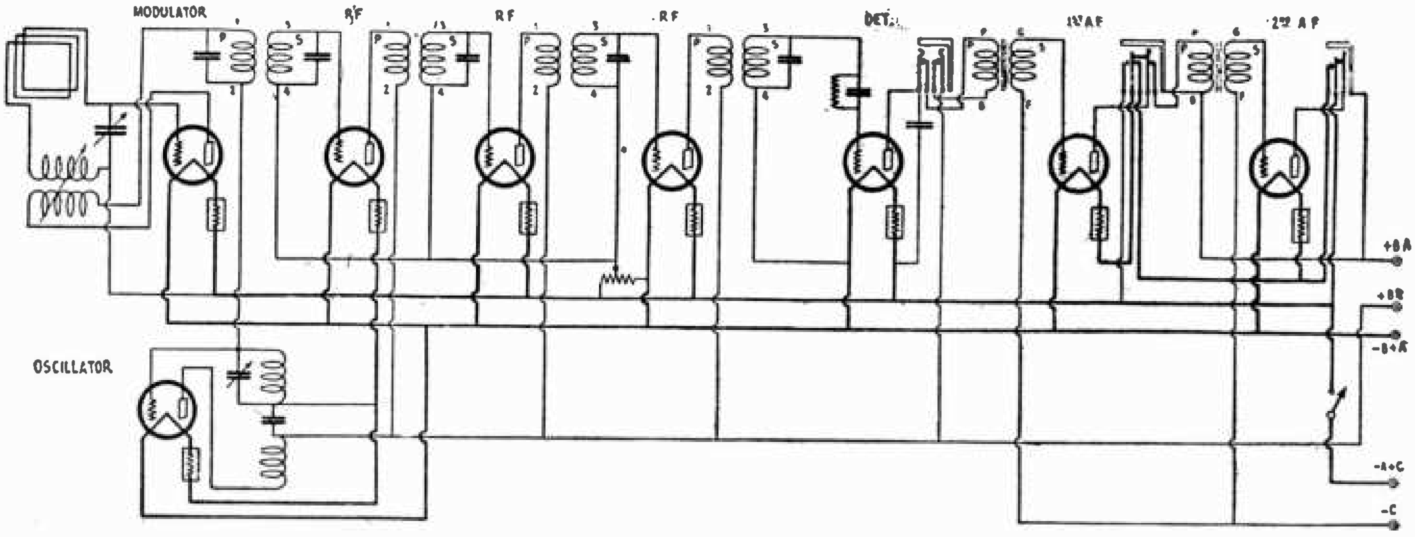

This shows the hook-up of the Ultradyne super-het. This set is not radically different from the average super-heterodyne. The heteroformers are of the tuned air core type and the filter is shown at the input side of the circuit. Aerial and ground as well as loop may be used with this set and instead of the first detector the inventor uses a tube known as the "modulator."

V

VACUUM TUBE—An evacuated bulb, or one containing a rare gas and having three elements, filament, plate and grid. (See *Three-Electrode Thermionic Tube, Triode,* also *Dynatron.*)

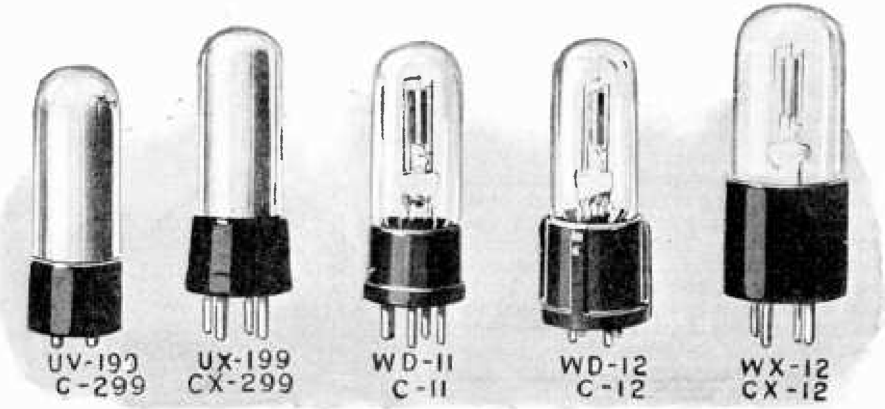

A group of dry-cell vacuum tubes. The first two differ only as to the type of base; both are intended to be operated from a 4½-volt "A" battery, that is, three dry-cells connected in series. The filaments draw very little current, as can be determined from the chart. The last three tubes also have identical characteristics, but are designed to fit different types of sockets. These tubes operate from a single 1½-volt dry cell and draw approximately ¼ ampere of current. The filaments are oxide-coated and very rugged.

VACUUM TUBE AMPLIFIER, THEORY OF—The vacuum tube may be used as an amplifier or magnifier because of the fact that under proper conditions, a small variation of the electromotive force applied between the filament and the grid produces a very large variation in the current flowing in the plate circuit. It is usual to use several amplifying tubes together, suitably coupling the plate of one tube to the grid of the next, thus increasing the amplification to the desired value.

VACUUM TUBE CHARACTERISTICS Performance curves of vacuum tubes. These are obtained experimentally by keeping the filament current constant and applying various known voltages between the plate and the filament and between the grid and the filament and reading the resultant currents that flow to the plate and the grid. The various values obtained are then plotted on cross-section or graph paper. Such curves are convenient for determining the operating characteristics of a particular vacuum tube under consideration. (See *Characteristic Curve.*)

VACUUM TUBE GENERATOR—The vacuum tube can be used to produce sustained oscillations without the necessity of supplying potential variations to the grid from an external source. This can be accomplished as a direct result of the amplification properties of the vacuum tube, since the energy of the output circuit is greater than the energy of the input circuit and part of this output energy may be returned to the input to produce constant reamplification.

VACUUM TUBE, POWER—A tube of special design, made for use in the last audio stage of a radio receiving set, which will handle greater current than the ordinary amplifying tube, thus resulting in more volume and at the same time clearer and more satisfactory tone reproduction. Power tubes now on the market require higher plate voltages than conventional tubes, ranging from 135 volts upwards. The amount of grid bias or "C" battery ranges from approximately 4½ volts to 45 volts, depending upon the type of power tube. Tubes suitable for developing considerable power, in radio transmission are also referred to as power tubes.

By increasing the surface area and hence the filament emission or by increasing the degree of evacuation of the bulb, thus allowing a higher plate voltage, the output of the tube may be increased. A form of transmitting power tube is described under the heading *Pliotron.* The illustration below shows one of the new power tubes used for reception.

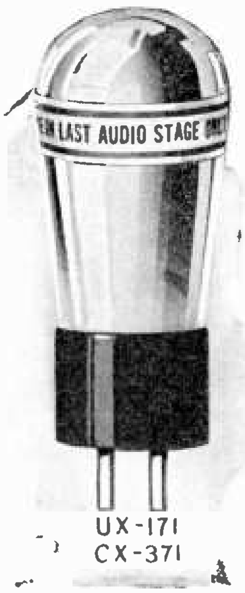

A new storage-battery type power amplifier tube, for use in the last stage audio amplifier only. This tube takes the same filament voltage and filament current as the No. 112 tube but employs a "B" voltage of 180 and 40½-volt "C" battery. Special means should be provided for handling its heavy output.

VACUUM TUBE RECTIFIER—A vacuum tube designed or used to change alternating current to direct current. (See *Rectifying Tube, Kenotron,* also *Tube Rectifier.*)

VACUUM TUBE, THEORY OF OPERATION—See *Theory of Operation of Vacuum Tubes.*

CHART OF VACUUM TUBE TYPES AND CHARACTERISTICS

TYPE	USE	"A" battery volts (Supply)	Filament Terminal Volts	"A" Battery Current Amperes	"B" Battery Volts		Negative "C" Battery Volts	Voltage Amplification Factor	Output Resistance Ohms
					Det.	Amp.			
UV. UX-199 C. CX-299	Det. or Amp.	4.5	3	.06	45	90	4.5	6.25	15,000
UV. UX-200 C. CX-300	Det. only	6	5	1.0	16 to 22½				
UX-200-A CX300-A	Det. only	6	5	.25	Max. 45				
UV. UX.201-A C. CX. 201-A	Det. or Amp.	6	5	.25	45	90 to 135	4.5 9.0	8 8	12,000 11,000
UX-120 CX-220	Pow. Amp. (Lst. stg. only)	4.5	3	.125		135	22.5	3.3	6,600
UX-112 CX-112	Det. or Amp.	6	5	.5	22½ to 45	135	9.0	Var.	Var.
UX-171 CX-371	Pow. Amp. (Lst. stg. only)	6	5	.5		180	40.5		
UX-210 CX-310	Pow. Amp. Oscillator	6	6	1.1		90 to 425	4.5 to 35	Var.	Var.
WD-11 C-11	Det. or Amp.	1.5	1.1	.25	22½	90	4.5	5.6	14,000
WD. WX-12 WX. CX-12	Det. or Amp.	1.5	1.1	.25	22½	90	4.5	5.6	14,000
3VB-199 3VBX-199	Det. or Amp.	4.5	3	.06	20	80	4.5	6.0	
3V-A 3VAX	Det. or Amp.	4.5	3	.12	20	90	4.5	6.5	
5V-A 5VAX	Det. or Amp.	6	5	.25	20	100	4.5 to 9.0	9.4	9,400
5VC 5VX	Pow. Amp. or Det.	6	5	.5	22½	90 to 157½	6 to 10.5	8.6	5,900
99 99X	Det. or Amp.	4.5	3	.06	22½	90 to 150	3 to 12		
O1A O1X	Det. or Amp.	6	5	.25	22½	90 to 150	3 to 12		
MU-20	Audio Amp.	6	6	.25		90 to 150	4.5 to 10.5	20	40,000
MU-6	Pow. Amp. (Lst. stg. only)	6	6	.5		90 to 150	4.5 to 10.5	6	5,000
B-6	Det. only	6	5	.25	16 to 22½				
A	Det. or Amp.	6	5	.25	20	120	4.5 to 9		
B C	Det. or Amp.	4.5	3	.06	20	80	4.5		
E	Pow. Amp. (Lst. stg. only)	4.5	3	.125		135	22.5		
F	Pow. Amp. (Lst. stg. only)	6	5	.5		90 to 180	4.5 to 9		
G	Audio Amp.	6	5	.25		90 to 180	4.5 to 9		
DC—	Det. or Amp.	4.5	3	.06	45	90	4.5	6.3	16,500
DC—	Det. or Amp.	6	5	.25	45	90	4.5	8.5	10,000
DC—	Pow. Amp. (Lst. stg. only)	4.5	3	.125		112 to 135	13 to 22.5	3.3	6,300
DC—	Pow. Amp. (Lst. stg. only)	6	5	.5		90 to 157.5	6 to 10.5	8.0	8,500

VACUUM TUBES, TYPES OF—Vacuum tubes are of various types and designs depending upon the work which they must do. They may be divided into two general classes, receiving tubes and transmitting tubes.

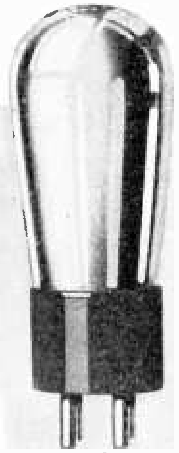

A new type of detector tube which has greater sensitivity than former types, due to the use of an alkaline vapor. It can be used as a detector in any type of receiving circuit.

Tubes for receiving purposes are designed either for dry cell operation or else are of the storage battery type.

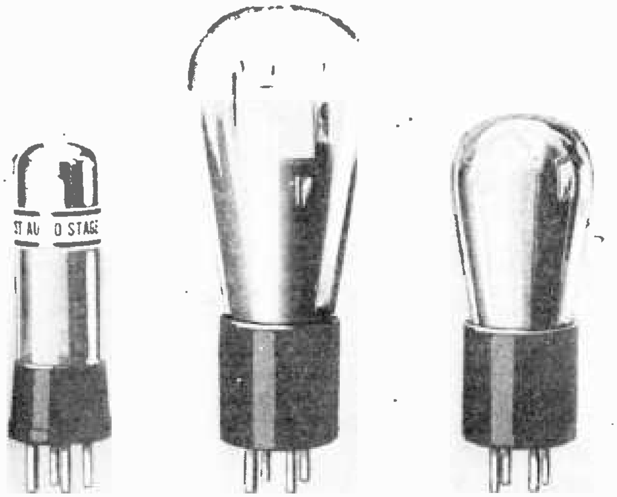

These power amplifier tubes are designed for use in the last stage of an audio frequency amplifier. On the left is a dry cell tube which requires a 4½-volt "A" battery. The center tube employs exceptionally high "B" battery voltage, and is usually employed with an amplifier that operates from the house current. This tube can also be used for transmitting purposes. The right-hand tube is of the storage battery type and has a very rugged filament. It makes an excellent detector tube as well as a power amplifier.

They may be further classified as detector tubes, amplifying tubes or power tubes. There are many special types of tubes, such as *kenotrons* (q.v.), *pliotrons* (q.v.), etc.

VALENCY—The property possessed by elements in combining with or replacing other elements in a certain definite proportion. Also referred to as *valence*. (See *Electro-Chemical Equivalent.*)

VALVE—This generally refers to a vacuum tube used as a detector or rectifier. The name *valve* is used in many foreign countries to refer to the thermionic vacuum tube, as used for any purpose.

VALVE DETECTOR—See *Valve.*

VARIABLE CONDENSER—A condenser having rotable or movable metal plates and an air dielectric. Both the stationary and rotating plates may be made of brass, copper or aluminum. Bakelite, fibre or composition are used to insulate the stationary from the movable plates. Turning the movable plates either increases or decreases the capacity of the condenser. Variable condensers are primarily used in radio for *tuning* (q.v.). They are constructed so as to have straight line frequency characteristics either by undercutting a portion of the movable plates or else by making these of a variable thickness.

VARIABLE GRID LEAK—A variable high resistance unit designed to be placed in the grid circuit of a vacuum tube used as a detector. Certain types of variable grid leaks may have their resistance varied from ¼ megohm to 10 or more megohms by merely tightening a thumb-screw. (See *Grid Leak*, also *Adjustable Grid Leak.*)

VARIABLE RATIO TRANSFORMER—A transformer in which the ratio between the primary and the secondary windings can be varied, usually by means of suitably located taps.

VARIABLE RESISTANCE—Both the *rheostat* and the potentiometer used in radio work, are variable resistances. Variable resistances are usually of high resistance wire type, with a sliding contact so that the resistance can be easily altered.

VARIO COUPLER—A tuner formerly used in radio receiving sets, having a primary and a secondary coil inductively coupled, with the secondary arranged so as to be rotable within the primary. The amount of coupling depends upon the variation of the angle between the axes of the two coils.

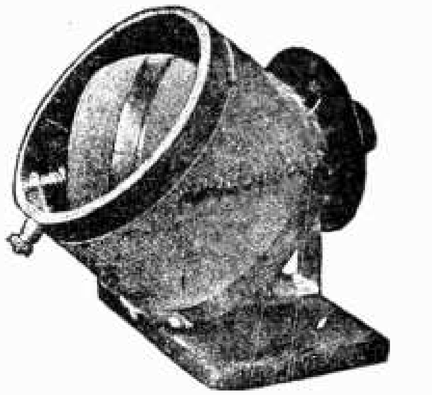

A standard form of vario coupler.

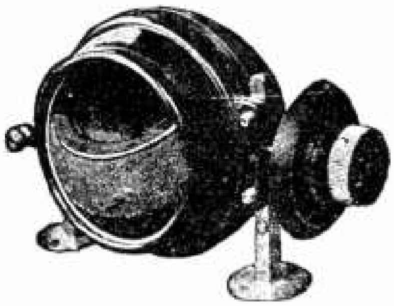

A variometer.

VARIOMETER—A tuner, similar to the vario-coupler, except that the primary and secondary coils are electrically connected.

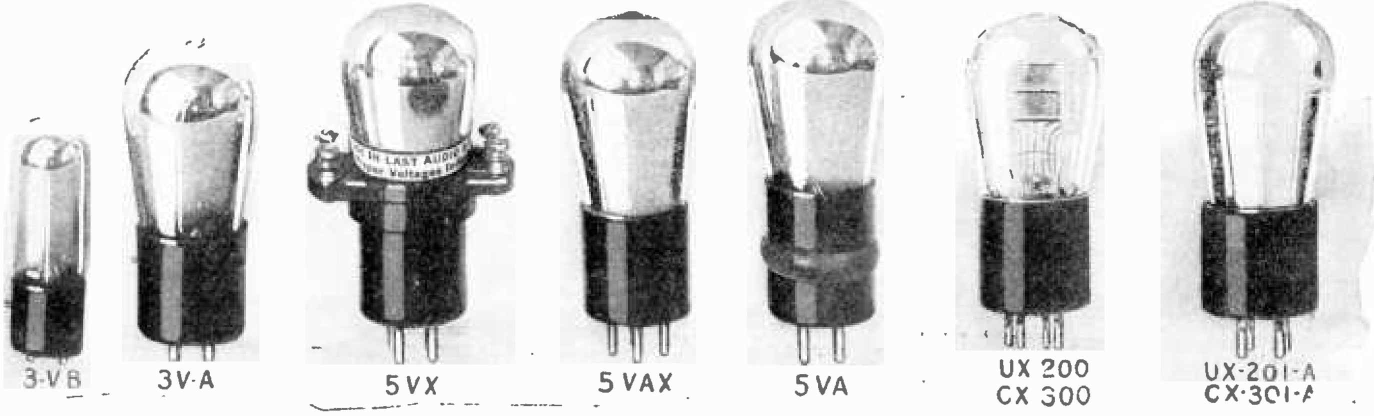

The characteristics of these vacuum tubes are given in the vacuum tube chart on page 187. The first tube is of the dry-cell type, requiring a 4½-volt "A" battery for lighting the filament. 3V-A is a standard 5-volt tube. 5VX is a power amplifier which can be used in the last stage of any audio-frequency amplifier. There is no necessity of changing any of the wiring in the set for the addition of more "B" battery, as binding posts are included on the tube itself. 5VA has a sponge rubber ring included as a part of the base which tends to absorb all vibrations which might otherwise cause the tube to become noisy. The next type of tube is for use as a detector only. The last is the well-known 201-A type, which can be employed as a detector or an amplifier.

VECTOR—A graphical illustration used in mathematical calculations, consisting of a line with an arrow-head at one end, used to show by means of its length, direction and the angle between itself and another vector or vectors, the magnitude, direction and phase angle of alternating current quantities.

VELOCITY—The distance passed through in a certain time. Velocity is measured in feet per second, miles per hour, etc.

VIBRATION—A to and fro motion. An oscillating or swinging motion.

VITREOUS—Consisting of or pertaining to glass. (See *Resinous Electricity*.)

VOLT—The practical unit of *electromotive force* (q.v.). The volt is a measure of the electromotive force which will cause a current of one ampere to flow through a resistance of one ohm. It is equal to 10^8 absolute electromagnetic units (Abvolts).

VOLTAGE—The electrical pressure or the electromotive force between two points in an electrical circuit, measured in volts.

VOLTAGE, CONSTANT—A steady unvarying electrical pressure or electromotive force, as differentiated from a pulsating or fluctuating voltage.

VOLTAGE DROP—The fall in potential caused by the resistance of the conductor through which the current is flowing. The longer the conductor, the greater will be the voltage drop. The greater the cross-section of the conductor, the less will be the voltage drop.

VOLT-AMPERES—The product of volts as measured with a voltmeter, by the current as shown on an ammeter, in an alternating current circuit, gives the apparent power, or in other words the apparent watts. (See *Instantaneous Values*.)

VOLTMETER—An instrument for measuring voltage drop or potential difference. Most of the voltmeters used in radio work operate on the moving coil principle. (See *Galvanometer*.)

A plug-in type of voltmeter.

Panel-type of voltmeter.

VOLTMETER CALIBRATION—A method of correctly marking the voltage readings on the scale of a voltmeter or of checking up the voltmeter readings to make sure that they are accurate. The usual method of calibration is to check the voltmeter reading by comparison with a standard voltmeter. (See *Calibration*.)

V.T.—Abbreviation for *vacuum tube*.

VULCANITE — (EBONITE) — A hard black substance produced by vulcanizing rubber with about 25 per cent sulphur. It is readily polished and has excellent insulating properties.

VULCANIZED RUBBER—Pure rubber, mixed with five per cent sulphur and baked at a temperature of about 150 degrees Centigrade. Vulcanized rubber must not be allowed to come into contact with copper, which is chemically attacked by the sulphur.

W

WALL TUBE—See *Partition Insulator*.

WATCHCASE RECEIVER—A telephone receiver having a shape somewhat similar to a watch. Watch case receivers are compact in construction and are the usual type used in the radio telephone *head set*.

WATT—The unit of electric power. One *Joule* (q.v.) per second. To find power in a direct current circuit, multiply voltage by amperage. The *Kilowatt* (q.v.) equals one thousand watts. Seven hundred and forty-six watts are equivalent to one electrical horsepower.

WATT HOUR—The commercial unit of electrical work. The work done in one hour by a current of one ampere flowing between two points of a conductor having a difference of potential of one volt.

WATT-HOUR METER—An integrating meter which measures energy in watt-hours or in kilowatt-hours. (See *Integrating Wattmeter*.)

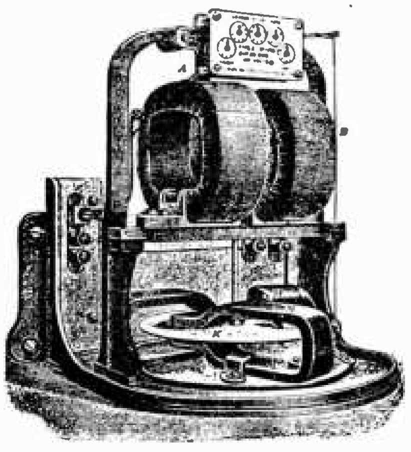

A direct current watt-hour meter.

WATT, INTERNATIONAL—symbol W —The energy expended per second by an unvarying electric current of one International ampere under an electrical pressure of one International volt. IW = 1 Joule per second = 10^7 ergs per second.

WATTLESS COMPONENT—WATTLESS CURRENT—The component of an alternating current which is in quadrature with the voltage. (See *Reactive Component*.)

WATTMETER—An instrument designed to measure the power being expended in an electrical circuit. The most common form of direct current wattmeter is of the dynamometer type. This has two coils, one fixed and one movable, one coil being connected so as to exert a force proportional to the current flowing in the circuit, and the other coil exerting a force proportional to the electromotive force. The *induction* type of wattmeter is in general use for alternating current circuits. (See *Electro-Dynamometer*, *Dynamometer*, also *Induction Wattmeter*.)

WATT SECOND — A unit of electrical energy, representing the energy expended by one watt flowing for a second. The watt second is the same as the *Joule* (q.v.) and is usually applied to the measurement of heat developed by an electric current.

WAVE—A periodic alternation of an alternating current. As applied to an electric or electromagnetic disturbance, an electric wave is an undulatory movement of the ether, radiated from conductors carrying electrical oscillations.

WAVE ANALYSIS — A study of the wave form of an alternating current or other type of wave. The wave form of commercial alternating current used for lighting purposes approximates a *sine curve* (q.v.). A convenient instrument for the analysis and study of alternating current wave form is known as the *oscillograph* (q.v.).

WAVE ANTENNA—A horizontal antenna, the physical length of which is approximately equal to the length of signaling waves to be received, and which is so used as to be strongly directional.

WAVE DISTORTION—Alteration in the wave form of a wave after it has traversed a considerable distance. (See *Distortion*.)

WAVE FORM—Referring to an alternator, the "shape" of the curve of the current generated.

WAVE FREQUENCY—See *Frequency*.

WAVE LENGTH—symbol λ (lambda) —The distance between two successive *antinodes* (q.v.) in the same direction. In referring to the wave length of electrical oscillations in a circuit, this means the length of the waves in free space that would have a frequency corresponding to the given oscillations. Electromagnetic waves used in radio work have frequencies of from 10,000 to 3,000,000 cycles per second. Since these waves have a velocity of approximately 300,000,000 meters per second, it is possible to calculate the length of a wave by dividing the velocity by the frequency. In other words meters per second divided by cycles per second gives length of the wave in meters. Thus a wave having a frequency of 100 kilocycles (100,000 cycles) will have a wave length of 300,000,000 divided by 100,000 or 3,000 meters.

WAVE LENGTH ALLOCATIONS — In order to prevent radio transmitting stations from interfering with each other, or creating interference at a receiving station, various wave length bands have been allotted to each type of service. The following table shows the present short wave assignments and the service for which they are used:

Wave-length in Meters	*Service*
109–105	Relay broadcasting only.
105–85.7	Public toll service, Government mobile, and point-to-point communication by electric power supply utilities, and point-to-point and multiple address message service by press organizations only.
85.7 –75.0	Amateur, Army mobile, naval aircraft, and naval vessels working aircraft only.
75.0 –66.3	Public toll service, mobile Government point-to-point and point-to-point public utilities.
66.3 –60.0	Relay broadcasting only.
60.0 –54.5	Public toll service only.
54.5 –52.6	Relay broadcasting only.
52.6 –42.8	Point-to-point only.
42.8 –37.5	Amateur and Army mobile only.
37.5 –33.1	Public toll service, mobile, Government point-to-point, and point-to-point public service utilities.
33.1 –30.0	Relay broadcasting only.
30.0 –27.3	Public toll service only.
27.3 –26.3	Relay broadcasting only.
26.3 –21.4	Public service, mobile, and Government point-to-point.
21.4 –18.7	Amateur only.
18.7 –16.6	Public toll service, mobile and Government point-to-point.
16.6 – 5.35	Experimental.
5.35 – 4.69	Amateur.
4.69 – 0.7496	Experimental.
0.7496– 0.7477	Amateur.

Radio broadcast stations work within the wave length band of 200 to 545 meters at intervals of 10 kilocycles. 600 meters is reserved for distress signals. The higher wave lengths are allocated to various marine, government, university, aircraft and other uses.

WAVE LENGTH CALCULATION FOR ANTENNA—For the ordinary vertical wire grounded antenna, the fundamental wave length is slightly greater than four times the length of the wire. A constant suggested is 4.2 and applies approximately also to flat top antennae having vertical lead ins. For other calculations see *Fundamental Wave Length.*

WAVE LENGTH, FUNDAMENTAL—See *Fundamental Wave Length.*

WAVE LENGTH, NATURAL—In a loaded antenna (that is with series inductance or capacity) the natural wave length corresponds to the lowest free oscillation. (See *Natural Wave Length.*)

WAVE LENGTH OF STATIONARY S. H. M. WAVES—On a straight wire, the smallest distance between two points where the disturbance is of the same amplitude and phase; or, since consecutive loops (q.v.) are in opposite phases, the wave length is double the distance between consecutive loops or consecutive *nodes* (q.v.). (See *S.H.M.*)

WAVEMETER—A radio instrument for measuring frequency. A calibrated resonator of variable frequency, capable of easy excitation by the oscillation under test, combined with a means of indicating the attainment of resonance. The calibration may be given directly in wave lengths or frequencies or both.

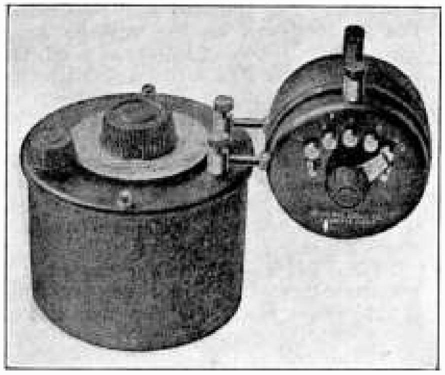

Photo by courtesy of General Radio Co.

The selectivity of the radio set may be greatly increased by the use of this wavemeter as a radio filter.

WAVEMETER CALIBRATION—Wavemeters may be calibrated by comparison with a standard wavemeter. An accurate method of calibrating a wavemeter, known as the *Three-Way Method* is shown in the accompanying diagram. In this illustration A is the standard wavemeter, B the wavemeter to be calibrated, and C is a condenser and coil of size permitting resonance with A. To calibrate, the buzzer should be put in operation, varying the capacity of the condenser C^1 until L^1C^1 is in resonance with L^2C^2 as noted in the head set. The wavemeter B should then be put in inductive relation to C and the condenser C^2 varied until the wavemeter A resonates with C. The wave length of B is now the same as A. This procedure should be carried out over the whole range of the condenser C^3.

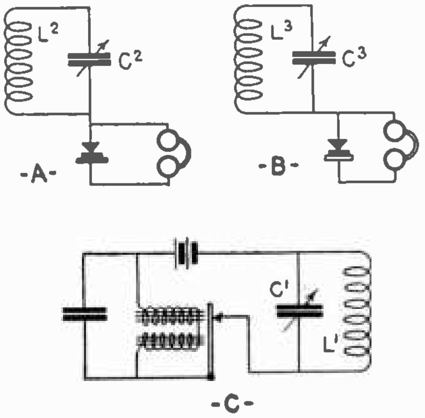

Diagram showing three-way method of wavemeter calibration.

WAVE MOTION—A disturbance of the equilibrium of a medium or body, extended or propagated from point to point with a continuous motion, each particle vibrating only on each side of its position of equilibrium, while each phase of vibration moves onward. Examples of wave motion are the propagation of electromagnetic waves, or the waves on the surface of a body of water.

WAVES, CONTINUOUS—See *Continuous Waves.*

WAVES, CONTINUOUS, KEY MODULATED—See *Continuous Waves, Key Modulated.*

WAVES, CONTINUOUS, MODULATED AT AUDIO FREQUENCY—See *Continuous Waves at Audio Frequency.*

WAVES, ELECTROMAGNETIC—Periodic electromagnetic disturbances progressive through space. (See *Wave.*)

WAVES, SUSTAINED—See *Sustained Waves,* also *Continuous Waves.*

WAVE TRAIN—A series of waves. An example of a wave train is the series of disturbances set up by each spark in a spark system of radio telegraphy.

WAVE TRAP—A device consisting of a variable condenser and coil, usually connected in the antenna circuit of a receiving set, to eliminate interference from undesired stations. (See *Filter, Filter Condenser, Acceptor,* also *Rejector.*)

WAVE-WOUND—A class of drum armature in which connections produce a "stepping forward" in a zig-zag wavy line all the time. Also called *Series* and *Two Circuit Windings.*

WEBER—The *Practical Unit* (q.v.) of magnetic flux. The engineers' unit is the "line" which is also the electromagnetic unit. 1 Weber is equal to 108 electromagnetic units (or "lines"). The Weber is produced by a current of one ampere flowing through a circuit having one henry of inductance.

WEHNELT BREAK—An *electrolytic interrupter* (q.v.).

WEIGHT—The force with which the earth attracts a body. A body varies in weight according to its distance from the centre of the earth, although its mass remains constant.

WESTON CELL—The present standard cell. It consists of mercury with a paste of mercurous and cadmium sulphates which form the cathode. The Weston cell has an anode of 12.5 per cent cadmium amalgan in an electrolyte of saturated solution of cadmium sulphate. It has a constant electromotive force of 1.0125 volts at 20 degrees Centigrade. (See *Electromotive Force,* also *Cadmium.*)

WHEATSTONE BRIDGE—An instrument for measuring resistance in which the current from a battery divides into two parallel circuits, each

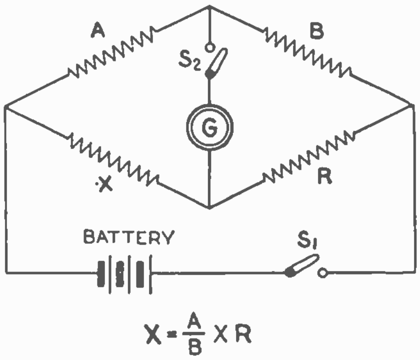

Diagram illustrating principle of the Wheatstone Bridge.

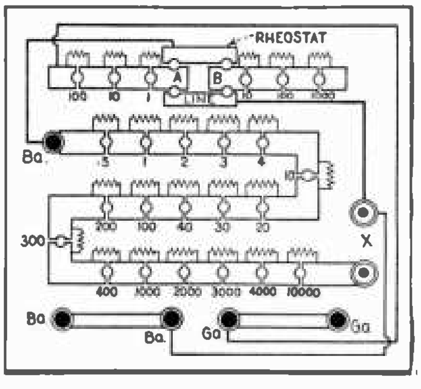

Post-Office type of Wheatstone Bridge.

divided into two "arms" capable of being adjusted so that there is no difference of potential between the dividing points on both sides. The point of zero difference of potential is indicated by the absence of deflection of a galvanometer placed across the dividing points. The ratio of the resistance of the two parts of the arms is then the same, so that if the resistance of three of the arms is known, that of the fourth arm (i.e. the resistance to be measured) can be determined.

WHEATSTONE TRANSMITTER—An apparatus for delivering telegraphic currents to a line at high speed. Control is effected by aid of a moving perforated paper tape, prepared according to a code.

WHIPPING—The binding of string or small wire round the end of a rope or multiple wire to prevent the ends from fraying out.

WIMSHURST—A type of Induction or Influence Machine consisting, in a simple form, of two discs of insulating material revolving in opposite directions. These carry a number of equal sectors which form combined inductors and carriers. There are usually two collectors and two pair of brushes. (See *Induction Machine.*)

WINDOW LEAD-IN—A form of insulator for passing an aerial *lead-in* through a window. In some cases porcelain tubes are used as window lead-in's. A common form of window lead-in consists of a flat copper conductor covered with woven cotton insulation.

WIRED WIRELESS—A method of transmitting radio messages, in which the waves are guided by wires instead of radiating freely through the ether. (See *Line Radio.*)

WIRE GAUGE—A device for measuring the diameter of a round wire in accordance with a predetermined scale. The gauge commonly used in this country is the Brown and Sharpe (B. & S.) or American gauge.

WIRELESS TELEGRAPHY—Also known as RADIO TELEGRAPHY (q.v.).—A system of telegraphy, utilizing electromagnetic waves set up by oscillating currents as the means of transmission. The waves generated at the transmitting station spread out from the *aerial* (q.v.) in all directions. They are intercepted by the receiving aerial, detected by a rectifying device such as a crystal or vacuum tube, amplified and delivered to a head set or loud speaker. A key is used at the transmitter to control the duration of the waves so that they will correspond to the dots and dashes of the *continental code* and these dots and dashes are reproduced in the phones at the receiving station.

WOLLASTON WIRE—Platinum wire, drawn extremely fine by the process of coating fine platinum wire with ductile material, drawing the whole down together, and (if required) removing the coating by dissolving it off.

WORK—The action of a force upon a body to overcome resistance. Work is measured by the product of the force exerted and the distance moved. Work may be measured by *ergs* (q.v.), *foot-pounds, horsepower*, etc.

WOOD'S METAL—A soft metallic alloy having a low melting point. This alloy consists of two parts of lead, one part of tin, four parts of bismuth, and one part of cadmium, all by weight. This metal has been used extensively for setting detector crystals in their cups, so as to make a good electrical contact.

X, Y, Z

X—Symbol for *Reactance* (q.v.).

"X" RAYS—Also called *Roentgen Rays.* Electric waves of much higher frequency than light, produced by the striking of *cathode rays* upon a solid substance. These rays are not deflected by electric or magnetic fields and possess the property of penetrating solid substances which could not be penetrated by ordinary light rays. The use of the "X" Rays permit photographs to be taken of bones or other parts of the human body which could not be seen in any other way.

X'S—Disturbances of an erratic nature heard from time to time in the 'phones or loud speaker of a radio receiving set, due to storms, electric disturbances, etc. (See *Atmospherics, Natural Electric Waves, Static,* also *Strays.*)

YAGI SPARK GAP—A form of spark gap originated by H. Yagi. It is a quenched spark gap, the electrodes of which are aluminum and brass. The gap functions in an atmosphere of coal gas.

"Y" GROUPING—See *Star Grouping.*

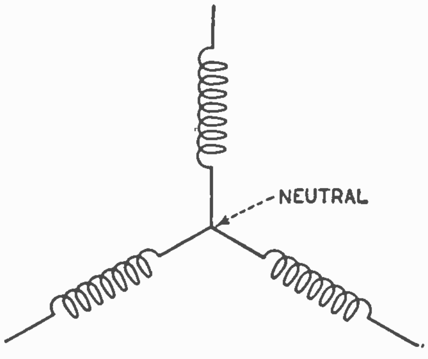

A three-phase "Y" connection.

YOKE—A piece of soft iron used in certain forms of electromagnets to yoke two parallel cores together magnetically. The cores of the ordinary electric bell are fastened together at one end by means of a yoke.

YOKOJAMA, EITARO—Japanese radio expert. He was born in 1883 and was educated at the Engineering College of the Tokyo Imperial University. While at college, he specialized in radio. He was appointed to the Electro-technical Laboratory of the Japanese Ministry of Communications, to carry out research work in wireless telegraphy and telephony. He was one of the inventors of the T.Y.K. oscillation gaps for radio telephony, for which he received many distinctions. In 1910 he was appointed head of the Radio Section. Yokojama, who is one of the most brilliant Japanese radio experts, is a member of the Institute of Radio Engineers and of many other scientific societies.

ZEEMAN EFFECT—Doubling of the spectrum lines of light sources when placed in a strong magnetic field.

ZENNECK, J.—German wireless expert. He was born April 15th, 1871, at Wurtemberg and was educated at Tuebingen. In 1895 he was appointed assistant in the Physical Institute in Strassburg, a post he held until 1899, when he carried out a series of tests in radio telegraphy in the North Sea. In 1905 he was appointed assistant professor of physics at the Institute of Technology, Brunswick and was appointed professor at Munich in 1913. Professor Zenneck has written a number of authoritative books on wireless and also a large number of articles on electro-magnetic oscillations.

ZERO BEAT RECEPTION—The detection of continuous modulated waves using a local source of high frequency current having a frequency equal to that of the incoming wave. When this continuous voltage is impressed on the incoming modulated voltage, (both of the same frequency) the output of the detector contains a current of audio frequency similar in character to the modulated current at the transmitter.

ZERO METHOD—A method of measurement in which various adjustments are made until the current flowing through a galvanometer is reduced to zero, as in a *Wheatstone Bridge* (q.v.). (See *Null Method.*)

ZINC—A metallic element. Its chemical symbol is Zn. In color, zinc is bluish-gray. It is practically non-corrosive in the atmosphere, is capable of taking a high polish, is unaffected by water, but is soluble in nitric acid and in soda and potash solutions. Pure zinc is attacked very slowly by sulphuric acid, but this feature is one of the greatest in the application of zinc in radio work and in electrical work generally.

Zinc is one of the most important components of most dry cells. It forms the negative terminal in most "B" batteries. Zinc in rod form is used in most forms of wet cells.

ZINCITE—An oxide of zinc. Zincite crystals can be distinguished by their red color, often broken up by orange-yellow streaks. This crystal, in combination with several other crystals makes a very excellent detector for radio purposes. The *perikon detector* (q.v.) is a combination of zincite and chalcopyrites. Zincite is also used in combination with bornite, galena, copper, iron pyrites, tellurium and silicon.

ZIRCONIUM—One of the metallic elements. Its chemical symbol is Zr, and its atomic weight is 90.6. Zirconium is an iron-gray powder in one form, or it may be made to crystallize. The crystals look like antimony, are very brittle and extremely hard, being capable of scratching glass and rubies. Zirconium resembles thorium in many of its chemical properties. For the control of the vacuum in high vacuum tubes, a small quantity of thorium or zinconium is included in the tube. These metals combine with hydrogen, oxygen, nitrogen, etc., to form compounds of very low vapor pressure.

United States Patent [19]

Rogers, Sr.

[11] **4,292,804**

[45] **Oct. 6, 1981**

[54] METHOD AND APPARATUS FOR OPERATING AN ENGINE ON COMPRESSED GAS

[76] Inventor: **Leroy K. Rogers, Sr.,** #5 Capistrano Ct., Ft. Myers, Fla. 33908

[21] Appl. No.: **158,303**

[22] Filed: **Jun. 10, 1980**

[51] **Int. Cl.**^{3} **F15B 11/06**

[52] **U.S. Cl.** **60/407**; 91/187; 91/275

[58] **Field of Search** 60/407, 412; 91/187, 91/275, 364

[56] References Cited

U.S. PATENT DOCUMENTS

3,881,399	5/1975	Sagi et al.	91/187 X
3,885,387	5/1975	Simington	60/407 X
4,018,050	4/1977	Murphy	60/412 X

Primary Examiner—Allen M. Ostrager
Attorney, Agent, or Firm—Burns, Doane, Swecker & Mathis

[57] ABSTRACT

The present invention relates to a method and apparatus for operating an engine having a cylinder and a piston reciprocable therein on compressed gas. The apparatus comprises a source of compressed gas connected to a distributor which distributes the compressed gas to the cylinder. A valve is provided to selectively admit compressed gas to the cylinder when the piston is in an approximately top dead center position. In one embodiment of the present invention the timing of the opening of the valve is advanced such that the compressed gas is admitted to the cylinder progressively further before the top dead center position of the piston as the speed of the engine increases. In a further embodiment of the present invention a valve actuator is provided which increases the length of time over which the valve remains open to admit compressed gas to the cylinder as the speed of the engine increases. A still further embodiment of the present invention relates to an apparatus for adapting a conventional internal combustion engine for operation on compressed gas.

22 Claims, 8 Drawing Figures

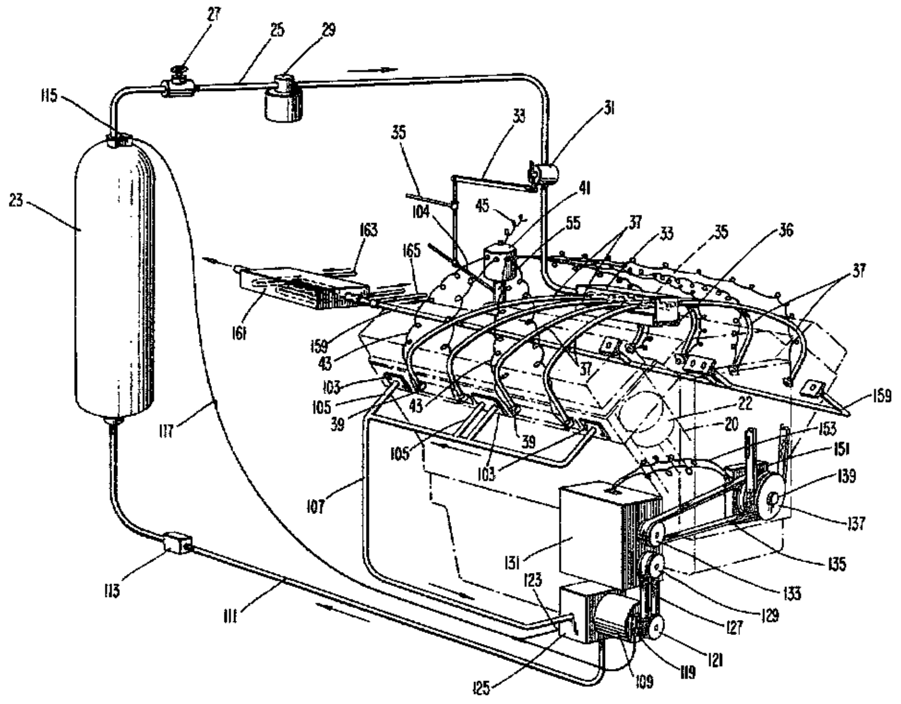

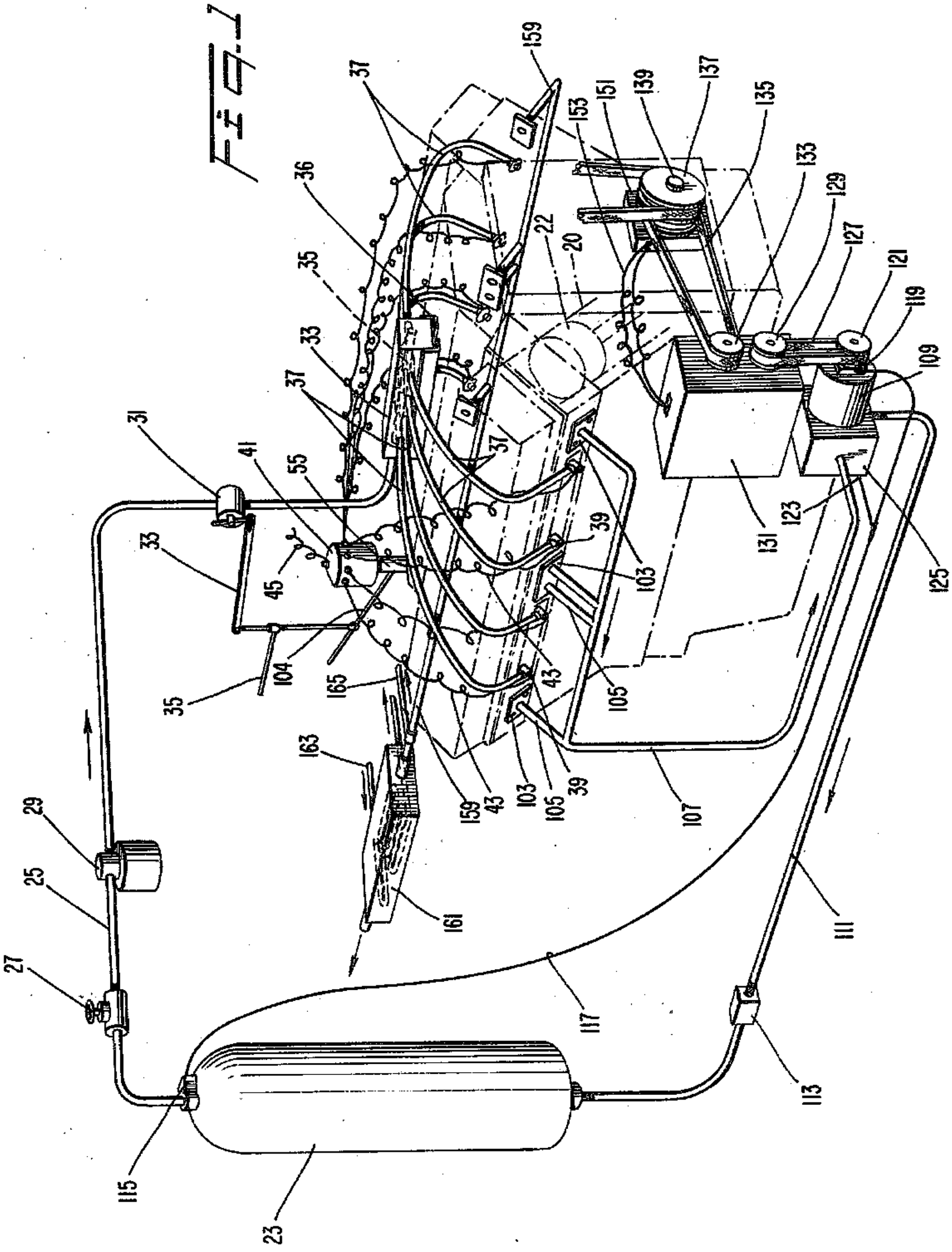

FIG. 1

Fig. 6

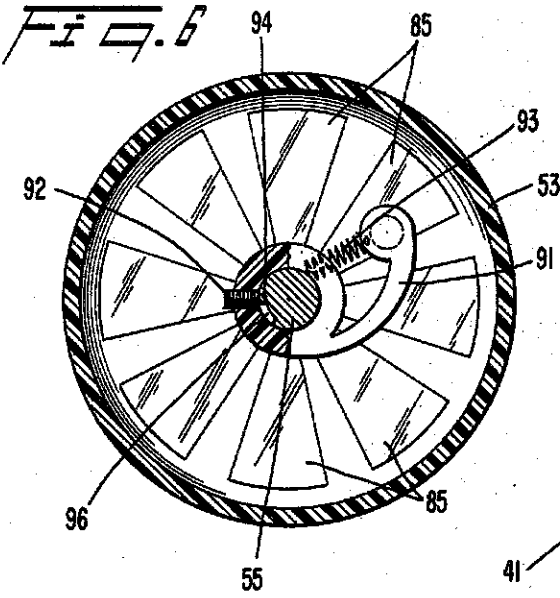
94
85
93
92
53
91
96
55
85

Fig. 7

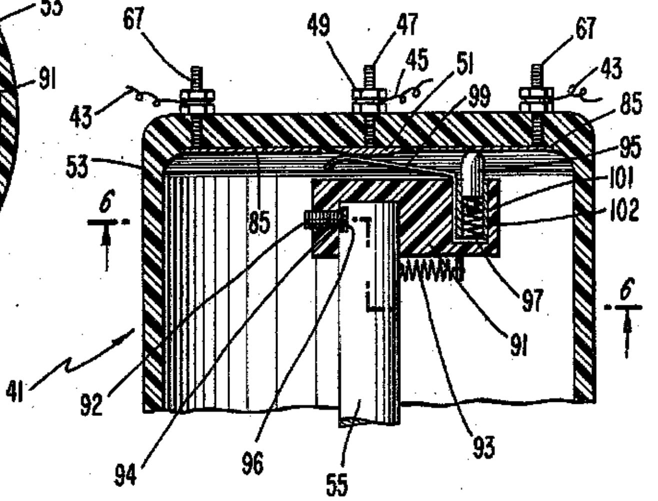
67
49
47
45
51
67
43
43
99
85
53
95
101
6
85
102
97
6
91
41
92
94
96
55
93

Fig. 3

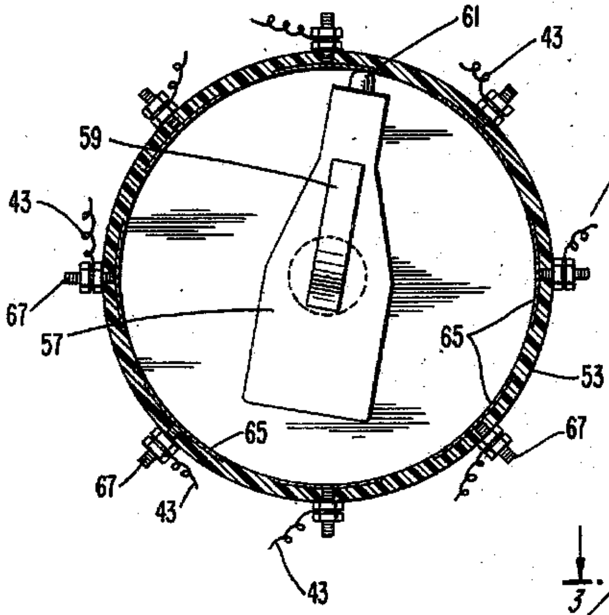
61
43
59
43
67
57
65
53
65
67
67
43
43

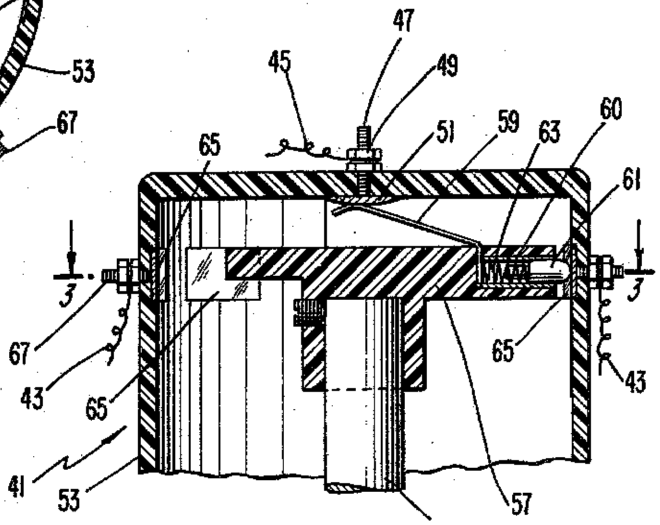
47
45
49
51
59
63
60
65
61
3
3
67
43
65
65
43
41
53
57
55

Fig. 2

Fig. 5

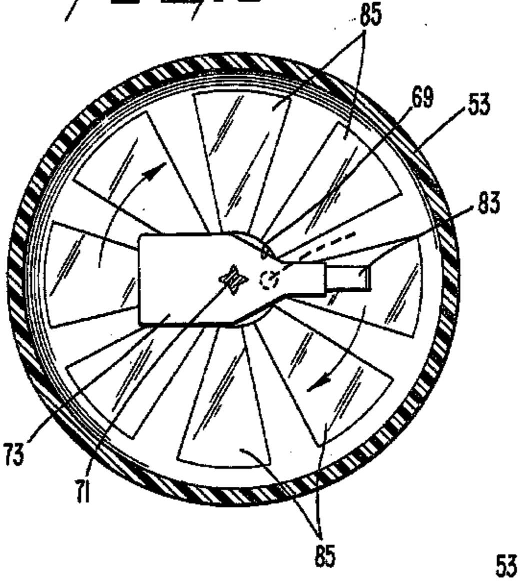
85
69
53
83
73
71
85

Fig. 4

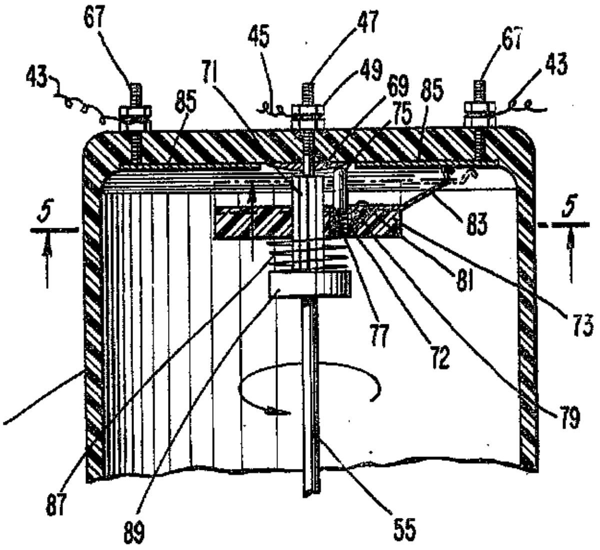
67
45
47
67
43
43
71
49
69
85
75
85
5
5
83
81
73
77
72
79
53
87
89
55

Fig. 6

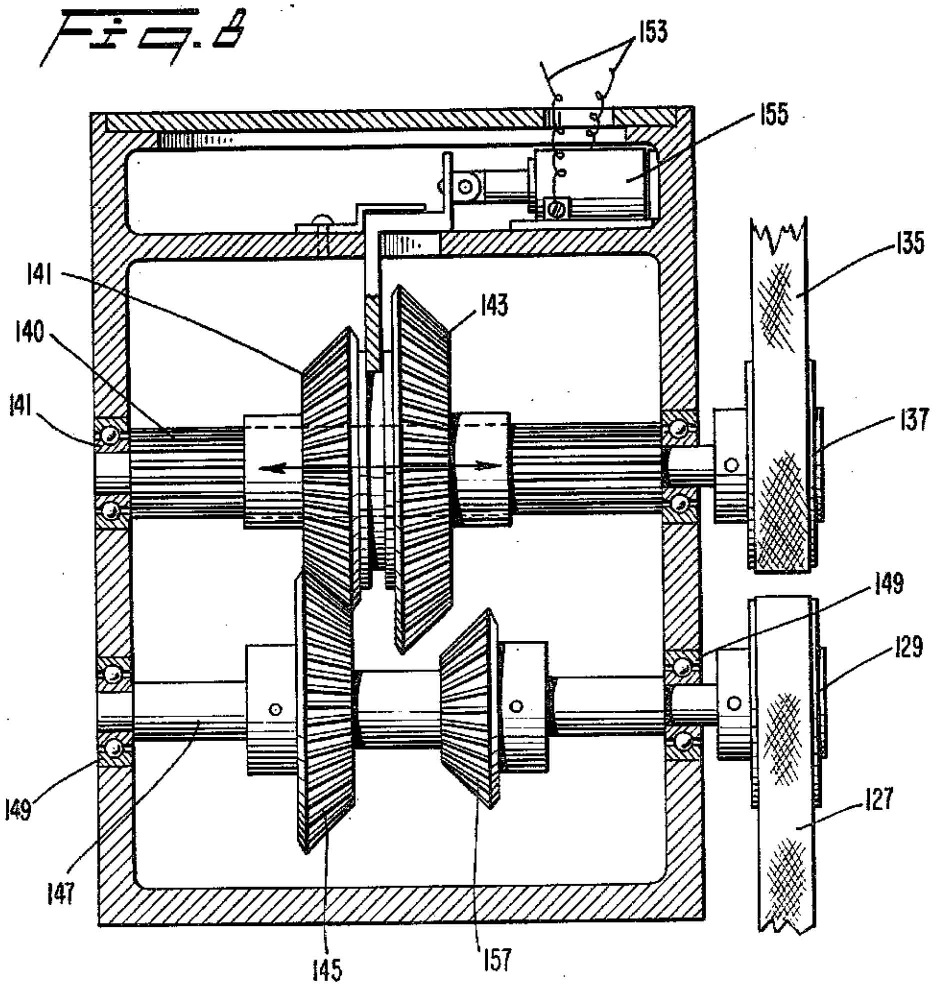
153
155
141
140
141
143
135
137
149
129
127
149
147
145
157

METHOD AND APPARATUS FOR OPERATING AN ENGINE ON COMPRESSED GAS

BACKGROUND AND SUMMARY OF THE PRESENT INVENTION

The present invention relates to a method and apparatus for operating an engine using a compressed gas as the motive fluid. More particularly, the present invention relates to a apparatus for adapting a pre-existing internal combustion engine for operation on a compressed gas.

Air pollution is one of the most serious problems facing the world today. One of the major contributors to air pollution is ordinary internal combustion engine which are used in most motor vehicles today. Various devices, including many items mandated by legislation, have been proposed in an attempt to limit the pollutants which an internal combustion engine exhausts to the air. However, most of these devices have met with limited success and are often both prohibitively expensive and complex. A clean alternative to the internal combustion engine is needed to power vehicles and other machinery.

A compressed gas, preferably air, would provide an ideal motive fluid for a engine since it would eliminate the usual pollutants exhausted from an internal combustion engine. An apparatus for converting an internal combustion engine for operation on compressed air is disclosed in U.S. Pat. No. 3,885,387 issued May 27, 1975 to Simington. The Simington patent discloses an apparatus including a source of compressed air and a rotating valve actuator which opens and closes a plurality of mechanical poppet valves. The valves deliver compressed air in timed sequence to the cylinders of an engine through adapters located in the spark plug holes. However, the output speed of an engine of this type is limited by the speed of the mechanical valves and the fact that the length of time over which each of the valves remains open cannot be varied as the speed of the engine increases.

Another apparatus for converting an internal combustion engine for operation on steam or compressed air is disclosed in U.S. Pat. No. 4,102,130 issued July 25, 1978 to Stricklin. The Stricklin patent discloses a device which changes the valve timing of a conventional four stroke engine such that the intake and exhaust valves open once for every revolution of the engine instead of once every other revolution of the engine. A reversing valve is provided which delivers live steam or compressed air to the intake valves and is subsequently reversed to allow the exhaust valves to deliver the expanded steam or air to the atmosphere. A reversing valve of this type however does not provide a reliable apparatus for varying the amount of motive fluid injected into the cylinders when it is desired to increase the speed of the engine. Further, a device of the type disclosed in the Stricklin patent requires the use of multiple reversing valves if the cylinders in a multi-cylinder engine were to be fired sequentially.

Therefore, it is an object of the present invention to provide a reliable method and apparatus for operating an engine or converting an engine for operation with a compressed gas.

A further object of the present invention is to provide a method and apparatus which is effective to deliver a constantly increasing amount of compressed gas to an engine as the speed of the engine increases.

A still further object of the present invention is to provide a method and apparatus which will operate an engine using compressed gas at a speed sufficient to drive a conventional automobile at highway speeds.

It is still a further object of the present invention to provide a method and apparatus which is readily adaptable to a standard internal combustion engine to convert the internal combustion engine for operation with a compressed gas.

Another object of the invention is to provide a method and apparatus which utilizes cool expanded gas, exhausted from a compressed gas engine, to operate an air conditioning unit and/or an oil cooler.

These and other objects are realized by a method and apparatus according to the present invention for operating an engine having at least one cylinder and a reciprocating piston therein using compressed gas as a motive fluid. The apparatus includes a source of compressed gas and a distributor connected with the source of the compressed gas for distributing the compressed gas to the at least one cylinder. A valve is provided for admitting the compressed gas to the cylinder when the piston is in approximately a top dead center position within the cylinder. An exhaust is provided for exhausting the expanded gas from the cylinder as the piston returns to approximately the top dead center position.

In a preferred embodiment of the present invention a device is provided for varying the duration of each engine cycle over which the valve remains open to admit compressed gas to the cylinder dependent upon the speed of the engine. In a further preferred embodiment of the present invention, an apparatus for advancing the timing of the opening of the valve is arranged to admit the compressed gas to the cylinder progressively further before the top dead center position of the piston as the speed of the engine increases.

Further features of the present invention include a valve for controlling the amount of compressed gas admitted to the distributor. Also, a portion of the gas which has been expanded in the cylinder and exhausted through the exhaust valve is delivered to a compressor to be recompressed and returned to the source of compressed gas. A gear train is selectively engagable to drive the compressor at different operating speed depending upon the pressure maintained at the source of compressed air and/or the speed of the engine. Still further, a second portion of the exhaust gas is used to cool a lubricating fluid for the engine or to operate an air conditioning unit.

In a preferred embodiment of the present invention, the valve for admitting compressed gas to the cylinder is electrically actuated. The device for varying the duration of each engine cycle over which the intake valve remains open as the speed of the engine increase comprises a rotating element whose effective length increases as the speed of the engine increases such that a first contact on the rotating element is electrically connected to a second contact for a longer period of each engine cycle. The second contact actuates the valve whereby the valve remains in an open position for a longer period of each engine cycle as the speed of the engine increases.

Still further features of the present invention include an adaptor plate for supporting the distributor above an intake manifold of a conventional internal combustion engine after a carburetor has been removed to allow air

to enter the cylinders of the engine through the intake manifold and conventional intake valves. Another adaptor plate is arranged over an exhaust passageway of the internal combustion engine to reduce the cross-sectional area of the exhaust passageway.

BRIEF DESCRIPTION OF THE DRAWINGS

Preferred embodiments of a method and apparatus for operating an engine according to the present invention will be described with reference to the accompanying drawings wherein like members bear like reference numerals and wherein:

FIG. 1 is a schematic representation of an apparatus according to the present invention arranged on an engine;

FIG. 2 is a side view of one embodiment of a valve actuator according to the present invention;

FIG. 3 is a cross-sectional view taken along the line 3—3 in FIG. 2;

FIG. 4 is a cross-sectional view of a second embodiment of a valve actuator according to the present invention;

FIG. 5 is a view taken along the line 5—5 in FIG. 4;

FIG. 6 is a cross-sectional view of a third embodiment of a valve actuator according to the present invention;

FIG. 7 is a view taken along the line 7—7 in FIG. 6;

FIG. 8 is a cross-sectional view of a gearing unit to drive a compressor according to the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

With reference to FIG. 1, an engine block 21 (shown in phantom) having two banks of cylinders with each bank including cylinders 20 having pistons 22 reciprocable therein (only one of which is shown in phantom) in a conventional manner. While the illustrated engine is a V-8 engine, it will be apparent that the present invention is applicable to an engine having any number of pistons and cylinders with the V-8 engine being utilized for illustration purposes only. A compressed gas tank 23 is provided to store a compressed gas at high pressure. It may also be desirable to include a small electric or gas compressor to provide compressed gas to supplement the compressed gas held in the tank 23. In a preferred embodiment, the compressed gas is air which can be obtained from any suitable source.

A line 25 transports the gas withdrawn from the tank 23 when a conventional shut off valve 27 is open. In addition, a solenoid valve 29 preferably operated by a suitable key operated switch (not shown) for the engine is also arranged in the line 25. In normal operation, the valve 27 is maintained open at all times with the solenoid valve 29 operating as a selective shut off valve to start and stop the engine 21 of the present invention.

A suitable regulating valve 31 is arranged downstream from the solenoid valve 29 and is connected by a linkage 33 to a throttle linkage 35 which is operator actuated by any suitable apparatus such as a foot pedal (not shown). The line 25 enters an end of a distributor 33 and is connected to an end of a pipe 35 which is closed at the other end. A plurality of holes, which are equal to the number of cylinders in the engine 21, are provided on either side of the pipe 35 along the length of the pipe 35.

When the present invention is used to adapt a conventional internal combustion engine for operation on compressed gas, an adaptor plate 36 is provided to support the distributor 33 in spaced relation from the usual intake opening in the intake manifold of the engine after a conventional carburetor has been removed. In this way, air is permitted to enter the internal combustion engine through the usual passageways and to be admitted to the cylinders through suitable intake valves (not shown). The adaptor plate 36 is secured to the engine block 21 and the distributor 33 by any suitable apparatus, e.g., bolts.

Each of the holes in the pipe 35 is connected in fluid-tight manner to a single line 37. Each line 37 carries the compressed gas to a single cylinder 20. In a preferred embodiment, each of the lines 37 is ½ inch high pressure plastic tubing attached through suitable connectors to the distributor 33 and the pipe 35. Each of the lines 37 is connected to a valve 39 which is secured in an opening provided near the top of each of the cylinders 20. In the case of a conversion of a standard internal combustion engine, the valves 39 can be conveniently screwed into a tapped hole in the cylinder 20 typically provided for a spark plug of the internal combustion engine. In a preferred embodiment, the valves 39 are solenoid actuated valves in order to provide a fast and reliable opening and closing of the valves 39.

Each of the valves 39 is energized by a valve actuator 41 through one of a plurality of wires 43. The valve actuator 41 is driven by a shaft of the engine similar to the drive for a conventional distributor of an internal combustion engine. That is, a shaft 55 of the valve actuator 41 is driven in synchronism with the engine 21 at one half the speed of the engine 21.

A first embodiment of the valve actuator 41 (FIGS. 2 and 3) receives electrical power through a wire 45 which is energized in a suitable manner by a battery, and a coil if necessary (not shown) as is conventional in an internal combustion engine. The wire 45 is attached to a central post 47 by a nut 49. The post 47 is connected to a conducting plate 51 arranged within a housing 53 for the valve actuator 41. Within the housing 53, the shaft 55 has an insulating element 57 secured to an end of the shaft 55 for co-rotation therewith when the shaft 55 is driven by the engine 21. A first end of a flexible contact 59 is continuously biased against the conducting plate 51 to receive electricity from the battery or another suitable source. A second end of the contact 59 is connected to a conducting sleeve 60 which is in constant contact with a spring biased contact 61 which is arranged within the sleeve 60. The contact 61 is biased by a spring 63 which urges the contact 61 towards a side wall of the housing 53.

With reference to FIG. 3, a plurality of contacts 65 are spaced from one another and are arranged around the periphery of the housing 53 at the same level as the spring biased contact 61. Each contact 65 is electrically connected to a post 67 which extends outside of the housing 53. The number of contacts 65 is equal to the number of cylinders in the engine 21. One of the wires 43, which actuate the valves 39, is secured to each of the posts 67.

In operation, as the shaft 55 rotates in synchronism with the engine 21, the insulating element 57 rotates and electricity is ultimately delivered to successive ones of the contacts 65 and wires 43 through the spring biased contact 61 and the flexible contact 59. In this way, each of the electrical valves 39 is actuated and opened in the proper timed sequence to admit compressed gas to each of the cylinders 20 to drive the pistons 22 therein on a downward stroke.

The embodiment illustrated in FIGS. 2 and 3 is effective to actuate each of the valves 39 to remain open for a long enough period of time to admit sufficient compressed gas to each of the cylinders 20 of the engine 21 to drive the engine 21. The length of each of the contacts 65 around the periphery of the housing 53 is sufficient to permit the speed of the engine to be increased when desired by the operator by moving the throttle linkage 35 which actuates the linkage 33 to further open the regulating valve 31 to admit more compressed gas from the tank 23 to the distributor 33. However, it has been found that the amount of air admitted by the valves 39 when using the first embodiment of the valve actuator 41 (FIGS. 2 and 3) is substantially more than required to operate the engine 21 at an idling speed. Therefore, it may be desirable to provide a valve actuator 41 which is capable of varying the duration of each engine cycle over which the solenoid valves 39 are actuated, i.e., remain open to admit compressed gas, as the speed of the engine 21 is varied.

A second embodiment of a valve actuator 41 which is capable of varying the duration of each engine cycle over which each of the valves 39 remains open to admit compressed gas to the cylinders 20 dependent upon the speed of the engine 21 will be described with reference to FIGS. 4 and 5 wherein members corresponding to those of FIGS. 2 and 3 bear like reference numerals. The wire 45 from the electrical source is secured to the post 47 by the nut 49. The post 47 has a annular contact ring 69 electrically connected to an end of the post 47 and arranged within the housing 53. The shaft 55 rotates at one half the speed of the engine as in the embodiment of FIGS. 2 and 3.

At an upper end of the shaft 55, a splined section 71 slidably receives an insulating member 73. The splined section 71 of the shaft 55 positively holds the insulating member 73 for co-rotation therewith but permits the insulating member 73 to slide axially along the length of the splined section 71. Near the shaft 55, a conductive sleeve 72 is arranged in a bore 81 in an upper surface of the insulating element 73 generally parallel to the splined section 71. A contact 75, biased towards the annular contact ring 69 by a spring 77, is arranged within the conductive sleeve 72 in contact therewith. The conductive sleeve 72 also contacts a conductor 79 at a base of the bore 81.

The conductor 79 extends to the upper surface of the insulating element 73 near an outer periphery of the insulating element 73 where the conductor 79 is electrically connected to a flexible contact 83. The flexible contact 83 selectively engages a plurality of radial contacts 85 arranged on an upper inside surface of the housing 53. A weak spring 87 arranged around the splined section 71 engages a stop member 89 secured on the shaft 55 and the insulating element 73 to slightly bias the insulating element 73 towards the upper inside surface of the housing 53 to ensure contact between the flexible contact 83 and the upper inside surface of the housing 53. As best seen in FIG. 5, the radial contacts 85 on the upper inside surface of the housing 53 are arranged generally in the form of radial spokes extending from the center of the housing 53 with the number of contacts being equal to the number of cylinders 20 in the engine 21. The number of degrees covered by each of the radial contacts 85 gradually increases as the distance from the center of the upper inside surface of the housing 53 increases.

In operation of the device of FIGS. 4 and 5, as the shaft 55 rotates, electricity flows along a path through the wire 45 down through post 47 to the annular contact member 69 which is in constant contact with the spring biased contact 75. The electrical current passes through the conductive sleeve 72 to the conductor 79 and then to the flexible contact 83. As the flexible contact 83 rotates along with the insulating member 73 and the shaft 55, the tip of the flexible contact 83 successively engages each of the radial contacts 85 on the upper inside of the housing 53. As the speed of the shaft 55 increases, the insulating member 73 and the flexible contact 83 attached thereto move upwardly along the splined section 71 of the shaft 55 due to the radial component of the splines in the direction of rotation under the influence of centrifugal force. As the insulating member 73 moves upwardly, the flexible contact 83 is bent such that the tip of the contact 83 extends further radially outwardly from the center of the housing 53 (as seen in phantom lines in FIG. 4). In other words, the effective length of the flexible contact 83 increases as the speed of the engine 21 increases.

As the flexible contact 83 is bent and the tip of the contact 83 moves outwardly, the tip remains in contact with each of the radial contacts 85 for a longer period of each engine cycle due to the increased angular width of the radial contacts with increasing distance from the center of the housing 53. In this way, the length of time over which each of the valves 39 remains open is increased as the speed of the engine is increased. Thus, a larger quantity of compressed gas or air is injected into the cylinders as the speed increases. Conversely, as the speed decreases and the insulating member 73 moves downwardly along the splined section 71, a minimum quantity of air is injected into the cylinder due to the shorter length of the individual radial contact 85 which is in contact with the flexible contact 83. In this way, the amount of compressed gas that is used during idling of the engine 21 is at a minimum whereas the amount of compressed gas which is required to increase the speed of the engine 21 to a level suitable to drive a vehicle on a highway is readily available.

With reference to FIGS. 6 and 7, a third embodiment of a valve actuator 41 according to the present invention includes an arcuate insulating element 91 having a first end pivotally secured by any suitable device such as screw 92 to the shaft 55 for co-rotation with the shaft 55. The screw 92 is screwed into a tapped hole in the insulating element 91 such that a tab 94 at an end of the screw 92 engages a groove 96 provided in the shaft 55. In this way, the insulating element 91 positively rotates with the shaft 55. However, as the shaft 55 rotates faster, a second end 98 of the insulating element 91 is permitted to pivot outwardly under the influence of centrifugal force because of the groove 96 provided in the shaft 55. A spring 93 connected between the second end 98 of the element 91 and the shaft 55 urges the second end of the element 91 towards the center of the housing 53.

A contact 99 similar to the contact 59 (FIG. 2) is arranged such that one end of the contact 99 is in constant contact with the conducting plate 51 located centrally within the housing 53. The other end of the contact 99 engages a conductive sleeve 101 arranged in bore 102. A contact element 95 is arranged in the conductive sleeve 101 in constant contact with the sleeve 101. The bore 102 is arranged generally parallel to the shaft 55 near the second end of the arcuate insulating

element **91**. The contact **95** is biased by a spring **97** towards the upper inside surface of the housing **53** for selective contact with each of the plurality of radial contacts **85** which increase in arc length towards the outer peripheral surface of the housing **53** (FIG. **6**).

In operation of the device of FIGS. **6** and **7**, as the shaft **55** rotates the arcuate insulating element **91** rotates with the shaft **55** and the second end **98** of the insulating element **91** tends to pivot about the shaft **55** due to centrifugal force. Thus, as the effective length of the contact **95** increases, i.e., as the arcuate insulating element **91** pivots further outwardly, the number of degrees of rotation over which the contact **95** is in contact with each of the radial contacts **85** on the upper inside surface of the housing **53** increases thereby permitting each of the valves **39** to remain open for a longer period of each engine cycle to admit more compressed gas to the respective cylinder **20** to further increase the speed of the engine **21**.

With reference to FIG. **1**, a mechanical advance linkage **104** which is connected to the throttle linkage **35**, advances the initiation of the opening of each valve **39** such that compressed gas is injected into the respective cylinder further before the piston **22** in the respective cylinder **20** reaches a top dead center position as the speed of the engine is increased by moving the throttle linkage **35**. The advance linkage **104** is similar to a conventional standard mechanical advance employed on an internal combustion engine. In other words, the linkage **104** varies the relationship between the angular positions of a point on the shaft **55** and a point on the housing **53** containing the contacts. Alternatively, a conventional vacuum advance could also be employed. By advancing the timing of the opening of the valves **39**, the speed of the engine can more easily be increased.

The operation of the engine cycle according to the present invention will now be described. The compressed gas injected into each cylinder of the engine **21** drives the respective piston **22** downward to drive a conventional crankshaft (not shown). The movement of the piston downwardly causes the compressed gas to expand rapidly and cool. As the piston **22** begins to move upwardly in the cylinder **20** a suitable exhaust valve (not shown) arranged to close an exhaust passageway is opened by any suitable apparatus. The expanded gas is then expelled through the exhaust passageway. As the piston **22** again begins to move downwardly a suitable intake valve opens to admit ambient air to the cylinder. The intake valve closes and the ambient air is compressed on the subsequent upward movement of the piston until the piston reaches approximately the top dead center position at which time the compressed gas is again injected into the cylinder **20** to drive the piston **22** downward and the cycle begins anew.

In the case of adapting a conventional internal combustion engine for operation on compressed gas, a plurality of plates **103** are preferably arranged over an end of the exhaust passageways in order to reduce the outlet size of the exhaust passageways of the conventional internal combustion engine. In the illustrated embodiment, a single plate having an opening in the center is bolted to the outside exhaust passageway on each bank of the V-8 engine while another single plate having two openings therein is arranged with one opening over each of the interior exhaust passageways on each bank of the V-8 engine. A line **105** is suitably attached to each of the adaptor places to carry the exhaust to an appropriate location. In a preferred embodiment, the exhaust lines **105** are $1\frac{1}{2}''$ plastic tubing.

In a preferred embodiment, the exhaust lines **105** of one bank of the V-8 engine are collected in a line **107** and fed to an inlet of a compressor **109**. The pressure of the exhaust gas emmanating from the engine **21** according to the present invention is approximately 25 p.s.i. In this way, the compressor **109** does not have to pull the exhaust into the compressor since the gas exhausted from the engine **21** is at a positive pressure. The positive pressure of the incoming fluid increases the efficiency and reduces wear on the compressor **109**. The exhaust gas is compressed in the compressor **109** and returned through a line **111** and a check valve **113** to the compressed gas storage tank **23**. The check valve **113** prevents the flow of compressed gas stored in the tank **23** back towards the compressor **109**.

A suitable pressure sensor **115** is arranged at an upper end of the tank **23** and sends a signal along a line **117** when the pressure exceeds a predetermined level and when the pressure drops below a predetermined level. The line **117** controls an electrically actuated clutch **119** disposed at a front end of the compressor **109**. The clutch **119** is operative to engage and disengage the compressor **109** from a drive pulley **121**. Also, the signal carried by the line **117** actuates a suitable valve **123** arranged on a compressor housing **125** to exhaust the air entering the compressor housing **125** from the line **107** when the clutch **119** has disengaged the compressor **109** from the drive pully **121**.

In a preferred embodiment, when the pressure is the tank **23** reaches approximately 600 p.s.i., the clutch **119** is disengaged and the compressor **109** is deactivated and the valve **123** is opened to exhaust the expanded gas delivered to the compressor **109** from the line **107** to the atmosphere. When the pressure within the tank **23** drops below approximately 500 p.s.i., the sensor **115** sends a signal to engage the clutch **119** and close the valve **123**, thereby operating the compressor **109** for supplying the tank **23** with compressed gas.

The pulley **121** which drives the compressor **109** through the clutch **119** is driven by a belt **127** which is driven by a pulley **129** which operates through a gear box **131**. With reference to FIGS. **1** and **8**, a second pulley **133** on the gear box is driven by a belt **135** from a pulley **137** arranged on a drive shaft **139** of the engine **21**. The pulley **137** drives a splined shaft **140** which has a first gear **141** and a second larger gear **143** arranged thereon for rotation with the splined shaft **140**. The splined shaft **140** permits axial movement of the gears **141** and **143** along the shaft **140**.

In normal operation (as seen in FIG. **8**), the first gear **141** engages a third gear **145** arranged on a shaft **147** which drives the pulley **129**. The shafts **140** and **147** are arranged in suitable bearings **149** arranged at each end thereof. When the speed of the engine **21** drops below a predetermined level, a suitable sensor **151** responsive to the speed of the drive shaft **139** of the engine **21** generates a signal which is transmitted through a line **153** to a solenoid actuator **155** arranged within the gear box **131**. The solenoid actuator **155** moves the first and second gears **141**, **143** axially along the splined shaft **140** to the right as seen in FIG. **8** such that the second, larger gear **143** engages a fourth smaller gear **157** which is arranged on the shaft **147**. The ratio of the second gear **143** to the fourth gear **157** is preferably approximately 3 to 1.

In this way, when the speed of the engine **21** drops below the predetermined level as sensed by the sensor **151** (which predetermined level is insufficient to drive the compressor **109** at a speed sufficient to generate the 500–600 pounds of pressure which is preferably in the tank **23**), the solenoid actuator **155** is energized to slide the gears **143**, **141** axially along the splined shaft **140** so that the second, larger gear **143** engages the fourth, smaller gear **157** to drive the pulley **129** and hence the compressor **109** at a higher rate of speed to generate the desired pressure. When the speed of the engine increases above the predetermined level, in a preferred embodiment approximately 1500 rpm, the solenoid actuator **155** is deactivated by the sensor **151** thereby moving the gears **143** and **141** to the left as seen in FIG. **8** such that the first gear **141** re-engages with the third gear **145** to effectuate a 1 to 1 ratio between the output shaft **139** of the engine **21** and the pulley **129**.

The other bank of the V-8 engine has its exhaust ports arranged with adapter plates **103** similar to those on the first bank. However, the exhaust from this bank of the engine **21** is not collected and circulated through the compressor **109**. In a preferred embodiment, a portion of the exhaust is collected in a line **159** and fed to an enlarged chamber **161**. A second fluid is fed through a line **163** into the chamber **161** to be cooled by the cool exhaust emmanating from the engine **21** in the line **159**. The second fluid in the line **163** may be either transmission fluid contained in a transmission associated with the engine **21** or a portion of the oil used to lubricate the engine **21**. A second portion of the exhaust from the second bank of the V-8 engine is removed from the line **159** in a line **165** and used as a working fluid in an air conditioning system or for any other suitable use.

It should be noted that the particular arrangement utilized for collecting and distributing the gas exhausted from the engine **21** would be determined by the use for which the engine is employed. In other words, it may be advantageous to rearrange the exhaust tubing such that a larger or smaller percentage of the exhaust is routed through the compressor **109**. It should also be noted that since the exhaust lines **105** are plastic tubing, a rearrangement of the lines for a different purpose is both simple and inexpensive.

In operation of the engine of the present invention, the engine **21** is started by energizing the solenoid valve **29** and any suitable starting device (not shown), e.g., a conventional electric starter as used on an internal combustion engine. Compressed gas from the full tank **23** flows through the line **25** and a variable amount of the compressed gas is admitted to the distributor **33** by controlling the regulator valve **31** through the linkage **33** and the operator actuated throttle linkage **35**. The compressed gas is distributed to each of the lines **37** which lead to the individual cylinders **20**. The compressed gas is admitted to each of the cylinders **20** in timed relationship to the position of the pistons within the cylinders by opening the valves **39** with the valve actuator **41**.

When it is desired to increase the speed of the engine, the operator moves the throttle linkage **35** which simultaneously admits a larger quantity of compressed gas to the distributor **33** from the tank **23** by further opening the regulator valve **31**. The timing of the valve actuator **41** is also advanced through the linkage **104**. Still further, as the speed of the engine **21** increases, the effective length of the rotating contact **83** (FIG. 4) or **95** (FIG. 6) increases thereby electrically contacting a wider portion of one of the stationary radial contacts **85** to cause each of the valves **39** to remain open for a longer period of each engine cycle to admit a larger quantity of compressed gas to each of the cylinders **20**.

As can be seen, the combination of the regulating valve **31**, the mechanical advance **104**, and the valve actuator **41**, combine to produce a compressed gas engine which is quickly and efficiently adaptable to various operating speeds. However, all three of the controls need not be employed simultaneously. For example, the mechanical advance **104** could be utilized without the benefit of one of the varying valve actuators **41** but the high speed operation of the engine may not be as efficient. By increasing the duration of each engine cycle over which each of the valves **39** remains open to admit compressed gas to each of the cylinders **20** as the speed increases, conservation of compressed gas during low speed operation and efficient high speed operation are both possible.

After the compressed gas admitted to the cylinder **20** has forced the piston **22** downwardly within the cylinder to drive the shaft **139** of the engine, the piston **22** moves upwardly within the cylinder **20** and forces the expanded gas out through a suitable exhaust valve (not shown) through the adapter plate **103** (if employed) and into the exhaust line **105**. The cool exhaust can then be collected in any suitable arrangement to be compressed and returned to the tank **23** or used for any desired purpose including use as a working fluid in an air conditioning system or as a coolant for oil.

When using the apparatus and method of the present invention to adapt a ordinary internal combustion engine for operation with compressed gas it can be seen that considerable savings in weight are achieved. For example, the ordinary cooling system including a radiator, fan, hoses, etc. can be eliminated since the compressed gas is cooled as it expands in the cylinder. In addition, there are no explosions within the cylinder to generate heat. Further reductions in weight are obtained by employing plastic tubing for the lines which carry the compressed gas between the distributor and the cylinders and for the exhaust lines. Once again, heavy tubing is not required since there is little or no heat generated by the engine of the present invention. In addition, the noise generated by an engine according to the present invention is considerably less than that generated by an ordinary internal combustion engine since there are no explosions taking place within the cylinders.

The principles of preferred embodiments of the present invention have been described in the foregoing specification. However, the invention which is intended to be protected is not to be construed as limited to the particular embodiments disclosed. The embodiments are to be regarded as illustrative rather than restrictive. Variations and changes may be made by others without departing from the spirit of the invention. Accordingly, it is expressly intended that all such variations and changes which fall within the spirit and the scope of the present invention as defined in the appended claims be embraced thereby.

What is claimed is:

1. An apparatus for operating an engine having at least one cylinder and a reciprocating piston therein comprising:

a source of compressed gas;

distributor means connected with the source of compressed gas for distributing the compressed gas to the at least one cylinder;

valve means for admitting the compressed gas to the at least one cylinder when the piston is in approximately a top dead center position within the cylinder;

altering means for increasing the duration of each engine cycle over which the valve means admits compressed gas to the at least one cylinder as the speed of the engine increases; and

exhaust means for exhausting gas as the piston subsequently approaches approximately the top dead center position.

2. The apparatus of claim **1** further comprising control means for controlling the amount of compressed gas admitted to the distributor means.

3. The apparatus of claim **1** wherein the valve means is a solenoid valve secured in an opening in the cylinder above the level of the piston at the top dead center position.

4. The apparatus of claims **1** or **2** further comprising means for advancing the timing of the valve means as the speed of the engine increases such that compressed gas is admitted progressively further before the top dead center position as the speed of the engine increases.

5. The apparatus of claim **4** wherein the means for advancing the timing comprises a mechanical linkage connected to an operator actuated accelerator linkage.

6. The apparatus of claim **1** wherein a portion of the gas exhausted through the exhaust means is compressed in a compressor driven by an output shaft of the engine and is returned to the source of compressed gas.

7. The apparatus of claim **1** wherein a portion of the gas exhausted through the exhaust means is used to cool transmission fluid for a transmission associated with the engine.

8. The apparatus of claim **1** wherein a portion of the gas exhausted through the exhaust means is used as a working fluid in an air conditioning system.

9. The apparatus of claim **6** further comprising first gearing means interposed between the output shaft of the engine and the compressor for increasing the speed at which the compressor is driven.

10. The apparatus of claim **6** further comprising clutch means attached to the compressor both for disengaging the compressor from the output shaft of the engine when a first predetermined pressure at the source of compressed gas is exceeded and for engaging the compressor with the output shaft of the engine when the pressure at the source of compressed gas drops below a second predetermined pressure.

11. The apparatus of claim **9** further comprising means for both disengaging the first gearing means when a predetermined speed of the engine is exceeded and engaging a second gearing means for driving the compressor at a speed slower than the first gearing means when the predetermined speed of the engine is exceeded.

12. The apparatus of claim **1** wherein the valve means is electrically actuated and wherein the altering means comprises:

a rotating member timed with the at least one cylinder and arranged within a housing;

first and second contacts arranged on a first end of the rotating member and on an inside surface of the housing, respectively;

means for increasing the distance of the first contact from the rotational axis of the rotating member as the speed of the engine increases such that the first contact moves radially outwardly within the housing; and

said second contact presenting a longer arc length to the first contact as the distance of the first contact from the rotational axis of the rotating member increases.

13. The apparatus of claim **12** wherein the rotating member comprises an arcuate arm and wherein the means for increasing the distance of the first contact comprises pivotally mounting a second end of the arcuate arm about the axis of rotation of the rotating member and spring means for biasing the first end of the arcuate arm towards a radially inward position whereby the first end of the arcuate arm pivots radially outwardly as the speed of the engine increases.

14. The apparatus of claim **12** wherein the rotating member is axially slidably received on a rotating shaft for co-rotation therewith, said shaft having splines with a radial component in the direction of rotation, and wherein the first contact comprises a flexible contact located on an upper surface of the rotating member, said flexible contact being biased against the inside surface of the housing which carries the second contacts whereby as the speed of the engine increases the rotating member is urged axially along the splined shaft towards the inside surface of the housing such that the flexible contact is forced radially outwardly along the inside surface.

15. The apparatus of claim **12** wherein the second contact comprises of radially extending conductor arranged on an upper inside surface of the housing, said conductor increasing in arc length as the conductor extends radially outwardly from a central portion of the housing.

16. An apparatus for adapting an internal combustion engine for operation with compressed gas, the internal combustion engine having at least one cylinder, a piston reciprocable within the at least one cylinder, intake and exhaust means disposed in the at least one cylinder, and a tapped hole in the at least one cylinder adapted to receive a spark plug, the apparatus comprising:

a source of compressed gas;

distributor means connected with the source of compressed gas for distributing the compressed gas to the at least one cylinder;

valve means arranged in the tapped hole for admitting the compressed gas to the at least one cylinder when the piston is in approximately a top dead center position within the cylinder; and

altering means for increasing the duration of each engine cycle over which the valve means remains open to admit the compressed gas as the speed of the engine increases.

17. An apparatus as in claim **16** further comprising first adapter plate means for supporting the distributor means above an intake manifold of the engine, which adaptor plate means allows ambient air to enter through the intake manifold.

18. The apparatus of claim **16** further comprising second adapter plate means for reducing the exit area of the exhaust means.

19. A method of operating an engine on compressed gas, said engine having at least one cylinder and a piston reciprocable therein comprising the steps of:

delivering compressed gas from a source to a distributor;

distributing the compressed gas to the at least one cylinder;

admitting compressed gas to the at least one cylinder through an intake valve when the piston is at approximately a top dad center position;

increasing the duration of each engine cycle over which compressed gas is admitted to the at least one cylinder as the engine speed increases; and

exhausting the remaining gas when the piston subsequently reaches approximately the top dead center position.

20. The method of claim **19** further comprising the step of controlling the amount of compressed gas which is delivered to the distributor.

21. The method of claim **19** further comprising the step of advancing the timing of the opening of the intake valve as the speed of the engine increases.

22. An apparatus for operating an engine having at least one cylinder and a piston reciprocable therein on compressed gas comprising:

a source of compressed gas;

distributor means connected with the source of compressed gas for distributing the compressed gas to the at least one cylinder;

electrically actuated valve means secured in an opening in the at least one cylinder for selectively admitting compressed gas to the at least one cylinder when the piston is in approximately a top dead center position; and

means for advancing the timing of the valve means as the speed of the engine increases whereby compressed gas is admitted progressively further before the top dead center position as the speed of the engine increases.

* * * * *

PRELIMINARY PLANS TO RUN YOUR CAR ON TAP WATER!

IT ALSO WORKS ON YOUR
TRUCK / RV / MOTORCYCLE / AIRPLANE (ETC)

Will This Work?

These plans were sent to the Spirit of Ma'at anonymously, from someone who does not want his or her name printed (for obvious reasons).
We have had them checked by an expert who believes that they are real.
We also have talked with another individual who has patented a similar device, and we know by personal experience that the technology is sound.
So although we cannot guarantee it, we believe these plans will enable you to build a car that runs on water. If you test it out, though, do as the writer suggests and use an old car that doesn't represent a loss of value if you can't make it work. And leave everything intact so that you can always reconnect back to gas if you have to.

But if you do get it working, please send us your experience for our readers. You could be a national hero and help save our country and our world.

We know for certain that an automobile will run on water. So this could be an interesting project for you mechanical types, with a great reward of never having to purchase gasoline for the rest of your life - and helping humanity at the same time.

Re The Need to Rustproof Your Exhaust System

It is possible to make a hybrid of both gas and water (a system that is being tested now in Mexico), which would eliminate the need to open the head and remove the exhaust system. Just a thought. It takes only a small amount of gas to keep the system dry.

The text sent from the anonymous individual was edited slightly for better reading. The following is his/her words and drawings, which has been given into the public domain.
- Spirit of Ma’at LLC

Introduction

It is suggested you try this out to begin with on a second vehicle you own, one that you don't need to live with everyday, until you perfect this technology.

Do-it-yourself plans allow the individual (that's you and me, folks) to make a difference. This is the easiest and lowest-cost way to convert your car to run on (relatively) free energy.

Now, with existing technology, anyone can stand up and make a difference by reducing the local automotive pollution, eliminate gasoline expenses, help restore our atmosphere, and breathe a little easier.

In putting these plans into operation, you will be making use of your entire existing system except for the fuel tank and the catalytic converter.

The Plan

Build and install a low-cost alternative method for running your vehicle (internal combustion engine) on tap water, using off-the-shelf components.

This is simply an efficient way to convert ordinary tap water into gaseous hydrogen and oxygen, and then burn these vapors in the engine, instead of gasoline.

This "minisystem" runs easily from your existing battery and electrical system, and it plugs into your carburetor with simple off-the-shelf fittings.

You will be installing a plastic water tank, a control circuit, a reaction chamber, a hi-pressure carb/FI fitting, and 3 gauges, and then hooking into your existing carb/FI.

The simplicity comes from its being an "on-demand" system requiring no fancy storage or plumbing. You crank the gas pedal or throttle, and you electrically create more vapor for immediate consumption, on demand; low-high flow rate as needed, from idle to maximum power. The only real change is that you are using tap water as fuel, instead of the traditional petroleum-based fuel.

Given a choice, which way would you choose?

Frequently Asked Questions

Q: Does it really work ?
A: Yes; this is well-established technology dating back to stainless steel. But be sure to follow these instructions using the proper mechanical and electrical assembly techniques, as this plan incorporates the best qualities of several techniques.

Q: How does it qualify as "free energy"?
A: If you're paying someone for the water you use, then it is not strictly free. But the alternative is to keep buying into expen$ive ga$oline and its resultant hydrocarbon pollution.

Q: Is it safe?
A: Technically, it is safer than running on fossil fuel because you are no longer choking on your own emissions (health-wise). In general, it is practically as safe as your current gasoline arrangement. You will be installing a few simple safety devices, using current automotive standards.

Q: What kind of performance can I expect?
A: Properly adjusted, your modified vapor-only fuel system will run cooler, and at a modestly higher power level. The mileage performance expected from this design ranges from 50-300 mpg (of water), depending on your adjusting skills.

Q: Can I do the modification myself?
A: Why not? If you don't have any mechanical skills, and you know someone with basic mechanical and/or electrical skills, you can even delegate some of the construction. If you are using a fuel-injected engine, you may have to get a mechanic's opinion. [There will have to be an adapter inserted into the fuel-injection system, just as you would have to do if you were going to run on propane, hydrogen, or natural gas. Ed.)

Q: What is the environmental impact that my vehicle will have?
A: It will be producing H20 steam (water vapor) and unburnt O2 (Oxygen). Hence, it will be cleaning the environment, rather than dumping nauseous toxins into it. Plus you will be helping to save our dwindling supply of atmospheric oxygen. Any excess vapor in the reaction becomes either steam or oxygen. You can also expect to be receiving more than casual interest from those around you.

Q: Isn't this really a steam engine?
A: No. Really. Exceedingly high temperature and pressure are not used. This is strictly an internal-combustion engine (burning orthohydrogen) with residual steam in the exhaust as a by-product.

There are a few things you should know about gasoline:

Gasoline as a fuel is not necessary; it is optional.

Gasoline versus Water

There is a lot of thermochemical energy in gasoline, but there is even more energy in water. The DOE (Department of Energy) has quoted about 40%, so it is probably much more than that.

Most people are unaware that "internal combustion" is defined as "a thermo-vapor process" - as in "no liquid in the reaction." Most of the gasoline in a standard internal combustion engine is actually consumed, (cooked, and finally, broken down) in the catalytic converter after the fuel has been not-so-burnt in the engine. Sadly, this means that most of the fuel we use in this way is used only to cool down the combustion process, a pollution-ridden and inefficient means of doing that.

How It Works

Exceedingly simple. Water is pumped as needed to replenish and maintain the liquid level in the chamber. The electrodes are vibrated with a 0.5-5A electrical pulse which breaks $2(H_2O) \Rightarrow 2H_2 + O_2$. When the pressure reaches say 30-60 psi, you turn the key and go. You step on the pedal, you send more energy to the electrodes, and thus more vapor to the cylinders; i.e. fuel vapor on demand.

You set the idle max-flow rate to get the most efficient use of power, and you're off to the races.

In the big picture, your free energy is coming from the tap water in an open system, as the latent energy in the water is enough to power the engine and hence drive the alternator and whatever belt-driven accessories. And the alternator is efficient enough to run the various electrical loads (10 - 20 amps), including the additional low current to run this vapor reaction. No extra batteries are required.

STEP BY STEP CONSTRUCTION (Please refer to diagrams at the end of this document)

OVERVIEW - Here is the suggested sequence of steps:

1. Install the CHT (or EGT) gauge and measure your current operating temp range (gasoline), for comparison.

2. Build and test the controller to verify the correct pulse output.

3. Build the reaction chamber and test it with the controller (i.e pressure out).

4. Install the tank, controller, chamber, and pressure fittings.

5. Run engine and adjust the control circuit as necessary for best performance.

6. Install the stainless steel valves and get the pistons/cylinders coated with ceramic.

7. Coat the exhaust system with ceramic without the catalytic converter (or let it rust out and then replace the whole dang thang with stainless steel pipe sections).

YOU WILL NEED

plastic water tank with pump and level sensor.
control circuit, wiring, connectors, and epoxy.
reaction chamber with electrodes and fittings.
3/8" stainless steel flex-tubing, fittings and clamps.
carb/FI vapor-pressure fitting kit. - pressure, CHT (or EGT), & level gauges.
stainless steel valves.
copper mesh junction.
ceramic surface treatment for cylinders & pistons.
stainless steel or ceramic treated exhaust assembly.

BASIC TOOLS

drill, screwdriver and pliers
hole cutter
wire-wrap, solder-iron and clippers
DVM and oscilloscope.

REACTION CHAMBER

Construct as shown in the diagrams. Use a section of 4" PVC waste pipe with a threaded screw-cap fitting on one end and a standard end-cap at the other. Make sure to drill-and-epoxy or tap threads thru the PVC components for all fittings. Set and control the water level in the chamber so that it well submerses the pipe electrodes; yet leave some headroom to build up the hydrogen/oxygen vapor pressure. Use stainless steel wires inside the chamber or otherwise use a protective coating; use insulated wires outside. Ensure that the epoxy perfects the seal, or otherwise lay down a bead of water-proof silicone that can hold pressure.

The screw fitting may require soft silicone sealant, or a gasket; its purpose is to hold pressure and allow periodic inspection of the electrodes. No leaks, no problems. Make sure you get a symmetric 1-5mm gap between the 2 stainless steel pipes. The referenced literature suggests that the closer to 1mm you get, the better. You will want to get your chamber level sensor verified before you epoxy the cap on.

Make your solder connections at the wire/electrode junctions nice, smooth, and solid; then apply a water-proof coating, e.g. the epoxy you use for joining the pipes to the screw cap. This epoxy must be waterproof and be capable of holding metal to plastic under pressure. You will want to get your chamber level sensor verified before you epoxy the cap on.

CONTROL CIRCUIT

The diagrams show a simple circuit to control and drive this mini-system. You are going to make a 'square-pulse' signal that 'plays' the electrodes like a tuning fork; which you can watch on an oscilloscope. The premise given by the literature is: the faster you want do go down the road, the 'fatter' you make the pulses going into the reaction chamber. Duty cycle will vary with the throttle in the vicinity of 90%MARK 10%SPACE (OFF/ON).

There is nothing sacred about how the pulse waveform is generated; there are many ways to generate pulses, and the attached diagrams show a few. The diagram shows the NE555-circuit approach from the referenced patent. The output switching transistor must be rated for 1-5 amps @ 12VDC (in saturation).

Go with a plan that works for you or your friendly neighborhood technoid or mechanic, and go get all the circuit elements from your local electronics store, such as Radio-Shack or Circuits-R-Us, including the circuit board, IC sockets, and enclosure/box.

DigiKey has better selection, service, and knowledge; plus they have no minimum order. Be sure to use a circuit board with a built-in ground plane, and to accommodate room for mounting 2 or 3 of the gauges. Mounting the reaction chamber in the engine compartment will require running a stub to your pressure gauge where you can watch it.

You can easily make 30-gauge wire-wrap connections between the socket pins and thru-hole discrete components having wire leads. Also make sure to get spec sheets on any IC you use. More details of the best circuits to use will be announced pending prototype testing. You will want to get your chamber level sensor verified before you epoxy the cap on.

Throttle Control

If you have a throttle position sensor, you should be able to access the signal from the sensor itself OR from the computer connector. This signal is input to the circuit as the primary control (i.e. throttle level = pulse width = vapor rate).

If you don't have such a signal available, you will have to rig a rotary POT (variable resistor) to the gas linkage (i.e. coupled to something at the gas pedal or throttle cable running to the carb or FI. If you make the attachment at the carb/FI, be sure to use a POT that can handle the engine temp cycles. Don't use a cheezy-cheapy POT; get one rated for long life and mechanical wear; mount it securely to something sturdy and stationary that will not fall apart when you step on the gas.

Control Range. The full throttle RANGE (idle-max) MUST control the vapor rate, i.e. pulse-width (duty). The resistor values at the throttle signal must allow the throttle signal voltage, say 1-4 Volt swing, to drive the VAPOR RATE. You will be using this voltage swing to generate a 10% ON 'square' pulse. The patent implies using a 'resonant' pulse in the 10-250 KHz frequency range; but it is not explicitly stated so.

In this circuit, you will simply tune to whatever frequency makes the most efficient vapor conversion. You will have to get into the specs for each IC you use, to insure you connect the right pins to the right wires, to control the frequency and pulse width. You can use spare sockets to try out different discrete component values. Just keep the ones that are spec-compatible in the circuit, and get the job done.

You crank up the throttle signal and put more electrical energy (fatter pulses) into the electrodes; verify you can get 10% duty on the scope (2 - 100 usec on the horizontal time-base). Your averaging DVM will display the 90%-10% DC voltage across the output transistor (Vce or Vds or Output to Ground). Set and connect DVM in the supply current and measure .5 - 5 amps, without blowing the DVM fuse. Now verify that you got everything you wanted.

Verify your wiring connections using your DVM as a continuity detector. Check your wiring 1 at a time and yellow line your final schematic as you go. You can best use board-mount miniature POTs for anything you want to set-and-forget. The LEDs are there to give you a quick visual check of normal vs abnormal operation of your new creation. You will want to get your chamber level sensor verified before you epoxy the cap on.

CARB/FI CONNECTION

The diagram also shows that fittings are required to the carb/FI l. There are ready-made kits (such as by Impco) available for making your pressure fittings to the carburetor or fuel-injector as the case may be. You will necessarily be sealing the built-in vents and making a 1-way air-intake.

The copper mesh comprises the inadvertent backfire' protection for the reaction chamber. Make sure that all vapor/duct junctions are air-tight and holding full pressure without leakage. Your new 'system' is considered successful and properly adjusted when you get the full power range at lower temp and minimum vapor flow without blowing the pressure safety valve.

CHT (or EGT)

Monitor your engine temp with the CHT (cylinder head temp) or EGT (exhaust gas temp) instead of your original engine temp indicator (if any). Your existing gauge is too slow for this application and will not warn you against overheating until after you have burnt something. Make sure that your engine runs no hotter than in the gasoline arrangement. VDO makes a CHT gauge with a platinum sensor that fits under your spark plug against the cylinder head (make sure it is really clean before you re-install your spark plug (as this is also an electrical ground).

ENGINE/EXHAUST TREATMENT

Get the valves replaced with stainless steel ones and get the pistons/cylinders ceramic-treated ASAP when you have successfully converted and run your new creation. Do not delay as these items will rust, either by sheer use or by neglect (i.e. letting it sit). You could make max use of your current exhaust system by using it with your new deal until it rusts through, then have your mechanic or welder friend to fit a stainless steel exhaust pipe (no catalytic converter is required). But it could be easier and cheaper to send your existing exhaust system out for the ceramic treatment, and then simply re-attach it to the exhaust ports.

GENERAL

1. Do not discard or remove any of the old gasoline setup components, e.g. tank, carb/FI, catalytic converter, unless necessary. Better to always leave an easy way to revert back to something that at least runs, just in case. Some people are leaving their gasoline setup completely intact, and switching back and forth at will, just to have a backup plan.

2. Set your throttle circuit so that you get minimum vapor flow at idle, and maximum vapor flow at full power without blowing the pressure relief valve. In this way, you control how 'lean' your mixture is by the strength of the pulse (i.e. “fatness” at the optimum pulse frequency).

3. If you just don't get enough power (at any throttle setting), it means that you need to (1) change the pulse frequency, (2) change the gap between the electrodes, (3) change the size (bigger) electrodes, or (4) make a higher output pulse voltage (last resort). Always use an output transistor, such as a MOSFET, that is rated for the voltage and current you need to get the job done. OK so you might have to play around with it some. Isn't that where all the Fun is anyhow?

4. If you get any engine knock our loud combustions (not compensated by adjusting the timing), it means that you need to install an additional coil in the chamber, and drive the coil with an additional pulse signal (about 19 Hz on the .1sec time base (see diagram). Here,

you will be slowing down the burn rate just enough so that the vapors burn thru out the power stroke of the piston. Be sure to include a board-mount POT to set the correct strength of this 2nd pulse signal into the coil. This is a stainless steel coil of about 1500 turns (thin wire) that you can arrange like a donut around the center pipe (but NOT touching either electrode), directly over the circular 1-5mm gap. You want no knocking at any power/throttle setting; smooth power only, but also no excess hydrogen leftover from the combustion.

5. Build the canister(s) as tall as you can without compromising your ability to mount them conveniently near the dash panel, or in the engine compartment, as the case may be. This way, you can always make the electrodes bigger, if necessary without undue hardship. Remember that anything in the engine compartment should be mounted in a bullet-proof, vibration and temperature tolerant fashion.

6. If you have to drill a thru-hole for wiring or plumbing thru metal, make sure to also install a grommet for protection against chafing. Always watch your chamber pressure range from IDLE (15-25 psi) - FULL POWER (30-60 psi). Set your safety-pressure relief-valve to 75 psi and make sure it's rated for much higher.

7. Shut OFF the power switch and pull over if there is any malfunction of the system. Your engine will last longest when it still develops FULL POWER+ at some minimum temperature that we are sure you can find, by leaning back the Royal Vapor Flow and/or by making use of the water-vapor cooling technique (see diagram). Keep good mpg performance records, and periodic maintenance/inspection. Keep it clean; save some money; clean the air; heal the planet; happy motoring; tell a friend; enjoy your freedom and self-empowerment.

8. There lacks documented material for perfecting this vapor system thru a fuel injector; there may be some details you will discover on your own as working prototypes progress. For example, you may be restricted to inject the hydrogen/oxygen vapor without any water vapor, as it may rust the injectors. If engine temp and CHT is a problem, then you will want to re-think your plan, e.g. ceramic-coating the injectors. There is always “replacing the FI system with a Carb.”

9. If you install the water-vapor system (for lower operating temp/stress), you will want to lean the mixture (vapor/air) for minimum vapor flow rate to achieve any given throttle position (idle - max). Make sure that you get a minimum flow for IDLE and a modestly sufficient flow for MAX, that does the cooling job without killing the combustion.

10. If you cannot find stainless steel pipe combinations that yield the 1-5mm gap, you can always regress back to alternating plates of +/- electrodes.

11. If you are concerned about the water freezing in your system, you can (a) add some

98% isopropyl alcohol and re-adjust the pulse frequency accordingly; or (b) install some electric heating coils.

12. Do not let ANYONE ever compromise your dream, your freedom, your independence or your truth.

REFERENCES

Stephen Chambers 'Apparatus for Producing Orthohydrogen and/or Parahydrogen' US Patent 6126794, uspto.gov
Stanley Meyer 'Method for the Production of a Fuel Gas' US Patent 4936961, uspto.gov
Creative Science & Research, 'Fuel From Water', fuelless.com
Carl Cella "A Water-Fuelled Car" Nexus Magazine Oct-Nov 1996
Peter Lindemann "Where in the World is All the Free Energy", free-energy.cc <http://www.free-energy.cc/>
George Wiseman "The Gas-Saver and HyCO Series" eagle-research.com <http://www.eagle-research.com/>
C. Michael Holler "The Dromedary Newsletter" and "SuperCarb Techniques"
Stephen Chambers "Prototype Vapor Fuel System" xogen.com <http://www.xogen.com/>

COMMON LAW COPYRIGHT #285714: All rights to the use and duplication of these plans are hereby reserved for the People, in their efforts to heal and restore the environment. Dare to express your uniqueness and environmental ideals. This technology is an exercise in responsible self-determination.

DISCLAIMER: The Spirit of Ma'at LLC and the Spirit of Ma'at ezine and the author of this document assumes no liability for the use or misuse of this information; which is made available as public-domain information and free of charge, for the purposes of education, ecology, health, well-being, freedom, liberty, and pursuit of happiness.

CONVERT YOUR ENGINE TO BURN HYDROGEN AND OXYGEN VAPOR FROM WATER....... ON DEMAND + POLLUTION-FREE

FIGURE 1

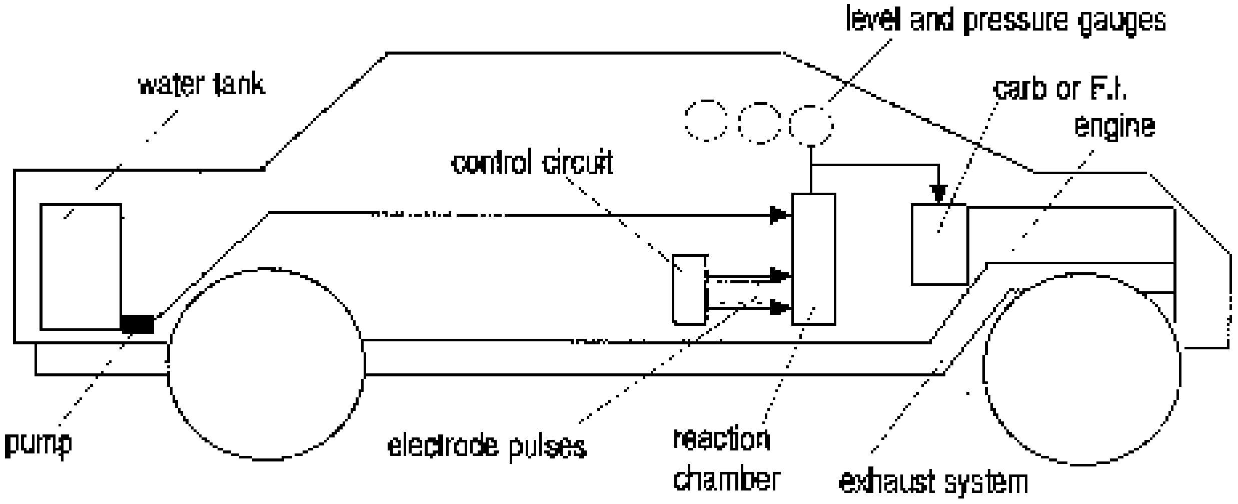
level and pressure gauges
water tank
carb or F.I.
engine
control circuit
pump
electrode pulses
reaction
chamber
exhaust system

BUILD YOUR OWN WATER VAPOR SYSTEM CONTROLLER FROM THE-SHELF MATERIALS

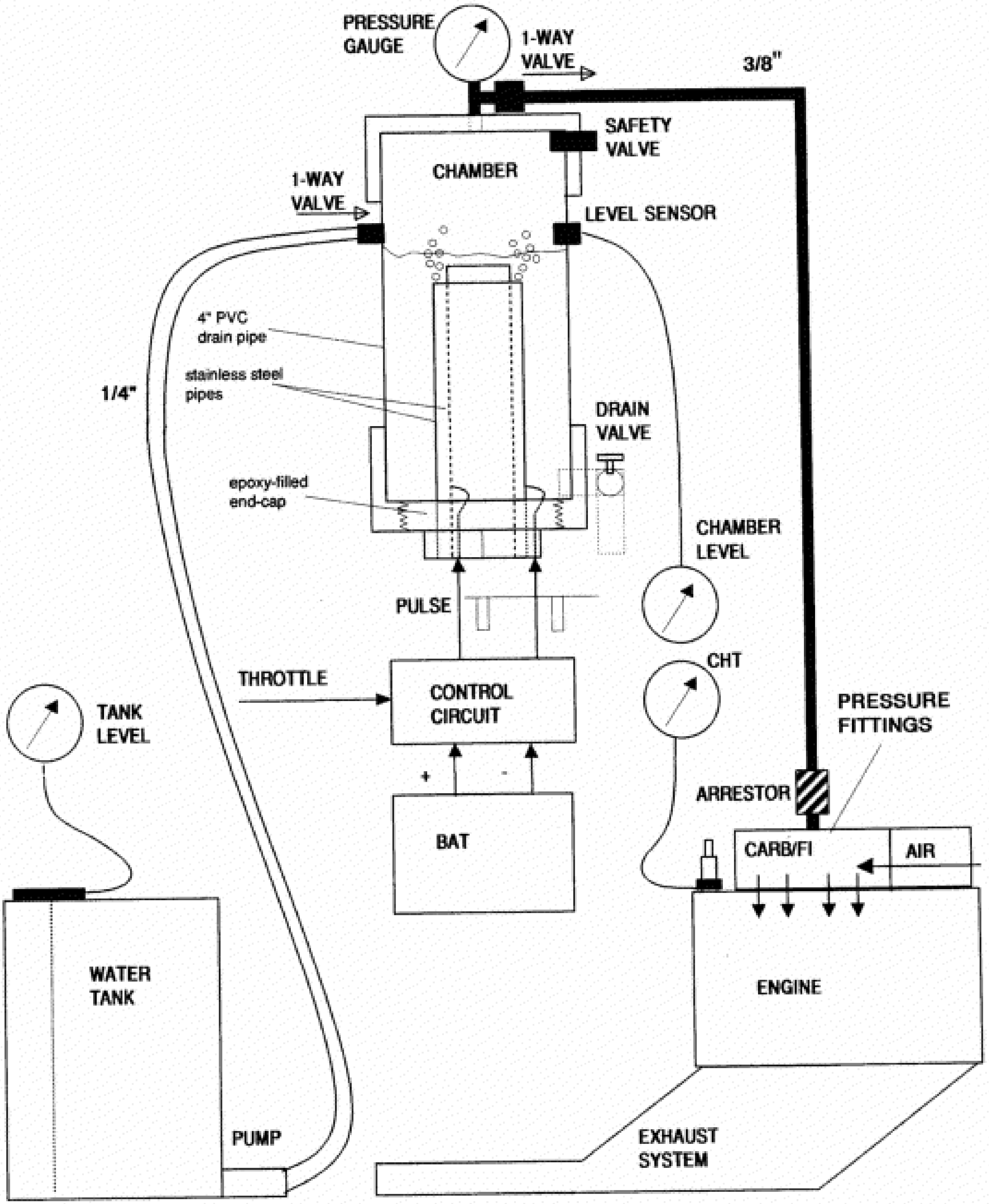
PRESSURE GAUGE
1-WAY VALVE
3/8"
SAFETY VALVE
CHAMBER
1-WAY VALVE
LEVEL SENSOR
4" PVC drain pipe
stainless steel pipes
1/4"
DRAIN VALVE
epoxy-filled end-cap
CHAMBER LEVEL
PULSE
THROTTLE
CONTROL CIRCUIT
CHT
PRESSURE FITTINGS
TANK LEVEL
+
-
ARRESTOR
BAT
CARB/FI
AIR
WATER TANK
ENGINE
PUMP
EXHAUST SYSTEM

FIGURE 2 BLOC DIAGRAM

BUILD YOUR OWN WATER VAPOR SYSTEM CONTROLLER FROM THE-SHELF MATERIALS

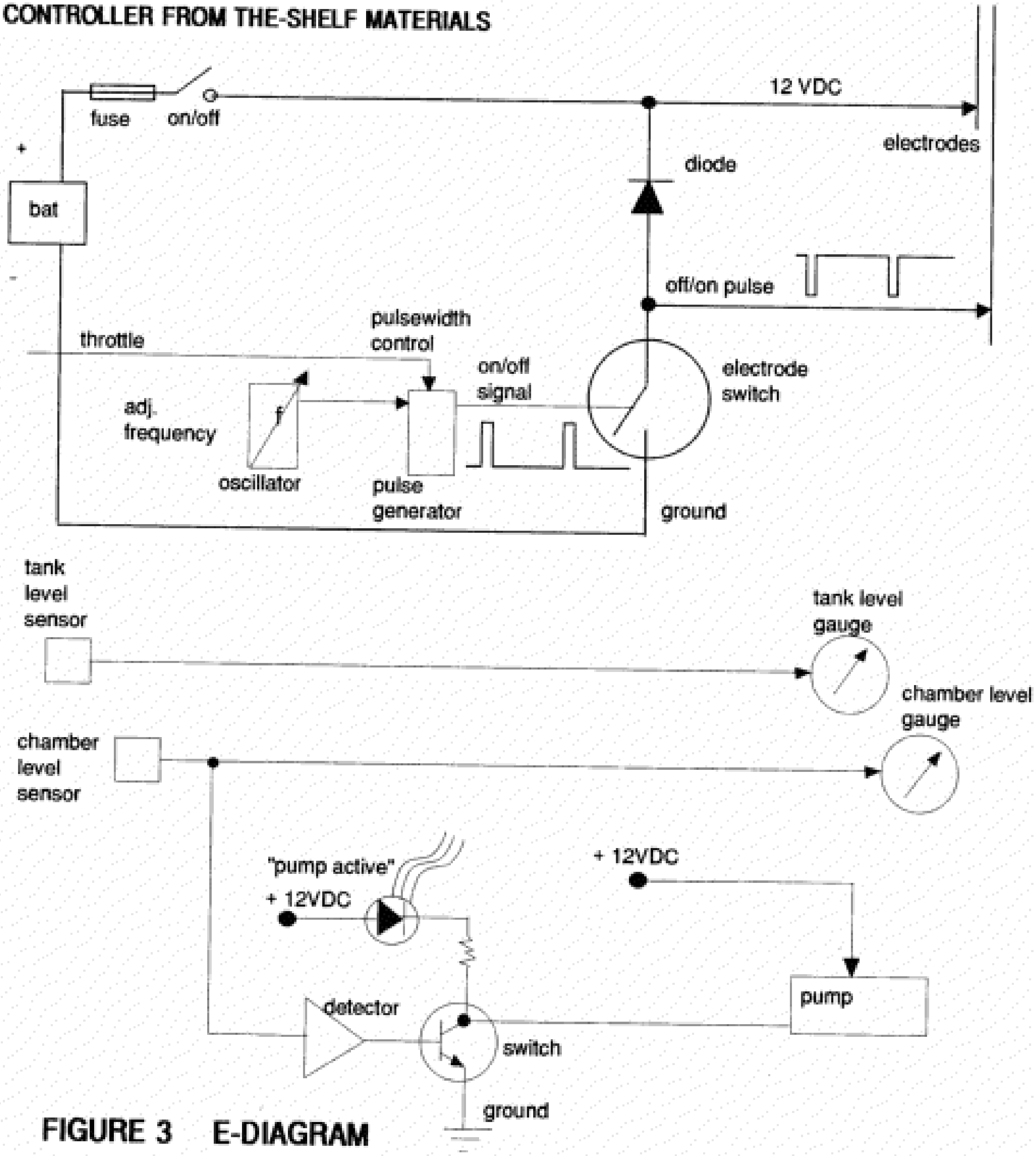
fuse
on/off
12 VDC
electrodes
+
bat
-
diode
off/on pulse
throttle
pulsewidth
control
on/off
signal
electrode
switch
adj.
frequency
f
oscillator
pulse
generator
ground
tank
level
sensor
tank level
gauge
chamber level
gauge
chamber
level
sensor
"pump active"
+ 12VDC
+ 12VDC
pump
detector
switch
ground

FIGURE 3 E-DIAGRAM

BUILD YOUR OWN WATER VAPOR SYSTEM CONTROLLER FROM THE-SHELF MATERIALS

REV. A

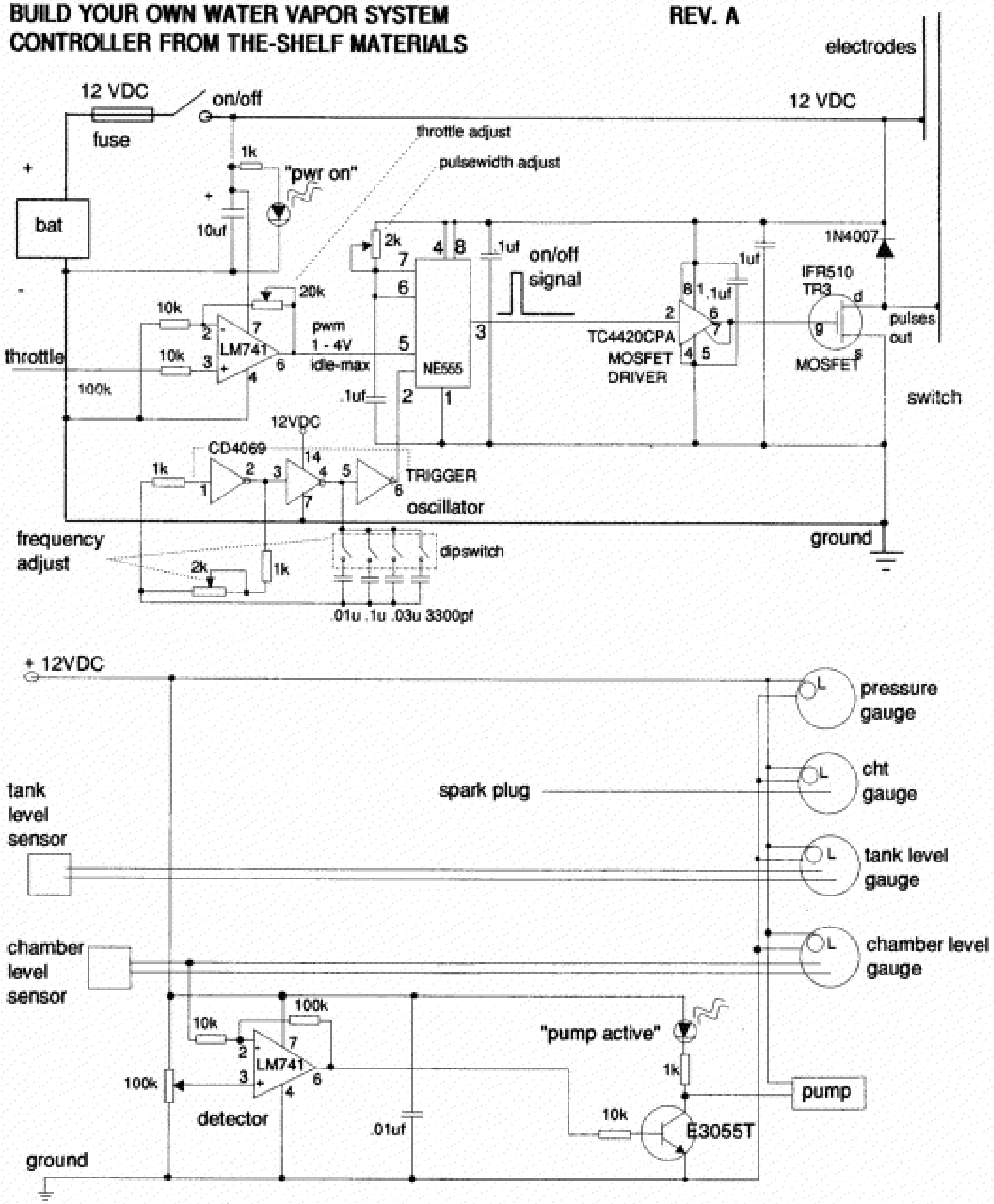
electrodes
12 VDC
on/off
fuse
12 VDC
+
bat
-
1k
"pwr on"
+
10uf
throttle adjust
pulsewidth adjust
2k
7
4
8
.1uf
on/off
signal
1N4007
IFR510
TR3
d
g
s
pulses
out
MOSFET
switch
.1uf
8
1
.1uf
2
6
7
4
5
TC4420CPA
MOSFET
DRIVER
20k
10k
2
-
7
LM741
3
+
6
4
10k
throttle
100k
pwm
1 - 4V
idle-max
6
5
3
NE555
.1uf
2
1
12VDC
CD4069
1k
1
2
3
14
4
7
5
6
TRIGGER
oscillator
ground
frequency
adjust
2k
1k
dipswitch
.01u
.1u
.03u
3300pf
+ 12VDC
pressure
gauge
L
cht
gauge
spark plug
tank
level
sensor
tank level
gauge
chamber
level
sensor
chamber level
gauge
100k
10k
2
-
7
LM741
3
+
6
4
100k
detector
.01uf
"pump active"
1k
pump
10k
E3055T
ground

FIGURE 4 E-SCHEMATIC

BUILD YOUR OWN WATER VAPOR SYSTEM
CONTROLLER FROM THE-SHELF MATERIALS

REV. A

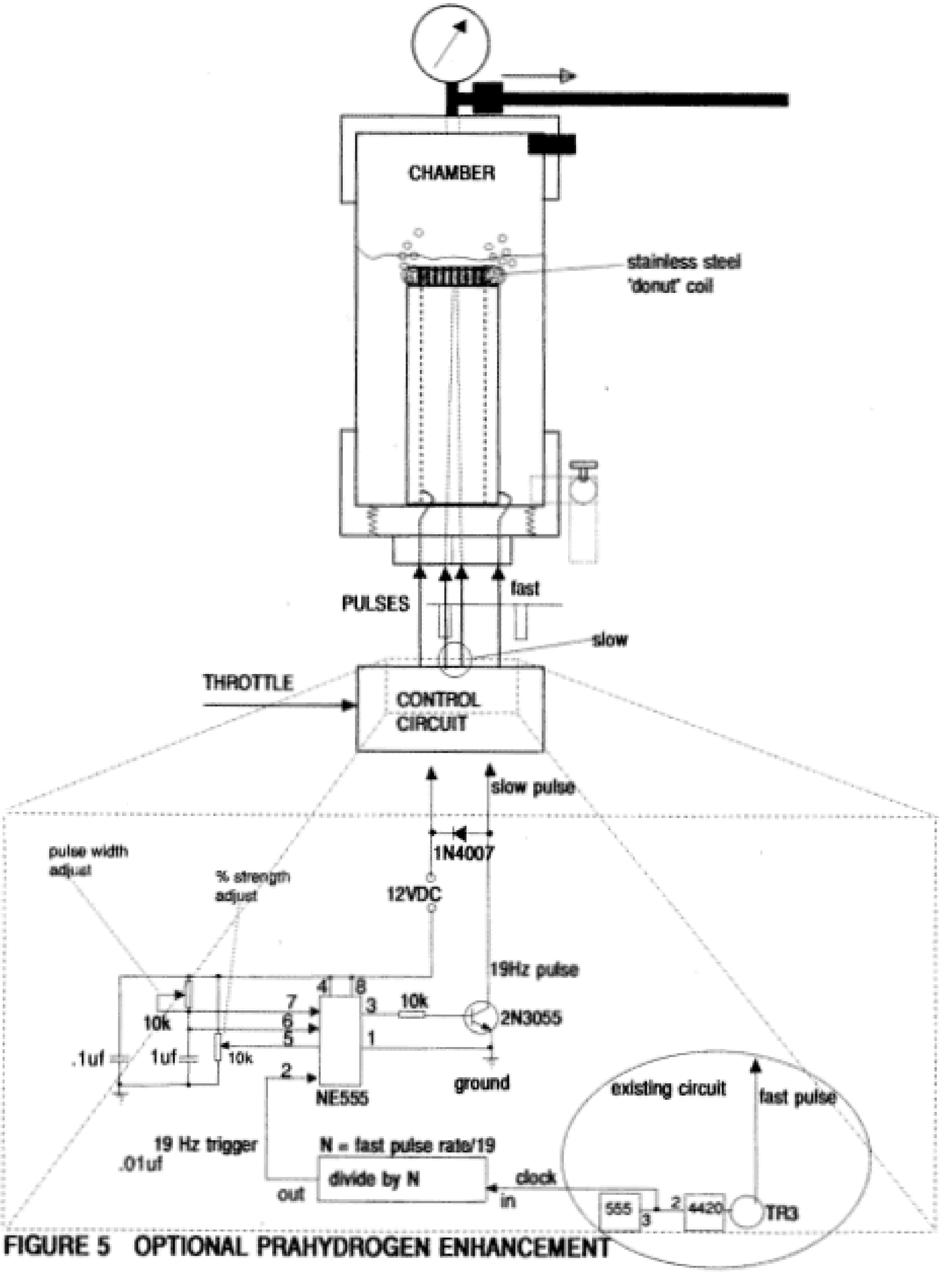
CHAMBER
stainless steel 'donut' coil
PULSES
fast
slow
THROTTLE
CONTROL CIRCUIT
slow pulse
1N4007
pulse width adjust
% strength adjust
12VDC
19Hz pulse
4
8
7
6
5
3
1
2
10k
10k
10k
2N3055
.1uf
1uf
10k
ground
NE555
19 Hz trigger
.01uf
N = fast pulse rate/19
divide by N
out
clock
in
existing circuit
fast pulse
555
3
2
4420
TR3

FIGURE 5 OPTIONAL PRAHYDROGEN ENHANCEMENT

BUILD YOUR OWN WATER VAPOR SYSTEM CONTROLLER FROM THE-SHELF MATERIALS

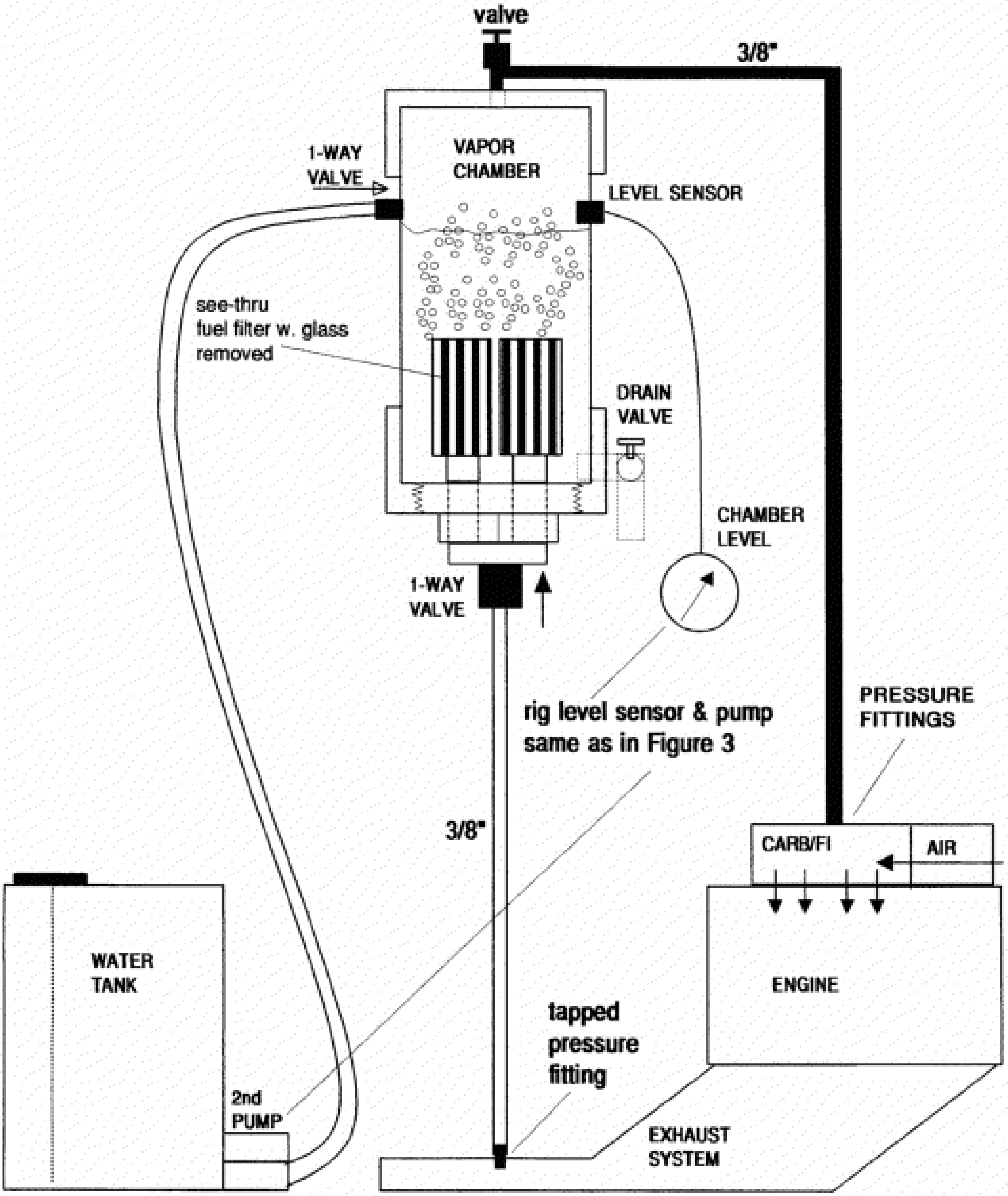
valve
3/8"
VAPOR CHAMBER
1-WAY VALVE
LEVEL SENSOR
see-thru fuel filter w. glass removed
DRAIN VALVE
CHAMBER LEVEL
1-WAY VALVE
rig level sensor & pump same as in Figure 3
PRESSURE FITTINGS
3/8"
CARB/FI
AIR
WATER TANK
ENGINE
tapped pressure fitting
2nd PUMP
EXHAUST SYSTEM

FIGURE 6 OPTIONAL WATER COOLING

An idle telephone line tempts designers to use its 48V potential as a power source. However, Part 68 of the US Federal Communications Commission's telecommunications regulations states that any device that connects to the phone line and is not actively communicating must present a resistance of at least 5 MΩ (**Reference 1**). To meet this requirement, a device's continuous-current drain must not exceed 10 μA. Fortunately, many devices that connect to the phone line do not require continuous power and can remain off for long intervals, awakening only for a short time before relapsing into power-off mode. Providing power for these applications from the phone line presents obvious advantages by eliminating the need for a battery or another power source and the cost of battery maintenance.

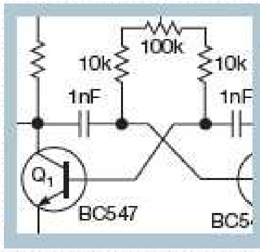
10k
100k
10k
1nF
1nF
Q_1
BC547
BC5

Click here to find out how to submit your own Design Idea.

The circuit in **Figure 1** charges a 1.5F supercapacitor, C_1, from the phone line through a diode bridge and a 5.6-MΩ resistor. A **Maxim MAX917** nanopower comparator, IC_1, consumes only 0.75 μA from its power supply. Resistors R_2 and R_3 halve the voltage across C_1 and apply it to IC_1's positive input voltage at Pin 3 for comparison with its built-in 1.245V reference. For voltages across C_1 that do not exceed 2.49V, IC_1's output at Pin 6 remains low. When C_1's voltage reaches 2.5V, Pin 3's voltage exceeds the reference voltage, and IC_1's output goes high, turning on Q_1 and Q_2.

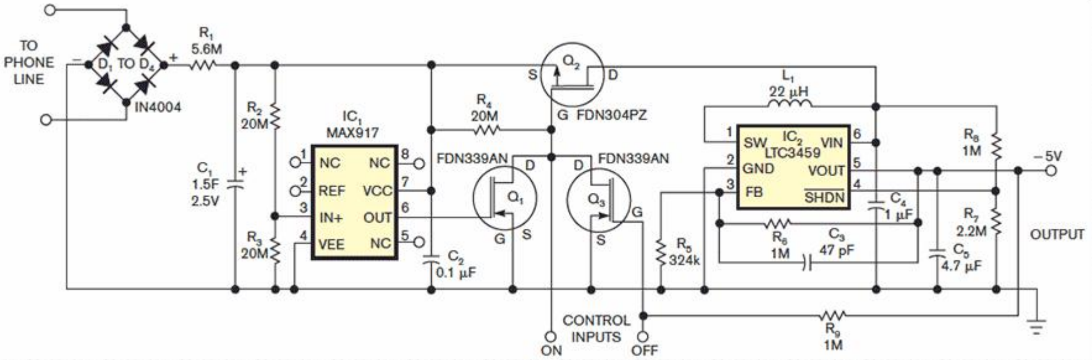
TO PHONE LINE
D_1 TO D_4
IN4004
R_1 5.6M
R_2 20M
C_1 1.5F 2.5V
R_3 20M
IC_1 MAX917
NC
REF
IN+
VEE
NC
VCC
OUT
NC
C_2 0.1 μF
R_4 20M
Q_2
S
D
G FDN304PZ
FDN339AN
Q_1
FDN339AN
Q_3
R_5 324k
L_1 22 μH
IC_2 LTC3459
SW
GND
FB
VIN
VOUT
SHDN
R_6 1M
C_3 47 pF
C_4 1 μF
R_8 1M
R_7 2.2M
C_5 4.7 μF
–5V
OUTPUT
CONTROL INPUTS
ON
OFF
R_9 1M

Figure 1 This power-conversion circuit delivers intermittent bursts of regulated voltage from a supercapacitor charged by a trickle of current from a telephone line.

Several days must elapse before C_1 becomes fully charged, given its huge capacitance and a charging current of less than 10 µA. The voltage on C_1 can never exceed 2.5V because, once it reaches 2.49V, Q_1 and Q_2 turn on, connecting C_1 to a switched-mode-power-supply circuit. Because the power-supply current exceeds the charging current, the voltage across C_1 starts to decrease when Q_2 turns on. Transistor Q_3 holds Q_2 on when C_1's decreasing voltage causes Q_1 to turn off.

The switched-mode-power-supply circuit comprises a **Linear Technology LTC3459** micropower boost converter, IC_2, and its associated components, which deliver 5V at 10 mA. A fully charged C_1 can supply power to a 10-mA load for approximately 40 sec. With no load, the circuit can sustain its 5V output for more than 10 hours. For greater output current and shorter operating time, select another boost converter that can operate at a low input voltage.

READ MORE designideas

Mechanical switches, open-drain MOSFETs, open-collector transistors, or a microcontroller's open-drain output pins can drive two external control inputs to force the circuit on and off. Pulling the On input low forces Q_2 to turn on and deliver power from C_1 to the power converter, and pulling the Off input low turns off Q_2 and removes power from the converter. Note that the power converter's output-return line connects to the telephone line and thus should not connect to an earth ground or to grounded equipment.

Reference

1. "**Part 68**," Federal Communications Commission.

Share G+ Tweet Like 0

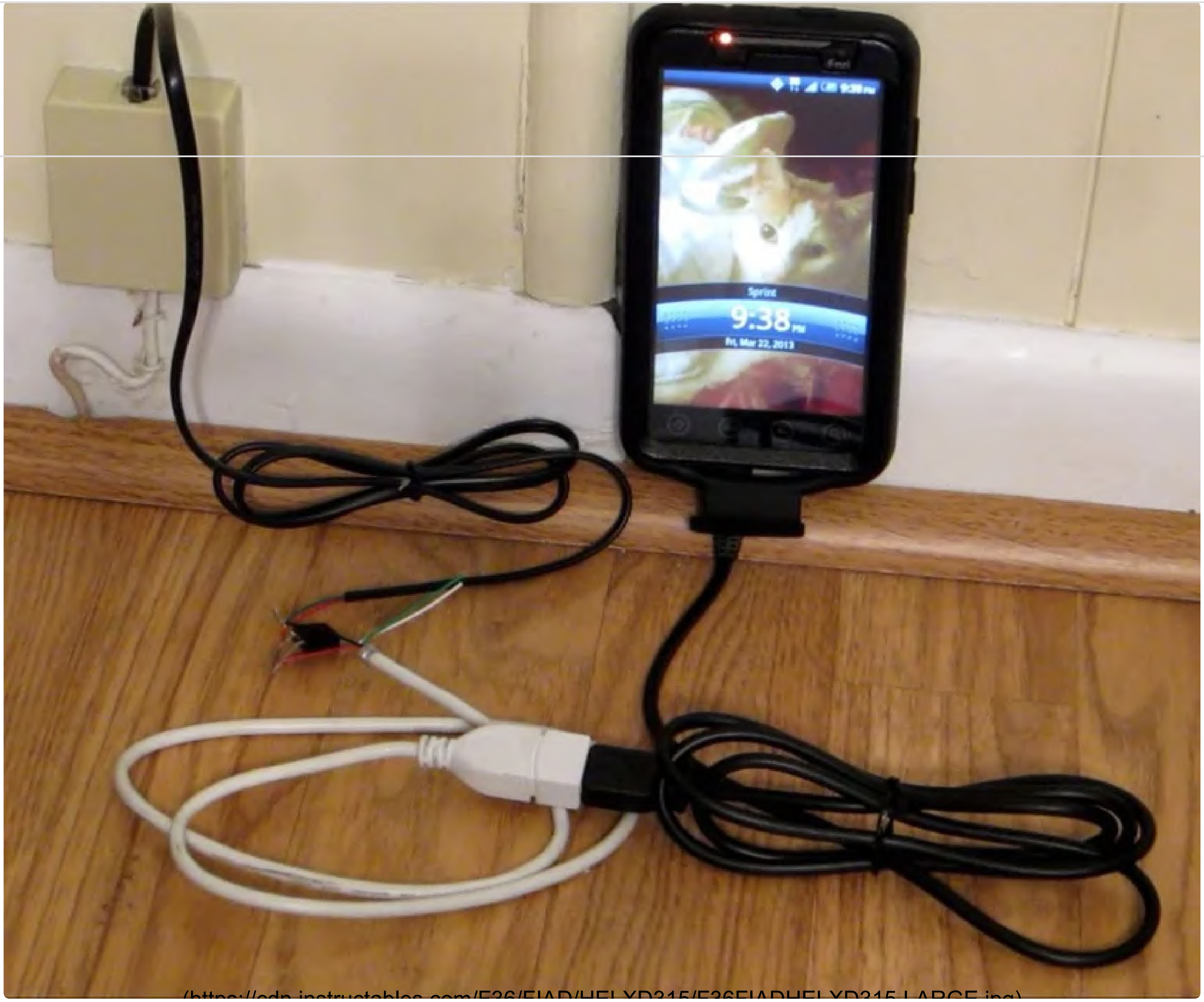

(https://cdn.instructables.com/F36/FIAD/HELXD315/F36FIADHELXD315.LARGE.jpg)

Download Favorite

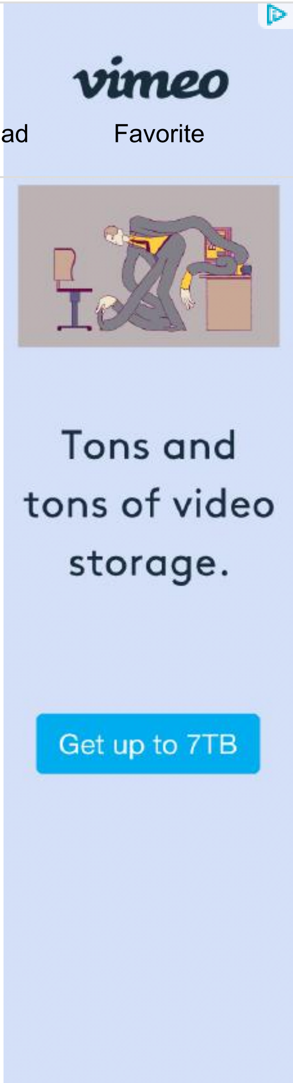
vimeo
Tons and tons of video storage.
Get up to 7TB

advertisement

(/member/DIY%20Hacks%20and%20How%20Tos/)

By **DIY Hacks and How Tos (/member/DIY%20Hacks%20and%20How%20Tos/)**

DIY Hacks & How Tos on YouTube (http://www.youtube.com/user/DIYHacksandHowTos)

Follow

More by the author:

HELLO
I AM BAYMAX
YOUR PERSONAL
HEALTHCARE
COMPANION
BIG

6 THINGS TO COOK
IN A COFFEE MAKER

Festivus
Pole

not in use, this is a constant DC signal (about 50-60 volts). When the phone rings, the signal is a 20 hertz AC signal (about 90 volts). When in use it is a modulated DC signal (between 6 and 12 volts).

The phones lines even have power during a blackout in most cases. This is because the phone company maintains their own backup power system. Your phone lines may be powered even if you don’t have a land line service set up.

Add Tip Ask Question Comment Download

Step 2: Check the Phone Line With a Multimeter

Download Favorite

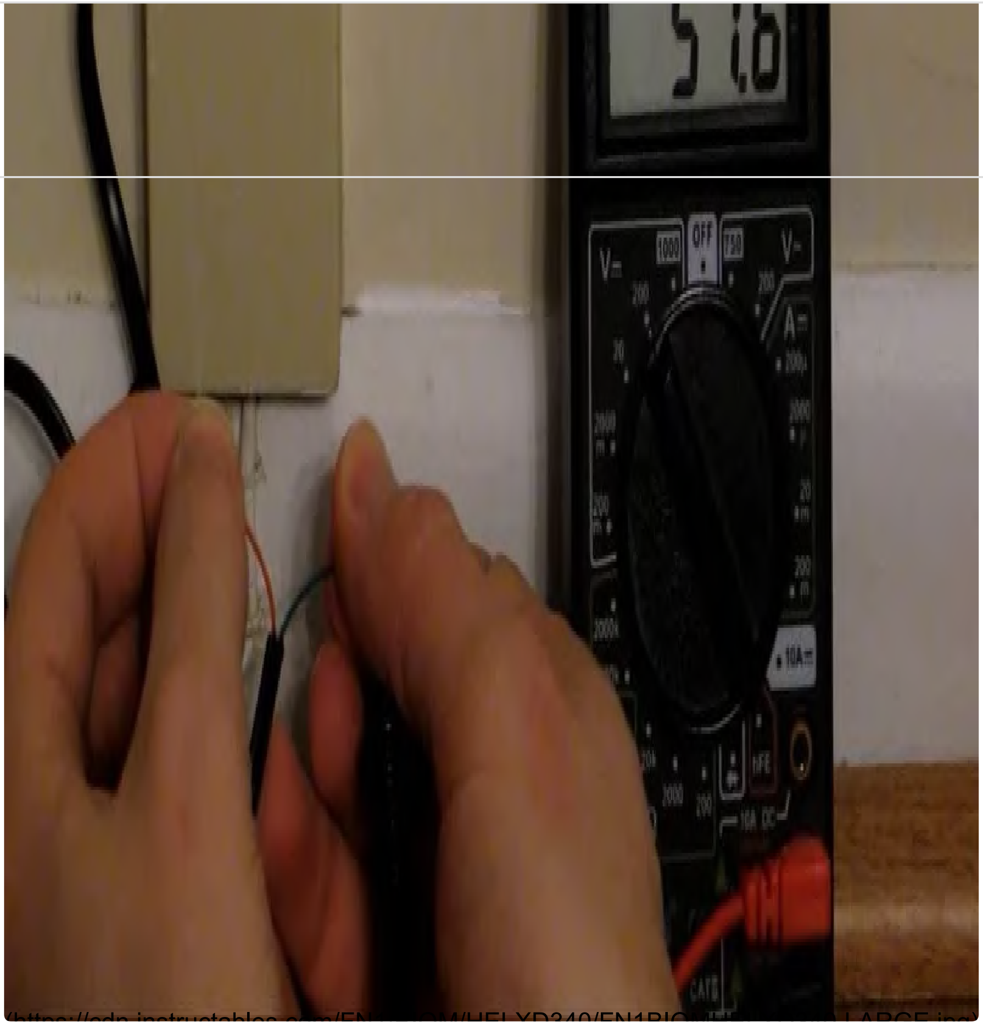

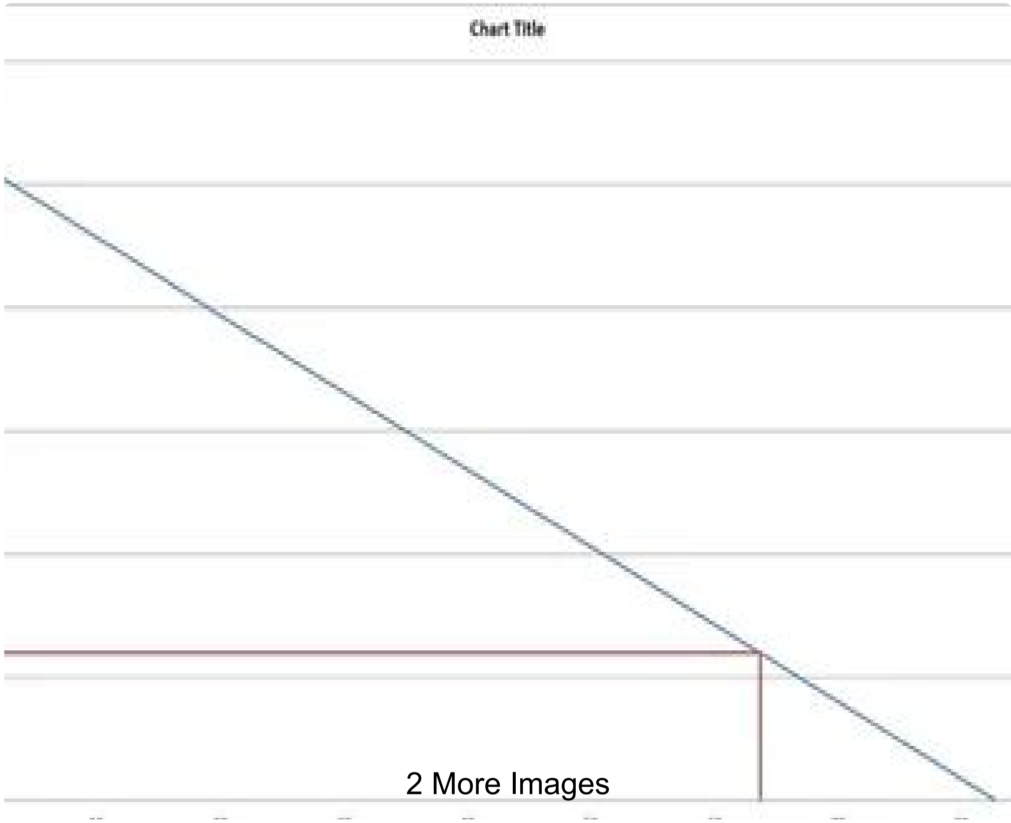

2 More Images

Before you try to tap into the electricity in the phone line, you should check it with a multimeter to see what you are working with.

Start by cutting open a phone cord and separating the internal wires. In most cases you will have one red wire and one green wire. Strip the insulation off the ends. Then plug the cord into a phone jack and use a multimeter to measure the output voltage. At my house, I measured an open-circuit (no load) voltage (http://en.wikipedia.org/wiki/Open-circuit_voltage) of 52 volts DC.

Then I hooked up various resistors to see what the output would be with different loads. I determined that the supply voltage isn't regulated. This means that the voltage changes depending on the resistance of the circuit that it is powering. After some calculating, I worked out that the base signal coming out of my phone jack pretty closely resembles a 52 Volt DC source (http://en.wikipedia.org/wiki/Voltage_source) with a 628 ohm internal resistance (http://en.wikipedia.org/wiki/Internal_resistance).

Basically this means that I can run a 12V circuit at 64mA, a 9V circuit at 68mA, or a 5V circuit at 75mA. This isn't a lot. But it is enough to charge a cell phone.

Add Tip Ask Question Comment Download

Step 3: Construct a Simple Voltage Regulator Circuit

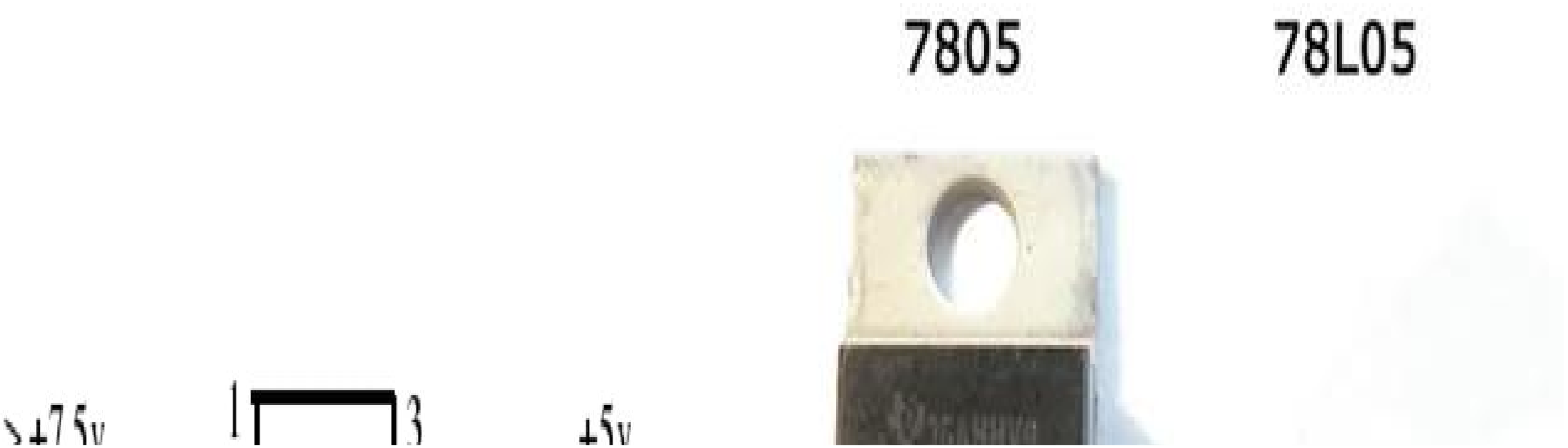

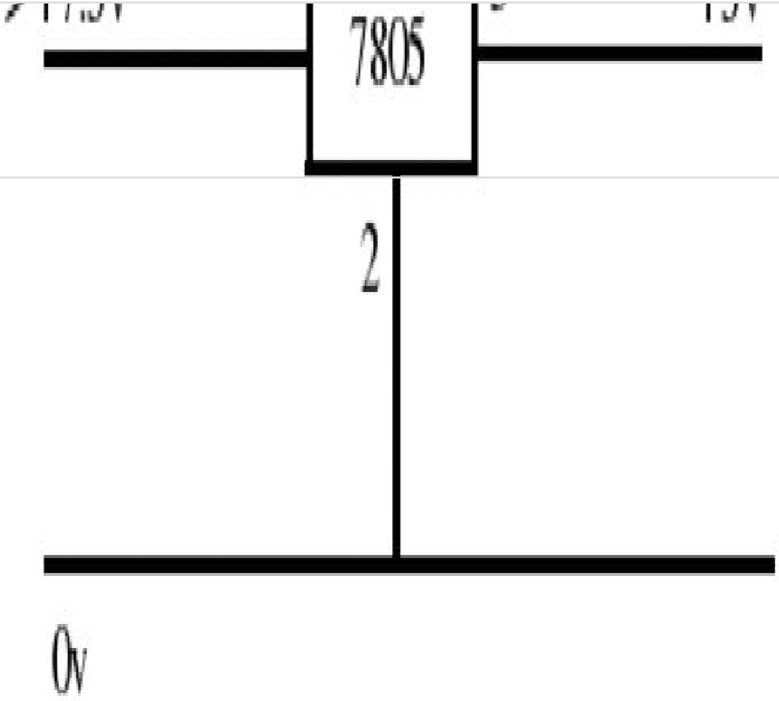

(https://cdn.instructables.com/FFV/H9M3/HELXD34J/FFVH9M3HELXD34J.LARGE.gif)

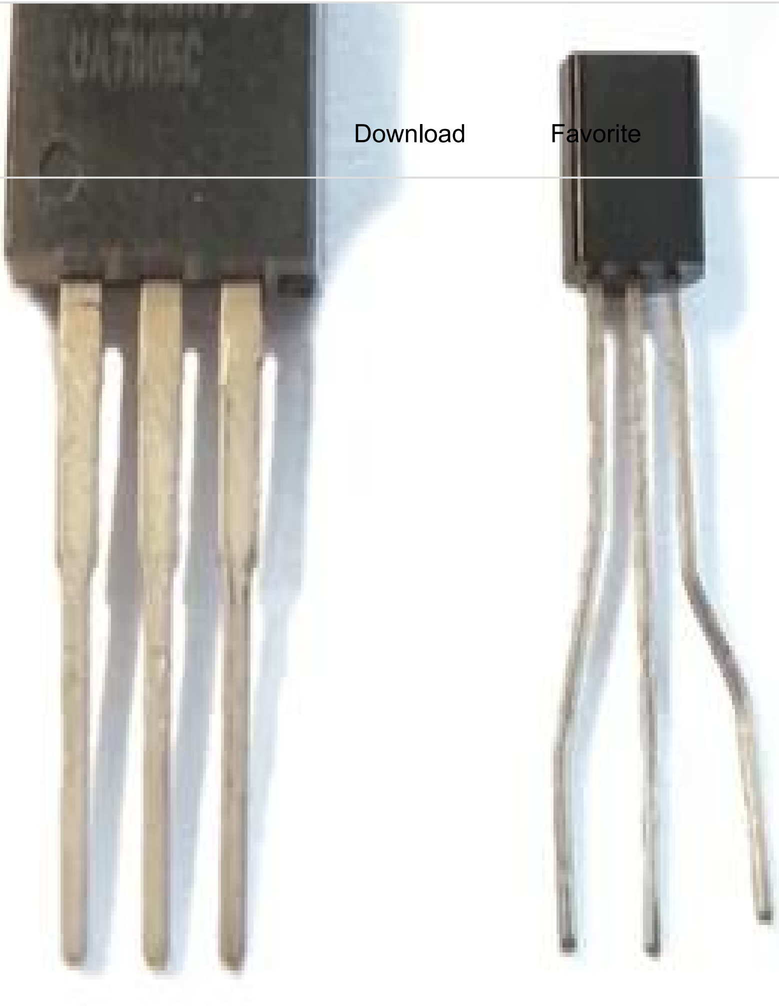

(https://cdn.instructables.com/FXB/5YW1/HEI4EJY4/FXB5YW1HEI4EJY4.LARGE.jpg)

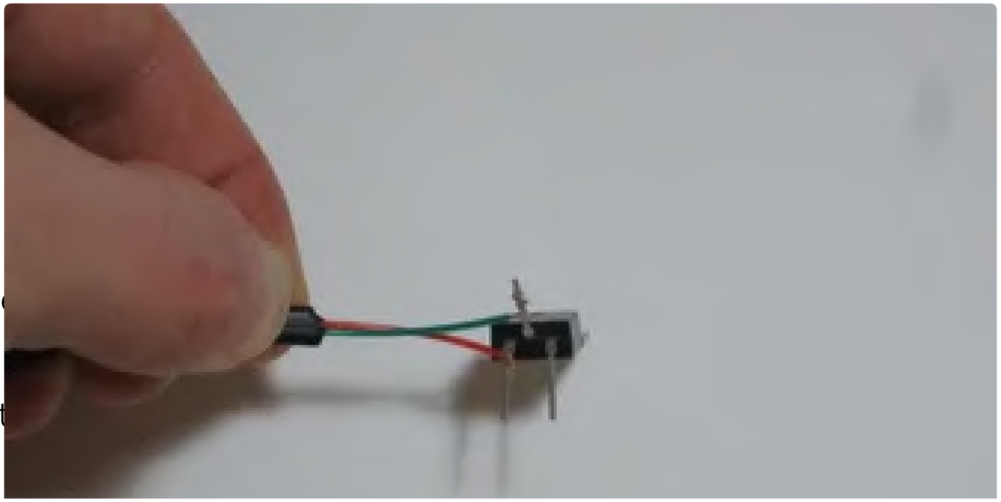

We know that the phone needs 5 volts in order to charge. But much current it draws or it's equivalent [load resistance (http://en.wikipedia.org/wiki/Electrical_load)](http://en.wikipedia.org/wiki/Electrical_load). So we can't charg from the phone line. We need to use a [voltage regulator (http://en.wikipedia.org/wiki/Voltage_regulator)](http://en.wikipedia.org/wiki/Voltage_regulator) to bring the out down to 5 volts and keep it there. A LM7805 5 volt regulator s

3 More Images
Download Image
Favorite
(https://cdn.instructables.com/FAB/XDW2/HFLXD34Y/FABXDW2HFLXD34Y.LARGE.jpg)

To make this simple phone line adapter you will need the phon been working with, the 5V voltage regulator and a USB connector cable with a female end. Just connect the red wire from the phone line to the first lead on the regulator and connect the green wire from the phone line to the second lead. Then connect the black wire from the USB cable to the second lead on the regulator and connect the red wire from the USB cable to the third lead on the regulator. If you can't solder the wires together (because the power is out), you can just wrap the wires around each lead. If you do this, you should bend the leads of the regulator away from each other. This will help you avoid accidentally crossing the wires.

This simple regulator circuit is able to safely convert the base phone signal into something that can be used to charge your phone. However, many voltage regulators are not able to handle the AC signal that they would receive if the phone rang. So if you are worried that you might receive a call while your regulator is hooked up to the phone line , then you may wish to add a diode between the red wire from the phone line and the first pin on the voltage regulator. This will protect your circuit from problems that may be caused by reverse polarity.

Add Tip Ask Question Comment Download

Step 4: Use Other Regulators for Other Output Voltages

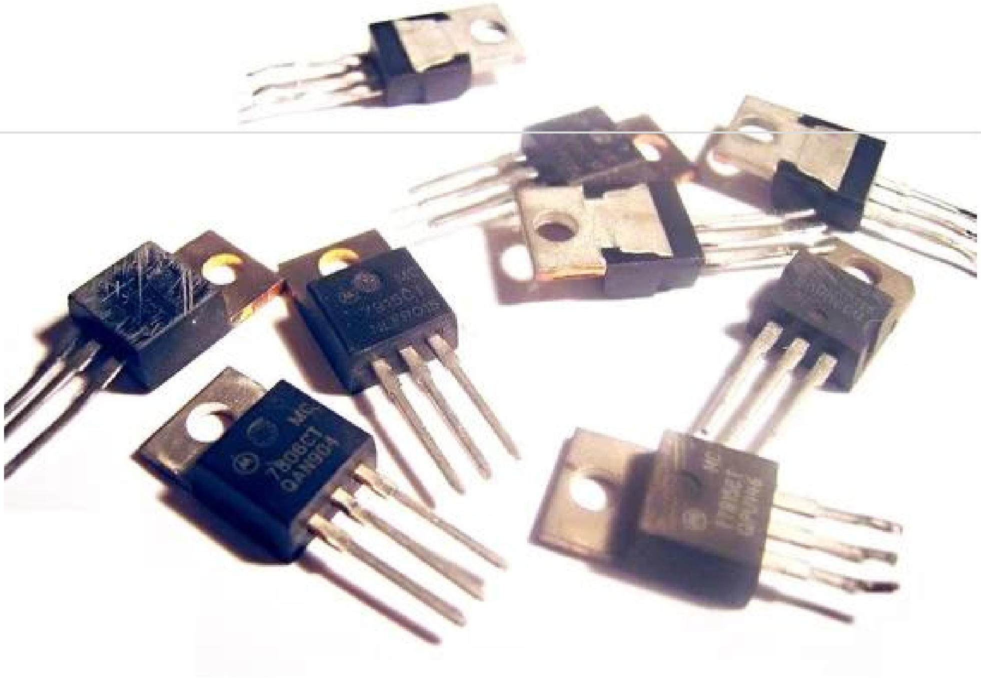

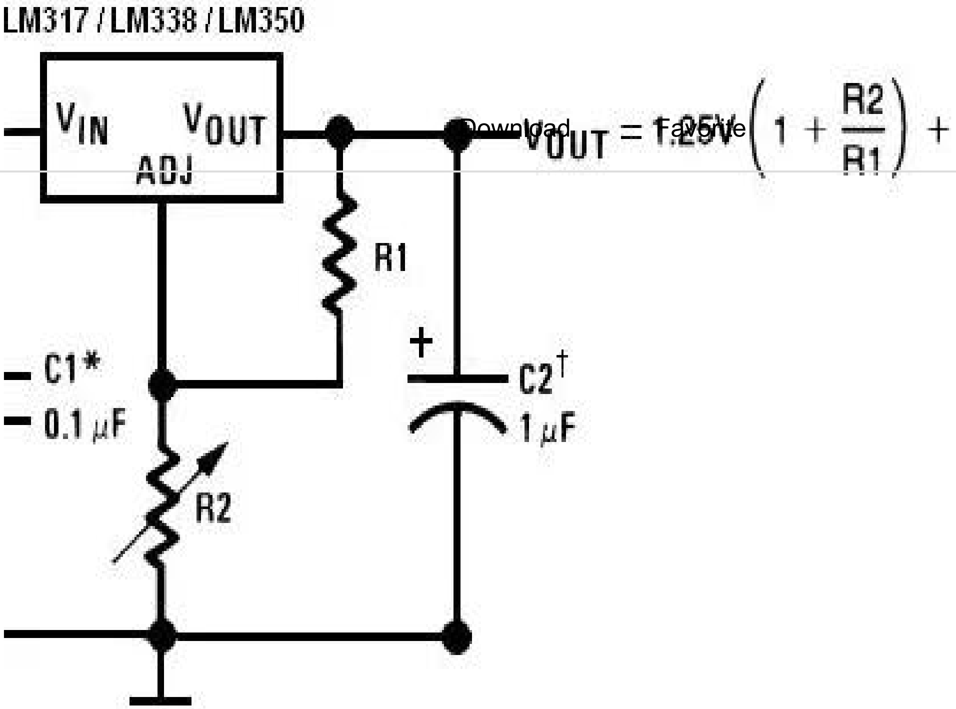

A 7805 regulator will work if you need an output of 5 volts but other kinds of voltage of voltage regulators are also available. Other voltages in the 78xx series include 6V, 8V, 9V, 10V, 12V, 15V, 18V, and 24V. In addition to these fixed value regulators, there are also variable regulators that let you set the voltage level with the use of a few external components. One such variable voltage regulator is the LM317 (http://www.ti.com/lit/ds/symlink/lm117.pdf). These are what you would use if you needed a different output voltage.

Add Tip Ask Question Comment Download

Step 5: Finished Phone Line Adapter Tool

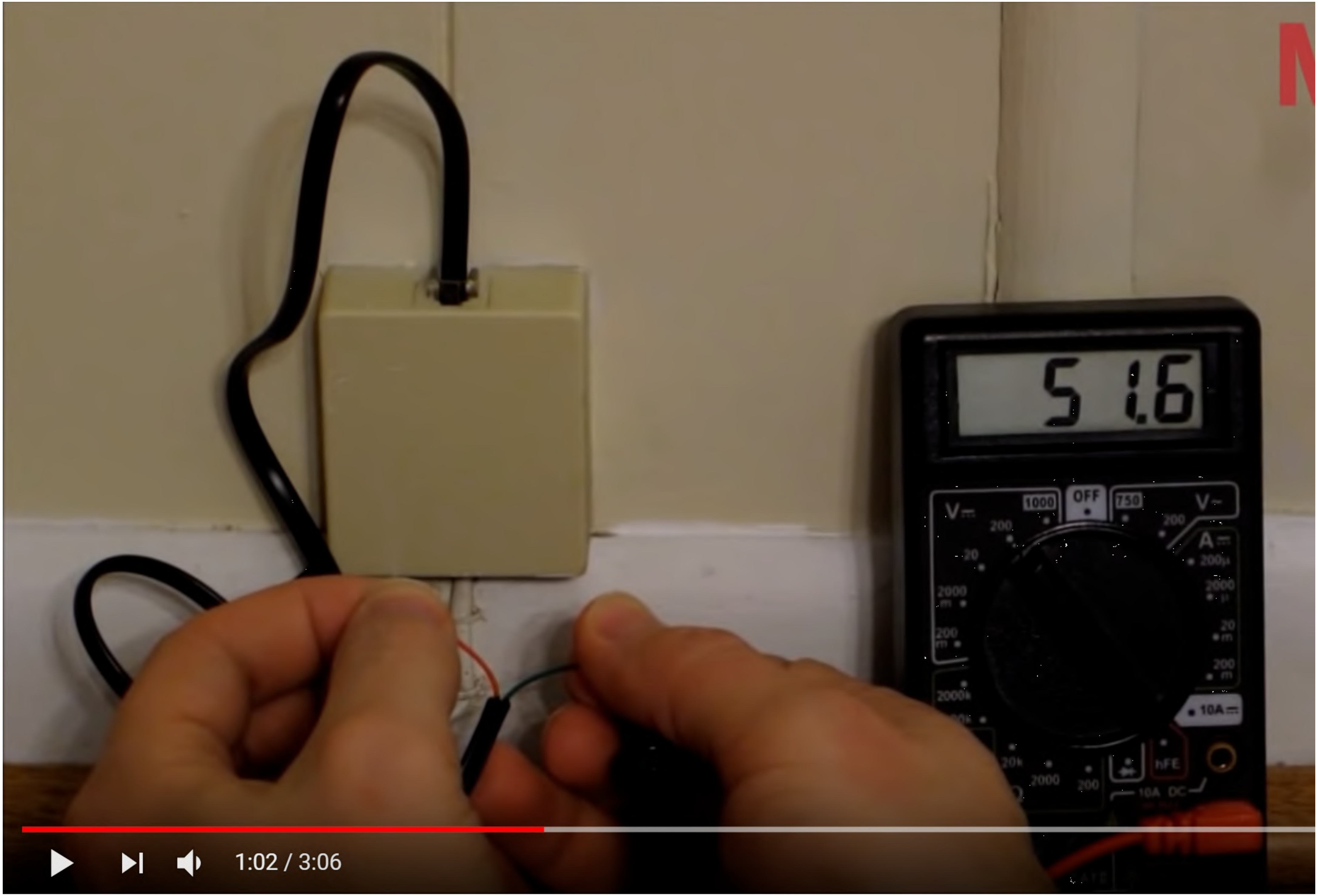
51.6
OFF
1000
750
V⎓
V~
200
20
2000 m
200 m
A⎓
200µ
2000 µ
20 m
200 m
10A⎓
2000k
20k
2000
200
hFE
10A DC
1:02 / 3:06

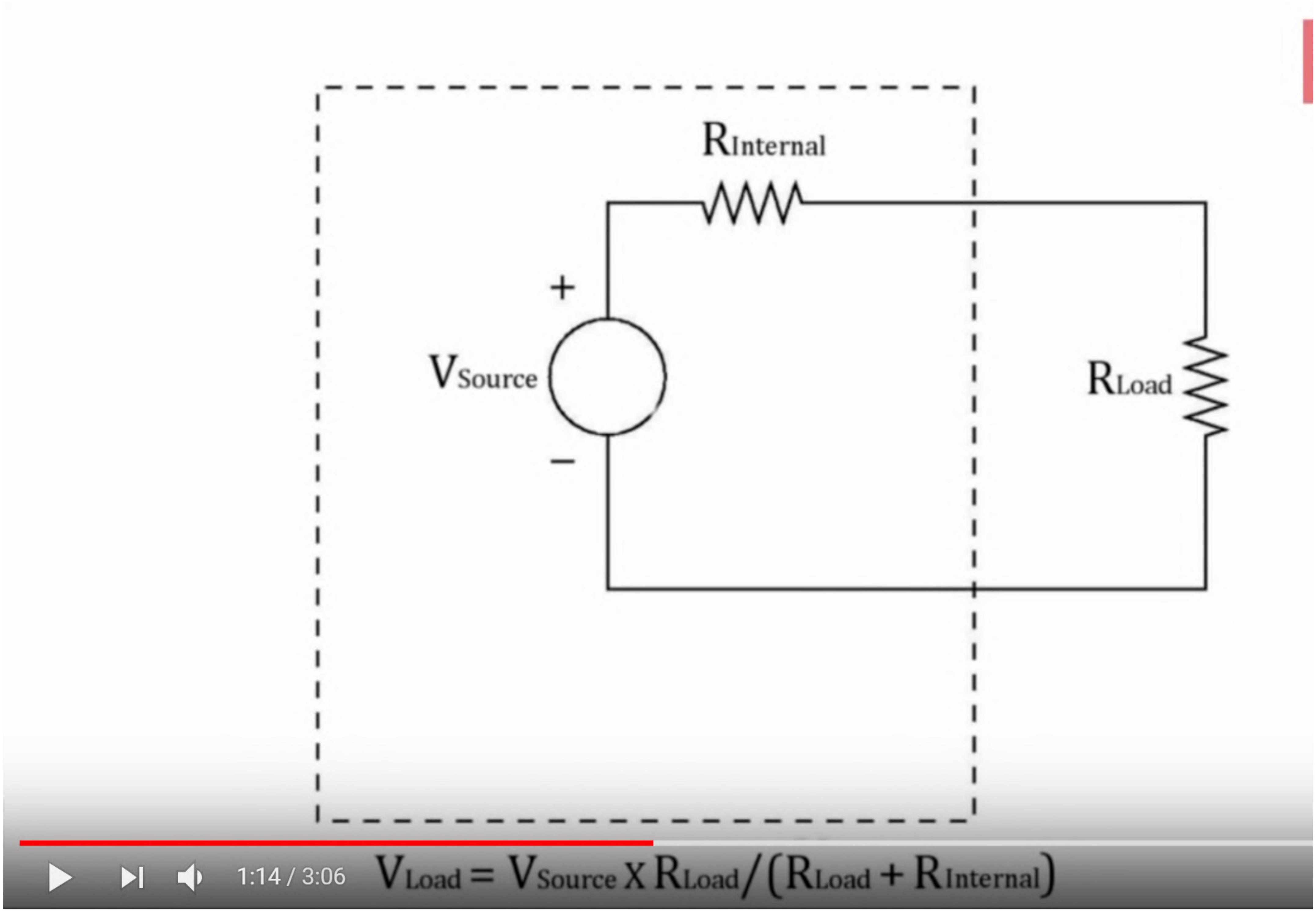
$R_{Internal}$
+
V_{Source}
−
R_{Load}
1:14 / 3:06

$$V_{Load} = V_{Source}\ \mathrm{X}\ R_{Load}/(R_{Load} + R_{Internal})$$

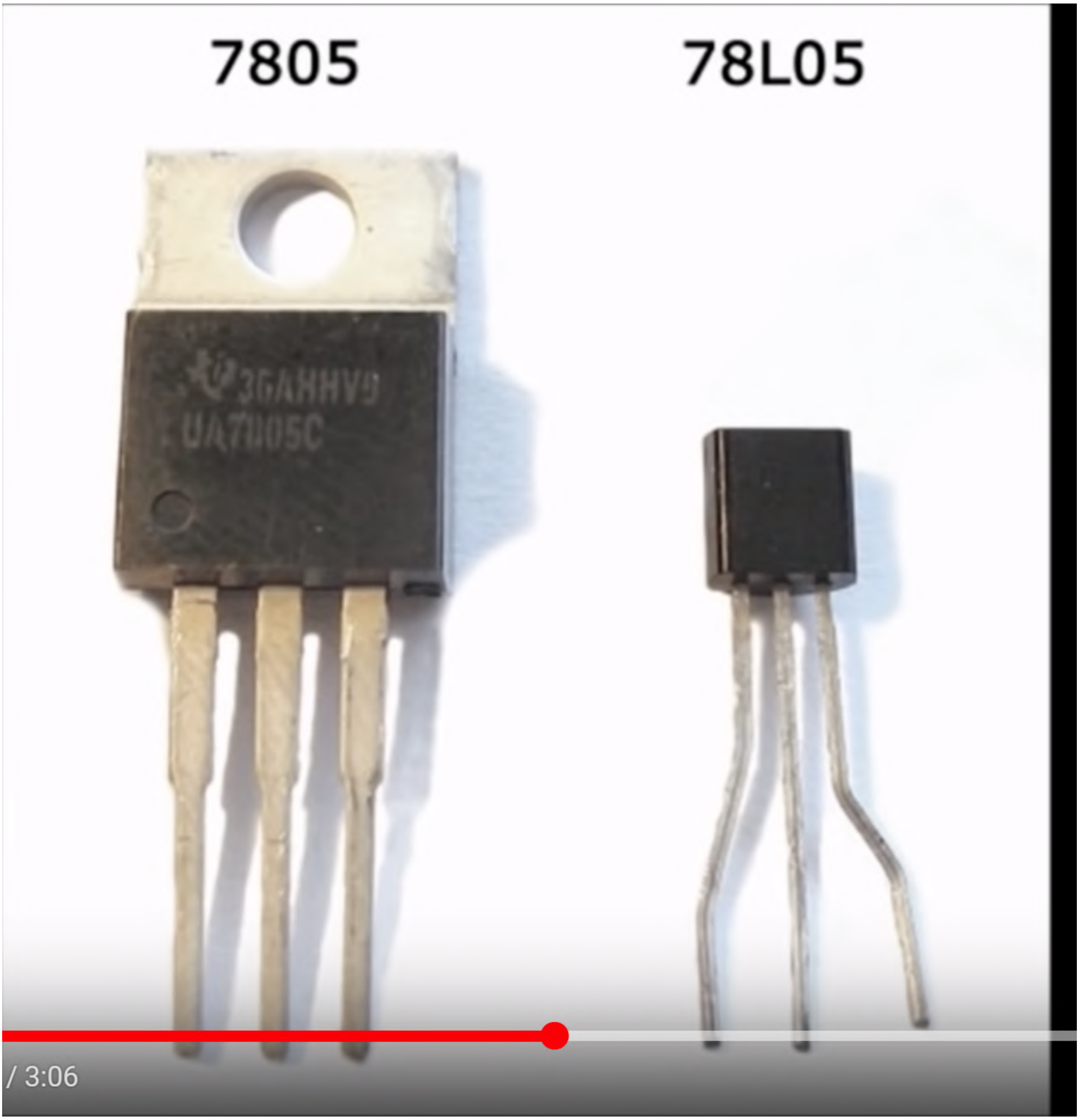
7805
78L05
36AHHV9
UA7805C
/ 3:06

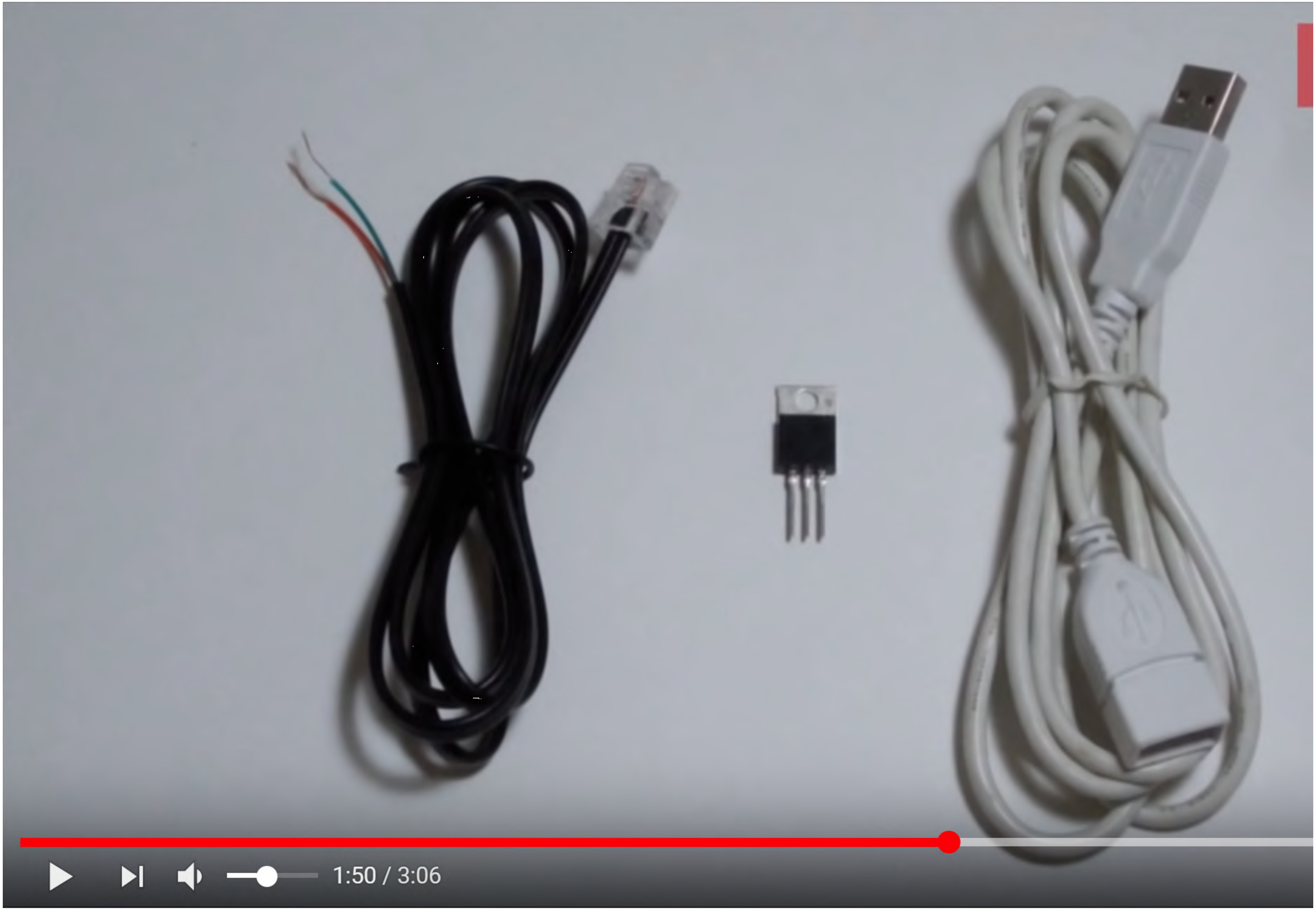
1:50 / 3:06

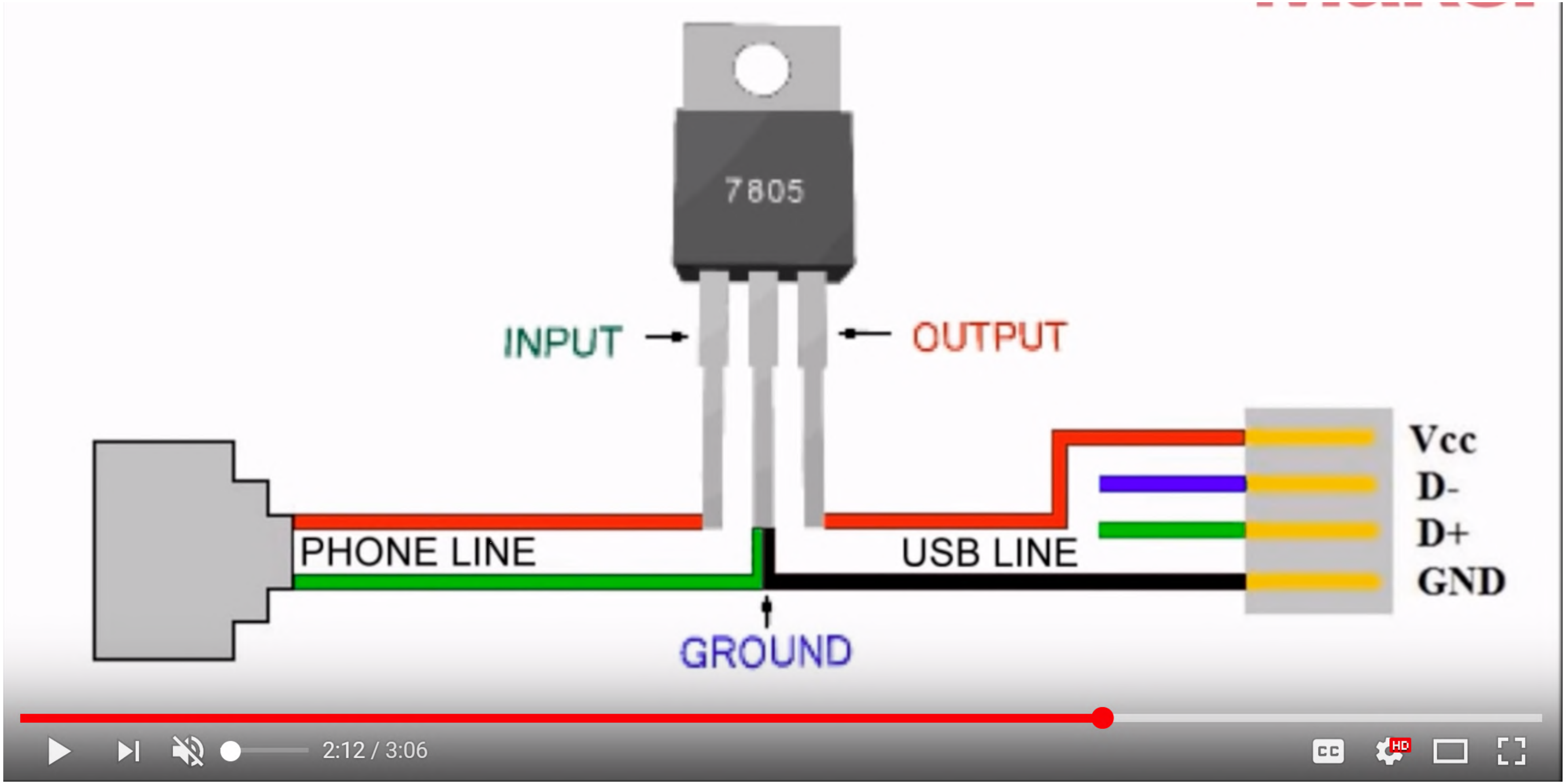
7805
INPUT
OUTPUT
Vcc
D-
D+
GND
PHONE LINE
USB LINE
GROUND
2:12 / 3:06
HD

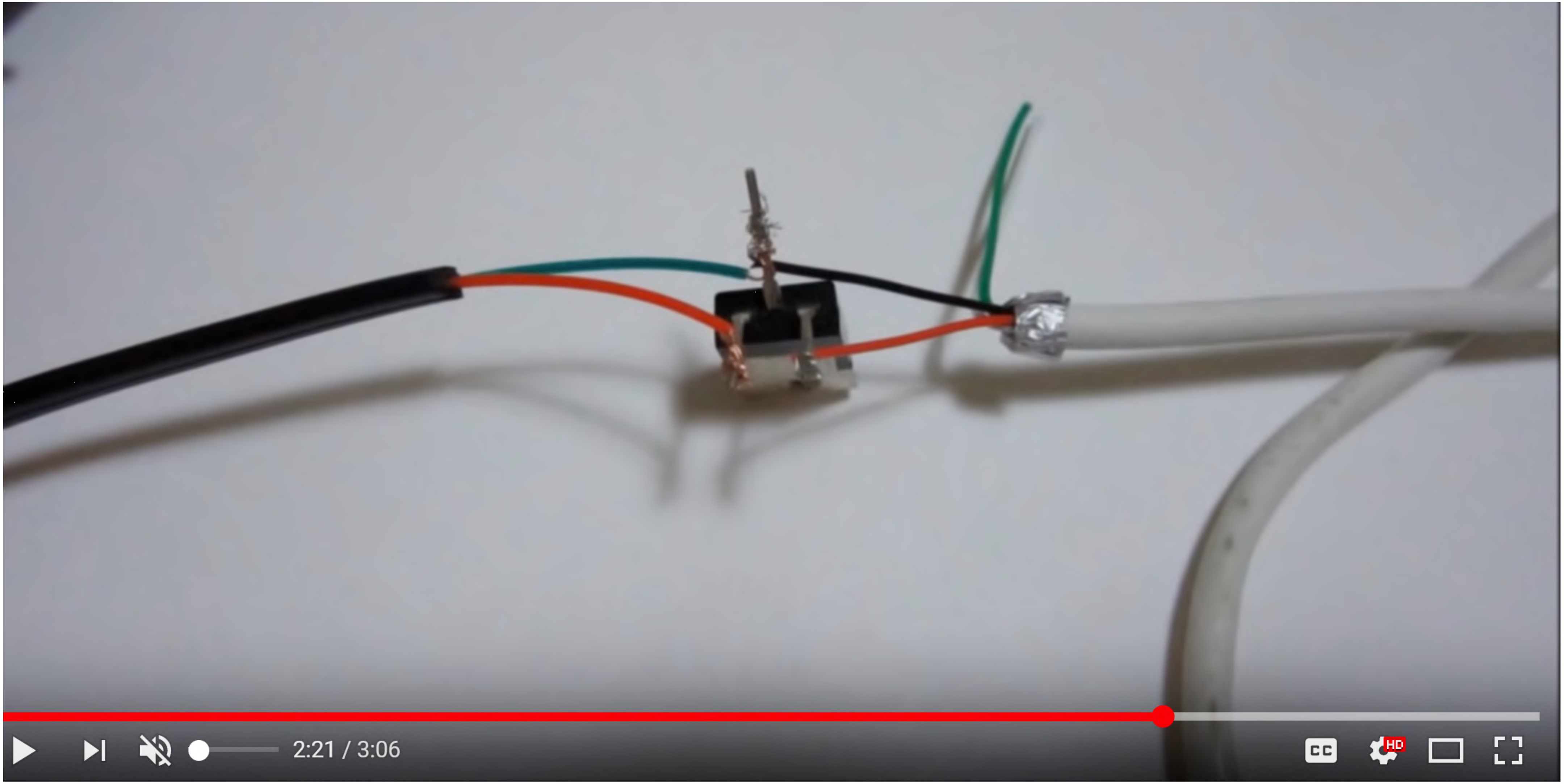
2:21 / 3:06
HD

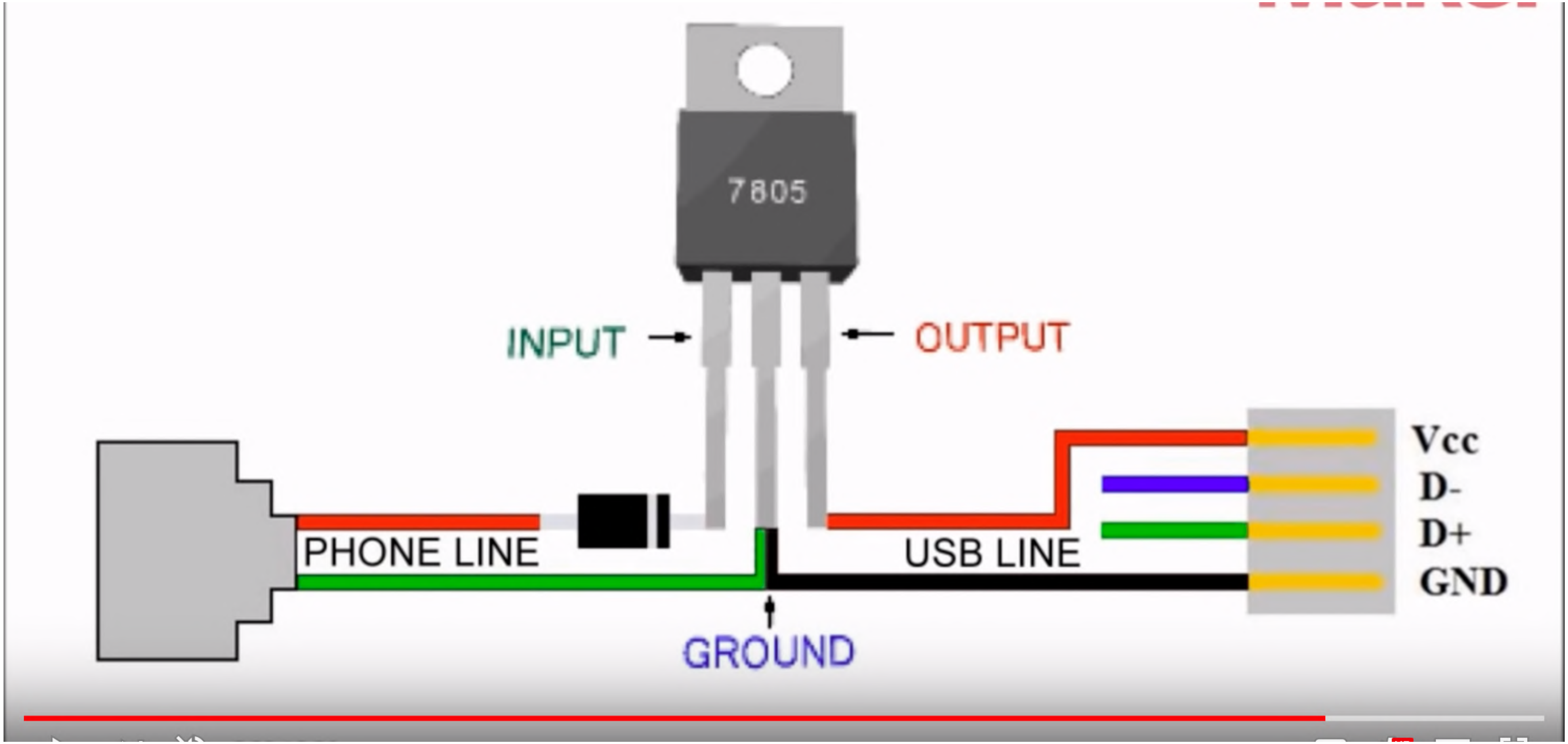
7805
INPUT
OUTPUT
PHONE LINE
USB LINE
GROUND
Vcc
D-
D+
GND

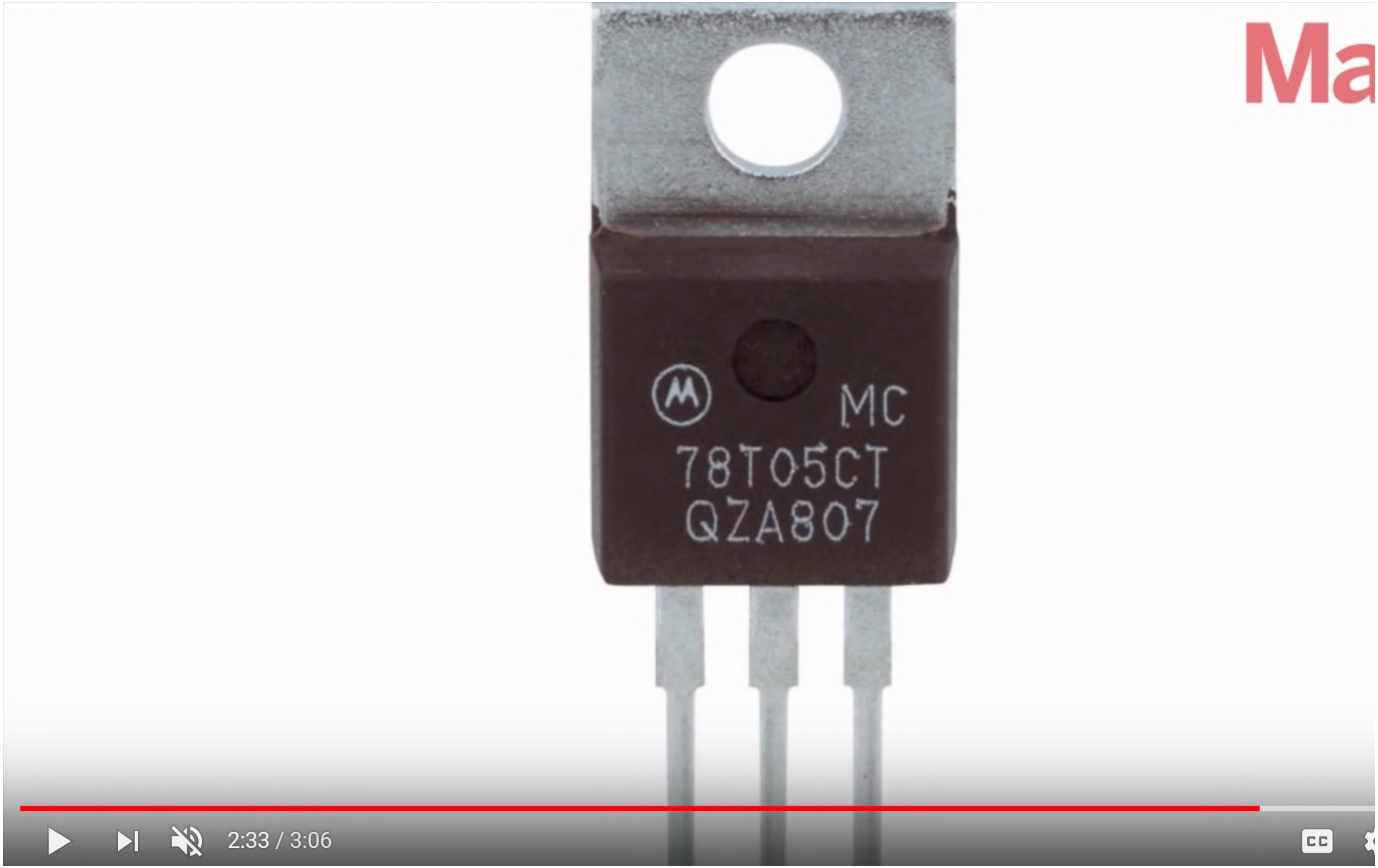
Ma
MC
78T05CT
QZA807
2:33 / 3:06
CC

BUILDING A DIY VIDEO TRANSMITTER

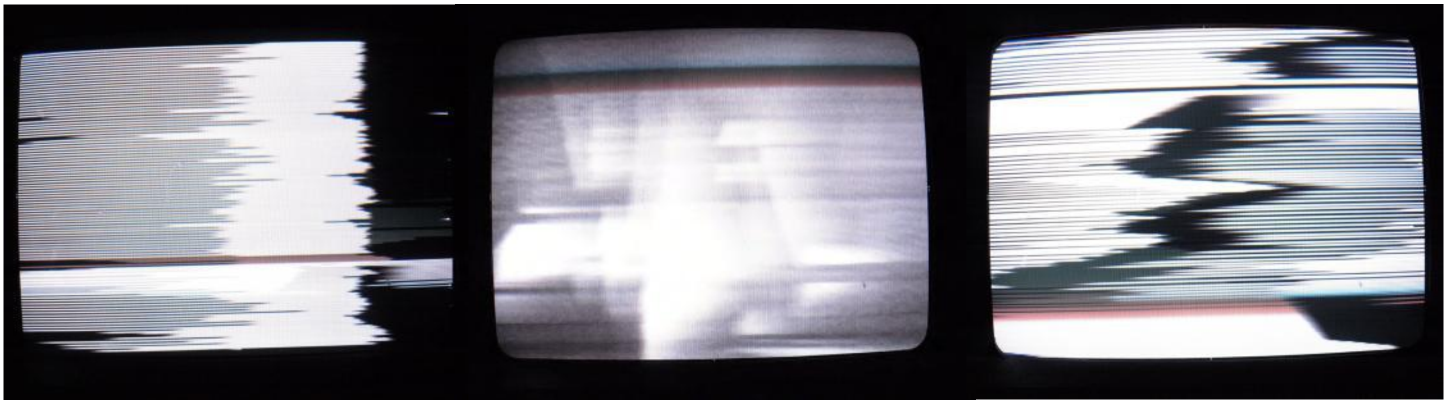

WHAT YOU NEED

- Blank PCB Copper Plate (jameco.com part #169279)
- Capacitors: 2x 0.1uf
  - 1x 100u
  - 2x 6-70pf Variable Caps (jameco.com part #32855)
- Resistors: 27kOhm
  - 10kOh
- Transistor-MPSA18 (jameco.com part #210681)
- 1k potentiometer
- Small amount (~4”) of magnet wire 24 AWG or close (jameco.com part #2098419)
- Sand paper
- Stranded hookup wire
- Hacksaw or dremel tool
- Headphones with 1/8” plug (from iPod or MP3 player)
- Hot glue or super glue
- Soldering Iron
- CRT “tube” TV with an antenna (the transmitter won’t transmit to newer flat screen TV’s)

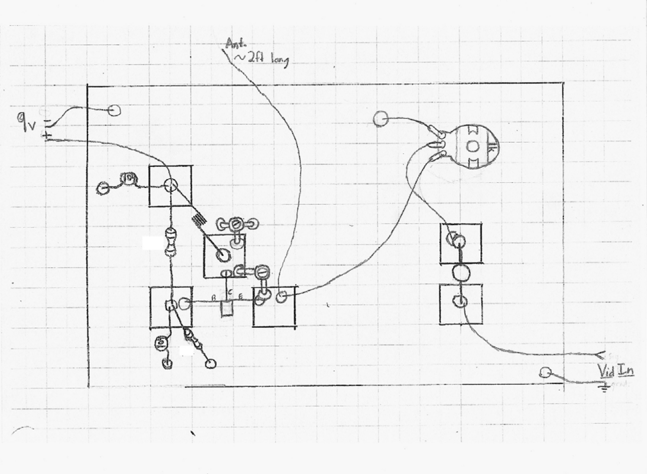

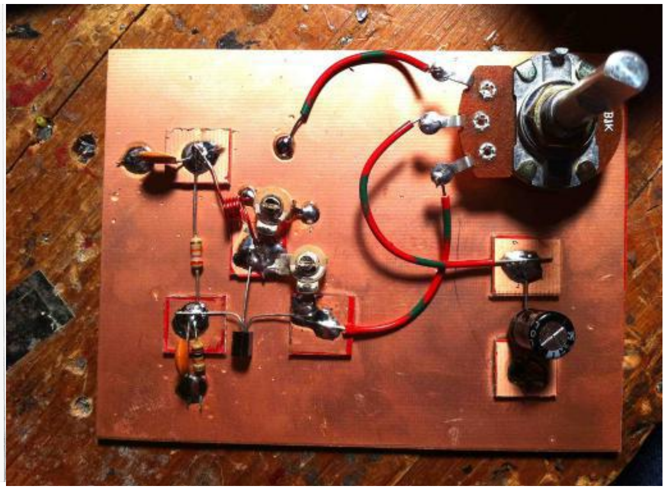

INTRODUCTION

This micro video transmitter was designed by the DIY radio and micro TV artist Tetsuo Kogawa. His website (http://anarchy.translocal.jp/) is full of great info on DIY radio, and all credit for this circuit goes to him.The transmitter can be used to transmit a relatively clear video signal or abstract visualizations, depending on the desired output. It will usually function at a distance of about 20 feet. If you have any previous experience with electronics or circuit building, you will notice that this method of building is different. The circuit is constructed on top of a copper “ground plate” which is common in radio applications. This tutorial uses step-by-step sketches and photos created based on Tetsuo Kogawa’s design.

BUILDING THE TRANSMITTER

1. PREPARING THE COPPER PLATE

- Take the bare copper plate and cut off 1inch of material using a dremel tool or hacksaw.
- Use this material to cut out 6 square ½ in. x ½ in. pieces.
- Arrange the 6 pieces on the larger copper plate as shown in the image below…

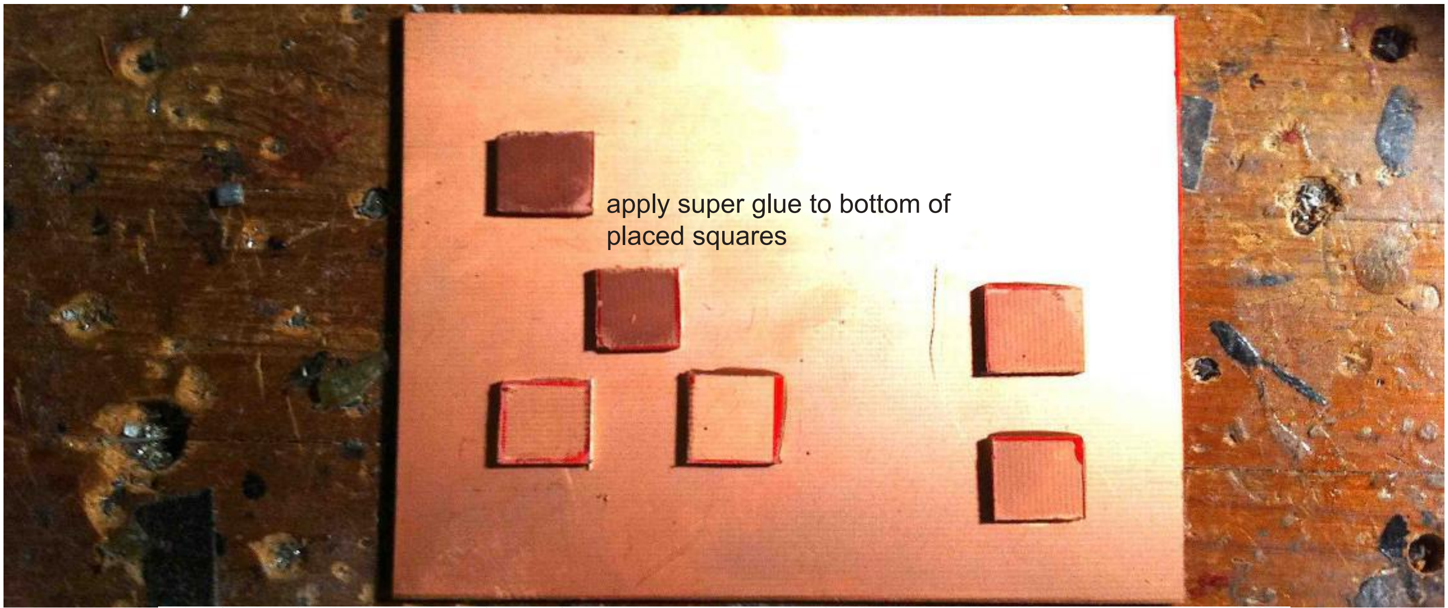

apply super glue to bottom of placed squares

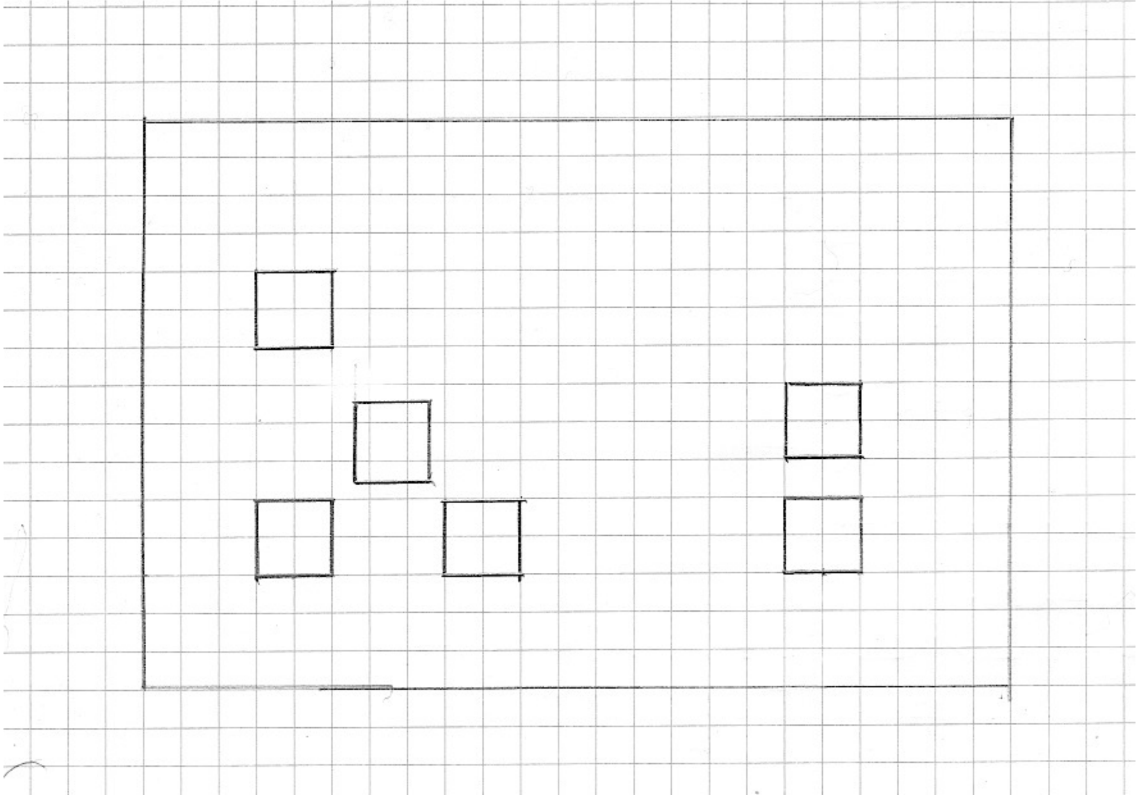

NOTE Be sure to arrange them so that all the component leads will be able to reach the squares. Look ahead to see what components will go where so you know exactly how to space them.

- Once you have them arranged, put a dab of hot or super glue on the bottom of each square and stick them in place on the copper plate.

2. ADDING THE 100uF CAPACITOR (the cylinder with “100uF” written on the side)

- Melt a blob of solder onto the two squares furthest to the right. To do this, apply your iron to each square to transfer heat for an extended period of time until the solder flows onto the surface.

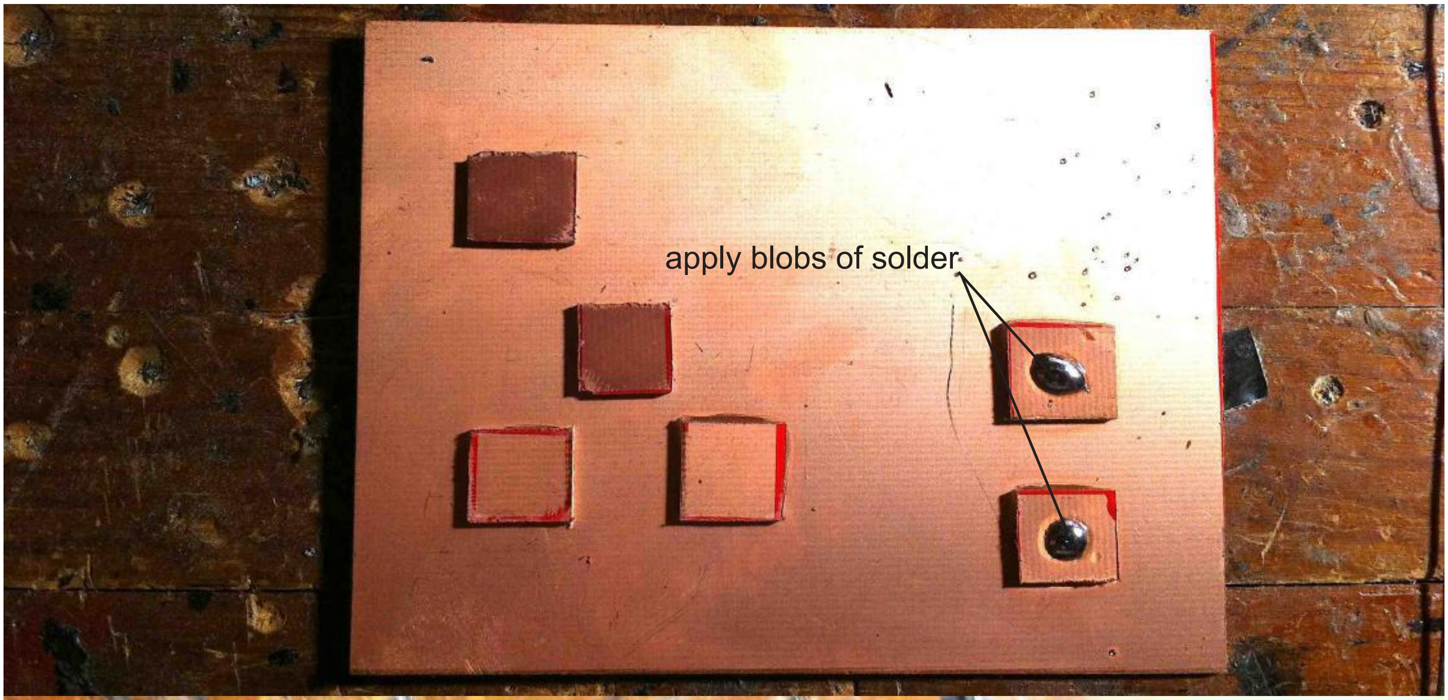
apply blobs of solder

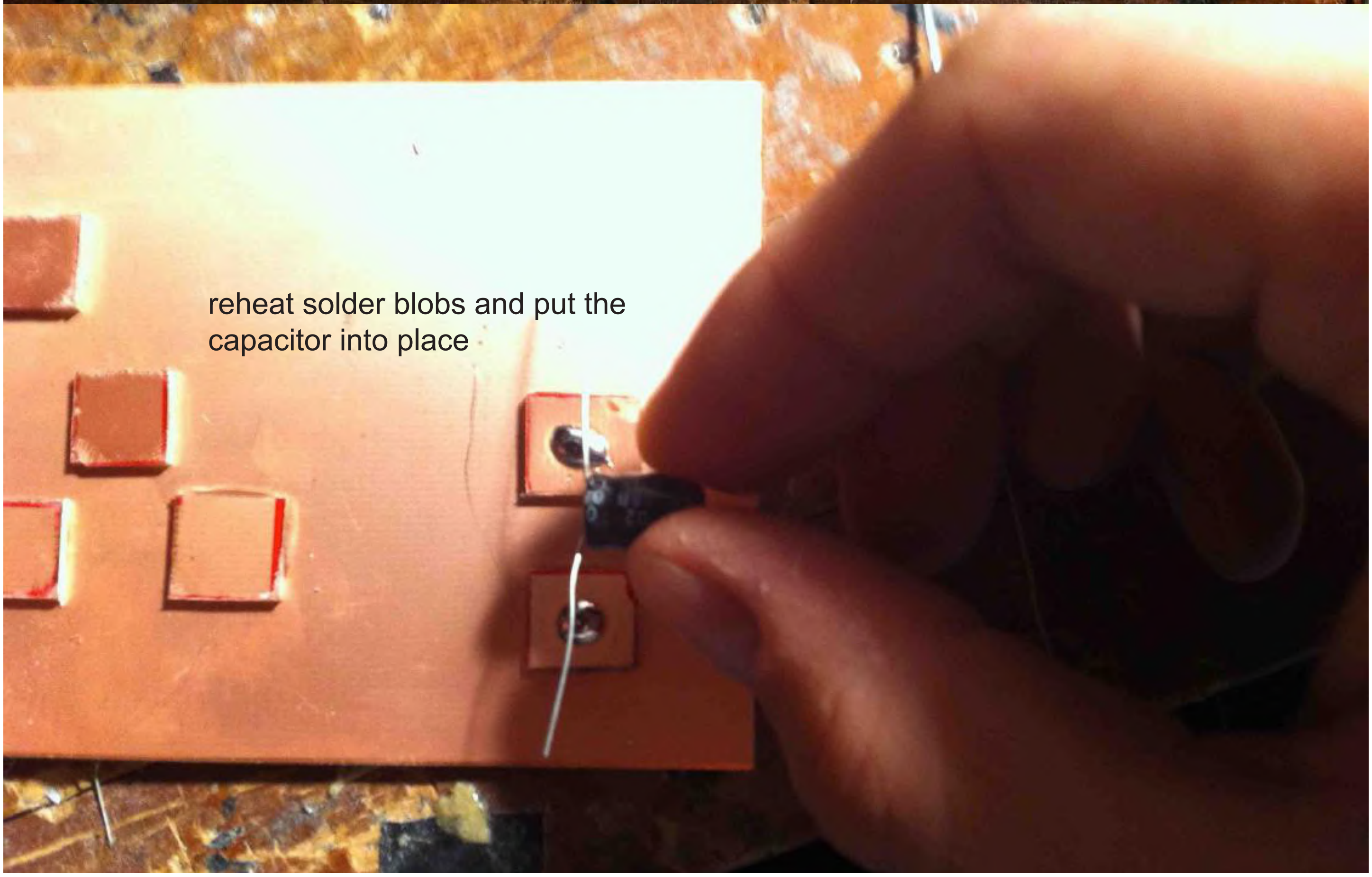
reheat solder blobs and put the capacitor into place

- Bend the legs out of the 100uf capacitor and straddle them between the two squares making sure that the leads are only touching the two small squares (not touching the ground plate).
- Cut the excess off of the leads.
- One at a time, re-heat the solder blobs and place the capacitor leads into them, allowing them to cool down until they stay in place.

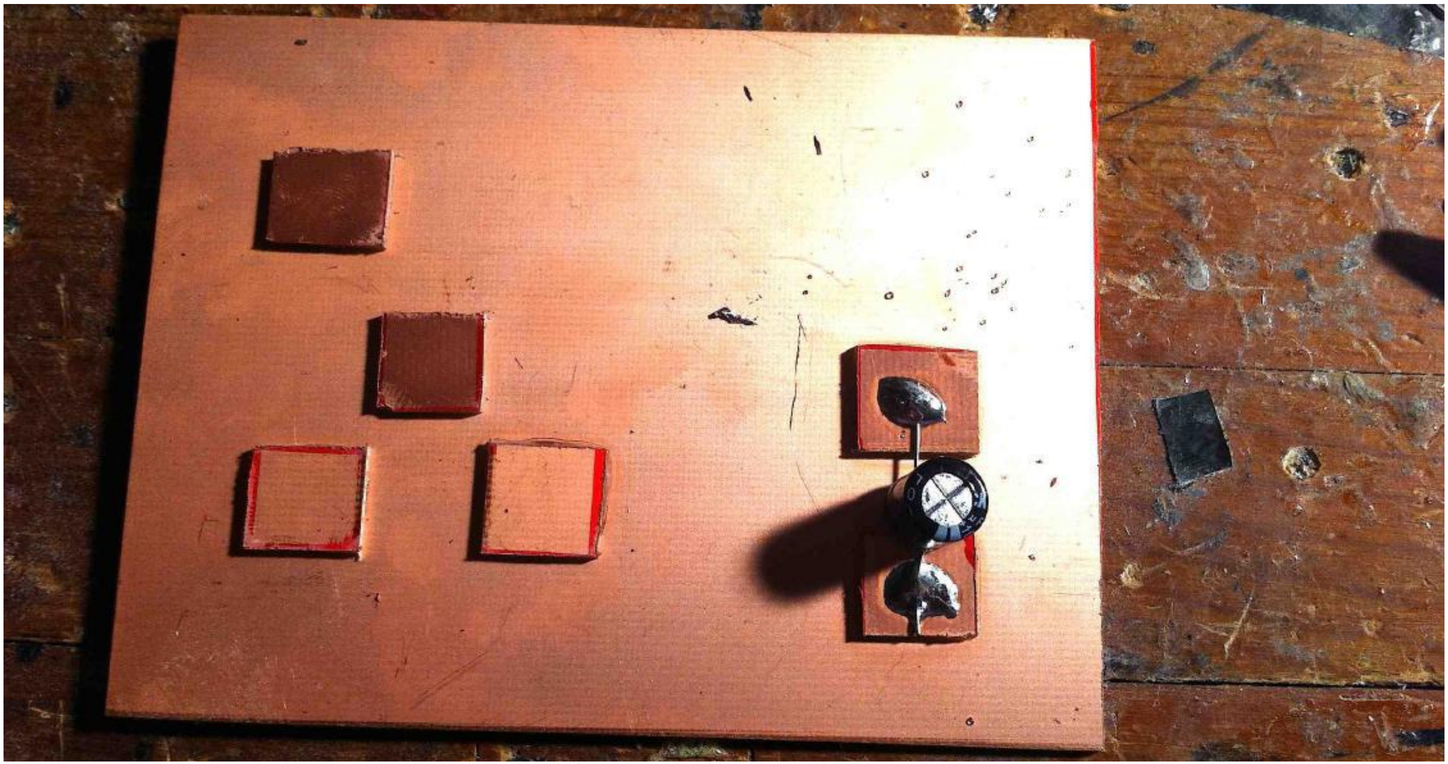

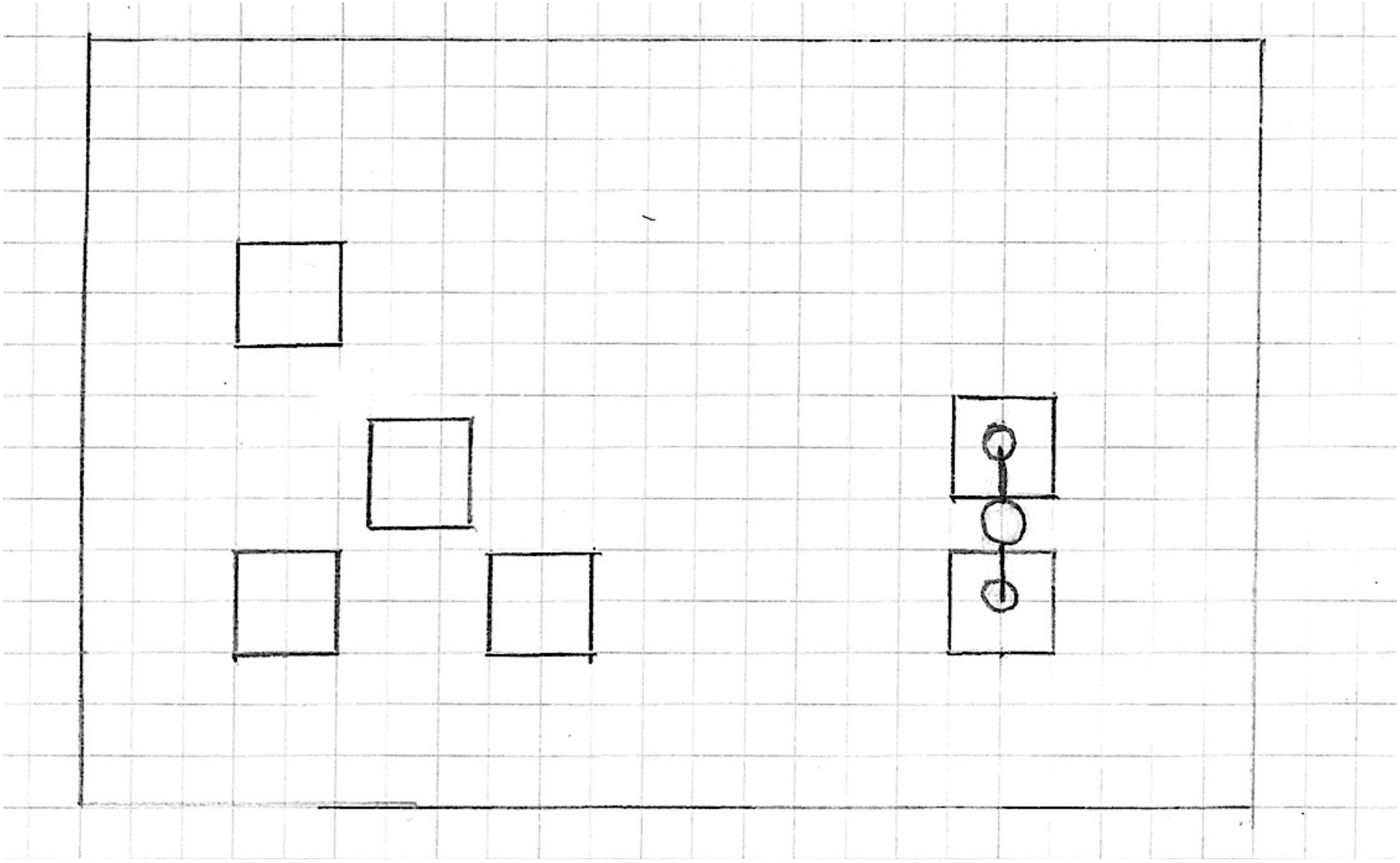

*****NOTE***** - This capacitor is polarized meaning that it has to be facing in the appropriate direction. Make sure you solder the NEGATIVE lead to the LOWER of the two squares. The negative side has the shorter lead and a vertical stripe along the side of the capacitor.

3. ADDING THE .1uF CAPACITORS (the two discs with “104” written on them)

- These capacitors will be positioned between the upper and lower left most squares and the ground plate. Melt a blob of solder on these two squares and two blobs on the ground plate beside each. (Reference the pictures for exact positioning).
- Re-heat the blobs and position the capacitors accordingly as in step 2.

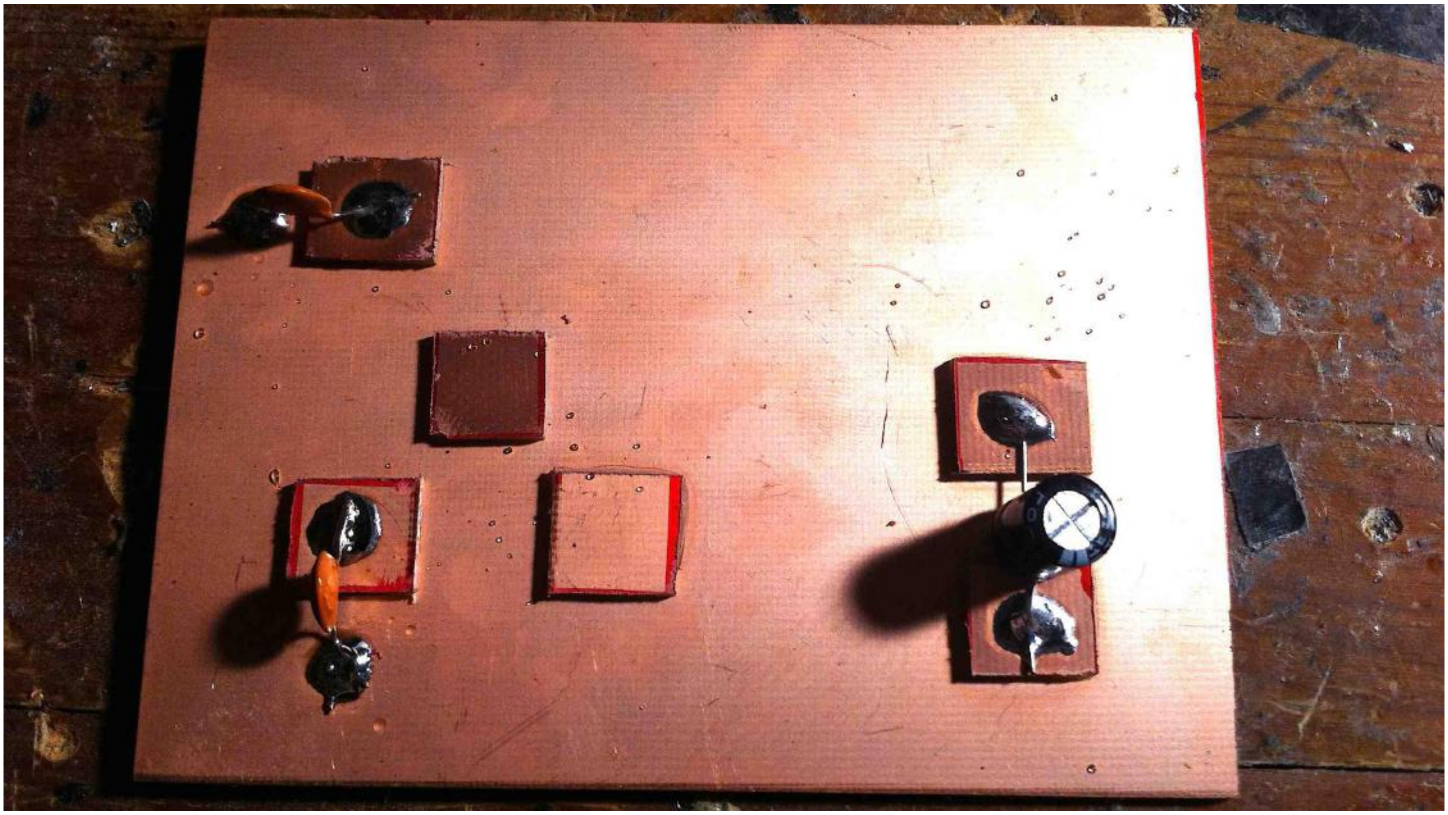

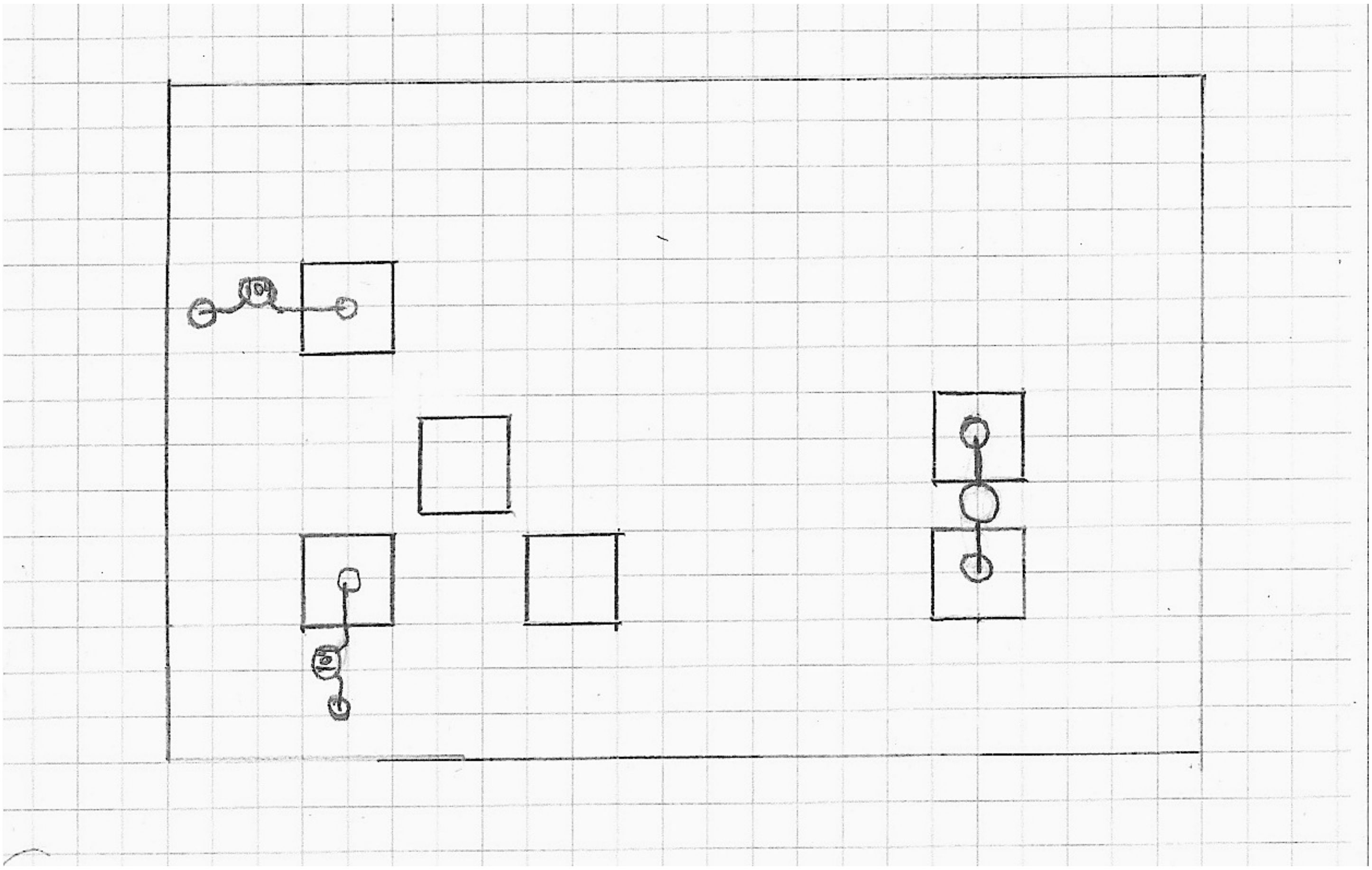

4. ADDING THE RESISTORS (10k and 27k)

- Solder the 10k resistor (brown, black and orange stripes) between the bottom left square and ground.
- Solder the 27k resistor (red, violate and orange stripes) between the two left squares.

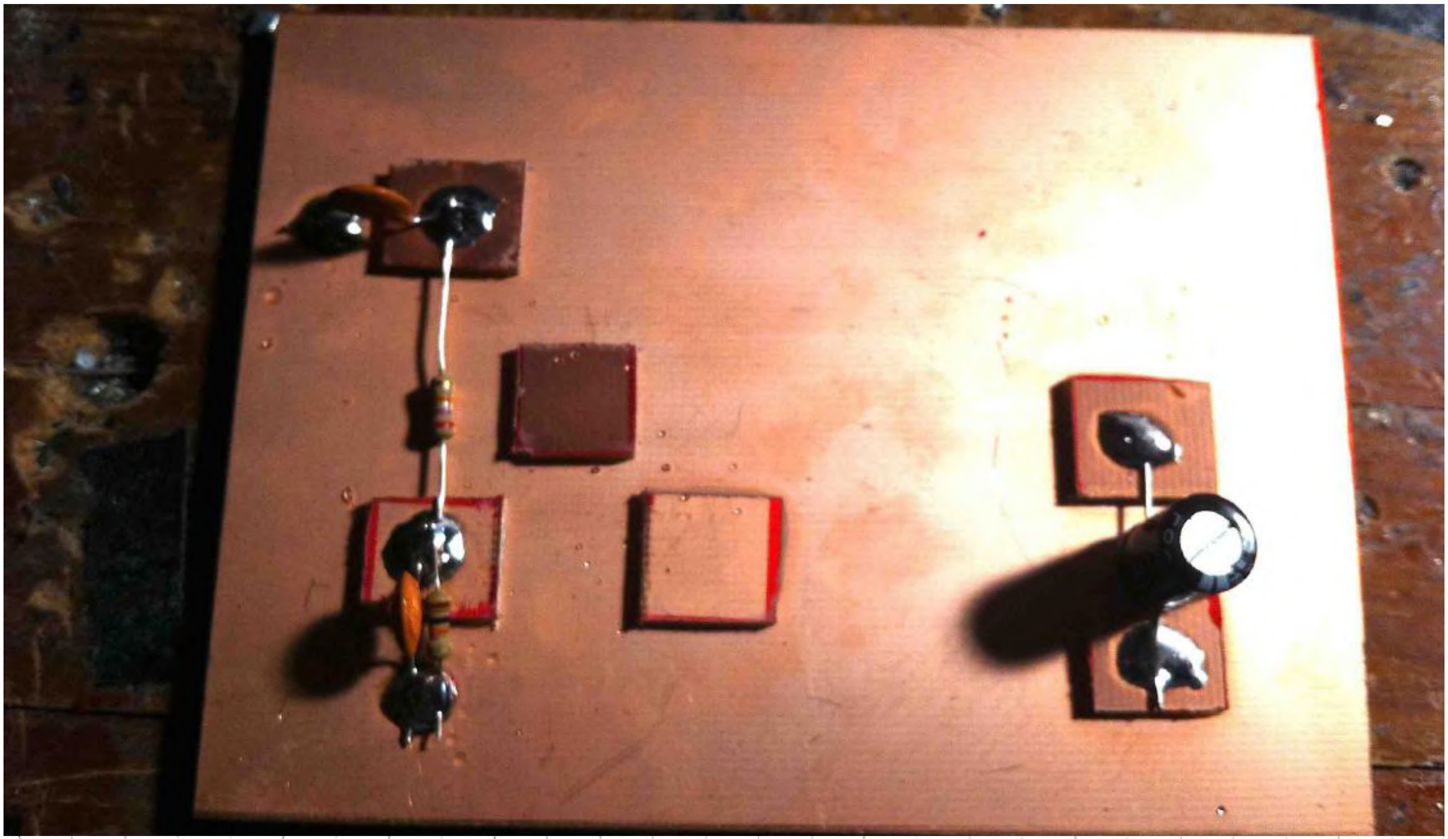

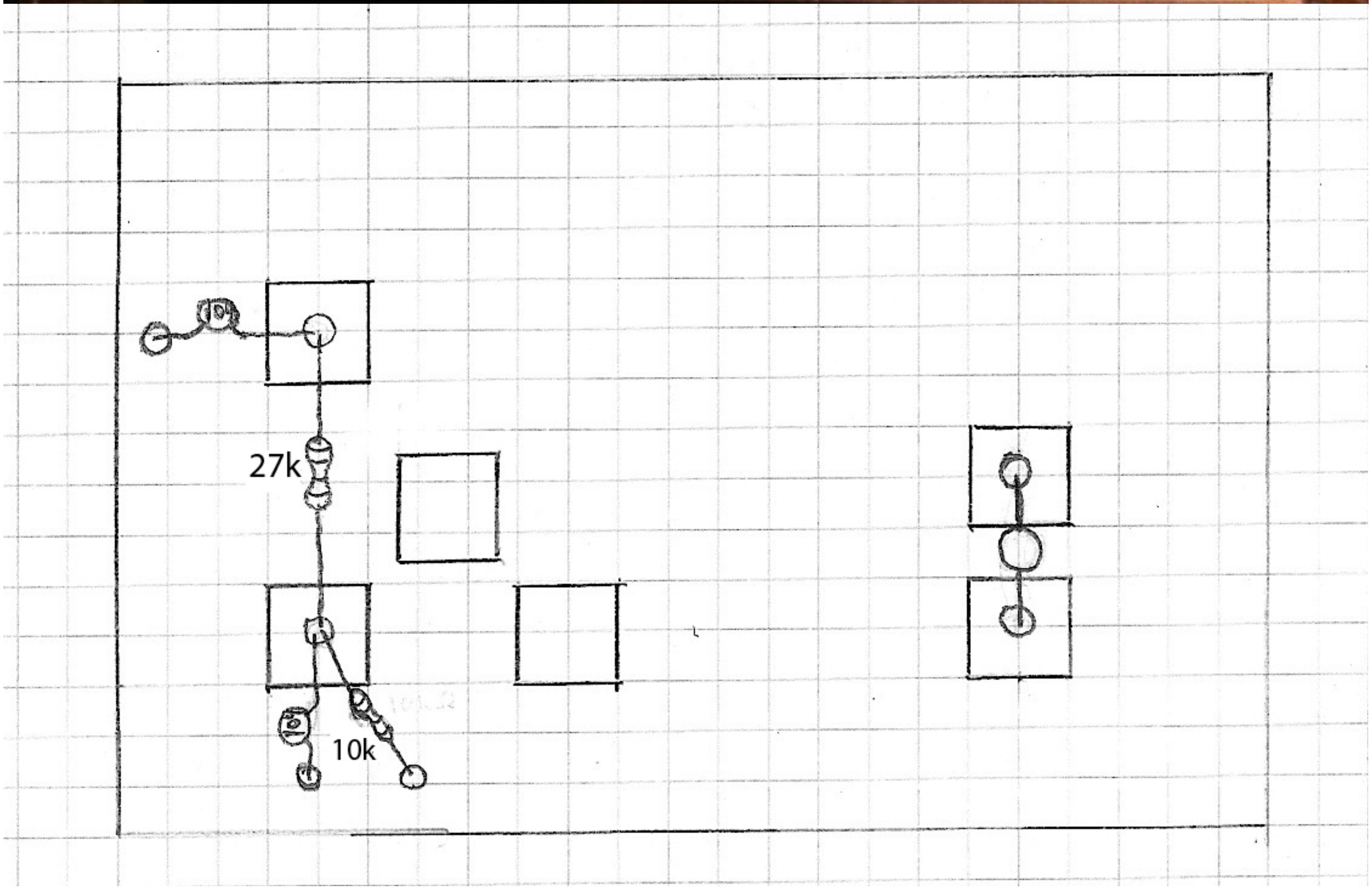

5. ADDING THE COIL

- Cut about 4” of the magnet wire (24 AWG or close)
- Use the plug from your headphones to wrap the wire around. Make 5 turns.

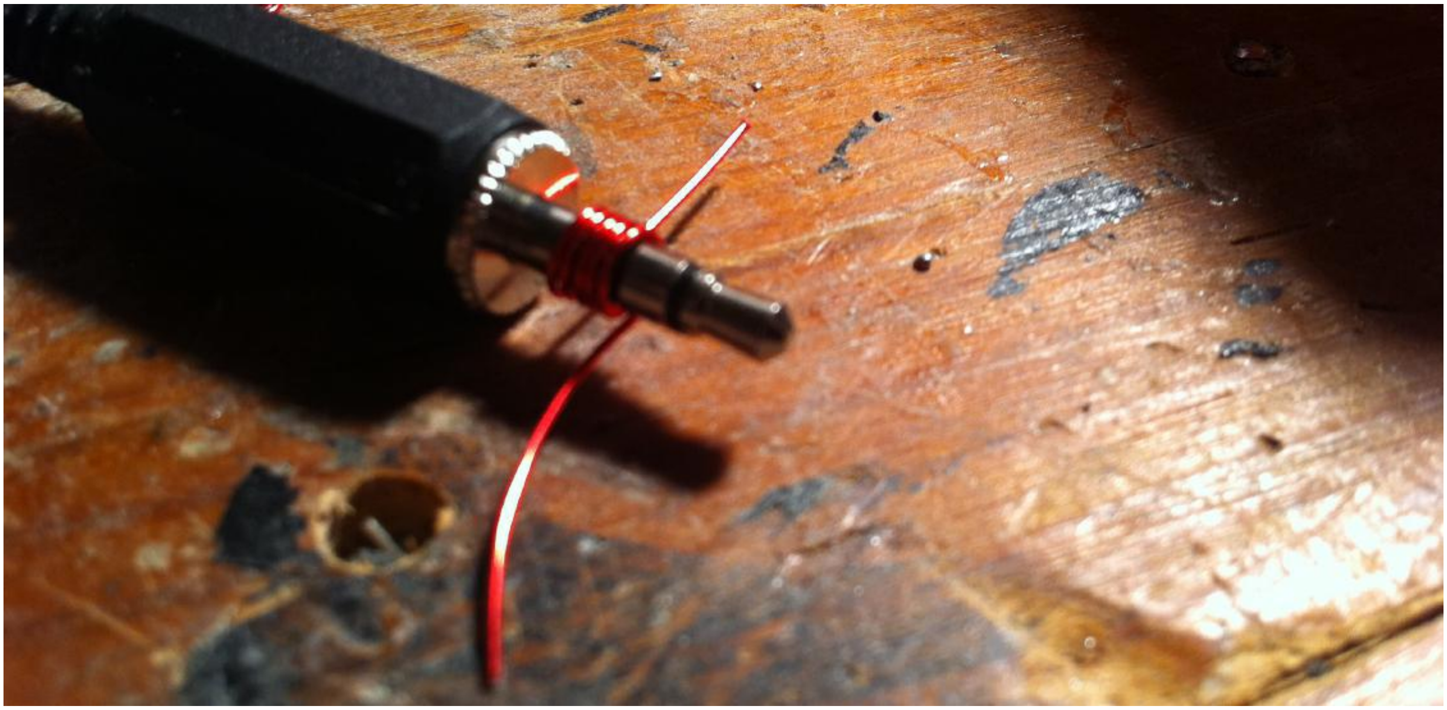

- With the wire still wrapped around the headphone plug, take a small piece of sand paper and rub off the colored insulation of each end until the bare copper is exposed (If the copper is not exposed, you won’t be able to make a connection).

- Solder the coil into place bridging the two squares as seen below

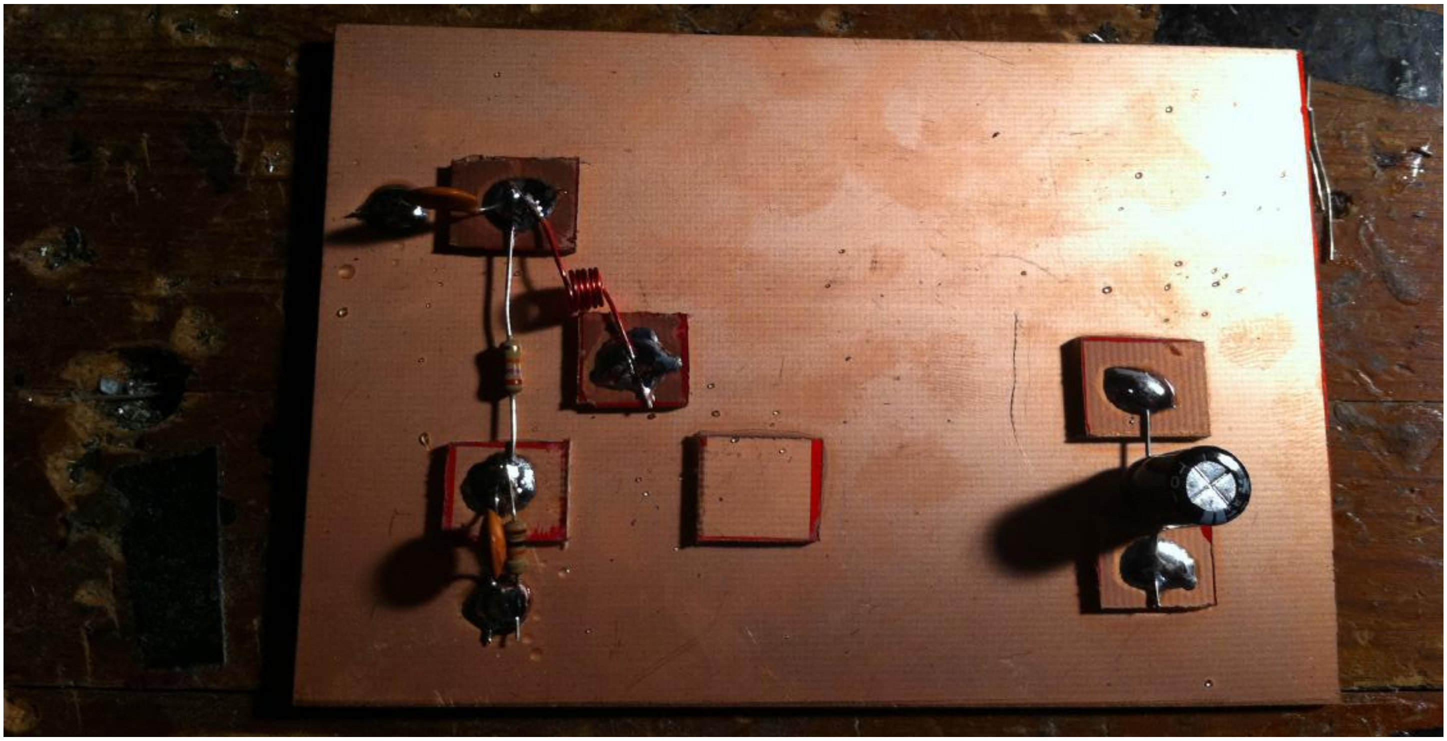

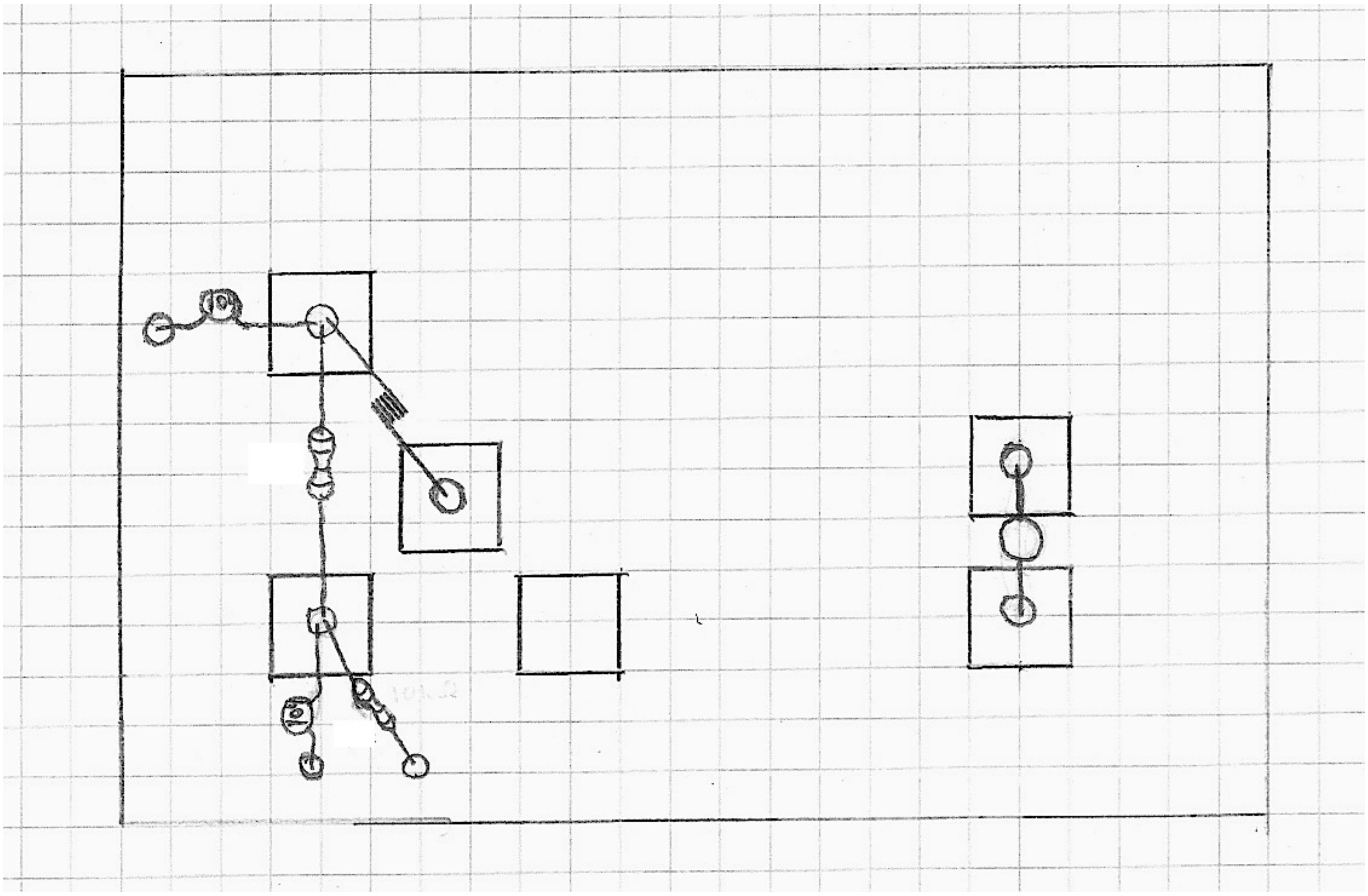

6. ADDING THE VARIABLE CAPACITORS

NOTE Once the component leads start to pile up on each other on each square, it helps to add additional solder.

- The first variable capacitor will be soldered between the top square of the three that make up the bottom triangle and the ground plate.

- Solder the “nose” of the capacitor (the lead that is different from the other two) to the square, and the other two side leads to the ground plate.

NOTE You will have to bend the leads in a certain way in order to get everything positioned properly. Just play around with it till you get it.

- The second variable capacitor will bridge between top and right squares of the triangle. Only two leads are needed here so use your clippers to remove one of the side leads, leaving the nose and only one side. (See the pic below)

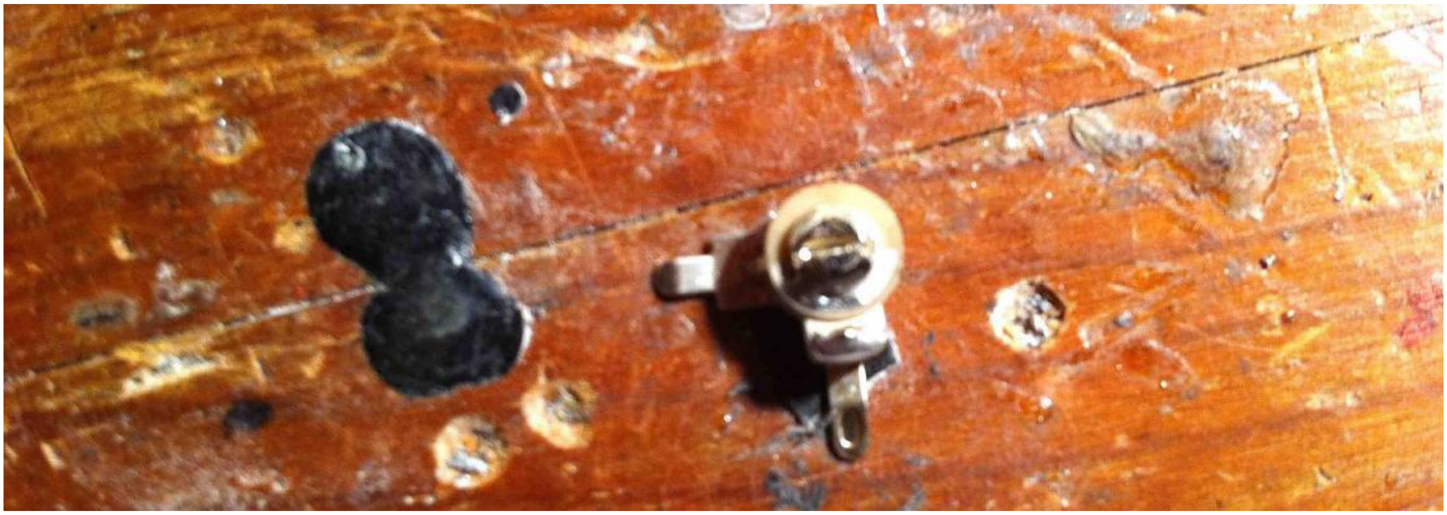

The variable capacitors soldered into place…

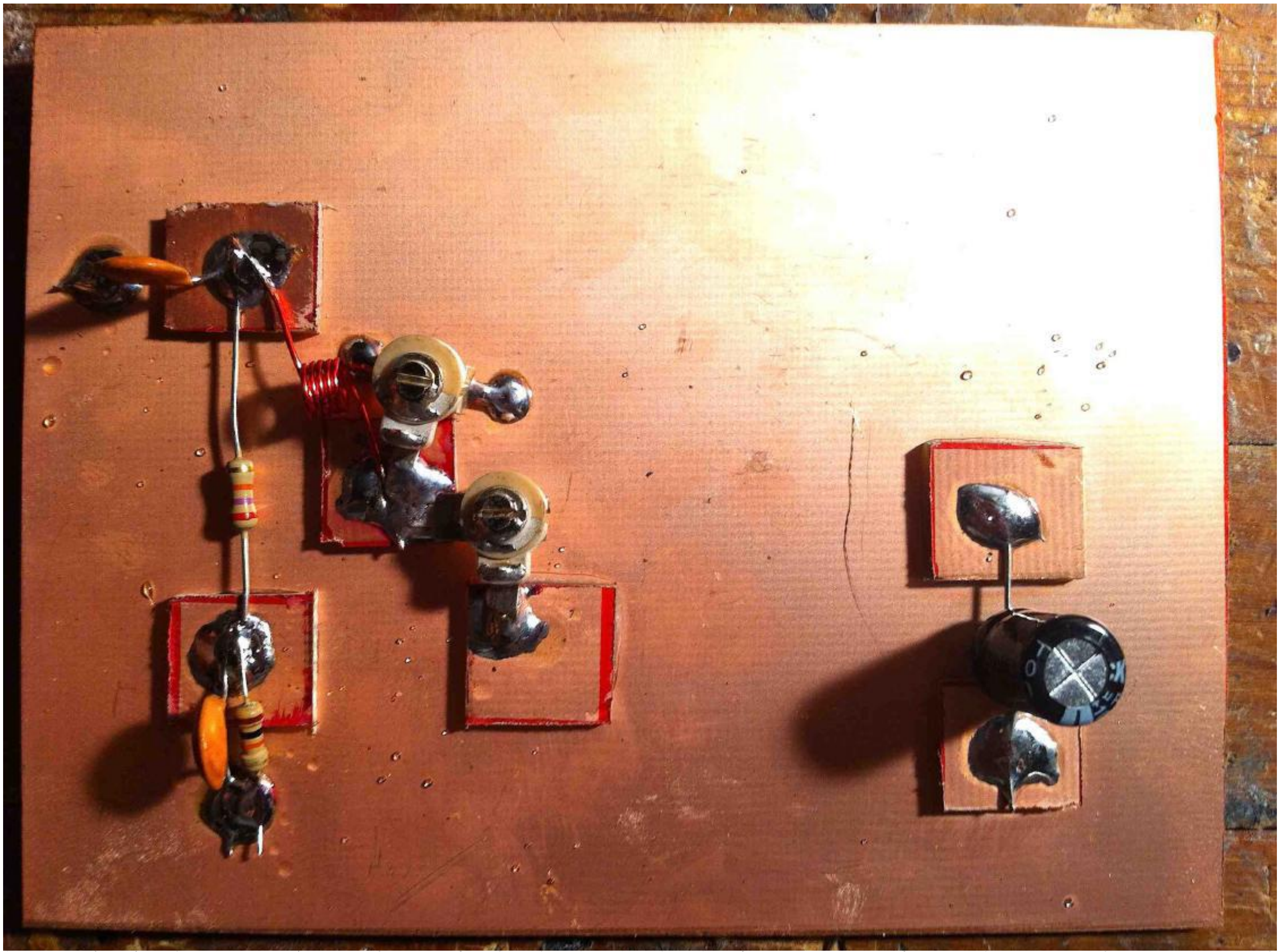

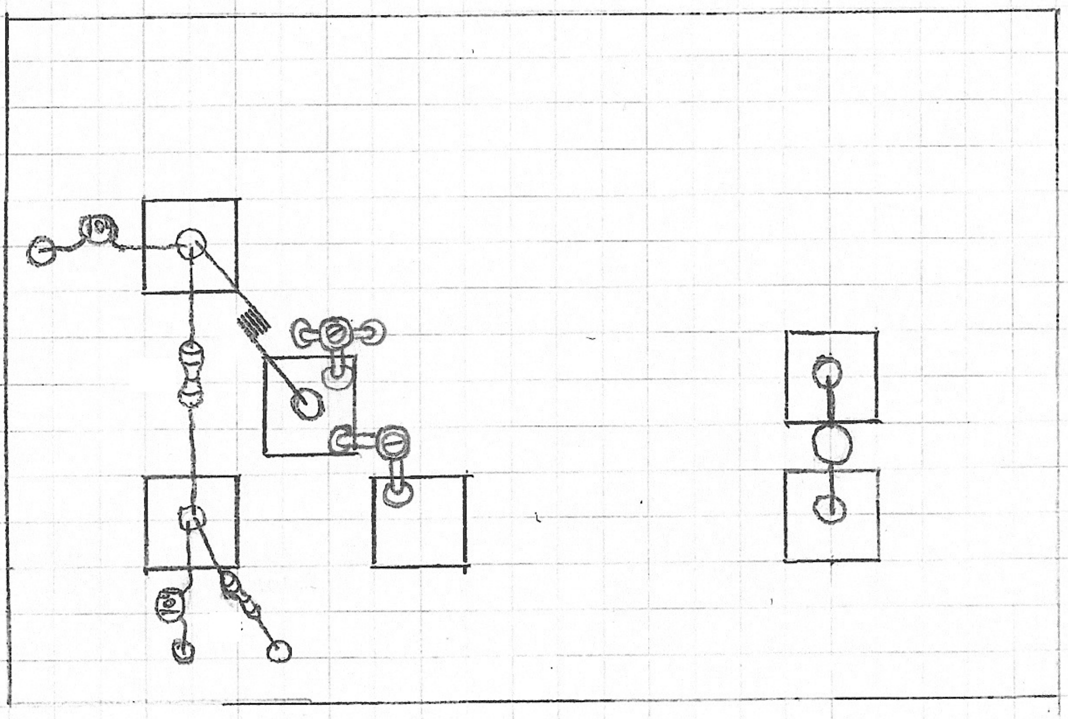

7. ADDING THE TRANSISTOR (MPSA18)

NOTE This transistor is sensitive to heat and can be destroyed by the heat from your iron. Be sure to follow the necessary precautions below. But don't worry if you do kill one, they are only ~$.10 a piece.

- The three leads of the transistor all do specific things and must go to the appropriate places. Here is a pic of the orientation of this transistor…

C=Collector B=Base E=Emitter

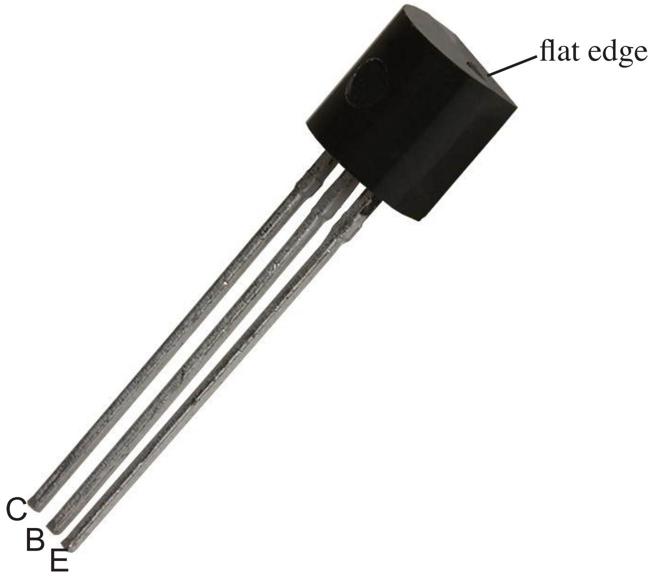
flat edge
C
B
E

- Each lead from the transistor will be soldered to the three different squares that make up the triangle, so you will have to bend the leads accordingly.
- The B lead (base) goes to the left square of the triangle, the C lead (collector) goes to the top square, and the E lead (emitter) goes to the right square, as in the image below.

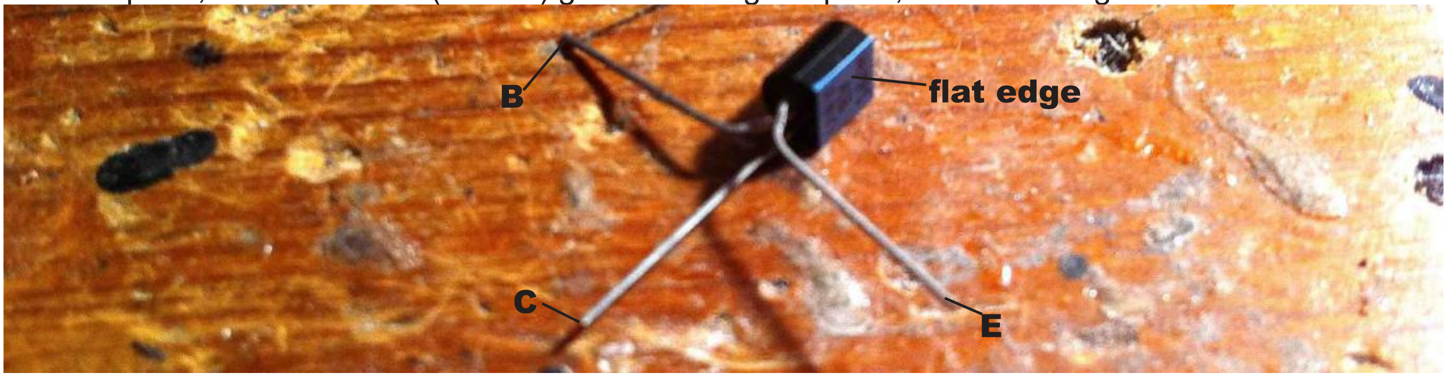
B
flat edge
C
E

- Before you solder the transistor in place, you need to have a "heat sink" attached to protect it. This will draw some of the heat from the transistor to something else. To do this, simply take an alligator clip and attach it to the lead you are soldering…

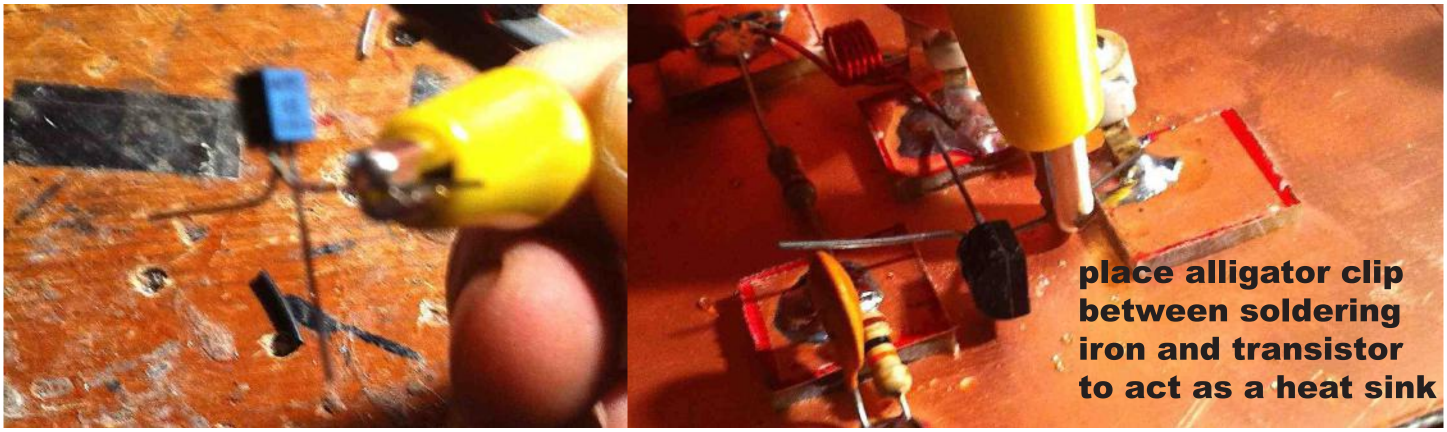
place alligator clip
between soldering
iron and transistor
to act as a heat sink

- Solder each lead in place attaching the clip lead to each one before doing so.

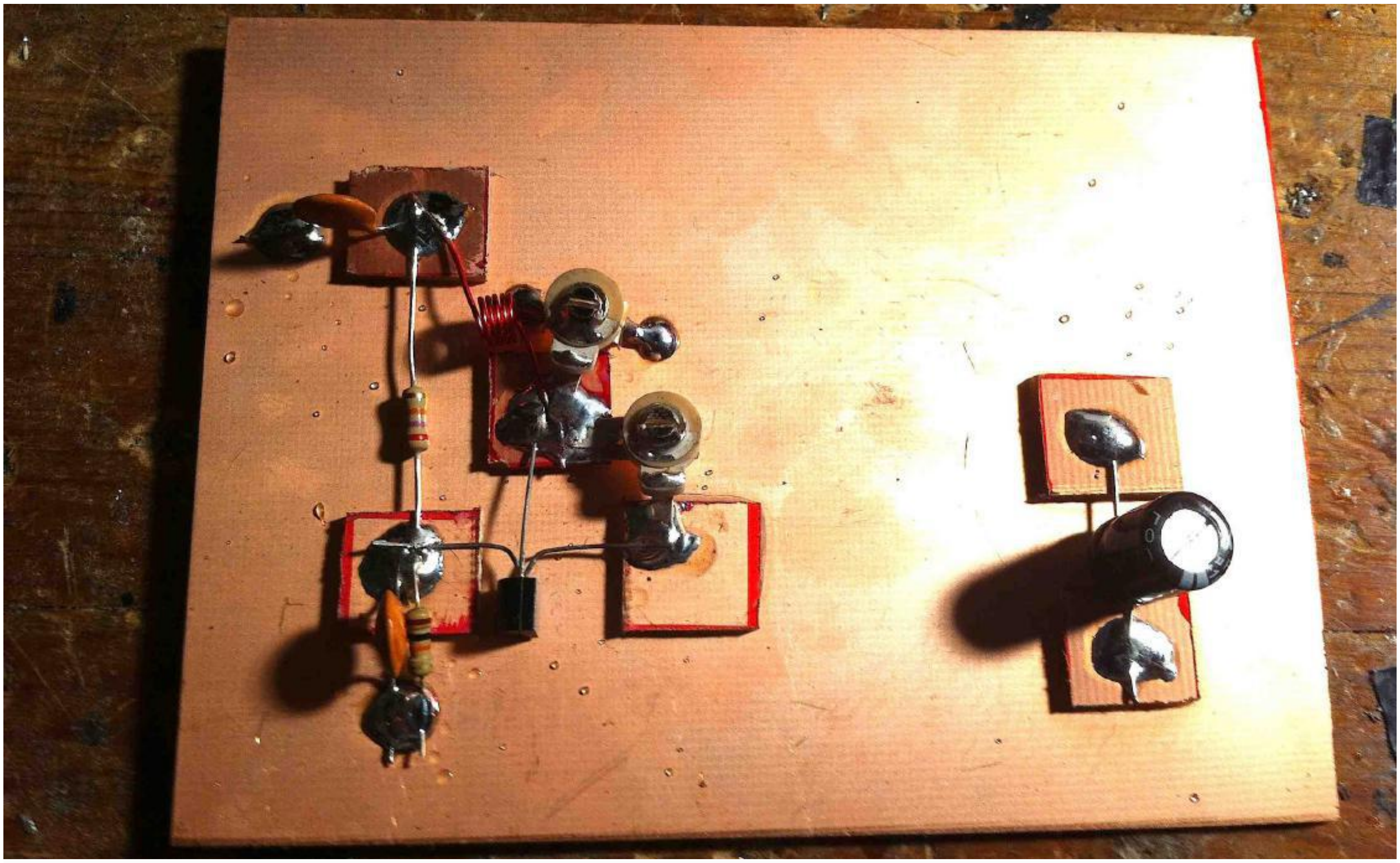

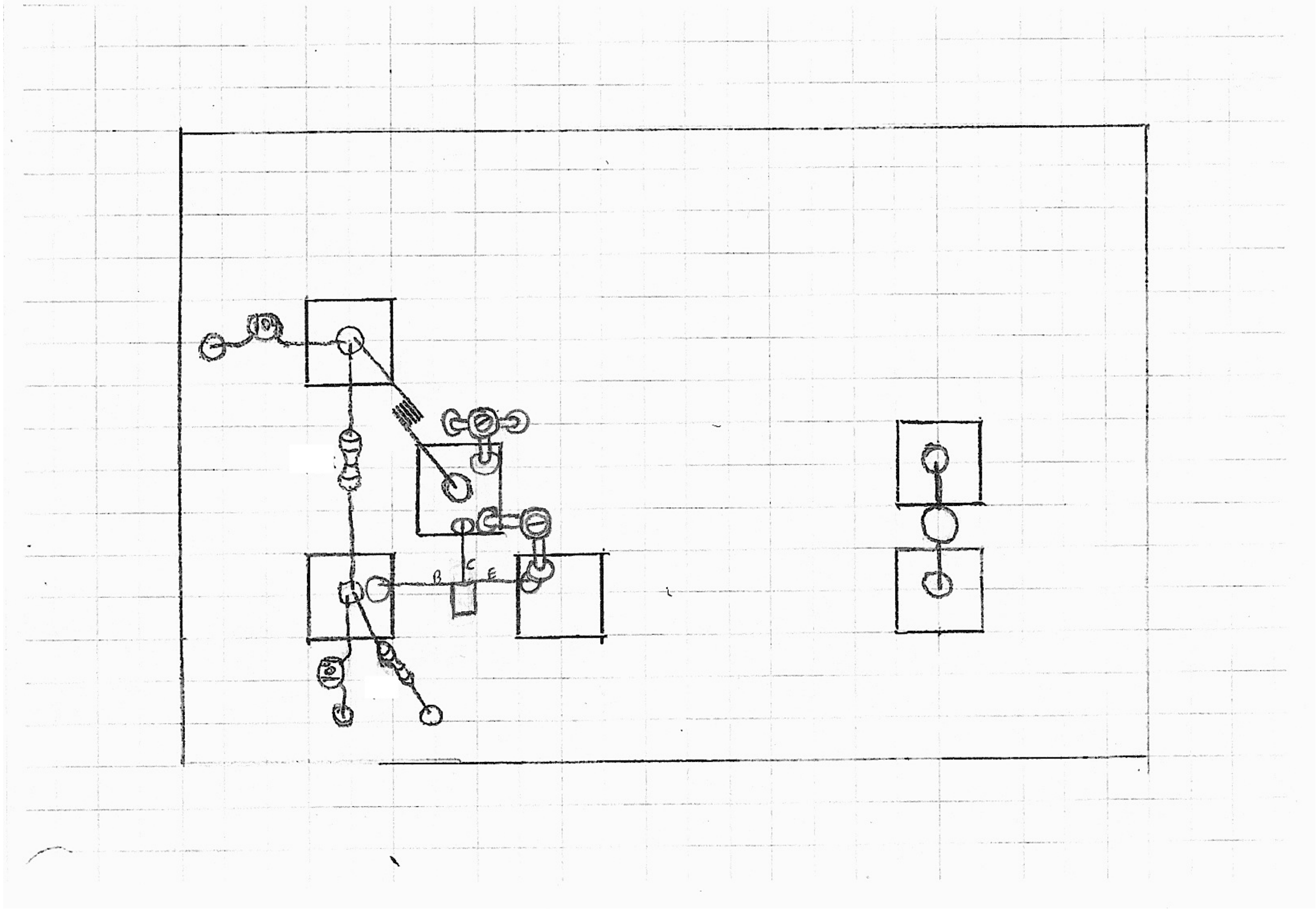

8. ADDING THE 1k POTENTIOMETER

- Put a bit of super glue or hot glue on the bottom of the potentiometer and stick it to the upper right corner with the leads facing towards the left of the plate.
- Take a small amount of stranded hookup wire and strip both ends. Solder one side to the upper most lead and the other side to the ground plate.
- Repeat the process by connecting the middle lead of the potentiometer to the upper most of the two right squares. Then connect the lower lead of the potentiometer to the right most square of the triangle.

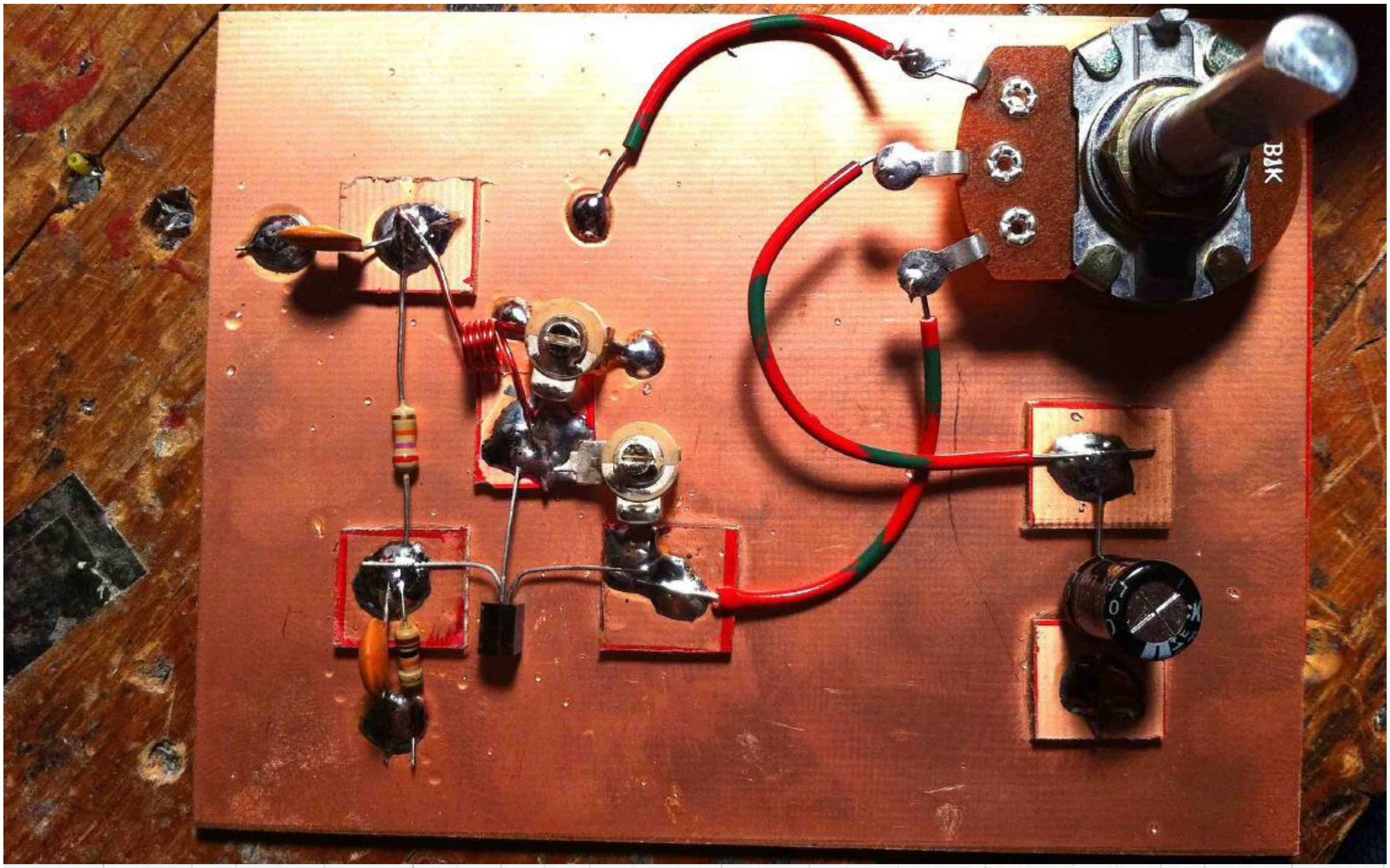

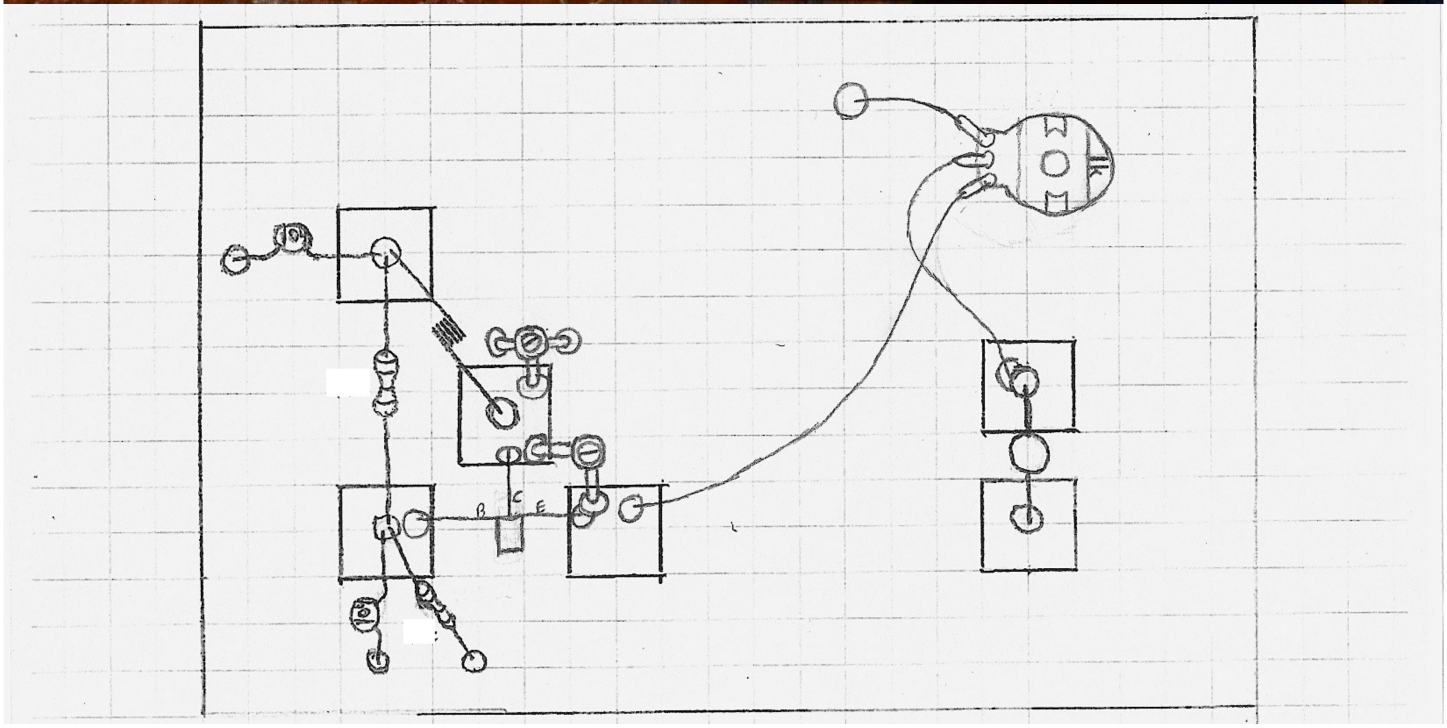

9. ADDING THE INPUT WIRES

- Cut two small lengths (~4-5”) of your stranded wire and strip both sides
- Solder one of the wires to the ground plate (this will be the ground input from your video source)
- Solder the other wire to the bottom square of the two right squares. (This will be the signal input from your video source)

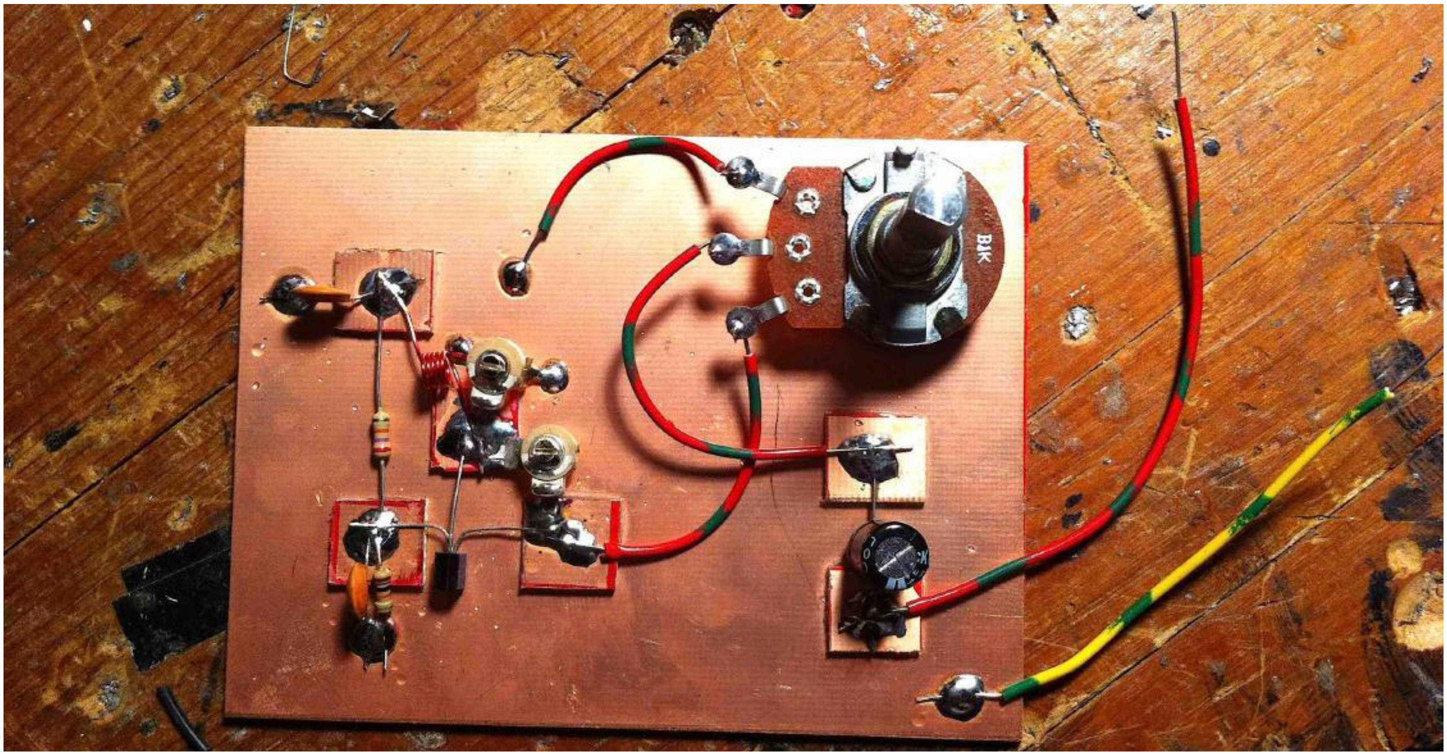

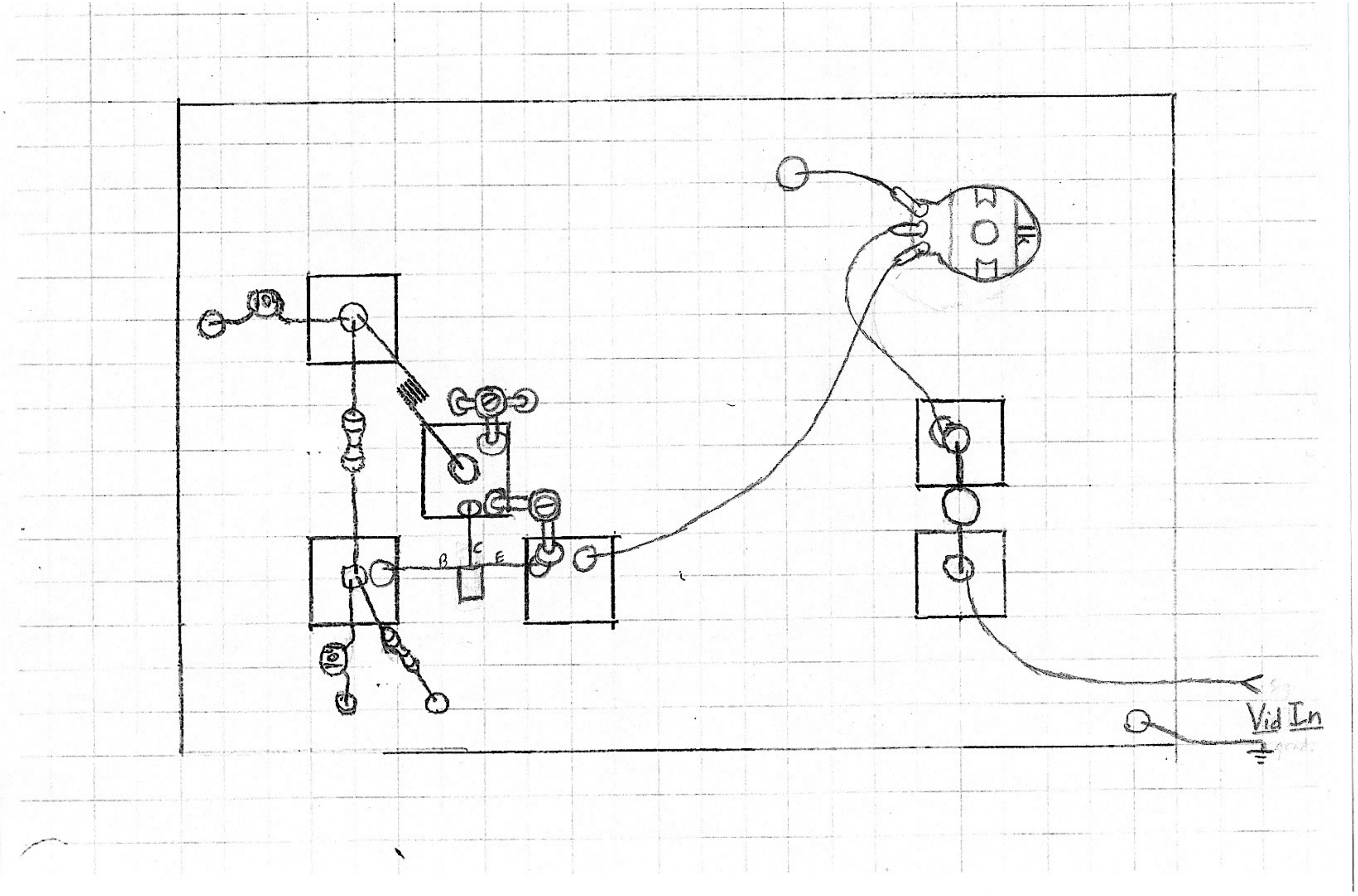

10. ADDING THE BATTERY INPUT WIRES

- Just like the last step, cut two shot pieces of wire and strip both ends.
- Solder one of the wires to the ground plate (this will be the negative input from your 9volt battery).
- Solder the other wire to the upper left square (this will be the positive input from your 9volt battery).

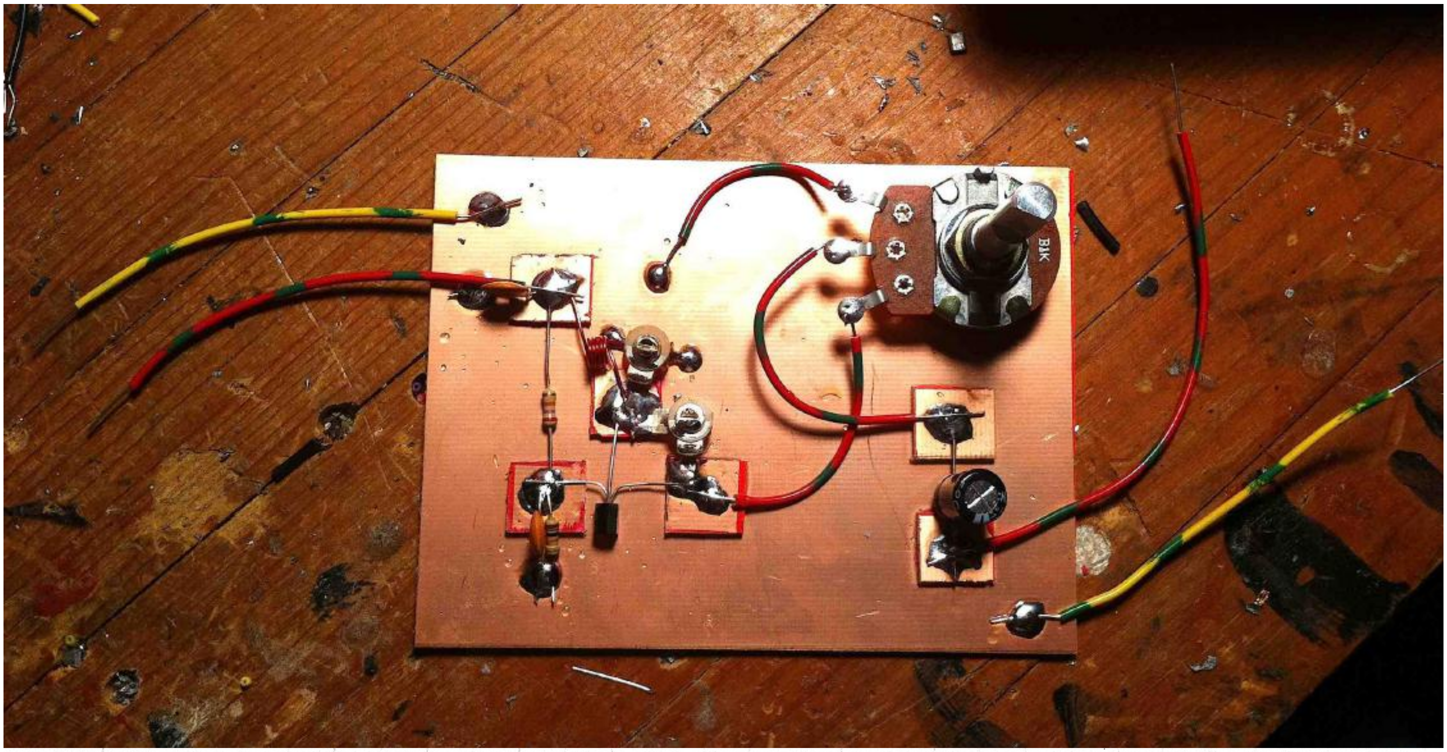

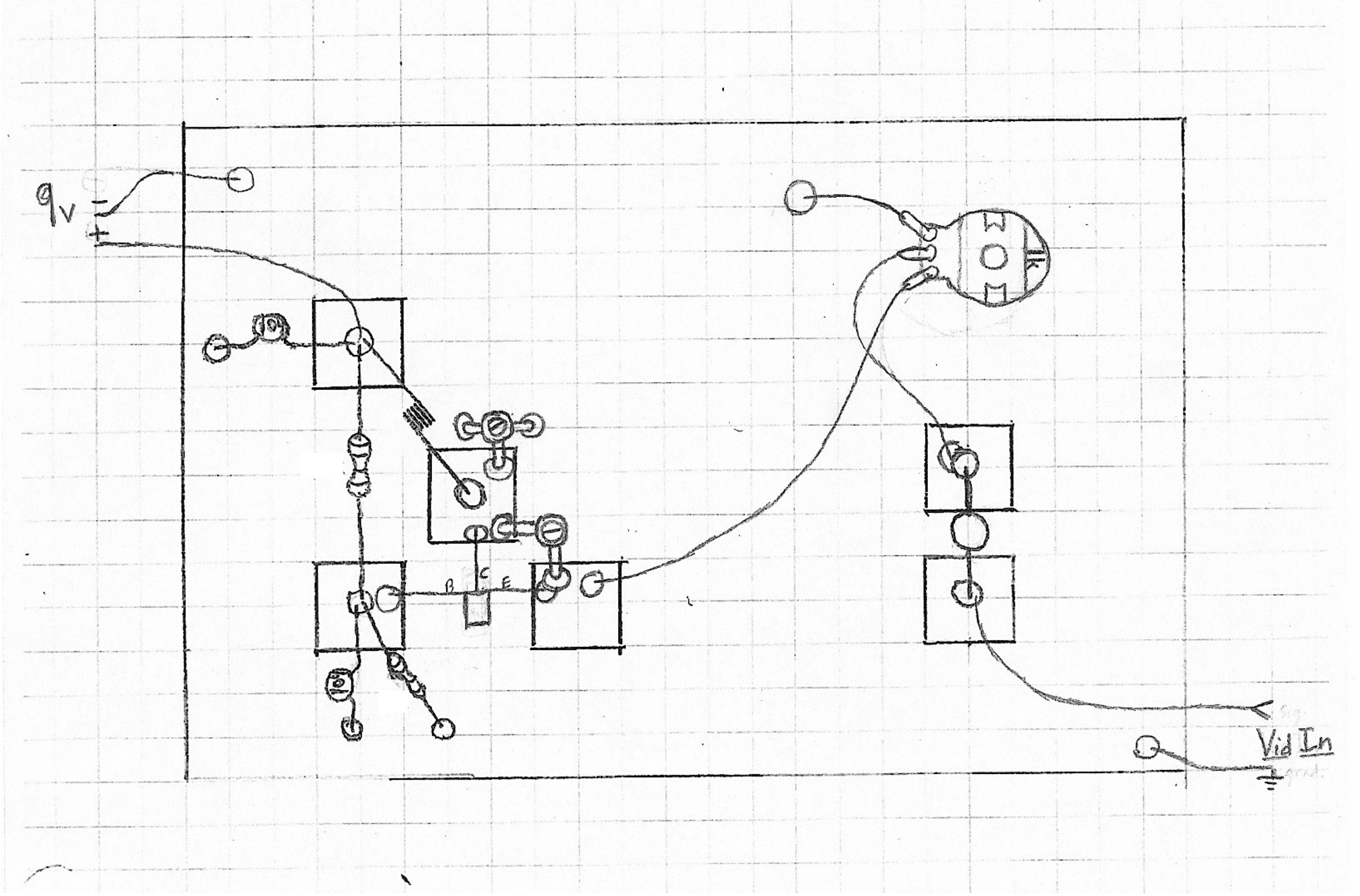

11. ADDING THE ANTENNA

- Cut about 2 feet of wire and strip one end.
- Solder the wire to the right most square of the triangle.

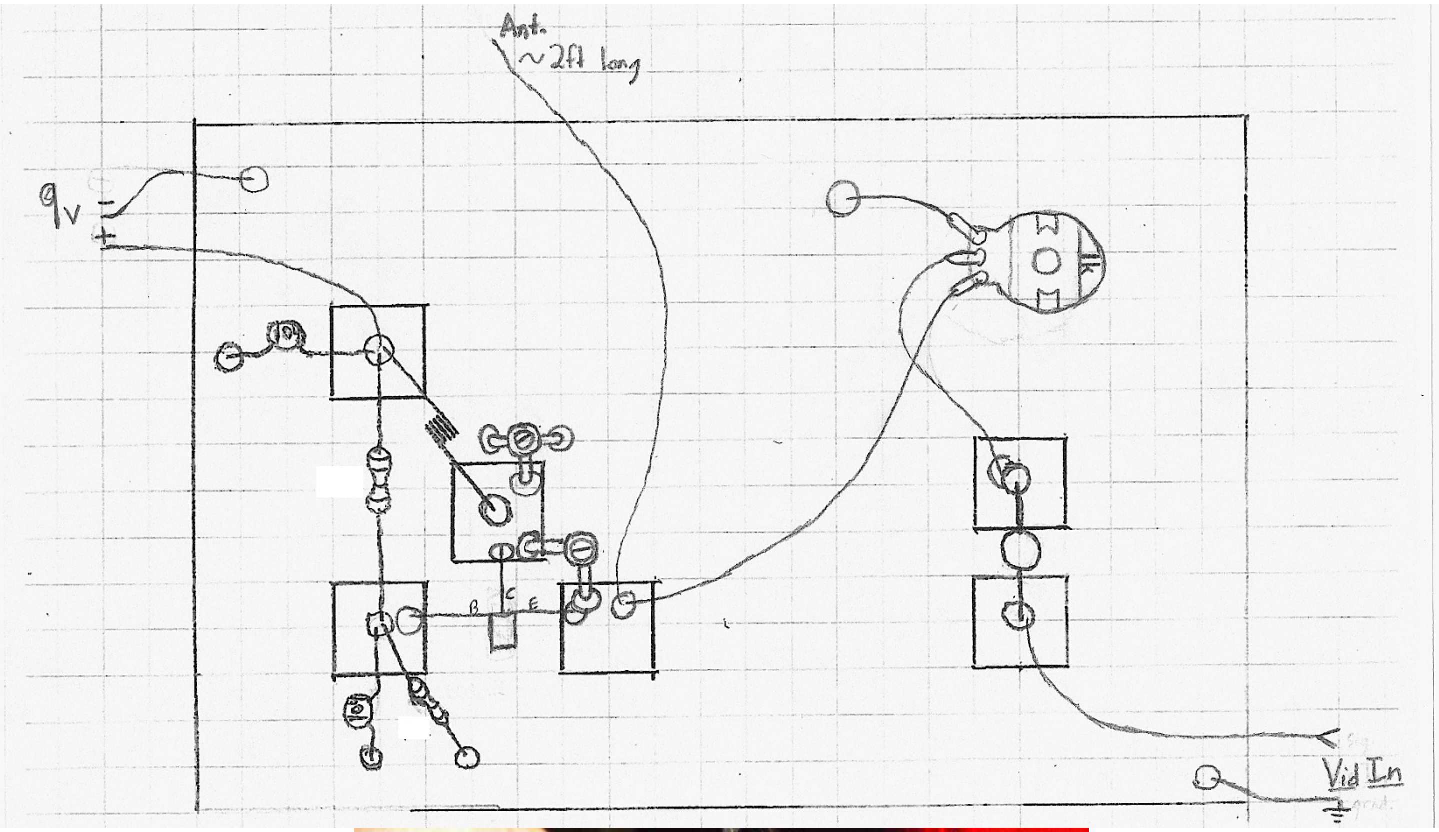
Ant.
~2ft long
9v
-
+
B
C
E
K
Vid In
sig
gnd.

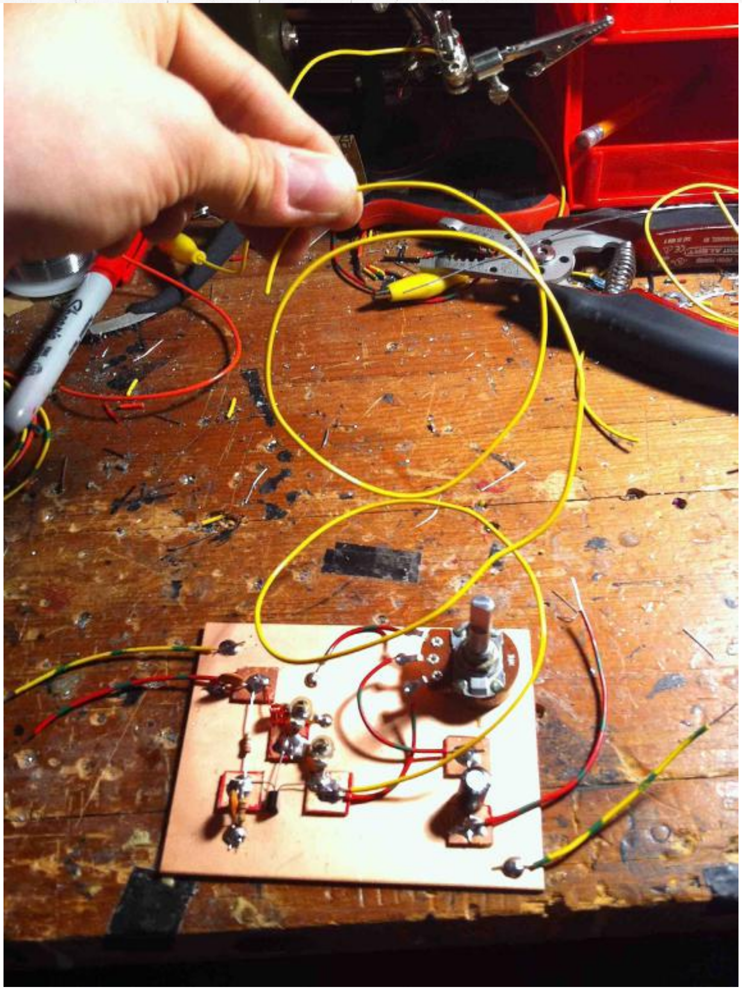

OPERATING

- Find a video source (DVD player, VHS player, iPod etc.) that has a composite video output.
- Use an alligator clip lead to connect the ground from your video source to the ground input of the transmitter. Using another alligator clip lead, attach the signal output from your source to the signal input of the transmitter.
- Use two more alligator clips to connect the positive and negative of a 9v battery to the positive and negative battery inputs of the transmitter.
- Turn on a CRT “tube style” TV and tune it to channel 2 or 3.
- Turn your potentiometer to somewhere around 50%
- Using a small screwdriver, tune the upper most variable cap slowly until you see the TV make a flicker or change.
- Tune the second variable capacitor until you start to see some strong signal coming through
- Tune the potentiometer and the two variable capacitors until you get the imagery you wan

Send any comments, questions or suggestions to yaktronix.online@gmail.com

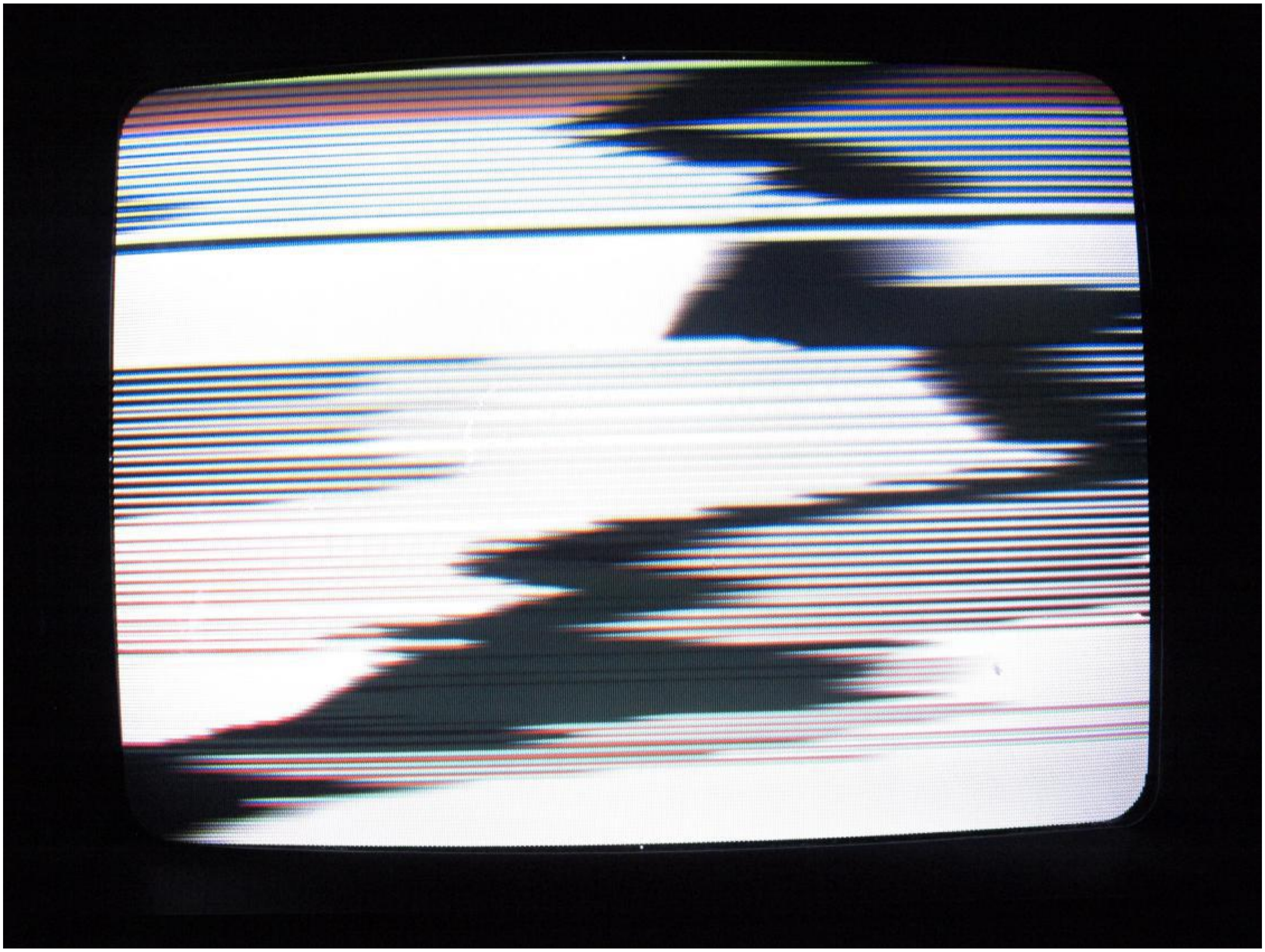

rexresearch.com

Paul BAUMANN

Testatika Generator

life.augmented
INNOVATIVE
SEMICONDUCTOR
SOLUTIONS
Digi-Key ELECTRONICS
Learn More

Methernitha Group: CH-3517 Linden, Switzerland

Photo Gallery & Videos

Links:

Methernitha ~ http://www.methernitha.com
http://utenti.tripod.it/testatikmachine/
Paul E. Potter ~ http://website.lineone.net/~aarekhu/index.html

Hans Holzerr: Testatika Demonstration Notes

Paul Potter: Back-Engineered Testatika

P. Potter: Principle Experiment

Anonymous: Believe It or Not --- Here It is!

Cyril Smith: Testatika Generator and Over-Unity

Methernitha: The Research Work of Methernitha in the Field of the so-called Free Energy"

P. Potter: Transcript of Methernitha Testatika Video

Testatika Demonstration (4 August 1999)

Translation by Stefan Hartmann < harti@harti.com > & Hans Holzherr

From: Hans Holzherr
To: Stefan Hartmann

Recently, over 30 technicians and engineers (most of them retired) were allowed to visit the Methernitha group in Linden, Switzerland, where they witnessed a demo of the different Testatika machines.

Here is a report from Hans Holzherr from Switzerland who was present:

Hello Mr. Hartmann,

To your questions: > Have you seen live a machine with a load? If so, what load?

I am referring in the following to the model with the 50-cm diameter disks. This machine was already running when the visitors stepped into the room, and was not halted during the whole time --- we were there for about 1.5 hrs. As a first load a 1000-Watt lamp was connected for approximately 10 seconds whose brightness did NOT diminish --- the corresponding sequence on the Testatika film is just an effect of the camera aperture's automatic adjusting to the sudden brightness! The second load was a U-shaped heating element, that Mr.Baumann handed to me. It became so hot within one second that I had to put it down immediately! What was particularly impressive was that while he pulled back one of the contact wires (that was with the lamp, I believe), a 1-cm long arc appeared between the output electrode and the connecting wire for approximately one second. The apparatus was under a plexiglass hood. Near the base it had two holes which Baumann used to insert the contact wires to touch the output electrodes.

How do you think the high wattage is produced ?

Good question! I'd love to know the answer, too!

Did the disks slow down when a load was placed across the output electrodes ?

I did not notice that (nobody else did), but of course you tend to turn your look to 'where the action is' (the lamp etc.) The disks turned with 15 rpm, which is quite slow. The spin rate was regulated magnetically.

What general impression did you have ?

It was really impressive! One can hardly believe it, with this slow rotation. In any case, this cannot be explained in terms of bare electrostatics in the sense of the Wimshurst machine. The perforated sheets seem to have a key function... Beside the pick-up and the drive electrodes there are a number of small plexiglass blocks with glued-on perforated sheets, whose function is unknown.

As Adolf Schneider already mentioned, my colleague Bernhard XXX and I want to try to copy the principle experiment shown by Baumann --- without much hope to find anything extraordinary, though.

The device consists of a horizontal swiveling plexiglass arm with a small rectangular plexiglass plate at both ends glued to the lower side of the arm. The lower side of the arm is covered with perforated aluminum sheets (square holes), while the bottom of the plates is covered with brass wire mesh. Beneath each plate five additional plates are glued onto the base plate. There is also wire mesh between each pair of plates in the two blocks. From the mesh layer between the lowest plate and the base a wire goes to the two capacitors, which are connected in parallel . Baumann seized the arm with both hands and turned it about ten times back and forth (a full rotation was not possible, because the capacitors were in the way), then measured the DC voltage with a digital measuring instrument: 60 Volts. Then, as he short-circuited the condensers a loud crack could be heard. I don´t know if that already is an abnormal result...

On my question Baumann replied that with metal foil (instead of wire mesh) the device would not produce that effect.

For more information, Contact: leo@zelator.in-berlin.de

"Back-Engineered Testatika"

by

Paul E. Potter

http://website.lineone.net/~aarekhu/index.html

(Copyright 2000)

That the Swiss Methernitha group's Testatika machine is thought to be based on a Wimshurst electrostatic generator, is only a sparse approximation of the truth --- of the great multitude of electrostatic influence machines developed around the 1900's it more closely follows the charge-separation-and-collection system used by the 1898 Pidgeon machine [note 1] for its electrical circuit.

Its 50-per-disc steel grilles or 'gitter-grilles' are plainly unique to the Methernitha (see fig.1) but in principle follow on from previous research and patents for corrugated sectors which were found to be more efficient charge carriers [note 2] than flat ones, and from a similar example in more recent times of aluminium rods extending out like wheel spokes from an insulating hub of perspex [note 3].

Another unique function of these perforated grilles attached to the discs is how they induce charge from the rotating discs onto the special collecting pads, or 'tasten' antennae keys (which are also perforated --- so as to more readily pick up charge); for in a Wimshurst you had conductive brushes or rails of sharp points which actually touched the discs or were placed very close to them, but in the Methernitha the charge has to be made to traverse a parallel air-gap to the pads and for this purpose the metal gitter-grilles are so designed to create miniature eddy-currents of charged air which circulate in and out of the perforated metal's surface charges, and are more easily bounced out to the collecting pads. This process is categorised as VARIABLE CAPACITANCE electrostatic generation.

Careful note needs to be made of how the Methernitha uses its basically Pidgeon setup with regard to its neutralising rods (that equalise and stabilise the opposite charges --- see fig.2), and how charges are picked up from one area and accumulated at others, so that the polarities of charge are distributed correctly to specific areas on both discs [note 4].

And although there have been some fanciful claims, or misinformation, that it uses all sorts of radioactive materials to achieve its pulsed output I most strongly believe that the auxiliary electromagnetic circuit, that wraps itself around the rotating discs, portrays a simple electronic approach; afterall, who would use radium radioactive emission alongside leyden jar capacitors ! Indeed, the more you look into certain elements of its construction the more they point to three main eras of electronics development, the 1900's, the 1920's and the 1950/60's. The authentic Methernitha was designed and developed by purists who believed they had discovered a previously unknown electronic phenomenon, but they wanted to keep an integrity to the early pioneering days of the Pidgeon, Wimshurst and Holtz electrostatic machinery; they would not use such modern devices as transistors or IC chips (more's the pity) --- but they do use some pretty uncommon electronic engineering in their circuit [note 5].

Obviously, the electronics are in two parts; one --- the electrostatic generator and its particular technologies of how to direct what charge where, and two --- the very unique auxiliary electromagnetic circuit of inductances, capacitances and rectification that mobilises that 'static' electricity. To understand how they convert static energy into an electromotive force you would do well to go back to the earliest years of radio. From the pages of spark radio you soon appreciate just how important oscillation circuits and their valve rectifiers were, and moreover, how difficult it proved to engineer them. For although radio transmitters and receivers from the 1900's used resonating circuits their oscillations were controlled by sparks between two contacts and, of course, they were relatively inefficient. Not until the 1920's did the first electric current oscillations become an observable controlled phenomenon when someone coupled a rectifier valve, a capacitor, and a resistor together [note 6]. The early 1920's also saw the best era of experimentation and invention for novel devices that turned static energy into useable electromagnetic energy; it was in a 1921 patent that we see a German physicist Hermann Plauson describe in great detail his methods to convert static power, not only from rotary influence machines but also from balloons collecting atmospheric electricity up in the sky; and by using thermionic rectifiers, leyden jar capacitors and inductor coils he proposed a free-energy network that was to power the whole of Germany [note 7] ! The thermionic rectifier valve heralded a new era for radio and high voltage physics, and as it was then subjected to such a broad array of experiments and modifications to improve its efficiency so it paved the way for all sorts of new avenues in electronics. Indeed, with such a technical catalogue of similarities with what we see in the available photographs of Testatika it can be assumed without doubt that the horizontal glass tube which sits on top of the Methernitha machines is exactly what a home-made vacuum thermionic rectifying valve would look like; with its internal anode mesh-plate, surrounded by a coiled copper grid, fed by a glowing (heated) cathode wire running horizontally across its centre and capped by two black end-pieces, which are too big and bulbous to be mere end-caps and must surely be black rubber vacuum seals to seal the glass tube and the input/output wires [note 8].

With such a rectifier, some induction coils, and some leyden jar capacitors you have a circuit that oscillates, and that's what has to happen with a Methernitha, the electromagnetic circuit has to oscillate for it to work, and then the oscillations have to be rectified (or even modulated) so that the resulting single-pole pulses can be channeled through the big cans, which are basically high-efficiency transformers, and outputted as reduced voltage higher current DC pulses (see fig.3).

The precise components used to oscillate the primary oscillating circuit are, I believe, not to be seen in any of the available photographs, but there are various hints for their approximate whereabouts on the machine. Firstly, according to electronic design there should be a capacitor and coil configuration in close proximity to the rectifier. Well, from the picture "3KWREAR" can be seen the two long upright tubes which, according to those who have seen them first-hand, comprise a spirally turned aluminium strip (which indicates they are chokes [note 9]) inside a glass tube, inside the same sort of outer shielding that the big cans have (which indicates they are electrostatic shields), inside yet another glass tube, and are terminated at the top with a brass connecting rod which does a right-angled turn and passes into the side of the tower --- but only two-thirds up the height of the tower. These two assemblies must form a connection to the rectifier, because the rectifier is at the top of the tower, so why don't these electrostatically sensitive tubes extend all the way up to it ? Again, from the photographs of the rear and front of the Methernithas there is a wire that comes out of the tower's side wall at about 4 inches above the upright's brass terminals and this wire then passes through a short black tube and on to the rectifier valve. This, of course, would happen on both sides of the tower, enabling a connection to both ends of the rectifier. But why have this 4 inch gap of connections at the top of the tower ? Something is placed inside the top of the tower in this intermediate space which is very necessary to the circuit, and I think it must be the location of the capacitor/inductor configuration to oscillate the circuit. This (fig.4) is how I would see the inside of the top of the towers [note 10].

I've seen some of the patented inventions that rotate discs --- by using magnets (i.e., H.Rosenberg's permanent magnet excited rotational machine, US Patent # 3,411,027), and by utilising inscribed metalised discs (US Patent # 3,239,705 for instance), but there simply isn't enough room for these to be located in the Methernitha disc setup --- also, you don't want to interfere with the ES fields that zip around the revolving discs: From the reports of those who have seen the small machines working it appears their discs were rotated by small DC electric motors after they were hand-started, some re-wound with thinner wire (to presumably increase their torque) and powered directly from the discs' generated electricity --- but I have also seen how two discs can continue to rotate simply by careful placement of curved electrodes [note 11] which would act on the charges on the discs --- like the 3kw Testatica Distatica generators.

After reading through the many early accounts of electrostatic rotary machines, and some of the more recent ones, you can't help but be puzzled by the Methernitha's incredibly low rotational speed of just 60 rpm (and in the 1999 engineers report as low 15 rpm !). Most other early experimenters boasted up to 3000 rpm. J.G.Trump in his work on high voltage generation in space [note 12] spun his rotary machine at 10,000 rpm (to produce 433 Watts at 24 KV no less). One reason for this low speed might be to do with the close proximity of the 50 lamellas (gitter-grilles) on the discs at their inner ends, they are very close together, I think too close. Air, normally an insulator, breaks down and conducts at around 25-35 KV (this figure has been fairly constant from day-one of electrostatic machine experiments right through to the present day --- because air has a breakdown field strength of 3x106 volts/metre) and short-circuits the circuit. I feel that because this design of grilles is prone to short-circuiting at high voltages the Methernitha people have limited their rotational speed so as to ensure a low operating voltage --- of what I'd guesstimate to be only 12 to 24KV.

But, is this a waste of extra potential ? Not necessarily, for I don't think that the main power output comes solely from what the two counter-rotating discs supply.

There is, I believe, a far more important power generator --- the electron cascade generator, and the Methernitha has two of them, held inside the two horseshoe magnets, and providing the circuits to the magnets are made to oscillate at the right frequency at a high enough voltage then these metalised-perspex laminated blocks can enmass A MUCH LARGER AMOUNT OF ELECTRICITY THAN WHAT IS PUT INTO THEM.

This, perhaps, is the previously unknown electronic phenomenon that the Methernitha group have so zealously been trying to protect against unscrupulous entrepreneurs. But I would say that this copious supply of free energy is already known to the world --- it is not readily available - and its principles are not fully understood, as yet, but it is known about.

As the descriptions say (on the Testatika website), between the horseshoe magnet legs are four blocks of transparent 'plexiglass' type material alternated with copper and aluminium plates (that may or may not be perforated), in the sequence c-p-a-c-p-a-c-p-a-c-p-a (also see fig.6). And according to the Linden Experiment, where Paul Baumann induces a resonance of about 80-140 MHz in a coiled horseshoe and then has an aluminium-insulator-copper block moved between the horseshoe legs, a voltage could be taken off the plates of the block which measured 700 volts (DC presumably) [note 13]. This incredible phenomenon has never been replicated by any 'outside researcher', and is said to be the basis by which the Methernitha machine could be understood how to work [the clue, possibly, to this Principle Experiment may be variable-capacitance and dielectric-absorpsion].

But what, I hear you say, is an electron cascade... Well, it was only by chance, very recently, that I happened to listen to an audio tape by a Dr. Flanagan about crystal water; when I switched the tape over after the end of side one Dr. Flanagan then began talking about an electronic configuration that applied a high frequency, high voltage alternating field across an insulator --- that created what he called an electron cascade effect --- Yes, I thought, here is the answer to the Methernitha Machine.

The electron cascade or avalanche effect is where air molecules are accelerated to the device at such a high velocity that they collide with other molecules and atoms in the air to liberate new electrons which in turn also collide and liberate even more 'free electrons' from other air molecules (see fig.5), all of which become accelerated by the electric field, and an avalanche of electron-multiplications progresses throughout the whole immediate environment [note 14]. It's a chain reaction, and an entirely safe one, it happens in a more ferocious way in lightning strikes, and is a natural phenomenon. And, as in this case, the environment actually becomes part of the circuit [note 15] because the process is actually negatively-ionising the air surrounding the Methernitha machines, and that is why those who have been near these generators when working say the air around about them is cool and fresh [note 16].

In view of the fact that it's designers have chosen to wind insulated wire (which may or may not be bifilar) around the horseshoe metal it is likely that the horseshoes are used for some form of induction [note 17], it would also be very possible to draw directly from this part of the circuit the extra electric current produced from the electron cascade blocks, with suitable connections that might lead downward into the wooden base (where it is believed that an alternate layering of perforated metal plates and insulating plates - making up a large high-voltage storage capacitor - is located). This power could then be discharged as a pulsed output of high wattage, especially if the final output part of the electronic circuit is configured as a Pulse Forming Network of multiple sections of inductor / capacitor combinations [note 18].

The two big cans at the side, the big capacitors, are probably not highly technical (see fig.7), once the fundamental formula has been decided upon all models of a Testatika generator would follow a similar construction process. The written descriptions are a little contradictory but they seem to suggest a central input rod, or tube, connecting at the bottom of the cans to a stack of inter-linked pancake coils, that are wound secondary-outside primary-inside, fitted around a core of 6 hollow donut-ring magnets stacked in such a way with plastic spacers as to allow air gaps between them, and then finally the output of each can is a connection from the top coil of the secondaries of the pancake coils to a brass ring around the centre of the black plastic top lid --- and from the photographs can be seen a large diameter wire or tube [note 19] connecting that polarity's output terminal to the top lid's brass ring via a brass screw terminal. I would suggest that the ring magnets (of anistropic ferrite perhaps) are gapped in this way to prevent the magnetic flux fields of the pancake primaries co-joining as one sprawling field, because it would be more advantageous, and safer, to have each separate pancake's magnetic flux cut it's own adjoining secondary coil, and divide the secondary output voltage into smaller amounts of potential, thus depending less on complicated insulating procedures that accompany high voltage single primary / single secondary transformers.

The use of aluminium mesh and solid copper sheeting is commonly used in electronic construction; the outer aluminium mesh cylinder would be used to shield stray electrostatic charges, and the solid copper cylinder is to shield the large amount of stray electromagnetic fields produced by the transforming process from high voltage/low current to lower voltage/higher current [note 20]. Obviously they don't want field contamination taking place between the sensitive electrostatic generator and the transformers.

Within these two outside shielding-cylinders are 'grid condensers' which, according to the 1999 report by the 30 engineers, can be as many as 20 layers of perforated sheet (presumably as concentric cylinders) - which I have indicated (in fig.7 for instance) as being electrically connected

BETWEEN each separate secondary winding - in the fashion of an old discovery from the early days of wireless telegraphy and based on the 'disruptive discharge coil' devised by Nikola Tesla, that such a condenser connected in the center of a secondary coil collects the maximum amount of voltage created by that secondary. This configuration of one condenser inside another inside another etc etc, has a striking similarity to the layout of a pulse forming network [see note 18].

In the red wired can the transformer is wired to output negative, and the blue wired can's transformer is wired to output positive polarity. Special note should be made of a similar arrangement for divided primary / secondary windings devised by Van de Graaff in his 'High Voltage Electromagnetic Charged-Particle Accelerator Apparatus Having an Insulating Magnetic Core' [note 21] with respect to magnetic reluctance gaps.

Whilst it has been said that the clear perspex disc was designated the 'cloud' disc, and the (rear) dark disc the 'ground' disc I would think this relates to different types of acrylics or plastics that might become charged to different polarities, as in the triboelectric series, where frictional charging of different plastics --- and then bringing them close together, might cause donation or acceptance from one to the other; I would think from the above that cloud represents a donator (positive charge) and that ground must mean an acceptor (negative charge). Has anyone tried the combination of a teflon disc (extremely negative charge) with a glass disc (highly positive charge) ?

Or discs doped with paramagnetic particles perhaps [note 22] ?

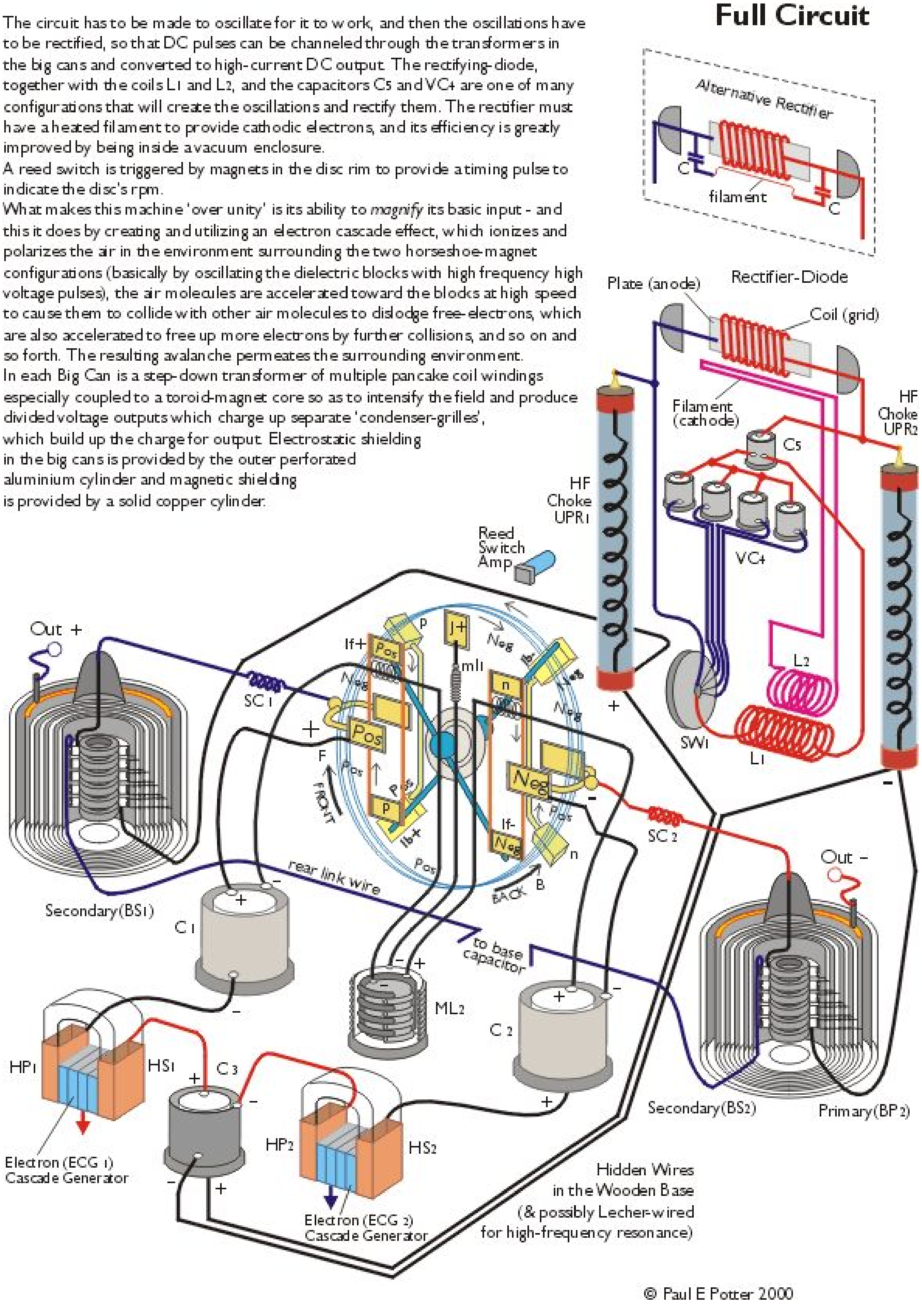

The circuit has to be made to oscillate for it to work, and then the oscillations have to be rectified, so that DC pulses can be channeled through the transformers in the big cans and converted to high-current DC output. The rectifying-diode, together with the coils L1 and L2, and the capacitors C5 and VC4 are one of many configurations that will create the oscillations and rectify them. The rectifier must have a heated filament to provide cathodic electrons, and its efficiency is greatly improved by being inside a vacuum enclosure.

A reed switch is triggered by magnets in the disc rim to provide a timing pulse to indicate the disc's rpm.

What makes this machine 'over unity' is its ability to *magnify* its basic input - and this it does by creating and utilizing an electron cascade effect, which ionizes and polarizes the air in the environment surrounding the two horseshoe-magnet configurations (basically by oscillating the dielectric blocks with high frequency high voltage pulses), the air molecules are accelerated toward the blocks at high speed to cause them to collide with other air molecules to dislodge free-electrons, which are also accelerated to free up more electrons by further collisions, and so on and so forth. The resulting avalanche permeates the surrounding environment.

In each Big Can is a step-down transformer of multiple pancake coil windings especially coupled to a toroid-magnet core so as to intensify the field and produce divided voltage outputs which charge up separate 'condenser-grilles', which build up the charge for output. Electrostatic shielding in the big cans is provided by the outer perforated aluminium cylinder and magnetic shielding is provided by a solid copper cylinder.

Full Circuit
Alternative Rectifier
C
filament
C
Plate (anode)
Rectifier-Diode
Coil (grid)
Filament (cathode)
HF Choke UPR2
C5
HF Choke UPR1
VC4
Reed Switch Amp
Out +
SC1
If+
Pos
P
J+
Neg
mlI
n
Ib-
Neg
Neg
Pos
+
F
Pos
FRONT
Pos
P
Ib+
Neg
If-
Neg
Pos
n
Pos
BACK
B
+
SW1
L2
L1
-
SC2
Out -
rear link wire
Secondary(BS1)
+
-
C1
to base capacitor
-
+
ML2
-
+
-
C2
-
HP1
HS1
+
C3
-
HP2
HS2
Electron (ECG 1) Cascade Generator
-
+
Electron (ECG 2) Cascade Generator
+
Secondary(BS2)
Primary(BP2)
Hidden Wires in the Wooden Base (& possibly Lecher-wired for high-frequency resonance)

© Paul E Potter 2000

Back-Engineered Methernitha --- Notes

Note 1 --- For more information on the Pidgeon machine see "Electrical Influence Machines" by John Gray, 1903 p. 206 & "Philosophical Magazine" Dec 1898, p. 564, and of course the Pidgeon patents.

Note 2 --- See "Modern High Speed Influence Machines" by V.E. Johnson, 1921, p. 76. Johnson was not only a researcher of electrostatic machines but was also an innovative constructor of them, and as such was keen to try any technique that made his generators more powerful than even the specialised Wommelsdorf multi-disc condenser machines. This book is an absolute must for those who wish to work in this field. Another 'must' is the website of Antonio Carlos M. de Queiroz (http://www.coe.ufrj.br/~acmq/electrostatic.html) which is absolutely full of information about (and with links to) present-day developments in electrostatic machines.

Note 3 --- See "Self-Excited, Alternating, High-Voltage Generation Using A Modified Electrostatic Influence Machine" by M.Zahn, et al., American Journal of Physics, Vol 42 (1974) p. 289.

Note 4 The Methernitha designers have taken a basic Pidgeon electric field system and added a few modifications of their own, partly to lock a certain polarity of charge to a certain area so as to stabilise it, and also to boost certain areas with charge. As in their use, for example, of an extra field plate located at the top-centre in front of the front disc (just under the rectifying valve), note also that this plate, or antennae key, is indirectly coupled to the rest of the circuit, via a coil setup. Much the same occurs with the two plates slightly below it, these plates are connected to a brass terminal which connects to a copper wire that goes down and winds in a coil shape around a hollow plastic tube, and inside the tube will be another wire or small coil that draws off the electric charge. So these three plates are using not direct connection but induction to get their charge.

Note 5 ------- By looking at how each of the photographed machines have been constructed you can see that these are high quality crafted structures. I would think each would start off as sub-assemblies fitted together by pairs or small groups of members, those sub-assemblies of wooden base, big cans, perspex framework, discs with bearings and axles, when completed would be passed on to the electrical engineers of the community who would then fit the wiring connections, vacuum tube rectifier and make sure that not only did they work but that they looked like a work of art.

Note 6 --- The Fleming valve had been around since 1905 and while it progressed to the thermionic valve and audion, by 1922 the ‘Pearson and Anson Effect’ was discovered whereby oscillating currents could be produced with a resistor, capacitor and thermionic valve coupled together.

Note 7 --- See US Patent 1,540,998 (9 June 1925) Conversion of Atmospheric Electric Energy by Hermann Plauson. He also wrote a book of the subject titled "Gewinnung und Verwertung der Atmospharischen Elektrizitat" in 1920 in German (which is currently held in the British Library).

Note 8 --- Whilst some have seen the smaller 300 Watt machine’s discharger/rectifier quite open and not encased in a vacuum tube the vacuum tube models would be much more efficient and would waste less current. Also, the rectifier tube must have a heated filament (which on the 3KW machines can be seen as a glowing line running the whole length of the grid and coil assembly between the two black end caps, and in the films you can see faint flashes coming from behind the rectifier so possibly the filament is wrapped around the other side of the grid/coil assembly as well). Coolridge, back in the 1900’s, discovered that no discharge from the cathode to the anode would occur, even at 100,000 volts, unless the filament was heated (Physics Review, Vol. 2, Dec 1913, p. 418). Aluminium mesh will give off electrons quite readily and can be used as a cold cathode --- but a heated cathode offers the advantage of being able to control the oscillations.

Note 9 --- The two long upright tubes are without doubt choke coil assemblies in precisely the right place to slow down the current where it gets oscillated and rectified. In a choke the higher the flow of current the greater will be its resistance to that current flow. An even better form of choke will have some form of iron core inside it.

Note 10 --- I have come up with 6 different circuits for this oscillation section, some of which include small quartz crystals. (See notes 13 and 16 on frequency of oscillation). The black dial at the rear of the 3kw machine is most likely to select a variety of capacitances so as to control the oscillations of the circuit, which in turn control disc rotational speed.

Note 11 --- The phenomenon of electrostatic motors has been well researched over the years (see "Electrostatic Motors" O. Jefimenko in "Physics Teacher" Vol. 9, March 1971, p. 121-9, and in "Electrostatics And Its Applications" by A.D. Moore (1973) p. 131-147; "Electrostatic Motors" by B.Bollee in "Philips Tech. Review" Vol 30 (1969), p. 178-194). The Methernitha Testatika generators (see a recent report by 30 engineers) auto-rotate, after they have been started by hand, by the same principles of these ES motors.

Note 12 --- J.G.Trump worked for the US Air Force and pioneered some highly efficient electrostatic machines around the 1960's (see"Electrostatic Sources of Electric Power" in "Elec. Eng." 66:525, June 1947; and "High Voltage Generation in Space:The Parametric Electrostatic Machine" in "Progr. Astronaut. Rocketry" (vol 3 --- Energy Conversion for Space Power) 1961 p745).

Note 13 --- Although the 'Linden Experiment' was thought to register a frequency of 80-140 MHz this does not necessarily mean that the Methernitha generators would oscillate at that rate also. Such a frequency seems unnecessarily high.

Note 14 --- See "Plasma --- The Fourth State of Matter" by D.A. Frank-Kamenetskii (1972) pp10, and Dr.Patrick Flanagan's US patents 4,743,275 (May 10, 1988) and 4,391,773 (Jul 5, 1983).

Note 15 --- The effect is very similar to the converging forces in a non-uniform field, the oscillating perspex blocks become one 'electrode' and the surrounding air in the room becomes the opposite 'electrode', and by the processes of electrophoresis and dielectrophoresis the electrically charged particles in the air (the electrons and negative ions) are drawn toward the central electrode, which in this case is the perspex block assembly (see "Nonuniform Electric Fields" by Herbert A. Pohl in "Scientific American" (Dec 1960) p. 107-8). I am much more inclined to believe that the ingenuity of the design of these types of machines comes from physicists and not electronic engineers.

Note 16 --- Dr. Flanagan actually uses the electron field generator in his own special ionizer (see Method of Purifying Air and Negative Field Generator US Patent 4,391,773).

How does an electron cascade generator work, I would think that while you have an alternating electron movement (and Dr. Flanagan reckons this effect occurs with a high voltage field alternating at above 20 KHz) at the metal electrodes, the perspex blocks sandwiched between them would transfer the electricity not through their mass but around it, as surface charge --- actually in the layer of air right next to the insulator's surface. Its the same principle as dielectric absorption - the perspex blocks don't discharge themselves fast enough to keep up with the alternating voltage and so they accumulate more and more charge --- until it forms as a layer of charge on the insulator's surface. This means that at a high enough frequency the surface-air molecules polarize, with the more mobile electrons separating from the slower bulks of those molecules and while the electrons get thrusted back and forth a secondary layer of (slower) positive air ions develops, and so on, and the process of high voltage high frequency polarization triggers the electron avalanche effect.

In the event that the perspex blocks are indeed ELECTRETS (as free-energy researcher Geoff Egel and others suggest) I would think that they would work in a similar fashion to the above process, of dielectric absorption that charges up the blocks before they produce the cascade-effect. Because in the electret the electrons charged into the perspex/plastic, and the positive ions, would still be manipulated by the reversing electromagnetic field in such a way as to orientate (as with dipoles) back and forth, to eventually reach the point (if the whole circuit is tuned properly) where they would attain resonance with the immediate air surrounding them. And if this effect is similar to an inductance then possibly a back-emf will result also, to increase the output voltage. Either way I believe the effect will still be an electron cascade through the environment and the product of this oscillating output (at the blocks) could similarly be drawn off and accumulated in the multi-layered base capacitor network.

I would suggest that a test program to find the best type of blocks would be; One --- try different types of plastic/acrylic/ceramic materials for the blocks.
Two --- try different methods of electrifying the plastics (as with electrets). Three --- try plastics doped with semiconducting particles. Four - try plastics doped with paramagnetic particles. Five - try hollow plastic blocks containing an electrolytic fluid. More information on plastics will be found on the Electret vs Dielectric Absorption Page.

Note 17 --- There are several definitions of Bifilar, one where the wires cancel out their magnetic fields, and one where the wires are wound to ensure a tight low-loss magnetic flux coupling, in this case you need all the magnetic flux you can get so it must be the latter --- See "Transformers For Electronic Circuits" by Nathan R. Grossner (1967), p. 224 etc.

Most commonly used magnetic metal is Mumetal, which is an easily saturable magnetic material, routing magnetic flux through it rather than in the surrounding air, so as to enhance the mutual induction between the two coilings of red wire around the horseshoe legs.

Note 18 --- So that the machine's output voltage doesn't drain away when it is connected up to a large load, what is needed here is a Pulse Forming Network (or artificial delay line). "Such a network is an improvement on simple capacitor storage because of the cascading action from one capacitor to the next along the chain. At the beginning, all capacitors are charged to the same voltage but as soon as the first one starts to loose voltage, the one behind it is then free to discharge into it. This topping-up action, which trickles down the network from capacitor to capacitor, is the mechanism by which the voltage across the output terminals tends to hold onto its original level." (see "High Energy Discharge Systems" A.P.Stephenson, Electronics Today International, March 1992, pp. 24-26).

Note 19 --- When voltage of a high potential and high frequency flows along a wire it does so on the outer surface (called the ‘skin effect’) and so the Methernitha would use thick wiring or even 1/8" tubing to connect its circuit.

Note 20 --- Two references for shielding are: "A Shielded Loop" by S. Goldman in "Electronics" Vol 11 (1938), pp. 20-22; and "Measurements in Radio Engineering" by F.E. Terman (1935), p. 218 & p. 341.

Note 21 --- For information on maximum voltage in the center of a secondary coil see "A Handbook of Wireless Telegraphy" by J. Erskine-Murray (1913) p. 42; and an article called "Dielectric Hysteresis at Radio Frequencies" by E.F.W. Alexanderson in "Proc. I.R.E., Vol. 2 (June 1914) p137-157. For Van de Graaff's transformer see US patents 3,323,069 (May 30, 1967) and 3,187,208 (June 1, 1965). These patents were not just for a Van de Graaff high voltage generator, they were for a special system devised by Van de Graaff long after his generator had been in use to convert static electricity into current electricity. This system may be a little too complicated for the Methernitha but, nevertheless, the principles he used for multiple primary / secondary windings may be of some interest.

Note 22 --- Dr. Flanagan modified his insulator blocks, made of resin, by doping them with paramagnetic granules (such as silicon carbide) to enhance even more the electron cascade effect; which is an idea that the physicist Thomas Townsend Brown first experimented with (by using lead oxide granules) in his US Patent 3,187,206 (June 1, 1965) to good effect. The surrounding air could also be ‘enhanced’ in similar fashion to polarise it's electric charge and improve its side of the performance (for those interested in the ‘physics’ of this see an article by W.A.Douglas Rudge "On Some Sources of Disturbance of the Normal Atmospheric Potential Gradient" in Proc. Royal Soc. A, Vol. 90 (1914) pp. 571, etc).

Note 23 --- Some other generators with similarities to the Testatika machine are the "Electrostatic Energy Field Power Generating System" invented by William W. Hyde (US Patent 4,897,592 of Jan 30, 1990) is a rotor/stator variable capacitance machine capable of producing 300 KV. Other such generators are; "Parametric Electric Machine" invented by Ferdinand Cap (US Patent 4,622,510 of Nov 11, 1986) which has a series resonant (LCR) circuit structured into it so that it oscillates --- and indeed operates AT RESONANCE to ensure a high output; "Electrostatic Generator" invented by Dan B. Le May, et al. (US Patent 3,094,653 of Jun 18 1963) is a very ingenious system of variable capacitance; the "Electrostatic Machine" by Noel Felici (US Patent 2,522,106 of Sep 12, 1950) is a good standard which utilizes a valve rectifier; and the "Electrostatic Generator" by William S. Spencer (US Patent 1,415,779 of May 9, 1922) is an early rotor/stator generator which transferred its electric impulses through a transformer to produce a higher current output.

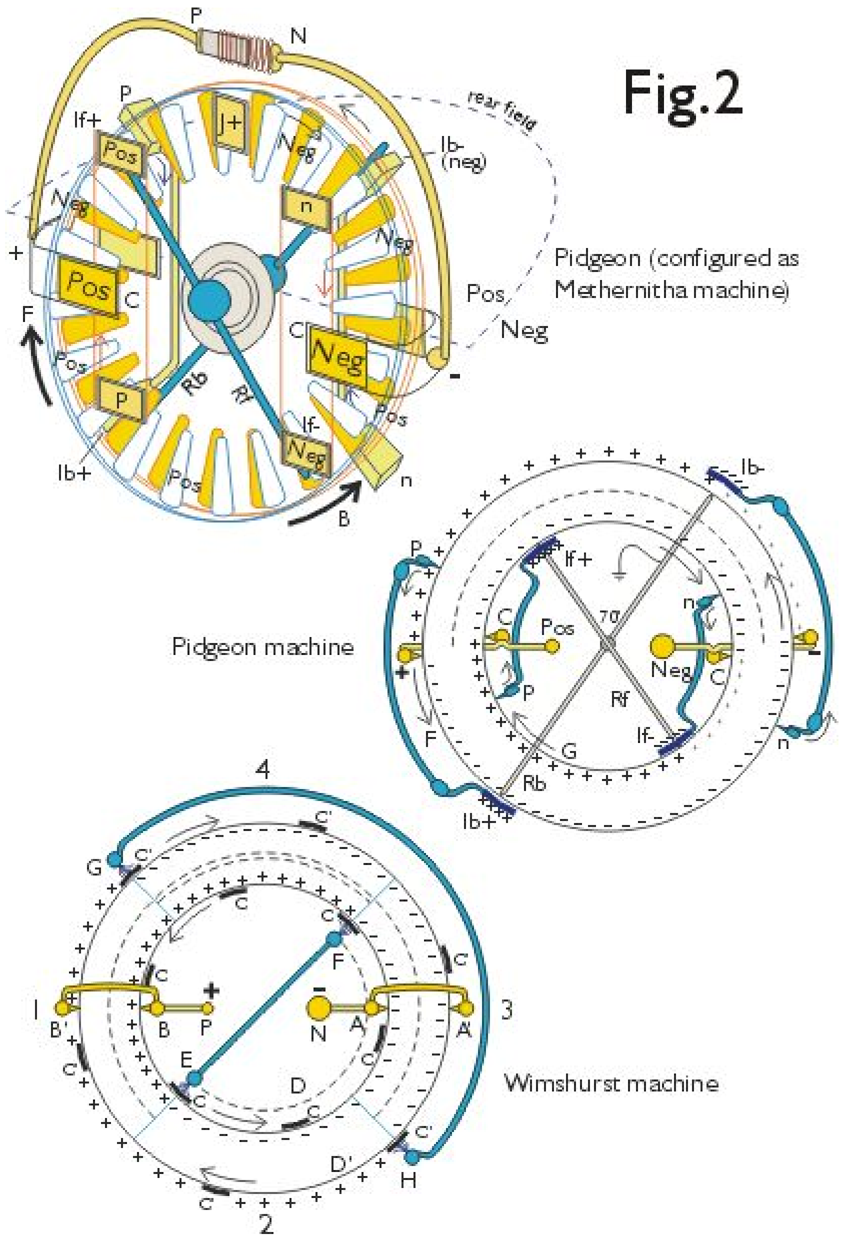
Fig.2

P
N
P
If+
J+
Neg
rear field
Ib-
(neg)
Pos
n
Neg
Neg
+
Pos
C
F
Pos
P
Rb
Rf
If-
Neg
Ib+
Pos
n
B
Pos
C
Neg
Pos
Neg
-
Pidgeon (configured as Methernitha machine)
Ib-
P
If+
70
C
Pos
n
Neg
C
P
Rf
F
G
If-
n
Rb
Ib+
Pidgeon machine
4
C'
G
C'
C
C
F
C
C
1
B'
B
P
N
A
A
3
C'
E
C
D
C
C
C'
D'
H
C
2
Wimshurst machine

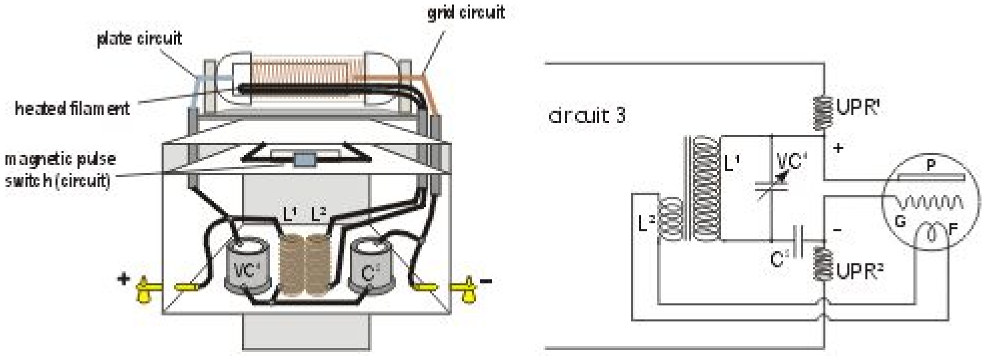
grid circuit
plate circuit
heated filament
magnetic pulse
switch (circuit)
L¹ L²
VC¹
C¹
+
-
circuit 3
UPR¹
L¹
VC¹
+
L²
C¹
-
UPR²
P
G
F

Fig. 4 - Tower Circuits

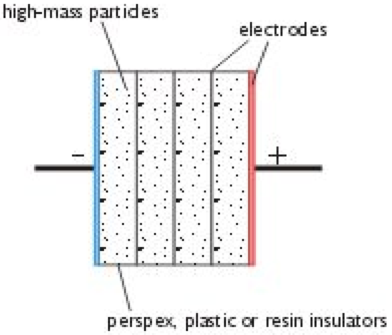
high-mass particles
electrodes
-
+
perspex, plastic or resin insulators

Fig. 6 - ECG Doped

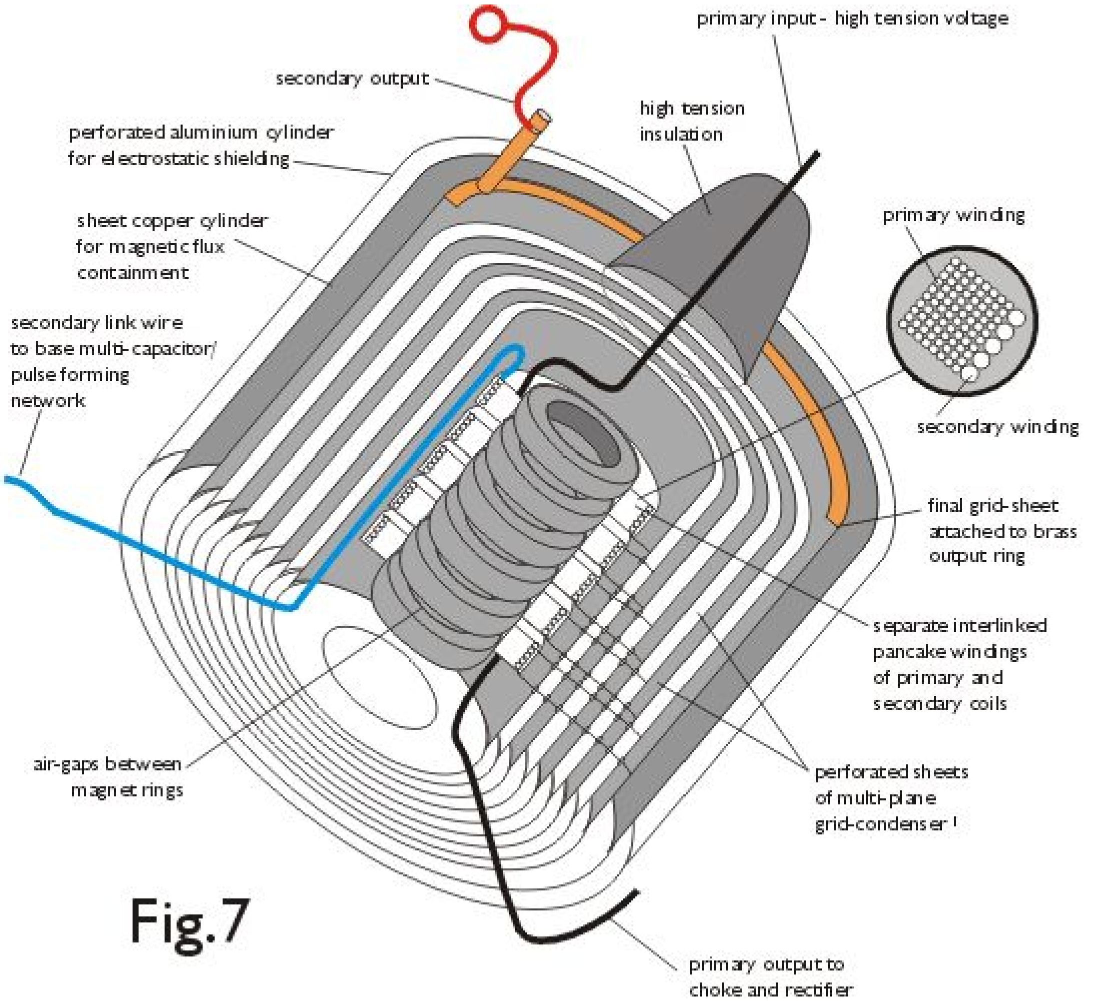
primary input - high tension voltage
secondary output
high tension insulation
perforated aluminium cylinder for electrostatic shielding
sheet copper cylinder for magnetic flux containment
primary winding
secondary link wire to base multi-capacitor/ pulse forming network
secondary winding
final grid-sheet attached to brass output ring
separate interlinked pancake windings of primary and secondary coils
air-gaps between magnet rings
perforated sheets of multi-plane grid-condenser ¹
primary output to choke and rectifier

Fig.7

[1] In the Testatika-Distatica Video it has... "By means of **grid-condensers** the energy is stored and it is then uniformly discharged, at the same time reducing the high-voltage and building up power with additional devices."

"Principle Experiment"

by

Paul E. Potter

With respect to the 30 engineer's report (of 1999) the 'principle experiment' was set up in such a way (see diag 1) that by swivelling the cross-piece over two other plexiglas blocks a current was transferred into two capacitors. Possibly when someone grabs hold of the swivelling cross-piece they transfer some static electricity from their body to the grilles of the cross-piece which, when vigorously turned back and forth will initiate a transfer of electric charge to the two stacks of blocks below them attached to the base. This may either be because anyone walking over a carpet or wearing man-made fibres will already hold thousands of volts of electrostatic charge - which will be enough to 'prime' this apparatus after touching it, to get it to work, much the same as when a rotating influence machine has to be primed with an electric charge before it will work; or, as many researchers have surmised already, that the plexiglas blocks glued to the swivelling cross-piece had been charged up with electricity as an ELECTRET.

The use of mesh around the cross-piece guarantees a variable capacitance to the moving arms --- and as these are moved quickly over the base blocks (again, alternated with mesh and Plexiglas) an amount of electric charge will be transferred to them EACH TIME the arms go backward-and-forward over them. Electronically, you have one variable capacitance acting upon another variable capacitance - which is why Baumann said that if metal foil were used instead of the mesh it would not produce the same effect.

What happens next is called "dielectric absorption", and the following extract, from "Dielectrics" by P.J.Harrop (1972) pp71, will explain;

"When a given dielectric [i.e., plexiglas] has a field put across it, and this field is then removed, the electrodes are briefly shorted out (not enough for all the ions and particles to relapse back) and the device left on open circuit one finds subsequently that it has partially charged up again.

This is due to the slow-moving charge carriers remaining in position. It can be a frightening phenomenon with practical dielectrics since a person who has briefly discharged a charged insulator may subsequently touch it and be severely shocked." Some dielectric materials being more prone to this effect than others, and so obviously some experimentation will be needed to choose the best perspex, plexiglas, plastic, or acrylic materials so as to take the most advantage from this dielectric absorption effect.

Once the blocks are being charged up the electric charge simply flows into the two interconnected capacitors and accumulates.

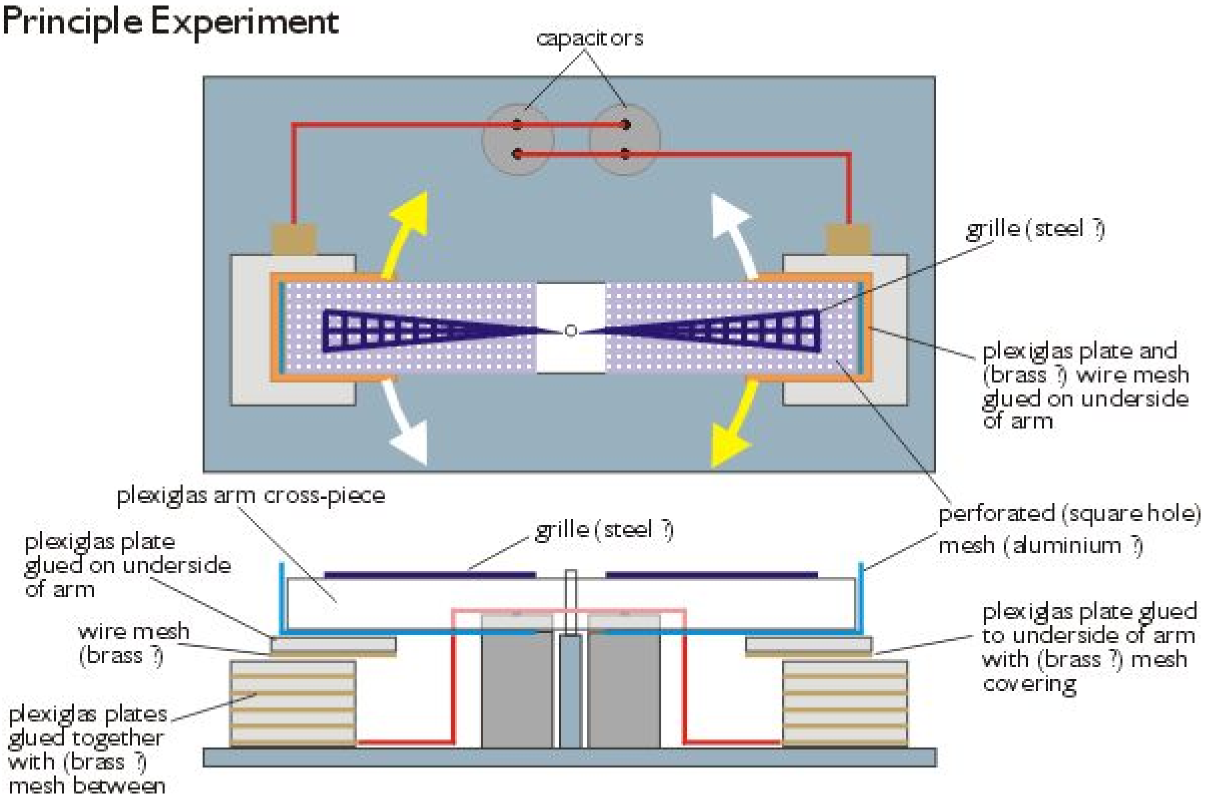
Principle Experiment
capacitors
grille (steel ?)
plexiglas plate and (brass ?) wire mesh glued on underside of arm
plexiglas arm cross-piece
grille (steel ?)
perforated (square hole) mesh (aluminium ?)
plexiglas plate glued on underside of arm
wire mesh (brass ?)
plexiglas plate glued to underside of arm with (brass ?) mesh covering
plexiglas plates glued together with (brass ?) mesh between

Linden Experiment ~

A look at the only available diagram of the Linden experiment (see diag 2) shows a horseshoe magnet coupled to what seems to be a closed-ended wire. It can also be seen in this diagram that the wire coils around the magnet are sufficiently spaced apart so as to provide capacitance (between each coil of the wire) in the circuit. So, with capacitance, coils (for induction), and magnetic flux, you have all the ingredients needed to make an oscillating circuit --- provided there is a supply of voltage and some sort of interruption mechanism - then you have a circuit that will resonate. And this leads us back to that oft repeated question --- is the block used in the Linden Experiment an electret. The first person to PUBLISH how they have duplicated this experiment will be the one to answer that question...

But there is something wrong with this diagram --- the resonant frequency of this simple circuit --- depending only on the values of inductance and capacitance included in the circuit would consequently enable it to oscillate at only a few kilohertz at best. No circuit of capacitance/inductance can be made to resonate at such a high frequency as claimed, not even if it included any form of quartz crystal. If, as the diagram says, it resonates at 140 MHz then there must be a Lecher wire system (the running of two conducting wires in a parallel line, about a foot apart or less, for several feet) to make it resonate at such a high frequency. It would seem to me also that a required modification of the Linden Experiment diagram would be to put a different metal into this circuit where the wire is said to be "closed-ended", say for instance a zinc plate, then there will be a contact voltage established between the copper wire and the zinc plate. Such a voltage would start the circuit resonating, for the reasons stated above.

And as with the Tini setup if a block comprising two metal plates separated by a suitable dielectric material (such as plexiglas) is placed in an oscillating electric field the electric charge on the metal plates will permeate into the dielectric, and for the same reasons as detailed above in the Principle Experiment, will temporarily accumulate voltage because of dielectric absorption. As I have already said elsewhere there is a great similarity between the phenomenon of the Testatika horseshoe blocks (and the Tini setup) and the principle behind Patrick Flanagan's Electron Field Generator which energizes an electron cascade mechanism throughout the immediate environment.

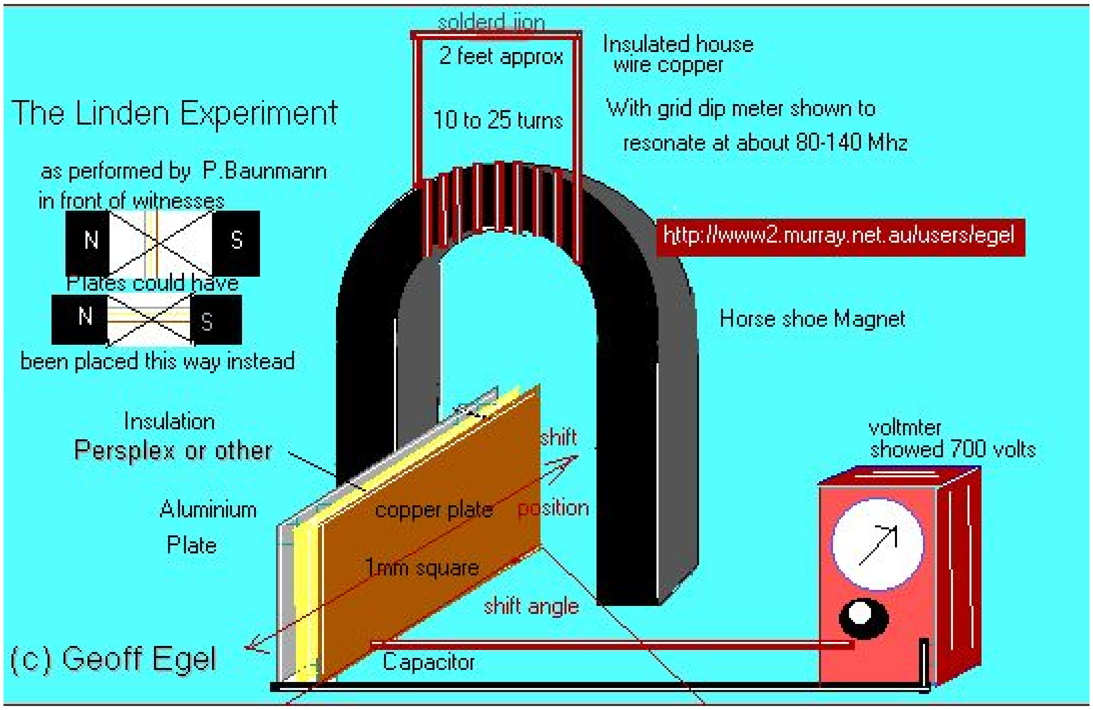
solderd jion
2 feet approx
10 to 25 turns
Insulated house wire copper
With grid dip meter shown to resonate at about 80-140 Mhz
The Linden Experiment
as performed by P.Baunmann in front of witnesses
N
S
Plates could have
N
S
been placed this way instead
http://www2.murray.net.au/users/egel
Horse shoe Magnet
Insulation
Persplex or other
Aluminium Plate
copper plate
shift
position
1mm square
shift angle
voltmter showed 700 volts
(c) Geoff Egel
Capacitor

Geoff Egel (1997) --- http://www2.murray.net.au/users/egel ~

Rectifier Circuit ~

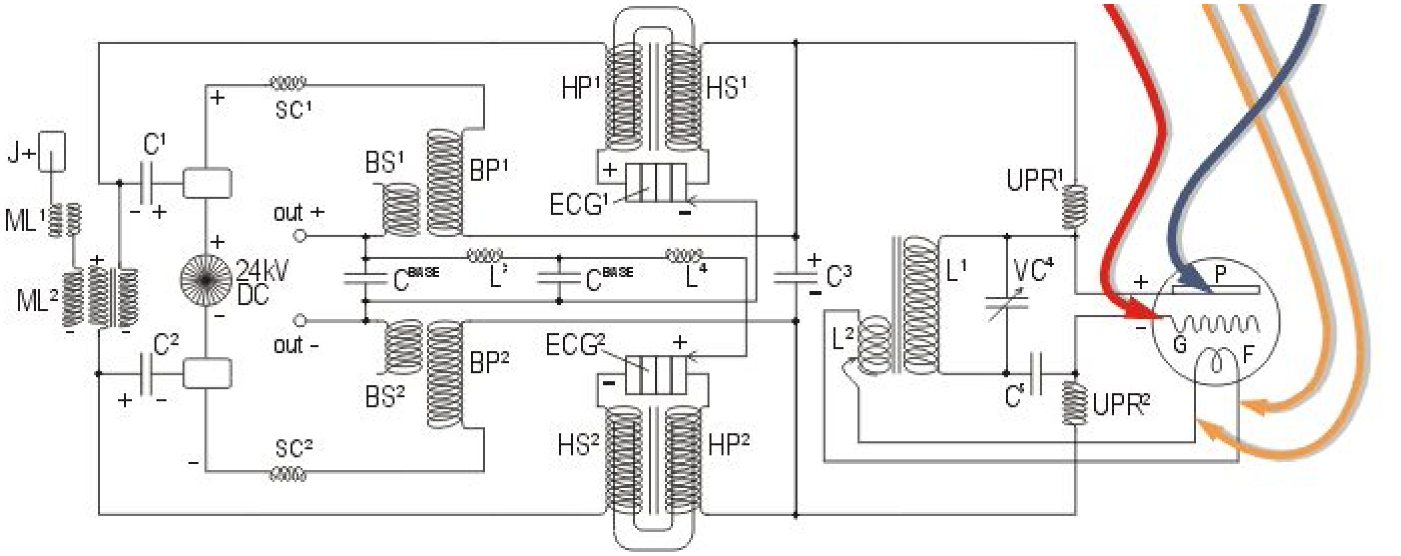
J+
ML¹
ML²
C¹
C²
SC¹
SC²
24kV DC
out +
out -
BS¹
BP¹
BS²
BP²
HP¹
HS¹
HS²
HP²
ECG¹
ECG²
C^BASE
L³
C^BASE
L⁴
C³
L¹
L²
VC⁴
C⁵
UPR¹
UPR²
P
G
F

The rectifier forms the basis for Testatika's operation For whilst the LC circuits L^1, VC^4, UPR^1, UPR^2 and C^5 set up oscillations from the generator's high voltage output the rectifier controls the oscillations to provide positive dc pulses (which are then directed through a step-down transformer BP^1/BS^1 and BP^2/BS^2), to supply power at a higher current rate and lower voltage.

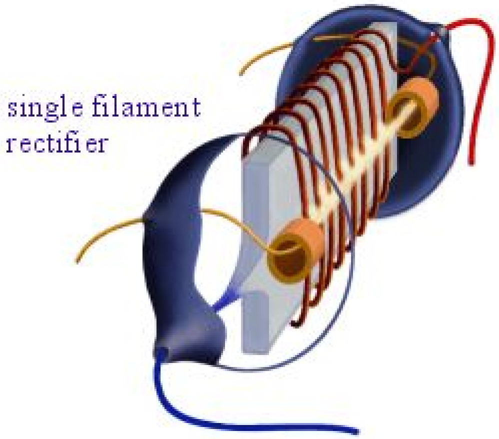
single filament rectifier

Unidentified European Report ~

"Believe or Not, Here It Is!"

The first really running "free energy" machine coming from Switzerland, Europe. It was developed over a 20 years Research period by a religious group called: METHERNITHA. This group lives in CH-3517 Linden, Switzerland. The inventor of this superb machine, Mr.Paul Baumann claims, its running principle was found by studying the lightning effects from nature.

And here are the facts :

* Testatika is an influence-type Wimhurst machine which runs on it's own energy, once started by counterclock revolving it's 2 dics by hand !!!

* Testatika not only runs on it' own energy, but produces also a huge amount of excess power, at least 3 KWATTS of power! That is enough to supply a one-person apartment with one machine!

* The machine you see during this demo-animation is only about 70 cm wide, 40 cm deep and about 60 cm tall !

* It delivers DC-Voltage ranging from 270 til 320 Volts, only depending on the dryness of the air! At this voltage it can supply at least 10 Amperes of DC-current !

* Testatika is not a perpetuum mobile, but an energy machine that collects it's huge amount of "FREE" energy from the charged and ionized air particles. But there are still some technological tricks implemented to overcome the normal drag-resistance of a conventional Wimhurst machine, which is still the secret of the Methernitha group !!!

When will this technology be available for everyone?

The Testatika machines, (there already exists also some smaller units which only deliver about 200-300 Watts) are not yet mass-production type models! They are still laboratory prototype units, although they are build with a very good craftmansship !

For more info contact directly:

METHERNITHA, CH-3517 LINDEN, Switzerland
TEL.-Nr.: ++ 41 31 97 11 24

The pictures are from some photos and from a video-tape Methernitha sells. It shows the machines (also the smaler ones) and explains also Methernitha's spiritual aims. The tape has a running length of about 40 minutes and is really worth looking at it! It is also available in VHS-NTSC and American language. It also shows Methernitha's earlier research developments like huge wind generators.

With the Testatika there is no fraud, hidden batteries or any other hidden energy source. It is really running! They already have the technology of ten years ahead! No more oil, no more pollution, no more atomic waste products, no more hunger in this world! This is the machine the mankind has dreamed about for centuries. Now it has come true!

More information about the Methernitha machine is also available in the German magazine *RAUM & ZEIT*, issue Nr.40 (8).

Jean L. Naudin Labs (4 February 1998) ~

"Testatika Generator and Over-Unity"

by

Cyril Smith

Having read Nelson Camus' article purporting to "explain" the Testatika over-unity generator (http://members.aol.com/overunity2/nelson/testatic.htm), I am of the opinion that the real explanation is hidden amongst the 19th century electrical mumbo jumbo which makes up much of the machine (Leyden Jars, Horseshoe Magnets, Wimshurst Generator and so on). From the description of spiral-wound foil capacitors containing radioactive material and placed within current carrying coils it strikes me that perhaps the real energy comes from Beta particles (fast electrons) absorbed into the electrical circuit.

To extract energy from Beta particles it is not just enough to capture them. Clearly this will result in a current flow according to the capture rate, but current by itself is not power. The kinetic energy of the particle must be captured and this will result in potential difference or voltage. If a Beta particle enters a thin foil conductor normal to its surface and is captured, then the potential difference associated with giving up its kinetic energy can be expected to occur across the opposite faces of the foil. If however the Beta particle were turned so as to enter at a shallow grazing angle to the surface, not only would the probability of capture be increased (since the particle will travel further within the material) but the potential difference will occur along the length of the foil. Beta particles travelling even at relavistic velocities can be turned within short distances by relatively weak magnetic fields, so the concept shown in the following Figure suggests itself.

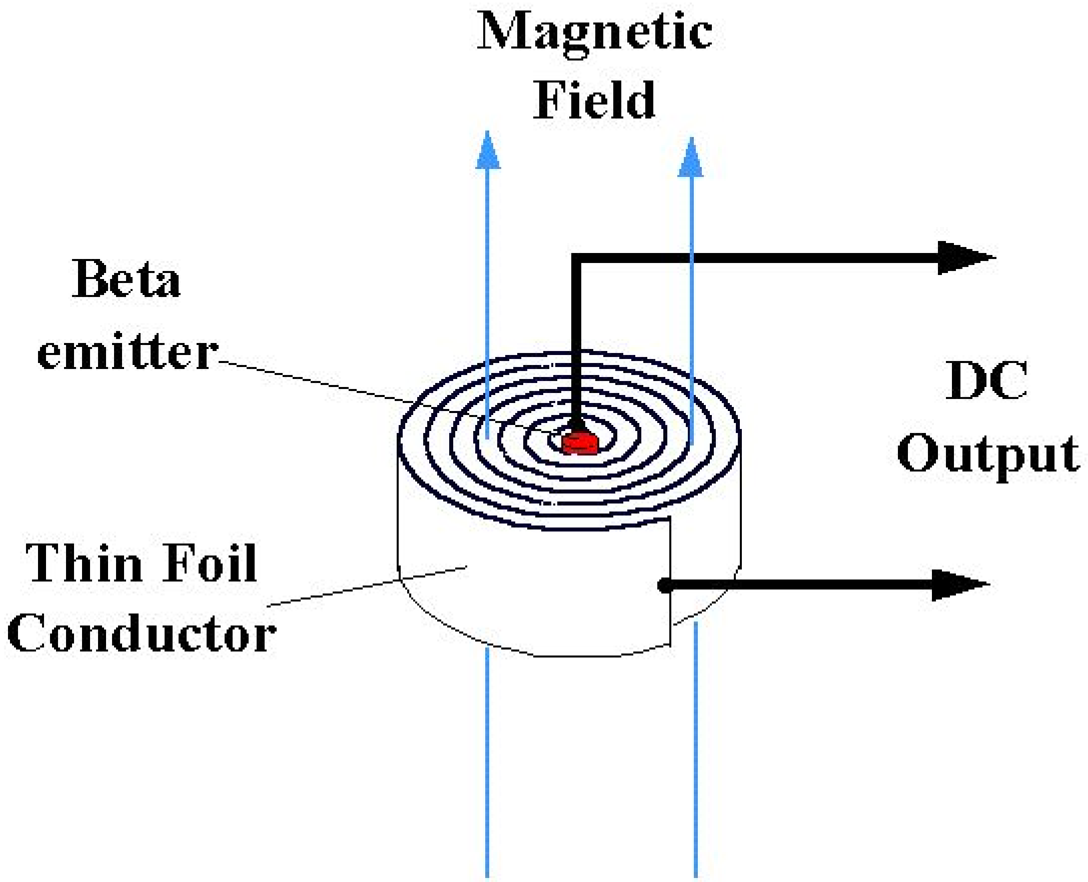

Methernitha Statement of Purpose ~

"The Research Work of Methernitha in the Field of the so-called Free Energy"

Ever since the foundation of Methernitha there existed a department for research, development and electronics, which was concerned with the problem of alternative energy sources, namely with technologies, which were suited to exploit the inherent forces of nature and thus to unlock sources of energy without disturbing natures ecological balance in any negative way. Any technology man invents should serve him in short as well as in long terms, and this condition is not fulfilled as soon as it opposes nature in any way.

This team of research within Methernitha works completely autonomous and is financed out of the co-operatives own resources, without any outside support.

The efficient utilisation of wind energy was one of the first objectives of Methernitha's research program. At the beginning generators with special excitation were developed, which allowed to load the cells of accumulators even at low RPM at times when there are only moderate air movements.

Utilising the kinetic energy of water currents was another field of interest of the development team, but it was pursued more as hobby. The key problem was here to transfer the slow revolutions of the waterwheel to an extent that the excitation threshold of the generator could be surpassed by a minimal loss of energy.

Also solar cells and solar heat collectors attracted the attention of our researchers since a long time. But since in these fields other institutions have attained outstanding results, Methernitha started, and this was already more than 30 years ago, to concentrate its efforts on lesser known and even generally unknown sources of energy. The result of this scientific work is the Thestatica machine.

The question arises: How it comes, that Methernitha, nothing more than a private organisation, could invest so much time, engagement, perseverance and financial resources in this kind of research.

Research and development are integrated parts of the general idealistic concept of Methernitha. To make you understand these ideals and goals, we will now introduce you into the practical sides of this spiritual community Methernitha, as it functions as a co-operative of people, living and working together.

Linden is a calm village of farmers just as dozens of others in the region of the valley called Emmental. However Linden is also the home of a special form of human life, which is capable of attracting more and more people from all over the globe, and which has repeatedly caused astonishment through its excellent functioning, even in extremely difficult situations.

There is no doubt that the formation of Methernitha is also a good part of the life story of Paul Baumann. Through his extraordinary technical capabilities, but also through his astonishing practicable wisdom he opened relations to all kind of people with ideals and in many places, and pretty soon, the idea to form a co-operative enterprise took shape.

"In the name of the God Almighty" people of equal thinking gathered and founded a co-operative, which could serve as the economic basis of this spiritual community.

"We want to be a united group of brothers, and never separate however severe the burden may be", this was the solemn vow.

Renunciation of alcohol and smoking and the will to realise a harmonious community life without dispute and discord like in the original Christian communities were the prerequisites to become a member. For the members Methernitha is an ideal opportunity to lead a fulfilled life by practising charity.

A workshop was erected and soon, one house after the other grew on these premises. Also all construction was done out of own resources and solely with the savings from the common work.

Today Methernitha is a social model proving the quality of its ideals through its well functioning just more than forty years. The people working here are almost without exception members of Methernitha.

Not even the prospering development that took place over the following decades could impair or even suppress the global idealistic aims. On the contrary, everybody works in his own interest with diligence and great joy towards the erection and the support of their new homeland and according to the principle:

One for all and all for one!

With this word one can conquer anything. Also the ancient truths appear again and attain respect.

The people living here feel themselves as members of a family, like a group sitting in the same boat and proud, but also grateful proprietors of their own beautiful homeland, which they may shape exactly as they wish to. Evidently, this form of human social life can function only on the base of idealistic principles.

The question arises, how the realisation of a sincere religious philosophy of life may be brought in harmony with a successful economic management. It is not obvious at all, that this is attainable.

Nowadays there are many amongst us that are caught in the world-wide and dense network of social and economic dependencies and obligations and many also feel the confrontation with its problems.

One member of the economic directory has recently phrased this in the following way:

"The fact that all essential functions of Methernitha are fulfilled without any external force, driven solely through inner conviction, which causes everybody to help and take care of the other, this is for me the most astonishing effect which is produced by this form of living together. It seems to be a miracle."

Another miracle within Methernitha is the Thestatica, which is the result of more than 20 years of research.

This wonder machine is lurked from nature, nothing else. Nature is the greatest source of power as well as knowledge which man has, and it still conceals many secrets, which are only revealed to those, who approach and tie in with them with highest respect and responsibility.

To understand nature and to perceive its voice, man is obliged to experience silence and solitude, and it was there, where the knowledge about this technology was obtained.

For these reasons it was always a great concern of Methernitha to acquire properties, untouched by man as far as possible, be it in valleys or forests, in the mountains or ashore of lakes, where one could study nature, ones own being and the creator of all this universe in silence and concentration and without being disturbed.

The public never understood this properly, rather interpreted it wrongly as an act of seclusion, assuming we had to hide something unclean. We actually had --- and still have --- to take great troubles to realise undisturbed all the things we intended to accomplish.

Such undertaking in research and development necessitate considerable financial expenditures. Therefore we are quite often obliged to construct things with most primitive means and materials. What was thrown away by our affluent society, we collect and possibly set up with it a cornerstone in the discovery of new forces and truth.

We are fortunate in gaining the experience that paradoxically the most beautiful and useful results can be achieved by just using the most simple means. Never did we use any borrowed capital because we want to stay free Swiss citizens and do not want to be hindered or even bound in any way in the pursuance of our aims.

The two conterrotating discs generate an electrostatic charge. One disc represents the earth, the other the cloud. Using grid electrodes the charges are bound. After that they are collected by non-contacting so-called antenna keys and then sorted.

After being initially turned on by hand, the discs rotate by themselves according to the electrostatic laws about attraction and repulsion. A rectifying diode keeps the cycles in steady state. Otherwise the impulses of attraction and repulsion would accumulate and cause the discs to run faster and faster. The correct speed is of great importance and for optimal power generation the discs have to run quite steady and slow.

By means of grid condensers the energy is stored and then uniformly discharged, at the same time reducing the high voltage and building up power with additional devices. Finally the machine supplies a uniform direct current, which varies according to the size of the model. The machine furnishes about 3-4kW permanent output, depending on humidity, whereby the electric potential ranges from 270 to 320Volt. High humidity of the atmosphere prevents the build-up of electric potential. The drier the air, the better.

No doubt, through the so far achieved results one main objective has been reached, namely to prove that it is possible to use Free Energy. Nevertheless the research work is not yet completed.

To the educated physicist many a thing of this machine may seem impossible, maybe even crazy. Maybe he is also offended by the conceptions used to explain the whole. Only partly we could use the concepts of conventional physical terminology to explain and define only approximately the functions and properties of the various parts of the machine.

After all it will be necessary to create some more new concepts like the one we have already used before, when we termed the non-contacting collectors of electric charges as antenna keys.

This machine puts experts, which are just trained in conventional physics to a very hard test, because its mode of action is not explainable with the state of the art of officially accepted physical knowledge, or at the most only partially explainable. However also a trained specialist should remain free and independent in his thinking, and should avoid to be limited by the temporal framework of publicly admitted knowledge in any science.

It has to be remembered that the established science was already many times forced to change or give up some of its very fundamental concepts. Think about Galilee, to name only one example. Our human society almost condemned this man as a sorcerer and magician, just because he investigated and discovered a truth that seemed unacceptable by the established science of the days.

The book knowledge of any times is not wrong, but it is incomplete, and therefore allows to draw wrong conclusions.

We are part of a new area which brings to light many new facts and new knowledge. The clothes of nowadays science have become too tight and should be stripped off, just as the larva of an insect does with its skin. Only this will allow a true metamorphosis to take its course, and finally, at the limits and at the destination of all worldly knowledge, an universal and unlimited spiritual science, radiating and beautiful, just as the completed imago of an insect, may give its blessings and benediction to a renovated humanity.

To ever reach there, a universal development of man has to take place. This however is only possible when man becomes aware of his true role within the whole creation, and again learns to recognise his true tasks, because the whole universe functions within a strict and precisely structured order according to the will and word of the creator. Therefore also man should recognise and realise these universal laws that are valid within the whole and within every part of this creation.

The hard facts rather show how far man has left the divine order through his self willed and authoritarian way of action and that he has become the actual cause of all discord and evil on this planet.

Unfortunately the ruling bodies that should be responsible for the well-being of the people work too often with the target to make life more and more difficult and to render impossible every free spiritual development.

Instead of utilising the achievements of science and technology for the benefit and preservation of all form of life, they are abused carelessly and irresponsibly to destroy and to kill, and thus turn them into a curse upon mankind.

To change all this, the evolution of a new technology is not enough, even if it were the most ecological and ingenious. To change this present status one has to go much deeper down, to the root-cause of all this evil, and this is mans way of thinking, his state of mind.

The ancient divine commandments are still valid today and also show today the way and direction humanity should go, as clearly spoken by the prophet Mica:

God has told you what is good. And what is it that the Lord asks of you? Only to act justly, to love loyalty, to walk wisely before your God.

"Testatika-Distatica Machine"

(Transcribed from the Methernitha Testatika video by Paul E Potter)

"... The efficient utilization of wind energy was one of the first objectives of Methernitha's research program.

At the beginning generators with special excitation were developed which allowed to load the cells of accumulators even at low rpm, at times when there are only moderate movements.

Utilizing the kinetic energy of water currents was another field of interest of the development team, but it was pursued more as hobby. The key problem was here to transfer the slow revolutions of the water-wheel to an extent that the excitation threshold of the generator could be surpassed by a minimal loss of energy. Also solar cells and solar heat collectors attracted the attention of our researchers since a long time. But since in these fields other institutions have obtained outstanding results Methernitha started, and this was already more than 20 years ago [since 1960], to concentrate its efforts on lesser-known and even generally unknown sources of energy, the result of this scientific work is the Testatika machine of which most of you may have heard already.

The two counter-rotating discs generate an electrostatic charge. One disc represents the earth, the other the cloud. Using grid-electrodes* the charges are bound. After that they are collected by non-contacted, so-called, antennae keys** and then sorted [shorted ?].

After being initially turned on by hand the discs rotate by themselves according to the electrostatic laws about attraction and repulsion.

A rectifying-diode keeps the cycles in steady state, otherwise the impulses of attraction and repulsion would accumulate and cause the disc to run faster and faster. The correct speed is of great importance, and for optimum power generation the discs have to run quite steady and slow.

By means of grid-condensers* the energy is stored and it is then uniformly discharged, at the same time reducing the high-voltage and building up power with additional devices.

Finally, the machine supplies a uniform direct current which varies according to size of the model.

The machine furnishes about 3 to 4 Kilo-Watt permanent output, depending on humidity, whereby the electrical potential ranges from 270 to 320 volt. High humidity of the atmosphere prevents the build up of electric potential. The dryer the air the better.

No doubt, through the so-far achieved results, one main objective has been reached, namely, to prove that it is possible to use free energy. Nevertheless, the research work is not yet completed: To finish a model which can be handed out more or less to anybody and without any haphazards, also to non-specialists, much work and also time will still be needed.

Only partly we could use the concepts of conventional physical terminology in order to explain and define only approximately the functions and properties of the various parts of the machine. After all, it will be necessary to create some more new concepts, like the one we have already used before, when we termed the non-contacting collectors of electric charge as antennae keys**. This machine puts experts who are just trained in conventional physics to a very hard test..."

* The German original has gitter-condensers, with gitter defined as wire lattice, grille, grate.

** the original word used is tasten or taster, which means antenna, probe, key.

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

Started prototyping receiver to go with Michigan Mighty Mite Transmitter

wandrson Jul '16

Jul 2016
1 / 2
Jul 2016
May 17

In August, there will be two classes where we build a small one transistor morse code radio transmitter. The second class still has four tickets available, so if interested email me to reserve your seat!

August 20th Build a Michigan Mighty Mite Transmitter

In preparation for a couple of classes in September, I started prototyping a matching radio receiver for that transmitter. It will be a very basic direct conversion receiver with crystal control to match the frequency of the transmitter. With luck I will have the prototype completed by the first class on August 3rd

While waiting for some parts, I started assembling the front end to the receiver on some scrap pcb material.

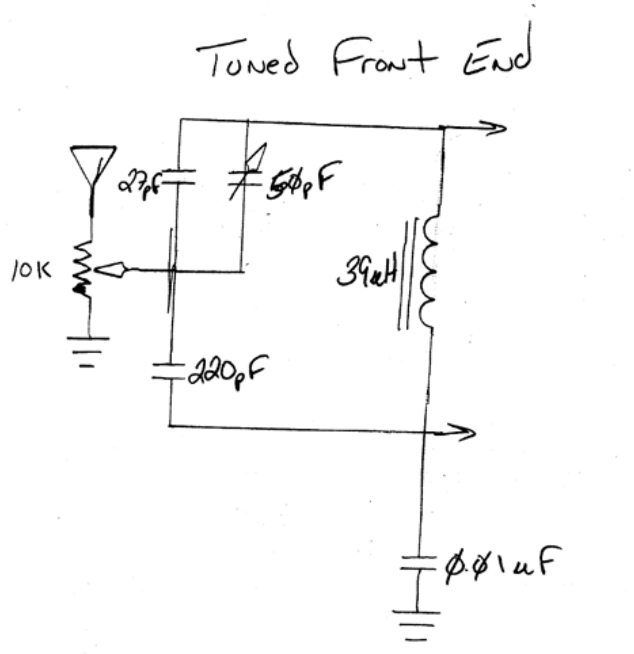

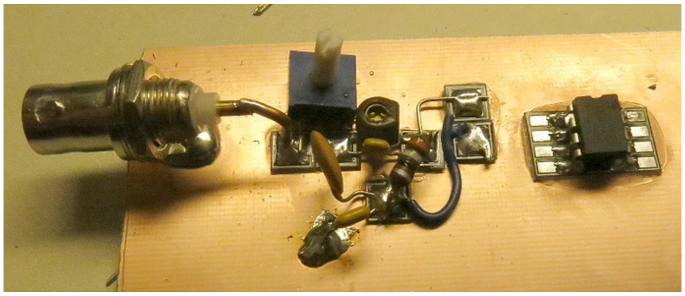

When working with projects like this, it is best to test (and if nescessary) fix the project in stages. So the first stage is a simple tuned front end (inductor-capacitor tank circuit). To ensure I did the math correctly, I plugged one of the crystals into a test oscillator I have and fed that into the antenna input and hooked up my scope to both the input and output of the tuned circuit as shown below.

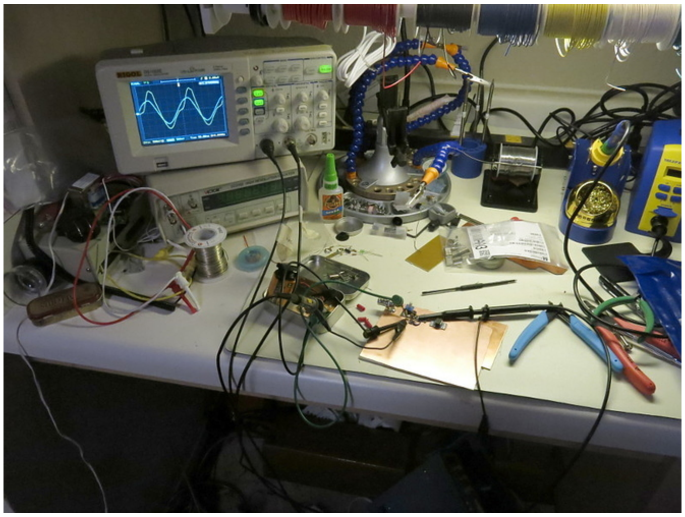

The circuit is aligned by using a non-conductive alignment tool (@artg_dms the Electronics committee could use a couple of sets of these) to rotate the little trimmer capacitor to maximize the amplitude of the output.

Here is a short video demonstrating the process:

If this kind of thing interests you, please sign up for the class (we can/will offer repeats of the transmitter class if there is interest) and come out to our monthly special interest group meetings for the Amateur Radio SIG!

Dallas Makerspace Show and Tell - August 2017

---

zmetzing May 17

> wandrson:
>
> If this kind of thing interests you, please sign up for the class (we can/will offer repeats of the transmitter class if there is interest) and come out to our monthly special interest group meetings for the Amateur Radio SIG!

I'm interested. 😉

Another interesting design for a 80m/40m receiver is here:

http://www.schripsema.org/pa3hdf/projects/dc-receiver/dc-rx.html

No. 96

CRYSTAL SET CONSTRUCTION

SIMPLE INSTRUCTIONS FOR MAKING NINE CRYSTAL SETS

by B. B. BABANI

There is to-day a marked revival of interest in crystal receivers, and this book has been specially prepared to introduce the home constructor to the design and building of the various types. The text is simple and free from technicalities, so that even the young reader may confidently embark on the construction of any of the receivers shown.

CONTENTS

1. AERIALS.
2. EARTHS.
3. CRYSTAL DETECTORS.
4. COILS.
5. THE BEGINNERS' CRYSTAL RECEIVER.
6. FRAME AERIAL RECEIVER.
7. LONG DISTANCE RECEIVER.
8. HIGH GAIN RECEIVER.
9. HIGH SELECTIVITY RECEIVER.
10. BATTERY AIDED CRYSTAL RECEIVER.
11. MEDIUM AND LONG BAND RECEIVER.
12. ULTRA SENSITIVE ALL WAVE RECEIVER.
13. AN ADVANCED CRYSTAL DIODE RECEIVER.

THERE ARE 12 DIAGRAMS

LONDON: BERNARDS (PUBLISHERS) LIMITED

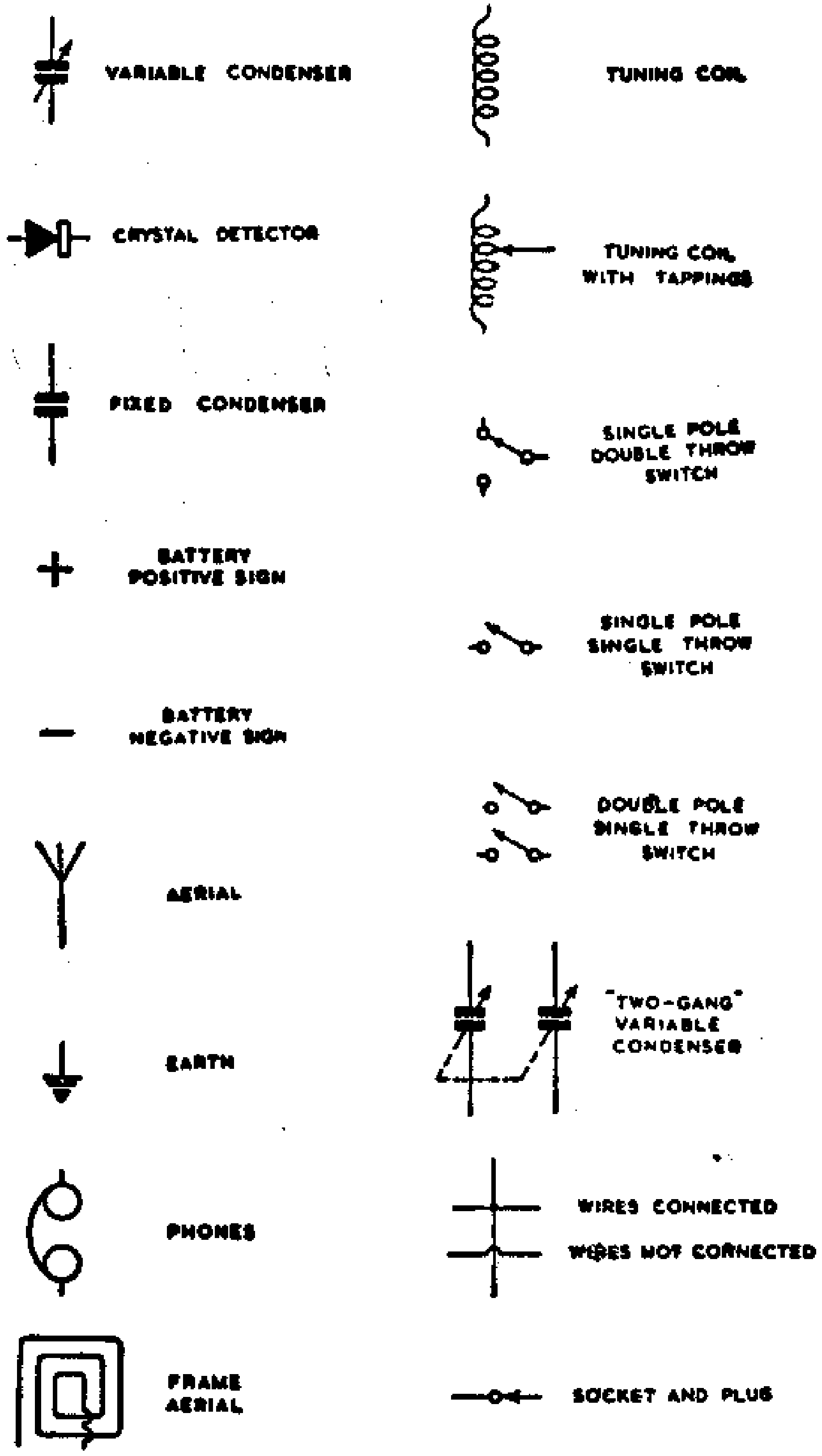
VARIABLE CONDENSER
CRYSTAL DETECTOR
FIXED CONDENSER
BATTERY POSITIVE SIGN
BATTERY NEGATIVE SIGN
AERIAL
EARTH
PHONES
FRAME AERIAL
TUNING COIL
TUNING COIL WITH TAPPINGS
SINGLE POLE DOUBLE THROW SWITCH
SINGLE POLE SINGLE THROW SWITCH
DOUBLE POLE SINGLE THROW SWITCH
"TWO-GANG" VARIABLE CONDENSER
WIRES CONNECTED
WIRES NOT CONNECTED
SOCKET AND PLUG

FIG. I SYMBOLS USED IN THE CIRCUITS DESCRIBED.

1. Aerials.

One of the essentials for satisfactory reception with crystal sets is an efficient aerial. Four different types are suggested and are shown in Fig. 2.

There are several important points to be noted when erecting an aerial for a crystal set.

(1) See that it is suspended as high as possible and is not screened by tall buildings, obstructions, etc.

(2) The aerial itself and the lead-in from the aerial must be adequately insulated. Be sure that the porcelain or plastic insulators do not permit the aerial or lead-in to touch the building, tree, or post from which they are suspended. The lead-in should be brought into the house through an insulated rod.

(3) It is exceedingly important to use insulated copper wire, either single or multi-stranded, for both the aerial and the lead-in.

2. Earths.

An adequate and efficient earth contributes to the good performance of any well-designed crystal set and Fig. 3 shows the best method of obtaining an efficient earth where there is a garden or patch of ground available at a short distance from the receiver.

Obtain a 3-ft. length of ½-in. galvanised iron pipe and at one end drill a hole right through. Insert a bolt with spring washer and locking nut in the hole. Connect the earth lead-in wire to the bolt; the other end of the lead-in to the earth connection on the receiver. Leave both ends of the pipe open. Insert the end opposite the bolt into the ground vertically until only 3-in. are showing above the surface of the surrounding earth.

It is important to see that the surrounding ground is kept damp, if necessary by occasionally pouring a pint or two of water down the open projecting end of the pipe.

Ready-made copper earthing rods with a pointed end for easy hammering into the earth are sometimes obtainable from radio dealers.

Should a garden not be available, an alternative earth is a cold water pipe coming from a rising main. Simply file very lightly the pipe at the back of the tap and wind the copper wire earth lead-in round the filed part and cover the joint with insulating tape.

In no circumstances should a gas pipe be used for an earth.

3. Crystal Detectors.

Most radio dealers can supply very efficient permanent or semi-permanent crystal detectors of proprietary makes. However, there are always available quite a number of ex-Government surplus units which are called crystal valves and, though these were originally designed for a very different use in service equipment, they are easily adaptable and will give extremely efficient results when used in circuits designed for them. To enable the constructor to purchase suitable types for this purpose—and they cost very little—the following list gives the Government reference numbers.

CV 101, CV 102, CV 103, CV 111, CV 112, CV 113, CV 226, CV 241, CV 246, CV 247, CV 253, CV 291, CV 351, CV 361, CV 364, CV 367, CV 727, CV 749, and CV 1785.

One of the satisfactory proprietary makes of this class of crystal detector is the BTH CS7A, which, incidentally, is the commercial equivalent of the CV 253.

For all these Government types, one connection should be made to each end by arranging a holder to fix them into or by carefully soldering the appropriate connecting wires to each end.

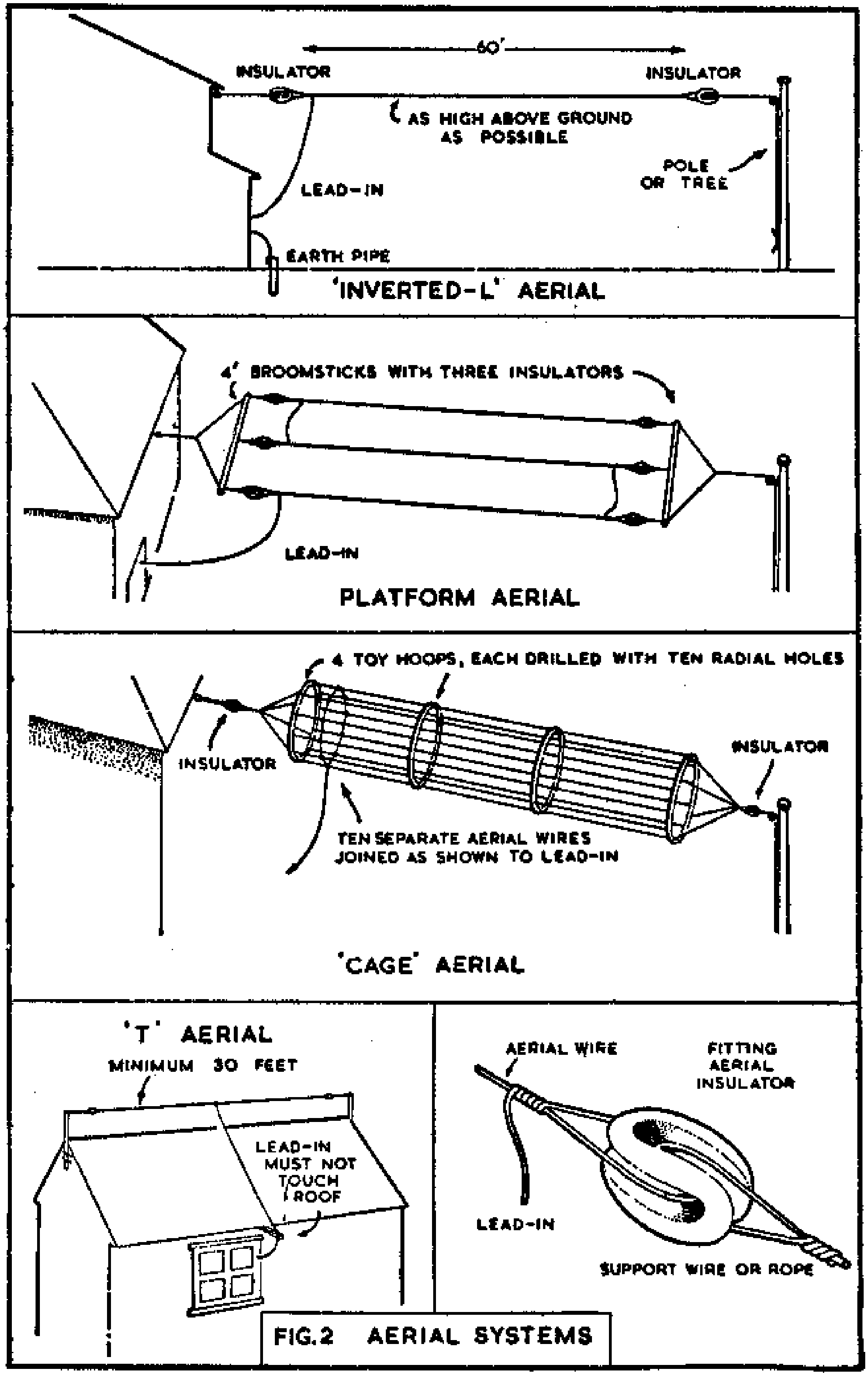
60'
INSULATOR
INSULATOR
AS HIGH ABOVE GROUND AS POSSIBLE
POLE OR TREE
LEAD-IN
EARTH PIPE
'INVERTED-L' AERIAL
4' BROOMSTICKS WITH THREE INSULATORS
LEAD-IN
PLATFORM AERIAL
4 TOY HOOPS, EACH DRILLED WITH TEN RADIAL HOLES
INSULATOR
INSULATOR
TEN SEPARATE AERIAL WIRES JOINED AS SHOWN TO LEAD-IN
'CAGE' AERIAL
'T' AERIAL
MINIMUM 30 FEET
LEAD-IN MUST NOT TOUCH ROOF
AERIAL WIRE
FITTING AERIAL INSULATOR
LEAD-IN
SUPPORT WIRE OR ROPE

FIG.2 AERIAL SYSTEMS

4. Coils.

In Fig. 3 two methods are shown of fixing the coil to the chassis. However, the ingenious constructor may evolve another arrangement.

The important thing to remember is that the coil should be fixed so that it is completely insulated from the metal chassis that may be used.

It is also important to remember that if a metal chassis is used, one must see that all components and wiring, with the exception of earth connections, are adequately insulated from this chassis so as to make sure that no short circuits take place. It is recommended that all crystal sets should be built on a panel, and that both panel and chassis should be of bakelite, paxolin or perspex, as these three materials are efficient insulators in themselves.

5. The Beginners' Crystal Receiver.

The popularity of the crystal receiver among beginners is undoubtedly due to the very modest outlay necessary, and the ease with which such receivers may be built. It is not always appreciated that, to obtain worth while results,

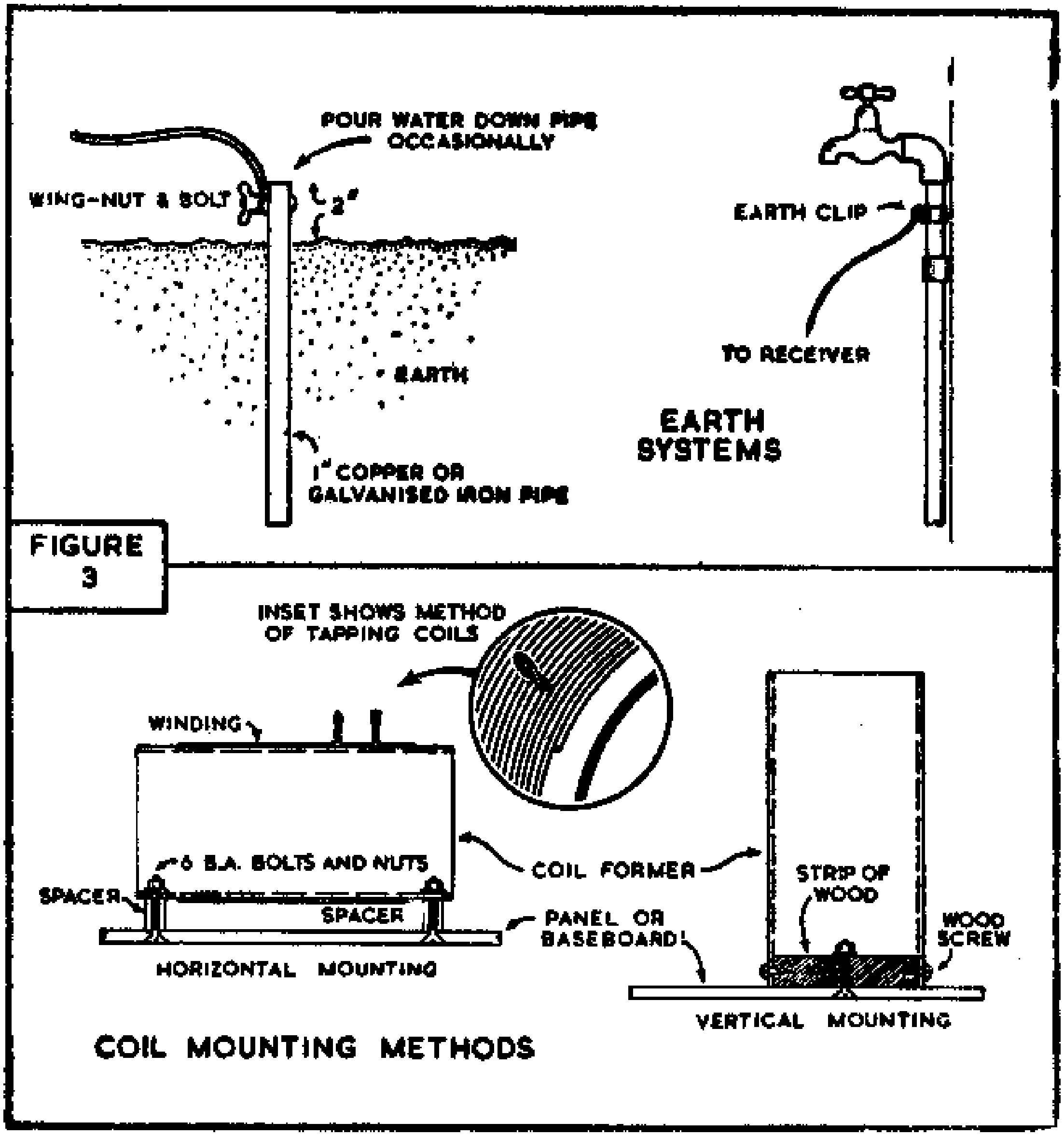
POUR WATER DOWN PIPE OCCASIONALLY
WING-NUT & BOLT
2"
EARTH
1" COPPER OR GALVANISED IRON PIPE
EARTH CLIP
TO RECEIVER
EARTH SYSTEMS
INSET SHOWS METHOD OF TAPPING COILS
WINDING
6 B.A. BOLTS AND NUTS
SPACER
SPACER
COIL FORMER
PANEL OR BASEBOARD
STRIP OF WOOD
WOOD SCREW
HORIZONTAL MOUNTING
VERTICAL MOUNTING
COIL MOUNTING METHODS

FIGURE 3

far more care in avoiding losses is necessary than with a three or four valve receiver. One reason for this is the fact that the crystal receiver can offer no help by way of amplification and it is therefore dependent on an efficient aerial/earth system, low loss design within the receiver itself, and the use of sensitive headphones. The experience gained is well worth the time expended since, when more ambitious receivers are attempted, the constructor will remember the benefits gained by efficient construction and installation with the result that each receiver attempted will be an instrument capable of giving a first class performance.

Losses within the receiver itself may be considerably reduced by efficient coil design. Unfortunately, selectivity is a factor which must not be overlooked; it is this factor which governs the sharpness of the tuning and permits the separation of adjacent powerful stations. The greater the selectivity, the greater the ease with which signals may be separated; but at maximum selectivity the received signals are at minimum strength. In Fig. 4, the coil is tapped at four points, which enables the best point between selectivity and sensitivity to be chosen. The nearer the aerial tap to the earth end of the coil, the greater the selectivity. The pre-set condenser C1 also aids selectivity; at minimum capacity the selectivity is high, and *vice versa*.

Used in conjunction with the coil taps, the constructor should have no difficulty in selecting a condenser setting and tapping point which permit the reception of strong signals and at the same time effect the required separation of powerful stations.

The coil is constructed on a former 3-in. diameter × 2-in., and consists of 50 turns 36 S.W.G. enamelled wire close wound and tapped at every tenth turn. These taps are brought out in the form of loops twisted firmly so that there is no actual break in the winding continuity. After completion the winding may be painted with Durafix to prevent any movement, and the taps and two free ends cleaned with emery cloth to remove the enamel. The layout and wiring is shown in Fig. 4. The headphones should have a resistance of some 4,000 ohms or more and should be of reputable manufacture. Any of the aerials previously described are suitable for use with this receiver.

C2 .0005 variable condenser.
C1 .0002 pre-set condenser.
C3 .001 mica condenser.
1 Crystal detector, semi-permanent pattern.
2 oz. 30 S.W.G. enamelled copper wire.
1 Coil former 3-in. dia. × 2-in.
1 Ebonite or Bakelite panel 6-in. × 5-in. × ¼-in.
4 terminals, connecting wire, etc.

6. Frame Aerial Receiver for Local Station use on Medium wave bands.

This receiver was specially designed for constructors who are within 5 to 10 miles of a powerful station and who have no facilities for constructing a good outdoor aerial.

The frame itself consists of two ¾-in. × ¾-in. square section wooden rods each 36-in. in length. These are placed together at their centre line to form a cross; they are fastened by a nut and bolt and these in turn pass through an upright piece of wood ½-in. × 1-in. section 2-ft. in length. This upright support is then screwed into a baseboard 8-in. × 8-in. × 1-in. thick. At each of the arms of the 36-in. cross a 4-in. strip of ½-in. square section wood is placed. Reference to Fig. 5B will give the details of the construction of this aerial quite clearly.

It will be noted that approximately 85-ft. of 20 S.W.G. double cotton-covered wire are required for the inductance to cover the medium wave broadcast band. The wire should be wound around the four crossbars, thereby creating

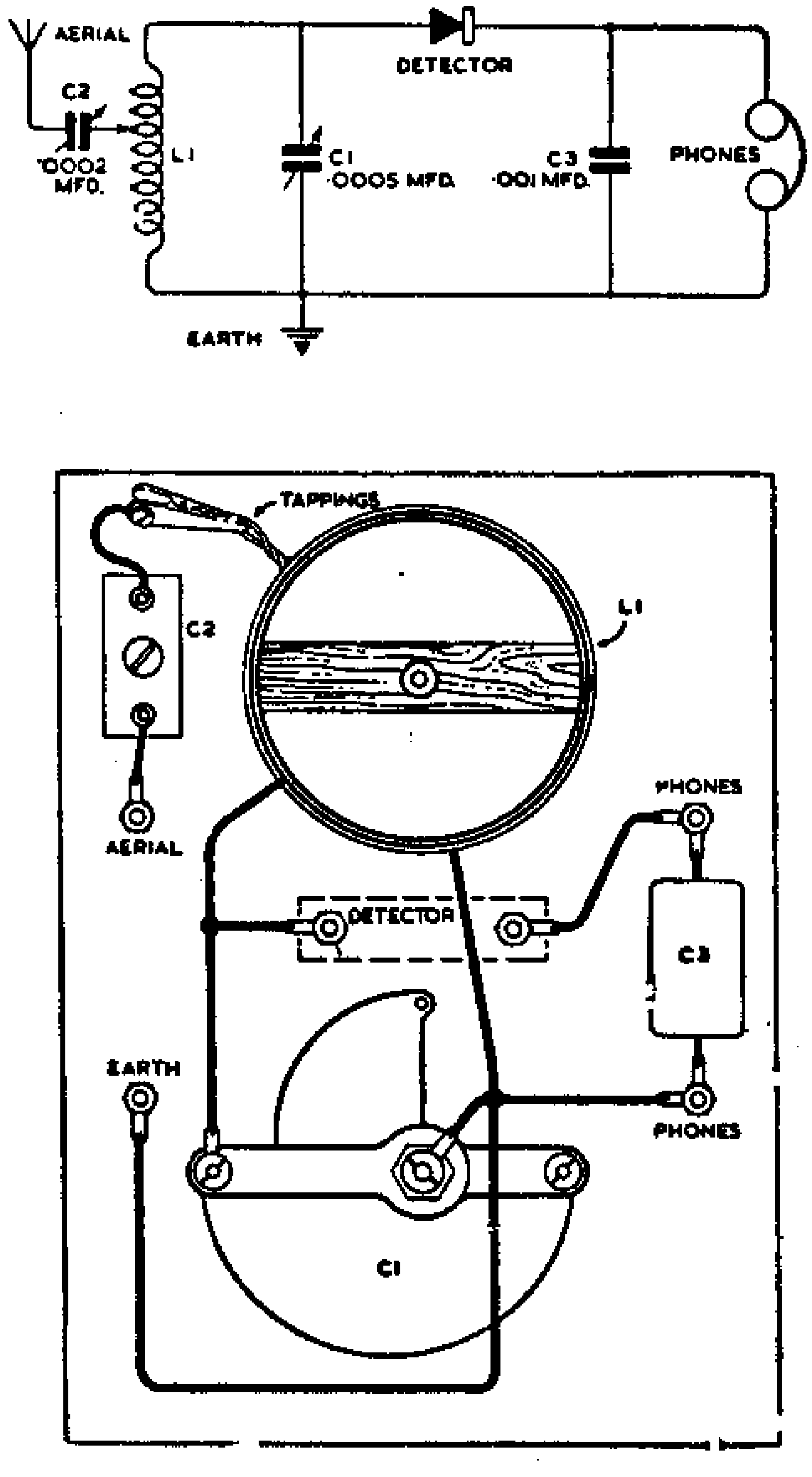
AERIAL
C2
·0002 MFD.
L1
DETECTOR
C1 ·0005 MFD.
C3 ·001 MFD.
PHONES
EARTH
TAPPINGS
C2
L1
AERIAL
PHONES
DETECTOR
C3
EARTH
PHONES
C1

FIG. 4 THE BEGINNER'S CRYSTAL RECEIVER.

a square wire frame, and each turn of the wire should be spaced approximately ⅛-in. The best way to effect this spacing accurately is to file on the 4-in. crossbars a number of small nicks at ⅛-in. intervals. The two ends of the wire should then be brought down on the two terminals which may be screwed into the square baseboard and two leads taken from there to the crystal set itself.

The actual set may be constructed on a small ebonite, paxolin or perspex panel 8-in. × 5-in. which may be screwed to one side of this base supporting piece of wood, thereby making the set virtually self-contained except for the earphones. Fig. 5A shows the circuit of the receiver.

To obtain the best results with this receiver, the aerial should be rotated by hand until the signals become loudest.

1 Semi-permanent crystal detector.
C1 1 .0005 mfd. variable condenser.
C2 1 .0005 mfd. fixed condenser.
1 condenser dial marked 0 to 100 or 0 to 180.
1 pair of high sensitivity earphones.
85-ft. to 90-ft. of 20 S.W.G. double cotton covered wire. (8oz.)
Sufficient connecting wire for wiring purposes.

7. Long Distance Receiver.

This receiver has been specially designed to achieve high sensitivity. With a sufficiently good outdoor aerial effective reception has been obtained up to 150 miles.

The circuit is quite straightforward, as can be seen from Fig. 6 and the only special point is in winding the coil L1.

This coil consists of a 2½-in. diameter former on which are wound a total of 51 turns using 24 S.W.G. double cotton wire. Starting at one end, tap every two turns until there are eight taps; wind a further 15 turns and tap; then a further 10 turns and tap; finally, add 10 turns to end the winding. The wire should be close wound, *i.e.*, with turns touching.

To obtain the best results, the two crocodile clips A and B should be tried on the various coil taps until the loudest reception is obtained. It may be found in use that the loudest reception of any particular station causes another station to be heard at the same time; care should, therefore, be taken to select the appropriate tapping point to separate any two powerful stations that may interfere with one another.

4 oz. 24 S.W.G. DCC copper wire.
Coil former 2½-in. dia. × 3-in.
C1-2 single variable condensers .00035 mfd.
2 Condenser dials marked from 0 to 100 or 0 to 180.
1 Permanent crystal detector.
C3 .001 fixed condenser.
1 pair of high sensitivity earphones (preferably between 4,000 ohms and 8,000 ohms impedance).
2 Crocodile clips.

8. High Gain Receiver.

This receiver has been designed to give good reception up to 50 miles distance from any normal power broadcasting station in the medium band. It is a very selective circuit and will therefore enable the user to separate closely situated stations.

It will be noticed in the circuit (Fig. 7) that the variable condenser C1, has an on/off switch S1 to throw it out of circuit if necessary. It is used to increase the selectivity of the receiver and, in operating, it should be tried with this condenser either in or out of circuit on the various tappings on the coil L1.

L1 consists of 90 turns of 22 S.W.G. double cotton covered wire wound on a 2-in. diameter former and tapped every ten turns. L2 separately wound on a 2-in. former, consists of 90 turns of 22 S.W.G. double cotton covered wire, also tapped as L1. The components required are given on page [illegible]

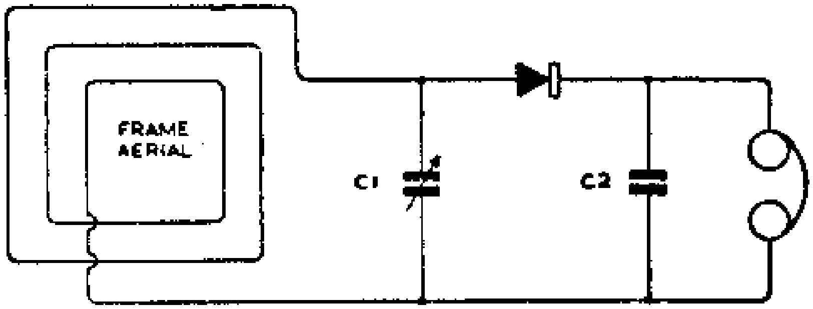
FRAME AERIAL
C1
C2

FIG. 5A FRAME AERIAL CRYSTAL RECEIVER

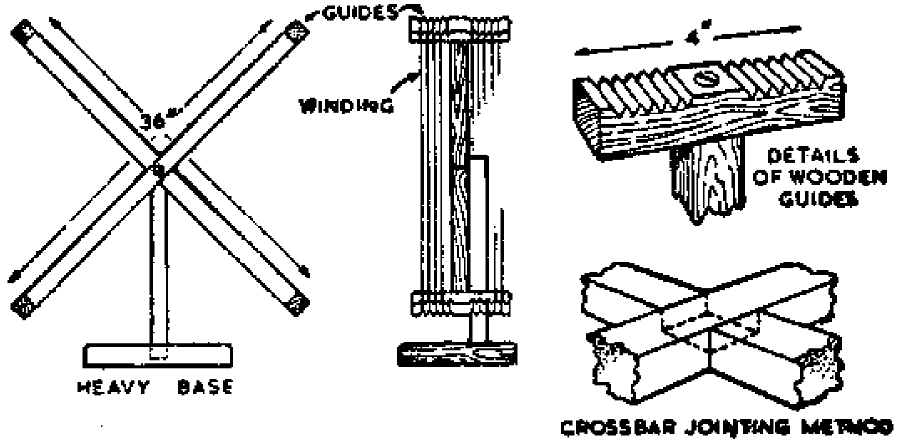
GUIDES
36"
WINDING
HEAVY BASE
4"
DETAILS OF WOODEN GUIDES
CROSSBAR JOINTING METHOD

FIG. 5B. DETAILS OF FRAME AERIAL

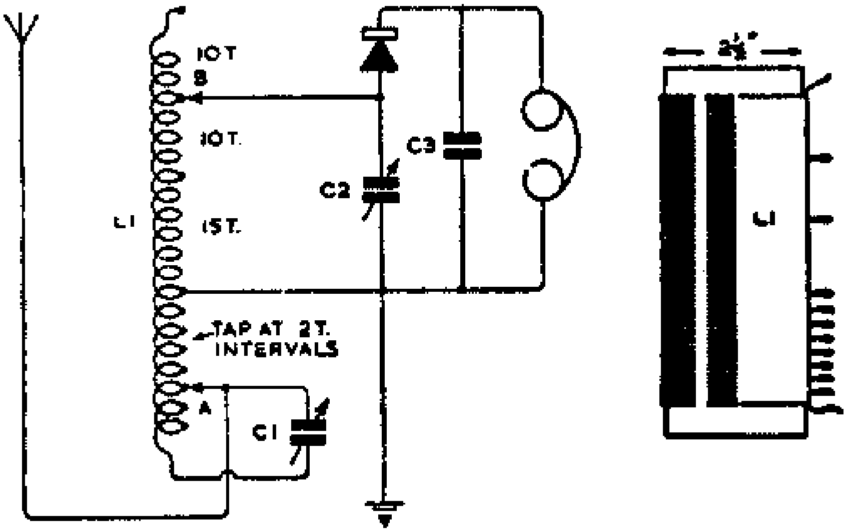
10T
B
10T.
C3
C2
L1
15T.
TAP AT 2T. INTERVALS
A
C1
2½"
L1

FIG. 6 LONG-DISTANCE CRYSTAL RECEIVER

C1-2 Single variable condensers .00035 mfd.
S1 Single pole single throw switch.
C5 .0001 preset condenser.
1 Permanent crystal detector.
C3-4 .001 mfd. fixed condenser.
1 Pair of high sensitivity earphones.
8 oz. 22 S.W.G. DCC.
2 Dials for variable condenser marked 0 to 100 or 0 to 180.

9. High Selectivity Receiver.

This circuit (see Fig. 8) is for use in high saturation strength areas where it is desired to separate powerful stations. The circuit is not complicated and is extremely effective.

It will be noted that C1 and C2, each .00035 mfd. capacity, are 2 gang unit.

The only special point in this set is the design of coils L1, 2 and 3, which are constructed as follows: L1, 60 turns of 22 S.W.G. double cotton covered wire on a 3 × 4 inches former tapped every five turns; L2, 15 turns of 22 S.W.G. double cotton covered wire on a 3 × 6 former; and L3, 85 turns of 22 S.W.G. double cotton covered wire wound on the same former as L2, the two windings being separated by approximately ¼-in. (see Fig. 8).

Note that the aerial lead is brought down to one of the tappings on wire L1. This should be tried on each of the tappings and the one giving the required selectivity should be used.

C1-2 2-gang variable condenser .00035 mfd.
2 3-in. formers for coils L1-2-3.
8 oz. 22 S.W.G. DCC.
C3 .0005 fixed condenser.
1 Semi-permanent crystal detector.

10. Battery-aided Crystal Receiver.

Fig. 9 shows a design which utilises a 9-volt grid bias battery to improve performance.

The charge built up in the .02 mfd. fixed condenser is changed to a positive current when it passes through the coil L1. This, in turn, is superimposed on the positive charge that is commencing to be rectified by the crystal detector and thereby increases the signal strength.

Coil L1 consists of a 4-in. diameter former wound with 85 turns of 24 S.W.G. double cotton covered wire tapped at 5, 25, 45 and 65 turns from the earth end. The earth end and these four taps are brought out to five small sockets and the connection from the earth is taken to a small socket. The plug should be tried in each of the five positions to find which one gives the greatest signal strength.

The 9-volt battery is connected across the .02 mfd. condenser and special care should be taken to see that the battery is connected correctly, *i.e.*, negative pole to earth.

1 Permanent crystal detector.
1 4-in. dia. × 3-in. former.
8 oz. 20 S.W.G. DCC copper wire.
C1 Variable condenser .0005 mfd.
C2 Fixed condenser .005 mfd.
C3 Fixed condenser .02 mfd.
1 pair of high sensitivity headphones.
1 Clix plug.
5 Clix sockets.
1 9-volt grid bias battery.

11. Medium and Long wave band Receiver.

This set has been designed for reception on two wave bands.

Care must be taken in winding the coils L1, 2, 3 and 4. These are wound on a single 3-in. diameter former. L1 consists of 30 turns of 28 S.W.G. DCC tapped at 8, 16 and 24 turns. L2, 60 turns of 32 S.W.G. DCC. L3, 40 turns of 28 S.W.G. DCC, and L4, 80 turns of 32 S.W.G. DCC. All are close wound and are separated from each other as follows: between L1 and L2, 3/16-in.;

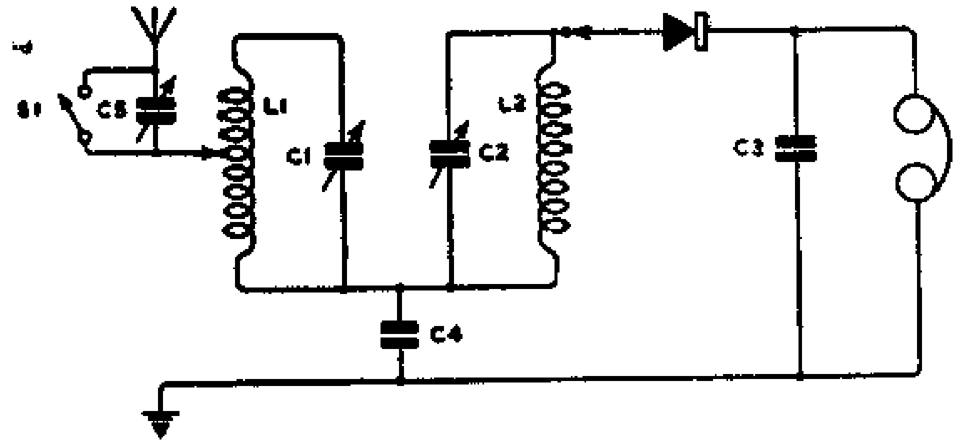
S1
C5
L1
C1
C2
L2
C3
C4

FIG. 7 HIGH-GAIN CRYSTAL RECEIVER

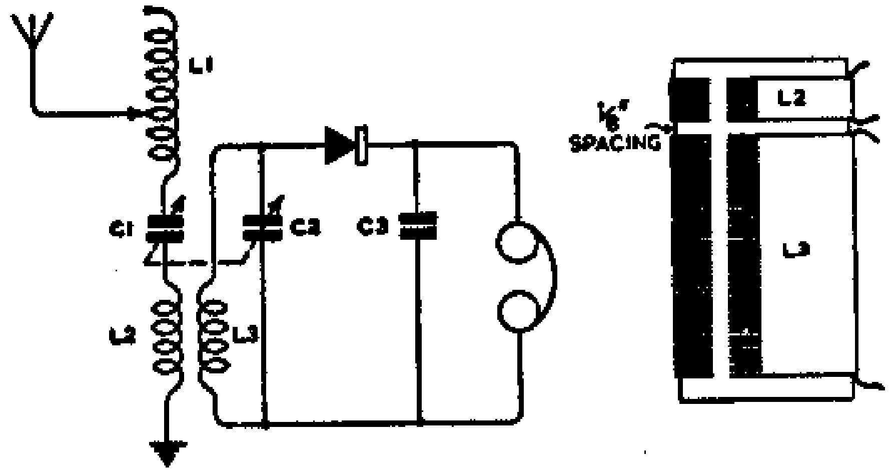
L1
C1
C2
C3
L2
L3
1/8" SPACING
L2
L3

FIG. 8 HIGH SELECTIVITY CRYSTAL RECEIVER

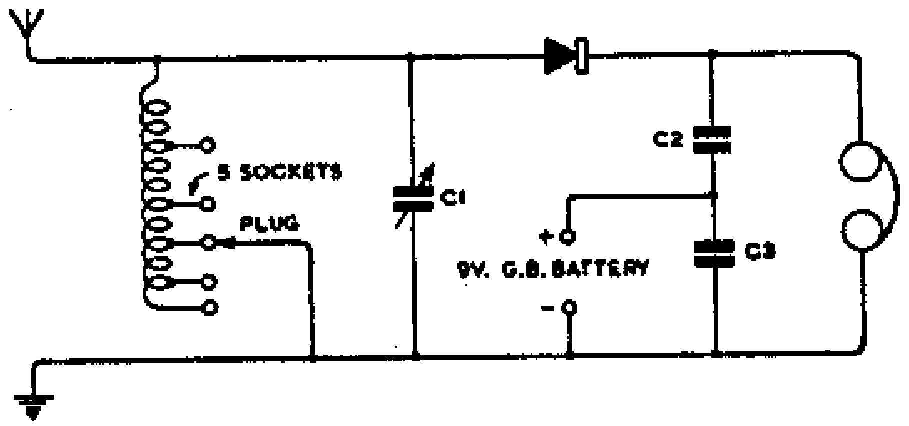
5 SOCKETS
PLUG
C1
C2
C3
+
9V. G.B. BATTERY
-

FIG. 9 BATTERY-AIDED CRYSTAL RECEIVER

between L2 and L3, 1-in.; and between L3 and L4, 3/16-in. See Fig. 10.

C1-2 single variable condensers .0005 mfd.
2 Variable condenser dials marked 0 to 100 or 0 to 180.
1 Clix plug.
3 Clix sockets.
S1 Single pole double throw switch.
C3 .001 mfd. fixed condenser.
1 Pair of high sensitivity earphones.
1 3-in. dia. × 8-in. former.
2oz of both 28 and 32 S.W.G. double cotton covered wire to make these 4 coils.
1 Permanent crystal detector.

In use, the plug should be tried in each of the three sockets to find the point which gives the best results. The switch is used as a wave band change switch.

12. Ultra Sensitive All Wave Receiver.

The receiver shown in Fig. 11 is for long distance reception. It is a novel circuit and, though apparently complicated in construction, it is really very simple to assemble with the minimum of tools and labour.

It will be seen from the circuit diagram that there are two banks of sockets, five black and six red. When using this set the appropriate black and red plugs should be inserted in whichever of the similar coloured sockets give the best results. The coil is quite easy to construct and consists of a primary winding which is inserted at the earth end of the secondary winding. For the medium wave band, the secondary winding L2 consists of 54 turns of 20 S.W.G. enamelled or cotton covered wire. This coil has four taps in addition to the starting point and the end connection, and these taps are taken at six turns, 14 turns, 27 turns and 40 turns from the earth end. The taps are taken out and a connection is made to an appropriate socket, as shown in the diagram.

The coil is wound on a 3-in. diameter former which may be a paper tube, bakelite, paxolin or perspex, and spacing is approximately 17 turns to the inch.

For short wave work down to about 30 metres, L2, the secondary coil, will consist of the same diameter former wound with 15 turns of the same gauge wire, the four taps being taken at 3, 6, 9 and 12 turns from the earth end.

L1, the primary coil, is used for both wave bands and consists of 11 turns of similar gauge wire to that used for L2, wound on a 2-in. × 1-in. former, the winding to be spaced the same as L2. The coil is then inserted at the earth end of whichever secondary coil L2 is being used.

The crystal detector may be one of the semi-permanent types freely available to-day; the old-fashioned cat's whisker pattern; or one of the silicon permanent detectors which are sold by many surplus dealers.

The 'phones should be of a standard type 4,000 ohms impedance although 'phones of other values will probably operate quite satisfactorily with this powerful receiver. It should, however, be stressed that the better the headhones used, the more satisfactory the reception.

Coils for other wave bands may be experimentally designed by the constructor, remembering that four tappings are advisable for the secondary coil L2.

A long aerial is very desirable with this receiver, placed as high as possible. It is recommended that a minimum length of 70-ft. be allowed, to include the aerial and lead-in.

The values of the components are shown in the following components list. The constructor should experience no difficulty in obtaining excellent results from this cleverly designed receiver.

C1 .0005 variable condenser.
C2 .002 fixed condenser.
1 semi-permanent crystal detector.
1 coil former 3-in. dia. × 4½-in.
1 coil former 2-in. dia. × 1-in.
½ lb. 20 S.W.G. enamelled or DCC copper wire.
6 red wander plug sockets.
6 black wander plug sockets.
1 red and 1 black wander plug.

Headphones, terminals, connecting wire

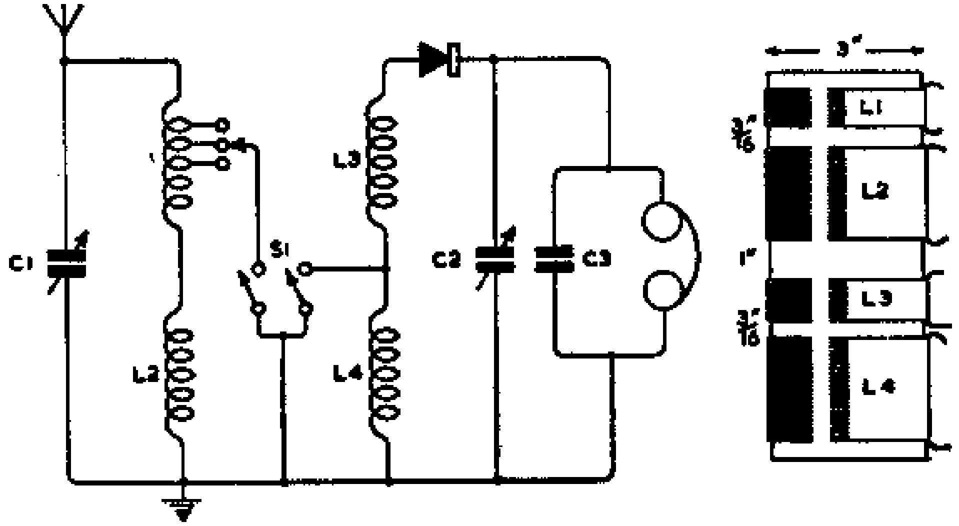
C1
L2
S1
L3
L4
C2
C3
3"
L1
L2
3/16"
1"
L3
3/16"
L4

FIG. 10 MEDIUM AND LONG-WAVE CRYSTAL RECEIVER

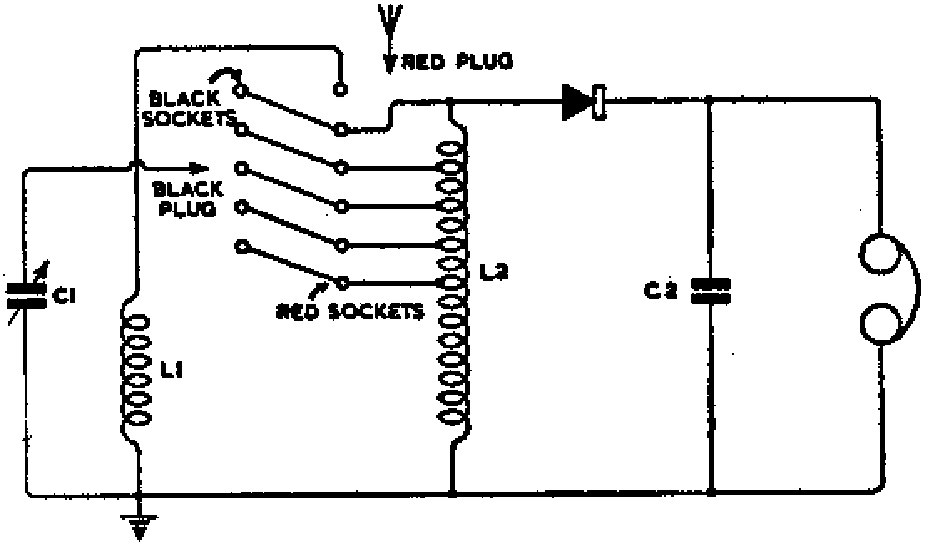
RED PLUG
BLACK SOCKETS
BLACK PLUG
C1
L1
RED SOCKETS
L2
C2

FIG. 11 ULTRA-SENSITIVE ALL-WAVE CRYSTAL RECEIVER

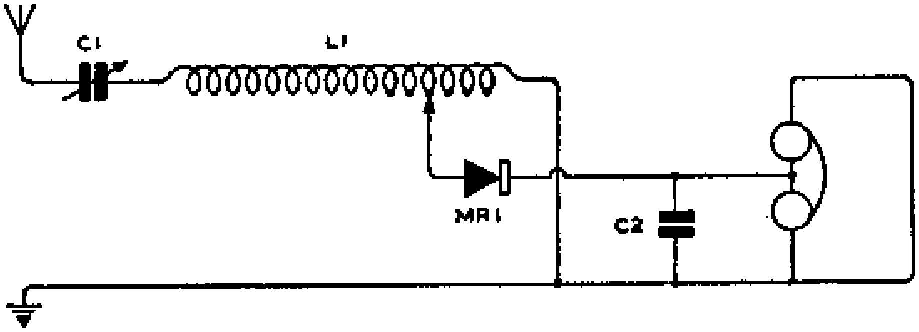

FIG. 12 ADVANCED CRYSTAL DIODE RECEIVER

13. An Advanced Crystal Diode Receiver.

The modern crystal diode of the radar type does not require any adjustment and thus has a distinct advantage over the ordinary crystal detector which needs delicate adjustment—even the semi-permanent type.

Such crystals are now obtainable on both the regular and the surplus markets, and a popular type is the B.T.H. CS7A silicon crystal (also coded as CV 253).

A rather different circuit from the usual crystal set arrangement is needed to suit the characteristics of a crystal diode. The diode must be tapped on to the tuning coil, and it is found that the tuning circuit itself gives best results if a series-tuned acceptor circuit is used. The circuit of a radar crystal receiver is shown in Fig. 12. The type of aerial used with this circuit has a very great bearing on the behaviour of the receiver for, in effect, the tuned circuit, with a series resistance equivalent to the reflected crystal load resistance, is in series with the capacitance of the aerial to earth and the aerial's effective series resistance.

At resonance—when the combination is tuned to any particular signal—the inductance resonates with the capacitances of the tuning condenser and the aerial in series, and the final effect is that the reflected load of the crystal, in series with the coil's R.F. resistance, is paralleled across the aerial's series resistance.

For maximum power transfer, the crystal load resistance must be made equal to the sum of the aerial series resistance and the coil R.F. resistance, and so the method in which the crystal is tapped into the tuning coil, and the exact capacitance required to tune the receiver to any required signal, must depend to a very great extent on the aerial itself.

At the same time the crystal resistance varies with the signal strength, the resistance being high for weak signals and dropping by as much as 50 per cent. and more for strong signals, so that this effect also has a bearing on the correct coil tap. The output impedance of the crystal also varies similarly, affecting the matching of the headphones into the crystal diode, and so for any set of conditions, the receiver requires to be matched up to both the aerial and the signal being received for best results. This would mean a series of coil taps and (theoretically, not practically) a matching transformer between the diode and the headphones; but in practice it will be found that a receiver using standard parts can be built up to give very good results under various conditions.

It has been shown already that the tuning of the receiver depends to a great extent on the characteristics of the aerial, and while a .0005 mfd. variable capacitor is shown in Fig. 12 as the tuner, the constructor must be prepared to experiment with different capacitance values until the required station is tuned in. The range of reception given by the receiver is quite good, if a really long and high aerial, and a good earth connection, are used; but no more than the local station signal can be expected, and the tuning therefore must be adjusted to suit the station frequency.

It must be mentioned that the headphones are shown parallel connected. High resistance headphones of the 4,000 ohm type must be used, and if two of these are connected in parallel rather than in the more usual series method, they will provide a roughly accurate match to the crystal. If more than one pair of headphones are to be connected in, then the pairs of headphones may be left series connected in the usual manner, the sets of headphones being connected in parallel.

The capacitor across the headphone terminals completes the crystal R.F. circuit, and any value between about 0.001 and 0.005 mfd. will serve. The higher capacitance will, of course, by-pass some of the higher audio frequencies, so that if good steady reception is obtained some experiment with this condenser is also worth while.

C1, .0005 mfd. variable condenser.
C2, .005 mfd. fixed condenser.
4oz. 26 S.W.G. enamelled copper wire.
M.R.1 CV 253 Crystal diode.
Coil former 1½-in. dia. × 4½-in.

Coil details are as follows: 150 turns of 26 S.W.G. tapped at 20, 25, 30, 35, 40 and 45 turns from the earth end. At each tapping-point the wire should be twisted up into a loop and the winding then continued without breaking the wire; when the coil is completed and the ends anchored the tapping loops can be bared.

BERNARDS RADIO BOOKS

No.		Price
56.	Radio Aerial Handbook	2/6
57.	Ultra-Shortwave Handbook	2/6
58.	Radio Hints Manual	2/6
64.	Sound Equipment Manual	2/6
68.	Frequency Modulation Receivers' Manual	2/6
73.	Radio Test Equipment Manual	2/6
83.	Radio Instruments and their Construction	2/6
96.	Crystal Set Construction	1/-
99.	One Valve Receivers	1/6
100.	A Comprehensive Radio Valve Guide, Book 1	5/-
103.	"Radiofolder" A. The Master Colour Code Index for Radio and T.V.	1/6
104.	Three Valve Receivers	1/6
107.	Four Valve Circuits	1/6
108.	Five Valve Circuits	2/6
121.	A Comprehensive Radio Valve Guide, Book 2	5/-
123.	"Radiofolder" F. The Beginners' Push-Pull Amplifier	1/6
126.	Boys' Book of Crystal Sets and Simple Circuits	2/6
129.	Universal Gram Motor Speed Indicator	1/-
134.	F.M. Tuner Construction	2/6
135.	All Dry Battery Portable Construction	2/6
138.	How to Make F.M. and T.V. Aerials, Bands 1, 2 and 3	2/6
141.	Radio Servicing for Amateurs	3/6
143.	A Comprehensive Radio Valve Guide, Book 3	5/-
145.	Handbook of AM/FM Circuits and Components	2/-
146.	High Fidelity Loudspeaker Enclosures	5/-
147.	Practical Tape Recording Handbook	5/-
148.	Practical Transistor Receivers, Book 1	5/-
149.	Practical Stereo Handbook	3/6
150.	Practical Radio Inside Out	3/6
151.	Transistor Superhet Receivers	7/6
155.	Portable Transistor Radio and Radiogram	2/6
156.	Transistor Circuits Manual, No. 1	2/6
157.	A Comprehensive Radio Valve Guide, Book 4	5/-
158.	Radio, Television, Industrial Tubes, Semiconductor and Diodes Equivalents Handbook (208 pages)	9/6
159.	Realistic High Fidelity	5/-
160.	Coil Design and Construction Manual	5/-
161.	Radio, Television and Electronics Data Manual	3/6
162.	High Fidelity Stereo Gramophone	5/-
163.	Transistor Circuits Manual, No. 2	2/6
164.	High Fidelity Tape Recorder for the Home Constructor	2/6
165.	Hartley on Hi-Fi, Book 1, Radio Tuners	5/-
166.	Public Address Systems	2/6
167.	Transistor Circuits Manual, No. 3	2/6
168.	Transistor Circuits Manual, No. 4	2/6
169.	Hi-Fi Transistor F.M. Tuner for the Home Constructor	3/6
170.	Transistor Circuits for Radio Controlled Models	7/6
171.	Super Sensitive Transistorised Pocket Set Construction	3/6
172.	International Radio Stations List	2/6
173.	Practical Transistor Audio Amplifiers	3/6
174.	Transistor Subminiature Receivers Handbook	5/-
175.	Transistorised Test Equipment and Servicing Manual	3/6
176.	Transistor Audio Amplifiers Manual	6/-
177.	Modern Transistor Circuits for Beginners	7/6
178.	Comprehensive Radio Valve Guide, Book 5	6/-
179.	Transistor Circuits Manual No. 5	5/-
180.	British Semiconductor Survey	15/-
181.	22 Tested Circuits Using Micro Alloy Transistors	5/-
182.	"At A Glance" Radio Valve & T.V. Tube Equivalents	3/6
183.	How to Receive Foreign T.V. Programmes on your Television Set by simple Modifications	5/-
	Resistor Colour Code Disc Calculator	1/6
	Engineers' Reference Tables	1/6
	International Radio Tube Encyclopaedia—3rd Edition	63/-

A Simple Crystal Radio Set

Whether you are a beginner or a seasoned expert, you would probably agree that there is nothing quite like using this simple design to make a crystal radio and enjoying it, listening to stations both near and far. Add to this the "something for nothing" advantage of not having to power this set with batteries or AC line and you can see why this simple design will be used and enjoyed for a long time to come. Almost like magic... Let' build one!

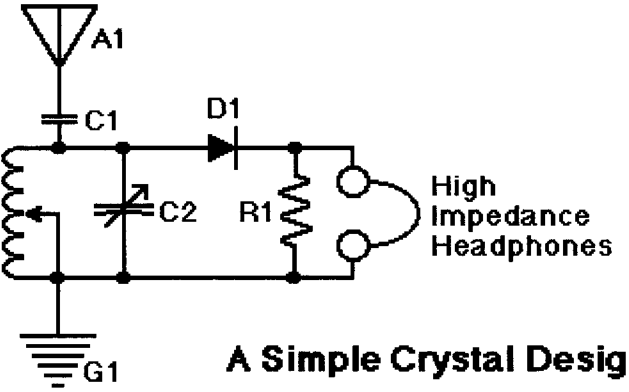
A1
C1
D1
L1
C2
R1
High Impedance Headphones
G1

A Simple Crystal Design

Components:

- **A1- 50-100 foot longwire antenna**
- **C1 - 50 pf ceramic capacitor OSE p/n CD-5**
- **C2 - 365 pf variable capacitor OSE p/n BC14400**
- **D1 - 1N34A diode OSE p/n 1N34A**
- **G1 - Ground (see instructions below)**
- **L1 - 100 turns of #20 AWG enamel coated copper wire OSE p/n MW20, adjustable tap (see instrctions below)**
- **R1 - a 100k ohm resistor OSE p/n RQ100K**
- **Crystal Earphones OSE p/n CE748**
- **Fahnestock clips OSE p/n 2090**

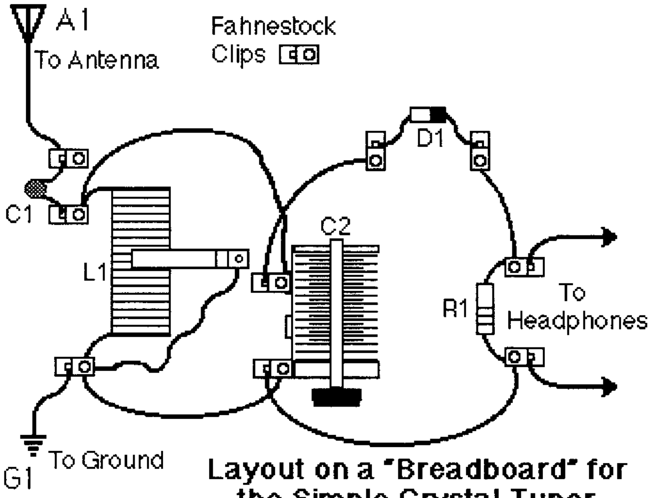
A1
To Antenna
Fahnestock Clips
C1
L1
D1
C2
R1
To Headphones
G1
To Ground

Layout on a "Breadboard" for the Simple Crystal Tuner

FREE-POWER AM RECEIVER (P,E)

This project uses a strange way to get its power. It uses power "extracted" from the tuned station to give to the amplifier transistor. With a reasonably strong applied modulation envelope, this type of detector produces a strong demodulated output.

Naturally, the closer you are to a strong station, the more current your radio will be able to supply.

For best results, you must do everything possible to deliver a strong signal to the transistor detector. We recommend a good antenna and ground, the latter preferably being made to a water pipe or solid external ground composed of a pipe driven at least 5 feet into moist earth. This is important for ensuring maximum signal pickup.

If you have plenty of space available, the longer the antenna (up to about 100 feet), the better the results.

Figure 1 shows the schematic diagram of the Free-Power AM receiver. Observe that the heart of the circuit is a germanium transistor that works as a detector and audio amplifier.

Components placement on a terminal strip used as a chassis is shown in ***Figure 2***. The terminal strip can be fixed on an experiment board that can be constructed using some common tools and materials.

Q1 is any germanium transistor, such as GE-2 or 2N107. You can find germanium transistors in old nonworking AM transistor radio receivers. Nonworking AM radios can also supply the variable capacitor and the loopstick (ferrite rod).

Earphone must be a high-impedance crystal type. Low-impedance earphones don't function in this circuit.

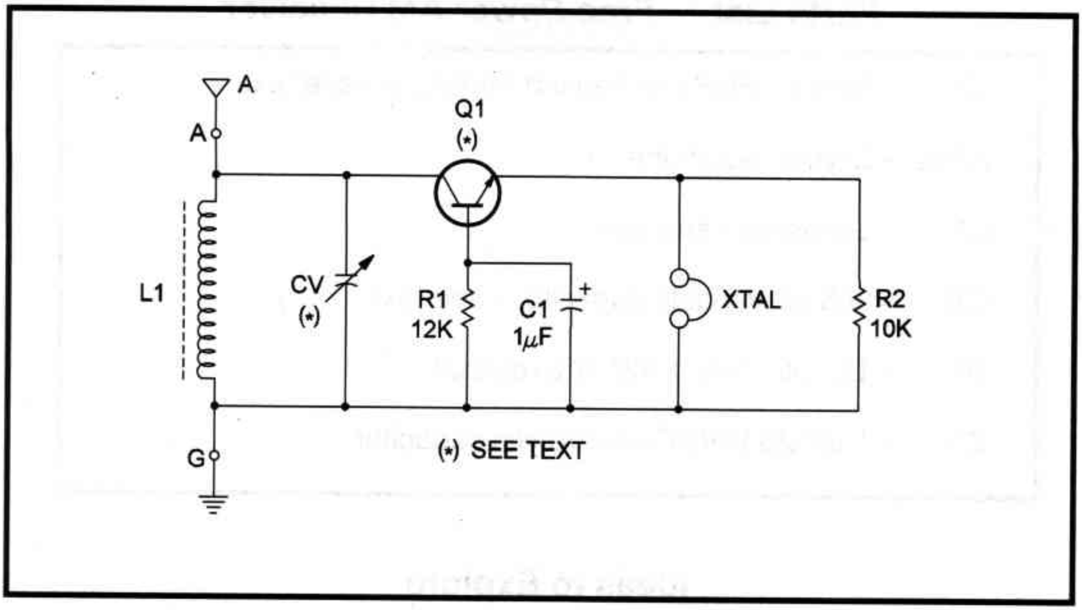
A
A
G
Q1
(*)
L1
CV
(*)
R1
12K
C1
1μF
+
XTAL
R2
10K
(*) SEE TEXT

Figure 1

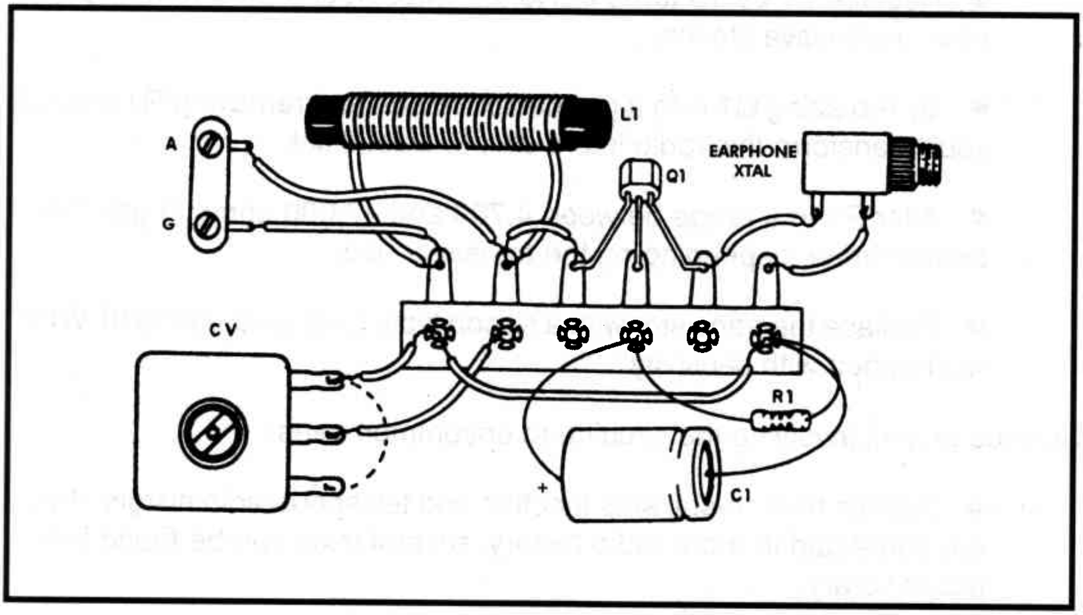
A
G
L1
EARPHONE
XTAL
Q1
CV
R1
+
C1

Figure 2

Parts List — Free-Power AM Receiver

Q1	- NPN or PNP germanium transistor - see text
XTAL	- Crystal earphone
L1	- Loopstick - see text
CV	- 365 pF variable capacitor - see text
R1	- 12,000 ohm, 1/4W, 5% resistor
C1	- 1 uF, 25 WVDC electrolytic capacitor

Ideas to Explore

To learn more about the circuit:

- Replace L1 by 15 turns of No. 28 wire, wound on a ferrite rod, to tune short-wave stations.
- By replacing Q1 with a germanium diode and removing R1 and C1 you'll transform this radio into a simple crystal set.
- Alter R1 in a range between 4,700 and 47,000 ohms to get better performance, depending on the transistor used.
- Replace the transistor with a silicon type, such as the BC548. What will happen with sensivity?

Science project involving the circuit and uncommon uses:

- Explain how crystal sets function and tell about radio history. If you are interested in more radio history, several texts can be found in the public library.
- If you want to learn more about radio receivers, try to find information about heterodyne and super-hereterodyne types.

Electronic Projects for Beginners

F. G. RAYER, T.Eng. (CEI), Assoc. IERE

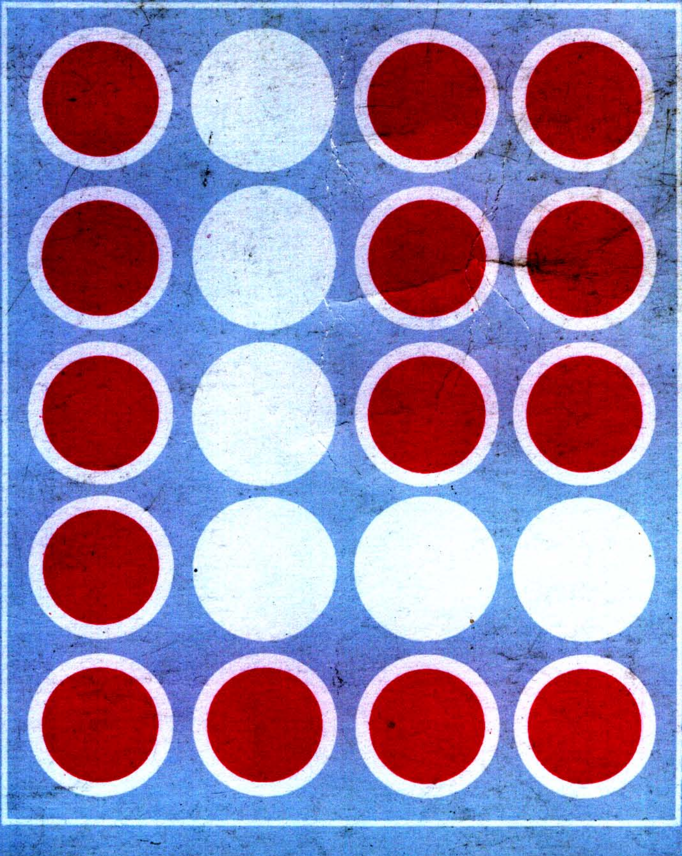

speaker SP2. SP1 then acts as a microphone, and sounds are reproduced in SP2.

With the switch in the opposite position, S1A selects SP2 as microphone, and amplified signals pass via S1B to SP1, so that speech can be in the other direction.

The switch, amplifier, battery and SP2 are assembled as the master control unit, and twin leads run to SP1. The person at SP1 should say "Over" so that the other person can operate the switch, and he in turn should say "Over" to signify that SP1 is again acting as microphone.

A simplified home-made change over switch is shown in Fig. 5. This is two strips of metal, pivoted so that they can be swung from side to side, and bearing on two of the three round-headed screws. A strip of insulating material is fitted so that both strips move together. In Fig. 5, this switch is in its "Off" position.

Ready made 2-pole 2-way switches can be obtained and used instead (see Fig. 49). Spring-loaded push switches are also available, and can be wired to provide "press to talk". A switch without a central "Off" position can also be used, a single pole on-off switch then being added in one battery connection.

The master unit can call SP1 by switching to speak, and allowing SP2 to be used as microphone. If calling in the other direction is required, a third connection, with bell push, battery and buzzer can be used. (An alternative, requiring more circuitry, but avoiding the need for a third wire, will be found in "Two Transistor Electronic Projects", (BP30, Babani Press.)

Fig. 5B shows assembly of the intercom in a wooden case. A 3½ inch speaker is suitable, so the whole can be quite a compact unit. The home-constructed switch, using the bolts below the speaker, can be replaced by a ready-made switch, as described, if preferred.

Best results are obtained with speaker units of quite high impedance – say 45 ohm to 75 ohm or so. The few other components are assembled on a small insulated board, as for earlier circuits.

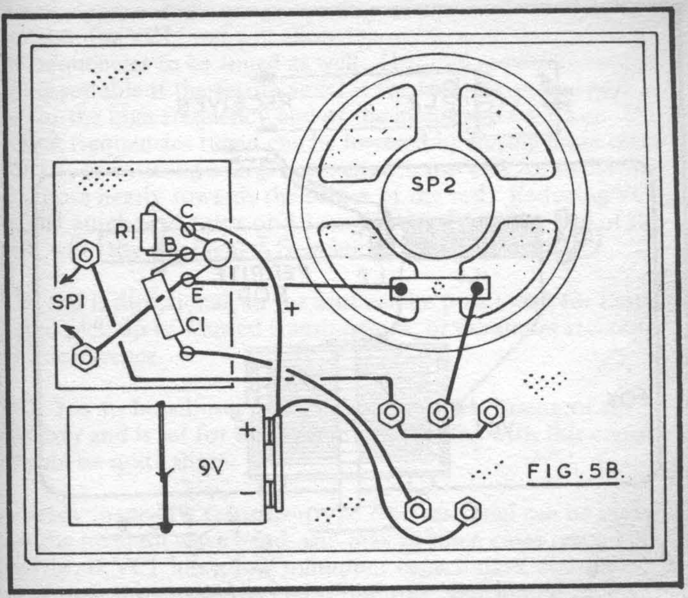

FIG.5B.

The speaker at the remote point also requires a small case. Dimensions are not important, and it is possible to utilise plastic boxes of suitable type for each unit. Connections between units are by twin bell wire, or any similar twin conductor. With a transistor of other than the type shown, the value of R1 can be chosen to give maximum amplification.

Tuned Ferrite Aerial

This simple device will allow "portable" reception with the old type of receiver which normally requires an external aerial, and it can also give improved reception in terms of freedom from interference on the 160 metre amateur band and adjacent frequencies. As shown in Fig. 6, it consists of a tuned circuit with ferrite rod, and a coupling winding to connect to the receiver aerial and earth sockets. It will be found very useful with a communications type receiver, where a long extended outdoor or similar aerial results in a high level of noise and interference.

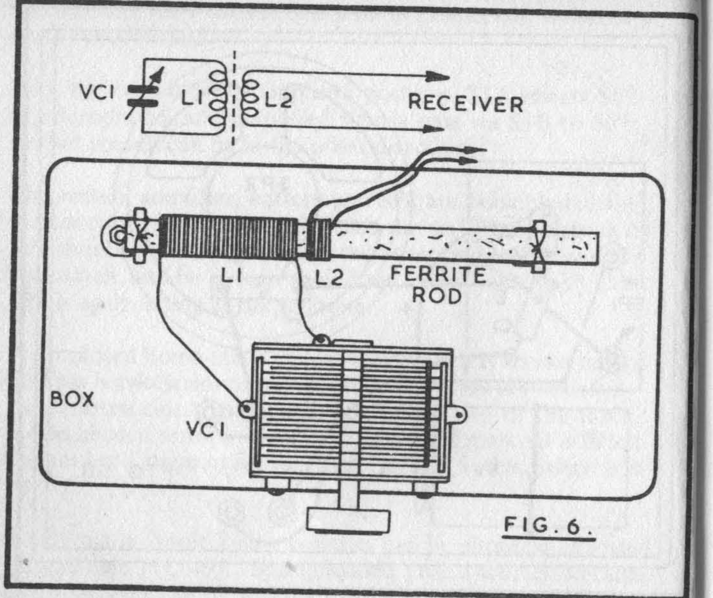

FIG. 6.

The device is constructed in a plastic lunch box about 6 x 4 x 4 inch or other *insulated* case. Some boxes of this kind, though strong, are of brittle material. These must be drilled without too much pressure, or they may crack. Drill points should also be sharp.

The rod is 6 inch long and 3/8 inch in diameter, and it is supported by two strips of insulating material, which are fixed with brackets. Each strip has a V-shaped notch, so that the rod can be held with thin string. L1 is the tuned winding, and it had sixty turns of 24 swg enamelled wire, side by side. Connections can be made to the frame (moving plates) and fixed plates of the capacitor VC1 by means of small bolts, or may be lightly twisted.

L2 is the coupling winding. Many communications type receivers have a low impedance input circuit, and twelve turns will be adequate for L2. The same wire can be used as for L1. With other receivers it may be worth while to experiment with the number of turns on L2, for best reception.

An air-spaced single gang capacitor of about 315pF is most suitable for VC1, and will allow the usual medium wave band of frequencies to be tuned as well. However, a smaller capacitor is suitable if the ferrite aerial is for 160m reception only, or for the high frequency end of the medium wave band. The actual frequencies tuned can be lowered by having more turns for L1, or by using a larger capacitance for VC1, or by having L1 more nearly towards the centre of the rod. Reducing VC1, or the number of turns on L1, or moving L1 to the end of the rod, will raise the highest frequency tuned.

The rod is directional, so the unit can be positioned for best signal pick up of wanted transmissions, or minimum reception of interference.

VC1 has to be adjusted in conjunction with tuning of the receiver and is set for best reception. Tuning with this control should be quite sharp.

A ready made MW transistor type receiver aerial can be useful for the medium wave band, and may in some cases reach 160 metres, if VC1 has a low minimum capacitance, and the winding is quite near the end of the rod. But should tuning not reach a high enough frequency, some turns will have to be removed.

Diode Radio

A lot of interest can be obtained from a diode radio receiver. It requires few components, no battery or other supply, and can give good headphone reception of a few of the more powerful transmitters. It can also be used as a tuner, as shown later, and will provide excellent audio quality for local station reception.

The circuit in Fig. 7 is intended for medium wave reception. That is, from about 200 metres to 500 metres, or 1500 kilohertz to 600 kilohertz.

The coil L1 is tuned by the variable capacitor VC1. At resonance with the wanted transmission, a signal voltage is developed across the coil. Detection or demodulation by the diode D1 allows the audio component of the signal to be recovered, and taken to the phones or to an amplifier. Such "crystal sets" were used in very many homes, in the early days of radio.

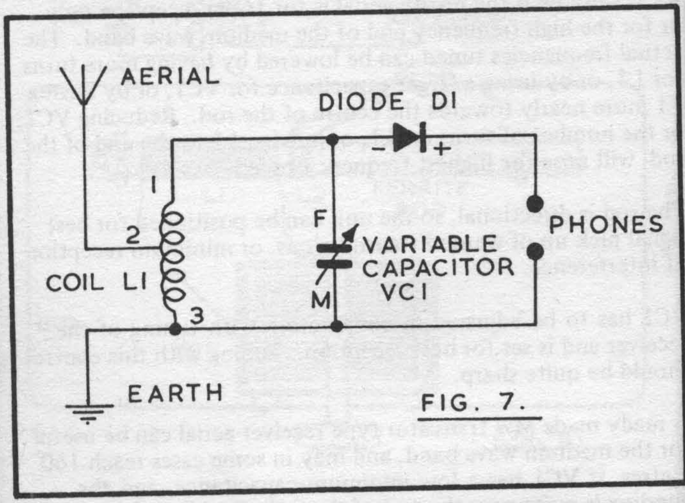

FIG. 7.

Fig. 7B shows the practical layout. It is convenient to use an insulated board larger than necessary, so that a transistor amplifier can be added later. A board about 5 x 8 inch is suggested, and it can be supported on 1 inch high runners (strips of wood) so that there is clearance for screws or terminals.

Coil L1 is wound with 26swg or similar wire, and has seventy turns, side by side. Secure the wire by passing it through two small holes, wind on about thirty five turns, and form a loop. Continue to wind in the same direction for a further thirty five turns, cut off the wire, and secure the end through two holes. The tube used can be 1½ inch to 2 inch in diameter, and can be paxolin, or can be made by winding thin card round a suitable object, smearing a little adhesive on to form a tube.

Point 1, the beginning of the coil, is taken to the fixed plates tag F of the variable capacitor, which is also connected to diode D1. The tapping 2 goes to the aerial terminal. End 3 is taken to the moving plates connection or metal frame of the capacitor. Ends 1 and 2, and loop 3, are left long enough to reach these points, and the insulation is carefully scraped off the wire where connections are to be made. The capacitor VC1 can best be about 500pF, but this exact value is not essential.

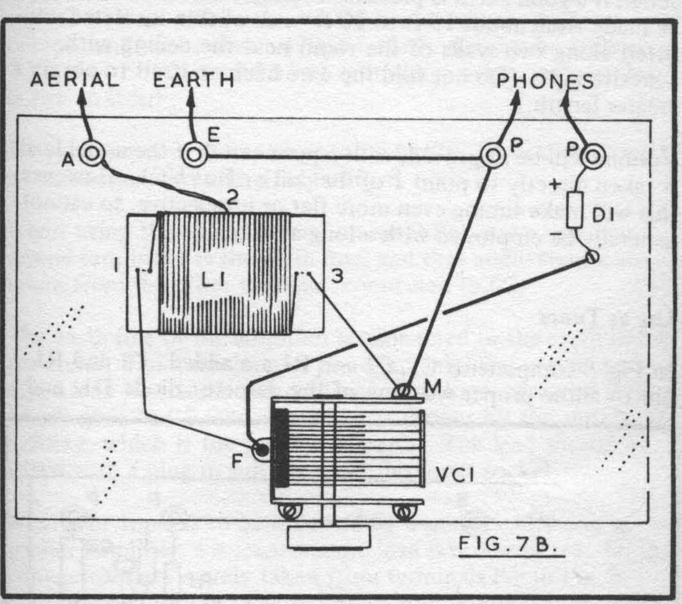

FIG.7B.

Four terminals, for Aerial, Earth, and Phones, will complete the receiver. Best results will be had by connecting a good headset, of about 500 ohm to 2,000 ohm. Very low impedance earpieces, such as are employed with some transistor receivers, are not suitable here.

With a simple receiver of this kind it is necessary to employ a reasonably effective aerial and earth. Only in some circumstances will volume be great enough with no earth, or with a short aerial. The earth connection can be taken to a metal spike in the ground, or to a metal object buried in soil, or to any other available earth other than a mains circuit earth. (No connection must be made to any mains plug or other mains circuit.) A cold water pipe may provide a good earth. Gas pipes should not be used.

For the aerial, a length of aerial wire some 30 ft to 75 ft or so in all may form the aerial itself and lead-in. If this can be suspended quite high, and clear of buildings or earthed objects, signal strength will be improved. However, satisfactory reception may be obtained from an indoor aerial, or shorter aerial, if a good earth is present. A simple indoor aerial can be made from about 15 ft to 20 ft or so of thin insulated wire, fitted along two walls of the room near the ceiling with adhesive tape. (Do not fold the wire back on itself to obtain a greater length.)

Volume will be improved, with a poor aerial, if the aerial lead is taken directly to point 1 of the coil or F at VC1. However, this will make tuning even more flat or unselective, so cannot generally be employed with a long aerial.

Use as Tuner

In Fig. 8 components C1, C2 and R1 are added. C1 and R1 are to allow proper working of the detector diode D1, and may be considered as replacing the headphones. C2 is a coupling capacitor passing audio signals to the following amplifier.

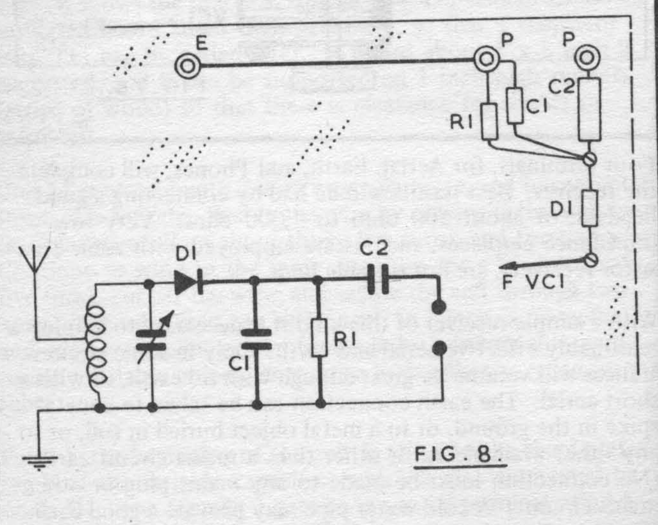

FIG. 8.

As with the previous circuit, any general purpose detection type diode is satisfactory for D1, such as the OA91 and many similar devices. The value of C1 is not critical, and this component can be about 470pF or 1nF. R1 can be 100k to 220k or so. C2 should be of quite large value, and can be 0.1μF to 0.5μF.

D1 can be moved from the position in Fig. 7B, and the few extra components added, as in Fig. 8.

When using the tuner to feed an amplifier, note that one phone terminal P is the earth line, and that audio signals are taken from the other terminal, connected to C2.

The earth line of the amplifier is connected to the earth line of the tuner. For a sensitive or powerful amplifier, a screened audio lead is recommended. This has an inner screened lead, which goes to C2, and a conductor former by the outer brading, which is the ground or earth. The lead should be fitted with a plug to suit the amplifier input socket.

Where the tuner is to be followed by a small headphone or similar amplifier, a screened audio lead is not required. So then connections are merely taken from terminals P-P in Fig. 8, remembering to join amplifier earth and tuner earth circuits as explained.

Medium and Long Wave

The addition of long wave coverage is worthwhile in some areas, and the coil can be modified as in Fig. 9 so that both MW and LW can be tuned.

A tube about 3 inch long and approximately 1½ inch in diameter will be necessary. The section L1 is wound as before. L2 consists of about two hundred turns of 34swg or 36swg wire, in two compact piles of one hundred turns each. All turns throughout are in the same direction. The junction of L1 and L2 goes to the waveband switch S1. The second switch terminal goes to 4 and earth line. MW tuning is obtained with S1 closed, and LW tuning with S1 open.

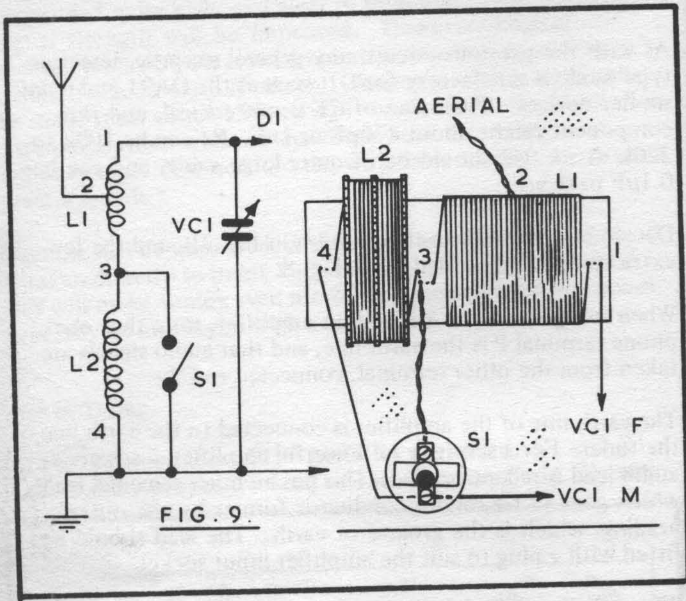

FIG. 9.

For accurate coverage of the usual MW and LW bands, and using a 1¼ inch diameter tube, L1 can be eighty-four turns of 32swg enamelled wire, with tapping 2 sixty turns from point 1. Leave a space of 5/8 inch and glue on a card disc cut to fit the tube. Also glue on two further discs, leaving spaces of ¼ inch between them. L2 is wound with 32swg double cotton covered wire, with one hundred and twenty turns in each slot (making 240 in all for L2). The coil can be fixed with two screws or small bolts, with spacers or extra nuts to hold it a little clear of the board. The windings should not be varnished, covered with adhesive, or wax.

With circuits of this kind there is no ganged tuning, or need for exact tuning of a particular frequency band. This means that coil winding details (wire gauge and covering, and diameter, or number of turns) can be varied to suit materials to hand. Ready-made coils can also be tried. Some commercially made coils with cores can give a slightly better performance than the air-cored home made type of coil. Coils can also be wound in a similar way to that in Fig. 6.

One Transistor Amplifier

Headphone volume obtained from the crystal diode receiver can be considerably boosted by adding a single transistor audio amplifier. Space for this can be found on the circuit board.

In Fig. 10, resistor R2 and the transistor have been added, with a switch S2 for on-off switching of the battery. R1, C1, D1 and C2 are already present, as shown in Fig. 8.

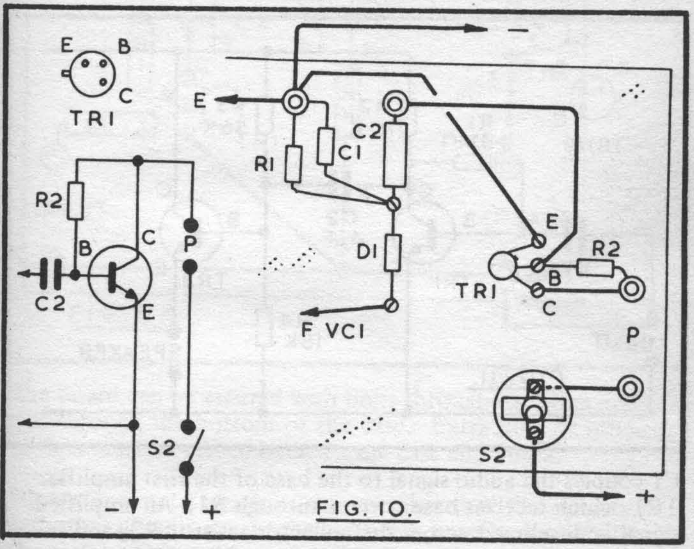

FIG. 10.

TR1, R2 and S2 are connected as shown. For TR1, a BC108, 2N706, or similar general purpose NPN transistor will be satisfactory. R2 can generally be 1 megohm with satisfactory results, but it can be worth while trying values from about 100k to 2.2 megohm here, as performance depends on the transistor, phones, and battery voltage.

Medium and high impedance phones, such as a 600 ohm DC headset, will be most suitable. The battery can be 4.5v to 9v.

If the phones are first connected to terminals P-P in Fig. 8, then taken to terminals P in Fig. 10, a very considerable increase in volume should be obtained.

If a PNP transistor is to be tried in this circuit, the polarity of the battery *must* be reversed.

Two Transistor Amplifier

Fig. 11 is the circuit for an easily made 2-transistor amplifier which can provide reasonable loudspeaker volume. The audio signal is taken to the input terminals, battery negative being the grounded side of this circuit.

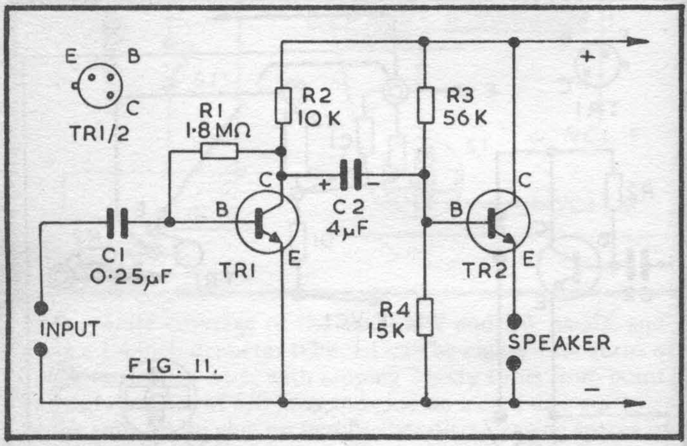

FIG. 11.

C1 couples the audio signal to the base of the first amplifier TR1, which receives base current through R1. An amplified signal is developed across the collector resistor R2, and is coupled to the base of the following amplifier TR2 by means of C2.

The two resistors R3 and R4 set the base operating conditions for TR2, and the output is taken from TR2 emitter.

Fig. 11B shows assembly of the components on a small insulated board. If preferred, connections between components can be on the underside of the board, in the way shown for later circuits. Fit red and black flexible leads for battery connections. The amplifier, with its battery, can be fixed in an insulated box or metal case. Sockets can then be used for input and output circuits, and an on-off switch can be placed in one battery lead.

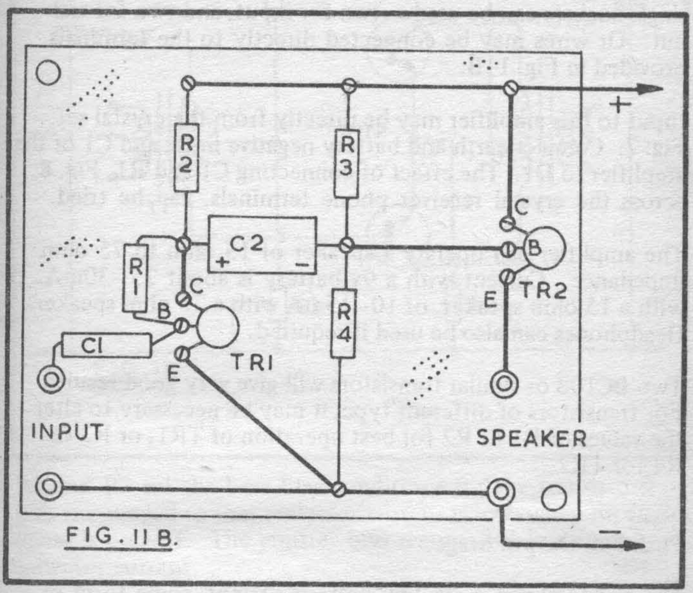

FIG. 11 B.

The board can be secured with bolts through its fixing holes, and through the bottom of the case. Extra nuts or other spacers must be placed between the case and board.

Whenever a metal box is used, remember that all connections and other circuits except the earth line must be insulated from the metal. The metal case itself is normally connected to the earth line, and a lead can be put under one of the fixing screws for this purpose. In Fig. 11, both input and speaker are returned to the earth line. So 3.5mm jack sockets of the type in which the outer or sleeve contact is common to the fixing bush could be used, as these points will be connected to battery negative and the earth line and metal box. But with some circuits it is necessary to use an insulated jack socket, or to insulate this from the metal. An example of this would arise if the speaker were placed in the collector circuit of TR2, as the outer or sleeve circuit would then need to go to battery positive.

Instead of the type of socket which takes a jack plug, ordinary single sockets can be used – two for input, and two for output. Or wires may be connected directly to the terminals provided in Fig. 11B.

Input to this amplifier may be directly from the crystal set, Fig. 7. Connect earth and battery negative lines; and C1 of the amplifier to D1. The effect of connecting C1 and R1, Fig. 8, across the crystal receiver phone terminals, can be tried.

The amplifier can operate a speaker of 15 ohm to 75 ohm impedance. Current with a 9v battery is about 25–30mA with a 15 ohm speaker, or 10–15mA with a 75 ohm speaker. Headphones can also be used if required.

Two BC108 or similar transistors will give very good results. For transistors of different type, it may be necessary to alter the values of R1 or R2 for best operation of TR1, or R3 or R4 for TR2.

Push-Pull Output Stage

For good volume, with low battery current, some form of push-pull output stage is generally used. This has two transistors, and each amplifies one half of the audio signal. Base bias conditions of the two transistors are so arranged that only a low current flows during silent intervals. Overall battery drain is thus much lower than if a single transistor were used to obtain a similar level of volume.

Fig. 12 is a typical push-pull output stage. T1 is the input or driver transformer, and an audio signal is taken to its primary P. The secondary S is centre tapped. So for one half of the audio cycle one transistor base is driven positive and the other transistor base is driven negative, this being reversed for the remaining half cycle. Thus each transistor conducts alternately, and the centre-tapped output transformer T2 combines the collector currents, to operate the speaker connected to the secondary S.

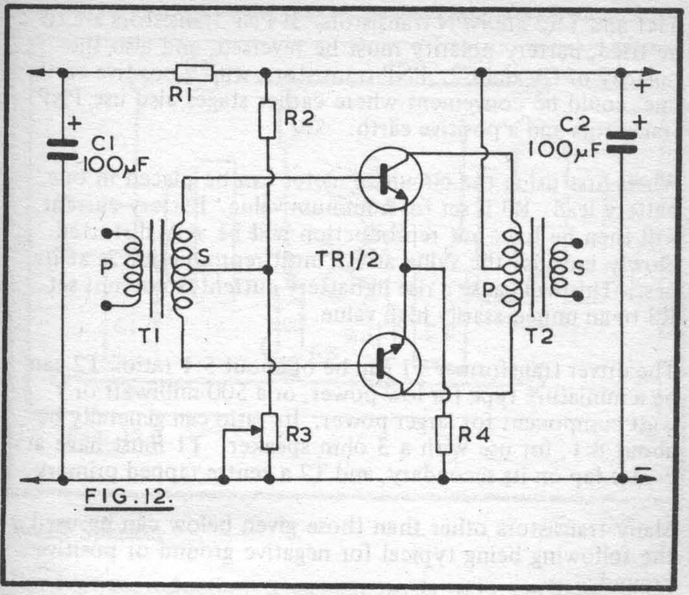

FIG.12.

R2 and R3 set the base bias conditions for the transistors. Bias is arranged so that collector current is low, when no audio signal is present. The emitter bias resistor R4 limits peak collector current.

If R3 is a pre-set resistor, as in Fig. 12, operating conditions can be arranged to suit the transistors fitted. For best results TR1 and TR2 should be a matched pair. This means that they will have similar gain and current characteristics. However, two transistors of the same type number will generally be satisfactory.

It is assumed that the same battery will supply earlier stages of the amplifier, and current for this purpose is taken from C1 and R1. R1 and C1 prevent audio signals present in the positive line passing back to earlier amplifier stages. The by-pass capacitor C2 also helps in this respect. With simple circuits not giving much overall amplification, or small power outputs, R1, C1 and C2 might all be omitted.

TR1 and TR2 are NPN transistors. If PNP transistors are to be tried, battery polarity must be reversed, and also the polarity of C1 and C2. PNP transistors, with a positive earth line, could be convenient where earlier stages also use PNP transistors and a positive earth.

When first using the circuit, a meter can be placed in one battery lead. R3 is set for minimum value. Battery current will then be low, but reproduction will be very distorted. Slowly increase the value at R3 until reproduction is at its best. This will cause a rise in battery current, so do not set R3 to an unnecessarily high value.

The driver transformer T1 can be of about 5:1 ratio. T2 can be a miniature type for low power, or a 500 milliwatt or 1 watt component for larger power. Its ratio can generally be about 8:1, for use with a 3 ohm speaker. T1 must have a centre tap on its secondary, and T2 a centre tapped primary.

Many transistors other than those given below can be used, the following being typical for negative ground or positive ground.

Negative earth (NPN transistors). 2 x BC108. R1 1.2k. R3 100 ohm fixed or 250 ohm pre-set. R4 4.7 ohm.

Positive earth (PNP transistors). 2 x AC128 or 2 x AC142. R2 4.7k. R3 82 ohm or 250 ohm pre-set. R4 4.7 ohm.

R1 will generally be 1k to 2.2k, depending on the earlier amplifier. The primary of T1 can be connected to terminals P in Fig. 10; or alternatively to any audio source where a reasonably strong signal is present.

Where signals are too strong, an audio gain control needs to be fitted. In Fig. 8 this can be done by omitting R1, and disconnecting C2 from D1. Connect the outer tags of a 100k volume control or log pot from D1 to earth line. Take C2 to the centre tag or slider of the potentiometer. Any required audio level can then be taken off by adjusting the control knob.

Fig. 12B shows the board layout.

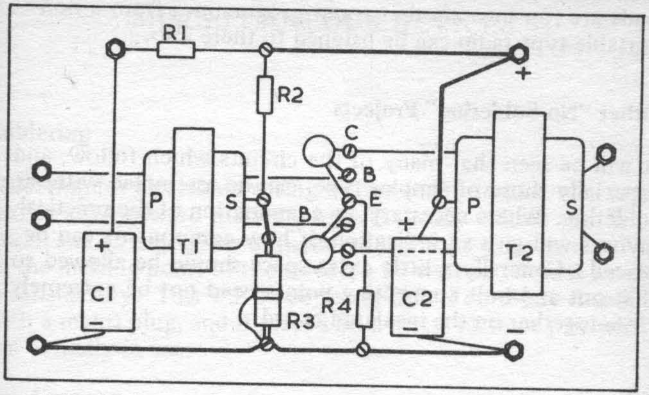

FIG. 12 B.

Boost Speaker

For best reproduction, a speaker needs to be enclosed in a cabinet or fitted to a baffle board. A cabinet can be made from wood, and can be about 6 x 10 inch and 4 inch deep, or as required for the speaker unit. Cut an aperture in the cabinet front to match the cone diameter, and glue thin fabric behind this opening. The speaker can be fixed with nuts and bolts, or short woodscrews driven in from behind.

Connect flexible leads, with separate plugs or a jack plug, to suit the equipment.

Small pocket receivers, which have a miniature speaker, can give much improved volume and reproduction when used with a larger speaker. If a headphone socket, operated from the receiver output stage, is present, it is only necessary to plug the speaker into this. A fairly large speaker, of sensitive type, will give best results. A unit 5 inch to 6 inch in diameter, or about 4 x 7 inch, will be convenient. The speaker has to be of approximately the correct impedance, which should be found in the receiver operating instructions. Where the speaker impedance is wrong, a matching transformer has to be employed between receiver and speaker.

Extension speakers are less frequently used today, but if two leads are run into another room programmes from a non-portable type radio can be listened to there also.

Other "No Soldering" Projects

It will be seen that many of the circuits which follow, and especially those of simpler type, can be assembled without soldering. Where necessary, an examination of the practical layouts will give an indication of how components can be placed. Generally a little extra space should be allowed so that nut and bolt connecting points need not be extremely close together on the insulated board.

SECTION 2
MISCELLANEOUS DEVICES

Soldering

Soldering for electronic wiring will be found to be a very easy and straightforward matter. An electrically heated iron is usual, and as large areas of metal do not have to be raised to the melting point of the solder, a 20 watt to 25 watt iron is satisfactory. This will be perfectly safe if correctly fitted with a mains plug, and if placed on a metal plate or tray when not actually in use.

It is best to use cored solder, as made for electronic wiring.

Leads, tags and other items to be joined should be clean and bright. They have to be raised in temperature to the melting point of the solder at the actual points where a joint is to be made, and the active core of the solder, or flux, must be present here when the joint is being made.

The iron should be tinned – that is, have melted solder on its tip – and have reached its correct temperature. Leads can be hooked or looped into position. The joint is then heated with the iron, and the solder is simultaneously applied to the joint, so that it melts and flows over the surfaces to be soldered. This usually takes place in a second or two. The solder and iron should be removed at once, as excessive heating may damage some components, and especially transistors or other semiconductors. The joint cools in a few seconds, and should be rigid, and solder having flown onto the items to be joined.

A probable cause of poor joints is insufficient heating of items to be soldered, or carrying solder to the joint on the iron, so that the activity of the flux is wasted.

Electronic items such as tags and the wire ends of components are generally designed for easy soldering. They will be copper, tinned copper, and similar metals and alloys which solder readily. For this reason, difficulties are unusual.

Some metals do not solder readily, or need special methods. Aluminium is one of these, and it is not used where popular cored solders intended for copper or general wiring are employed.

Make sure that heat from the iron does not damage furniture or other items, and keep the flexible cord away from the heated tip. Always switch the iron off at once when the work in progress is to be left.

One Transistor Audio Oscillator

This can be used instead of a bell or buzzer warning, or for code practice and similar purposes. It needs very few components, Fig. 13, and can be made to work satisfactorily with a wide range of transistors.

In Fig. 13, an NPN transistor is shown in the theoretical circuit, so battery polarity is as indicated, with negative to emitter. If a PNP transistor is fitted, battery polarity must be reversed.

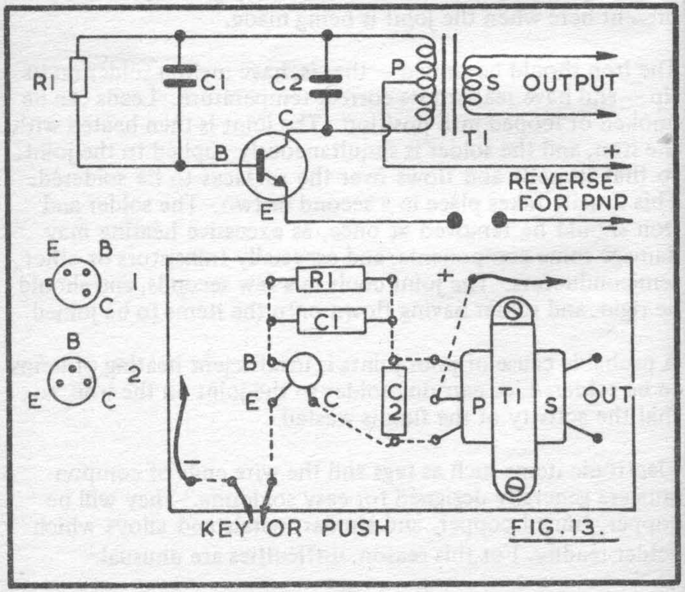

FIG.13.

(1) shows leads from a typical NPN transistor, such as the BC108, 2N706, and similar types. (2) is the base of the PNP AC128 transistor.

T1 can be a small speaker matching transformer, as used for push-pull stages. A centre tapped primary is necessary. The secondary S can be taken to a loudspeaker or headphones.

The values of R1, C1 and C2 can be varied considerably. C1 can be 0.1μF or so. R1 determines to some extent the power obtained, and can be about 22k, increased to 47k or 100k for lower power. The value of C2 has considerable influence on the note, or frequency of oscillation. Increasing the value of C2 lowers the pitch. C2 can be about 47nF to 4.7nF or thereabouts. A 9v battery can be used. Changes to the supply voltage will influence the note produced.

Fig. 13 also shows a layout for components, using either a fully perforated board, or a panel drilled to suit. Transformer T1 is fixed by screws. Small transformers of this type are usually wire ended, or have pins which pass down through holes in the board.

The wire ends of the resistors and capacitors are passed down through suitable holes, and bent over. The board is then turned so that the underside wiring can be carried out. In most places the wire ends of the components will be long enough to reach connecting points. Elsewhere, 24swg tinned copper wire, or similar connecting wire, will be used. Make the soldered joints as described, and snip off unwanted ends of wire. It is as well to leave the transistor to last, as this avoids heating it unnecessarily when soldering other items. Lengths of flex, anchored at holes in the board, are used for battery, key (or push) and speaker connections.

Multivibrator

Fig. 14 is the circuit of the multivibrator assembled in Fig. 2. Reference to Fig. 14B will show how this can also be prepared on a tag board.

The tag board has two rows of tags fitted to an insulated board, and it is five tags long in Fig. 14B. Boards of this type can be obtained in many sizes. It is also easy to cut a section from a long tag board, to suit the needs of the item to be assembled. Such boards are very handy indeed for electronic assemblies, as the tags support resistors, capacitors and other items, and a neat layout is readily achieved.

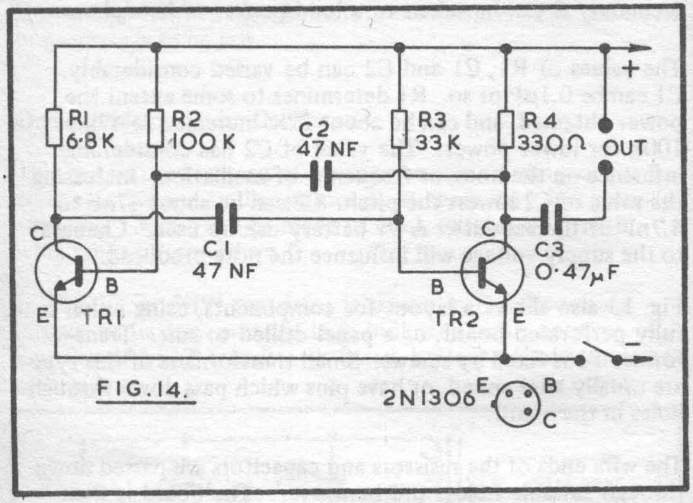

FIG. 14.

Tag strips are also useful for projects, and may have any number of tags required. Long tag strips can be cut down as necessary. Short tag strips have one tag for mounting the strip, longer strips having two tags formed as brackets for this purpose.

Place the components approximately as in Fig. 14B, and solder the wires to the tags. If any component values may be changed, do not twist the leads tightly round the tags before soldering, or removal will be difficult.

Wires which cross can be kept clear of each other, and then no insulated sleeving will be necessary. The transistors can be soldered on last, and can readily be changed without disturbing other parts.

The same layout can be used for the next project, though the component values are different.

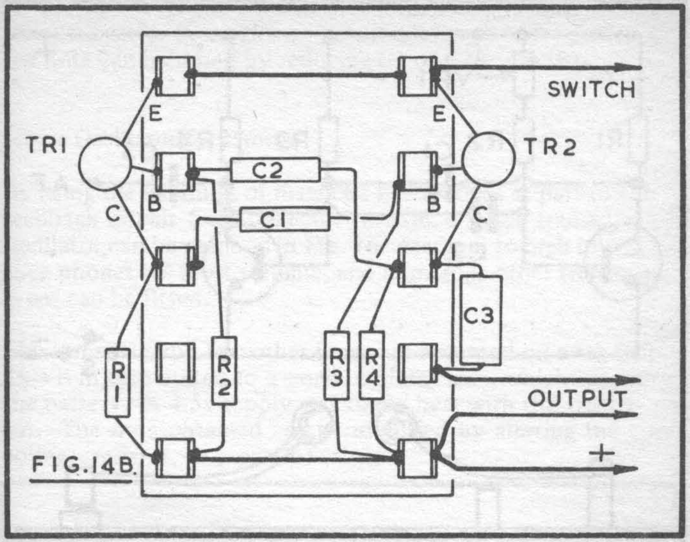

FIG. 14B.

Glissade

This device will produce swooping, rising and falling tones, for "effects" purposes, or can with additional skill be used to obtain musical notes through several octaves. The circuit is shown in Fig. 15, and reference can be made to Fig. 14B for a practical layout on a tag board.

In Fig. 15, suitable component values are:

R1	10k	C1	22nF
R2	12k	C2	22nF
R3	100k	C3	22nF
R4	10k	VR1	100k linear pot.

Transistors: 2 x BC108 or similar types.

The circuit is a wide frequency audio oscillator, in which the pitch can be controlled by VR1, and "on" switching is obtained by S. The oscillator is placed in a box, so that switch S and VR1 can be mounted on the top. The switch is a strip of springy metal, fitted with a spacer and bolt, so that it is just clear of a further bolt. A small downwards movement of the arm or knob completes the circuit.

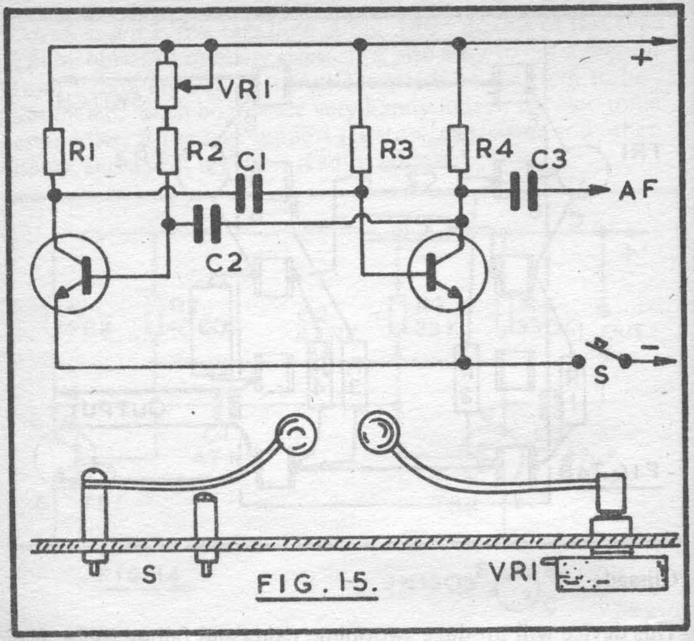

FIG. 15.

VR1 has a ¼ inch shaft coupler, into which is screwed a length of 4ba threaded rod. A knob is also fitted to the end of this rod. It can be swung backwards or forwards, to change the resistance of VR1. Connect VR1 so that movement of the knob towards the right raises the pitch.

Output from C3 can go to high resistance headphones, for amusement, but should be taken to an audio amplifier. Gain of the amplifier will have to be turned back somewhat, as the output from the glissade is at a quite high level.

Swooping tones are produced by holding down the switch S and swinging the control knob attached to VR1 from side to side. For notes which do not swoop to frequency, switch S must be released while VR1 is rotated. A slight quivering motion of the hand controlling VR1 will produce a vibrate effect. It is useful to have an ordinary switch in parallel with the knob control S, and to have the amplifier audio gain control near the left hand, so that fading effects can be produced as well.

About 4.5v to 9v can be used, this influencing volume. Should other transistor types give an unsuitable range of frequencies, the note can be raised by reducing C1 or C2 (or both).

Morse Oscillator for Phones

By using the windings of magnetic headphones as part of the feedback circuit from collector to base, a single transistor oscillator can be made as in Fig. 16. Medium to high impedance phones are most suitable, and numerous other transistor types can be fitted.

The transistor and few other items are mounted on a tag strip. This is in turn bolted to a small insulated box, which can carry the battery. A 4.5v supply was found best with this oscillator. The note obtained can be modified by altering the voltage, resistor, or capacitor values.

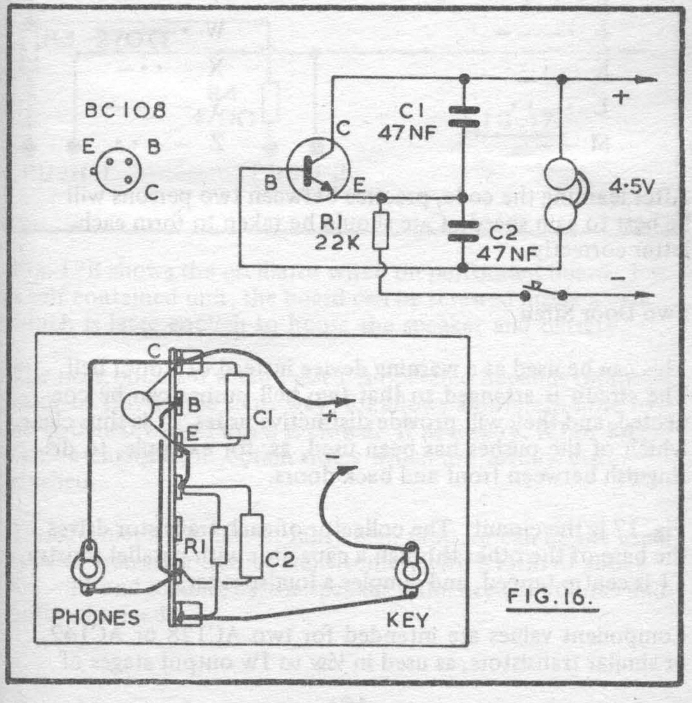

FIG. 16.

For serious code practice a good Morse key is required. It is adjusted so that only a small movement of the wrist is needed to operate it. Each letter consists of short and long sounds, shown as (•) and (–). The (–) has three times the length of the (•)

A •–
B –•••
C –•–•
D –••
E •
F ••–•
G ––•
H ••••
I ••
J •–––
K –•–
L •–••
M ––

N –•
O –––
P •––•
Q ––•–
R •–•
S •••
T –
U ••–
V •••–
W •––
X –••–
Y –•––
Z ––••

After learning the code, practice between two persons will be best to gain speed. Care should be taken to form each letter correctly.

Two Door Siren

This can be used as a warning device instead of a door bell. The circuit is arranged so that two bell pushes can be connected, and they will provide distinctive notes. It is thus clear which of the pushes has been used, as, for example, to distinguish between front and back doors.

Fig. 17 is the circuit. The collector of each transistor drives the base of the other through a capacitor with parallel resistor. T1 is centre tapped, and couples a loudspeaker.

Component values are intended for two AC128 or AC142, or similar transistors, as used in ½w to 1w output stages of

receivers and small amplifiers. Other transistors will of course operate in this circuit. The pitch obtained can be modified by changing C3. The circuit shown is for PNP transistors, and battery polarity has to be reversed for NPN transistors.

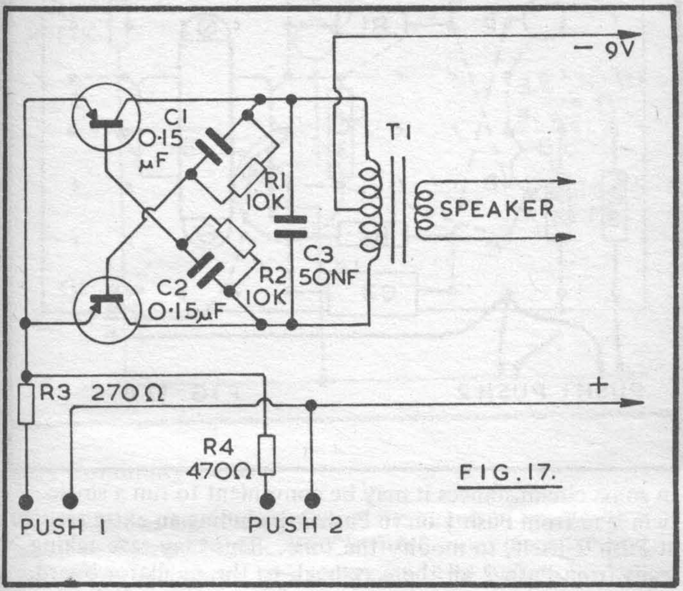

FIG.17.

Fig. 17B shows the oscillator wired on perforated board. For a self contained unit, the board can be screwed inside a case which is large enough to house the speaker and battery.

The note obtained when Push 1 is operated depends on the value of R3. In the same way, R4 allows a different note to be obtained when Push 2 is closed. If necessary, R3 or R4 can be changed, to obtain notes which are readily distinguished.

Transformer T1 is best of fairly high ratio, or can be the usual type of transistor output transformer, with a ratio of about 8:1. T1 and loading by the speaker influence the volume and notes obtained.

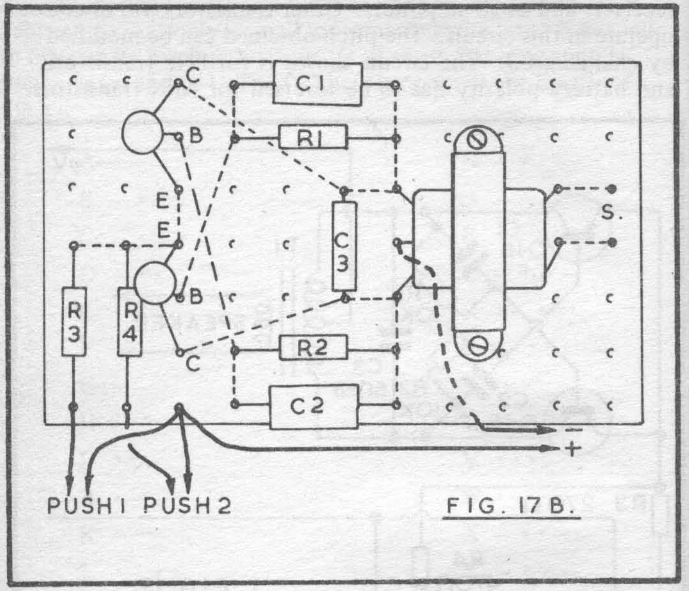

FIG. 17B.

In some circumstances it may be convenient to run a single twin lead from Push 1 on to Push 2, including an extra resistor at Push 2 itself, to modify the tone. This may save taking leads from Push 2 all the way back to the oscillator board.

UJT Oscillator

The uni-junction transistor is a device having emitter, Base 1, and Base 2, Fig. 18. B2 receives positive bias, in this case through the headphones. When forward bias on the emitter E is reached the UJT conducts.

In Fig. 18, the emitter potential rises as C1 charges through the resistor R1. When the UJT conducts, C1 is discharged, and the process is repeated. This can take place over a very wide range of frequencies, depending on the values used for R1 and C1. Here, these are chosen to produce an audio tone.

A battery supply of 3v to 9v will be convenient. For use of the UJT oscillator during code practice, place the key in one battery lead.

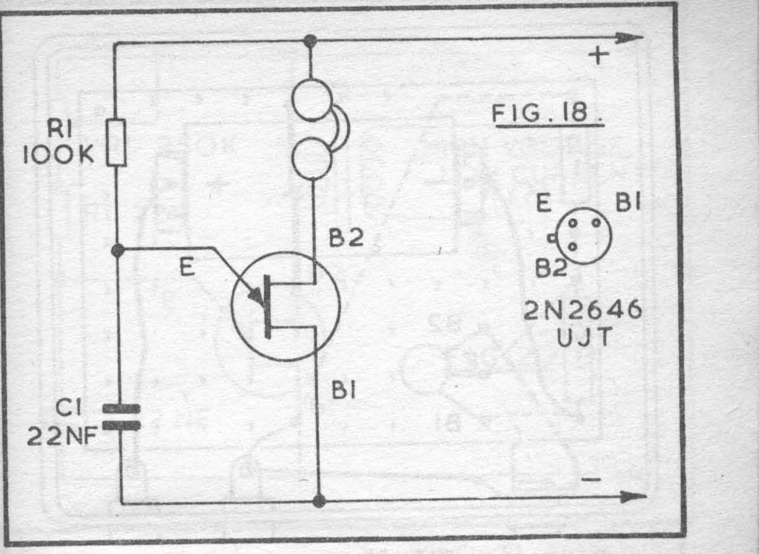

FIG. 18.

UJT Continuity Tester

Fig. 19 shows actual assembly of the circuit in Fig. 18 on a small perforated board. This is fitted in an insulated box, which carries a phone jack socket, and two sockets for Morse key or test leads. A 9v battery is used, secured with string or elastic.

For continuity tests, take prods or clips on the test leads to the suspected item. If the circuit is in order, the UJT oscillator will produce an audio tone. Oscillation will be obtained with a series resistance between the prods of up to 20k or more, so various components can also be tested, and some idea of the resistance in circuit can be gained from the sound obtained. The tester can also be used to check bulbs, lamps, fuses, extension and other leads, and similar items – these must of course be removed or disconnected from the equipment, before being tested.

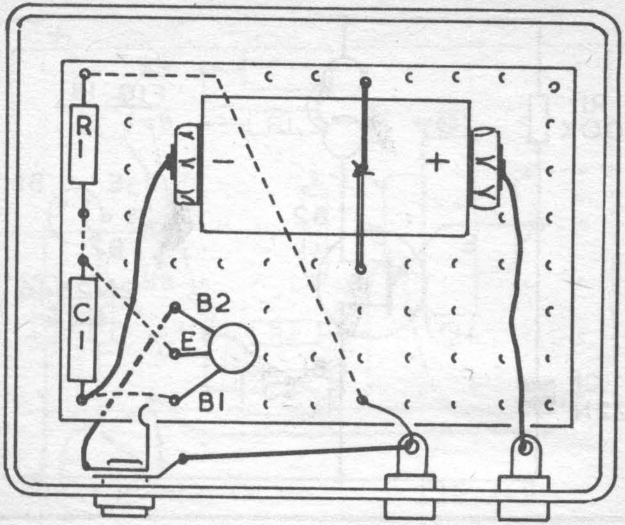

FIG. 19.

UJT High Voltage Producer

The UJT oscillator can be used to produce a high voltage, at very low current, by connecting a transformer as in Fig. 20. A loudspeaker transformer, as would be used in valve equipment, is suitable, and will have a ratio of about 45:1 or so. The usual secondary, or low-resistance winding, is connected to B2 and positive, so that a step-up of voltage is obtained. This should be found to light a small neon lamp.

VR1 is fitted to allow some adjustment, to find the most suitable frequency of oscillation. For maximum possible high voltage output, if required, operating voltage and current should be raised. No rating of the UJT must be exceeded. For the 2N2646, the supply should not exceed 30v, emitter current should not be over 50mA, and dissipation of power in the device should not be more than 300 milliwatts. (Dissipation can be found from Voltage x Current.)

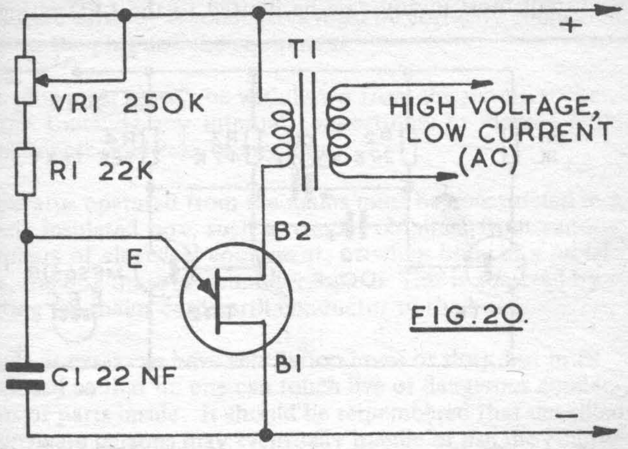

FIG. 20.

A small mains transformer can be used for T1, with a low voltage winding, such as that for a 6.3v heater circuit, as primary, and the 200/250v winding as secondary. The high voltage output is alternating current.

Signal Blinker

Fig. 21 is a multivibrator in which the feedback capacitors C1 and C2 are of high value, so that conduction alternates between TR1 and TR2 at a low rate. A light emitting diode is placed in the collector circuit of TR2, and so flashes on and off.

This device can be used for various instruments and other equipment, and a flashing indicator draws attention to a greater extent than does one continuously illuminated. The LED can be located at any required point, with leads running to it.

Transistors other than the types shown can be utilised. R4 can usually be about 150 ohm for a 3v supply, 470 ohm to 1.2k for 6v, or 680 ohm to 1.5k for 9v. This depends to some extent on the amount of light required when the indicator is operating. Lower resistor values at R4 will increase light, but current must in any case be limited to the LED rating.

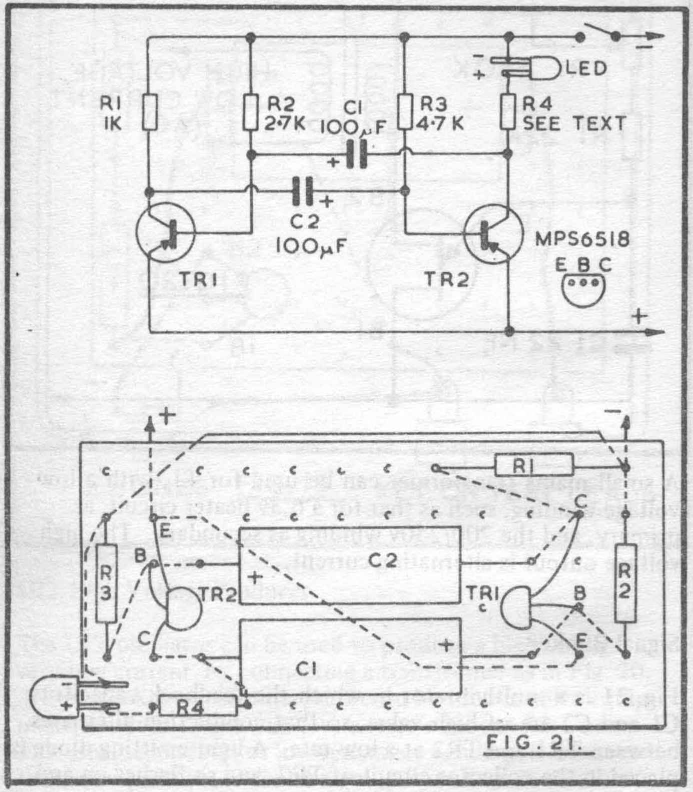

FIG. 21.

C1 and C2 need only be 6.4v working, and the device can be assembled on a very small perforated board. See p.61 for LED polarity.

Mains Wiring

All the devices described so far have operated from low voltage batteries, so than no shock hazard exists with them. However, it is necessary to employ mains power for some units, and proper care must then be taken to avoid any danger.

Power should be drawn from a 3-pin plug, and this should have a 3A or other low rating fuse installed. The live or L, neutral or N, and earth or E conductors must be correctly connected, both at the plug and the equipment.

The plug must always be withdrawn from the supply socket, before touching any internal connections, or making any adjustments or checks of any kind.

Apparatus operated from the mains must be constructed in a strong insulated box, such as can be obtained from various suppliers of electrical equipment; or when built in a metal box, the box must be soundly earthed. This is achieved by bolting the mains cord earth conductor to the box.

Boxes or cases can have ventilation holes or slots, but must be closed so that no one can touch live or dangerous connections or parts inside. It should be remembered that unskilled or unaware persons may eventually handle or use the equipment, so it must be arranged to present no unexpected hazards such as live terminals or parts, or openings for the fingers.

If these points are remembered, mains operated equipment can be safe to use and handle. No switch or other means of interrupting the circuit is put in the earth conductor, and any switch or equivalent device is placed in the live or L conductor. This leaves both earth and neutral permanently connected, for maximum safety.

Dimmer/Heat Control

It is often convenient to dim a filament lamp, or reduce the heat of a soldering iron to be left on for long periods, or the power of a low wattage heater. The silicon controlled rectifier circuit in Fig. 22 can be used for these purposes.

When the switch marked "Full" is closed, current for the load passes through this, so that full power is obtained.

When this switch is opened, current passes through the SCR. This device conducts only when the gate has been triggered, and the instantaneous point in the alternating current cycle where this arises is determined by the setting of the potentiometer VR1. Triggering is advanced as the wiper of VR1 is moved towards R1. The average power supplied to the load can thus be adjusted between wide limits. Since the device is operating as a half wave controller, full power is not obtained with VR1 at maximum, so the switch is closed as described.

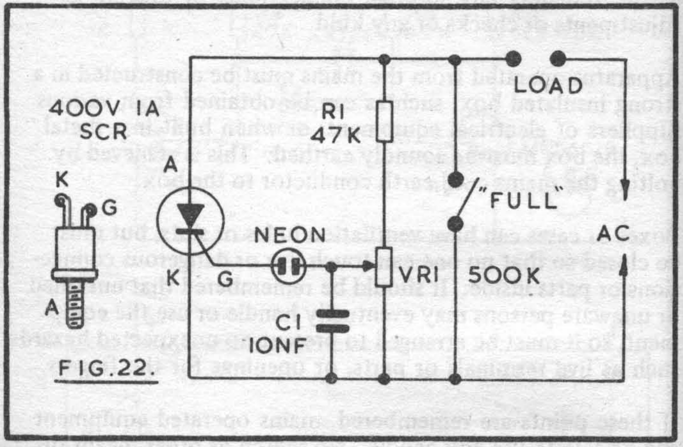

FIG.22.

Fig. 22B shows construction of the SCR controller in a metal box. The front of this carries a socket outlet, and the mains cord has a 3-pin plug, wired as explained earlier. It is thus easy to plug the controller into an outlet, and insert the plug of the lamp or other item to be used in the new outlet.

The mains cord is secured with a clamp, and passes through an insulated grommet. The tag strip supporting the small items is mounted by an unused tag. A strip of metal is cut to clamp round the body of the SCR, and is soldered to one tag, as shown. Run earth and neutral leads from the cord up through a hole, and connect them to the E and N terminals of the socket outlet. Also earth the case, at E. A well insulated wire runs from the L terminal of the outlet, to the "Full" switch. The outlet can be secured with bolts and nuts.

R1 should be 1 watt, and C1 is a 500v capacitor. The neon is a small type as used for indicators, etc.

After checking that the controller operates correctly, the back of the box should be screwed in place. A suitable size is about 4 x 5 x 2 inch deep.

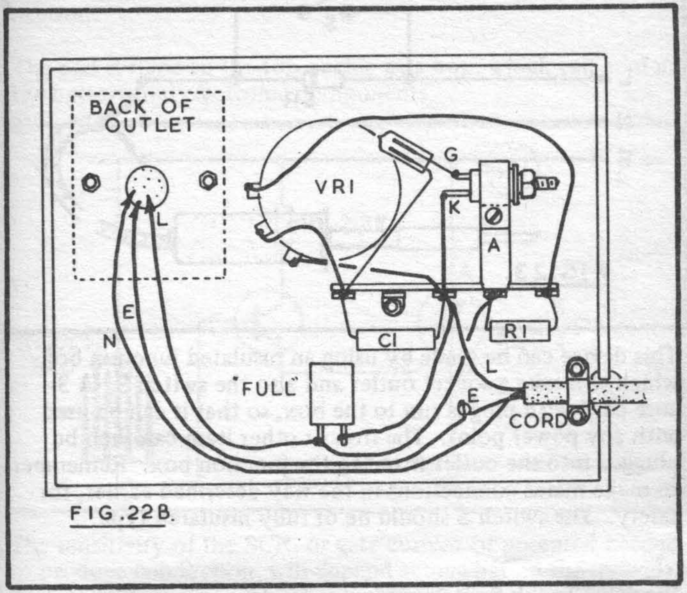

FIG.22B.

If a fully insulated box is used instead, then the bush of VR1 and of the switch should both be earthed by wires, except when these items are constructed wholly with insulated material. (This precaution is necessary to avoid any chance of metal parts outside the box becoming "live".)

Two Power

A circuit which will allow half or full power can be useful to reduce the heat of a soldering iron, or for dimming a filament lamp.

Fig. 23 is an easy circuit of this type. Full power is obtained when the switch S is closed. With S open, the silicon rectifier allows current to flow in one direction only, so that power is approximately halved. A 600v 1A or 1000v 1A rectifier is suitable.

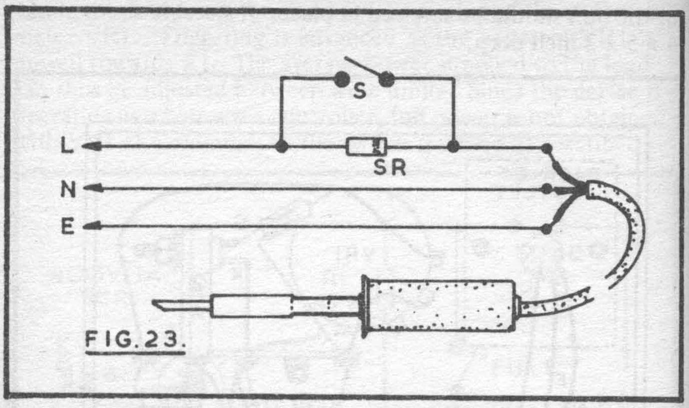

FIG. 23.

This device can be made by using an insulated junction box which will take a socket outlet and also the switch S. A 3-core cord with plug is run to the box, so that it can be used with any power point. The iron or other item can then be plugged into the outlet fitted to the junction box. Remember to make mains connections in the way described earlier, for safety. The switch S should be of fully insulated type.

Sensitive Touch Switch

In Fig. 24, finger contact with the touch pad switches on a low voltage bulb. This circuit is suitable for a battery operated portable bedside light, and for some other devices.

Normally the silicon controlled rectifier is not conducting, so the lamp is extinguished. Finger contact with the pad provides positive base bias for the transistor so that it conducts. This moves the transistor emitter and SCR gate positive, triggering the SCR. Current then flows from cathode K to anode A, and through the bulb. The SCR remains in the conducting condition until the battery circuit is interrupted by opening the switch. A push switch, normally closed, is most suitable here. Pressing it then switches off the bulb.

This touch pad is easily made from a piece of circuit board with foil strips. It can be about 2 x 1 inch, or of any convenient size. Join together alternate foils – e.g., 1, 3, 5 and 7 – by soldering a connecting wire through the holes. This will provide one connection to the pad. The other foils – 2, 4 and 6 – are similarly joined together for the other pad connection.

The pad is fixed to the top or side of a box, which can contain the battery and electronic components.

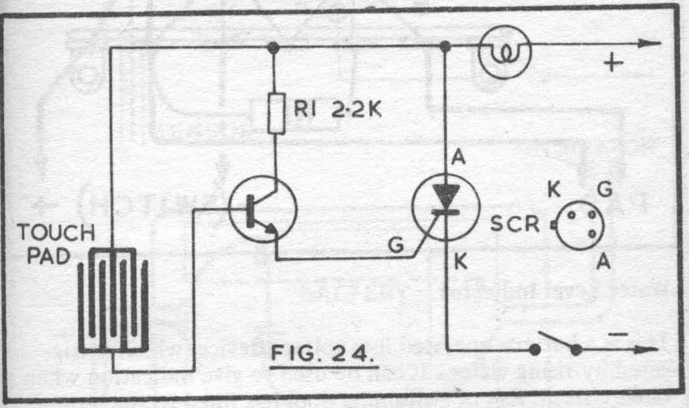

FIG. 24.

The sensitivity of the SCR, or gate current or potential needed to produce conduction, will depend somewhat on the individual device, while the gain of individual transistors will also be different. Numerous NPN types are suitable. If it is found that the switch is too sensitive, this can be corrected by placing a resistor from transistor base to SCR cathode or negative line. This resistor can be about 470k to 1 megohm or so. Excessive sensitivity will be apparent if the circuit is switched on by dust, damp air, or similar effects. The pad should be kept clean by wiping it with a dry cloth.

Fig. 24B shows assembly of the components on a tag strip. A small 50v 1A SCR is more than adequate. The bulb can be a screw in type, and can be 6.3v 0.15A, or 6.3v 0.3A. A 6v 3W (0.5A) bulb is also suitable, provided the battery is fairly large.

It is convenient to make a wooden box which will take the battery, and leave a little space for the tag strip assembly.

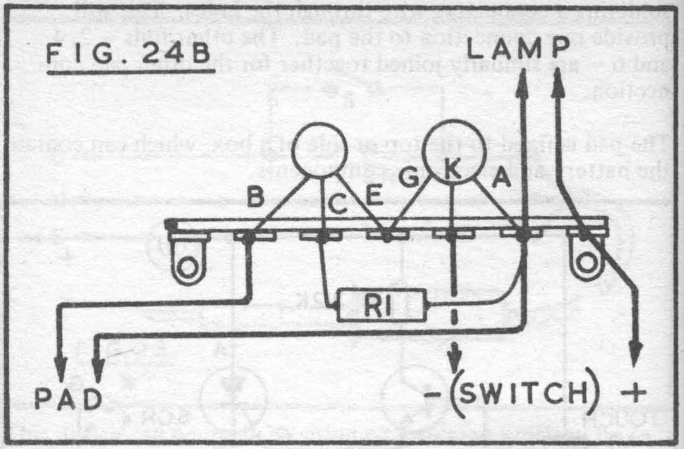
FIG. 24B

Water Level Indicator

This is a battery operated low voltage device, which is triggered by rising water. It can be used to give indication when a tank, cistern, fish or swimming pool has filled to the required level.

Fig. 25 shows the circuit, and also assembly on an insulated board. This also carries the battery, and it is fitted in a metal or plastic box.

In use, the base normally has no positive bias, as the sensor passes no current. When water reaches the sensor, it conducts, and the transistor passes collector current, to operate the indicator. R1 limits base current.

This circuit will operate satisfactorily with a wide range of PNP or NPN transistors, but battery polarity must be reversed for the PNP types.

The indicating or warning device can be one of the multivibrator or other audio oscillators shown earlier. These will give an audible warning, and will draw only a small current, so that a large transistor is not required.

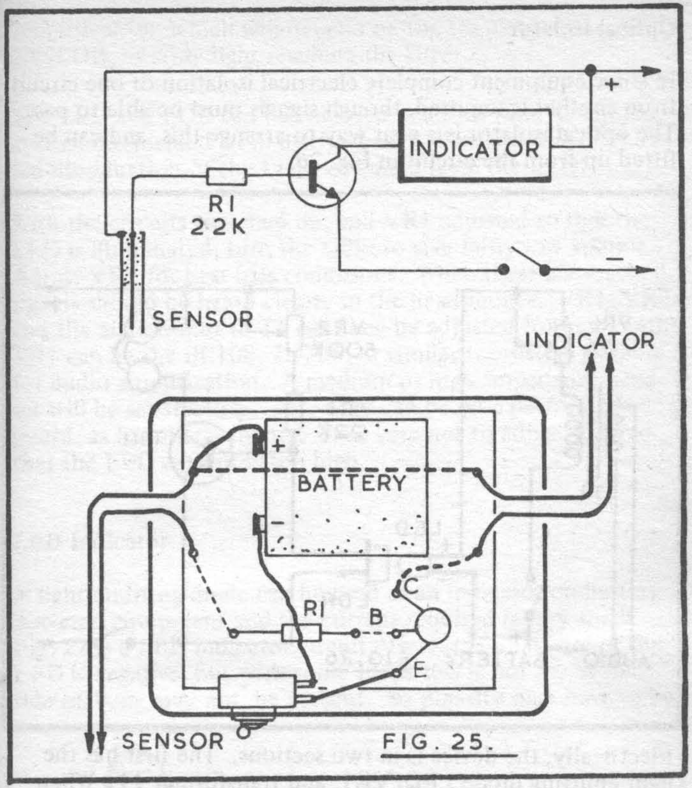

FIG. 25.

For indoor use or in a location protected from rain, the sensor can be two bare wires, about ¼ inch apart, bolted to a strip of insulated material. Where rain could operate the device, it needs to be covered. Alternatively, it can be made from a piece of metal tube, with a central wire, as in Fig. 25. The wire is bound with insulating tape for about half its length, to fit tightly in the tube. Flexible leads are soldered on, and the joints covered with tape. The sensor can be suspended inside the tank or pool, with its bottom at the required water level.

If the transistor is a high gain one, so that damp operates the indicator, sensitivity can be reduced by connecting a 100k resistor from base to emitter.

Optical Isolator

In some equipment complete electrical isolation of one circuit from another is required, though signals must be able to pass. The optical isolator is a neat way to arrange this, and can be fitted up from the circuit in Fig. 26.

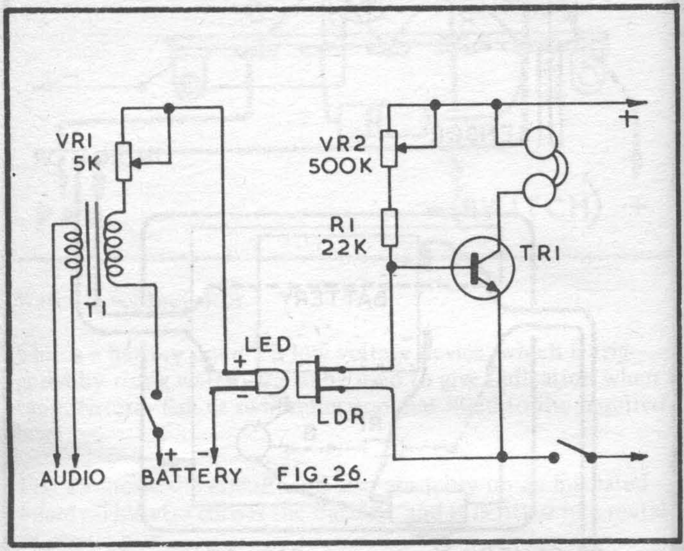

FIG.26.

Electrically, the device is in two sections. The first has the light emitting diode LED, VR1, and transformer T1. When the battery is switched on, current flowing through the LED is adjusted by VR1, so that the LED glows dimly. A signal voltage in the secondary of T1 then modulates the LED light intensity.

To operate this circuit from a radio having a 2/3 ohm or similar speaker, a speaker matching transformer with a ratio of about 45:1 can be used. (The actual ratio is not very important.) Connect the usual transformer secondary to the radio speaker, so that T1 provides a step up effect.

The second part of the equipment has the light dependent resistor LDR, forming part of the network providing TR1 with base bias. VR2 allows adjustment of conditions for best amplification, which will depend on the transistor, and also the LDR, or stray light reaching the latter.

The LED and LDR can be enclosed in a card tube, or should at least be shielded from light, which will swamp the modulated illumination of the LDR, obtained from the LED.

With the circuits switched on, and VR1 adjusted so that the LED is illuminated, turn the radio to give fairly low volume. Adjust VR2 for best bias conditions. When these are reached, signals should be heard clearly in the headphones. VR1, VR2 and the audio input to T1 can then be adjusted for best results. TR1 can be the BC108, 2N706, or similar transistors suitable for audio amplification. A medium or high impedance headset will be satisfactory. Assembly can be on a perforated board, as for other circuits. Take care not to adjust VR1 so that the LED current is too high.

LED Indicator

A light emitting diode can be used as an indicator on battery powered equipment and the current required is very small. Fig. 27 is a LED indicator circuit. Here, the long wire of the LED is negative, but with some LEDs this is not so. A flat side may, or may not, be present. So polarity may have to be

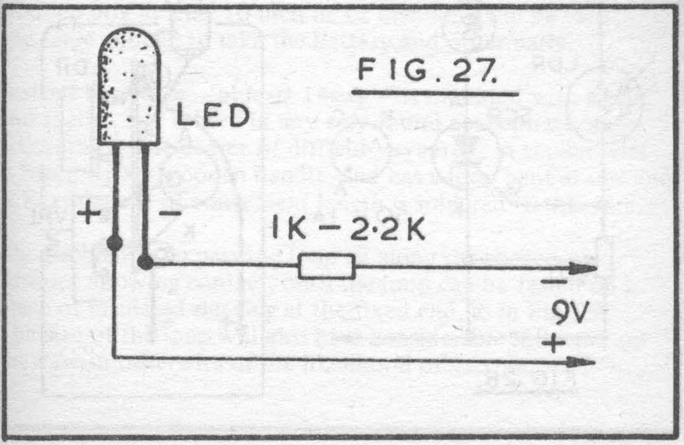

FIG. 27.

found by trial – the correct connections are those allowing the LED to light. The series resistor may be about 150 ohm to 1k or more, depending on the actual voltage, and brightness required. Current drawn varies with different LEDs, and with voltage and series resistance, but can be about 5mA or less, without reducing the brightness too far.

The LED indicator with resistor is connected to the circuit where voltage is present when the equipment is on. LEDs are available in red, yellow, green and other colours, with grommets or clips for panel mounting. Different circuits may be identified by a range of colours.

"Magic Candle"

The triggering of a silicon controlled rectifier can be used for slave photoflash and similar purposes. In Fig. 28 it is employed to switch on a light when a match or lighter is brought near. Sensitivity is best in somewhat subdued light, and it will then be found that striking a match at some distance can bring on the light.

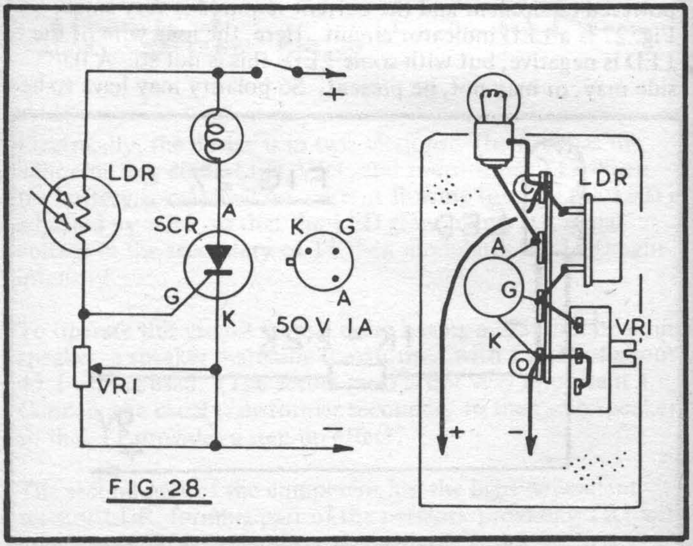

FIG.28.

The LDR has a high resistance when in darkness or subdued light, so that the gate potential of the SCR is not sufficiently high to trigger this device. When light reaches the LDR, its resistance falls, and the SCR gate voltage rises. The SCR then moves into its conducting condition, so that current flows through the bulb. The sensitivity of the circuit, or level of illumination needed to trigger the SCR, is set by VR1.

Components can be assembled on a small tag strip. This, with a battery, can occupy a tubular case, with the lamp on top. An opening is necessary opposite the LDR, and a hole is also made so that VR1 can be adjusted with a screwdriver. Here, a miniature pre-set 50k or 100k linear potentiometer is most suitable. It is necessary to have a small push switch to interrupt the battery current, as the SCR remains in the conducting condition until the circuit is momentarily broken. A 6V battery and bulb can be used. SCRs may fail to operate at all if attempts are made to use them with extremely low voltages, such as a single 1.5v cell.

"Nerve Tester"

This device tests the nerve, concentration, and steadiness of hand, and always provides amusement.

Connections and construction will be seen from Fig. 29. A wooden box at least 10 inch or 12 inch long will be used, and large enough to take the battery and other parts.

A stout bare wire – at least 14swg – is mounted with a bolt and spacer, and is bent in any way found appropriate, and according to the degree of difficulty wanted. A similar wire is fitted with a wooden handle, and has a loop bent at one end. A flexible lead of convenient length is soldered to this wire.

The problem is to pass the loop all along the shaped wire, without allowing contact, until the loop can be rested on a piece of insulated sleeving at the fixed end, as in Fig. 29. The size of the loop will also have considerable influence on the ease or otherwise of the likelihood of success.

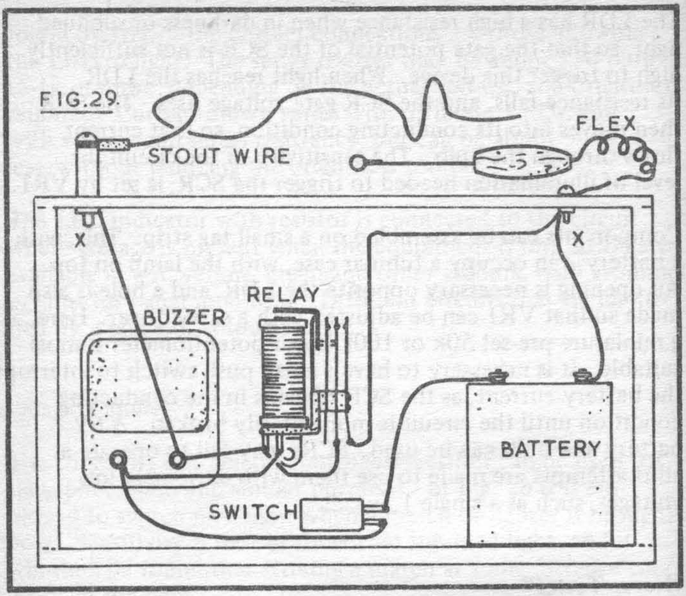

FIG.29.

If loop and wire touch, current flows from the battery to the buzzer, which sounds. Current also flows to the relay winding, so that the relay contacts close. Even a momentary contact between loop and wire will thus cause the buzzer to sound until the switch is used to break the circuit.

Any 4.5v, 6v or similar battery and buzzer will be satisfactory, with a relay which operates on a similar voltage. Connections are to the relay contacts which remain open until the relay is energised.

In use, the device can stand upright. A back is attached with small screws.

Burglar Alarm

The same circuit arrangement as in Fig. 29 will provide a burglar alarm operated by pressure mats. The latter are available in various sizes, and consist of a thin pad which can be placed under a carpet or ordinary mat.

Leads from the mat or mats will run to the terminals at X–X. The buzzer should be changed for a bell, which need not be fitted in the box, but can be situated where it is best heard. Momentary weight on any of the mats will then switch on the bell, which will be locked on by the relay as described, and will ring until the unit is switched off.

Transistor Tester

A very useful check of transistors can be made with the aid of a multi-range test meter, with a selection of base resistors to provide base current, as in Fig. 30. This method does not give an exact indication of transistor gain, but it does allow defective transistors to be eliminated, and provides a good idea of the performance to be obtained from surplus or other unknown transistors.

A 6-way switch allows the base to be open circuit, or to receive current through one of the resistors R1 to R5. These values are R1 330k, R2 110k, R3 33k, R4 11k and R5 3.3k. After allowing for the base-emitter potential, switch positions and corresponding base currents are near the following: 1, zero; 2, 10μA; 3, 30μA; 4, 100μA; 5, 0.3mA, and 6, 1mA.

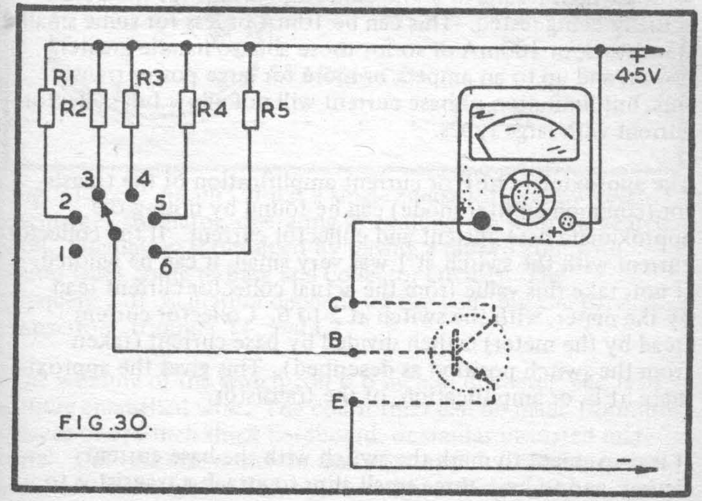

FIG.30.

Polarity is shown for NPN transistors. Reverse polarity of both meter and battery for PNP transistors.

For all tests, initially set the switch to 1, and the meter to its highest range (say 1 ampere). Afterwards, switch the meter to a lower range, after adjustment of the 6-way switch, as may prove to be necessary to obtain an easy reading.

With the switch at 1, a high reading on the meter shows that the transistor has a short circuit from emitter to collector, and is defective.

With germanium transistors, collector current (with the switch at 1) may be 0.1mA or so up to perhaps 0.5mA for small types, and up to some milliamperes for larger types. This small leakage current does not show that the transistor is defective.

In the case of silicon transistors, the leakage current (with switch at 1) will be zero or extremely small, if the device is in good condition.

The switch is then turned to one of the positions 2 to 6. If collector current does not rise, the transistor is defective. If current rises at each step, do not go beyond the switch setting which results in a safe collector current for the device actually being tested. This can be 10mA or less for some small transistors, or 100mA or so for those able to handle greater power, and up to an ampere or more for large power transistors, but limitation of base current will not allow full collector current with large types.

The approximate hFE or current amplification of the transistor (common emitter mode) can be found by noting the approximate base current and collector current. If the collector current with the switch at 1 was very small, it can be ignored. If not, take this value from the actual collector current read by the meter, with the switch at 2 to 6. Collector current (read by the meter) is then divided by base current (taken from the switch position as described). This gives the approximate hFE, or amplification, of the transistor.

It is convenient to mark the switch with the base current figures, and to have three small clips to attach a transistor to the circuit. Switch, clips, and battery can all be fitted to a small box. A development of the circuit would be to include a current meter, with switch and shunts, to replace the external general purpose test instrument.

Treasure Locator

This locator is among the simplest that can be constructed, as it uses an ordinary small portable radio as part of the detection circuitry. No modifications or connections to the radio are needed.

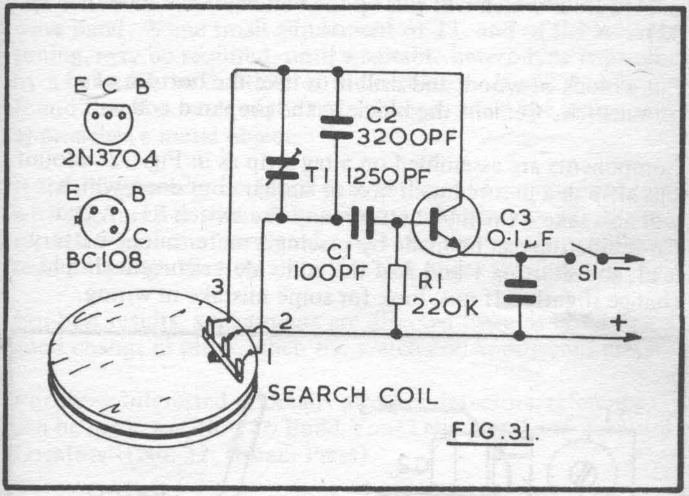

FIG.31.

Fig. 31 is the circuit. It has a single transistor as a low radio-frequency oscillator, operating in conjunction with the search coil. Oscillation is obtained by having a tapped coil, with outer leads going to base and collector circuits. The frequency of oscillation depends on the coil, and the parallel capacitors, trimmer T1, and C2.

The winding of the search coil is 6 inch in diameter, and is of 34swg enamelled wire. The coil former can be made from three pieces of 1/8 inch thick hardboard, or similar insulated material. Cut one disc 6 inch in diameter, and two others about 6½ inch in diameter. Smooth the edges with glasspaper, and cement them securely together, with the small disc between the larger discs. This provides a former with a channel to take the winding. The adhesive should be allowed to dry with the discs under pressure, so that there is no gap into which turns can fall. The former should also be varnished, and allowed to dry, before winding.

The wire is passed up through a small hole, this being point 1. Wind on fifteen turns. Make a loop and pass this up through a hole, for point 2. Wind a further twenty-five turns in the same direction, finishing off at point 3. These leads are scraped and soldered to a small tag strip. This is an anchor point for insulated leads which will run up the handle of the locator (Fig. 32).

Cut a block of wood, and drill it to take the bottom of a broomstick. Cement the block to the prepared coil.

Components are assembled on a tag strip as in Fig. 32. Mount this strip in a plastic lunch box or similar container, which will also take a small 9v battery, and the switch S1. A test for oscillation can be made by placing a meter in one battery lead, and shorting 1 and 3 of the coil. Meter current should change slightly. If not, look for some mistake in wiring.

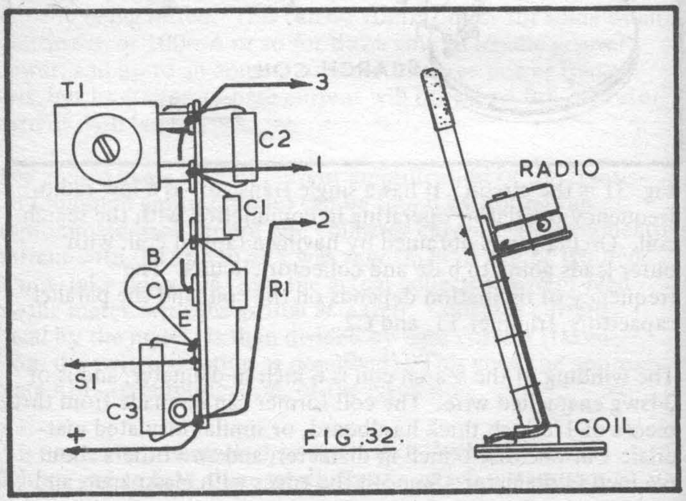

FIG.32.

A shelf is bolted to the handle, and the detector rests on this, with connections passing down to the search coil. A few loops of tape or string will secure these to the handle.

Operation in Great Britain can be on frequencies under 100kHz. Initially tune the radio receiver to 200kHz on long wave, and place it near the coil. Adjust T1 until a strong whistle is heard in the receiver speaker. This is the 2nd harmonic of the oscillator. T1 is then screwed down very slightly from this position.

The small portable radio is placed on top of the locator box, and can be held here with elastic. It is tuned to the medium wave band. Some small adjustment of T1, and of the receiver tuning, may be required, until a suitable heterodyne is produced by a harmonic or multiple of the oscillator. It will then be found that the pitch of the note changes when the search coil approaches a metal object.

It should generally be found that sufficient coupling is obtained with the receiver when they are situated as in Fig. 32. If not, run a few inches of insulated wire from the transistor collector tag up through the box and place this under the receiver.

For best results, adjustments are directed towards obtaining a good change in pitch, when the search coil approaches metal.

For those interested in details of other detectors, reference can be made to "How to Build Your Own Metal and Treasure Locators" (No. 32, Babani Press).

SECTION 3
RADIO AND AUDIO FREQUENCY

MW Booster

This unit can be employed in conjunction with any ordinary portable having a ferrite rod aerial, and will provide a considerable increase in range and sensitivity to weak signals. It has two functions – to allow the use of an external aerial and earth, and to amplify and couple signals to the portable receiver.

Tuning coverage is for medium waves. That is, about 200–550 metres, or 1500–550kHz. Many overseas and other stations can be picked up on this band, and distant reception can be particularly common on the higher frequencies, during the hours of darkness.

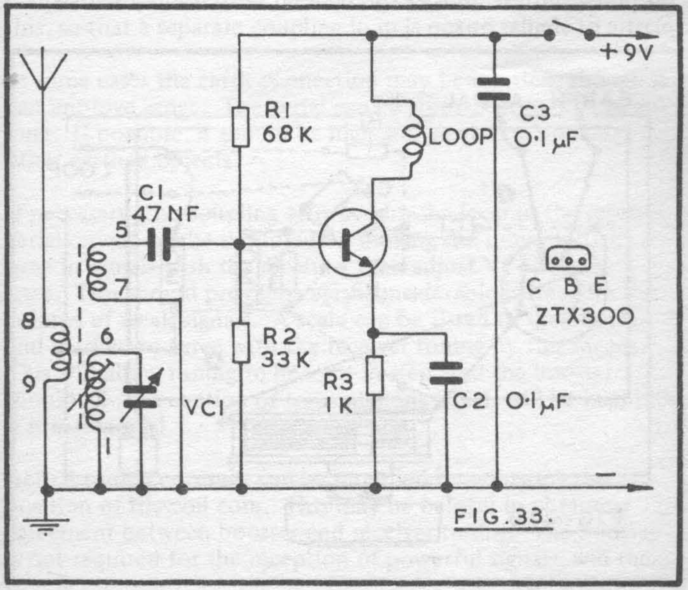

FIG.33.

Fig. 33 is the circuit, and it uses a cored tuning coil with aerial coupling winding 8–9, tuned winding 6–1, and base coupling coil 5–7. These numbers are for the pins of the Denco (Clacton) "Blue" Range 2 component. It will be apparent that other coils, or surplus or home-wound inductors, can be substituted.

For the coil mentioned, VC1 is 315pF. However, this value is not too critical, provided the capacitor fitted allows tuning the wanted frequencies. A solid dielectric capacitor can be used, though the air spaced type is likely to be a little more efficient.

Coupling to the base is by C1, and resistors R1 and R2 set the base bias conditions, with R3 for emitter bias. These values will suit numerous RF transistors.

Fig. 33B shows assembly of the components on an insulated board. The coil is mounted by its threaded bush. When the board has been wired, it can be mounted by means of bolts with extra nuts. The case is preferably metal, for screening purposes, but the booster can be used satisfactorily with a plastic or similar box.

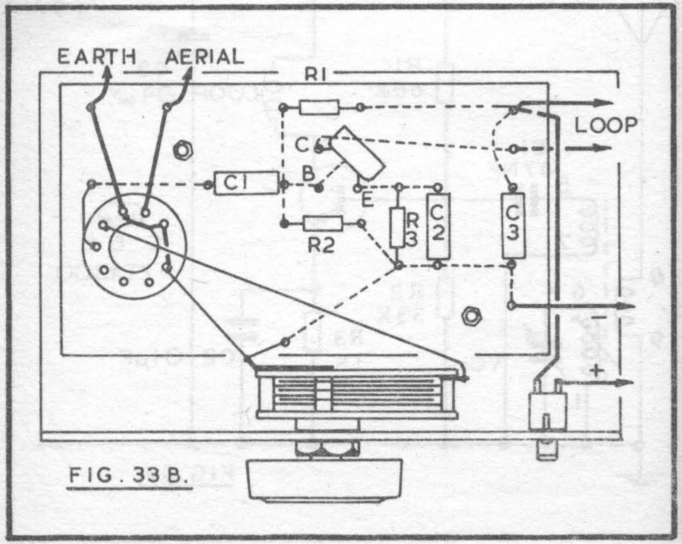

FIG. 33 B.

Fit a small slide switch to the box front, and aerial and earth sockets to the back, near the coil. A 9v battery is accommodated in the case.

Coupling to the transistor receiver is by means of a loop which is placed on the receiver ferrite rod, or near the rod. The loop can be about twenty turns of thin insulated wire, about ½ inch in diameter, and bound with cotton to keep the turns together The ends can be left long enough to reach the two connecting points in Fig. 33B.

If the case is metal, the booster and receiver can be near each other, so the loop leads will only need to be a few inches long. But when an insulated case is used, coupling from the receiver ferrite aerial or other circuits, to the booster coil, may cause instability. This will become apparent as whistles accompanying most signals. If this arises, the booster and receiver should be moved some inches apart. The aerial lead must not run near the loop connections.

Some receivers have an internal winding which is intended to couple an external aerial, and it may be possible to connect to this, so that a separate coupling loop is not required.

In some cases the earth connection may be omitted, though it can improve range. The aerial can be some 20ft to 40ft or so long. If possible, it should be high and clear of walls and other earthed objects.

If necessary, the coupling afforded by the loop to the receiver aerial circuit can be modified by moving the loop. Initially, tune in signals with the receiver, then adjust VC1 for resonance. This should provide a very considerable increase in the volume of weak signals. A scale can be fitted to the booster, and marked to agree with the receiver tuning by this means. This will allow tuning to be kept in step, and the booster should allow reception of transmissions which will be inaudible if it is removed.

Actual tuning coverage can be modified by adjusting the position of the coil core. This may be helpful in obtaining agreement between booster and receiver tuning. The booster is not required for the reception of powerful signals, and the receiver automatic gain control circuit will tend to mask any increase in signal strength obtained with such transmissions.

Midget MW Diode Radio/Tuner

A small, compact tuner can be of utility in allowing radio reception through an amplifier, and a diode receiver can give excellent quality of reception. Such a unit can also be employed alone, with headphones, for personal listening.

Fig. 34 is a receiver/tuner which can be built in a very small plastic or other insulated box. VC1 is a compression type trimmer, fitted with a small knob for adjustment by hand. To do this, it is necessary to remove the short screw (generally 6ba) and replace this with a longer one, which projects to take the knob. The latter can be a terminal head, or the cap from an empty tube of adhesive, secured with lock nuts.

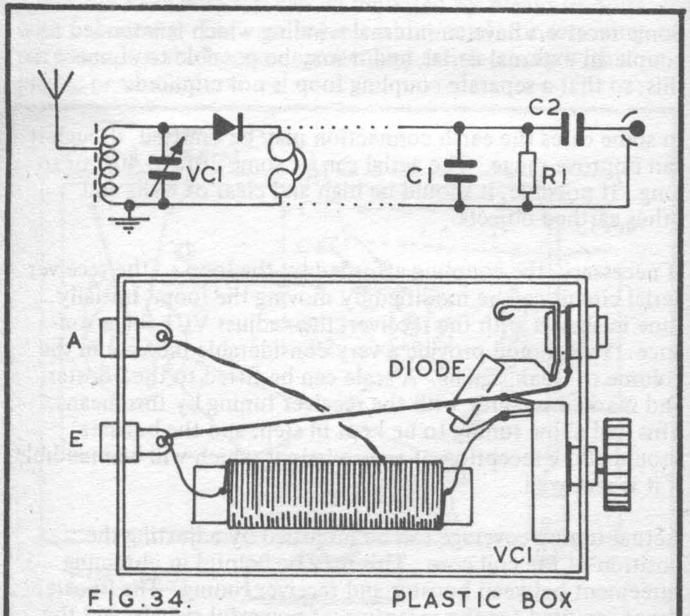

FIG.34.

For best efficiency and small size, the tuning coil is wound on a piece of ferrite rod. It consists of seventy turns of 24swg or similar enamelled wire, on a rod about 1½ inch long and 3/8 inch in diameter. Actual details will not be critical. A piece can be obtained from a longer rod by cutting a notch round the rod with a file and snapping at this point with the fingers. A few touches of adhesive will hold the ends of the winding secure. The loop for the aerial connection is about one-third the total number of turns from the diode end of the coil. (Moving this tapping towards the earthed end of the coil will increase selectivity or sharpness of tuning, but reduce volume.)

Aerial and earth sockets, and a 3.5mm jack outlet, complete the receiver. Headphones can be plugged directly in here.

The aerial and earth can be provided as explained for the diode radio in Fig. 7.

To use the radio as a tuner, C1, R1 and C2 may be added. Suitable values are C1 470pF, R1 220k, and C2 0.25μF. A screened lead with jack plug should be made up to connect the tuner to the amplifier input socket.

One Transistor VHF Receiver/Converter

The super-regenerative receiver is well known for its great sensitivity, which equals that of a multi-stage superhet. It has no alignment difficulties, and few components, so is an interesting project. Its main disadvantages are its background hiss, but this ceases when a signal is tuned in, and its lack of selectivity. The latter is however relatively unimportant on the VHF bands. The circuit here is particularly intended for 2m reception, but is easily modified. It is not intended for FM signals.

In Fig. 35 the detector is a VHF FET, and the field effect transistor will be found very satisfactory in circuits of this kind. T1 is the pre-set tuning capacitor, which allows adjustment over a fairly wide band of frequencies. Tuning is by VC1, which does not require a reduction drive.

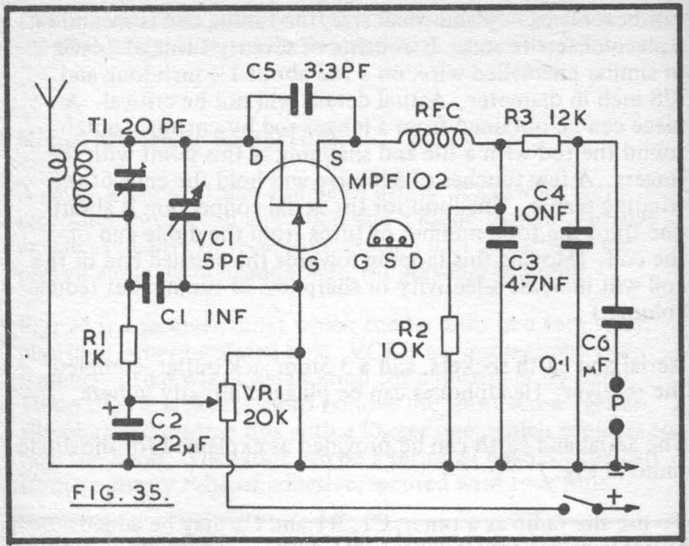

FIG. 35.

Control of super-regeneration is by means of the potentiometer VR1, which adjusts the FET drain potential. The value of VR1 is not very important, but a linear potentiometer is necessary.

C5 provides feedback from source to drain, for regeneration. R3 and the associated components form the quench network, with C6 feeding headphones, or an audio amplifier.

For best results, VR1 should enable the transistor to go smoothly into super-regeneration, which is shown by the hiss mentioned. Achieving this depends somewhat on the transistor, value of C5, adjustment of T1, and aerial loading. However, matters are not too critical.

Transistors unsuitable for VHF cannot be used. Individual FETs of the same type will also differ somewhat, so that in some cases it might be necessary to omit C5, or change the value. Alternatively, a small pre-set (not over 10pF) may be fitted. It will be found that a very small capacitance is required.

Another cause of failing to obtain super-regeneration is too tight aerial coupling, and this is easily corrected by moving the aerial coupling winding a little further from the larger winding.

The RF choke can be a VHF type, or can consist of seventy five turns of 36swg enamelled wire, side by side on a ¼ inch diameter insulated rod. Touches of adhesive may hold the end turns only.

Short leads are essential, and assembly on a tag board is shown in Fig. 35B. The variable capacitor is fitted first, and it can later attach the finished board to a panel, by means of an extra nut.

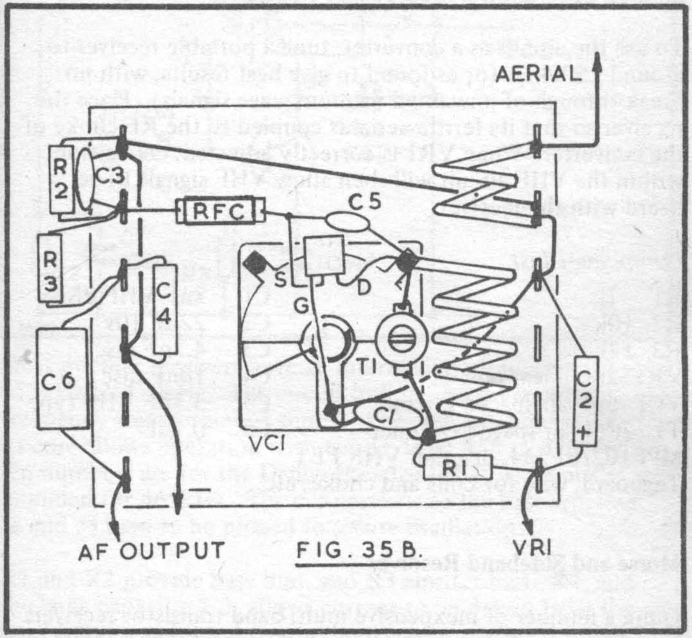

FIG. 35B.

The larger coil has five turns of 18swg wire, and is ½ inch in diameter and ¾ inch long. The smaller coil has two turns. Wiring to VC1, T1, C5, the coil and C1 must be as short as can be arranged. Connections elsewhere are reasonably direct. A ceramic disc capacitor may be used at C1. T1 should be a low loss trimmer.

The board can be fitted in a case which will carry VR1, and large enough for the battery and on-off switch. The case can be used to support a telescopic aerial, some 30 inch or so when extended.

Frequencies tuned by VC1 can be raised by unscrewing T1, or by stretching the larger coil to separate the turns more, or by using a coil with fewer turns, or of smaller diameter. The receiver is most suitable for use over the 70MHz to 200MHz range.

For the single transistor alone, a sensitive pair of medium or high resistance phones will be needed. Audio ouptut from C6 can be taken to an amplifier, for more powerful reception with phones, or to allow a loudspeaker to be used.

To use the circuit as a converter, tune a portable receiver to around 1500kHz (or as found to give best results, with no break through of unwanted medium wave signals). Place the receiver so that its ferrite aerial is coupled to the RF choke of the converter. When VR1 is correctly adjusted, conversion within the VHF circuit will then allow VHF signals to be heard with the receiver.

Component List

R1	1k	C1	1nF VHF disc
R2	10k	C2	22μF 10v
R3	12k	C3	4.7nF disc
VR1	20k linear pot.	C4	10nF disc
VC1	5pF air-spaced variable	C5	3.3pF VHF type
T1	20pF air spaced trimmer	C5	0.1μF

MPF102/BF244, or other VHF FET
Tagboard, wire for coils and choke, etc.

Morse and Sideband Resolver

Quite a number of inexpensive multi-band transistor receivers have frequency coverage which includes one or more Amateur bands. That is, 1.8, 3.5, 7, 14, 21, or 28MHz. On these bands transmissions are often in Morse code or CW, or single sideband. CW or SSB can be resolved by the specialised Amateur receivers and transceivers, but not with the general purpose home receiver. To receive CW or SSB with the latter type of radio, it is necessary to add an oscillator. This either generates a heterodyne or beat note for code reception, or replaces the carrier which was eliminated from the single sideband transmission.

The circuit in Fig. 36 may be used for these purposes. It generates a carrier of about the same frequency as that used in the intermediate frequency stages of the receiver, exact adjustment of frequency being by VC1,

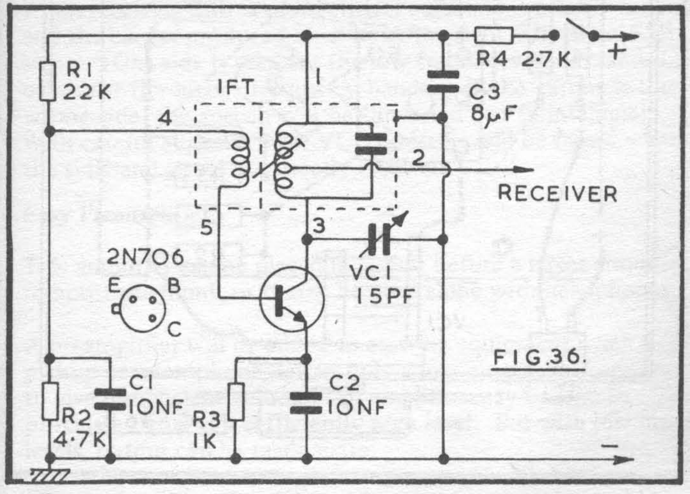

FIG.36.

Most popular receivers have an intermediate frequency of about 455kHz to 470kHz. The circuit in Fig. 36 uses an intermediate frequency transformer of similar type, and adjustment of its core allows operation with any IF from 455 to 470kHz. Pin numbers are for the Denco (Clacton) IFT13, normally intended for 465kHz. The connections to the base winding (4 and 5) have to be phased to secure oscillation.

R1 and R2 provide base bias, and R3 emitter bias. R4, and the large capacitor C3, allows current to be drawn from the same battery as is used to run the receiver, if wished. Alternatively, a separate 9v battery may be employed here.

Fig. 36B shows the practical layout of components, and wiring to them, using a perforated board. Connections and joints are kept close against the underside of the board. When wiring is finished, the board is mounted in a metal box, using ½ inch bolts with extra nuts for clearance.

VC1 can best have a dial with a central zero mark, this showing when the plates are half meshed.

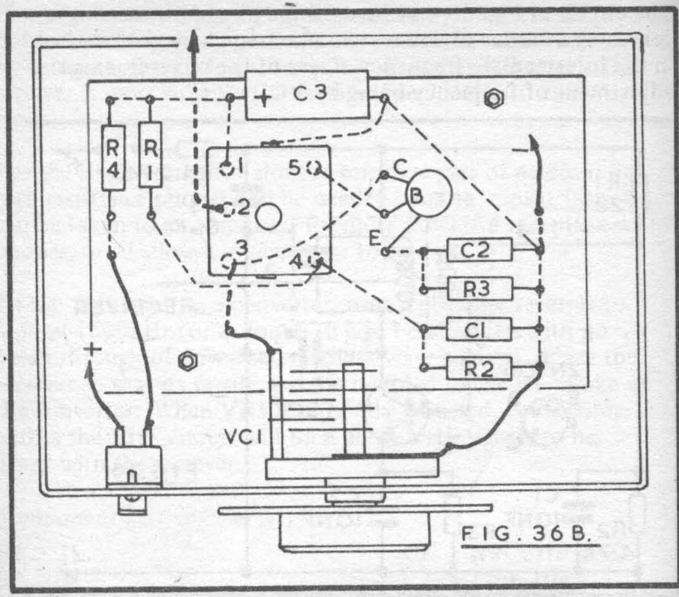

FIG. 36B.

An insulated lead is run from 2 on the IFT, and is placed near the receiver. This lead may be so positioned as to give best coupling into the IF stages of the receiver. If coupling is too weak, the signal from the CW/SSB Resolver will be inadequate. On the other hand, tight coupling will allow the carrier to swamp weak signals tuned in on the receiver. Fortunately the degree of coupling is not at all critical.

To set up the oscillator, first of all tune in any normal broadcast with the receiver. Set VC1 half closed and switch on. The core of the IFT should now be slowly adjusted until a loud heterodyne or beat note is heard accompanying the signal received. Set the core so that this note falls to about zero, or to a low pitch. It will then be possible to rotate VC1 either way from this central zero beat position, in each case producing a note which rises in frequency as VC1 is rotated.

To receive CW or Morse code, it is only necessary to tune in the code signal with the receiver, and then to adjust VC1 to obtain a note of the most suitable pitch. VC1 can be rotated either way from its central or zero beat position, to obtain best freedom from interference.

When receiving SSB, a more critical adjustment of VC1 is needed and the carrier produced must be at the right side of the SSB signal. (One side is used for the low frequency bands, and the other for the higher frequency bands.) If the carrier is the wrong side, the speech will be "inverted" or unintelligible. With careful adjustment of VC1, a setting will be found where the sideband signal is correctly resolved.

Easy Preamplifier

This amplifier can be placed in circuit before a larger amplifier, to boost the input, or it may be used alone with headphones.

A preamplifier will be added to existing equipment when a pickup or microphone having only a low output has been found to give insufficient volume. An amplifier may be able to provide a signal at a sufficiently high level. But with low input levels, output can be inadequate.

The stage in Fig. 37 can provide considerable gain. Audio input is to C1, and R1 provides base current for the transistor. Amplified signals are obtained across R2, and pass to the main amplifier (or phones) via the isolating capacitor C2.

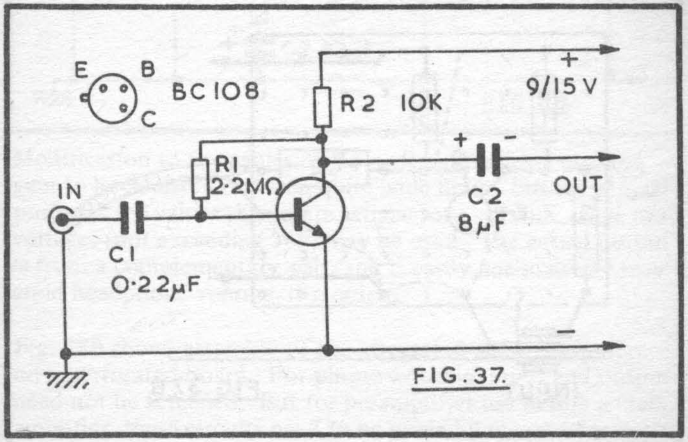

FIG. 37.

With a BC108 or similar transistor, the supply can be about 9v to 15v. Numerous NPN transistors, intended for preamplifier use and of low noise type, will perform well. Current may be from a separate battery. Or it may in many cases be drawn from the main amplifier. To avoid instability or similar troubles, the positive supply should be from a well decoupled and smoothed point in the main amplifier. If this is not available, connect a 220μF capacitor from positive to negative in Fig. 37, and draw the positive supply through a 1k resistor. Negative is returned to the negative line of the main amplifier.

Should the main amplifier have a positive earth line, then the transistor can be a PNP type, R1 and R2 being changed to suit.

Fig. 37B shows assembly on a perforated board. When the circuit provides input to a main amplifier, it should generally be in a metal case. Input is by a screened lead, and a screened lead is also fitted for the audio circuit to the main amplifier. The case should be earthed to the negative line. This can be by means of one of the bolts securing the board.

Screening to avoid hum or instability will generally prove to be unnecessary when using the stage to operate headphones. So two sockets, or a jack socket, will be more convenient for the phone circuit.

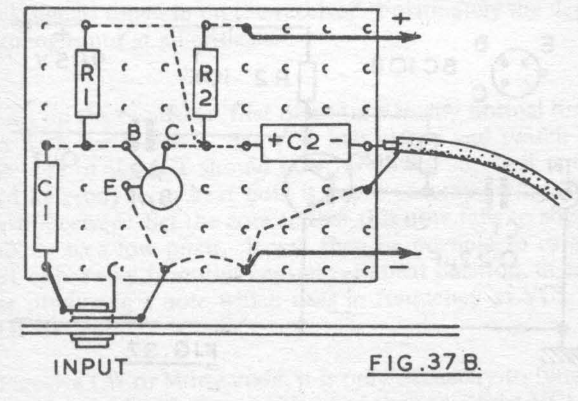

FIG.37B.

741 Preamplifier

Integrated circuits have several stages of amplification, formed as a single chip and enclosed in a small case. The 741 is an easily obtained audio IC which can be used in many circuits. The uA741CE or metal can 741 is convenient, as its leads may be spaced to suit any ready perforated board. The dual in line 741 is electrically similar.

Fig. 38 shows a preamplifier using the metal can 741. Input is via C1 to 3. As shown, 7 and 4 are positive and negative, while output is from 6. Negative feedback is applied to 2, via R4. C3 is an isolating capacitor.

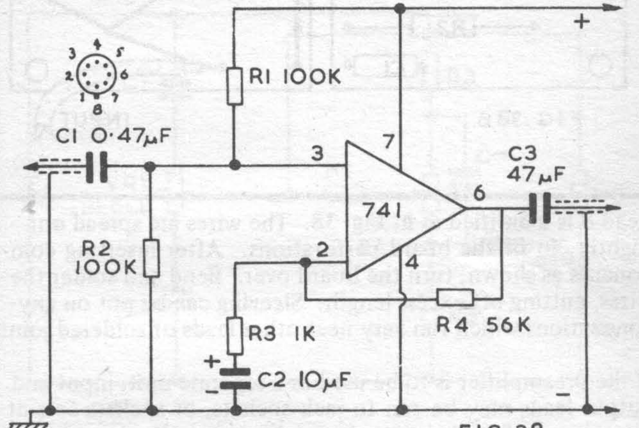

FIG.38.

Modification to the values of R3 or R4 allow feedback and gain to be adjusted between quite wide limits, but for general purposes the values shown are satisfactory. A wide range of voltages (not exceeding 36v) may be used. The actual output is from a complementary pair, and is easily adequate for very good headphone volume, if required.

Fig. 38B shows assembly of the integrated circuit amplifier on a perforated board. For phone working, input and output need not be screened. But for preamplifier use before a main amplifier, these circuits need to be made by means of screened leads.

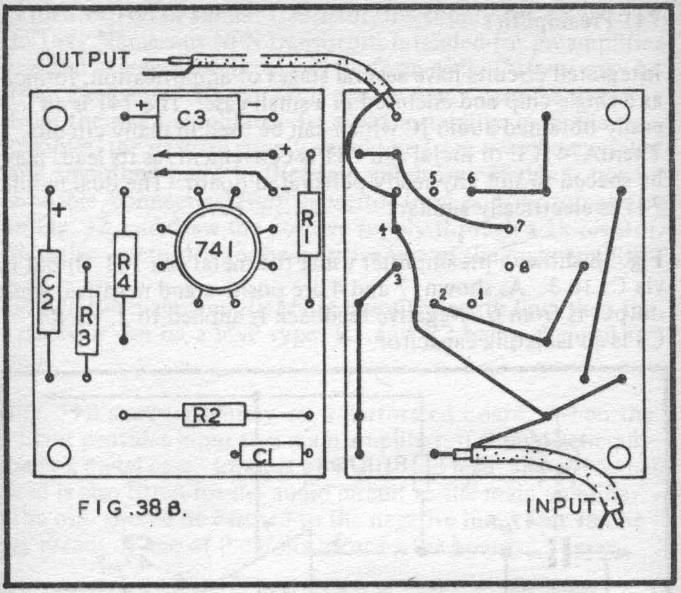

FIG.38B.

Lead 8 is identified as in Fig. 38. The wires are spread out slightly, to fit the board perforations. After inserting components as shown, turn the board over. Bend and solder the wires, cutting off excess length. Sleeving can be put on any connections which run very near other leads or soldered joints.

If the preamplifier is to be used as a separate unit, input and output leads may be run to jack sockets, or sockets to suit other items of equipment. A metal case is advised, as this will help to avoid any pick up of hum, or other troubles due to coupling to the board wiring. The board can be secured to the metal box or case with ½ inch 6ba bolts, and extra nuts to give a little clearance.

The amplifier may be run from its own battery; or from a supply used with a main amplifier. For the latter, it is wise to provide a decoupled point, in a similar manner to that described for the transistor preamplifier. This consists of a quite large capacitor (say 220μF to 470μF) from positive to negative across the preamplifier supply circuit, and a resistor of 1k to 2.2k or so from preamplifier positive to positive on the main amplifier.

Tone Control for Your Discs Etc.

The preamplifiers shown have a substantially flat response. That is, they do not purposely emphasize bass or treble, or have similar effects in modifying the audio frequencies present. In some circumstances it is necessary to boost treble or bass, or reduce the emphasis of these frequencies, in order to obtain the most satisfactory reproduction.

This may be achieved by making a fully adjustable tone control, and Fig. 39 is a suitable circuit. A volume control is also incorporated, as this has not been present in the preamplifiers.

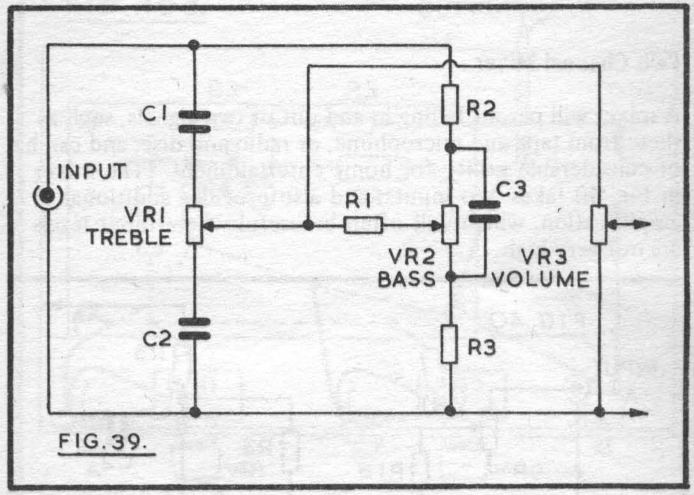

FIG.39.

The tone control is a passive circuit, needing no battery, and it can be placed between the source of audio signals (generally a microphone, pick-up, or radio tuner) and the amplifier or preamplifier. It may also be located between preamplifier and main amplifier, if wished.

Results are substantially flat with the treble control VR1 and base control VR2 in their central positions. Adjustment of VR1 and VR2 will provide separate boost or cut of treble, or boost or cut of bass frequencies. VR3 is the volume control.

The degree of control afforded by circuits of this type may be adjusted by alterations to the component values. In practice, it will be found that there is little point in having extreme levels of adjustment, which will never be required. The following values may be used: C1 and C2, 270pF; R1, 390k; R2 and R3, 100k; C3 680pF; VR1 and VR2, 2 megohm linear potentiometers; VR3, 2 megohm log potentiometer.

Components should be assembled in a metal case, for screening. The case is grounded to the earth line. Input and output sockets can be near the ends of the case, with VR1, VR2 and VR3 between. Components can be soldered directly to the potentiometer tags.

Two Channel Mixer

A mixer will permit fading in and out of two signals, such as these from tape and microphone, or radio and disc, and can be of considerable utility for home entertainment. The mixer in Fig. 40 takes two inputs, and also provides additional amplification, which will often be useful where input levels are not very high.

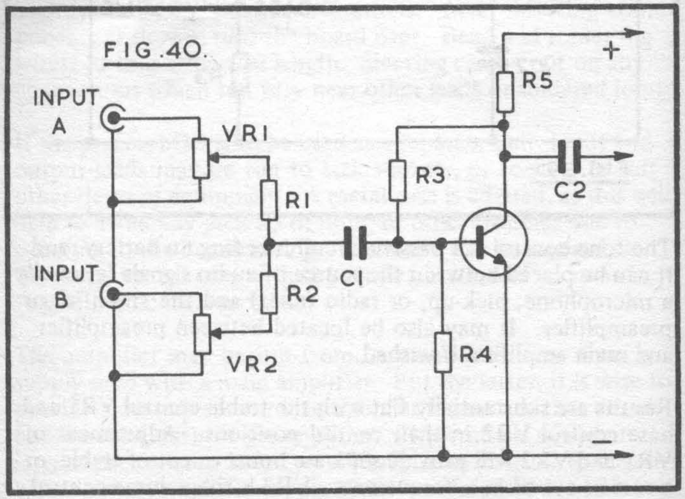
FIG. 40.

Input A has its individual volume control VR1, and Input B is similarly provided with VR2. Resistors R1 and R2 prevent short circuiting one input when the other potentiometer is set at zero. Coupling to the transistor base is by C1. R3 and R4 set base bias conditions, with R5 as the collector load and audio output from C2.

Component values are not very critical, though R3, R4 and R5 need to be selected to suit the transistor fitted. The mixer is connected and used in a similar way to that explained for the preamplifiers.

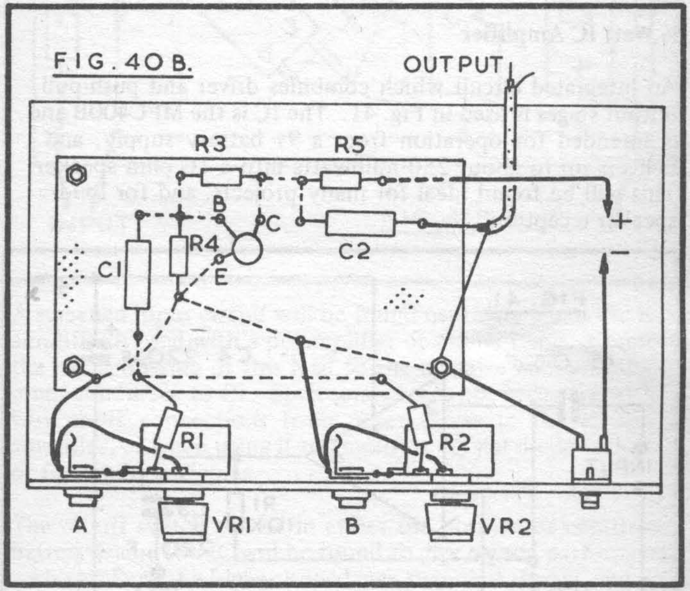
FIG. 40 B.

Fig. 40B is the layout, using a metal case or box. A screened output lead is provided. Input sockets A and B may be chosen to suit the type of plugs on existing equipment. The on-off switch is separate from the potentiometers, so that these can be left at pre-set levels if wished.

Component values may be chosen to suit low or high impedance inputs, if wished. Or the following may be used for general purposes:

Components List:

VR1, VR2, 1 megohm log potentiometers.
R1, R2, 200k
C1 0.5μF
R3 330k
R4 22k
R5 10k

C2 0.5μF
BC109

This is when using a 9v supply.

¼ Watt IC Amplifier

An integrated circuit which combines driver and push-pull output stages is used in Fig. 41. The IC is the MFC400B and is intended for operation from a 9v battery supply, and delivers up to about 250 milliwatts into a 16 ohm speaker. This will be found ideal for many projects, and for loudspeaker reception.

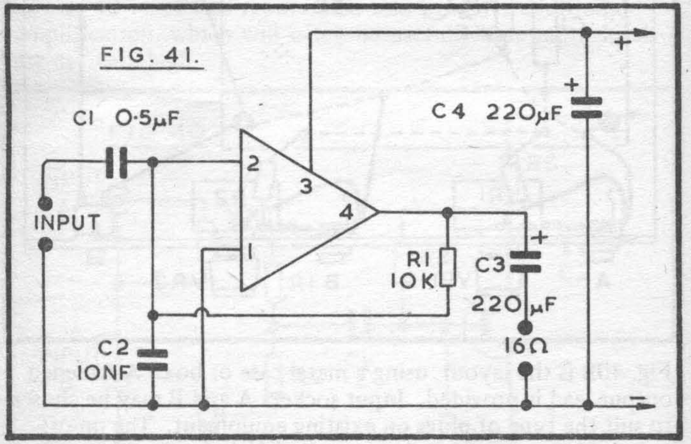

FIG. 41.

Audio input is to 2 of the IC, from the isolating capacitor C1. C2 is a by-pass capacitor to maintain stability. The IC has only four connections, 1 going to the negative line, and 3 to the positive supply. Audio output is from 4, and to the loudspeaker via C3. R1 provides negative feedback and bias to point 2.

The assembly of this amplifier on 0.15 inch matrix board is shown in Fig. 41B. Pin numbering is when looking at the top of the IC and the component side of the board.

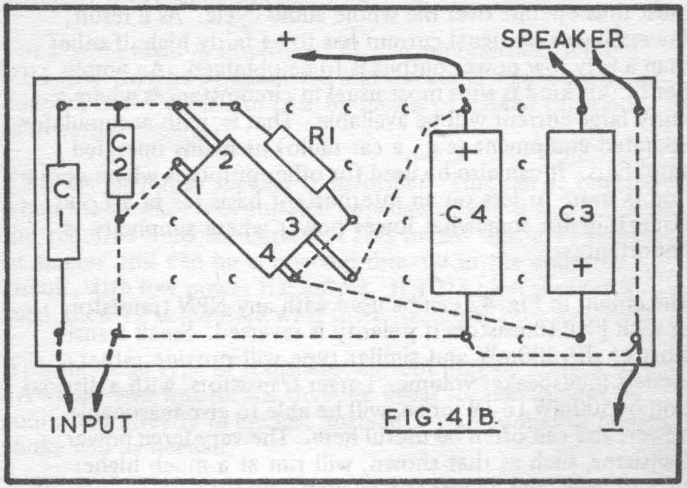

FIG.41B.

A screened input circuit will be found necessary when the IC amplifier is used with a preamplifier or similar items. Connect the outer brading of this lead to the negative line, and the inner conductor to C1. Such screening is not necessary with very short connections from other stages to the IC amplifier, or when using it to amplify a crystal diode radio, or for similar purposes.

The on-off switch can be in either the positive or negative battery lead. The IC will be found to give a good performance with speakers of a higher impedance than that shown, and up to 75 ohm may be used (with some small loss in output). Units of lower impedance than 16 ohms should not be fitted.

The small size of this IC and the few extra components needed make is very suitable for miniature apparatus, probably with a 2½ inch speaker. It will also operate a large speaker well (see "Boost Speaker") and at a volume level adequate for many purposes.

Class A Amplifier

A Class A amplifier uses a single transistor, Fig. 42, which must thus operate over the whole audio cycle. As a result, the resting or no-signal current has to be fairly high, if other than a very low power output is to be obtained. An amplifier of this kind is thus most usual in circumstances where a fairly large current will be available. That is, with accumulator operated equipment (e.g., a car radio) or mains operated amplifiers. It can also be used for other purposes where operation is more or less on an intermittent basis (as in a Loud Hailer), or for somewhat lower power where simplicity is important.

The circuit in Fig. 42 can be used with any NPN transistor, or with PNP transistors if polarity is reversed. Small transistors of the 300mW and similar type will provide rather modest loudspeaker volume. Larger transistors, with a dissipation of 500mW to 1W or so, will be able to give reasonable power, and can often be useful here. The very large power transistor, such as that shown, will run at a much higher current level, and deliver several watts output.

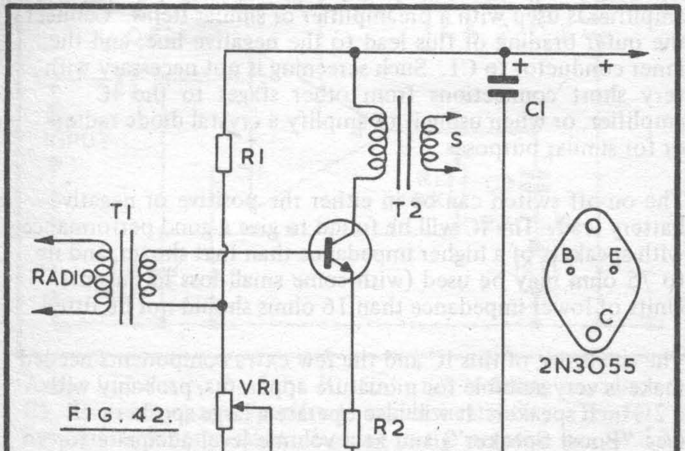

FIG. 42.

T1 is the coupling or driver transformer, which may receive audio signals from a radio receiver, smaller amplifier, or other equipment. A step down ratio of about 5:1 to 10:1 will generally prove to be suitable.

For smaller Class A amplifiers, T1 may be omitted. R1 and VR1 may then be of about 1k, and the audio signal can be coupled to the transistor base by means of a fixed capacitor, as in earlier circuits.

T2 is the output or speaker transformer. This can be omitted when the speaker is of suitable impedance, and is able to carry the required collector current. This means that a 75 ohm or similar unit can be connected directly in the collector circuit, with low power transistors. If a 2/3 ohm speaker is used, then T2 is required with these for impedance matching.

With the large type of transistor, current may be 500mA to 1A or so, and this means that the usual speaker cannot be connected directly in circuit. Instead, a transformer or tapped choke will be needed.

The circuit is often operated from 12v (a vehicle supply) but can be used with 9v, or more than 12v.

R1 and VR1 set bias conditions, and R2 limits peak collector current. For the 2N3055, R2 may be 1 ohm. R1 can be 100 ohm, and VR1 25 ohm. T2 can be 2:1 for a 3.5 ohm speaker. C1 is 3200μF. Begin with VR1 at minimum value, slowly increasing this until reproduction is satisfactory. (This will set collector current at about ½ ampere.)

With other transistors, adjustment can be made in the same way, by slowly increasing the base bias until best results are secured. Raising the value of VR1 in Fig. 42, or reducing the value of R1, will increase collector current, and thus raise the power which can be handled. However, current is limited by the power handling capacity of the transistor, so a meter should be put in one battery lead, and a check made that the power dissipation of the transistor is not exceeded. With small transistors, current can be limited to 50mA or so.

Fig. 42B shows assembly of the circuit in Fig. 42. The power transistor is mounted on an aluminium plate, which acts as a heat sink to carry away heat from the transistor itself. Note that this plate is common to the collector. Base and Emitter pins pass through clearance holes. Smooth off any burr round the fixing and clearance holes, so that the transistor can rest flat on the metal. Where it is awkward to have the plate or case common to the collector, the usual insulation set can be fitted. This consists of bushes for the fixing bolts, and a thin mica washer to go under the transistor, so that electrical isolation is obtained.

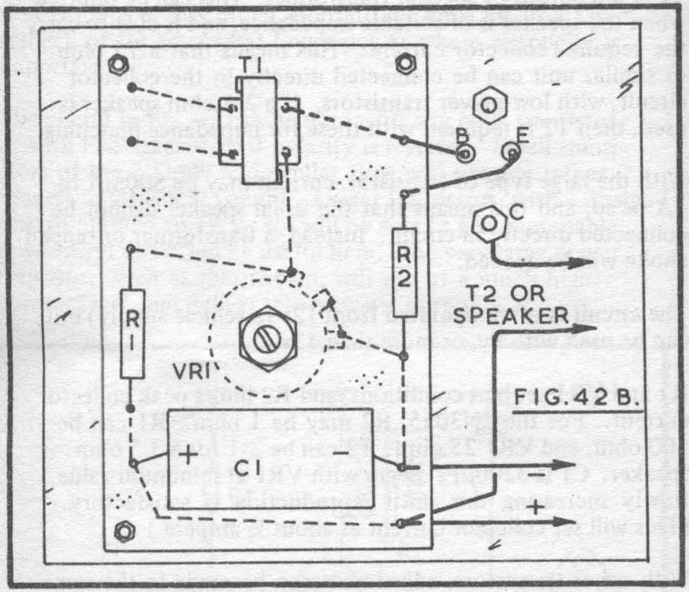

FIG.42 B.

The insulated board carrying the other components is mounted by long bolts with extra nuts. A large Class A stage transistor of this type will grow very warm to the touch, if run at anything like maximum power rating. This does not indicate any defect.

Two Stage Amplifier

The circuit of a 2-stage amplifier is shown in Fig. 43, and this can be used for loudspeaker reproduction. Input to the driver stage is by the isolating capacitor C1. Operating conditions in this stage depend largely on the values of R2, R3, and the emitter bias resistor R5.

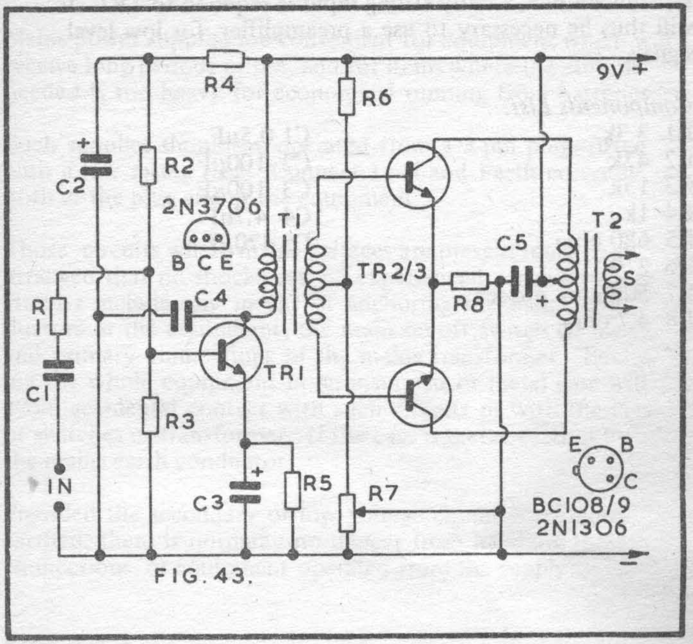

FIG. 43.

T1 is the driver transformer, with centre tapped secondary, so that TR2/3 may be operated in push-pull (see Fig. 12). T2 couples output to a 2/3 ohm speaker.

With the values shown, a BC109 will give very good results in the TR1 position, with 2 x BC108s for the output stage. R7 is set so that battery current is around 10mA, with no signal present. Current rises from 20mA to 60mA or so, according to volume, when the amplifier is driven.

A pair of 2N1306 transistors may be used for the output stage. Other transistors are also possible here, as operating conditions can be adjusted by R7.

The 2N3706 will operate in the TR1 position, though with lower gain than the BC109. Numerous other small audio transistors can also be used here.

For full volume, a fairly strong input is required to TR1. It will thus be necessary to use a preamplifier, for low level signals.

Components List:

R1 3.3k
R2 47k
R3 15k
R4 1k
R5 680 ohm
R6 2.7k
R7 500 ohm pre-set
R8 4.7 ohm

C1 0.5μF
C2 100μF
C3 100μF
C4 4.7nF
C5 470μF

SECTION 4
POWER SUPPLIES

Mains power supplies are convenient for equipment which will receive long periods of use, and for items where the current needed is too heavy for economical running from batteries.

Such supplies should be operated from a 3-pin plug, fitted with a low rating fuse. Connect L, N and Earth correctly, both at the plug and at the equipment.

Those circuits where mains voltages are present must be so arranged that no shocks can be experienced. Mains voltage circuits include any means of anchoring the mains conductors in the equipment, the main on-off switch (if used), and primary connections of the mains transformer. Enclosing the whole equipment in an insulated or metal case will avoid accidental contact with such circuits or with the tags of switches or transformer. If the case is metal, earth it by the mains earth conductor.

Provided the secondary or low voltage circuit is correctly earthed, there is normally no danger from handling these connections, or equipment operated from the supply.

Easy PSU

A useful power supply suitable for many purposes can use the circuit in Fig. 44. Mains L (live) and N (neutral) conductors go to the primary P of the transforer. A switch is placed in the L circuit.

The secondary S can delivery any wanted low voltage, and also give isolation from the mains.

Almost all equipment will need direct current. The four separate rectifiers, or an equivalent single device, produce pulsating DC. The reservoir capacitor C1 smoothes this. Resistor R1, with the second capacitor C2, gives additional smoothing. Output from the supply is then substantially smooth DC, as would be obtained from a battery.

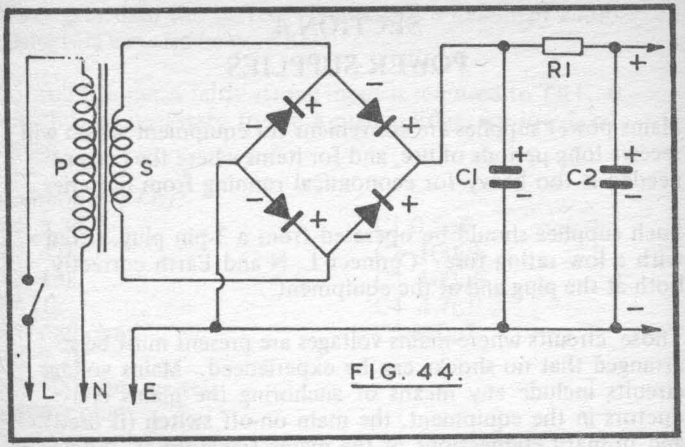

FIG. 44.

Sometimes R1 and C2 are omitted. Should hum accompany reproduction from an amplifier, this shows that smoothing is probably inadequate, and if so the cure is to add R1 and C2.

Any wanted voltage can be obtained by using a transformer with an appropriate secondary, with rectifiers and capacitors to suit.

C1 may charge up to approximately 1.4 times the RMS or usual voltage rating of the secondary S. Thus, if S were a 6.3v winding, about 8.8v would be expected across C1. But should R1 be present, a voltage drop will arise in it, the figure depending on the resistance value of R1, and current flowing.

For these reasons, the actual output may not be exactly the figure expected. This will often be unimportant, and such power supplies are largely used for all sorts of equipment.

For apparatus generally operated from a 9v or similar battery, a bell transformer with 3v, 5v and 8v output tappings is suitable. If R1 is 22 ohm, this will give around 9v at 100mA, with the 8v tap in use.

It is convenient to use 50v or 100v rectifiers, and the rating can be 1 ampere, with a 1 ampere transformer secondary. Four individual wire ended rectifiers can be mounted on a tag strip, and connected as in Fig. 44. A combined rectifier will have AC, positive, and negative tags (see Fig. 45) and can be fitted instead.

The capacitors need to be of large value, and have a voltage rating higher than the actual voltage which will be present. The 12v capacitors could be fitted in the circuit described. C1 can be 2500μF or larger, and C2 can be of similar value.

Easy Regulated PSU

With the circuit in Fig. 44, the exact voltage will depend on the current drawn, rising when current is low. To avoid this, some means of regulation may be employed.

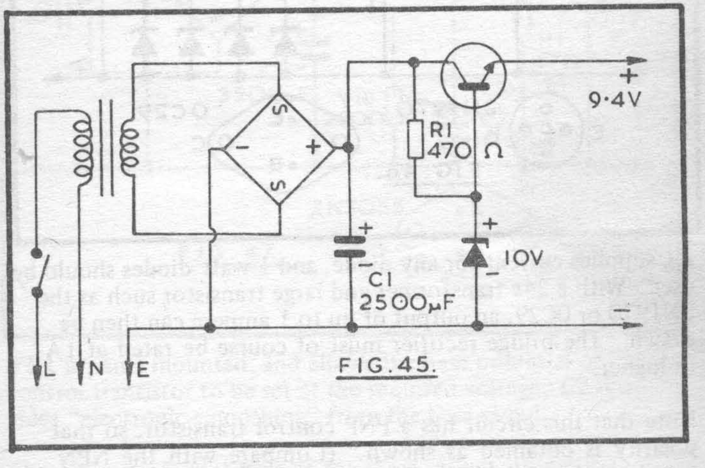

FIG.45.

Fig. 45 shows a typical regulation circuit, giving an output of approximately 9.4v. The output from the rectifier needs to be of somewhat higher voltage than required at the output, and a transformer with a 10v secondary is suitable. Current through R1 supplies the Zener diode, which holds the base of the transistor at virtually the same voltage for all current loads. Some voltage is lost in the transistor, so the output is nearly the same, for any current load. For up to about 200mA, a BFY51 or equivalent transistor is adequate, so this will do well for many purposes.

Multi-Voltage Regulated PSU

Fig. 46 is a somewhat similar circuit, but it has a switch, to select any one of the four Zener diodes. These may be chosen so that 6v, 9v, 12v and 15v will be available.

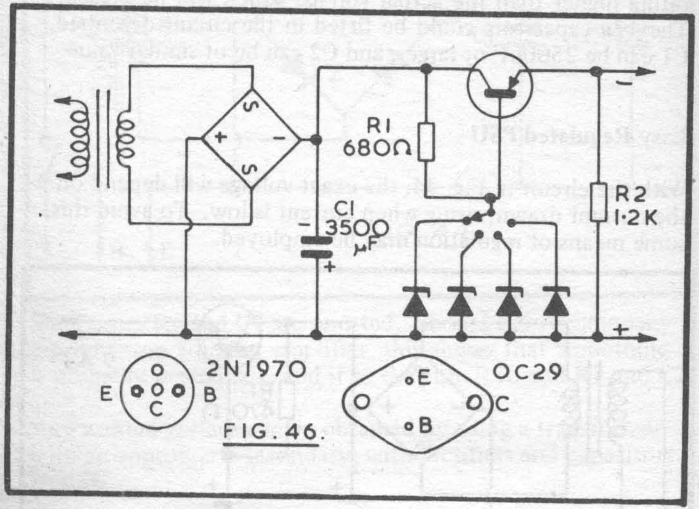

FIG. 46.

R1 supplies current for any diode, and 1 watt diodes should be used. With a 24v transformer and large transistor such as the 2N1970 or 0C29, an output of up to 1 ampere can then be drawn. The bridge rectifier must of course be rated at 1A or higher.

Note that this circuit has a PNP control transistor, so that polarity is obtained as shown. (Compare with the NPN transistor in the previous circuit.) However, either positive or negative can be grounded, and a separate earth terminal can be provided to allow this. A flying lead is placed between earth terminal and positive, or earth and negative, as required. The permanent negative earth connection shown in Fig. 45 must of course then be omitted.

More details of actual construction will be found later.

One Transistor Adjustable PSU

Fig. 47 is the circuit of a simple power supply unit which can be set to provide any output from zero to 12 volts. The transformer has a 17v secondary, rated at 1 ampere or higher. After rectification, DC is obtained from the reservoir capacitor C1.

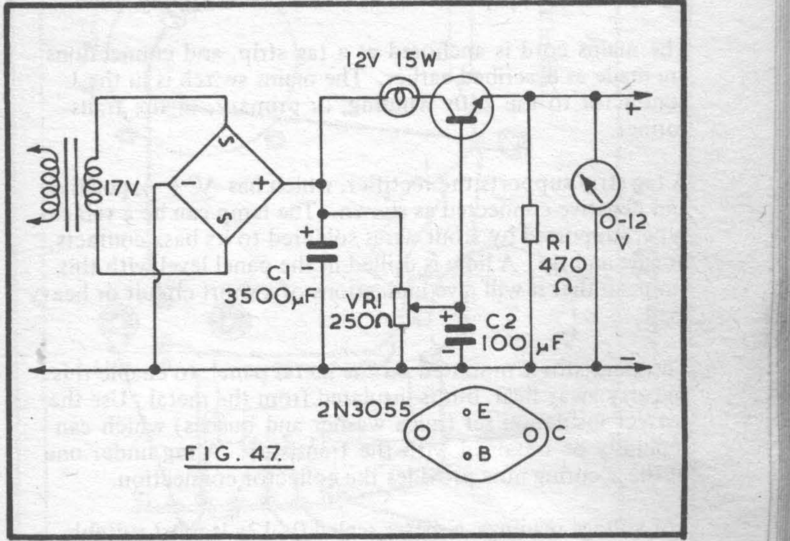

FIG. 47.

VR1 is panel mounted, and allows the base potential of the control transistor to be set at the required voltage. C2 provides "electronic smoothing" from the base circuit.

The 12 volt 15 watt bulb is for short circuit protection. With relatively low currents, little voltage is lost here. But if a short circuit arises in the output circuit, current is limited to a little over 1A by the bulb. This reduces the chances of any damage due to temporary shorts in leads or equipment.

R1 is a bleeder resistor, drawing some current even when no external load is connected. The meter is the usual high resistance type, scaled from zero to 12v. VR1 is adjusted until this instrument shows the wanted voltage.

The regulation of this supply is not so good as that of supplies using a number of transistors. But it will be found suitable for all ordinary purposes, and to supply power for circuits in this book, as only a small change in voltage arises, with changes to the load, or current drawn.

Fig. 47B shows practical construction of this PSU. A similar layout can be adopted for the other power supply units.

The mains cord is anchored at a tag strip, and connections are made as described earlier. The mains switch is in the L conductor to the 240v winding, or primary, of the transformer.

A tag strip supports the rectifier, which has AC (~), positive and negative connected as shown. The lamp can be a vehicle type, supported by stout wires soldered to its base contacts, or side and tip. A hole is drilled in the panel level with this lamp, so that it will give indications of a short circuit or heavy load.

The transistor is mounted on the metal panel, to enable this to carry away heat, but is insulated from the metal. Use the correct insulation set (mica washer and bushes) which can generally be obtained with the transistor. A tag under one of the securing nuts provides the collector connection.

For voltage readings, a meter scaled 0–12v is most suitable. Or a 1mA meter, with 12k 1% series resistor, may be used. If so, a new scale, marked 0–12, can be drawn for it.

A 3 watt linear potentiometer is used. Earth and negative are connected to the metal panel, for equipment with a negative ground line. The positive output terminal or socket is insulated from the metal.

The case can be metal, or wood. It should have a few ventilation holes. The transistor can be expected to run quite warm, at maximum current. Heating here rises as the output voltage is reduced by VR1, as in these circumstances more voltage is dropped in the transistor.

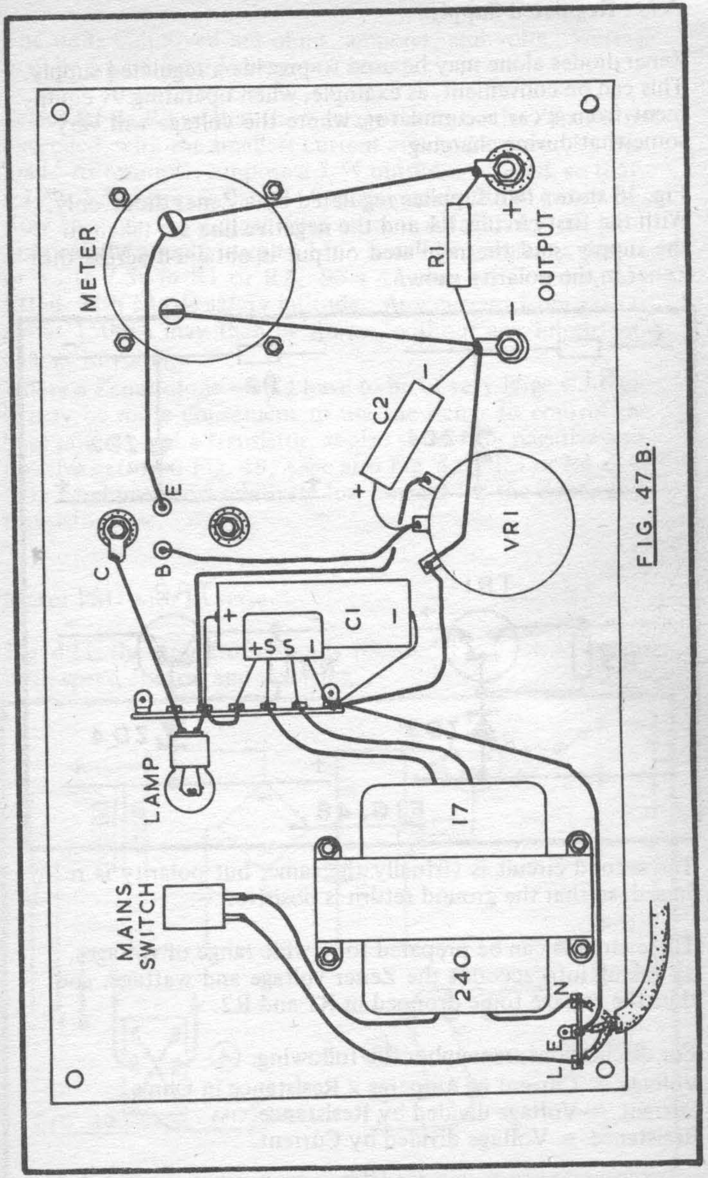

FIG. 47B.

Zener Regulated Supply

Zener diodes alone may be used to provide a regulated supply. This can be convenient, as example, when operating 9v equipment from a car accumulator, where the voltage will vary somewhat during charging.

Fig. 48 shows two supplies regulated by a Zener diode only. With the first circuit, R1 and the negative line are taken to the supply, and the regulated output is obtained across the Zener in the polarity shown.

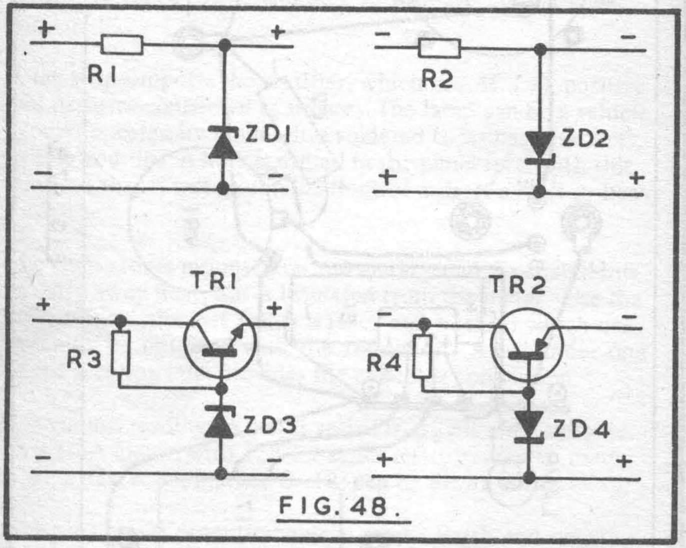

FIG. 48.

The second circuit is virtually the same, but polarity is reversed, so that the ground return is positive.

These circuits can be prepared for a wide range of voltages, by taking into account the Zener voltage and wattage, and thus the voltage to be dropped in R1 and R2.

For calculations, remember the following:
Voltage = Current in Amperes x Resistance in Ohms.
Current = Voltage divided by Resistance.
Resistance = Voltage divided by Current.

The units employed are ohms, amperes, and volts. Wattage may be found from Voltage x Current.

Select R1 or R2 so that the wattage rating of the Zener is not exceeded, with the smallest current ever to be drawn by the load. As example, suppose a 7.5v output is wanted, so that a 1½w 7.5v Zener is fitted. Approximately 200mA may thus flow through the Zener, if no current is taken by the load. Suppose the supply is up to 15v. It will be necessary to drop or waste 7.5v in R1 or R2. So a 42 ohm resistor may be fitted, with a little safety latitude. Any current from zero to about 150mA may then be drawn, without any important change in voltage.

Where a Zener diode would have to be of very large wattage, it may be more convenient to use the Zener to control the base potential of a transistor, as also shown for negative and positive earths in Fig. 48. (See also Fig. 45.) R3 or R4 then only need provide a relatively low current for the Zener and transistor base.

Motor PSU with Reverse

Fig. 49 is the circuit of a supply for model motors, and giving both speed control and reversing.

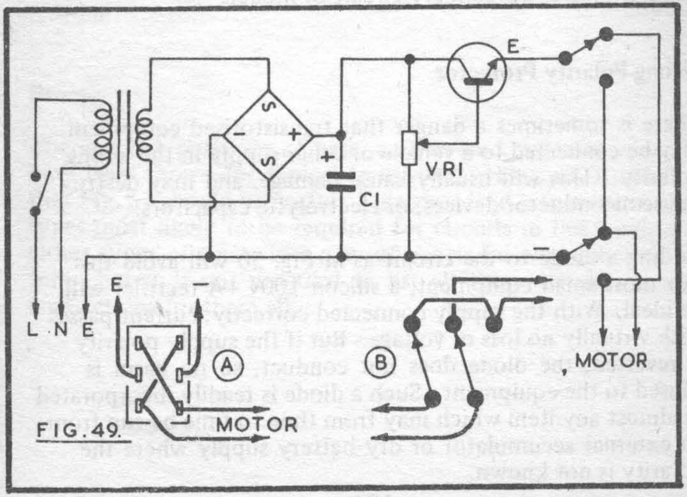

FIG. 49.

A transformer with a 12v secondary will be suitable for the usual model requiring up to 12v. The circuit is similar to that of the supply in Fig. 47. The potentiometer VR1 gives speed control. C1 can be around 2,000μF.

"A" is one type of ready-made switch, and it can be obtained with a central off position. "B" shows a home-built switch, similar to that described in Fig. 5. The switch must not short circuit the supply when changing over polarity, and one with central off is recommended.

Motors reversed in this way must be of the permanent magnet type, in which a change of polarity causes the armature to run in the opposite direction. Train and many other model motors are of this construction. Motors with a wound field will not be reversed by changing the polarity of supply.

A controller of this kind should be constructed as shown earlier, so that it is safe to use, and so that no joints in the mains circuits can be touched.

Earthing can only be omitted when a double-insulated transformer, of the type specifically made to avoid any need for earthing, has been used. The whole controller should be so constructed that, even if a fault develops, mains voltages cannot reach the secondary circuit or model.

Wrong Polarity Protector

There is sometimes a danger that transistorised equipment may be connected to a vehicle or other supply in the wrong polarity. This will usually cause damage, and may destroy the semiconductor devices, or electrolytic capacitors.

Adding a diode to the circuit as in Fig. 50 will avoid this. For most small equipment, a silicon 100v 1A rectifier will be ideal. With the supply connected correctly, current passes with virtually no loss of voltage. But if the supply polarity is reversed, the diode does not conduct, so no harm is caused to the equipment. Such a diode is readily incorporated in almost any item which may from time to time be run from an external accumulator or dry battery supply where the polarity is not known.

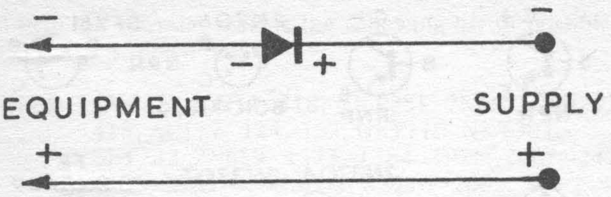

FIG. 50.

Component Tolerances

For circuits in this book, exact component values are generally not too important. Capacitor values such as 47nF (0.047μF) and 0.05μF may be regarded as interchangeable. In the same way, a 125μF capacitor may replace 100μF, or 200μF can be used instead of 220μF, and so on.

Resistors will generally be the standard 5% tolerance types. In some circuits exact resistor values are not at all critical. In others, and especially base, emitter and collector circuits, the values which provide best results can have much less latitude.

Bases

The lead-out positions of base, emitter and collector wires have been shown for typical transistors. However, it is felt that Fig. 51 should be useful. This gives lead information for types most likely to be required for circuits in this book. Other types of transistors may of course be used successfully, but it is not practical to list all these, or show connections for them all.

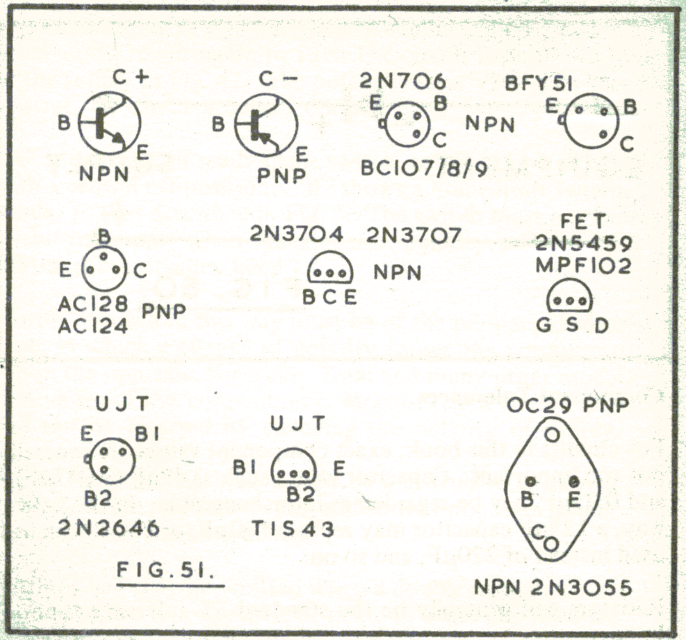
C +
B
E
NPN
C −
B
E
PNP
2N706
E
B
C
NPN
BC107/8/9
BFY51
E
B
C
B
E
C
AC128
AC124
PNP
2N3704
2N3707
BCE
NPN
FET
2N5459
MPF102
G S D
UJT
E
B1
B2
2N2646
UJT
B1
E
B2
TIS43
OC29 PNP
B
E
C
NPN 2N3055

FIG. 51.

EQUIVALENT SEMICONDUCTORS

To help our overseas readers the following list of possible equivalents is shown below: –

AC128	CV9778, OC74–318, NKT281, 2N467, 2SB222–415, AC117–124–153, GET110, NKT302, SFT232, 2N659–1373–1384–1926–2001–2431–4106, 2SB370, SK3004, RS276–2005.
ACY39	CV9475, NKT243, ACZ10, 2N652–2303–2801.
AF117	2N642, 2SA155, AF102–105–127–133–137, BFX48, BSW19–73, GM1213B, NKT603F, SFT354, SK3006, 2N2273–2635–3324–4916–5354, RS276–2003.
BC107	CV9780, AM251, BC147–167–207–317, MPS6566, SK3020–3122, ZTX107, TT107, RS276–2009.
BC108	CV10541, AM252, BC148–168–208–318, MPS6520, SK3020, ZTX108, TT108, RS276–2009.
BC109	CV10769–10806, AM253, BC149–169–209–319, MPS6521, SK3020, ZTX109, TT109, RS276–2009.
BC147	BC171–182/L–207–317, MPS6566, SK3020, RS276–2013,
BC148	BC172–183/L–208–318, MPS6520, SK3020, BS276–2009.
BC149	BC184/L–209–173–319, MPS6521, SK3020, RS276–2009.
BF200	BF162–173–225–271–311–314, SE5020/1/2/3/4, SK3018, 2N616–2708–3288–3693, RS276–2011.
BFY51	CV7723–7726–10827–11238, BFX51–68A, BFR19, BSX45/X, BSY46–85, SK3024, 40347, 2SC708, ME6102 2N2193/4–2218–2410–3053–3252–4046, RS276–2009.
BSY95A	CV7648–9379, BC108A, BSX51–73–91, BFX96, BSY20–58–62, ME9022, SK3122, 2SC321, 2N744–914–1708–2218–2369–5187, RS276–2011.
MPF102	E305, SK3116.
OA81	AA117–118–132–144, AAZ15, OA91–95–161, SD38, 1N38–56–476–618, 1S33.
OC26	CV8560, NKT452, 2N257–297–458–553–1021, 2SB3, AD130–131–138–140–149, OC28, SK3009, TF80, T1156–3027, 2N174–1073–2870–RS276–2006.

OC29	CV7083–8356, NKT402, 2N457, 2SB86–425, AD150, ASZ16, AUY21, OC28–35–36, SK3009, T13028, 2N174–1073–2870, RS276–2006.
OC44	CV5710–7003, OC170–171–410–613, 2N1303, 2SA15, AF101–117–126, ASY55, AC191, BC126, NKT211, SKT358, SK3005, 2N36–1191/2–1352–1373–1384, RS276–2003.
OC70	CV5457, OC303–602, NKT214, 2N279–1305, 2SB75, AC122–125–151–163, ACY27–34, ASY27, BC213–206A, BSX36, SK3003, 2N1193–1352–1375–2429, RS276–2004.
OC71	CV5712–7005–8344, OC70–304–604, NKT214, 2N280–1305, 2SB77, AC122–125–151–163, ACY35, BSX36, OC3041, SK3003, 2N1193–1352–1375–1384–2429, RS276–2004.
OC72	AC122/3–125/6–131–151 IV–152–162–191, BC126–213–206A, NKT211, SFT353, 2N282–1190–1352–1371–1384–1991–2431, CV5713–7006–8440, RS276–2005.
OC76	CV8314, OC307–6025, 2N284–1305, NKT212, 2SB89, AC128–131–151–152, ACY28, ASY58–76, BC126–213–260A, SK3004, 2N394–1191–1352–1371–1384–1991, RS276–2004.
ORP12	RS276–116, LDR03, RPY25.
1N4001	BY127, EM502, RS276–1101/1135, 1S100.
1N4002	BY127, EM502, RS276–1102/1136, 1S100.
2N1306	CV7350–10686, NKT736, 2N634–635–636–1891–1993–1994–1995, ASY26–29, BSX19–20, RS276–2001.
2N1613	CV7440–8843, BFY33–34–67, BSY44, 2N698–699–717–1711–1889–1890–1893, BFY41, BSX45/6, BSY53, MPS6530, 2SC708, 2N2101.

N.B.: Although equivalent semiconductors may have similar electrical properties, physical dimensions may be different and this must be born in mind if space is tight and for mounting details.

Remember polarities if replacing PNP with NPN types and vice versa. If in doubt always be advised by your dealer for suitable equivalent semiconductors.

Notes

by Peter J. Bertini, radioconnection@juno.com

Building An FM Crystal Receiver!

Our earlier crystal receiver projects proved popular, so here's another crystal receiver with an odd twist: it's intended to intercept stations on the 88- to 108-MHz FM band! With some modification, it'll also copy nearby VHF AM aviation band traffic, just the ticket for eavesdropping while watching the activity at your local airfield! FM crystal receiver projects aren't new; this project is based on material from an article that ran back in the 1950s, but with some modern refinements to improve performance. Most of the parts you'll need to build this project can be found in your junkbox, but I'll also provide some mail-order sources for those of you who don't have goodies hoarded away!

Monitoring FM signals is more challenging than tuning in your local AM stations with a simple crystal set. There are two reasons: AM signals are readily detected by a crystal-diode detector, and a simple longwire antenna will usually suffice in providing enough signal level so you can hear at least a few of the local stations. FM is a different animal. A crystal detector recovers both a DC and audio signal from an AM carrier, a simple diode detector will only produce a DC signal component when monitoring a FM signal. Since there are no AM sidebands to mix with the carrier in the detector, none of the audible information is recovered.

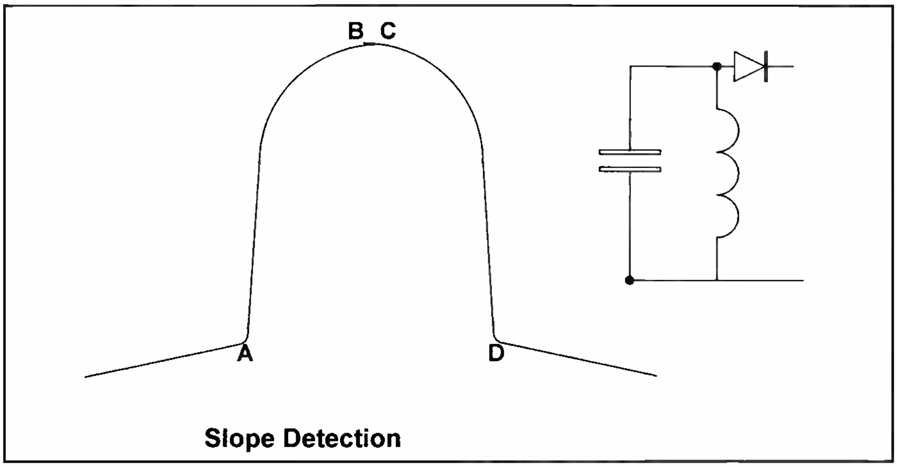

Figure 1. Here's a representative tuning curve for an LC tuned circuit as found in our FM Crystal Set. Slope detection occurs when the FM carrier straddles either side of the tuning curve between points A and B or C and D. An AM carrier would be centered between points B and C for best AM detection.

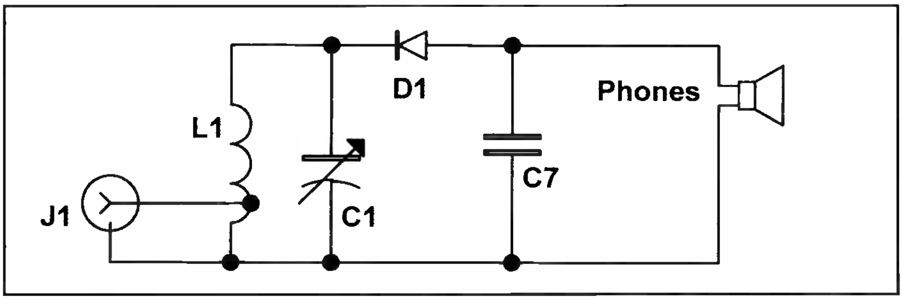

Figure 2. Here's the schematic for the simple and basic FM crystal receiver. FM performance is about 20 dB down from AM signals in the same tuning range. It makes a good companion at the local airfield to monitor AM aircraft and tower traffic in the 120-MHz region.

Slope Detection

So, how's it possible to make a working FM crystal radio? Good question. The answer lies in using a technique called "Slope Detection." FM is simply a carrier being shifted in frequency at an audible rate. The rate of shift corresponds to the frequency of the audio signal while the excursion, or carrier deviation, represents the audio signal level, or amplitude. U.S. FM broadcasters are allowed a deviation of ± 75 kHz, or 150 kHz total. This means we can hear FM signals on a simple crystal radio, providing the tuned circuit has enough Q for the recovered audio to be great enough to drive a pair of headphones. Whereas a crystal detector will detect AM signals without tuned circuits, the FM set relies on a very sharply tuned high-Q circuit to work. If we tune the FM station so it falls on either side of the slope, the modulated FM carrier will be converted to AM since its signal amplitude will vary depending on where it falls on the tuning slope at any given instant.

Many early 2-meter amateur FM repeater operators used AM transceivers, modified with simple external FM modulators, while using slope detection to hear the FM repeaters. Most of these early AM receivers were 15 or 30 kHz wide; the IF bandpass curves easily recovered the ± 5-kHz FM deviation! The trick was to tune to one side of the signal or the other, carefully centering the FM carrier on the center of the tuning skirt, not dead on at the peak! In short: The steeper the slope, the greater the recovered audio for a given deviation.

Figure 1 illustrates how slope detection works. The curve on the drawing represents a typical, but not exact, representation of how a tuned circuit responds as it is tuned across a carrier. Points A and D denote the bandwidth of the tuned LC circuit. The areas between points A and B and C and D represent the slopes of the tuning curve. Ideally, an AM signal would be tuned so its carrier is at the apex, or center, of the tuning curve. If we center the carrier frequency of the FM transmitter to the same spot, the carrier will swing over both slopes, and any recovered audio would be canceled. Instead, the "sweet spot" is where the FM carrier falls near the midpoint of the slope between A and B, or likewise on the slope of the curve between points C and D. The AM detector output will then swing in correspondence with the aural information carried on the FM carrier.

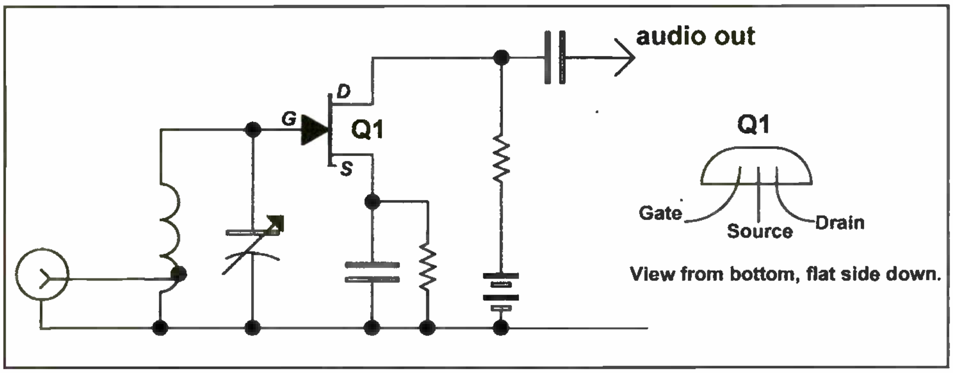

Figure 3. The infinite-impedance detector offers several advantages over a germanium diode detector. It has gain, doesn't load the tuned circuit, and it features better weak-signal detection.

Ideally, for maximum audio recovery, the FM deviation will "sweep" over the largest area of one side of the slope as is possible. It's hard to match the tuning curve to the FM deviation, and also any non-linearity in the slope along the rise or fall on the curve (a straight line, like a saw tooth, is ideal) will cause distortion. Too wide a slope and the FM deviation will fall on a very small area on the slope, giving very poor recovered audio. Too much deviation, where the carrier swings beyond the cutoff and into the opposing slope, yields large amounts of distortion on the recovered audio. In practice, these sets do work, but don't expect anything approaching high-fidelity reception!

The Simplest Receiver

There are several key elements needed for this project to work, most importantly, having a very strong local FM station or two or listen to. Most urban locations are so blessed. You'll also probably need a good outdoor FM or TV antenna to capture enough signal for the set to work, and it's a big plus if your system includes an inline VHF/FM preamplifier.

Let's get started by looking at **Figure 2**, the starting point for our project. It closely resembles the primitive FM crystal set projects shown in those early magazines. The heart of the receiver is the high-quality tuning capacitor and inductor. The tuning capacitor can be salvaged from an old FM tuner or converter, if need be. The tuning capacitor (C1 in the drawings) is a 35-pF capacitor with a built in 3:1 vernier dial reduction, and was salvaged from my junkbox. Any capacitor with a maximum capacity of 25 to 45 pF will work.

The set tunes rather broadly, thus the vernier tuning isn't a necessity, but it is nice to have. One source for the cap is Fair Radio; its 3G-25 variable cap has a vernier drive and three sections with 4-25, 4-20, and 3-17 pF. Two of these sections could be paralleled to yield a larger tuning range. The inductor (L1 in all drawings) is wound from a length of #8 copper ground wire, available from any hardware store. You can substitute a smaller gauge if necessary. The circuit Q could be improved by silver-plating the coil—if you have a means to do so try it. Otherwise, I suggest polishing the copper to a sheen, followed by applying a coat of clear Krylon to prevent tarnishing. The coil consists of six turns; I used a half-inch wood dowel as a form to wind the coil. The antenna tap is made at one-half turn above ground on the coil. **Photo A** shows the tap point and also the tuning capacitor. **Table 1** lists the parts values and sources for all the components used in three versions of the receiver shown in this month's column.

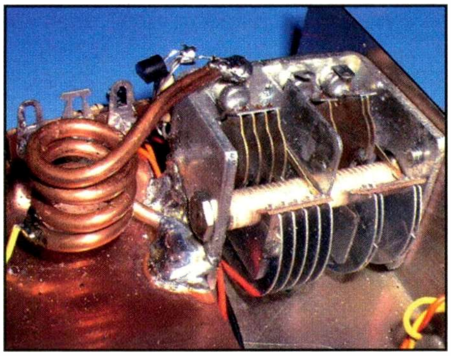

Photo A. The tuning capacitor and coil form the heart of the crystal set. The Q of these parts determines the tuning slope, and ultimately audio recovery on FM signals. The capacitor should be the best quality available. Often military surplus capacitors with ceramic insulation are available on the surplus market or at hamfests. If you can find one, use it. The tap point for the antenna is visible. Use a short lead between the tap and J1 connector.

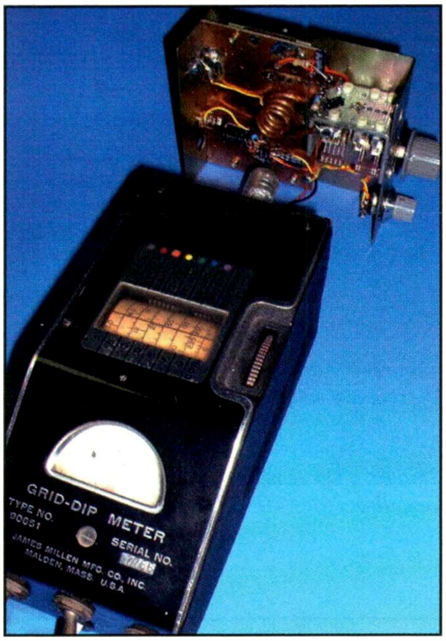

Photo B. Here's the Millen Grid Dip Meter in action! When tuned to the resonant frequency of the capacitor and coil, the meter on the Millen will dip to a lower value. Fifty or 60 years ago, every ham owned one of these useful pieces of test equipment. That was a time when hams built their own gear. Now using one has become an archaic art.

I know finding parts can be difficult. I listed a few resources in **Table 1**, including one for a basic kit, sans headphones, for building the basic FM crystal set. This five-dollar postpaid kit can serve as the nucleus for the more advanced versions shown here as well.

Checking The Tuning Range

A grid-dip meter is the easiest way to check the tuning range of your coil and tuning capacitor combination. A ham friend who is also a builder is a likely candidate to have one of these units in his workshop. **Photo B** shows my venerable James Miller *'dipper* being used to check that my set will tune the entire FM band. These devices are variable oscillators with calibrated tuning ranges. A set of companion plug-in coils allows covering from the broadcast band up through 300 MHz. When the *'dipper's* coil is held near a tuned circuit, and the grid dipper is tuned to that frequency, the grid current in the *'dipper* will drop (or "dip"), as shown on the *'dipper's* built-in meter.

Alternately, a signal generator would also do for checking the tuning range. Compressing or expanding the coil slightly will shift the range: squeezing the coil will allow lower frequency tuning, while

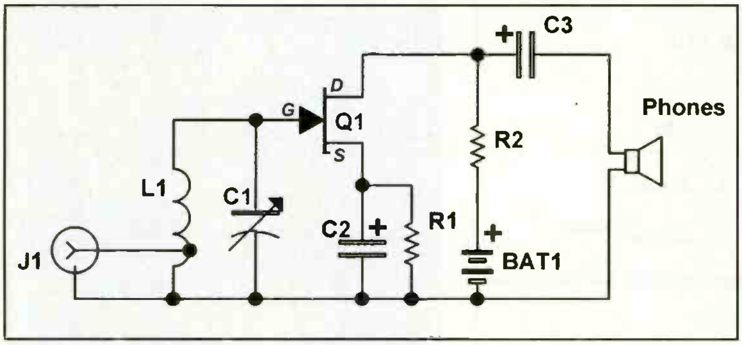

Figure 4. Adding the infinite-impedance detector to the basic FM crystal receiver might make the difference between hearing a station or not!

Photo C. The leads for the infinite-impedance detector should be kept short and direct for VHF. The J-FET leads and associated parts are mounted to the capacitor terminals and frame. ⟶

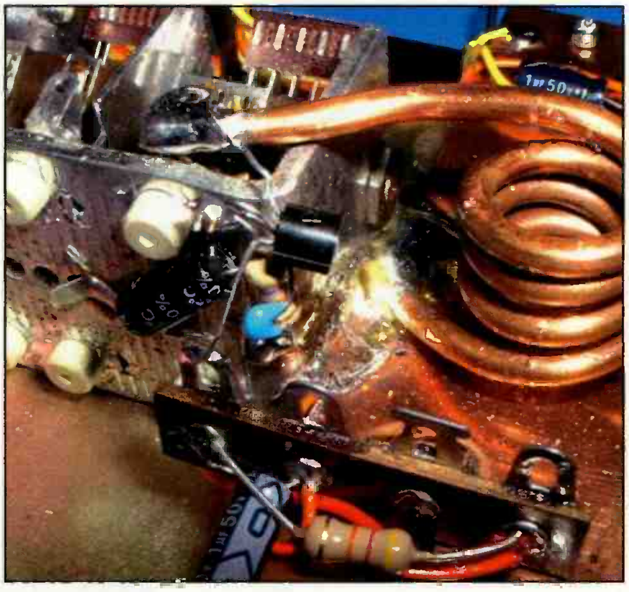

expanding the coil will increase the higher end of the tuning range. A 40-pF variable capacitor should tune from 80 MHz through at least 130 MHz into the aircraft band. At worst, you may have to wind a new coil with a turn added or removed before you're able to tune the desired range. Adding a full or partial turn will lower the frequency. **Photo B** shows my Millen dip meter in action.

Improving The Detector

Crystal diodes make poor detectors. Germanium diode detectors don't work well on weak signals; for a weak enough signal they act more like resistors than diodes. **Figure 3** shows a circuit for a neat device called an Infinite Impedance Detector. This detector offers some gain and doesn't load the tuned circuit as much as a diode detector would; hence the Q and selectivity are also improved. The J-FET gate adds some capacity to the tuned circuit, and this can reduce the high-frequency tuning by several MHz.

Photo C shows how the J-FET and associated parts for the infinite-impedance detector are mounted using short leads. The gate is tied directly to the tuning capacitor stator terminal, and the FET's *source* lead bypass caps and biasing resistor are tied directly to the capacitor frame for good VHF performance.

Figure 4 shows how the infinite-impedance detector is incorporated in the FM crystal receiver. Parts values are given in **Table 1**. The detector will drive a pair of vintage high-impedance headphones (2000 ohms or better) directly. I use a pair of military sound-powered headphones with a matching transformer. Adding the

Table 1. Parts List And Recommended Suppliers

L1	#8 copper wire, see text
C1	variable capacitor, 5 to 50 pF max. with knob
D1	1N34A germanium diode, or equivalent
J1	RCA phono jack or suitable RF connector for antenna
C2	10 µfd @ 16 volts electrolytic
C3	10 µfd @ 16 volts electrolytic
C4	10 µfd @ 16 volts electrolytic
C5	0.1-mfd ceramic capacitor @ 16 volts
C6	220 to 470 µfd @ 16 volts electrolytic
C7	500-pF ceramic disc or silver mica capacitor
R1	47k-ohm 1/4-watt resistor
R2	47k-ohm 1/4-watt resistor
R3	10k-ohm variable potentiometer, audio taper/with knob
BAT1	9-volt transistor battery with battery clip
Phones	2000-ohm vintage headset or crystal earpiece
Speaker	8-ohm speaker in enclosure
Q1	J-FET, type J-310
U1	LM-386 DIP package 8-pin audio IC

Resources

• Dan's Small Parts, PO Box 3634, Missoula, MO 59806-3634, (406) 258-2782, www.danssmallpartsandkits.net

This is an excellent resource for all components (ICs, FETs, small parts, and variable caps). Dan's features low prices and is set up to deal with home experimenters. Dan's offers an online catalog and ordering system. Minimum order might apply, check with seller.

• Bill Turner, 1117 Pike Street, St. Charles, MO 63301, (636) 949-2210, www.dialcover.com

Bill offers a kit and copy of the original vintage magazine article for the FM Crystal receiver for $5 postpaid. No frills or headphones, just the basic parts. No minimum order.

• Fair Radio Sales, 2935 St. Johns Road, PO Box 1105, Lima, OH, (419) 227-6573, www.fairradio.com

Surplus stalwart Fair Radio has a nice three-section variable (catalog #3G-25) that will work well. Price is $6. They also offer some high-Q ceramic military versions, but these are more expensive. Minimum orders might apply, check with seller.

Photo D. The audio amp IC is mounted using "dead bug" construction techniques. The larger parts are glued to the pc board to alleviate stress on the IC leads.

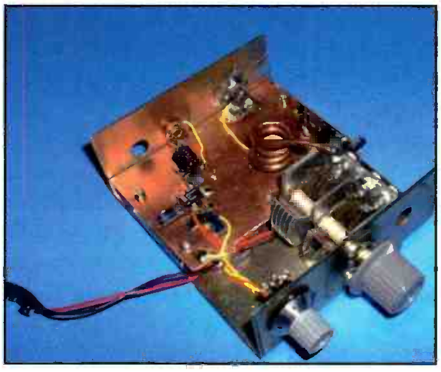

Photo E. Here's a top view of the completed receiver. Ugly, but it works—and it was intended as an experimental prototype to begin with.

infinite-impedance detector allowed me to hear some FM stations from my rural home when using some high-gain tower-mounted VHF antennas.

You might want to try this detector on your AM BCB crystal sets; it makes a big improvement!

Adding An Audio Amplifier

Figure 5 shows the final phase of the FM crystal receiver. A small 8-pin LM-386 audio IC gives enough gain for loudspeaker level volume, or to drive more contemporary low-impedance headphones. I mounted the IC and supporting parts "dead bug" fashion on the pc board. If you're not familiar with this breadboarding technique, the parts are soldered directly to the copper side of an unetched pc board for ground, and are self-supporting or glued in place as needed. (ICs were sometimes mounted on their backs, looking like little dead bugs; hence the name.)

It's not very pretty, as shown in **Photo D**, but it works and is good for quickly prototyping circuits on the fly. I think some builders call this technique *Ugly Board* construction. Hot glue assisted in mounting and supporting the IC and larger components to the board. **Photo E** shows an overall view of the entire receiver. Again, it isn't pretty, since I considered this to be an experimental project. An off/on switch is needed to preserve battery life.

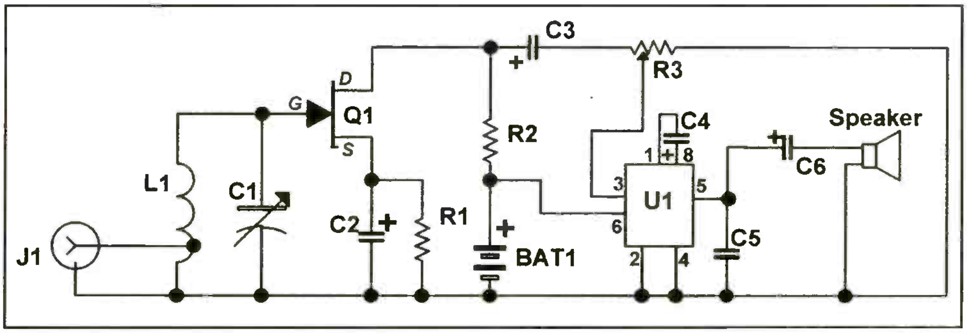

Figure 5. Adding the high-gain audio stage gives loudspeaker volume on strong stations.

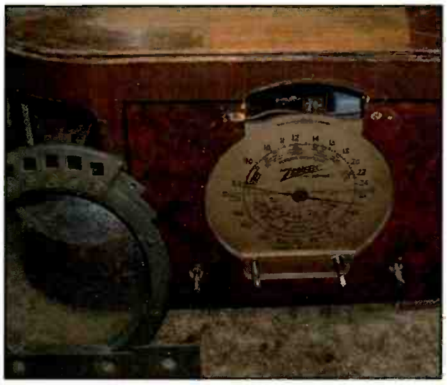

Photo F. This is my 6S321 Stars-and-Bars Zenith! Sorry for the mix up! As you can see, this cabinet needs a bit of work. I'd started scraping the old lacquer off the top when the photo was taken. Chemical strippers risk damaging the photo-etched lacquered patterns on the front, which I intend to restore and save.

An Apology

Our opening photograph in the July 2004 issue purportedly showed the deplorable condition of my Zenith 6S321 cabinet. What accidentally got forwarded for publication was a photo of a nicely restored example from the webpages of Mike Urban's Internet site at www.urban-antiqueradio.com. I had saved his photo as an example to follow when restoring my Zenith. Hopefully **Photo F** is the correct photo. Anyway, visit Mike's website to see some nicely restored radios in his collection, and to check out a few he's offering for sale.

Well, that's it for this month. If you build one of these sets, I'd like to hear about it, and photos would be great! Let me hear about your other projects, as well. Until next month, keep those soldering irons warm and those letters coming in! ■

If you enjoy radio communications you'll love....

POPULAR COMMUNICATIONS

The World's most authoritative monthly magazine for Shortwave Listening and Scanner Monitoring. Read by more active listeners world-wide.

*Get fast home delivery of **Popular Communications** and save $30.93 a year over the newsstand price. Save even more on 2 or 3 year subs.*

☐ 1 year - 12 issues $28.95 (Save $30.93)

☐ 2 years - 24 issues $51.95 (Save $67.81)

☐ 3 years - 36 issues $74.95 (Save $104.69)

SAVE UP TO 58%

SUBSCRIBE TODAY!

Canada/Mexico–One year $38.95, two years $71.95, three years $104.95;
Foreign Air Post–one year $48.95, two years $91.95, three years $134.95,

Payable in US dollars onlyNote: Allow 6-8 weeks for delivery of first issue.

Mail orders to:
Popular Communications, 25 Newbridge Rd, Hicksville, NY 11801
FAX 516-681-2926

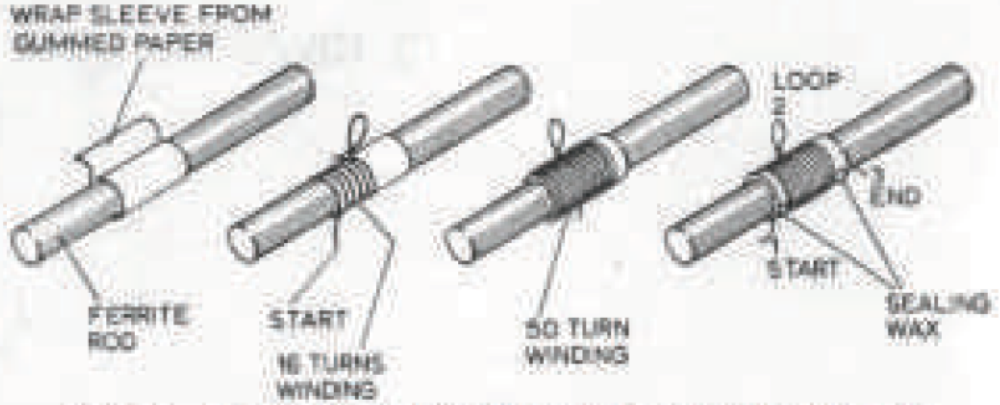

Fig. 3.1 Aerial tuning coil wound from 38-s.w.g. wire on a ⅜-in. diameter ferrite rod.

Fig. 3.1 details the construction of a suitable coil for a tuned circuit, matched to a ferrite rod of ⅜ in. diameter and about 4 in. long (see Chapter 4 for alternative coil windings on different rod sizes). If a rod is purchased longer than 4 in., it can be used as it is, or cut down by marking around with a file and then breaking off the surplus length.

Cut seven 1-in. lengths of gumstrip. Moisten one and wrap around the rod gummed side up. Now add about another half a dozen wrappings of similar length over the first, this time with gummed side *down* to form a reasonably rigid tube. Make sure that the paper tube is a *sliding* fit on the ferrite rod and leave to dry thoroughly (preferably removed from the rod so that it cannot become stuck to it).

When the paper tube is quite dry it should be rigid, when the coil windings can be applied. The wire to be used is 38-s.w.g. enamelled copper wire, the number 38 referring to the actual diameter of the wire according to the standard wire gauge (s.w.g.).*

Starting about ¼ in. in from one end of the paper tube, wind the wire carefully round the tube, with each turn tight against the one before it, until sixteen full turns have been completed. Then make a loop in the wire, as shown, and carry on winding, with succeeding turns touching, until fifty turns in all have been completed. The two loose ends of the coil (the start and finish) can be secured with a dab of sealing wax whilst the projecting loop can be twisted together (e.g. by putting a pencil through the loop and twisting up). Cut off the loop, leaving about ½ in. protruding from the main coil, bare the wire ends and solder together. This forms point 2 on the coil; the start is point 1, and the end point 3–see Fig. 3.1. It will be easy to remember these without marking since the loop or tapping point (2) comes much closer one end (1) than the other (3).

---

* American wire gauge (AWG) is about two sizes smaller than the British standard wire gauge (s.w.g.). For example, 38 s.w.g. is approximately the same as 36 AWG.

Cut a panel of *Paxolin sheet* to about the size shown in Fig. 3.2, using a hacksaw.* On this secure a *tag strip*, as shown, and drill a hole to mount a miniature or small-size 500 pF *variable capacitor*.

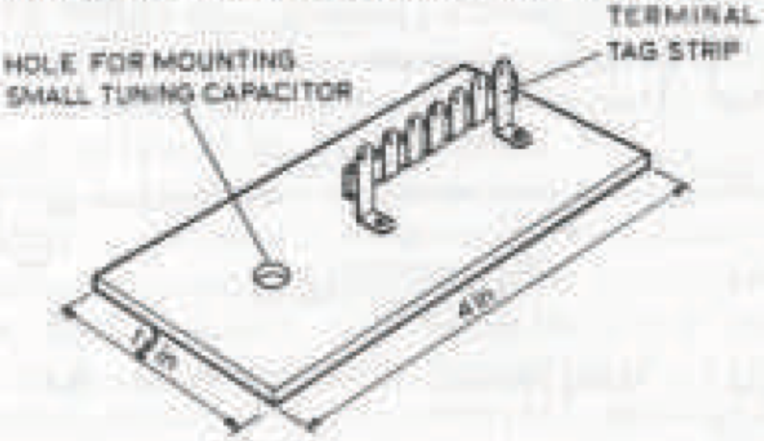

Fig. 3.2 Paxolin panel and tag strip for crystal set circuits.

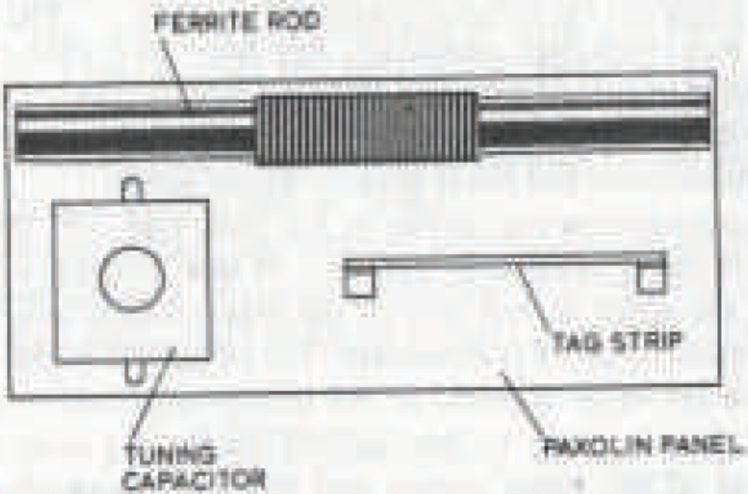

Fig. 3.3 Layout of components on Paxolin panel.

The aerial coil is then mounted on the panel as shown in Fig. 3.3, gluing the coil on to the Paxolin with two or three dabs of sealing wax, or some other suitable adhesive. *Note:* the ferrite rod must be free to slide in the paper tube for 'tuning' adjustments.

Virtually any miniature germanium or silicon diode will be suitable for the detector. Recommended types, which are readily available, are 1N34 and 1N914.

Earphones must be of *high-impedance* type, which need not be expensive to buy, and the higher the impedance the better the reception.

---

* A phenolic or fiberglass sheet may also be used.

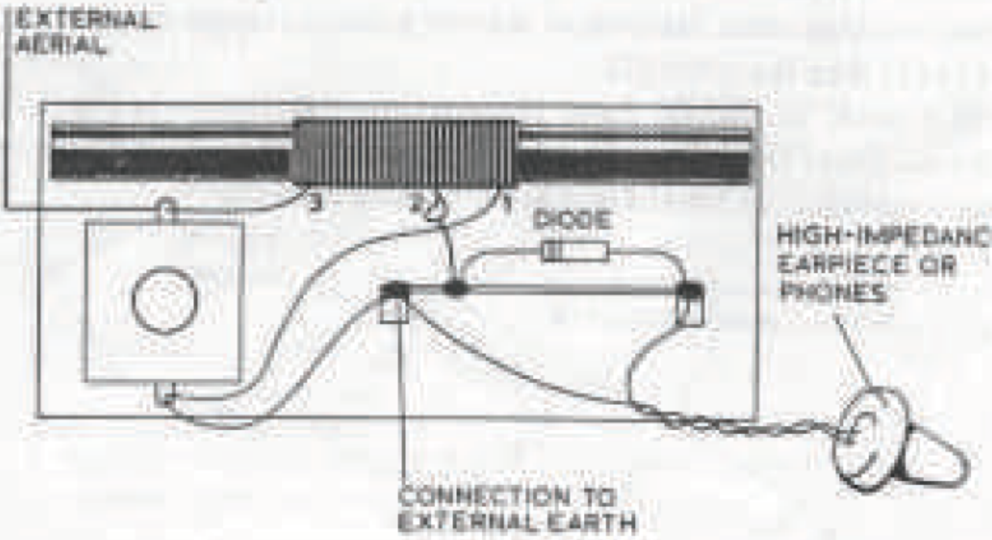
EXTERNAL AERIAL
3
2
1
DIODE
HIGH-IMPEDANCE EARPIECE OR PHONES
CONNECTION TO EXTERNAL EARTH

Fig. 3.4 Wiring connections to complete the crystal set.

Alternatively a deaf-aid type earpiece can be used; this will not give the same volume or quality of reproduction as headphones, but is a less expensive component. This should preferably be of the *high-impedance magnetic* type, with high sensitivity. Any high-impedance earpiece will suffice, but if of crystal type will require a resistor, connected across it, to complete the circuit. This will reduce the amount of current flowing through the earpiece and lower the strength of signal.

Wiring connections are shown in Fig. 3.4. End 3 of the aerial coil (the end of the 50-turn coil) connects to one terminal of the tuning capacitor, the aerial or 'hot' end of the tuned circuit, and the point to which an external aerial is connected. The other end of the coil (end 1) connects to the other terminal of the tuning capacitor, from which an additional wire is taken to the first tag on the tag strip. This is the 'earthy' end of the tuned circuit, and the point to which an external earth is connected. Leave plenty of slack wire between the coil and tuning capacitor.

The other connections are then as follows:

(i) Tapping point of the coil wire bared and connected to the second tag.
(ii) The diode also soldered to this same tag, and to any other free tag.
(iii) Headphone (or earpiece) leads to the 'earthing' tag, and to the 'free' tag to which the above diode has been connected. All connections should be made with *soldered* joints.

The wiring-up can be checked against the *circuit diagram* shown in Fig. 3.5 (ignoring the components shown with broken lines). The set should now be 'working'.

In areas of strong signal strength, no external aerial or earth connections should be necessary. Performance will, however, be improved in any area by attaching an aerial wire (which can be any thin wire, e.g. using the same wire as for the coil winding), of up to 160 feet. The longer the aerial the better the reception, provided it is led away from the receiver to as *high* a point as possible.

An earth connection may further improve aerial performance; by this we mean a connection to some conductor positively in contact with the ground (preferably buried), an excellent example being a metal water-pipe. Thus, if an earth connection is found to be necessary (or you want to try one to see how performance is affected), connect a wire from the 'earthing' tag on the receiver to a convenient water-pipe.

This question of obtaining a *good* aerial and earth is a most important one in areas of poor signal strength. Linking up to a television aerial is often a good plan, since TV aerials are also usually mounted as high as possible. If bare wire is used, it is also important that the upper (free) end of the aerial is not made fast to something which could produce an earth connection (e.g. a damp tree), or at least is suitably insulated from such a support. String is not an efficient insulator; that, too, can conduct when wet.

Quite good results are often obtained by using the springs of a bed as an aerial, in which case an earth connection is usually not necessary. Sometimes, too, when other attempts to yield a good signal strength in

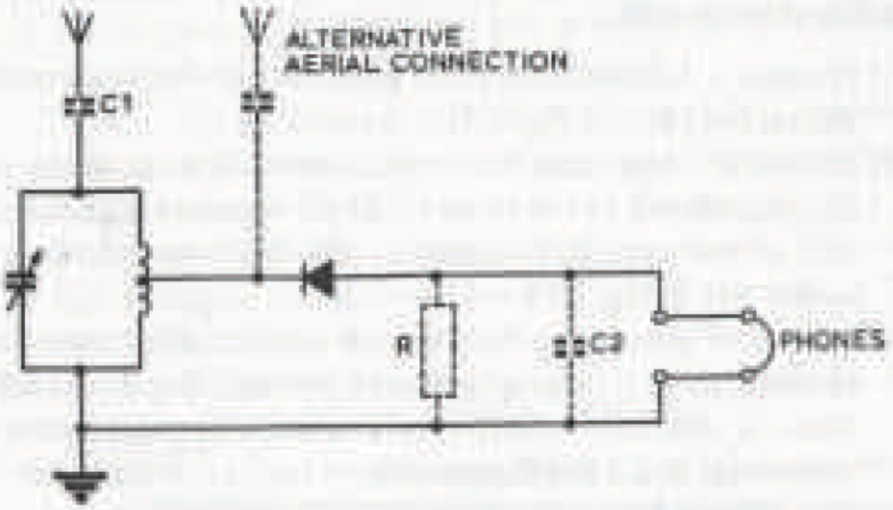

Fig. 3.5 *Circuit 1.* Circuit diagram of basic crystal set.

the aerial have failed, connecting the *aerial* side of the tuning coil to a *good earth* (a water-pipe) can produce better results, the normal earth connection being left off.

Tuning

The receiver is adjusted as follows. Turn the tuning capacitor to fully close the vanes, then open about half a turn on the spindle or knob. (If you are using a trimmer as a tuning capacitor, screw right down and then open half a turn.)

The tuning coil should now be slid up and down the ferrite rod (the coil leads were left a fairly long time to give the necessary freedom of movement) until BBC Radio 3 is heard.*

It may be necessary to slightly alter the adjustment of the variable condenser to tune in to this programme. Also, because of the ferrite-rod aerial the set will be directional, that is, the signal strength received will depend to some extent on the direction in which the aerial rod is pointing, so position the set to pick up the maximum volume.

Having established the best position of the tuning coil on the rod to receive BBC Radio 3, fix permanently with a dab of sealing wax. You should then find it possible to tune in to further stations by altering the setting of the variable capacitor—e.g. typically Radio 4 in about the middle of the capacitor travel and Radio 1 towards the other end.

Any reception you get will almost certainly be very weak and (unless you live close to a broadcast station) you can really feel satisfied if you get any station at all at audible strength. But it is surprising how, sometimes, even quite distant stations can be heard. Also you can often improve the reception and listening strength by quite simple modifications. Try these in order:

(i) Connect a 1,000-pF capacitor across the headphone (earpiece) connections (C2 in Fig. 3.5).

(ii) Instead of connecting the external aerial directly to the tuned circuit, connect one lead to a 220-pF capacitor, and the other end of the capacitor circuit to the 'hot' end of the tuned circuit (C1 in Fig. 3.5).

(iii) Instead of connecting the external aerial to the tuned circuit, connect to the tapping point of the coil (tag to which the diode is also connected). Try a direct connection, and also connecting in a 220-pF capacitor.

---

* Tune to an AM broadcast station near the lower end of the dial—close to 550 kHz.

(iv) Try connecting a 1.2-k resistor (or higher value) across the phone connection (R in Fig. 3.5). You may be using the wrong type of phones or earpiece, which do not provide a proper load or complete the circuit.

If there is a complete lack of response, check for faulty wiring-up. A more likely cause, however, is lack of an external aerial or earth connection in an area where these are strictly necessary for adequate reception; or an inefficient aerial (too short) or poor earth connection (bad electrical contact to a good earth point, or connection to a bad earth point).

Another possible cause of apparent failure may be too much outside noise entering the ear so that it is impossible to detect the very weak radio signal as it is being tuned in. Headphones are better than a single deaf-aid type of earpiece in this respect but, in any case, a really quiet room is virtually essential for initial setting up and tuning adjustments. Also, if your adjustment of the tuning control is too coarse, you may completely miss the setting for the station you are looking for, without realizing it.

Reception will also tend to vary with weather conditions. Some days it may be so poor that what was normally a strong station is hardly heard at all. The simple basic receiver has many limitations but, since it costs very little to construct and nothing at all to operate, this must be regarded as inevitable.

Providing you can hear something–even if too weak a signal to distinguish properly–you can certainly improve the performance of your basic set by further experimentation with tuned circuits (see Chapter 4) and/or the addition of amplification to the circuit. You can also try other types of basic crystal set, as described in the following projects.

Circuit 2 (Fig. 3.6) is identical to *Circuit 1* except that, instead of a diode, a transistor is used as a detector. Only two of the transistor leads are connected–the emitter (*e*) connection to the tapping part of the coil, and the base (*b*) to the 'earthy' end of the circuit. The collector lead of the transistor is ignored (bend it out of the way so that it cannot accidentally short out the other leads).

You can try almost any type of low-cost AF transistor; recommended types are OC42 2N370, 2N2925, 2N5088.

With the addition of two more components, Circuit 2 can be modified to work the transistor both as a detector and an amplifier, to give stronger signals through the headphones. Using the same transistor type

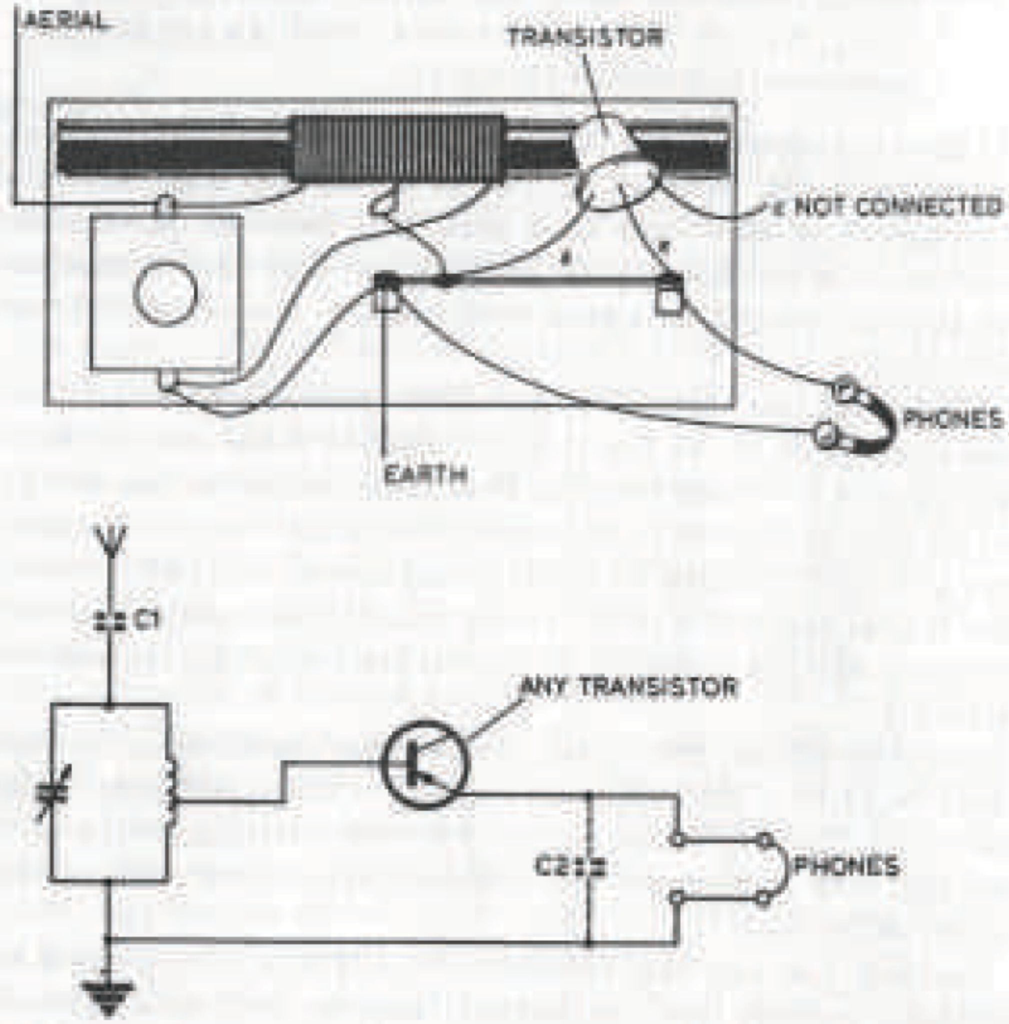

Fig. 3.6 *Circuit 2.* Crystal set using transistor instead of a diode.

as above (or near-equivalent) resistor R should be 15 k and capacitor C3 1 μF or higher.

This time a battery is also required to supply power for the transistor to work as an amplifier. This can be from 1.5 V up to 9 V. Remember the rules for polarity of connection: those shown in Fig. 3.7 are for a p-n-p transistor; an n-p-n transistor would need the battery connected the opposite way round. Battery polarity also affects the connections of capacitor C3 (if an electrolytic or polarized type).

Experiment further by trying the effect of using additional capacitors in the circuit, e.g. C1, 220 pF; C2, 0.001 μF (try other values as well); C4, 0.001 μF (try other values as well).

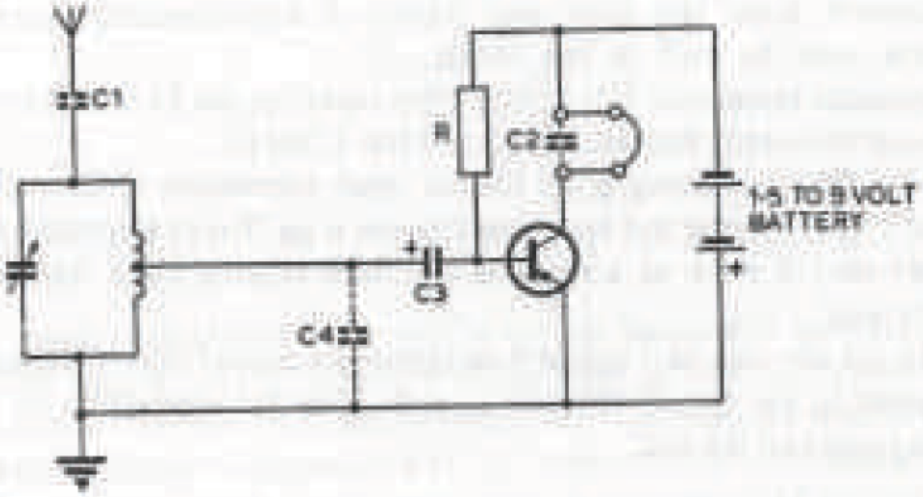
C1
R
C2
C3
C4
1·5 TO 9 VOLT BATTERY
+
+

Fig. 3.7 *Circuit 3.* Crystal set with amplification.

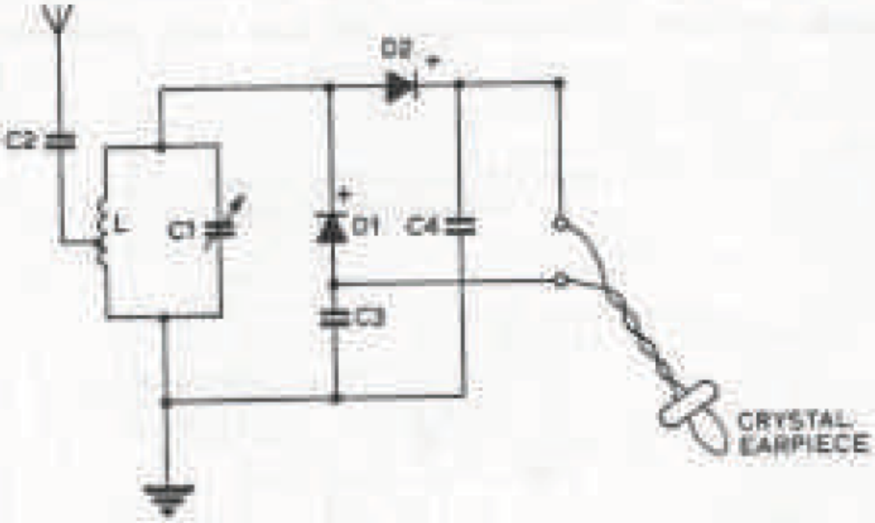
D2
+
C2
L
C1
+
D1
C4
C3
CRYSTAL EARPIECE

Fig. 3.8 *Circuit 4.* Double diode crystal set.

This circuit incorporates 'voltage doubling' to improve the signal volume and should give better performance than a single diode circuit. L and C1 are the usual tuned circuit, but an unusual feature is that the *aerial* is connected to the coil tapping point. Any type of germanium diodes can be used (they should preferably be the same), making sure to connect them the right way round. A high-impedance *crystal* earpiece must be used in this circuit.

Capacitor values are: C2, 220 pF (this capacitor can be omitted–try with and without in the circuit); C3 and C4, 1,000 pF.

The optimum tapping point for the aerial connection to the coil L is best found by trial and error (see Chapter 4 on 'Tuned Circuits'), but the set should work at a nominal one-third tapping point from the 'earthy' end.

This set can also be tried with conventional tuned-circuit coupling–*i.e.* aerial to the top of the coil L, and diode D2 connection to the tapping point on the coil.

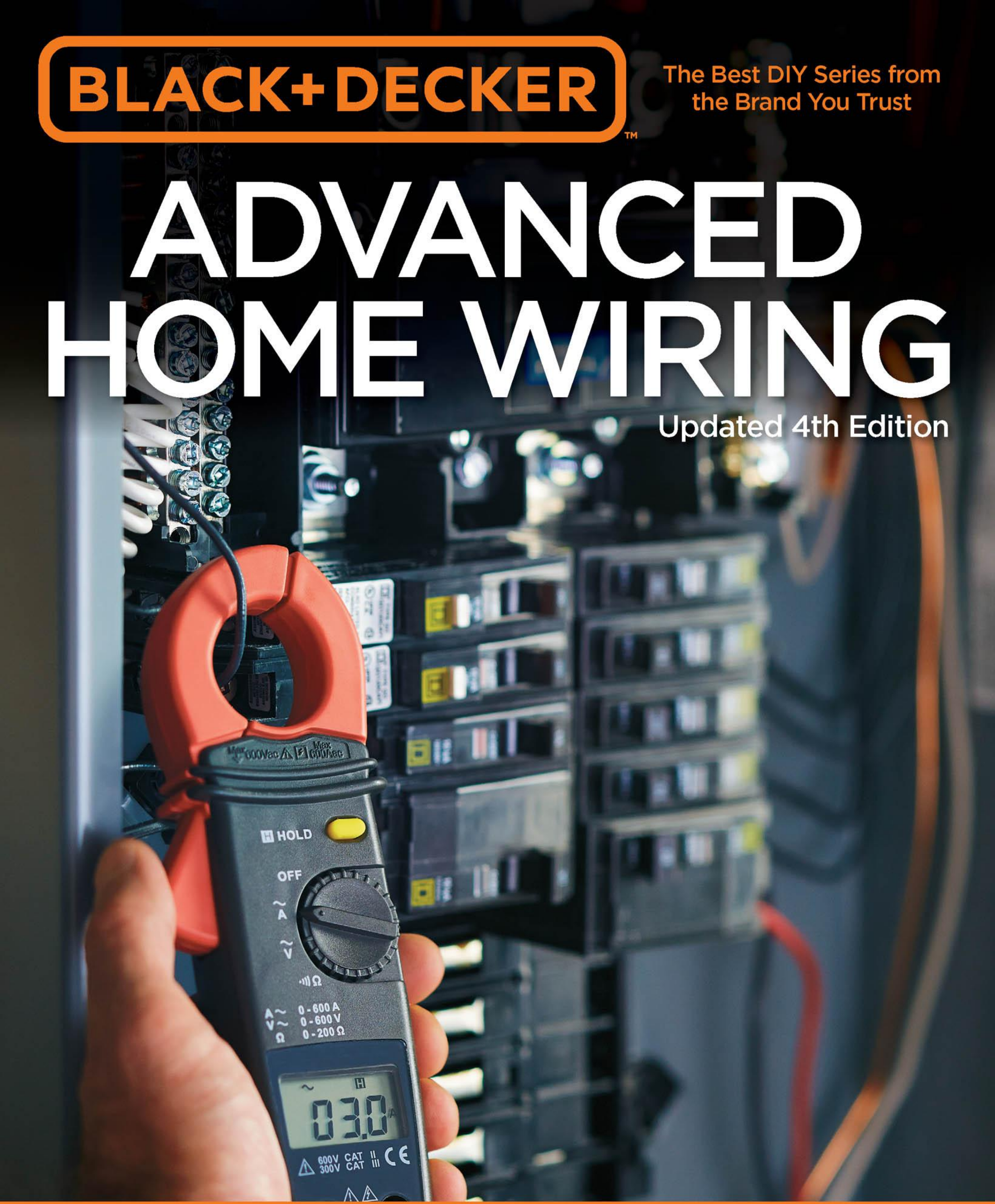
HOLD
OFF
0-600 A
0-600 V
0-200 Ω
03.0

BLACK+DECKER™

The Best DIY Series from
the Brand You Trust

ADVANCED HOME WIRING

Updated 4th Edition

DC Circuits • Transfer Switches
Panel Upgrades • Circuit Maps • More

Introduction

Experienced home electricians understand the need to keep up with changes in the world of wiring. Wiring projects, and more advanced projects in particular, almost always require a permit from your municipality and typically an on-site inspection or two as well. If you aren't up-to-date with wiring codes, there is a likelihood that your project will not pass inspection. But beyond the practicality of passing inspections, the codes that govern wiring practices are updated for good reason: they improve safety. And when you're talking about your own home and family, that's worth paying attention to.

This newest edition of BLACK+DECKER *Advanced Home Wiring* has been reviewed and revised to reflect the many changes to wiring code published in the 2014 edition of the National Electrical Code (NEC), which is updated every three years. And this time around there were an unusually high number of changes. They include the requirement that an available neutral wire be present in every switch box, expansion of the types of circuits that require AFCI protection, a new prohibition on tying into a garage receptacle, and more. Several of these changes are reflected in the updated information you'll find here.

Almost all of the advanced wiring projects featured in this book involve new circuitry, panel upgrades, or troubleshooting with diagnostic equipment. Among the high-level projects: making a direct-current, solar-electric circuit; upgrading the grounding and bonding on your new 200-amp or larger home circuit; installing an automatic transfer switch for your backup power supply; wiring a room addition; and using a multimeter to precisely locate an open neutral in a home circuit.

Because the projects found in this book are advanced in nature, do not attempt any of them unless you are confident in your abilities. Consult a professional electrician if you have any concerns—in many cases your best solution might be to do some of the work yourself, such as pulling new sheathed cable through walls, and to have the electrical contractor do the other work, such as making the connections. But do keep in mind that home wiring can be a fun and fascinating pursuit, and successfully accomplishing a major project is personally gratifying and can also save you substantial amounts of money.

1st
proofs for reference only

Edition
Page: 4
41940 - Advanced Home Wiring_001-063.indd 5
10/10/14 9:06 AM
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
Page: 5
Pro-Vision C M Y K Text

Wiring Safety

Safety should be the primary concern of anyone working with electricity. Although most household electrical repairs are simple and straightforward, always use caution and good judgment when working with electrical wiring or devices. Common sense can prevent accidents.

The basic rule of electrical safety is: Always turn off power to the area or device you are working on. At the main service panel, remove the fuse or shut off the circuit breaker that controls the circuit you are servicing. Then check to make sure the power is off by testing for power with a voltage tester. *Tip: Test a live circuit with the voltage tester to verify that it is working before you rely on it.* Restore power only when the repair or replacement project is complete.

Follow the safety tips shown on these pages. Never attempt an electrical project beyond your skill or confidence level.

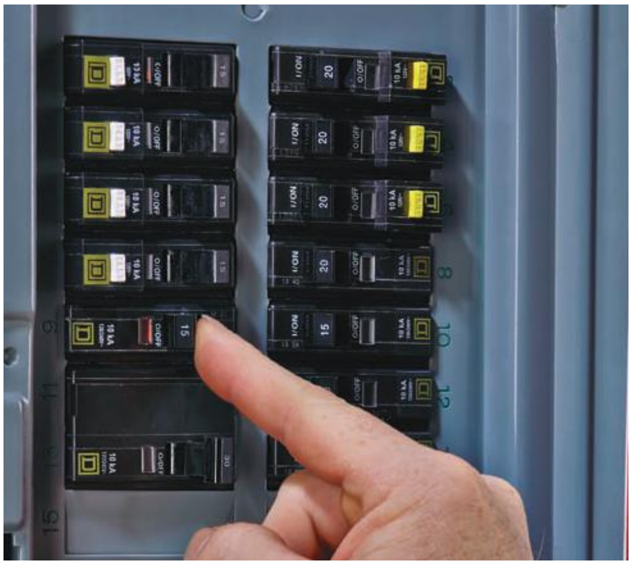

Shut power OFF at the main service panel or the main fuse box before beginning any work.

Create a circuit index and affix it to the inside of the door to your main service panel. Update it as needed.

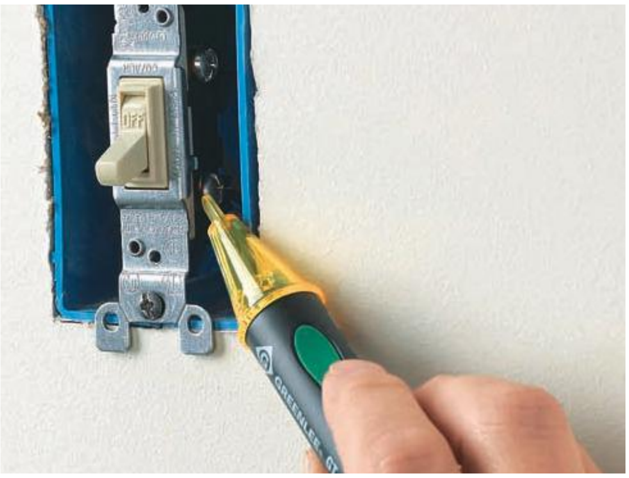

Confirm power is OFF by testing at the outlet, switch, or fixture with a voltage tester.

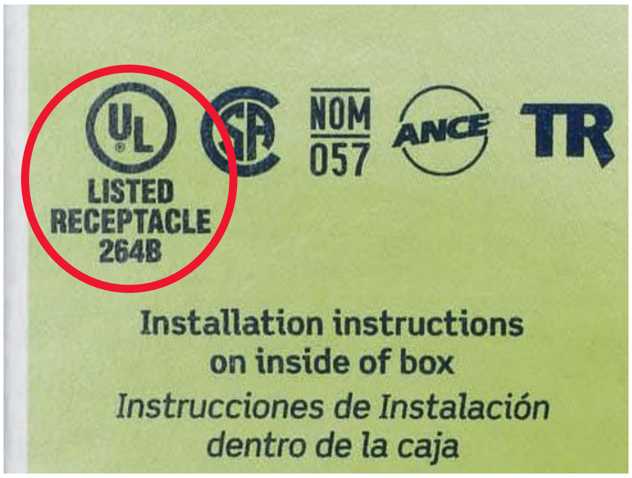

Use only UL-approved electrical parts or devices. These devices have been tested for safety by Underwriters Laboratories.

1st proofs for reference only
41940 - Advanced Home Wiring_001-063.indd 6
Pro-Vision C M Y K Text
10/30/14 10:06 AM
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
Page: 6

Wear rubber-soled shoes while working on electrical projects. On damp floors, stand on a rubber mat or dry wooden boards.

Use fiberglass or wood ladders when making routine household repairs near the service mast.

Extension cords are for temporary use only. Cords must be rated for the intended usage.

Electracraft
plug
fuses
TIME DELAY FUSE

Breakers and fuses must be compatible with the panel manufacturer and match the circuit capacity.

Never alter the prongs of a plug to fit a receptacle. If possible, install a new grounded receptacle.

Do not penetrate walls or ceilings without first shutting off electrical power to the circuits that may be hidden.

1st proofs for reference only
Edition
Page: 6
41940 - Advanced Home Wiring_001-063.indd 7
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 9:06 AM
Page: 7

Planning Your Project

Careful planning of a wiring project ensures you will have plenty of power for present and future needs. Whether you are adding circuits in a room addition, wiring a remodeled kitchen, or adding an outdoor circuit, consider all possible ways the space might be used, and plan for enough electrical service to meet peak needs.

For example, when wiring a room addition, remember that the way a room is used can change. In a room used as a spare bedroom, a single 15-amp circuit provides plenty of power, but if you ever choose to convert the same room to a family recreation space, you will need additional circuits.

When wiring a remodeled kitchen, it is a good idea to install circuits for an electric oven and countertop range, even if you do not have these electric appliances. Installing these circuits now makes it easy to convert from gas to electric appliances at a later date.

A large wiring project adds a considerable load to your main electrical service. In about 25 percent of all homes, some type of service upgrade is needed before new wiring can be installed. For example, many homeowners will need to replace an older 60-amp electrical service with a new service rated for 100 amps or more. This is a job for a licensed electrician but is well worth the investment. In other cases, the existing main service provides adequate power, but the main circuit breaker panel is too full to hold any new circuit breakers. In this case it is necessary to install a circuit breaker subpanel to provide room for hooking up added circuits. Installing a subpanel is a job most homeowners can do themselves (see pages 65 to 67).

This chapter gives an easy five-step method for determining your electrical needs and planning new circuits.

Five Steps for Planning a Wiring Project

1
200
MAIN PRINCIPAL

Examine your main service panel (see page 10). The amp rating of the electrical service and the size of the circuit breaker panel will help you determine if a service upgrade is needed.

2
NFPA 70
2014
nec
National Electrical Code
International Electrical Code Series
NFPA

Learn about codes (see pages 11 to 13). The National Electrical Code (NEC), and local electrical codes and building codes, provide guidelines for determining how much power and how many circuits your home needs. Your local electrical inspector can tell you which regulations apply to your job.

3

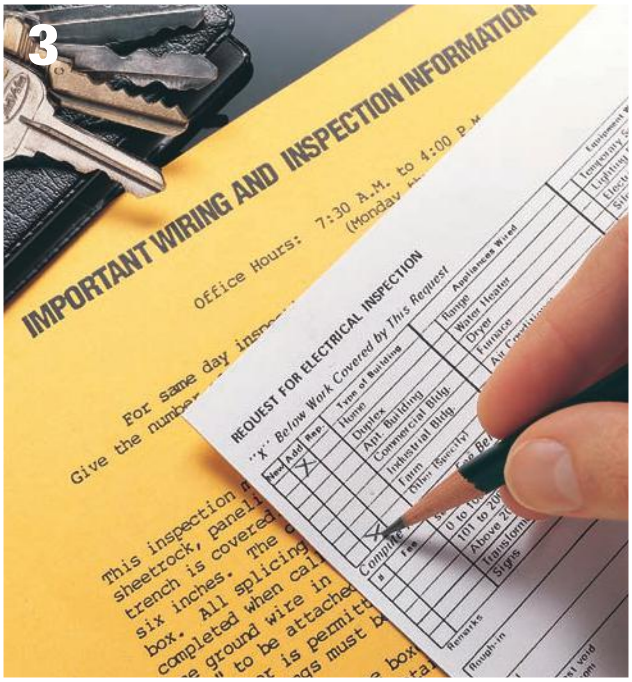
IMPORTANT WIRING AND INSPECTION INFORMATION
Office Hours: 7:30 A.M. to 4:00 P.M.
REQUEST FOR ELECTRICAL INSPECTION

Prepare for inspections (see pages 14 to 15). Remember that your work must be reviewed by your local electrical inspector. When planning your wiring project, always follow the inspector's guidelines for quality workmanship.

4

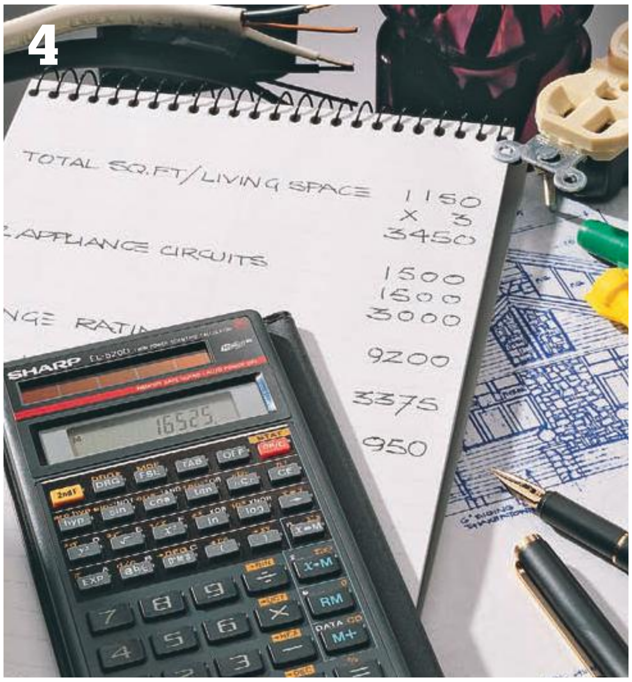
TOTAL SQ.FT/LIVING SPACE
APPLIANCE CIRCUITS
1150
x 3
3450
1500
1500
3000
9200
3375
950
SHARP

Evaluate electrical loads (see pages 16 to 19). New circuits put an added load on your electrical service. Make sure that the total load of the existing wiring and the planned new circuits does not exceed the service capacity or the capacity of the panel.

5

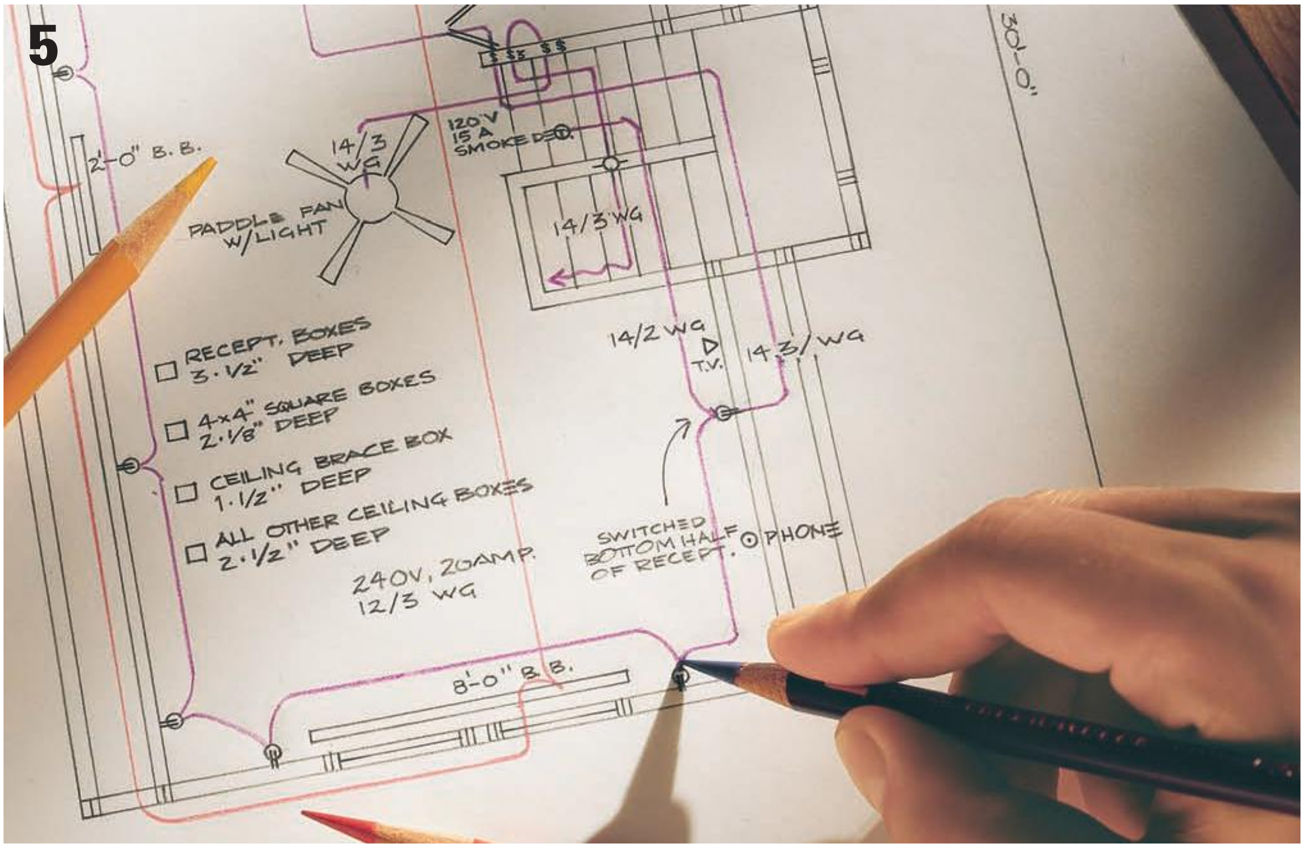
2'-0" B.B.
14/3 WG
120 V 15 A SMOKE DET.
PADDLE FAN W/LIGHT
14/3 WG
14/2 WG
T.V.
14/3/WG
☐ RECEPT. BOXES 3-1/2" DEEP
☐ 4×4" SQUARE BOXES 2-1/8" DEEP
☐ CEILING BRACE BOX 1-1/2" DEEP
☐ ALL OTHER CEILING BOXES 2-1/2" DEEP
240V, 20AMP 12/3 WG
SWITCHED BOTTOM HALF OF RECEPT.
PHONE
8'-0" B.B.
30'-0"

Draw a wiring diagram and get a permit (see pages 20 to 21). This wiring plan will help you organize your work.

Examine Your Main Service Panel

The first step in planning a new wiring project is to look in your main circuit breaker panel and find the size of the service by reading the amperage rating on the main circuit breaker. As you plan new circuits and evaluate electrical loads, knowing the size of the main service helps you determine if you need a service upgrade.

Also look for open circuit breaker slots in the panel. The number of open slots will determine if you need to add a circuit breaker subpanel.

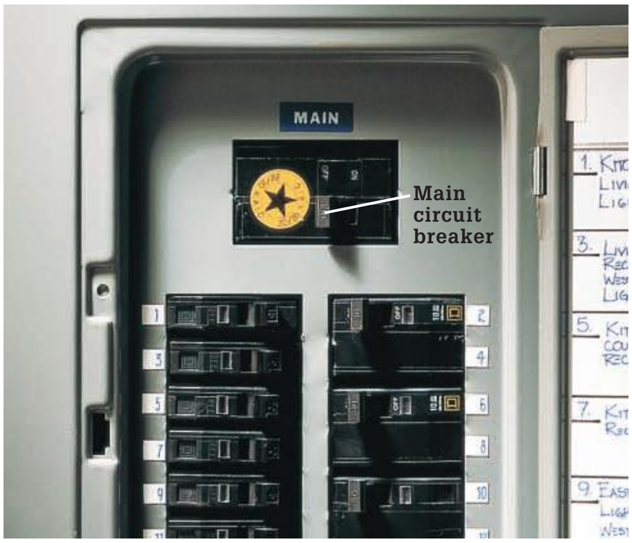

Find the service size by opening the main service panel and reading the amp rating printed on the main circuit breaker. In most cases, 100-amp service provides enough power to handle the added loads of projects such as the ones shown in this book. A service rated for 60 amps or less should be upgraded. Note: In some homes the main circuit breaker is located in a separate box.

Older service panels use fuses instead of circuit breakers. Have an electrician replace this type of panel with a circuit breaker panel that provides enough power and enough open breaker slots for the new circuits you are planning.

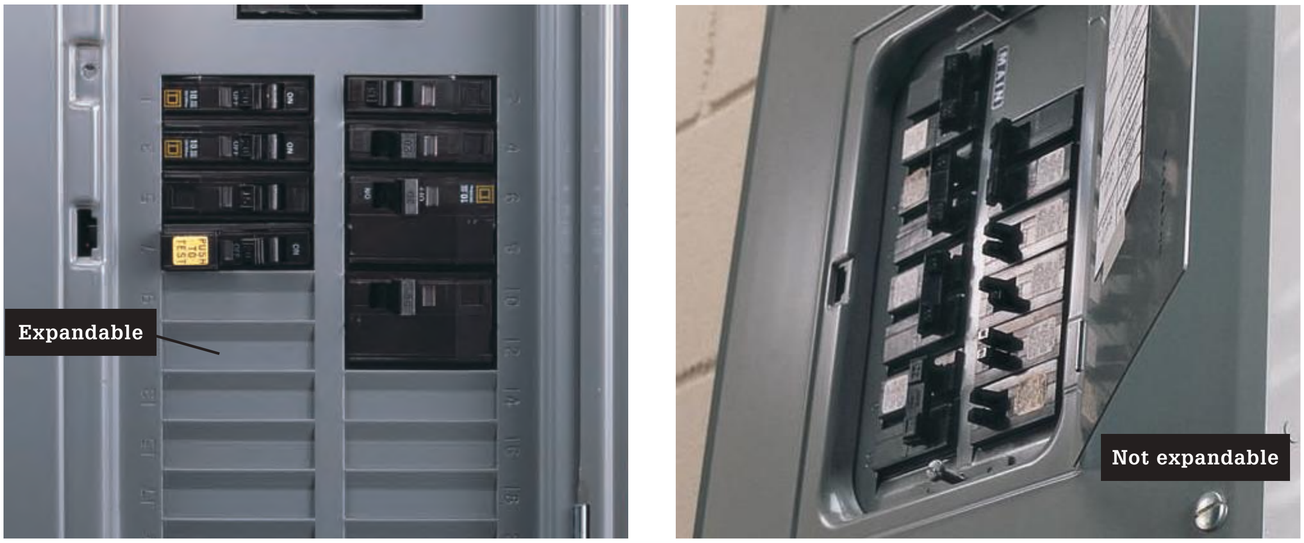

Look for open circuit breaker slots in the main circuit breaker panel or in a circuit breaker subpanel, if your home already has one. You will need one open slot for each 120-volt circuit you plan to install and two slots for each 240-volt circuit. If your main circuit breaker panel has no open breaker slots, install a subpanel (see pages 65 to 67) to provide room for connecting new circuits.

1st
proofs for reference only
41940 - Advanced Home Wiring_001-063.indd 10
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 9:06 AM
Page: 10
Pro-Visi

Learn About Codes

To ensure public safety, your community requires that you get a permit to install new wiring and have the work reviewed by an inspector. Electrical inspectors use the National Electrical Code (NEC) as the primary authority for evaluating wiring, but they also follow the local building code and electrical code standards.

Most communities use a version of the NEC that is not the most current version. Also, many communities make amendments to the NEC, and these amendments may affect your work.

As you begin planning new circuits, call or visit your local electrical inspector and discuss the project with him or her. The inspector can tell you which of the national and local code requirements apply to your job and may give you a packet of information summarizing these regulations. Later, when you apply to the inspector for a work permit, he or she will expect you to understand the local guidelines as well as a few basic NEC requirements.

The NEC is a set of standards that provides minimum safety requirements for wiring installations. It is revised every three years. The national code requirements for the projects shown in this book are thoroughly explained on the following pages. For more information, you can find copies of the current NEC, as well as a number of excellent handbooks based on the NEC, at libraries and bookstores.

In addition to being the final authority of code requirements, inspectors are electrical professionals with years of experience. Although they have busy schedules, most inspectors are happy to answer questions and help you design well-planned circuits.

Basic Electrical Code Requirements

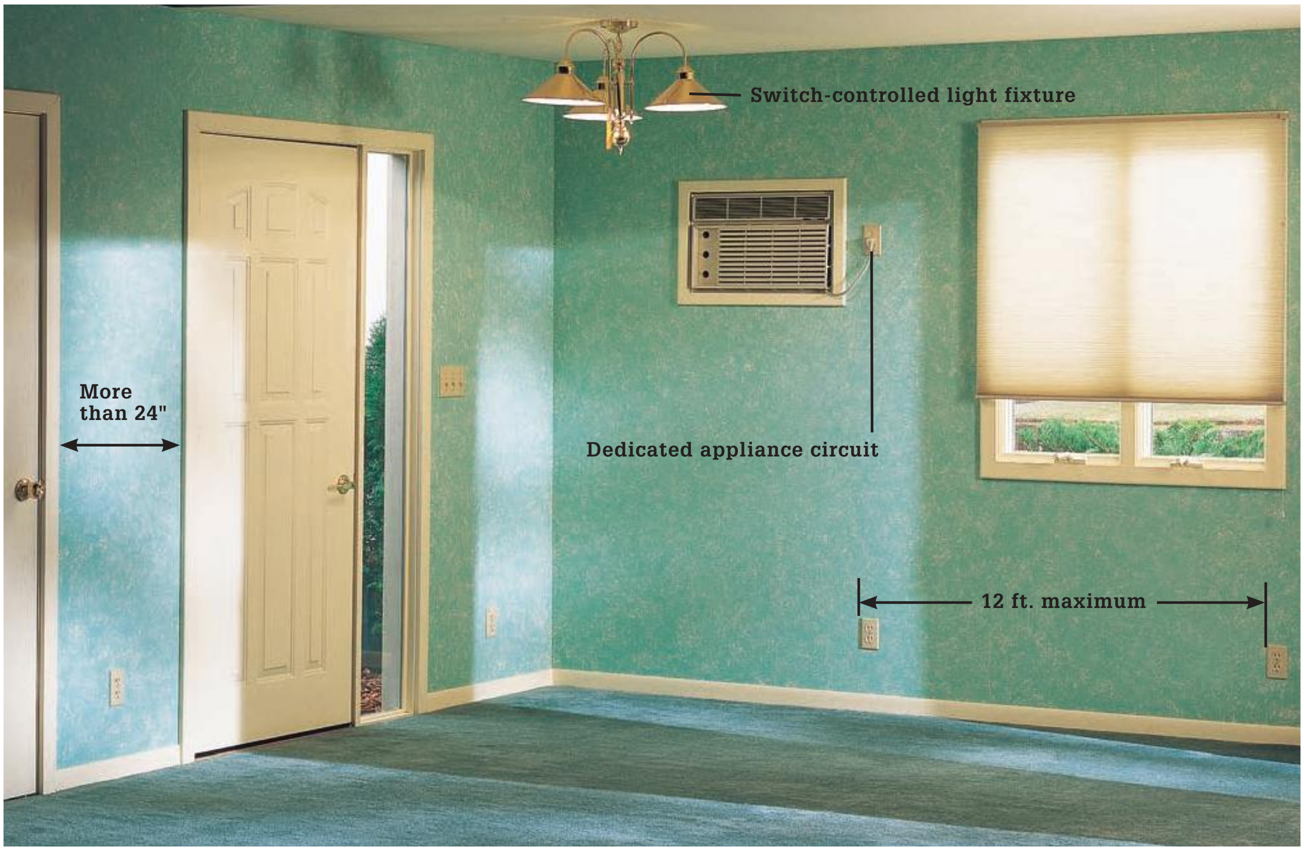

Electrical code requirements for living areas: Living areas need at least one 15-amp or 20-amp basic lighting/receptacle circuit for each 600 sq. ft. of living space and should have a dedicated circuit for each type of permanent appliance, such as an air conditioner, or a group of baseboard heaters. Receptacles on basic lighting/receptacle circuits should be spaced no more than 12 ft. apart. Many electricians and electrical inspectors recommend even closer spacing. Any wall more than 24" wide also needs a receptacle. Every room should have a wall switch at the point of entry to control either a ceiling or wall-mounted light or plug-in lamp. Kitchens and bathrooms must have a ceiling or wall-mounted light fixture.

1st proofs for reference only
Edition
age: 10
41940 - Advanced Home Wiring_001-063.indd 11
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 9:06 AM
Page: 11

Highlights of the National Electrical Code ▸

BY MATERIAL

Panels

- Maintain a minimum 30" wide by 36" deep of clearance in front of the panel.
- Match the amperage rating of the circuit when replacing fuses.
- Use handle ties on all 240-volt breakers and on 120-volt breakers protecting multi-wire branch circuits.
- Close all unused panel openings.
- Label each fuse and breaker clearly on the panel.

Electrical Boxes

- Use boxes that are large enough to accommodate the number of wires and devices in the box.
- Install all junction boxes so they remain accessible.
- Leave no gaps greater than ⅛" between wallboard and the front of electrical boxes.
- Place receptacle boxes flush with combustible surfaces.
- Leave a minimum of 3" of usable cable or wire extending past the front of the electrical box.

Wires & Cables

- Use wires that are large enough for the amperage rating of the circuit.
- Drill holes at least 2" from the edges of joists. Do not attach cables to the bottom edge of joists.
- Do not run cables diagonally between framing members.
- Use nail plates to protect cable that is run through holes drilled or cut into studs less than 1¼" from the front edge of a stud.
- Do not crimp cables sharply.
- Contain spliced wires or connections entirely in a plastic or metal electrical box.
- Use wire connectors to join wires.
- Use staples to fasten cables within 8" of an electrical box and every 54" along its run.
- Leave a minimum ¼" (maximum 1") of sheathing where cables enter an electrical box.
- Clamp cables and wires to electrical boxes with approved clamps. No clamp is necessary for one-gang plastic boxes if cables are stapled within 8".
- Connect only a single wire to a single screw terminal. Use pigtails to join more than one wire to a screw terminal.

Switches

- Use a switch-controlled receptacle in rooms without a built-in light fixture operated by a wall switch.
- Use three-way switches at the top and bottom on stairways with six risers or more.
- Use switches with grounding screws with plastic electrical boxes.
- Locate all wall switches within easy reach of the room entrance and not behind the door.
- Install a neutral wire in switch boxes.
- Use black or red wires to supply power to switched devices.

Receptacles

- Install receptacles on all walls 24" wide or greater.
- Install receptacles so a 6-ft. cord can be plugged in from any point along a wall or every 12 ft. along a wall.
- Include receptacles in any hallway that is 10 ft. long or longer.
- Use three-slot, grounded receptacles for all 15- or 20-amp, 120-volt branch circuits.
- Include a switch-controlled receptacle in rooms without a built-in light fixture operated by a wall switch.
- Install GFCI-protected circuits in bathrooms, kitchens, garages, crawl spaces, unfinished basements, and outdoor receptacle locations.

Light Fixtures

- Use mounting straps that are anchored to the electrical boxes to mount ceiling fixtures.
- Keep non-IC-rated recessed light fixtures 3" from insulation and ½" from combustibles.
- Include at least one switch-operated lighting outlet in every room.

Grounding

- Ground receptacles by connecting receptacle grounding screws to the circuit grounding wires.
- Use switches with grounding screws whenever possible. Always ground switches installed in plastic electrical boxes and all switches in kitchens, bathrooms, and basements.

1st proofs for reference only

41940 - Advanced Home Wiring_001-063.indd 12
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 9:08 AM
Page: 12

BY ROOM

Kitchens/Dining Rooms

- Install at least two 20-amp small appliance receptacle circuits.
- Install dedicated 15-amp, 120-volt circuits for dishwashers and food disposals (required by many local codes).
- Install GFCI protection for all countertop receptacles; receptacles behind fixed appliances do not need to be GFCI protected.
- Position receptacles for appliances that will be installed within cabinets, such as microwaves or food disposals, according to the manufacturer's instructions.
- Include receptacles on all counters wider than 12".
- Space receptacles a maximum of 48" apart above countertops and closer together in areas where many appliances will be used.
- Locate receptacles on the wall above the countertop not more than 20" above the countertop.
- Install at least one receptacle not more than 12" below the countertop on islands and peninsulas that are 12" × 24" or greater.
- Do not connect lights to the small appliance receptacle circuits.
- Install at least one wall or ceiling-mounted light fixture.

Bathrooms

- Install a separate 20-amp GFCI-protected circuit only for bathroom receptacles.
- Ground switches in bathrooms.
- Install at least one receptacle not more than 36" from each sink.
- Install at least one ceiling- or wall-mounted light fixture.

Utility/Laundry Rooms

- Install a separate 20-amp circuit for a washing machine.
- Install approved conduit for wiring in unfinished rooms.
- Use GFCI-protected circuits for 120-volt receptacles within 6 feet from a sink (including the washing machine receptacle).

Living, Entertainment, Bedrooms

- Install at least one 15- or 20-amp lighting/receptacle circuit for each 600 sq. ft. of living space.
- Install a dedicated circuit for each permanent appliance, such as an air conditioner or group of electric baseboard heaters.
- Use electrical boxes listed and labeled to support ceiling fans.
- Include receptacles on walls 24" wide or more.
- Space receptacles on walls in living and sleeping rooms a maximum of 12 ft. apart.
- Check with your local electrical inspector about requirements for installing smoke and carbon monoxide alarms during remodeling.

Outdoors

- Check for underground utilities before digging.
- Use UF cable or other wiring approved for wet locations for outdoor wiring.
- Run cable and wires in schedule 80 PVC plastic and other approved conduit, as required by local code.
- Install in-use rated weatherproof receptacle covers.
- Bury cables and wires run in conduit at least 18" deep; cable not in conduit must be buried at least 24" deep.
- Use weatherproof electrical boxes with watertight covers.
- Install GFCI-protected circuits for receptacles.
- Support boxes that are not attached to a building and that contain switches or receptacles using at least two pieces of conduit. Secure the conduit not more than 18 feet from the box. Locate the box at least 12" above the ground.

Stairs/Hallways

- Use three-way switches at the top and bottom on stairways with six risers or more.
- Include receptacles in any hallway that is 10 ft. long or longer.
- Position stairway lights so each step and landing is illuminated.

1st proofs for reference only
Edition
age: 12
41940 - Advanced Home Wiring_001-063.indd 13
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 9:08 AM
Page: 13

1st proofs for reference only

Prepare for Inspections

Electrical inspectors who issue the work permit for your wiring project will also visit your home to review the work. Make sure to allow time for these inspections as you plan the project. For most projects, inspectors make two visits.

The first inspection, called the rough-in, is done after the cables are run between the boxes but before the insulation, wallboard, switches, and fixtures are installed. The second inspection, called the final, is done after the walls and ceilings are finished and all electrical connections are made.

When preparing for the rough-in inspection, make sure the area is neat. Sweep up sawdust and clean up any pieces of scrap wire or cable insulation. Before inspecting the boxes and cables, inspectors will check to make sure all plumbing and other mechanical work is completed. Some electrical inspectors will ask to see your building and plumbing permits.

At the final inspection, inspectors check random boxes to make sure the wire connections are correct. If they see good workmanship at the selected boxes, the inspection will be over quickly. However, if they spot a problem, inspectors may choose to inspect every connection.

Inspectors have busy schedules, so it is a good idea to arrange for an inspection several days in advance. In addition to basic compliance with code, inspectors expect your work to meet their own standards for quality. When you apply for a work permit, make sure you understand what the inspectors will look for during inspections.

You cannot put new circuits into use legally until an inspector approves them at the final inspection. If you have planned carefully and done your work well, electrical inspections are routine visits that give you confidence in your own skills.

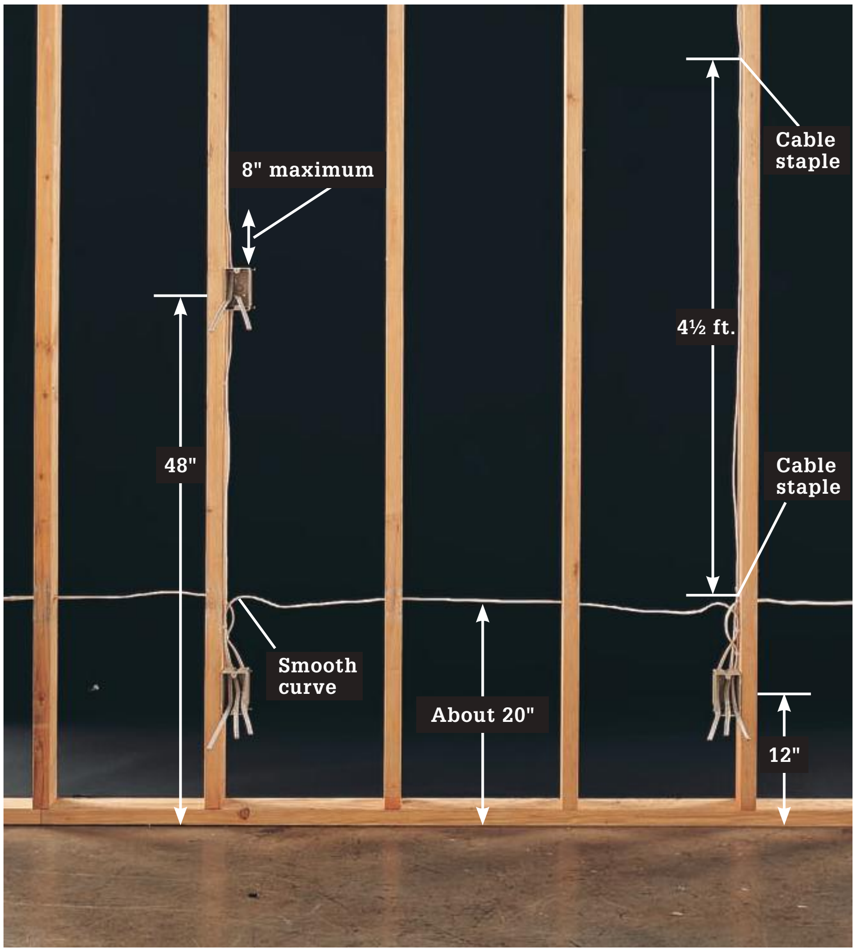

Inspectors may measure to see that electrical boxes are mounted at consistent heights. Height may not be dictated by code, but consistency is a sign of good workmanship. Measured from the center of the boxes, receptacles in living areas typically are located 12" above the finished floor and switches at 48". For special circumstances, inspectors allow you to alter these measurements. For example, you can install switches at 36" above the floor in a child's bedroom, or set receptacles at 24" to make them more convenient for someone using a wheelchair.

Inspectors will check cables to see that they are anchored by cable staples driven within 8" of each box and every 4½ ft. thereafter when they run along studs. When bending cables, form the wire in a smooth curve. Do not crimp cables sharply or install them diagonally between framing members. Some inspectors specify that cables running between receptacle boxes should be about 20" above the floor.

41940 - Advanced Home Wiring_001-063.indd 14
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 9:06 AM
Page: 14
Pro-Vis

What Inspectors Look for

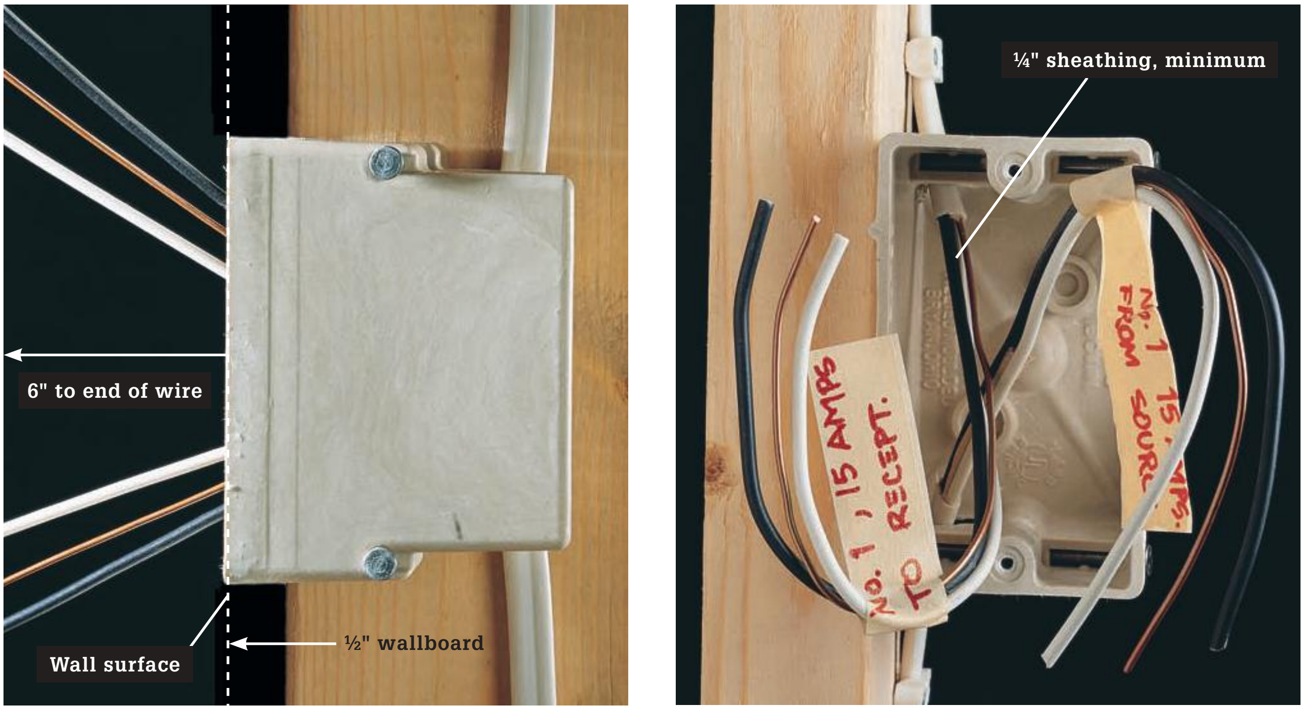
6" to end of wire
Wall surface
½" wallboard
¼" sheathing, minimum
No. 1, 15 AMPS TO RECEPT.
No. 1 FROM SOURCE 15 AMPS.

Electrical box faces should extend past the front of framing members so the boxes will be flush with finished walls (left). Inspectors will check to see that all boxes are large enough for the wires they contain. Cables should be cut and stripped back so that at least 3" of usable length extends past the front of the box and so that at least ¼" of sheathing reaches into the box (right). Label all cables to show which circuits they serve: inspectors recognize this as a mark of careful work. The labels also simplify the final hookups after the wallboard is installed.

Is your Receptacle Spacing Correct? ▸

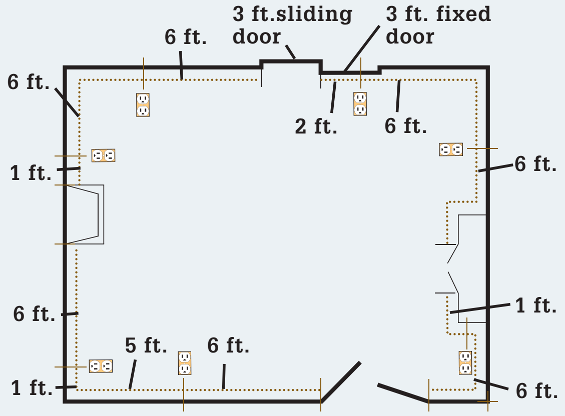
3 ft.sliding door
3 ft. fixed door
6 ft.
6 ft.
2 ft.
6 ft.
1 ft.
6 ft.
6 ft.
1 ft.
5 ft.
6 ft.
1 ft.
6 ft.

Example of receptacle spacing requirements in a typical room. Measure receptacle spacing distance along the wall line. Install receptacles along partial height walls and along balcony guards in lofts and similar areas.

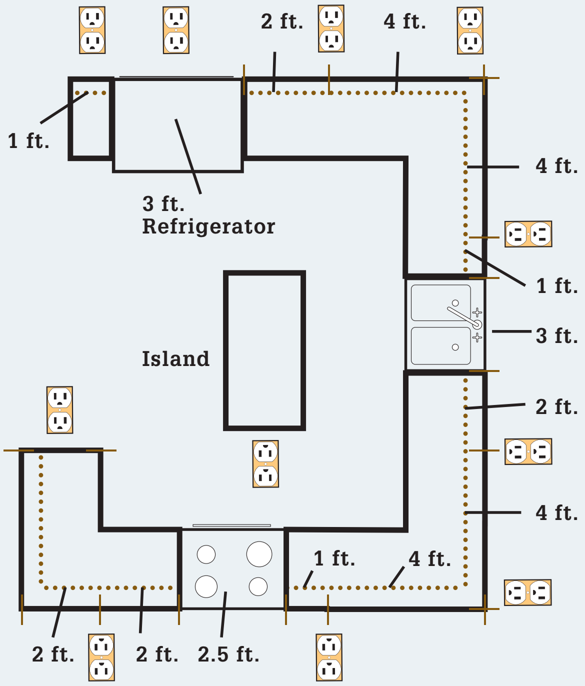
2 ft.
4 ft.
1 ft.
3 ft.
Refrigerator
4 ft.
1 ft.
3 ft.
Island
2 ft.
4 ft.
1 ft.
4 ft.
2 ft.
2 ft.
2.5 ft.

Example of countertop receptacle spacing in a typical kitchen (right).

1st proofs for reference only
Edition
age: 14
41940 - Advanced Home Wiring_001-063.indd 15
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 9:06 AM
Page: 15

Evaluate Electrical Loads

Before drawing a plan and applying for a work permit, make sure your home's electrical service provides enough power to handle the added load of the new circuits. In a safe wiring system, the current drawn by fixtures and appliances never exceeds the main service capacity.

To evaluate electrical loads, use whatever evaluation method is recommended by your electrical inspector. Include the load for all existing wiring as well as that for proposed new wiring when making your evaluation.

Most of the light fixtures and plug-in appliances in your home are evaluated as part of general allowances for basic lighting/receptacle circuits and small-appliance circuits. However, appliances that are permanently installed require their own dedicated circuits. The electrical loads for these appliances are added in separately when evaluating wiring.

If your evaluation shows that the load exceeds the main service capacity, you must have an electrician upgrade the main service before you can install new wiring. An electrical service upgrade is a worthwhile investment that improves the value of your home and provides plenty of power for present and future wiring projects.

Amperage ▸

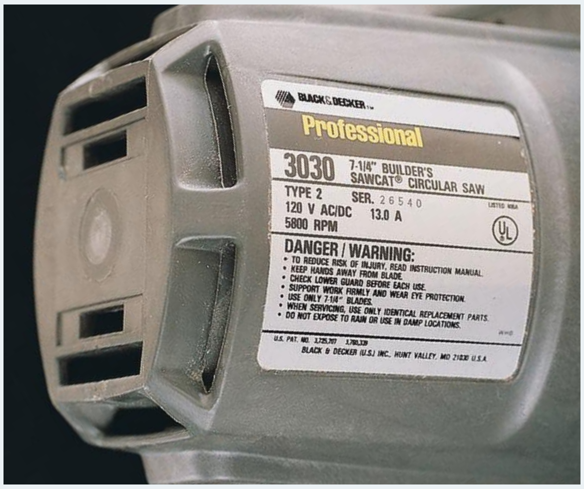

Amperage rating can be used to find the wattage of an appliance. Multiply the amperage by the voltage of the circuit. For example, a 13-amp, 120-volt circular saw is rated for 1,560 watts.

AMPS × VOLTS	TOTAL CAPACITY	SAFE CAPACITY
15 A × 120 V =	1,800 watts	1,440 watts
20 A × 120 V =	2,400 watts	1,920 watts
25 A × 120 V =	3,000 watts	2,400 watts
30 A × 120 V =	3,600 watts	2,880 watts
20 A × 240 V =	4,800 watts	3,840 watts
30 A × 240 V =	7,200 watts	5,760 watts

1st proofs for reference only
41940 - Advanced Home Wiring_001-063.indd 16
10/10/14 9:06 AM
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
Page: 16
Pro-Vision C M Y K Text

Calculating Loads

Add 1,500 watts for each small appliance circuit required by the local electrical code. In most communities, three such circuits are required—two in the kitchen and one for the laundry—for a total of 4,500 watts. No further calculations are needed for appliances that plug into small-appliance or basic lighting/receptacle circuits.

S.A.49022
KILOWATTS
240 VOLTS
12.3
LIS
923
Nameplate

If the nameplate gives the rating in kilowatts, find the watts by multiplying kilowatts times 1,000. If an appliance lists only amps, find watts by multiplying the amps times the voltage—either 120 or 240 volts.

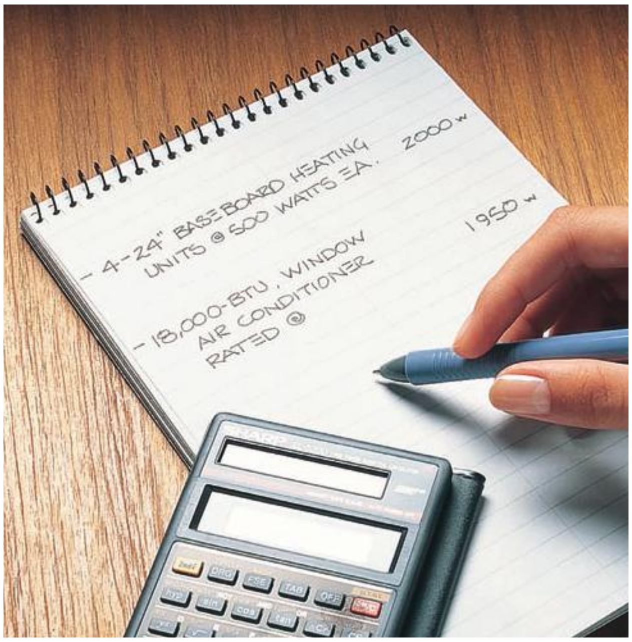
- 4-24" BASEBOARD HEATING UNITS @ 500 WATTS EA.
2000 w
- 18,000-BTU, WINDOW AIR CONDITIONER RATED @
1950 w

Air-conditioning and heating appliances are not used at the same time, so figure in only the larger of these two numbers when evaluating your home's electrical load.

Fixed Devices

Do not connect one or more fixed devices that in total exceed 50 percent of a multiple outlet branch circuit's amperage rating. Fixed devices do not include light fixtures. This means that that all fixed devices (such as a permanently wired disposal or hot water circulating pump) on a multiple outlet branch circuit may not exceed 7.5 amps (about 900 watts) on a 15-amp multiple outlet branch circuit and may not exceed 10 amps (about 1,200 watts) on a 20-amp multiple outlet branch circuit.

1st proofs for reference only
Edition
age: 16
41940 - Advanced Home Wiring_001-063.indd 17
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 9:08 AM
Page: 17

1st
proofs for reference only

Locating Wattage

23W 120V 60Hz
335mA 2700K
CF23EL/MINI/827
Hg

Light bulb wattage ratings are printed on the top of the bulb. If a light fixture has more than one bulb, remember to add the wattages of all the bulbs to find the total wattage of the fixture.

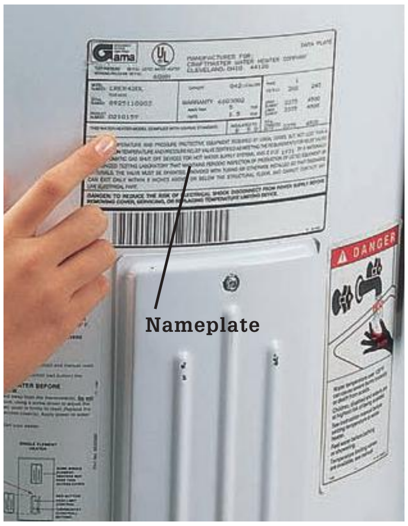

Electric water heaters are permanent appliances that require their own dedicated 30-amp, 240-volt circuits. Most water heaters are rated between 3,500 and 4,500 watts. If the nameplate lists several wattage ratings, use the one labeled "Total Connected Wattage" when figuring electrical loads.

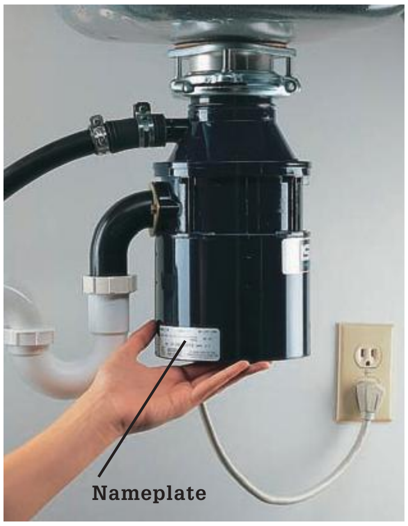

Food disposers are considered permanent appliances and may require their own dedicated 15-amp, 120-volt circuits. Most disposers are rated between 500 and 900 watts.

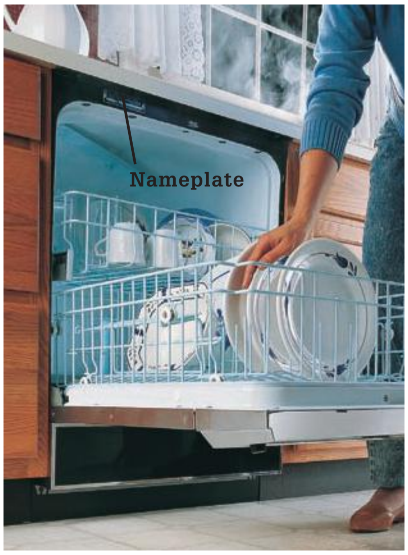

Dishwashers installed permanently under a countertop may need a dedicated 15-amp, 120-volt circuits. Dishwasher ratings are usually between 1,000 and 1,500 watts. Portable dishwashers are regarded as part of small appliance circuits and are not added in when figuring loads.

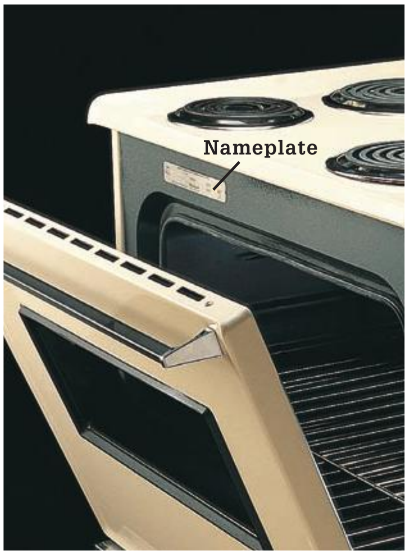

Electric ranges can be rated for as little as 3,000 watts or as much as 12,000 watts. They require dedicated 120/240-volt circuits. Find the exact wattage rating by reading the nameplate found inside the oven door or on the back of the unit.

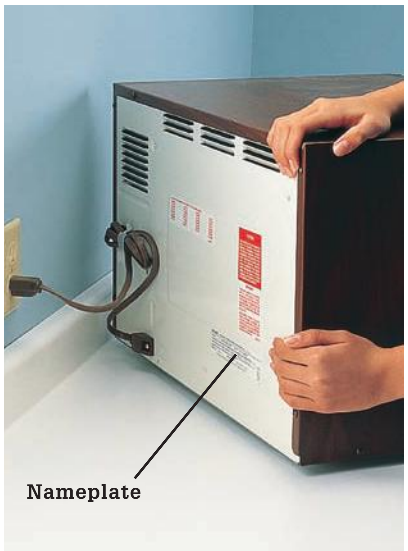

Microwave ovens are regarded as permanent appliances. Add in its wattage rating when calculating loads. The nameplate is found on the back of the cabinet or inside the front door. Most microwave ovens are rated between 500 and 1,200 watts.

41940 - Advanced Home Wiring_001-063.indd 18
10/10/14 9:06 AM
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
Page: 18
Pro-Vis

Freezers are permanent appliances that may need a dedicated 15- or 20-amp, 120-volt circuits. Freezer ratings are usually between 240 and 480 watts. But combination refrigerator-freezers rated for 1,000 watts or less are plugged into small appliance circuits and do not need their own dedicated circuits. The nameplate for a freezer is found inside the door or on the back of the unit, just below the door seal.

Electric clothes dryers are permanent appliances that need dedicated 30-amp, 120/240-volt circuits. The wattage rating is printed on the nameplate inside the dryer door. Use 5,000 watts as a minimum, regardless of the printed rating. Washing machines and gas-heat clothes dryers with electric tumbler motors do not need dedicated circuits. They plug into the 20-amp small-appliance circuit in the laundry room.

Forced-air furnaces and heat pump air handlers have electric fans and are considered permanent appliances. They require dedicated 15-amp, 120-volt circuits. Include the fan wattage rating, printed on a nameplate inside the control panel, when figuring wattage loads for heating. You should also include the wattage rating for heat pump backup heating coils.

A central air conditioner requires a dedicated 240-volt circuit. Estimate its wattage rating by adding the numbers labeled RLA and FLA on the air conditioner's metal plate. Multiply the RLA+FLA by 240.

Window air conditioners, both 120-volt and 240-volt types, are permanent appliances that require dedicated circuits. The wattage rating, which can range from 500 to 2,000 watts, is found on the nameplate located inside the front grill. Make sure to include all window air conditioners in your evaluation.

Electric room heaters that are permanently installed require a dedicated circuit and must be figured into the load calculations. Use the maximum wattage rating printed inside the cover. In general, 240-volt baseboard-type heaters are rated for 180 to 250 watts for each linear foot.

1st
proofs for reference only

Edition
age: 18
41940 - Advanced Home Wiring_001-063.indd 19
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 9:06 AM
Page: 19

GFCI & AFCI Breakers

Understanding the difference between GFCI (ground-fault circuit interrupter) and AFCI (arc fault circuit interrupter) is tricky for most homeowners. Essentially it comes down to this: Arc-fault interrupters keep your house from burning down; ground-fault interrupters keep people from being electrocuted.

The National Electric Code (NEC) requires that an AFCI breaker be installed on most branch circuits that supply outlets or fixtures in newly constructed homes. The NEC also requires adding AFCI protection to these circuits when you add new circuits and modify or extend existing circuits. They're a prudent precaution in any home, especially if it has older wiring. AFCI breakers will not interfere with the operation of GFCI receptacles, so it is safe to install an AFCI breaker on a circuit that contains GFCI receptacles. For a discussion on codes that concern AFCI and GFCI breakers see pages 120 to 121.

GROUND-FAULT CIRCUIT-INTERRUPTERS

A GFCI is an important safety device that disconnects a circuit in the event of a ground fault (when current takes a path other than the neutral back to the panel).

On new construction, GFCI protection is required for receptacles in these locations: kitchen counter tops, bathrooms, garages, unfinished basements, crawlspaces, outdoors, within six feet of sinks, and in unfinished accessory buildings such as storage and work sheds. In general it is a good practice to protect all receptacle and fixture locations that could encounter damp or wet circumstances.

Tools & Materials ▸

Insulated screwdriver
Circuit tester
Combination tool
AFCI or GFCI breaker

ARC-FAULT CIRCUIT INTERRUPTERS

AFCIs detect arcing (sparks) that can cause fires between and along damaged wires. AFCI protection is required for 15- and 20-amp, 120-volt circuits that serve living rooms, family rooms, dens, parlors, libraries, dining rooms, bedrooms, sun rooms, kitchens, laundry areas, closets, and hallways. AFCI protection is not required for circuits serving bathrooms, garages, the exterior of the home, appliances such as furnaces and air handlers.

The easiest way to provide AFCI protection for a circuit is to install an AFCI circuit breaker labeled as a "combination" device in the electrical panel. The 2014 NEC allows several alternate methods of providing AFCI protection, but you should consult an electrician before using these alternate methods. You should install combination AFCI circuit breakers when installing new circuits that require AFCI protection. You should install either combination AFCI circuit breakers or AFCI receptacles when you modify, replace, or extend an existing circuit that requires AFCI protection.

AFCI breakers (left) are similar in appearance to GFCI breakers (right), but they function differently. AFCI breakers trip when they sense an arc fault. GFCI breakers trip when they sense fault between the hot wire and the ground.

An AFCI-protected receptacle

1st proofs for reference only

41940 - Advanced Home Wiring_001-063.indd 48
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/30/14 10:06 AM
Page: 48
Pro-Vis

How to Install an AFCI or GFCI Breaker

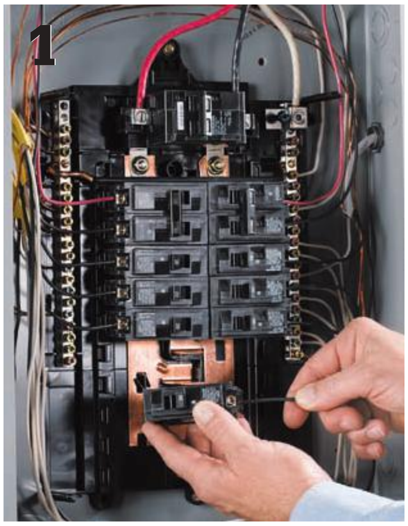

Locate the breaker for the circuit you'd like to protect. Turn off the main circuit breaker. Remove the cover from the panel, and test to ensure that power is off. Remove the breaker you want to replace from the panel. Remove the black wire from the LOAD terminal of the breaker.

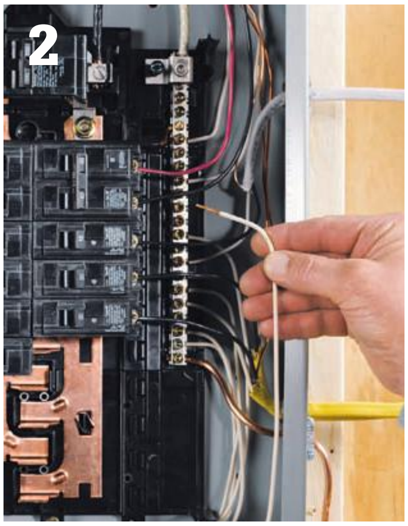

Find the white wire on the circuit you want to protect, and remove it from the neutral bus bar.

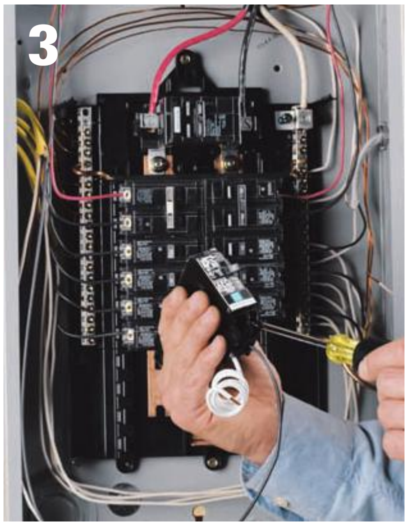

Flip the handle of the new AFCI or GFCI breaker to OFF. Loosen both of the breaker's terminal screws. Connect the white circuit wire to the breaker terminal labeled PANEL NEUTRAL. Connect the black circuit wire to the breaker terminal labeled LOAD POWER.

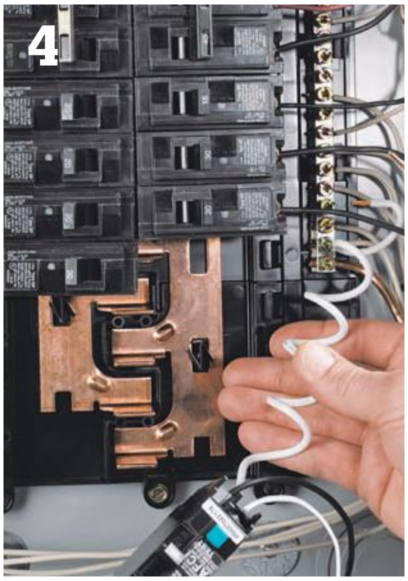

Connect the new breaker's coiled white wire to the neutral bus bar on the service panel.

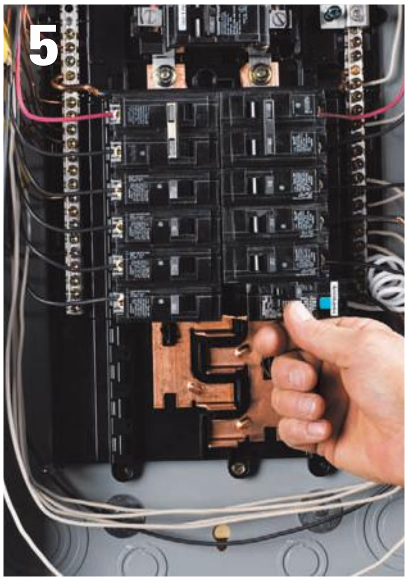

Make sure all the connections are tight. Snap the new breaker into the bus bar.

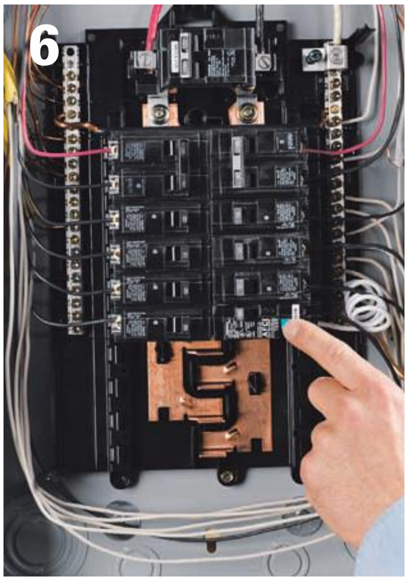

Turn the main breaker on. Turn off and unplug all fixtures and appliances on the AFCI or GFCI breaker circuit. Turn the AFCI or GFCI breaker on. Press the test button. If the breaker is wired correctly, the breaker trips open. If it doesn't trip, check all connections or consult an electrician. Replace the panel cover.

1st
proofs for reference only

Edition
age: 48
41940 - Advanced Home Wiring_001-063.indd 49
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 9:06 AM
Page: 49

Replacing a Service Panel

Only a generation ago, fuse boxes were commonplace. But as our demands for power increased, homeowners replaced the 60-amp boxes with larger, safer, and more reliable circuit breaker panels. Typical new homes were built with perfectly adequate 100-amp load centers. But today, as average home size has risen to more than 2,500 sq. ft. and the number of home electronics has risen exponentially, 100 amps is often inadequate service. As a result, many homeowners have upgraded to 200-amp service, and new single-family homes often include 250 amps or even 400 amp service.

Upgrading your electrical service panel from 100 amps to 200 amps is an ambitious project that requires a lot of forethought. The first step is to obtain a permit. When you are ready to begin, you will need to have your utility company disconnect your house from electrical service at the transformer that feeds your house. When you schedule this, talk to your utility company about the size of your service drop or lateral. That may need to be upgraded too. Not only does this involve working them into your schedule, it means you will have no power during the project. You can rent a portable generator to provide a circuit or two, or you can run a couple extension cords from a friendly neighbor. But unless you are a very fast worker, plan on being without power for at least one to two days while the project is in process.

Also check with your utility company to make sure you know what equipment is theirs and what belongs to you. In most cases, the electric meter and everything on the street side belongs to the power company, and the meter base and everything on the house side is yours. Be aware that if you tamper with the sealed meter in any way, you likely will be fined. Utility companies will not re-energize your system without approval from your inspecting agency.

Upgrading a service panel is a major project. Do not hesitate to call for help at any point if you're unsure what to do.

Tools & Materials ▸

- 200-amp load center (service panel)
- 200-amp bypass meter base
- Circuit breakers (AFCI if required by local code)
- Schedule 80 or RMC conduit and fittings
- Weatherhead
- Service cable
- Circuit wires
- Plywood backer board
- Screwdrivers
- Drill/driver
- Tape
- Allen wrench
- Circuit tester
- Multimeter

Modern homeowners consume more power than our forebears, and it is often necessary to upgrade the electrical service to keep pace. While homeowners are not allowed to make the final electrical service connections, removing the old panel and installing the new panel and meter base yourself can save you hundreds or even thousands of dollars.

1st
proofs for reference only

41940 - Advanced Home Wiring_001-063.indd 50
10/30/14 10:06 AM
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
Page: 50
Pro-Vis

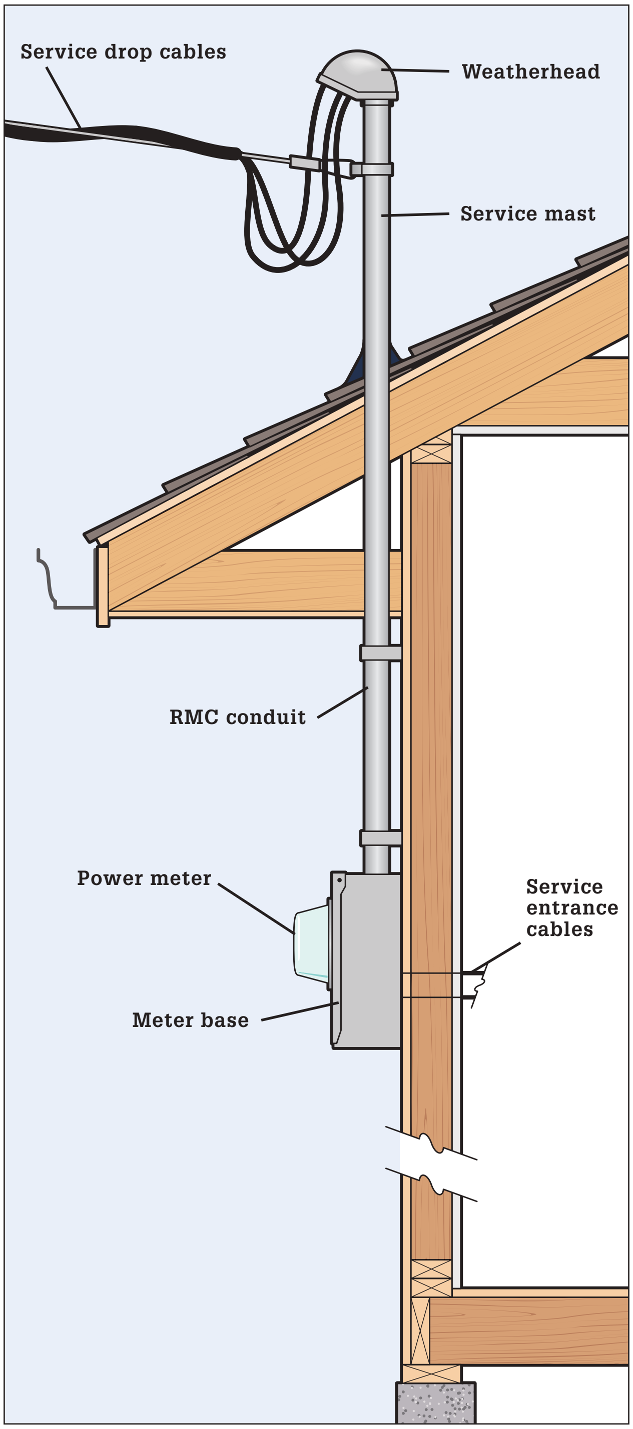
Service drop cables
Weatherhead
Service mast
RMC conduit
Power meter
Service entrance cables
Meter base

Aboveground service drop. In this common configuration, the service cables from the closest transformer (called the service drop) connect to service entrance wires near the weatherhead. This connection is called the service point and is where your property usually begins. The service entrance wires from the weatherhead are routed to a power meter that's owned by your utility company but is housed in a base that's considered your property. From the meter the service entrance wires enter your house through the wall and are routed to the main service panel, where they are connected to the main circuit breaker.

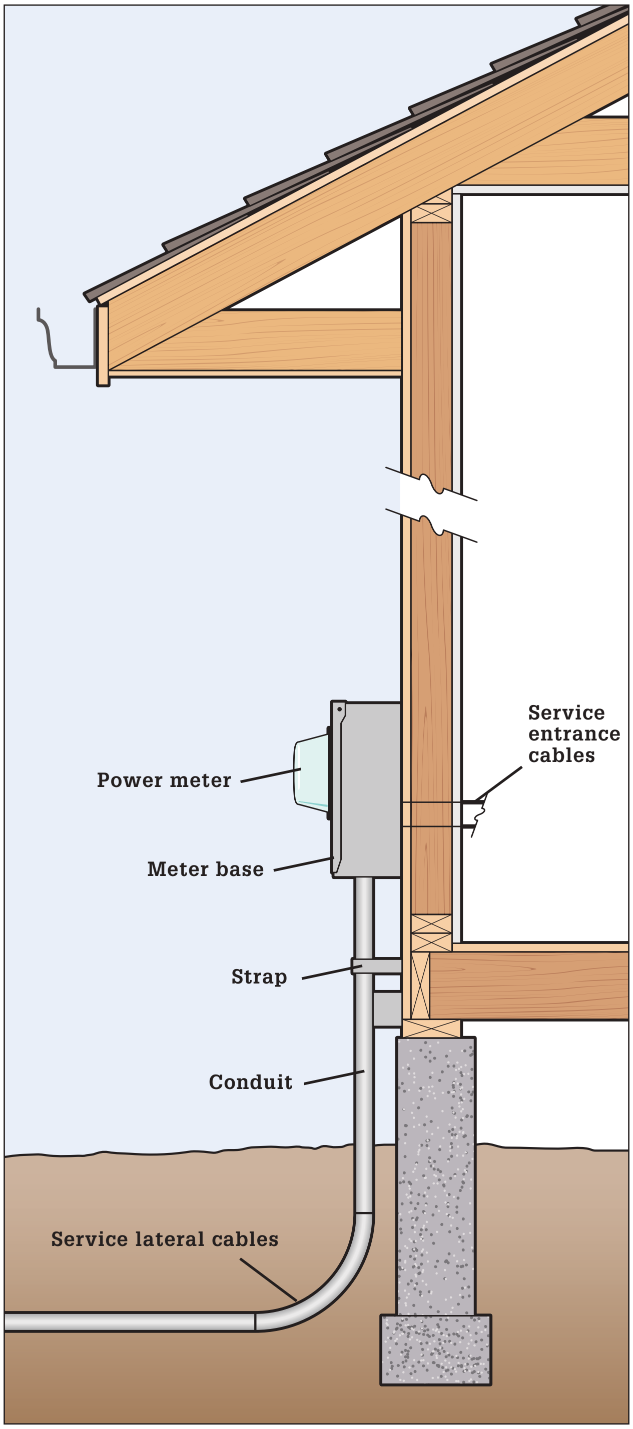
Power meter
Service entrance cables
Meter base
Strap
Conduit
Service lateral cables

Underground service lateral. Increasingly, homebuilders are choosing to have power supplied to their new homes underground instead of an overhead service drop. Running the cables in the ground eliminates problems with power outages caused by ice accumulation or fallen trees, but it entails a completely different set of cable and conduit requirements. For the homeowner, however, the differences are minimal, because the hookups are identical once the power service reaches the meter.

1st proofs for reference only
Edition
age: 50
41940 - Advanced Home Wiring_001-063.indd 51
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 9:06 AM
Page: 51

Locating Your New Panel ▸

Local codes dictate where the main service panel may be placed relative to other parts of your home. Although the codes vary (and always take precedence), national codes stipulate that a service panel (or any other distribution panel) may not be located near flammable materials, in a bathroom, clothes closet or other area designated for storage, above stairway steps, or directly above a workbench or other permanent work station or appliance. The panel also can't be located in a crawl space. If you are installing a new service entry hookup, there are many regulations regarding height of the service drop and the meter. Contact your local inspections office for specific regulations.

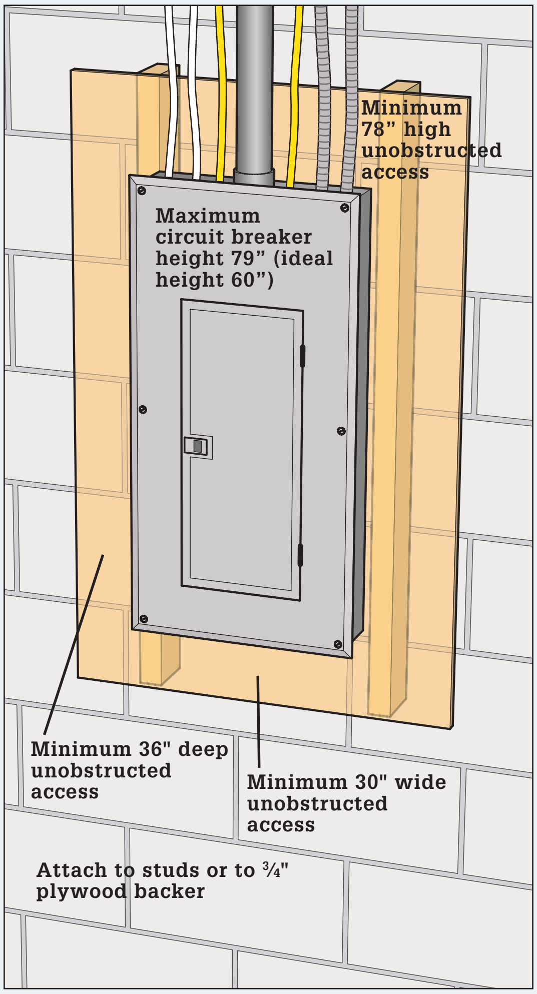
Minimum 78" high unobstructed access
Maximum circuit breaker height 79" (ideal height 60")
Minimum 36" deep unobstructed access
Minimum 30" wide unobstructed access
Attach to studs or to ¾" plywood backer

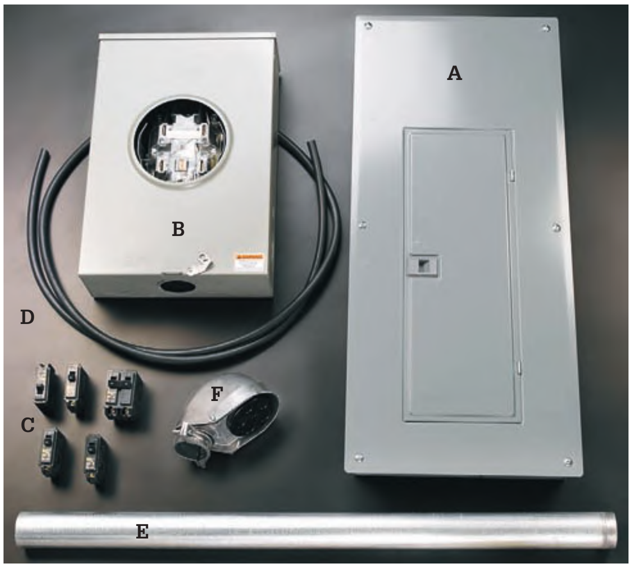

All the equipment you'll need to upgrade your main panel is sold at most larger building centers. It includes (A) a new 200-amp panel; (B) a 200-amp bypass meter base (also called a socket); (C) individual circuit breakers (if your new panel is the same brand as your old one you may be able to reuse the old breakers); (D) new, THW, THHW, THWN-2. RHW, RHW-2, XHHW (2/0 copper seen here); (E) 2" dia. rigid metallic conduit; (F) weatherhead shroud for mast.

Meter
Shutoff switch

The main circuit breaker (called the service equipment) may need to be located outside next to the electric meter if your main panel is too far away from the point where the service cable enters your house. The maximum distance allowed varies widely, from as little as 3 ft. to more than 10 ft. Wiring the service cable through the shutoff has the effect of transforming your main panel into a subpanel, which will impact how the neutral and ground wires are attached (see Subpanels, pages 64 to 65).

1st
proofs for reference only
41940 - Advanced Home Wiring_001-063.indd 52
10/10/14 9:06 AM
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
Page: 52
Pro-Vis

How to Replace a Main Panel

Shut off power to the house at the transformer. This must be done by a technician who is certified by your utility company. Also have the utility worker remove the old meter from the base. It is against the law for a homeowner to break the seal on the meter.

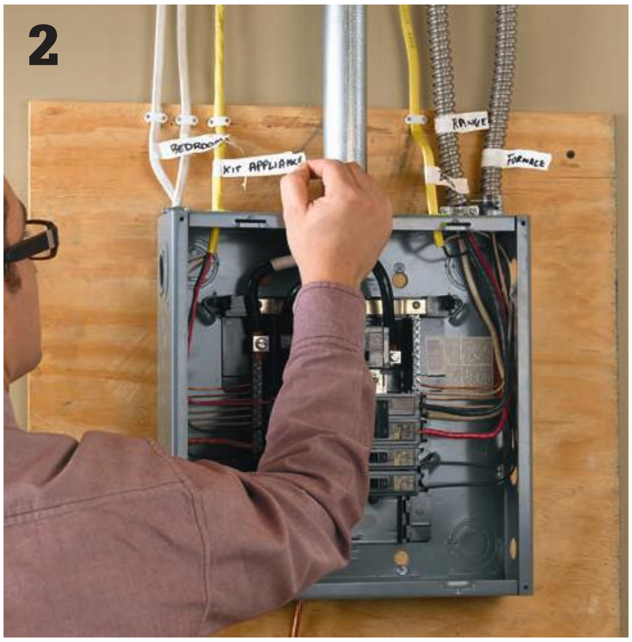

Label all incoming circuit wires before disconnecting them. Labels should be written clearly on tape that is attached to the cables outside of the existing service panel. Test the circuits before starting to make sure they are labeled correctly.

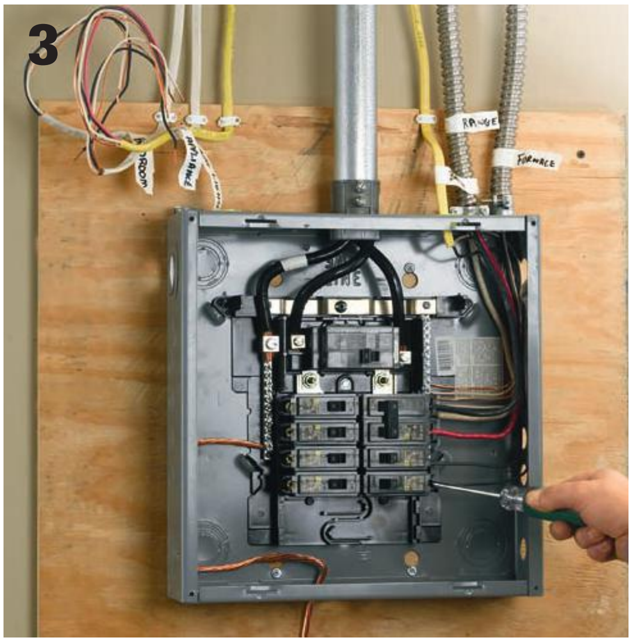

Disconnect incoming circuit wires from breakers, grounding bar, and neutral bus bar. Also disconnect cable clamps at the knockouts on the panel box. Retract all circuit wires from the service panel and coil it up neatly, with the labels clearly visible.

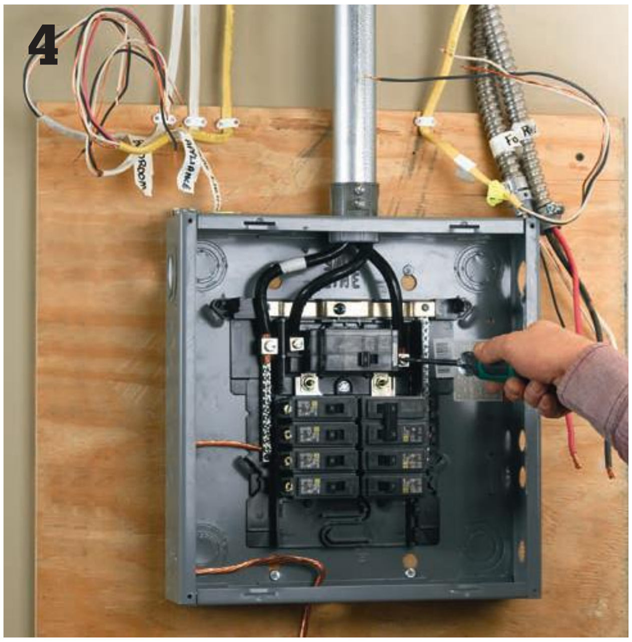

Unscrew the lugs securing the service entry cables at the top of the panel. For 240-volt service you will find two heavy-gauge SE cables, probably with black sheathing. Each cable carries 120 volts of electricity. A neutral service cable, usually of smaller gauge than the SE cables, will be attached to the neutral bus bar. This cable returns current to the source.

(continued)

1st
proofs for reference only
Edition
age: 52
41940 - Advanced Home Wiring_001-063.indd 53
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 9:06 AM
Page: 53

5

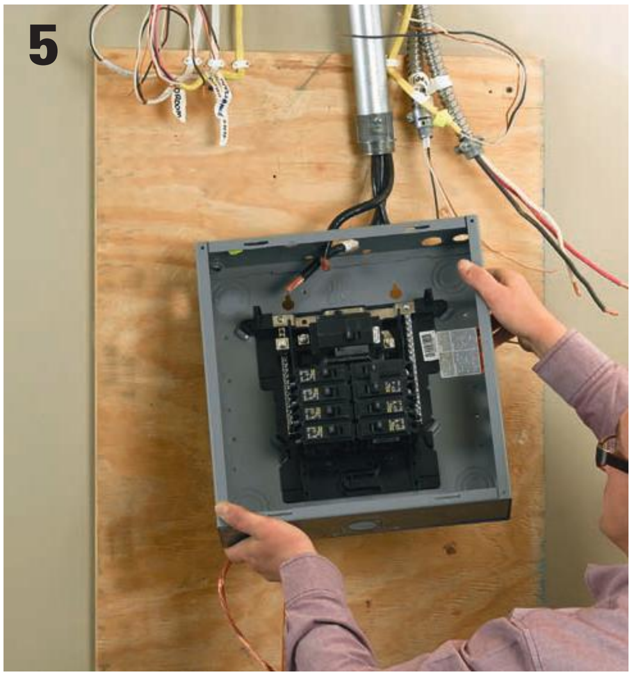

Remove the old service panel box. Boxes are rated for a maximum current capacity; and if you are upgrading, the components in the old box will be undersized for the new service levels. The new box will have a greater number of circuit slots as well.

6

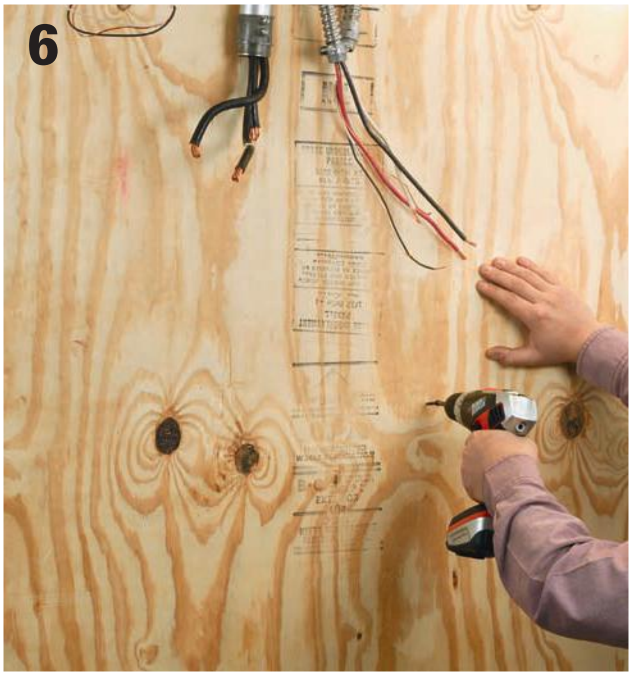

Replace the old panel backer board with a larger board in the installation area (see sidebar, page 52). A piece of ¾" plywood is typical. Make sure the board is well secured at wall framing members.

7

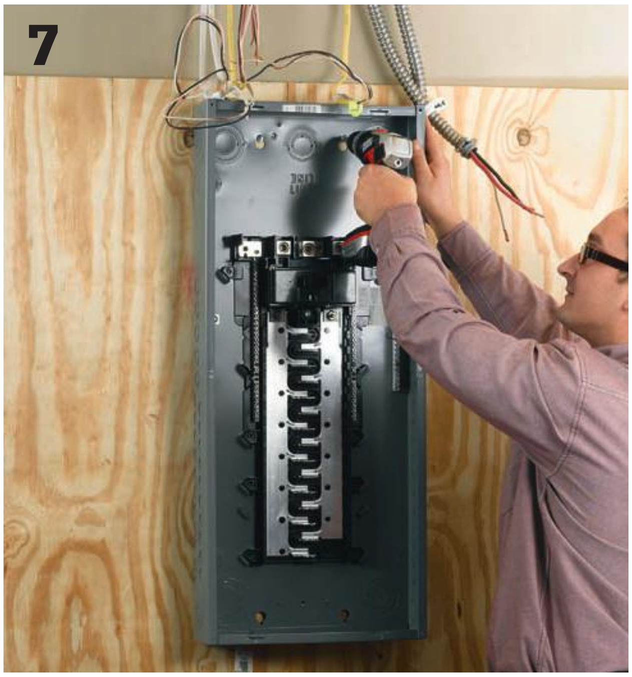

Attach the new service panel box to the backer board, making sure that at least two screws are driven through the backer and into wall studs. Drill clearance holes in the back of the box at stud locations if necessary. Use roundhead screws that do not have tapered shanks so the screwhead seats flat against the panel.

8

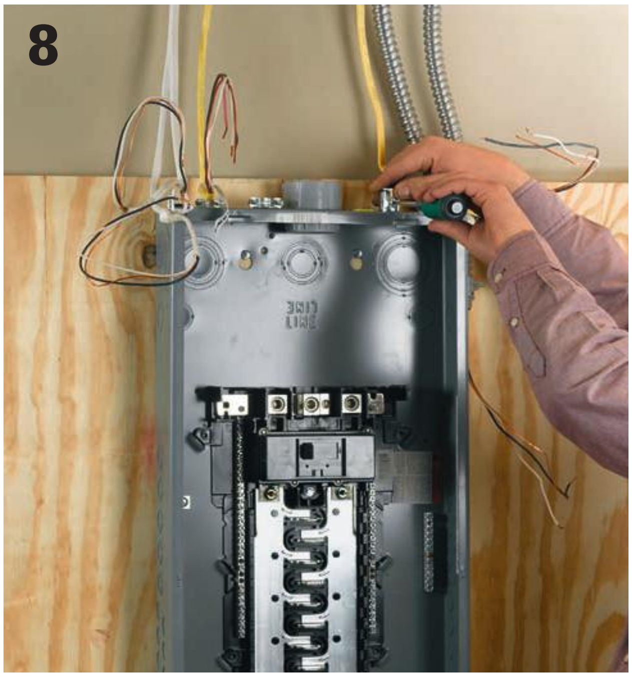

Attach properly sized cable clamps to the box at the knockout holes. Install one cable per knockout in this type of installation and plan carefully to avoid removing knockouts that you do not need to remove (if you do make a mistake, you can fill the knockout hole with a plug).

1st proofs for reference only

41940 - Advanced Home Wiring_001-063.indd 54

Pro-Vision C M Y K Text

10/10/14 9:08 AM

Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition

Page: 54

Pro-Vis

Splicing in the Box ▸

Some wiring codes allow you to make splices inside the panel box if the circuit wire is too short. Use the correct wire cap and wind electrical tape over the conductors where they enter the cap. If your municipality does not allow splices in the panel box, you'll have to rectify a short cable by splicing it in a junction box before it reaches the panel and then replacing the cable with a longer section for the end of the run. Make sure each circuit line has at least 12" of slack.

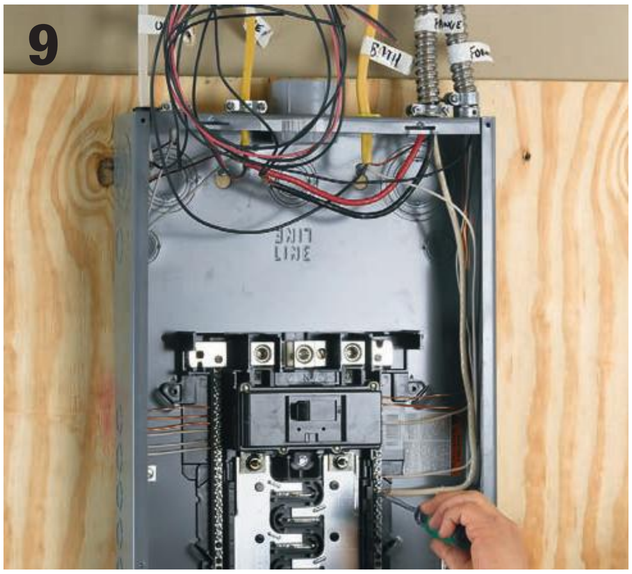

9

Attach the white neutral from each circuit cable to the neutral bus bar. Most panels have a preinstalled neutral bus bar, but in some cases you may need to purchase the bar separately and attach it to the panel back. The panel should also have a separate grounding bar that you also may need to purchase separately. Attach the grounds as well.

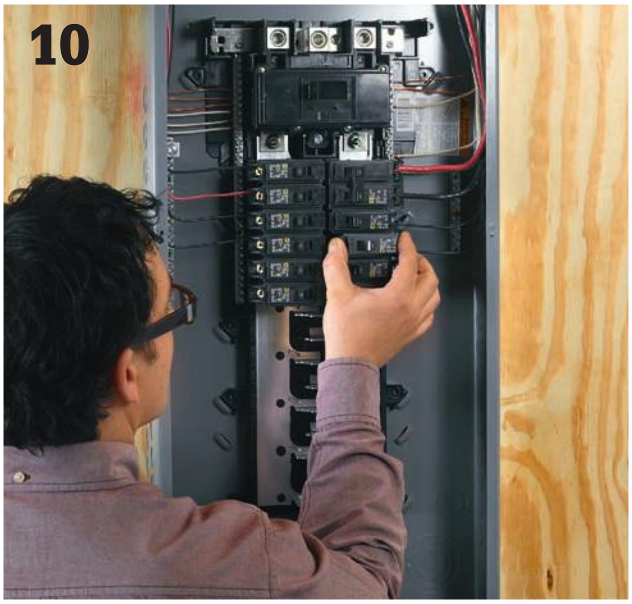

10

Attach the hot lead wire to the terminal on the circuit breaker, and then snap the breaker into an empty slot. When loading slots, start at the top of the panel and work your way downward. It is important that you balance the circuits as you go to equalize the amperage. For example, do not install all the 15-amp circuits on one side and all the 20-amp circuits on the other.

11

Create an accurate circuit index and affix it to the inside of the service panel door. List all loads that are on the circuit as well as the amperage. Once you have restored power to the new service panel (see step 18), test out each circuit to make sure you don't have any surprises. With the main breakers on, shut off all individual circuit breakers, and then flip each one on by itself. Walk through your house and test every switch and receptacle to confirm the loads on that circuit.

(continued)

1st proofs for reference only

Edition
age: 54

41940 - Advanced Home Wiring_001-063.indd 55
Pro-Vision C M Y K Text

Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
Page: 55
10/10/14 9:06 AM

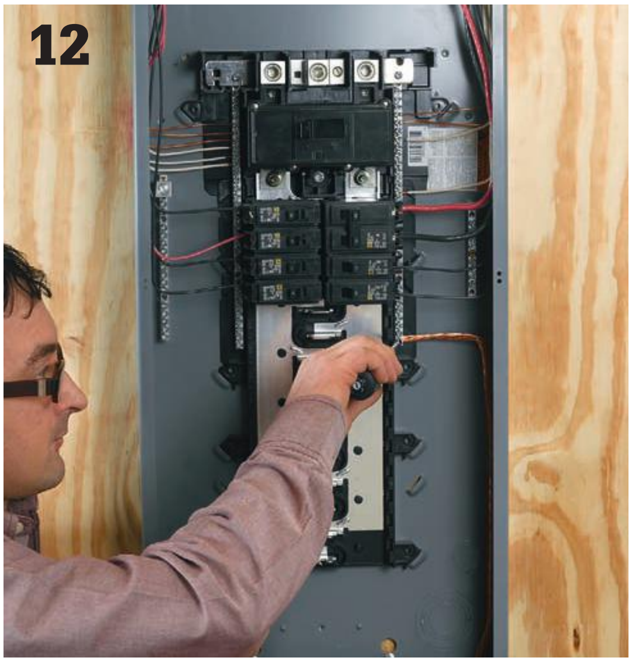

12

Install grounding conductors (see pages 58 to 59). Local codes are very specific about how the grounding and bonding needs to be accomplished. For example, some require multiple rods driven at least 6 ft. apart. Discuss your grounding requirements thoroughly with your inspector or an electrician before making your plan.

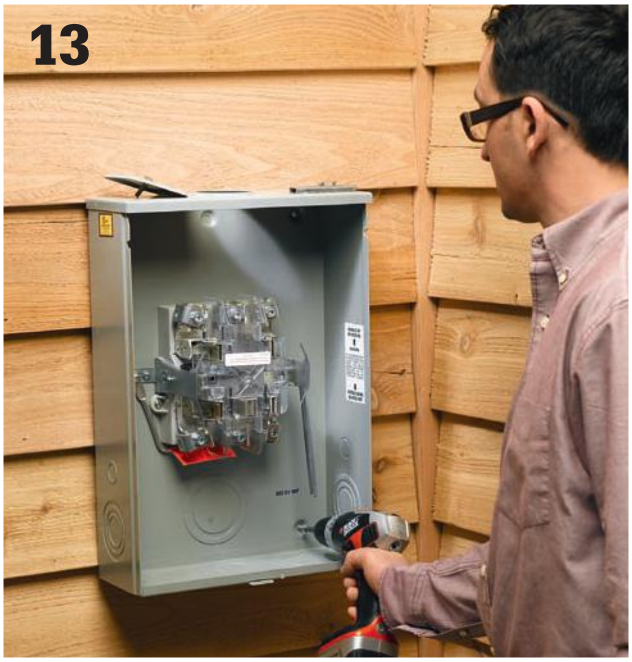

13

Replace the old meter base (have the utility company remove the meter when they shut off power to the house, step 1). Remove the old meter base, also called a socket, and install a new base that's rated for the amperage of your new power service. Here, a 200-amp bypass meter base is being installed.

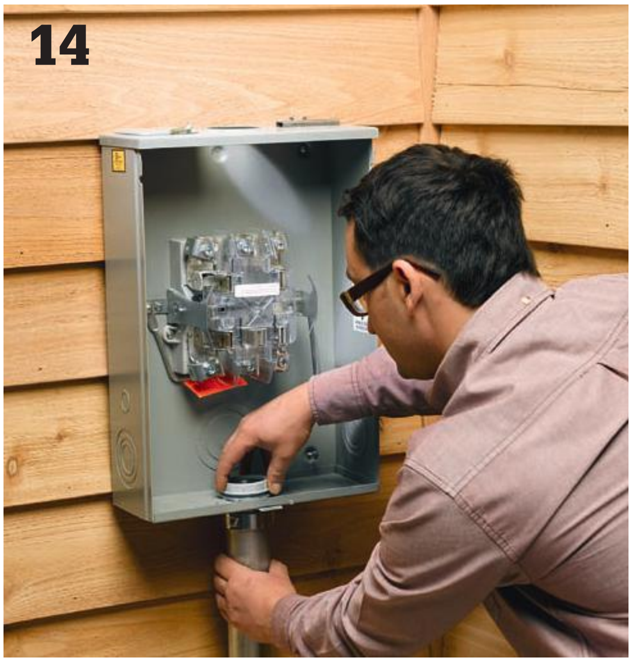

14

Update the conduit that runs from your house to the bottom of the meter base. This should be 2" rigid conduit in good repair. Attach the conduit to the base and wall with the correct fittings. Rigid metal conduit is a good option, but Schedule 80 PVC is probably the best choice for housing the service entrance wires.

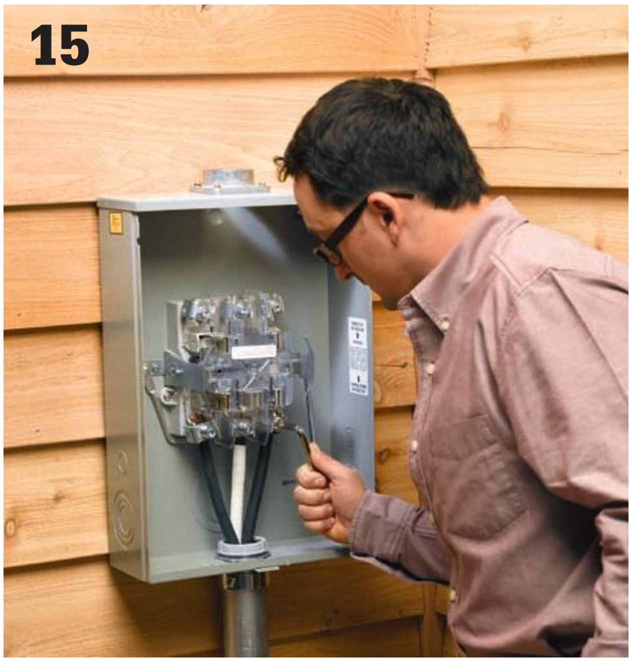

15

Install new service entrance wires. Each wire carries 120 volts from the meter to the service wire lugs at the top of your service panel. Code is very specific about how these connections are made. In most cases, you'll need to tighten the terminal nuts with a specific amount of torque that requires a torque wrench to measure. Also attach the sheathed neutral wire to the neutral/grounding lug.

1st proofs for reference only

41940 - Advanced Home Wiring_001-063.indd 56

Pro-Vision C M Y K Text

Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 9:06 AM

Page: 56

Pro-Vis

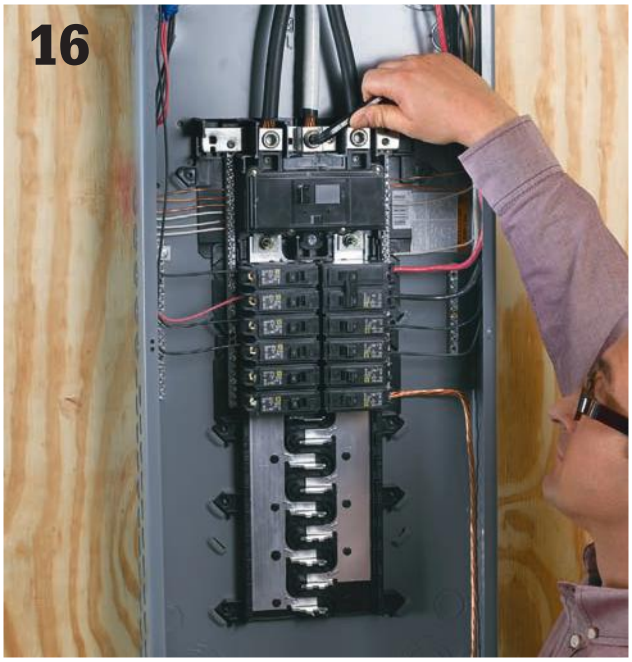

16

Attach the SE wires to the lugs connected to the main breakers at the top of your service entry panel. Do not remove too much insulation on the wires—leaving the wires exposed is a safety hazard. The neutral service entry wire is attached either directly to the neutral bus bar or to a metal bridge that is connected to the neutral bonding bus bar. Install the green grounding screw provided with the panel.

17

Install service entrance wires from the meter to the weatherhead, where the connections to the service drop wires are made. Only an agent for your public utility company may make the hookup at the weatherhead.

Tall Mast, Short Roof ▸

The service drop must occur at least 10 ft. above ground level, and as much as 14 ft. in some cases. Occasionally, this means that you must run the conduit for the service mast up through the eave of your roof and seal the roof penetration with a boot.

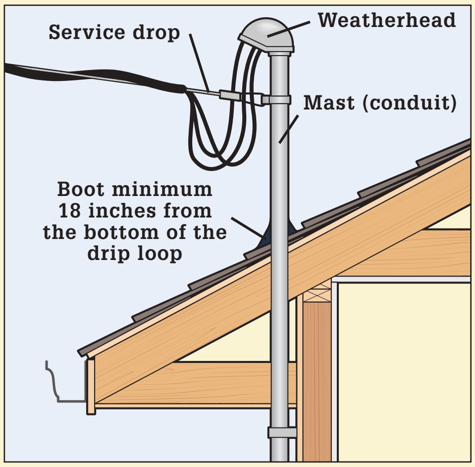

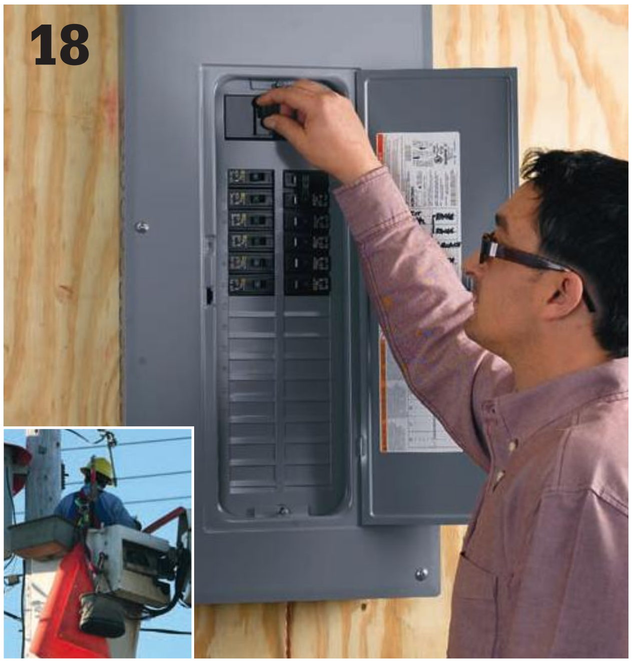

18

Have the panel and all connections inspected and approved by your local building department, and then contact the public utility company to make the connections at the power drop. Once the connections are made, turn the main breakers on and test all circuits.

1st proofs for reference only

Grounding & Bonding a Wiring System

All home electrical systems must be bonded and grounded according to code standards. This entails two tasks: the metal water and gas pipes must be connected electrically to create a continuous low resistance path back to the main electrical panel; and the main electrical panel must be grounded to a grounding electrode such as a ground rod or rods driven into the earth near the foundation of your house. Although the piping system is bonded to the ground through your main electrical service panel, the panel grounding and the piping bonding are unrelated when it comes to function. The grounding wire that runs from your electrical panel to grounding electrode helps even out voltage increases that often occur because of lightning and other causes. The wires that bond your metal piping are preventative, and they only become important in the unlikely event that an electrical conductor energizes the pipe. In that case, correct bonding of the piping system will ensure that the current does not remain in the system, where it could electrocute anyone who touches a part of the system, such as a faucet handle. Bonding is done relatively efficiently at the water heater, as the gas piping and water piping generally there.

Gas pipe in older homes is usually steel or copper. The bonding connection point for these pipes can be at any accessible location, such as at the water heater or at the gas meter. Gas pipe in some new homes is a flexible material called corrugated stainless steel tubing (CSST). The bonding point for CSST must be at the first piece of steel or copper pipe where the gas service enters the home. This is because lightning can blow holes in CSST, causing a gas leak.

Tools & Materials ▸

- Hammer
- Straight edge screwdriver
- Drill
- ½" drill bit
- A length of ground wire
- Some wire staples
- 3 pipe ground clamps
- Eye and ear protection
- Work gloves
- Grounding rods
- 5-lb. maul
- Caulk

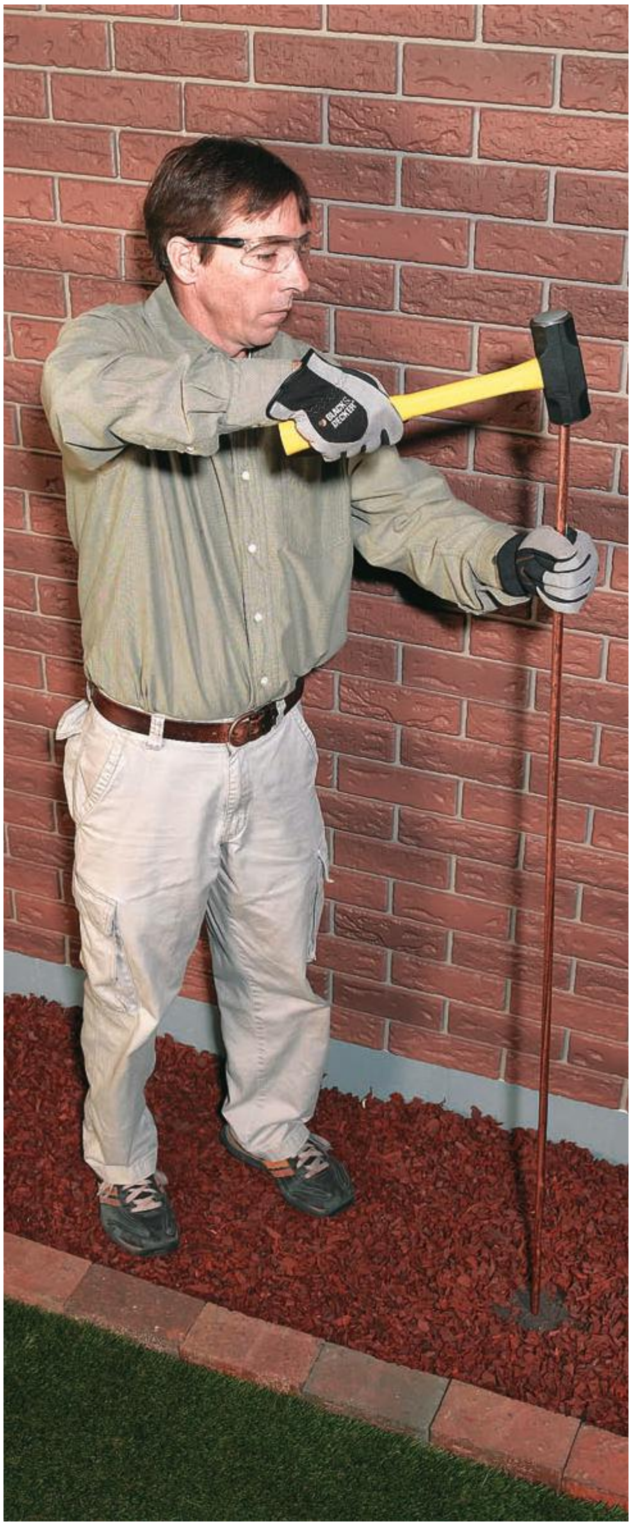

A pair of 8-ft.-long metal ground rods are driven into the earth next to your house to provide a path to ground for your home wiring system.

1st proofs for reference only

41940 - Advanced Home Wiring_001-063.indd 58
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/30/14 10:06 AM
Page: 58

How to Bond Metallic Piping

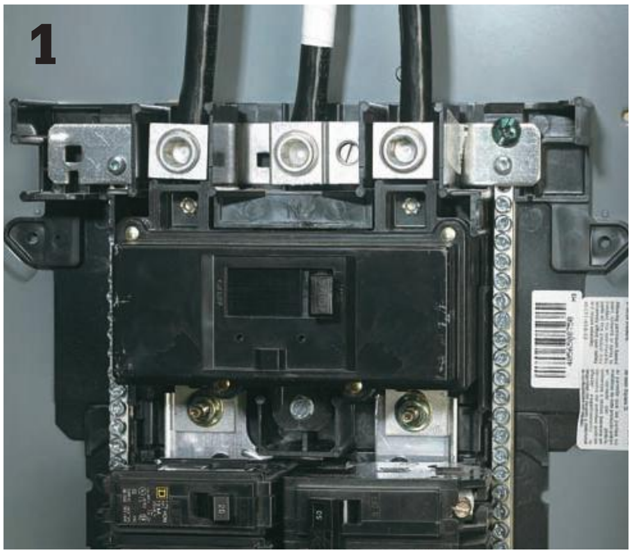

Determine the amperage rating of your electrical service by looking at your main breakers. The system amperage (usually 100 or 200 amps) determines the required gauge of the bonding wire you need. #4 copper wire is sufficient for service not exceeding 200 amps. Smaller, less expensive copper wire is allowed for services between 100 and 175 amps. Check with your electrical inspector if you want to use wire smaller than #4.

Run the bonding wire from a point near your water heater (a convenient spot if you have a gas-fueled water heater) to an exit point where the wire can be bonded to the grounding wire that leads to the exterior grounding electrodes. This is frequently done at the service panel. Run this wire as you would any other cable, leaving approximately 6 to 8 ft. of wire at the water heater. If you are running this wire through the ceiling joists, drill a ½" hole as close to the center as possible to not weaken the joist. Staple the wire every 2 ft. if running it parallel to the joists.

Install pipe ground clamps on each pipe (hot water supply, cold water supply, gas), roughly a foot above the water heater. Do not install clamps near a union or elbow because the tightening of the clamps could break or weaken soldered joints. Also make sure the pipes are free and clear of any paint, rust, or any other contaminant that may inhibit a good clean connection. Do not overtighten the clamps. Use clamps that are compatible with the pipe so that corrosion will not occur. Use copper or brass clamps on copper pipe. Use brass or steel clamps on steel pipe.

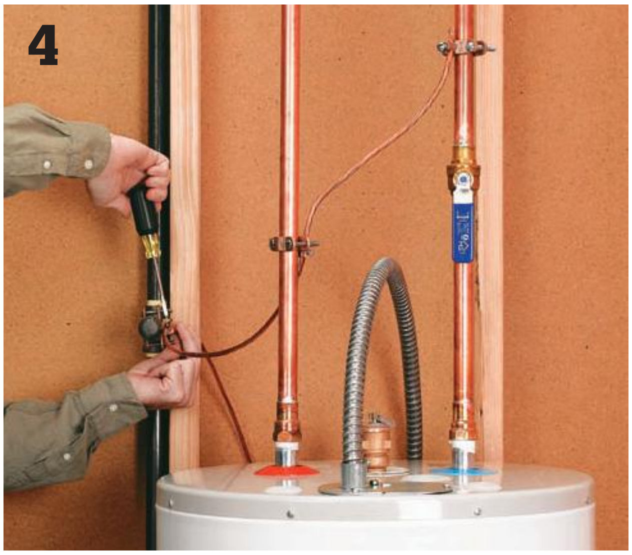

Route the ground wire through each clamp wire hole and then tighten the clamps onto the wire. Do not cut or splice the wire: The same wire should run through all clamps.

(continued)

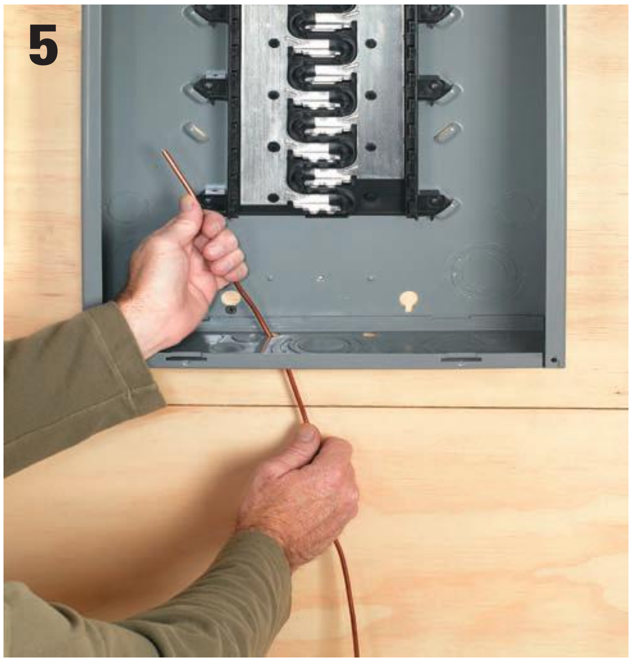

At the panel, turn off the main breaker. Open the cover by removing the screws, and set the cover aside. Route the ground wire through a small ⅜" hole provided towards the rear of the panel on the top or bottom. You will usually have to knock the plug out of this hole by placing a screwdriver on it from the outside and tapping with a hammer. Make sure the ground wire will not come into contact with the bus bars in the middle of the panel or any of the load terminals on the breakers.

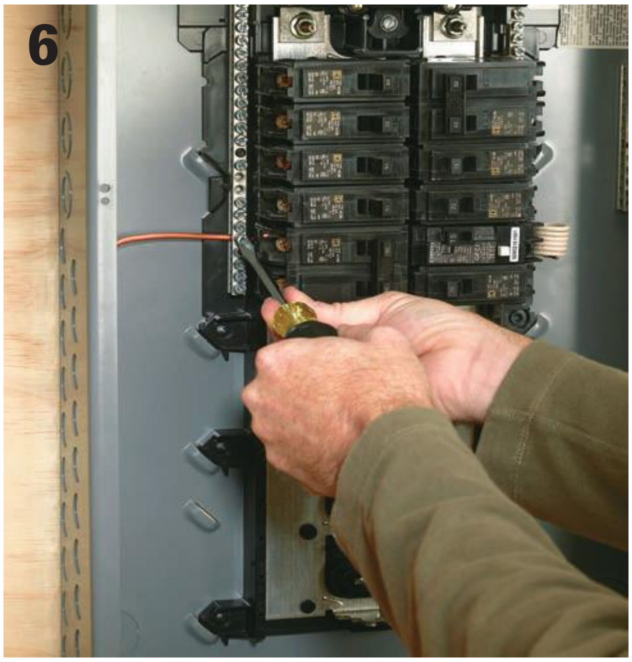

Locate an open hole on your ground and neutral bus and insert the ground wire. These holes are large enough to accommodate up to a #4 awg wire, but it may be difficult at times. If you're having trouble pushing the wire in, trim a little wire off the end and try with a clean cut piece. Secure the set screw at the lug. Replace the panel cover and turn the main breaker back on.

Tips for Grounding the Main Service Panel

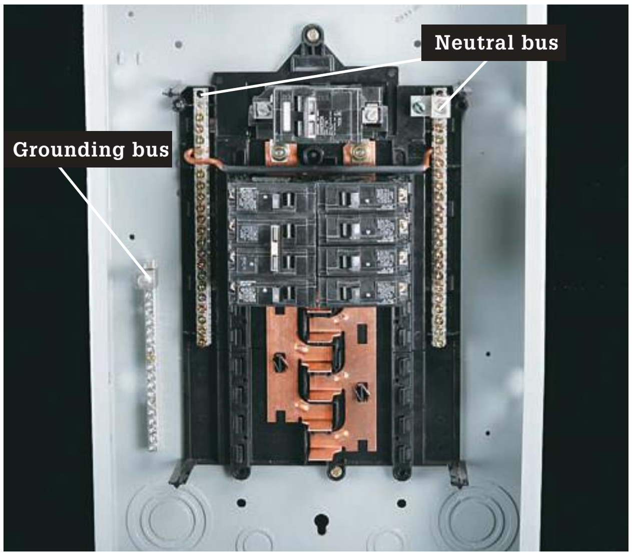

The neutral and grounding wires should not be connected to the same bus in most subpanels. The grounding bus should be bonded to the subpanel cabinet. The neutral bus should not be bonded to the subpanel cabinet.

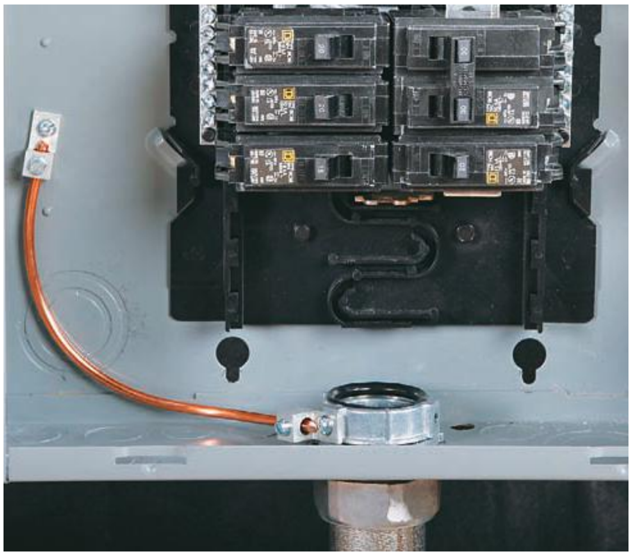

Metallic conduit must be physically and electrically connected to panel cabinets. A bonding bushing may be required in some cases, where not all of a knockout is removed.

Ground Rod Installation

The ground rod is an essential part of the grounding system. Its primary function is to create a path to ground for electrical current, such as lightning, line surges, and unintentional contact with high voltage lines. If you upgrade your electrical service you likely will need to upgrade your grounding wire and rods to meet code.

Note: Different municipalities have different requirements for grounding, so be sure to check with the AHJ (Authority Having Jurisdiction) first before attempting to do this yourself.

Call before you dig! Make sure the area where you will be installing the ground rods is free and clear from any underground utilities.

Exercise Your Breakers ▸

Your breakers (including the main) should be "exercised" once a year to ensure proper mechanical function. Simply turn them off and then back on. A convenient time to perform the exercise is at daylight savings time, when you'll need to reset all of your clocks anyway.

How to Install a Grounding Electrode System

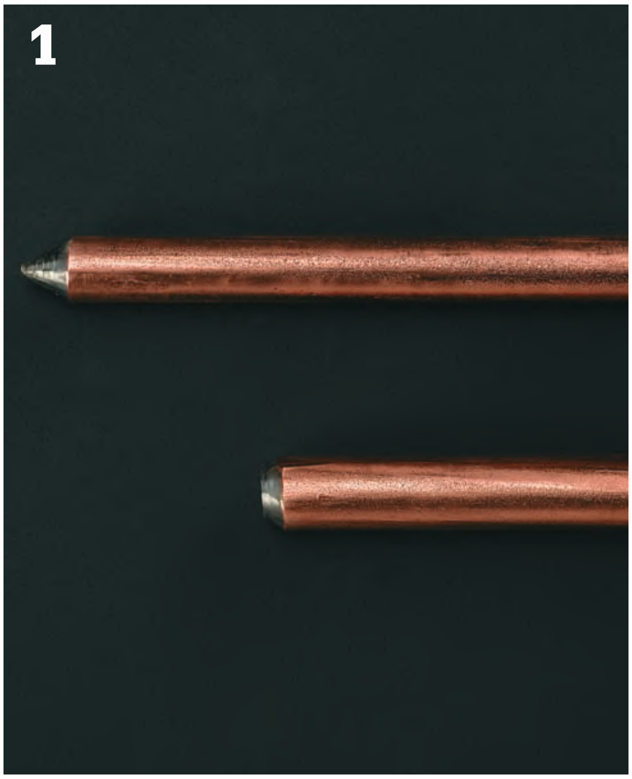

Begin by purchasing two copper-coated steel ground rods 5/8" diameter by 8' long. Grounding rods have a driving point on one end and a striking face on the other end.

2
BLACK&DECKER

Drill a 5/16" hole in the rim joist of your house, as close as practical to the main service panel to the outside of the house above the ground level at least 6".

(continued)

1st
proofs for reference only
Edition
age: 60
41940 - Advanced Home Wiring_001-063.indd 61
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 9:06 AM
Page: 61

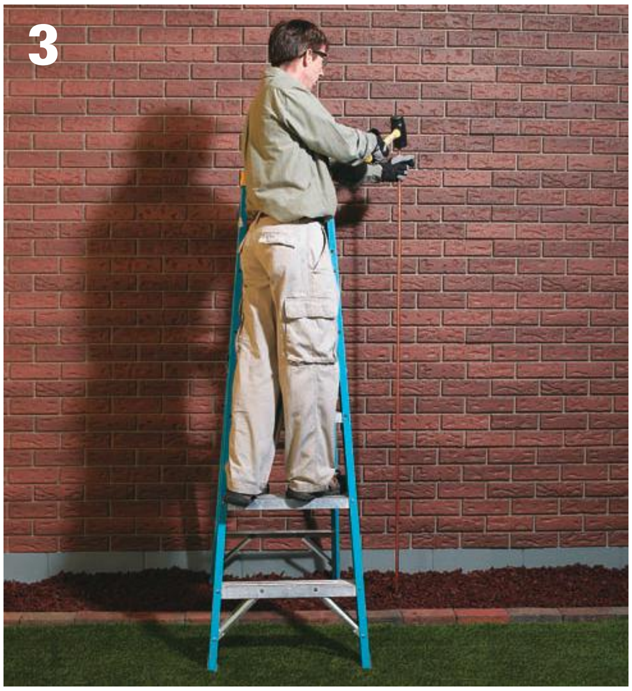

About a foot from the foundation of the house, pound one ground rod into the earth with a five-pound maul. If you encounter a rock or other obstruction, you can pound the ground rod at an angle as long as it does not exceed 45°. Drive until only 3" or 4" of the rod is above ground. Measure at least 6 ft. from the first ground rod and pound in another one.

Run uninsulated #4 copper wire from the ground bus in your main service panel through the hole in the rim joist and to the exterior of the house, leaving enough wire to connect the two ground rods together.

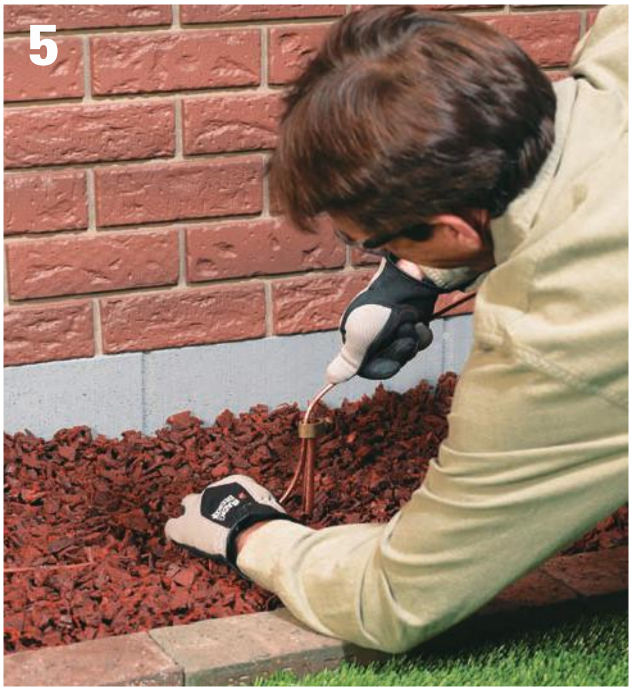

Using a brass clamp commonly referred to as an acorn, connect the wire to the first ground rod, pulling the wire taut so no slack exists. Continue pulling the wire to reach the second grounding rod, creating a continuous connection.

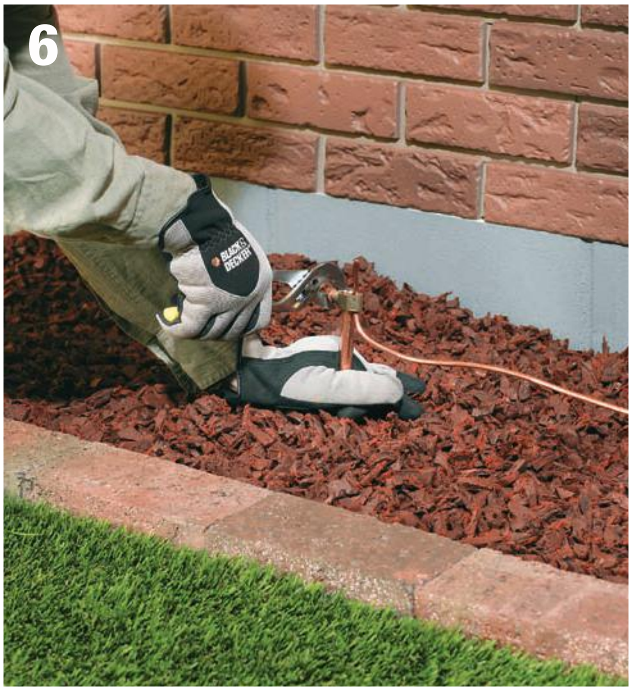

Connect the second ground rod with another acorn to the uncut grounding wire previously pulled through the first acorn. Trim the excess wire.

1st
proofs for reference only
41940 - Advanced Home Wiring_001-063.indd 62
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 9:06 AM
Page: 62
Pro-Visi

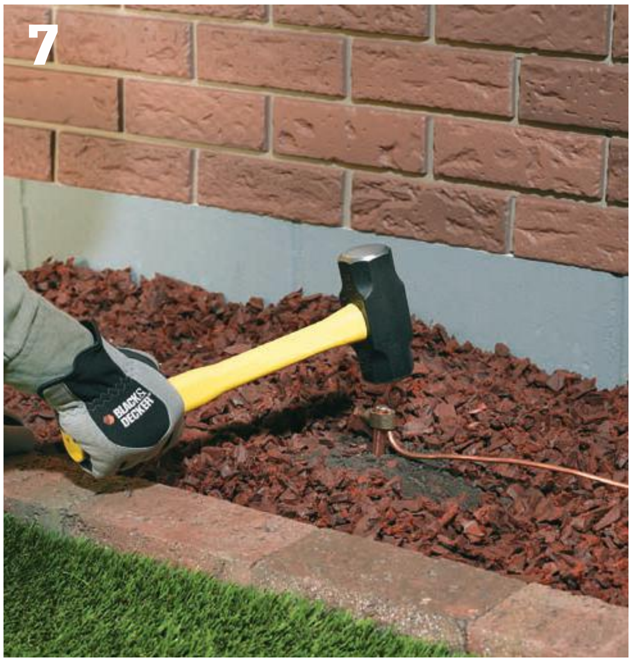

Dig out a few inches around each rod to create clearance for the five-pound maul. Creating a shallow trench beneath the grounding wire between the rods is also a good idea. Drive each rod with the maul until the top of the rod is a few inches below grade.

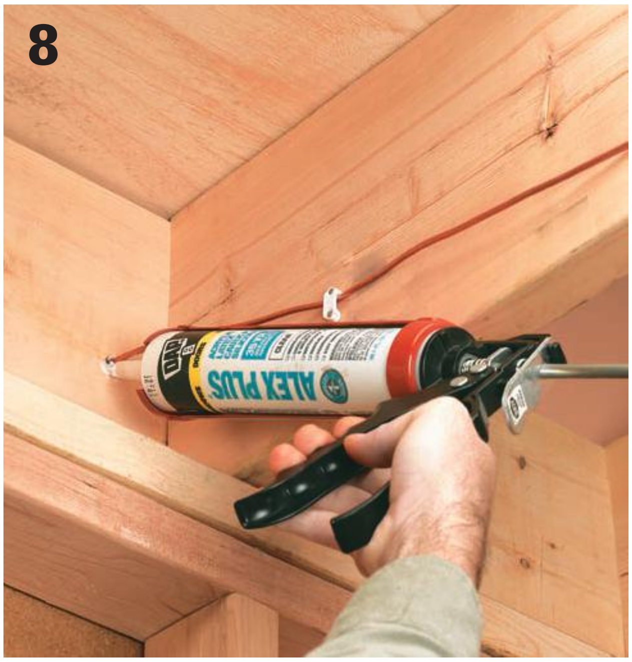

Inject caulk into the hole in the rim joist on both the interior and exterior side.

Tips for Grounding

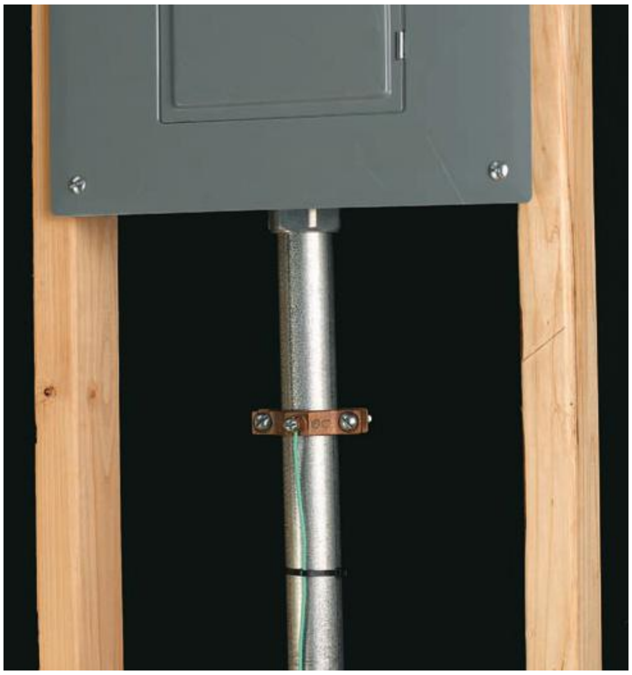

A listed metal strap may be used to ground indoor communication wires such as telephone and cable TV if an intersystem bonding terminal is not available.

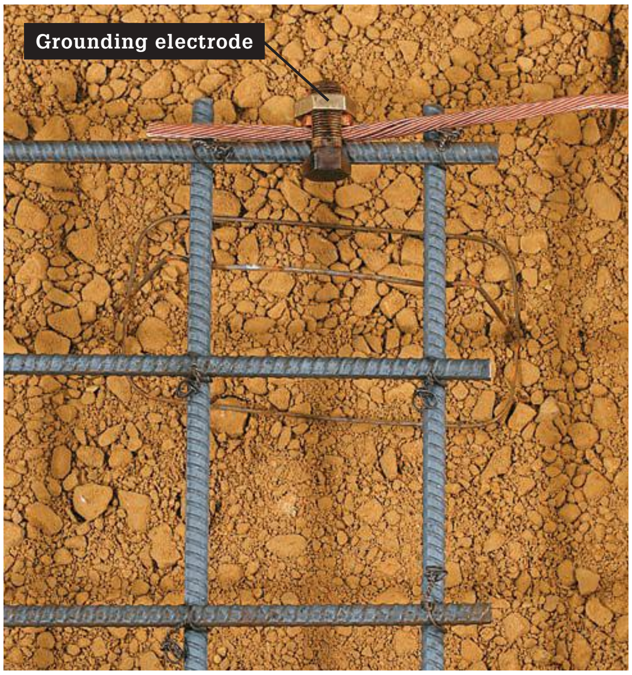

A piece of reinforcing bar encased in a concrete footing is a common grounding electrode in new construction. Called an ufer, the electrode must be No. 4 or larger rebar and at least 20 ft. long. (Shown prior to pouring concrete.)

1st proofs for reference only

Edition
age: 62

41940 - Advanced Home Wiring_001-063.indd 63
Pro-Vision C M Y K Text

Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 9:06 AM
Page: 63

Installing a Subpanel

Install circuit breaker subpanels if the main circuit breaker panel does not have enough open breaker slots for the new circuits you are planning. Subpanels serve as a second distribution center for connecting circuits. They receive power from a double-pole circuit breaker you install in the main circuit breaker panel.

If the main service panel is so full that there is no room for the double-pole subpanel breaker, you can reconnect some of the existing 120-volt circuits to special slimline breakers (photos below).

Plan your subpanel installation carefully, making sure your electrical service supplies enough power to support the extra load of the new subpanel circuits. Assuming your main service is adequate, consider installing an oversized subpanel breaker in the main panel to provide enough extra amps to meet the needs of future wiring projects.

Also consider the physical size of the subpanel, and choose one that has enough extra slots to hold circuits you may want to install later. The smallest panels have room for up to six single-pole breakers (or three double-pole breakers), while the largest models can hold 20 single-pole breakers or more.

Subpanels often are mounted near the main circuit breaker panel. Or, for convenience, they can be installed close to the areas they serve, such as in a new room addition or a garage. In a finished room, a subpanel can be painted or housed in a decorative cabinet so it is less of a visual distraction. If it is covered, make sure the subpanel is easily accessible and clearly identified.

Tools & Materials ▸

- Hammer
- Screwdriver
- Circuit tester
- Cable ripper
- Combination tool
- Screws
- Cable clamps
- Three-wire NM cable
- Cable staples
- Double-pole circuit breaker
- Circuit breaker subpanel
- Slimline circuit breakers

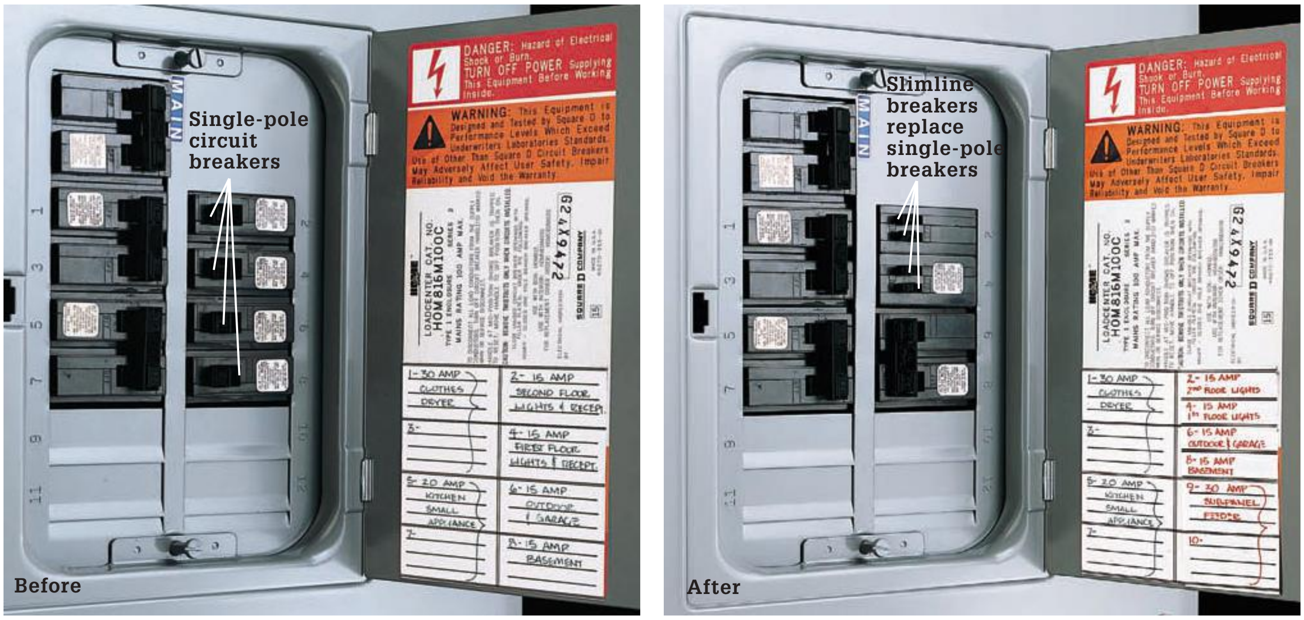
Single-pole circuit breakers
Before
Slimline breakers replace single-pole breakers
After

To conserve space in a service panel, you can replace single-pole breakers with slimline breakers. Slimline breakers take up half the space of standard breakers, allowing you to fit two circuits into one single slot on the service panel. In the service panel shown above, four single-pole 120-volt breakers were replaced with slimline breakers to provide the double opening needed for a 30-amp, 240-volt subpanel feeder breaker. Use slimline breakers (if your municipality allows them) with the same amp rating as the standard single-pole breakers you are removing, and make sure they are approved for use in your panel. If your municipality and panel allow slimline breakers, there may be restrictions on the quantity and location where they may be installed on the panel.

1st proofs for reference only

41940 - Advanced Home Wiring_064-128.indd 64
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 64

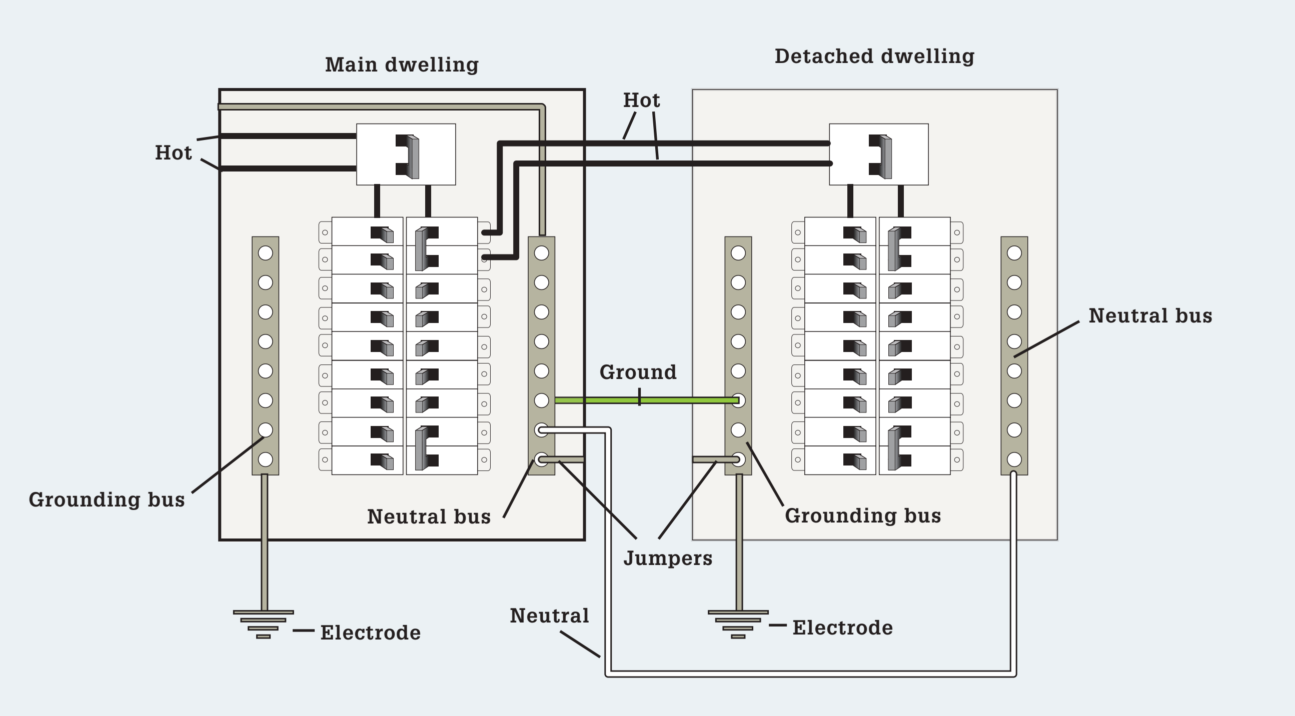
Main dwelling
Detached dwelling
Hot
Hot
Neutral bus
Ground
Grounding bus
Neutral bus
Grounding bus
Jumpers
Neutral
Electrode
Electrode

Wiring diagram for wiring a feeder from the main service panel to a subpanel in a separate building.

How to Install a Subpanel

Subpanels are subject to the same installation and clearance rules as service panels. The subpanel can be mounted to the sides of studs or to plywood attached between two studs. The panel shown here extends ½" past the face of studs so it will be flush with the finished wall surface.

Open a knockout in the subpanel using a screwdriver and hammer. Run the feeder cable from the main circuit breaker panel to the subpanel, leaving about 2 ft. of excess cable at each end.

Attach a cable clamp to the knockout in the subpanel. Insert the cable into the subpanel, and then anchor it to framing members within 8" of each panel and every 54" thereafter.

(continued)

1st
proofs for reference only

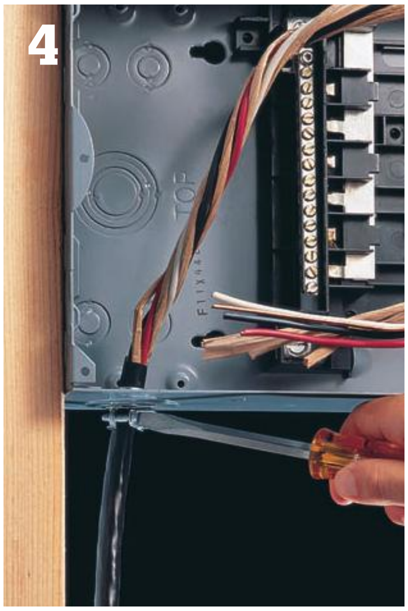

Strip away outer sheathing from the feeder cable using a cable ripper. Leave at least ¼" of sheathing extending into the subpanel. Tighten the cable clamp screws so the cable is held securely, but not so tightly that the wire sheathing is crushed.

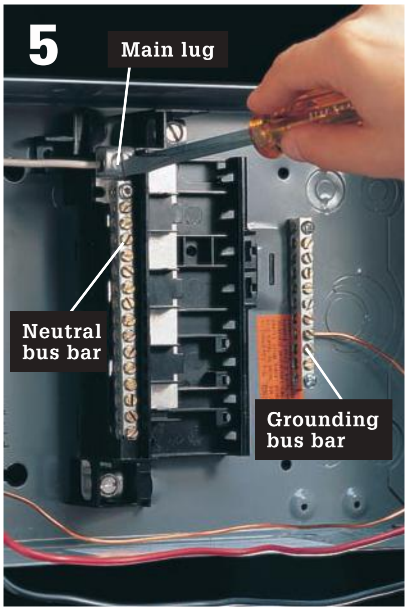

Strip ½" of insulation from the white neutral feeder wire, and attach it to the main lug on the subpanel neutral bus bar. Connect the grounding wire to a setscrew terminal on the grounding bus bar. Fold excess wire around the inside edge of the subpanel.

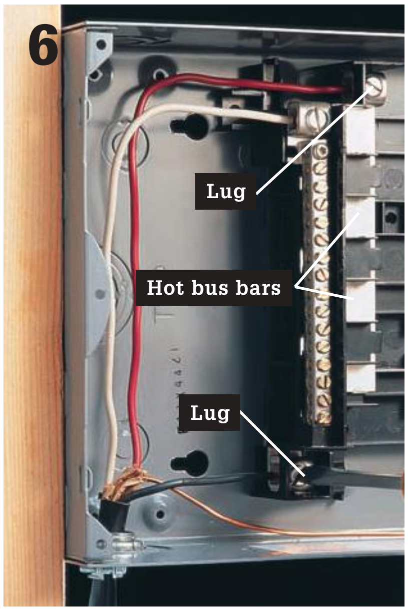

Strip away ½" of insulation from the red and the black feeder wires. Attach one wire to the main lug on each of the hot bus bars. Fold excess wire around the inside edge of the subpanel.

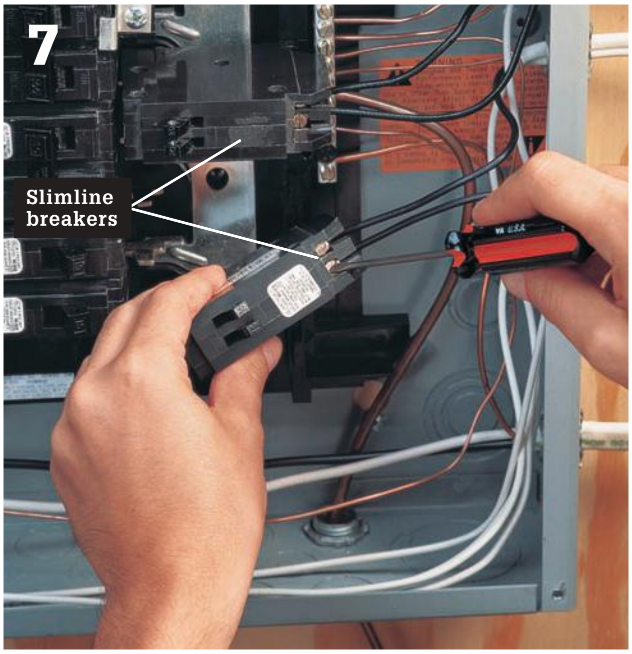

At the main circuit breaker panel, shut off the main circuit breaker, and then remove the coverplate and test for power. If necessary, make room for the double-pole feeder breaker by removing single-pole breakers and reconnecting the wires to slimline circuit breakers. Open a knockout for the feeder cable using a hammer and screwdriver. Note: some panels do not allow slimline breakers and some restrict where slimline breakers can be installed. Read the instructions on the panel cover.

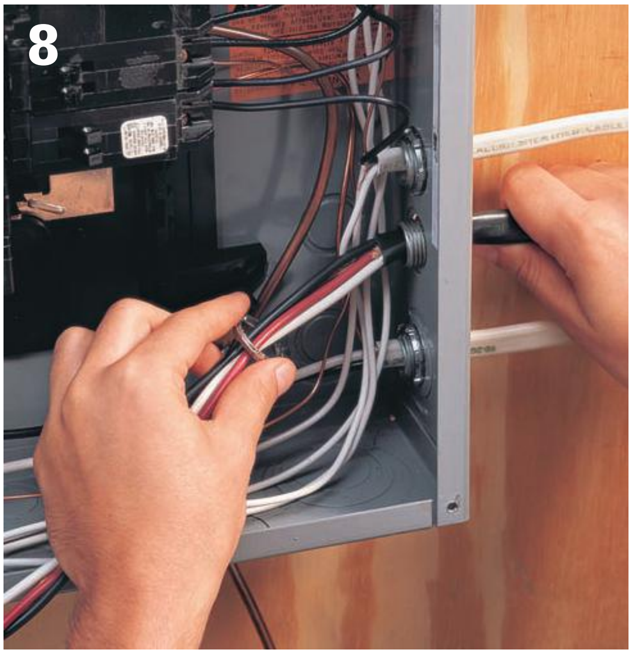

Strip away the outer sheathing from the feeder cable so that at least ¼" of sheathing will reach into the main service panel. Attach a cable clamp to the cable, and then insert the cable into the knockout, and anchor it by threading a locknut onto the clamp. Tighten the locknut by driving a screwdriver against the lugs. Tighten the clamp screws so the cable is held securely, but not so tightly that the cable sheathing is crushed.

41940 - Advanced Home Wiring_064-128.indd 66
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 12:05 AM
Page: 66
Pro-Vis

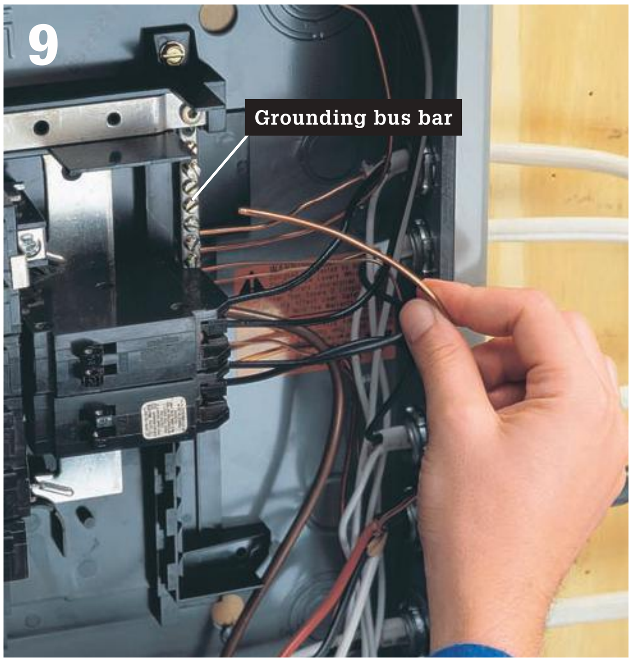

Bend the bare copper wire from the feeder cable around the inside edge of the main circuit breaker panel, and connect it to one of the setscrew terminals on the grounding bus bar.

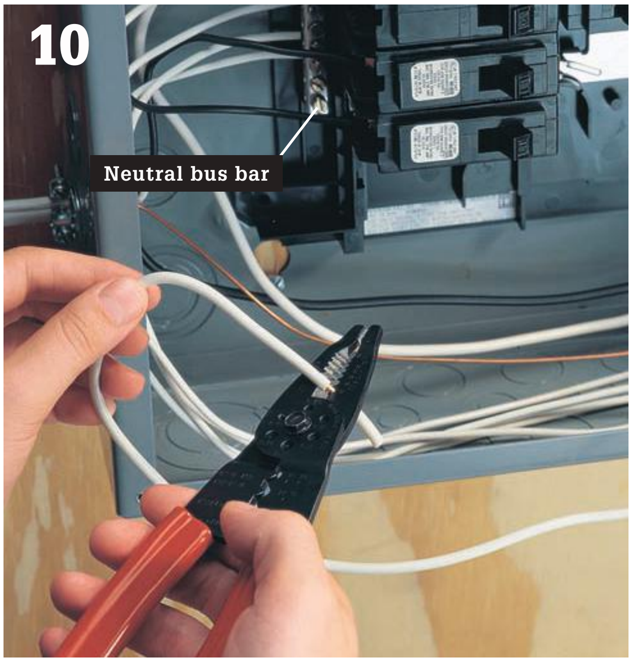

Strip away ½" of insulation from the white feeder wire. Attach the wire to one of the setscrew terminals on the neutral bus bar. Fold excess wire around the inside edge of the service panel.

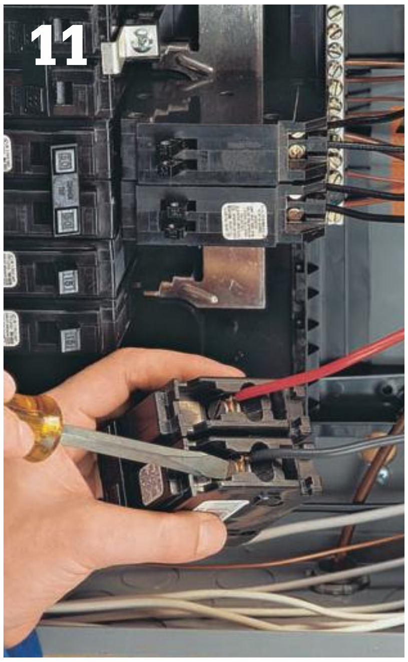

Strip ½" of insulation from the red and the black feeder wires. Attach one wire to each of the setscrew terminals on the double-pole feeder breaker. *Note: If your subpanel arrived with a preinstalled grounding screw in the panel back, remove and discard it.*

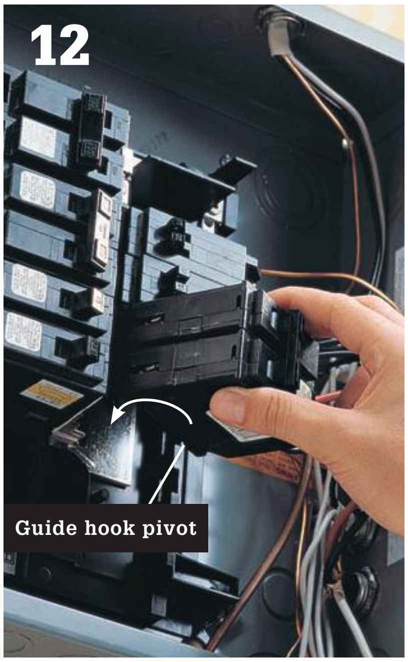

Hook the end of the feeder circuit breaker over the guide hooks on the panel, and then push the other end forward until the breaker snaps onto the hot bus bars (follow manufacturer's directions). Fold excess wire around the inside edge of the circuit breaker panel.

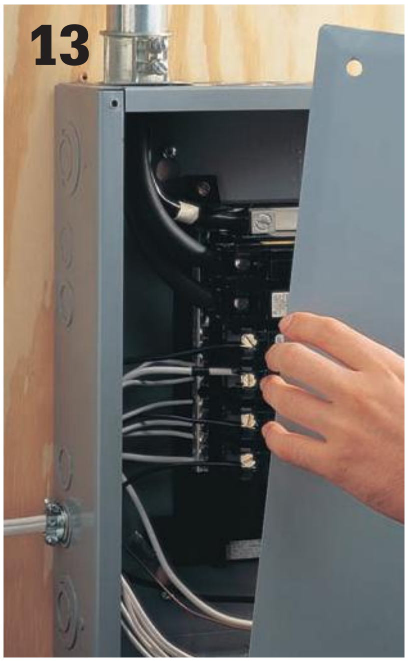

If necessary, open two tabs where the double-pole feeder breaker will fit, and then reattach the cover plate. Label the feeder breaker on the circuit index. Turn the main breaker on, but leave the feeder breaker off until all subpanel circuits have been connected and inspected.

1st
proofs for reference only
Edition
age: 66
41940 - Advanced Home Wiring_064-128.indd 67
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 12:05 AM
Page: 67

Baseboard Heaters

Baseboard heaters are a popular way to provide additional heating for an existing room or primary heat to a converted attic or basement.

Heaters are generally wired on a dedicated 240-volt circuit controlled by a thermostat. Several heaters can be wired in parallel and controlled by a single thermostat (see circuit map 15, page 39).

Baseboard heaters are generally surface-mounted without boxes, so in a remodeling situation, you only need to run cables before installing wallboard. Be sure to mark cable locations on the floor before installing drywall. Retrofit installations are also not difficult. You can remove existing baseboard and run new cable in the space behind. Baseboard heaters (and other heating equipment) get very hot and can ignite nearby combustible materials. Maintain the manufacturers recommended distance between the heater and materials such as curtains, blinds, and wood.

Tools & Materials ▸

Drill/driver
Wire stripper
Cable ripper
Wallboard saw
Baseboard heater or heaters
240-thermostat (in-heater or in-wall)
12/2 NM cable
Electrical tape
Basic wiring supplies

Baseboard heaters can provide primary or supplemental heat for existing rooms or additions. Install heaters with clear space between the heater and the floor.

1st proofs for reference only

41940 - Advanced Home Wiring_064-128.indd 68
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 68

Baseboard Thermostats

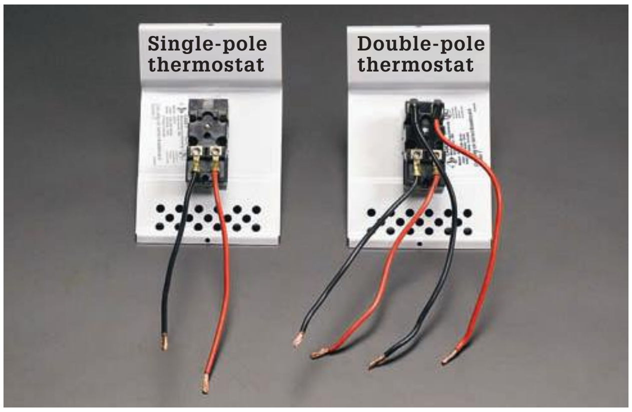

Single-pole and double-pole thermostats work in a similar manner, but double-pole models are safer. The single-pole model will open the circuit (causing shutoff) in only one leg of the power service. Double-pole models have two sets of wires to open both legs, lessening the chance that a person servicing the heater will contact a live wire.

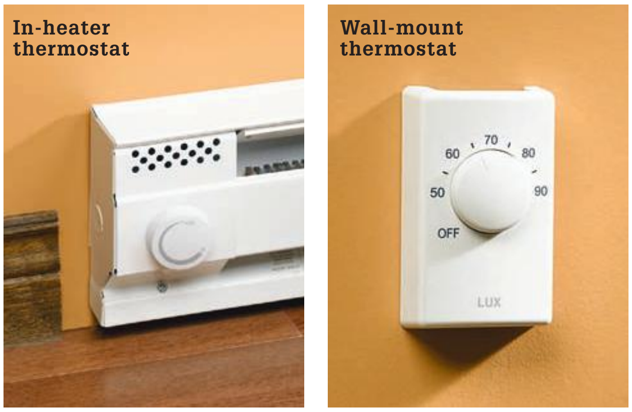

In-heater and wall-mount are the two types of baseboard thermostats you can choose from. If you are installing multiple heaters, a single wall-mount thermostat is more convenient. Individual in-heater thermostats give you more zone control, which can result in energy savings.

How Much Heater Do You Need?

If you don't mind doing a little math, determining how many lineal feet of baseboard heater a room requires is not hard.

1. Measure the area of the room in square feet (length × width): ________
2. Multiply the area by 10 to get the baseline minimum wattage: ________
3. Add 5% for each newer window or 10% for each older window: ________
4. Add 10% for each exterior wall in the room: ________
5. Add 10% for each exterior door: ________
6. Add 10% if the space below is not insulated: ________
7. Add 20% if the space above is not well insulated: ________
8. Add 10% if ceiling is more than 8 ft. high: ________
9. Total of the baseline wattage plus all additions: ________
10. Divide this number by 250 (the wattage produced per foot of standard baseboard heater): ________
11. Round up to a whole number. This is the minimum number of feet of heater you need. ________

Note: It is much better to have more feet of heater than is required than fewer. Having more footage of heater does not consume more energy; it does allow the heaters to work more efficiently.

Planning Tips for Baseboard Heaters

- Baseboard heaters require a dedicated circuit. A 20-amp, 240-volt circuit of 12-gauge copper wire will power up to 16 ft. of heater.
- Do not install a heater beneath a wall receptacle. Cords hanging down from the receptacle are a fire hazard.
- Do not mount heaters directly on the floor. You should maintain at least 1" of clear space between the baseboard heater and the floor covering.
- Installing heaters directly beneath windows is a good practice.
- Locate wall thermostats on interior walls only, and do not install directly above a heat source.

1st
proofs for reference only

Edition
age: 68
41940 - Advanced Home Wiring_064-128.indd 69
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 69

How to Install a 240-Volt Baseboard Heater

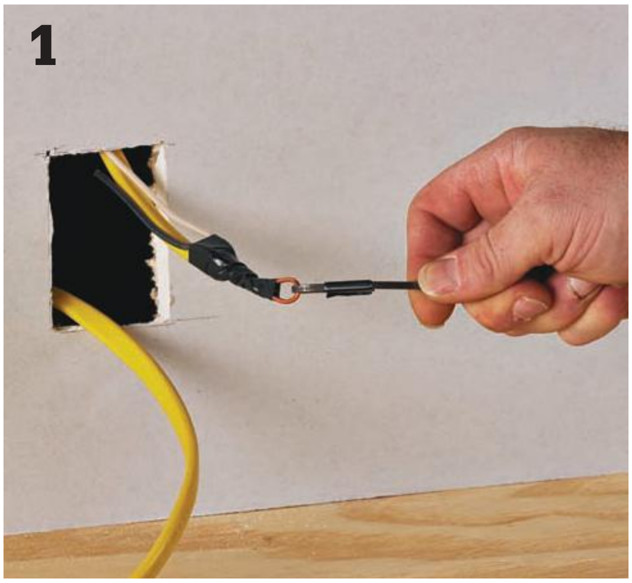

At the heater locations, cut a small hole in the drywall 3" to 4" above the floor. Pull 12/2 NM cables through the first hole: one from the thermostat, the other to the next heater. Pull all the cables for subsequent heaters. Middle-of-run heaters will have two cables, while end-of-run heaters have only one cable. (See also circuit map 15, page 39.)

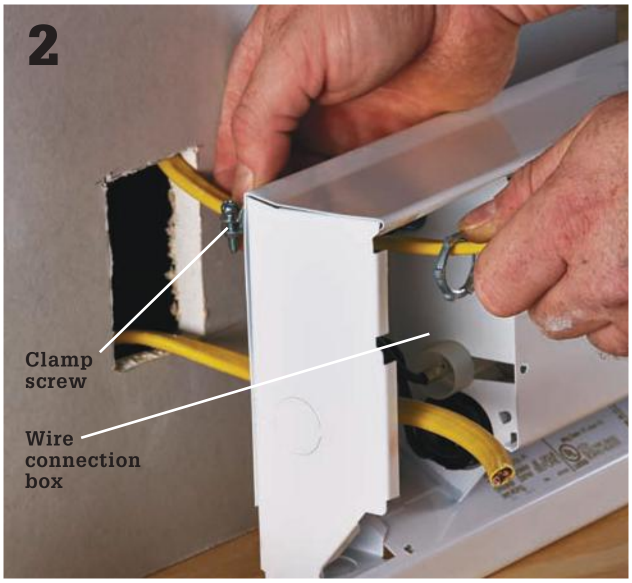

Remove the cover on the wire connection box. Open a knockout for each cable that will enter the box, and then feed the cables through the cable clamps and into the wire connection box. Attach the clamps to the wire connection box, and tighten the clamp screws until the cables are gripped firmly.

Anchor the heater against wall about 1" off floor by driving flathead screws through the back of the housing and into studs. Strip away cable sheathing so at least ½" of sheathing extends into the heater. Strip ¾" of insulation from each wire using a combination tool.

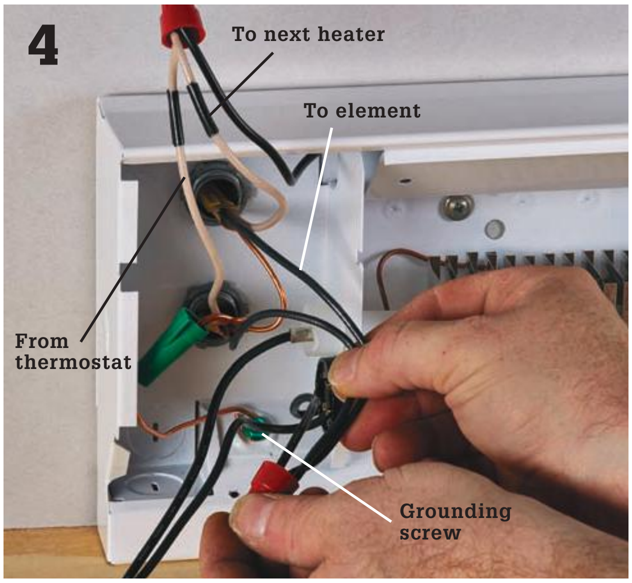

Make connections to the heating element if the power wires are coming from a thermostat or another heater controlled by a thermostat. See the next page for other wiring schemes. Connect the white circuit wires to one of the wire leads on the heater. Tag white wires with black tape to indicate they are hot. Connect the black circuit wires to the other wire lead. Connect a grounding pigtail to the green grounding screw in the box, and then join all grounding wires with a wire connector. Reattach the cover.

1st
proofs for reference only

41940 - Advanced Home Wiring_064-128.indd 70
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 12:05 AM
Page: 70
Pro-Vis

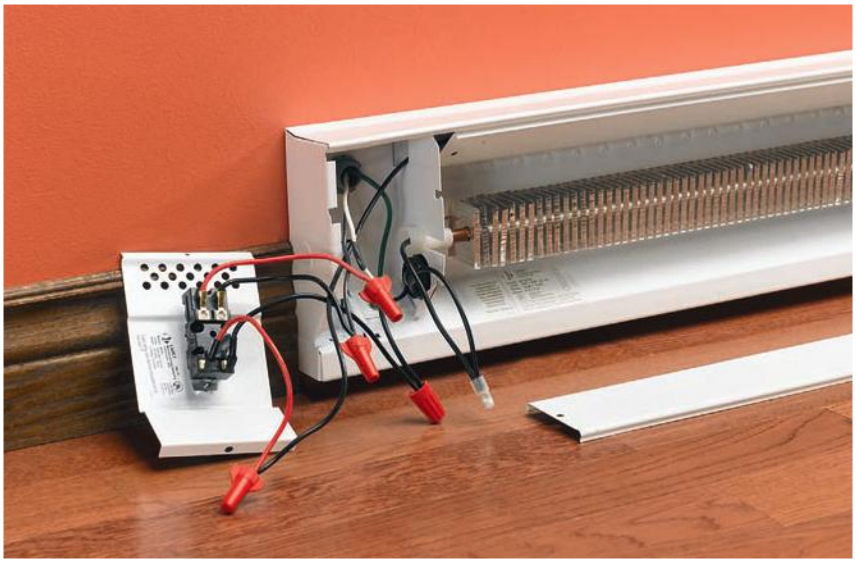

One heater with end-cap thermostat. Run both power leads (black plus tagged neutral) into the connection box at either end of the heater. If installing a single-pole thermostat, connect one power lead to one thermostat wire and connect the other thermostat wire, to one of the heater leads. Connect the other hot LINE wire to the other heater lead. If you are installing a double-pole thermostat, make connections with both legs of the power supply.

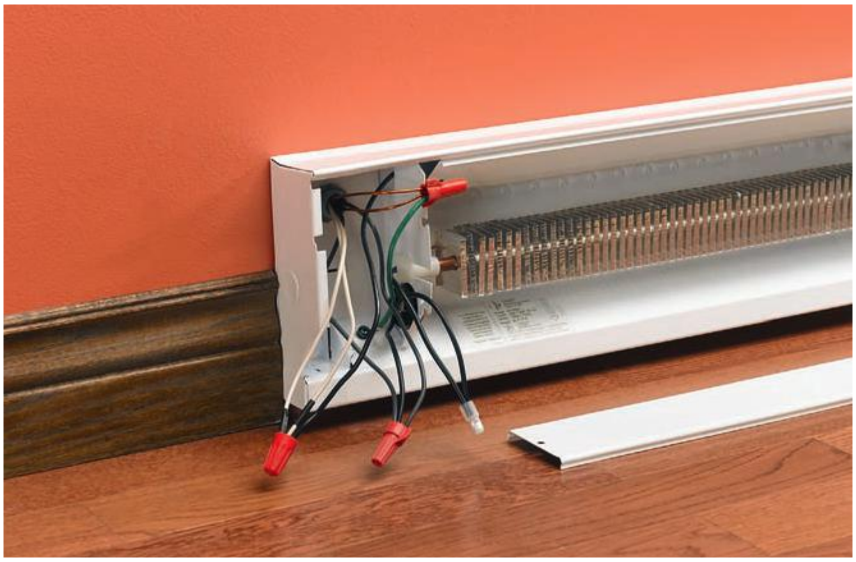

Multiple heaters. At the first heater, join both hot wires from the thermostat to the wires leading to the second heater in line. Be sure to tag all white neutrals hot. Twist copper ground wires together and pigtail them to the grounding screw in the baseboard heater junction box. This parallel wiring configuration ensures that power flow will not be interrupted to the downstream heaters if an upstream heater fails.

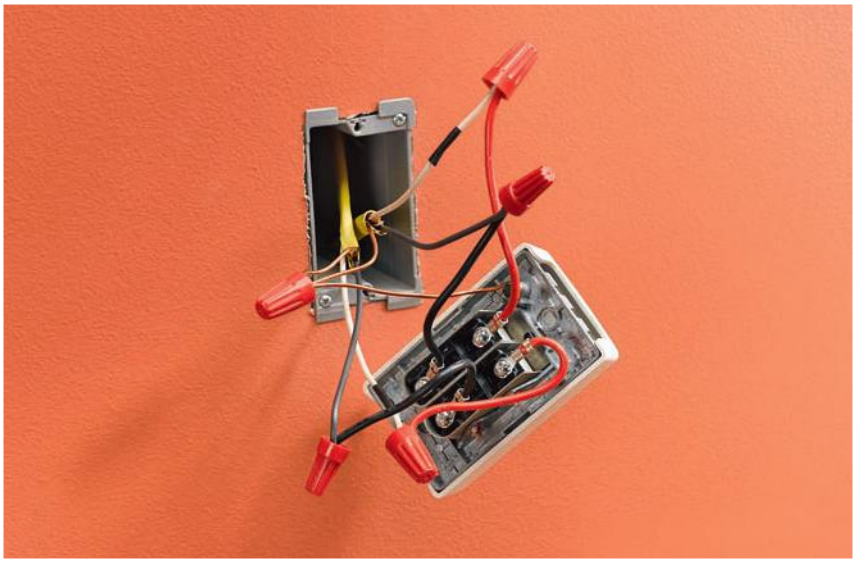

Wall-mounted thermostat. If installing a wall-mounted thermostat, the power leads should enter the thermostat first and then be wired to the individual heaters singly or in series. Hookups at the heater are made as shown in step 4. Be sure to tag the white neutral as hot in the thermostat box as well as in the heater box.

Wall Heaters

Installing a wall heater is an easy way to provide supplemental heat to a converted attic or basement without expanding an existing HVAC system.

Wall heaters are easy to install during a remodel (most have a separate can assembly that you attach to the framing before the drywall is installed). They can also be retrofitted.

Most models available at home centers use 120-volt current (shown below), but 240-volt models are also available.

Tools & Materials ▸

Drywall saw
Drill
Fish tape
Combination tool
Screwdrivers
12/2 NM cable
Wire connectors
Wall heater
Thermostat
(optional)
Wallboard saw

Wall heaters are an easy-to-install way to provide supplemental heat. Some models have built-in thermostats, while others can be controlled by a remote thermostat.

How to Install a Wall Heater in a Finished Wall

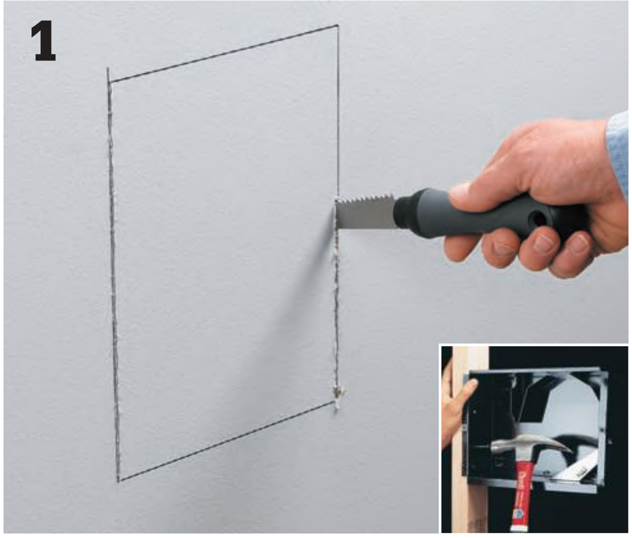

Make an opening in the wall for the heater. Use a stud finder to locate a stud in the area where you want to install the heater. Mark the opening for the heater according to the manufacturer's guidelines so that one side of the heater sits flush with a stud. Pay attention to clearance requirements. Cut the opening with a wallboard saw. If the wall is open, install the heater can before hanging drywall (inset).

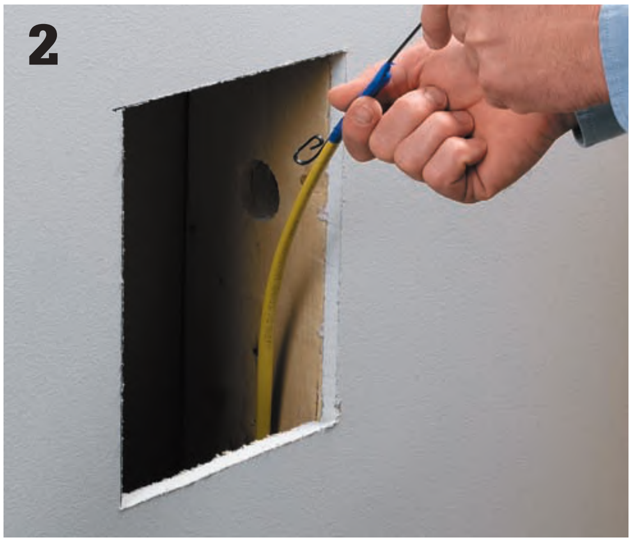

Turn power off, and test for power. Pull 12/2 NM cable from the main panel to the wall opening. If the heater is controlled by a separate thermostat, pull cable to the thermostat, and then run another cable from the thermostat to the heater location.

1st
proofs for reference only
41940 - Advanced Home Wiring_064-128.indd 72
Pro-Vision C M Y K Text
10/10/14 12:05 AM
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
Page: 72

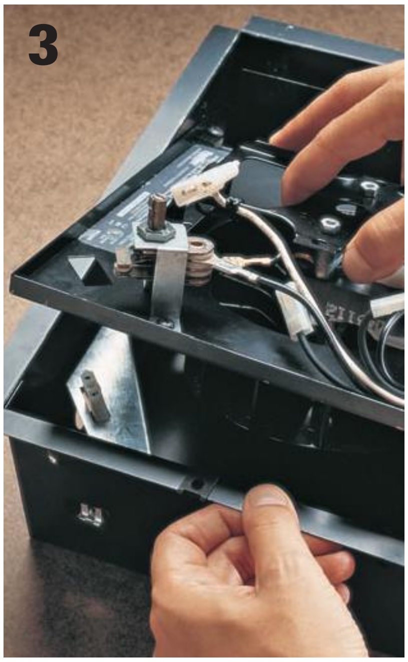

Disconnect and remove the motor unit from the heater can. Remove a knockout from the can, and route the cable into the can.

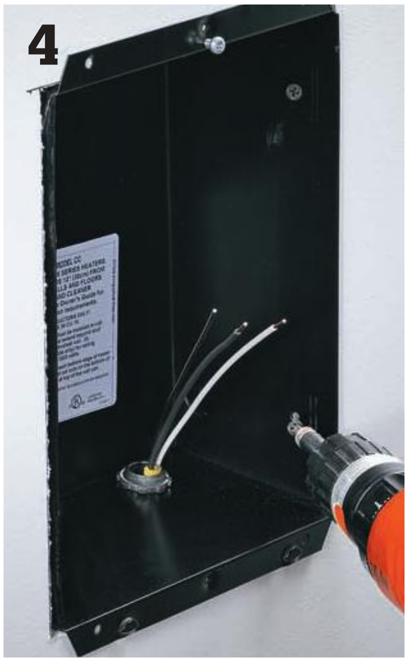

Install the can in the opening. Secure the cable with a clamp, leaving 8" to 12" of cable exposed. Attach the can to the framing as directed by the manufacturer.

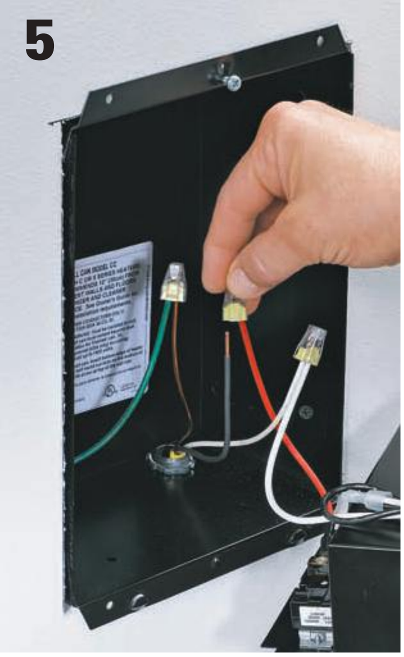

Wire the heater. Connect the black circuit wire to one of the black heater leads. Connect the white circuit wire to the other lead. Connect the grounds.

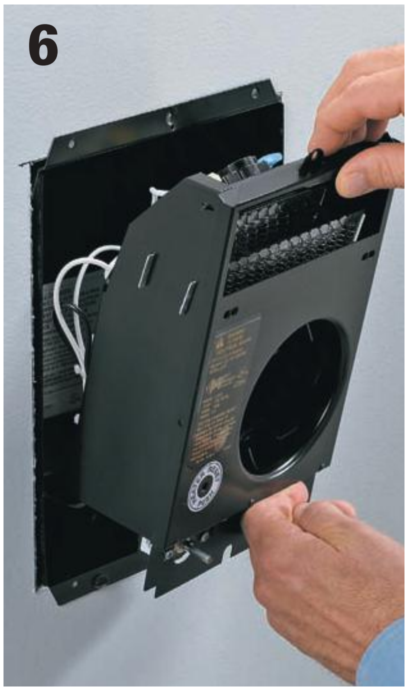

Secure the heater unit in the can as directed by the manufacturer. Reconnect the motor if necessary. Attach the grill and thermostat knob as directed. Connect the new circuit breaker at the main panel.

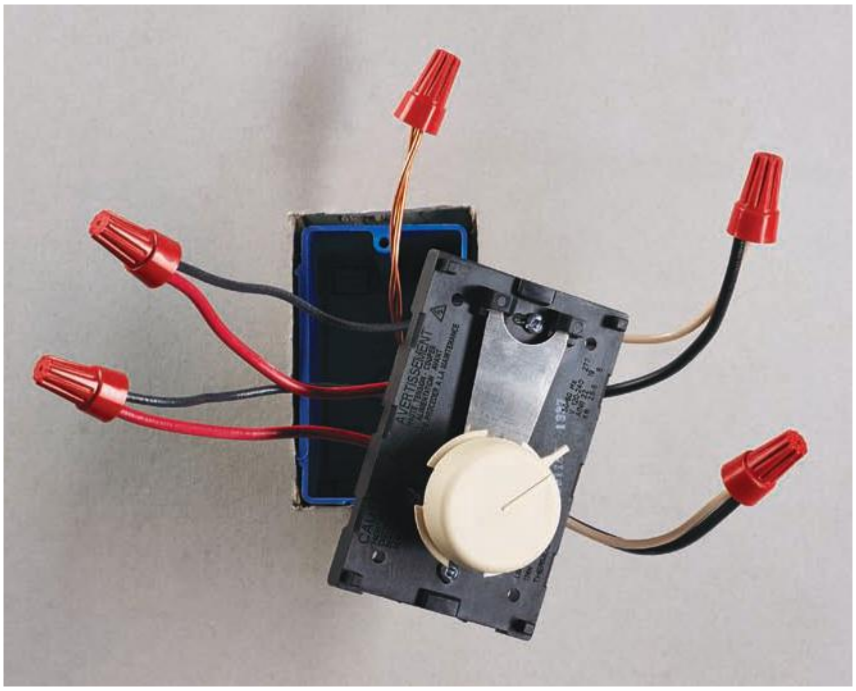

Variation: Connect a thermostat to control a wall heater. Some wall heaters do not use built-in thermostats. Install a thermostat in the heater circuit before the wall heater. Connect the black and the white wires coming from the main panel to the red leads on the thermostat. Connect the wires going to the heater to the black leads on the thermostat. Connect the grounds.

1st proofs for reference only

Edition
age: 72
41940 - Advanced Home Wiring_064-128.indd 73
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 73

Backup Power Supply

Installing a backup generator is an invaluable way to prepare your family for emergencies. The simplest backup power system is a portable gas-powered generator and an extension cord or two. A big benefit of this approach is that you can run a refrigerator and a few worklights during a power outage with a tool that can also be transported to remote job sites or on camping trips when it's not doing emergency backup duty. This is also the least expensive way to provide some backup power for your home. You can purchase a generator at most home centers and be up and running in a matter of hours. If you take this approach, it is critically important that you make certain any loads being run by your generator are disconnected from the utility power source.

The next step up is to incorporate a manual transfer switch for your portable generator. Transfer switches are permanently hardwired to your service panel. They are mounted on either the interior or the exterior of your house between the generator and the service panel. You provide a power feed from the generator into the switch. The switch is wired to selected essential circuits in your house, allowing you to power lights, furnace blowers, and other loads that can't easily be run with an extension cord. But perhaps the most important job a transfer switch performs is to disconnect the utility power. If the inactive utility power line is attached to the service panel, "backfeed" of power from your generator to the utility line can occur when the generator kicks in. This condition could be fatal to line workers who are trying to restore power. The potential for backfeed is the main reason many municipalities insist that only a licensed electrician hook up a transfer switch. Using a transfer switch not installed by a professional may also void the warranty of the switch and the generator.

Automatic transfer switches turn on the generator and switch off the utility supply when they detect a significant drop in line voltage. They may be installed with portable generators, provided the generator is equipped with an electric starter.

Large standby generators that resemble central air conditioners are the top of the line in backup power supply systems. Often fueled by home natural gas lines or propane tanks that offer a bottomless fuel source, standby generators are made in sizes with as much as 20 to 40 kilowatts of output—enough to supply all of the power needs of a 5,000-sq.-ft. home.

Generators have a range of uses. Large hard-wired models can provide instant emergency power for a whole house. Smaller models (below) are convenient for occasional short-term backup as well as job sites or camping trips.

HONDA
EU3000is

proofs for reference only 1st
41940 - Advanced Home Wiring_064-128.indd 74
Pro-Vision C M Y K Text
10/10/14 12:05 AM
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
Page: 74

Options for Backup Generators

A 2,000- to 5,000-watt gas-powered generator and a few extension cords can power lamps and an appliance or two during shorter-term power outages. Appliances must not be connected to household wiring and the generator simultaneously. Never plug a generator into an outlet. Never operate a generator indoors. Run extension cords through a garage door.

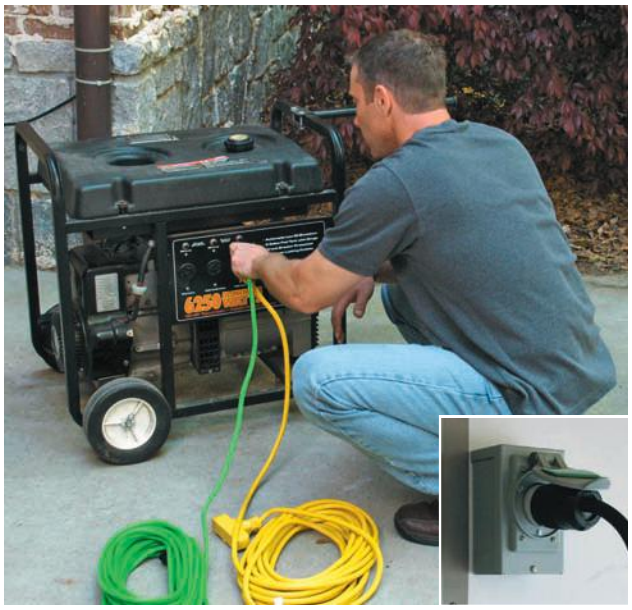

A permanent transfer switch patches electricity from a large portable generator through to selected household circuits via an inlet at your service panel (inset), allowing you to power hardwired fixtures and appliances with the generator.

GUARDIAN QUIETSOURCE

For full, on-demand backup service, install a large standby generator wired through to an automatic transfer panel. In the event of a power outage, the household system instantly switches to the generator.

A Typical Backup System

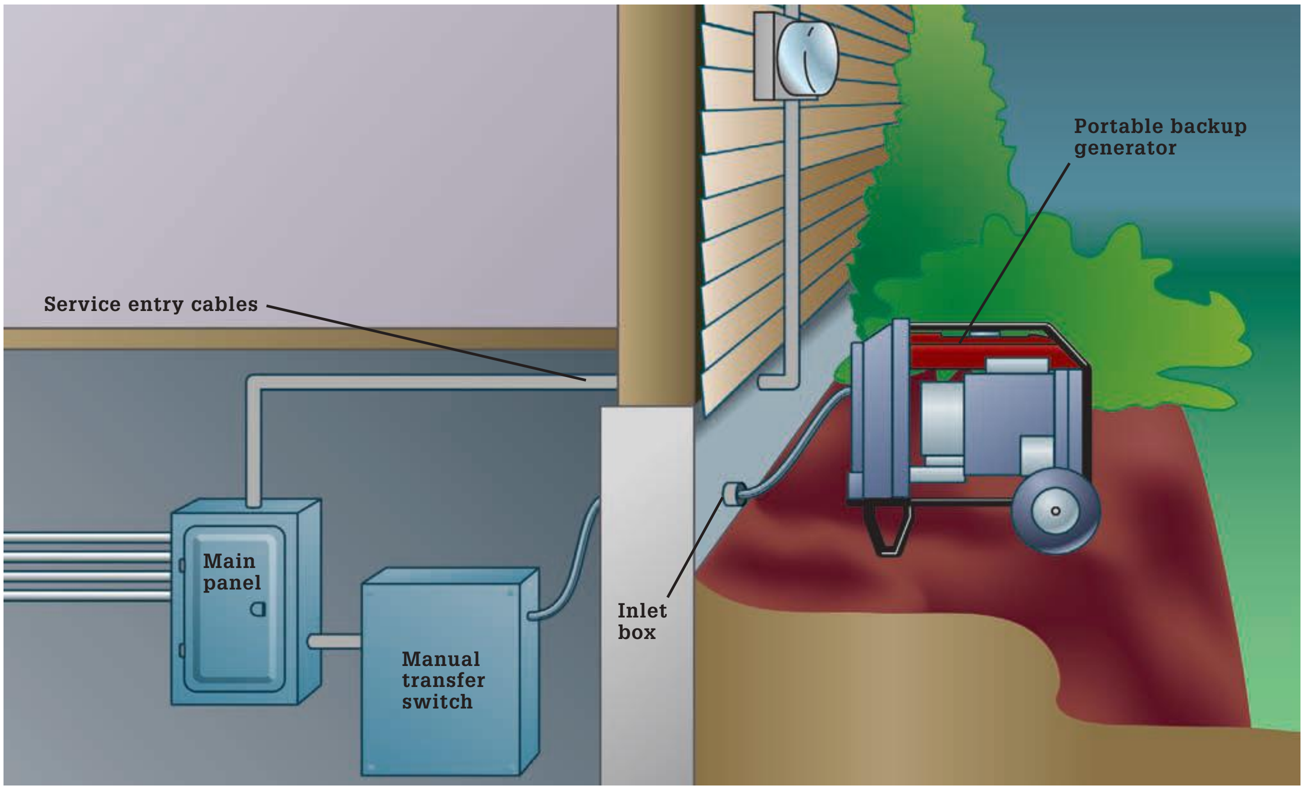

Backup generators supply power to a manual transfer switch, which disconnects the house from the main service wires and routes power from the generator through selected household circuits.

Choosing a Generator

Choosing a generator for your home's needs requires a few calculations. The chart below gives an estimate of the size of generator typically recommended for a house of a certain size. You can get a more accurate number by adding up the power consumption (the watts) of all the circuits or devices to be powered by a generator. It's also important to keep in mind that, for most electrical appliances, the amount of power required at the moment you flip the ON switch is greater than the number of watts required to keep the device running. For instance, though an air conditioner may run on 15,000 watts of power, it will require a surge of 30,000 watts at startup (the power range required to operate an appliance is usually listed somewhere on the device itself). These two numbers are called run watts and surge watts. Generators are typically sold according to run watts (a 5,000-watt generator can sustain 5,000 watts). They are also rated for a certain number of surge watts (a 5,000-watt generator may be able to produce a surge of 10,000 watts). If the surge watts aren't listed, ask, or check the manual. Some generators can't develop many more surge watts than run watts; others can produce twice as much surge as run wattage.

It's not necessary to buy a generator large enough to match the surge potential of all your circuits (you won't be turning everything on simultaneously), but surge watts should factor in your purchasing decision. If you will be operating the generator at or near capacity, it is also a wise practice to stagger startups for appliances.

SIZE OF HOUSE (IN SQUARE FEET)	RECOMMENDED GENERATOR SIZE (IN KILOWATTS)
Up to 2,700	5–11
2,701–3,700	14–16
3,701–4,700	20
4,701–7,000	42–47

1st
proofs for reference only
41940 - Advanced Home Wiring_064-128.indd 76
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 76
Pro-Vis

Types of Transfer Switches

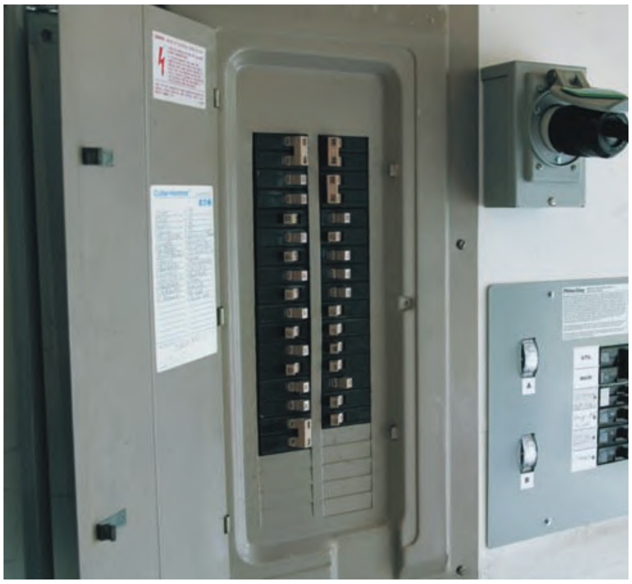

Cord-connected transfer switches (shown above) are hard-wired to the service panel (in some cases they're installed after the service panel and operate only selected circuits). These switches contain a male receptacle for a power supply cord connected to the generator. Automatic transfer switches (not shown) detect voltage drop-off in the main power line and switch over to the emergency power source.

When using a cord-connected switch, consider mounting an inlet box to the exterior wall. This will allow you to connect a generator without running a cord into the house.

Generator Tips ▸

HONDA
CIRCUIT PROTECTOR
AC OUT120V
120/240V
ENGINE SW.
FUEL VALVE
ECO THROTTLE
VOLTAGE SELECTOR

If you'll need to run sensitive electronics such as computers or home theater equipment, look for a generator with power inverter technology that dispenses "clean power" with a stable sine wave pattern.

A generator that will output 240-volt service is required to run most central air conditioners. If your generator has variable output (120/240) make sure the switch is set to the correct output voltage.

1st proofs for reference only

Edition
age: 76
41940 - Advanced Home Wiring_064-128.indd 77
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 12:05 AM
Page: 77

Running & Maintaining a Backup System

Even with a fully automatic standby generator system fueled by natural gas or propane, you will need to conduct some regular maintenance and testing to make sure all systems are ready in the event of power loss. If you're depending on a portable generator and extension cords or a standby generator with a manual transfer switch, you'll also need to know the correct sequence of steps to follow in a power emergency. Switches and panels also need to be tested on a regular basis, as directed in your owner's manual. And be sure that all switches (both interior and exterior) are housed in an approved enclosure box.

Pull-cord starter

Smaller portable generators often use pull-cords instead of electric starters.

ANATOMY OF A PORTABLE BACKUP GENERATOR

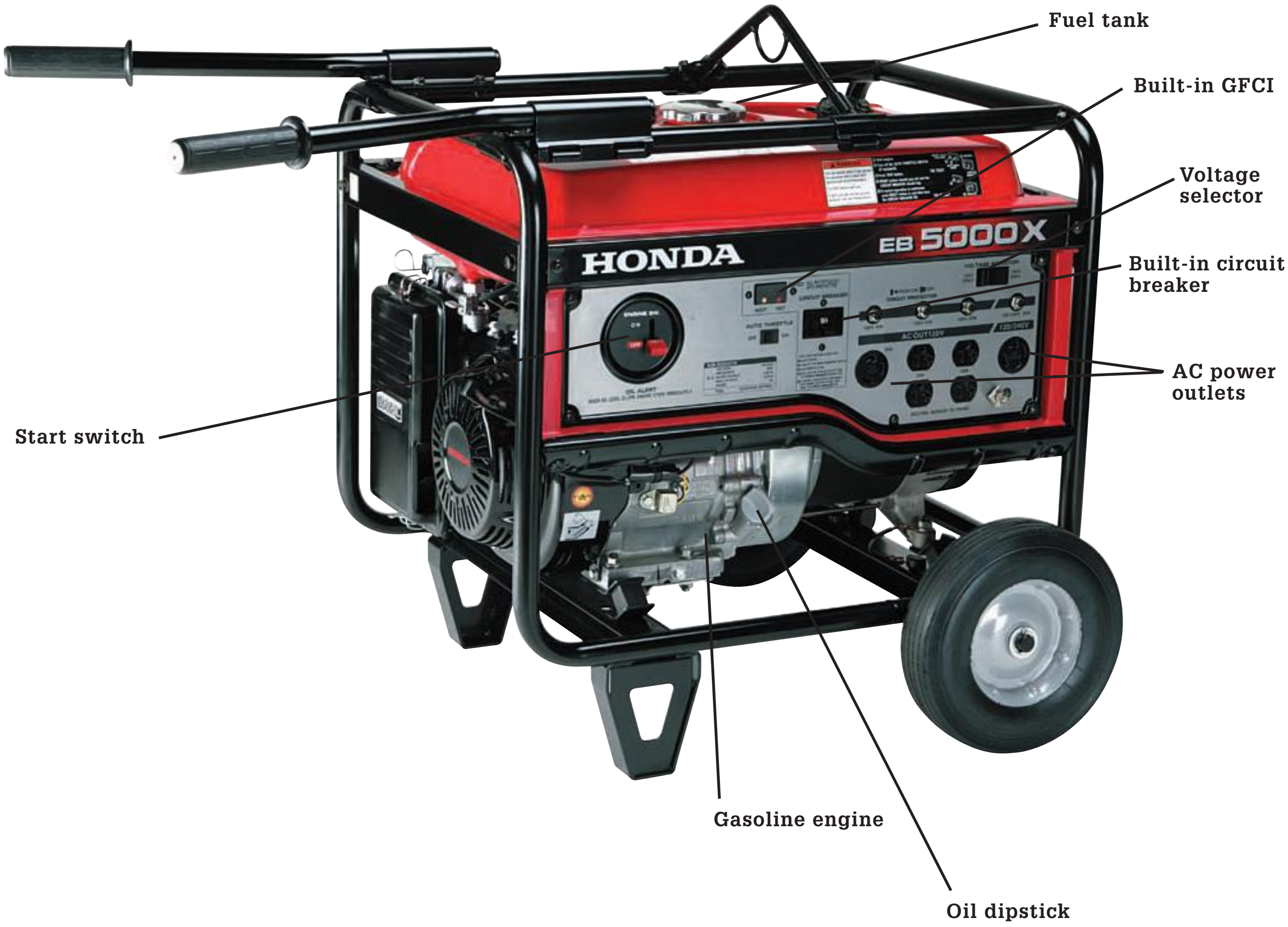

Portable generators use small gasoline engines to generate power. A built-in electronics panel sets current to AC or DC and the correct voltage. Most models will also include a built-in circuit breaker to protect the generator from damage in the event it is connected to too many loads. Better models include features like built-in GFCI protection. Larger portable generators may also feature electric starter motors and batteries for push-button starts.

proofs for reference only 1st
41940 - Advanced Home Wiring_064-128.indd 78
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 12:05 AM
Page: 78
Pro-Vis

Operating a Manual System During an Outage

Plug the generator in at the inlet box. Make sure the other end of the generator's outlet cord is plugged into the appropriate outlet on the generator (120-volt or 120/240-volt AC) and the generator is switched to the appropriate voltage setting.

Start the generator with the pull-cord or electric starter (if your generator has one). Let the generator run for several minutes before flipping the transfer switch.

Flip the manual transfer switch. Begin turning on loads one at a time by flipping breakers on, starting with the ones that power essential equipment. Do not overload the generator or the switch, and do not run the generator at or near full capacity for more than 30 minutes at a time.

Maintaining & Operating an Automatic Standby Generator

If you choose to spend the money and install a dedicated standby generator of 10,000 watts or more and operate it through an automatic transfer switch or panel, you won't need to lift a hand when your utility power goes out. The system kicks in by itself. However, you should follow the manufacturer's instructions for testing the system, changing the oil, and running the motor periodically.

Installing a Transfer Switch

A transfer switch is installed next to the main service panel to override the normal electrical service with power from a backup generator during a power outage. Manual transfer switches require an operator to change the power source, while automatic switches detect the loss of power, start the back-up generator, and switch over to the backup power feed. Because the amount of electricity created by a backup generator is not adequate to power all of the electrical circuits in your house, you'll need to designate a few selected circuits to get backup current (see page 81).

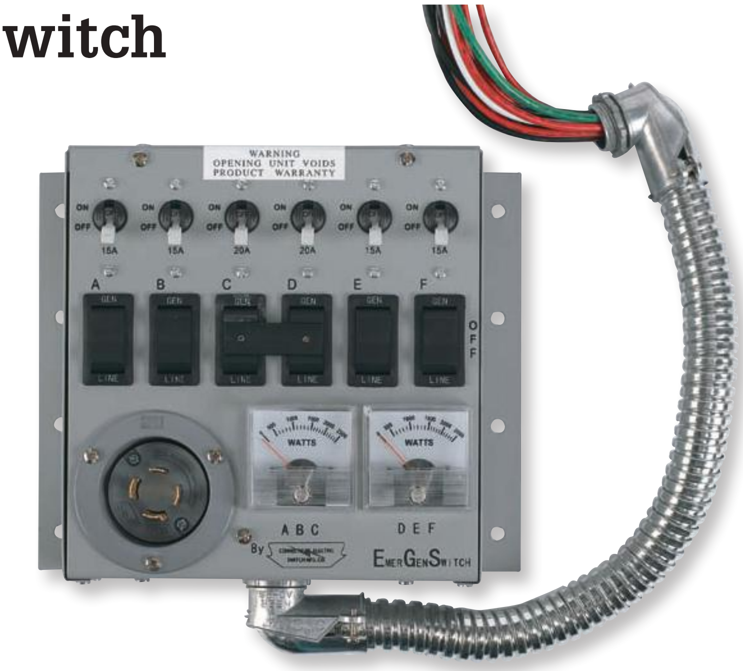

A manual transfer switch connects emergency circuits in your main panel to a standby generator.

Tools & Materials ▸

Circuit tester
Drill/driver
Screwdrivers
Hammer
Wire cutters
Cable ripper
Wire strippers

Level
Manual transfer switch
Screws
Wire connectors (yellow)
Standby power generator

One flip of a switch reassigns the power source for each critical circuit so your backup generator can keep your refrigerator, freezer, and important lights running during an outage of utility power.

Selecting Backup Circuits ▸

Before you purchase a backup generator, determine which loads you will want to power from your generator in the event of a power loss. Generally you will want to power your refrigerator, freezer, and maybe a few lights. Add up the running wattage ratings of the appliances you will power up to determine how large your backup generator needs to be. Because the startup wattage of many appliances is higher than the running wattage, avoid starting all circuits at the same time—it can cause an overload situation with your generator. Here are some approximate running wattage guidelines (see page 132 to 137 for more information on calculating electrical loads):

- Refrigerator: 750 watts
- Forced air furnace: 1,100 to 1,500 watts
- Incandescent lights: 60 watts per bulb (CFL and LED lights use less wattage)
- Sump pump: 800 to 1,000 watts
- Garage door opener: 550 to 1,100 watts
- Television: 300 watts

Add the wattage values of all the loads you want to power, and multiply the sum by 1.25. This will give you the minimum wattage your generator must produce. Portable standby generators typically output 5,000 to 7,500 watts. Most larger, stationary generators can output 10,000 to 20,000 watts (10 to 20 kilowatts).

How to Install a Manual Transfer Switch

1

Turn off the main power breaker in your electrical service panel. CAUTION: The terminals where power enters the main breakers will still be energized.

2

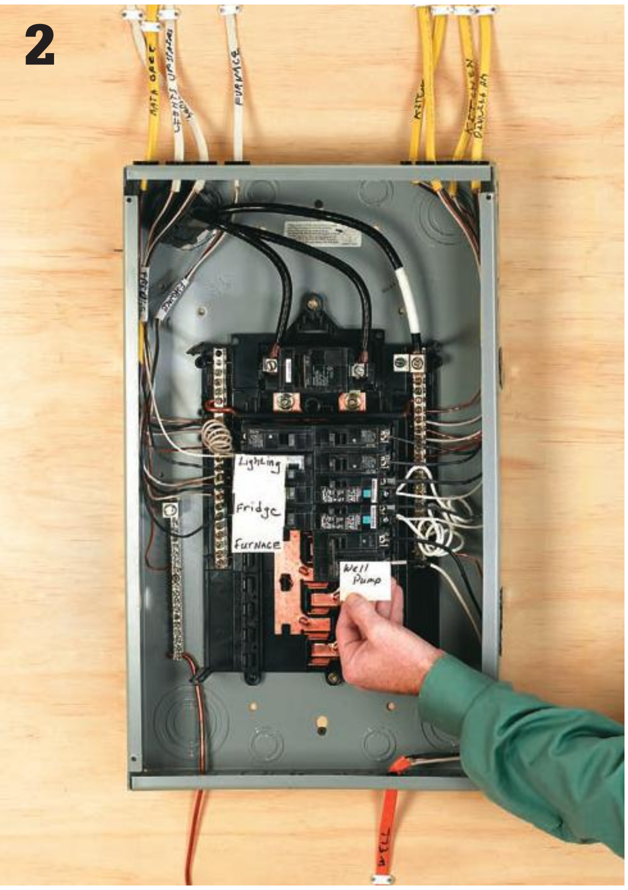

Determine which household circuits are critical for emergency usage during a power outage. Typically this will include the refrigerator, freezer, furnace, and at least one light or small appliance circuit.

(continued)

1st
proofs for reference only

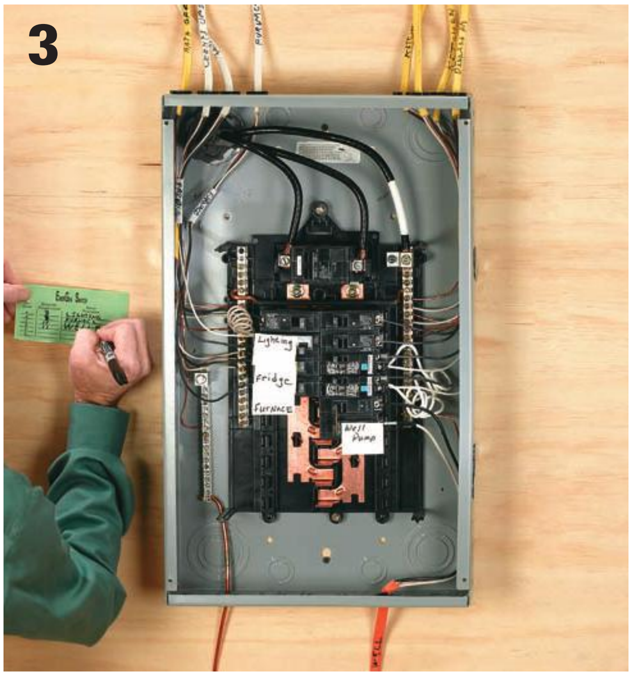

Match your critical circuits with circuit inlet on your pre-wired transfer switch. Try to balance the load as best you can in the transfer switch: For example, if your refrigerator is on the leftmost switch circuit, connect your freezer to the circuit farthest to the right. Double-pole (240-volt) circuits will require two 120-volt circuit connections. Also make sure that 15-amp and 20-amp circuits are not mismatched with one another.

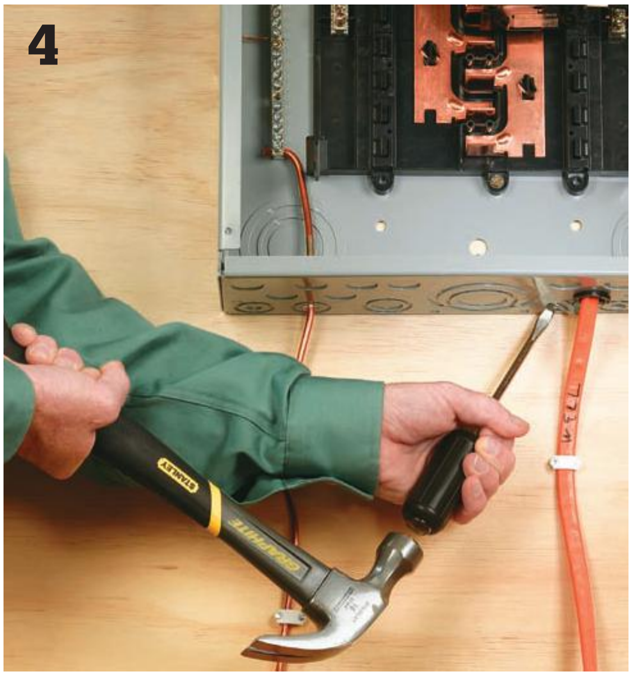

Select and remove a knockout at the bottom of the main service panel box. Make sure to choose a knockout that is sized to match the connector on the flexible conduit coming from the transfer switch.

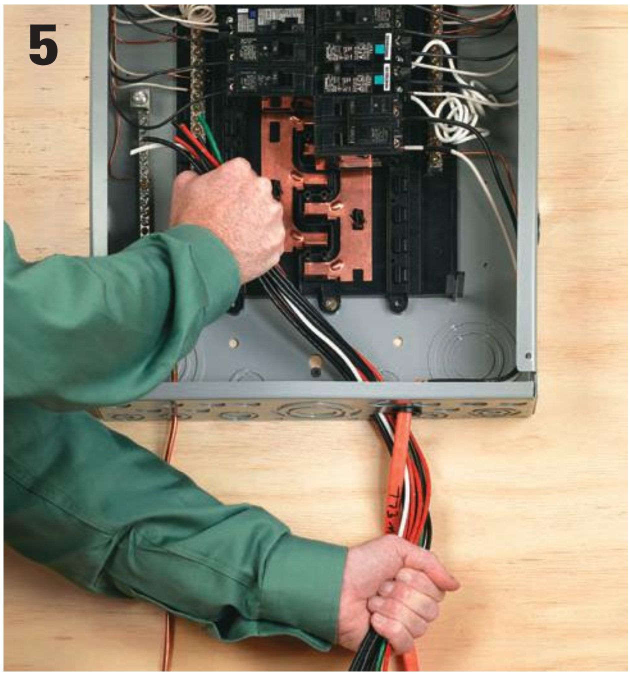

Feed the wires from the transfer switch into the knockout hole, taking care not to damage the insulation. You will note that each wire is labeled according to which circuit in the switch box it feeds.

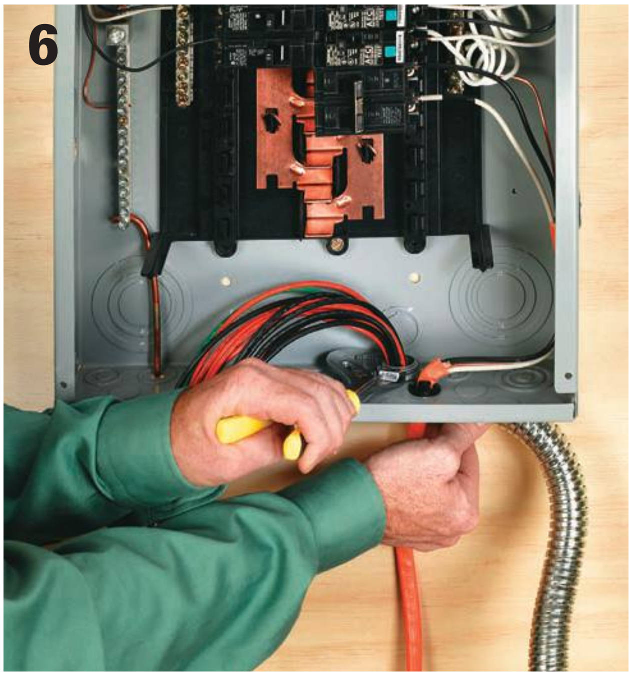

Secure the flexible conduit from the switch box to the main service panel using a locknut and a bushing where required.

41940 - Advanced Home Wiring_064-128.indd 82
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 12:05 AM
Page: 82

7

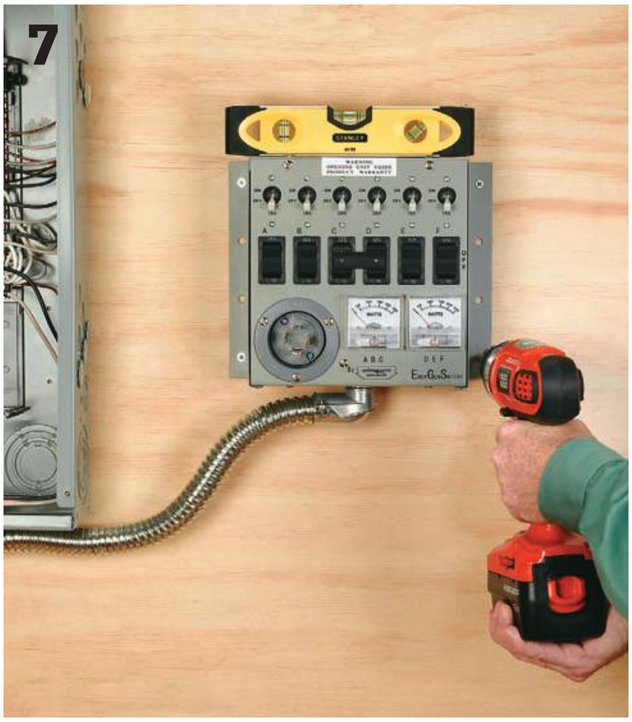

Attach the transfer switch box to the wall so the closer edge is about 18" away from the center of the main service panel. Use whichever connectors make sense for your wall type.

8

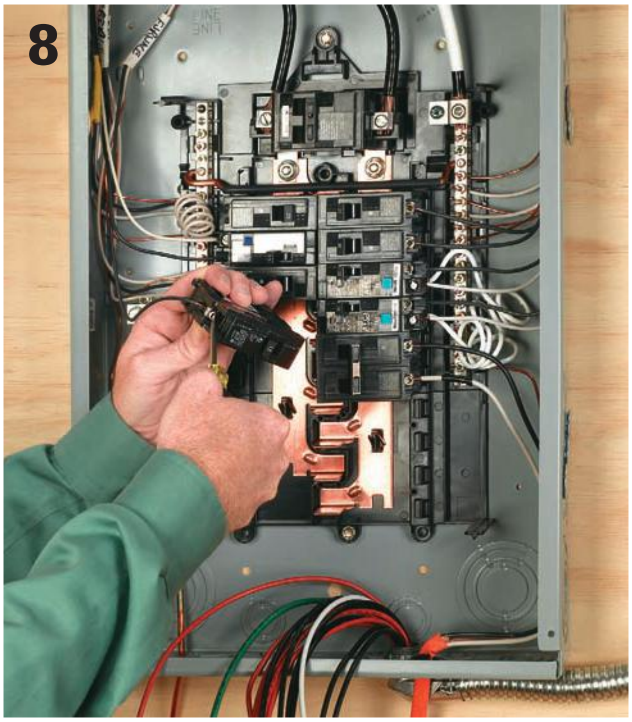

Remove the breaker for the first critical circuit from the main service panel box, and disconnect the hot wire lead from the lug on the breaker.

9

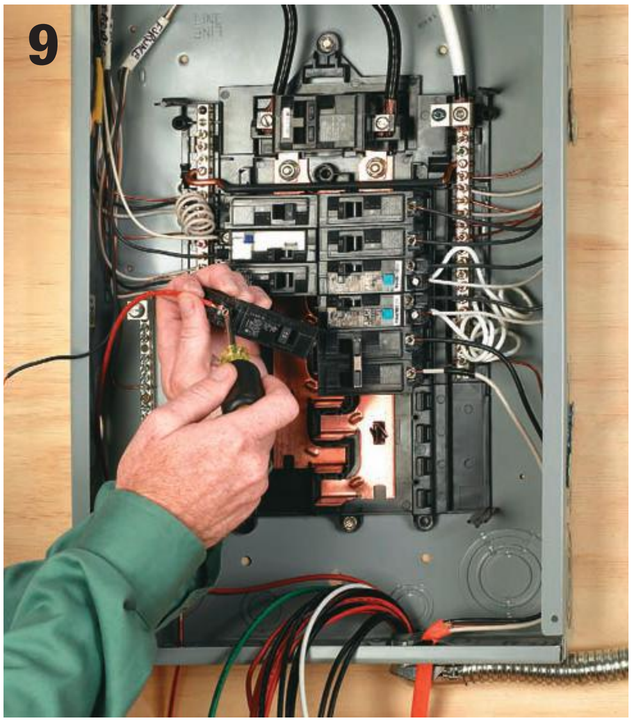

Locate the red wire for the switch box circuit that corresponds to the circuit you've disconnected. Attach the red wire to the breaker you've just removed, and then reinstall the breaker.

10

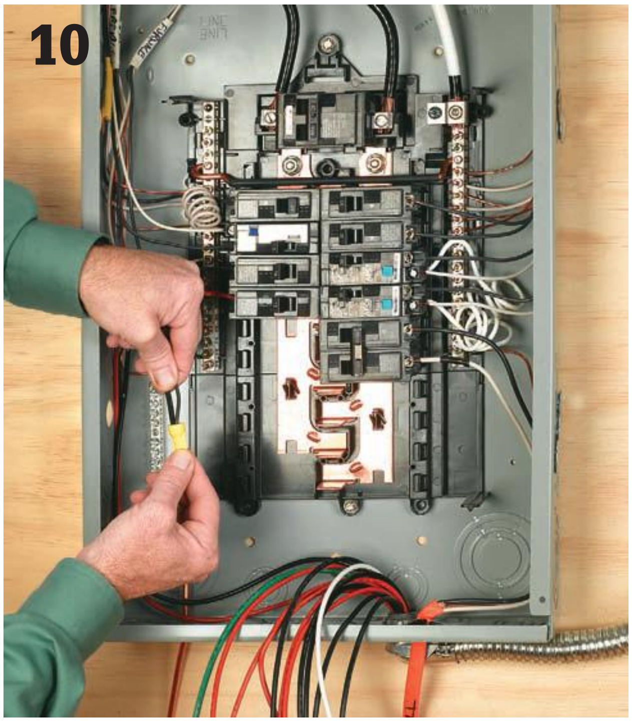

Locate the black wire from the same transfer switch circuit, and twist it together with the old feed wire, using a yellow wire connector. Tuck the wires neatly out of the way at the edges of the box. Proceed to the next circuit, and repeat the process.

(continued)

1st
proofs for reference only

Edition
age: 82
41940 - Advanced Home Wiring_064-128.indd 83
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 83

proofs for reference only 1st

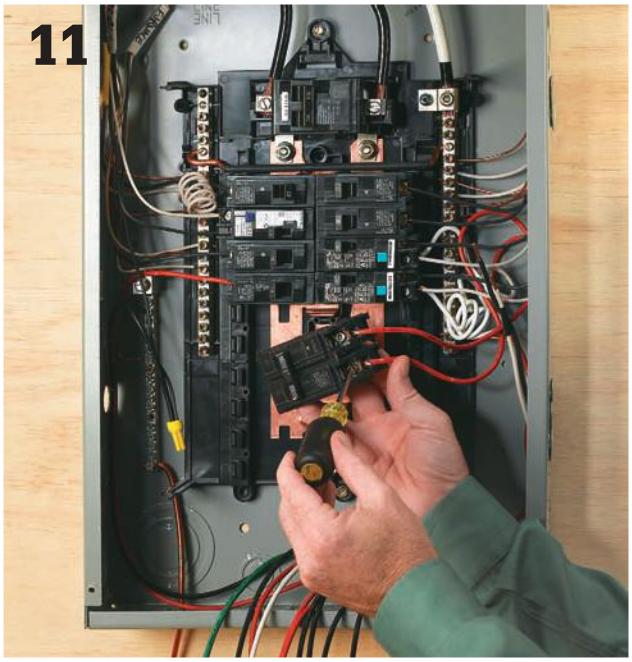

If any of your critical circuits are 240-volt circuits, attach the red leads from the two transfer switch circuits to the double-pole breaker. The two circuits originating in the transfer switch should be next to one another, and their switches should be connected with a handle tie. If you have no 240-volt circuits you may remove the preattached handle tie and use the circuits individually.

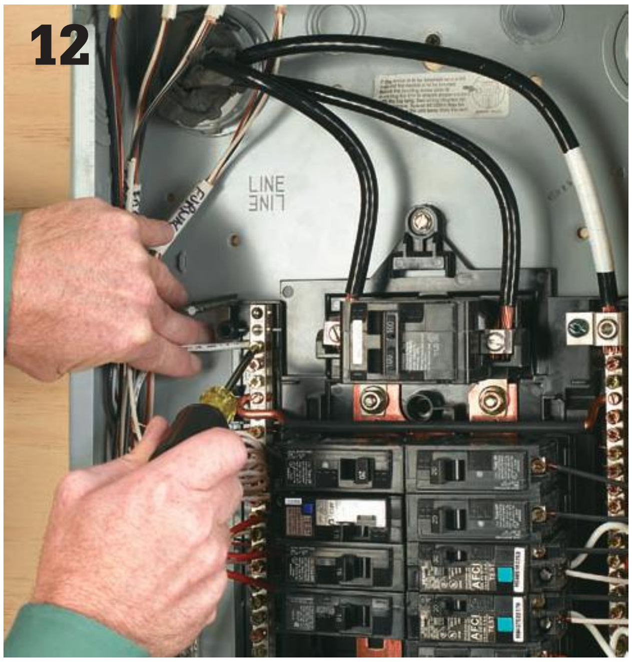

Once you have made all circuit connections, attach the white neutral wire from the transfer switch to an opening in the neutral bus bar of the main service panel.

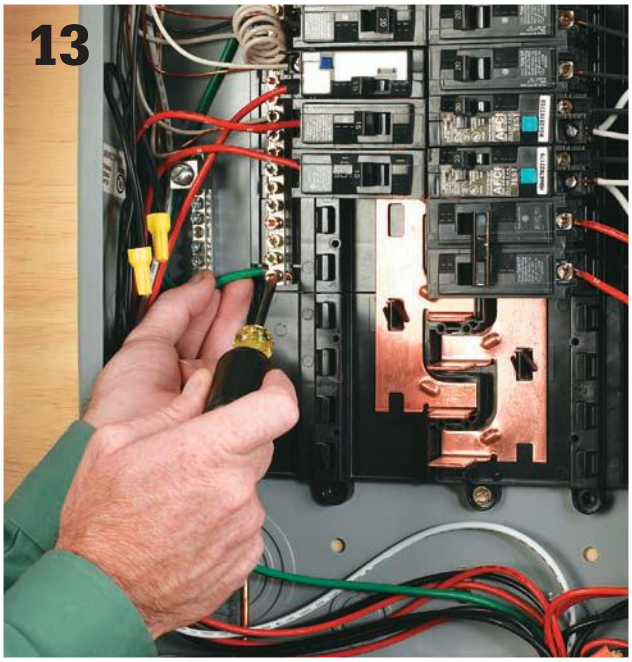

Attach the green ground wire from the transfer switch to an open port on the grounding bar in your main service panel. This should complete the installation of the transfer switch. Replace the cover on the service panel box, and make sure to fill in the circuit map on your switch box.

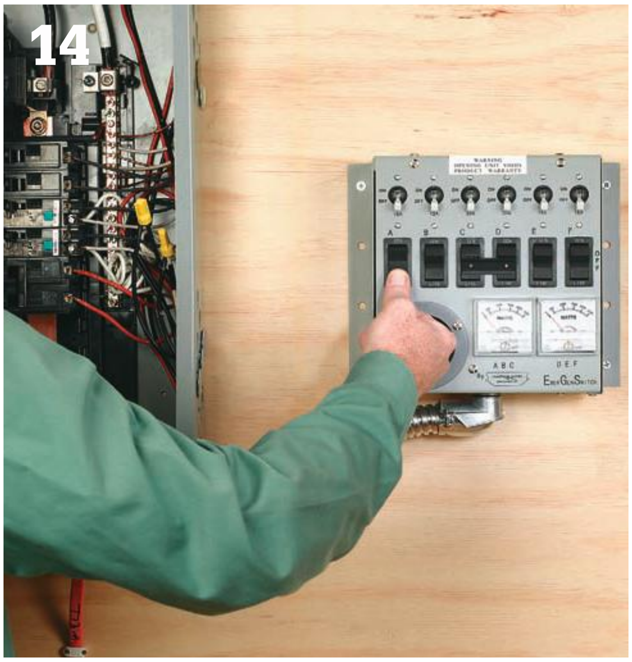

Begin testing the transfer switch by making sure all of the switches on it are set to the LINE setting. The power should still be OFF at the main panel breakers.

41940 - Advanced Home Wiring_064-128.indd 84
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 84

Standby Generators ▸

KOHLER. POWER SYSTEMS

Make sure your standby generator is operating properly and has been installed professionally. See page 76 for information on choosing a generator that is sized appropriately for your needs.

15

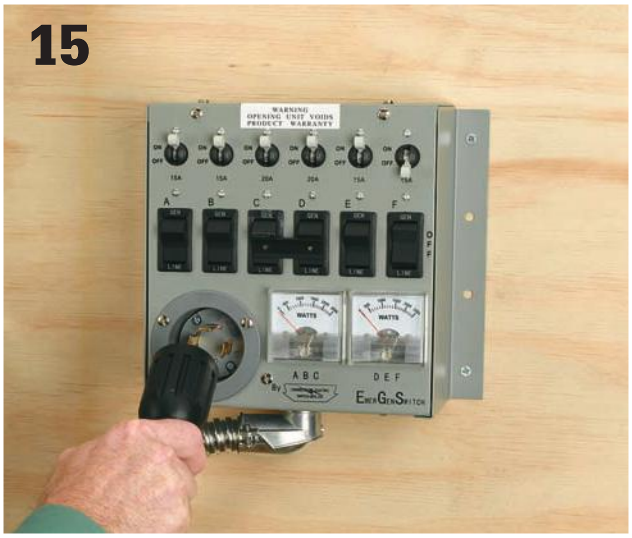

Before turning your generator on, attach the power cord from the generator to the switch box. Never attach or detach a generator cord with the generator running. Turn your standby power generator on, and let it run for a minute or two.

16

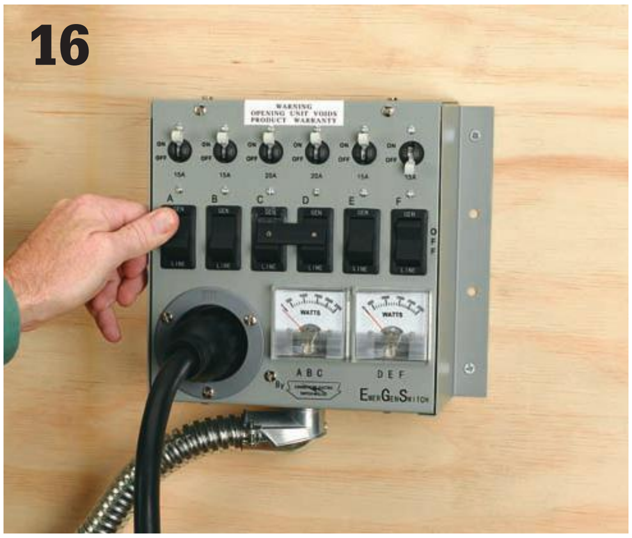

Flip each circuit switch on the transfer switch box to GEN, one at a time. Try to maintain balance by moving back and forth from circuits on the left and right side. Do not turn all circuits on at the same time. Observe the onboard wattage meters as you engage each circuit, and try to keep the wattage levels in balance. When you have completed testing the switch, turn the switches back to LINE, and then shut off your generator.

1st proofs for reference only

Edition
age: 84
41940 - Advanced Home Wiring_064-128.indd 85
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 85

Outbuildings

Nothing improves the convenience and usefulness of an outbuilding more than electrifying it. Running a new underground circuit from your house to an outbuilding lets you add receptacles and light fixtures both inside the outbuilding and on its exterior. If you run power to an outbuilding, you are required to install at least one receptacle.

Adding one or two 120-volt circuits is not complicated, but every aspect of the project is strictly governed by local building codes. Therefore, once you've mapped out the job and have a good idea of what's involved, visit your local building department to discuss your plans and obtain a permit for the work.

This project demonstrates standard techniques for running a circuit cable from the house exterior to a shed, plus the wiring and installation of devices inside the shed. To add a new breaker and make the final circuit connections to your home's main service panel, see page 52. If you run power to an outbuilding, you are required to install at least one receptacle.

First, determine how much current you will need. For basic electrical needs, such as powering a standard light fixture and small appliances or power tools, a 120-volt, 15-amp circuit should be sufficient. A small workshop may require one or two 120-volt, 20-amp circuits. If you need any 240-volt circuits or more than two 120-volt, 20-amp circuits, you will need to install at least a 60-amp subpanel with appropriate feeder wires. Installing a subpanel in an outbuilding is similar to installing one inside your home, but there are some important differences.

You may use #14 copper wire for one 120-volt, 15-amp circuit or #12 copper wire for one 120-volt, 20-amp circuit. Use #10 copper wire for two 120-volt, 20 amp circuits. Also, if the shed is more than 150 ft. away from the house, you may need heavier-gauge cable to account for voltage drop.

Most importantly, don't forget to call before you dig. Have all utility and service lines on your property marked even before you make serious project plans. This is critical for your safety of course, and it may affect where you can run the circuit cable.

Adding an electrical circuit to an outbuilding such as this shed greatly expands the activities the building will support and is also a great benefit for home security.

1st
proofs for reference only

41940 - Advanced Home Wiring_064-128.indd 86
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 86
Pro-Vis

Tools & Materials ▸

- Spray paint
- Trenching shovel (4" wide blade)
- 4" metal junction box
- Metal L-fittings (2) and conduit nipple for conduit
- Wood screws
- Conduit with watertight threaded and compression fittings
- Wrenches
- Hacksaw
- 90° sweeps for conduit (2)
- Plastic conduit bushings (2)
- Pipe straps
- Silicone caulk and caulk gun
- Double-gang boxes, metal (2)
- One exterior receptacle box (with cover)
- Single-pole switches (2)
- Interior ceiling light fixture and metal fixture box
- Exterior motion-detector fixture and plastic fixture box
- EMT metal conduit and fittings for inside the shed
- Utility knife
- UF two-wire cable (12 gauge)
- THNN wire (12 gauge)
- 20-amp GFCI-protected circuit breaker
- Wire stripper
- Pliers
- Screwdrivers
- Wire connectors
- Hand tamper
- Schedule 80 conduit
- Eye protection

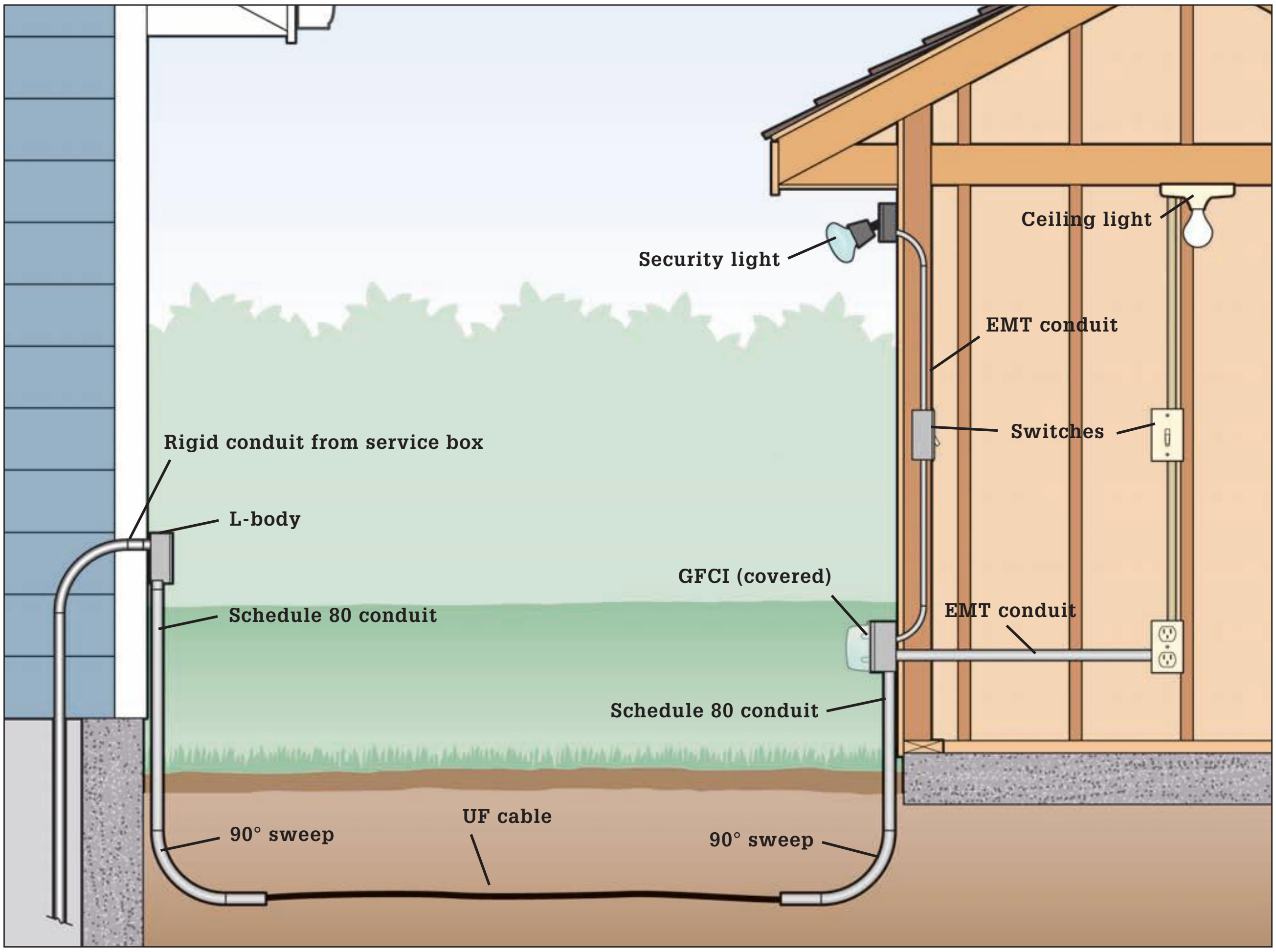
Ceiling light
Security light
EMT conduit
Switches
Rigid conduit from service box
L-body
GFCI (covered)
Schedule 80 conduit
EMT conduit
Schedule 80 conduit
UF cable
90° sweep
90° sweep

A basic outdoor circuit starts with a waterproof fitting at the house wall connected to a junction box inside. The underground circuit cable—rated UF (underground feeder)—runs in a 24"-deep trench and is protected from exposure at both ends by metal or PVC conduit. Inside the shed, standard NM cable runs through metal conduit to protect it from damage (not necessary if you will be adding interior wallcoverings). All receptacles in the shed must be GFCI protected.

1st
proofs for reference only
Edition
age: 86
41940 - Advanced Home Wiring_064-128.indd 87
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 12:05 AM
Page: 87

How to Wire an Outbuilding

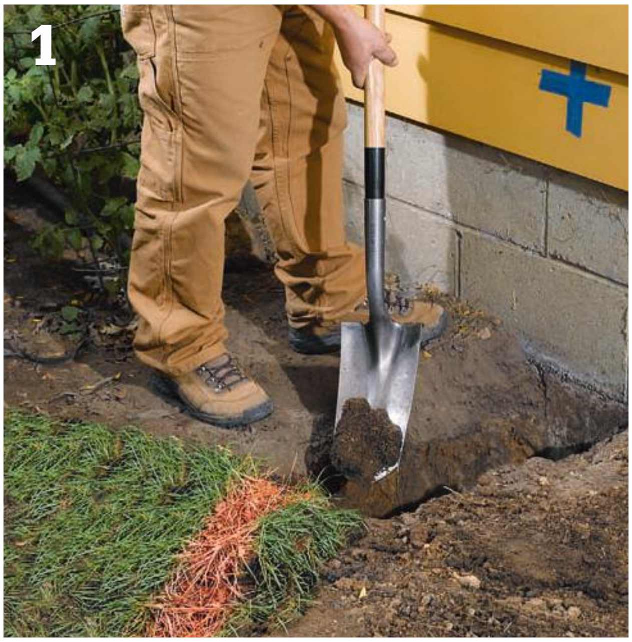

Identify the circuit's exit point at the house and entry point at the shed and mark them. Mark the path of the trench between the exit and entry points using spray paint. Make the route as direct as possible. Dig the trench to the depth required by local code (24") using a narrow trenching shovel.

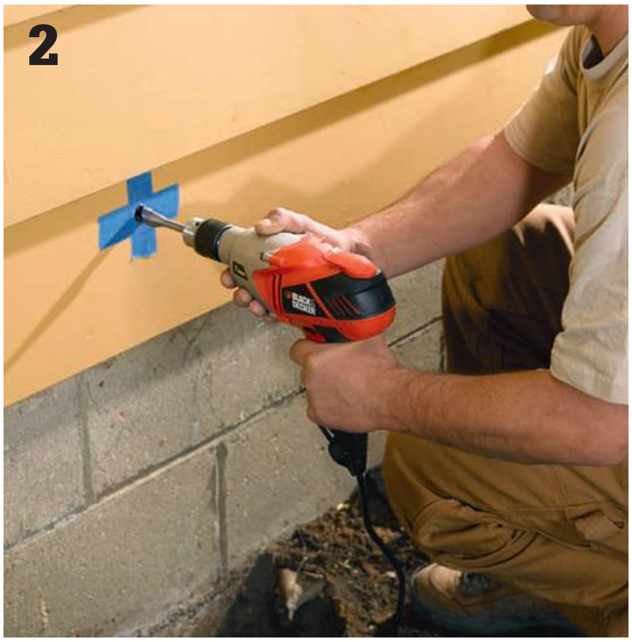

From outside, drill a hole through the exterior wall and the rim joist at the exit point for the cable (you'll probably need to install a bit extender or an extra-long bit in your drill). Make the hole just large enough to accommodate the L-body conduit fitting and conduit nipple.

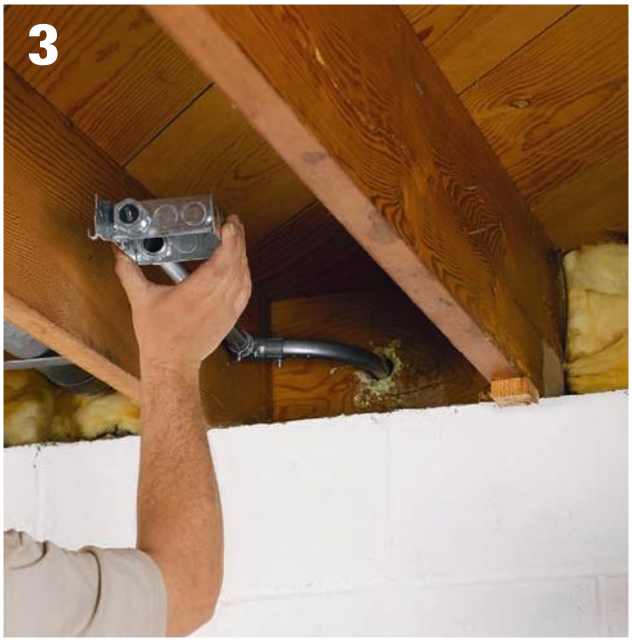

Assemble the conduit and junction box fittings that will penetrate the wall. Here, we attached a 12" piece of ¾" IMC (intermediate metallic conduit) and a sweep to a metal junction box with a compression fitting and then inserted the conduit into the hole drilled in the rim joist. The junction box is attached to the floor joist.

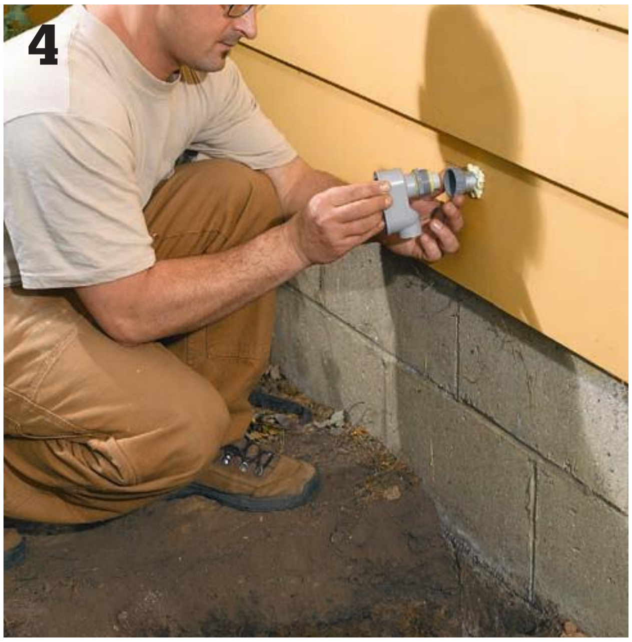

From outside, seal the hole around the conduit with expandable spray foam or caulk, and then attach the free end of the conduit to the back of a waterproof L-body fitting. Mount the L-body fitting to the house exterior with the open end facing downward.

1st proofs for reference only
41940 - Advanced Home Wiring_064-128.indd 88
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 88
Pro-Vis

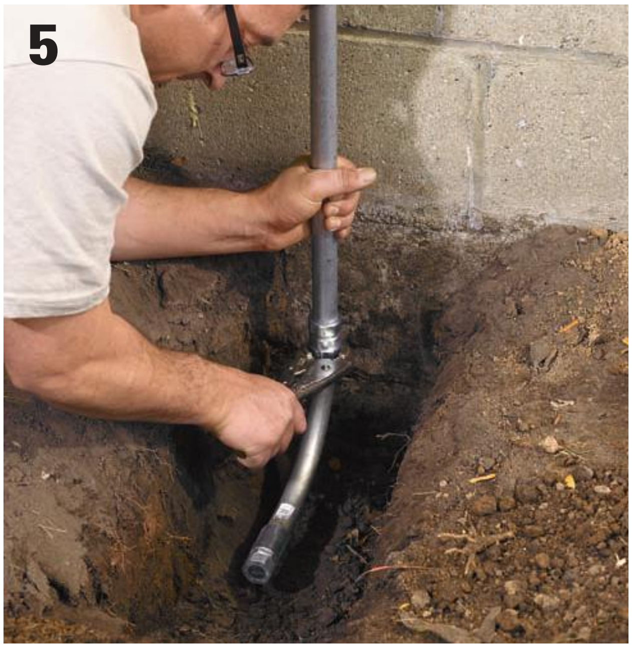

Cut a length of IMC to extend from the L-fitting down into the trench using a hacksaw. Deburr the cut edges of the conduit. Secure the conduit to the L-fitting, and then attach a 90° sweep to the bottom end of the conduit using compression fittings. Add a bushing to the end of the sweep to protect the circuit cable. Anchor the conduit to the wall with a corrosion-resistant pipe strap.

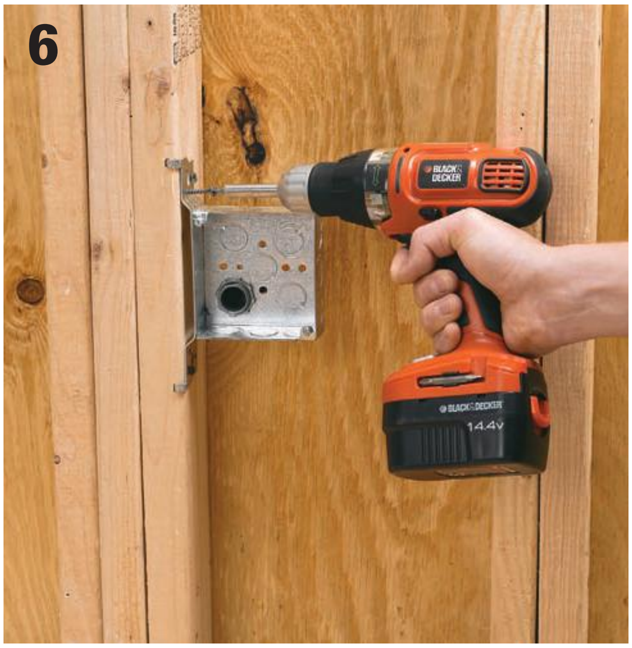

Inside the shed, drill a ¾" dia. hole in the shed wall. On the interior of the shed, mount a junction box with a knock-out removed to allow the cable to enter through the hole. On the exterior side directly above the end of the UF trench, mount an exterior-rated receptacle box with cover. The plan (and your plan may differ) is to bring power into the shed through the hole in the wall behind the exterior receptacle.

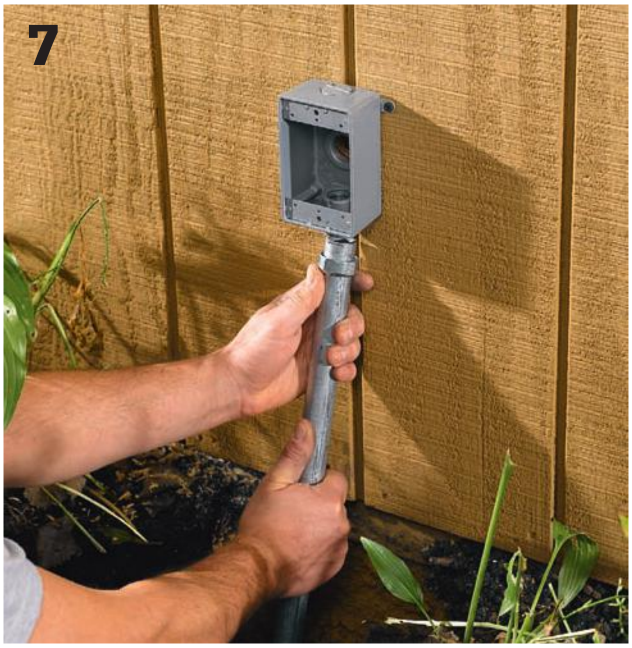

Run conduit from the exterior box down into the trench. Fasten the conduit to the building with a strap. Add a 90° sweep and bushing, as before. Secure the conduit to the box with an offset fitting. Anchor the conduit with pipe straps, and seal the entry hole with caulk.

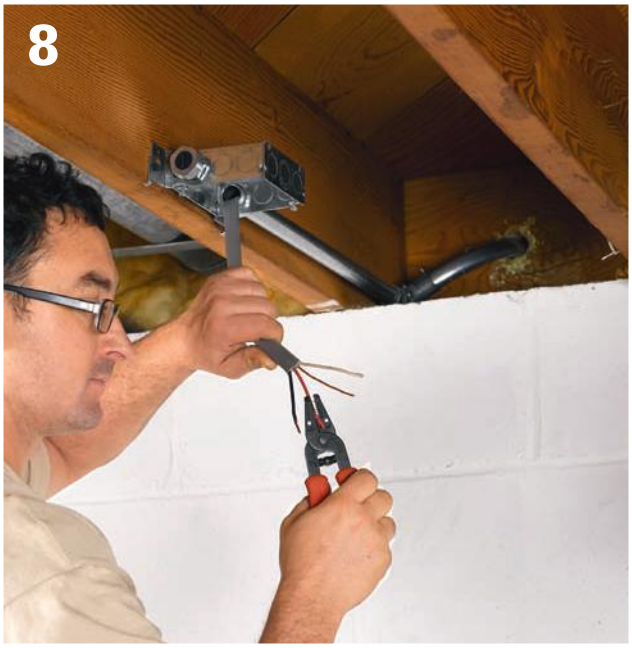

Run UF cable from the house to the outbuilding. Feed one end of the UF circuit cable up through the sweep and conduit and into the L-fitting at the house (the back or side of the fitting is removable to facilitate cabling). Run the cable through the wall and into the junction box, leaving at least 12" of extra cable at the end.

(continued)

1st
proofs for reference only

Edition
age: 88
41940 - Advanced Home Wiring_064-128.indd 89
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 12:05 AM
Page: 89

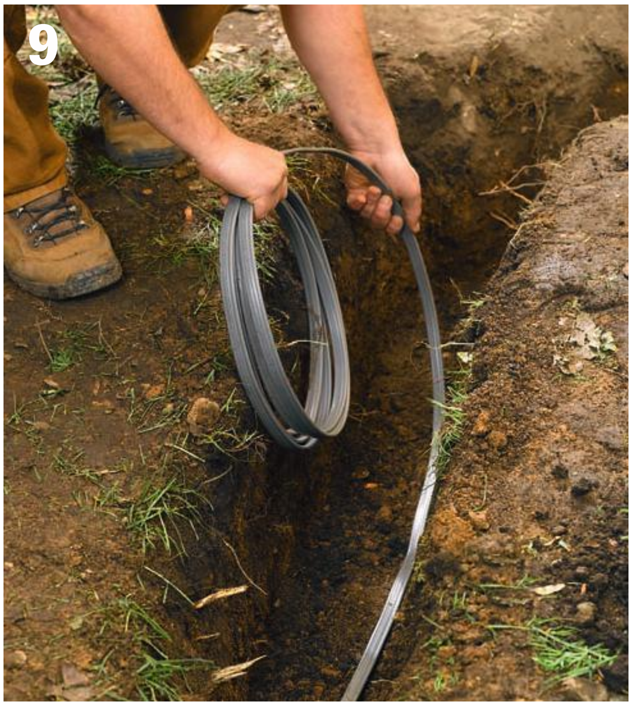

Lay the UF cable into the trench, making sure it is not twisted and will not contact any sharp objects. Roll out the cable, and then feed the other end of the cable up through the conduit and into the receptacle box in the shed, leaving 12" of slack.

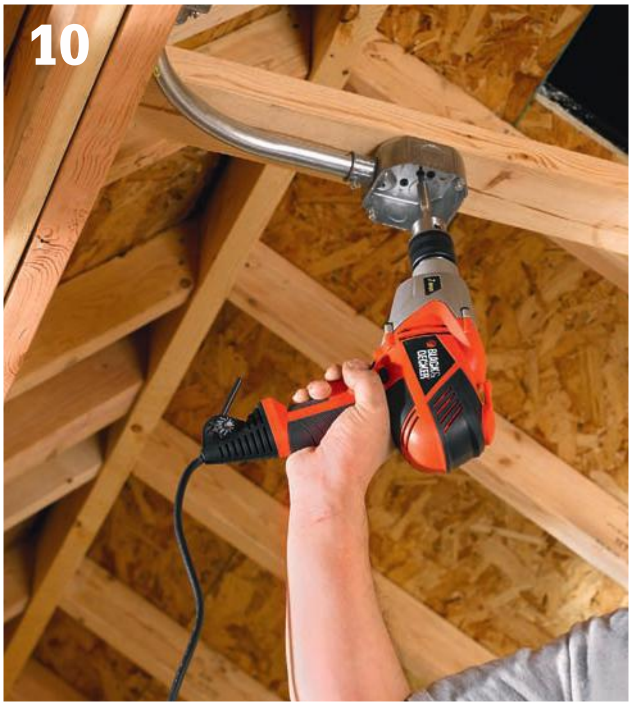

Inside the outbuilding, install the remaining boxes for the other switches, receptacles, and lights. With the exception of plastic receptacle boxes for exterior exposure, use metal boxes if you will be connecting the boxes with metal conduit.

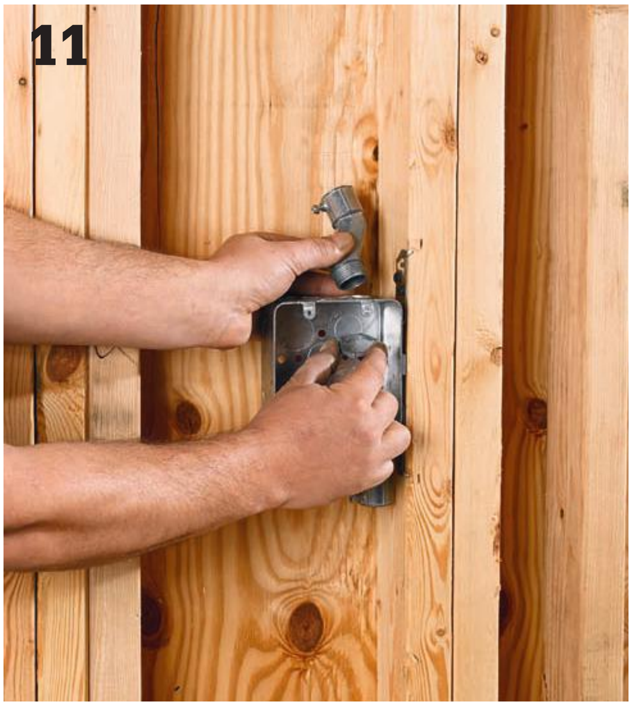

Connect the electrical boxes with conduit and fittings. Inside the outbuilding, you may use inexpensive EMT to connect receptacle, switch, and fixture boxes. Once you've planned your circuit routes, start by attaching couplings to all of the boxes.

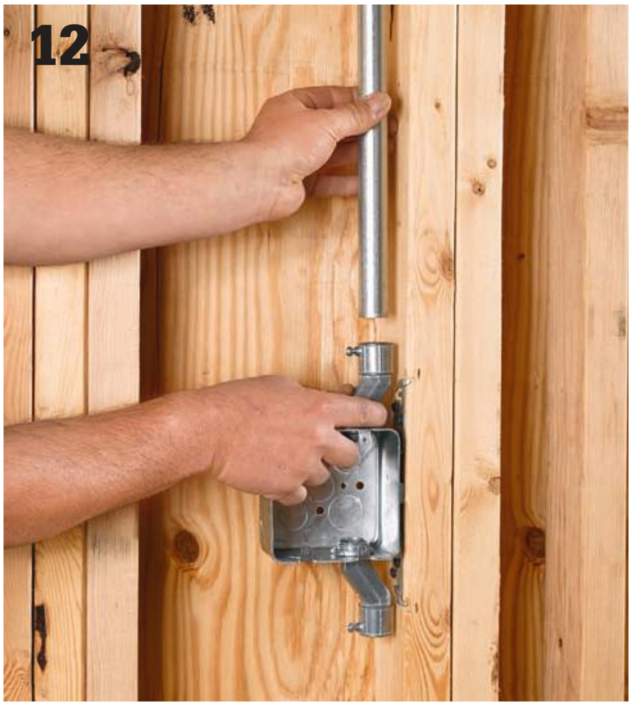

Cut a length of conduit to fit between the coupling and the next box or fitting in the run. If necessary, drill holes for the conduit through the centers of the wall studs. Attach the conduit to the fitting that you attached to the first box.

1st
proofs for reference only

41940 - Advanced Home Wiring_064-128.indd 90
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 90
Pro-Vis

13

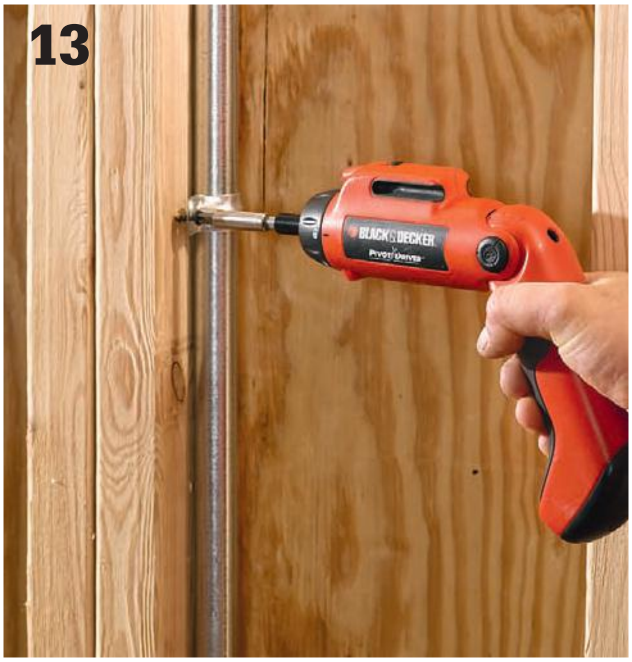

If you are surface-mounting the conduit or running it up or down next to wall studs, secure it with straps no more than 3 ft. apart. Use elbow fittings for 90° turns and setscrew couplings for joining straight lengths as needed. Make holes through the wall studs only as large as necessary to feed the conduit through.

14

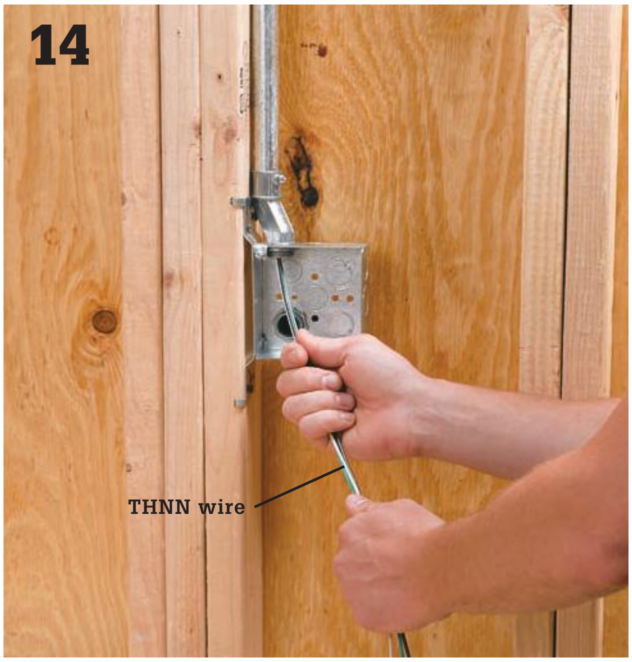

Measure to find how much wire you'll need for each run, and cut pieces that are a foot or two longer. Before making L-turns with the conduit, feed the wire through the first conduit run.

15

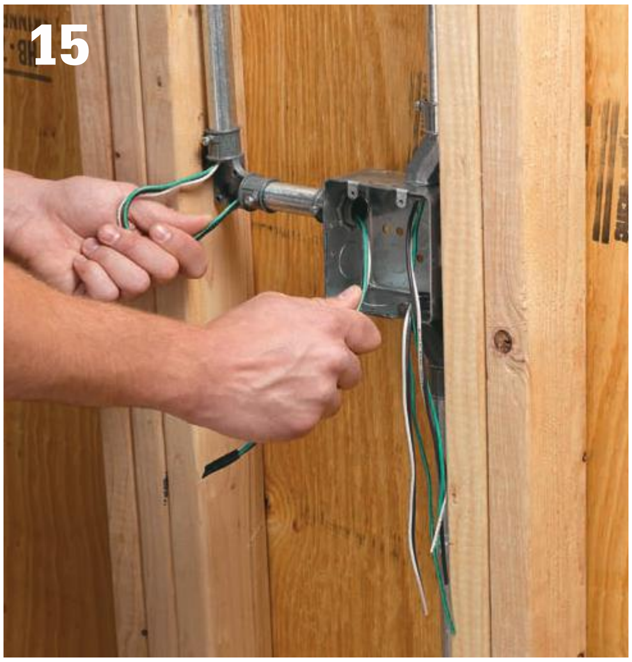

Feed the other ends of the wires into the next box or fitting in line. It is much easier to feed wire into 45° and 90° elbows if they have not been attached to the conduit yet. Continue feeding wire into the conduit and fitting until you have reached the next box in line.

16

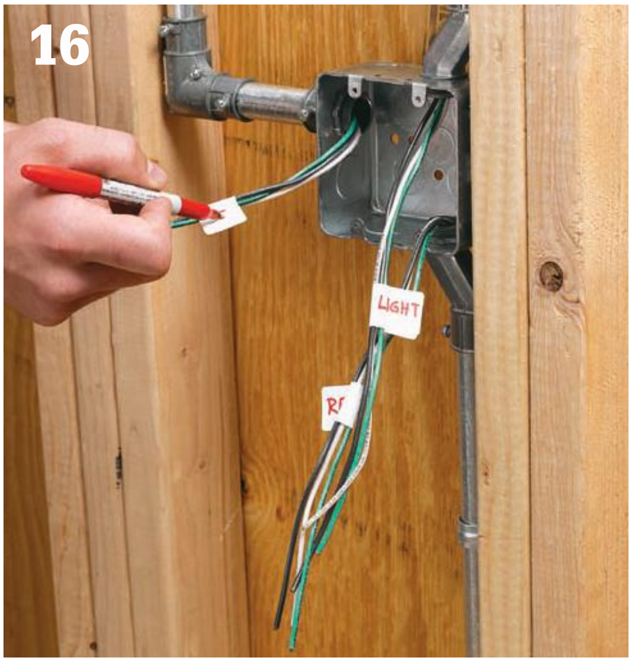

Once you've reached the next box in line, coil the ends of the wires and repeat the process with new wire for the next run. Keep working until all of the wire is run and all of the conduit and fittings are installed and secured. If you are running multiple feed wires into a single box, write the origin or destination on a piece of masking tape and stick it to each wire end.

(continued)

1st proofs for reference only
Edition
age: 90
41940 - Advanced Home Wiring_064-128.indd 91
Pro-Vision C M Y K Text
10/10/14 12:05 AM
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
Page: 91

17

Note: Your Code may require an in-use rated receptacle box cover.

Make the wiring connections at the receptacles. Strip ¾" of insulation from the circuit wires using a wire stripper. Connect the white (neutral) wire and black (hot) wire of the UF cable to the LINE screw terminals on the receptacle. Connect the white (neutral) and black (hot) wires from the NM cable to the LOAD terminals. Pigtail the bare copper ground wires and connect them to the receptacle ground terminal and the metal box. Install the receptacle and cover plate.

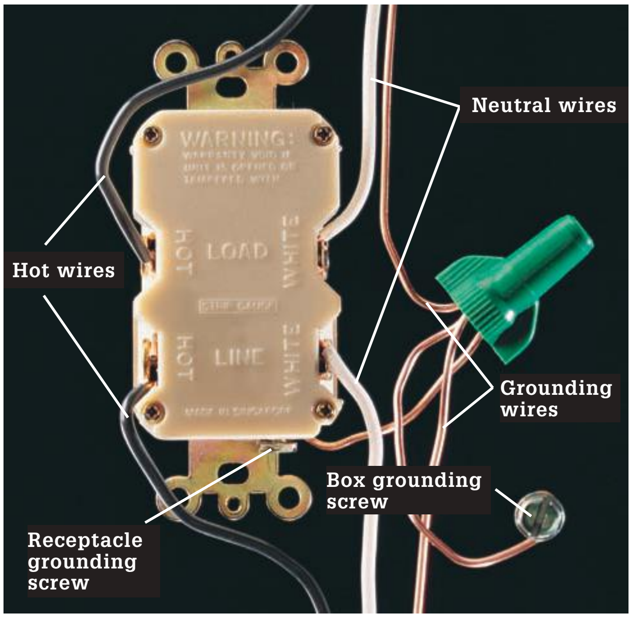

Variation: Installing a GFCI-protected breaker for the new circuit at the main service panel is the best way to protect the circuit and allows you to use regular receptacles in the building, but an alternative that is allowed in many areas is to run the service into a GFCI-protected receptacle and then wire the other devices on the circuit in series. If you use this approach, only the initial receptacle needs to be a GFCI receptacle.

18

Continue installing receptacles in the circuit run, and then run service from the last receptacle to the switch box for the light fixture or fixtures. (If you anticipate a lot of load on the circuit, you should probably run a separate circuit for the lights). Twist the white neutral leads and grounding leads together and cap them. Attach the black wires to the appropriate switches. Install the switches and cover plate.

19

Install the light fixtures. For this shed, we installed a caged ceiling light inside the shed and a motion-detector security light on the exterior side.

1st proofs for reference only
41940 - Advanced Home Wiring_064-128.indd 92
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 92
Pro-Vis

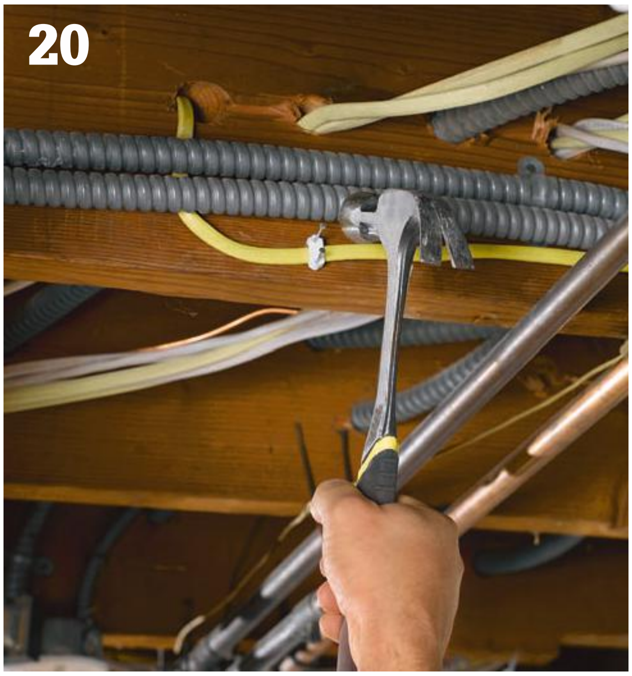

Run NM cable from the electrical box in the house at the start of the new circuit to the main service panel. Use cable staples if you are running the cable in floor joist cavities. If the cable is mounted to the bottom of the floor joists or will be exposed, run it through conduit.

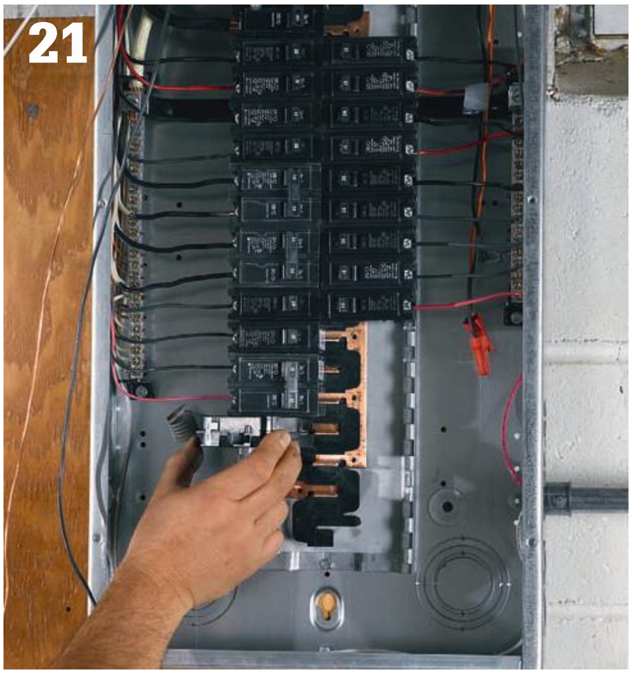

At the service panel, feed the NM cable in through a cable clamp. Arrange for your final electrical inspection before you install the breaker. Then attach the wires to a new circuit breaker, and install the breaker in an empty slot. Label the new circuit on the circuit map.

Turn on the new circuit, and test all of the receptacles and fixtures. Depress the Test button and then the Reset button if you installed a GFCI receptacle. If any of the fixtures or receptacles is not getting power, check the connections first, and then test the receptacle or switch for continuity with a multimeter. Backfill the trench.

1st proofs for reference only
Edition
age: 92
41940 - Advanced Home Wiring_064-128.indd 93
Pro-Vision C M Y K Text
10/10/14 12:05 AM
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
Page: 93

Standalone Solar Lighting System

A self-contained electrical circuit with dedicated loads, usually 12-volt light fixtures, is one of the most useful solar amenities you can install. A standalone system is not tied into your power grid, which greatly reduces the danger of installing the components yourself. Plus, the fact that your light fixtures are independent of the main power source means that even during a power outage you will have functioning emergency and security lights.

Installing a single solar-powered circuit is relatively simple, but don't take the dangers for granted. Your work will require permits and inspections in most jurisdictions, and you can't expect to pass if the work is not done to the exact specifications required.

Solar panels that convert the sun's energy into electricity are called photovoltaic (PV) panels, and they produce direct current (DC) power. PV solar panel systems can be small and designed to accomplish a specific task, or they can be large enough to provide power or supplementary power to an entire house. Before you make the leap into a large system, it's a good idea to familiarize yourself with the mechanics of solar power. The small system demonstrated in this project is relatively simple and is a great first step into the world of solar. The fact that the collector, battery, and lights are a standalone system makes this a very easy project to accomplish. By contrast, installing panels that provide direct supplementary power through your main electrical service panel is a difficult wiring job that should be done by professional electricians only.

This 60-watt solar panel is mounted on a garage roof and powers a self-contained home security lighting system. Not only does this save energy costs, it keeps the security lights working even during power outages.

1st
proofs for
reference
only

41940 - Advanced Home Wiring_064-128.indd 94
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 12:05 AM
Page: 94
Pro-Vis

Schematic Diagram for an Off-the-Grid Solar Lighting System

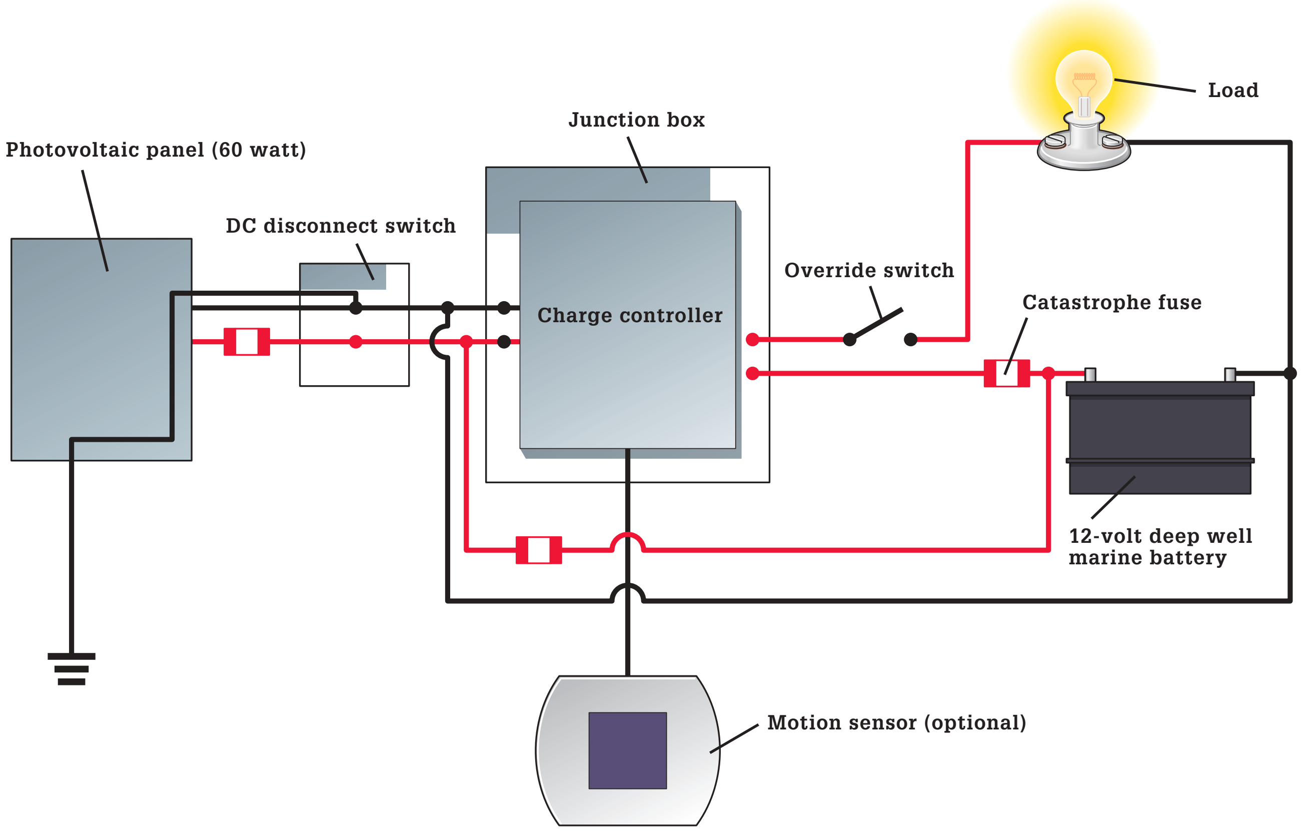
Load
Junction box
Photovoltaic panel (60 watt)
DC disconnect switch
Charge controller
Override switch
Catastrophe fuse
12-volt deep well marine battery
Motion sensor (optional)

Tools & Materials ▸

Tape measure
Drill/driver with bits
Caulk gun
Crimping tool
Wiring tools
Metal-cutting saw
Photovoltaic panel (50 to 80 watts)
Charge controller
Catastrophe fuse
Battery sized for 3 day autonomy
Battery case
Battery cables
12-volt LED lights including motion-sensor light
Additional 12-volt light fixtures as desired
20 ft. Unistrut 1⅞" thick U-channel (See Resources, page 125)
(4) 45° Unistrut connectors
(2) 90° Unistrut angle brackets
(4) Unistrut hold down clamps
(12) ⅜" spring nuts
(12) ⅜"-dia. × 1"-long hex-head bolts with washers
DC-rated disconnect or double throw snap switch
6" length of ½"-dia. liquid-tight flexible metallic conduit
(2) ½" liquid tight connectors
(2) Lay-in grounding lugs
(2) Insulated terminal bars to accept one 2-gauge wire and 4 12-gauge wires
(2) Cord cap connectors for ½"-dia. cable
½" ground rod and clamp
Copper wire (6, 12-gauge)
Green ground screws
½" Flexible metallic conduit or Greenfield
½" Greenfield connectors
(4) 11⁄16" junction boxes with covers
(4) square boxes with covers
PVC 6"× 6" junction box with cover
14/2 UF wire
¼" × 20 nuts and bolts with lock washers
Roof flashing boot
Roof cement
Silicon caulk
Eye protection

Mounting PV Panels

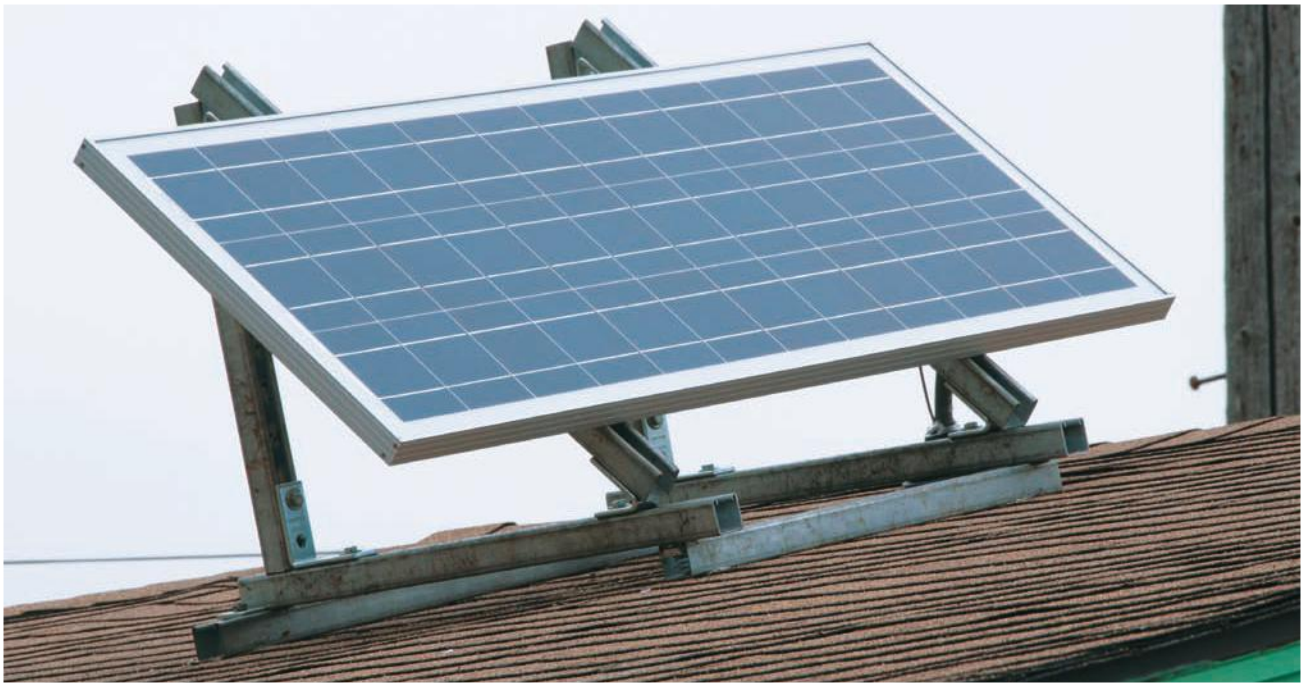

The mounting stand for the PV panel is constructed from metal U-channel (a product called Unistrut is seen here. See Resources page 125) and pre-bent fasteners. Position the solar panel where it will receive the greatest amount of sunlight for the longest period of time each day—typically the south-facing side of a roof or wall. For a circuit with a battery reserve that powers two to four 12-volt lights, a collection panel rated between 40 and 80 watts of output should suffice. These panels can range from $200 to $600 in price, depending on the output and the overall quality.

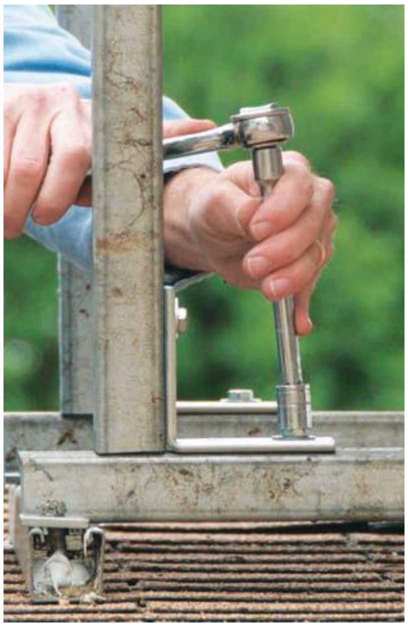

The stand components are held together with bolts and spring-loaded fasteners. The 45° and 90° connectors are manufactured specifically for use with this Unistrut system.

Connections for the feed wires that carry current from the collector are made inside an electrical box mounted on the back of the collector panel.

An EPDM rubber boot seals off the opening where the PVC conduit carrying the feed wires penetrates the roof.

1st
proofs for reference only

41940 - Advanced Home Wiring_064-128.indd 96
Pro-Vision C M Y K Text
10/10/14 12:05 AM
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
Page: 96
Pro-Vis

How to Wire a DC Lighting Circuit

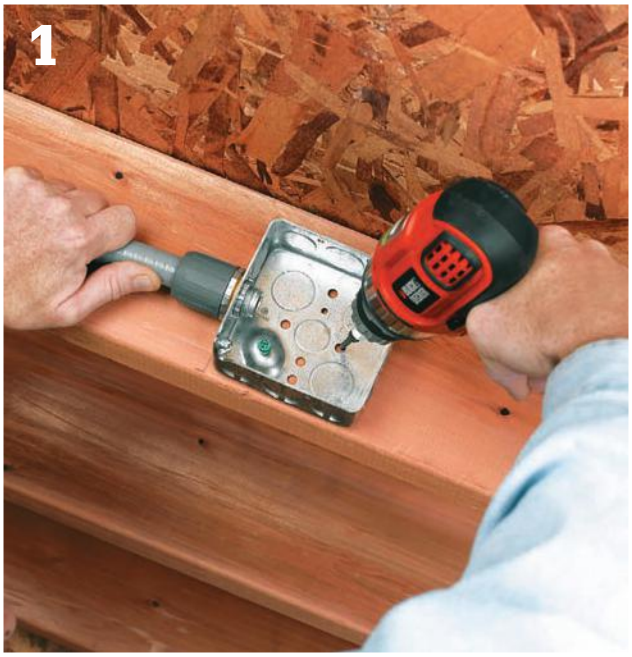

Mount a junction box inside the building where the conduit and wiring enter from the power source. Secure the box to the conduit with appropriate connectors. Run two #14 awg wires through the conduit and connect them to the positive and negative terminals on the panel (see previous page).

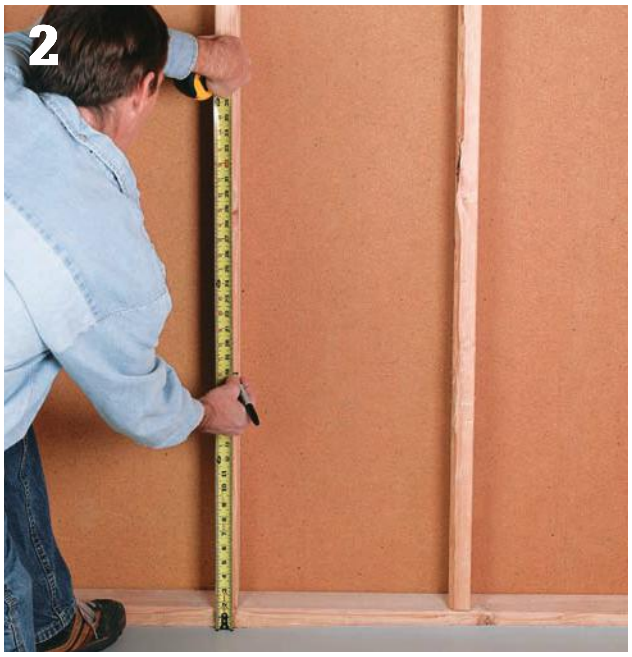

Plan the system layout. Determine the placement of the battery, and then decide where you will position the charge controller and DC disconnect. The battery should be placed at least 18" off the floor, in a well-ventilated area where it won't be agitated by everyday activity. Mark locations directly on the wall.

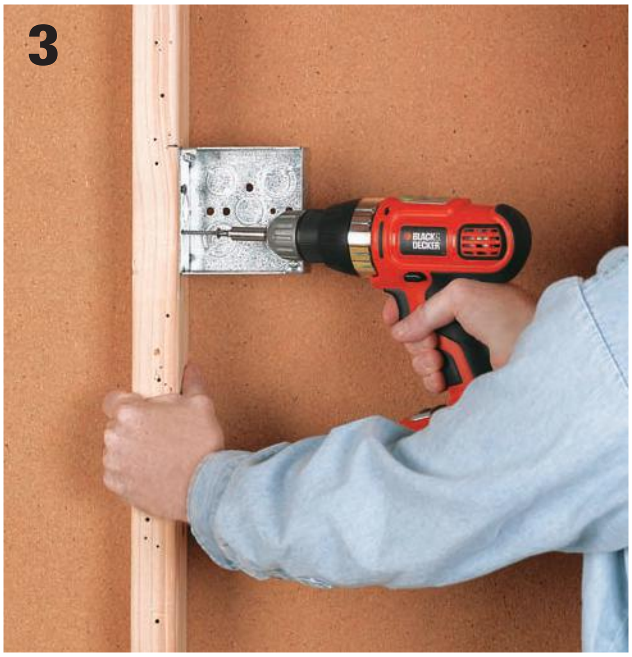

Attach a junction box for enclosing the DC disconnect, which is a heavy-duty switch, to a wall stud near the battery and charge controller location. Use a metal single-gang box with mounting flanges.

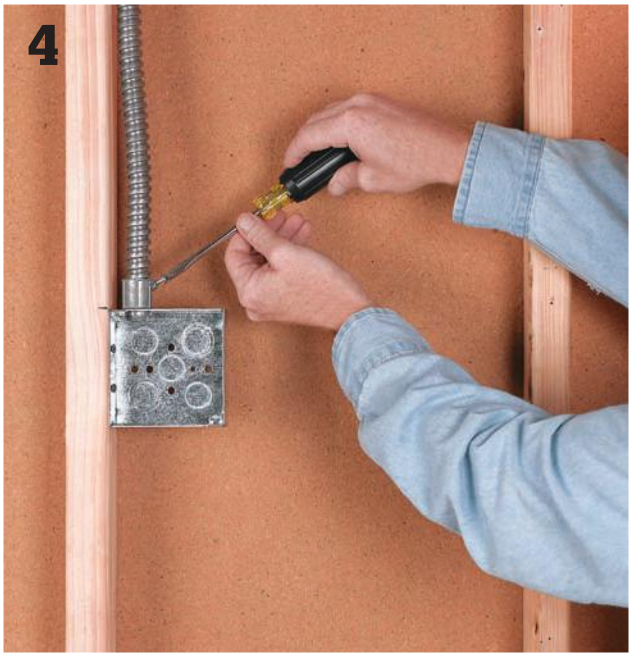

Run flexible metal conduit from the entry point at the power source to the junction box for the DC disconnect box. Use hangers rated for flexible conduit.

(continued)

1st
proofs for reference only

Edition
age: 96
41940 - Advanced Home Wiring_064-128.indd 97
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 12:05 AM
Page: 97

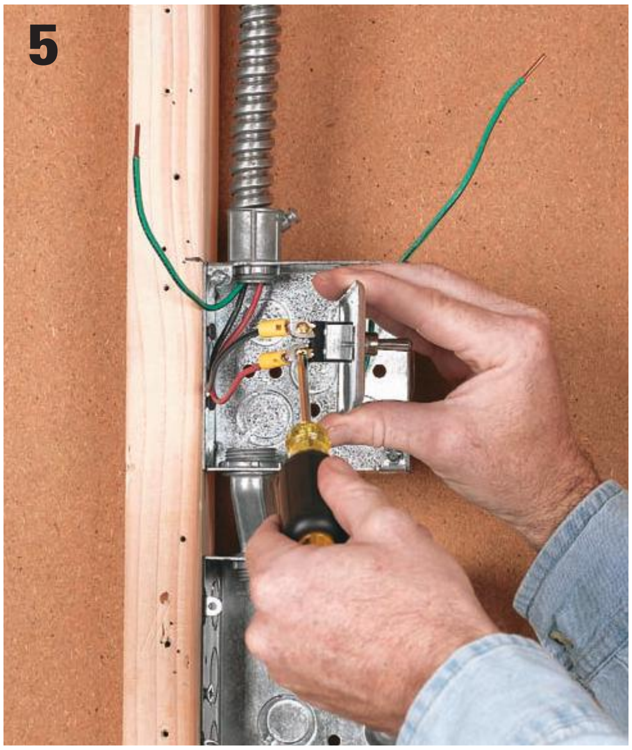

Attach the DC disconnect switch to the wire leads from the power source.

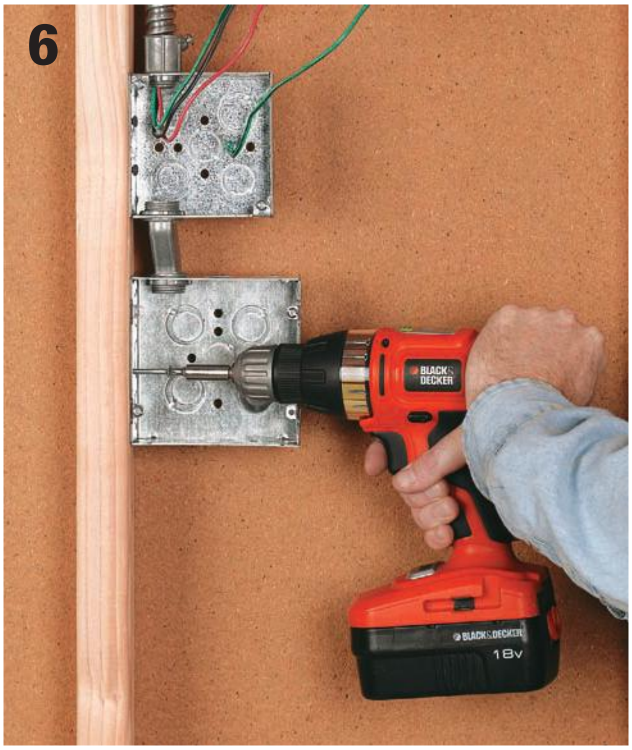

Attach a double gang metal junction box to the building's frame beneath the DC disconnect box to enclose the charge controller.

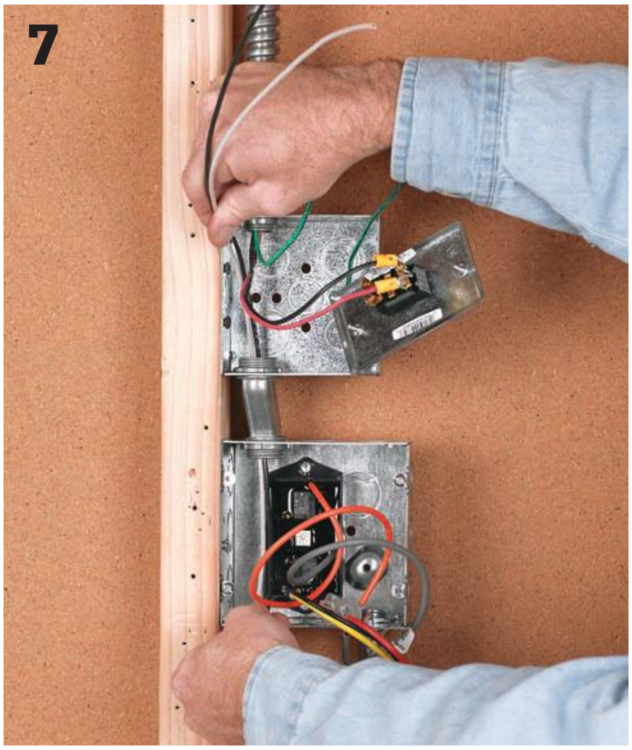

Install the charge controller inside the box. Run flexible conduit with connectors and conductors from the disconnect box and to the charge controller box.

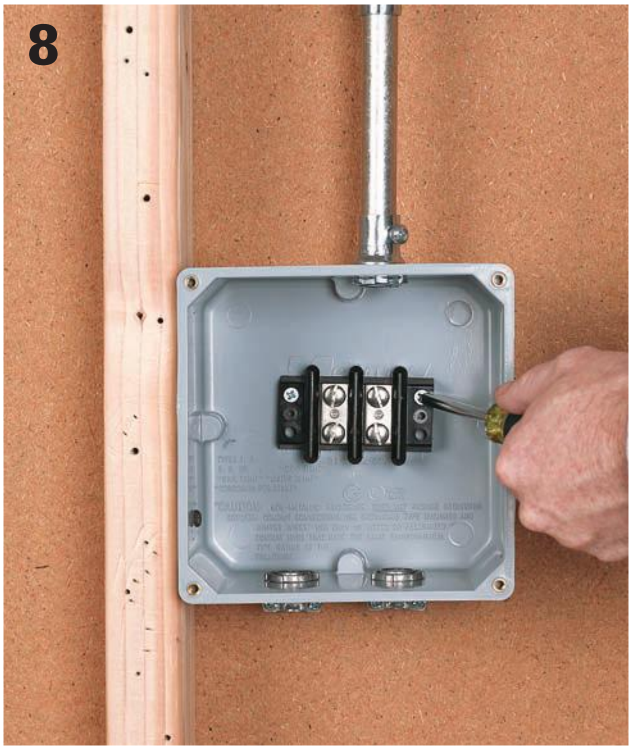

Mount a PVC junction box for the battery controller about 2 ft. above the battery location, and install two insulated terminal bars within the box.

1st
proofs for reference only
41940 - Advanced Home Wiring_064-128.indd 98
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 12:05 AM
Page: 98
Pro-Vis

9

Build a support shelf for the battery using 2 × 4s. The shelf should be at least 18" above ground. Set the battery on the shelf in a sturdy plastic case.

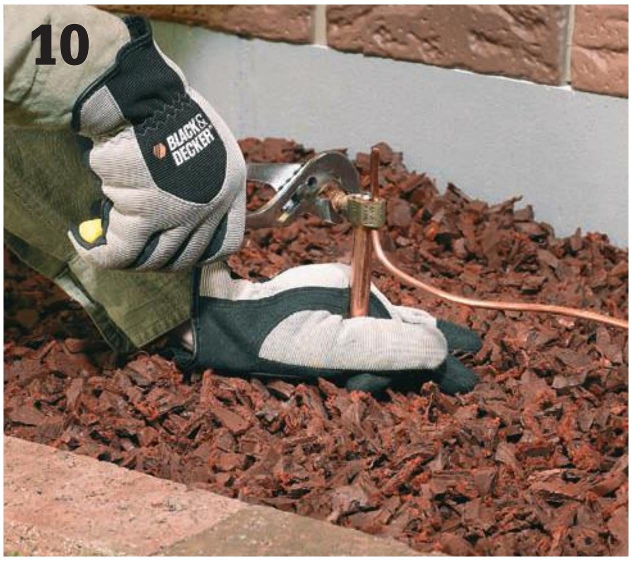
10

Set up grounding protection. Pound an 8-ft. long, ½"-dia. ground rod into the ground outside the building, about 1 ft. from the wall on the opposite side of the charge controller. Leave about 2" of the rod sticking out of the ground. Attach a ground rod clamp to the top of the rod. Drill a 5⁄16" hole through the garage wall (underneath a shake or siding piece) and run the #6-gauge THWN wire to the ground rod. This ground will facilitate lightning protection. See pages 58 to 63 for more information on grounding the system.

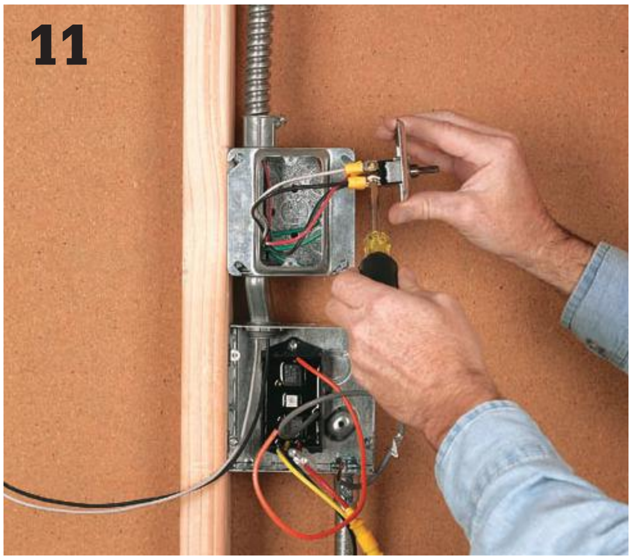
11

Wire the DC disconnect. Attach the two #14-gauge wires to the two terminals labeled “line” on the top of the DC disconnect switch.

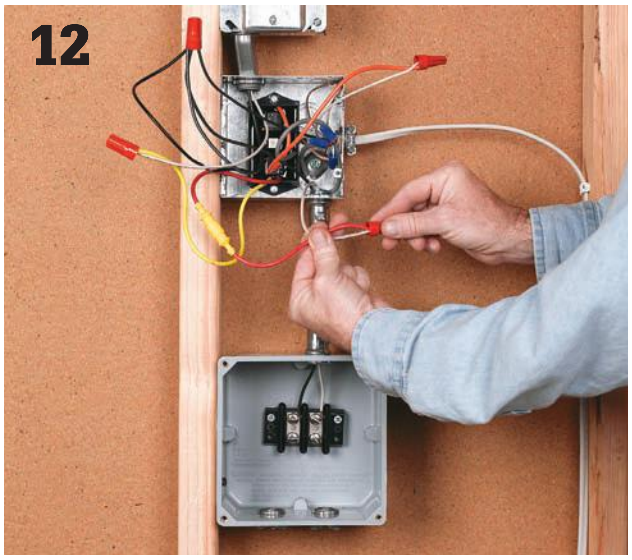
12

Wire the charge controller. Route two more #14-gauge wires from the bottom of the DC disconnect terminals into the 4" × 11⁄16" junction box and connect to the “Solar Panel In” terminals on the charge controller. The black wire should connect to the negative terminal in the PVC box and the red to the positive lead on the charge controller. Finish wiring of the charge controller according to the line diagram provided with the type of controller purchased. Generally the load wires connect to the orange lead, and the red wire gets tied to the battery through a fuse.

(continued)

1st proofs for reference only

Q.C.
PASSED
0836

OPTION: Attach a motion sensor. Some charge controllers come equipped with a motion sensor to maximize the efficiency of your lighting system—these are especially effective when used with security lighting. The motion sensor is typically mounted to a bell box outside and wired directly to the charge controller with an 18-gauge × 3-conductor insulated cable. A system like this can support up to three motion sensors. Follow the manufacturer's directions for installing and wiring the motion sensor.

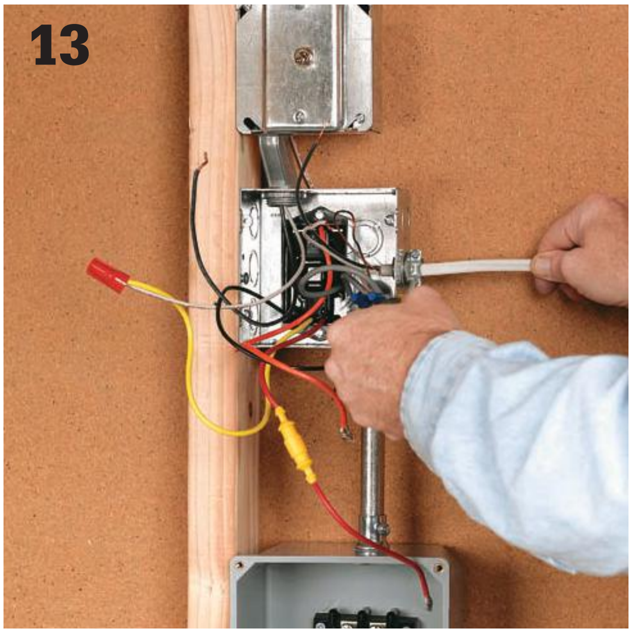

13

Run wiring to the loads (exterior DC lighting fixtures in this case) from the charge controller. DC light fixtures (12-volt) with LED bulbs can be purchased at marine and RV stores if you can't find them in your home center or electrical supply store.

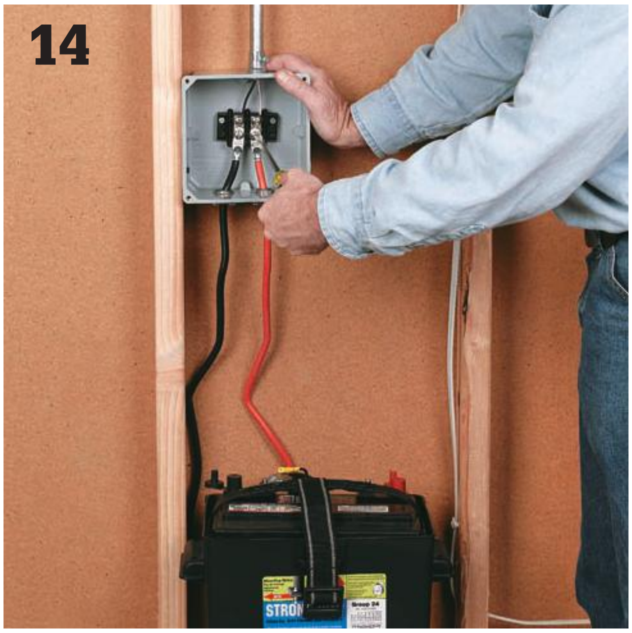

14

Install the battery. Here, a deep-cell 12-volt marine battery is used. First, cut and strip each of the two battery cables at one end and install into the battery control junction box through cord cap connectors. Terminate these wires on two separate, firmly mounted insulated terminal blocks.

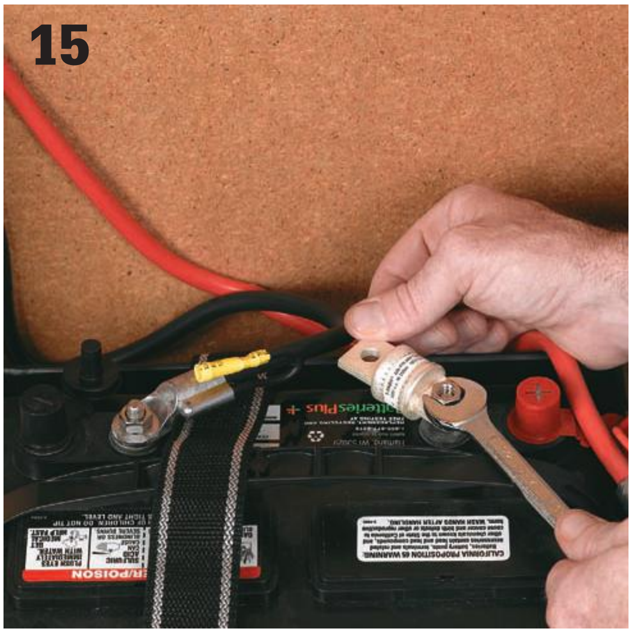

15

Install the catastrophe fuse onto the positive terminal using nuts and bolts provided with the battery cables. Connect the battery cables to the battery while paying close attention to the polarity (red to positive and black to negative). Make sure all connections have been made and double checked.

Troubleshooting & Repairs

Running new circuits and hooking up new fixtures are fairly predictable projects when it comes to estimating time and expense. This is less true with repairing problems in your system and fixtures. In some cases, a repair is as simple as opening an electrical box, spotting a loose wire connector and remaking the connection. But there are also times when fixing a dead circuit or device is a highly frustrating proposition. Such cases are almost always caused by tricky diagnostic challenges. Wires are hidden behind walls and there very often are no visual clues to system breakdowns. So essentially, minimizing repair frustration boils down to learning to deploy logical, systematic diagnostics. Educated troubleshooting, you could say.

In this project you'll learn how to use the most important diagnostic tool in any electrician's toolkit: the multimeter. These handy devices come in a dizzying array of types and qualities, but for diagnostic purposes they are used to take readings for current (amperage), voltage and continuity (whether an electrical path is open or closed). Once you learn the basics of operating a multimeter, you can enlist it in a logical, deductive manner to track down the source of a wiring problem. Once located, correcting the problem is usually very simple.

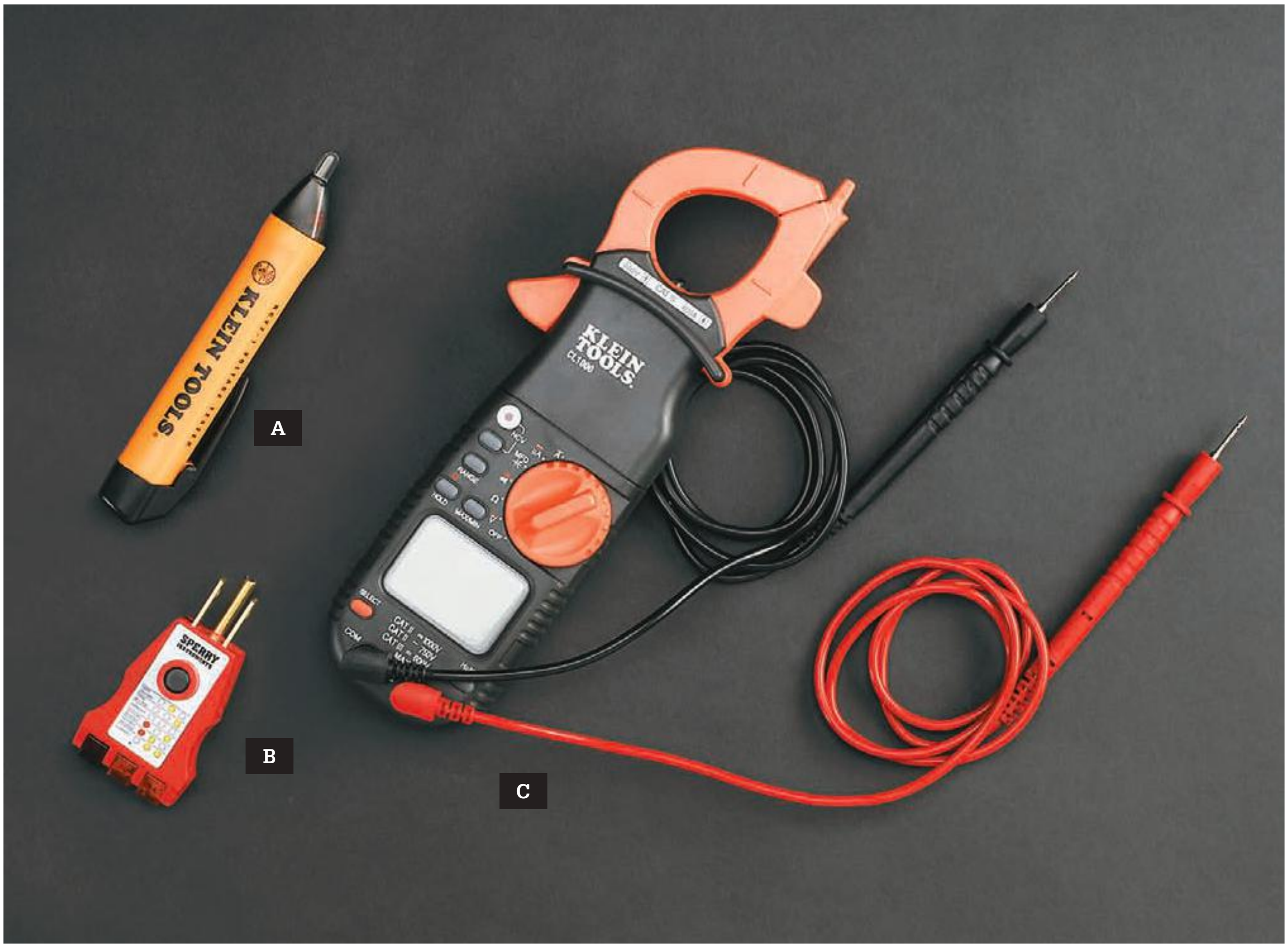

Diagnostic tools for home wiring use include: Touchless circuit tester (A) to safely check wires for current and confirm that circuits are dead; Plug-in tester (B) to check receptacles for correct polarity, grounding and circuit protection; Multimeter (C) to measure AC/DC voltage, AC/DC current, resistance, capacitance, frequency and duty cycle (model shown is an auto-ranging digital multimeter with clamp-on jaws that measure through sheathing and wire insulation).

Multimeters

Multimeters are nearly indispensible diagnostic tools for doing intermediate to advanced level electrical work (as well as automotive and electronics repair). They are used to measure voltage, current (amperage) and a few other conditions such as continuity, capacitance and frequency. For your home electrical system, by far the most used feature of a multimeter is testing voltage and current, although there are occasions where testing for resistance is needed. Among professional electricians, the most common and widely used multimeters have a clamp-on ammeter that measures current through the wire insulation so you don't have to disconnect the circuit and expose bare wire. Most clamp-on multimeters also are fitted with insertible probes with which you can measure voltage and continuity in the traditional way. An example of a clamp-on multimeter can be seen on the next page. Among homeowners, however, the most common multimeters these days are digital, auto-ranging tools that use probes or alligator clamps at the ends of wire leads for diagnostic work. Older multimeters that do not have autoranging capability must be pre-set to estimated calibration levels before use. Non-digital multimeters or ammeters usually have a dial gauge that gives readouts. These tools are somewhat more difficult to use and are less precise. Considering that digital, autoranging multimeters can be found for just a few dollars (the top of the line models cost over $100) there is really no good reason not to replace your old device with one that resembles the tools seen on these pages.

Time to Replace that Neon Tester ▸

Neon circuit testers are inexpensive and easy to use (if the light glows the circuit is hot), but they are less sensitive than multimeters and can be unsafe. In some cases, neon testers won't detect the presence of lower voltage in a circuit. This can lead you to believe that a circuit is shut off when it is not—a dangerous mistake. The small probes on a neon circuit tester also force you to get too close to live terminals and wires. For the most reliable readings, buy and learn to use a multimeter. At the very least, switch to a touchless tester like the one on page 107, Step 1.

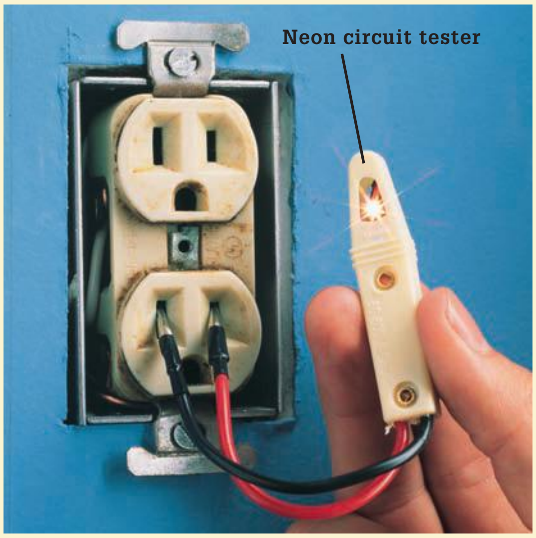
Neon circuit tester

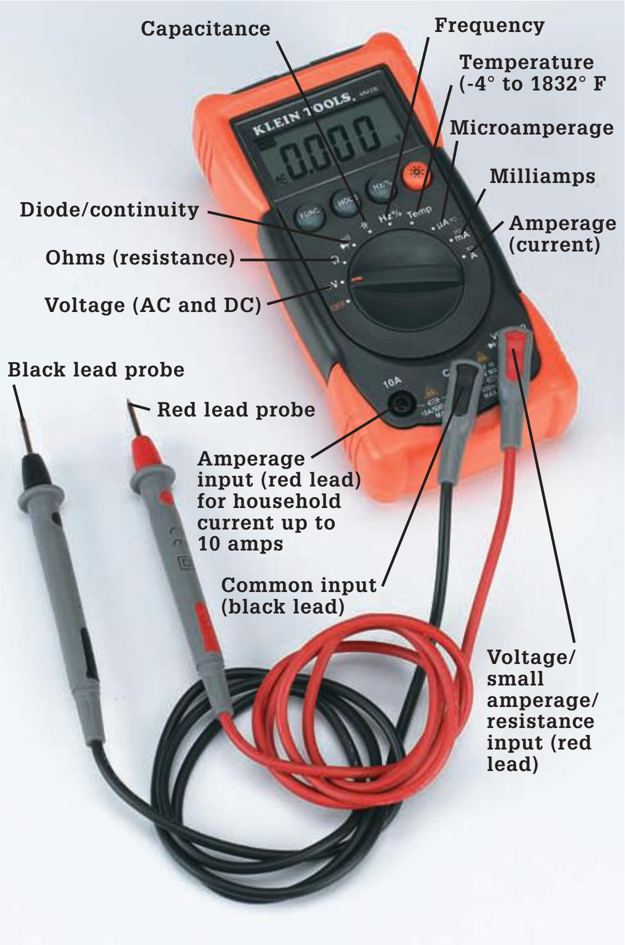
Capacitance
Frequency
Temperature (-4° to 1832° F
Microamperage
Milliamps
Diode/continuity
Amperage (current)
Ohms (resistance)
Voltage (AC and DC)
Black lead probe
Red lead probe
Amperage input (red lead) for household current up to 10 amps
Common input (black lead)
Voltage/ small amperage/ resistance input (red lead)
KLEIN TOOLS
0.000

A digital, autoranging multimeter must be adjusted to the proper setting for the reading you want to take. The probe leads also must be inserted into the correct inlet at the bottom of the tool. Inserting the red lead into the incorrect inlet can cause the tool to trip an internal fuse. Study your owner's manual carefully before using any tool.

1st
proofs for reference only

How to Measure Current

1

Create access to the wires you need to test. In most cases this requires that you remove the cover to an electrical service panel or an electrical box (inset).

2

KLEIN TOOLS
CL1000
NCV
RANGE
HOLD
MAX/MIN
OFF
SELECT
COM
CAT II ⎓1000V
CAT II ~750V
CAT III ⎓600V
MAX 600V
Hz/Duty

Set the multimeter to test for amperage (current is measured in amperes or amps). On some multimeters you need to select between amperage settings that are above or below 40 amps. Use the rated amperage of the circuit as a guide (amperage is printed on the circuit breaker switch).

3

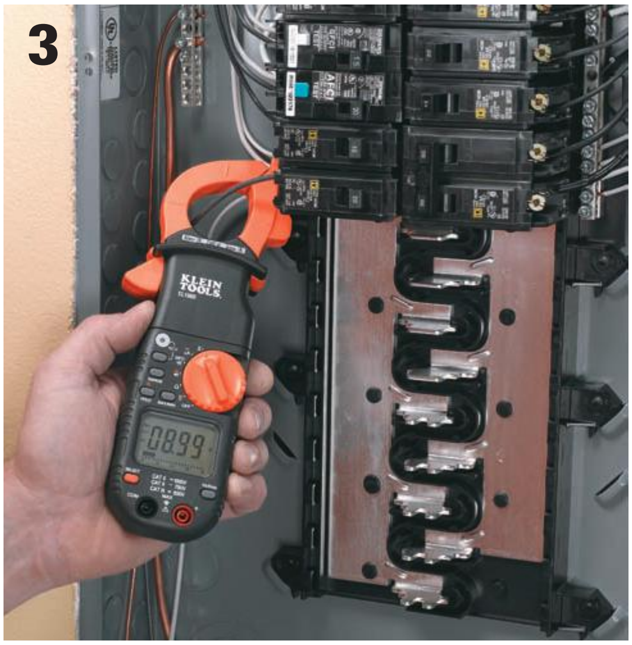

Clamp the jaws of a clamp-on multimeter onto the conductor or one of the conductors (if more than one) leading to a circuit breaker. If you are using a non-clamping multimeter, touch one probe to the screw terminal where the hot lead is attached to the breaker and touch the other probe to the metal panel box. The readout on your meter is the amount of current flowing in that circuit.

Taking Measurements at a Receptacle ▸

You may use a multimeter to measure for voltage at a wall receptacle. Regardless of whether the outlet is in service, if it is live you will get a voltage reading in the approximate range of the receptacle rating—here, 120 volts. To detect live current, measured in amps, the receptacle must be in use, with an appliance drawing from it. Taking an amperage reading in such an instance will only yield the amount of current being drawn, which is a factor of the appliance, not the circuit capacity.

1st
proofs for reference only
Edition
ge: 102
41940 - Advanced Home Wiring_064-128.indd 103
Pro-Vision C M Y K Text
10/10/14 12:05 AM
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
Page: 103

How to Measure Voltage

1
KLEIN TOOLS

To measure voltage using the multimeter, you will have to use the two probes provided with the multimeter and have access to a live terminal or slot as well as a grounded terminal or slot. If your meter has probe holders at the top, snap the probes into them. They are like extra hands.

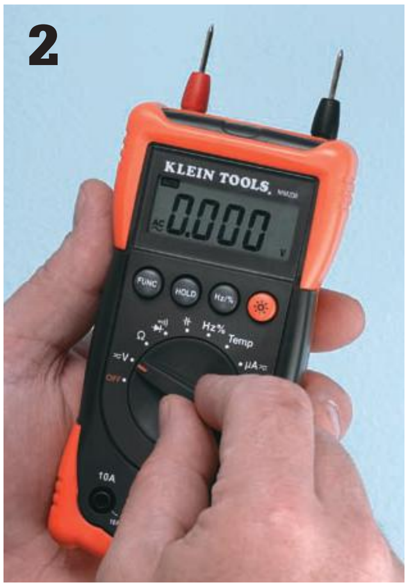

Turn the multimeter to the VAC setting to measure AC voltage that is found in your house. Set the multimeter to VDC if measuring DC voltage, such as in a car or a battery-fed device. On some multimeters, like the one above, you select "V" for voltage then change between AC and DC with the"FUNC" button.

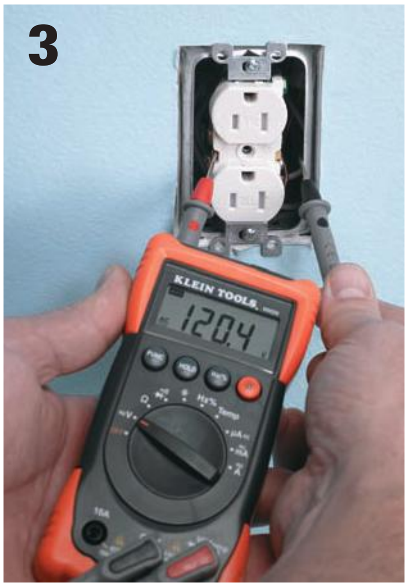

To measure the AC voltage, place one probe on a grounded surface, such as the metallic junction box or the bare ground wire. Place the other probe on the hot screw terminal or into the receptacle slot associated with the hot wire. The voltage readout should be in the range of 120 volts, plus or minus 5 volts (usually 120 volts in a residence in the US).

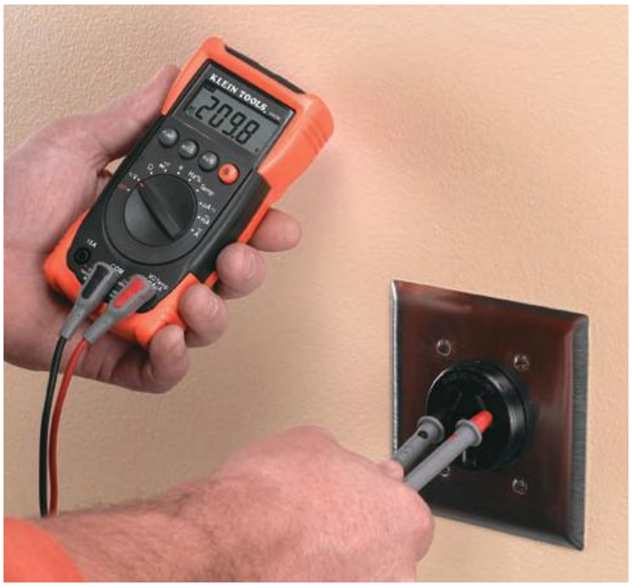

240 VOLTS. You can also measure voltage across the two hot leads to determine if you have 240 volts. This can be done at your range receptacle, dryer receptacle, or any other 240-volt receptacle. Place one probe in one of the small slots and the other probe in the other slot directly across from it. The voltage should read 240 volts, plus or minus 5 volts.

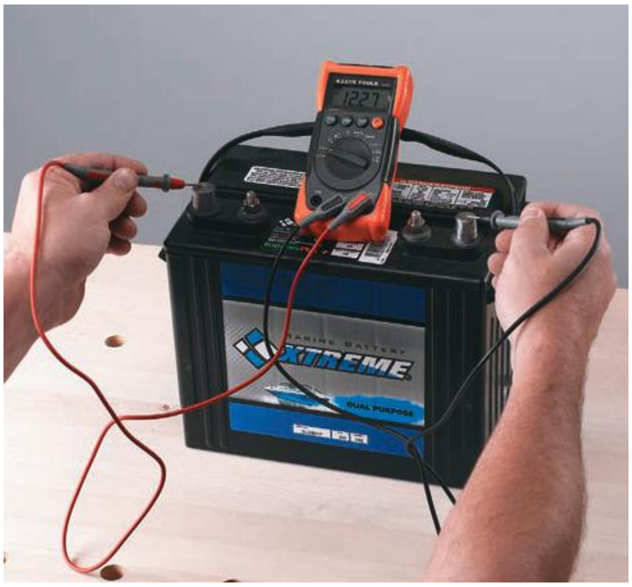

DC Voltage. When testing DC voltage, such as in a car battery, you can measure exactly the same way as for AC as long as the meter is set to the DC function. For more accurate results, test the voltage while the battery is in use.

1st
proofs for reference only

41940 - Advanced Home Wiring_064-128.indd 104
Pro-Vision C M Y K Text
10/10/14 12:05 AM
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
Page: 104
Pro-Vis

How to Test for Continuity

Continuity is a condition in a circuit where the conductors form an unbroken pathway through which current may flow. When measuring for continuity, always make sure there is no power present on the circuit you are testing or damage may occur to the meter. You can also measure the resistance in this mode as well.

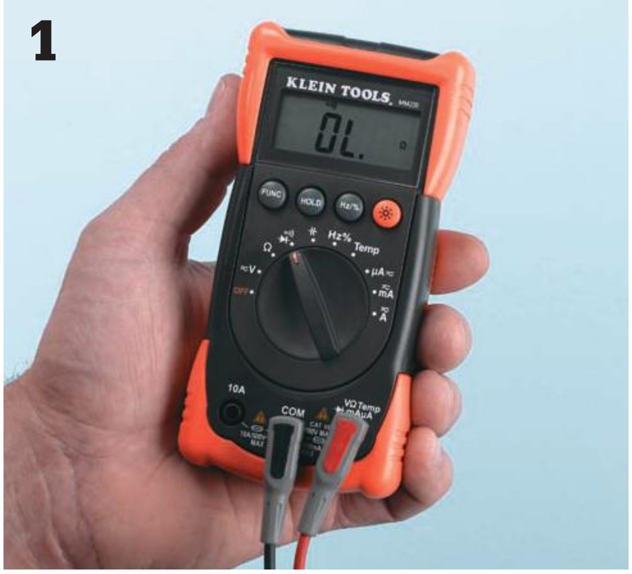

The setting for continuity is an "audible" or diode symbol display on the dial. Select this setting.

2
KLEIN TOOLS
0000

Verify that the continuity tester is functional by touching the two probes together. You should hear an alert sound and/or see a reading of zero ohms (Ohms is a value of the resistance to current flow).

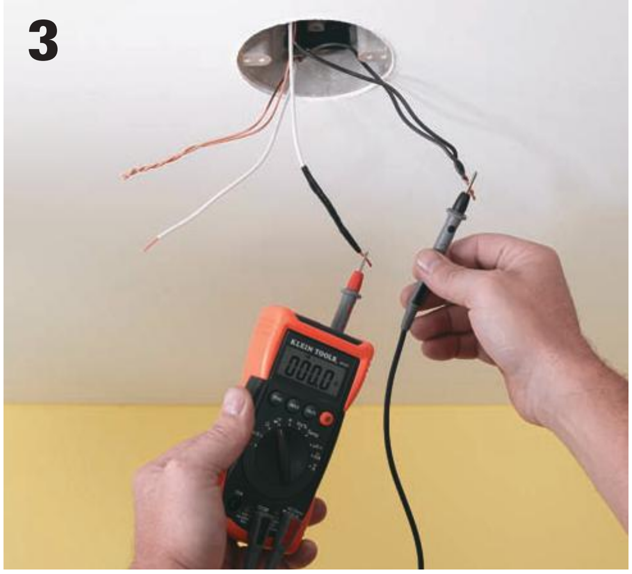

To test a circuit, touch one probe to one of the wires on a given circuit and the other to the second wire of the circuit. If you hear an audible sound or read a value of resistance other than zero, you have a complete or unbroken path for current to flow.

How to Test a 3-way Switch ▸

Remove the switch from the circuit and place one of the probes onto the common terminal and the other probe onto one of the other two terminals used for the traveler wires. If the meter indicates infinity ohms or there is no sound, flip the switch and if it is in working order the meter should read zero ohms or emit an audible sound. It should only work in one direction or the other, not both.

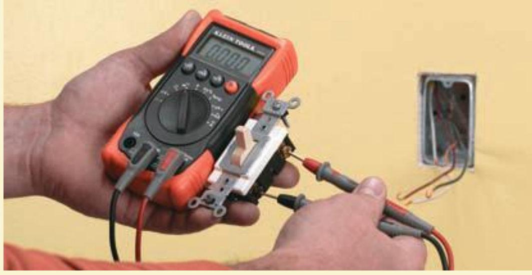

1st proofs for reference only

Edition
ge: 104
41940 - Advanced Home Wiring_064-128.indd 105
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 105

Troubleshooting an Open Neutral

An open neutral is an electrical problem where the circuit is broken on the return path wire or neutral (white wire). When this situation occurs anything plugged into or connected to this circuit can experience low or high voltage, which could damage voltage sensitive electronics such as a computer or flat screen TV. The lights will be dim or not work at all, depending upon where the problem lies within the circuit.

Possible Symptoms of an Open Neutral ▸

1. When a whole circuit does not work and the breaker associated with that circuit is operating normally.
2. When the neutral or white wire registers as a hot wire by using a non contact voltage tester when the circuit breaker is on. This most likely indicates a problem between the main service panel and the utility transformer. The condition should be readable on other receptacles as well.
3. If you register a voltage lower than 110 volts between the hot and neutral.
4. When the incandescent lights work, but are very dim.
5. When the fluorescent lights are barely lit and are flickering.
6. Discoloration of the wires or exposed copper turning green under the wirenut holding the neutral wires together.

1

KLEIN TOOLS
0899

2

3

KLEIN TOOLS
0924

6

1st proofs for reference only
41940 - Advanced Home Wiring_064-128.indd 106
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 106
Pro-Vis

How to Troubleshoot an Open Neutral

KLEIN TOOLS

1

Verify which lights and receptacles are on the circuit by turning the breaker off and by checking for power with a non-contact voltage tester. It is helpful to either draw a map of the house or place some tape on every affected opening.

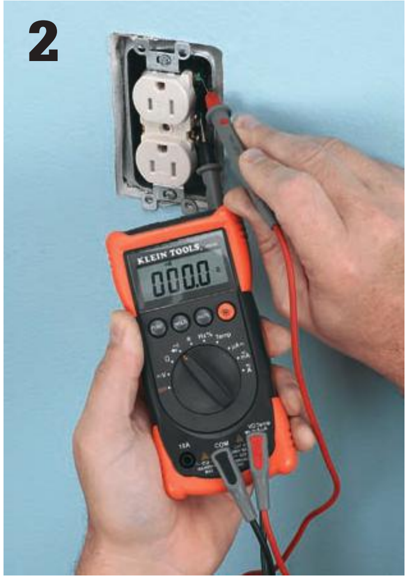

2

Start at the outlet nearest to the panel. With the breaker off and using a multimeter, check for continuity between the neutral (white wire) and the ground (bare or green wire). These two wires land at the same point electrically in your electrical panel. If there is an indication of continuity between these two wires, the neutral and ground connections are sound and you should proceed to the next outlet as you move away from the panel.

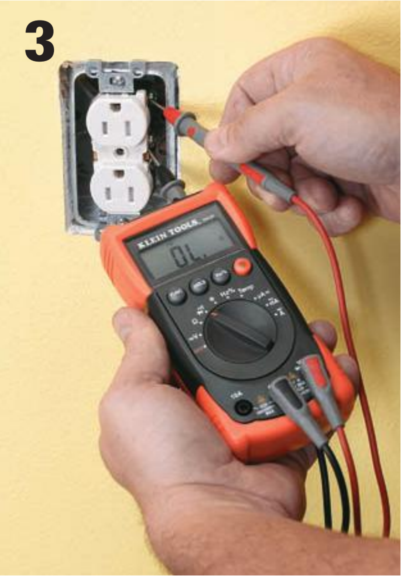

3

When you encounter a point at which you read infinity ohms or there is no continuity between the neutral and ground wires, the problem lies within the connections in that box or the box just upstream (toward) the panel from the one you are checking. Sometimes you will see evidence of arcing on the wire cap containing the connection which may include discoloration, or a blackish char near the copper.

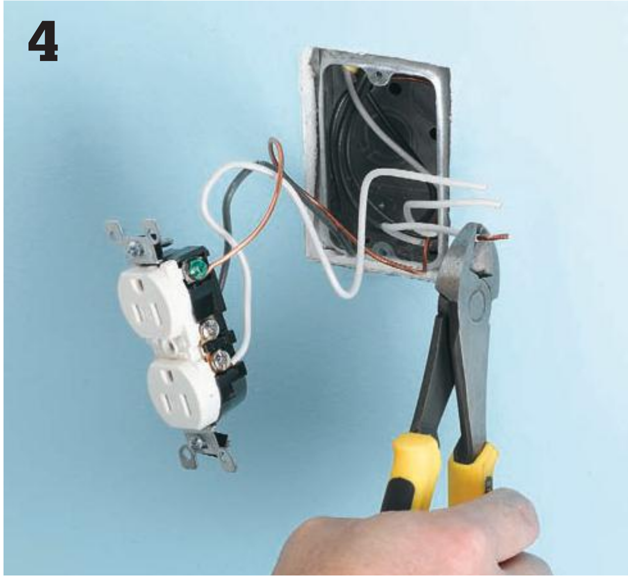

4

When you have found the problem connection, remove the wire cap and, if it is possible, cut the damaged portions of the wires off and restrip the wires to expose new copper. Line the wire ends up and twist on a new wire cap.

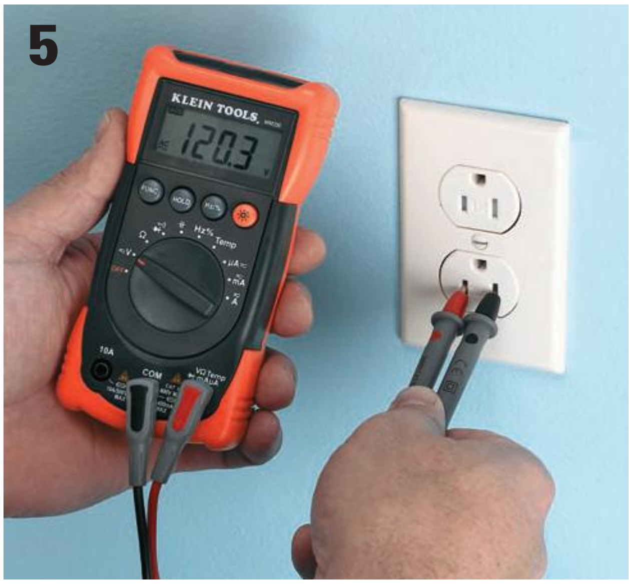

5

Turn the circuit breaker back on and verify the proper voltage is present at your receptacles by measuring with a multimeter.

1st
proofs for reference only
Edition
ge: 106
41940 - Advanced Home Wiring_064-128.indd 107
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 12:05 AM
Page: 107

1st
proofs for reference only

Troubleshooting a Short Circuit

Short circuits are a direct connection between the hot or power wire (black or red) and to either the neutral (white) or ground (bare) wire. This connection between the two will cause your circuit breaker or fuse to blow, which should interrupt power to the affected circuit.

Short circuits are a common problem and can usually be solved by taking the following steps. The idea behind electrical troubleshooting is to simplify the circuit by checking it at certain points, in order to narrow down the problem point by process of elimination. Generally, the problem is that there is a bare ground wire touching a hot terminal within a switch or an outlet box. There will usually be a black scorched mark or some sign of an electrical arc where the problem lies.

How to Troubleshoot a Short Circuit

1

Turn the power off at the affected breaker and verify with a non-contact voltage tester that there is no power present. Unplug everything from the receptacles and turn the lights off on the circuit that is affected.

2

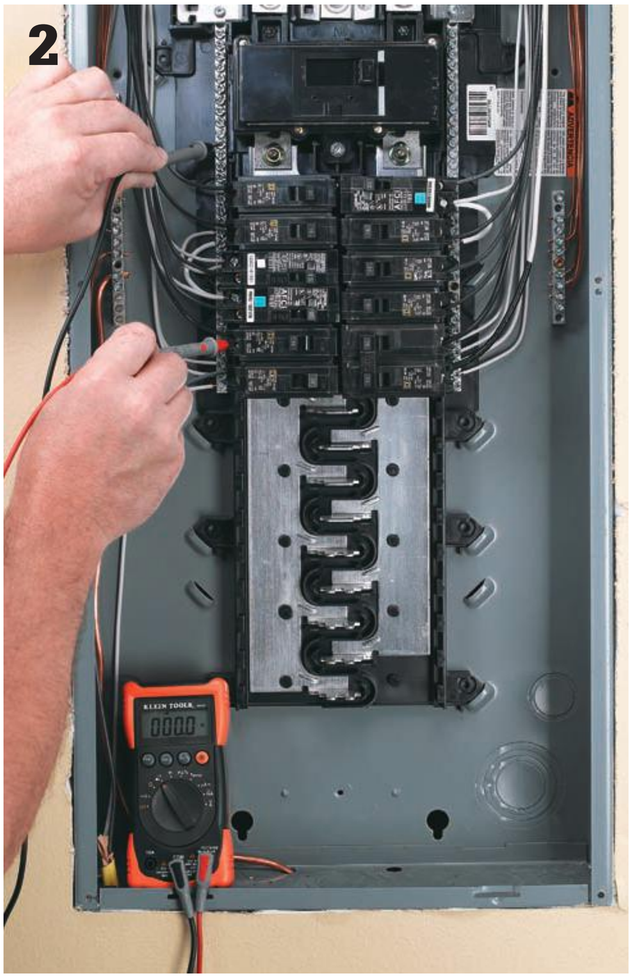

Using a multimeter set to the ohms or continuity setting, check the wires at the panel. Touch one of the probes to the hot or black wire and the other probe to the ground or bare wire. If the meter rings or indicates a low resistance value, you have a direct short to ground. If the meter does not ring or indicates a high resistance value the circuit is clear. If the meter does not ring, start by turning the switches on one-by-one and re-testing to verify the resistance value. If the meter indicates a low resistance value or a short circuit, the problem is downstream from the switch or within the light fixture itself.

41940 - Advanced Home Wiring_064-128.indd 108
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 108
Pro-Vis

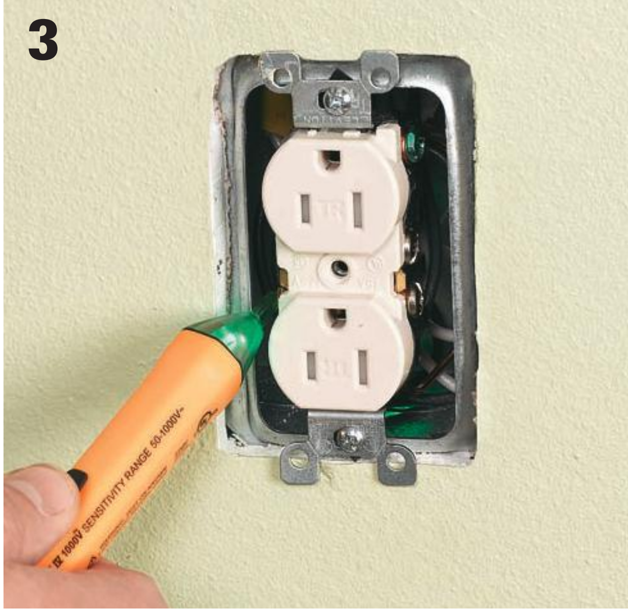

If the meter consistently rings or indicates a low resistive value, you will need to find the electrical box that contains the affected circuit. Choose a box that is convenient to open and preferably in the middle of the run, such as a receptacle. Verify there is no power present by touching all of the wires within the box with a non-contact voltage tester.

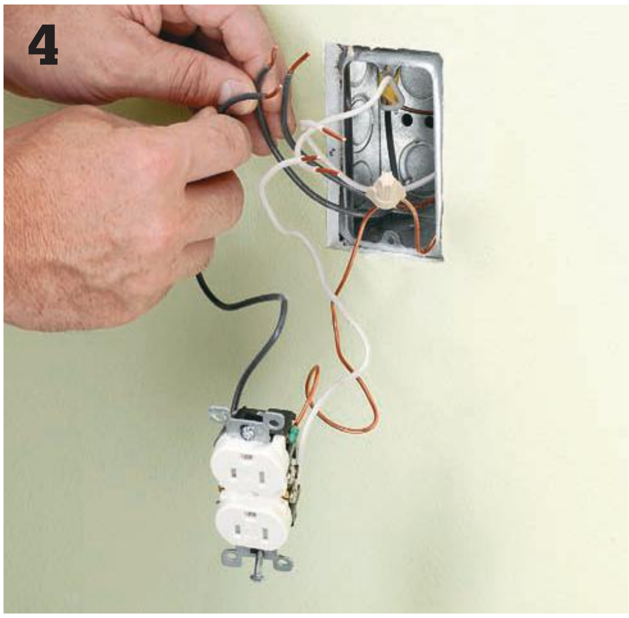

If the box you have chosen is in fact in the middle of the run, it will contain at least two cables. Remove the receptacle from the two cables and separate all of the wires.

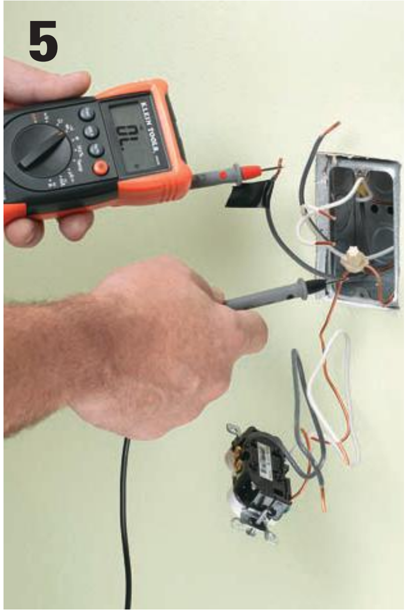

Check the resistance between the black and the ground on both sets of cables. One of the cables should cause the continuity alert to ring and the other should not. Mark the affected one with a piece of black tape and place wire caps over the exposed ends of the black wires.

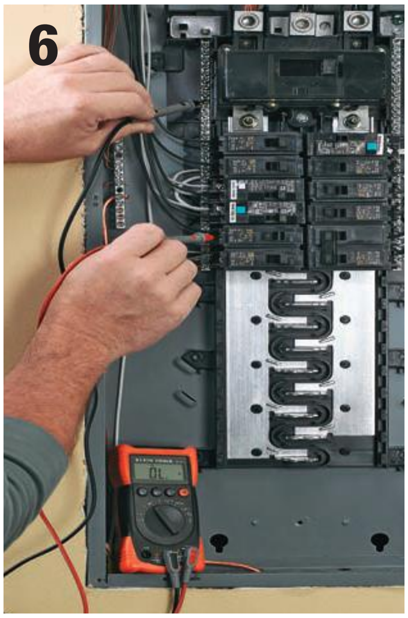

Check the wires at the panel to see if the short has cleared. If the short is clear, the problem lies down stream from the opened box and it is now safe to turn the breaker back on to help eliminate further problem points. If the short is still present, the problem lies between the opened electrical box and the panel.

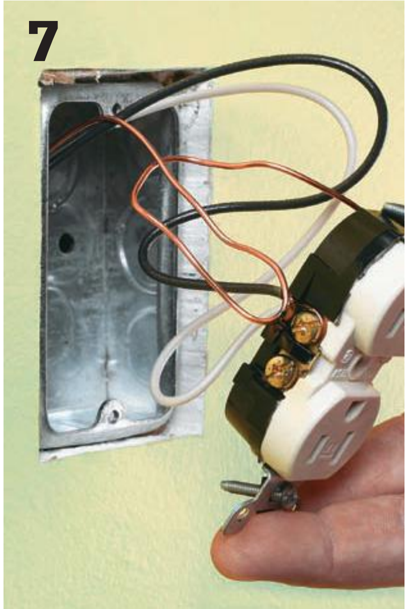

Choose another box in the middle of the affected circuit, there by narrowing down the possible problem areas until the short circuit can be positively identified and corrected. When you have discovered the short circuit, verify the wires are still in good shape and repair the connection.

1st
proofs for reference only
Edition
ge: 108
41940 - Advanced Home Wiring_064-128.indd 109
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 109

Types of Wall Switches

Wall switches are available in three general types. To re-connect or replace a switch, it is important to identify its type.

Single-pole switches are used to control a set of lights from one location. Three-way switches are used to control a set of lights from two different locations and are always installed in pairs. Four-way switches are used in combination with a pair of three-way switches to control a set of lights from three or more locations.

Identify switch types by counting the screw terminals. Single-pole switches have two screw terminals, three-way switches have three screw terminals, and four-way switches have four. Most switches include a grounding screw terminal, which is identified by its green color.

When replacing a switch, choose a new switch that has the same number of screw terminals as the old one. The location of the screws on the switch body varies depending on the manufacturer, but these differences will not affect the switch operation.

Whenever possible, connect switches using the screw terminals rather than push-in fittings. Some specialty switches (pages 118 to 119) have wire leads instead of screw terminals. They are connected to circuit wires with wire connectors.

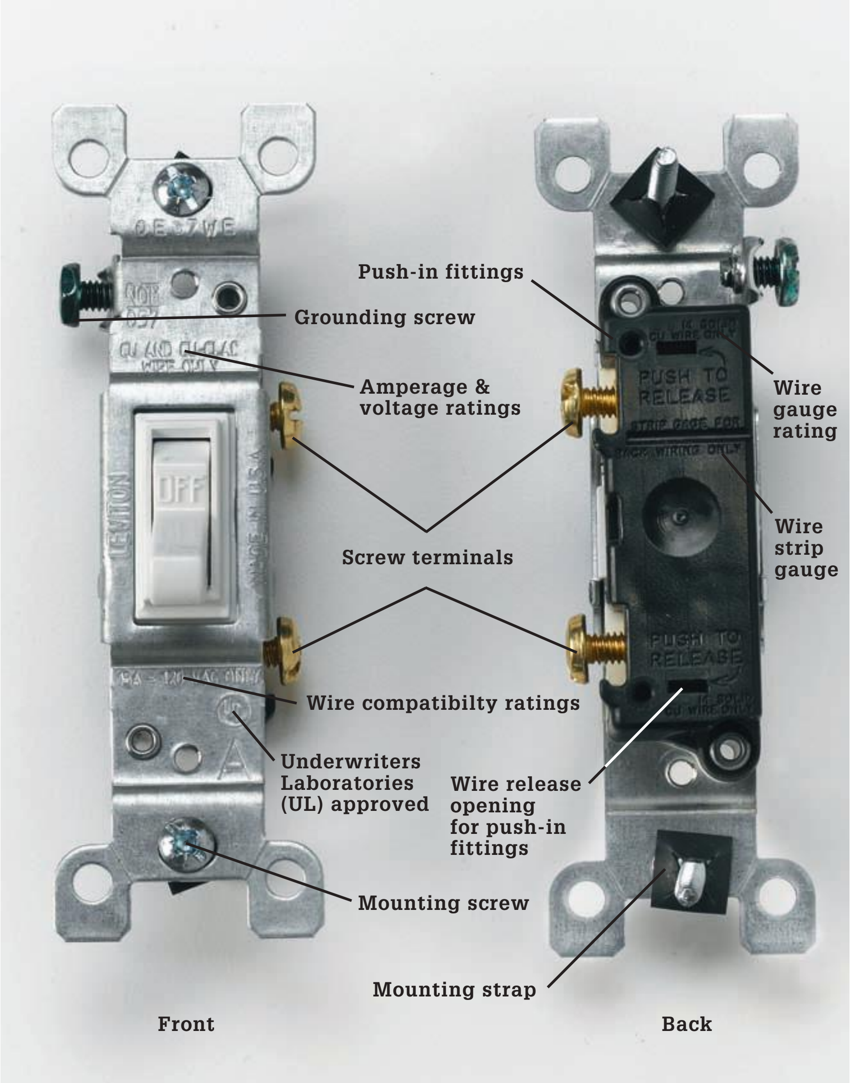
Push-in fittings
Grounding screw
Amperage & voltage ratings
Wire gauge rating
Screw terminals
Wire strip gauge
Wire compatibilty ratings
Underwriters Laboratories (UL) approved
Wire release opening for push-in fittings
Mounting screw
Mounting strap
Front
Back

A wall switch is connected to circuit wires with screw terminals or with push-in fittings on the back of the switch. A switch may have a stamped strip gauge that indicates how much insulation must be stripped from the circuit wires to make the connections.

The switch body is attached to a metal mounting strap that allows it to be mounted in an electrical box. Several rating stamps are found on the strap and on the back of the switch. The abbreviation UL or UND. LAB. INC. LIST means that the switch meets the safety standards of the Underwriters Laboratories. Switches also are stamped with maximum voltage and amperage ratings. Standard wall switches are rated 15A or 125V. Voltage ratings of 110, 120, and 125 are considered to be identical for purposes of identification.

For standard wall switch installations, choose a switch that has a wire gauge rating of #12 or #14. For wire systems with solid-core copper wiring, use only switches marked COPPER, CU, or CO/ALR. For aluminum wiring, use only switches marked CO/ALR. Note that while CO/ALR switches and receptacles are approved by the National Electrical Code for use with aluminum wiring, the Consumer Products Safety Commission does not recommend using these. Switches and receptacles marked AL/CU can no longer be used with aluminum wiring, according to the National Electrical Code.

1st proofs for reference only
41940 - Advanced Home Wiring_064-128.indd 110
10/10/14 12:05 AM
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
Page: 110
Pro-Vision C M Y K Text

Single-Pole Wall Switches

A single-pole switch is the most common type of wall switch. It has ON-OFF markings on the switch lever and is used to control a set of lights, an appliance, or a receptacle from a single location. A single-pole switch has two screw terminals and a grounding screw. When installing a single-pole switch, check to make sure the ON marking shows when the switch lever is in the up position.

In a correctly wired single-pole switch, a hot circuit wire is attached to each screw terminal. However, the color and number of wires inside the switch box will vary, depending on the location of the switch along the electrical circuit.

If two cables enter the box, then the switch lies in the middle of the circuit. In this installation, both of the hot wires attached to the switch are black.

If only one cable enters the box, then the switch lies at the end of the circuit. In this installation (sometimes called a switch loop), one of the hot wires is black, but the other hot wire usually is white. A white hot wire should be coded with black tape or paint.

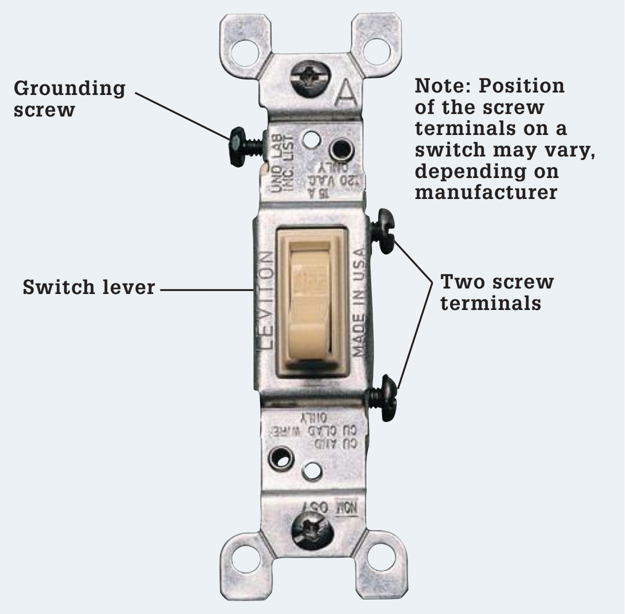
Grounding screw
Note: Position of the screw terminals on a switch may vary, depending on manufacturer
Switch lever
Two screw terminals
LEVITON
MADE IN USA

A single-pole switch is essentially an interruption in the black power supply wire that is opened or closed with the toggle. Single-pole switches are the simplest of all home wiring switches.

Typical Single-Pole Switch Installations

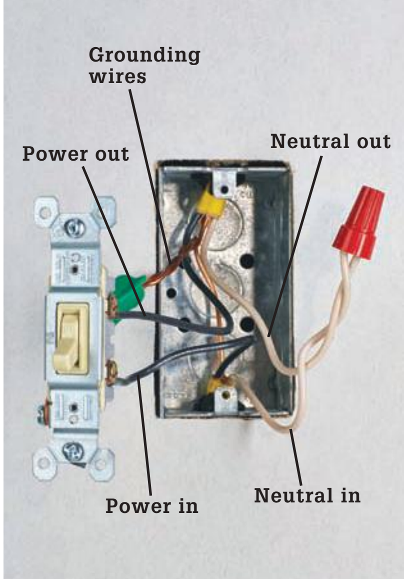
Grounding wires
Power out
Neutral out
Power in
Neutral in

Two cables enter the box when a switch is located in the middle of a circuit. Each cable has a white and a black insulated wire, plus a bare copper grounding wire. The black wires are hot and are connected to the screw terminals on the switch. The white wires are neutral and are joined together with a wire connector. Grounding wires are pigtailed to the switch.

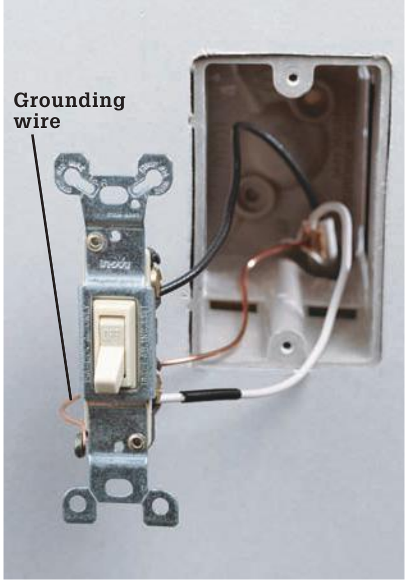
Grounding wire

Old method: One cable enters the box when a switch is located at the end of a circuit. In this installation, both of the insulated wires are hot. The white wire should be labeled with black tape or paint to identify it as a hot wire. The grounding wire is connected to the switch grounding screw.

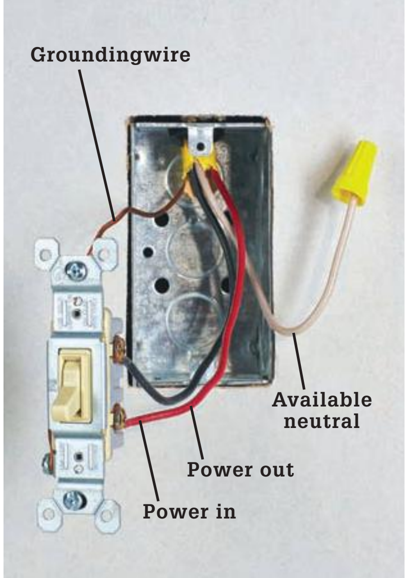
Groundingwire
Available neutral
Power out
Power in

Code change: In new switch wiring, the white wire should not supply current to the switched device and a separate neutral wire should be available in the switch box.

1st proofs for reference only
Edition
ge: 110
41940 - Advanced Home Wiring_064-128.indd 111
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 111

Three-Way Wall Switches

Three-way switches have three screw terminals and do not have ON-OFF markings. Three-way switches are always installed in pairs and are used to control a set of lights from two locations.

One of the screw terminals on a three-way switch is darker than the others. This screw is the common screw terminal. The position of the common screw terminal on the switch body may vary, depending on the manufacturer. Before disconnecting a three-way switch, always label the wire that is connected to the common screw terminal. It must be reconnected to the common screw terminal on the new switch.

The two lighter-colored screw terminals on a three-way switch are called the traveler screw terminals. The traveler terminals are interchangeable, so there is no need to label the wires attached to them.

Because three-way switches are installed in pairs, it sometimes is difficult to determine which of the switches is causing a problem. The switch that receives greater use is more likely to fail, but you may need to inspect both switches to find the source of the problem.

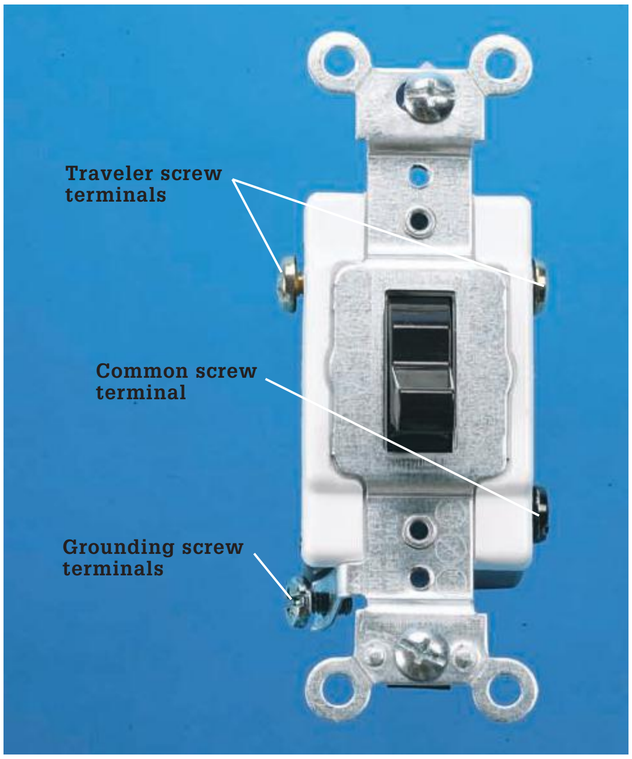

Typical Three-Way Switch Installation

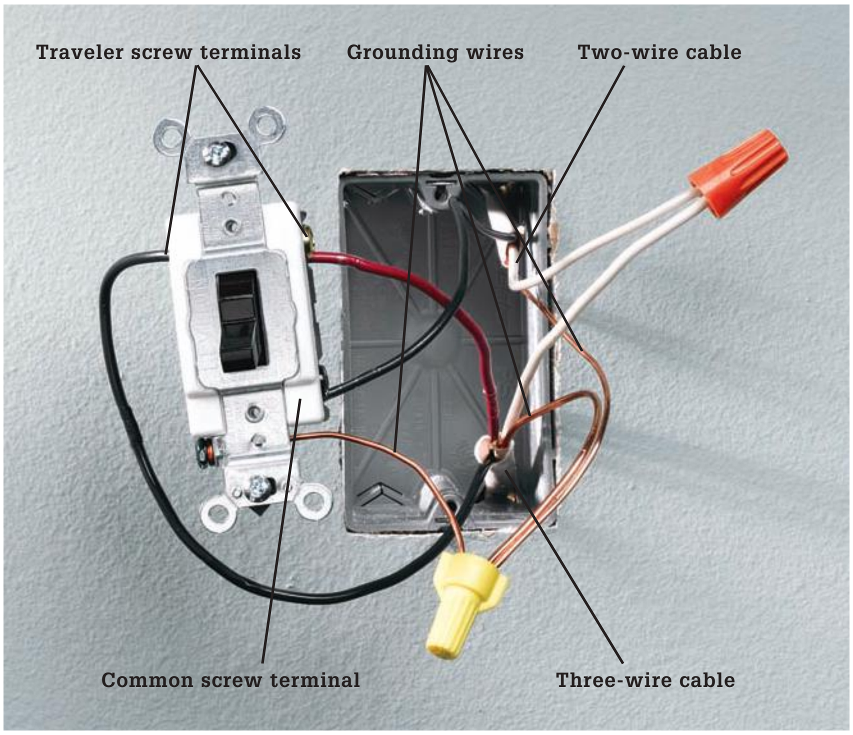

Two cables enter the box: one cable has two wires, plus a bare copper grounding wire; the other cable has three wires, plus a ground. The black wire from the two-wire cable is connected to the dark common screw terminal. The red and black wires from the three-wire cable are connected to the traveler screw terminals. The white neutral wires are joined together with a wire connector, and the grounding wires are pigtailed to the grounded metal box.

1st proofs for reference only
41940 - Advanced Home Wiring_064-128.indd 112
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 112

How to Replace a Three-Way Wall Switch

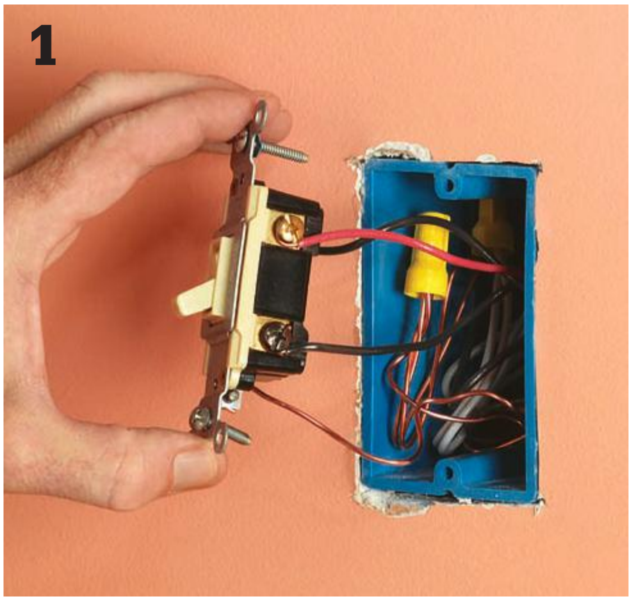

Turn off the power to the switch at the panel, and then remove the switch cover plate and mounting screws. Holding the mounting strap carefully, pull the switch from the box. Be careful not to touch the bare wires or screw terminals until they have been tested for power. *Note: If you are installing a new switch circuit, you must provide a neutral conductor at the switch.*

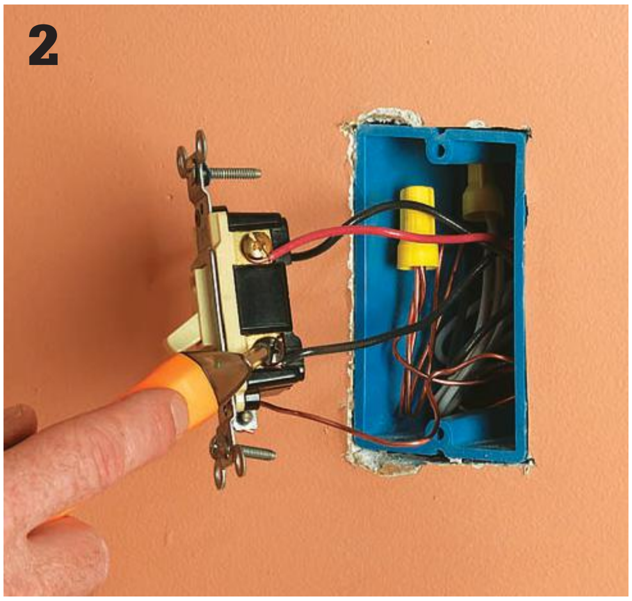

Test for power by touching one probe of the circuit tester to the grounded metal box or to the bare copper grounding wire and touching the other probe to each screw terminal. Tester should not glow. If it does, there is still power entering the box. Return to the panel, and turn off the correct circuit.

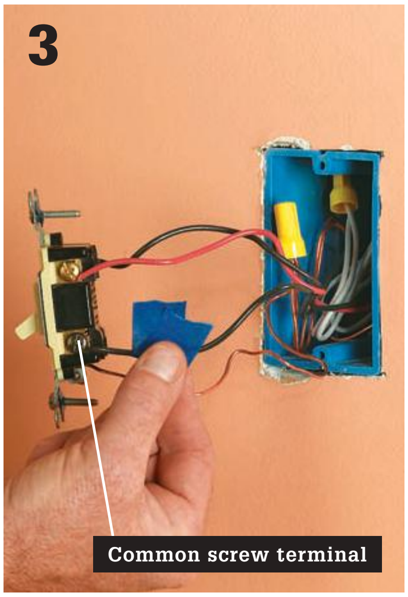

Locate the dark common screw terminal, and use masking tape to label the "common" wire attached to it. Disconnect wires and remove switch. Test the switch for continuity. If it tests faulty, buy a replacement. Inspect wires for nicks and scratches. If necessary, clip damaged wires and strip them.

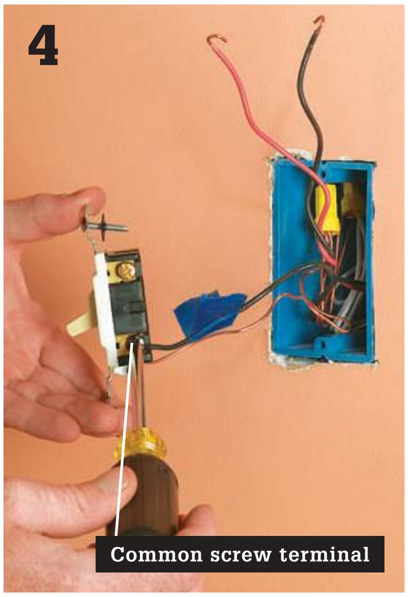

Connect the common wire to the dark common screw terminal on the switch. On most three-way switches, the common screw terminal is black. Or it may be labeled with the word COMMON stamped on the back of the switch. Reconnect the grounding screw, and connect it to the circuit grounding wires with a pigtail.

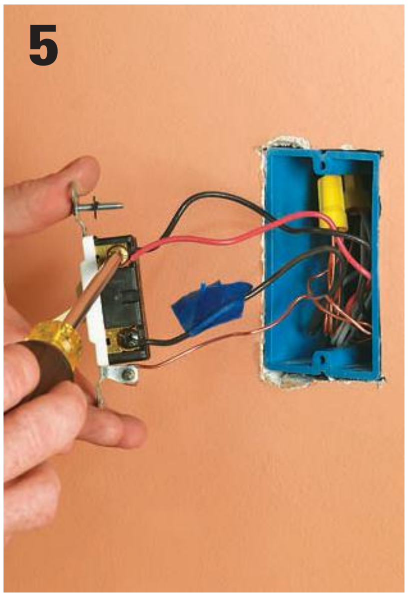

Connect the remaining two circuit wires to the screw terminals. These wires are interchangeable and can be connected to either screw terminal. Carefully tuck the wires into the box. Remount the switch, and attach the cover plate. Turn on the power at the panel.

proofs for reference only 1st

Edition
ge: 112

41940 - Advanced Home Wiring_064-128.indd 113
Pro-Vision C M Y K Text

Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 113

Four-Way Wall Switches

Four-way switches have four screw terminals and do not have ON-OFF markings. Four-way switches are always installed between a pair of three-way switches. This switch combination makes it possible to control a set of lights from three or more locations. Four-way switches are common in homes where large rooms contain multiple living areas, such as a kitchen opening into a dining room. Switch problems in a four-way installation can be caused by loose connections or worn parts in a four-way switch or in one of the three-way switches (facing page).

In a typical installation, there will be a pair of three-wire cables that enter the box for the four-way switch. With most switches, the white and red wires from one cable should be attached to the bottom or top pair of screw terminals, and the white and red wires from the other cable should be attached to the remaining pair of screw terminals. However, not all switches are configured the same way, and wiring configurations in the box may vary, so always study the wiring diagram that comes with the switch.

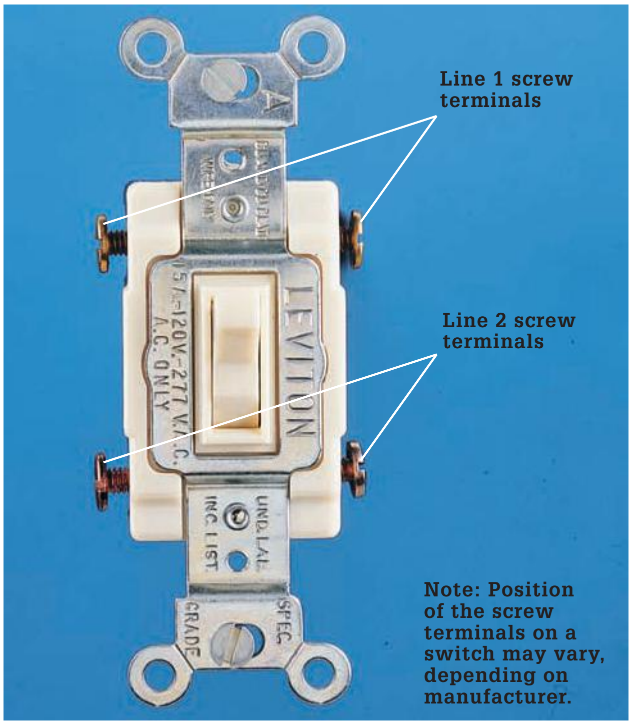
Line 1 screw terminals
Line 2 screw terminals
LEVITON
Note: Position of the screw terminals on a switch may vary, depending on manufacturer.

Typical Four-Way Switch Installation

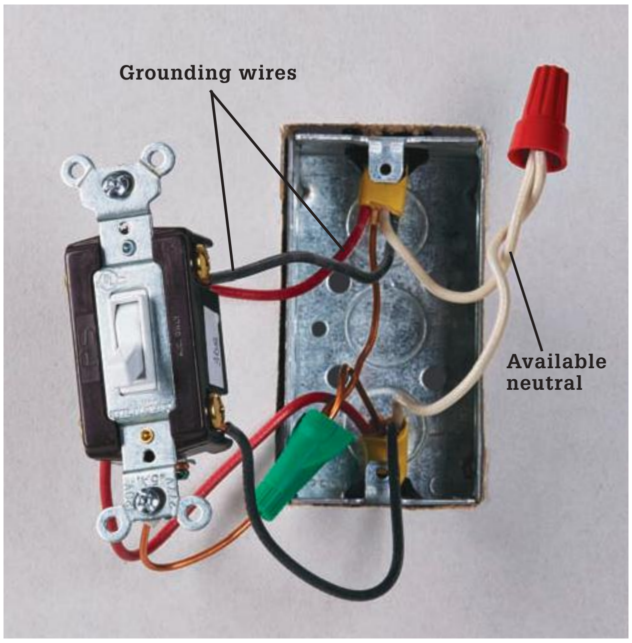

Four wires are connected to a four-way switch. The red and white wires from one cable are attached to the top pair of screw terminals, while the red and white wires from the other cable are attached to the bottom screw terminals. In new switch wiring, the white wire should not supply current to the switched device, and a separate neutral wire should be available in the switch box.

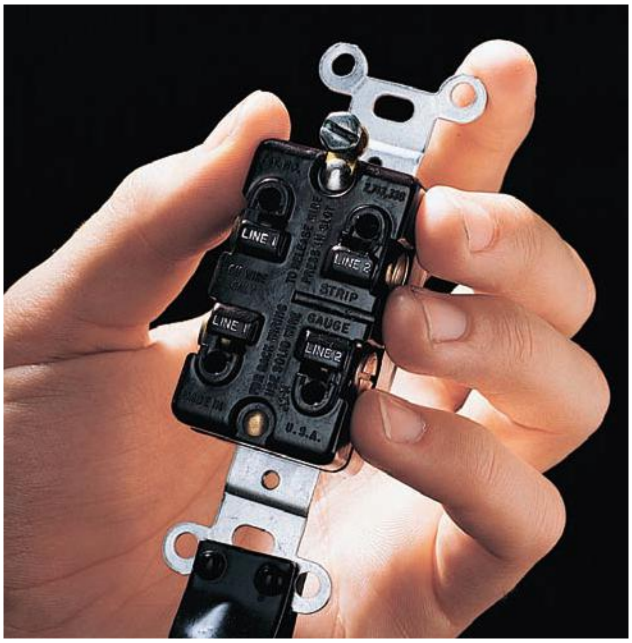

Switch variation: Some four-way switches have a wiring guide stamped on the back to help simplify installation. For the switch shown above, one pair of color-matched circuit wires will be connected to the screw terminals marked LINE 1, while the other pair of wires will be attached to the screw terminals marked LINE 2.

1st proofs for reference only
41940 - Advanced Home Wiring_064-128.indd 114
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 114
Pro-Vision C M Y K Text

How to Replace a Four-Way Wall Switch

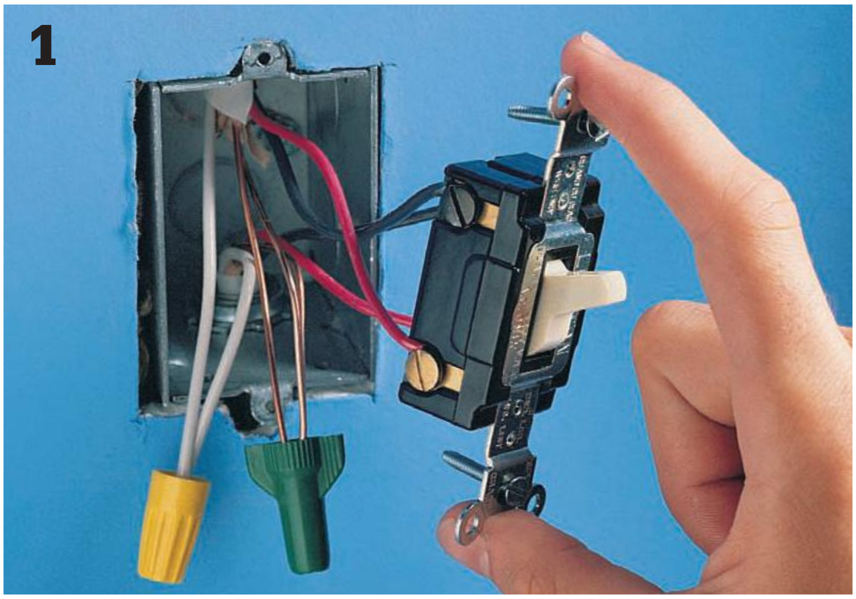

Turn off the power to the switch at the panel, and then remove the switch cover plate and mounting screws. Holding the mounting strap carefully, pull the switch from the box. Be careful not to touch any bare wires or screw terminals until they have been tested for power. Test for power by touching one probe of the neon circuit tester to the grounded metal box or bare copper grounding wire and touching the other probe to each of the screw terminals. The tester should not glow. If it does, there is still power entering the box. Return to the panel, and turn off the correct circuit.

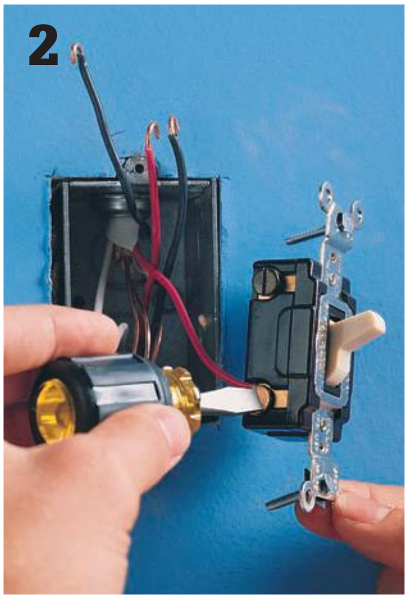

Disconnect the wires and inspect them for nicks and scratches. If necessary, clip damaged wires and strip them. Test the switch for continuity. Buy a replacement if the switch tests faulty.

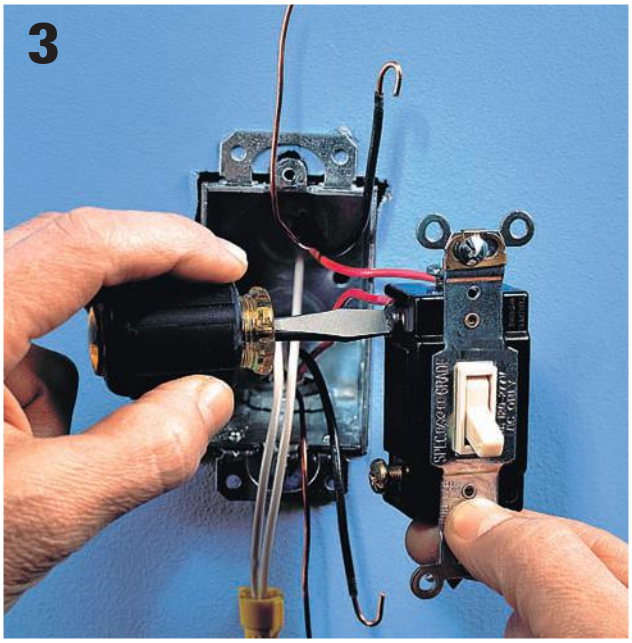

Connect two wires from one incoming cable to the top set of screw terminals.

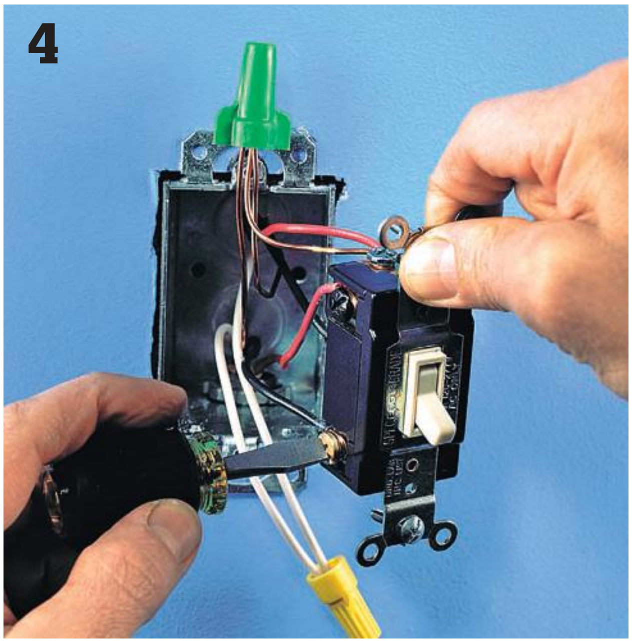

Attach remaining wires to the other set of screw terminals. Pigtail the grounding wires to the grounding screw. Carefully tuck the wires inside the switch box, and then remount the switch and cover plate. Turn on power at the panel.

1st proofs for reference only

Edition
ge: 114
41940 - Advanced Home Wiring_064-128.indd 115
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 12:05 AM
Page: 115

Double Switches

A double switch has two switch levers in a single housing. It is used to control two light fixtures or appliances from the same switch box.

In most installations, both halves of the switch are powered by the same circuit. In these single-circuit installations, three wires are connected to the double switch. One wire, called the feed wire (which is hot), supplies power to both halves of the switch. The other wires, called the switch leg, carry power out to the individual light fixtures or appliances.

In rare installations, each half of the switch is powered by a separate circuit. In these separate-circuit installations, four wires are connected to the switch, and the metal connecting tab joining two of the screw terminals is removed (see photo below).

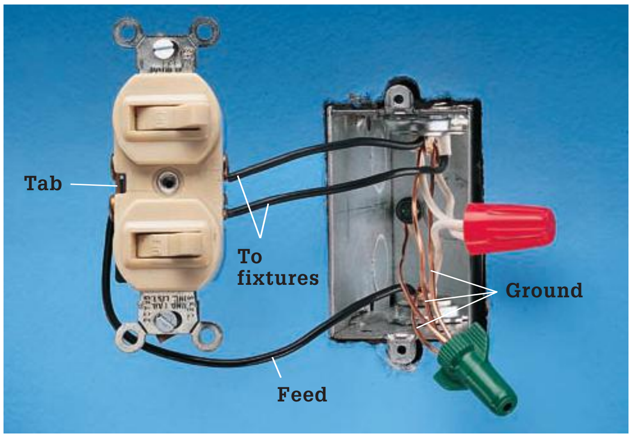

Single-circuit wiring: Three black wires are attached to the switch. The black feed wire bringing power into the box is connected to the side of the switch that has a connecting tab. The wires carrying power out to the light fixtures or appliances are connected to the side of the switch that does not have a connecting tab. The white neutral wires are connected together with a wire connector.

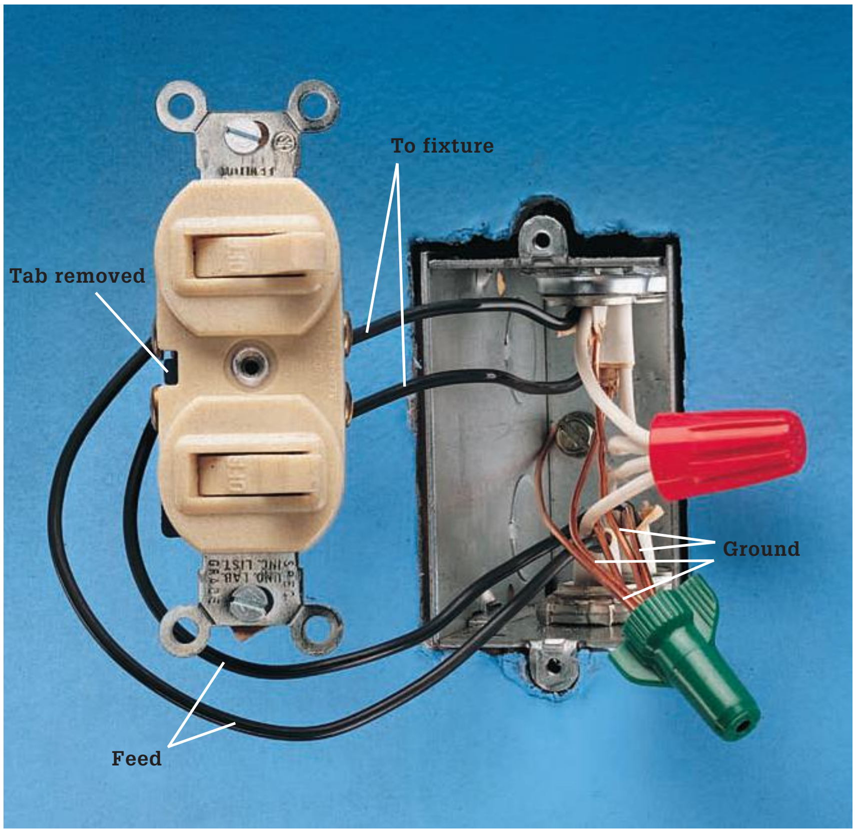

Separate-circuit wiring: Four black wires are attached to the switch. Feed wires from the power source are attached to the side of the switch that has a connecting tab, and the connecting tab is removed (photo, right). Wires carrying power from the switch to light fixtures or appliances are connected to the side of the switch that does not have a connecting tab. White neutral wires are connected together with a wire connector.

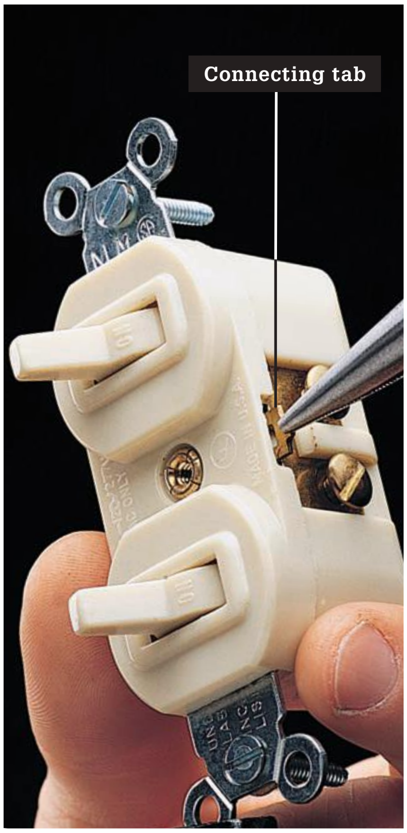

Remove the connecting tab on a double switch when wired in a separate-circuit installation. The tab can be removed with needlenose pliers or a screwdriver.

1st
proofs for reference only

41940 - Advanced Home Wiring_064-128.indd 116
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 116
Pro-Vis

Pilot-Light Switches

A pilot-light switch has a built-in bulb that glows when power flows through the switch to a light fixture or appliance. Pilot-light switches often are installed for convenience if a light fixture or appliance cannot be seen from the switch location. Basement lights, garage lights, and attic exhaust fans frequently are controlled by pilot-light switches.

A pilot-light switch requires a neutral wire connection. A switch box that contains a single two-wire cable has only hot wires and cannot be fitted with a pilot-light switch.

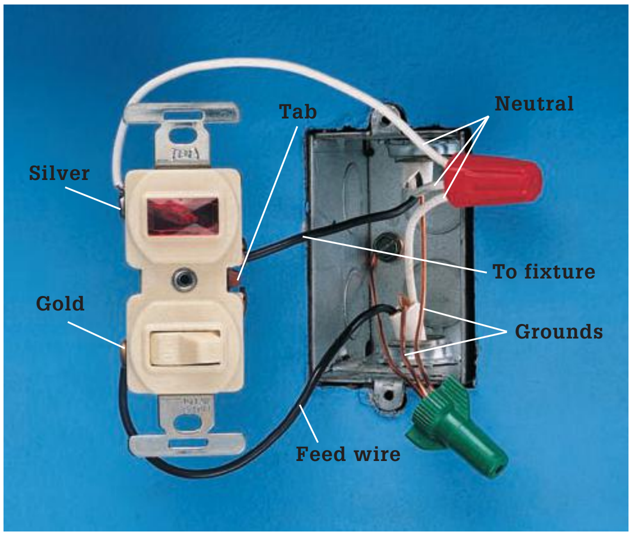

Pilot-light switch wiring: Three wires are connected to the switch. One black wire is the feed wire that brings power into the box. It is connected to the brass (gold) screw terminal on the side of the switch that does not have a connecting tab. The white neutral wires are pigtailed to the silver screw terminal. The black wire carrying power out to a light fixture or appliance is connected to the screw terminal on the side of the switch that has a connecting tab.

Switch/Receptacles

A switch/receptacle combines a grounded receptacle with a single-pole wall switch. In a room that does not have enough wall receptacles, electrical service can be improved by replacing a single-pole switch with a switch/receptacle.

A switch/receptacle requires a neutral wire connection. A switch box that contains a single two-wire cable has only hot wires and cannot be fitted with a switch/receptacle.

A switch/receptacle can be installed in one of two ways. In the most common installations, the receptacle is hot even when the switch is off (photo, right).

In rare installations, a switch/receptacle is wired so the receptacle is hot only when the switch is on. In this installation, the hot wires are reversed, so that the feed wire is attached to the brass screw terminal on the side of the switch that does not have a connecting tab.

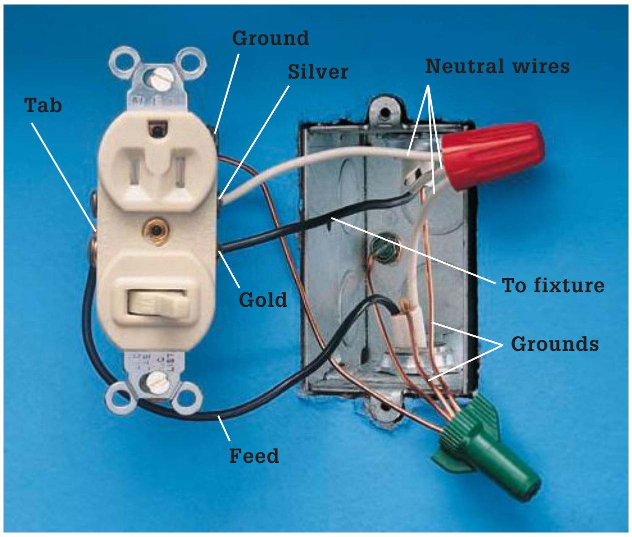

Switch/receptacle wiring: Three wires are connected to the switch/receptacle. One of the hot wires is the feed wire that brings power into the box. It is connected to the side of the switch that has a connecting tab. The other hot wire carries power out to the light fixture or appliance. It is connected to the brass screw terminal on the side that does not have a connecting tab. The white neutral wire is pigtailed to the silver screw terminal. The grounding wires must be pigtailed to the green grounding screw on the switch/receptacle and to the grounded metal box.

1st
proofs for reference only
Edition
ge: 116
41940 - Advanced Home Wiring_064-128.indd 117
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 12:05 AM
Page: 117

Specialty Switches

Your house may have several types of specialty switches. Dimmer switches (pages 96 to 97) are used frequently to control light intensity in dining and recreation areas. Timer switches and time-delay switches (below) are used to control light fixtures and exhaust fans automatically. Electronic switches provide added convenience and home security, and they are easy to install. Electronic switches are durable, and they rarely need replacement.

Most specialty switches have preattached wire leads instead of screw terminals and are connected to circuit wires with wire connectors. Some motor-driven timer switches require a neutral wire connection and cannot be installed in switch boxes that have only one cable with two hot wires. It is precisely due to the rise in popularity of "smart" switches that the NEC Code was changed in 2014 to require an available neutral wire in newly-installed switch boxes.

If a specialty switch is not operating correctly, you may be able to test it with a continuity tester. Timer switches and time-delay switches can be tested for continuity, but dimmer switches cannot be tested. With electronic switches, the manual switch can be tested for continuity, but the automatic features cannot be tested.

Timer Switches

Countdown timer switches can be set to turn lights or fans on and off automatically once each day. They are commonly used to control outdoor light fixtures.

Timer switches have three preattached wire leads. The black wire lead is connected to the hot feed wire that brings power into the box, and the red lead is connected to the wire carrying power out to the light fixture. The remaining wire lead is the neutral lead. It must be connected to any neutral circuit wires. A switch box that contains only one cable has no neutral wires, so it cannot be fitted with a timer switch.

After a power failure, the dial on a timer switch must be reset to the proper time.

Countdown timer switch. This rocker-type switch gives you the option to easily program the switch to shut off after a specified time: from 5 to 60 minutes. Garage lights or basement lights are good applications: anywhere you want the light to stay on long enough to allow you to exit, but not to stay on indefinitely. These switches often are used to control vent fans.

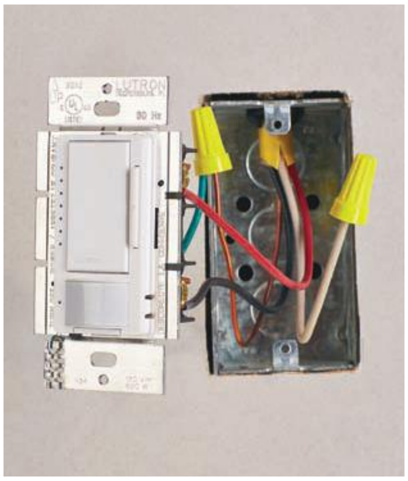

Occupancy sensor. Many smart switches incorporate a motion detector that will switch the lights on if they sense movement in the room and will also shut them off when no movement is detected for a period of time. The model shown above also has a dimmer function for further energy savings.

Programmable timer switch. A dial-type timer allows you to program the switch to turn on for specific time periods at designated times of day within a 24-hour cycle. Security lights, space heaters, towel warmers, and radiant floors are typical applications.

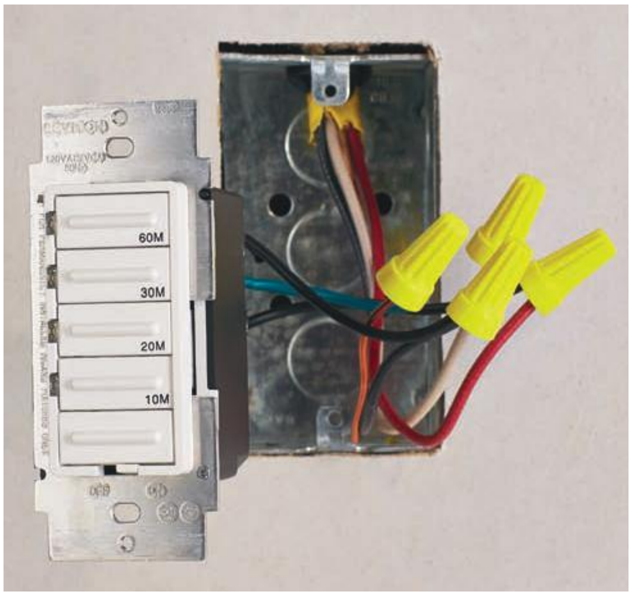

Preset timer switch. This lets you turn on lights, heat lamps, and other loads for a designated amount of time (10 to 60 minutes) with one easy push of a button. The green LED at the bottom of this unit provides a readout of how much time is left before the switch shuts off. The model shown is not compatible with fluorescent ballasts.

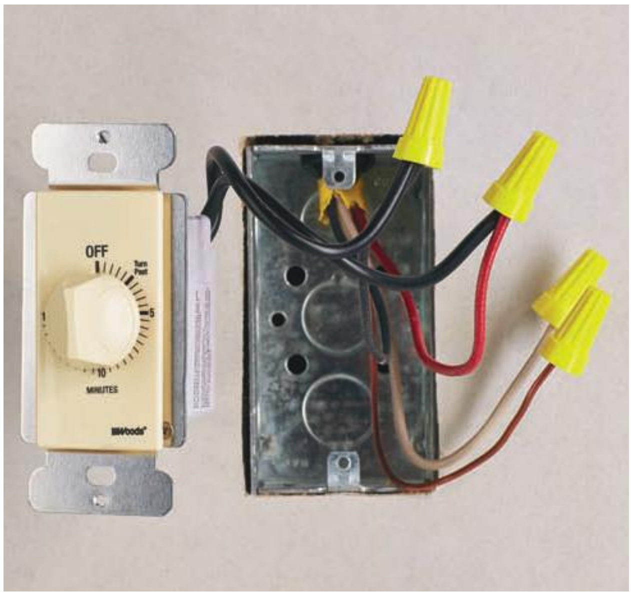

Spring-wound timer switch. A relatively simple device, this timer switch functions exactly like a kitchen timer, employing a hand-turned dial to and spring mechanism to shut the switch off in increments up to 15 minutes.

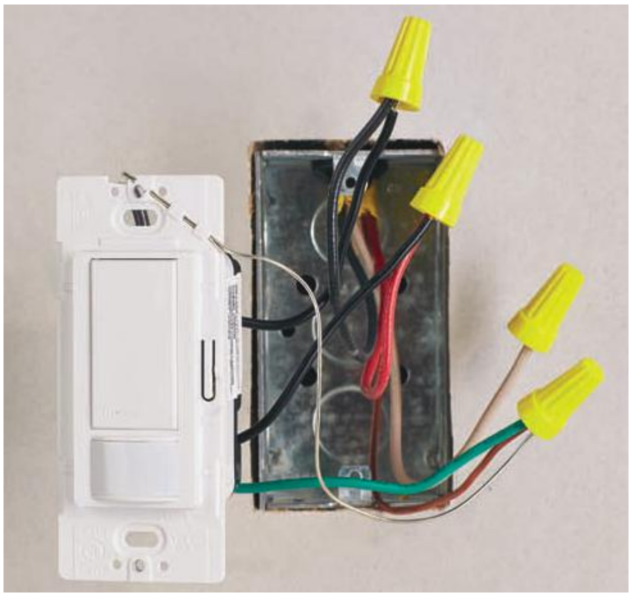

Daylight sensor switch. This switch automatically turns on when light levels drop below a proscribed level. It can also be programmed as an occupancy sensor to shut off when the room is vacant and turn on when the room is entered.

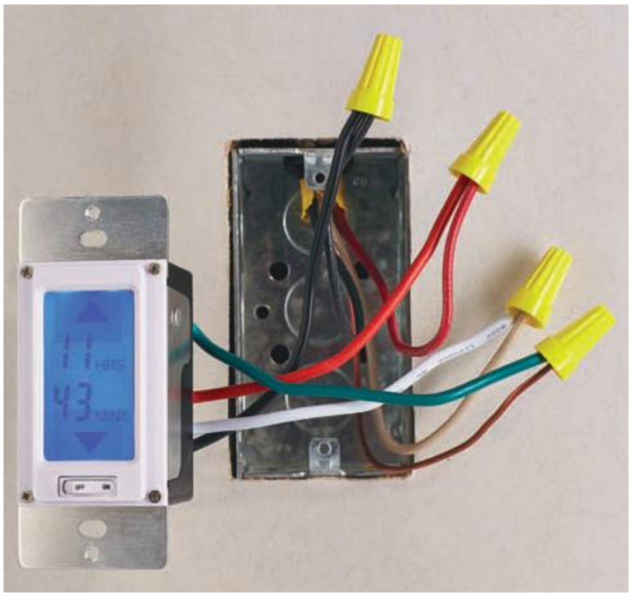

Backlit countdown timer. This digital switch lets you program lights or other devices to stay on for up to 24 hours and then shut off automatically. The backlit, LED readout gives a countdown, in minutes, of the amount of time left in the "on" cycle. Up and down buttons let you raise or lower the remaining time easily, and a manual override button will shut off the switch until it is turned back on.

1st
proofs for reference only
Edition
ge: 118
41940 - Advanced Home Wiring_064-128.indd 119
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 12:05 AM
Page: 119

Ground-fault (GFCI) & Arc-fault (AFCI) Protection

GROUND-FAULT LOCATION REQUIREMENTS

1. Kitchen receptacles. Install ground-fault circuit interrupt (GFCI) protection on all 120-volt receptacles that serve kitchen countertops. This does not include receptacles under the kitchen sink receptacles located on kitchen walls that do not serve the countertop and receptacles that are not within six feet of a sink.
2. Kitchen. Install ground-fault circuit interrupt (GFCI) protection on the outlets that supply dishwashing machines.
3. Bathroom receptacles. Install ground-fault circuit interrupt (GFCI) protection on all 120-volt receptacles located in bathrooms. This applies to all receptacles regardless of where they are located in the bathroom and includes receptacles located at countertops, inside cabinets, and along bathroom walls. This also applies to bathtubs and shower stalls that are not located in a bathroom. Install ground-fault circuit interrupt (GFCI) protection on all circuits serving electrically heated floors in bathrooms, kitchens, and around whirlpool tubs, spas, and hot tubs.
4. Garage and Accessory Building receptacles. Install ground-fault circuit interrupt (GFCI) protection on all 120-volt receptacles located in garages and grade-level areas of unfinished accessory buildings.
5. Exterior receptacles. Install ground-fault circuit interrupt (GFCI) protection on all 120-volt receptacles located outdoors. This does not apply to receptacles that are dedicated for deicing equipment and are located under the eaves. This applies to holiday lighting receptacles located under the eaves.
6. Basement receptacles. Install ground-fault circuit interrupt (GFCI) protection on all 120-volt receptacles located in unfinished basements. An unfinished basement is not intended as habitable space and is limited to storage and work space.
7. Crawl space receptacles. Install ground-fault circuit interrupt (GFCI) protection on all 120-volt receptacles located in crawl spaces. Receptacles in crawl spaces are not required unless equipment requiring service is located there.
8. Sink receptacles. Install ground-fault circuit interrupt (GFCI) protection on all 120-volt receptacles that are located within six feet of the outside edge of a sink. This includes wall, floor, and countertop receptacles.
9. Boathouse receptacles. Install ground-fault circuit interrupt (GFCI) protection on all 120-volt receptacles located in boathouses.

LEVITON
RESET
TEST
15A 125V RECEPTACLE

Ground-fault receptacles and circuit breakers detect unwanted current running between an energized wire and a grounded wire.

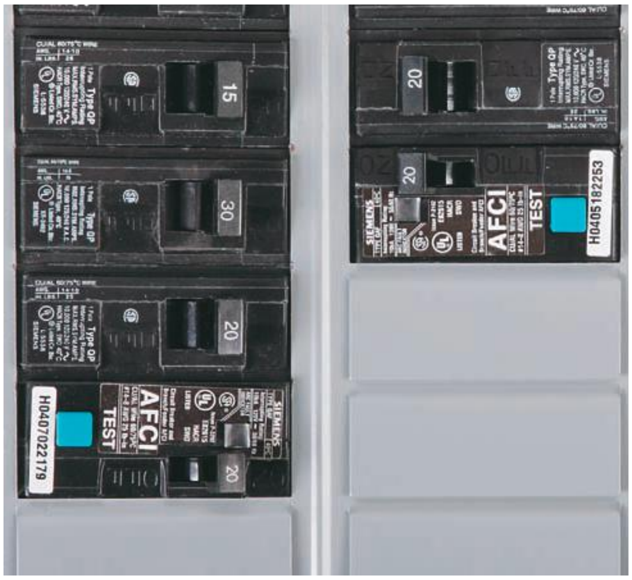

A combination ARC-fault circuit breaker detects sparking (arcing) faults along damaged energized wires and detects these faults between wires. A branch ARC-fault circuit breaker only detects arcing faults between wires.

1st proofs for reference only
41940 - Advanced Home Wiring_064-128.indd 120
10/13/14 10:50 AM
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
Page: 120
Pro-Vision C M Y K Text

10. Spas, tubs, and other circuits requiring ground-fault protection. Install ground-fault circuit interrupt (GFCI) protection on all circuits serving spa tubs, whirlpool tubs, hot tubs, and similar equipment. Refer to the general codes for more information about receptacles serving these components.
11. Install GFCI circuit breakers and receptacles so that they are readily accessible.

ARC-FAULT LOCATION REQUIREMENTS

1. Install a combination type or an outlet (receptacle) type arc-fault circuit interrupter (AFCI) on all 15- and 20-amp, 120-volt branch circuits serving sleeping, family, dining, living, sun, and recreation rooms, kitchens, laundry areas, and parlors, libraries, dens, hallways, closets, and similar rooms and areas. This means that 15-and 20-amp, 120-volt branch circuits serving most interior spaces in a home are required to have AFCI protection. Note that garages, basements, utility and mechanical rooms, and exterior branch circuits are not included in this list, although local building officials may include these areas by interpretation.
2. You may provide AFCI protection for the entire branch circuit by installing a combination-type AFCI circuit breaker in the electrical panel where the branch circuit originates.
3. You may provide AFCI protection to a branch circuit using several different combinations of branch-circuit type AFCI circuit breakers and branch-circuit type AFCI receptacles. Refer to general codes or your local building inspector for details about these alternate methods.
4. Provide AFCI for branch circuits that are modified, replaced, or extended. You may use either of the following methods: (a) install a combination-type AFCI circuit breaker in the electrical panel where the branch circuit originates, or (b) install a branch-circuit type AFCI receptacle at the first receptacle in the existing branch circuit.
5. Install AFCI circuit breakers and receptacles so that they are readily accessible.

Receptacles for whirlpool tubs must be GFCI protected.

1st proofs for reference only

Edition
ge: 120
41940 - Advanced Home Wiring_064-128.indd 121
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 121

Junction Boxes, Device Boxes & Enclosures

All electrical boxes are available in different depths. A box must be deep enough so a switch or receptacle can be removed or installed easily without crimping and damaging the circuit wires. Replace an undersized box with a larger box using the Electrical Box Fill Chart (see page 124) as a guide. **The NEC also says that all electrical boxes must remain accessible. Never cover an electrical box with drywall, paneling, or wall coverings.**

NONMETALLIC BOX INSTALLATION

1. Use nonmetallic boxes only with NM type cable or with nonmetallic conduit or tubing. You may use nonmetallic boxes with metallic conduit or tubing if you maintain the electrical continuity of the metallic conduit or tubing by installing a bonding jumper through the box. In many situations it is easier to use a metallic box with metallic conduit or tubing.
2. Extend NM cable sheathing at least ¼ inch into a nonmetallic box knockout opening.
3. Secure NM cable, conduit, and tubing to each box. You may secure NM cable with cable clamps inside the box or with compression tabs provided where the cable enters the box. You do not need to secure NM cable to a standard single-gang box (2¼ by 4 inches) mounted in a wall or ceiling if you fasten the cable not more than eight inches from the box and if the sheathing enters the box at least ¼ inch. Measure the eight inches along the length of the sheathing, not from the outside of the box.

LIGHT FIXTURE BOX INSTALLATION

1. Use boxes designed for mounting light fixtures if a light fixture is to be mounted to the box. These boxes are usually four-inch round or octagonal.
2. You may use other boxes to mount light fixtures on walls if the fixture weighs less than 6 pounds and if the fixture is secured to the box using at least #6 screws.
3. Support light fixtures weighing at least 50 pounds independently from the light fixture box. You may use the light fixture box to support light fixtures weighing less than 50 pounds. Note that ceiling fans are not light fixtures.

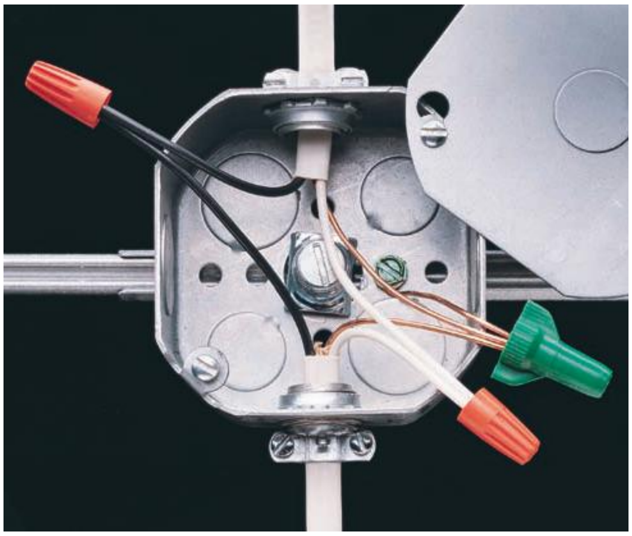

Box shape is directly related to function, as electrical fixtures are created to fit on boxes of a particular shape. Octagonal and round boxes generally are designed for ceiling mounting, while square and rectangular boxes are sized for single-pole, duplex, and other standard switch and receptacle sizes.

Do not support heavy light fixtures using only the light fixture electrical box. The eye hook supporting this chandelier is driven into the same ceiling joist to which the electrical box is mounted.

1st proofs for reference only
41940 - Advanced Home Wiring_064-128.indd 122
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM
Page: 122

BOX CONTENTS LIMITATIONS

1. Limit the number of wires, devices (such as switches and receptacles), and fittings in a box. This limitation is primarily based on the heat generated by the wires and devices in the box. The actual size of the box relative to its contents is a secondary consideration.
2. Use the cubic inch volume printed on the box or provided in the box manufacturer's instructions to determine box volume. Do not attempt to measure the box volume. Do not estimate box volume from the volume of similar size boxes. You will probably not get the same volume as provided by the manufacturer.
3. Use table "Wire Volume Unit" to determine the volume units required by wires, devices, and fittings in a box.

BOX INSTALLATION TOLERANCES

1. Install boxes in non-combustible material, such as masonry, so that the front edge is not more than ¼ inch from the finished surface.
2. Install boxes in walls and ceilings made of wood or other combustible material so that the front edge is flush with the finished surface or projects from the finished surface.
3. Cut openings for boxes in drywall and plaster so that the opening is not more than ⅛ inch from the perimeter of the box.

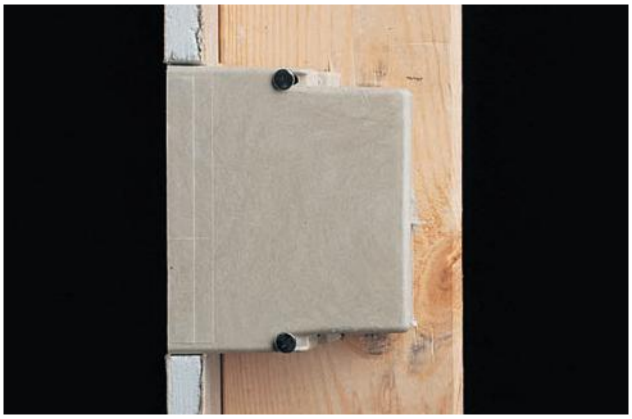

Boxes must be installed so the front edges are flush with the finished wall surface, and the gap between the box and the wall covering is not more than ⅛".

Wire Volume Unit ▸

WIRE SIZE (AWG)	WIRE VOLUME
14	2.00 in.^{3}
12	2.25 in.^{3}
10	2.50 in.^{3}
8	3.00 in.^{3}
6	5.00 in.^{3}

Volume Units ▸

Calculate the volume units required by wires, devices, and fittings based on the following definitions:

Volume units for current-carrying wires. Allow one volume unit for each individual hot (ungrounded) and neutral (grounded) wire in the box. Use Table 47 to determine the volume units of common wire sizes. Example: two pieces of #14/2 NM are in a box. Each piece of this cable contains one hot (ungrounded) and one neutral (grounded) wire and one grounding wire. From table "Wire Volume Unit", each #14 wire uses 2.00 cubic inches in the box. The total volume units required by the hot (ungrounded) and neutral (grounded) wires is eight cubic inches.

Volume units for devices. Allow two volume units for each device (switch or receptacle) in the box. Base the volume units on the largest hot (ungrounded) or neutral (grounded) wire in the box. Example: NM cable size #14 and #12 are in a box. From Table 47, #14 wire uses 2.00 cubic inches and #12 wire uses 2.25 cubic inches. Allow 4.5 cubic inches volume units (2 × 2.25 cubic inches) for each switch or receptacle in the box based on the volume of the larger #12 NM cable.

Volume units for grounding wires. Allow one volume unit for all grounding wires in the box. Base the volume unit on the largest hot (ungrounded) or neutral (grounded) wire in the box.

Volume units for clamps. Allow one volume unit for all internal cable clamps in the box, if any. Base the volume unit on the largest hot (ungrounded) or neutral (grounded) wire in the box.

Volume units for fittings. Allow one volume unit for all fittings in the box, if any. Base the volume unit on the largest hot (ungrounded) or neutral (grounded) wire in the box.

1st proofs for reference only
41940 - Advanced Home Wiring_064-128.indd 123
Pro-Vision C M Y K Text
Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition
10/10/14 12:05 AM
Page: 123
Edition
ge: 122

Electrical Box Fill Chart ▸

BOX SIZE AND SHAPE	**MAXIMUM NUMBER OF VOLUME UNITS PERMITTED (SEE NOTES BELOW)**			
(If volume not labeled by manufacturer)	**14 AWG**	**12 AWG**	**10 AWG**	**8 AWG**
JUNCTION BOXES				
4 × 1¼" R or O	6	5	5	4
4 × 1½" R or O	7	6	6	5
4 × 2⅛" R or O	10	9	8	7
4 × 1¼" S	9	8	7	6
4 × 1½" S	10	9	8	7
4 × 2⅛" S	15	13	12	10
4¹¹⁄₁₆ × 1¼" S	12	11	10	8
4¹¹⁄₁₆ × 1½" S	14	13	11	9
4¹¹⁄₁₆ × 2⅛" S	21	18	16	14
DEVICE BOXES				
3 × 2 × 1½"	3	3	3	2
3 × 2 × 2"	5	4	4	3
3 × 2 × 2¼"	5	4	4	3
3 × 2 × 2½"	6	5	5	4
3 × 2 × 2¾"	7	6	6	4
3 × 2 × 3½"	9	8	7	6
4 × 2⅛ × 1½"	5	4	4	3
4 × 2⅛ × 1⅞"	6	5	5	4
4 × 2⅛ × 2⅛"	7	6	5	4

Notes:

- R = Round; O = Octagonal; S = Square or rectangular
- Each hot or neutral wire entering the box is counted as one volume unit
- Grounding wires are counted as one volume unit in total—do not count each one individually.
- Raceway fittings and external cable clamps do not count. Internal cable connectors and straps count as one volume unit.
- Devices (switches and receptacles mainly) each count as two volume units.
- When calculating total volume units, any non-wire components should be assigned the gauge of the largest wire in the box.
- For wire gauges not shown here, contact your local electrical inspections office.

BOX SUPPORT IN WALLS, CEILINGS & FLOORS

1. Provide support for boxes that rigidly and securely fasten them in place. You may use nails or screws to support these boxes.
2. Protect screws inside boxes so that the threads will not damage the wires.
3. Wood braces used to support boxes must be at least one by two inches.
4. Use “cut-in” or “old work” retrofit boxes only if they have approved clamps or anchors that are identified for the location where they are installed.

DAMP LOCATIONS

1. Install a receptacle box cover that is weatherproof when the cover is closed and a plug is not inserted into a receptacle located in a damp location. This applies to 15-amp and 20-amp receptacles. A damp area is protected from direct contact with water. Refer to the definition of damp location. You may use a receptacle cover suitable for wet locations in a damp location.
2. Install a watertight seal between a flush-mounted receptacle and its faceplate. This will require a gasket or sealant between the finished surface (such as stucco, brick, or siding) and the faceplate.

WET LOCATIONS

1. Install a receptacle box cover that is weatherproof when the cover is closed on any receptacle located in a wet location. This applies to 15-amp and 20-amp receptacles in any indoor or outdoor wet location. This applies regardless of whether or not a plug is inserted into the receptacle.
2. Install a watertight seal between a flush-mounted receptacle and its faceplate. This will require a gasket or sealant between the finished surface (such as stucco, brick, or siding) and the faceplate.

1st proofs for reference only

41940 - Advanced Home Wiring_064-128.indd 124

Pro-Vision C M Y K Text

Job: 41940 Title: CSP - Advanced Home Wiring-4th Edition 10/10/14 12:05 AM

Page: 124

Pro-Visi

The Crystal Radio

QST December 1997, pp. 56-57

Copyright © 1997 by the American Radio Relay League, Inc. All rights reserved

By Dave Evison, W7DE

The Crystal Radio

What's old is new again!

Even a casual auditor of the amateur bands will conclude that the hobby is largely populated by wizened old geezers—and I'm one of them. This is a sad situation because most of us entered Amateur Radio as kids. If ham radio is to continue for generations to come, we must get youngsters interested in radio communication while they're still in grammar school and still capable of wonder. We need to reach them while they are still asking questions about everyday matters (that all too often go unanswered) such as, "Why does it get light in here when I push this thing?"

Reaching today's kids won't be easy. With the Internet, cellular telephones and other forms of global communication commonplace in their lives, the true magic of radio lies buried beneath the technology. We must strip away layers upon layers of complicated hardware and reveal radio in its purest form. The perfect tool to achieve this goal is the "lowly" crystal receiver.

Imagine the curiosity you'll inspire when you demonstrate a radio that has no obvious power source whatsoever! The simple crystal receiver described in this article works surprisingly well. It's selective and produces robust headphone volume. I've been able to tune several local AM broadcast stations with ease. Best of all, it's inexpensive, uses readily available parts, and can be built by a youngster (with a little guidance). Of course, adults will love this little receiver, too.

There is a fascinating resurgence of interest in crystal radios. Even hams normally accustomed to owning rigs with four-digit price tags are rediscovering the joys of simple, "powerless" reception. To get a taste of what is going on, pick up a copy of *Crystal Set Projects* from the Xtal Set Society, Box 3026, St Louis, MO 63130; tel 314-725-1172, or visit their Web site at **http://www.midnightscience.com**. (The book was reviewed in the August, 1997 *QST*, page 102.)

About the Circuit

The circuit shown in Figure 1 was developed in the early '30s by Elmer G. Osterhoudt. Selectivity problems plagued most crystal sets of the day due to the proliferation of broadcast stations. Elmer devised a solution. He determined that poor crystal receiver selectivity was caused by the headphones swamping the tuned circuit and lowering its Q. The net result was an inability to separate individual signals. Elmer improved selectivity considerably by connecting the detector near the cold (ground) end of the tuned circuit to minimize the swamping.

Capacitor C2 is used to match the antenna system, and C3 is the main tuning capacitor. Absent from this circuit is the finicky galena-and-catwisker detector. It has been supplanted by a 1N34 germanium diode that is connected near the grounded end of the tuned circuit that includes coil L1. C1 is an RF bypass capacitor.

Headphones

You must use high-impedance headphones (2000 Ω or greater) with crystal receivers. If you attempt to use common 8-Ω headphones, you'll have very disappointing results. You'll find high-impedance headphones at hamfest flea markets, but if you're not up to the hunt, you can buy them from Antique Electronic Supply (see Figure 1).

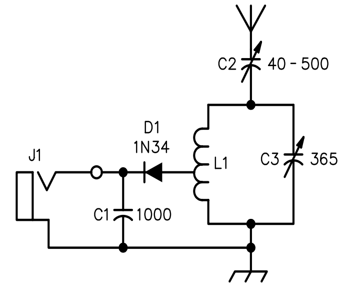

Figure 1—Schematic diagram of the crystal receiver. Parts are available from Radio Shack or Antique Electronic Supply (6221 South Maple Ave, Tempe, AZ 85283; tel 602-820-5411) as noted.

C1—1000 pF mica or disc ceramic trimmer (Radio Shack 272-123)
C2—40 to 500 pF trimmer capacitor (Antique Electronic Supply CV-263)
C3—365 pF variable capacitor (Antique Electronic Supply CV-231)
D1—1N34 germanium diode (Radio Shack 276-1123)
J1—1/4-inch, 2-conductor headphone jack (Radio Shack 274-252)
L1—35 feet of 22-gauge enamel wire. See text. (Radio Shack 278-1345)
Fahnestock clips (Antique Electronic Supply SH-11-4034)
2000-Ω headphones (Antique Electronic Supply PA-466)

The Importance of a Good Antenna System

Since the crystal receiver depends completely on the energy from the transmitted signal, you need to gather as much signal out of the air as possible! Remember that this receiver is designed to tune the AM broadcast band. A half-wavelength dipole cut for about 1 MHz would be ideal, but few of us own enough real estate to accommodate a 500-foot long antenna. The practical solution is

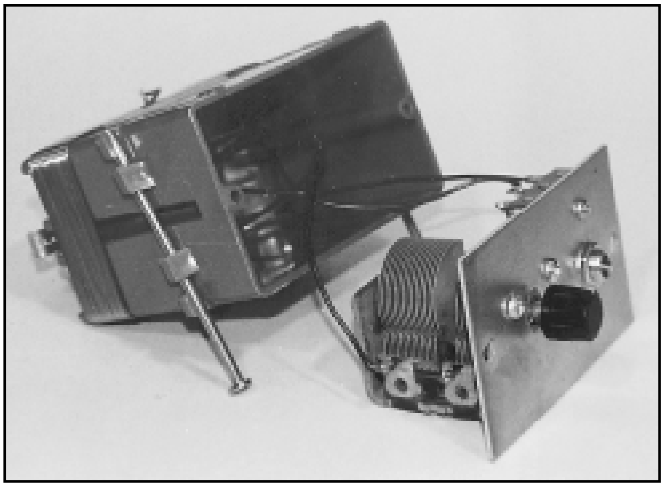

Plastic electrical outlet boxes make terrific enclosures. Note that I've mounted the main tuning capacitor, C3, and the headphone jack on the front using copper-clad PC board.

to simply use an end-fed wire about 100 feet in length. You'll also need to provide a connection to an Earth ground (a cold-water pipe may suffice). With the proper antenna system, you should be able to receive most of the AM stations in your area. If you live in a rural location without an AM broadcaster nearby (within about 50 miles), try using even more antenna wire. Just string it over trees, bushes, or any other supports.

Construction

You can build a crystal receiver into almost anything. I used plastic electrical outlet boxes. They're available at home centers and hardware stores. They are attractive, rugged and inexpensive. They even serve as forms for winding the coil. I made the front panel of my box from a piece of copper-clad PC board. You can cut this material into the proper shape using a hacksaw, and it is easily drilled for mounting parts.

For C3, a miniature variable capacitor salvaged from an old AM-only radio will work, but a single-section air variable is better. If you don't have these capacitors in your junk box, don't despair. You can order one from Antique Electronic Supply.

L1 is wound on the outside rear portion of the enclosure and consists of 32 turns of 22-gauge enamel-covered wire. The coil is tapped nine turns from the grounded end. The tap is made by tightly twisting about 1 inch of the magnet wire and scraping off the enamel insulation. You can secure the windings to the enclosure with transparent plastic tape.

The antenna trimmer capacitor (C1) and the Fahnestock clips for the antenna and ground connections are mounted on the rear of the enclosure. The tuning capacitor, phone jack and detector are mounted on the front panel. Interconnecting wires are fed through holes in the enclosure and soldered to the trimmer capacitor, ground and cold connections.

Operating Adjustments

Connect the antenna, ground and headphones. Tighten the antenna trimmer capacitor C2. Slowly adjust C3 to tune in the stations. Note the number of stations and log their call letters and frequencies. Readjust C1 to provide the best trade-off between volume and selectivity. If you're using your crystal radio near a particularly strong broadcast station, removing the ground connections will often improve selectivity (with some reduction in volume).

Use Your Radio as a Teaching Tool!

You can use your crystal radio to teach a number of important concepts. For example:

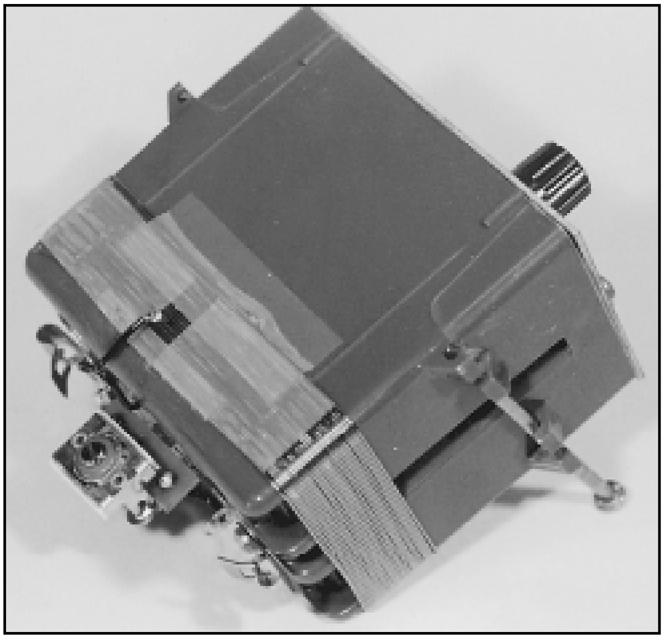

The rear view of the crystal radio enclosure. You can easily see the windings of coil L1 held in place with transparent tape.

Inductance: Applying the "right hand rule," the coil for the tuned circuit can be used to explain how a magnetic field is developed by electron flow through a conductor, and how the magnetic field is increased by forming the conductor into a coil.

Capacitance: Varying capacitance may be illustrated by simply observing how the total area increases as the rotor plates mesh with the stator plates. The compression trimmer capacitor (C1) illustrates the function of a dielectric other than air. It also shows how the capacitance can be varied by changing the separation between the plates.

Resonance, modulation and detection: You can teach these concepts by inviting your student to tune in a station while you display the modulated RF on an oscilloscope! Move the 'scope probe from one side of diode D1 to the other to illustrate "before" and "after."

Magnetism and transduction: High-impedance headphones are ideal instructional aids to demonstrate these ideas. Disassemble a pair of headphones and show your students what's inside. Touch the permanent magnet with a screwdriver so they can see that it is indeed magnetic. Explain how the detected signal flows through the tiny headphone coil, creating a magnetic field that adds or subtracts from the field produced by the permanent magnet. This causes the metal diaphragm to vibrate, creating sound! What began as electrical energy has become mechanical energy—*transduction*.

Summary

Throughout the years, the simple crystal radio has spawned tens of thousands of radio amateurs, engineers and electronic hobbyists. And despite the antiquity of its design, it is still an effective tool for recruitment and instruction. A crystal radio is a delightful device, and the perfect way to introduce the wonder of radio to a young mind—or rekindle it in your own!

---

153 Park Ave
Palo Alto, CA 94306
e-mail w7de@aol.com

COBISS: 1.01

PRELIMINARY U/TH DATING AND THE EVOLUTION OF GYPSUM CRYSTALS IN NAICA CAVES (MEXICO)

PREDHODNI REZULTATI DATIRANJA Z METODO U-TH IN RAZVOJ SADRINIH KRISTALOV V JAMI NAICA (MEHIKA)

Laura SANNA[1,*], Paolo FORTI[2] & Stein-Erik LAURITZEN[1,3]

Abstract UDC 551.435.84(72):553.635

Laura Sanna, Paolo Forti & Stein-Erik Lauritzen: Preliminary U/Th dating and the evolution of gypsum crystals in Naica caves (Mexico).

The origin and the evolution of giant selenite crystals in Naica caves, together with the understanding of their growth mechanisms, is one of the aims of the international multidisciplinary research, called the "Naica Project". In this context, the exact timing of when the gypsum nucleation started and whether its growth has been constant over time, have been investigated. The preliminary data obtained with the U–Th disequilibrium method show significant differences in ages for gypsum (between 191 ± 13 kyr for one of the Ojo de la Reina cave crystals and 57 ± 1.7 kyr for the base of Espadas cave's spar) and have produced a coarse chronological interval of growth. The crystal depositional rates vary from 0.56 to 1.22 mm/kyr, in excellent agreement with the laboratory tests for gypsum deposition under present conditions performed in the deepest part of the mine. These results are also consistent with a multistage precipitation started at different times in the Naica caves (first in caves at the upper level, where gypsum was subsequently dissolved, and only later in the deeper part of the aquifer under stable conditions) and allow us to improve the knowledge on the speleogenetic evolution of these caves.

Keywords: U-Th dating, gypsum speleothems, mine caves, crystal growth rate, Naica, Mexico.

Izvleček UDK 551.435.84(72):553.635

Laura Sanna, Paolo Forti & Stein-Erik Lauritzen: Predhodni rezultati datiranja z metodo U-Th in razvoj sadrinih kristalov v jami Naica (Mehika)

Izvor, razvoj in mehanizmi rasti velikih selenitnih kristalov v jamah Naica, so del raziskav v okviru mednarodnega projekta »Naica«. V okviru teh smo določali natančen čas začetka nukleacije sadre in stalnost rasti kristalov v času. Prvi podatki datiranj z uran-torijevo metodo kažejo na pomembne razlike v starosti kristalov sadre (med 191 ± 13 ky za kristale v jami Ojo de la Reina in 57±1.7 ky v korenu kristala Espada) in veliko časovno obdobje rasti. Hitrost rasti kristalov je med 0,56 in 1,22 mm/ky, kar se odlično ujema z eksperimentalnimi rezultati v trenutnih pogojih v najglobljem delu rudnika. Rezultati podpirajo tezo o več stopenjskem izločanju kristalov (prvotno v jamah gornjega nivoja, kjer je bila sadra naknadno raztopljena in kasneje v stabilnih razmerah globoko v vodonosniku) in omogočajo nove interpretacije speleogeneze na obravnavanem območju.

Ključne besede: Datiranje z metodo U-Th, sadra, jame v rudniku, hitrost rasti kristalov, Naica, Mehika.

---

[1] GEO, Department of Earth Sciences, Bergen University, Norway, e-mail: speleokikers@tiscali.it

[2] Earth and Geological Environmental Sciences, University of Bologna, Italy

[3] Department of Plant and Environmental Sciences, University of Ås, Norway

Received/Prejeto: 22.9.2010

INTRODUCTION

Giant selenite crystals discovered inside natural caves in the Naica Mine (State of Chihuahua, Northwest Mexico) are among the most interesting geological revelations of recent years (Badino & Forti 2005). In 2006, a multidisciplinary research project was started by Speleoresearch & Film from Mexico with the co-operation of La Venta Exploring Team (Forti 2006). The aim of this project is to investigate the origin and the evolution of these gypsum crystals, together with the understanding of their growth mechanisms, and the kind of relationships existing between the ore bodies and these crystals. Even though some of this research has already brought important scientific results (Bernabei *et al.* 2009), e.g. the detailed survey of Cueva de los Cristales (Crystals Cave) (Badino *et al.* 2009), the comprehension of the genesis and the evolution of the caves in the Naica Mine (Bernabei *et al.* 2007; Forti *et al.* 2007), and the minerogenetic environments (Forti *et al.* 2008, 2009a, b; Garcìa-Ruiz *et al.* 2007; Garofalo *et al.* 2007; Panieri *et al.* 2008), the exact timing when the nucleation started and whether growth has been constant over time was still to be investigated. For this reason in 2006 an underground laboratory was set up at the -590 Level in the Naica Mine with the goal of investigating the crystal growth rate at the present conditions (Forti & Lo Mastro 2008), and in 2007 some selenite samples from the three main caves (Cueva de los Cristales, Ojo de la Reina and Cueva de las Espadas) were collected for dating purposes. A U and Th extraction procedure using isotope dilution for radiometric dating intent of gypsum has been developed in the U-series Dating Laboratory of Bergen University (Sanna *et al.* 2010). This paper shows preliminary age results on the gypsum Naica samples and their relationship with the local evolution of these crystallizations.

EXPERIMENTAL OBSERVATIONS

Besides the wide desolate landscape around the Sierra de Naica (Chihuahua, Mexico), small amazing geological secrets (Foshag 1927; Shagun 2001; London 2003) are hidden inside this anticline structure of Cretaceous carbonate rocks (Megaw *et al.* 1988). During the last century, the depths of this NW-oriented mountain have been penetrated by several kilometres of mine galleries, and the extensive silver exploitation has intercepted some natural cavities at different levels inside the mountain (Fig. 1), most of which host huge selenite crystals grown in underwater conditions (Forti *et al.* 2007). The gypsum deposition occurred in three different environments (deep phreatic, epiphreatic and aerate) starting in the upper part of the mountain profile, where the water temperature first decreased below 59 °C (Forti *et al.* 2009a). Gypsum growth is controlled by the dynamic evolution of a stable supersaturated flow system and went on until some 25 years ago, when the water eduction for mine exploitation purposes caused the lowering of the watertable from -130 m to -760 m, thus dewatering the caves (Garofalo *et al.* 2007). All these caves developed along the main faults, which guided the uplift of these thermal fluids (Forti & Sanna 2010). Fluids still rise in the mine galleries of level -590 and have a temperature of 51-54 °C.

The most famous explored cavities in the Naica Mine are Cueva de los Cristales (Crystals Cave), Ojo de la Reina (Queen's Cave) and Cueva de las Espadas (Swords Cave), and all of those host euhedral gypsum crystals for which post-depositional alteration and open systems - necessary conditions for dating purposes - are excluded.

Cueva de los Cristales is a big chamber of about 30 m diameter and 20 m height, intercepted by the mine galleries at the -290 Level (Fig. 2b), that hosts a network of the largest known, prismatic shaped, gypsum single crystals in the world (Fig. 2a) (the largest measures 11 m in length and 1 m in diameter) (Badino *et al.* 2009). From Cueva de los Cristales we sampled two broken crystals (N1 and N07) found on the cave floor. The dated subsamples were collected from slices cut perpendicularly to their elongation axis: a) N07-10 was drilled 13 cm from the outer surface on a slide about 40 cm in diameter on the N07 crystal (Fig. 2); b) N1 was collected 5 cm from the outer surface on a slide about 20 cm in diameter on the N1 crystal (Figure 2d).

Ojo de la Reina is a small cave consisting of a narrow fracture discovered close to Crystals Cave (Fig. 2f). It is completely filled by transparent and pinacoid-shaped crystals grown in several steps (Fig. 2e), as demonstrated by black sulphur layers in the 4-m high exposed crystal wall below the entrance (Fig. 2g). The Reina sample (N01-1) is composed of loosely bound laminae of completely transparent gypsum, taken at the bottom of this crystal wall almost at the contact with the dolomitic limestone: therefore this sample should represent to the

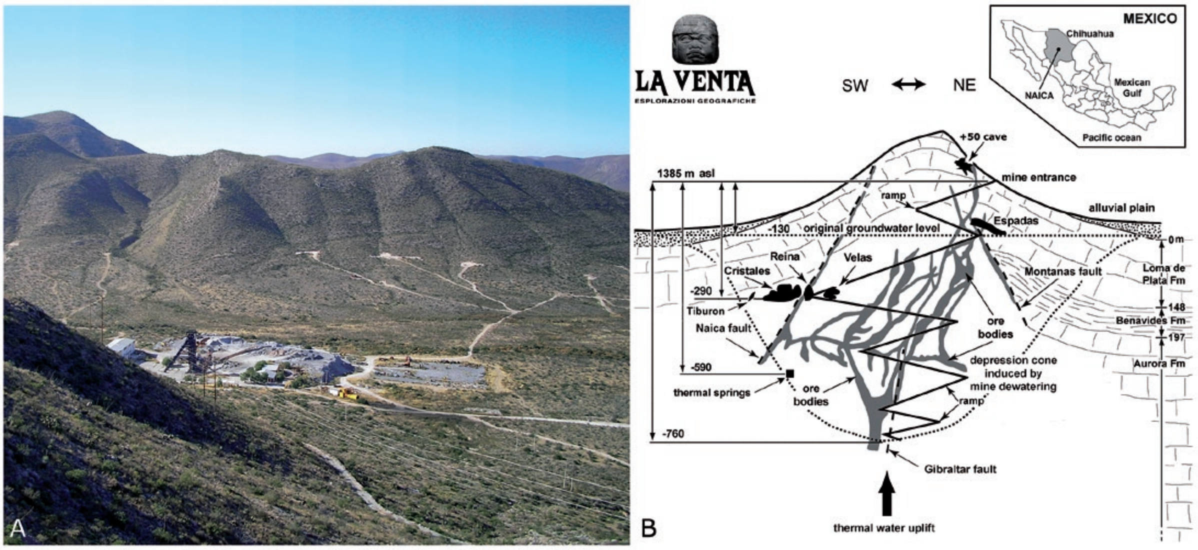
A
B
LA VENTA
ESPLORAZIONI GEOGRAFICHE
SW ↔ NE
MEXICO
Chihuahua
NAICA
Mexican Gulf
Pacific ocean
+50 cave
1385 m asl
mine entrance
ramp
alluvial plain
Espadas
-130 original groundwater level
0m
Reina
Velas
Cristales
Montanas fault
Loma de Plata Fm
-290
Tiburon
148
Benavides Fm
Naica fault
ore bodies
197
Aurora Fm
depression cone induced by mine dewatering
-590
thermal springs
ore bodies
ramp
-760
Gibraltar fault
thermal water uplift

Fig. 1: a) The desert landscape of the Sierra de Naica in a general view of the south-western side of the anticline structure (Photo: L. Sanna, La Venta & S/F Archives); b) Sketch of the Naica mine with the locations of the main natural cavities - not to scale (after Badino & Forti 2007, modified).

very beginning of the gypsum deposition in the -290 Level. Actually, due to the disposition of the cave (a subvertical fracture) and of the crystals, one cannot be sure that the sample was collected just in the inner (oldest) core of the crystal bulk. Thus the gypsum deposition date may somehow appear older than the age obtained from this sampling.

Cueva de las Espadas is a distinct sub-horizontal conduit at the -120 Level where single crystals up to 2 m in length occur, while other, smaller, crystals completely cover its walls (Fig. 2h & 2j). This cave develops along a fault, and there are clear evidences of a complex speleogenesis of hydrothermal and meteoric behaviour (Panieri *et al.* 2008). The Espadas sample is a subaqueous speleothem found just in the bottom of the cave. Its nucleus consists of a prismatic gypsum crystal enclosed inside two layers of white acicular aragonite crystals (pointing to the partial emersion of the cave), alternated with two layers of gypsum macrocrystals (Fig. 2i). The surface of the speleothem was covered by a thin layer of hazel-brown calcite and by scarcely cemented clay-silty deposits (Forti 2007). Three different subsamples were collected: a) a gypsum subsample drilled about 6 cm from the core (ESP1-1); b) a 2-mm thick subsample of aragonite layer in contact with the gypsum core (ESP1-2); c) a 1-mm thick calcite cover (Fig. 2k).

EXPERIMENTAL DATING OF GYPSUM

The U-Th method is the most widely-used dating technique applied to carbonate speleothems (Walker 2008), but none of the published papers indicate that this method is suitable for dating of gypsum, due to its very low U content and owing to the difficulty in separation and purification of the samples and the high solubility of sulphate that produces post-depositional alteration (Sancho *et al.* 2004). For these reasons some authors prefer the Electron Spin Resonance technique (Mathew *et al.* 2004; Ikeda & Ikeya 1991). However, few studies have been published on the determination of U/Th ratio in a variety of gypsum samples using different procedures, such as the α-spectrometric technique (Hendy *et al.* 1979), isochron dating method (Luo & Ku 1991), thermal ionization mass spectrometry (TIMS) (Peng *et al.* 2001) and inductively coupled plasma mass spectrometry (ICP-MS) (Gamble 2008). There is only one documented test on single gypsum speleothem dating, but the ensuing age showed a large standard error and thus has been regarded as unreliable (Goede *et al.* 1990). Although the high precision analysis of U isotopes with multi-collector ICP-MS (MC-ICP-MS) has been well established (Becker 2003), TIMS has been a widely employed technique for accomplishing the highest accuracy and precision of isotope ratios. Only recently, several studies have focused on the capability of MC-ICP-MS using isotope dilution (ID) for U/Th activity ratios in samples with extremely low U content (Lu 2009), such as in the gypsum crystals of Naica.

Despite the fact that the first chronological data on the giant selenite crystals of Naica were obtained with the traditional TIMS analytical approach (Lauritzen *et*

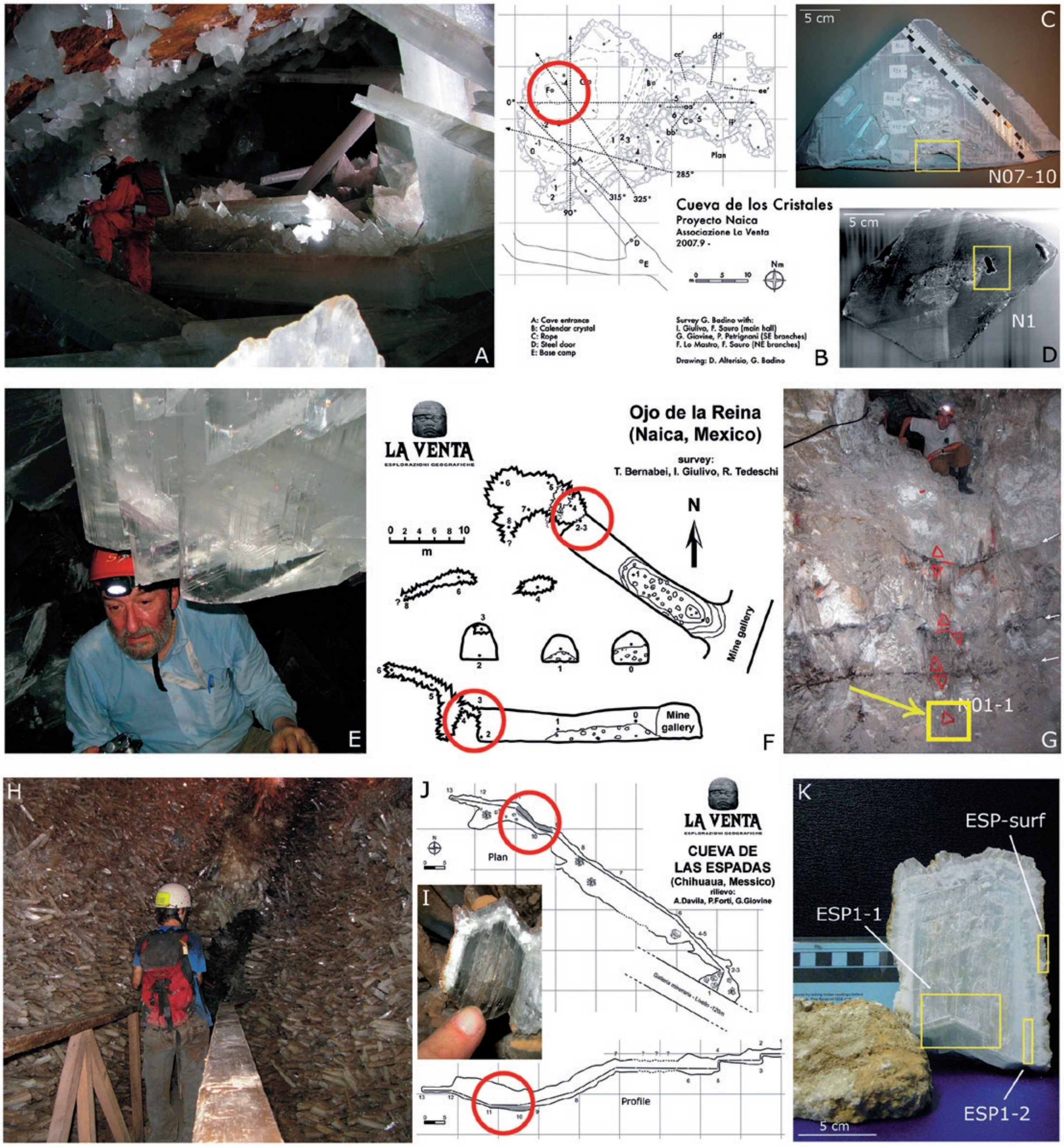
A
B
Cueva de los Cristales
Proyecto Naica
Associazione La Venta
2007.9 -
Plan
A: Cave entrance
B: Calendar crystal
C: Rope
D: Steel door
E: Base camp
Survey G. Badino with:
I. Giulivo, F. Sauro (main hall)
G. Giovine, P. Petrignani (SE branches)
F. Lo Mastro, F. Sauro (NE branches)
Drawing: D. Alterisio, G. Badino
C
5 cm
N07-10
D
5 cm
N1
E
F
LA VENTA
Ojo de la Reina
(Naica, Mexico)
survey:
T. Bernabei, I. Giulivo, R. Tedeschi
N
0 2 4 6 8 10
m
Mine gallery
G
N01-1
H
J
Plan
Profile
I
LA VENTA
CUEVA DE
LAS ESPADAS
(Chihuaua, Messico)
rilievo:
A.Davila, P.Forti, G.Giovine
K
ESP-surf
ESP1-1
ESP1-2
5 cm

Fig. 2: The three main caves of Naica Mine. a) View from the west of the giant selenite crystals (caver for scale) of Cueva de los Cristales. The two samples of already broken crystals (c and d) have been found on the north-western side of the cave floor, red circle in the map (b). e) Well-formed transparent and pinacoid-shaped gypsum crystals hosted inside the narrow fracture of Ojo de la Reina and (g) the 4-m high exposed crystal wall below the entrance, point 2 in the survey (f). The white arrows indicate the black sulphur layers. h) The gypsum speleothems on the wall of Cueva de las Espadas and (i) an euhedral gypsum crystal covered by a layer of carbonate. The dated spar (k) was collected on the floor of a small cave lake, corresponding to the grey area in the map (j). Yellow squares show dated sub-samples (Photos: La Venta & S/F Archives).

al. 2007), this method was time-consuming and expensive, and the results were not satisfying (very low U concentration), so a simpler method with the TRU Resin separation for high resolution ICP-MS detection has been adopted (Sanna *et al.* 2010). After determining the solubility of gypsum in different solutions, to optimize gypsum digestion, and then increasing the amount of the sample up to 5 g to have a greater quantity of actinides, a mixed procedure (after Hellstrom 2003 with several modifications) has been tested using the semi-quantitative U concentration provided by TIMS preliminary analysis, to balance the isotopic tracers (spike). This quantity is one tenth of the normal spike amount usually used for carbonate analyses, since overspiked samples cannot be dated. In the first experiment, with a 5 g sample and 10 μL spike, the yield was very bad. Almost all actinides were lost during the process, but the spiking level was quite good. Contamination of ^{232}Th was also very high. In the following experiment, prior to further raising the quantity of the gypsum sample, the acid concentration was adjusted to amplify the uranium and thorium purification, but during ICP-MS acquisition no actinide signal occurred. Finally, after increasing the selenite amount up to 10 g (in spite of a loss in the stratigraphic resolution), the procedure worked satisfactorily.

For each dating, gypsum subsamples were separated by means of a cutting disk for dentists, from a 20x8 mm thick rod (cut along the elongation axis of crystal), and the gypsum powder was dissolved completely by digestion with 600 mL of 1M HNO_3. The spike was added prior to U and Th purification. After centrifugation, the small insoluble residue was discarded, and the separation of actinides was performed using a preconditioned TRU Resin®. Each sample was loaded into its own columns. The U and Th were eluted by 0.1 M HCl – 0.2 M HF. Afterwards, the U and Th fractions were dried.

In order to carry out the isotopic analysis the actinides dried fraction was dissolved in 2% v/v HNO_3: the isotopic measurements were performed on a Nu Plasma HR multicollector ICP-MS with a U-Pb collector block at the Department of Geology, University of Oslo. Analyses were done in dry plasma using a DSN-100 desolvating nebuliser with a sample uptake rate of 0.1 mL/min (Sanna *et al.* 2010).

EXPERIMENTAL ACCRETION OF GYPSUM

Another method to define the age of the largest crystals is to measure the present rate of growth, but 25 years ago their accretion stopped due to mine dewatering that dried up the caves. Thermal water (51 °C) is now spilling inside the mine galleries only 300 m below the known caves at the -590 Level (Forti & Lo Mastro 2008).

Within Naica caves the principal genetic mechanism for the giant crystals was ruled only by the solubility disequilibrium between anhydrite and gypsum under supersaturated aqueous solution within a temperature range 55-58 °C (Garcìa Ruiz *et al.* 2007; Panieri *et al.* 2008). To restore these conditions the contact between water and air should be avoided, and both evaporation and temperature losses should be kept as low as possible: for this purpose a specific device was planned and realized (Forti & Lo Mastro 2008). In 2006 a vessel (the main component of this apparatus) was placed in a dangerously hot location where original thermal water with a temperature of more than 51 °C is still dripping out from the mine wall rock and 100% relative humidity is present. It was designed to stay there for at least 2 years (Fig. 3). Inside the vessel 11 prismatic polycrystalline samples of Messinian gypsum from the Emilia-Romagna formation (Italy) have been suspended to act as crystallization support in thermal water coming from the spring in the mine gallery. The average surface of the tablets is 66 cm^2. All the tablets were carefully measured and weighed before their use.

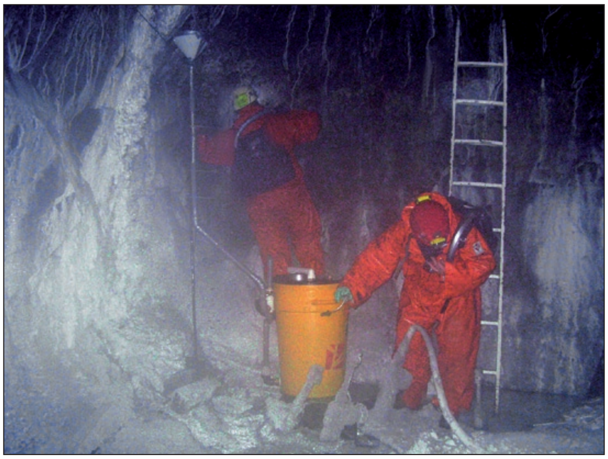

Fig. 3: The experimental laboratory: the vessel was placed in a place with a temperature of more than 51 °C and relative humidity of 100% (Photo: L. Sanna, La Venta & S/F Archives).

RESULTS

ANALYSES OF THE NAICA CRYSTALS

According to Sanna *et al.* (2010) the giant crystals display significant differences in ages (Tab. 1). The core of the broken giant selenite from Cristales began to grow at an imprecise age between 106 – 260 kyr ago, while the first U/Th dating was made by TIMS technique some 5 cm

Tab. 1: Uranium concentration, measured U and Th activity ratios and ages of subsamples from Naica crystals (modified after Sanna et al. 2010). For the subsample locations see Fig. 2.

ID Lab	*ID sample*	*Cave*	*U (ppm)*	*$^{234}U/^{238}U$*	*$^{230}Th/^{234}U$*	*$^{230}Th/^{232}Th$*	*Age, kyr*	*$2\sigma^+$*	*$2\sigma^-$*	*Corr. age[1]*	*$2\sigma^+$*	*$2\sigma^-$*
850	N01-1	Ojo	0.001	1.05741 ±0.0129	0.87149 ± 0.0133	7 ± 0.49	213.7	12.53	11.03	191.018	13.75	12.5
788	N1	Cristales	0.004	0.76927 ± 0.0020	0.30282 ± 0.0049	10 ± 0.17	40.071	0.82	0.81	34.544	0.82	0.81
853b	N07-10	Cristales	0.000	1.33974 ± 0.0933	0.82823 ± 0.1406	12 ± 1.85	168.838	101.14	51.8	158.526	101.64	51.96
858	ESP1-1	Espadas	0.046	2.36105 ± 0.0109	0.44666 ± 0.0773	19 ± 18.62	60.457	0.07	0.07	57.01	1.77	1.77
863b	ESP1-2	Espadas	0.163	2.97474 ± 0.0294	0.13186 ± 0.0334	29 ± 7.43	15.209	4.14	4.02	14.491	4.15	4.03
796	ESP-surf	Espadas	0.2	3.42787 ± 0.0067	0.24131 ± 0.00289	949 ± 20	7.874	0.04	0.04	7.863	0.04	0.04

[1] Correction for detrital ^{230}Th contamination, assuming "world mean" initial ^{230}Th/^{232}Th of 1.5 (Richards & Dorale, 2003)

below the surface of another broken crystal and gave an age of 34.5 ± 0.8 kyr (Lauritzen et al. 2007). The Ojo de la Reina sample is the oldest crystal analysed (191 ± 13 kyr). The depositional sequence of the Espadas speleothem shows a gypsum core 57 ± 1.7 kyr in age, covered by the aragonite layer dated to 14.5 ± 4 kyr and finally a 7.8 ± 0.04 kyr old calcite surface.

CURRENT SPELEOTHEM GROWTH AT THE -590 LEVEL

After six months gypsum crystals developed showing no trace of calcium carbonate. This demonstrates that as reported by Forti & Lo Mastro (2008), the measured increase in weight (Δ*w*) in the first four recovered tablets (covering an interval of 485 days) is in good agreement with the time of immersion that they spent inside the vessel (Fig. 5).

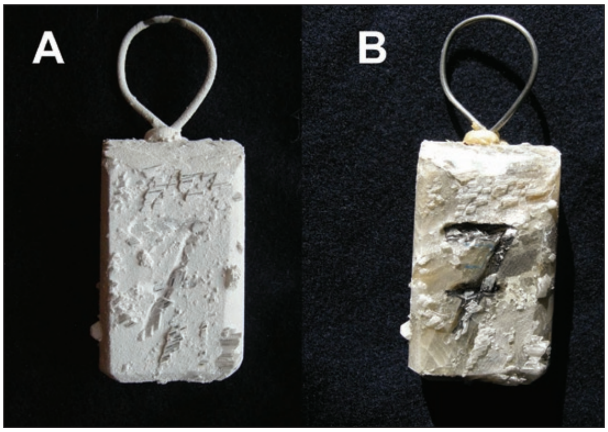

Fig. 4: a) One of the gypsum tablets recovered in February 2010: the presence of a calcite microcrystalline layer covering the whole tablet is evident; b) the same tablet after the chemical removal of the calcite.

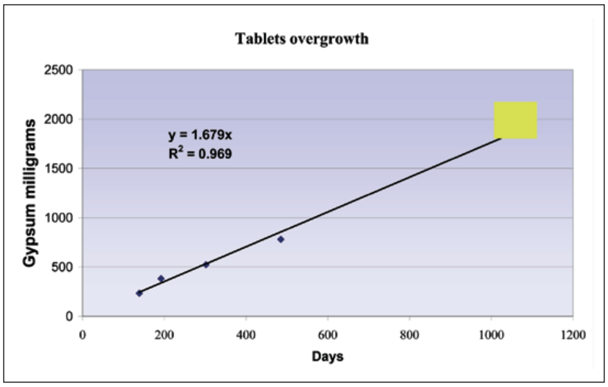

Fig. 5: Correlation between measured gypsum overgrowth and time (data from table 2). The yellow square marks the uncertainty of the data in which the vessel was open (September or December 2009). The equation and the correlation line were obtained by using only the data from the first 4 tablets.

For about two years, no scientific expedition to Naica was allowed, and in February 2010 the vessel was found open and all the remaining tablets partially covered by a thin calcite layer (Fig. 4). It is possible to state that in 2009 the laboratory at -590 Level was visited twice (in September and December, respectively) by two different filming teams: therefore the opening of the vessel surely occurred on one of these two occasions, but it has not been possible to fix the exact date the vessel was left open. The scarce amount of calcium carbonate covering the tablets suggests that the contact between the water in the vessel and the atmosphere was induced very recently, thus December 2009 is the most probable time. As expected (Forti *et al.* 2008), the contact with the atmosphere caused a rather immediate supersaturation with respect to calcium carbonate and the development of a microcrystalline crust over the tablets. Even if the termination of gypsum deposition was fast, it was not synchronous with the calcite precipitation, thus a small amount of calcium carbonate was trapped within the last accretion layers of the

Tab. 2: Growth data of gypsum tablets and the relative average deposition correction for temperature effects on the supersaturation. The tablets recovered in February 2010 stopped their growth when the vessel was left open, probably in December 2009, and the calcite trapped in the gypsum crystals slightly contributes to increase their weight.

ID Sample	Weight (g) at 18/11/06	Date	N. Days	Weight (g)	Δw	Δw/yr (g/yr)	Δh/yr (mm/yr)	mM/cm^2 s	[1]Corr. mM/cm^2 s
5	105.130	27/3/2007	139	105.364	0.234	0.614	0.040	1.71E-9	1.56E-10
2	109.969	18/5/2007	191	110.351	0.382	0.730	0.048	2.04E-9	1.85E-10
6	110.742	6/9/2007	302	111.263	0.524	0.633	0.041	1.77E-9	1.61E-10
3	107.209	28/2/2008	485	107.989	0.780	0.587	0.038	1.64E-9	1.49E-10
1	108.833	03/02/2010 (03/12/2009)	1103	*110.710	1.877	0.604	0.040	1.73E-9	1.58E-10
4	104.881	03/02/2010 (03/12/2009)	1103	*106.884	2.003	0.738	0.048	1.85E-9	1.68E-10
7	98.648	03/02/2010 (03/12/2009)	1103	*100.858	2.210	0.711	0.047	2.04E-9	1.86E-10
8	99.323	03/02/2010 (03/12/2009)	1103	*101.312	1.989	0.640	0.041	1.84E-9	1.67E-10
10	107.627	03/02/2010 (03/12/2009)	1103	*109.505	1.878	0.600	0.038	1.73E-9	1.58E-10
11	109.445	03/02/2010 (03/12/2009)	1103	*111.337	1.892	0.608	0.039	1.75E-9	1.59E-10
9	104.867	Lost	---	---	---	---	---	---	---
Average						**0.647**	**0.042**	**1.81E-9**	**1.65E-10**

[1] Correction factor for temperature lowering with respect to caves (from 57 to 51 °C): 1/11

*Weights obtained after the chemical removal of $CaCO_3$

gypsum crystals. At any rate this event was fast enough to allow the use of these six tablets also, after the chemical removal of the calcite crust, at least to validate the accretion data obtained by the first four tablets. In fact (see Fig. 5), all six obtained values are only slightly higher than those predicted by the trend line based on the first four, unperturbed, values.

The rate of growth for a giant crystal, as derived from gypsum tablets accretion data, can be calculated by converting the gypsum density in lateral accretion using the experimental average deposition (0.647 g/yr) (Tab. 2); thus the average yearly deposition corresponds to 0.0097 g/cm^2 yr, which is equivalent to an average linear growth rate of 0.042 mm/yr or 0.058 mM/cm^2 yr.

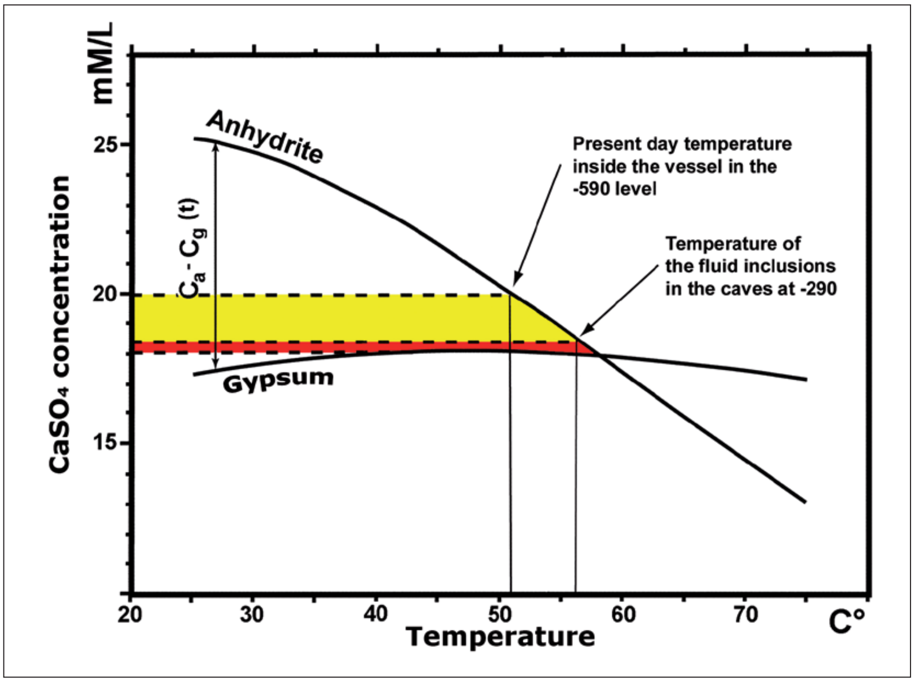

Fig. 6: Water anhydrite-gypsum solubility as a function of the temperature in the present-day sulphate concentration conditions at Naica mine: the temperature inside the vessel is lower than that which characterized the caves during growth of crystals (51 °C instead of 57 °C). This parameter affects the growth rate with a factor of about 10 (modified after Garcìa-Ruiz et al. 2007).

At any rate, this value cannot be directly used to evaluate the age of the largest crystals, because a fundamental parameter affecting the crystal growth rate has not yet been considered: the temperature inside the vessel. In fact the present temperature inside the device is lower than that which characterized the caves during the growth of their giant crystals: 51 °C in the present-day thermal spring instead of 57 °C in the past inside the caves (Garofalo *et al.* 2010). The anhydrite/gypsum solubility difference is 0.2 mM/L at 57 °C, whereas it increases to a value of 2.2 mM/L at 51 °C (Fig. 6). By taking into consideration only the temperature effect over the supersaturation on the evaluation of the real speed of growth for giant crystals (correction factor 0.2 mM/L / 2.2 mM/L = 1/11) the value of the average deposition at that time should be 5.18×10^{-3} mM/cm^2 yr or 0.0037 mm/yr, giving a growing age of the largest crystals of about 245,000 yr, which is fairly close to that obtained by the MC-ICP-MS U/Th dating, with a corresponding rate for 13 cm of accumulation between 0.001 and 0.005 mm/yr (Sanna *et al.* 2010), but in coarse agreement with the first available absolute date (600,000 years) obtained by TIMS measurements, with an average growth of 1.447 mm/kyr (Lauritzen *et al.* 2007).

However the extrapolated data from the first TIMS measurement suffer from a rather large uncertainty, because they are related to crystal growth during a period (the last 40,000 years) in which the groundwater level inside the Naica ridge started oscillating as strong differences in the physico-chemical (mainly temperature and salinity) parameters of the fluid inclusions seem to indicate (Garofalo *et al.* 2010).

Further, it must be noted here that other second order factors exist in the experimental laboratory, with respect to the natural caves, which may partially justify these differences. Significant parameters may be the ionic effect due to salinity variation, the evaporation that affects the supersaturation value (the consequence of which may influence the growth rate for no more than 3-4%) and, finally, the flow effect due to water renewal that in the vessel is about 1-2 minutes, but in the Cueva de los Cristales probably lasted from 100 to 1000 times more. The result of this last parameter has not been evaluated, but it may have affected the growth rate even more than the first two (Forti & Lo Mastro 2008). On the basis of the present available data on the crystal rate of growth, one can estimate the age of the oldest crystals in the Naica mine to converge to a value close to 300,000 years.

DISCUSSION

On the basis of the preliminary assessment of the ages of the giant selenite crystals hosted in the Naica caves, and of the growth rate evaluation from the experimental laboratory data, it is now possible to reconstruct the coarse minerogenetic sequence of gypsum deposition stages (Fig. 7). The first speleogenetic step was performed by hydrothermal fluids and was common to all the caves; later, the uplifting water became supersaturated with respect to gypsum and the nucleation of huge selenite speleothems started very slowly with minor differences among the various hypogenic cavities, depending on a relatively long time interval needed to reach a temperature range between 55-58 °C (Panieri *et al.* 2008).

At the -290 Level, this temperature range was reached some 200,000 – 250,000 years ago: in fact the appearance of gypsum crystals inside the Ojo de la Reina cave (191 ± 13 kyr) provides an indirect proof of the beginning of the crystallization, at that time, in this deep phreatic environment. Although this process was interrupted several times by the oxide-hydroxide depositions, as testified by the black layers preserved within the gypsum wall (at the cave entrance), the deposition mainly

Fig. 7: Chronology of evolution stages of gypsum deposition inside the Naica caves: a) when the temperature of the hydrothermal fluids uplifting along the three main faults lowered below 100-120 °C, hypogenic speleogenesis started at different levels inside the aquifer; b) the thermal water became supersaturated with respect to gypsum, and its nucleation took place extremely slowly in a phreatic environment at the -120 Level, where water first cooled down to the gypsum-anhydrite equilibrium temperature; c) huge selenite speleothems began to grow also at the -290 Level while fluctuation of the thermal groundwater in epiphreatic conditions produced gypsum corrosion phenomena at the -120 Level; d) a new period of phreatic gypsum nucleation was common in all caves; e) at the -120 Level further changes in the thermal watertable produced two cycles of epiphreatic phase of partial emersion with aragonite layer deposition on submerged gypsum crystals, alternated by completely saturated steps; f) a short period of meteoric seepage was active in a vadose fast-cooling environment with deposition of a thin calcite cover; g) the calcite deposition at the -120 Level stopped by the thermal groundwater level lowering while deeper caves were still in phreatic conditions; h) at the -290 Level gypsum growth went on until 1985 when a sudden change from deep phreatic to vadose conditions was caused by the depression cone associated with mine exploitation.

▶

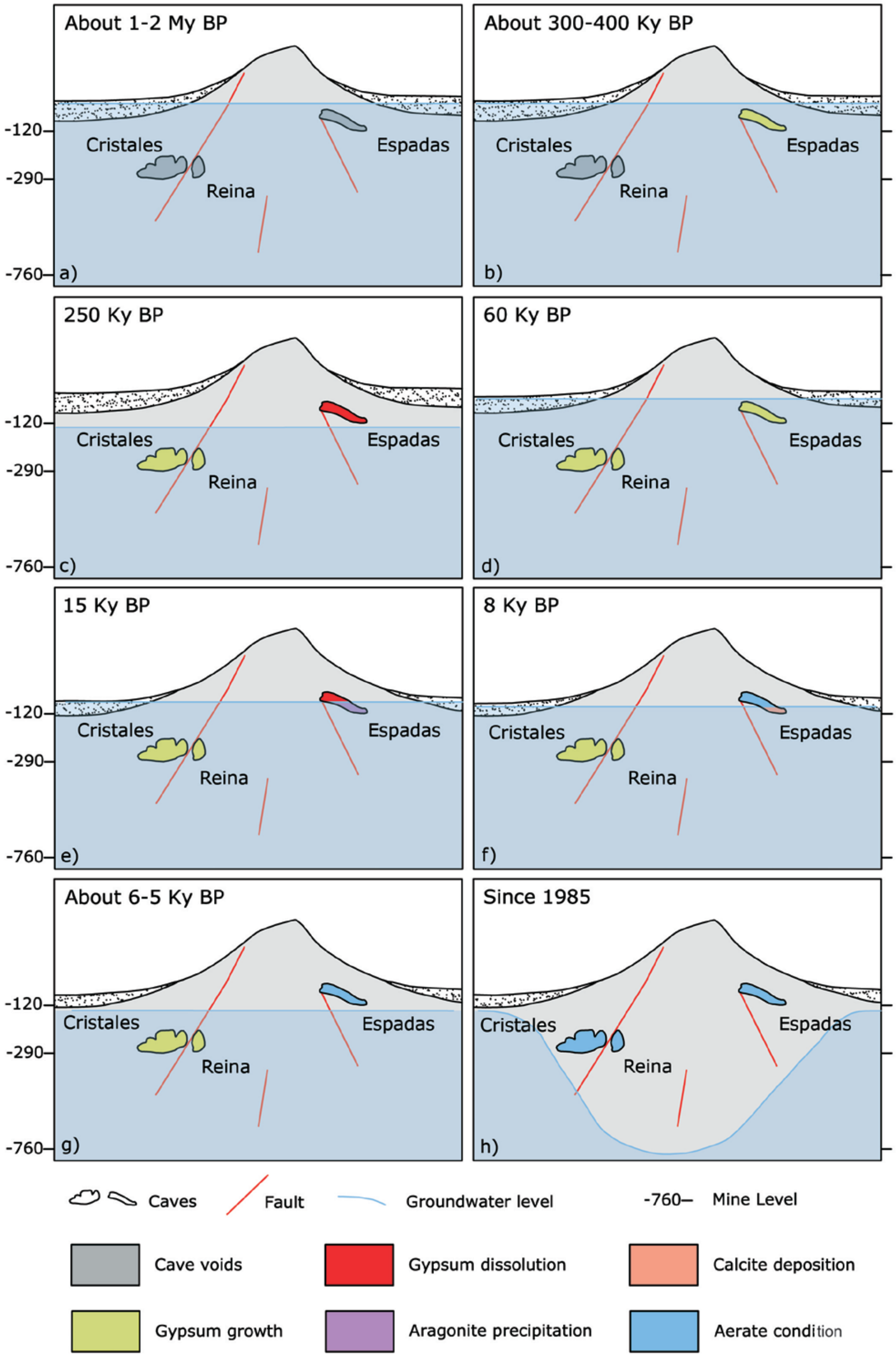
About 1-2 My BP
About 300-400 Ky BP
250 Ky BP
60 Ky BP
15 Ky BP
8 Ky BP
About 6-5 Ky BP
Since 1985
a)
b)
c)
d)
e)
f)
g)
h)
-120
-290
-760
Cristales
Reina
Espadas
Caves
Fault
Groundwater level
-760– Mine Level
Cave voids
Gypsum dissolution
Calcite deposition
Gypsum growth
Aragonite precipitation
Aerate condition

occurred under phreatic conditions until 25 years ago when dewatering abruptly lowered the watertable below the -290 Level.

In spite of the purity of giant crystals of Cueva de los Cristales, the information recorded on the naturally broken samples we collected is not so clear, due to the ambiguity of stratigraphic position along the crystal length and even more to the wide gap in age of the inner core, which is dated between 106-260 kyr. At any rate it must be noticed that the obtained age interval is in agreement with the supposed starting of the gypsum evolution within the -290 Level.

As a consequence of the uncertainty about the age of the subsample taken close to the central part of the crystal, the corresponding growth rate is not well determined, ranging between 0.5 and 1.22 mm/kyr. However the comparison of these data to that obtained from the experimental growth rate (0.4 mm/kyr) suggests that the "true" age of this sample is probably close to the oldest obtained value.

On the contrary, the growth rate of 1.45 mm/kyr, obtained by the other sample (collected within Cueva de los Cristales some 5 cm below its surface) is very different but, as pointed out before, it is related to the relatively recent stage in the development of the crystals. This suggests that although the development of the giant crystals was uninterrupted until the cave was drained, the growth was not uniform during time at least between about 50 and 35 kyr, and was affected by local factors, as suggested by the study of the fluid inclusions (Garofalo *et al.* 2010).

Large gypsum crystals developed also at the -120 Level in the Naica mine, but only a few of them are still preserved inside the Cueva de las Espadas. It was supposed that the deposition process was first active at this level, where the temperature cooled down earlier with respect to the deepest part of the system (Forti 2010; Forti & Sanna 2010), and its evolution was controlled by subsequent fluctuations in the thermal groundwater level inside the cave, which led to several changes between phreatic and vadose environments (Forti 2007). Thus, it is hypothesized that the huge crystals grew over a time of around 300-400 thousand years, but almost all of these were re-dissolved during successive stages, and thus no dating of them is currently available. On the other side, a dating sequence can be obtained from the speleothem collected at the bottom of a small dry lake in this cave. Its gypsum-aragonite-gypsum-aragonite-gypsum-calcite stratification allows reconstruction in detail of the latter steps in the evolution of the Naica thermal aquifer: after a new period of gypsum precipitation in phreatic conditions (dated back about 57 ± 1.7 kyr ago), an epiphreatic phase of partial emersion of the cave started about 14.5 ± 4 kyr ago (when the nucleation of aragonite occurred from thermal groundwater), and was followed by another period of complete saturation. All through this cycle, which was repeated two times during the cave evolution, some crystals (gypsum hooks) in the upper part of the cave were partially re-dissolved by condensation and bent by strong heating of their tips (Forti 2007) (Fig. 8).

Finally, since 7.8 ± 0.04 kyr a short period of meteoric water seepage occurred with the deposition of calcite as a consequence of the diffusion of CO_2 from the cooling atmosphere into the gypsum-saturated water, as documented by the current calcium carbonate precipitation at the -590 Level (Forti *et al.* 2008). This step was active only for a very short span of time (the thickness of the calcite layer suggests that deposition lasted no more that 1-2 thousand years) after which the seepage of meteoric water stopped completely, thus inducing the end of the calcite deposition a few thousand years before the beginning of the mine activities.

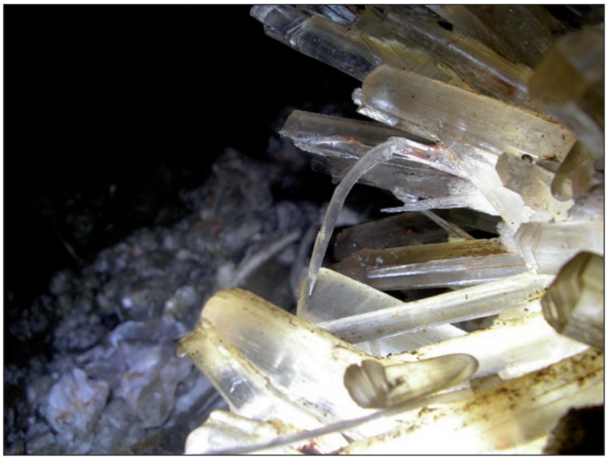

Fig. 8: Gypsum hooks from Cueva de las Espadas: re-dissolved and bent gypsum speleothems formed by condensation and strong heating of the crystal tips (Photo: Laura Sanna, La Venta & S/F Archives).

ACKNOWLEDGEMENTS

This research is a part of the "Naica Project" carried out by Speleoresearch & Films of Mexico City in co-operation with La Venta Exploring Team of Italy and partly funded by the international collaboration project "Paleogyp" (CGL2006-01707/BTE Ministry of Science and Innovation, Spain). The author L.S. benefits from a training grant supported by the European Social Funding, through the Regione Autonoma della Sardegna ("Master & Back" Program – code T2-MAB-A2008-285). The authors thank Peñoles Company for allowing access inside the Naica Mine and for their support given during the field work. We greatly appreciated the comments provided by Jo De Waele and by the two reviewers (Dino Aquilano and the anonymous referee) that have been very useful in improving a first draft of this paper.

REFERENCES

Badino, G., A. Ferreira, P. Forti, G. Giovine, I. Giulivo, G. Infante, F. Lo Mastro, L. Sanna & R. Tedeschi, 2009: The Naica caves survey.- In: White, W.B. (ed.) *Proceedings of 15th International Congress of Speleology "Karst Horizons"*, vol. III, pp. 1764-1769, Kerrville Texas-USA.

Badino, G. & P. Forti, 2005: L'eccezionale ambiente della Cueva de los Cristales, Miniera di Naica, Messico: problemi genetici ed esplorativi. [The exceptional environment of the Crystal Cave, Naica Mine, Mexico].- Memorie Istituto Italiano di Speleologia, XVII, s. II, 87-92.

Badino G. & P. Forti P, 2007: The Exploration of the Caves of the Giant Crystals (Naica, Mexico).- National Speleological Society News, 65, 2, 12-18.

Becker, J.S., 2003: Mass spectrometry of long-lived radionuclides.- Spectrochimica Acta, B, 58, 1757-1784.

Bernabei, T., G. Casagrande, A. Davila, A. De Vivo, A. Ferreira, G. Giovine, G. Infante & F. Lo Mastro, 2009: The Naica Project.- In: White, W.B. (ed.) *Proceedings of 15th International Congress of Speleology "Karst Horizons"*, vol. I, pp. 283-288, Kerrville Texas-USA.

Bernabei, T., P Forti. & R. Villasuso, 2007: A new type of gypsum speleothems from Naica (Chihuahua, Mexico): the sails.- International Journal of Speleology, 36, 1, 23-30.

Forti, P., 2006: Una foresta di cristalli di gesso nel profondo della miniera di Naica. [A forest of gypsum crystals in the depth of Naica mine].- Geoitalia 18, 29-34.

Forti, P., 2007: Studio della struttura interna di una stalagmite della Grotta delle Spade (Naica, Messico). [Study of internal structure of Sword Cave stalagmite (Naica, Mexico)].- Grotte e Dintorni, 6, 13, 3-20.

Forti, P., 2010: Genesis and evolution of the caves in the Naica Mine (Chihuahua, Mexico).- Zeitschrift für Geomorphologie, 54, sup. 2, 285-306.

Forti, P., E Galli. & A. Rossi, 2007: The mineralogical study on the Cueva de las Velas (Naica, Mexico).- Acta Carsologia, 36, 2, 75-85.

Forti, P., E Galli. & A. Rossi, 2008: Il sistema Gesso-Anidrite-Calcite: nuovi dati dalle concrezioni della miniera di Naica (Messico). [Gypsum-Anhydrite-Calcite system: new data from speleothems of the Naica mine (Mexico)].- Memorie Istituto Italiano di Speleologia XX, s. II, 139-149.

Forti, P., E Galli. & A. Rossi, 2009a: Minerogenesis in the Naica caves (Chihuahua, Messico).- In: White, W.B. (ed.), *Proceedings of 15th International Congress of Speleology "Karst Horizons"*, vol. I, pp. 300-305, Kerrville Texas-USA.

Forti, P., E Galli. & A. Rossi, 2009b: Le grotte di Naica: non solo giganteschi cristalli di gesso. [Naica caves: not only giant gypsum crystals].- Rivista Italiana di Mineralogia, 3/2009,180-196.

Forti, P. & F. Lo Mastro, 2008: Il laboratorio sperimentale di -590 nella Miniera di Naica (Chihuahua, Messico). [The experimental laboratory at -590 in the Naica Mine (Chihuahua, Mexico)].- Mondo Sotterraneo, 31, 1-2, 11-26.

Forti, P. & L. Sanna, 2010: The Naica project. A multidisciplinary study of the largest gypsum crystal of the world.- Episodes, 33, 1, 23-32.

Foshag, W., 1927: The selenite caves of Naica, Mexico.- American Mineralogist, 12, 252-232.

Gamble, V.C., 2008: *An application of U-Th dating to lacustrine systems: lake Petén-Itzà, Guatemala and lake Malawi, Malawi*.- Master thesis. University of Minnesota, pp. 98.

Garcìa-Riuz, J.M., R. Villasuso, C. Ayora, A. Canals & F. Otàlora, 2007: The formation of natural gypsum megacrystals in Naica (Mexico).- Geology, 35, 4, 327-330.

Garofalo, P.S., D. Günther, P. Forti, S.-E. Lauritzen & S. Constantin, 2007: The fluids of the giant selenite crystals of Naica (Chiuhahua, Mexico).- Proceeding of E.C.R.O.F.I., XIX European Current Research On Fluid Inclusions, 128.

Garofalo, P.S., M. Fricker, D. Günther, P. Forti, A.M. Mercuri, M. Loreti, B. Capaccioni, 2010: Climatic control on the formation of gigantic gypsum crystals within hypogenic caves (Naica mine, Mexico)?- Earth & Planetary Science Letters, 289, 3/4, 560-569.

Goede, A., R.S. Harmon, T.C. Atkinson & P.J. Rowe, 1990: Pleistocene climatic change in Southern Australia and its effect on speleothem deposition in some Nullarbor caves.- Journal of Quaternary Science, 5, 1, 29-38.

Hellstrom, J., 2003: Rapid and accurate U/Th dating using parallel ion-counting multi-collector ICP-MS.- Journal of Analytical Atomic Spectrometry, 18, 1346-1351.

Hendy, C.H., T.R. Healy, E.M. Rayner, J. Shaw & A.T. Wilson, 1979: Late Pleistocene glacial chronology of the Taylor Valley, Antarctica and the global climate.- Quaternary Research, 11, 172.

Ikeda, S. & M. Ikeya, 1991: Electron Spin Resonance (ESR) signals in natural and synthetic gypsum: an application of ESR to the age estimation of gypsum precipitates from the San Andreas Fault.- Japanese Journal of Applied Physics, 31, L136-L138.

Lauritzen, S.-E., S. Constantin & P. Forti, 2007: Chronology and growth rate of the Naica gypsum crystals.- In: Forti, P. (ed.) *Le Grotte di Naica: esplorazione, documentazione, ricerca*. pp. 49-50, Bologna.

London, D., 2003: New "Cave of the Crystals" at Naica, Chihuahua, Mexico.- Earth Science Magazine, 24-27.

Lu, Y., 2009: Accurate and precise determination of isotopic ratios by MC-ICP-MS: a review.- Mass Spectrometry Reviews, 28, 990-1011.

Luo, S.D. & T.L. Ku, 1991: U-series isochronal dating: A generalized method employing total-sample dissolution.- Geochimica et Cosmochimica Acta, 55, 555.

Mathew, G., T.K. Gundu Rao, P.S. Sohoni & R.V. Karanth, 2004: ESR dating of interfault gypsum from Katrol hill range, Kachchh, Gujarat: Implications for neotectonism.- Current Science, 87, 9, 1269-1274.

Megaw, P.K.M., J. Ruiz & S.R. Titley, 1988: High-temperature, carbonate-hosted Pb-Zn-Ag(Cu) deposits of northern Mexico.- Economic Geology, 83, 1856-1885.

Panieri, G., P. Forti, G. Gasparotto & L. Soliani, 2008: Studio delle inclusioni solide presenti nei cristalli di gesso della grotta delle Spade (Naica, Messico). [Study of solid inclusions in gypsum crystals of Sword Cave (Naica, Mexico)].- Memorie Istituto Italiano di Speleologia, XIX, s. II, 335-343.

Peng, Z., W. Liu, P. Zhang, Z. Zhang & J. Zhou, 2001: Precise timing of lacustrine gypsum in Luobubo, Xinjiang using the thermal ionization mass spectrometry U-series method.- Chinese Science Bulletin, 46, 1538-1541.

Richards, D.A. & J.A. Dorale, 2003: U-series chronology and environmental applications of speleothems.- In: Bourdon B., Henderson G.M., Ludstrom C. & Turner S. (eds.) *Uranium-series Geochemistry*. Reviews in Mineralogy and Geochemistry, 52, 407-460.

Sancho C., J.L. Peña, R. Mikkanc, C. Osácara & Y. Quinif, 2004: Morphological and speleothemic development in Brujas Cave (Southern Andean Range, Argentine): palaeoenvironmental significance.- Geomorphology 57, 367-384.

Sanna, L., F. Saez, S.L. Simonsen, S. Constantin, J.M. Calaforra, P. Forti & S.-E. Lauritzen, 2010: Uranium-series dating of gypsum speleothems: methodology and examples.- International Journal of Speleology, 39, 1, 35-46.

Shagun, C.L., 2001: Naica's subterranean marvels.- NSS News, 59, 6, 166-169.

Walker, M., 2008: *Quaternary dating methods*.- Wiley & Sons, pp. 286, Chichester UK.

Formation of natural gypsum megacrystals in Naica, Mexico

Juan Manuel García-Ruiz* Instituto Andaluz de Ciencias de la Tierra, Consejo Superior de Investigaciones Científicas-Universidad de Granada, Campus Fuentenueva, E-18002 Granada, Spain
Roberto Villasuso Compañía Peñoles, Unidad Naica, Naica, Chihuahua, Mexico
Carlos Ayora Institut de Ciències de la Terra Jaume Almera, Consejo Superior de Investigaciones Científicas, Lluis Solé Sabarís s/n, E-08028 Barcelona, Spain
Angels Canals Departament de Cristal.lografia, Mineralogia i Dipòsits Minerals, Universitat de Barcelona, Martí Franquès s/n, E-08028 Barcelona, Spain
Fermín Otálora Instituto Andaluz de Ciencias de la Tierra, Consejo Superior de Investigaciones Científicas-Universidad de Granada, Campus Fuentenueva, E-18002 Granada, Spain

ABSTRACT

Exploration in the Naica mine (Chihuahua, Mexico) recently unveiled several caves containing giant, faceted, and transparent single crystals of gypsum ($CaSO_4{\bullet}2H_2O$) as long as 11 m. These large crystals form at very low supersaturation. The problem is to explain how proper geochemical conditions can be sustained for a long time without large fluctuations that would trigger substantial nucleation. Fluid inclusion analyses show that the crystals grew from low- salinity solutions at a temperature of ~54 °C, slightly below the one at which the solubility of anhydrite equals that of gypsum. Sulfur and oxygen isotopic compositions of gypsum crystals are compatible with growth from solutions resulting from dissolution of anhydrite previously precipitated during late hydrothermal mineralization, suggesting that these megacrystals formed by a self-feeding mechanism driven by a solution-mediated, anhydrite-gypsum phase transition. Nucleation kinetics calculations based on laboratory data show that this mechanism can account for the formation of these giant crystals, yet only when operating within the very narrow range of temperature identified by our fluid inclusion study. These singular conditions create a mineral wonderland, a site of scientific interest, and an extraordinary phenomenon worthy of preservation.

Keywords: gypsum, crystallization, mineral growth, anhydrite, Naica mine, nucleation.

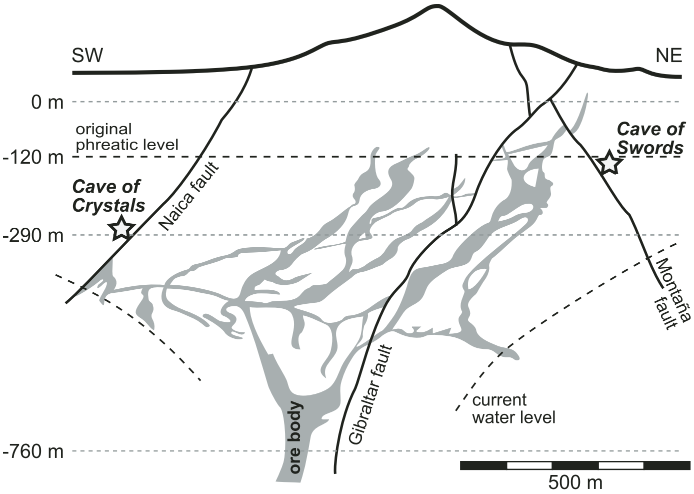

Figure 1. Cross section of Naica mine. Mine exploits hydrothermal Pb-Zn-Ag deposit with irregular manto and pipe morphologies entirely enclosed in subhorizontally dipping carbonates (not represented for simplicity). Cavities of gypsum crystals are located in carbonates close to main and secondary faults. Galleries have been excavated down to –760 m, requiring average pumping rate of 55 m^3/min to depress groundwater to –580 m with respect to phreatic level located at –120 m; Naica and Gibraltar faults act as main drains.

GEOLOGIC SETTING

The Naica mine is located on the northern side of the Sierra de Naica, 112 km SE of Chihuahua city in northern Mexico. It constitutes one of the most important lead and silver deposits in the world. The area is dominated by a 12-km-long, 7-km-wide dome structure oriented NW-SE, standing out from an extensive alluvial plain. The Naica mining district is entirely formed by sedimentary rocks, consisting of a sequence of Albian limestones with some lutitic interbeddings overlying an Aptian evaporitic sequence that has been found in the surrounding region, but not in the mining district. The sedimentary rocks are intruded by Tertiary acidic dikes dated as 26.2 and 25.9 Ma (Megaw et al., 1988). An igneous intrusion 4 km south of the Naica mine at a depth between 2 and 5 km was revealed by magnetometric studies.

The hydrothermal fluid circulation associated with the Tertiary dikes formed the Ag-Pb-Zn deposits (Lang, 1995). These deposits display chimney and manto shapes and developed in the dikes and in the enclosing carbonates, which appear to be intensively transformed to calc-silicates. A set of faults acts as the main structural control on the mineralization; Gibraltar, Montaña, and Naica are the most important (Fig. 1). The ore assemblage comprises pyrite-pyrrhotite, sphalerite, galena, and chalcopyrite, and formed from high-temperature hypersaline brines, according to published fluid inclusion data (Erwood et al., 1979; Lang, 1995). During the late hydrothermal stage, sulfuric acid formed by oxidation of the underlying sulfides, and this reacted with limestone to form calcium-sulfate–rich waters that precipitated the anhydrite. Anhydrite occurs massively in Naica at all levels below –240 m. Late quartz-calcite-anhydrite veins cut the ore (Stone, 1959).

GIANT SELENITE CRYSTALS

Since early in the twentieth century, the excavation of new galleries and tunnels at the Naica mine has occasionally led to the discovery of cavities and caves containing meter-sized crystals of selenite, the colorless crystalline variety of gypsum. The caves are always located near fractures and their walls are covered by a red coating of calcite and celestite with minor amounts of iron oxide, montmorillonite, chlorite, and illite. Until 2000, the most famous of these caves was the Cueva de las Espadas (Cave of Swords), so named because of meter-sized selenite crystals (described by Foshag, 1927); that year several caves were discovered at the –290 m level in secondary faults associated with the Naica fault (Fig. 1). One of these caves, the Cueva de los Cristales (Cave of the Crystals), contains selenite crystals as long as 11 m and as thick as 1 m, much larger than any gypsum crystal previously reported (Foshag, 1927; Palache, 1932; Rickwood, 1981; García-Guinea et al., 2002). In addition, at that depth the impact of muddy, oxygen-enriched waters is very unlikely, and consequently the gypsum crystals contain very few solid inclusions of iron and

---

*E-mail: jmgruiz@ugr.es.

© 2007 The Geological Society of America. For permission to copy, contact Copyright Permissions, GSA, or editing@geosociety.org.
Geology, April 2007; v. 35; no. 4; p. 327–330; doi: 10.1130/G23393A.1; 4 figures; Data Repository item 2007080.

manganese oxides, and are almost transparent, increasing their beauty and singularity. While several hypotheses have been suggested for the origin of these crystals, to date none of them has been thoroughly investigated (Marín Herrera et al., 2006).

The walls of the Cueva de los Cristales and particularly the floor are sprinkled with blocky single crystals that in some cases cluster to form parallel aggregates (see GSA Data Repository Fig. DR6[1]). Rather than having the characteristic platy or tabular habit of gypsum crystals, they are euhedral, displaying well-developed {010}, {1k0}, and {–111} forms (Fig. DR2; see footnote 1). The {010} pinacoid is less developed than in the usual growth morphology of gypsum crystals. The {1k0} prism appears as striated surfaces made of segments of prismatic forms with k = 2, 4, or 6, as shown by laser reflection goniometry and scanning electron microscopy (Figs. DR3 and DR4 and Table DR1). The resulting nonsingular face has an orientation close to {140}.

Giant elongate selenite crystals grow from some of these groups of blocky crystals or directly from the floor, and some cross the cave from side to side (Fig. 2). Most of these crystals are 6 m long and several of them reach 10 m. Consequently, the miners correctly term them vigas, or beams, instead of swords. In addition to the aspect ratio, there is only one morphological difference between the beams and the blocky crystals, i.e., the {010} pinacoid faces of beams disappear in most cases. The crystalline beams are elongated in the direction of the c axis, defined by huge prismatic {1k0} faces and capped by {–111} prisms (Fig. DR5). The density of crystals in the cave is very small (Fig. 2), creating a natural scenario of unparalleled beauty, and is critical to explain its formation, because this can only result from an extremely low nucleation rate for the entire history of the crystal growth. This must be the first boundary condition for any genetic model.

GEOCHEMICAL CONDITIONS

Gypsum is slightly soluble in water (13.78 × 10^{-3} mol L^{-1} at 25 °C and atmospheric pressure) (Partridge and White, 1929). Its solubility shows a positive dependence on temperature in cool waters, reaching a maximum at ~58 °C and then decreasing (Posnjak, 1938). The solubility of gypsum also has a positive variation with salinity, but the temperature of maximum solubility near 58 °C does not change significantly (see the section "Nucleation and Growth of the Megacrystals," below). At atmospheric pressure, anhydrite is the stable thermodynamic phase at temperatures hotter than 56–58 °C (Hardie, 1967; Blount and Dickson, 1973), but below this temperature anhydrite becomes unstable with respect to gypsum.

The mine area is still a thermal anomaly: water samples collected at different times and mine locations show that the groundwater is ~53 °C in the deep levels of the mine (Table DR4; see footnote 1), although temperatures as high as 59 °C have been measured by miners. Cooling of the system and contribution of meteoric waters yield sulfate- and carbonate-rich water with a temperature in the range of 48–59 °C, and a neutral to slightly basic pH (7.0–7.8), like the one today. According to their oxygen and deuterium isotopic values (Table DR3), present-day waters have a meteoric origin with more than 50 yr of residence time in the aquifer, as indicated by the lack of tritium.

Upon cooling of the Naica district, hydrothermal anhydrite produced during the late stage of the formation of the ore deposit became unstable and dissolved, thus explaining the high content of sulfate anions in the present-day waters. The dissolution of anhydrite below 58 °C supplies Ca and SO_4, thus keeping the water slightly supersaturated in gypsum and promoting its crystallization. Examples of this transition are also observed at a microscopic scale in some anhydrite masses close to the Cave of the Crystals (Fig. DR11). Supersaturation values ($S = c/c_e$, where c is the actual concentration and c_e is the equilibrium concentration for a given temperature and salinity) calculated with the speciation code PHREEQC (Parkhurst and Appelo, 2003) for present-day waters yield values slightly supersaturated for gypsum and undersaturated for anhydrite. Such a small difference between the solubility of gypsum and anhydrite slightly below 58 °C is therefore a continuous source of calcium and sulfate ions in the Naica area and provides a very stable mechanism to constantly maintain the crystallizing system close to equilibrium. This is the mechanism we propose for the formation of the selenite megacrystals of Naica. To check this hypothesis, geochemical evidence is required to prove that selenite crystals grew from low-salinity solutions and that these solutions were produced by dissolution of anhydrite.

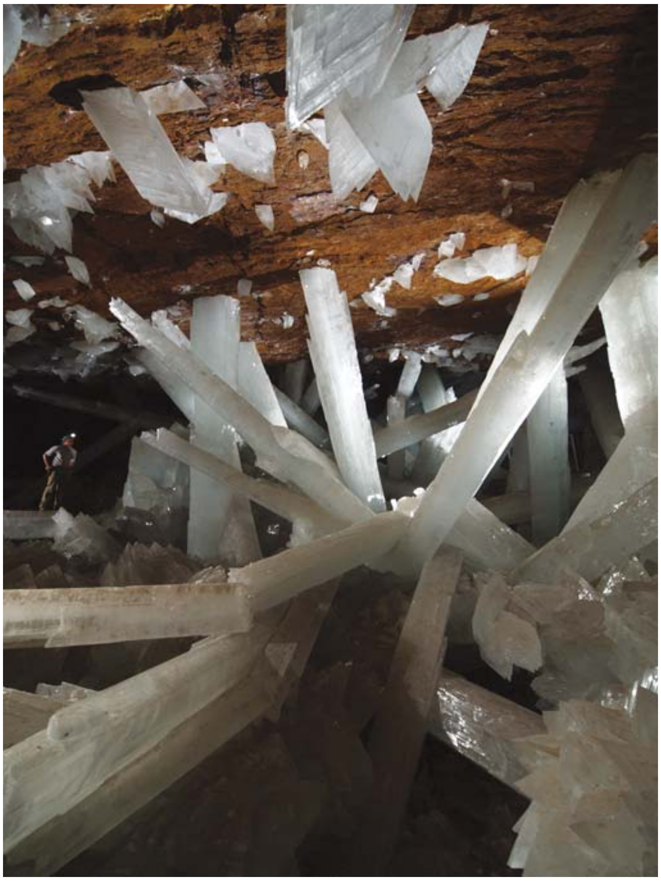

Figure 2. View of Cave of Crystals showing characteristic moonlight luster for which gypsum crystals are named (i.e., selenite). Two crystal morphologies include blocky crystals displaying prisms {1k0} and {–111}, and pinacoid {010} and beams, which are crystals elongated along c axis displaying only prismatic crystal faces. Note that population of crystals is much lower in ceiling of cave than on floor. For very long waiting time for nucleation, this gradation is due to a density layering of the solution filling cave (Turner, 1985), which increases the probability of nucleation from the solution located in lower part of cavern.

The salinity of the liquid from which selenite crystals grew was characterized by measuring the melting temperatures of ice in 33 selected fluid inclusions that contained liquid and vapor phases at room temperature (Fig. DR9 and Table DR2). The gypsum crystals contain very scarce fluid inclusions. Primary fluid inclusions parallel to crystal faces forming growth bands were selected (Fig. DR7). Measured temperatures of ice melting show a maximum at –0.1 °C and a lower limit at –1.1 °C. This range of temperatures indicates a low-salinity fluid (0.17 mass% equivalent NaCl typical and 1.9 mass% equivalent NaCl maximum) compatible with today's meteoric and mine waters (Bodnar, 1993; Bakker, 2003). That means that the original high-salinity, sulfur-rich, acidic magmatic fluids (Lang, 1995; Erwood et al., 1979) that caused the skarn formation had practically lost their salinity by the time the gypsum crystals grew. The dispersion of ice melting temperatures within an inclusion band is the same order of magnitude as the variations along the whole set of samples, and therefore no significant salinity variation with crystal growth can be established.

---

[1] GSA Data Repository item 2007080, Figures DR1–DR11, and Tables DR1 and DR2 (data sets, photos, and experimental details), is available online at www.geosociety.org/pubs/ft2007.htm, or on request from editing@geosociety.org or Documents Secretary, GSA, P.O. Box 9140, Boulder, CO 80301, USA.

To check for the origin of the solutions, several samples of anhydrite and gypsum from different locations in the mine were collected for stable isotope analysis. The $\delta^{34}S$ and $\delta^{18}O$ values of two gypsum samples from the Cave of the Crystals show no variation, despite the differing habit of the crystals (beams or blocky), with a $\delta^{34}S$ of 16.9‰ and $\delta^{18}O$ of 17.6‰. As shown in Figure 3, the isotopic values of these gypsum crystals are consistent with their formation from waters with dissolved sulfate similar to those found in present-day groundwater collected at the –530 level. The isotopic values of sulfate dissolved in the water can be explained by mixing of sulfates from the dissolution of anhydrite at different locations of the mine: those from the ore deposit area (samples 23, 24, 25, and 34 in Table DR3) and those from upper levels in the mine (samples 04 and 21). Furthermore, these values, along with the absence of alunite and kaolinite minerals, eliminate a sulfuric speleogenesis mechanism (Polyak and Güven, 1996).

Additional evidence for the dissolution of anhydrite is the presence of celestite, a strontium sulfate found with clays and iron oxide in a thin layer coating the walls of the Cave of the Crystals. It has been demonstrated that the anhydrite structure can host more strontium than gypsum structure (Butler, 1973). Therefore, the celestite crystals likely result from the excess of strontium hosted in the dissolving anhydrite with respect to crystallizing gypsum (Table DR4).

NUCLEATION AND GROWTH OF THE MEGACRYSTALS

Geochemical evidence shows that close-to-equilibrium dissolution of anhydrite and crystallization of gypsum under the conditions of the Cave of Crystals can explain the origin of these giant crystals. The question remains if these values can also explain the low number of crystals observed, and as a consequence, the final size reached by each of them. The question is therefore how much cooling the system may have undergone to avoid the massive nucleation that would have packed the caves with a large number of smaller crystals. Classical crystal growth theory shows that the nucleation rate, J, is described by the equation:

$$J = \mathrm{A}exp - \frac{16\pi\gamma^3\nu^2}{3\mathrm{k}^3T^3ln^2S}, \qquad (1)$$

where γ is the crystal-liquid interfacial tension, ν is the molecular volume, k is the Boltzmann constant, T is the absolute temperature, and A is a pre-exponential factor that can be considered constant in this discussion (Mullin, 1993). As the nucleation rate J varies with the second power of the logarithm of supersaturation S, low values of J mean that supersaturation was kept in the metastable zone and very close to equilibrium for the history of crystallization in the cave. Because the induction time t_{ind} (time to induce formation of detectable crystals) for nucleation is inversely proportional to J, the equation can be written as a relationship:

$$ln\ t_{ind} \propto \frac{\gamma^3}{T^3 \log^2 S}. \qquad (2)$$

Using experimental data available from gypsum nucleation studies (He et al., 1994; Lancia et al., 1999), we calculated (for both homogeneous and heterogeneous nucleation) the induction time for three salinities: calcium-sulfate and water molecules (labeled 0), salinity values (in terms of Na, Cl, and Mg concentrations) identical to those found in Naica waters (labeled 1), and twice these concentrations (labeled 2). Solubility values calculated with PHREEQC (Parkhurst and Appelo, 2003) were used to compute the supersaturation values and nucleation induction times plotted in Figure 4B. For any geological reservoir where crystallization occurs, heterogeneous nucleation is unavoidable because it requires lower critical supersaturation values to surpass the nucleation barrier. As shown in Figure 4B, the induction time for heterogeneous nucleation is thousands of years for temperatures of 47–48 °C. For any temperature below 44–45 °C, we found a very large probability for nucleation within a time scale of months or days, inconsistent with the small number of crystals found in the caves.

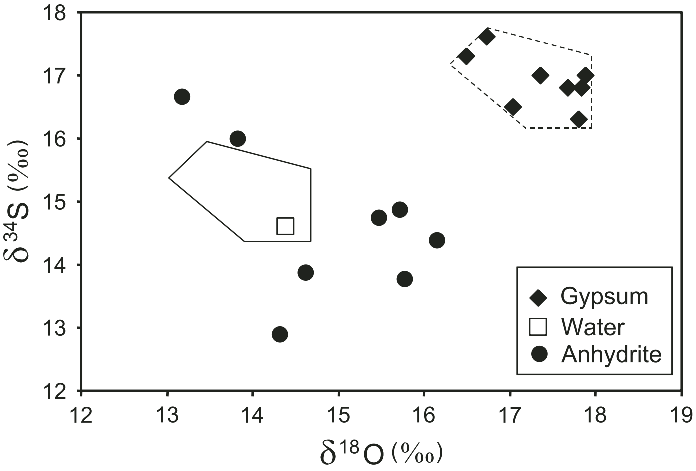

Figure 3. Isotope composition of sulfate molecules in gypsum, anhydrite, and water collected at Naica mine. Polygonal surface enveloping the isotope composition of water from which the gypsum crystals grew is obtained by subtracting from the gypsum values the enrichment factor of 1.65‰ for sulfur and 3.3‰ for oxygen (Thode and Monster, 1965).

To support the proposed mechanism, it is mandatory that the temperature during crystal growth is within the given range. We have used fluid inclusions to check this. Homogenization temperature (T_h) data from 31 fluid inclusions show values between 46 °C and 60 °C with a maximum of ~54 ± 1 °C (Fig. 4C; Table DR2 [see footnote 1]). According to the proposed mechanism, this implies that most of the crystallization process took place at a temperature for which t_{ind} is longer than 1 m.y. (Fig. 4B). These close-to-equilibrium conditions were evidently maintained for a long time with some episodes of cessation of growth, as suggested by the occasional solid inclusions of manganese and iron oxihydroxides on {1k0} and {–111} growth fronts. Eventually, as indicated by lower values of T_h down to 47 °C, the system cooled due to a stronger influence of meteoric waters; the typical time for nucleation was thousands of years. It follows from this mechanism that caves such as the Cave of the Swords (Foshag, 1927) that are located closer to the surface had lower temperatures and are therefore filled with large numbers of smaller crystals (Fig. 1; Fig. DR1).

CONCLUSION

We offer an original explanation for the formation of large selenite crystals based on a self-feeding mechanism driven by solution-mediated anhydrite-gypsum phase transition at a temperature slightly below the gypsum-anhydrite equilibrium temperature. This explanation is supported by the following evidence. (1) Anhydrite was readily available from the late hydrothermal stage after the ore mineralization. (2) Temperature during the growth of the crystals was kept slightly below 58 °C, the value at

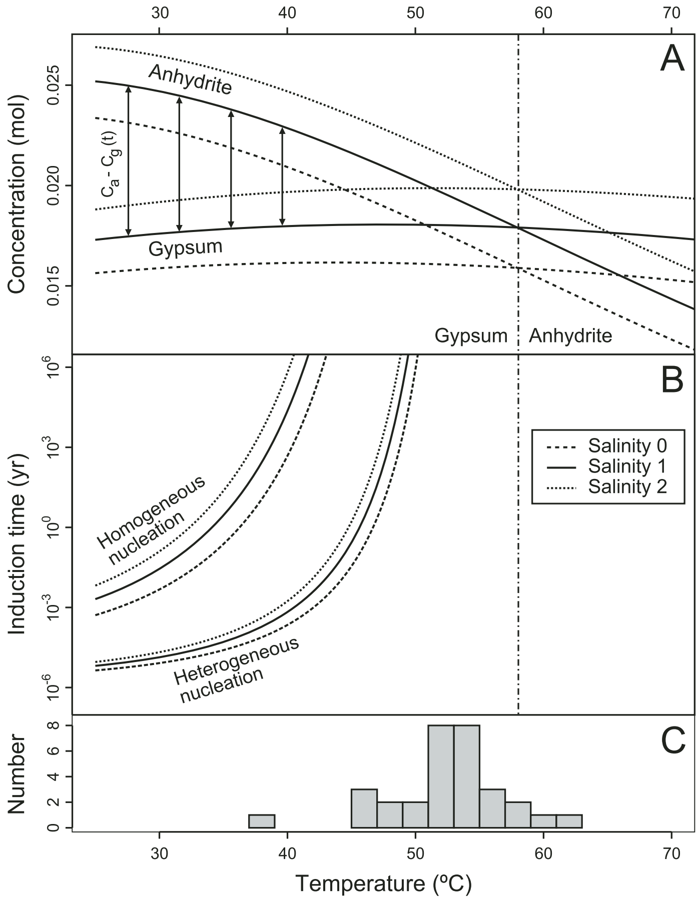

Figure 4. A: Variation of anhydrite and gypsum solubility with temperature for three different values of salinity: (0) for pure calcium sulfate solutions; (1) for salinity equal to and (2) twice that found in present-day water collected at Naica mine. B: Induction time for homogeneous and heterogeneous nucleation calculated for supersaturated solutions forming by differences in solubility between anhydrite and gypsum at different temperatures. C: Homogenization temperatures of 31 fluid inclusions showing actual temperature of growth.

which the solubility of anhydrite equals that of gypsum. (3) Giant selenite crystals grew from low-salinity solutions with isotopic compositions compatible with those formed by dissolving the anhydrite found in the mine. (4) Gypsum nucleation kinetics imply induction times longer than 1 m.y. for the typical temperature (54 °C) and ~1 k.y. for the lower-temperature episodes (down to 47 °C). This mechanism provides a level of supersaturation that is not only small and sustained over time, but is also virtually free of fluctuations (even small amplitude ones).

It can be predicted that other caves with similar or even larger selenite crystals exist among the tangle of underground galleries in the Naica mine area. They may constitute one of the oddest locations of the mineral world and are worthy of preservation; however, there are two problems that can be envisaged for that task. The cave must be kept wet and hot, but equally important is the development of a method to maintain the integrity of the long beams that bend and break once the buoyancy support of the cave water is lost. The second problem is in the long term but more critical: unless water pumping continues after mining, the caves will be again flooded to their natural state. According to the proposed mechanism, the selenite crystals may well start to grow again, but they will be hidden for future generations.

ACKNOWLEDGMENTS

We gratefully acknowledge Compañía Peñoles for the facilities provided during the field studies performed in the Naica mine, and the Ministerio de Educación y Ciencia of Spain for financial support.

REFERENCES CITED

Bakker, R.J., 2003, Package FLUIDS 1. New computer programs for the analysis of fluid inclusion data and for modelling bulk fluid properties: Chemical Geology, v. 194, p. 3–23, doi: 10.1016/S0009–2541(02)00268–1.

Blount, C.W., and Dickson, F.W., 1973, Gypsum-anhydrite equilibria in systems $CaSO_4$–H_2O and $CaSO_4$-NaCl-H_2O: American Mineralogist, v. 58, p. 323–331.

Bodnar, R.J., 1993, Revised equation and table for determining the freezing point depression of H_2O-NaCl solutions: Geochimica et Cosmochimica Acta, v. 57, p. 683–684, doi: 10.1016/0016–7037(93)90378-A.

Butler, G.P., 1973, Strontium geochemistry of modern and ancient calcium sulphate minerals, *in* Purser, B.H., ed., The Persian Gulf: Berlin, Springer-Verlag, p. 423–452.

Erwood, R.J., Kesler, S.E., and Cloke, P.L., 1979, Compositionally distinct, saline hydrothermal solutions, Naica Mine, Chihuahua, Mexico: Economic Geology and Bulletin of the Society of Economic Geologists, v. 74, p. 95–108.

Foshag, W., 1927, The selenite caves of Naica, Mexico: American Mineralogist, v. 12, p. 252–256.

García-Guinea, J., Morales, S., Delgado, A., Recio, C., and Calahorra, J.M., 2002, Formation of gigantic gypsum crystals: Geological Society [London] Journal, v. 159, p. 347–350.

Hardie, L.A., 1967, The gypsum-anhydrite equilibrium at one atmosphere pressure: American Mineralogist, v. 52, p. 171–200.

He, S., Oddo, J.E., and Tomson, M.B., 1994, The nucleation kinetics of calcium sulfate dihydrate in NaCl solutions up to 6 m and 90°C: Journal of Colloid and Interface Science, v. 162, p. 297–303, doi: 10.1006/jcis.1994.1042.

Lancia, A., Musmarra, D., and Prisciandaro, M., 1999, Measurement of the induction period for calcium sulfate dihydrate precipitation: American Institute of Chemical Engineers Journal, v. 45, p. 390–397.

Lang, J.R., 1995, A geological evaluation of the Naica deposit, Chihuahua, Mexico: Internal Report of Compañia Fresnillo, Mexico, 109 p.

Marín Herrera, B.R., Vogel González, F., and Echegoyén Guzmán, R., 2006, Las megaselenitas del distrito minero de Naica, Chihuahua, una ocurrencia mineralógica anómala: Boletín de Mineralogía, v. 17, p. 139–148.

Megaw, P.K.M., Ruiz, J., and Titley, S.R., 1988, High-temperature, carbonate-hosted Pb-Zn-Ag (Cu) deposits of northern Mexico: Economic Geology and Bulletin of the Society of Economic Geologists, v. 83, p. 1856–1885.

Mullin, J.W., 1993, Crystallization (third edition): Oxford, Butterworth-Heinemann, 527 p.

Palache, C., 1932, The largest crystals: American Mineralogist, v. 17, p. 362–363.

Parkhurst, D.L., and Appelo, C.A.J., 2003, PHREEQC—A hydrogeochemical transport model: http://wwwbrr.cr.usgs.gov/projects/GWC_coupled/phreeqc/index.html (September 2006).

Partridge, E.P., and White, A.H., 1929, The solubility of calcium sulfate from 0 to 200°: American Chemical Society Journal, v. 51, p. 360–370, doi: 10.1021/ja01377a003.

Polyak, V.J., and Güven, N., 1996, Alunite, natroalunite and hydrated halloysite in Carlsbad Cavern and Lechuguilla Cave, New Mexico: Clays and Clay Minerals, v. 44, p. 843–850, doi: 10.1346/CCMN.1996.0440616.

Posnjak, E., 1938, The system $CaSO_4$-H_2O: American Journal of Science, v. 235A, p. 247–272.

Rickwood, P.C., 1981, The largest crystals: American Mineralogist, v. 66, p. 885–907.

Stone, J.G., 1959, Ore genesis in the Naica District, Chihuahua, Mexico: Economic Geology and Bulletin of the Society of Economic Geologists, v. 54, p. 1002–1034.

Thode, H.G., and Monster, J., 1965, Sulfur-isotope geochemistry of petroleum, evaporites, and ancient seas, in Young, A., and Galley, J.E., eds., Fluids in subsurface environments: American Association of Petroleum Geologists Memoir 4, p. 367–377.

Turner, J.S., 1985, Multicomponent convection: Annual Review of Fluid Mechanics, v. 17, p. 11–44, doi: 10.1146/annurev.fl.17.010185.000303.

Manuscript received 25 September 2006
Revised manuscript received 3 December 2006
Manuscript accepted 7 December 2006

Printed in USA

Unusual Places

NORTH AMERICA EUROPE AFRICA ASIA AUSTRALIA MIDDLE EAST SOUTH AMERICA

Search the site Search

Home > Mexico > North America > The Giant Crystal Cave: a Mexican Wonder Underground

- Contact
- Submit
- DMCA Policy

The Giant Crystal Cave: a Mexican Wonder Underground

You know your business. We know construction software.
Let's talk.

Software Advice

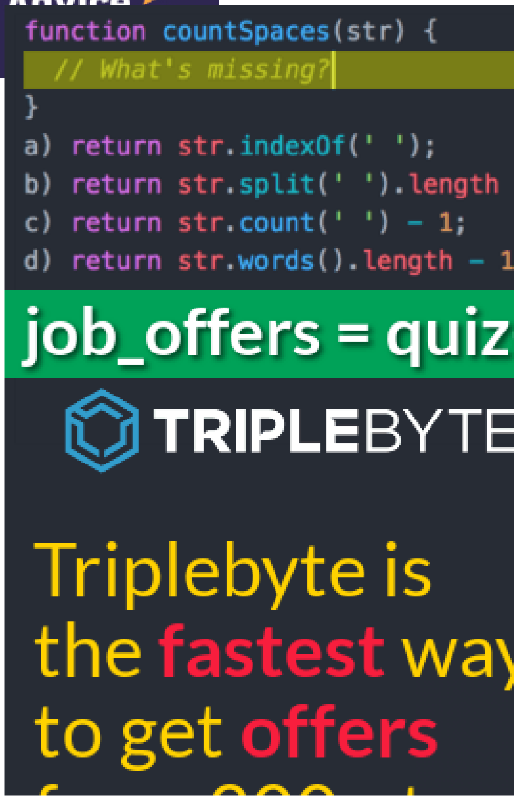

```
function countSpaces(str) {
  // What's missing?
}
a) return str.indexOf(' ');
b) return str.split(' ').length
c) return str.count(' ') - 1;
d) return str.words().length - 1
```

job_offers = quiz

TRIPLEBYTE

Triplebyte is the fastest wa
to get offers

The city of Chihuahua, Mexico, hides one of the world's greatest treasures, which became a landmark for the location. In the Naica mine dozens of large crystals have been lying for over a century. Since most of the crystals weight over 55 tons, they hold a record for the largest selenite crystals in the world. Even though it's been over 100 years since the discovery, the seemingly surreal location kept its title.

Back in the 18th century, when silver was worth a lot in Mexico, the Naica mine was discovered, when workers struck a silver vein. Later on, at the beginning of the 20th century, large scale mining began. Just before the Naica mine was officially closed in 1911, the Giant Crystal Cave was discovered. Because of it spectacular natural location, locals decided to keep it safe from being destroyed, so the cave is protected by law.

However, it wasn't until 2000 that two researchers, Juan and Pedro Sanchez, rediscovered the cave again while searching the Naica cave; and what they saw was unforgettable. Over 900 feet underground, lies a large chamber filled with dozens of giant crystals.

Each crystal has a diameter of at least 12 feet and weights over 55 tons. Apparently, the crystals created due to the magma underground and are over 500,000 years old. In time, natural phenomena shaped them uniquely and the result is definitely amazing to anybody who gets to see them.

While the cave provides anyone with a once-in-a-lifetime view, staying there is nearly impossible without the proper equipment. The average temperature in the main chamber is 58 degrees Celsius(136.4), which may cause death for anyone, who stays there longer than 10 minutes. On top of that, climbing the crystals is quite difficult, which is why the cave can be visited only with specialists nearby.

The legend says that selenite, which is gypsum crystal, provides people with healing benefits and makes people, who touch them look younger. The Greek goddess of the moon, named Selenite, offered luck and emotional stability, which is why the Giant Crystal Cave is thought to be not only a spectacular, unique place, but also a magical one.

Submitted by Hamid Bagha. Thank you

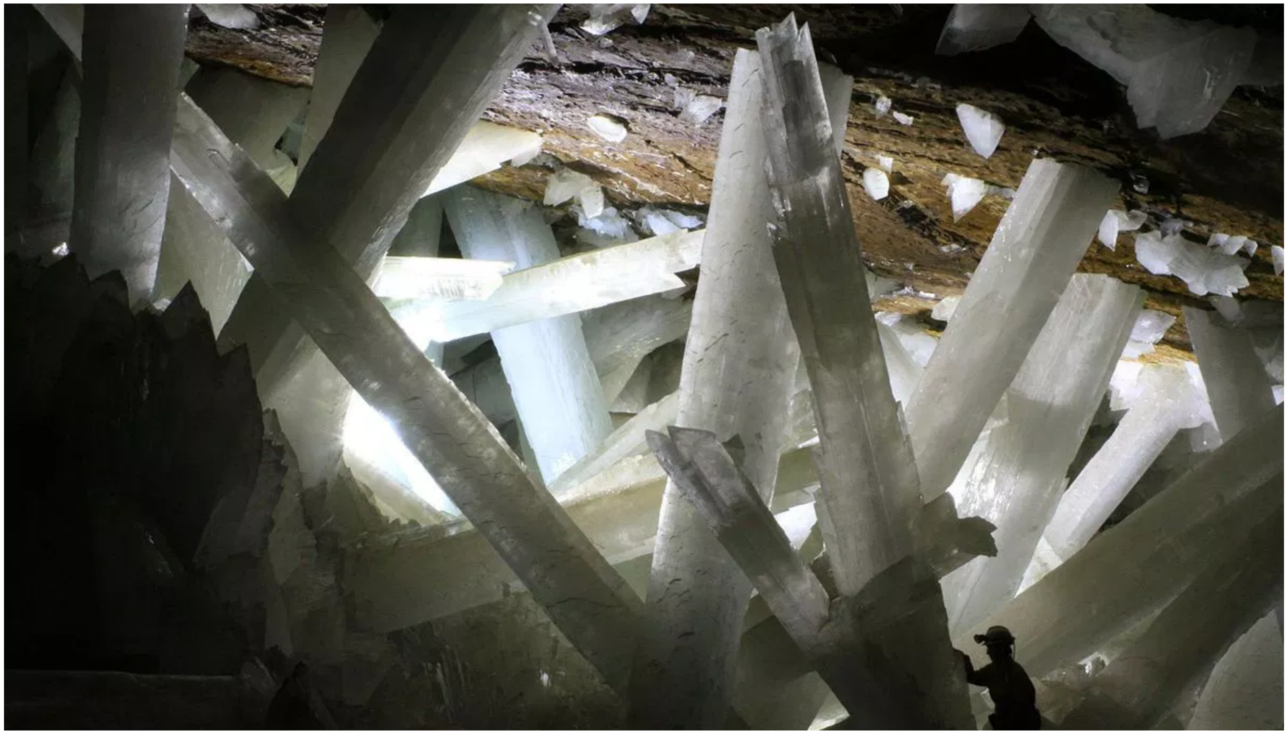

from 200+ top
tech companie

Take the quiz.

Follow Us

Subscribe via Email

Enter your email address to subscribe to this blog and receive notifications of new posts by email.

Join 339 other subscribers

Email Address

Subscribe

Image credit: Wikipedia

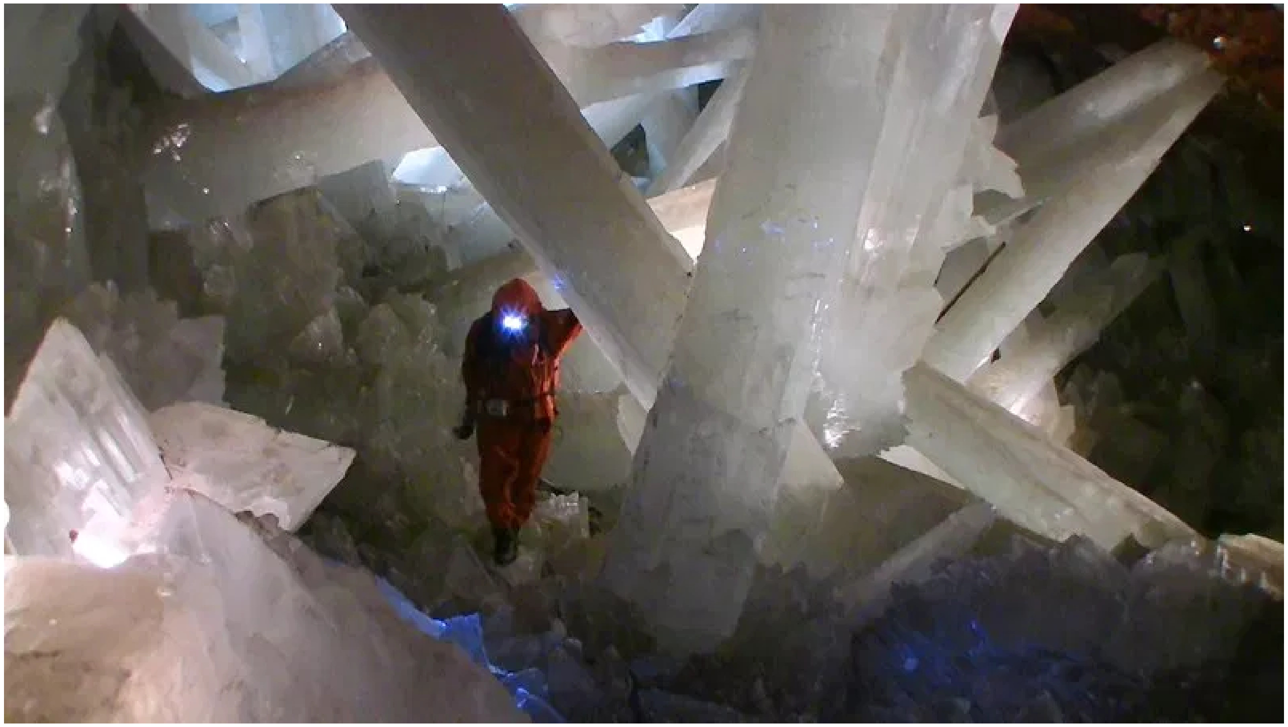

Image credit: stormchaser.ca

Image credit: stormchaser.ca

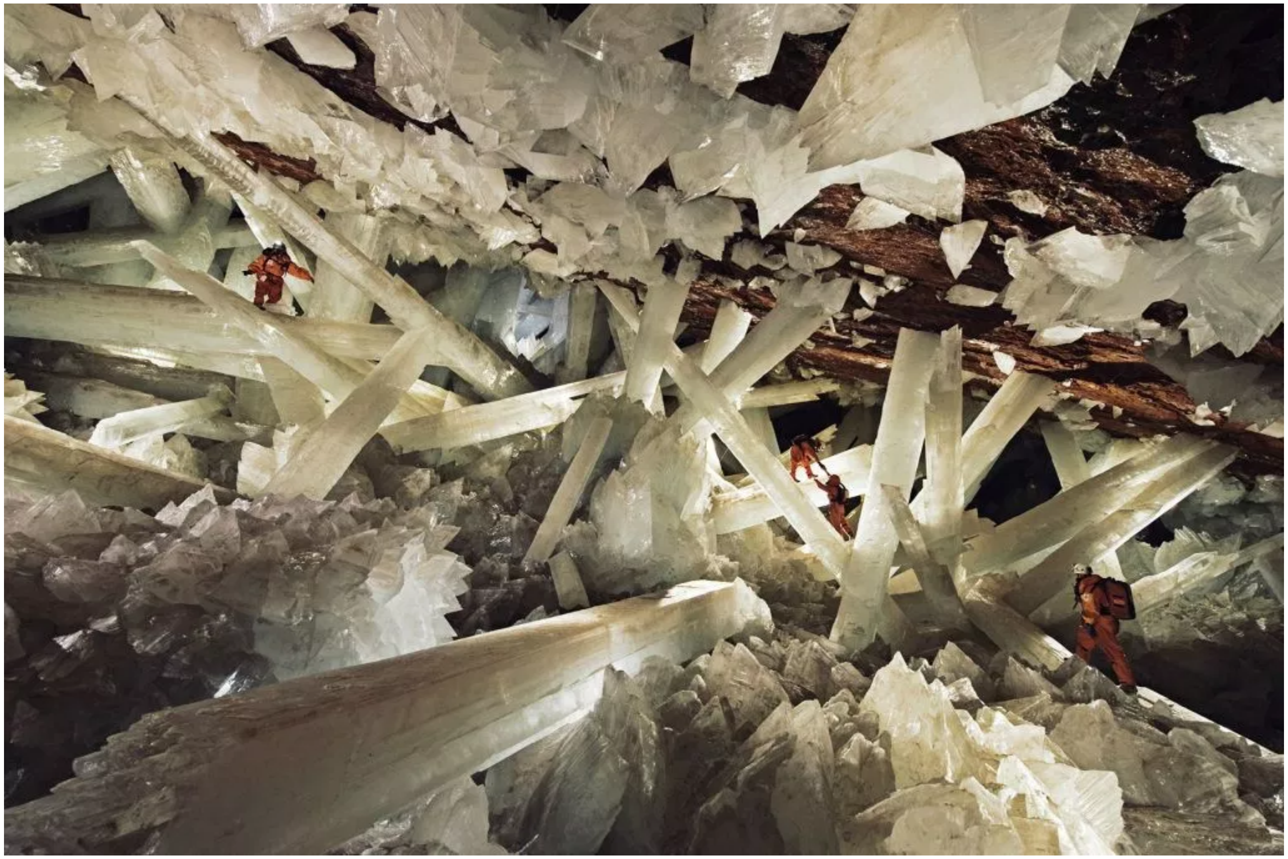

Image credit: nationalgeographic.com

The Choccy Block Crystal Set Radio Kit

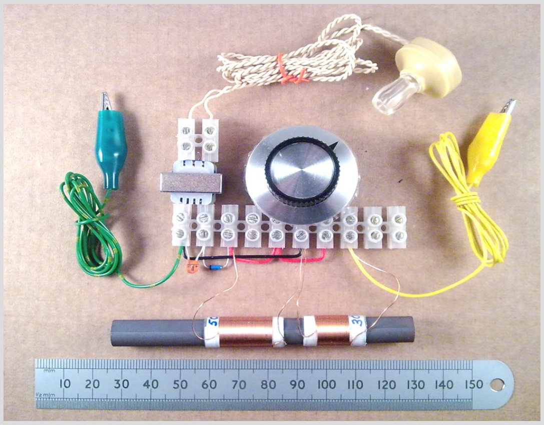
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150

I've sold quite a few of these now and people are having success with them. If you're thinking of buying one of the commercial plastic sets without an aerial coupling coil, output transformer and with an untested earphone, please consider this one instead. I can confirm that it is much better and you won't risk the 40% chance of getting a dud earphone that are the current odds. (See The Problem With Crystal Earphones)

A kit of electronic parts is available on Ebay here: Useful Components Ebay Shop

This is a small crystal radio which can be constructed without soldering using 3A terminal strip or "Choccy Block." It's best to read right through the page first before starting out.

How To Build The Radio

Here is a pdf electronic schematic diagram. If you're familiar with electronics this will be helpful. If not, you can just look at the pictures. In addition to the parts provided in the picture you will need a screwdriver for the terminal block, some pliers, some wire cutters and some form of wire stripper. To make the paper tubes for the coils you will need some paper, glue, scissors and sticky tape.

Schematic Diagram

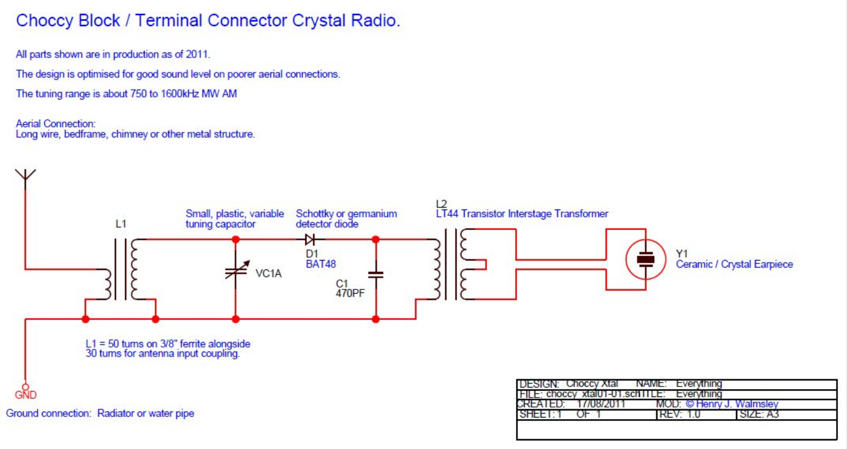
Choccy Block / Terminal Connector Crystal Radio.

All parts shown are in production as of 2011.

The design is optimised for good sound level on poorer aerial connections.

The tuning range is about 750 to 1600kHz MW AM

Aerial Connection:
Long wire, bedframe, chimney or other metal structure.

L1 = 50 turns on 3/8" ferrite alongside 30 turns for antenna input coupling.

Ground connection: Radiator or water pipe

DESIGN: Choccy Xtal NAME: Everything
FILE: choccy xtal01-01.sch TITLE: Everything
CREATED: 17/08/2011 MOD: © Henry J. Walmsley
SHEET: 1 OF 1 REV: 1.0 SIZE: A3

L1
Small, plastic, variable tuning capacitor
VC1A
Schottky or germanium detector diode
D1
BAT48
C1
470PF
L2
LT44 Transistor Interstage Transformer
Y1
Ceramic / Crystal Earpiece
GND

Link to Schematic Diagram of Choccy Block Crystal Set Radio

Bill Of Materials:

* 3 Amp terminal strip

* LT44 transformer L2
* variable capacitor VC1
* 2 M2.5 4mm screws
* 3 M2.5 20mm screws
* 3 M2.5 nuts
* 5 M2.5 washers
* BAT48 schottky diode D1
* 470pF ceramic disc capacitor C1
* ceramic / crystal earpiece Y1
* 30SWG / 0.315mm enameled copper wire 4m
* 1m yellow aerial connection wire
* 1m green earth connection wire
* 1 yellow insulated crocodile clip
* 1 green insulated crocodile clip
* control knob for short shaft
* 100mm long, 9.5mm diameter ferrite rod
* 100mm single core red hookup wire
* 100mm single core black hookup wire

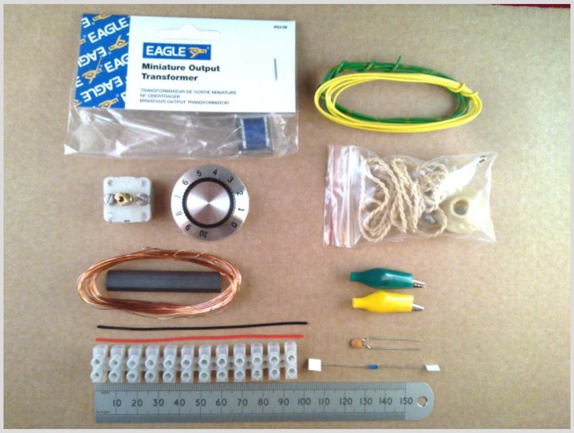

General Connections

[Link to Large Photo of Choccy Block Crystal Radio Wiring]

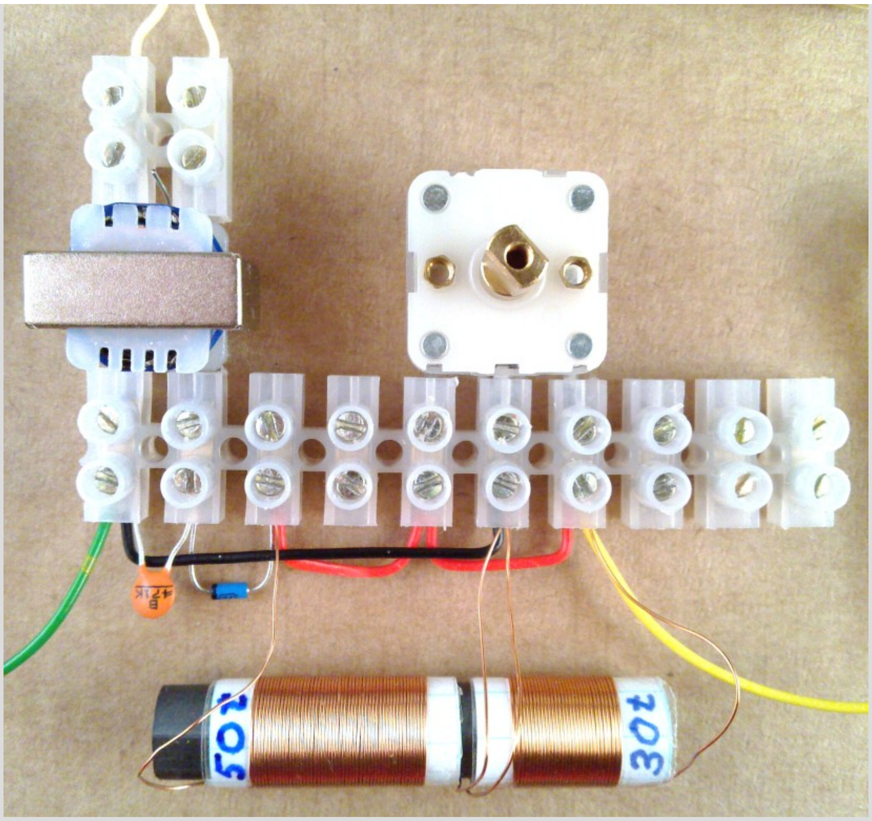

In the photo above you can see each component and where it is connected quite easily, using the large wiring picture if needed. You need to cut a group of two terminals from one end of the block for the connections to the earpiece. The red and black link wires, the aerial and earth wires, and the crystal earpiece wires need to be stripped of plastic insulation where they connect into the terminal block. The coil wire ends need to have the enamel insulation scraped off where they connect in. The LT44 transformer should be connected so that the outer two wires of the end with three wires are connected to the earpiece wires. You can leave the middle wire disconnected. The transformer is best butted right up against the terminal strip which gives it a little bit of support to stop the wires bending. The same applies to the variable capacitor.

The Variable Capacitor Connections

Looking at the variable capacitor from the bottom side, the side without the control shaft, you will see two trimmer screws and the three strip connections. The two trimmer screws add extra fixed capacitance to the variable elements. They should be adjusted as shown in the third photograph for minimum extra capacitance. This is the position where the semi-circular moving plates attached to the screws are clear of the lower metal plates.

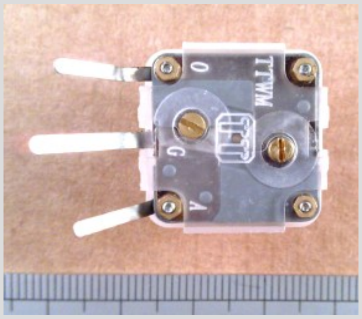

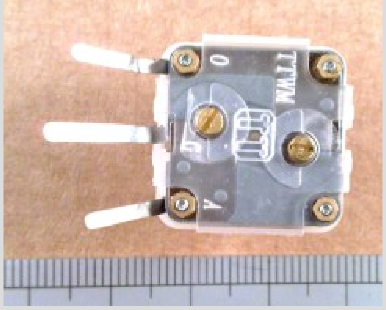

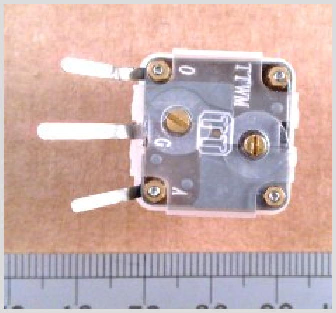

The middle strip is the common ground for the two capacitors and the outer strips are the remaining connections. The variable capacitor is actually two capacitors which are connected together in the terminal strip circuit. As supplied, the connection strips exit at the side of the top face of the capacitor. To fit it into the terminal strip you need to remove the transparent plastic case of the variable capacitor, and unfold the strips so that they go out of the holes in the bottom face, as shown in the picture sequence.

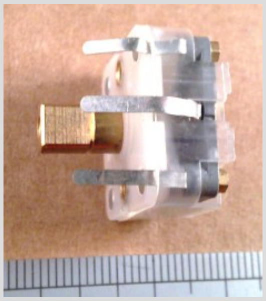

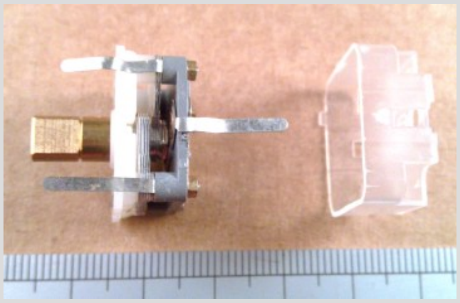

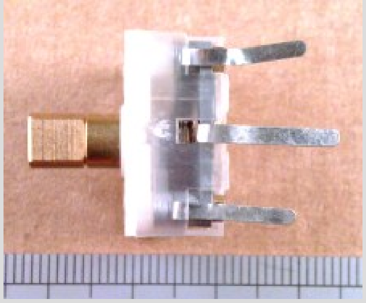

Making the Coils

To wind the coils you need the wire supplied, some sellotape, scissors, ordinary paper and some glue. The process in shown in the pictures. You first need to make two paper tubes by cutting a piece of paper 25mm wide and about 100mm long and another about 20mm wide by 100mm long. The tubes are formed by applying glue to two thirds of the length of the paper and wrapping it loosely around the ferrite rod. The tubes need to be able to slide up and down on the rod, so don't wrap it too tightly. It's best to leave the glue to dry before winding the wire on so that the tube doesn't crush and bind on the rod. The wire is secured at one end of the paper former with a thin strip of tape and then for the larger main coil, wind 50 turns onto the tube. The aerial coupling coil requires just 30 turns. It is better to turn the tube rather than winding the wire around by hand, because winding it around puts a twist in the wire. Again, you need to wind the turns just tightly enough to stay in place so that the coil and tube can still move on the rod afterwards. At the end of each coil, another thin strip of tape holds the far end of the wire in place. If you lose exact count of the number of turns don't worry. It need only be approximate.

Update: The ferrite rod provided is now 100mm long, so it's a bit easier to handle and you can make the paper tubes a bit wider. You still want to be able to slide the antenna coil close up to the tuning coil though.

mm
10 20 30 40 50 60
½mm

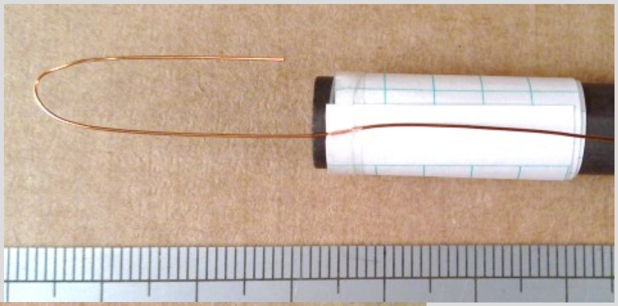

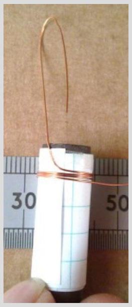

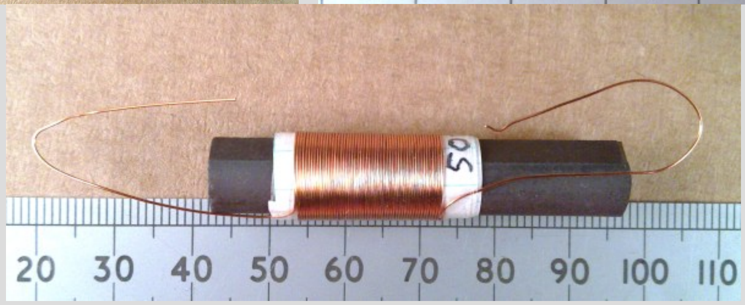

Connecting to the Crocodile Clips

The crocodile clips are much more convenient for connecting to aerials and earth connections than bare wires. You need to strip the ends of the yellow and green wires first, then remove the insulating boot from the clips as shown. The metal flaps on the ends of the clips can then be crushed down on the folded-back stripped wire ends to make a firm connection before putting the boot back on from the other end of the wire. To remove and replace the insulating boot, hold the jaws open by clipping them onto some small object.

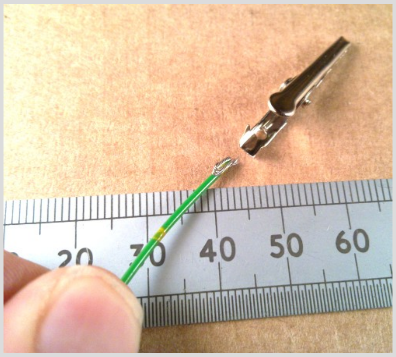

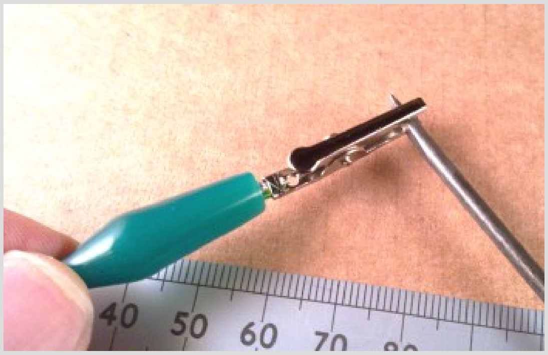

Finishing Off

It is worth checking that all of the wires are trapped under each screw correctly and that none of them will pull out. When everything is wired as in the main picture, the control knob can be fitted to the shaft of the variable capacitor and you are ready to test.

Testing and Operation

Connect the green earth wire to your earth connection and the yellow aerial wire to your aerial. With the coils in about the position shown, turn the tuning knob until a station is heard. You may find that you can increase the loudness of a particular station a bit by moving the coils around on the ferrite rod, but the adjustment is not critical. There's very little to go wrong, so troubleshooting should be confined to checking that the connections are made as shown. That concludes the building instructions and should get you going. Once it's working you can fix the terminal strips down to a wooden board by carefully banging in nails down through the holes or using the M2.5 screws provided. If you decide to fix the variable capacitor to a front panel, only use the very short 4mm screws provided. These are guaranteed short enough not to foul the insides of the capacitor.

If you want to get on and build it, that's all you need to know but there's more information below which might be useful and some technical discussion.

More Information

Aerials and Earths

I have successfully used this to pick up the main MW station in the UK just by connecting to a central heating radiator for the earth and to the outer casing of a television antenna cable, disconnected from the TV, for the aerial. This is not ideal because the TV cable is nailed to the house which both blocks and drains away a lot of the MW radio signal. It is much better if you can erect a long wire high up in a garden and there are many places online which will show you how to do that, taking the usual precautions against lightning strikes and going near to power cables. Not everyone has a garden though, and you don't want the TV aerial being disconnected on a regular basis, so there are other interesting ways to get an aerial and earth.

Farm and Other Fences

While out walking I wondered if there were any useful MW signals being picked up on farm fences. They are not good aerials for a number of reasons, but they can be quite long which can make up for that. You need to connect the earth clip to the lowest wire and the aerial clip to the highest wire. They will always be a bit corroded so you have to wiggle the clips to get a good connection. If you find a fence in good condition, outside of a deep valley which has dry posts and isn't covered in wet bushes, you will almost always get reception of one or more station. Station pickup depends on the direction of the fence as well, so it's worth trying a few if you are out walking.

Some fences are electrified and it will be pretty clear which those are, those being a single wire with insulated posts. Obviously, I have to say it: Don't use those.

Update September 2011: If you get a high-ish spot and a decent dry fence this can work even better than I expected. I walked up Clee Hill in Shropshire recently and attached the set to an ordinary fence with the ground wire clipped to the standard square mesh below and the barbed wire on top. I wasn't expecting much because the mesh and wire sections have little separation, only a few inches. Signals from Absolute radio, Five Live and another station were good and loud. I needed to move the aerial coupling coil away from the main coil on the rod to help separate the stations. The air traffic radar station in the background is coincidental.
Update October 2011: The New Forest campsite at Holmsley is neither particular high, nor is the fence all that uncluttered; Still I picked up three stations from there and found that it was best using either myself holding the green clip for the ground or the very bottom fence wire which was partially buried.

Other Structures

If you see any two metal structures which are not obviously connected together you can see if there is reception available from them. For example at various times in the past, crystal radios have been used in prisons connected to a water pipes for an earth and a bed frame for an aerial. Always stay away from using the electrical mains earth and avoid connecting to household appliances.

Boxing it Up

If you want to carry this around on walks, it's a bit inconvenient in the basic form which was intended to be the easiest way to build it up. If you want to put it in a small box, the first thing to do is to change the variable capacitor connection strips back around so that they once more exit at the same side as the shaft. You can then fit the variable capacitor into the terminal strip upside down so that the shaft faces downwards, or outwards from your box, and the terminal strip and transformer will all be flush with the same surface. I've shown an example below where I have used an old cassette case. This isn't very good because the transformer is too big so you have to cut a hole for it in the back. It doesn't matter which terminal strip connection along the block each circuit node connects into, so you can shuffle the parts along towards the left to save space, and cut off the unused terminals. The tuning knob is specially selected as one which works well with this tuning capacitor: The grub screw is quite near to the bottom of the knob and it has a recess underneath which clears the mounting screws. This allows it to fit well on the short tuning shaft.

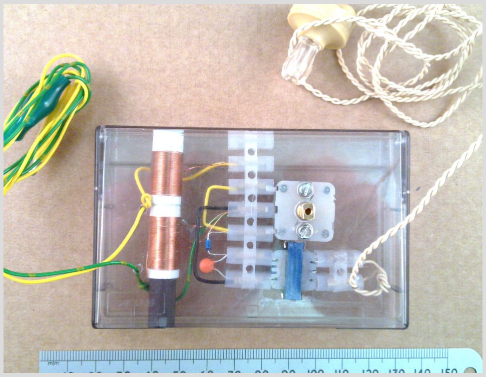

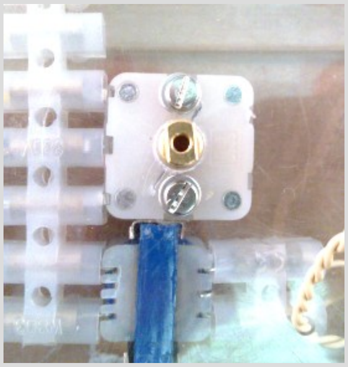

Soldering

If you're happy soldering the wires directly together it is possible to squeeze things up even further. If you shorten the ferrite rod and wind the coils in a small pile you can just about fit the whole thing into an old 35mm film canister. The tuning knob sits neatly at one end. Something a bit like this was used in the Maze prison in the early 1980s. If you want to shorten the ferrite rod to fit it in a box, first use a hacksaw to cut a shallow notch in it. Then wrap the rod in a tea towel and snap it into two pieces at the notch by hand.

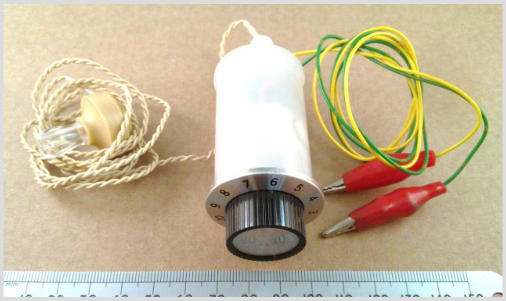

Further Technical Discussion

Circuit Operation - Earpieces

Anyone who's played with crystal sets a bit will see that everything in the circuit is pretty standard. There are some small points to note. The beige coloured ceramic / crystal earpieces are quite sensitive and in circuit model terms appear much like a 26nF capacitor in series with some lossy resistance. 26nF is about 6K Ohms impedance at 1kHz. The LT44 is used in reverse to provide a higher impedance audio match to the top of the tuning coil via the diode. Looking into the diode side of the transformer, the impedance actually increases when the the earpiece is connected due to a fortunate broad resonance with the earpiece and transformer leakage inductances. The result is audio which is a bit restricted on bandwidth but which has a fairly good match and efficiency. Some of the ceramic earpieces in the supply chain are very poor at the moment and I recently sent back 67 out of an order of 100 as completely dead on arrival. The ones supplied with the kit are 100% tested.

Diodes

You can use a 1965 Mullard OA81 or OA91 germanium point contact diode if you really want to. In this circuit I've found that the modern low Vf silicon Schottkys work as well as the best germanium diode out of the junk box, and a lot better than a poor one of which there were plenty.

Coils

Crystal Detectors to Point Contact and Layered Signal Diodes in Electronics

OddMix.com - Technical Note - Crystal Radio - TN080417 - Karl Nagy

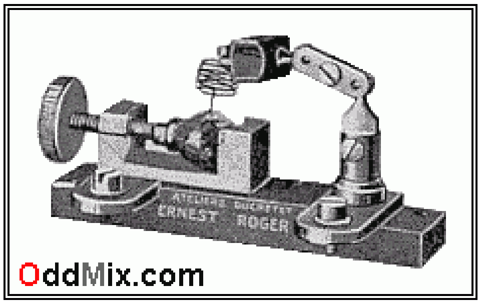

Picture 1. "Ernest Roger" Crystal detector from 1920

Many of the **younger generation** never had the pleasure to have, or to even see a real beautiful crystal detector as shown on **Picture 1**. A detector like that was the product of the early 1920's, and it was an expensive, quality, well made product. They were usually hands assembled with machine made components.

This particular "**Ernest Roger**" detector device is built on an insulated base, either wood or most likely of Bakelite, the most widely used and only available plastic material at the times.

Many of the younger generation never had the pleasure to have, or to even see a real beautiful crystal detector as shown on **Picture 1**. The new generations are much too involved with technology and playing on their computers or going to an **online University**. This excellent detector holder used copper, brass and bronze generously for the crystal holder and for all of the **electrodes**.

The most often used crystal in this and similar detectors - then and now - was **pyrite** or galena as shown in the 1925 listing in a table in the **"American Mineralogist"** publication, that has a listing of thirty three minerals catalogued with rectifying detector properties.

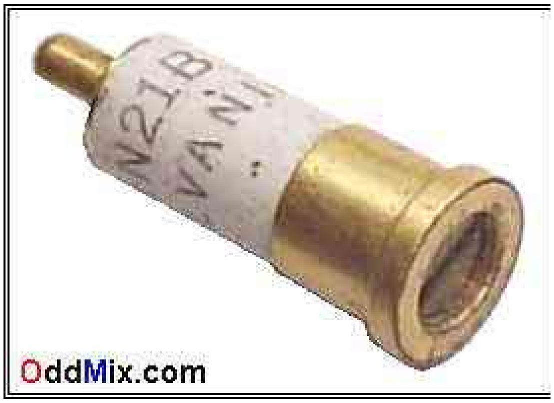

Picture 2. Early point contact diode with adjustable srew

Among the very first commercially produced diodes is the 1N21B shown on **Picture 2**. It is more than curious, that this "diode" is enclosed in a case on which there is a slotted screw-head is just visible on the wide side (lower right). That screw is connected to a fine "cats whisker" steel wire, a few turns of a spring-like device, terminating in a point that is in touch of the germanium semiconductor material connected to the top left diode terminal.

The arrangement is much smaller, and more diode like, then the detector on **Picture 1**., but it made in a very similar arrangement, which has became known as the point contact diode. From these and similar germanium diodes, evolutionary progress leads us to current PN junction Silicon, Gallium Arsenide or other more exotic semiconductor materials which are the achievement of the latest scientific age and many years of steady experimentation, research and development.

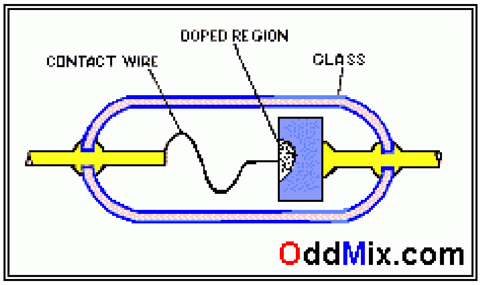

Picture 3. Point contact diode

Picture 3 shows the cross section of a hermetically glass enclosed, point contact diode. If the left side, point contact terminal would be attached to a screw and a short spring, and the glass envelope would have a threaded metal part in it, the **Picture 2** and **Picture 3** devices would be nearly identical.

These point contact "cat's whisker" devices, are still made occasionally because of their very small capacitance. As it was discovered early on, they are highly useful in high frequency electronics. All the way up to microwave frequencies they useable. The simplest radar detectors usually made with a simple horn **antenna**, a diode similar to **Picture 2** and **Picture 3** and a transistor amplifier. If the designer keeps it simple, and don't use a local oscillator, such radar detectors are not even detectable by the police.

Previous Page | **All Circuit List**

Google [ ] Search

● Web ○ www.oddmix.com

056522

Copyright © 2015 K Nagy - http://www.OddMix.com - All Rights Reserved
Page Revised: 2015-01-01 - - Privacy Policy - Site Map - Support

Hosted by 1&1

Crystal Detectors

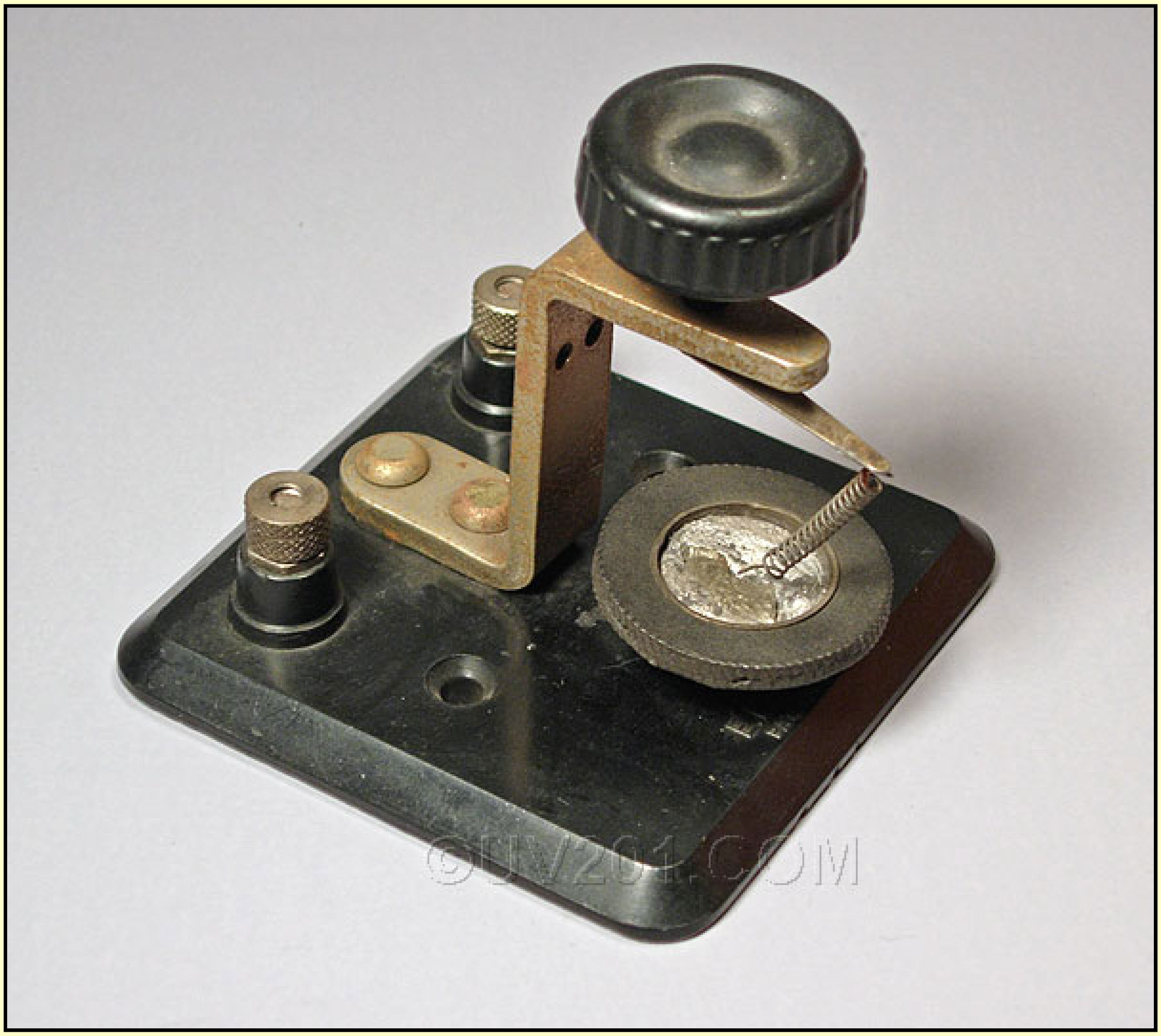

This page presents an assortment of crystal detector units which were made between 1910 and 1925.

In most cases, the "crystal" is a piece of galena (lead sulfide), though other minerals were used as well, including silicon, pyrite ("fool's gold"), and a number of others. A mount was required to hold the crystal, forming one connection, and to provide a second contact which was most often through a stiff wire that lightly touched the surface of the crystal. This wire came to be called a "cat's whisker". The standard electrical symbol for the semiconductor diode immortalizes this arrangement.

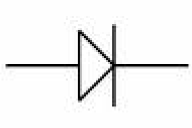

With these primitive devices, it was necessary to experimentally find a "sensitive spot", so a method of moving the crystal and/or the contact was provided, and to alter the pressure of the contact.

The detector shown above was made and sold by the Electro-Importing Company of New York. It is marked "E. I. Co. N.Y. Pat. Appl. For". This company was one of the pioneering distributors of wireless, radio, and electrical apparatus, and was founded in 1904 by author and publisher Hugo Gernsback. This detector is shown below in a scan from their 1918 catalog.

The binding post caps may not be original, as they don't match any of the catalog illustrations, though that is by no means conclusive. The base size (3 inches square) does not match the catalog description either.

No patent appears to have been granted for this design, but Hugo Gernsback was granted a patent for a relay using a similar design. Patent 978,999 was filed and issued in 1910, so it is likely that he attempted to patent this detector in that same period of time.

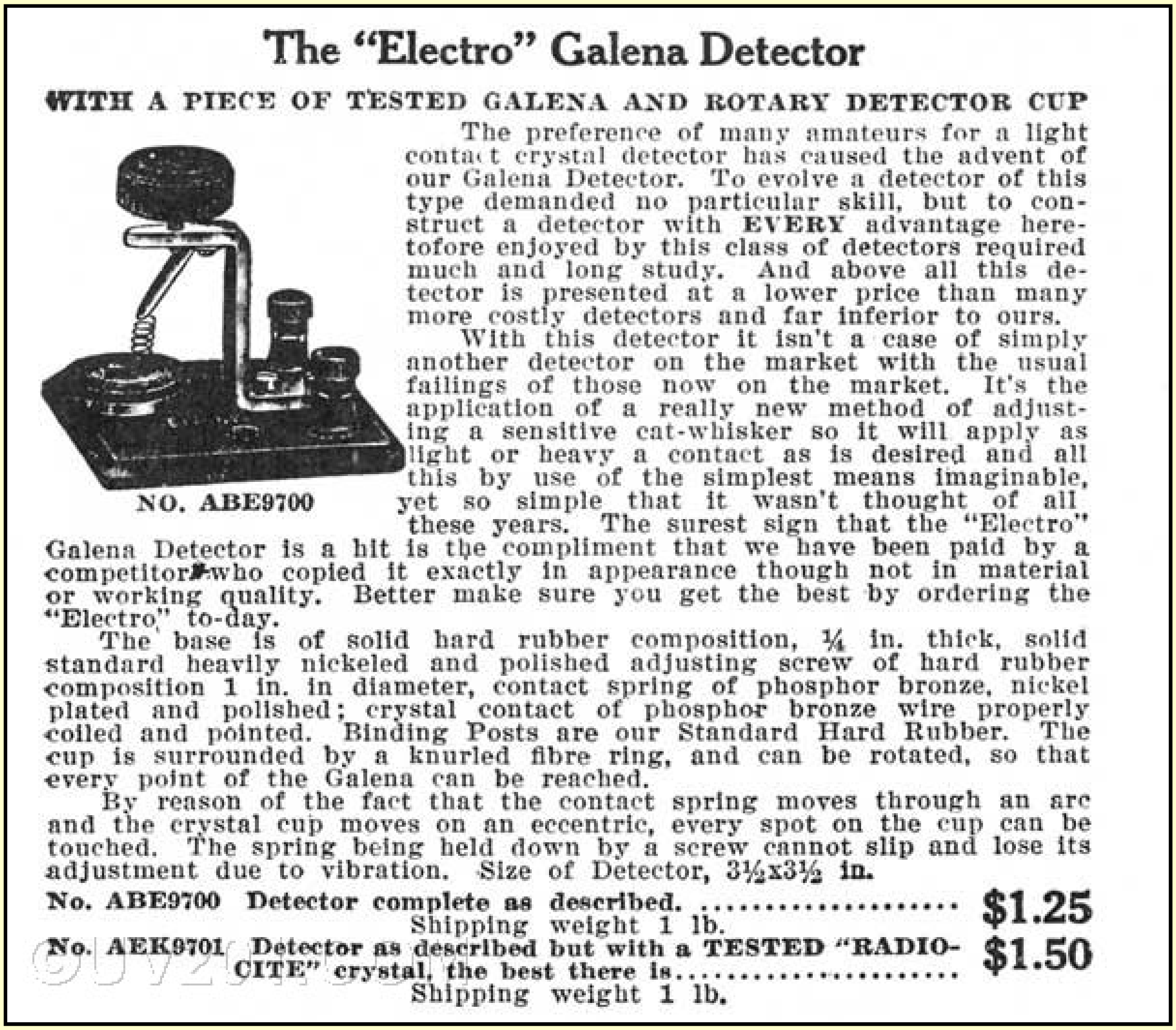

The "Electro" Galena Detector

WITH A PIECE OF TESTED GALENA AND ROTARY DETECTOR CUP

NO. ABE9700

The preference of many amateurs for a light contact crystal detector has caused the advent of our Galena Detector. To evolve a detector of this type demanded no particular skill, but to construct a detector with **EVERY** advantage heretofore enjoyed by this class of detectors required much and long study. And above all this detector is presented at a lower price than many more costly detectors and far inferior to ours.

With this detector it isn't a case of simply another detector on the market with the usual failings of those now on the market. It's the application of a really new method of adjusting a sensitive cat-whisker so it will apply as light or heavy a contact as is desired and all this by use of the simplest means imaginable, yet so simple that it wasn't thought of all these years. The surest sign that the "Electro" Galena Detector is a hit is the compliment that we have been paid by a competitor* who copied it exactly in appearance though not in material or working quality. Better make sure you get the best by ordering the "Electro" to-day.

The base is of solid hard rubber composition, ¼ in. thick, solid standard heavily nickeled and polished adjusting screw of hard rubber composition 1 in. in diameter, contact spring of phosphor bronze, nickel plated and polished; crystal contact of phosphor bronze wire properly coiled and pointed. Binding Posts are our Standard Hard Rubber. The cup is surrounded by a knurled fibre ring, and can be rotated, so that every point of the Galena can be reached.

By reason of the fact that the contact spring moves through an arc and the crystal cup moves on an eccentric, every spot on the cup can be touched. The spring being held down by a screw cannot slip and lose its adjustment due to vibration. Size of Detector, 3½x3½ **in.**

No. ABE9700 Detector complete as described. **$1.25**
Shipping weight 1 lb.

No. AEK9701 Detector as described but with a TESTED "RADIO-CITE" crystal, the best there is...................... **$1.50**
Shipping weight 1 lb.

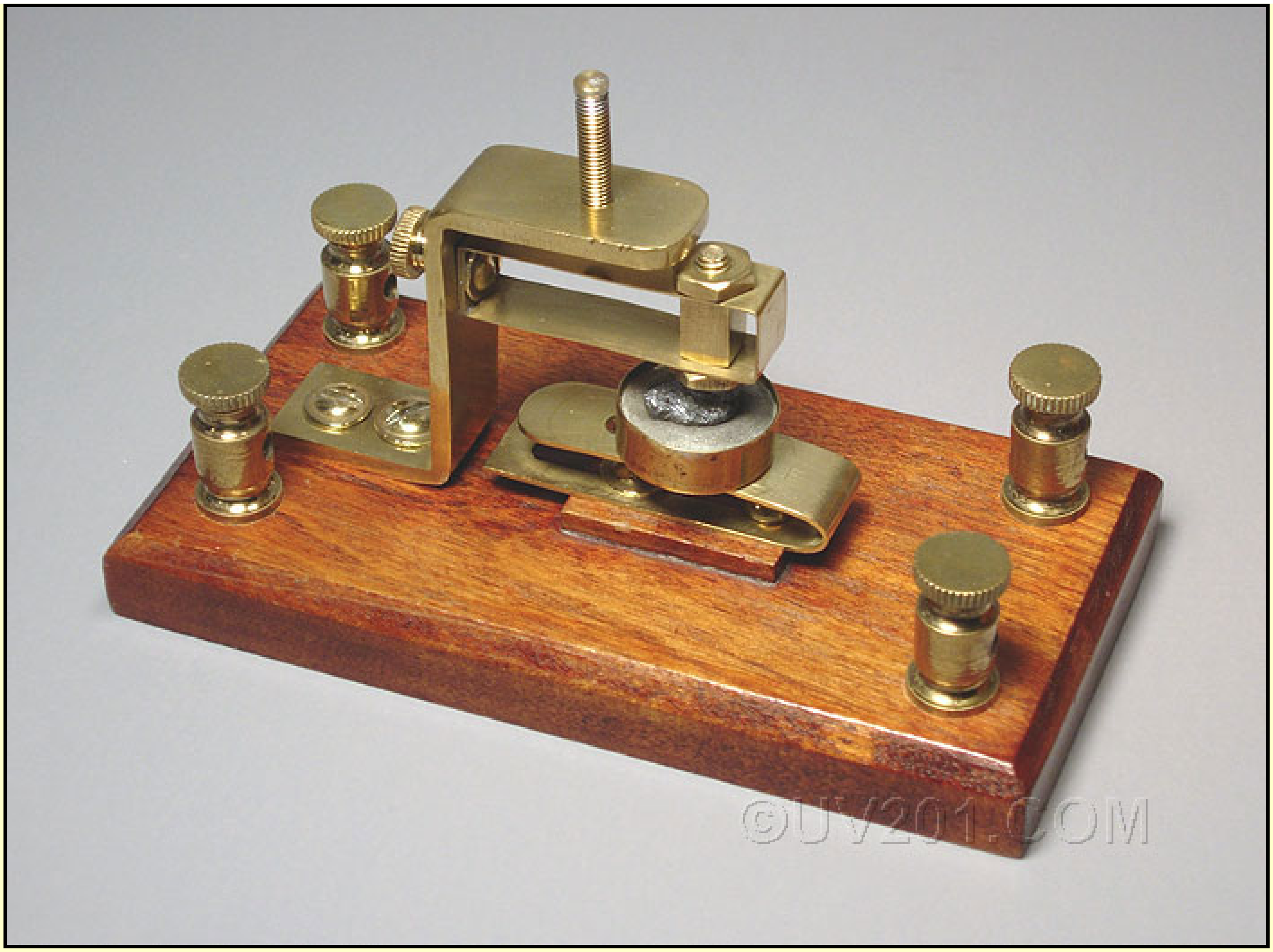

There are no markings on this early detector unit. As a guess, I would say that it dates to about 1912-1915. It is well made, so it is either a commercial product, or one that was fashioned by an above average craftsman. It is the only detector that I have seen that used a wooden base.

It's design is very similar to one that was sold by the Electro Importing Co. as shown below in a scan from their 1913 catalog.

The threaded rod at the top is used to apply and adjust the pressure of a blunt point contact on the surface of the crystal. There was originally a knob on the end of this rod, probably made of hard rubber, which is now missing. The crystal can be rotated and slid along its mount to give additional degrees of adjustment. The detector is a two-terminal device, so there is no specific need for four binding posts. They are connected together in pairs, so this may only have been done to facilitate connections. The base is about 5 inches long and 2 1/2 inches wide.

The crystal in this unit may be silicon, which required a heavier contact than galena, hence the use of a blunt point contact.

I would appreciate an email from anyone who recognizes this detector.

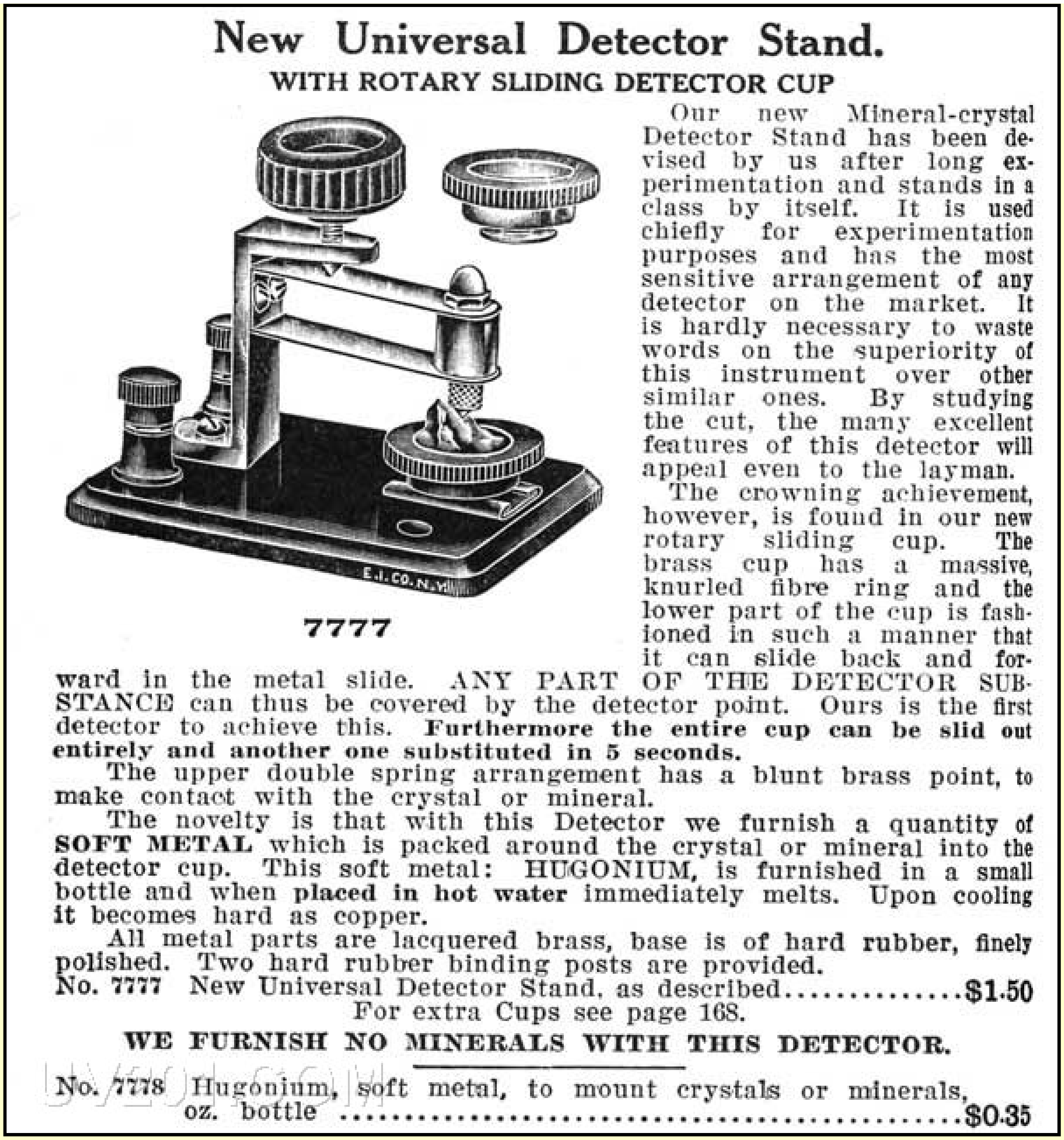

New Universal Detector Stand.

WITH ROTARY SLIDING DETECTOR CUP

7777

Our new Mineral-crystal Detector Stand has been devised by us after long experimentation and stands in a class by itself. It is used chiefly for experimentation purposes and has the most sensitive arrangement of any detector on the market. It is hardly necessary to waste words on the superiority of this instrument over other similar ones. By studying the cut, the many excellent features of this detector will appeal even to the layman.

The crowning achievement, however, is found in our new rotary sliding cup. The brass cup has a massive, knurled fibre ring and the lower part of the cup is fashioned in such a manner that it can slide back and forward in the metal slide. ANY PART OF THE DETECTOR SUBSTANCE can thus be covered by the detector point. Ours is the first detector to achieve this. **Furthermore the entire cup can be slid out entirely and another one substituted in 5 seconds.**

The upper double spring arrangement has a blunt brass point, to make contact with the crystal or mineral.

The novelty is that with this Detector we furnish a quantity of **SOFT METAL** which is packed around the crystal or mineral into the detector cup. This soft metal: HUGONIUM, is furnished in a small bottle and when **placed in hot water** immediately melts. Upon cooling it becomes hard as copper.

All metal parts are lacquered brass, base is of hard rubber, finely polished. Two hard rubber binding posts are provided.

No. 7777 New Universal Detector Stand, as described..............**$1.50**

For extra Cups see page 168.

WE FURNISH NO MINERALS WITH THIS DETECTOR.

No. 7778 Hugonium, soft metal, to mount crystals or minerals, oz. bottle **$0.35**

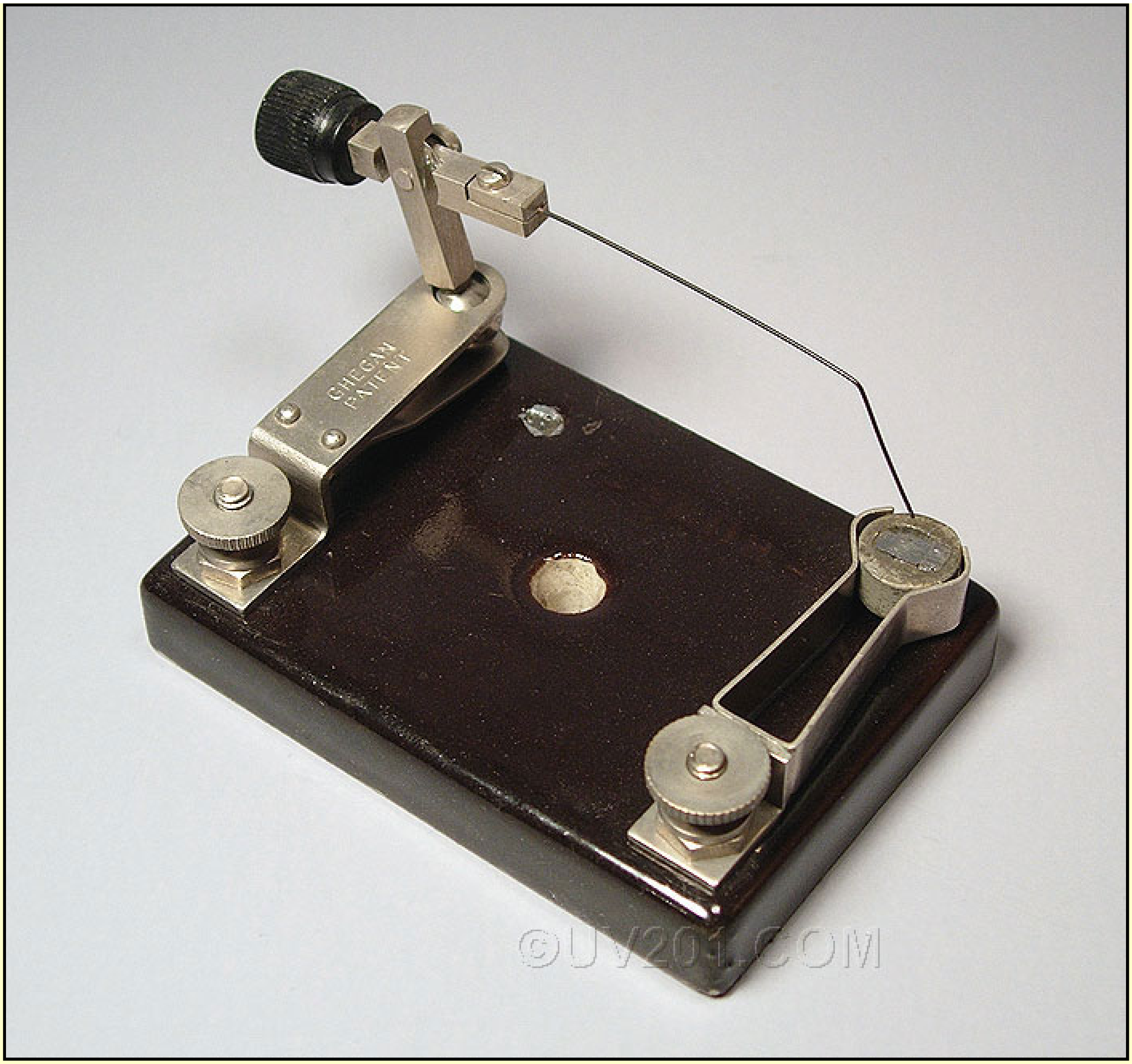

This detector was made by the J. H. Bunnell company circa 1916, and was known as the "Jove" detector. Bunnell was the largest U.S. manufacturer of telegraph equipment, and as wireless telegraphy came into widespread use, they expanded into that realm, though they were never a major player in that market.

The Jove detector was available either with a porcelain or Bakelite base. The porcelain base, as seen above, was heavily coated with a very dark brown glaze. The Bunnell name appeared in raised letters directly below the cat's whisker, but the glaze was so thick that the name was almost illegible. The Bakelite version had only the word "Jove" to identify it. Both versions were marked with the words "Ghegan Patent", a reference to patent number 1,162,765, issued on Dec. 7, 1915. This was John Ghegan, one of Bunnell's most prolific designers, who became president of the company in 1903.

The base measures about 3 1/4 by 2 1/4 inches. The adjusting mechanism is particularly nice, and is what you might expect from a company whose primary business was building precision mechanical devices.

They appeared to have used two styles of binding posts on these units. A Jove detector can be seen in a vintage photograph elsewhere on this site. That detector uses the same binding posts as seen in the Bakelite based detector shown below, which tends to confirm that these are original parts.

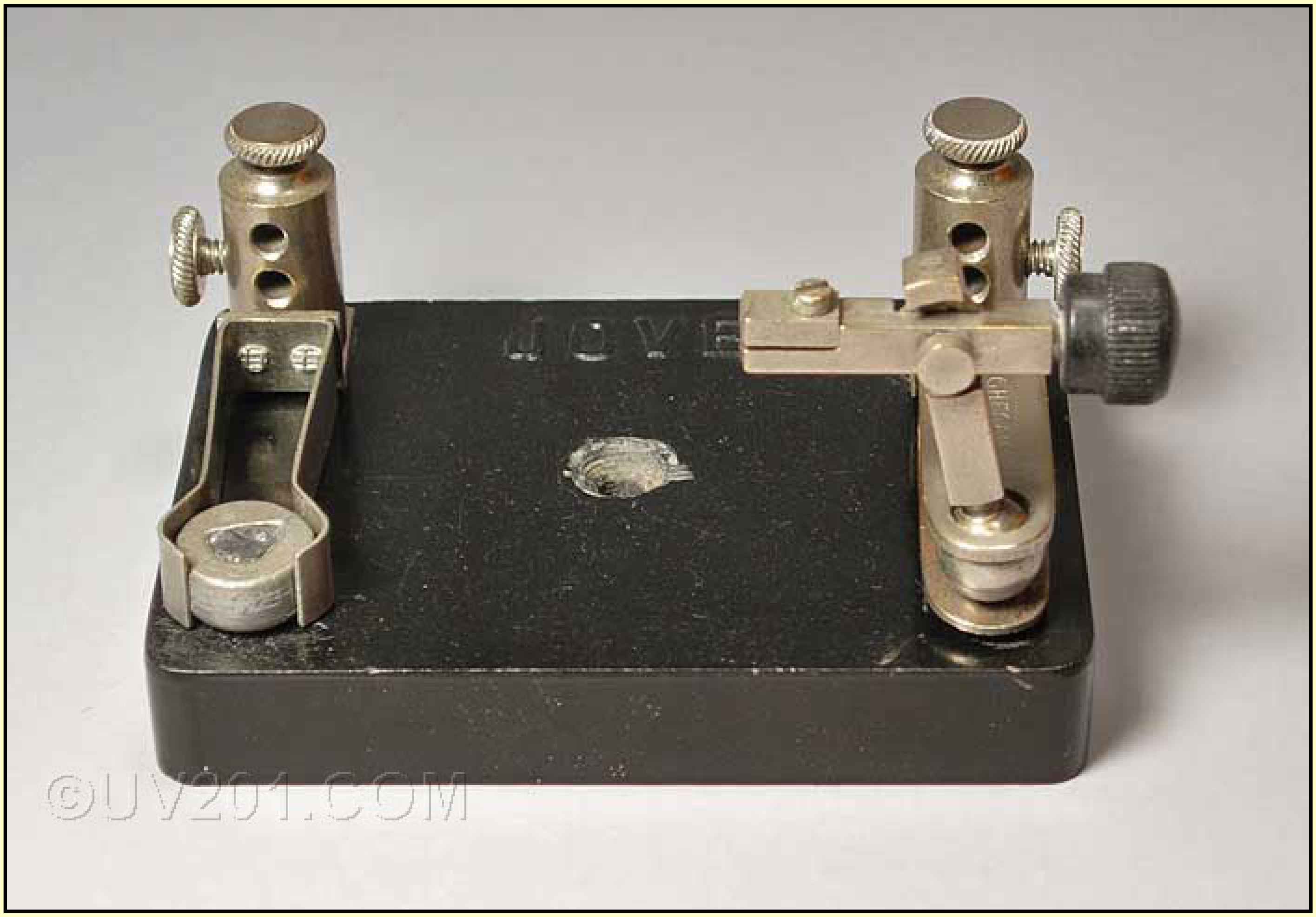

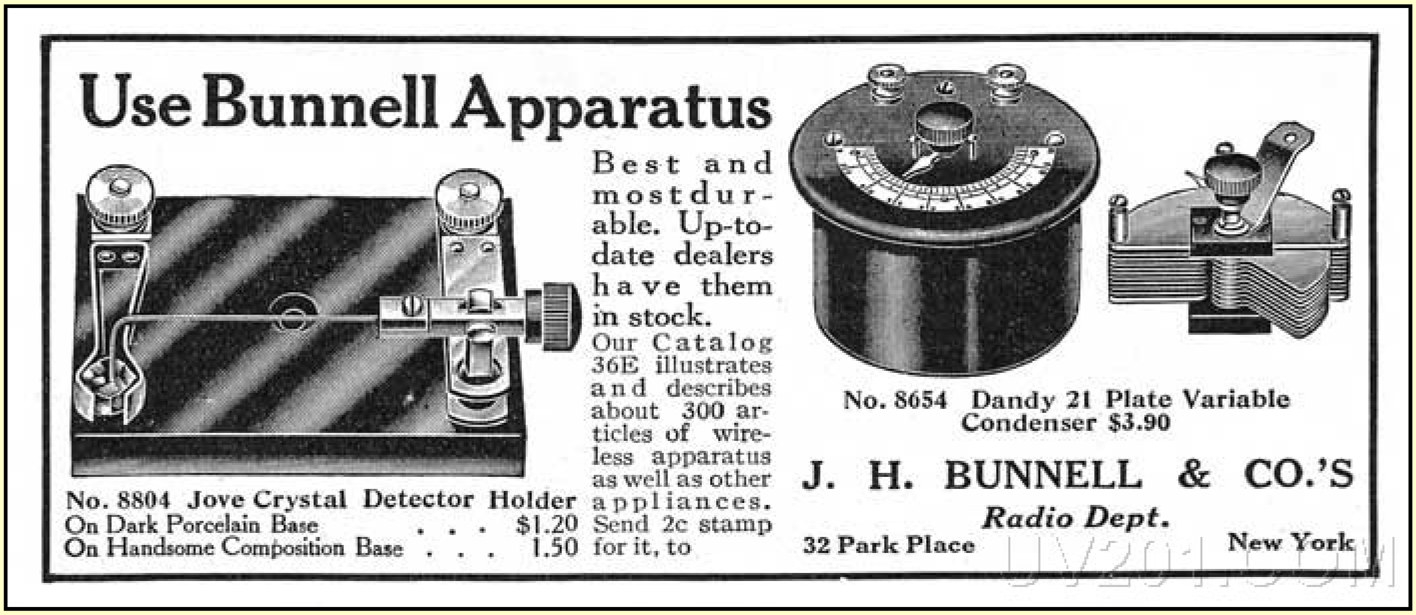

Use Bunnell Apparatus

Best and most durable. Up-to-date dealers have them in stock. Our Catalog 36E illustrates and describes about 300 articles of wireless apparatus as well as other appliances. Send 2c stamp for it, to

No. 8804 Jove Crystal Detector Holder

On Dark Porcelain Base . . .	$1.20
On Handsome Composition Base . . .	1.50

No. 8654 Dandy 21 Plate Variable Condenser $3.90

J. H. BUNNELL & CO.'S
Radio Dept.
32 Park Place New York

This advertisement for the Bunnell Jove detector appeared in the February, 1917 issue of the Electrical Experimenter magazine.

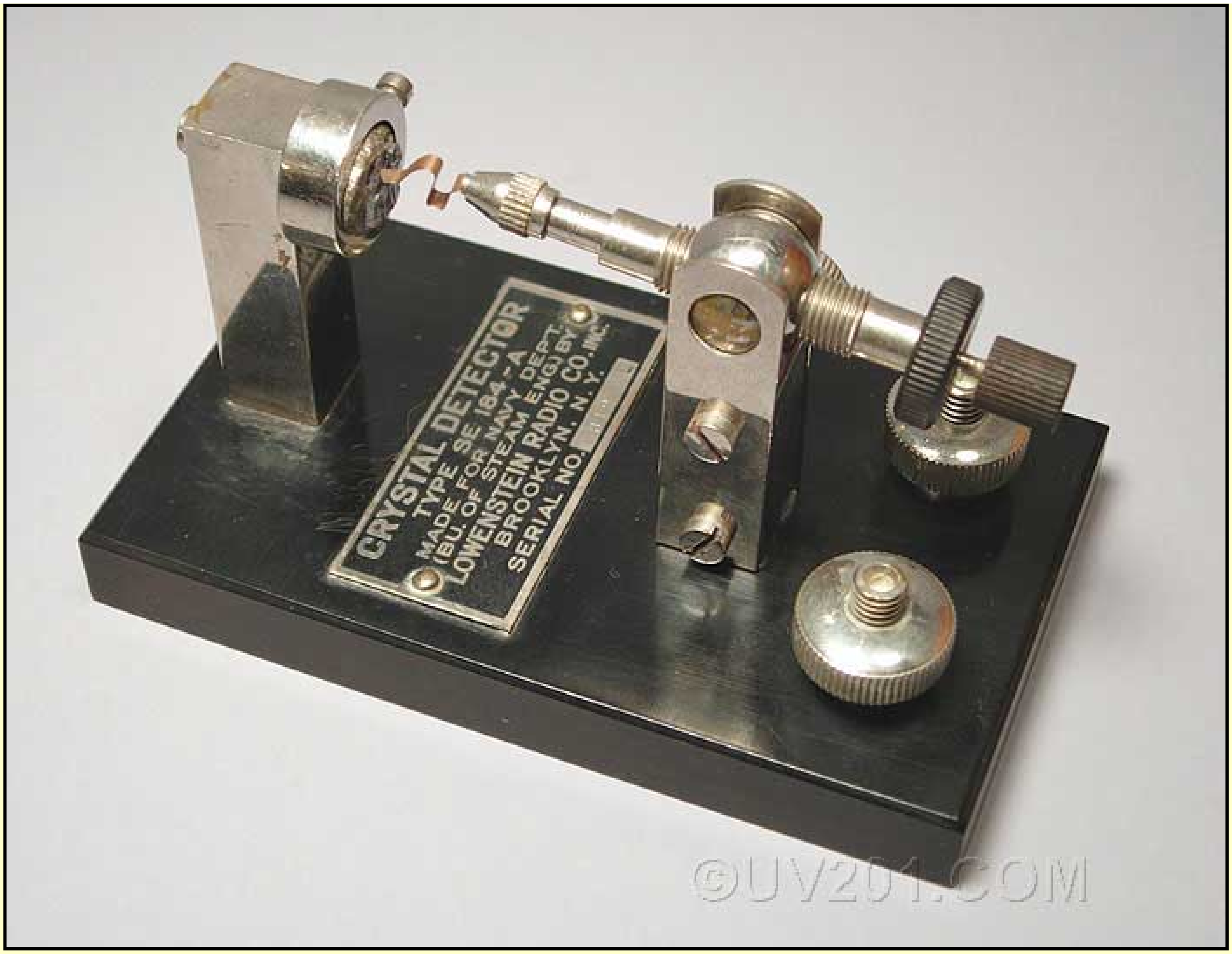
CRYSTAL DETECTOR
TYPE SE 184-A
MADE FOR NAVY DEPT.
(BU. OF STEAM ENG.) BY
LOWENSTEIN RADIO CO. INC.
BROOKLYN, N. Y.
SERIAL NO.
©UV201.COM

This beautifully made and very robust type SE-184-A detector unit was made by the Lowenstein Radio Company for the U. S. Navy, circa 1918. The Bakelite base is about 4 1/4 by 2 1/2 inches and is 1/2 inch thick.

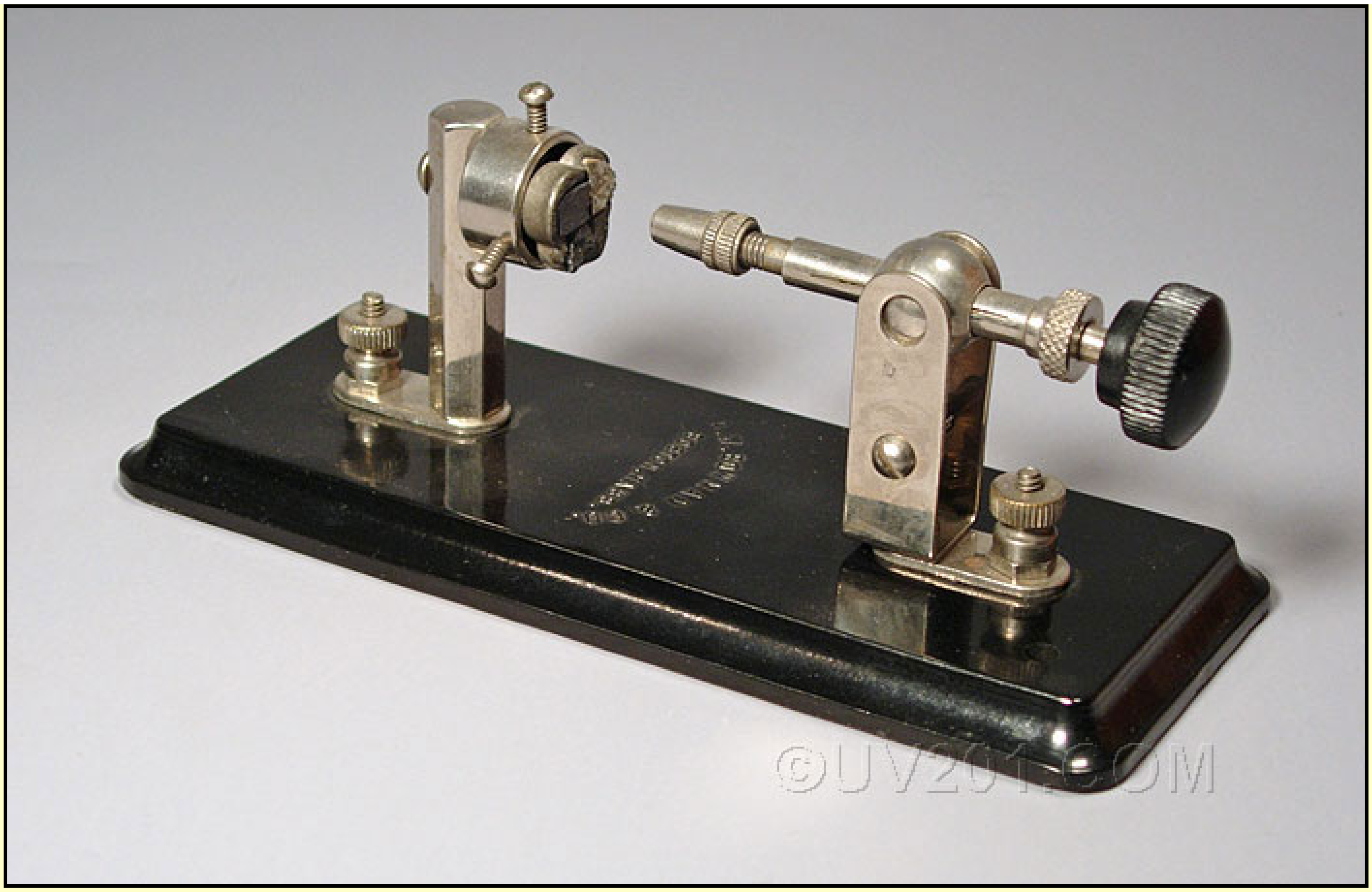

This detector was made by A. W. Bowman & Company of Boston, Massachusetts. The Bakelite base measures 6 by 2 1/2 inches. Though not original, this detector is equipped with three different minerals which could be selected for use by positioning the cat's whisker (which is missing) onto the surface of the desired crystal. This detector was sold by the Sears Roebuck company, and can be seen in an excerpt from a 1924 Sears radio catalog below. Click here or on the image to see the whole page.

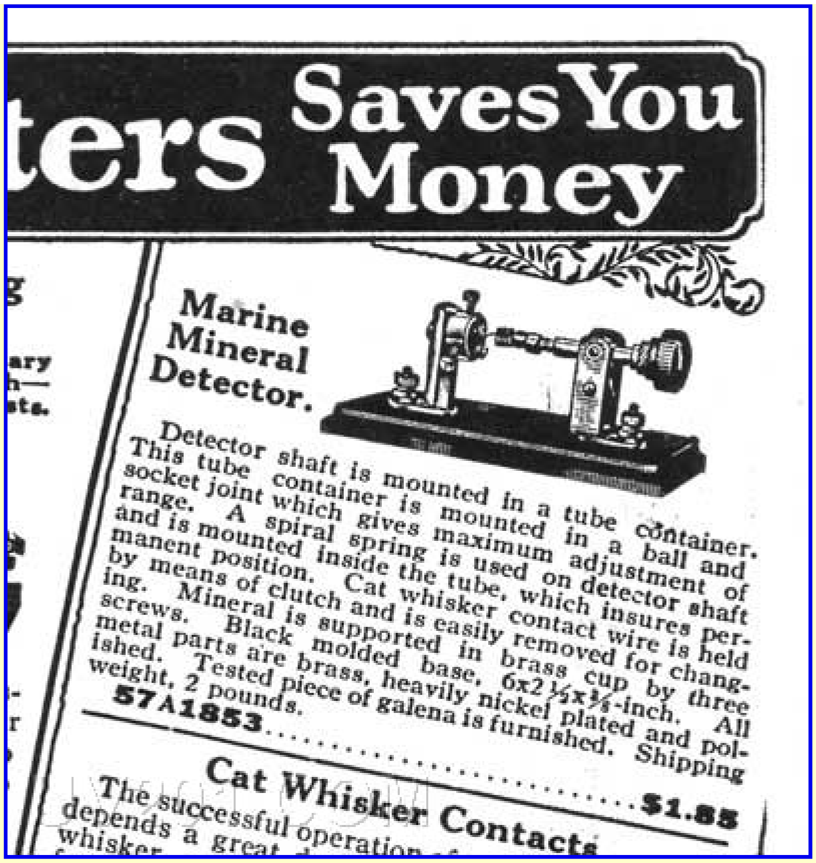

ters Saves You Money

Marine Mineral Detector.

Detector shaft is mounted in a tube container. This tube container is mounted in a ball and socket joint which gives maximum adjustment of range. A spiral spring is used on detector shaft and is mounted inside the tube, which insures permanent position. Cat whisker contact wire is held by means of clutch and is easily removed for changing. Mineral is supported in brass cup by three screws. Black molded base, 6x2½x⅜-inch. All metal parts are brass, heavily nickel plated and polished. Tested piece of galena is furnished. Shipping weight, 2 pounds.

57A1853........................**$1.85**

Cat Whisker Contacts

The successful operation
depends a great
whisker

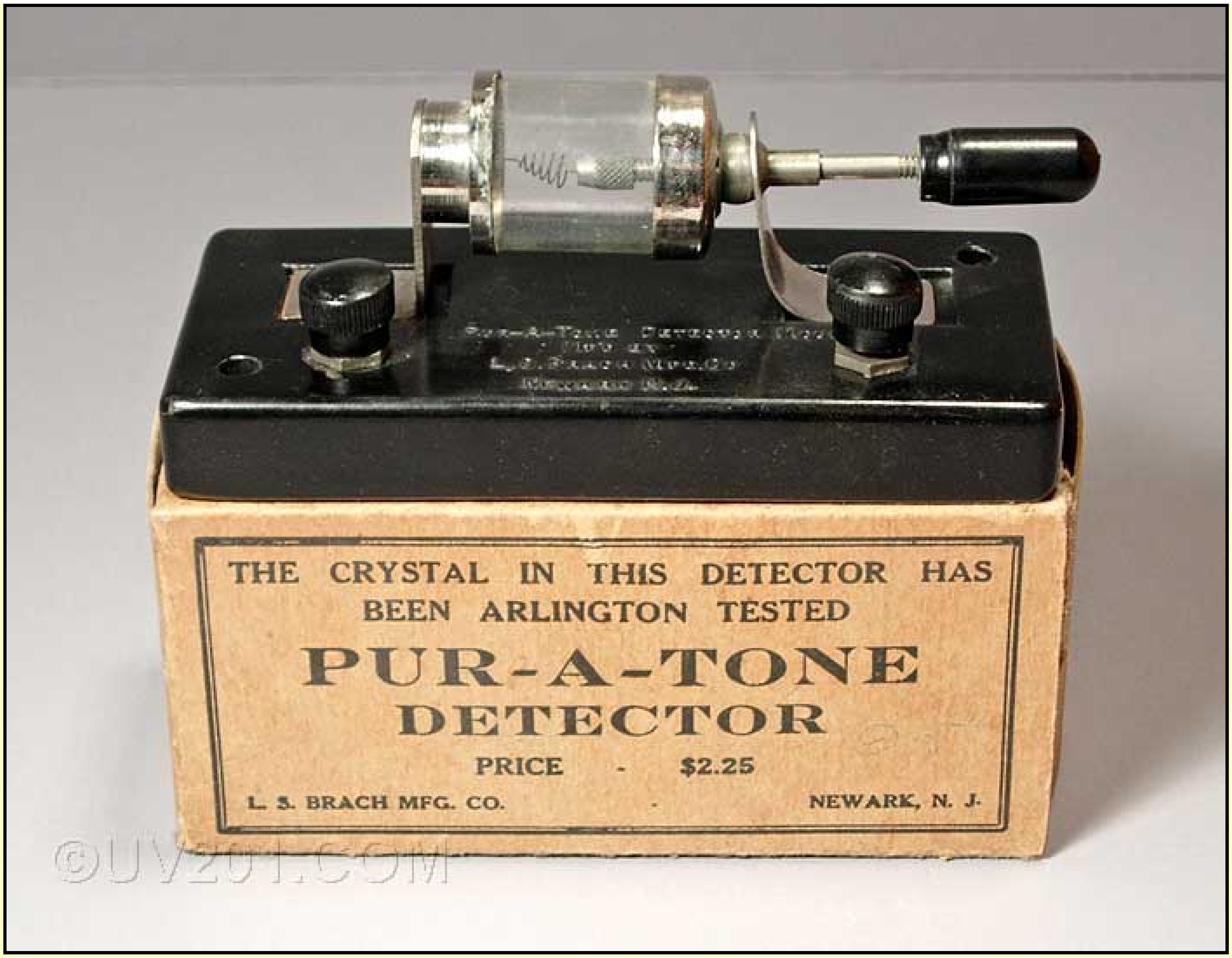
THE CRYSTAL IN THIS DETECTOR HAS
BEEN ARLINGTON TESTED
PUR-A-TONE
DETECTOR
PRICE - $2.25
L. S. BRACH MFG. CO.
NEWARK, N. J.
©UV201.COM

This Pur-A-Tone detector dates to the early 1920's, and is shown with its original box. The protective glass sleeve was commonly used to keep dirt and finger oils off the crystal.

The Navy transmitted time signals from its station NAA located in Arlington, Virginia, at 2500 meters (33 kHz) starting in 1913. It's transmissions were so reliable that they were used as a test signal for radio equipment. If you couldn't get NAA, you knew you had a problem.

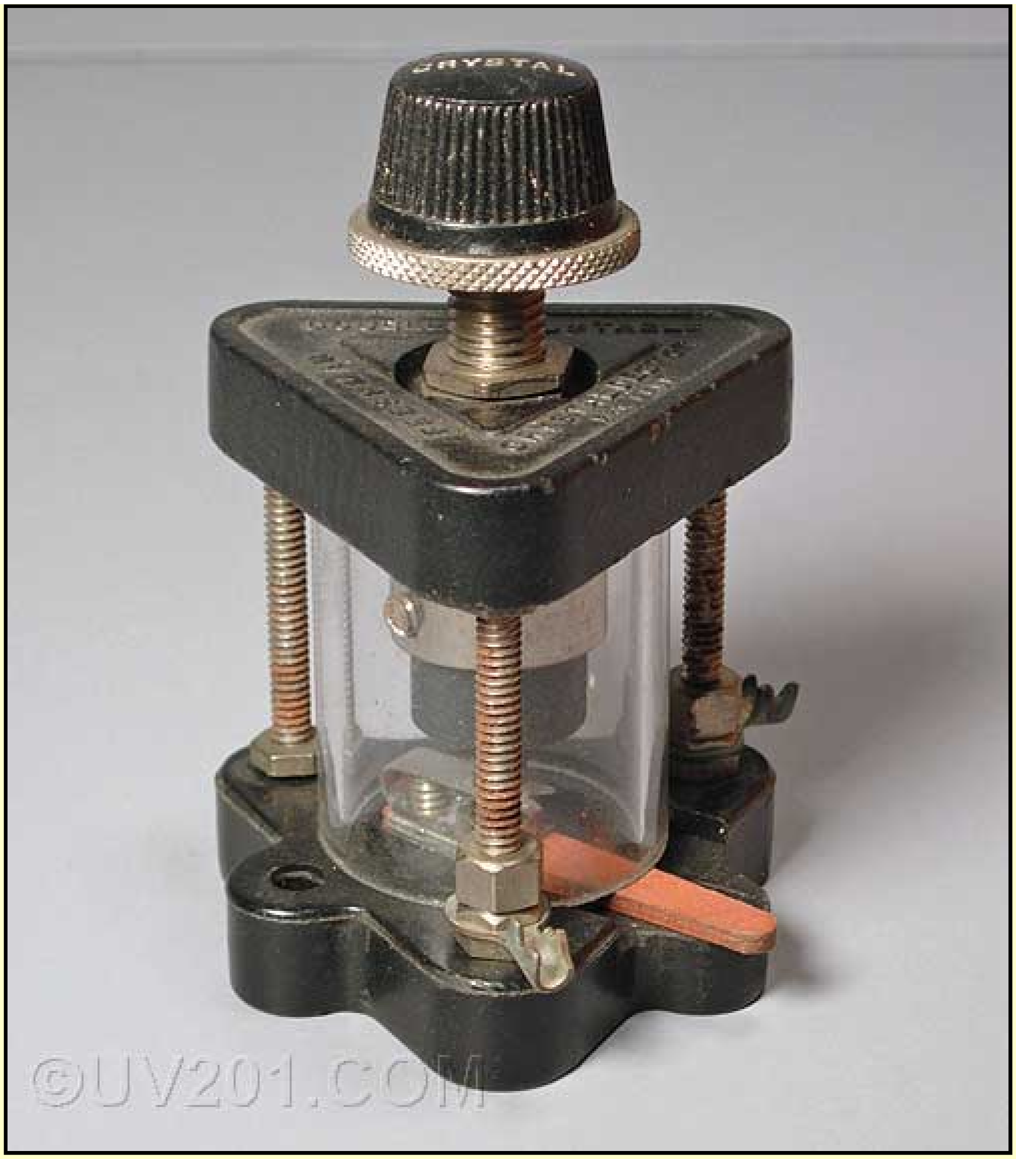

This detector was made by the Charles Freshman company in the early 1920's. The knob at the top rotated the crystal holder, and the lever at the bottom adjusted the position of the cat's whisker.

Home

CRYSTAL ENERGY RECEIVER©

by Drew Paul

Build Your Own Crystal Energy Receiver©

Crystal Energy Receiver©

FREE ENERGY
FROM THIN AIR!
INVOLVES CRYSTALS AND DANGEROUS VOLTAGES!
Copyright 2016. Drew Paul Designs. All Rights Reserved.

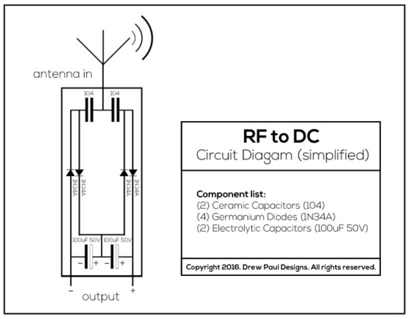
antenna in
104
104
1N34A
1N34A
1N34A
1N34A
100uF 50V
100uF 50V
output
-
+
RF to DC
Circuit Diagam (simplified)
Component list:
(2) Ceramic Capacitors (104)
(4) Germanium Diodes (1N34A)
(2) Electrolytic Capacitors (100uF 50V)
Copyright 2016. Drew Paul Designs. All rights reserved.

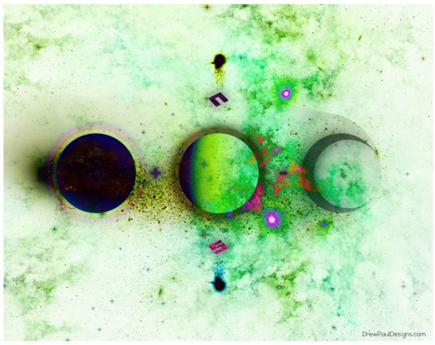
DrewPaulDesigns.com

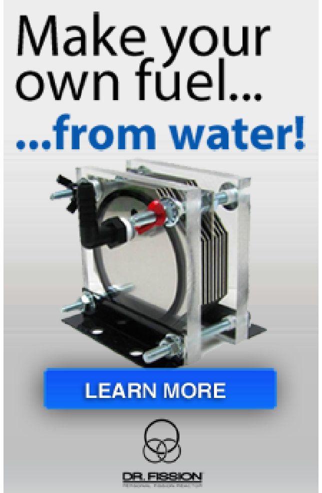
Make your own fuel...
...from water!
LEARN MORE
DR. FISSION

"Dendera Light"
circa 1500 BC
Image: creative commons, Wikipedia

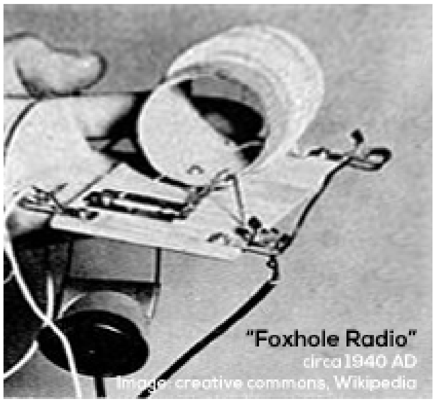
"Foxhole Radio"
circa 1940 AD
Image: creative commons, Wikipedia

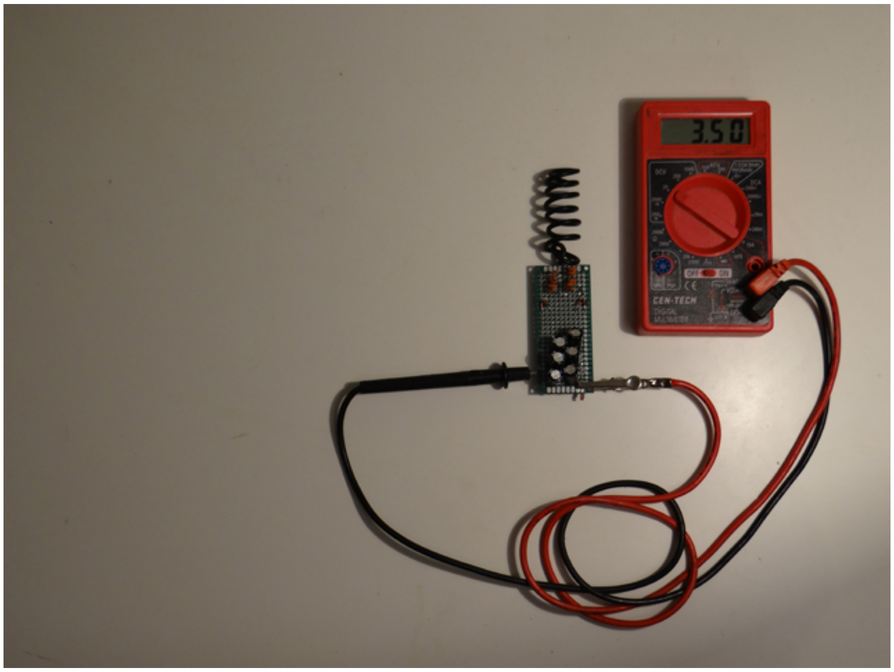

"Free energy from the air? Yea, right!"

Sardonic skepticism was my first reaction to this unusual concept, as well.

Though, its not so far out there, in fact. Light can be converted to DC current with solar panels, electricity can be converted to magnetism as I did in my last article, in a microphone sound waves are converted (by vibrating a magnet near a coil) to an electrical signal, solar rays can even be focused and converted to heat in devices like this. When we think about it, energy is all around us and can be harvested in an enumerable many of ways.

Today, we are going to take a rather novel approach to this venture. The earth is magnetic and anyone who has ever used a compass knows this. Magnetic bodies in motion produce electricity, we can see this in any alternator, like the one in your car. So, therefore the earth is electric as well as magnetic, by their very definitions.

Can we detect this energy? Yes, we sure can! Ever turn on a radio in the middle of nowhere and heard static? That is your radio picking up naturally occurring energy in the RF spectrum!

Can we use this energy to do work? Absolutely! This has been known for a long time. Crystal radios have been around since before the 1930's and can run with no input energy other than the radio signal. Even when completely isolated, but from the atmosphere, a crystal radio will produce a voltage in the earpiece resulting in a sound (albeit and undesirable one).

Well, this is where it gets interesting...

Can we replicate this effect? Yea, and with modern components like the high quality crystals found in germanium diodes, we can even increase efficiency. By applying this concept as a contemporary crystal energy receiver, we can also take advantage of a wide range of energetic frequencies rather than tuning in to just one. Can we scale it up? Definitely. Things like micro germanium diodes, high efficiency antennas and compact contemporary capacitors make the components that are required to build a crystal receiver fit in the palm of your hand. While there may or may not be a more efficient way, this renewable energy solution is simple to employ and can be scaled up, or down, indefinitely.

It sounds like we can build a Crystal Energy Receiver.© Let's give it a shot...

What You'll Need

Get the Kit

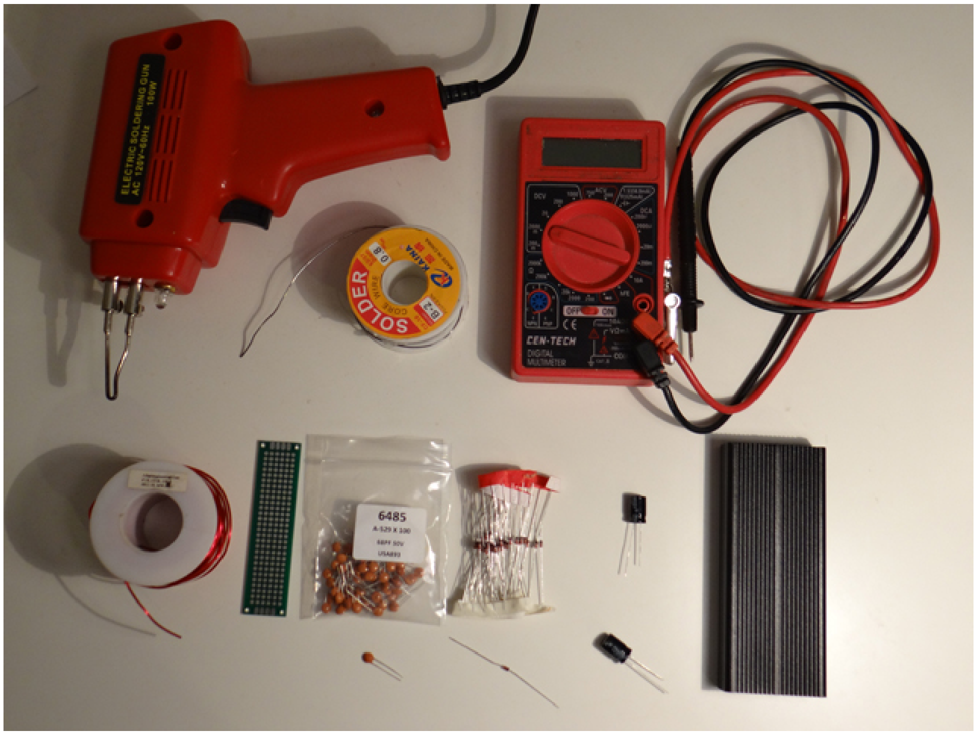

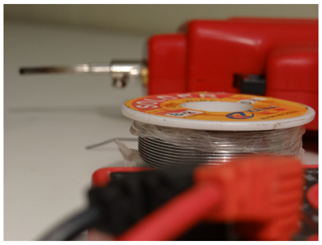

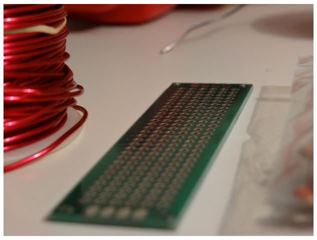

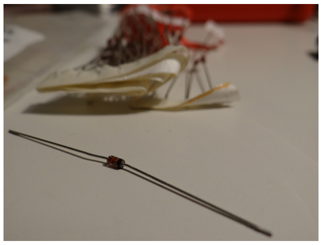

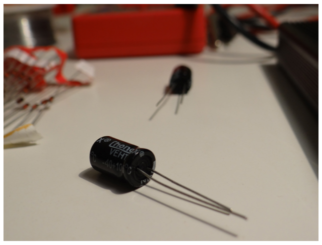

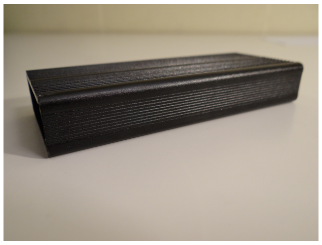

One of the reasons this particular renewable energy harvesting method is so viable is the relatively few and easy to obtain materials required.

The simplest crystal receiver design needs no power and can be built with only three parts: a coil, a crystal and a resistor. We're going to optimize that design in order to produce a cleaner and more reliable output signal by first polarizing the input amplitude, then rectifying and filtering the signal. Then we'll add an antenna, case and connections.

The parts for the circuit include:
(1) Circuit Board
(1) 10-18 gauge Copper Wire
(2-12+) Ceramic Capacitors (matched)
(2-6+) Electrolytic Capacitors (matched)
*note various types of capacitors can be used
(4) Germanium Crystal Diodes (1A+)
Total Cost: +/-**$0.40** (USD, scaled for volume)

In addition, you'll probably want to get:
(1) Project box (optional)
(1) Antenna (a loop antenna or elevated antenna is recommended and can be made with copper wire)

Get the Kit

The tools you'll need are:
Soldering Iron/ Solder (optional)
Multimeter
Oscilloscope (optional)

That's it. Yup, that's all. Once we've got it all, let's begin.

Build the Circuit

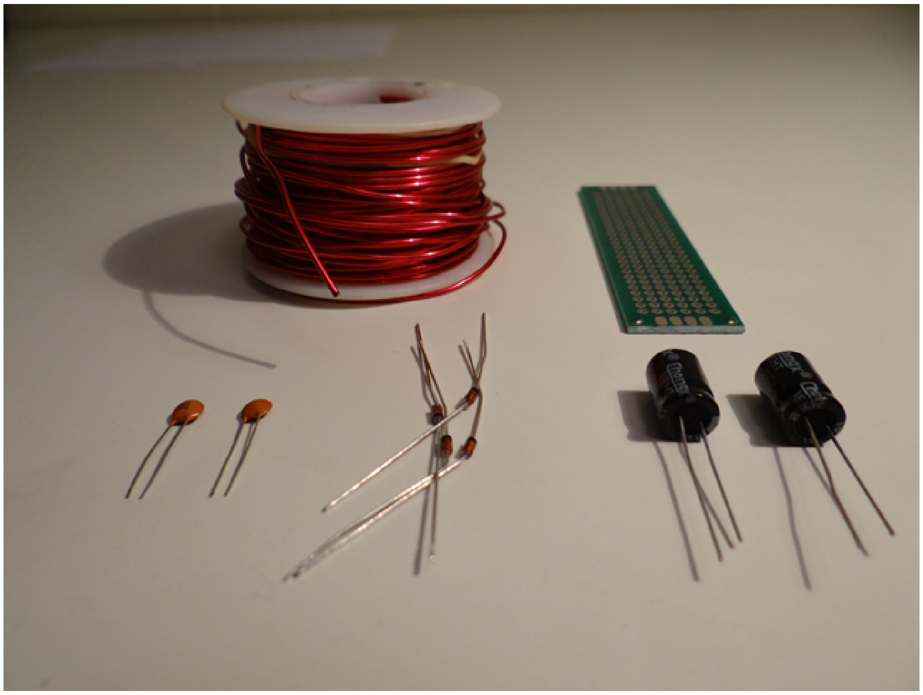

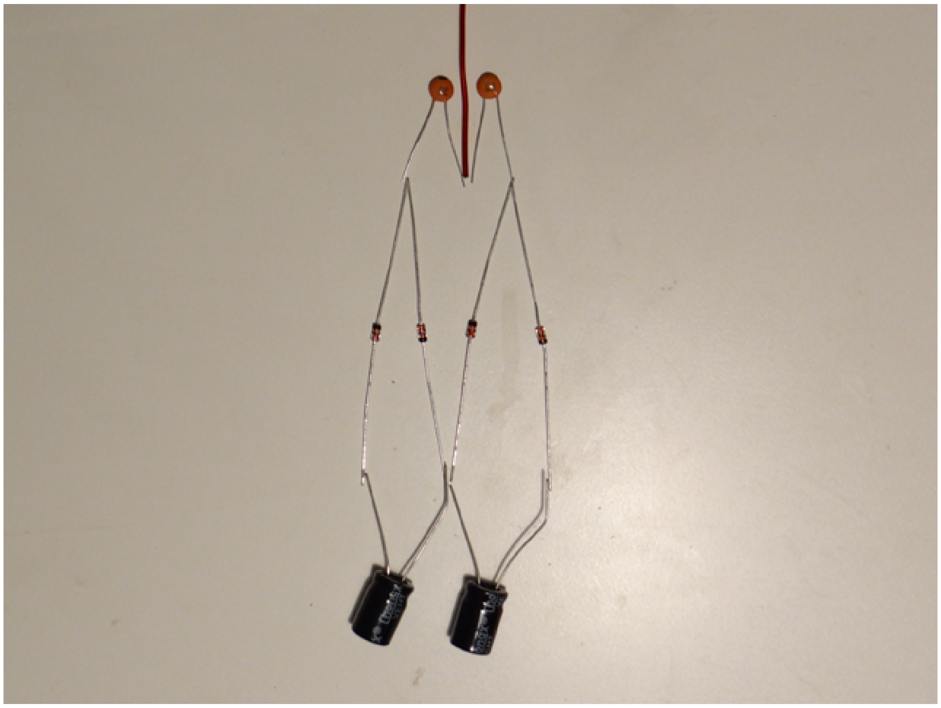

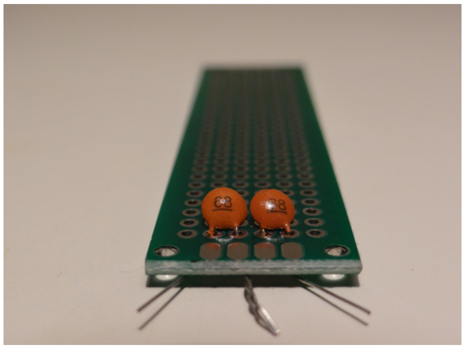

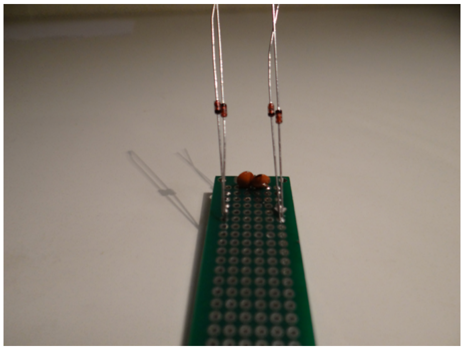

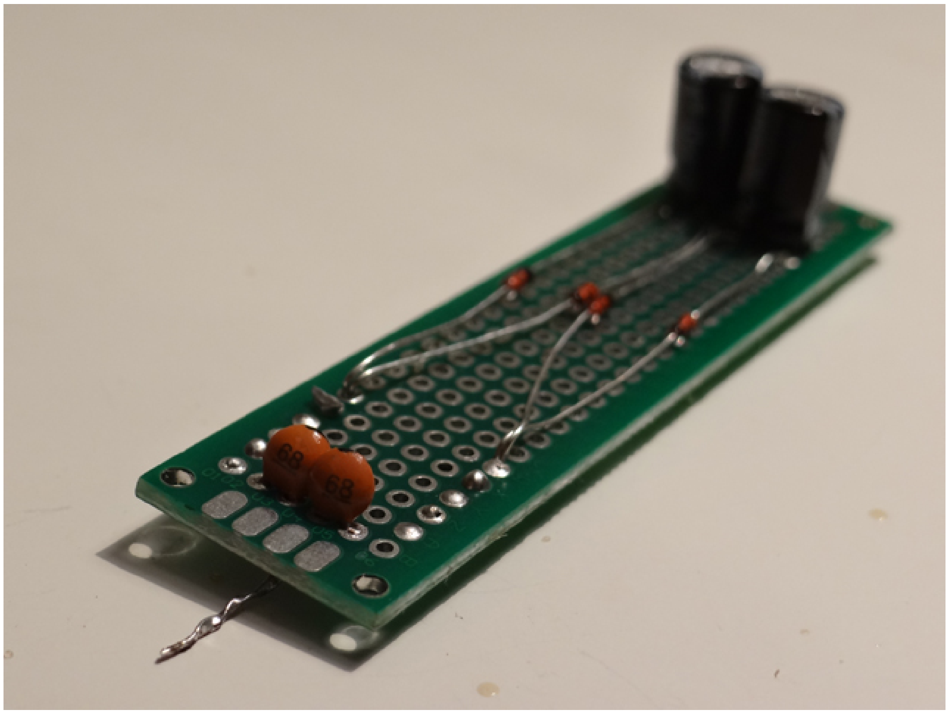

We're going to build the simplest version of this circuit variation in order to understand how each component interacts and as a proof of concept.

There are three simple systems at work in the circuit that are composed of capacitors, which store energy, and diodes that direct it.

Energy in the band of radio waves, among others, will vibrate a wire antenna on an atomic level, sending a discernible signal to its lead. This signal will then meet the junction between two ceramic capacitors wired in series. This junction will force positive charge from the wave to travel in one direction and negative charge in the other direction which, when collected again, makes the signal uniform and polar. Connecting the two capacitors in series creates leads on each end; the now positively charged side of one and the now negatively charged side of the other creates a two cell battery. This stage of the circuit takes a signal with a net value of zero, adds the absolute values of the positive and negative amplitudes with respect to the origin and produces a positive integer. This concept can be thought of as taking:

$$(+1) + (-1) + (+1) + (-1) = \mathbf{0}$$

and converting it to:

$$| [(+1) + (+1)] | = \mathbf{2}$$
$$\mathbf{+}$$
$$| [(-1) + (-1)] | = \mathbf{2}$$
$$\mathbf{= 4}$$

Isn't math fun?

To each of these leads from our two capacitors in series, we will connect two crystal diodes, one facing each direction, to form what is called a bridge rectifier. A bridge rectifier is a configuration which will convert an alternating current to a direct one by cleverly rerouting the signal.

By connecting the bridge rectifier as shown in the circuit diagram, this direct current from the diodes then charges the electrolytic capacitors. This stage normalizes the amplitude, making the current constant and usable.

Components can easily be twisted together for testing and then soldered to a circuit board to secure.

Test and Optimize

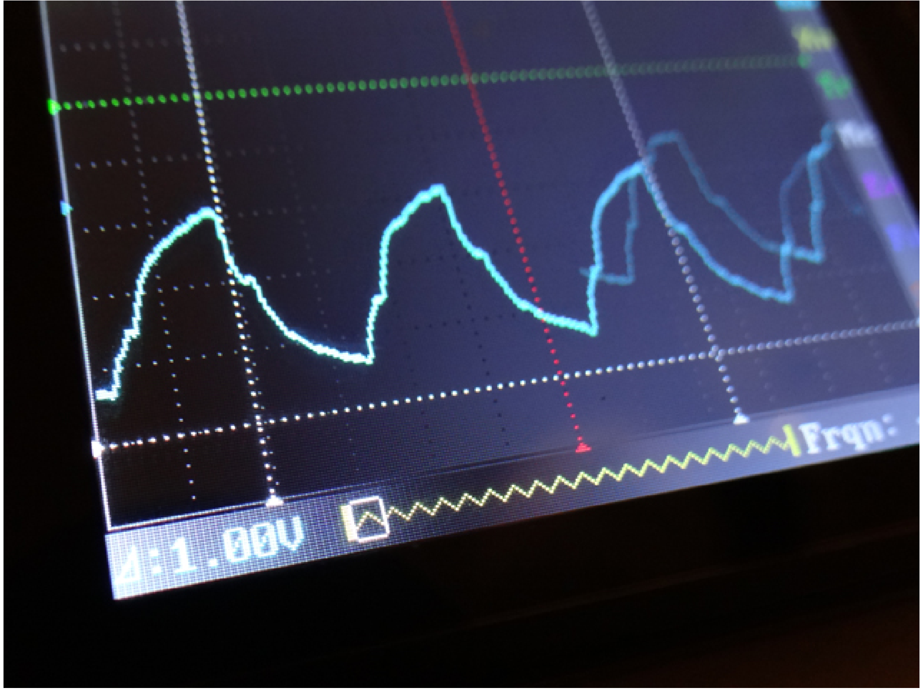

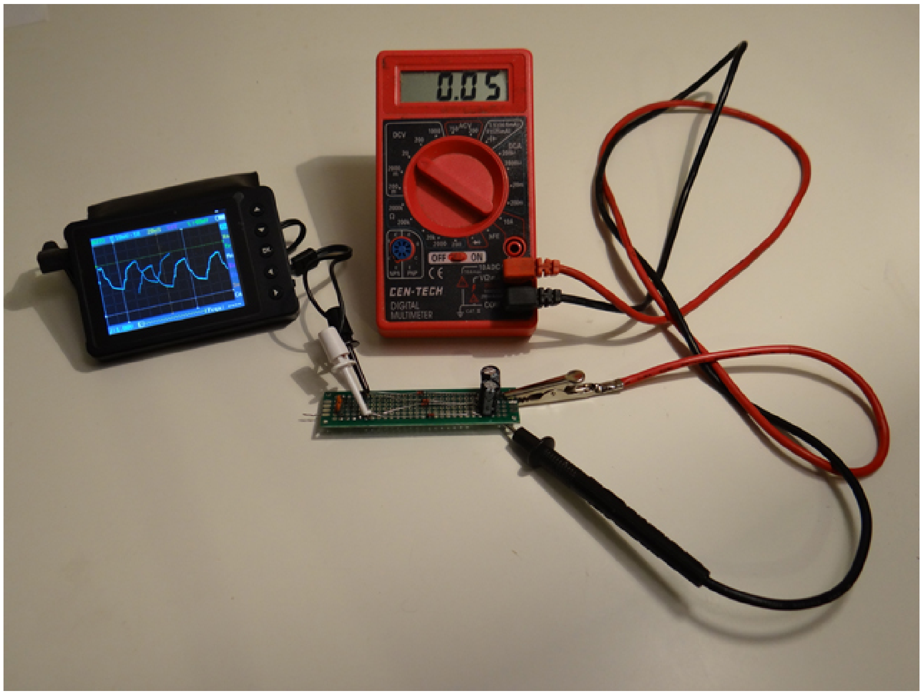
0.05
CEN-TECH
DIGITAL
MULTIMETER
OFF
ON

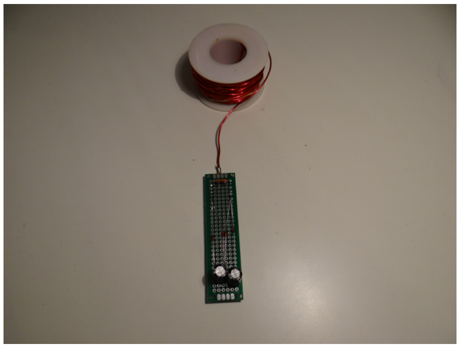

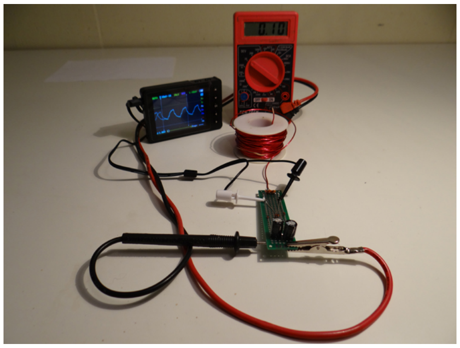
0.10

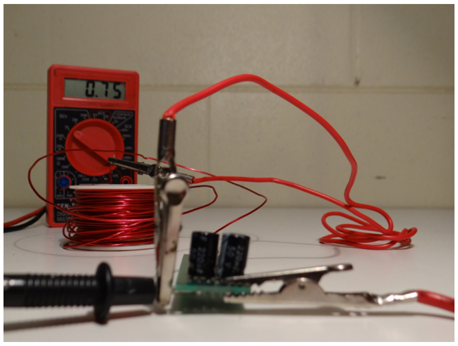
0.75

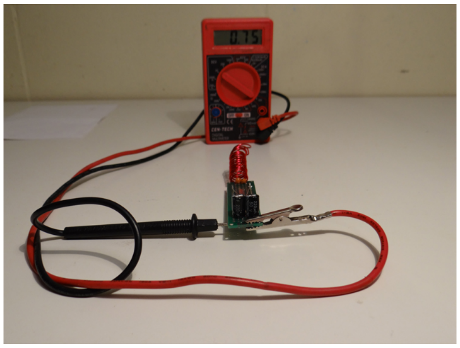
0.75

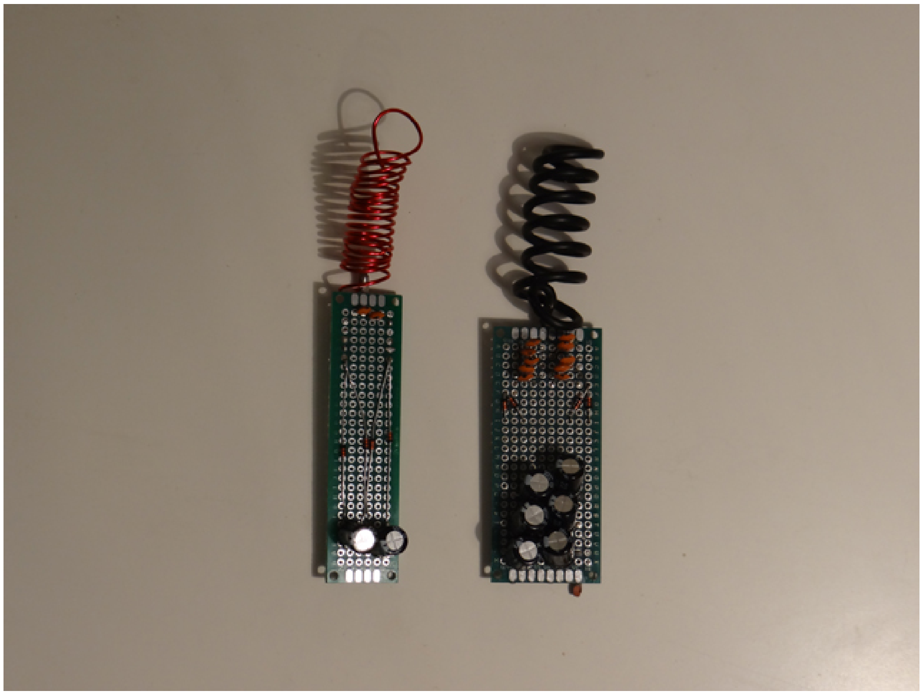

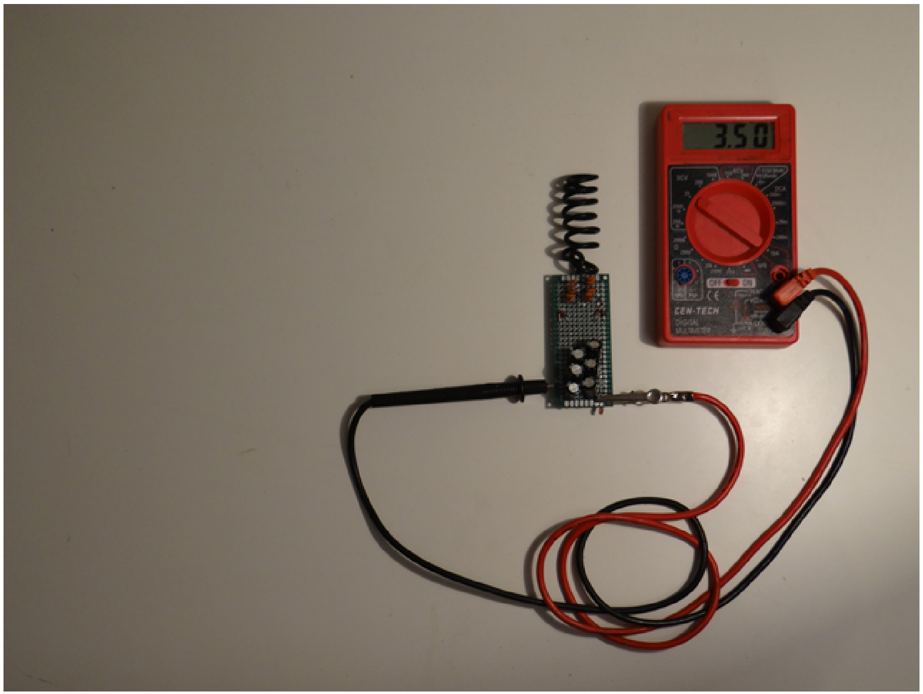

Note: Photos depict tests performed in an open-air urban environment.

To test and analyze our circuit, we'll be using a digital voltmeter and oscilloscope.

Then, by connecting a voltmeter to the output, we'll immediately begin to see a small voltage climbing in the 10-100mV range. If not, we'll want to check our connections and make sure the circuit is not isolated from the environment by taking it outside to a clear area.

By connecting an oscilloscope to the outside leads of our two ceramic capacitor bank, we will see the the polarized signal being captured from the air around us. We can then connect after the diodes to see our varying direct current and then to after the electrolytic capacitors to see a normalized, usable direct current at our output.

We can then optimize the input resistance in two ways. Firstly, we can add additional ceramic capacitors in parallel to our original two and make sure our soldered connections are consistent and thick in this area.

We can optimize the circuit's capacity by adding electrolytic capacitors in parallel to our original two which will allow this circuit to charge slightly when not in use. For this purpose, a charging circuit can also be added here in order to incorporate an optional battery bank.

We can optimize the antenna by attaching loops and coils of copper wire in various positions, store-bought antennas or by stringing some wire up to the highest point you can reach.

We don't have to stop there, either. We can now connect multiple circuits in series to increase voltage or in parallel to increase current. This can be done indefinitely.

Add a Case and Antenna

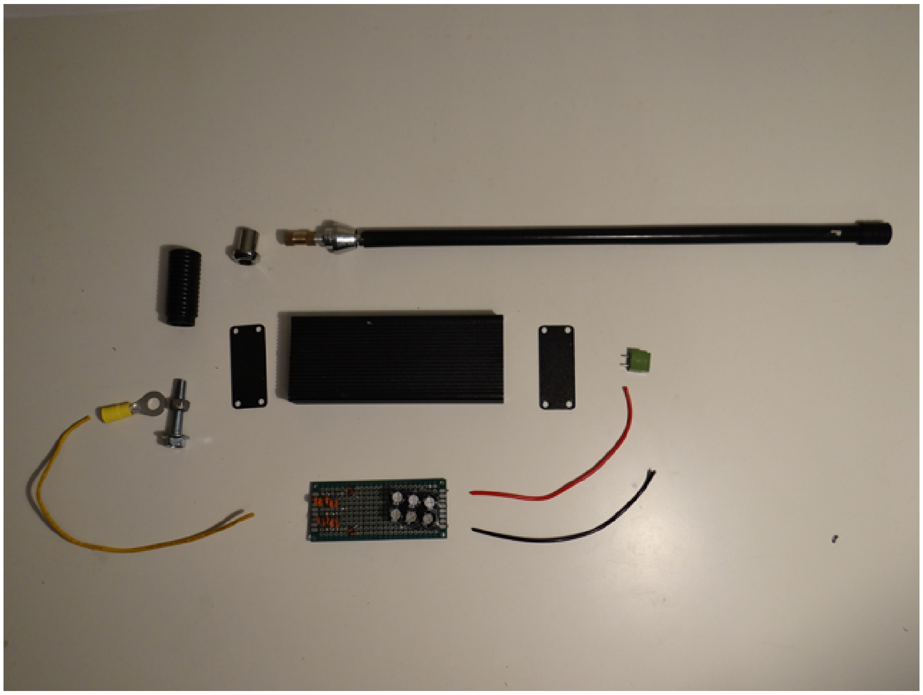

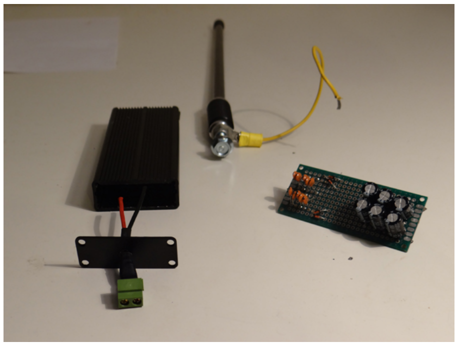

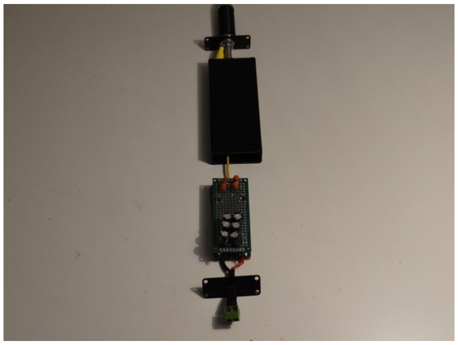

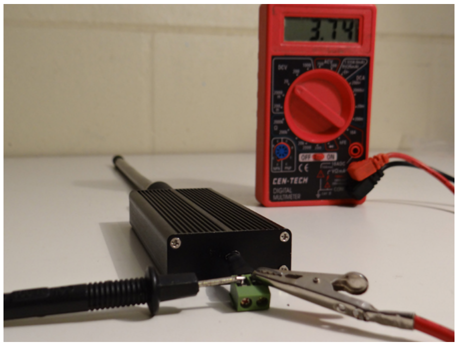
3.741
CEN-TECH

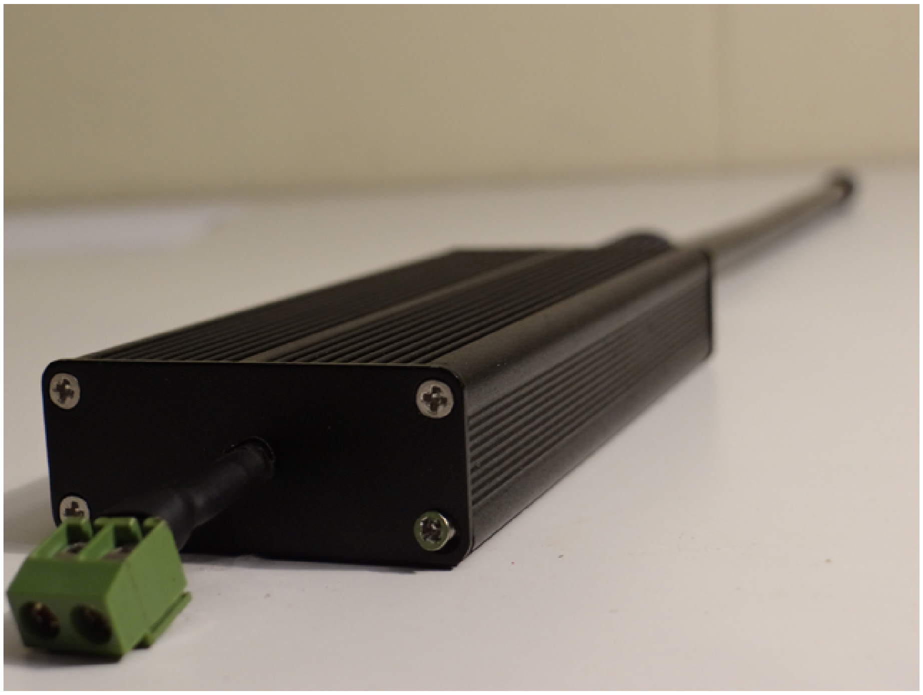

After choosing an antenna in the last step we'll now want to permanently wire it. Whether you choose a compact antenna for portability or a tall fixed antenna for power and range, we will wire it in the same manner according to the diagram in the previous step. Note that the input on the configuration here is grounded to the metallic case, and thus the users hand, and incorporation of longer antennas will require proportionally more substantial grounding.

We will then attach a terminal to the output to allow us to connect this circuit to an electrical device or charging circuit and battery bank.

Next, we will add a case, making sure to isolate exposed leads with non-conductive material especially if mounting in a conductive case. A piece of cardboard secured with glue is sufficient for the circuit's bottom and shrink wrap or electrical tape can be used in the case of any additional exposed leads. Drill two holes in your enclosure, one for the antenna or antenna lead and another for your output terminals. You can then insert your components, fasten the enclosure and your device is ready to use!

Your Crystal Energy Receiver© is Complete

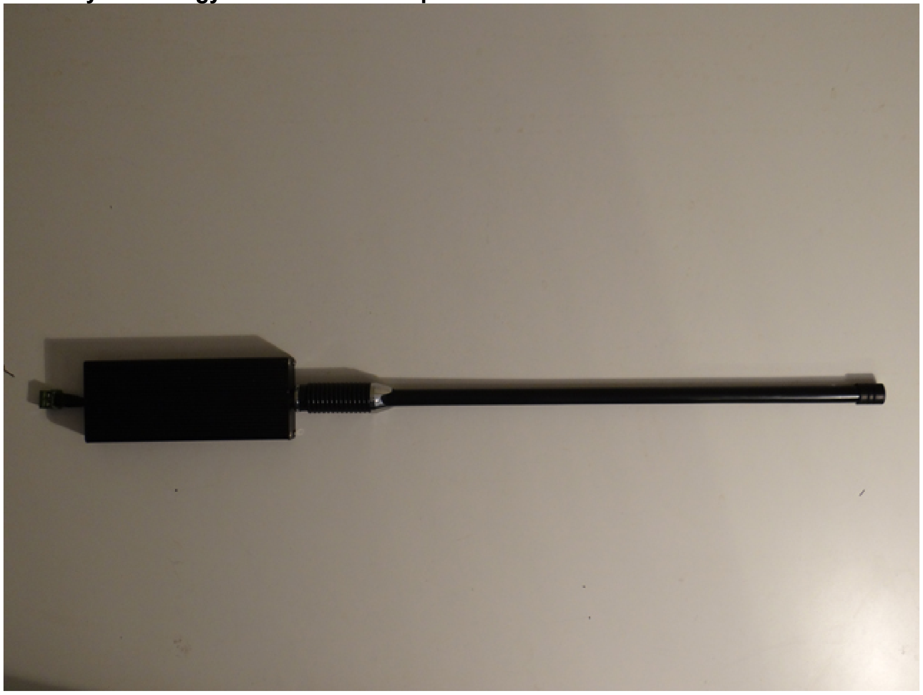

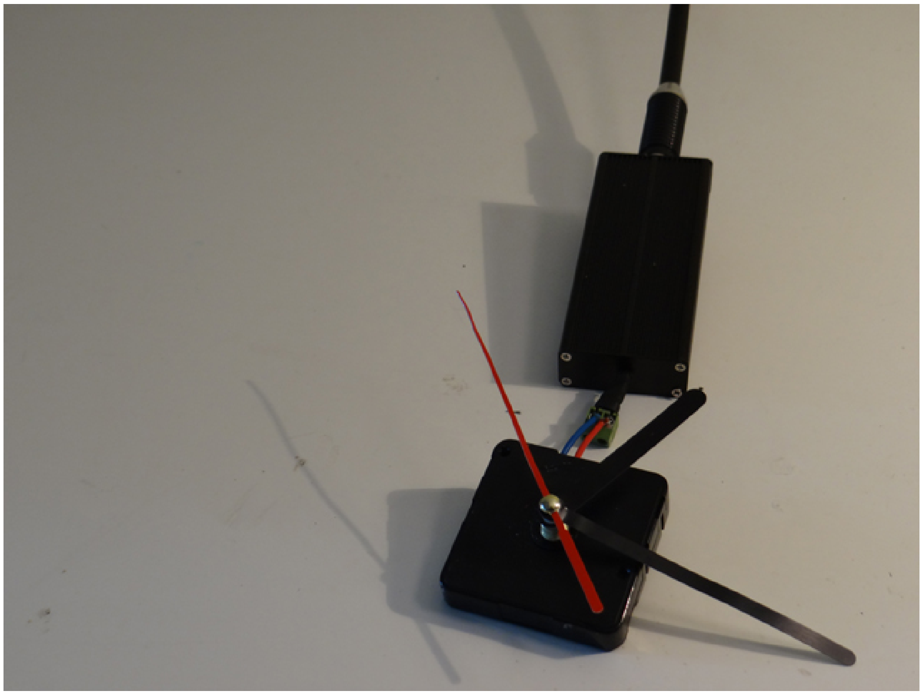

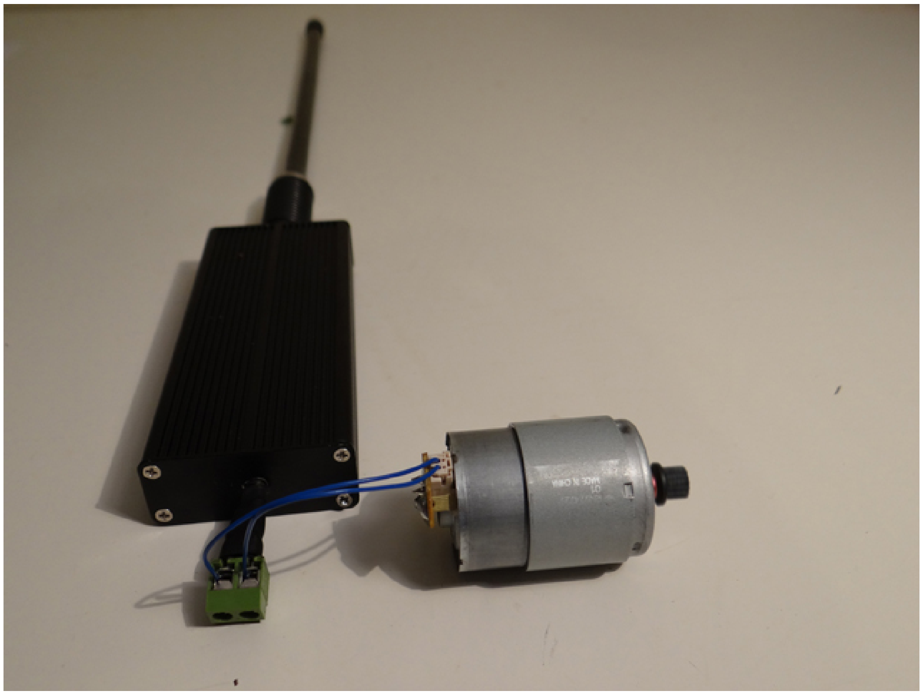

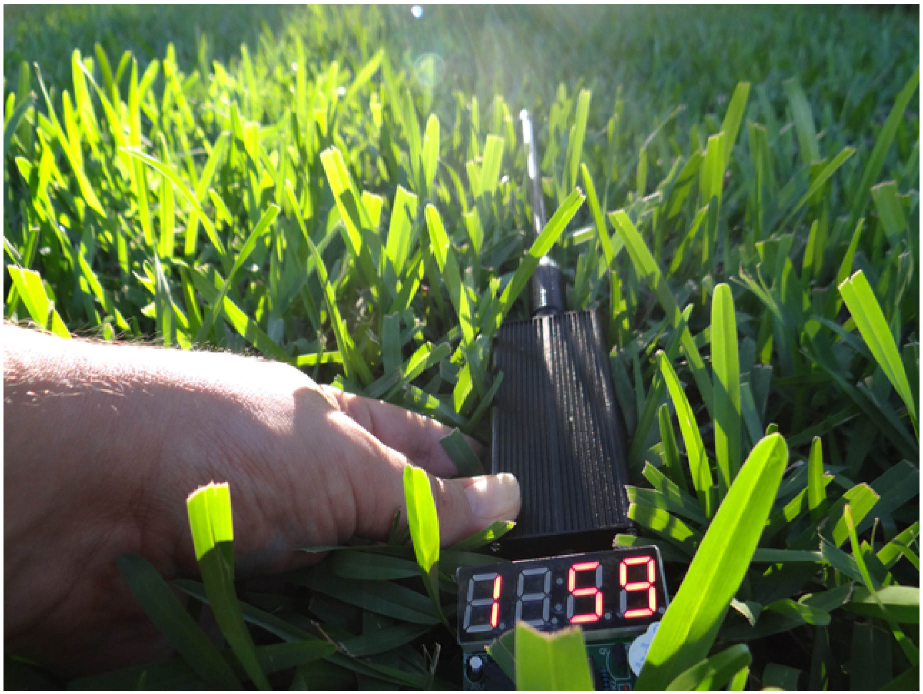

It is hour 11:59; the energy revolution is upon us!

Your Crystal Energy Receiver© is now complete and ready to use!

I built a portable version, for proof of concept and demonstration purposes. However, you can go as big as you want- to passively charge batteries or run equipment remotely; or go as small as you want- to power sensors, RFID devices, small electronics and more.

I used this harvested energy to easily power a low-consumption quartz clock, a digital chronograph with integrated circuits and LCD and was even able to momentarily rotate a small dc motor.

Because of its simplicity this device is a durable, efficient and reasonably effective method of harvesting radiant energy in a simple, replicable and sustainable way. I humbly hope that the contributions made here, and by those reading, can be one day used by people worldwide to conveniently capture free energy.

Want to build it? **Get the Kit!**

See more **Projects**

Copyright 2016. Drew Paul. All rights reserved.

Please take caution. User is responsible for safe operation of components.

WEB | PRINT | MARKETING | MEDIA
by
blabor.com

Donald Lee Smith

Don Smith is elderly and has suffered several serious strokes in the last few years. He is no longer able to respond to e-mails and his web site has been taken over by people who appear very keen to ensure that his work and information can no longer be accessed by members of the public.

At this point in time there is very little information available on Don's achievements. What I have been able to locate is four video recordings of lectures which he gave, plus a copy of his pdf document entitled "Resonance Energy Systems". There is also a web site with limited information. This site appears to have been set up by Don's son whom I suspect does not fully understand how his father's devices work. Don has produced at least forty eight different devices which draw energy from what Don prefers to call "the ambient background". His devices are capable of supplying kilowatts of excess energy and in most cases they do not require any input energy to be supplied by the user.

It is said that 40,000 copies of Don's pdf document have been sold worldwide, but it appears that it is no longer possible to buy a copy, and for that reason, a copy is contained within this document. Please be aware that Don states quite openly that he does not disclose all of the details on any of his devices in his public discussions. This is because the rights to each device have been assigned to a different company in which Don has a financial interest, and so it is not in his interests to disclose the full details. However, ha says that he discloses enough for somebody who is experienced in radio-frequency electronics to be able to deduce the things which he does not disclose and so build a device for his own use. If that is the case, then anybody who has succeeded in doing so has kept very quiet about it afterwards (which is understandable). Having said that, Tariel Kapanadze of Georgia appears to have replicated one of the designs although it is quite likely that Tariel deduced the operating principles for himself. These principles are clearly based on the work of Nikola Tesla.

In Don's work, he refers to the educational kit entitled "Resonant Circuits" No 10 - 416 from The Science Source at www.thesciencesiource.com. That kit now costs US$100 and they don't mention the fact that you need a signal generator capable of producing a 10 volt output at up to 1 MHz in order to be able to do resonance tests, and the kit now uses an LED rather than a neon bulb.

There is a "book" of Don's entitled "An Answer to America's Energy Deficit" but, being an earlier work than his pdf document, it does not appear to contain anything extra of any significance. Here is Don's pdf document, in it's original American-language wording with just minor clerical corrections:

RESONANCE ENERGY METHODS

Donald L. Smith
TransWorld Energy, CEO
September 23, 2002

Fax/Phone 281-370-4547 and e-mail donsm1@earthlink.net

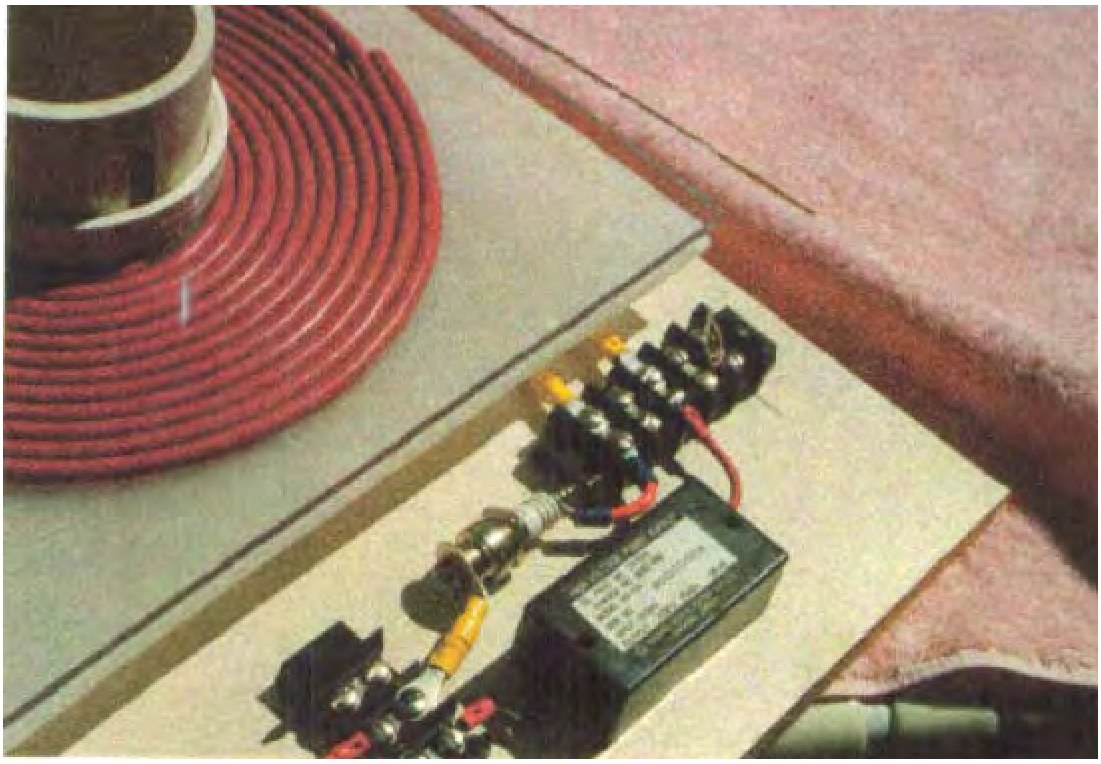

DANGER
HIGH VOLTAGE

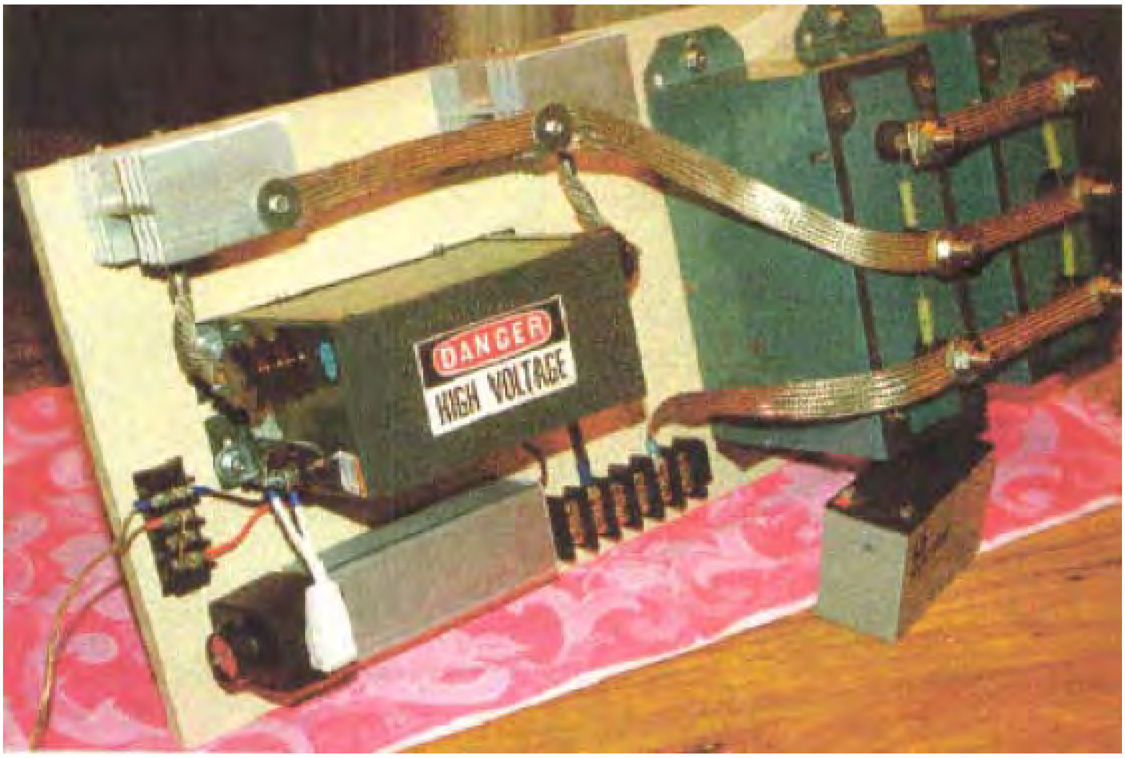
DANGER
HIGH VOLTAGE

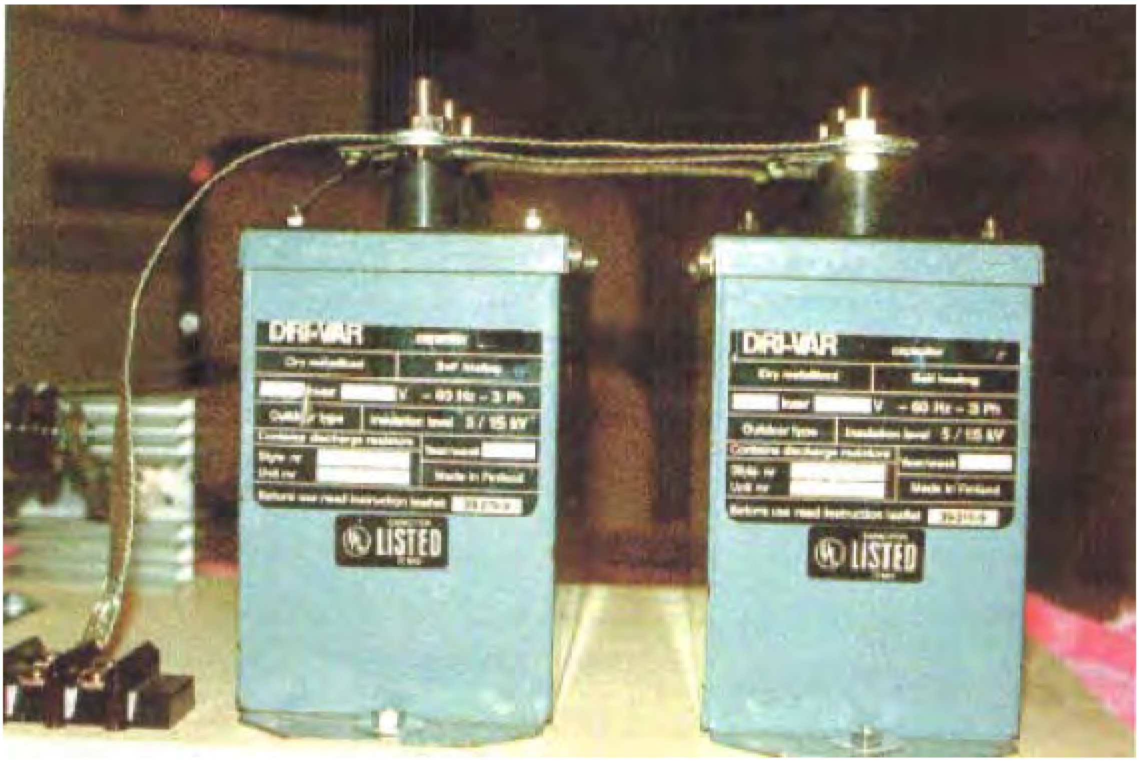
DRI-VAR capacitor
Dry metallized
Self healing
V - 60 Hz - 3 Ph
Outdoor type
Insulation level 5 / 15 kV
Contains discharge resistors
Made in Finland
Before use read instruction leaflet
LISTED

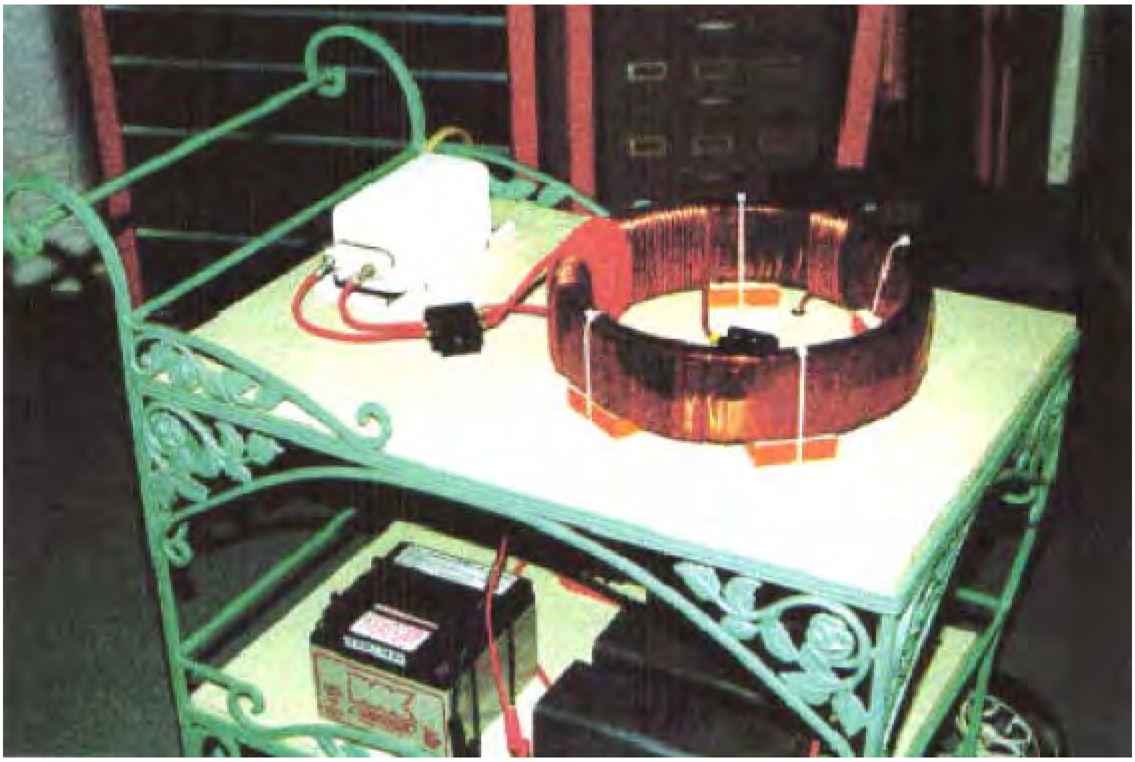

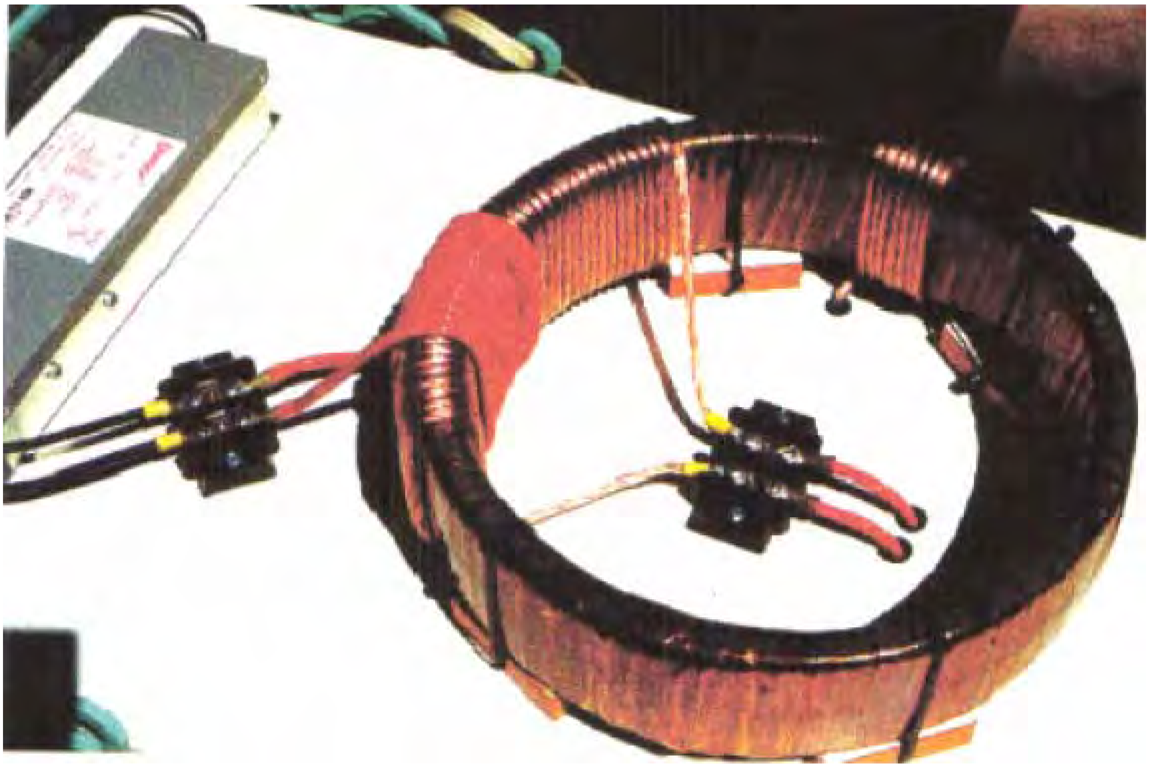

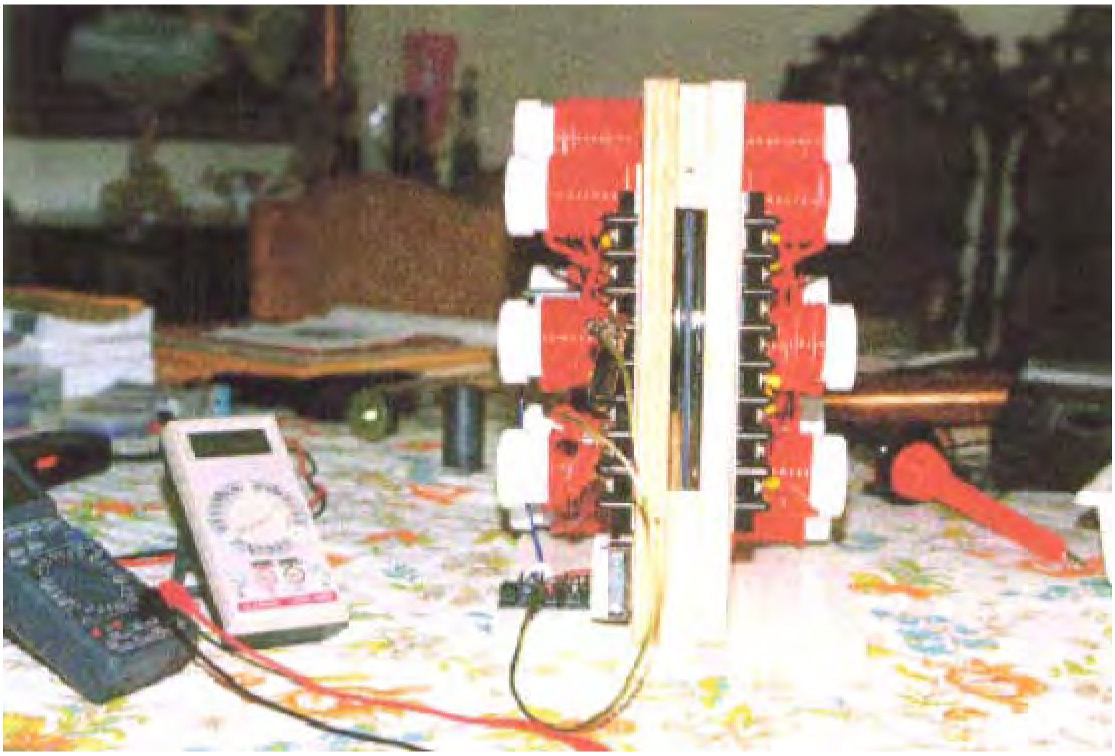

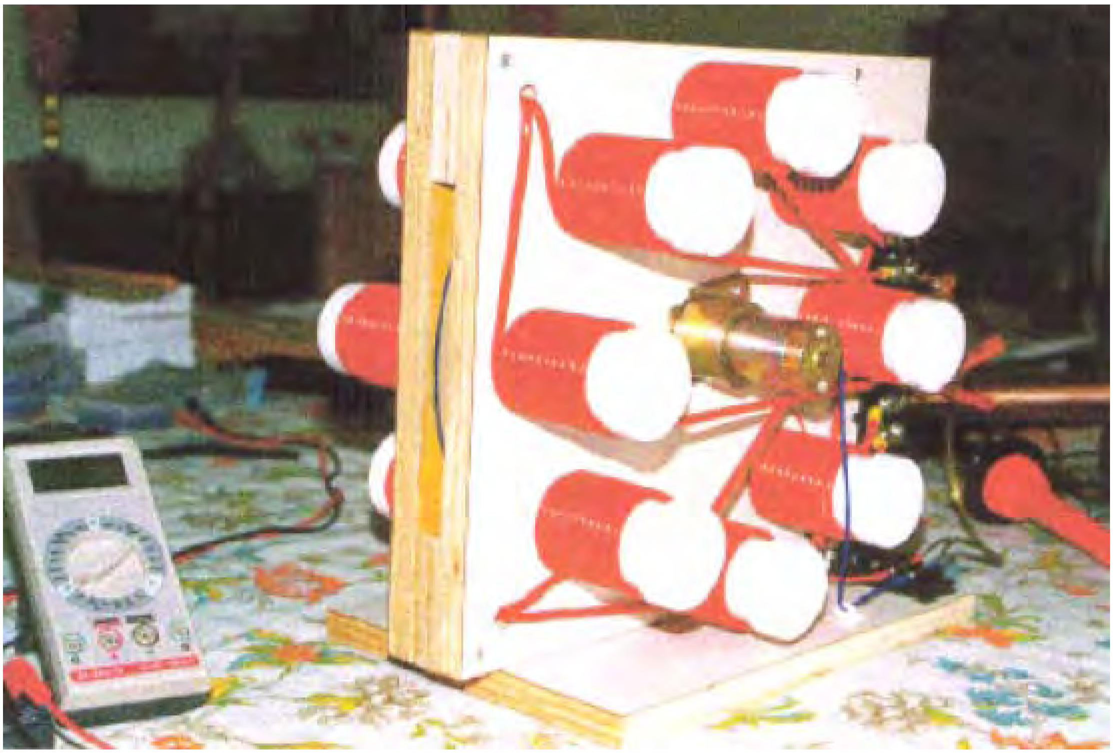

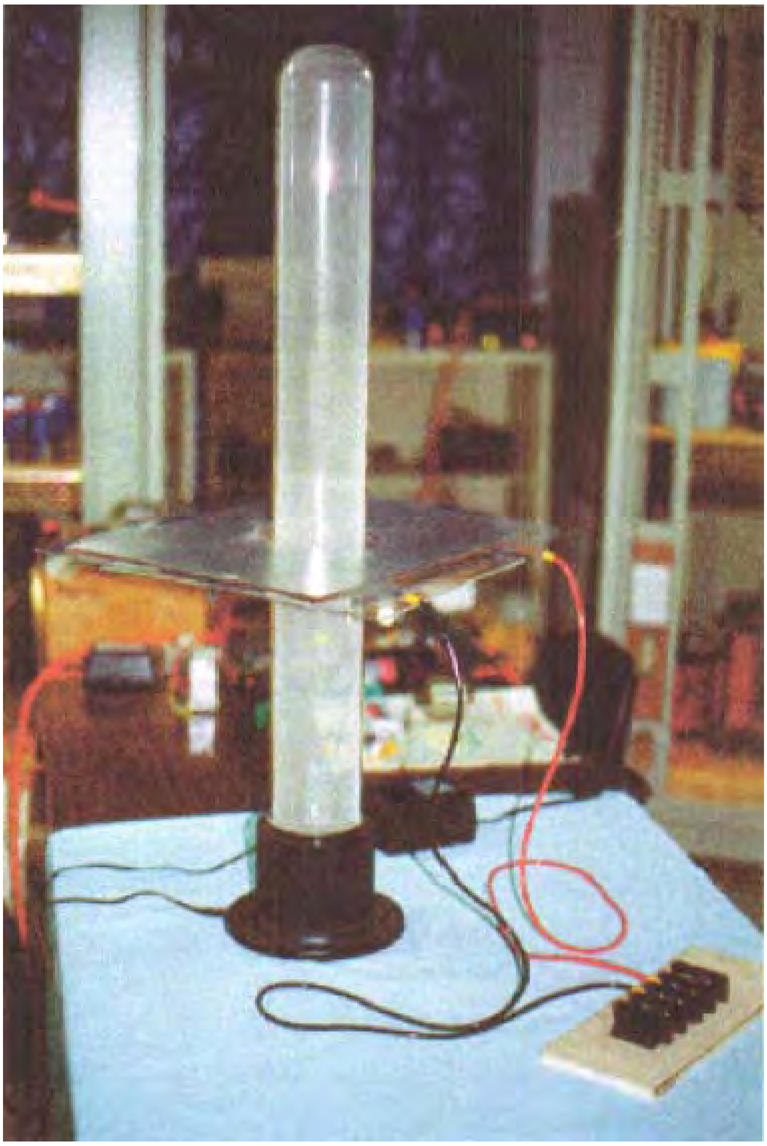

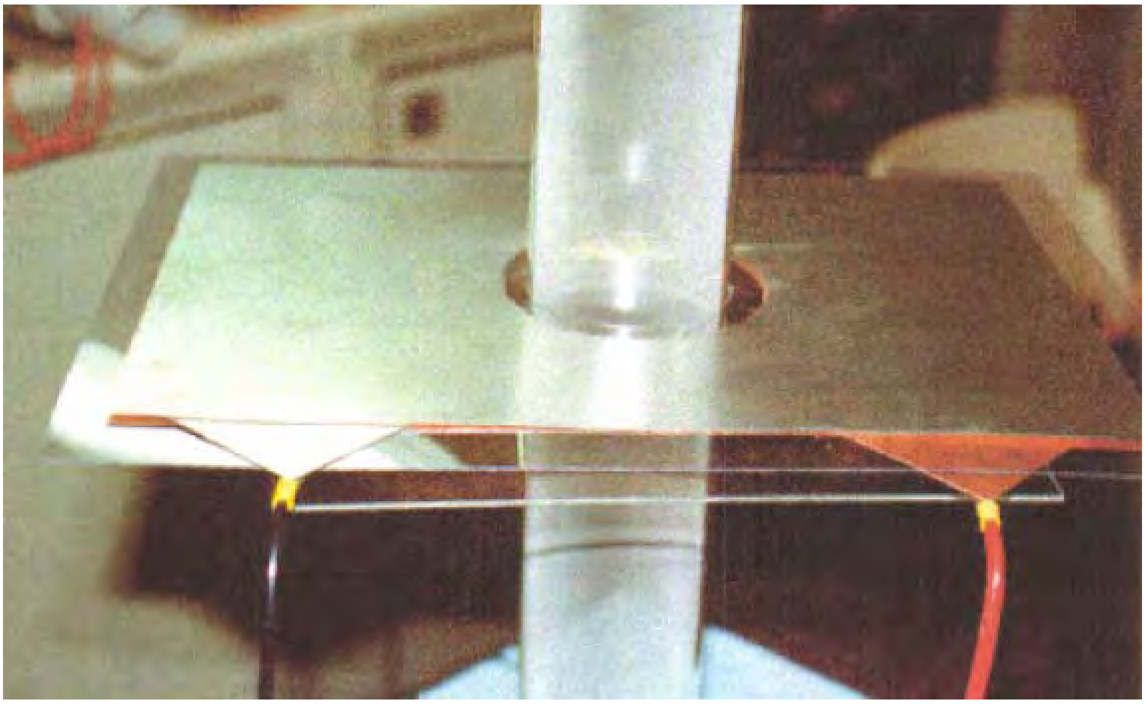

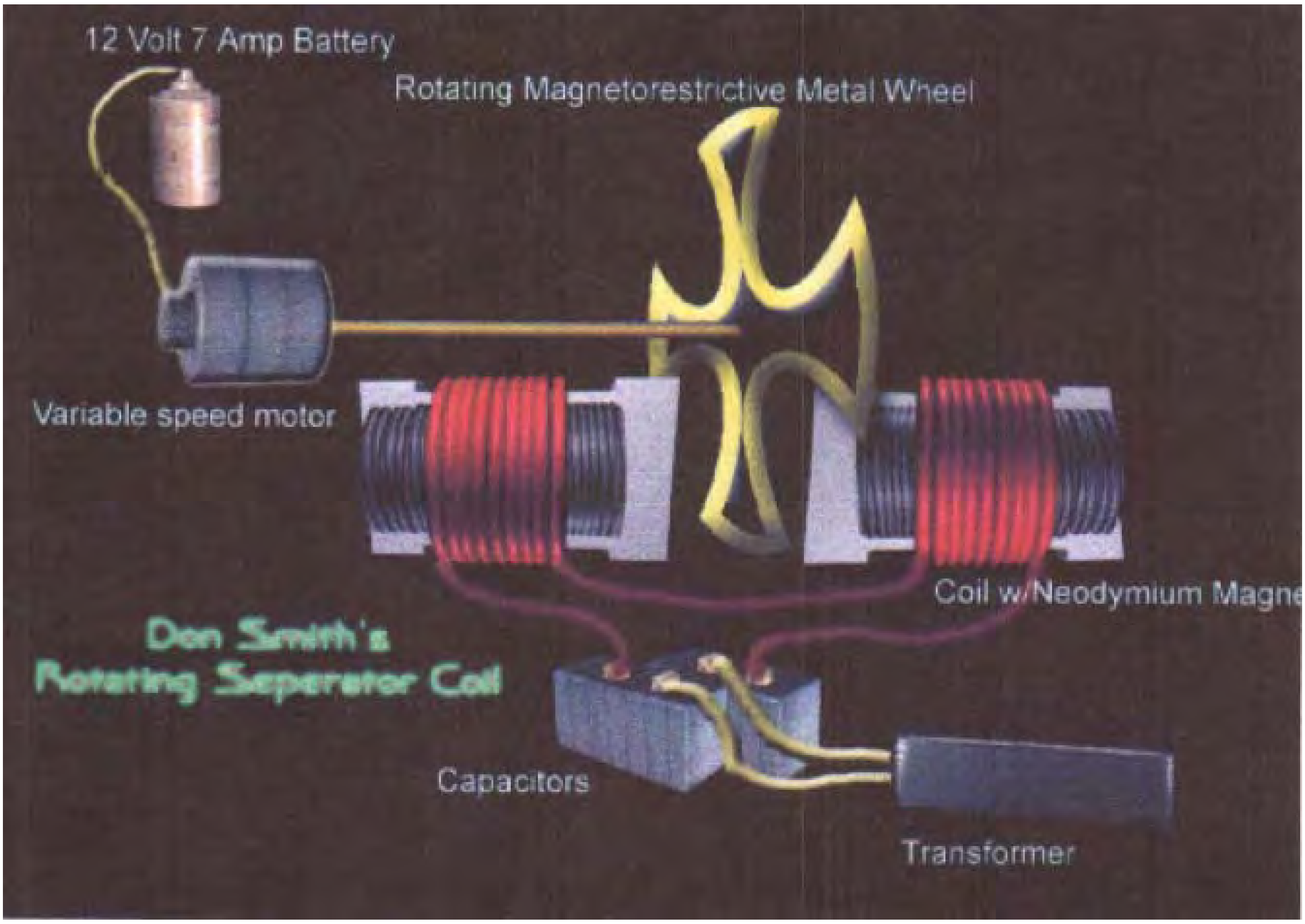
12 Volt 7 Amp Battery
Rotating Magnetorestrictive Metal Wheel
Variable speed motor
Coil w/Neodymium Magne
Don Smith's
Rotating Seperator Coil
Capacitors
Transformer

DIPOLE TRANSFORMER GENERATOR
DESCRIPTION

TECHNICAL FIELD:

The Invention relates to loaded Dipole Antenna Systems and their Electromagnetic radiation. When used as a transformer with an appropriate energy collector system it becomes a transformer generator. The invention collects and converts energy which, with conventional devices, is radiated and wasted.

BACKGROUND ART:

An International search of Patent Databases for closely related methods did not reveal any prior Art with an Interest in conserving radiated and wasted magnetic waves as useful energy.

DISCLOSURE OF INVENTION:

The Invention is a new and useful departure from transformer generator construction, such that radiated and wasted magnetic energy changes into useful electrical energy. Gauss Meters show that much energy from conventional electromagnetic devices is radiated back into the ambient background and wasted. In the case of conventional transformer generators, a radical change in the physical construction, allows better access to the energy available. It is found that creating a dipole and Inserting capacitor plates at right angle to the current flow, allows magnetic waves to change back to useful electrical (coulombs) energy. Magnetic waves passing through the capacitor plates do not degrade and the full impact of the available energy is accessed. One, or many sets of capacitor plates, may be used as desired. Each set of plates makes an exact copy of the full force and effect of the energy present in the magnetic waves. The originating source is not depleted or degraded as is common in conventional transformers.

BRIEF DESCRIPTION OF THE DRAWINGS:

The Dipole at right angle allows the magnetic flux surrounding it to intercept the capacitor plate, or plates, at right angles. The electrons present are spun in such a way that the electrical component of the electrons is collected by the capacitor plates. Essential parts are the South and North component of an active Dipole. Examples presented here, exist as fully functional prototypes, and were engineer constructed and fully tested for utility by the Inventor. Corresponding parts are utilized in each of the three examples as shown in the Drawings.

DRAWING 1 OF 4: VIEW OF THE METHOD

N = North and S = South of the Dipole

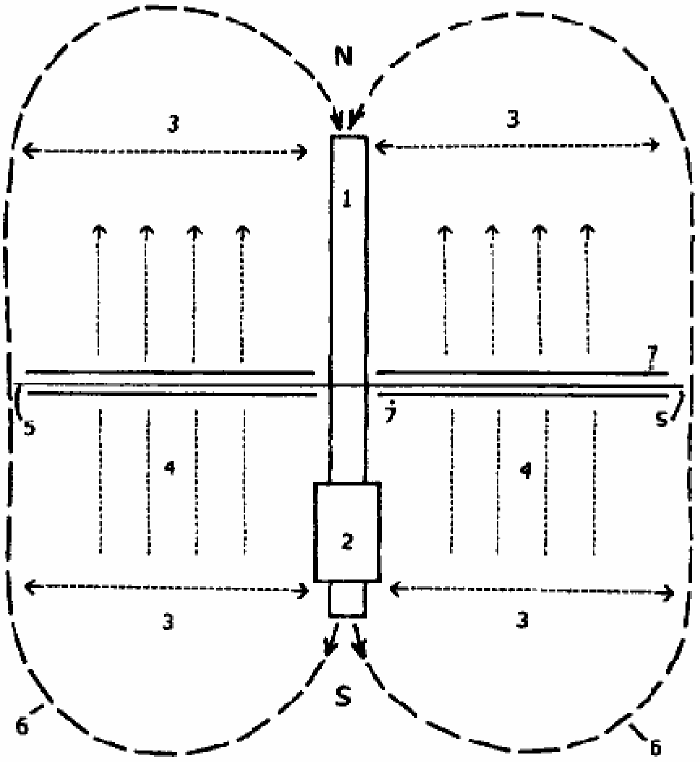

1. North and South component of the Dipole.
2. Resonate High Voltage induction coil.
3. Dipole's electromagnetic wave emission.
4. Heaviside current component.
5. Dielectric separator for the capacitor plates
6. For purposes of the drawing, a virtual limit of the electromagnetic wave energy.
7. Capacitor plates, with dielectric in between.

DRAWING 2 OF 4: COMPONENTS, 2A and 2B

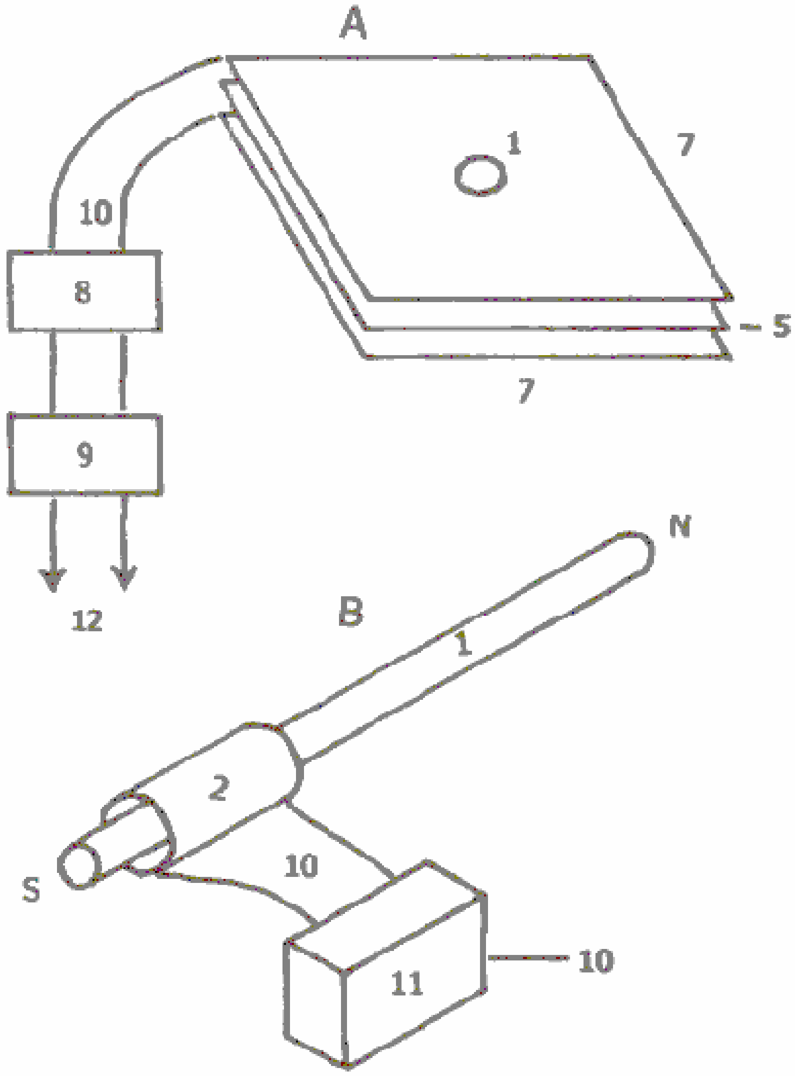

Fig.2-A:

1. Hole for mounting Dipole **B-1**.
2. Resonate high voltage induction coil.
5. Dielectric separator, a thin sheet of plastic separating the capacitor plates.
7. Capacitor plates, upper plate is aluminum and lower plate is copper.
8. Battery system, deep cycle.
9. Inverter. Input: Direct Current, output: 120 Volts at 60 cycles per second.
10. Connector wires.
12. Output to point of use being the load.

Fig.2-B N = North and S = South component of the Dipole

1. Metal rod, being soft magnetic metal such as iron.
2. Resonate high voltage induction coil.
10. Connector wires.
11. High Voltage input energy source such as a neon tube transformer.

DRAWING 3 OF 4 : Proof of Principle Device, using a Plasma Tube as an active Dipole.
N = North and S = South Components of the active Dipole.

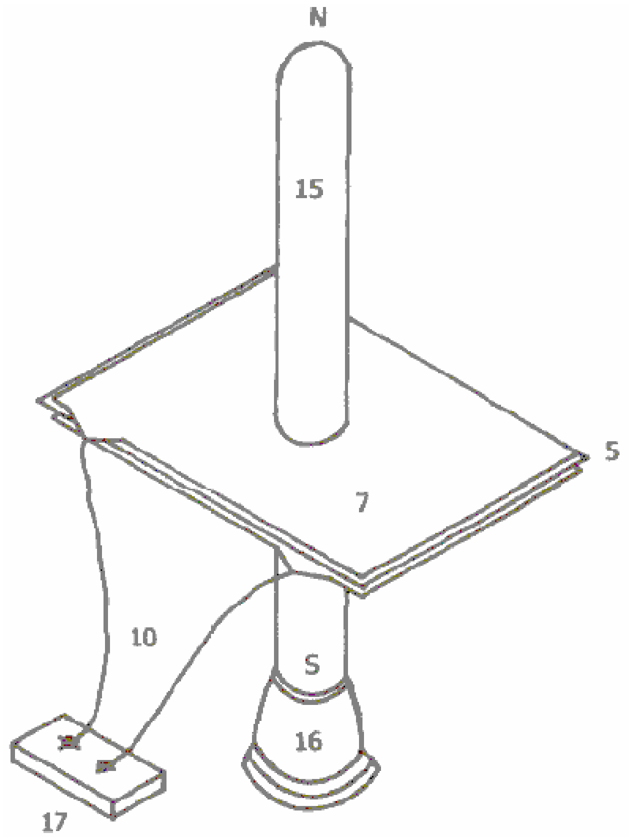

5. Dielectric separator of the capacitor plates.
7. Upper capacitor plate: upper plate is aluminum and lower plate is copper.
10. Connector wires.
15. Plasma Tube, 4 feet long and 6 inches in diameter.
16. High Voltage Energy source for the active Plasma Dipole.
17. Connector block: outlet for testing and use.

DRAWING 4 OF 4: Manufactures Prototype, Constructed and fully tested.

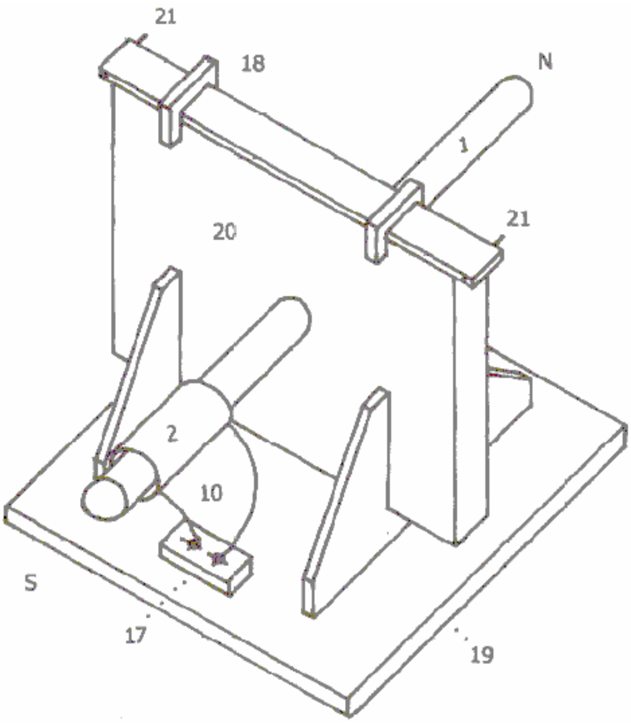

1. Metal Dipole rod.
2. Resonate High Voltage induction coil.
10. Connector wires.
17. Connector block for Input from high voltage energy source.
18. Clamps for upper edge of capacitor packet.
19. Support Device for The Dipole Transformer Generator.
20. Packet of Capacitor Plates.
21. Output connectors of the capacitor, producing energy into a deep cycle battery which then powers the inverter.

BEST METHOD OF CARRYING OUT THE INVENTION:

The Invention is applicable to any and all electrical energy requirements. The small size and high efficiency makes It an attractive option. It is particularly attractive for remote areas, homes, office buildings, factories, shopping centers, public places, transportation, water systems, electric trains, boats, ships and all things small or great Construction materials are readily available and the skill level required is moderate.

CLAIMS:

1. Radiated magnetic flux from the Dipole, when intercepted by capacitor plates at right angles, changes to useful electrical energy.
2. A Device and method for converting for use, normally wasted electromagnetic energy.
3. The Dipole of the Invention is any resonating substance such as Metal Rods, Coils and Plasma Tubes which have interacting Positive and Negative Components.
4. The Resulting Heaviside current component Is changed to useful electrical energy.

ABSTRACT

An Electromagnetic Dipole Device and Method, wherein, radiated and wasted energy is transformed into useful energy. A Dipole as seen in Antenna Systems is adapted for use with capacitor plates such that the Heaviside Current Component becomes a useful source of electrical energy.

TransWorld Energy
227 W. Airtex Blvd.
Houston, Tx. 77090

September 23.2002

Phone: (281)876-9200
Fax: (281)876-9216

Web-site: altenergy-pro.com
E-mail: donsm1@earthlink.net

Dear Reader:

TransWorld Energy is dedicated to improving the Human Condition in the Field of Energy which, at the same time, makes possible Healthy Water and increases the food Supply. A never-ending source of energy found throughout the universe is easily accessed with the minimum of effort and cost. The technology for doing this has been around since the 1820s. Selfish special interests have made sure that the technology remains discredited. People who control the Energy Sources control the World.

Extensive research and development by TransWorld and Associates has been progressing for more than 15 years. Numerous successful Energy Producing Devices have been produced and demonstrated throughout the World. Some of these can be viewed by the Web Site located using any major search engine (such as Lycos, Yahoo, Altavista, NorthenLight and more than 2,000 others throughout the World).

The Book which You are viewing has more than 40,000 copies in circulation. It has been translated and distributed in all major languages including Japanese, Arabic, Portuguese, French, Italian, Russian, Chinese, German, Spanish and many more. There are seven editions in circulation. An enormous interest is evident in the subject matter. An average of about fifty e-mails per day are received from the ends of the Earth (that is about 1,500 per month).

Once the Web Site and the book are viewed, it will become evident that abundant, self-sustainable energy is available everywhere for the taking. This is natural energy which does not harm the environment or those using it. The proper Device for Collecting is all that's required.

The Good News is that the problem is solved and with assistance, an ultimate source of energy which is environmentally benign, abundant throughout the universe and inexpensive to capture, is there for the taking.

Thank You for your consideration

Donald L. Smith, CEO

MEM

Electrical Energy Generating System

Description and Function:

The Generation of Electrical Power requires the presence of electrons with various methods of stimulation, yielding magnetic and electrical impulses, collectively resulting in Electrical Energy (Power). In place of the mechanical - coils and magnet system, present in conventional electrical power generation, visible moving parts are replaced by resonate magnetic induction, using radio frequency. Transfer of energy by resonate induction is related to the ratio of the square of the cycles per second.

The Energy System, presented here, operates at millions of cycles per second verses the conventional 60 C.P.S. This tells us that it has a frequency advantage over conventional methods. This same advantage applies to the amount of electrical energy output. Therefore the Device is small in size and produces large amounts of Electrical Energy. The Electrons acquired, are from the surrounding Air and Earth Groundings, being the same source as in conventional methods. This is accomplished by magnetic resonate radio induction.

Applications:

This Electrical System adapts nicely to all Energy Requirements. It is a direct replacement for all existing Energy Systems. This includes such things as Manufacturing, Agricultural, Home Usage, Office Complexes, Shopping Centers, Rail Transportation, Automobiles, Electrical Power Grids, Municipalities, Subdivisions, and Remote Areas. Briefly, the only limiting factor is the imagination.

Economic Possibilities:

No Historical Reference Point exists for a comparison of the Possibilities of this System. One can see from the impacted applications listed above, that the magnitude exceeds any known invention, presently a part of the Human Experience.

Present and Future Plans:

The Energy System has been in the developmental stage during the past seven years. It is Patent Pending # 08/100,074 with the Patent Office. No prior art exists according to the Patent Office's response. The System is presently being introduced into the World Market.

Useful energy occurs as the result of imbalances in the ambient background energy, which is a transient phenomena. In the electrical field, it is a closed system subject to heat death, which severely limits it's utility. The flip side of the electron, produces magnetic waves which are an open system, not subject to heat death. These waves, being unrestricted, are the universal source of energy when unlimited resonate duplicates from this one source are available. Therefore, the key to unlimited energy, is Magnetic Resonance. In order to understand this, requires putting a stake through the Heart of Antique Physics. Non-linear and Open Systems are universally available in Magnetic Resonance Systems, Explosions of any sort [including Atomic Explosions] and Combustibles of any type. Mechanical equivalents would be levers, pulleys and hydraulics. A highly obvious example is the Piano where the Key impacts the one note giving one sound level, which resonates with it's two side keys providing a much higher sound level. Magnetic Resonance Energy clearly amplifies itself, demonstrating more energy out, than in.

Ohmic resistance does not apply to Magnetic Resonance which travels unrestricted for great distances, therefore multitudes of electrons are disturbed, and their back-spin translates magnetic into usable electric energy. The right angle component which the magnetic flux provides, translates into useful electrical energy. Taken at right angles, the Magnetic Dipole provides an unlimited source of electrical energy. The writer is recognized world-wide for his knowledge and experience. See his Web Site at altenergy-pro.com .

Gravity is a function of spin phenomenon as observed in gravity separation of liquids. When spun, milk and cream separate. Therefore, relative specific gravity is function of mass versus spin. Magnetics and gravity are both spin related. In part, a top levitates when spun. Therefore, spinning magnetic fields are a functional motor source as in flying saucers.

ABSTRACT: Technology of New Energy:

Developments in the understanding of Electricity, along with Materials which were not previously available, allows the construction of Devices which collect energy in large quantities, from the Earth's Ambient Electrical Background. This Energy is naturally occurring, environmentally benign and is available everywhere. It is available wherever and whenever it is required. New Devices use Resonate Magnetic Waves which replicate upon spinning the locally present electrons, providing multiple duplicate copies of the Energy Present. Each electron when spun yields both magnetic and electric waves in equal proportion. The electrical component is a closed system limited by Ohms Law. The magnetic component is an open system not limited and it replicates multiple copies of the energy present.

Special materials and recent developments allow the magnetic energy to reproduce, through resonance, unlimited duplicate copies acquired from the ambient background. These Devices harvest the energy that has been, and is always present universally. Conventional methods consist of coils and magnets systems. Upon moving past each other, the magnetic flux field disturbs electrons which yield electricity, which is collected by the coils system. This is accomplished electronically with the new technology, without any moving parts and the energy is multiplied such that the Device becomes self-sustaining once it is started. This Technology, already presented Worldwide, will be shown at the Conference.

Dr. Smith
www.altenergy-pro.com
e-mail donsml@earthlink.net

"Putting a stake through the Heart" and thus removing the mental block created by antique physics is required. Conditions wherein this becomes necessary are non-linearity, resonance and explosions of any sort. Combustibles of any sort such gasoline and atomic explosions are good examples wherein more energy out than in, is obvious. You can add to that the non-linearity found in pulleys, hydraulics, steam power and suchlike. Magnetic resonance is a highly obvious source for multiplying energy output. The sound system present in the piano, demonstrates this very clearly. Energy amplification clearly present in the above, demonstrates the silliness attested to by many Physicists.

Ohmic resistance does not apply to magnetic resonance which travels unrestricted for great distances, therefore multitudes of electrons are disturbed, and their back-spin converts from magnetic energy to usable electric energy. These same electrons have been around from the beginning of time and they are undiminished and will remain so until the end of time.

ELECTRICAL ENERGY SYSTEMS PREFACE

Useful Electrical Energy is obtained directly from electron spin induced by incoming magnetic waves, or indirectly through mechanical exchange as in dynamo type devices. Simply put, electron spin converts from magnetic to electrical energy and vice versa. Nature provides grand scale magnetic wave induction throughout the universe, for free. In Electrical Systems, movement is at right angles to the direction of current movement. This explains the rotary movement of the Earth and other related Systems. The rate of Spin for the Earth is known as well as the mass (5.98 x 10^{24} Kg - "Physics for Scientist and Engineers", by Raymond A. Serway, Saunders College Publishing, 2nd Ed. page 288, Table 14.2), therefore the amount of incoming Electrical Energy which produces this action can be calculated.

It can be seen quite easily, that the incoming magnetic wave energy is Vast and Continuous. As an accretion mass, the Earth is an Energy Sink, getting it's energy from elsewhere, being Cosmic, Galactic and Solar. Conversion of incoming magnetic waves into electrical energy provides an unending, inexpensive and environmentally friendly source available to all. Cosmic and Galactic Energy is available twenty four hours per day. Large amounts of this Energy accumulates in the Earth's radiation belts. This Giant Energy Storage, when properly understood, provides a major source of free unending electrical energy. Each of My Inventions plugs into this vast energy source.

A perverse, Intentional Ignorance on the part of the Establishment, prevents recognition of the importance of the Energy Systems shown here. Any new system which is favorable towards the masses, is considered as disruptive, and therefore not allowed. Those who have the (Gold) Energy Rule (Golden rule) Mandated Destruction of all Humanity is not a consideration.

This Presenter will remove some of the Fog placed with the intention of preventing the recognition of this unending, environmentally clean, electrical energy Source, which is present everywhere throughout the Universe. The Cost of Harvesting and Using this Free Energy is a function of Human Stupidity.

RESONANCE CIRCUITS DEMO

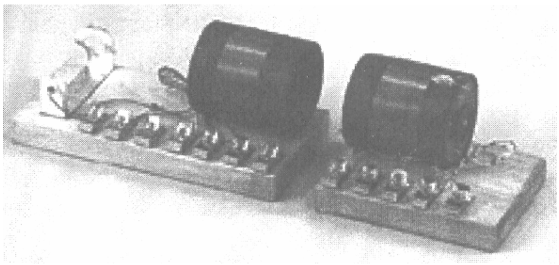

Used to demonstrate electromagnetic radiation between two UC circuits - one a transmitter and the other a receiver. When the 1.5 volt power transmitter is pulsed, the radiated signal is picked up by the remote receiver circuit which then lights up a 70 volt neon lamp.

With this apparatus, the student quickly understands some basic principles governing wireless communication, broadcasting, etc.

Kit: #10-416 $49.95

THE SCIENCE SOURCE
WALDOBORO, MAINE 04572
P.O. BOX 727

Tel. 1-800-299-5469
info@thesciencesource.com

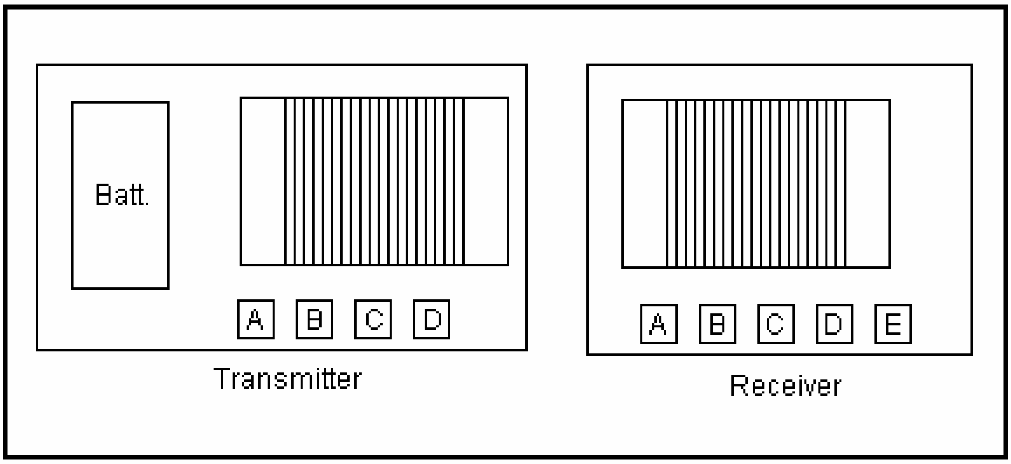

Diagram of transmitter and receiver coils

ULTIMATE ENERGY SOURCES

A human is a speck of dust on Earth, the Solar System is a speck of dust in the Galaxy and in turn, the Galaxy is a speck of dust in the Universe (Cosmos). All of these respectively represent vast ambient energy reservoirs. Awareness of the Sun, opens doors into other energy sources. Electromagnetic Energy which is present everywhere throughout the Universe, is accessed by catalytic activity, directly as in Solar Cells or indirectly as by mechanical means. Resonate, Magnetic Waves (Faraday's "Action at a Distance") allow Energy Activation Transfer to remote points of usage. The method of capture and use of this Energy is optional, and therefore it's cost is a function of Human Stupidity (Free-Energy).

Direct access is more desirable, and technology transfer from Solar Cell-type Devices provides the Catalyst. Enormously high Ambient Energy Levels are not detected by instruments that use the Ambient Background as a Reference Plane. A spoonful of water lifted from the Ocean does not define the Ocean. Incoming magnetic waves are reflected, Deflected or absorbed. Deflected Magnetic Waves spin electrons sideways producing useful Electrical Energy. Absorbed Wave Energy produces heat, therefore a hot interior for the Earth. In Electrical Systems physical movement is in the direction of current flow, frictional drag from inflow current defines gravity. Accretion masses resulting from Energy Sinks, provide all solid entities with their respective gravitational effect.

Increasing the tolerance level for Intellectual Awakening opens Doors of Reality. These doors blink into, and out of existence, and upon recognition, benefit Mankind. Opening some of these Doors, which at the present time are seen through a deep fog, is our purpose. Exploring Unrecognized Energy Sources, which are a Part of the Ambient Background, is another goal. Our Available Instruments do not use reference planes which allow recognition of this energy, as we shall see, vast Energy Sources that totally surround us are available through Technology Transfer. They are inexpensive (Free), fully self-renewable and environmentally benign.

Incoming Magnetic Wave Energy with Faraday's "action at a distance" will be looked at closely. Particle Physics will be left for the Astrophysics. Excited Electrons at point "A" the Sun (including the Galaxy and Cosmos) do not travel to point "B" the Earth, however a corresponding action occurs at point "B". The Electrons being disturbed at the Central Power Plant, in the same manner excite the Electrons at Your House, upon switching into an Earth grounding (known as "flipping the switch"). Correspondingly, there are Four Major Power Sources providing enormous amounts of Ambient Background Magnetic Wave Energy. They are The Cosmic, Galactic, Solar and Earth's Ambient Electromagnetic Backgrounds. The Earth's Electromagnetic Field comes from reflection, deflection and absorption as a result of action at a distance from the above.

Prescription Physics mandates that the Earth's background is of little interest. When we have Considered the evidence herein, it will become obvious that Special Interest's effort at keeping the People ignorant has, until now, largely succeeded.

Information for the entire World is available regarding the Magnetic Flux Background of the Earth's Surface (United State's Geological Survey, Colorado, USA, Office). When examined and properly understood, these Maps yield important information regarding reflection, deflection and absorption of incoming Magnetic Waves, plus action at a distance. When properly understood, these Maps reveal a very large Ambient Electromagnetic Energy Source. This is the Part of the Earth's Energy System that relates to the Bird on the High Voltage Line. When deflected, magnetic flux from electrons changes to electrical flux, providing the Motor System that spins or rotates the Earth. Physical movement by electrical systems is from inflow current movement. What level of current movement is required to spin the Earth? The Earth's Mass is 5.98 x 10^{24}. From this Information, the Watts of Electricity Required may be calculated! Absorbed microwave flux energy heats from the inside out, therefore a hot interior of the Earth results. Water is strongly diamagnetic, and on windless days, ocean waves provide visible Proof of the overhead incoming magnetic flux. From the information above, the Earth's weight and rate of spin allows the calculation of the amount of incoming ambient background energy required. As You can see, it is not inconsequential as Prescription Physics mandates.

Astrophysicist are concerned with charged particles that whiz by, once every one hundred years, rather than Wave Phenomenon associated with action at a distance. This highly Active Wave Energy translates into Electrical Energy at point "B". The Galaxy is alive With Energy which is billions of times greater than that of the Sun. Visible Light is a very tiny part of the Electromagnetic Energy Spectrum. Frequencies present in the Galaxy and Cosmos allow Radio Telescope photographs of their existence and magnitude. One such 408 MHz photograph of the Electromagnetic Energy Spectrum shows that the Earth is a tiny speck of dust in this Enormous Ocean of Energy, and can be seen near the left end of the Central High Energy Area.

This Energy extends in all directions. Accretion and formation of Planets, Suns and Galaxies are results of energy sinks and variable sized black holes. Mass retains heat, and is cooked from the inside out by the microwave background energy provided by the Universe. Flux movement into energy sinks, provides the frictional force know as gravity. Spinning mass in the presence of incoming flux amplifies the gravitational effect.

At present, only Solar Energy is recognized. It is inconsistent, flaky and a very small Part of the Magnetic Wave Energy Present. Technology Transfer from Solar Power provides uncomplicated and inexpensive, direct access to the Other Greater Energy Sources. All Electromagnetic Energy harvesting methods include a Catalyst, a Collector and a Pump. Catalysts include sensitization through doping with certain elements, air and earth groundings. Collectors include temporary storage as in Capacitors, Coils and Transformers. The Pump System includes induced movement onward to the point of use. Conventional rotating coils and magnet systems activate electrons present, such that action at a distance can occur, therefore it is an energy activation pump. In Direct Access Systems such as Solar Cells, the same occurs without mechanical action. Direct access occurs when Magnetic Waves impact a catalyst, spinning the local electrons sideways, producing useful electrical energy.

Indirect acquisition of electrical energy by mechanical means is wasteful, troublesome, expensive and degrades the environment. The dynamo is a combination collector and pump of energy which is collected from the Earth's Ambient Energy Background. Generators do not. make electricity, they collect it from the Ambient Background and forward it, as in Faraday's "action at a distance". Energy Conservation Laws relating to these systems, relate to gray areas, and when understood, are excluded because of the existence of External forces, open and non-linear systems as per Einstein. The Knowledge Base just viewed, provides a Direct Understanding of the Requirements for Harvesting of unending, fully renewable, environmentally benign Sources of Electrical Energy.

Magnetic Resonance Power System
Suggestions for Construction

This is the Basic Sonar Power System which permits submarines to see approximately 50 miles distance. What is not commonly known is that it works better at higher frequencies in the Gigahertz range. Any Device that can radiate 50 miles plus, is producing an enormous electromagnetic disturbance from a small input into a rod of magnetostrictive material. Disturbing the Earth's Ambient Background plus the strong dipole being produced, turns the magnetostrictive rod into a combination of a receiving antenna and a vastly superior output transformer.

The Drawing is only the Key Unit. A power input module and an output inverter circuit (diode bridge plus output transformer) are also required. The metal core and the wire size of the output transformer, plus adjusting the Earth Grounding of the Load, will determine the Amperage.

The Ideal rod material is Terfenol-D (check the internet). However a 1.5" diameter 10" long rod, costs over $5,000 each. Less expensive alternatives are obvious. When constructing, use PVC tubing with removable caps. Wind the coils on it and insert the experimental rod. Use only magnetostrictive material. When you get it right, you will have exactly what the Doctor ordered:

Magnetic Resonance Power System for Water Systems
Donald L. Smith

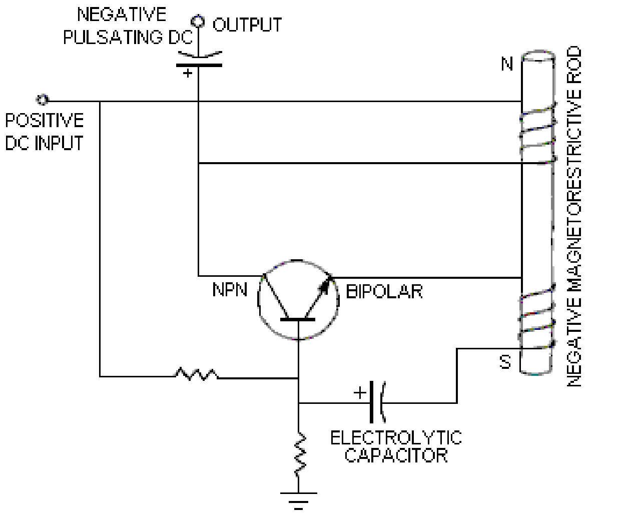

Magnetostriction oscillators work by magnetic resonance in a rod of magnetostriction material. This rod serves two purposes: It vibrates at the frequency of resonance oscillation, and it becomes the feedback transformer. Frequency is determined by items 4, 5, 6 and 8. The diameter, length and volume of the rod and output windings, determines the output. Item 2 provides feedback into the system. The negative magnetic character of item 8 plus the windings 2, in reaction to the magnetic flux field provided by 9, increases (amplifies or magnifies) the output. Magnetic permeability is the counterpart of negative resistance. Resonating with negative magnetic resistance, it pumps energy from the Earth's ambient background. Magnetic permeability is the ratio of flux density (Earth's B field) to the magnetizing force (H) in oersteds.

Magnetostrictive materials are piezoelectric in character, and have a very high resistance to electrical current flow. Examples are:

1. Permealloy Negative Magnetic Permeability	> 80,000
2. Sendust Negative Magnetic Permeability	30,000 -120,000
3. Metglas Negative Magnetic Permeability	> 200,000
4. Iron with (34%) Cobalt Magnetic Permeability	13,000
5. New Technology Magnetic Permeability	> 1,000,000

ELECTRICAL ENERGY SYSTEMS METHODS

1. DIRECT - Faraday's "Action at a Distance" incoming magnetic wave conversion to useful electrical energy. This includes Cosmic, Galactic, Solar and Magnets. Technology Transfer is from Solar Cell Technology.

2. INDIRECT - Electron Stimulation-Induced Electron Spin Systems, Electron Avalanche Pumping Systems

Primitive, Indirect Conversion from another form of energy. Coils and Magnet as in Dynamo Systems (Closed Systems). Chemical Systems, Atomic, Pons & Fleischman and etc.

Advanced, Direct Conversion, Magnetic Wave (Open Systems).

Ambient Sources
Air Core Coil Systems
Gaseous Tube Systems,
Solid State Marx Generator Avalanche Type Systems.
Leyden Bottle Capacitor Types inserted in Lakes and other.
Magnet Systems
Electron Beam Antenna Systems

3. TRANSFER MECHANISMS

Solids - as in metal conductors

Gaseous as in radio wave transmission, a form of ionization.

Sensitizing of Systems by use of Trace Doping with Radioactive elements, includes metal surfaces.

Open Systems, non-linear with external forces. Albert Einstein in a direct quote from his biography states that these are excluded from the conservation of energy laws.

Closed Systems Maxwellian Type Systems. Mathematics are predictable requiring deductive reasoning. Ohm's Law is King, and Establishment Intellectuals being comfortable with this, brand all else as a violation of the Laws of Nature by obtaining something for nothing. This is Dishonesty grand mal.

AMBIENT ENERGY SOURCES

Radiation System	*Diffusion Method*	*Magnetic Wave Energy*
1. Cosmic	**Reflection, Deflection and Absorption**	**Ultraviolet**
2. Galactic	**Reflection, Deflection and Absorption**	**Infrared**
3. Solar	**Reflection, Deflection and Absorption**	**Visible Light**
4. Earth	**Reflection, Deflection, Absorption, Faraday's "Action at a Distance" also, a Composite of all of the above**	**Earth's Electrical**

A deep fog pervades the entire Scientific Community with regards to the Significance of the Above Energy Sources. Magnetic Waves convert directly into Electrical Waves (useful electricity). Two sides of the electromagnetic system are always present and never separate. Local electron spin provides (action at a distance) the flip side of the incoming magnetic wave energy.

Enormous amounts of incoming magnetic wave energy becomes a part of the Ambient Background, and as such, cannot be measured directly. Reconstruction from indirect information, allows us to establish the actual energy levels which are present. Instruments provided by the Scientific Community measure only point "A" to "B", and when both are ambient, no potential energy is shown. This is the "bird sitting on the million volt power line and sensing nothing" approach. The Earth's actual ambient background has as it's Energy level multi-billions of Volts, which are conveniently and obliviously ignored by the scientific community. When properly understood, this enormous, never-ending source of environmentally-friendly energy becomes available.

ELECTRICAL ENERGY GENERATING SYSTEM

Patent Pending 08 / 100,074

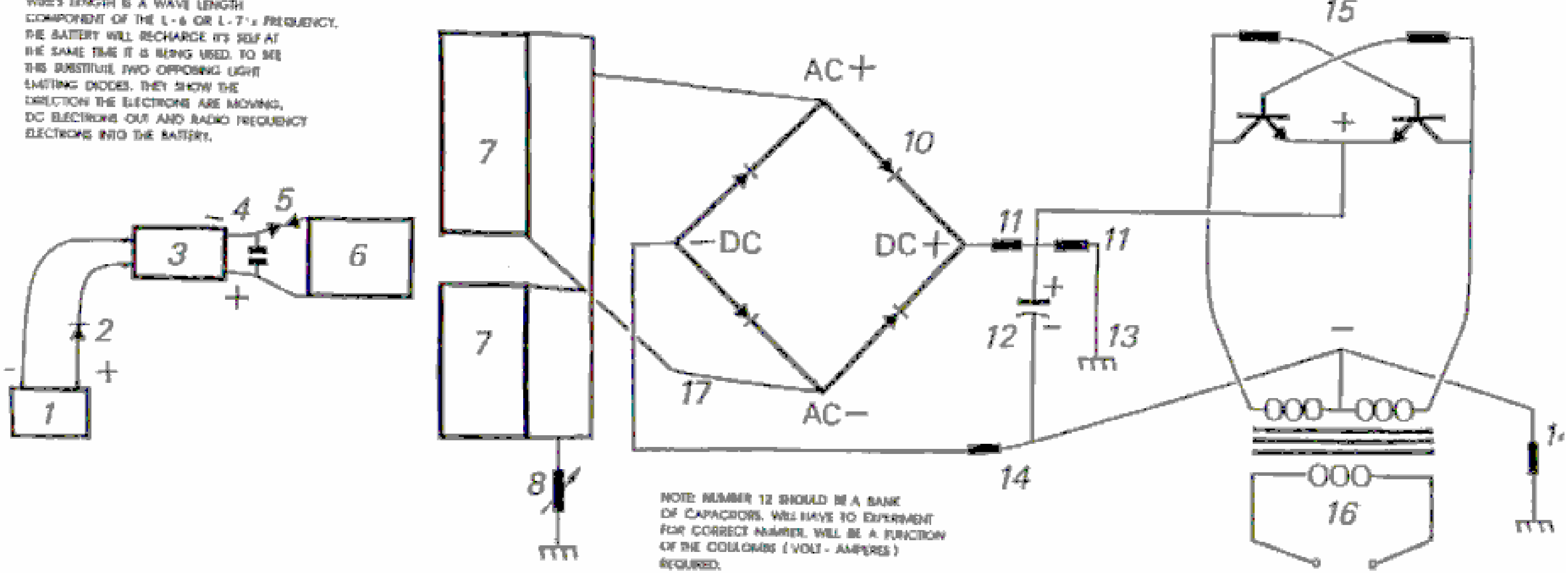

NOTE: IF THE PLUG CONNECTOR (2) WIRE'S LENGTH IS A WAVE LENGTH COMPONENT OF THE L-6 OR L-7's FREQUENCY, THE BATTERY WILL RECHARGE IT'S SELF AT THE SAME TIME IT IS BEING USED. TO SEE THIS SUBSTITUTE TWO OPPOSING LIGHT EMITTING DIODES. THEY SHOW THE DIRECTION THE ELECTRONS ARE MOVING, DC ELECTRONS OUT AND RADIO FREQUENCY ELECTRONS INTO THE BATTERY.

NOTE: NUMBER 12 SHOULD BE A BANK OF CAPACITORS. WILL HAVE TO EXPERIMENT FOR CORRECT NUMBER. WILL BE A FUNCTION OF THE COULOMBS (VOLT - AMPERES) REQUIRED.

1. Gelcel, 6 or 12 Volt,
2. Diode, Poss. use a Varactor,
3. High Voltage Module, Constituting the L-1 and L-2 Coils.
4. Capacitor, TDK 10.9 Pf., 30 KV,
5. Spark Gap, Small Engine Spark Plug, Gap = .0025 In.
6. Induction Transfer Coil L-3.
7. Induction Receiving Coil L-4.
8. Voltage Control Shunt.
9. Frequency Adjustor, prevents derating by Diode Bridge
10. Diode Bridge, 200 Nanosecond, R.F., > 100 KV .
11. Voltage Divider Circuit, corrects voltage for next stage.
12. Capacitor, electrolytic, smooths out DC + ripple effect.
13. Earth Ground.
14. Voltage Divider Curcit, corrects voltage for Transformer
15. Inverter Circuit, DC + in and 60 CPS to Transformer
16. Output from Transformer to Load (Work).
17. Center Tap

17 JANUARY, 1997

Electrical Energy System

Don L. Smith, Energy Consultant

At a meeting between J.P. Morgan, Edison and Tesla, Tesla proposed an Electrical Energy System which could he connected into directly, without using a meter. Tesla's Idea of "Free Energy" was not compatible with their thinking. Courtesy of Morgan and Edison, from that day foreword, a complete and total bastardization of the Idea has been in progress. Agents for Morgan and Friends include the U.S. Patent Office and Academia. Academia's bad habit of incestuous quoting of each other, eliminates them as a possibility in cleaning up the mess. This selective ignorance, permeates throughout the study of electricity.

Many people, otherwise known as "intellectuals", have a total blackout and become jabbering idiots when "free-energy" is mentioned. The term has been amended to say, "something which was never there is being harvested and that this violates the laws of physics". For the selectively ignorant, this seems the way to run. Those who choose Morgan's drum beat, have severely limited the possibilities built into electricity.

This paper will be an exercise in creative understanding, in placing updated knowledge at your disposal. Whether it becomes a useful tool or is selectively ignored is your choice.

Electrons are defined as being the practical source of electrical and magnetic energy. The electron as a particle, was postulated by professor J. Thompson in early 1900's. It is now universally accepted that the electron exists and that it is the source of electricity. When the electron is agitated it produces magnetic and negative electrical energy. Physics as it exists today, cannot explain why the electron remains intact and is not diminished by the energy it releases. This is a part of the built-in ignorance provided by the Morgan and Edison Camp.

One volts worth of electrons, when cycled, yields one volts worth of electricity. This can be repeated continuously forever and it never deplete or diminishes the electrons in question. They simply return to their air and/or earth source, waiting to do the whole thing again and again. Therefore, electrical energy is available, anywhere and everywhere humans go. People who intercede for profit, set the cost of electrical energy. Otherwise, all electrical energy is free, Morgan and Edison be damned.

Improving upon Professor Thompson's postulation, other obvious characteristics can be seen to further define the electron. It has both magnetic and electrical emanations resulting from a right-hand and left-hand spin. Since magnetism and amperage come as one package, this suggest, that electrons in their natural non-ionic state, exist as doublets. When pushed apart by agitation one spins and supplies electricity and the other spins and provides magnetic (amperage) energy. When they reunite, we have Volts x Amperage = Watts. This Idea, until now, has been totally absent from the knowledge base.

The number of times that an electron is cycled, sets the collective energy potential present. The electrical equivalent of $E = mC^2$ is $E = (\text{Volts x Amperes}) \text{ x } (\text{Cycles Per Second})^2$. Those who choose, are now free to head for the bushes and make their usual contribution to humanity.

Prior to Tesla, there was a large group of people in Europe, who were building resonant coil systems for medical use. Amperage was dangerous in their coil systems. The Tesla Coil is only the Voltage half of their coil system, as will be demonstrated.

A short list of those (from 1860 onwards) active in resonate high frequency coil systems include; the Curies, Roentgen, Ruhmkoff, Oudin, Hertz, Levassor, Dumont, D'Arsonval and many others.

Peugeot, Panhard-Levassor, Bollee, Renault and others had successful electric automobiles in production using A C. motors. Various electrically-powered airships, including the Dirigible "France" were in service.

D'Arsonval, Professor of Experimental Medicine at the College of France, invented the electrocardiograph, oscilloscope, amp and volt meters, thermography and numerous other medical applications of high frequency electricity. As early as 1860, he was building high frequency coil systems, which he used in his experimental work. There is a strong connection between the work of Tesla and the people mentioned above.

Electric vehicles of all sorts, dominated until the 1920s, when the electric starter motor made the internal combustion engine practical. Prior to that, upon cranking, it frequently would break the owner's arm. At that point the use of batteries as a source of power was replaced by oil.

The establishment's carpet has some rather large lumps under it. Coulomb's and Newton's inverse square law is politely ignored and it's opposite is allowed to have only the most abstract status. Without opposites we have no definition.

The source value of a remote flux reading, requires the squaring of the distance, times the remote reading, to obtain the original value. The opposite of this, being the derivations relate to Energy equals Mass times the Velocity constant squared. The electrical equivalent, being Energy equal capacitance times voltage squared and Energy equals induction times amperes squared. Flux lines increase as the law of squares and then activate electron energy which was not previously a part of the sum. The cumulative capacitance and inductance increase as the outer ends of a Tesla coil are approached, and this results in output energy being greater than the input energy present. This Energy is real. It can be safely measured by magnetic flux methods and electrostatic voltmeters, based on the inverse square law

As seen above, flux lines result both from induction-henrys-amperage and capacitance-coulombs-volts, and define electrical energy. The non-linearity of this system does not obey Ohm's law, which is replaced with impedance and reactance for alternating current systems. Impedance is the sum of the system's resistance to AC current flow, and this becomes zero at resonance. In resonant induction systems, a cycles-per-second increase, invokes a second round for the law of squares.

The degree to which flux lines are present, disturbs an equal amount of electrons, upsetting the ambient background energy, resulting in useful electrical energy being obtained. The frequency at which the disturbance occurs, increases the useful energy available, and it obeys the law of squares. Two square-law components, *flux density* and *frequency* are involved. Enter resonance which cancels the resistive effect.

Only the electrical energy which is either above or below the ambient level is useful. For the Central U.S. going east to west, ambient as approximated by electro-static voltmeters and flux methods is 200,000 volts on a solar-quiet day. At night time, the ambient energy level drops to about one half of the daytime value. On a solar-active day, it may reach more than five times that of a solar-quiet day. Ambient background energy at the polar regions, is approximately 500,000 volts on a solar-quiet day. The background energy varies as it relates to the North-South component and the East-West continuum.

This leaves us with an interesting problem. Electrons, when disturbed, first produce magnetic flux and then produce electrical flux when they spin back to their normal position. Therefore any electron movement produces above ambient energy, being over unity.

ELECTRICAL ENERGY WITH ASSOCIATED PHENOMENA

1. Current-amperes results from the unequal distribution of negativity (electrons).

2. Electron spin causes electrical current and magnetic lines of force.

3. Magnetic imbalance causes the gravitational effect. This is evidenced in electric motors by magneto-gravitational displacement of mess, which causes the motor to rotate.

ENERGY LINES OF FLUX (FORCE)
FIELDS & WAVES *

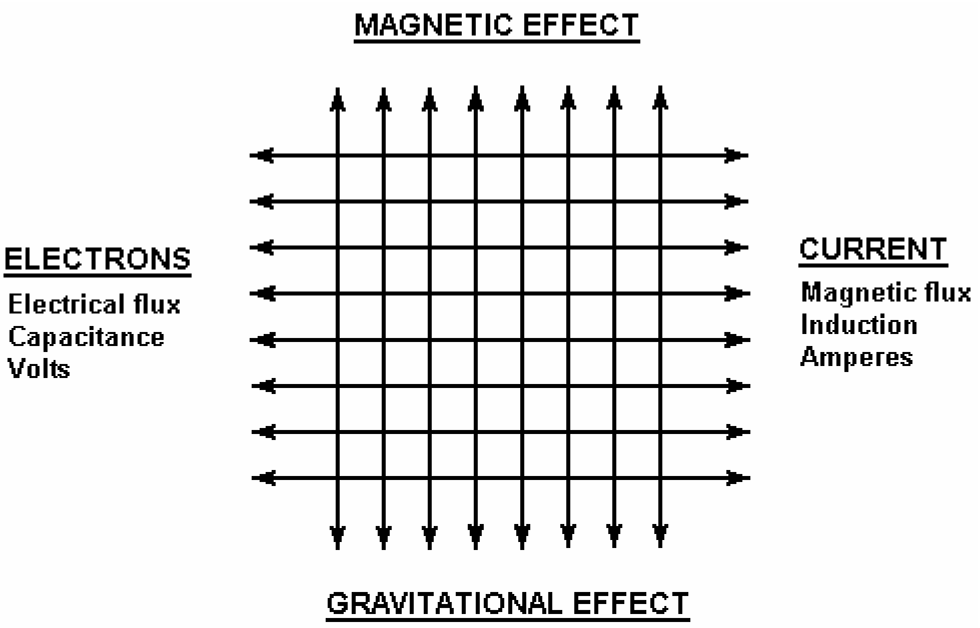
MAGNETIC EFFECT
ELECTRONS
Electrical flux
Capacitance
Volts
CURRENT
Magnetic flux
Induction
Amperes
GRAVITATIONAL EFFECT

* Below 20,000 Cycles Per Second = **Fields**
Above 20,000 Cycles Per Second = **Waves** (Radio Frequency)

Derivation of Magnetic and Electrical Power

Analogous Relationships:

1. Potential Power is present in a bar magnet as shown:

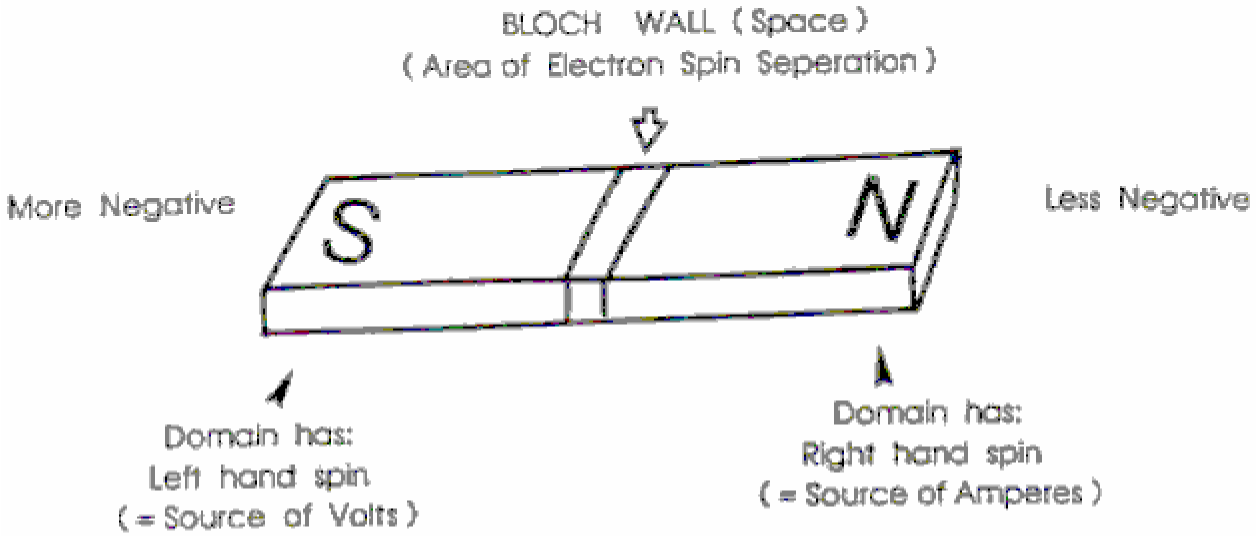
BLOCH WALL (Space)
(Area of Electron Spin Seperation)
More Negative
S
N
Less Negative
Domain has:
Left hand spin
(= Source of Volts)
Domain has:
Right hand spin
(= Source of Amperes)

2. The Source of these Electrons being from the Solar Plasma, are non-ionic and occupy all Free Space. They are commonly obtained from Earth and Air Groundings. They exist in Doublet Pairs, one being more negative than the other. The more negative one has a Left Hand Spin. The less negative one has a Right Hand Spin.

3. Resonate Electrical Coil Systems (Tesla) are Analogous to the System observed in the Bar Magnet (above). The Bloch Wall Area is Located at the base of the L-2 Coil. The Left Spin portion of the coil (Voltage Only) Coil predominates. The Right hand Spin portion of the coil (Magnetic-Amperage) is mostly absent.

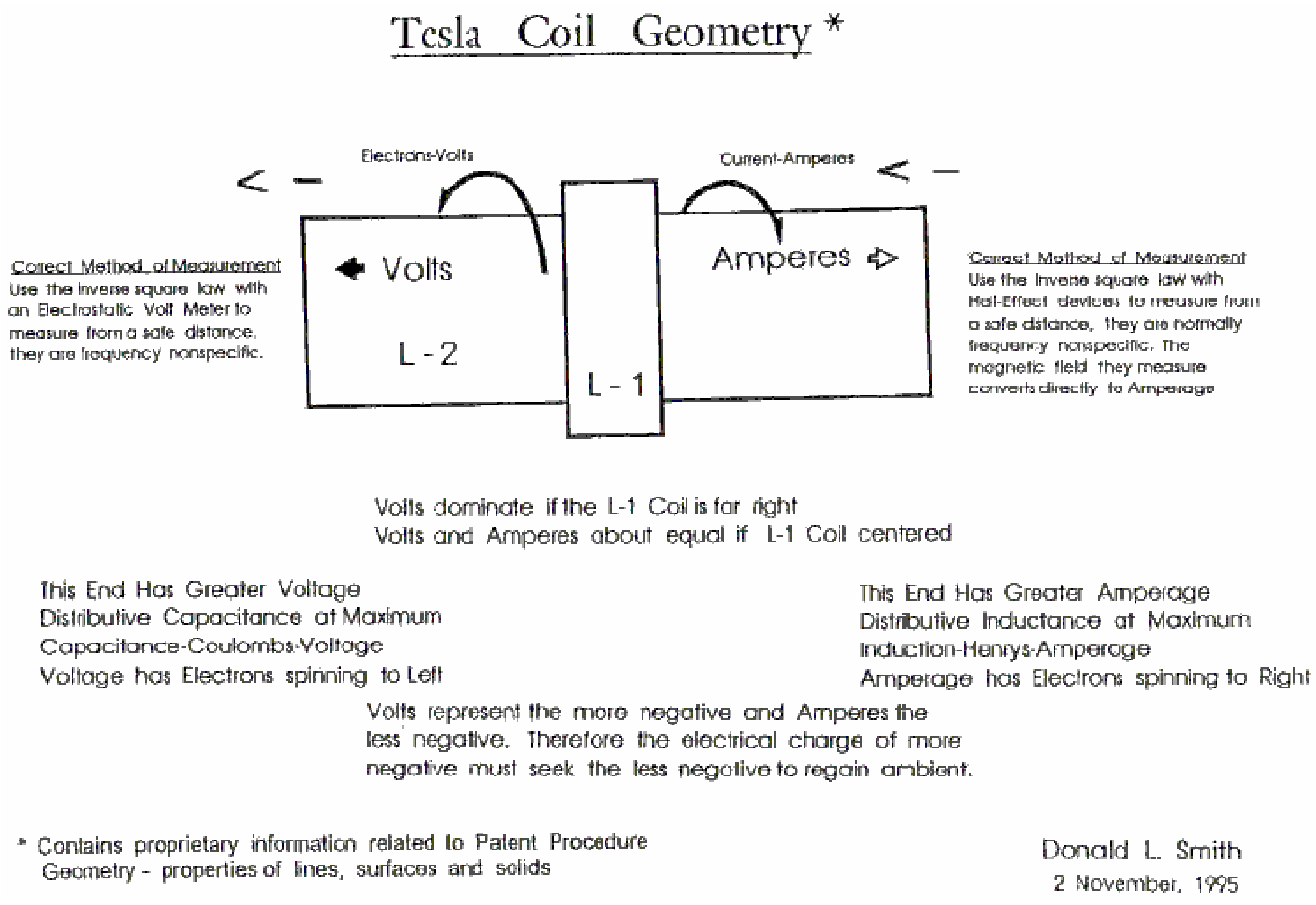

Tesla Coil Geometry *

Correct Method of Measurement
Use the inverse square law with
an Electrostatic Volt Meter to
measure from a safe distance.
they are frequency nonspecific.

Correct Method of Measurement
Use the inverse square law with
Hall-Effect devices to measure from
a safe distance, they are normally
frequency nonspecific. The
magnetic field they measure
converts directly to Amperage

Volts dominate if the L-1 Coil is far right
Volts and Amperes about equal if L-1 Coil centered

This End Has Greater Voltage
Distributive Capacitance at Maximum
Capacitance-Coulombs-Voltage
Voltage has Electrons spinning to Left

This End Has Greater Amperage
Distributive Inductance at Maximum
Induction-Henrys-Amperage
Amperage has Electrons spinning to Right

Volts represent the more negative and Amperes the
less negative. Therefore the electrical charge of more
negative must seek the less negative to regain ambient.

* Contains proprietary information related to Patent Procedure
Geometry - properties of lines, surfaces and solids

Donald L. Smith
2 November, 1995

Induced Electrical Energy System

Collection and transfer of energy requires temporary storage, which occurs as capacitors and coils of a resonant circuit are cycled, on and off. The frequency at which the capacitors and coils are pumped, determines the amount of electrical energy that moves onward.

The amount of Energy transferred relates directly to the density of lines of flux present. The Kinetic Energy Formula is helpful in establishing the amount of energy present. This formula squares the velocity times mass. In the case of electrical energy, the intensity of voltage and amperes multiplied by the cycles per second, replace the velocity component.

Note that the "acceleration" of the Voltage "E" and Amperage "I", which increase as non-linear components, then obeys the Law of Squares.

Each unit of increase, causes a squaring of the flux lines present. The amount of energy transfer caused by this increase in flux lines is demonstrated below.

Increase in Flux Lines Present Symbolized

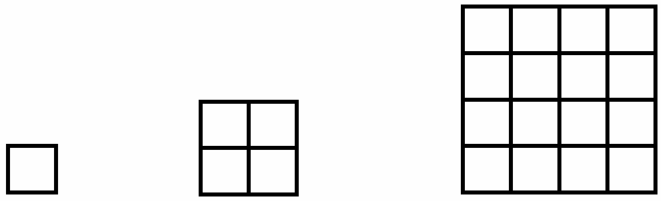

One Volt **Two Volts = Four** **Four Volts = Sixteen**

In resonant air-core coil energy transfer, the increase in flux lines present disturbs more electrons than previously, resulting in over-unity energy being present and available.

Energy stored, times the cycles per second, represents the energy being pumped by the system. Capacitors and inductors store electrons temporarily.

Capacitor formula: **W = 0.5 x C x E x Cycles per second** where:

W = energy in Joules (Watt Seconds)
C = capacitance in farads
E = applied potential in volts squared.

Inductor (Coil) formula: **W = 0.5 x L x I x Cycles per second** where:

W = energy in Joules (Watt Seconds)
L = inductance in henrys
I = current in amperes squared

Both one henry, and one farad, equal one volt. The higher the cycles per second, including the squaring of the flux lines, cause a large increase in the amount of energy being produced.

The above combined with a resonant energy induction system (where all electrons are moving in the same direction at the same time), make the next move into over-unity practical.

The dampening process of conventional electrical power generation, has all the available electrons bouncing randomly, mostly cancelling out each other. In that System, the useful energy available is a very small percentage of the energy which is present.

In the resonant induction system, a very high percentage of the energy present is useful. At resonance, (ohms-impedance-Z) becomes zero and all of the energy present is not degraded and becomes available to do useful work. "Ohms" is load or wasted energy, and "amperes" is the rate of that wasting of energy.

Using the previous information, if we now apply it to an air-core coil, resonant transformer energy system. L-1 and L-2 coils are now present. L-1 has a smaller number of turns and is several times the diameter of L-2. Input from a 12 volt high-voltage laser driver source, produces 8,000 volts with a low level of wasted energy, pushing amperage into, say, 4 turns of coil L-1. Each turn of the L-1 coil then acquires 2,000 volts of resonant potential. Consequently, each turn of L-2 is then exposed to the electric flux of 2,000 volts. Each turn at the bottom end of L-2 acquires 2,000 volts. The flux lines are squared and are additive as the voltage and amperage progresses towards the top end of L-2's large number of turns.

A huge number of additional flux lines which were not previously present become present at the top end of the L-2 coil. These flux lines excite the nearby electrons in it's earth and air and groundings. This high level of excitement above the ambient, causes a large number of electrons which were not previously a part of the energy present, to become available for use. At this point over-unity is present in large amounts.

The "bubble gum between the ears" response to this is: "this must be lots of volts but no amperes". Please recall that amperage is wasted energy, and that until that wasting occurs, there are no amperes. A good way to demonstrate this, would be to let the bubble gum crowd put their hands on the high-voltage end of the device while standing on wet ground (a people zapper). **Note: don't do this.**

This over-unity device produces energy at radio frequencies which range into the megahertz band. This allows the device to be small in size, and yet produce large amounts of energy. A megawatt-sized unit will sit comfortably on a breakfast table. This energy is changed to Direct Current and then switched to produce the desired working frequency AC.

Power Triangle

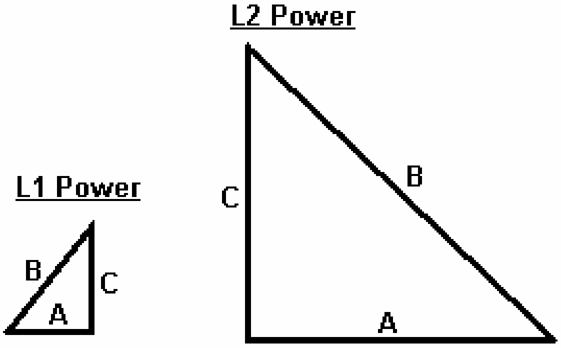

A: Volts x Amperes (the Available Power)

B: Volts x Ampres x Time (the Used Power)

C: Volts x Amperes x Recative (the Resinant Power)

1. Random movement of electrons in "A" and "B", mostly cancel each other out. This dampening, or wasteful concept of energy, is a source of much pleasure for the establishment.

2. "C" (Volt, Amperes, Reactive "V.A.R."), is the situation where all of the electrons move in the same direction at the same time. This results in near-unity energy output by resonant induction transfer.

3. Resonant induction transfer from one isolated power system, allows other resonant induction systems to duplicate the original source, which in no way diminishes the original source. Air-core coils (isolation-transformers) confirm this when they are a part of one of these functioning systems. A less perfect illustration would be the fact that the number of radio sets tuned to a particular radio transmission, does not alter the power required at the radio transmitter.

4. Resonant induction transfer, disturbs a large number of adjacent electrons which were not a part of the original input power source. The pulsating-pumping effect then draws in the newly available additional electrons into the on-going energy generation system. A near unity energy system of resonant air-core coils and the extra acquired electron-energy source constitute an over-unity system.

Electrical Power Generation / Points of Reference

Useful Electrical Power is Generated when Electrons from Earth and Air Groundings are disturbed by the movement of coils and magnets with reference to each other. The resulting electrical and magnetic energy is then changed to joules [watt-seconds: Volts x Amps x Seconds]. Each forward electron movement results in a magnetic impulse and each return movement causes an electrical impulse. The composite of the electrical energy impulses from these electrons yields useful energy [Power].

Let the above electron movement be represented by a room full of ping pong balls bouncing randomly. Most of the energy present cancels out by random impacts. This is the Classic Under-Unity approach to Electrical Power Generation, sanctioned by the Establishment.

In contrast to that, in the Electrical Energy Generation System presented here, the resonant Electrons are all moving in the same direction at the same time. This allows Near-Unity Electrical Power to Develop. This is the room-temperature equivalent of super conductivity.

The Energy System presented here, consists of a properly-adjusted and functional resonant air-core coil tank. The magnetic energy is stored in the coil system and the Electrical Energy is stored in capacitors. From Maxwell and others, we know that electrical-related energy has an equal amount of magnetic energy associated with it.

"The formula which establishes the Useful Energy of the System":

Joules = 0.5 x **C** x **V** squared x **C.P.S.** squared
units:
Joules (Volts x Amps x Seconds) Watt Seconds

C = Capacitance in microfarads
V = Potential in Volts
C.P.S. = Cycles per second

The transfer of Electrical Power by Resonant Induction is a direct function of the squaring of the cycles per second. For example, square 60 C.P.S. and then square the radio frequency C.P.S.s of the System here presented. Obviously, One Million Cycles per Second transfers more energy than Sixty Cycles per second. The Sanctioned Method of Electrical Power Generation uses the 60 C.P.S. Method. Using 60 C.P.S. and the random scattering of the Electrons System, assures the Establishment of it's desired Under-Unity Goal.

This random bouncing of the Electrons is the Ohms of Ohm's Law and is used to establish the rate of dissipation and/or Load [Work].

In the Resonant Tank Induction Energy Transfer System presented here, Impedance [system resistance] replaces the conventional ohm's usage. At Resonance, impedance becomes zero and the full force and effect of the Energy Transfer occurs. This is superconducting conditions at room temperature. At radio frequency the Electrons do not pass through the conductor as they do at lower frequencies. Instead, these Electrons encircle the conductor and are free of the conductor's resistance.

Let the Establishments Power Generation System be called 'A" and the System presented here be called "B".

With "A": Given 60 C.P.S. at 120 Volts using a 10 microfarad Capacitor:

Joules = [0.5 x 0.000010 x 120 squared] x C.P.S.s squared

(120 x 120 = 14,400)
[0.000010 x 14,400 = 0.144]
[0.144 x 0.5 = .072]
(0.072 x 3,600 = **259.2**]

Using the Inventor's Resonant Induction System, the Electrical Power available would then be 259.2 Joules [Watt-Seconds]. Using the Establishment's method only permits less than 10 Watt-Seconds of Useful Electrical Energy.

"B". Given One Million Cycles per second at 100,000 Volts, using a 10 microfarad Capacitor.

Joules = [0.5 x 0.000010 x 100,000 squared] x C.P.S.s squared

(100,000 x 100,000 = 10,000,000,000)
[0.000010 x 10,000,000,000 = 100,000]
(100,000 x 0.5 = 50,000]
(50,000 x One Million squared = **50,000,000,000,000,000**)

The useful Electrical Energy available is greater than 50 Mega Watts. Since the Resonant Electrons are non-impacting, all of the Energy is available for direct usage.

Benefits of the Inventor's System

1. Induction Energy transfer is enhanced by the squaring of the cycles per second produced by the System.

2. Induction Energy transfer is enhanced by the squaring the input voltage and amperage.

3. The increase of the flux lines occurring from the above, disturbing more electrons, causes more electrical energy to become available.

4. Resonant Induction has all of the Electrons moving unimpeded, resulting in superconductor conditions at room temperature.

5. A smaller amount of energy is used to disturb a larger number of Electrons. Electrons not originally a part of the System then contribute their energy, resulting in a net gain in available usable power.

6. The physical size of the System [Device] is small. The Device described in "B" sits comfortable on a breakfast table.

7. A small energy source is used to start the device and that source remains fully charged at all times by the System.

The Evidence Against Under Unity

1. Use of Logarithmic Scales on electrical measurement instruments. Linear measurement works fine where Ohm's Law applies (direct current). In alternating current, ohms are replaced by impedance and the measurements become non-linear.

2. Infinite "Q" at resonance confirms that voltage and amperage is squared, as in the kinetic energy formula. See the formulas of this report.

3. Square waves are clipped infinite "Q"s.

4. Maxwell and others show that magnetic-inductance-amperage and electrical-capacitance-voltage are two sides of the same coin. Magnetic-inductance is directly equal to amperage. Both obey the Law of Squares, which has over-unity built in.

5. Magnetic and electrical flux are present in enormous amounts at the outer ends of an operating Tesla Coil.

6. Ignorance of how to measure and relate magnetic and electrical flux, is the chief weapon of the under-unity gaggle.

7. The Cumulative inductance and capacitance of the Tesla Coil grounds itself out, if not properly utilized. See this report for the temporary energy storage accessible, when properly managed.

8. The Patent Office refers devices related to over-unity to their metering group, which is a sure indication that they are aware and accept the logarithmic measuring devices. This is direct and absolute evidence that they accept the square law as it relates to kinetic energy. This also indicates they are aware that over-unity exists. Since their bureaucratic brain is improperly motivated they continue to badger inventors who are working in the over-unity arena. Their level of intellectual dishonesty is sanctioned by, and is a real part of doing business with, a government which prides itself in being a hooliganistic bureaucracy.

Reading List

1. Alston, L.L. (Ed.), 1968, High-Voltage Technology, Oxford Univ. Press, London
2. Beck, E. 1954, Lightning Protection for Electric Systems, McGraw-Hill, N Y.
3. Bowdler, GAV. 1973, Measurements in High-voltage Test Circuits Pergamom, Oxford. L C. 72-86488
4. Brewley, L.V., 1951, Traveling Waves in Transmission Lines Pub: John Wiley. N.Y.
5. Chapman, S. and Barteis, J . 1940, Geomagnetism, two volumes, 1,000 plus pages, Oxford at the Clarendon press, England
6. Craggs. J.D. and Meek, J.M. 1954, High-Voltage laboratory Technique., Butterworth, London
7. EHV Transmission Line Reference Book, 1968 Edison Elcctric Institute, N.Y.
8. Farno, R.M., Chu. L.J., and Adler, R.B., 1968, Electromagnetic Fields, Energy and Forces, Pub: M.I.T Press, Cambridge, Mass.
9. Feinberg, R. (Ed.) 1979, Modern Power Transformer Practice, McMillan. London
10. Frungel, F., High Speed Pulse Technology, Vols 1 and 2, Pub: Academic Press. 1965, N.Y., and London
11. Gallagher, T.J. and Pearmain, AJ. 1983, ISBN 0-471-90096-6 High Voltage Measurement and Design, John Wiley, N.Y.
12. Hague, B., Alternating-Current Bridge Methods, 5 th. ed. Pub: Sir lsaac Pitman and Sons, 1959, London.
13. Hawley, W.G. 1959, Impulse-voltage Testing, Chapman and Hill, London.
14. Hayashi, Ch., Non-linear Oscillations in Physical Systems, Pub: McGraw-Hill, 1964. N.Y.
15. Henny, K. 1933-1959, Editor-in-Chief, Radio Engineering Handbook, five different copyrights, L.C. 58-11174, McGraw-Hill. N.Y.
16. Hudlestone, R.H. and Leonard, S.L., Plasma Diagnostics Techniques, Pub: Academic Press. N.Y.
17. Jacobs, J A., Editor, Geomagnetism, (a massive work) 3 large volumes, Pub: Academic Press, London.
18. Jeans, J.H. 1925, The Mathematics of Electricity and Magnetism, 5 th. Ed., Cambridge University Press,
19. Jones, B. 1972, New Approaches to the Design and Economics of EHV Transmission Plant, Pergamom, London.
20. Kind, D. 1978, An Introduction to High-voltage Experimental Technique, Viewieg, Braunschweig ISBN 3-528-08383-2
21. Knoepfel. H. 1970. Pulsed High Magnetic Fields, North-Holland, Amsterdam.
22. Kreuger, F H. 1964, Discharge Detection in High-Voltage Equipment, Temple Press, Heywood, London.
23. Kuffel, E. and Zacngl. W S. 1984, High-Voltage Engineering, Pergamom, ;London. ISBN 0-08-024213-8
24. Kupfmuller, K. 1957, Introduction to the Scientific Basis of Electrical Engineering, Pub Spinger. Berlin.
25. Lemon. H.B. and Ference, M. Jr., Analytical Experimental Physics " a major work" from The Ryerson Physical Laboratory at the University of Chicago, numerous copyrights 1933-1944. Printed as a text book by the University of Chicago Press
26. Lewis, I.A.D., and Well, F.H., Milli-microsecond Pulse Techniques Pub: Pergamom Press, N.Y. and London.
27. Malan, D.J. 1963, Physics of Lightning, English University Press, London.
28. Martin, T.L Jr., Physical Basis for Electrical Engineering, Prentice-Hall, N.J.
29. Matsusta, S. and Campbell, W.H . 1967 Physics of Geomagnetic Phenomena, Massive work in two volumes of 700 pages each. Academic Press, London.
30. National Physical laboratory, 1956, Notes on Applied Science ii 17, High Voltage Impulse Testing. HMSO.
31. Rather, H.. 1961, The Electron Avalanche and it's Generation, Pub: Vo. 33, Springer, Berlin.
32. Rokityansky, I.I., 1982. Geo-electromagnetic Investigation of the Earth's Crust and Mantle, Pub: Springer-Verlag, Berlin.
33. Salge, J., Pcier, D , Brilka, R., Schneider, D„ 1970, Applications of Inductive Energy Storage for the Production of Intense Magnetic Fields, Pub: Procedure of 6 th. Symp. on Fusion Technology, Aachen
34. Schulz, E.H., Anderson, L.T., and Leger, R.M., Experiments in Electronics and Communication Engineering, Copyrights, various, from 1943 thru 1954 LC 55-5033 Harper and Brothers, N.Y.
35. Schwab, AJ 1972, High-Voltage Measurement Techniques, The M I T. Press, Cambridge. Mass., ISBN 0262-19096-6
36. Sevin, L., 1965, Field Effect Transistors, McGraw-Hill, London.
37. Silsbee, F.B., 1942, Static Electricity, U.S. Department of Commerce, Washington. D C , U.S. Government Printing Office, (N.B.S. Circular C 438)
38. Smith, A. A., 1977, Coupling of External Electromagnetic Fields to Transmission Lines, ISBN 0-471-01995-X
39. Terman. F.E., Electronics and Radio Engineering, 4 th. ed., Copyrights 1932, 1937, 1947 and 1955 LC 55-6174 McGraw-Hill Book Co., N.Y
40. Thomas, R.T.,1970 High-Impulse Current and Voltage Measurement, Trans. I.E.E.E. EV1-19, pages 102-107
41. Thompson, Professor W, (titled Lord Kelvin) Reprint of Papers or. Electrostatics and Magnetism (original 1872) London.
42. Traister, R., 1983, The Experimenter's Guide to Solid State Diodes, Prentice-Hall, Inc. ISBN 0-13-295444
43. Valley, G.E, and Wallman. II . 1948, Vacuum Tube Amplifiers, McGraw-Hill, N.Y.
44. Weedy, B.M. 1980, Underground Transmission of Electrical Power, Wiley, London,
45. Weeks, W L. 1981, Transmission and Distribution of Electrical Energy, Harper and Row, N.Y.
46. Winch, R_P. Electricity and Magnetism (several printings) Composed in Copenhagen, Denmark, Prentice-Hall Physics Series
47. Zijlstra, H., 1957, Experimental Methods in Magnetism, two volumes, the North-Holland Publishing Co.. Amsterdam.

An Answer to America's Energy Deficit

Donald L. Smith

Energy Consultant

Energy, energy everywhere and not a Joule to Jounce. Conventional wisdom, when properly tuned will appreciate the nature of energy, as here presented. The basic unit of electricity (the electron) upon encountering a moving magnetic field (or wave) spins, giving off an electric impulse. When this impulse collapses, it spins back to it's natural position, giving off a magnetic impulse. Therefore, magnetic and electric are two sides of the same coin. When the magnetic side is pulsed, it yields electricity and conversely, pulsing of the electrical side yields a magnetic field. Moving one in relation to the other produces useful energy. When done consecutively, each cycle pushes (current) forward, while pulling electrons into the system... in much the same way as a water pump moves water. These electrons are obtained from Earth and air grounding.

The word "electric" comes from the Latin word electron "amber". When rubbed, amber develops an electrical charge, which can be transferred to a dissimilar substance. During the seventeenth and eighteenth centuries, a great deal of attention was centered on this attribute of amber. Amber was used to differentiate the non-metals. Carbon-related substances and other non-metals, when subjected to friction, give up negative electrical charges. On the other hand, metals when subjected to friction, simply conduct the charge. It is important to note that approximately 70% of the Earth's exposed crustal portions (surface) consist of silicone-related non-metals (electron donors) and become a direct source of electrical energy when properly agitated.

Useful electrical energy can be obtained by grounding into the Earth's non-metal crust and into it's atmosphere as a natural source of electrons. These electrons have accumulated from the solar plasma during the aging of the Earth for more than 4.5 billion years, at a rate exceeding 3.9 exajoules per year This indicates that the Earth's electrical field contains in excess of 17.6 x 10^{18} of cumulative exajoules of energy. One exajoule is the approximate energy equivalent of 125 million barrels of oil. The electrical energy in one display of lightning is approximately ten trillion joules. During each 24 hour period, the land portions of Earth's surface yields in excess of 200,000 emissions, which involves more than 2,000 quadrillion watts.

C.F. Gauss (1777-1855) and H.C. Oersted (1777-1851) were each separately trying to define the Earth's electrical field with all external influences removed. These external influences being solar-quiet periods and being remote from the land's surface. The air electricity background which they measured varies with latitude. Their European measurements correspond to approximately the latitude of Washington. D C. They were measuring magnetic field flux as an indicator of negative electron energy active and present. A related family of measurement are now presented. Units of measurement used to define flux fields include Gauss (one unit = 100,000 volts), Oersted (one unit = 50,000 volts), Tesla (one unit = 10,000 Gauss) and Gamma (one unit = 1/10,000 of a Gauss). Much confusion exists in electrical related publications about these units. As presented here they are correct with values taken from their original definitions.

The entire surface of the Earth has been surveyed by aerial magnetometer, in most cases using gamma or nano teslas. One gamma is the magnetic flux equivalent of 10 active volts of electricity. When the data is corrected for flight height it becomes obvious that there are numerous areas where the gamma readings exceed one trillion gammas. Lightning strikes from the ground up are in that energy range. With knowledge of these electron enriched areas, the quality of Earth grounding, becomes enhanced. The correction necessary for land surface data when acquired from aerial magnetometer maps (using Coulomb's law) requires that the remote distance be squared and then multiplied by the remote reading. As an example, if the remote reading is 1,600 gammas and the flight height being 1,000 feet. Take 1,000 x 1,000 = 1,000,000 x 1,600 gammas = 1.6 trillion gammas x 10 volts = 16 trillion volts equivalent for land surface data. Present day methodology requires mechanical energy in exchange for electrical energy. Once obtained, this energy is subject to Ohm's Law. Present Methodology obtains it's electrical energy from it's non-metal and air groundings.

This same energy can be obtained without the wasteful mechanical approach and at a much, much lower cost. Any required amount of electricity is available by resonant induction transfer from the

Earth's magnetic and electrical fields. The major difference is in the functioning of Ohm's Law in relation to resonant circuits. In the resonant induction system suggested here, system resistance (Z) becomes zero at resonance.

Therefore, Volts and Amperes are equal (V.A.R.) until work (load) is introduced.

Each cycling of this resonant induction system pulls in additional electrons from the Earth's electrical field, generating electrical energy in any required amount. In this system, a small amount of electrical energy is used to activate and pull a much larger amount of energy into the system.

This electrical advantage corresponds to the pulley and lever of the mechanical world. The electrical system presented here is extremely efficient. Using present methodology as a basis for comparison, with it's 60 cycles per second system. The resonant induction system, cycling at 60 million times per second produces one million times the energy which is produced by the present energy systems. A single small size unit of the resonant induction system has more usable electrical output than a major conventional unit. The radio frequency energy produced is easily changed to Direct Current, and then to the present 60 cycles per second system in preparation for commercial usage.

The Patent Pending on this system is #08/100,074, "Electrical Energy Generating System", dated
4 February, 1992.

Definitions: One **Joule** is one watt for one second
One **Watt** is one volt ampere
V.A.R. is Volt Amperes Reactive

Additional Reading:

Electricity and Magnetism by B.I. Bleany and B. Bleany Oxford University Press 1991 ISBN. 0-19-851172-8

Engineering Electromagnetics by W. H. Hayt. Jr. McGraw-Hill 1989 ISBN. 0-07-027406-1

Energy Methods in Electromagnetism by P. Hammond Oxford University Press 1986 ISBN. 0-19-859368-6

Energy in Electromagnetism by H. G. Booker Institution of Electrical Engineers by Peter Peregrinus Ltd. 1982 ISBN 0-900040-59-1

The American Radio Relay League Handbook for 1992 and 1993. 69th and 70th editions. Published by The American Radio Relay League. (For V.A.R. information) ISBN. 0-87259-169-7

Electron Paramagnetic Resonance, Techniques & Applications by R. S. Alger, U. S. Naval Radiological Defence Laboratory, San Francisco, California. Pub. Interscience Lib. Congress #67-20255

Geomagnetic Diagnosis of the Magnetosphere by A. Nishida, University of Tokyo 1978 Pub: Springer-Verlag ISBN. 0-387-08297-2

Energy and The Missing Resource by I. Dostrovsky. Pub: Cambridge University Press 1988 ISBN. 0-521-26592-4

High Voltage Measurement Techniques By A. Schwab, M.I.T.. 1971 I S.B.N. 0-262-19096

Environmental Magnetism by R. Thompson & F. Oldfield Pub: Allen & Unwin, London 1986 ISBN. 0-04-538003-1

Geo-electromagnetic Investigation of the Earth's Crust and Mantle. Translated from Russian, By I. I. Rokityansky. Institute of Geophysics, Kiev, U.S.S R. Pub: Springer-Verlag 1082. ISBN. 3-540-10630-8

Electron Paramagnetic Resonance of Transition Ions by A. Abragam and B. Bleaney Dover Publications, New York, N.Y. 1986

The Electromagnetic Field by A. Shadowitz, Dover Publications, New York, N.Y.

Geomagnetism, Several Volumes, Pub. Periodically by J.A. Jacobs, Institute of Earth Studies, Dyfed, U.K. Pub: Academic Press 1989-1990's.

Geomagnetism by S. Chapman and J. Bartels, 3 Volumes Oxford University Press, 1940

Physics of Geomagnetic Phenomena, Several Volumes by S. Matsushita and W. H. Campbell National Center for Atmospheric Research, Boulder, Colorado Library of Congress #67-23168 Pub: Academic Press, 1967

Physics Problems and How to Solve Them by C. E. Bennett, Professor Emeritus of Physics. University of Main (Particularly the sections on Electricity and Magnetism, and Units of Measure). Pub: Harper & Row ISBN. 0-06-460203-6

Units and Standards for Electromagnetics By P. Vigoureux, National Physical Laboratory Pub: Springer-Verlag 1071 ISBN. 0-387-91077-8

Surveyor's Guide to Electromagnetic Distance Measurement. Edited by J. J. Saastamoinen, Canada Pub: University of Toronto Press

Electromagnetic Distance Measurement by C. D. Burnside Pub: Granada, London 1971 ISBN. 0-258-96793-5

Der Magnetische Kreis "The Magnetic Circuit" By Von Heinz Rieger of Siemens AG. 1970 Berlin and Munchen, Germany I.S.B.N. 3-8009-4719 6

Electronic Modelling of Power Electronic Converters by J. A Ferreira Pub: Kluwer Academic 1989 33 AH Dordrecht, The Netherlands I.S.B.N. 0-7923-9034-2

Alternating Impact Energy Transfer Analogy

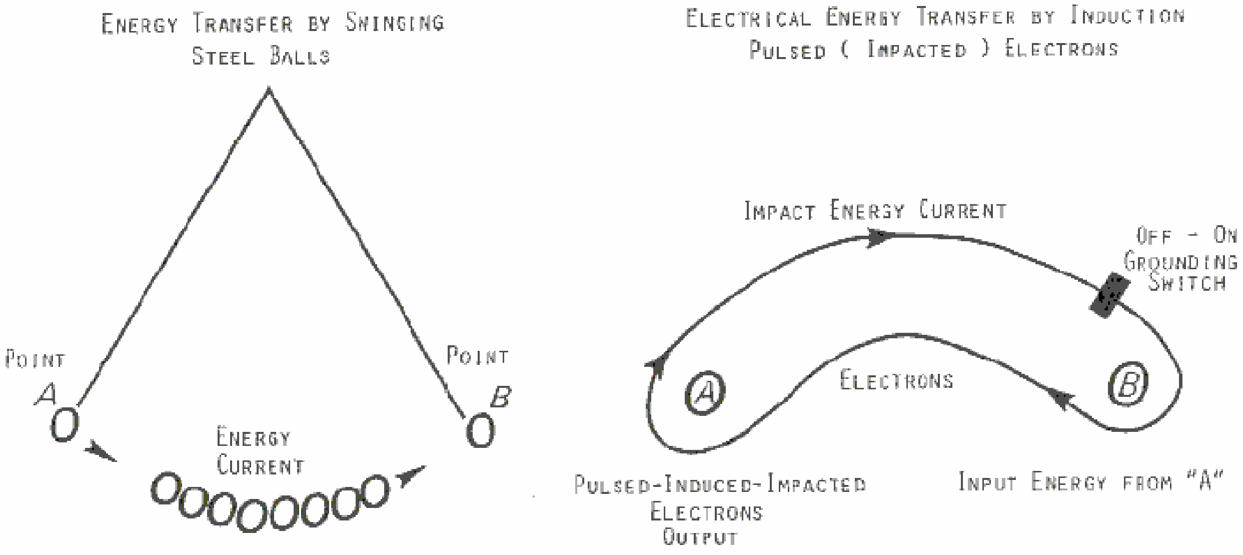
Energy Transfer by Swinging
Steel Balls
Point A
Point B
Energy Current
Electrical Energy Transfer by Induction
Pulsed (Impacted) Electrons
Impact Energy Current
Off - On Grounding Switch
A
B
Electrons
Pulsed-Induced-Impacted Electrons Output
Input Energy from "A"

In Alternating Electrical Energy Systems, Only the Potential-Impact Level Replicates It's Self at Point "B". Electrons do not Travel From Point "A" to Point "B".

E.E.S. II, BACKGROUND INFORMATION & CONCEPT

With alternating electrical current, electrons do not move from point "A" to point "B" as commonly envisioned! Electrical potential (oscillating electrons) at point "A", results in harmonic electron activity at point "B", when the grounding switch (circuit) is closed. That is to say, point "B" supplies it's own electrons and mirrors the activity of point "A". Impulsing (turbulence) by magnetic induction causes electrons to be pulled into the system, which then oscillates. When the magnetic field collapses (becomes absent) the electrical potential returns to it's natural background level.

Several major flaws are present in the conventional 60 cycles per second method of electrical power generation and it's iron core transformer system. ***This system is handcuffed by the inverse relationship of volts to amperes.*** This represents a stodgy, inflexible inheritance, courtesy of Mr. T.A. Edison and his concept of electrical power generation.

Nikola Tesla stood, almost alone, against Edison and managed to prevail with his Alternating Current system. Without the alternating current system, electronic things in the modern sense would not exist.

This report will be concerned with some of the extensions and benefits of the alternating current electrical system. ***This study will limit it's scope to air core coil transformers at radio frequency and upwards. The electrical power produced by this method is inverted to Direct Current and then to Alternating Current as required for popular usage.*** There are several important advantages of this system over conventional power generation.

Start with two coils (separate-apart), one being a reactor coil (L-1) and a second coil (L-2), being the reactant coil. Magnetic field fluxing (off-on of the electrical source) causes inductive reactance of L-1 which replicates by induction in L-2. Pulsing of the magnetic field (from L-1) in the presence of L-2, generates electrical potential. For example, should the L-1 coil have ten turns, with an imposed AC. potential of 1,200 volts. This results in each turn of L-1 acquiring 120 volts of potential. This induced magnetic field, then replicates itself in each turn of the L-2 coil. The L-2 coil may have one or many hundreds of turns. Modern encapsulation techniques makes high frequency and high energy controllable.

Let's take another important step in this air-core transformer process. For purpose of discussion, let the value of inductive reactance at 60 cycles per second, equal one. Each time the frequency is doubled, the effectiveness of induction is squared. At about 20,000 Hz, ***when radio frequency is achieved, the electrons begin spinning free, outside of the inductor and they become increasingly free of the inverse relationship of volt-amperes.*** From this point on, they replicate by the inductive process as V.A.R.. *That is to say,* ***volts and amperes are equal, until resistance (work) is introduced. Therefore, additional, not previously available electrons become incorporated for a very large net gain in potential. This gain is real !***

The quality of the grounding system determines the effectiveness of this method of producing electricity. A handy reference to locate the negative grounding areas for power generation can be found in the Aeromagnetic Map Studies of the US Geological Survey. They provide an excellent method for locating the best sites for optimum negative grounding areas.

When this method is combined with the induction coil system, already described, it provides an electrical power generating system millions of times more efficient than any known conventional method.

This new system ("E.E.S. II") is uncomplicated, physically small and it is inexpensive to build. The technology required for it's construction already exists. Maintenance is near zero, as there are no moving parts. Once operating, this system could last forever.

Small mobile E.E.S. II units are already available as replacements for the batteries used in electric automobiles. Larger E.E.S. II units can be provided as a replacement source of power for hotels, office buildings, subdivisions, electric trains, manufacturing, heavy equipment, ships, and generally speaking, any present day application of electrical power.

Earth Electrical System II, Modular Units

The system consists of three separate modules. Reverse engineering is used in matching the modules to the desired usage.

HIGH VOLTAGE INDUCTION TRANSFORMER MODULE:

1. Preferably an off-the-shelf-unit similar to a TV flyback and/or automobile ignition type related coil (transformer).

2. Ratio of input to output may be from less than 1:100 to greater than 1:1,000 A voltage tripler may then be used.

3. A connection allowing the high voltage output to pass onward through the induction coil L-1 and then to it's grounding.

AN AIR CORE INDUCTION COIL TRANSFORMER MODULE:

1. There are two coils: the reactor coil L-1 and the reactant coil L-2. L-1 has a high voltage radio frequency capacitor between it and it's grounding.

2. Input into the L-1 inductor is divided by the number of turns in it. The magnetic flux field provided from each turn of L-1 replicates itself as an electrical potential in each turn of L-2.

3. L-2 may have one turn or many hundreds of turns. The net gain depends upon the number of turns in L-2. Output from L-2 is in V.A.R. ***With this type of output, volts and amperes are the same until work(resistivity) is introduced.***

THE INVERTER MODULE:

1. Inverts to direct current (D C.)

2. Inverts to alternating current (A C), as desired.

3. Provides customized output of electrical power ready for designated usage

Earth Electrical System II, Domestic Use Range Module
Up to Two Million Volt-Amperes-Reactive Output *
Plan "A", With Variable Controls
Not to Scale

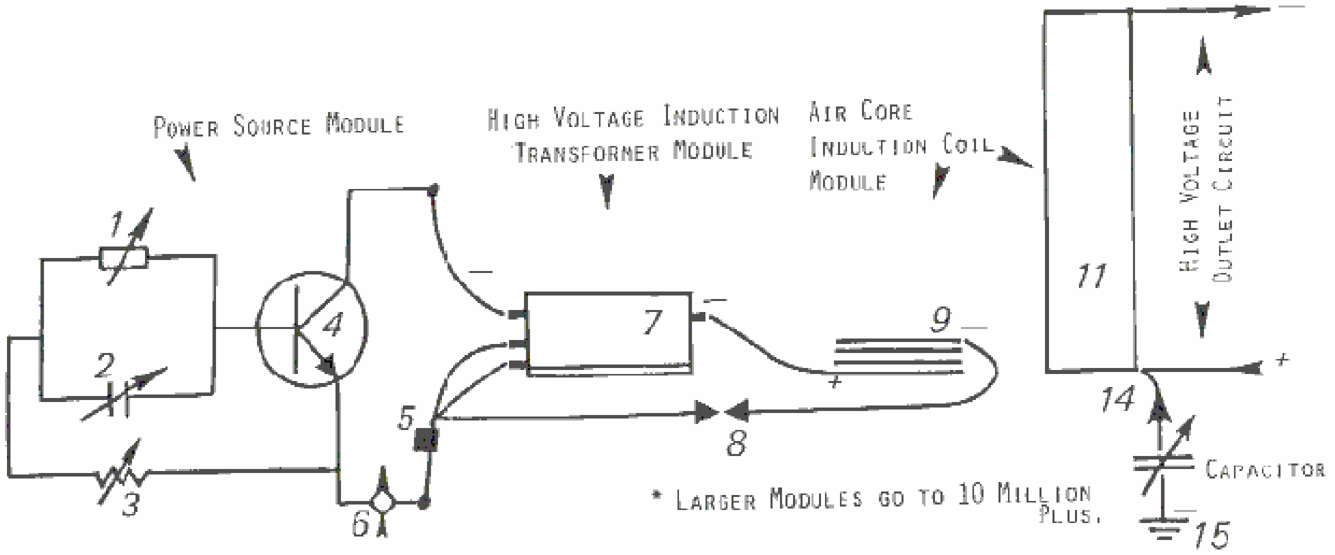
Power Source Module
High Voltage Induction Transformer Module
Air Core Induction Coil Module
12
High Voltage Outlet Circuit
11
1
2
3
4
5
6
7
8
9
14
15
Capacitor
—
+

* Larger Modules go to 10 Million Plus.

Parts: 1. Coil, Variable, 2. Capacitor, Variable, 3. Resistor, Variable, 4. Transistor, R.F., 5. Battery, Rechargeable, 6. Off-On Switch, Variable, 7. High Voltage Transformer, 8. Feed Back with Spark Gap, 9. Reactor, Induction Coil, 10. Feed Back with Spark Gap, 11. Reactant Coil, 12. Output for # 11, 13. Input for eleven, 14. Grounding for eleven.

Earth Electrical System II, Domestic Use Range Module
Up to Two Million Volt-Amperes-Reactive Output *
Plan "B", Electrical Automobile Energy Source

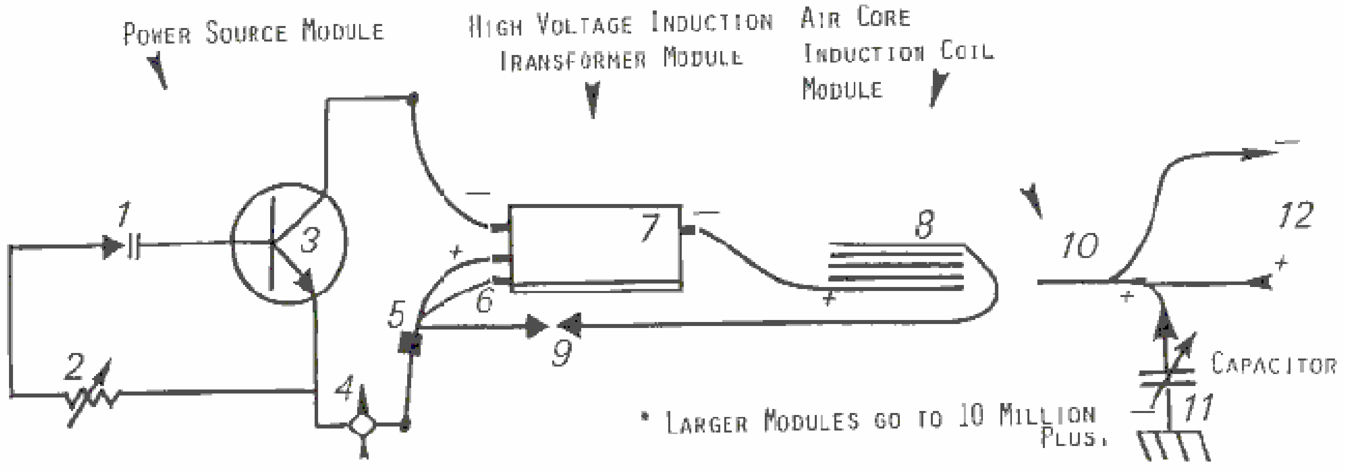
Power Source Module
High Voltage Induction Transformer Module
Air Core Induction Coil Module
1
2
3
4
5
6
7
8
9
10
11
12
Capacitor
—
+

* Larger Modules go to 10 Million Plus.

Parts: 1. Varactor, Radio Frequency, 2. Resistor, 3. Transistor, Radio Frequency, 4. Off-On Switch, Multi-Position, 5. Battery, Rechargeable, 6. Transformer Grounding, 7. High Voltage Induction Transformer, 8. Reactor, Induction Coil, 9. Feed Back with Spark Gap, 10. Reactant, Induction Coil, 11. Grounding for # 10, Out-put Circuit, In Volt-Amperes-Reactive.

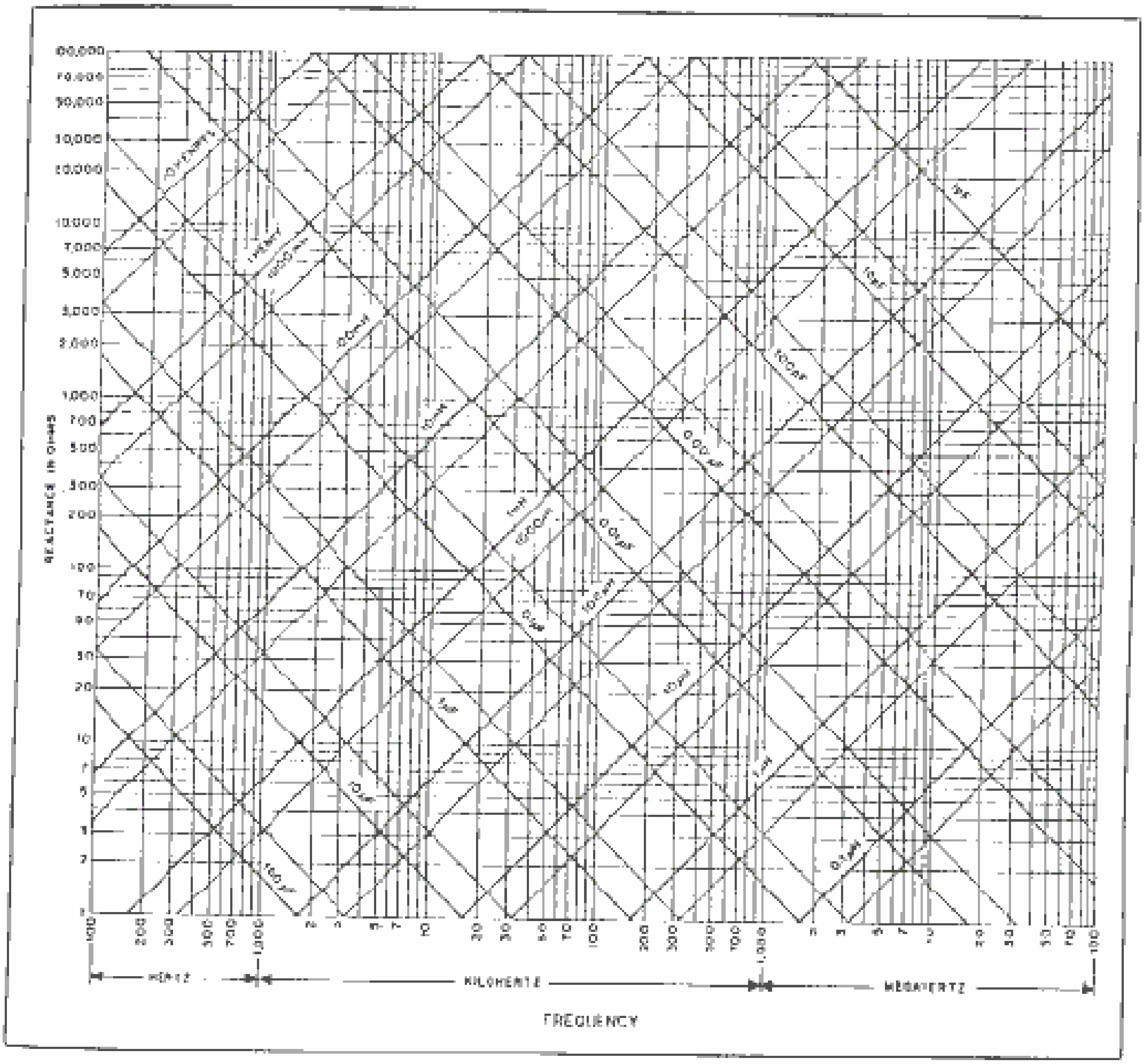

Fig. 44 — Inductive and capacitive reactance vs. frequency. Heavy lines represent multiples of 10, intermediate light lines multiples of five. For example, the light line between 10 µH and 100 µH represents 50 µH; the light line between 0.1 µF and 1 µF represents 0.5 µF, and so on. Intermediate values can be values within the chart range. For example, the reactance of 10 henrys at 60 Hz can be found by taking the reactance of 10 henrys at 600 Hz and dividing by 10 for the 10× times decrease in frequency.

CHART, COURTESY OF A.R.R.L., 1992 ED.

Speech presented the evening of 23 July. 1994 at the
International Tesla Society Convention at Colorado Springs. Colorado.

DONALD L. SMITH
ENERGY CONSULTANT
8110 BENT OAK LANE
SPRING. TEXAS 77379

ELECTRICAL ENERGY REFERENCE POINTS

Electrical Energy Generating System
Patent Pending # 08/100.074. 2/4/92

The word "electric" comes from the Latin word electron "amber". When rubbed, amber develops an electrical charge, which can be transferred to a dissimilar substance. During the seventeenth and eighteenth centuries, a great deal of attention was centered on this attribute of amber. Amber was used to differentiate the non-metals. Carbon-related substances and other non-metals, when subjected to friction, give up negative electrical charges. On the other hand, metals when subjected to friction, simply conduct the charge. It is important to note that approximately 70% of the Earth's exposed crustal portions (surface) consist of silicone related non-metals (electron donors) and therefore becomes a direct source of electrical energy when properly agitated.

Useful electrical energy is obtained by grounding into the Earth's non-metal crust and into it's atmosphere as a natural source of electrons. These electrons have accumulated from the solar plasma during the aging of the Earth for more than 4.5 billion years, at a rate exceeding 3.9 exajoules per year. This indicates that the Earth's electrical field contains in excess of 17.6 x 10^{18} power of cumulative exajoules of energy. One exajoule is the approximate energy equivalent of 125 million barrels of oil. The electrical energy in one display of lightning is approximately ten trillion joules. During each 24 hour period, the land portions of the Earth's surface yields in excess of 200,000 emissions, which involves more than 2,000 quadrillion watt-seconds of active energy on display.

This physical phenomenon indicates that the Earth's crust is an unending source of electrical energy. The surface area involved is a very small portion of the Earth's crust.

J.C. Maxwell (1891) suggested that an active electron field gives rise to an associated magnetic field. Therefore, both are present with pulsating current. Early studies, involving observation of compass needles by microscopy, revealed that the needle vibrates as with alternating current. More recent studies by A. Nishida and others, confirm that alternating current is common in the Earth's crust.

C.F. Gauss (1777-1855) and H.C. Oersted (1777-1851), both were separately trying to define the Earth's electrical field with all external influences removed. These external influences being solar-quiet periods and being remote from the land's surface. The air electricity background which they measured varies with latitude. Their European measurements correspond to approximately the latitude of Washington, D.C. They were measuring magnetic field flux as an indicator of negative electron energy active and present.

A related family of measurement is now presented. Units of measurement used to define flux fields include Gauss (one unit = 100,000 volts), Oersted (one unit - 50,000 volts), Tesla (one unit = 10,000 Gauss) and Gamma (one unit = 1/10,000 th of a Gauss). Much confusion exists in electrical related publications about these units. As presented here, they are correct with values taken from their original definitions.

The entire surface of the Earth has been surveyed by aerial magnetometer, in most cases using gamma or nano teslas. One gamma is the magnetic flux equivalent of 10 active volts of electricity. When this data is corrected for flight height, it becomes obvious that there are numerous areas where the gamma readings exceed one trillion gammas. Lightning strikes from the ground up are in that energy range. With knowledge of these electron enriched areas, the quality of Earth grounding, becomes enhanced.

The correction necessary for land surface data when acquired from aerial magnetometer maps (using the inverse square law) requires that the remote distance be squared and then multiplied by the remote reading. For example, if the reading is 1,600 gammas and the flight height is 1,000 feet. Take 1,000 x 1,000 = 1,000,000 x 1,600 gammas = 1.6 trillion gammas x 10 volts = 16 trillion volts equivalent for land surface data.

Present day methodology requires mechanical energy to be expended in exchange for electrical energy. Any required amount of electricity is available by resonant induction transfer from the Earth's magnetic and electrical fields. Each cycling of this resonant induction system pulls in additional electrons, generating energy in any required amount. A small amount of electrical energy is used to activate and pull into the system a much larger amount of energy.

ENERGY VERSUS MASS

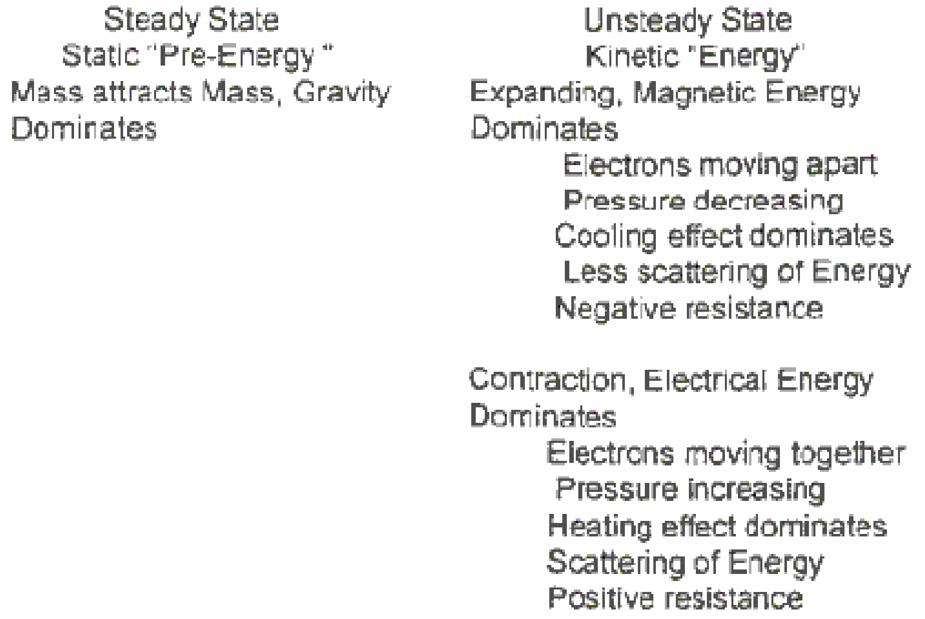

Steady State
Static "Pre-Energy "
Mass attracts Mass, Gravity
Dominates

Unsteady State
Kinetic "Energy"
Expanding, Magnetic Energy
Dominates
Electrons moving apart
Pressure decreasing
Cooling effect dominates
Less scattering of Energy
Negative resistance

Contraction, Electrical Energy
Dominates
Electrons moving together
Pressure increasing
Heating effect dominates
Scattering of Energy
Positive resistance

Functions of active Electrons

Electrons become active when placed inside the critical distance allowed by their negativity.

Active Electrons provide:

1. Electricity
2. Magnetics
3. Gravitational thrust as in Electric Motors
4. The source of Visible Light
5. It's charge is Negative

They move in a closed loop as seen in the Icon for infinity, not in a circle as shown in many books. One half of the loop consist of a magnetic impulse and the return half consist of the electrical impulse. This is seen as the classic sine wave of alternating electrical energy.

A flash of light occurs when two electrons suddenly find they are too close together. Daylight results from the impingement of Electrons in the Earth's atmosphere with the Electrons of the Solar Plasma.

My Concept of the Forces of Nature differs from the conventional. It consist of a weak and a strong force, each being additionally composed of electrical, magnetic and gravitational (fields and waves). Any two of the three constitute the third member; Gravity "B" of the weak force competes with humans on a daily basis. Gravity "A" of the strong force is the force that holds the Solar System and the Universe in place. Energy from the Electrons represent the weak force. Energy inside the Atom represents the strong force "A". Controlled resonant induction of any two of the three, changes into the third and is the motor that runs the Universe. We see this in the electrically-induced magnetic thrust against gravity in electric motors.

Weak force is required to dislodge electrons and strong force (atomic) to dislodge protons. Unless dislodged, these particles are of little value in producing Conventional Electrical Energy.

Therefore, in conventional electrical energy production, the particle of importance is the negative electron. Electrons have a "grudging" relationship with other electrons. They like each other, especially at arms length. Like potentials repel each other, and unlike potentials attract. To demonstrate this, take two batteries of the same type, but of a different charge level (unequal potentials). Put the plus and minus ends facing the same direction. Then with a volt meter, measure the electrical potential between the two negative ends and then the two positive ends. It is obvious that the "more negative" moves to the "less negative" is the correct concept for electrical energy generation. Electrical Energy flow consist of a higher concentration of electrons moving to an area of lesser concentration.

OHM'S LAW WITH CORRECTIONS:

A major obstruction in reference to the correct function of electrical energy is the establishment's incorrect interpretation of Ohm's Law. The corrected version is:

Volts = Energy Available (Potential)

Ohm = Scattering, dissipation of Energy (Load)

Ampere = the rate of, dissipation / scattering of energy

It is important to note that Ohm and Ampere are after the fact, and are not decisive except for the dissipation factor. High Voltage at low amperage simply means that the High Voltage is still intact for future usage. In no way is the potential diminished by low amperage.

EXAMPLES OF OVERUNITY

Dominos did not exist in England when the Laws of Conservation were originally put in place. Otherwise they might have been very different. For example, let us take a long row of upright dominos, (many thousands) and flip number one. The Energy required to flip the first domino must now be added with that of thousands more in order to have a correct assessment.

The Electron itself is an excellent example of over-unity. The electron provides various forms of energy continuously throughout eternity and is in no way diminished. It simply cycles through the system and is available thereafter.

In Electrical Systems, Electrons active at point "A" are not the same Electrons active at point "B". That is to say, the Electrons activated at the Central Electrical Energy Station are not the ones used at your house. When you ground your system by flipping the wall switch, you use your own electrons. In closed energy systems, electrons communicate with and replicate the activity of the overbalanced potential, when provided with Earth and or Air Groundings.

The number of Radio sets and Television sets running at any one time do not diminish, in any way the electrical output of the source station.

For example, let now use an Air Coil Resonant Induction System for the purpose of flipping some electrons. The flipping device (reactor coil L-1) is pulsed, which then provides a resonant induction pulse. In turn, this flips the electrons present at the (reactant L-2) Coil. The energy input in L-1 is divided by the number of turns present. The induced magnetic pulsing in turn flips the electrons in each turn of L-2. If more turns are present in L-2 than L-1, there is a net gain in the Energy present, as demonstrated by the dominos above. The farads and henrys of the resonant system provide the resonant frequency when pulsed by an external energy system. A system shunt in the resonant circuit sets the containment level for energy potential.

The Induction Process itself provides an excellent example of over-unity. When comparing rate of induction, the cycles per second must be squared and then compared to the square of the second System. Let us then compare the 60 c.p.s. System with my 220 MHz Device. Energy produced at radio frequency has several major advantages over the conventional system. Ohm's Law does not apply to a resonant air-core radio frequency system.

For example: When the system is resonant, the following is true:

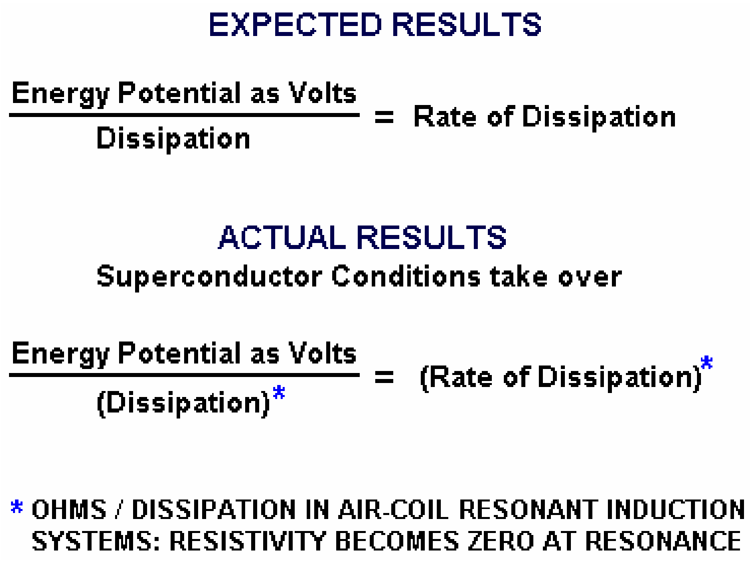

EXPECTED RESULTS

$$\frac{\text{Energy Potential as Volts}}{\text{Dissipation}} = \text{Rate of Dissipation}$$

ACTUAL RESULTS

Superconductor Conditions take over

$$\frac{\text{Energy Potential as Volts}}{(\text{Dissipation})^{*}} = (\text{Rate of Dissipation})^{*}$$

* OHMS / DISSIPATION IN AIR-COIL RESONANT INDUCTION SYSTEMS: RESISTIVITY BECOMES ZERO AT RESONANCE

This is named the V.A.R. (Volt Amperes Reactive) System.

When compared to the Conventional Under-Unity iron-core transformer system, the results are over-unity. It is strange that mechanical advantage as in pulleys, gears, levers and others which correspond to the electrical advantage above mentioned, are not considered over-unity devices.

Let us take a closer look at resonant induction. As an example, let a room full of ping pong balls randomly bouncing at a high speed represent the Conventional method of under-unity energy generation. Suppose that by resonant induction the balls all move in the same direction at the same time. When this occurs a huge amount of energy not previously available is present. The resonant air-core coil system lines up the electrons in such a manner that the energy factor is nearly 100 % , and not the 2% or 3% of Conventional under-unity devices sanction by the establishment.

Some other devices where overunity is common would be resonant induction circuits present in conventional radio tubes (high plate voltage), negative-feedback systems found in Op-Amps and possibly others.

SUMMARY

Useful electrical energy is achieved when the electron density at point "A" becomes greater than at point "B", (being the more-negative moving to the less-negative concept). Coils moving through a magnetic field or vice versa causes this imbalance.

The mindset of the professional Electrical Engineer is restricted to non-resonant and iron-core coil resonant systems. Ohm's Law, when applied to resonant air-core induction systems, becomes, system resistivity (impedance, Z). "Z" becomes zero at resonance. Therefore, in this system, volts and amperes are equal until load (resistivity) is introduced. This is called the Volt Ampere Reactive (V.A.R.) System. With impedance being zero, the System grounding is coupled directly into the Earth's immense electrical potential. Efficiency of induction relates to the square of the cycles per second. Compare the ratio of the conventional 60 c.p.s. System and the 220 million plus cycles of my Earth Electrical System II.

Electrons which cycle through this system, after being used, are returned intact to their former state for future usage.

Electron spin causes electrical current and magnetic lines of force

The effect of current, results from the unequal distribution of negativity (electrons).

Magnetic imbalance causes the gravitational effect. This is evidenced in electric motors by magnet-gravitational displacement of mass which causes the motor to rotate.

The System is an extension of present technology.

The System and it's source utilizes magnetometer studies.

This System (Earth Electrical System II. "EES.II") utilizes a fully renewable energy source.

This System utilizes a non-polluting energy source.

This System utilizes an universally available energy source.

Endorsement and Certification of The System can be anticipated by States with pollution problems.

AIR CORE INDUCTION COIL BUILDERS GUIDE

DONALD L. SMITH

Energy Consultant

1. Decide frequency. Considerations are: (economy of size)

- **a.** Use radio frequency upward (above 20,000 Hz).
- **b.** Use natural frequency (coils have both capacitance and inductance), that is match the wire length of the wire in the coil to the desired frequency.
- **c.** Wire length is either one quarter, one half or full wave length.
- **d.** To obtain the wire length (in feet) use the following: If using one quarter wave length divide 247 by the desired frequency (megahertz range is desirable). If using one half wave length divide 494 by the desired frequency. If using full wave length divide 998 by the desired frequency.

2. Decide number of turns, ratio of increase in number of turns sets the function. In the case of the L-1 coil, each turn divides the input voltage by the number of turns. In the case of L-2 coil, the resulting voltage in each turn of L-1 is induced into each turn of L-2, adding up with each turn. For example if the input into L-1 from a high voltage, low amperage module is 2,400 volts, and L-1, for example, has 10 turns, then each turn of L-1 will have 240 volts of magnetic induction which transfers 240 volts of electricity to each turn of L-2.   L-2 may be one turn or many turns, such as 100 to 500 or more turns. At 100 turns, 24,000 volts would be produced.   At 500 turns, 120,000 volts would be produced.

3. Decide the height and diameter of the coil system. The larger the diameter of the coil, the fewer turns are required, and the coil has a lesser height. In the case of L-2 this results in lowering the amplification of the induced voltage from L-1.

4. For example, if 24.7 MHz is the desired frequency output from L-2. One quarter wave length would be 247 divided by 24.7 which equals 10 feet of wire. The number of turns will be the amplification factor. The coil may be wound on standard size P.V.C. or purchased from a supplier. The supplier is normally a ham radio supply source. Once the length is determined and the number of turns decided, move to the next step.   For example, let each turn of L-1 have 24 volts and desired output of L-2 be 640 volts. Therefore L-2 needs 26.67 turns. It has been determined that the wire length for one quarter wave length is 10 feet. The number of inches in 10 feet is 120. Using Chart "A" supplied look for next higher number of turns showing (being between 20 and 30 turns with a 2" diameter coil). This tells us to use a 2" coil. If ready-made as in the case of Barker and Williamson, 10 Canal Street, Bristol, Penna., 215-788-5581, the coils come in standard sizes of 4, 6 and 10 turns per inch. For higher "Q" use wider spacing of the turns. These coils come in a ready-made length of 10 inches. Select from the coil 30 turns and put input clamps on the base of the coil and at 30 turns. For exact determination of the correct position of the output clamp, use an externally grounded voltage probe. The node of maximum intensity, being the natural resonant point. Off the shelf multimeters are not radio frequency responsive. The easiest way to accomplish the above is to get from the hardware store or Radio Shack a voltage detector having a neon bulb system (Radio Shack Cat. No. 272-1100b, NE2-Neon Lamps) will work. With your hand as a ground, move the wire extension of the neon lamp along the coil surface until the neon is brightest. This is the desired point of resonance and it is the optimum connection point.

5. The input power now needs consideration. A 2,400 High Voltage module has been previously selected. This module can be made from a diode bridge or any combination of voltage amplifiers. The one used here is an off-the-shelf type, similar to those used for laser technology.

6. Construction of the input L-1 coil. It has already been decided that there will be 10 turns. The length of the wire here is not critical. Since the L-2 coil is 2-inches in diameter, the next off-the-shelf larger size may be used for L-1. Use a 3" diameter off-the-shelf coil which has 10 turns to the inch. Remove (cut) a 10 turn portion from the larger coil. Use an L.C.R. meter and measure the natural farads (capacitance) and henrys (inductance) values of the L-2 coil. Now do the same for the L-1 coil. It will be necessary to put a capacitor across the voltage input of L-1 in order to match the L-1 coil to the L-2 coil. A spark gap across L1 is also required to deal with the return voltage from L-1. A tuneable capacitor of the pad ("trimmer") type for L-1 is desirable.

7. The performance of the L-2 coil can be further enhanced by having an Earth grounding from the base of the coil. The maximum voltage output will be between the base and the top of the L-2 coil. Lesser voltages can be obtained at intermediate points along the length of the L-2 coil.

SUPPLY SOURCES

1. HAM RADIO SUPPLY STORES

2. COILS, AIR INDUCTOR IN HOUSTON

3. BAKER AND WILLIAMSON (READY MADE), BRISTOL. PENNA.
ALSO R.F. DUMMY LOADS AND WATTMETERS.

NOTES

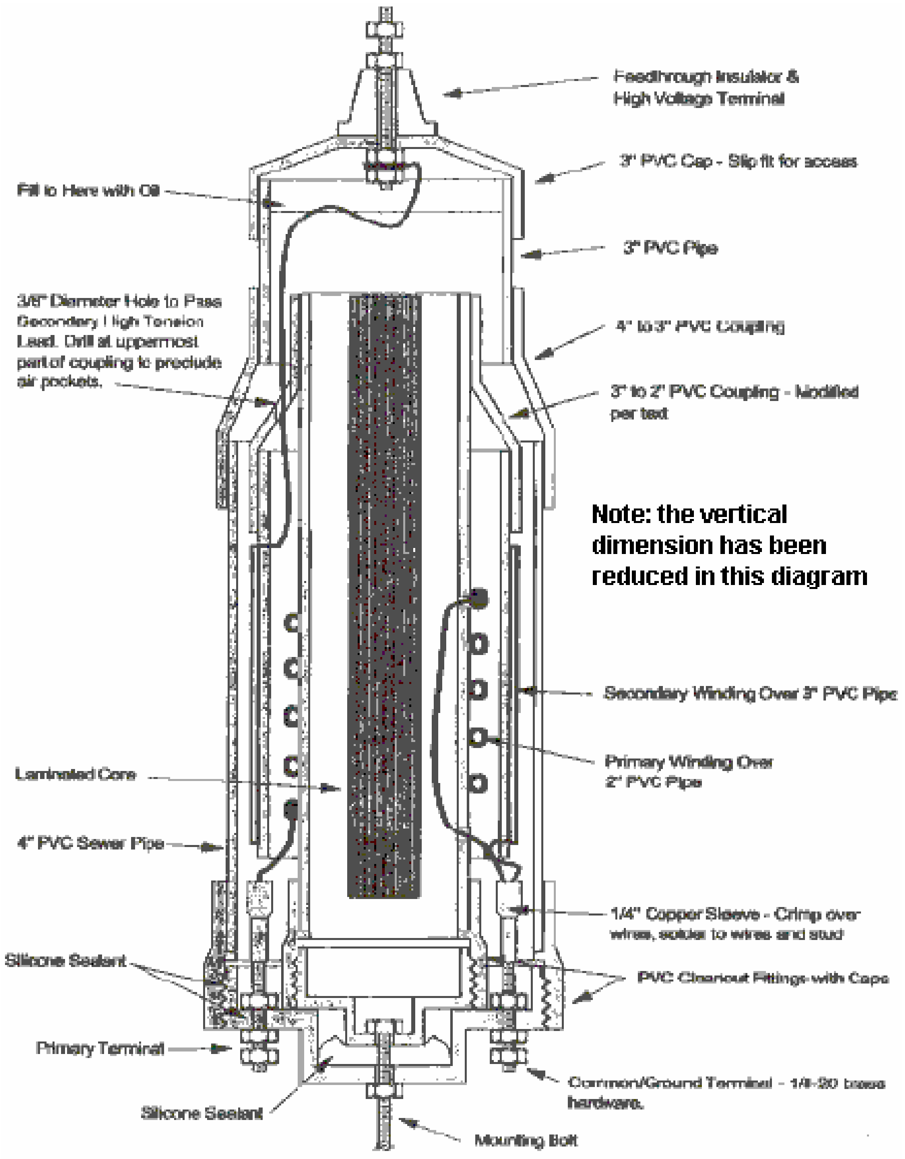
Feedthrough Insulator &
High Voltage Terminal
3" PVC Cap - Slip fit for access
Fill to Here with Oil
3" PVC Pipe
3/8" Diameter Hole to Pass
Secondary High Tension
Lead. Drill at uppermost
part of coupling to preclude
air pockets.
4" to 3" PVC Coupling
3" to 2" PVC Coupling - Modified
per text
Secondary Winding Over 3" PVC Pipe
Primary Winding Over
2" PVC Pipe
Laminated Core
4" PVC Sewer Pipe
1/4" Copper Sleeve - Crimp over
wires, solder to wires and stud
Silicone Sealant
PVC Cleanout Fittings with Caps
Primary Terminal
Common/Ground Terminal - 1/4-20 brass
hardware.
Silicone Sealant
Mounting Bolt

Note: the vertical dimension has been reduced in this diagram

Impulse Transformer Assembly

ELECTRICAL PRINCIPLES: TERMINOLOGY & SAFETY

The use of electricity is so commonplace that most people assume that it will always be available on demand. To fully realize our dependence upon electricity, consider the ways in which electricity is being used each day in the home, on the farm and the ranch. Electricity is doing more to increase work efficiency and promote enjoyable living than any other single factor. The use of electricity has grown to the extent that an increasing portion of the home or business budget, is used in paying for this source of energy.

1. Definition of Electricity

Electricity can be defined in several ways. The layman defines electricity as a source of energy that can be converted to light, heat, or power. Electrical Engineers define electricity as a movement of electrons caused by electrical pressure or voltage. The amount of energy produced depends on the number of electrons in motion.

2. The Manufacture and Distribution of Electricity

Electricity is produced from generators that are run by water, steam, or internal combustion engines. If water is used as a source of power to turn generators, it is referred to as hydroelectric generation. There are a number of this type located in areas where huge dams have been built across large streams.

Steam is used as a source of power for generating much of today's electricity. Water is heated to a high temperature, and the steam pressure is used to turn turbines which generate electricity. These are referred to as thermal-powered generators. Fuels used to heat the water are coal, natural gas, and/or fuel oil.

Generators at the power plant generate from 13,800 to 22,000 volts of electricity. From the power plant, electricity is carried to a step-up sub-station which, through the use of transformers, increases the voltage from 69,000 to 750,000 volts. This increase in voltage is necessary for the efficient transmission of electricity over long distances. From the step-up sub-station, the electricity is carried on transmission lines to a step-down sub-station which reduces the voltage to 7,200 to 14,000 volts for distribution to rural and city areas.

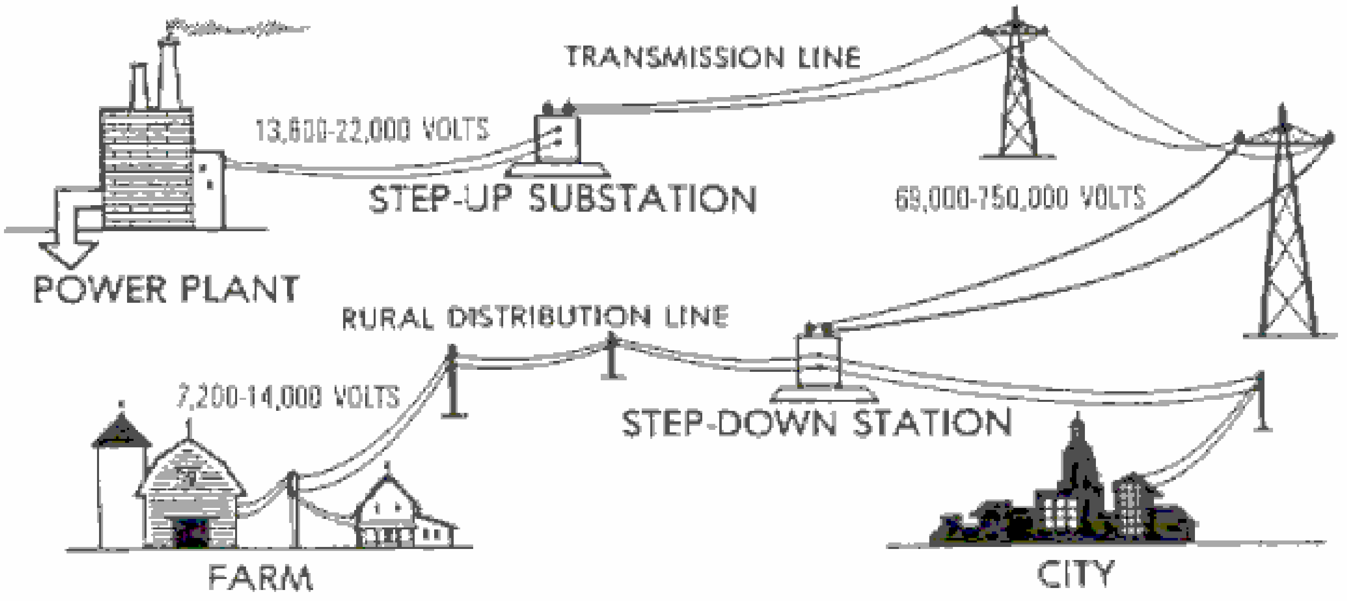
TRANSMISSION LINE
13,800-22,000 VOLTS
STEP-UP SUBSTATION
69,000-750,000 VOLTS
POWER PLANT
RURAL DISTRIBUTION LINE
7,200-14,000 VOLTS
STEP-DOWN STATION
FARM
CITY

Transformers at the business or residence reduce the voltage to 120 or 240 volts to supply the meter of the customer:

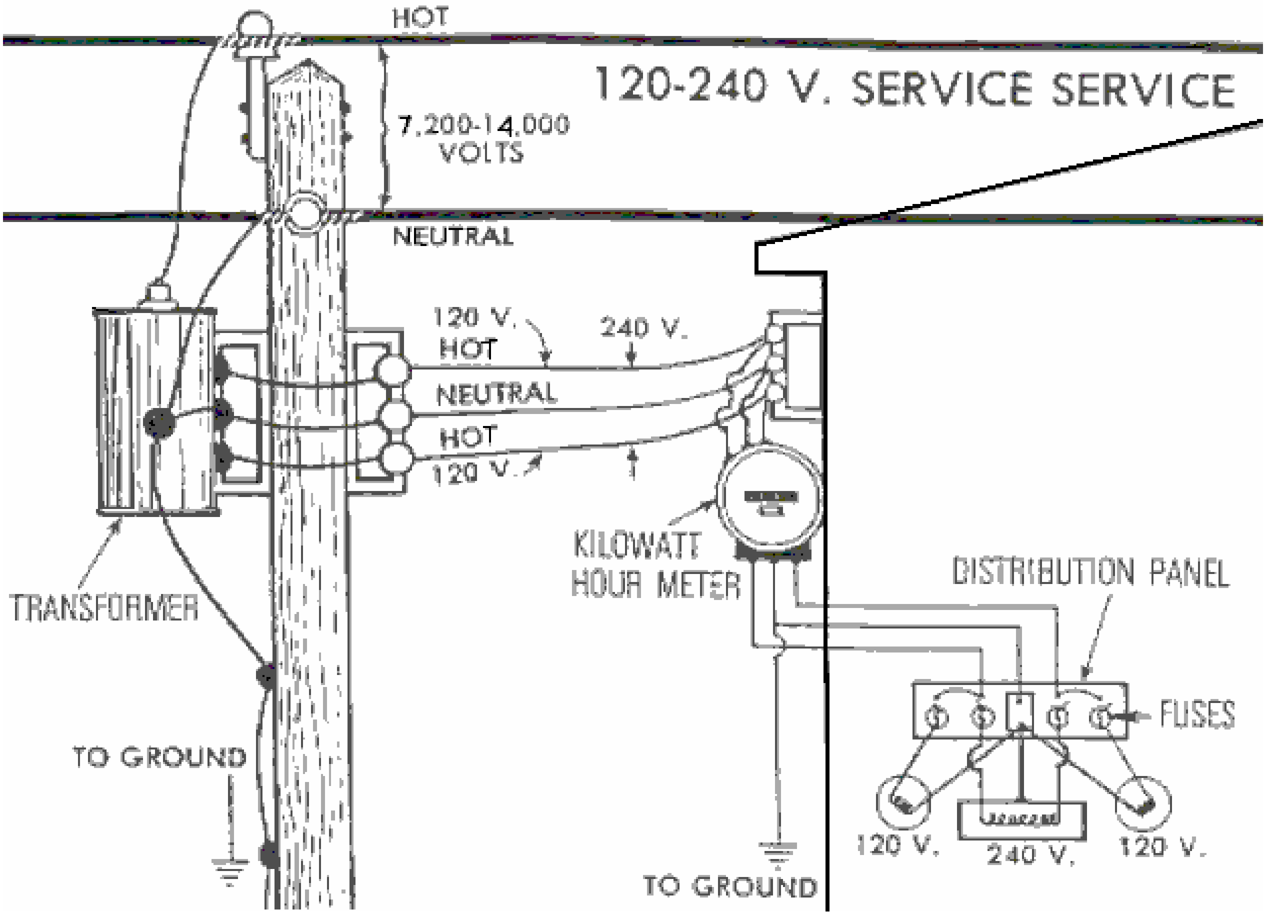
HOT
120-240 V. SERVICE SERVICE
7,200-14,000 VOLTS
NEUTRAL
120 V. HOT
240 V.
NEUTRAL
HOT
120 V.
KILOWATT HOUR METER
DISTRIBUTION PANEL
FUSES
TRANSFORMER
TO GROUND
TO GROUND
120 V.
240 V.
120 V.

3. Common Electrical Terms

In order to work safely and efficiently with electricity and have the ability to converse on the subject, the following terms should be understood:

Ampere (Amp) - A measurement in units of the rate of flow of electrical current. This may be compared with the rate of flow of water in gallons per minute.

Example: A 60-watt incandescent lamp on a 120V circuit would pull 1/2 ampere of electricity (60 divided by 120 = 0.5 or 1/2, Formula: Amperes = Watts / Volts

Volt (V) - A unit of measure of electrical pressure. A given electrical pressure (V) causes a given amount of electrical current (Amps) to flow through a load of given resistance. Voltage may be compared to water pressure in pounds per square inch in a water system. Common service voltages are 120 volts for lighting and small appliance circuits and 240 volts for heating, air conditioning, and large equipment circuits.

Watt (W) - A unit of measure of electrical power. When applied to electrical equipment, it is the rate that electrical energy is transformed into some other form of energy such as light. Watts may be compared to the work done by water in washing a car. (Formula: Volts x Amps = Watts)

Kilowatt (KW) - A unit of measurement used in computing the amount of electrical energy used. Kilowatts are determined by dividing the number of watts by 1000 as 1 kilowatt = 1,000 watts.

Kilowatt-Hour (KWH) - A measure of electricity in terms of power in kilowatts and time in hours. One KWH is 1000 watts used for one hour.

Alternating Current (A.C.) - Electrical current that alternates or changes direction several times per second. The direction current moves depends on the direction in which the voltage forces it.

Cycle - The flow of electricity in one direction, the reverse flow of electricity in the other direction, and the start of the flow back in the other direction. The cycles per second are regulated by the power supplier and are usually 60 in America. Most electric clocks are built to operate on the mains frequency. More or fewer cycles per second would cause mains-operated clocks to gain or lose time. The present practice is to use the term Hertz (Hz) rather than "cycles per second".

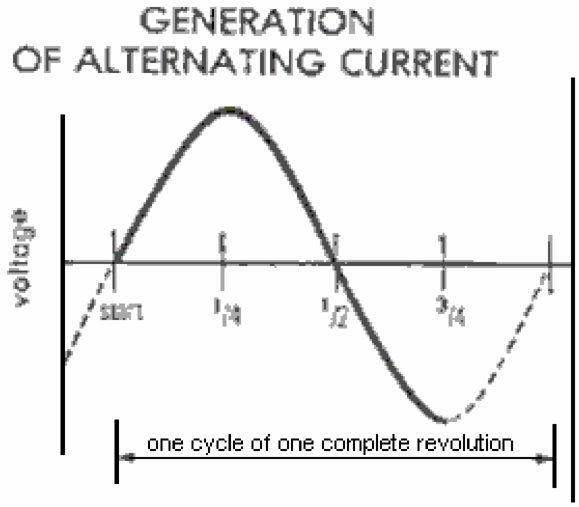
GENERATION
OF ALTERNATING CURRENT
voltage
start
1/4
1/2
3/4
one cycle of one complete revolution

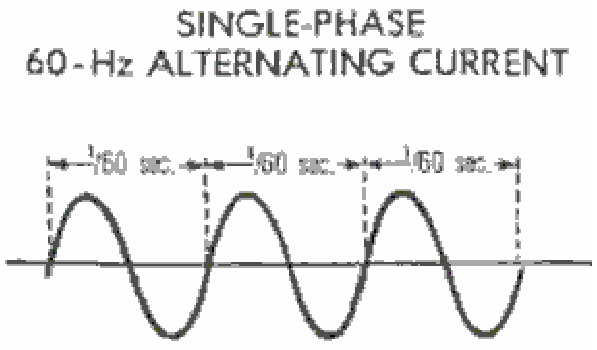
SINGLE-PHASE
60-Hz ALTERNATING CURRENT
1/60 sec.
1/60 sec.
1/60 sec.

Direct Current (D.C.) - Electrical current flowing in one direction. Example: electrical circuits in automobiles and tractors.

Transformer - A device used to increase or decrease voltage.

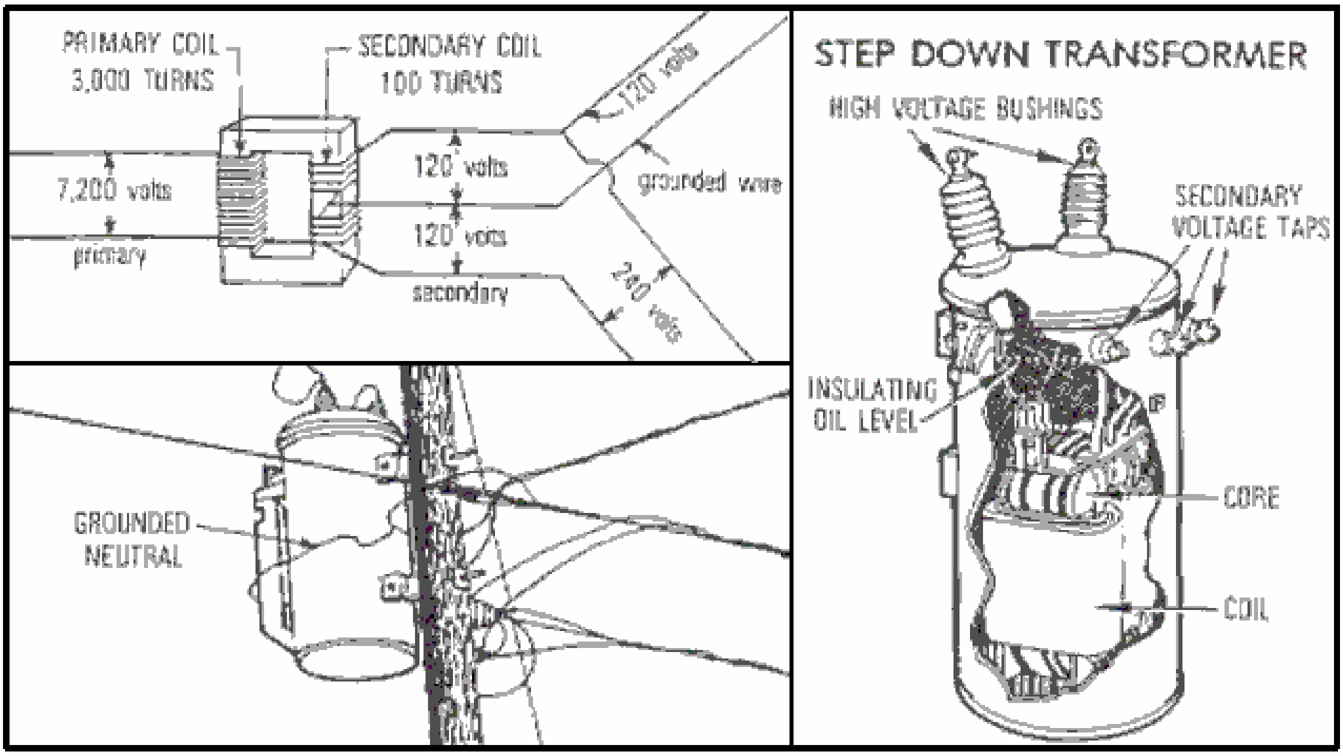
PRIMARY COIL
3,000 TURNS
SECONDARY COIL
100 TURNS
7,200 volts
primary
120 volts
120 volts
secondary
120 volts
grounded wire
240 volts
GROUNDED
NEUTRAL
STEP DOWN TRANSFORMER
HIGH VOLTAGE BUSHINGS
SECONDARY
VOLTAGE TAPS
INSULATING
OIL LEVEL
CORE
COIL

Single Phase - The most common type of electrical service or power available to consumers. One transformer is used between the distribution line and the meter. Usually three wires, two "hot" and one neutral, are installed to provide 120V and 240V single-phase service. Single-phase service may also be supplied with a three-phase service.

Three-Phase - This type of service is designed especially for large electrical loads. It is a more expensive installation due to three wires and three transformers being required. The important advantage of three-phase power is that the total electrical load is divided among the three phases, consequently, the wire and transformers can be smaller. Other advantages exist in the design of three-phase motors.

Short Circuit - A direct connection (before current flows through an appliance) between two "hot" wires, between a "hot" and neutral wire, or between a "hot" wire and ground.

Voltage Drop - A reduction of current between the power supply and the load. Due to resistance, there will be a loss of voltage any time electricity flows through a conductor (wire). Factors that influence voltage drop are size of wire, length of wire, and the number of amps flowing. A drop in voltage may cause a loss of heat, light, or the full power output of a motor. It could cause motor burn-out unless the motor is properly protected (time-delay fuse).

Fuse - A device used to protect circuits from an overload of current.

Circuit Breaker - A device used to protect circuits from an overload of current. May be manually reset.

Time-Delay Fuse - A fuse with the ability to carry an overload of current for a short duration without disengaging the contacts or melting the fuse link.

Horsepower (hp) - A unit of mechanical power equal to 746 watts of electrical power (assuming 74.6% electric motor efficiency). Motors of one horsepower and above are rated at 1000 watts per hp while motors below one horsepower are rated at 1,200 watts per hp.

Conductor - The wire used to carry electricity (typically, copper or aluminum). Copper and aluminum should not be spliced together due to their incompatibility resulting in deterioration and oxidation.

Insulator - A material which will not conduct electricity and is usually made of glass, Bakelite, porcelain, rubber, or thermo-plastic.

"Hot" Wire - A current-carrying conductor under electrical pressure and connected to a fuse or circuit breaker at the distribution panel. (Color Code: usually black or red)

Neutral Wire - A current-carrying conductor not under electrical pressure and connected to the neutral bar at the distribution panel. (Color Code: usually white)

Grounding - The connection of the neutral part of the electrical system to the earth to reduce the possibility of damage from lightning and the connection of electrical equipment housings to the earth to minimize the danger from electrical shock. (Color Code: Can be green or bare wire).

Underwriters' Laboratory (U.L.) - An American national organization which tests all types of wiring materials and electrical devices to insure that they meet minimum standards for safety and quality.

National Electric Code (N.E.C.) - Regulations approved by the National Board of Fire Underwriters primarily for safety in electrical wiring installations. All wiring should meet the requirements of the national as well as the local code.

4. Computing Electrical Energy Use and Cost

If an estimate of cost for electricity used is desired, the name plate data on appliances and equipment and an estimate of operating time may be used. The following formulas should be used for determining watts, amps, volts, watt-hours, kilowatt-hours, and cost.

Watts = Volts x Amperes

Amperes = Watts / Volts

Volts = Watts / Amperes

Watt-Hours = Watts x Hours of operation

Kilowatt-Hours = Watt-Hours / 1000

Cost = Kilowatt-Hours x Local Rate per Kilowatt-Hour (or per "Unit")

Example:

Local electricity rate per Kilowatt-Hour: 8 cents
Equipment plate data: 120 Volts 5 Amps
Monthly hours of operation: 10

1. Watts = Volts x Amperes, so Watts = 120 x 5 = 600 watts
2. Watt-Hours = 600 x 10 = 6,000 watt-hours
3. Kilowatt-Hours = 5,000 / 1,000 = 6 kilowatt-hours (or 6 Units)
4. Cost = 6 x 8 = 48 cents

5. Electrical Circuits

An Electrical Circuit is a completed path through which electricity flows. Insulated conductors (wires) provide the path for the flow of electricity. A water system and an electrical circuit are similar in many respects. Water flows through pipes and is measured in gallons per minute, and electricity flows through conductors and is measured in amperes. A simple circuit is shown here:

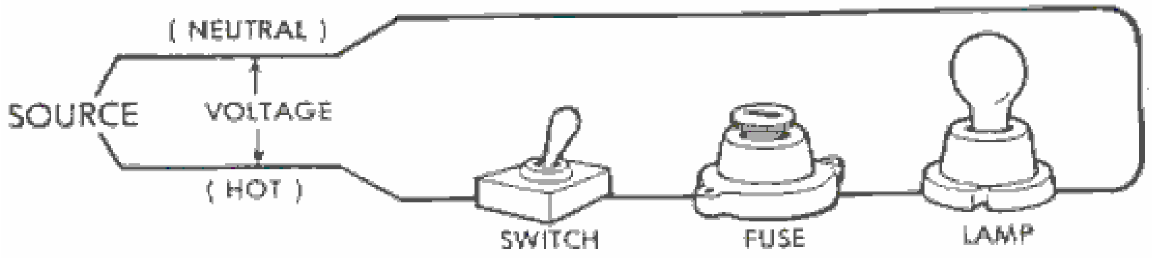

A circuit includes a "hot" wire (red or black) carrying current from the source through a switch, circuit protector (fuse or circuit-breaker), and an appliance. The neutral wire (white) conducts the current from the appliance to the source (ground).

There arc two methods for connecting devices in a circuit - "in series" or "in parallel". In a series circuit, all of the current must flow through each device in the circuit. Removing any one of the devices in a series circuit will stop the flow of current. In parallel circuits, the load (lights or appliances) are connected between the two wires of the circuit providing an independent path for the flow of current, and removing a lamp has no effect on the other lamps in the circuit.

Switches, fuses, and circuit breakers are always connected in series. In most cases, except for some Christmas tree lights, appliances and lights are connected in parallel.

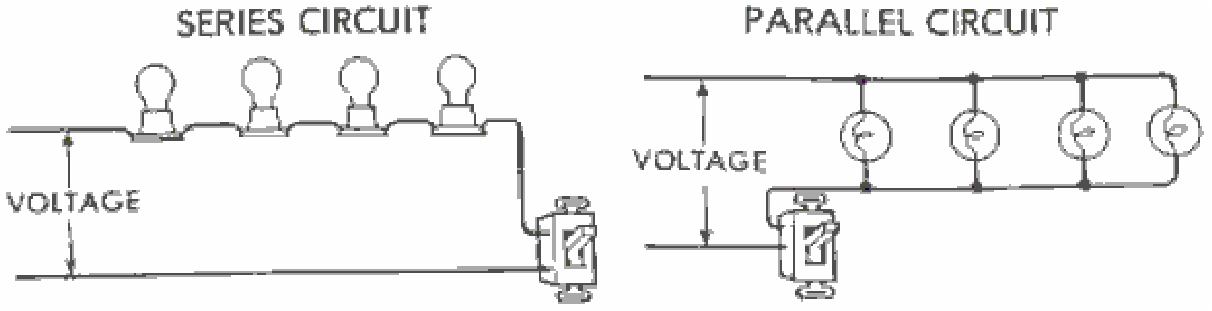

6. 120 Volt and 240 Volt Circuits

The 120V circuit has one "hot" and one neutral wire, with the switch and circuit protector in the hot line. The neutral wire from the appliance is connected to the neutral bar in the fuse or breaker box. For safety, the neutral wire should never be broken or interrupted with a switch or fuse.

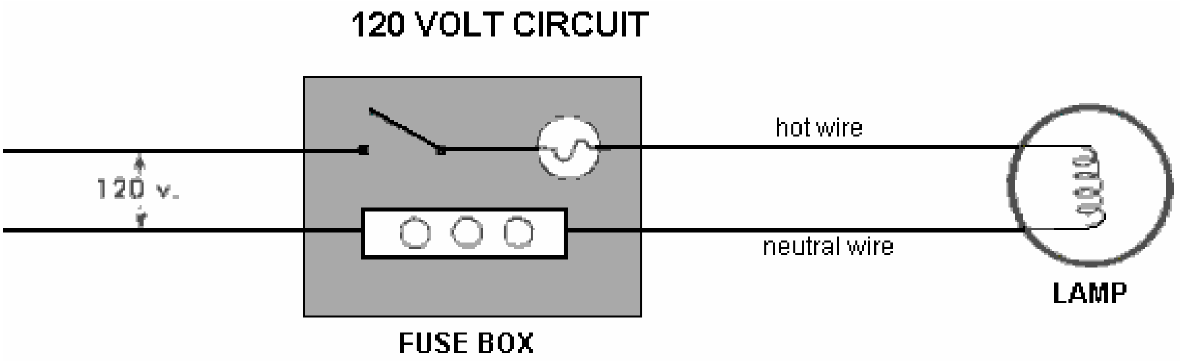

The voltage in a 120V circuit is measured with a voltmeter with one lead on the hot terminal and the other lead on the neutral bar. The number of amperes flowing may be measured with a clamp-on ammeter by encircling the hot or neutral wire with the jaws of the ammeter.

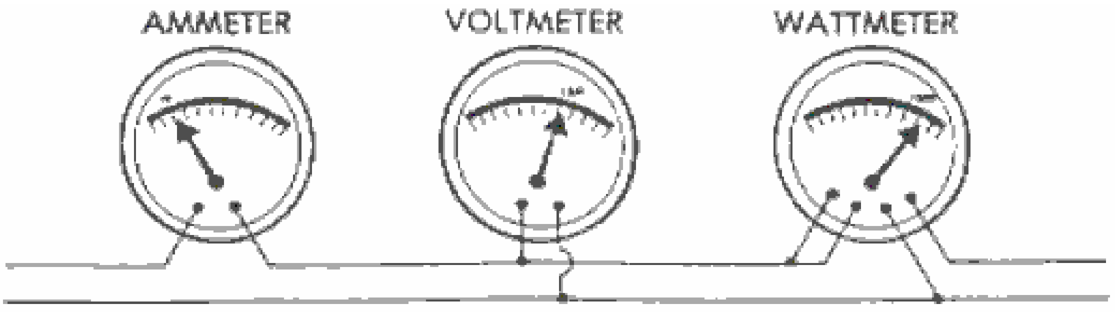

The 240V circuit has two hot wires and one safety-ground wire. Switches and fuses are installed in the hot lines. The two hot wires arc necessary for the operation of 240V welders and motors. The safety-ground wire, connected to the metal frame of the equipment or motor and to the neutral bar, does not carry current unless a "short" develops in the motor or welder. If a short should occur, one of the circuit protectors will burn-out or open, thus opening the circuit.

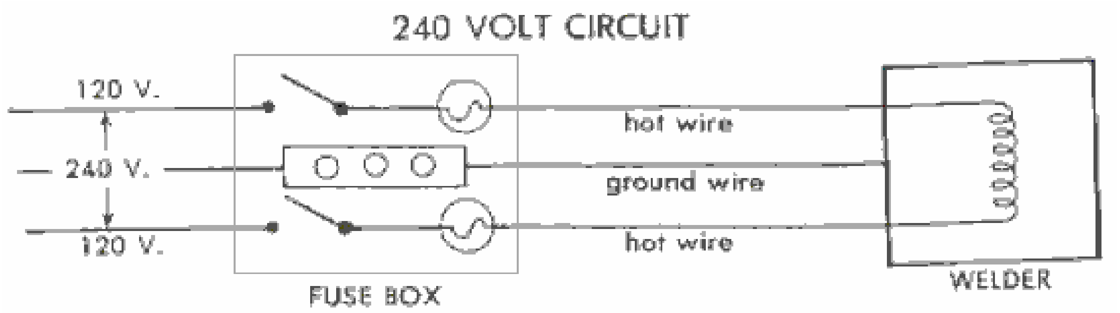
240 VOLT CIRCUIT
120 V.
240 V.
120 V.
hot wire
ground wire
hot wire
FUSE BOX
WELDER

The voltage on a 240V circuit is measured by fastening a lead on the voltmeter to each of the hot wires. Voltage between either hot terminal and the neutral bar will be one-half of the voltage between the two hot wires. The number of amperes flowing can be measured by clamping an ammeter around either of the hot wires.

7. Safety Grounding of Electrical Equipment

Refer back, to the 240V circuit and note the ground wire from the metal frame to the neutral bar. The following illustration shows proper safety grounding when operating a drill in a 120V circuit. The safety-ground wire may be bare, but a three-wire cable is recommended. Safety-ground wire in three-wire cable is usually green in color. A current-carrying neutral wire should never be used for a safety-ground. Likewise, a safety-ground wire should never be used as a current-carrying hot or neutral wire.

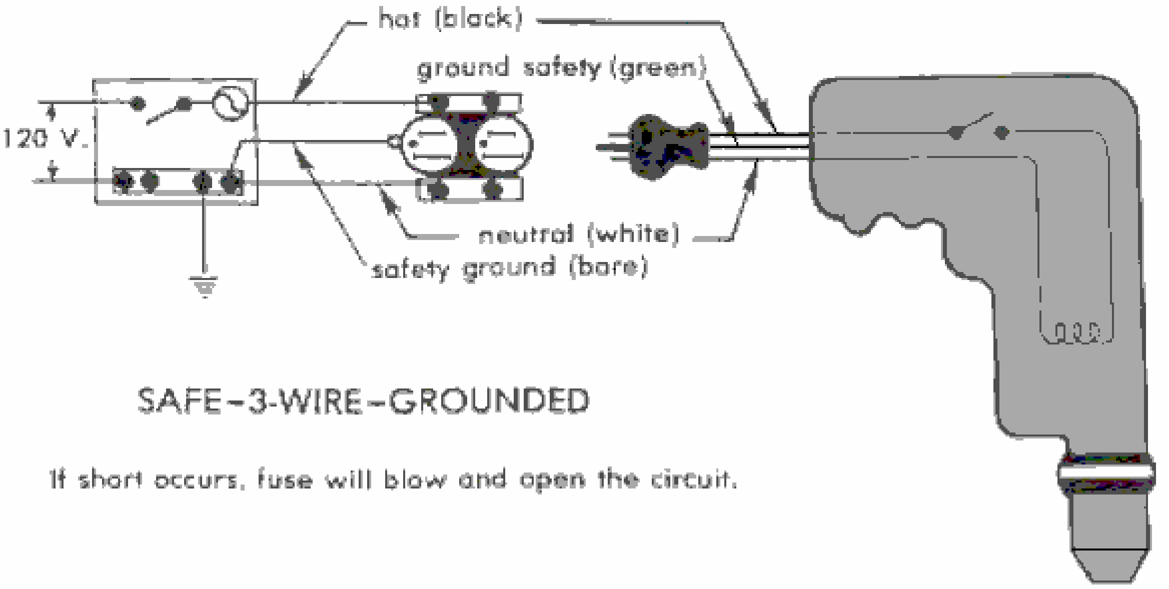
hot (black)
ground safety (green)
120 V.
neutral (white)
safety ground (bare)
SAFE-3-WIRE-GROUNDED
If short occurs, fuse will blow and open the circuit.

Using grounded receptacles and a safety-ground on all circuits will allow the safety-grounding of appliances when they are plugged into the outlet. An adapter must be used to properly ground appliances connected to receptacles which are not safety-grounded. If an adapter is used, the green pigtail wire must be connected to a known ground to give protection from electrical shock should a short-circuit occur.

GROUND PLUG AND ADAPTER

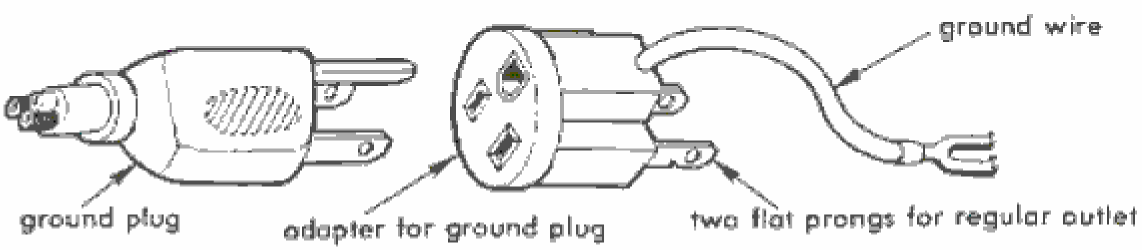

A test lamp can be used to check a circuit completed between a "hot" wire and a neutral wire. Use the test lamp to check appliances for shorts. With the appliance plugged into an outlet, touch the appliance frame with one lead of the test lamp while the other lead of the test lamp is grounded to a water or gas pipe. If the test light does not burn, reverse the appliance plug and check with the test lamp again. If the light burns, a short exists (the hot wire is touching the frame of the appliance). Unplug the appliance and repair or discard it.

8. Electrical Circuit Protection

Electrical circuits should be protected from an overload of amperes. Too many amperes flowing through an unprotected circuit will generate heat, which will deteriorate or melt the insulation and possibly cause a fire. The number of amperes that a given conductor can safely carry, depends upon the kind and size of wire, type of insulation, length of run in feet, and the type of installation. Charts are available in reference texts giving allowable current-carrying capabilities of various conductors.

The four types of circuit protection are: common fuses, fusetrons (time-delay), fustats (two-part time-delay), and circuit-breakers. Fuses are of two basic types: plug, and cartridge.

Common fuses contain a link made from a low-temperature melting alloy which is designed to carry current up to the rating of the fuse. Current higher than the amperage rating causes the link to heat above it's melting point. When the fuse "blows", the link melts and opens the circuit.

PLUG TYPE FUSE

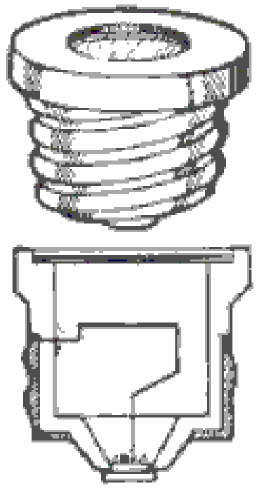

CARTRIDGE TYPE FUSE

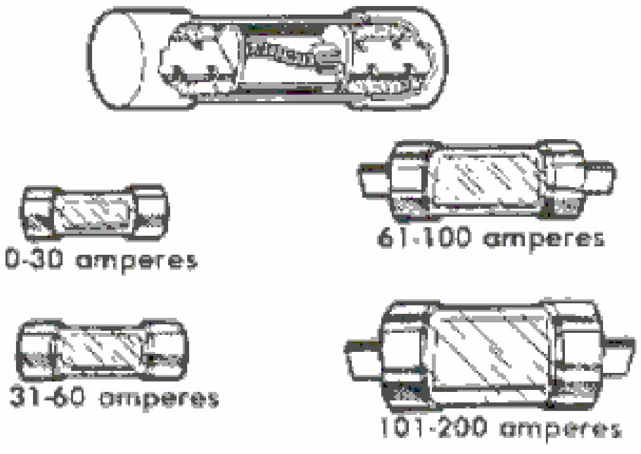

Fusetrons (time-delay fuses) are made to carry a temporary overload, such as the overload caused by the starting of an electric motor. The fuse, however, still provides protection for the circuit, and a short-circuit will melt the fuse link. If a common fuse is used, the fuse link will melt every time an electric motor starts. The use of a larger ampere common fuse will prevent the "blow" resulting from the temporary overload, but will not provide protection for the motor or the circuit.

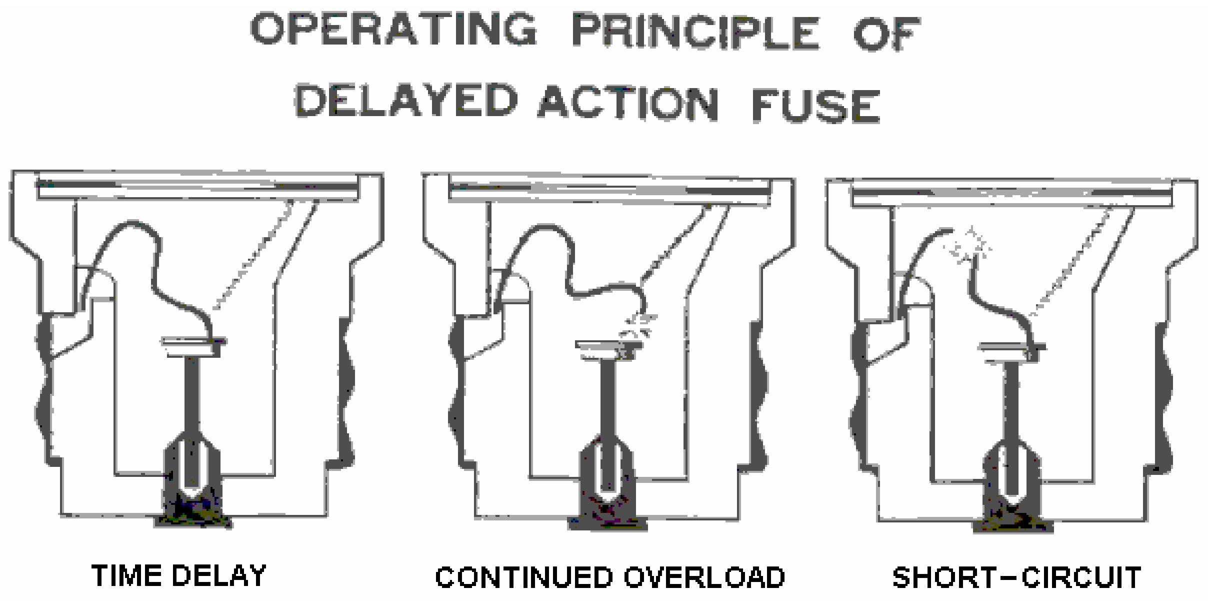
OPERATING PRINCIPLE OF DELAYED ACTION FUSE

Fustats, non-tamperable fuses of the time-delay type, have a different size base and require a special adapter which is screwed into the standard fuse socket. After the adapter is installed, it cannot be removed. For example, the installation of a 15-ampere adapter allows only the use of 15-ampere or smaller fuse.

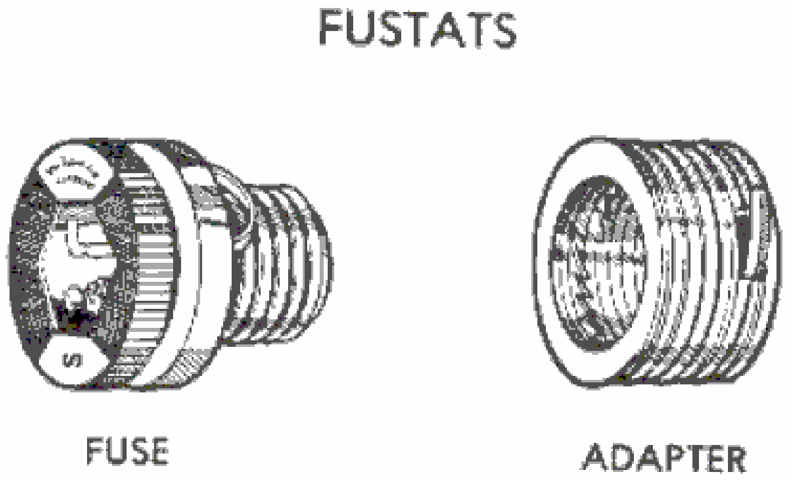
FUSTATS

Circuit breakers eliminate the replacement of fuses and are commonly used even though a circuit breaker box costs more than a fuse box. Circuit breakers are of two types, thermal and magnetic. The thermal breaker has two contacts held together by a bi-metal latch. A current overload causes the bi-metallic strip to become heated, the latch releases, and the points spring open. After the bi-metallic strip cools, the switch is reset and service is restored.

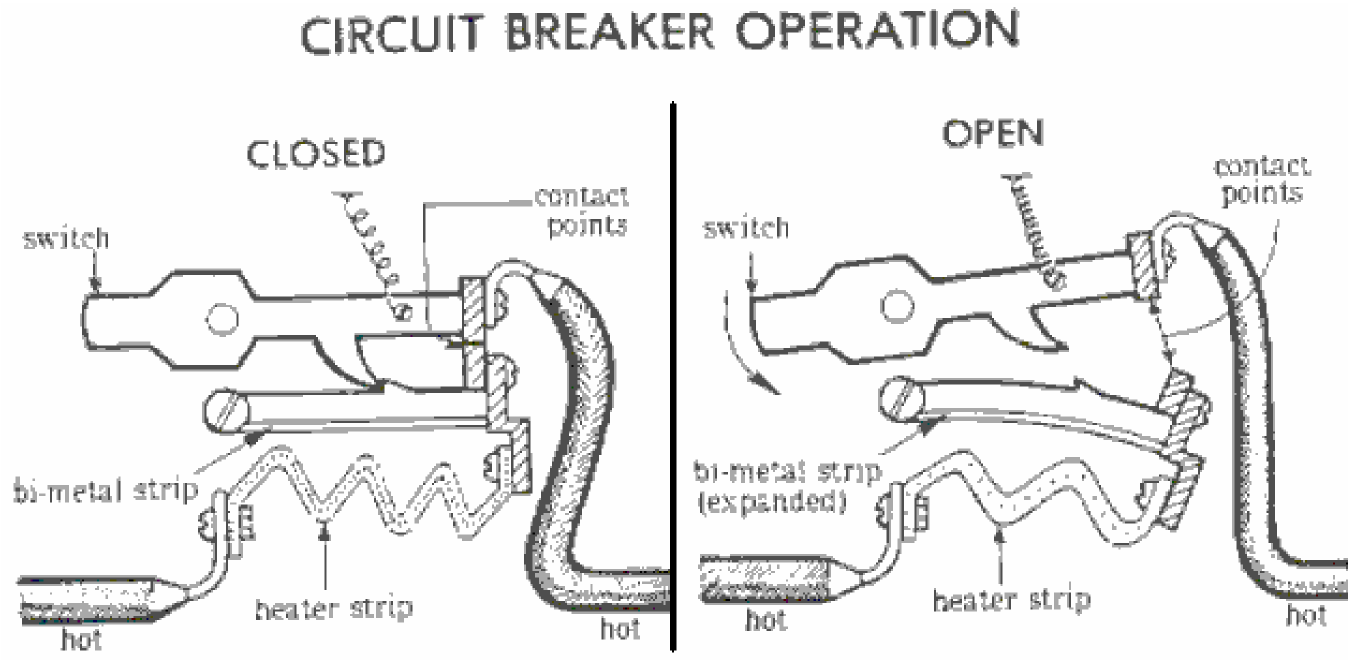
CIRCUIT BREAKER OPERATION
CLOSED
switch
contact points
bi-metal strip
heater strip
hot
hot
OPEN
switch
contact points
bi-metal strip (expanded)
heater strip
hot
hot

The magnetic breaker has contacts that are held together by a latch which is released by the action of an electromagnet. The amount of current flowing through the circuit will determine the size of the electromagnet. This type of circuit-breaker is reset by moving the toggle switch to the "on" position.

The following diagram shows the parts of a circuit breaker.

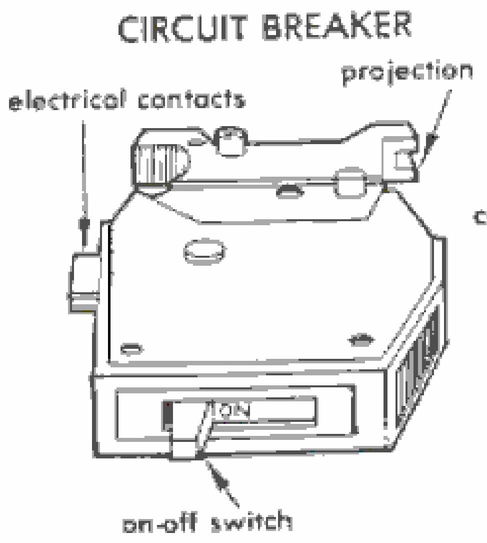
CIRCUIT BREAKER
projection
electrical contacts
ON
on-off switch

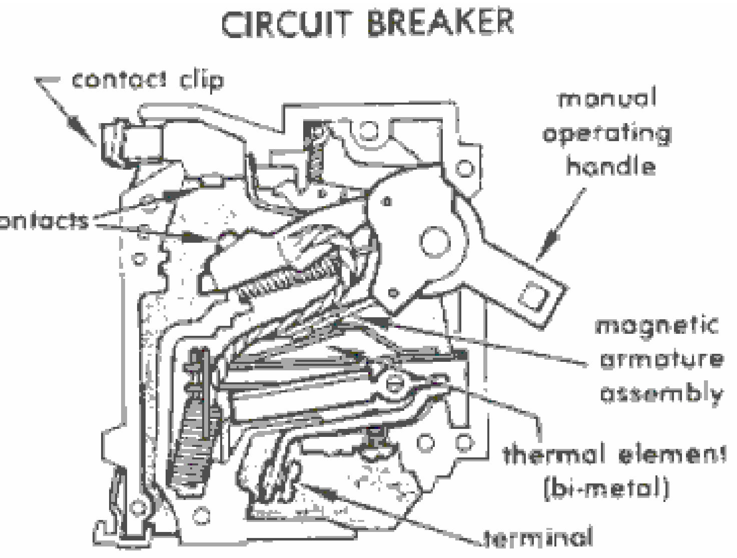
CIRCUIT BREAKER
contact clip
manual operating handle
contacts
magnetic armature assembly
thermal element (bi-metal)
terminal

9. No Fault Grounding

Fuses and circuit-breakers are safety devices which limit current (amperage) in a circuit. Their main function is to protect equipment and wiring from overload. Ground fault circuit interrupters (GFI) are designed to protect humans, equipment, and/or electrical systems from injury or damage if electricity flows in an unintended path (a short-circuit).

A GFI is a very sensitive device that functions by comparing the current moving in the "hot" wire with that in the neutral wire. If these two currents are not equal, a fault exists, and current is "leaking" out of the circuit. If the difference in current between the two wires is 5/1000 of an ampere or greater, the GFI will open the circuit, shutting off the power and eliminating any shock hazard.

The National Electrical Code requires GFl's for all 120V, single phase, 15 and 20 amp receptacles installed outdoors, in bathrooms, and in garages for residential buildings. A GFI is required at construction sites and some other applications. After correcting a circuit fault, the GFI may be reset for further use.

A variety of GFI equipment is made for 120 and 240 volt circuits:

GFI RECEPTACLE FOR CONVENTIONAL OUTLET

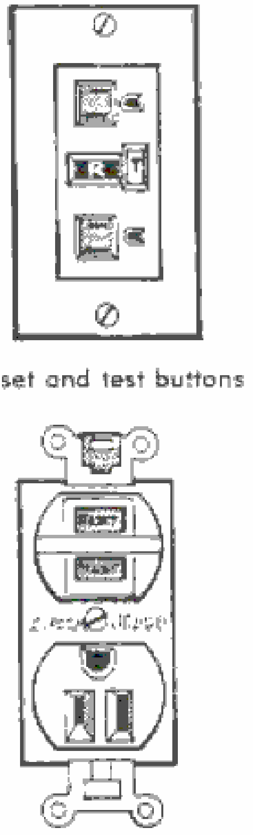

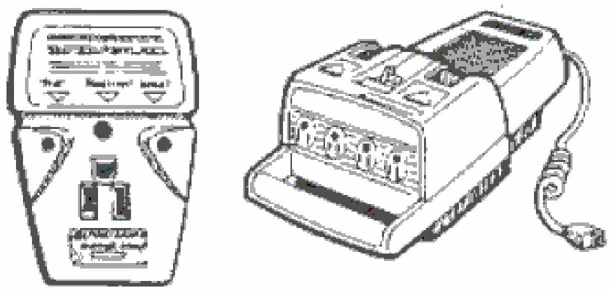

PORTABLE GFI TO PLUG INTO OUTLET

GFI TO INSTALL IN BRANCH CIRCUIT

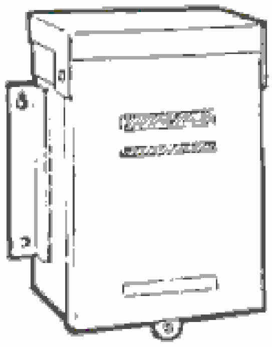

GFI TO REPLACE CIRCUIT BREAKER

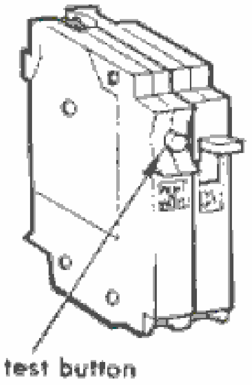

REFERENCES:

COOPER, ELMER L., *AGRICULTURAL MECHANICS: FUNDAMENTALS AND APPLICATIONS*. DELMAR PUBLISHERS INC., ALBANY, NEW YORK

ELECTRICAL WIRING - RESIDENTIAL, UTILITY BUILDINGS, SERVICE AREAS, AAVIM, ATHENS, GEORGIA.

Note: This electrical information does not apply directly to areas outside America and local regulations for electrical supply should be checked.

The following material is the section originally in Chapter 3 of "The Practical Guide to Free-Energy Devices" which deals with Don Smith's technology:

Don Smith. One of most impressive developers of free-energy devices is Don Smith who has produced many spectacular devices, generally with major power output. These are a result of his in-depth knowledge and understanding of the way that the environment works. Don says that his understanding comes from the work of Nikola Tesla as recorded in Thomas C. Martin's book "The Inventions, Researches, and Writings of Nikola Tesla" ISBN 0-7873-0582-0 available from http://www.healthresearchbooks.com and various other book companies. This book can be downloaded from http://www.free-energy-info.com as a pdf file, but a paper copy is much better quality and easier to work from.

Don states that he repeated each of the experiments found in the book and that gave him his understanding of what he prefers to describe as the 'ambient background energy' which is called the 'zero-point energy field' elsewhere in this eBook. Don remarks that he has now advanced further than Tesla in this field, partly because of the devices now available to him and which were not available when Tesla was alive.

Don stresses two key points. Firstly, a dipole can cause a disturbance in the magnetic component of the 'ambient background' and that imbalance allows you to collect large amounts of electrical power, using capacitors and inductors (coils). Secondly, you can pick up as many powerful electrical outputs as you want from that one magnetic disturbance, without depleting the magnetic disturbance in any way. This allows massively more power output than the small power needed to create the magnetic disturbance in the first place. This is what produces a COP>1 device and Don has created nearly fifty different devices based on that understanding.

Although they get removed quite frequently, there is one video which is definitely worth watching if it is still there. It is located at http://www.metacafe.com/watch/2820531/don_smith_free_energy/ and was recorded in 2006. It covers a good deal of what Don has done. In the video, reference is made to Don's website but you will find that it has been taken over by Big Oil who have filled it with innocuous similar-sounding things of no consequence, apparently intended to confuse newcomers. A website which I understand is run by Don's son is http://www.28an.com/altenergypro/index.htm and it has brief details of his prototypes and theory. You will find the only document of his which I could locate, here http://www.free-energy-info.com/Smith.pdf in pdf form and it contains the following patent on a most interesting device which appears to have no particular limit on the output power. This is a slightly re-worded copy of that patent as patents are generally worded in such a way as to make them difficult to understand.

Patent NL 02000035 A **20th May 2004** **Inventor: Donald Lee Smith**

TRANSFORMER GENERATOR MAGNETIC RESONANCE INTO ELECTRIC ENERGY

ABSTRACT

The present invention refers to an Electromagnetic Dipole Device and Method, where wasted radiated energy is transformed into useful energy. A Dipole as seen in Antenna Systems is adapted for use with capacitor plates in such a way that the Heaviside Current Component becomes a useful source of electrical energy.

DESCRIPTION

Technical Field:

This invention relates to loaded Dipole Antenna Systems and their Electromagnetic radiation. When used as a transformer with an appropriate energy collector system, it becomes a transformer/generator. The invention collects and converts energy which is radiated and wasted by conventional devices.

Background Art:

A search of the International Patent Database for closely related methods did not reveal any prior art with an interest in conserving radiated and wasted magnetic waves as useful energy.

DISCLOSURE OF THE INVENTION

The invention is a new and useful departure from transformer generator construction, such that radiated and wasted magnetic energy changes into useful electrical energy. Gauss meters show that much energy from conventional electromagnetic devices is radiated into the ambient background and wasted. In the case of conventional transformer generators, a radical change in the physical construction allows better access to the energy available. It is found that creating a dipole and inserting capacitor plates at right angles to the current flow, allows magnetic waves to change back into useful electrical (coulombs) energy. Magnetic waves passing through the capacitor plates do not degrade and the full impact of the available energy is accessed. One, or as many sets of capacitor plates as is desired, may be used. Each set makes an exact copy of the full force and effect of the energy present in the magnetic waves. The originating source is not depleted of degraded as is common in conventional transformers.

BRIEF DESCRIPTION OF THE DRAWINGS

The Dipole at right angles, allows the magnetic flux surrounding it to intercept the capacitor plate, or plates, at right angles. The electrons present are spun such that the electrical component of each electron is collected by the capacitor plates. Essential parts are the South and North component of an active Dipole. Examples presented here exist as fully functional prototypes and were engineer constructed and fully tested in use by the Inventor. In each of the three examples shown in the drawings, corresponding parts are used.

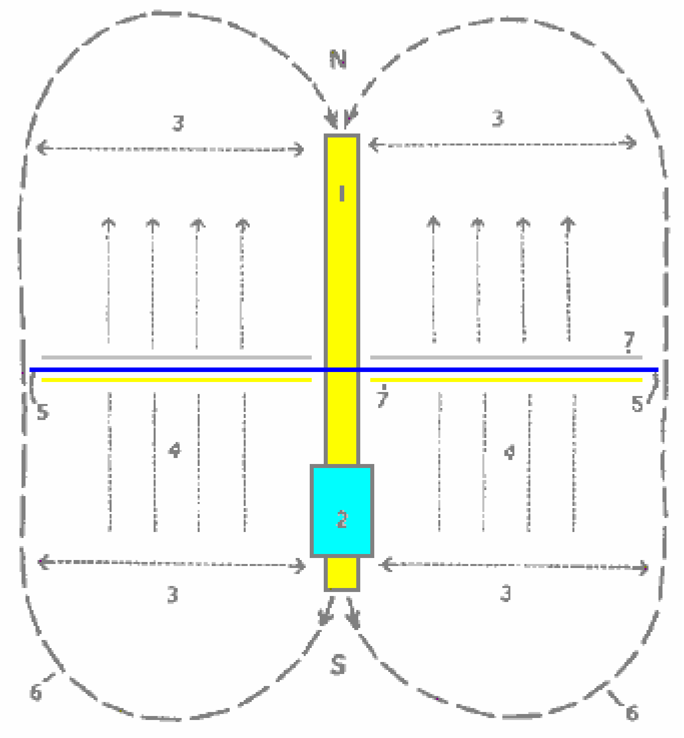

Fig.1 is a View of the Method, where **N** is the North and **S** is the South component of the Dipole.

Here, **1** marks the Dipole with its North and South components. **2** is a resonant high-voltage induction coil. **3** indicates the position of the electromagnetic wave emission from the Dipole. **4** indicates the position and flow direction of the corresponding Heaviside current component of the energy flow caused by the induction coil **2**. **5** is the dielectric separator for the capacitor plates **7**. **6** for the purposes of this drawing, indicates a virtual limit for the scope of the electromagnetic wave energy.

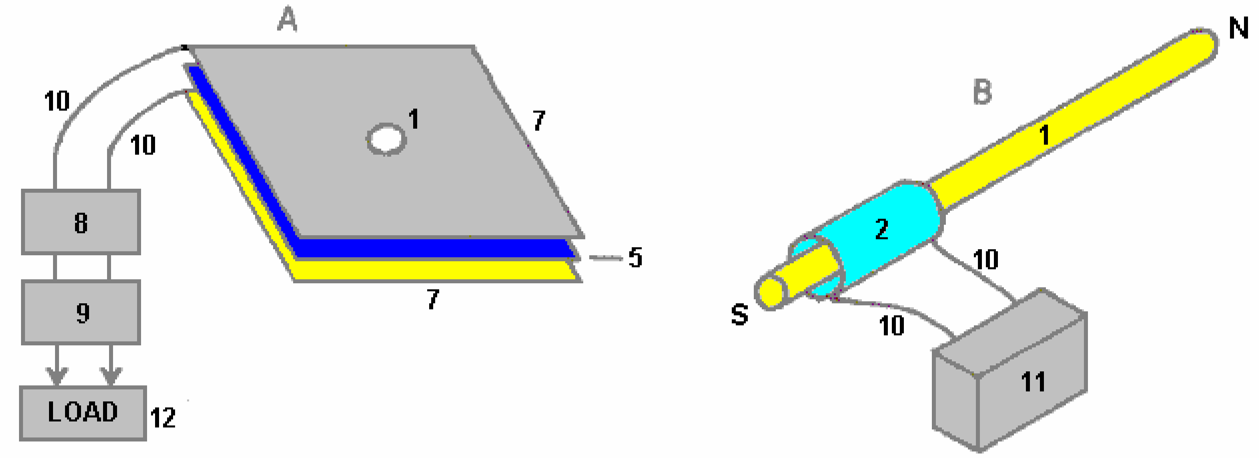

Fig.2 has two parts **A** and **B**.

In **Fig.2A** **1** is the hole in the capacitor plates through which the Dipole is inserted and in **Fig.2B** it is the Dipole with its North and South poles shown. **2** is the resonant high-voltage induction coil surrounding part of the Dipole **1**. The dielectric separator **5**, is a thin sheet of plastic placed between the two capacitor plates **7**, the upper plate being made of aluminium and the lower plate made of copper. Unit **8** is a deep-cycle battery system powering a DC inverter **9** which produces 120 volts at 60 Hz (the US mains supply voltage and frequency, obviously, a 240 volt 50 Hz inverter could be used here just as easily) which is used to power whatever equipment is to be driven by the device. The reference number **10** just indicates connecting wires. Unit **11** is a high-voltage generating device such as a neon transformer with its oscillating power supply.

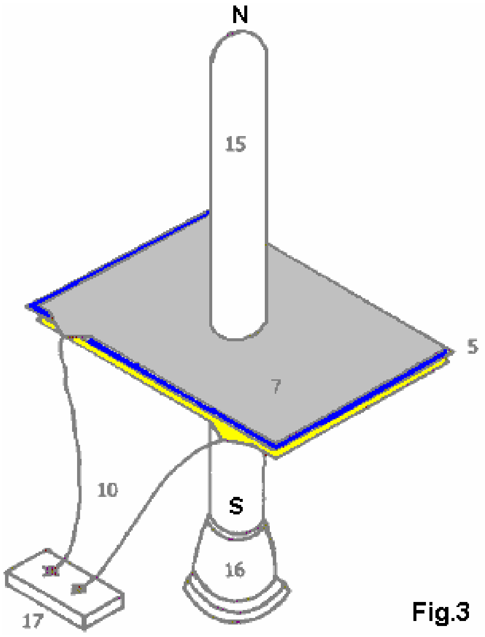

Fig.3

Fig.3 is a Proof Of Principal Device using a Plasma Tube as an active Dipole. In this drawing, **5** is the plastic sheet dielectric separator of the two plates **7** of the capacitor, the upper plate being aluminium and the lower plate copper. The connecting wires are marked **10** and the plasma tube is designated **15**. The plasma tube is four feet long (1.22 m) and six inches (100 mm) in diameter. The high-voltage energy source for the active plasma dipole is marked **16** and there is a connector box **17** shown as that is a convenient method of connecting to the capacitor plates when running tests on the device.

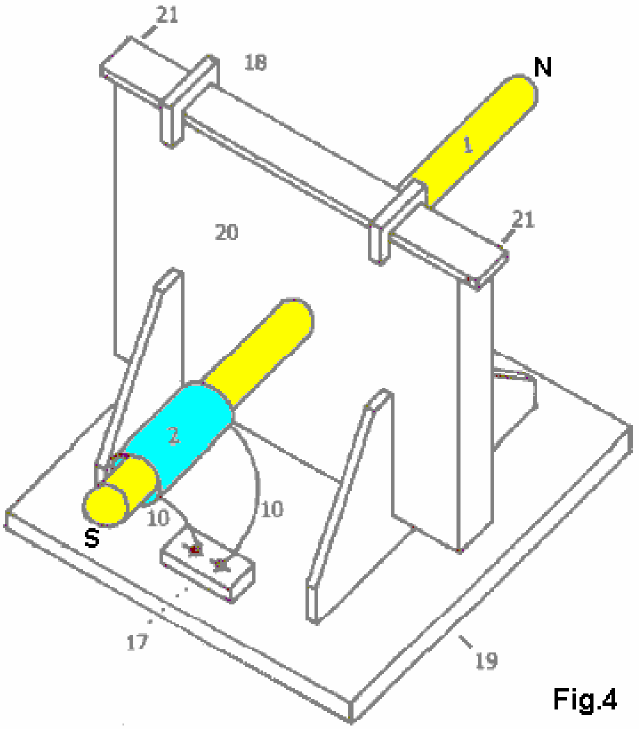

Fig.4

Fig.4 shows a Manufacturer's Prototype, constructed and fully tested. **1** is a metal Dipole rod and **2** the resonant high-voltage induction coil, connected through wires **10** to connector block **17** which facilitates the connection of it's high-voltage power supply. Clamps **18** hold the upper edge of the capacitor packet in place and **19** is the base plate with it's supporting brackets which hold the whole device in place. **20** is a housing which contains the capacitor plates and **21** is the point at which the power output from the capacitor plates is drawn off and fed to the DC inverter.

BEST METHOD OF CARRYING OUT THE INVENTION

The invention is applicable to any and all electrical energy requirements. The small size and it's high efficiency make it an attractive option, especially for remote areas, homes, office buildings, factories, shopping centres, public places, transportation, water systems, electric trains, boats, ships and 'all things great and small'. The construction materials are commonly available and only moderate skill levels are needed to make the device.

CLAIMS

1. Radiated magnetic flux from the Dipole, when intercepted by capacitor plates at right angles, changes into useful electrical energy.

2. A Device and Method for converting for use, normally wasted electromagnetic energy.

3. The Dipole of the Invention is any resonating substance such as Metal Rods, Coils and Plasma Tubes which have interacting Positive and Negative components.

4. The resulting Heaviside current component is changed to useful electrical energy.

****

This patent does not make it clear that the device needs to be tuned and that the tuning is related to its physical location. The tuning will be accomplished by applying a variable-frequency input signal to the neon transformer and adjusting that input frequency to give the maximum output.

Don Smith has produced some forty eight different devices, and because he understands that the real power in the universe is magnetic and not electric, these devices have performances which appear staggering to people trained to think that electrical power is the only source of power. One device which I understand is commercially produced in Russia, is shown here:

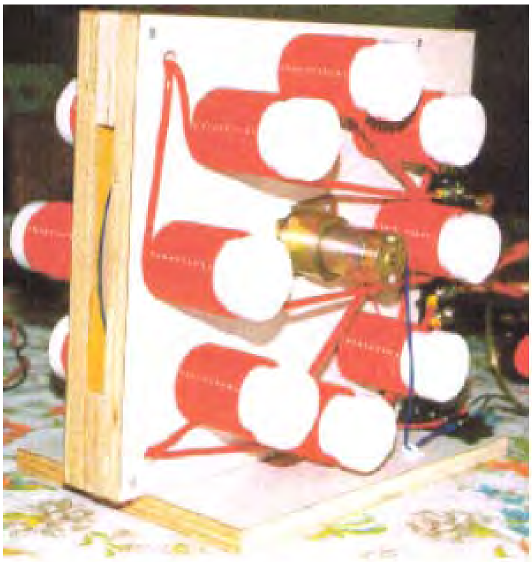

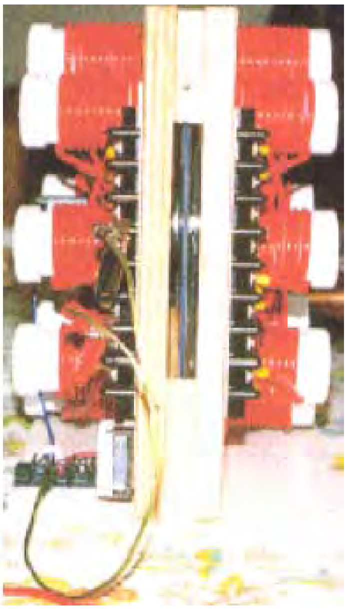

This is a small table-top device which looks like it is an experiment by a beginner, and something which would be wholly ineffective. Nothing could be further from the truth. Each of the eight coils pairs (one each side of the rotating disc) produces 1,000 volts at 50 amps (fifty kilowatts) of output power, giving a total power output of 400 kilowatts. It's overall size is 16" x 14.5" x 10" (400 x 370 x 255 mm). In spite of the extremely high power output, the general construction is very simple:

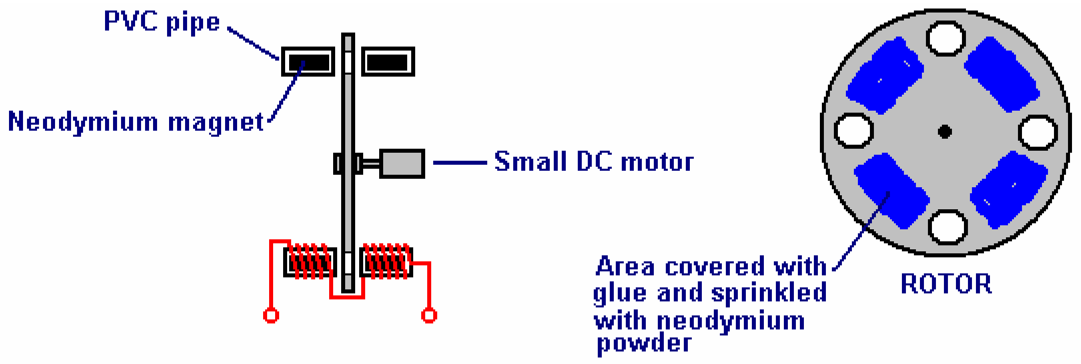

The device operates on a fluctuating magnetic field which is produced by a small low-power DC motor spinning a plastic disc. In the prototype shown above, the disc is an old vinyl record which has had holes cut in it. Between the holes is an area which was covered with glue and then sprinkled with powdered neodymium magnet material. It takes very little power to spin the disc, but it acts in a way which is very much like the Ecklin-Brown generator, repeatedly disrupting the magnetic field. The magnetic field is created by a neodymium magnet in each of the sixteen plastic pipes. It is important that the change in the magnetic flux between the matching magnets on each side of the disc is as large as possible. The ideal rotor material for this is "Terfenol-D" (tungsten zirconate) with slots cut in it but it is so expensive that materials like stainless steel are likely to be used instead. Please understand that all of Don's designs rely on resonant operation and so the coil impedance has to be matched to the pulse frequency used to drive the coil.

For Don Smith, this is not an exceptional device. The one shown below is also physically quite small and yet it has an output of 160 kilowatts (8000 volts at 20 amps) from an input of 12 volts 1 amp (COP = 13,333):

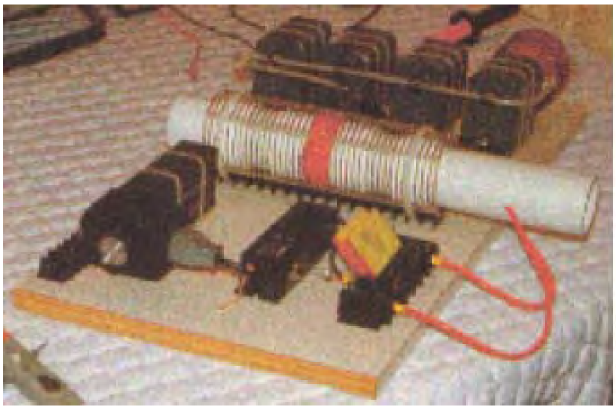

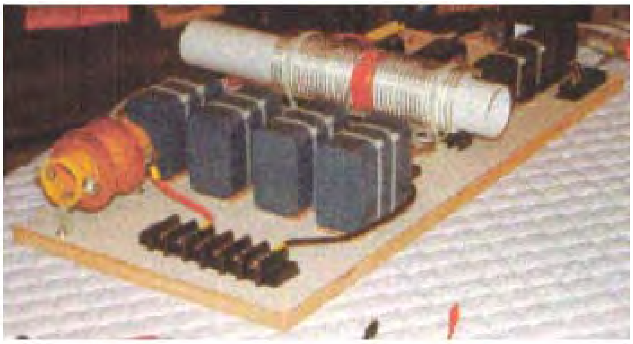

Again, this is a device which can be placed on top of a table and is not a complicated form of construction, having a very open and simplistic layout. However, some components are not mounted on this board. The twelve volt battery and connecting leads are not shown, nor is the ground connection, the step-down isolation transformer and the varistor used to protect the load from over-voltage by absorbing any random voltage spikes which might occur, but more of these things later on when a much more detailed description of this device is given. Again, please understand that Don does not reveal all of the details of any of his designs, and he deliberately omits to mention various important details, leaving us to deduce what is missing from our own understanding of how these devices work.

The device shown above is a typical example of this with various subtle points glossed over in spite of this being one device which Don says that we should be able to reproduce ourselves. Let me state here that reproducing this seemingly simple design of Don's is not an easy thing to do and it is not something which can be thrown together by a beginner using whatever components happen to be at hand at the time. Having said that, with careful study and commonsense application of some obvious facts, it should be possible to make one of these devices.

Another of Don's devices is shown here:

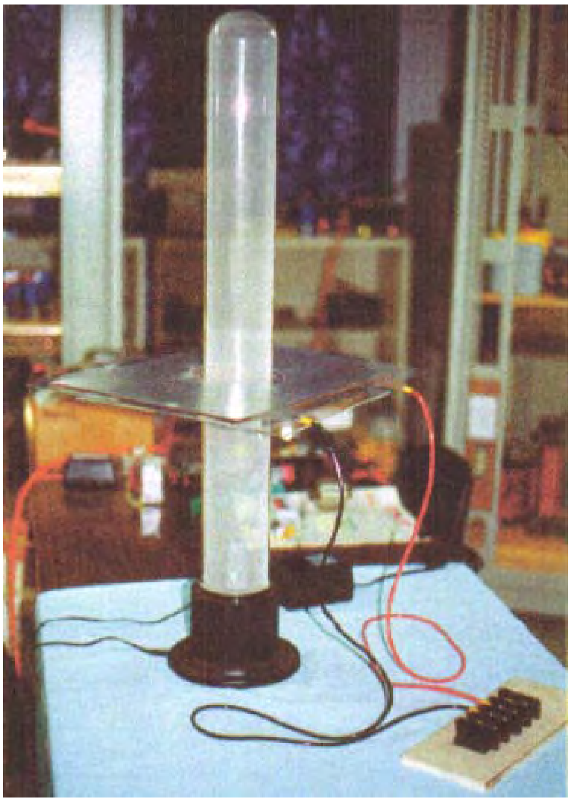

This is a larger device which uses a plasma tube four feet (1.22 m) long and 6 inches (100 mm) in diameter. The output is a massive 100 kilowatts. This is the design shown as one of the options in Don's patent. Being an Electrical Engineer, none of Don's prototypes are in the "toy" category. If nothing else is taken from Don's work, we should realise that high power outputs can be had from very simple devices.

There is one other brief document "Resonate Electrical Power System" from Don Smith which says:

Potential Energy is everywhere at all times, becoming useful when converted into a more practical form. There is no energy shortage, only grey matter. This energy potential is observed indirectly through the manifestation of electromagnetic phenomenon, when intercepted and converted, becomes useful. In nonlinear systems, interaction of magnetic waves amplify (conjugate) energy, providing greater output than input. In simple form, in the piano where three strings are struck by the hammer, the centre one is impacted and resonance activates the side strings. Resonance between the three strings provides a sound level greater than the input energy. Sound is part of the electromagnetic spectrum and is subject to all that is applicable to it.

"Useful Energy" is defined as "that which is other than Ambient". "Electric Potential" relates to mass and it's acceleration. Therefore, the Earth's Mass and Speed through space, gives it an enormous electrical potential. Humans are like the bird sitting unaware on a high voltage line. in nature, turbulence upsets ambient and we see electrical displays. Tampering with ambient, allows humans to convert magnetic waves into useful electricity.

Putting this in focus, requires a look at the Earth in general. During each of the 1,440 minutes of each day, more than 4,000 displays of lightning occur. Each display yields more than 10,000,000 volts at more than 200,000 amperes in equivalent electromagnetic flux. This is more than 57,600,000,000,000 volts and 1,152,000,000,000 amperes of electromagnetic flux during each 24 hour period. This has been going on for more than 4 billion years. The USPTO insist that the Earth's electrical field is insignificant and useless, and that converting this energy violates the laws of nature. At the same time, they issue patents in which, electromagnetic flux coming in from the Sun is converted by solar cells into DC energy. Aeromagnetic flux (in gammas) Maps World-Wide, includes those provided by the US Department of Interior-Geological Survey, and these show clearly that there is present, a spread of 1,900 gamma above Ambient, from reading instruments flown 1,000 feet above the (surface) source. Coulomb's Law requires the squaring of the distance of the remote reading, multiplied by the recorded reading. Therefore, that reading of 1,900 gamma has a corrected value of 1,900 x 1,000 x 1,000 = 1,900,000,000 gamma.

There is a tendency to confuse "gamma ray" with "gamma". "Gamma" is ordinary, everyday magnetic flux, while "gamma ray" is high-impact energy and not flux. One gamma of magnetic flux is equal to that of 100 volts RMS. To see this, take a Plasma Globe emitting 40,000 volts. When properly used, a gamma meter placed nearby, will read 400 gammas. The 1,900,000,000 gamma just mentioned, is the magnetic ambient equivalent of

190,000,000 volts of electricity. This is on a "Solar Quiet" day. On "Solar Active" days it may exceed five times that amount. The Establishment's idea that the Earth's electrical field is insignificant, goes the way of their other great ideas.

There are two kinds of electricity: "potential" and "useful". All electricity is "potential" until it is converted. The resonant-fluxing of electrons, activates the electrical potential which is present everywhere. The Intensity/CPS of the resonant-frequency-flux rate, sets the available energy. This must then be converted into the required physical dimensions of the equipment being used. For example, energy arriving from the Sun is magnetic flux, which solar cells convert to DC electricity, which is then converted further to suit the equipment being powered by it. Only the magnetic flux moves from point "A" (the Sun) to point "B" (the Earth). All electrical power systems work in exactly the same way. Movement of Coils and Magnets at point "A" (the generator) fluxes electrons, which in turn, excite electrons at point "B" (your house). **None of the electrons at point "A" are ever transmitted to point "B"**. In both cases, the electrons remain forever intact and available for further fluxing. This is not allowed by Newtonian Physics (electrodynamics and the laws of conservation). Clearly, these laws are all screwed up and inadequate.

In modern physics, USPTO style, all of the above cannot exist because it opens a door to overunity. The good news is that the PTO has already issued hundreds of Patents related to Light Amplification, all of which are overunity. The Dynode used to adjust the self-powered shutter in your camera, receives magnetic flux from light which dislodges electrons from the cathode, reflecting electrons through the dynode bridge to the anode, resulting in billions of more electrons out than in. There are currently, 297 direct patents issued for this system, and thousands of peripheral patents, all of which support overunity. More than a thousand other Patents which have been issued, can be seen by the discerning eye to be overunity devices. What does this indicate about Intellectual Honesty?

Any coil system, when fluxed, causes electrons to spin and produce useful energy, once it is converted to the style required by its use. Now that we have described the method which is required, let us now see how this concerns us.

The entire System already exists and all that we need to do is to hook it up in a way which is useful to our required manner of use. Let us examine this backwards and start with a conventional output transformer. Consider one which has the required voltage and current handling characteristics and which acts as an isolation transformer. Only the magnetic flux passes from the input winding to the output winding. No electrons pass through from the input side to the output side. Therefore, we only need to flux the output side of the transformer to have an electrical output. Bad design by the establishment, allowing hysteresis of the metal plates, limits the load which can be driven. Up to this point, only potential is a consideration. Heat (which is energy loss) limits the output amperage. Correctly designed composite cores run cool, not hot.

A power correction factor system, being a capacitor bank, maintains an even flow of flux. These same capacitors, when used with a coil system (a transformer) become a frequency-timing system. Therefore, the inductance of the input side of the transformer, when combined with the capacitor bank, provides the required fluxing to produce the required electrical energy (cycles per second).

With the downstream system in place, all that is needed now is a potential system. Any flux system will be suitable. Any amplification over-unity output type is desirable. The input system is point "A" and the output system is point "B". Any input system where a lesser amount of electrons disturbs a greater amount of electrons - producing an output which is greater than the input - is desirable.

At this point, it is necessary to present updated information about electrons and the laws of physics. A large part of this, originates from me (Don Smith) and so is likely to upset people who are rigidly set in the thought patterns of conventional science.

Non - Ionic Electrons

As a source of electrical energy, non-ionic electrons doublets exist in immense quantities throughout the universe. Their origin is from the emanation of Solar Plasma. When ambient electrons are disturbed by being spun or pushed apart, they yield both magnetic and electrical energy. The rate of disturbance (cycling) determines the energy level achieved. Practical methods of disturbing them include, moving coils past magnets or vice versa. A better way is the pulsing (resonant induction) with magnetic fields and waves near coils.

In coil systems, magnetic and amperage are one package. This suggests that electrons in their natural non-ionic state, exist as doublets. When pushed apart by agitation, one spins right (yielding Volts-potential electricity) and the other spins left (yielding Amperage-magnetic energy), one being more negative than the other. This further suggests that when they reunite, we have (Volts x Amps = Watts) useful electrical energy. Until now, this idea has been totally absent from the knowledge base. The previous definition of Amperage is therefore flawed.

Electron Related Energy

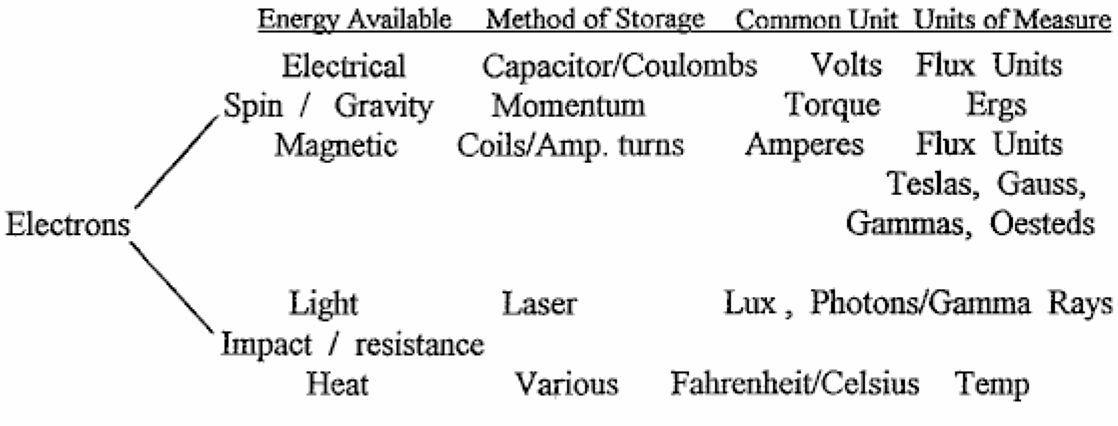

	Energy Available	Method of Storage	Common Unit	Units of Measure
Electrons	Electrical	Capacitor/Coulombs	Volts	Flux Units
	Spin / Gravity	Momentum	Torque	Ergs
	Magnetic	Coils/Amp. turns	Amperes	Flux Units Teslas, Gauss, Gammas, Oesteds
	Light	Laser	Lux ,	Photons/Gamma Rays
	Impact / resistance			
	Heat	Various	Fahrenheit/Celsius	Temp

Left hand spin of electrons results in Electrical Energy and right hand spin results in Magnetic Energy. Impacted electrons emit visible Light and heat.

Useful Circuits, Suggestions for Building an Operational Unit

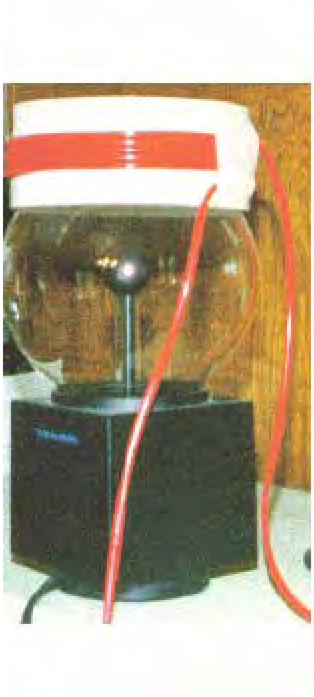

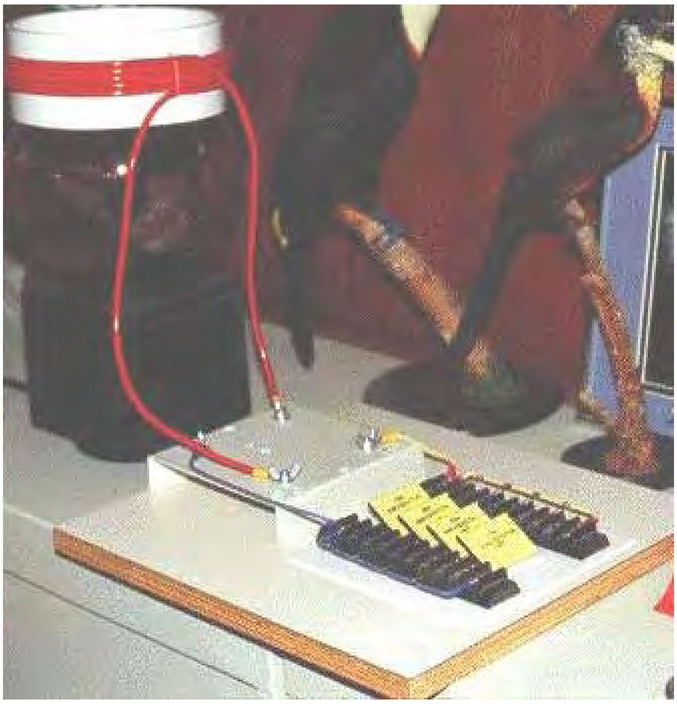

1. Substitute a Plasma Globe such as Radio Shack's "Illumna-Storm" for the source-resonant induction system. It will have about 400 milligauss of magnetic induction. One milligauss is equal to 100 volts worth of magnetic induction.

2. Construct a coil using a 5-inch to 7-inch (125 to 180 mm) diameter piece of PVC for the coil former.

3. Get about 30 feet (10 m) of Jumbo-Speaker Cable and separate the two strands. This can be done by sticking a carpet knife into a piece of cardboard or wood, and then pulling the cable carefully past the blade to separate the two insulated cores from each other. (PJK Note: "Jumbo-Speaker Cable" is a vague term as that cable comes in many varieties, with anything from a few, to over 500 strands in each core. As Don points out that the output power increases with each turn of wire, it is distinctly possible that each of these strands acts the same as individual insulated turns which have been connected in parallel, so a 500-strand cable may well be far more effective than a cable with just a few strands).

4. Wind the coil with 10 to 15 turns of wire and leave about 3 feet (1 m) of cable spare at each end of the coil. Use a glue gun to hold the start and finish of the coil.

5. This will become the "L - 2" coil shown in the Circuits page.

6. When sitting on top of the Plasma Globe (like a crown) you have a first-class resonant air-core coil system.

7. Now, substitute two or more capacitors (rated at 5,000 volts or more) for the capacitor bank shown on the Circuits page. I use more than two 34 microfarad capacitors.

8. Finish out the circuit as shown. You are now in business !

9. Voltage - Amperage limiting resistors are required across the output side of the Load transformer. These are used to adjust the output level and the desired cycles per second.

Useful Circuits from Nikola Tesla

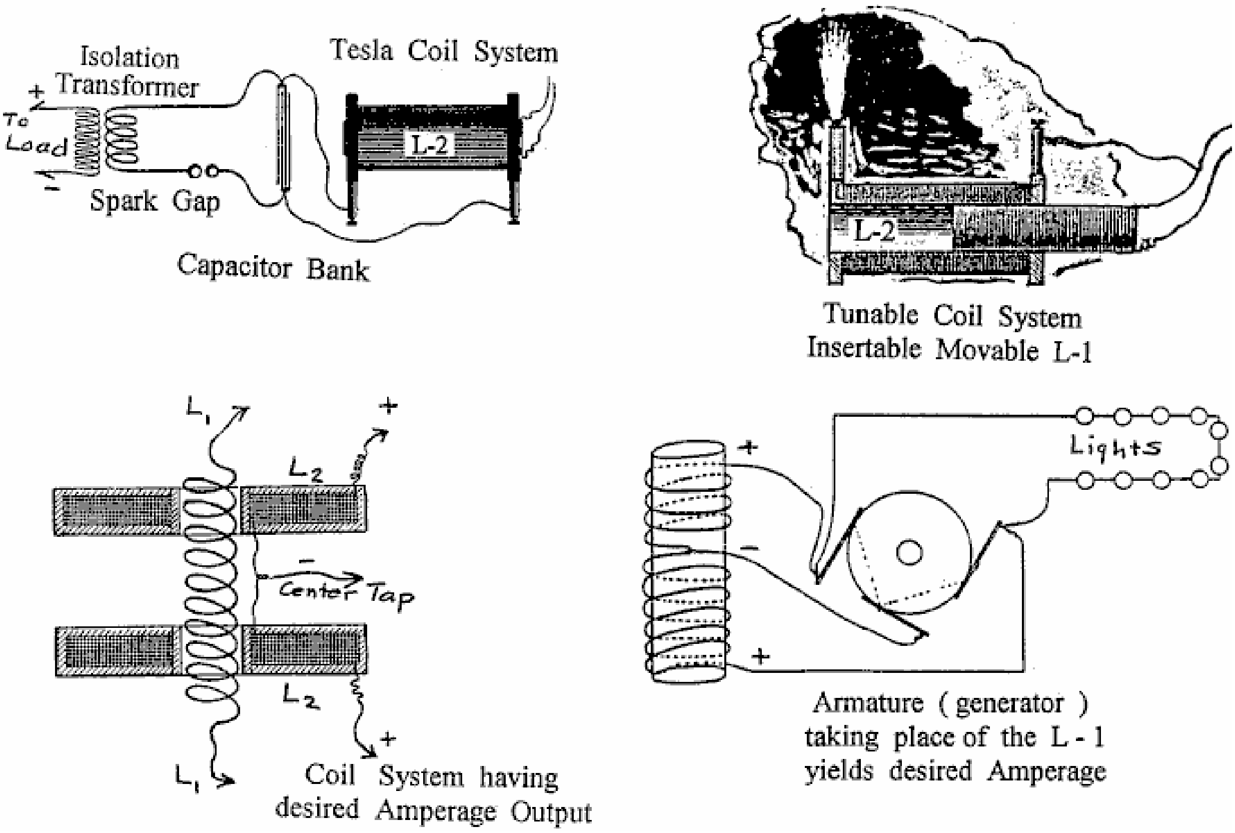
Isolation Transformer
+
To Load
-
Spark Gap
Tesla Coil System
L-2
Capacitor Bank
L-2
Tunable Coil System
Insertable Movable L-1
L_1
+
L_2
-
Center Tap
L_2
+
L_1
Coil System having desired Amperage Output
Lights
+
-
+
Armature (generator) taking place of the L - 1 yields desired Amperage

Don Smith's Suggestions: Get a copy of the "Handbook of Electronic Tables and Formulas", published by Sams, ISBN 0-672-22469-0, also an Inductance/Capacitance/Resistance meter is required. Chapter 1 of Don's pdf document has important time-constant (frequency) information and a set of reactance charts in nomograph style (*"nomograph": a graph, usually containing three parallel scales graduated for different variables so that when a straight line connects values of any two, the related value may be read directly from the third at the point intersected by the line*) which makes working, and approximating of the three variables (capacitance, inductance and resistance) much easier. If two of the variables are known, then the third one can be read from the nomograph.

For example, if the input side of the isolation transformer needs to operate at 60 Hz, that is 60 positive cycles and 60 negative cycles, being a total of 120 cycles. Read off the inductance in Henries by using the Inductance meter attached to the input side of the isolation transformer. Plot this value on the (nomographic) reactance chart. Plot the needed 120 Hz on the chart and connect these two points with a straight line. Where this line crosses the Farads line and the Ohms line, gives us two values. Choose one (resistor) and insert it between the two leads of the transformer input winding.

The Power Correction Factor Capacitor (or bank of more than one capacitor) now needs adjusting. The following formula is helpful in finding this missing information. The capacitance is known, as is the desired potential to pulse the output transformer. One Farad of capacitance is one volt for one second (one Coulomb). Therefore, if we want to keep the bucket full with a certain amount, how many dippers full are needed? If the bucket needs 120 volts, then how many coulombs are required?

$$\frac{\text{Desired Voltage}}{\text{Capacitance in Microfarads}} = \text{Required frequency in Hz}$$

Now, go to the nomograph mentioned above, and find the required resistor jumper to place between the poles of the Correction Factor Capacitor.

An earth grounding is desirable, acting as both a voltage-limiter and a transient spike control. Two separate earths are necessary, one at the Power Factor Capacitor and one at the input side of the isolation transformer. Off-the-shelf surge arrestors / spark gaps and varistors having the desired voltage/potential and amperage control are commonly available. Siemens, Citel America and others, make a full range of surge arrestors, etc. Varistors look like coin-sized flat capacitors. Any of these voltage limiters are marked as "V - 1" in the following text.

It should be obvious that several separate closed circuits are present in the suggested configuration: The power input source, the high-voltage module, a power factor capacitor bank combined with the input side of the isolation transformer. Lastly, the output side of the isolation transformer and its load. None of the electrons active at the power source (battery) are passed through the system for use downstream. At any point, if the magnetic flux rate should happen to vary, then the number of active electrons also varies. Therefore, controlling the flux rate controls the electron (potential) activity. Electrons active at point "A" are not the same electrons which are active at point "B", or those at point "C", and so on. If the magnetic flux rate (frequency Hz) varies, then a different number of electrons will be disturbed. This does not violate any Natural Law and it does produce more output energy than the input energy, should that be desirable.

A convenient high-voltage module is a 12 volt DC neon tube transformer. The Power Factor Correction Capacitors should be as many microfarads as possible as this allows a lower operating frequency. The 12-volt neon tube transformer oscillates at about 30,000 Hz. At the Power Correction Factor Capacitor bank we lower the frequency to match the input side of the isolation transformer.

Other convenient high-voltage sources are car ignition coils, television flyback transformers, laser printer modules, and various other devices. Always lower the frequency at the Power Factor Correction Capacitor and correct, if needed, at the input side of the isolation transformer. The isolation transformer comes alive when pulsed. Amperage becomes a part of the consideration only at the isolation transformer. Faulty design, resulting in hysteresis, creates heat which self-destructs the transformer if it is overloaded. Transformers which have a composite core instead of the more common cores made from many layers of thin sheets of soft iron, run cool and can tolerate much higher amperage.

RESONATE ELECTROMAGNETIC POWER SYSTEM

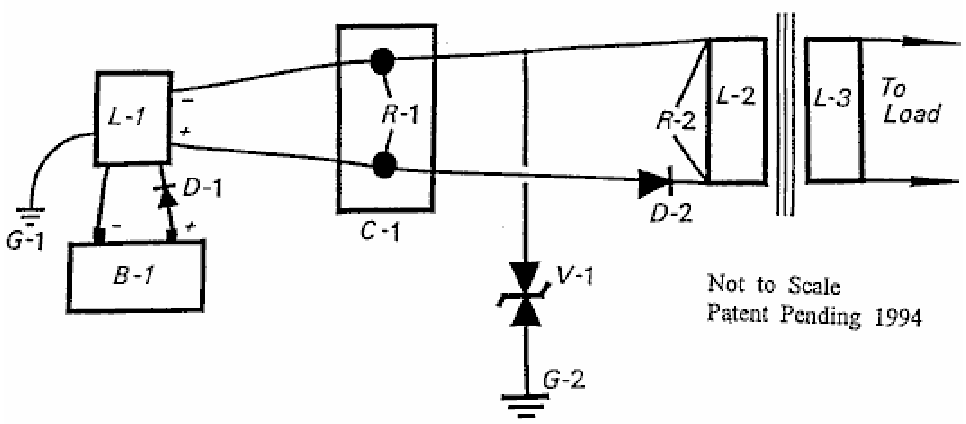

Not to Scale
Patent Pending 1994

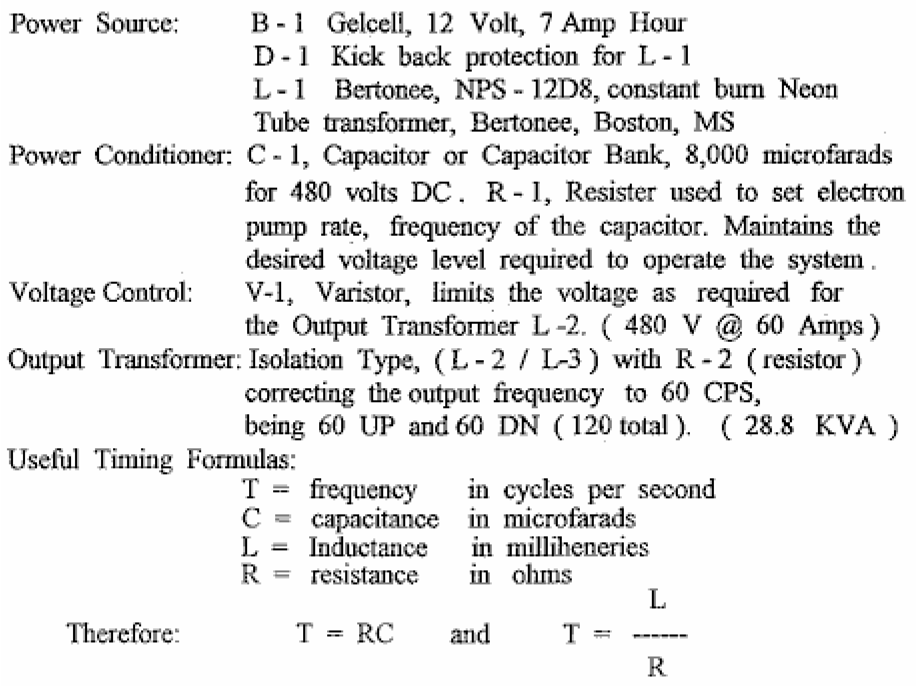

Power Source: B - 1 Gelcell, 12 Volt, 7 Amp Hour
D - 1 Kick back protection for L - 1
L - 1 Bertonee, NPS - 12D8, constant burn Neon Tube transformer, Bertonee, Boston, MS

Power Conditioner: C - 1, Capacitor or Capacitor Bank, 8,000 microfarads for 480 volts DC . R - 1, Resister used to set electron pump rate, frequency of the capacitor. Maintains the desired voltage level required to operate the system .

Voltage Control: V-1, Varistor, limits the voltage as required for the Output Transformer L -2. (480 V @ 60 Amps)

Output Transformer: Isolation Type, (L - 2 / L-3) with R - 2 (resistor) correcting the output frequency to 60 CPS, being 60 UP and 60 DN (120 total). (28.8 KVA)

Useful Timing Formulas:

T = frequency in cycles per second
C = capacitance in microfarads
L = Inductance in milliheneries
R = resistance in ohms

Therefore: $T = RC$ and $T = \frac{L}{R}$

The information shown above, relates to the small Suitcase Model demonstrated at the 1996 Tesla Convention, presented as Don Smiths' Workshop. This unit was a very primitive version and newer versions have atomic batteries and power output ranges of Gigawatts. The battery requirement is low level and is no more harmful than the radium on the dial of a clock. Commercial units of Boulder Dam size are currently being installed at several major locations throughout the world. For reasons of Don's personal security and contract obligations, the information which he has shared here, is incomplete.

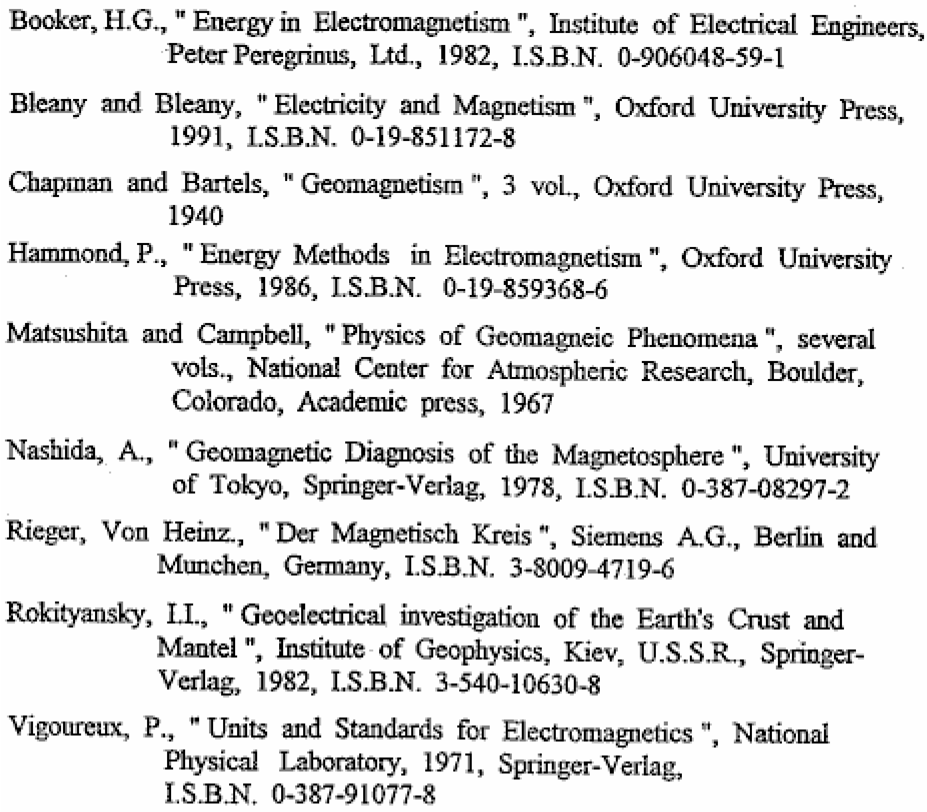

Booker, H.G., " Energy in Electromagnetism ", Institute of Electrical Engineers, Peter Peregrinus, Ltd., 1982, I.S.B.N. 0-906048-59-1

Bleany and Bleany, " Electricity and Magnetism ", Oxford University Press, 1991, I.S.B.N. 0-19-851172-8

Chapman and Bartels, " Geomagnetism ", 3 vol., Oxford University Press, 1940

Hammond, P., " Energy Methods in Electromagnetism ", Oxford University Press, 1986, I.S.B.N. 0-19-859368-6

Matsushita and Campbell, " Physics of Geomagneic Phenomena ", several vols., National Center for Atmospheric Research, Boulder, Colorado, Academic press, 1967

Nashida, A., " Geomagnetic Diagnosis of the Magnetosphere ", University of Tokyo, Springer-Verlag, 1978, I.S.B.N. 0-387-08297-2

Rieger, Von Heinz., " Der Magnetisch Kreis ", Siemens A.G., Berlin and Munchen, Germany, I.S.B.N. 3-8009-4719-6

Rokityansky, I.I., " Geoelectrical investigation of the Earth's Crust and Mantel ", Institute of Geophysics, Kiev, U.S.S.R., Springer-Verlag, 1982, I.S.B.N. 3-540-10630-8

Vigoureux, P., " Units and Standards for Electromagnetics ", National Physical Laboratory, 1971, Springer-Verlag, I.S.B.N. 0-387-91077-8

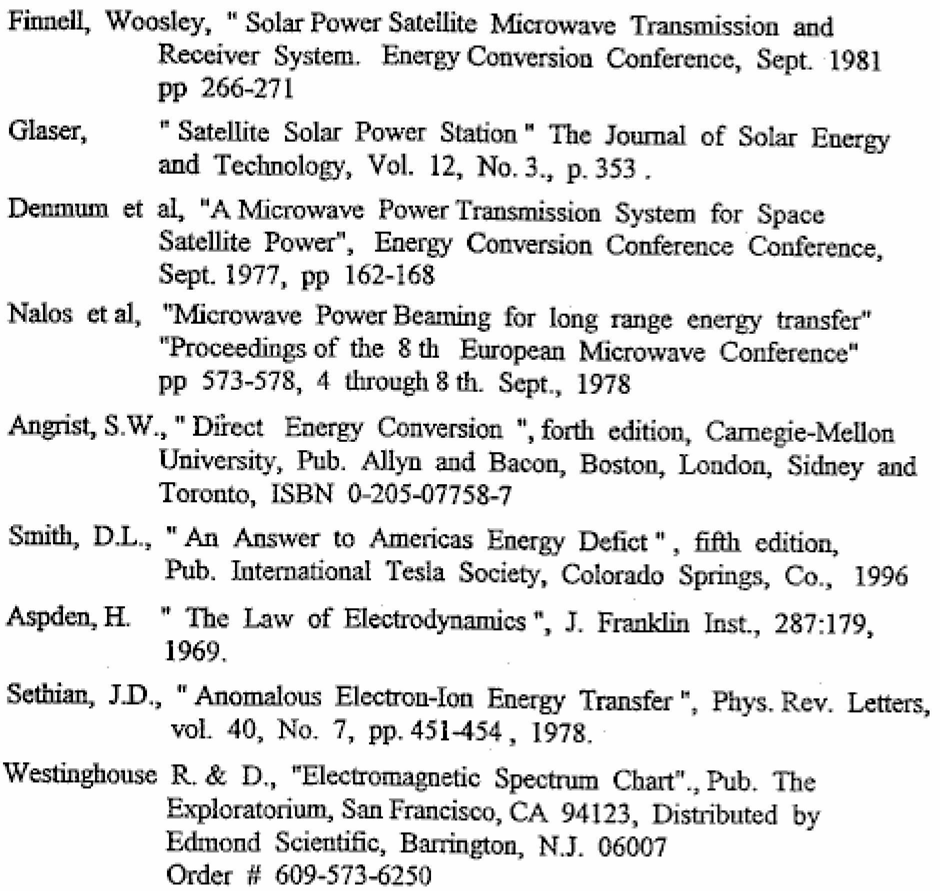

Finnell, Woosley, " Solar Power Satellite Microwave Transmission and Receiver System. Energy Conversion Conference, Sept. 1981 pp 266-271

Glaser, " Satellite Solar Power Station " The Journal of Solar Energy and Technology, Vol. 12, No. 3., p. 353 .

Denmum et al, "A Microwave Power Transmission System for Space Satellite Power", Energy Conversion Conference Conference, Sept. 1977, pp 162-168

Nalos et al, "Microwave Power Beaming for long range energy transfer" "Proceedings of the 8 th European Microwave Conference" pp 573-578, 4 through 8 th. Sept., 1978

Angrist, S.W., " Direct Energy Conversion ", forth edition, Carnegie-Mellon University, Pub. Allyn and Bacon, Boston, London, Sidney and Toronto, ISBN 0-205-07758-7

Smith, D.L., " An Answer to Americas Energy Defict " , fifth edition, Pub. International Tesla Society, Colorado Springs, Co., 1996

Aspden, H. " The Law of Electrodynamics ", J. Franklin Inst., 287:179, 1969.

Sethian, J.D., " Anomalous Electron-Ion Energy Transfer ", Phys. Rev. Letters, vol. 40, No. 7, pp. 451-454 , 1978.

Westinghouse R. & D., "Electromagnetic Spectrum Chart"., Pub. The Exploratorium, San Francisco, CA 94123, Distributed by Edmond Scientific, Barrington, N.J. 06007 Order # 609-573-6250

I am most definitely not an expert in this area. However, it is probably worth mentioning some of the main points which Don Smith appears to be making. There are some very important points being made here, and grasping these may make a considerable difference to our ability to tap into the excess energy available in our local environment. There are four points worth mentioning:

1. Voltage
2. Frequency
3. Magnetic / Electric relationship
4. Resonance

1. Voltage. We tend to view things with an 'intuitive' view, generally based on fairly simple concepts. For example, we automatically think that it is more difficult to pick up a heavy object than to pick up a light one. How much more difficult? Well, if it is twice as heavy, it would probably be about twice as much effort to pick it up. This view has developed from our experience of things which we have done in the past, rather than on any mathematical calculation or formula.

Well, how about pulsing an electronic system with a voltage? How would the output power of a system be affected by increasing the voltage? Our initial 'off-the cuff' reaction might be that the power output might be increased a bit, but then hold on… we've just remembered that Watts = Volts x Amps, so if you double the voltage, then you would double the power in watts. So we might settle for the notion that if we doubled the voltage then we could double the output power. If we thought that, then we would be wrong.

Don Smith points out that as capacitors and coils store energy, if they are involved in the circuit, then the output power is proportional to the **square** of the voltage used. Double the voltage, and the output power is four times greater. Use three times the voltage and the output power is nine times greater. Use ten times the voltage and the output power is one hundred times greater !

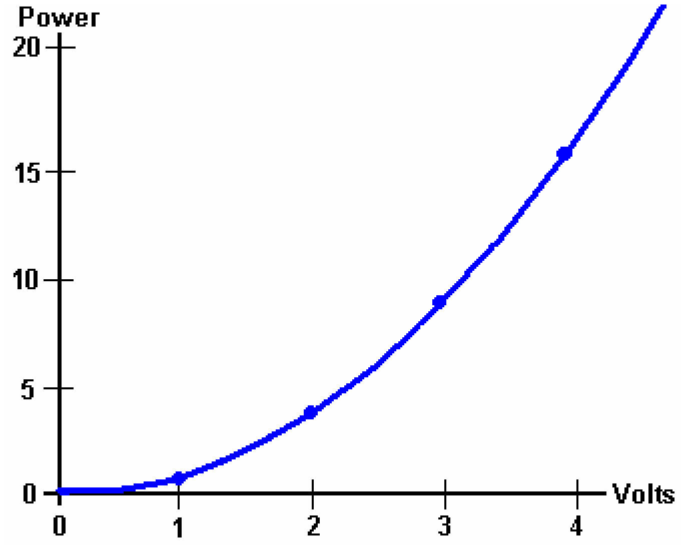

Don says that the energy stored, multiplied by the cycles per second, is the energy being pumped by the system. Capacitors and inductors (coils) temporarily store electrons, and their performance is given by:

Capacitor formula: **$W = 0.5 \times C \times V^2 \times Hz$** where:

W is the energy in Joules (Joules = Volts x Amps x seconds)
C is the capacitance in Farads
V is the voltage
Hz is the cycles per second

Inductor formula: **$W = 0.5 \times L \times A^2 \times Hz$** where:

W is the energy in Joules
L is the inductance in henrys
A is the current in amps
Hz is the frequency in cycles per second

You will notice that where inductors (coils) are involved, then the output power goes up with the square of the current. Double the voltage **and** double the current gives four times the power output due to the increased voltage and that increased output is increased by a further four times due to the increased current, giving sixteen times the output power.

2. Frequency. You will notice from the formulas above, that the output power is directly proportional to the frequency "Hz". The frequency is the number of cycles per second (or pulses per second) applied to the circuit. This is something which is not intuitive for most people. If you double the rate of pulsing, then you double the power output. When this sinks in, you suddenly see why Nikola Tesla tended to use millions of volts and millions of pulses per second.

However, Don Smith states that when a circuit is at it's point of resonance, resistance in the circuit drops to zero and the circuit becomes effectively, a superconductor. The energy for such a system which is in resonance is:

Resonant circuit: **$W = 0.5 \times C \times V^2 \times (Hz)^2$** where:

W is the energy in Joules
C is the capacitance in Farads
V is the voltage
Hz is the cycles per second

If this is correct, then raising the frequency in a resonating circuit has a massive effect on the power output of the device. The question then arises: why is the mains power in Europe just fifty cycles per second and in America just sixty cycles per second? If power goes up with frequency, then why not feed households at a million cycles per second? One major reason is that it is not easy to make electric motors which can be driven with power delivered at that frequency, so a more suitable frequency is chosen in order to suit the motors in vacuum cleaners, washing machines and other household equipment.

However, if we want to extract energy from the environment, then we should go for high voltage and high frequency. Then, when high power has been extracted, if we want a low frequency suited to electric motors, we can pulse the already captured power at that low frequency.

It might be speculated that if a device is being driven with sharp pulses which have a very sharply rising leading edge, that the effective frequency of the pulsing is actually determined by the speed of that rising edge, rather than the rate at which the pulses are actually generated. For example, if pulses are being generated at, say, 50 kHz but the pulses have a leading edge which would be suited to a 200 kHz pulse train, then the device might well see the signal as a 200 kHz signal with a 25% Mark/Space ratio, the very suddenness of the applied voltage having a magnetic shocking effect equivalent to a 200 kHz pulse train.

3. Magnetic / Electric relationship. Don states that the reason why our present power systems are so inefficient is because we concentrate on the electric component of electromagnetism. These systems are always COP<1 as electricity is the 'losses' of electromagnetic power. Instead, if you concentrate on the magnetic component, then there is no limit on the electric power which can be extracted from that magnetic component. Contrary to what you might expect, if you install a pick-up system which extracts electrical energy from the magnetic component, you can install any number of other identical pick-ups, each of which extract the same amount of electrical energy from the magnetic input, **without** loading the magnetic wave in any way. Unlimited electrical output for the 'cost' of creating a single magnetic effect.

The magnetic effect which we want to create is a ripple in the zero-point energy field, and ideally, we want to create that effect while using very little power. Creating a dipole with a battery which has a Plus and a Minus terminal or a magnet which has North and South poles, is an easy way to do create an electromagnetic imbalance in the local environment. Pulsing a coil is probably an even better way as the magnetic field reverses rapidly if it is an air-core coil, such as a Tesla Coil. Using a ferromagnetic core to the coil can create a problem as iron can't reverse it's magnetic alignment very rapidly, and ideally, you want pulsing which is at least a thousand times faster than iron can handle.

Don draws attention to the "Transmitter / Receiver" educational kit "Resonant Circuits #10-416" supplied by The Science Source, Maine. This kit demonstrates the generation of resonant energy and it's collection with a receiver circuit. However, if several receiver circuits are used, then the energy collected is increased several times without any increase in the transmitted energy. This is similar to a radio transmitter where hundreds of thousands of radio receivers can receive the transmitted signal without loading the transmitter in any way.

This immediately makes the Hubbard device spring to mind. Hubbard has a central "electromagnetic transmitter" surrounded by a ring of "receivers" closely coupled magnetically to the transmitter, each of which will receive a copy of the energy sent by the transmitter:

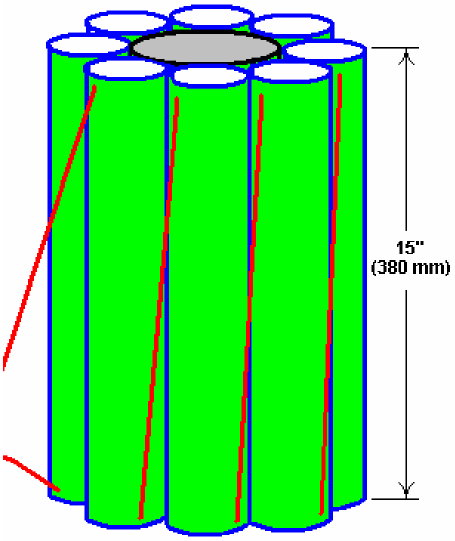

Don points to an even more clearly demonstrated occurrence of this effect in the Tesla Coil. In a typical Tesla Coil, the primary coil is much larger diameter than the inner secondary coil:

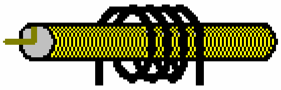

If, for example, 8,000 volts is applied to the primary coil which has four turns, then each turn would have 2,000 volts of potential. Each turn of the primary coil transfers electromagnetic flux to every single turn of the secondary winding, and the secondary coil has a very large number of turns. Massively more power is produced in the secondary coil than was used to energise the primary coil. A common mistake is to believe that a Tesla Coil can't produce serious amperage. If the primary coil is positioned in the middle of the secondary coil as shown, then the amperage generated will be as large as the voltage generated. A low power input to the primary coil can produce kilowatts of usable electrical power as described in chapter 5.

4. Resonance. An important factor in circuits aimed at tapping external energy is resonance. It can be hard to see where this comes in when it is an electronic circuit which is being considered. However, everything has it's own resonant frequency, whether it is a coil or any other electronic component. When components are connected together to form a circuit, the circuit has an overall resonant frequency. As a simple example, consider a swing:

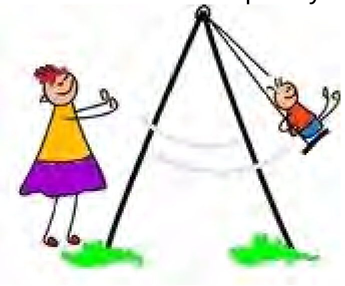

If the swing is pushed before it reaches the highest point on the mother's side, then the push actually detracts from the swinging action. The time of one full swing is the resonant frequency of the swing, and that is determined by the length of the supporting ropes holding the seat and not the weight of the child nor the power with which the child is pushed. Provided that the timing is exactly right, a very small push can get a swing moving in a substantial arc. The key factor is, matching the pulses applied to the swing, to the resonant frequency of the swing. Get it right and a large movement is produced. Get it wrong, and the swing doesn't get going at all (at which point, critics would say "see, see …swings just don't work - this proves it !!").

Establishing the exact pulsing rate needed for a resonant circuit is not particularly easy, because the circuit contains coils (which have inductance, capacitance and resistance), capacitors (which have capacitance and a small amount of resistance) and resistors and wires, both of which have resistance and some capacitance. These kinds of circuit are called "LRC" circuits because "L" is the symbol used for inductance, "R" is the symbol used for resistance and "C" is the symbol used for capacitance.

Don Smith provides instructions for winding and using the type of air-core coils needed for a Tesla Coil. He says:

1. Decide a frequency and bear in mind, the economy of the size of construction selected. The factors are:

(a) Use radio frequency (above 20 kHz).
(b) Use natural frequency, i.e. match the coil wire length to the frequency - coils have both capacitance and inductance.
(c) Make the wire length either one quarter, one half of the full wavelength.
(d) Calculate the wire length in feet as follows:
    If using one quarter wavelength, then divide 247 by the frequency in MHz.
    If using one half wavelength, then divide 494 by the frequency in MHz.
    If using the full wavelength, then divide 998 by the frequency in MHz.
  For wire lengths in metres:
    If using one quarter wavelength, then divide 75.29 by the frequency in MHz.
    If using one half wavelength, then divide 150.57 by the frequency in MHz.
    If using the full wavelength, then divide 304.19 by the frequency in MHz.

2. Choose the number of turns to be used in the coil when winding it using the wire length just calculated. The number of turns will be governed by the diameter of the tube on which the coil is to be wound. Remember that the ratio of the number of turns in the "L - 1" and "L - 2" coils, controls the overall output voltage. For example, if the voltage applied the large outer coil "L - 1" is 2,400 volts and L - 1 has ten turns, then each turn of L - 1 will have 240 volts dropped across it. This 240 volts of magnetic induction transfers 240 volts of electricity to every turn of wire in the inner "L - 2" coil. If the diameter of L - 2 is small enough to have 100 turns, then the voltage produced will be 24,000 volts. If the diameter of the L - 2 former allows 500 turns, then the output voltage will be 120,000 volts.

3. Choose the length and diameter of the coils. The larger the diameter of the coil, the fewer turns can be made with the wire length and so the coil length will be less, and the output voltage will be lower.

4. For example, if 24.7 MHz is the desired output frequency, then the length of wire, in feet, would be 247 divided by 24.7 which is 10 feet of wire (3,048 mm). The coil may be wound on a standard size of PVC pipe or alternatively, it can be purchased from a supplier - typically, an amateur radio supply store.

If the voltage on each turn of L - 1 is arranged to be 24 volts and the desired output voltage 640 volts, then there needs to be 640 / 24 = 26.66 turns on L - 2, wound with the 10 feet of wire already calculated.

Note: At this point, Don's calculations go adrift and he suggests winding 30 turns on a 2-inch former. If you do that, then it will take about 16 feet of wire and the resonant point at 10-feet will be at about 19 turns, giving an output voltage of 458 volts instead of the required 640 volts, unless the number of turns on L - 1 is reduced to give more than 24 volts per turn. However, the actual required diameter of the coil former (plus one diameter of the wire) is 10 x 12 / (26.67 x 3.14159) = 1.43 inches. You can make this size of former up quite easily if you want to stay with ten turns on the L - 1 coil.

5. Connect to the start of the coil. To determine the exact resonant point on the coil, a measurement is made. Off-the-shelf multimeters are not responsive to high-frequency signals so a cheap neon is used instead. Holding one wire of the neon in one hand and running the other neon wire along the outside of the L - 2 winding, the point of brightest light is located. Then the neon is moved along that turn to find the brightest point along that turn, and when it is located, a connection is made to the winding at that exact point. L - 2 is now a resonant winding. It is possible to increase the ("Q") effectiveness of the coil by spreading the turns out a bit instead of positioning them so that each turn touches both of the adjacent turns.

6. The input power has been suggested as 2,400 volts. This can be constructed from a Jacob's ladder arrangement or any step-up voltage system. An off-the-shelf module as used with lasers is another option.

7. Construction of the L - 1 input coil has been suggested as having 10 turns. The length of the wire in this coil is not critical. If a 2-inch diameter PVC pipe was used for the L - 2 coil, then the next larger size of PVC pipe can be used for the L - 1 coil former. Cut a 10-turn length of the pipe (probably a 3-inch diameter pipe). The pipe length will depend on the diameter of the insulated wire used to make the winding. Use a good quality multimeter or a specialised LCR meter to measure the capacitance (in Farads) and the inductance (in henrys) of the L - 2 coil. Now, put a capacitor for matching L - 1 to L - 2 across the voltage input of L - 1, and a spark gap connected **in parallel** is required for the return voltage from L - 1. A trimmer capacitor for L - 1 is desirable.

8. The performance of L - 2 can be further enhanced by attaching an earth connection to the base of the coil. The maximum output voltage will be between the ends of coil L - 2 and lesser voltages can be taken off intermediate points along the coil if that is desirable.

Don provides quite an amount of information on one of his devices shown here:

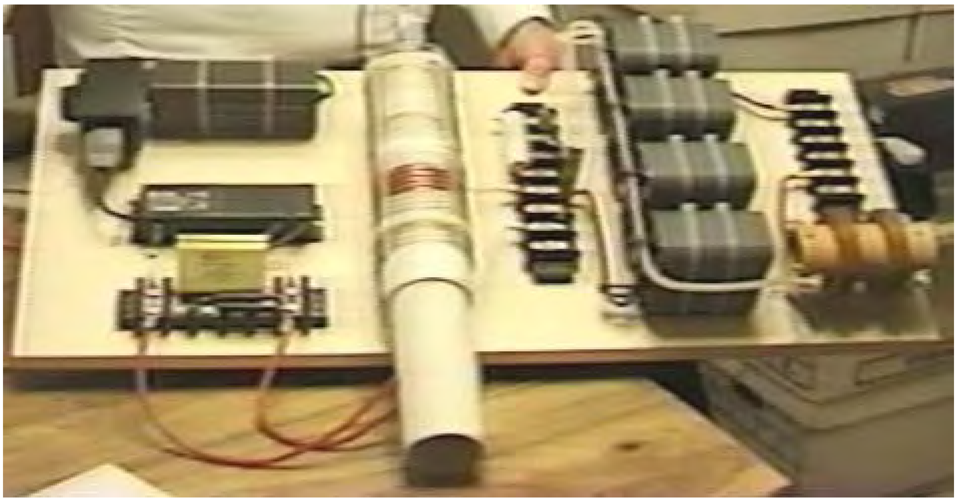

Without his description of the device, it would be difficult to understand it's construction and method of operation. As I understand it, the circuit of what is mounted on this board is as shown here:

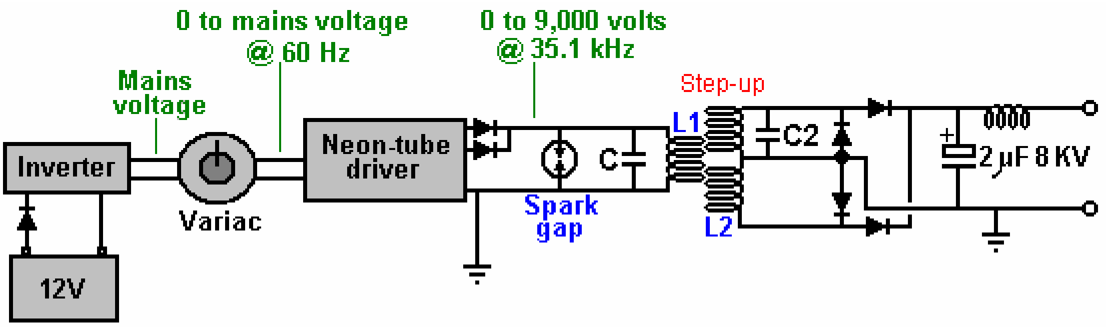
0 to mains voltage
@ 60 Hz
0 to 9,000 volts
@ 35.1 kHz
Mains
voltage
Step-up
Inverter
Variac
Neon-tube
driver
Spark
gap
C
L1
C2
L2
2 μF 8 KV
12V

This arrangement has bothered some readers recently as they feel that the spark gap should be in series with the L1 coil, like this:

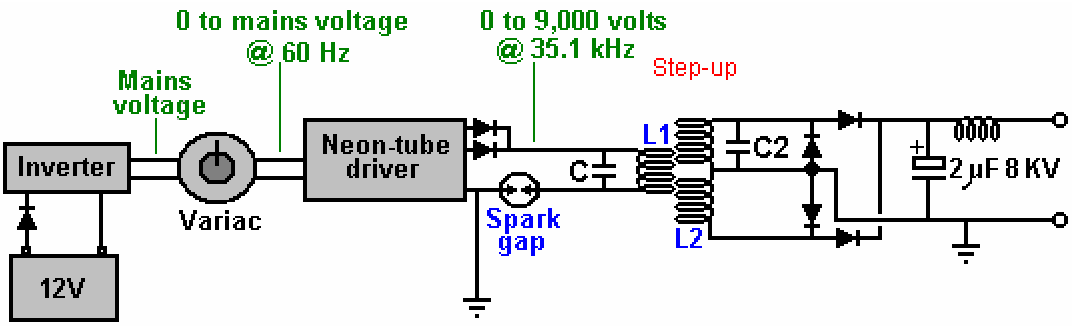
0 to mains voltage
@ 60 Hz
0 to 9,000 volts
@ 35.1 kHz
Step-up
Mains
voltage
Inverter
Variac
Neon-tube
driver
C
Spark
gap
L1
C2
L2
2 μF 8 KV
12V

This is understandable, as there is always a tendency to think of the spark gap as being a device which is there to protect against excessive voltages rather than seeing it as an active component of the circuit, a component which is in continuous use. In 1925, Hermann Plauston was granted a patent for a whole series of methods for converting the high voltage produced by a tall aerial system into useable, standard electricity. Hermann starts off by explaining how high voltage can be converted into a convenient form and he uses a Wimshurst static electricity generator as an example of a constant source of high voltage. The output from a rectified Tesla Coil, a Wimshurst machine and a tall aerial are very much alike, and so Hermann's comments are very relevant here. He shows it like this:

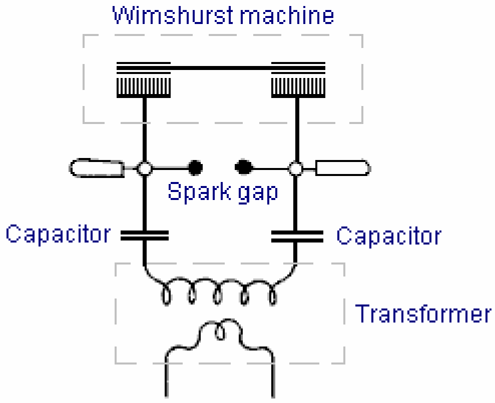
Wimshurst machine
Spark gap
Capacitor
Capacitor
Transformer

Here, the output of the Wimshurst machine is stored in two high-voltage capacitors (Leyden jars) causing a very high voltage to be created across those capacitors. When the voltage is high enough, a spark jumps across the spark gap, causing a massive surge of current through the primary winding of the transformer, which in his case is a step-down transformer as he is aimed at getting a lower output voltage. Don's circuit is almost identical:

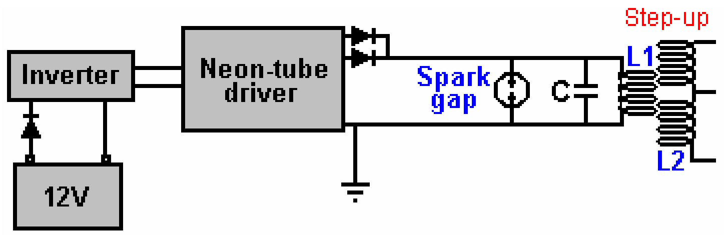
Step-up
Inverter
Neon-tube
driver
Spark
gap
C
L1
L2
12V

Here the high voltage comes from the battery/inverter/neon-tube driver/rectifiers, rather than from a mechanically driven Wimshurst machine. He has the same build up of voltage in a capacitor with a spark gap across the capacitor. The spark gap will fire when the capacitor voltage reaches its designed level. The only difference is in the positioning of the capacitor, which if it matched Hermann's arrangement exactly, would be like this:

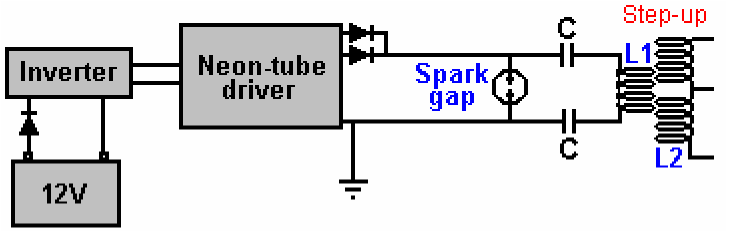

which would be a perfectly viable arrangement as far as I can see. You will remember that Tesla, who always speaks very highly of the energy released by the very sharp discharge produced by a spark, shows a high-voltage source feeding a capacitor with the energy passing through a spark gap to the primary winding of a transformer:

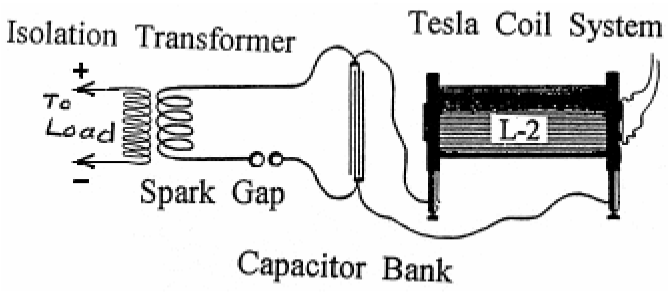

However, with Don's arrangement, it can be a little difficult to see why the capacitor is not short-circuited by the very low resistance of the few turns of thick wire forming the L1 coil. Well, it would do that if we were operating with DC, but we are most definitely not doing that as the output from the neon-tube driver circuit is pulsing 35,000 times per second. This causes the DC resistance of the L1 coil to be of almost no consequence and instead, the coil's "impedance" or "reactance" (effectively, it's AC resistance) is what counts. Actually, the capacitor and the L1 coil being connected across each other have a combined "reactance" or resistance to pulsing current at this frequency. This is where the nomograph diagram comes into play, and there is a much easier to understand version of it a few pages later on in this document. So, because of the high pulsing frequency, the L1 coil does not short-circuit the capacitor and if the pulsing frequency matches the resonant frequency of the L1 coil (or a harmonic of that frequency), then the L1 coil will actually have a very high resistance to current flow through it. This is how a crystal set radio receiver tunes in a particular radio station, broadcasting on it's own frequency.

Anyway, coming back to Don's device shown in the photograph above, the electrical drive is from a 12-volt battery which is not seen in the photograph. Interestingly, Don remarks that if the length of the wires connecting the battery to the inverter are exactly one quarter of the wave length of the frequency of the oscillating magnetic field generated by the circuit, then the current induced in the battery wires will recharge the battery continuously, even if the battery is supplying power to the circuit at the same time.

The battery supplies a small current through a protecting diode, to a standard off-the-shelf "true sine-wave" inverter. An inverter is a device which produces mains-voltage Alternating Current from a DC battery. As Don wants adjustable voltage, he feeds the output from the inverter into a variable transformer called a "Variac" although this is often made as part of the neon-driver circuit to allow the brightness of the neon tube to be adjusted by the user. This arrangement produces an AC output voltage which is adjustable from zero volts up to the full mains voltage (or a little higher, though Don does not want to use a higher voltage). The use of this kind of adjustment usually makes it essential for the inverter to be a true sine-wave type. As the power requirement of the neon-tube driver circuit is so low, the inverter should not cost very much.

The neon-tube (or "gas-discharge" tube) driver circuit is a standard off-the-shelf device used to drive neon tube displays for commercial establishments. The one used by Don contains an oscillator and a step-up transformer, which together produce an Alternating Current of 9,000 volts at a frequency of 35,100 Hz (sometimes written as 35.1 kHz). The term "Hz" stands for "cycles per second". Don lowers the 9,000 volts as he gets great power output at lower input voltages and the cost of the output capacitors is a significant factor. The particular neon-tube driver circuit which Don is using here, has two separate outputs, so Don connects them together and uses a blocking diode in each line to prevent either of them affecting the other one. Not easily seen in the photograph, the high-voltage output line has a very small, encapsulated, spark gap in it and the line is also earthed. This device is commonly used as a lightning strike protection component and in Don's circuit it lights continuously when the device is running. It looks like this:

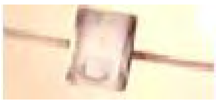

The output of the neon-tube driver circuit is used to drive the primary "L1" winding of a Tesla Coil style transformer. This looks ever so simple and straightforward, but there are some subtle details which need to have attention paid to them.

The operating frequency of 35.1 kHz is set and maintained by the neon-tube driver circuitry, and so, in theory, we do not have to do any direct tuning ourselves. However, we want the resonant frequency of the L1 coil and the capacitor across it to match the neon-driver circuit frequency. The frequency of the "L1" coil winding will induce exactly the same frequency in the "L2" secondary winding. However, we need to pay special attention to the ratio of the wire lengths of the two coil windings as we want these two windings to resonate together. A rule of thumb followed by most Tesla Coil builders is to have the same weight of copper in the L1 and L2 coils, which means that the wire of the L1 coil is usually much thicker than the wire of the L2 coil. If the L1 coil is to be one quarter of the length of the L2 coil, then we would expect the cross-sectional area of the L1 coil to be four times that of the wire of the L2 coil (as the area is proportional to the square of the radius, and the square of two is four)

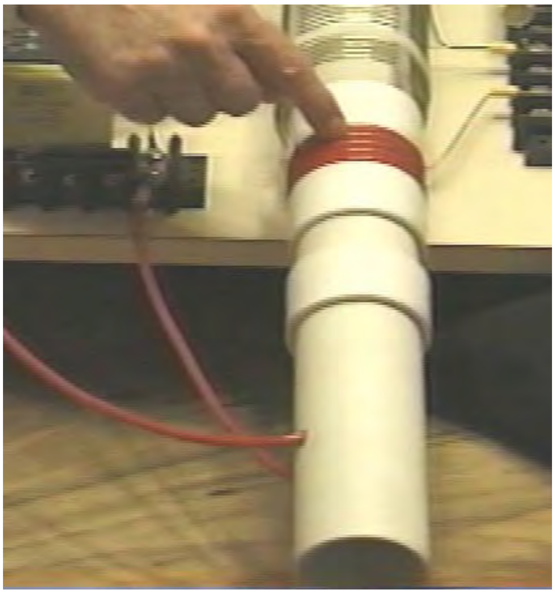

Don uses a plastic tube as the former for his "L1" primary coil winding. As you can see here, the wire is fed into the former, leaving sufficient clearance to allow the former to slide all the way into the outer coil. The wire is fed up inside the pipe and out through another hole to allow the coil turns to be made on the outside of the pipe. There appear to be five turns, but Don does not always go for a complete number of turns, so it might be 4.3 turns or some other value. The key point here is that the length of wire in the "L1" coil turns should be exactly one quarter of the length of wire in the "L2" coil turns.

The "L2" coil used here is a commercial 3-inch diameter unit from Barker & Williamson, constructed from uninsulated, solid, single-strand "tinned" copper wire. Don has taken this coil and unwound four turns at the centre of the coil in order to make a centre-tap. He then measured the exact length of wire in the remaining section and made the length of the "L1" coil turns to be exactly one quarter of that length. The wire used for the "L1" coil looks like Don's favourite "Jumbo Speaker Wire" which is a very flexible wire with a very large number of extremely fine uninsulated copper wires inside it.

You will notice that Don has placed a plastic collar on each side of the winding, matching the thickness of the wire, in order to create a secure sliding operation inside the outer "L2" coil, and the additional plastic collars positioned further along the pipe provide further support for the inner coil. This sliding action allows the primary coil "L1" to be positioned at any point along the length of the "L2" secondary coil, and that has a marked tuning effect on the operation of the system. The outer "L2" coil does not have any kind of tube support but instead, the coil shape is maintained by the stiffness of the solid wire plus four slotted strips. This style of construction produces the highest possible coil performance at radio frequencies. With a Tesla Coil, it is most unusual to have the L1 coil of smaller diameter than the L2 coil.

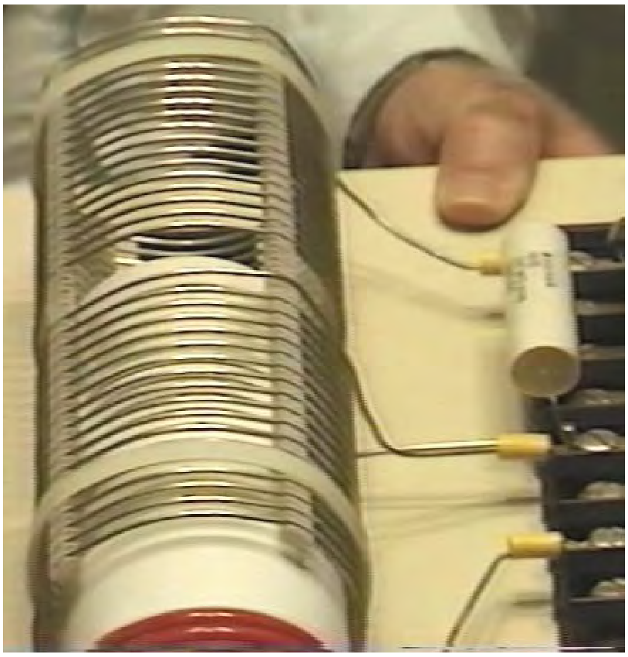

The "L2" coil has two separate sections, each of seventeen turns. One point to note is the turns are spaced apart using slotted strips to support the wires and maintain an accurate spacing between adjacent turns. It must be remembered that spacing coil turns apart like this alters the characteristics of the coil, increasing it's "capacitance" factor substantially. Every coil has resistance, inductance and capacitance, but the form of the coil construction has a major effect on the ratio of these three characteristics. The coil assembly is held in position on the base board by two off-white plastic cable ties. The nearer half of the coil is effectively connected across the further half as shown in the circuit diagram above.

One point which Don stresses, is that the length of the wire in the "L1" coil and the length of wire in the "L2" coil, must be an exact even division or multiple of each other (in this case, the "L2" wire length in each half of the "L2" coil is exactly four times as long as the "L1" coil wire length). This is likely to cause the "L1" coil to have part of a turn, due to the different coil diameters. For example, if the length of the "L2" coil wire is 160 inches and "L1" is to be one quarter of that length, namely, 40 inches. Then, if the "L1" coil has an effective diameter of 2.25 inches, (allowing for the thickness of the wire when wound on a 2-inch diameter former), then the "L1" coil would have 5.65 (or 5 and 2/3) turns which causes the finishing turn of "L2" to be 240 degrees further around the coil former than the start of the first turn - that is, five full turns plus two thirds of the sixth turn.

The L1 / L2 coil arrangement is a Tesla Coil. The positioning of the "L1" coil along the length of the "L2" coil, adjusts the voltage to current ratio produced by the coil. When the "L1" coil is near the middle of the "L2" coil, then the amplified voltage and amplified current are roughly the same. However, Don stresses that the "height" length of the coil (when standing vertically) controls the voltage produced while the coil "width" (the diameter of the turns) controls the current produced.

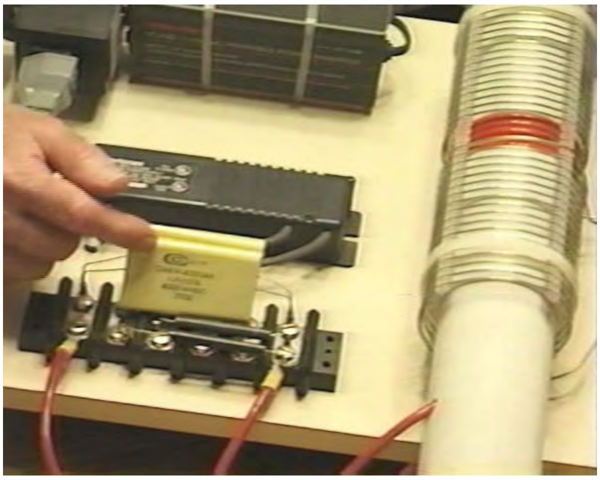

The exact wire length ratio of the turns in the "L1" and "L2" coils gives them an almost automatic synchronous tuning with each other, and the exact resonance between them can be achieved by the positioning of the "L1" coil along the length of the "L2" coil. While this is a perfectly good way of adjusting the circuit, in the 1994 build shown in the photograph, Don has opted to get the exact tuning by connecting a capacitor across "L1" as marked as "**C**" in the circuit diagram. Don found that the appropriate capacitor value for his particular coil build, was about 0.1 microfarad (100 nF) and so he connected two 47 nF high-voltage capacitors in parallel to get the value which he wanted. It must be remembered that the voltage across "L1" is very high, so a capacitor used in that position needs a voltage rating of at least 9,000 volts. Don remarks that the actual capacitors seen in the photograph of this prototype are rated at fifteen thousand volts, and were custom made for him using a "self-healing" style of construction.

Don has also connected a small capacitor across the "L2" coil, and that optional component is marked as "C2" in the circuit diagram and the value used by Don happened to be a single 47nF, high-voltage capacitor. As the two halves of the "L2" coil are effectively connected across each other, it is only necessary to have one capacitor for "L2":

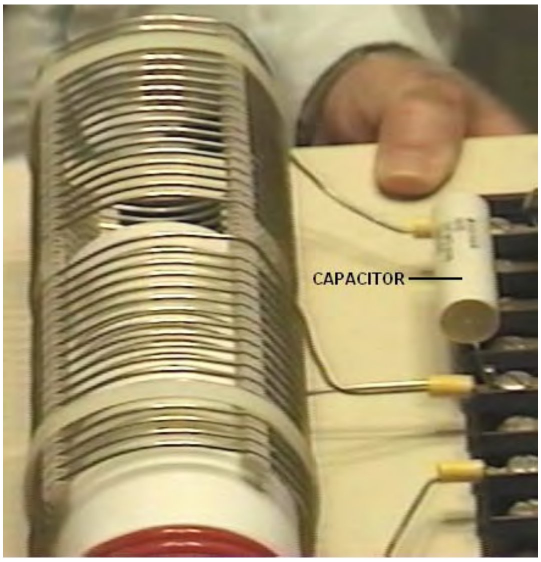

There are various ways of dealing with the output from the "L2" coil in order to get large amounts of conventional electrical power out of the device. The method shown here uses the four very large capacitors seen in the photograph. Each of these four capacitors are said to be 8 microfarad capacity with a 2,000 volt rating. In spite of the fact that they appear to be wired in parallel, Don states that they are in fact, wired in series to make a single capacitor of 2 microfarad capacity with an 8,000 volt rating. Don remarks that he has to be very careful to keep the voltage to the neon-tube driver circuit turned down in order to avoid getting more than 8,000 volts on these output storage capacitors.

This capacitor bank is fed through a diode which is rated for both high voltage and high current, as Don states that the device produces 8,000 volts at 20 amps, in which case, this rectifying diode has to be able to handle that level of power, both at start-up when the capacitor bank is fully discharged and "L2" is producing 8,000 volts, and when the full load of 20 amps is being drawn. The actual diodes used by Don happen to be rated at 25 KV but that is a far greater rating than is actually needed.

In passing, it might be remarked that the average home user will not have an electrical requirement of anything remotely like as large as this, seeing that 10 kW is more than most people use on a continuous basis, while 8 KV at 20 A is a power of 160 kilowatts. As the neon-tube driver circuit can put out 9,000 volts and since the L1 / L2 coil system is a step-up transformer, if the voltage fed to the capacitor bank is to be kept down to 8,000 volts, then the Variac adjustment must be used to reduce the voltage fed to the neon-tube driver circuit, in order to lower the voltage fed to the L1 / L2 coil pair, typically, to 3,000 volts.

A very astute and knowledgeable member of the EVGRAY Yahoo EVGRAY forum whose ID is "silverhealtheu" has recently pointed out that Don Smith says quite freely that he does not disclose all of the details of his designs, and it is his opinion that a major item which has not been disclosed is that the diodes in the circuit diagrams shown here are the wrong way round and that Don operates his voltages in reverse to the conventional way. In fact, the circuit diagram should be:

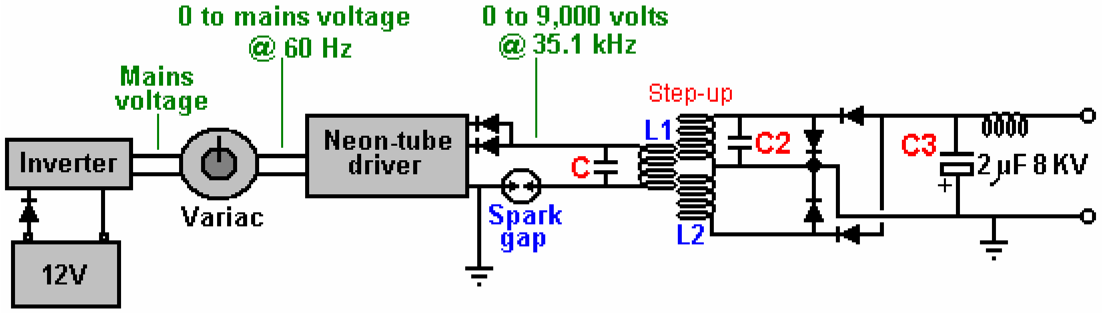

He comments: "the diodes leaving the Neon-tube Driver may need to be reversed as we want to collect the negative polarity. The spark gap will then operate on ambient inversion and the spark will look and sound totally different with a much faster crack and producing very little heat and even becoming covered in frost is possible.

The Variac should be raised up just enough to get a spark going then backed off slightly. Any higher voltage is liable to make the Neon-tube Driver think that it has a short-circuit condition, and the new electronic designs will then shut down automatically and fail to operate at all if this method is not followed.

When running, C, L1 and L2 operate somewhere up in the Radio Frequency band because the Neon-tube Driver only acts as a tank-circuit exciter. The large collection capacitor C3, should fill inverted to earth polarity as shown above. The load will then be pulling electrons from the earth as the cap is REFILLED back to ZERO rather than the joules in the capacitor being depleted.

Also remember that the Back-EMF systems of John Bedini and others, create a small positive pulse but they collect a super large NEGATIVE polarity spike which shoots off the bottom of an oscilloscope display. This is what we want, plenty of this stored in capacitors, and then let the ambient background energy supply the current when it makes the correction."

This is **a very important point** and it may well make a really major difference to the performance of a device of this nature.

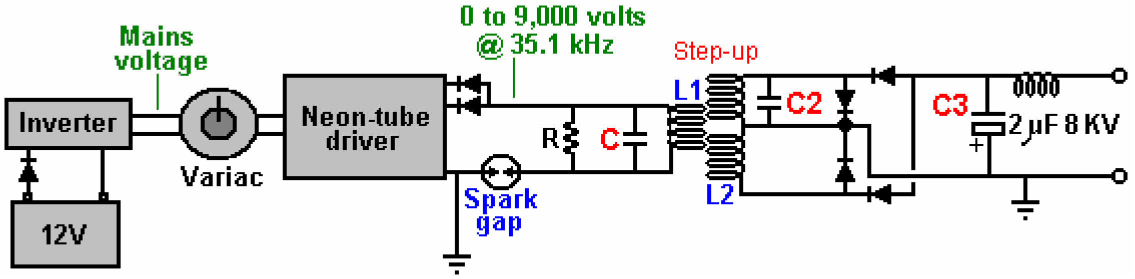

One reader has drawn attention to the fact that Don's main document indicates that there should be a resistor "R" across the L1 coil as well as the capacitor "C" and he suggests that the circuit should actually be as shown above, considering what Don said earlier about his "suitcase" design. Another reader points out that the wire in the output choke shown in the photograph below appears to be wound with wire that is far too small diameter to carry the currents mentioned by Don. I seems likely that a choke is not needed in that position, but a more powerful choke can easily be wound using larger diameter wire.

When the circuit is running, the storage capacitor bank behaves like an 8,000 volt battery which never runs down and which can supply 20 amps of current for as long as you want. The circuitry for producing a 220 volt 50 Hz AC output or a 110 volt 60 Hz AC output from the storage capacitors is just standard electronics. In passing, one option for charging the battery is to use the magnetic field caused by drawing mains-frequency current pulses through the output "choke" coil, shown here:

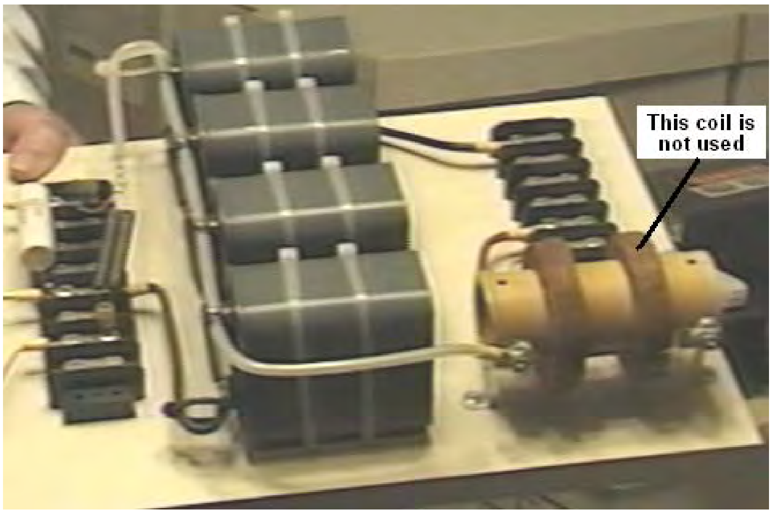
This coil is not used

The output current flows through the left hand winding on the brown cylindrical former, and when the photograph was taken, the right-hand winding was no longer in use. Previously, it had been used to provide charging power to the battery by rectifying the electrical power in the coil, caused by the fluctuating magnetic field caused by the pulsing current flowing through the left hand winding, as shown here:

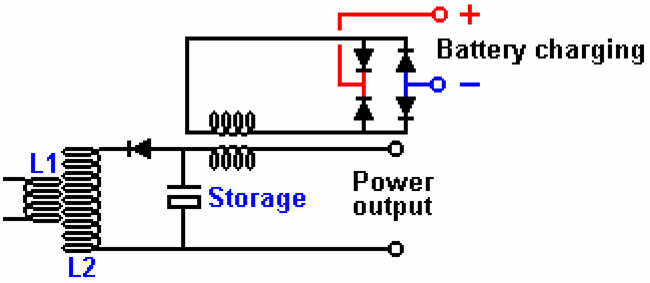
+
Battery charging
–
L1
L2
Storage
Power output

The DC output produced by the four diodes was then used to charge the driving battery, and the power level produced is substantially greater than the minor current drain from the battery. Consequently, it is a sensible precaution to pass this current to the battery via a circuit which prevents the battery voltage rising higher than it should. A simple voltage level sensor can be used to switch off the charging when the battery has reached its optimum level. Other batteries can also be charged if that is wanted. Simple circuitry of the type shown in chapter 12 can be used for controlling and limiting the charging process. The components on Don's board are laid out like this:

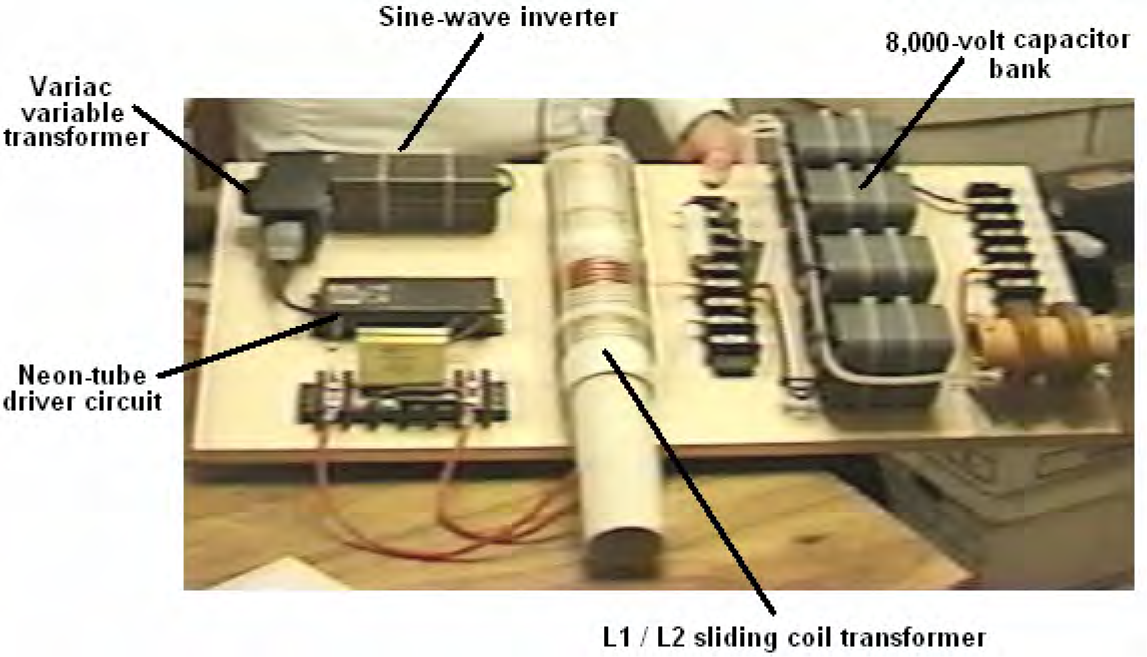
Sine-wave inverter
8,000-volt capacitor bank
Variac variable transformer
Neon-tube driver circuit
L1 / L2 sliding coil transformer

Don draws attention to the fact that the cables used to connect the output of "L2" to the output of the board, connecting the storage capacitors on the way, are very high-voltage rated cables with special multiple coverings to ensure that the cables will remain sound over an indefinite period. It should be remarked at this point, that the outer 3" diameter coil used by Don, is not wound on a former, but in order to get higher performance at high frequencies, the turns are supported with four separate strips physically attached to the turns - the technique described later in this document as being an excellent way for home construction of such coils.

Please bear in mind that the voltages here and their associated power levels are literally lethal and perfectly capable of killing anyone who handles the device carelessly when it is powered up. When a replication of this device is ready for routine use, it must be encased so that none of the high-voltage connections can be touched by anyone. This is not a suggestion, but it is a mandatory requirement, despite the fact that the components shown in the photographs are laid out in what would be a most dangerous fashion were the circuit to be powered up as it stands. Under no circumstances, construct and test this circuit unless you are already experienced in the use of high-voltage circuits or can be supervised by somebody who is experienced in this field. This is a "one hand in the pocket at all times" type of circuit and it needs to be treated with great care and respect at all times, so be sensible.

The remainder of the circuit is not mounted on the board, possibly because there are various ways in which the required end result can be achieved. The one suggested here is perhaps the most simple solution:

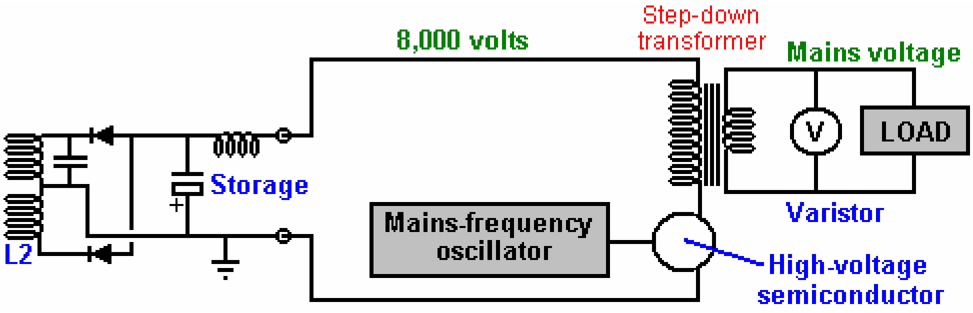

The voltage has to be dropped, so an iron-cored mains-frequency step-down transformer is used to do this. To get the frequency to the standard mains frequency for the country in which the device is to be used, an oscillator is used to generate that particular mains frequency. The oscillator output is used to drive a suitable high-voltage semiconductor device, be it an FET transistor, an IGBT device, or whatever. This device has to switch the working current at 8,000 volts, though admittedly, that will be a current which will be at least thirty six times lower than the final output current, due to the higher voltage on the primary winding of the transformer. The available power will be limited by the current handling capabilities of this output transformer.

As the circuit is capable of picking up additional magnetic pulses, such as those generated by other equipment, nearby lightning strikes, etc. an electronic component called a "varistor" marked "V" in the diagram, is connected across the load. This device acts as a voltage spike suppressor as it short circuits any voltage above its design voltage, protecting the load from power surges.

Don also explains an even more simple version of the circuit as shown here:

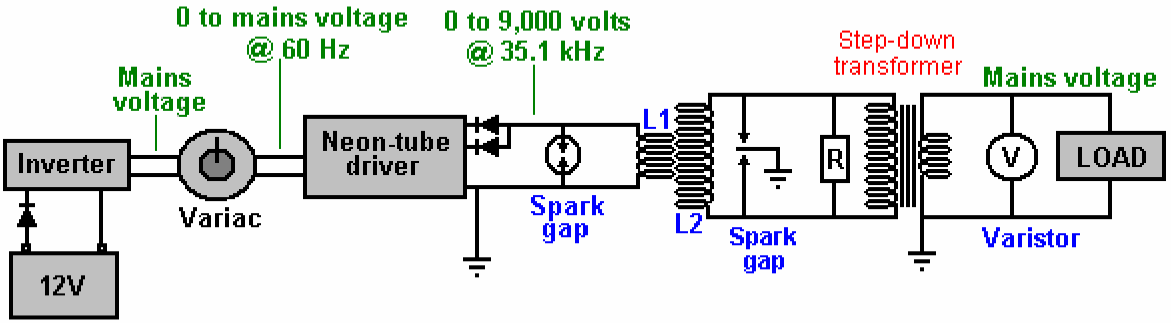

This simplified circuit avoids the need for expensive capacitors and the constraints of their voltage ratings, and the need for electronic control of the output frequency. The wire length in the turns of coil "L2" still needs to be exactly four times the wire length of the turns in coil "L1", but there is only one component which needs to be introduced, and that is the resistor "**R**" placed across the primary winding of the step-down isolation transformer. This transformer is a laminated iron-core type, suitable for the low mains frequency, but the output from "L2" is at much higher frequency. It is possible to pull the frequency down to suit the step-down transformer by connecting the correct value of resistor "R" across the output transformer (or a coil and resistor, or a coil and a capacitor). The value of resistor needed can be predicted from the American Radio Relay League graph (shown as Fig.44 in Don's .pdf document which can be downloaded from the www.free-energy-info.com website). The sixth edition of the Howard Sams book "Handbook of Electronics Tables and Formulas" (ISBN-10: 0672224690 or ISBN-13**:** 978-0672224690) has a table which goes down to 1 kHz and so does not need to be extended to reach the frequencies used here. The correct resistor value could also be found by experimentation. You will notice that an earthed dual spark gap has been placed across "L2" in order to make sure that the voltage levels always stay within the design range. Don remarks that he intends this particular device to be built by anyone who wants to, providing power for that person's needs and he states that some two hundred replications have already been built.

Don also explains an even more simple version which does not need a Variac, high voltage capacitors or high voltage diodes. Here, a DC output is accepted which means that high-frequency step-down transformer operation can be used. This calls for an air-core transformer which you would wind yourself from heavy duty wire. Mains loads would then be powered by using a standard off-the-shelf inverter. In this version, it is of course, necessary to make the "L1" turns wire length exactly one quarter of the "L2" turns wire length in order to make the two coils resonate together. The operating frequency of each of these coils is imposed on them by the output frequency of the neon-tube driver circuit. That frequency is maintained throughout the entire circuit until it is rectified by the four diodes feeding the low-voltage storage capacitor. The target output voltage will be either just over 12 volts or just over 24 volts, depending on the voltage rating of the inverter which is to be driven by the system. The circuit diagram is:

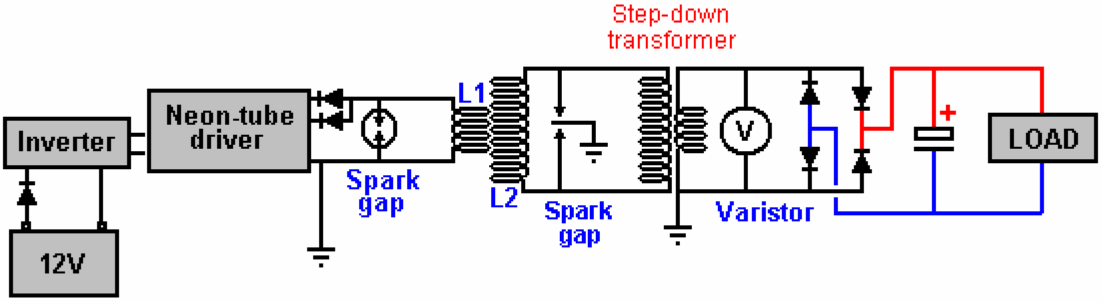

As many people will find the nomograph chart in Don's pdf document very difficult to understand and use, here is an easier version:

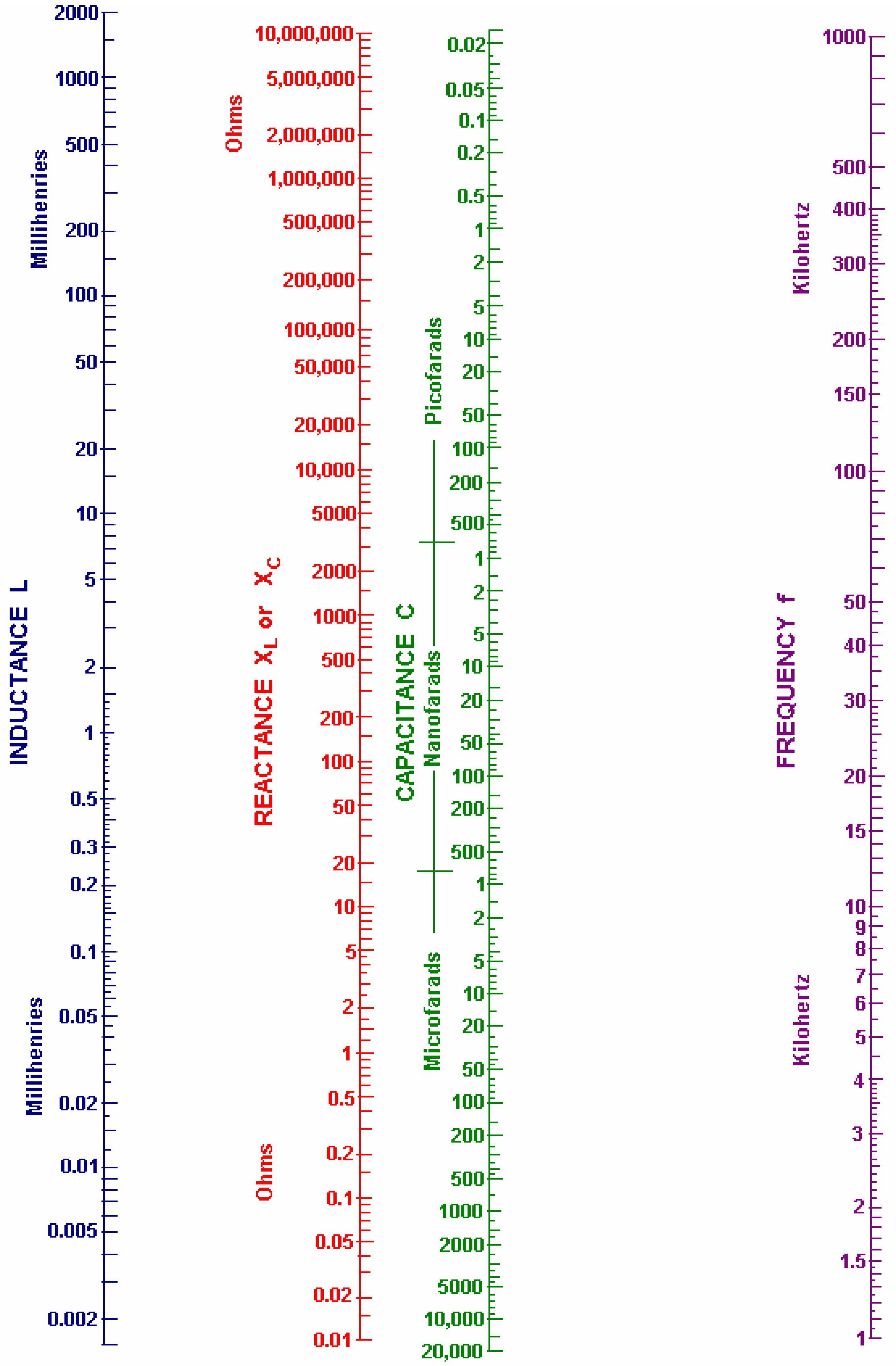

The objective here is to determine the "reactance" or 'AC resistance' in ohms and the way to do that is as follows:

Suppose that your neon-tube driver is running at 30 kHz and you are using a capacitor of 100 nF (which is the same as 0.1 microfarad) and you want to know what is the AC resistance of your capacitor is at that frequency. Also, what coil inductance would have that same AC resistance. Then the procedure for finding that out is as follows:

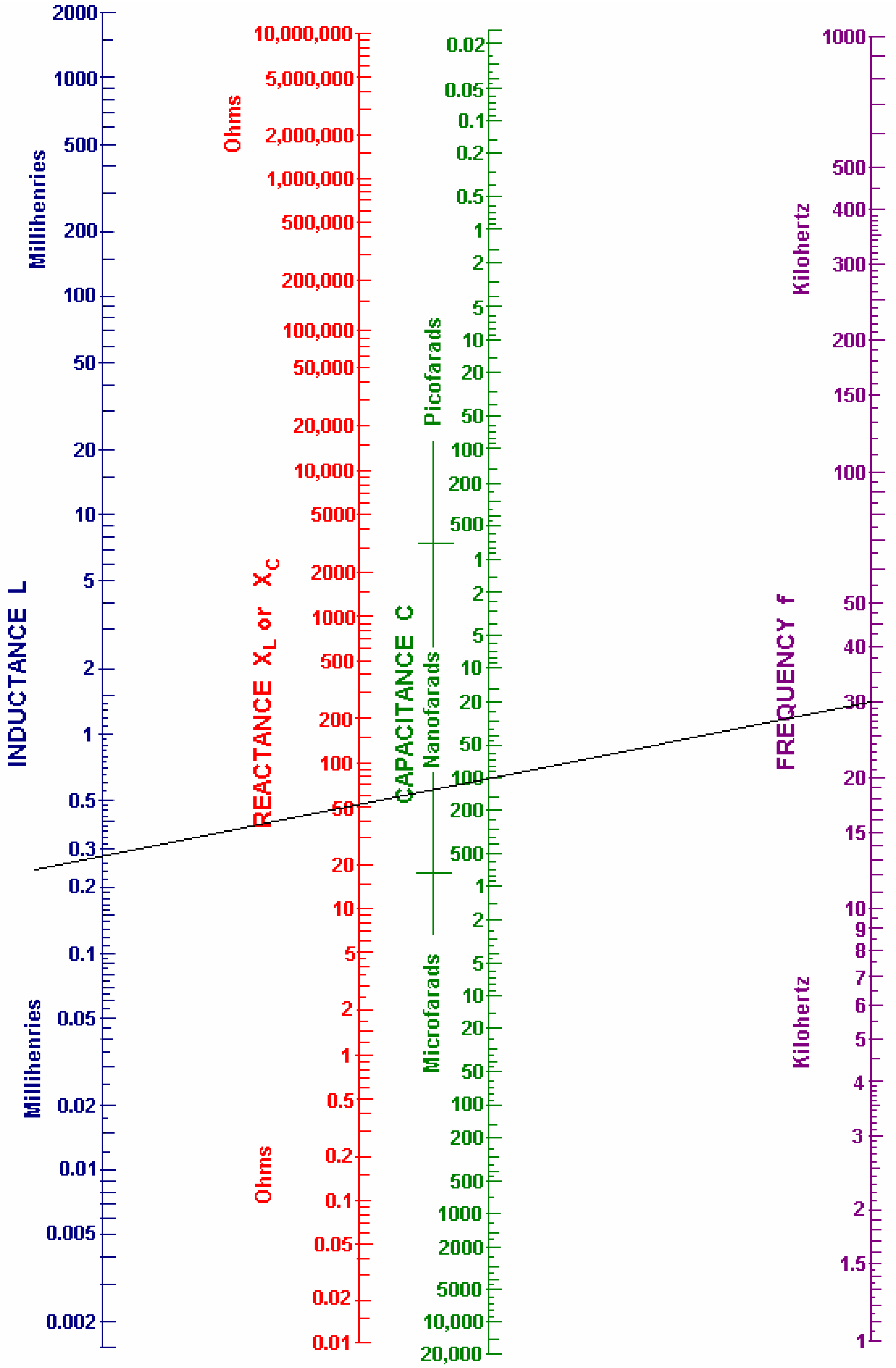

Draw a straight line from your 30 kHz frequency (purple line) through your 100 nanofarad capacitor value and carry the line on as far as the (blue) inductance line as shown above.

You can now read the reactance ("AC resistance") off the red line, which looks like 51 ohms to me. This means that when the circuit is running at a frequency of 30 kHz, then the current flow through your 100 nF capacitor will be the same as through a 51 ohm resistor. Reading off the blue "Inductance" line that same current flow at that frequency would occur with a coil which has an inductance of 0.28 millihenries.

I have been passed a copy of Don's circuit diagram for this device, and it is shown here:

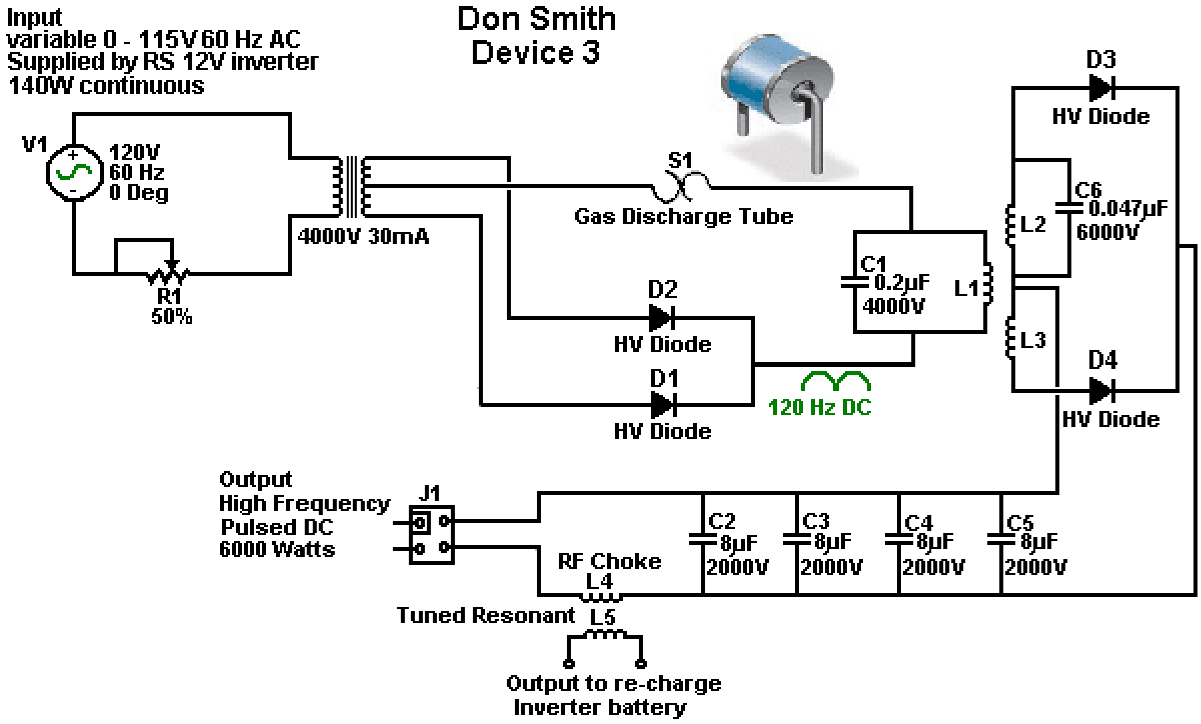

Input
variable 0 - 115V 60 Hz AC
Supplied by RS 12V inverter
140W continuous
Don Smith
Device 3
V1
120V
60 Hz
0 Deg
R1
50%
4000V 30mA
S1
Gas Discharge Tube
C1
0.2μF
4000V
L1
D2
HV Diode
D1
HV Diode
120 Hz DC
D3
HV Diode
L2
C6
0.047μF
6000V
L3
D4
HV Diode
Output
High Frequency
Pulsed DC
6000 Watts
J1
RF Choke
L4
Tuned Resonant
L5
C2
8μF
2000V
C3
8μF
2000V
C4
8μF
2000V
C5
8μF
2000V
Output to re-charge
Inverter battery

The 4000V 30mA transformer shown in this circuit diagram, may use a ferrite-cored transformer from a neon-tube driver module which steps up the voltage but it does not raise the frequency as that is clearly marked at 120 Hz pulsed DC. You will notice that this circuit diagram is drawn with Plus shown below Minus (which is most unusual).

Please note that when an earth connection is mentioned in connection with Don Smith's devices, we are talking about an actual wire connection to a metal object physically buried in the ground, whether it is a long copper rod driven into the ground, or an old car radiator buried in a hole like Tariel Kapanadze used, or a buried metal plate. When Thomas Henry Moray performed his requested demonstration deep in the countryside at a location chosen by the sceptics, the light bulbs which formed his demonstration electrical load, glowed more brightly with each hammer stroke as a length of gas pipe was hammered into the ground to form his earth connection.

Don also explains an even more simple version of his main device. This version does not need a Variac (variable voltage transformer) or high voltage capacitors. Here, a DC output is accepted which means that high-frequency step-down transformer operation can be used. This calls on the output side, for an air-core (or ferrite rod core) transformer which you would wind yourself from heavy duty wire. Mains loads would then be powered by using a standard off-the-shelf inverter. In this version, it is of course, very helpful to make the "L1" turns wire length exactly one quarter of the "L2" turns wire length in order to make the two coils automatically resonate together. The operating frequency of each of these coils is imposed on them by the output frequency of the neon-tube driver circuit. That frequency is maintained throughout the entire circuit until it is rectified by the four diodes feeding the low-voltage storage capacitor. The target output voltage will be either just over 12 volts or just over 24 volts, depending on the voltage rating of the inverter which is to be driven by the system.

As the circuit is capable of picking up additional magnetic pulses, such as those generated by other equipment, nearby lightning strikes, etc. an electronic component called a "varistor" marked "V" in the diagram, is connected across the load. This device acts as a voltage spike suppressor as it short-circuits any voltage above its design voltage, protecting the load from power surges. A Gas-Discharge Tube is an effective alternative to a varistor.

This circuit is effectively two Tesla Coils back-to-back and the circuit diagram might be:

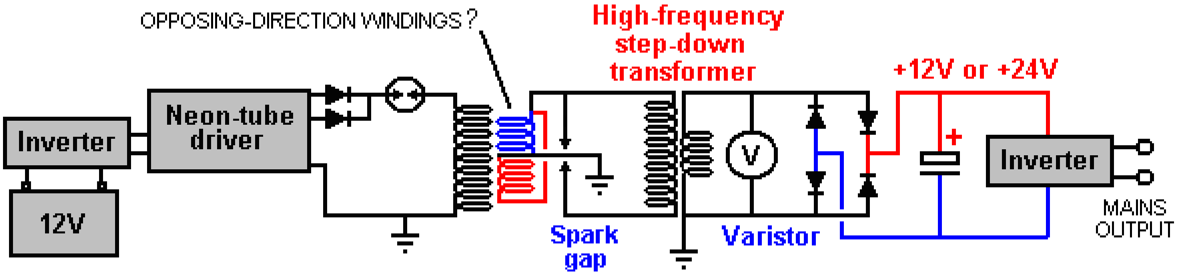

It is by no means certain that in this circuit, the red and blue windings are wound in opposing directions. The spark gap (or gas-discharge tube) in series with the primary of the first transformer alters the operation in a somewhat unpredictable way as it causes the primary to oscillate at a frequency determined by it's inductance and it's self-capacitance, and that may result in megahertz frequencies. The secondary winding(s) of that transformer **must** resonate with the primary and in this circuit which has no frequency-compensating capacitors, that resonance is being produced by the exact wire length in the turns of the secondary. This looks like a simple circuit, but it is anything but that. The excess energy is produced by the raised frequency, the raised voltage, and the very sharp pulsing produced by the spark. That part is straightforward. The remainder of the circuit is likely to be very difficult to get resonating as it needs to be in order to deliver that excess energy to the output inverter.

When considering the “length” of wire in a resonant coil, it is necessary to pay attention to the standing wave created under those conditions. The wave is caused by reflection of the signal when it reaches the end of the wire OR when there is a sudden change in the diameter of the wire as that changes the signal reflection ability at that point in the connection. You should pay attention to Richard Quick's very clear description of this in the section of his patent which is included later on in this chapter. Also, remember what Don Smith said about locating the peaks of the standing wave by using a hand-held neon lamp.

One very significant thing which Don pointed out is that the mains electricity available through the wall socket in my home, does **not** come along the wires from the generating station. Instead, the power station influences a local ‘sub-station’ and the electrons which flow through my equipment actually come from my local environment because of the influence of my local sub-station. Therefore, if I can create a similar influence in my home, then I no longer need that sub-station and can have as much electrical energy as I want, without having to pay somebody else to provide that influence for me.

A Practical Implementation of one of Don Smith's Designs

The objective here, is to determine how to construct a self-powered, free-energy electrical generator which has no moving parts, is not too expensive to build, uses readily available parts and which has an output of some kilowatts. However, under no circumstances should this document be considered to be an encouragement for you, or anyone else to actually build one of these devices. This document is presented solely for information and educational purposes, and as high voltages are involved, it should be considered to be a dangerous device unsuited to being built by inexperienced amateurs. The following section is just my opinions and so should not be taken as tried and tested, working technology, but instead, just the opinion of an inexperienced writer.

However, questions from several different readers indicate that a short, reasonably specific description of the steps needed to attempt a replication of a Don Smith device would be helpful. Again, this document must not be considered to be a recommendation that you actually build one of these high-voltage, potentially dangerous devices. This is just information intended to help you understand what I believe is involved in this process.

In broad outline, the following steps are used in the most simple version of the arrangement:

1. The very low frequency and voltage of the local mains supply is discarded in favour of an electrical supply which operates at more than 20,000 Hz (cycles per second) and has a voltage of anything from 350 volts to 10,000 volts. The higher voltages can give greater overall output power, but they involve greater effort in getting the voltage back down again to the level of the local mains voltage in order for standard mains equipment to be used.

2. This high-frequency high voltage is used to create a series of very rapid sparks using a spark gap which is connected to a ground connection. Properly done, the spark frequency is so high that there is no audible sound caused by the sparks. Each spark causes a flow of energy from the local environment into the circuit. This energy is not standard electricity which makes things hot when current flows through them, but instead this energy flow causes things to become cold when the power flows through them, and so it is often called “cold” electricity. It is tricky to use this energy unless all you want to do is light up a series of light bulbs (which

incidentally, give out a different quality of light when powered with this energy). Surprisingly, the circuit now contains substantially more power than the amount of power needed to produce the sparks. This is because additional energy flows in from the ground as well as from the local environment. If you have conventional training and have been fed the myth of "closed systems", then this will seem impossible to you. So, let me ask you the question: if, as can be shown, all of the electricity flowing into the primary winding of a transformer, flows back out of that winding, then where does the massive, continuous flow of electricity coming from the secondary winding come from? None of it comes from the primary circuit and yet millions of electrons flow out of the secondary in a continuous stream which can be supplied indefinitely. So, where do these electrons come from? The answer is 'from the surrounding local environment which is seething with excess energy' but your textbooks won't like that fact as they believe that the transformer circuit is a 'closed system' – something which probably can't be found anywhere in this universe.

3. This high-voltage, high-frequency, high-power energy needs to be converted to the same sort of hot electricity which comes out of a mains wall socket at the local voltage and frequency. This is where skill and understanding come into play. The first step is to lower the voltage and increase the available current with a step-down resonant transformer. This sounds highly technical and complicated, and looking at Don Smith's expensive Barker & Williamson coil, makes the whole operation appear to be one for rich experimenters only. This is not the case and a working solution can be cheap and easy. It is generally not convenient to get the very high voltage all the way down to convenient levels in a single step, and so, one or more of those resonant transformers can be used to reach the target voltage level. Each step down transformer boosts the available current higher and higher.

4. When a satisfactory voltage has been reached, we need to deal with the very high frequency. The easiest way to deal with it is to use high-speed diodes to convert it to pulsing DC and feed that into a capacitor to create what is essentially, an everlasting battery. Feeding this energy into a capacitor converts it into conventional "hot" electricity and a standard off-the-shelf inverter can be used to give the exact voltage and frequency of the local mains supply. In most of the world, that is 220 volts at 50 cycles per second. In America it is 110 volts at 60 cycles per second. Low-cost inverters generally run on either 12 volts or 24 volts with the more common 12 volt units being cheaper.

So, let's take a look at each of these step in more detail and see if we can understand what is involved and what our options are:

1. We want to produce a high-voltage, high-frequency, low-current power source. Don Smith shows a Neon-Sign Transformer module. His module produced a voltage which was higher than was convenient and so he used a variable AC transformer or "Variac" as it is commonly known, to lower the input voltage and so, lower the output voltage. There is actually no need for a Variac as we can handle the higher voltage or alternatively, use a more suitable Neon-Sign Transformer module.

   However, we have a problem with using that technique. In the years since Don bought his module, they have been redesigned to include circuitry which disables the module if any current flows out of it directly to earth, and as that is exactly what we would want to use it for, so most, if not all of the currently available neon-sign transformer modules are not suitable for our needs. However, I'm told that if the module has an earth wire and that earth wire is left unconnected, that it disables the earth-leakage circuitry, allowing the unit to be used in a Don Smith circuit. Personally, I would not recommend that if the module is enclosed in a metal housing.

   A much cheaper alternative is shown here: http://www.youtube.com/watch?v=RDDRe_4D93Q where a small plasma globe circuit is used to generate a high-frequency spark. It seems highly likely that one of those modules would suit our needs:

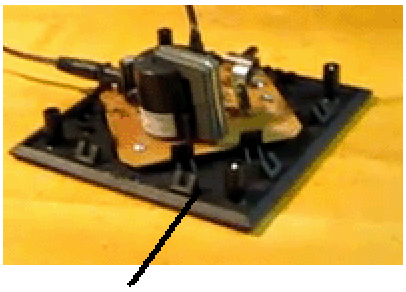

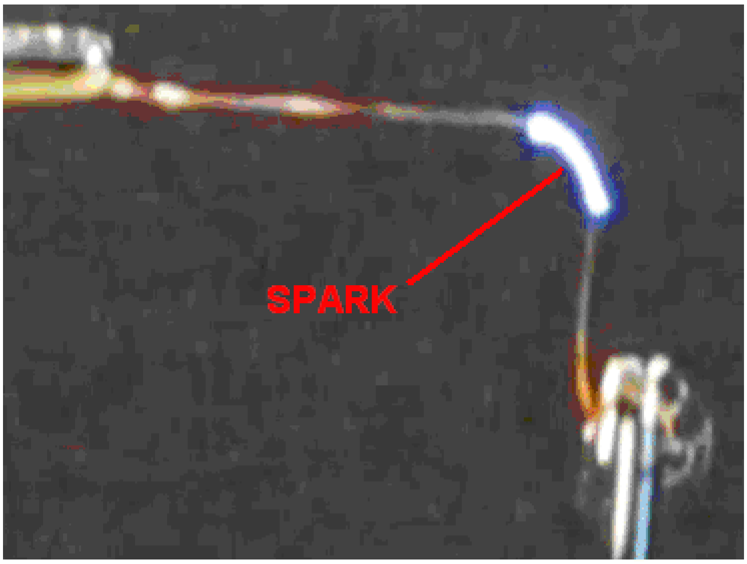

An alternative method is to build your own power supply from scratch. Doing that is not particularly difficult and if you do not understand any electronics, then perhaps, reading the beginner's electronics tutorial in chapter 12 (http://www.free-energy-info.com/Chapter12.pdf) will fill you in on all of the basics needed for understanding (and probably designing your own) circuits of this type. Here is a variable frequency design for home-construction:

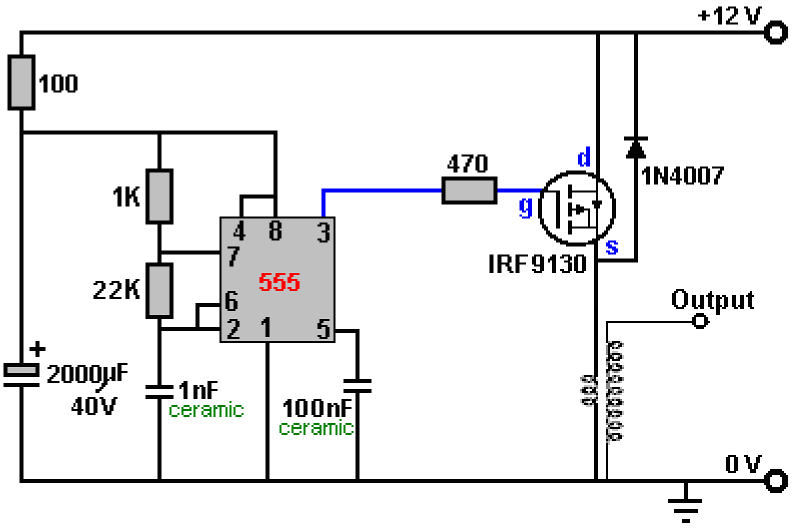

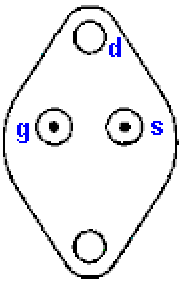

One advantage of this circuit is that the output transformer is driven at the frequency set by the 555 timer and that frequency is not affected by the number of turns in the primary winding, nor it's inductance, wire diameter, or anything else to do with the coil. While this circuit shows the rather expensive IRF9130 transistor, I expect that other P-channel FETs would work satisfactorily in this circuit. The IRF9130 transistor looks like this:

The circuit has a power supply diode and capacitor, ready to receive energy from the output at some later date if that is possible and desired. The 555 circuit is standard, giving a 50% Mark/Space ratio. The 10 nF capacitor is there to maintain the stability of the 555 and the timing section consists of two variable resistors, one fixed resistor and the 1 nF capacitor. This resistor arrangement gives a variable resistance of anything from 100 ohms to 51.8K and that allows a substantial frequency range. The 47K (Linear) variable resistor controls the main tuning and the 4.7K (Linear) variable resistor gives a more easily adjustable frequency for exact tuning. The 100 ohm resistor is there in case both of the variable resistors are set to zero resistance. The output is fed through a 470 ohm resistor to the gate of a very powerful P-channel FET transistor which drives the primary winding of the output transformer.

The output transformer can be wound on an insulating spool covering a ferrite rod, giving both good coupling between the windings, and high-frequency operation as well. The turns ratio is set to just 30:1 due to the high number of primary winding turns. With a 12-volt supply, this will give a 360-volt output waveform, and by reducing the primary turns progressively, allows the output voltage to be increased in controlled steps. With 10 turns in the primary, the output voltage should be 3,600 volts and with just 5 turns 7,200 volts. The higher the voltage used, the greater the amount of work needed later on to get the voltage back down to the output level which we want.

Looking at the wire specification table, indicates that quite a small wire diameter could be used for the oscillator output transformer's secondary winding. While this is perfectly true, it is not the whole story. Neon Tube Drivers are very small and the wire in their output windings is very small diameter indeed. Those driver modules are very prone to failure. If the insulation on any one turn of the winding fails and one turn becomes a short-circuit, then that stops the winding from oscillating, and a replacement is needed. As there are no particular size constraints for this project, it might be a good idea to use enamelled copper wire of 0.45 mm or larger in an attempt to avoid this insulation failure hazard. No part of the transformer coil spool should be metal and it would not be any harm to cover each layer of secondary winding with a layer of electrical tape to provide additional insulation between the coil turns in one layer and the turns in the layer on top of it.

A plug-in board layout might be:

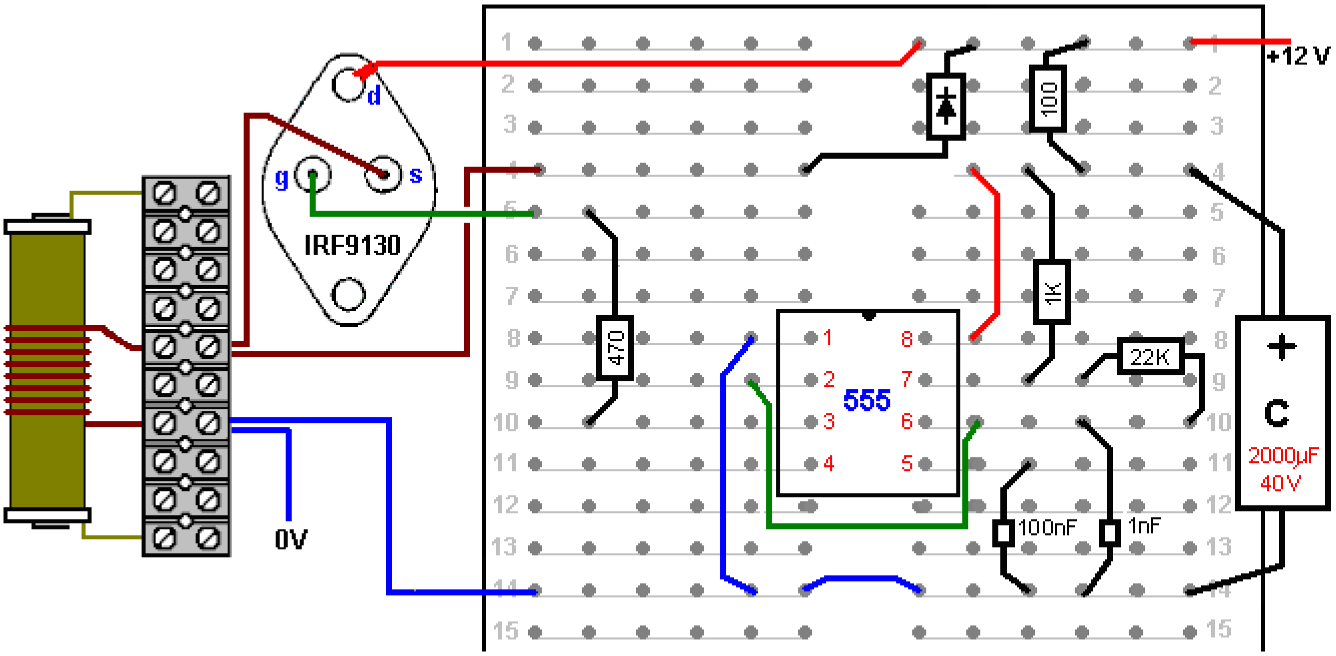

Please remember that you can't just stick your average voltmeter across a 4 kV capacitor (unless you really do want to buy another meter) as they only measure up to about a thousand volts DC. So, if you are using high voltage, then you need to use a resistor-divider pair and measure the voltage on the lower resistor. But what resistor values should you use? If you put a 10 Megohm resistor across your 4 kV charged capacitor, the current flowing through the resistor would be 0.4 milliamps. Sounds tiny, doesn't it? But that 0.4 mA is 1.6 watts which is a good deal more than the wattage which your resistor can handle. Even using this arrangement:

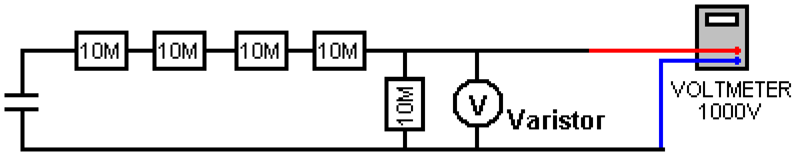

the current will be 0.08 mA and the wattage per resistor will be 64 mW. The meter reading will be about 20% of the capacitor voltage which will give a voltmeter reading of 800 volts. The input resistance of the meter needs to be checked and possibly, allowed for as the resistance in this circuit is so high (see chapter 12). When making a measurement of this type, the capacitor is discharged, the resistor chain and meter attached, and then, and only then, is the circuit powered up, the reading taken, the input power disconnected, the capacitor discharged, and the resistors disconnected. High-voltage circuits are highly dangerous, especially so, where a capacitor is involved. The recommendation to wear thick rubber gloves for this kind of work, is not intended to be humorous. Circuits of this type are liable to generate unexpected high-voltage spikes, and so, it might be a good idea to connect a varistor across the meter to protect it from those spikes. The varistor need to be set to the voltage which you intend to measure and as varistors may not be available above a 300V threshold, two or more may need to be connected in series where just one is shown in the diagram above. The varistor should not have a higher voltage rating than your meter.

2. We now need to use this high voltage to create a strategically positioned spark to a ground connection. When making an earth connection, it is sometimes suggested that connecting to water pipes or radiators is a good idea as they have long lengths of metal piping running under the ground and making excellent contact with it. However, it has become very common for metal piping to be replaced with cheaper plastic piping and so any proposed pipe connection needs a check to ensure that that there is metal piping which runs all the way into the ground.

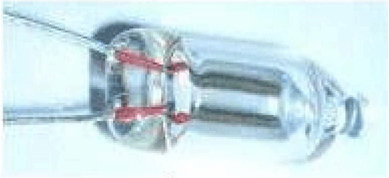
Neon

Gas-Discharge Tube

The spark gaps shown can be commercial high-voltage gas discharge tubes, adjustable home-made spark gaps with stainless steel tips about 1 mm apart, car spark plugs, or standard neon bulbs, although these run rather hot in this application. A 15 mm x 6 mm size neon bulb operates with only 90 or 100 volts across it, it would take a considerable number of them connected in series to create a high voltage spark gap, but it is probably a misconception that the spark gap itself needs a high voltage. Later on in this chapter, there is an example of a very successful system where just one neon bulb is used for the spark gap and an oscillating magnetic field more than a meter wide is created when driven by just an old 2,500 volt neon-sign transformer module. If using a neon bulb for the spark gap, then an experienced developer recommends that a 22K resistor is used in series with the neon in order to extend it's working life very considerably.

This circuit is one way to connect the spark gap and ground connection:

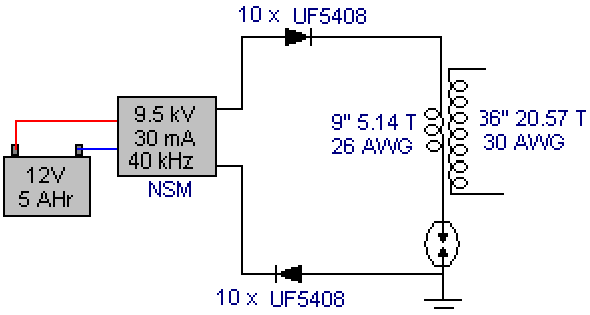

This is an adaption of a circuit arrangement used by the forum member "SLOW-'N-EASY" on the Don Smith topic in the energeticforum. Here, he is using a 'LowGlow' neon transformer intended for use on a bicycle. The diodes are there to protect the high-voltage power supply from any unexpected voltage spikes created later on in the circuit. The spark gap is connected between the primary winding of a step-up transformer and the earth connection. No capacitor is used. Seeing this circuit, we immediately think of Don Smith's large and expensive coils, but this experimenter does not use anything like that. Instead, he winds his transformer on a simple plastic former like this:

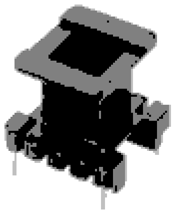

Ho Sung International. EI-2820 nylon bobbin. Core is 10 mm x 13 mm x 10 mm high. Top is 18.5 mm x 21.5 mm. Base is 22 mm x 26 mm. Four leads, 15 mm and 20 mm spacing

And to make matters 'worse' the primary winding wire is just 9 inches (228.6 mm) long and the secondary just 36 inches (914.4 mm) long, the primary being wound directly on top of the secondary. Not exactly a large or expensive construction and yet one which appears to perform adequately in actual tests.

This is a very compact form of construction, but there is no necessity to use exactly the same former for coils, nor is there anything magic about the nine-inch length of the L1 coil, as it could easily be any convenient length, say two feet or 0.5 metres, or whatever. The important thing is to make the L2 wire length exactly four times that length, cutting the lengths accurately. It is common practice to match the weight of copper in each coil and so the shorter wire is usually twice the diameter of the longer wire.

The circuit above, produces a cold electricity output of high voltage and high frequency. The voltage will not be the same as the neon transformer voltage, nor is the frequency the same either. The two coils resonate at their own natural frequency, unaltered by any capacitors.

3. The next step is to get the high voltage down to a more convenient level, perhaps, like this:

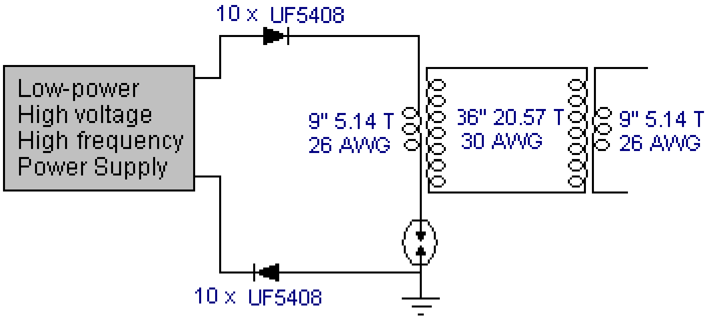

Here, an identical transformer, wound in exactly the same way, is used in reverse, to start the voltage lowering sequence. The wire length ratio is maintained to keep the transformer windings resonant with each other.

Supposing we were to wind the L2 coil of this second transformer in a single straight winding and instead of winding just one L1 winding on top of it, two or more L1 identical windings were placed on top of it – what would happen?:

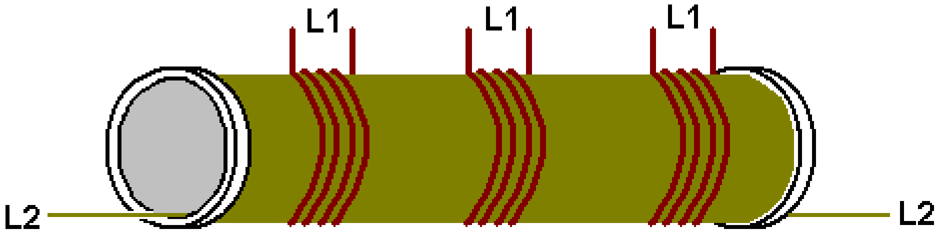

Now for a comment which will seem heretical to people steeped in the present day (inadequate) level of technology. The power flowing in these transformers is cold electricity which operates in an entirely different way to hot electricity. The coupling between these coils would be inductive if they were carrying hot electricity and in that case, any additional power take-off from additional L1 coils would have to be 'paid' for by additional current draw through the L2 coil. However, with the cold electricity which these coils are actually carrying, the coupling between the coils is magnetic and not inductive and that results in no increase in L2 current, no matter how many L1 coil take-offs there are. Any additional L1 coils will be powered for free. However, the position of the coils relative to each other has an effect on the tuning, so the L1 coil should be in the middle of the L2 coil, which means that any additional L1 coils are going to be slightly off the optimum tuning point.

4. Anyway, following through on just one L1 coil, there is likely to be at least one further step-down transformer needed and eventually, we need conversion to hot electricity:

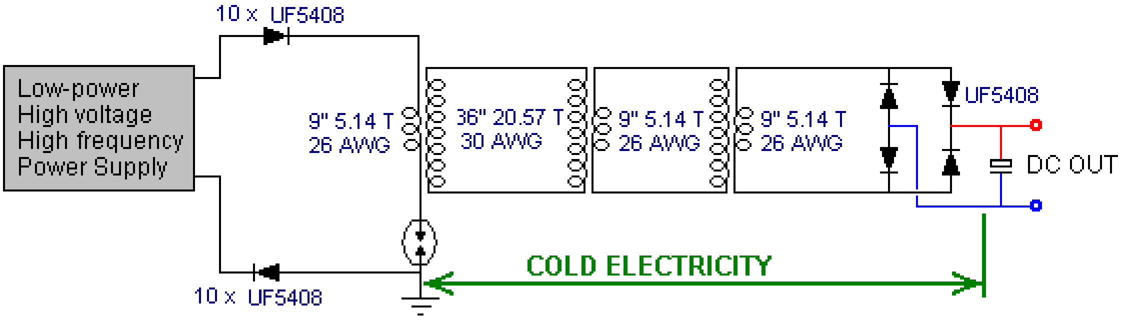

Probably the easiest conversion is by feeding the energy into a capacitor and making it standard DC. The frequency is still very high, so high-speed diodes (such as the 75-nanosecond UF54008) are needed here although the voltage level is now low enough to be no problem. The DC output can be used to power an inverter so that standard mains equipment can be used. It is not necessary to use just one (expensive) large-capacity inverter to power all possible loads as it is cheaper to have several smaller inverters, each powering it's own set of equipment. Most equipment will run satisfactorily on square-wave inverters and that includes a mains unit for powering the input oscillator circuit.

PVC pipe is not a great material when using high-frequency high-voltage signals, and grey PVC pipe is a particularly poor coil former material. The much more expensive acrylic pipe is excellent, but if using PVC, then performance will be better if the PVC pipe is coated with an insulating lacquer (or table tennis balls dissolved in acetone as show on YouTube).

However, there are some other factors which have not been mentioned. For example, if the L1 coil is wound directly on top of the L2 coil, it will have roughly the same diameter and so, the wire being four times longer, will have roughly four times as many turns, giving a step-up or step-down ratio of around 4:1. If, on the other hand, the coil diameters were different, the ratio would be different as the wire lengths are fixed relative to each other. If the L2 coil were half the diameter of the L1 coil, then the turns ratio would be about 8:1 and at one third diameter, 12:1 and at a quarter diameter 16:1 which means that a much greater effect could be had from the same wire length by reducing the L2 coil diameter. However, the magnetic effect produced by a coil is linked to the cross-sectional area of the coil and so a small diameter is not necessarily at great advantage. Also, the length of the L1 coil wire and number of turns in it, affect the DC resistance, and more importantly, the AC impedance which affects the amount of power needed to pulse the coil.

It is also thought that having the same weight of copper in each winding gives an improved performance, but what is not often mentioned is the opinion that the greater the weight of copper, the greater the effect. You will recall that Joseph Newman (chapter 11) uses large amounts of copper wire to produce remarkable effects. So, while 9 inches and 36 inches of wire will work for L1 and L2, there may well be improved performance from longer lengths of wire and/or thicker wires.

We should also not forget that Don Smith pointed out that voltage and current act (out of phase and) in opposite directions along the L2 coil, moving away from the L1 coil:

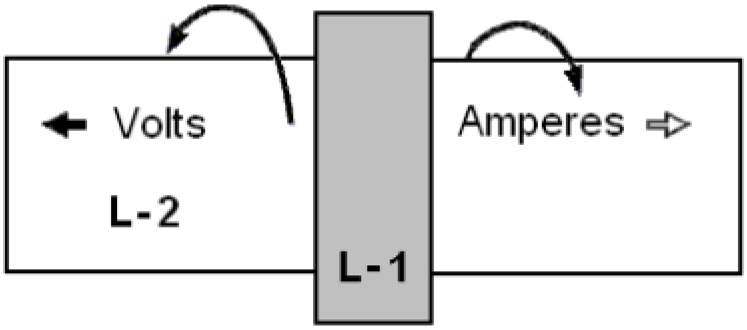

It has been suggested that a greater and more effective power output can be obtained by splitting the L2 coil underneath the L1 coil position, winding the second part of L2 in the opposite direction and grounding the junction of the two L2 windings. Don doesn't consider it necessary to reverse the direction of winding. The result is an L2 winding which is twice as long as before and arranged like this:

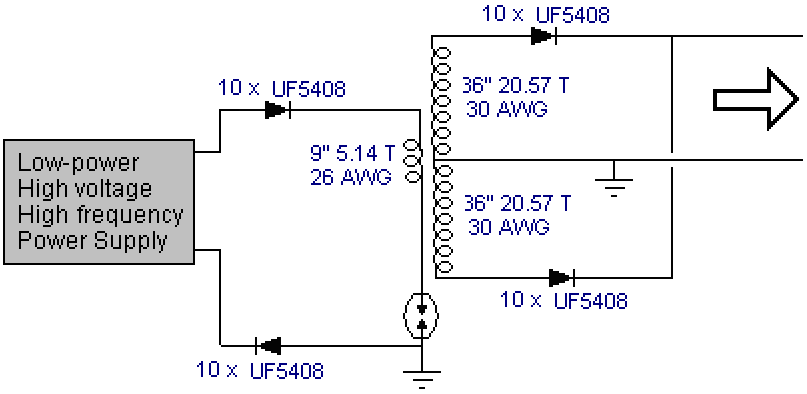
10 x UF5408
10 x UF5408
36" 20.57 T
30 AWG
9" 5.14 T
26 AWG
Low-power
High voltage
High frequency
Power Supply
36" 20.57 T
30 AWG
10 x UF5408
10 x UF5408

Here, the additional high-voltage diodes allow the two out of phase windings to be connected across each other. You will notice that this arrangement calls for two separate earth connections, both of which need to be high-quality connections, something like a pipe or rod driven deeply into moist soil or alternatively, a metal plate or similar metal object of substantial surface area, buried deep in moist earth, and a thick copper wire or copper braid used to make the connection. These earthing points need to be fairly far apart, say, ten metres. A single earth connection can't be used as that would effectively short-circuit across the L1/L2 transformer which you really do not want to do.

With this arrangement, the outline circuit becomes:

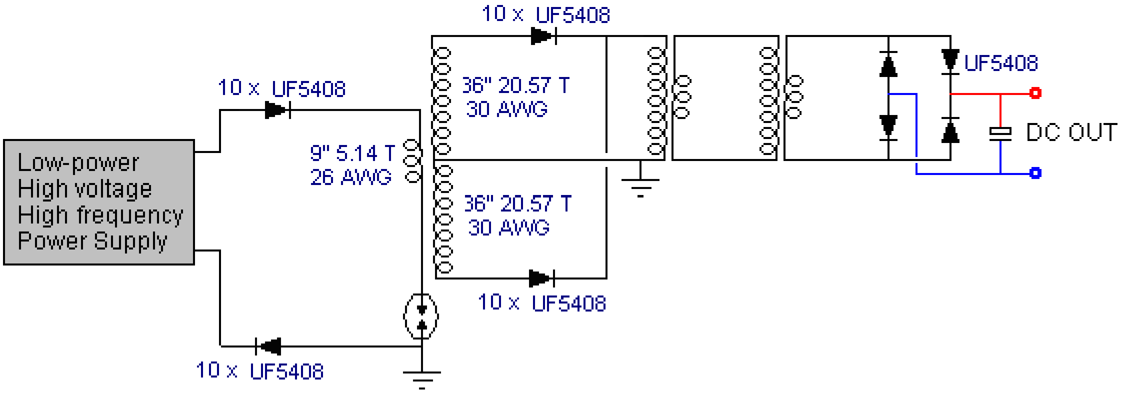
10 x UF5408
10 x UF5408
36" 20.57 T
30 AWG
9" 5.14 T
26 AWG
UF5408
DC OUT
Low-power
High voltage
High frequency
Power Supply
36" 20.57 T
30 AWG
10 x UF5408
10 x UF5408

The thick earth wiring is helpful because in order to avoid the earth wire being included in the resonant wire length, you need a sudden change in wire cross-section:

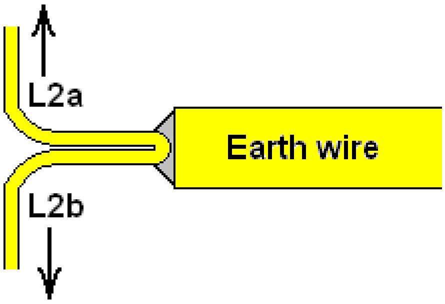
L2a
Earth wire
L2b

These are just some ideas which might be considered by some experienced developer who may be thinking of investigating Don Smith style circuitry.

To give you some idea of the capacity of some commercially available wires when carrying hot electricity, this table may help:

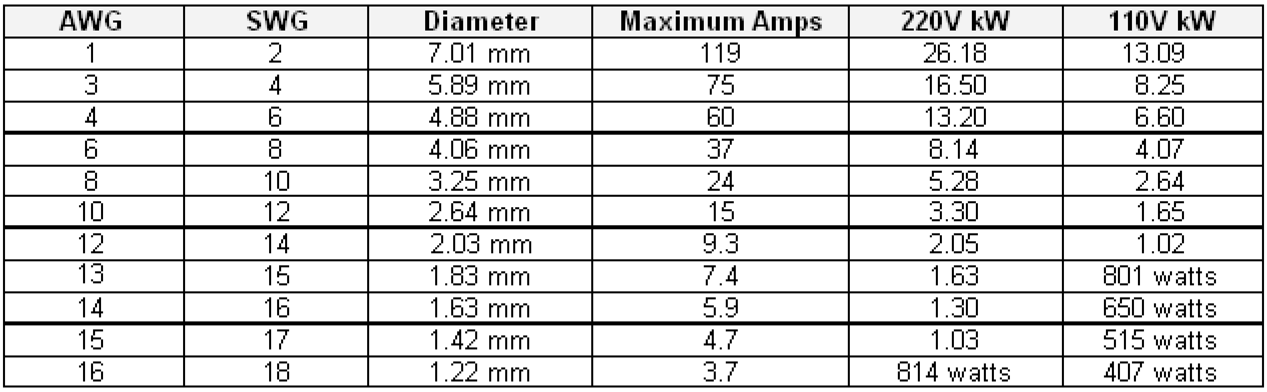

AWG	SWG	Diameter	Maximum Amps	220V kW	110V kW
1	2	7.01 mm	119	26.18	13.09
3	4	5.89 mm	75	16.50	8.25
4	6	4.88 mm	60	13.20	6.60
6	8	4.06 mm	37	8.14	4.07
8	10	3.25 mm	24	5.28	2.64
10	12	2.64 mm	15	3.30	1.65
12	14	2.03 mm	9.3	2.05	1.02
13	15	1.83 mm	7.4	1.63	801 watts
14	16	1.63 mm	5.9	1.30	650 watts
15	17	1.42 mm	4.7	1.03	515 watts
16	18	1.22 mm	3.7	814 watts	407 watts

It is recommended that the wire have a current carrying capacity of 20% more than the expected actual load, so that it does not get very hot when in use. The wire diameters do not include the insulation, although for solid enamelled copper wire, that can be ignored.

There is a most impressive video and circuit shown at http://youtu.be/Q3vr6qmOwLw where a very simple arrangement produces an immediately successful performance for the front end of Don's circuitry. The circuit appears to be:

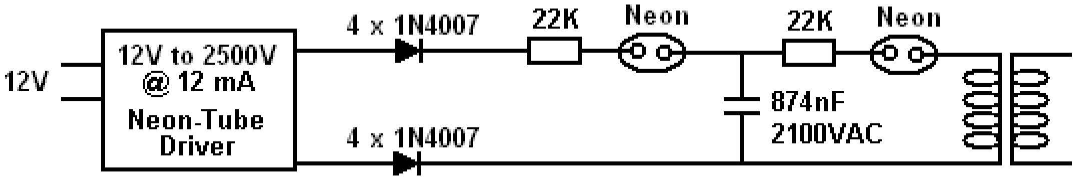

Here, a simple Neon Sign Transformer module which has no earth connection, is used to produce a 2.5 kV voltage with a frequency of 25 kHz and a maximum output current capacity of 12 mA. There is no difficulty in constructing the equivalent to that power supply unit. The two outputs from the module are converted to DC by a chain of four 1N4007 diodes in series in each of the two outputs (each chain being inside a plastic tube for insulation).

This output is fed through an optional 22K resistor via a neon lamp to a microwave oven capacitor which happens to be 874 nF with a voltage rating of 2,100 volts. You might feel that the voltage rating of the capacitor is too low for the output voltage of the neon sign module, but the neon has a striking voltage of just 90 volts and so the capacitor is not going to reach the output voltage of the power supply. The resistors are solely to extend the life of the neons as the gas inside the tube gets a considerable jolt in the first nanosecond after switch-on. It is unlikely that omitting those resistors would have any significant effect, but then, including them is a trivial matter. The second neon feeds the primary of the resonant transformer which is only shown in notional outline in the diagram above as the developer suggests that the primary acts as a transmitter and that any number of receiving coils can be used as individual secondaries by being tuned to the exact frequency of that resonating primary.

In the video showing this arrangement, the developer demonstrates the fluctuating, high-frequency field which extends for some four feet (1.2 m) around the coil. He also remarks that the single neons in his arrangement could each be replaced with two neons in series. In test which I ran, I found that I needed two neons in series ahead of the capacitor in order to get continuous lighting of the output neon. Also, one of the diodes needed to be reversed so that one faced towards the input and one away from it. It did not matter which diode was reversed as both configurations worked. Again, please note that this presentation is for information purposes only and it is **NOT** a recommendation that you should actually build one of these devices. Let me stress again that this is a high-voltage device made even more dangerous by the inclusion of a capacitor, and it is quite capable of killing you, so, don't build one. The developer suggests that it is an implementation of the "transmitter" section of Don's Transmitter/multiple-receivers design shown below. However, before looking at that design, there is one question which causes a good deal of discussion on the forums, namely, if the centre-tap of the L2 secondary coil is connected to ground, then should that earth-connection wire length be considered to be part of the quarter length of the L1 coil? To examine this possibility in depth, the following quote from Richard Quick's very clear explanation of resonance in his US patent 7,973,296 of 5th July 2011 is very helpful.

However, the simple answer is that for there to be exact resonance between two lengths of wire (whether or not part, or all of those lengths of wire happen to be wound into a coil), then one length needs to be **exactly** four times as long as the other, and ideally, half the diameter as well. At both ends of both lengths of wire, there needs to be a sudden change in wire diameter and Richard explains why this is. But, leaving that detailed explanation for now, we can use that knowledge to explain the above simplified system in more detail. Here is the circuit again:

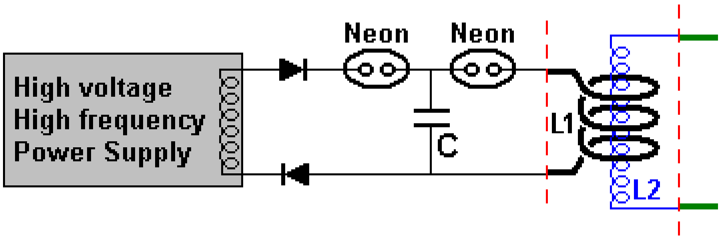

One very important point to note is that no earth connection is required and in spite of that, the performance shown on video is very impressive. While an earth connection can feed substantial power into the circuit, not needing one for the front end is an enormous advantage and potentially, opens the way for a truly portable device. Another very important point is the utter simplicity of the arrangement where only cheap, readily available components are used (and not many of those are needed). The resistors for extending the life of the neon bulbs are not shown, but they can be included if desired and the circuit operation is not altered significantly by having them there. If a higher spark voltage is wanted, then two or more neon bulbs can be used in series where these circuit diagrams show just one.

A point to note is that the lower diode is shown reversed when compared to the previous diagram. This is because the power supply shown is any generic power supply which drives a simple output coil which does not have a centre tap. The neon supply of the earlier diagram appears to have two separate outputs which will, presumably, be out of phase with each other as that is common practice for neon-sign driver modules. If you wish, the two diodes shown here could be replaced by a diode bridge of four high-voltage, high-speed diodes.

The wire lengths of L1 and L2 are measured very accurately from where the wire diameter changes suddenly, as indicated by the red dashed lines. The L2 wire length is exactly four times as long as the L1 wire length and the L2 wire diameter is half of the L1 wire diameter.

How long is the L1 wire? Well, how long would you like it to be? It can be whatever length you want and the radius of the L1 coil can be whatever you want it to be. The theory experts will say that the L1 coil should resonate at the frequency of the power feeding it. Well, good for them, I say, so please tell me what frequency that is. It is not going to be the frequency of the power supply as that will be changed by at least one of the neon bulbs. So, what frequency will the neon bulb produce? Not even the manufacturer could tell you that as there is quite a variation between individual bulbs which are supposedly identical.

Actually, it doesn't matter at all, because the L1 coil (and the L2 coil if you measure them accurately) has a resonant frequency all of its own and it will vibrate at that frequency no matter what the frequency feeding it happens to be. A coil resonates in very much the same way that a bell rings when it is struck. It doesn't matter how hard you strike the bell or how rapidly you strike it – the bell will ring at it's own natural frequency. So the L1 coil will resonate at it's own natural frequency no matter what rate the voltage spikes striking it arrive, and as the L2 coil has been carefully constructed to have exactly that same frequency, it will resonate in synchronisation with the L1 coil.

This means that the length of the wire for the L1 coil is the choice of the builder, but once that length is chosen it determines the length of the wire for the L2 coil as that is exactly four times as long, unless the builder decides to use an arrangement which has L2 wound in both the Clockwise and counter-clockwise directions, in which case, each half of the L2 coil will be four times the length of the wire in the L1 coil, like this:

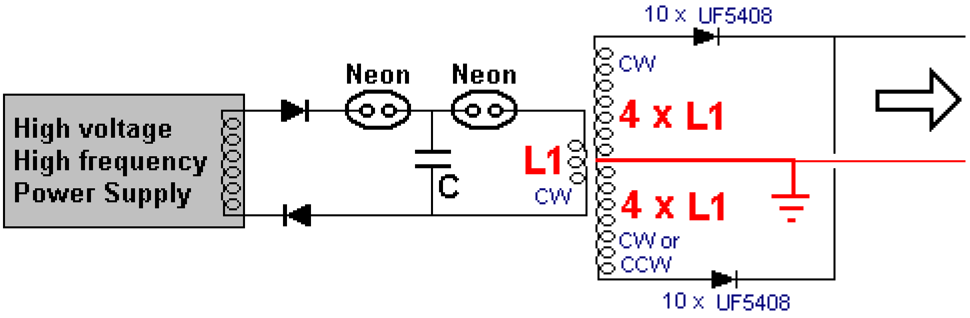
10 x UF5408
Neon
Neon
High voltage
High frequency
Power Supply
C
L1
CW
CW
4 x L1
4 x L1
CW or
CCW
10 x UF5408

Mind you, there is one other factor to be considered when deciding what the most convenient wire length for L1 might be, and that is the number of turns in the L1 coil. The larger the ratio between the turns in L1 and the turns in L2, the higher the voltage boost produced by the L1/L2 transformer, and remember that the length of L2 is fixed relative to the length of L1.

So, a possible circuit style might be:

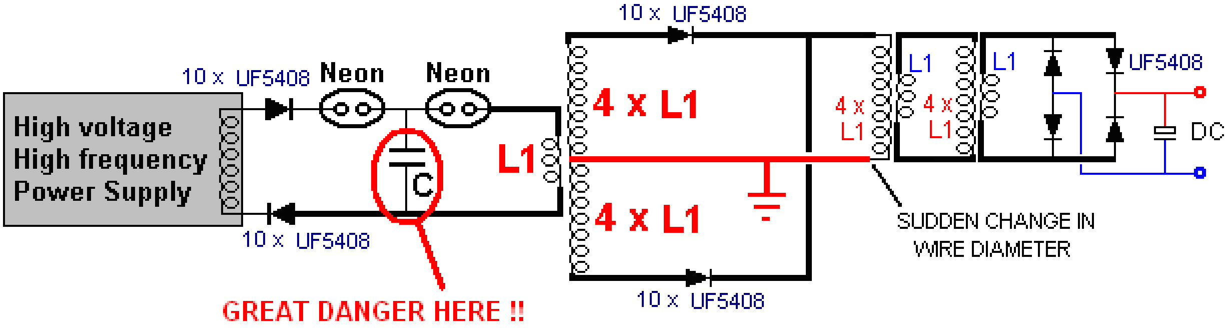
10 x UF5408
10 x UF5408
Neon
Neon
High voltage
High frequency
Power Supply
C
L1
10 x UF5408
GREAT DANGER HERE !!
4 x L1
4 x L1
10 x UF5408
4 x L1
4 x L1
L1
L1
UF5408
DC
SUDDEN CHANGE IN WIRE DIAMETER

There are some important points to remember. One is that there must be a sudden change of wire diameter at both ends of each L1 coil and at the ends of each L2 coil. If there isn't, then the connecting wire length will form part of the coil and if there is some change in diameter but not very much, then it is anybody's guess what the resonant wire length for that coil will be. There can be as many step-down isolation air-core L1/L2 transformers as desired and these do not need to be particularly large or expensive.

The builder of this circuit put it together in just a few minutes, using components which were to hand, including the microwave oven capacitor marked "C" in the diagrams above. That capacitor is isolated on both sides by the neon bulb spark gaps and so it will have no modifying effect on the resonant frequency of any of the coils in this circuit. But it is vital to understand that the energy stored in that capacitor can, and will, kill you instantly if you were to touch it, so let me stress once again that this information is **NOT** a recommendation that you actually build this circuit. The DC output from the circuit is intended to power a standard inverter, which in turn, would be perfectly capable of powering the high voltage, high frequency input oscillator.

One final point is that as demonstrated in the video, the oscillating magnetic field produced by the L1 coil can power several identical L2 coils, giving several additional power outputs for no increase in input power, because the coupling is magnetic and not inductive as mentioned earlier in this chapter. Please notice that neither the L1 coil nor the L2 coil has a capacitor connected across it, so resonance is due solely to wire length and no expensive high-voltage capacitors are needed to get every L1/L2 coil pair resonating together. One possible arrangement might be like this:

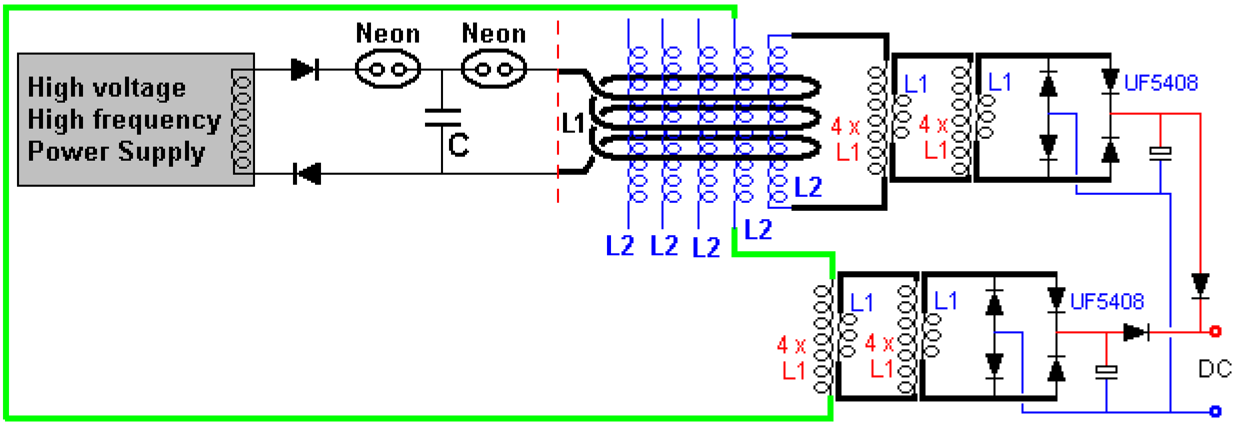
Neon
Neon
High voltage
High frequency
Power Supply
C
L1
L2
L2
L2
L2
L2
4 x L1
4 x L1
L1
L1
UF5408
4 x L1
4 x L1
L1
L1
UF5408
DC

Where two of the L2 coils are shown connected together to give increased output power. This arrangement uses low-voltage inexpensive components for the output stages and there is no obvious limit to the amount of output power which could be provided. As the circuit operates at high frequency throughout, there is no particular need for additional L2 coils to be placed physically inside the L1 coil:

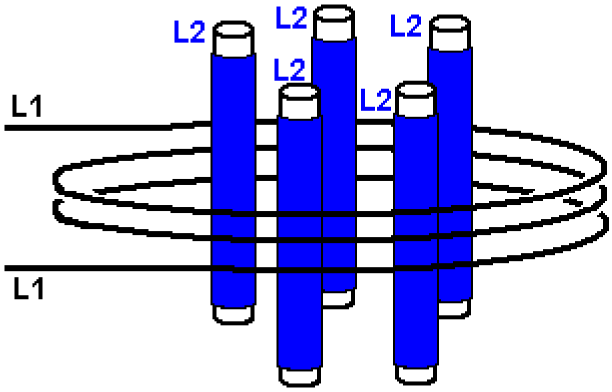

However, there can be an advantage to this arrangement in that the wire length of the L1 coil is greater, which in turn makes the wire length of each L2 coil greater (being four times longer). This gives greater flexibility when planning the turns ratio of the L1/L2 transformer. The voltage step-up or step-down of that transformer happens to be in the ratio of the turns, in spite of the fact that this is not inductive coupling and so standard transformer technology does not apply.

When you choose the number of turns and coil diameter for L1, that also gives the length of the L2 wire. In order to get the desired output voltage, if perhaps, the step-down ratio is needed to be an amount of 46:1, then you need 46 times the number of L1 turns on the L2 coil. That means that you know both the wire length and number of turns wanted in the L2 coil. But, as each turn will have a length of 3.14159 times the diameter, it follows then that the wanted diameter is the wire length per turn, divided by 3.14159. The wire sits on top of the tube on which it is wound and so has a greater diameter by one wire thickness, so the calculated tube diameter needs to be reduced by one wire diameter. For example, if the length per turn is 162 mm and the wire diameter 0.8 mm, then the tube diameter would be 162 / 3.14159 – 0.8 which is 50.766 mm (just over two inches).

Now for Richard's explanation of the resonant frequency of any length of wire:

"Quarter-Wave" Resonance; Standing Electromagnetic Waves"

One of the two main types is electrical resonance is referred to here as quarter-wave resonance. This type of resonance depends almost entirely on the length of a wire element For reasons described below, if a segment or length of wire is one quarter as long as the "voltage waves" which are travelling through the wire, then a set of "reflected" waves will be added to the emitted waves, in a synchronised alignment which creates stronger "superimposed waves".

Accordingly, an understanding of the "quarter-wave" phenomenon will help a reader understand how a straightforward and easily-controlled factor (i.e., the length of a wire ribbon which will be used to form a spiral coil) can help create a "quarter-wave" resonant response, which will create the types of electromagnetic pulses and fields referred to as "standing waves".

The speed at which a voltage impulse is transmitted through a metal wire is extremely fast. It is essentially the same as the speed of light, which travels 300 million meters (186,000 miles) in a single second (that distance would circle the earth more than 7 times).

If wavelength (in meters) is multiplied by frequency (cycles per second), the result will be the speed of light, 300 million meters/second. Therefore, the wavelength of an "alternating current" (AC) voltage, at some particular frequency, will be the speed of light, divided by which frequency.

Therefore, using simple division, if an alternating voltage operates at a frequency of 1 megahertz (MHz), which is a million cycles per second, then the "wavelength" at that frequency will be 300 meters. If the frequency halves become 500 kilohertz, the wavelength becomes twice as long (600 meters); and, if the frequency were to increase to 2 megahertz, the wavelength drops to 150 meters.

It should be noted which the term "cycles" is what scientists call "a dimensionless unit", which drops out and becomes silent when other physical terms are multiplied or divided.

At AC frequencies of 10 kilohertz or greater, the common references to "alternating current" (AC) voltage begin using a different term, which is "radio-frequency" (RF) voltage. Accordingly, RF voltage is a form (or subset) of AC voltage, which operates at frequencies higher than 10 kilohertz. RF power generators are readily available, and are sold by numerous companies which can be easily located by an Internet search, using the term "RF power generator". For example, Hotek Technologies Inc. (hotektech.com) sells two RF power generators, called the AG 1024 and AG 1012 models, which can provide output power at frequencies ranging from 20 kHz to 1 MHz; the 1012 model has a power output of 1000 watts, while the 1024 model has a power output of 2000 watts. The output frequency of any such RF power supply can be adjusted and "tuned" across the entire range of operating frequencies, merely by turning knobs or manipulating other controls in a power supply of this type.

In a wire having a fixed and unchanging length, the easiest way to create a "standing wave" is to adjust the RF frequency emitted by a power supply with an adjustable frequency, until the "tuned" frequency creates a wavelength which is 4 times as long as the wire. This principle is well-known to physicists, and it is commonly referred to as "quarter-wave" behaviour, since the length of the wire segment must be one quarter as long as the wavelength. Since it is important to this invention, the principles behind it are illustrated in a series of drawings provided in **Fig.1** to **Fig.4**, all of which are well-known prior art.

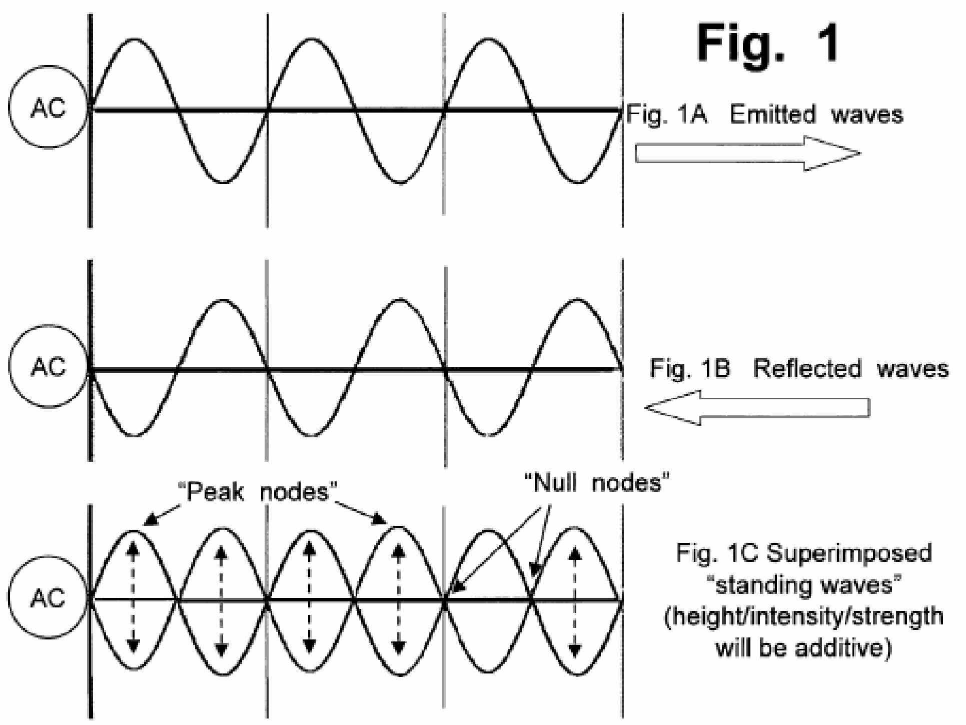
Fig. 1
AC
Fig. 1A Emitted waves
AC
Fig. 1B Reflected waves
"Peak nodes"
"Null nodes"
AC
Fig. 1C Superimposed "standing waves" (height/intensity/strength will be additive)

Fig.1A indicates an idealized wavelength of an alternating voltage, depicted by a sine wave which is being sent from an AC power supply (shown by a circle at the left end of a horizontal straight wire) into the "input" end of the wire. The voltage waves travel through the wire towards the right, as indicated by the block arrow in **Fig.1A**. When the waves reach the end of the wire, they cannot leave the wire (at least, not in a simplified and "ideal" system, which is being assumed and used here to explain the principle of how a simple straight wire can create a standing wave). Therefore, the voltage wave will effectively "bounce" or "reflect" back from the tip of the wire, and the "reflected wave" will begin travelling back through the wire, going in the opposite direction, as indicated by the left-pointing block arrow in **Fig.1B**.

Because of the laws of conservation of energy, the reflection and "return travel" of these types of waves, when they bounce off the tip of a wire, is actually quite good, and rather efficient, as discussed below, provided which the wire tip does not emit sparks, arc discharges, or other forms of "escaping" electrical energy.

Accordingly, **Fig.1A** depicts a set of "emitted waves" travelling towards the right, while **Fig.1B** depicts an idealised set of "reflected waves" travelling toward the left along the same wire.

Fig.1C illustrates what happens when both sets of waves (emitted and reflected) are superimposed on each other. Since the two sets of waves are travelling at exactly the same speed, and since they have exactly the same wavelength, they will create a "standing wave" pattern when they are added together. As can be visualised from **Fig.1C**, there will be a set of locations, along the length of the wire, which can be referred to as "peak nodes", where the AC voltage reaches it's maximum.

At a location halfway between a pair of adjacent "peak nodes", there will be a spot which can be called a "null node", a "zero node", a trough or valley node, or similar terms. At each "null node" location, the AC voltage will appear to be not fluctuating at all. Those are the sites, along the length of the wire, where each "positive" hump (created by a sine wave travelling toward the right) will be counter-balanced and offset by a "negative hump" with exactly the same height, travelling at an identical speed toward the left.

As a result, this type of response within a wire creates a "standing wave". If the instantaneous voltage is measured at a "null node", it would appear that nothing is happening, in terms of fluctuating voltage. Furthermore, the "null node" will not be moving, along the length of the wire; instead, it will appear to be standing still.

This can be demonstrated, in a coil, by using a "grounded lead" to test for voltages along the length of a coil. If a "grounded lead" coupled to a volt meter is used to touch the surfaces of a series of strands in a non-insulated coil (such as a coil made of thin copper tubing, wrapped around a plastic cylindrical shape, as used in the types of large transformers used by hobbyists to create "Tesla coils" which will emit large and visually impressive electrical arcs), the "test lead" will detect no apparent voltage at a null node, which will occur at some particular strand in the coil. At a different strand of the coil, the "test lead" will detect an alternating voltage which has twice the strength and intensity of the voltage being emitted by the power supply.

If voltage is measured at a "peak node", the voltage will be doing something which can be called, using vernacular or laymen's terms, "the full-tilt boogie". The AC voltage levels will be moving back and forth, between: (i) a very high and intense positive voltage, to (ii) an equally intense negative voltage. This is indicated by the "bubble" shapes shown along the wire in **Fig.1C**.

The "bubbles" which are shown in **Fig.1C** can help someone understand how standing waves are created, and how they act in a synchronised manner. However, which drawing fails to show another result which is very important in what actually happens in a standing wave. For purposes of description and analysis at this introductory level, the system can be assumed to be "ideal", which implies a perfect "mirror-image" reflection of each wave from the right end of the wire. An "ideal" system also implies that no reflections occur at the left hand end of the wire where the power supply is located, and all "reflected" wave activity simply ceases. In real circuits and wires of this type, second and third order reflections do in fact occur, and they are used to further increase the strength and power output of these types of systems; however, those additional factors and "harmonics" should be ignored until after the basic principles of this type of system have been grasped and understood.

In an ideal system, when the reflected waves (which are travelling toward the left, in the wire segments illustrated in **Fig.1**) are "superimposed" on the emitted waves (travelling toward the right), the "peak" positive voltage which will be instantaneously reached, at the highest point of each "bubble" shown in **Fig.1C**, will occur when the positive peak of an emitted wave crosses a mirror-image positive peak of a reflected wave, travelling in the opposite direction. Accordingly, when those two "positive peak" values are added to each other, the instantaneous positive peak voltage which will occur, in the wire, will actually be twice as intense as the "positive peak" voltage being emitted by the AC power supply.

An instant later, at that exact point on that segment of wire, a negative peak voltage will be created, which will be the sum of (i) the negative peak voltage emitted by the power supply, and (ii) the negative peak voltage of a reflected wave also will pass through, travelling toward the left. At which instant, when those two negative peak voltages are added to each other, the instantaneous negative voltage which will occur, in the wire, will be twice as intense as the "negative peak" voltage being generated by the AC power supply.

A more accurate and representative visual depiction of a "standing wave" in a wire would actually show the heights of the peaks as being twice as tall as the peaks of the emitted voltage waves, and the reflected voltage waves. However, which depiction might confuse people, so it usually is not shown in drawings of "standing waves".

Accordingly, the instantaneous response in the wire, at a location halfway between two "null nodes", is doing something which can fairly and properly be called "the full-tilt double double boogie". The "double double" phrase (note which it contains not just one but two "doubles") was added to that phrase, for two reasons:

(i) To emphasise the fact that each and every voltage peak (maximum positive, and maximum negative) will be twice as strong, and twice as intense, as the maximum positive and negative peak voltages emitted by the power supply; and,

(ii) to point out that the frequency of the superimposed "bubbles", shown in **Fig.1C**, is actually twice as fast as the frequency of the AC cycle which is emitted by the power supply, as discussed below.

The "twice the intensity" result is directly comparable to what an observer will see, if a large mirror is placed behind a light bulb in an otherwise dark room. The mirror effectively keeps the room dark, everywhere behind the mirror, so there is no "magical doubling" of the light in the room; which would violate the basic law of conservation of energy. Instead, what the mirror does is to shift light away from the backside of the mirror, and keep that light energy on the reflective side of the mirror. Anyone standing in front of the mirror will see two apparent light bulbs. Both of those light bulbs (the original bulb, and the reflected image) will have the same brightness (if the mirror is perfect). Therefore, the mirror will double the intensity of the light energy reaching the observer.

That same effect, in a circuit, will happen if the end of a wire acts like a mirror. If a wire does not have any components which will cause it to become an active "emission source" (which is the behaviour of transmission antennas and certain other components), in a way which efficiently releases voltage-created energy into the atmosphere, then the basic rules which require conservation of energy will prevent that energy from simply disappearing and ceasing to exist. As a result, even if the end of a wire is not designed to be a perfect reflector, a large portion of the voltage wave will indeed reflect off the wire tip, and travel back through the same wire, in a "second pass".

To understand adequately, the type and amount of "wave reflection" which occurs at a wire tip, consider what happens if a light bulb is shining in a room which has shiny, glossy white paint on all the walls and ceilings; then, consider how it would look if the same light bulb were located in a room with all of the walls and ceilings painted "matt black". The total amount of light which would be available, to carry out a task such as reading a newspaper, clearly would be much greater in the white room, because light reflects off white paint, even though white paint does not even begin to approach the type of "reflection quality or clarity" which a mirror creates. The difference in what happens, when light intensity in a room painted matt black is compared to a room painted a glossy white, does not arise from the presence or absence of "reflection quality or clarity"; instead, it is governed by the laws of conservation of energy. When light shines on to a surface which is painted matt black, the light energy is absorbed by the paint, and it literally warms the paint up. In contrast to that, glossy white paint will not absorb light energy, so it reflects the light back out, for a "second pass" through the air which fills a room.

Because of the laws of conservation of energy, and without depending on any "quality of reflectance" characteristic of wire tips, electrical energy cannot simply disappear, when it reaches the end of a wire. Instead, there are only two things which can happen to that energy:

(i) the electrical energy can be emitted into the surroundings, such as by emitting sparks, arcs, or radio-frequency signals which will carry energy; or

(ii) if the energy is not emitted by the tip of the wire, then, by simple necessity and because of the basic law of conservation of energy, it must be reflected back into the wire, and it will be forced to travel back through the wire again.

If a wire has a long and tapered tip, then the reflected wave might become somewhat diffused, and it might lose some portion of the "clarity" of the wave. However, since wavelengths in the frequencies of interest here are hundreds of meters long, the type of tip created by a conventional wire cutter will not create any significant diffusion, in a reflected wave. And, unlike the white-painted walls of a room, there is not a large area which is available, at the tip of a wire, which can create scatter, spread, or diffusion. As a result, the tip of a wire will be a relatively efficient mirror-type reflector, when an AC voltage is "pumped" into one end of the wire.

The second factor mentioned above, when the "double-double" boogie phrase was mentioned, relates to a doubling of the frequency of a standing wave. When a standing wave is created in a wire by reflection of an emitted AC voltage wave, the frequency of the standing wave is, quite literally, double the frequency of the emitted wave.

This can be seen, visually, by noting that in the emitted AC voltage, shown in **Fig.1A**, a single complete wavelength contains both a "positive hump" and a "negative hump". Accordingly, three complete sine waves, divided into three segments by the imaginary vertical lines, are shown in **Fig.1A**.

By contrast, each and every "bubble" shown in **Fig.1C** depicts a complete and total "wavelength", in a standing wave. Six of those standing wave "bubbles" fit into exactly the same length of wire which holds only 3 emitted wavelengths from the power supply.

The "frequency doubling" effect of standing waves is important, because AC systems can convey and release energy in a manner which increases, as the frequency of the AC voltage supply increases. To some extent, this is analogous to saying that, if a motor can be run at twice the speed (while still generating the same torque), then the work output of that motor can be twice as great, at the higher speed. That analogy is not entirely accurate, since work output from an electric device which uses AC power depends on "area of the curve" functions which occur when sine waves are involved. Nevertheless, as a general principle, if the frequency of the voltage peaks increases, then power output will also increase, in many types of electric circuit components.

In the three panels of **Fig.1**, the wire length is three times as long as the wavelength of the voltage from the power supply. However, to create standing waves, a wire length does not need to be any particular multiple of the wavelength of an AC voltage. As can be seen by considering **Fig.1C**, the same types of "bubbles" would be created: (i) if the wire length were exactly twice as long as the wavelength; or, (ii) if the wire length were the same length as the wavelength.

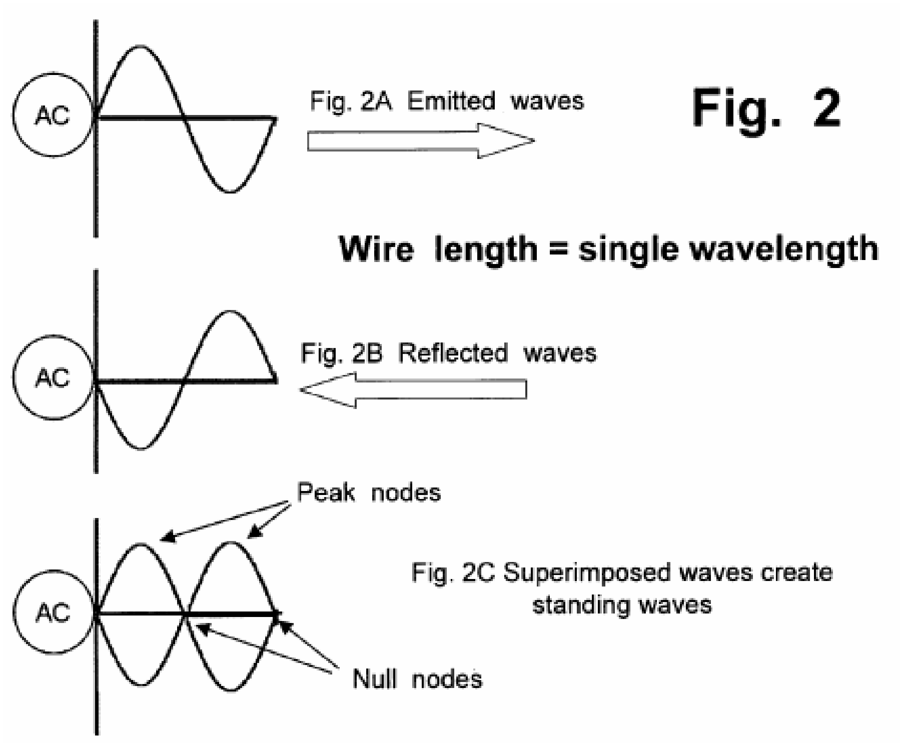

Fig. 2

AC
Fig. 2A Emitted waves
Wire length = single wavelength
AC
Fig. 2B Reflected waves
Peak nodes
AC
Fig. 2C Superimposed waves create standing waves
Null nodes

Accordingly, **Fig.2** (which includes **Fig.2A** showing an emitted wave, **Fig.2B** showing a reflected wave, and **Fig.2C** showing the superimposed “bubbles”) shows what happens in a wire segment which has a length which is equal to a single wavelength from an AC voltage at a fixed frequency. A resonant standing wave will be formed, with a frequency which is double the frequency of the input AC voltage. which same result will apply, in a wire having any length which is an exact (integer) multiple (such as 1x, 2x, 3x, etc.) of the wavelength of the AC voltage being pushed (or forced, driven, pumped, etc.) into the wire segment.

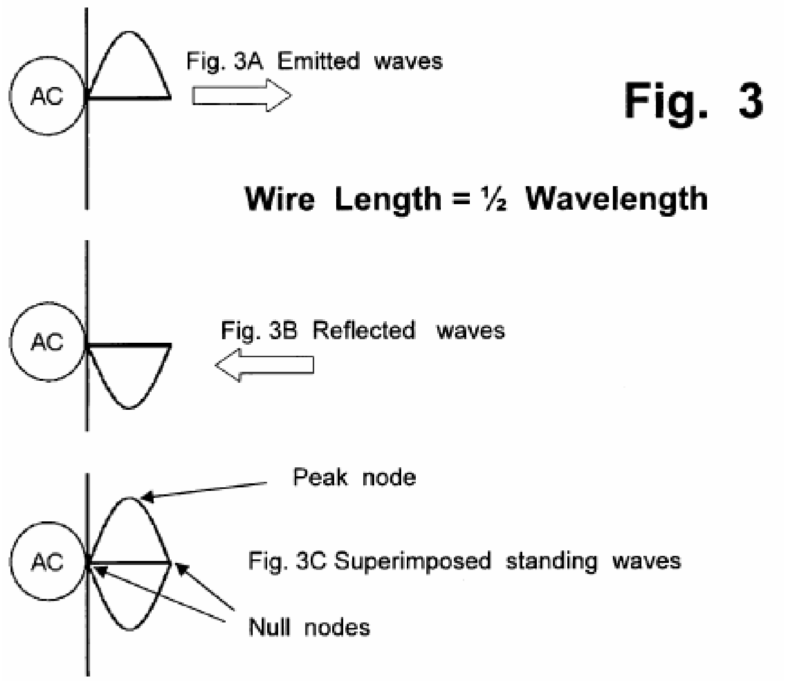

Fig. 3

AC
Fig. 3A Emitted waves
Wire Length = ½ Wavelength
AC
Fig. 3B Reflected waves
Peak node
AC
Fig. 3C Superimposed standing waves
Null nodes

Moving to still shorter wires, the same principle also applies to any wire with a length equal to one half of an AC voltage wavelength. As shown in **Fig.3** (which includes **Fig. 3A** showing an emitted wave, **Fig. 3B** showing a reflected wave, and **Fig. 3C** showing the superimposed “bubbles”), if the wire length is one half of the wavelength, a natural and resonant standing wave will still form, with a frequency which is double the frequency of the input AC voltage.

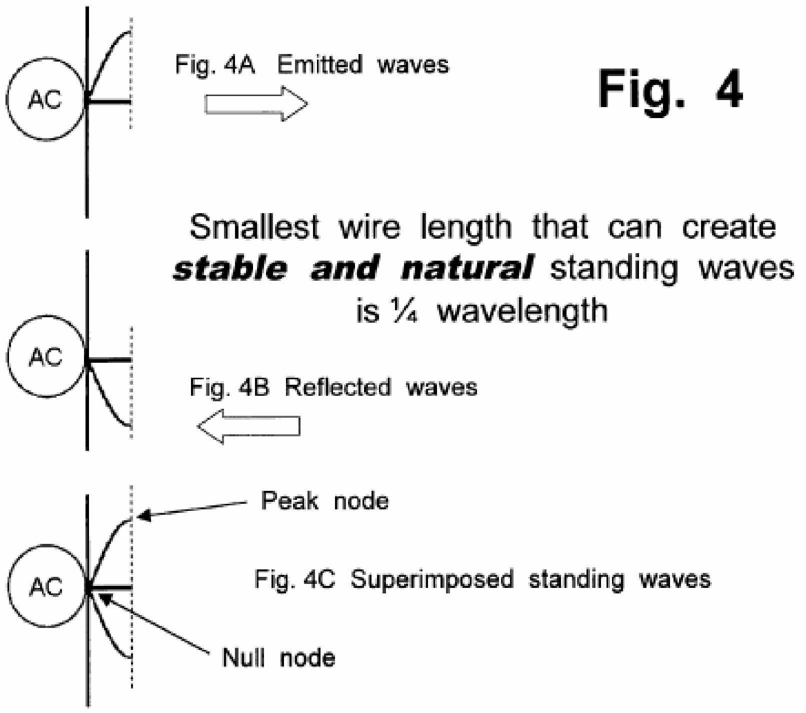
AC
Fig. 4A Emitted waves
Fig. 4
Smallest wire length that can create
stable and natural standing waves
is ¼ wavelength
AC
Fig. 4B Reflected waves
Peak node
AC
Fig. 4C Superimposed standing waves
Null node

Finally, moving to a still shorter wire, the same principle also applies to any wire which has a length equal to one quarter of an AC voltage wavelength, as shown in **Fig.4A**, **Fig.4B**, and **Fig.4C** Even though it does not stretch across or cover a complete "bubble", the standing wave shown in **Fig.4C** is nevertheless a stable, natural, and resonant "standing wave", with a frequency which is exactly twice the frequency of the input AC voltage.

It is possible to create partially stable and semi-resonant responses, using one eighth, one sixteenth, or shorter lengths of wire, by using additional devices which can remove electrical power from the system, or which can generate effects which are usually called "harmonics". However, those are not the types of natural and stable responses which can be created by a simple, basic system consisting of nothing more than: (i) a wire having a fixed length and a "reflective" tip; and (ii) an AC power source with a frequency which can be "tuned" until it creates a resonant response in any wire segment having a suitable length.

Therefore, since quarter-wave wire lengths are the shortest lengths which can create natural and stable standing waves, the conventional term which is commonly used, to describe what happens when a wire creates a resonant standing-wave response, is a "quarter-wave" response.

In some devices, telescoping components (or other elements which can alter the effective length of a wire-type element) can be used to alter the ability of the element to respond to a fixed wavelength. Many types of antennas use this approach, if they need to process signals which are being transmitted on fixed and known frequencies. However, those examples are not relevant to spiral coil reactors, which will use an approach which involves tuning and adjusting the frequency of the voltage which is being supplied to a reactor, until a resonant response is observed in coils with fixed and unchanging lengths.

It should also be noted that certain types of "tuning" elements (such as capacitors, which can have either fixed or adjustable capacitance levels) can also be coupled electrically to a wire, in a manner which "emulates" adding more length to that wire. This approach can be used to alter (or increase the range of) the frequencies to which a wire circuit will respond resonantly.

So, if we have resonant standing-wave voltages in our L2 coil and some of that signal passes through the wire connecting one end of the coil to the earth, then what will happen? The best way to check it is to test the way which a prototype behaves, however, if I may express an opinion, I would suggest that the signal passing down the earth wire will be absorbed when it reaches the earth and that will prevent the signal being reflected back to the L2 coil to upset it's operation.

Another device of Don's is particularly attractive because almost no home-construction is needed, all of the components being available commercially, and the output power being adaptable to any level which you want. Don particularly likes this circuit because it demonstrates COP>1 so neatly and he remarks that the central transmitter Tesla Coil on its own is sufficient to power a household.

Another device of Don's is particularly attractive in that almost no home-construction is needed, all of the components being available commercially, and the output power being adaptable to any level which you want. Don particularly likes this circuit because it demonstrates COP>1 so neatly and he remarks that the central transmitter Tesla Coil on its own is sufficient to power a household.

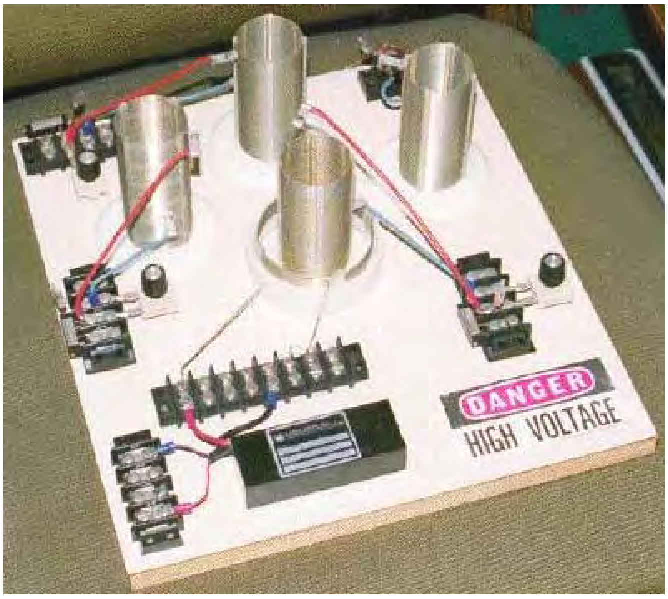

The coil in the centre of the board is a power transmitter made from a Tesla Coil constructed from two Barker & Williamson ready-made coils. Three more of the inner coil are also used as power receivers. The outer, larger diameter coil is a few turns taken from one of their standard coils and organised so that the coil wire length is one quarter of the coil wire length of the inner coil ("L2").

As before, a commercial neon-tube driver module is used to power the "L1" outer coil with high voltage and high frequency. It should be understood that as power is drawn from the local environment each time the power driving the transmitter coil "L1" cycles, that the power available is very much higher at higher frequencies. The power at mains frequency of less than 100 Hz is far, far less than the power available at 35,000 Hz, so if faced with the choice of buying a 25 kHz neon-tube driver module or a 35 kHz module, then the 35 kHz module is likely to give a much better output power at every voltage level.

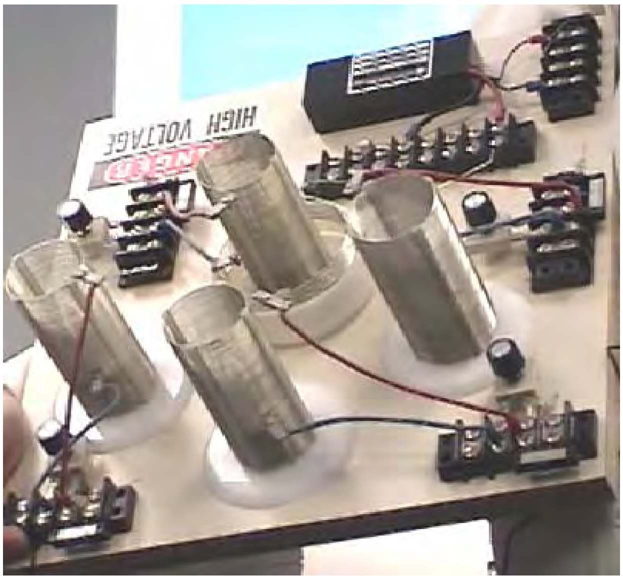

The "L1" short outer coil is held in a raised position by the section of white plastic pipe in order to position it correctly relative to the smaller diameter "L2" secondary coil. Again, it appears to have five turns:

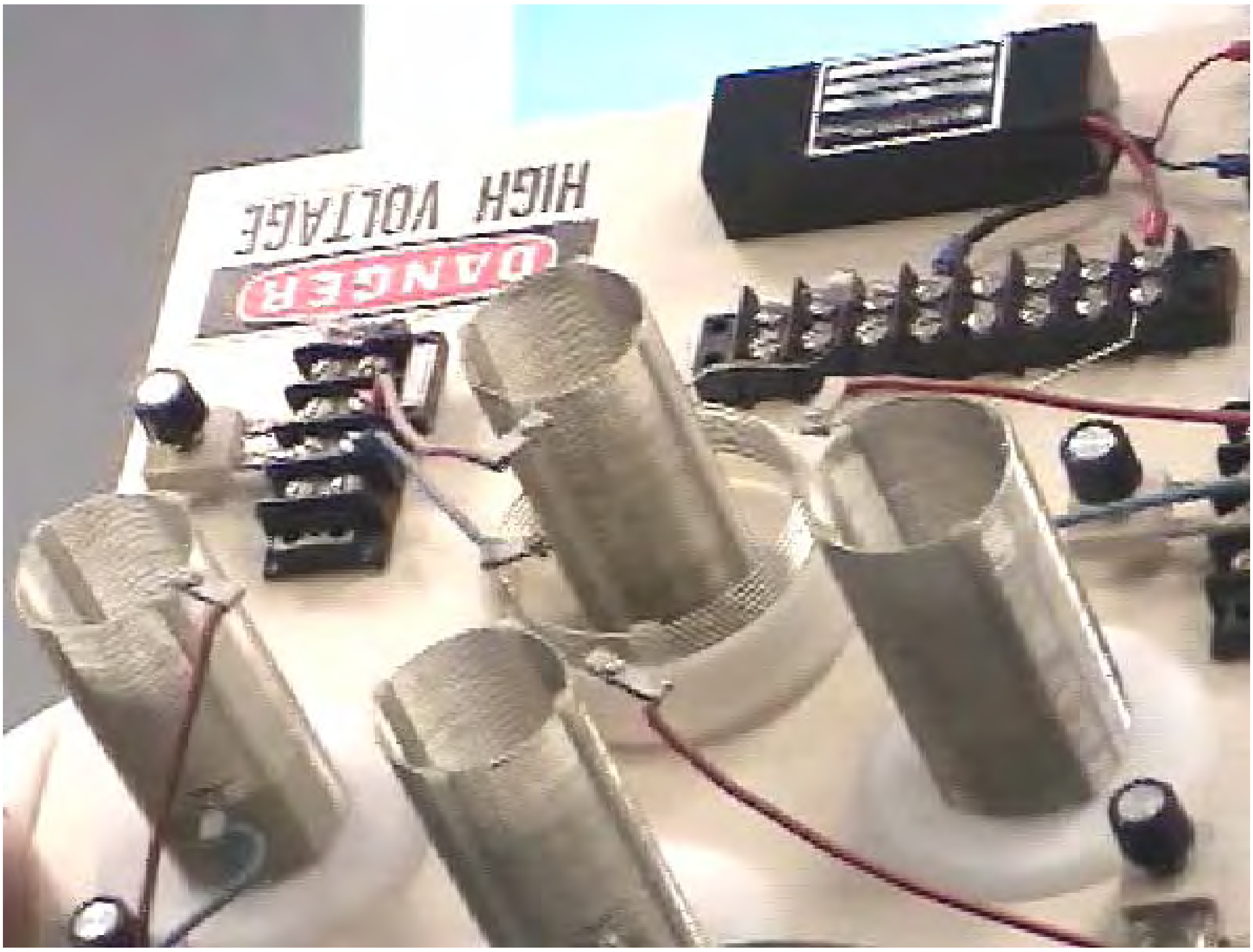

The secondary coils are constructed using Barker & Williamson's normal method of using slotted strips to hold the tinned, solid copper wire turns in place.

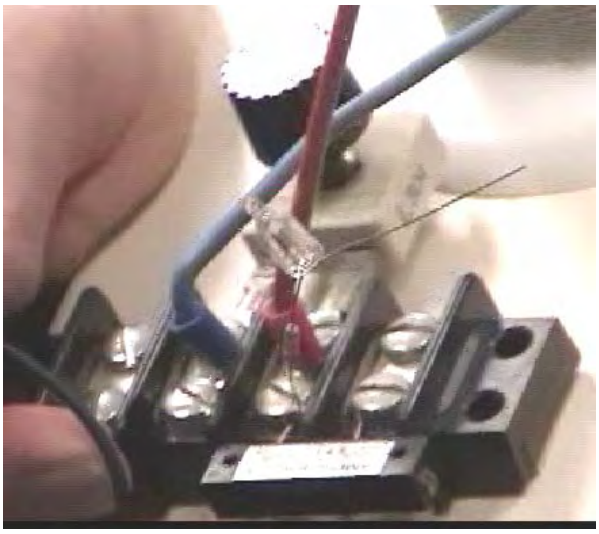

As there are very slight differences in the manufactured coils, each one is tuned to the exact transmitter frequency and a miniature neon is used to show when the tuning has been set correctly.

The key feature of this device is the fact that any number of receiver coils can be placed near the transmitter and each will receive a full electrical pick up from the local environment, without altering the power needed to drive the Tesla Coil transmitter - more and more output without increasing the input power - unlimited COP values, all of which are over 1. The extra power is flowing in from the local environment where there is almost unlimited amounts of excess energy and that inflow is caused by the rapidly vibrating magnetic field generated by the central Tesla Coil. While the additional coils appear to just be scattered around the base board, this is not the case. The YouTube video http://www.youtube.com/watch?v=TiNEHZRm4z4&feature=related demonstrates that the pick-up of these coils is affected to a major degree by the distance from the radiating magnetic field. This is to do with the wavelength of the signal driving the Tesla Coil, so the coils shown above are all positioned at exactly the same distance from the Tesla Coil. You still can have as many pick-up coils as you want, but they will be mounted in rings around the Tesla Coil and the coils in each ring will be at the same distance from the Tesla Coil in the centre.

Each of the pick up coils act exactly the same as the "L2" secondary coil of the Tesla Coil transmitter, each picking up the same level of power. Just as with the actual "L2" coil, each will need an output circuit arrangement as described for the previous device. Presumably, the coil outputs could be connected in parallel to increase the output amperage, as they are all resonating at the same frequency and in phase with each other. Each will have its own separate output circuit with a step-down isolation transformer and frequency adjustment as before. If any output is to be a rectified DC output, then no frequency adjustment is needed, just rectifier diodes and a smoothing capacitor following the step-down transformer which will need to be an air core or ferrite core type due to the high frequency. High voltage capacitors are very expensive. The http://www.richieburnett.co.uk/parts.html web site shows various ways of making your own high-voltage capacitors and the advantages and disadvantages of each type.

There are two practical points which need to be mentioned. Firstly, as the Don Smith devices shown above feed radio frequency waveforms to coils which transmit those signals, it may be necessary to enclose the device in an earthed metal container in order not to transmit illegal radio signals. Secondly, as it can be difficult to obtain high-voltage high-current diodes, they can be constructed from several lower power diodes. To increase the voltage rating, diodes can be wired in a chain. Suitable diodes are available as repair items for microwave ovens. These typically have about 4,000 volt ratings and can carry a good level of current. As there will be minor manufacturing differences in the diodes, it is good practice to connect a high value resistor (in the 1 to 10 megohm range) across each diode as that ensures that there is a roughly equal voltage drop across each of the diodes:

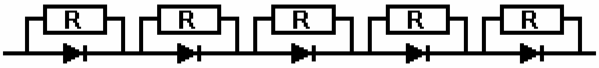

If the diode rating of these diodes were 4 amps at 4,000 volts, then the chain of five could handle 4 amps at 20,000 volts. The current capacity can be increased by connecting two or more chains in parallel.

Two DVDs containing video recordings of Don Smith's lectures are available for purchase via the website link: https://secure.netsolhost.com/nuenergy.org/product_catalog.htm

Various questions from readers indicate that the operation of AC circuits is not really understood, so electronics experts can skip this next section.

AC Circuits. This is a lightweight introduction to Alternating Current circuits and pulsed DC circuits for people who have not read Chapter 12 which is an electronics tutorial. Let me say again, that I am mainly self-taught, and so this is just a general introduction based on my present understanding.

Alternating Current, generally called “AC” is called that because the voltage of this type of power supply is not a constant value. A car battery, for instance, is DC and has a fairly constant voltage usually about 12.8 volts when in it’s fully charged state. If you connect a voltmeter across a car battery and watch it, the voltage reading will not change. Minute after minute it says exactly the same because it is a DC source.

If you connect an AC voltmeter across an AC power supply, it too will give a steady reading, but it is telling a lie. The voltage is changing all the time in spite of that steady meter reading. What the meter is doing is **assuming** that the AC waveform is a sine wave like this:

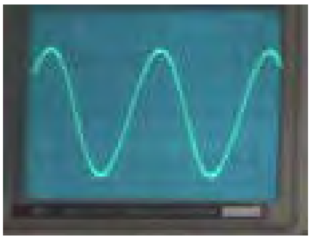

and based on that assumption, it displays a voltage reading which is called the “Root Mean Square” or “RMS” value. The main difficulty with a sine wave is that the voltage is below zero volts for exactly the same length of time as it is above zero volts, so if you average it, the result is zero volts, which is not a satisfactory result because you can get a shock from it and so it can't be zero volts, no matter what the arithmetical average is.

To get over this problem, the voltage is measured thousands of times per second and the results squared (that is, the value is multiplied by itself) and then those values are averaged. This has the advantage that when the voltage is say, minus 10 volts and you square it, the answer is plus 100 volts. In fact, all of the answers will be positive, which means that you can add them together, average them and get a sensible result. However, you end up with a value which is far too high because you squared every measurement, and so you need to take the square root of that average (or "mean") value, and that is where the fancy sounding "Root Mean Square" name comes from – you are taking the (square) root of the (average or) mean value of the squared measurements.

With a sine wave like this, the voltage peaks are 41.4% higher than the RMS value which everyone talks about. This means that if you feed 100 volts AC through a rectifier bridge of four diodes and feed it into a capacitor the capacitor voltage will **not** be 100 volts DC but instead it will be 141.4 volts DC and you need to remember that when choosing the voltage rating of the capacitor. In that instance I would suggest a capacitor which is made to operate with voltages up to 200 volts.

You probably already knew all of that, but it may not have occurred to you that if you use a standard AC voltmeter on a waveform which is **not** a sine wave, that the reading on the meter is most unlikely to be correct or anywhere near correct. So, please don't merrily connect an AC voltmeter across a circuit which is producing sharp voltage spikes like, for instance, one of John Bedini's battery pulsing circuits, and think that the meter reading means anything (other than meaning that you don't understand what you are doing).

You will, hopefully, have learned that power in watts is determined by multiplying the current in amps by the voltage in volts. For example, 10 amps of current flowing out of a 12 volt power supply, represents 120 watts of power. Unfortunately, that only holds true for circuits which are operating on DC, or AC circuits which have only resistors in them. The situation changes for AC circuits which have non-resistive components in them.

The circuits of this type which you are likely to come across are circuits which have coils in them, and you need to think about what you are doing when you deal with these types of circuit. For example, consider this circuit:

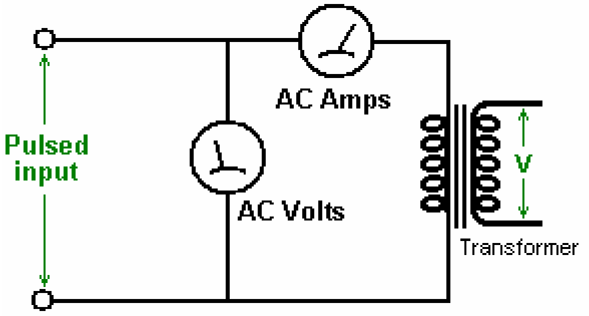

This is the output section of a prototype which you have just built. The input to the prototype is DC and measures at 12 volts, 2 amps (which is 24 watts). Your AC voltmeter on the output reads 15 volts and your AC ammeter reads 2.5 amps and you are delighted because 15 x 2.5 = 37.5 which looks much bigger than the 24 watts of input power. **But**, just before you go rushing off to announce on YouTube that you have made a prototype with COP = 1.56 or 156% efficient, you need to consider the real facts.

This is an AC circuit and unless your prototype is producing a perfect sine wave, then the AC voltmeter reading will be meaningless. It is just possible that your AC ammeter is one of the few types that can accurately measure the current no matter what sort of waveform is fed to it, but it is distinctly possible that it will be a digital meter which assesses current by measuring the AC voltage across a resistor in series with the output, and if that is the case, it will probably be assuming a sine wave. The odds are that both readings are wrong, but let's take the case where we have great meters which are reading the values perfectly correctly. Then the output will be 37.5 watts, won't it? Well, actually, no it won't. The reason for this is that the circuit is feeding the transformer winding which is a coil and coils don't work like that.

The problem is that, unlike a resistor, when you apply a voltage across a coil the coil starts absorbing energy and feeding it into the magnetic field around the coil, so there is a delay before the current reaches it's maximum value. With DC, this generally doesn't matter very much, but with AC where the voltage is changing continuously, it matters a great deal. The situation can be as shown in this graph of both voltage and current:

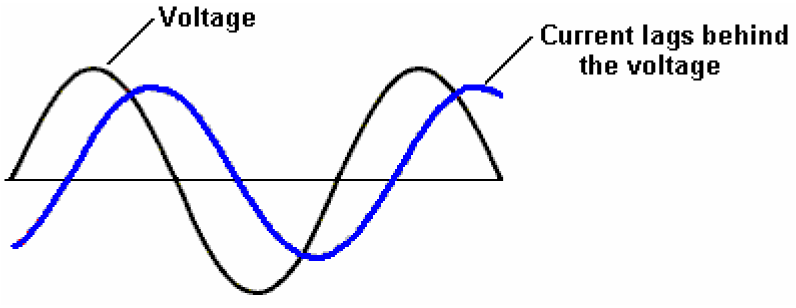
Voltage
Current lags behind the voltage

At first, this does not look like any great problem, but it has a very significant effect on the actual power in watts. To get the 37.5 watts output which we were talking about earlier, we multiplied the average voltage level by the average current level. But these two values do not occur at the same time and that has a major effect.

As this can be a little difficult to see, let's take the peak values rather than the averages as they are easier to see. Let's say that in our example graph that the voltage peak is 10 volts and the current peak is 3 amps. If this were DC we would multiply them together and say that the power was 30 watts. But with AC, this does not work due to the timing difference:

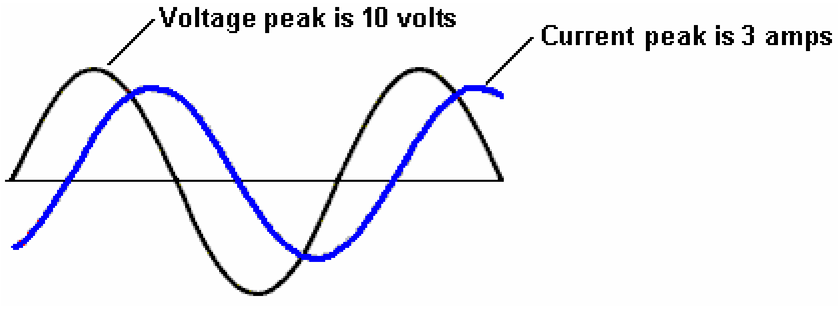
Voltage peak is 10 volts
Current peak is 3 amps

When the voltage is peaking, the current is nowhere near it's peak value of 3 amps:

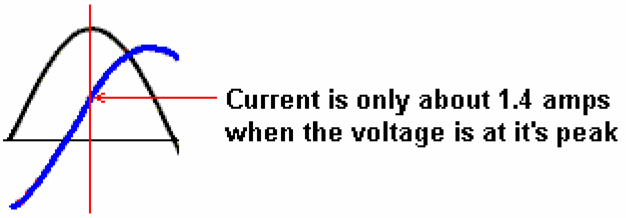
Current is only about 1.4 amps
when the voltage is at it's peak

As a result of this, instead of getting our expected peak power at the top of the voltage peak, the actual power in watts is very much lower – less than half of what we were expecting. Not so good, but it gets worse when you look at the situation more closely. Take a look at what the voltage is when the current crosses the zero line, that is, when the current is zero. The output power is zero when the current is zero but this occurs when the voltage is at a very high value:

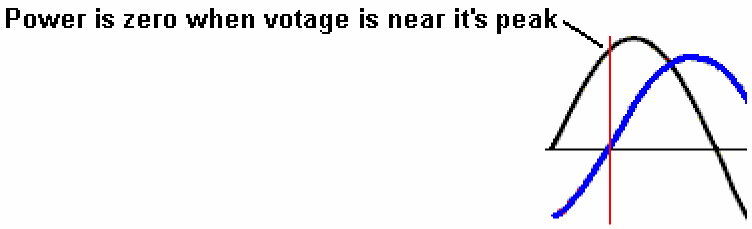
Power is zero when votage is near it's peak

The same goes for when the voltage is zero. When the voltage is zero, then the power is also zero, and you will notice that this occurs when the current is at a high value:

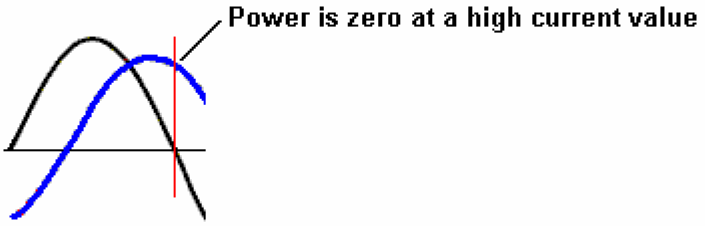

The power is **not** the average current multiplied by the average voltage if there is a coil involved in the circuit – it will be less than that by an amount known as the "power factor" and I'll leave you to work out why it is called that.

So, how do you determine what the power is? It is done by sampling the voltage and current many times per second and averaging those combined results:

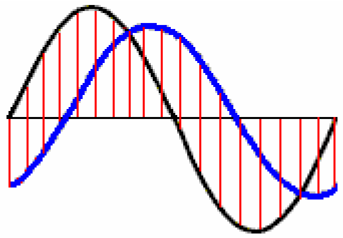

Both the voltage and the current are sampled at the times indicated by the vertical red lines and those figures are used to calculate the actual power level. In this example, only a few samplings are shown, but in practice, a very large number of samples will be taken. The piece of equipment which does this is known as a wattmeter as it measures watts of power. The sampling can be done by windings inside the instrument, resulting in an instrument which can be damaged by overloading without the needle being anywhere near full deflection, or it can be done by digital sampling and mathematical integration. Most digital sampling versions of these meters only operate at high frequencies, typically over 400,000 cycles per second. Both varieties of wattmeter can handle any waveform and not just sine waves.

The power company supplying your home, measures the current and assumes that the full voltage is present during all of the time that the current is being drawn. If you are powering a powerful electric motor from the mains, then this current lag will cost you money as the power company does not take it into account. It is possible to correct the situation by connecting one or more suitable capacitors across the motor to minimise the power loss.

With a coil (fancy name "inductor" symbol "L"), AC operation is very different to DC operation. The coil has a DC resistance which can be measured with the ohms range of a multimeter, but that resistance does not apply when AC is being used as the AC current flow is **not** determined by the DC resistance of the coil. Because of this, a second term has to be used for the current-controlling factor of the coil, and the term chosen is "impedance" or for people who like to make everything sound unduly complicated "reactance". I will stick with the term "impedance" as it is clear that it is the feature of the coil which "impedes" AC current flow through the coil.

The impedance of a coil depends on it's size, shape, method of winding, number of turns and core material. It also depends on the frequency of the AC voltage being applied to it. If the core is made up of iron or steel, usually thin layers of iron which are insulated from each other, then it can only handle low frequencies. You can forget about trying to pass 10,000 cycles per second ("Hz") through the coil as the core just can't change it's magnetic poles fast enough to cope with that frequency. A core of that type is ok for the very low 50 Hz or 60 Hz frequencies used for mains power, which are kept that low so that electric motors can use it.

For higher frequencies, ferrite can be used for a core and that is why some portable radios use ferrite-rod aerials, which are a bar of ferrite with a coil wound on it. For higher frequencies (or higher efficiencies) iron dust encapsulated in epoxy resin is used. An alternative is to not use any core material and that is usually referred to as an "air-core" coil. These are not limited in frequency by the core but they have a very much lower inductance for any given number of turns. The efficiency of the coil is called it's "Q" (for "Quality") and the higher the Q factor, the better. The resistance of the wire lowers the Q factor.

A coil has inductance, and resistance caused by the wire, and capacitance caused by the turns being near each other. However, having said that, the inductance is normally so much bigger than the other two components that we tend to ignore the other two. Something which may not be immediately obvious is that the impedance to AC current flow through the coil depends on how fast the voltage is changing. If the AC voltage applied to a coil completes one cycle every ten seconds, then the impedance will be much lower than if the voltage cycles a million times per second.

If you had to guess, you would think that the impedance would increase steadily as the AC frequency increased. In other words, a straight-line graph type of change. That is not the case. Due to a feature called resonance, there is one particular frequency at which the impedance of the coil increases massively. This is used in the tuning method for AM radio receivers. In the very early days when electronic components were hard to come by, variable coils were sometimes used for tuning. We still have variable coils today, generally for handling large currents rather than radio signals, and we call them "rheostats" and some look like this:

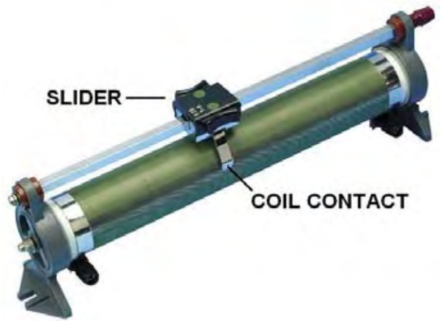

These have a coil of wire wound around a hollow former and a slider can be pushed along a bar, connecting the slider to different winds in the coil depending on it's position along the supporting bar. The terminal connections are then made to the slider and to one end of the coil. The position of the slider effectively changes the number of turns of wire in the part of the coil which is being used in the circuit. Changing the number of turns in the coil, changes the resonant frequency of that coil. AC current finds it very, very hard to get through a coil which has the same resonant frequency as the AC current frequency. Because of this, it can be used as a radio signal tuner:

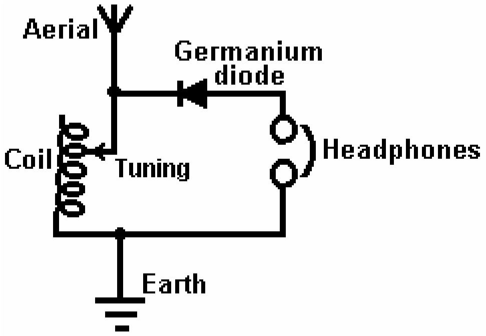

If the coil's resonant frequency is changed to match that of a local radio station by sliding the contact along the coil, then that particular AC signal frequency from the radio transmitter finds it almost impossible to get through the coil and so it (and only it) diverts through the diode and headphones as it flows from the aerial wire to the earth wire and the radio station is heard in the headphones. If there are other radio signals coming down the aerial wire, then, because they are not at the resonant frequency of the coil, they flow freely through the coil and don't go through the headphones.

This system was soon changed when variable capacitors became available as they are cheaper to make and they are more compact. So, instead of using a variable coil for tuning the radio signal, a variable capacitor connected across the tuning coil did the same job:

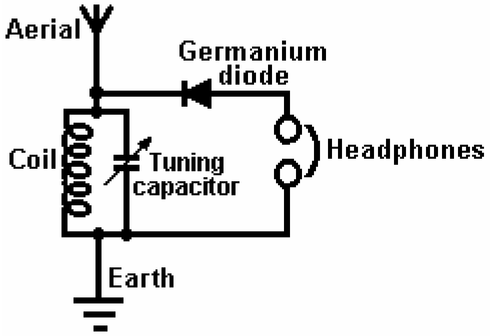

While the circuit diagram above is marked "Tuning capacitor" that is actually quite misleading. Yes, you tune the radio receiver by adjusting the setting of the variable capacitor, **but**, what the capacitor is doing is altering the resonant frequency of the coil/capacitor combination and it is the resonant frequency of that combination which is doing exactly the same job as the variable coil did on it's own.

This draws attention to two very important facts concerning coil/capacitor combinations. When a capacitor is placed across a coil "in parallel" as shown in this radio receiver circuit, then the combination has a very high impedance (resistance to AC current flow) at the resonant frequency. But if the capacitor is placed "in series" with the coil, then there is nearly zero impedance at the resonant frequency of the combination:

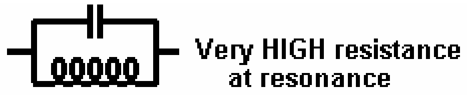

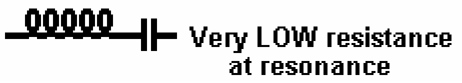

This may seem like something which practical people would not bother with, after all, who really cares? However, it is a very practical point indeed. Remember that Don Smith often uses an early version, off-the-shelf neon-tube driver module as an easy way to provide a high-voltage, high-frequency AC current source, typically, 6,000 volts at 30,000 Hz. He then feeds that power into a Tesla Coil which is itself, a power amplifier. The arrangement is like this:

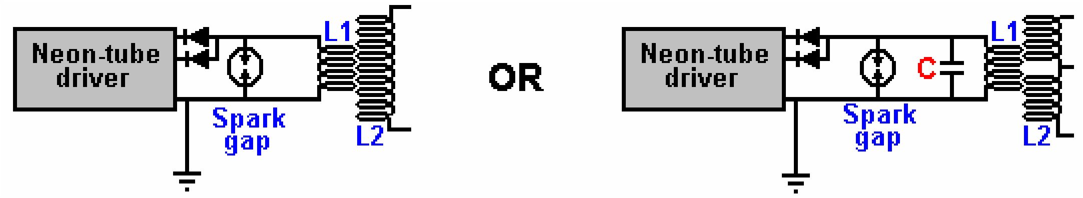

People who try to replicate Don's designs tend to say "I get great sparks at the spark gap until I connect the **L1** coil and then the sparks stop. This circuit can never work because the resistance of the coil is too low".

If the resonant frequency of the **L1** coil does not match the frequency being produced by the neon-tube driver circuit, then the low impedance of the **L1** coil at that frequency, will definitely pull the voltage of the neon-tube driver down to a very low value. But if the **L1** coil has the same resonant frequency as the driver circuit, then the **L1** coil (or the **L1** coil/capacitor combination shown on the right, will have a very high resistance to current flow through it and it will work well with the driver circuit. So, no sparks, means that the coil tuning is off. It is the same as tuning a radio receiver, get the tuning wrong and you don't hear the radio station.

This is very nicely demonstrated using simple torch bulbs and two coils in the YouTube video showing good output for almost no input power: http://www.youtube.com/watch?v=kQdcwDCBoNY and while only one resonant pick-up coil is shown, there is the possibility of using many resonant pick-up coils with just the one transmitter.

As it would not be unusual for many readers to feel that there is some "black magic" about the neon-driver circuit used by Don to drive the Tesla Coil section of his circuitry and that if a suitable unit could not be purchased then the circuit could not be reproduced or tested, it seems reasonable to show how it operates and how it can be constructed from scratch:

The circuit itself is made up of an oscillator to convert the 12-volt DC supply into a pulsating current which is then stepped up to a high voltage by a transformer. Here is a circuit which has been used for this:

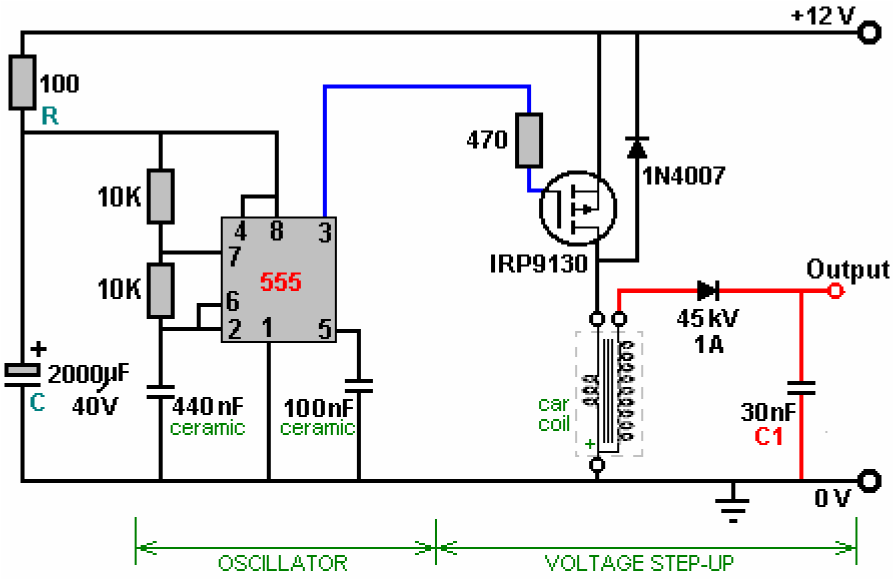
+12 V
100
R
470
10K
1N4007
4 8 3
7
555
10K
6
2 1 5
IRP9130
Output
45 kV
1A
+
2000μF
C
40V
440 nF
ceramic
100nF
ceramic
car
coil
+
30nF
C1
0 V
OSCILLATOR
VOLTAGE STEP-UP

The supply for the 555 timer chip is protected against spikes and dips by the resistor "R" and the capacitor "C". The 555 timer chip acts as an oscillator or "clock" whose speed is governed by the two 10K resistors feeding the 440 nF capacitor. The step-up transformer is an ordinary car coil and the drive power to it is boosted by the IRP9130 FET transistor which is driven by the 555 chip output coming from it's pin 3.

The output from the (Ford Model T) car coil is rectified by the diode, which needs to have a very high voltage rating as the voltage at this point is now very high. The rectified voltage pulses are stored in a very high-voltage capacitor before being used to drive a Tesla Coil. As a powerful output is wanted, two car coils are used and their outputs combined as shown here:

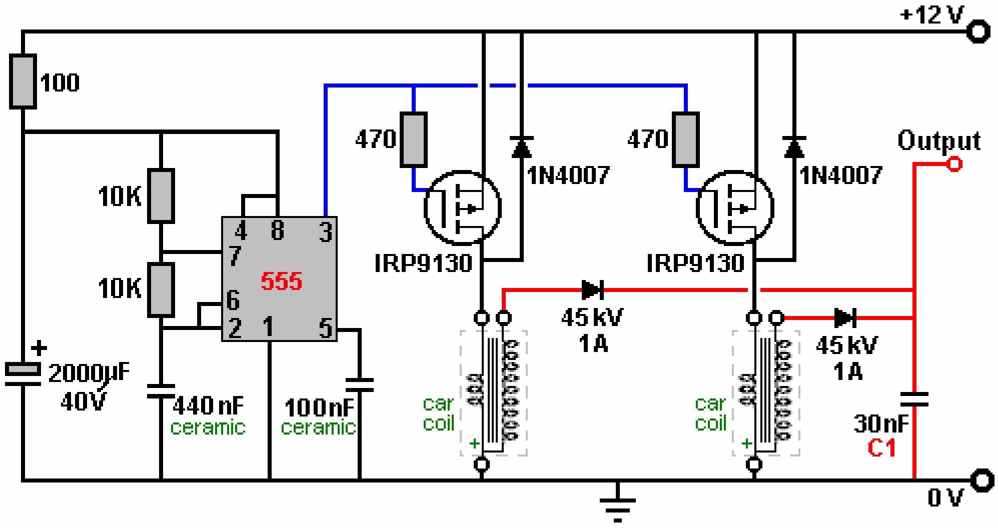
+12 V
100
470
470
1N4007
1N4007
Output
10K
4 8 3
7
555
10K
6
2 1 5
IRP9130
IRP9130
45 kV
1A
45 kV
1A
+
2000μF
40V
440 nF
ceramic
100nF
ceramic
car
coil
+
car
coil
+
30nF
C1
0 V

You will notice that the car coil has only three terminals and the terminal marked "+" is the one with the connection common to both of the coils inside the housing. The coil may look like this:

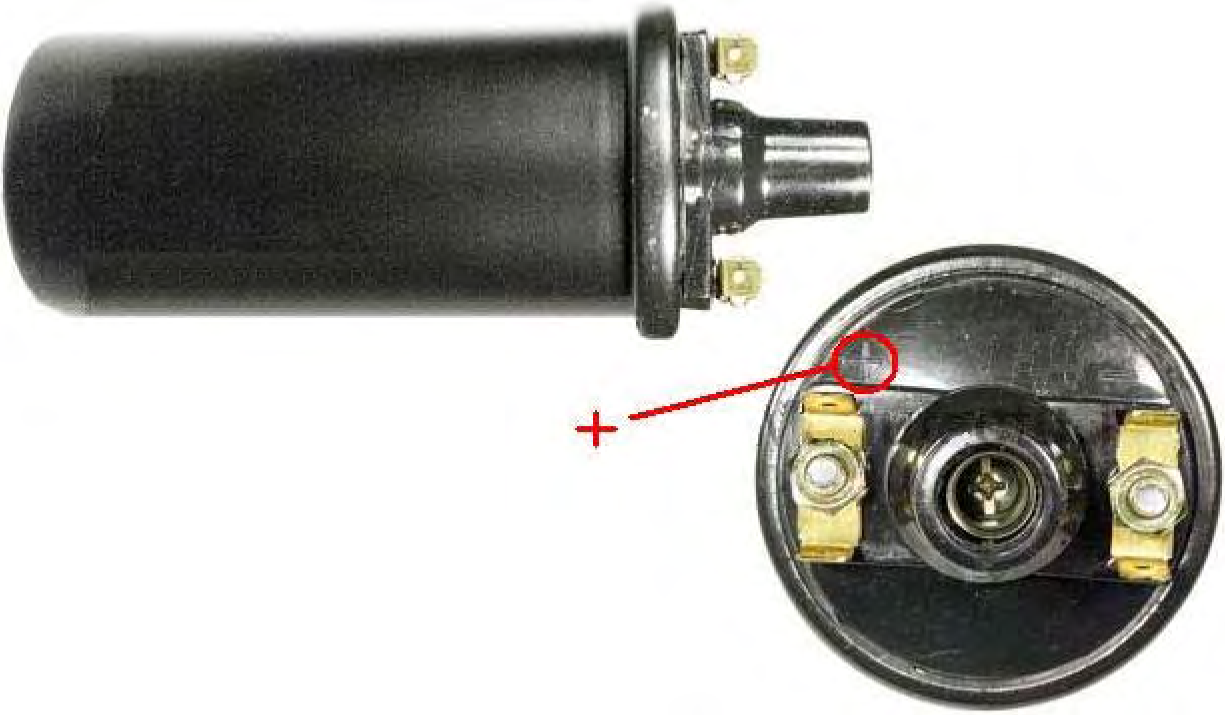

and the "+" is generally marked on the top beside the terminal with the two internal connections running to it. The circuit described so far is very close to that provided by a neon-tube driver circuit and it is certainly capable of driving a Tesla Coil.

There are several different way of constructing a Tesla Coil. It is not unusual to have several spark gaps connected in a chain. This arrangement is called a "series spark gap" because the spark gaps are connected "in series" which is just a technical way of saying "connected in a row". In the chapter on aerial systems, you will see that Hermann Plauston uses that style of spark gap with the very high voltages which he gets from his powerful aerial systems. These multiple spark gaps are much quieter in operation than a single spark gap would be. One of the possible Tesla Coil designs uses a pancake coil as the "L1" coil as that gives even higher gain. The circuit is as shown here:

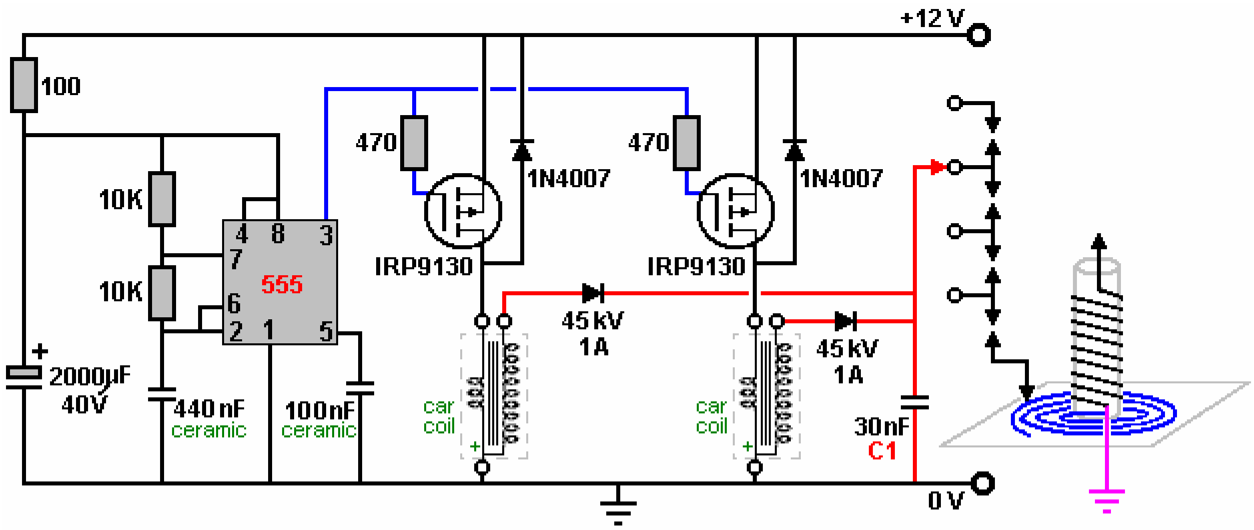
+12 V
100
10K
10K
4 8 3
7
555
6
2 1 5
+
2000µF
40V
440 nF
ceramic
100nF
ceramic
470
1N4007
IRP9130
470
1N4007
IRP9130
45 kV
1A
45 kV
1A
car
coil
+
car
coil
+
30nF
C1
0 V

The connection to the pancake coil is by a moveable clamp and the two coils are tuned to resonance by careful and gradual adjustment of that connection. The series spark gap can be constructed in various ways, including using car spark plugs. The one shown here uses nuts and bolts projecting through two strips of a stiff, non-conducting material, as that is much easier to adjust than the gaps of several spark plugs:

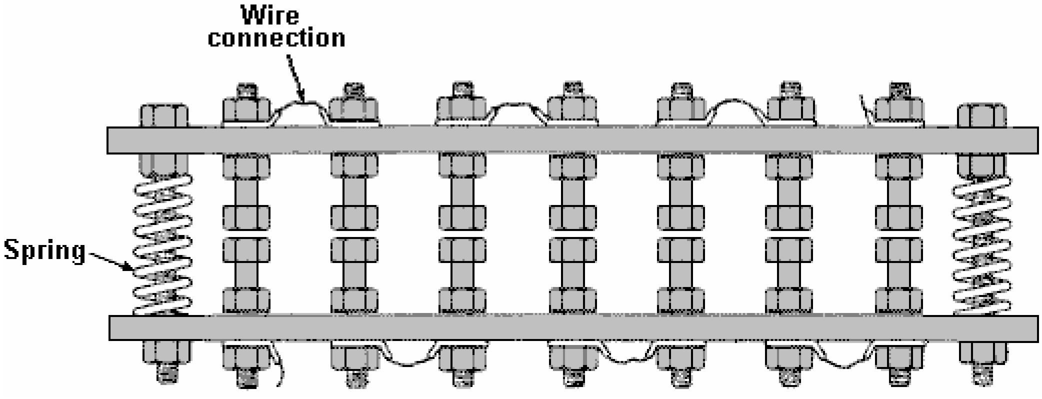

Tightening the bolts which compress the springs moves the bolt heads closer together and reduces all of the spark gaps. The electrical connections can be made to the end tags or to any of the intermediate wire connection straps if fewer spark gaps are required in the chain.

Let me remind you again that this is not a toy and very high voltages will be produced. Also, let me stress again that if you decide to construct anything, then you do so entire on your own responsibility. This document is only provided for information purposes and must not be seen as an encouragement to build any such device nor is any guarantee given that any of the devices described in this eBook will work as described should you decide to attempt to construct a replication prototype of your own. Generally, it takes skill and patience to achieve success with any free-energy device and Don Smith's devices are some of the most difficult, especially since he admits quite freely that he does not disclose all of the details.

The output capacitor marked "C1" in the circuit diagram above has to be able to handle very high voltages. There are various ways of dealing with this. Don dealt with it by getting very expensive capacitors manufactured by a specialist company. Some home-based constructors have had success using glass beer bottles filled with a salt solution. The outside of the bottles are wrapped in aluminium foil to form one of the contacts of the capacitor and bare wires are looped from deep inside each bottle on to the next one, looping from the inside of one bottle to the inside of the next one, and eventually forming the other contact of the capacitor. While that appears to work well, it is not a very convenient thing to carry around. An alternative is just to stand the bare bottles in a container which is lined with foil which forms the second contact of the capacitor.

One method which has been popular in the past is to use two complete rolls of aluminium foil, sometimes called "baking foil", laying them one flat, covering it with one or more layers of plastic cling film and laying the second roll of foil on top of the plastic. The three layers are then rolled up to form the capacitor. Obviously, several of these can be connected together in parallel in order to increase the capacitance of the set. The thicker the plastic, the lower the capacitance but the higher the voltage which can be handled.

The November 1999 issue of Popular Electronics suggests using 33 sheets of the thin aluminium used as a flashing material by house builders. At that time it was supplied in rolls which were ten inches (250 mm) wide, so their design uses 14" (355 mm) lengths of the aluminium. The plastic chosen to separate the plates was polythene sheet 0.062 inch (1.6 mm) thick which is also available from a builders merchants outlet. The plastic is cut to 11 inch (280 mm) by 13 inch (330 mm) and assembly is as follows:

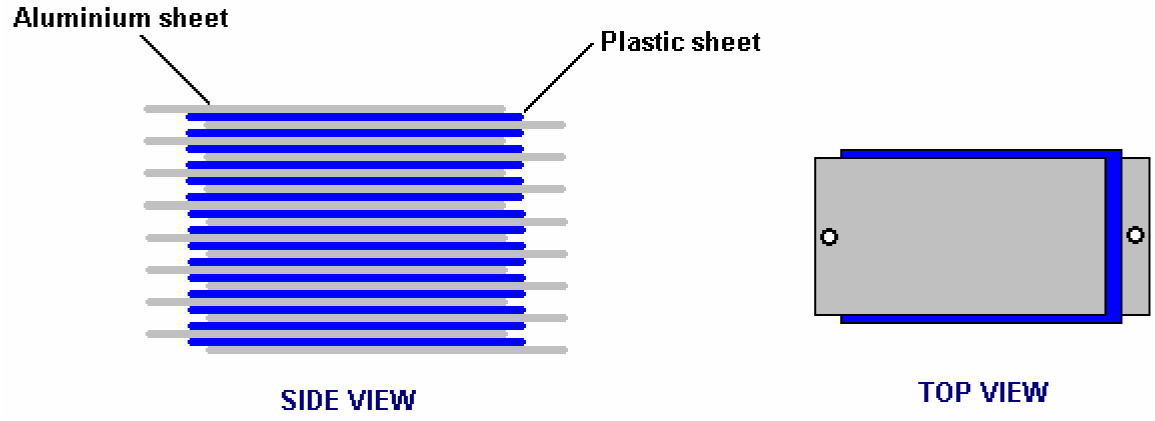

The sandwich stack of sheets is then clamped together between two rigid timber sheets. The tighter that they are clamped, the closer the plates are to each other and the higher the capacitance. The electrical connections are made by running a bolt through the projecting ends of the plates. With two thicknesses of plastic sheet and one of aluminium, there should be room for a washer between each pair of plates at each end and that would improve the clamping and the electrical connection. An alternative is to cut a corner off each plate and position them alternatively so that almost no plate area is ineffective.

As Don Smith has demonstrated in one of his video presentations, Nikola Tesla was perfectly correct when he stated that directing the discharge from a Tesla Coil on to a metal plate (or in Don's case, one of the two metal plates of a two-plate capacitor where a plastic sheet separates the plates just as shown above), produces a very powerful current flow onwards through a good earth connection. Obviously, if an electrical load is positioned between the plates and the earth connection, then the load can be powered to a high level of current, giving a very considerable power gain.

Coil Construction: The Barker & Williamson coils used by Don in his constructions are expensive to purchase. Some years ago, in an article in a 1997 issue of the “QST” amateur radio publication, Robert H. Johns shows how similar coils can be constructed without any great difficulty. The Electrodyne Corporation research staff have stated that off-the-shelf solid tinned copper wire produces three times the magnetic field that un-tinned copper does, so perhaps that should be borne in mind when choosing the wire for constructing these coils.

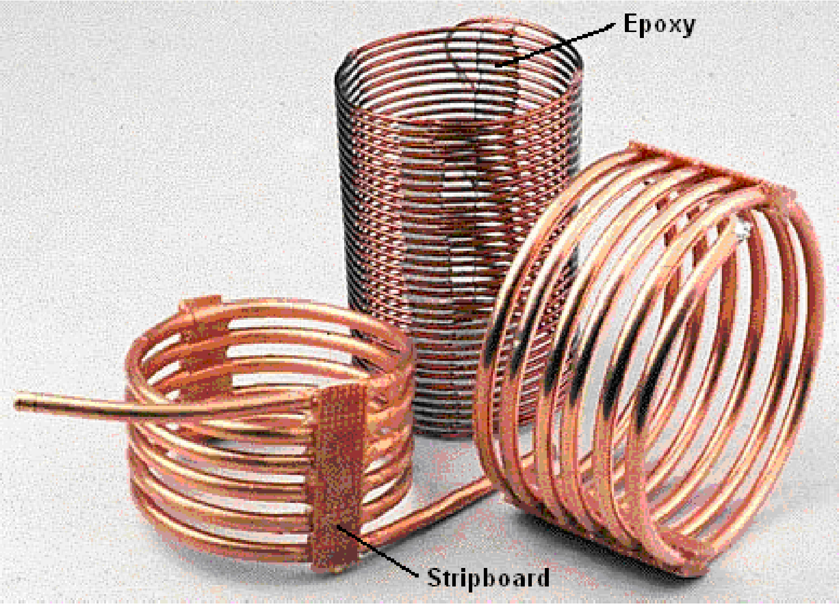

These home-made coils have excellent “Q” Quality factors, some even better than the tinned copper wire coils of Barker & Williamson because the majority of electrical flow is at the surface of the wire and copper is a better conductor of electricity than the silver tinning material.

The inductance of a coil increases if the turns are close together. The capacitance of a coil decreases if the turns are spread out. A good compromise is to space the turns so that there is a gap between the turns of one wire thickness. A common construction method with Tesla Coil builders is to use nylon fishing line or plastic strimmer cord between the turns to create the gap. The method used by Mr Johns allows for even spacing without using any additional material. The key feature is to use a collapsible former and wind the coil on the former, space the turns out evenly and then clamp them in position with strips of epoxy resin, removing the former when the resin has set and cured.

Mr Johns has difficulty with his epoxy being difficult to keep in place, but when mixed with the West System micro fibres, epoxy can be made any consistency and it can be applied as a stiff paste without any loss of it’s properties. The epoxy is kept from sticking to the former by placing a strip of electrical tape on each side of the former.

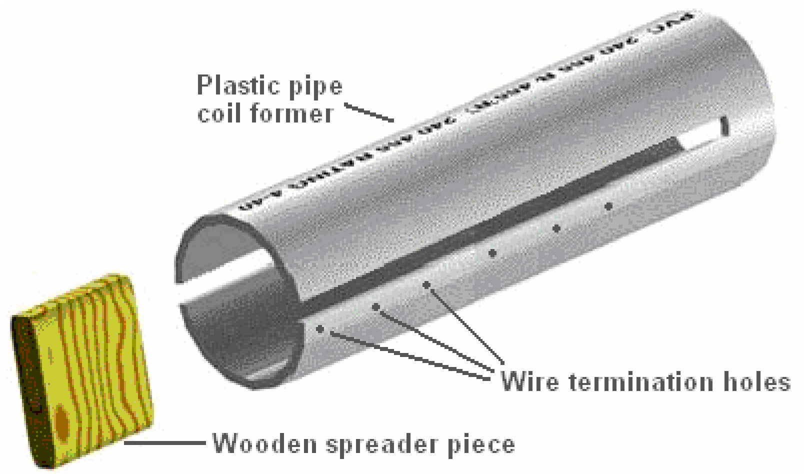

I suggest that the plastic pipe used as the coil former is twice the length of the coil to be wound as that allows a good degree of flexing in the former when the coil is being removed. Before the two slots are cut in the plastic pipe, a wooden spreader piece is cut and it's ends rounded so that it is a push-fit in the pipe. This spreader piece is used to hold the sides of the cut end exactly in position when the wire is being wrapped tightly around the pipe.

Two or more small holes are drilled in the pipe beside where the slots are to be cut. These holes are used to anchor the ends of the wire by passing them through the hole and bending them. Those ends have to be cut off before the finished coil is slid off the former, but they are very useful while the epoxy is being applied and hardening. The pipe slots are cut to a generous width, typically 10 mm or more.

The technique is then to wedge the wooden spreader piece in the slotted end of the pipe. Then anchor the end of the solid copper wire using the first of the drilled holes. The wire, which can be bare or insulated, is then wrapped tightly around the former for the required number of turns, and the other end of the wire secured in one of the other drilled holes. It is common practice to make the turns by rotating the former. When the winding is completed, the turns can be spaced out more evenly if necessary, and then a strip of epoxy paste applied all along one side of the coil. When that has hardened, (or immediately if the epoxy paste is stiff enough), the pipe is turned over and a second epoxy strip applied to the opposite side of the coil. A strip of paxolin board or strip-board can be made part of the epoxy strip. Alternatively, an L-shaped plastic mounting bracket or a plastic mounting bolt can be embedded in the epoxy ready for the coil installation later on.

When the epoxy has hardened, typically 24 hours later, the coil ends are snipped off, the spreader piece is tapped out with a dowel and the sides of the pipe pressed inwards to make it easy to slide the finished coil off the former. Larger diameter coils can be wound with small-diameter copper pipe.

The coil inductance can be calculated from:

Inductance in microhenrys $\mathbf{L} = \mathbf{d}^2\mathbf{n}^2 / (18\mathbf{d} + 40\mathbf{l})$

Where:
d is the coil diameter in inches measured from wire centre to wire centre
n is the number of turns in the coil
l is coil length in inches (1 inch = 25.4 mm)

Using this equation for working out the number of turns for a given inductance in microhenrys:

$$n = \frac{\sqrt{L(18d + 40l)}}{d}$$

Patrick Kelly
engpjk@gmail.com
http://www.free-energy-info.co.uk
http://www.free-energy-devices.com

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/319136703

Radio Frequency Energy Harvesting From Ambient FM Signals for Making Battery-Less Sensor Nodes For Wireless Sensor Networks

Conference Paper · November 2017

DOI: 10.1109/APMC.2017.8251487

CITATIONS
0

READS
678

4 authors, including:

Mutee Urrehman
National University of Sciences and Technology

6 PUBLICATIONS **2** CITATIONS

SEE PROFILE

Muhammad Ibrahim Qureshi
LUMS

4 PUBLICATIONS **2** CITATIONS

SEE PROFILE

Waleed Ahmad
Lahore University of Management Sciences

8 PUBLICATIONS **12** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Planar Inverted-F Antenna for 5 G Applications View project

Project Energy Harvesting View project

All content following this page was uploaded by Waleed Ahmad on 16 August 2017.

The user has requested enhancement of the downloaded file.

Radio Frequency Energy Harvesting From Ambient FM Signals for Making Battery-Less Sensor Nodes For Wireless Sensor Networks

Mutee-ur-Rehman, Muhammad Ibrahim Qureshi, Waleed Ahmad and Wasif Tanveer Khan
Lahore University of Management Sciences Pakistan
mutee@lums.edu.pk
qureshi.mi13@gmail.com
waleed.ahmed@lums.edu.pk
wasif.tanveer@lums.edu.pk

***Abstract*—This paper presents an efficient rectifier for Radio Frequency (RF) energy harvesting applications in FM band (88MHz to 108MHz). The proposed rectifier is based on Modified Greinacher Rectification topology. The circuit is designed on Keysight Advance Design System (ADS) and realized on a 1.6mm thick FR4 sheet with 35um of copper cladding. Schottky diode, HSMS2850 is used as rectification element and lumped components are used for impedance matching. Power levels of interest are from -30dBm to 0dBm. The rectifier operates optimally for a load range of 10kΩ to 75kΩ with maximum RF to DC conversion efficiency of 74.8% for 5dBm input power at 96 MHz.The conversion efficiency for entire FM band for -10 dBm input is around 50% which is the typical input power level of ambient wireless signals. Measurements show excellent agreement with the simulated results. The proposed rectifier has harvested 12.77 uW of DC power from an input of -14.2dBm channel power with an overall efficiency of 33.59% in FM band.**

Keywords—Ambient sources;FM band;impedance matching;rectifier;RF energy harvesting

I. INTRODUCTION

With the advent of Wireless Sensor Networks (WSNs), development of self- sustainable battery-less sensor nodes has become an active research area. Different mechanisms are being investigated to supply wireless power to the sensor nodes of WSNs and replace batteries from them which will save the maintenance cost and labor. Long range coverage and omnipresence of ambient RF signals make RF Energy Harvesting as one of the most suitable ways to achieve this objective [1]. Investigations have been done to design various rectennas to harvest DC power from different RF bands present in the ambient environment. The authors in [2] have featured a dual band rectenna that harvested DC voltage from GSM1800 and UMTS2100 bands. By using the multiband approach in [3] a stacked RF energy harvester was proposed which can harvest energy from GSM900, GSM1800, UMTS2100 and Wi-Fi bands. Authors in [4] have featured a novel six-band rectenna which also incorporated lower bands of 550MHz and 750MHz along with the four others published in [3].

FM band is used to broadcast radio channels throughout the world. Its omnipresence, lesser free space path loss, easier circuit design and cheap electronic components are attractive features to use it for RF energy harvesting but in these recent developments FM band is not utilized. Authors in [5] have proposed a RF energy harvester for FM band with a bandwidth of 3MHz (81.9 MHz to 84.7MHz). According to the best of authors' knowledge no rectifier design has been reported which can operate at low power levels for the entire bandwidth of FM band (88MHz to 108MHz). This paper focuses on development of a highly efficient, low threshold and wideband rectifier to harvest DC power from ambient FM band signals. The rest of the paper is organized as follows: Section II describes the design of the rectifier, Section III analyzes the performance of the proposed design, Section IV shows the results of measurement and Section V concludes the paper.

II. DESIGN OF RECTIFIER

A. Topology of Rectifier

Voltage multipliers are considered as best choice for RF energy harvesting because of their high efficiency and simple designs [1]. Due to the same merits, a Modified Greinacher Rectifier topology (shown in Fig.1) was chosen in this study. Two voltage doublers are differentially connected to form this topology. Modified Greinacher Rectifier has smaller input impedance as compared to simple voltage doubler which results in ease in impedance matching **[6]**. The differential configuration results in reduced harmonic content, better power handling capability and lesser loss in capacitors **[6]**. The diode used for rectification is HSMS-2850 by Avago Technologies.

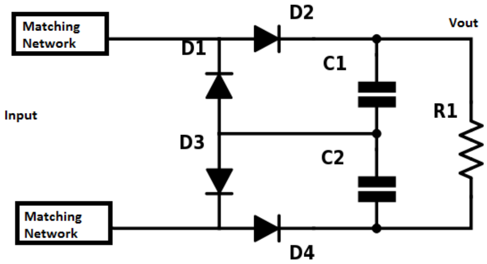

Fig. 1 Modified Greinacher Rectifier Topology with Impedance Matching

B. Impedance Matching

Impedance matching is a crucial stage in rectifier design because of the non-linear behavior of rectifier. If the impedance of the rectifier is not matched to the antenna (which is generally 50Ω) the incident wireless signals will be reflected to the environment which shall result in decrease in efficiency of the RF energy harvester. The input impedance of a rectifier is function of frequency of operation, input power and load connected at the output. To get a better insight of this non-linear behavior a Modified Greinacher Rectifier was designed in ADS. Non-linear model of the diode HSMS2850 was imported in design from HF Diode Library of ADS. Chip capacitors (C1 and C2) of 100nF were used in this design and were modeled by using SPICE files provided by Murata. The layout for rectifier was designed in Momentum and was co-simulated to include the non-linear effects. The circuit was simulated using Harmonic Balance (HB) simulation controller. Real and imaginary values of impedance were noted for power levels -30dBm to 0dBm in the desired frequency range (88MHz to 108MHz). A third order lumped component based matching network shown in Fig. 2 was chosen for each branch. Initial parameters of the matching network were calculated for the following values of input impedance; 1636-j1298Ω @88MHz and 1378-j1325 Ω for -20 dBm and 1685-1622jΩ@ 88MHz and 1346-1585jΩ @108 MHz for -30 dBm input power for a load resistor of 10kΩ using Smith Chart Utility in ADS. In the next step network was optimized for entire power range (-30dBm to 0dBm) and load range (10kΩ to 75Ω) using Optimization Utility in ADS. The optimized values for components of matching network are shown in Table I. The realized prototype of the proposed rectifier is shown in Fig 3.

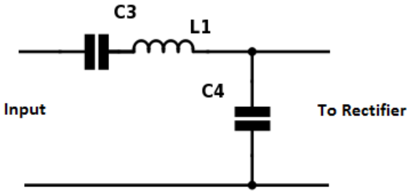

Fig. 2 Proposed Impedance Matching Network for each branch

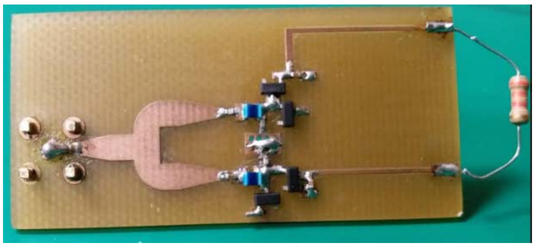

Fig 3. Fabricated Prototype of the Proposed Rectifier

TABLE I

OPTIMIZED PARAMETERS FOR THE PROPOSED MATCHING NETWORK

Component	**Upper Branch**	**Lower Branch**
C3	8.2pF	47000pF
L1	820nH	820nH
C4	3pF	2.7pF

III. PERFORMANCE ANALYSIS

The simulated and measured reflection coefficients (S11) for the proposed rectifier are shown in Fig. 4. It can be seen in fig.4a that the proposed rectifier is matched in FM band for multiple power levels (-10dBm to-30 dBm) and hence it can be used with ambient FM band signals which always have a varying power level. Fig4b shows that the proposed rectifier is matched for a diverse load range (10kΩ to 75kΩ) and it can supply DC power to variety of wireless sensors and loads. Both traits make the proposed rectifier a perfect fit for RF energy harvesting applications.

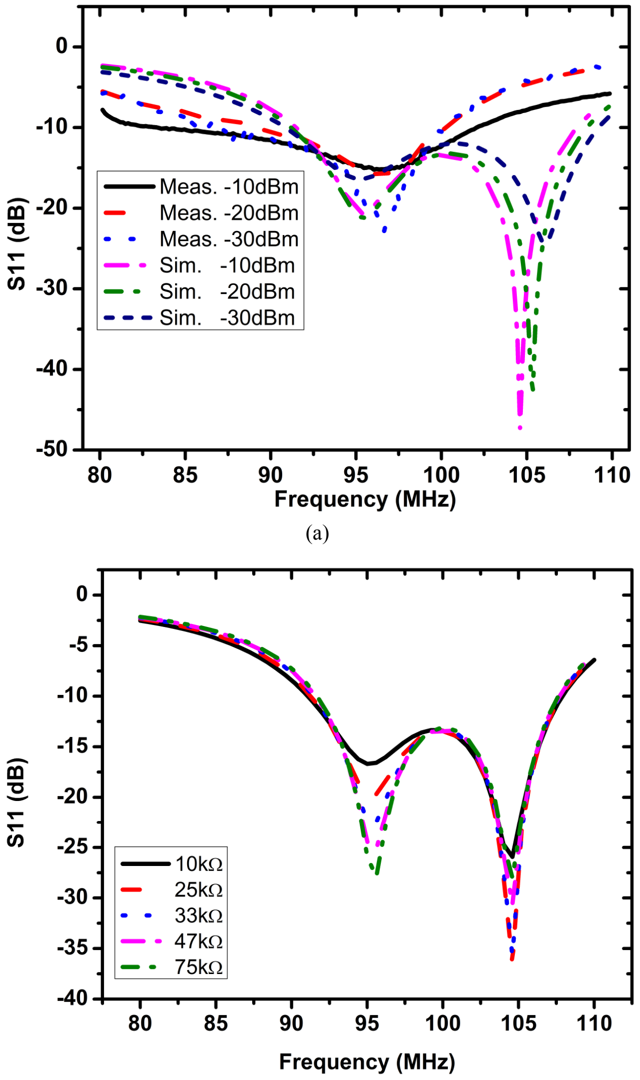

(a)

(b)

Fig. 4 S11(dB) of the proposed rectifier (a) measured and simulated for multiple input power levels when R_L=10kΩ (b) simulated for multiple loads when Pin=-15 dBm

The key performance indicator of a rectifier is its RF to DC conversion efficiency (η_{conv}) which is described in “equation (1)”.

It shows the capability of a rectifier to produce output DC power.

$$\eta_{conv} = P_{out}/P_{in} \qquad (1)$$

Where, P_{in} is the input power to the rectifier and P_{out} is the output DC power generated by the rectifier and is described by "equation (2)"

$$P_{out} = (V_{out.}I_{out})/R_{Load} \qquad (2)$$

V_{out} is output voltage across the load resistor and I_{out} is the current through it.

The simulated and measured conversion efficiency of the proposed rectifier is shown in Fig 5. It can be seen that in fig.5a that the conversion efficiency is almost constant for entire FM band at different power levels and hence it can maintain excellent performance at different frequencies and varying power levels and is fit for real world conditions. The conversion efficiency is around 50% for entire FM band at -10dBm power level which implies that it is very efficient at low power and makes it a very good candidate for energy harvesting applications. Fig5b shows that the maximum conversion efficiency of the proposed rectifier is 74.8% for 5dBm input power. This implies that the proposed rectifier is also efficient for higher power levels and it can also be used for RF to DC conversion in applications which involves dedicated power sources and wireless power transfer. The efficiency reduces at higher power levels (above 5dBm) due to the break-down of the diode.

The literature available for RF to DC conversion in FM band is very limited because it hasn't been investigated for RF energy harvesting. Table II shows comparison of this work with some related designs and in terms of varying power level, bandwidth of operation and output load range the proposed rectifier is better than the other published designs because they haven't reported the performance under changing power levels, output loads and are not matched for entire FM band.

TABLE II
COMPARISON WITH THE RELATED WORK

Ref	**Bandwidth**	**Power Levels of Interest**	**Load Range**	**Efficiency at Pin=-10 dBm**
[5]	81.9 to 84.7 MHz	-20 dBm	NR	NR
[6]	100 MHz	0dBm to 35dBm	NR	NR
[This Work]	**88 to 108 MHz**	**-35dBm to 0dBm**	**10-75 kΩ**	**53% @ 90MHz** (Max. Efficiency 74% @ 5dBm input power)

* NR: Not Reported

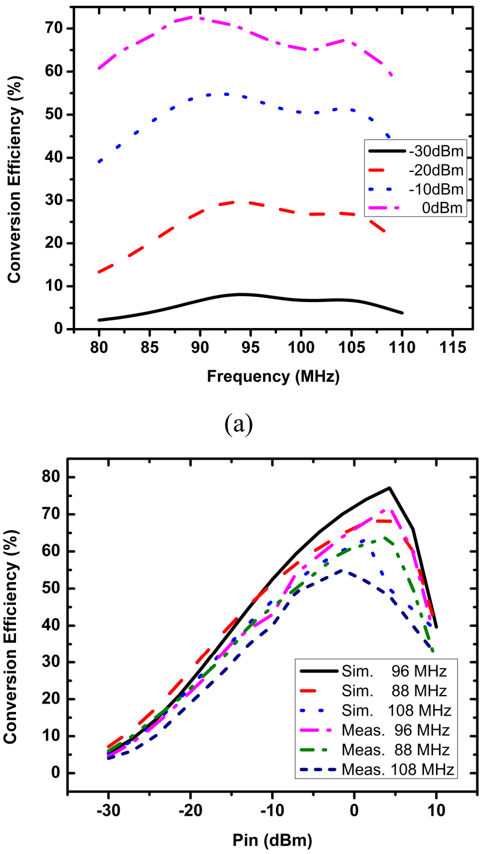

(a)

(b)

Fig.5 Conversion Efficiency of the proposed rectifier (a) Conversion Efficiency vs Freq (b) Conversion Efficiency vs Input Power for Multiple Frequencies

IV. MEASUREMENT WITH ANTENNA

Having optimized the rectifier to work with low power signals of FM band and for a wide output load range the proposed rectifier was tested with an antenna to analyze its performance with wireless signals. First, ambient signal strength of FM spectrum was measured on top of a building with a handheld spectrum analyzer (PSA-2702). The received spectrum is shown in Fig. 6. The received spectrum received was exported to MATLAB for calculating the channel power using the 'bandpower' function. The total channel power residing in FM band was -14.2 dBm. The ambient signal strength is always varying because the incident signals are suffering various multipath reflections. To test the proposed rectifier and analyze its performance, FM spectrum with similar channel power was emulated in lab using Rhode and Schwarz Vector Signal Generator. A general-purpose antenna, PSA Ant2 (gain 2.15 dBi) was connected to the rectifier and it was placed at 10m distance from the transmitting end. The total transmit power was 20dBm which was distributed in 8 tones at multiple frequencies. The spectrum received at 10m distance from the transmitting end is shown in fig.6. At this distance, the channel power received by the antenna was roughly the same as measured in ambient environment and the output voltage measured was 0.619V (shown in Fig. 7) across a 30kΩ resistor. The total DC power harvested was 12.77uW with an overall efficiency of 33.59%. The harnessed DC power can be used to power a variety of low power sensors.

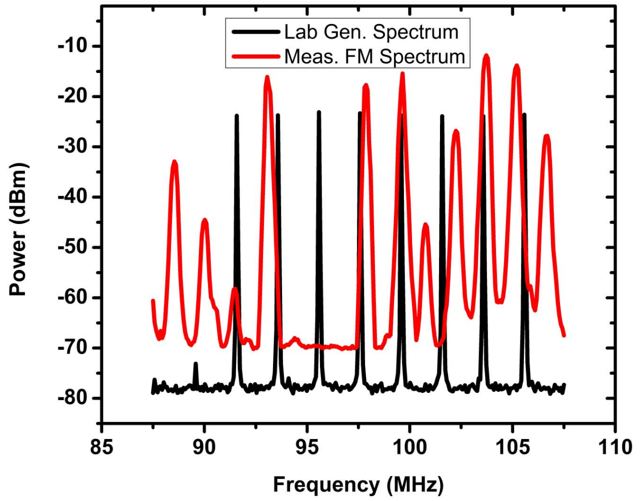

Fig. 6 Measured FM band spectrum in ambient environment and lab generated FM band for testing of the proposed rectifier

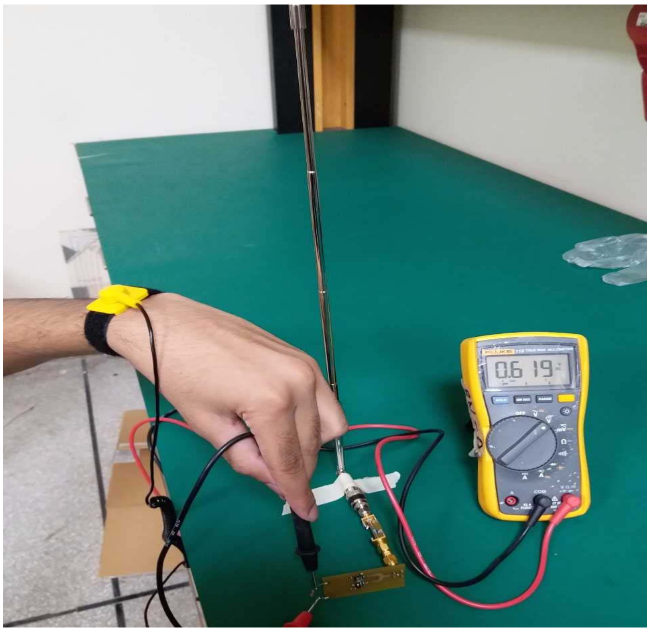

Fig. 7 Measurement of output voltage generated by proposed rectifier for -14.2 dBm input power

V. CONCLUSION

A highly efficient and low threshold rectifier is presented for harnessing DC power from ambient wireless signals of FM band. The proposed design is optimized for operation in various conditions such as varying power levels of signals and wide range of loads. The rectifier operates optimally at lower power levels and maintains a flat efficiency of 50% at -10dBm input power level for all the frequencies of FM band (88MHz to 108MHz). The proposed design has an excellent power handling capability it has a maximum RF to DC conversion efficiency of 74.8% for 5dBm input power at 96 MHz. To the best of authors' knowledge this work is the first study which has proposed a rectifier which maintains excellent performance in various conditions for entire FM band. This work will pave the way for developing various real world wireless energy harvesting circuits and battery-less sensor modules for wireless sensor networks.

REFERENCES

[1]. Christopher R. Valenta, Gregory D. Durgin,"*Harvesting Wireless Power,"* IEEE Microwave Magazine, vol. 15,no.4, pp.108-120, June,2014

[2]. H. Sun, Y. x. Guo, M. He and Z. Zhong, *"A Dual-Band Rectenna Using Broadband Yagi Antenna Array for Ambient RF Power Harvesting,"* in *IEEE Antennas and Wireless Propagation Letters*, vol. 12, no. , pp. 918-921, 2013.

[3]. V. Kuhn, C. Lahuec, F. Seguin and C. Person, *"A Multi-Band Stacked RF Energy Harvester With RF-to-DC Efficiency Up to 84%,"* in IEEE Transactions on Microwave Theory and Techniques, vol. 63, no. 5, pp. 1768-1778, May 2015.

[4]. C. Song *et al*., "*A Novel Six-Band Dual CP Rectenna Using Improved Impedance Matching Technique for Ambient RF Energy Harvesting*," in *IEEE Transactions on Antennas and Propagation*, vol. 64, no. 7, pp. 3160-3171, July 2016.

[5]. A. Noguchi and H. Arai, "*Small loop rectenna for RF energy harvesting,*" 2013 Asia-Pacific Microwave Conference Proceedings (APMC), Seoul, 2013, pp. 86-88.

[6]. J. P. Curty, N. Joehl, F. Krummenacher, C. Dehollain and M. J. Declercq, *"A model for µ-power rectifier analysis and design,"* in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 12, pp. 2771-2779, Dec. 2005

View publication stats

S. Gernsback's
Radio Encyclopedia

S. GERNSBACK'S

RADIO ENCYCLOPEDIA

SIDNEY GERNSBACK
PUBLISHER
NEW YORK

All rights reserved including that of translating
into foreign languages including the Scandinavian

Copyright 1927
by
SIDNEY GERNSBACK

Printed in the United States of America

A

“A” BATTERY—An appliance for furnishing electrical current to light the *filaments* of a *vacuum tube*, and cause emission of *electrons* (see *Theory of Vacuum Tube Operation*). This battery may be in any of several forms, the two in popular use being the storage type (see *Storage Battery*), and the dry cell (see *Dry Cell*). Fig. 1 shows method of connecting the “A” Battery in a vacuum tube circuit.

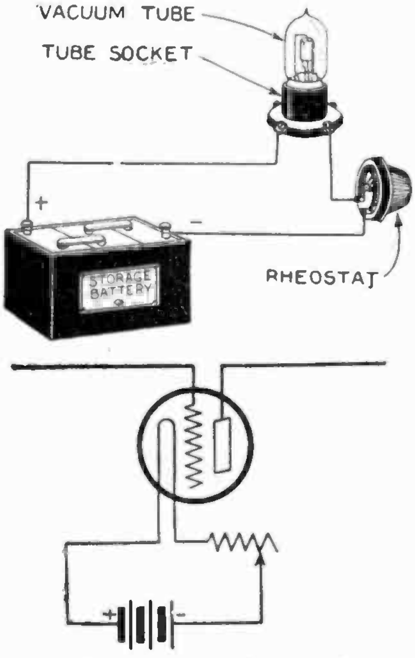
VACUUM TUBE
TUBE SOCKET
+
−
STORAGE BATTERY
RHEOSTAT
+
−

Fig. 1. Method of connecting the “A” battery to a vacuum tube in radio receiving circuits. The lower diagram shows the “A” battery as represented in schematic drawings

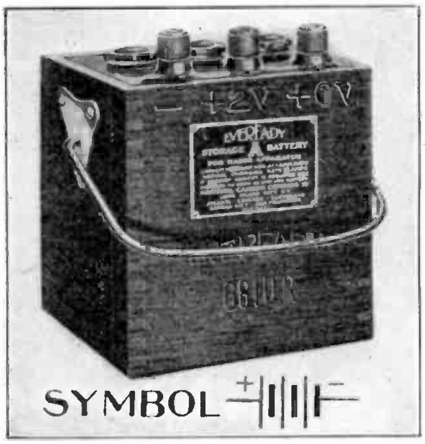
− +2V +6V
EVEREADY
STORAGE A BATTERY
SYMBOL
+
−

Photos by courtesy of National Carbon Co., Inc.

Fig. 2. A storage battery such as used for “A” battery current supply

In Fig. 2 a popular type of storage battery is shown. Batteries of this type come in many individual forms, the usual voltage being standardized at 6 volts. However, for use with certain tubes, 2 and 4 volt storage batteries are available. For general radio work, using the low-consumption tubes, the battery, as shown in Fig. 2, can hold a charge of from 60-80 ampere hours. (*See Storage Battery.*)

The other common form of “A” Battery is the dry cell, usually having a difference of potential (voltage) of 1.5 volts. Fig. 3 illustrates a popular type of dry cell. In connecting a number of cells for use with a vacuum tube, the method of connecting will depend on the voltage rating of the particular tube. If the tube is rated at 1.5 volts, a single dry cell may be used as in Fig. 4, or several may be connected in *parallel* (*q.v.*) shown in Fig. 5, the total voltage however, being only the voltage of one cell. Three dry cells connected in *series* (*q.v.*) will give 4½ volts, Fig. 6. A connection for dry cells, known as *series-parallel* or *multiple* (*q.v.*), is shown in Fig. 7. This connection will also produce 4½ volts and at the same time increase the life of the cells. Four dry cells connected in series will give six volts for use with a standard six-volt tube, Fig. 8. Dry cells are only used with six-volt tubes where the tube has a current rating of ¼ ampere, and seldom more than one tube is used in this manner due to the relatively high discharge and consequent short life of the battery.

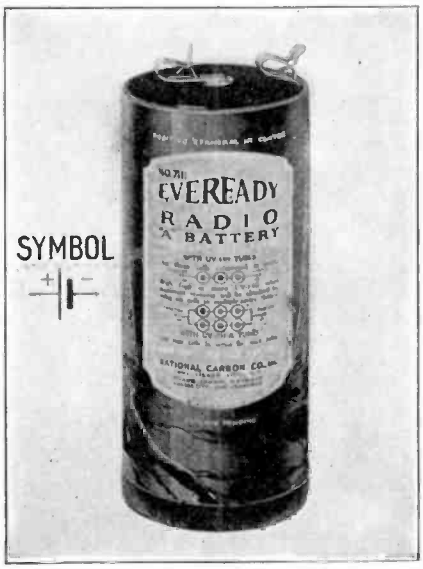
SYMBOL
+
−
No. 711
EVEREADY
RADIO
“A” BATTERY
NATIONAL CARBON CO. INC.

Fig. 3. A dry cell “A” battery

A storage battery can be used for the “A” Battery current with a 4-volt tube if the connection is made as in Fig. 9. Here a standard 6-volt storage battery is tapped at the four volt terminal.

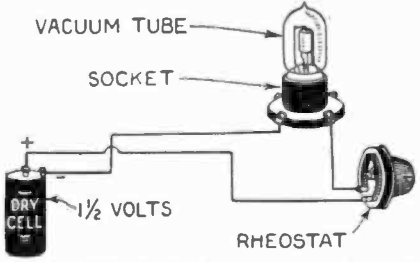
VACUUM TUBE
SOCKET
+
DRY CELL
1½ VOLTS
RHEOSTAT

Fig. 4. This circuit shows how one dry cell is used as the “A” battery for types of tubes requiring 1½ volts filament current

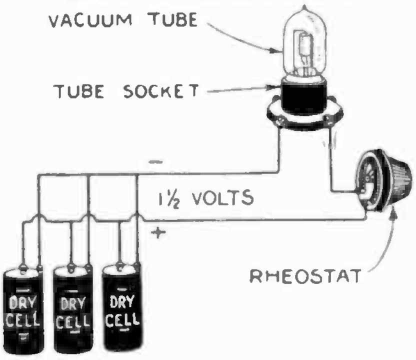
VACUUM TUBE
TUBE SOCKET
−
1½ VOLTS
+
RHEOSTAT
DRY CELL
DRY CELL
DRY CELL

Fig. 5. Several dry cells may be connected in parallel for increased amperage or longer life of the “A” battery

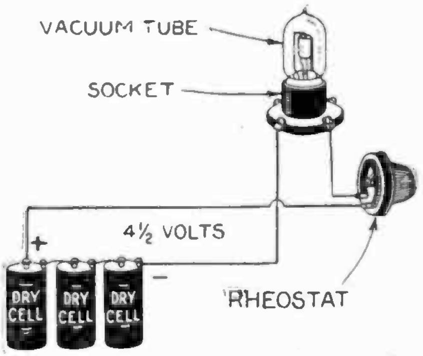
VACUUM TUBE
SOCKET
4½ VOLTS
+
−
DRY CELL
DRY CELL
DRY CELL
RHEOSTAT

Fig. 6. Three dry cells connected in series give 4½ volts, serving as the “A” battery for the U.V.199 or C299 type of tube

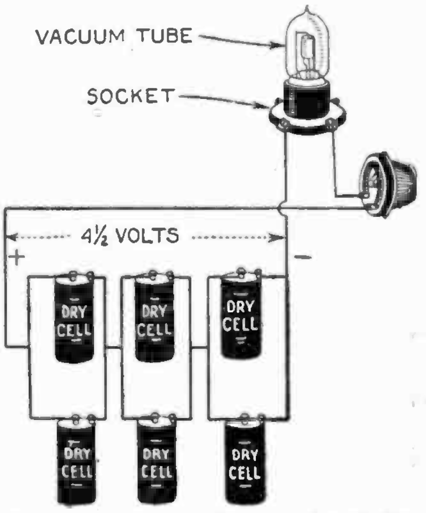
VACUUM TUBE
SOCKET
4½ VOLTS
+
−
DRY CELL
DRY CELL
DRY CELL
DRY CELL
DRY CELL
DRY CELL

Fig. 7. A series—parallel connection of three dry cells for longer life of the “A” battery with 4½ volt tubes

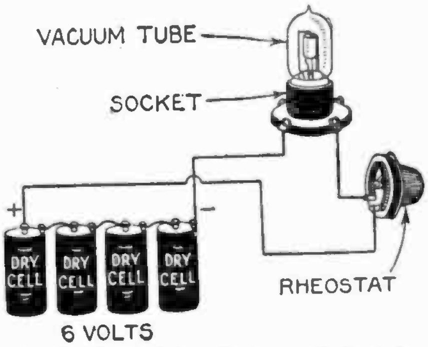
VACUUM TUBE
SOCKET
+
−
DRY CELL
DRY CELL
DRY CELL
DRY CELL
RHEOSTAT
6 VOLTS

Fig. 8. Four dry cells connected in series give: 6 volts for low consumption tubes

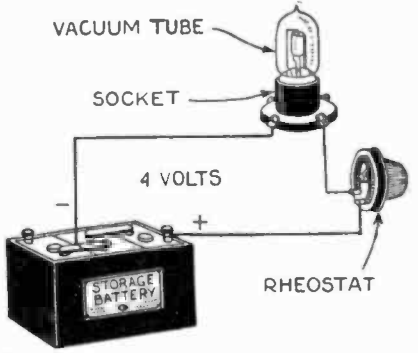
VACUUM TUBE
SOCKET
4 VOLTS
−
+
STORAGE BATTERY
RHEOSTAT

Fig. 9. Showing method of tapping at 4-volt terminal of a 6-volt storage battery for type U.V.199 or C299 tubes or other 4-volt tubes

Another connection for low voltage tubes can be made as illustrated in Fig. 10. Here only one cell of the storage battery is used. This cell can be used until exhausted when the other cells can be used in the same manner. When the three cells are exhausted, the storage battery may be charged by means of a *charger* (*q.v.*).

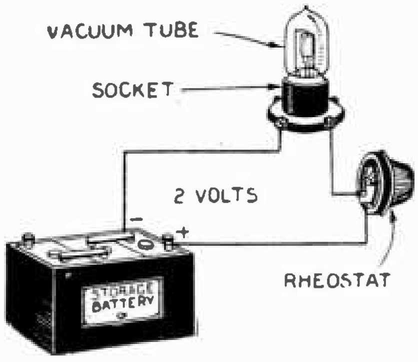

Fig. 10. How a single cell of a 6-volt storage battery may be employed to furnish "A" battery current for 1½ volt tubes

ABAMPERE—A unit of current used in special theoretical work. It has a value of 10 *amperes* (*q.v.*) and is not generally used in electrical work.

ABBREVIATIONS, AMATEUR—The transmitting amateurs of the United States use their own peculiar abbreviations in their inter-communication work. These are used to lessen the interference between transmitting stations as it permits conversation in a much quicker manner. The abbreviations generally used are listed below. Simplified spelling is always resorted to, words being spelled according to their sounds. Some such words are given, but others are so obvious that they are not listed. There are also the standard "Q" signals adopted by the International Radiotelegraphic Convention, to be found in official publications and given under the heading "Abbreviations, International Radiotelegraphic Convention."

ACCW—A transmitter using rectified alternating current as the plate supply.

AFTRN or P.M.—Afternoon.

AMMTR—Ammeter.

ARRL—American Radio Relay League.

BCL—Broadcast listener, one who does not have a transmitting station.

BC STATION—Broadcast Station.

BFRE—or B4—Before.

BLO—Blow.

BOTTLE or V.T.—Vacuum Tube.

C—See or Call.

CGE—Cage.

CM—City Manager of A. R. R. L.

CTPSE—Counterpoise.

CUL—See you later or call you later.

CW—Continuous wave transmitter.

DCCW—A transmitter using direct current as the supply for the plate.

DM—District Manager of A. R. R. L.

DOPE—Information, or in another sense, a varnish for coating inductance coils.

DS—District Superintendent of the A. R. R. L.

DX—Distance, receiving or transmitting.

FB—Fine Business.

FONE—Phone.

GA—Good afternoon.

GLD—Glad.

GM—Good morning.

GN—Good night.

FRQ—Frequency.

HAM—Being a term for the transmitting amateur.

HI—High or I am laughing—it's funny, according to use in sentence.

HI LOS—High loss.

HR—Here.

HV—Have.

KCK BCK—Kick back.

KONGRATS—Congratulations.

KY—Key.

LW LOS—Low loss.

LW WVE—Low waves.

MGHT—Might.

MIKE—Microphone.

MSG—Message.

NITE—Night.

OM—Old man.

OP—Operator.

OPS—Operators.

ORS—Official Radio Station.

OT—Oscillation transformer.

OW—Female operator.

PEEP-PEEP—Continuous wave transmitter.

QCW—Listen for my CW transmitter.

QCW?—Shall I listen to your CW transmitter?

QRMER—An operator who persists in interfering with other stations after being warned to stop.

QSR—I will relay the message.

QSR?—Will you relay the message?

QSS—You are fading.

QSS?—Is my signal fading?

RDO—Radio.

RECVR—Receiver.

RELY—Relay.

ROCKCRUSHER—Spark transmitter.

SA—Say.

SIGS—Signals.

SPIKE—Message file.

SPK—Speak.

SUREFIRE—Something positive in action.

THG—Thing.

THT—That.

TLL—Till.

TM—Traffic manager of A. R. R. L.

TMR—Tomorrow.

TRAFFIC—The handling of messages among the stations.

U—You.

UNLIS—Unlicensed.

UR—Your.

VLTMTR—Voltmeter.

WL—Will.

WRLS—Wireless.

WVL—Wavelength.

"X" LICENSE—Experimental License.

YL—Young lady.

"Z"—Special license.

73—Best regards.

ABBREVIATIONS — RADIO — ELECTRICAL—A letter or group of letters of the alphabet arranged so that there is a definite electrical meaning when used in radio telegraphy or telephony. These abbreviations are the means by which technical writings are simplified, *i.e.*, Alternating Current is simply A.C. when used in general telegraphy. As most of the latest radio text books use these abbreviations, the following are given to aid the reader in understanding the various abbreviations:

A.C.—Alternating Current.

A.F.—Audio Frequency.

B.E.M.F.—Back Electromotive Force.

B. & S.—Brown & Sharpe Wire Gauge.

B.S.W.G.—Birmingham Standard Wire Gauge.

C.—Capacity.

C.G.S. — Centimetre-gramme-second (*q.v.*).

C.W.—Continuous Waves.

D.C.—Direct Current.

D.C.C. — Double Cotton Covered (Wire).

D.S.C.—Double Silk Covered.

E.—Induced Pressure (See Abvolt).

E.M.F.—Electromotive Force.

H.F.—High Frequency.

H.F.C.—High Frequency Current.

I.R.E.—Institute of Radio Engineers.

L.—Inductance (in Formulae).

L.F.—Low Frequency.

L.F.C.—Low Frequency Current.

M.A.—Milliampere.

Mfd.—Microfarad.

P.D.—Potential Difference.

R.—Electrical Resistance.

R.F.—Radio Frequency.

R.P.M.—Revolutions per Minute.

R.P.S.—Revolutions per Second.

S.G.—Specific Gravity.

S.H.M.—Simple Harmonic Motion.

S.S.C.—Single Silk Covered.

S.W.G.—Standard Wire Gauge.

W/L—Wavelength.

X's—Atmospherics.

ABBREVIATIONS, INTERNATIONAL RADIO TELEGRAPHIC CONVENTION—All messages and conversations are not transmitted in regular form, various forms of abbreviations being widely used to increase the speed of transmission and to reduce cost. It also has the tendency to decrease the interference as messages are sent swiftly in half the time. In wireless telegraphy, abbreviations have been adopted as international, and today there are very few ships or shore stations that do not use at least part of them daily. Below we give the International Radiotelegraphic Abbreviations as adopted at The International Radiotelegraphic Convention. An example of the use of the abbreviations is as follows:

A station wishing to learn the name of a ship which is seen off the coast, sends in code the letters QRA?, which means, what is the name of your ship? The station or ship answers with the letters, QRA S.S. Lapland, thus in a few seconds the question had been answered.

Abbreviation	*Question*	*Answer or Notice*
CQ	—.—.—.—.—	Signal of inquiry made by a station desiring to communicate.
TR	— .—.	Signal announcing the sending of particulars concerning a station on shipboard.
	— —. .— —	Signal indicating that a station is about to send a high power.
PRB	Do you wish to communicate by means of the International Signal Code?	I wish to communicate by means of the International Signal Code.
QRA	What ship or coast station is that?	This is ....
QRB	What is your distance?	My distance is ....
QRC	What is your true bearing?	My true bearing is.........degrees.
QRD	Where are you bound for?	I am bound for.............
QRF	Where are you bound from?	I am bound from............
QRG	What line do you belong to?	I belong to the.............Line.
QRH	What is your wave length in meters?	My wave length is..........meters.
QRJ	How many words have you to send?	I have .... words to send.
QRK	How do you receive me?	I am receiving well.
QRL	Are you receiving badly? Shall I send 20? ...—. for adjustment?	I am receiving badly. Please send 20. ...—. for adjustment.
QRM	Are you being interfered with?	I am being interfered with.
QRN	Are the atmospherics strong?	Atmospherics are very strong.
QRO	Shall I increase power?	Increase power.
QRP	Shall I decrease power?	Decrease power.
QRQ	Shall I send faster?	Send faster.
QRS	Shall I send slower?	Send slower.
QRT	Shall I stop sending?	Stop sending.
QRU		I have nothing to transmit. I have nothing for you.
QRV	Are you ready?	I am ready. All right now.
QRW	Are you busy?	I am busy (or, I am busy with.... Please do not interfere).
QRX	Shall I stand by?	Stand by. I will call you when required.
QRY	When will be my turn?	Your turn will be No. ....
QRZ	Are my signals weak?	Your signals are weak.
QSA	Are my signals strong?	Your signals are strong.
QSB	Is my tone bad? Is my spark bad?	The tone is bad. The spark is bad.
QSC	Is my spacing bad?	Your spacing is bad.

(When an abbreviation is followed by a mark of interrogation, it refers to the question indicated for that abbreviation)

ABSCISSA—A term in geometry adapted to radio use in making *curves* to show various values. The abscissae are the horizontal lines, and the ordinates are the vertical lines. In the diagram, the line AY is the axis of *ordinates* (*q.v.*) and the line AX the axis of abscissae. Fig. 2 shows a simple curve indicating the relation between voltage and current (amperes) in an *arc* (*q.v.*). The volts are given as ordinates and the amperes are shown as abscissae.

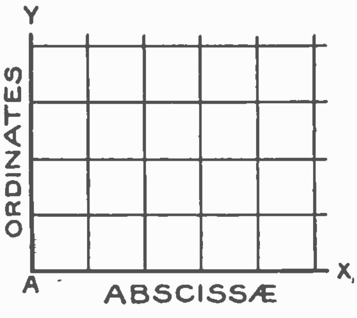

Fig. 1. Indicating axis of abscissae by the line A to X

The distance of any point from the axis of ordinates, measured on a line parallel to the other axis represents the value of the quantity and locates a point on the curve. Where abscissae and ordinates are known, it is a simple matter to plot a series of points and then draw the curve through these points. (See *curve.*) Thus, in Figure 2, the curve shows the operation of the arc to be such that when the voltage is zero, the current is 20 amperes; when the voltage is 60, the current is 6 amperes, etc.

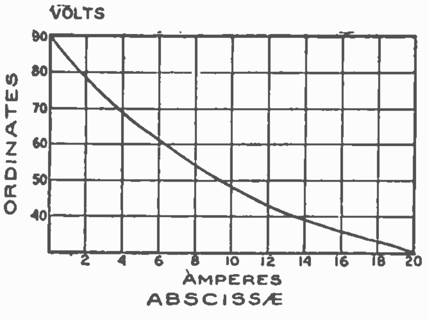

Fig. 2. A simple curve to illustrate the relation between volts and amperes

ABSORPTION, ELECTRIC—(See *Soaking-in.*)

ABSORPTION MODULATION—In the production of modulated (see *modulation*) waves, used to transmit speech in radio, it is necessary first to produce *undamped* (*q.v.*) waves of high or *radio frequency* (*q.v.*). Then other waves of low or *audio frequency* (*q.v.*) are superposed or run together with them. Now, if we insert a *microphone* (*q.v.*) in the *aerial circuit* (*q.v.*) of the transmitting station, providing that the *resistance* (*q.v.*) of that instrument is such as to reproduce accurately the vibrations of the voice or music, it is possible to vary the current output of the system producing the high frequency currents, so as to transmit the voice vibrations. If the microphone has resistance comparably low with the antenna or *antenna resistance* (*q.v.*), then the high frequency power output is handled (apportioned) equally between the aerial and microphone. The result is the production of fairly well modulated waves. There are, however, better methods of modulating the waves for which see *Modulation.*

ABVOLT—The *CGS unit* (*q.v.*) indicating the pressure or induced E.M.F. (*electromotive force*) (*q.v.*) set up between the ends of a wire one centimeter long, moving with a speed of one centimeter per second across the magnetic lines of force (*flux lines*) (*q.v.*) of a uniform magnetic field of unit intensity or flux density. An abvolt is therefore a very small unit and 100,000,000 (10^8) are required to equal one volt. In all formulae for determining the output or E.M.F. that can be expected to be produced by a certain *alternator* (*q.v.*), the result is in abvolts, and the complete formula usually provides for obtaining the actual volts. This is done by dividing the result in abvolts by 100,000,000 or 10^8 as it is generally expressed. (See *Alternator* also *abampere.*)

A. C.—Abbreviation for Alternating Current. (See *Alternating Current.*)

ACCELERATION—The rate at which the speed of a body in motion increases within a certain period of time. If we assume a train running at a certain rate of speed when the brakes are applied, the speed will decrease, and if measured over a period of one second, there will be a definite amount of decrease or a rate of decrease. This is known as *negative* acceleration or *deceleration.* When the train is speeded up there will be a definite change (increase) in velocity. The rate of change of velocity per unit of time is the acceleration. (*See Velocity* also *Frequency.*)

ACCEPTOR—A supplementary combination of *inductance* (*q.v.*) and *capacity* (*q.v.*) tuned to the frequency of the desired signal and connected in series in the receiving antenna. Generally, a *coil* (*q.v.*) and *condenser* (*q.v.*) of a value to correspond with a certain given wavelength—thus literally to "accept" the desired signal. (See also *Rejector.*)

A series combination of *inductance* (*q.v.*) and *capacity* (*q.v.*) connected in series with a high-frequency C.M.F. When tuned to resonance with the impressed E.M.F., the impedance drops to a low value, limited by the resistance of the system

ACCUMULATION OF ELECTRICITY—The process of storing electrical energy as (1)—a *storage battery* (*q.v.*) in which electric charges are placed to be later withdrawn and expended at will for various purposes, or (2)—the collection of energy by a *condenser* (*q.v.*) wherein the *electrons* (*q.v.*)

or minute charges of electricity are stored up momentarily on the plates or electrodes of the *condenser* (*q.v.*) and then released into the circuit.

ACCUMULATOR—(See *Storage Battery.*)

ACID—An active chemical compound formed by combining hydrogen with various acid radicals. Acids are much used in radio, particularly in *storage batteries* (*q.v.*). The most commonly used acid in this connection is sulphuric acid—a compound or union of hydrogen, sulphur and oxygen in certain definite proportions as indicated by the chemical symbol H_2SO_4. (See *Flux, Soldering.*)

ACIDIMETER or ACIDOMETER—An instrument used to determine the *specific gravity* (*q.v.*) of an acid solution. It is similar to a hydrometer and used in the same manner. Much used for testing strength or purity of acids or solutions. Operates on the principle that the strength of an acid will be directly proportional to the quantity of carbonic acid gas which it will liberate from a carbonate of soda or potash. (See *Hydrometer.*)

ACLINIC LINE—An imaginary line on the earth's surface assumed as passing through points where there is no magnetic inclination or dip of the needle of a compass. (See *Agonic Line.*)

ACOUSTICS—The science of sound or the study of the cause and effect of vibrations which effect the hearing. Generally speaking, the production and transmission of sound.

ACOUSTIC WAVE—The term occasionally used to denote a *sound wave.*

ACTINIC RAY—A light ray or beam of invisible *radiant energy* (*q.v.*). Rays of light, generally considered as lying at the blue or extreme left of the *spectrum* (*q.v.*) and being of such short *wavelength* as to be invisible to the eye. These rays have the power to induce or bring about chemical action. The most powerful actinic rays are the violet (X Ray) and the *ultra-violet.*

ACTIVE CONDUCTOR—(See *Conductor.*)

ACTIVE MATERIAL—The spongy part of a plate of a storage battery. The part which changes in nature and appearance due to the flow of electric current and having the ability to redevelop the current by a secondary chemical change.

ACTIVE PRESSURE—The active *electromotive force* (*q.v.*) or the pressure which produces a current. The term is used to distinguish this pressure from one impressed on the circuit.

A component in phase with the current in an *alternating current* (*q.v.*)

ACTIVE SPARK—A spark produced from energy contained in a charged condenser, which produces active *oscillations* (*q.v.*).

Thus, the spark discharge from a coil without any method of storing energy will be "inactive." When a condenser is inserted in such manner as to allow it to store up energy and then discharge, the resultant spark will be "active."

Fig. 1 shows a simple spark coil circuit with the secondary discharging across a gap. This is an Inactive Spark. Fig. 2 represents the same coil with a condenser in the circuit to store up the energy before discharging across the *gap* (*q.v.*). This is an Active Spark.

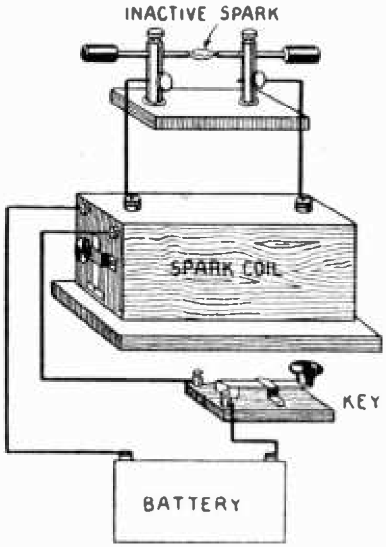

Fig. 1. Showing how an inactive spark is produced

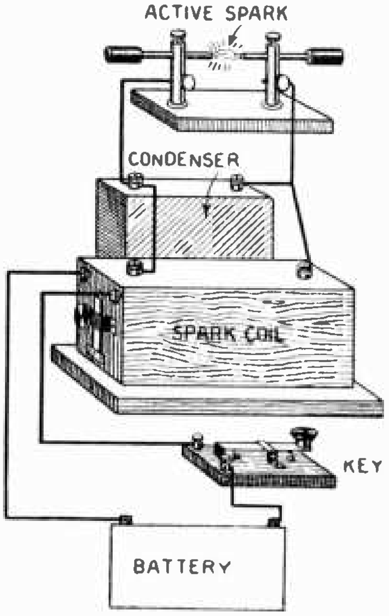

Fig. 2. The active spark is produced by means of a condenser in the circuit as shown

ADAPTER, AERIAL—A device for utilizing an electric light line as an aerial for the reception of radio signals. Essentially, two or more fixed mica condensers of comparatively high capacity, in series with two metal legs attached rigidly to a standard plug for insertion in socket of electric light fixture. Terminals are provided for connection to receiving apparatus. The illustration shows one of the popular types of Aerial Adapters which consists of merely a condenser to prevent the voltage of the line from passing directly to, or through, the instruments.

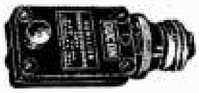

Illustration by courtesy of Dubilier Condenser and Radio Corp.

ADAPTER, TUBE—A device similar in form to a vacuum tube socket. Usually comprises a cylindrical form or shell of either metal or insulating material with legs identical with the tips or legs of a vacuum tube. A receptacle is allowed for the insertion of a tube, and the lead from each tip recess is so connected to a leg or tip at its base, as to permit the use of a non-standard base tube in a standard *socket* (*q.v.*). The illustration shows adapter for U.V. 199 or C-299 type of tube. When it is desired to employ a tube having a non-standard base in a set already equipped with standard sockets, without the necessity of changing the sockets, an adapter may be used. (See *Socket.*)

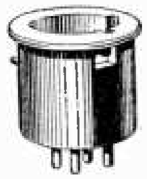

Illustsation by courtesy of Pacent Elec. Co., Inc.

ADHESION, ELECTRIC—Affinity of one body for another due to dissimilar charges of electricity passing through or being carried by them. The electrical equivalent of magnetic attraction.

ADJUSTABLE CONDENSER—(See *Variable Condenser.*)

ADJUSTABLE GRID LEAK—A form of *grid leak* (*q.v.*) that can be easily adjusted over a wide range of values without removal from the circuit. The illustration shows one type of Adjustable Grid Leak.

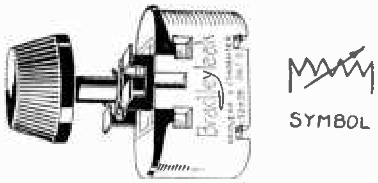

Illustration by courtesy of Allen-Bradley Co.

ADJUSTABLE RESISTANCE—(See *Variable Resistance.*)

ADMITTANCE—The inverse of *impedance* (*q.v.*). In an *alternating current* (*q.v.*) the Admittance acts in the opposite manner to impedance. A circuit having low impedance is said to have relatively high Admittance.

AERIAL—A system of wires suspended in the air or in any form in which they may be insulated or kept free from surrounding objects, used for the purpose of receiving or transmitting impulses as in radio transmission and reception. The term Aerial is actually identical with *Antenna*, but due to the general English definition of Antenna, authorities are gradually coming to distinction between transmission and reception Aerial systems by referring to a receiving device of this nature as an Antenna, and when used to transmit, as an Aerial.

There are different types and forms of Aerials, varying mainly according to purpose and facilities. The standard and perhaps more commonly used receiving Aerial consists of a single wire as in Fig. 1 or several wires as in Fig. 2. The wire used is usually of copper, as this metal combines low resistance with economy, strength, and durability. The theory of the action of the waves either received by or transmitted from an Aerial system will be found under the heading "Electromagnetic Waves."

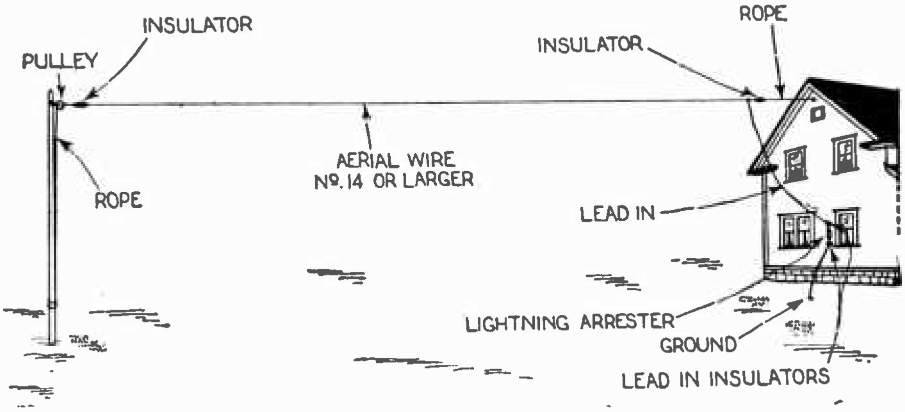

Fig. 1. A single wire aerial in the form of an inverted L. If a long wire is used, the lead-is may be connected in the center, forming a T type

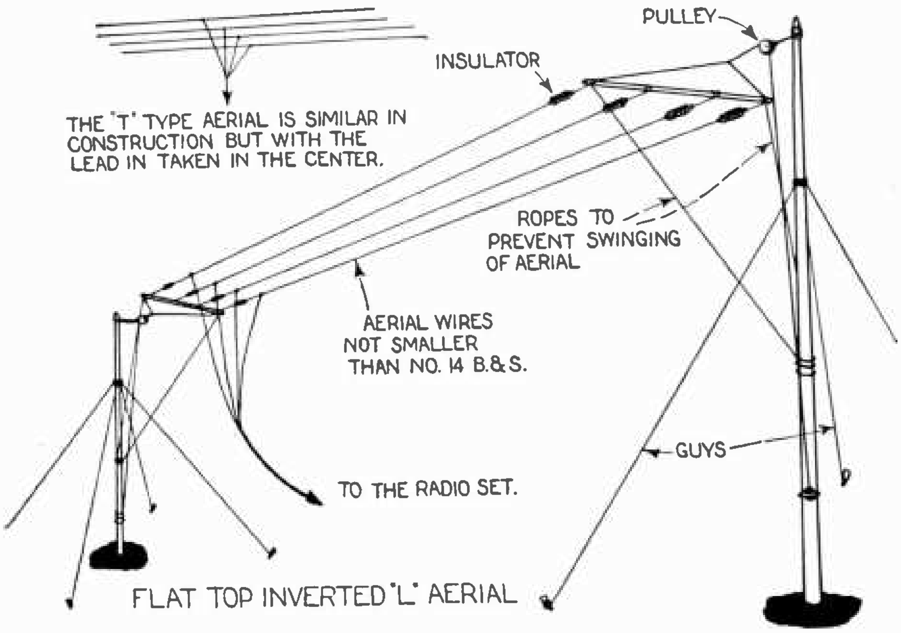
PULLEY
INSULATOR
THE 'T' TYPE AERIAL IS SIMILAR IN CONSTRUCTION BUT WITH THE LEAD IN TAKEN IN THE CENTER.
ROPES TO PREVENT SWINGING OF AERIAL
AERIAL WIRES NOT SMALLER THAN NO. 14 B.&S.
GUYS
TO THE RADIO SET.
FLAT TOP INVERTED 'L' AERIAL

Fig. 2. Method of erecting a multi-wire aerial. This type of aerial may be made in either the inverted L or T form as shown

The exact type and extent of an Aerial will depend entirely on the available facilities and the particular purpose for which it is intended. These various forms will be taken up under their respective headings. In the receiving Aerials, Figs. 1 and 2, a certain definite *wavelength* (*q.v.*) is obtained by arranging a certain length to the wires of the Aerial system including the *ground* and *lead-in.* (See *Fundamental Wave-Length of Aerials.*) In the single wire Aerial, the lead-in is usually taken from the end nearest the receiving apparatus, but where a particularly long Aerial is to be used, it is customary to attach the lead-in to the centre of the suspended wire. The former is known as an *inverted L* type and the latter as a *T* type Aerial. These designations will apply as well where a number of wires are used as in Fig. 2.

Generally speaking, an Aerial should be as high above the ground as possible, of course within practical limits. From 25 to 100 feet is the most practical height to use. Where the Aerial is suspended at a substantial height the incoming waves will be less likely to be obstructed by surrounding trees or buildings, and in the case of transmission the same factor will apply, inasmuch as the waves may be obstructed in their transmission, if buildings or trees closely surround the Aerial.

The insulation of an Aerial is of particular importance, and especially so in the case of a transmitting Aerial. The reason for insulating an Aerial is to prevent the currents from escaping to the ground by way of trees, buildings, or the masts to which the Aerial is attached.

Fig. 3 shows a method of arranging an indoor Antenna system where it is not possible to use the out-door type. The construction may be essentially the same as the usual form of outdoor Aerial and can be arranged in a number of different ways. Very often a wire is merely suspended around two sides of a room and insulated from the moulding. One end will be connected as a lead-in to the receiving apparatus. In constructing an indoor Antenna, it is well to remember that the effective length of the Antenna is not much greater than the direct distance from the receiving set to the farthest point of the Antenna.

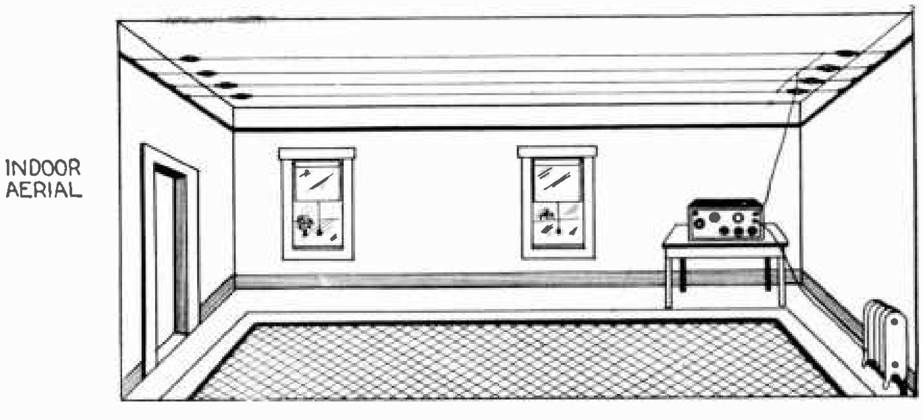

Fig. 3. Showing one method of installing an indoor aerial. While this illustrates a multi-wire type, a single wire run around the moulding of the room may give excellent results

Fig. 4 shows a different type of Aerial known as the *Cage* type. The illustration shows an inverted L, and the T arrangement. Fig. 5 shows a *Fan Aerial* and Fig. 6 the *umbrella* type. The illustration, Fig. 7, shows one type of *loop Aerial* (*q.v.*) and Fig. 8 the same general type in somewhat different form. Owing to the fact that the theory of operation of a *loop Aerial* is somewhat different than the ordinary types, it will be described more fully under the heading "Loop Aerial." Fig. 9 illustrates a form of *Aerial Adapter* which is arranged to allow the electric light wires to be used as an Aerial. (See *Adapter Aerial.*) Fig. 10 shows an *underground Aerial* system designed by Dr. Rogers and which is not appreciably affected by atmospheric disturbances. (See *Strays.*) This form of Aerial, however, is generally impractical due to the difficulties in construction.

Aerials for transmitting may be of the same general types as receiving Antennas arranged outdoors, with the exception as explained previously, that heavy insulators must be used owing to the high voltage employed. Aerial for transmitting will depend for size on the wave-length desired.

Fig. 11 illustrates several types and arrangements of *counterpoise* (*q.v.*). In the case of the receiving Aerial a single wire may be from 75 to 150 feet in length, or if a number of wires are used, the over-all length may be considerably less. (See *Fundamental Wave-length.*)

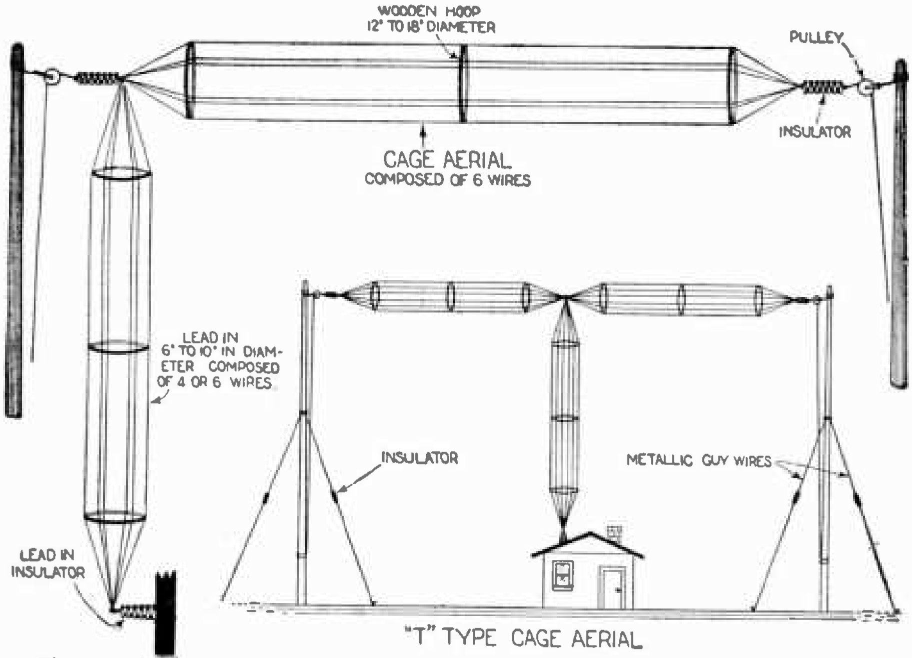
WOODEN HOOP
12" TO 18" DIAMETER
PULLEY
INSULATOR
CAGE AERIAL
COMPOSED OF 6 WIRES
LEAD IN
6" TO 10" IN DIAMETER COMPOSED OF 4 OR 6 WIRES
LEAD IN INSULATOR
INSULATOR
METALLIC GUY WIRES
"T" TYPE CAGE AERIAL

Fig. 4. Cage aerials are usually employed for transmitting purposes. The inverted L and T types are erected as shown

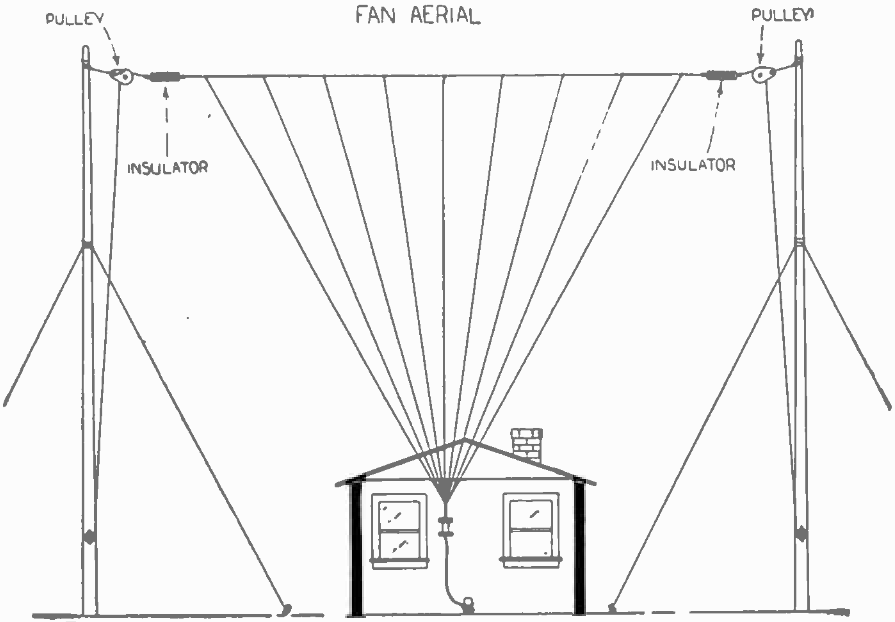
PULLEY
FAN AERIAL
PULLEY
INSULATOR
INSULATOR

Fig. 5. The fan aerial has its wires spread apart at the farthest point, forming a fan

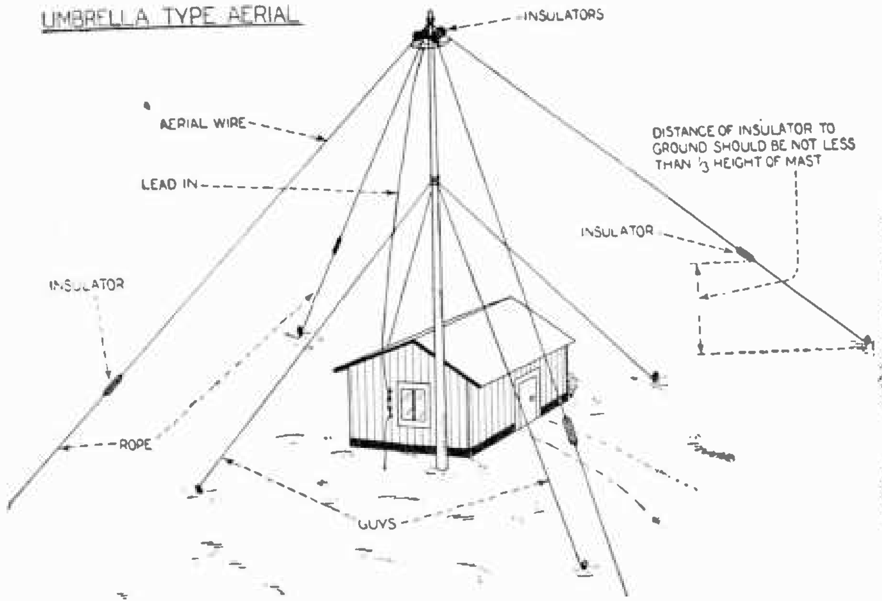
UMBRELLA TYPE AERIAL
INSULATORS
AERIAL WIRE
LEAD IN
DISTANCE OF INSULATOR TO GROUND SHOULD BE NOT LESS THAN ⅓ HEIGHT OF MAST
INSULATOR
INSULATOR
ROPE
GUYS

Fig. 6. The umbrella type of aerial derived its name from the arrangement of its wires in the form of an umbrella

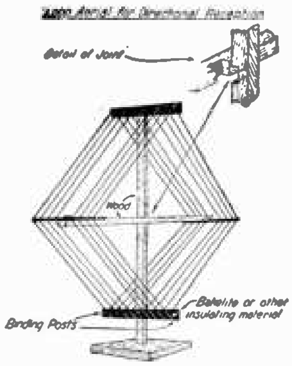
Loop Aerial for Directional Reception
Detail of Joint
Wood
Bakelite or other insulating material
Binding Posts

Fig. 7. The selenoid type loop aerial

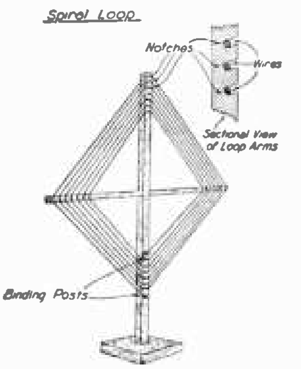
Spiral Loop
Notches
Wires
Sectional View of Loop Arms
Binding Posts

Fig. 8. The spiral type loop aerial

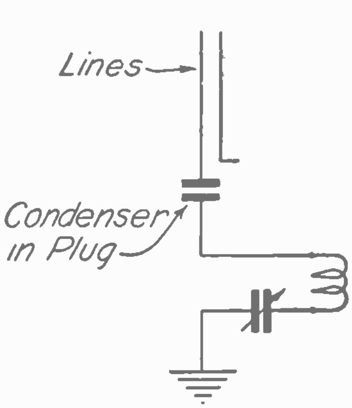
Lines
Condenser in Plug

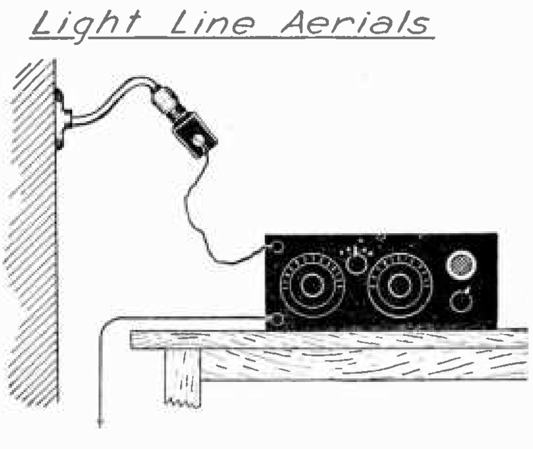
Light Line Aerials

Fig. 9. Method of employing the electric light line as an aerial with an aerial adapter. Above is shown the diagram of the condenser in connection with the aerial circuit of a radio set

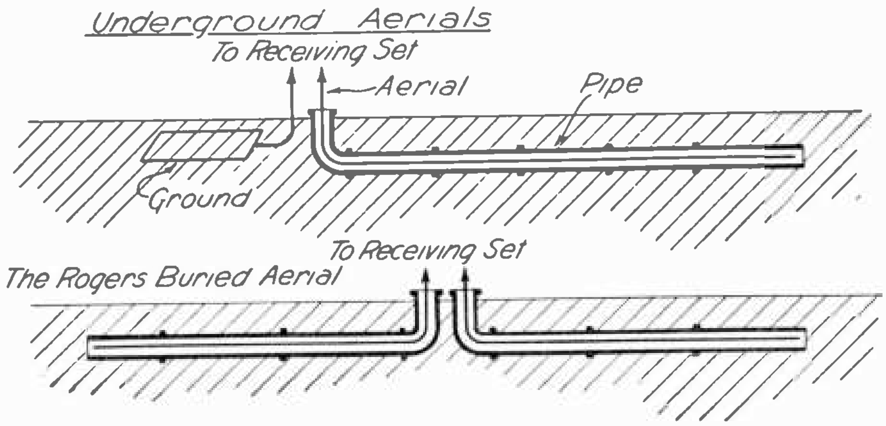

Fig. 10. Showing how wires are buried in the ground for aerials

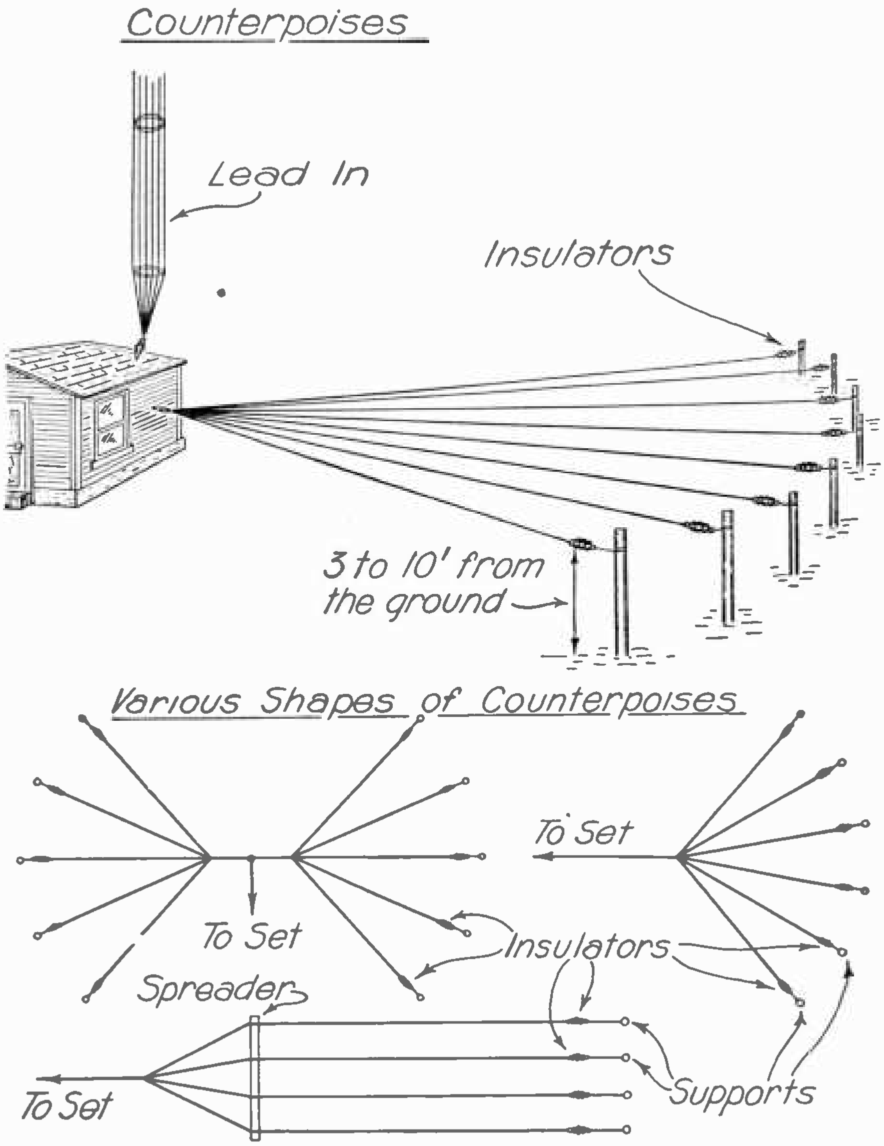

Fig. 11. Different methods of arranging counterpoises for either transmission or reception

AERIAL CAPACITY—(See *Capacity of Antenna.*)

AERIAL CIRCUIT—Consists of *aerial* and *earth* or *ground*, including all *coils* or *inductances*, *condensers*, etc., which may be connected with the aerial and earth and forming a direct path between those points.

AERIAL SWITCH—A device to transfer the *aerial* and *ground* connections from the transmitting circuit to the receiving circuit or vice versa. Its object is to separate the high tension current of the transmitter from the low tension of the receiver. The switch is usually installed in a convenient position so that the operator may change from transmitting to receiving by one motion of the hands. A simple form of aerial switch is illustrated in Fig. 1. The circuit for same is shown in Fig. 2. A more efficient circuit is the one shown in Fig. 3. This circuit requires another blade to control the primary current to the transformer. The latter circuit is most satisfactory as it entirely disconnects the current to the transformer, doing away with the possibility of dangerous shocks to the operator. The switch commonly used in such circuits is shown in Fig. 4.

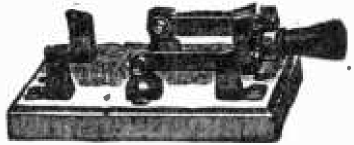

Fig. 1. A simple Aerial Switch

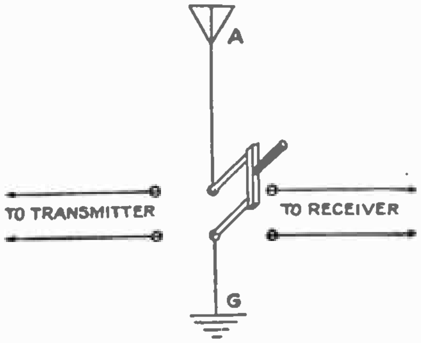

Fig. 2. Circuit for a simple Aerial Switch

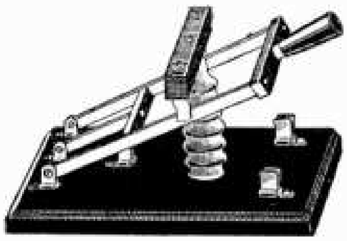

Fig. 4. An Aerial Switch with an extra blade

The switches used in the various commercial wireless stations are of course more complicated. There is a need to control many circuits with the simplest possible method. In order to do this it was found more convenient to only change the aerial connections, and allow the ground connection to both the receiver and transmitter to remain permanently. This of course necessitates a small air gap in the aerial for the transmitter. This is called an *anchor gap* (*q.v.*). This small appliance allows the transmitter to be connected continuously to the aerial as the high voltage from the transformer discharges across the small air gap, as provided for in the *anchor gap* (*q.v.*). Fig. 5 shows the circuit when this type of aerial transfer is to be used. When the switch handle is pushed down it makes connections with the receiver, when it is raised, the transmitter is ready to be used as the primary current is connected also.

AERIAL TUNING CONDENSER—A *condenser* (*q.v.*), usually variable, connected in the *aerial circuit* (*q.v.*) for the purpose of adjusting the natural period of the receiving circuit to the period of the incoming waves (see *Wavelength*) or signals coming from the *aerial*. In the illustration is shown an Aerial Tuning Condenser.

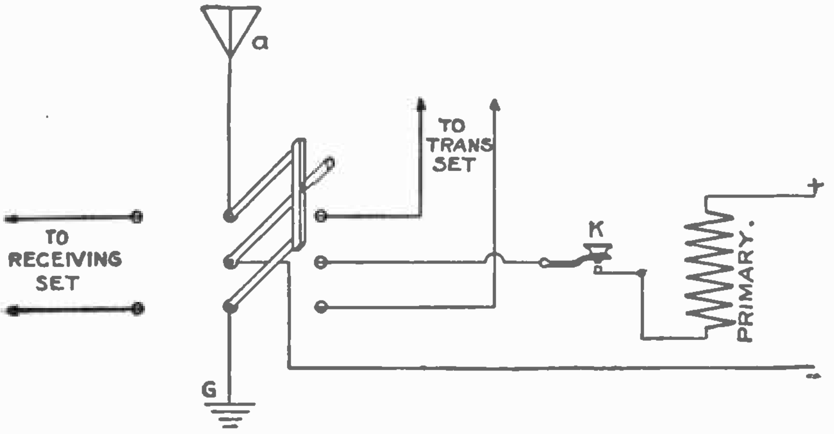

Fig. 3. Circuit for the Aerial Switch shown in Fig. 4 on page 7

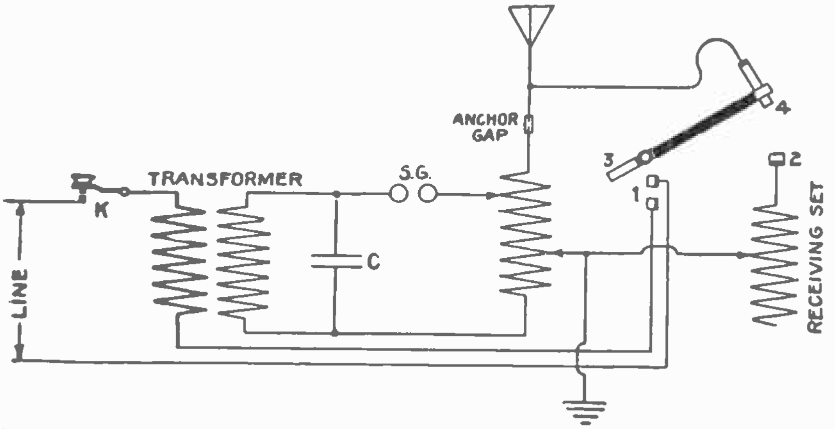

Fig. 5. An Aerial Switch employed in conjunction with an Anchor Gap as explained on page 7

Any *capacitive* (*q.v.*) means of establishing or varying the *oscillation constant* (*q.v.*) of the receiver.

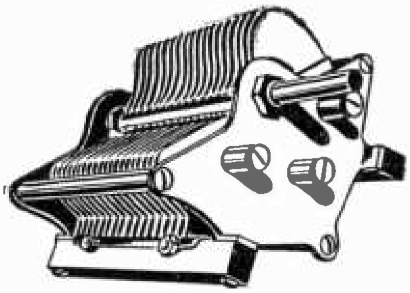

A variable condenser such as commonly used for an Aerial Tuning Condenser (see page 7).

AERIAL WIRE—The wire forming the *aerial* (*q.v.*). Copper is generally used for this purpose as it offers very little *resistance* (*q.v.*) to the feeble currents in receiving.

The illustrations show various types of wire as used for aerials. Among them there are the solid copper wire, copper clad steel wire, stranded wire, braided wire and brass or copper strip.

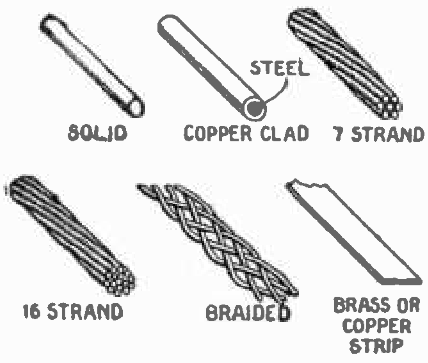

AERO-FERRIC INDUCTANCE—The *inductance* (*q.v.*) of coils wherein the *magnetic circuit* (*q.v.*) is completed through both air and iron.

AEROPLANE ANTENNA—Receiving or transmitting *antenna* (*q.v.*) used on an aeroplane. The commonest form is a suspended wire shown as A in the illustration. The *ground* or *counterpoise* (*q.v.*) in this case may be another wire suspended on and insulated from the fuselage as indicated at B. When the plane is in flight, the hanging wire trails up and backward, depending on the speed of the plane and the size of the weight used. This type of antenna will be slightly *directional* (*q.v.*) in the direction the plane is traveling. If it is impractical to use a trailing wire antenna for any reason, a loop or coil of wire may be used, the results of course being less satisfactory. The metal frame of the plane is often used as a *counterpoise* (*q.v.*)

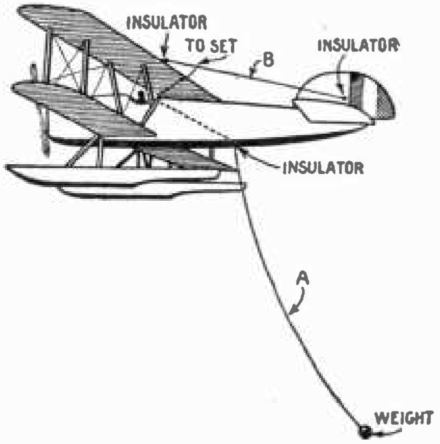

A. F.—Abbreviation for *Audio-Frequency* (*q.v.*)

AFTER-GLOW—Fluorescent phenomena in a *vacuum tube* after current has been withdrawn.

AGONIC LINES—Lines imagined as passing through points on the earth's surface where the magnetic inclination or dip of a magnetic needle (compass) is zero. (See *Aclinic Lines.*)

AGING, VACUUM TUBE—Gradual diminishing of the brilliancy of a tube due to the deterioration of the *filament* (*q.v.*) and the coating or deposit on the bulb. A tube is said to be Aging when it gradually loses its power to emit *electrons* (*q.v.*)

AIR CONDENSER—*Condenser* utilizing air as the *dielectric*. The majority of variable condensers used in radio work employ the air dielectric, that is, the space between plates is merely an air gap. Most *fixed condensers* (*q.v.*) use some substance such as mica, or sheets of waxed paper as a dielectric.

AIR CORE TRANSFORMER—A *transformer* used in radio, having no metal core, *the lines of force* having their path through air. (See *Amplifier, Radio Frequency.*)

ALBIZ, Count—Managing Director of the Compañia National de Telegrafia sin Hiols in 1910, of Scotch origin.

He was born in Madrid in 1858 and received a schooling at the Madrid University and London University College. He was a Tory member of Parliament.

ALEXANDERSON, Ernst Fredrick Werner (1878)—Radio engineer and inventor. Born at Upsala, Sweden, January 25th, 1878, he was educated at the High School and University of Lund, Sweden, at the Royal Institute of Technology, Stockholm, and at Berlin. In 1902 he joined the General Electric Company, and has been for some years their consulting engineer. He holds the post of chief engineer to the Radio Corporation of America, and

Photo by courtesy of General Elec. Co.

E. F. W. Alexanderson

is a member of the American Institute of Radio Engineers. Alexanderson has read many papers on electrical subjects before the chief technical societies of America.

Alexanderson is famous for his work on *high-frequency alternators* used in Radio telegraphy. The Alexanderson alternator is connected directly or *inductively* to the aerial and earth, and constitutes the simplest possible connection for producing *continuous* or *undamped waves*. It is a machine of great speed and many *field poles*, and frequencies as high as 200,000 *cycles* are obtained. For long distance work, employing very *long wavelengths*, this high-frequency alternator is largely used. Alexanderson is also responsible for the *magnetic amplifier*, patented in 1913, and has carried out successful experiments on *duplex wireless telephony*. In the Alexanderson *microphone transmitter* the *modulation* is mainly effected by variations in the tuning of the *aerial circuit*.

ALEXANDERSON ALTERNATOR—A type of *high-frequency alternator* (*q.v.*) which produces *continuous oscillations* (*q.v.*)

ALLOY—A compound of two or more metals. (See *Woods Metal.*)

ALTERNATING CURRENT—An electric current that does not flow steadily in one direction—as from *positive* to *negative*—but completely reverses its direction or *polarity* (*q.v.*) at certain definite intervals. In other words, the current flows first in one direction and then reverses and flows in the opposite direction—these changes being referred to as changes of *phase* (*q.v.*). An electric current is either *direct* (*q.v.*), alternating, or *pulsating*. The latter is an alternating current that has been rectified. (See *Rectified Current.*) Practically the only direct current that is employed or appears in a radio receiver is that supplied by the *"A"* and *"B" Batteries* (*q.v.*). In an alternating current, the change in direction or polarity takes place in a steady, even manner.

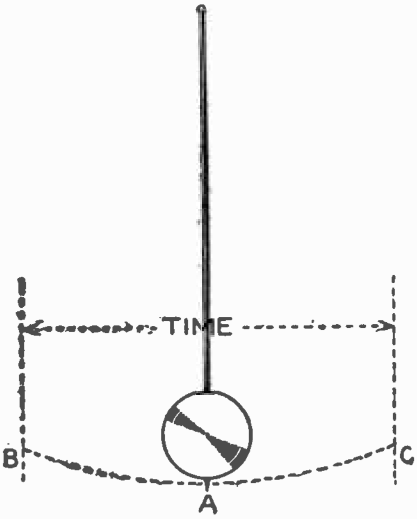

Fig. 1

An understandable analogy for alternating current is its comparison with a pendulum. In Fig. 1 the pendulum is supposed to be the alternating current or impulse. Now when the bob is swung from its natural (neutral) position A, which may be considered as zero value of voltage or current, it will move to a maximum point, either B or C. If we assume B to be the *positive* side, then the bob will move to B which will be the maximum positive *amplitude* (*q.v.*). It will then fall back to zero again, and with the power behind it, the momentum in this case, will rise on the other side to C, which is the maximum *negative* amplitude. The bob will then fall back to A once more and will have undergone a complete change of direction and values; it will have gone through a *cycle* (*q.v.*). That is to say, it will have risen from zero A to positive maximum B, then back to A, rising on the other side to negative maximum C and back to A. Now the length of time required for this complete cycle (see *Alternation*), is known as the *time period* of vibration or *phase change*. If, for example, it required one-sixtieth of a second for this change to take place, then we say that the *frequency* (*q.v.*) is sixty cycles, because obviously, a change requiring one-sixtieth of a second will occur sixty times in one second.

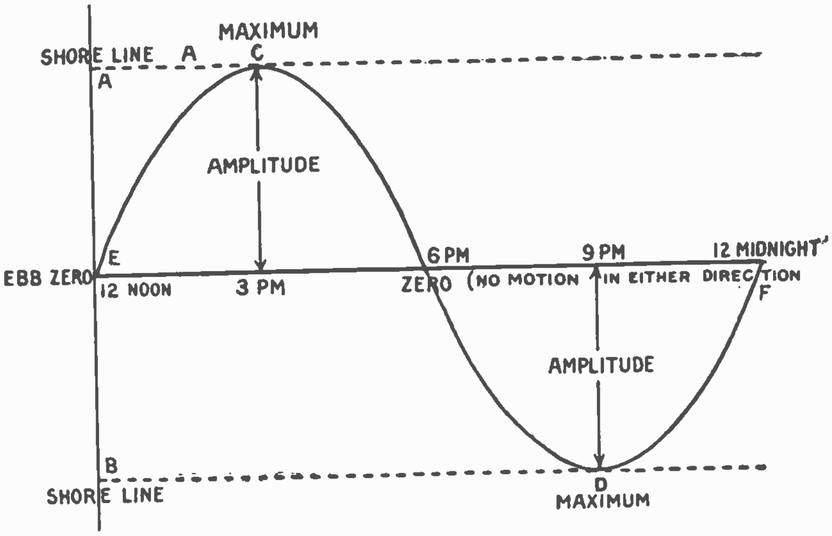

Fig. 2

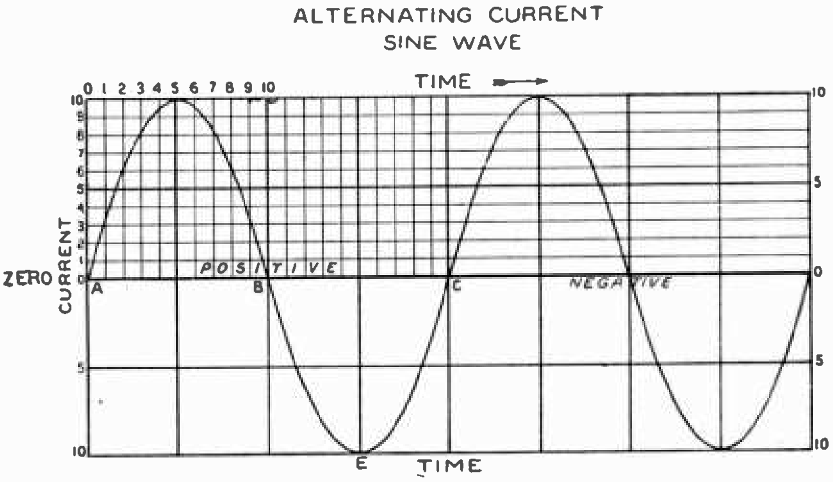

Fig. 3

The action of alternating currents may also be likened to that of ocean tides. In Fig. 2 the curved line represents the water in motion and is to be compared with current in motion in an alternating circuit. For the sake of clarity, we can assume that the tide rises and falls between two shore lines, A and B. At 12 noon the tide is at zero or "ebb" and there is no motion in either direction. Now the flow starts toward shore line A and gradually rises toward that side until at 3 P. M. it has reached its maximum height. In the case of an alternating current, we may consider that at 3 P. M. the current has reached maximum value on the positive side or in a positive direction. It then reverses and falls toward zero until at 6 P. M. it has reached the zero point. In the case of the current, it will have become zero. Motion again starts in the other direction and the tide rises once more, but in the opposite direction, toward shore line B. By 9 P. M. it will have reached maximum height in this direction, or in the case of the current it has reached maximum value or *amplitude* (*q.v.*) at D on the negative side. It then falls back again to zero, and at 12 midnight has come back to its original starting point, having gone through all stages of value or height from zero to maximum at one shore and then back to zero; then from there to maximum at the other shore and back to zero. The action of the alternating current then will have been to flow from zero to positive maximum, back to zero and reversing, from zero to maximum on the negative side, then back to zero. It will have completed one cycle. The time required for the tide to go through these changes E to F is 12 hours and is known as the *time period* or *period* (*q.v.*). Therefore, the time required for a complete change in the case of alternating current will be referred to as the period. The number of times that these changes take place within a certain time limit is known as the *frequency* (*q.v.*)

The current value in Fig. 3 is assumed to be 10 amperes, and the current values at any moment are shown as *ordinates* (*q.v.*), (vertical) while the time is shown in the *Abscissas* (*q.v.*) (horizontals). If the above curve is to be assumed as representing the Sine wave of an Alternating current having a frequency of 60 cycles per second, the total time in the chart would be one-thirtieth of a second. Thus, Time A to C would be one-sixtieth second, the Time of one complete cycle. A B is a positive Alternation and B C a negative Alternation, D and E represent maximum current values in each direction.

If the above Sine Wave is to be used as an example of a high frequency oscillation as in radio, the Time of one cycle can be assumed as about $\frac{1}{1,000,000}$ second. (See *Frequency-Oscillation - Sine - Phase - Characteristic of Alternating Current.*)

The frequency of an alternating current depends on the number of poles (field magnets) and the speed of rotation when the A.C. is generated by an alternator. (See *Alternator.*)

ALTERNATING CURRENT, THEORY OF PRODUCTION OF—Electric current in motion along a conductor always has a magnetic field (*magnetic flux*) surrounding it. If this current is passed near a number of turns of wire, the lines of flux will link with the wire turns in the circuit. The number of flux lines through a coil will depend upon the current. A change of current will change the number of linkages. If two turns of wire are used, the circuit will link twice with the same magnetic flux. Thus, a change in the current or a change in the number of turns of wire will alter the linkages.

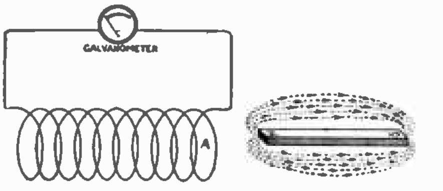

Fig. 1

Now in the case of Fig. 1, the magnet being placed near the coil causes lines of force to pass through the coil. Each turn of wire will link the flux lines. Now any *change* in these linkages between the flux lines and wire turns will produce an E.M.F. (*electro motive force*) in the circuit of the wire turns. This is called an "induced current" because it is induced by the change in linkages. In Fig. 1, we have a coil of wire in a circuit with some sensitive measuring instrument for minute currents, in this case a *galvanometer*. Near it we have a bar magnet. When this magnet is moved toward or within the coil, a current is induced in the circuit by reason of the change in linkages. This current is *always* in a direction to *oppose* the change that causes it. (See *Lenz Law.*) "Whenever an induced current arises by reason of some change in the linkages, the magnetic field about the induced current is in such direction as to oppose the change." When the magnet is moved toward the coil, the change in linkages creates or induces a current that flows in the circuit in such manner as to make A act as a North pole. As the North pole of the magnet is toward this point A, the similar poles will oppose each other according to the simple law of magnetic attraction and repulsion. Now if we attempt to withdraw the magnet, the change will alter the linkages again, this time creating an induced current that makes A act as a South pole, thus *attracting* the magnetic North pole and opposing in the opposite manner, the change in linkages that have created the current.

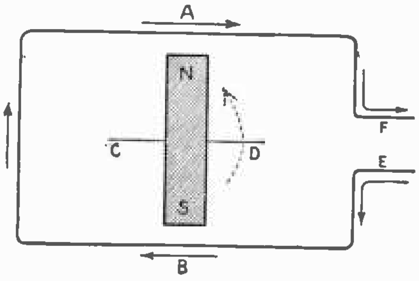

Fig. 2

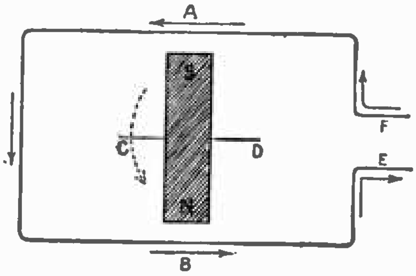

Fig. 3

We already know that if a conductor is moved across a magnetic field, an E.M.F. is induced in the conductor. The action is the same whether the conductor is moved across the magnetic field or vice versa. In Fig. 2, a magnet is shown in position on an axis—C D for rotation within the loop A B. As the magnet is rotated, the North pole passes across A, an E.M.F. is created toward F in the same direction. As the position of the magnet reverses in rotation and S passes A while N passes B, a current is induced in the *opposite* direction, both times in accordance with the law previously stated. Thus, if a meter is connected between the ends of the loop at E F and the magnet rotated, a current will be registered. With each half turn of the magnet, this current will reverse its direction. Therefore, we have produced an alternating current that changes direction with each reversal of the position of the magnets. This is the simple *AC generator* or *alternator.* The illustration, Fig. 4, shows a standard type of low frequency alternator.

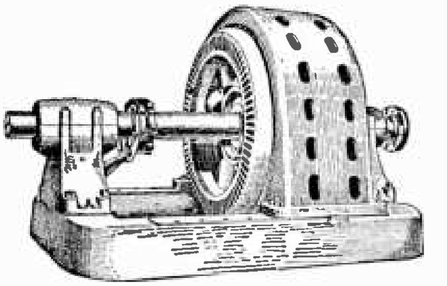

Fig. 4

ALTERNATION—In an alternating current, is the rise from zero to maximum *amplitude* and back to zero on either side, i.e., *positive* or *negative* directions. Two alternations—one positive and one negative—make a complete *cycle.* (See *Alternating Current.*)

ALTERNATOR—A machine for the purpose of changing mechanical energy into electrical energy. Fig. 1 shows an alternator in its elementary or simplest form—merely a conductor, B C moved across a *magnetic flux* (*q.v.*) (Note: For a complete explanation of the production of alternating currents, see *"Alternating Current, Theory of Production of."*) Here the conductor B C is moved to the right *across* the magnetic flux shown as dotted lines between N and S, the poles of the magnets. As the conductor moves across the flux lines of the magnet, an *electro-motive force* (*q.v.*) *or electric* current is set up or induced in the conductor, this force being in such a direction that the current produced will create a flux or magnetic field, in the same direction as the flux on the side of the conductor that comes first into contact with the flux lines N S on moving the conductor. In this case E.M.F. (electromotive force) is in the direction B to C in the conductor. In order to establish a circuit and utilize the current pro-

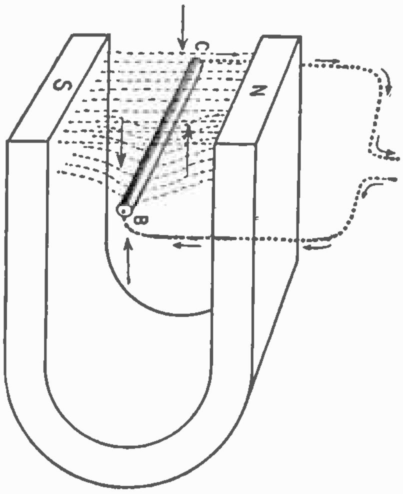

Fig. 1

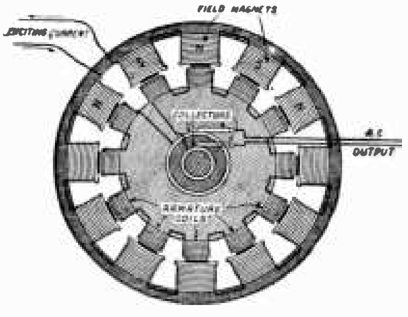

Fig. 2

duced in this manner, it is only necessary to attach wires to both ends of the conductor as shown by the dotted loop. Now at this moment the conductor is being traversed by a current having a definite polarity, that is to say, it is flowing in one direction only. When the direction of the flux lines N S are reversed, as by reversal of the magnet, or the conductor is turned so that the other end comes into contact with the flux lines, then the direction of the current in the conductor will reverse and it will flow in the opposite direction. In other words, with every reversal of the position of the conductor in its relation to the magnetic flux N S, or the reversal of the direction of the flux lines, with respect to the conductor, the E.M.F. that is created will change its direction of flow. (See *Alternating Current.*)

In an alternator it has been shown that the result is approximately the same whether the magnet is rotated or the conductor is rotated through the magnetic flux. In either case the moving member is known as the *rotor* (*q.v.*) and the stationary member is called the *stator* (*q.v.*). The magnets producing the flux lines may be stationary or they may revolve; in either case they are referred to as the *"field"* (*q.v.*).

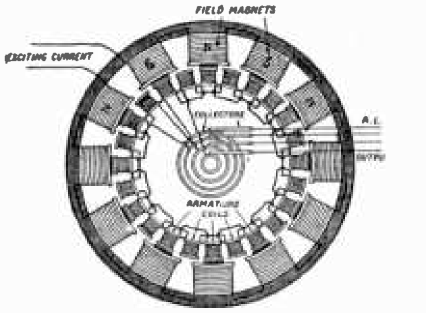

Fig. 3

It has been shown that with every change in the relative position of the field magnets and the *armature* (*q.v.*), irrespective of which member is the rotating one, a change of direction of the induced E.M.F. takes place. Then it is apparent that in order to increase the number of changes, that is, to increase the *frequency* (*q.v.*), it is only necessary to use (1) a greater number of field magnets, or (2) have the armature or conductor pass the magnets more frequently by means of greater speed in the rotating member, or by means of both. Now each of these changes is an alternation (see *Alternating current*) and two such changes will be a *cycle* (*q.v.*). Let us suppose that the moving member of a certain alternator has a rotary speed of 720 revolutions per minute and there are 10 poles or magnets. As frequency is figured by the changes per second, it will be necessary to reduce this number to the revolutions per second. Thus, there will be 12 revolutions per second. This figure multiplied by the number of *poles* (*q.v.*), which we assume as 10, indicates 120 alternations per second. As there are two alternations to each *cycle* (*q.v.*), the frequency in this case will be 60 cycles. The equation is as follows:

$$F = \frac{P}{2} \times \frac{N}{60}$$

where F is the frequency, P, the number of poles and N the speed of rotation per minute. The exact voltage or E.M.F. obtained will depend on the number of poles and the speed of rotation, and also on the distribution of the field windings and total flux from all the poles or magnets. In designing alternators the effective E.M.F. (electromotive force) is expressed by the formula:

$$E = \frac{K P \phi N Z}{10^8}$$

E is the effective electromotive force of the alternator.
K represents the E.M.F. factor of the alternator. This factor depends upon certain charcteristics of the poles and also the distribution of the windings on the armature.
P is the number of poles of the field magnates.
ϕ is the number of lines of magnetic flux that flow from one pole across the gap to the armature—known as the useful magnetic flux per pole; (see *Maxwells*).
N is the speed of rotation of the rotor in revolutions per second.
Z is the total number of conductors on the surface of the armature.
10^8 represents the number 100,000,000—the calculation given would be determined in *abvolts* (*q.v.*), $\frac{1}{100{,}000{,}000}$ volt and it is therefore necessary to reduce it to volts by division.

In the "revolving field" alternator, as explained above, the field magnets are rotated and the conductor remains stationary. The windings in which the E.M.F. is produced, are placed in slots in the inner surface of the stator, the stator and the windings composing it being the *armature.*

In the alternators of the revolving armature type, the conductors or armatures are the rotating member and the field magnets remain stationary. Fig. 2 shows the revolving armature type of alternator in diagrammatic form. A small direct current *generator* (*q.v.*) may be used to excite or build up the field. That is to say, a current is sent through the field magnets (windings in this case) for the purpose of increasing or building up the *flux density* (*q.v.*). This is known as an *"exciter"* (*q.v.*). The exciter may be mounted directly on the end of the rotating shaft of the alternator, or mounted separately and connected to the alternator rotor shaft by means of a belt. This permits a heavy field to be built up by the electro-magnets used in place of the ordinary magnet shown in Fig. 1. If the machine is to be *single phase* (*q.v.*) there will be two *collector rings* (*q.v.*), the two phase type will have four rings, and the three phase six rings. The illustration Fig. 2 shows the general design of a single phase alternator. In this case there will be of course only two rings to collect the E.M.F. set up by the machine. Fig. 3 illustrates the general scheme of armature winding for a two phase machine. Here there are actually two separate E.M.F.'s set up by the separate windings and the ends of each of these windings are connected to collector rings as shown. In the three phase arrangement there may be three separate windings and six collector rings.

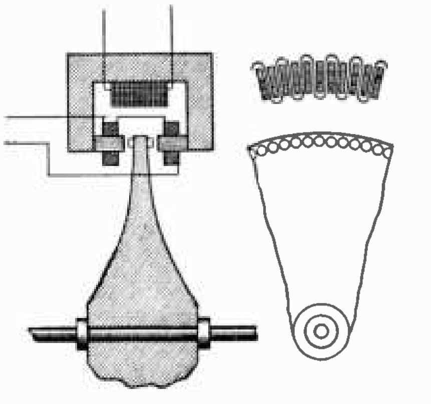

Fig. 4

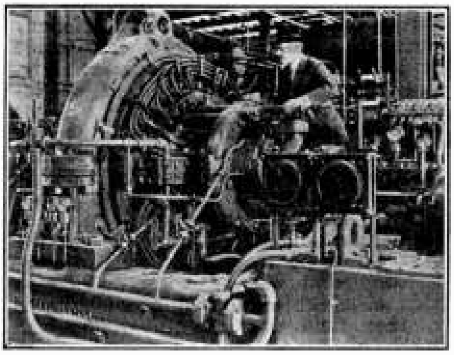

Fig. 5

The use of the revolving armature type for comparatively small voltages is due, as mentioned previously, to the fact that they are easily constructed, and also the advantage that the exciter need not be an external source, but may be a separate winding with a *commutator* (*q.v.*) mounted on the main shaft. The revolving field type of course allows much greater space for the armature winding, inasmuch as the stationary member is not limited to any particular space, whereas the moving member must be arranged as small as possible to do away with mechanical troubles. Thus, the stationary armature can be wound and properly insulated for great voltages. For this reason the revolving field is generally used in high power work. Another point of importance is that where the armature revolves, the windings must be made very carefully as the centrifugal force due to rapid rotation of the members places a great strain on them.

One of the widely used forms of alternator for extremely high frequencies, is the Alexanderson, named after its originator, *Ernst Alexanderson* (*q.v.*), a noted engineer. This machine is of the inductor type. It is a well-known fact that magnetism is more readily established through iron than through air, and therefore when iron is placed in the air gap of a magnetic circuit, the lines of force or flux lines are greatly increased. Then when the iron is withdrawn, the flux lines fall to their original value. We have seen that as changes of flux take place throughout a magnetic circuit an electromotive force is induced in a coil or conductor (armature) surrounding the magnetic circuit. Now, if iron is periodically inserted into and withdrawn from the magnetic field, an alternating current will be set up in the coil surrounding the field. This principle is made use of in the inductor alternator and the field and armature thus remain stationary, contrary to the usual form of an alternator. Now if we arrange to have pieces of iron moved rapidly in the gap between field and armature, changes of flux will take place. Obviously, it is easy in this manner to create many changes of flux within a certain time, and therefore the alternating current set up (E.M.F.) will have a high frequency as seen by the explanation of this phenomena given previously. These moving pieces are so arranged that they practically close the gap between field and armature with only enough clearance to prevent the rotating member striking the stationary members. This is briefly the principle of the Alexanderson alternator. Fig. 4 shows the general plan of an inductor alternator. On the right shows the scheme of winding in slots around the entire form, and as there are necessarily many windings in order to obtain the high frequency, they are generally single turns of wire, each turn having a separate slot. On the left is given a general idea of the arrangement of the rotor, which in this case, as explained previously, is neither field nor armature, but iron teeth arranged on the periphery of the moving disk. The field is increased or excited in the usual manner by means of a source of direct current. In the assembly, the relation of the moving disk to the field and armature is shown. Fig. 5 is a commercial type of Alexanderson Alternator having a frequency (*q.v.*) of 100,000 cycles used for high power radio transmission (*q.v.*).

ALTERNATOR, SINGLE PHASE—(See *Alternator.*)

ALTERNATOR, POLYPHASE—(See *Alternator.*)

AMMETER OR AMPERE METER—An instrument for measuring current in amperes in a circuit. Is connected in series with a circuit. Exists in a variety of forms. The operation of the most commonly used type depends upon the fact that the force a magnet exerts depends upon the number of *ampere turns* (*q.v.*). Therefore, the greater the current sent through its coils, the greater will be its attraction for a balanced armature. Fig. 1 shows a common form of ammeter.

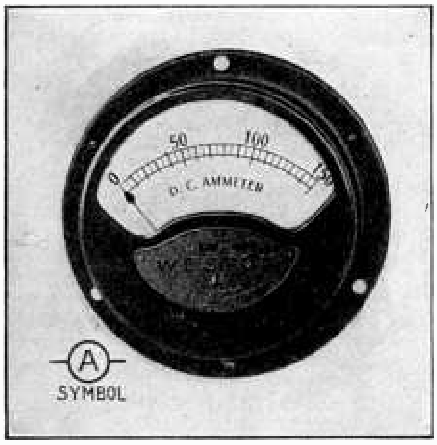

Photo by courtesy of Weston Elec. Inst. Co.

Fig. 1

AMMETER, THERMAL — Commonly known as "*Hotwire Meter.*" Also called "Aerial Ammeter" and "Radio Frequency Ammeter." Operates on the principle of thermal expansion, i.e., the tendency of a wire to expand with heat of current passing through it. Is used to measure current in amperes at extremely high voltages as in radio transmission. Figure 2 shows a Thermal Ammeter which is generally employed in radio work.

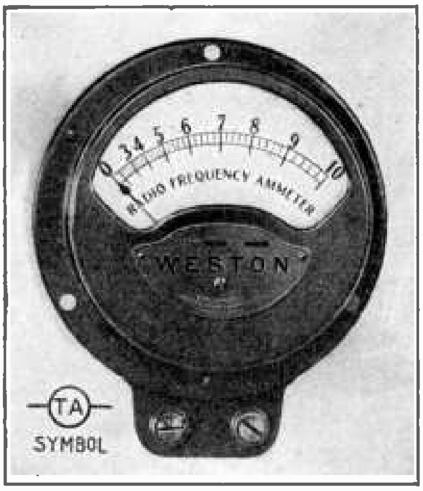

Photo by courtesy of Weston Elec. Inst. Co.

Fig. 2

AMP.—Abbreviation for *Ampere.*

AMPERE—The practical unit of the rate of flow of electricity. It is the unit current, or the rate of flow when one *coulomb* (*q.v.*) flows each second. The value or strength of electric current is almost invariably given in amperes. This unit is used to express the flow of a definite quantity of electricity in a second. If, for example, ten coulombs pass a given point in a circuit in one second, then we say that the current in that circuit is ten amperes. (See *Coulomb, Milli-ampere, Micro-ampere.*)

AMPERE, Andre Marie—A distinguished French physicist, born 1775, died 1836. Noted for his researches in the field of *electro-dynamics.*

AMPERE-HOUR—Commercial unit of quantity. Is that quantity which flows in one hour through a circuit carrying a current of one *ampere.* It is equal to 3,600 coulombs.

AMPERE, INTERNATIONAL—Symbol a—is defined as the unvarying electric current which deposits silver at the rate of 0.00111800 gram per second from a specified solution of nitrate of silver in water.

AMPERE'S RULE FOR DEFLECTION OF NEEDLE—"If one swims with the current and looks at the plus (+) or *north seeking pole* of a magnetic needle it will be deflected to the left, while the negative (—) or *south seeking pole* will be urged to the right."

AMPERE'S THEORY OF MAGNETISM—The theory advanced by *Ampere* that an *electric current continually circulates around each molecule of a magnetic substance and that the process of magnetization is one of arranging these currents in such manner that they all take the same direction.*

AMPERE TURNS—Expressed by the product of number of turns of, and the number of amperes flowing through, the coils of an electro-magnet. Thus, one ampere turn would be one ampere flowing through one turn. The ampere turn is very frequently used as a unit of *magneto-motive force* (*q.v.*) due to the ease with which it can be obtained.

AMPERE VOLT OR VOLT AMPERE—The expression sometimes used to denote *watts* (*q.v.*). For example 1 volt ampere will be one watt, as watts are equal to the product of the volts and amperes in D.C. circuits. (See *Efficiency,* also *Watts.*)

AMPLIFICATION CONSTANT — The factor expressing the maximum voltage *amplification* (*q.v.*) that it is possible to obtain with a given vacuum tube. Thus, if a certain tube will permit amplification of eight times the original voltage, the amplification constant is said to be 8. (See *Vacuum Tube.*)

AMPLIFICATION FACTOR—The ratio of the change of instantaneous voltage, between filament and plate of a *vacuum tube* to a small change of instantaneous voltage between filament and grid for a given constant plate current:

$$y = \frac{\delta \, Vp}{\delta \, Vg}$$

The ratio of power (radio signals), voltage or current output of an amplifying device to the power, voltage or current delivered to the input terminals. Generally speaking the degree of increase in amplitude or volume by insertion of an amplifying device in the circuit. (See *Amplification Voltage.*)

AMPLIFICATION, VOLTAGE — In a *vacuum tube* the ratio of the voltage apparent between the terminals of an impedance in series with the *plate,* to the voltage impressed on the *grid.* (See *Amplifier.*)

AMPLIFIER—One of the most important features of the *vacuum tube* (*q.v.*) is its ability under certain conditions to amplify or increase the intensity of electric currents. Such tubes are generally referred to as amplifier tubes and an amplifier is understood to be one or more such tubes with the necessary associated circuits to accomplish amplification.

X—In the accompanying diagram a circuit of a typical one stage *audio frequency* (*q.v.*) amplifier is shown in conjunction with the *detector circuit* (*q.v.*). In this case the *radio frequency signals* (*q.v.*) are impressed at the points A-B from the aerial and ground circuit. The resulting signals at points C-D are audio frequency or voice frequency because they have undergone a change or *rectification* (*q.v.*). Now, if we consider only the solid lines, the circuit will be an ordinary detector circuit and the signals at C-D can be heard by merely connecting the phones at that point in place of the primary of the transformer—P-B. However, by sending these currents through the transformer and impressing them on the second tube it is possible to obtain signals at points E-F of much greater volume and yet having the same general characteristics or sound. The signals or currents will have undergone amplification—the effect of the amplified tube on a local source of power increase the amplitude, but the form or nature of the signals or waves will be the same. (This statement is not entirely true as amplification is generally accomplished only at the expense of some slight changes in the nature of the waves). They will, however, sound about the same in the phones at E-F as they will when the phones are placed at C-D, the volume at E-F being greater due to the amplification. (See *Distortion*). Now it must not be assumed that an amplifier in itself actually has the ability to increase power without some assistance. It is no more possible to obtain something for nothing in electricity than it is in mechanics. What actually takes place is that the amplifier tube and associated circuits control local power—in this case the voltage of "B" battery—in such a manner as to boost or amplify the power of the signals without materially changing their nature. Therefore, an amplifier may be considered as a device which modifies the effect of a local source of power in accordance with the variations of input power, and produces an increased output power or a means of stepping-up or increasing the amplitude or volume of a signal or series of signals. (See *Amplifier Action or Vacuum Tube.*) (See *Amplifier, Audio Frequency, Amplifier Radio Frequency,* also *Transformer.*)

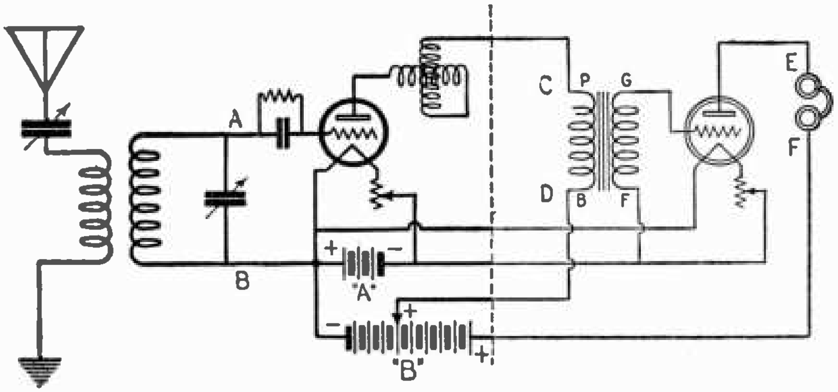

A single stage radio frequency amplifier as used in conjunction with a tuner and detector
Showing the method of amplifying rectified signals at audio frequency

AMPLIFIER ACTION OF VACUUM TUBE—The tendency of a vacuum tube to increase the power or voltage of incoming signals under proper conditions. In the illustration is shown a simple vacuum tube circuit. The incoming signals are impressed on the tube at points A and B. Now it has been noted that small changes of voltage at the *grid* (*q.v.*) will produce much greater changes at the *plate* (*q.v.*). (See *Vacuum Tube, Theory of Operation.*) In other words, the constant changes in value of the incoming signals that are impressed at A-B will produce comparatively larger changes in the current at the plate and therefore larger changes of voltage in the output circuit at C-D. These changes will of course manifest themselves in the form of signals in the phones placed at C-D. Now if these voltages are impressed on a *resistance* (*q.v.*) of suitable value—such as the primary of an amplifying transformer—they can be sent to another tube and again amplified. (See *Amplifier, Audio Frequency.*) In other words the amplifying action of a vacuum tube means the ability of that tube to use input power—the power of the signals delivered to the tube—to control a local source of power such as the "B" battery voltage used on the plate of the tube, and deliver increased power. The signals to be increased or amplified in this manner may be *Radio Frequency* (*q.v.*) or *Audio Frequency* (*q.v.*), each instance requiring a different type of *amplifier* (*q.v.*). The radio frequency signals are amplified by means of a tube and suitable apparatus (transformer or resistance) before they have been changed to audio frequency. The signals that have been *rectified* (*q.v.*) or changed to audio frequency may be amplified by means of a similar arrangement of somewhat different type. The latter would be known as audio frequencv amplification. The increase in power is accomplished without any material change in the nature of the signals or wave form (*q.v.*)

AMPLIFIER, INTERMEDIATE FREQUENCY—An arrangment of a *radio frequency amplifier* to increase the strength of signals at the *intermediate frequency* (*q.v.*) as used in a *Super-Heterodyne receiver*. This is usually a radio frequency amplifier using *transformer coupling* (*q.v.*) the transformers being designed to cover extremely high wave-lengths. (See *Super-Heterodyne*, also *Transformer.*)

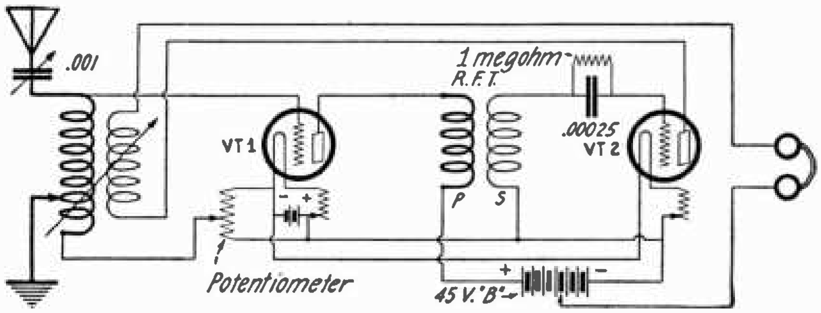

A single stage radio frequency amplifier as used in conjunction with a tuner and detector

AMPLIFIER, RADIO FREQUENCY — An amplifier used to increase the volume of the incoming signals at *radio frequency* (*q.v.*) before they have been *rectified* by the detector. In the illustration is shown a single stage of radio frequency amplification. This is the transformer coupled type. In this case the incoming signals of radio frequency are increased in strength and then passed on to the detector tube VT2. If further amplification is desired, more stages may be used, the number being held within certain limits due to *distortion* (*q.v.*). If more than three stages are used, it is often difficult to control the *feed-back effect* (q.v.) and efficiency is cut down.

AMPLITUDE—The current or voltage on an A.C. or H.F. circuit at any instant.

The greatest vertical distance above or below zero of the crest or trough of a wave is the maximum amplitude of the wave. In the accompanying illustration the numerals at the left (*ordinates q.v.*) may be assumed to be volts, current (amperes), or voice vibrations. A is the zero point; B the maximum *positive*, C the maximum *negative*, and D the length of the wave. Thus, A. B will be the maximum *positive* amplitude, A C the maximum *negative* amplitude, and D the wave length. B is known as the crest or Node—C is the trough or antinode.

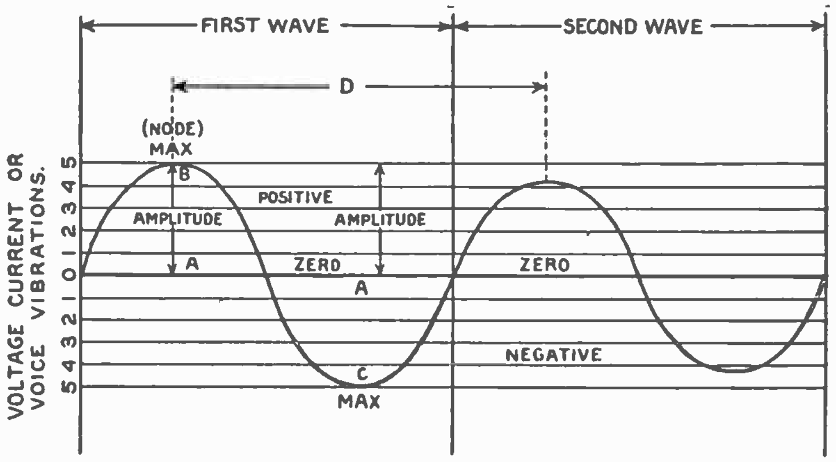

Line wave showing two cycles of an alternating current, indicating the amplitude and damping

X—Now if in a certain train or group of waves, the amplitude of one wave is less than the one immediately preceding it, there is said to be *damping* (*q.v.*), the degree of this damping or diminishing in amplitude being referred to as *decrement* (*q.v.*). Thus, if we assume that in Fig. 1 the current or amplitude of the first wave is 10 amperes, and the next wave has an amplitude of only 5 amperes, then the damping or decrement is 50 per cent, because the amplitude has decreased that much. Similarly, if the first wave amplitude were 10, and the second were 9 amperes, then the decrease

Super Regenerative and *Super Heterodyne*. The principle of regeneration is one that has revolutionized the art of radio reception and broadcasting. The vast majority of present day receivers use regeneration in one form or another. Basically, regeneration is the process of returning a certain portion of the unrectified or radio frequency (*q.v.*) current from the plate circuit of the detector tube to the grid circuit. This has the effect of reducing the resistance of the grid circuit and thus permitting more ready passage of the weak impulses. The result is not only to greatly increase the sensitivity of the detector to weak signals, but actually to step up or amplify the volume of the received signals. In Fig. 1, coils L1 and L2 are the aerial tuning units in conjunction with the variable condenser, C1; C2 and C3 are variable condensers of low capacity; L3 is the plate control for regenerative action; T is the telephone receiver, G, P and F are respectively the grid, plate and filament elements of the tube; R the filament rheostat, and A and B the batteries for filament and plate. In the illustration Fig. 2, C1 is the aerial tuning condenser, L1 the primary coil of a standard vario coupler; V1 the grid variometer and V2 the plate or regeneration variometer. L2 is the secondary of the vario coupler, and the balance of the elements are as in Fig. 1, GL being the customary grid leak and condenser. Regenerative receivers have come into some disfavor owing to their tendency to radiate annoying signals into the antenna, and also the howling within the set when not properly adjusted. While regeneration in some form or other is used in the majority of receivers, it is not widely used now in its original state. (See *Neutrodyne, Tuned Radio Frequency,* etc.)

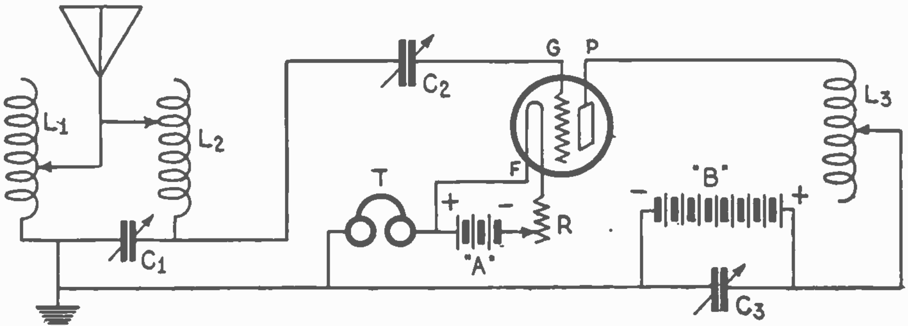

Fig. 1. The Original Armstrong Circuit.

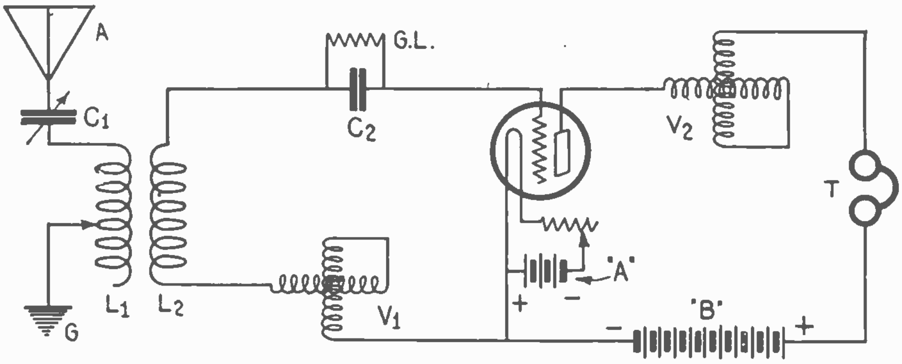

Fig. 2. The Armstrong 3-Circuit Regenerative Receiver.

There are many types of regenerative set, some extremely simple and others more complicated. The original Armstrong regenerative circuit is shown in Fig. 1. Fig. 2 shows a more modern application of the regenerative principle and at the same time a more complicated one. (See *Regeneration, Super Regeneration* and *Super Heterodyne.*)

ARRESTER—A device used for by-passing heavy electrical discharges in the atmosphere in the vicinity of the antenna when they strike the antenna. It ordinarily consists of a small gap between two metallic points connected in series with the antenna and ground in order to protect the apparatus and property against danger. (See *Lightning Arrester.*)

A. R. R. L.—Abbreviation for American Radio Relay League.

ARTIFICIAL ANTENNA — See *Mute Antenna.*

ARTIFICIAL MAGNET—A magnet produced by magnetizing a piece of steel that previously had no magnetic attraction. May be produced by rubbing a *natural magnet* or another artificial magnet across a piece of hard steel in the same direction a number of times. (See *Electro-magnet.*)

ASSUMED DIRECTION OF CURRENT FLOW—The direction which an electric current is assumed to take in its flow. Current is generally considered as leaving the *positive* terminal of its source, flowing through the circuits external to the source, and thence to the *negative* terminal of the source. The direction of flow within the source itself is from the negative to the positive terminal. (See *Theory of Current Flow.*)

ASTATIC COILS—Coils wound in such a manner that, when connected together, they neutralize each other's effect and produce no external *magnetic field.* They are used in the measurement of *inductance.*

ASYMMETRIC CONDUCTOR — One that permits the flow of greater current in one direction than in another. Non-symmetrical as to conductivity.

ASYMMETRICAL EFFECT—The lack of symmetry in the directional effect in a loop or frame aerial due to lack of symmetry in construction. (See *Loop Aerial.*)

ASYNCHRONOUS—A term used in referring to AC (Alternating Current) motors or generators and also to *dischargers* used in radio transmission. As applied to alternating current machines it is signified that the speed of rotation of the machine does not have any definite relation to the *frequency* of the currents produced—thus, out of *synchronism.* (See *Synchronous.*)

An asynchronous radio circuit is one which is not tuned to, or in syntony with, the frequency of the *oscillations* impressed on it. (See *Aperiodic.*)

ASYNCHRONOUS ROTARY DISCHARGE—A spark discharge produced at electrodes of a *rotary discharger* (*q. v.*) for transmitting radio signals, in which the spark rate or number of sparks per second has no relation to the *frequency* of the *alternating current.* (See *Spark Discharge.*)

ASYNCHRONOUS SPARK GAP—See *Spark Gap.*

ATMOSPHERIC ABSORPTION — The portion of the total reduction of radiated power due to atmospheric conductivity, reflection and refraction. The amount of energy loss in radio transmission due to atmospheric factors. This applies only to *electromagnetic waves* (*q.v.*) in *motion in the ether* and should not be confused with the reception losses due to atmospheric disturbances.

ATMOSPHERICS — See *X's, Strays, Static.*

ATTENUATION — of the electric or magnetic intensity or of the average energy density in electric waves refers to the reduction in strength with increase of distance traversed due to absorption or equivalent losses, as distinguished from the reduction in strength due to geometrical divergence. In free space, the geometrical divergence of waves from a small source involves a diminution of the average energy density in accordance with the inverse square law; plane waves (e. g., along wires) suffer no reduction by divergence, but undergo attenuation due to resistance and other line faults. (See *Attenuation, Geometrical Divergence*).

ATTENUATION, GEOMETRICAL DIVERGENCE—The gradual decrease in strength of electromagnetic waves (radio signals) as the distance which they have traversed increases. The effect on the waves as they travel through the air is essentially the same as in the case of light waves. Now if a beam of light is viewed one mile from its source it will have a certain intensity, but if it is viewed again from a distance of two miles the intensity will be much less. In the case of radio waves, the intensity is said to decrease as the square of the increased distance traveled; that is to say, if they have a certain power at a distance of five miles from the transmitter, the same waves at twice the distance or ten miles will be found to have a strength only one-quarter as great. Similarly, if the waves travel fifteen miles or three times the distance, the signal strength will be only one-ninth that of the same waves at a distance of five miles from the transmitter.

AUDIBILITY—The measure of strength of received radio signals as heard in a telephone receiver connected to the radio receiving apparatus. A radio telegraph signal which is at all audible can be read, i.e., the dots and dashes distinguished from one another; radio broadcast programs heard with at least some degree of clarity.

AUDIBILITY (Radio Telegraph)—A measure of the ratio of the telephone current producing a signal in a telephone receiver to that producing a barely audible signal. (A barely audible signal is one which just permits the differentiation of the dot and dash elements of the letters sent in code.) In the simple shunted telephone method of measuring signals the audibility is defined as:

$$\frac{s+t}{s}$$

where s and t are the *impedances* of shunt and telephone respectively.

AUDIBILITY FACTOR—A measure of signal strength obtained by observing the *resistance* to be placed in shunt with the telephones of a receiving station to reduce the signals to unit audibility. Approximately, the audibility factor is proportional to the power absorbed from the signal waves by the receiving apparatus. Unit audibility is the strength of signal at which dots and dashes can just be discriminated.

AUDIBILITY METER—A device for comparison of the strength of received signals, either from different stations or with different circuits by means of a variable *resistance* shunted (placed) across the phones. This resistance is decreased until the signals are just audible in the phones. Now, if another station is tuned in, or the circuit changed in some manner, the signals may either be entirely inaudible or they may be much louder, this giving an idea of the *comparative* strength without actual measurement. A popular type of Audibility Meter is shown in the above illustration.

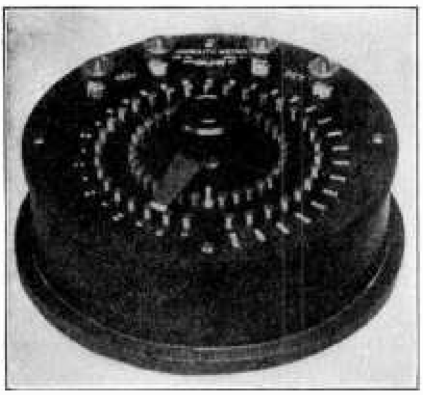

Photo by courtesy of General Radio Co.

An Audibility Meter.

For measurements with an audibility meter, the following shunt formula is used:

$$K = \frac{s+t}{s}$$

Where s is the *impedance* of the *shunt* resistance and t the impedance of the phones, then K is the audibility constant. In general practice a *series* resistance is devised to compensate for the reduction in resistance of the shunt circuit, this keeping the impedance of the plate circuit constant.

AUDIO AMPLIFIER—An *amplifier* used to increase the volume of signals of *audio frequency* (*q. v.*). Audio amplifiers are connected in the circuit of the detector in place of the phones (input) as shown in Fig. 1. This is the standard two stage audio amplifier. The separate stages each consist of a *vacuum tube, a transformer, socket* for the tube and a control *rheostat*. If desired one rheostat may be used for both amplifier tubes as the *filament* current (*q. v.*) can be the same if the tubes are of a similar design. This is known as "transformer coupled amplification." Another much used method is known as "resistance coupled amplification."

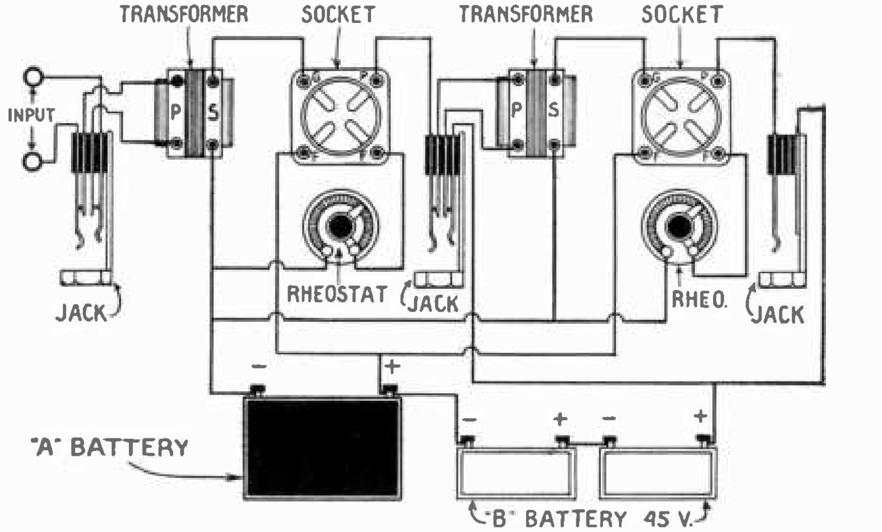

Fig. 1. A standard two stage Audio Amplifier employing phone jacks for detector, first stage, and second stage Audio Amplifiers.

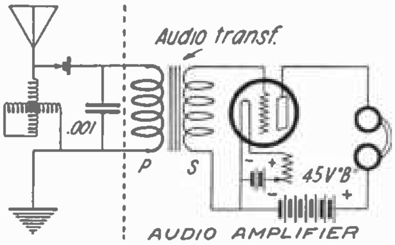

Fig. 2. A one stage Audio Amplifier connected to a crystal receiving set.

Fig. 2 shows one stage of audio frequency amplification connected to a crystal receiving set. There are numerous variations of audio amplification. (See *Push-Pull Amplification,* also *Resistance Coupled Amplifier,* and *Power Amplifier.*)

AUDIO FREQUENCY—A frequency corresponding to normally audible sound waves, i.e., between 40 and 10,000 per second. The limits of audibility are given variously from 16 to 20,000 cycles, but for ordinary usage may be taken as from 40 to 10,000 cycles per second. In audio frequency amplification as used in conjunction with the tuning and detector apparatus of the radio receiving set, the volume of the received sound is amplified or increased over that of the originally rectified sound waves. (See *Radio Frequency;* also *Amplifier.*)

AUDION—The name given by Dr. L. De Forest to a *vacuum tube detector* which possesses a glowing *cathode* and has a steady E. M. F. (*Electromotive force*) permanently applied between *anode* and hot *cathode* and two cold *anodes.* (See *Vacuum Tube.*)

AUSTIN, Louis Winslo, Ph.D.—Born in Ordwell, October 30th, 1867. Graduated at Middlebury College and studied at Clarke University and the Universities of Strasburg and Berlin. Noted for the fact that he is head of the U. S. Naval Telegraphic Laboratory, Washington, D. C., and for his work in measuring high frequency currents.

AUTO-COHERER—A coherer having automatic action to release the particles that cohere when circuit is established. Now obsolete. (See *Coherer.*)

AUTODYNE RECEPTION—See *Self-Heterodyne Reception.*

AUTO INDUCTIVE COUPLING—See *Coupling.*

AUTOMATIC INTERRUPTER—A vibrating electro-mechanical device to make and break an electric circuit by using the energy passing through it. The most common example is that of a *spark coil.* An interrupter may be used for several purposes. In the case of a spark coil it is used to produce sparks for transmission of signals. In another form it may appear in a mechanical battery *charger,* in this instance being used to interrupt the *alternating current* and permit passage of the current in one direction only. A *buzzer* uses a type of interrupter—a vibrating device producing a buzzing sound having a pitch dependent on the rapidity with which the vibrations take place.

AUTOMATIC TRANSMITTER—An apparatus for operating a sending *key* mechanically. (See *Wheatstone Transmitter.*)

AUTOPLEX—A type of radio receiver employing one vacuum tube, operating along the lines of the *Super-regenerative* principle and designed to operate a Loud-speaker. The circuit comprises two standard variometers, a 1500 turn duo-lateral coil, and the usual vacuum

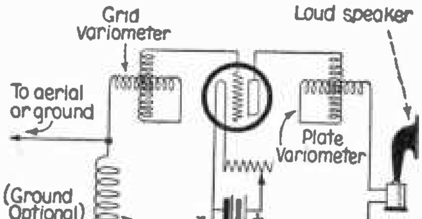

Circuit of the Autoplex radio receiver.

tube, socket, batteries, etc. It is extremely critical in operation, but if carefully assembled will give excellent volume under good conditions. Illustration shows the circuit diagram of the Autoplex receiver.

AUTO-RECEIVER—A device for the automatic reception and recording of radio signals. It is essential in the reception of signals transmitted by an Automatic transmitter (see *Wheatstone Transmitter*) as the speed is too high to permit reception by ear. There are three general classes of automatic receiver: a *resonance intensifier* in conjunction with an inking arrangement; a *phonographic recorder* and a recorder using *photographic means.* In the latter method, a powerful beam of light is projected on to a mirror which is arranged to respond to the signals in approximately the same manner as a mirror *galvanometer.* In the phonographic type, the signals are reproduced on a drum. (See *Resonance Intensifier*, *Photographic Recorder*, also *Phonographic Recorder.*)

AUTO TRANSFORMER—A device for changing voltage in a circuit carrying alternating currents. A form of *transformer* having but one winding, any part of which may be used as the *primary* and any part as the *secondary.* In Fig. 1 the two windings P and S are continuous, although different sizes of wire may be used according to the purpose for which intended. In this case the voltage taken from the secondary D E will be less than the voltage impressed at A B, because the secondary has fewer turns of wire than the primary. Now in Fig. 2, this conditions is reversed, and as the primary has only a few turns of wire and the secondary has a comparatively large number, the voltage obtained from D E will be greater than that impressed at A B. The voltage ratio is approximately proportional to the ratio of wire turns. Thus, if the primary has 100 turns of wire and a pressure of 100 volts is placed at A B, then the voltage at D E would be within certain limits about equal to the number of turns of the secondary. In other words, if the number of turns on the secondary is one-third the number of turns on the primary, then the voltage at D E will be approximately one-third that impressed at A B. The main advantage of the autotransformer is that a saving is accomplished by having the primary and secondary windings combined. The chief disadvantage is that where extremely high voltages are used there is always danger of the two windings becoming crossed or shorted. (See *Short.*) For this reason they are used mainly for comparatively low voltages. One of the most common forms is the step-down type, known as a "bell-ringing transformer." These are used to cut down the voltage of an electric light line to suitable value for use with small toys or bells. If the transformer is used to increase the voltage, it is known as a "step-up" transformer. (See *Transformer—Radio, Audio.*)

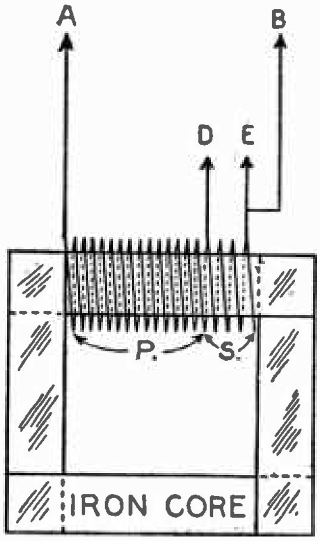

Fig. 1.

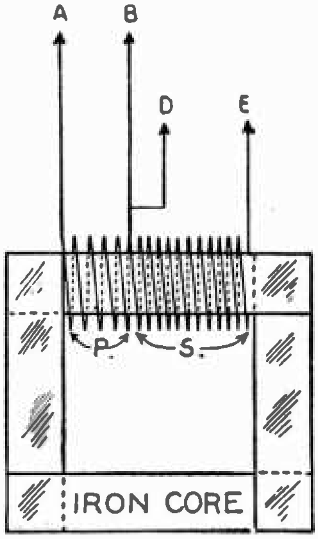

Fig. 2.

AUTO TRANSFORMER COUPLING—See *Coupling.*

B

B—Abbreviation for Baumé, the scale used in a *hydrometer* (*q.v.*)

B. A. AMPERE—The standard ampere fixed by the British Association for the Advancement of Science.

BACK COUPLING—See *Reactance Coils.*

BACK ELECTRO-MOTIVE FORCE—An *Electromotive force* (E.M.F.) which opposes the original electromotive force. The more common term is "*Counter E. M. F.*" In many arrangements of electrical circuits there will be an electromotive force set up in such manner as to oppose the impressed voltage or force. This will be true in the case of a *choke coil* in which any variation in the pressure or electromotive force applied to the coil will result, under proper conditions, in the production of a back or counter-electromotive force that will oppose the change in the original force.

This phenomena is frequently found in electrical circuits; at times introduced for a purpose and at others self-introduced, and therefore demanding consideration in design of apparatus. The rotation of the *armature* of a motor creates a counter E. M. F. opposed to the impressed E. M. F. and upon which the driving power of the motor depends. When an external E. M. F. is impressed on a circuit in which there exists a local E. M. F., the flow of the current through that portion of the circuit containing the local E. M. F. will result in an increase or decrease of energy depending on whether the local E. M. F. is opposed to, or in the same direction as, the external E. M. F. (See *Counter Electromotive force.*)

BACK OSCILLATION—When the *condenser* of a transmitting circuit of the spark type discharges across the *spark gap* some of the *high frequency* current may flow back through the secondary of the transformer instead of taking its normal path across the gap. This back flow is usually referred to as *Back oscillation* or kick back, and is apt to break down (see *Breaking Down of Insulation*) the insulation of the transformer if not prevented or controlled. Back oscillation is retarded ordinarily by means of coils of wire having a certain arbitrary *inductance* value. When the high frequency discharge from the condenser has any tendency to break through, the *reactance* of these coils becomes so great that they act in effect as insulators. As the capacity and inductance of these coils might create an *oscillatory circuit* of the same *natural period* as the primary oscillator, they are shunted with non-inductive carbon resistances. These coils are termed "*choke*" coils.

BAKELITE—An insulating material having widespread use in radio. It possesses great strength and has a high *dielectric* (insulating) value together with the ability to withstand high temperatures. This material is furnished in many forms and under a variety of trade names, being considered one of the best of the insulating compounds.

Chemically known as *Oxybenzylmethylenglycol anhydride.* It is produced by the union under heat and pressure of phenol and formaldehyde together with a small percentage of some alkaline agent. Is heat resisting up to about 500 degrees Fahrenheit—disintegration setting in somewhere above that point.

BALANCED CIRCUIT—An electric circuit arranged in such manner with relation to adjacent or neighboring circuits as to do away with the effect of *mutual induction* (*q.v.*).

BALANCED DETECTORS—A term applied to an arrangement of opposed rectifying *detectors* for reducing the effects of strong *strays* in receiving apparatus. The detectors must be balanced for violent *electro-motive forces*, but not balanced for the moderate E. M. F.'s due to the signals. Actually, a method of eliminating stray impulses in reception.

BALANCED METALLIC CIRCUIT—A circuit through conductors (wire or any metal) in which the total resistance is the same on each side of the circuit. (See *Circuit.*)

BALANCED ROTARY CONDENSER—A variable condenser having some arrangement for keeping the plates accurately balanced in rotation at all points. Fig. 1 shows a form known as the "balanced plate" type. In this case the plates are so arranged that half the movable plates are on one side and half on the other side of the shaft.

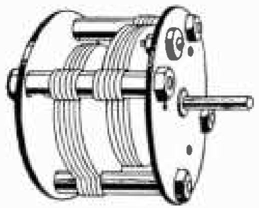

Fig. 1. A "Balanced-Plate" type Rotary Condenser.

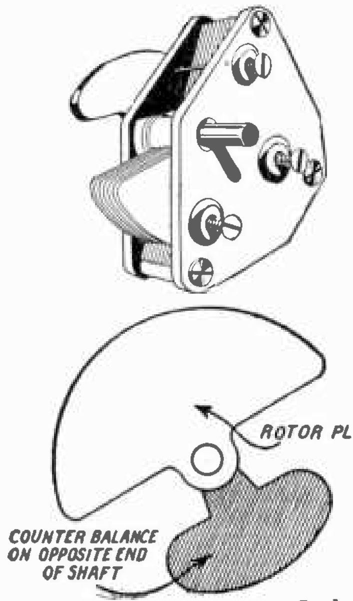

Fig. 2. A Counter-Balanced Rotary Condenser and illustration showing how the counter balance is attached to shaft.

This maintains the balance of the part. Fig. 2 shows another method known as "counter-balanced." The weight of the balance is the same as that of the movable plates and its position on the other side of the shaft serves to maintain balance. The idea is to prevent the plates slipping due to lack of balance and thus keep the setting at the desired point after adjustment.

BALANCING—Any method of eliminating interference by reducing the *amplification* of the undesired signals while permitting full amplification of the desired signals. (See *Interference Eliminator.*)

BALANCING AERIAL—An aerial system designed to permit reception of signals from distant points without interference from a powerful nearby transmitter. The system involves the use of two aerials, one for balancing, the other for receiving. The receiving aerial is tuned to the wavelength of the desired signals, while the balancing aerial is tuned to the wavelength of the nearby station (the interfering signals).

In operation the receiving circuit is coupled to each aerial *electro-magnetically*, and is acted on by both series of *oscillations* (distant and nearby waves). By careful adjustment of *coupling* and phase relation of the *induced currents*, the interfering wave may be balanced out in effect and the desired oscillations received.

BALDWIN RECEIVER—Perhaps the most sensitive type of telephone receiver thus far developed. It employs what is known as a "balanced armature," the signals being produced by the vibrations of a *diaphragm* which is not acted on directly by the magnets. The illustration shows the general plan of a receiver of this type. The *armature* (A) is of soft iron and is pivoted between two U-shaped soft iron pieces (M1, M2) mounted on the ring-shaped horseshoe magnet (M) as indicated. The armature is acted on by the magnets in response to incoming signals and the movement of this element in turn acts on the mica diaphragm (D) by means of a fine brass wire (C). The usual windings (R) are placed between the two pole pieces, the armature being mounted in a central slot. These receivers are much used both for headphones and for loud-speaker units. In the latter case the mica diaphragm is generally made heavier to handle the more powerful vibrations due to the *amplification* of the signals.

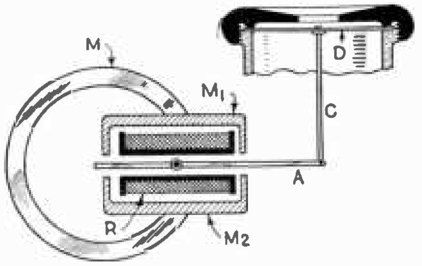

Principle plan of the Baldwin Receiver unit.

When no fluctuating impulses (signals) flow through the windings (R), there is no *magnetic stress* on the armature (A), because this member is suspended centrally between the pieces (M1) and (M2), the magnetic attraction of which are equal. Now when a fluctuating impulse flows through the windings (R) it produces a magnetic flux which combines with the flux of the permanent magnet (M) and the total flux is distributed asymmetrically or unevenly on both sides of the armature. The result is a rocking vibratory movement to this member which is in turn communicated by the juncture wire (C) to the diaphragm (D), producing the audible signals. (For a comparison of this action with the conventional telephone receiver action, see *Telephone Receiver.*)

BALKITE BATTERY CHARGER—An electrolytic battery charger which makes use of a sulphuric acid solution and tantalum as the valve metal. Used for charging *storage batteries* from *alternating current.* See illustration.

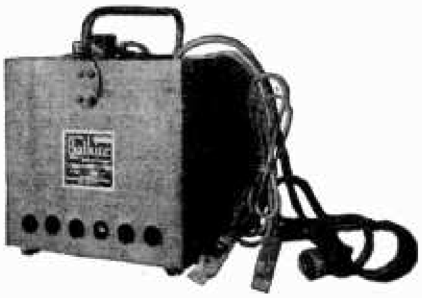

Photo by courtesy of Fansteel Products Co.

A Balkite Battery Charger.

Note: The name Balkite is derived from the name of the man who perfected the process of extraction of tantalum from its ores.

BALLAST TUBE—An automatic means of regulating the *filament current* of vacuum tubes. The device is enclosed in a glass tube and is arranged to permit passage of a certain fixed amount of current. The ballast element is a wire whose *resistance* changes with the current through it, due to the change of resistance with the temperature of the wire. When the *electromotive force* increases beyond normal, the increased current raises the temperature and hence the resistance, so that the current is held to the normal value.

B. and S.—Abbreviation for Brown and Sharpe Gauge (*q.v.*).

BAND OF FREQUENCIES—A continuous range of *frequencies* extending between two frequencies. (See *Band of Wavelengths;* also *Band Pass Filter.*)

BAND PASS FILTER—A *filter* arrangement designed to permit passage of all *frequencies* extending between two definite frequencies but excluding all other frequencies. (See *Band of Frequencies.*)

BAND OF WAVE-LENGTHS—A continuous range of *wave-lengths* extending between two definite wave-lengths. For example, the usual broadcast band of wave-lengths is approximately 250 meters to 550 meters.

BANKED BATTERY—A *battery* that has its *cells* connected in *parallel.*

BANK WINDING—A form of winding for coils in which the turns are staggered so that two or more layers may be used without having large distributed capacity losses. The illustration shows method of bank winding for two layers. Note that in this form the *electrostatic capacity* between any two turns is only that between adjacent turns, representing only a small part of the total winding, whereas on the usual form of winding the adjacent turns would represent a good portion of the total *inductance* and hence the effect of distributed capacity would then be much greater. (See *Distributed Capacity.*)

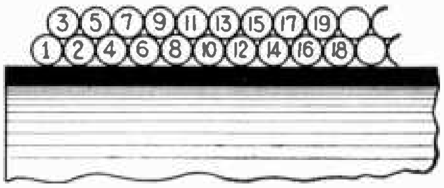

Method of hank winding wire on a coil form.

BARRETER—A receiving instrument consisting essentially of a small mass of conducting material that is heated by the passage of an *oscillatory* current, and arranged so that the consequent alteration of electrical *conductivity* affects an indicating instrument, such as a telephone receiver or *galvanometer.*

BASKET COIL—Variometers—couplers, etc. (See *Basket Wound.*)

BASKET WOUND—A method of winding coils, such as variometers or variocouplers. The winding is a latticework of wires in skeleton form, the object being to save weight and eliminate large losses such as occur in the types using a great deal of solid insulating material. The illustration shows a popular form of basket variometer.

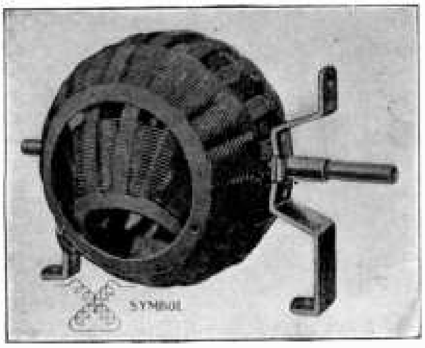

Photo by courtesy of Amer. Radio & Research Corp.
A Basket Wound Variometer.

BATTERY — In the electrical field, a group of *primary or storage cells, dynamos*, or *condensers* grouped together to form a single unit. The term is often used incorrectly to refer to a single cell—as "dry battery," the correct name for which is "dry cell." Thus a group of condensers joined together as one unit is termed a battery, also a number of dynamos arranged as a single unit for the production of electrical energy. (See *Dry Cell*, also *Battery of Alternators, Storage Battery, "A," "B" Battery*, etc.)

"B" BATTERY—An appliance for obtaining electricity of sufficient voltage to supply the plate of a *vacuum tube*.

There are two types of "B" batteries, viz., the *dry cell* type and the storage types. In order to obtain the high voltage required for vacuum tube plates, many small cells are connected in series. The usual method is to connect them together and seal in a square package. Often taps are brought out so that various voltages may be obtained for *critical* vacuum tubes. The most popular voltages which dry cell "B" batteries give are 22½ and 45 volts. The storage "B" battery is similar to the ordinary storage battery with the exception that the cells are much smaller, but contain more cells in series. These can be charged in the same manner as the "A" or *storage battery*. In Fig. 1, we have a popular type of "B" battery. This is tapped in order that different voltages may be obtained.

EVEREADY
RADIO

Photo by courtesy of National Carbon Co., Inc.
Fig. 1-A. A popular type of "B" Battery of 22½ volts.

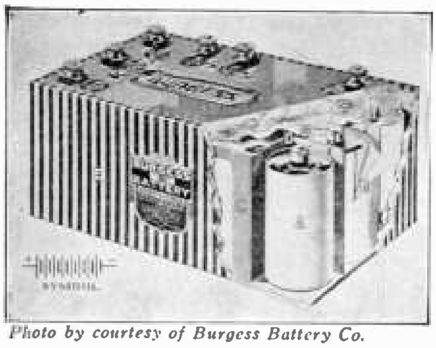

Photo by courtesy of Burgess Battery Co.
Fig. 1-B. A "B" Battery showing the contents of small cells.

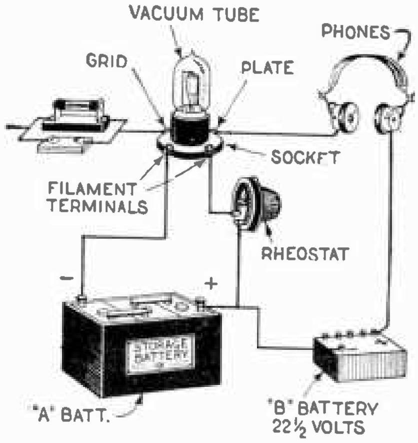

Fig. 2. Manner of connecting a "B" Battery in the circuit of a vacuum tube detector.

The manner of connecting up the "B" batteries in radio receiving is shown in Fig. 2. Here one battery of 22½ volts is used. In Fig. 3, three batteries are connected in series so that 67½ volts are obtained.

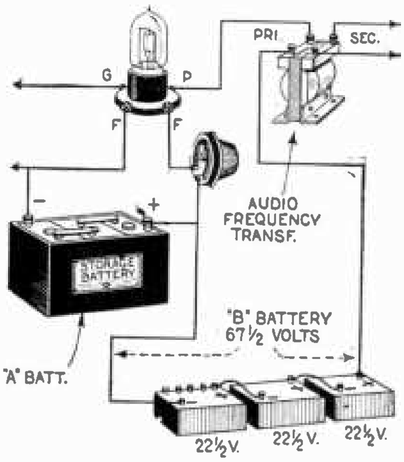

Fig. 3. "B" Battery of 67½ volts connected in an amplifier circuit.

A common "B" battery may be used for the detector tube and the amplifier as shown in Fig. 4.

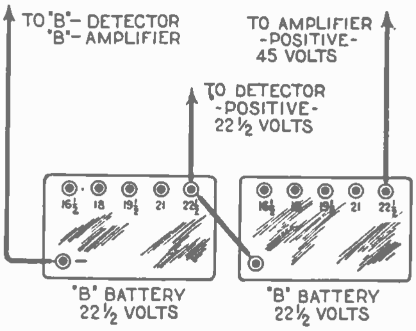

Fig. 4. How a common "B" Battery may be used both for detector and amplifier.

If separate batteries for each tube are desired, the circuit should be connected as in Fig. 5. This lengthens the life of the "B" battery as the drain on the battery is much less, i.e., only one vacuum tube is operated on each battery.

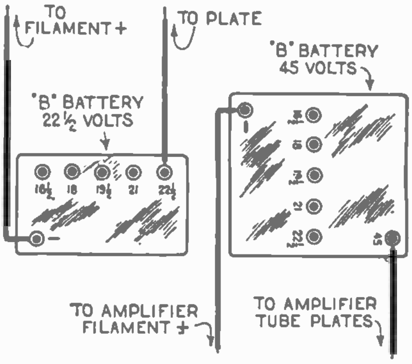

Fig. 5. Method of employing separate "B" Batteries for the detector and amplifier.

BATTERY OF ALTERNATORS — In ordinary electrical work such as house lighting, several *alternators* for the production of alternating current may be used together as one source of power. Also in high-power radio transmission, several *high-frequency alternators* may be used together to produce added power. (See *High-Frequency Alternator*.)

BATTERY TESTING INSTRUMENTS —The various devices used for determining the condition of batteries used as radio apparatus. (See *Ammeter, Voltmeter*, also *Hydrometer*.)

BEAT FREQUENCY—The *frequency* or number of vibrations per second of a series of *oscillations* representing the difference between two other series having different frequencies. It is numerically equal to the difference between the original two series of oscillations. Thus, if we have a series of oscillations having a frequency of 100,000 *cycles* per second, and combine it with another series having a frequency of 90,000 or 110,000 cycles, the *beat frequency* will be 10,000 cycles or the difference between the other two frequencies.

X. The frequency of recurrence per second of either maxima of addition or minima of opposition of two superposed periodic phenomena having the same nature but of different frequencies. The time measure of vibration of a beat or heterodyne. (See *Heterodyne*, also *Beats*.)

BEATS—When two sets of *oscillations* of different *frequencies* occur in the same system, the difference in the frequency or rapidity of vibration causes a new set of oscillations having a frequency equal to the difference between the original two sets. This may be accidental as in the case of certain types of receivers such as the *regenerative* type, in which case the beat oscillations are undesired and cause interference or disturbances in the set. Then again a beat frequency may be purposely introduced as in the case of the *Super-Heterodyne receiver*. Here we have the natural incoming oscillations from the antenna combined with a series of oscillations produced locally in the system. The result is a new set of oscillations where the incoming series and the locally produced series are not of the same frequency. (For more complete explanation see *Super-Heterodyne Receiver*.)

X. Where two *periodic phenomena* are *super-posed* or run together and the frequencies differ, the gradual change in *phase difference* produces a condition wherein the *amplitudes* are in opposition at one instant and in concurrence at a later instant with the various intermediate stages during the interval. (See *Heterodyne*, also *Beat Frequency*.)

BEG OHM—A *resistance* of one billion *ohms*. (See *Megohm*.)

BELIN, EDOUARD—French scientist, the inventor of what is considered the first practical system of radio transmission of photographs. The device can be used for sending photos by cable or telephone lines as well as radio. (See *Transmission of Photographs by Radio*.)

Brown photo.

Edouard Belin.

BELL, Alexander Graham (1847-1922)—Scottish scientist. Born in Edinburgh, March 3rd, 1847, he was educated at the High School and University and graduated as a doctor of medicine. In 1870 he went to Canada, and in 1872 became professor of vocal physiology in the University of Boston. In 1876 he exhibited his apparatus for the transmission of sound, afterwards developed into the telephone. Bell was experimenting with an electric invention by means of which he hoped to make speech visible to the deaf. A delicate metal reed was caused to vibrate by spoken speech and to transmit an electric current to the opposite end of a wire, where the vibrations of the first reed were reproduced in a second reed by a magnet. He found that it was possible to transmit not merely vibrations of the original reed, but to reproduce the sound itself in the vibrations of the second reed. In 1878 he invented the photophone, to enable sound to be transmitted by variations on a beam of light, and later a phonograph. Bell was the author of many scientific papers, was awarded the Albert Medal of the Royal Society of Arts in 1902 and the Hughes Medal of the Royal Society in 1918. He died Aug. 2nd, 1922.

Alexander Graham Bell.

BELLINI, Dr. Ettore—Born in Foligno, Italy, April 13th, 1876. Educated at the University of Naples. Noted as Electrical Engineer to the Royal Italian Navy, and chief of the Naval Electrical Laboratory at Venice. Joint inventor with Captain Tosi of the "Radiogoniometer" (*q. v.*), a device for finding the direction of transmitted radio signals.

BELLINI-TOSI AERIAL—See *Goniometer*.

BELLINI-TOSI DIRECTION FINDER—See *Goniometer*.

BEZEL—In mechanics, generally a groove and flange made to receive a beveled edge. In its adaptation to radio some liberty was taken with its real meaning. Its significance in this sense is a small metal ring having a wire mesh or glass center. It is fitted into a circular hole in the panel of a radio receiver and used as a peep hole to enable the operator to know at all times the conditions of the tubes, whether or not they are properly lighted.

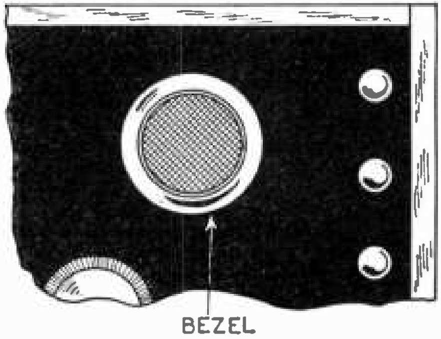

A Bezel on the panel of a radio receiving set.

BILLI CONDENSER—A variable *condenser* of low *capacity*, consisting of two brass tubes, one of which is arranged to slide in and out of the other for the purpose of varying the capacity. Such condensers were much used before the days of broadcasting, in conjunction with any circuit using a crystal and battery for detector. This type of condenser has long been obsolete in the United States.

BINDING POST—A screw and nut arrangement used on electrical units and radio apparatus to make convenient connections from an external source to the desired point within the apparatus. Binding posts are provided, for instance, to make handy connections from batteries to the set and from the aerial and ground leads. The illustration shows a popular type of binding post.

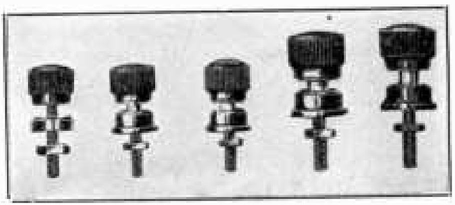

Photo by courtesy of The H. H. Eby Mfg. Co.

A group of Binding Posts as used on electrical and radio apparatus.

BI-POLAR—Having two *poles*. Usually a *dynamo* or motor whose *armature* rotates between a *field magnet* having only two poles. The most modern types are of the Multi-polar variety. (See *Bi-polar Magnetic Field*.)

BI-POLAR ARMATURE WINDING—An *armature* wound in a manner to permit its use with a dynamo having a bi-polar magnetic field.

BI-POLAR MAGNETIC FIELD—The magnetic field (*q.v.*) created between two magnetic poles. The illustration shows a common form of bi-polar magnet for a dynamo. Bi-polar magnetic fields are now used as a rule only for direct current machines of low power, such as those under 5 *Kilowatts*. (See *Dynamo*, also *Generator*.)

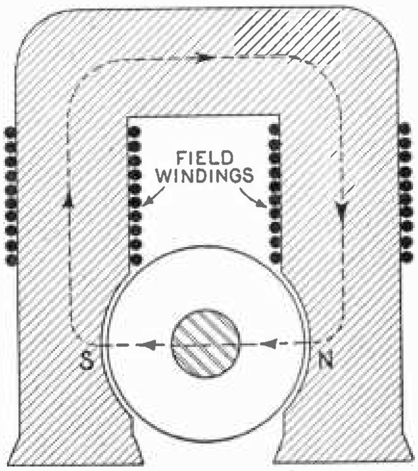

A Bi-Polar Magnet of a dynamo.

BI-TELEPHONE RECEIVER—A headset with two receivers or phones as generally used in radio. In the early days of radio transmission it was the custom to use only one phone.

BLONDEL, Andree—French electrical expert. Born at Chaumont, France, in 1863, he graduated at Paris University, and studied electric waves, on which subject he early contributed a number of papers to various scientific journals. In 1893 he invented the *oscillograph*, an instrument somewhat similar to a mirror *galvanometer*, for showing curves of *oscillating* or *alternating currents*. This invention opened up a fresh field in the study of alternating currents. In the same year Blondel explained for the first time mathematically the effect of inertia in the shunting of *alternators*. He is responsible for a system of acoustically syntonic wireless telegraphy, and for directed waves produced by a double aerial. In 1902 he patented a method for producing electric oscillations for wireless telephony, and has written many papers on microphonic control for transmitters, wireless telephony, the singing arc, etc.

BLONDLOT, Professor Prosper Rene—French wireless expert. Born at Nancy, France, in 1849, studied at

Paris, and became professor at the faculty of sciences, Nancy, and afterwards Honorary Professor and Correspondent of the Institute of France. Professor Blondlot is famous for his studies of electro-magnetic waves, particularly with regard to their speed, and the laws of *propagation* of wireless waves in various media.

BLOWER MOTOR—A motordriven fan used to deliver a high pressure blast at the *spark gap* to prevent *arcing*. (See *Spark Gap*, also *Spark Discharger*.)

BLOWOUT—See *Magnetic Blowout.*

BLUE GLOW—A condition within a *vacuum tube* when the vacuum has become poor. After continued use a tube will often hold a small percentage of gas, which causes a blue glow when current is passed through it. A tube in this condition is a poor *detector* or *amplifier* and should be replaced. The condition should not be confused with the glow often caused by an excessive voltage. In this latter case decreasing the voltage to the normal value will generally permit the tube to be operated efficiently. The blue glow in this case is due to *ionization* of the residual gas by the excessively high *potential* impressed across the elements of the tube.

BOARD OF TRADE UNIT—B.O.T. 1,000 Watt Hours. One and a third Horse Power.

BODY CAPACITY—The effect of the human body when tuning a radio receiving set. The hand when placed on or near the controls very often throws the receiver out of balance with the incoming signals. In the case of very *sharp tuning* this effect is more in evidence and is likely to cause howls due to *self-oscillation.* The howling is due to the production of an audible *beat frequency* in the system, caused by the combination of the local *oscillations* with the incoming signal oscillations. The remedy may be to shield the panel with metal foil or in some cases to merely alter the direction of the leads from the *tuning condenser.* The experienced operator seldom pays much attention to the phenomena, as a little practice enables the listener-in to compensate for the effect. To make this more clear, if, when tuning the set, the withdrawal of the hand from the dial *detunes* the set or throws it out of balance, it is due to the fact that the hand has acted as a certain amount of *capacity* in the circuit, and naturally its withdrawal is equivalent to a change in the condenser setting. A simple system is to turn the control beyond the point of maximum volume, then when the hand is withdrawn the capacity will drop to its proper value. This can be mastered very readily. If the tuning condenser is connected across the secondary, i.e., one side to the grid, and the other to the grid return, the stator plates of the condenser should be connected to the grid of the tube. If metal foil or thin copper sheeting is used for shielding it should be connected to the ground post of the set.

The tendency of the human body to insert an arbitrary capacity in the circuit. The phenomena is especially noted where maximum volume is obtained only by very critical control close to the oscillating point, in which case the circuit may be thrown out of *resonance*.

BOLITHO CIRCUIT—A *super-regenerative circuit* patented in England in 1919 by Captain J. B. Bolitho. The circuit was originally intended to operate a relay device. The circuit has been adapted to use as an *amplifier* for radio reception.

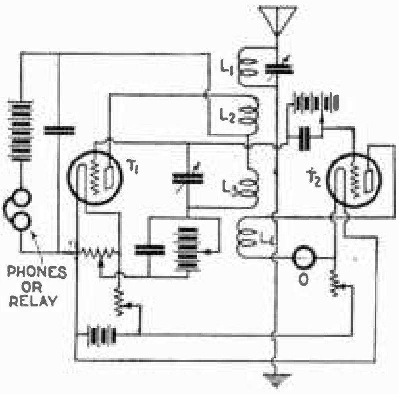

Diagram of the Bolitho Circuit.

In the illustration the phones may be replaced by a loud speaker or relay. In operation, T1 is held at a point just below that where *oscillation* sets in by means of T2 which is excited by an oscillator O in the plate circuit and coupled by means of coil L4 to the tuned grid circuit through L2 and L3, the coupling, as explained, being so arranged that the tube is always just below the oscillating point. A *reaction coil* L4 is placed in the plate circuit of T2 in such manner as to oppose the magnetic linkage between coils L2 and L3. The *frequency* of the generator O is lower than that of the received signals. As indicated, the grids of both tubes T1 and T2 are joined together and the tube T2 maintains the circuit in a receptive condition when tube T1 is tuned just below the oscillating point.

The oscillator O makes the *plate* of T2 alternately *positive* and *negative* with the following effect: When the plate of T2 is negatively charged there is no current in coil L4, this coil therefore not affecting coils L2 and L3, and permitting tube T1 to build up self oscillation. When the plate is made positive by the action of the oscillator, current flows through coil L4, and as it is coupled in opposition to coils L2 and L3 it neutralizes the coupling between L2 and L3 and prevents transfer of energy between the grid and plate circuits of the tube T1. This tends to stop the self oscillation and makes the circuit receptive to the energy (signal) produced by outside signals in the aerial coil L1. The frequency, of course, is determined by the frequency of the oscillator O. The system is especially adapted to the reception of continuous wave signals. (See *Super-regenerative circuit,* also *Feed-back* and *Regenerative circuits.*)

BOLOMETER—Type of *Wheatstone Bridge* having an easily heated *resistance,* such as a very fine wire in one arm. (See *Wheatstone Bridge.*)

BOOSTER—An expression signifying a small *dynamo* used in conjunction with main dynamo to temporarily raise, when necessary, its normal *pressure.* It is generally driven by a motor supplied with energy from the main generator and thus becomes in effect a *continuous current transformer.* Frequently used for charging *accumulators* of a generating plant. The term "Boost" is also used to denote increase or a stepping-up of any electrical quantity. (See *Amplification.*)

BORNITE—A *crystal rectifier* much used in radio reception. It is a natural sulphide of iron and copper, having a metallic blue lustre. This mineral is used in combination with zincite or copper pyrites as a crystal detector. Such combinations are generally known as "crystal to crystal" detectors to distinguish them from the ordinary variety using one mineral and a wire contact.

BORON—A non-metallic chemical element used in radio as one of the electrodes for the *T. Y. K. Arc.* Chemical symbol B, atomic weight 11.0, specific gravity 2.6.

B. O. T.—The customary abbreviation for *Board of Trade Unit.* The unit represents 1,000 *watt hours.* (See *Board of Trade Unit.*)

BOX AERIAL—A term occasionally applied to loop aerials. (See *Loop aerial,* also *Frame Aerial.*)

BRADFIELD INSULATOR—A particular form of *Lead-in Insulator,* consisting of an ebonite tube provided with zinc cone and ebonite spark discs, for breaking up continuous streams of rain running down outside which might cause the aerial to become grounded. The whole is held in position, half way through roof of operating room, by means of a stuffing box. The aerial is led in by means of a conducting rod through center of tube. (See *Petticoat insulator.*)

BRANLY COHERER—Early form of *Marconi Coherer.* (See *Coherer.*)

BRANLY, EDOUARD—French radio expert. Born at Amiens, France, Oct. 23rd, 1844. He was educated at Paris and afterwards became Fellow of the University, doctor of physical science, and doctor of medicine. Branley early made a study of electro-magnetic waves, and in 1890 and 1891 patented methods of operating a local relay circuit from a distance by means of wireless waves. In 1900 he was awarded the Grand Prix by the International Jury of Superior Precept Instruction for his exhibition of radio-conductors.

In 1890 Branly published an account of his very extensive series of observations on the electrical conductivity of loosely packed metal filings, and he made the extremely important observation that an electric spark at a dis-

Edouard Branly.

tance had the power of suddenly changing the electric *conductivity* of loose masses of powdered *conductors*. To Branley is due the *coherer* named after him.

BRAUN, FERDINAND—Professor at the University of Strasburg, and one of the leading world authorities on Radio transmission. As early as 1899 Braun was granted a patent for *closed oscillating* systems with an *inductively coupled* antenna. The system was claimed by Braun to possess a much greater efficiency than the directly coupled systems. The Braun transmitting set, as manufactured by Siemens and Halske, consisted of a large coil worked into an *electrolytic interrupter*, a set of *Leyden jars*, enclosed *spark gap*, *oscillation transformer* wound with insulated wire placed in oil, and for the receiving set the standard type of *coherer relay* and Morse register. In 1899 Braun established communication between Cuxhaven and Heligoland, using aerial wires 90 feet high, and the inductive coupled aerial connection for transmitting. In 1903 Braun joined with Slaby, von Arco, and Siemens to form the Telefunken system of transmission. Professor Braun was awarded the Nobel prize with Marconi in 1900, for his work in wireless. Braun has devised a method of directional wireless which depends upon the interference of *electric waves* travelling in the same direction but different in *phase*. Three simple vertical wire aerials are set up in positions corresponding to the angular points of an equilateral triangle, and *oscillations* are created in these which differ from one another in phase. In 1897 Professon Braun published a description of his cathode ray tube, and afterwards pointed out, in 1902, how such a tube could be used to trace the forms of alternating current waves.

BRITISH STANDARD WIRE GAUGE—The standard wire gauge of Great Britain. The table shows the various diameters in thousandths of an inch (mils.).

WIRE GAUGES IN MILS.

GAUGE	New British Standard
0 000 000	500
000 000	464
00 000	432
0 000	400
000	372
00	378
0	324
1	300
2	276
3	252
4	232
5	212
6	192
7	176
8	160
9	144
10	128
11	116
12	104
13	92
14	80
15	72
16	64
17	56
18	48
19	40
20	36
21	32
22	28
23	24
24	22
25	20
26	18
27	16.4
28	14.8
29	13.6
30	12.4
31	11.6
32	10.8
33	10.0
34	9.2
35	8.4
36	7.6
37	6.8
38	6.0
39	5.2
40	4.8

BREAK or BREAKER—See *Circuit Breaker.*

BREAKDOWN POTENTIAL or BREAKDOWN VOLTAGE—The voltage necessary to *break down* or puncture a *dielectric.* The ability of any *insulating* material to resist breakdown is an important matter. An insulator or dielectric material is usually given a certain rating or point above which it is likely to fracture. Thus, a condenser may have a breakdown rating of 500 volts. That is to say that voltages up to 500 and slightly over may be handled with safety, but any large increase above that figure may result in puncture of the material by the excessive potential or voltage. The relation between dielectric strength and breakdown voltage is shown in the following formula:

$$V = cd^{2/3}$$

Where d is the thickness of dielectric in millimeters

V the potential difference in volts

c-a constant representing the potential difference required to cause a dielectric breakdown of a sample of the material 1 mm. thick. (See *Dielectric Strength.*)

BREAKING DOWN OF DIELECTRIC—In a *condenser*, the gradual weakening of the *dielectric* or insulating sheets which eventually causes them to permit direct passage of the current. For example, if a force of 1000 volts is applied for a considerable period to a condenser that has been designed for a maximum of 500 volts, the insulating material between the plates would probably be punctured by the current. (See *Break-down Potential.*)

BREAKING DOWN OF INSULATION—The same effect as referred to under *Breaking Down of Dielectric.* The insulation weakens and permits the ready passage of electric currents. (See *Break-down Potential.*)

BRIDGES—See *Capacity Bridge*, also *Wheatstone Bridge.*

BRIDGING—A term occasionally used for *shunt* (q.v.) connections. Thus, a grid condenser is generally shunted or bridged by a *grid leak.*

BROADCAST—Transmission of music, news and other matters of general interest and entertainment by means of *radio telephony.* (See *Broadcasting, General Treatise on Methods.*)

BROADCASTING, GENERAL TREATISE ON METHODS—The application of radio telephony to the transmission through the ether of musical programs, speech or any form of news or entertainment, in such manner that it can be readily received by anyone possessing a radio receiver sufficiently sensitive and capable of being tuned to resonance with the particular waves being sent out. While broadcasting in its present world-wide scope is a comparatively new enterprise, it has actually been used to a more or less degree since the practical development of the wireless telephone. As near as records can be determined, the first actual broadcasting was done in 1912. It was, however, not until after the world war that it began to develop into an industry of prominence. Now it occupies a place of very definite importance in everday life; carrying news and entertainment into the remote corners of the earth, far removed from telephone or telegraph communication.

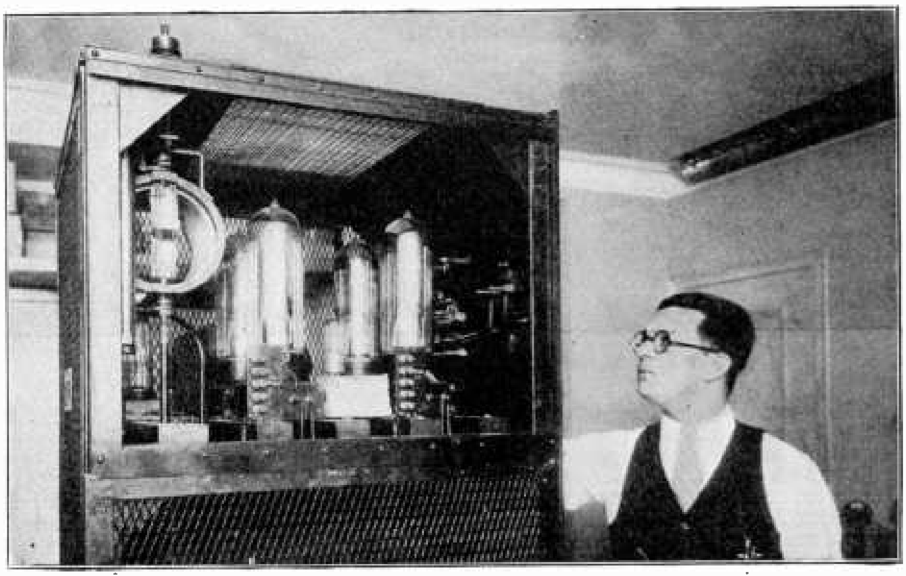

Fig. 1. A bank of 250-watt transmitting tubes in a Radio Broadcast Station. Note the small receiving tube for comparison of size.

The fundamental principles of broadcasting are essentially those of radio transmission, with the addition of means of combining the speech or music vibrations with the regular radio waves. First we have to consider the radio waves of high-frequency—alternating current that changes in direction many thousands of times each second. These high frequency waves are known as the carrier waves. The frequency of these waves varies according to the length of the waves (wavelength), in the case of American broadcasting stations, being between 250 and 550 meters in most cases, which means a frequency range of approximately 550,000 to 1,200,000 cycles.

Now by means of a microphone, a device similar in principle to the mouthpiece of an ordinary telephone, but designed to handle large currents, the air pressure waves, i.e., speech or music, are converted into variations of electric current. By use of suitable amplifying apparatus, the strength of the electric waves radiated from the antenna is varied to correspond electrically to the acoustic or sound variations due to the speech or music being impressed on the microphone. The speech or music waves are low frequency waves, generally ranging between 100 and 4000 per second. Thus, while the carrier wave oscillations are of high frequency, the variations in it are low frequency, corresponding to the ordinary speech or music vibrations.

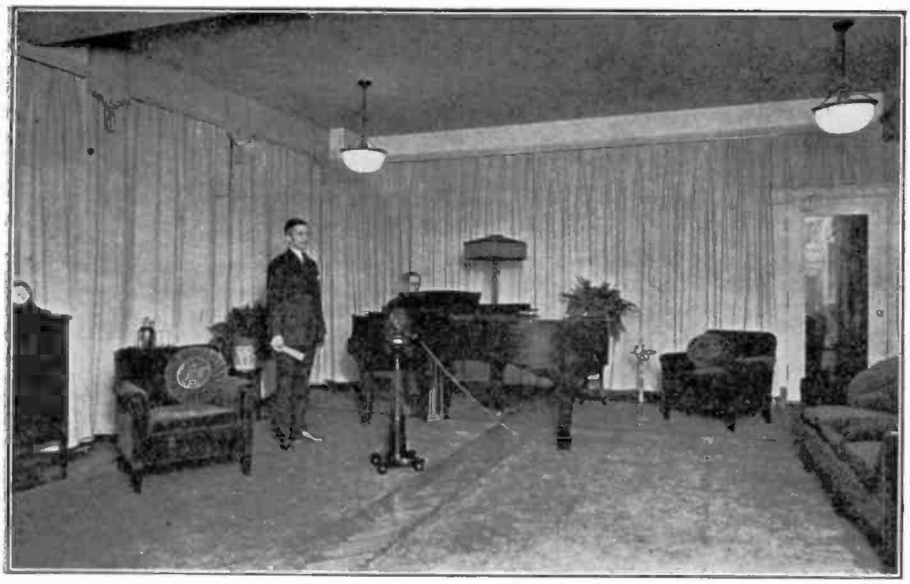

Fig. 2. Studio of a modern Radio Broadcast Station.

The process of varying the radiated wave in accordance with the sound waves is known as *modulation.* This adjustment is of the utmost importance, as defects here result in distorted or improperly modulated sound waves and the reproduction cannot be perfect. The reception of broadcast programs is essentially the same as that of ordinary radio signals, with the exception that care must be used at the receiving end to preserve the form of the original vibrations. When dot and dash signals are received, the tone is of comparatively little importance, whereas with voice or music it is essential that the tone quality be retained.

In broadcasting the high frequency oscillations are produced by means of a tube or group of tubes, practically identical with those used for reception, but designed to handle large currents and voltages. The majority of American stations use vacuum tubes of 250 watts power rating, one or more of them being used to obtain the total power required. These tubes are shown in the illustration, Fig. 1.

When a program is being broadcast, the action is somewhat as follows: The artists are disposed before the microphone with due regard to the effect of certain instruments. That is to say, one type of instrument may be placed near the microphone, while another must be placed farther away, owing to the differences in tone and volume. This is done to prevent any one tone or range of tones from predominating. These speech or music currents are converted into electrical variations by the microphone and then increased many times in power or amplitude by means of a speech amplifier system, comprising vacuum tubes and the necessary associated circuits.

These amplified currents are applied to the grid of a control tube or bank of tubes and result in large variations in the plate current of the tubes. (See *Theory of Vacuum Tube.*) The control tubes are supplied with filament current from storage batteries or special low voltage D. C. generators, and the high voltage applied to the plates of the tubes is furnished by generators. This power will vary in most cases from about 1,000 to 5,000 volts. In some broadcasting stations, the power is obtained through high voltage direct current generators, while in others, the alternating current supply is stepped up or transformed through high voltage transformers and then rectified by means of special vacuum tubes.

The high frequency waves are applied to the antenna system tuned circuits and carefully modulated in accordance with the variations of the speech or music being transmitted. The action of the transmitter and the effect of the modulation must be carefully observed at all times during the transmission of programs. (See *Harmonic Suppressor,* also *Oscillograph.*) Fig. 2 shows the studio of a modern broadcasting station. It is interesting to note that these studios are very carefully arranged and the walls cushioned to prevent troublesome acoustic effects. There must be no reverberation—echoes—or the transmitted program would be distorted by the additional or repeated tones. Fig. 3 shows the main control panels. Fig. 4 illustrates the power generating system of a typical broadcasting station.

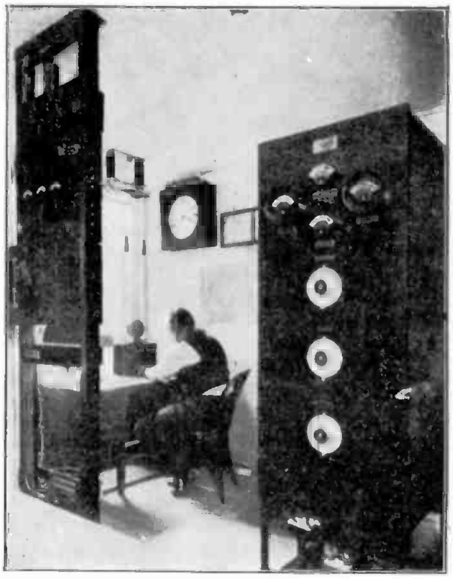

Fig. 3. Main control panels of a Radio Broadcast Station.

The radio waves generated and radiated from the broadcasting station travel away from the antenna at tremendous speed—that of light waves, or about 186,000 miles per second. These waves radiate off into space in

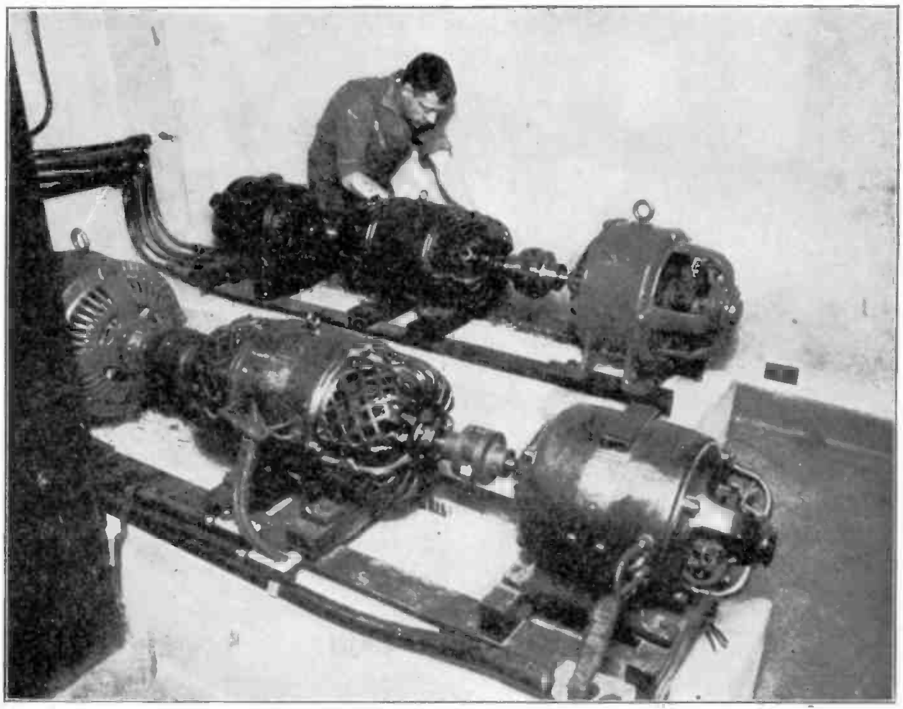

Photos by courtesy of Station WRNY (New York)

Fig. 4. A section of the generator room with power plant in a Radio Broadcast Station.

much the same manner as light waves, but a certain portion apparently follow the curvature of the earth. When these waves impinge on a receiving aerial they follow the path of least resistance, and being then in the nature of electrical impulses, they follow the aerial and lead-in wire to the receiving set, and thence to the ground. The receiver is arranged in such fashion that it corresponds electrically to the antenna circuit of the transmitting station. That is to say, the receiver is placed in resonance with the particular signals desired. The usual system of detecting, either with a simple crystal or by means of vacuum tubes, may be followed by special amplification which will make the signals audible on a loud speaker. In many cases the signals are amplified in their original form (radio frequency amplification), then rectified by the detector and further increased in volume at audio frequency. (See *Modulation*, *Speech Amplifier*, *Electromagnetic Waves*, *etc.*)

BROAD TUNING—A term used to designate the lack of *selectivity* in reception or in transmission, the use of several *waves* rather than a sharp, pure wave. If a receiving set is of such a nature that it is difficult or impossible to receive the desired signals without hearing any other, the set is said to be "broadly tuned." (See *Tuning*, also *Selective Tuning*.)

BROKEN CIRCUIT—See *Open Circuit*.

BROWN AMPLIFYING RELAY—An arrangement whereby the comparatively weak signals received by a *crystal detector*, usually of the *carborundum type*, may be *amplified* or increased in intensity, either to allow a recording machine to be used or merely to make the signals more readily audible in the ear phones. The device consists essentially of an arrangement for controlling and stepping up the vibrations in conjunction with a small *battery* or *dry cell*.

One example of such a relay is shown in the illustration. In this case the soft iron *cores* cc are *magnetized* by the *permanent magnets* mm, the poles of which are marked n and s. R1 is a fine wire winding such as is ordinarily used in ear phones and R2 is a larger winding placed over the main magnets mm. Directly above the soft core cc with the windings R1 is placed a steel vibrating tongue V having a contact D which touches lightly contact G. Contact G is of a special alloy of osmium and iridium and the contact D is a carbon button, which in contact with G forms a *microphone*. The winding R2 is joined with the telephone P and the battery B, all in series with the microphonic contact. When current (signal current) enters the leads A—E through the windings R1 the result is a change of *flux* in the core of the magnet which vibrates the tongue V. This in turn varies the current of the battery B and causes correspondingly greater sounds in the ear phones P. Several of these relays may be connected together to permit a much greater *amplification* of the original signals.

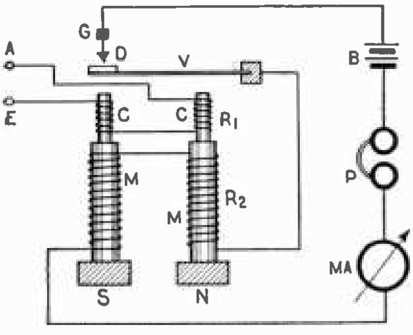

The Brown Amplifying Relay Circuit.

BROWN & SHARPE GAUGE—The American wire gauge adopted as standard for wires for electrical purposes. Commonly known as B. & S. The various sizes run from 40 to 0000, 40 having a diameter of .00314 inch, and 0000 being .46 inches in diameter. We refer to wire having a diameter of .00314 inch as being 40 B. & S. gauge. (*See page 26 for table.*)

BROWN, Sidney George—Born in Chicago, U. S. A., in 1873, of English parents, he was educated at Harrogate and London University. An early study was made by him of the subject of submarine telegraphy. During these investigations he invented the magnifying relay for cables. Another important invention was the cable drum relay. In the year 1898 he invented the magnetic shunt. Since that time many of his activities have been directed to the solution of problems in telephone and wireless telegraphy.

In particular the radio experimenter will know the telephone receivers bearing his name and the important developments of the *microphone relay*. In the field of land telegraphy and telephony, his activities have resulted in the invention of such items as the carbon telephone relay system, which is largely used on land trunk lines for the transmission and reception of telephony.

He is a Fellow of the Royal Society and Vice-President of the Radio Society of Great Britain. His writings on technical subjects are extensive, and numerous valuable patents have been taken out by him, including the vacuum tube *oscillation generator* in 1916 and ionic electric relays in 1918.

BRUSH—A device for collecting current from the *commutator* (*q.v.*) of a *dynamo* or supplying current to the commutator of a motor. Brushes are made in a great variety of forms, the most common being carbon blocks or brass or copper gauze. The illustration shows two different types.

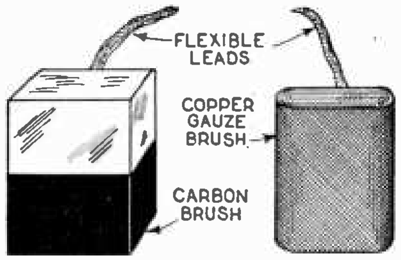

Two different types of Brushes. At the left is shown a carbon brush, and at the right a copper gauze brush.

BRUSH ANGLE—The angle formed by a brush in its contact with the *commutator*. The angle is usually 45°. If the brushes do not bed properly, sparking is apt to occur.

BRUSH DISCHARGE—A faintly luminous discharge which takes place on the surface of *conductors* charged to *high potential*, due to *ionization*. This effect can be noted at times on the *antenna* of a powerful transmitting station. (See *Corona*.)

BRUSH HOLDER—Metal clamp capable of adjustment which holds the *brush* in position on the *commutator* of a dynamo or motor. (See *Brush*.)

BRUSH LOSS—The loss in *watts* or power due to the friction of the brushes against the *commutator*. This loss is at a minimum when the brushes make perfect contact at the proper angle.

BRUSH PRESSURE—A term used to designate the pressure with which the brushes bear on the *commutator* of a dynamo machine and also referring to the voltage or *electrical pressure* delivered at the brushes.

B. S. G.—Abbreviation for British Standard Gauge of Wire, commonly known as B. S. (See *British Standard Wire Gauge*.)

B. T. U.—Abbreviation for "British Thermal Unit," being the commonly used heat unit. The amount of heat required to raise the temperature of a pound of water one degree Fahrenheit at ordinary atmospheric pressure.

BUCHER, Elmer E.—Born Akron, Ohio, Nov. 11, 1885. Educated at high school and private tutors. An American pioneer, Experimental Engineer with De-Forest Wireless Telegraph Company in 1903; constructed several high power stations in Middle West and Gulf Coast and later for United Wireless, which absorbed the Deforest Company. Organized training school for United Wireless and instituted first radio schools for Y. M. C. A. in New York City. Supervised commercial operations and conducted research work for United Wireless; later instructing engineer Marconi Wireless Tel. Co. Technical Editor "Wireless Age" 1913 to 1917. Appointed Commercial Engineer for newly formed Radio Corporation of America in 1920, and since 1922 has been in charge of general sales for that corporation. Author of numerous standard works on radio, including "Practical Wireless Telegraphy," "Wireless Experimenters' Manual," "Vacuum Tubes in Wireless Communication," and many others. Holds numerous U. S. patents on radio systems and devices.

BUCKLED DIAPHRAGM—Warping of the *diaphragm* of a phone as used in radio. Excessive voltages if applied direct to the head-phones will often bend the diaphragms. Any such defect may seriously impair the operation of the part and when it is noted that a diaphragm is bent or warped through rough usage or the application of too great voltages, it is best to replace it immediately.

BUCKLING OF PLATES—During discharge of a *storage battery* the plates gradually expand, owing to the fact that lead sulphate has about twice the volume of the same quantity of lead peroxide. Should this expansion or discharge take place too quickly, the plates will bend or buckle. (See *Storage Battery*.)

BUNSEN BURNER—A form of gas burner frequently used in radio construction for heating the soldering iron where electricity is not available. By special arrangement of the parts, the gas is combined with the proper amount of air before it reaches the burning part. This is done by driving the gas through a tube with small holes drilled to allow the air to be sucked in. This method promotes complete combustion and gives a non-luminous flame that heats the soldering iron without

THE BROWN & SHARPE (B. & S.) COPPER WIRE GAUGE—Below is a comprehensive copper wire table, which is the standard one in use in the United States. The resistance and cross-sectional areas of various sizes of conductors are given, so that considerable calculating can be done with the data at hand. A mil means 1-1000 (one thousandth) of an inch. Hence a wire with a diameter of 9 thousandths of an inch could also be stated as having a diameter of 9 mils. The area of cross-section of the wire in circular mils is found by squaring the diameter in mils; thus the 9 mil conductor would have 9 times 9, or 81 circular mils area. This is the end area of the wire, of course, and is used in all electrical figuring. If the resistance per 1,000 feet of a certain size wire is, say, 50 ohms, then 500 feet of the same size wire would have a proportionately less amount of resistance, or ½ of 50 ohms, that is, 25 ohms. The resistance per foot is found by dividing the resistance in ohms per 1,000 feet by 1,000. This wire table is for bare copper wire. Conductors, according to the B. & S. standard gauge, halve their sectional area in circular mils (Cir. Mils) for every 3 gauge sizes smaller wire; and double their sectional area in Cir. Mils for every 3 gauge sizes increase.

Gauge Number.	Size.		Weight and Length.			Resistance.			Carrying Capacity, 2,000 Amperes p. sq. in. section.
	Diameter in Mils.	Square of Diameter or circular Mils.	Grains per Foot.	Pounds per 1000 Feet.	Feet per Pound.	Ohms per 1000~~0~~ Feet.	Feet per Ohm.	Ohms per Pound.	Amperes.
0000	460.000	211600.0	4477.2	639.60	1.564	.051	19929.7	.0000785	430
000	409.640	167804.9	3550.5	507.22	1.971	.063	15804.9	.000125	262
00	364.800	133079.0	2815.8	402.25	2.486	.080	12534.2	.000198	208
0	324.950	105592.5	2236.2	319.17	3.133	.101	9945.3	.000815	165
1	289.300	83694.49	1770.9	252.98	3.952	.127	7882.8	.000501	130
2	257.630	66373.22	1404.4	200.63	4.994	.160	6251.4	.000799	103
3	229.420	52633.53	1113.6	159.09	6.285	.202	4957.3	.001268	81
4	204.310	41742.57	883.2	126.17	7.925	.254	3931.6	.002016	65
5	181.940	33102.16	700.4	100.05	9.995	.321	3117.8	.003206	52
6	162.020	26250.48	555.4	79.34	12.604	.404	2472.4	.005098	41
7	144.280	20816.72	440.4	62.92	15.893	.509	1960.6	.008106	32
8	128.490	16509.68	349.3	49.90	20.040	.643	1555.0	.01289	26
9	114.430	13094.22	277.1	39.58	25.265	.811	1233.3	.02048	20
10	101.390	10381.57	219.7	31.38	31.867	1.023	977.8	.03259	16
11	90.742	8234.11	174.2	24.89	40.176	1.289	775.5	.05181	13
12	80.808	6529.93	138.2	19.74	50.659	1.626	615.02	.08237	10.2
13	71.961	5178.39	109.6	15.65	63.898	2.048	488.25	.13087	8.1
14	64.084	4106.75	86.87	12.41	80.580	2.585	386.80	.20830	6.4
15	57.068	3256.76	68.88	9.84	101.626	3.177	306.74	.33133	5.1
16	50.820	2582.67	54.67	7.81	128.041	4.582	243.25	.52638	4.0
17	45.257	2048.19	43.33	6.19	161.551	5.183	192.91	.83744	3.2
18	40.303	1624.33	34.37	4.91	203.666	6.536	152.99	1.3312	2.5
19	35.390	1252.45	26.50	3.786	264.136	8.477	117.96	2.2392	1.96
20	31.961	1021.51	21.60	3.086	324.045	10.394	96.21	3.3438	1.60
21	28.462	810.09	17.14	2.448	408.497	13.106	76.30	5.3539	1.28
22	25.347	642.47	13.59	1.942	514.933	16.525	60.51	8.5099	1.08
23	22.571	509.45	10.77	1.539	649.773	20.842	47.98	13.334	.80
24	20.100	404.01	8.55	1.221	819.001	26.284	38.05	21.524	.63
25	17.900	320.41	6.77	.967	1044.126	33.135	30.18	34.298	.50
26	15.940	254.08	5.38	.768	1302.083	41.789	23.93	54.410	.40
27	14.195	201.49	4.26	.608	1644.737	52.687	18.98	86.657	.31
28	12.641	159.79	3.39	.484	2066.116	66.445	15.05	137.283	.25
29	11.257	126.72	2.69	.384	2604.167	83.752	11.94	218.104	.20
30	10.025	100.50	2.11	.302	3311.258	105.641	9.466	349.805	.16
31	8.928	79.71	1.67	.239	4184.100	133.191	7.508	557.286	.13
32	7.950	63.20	1.33	.190	5263.158	168.011	5.952	884.267	.098
33	7.080	50.13	1.06	.151	6622.517	211.820	4.721	1402.78	.078
34	6.304	39.74	.847	.121	8264.463	267.165	3.743	2207.98	.062
35	5.614	31.52	.658	.094	10638.30	336.81	2.969	3583.12	.049
36	5.000	25.00	.525	.075	13333.33	424.65	2.355	5661.71	.039
37	4.453	19.83	.420	.060	16666.66	535.33	1.868	8922.20	.031
38	3.965	15.72	.315	.045	22222.22	675.22	1.481	15000.5	.025
39	3.531	12.47	.266	.038	26315.79	851.789	1.174	22415.5	.020
40	3.144	9.88	.210	.030	33333.33	1074.11	.931	35803.8	.015

depositing soot as in the case of an ordinary luminous flame.

BURSTYN, Dr. W.—Born in Austria, 1877. Educated at Vienna University. Noted as: Developer of the *quenched spark system* in radio transmission, upon which he worked in conjunction with Baron Lepel in 1907 to 1912.

BUS BAR—A single bar, usually copper, which serves as a common connector for a number of pieces of apparatus. Also referred to as Omnibus Bar.

BUS BAR WIRE—Square copper wire, usually tinned, much used in making radio connections. This wire is generally furnished in lengths of two or two and a half feet and is particularly efficient for radio work, as it offers a low *resistance* connection, makes a neat appearance, and is suitable for soldering at joints.

BUSHING—Pieces of insulating material, usually fibre or rubber composition, used to *insulate* parts of radio apparatus or in mounting machine screws or shafts of *condensers* and other devices. The term is also applied to brass or other metal bearings through which the shaft of a piece of apparatus is run.

BUTT JOINT—A method joining two lengths of wire together by placing them end to end or at right angles and soldering the joint. The illustration shows a common butt joint.

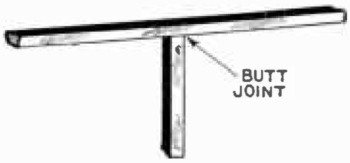

Showing how a Butt Joint is made.

BUZZER—An electric call signal device that makes a buzzing noise caused by the rapid vibrations of a contact breaker in a circuit using a small *battery* or *cell*. (See *Testing Circuit.*)

As the *current* flows from the posts A B the *electro-magnets* C and D are energized and draw down the vibrator spring F. This breaks the contact between E and F and as the current no longer flows the magnets release the spring F, which returns to its normal position and establishes the circuit once more. This making and breaking of the circuit takes place so rapidly that the vibrator makes a buzzing or humming sound, the speed with which it vibrates determining the pitch of the note. (See *Frequency*, also *vibration.*)

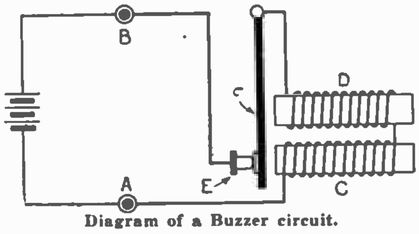

Diagram of a Buzzer circuit.

BUZZER EXCITER—A method of using a *buzzer* to produce local signals in a crystal receiving circuit to permit sensitive adjustment of the *detector*. The illustration shows a common method of producing signals to test the receiving set in this manner. In this case a small coil having a few turns of wire is placed in the buzzer circuit as indicated and a similar one is placed in the ground lead of the receiving set. When the button is depressed the buzzer is set in vibration and a transfer of energy takes place between L1 and L2. This causes an *oscillating* impulse (*q.v.*) to flow through the receiver circuit. The detector can then be adjusted to the most sensitive point at which it can be left until outside signals are heard. Another but less efficient method of exciting the circuit is to connect a short length of wire from the contact point of the buzzer to the ground connection of the receiving set. The former system is much more practical and furnishes better test signals.

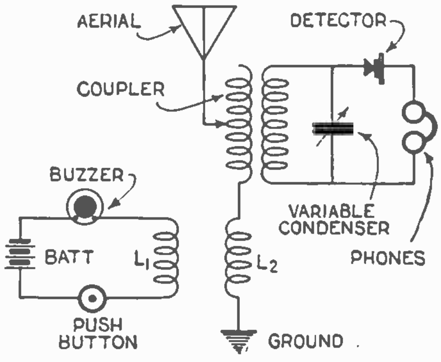

Buzzer Exciter.

Coil L1 being placed in *inductive relation* to coil L2, which is in the antenna circuit, undergoes a change of *flux* when the buzzer is *excited*. This sets up a *difference of potential* across coil L2 which charges the antenna ground system and produces oscillation at a definite *frequency* according to the adjustment. Now if the secondary circuit is tuned to resonance with the primary circuit, the signals will be audible in the phones when the detector is properly adjusted. Also called *Buzzer Wave Generator*.

BUZZER MODULATION—See *Modulation.*

BUZZER PRACTICE SET—A combination of a *buzzer* and signalling key arranged on a baseboard used for the purpose of practicing signalling.

B. W. G.—The abbreviation for Birmingham Wire Gauge—a system of gauging wire used in England. (See *B & S Gauge*, also *British Standard Wire Gauge.*)

BY-PASS CONDENSER — See *Condenser*, also *Filter.*

C

C—The chemical symbol for carbon. When printed in *italics*, it is also the symbol for *coulomb* in electrical practice. It is the international symbol for *capacity.*

CABLE—A number of wires stranded together, used to carry electrical currents, heavily insulated and often covered by a sheath or outer coating of lead, rubber, gutta percha, or braided silk or cotton. The term is used very often in a broad sense to signify any heavy wire or strand of wires that may be insulated more thoroughly than is ordinarily the case. (See *Conductor*, *Stranded Wire*, etc.)

CADMIUM—A chemical metallic element somewhat resembling zinc. In color, white with a bluish tinge. It is useful in fusible alloys because of its property for uniting readily with other metals, having a low melting point. Its electrical conductivity is about 25% of that of silver. It is very useful in the construction of a crystal detector, the molten metal holding the crystal firmly and as it melts at a low temperature the heat will not affect the sensitivity of the crystal. Cadmium is also used in making cells, notably the *Weston Cell.* (See *Weston Cell.*)

CAGE ANTENNA—An aerial having a number of wires arranged in cylindrical form, suspended either vertically or as a horizontal top. So named because of its resemblance to a cage. (See *Aerial.*)

CALCULATION OF CAPACITY—The mathematical or comparative determination of *capacity.* There are various formulas in use for calculating capacity under various conditions, such as for capacity of various types of condensers, or the capacity of a circuit. Rough approximations of capacity of condensers might be obtained by direct comparison with a known value. Other formulas are used for the calculation of antenna capacity. For determining the capacity of flat condensers, such as fixed condensers used for transmitting or receiving circuits, the following formula is generally used:

$$C = 0.0885\,K\,\frac{S}{r}$$

In this case, C is the capacity in micro-micro farads; to obtain result in the usual units of micro-farads, divide the result by one million. S is the surface area of one plate in square centimeters; r is the thickness of the dielectric in centimeters and K a constant depending on the material between the plates of the condenser. (See *Dielectric Constants.*)

CAL-ELECTRICITY — Electricity produced in the secondary of a transformer due to changes in temperature of the core. This current is in addition to the current normally induced in the secondary. (See *Transformer*, *Core*, also *Core Losses.*)

CALIBRATION — A process whereby instruments such as *galvanometers* *ammeters*, *wavemeters*, etc., are made to indicate certain values. Before such

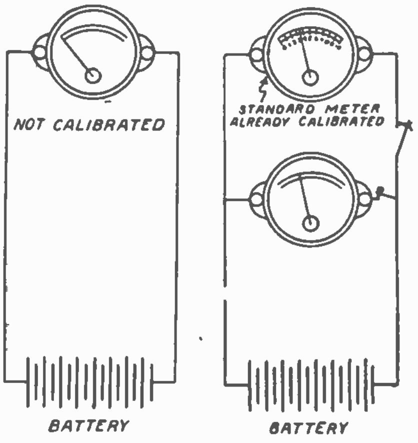

Fig. 1. Fig. 2.

Calibrating a Meter

instruments can be made use of, it is necessary to mark the dials in such a manner that the indicator showing on the dial or scale will have a definite meaning in terms of electrical values. Thus, a voltmeter before calibration is shown in Fig. 1. When it is placed across a battery, the pointer will move to some position on the white space, but without indicating any definite value. Now by testing with another meter which is already marked or calibrated in volts, we find, for example, that when the pointer is at the middle point in its swing, it indicates 10 volts. Similarly various other positions are found to indicate more or less voltage by comparison with the standard. This is shown in Fig. 2. It will be understood that one of the small switches must be open at all times. The reading of the calibrated meter is taken with the other meter out of the circuit (switch open), then the reading is marked on the other meter at the point indicated with the calibrated meter out of the circuit. Such a method would, of course, be inefficient and it is shown only to give a rough idea of the purpose and meaning of calibration. Another example of calibration is in the case of a dial on a receiver. Fig. 3 shows the usual form of dial used in tuning a receiving set. The dial is marked off to scale, from 1 to 100, the numbers having no particular significance. Now if we find that when the tuning dial is set at 20 it is adjusted for a wavelength of 260 meters, we can put that figure in place of the 20. The same procedure can be followed throughout the entire range of the dial, and when thus marked in terms of a definite quantity—wavelength in meters in this case—the dial is said to be calibrated.

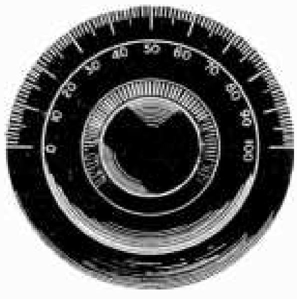

Fig. 3.

A dial as used on radio receiving sets may be Calibrated according to different wavelengths.

The actual method to be followed out in calibrating will depend on the equipment available and the accuracy desired. If a standard of comparison is available, it saves a great deal of work. Thus, a voltmeter can easily be calibrated by means of a standard meter, already calibrated, and a source of steady voltage controlled by a rheostat. An ammeter can be calibrated in the same manner by using a standard ammeter, and a wavemeter can be calibrated from another wavemeter of known values.

Calibration may be done in such a manner that the readings require mathematical solution to determine their value in terms of actual volts, amperes, meters and so on, or it may be *absolute*, in which case the readings are directly in terms of the quantity and require no solution. (See *Voltmeter Calibration, Wavemeter Calibration*, also *Rough Calibration.*)

CALIDO — An alloy containing nickel and chromium with a small percentage of iron. The melting point is very high, about 1550 degrees centigrade. Calido wire is much used as a *resistance*, particularly for heating devices. The chief characteristics are as follows: maximum working temperature 1100 degrees centigrade; microhms per cubic centimeter at 20 degrees centigrade, 100; temperature coefficient per degree centigrade, 0.00034; specific gravity, 8.2.

CALLAND CELL—A primary cell used in French telegraph work. It is a form of gravity cell having a negative electrode of copper and a positive electrode of amalgamated zinc, the electrolyte being zinc sulphate. Crystals of copper sulphate are used as a depolarizer. (See *Gravity Battery.*)

CALORIE — The unit of heat in the C. G. S. (Centimeter-Gram-Second) (q.v.) system. It is equivalent to the amount of heat necessary to raise the temperature of one gram of water from 0° to 1° centigrade. (See *Thermal*, also *C. G. S.*)

CALORIMETER—In electrical practice, an instrument used to measure the heat generated in a conductor carrying an electric current.

CAMBRIC, VARNISHED — Varnished muslin used as an insulating material. Also known as Empire Cloth. (See *Spaghetti.*)

CANAL RAYS—When an electric discharge takes place between the *anode* and a perforated *cathode* in a vacuum tube, fine pencils of light are seen to pass through the perforations in the cathode. These rays are called *canal rays*, and consist of positively charged particles. They produce phosphorescence on the wall of the tube.

Canal rays travel at a lesser velocity and in the opposite direction to the cathode rays. This fact and also the fact that the rays are deflected by powerful electric or magnetic fields in the opposite direction to cathode rays, is taken as proof that they consist of positively charged particles. (See *Cathode Rays.*)

CAPACITANCE — A term very often used as synonymous with *capacity*. Due to the fact that capacity may refer to the current carrying ability of a conductor and also to electrostatic capacity, it has been suggested that capacitance be used to refer only to *electrostatic capacity* of a body or device. Capacity would then refer to *current carrying ability.* (See *Capacity*, also *Electrostatic.*)

CAPACITATIVE REACTANCE—That part of the *reactance* of a circuit carrying alternating current which is due to the *capacity* in the circuit. (See *Reactance* also *Impedance.*)

CAPACITATIVE COUPLING — See *Coupling.*

CAPACITY — Generally speaking, the quantity of electricity in any form which a body is able to store or contain. The term is usually qualified to denote the particular case, such as *electrostatic capacity*, which is a measure of the ability of a condenser to store up energy in the form of electrostatic charges; also *current-carrying capacity* of a conductor, which is considered as the ability of a wire to carry a certain amount of electric current without overheating. The tendency in electrical literature is to use the term capacitance as applying to the electrostatic capacity of a condenser. Thus, a certain condenser may have a capacitance of 2 micro-farads. The unit of electrostatic capacity is the *farad.* This is understood as the capacity of a condenser that will store one *coulomb* of electricty under an electromotive force of one *volt.* When V is expressed in volts, C in farads, and Q in coulombs, $C = Q/V$.

CAPACITY CIRCUIT — An electrical circuit in which the *capacity* is very large compared with the *inductance.* The inductance may be considered negligible.

CAPACITY CONVERSION FACTORS —In the calculation of *capacity*, as the capacity of a condenser, there are several different units. For instance, a *farad* is a large unit of capacity and is equal to one million *microfarads*, or conversely, a microfarad is equal to one millionth of a farad. In the same way a farad is equal to one billionth of an *abfarad.* The four units used in capacity measurement are the *farad, micro-farad, micromicrofarad, statfarad* and *abfarad.* The following table gives the conversion factors for these various units:

$$1 \text{ abfarad} = \begin{cases} 10^{9} \text{ farads} \\ 10^{15} \text{ mfds.} \\ 9 \times 10^{20} \text{ stfds.} \end{cases}$$

$$1 \text{ farad} = \begin{cases} 10^{-9} \text{ abfd.} \\ 10^{6} \text{ mfds.} \\ 9 \times 10^{11} \text{ stfds.} \end{cases}$$

$$1 \text{ mfd.} = \begin{cases} 10^{-15} \text{ abfd.} \\ 10^{-6} \text{ farads} \\ 9 \times 10^{5} \text{ stfds.} \end{cases}$$

$$1 \text{ stfd.} = \begin{cases} 1/9 \times 10^{-20} \text{ abfd.} \\ 1/9 \times 10^{-11} \text{ farad} \\ 1/9 \times 10^{-5} \text{ mfd.} \end{cases}$$

$$1 \text{ mmfd.} = 10^{-6} \text{ mfd.}$$

In the table above, farad is abbreviated Fd.; abfarad is shown as abfd.; microfarad is mfd., micromicrofarad is mmfd. and statfarad is stfd. In dealing with large figures containing many ciphers the amount is shown as 1,000,000 is 10^6 and $\frac{1}{1{,}000{,}000}$ is 10^{-6}, etc. (See *Capacity, Farad*, also *Unit.*)

CAPACITY EARTH or CAPACITY GROUND—A substitute for the usual ground connection where the wires or plates are buried beneath the ground or attached to a water or steam pipe. It is in effect a second aerial, and is placed either under the regular aerial or to one side. It is more used in radio transmission than in reception. (See *Counterpoise.*)

CAPACITY ELECTROSTATIC — See *Capacity* also *Electrostatic Capacity.*

CAPACITY FREQUENCY FACTOR—The relation between the apparent capacity of a condenser and the *electrostatic capacity.* (See *Capacity, Condenser* also *Electrostatic.*)

CAPACITY OF CONDENSERS IN PARALLEL—When several condensers are connected in parallel the resultant capacity is the sum of the individual capacities. It is written:

$$\text{Capacity} = C_1 + C_2 + C_3, \text{etc.}$$

It will be evident that the resultant capacity of condensers connected in parallel is just the reverse of case for resistances, where the total resistance is less than that of the smallest individual resistance in the parallel connection. (See *Resistance Measurement* also *Capacity of Condenser in Series*, also *Condensers in Series* and *Condensers in Parallel.*)

CAPACITY OF CONDENSERS IN SERIES—When several condensers are connected in series with each other the resultant capacity is always less than the capacity of the smallest condenser in the group. The capacity of condensers in series is given as follows:

$$\text{Capacity} = \cfrac{1}{\cfrac{1}{C_1} + \cfrac{1}{C_2} + \cfrac{1}{C_3}, \text{ etc.}}$$

The resultant capacity is thus the reciprocal of the sum of the reciprocals of the several capacities. (See *Condenser, In Series.*)

CAPACITY MEASUREMENT OF ANTENNA—A simple method for the measurement of the capacity of an antenna as given by Austin follows: The illustrations show the general arrangement. The inductance L in Fig. A is a coil that will increase the wavelength (natural) of the antenna system 4 or 5 times; C is a calibrated variable condenser and W a standard wavemeter. In operation the coil L is connected in the antenna circuit, the buzzer is used to excite the antenna circuit and the wavelength is measured by w. The coil L is then taken out of the antenna circuit and shunted by the condenser C. The condenser C is then varied until the wavelength of the combined coil L and condenser C is the same as that obtained before with the coil in the antenna circuit but without the variable condenser. The capacity of the condenser now will be approximately that of the antenna, ignoring the distributed inductance of the antenna system. Fig. A shows the first operation, the measurement of the antenna with the coil in the circuit, and Fig. B shows the measurement of the wavelength of the coil and condenser. (See *Mutual Capacity, Wavemeter, Natural Wavelength,* also *Distributed Capacity.*)

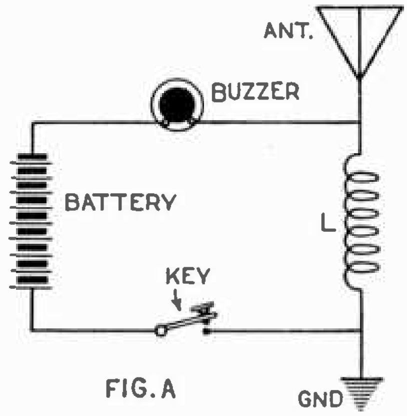

FIG. A

The first operation in the measurement of the antenna. The wavemeter is shown below.

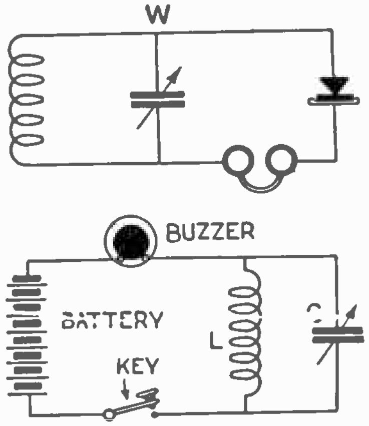

FIG. B

Circuit for the measurement of wavelength of coil and condenser.

CAPACITY MUTUAL—See *Mutual Capacity.*

CAPACITY RESISTANCE—The resistance to alternating currents offered by a body possessing electric capacity. (See *Capacity,* also *Resistance.*)

CAPACITY SWITCH—Any switch used in a circuit to introduce or cut out capacity (condenser).

CAPACITY (UNIT OF)—The unit of capacity is the farad. A condenser has a capacity of 1 farad when 1 coulomb is required to raise its potential from zero to 1 volt. Since the farad is very large, its millionth part, or the microfarad is generally used.

CARBON—A non-metallic chemical element used for many purposes in electricity and radio. It exists in three forms, two crystalline and one non-crystalline. The first two are diamond and graphite, the third charcoal. Carbon is used as an electrode in electric arcs, as one of the poles in certain types of primary cells, as a material in place of wire for various forms of resistance, and as brushes in motors and generators. (See *Diaphragm, Carbon.*)

CARBON RHEOSTAT—A rheostat or variable resistance using carbon in place of the customary wire. In its simplest form a carbon rheostat may be merely a piece of carbon having a sliding contact. This is shown in Fig. 1. Owing to the relatively high resistance of carbon used in this manner, such a controlling resistance will generally be termed a *potentiometer* (q.v.) and used as such.

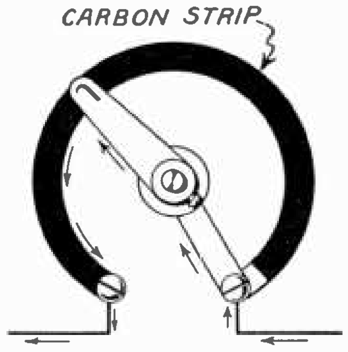

Fig. 1.—A carbon rheostat with a rotary sliding contact.

The most practical form of a carbon rheostat is known as a carbon pile rheostat. This type uses a number of flat pieces of carbon of relatively low resistance placed together and arranged with a knob and screw in such manner that the pressure can be varied. This change in pressure alters the resistance of the mass and allows control of the current for vacuum tubes. The same principle is used for very high resistances as well as for low values. The principle will be seen from the illustration Fig. 2.

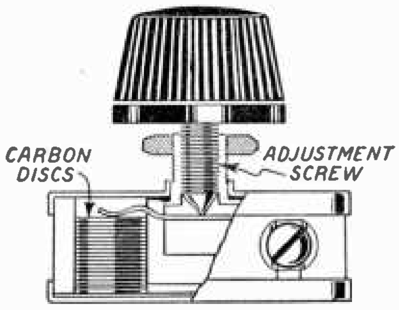

Fig. 2.—A carbon pressure type rheostat.

CARBORUNDUM—An abrasive material, a product of the electric furnace, composed of silica and carbon. It is used in radio as a crystal rectifier, although the strides of radio have practically rendered it obsolete, many other better crystals being in use. The chief difficulty with its use as a rectifier or crystal detector was the fact that for efficient operation it required a cell or battery and potentiometer to regulate the current flowing through it. The simple crystal circuit using a carborundum detector is shown in the illustration. C is a tuning coil, D the carborundum crystal and contact, P a potentiometer and T the head phones. (See *Rectifier,* also *Detector.*)

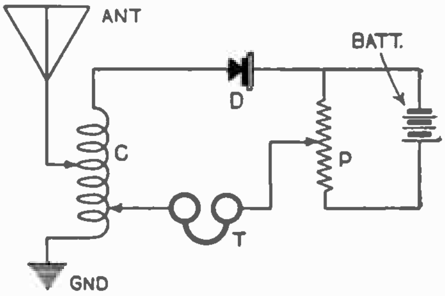

Diagram of a radio circuit employing a carborundum crystal detector.

CARDBOARD TUBING—Tubing made of laminations of paper pasted together. They are used for home made coils of all descriptions, and while not as efficient as the composition tubing, can be used for many purposes without any great loss. As a general rule it is better to use some of the numerous compositions, such as bakelite or rubber compounds.

CARPENTIER, JULES—French radio expert. Born in Paris, in 1854, he joined the Ecole Polytechnique in 1871, and in 1876 was appointed principal stores engineer of the Lyons Railway Company, making a special study of electricity. For his work in electricity he obtained in 1881 the cross of Chevalier of the Legion of Honor. One of the early pioneers of radio in France, he founded the Compagnie Generale Radiotelegraphique, afterwards absorbed in the Compagnie Generale de Telegraphie sans Fil. Carpentier is a member of the Academie des Sciences, Commander of the Legion of Honor, and President of many scientific societies.

CARRIER WAVE—The *continuous wave* (q.v.) of radio frequency generally thought of as carrying the voice or music waves from the radio broadcasting station. Actually it is a radio wave of high frequency, which is altered in amplitude by the music or speech transmitted. In action, this continuous wave has another wave super-imposed on it. This other wave having wave form and amplitudes determined by the voice or other sounds being transmitted. The illustration shows the carrier wave, the waves representing speech or music and the combination of the two. This effect of varying the amplitude or strength of a continuous wave by means of some sort of program is known as *modulation.* (q.v.) The wave representing the speech or music is imposed on the carrier wave resulting in the combined wave which reaches the listener and is translated by suitable apparatus into approximately the same sounds that are transmitted. (See *Modulation, Broadcasting*, also *Speech Vibrations*.)

CARRYING CAPACITY—See *Current Carrying Capacity.*

CARTRIDGE FUSE—A fuse in which the fuse wire is surrounded by some non-inflammable substance, enclosed in a cartridge-like tube and having brass lugs soldered to caps at ends. Used to prevent the hot wire from "flying" when fused. (See *Fuse*.)

CASCADE—A term applied to pieces of apparatus connected together in series, particularly vacuum tubes. In this case, the arrangement would be such that each vacuum tube would amplify the signal output of the preceding tube. The tubes will be so connected that the output of one is introduced as input to the next and so on the total result being greatly increased signal strength from the output of the last tube. This term may apply to either radio frequency or audio frequency amplification. (See *Amplifier, Radio Frequency*, also *Audio Frequency*.) A simple cascade arrangement is shown in the illustration. Fig. 1 is a circuit with two stages of cascade radio frequency amplification, and Fig. 2 a circuit using two stages of audio frequency amplification with the usual detector tube in each case. A combination of both radio frequency and audio frequency stages of amplification might also be used, the term cascade amplification also applying here.

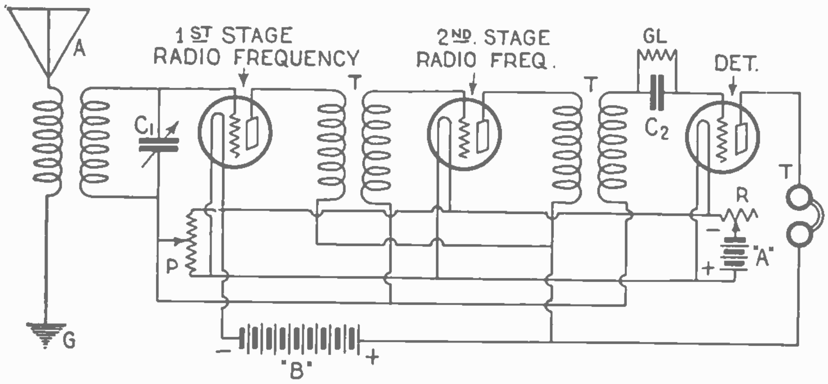

Fig. 1.—Circuit diagram showing two stages of cascade radio frequency amplification and detector.

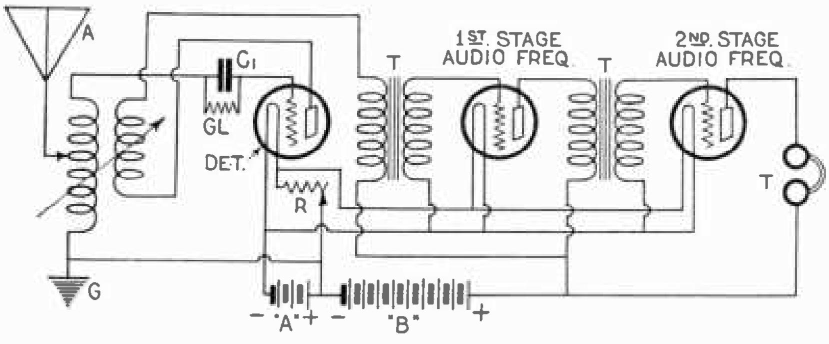

Fig. 2.—Diagram of detector and two stage cascade audio frequency amplifier.

CATHION or CATION—The charged particles which appear at the cathode or move toward it through the electrolyte of an electrolytic cell. (See *Anode*.)

CATHODE or KATHODE—A *negative electrode*. The term is often applied to the filament of a vacuum tube to distinguish it from the anode or plate. (See *Cell, Electrolysis*, also *Anode*.)

CATHODE RAY TUBE—A vacuum tube having a high vacuum, that is from which practically all air has been exhausted, used in the production of cathode rays. These tubes in their modern form are used for many purposes in studying the nature and form of electric waves. The illustration shows one of the latest types of tube, known as the *Cathode Ray Oscillograph*. This tube is used for the following purposes: examination of the wave forms of various rectifiers, study of wave forms of different types of generators, study of vacuum tube characteristics, examination of the characteristics of X-ray and other types of tubes, examination of radio frequency waves modulated as in broadcasting, measurement of hysteresis and dielectric loss in various materials, study of phenomena in arcs and sparks, together with numerous other associated effects. The development of this tube is one of the marvels of science, enabling the research engineer to virtually see the electric wave, and to study exhaustively, electric phenomena otherwise invisible.

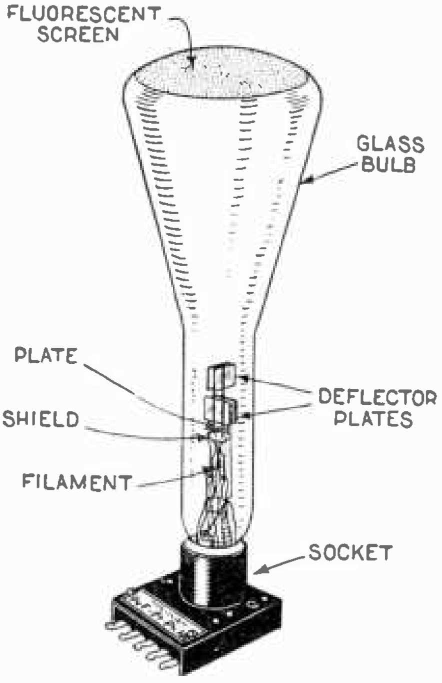

A cathode ray tube with socket.

The Western Electric modification of the Braun Cathode ray tube or oscillograph, is a three electrode glass tube about 30 centimeters (about twelve inches) long, in the form of a cylinder an inch and three-quarters wide, spreading out in conical shape, the tube being about three and a half inches wide at the top. The *cathode* is a filament coated with active oxides and arranged to emit the number of electrons required for the cathode rays, at a dull red heat. The *anode* or plate is a small platinum tube set very near the filament. Between the filament and the anodes is a small circular screen having a hole just a little smaller than the circular filament. A battery of several hundred volts is connected between the filament and plates with the positive terminal connected to the plate. The electrons emitted from the hot filament are controlled by this field in the same manner as in an ordinary vacuum tube. A small portion of these electrons pass completely through the tubular anode and constitute the *cathode* rays. These rays are passed between two pairs of deflecting plates set at right angles to each other, and fly against the large end of the tube. This end is covered with a fluorescent mixture which renders the rays visible. These rays are deflected in various ways. For example, if an alternating current is applied the visible spot on the screen will be drawn out into a line. (See *Electron, Vacuum Tube*, also *Wave Analysis*.)

CATHODE RAYS—The stream of electrons or electrical particles sent out from the *cathode* or filament of a vacuum tube. These rays are negatively charged. (See *Cathode, Electrons, Vacuum Tube* also *Oscillograph*.)

CAT-WHISKER—The fine wire used with certain crystal detectors to make contact with the crystal. Usually a springy metal such as phosphor bronze wire.

"C" BATTERY—One or more small cells having a voltage generally between 2 and 10 volts, used in the grid

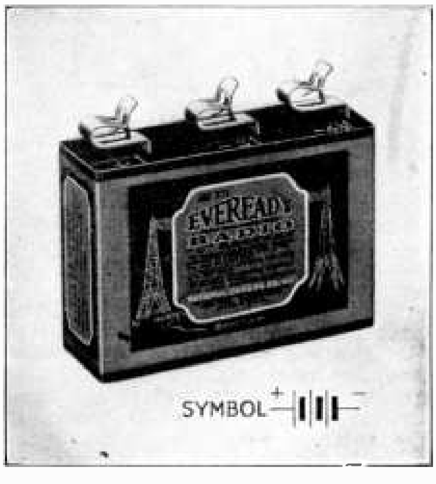

Fig. 1.—A 4½ volt "C" battery. This type of battery has a common positive terminal and is tapped at negative terminals for 3 and 4½ volts.

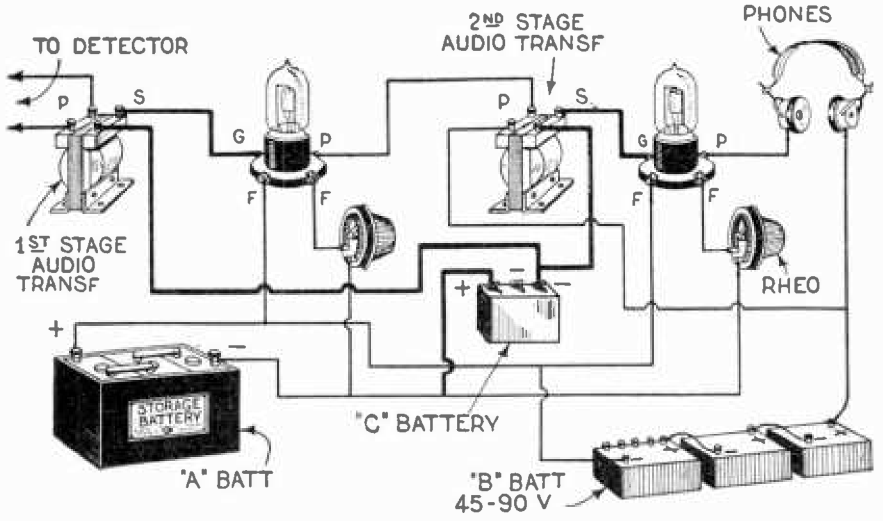

Fig. 2.—A diagram showing how the "C" battery is connected in the audio frequency amplifier.

circuit of an amplifier tube for the purpose of impressing a negative potential on the grid of the tube. The method of connecting it in the circuit is shown in Fig. 2. Fig. 1 shows a typical 4½ volt "C" battery. (See also *Grid Bias* and *"C" Battery Curve.*)

"C" BATTERY CURVE—In order to show the effect of a "C" battery on the consumption of current from the "B" battery in an audio amplifier, a *curve* (q.v.) may be drawn as in the illustration. In this case the test was made with a type UV 201A tube used as a one stage *audio frequency amplifier* in

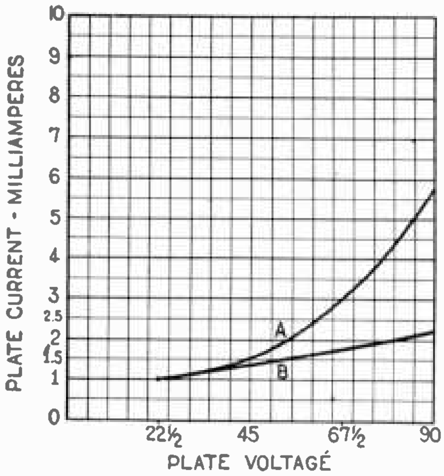

Test conducted on UV201A tube using single stage audio amplifier with conventional circuit. Plate Voltage—Abscissae. Plate Current in milliamperes—ordinates. A—Consumption curve without "C" Battery. B—Consumption curve with "C" Battery. (Note: 1.5 ma. at 45v plate rising to 5.75 ma. at 90v plate without "C". With "C" Battery consumption maximum is 2¼ ma. at 90v.)

a conventional circuit. It is readily apparent on referring to the lines A and B that when the "C" battery is in use the consumption of current by the plate is much less than without it.

CELL—A cell is one of the four chief sources of electric energy. In a cell the energy is created by an *electrochemical* process, using in its simple form two unlike metals immersed in a dilute acid or alkaline solution. In the illustration Fig. 1 the common form of electric cell is shown. C is a strip of *carbon* and Z a strip of *zinc.* These are placed in the jar in a conducting solution of sal ammoniac. If the exposed terminals from the carbon and zinc strips are connected together by a piece of wire, current will flow from one side of the cell to the other through the wire. The current flows from the carbon or *positive* pole to the *negative* or zinc pole through the wire, completing the circuit through the solution from zinc to carbon. (See *Current, Direction of Flow.*) The cell described is known as a *primary cell.*

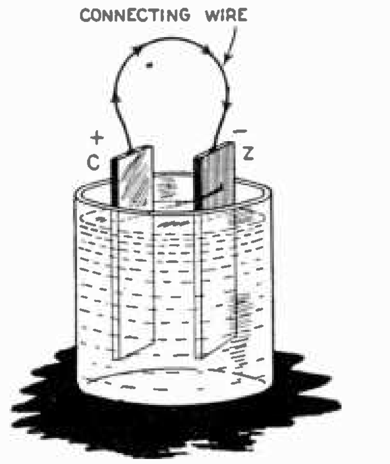

Fig. 1.—Common form of electric cell.

In Fig. 2 a simple *storage cell* or *secondary cell* is shown. In this case there are two lead plates immersed in a dilute solution of sulphuric acid. A source of primary current (primary cell) may be connected to the secondary cell in the manner shown in Fig. 3 and after being connected for a certain length of time it will be found that the energy of the primary cell has been transferred to the secondary cell and stored in it. This process of placing electric charges in a secondary cell is known as *charging*, and the current stored up in it is known as the *charge.* Now, if the wire is connected from one terminal to the other it will be found that electric energy can be withdrawn in a similar manner to that of the primary cell. A secondary cell may contain several plates and several of these cells may be connected together to form a storage battery. (See *Storage Battery, Current, Production of,* also *Anode* and *Cathode.*)

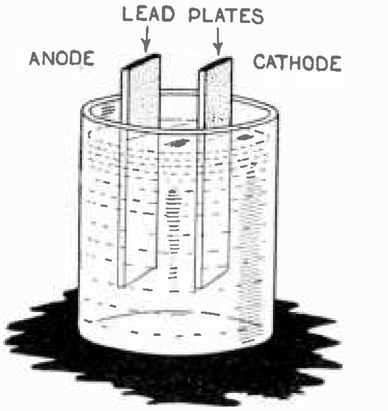

Fig. 2.—Storage or secondary cell.

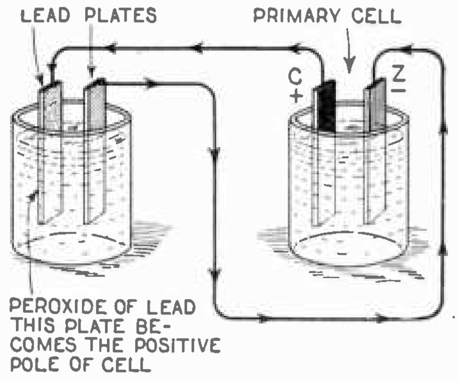
LEAD PLATES
PRIMARY CELL
C
+
Z
−
PEROXIDE OF LEAD THIS PLATE BECOMES THE POSITIVE POLE OF CELL

Fig. 3.—Primary cell connected to secondary or storage cell.

CELL, SECONDARY—When a *primary cell* is connected to a *storage* or *secondary cell,* the current from the primary cell flows into the secondary cell and in effect deposits energy in that cell. Then when the charging source (the primary cell) is withdrawn, the secondary cell can be used in practically the same manner as the primary cell. A wire connected from one terminal to the other will permit passage of current. The action of storing up energy in a storage cell is approximately as follows: When the charging current flows from plate to plate through the solution in the secondary or storage cell, the plate connected to the *positive* (+) *pole* (carbon or copper) of the primary cell, receives a brown coating of *peroxide of lead.* At the same time the other plate becomes spongy or porous. Now as one plate has received a coating while the other remains unchanged as far as its surface is concerned—it acts as a primary cell if the charging source is taken away and the two terminals connected. It now has all the essentials of an ordinary chemical cell. As long as the coating of peroxide of lead remains on the surface of one of the plates, the cell will be capable of delivering current.

When the coating has been worn off or eaten away the cell is said to be in a discharged condition. It will then be necessary to go through the same process of connecting a primary source to the terminals in order to again deposit peroxide of lead on one plate. It will therefore be evident that when the *storage cell* is charged, *current* is not actually stored in it, although in effect it has received potential energy. What has actually occurred is that the current supplied to the cell during the charging process has produced electrochemical changes, making the plates dissimilar and thus producing a difference of potential. (See *Current, Production of*). The *voltage* or *electromotive force* of primary cells varies from .06 to 1.5 volts according to the nature of the elements used and the grade of the *electrolyte* (q.v.). Secondary or *storage cells* of the lead plate type produce from 2.1 to 2.6 volts. (For more complete explanation of the theory and action of a secondary cell, see *Storage Battery.* See also *Electro Motive Force, Current* and *Voltage.*)

CELL, THEORY OF PRIMARY—It has been explained under the heading *Cell*, that a primary cell produces electrical energy by chemical action. More specifically the action of a simple cell is as follows: When the copper and zinc, or carbon and zinc placed in a cell are connected together by a conductor, (wire or other conducting object) the acid in the case of a sulphuric acid solution, attacks the surface of the zinc plate and forms a compound known as zinc sulphate. While this sulphate is being formed some of the hydrogen contained in the sulphuric acid is liberated in the form of bubbles which appear on the copper plate. Some of these bubbles escape into the surface of the solution, but others cling to the surface of the copper plate and gradually cover it with a film of hydrogen. It may be said here that the decomposition of the zinc plate in the acid solution furnishes the *electro-motive force* required to cause the flow of current between the plates, through the solution and through the conductor. Now, hydrogen is a non-conductor, and as the hydrogen gradually covers the copper plate, the surface of the plate in contact with the solution gradually decreases. The flow of *current* from zinc plate to copper plate in the solution is thus gradually diminished, until eventally we can imagine the copper plate as being entirely covered with hydrogen and no longer in contact with the solution. In addition to this insulating action of the hydrogen, it tends to set up a current within the cell in the opposite direction to the normal flow of current produced by the chemical action. A cell in this condition is said to be "polarized," and various means have been used to cut down or eliminate the *polarization effect.* The purpose of these various means is to prevent the hydrogen bubbles clinging to the surface of the copper plate, thus allowing them to escape to the top of the solution and thence into the surrounding air. (*See Polarization, Current, Voltage, also Storage Cells.*)

CENTIMETER—A measure of length in the metric system. It is one hundredth part of a meter, or approximately .3937 inch. It is used as the unit of length in the Centimeter Gram Second System (q.v.).

CENTIMETER GRAM SECOND or C. G. S. A system of units of measure employed in practically every branch of engineering, particularly in electrical practice. It is a decimal system, the centimeter being the unit of length, the gram the unit of weight and the second the unit of time. All other units are derived from these three. When the magnitude of a quantity is spoken of, it refers to the relative magnitude of the quantity when compared with some other quantity of the same nature. The magnitude used as a standard of comparison is termed a unit. Thus, if the length of a body is measured in feet, as for example, ten feet, it means that the magnitude of the body is ten times unity or ten times one foot. Similarly a *current* of ten *amperes*, is a current of magnitude ten times as great as the unit quantity—one ampere. Now an ampere is the practical unit of rate of flow of electric current. The unit has a certain definite value. (See Ampere.) The magnitude of any other similar quantity measured in amperes will mean a value equal to a certain number of units. In addition to the C. G. S. system of units, there is a practical system. This is commonly used for ordinary calculations. Under this system, the following are the main units: The *volt* is the practical unit of *electro-motive force;*—the force required to maintain a flow of current of one ampere through a resistance of one ohm. The unit of *current strength* or rate of flow of current is the *ampere;*—the strength of current maintained by a force of one volt through a resistance of one *ohm.* The unit of *resistance* is one *ohm*, and is the resistance of a *conductor* or *circuit* that will permit the passage of a current of one ampere under the force of one volt. The unit of *quantity* is the *coulomb*, which is the quantity of current flowing in a circuit when one ampere passes a given point during one second of time. The *watt* is the practical unit of *electrical power*, being the power of a current of one ampere flowing in a circuit under the pressure of one volt.

It will thus appear that the practical system is actually based on the C. G. S. system. Calculations of an involved nature necessarily make use of the C. G. S. system, and when multiples or sub-multiples are used they are expressed in Greek and Latin prefixes. For example, a million ohms is 10^6 ohms and is expressed as a *meg-ohm.* "Meg" being derived from *mega millian.* Similarly a millionth of an ampere is expressed as a *micro-ampere* or 10^{-6} AMPERE. The various units under the C. G. S. and practical systems are taken up in detail under the respective headings, as *JOULE. ERG, DYNE* and so forth. (See also *Inductance (Unit of); Capacity (Unit of); Electrostatic Units and Electromagnetic Units.*)

C. E. M. F.—The abbreviation for *counter electromotive force* (q.v.).

C. G. S. UNIT OF CURRENT—The absolute *electromagnetic unit of current.* It is the current which when passed through a conductor bent into a circle of one centimeter radius will attract or repel a unit magnetic pole placed at its center, with a force of one *dyne* (q.v.).

CERUSITE—A kind of mineral used as a rectifier in some types of *crystal detectors.* It is a carbonate of lead having a whitish-grey color. This mineral is not as good a rectifier as some of the others, but has the advantage of more uniform sensitivity over practically its entire surface. (See *Crystal Detectors.*)

CHALCOPYRITE—A crystal used in a *pericon detector*, which is composed of zincite and chalcopyrite in contact with each other. It is a combination detector suitable for use where it is subject to jarring, since it has nearly uniform characteristics over its entire surface. It works without the aid of an electric battery.

CHARACTERISTICS—A term broadly applied in electricity to refer to any distinguishing feature in electrical apparatus. In radio usage the term is generally applied to *vacuum tubes.* The characteristic of a tube is understood as the relation between the potential (voltage) at the *grid* (q.v.) and the resulting current obtained from the *plate.* (q.v.) It may also occasionally refer to other features of a tube, such as its current requirements or amplifying quality as distinct from the ratio of *grid voltage* to *plate current.* (See *Characteristic Curve, Vacuum Tube Characteristics;* also *Plate Current.*)

CHARACTERISTIC CURVE—A diagram showing graphically in the form of a curved line the relation of changing values. Thus, in the case of a *vacuum tube*, a change in the *grid* voltage produces certain changes in the *plate* current (q.v.) and these changes may be plotted as curves. The temperature of a wire carrying current increases with increase in the *current* passing through it, and this increase may be shown by a curve. (See *Vacuum Tube Characteristics.*)

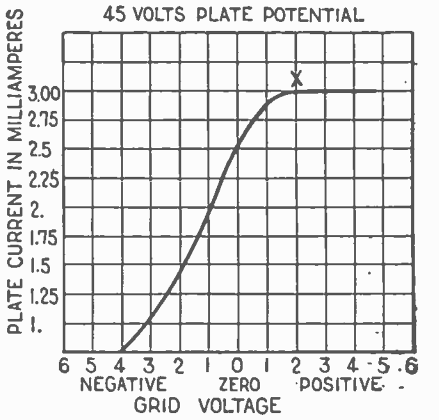

Characteristic chart of vacuum tube grid and plate voltage.

CHARGE—The presence of electricity in a body. Is usually defined as an excess or deficit of *electrons* (q.v.) on an insulated body. In the case of a *condenser* (q.v.) after it has accumulated electrical energy it is said to be in a charged condition, or to have an *electro-static charge.* The charge in a *condenser* is measured in *coulombs.* (q.v.) A *storage battery* is said to contain a charge when it has been connected to a source of direct current for a length of time, and the hydrometer test shows the electrolyte at proper specific gravity.

CHARGE, SPACE—See *Vacuum Tubes*, also *Space Charge.*

CHARGED BODY—Any body that has accumulated electrical energy. (See *Charge.*)

CHARGER, STORAGE BATTERY—Any means of controlling the current delivered from a primary source for the purpose of charging *storage batteries* or other *secondary cells.* Chargers are made in many forms and are designed for either direct or alternating current. Direct current chargers consist of certain resistances, with the necessary meters and switches. A simple form of direct current charger is shown in the illustration. In constructing a direct current charger there are several important details that must be

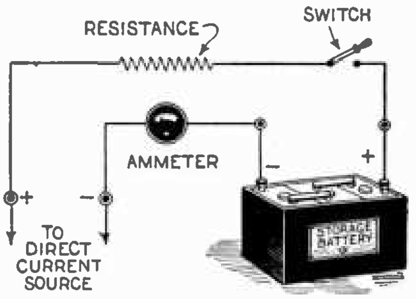

Charging storage battery from direct current source.

taken into consideration. The plates of a storage battery require certain charging rates in amperes, depending on the size and number of the plates. That is to say, the manufacturer specifies that when charging, a certain number of amperes must be passed to the battery from the charging source for a definite number of hours. This figure may be decreased but not increased. As an example, some storage batteries will take a charge of five amperes without injury to the plates, which rate would be ruinous for other batteries. Now the question of obtaining the proper amount of charging current arises. The majority of house lighting systems have a voltage of from 110 to 125 volts. (Reference is now made only to direct current systems, alternating current chargers being taken up later on.) According to *Ohm's Law* (q.v.) the current flowing in a circuit is directly proportional to the *voltage* and inversely proportional to the *resistance*. Voltage is the product of resistance times the current, the resistance in ohms and the current in amperes. Therefore if we desire a current of five amperes from a source of 110 volts, it is necessary to have a resistance of a value in ohms which, when multiplied by the current in amperes,—five in this case—will produce 110. Conversely, we can divide the voltage—110 by the current—5 to obtain the correct resistance. This would indicate a resistance of 22 ohms. There is another factor which enters into the problem however, the back pressure, or Counter Electromotive Force exerted on the charging source by the battery. In the case of a six volt storage battery, this will be six volts. This must be subtracted from the voltage of the charging source to obtain the effective value. Thus, the effective value in this case would be 110 — 6 or 104 volts. Then as explained before, dividing the voltage by the current required, we find that it is 104 ÷ 5 or approximately 21 ohms. Then by placing a resistance of 21 ohms in the circuit, as shown in the illustration, and connecting the battery as indicated, a charge of five amperes can be applied to the battery. The same result can be obtained by placing an ordinary electric lamp in series in place of the resistance. (See *Lamp Bank.*) The *internal resistance* of the battery is very small compared to that of the line resistance, and can be ignored. A point to remember is that the voltage of the charging source must always be higher than that of the storage battery, because, as explained above, the back or *counter electromotive force* of the battery, must be subtracted from the voltage of the charging source to indicate the effective voltage. Thus, if the battery exerts a counter electromotive force of six volts and the charging source is only six volts the counter force would be equal to the charging force and there would be no charging current. When *direct current* is not available and *alternating current* is used to charge the battery, it will be necessary to change it to direct current. For this purpose it is essential that some form of *rectifier* be used. This rectifier may be of a chemical, mechanical or electrical type. These various forms will be taken up under their separate headings. (See *Rectifier, Balkite, Tungar, Rectigon, Electrolytic Rectifier, Tube Rectifier.*)

CHARGING—The act of expending electrical energy to place a charge in or on a battery or body. The term is used generally to indicate the process of impressing electrical energy on the *storage battery* elements for the purpose of making the battery a secondary source of electricty. (See *Charger, Chemical Rectifier* also *Tube Rectifier.*)

CHARGING BOARD—A term used synonymously with *charging panel* to denote an arrangement for controlling the current used for charging *storage batteries.* It generally consists of an insulated panel with various *switches, fuses* and *resistances* thereon to limit the amount of current delivered, and at the same time furnish a ready means of connecting and disconnecting the various circuits connected through it. The fuses act as safety devices and break the circuit when too great a current is delivered. (See *Fuse, Switch, Charger.*)

CHEMICAL RECTIFIER—Device for changing alternating current to direct current. (*See Electrolytic Rectifier.*)

CHLORIDE BATTERY A type of storage battery for heavy duty operation. The positive plate is made of lead with a number of holes pierced in it and each filled with a coil of pure lead. This coil is forced into the holes during manufacture and is said to enable the battery to withstand heavy discharge without damage to the plates. (See *Storage Battery.*)

CHLORIDE CELL—A cell (q.v.) which uses in most cases, a positive plate having lead as the active material, the negative plate being made of metallic zinc. A solution of chloride of zinc is used, from which the name is derived. (See *Chloride Battery.*)

CHOKE OR CHOKING—The action of a device in a circuit wherein it tends to choke, or hold back, certain forms of current or certain *frequencies* while permitting others to pass freely, or to oppose fluctuations in the strength of the current passing through it in the circuit.

CHOKE COIL—A coil of wire possessing considerable *self-inductance* and relatively little *resistance.* These coils are used in various ways and for a variety of purposes. When connected in a direct current circuit, and possessing the proper value, they have a tendency to prevent any fluctuations in the current and thus keep the current smooth. An example of its use is with a *loudspeaker.* A *condenser* will permit passage of *high frequency* (*alternating*) *currents* such as we find in radio reception, but will effectively block direct currents. On the other hand a choke coil will pass direct current without offering any appreciable resistance, but will retard the *modulated currents* (q.v.) It will be apparent then that by a proper combination of the two (Choke Coil and Condenser) as in the illustration, the direct currents can be shunted around the loud speaker and thus prevent injury to it, while the necessary modulated currents are allowed to pass through without difficulty. Another use for the Choke is in a circuit carrying alternating currents. In this case the choke will limit the current in the circuit without loss, whereas an ordinary resistance, while it would limit the current, would dissipate a certain amount of energy. Chokes may be made with iron cores of the open or closed type, for *low frequency currents*, or may only have air for the core when used in *high frequency* circuits. The closed core type is the more common form. The formula most commonly used in determining the inductance of a choke coil in *henries* is as follows:

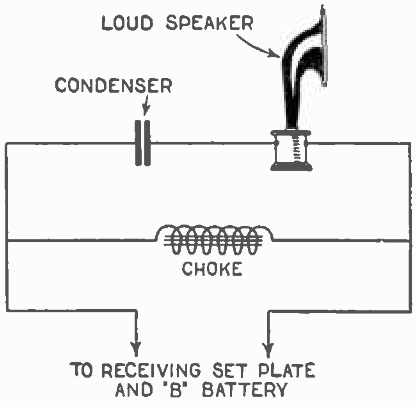

Method of using choke coil in connection with loud speaker.

$$L = \frac{1.257 \times \mu \times A \times N^2 \times 10^{-8}}{l}$$

Where L is the inductance in henries.

μ is the permeability. (This varies from about 1000 to 2000 in most cases).

A is the effective area of iron cross section (square centimeters).

N is the number of turns of wire.

l is the length of iron path (in centimeters).

CHOKE, HIGH FREQUENCY — A choke coil of high *self-inductance* (q.v.) used in transmitting circuits to prevent puncture of the secondary windings of the *transformer* (q.v.). When the condensers in a transmitting circuit are charged or discharged, a back pressure is exerted on the windings of the transformer supplying the high voltage. The illustration shows the position of these choke coils in the transmitting circuit. (See *Back Oscillation.*)

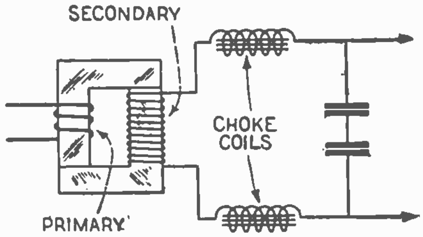

Position of choke coils in transmitting circuit.

CHOPPER—A device used in the aerial circuit of a *continuous wave* (q.v.) transmitter to break up the continuous trains of waves into separate groups and thus permit them to become audible. In the case of reception of ordinary *damped waves* (q.v.) the head phone produces signals that have a frequency of vibration equal to the number of wave trains per second. When the transmitted waves are con-

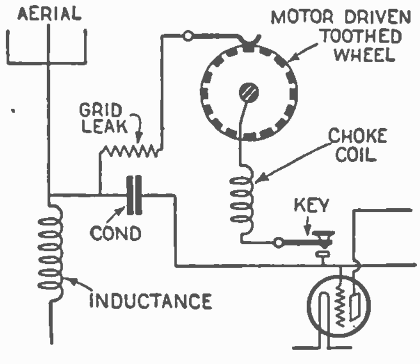

Showing how chopper is used to interrupt signals from continuous wave transmitter.

tinuous there is obviously no interval, that is, there are no groups of waves and they are therefore inaudible in the usual receiving apparatus. The *chopper* is only necessary when *crystal detectors* are used for reception. (See *Ticker, Beat Reception and Continuous Wave Transmission.*)

CIRCUIT—The path followed by an electric current passing from its source through a wire conductor or series of such *conductors*, and back again to its starting point. A *closed circuit* is one that is continuous and will permit the passage of current. An *open circuit* is one that is not continuous and, being broken or open at some point, will not permit passage of the current. The term is used very commonly, and broadly covers any system for the conducting of electric current from its source, through some electrical instrument or appliance and back to the starting point.

CIRCUIT BREAKER—Generally speaking, any device for automatically breaking or opening a circuit and thus preventing flow of current through the circuit. More specifically, it may be a device similar to a *switch*, used to automatically open the circuit when the current falls below a certain point, or rises above a certain fixed value. Thus an *overload circuit breaker* is an arrangement which will automatically open the circuit if the current becomes too great. This is shown in Fig. 1. It is a simple form and consists of an *electro magnet* which requires a certain force to operate it. When the current in the circuit reaches a certain point, the magnet operates, pulling down the bar "A" against the action of the spring "P," and thus breaking the circuit. As long as the current is not more than that fixed by the requirements of the circuit—the magnet being arranged in accordance with the particular requirements — the current flows, the breaker operating to open the circuit only when more than the desired current is introduced into the circuit. In the illustration we can consider "S1" and "S2" as the source of current, "M" as an electro-magnet in series with the motor "D," and "A-B" a switch, in which "A" has attached to it a piece of soft iron. Now if the current required for the motor is ten amperes, the magnet can be so designed that it will not act on the bar "A" until 15 amperes flow through it. Thus, as long as the current is less than this, the circuit is continuous, but if the current is increased to 15 amperes, the magnet operates and attracts the bar "A," breaking its contact with "B" and thus opening the circuit. In Fig. 2 a simple *underload breaker* is shown. In this case the operation is the reverse of that of the *overload breaker*. Here the bar "A" is held down in contact with "B" by the magnet "M" thus keeping the circuit closed through the motor "D". Now assume that the motor requires ten amperes and the magnet requires the same amount to hold the bar "A" in place against "B". If then the current in the circuit falls below ten amperes, the magnet can no longer hold the bar "A" against the action of the spring "P" and it rises, breaking the contact between "A" and "B" and thus opening the circuit. These circuits are of course only assumed for the purpose of illustrating the theory of circuit breakers. Actually the system is generally much more elaborate. Circuit breakers are used for many purposes—chiefly to prevent any sudden increase or decrease of current in a circuit, and to control the flow of current for charging batteries. (See *Charger-Storage Battery.*)

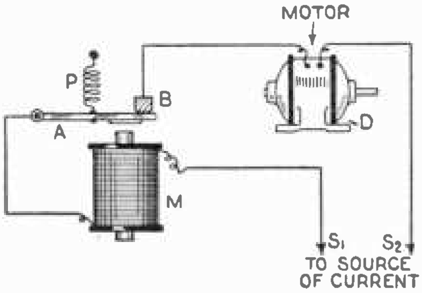

Fig. 1. Overload Circuit Breaker.

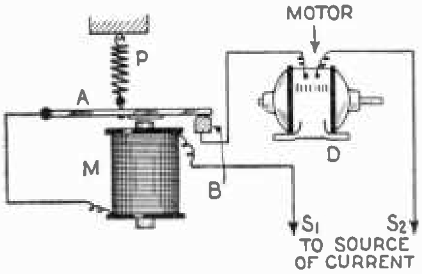

Fig. 2. Underload Circuit Breaker.

CIRCUIT, BROKEN—A circuit that is not continuous, having been opened either purposely, or due to some defect in the circuit. The term is usually applied where the circuit is opened by the breaking of a wire, or through a faulty contact.

CIRCUIT, OPEN — A circuit which is open and through which current cannot flow, or a circuit which is normally open but which can be closed at will by pressing a key or button. A common example of open circuit is in the case of an electric bell circuit. The illustration shows such a circuit. Normally it is open and current does not flow to the bell, but on pressing the button B, the circuit is closed and current passes to the bell. (See *Circuit*, also *Closed Circuit.*)

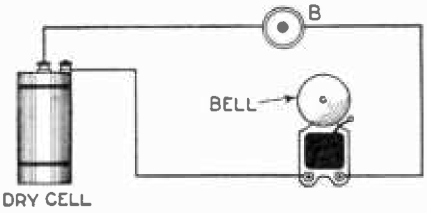

Open circuit which can be closed by pressing button "B".

CIRCUITS, PARALLEL — Circuits starting at a common point and ending at a common point. The illustration shows three *parallel circuits*. Actually each one is a part of the whole circuit. (See *Parallel Resistances.*)

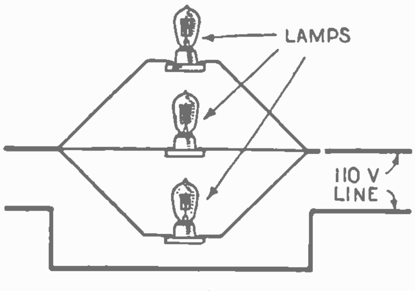

Three parallel circuits.

CIRCULAR MIL—The cross-sectional area of conductors (wires) is usually designated in terms of circular mils. A circular mil is the area of a circle having a diameter of one thousandth of an inch. Thus a wire having a diameter of one quarter of an inch is said to have an area of 250 circular mils. (See *Square Mil.*)

CLEAT—A form of porcelain *insulator* used in house wiring. They are generally arranged in two parts, with corresponding grooves in which the wire is held, and are nailed or otherwise fastened to the wall. Cleats are about the cheapest form of insulator and where the voltages used are low, are quite as effective as the more elaborate types. Cleats are much used in radio work, both for insulating lead in wires coming from the aerial, and for insulating the aerial itself.

CLERK-MAXWELL, JAMES—Born at Edinburgh, Scotland, in 1831, and educated at Edinburgh and Cambridge University, he became Professor of Natural Philosophy at Aberdeen, 1856-69, and Professor of Physics and Astronomy at King's College, London, 1860-65. In 1871 he became the first holder of the new chair of experimental physics at Cambridge, where he died, November 15, 1879. Clerk-Maxwell was recognized in his later years as one of the greatest authorities on physics of his time, and his fame has been steadily increasing since his death. Radio owes Clerk-Maxwell a deep debt. Electricity was the chief study of his lifetime, and his first important paper on the theory of *electromagnetism* was communicated to the Royal Society in 1867. In 1873 he published his "Electricity and Magnetism," a work on the subject which has never been surpassed. In it he formulated his famous *electro-magnetic theory* of light and his theories on electric waves, which developed into the modern system of wireless telegraphy and telephony through the experiments of Hertz.

James Clerk-Maxwell.

CLIMAX—A high resistance nickel steel alloy, used extensively in *rheostats* (q.v.). It is one of the cheapest and most practical alloys for low-temperature resistance wire. Its chief properties are as follows: Maximum working temperature 540 degrees centigrade, resistance in microhms per cubic centimeter at 20 degrees centigrade, 87.2; temperature coefficient per degree centigrade, .00054, and having a specific gravity of 8.14.

CLIPS—A clamp connector used for various purposes in radio receiving and transmitting circuits. (See *illustration.*) In a receiving set a clip may be used at the end of a flexible wire for the purpose of varying the connection to the various "B" battery taps; clips are very frequently used in making the connections to the storage "A" battery, and for many other purposes about the receiving set. One of their principal uses in the transmitter is in connection with the leads to the oscillation transformer. Flexible leads with clips attached come from the other units of the transmitter and are clamped at will along the various turns of the oscillation transformer until the point giving the desired wave length or highest radiation reading is found.

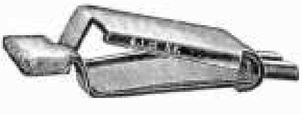

Type of clip in general use.

CLOCKWISE—A term used to indicate that the rotating part of an electrical machine or instrument moves from left to right (when facing it), or as the hands of a clock. In the illustration the needle or pointer of the meter moves in a clock-wise direction. In an electrical measuring instrument this is always true where the zero point is at the left part of the dial. The opposite is known as *counter-clockwise*. When looking at the *armature* of a motor or generator, if the rotating part moves from left to right it is said to have a clock-wise motion.

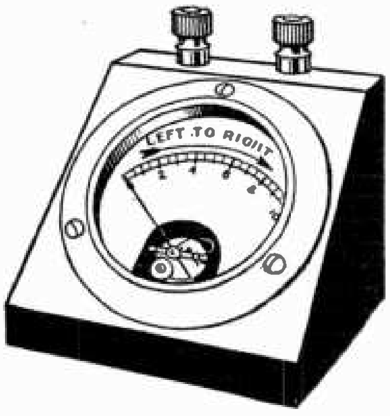

Indicator needle moves from left to right in clockwise motion.

CLOSE COUPLING The arrangement of two coils acting as *primary* and *secondary*, placed close together in such manner that the coupling or electrical relation between them is close. In the illustration the secondary coil is shown entirely within the primary. In other words the transfer of energy will be large in this case as the coupling is close. If the secondary is withdrawn a certain distance from the primary, the effect is referred to as loosening the coupling, and the coils are then said to be *loose coupled*, the transfer of energy from one to the other being considerably less than in the case of close or *tight coupling*. (See *Coupling*.)

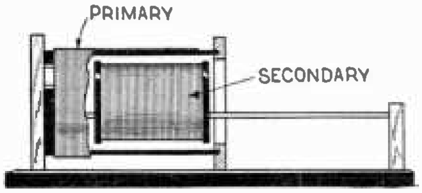

Loose coupler with secondary placed to produce close coupling.

CLOSED CIRCUIT—A circuit which permits the continuous flow of electricity. The illustration shows an ordinary dry cell connected in circuit with a small lamp. In Fig. 1A the circuit is closed and current is permitted to flow to the lamp. In Fig. 1B the circuit is open and no current can flow. A closed circuit is therefore an electrical circuit that is continuous. (See *Circuit*, also *Open Circuit*.)

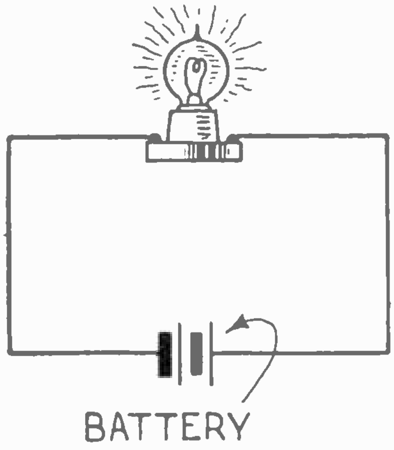

-1A-

Circuit closed.

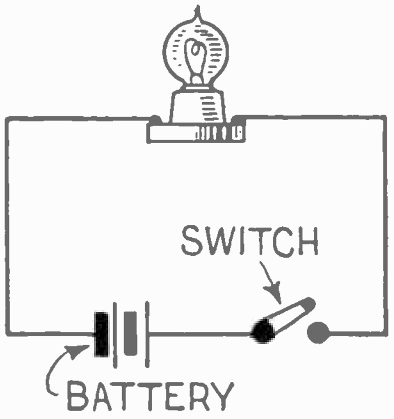

-1B-

Circuit open by means of switch.

CLOSED CIRCUIT CELL—A cell that can be used in a closed circuit for a considerable length of time without the impairing effects of *polarization* (q.v.) An *open circuit* cell is one that can only be used intermittently, as for example, with an electric bell where the circuit is only closed for a brief period. If an open circuit cell is used in a circuit that remains closed for some time, polarization sets in and ruins the cell in a very short time. An open circuit cell might be used with the *test buzzer* (q.v.) because in this case the circuit is closed intermittently as the key is depressed. (See *Cell.*)

CLOSED CORE—A core (q.v.) used in *transformers*, *chokes*, etc., which is continuous, forming a *closed magnetic path*. The illustration shows the difference between a closed core and an open core. In the open core type, one end is the north pole, the other the south pole, whereas in the closed core, there are obviously no poles. The closed core type is occasionally referred to as a *non-polar* core. (See *Core*, *Transformer*, also *Choke*.)

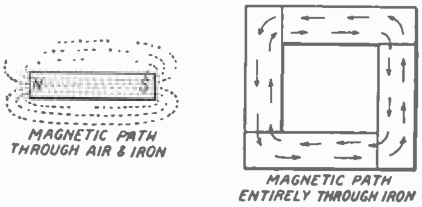
N
S
MAGNETIC PATH THROUGH AIR & IRON
MAGNETIC PATH ENTIRELY THROUGH IRON

Open core above on left, closed on right.

CLOSED CORE TRANSFORMER—A *transformer* (q.v.) having a *closed core*. In the illustration the magnetic path, or core, is continuous, being made of strips of uniform size. This form of core is known as *"laminated."* An efficient transformer is generally made with a closed core. A transformer may be for a variety of purposes, such as *step-up, step-down, audio frequency* or *radio frequency*. (See *Transformer*.)

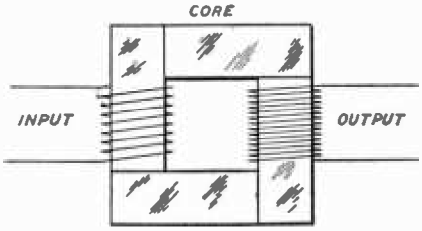

Closed core transformer.

C. M.—Abbreviation for circular mil. (q.v.)

Cm.—The abbreviation for *centimeter*, the unit of length in the C. G. S. system (q.v.).

COCKADAY CIRCUIT—A popular radio receiving circuit devised by the American experimenter, Lawrence M.

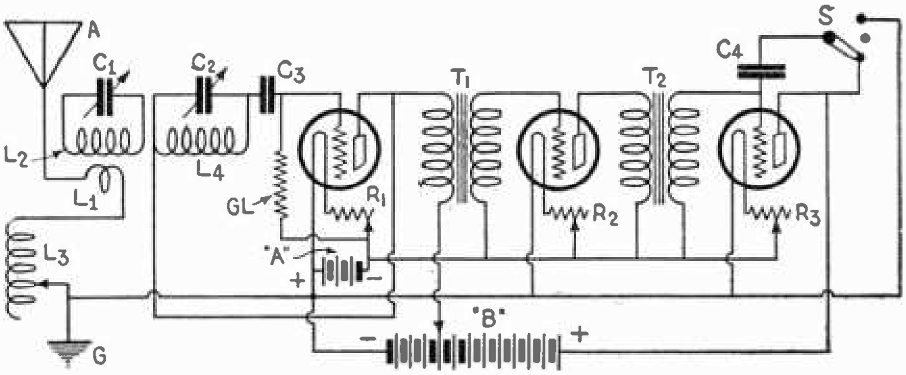

Fig. 1. Original 3 tube Cockaday Circuit.

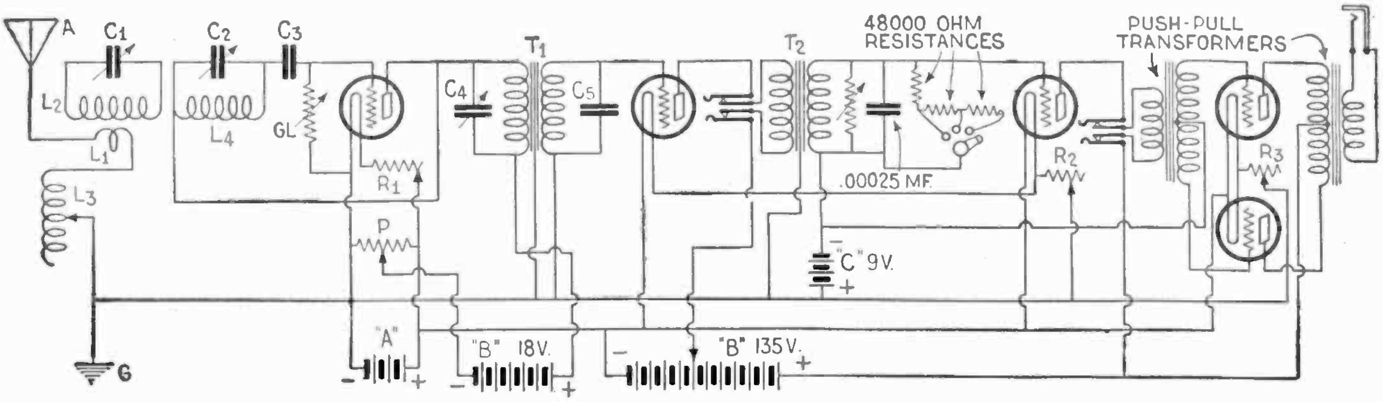

Fig. 2. Cockaday Circuit with push-pull amplifier.

Cockaday. It includes several radical features not found in the usual systems. Receivers using this circuit are generally made in three, four or five tube types. The chief value of this circuit lies in its extreme selectivity. It is usually referred to as a four circuit receiver, owing to the addition of a sensitizing circuit to the regular three circuit regenerative system. The diagram for the three tube receiver is given in Fig. 1. Here the primary or aerial tuning circuit is seen to have a single turn of wire, L1, wound around another coil L2, which is shunted by a variable condenser C1, thus forming a separate closed, tunable circuit, coupled very loosely with the antenna circuit. The primary, or antenna circuit, has in addition to the single turn, a bank wound coil L3 connected in the ground side of the open circuit and tapped to permit coarse tuning. Thus, the primary is coupled inductively to the separate closed circuit, which in turn is inductively coupled to the grid or secondary circuit containing a coil L4 shunted by the condenser C2, allowing tuning.

The tone of the music is varied by means of the novel arrangement using a condenser C4 connected to a three point switch. When the switch is placed in the center position the condenser is eliminated from the circuit and the wiring is standard. With the switch on the top tap, the grid of the tube is connected to the ground, thus by-passing a portion of the oscillations and giving a smooth soft tone with a certain loss of volume. With the switch arm on the bottom tap, the grid of the last tube is connected to the plate circuit, thus producing a different note due to the added grid-plate capacity. Coil L2 in the grid circuit is tuned by means of C1, which is a variable condenser of .0005 mfd. capacity.

A peculiarity of the circuit is that the length of the antenna has no bearing on the tuning. This is due to the *semi-aperiodic primary* and the fact that the sensitizing or pick-up coil is only one turn of wire, the inductance of which, compared to the whole inductance of the aerial circuit, is so small that any change in the over-all inductance has little, if any, effect.

This circuit, while originally having three tubes, has been arranged in conjunction with a standard *push-pull audio amplifier*, as shown in Fig. 2, making a five tube set with great power and possessing the characteristic of the original circuit with increased amplification. (See *Push-pull, Cockaday Coil* also *Regenerative Receiver.*)

COCKADAY COIL—The coil used in the *Cockaday circuits.* It comprises four windings, i.e. tapped primary, separate closed circuit coil, grid or secondary winding and a single turn looped around the closed circuit coil. (See *Cockaday Circuit.*)

CODE—Telegraph codes consist of characters formed by combinations of dots, dashes and spaces which represent letters, numerals, and punctuation marks. These characters are sent out by the radio operator with the aid of electrical impulses, and by using suitable receiving apparatus the receiving operator hears the incoming signal and is thus able to interpret the various combinations of dots and dashes and reproduce the original message. The Morse Code of characters came into general use in wire telegraphy shortly after the establishment of that means of communication. In this system, which is also called the "American Morse Code," some of the characters are made up with so-called spaces which are part of the group signal, and are necessary in distinguishing those characters. The International Morse code is a modified form of the American Morse code, and no spaces are used in the characters of the International Code. The International Morse Code is used all over the world for radio telegraphy, and for wire telegraphy in almost every country except the United States, Canada and parts of Australia. The American Morse, owing to the fact that there are fewer dashes in the characters, is about

LETTERS	MORSE	CONTINENTAL
A	●▬	●▬
B	▬●●●	▬●●●
C	●● ●	▬●▬●
D	▬●●	▬●●
E	●	●
F	●▬●	●●▬●
G	▬▬●	▬▬●
H	●●●●	●●●●
I	●●	●●
J	▬●▬●	●▬▬▬
K	▬●▬	▬●▬
L	▬▬	●▬●●
M	▬▬	▬▬
N	▬●	▬●
O	● ●	▬▬▬
P	●●●●●	●▬▬●
Q	●●▬●	▬▬●▬
R	● ●●	●▬●
S	●●●	●●●
T	▬	▬
U	●●▬	●●▬
V	●●●▬	●●●▬
W	●▬▬	●▬▬
X	●▬●●	▬●●▬
Y	●● ●●	▬●▬▬
Z	●●● ●	▬▬●●
&	● ●●●	
1	●▬▬●	●▬▬▬▬
2	●●▬●●	●●▬▬▬
3	●●●▬●	●●●▬▬
4	●●●●▬	●●●●▬
5	▬▬▬	●●●●●
6	●●●●●●	▬●●●●
7	▬▬●●	▬▬●●●
8	▬●●●●	▬▬▬●●
9	▬●●▬	▬▬▬▬●
0	▬▬▬	▬▬▬▬▬
.	●●▬▬●●	●● ●● ●●
,	●▬●▬	●▬●▬●▬
:	▬●▬●●	▬▬▬●●●
;	●●● ●●	▬●▬●▬●
?	▬●●▬●	●●▬▬●●

Morse and Continental Codes.

5% more rapid than the International Morse. The American Morse and International (Continental) Codes are given herewith.

COEFFICIENT—In algebra, a number or multiplier of a *symbol*. Thus in the expression 5Y, the coefficient is 5. In electricity, and in its radio application, it is a number expressing the relation or ratio between quantities, one of the quantities generally being unity (1) as a basis or standard for the coefficients. For example, let us assume a copper wire is carrying an electric current, the *resistance* of the wire at 32 degrees Fahrenheit being one *ohm*. If we consider the resistance of the wire at 32 degrees to be unity, or one, then its resistance at any other temperature will be some number or coefficient times one.

To make this clearer, let us take the ratio of an inch to a foot. An inch is one-twelfth, or approximately .0833, of a foot. Any number of inches multiplied by this fraction will give another number in feet. Therefore we say that the coefficient, or multiplier, of an inch as compared to a foot is .0833. A coefficient is thus any number by which a quantity must be multiplied to give another quantity.

In radio the most common cases of coefficients are the following. The *coefficient of coupling* is the ratio between the *mutual inductance* (q.v.) (between the two coils, *primary* and *secondary*) and the square root of the product of the total *self-inductances* (q.v.) in the two circuits containing the coils. Similarly, the coefficient of amplification is the ratio of the effect produced with an *amplifier* (q.v.) to the effect produced without an amplifier. (See *Mutual Induction Coefficient*, also *Self-induction Coefficient*.)

COHEN, LOUIS, Ph. D.—Born in Russia, Dec. 16, 1876. Educated at Armour Institute of Technology, University of Chicago and Columbia University. Noted for his formulae and tables for the calculation of alternating current problems. He has written a book on the subject. Has evolved many formulas for the measurement of inductance and capacity. Has secured a number of patents on radio apparatus.

COHERER—A device formerly employed for detecting *radio waves*. In its original form it consisted of two metal rods separated by a narrow space in a glass tube and the gap between the rods filled with nickel and silver filings. It was found that when the *electromagnetic waves* or signals were passed through the coherer the particles of filings clung together and became *conductors*. Then by using some device such as the tapper hammer of an electric bell to strike the glass tube, again separating the particles after each signal, it was possible to keep the coherer constantly in a state of change between a *conductive* and a *non-conductive* condition, thus making the signals audible. There were many variations of this type of *detector*, which, however, is now obsolete. (See *Marconi Coherer*.)

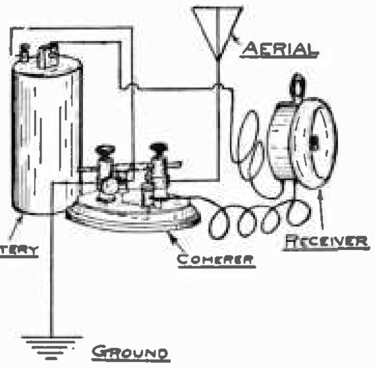

Coherer used in radio receiving circuit.

COIL—A term having very broad application in electricity and radio. It may be used to designate any arrangement of a number of turns of wire, usually copper, and used for almost any purpose in the production of electric and magnetic effects or phenomena. The most common significance of a "coil" in radio is in reference to a number of turns of insulated wire wound on various forms and used for *tuning* (q.v.). The term *inductance coil* is generally used, although actually, any coil of wire will have inductance. Coils come in countless numbers and variety for many purposes. A coil may be used not alone for tuning—that is to adjust a receiver or transmitter to resonance with incoming or outgoing signals; it may be used to connect two circuits, in which case it is known as a *coupling coil*; it may be used with a *core* (q.v.) of iron wire, or laminated pieces, as a *choke coil* (q.v.) or it may be used in connection with *high frequency* apparatus, such as used by physicians. Coils are generally referred to by individual names indicating their construction or use. (See *Inductance*, *Tuning Coil*, *Vario-coupler*, *Choke Coil*, *Honey Comb Coil*, *Pancake Coil*, *Spider Web Coil*, *Lorenz Coil*, *Duo Lateral Coils*.)

COIL ANTENNA—Or Loop Aerial. An antenna (aerial) constructed by winding the wire around a square form. The coil (or loop) antenna, because of its comparatively small size, is often mounted on a frame work which may be rotated. If one end of the coil is pointed toward the broadcasting station, the antenna (aerial) will pick up its signals with maximum intensity, but if the coil is rotated from that position the signals from that station grow weaker, until when the plane of the loop is at right angles with the intercepted signals no sound will be obtained. It is a part of the radio compass used at sea. (See *Loop-Aerial*.)

COLLECTOR RINGS—Insulated metal rings attached to the *armature* (q.v.) of an *alternator* (generator of alternating current) in such fashion that the alternating currents can be collected and communicated to the *brushes* without change, these rings are used to collect alternating current from that type of generator and brushes are used to collect direct current from a direct current generator.

COLLODION—A solution used by surgeons to cover slight abrasions or wounds, forming a protective coating similar to skin. It has been adapted to radio use as an insulating coating and is used as a binder on coils of various forms. It has been found more efficient than shellac for this purpose. A solution of pyroxylin or soluble guncotton in ether, with a certain amount of alcohol, depending on the purpose for which intended. When used as a coating for coils it serves as an insulator and also to hold the coils in place, but it should be sparingly applied.

COMBINATION DETECTOR—A detector composed of two different crystals in contact, instead of one crystal with a fine wire to search out the sensitive spots. Some combination detectors are composed of silicon and antimony; zincite and bornite; lenzite and cerusite, and zincite and chalcopyrite. These detectors all function without a battery and are simple to operate.

COMMON CONNECTIONS—A term often applied to a connection that joins several points. For example, in most radio receivers the binding posts for battery connections are arranged so that the "B—" and the "A+" or "A—" are connected to the same point. That is to say the post is a common connection for these leads. The term is also applied to a battery, as a *"B" battery*, which is used for both *detector* and *amplifier*. It is then known as a common "B" battery. The other method would, of course, be to use separate batteries for the detector and amplifier.

COMMUTATOR—In general electrical use, a device for reversing the direction of electric currents in a circuit. A *commutator* as used with a *generator* usually consists of a number of pieces called *segments*, mounted on a circular form but carefully insulated from each other. The illustration shows a common form. In the case of an ordinary motor, the commutator is employed to distribute the current to the windings and in a direct current generator it serves to furnish a direct current to the brushes although the current collected from the windings of the *armature* is alternating.

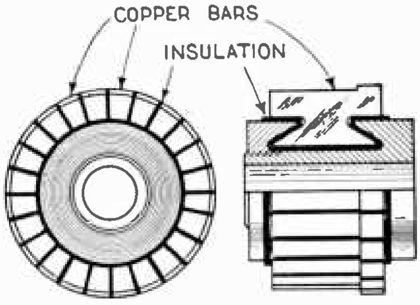

Front and side view of Commutator.

COMPASS, RADIO — See *Direction Finder*, *Goniometer*, also *Bellin-Tosi Direction Finder*.

COMPONENT—A part of the whole. In mechanics it is one of the parts of a stress or strain (mechanical force). In electricity, *component currents* are the several currents into which a single current may be assumed to be separated, so that if assumed to be acting together they would produce the equivalent effect of the actual current. Thus also, component *E. M. F.'s* (*electro-motive forces*) are understood as the several components into which any electro motive force may be divided. *Impedance* (q.v.) is said to consist of two components, the actual resistance and the apparent resistance present in the opposition offered to the flow of current termed the *reactance* (q.v.) by a circuit. For instance, the total current in the plate circuit of a vacuum tube responding to high *frequency grid voltages* may be considered as made up of two components, viz., a steady or constant current and a high frequency component superimposed on it.

COMPRESSED AIR CONDENSER—A condenser which uses compressed air as the *dielectric* or insulating material between the plates. This type of condenser is used chiefly in transmitting circuits. (See *Condenser*.)

COMPOUND MAGNETS—A combination of several *permanent magnets*. (See *Magnets, Compound.*)

COMPOUND WOUND—A *dynamo* or *motor field magnet* wound with two *field windings*, one of which is connected in series, the other in parallel with the *armature* (q.v.).

CONDENSER—In electricity or radio practice a *condenser* consists of two or more conducting surfaces placed in relation to each other and separated by an insulating medium such as air, impregnated paper, mica, etc. The conducting surfaces are generally referred to as *plates* or *electrodes* (q.v.) and the insulating material is known as the *dielectric* (q.v.). Condensers are used in radio for a variety of purposes, their broad use being to allow tuning. In action, a condenser permits charges of electricity known as *electrostatic charges* (q.v.) to be stored up and released periodically, according to the frequency desired. They may also be used to exclude certain types of energy. (See *Condenser, By Pass,* also *Filter.*) The action of a condenser may be compared to that of a spring when placed under a mechanical strain. In the illustration Fig. 1A a flat spring S is shown in a normal position, NB. In 1B a weight W is placed on the surface of the spring. This weight will force the spring down from its normal position to one of the positions (a), (b), (c) or (d), according to the force exerted and the strength of the spring. If a heavier weight is then placed on the spring it will depress still further until we can imagine a point P in Fig. 1C at which the elasticity of the spring is used up and the opposition is such that it will no longer depress and comes to rest. When the weight is removed the spring will fly back into normal position, Fig. 1D, and in doing so will vibrate back and forth a number of times, thus producing, or rather, returning the energy that was stored in it during the moment of stress.

Considered as analagous to a *condenser*, in the above case the weight can be likened to the electrical charge applied to the condenser, the *resistance* of the spring may be considered somewhat similar to the *capacity* of the condenser, or the power returned on release of the spring may be regarded as the electrical power released by the condenser when the charging force is removed. The capacity of the condenser then will be seen as the ability to store up energy placed in it.

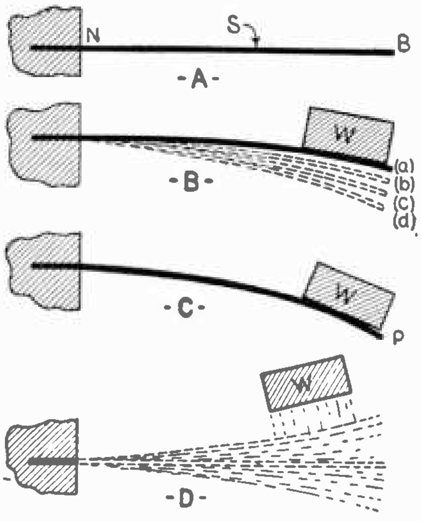

Spring analogy of a condenser.

In the case of the spring, if too great a force is exerted, it may break down or as would actually happen, the elasticity would be lost. The same applies in the case of a condenser. If the charging force is too great, it may exceed the dielectric strength of the condenser and result in puncturing the insulating material between the plates. (See *Breaking Down of Dielectric.*) The energy obtainable from a charged condenser will be equal to the charge placed in it providing there is no leakage or other loss. In a well-designed condenser the leakage will be so negligible that it need not be taken into consideration.

The relation between the charge in a *condenser*, the *voltage* applied and the *capacity of the condenser* is expressed as Q = C × E, where Q is the quantity of the charge in *coulombs*, E the potential in *volts* and C the capacity in *farads*. (See *Capacity, Condenser, Variable; Condenser Fixed; Condenser Transmitting,* also *Condenser Curves.*)

CONDENSER, AIR—See *Air Condenser.*

CONDENSER, ANTENNA—See *Aerial Tuning Condenser.*

CONDENSER, BY-PASS—A condenser, usually of the fixed (q.v.) type arranged in a circuit in such manner that currents of a certain frequency will pass freely around any obstructing object, as, for instance, a high resistance element necessary to the circuit. An example of the use of a condenser in this manner is shown in the case of a "B" battery where the battery has a high internal resistance. A fixed condenser is placed across the terminals of the battery without injury to it and offers a path of low resistance for the high frequency radio currents.

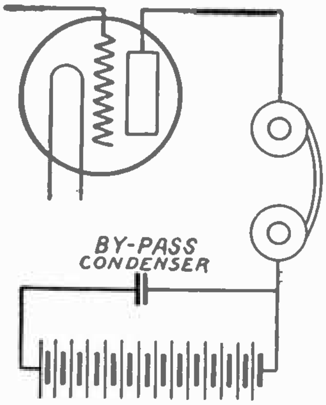

By-pass condenser across "B" battery.

CONDENSER CURVES—Curves showing the relation of variations in wavelength, frequency, and capacity with rotation of the dial of a variable condenser. Three types of condensers are shown here, the *compression*, Fig. 1, the *straight line capacity*, Fig. 2, and the *square law*, Fig. 3. In Fig. 1A is shown the relation of capacity and wavelength changes when the dial of a sample compression condenser was rotated. The condenser was shunted across a fixed coil. It will appear that the increase in capacity is gradual at the lower part of the range but increases very rapidly at the upper ranges. This is due mainly to the relation of the curvature of the upper plate to the lower plate as will be seen from the illustration, Fig. 1. These condensers are very satisfactory on about the lower two-thirds of their range, but through the remainder of their range a very slight adjustment of the knob causes an extremely large change in capacity, making them very difficult to handle near their maximum capacity. Fig. 2 shows the shape of the variable plates in the straight line capacity condenser and Fig. 2A illustrates the relation of wave-length and capacity changes with rotation of the dial. Here the capacity depends upon the overlapping area and is therefore proportional to the angle of rotation of the movable plates. The capacity curve appears as a practically straight line running diagonally across the chart. The wave-length, however, changes rapidly at the lower range and very gradually at the upper range as indicated by the wave-length curve. Fig.

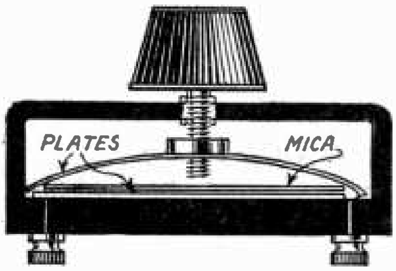

Fig. 1. Compression type condenser.

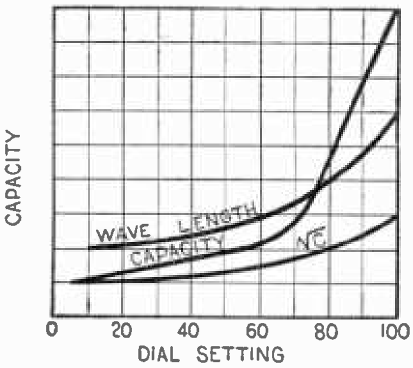

Fig. 1A. Curve for compression type condenser.

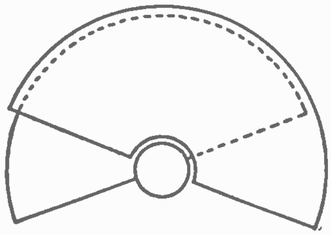

Fig. 2. Straight line condenser.

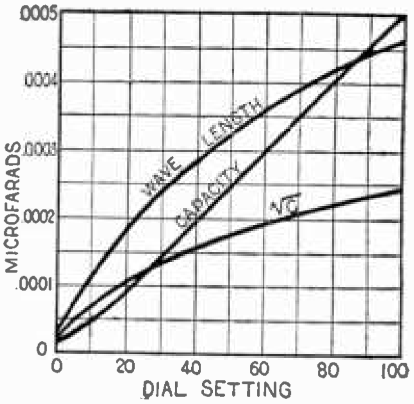

Fig. 2A. Curve for straight line condenser.

3 shows the shape and disposition of rotor and stator plates of a condenser so designed as to read directly in wavelength from the settings of the dial.

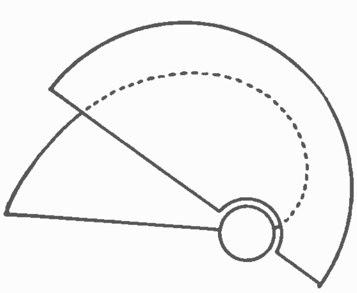

Fig. 3. Square law condenser.

This is known as a square law condenser, and the relational curves are shown in Fig. 3A. Here we find the wave-length curve a straight line and directly proportional to the angle through which the dial is turned. The line indicating the square root of the capacity √C is a straight line, and as the wavelength is the product of the square root of the capacity multiplied by a constant (varying with the individual condenser) the wavelength curve must also be a straight line.

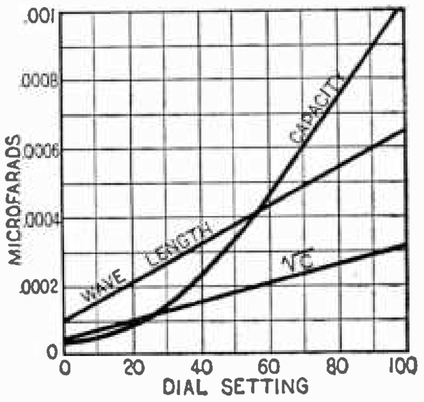

Fig. 3A. Curve for square law condenser.

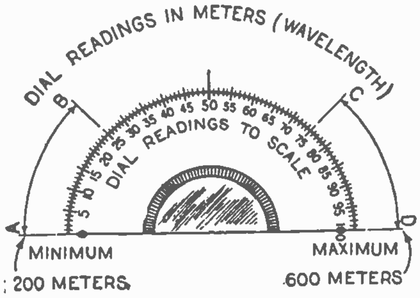

Fig. 3B. Showing effect of straight line condenser on dial reading.

Such condensers are much used in wave meters (q.v.). Condensers are also made so that the frequency curve is a nearly straight line, varying more or less directly as the angle of rotation of the plates. Fig. 3B illustrates the result when an ordinary straight line capacity condenser is arranged with a dial calibrated directly in meters. It appears that the readings are bunched at the lower range (A to B) and spread out at the upper ranges (C to D). The reason for this can readily be seen by glancing at the curve Fig. 2A.

CONDENSER, FIXED—A *condenser* which has a fixed or constant capacity, allowing no variation of the value. (See *Condenser*, also *Condenser, By-Pass.*)

CONDENSER, GRID—A condenser, generally of low capacity, of the fixed or variable type, used in series with the grid of the *detector tube* to prevent excessive charges from affecting the *grid*. These condensers generally have a capacity between .00025 and .0005 *micro farads*. Due to the blocking action of these condensers, a high resistance, known as a *grid leak* (q.v.) is placed across the condenser or from the grid to filament, thus allowing the accumulated charges to leak off the grid of the tube in time for the grid to be free for the succeeding electric charge. The grid leak may be considered as a valve controlling the amount of energy that the detector can efficiently take care of without overload. (See *Grid Leak.*)

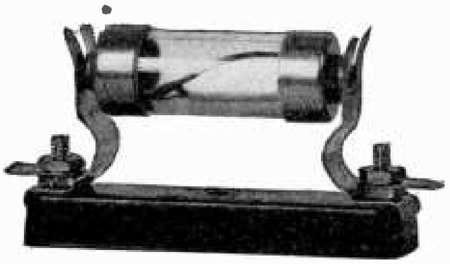

Grid leak and condenser.

CONDENSER, IN PARALLEL—A condenser arranged in a circuit in such manner that the current is divided between it and the other condenser or coil with which it is connected. The illustration shows a *variable condenser* connected in parallel with a coil, both placed in series between aerial and ground and forming a tuned primary or aerial circuit. (See *Parallel Connections.*)

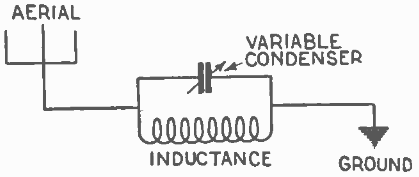

Coil and condenser in parallel.

CONDENSER, IN SERIES—A condenser connected in a circuit with other condensers or coils in such manner as to form a continuous path, allowing the current to pass through each unit without dividing. The illustration shows a condenser connected in series with a *tuning coil*, aerial and ground, thus forming a tuned primary or aerial circuit. (See *Condenser, in Parallel*, also *Series Connections.*)

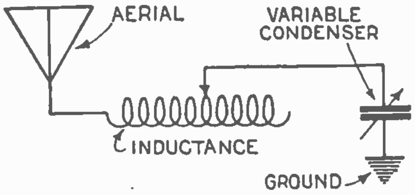

Coil and condenser in series.

CONDENSER, MICA—A condenser, usually of the fixed type, employing sheets of mica as the dielectric or insulating material between the plates. The condensers of small capacity used in receiving circuits are generally made with mica as a dielectric, but owing to its cost and thickness it is seldom used in low-power, high-capacity condensers. Mica is also used in condensers for high-power work such as transmitting, when the capacity is low. (See *Condenser.*)

CONDENSER, PHONE—A fixed condenser of comparatively low capacity used across the *telephone receivers* to allow a by-pass for a certain portion of the currents and thus prevent undue strain on the phones. (See *Condenser, By-Pass.*)

CONDENSER, TRANSMITTING—See *Condenser, Mica.*

CONDENSER, VARIABLE—Any condenser so arranged as to allow variation of its capacity. This may be accomplished in many ways, the usual method being to have the plates divided so that one series is movable with respect to the others, the capacity changing with the rotation of the variable plates. (See *Condenser.*)

CONDUCTANCE—Sometimes called the conductivity. The quality of a given conductor in virtue of which it facilitates the flow of an electric current. A *conductor* (q.v.) is a body having large *conductance* and an *insulator* (q.v.) is one having negligible conductance. *Conductance* is thus the opposite of *resistance*. The unit of conductance is the *mho*, the reciprocal of *ohm*. The symbol is obviously derived from ohm, reversed. It is the conductance of a column of mercury 106.3 cm. (centimeters) long, 1 square mm. (milimeter) in cross-sectional area and of a weight of 14.4521 grams.

CONDUCTIVE COUPLING—The association of one circuit with another by means of inductances (coils) mutual to both circuits.

CONDUCTIVITY, SPECIFIC—The reciprocal of Specific Resistance; being a standard of reference for comparing the conductances of different substances. (See *Specific Resistance, Conductance.*)

CONDUCTOR—Any substance that offers small *resistance* to the passage of electric currents may be termed a *conductor.* This term is the opposite of *insulator*, but there is actually no line of demarcation between the two, since any insulator will permit the passage of current when sufficiently high voltages are used. For all practical purposes, however, a conductor is considered as material that presents little difficulty to the passage of current. The most common conductor is copper, usually in the form of wire. An idea of the relative conductivity (power of carrying electric current possessed by various substances, pure copper being taken as standard) of various substances can be gained from the table following. The materials on the left are comparatively good conductors given in the order of their conductivity, while those on the right are poor conductors and generally used for resistances. (Note: A *resistance* is used to retard the flow of electricity without necessarily stopping it entirely, whereas an insulator is used to stop the passage of current.)

Silver
Copper
Brass
Gold
Aluminum
Zinc
Platinum
Iron
Nickel
Tin

German Silver
Platinum Silver
Manganin
Mercury
Graphite or carbon

There are, of course, many other substances that can be classed as good conductors, or as poor conductors, but the above are more generally known and used. It is a peculiar fact that while good conductors become less conductive as their temperature is raised, insulators lose a certain amount of their insulating properties as their temperature is raised. As too great current in a conductor will cause heat, and as explained above, cause lessening of conducting ability, it will be obvious that for the fullest efficiency a conductor will necessarily be arranged to carry the required current without undue heating. It is understood also that conductors in any part of a radio circuit should be carefully gauged as to size and the effective cross-section of the conductor determined, in order to offset the losses due to heating. It must not be assumed from the foregoing that the larger the wire the greater the efficiency—in fact a large wire is often less satisfactory than a small one in high frequency circuits.

When the conductor is to carry direct current the only consideration is to have the wire large enough to prevent heat losses and at the same time not be too bulky. Where the conductor is to carry *high-frequency* (q.v.) currents there are other factors to consider. High frequency currents are said to travel only on the outside of the wire (see *Skin Effect*) and therefore it is possible to use very fine wire for this purpose when the very low current values obtaining in reception are in question. A phenomenon that must be given attention, however, is that of *distributed capacity* (q.v.). (For more complete explanation of this subject see *Low-loss Coils*. See also *Bus bar.*)

CONNECTING UP or HOOKING UP—These general terms are used in radio to refer to the process of wiring a receiver or any particular part of a receiver or other circuit.

CONSTANT—A quantity or magnitude, which does not vary, derived from actual experiment, which is included as a factor in most formulas for the purpose of bringing theoretical calculations into agreement with experience. Constants are much used in radio calculations. As an illustration, different dielectric or insulating materials used in a condenser have different constants which must be taken into consideration when calculating the capacity of the *condenser*. Thus, as a dielectric, air at ordinary pressure is taken as the standard, the constant being stated as 1. Flint glass has a high dielectric value, its constant being stated as 10.10, other materials varying according to their comparison with air as a standard. (See *Dielectric.*)

CONSTANT CURRENT—An unvarying *current*. The term "constant" is applied to *voltage* when it is steady and unchanging, to *resistance* when it is fixed and unvarying, or to almost any value that does not change.

CONSUMPTION, CURRENT — The term usually employed to denote the current required or consumed by any part of an electrical circuit or by the circuit itself. A resistance in a circuit consumes a certain amount of energy because it dissipates it as heat energy. A filament in the case of a vacuum tube, requires a certain amount of current to heat it to brilliancy. If the filament thus requires one-quarter of an ampere we say its current consumption is one-quarter ampere. (See *Plate Consumption.*)

CONSUMPTION, PLATE—The current required to energize the plate of a vacuum tube. (See *Plate Consumption.*)

CONTACT BREAKER—A contrivance for quickly and automatically making or breaking an electrical circuit normally existing between two contacts. (See *Circuit Breaker.*)

CONTACT DETECTOR—A form of crystal detector in which a small wire, or "catwhisker," as it is called, makes contact with the crystal in the detector and acts as a *rectifier* of the *high frequency waves* received from the broadcasting station. (See *Crystal Detector—Theory of Operation.*)

CONTACT POINTS, or SWITCH POINTS—Generally refers to a small flat-headed machine screw used as a contact point for a *switch*. The illustration shows the types mostly used.

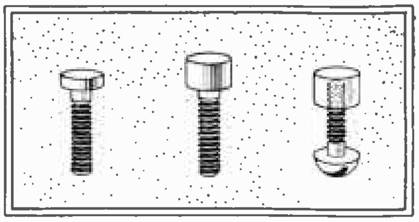

Various contact points.

Leads from various points of a tapped coil may be connected to the several contacts in order to permit variation of the *inductance*. A "contact point" may also mean any point in an electrical circuit at which contact is made to close the circuit.

CONTACT RECTIFIER—See *Contact Detector.*

CONTACT RESISTANCE—When two connecting surfaces in a circuit carrying electrical energy do not make absolute or perfect contact, a certain amount of *resistance* (q.v.) results. In any part of a radio circuit where two wires or *conductors* are joined to form a circuit, it is essential that they make a positive contact. The wires used for connections in a radio receiver are generally soldered or fastened tightly by nuts or bolts, or machine screws in order to form a good contact. If this is not done, energy may be wasted, due to the resistance at the point of contact.

CONTINUOUS OSCILLATIONS—Each oscillation or vibration being of the same amplitude as the preceding one. (See *Sine wave.*)

C. W.—Abbreviation for continuous wave (q.v.).

CONTINUOUS DIRECT CURRENT—A *direct current* which flows steadily without interruption or reversal. The current supplied from a battery for radio use is continuous direct current, while that portion of the *high frequency alternating current* which is rectified by a *vacuum tube* detector is called *pulsating direct current*, as there is an interval of time between the rectified half of each incoming wave, due to the other half not being rectified. (See *Current, Pulsating Direct, Pulsating Current,* also *Rectification.*)

CONTINUOUS WAVES—Also called *Undamped* or *Sustained waves* and usually referred to as *C. W. Electromagnetic waves*, the separate oscillations (each vibration) of which are of constant or unvarying amplitude. In Fig. A a series of *oscillations* (q.v.) are shown. It will appear that one oscillation or *cycle* is of the same amplitude as the preceding one—the amplitude (q.v.) being considered as the height of the wave above the straight line representing time or its depth below that line. Each loop above the line or below it represents an *alteration* (q.v.) (the increase in current to maximum and back to minimum in either direction)—the loops above the line being considered as *positive alternations* and those below the line as *negative alternations*. Now it is evident from the sketch that each succeeding alternation has the same strength (amplitude) as that preceding it, and the same is true of the complete cycle, consisting of one alternation in each direction. Thus the waves are continuous, or constant, as far as amplitude is concerned. Now in Fig. B the amplitude is not constant—it decreases with each succeeding alternation and hence with each succeeding cycle (two alternations). After a comparatively few oscillations the current has decreased to zero, that is, the oscillations have died out completely. In this case the percentage of decrease in amplitude from one cycle to another is known as *decrement* or *damping*. Thus continuous waves may be considered as waves in which, with any train or group of oscillations, each succeeding oscillation has the same amplitude as that preceding it.

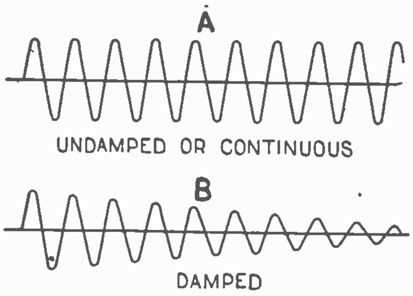

Series of undamped and damped waves.

CONTINUOUS WAVES—KEY MODULATED—Continuous waves (q.v.) which have been broken into dots and dashes of the telegraphic code by means of a *telegraphic key*, these signals being made audible at the receiving end by the use of an oscillating receiver.

CONTINUOUS WAVES—MODULATED AT AUDIO FREQUENCY—The process of variation of the frequency (vibration) of waves radiated by a broadcasting station in accordance with the sound waves is known as modulation. Frequencies from 10,000 down to about 32 are audible to the human ear and are known as audio frequencies. Hence continuous waves which are modulated at *audio frequencies* are waves which have been modulated (varied) by having the voice or music impressed on the *microphone* at the broadcasting station, which in turn causes the strength of the radiated electric wave to vary so as to correspond electrically to the sound vibrations at the microphone.

CONTINUOUS WAVE SIGNALS—See *Continuous Waves.*

CONTINUOUS WAVE TRANSMISSION—System of radio telegraphy using waves as described under "Continuous Waves."

CONVECTION CURRENTS—When a hot and a cold body are separated by a fluid which is free to circulate, heat is transferred from the hot to the cold body by currents of the fluid itself flowing from one to the other; similarly, all parts of a fluid which is being heated quickly come to approximately the same temperature. This transfer of heat by currents of the fluid itself is called "convection." In electricity "Convection Currents" are currents arising from the motion of charged particles thrown off in electrified streams, i.e., the flowing of charged air particles in streams from the pointed end of a highly electrified conductor, sometimes called "electric wind."

CONVENTIONAL—This term is much used in radio to denote a form established by custom. A conventional circuit for some particular purpose is understood as a circuit which has been generally used and which is generally considered as standard. Broadly, the most common form or use of anything. The conventional method of amplifying signals of *audio frequency* is by means of *audio frequency transformers*. Any other means is unconventional, although it may produce similar results.

CONVERSION FACTORS—A numerical factor which gives relation between the magnitude of a quantity expressed in terms of one system of units, and the same quantity expressed in terms of any other system of units. For example, the yard is regarded in the United States as 3600/3937 meter (a meter is approximately 39.37 inches). We say, therefore, that the conversion factor between the yard and the meter is 3600/3937 because any number of yards multiplied by this figure or factor will give the result in meters. (See *Capacity Conversion Factors.*)

CONVERTER—A term applied generally to any mechanical, electrical or electro-chemical device which will convert or change one form of energy to another. More specifically in electricity the term is used in reference to (1) a dynamotor designed with two *armatures*, mounted on one shaft, one being a motor armature driven by the original current, the other armature generates new current; this is also called a "motor-dynamo" and it can transform continuous (direct) current to higher or lower voltages, or, (2) a rotary electrical device used to change direct to alternating current or vice versa. (See *Rotary Converter.*)

COORDINATES—A thing of the same rank with another thing may be termed a coordinate. For example, lines by means of which the position of any point as of a curve may be defined with respect to certain fixed lines. In the illustration Fig. 1 lines known as axes of coordinates are shown. Such a system as this will be the basis for curves (q.v.) showing the relation between varying values. If we assume the horizontal line X as the axis of abscissas, all distances measured parallel to this line will be known as abscissas; distances measured parallel to the vertical line Y being referred to as ordinates. The line A-B is parallel to the axis of abscissas X and is therefore an abscissa. The line B-C is parallel to the axis of ordinates and is referred to as an ordinate. Now if we place a number of lines parallel to each axis (ordinates and abscissas) and assign to each of them a certain value, such as amperes, to each abscissa, and volts to each ordinate, a line can be drawn through a number of points to represent the various values of each quantity as related to the other. Thus, in Fig. 2 we have a purely hypothetical curve representing the assumed value of current in amperes under assumed voltages. In the example, the curve at point A shows approximately three amperes under a pressure of ten volts. This imaginary curve is, of course, shown merely for the purpose of illustrating the meaning of coordinates, ordinates, abscissa, etc. (See *Abscissa, Characteristic curve*, also *Graph.*)

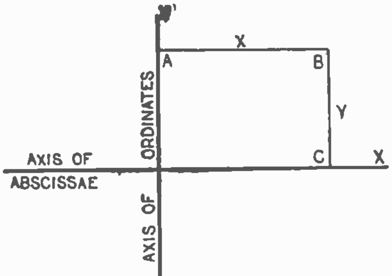

Fig. 1. Axis of coordinates.

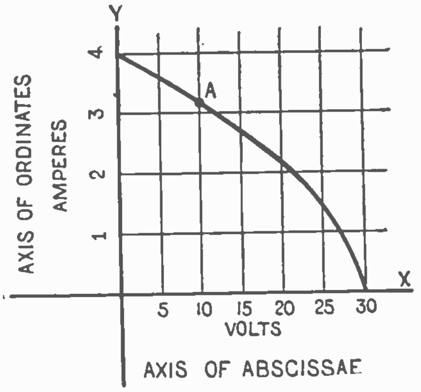

Fig. 2. Curve showing assumed value of current in amperes under assumed voltages.

COPPER PYRITES—Sometimes referred to as *Chalcopyrite*. It is used in radio reception as one of the elements of a *crystal rectifier* (detector) generally in conjunction with zincite or tellurium.

COPPER SULPHATE—A salt of copper. Also known as blue vitriol or copperas. A copper sulphate solution is sometimes used as a preservative for wooden aerial masts. It is used in certain forms of primary cells (devices for transforming chemical action into electric current) and also produces sulphuric acid when subjected to *electrolysis* (q.v.).

CORD, PHONE—The cord or flexible wire used to connect a *headset* (phone) or *loudspeaker* to the terminals of a receiver. Phone cord is usually copper tinsel wrapped around a cotton cord, the whole covered by braided silk or cotton threads. It cannot be used to carry high currents, but is efficient for the comparatively low currents flowing through the phone circuit of a radio receiver. Some phone cords are also formed of a number of fine copper wires covered with a suitable insulating material.

CORE—The central portion of a piece of certain electrical apparatus. Thus the center of a coil will be the *core*, irrespective of its nature, whether air or iron. The core of a *choke coil* (q.v.) is usually made of a bundle of soft iron wires or strips of iron. The core of the closed type is called a *laminated core*, being made up of laminations or layers of the core metal. (See *Choke or Choking Transformer, Closed Core.*)

CORE LOSSES—The losses due to the *stray currents* (q.v.) set up in an iron core while under the stress (strain) of a *magnetic field* (q.v.). Also losses due to *hysteresis* and *eddy-currents*. (See *Magnetic flux, Eddy current*, also *Hysteresis.*)

CORE TRANSFORMER (Iron)—A transformer having the wire wound over an *iron core*, the iron forming a path for the *magnetic lines of force*, as distinguished from a *transformer* in which there is no iron core—called *air core*. (See *Transformer.*)

CORONA—The Latin word meaning crown, originally applied to the crown shaped lights occurring in the *aurora borealis*. In electrical parlance it refers to the luminous discharge which occurs along high tension transmission lines or the aerial wires of a high power radio transmitting station. (See *Brush Discharge;* also *Corona Loss.*)

CORONA LOSS—The energy loss due to corona discharge. (See *Corona;* also *Brush Discharge.*)

CORRECTION FACTORS—A formula for the measurement of certain values is generally made by using certain conditions or types of apparatus as a standard. For example a formula for calculating the inductance of a coil is made by using a particular type and shape of coil as the standard. Therefore the formula will be essentially correct for this type of coil, assuming that its shape and other characteristics are the same. If, however, the nature of the coil to be tested differs from that of the standard coil in some way not covered by the formula, a system of correction must be used to adjust the formula to the particular case. Thus, the formula for inductance of coils takes into consideration the size of the wire or the number of turns, the length of the coil, its diameter and so on. However, it does not take into consideration the ratio between the length of the coil and its diameter. For this purpose a correction factor (necessary consideration) is needed to correct the result of the formula to fit the individual occasion. This correction factor is generally referred to as a *constant* (q.v.), its value being determined generally from a chart prepared to cover various ratios of length to diameter.

CORROSION—Chemical action which causes destruction of the surface of a metal. In the case of iron it is generally rust, corrosion of copper being due to oxidation, or perhaps the action of stray currents causing *electrolysis* (q.v.). This effect is very often noted in the case of an aerial wire where the wire is necessarily exposed to the action of the atmosphere, rain, etc. The effect is particularly important and undesirable where the lead in wire is joined to the aerial.

If the two wires are not soldered securely or at all, corrosion may set in, which will produce in time a high resistance connection and waste the valuable energy of the incoming signals. This is the main point of danger in an aerial system and the best protection is to solder the joint carefully. As a matter of fact the remainder of the wire may corrode to a considerable extent without disastrous effect. In the case of copper, the wire very quickly corrodes when exposed. For this reason insulated wire is sometimes used for the aerial. As the current received by an aerial travels only on the outer surface of the wire (see *Skin Effect*), and a corroded surface is not a particularly efficient conductor, there seems to be some basis for the use of insulated wire for this purpose. Another instance of corrosion is in a soldered joint within a receiving set. All soldered joints should be carefully wiped to remove any trace of *acid flux* (q.v.), which may remain to produce a destroying chemical effect on the wire and cause loss of energy. (See *Electrolysis*, also *High Resistance Joint.*)

COULOMB—The practical unit for electrical quantity. It is defined as the quantity of electricity delivered by a current of one *ampere* flowing for one second. Therefore a *coulomb* of electricity will pass each second in a circuit having one *ohm* of resistance and a pressure of one *volt.* It is also used to express the quantity of electricity which a condenser having one *farad* of capacity will absorb when placed under a pressure of one volt. (See *Capacity, Electro-static, Volt, Ampere Hour, Practical Units, CGS Units.*)

COUNTER ELECTROMOTIVE FORCE—Commonly known as *Counter E. M. F.* It is an opposing electromotive force which tends to resist or act against the original electromotive force. It may be introduced in a circuit to prevent any fluctuation of the original force. (See *Back Electromotive Force*, also *Choke Coil.*)

COUNTERPOISE—A *counterpoise* is generally a second aerial suspended either below the main aerial or to one side of it. It is used more often in transmitting stations in place of the usual ground connection, but is occasionally also used in reception where it is impossible to obtain an efficient ground. (See *Aerial.*)

COUPLE, THERMO ELECTRIC—See *Thermo-Couple.*

COUPLED CIRCUITS—The term generally used to refer to circuits that are joined by *coupling.* (q.v.) In the illustration the *primary circuit* (the aerial ground and tuning devices) is coupled or joined to the *secondary circuit* by *inductive* coupling. That is to say, the signals impressed on the *primary circuit* are transferred to the *secondary circuit* by means of *inductive coupling;* when the lines of magnetic force around the primary coil cut the secondary coil, a voltage is set up in this coil which causes a current to flow in the secondary circuit, and it is then said the coils are "inductively coupled", there being no connecting wires between them. Signals might be transferred to the secondary circuit by any of several means, but in each case the method of joining the two for the purpose of effecting a transfer of energy is referred to as "coupling the circuits." (See *Coupling.*)

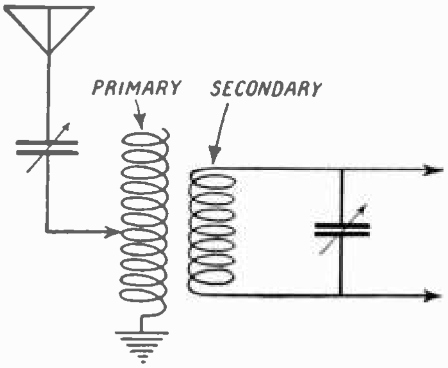

Circuit in which primary and secondary are joined by inductive coupling.

COUPLER—See "Coupler—(Loose)" and "Coupler—(Vario)".

COUPLER—(LOOSE)—A tuner used in a radio receiver, consisting of a primary and secondary coil so constructed that the secondary coil slides on rods in and out of the primary coil. When the secondary is entirely within the primary the coupling is said to be "tight," and as the secondary is withdrawn from the primary the coupling becomes "loose," depending upon the distance which the coils are separated.

COUPLER—(VARIO)—Form of tuner used in a radio receiver, having primary and secondary coils arranged in such a manner that the secondary rotates on a shaft within the primary. The degree of coupling is governed by the variation of the angle between the axes of the two coils.

COUPLING—The relation of one coil of wire carrying alternating currents to another in which energy is set up by the first coil. It is used to define the intimacy between the primary and secondary windings of various types of *oscillation transformers, couplers* and *tuning devices.* This transfer of energy may be produced in several ways, as by *inductive coupling, conductive coupling* or *electro-static coupling.* The more common method is to place the primary and secondary coils in close proximity and produce the energy in the secondary through induction. The illustration Fig. 1 shows a simple two slide tuning coil which is divided into a primary and secondary by *conductive* coupling. Here the transfer of energy from the aerial and ground, or primary circuit, to the secondary circuit (detector and phones) is accomplished with little loss of energy as the circuit is through wire conductors. This is also known as *direct coupling.* Now in Fig. 2 a simple form of *loose-coupler* is shown. This device is made in numerous forms, the principle of operation however, being the same in all cases. Here one coil, the secondary in this case, is made to fit into the larger coil which is the primary. The energy flowing in the primary coil induces or sets up a similar current in the secondary by reason of its proximity. The energy in the secondary, however, will not be as great as in the case of Fig. 1 where it is transferred directly through a wire conductor. The coupling shown in Fig. 2 will on the other hand, permit much closer adjustment to the desired signals.

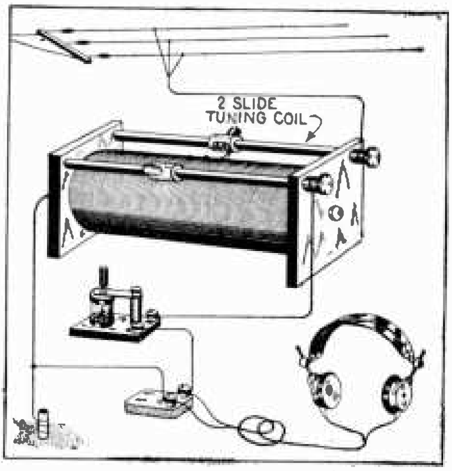

Fig. 1.—Two slide tuner divided into primary and secondary circuits by conductive coupling.

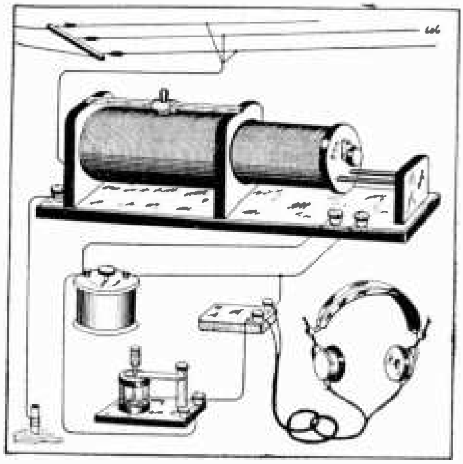

Fig. 2.—Circuit in which primary and secondary are inductively coupled by loose coupler.

The other method is known as *capacitive* or *electrostatic coupling.* This is shown in Fig. 3. Here the secondary coil is not in direct inductive relation to the primary, in fact it might well be almost any distance away were it not for the loss in efficiency due to long connections. The connection between the primary and secondary coils is by means of condensers. The transfer takes place through these and is determined by the setting of the condensers. It will appear that from a standpoint of maximum transfer of energy, the direct or *conductive coupling* system is best. The factor of tuning enters here however, and must be considered. In order to eliminate undesired signals the various systems of coupling are resorted to. It is known that *selectivity*—the ability to tune out undesired signals while receiving the desired signals—is often obtained only at the expense of a certain amount of energy or volume. It is a safe assumption therefore, that if there is any such thing as maximum efficiency in tuning apparatus, it must be a general term covering both selectivity and volume.

Coupling in its various forms is intended to refer not only to receiving circuits but to transmitting circuits as well. Its application in transmitting is covered under the heading *Spark transmission* (q.v.).

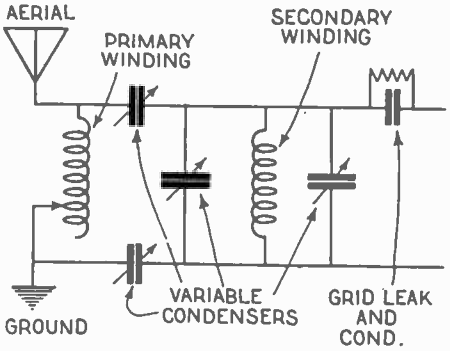

Fig. 3.—Electrostatic coupling between primary and secondary by means of variable condensers.

COUPLING COEFFICIENT — The closeness of *coupling* is specified by a quantity representing the ratio of the *mutual inductance* between two coils to the square root of the product of the two *self inductances.* In its practical sense coupling coefficient refers to the inductive relation of two coils. Coupling coefficient is defined as the ratio

$$\frac{X_m}{X.\ X_2}$$

Where Xm is the mutual or common reactance X is the total inductive or capacitive reactance in the primary circuit and X_2 the similar reactance in the secondary circuit. (See *Coupling, Mutual Inductance.*)

COUPLING, DEGREE OF—The extent of relationship between coupled circuits. In a great number of circuits the coupling is arranged so that it may be adjusted either close or loose, as the requirements of the reception may be. Variable coupling may be obtained either by varying the distance between the two coils of a *loose coupler*, or by changing the angle between the coils of a *variocoupler*. (See *Coupling.*)

COUPLING TRANSFORMER—(OSCILLATION TRANSFORMER) — This is the "tuner" of the transmitting circuit. It consists of primary and secondary windings usually of flat strips of copper, or copper ribbon, instead of wire, in order to reduce losses due to *skin effect* (q.v.). The primary coil receives the energy in the form of alternating currents from the generating device; these currents set up lines of force which interlink with the windings of the secondary coil connected in series with the antenna system, and in this manner the energy of the oscillation (alternating) current is transferred to the radiating system (aerial and ground) and thence into space. The purpose of the oscillation transformer is to produce a sharp wave which is readily tuned at the receiving end and causes a minimum of interference. The greater the distance between the primary and secondary coils the more sharp the wave becomes.

A device serving the same purpose in a receiving set, i.e. to enable the receiver to separate sharply the received signals, is sometimes called a receiving transformer, although commonly known as a tuner, coupler, vario-coupler, etc.

C. Q.—Abbreviation of inquiry made in code by a radio telegraph station desiring to communicate with another station. (See *Abbreviations.*)

C. Q. D.—The original Distress Signal (q.v.) established in 1904 by the Marconi Company. The origin of this signal is said to have developed from the use by telegraphers on land lines sending the C. Q. call when they wished everyone on their circuit to listen to the message being transmitted.

The signal C. Q. D. is now obsolete, having been replaced by S. O. S. at the International Radio Telegraph Convention held at Berlin, Germany, in July, 1908.

CRITICAL—A term often used in reference to the tuning qualities of a radio set, and, less frequently, to indicate a *vacuum tube* that *oscillates* too freely. If a receiving set is not critical in adjustment it is said to tune broadly, that is to say, not sharp in tuning. A critical set has a very narrow margin of control in tuning—the particular signals being receivable within only a very few points on either side of their true wave setting on the tuning dials. If a set is difficult to tune it is termed "too critical". This means that it tunes too sharply and it is difficult to get proper results, due to the possibility of passing by signals unnoticed when turning the dials. Obviously there is a middle point wherein a set may be selective and yet not too difficult to tune. If a vacuum tube goes into *self-oscillation* (q.v.) too readily it is said to be critical. This effect is not desired in a tube except in rare instances. (See *Resonance*, *Tuning*, *Broad, etc.*)

CRITICAL CURRENT—The current required to bring about some particular effect in an electrical circuit. In radio use, a common example is the exact amount of current needed by the filament of a vacuum tube to bring it close to the *oscillating point* (q.v.) without actually going into oscillation. (See *Critical.*)

CRITICAL POINT OF CURVE—The peak or point of maximum value in a curve showing the relation of any varying values. The illustration shows the curve indicating relation between voltage and signal strength in two types of vacuum tubes. The points marked "Peak" in each curve represent the critical points of the curves at which maximum audibility is obtained, any further increase in voltage not increasing the audibility. (See *Characteristic Curve*, also *Curve.*)

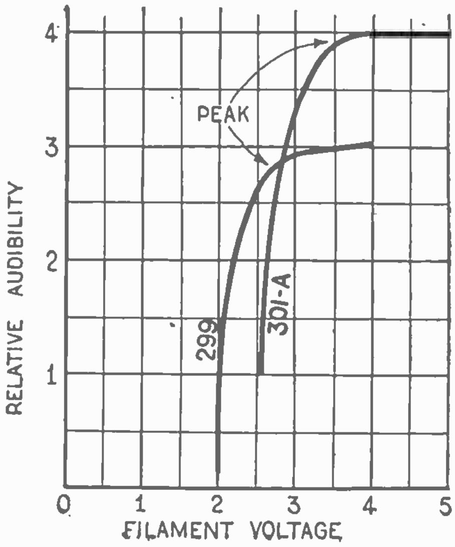

Illustration showing peaks or critical points of curve.

CROOKES, SIR WILLIAM (1832-1919)—English chemist and physicist. Born in London June 17th, 1832, Crookes was educated at the Royal College of Chemistry and afterwards became assistant in the meteorological department of the Radcliffe Observatory, Oxford. In 1855 he obtained a chemical post at Chester.

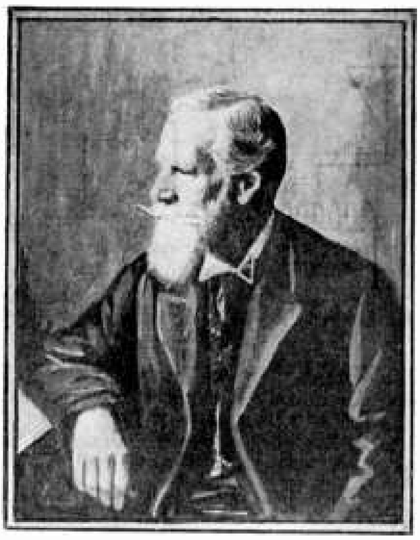

Sir William Crookes.

Crookes early began that series of brilliant researches which has left its mark on the scientific progress of the nineteenth century. In 1861 he isolated the new element *thallium*, during the investigation of the atomic weight of which he made the discovery of his *radiometer*. This led to his researches on the phenomena of electric discharges through highly exhausted tubes, or *Crookes tubes*. The illustration shows such a tube. When a current is passed through it, cathode particles travel along it. His discoveries on these phenomena led directly to the development by Sir J. J. Thomson of the now generally accepted *electron theory*. (q.v.) In 1883 he began his study of the rare earths, and in 1892 he forecast Radio telegraphy on the strength of Lodge's and Hughes' experiments.

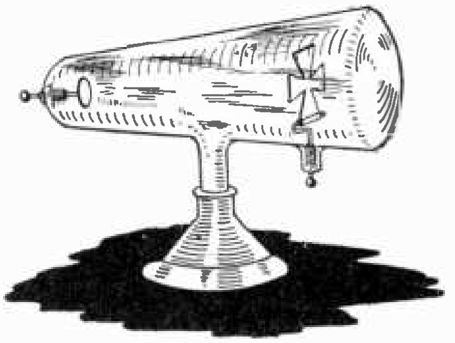

Crookes' tube.

Knighted in 1897, Crookes was awarded the Royal medal in 1875, the Davy medal in 1880, and the Copley medal of the Royal Society in 1904, and the O.M., in 1910. He died in London, April 4th, 1919.

CRYSTAL—A general radio term for a mineral used as a *detector* (q.v.) of radio signals. The Physicist F. Braun noted in 1874 that certain pairs of crystals when arranged so that only a small area of surface was in contact, offered high resistance to the passage of currents in one direction while permitting them to pass readily in the other. This was really the birth of the crystal detector, although radio was not then developed. There are a great many different minerals that will act as detectors for radio signals, some of which operate with a wire, or other metallic contact, others which require a combination of minerals, one in contact with the other. There are also numerous *synthetic crystals* such as *carborundum*. Then again some of the more common natural crystals are subjected to treatment by heat, or other means, to change their nature or to make them more sensitive.

The following crystals are most used in radio reception: *galena*, *silicon*, *molybdenum*, *carborundum*, *zincite*, *bornite*, *chalcopyrite*, and *cerusite*. All of these are taken up under their respective headings. (See *Crystal Detector*, also *Combination Detector.*)

CRYSTAL DETECTOR—A combination of a *crystal rectifier* with contact and supports, used to detect *radio frequency oscillations*. Detectors of this nature are made in a great variety of styles, and use various kinds of crystals. In the more widespread form

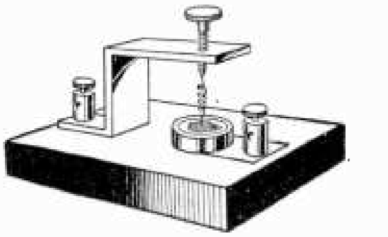

Fig. 1. Crystal detector using contact wire, or "cat-whisker."

the crystal detector consists of a crystal with rectifying properties, held in a cup or spring grip and arranged with a contact wire as shown in the illustration Fig. 1. This may be elaborated upon by supplying a series of flat springs to obtain very fine adjustment of the contact. Fig. 2. shows a combination detector, using two crystals, in contact with each other. This form has several advantages over the wire or *cat-whisker* (as the contact wire is called) type, chief among them the fact that it will withstand severe shocks without losing its adjustment. Fig. 3 illustrates a popular form of fixed crystal detector, in which the contact seldom requires adjustment. This type is of advantage providing it can be made sufficiently sensitive. (See *Carborundum, Vacuum Tube,* also *Crystal Receivers.*)

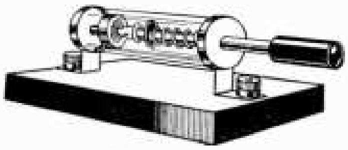

Fig. 2. Combination crystal detector.

Fig. 3. Fixed crystal detector.

CRYSTAL DETECTOR—THEORY OF OPERATION—Theory of operation of a device used to detect radio waves, usually comprising a crystal in combination with a metallic contact, or, in some cases, two crystals arranged to rest against each other. In action this device serves to rectify the incoming signals, that is to say, it changes their nature from an alternating current to pulsating direct current. The incoming waves or signals consist of a series of alternations, the current flowing first in one direction, then reversing and flowing in the opposite direction along the circuit of the receiver. The changes of direction take place many times each second and would not be audible in the head phones in their original state. The crystal, as used in radio reception, has the peculiar property of furnishing a ready path for the flow of current in one direction while offering high resistance to its passage in the other. Now an *alternation* (q.v.) is a rise in current from minimum to maximum in one direction and back to minimum or zero, repeating the process in the other direction. Two of these alternations, one in a positive direction and the other in the negative direction constitute a *cycle.* (q.v.) If the crystal will permit the passage of current in one direction only, it is apparent that only half of each cycle can pass through; the other half of each cycle being lost or blocked out by the action of the crystal.

The illustration "A" gives a graphic representation of the *radio frequency oscillations* as they move over the aerial circuit. The curves above the line show the current in a positive direction while those below the line show it in the opposite or negative direction. If the signals or waves in this state were sent through a telephone receiver or head phone in a radio receiving set no sounds would be reproduced. Then again if the incoming oscillations were continuous, of unvarying strength, the rectified current would merely act on the diaphragm of the phone and hold it in a certain position as long as the signal persisted. This of course would not create any audible or intelligible sounds. If, however, the incoming oscillations were in the forms of groups or trains of waves, in any one of which the successive oscillations were less than the previous ones, the current would be rectified as shown in B and the result would be a series of pulses in the phone as indicated in C. These pulses would occur as often as the trains of oscillations occur (for more complete explanation of the difference between continuous and damped waves, see Damped Waves, also Continuous Waves). It will therefore appear that a *crystal detector* can only be used to receive *damped waves.* When the rectified portion of a damped wave train reaches the phones, the *diaphragm* is depressed and held until the train has died out, at which point the diaphragm is released until another train reaches it. This creates the audible signals and the number of trains occcurring each second will determine the number of depressions of the diaphragm and hence the pitch of the audible signals. (See *Crystal Receiver, Crystal Detectors,* also *Tickler.*)

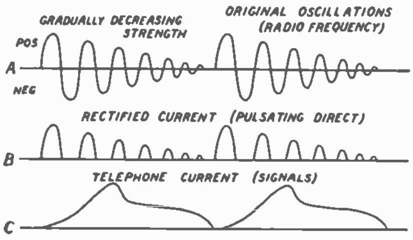

Rectifying process of radio signals.

CRYSTAL RECEIVERS—A crystal receiver is the complete set of apparatus for receiving broadcast programs or *spark transmission* (signals), employing a crystal as the *detector* or *rectifier.* The set comprises a crystal detector, tuning device and phones, together with the usual aerial and ground arrangement. In its simplest form it may be merely a crystal detector, coil of wire and phones as in Fig. 1.

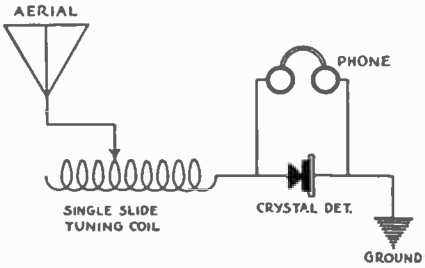

Fig. 1. Simple crystal set with single slide tuner.

Fig. 2 shows a more efficient form, using a two slide tuning coil. This method permits closer adjustment of the receiver to the particular signals desired. A still more efficient form is the *inductively coupled* type shown in Fig. 3. This method allows a very close adjustment and considerably reduces interference. The kind of detector to be used in any of the above types of receiver is immaterial as long as it is sufficiently sensitive. Such an arrangement as shown in Fig. 3 is very inexpensive to construct and when used within ten to twenty miles of the broadcasting station gives excellent results. A very simple arrangement for use with a crystal detector is shown in Fig. 4. This circuit uses a standard *variometer* and eliminates the necessity for a *condenser,* thus having only one control in addition to the adjustment of the crystal.

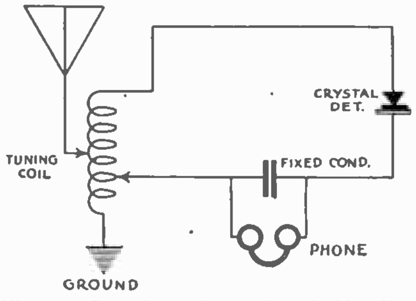

Fig. 2. Crystal receiver using double slide tuner.

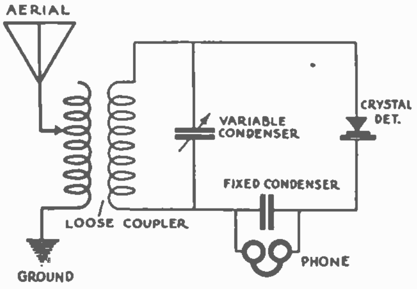

Fig. 3. Selective crystal receiver with loose coupler for tuning.

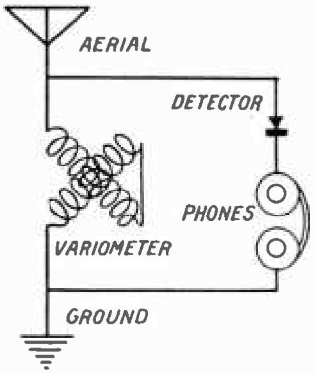

Fig. 4. A variometer is used as the tuning element in this crystal set.

CRYSTAL RECTIFIER—A device for changing an alternating current in a radio receiver into a pulsating direct current. This action is fully explained in "*Crystal Detectors (Theory of Operation).*"

CURRENT—A term used in electrical practise to signify the rate at which electricity flows from point to point in a circuit. The current in a circuit may be likened to the water flowing in a pipe. In this case the water would represent the electricity, the force or pressure represents the voltage, the pipe can be compared to the conductor or wire through which the electricity flows, and the current can be likened to the discharge from the pipe over a given period of time, that is the cubic feet of water per second. For practical purposes the unit of current is considered as the *Ampere.* This is the rate of flow of electricity when one *coulomb* (q.v.) passes a given point in the circuit each second. It will thus be apparent that while the ampere is often considered as the amount of current in a circuit, in reality it is the rate of flow of the electricity, the *coulomb* being the definite unit of quantity. Electric current is of course invisible. The only means of determining when there is current in a circuit is by observing its effect and it is always apparent by one or more of a number of effects. For example, if current is flowing in a wire or other conductor and a compass is placed near the conductor in such a position that the axis on which the needle revolves is parallel to the axis of the wire, the needle will be deflected. If the current reverses and flows in the opposite direction in the wire, the needle will be deflected in the opposite direction. The effect of current is also apparent by reason of the heat which it generates in passing through a wire conductor. If the wire is comparatively small and the current correspondingly large, the heat may be sufficient to be noticeable by placing the hand on the conductor. When the current is small, it can be detected by means of delicate instruments. This heating effect is made use of in various types of *ammeters*. The effect of current may also be noted in an incandescent lamp. In this case the current is made to flow through a fine wire inside a globe from which the air has been extracted, and the heat generated by the passing current causes the filament, or fine wire, to glow.

It is a well-known fact that heat causes expansion and this expansion of a fine wire is the principle on which one type of current measuring device operates.

Current is divided broadly into two divisions, *direct* and *alternating*. These two classes may again be separated into several further groups such as *high* and *low frequency alternating*, *continuous* and *pulsating*. Each phase or characteristic of current is taken up more in detail under its particular heading in this encyclopedia. (See *Current, Direction of Flow*, also *Current, Production of*.)

CURRENT, ALTERNATING—A current which does not flow steadily in one direction but changes its direction in the circuit periodically. (See *Alternating Current*.)

CURRENT CARRYING CAPACITY OF WIRE—The amount of current which a wire or other metallic conductor will carry without over-heating. Whenever current flows through a conductor, heat is generated. The amount of heat will be directly proportional to the resistance of the conductor and the square of the current in it. Now the resistance of a conductor depends on its cross-sectional area, length, and also its nature. It is therefore correct to state that the heating will depend on the amount of current flowing and the cross section and length of the wire. Obviously, a large wire will carry more current without over-heating, than will a small wire. Thus, it is necessary when choosing wire for a conductor to carry a certain amount of current, to select a wire of a cross section sufficient to carry the desired amount of current without undue heating.

The following formula for determining the current carrying capacity of various sizes and kinds of wire uses a certain factor T, which is the permissible rise in temperature above surrounding medium—such as air, earth or water. Where d is the diameter of the conductor in inches, T—the permissible temperature rise in degrees centigrade, N the resistance of conductor in ohms per mil-foot at final temperature, and I the current in amperes, then for solid conductors:

$I = \frac{K\sqrt{Td^3}}{r}$ and for stranded conductors: $I = 0.85\,K\frac{\sqrt{Td^3}}{r}$

K is a *constant* (q.v.) depending upon the condition of the surface of the wire and upon the amount of heat convection due to air currents. Values of the constant K for air differ according to different authorities from 800 to 1000, the former referring to still air and the latter to open air.

Table of current carrying capacities follows:

Wires and Cables, Insulated: Carrying Capacities, in Amperes, allowed by The Regulations of The National Board of Fire Underwriters for Interior Copper Conductors.

(For Aluminum 84 Per Cent of These Currents is allowed.)

Single Conductor Cables or Each Conductor of Multiple Conductor Cables.

A. W. G.	Area in Circular Mils	Table A. Rubber Insulation	Table B Varnished cloth	Table C. other insulation
18	1,624	3	....	5
16	2,583	6	....	10
14	4,107	15	(18)	20
12	6,530	20	(25)	25
10	10,380	25	(30)	30
8	16,510	35	(40)	50
6	26,250	50	60	70
5	33,100	55	65	80
4	41,740	70	85	90
3	52,630	80	95	100
2	66,370	90	110	125
1	83,690	100	120	150
0	105,500	125	150	200
00	133,100	150	180	225
000	167,800	175	210	275
	200,000	200	240	300
0000	211,600	225	270	325
	250,000	250	300	350
	300,000	275	330	400
	400,000	325	390	500
	500,000	400	480	600
	600,000	450	540	680
	700,000	500	600	760
	800,000	550	660	840
	900,000	600	720	920
	1,000,000	650	780	1000
	1,100,000	690	830	1080
	1,200,000	730	880	1150
	1,300,000	770	920	1220
	1,400,000	810	970	1290
	1,500,000	850	1020	1360
	1,600,000	890	1070	1430
	1,700,000	930	1120	1490
	1,800,000	970	1160	1550
	1,900,000	1010	1210	1610
	2,000,000	1050	1260	1670

Varnished cloth smaller than No. 6 may be used by special permission only.

CURRENT, CONVECTION—See *Convection Current*.

CURRENT, CRITICAL—See *Critical Current*.

CURRENT DENSITY—The number of amperes passing through a wire conductor per square inch of area of the conductor. For example, if a conductor is a wire or bar of copper having a cross section of one square inch and ten amperes are passed through it, the current density will be ten amperes per square inch.

CURRENT, DIRECT—Current that flows steadily, and only in one direction. (See *Direct Current*.)

CURRENT, DIRECTION OF FLOW (ASSUMED)—In electrical practice the *assumed direction of flow* of current is often referred to. This refers to the fact that direct current is assumed to flow from the *positive pole* of a battery, or other source of electrical energy, to the *negative pole* and thence back through the battery from negative to positive. This assumed action is shown in the illustration Fig. 1, where the arrows indicate the assumed direction of flow. Now while the current is stated as flowing from positive to negative outside the source, it is the generally accepted theory that the electron stream moves in the opposite direction, or from negative to positive. In other words, the current as we use the term, flows backward along the path of electrons. (See *Electron Flow*.) This reverse action of current and electrons is shown by the illustration Fig. 2 This apparent anomaly

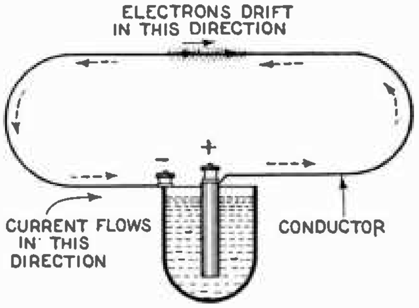
ELECTRONS DRIFT IN THIS DIRECTION
−
+
CURRENT FLOWS IN THIS DIRECTION
CONDUCTOR

Fig. 2. Showing reverse action of current and electrons.

can be likened to a salmon swimming up a rapid. The movement of electrons can be compared to that of the the water in the rapids and while the fish actually swim against the stream, and the electric current flows against the electron flow, the water and the electron stream nevertheless furnish the path of motion in each instance. The reason for the contradiction lies in the unfortunate arbitrary choice of positive and negative poles long before electrons were known. (See *Electron.*)

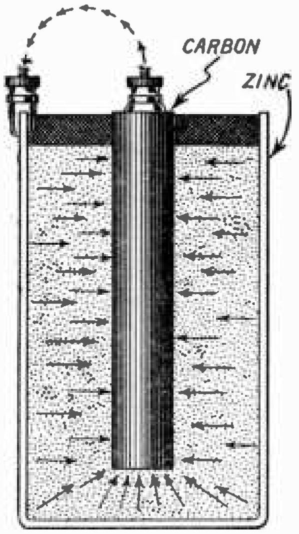

Fig. 1. Illustration showing the assumed direction of current flow in a dry battery.

CURRENT, EDDY—See *Eddy Currents.*

CURRENT, HIGH FREQUENCY—A current that changes direction many times each second. In radio usage, currents having a frequency over ten thousand cycles per second are considered as *high frequency* or *radio frequency currents.* (See *Frequency, Radio Frequency* also *Low Frequency.*)

CURRENT, HIGH TENSION—Alternate term *High Pressure.* When the voltage, tension or pressure, in a circuit is comparatively low, the current is said to be a low pressure or *low tension current.* When the voltage is high, the current is said to be of high tension or pressure. Thus, as applied to radio usage, a dry cell or storage battery giving, respectively, one and a half and six volts for a certain type, are said to produce low tension or low pressure current. A "B" battery (q. v.) giving from 22½ to 45 volts (the usual value) furnishes *high tension* or *high pressure current.* While there is not a very great difference in voltage between a small "B" battery and an "A" type storage battery, the ratio of voltage to current in a "B" battery is far greater than in the ordinary case of the "A" battery and this must be taken into consideration. In general electrical practice the terms are not so arbitrary, and high tension currents might be taken as those above a pressure of several thousand volts.

CURRENT, LOW FREQUENCY—An alternating current that changes its direction of flow a comparatively few times per second. In commercial electrical work any frequency up to about 500 cycles per second is considered as low frequency current. In radio phraseology *audio frequency* (q.v.) currents are termed low frequency. For ordinary radio practice low frequency may be considered as any frequency below about 10,000 cycles. (See *Oscillation* also *Oscillatory current.*)

CURRENT, OSCILLATORY—See *Oscillatory Current.*

CURRENT, PRIMARY—A current which flows directly from its source, such as a cell or generator. The current obtained from any source through a direct conductive circuit is termed a primary current. (See *Secondary Current,* also *Induced Current.*)

CURRENT, NATURE OF ELECTRIC—As explained under the heading *Current,* electric current is manifest or understood by its effects. It may be apparent through the production of heat or light, by producing mechanical action such as required to operate a bell or move an electric motor, or, again, by producing chemical changes (See *Electrolysis*), and still further in its disastrous effects on the human body when the current has sufficient intensity. Without any actual or visible form of manifestation it can best be understood from the *electron* basis. It will be understood that an *electron* (q.v.) small as it is, contains a charge of electricity. When a sufficient number of electrons are in motion in a conductor (Note: current consists of electrons in motion) the current becomes large enough to be measured. As an instance of this fact a current of one ampere intensity (one coulomb per second) requires that about 10^{19} electrons flow past a given point in the circuit each second. When the electrons are not progressing along a conductor, no current flows in the circuit. If no current is being carried by the conductor, the electrons may move about, but they do not progress along the conductor.

CURRENT, PRODUCTION OF ELECTRIC—An electric current is produced in a conductor by applying an *electro-motive force* (voltage) at points of the conductor or by maintaining a difference of *potential* (q.v.) between two points. In the first example we can assume a metal ring or closed coil of wire through which a magnet is pushed or perhaps in the core of which a magnet is being excited, or in the general sense any complete circuit through which the number of magnetic lines of force are being varied. In the second instance the difference of potential may be due to a *primary cell* (q.v.). Thus, the first case can be understood as referring to any mechanical means of producing current, such as *generators.* The other general method is by use of a *voltaic cell,* that is, by chemical action. There are four important methods of producing an *electro-motive force.* The first is by friction, the apparatus used in this connection being termed a static machine; the second is by chemical action, as a primary or secondary cell; third, by mechanical motion, i.e., *dynamos* and *generators,* and fourth, by thermal action—known as a *thermo-junction* which is the wiring of two unlike metals in a conductory circuit so that by heating one of the points of joining an electric current will be set up in the circuit. (See *Static Electricity, Generator, Thermo Couple,* also *Current* and *Electro-Motive Force.*)

CURRENT, PULSATING DIRECT—When an alternating current is rectified by means of any device which permits current to pass only in one direction, the resultant current will move only in one direction, hence it is known as a *direct current.* but unless both negative and positive halves of the *cycle* (q.v.) are rectified or changed to direct current, the resultant current will appear as a series of pulses in one direction. The most common application of this is found in the case of a *crystal rectifier,* or *detector* as used for receiving radio signals or broadcasting.

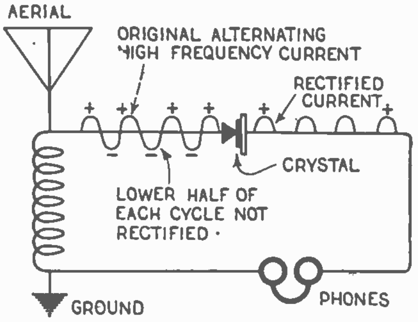
AERIAL
ORIGINAL ALTERNATING HIGH FREQUENCY CURRENT
RECTIFIED CURRENT
+
−
CRYSTAL
LOWER HALF OF EACH CYCLE NOT RECTIFIED
GROUND
PHONES

Diagram showing how crystal detector converts alternating high frequency current into pulsating direct current.

When the incoming waves, or high frequency *oscillations* (Q.v.), are impressed on the detector, due to its peculiar properties they can only pass through it in one direction. In other words, granting that the incoming signal, or current, reverses its direction a great many times each second, it will be obvious that at any instant in the *cycle* (q.v.) current is moving in a certain definite direction, as from positive to negative, or vice versa, as stated, due to the peculiar properties of the detector the incoming signal current can only pass through the detector when the current is flowing in the correct direction. This means that only half of the *wave* (q.v.) is passed through and rectified. The result is a series of direct current surges through the balance of the circuit, each surge of current being followed by a space or time interval while the opposite half of the *cycle* or opposite *alternation* (q.v.) is blocked by the uni-directional nature of the crystal. These surges are then audible in the *head phones,* as they will actuate the *diaphragm* (q.v.) whereas a high frequency alternating current such as the original impulse from the broadcasting station would be inaudible. For a more complete explanation of this action see Rectifying action of crystal detector. (See also, *Rectifier, Full Wave Rectification and Rectifying Tube.*)

CURRENT, SECONDARY—A current produced by a *primary current* (q.v.) acting on a *secondary curcuit* which does not receive its energy by direct connection with the primary source. The illustration shows both primary and secondary currents and the manner of production. It will be understood that the primary current in this case must be a varying current. (See *Induced Current,* also *Inductive Coupling.*)

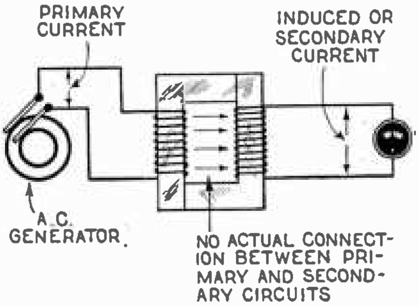
PRIMARY CURRENT
INDUCED OR SECONDARY CURRENT
A.C. GENERATOR
NO ACTUAL CONNECTION BETWEEN PRIMARY AND SECONDARY CIRCUITS

Illustration showing primary and secondary currents and method of production.

CURRENT, THEORY OF ELECTRIC—Under Electromotive force it will be explained that voltage or pressure is the result of a difference of potential (the power possessed by a charge of electricity for doing work) between two points in a conductor. Electric current will flow from the point of higher pressure or *voltage* to the point of lower pressure, the flow continuing as long as the difference is maintained. This action is the same as with water, which will always seek to move from a high point to a lower point. The flow of current is thus an effort to equalize the two *potentials*. (See *Current, Assumed Direction of Flow*, also *Pressure and Voltage*.)

CURVE—A straight or curved line showing the relation between various electrical phenomena or the characteristics (relation of changing values) of an electrical instrument. (See *Characteristic Curve*, also *Sine Wave*.)

CURVE OF SINES—A curved line representing the vibration of a body that oscillates like a pendulum. The illustration shows a *curve of sines* which may be used to represent an alternating current. (See *Alternating Current*, also *Sine Wave* and *Simple Harmonic Vibration*.)

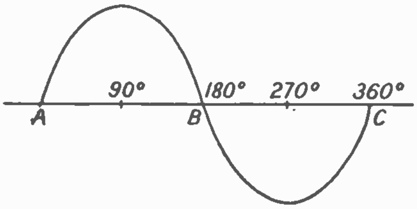

Sine curve.

CUT OUT—An electrical device to interrupt the flow of current through any particular piece of apparatus or instrument, either automatically or by hand; a switch.

CYCLE—In electricity, a period of time during which certain changes take place in an *alternating current*, (q.v.) the same changes occurring again in each successive cycle. We can imagine an alternating current as travelling along a wire in the form of a wave, one half of the wave being in one direction and the other half being in the opposite direction. That is to say, the current flows from positive to negative for one half of the cycle and then in effect, flows from negative to positive the other half of the cycle. (It will be understood that actually current always flows in the same direction, but the action of alternating current is due to a reversal of polarity at the source.) The difference between direct and alternating current is that, in the case of direct, the polarity always remains the same at the source and hence the direction of flow of the current is always in one direction, whereas with alternating current, the polarity changes twice each cycle due to changes at the source, and thus it flows in one direction for a certain fixed period, then reverses and flows in the opposite direction. In the illustration the current is shown as a *sine wave*. (q.v.) We assume that this represents an alternating current of 110 volts pressure such as in the average house lighting system. Now we consider the wave as a circle. The current flows in a positive direction toward its maximum value and at the 90 degrees it has reached the positive maximum. It then falls back to zero at 180 degrees and reverses, rising to maximum on the negative side, which it reaches at 270 degrees. It then falls back to zero at 360 degrees having completed a cycle or two alternations. The operation is repeated over and over a certain number of times each second. Thus, the rise and fall from A to B or 180 degrees is one *alternation*, the rise and fall from B to C is another alternation, the two alternations, one positive and one negative making a complete cycle. If sixty of these cycles occur each second we say that the frequency (q.v.) is sixty cycles. (See *Alternating Current, Phase, Sine Wave*.)

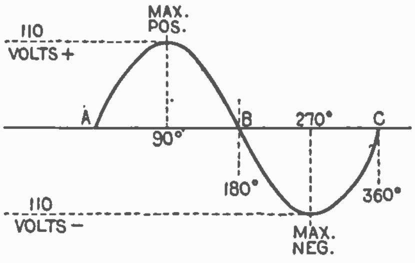

Curve showing action of 110 volts alternating current during one cycle.

CYMOMETER—The name given a type of *wave-meter* designed by Dr. J. A. Fleming. It is used to measure the wave-length of oscillatory (vibrating) circuits. (See *Wave-Meter*.)

CYMOSCOPE—A term used to designate any instrument which enables one to see the effect of electrical waves or to detect their presence. The original form was merely a loop of wire not completely closed as shown in Fig. 1. When held close to a transmitter of short waves a small spark will pass between the ends of the loop providing the *wave-length* is properly adjusted. This was the form used by Hertz (the discoverer of electro-magnetic waves which form the basis of radio telegraphy and telephony) in his early experiments. The more modern form is shown in Fig. 2 where a small lamp is made to light by holding the loop close to the transmitter. A complicated arrangement used to acually see the waves and study their form is called the *Oscillograph*. (See *Oscillograph*, also *Cathode Ray Tube*.)

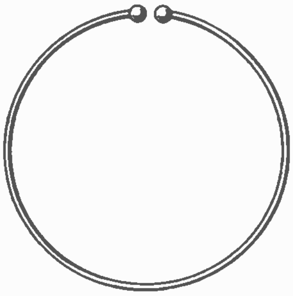

Fig. 1. Early type of cymoscope used to indicate resonance in a circuit.

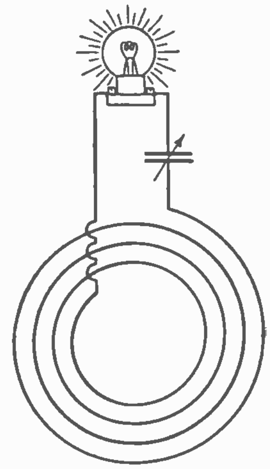

Fig. 2. More modern form of cymoscope.

D

D—The symbol of *electric displacement* (q.v.); also occasionally used as a symbol for diameter in electrical calculations.

DAMPED OSCILLATIONS — Electrical oscillations which die away, each succeeding oscillation in a group having lesser *amplitude* (q.v.) or strength than the preceding one. (See *Damped Waves*.)

DAMPED WAVES—Radio waves in the form of successive trains or groups, in each of which the amplitude or strength decreases with each successive wave. When electrical oscillations are caused by a single impulse, they do not continue indefinitely, but decrease in amplitude or die away. Thus, if oscillations are created in an antenna or other circuit by a discharge from a *condenser* (see spark discharge), each electric spark creates a train of oscillations which die away more or less gradually. Fig. 1 shows a pendulum which is given a momentary impulse. Now as this source of power is only momentary the pendulum will not swing from A to D indefinitely. If the first swing under the impulse carries the pendulum the entire distance A to D, the next swing will cover a lesser arc and so on until the movement has died out altogether.

Now in Fig. 2, the graphic illustration of a train of *damped waves* shows that each crest is a little lower than the one preceding it and after a short period the waves have died out alto-

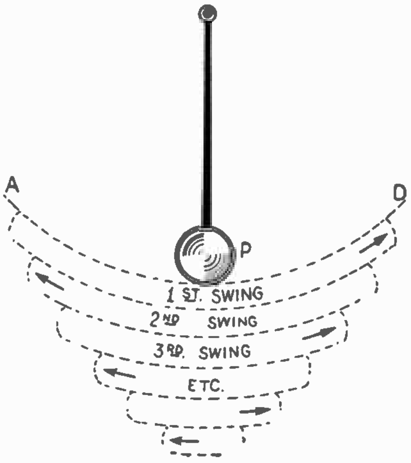

Fig. 1. Pendulum coming to rest after being given a momentary impulse.

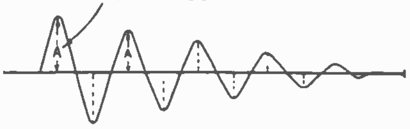

Fig. 2. Train of damped waves.

gether. These curved lines may be considered as an illustration of the path taken by the pendulum in its motion as shown in Fig. 1. The decrease in amplitude of one crest over that preceding is known as the damping, and the naperian logarithm of the ratio of amplitude of any oscillation to the one preceding it is known as the *logarithmic decrement.*

If the impulse of force causing the oscillations is steady—not intermittent—the result will be a sustained series of oscillations that will continue at constant strength or amplitude until the exciting force is withdrawn. Thus, in the case of the pendulum, we can very easily keep exciting it by an evenly exerted force and so insure that the swings will continue without diminishing, as in the case of a momentary impulse. Oscillations or waves that have constant amplitude are known as *Continuous Waves, Undamped Waves* or *Sustained Waves.* (q.v.) (See also *Decrement, Logarithmic;* also *Naperian Base.*)

DAMPING—A progressive decrease in amplitude or strength. (See *Damped Waves, Damping in Instruments.*)

DAMPING COILS—*Short circuited* (q.v.) coils used in electrical apparatus for the purpose of causing damping (i. e., decrease in the intensity of electrical oscillations). The reaction of induced currents in the shorted coil creates the damping effect. (See *Damping.*)

DAMPING DECREMENT—See *Decrement, Logarithmic.*

DAMPING FACTOR—The product of the *logarithmic decrement* (q.v.) and the frequency (rate of alternation) of a damped alternating current. (See *Decrement,* also *Amplitude.*)

DAMPING IN INSTRUMENTS — The term damping is applied in a mechanical sense as in the case of the needle of a measuring instrument such as a *galvanometer.* (q.v.)

In a delicate instrument for measuring current the needle is controlled by the current through the instrument, and when this current is withdrawn the needle is brought back to the neutral position by a fine spring. Now, unless some precaution is taken the needle will have a tendency to swing or vibrate when measurement is being taken. Various devices are used to bring the needle quickly to rest without undue vibration. If the needle comes to rest almost immediately it is said to be a highly damped or *dead beat instrument,* because the oscillations or vibrations die out rapidly the same as in the case of damped waves.

There are two principal methods of damping instruments. When the instrument is of the moving coil type, the coil is generally wound on a light aluminum frame which is mounted between the poles of a permanent magnet. A soft iron core fills the gap between the two poles, allowing space for free movement of the coil. The result is that the metal frame carrying the coil moves through a very strong *magnetic field* and *Eddy currents* (q.v.) are set up in the frame and serve to damp its motion. The other common method is to have a light metallic vane carried by the pointer in its motion and moving in an air chamber to stop vibration. While damping is highly essential in a measuring instrument, too high a degree of damping is not advisable. If there is a very slight tendency on the part of the needle to vibrate, it permits the operator to determine whether it is swinging freely. (See *Voltmeter, Wattmeter.*)

DAMPING MEASUREMENT—See *Decrement, Logarithmic.*

DAMPING WAVES—Term commonly used to denote the gradual decrease in amplitude of a train of oscillations as in radio transmission. If the oscillations persist at constant amplitude or strength, each oscillation being equal in amplitude to the one preceding, there is no damping. When the amplitude of each successive oscillation is less than that of the preceding one there is said to be damping and the oscillations will die out entirely, at a rate dependent on the degree of damping. If the oscillations die out very rapidly they are said to be highly damped; if they die out slowly the damping is said to be feeble.

DANIELL'S CELL—A two fluid *voltaic cell* containing a zinc plate immersed in dilute sulphuric acid, and a copper plate in a saturated solution of copper sulphate; the two solutions being separated by a porous cup. This cell has a constant voltage and shows only slight polarization. Was formerly much used in telegraph work.

D'ARSONVAL GALVANOMETER—A type of galvanometer in which a coil carrying the current to be measured is suspended in a *fixed magnetic field.* (See *Galvanometer.*)

DASH—The long stroke used in *Morse* and *Continental Code.* It is considered as equivalent in length to three dots. (See *Code.*)

D. C. C.—The customary abbreviation for double cotton covered wire.

D. C.—The abbreviation commonly used for *direct current.* (q.v.)

"D" COIL—Occasionally known as Figure Eight coil. A form of inductance, used particularly in tuned radio frequency circuits. The windings are in the shape of a figure eight, either in pancake form (flat wound) or wound on a slotted tube. Owing to the peculiar arrangement of the windings these coils have very little external field and thus can be used in radio frequency amplifiers without the usual oscillating due to conflicting fields. (See *Field, Inductance, Radio Frequency Coils.*)

DEAD BEAT INSTRUMENT—A measuring instrument arranged so that the needle comes quickly to rest with little or no vibration is said to be *dead beat* or highly *damped.* (See *Damping in Instruments.*)

DEAD CELL—A *cell* which has been discharged. (See *Cell.*)

DEAD CRYSTAL—A crystal used in a *crystal detector* (q.v.) that has lost its sensitivity. Very often a crystal becomes dirty or oily from contact with the fingers. In some cases they can be brought back to usefulness by cleaning with alcohol, or if the crystal is dead through causes other than dirt on its surface, it may be broken to present a new surface and thus find a new sensitive spot.

DEAD END EFFECT—The effect on a receiving circuit (oscillatory circuit) resulting from unused turns of wire in a coil. In the customary tuning coil or tapped coil a certain portion of the wire is not in use and this idle section, being in metallic connection with the main portion, but not actually in use, acts as a miniature oscillatory circuit. This tends to reduce the efficiency of the tuning unit as a whole. Dead end effect may be eliminated by using the entire coil at all times, doing the tuning solely by the variable condenser, or it may be reduced by the use of a *dead end switch* (q.v.). (See *Tuning Coil.*)

DEAD END SWITCH—A switch used to cut out unwanted turns of wire in a coil in order to eliminate or reduce the *dead end effect* (q.v.). Such a switch is generally arranged so that instead of taps being taken from various points on the coil, the winding is actually in sections and the switch blade connects into the circuit the desired number of sections.

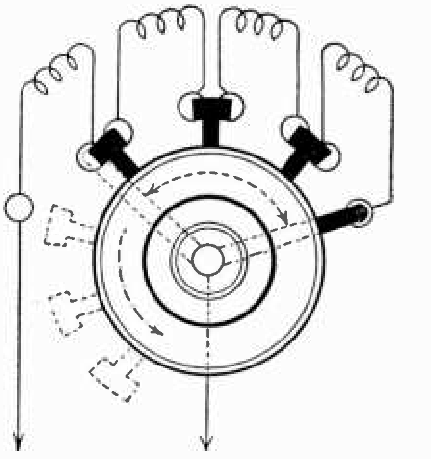

Dead end switch.

DEAD SPACE—In the case of *beat heterodyne* reception if the two frequencies are brought together or the difference is very slight there will be no beat. This space where the two frequencies are identical, or very nearly so, is known as the *dead space.* (See *Beats,* also *Heterodyne.*) The beat resulting from the superposition of two waves differing in frequency is the difference or the sum of the two frequencies. For instance, if two waves are present in the same circuit, of 1,000 and 1,005 kilocycles respectively, the two beat frequencies will be 5 and 2,005 kilocycles. If, however, the two frequencies are made identical, the difference between the two will be zero, and the sum of the two will be twice the original frequency. In the one case, where we take the difference of the frequencies, the beat note is zero, or there is no beat note at all; in the other case, where we add the two frequencies, the resulting frequency is the *harmonic* (q.v.) of the original frequencies, which is the same as an octave in music. In the present instance, this note will not be heard at all as it is above the range of audibility.

DECADENT WAVE—A damped wave. A wave, or oscillation, in a train of *damped waves* (q.v) which is of lesser amplitude or strength than the one preceding it. (See *Damped Waves,* also *Undamped Waves.*)

DECAY OF CURRENT—The gradual dying out of current in an *inductive circuit* (q.v.) after the *impressed potential* has been removed. The presence of inductance in a circuit causes the current to lag in time behind the voltage producing it. As a result, when a highly inductive circuit is broken by opening a switch or by other means, the current still tends to flow for a short interval, and the *induced voltage* in the circuit may become very high. This is the reason why a spark often occurs when the switch in a highly inductive circuit is opened. As a result the current decays gradually after the impressed voltage is removed.

It will be noted that the *decrement* (q.v.) or the rate of decay of the oscillations directly affects the number of oscillations in the wave train that will be useful in transmitting signals. For instance, if a certain oscillation in the figure should represent the smallest voltage that will operate a detector, it is evident that all oscillations in the wave train that have smaller amplitudes will be useless. Consequently, the lower the decrement of the wave, the greater will be the number of oscillations before the amplitude becomes too small for useful purposes. This likewise means that a greater amount of the energy in the wave can be utilized.

DECOHERER—In the type of detector, now obsolete, known as the *coherer*, the incoming signals caused the particles of nickel and silver filings in the glass tube of the device to cohere or cling together. (See *Coherer.*) In order to place the coherer in its original state after each impulse, it was necessary to have some means of separating the particles. These were known as *decoherers* and were in many different forms, some as separate attachments, others as a part of the detector or coherer. (See *Detector.*)

DECOMPOSITION, ELECTROLYTIC—The decomposition of chemicals taking place in a cell due to the action of the electrolyte. (See *Cell*, also *Electrolyte.*)

DECREMENT—A term used to indicate the rate of decay or dying out of an electrical oscillation that is subject to *damping.* (q.v.) In the illustration is shown a series of damped waves. It will be noted that the distance from zero to the crest of wave (C-2) is less than the one preceding it (C-1). This distance represents the amplitude and the ratio of any amplitude to the one preceding it is constant, that is the decay or decrease in amplitude is constant. The *Naperian Logarithm* (q.v.) of the ratio—of one wave to the one preceding it—is called the *logarithmic decrement.* The damping of a train of waves is an important consideration as it affects the tuning qualities to a great extent. This is covered by a U. S. Government Statute Concerning Transmitting Stations, decreeing that the logarithmic decrement per complete *oscillation* must not exceed 0.2, which means that for each single spark discharge from the transmitter there must be not less than 24 complete oscillations in the *antenna system.* (See *Oscillations, Damping, Naperian Logarithm, Sharp Wave.*)

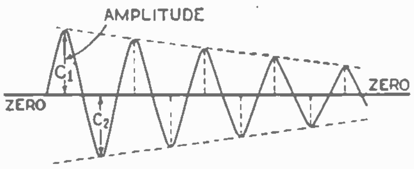

Damped waves gradually dying out.

DECREMETER—An instrument used to measure the *decrement* or degree of damping in an oscillatory current. Such meters are made in several different forms. Some types are direct reading in terms of the logarithmic decrement, while others require mathematical reduction from the readings of a dial. All decremeters operate on the basis of comparison of the resonance current in a tuned circuit with that in a circuit out of tune by a known or unknown percentage. (See *Logarithmic Decrement*, also *Damping.*)

DE FOREST, DR. LEE—Electrical engineer; b. Council Bluffs, Ia., Aug. 26, 1873. Grad. Yale (Sheffield Scientific), Ph. B., 1896; Ph. D., 1899. Inventor De Forest system wireless telegraphy and founder Am. De Forest Wireless Telegraph Co.; system originally adopted largely by U. S. Govt.; demonstrated to the British, Danish, German, Russian and Indian Govts., first used by the U. S. Signal Corps in the war manoeuvres and inst. in Alaska. 1905 invented the Audion or three-electrode vacuum tube. 1906-09 was the pioneer in Radio Telephony. First to broadcast music and opera. 1908 equipped the 24 battleships and destroyers of Admiral Evans for their historic round-the-world cruise. 1913 incorporated De Forest Radio Tel. & Tel. Co., which developed the audion amplifier and oscillator for commercial purposes. These De Forest devices completely revolutionized the radio art and made possible the Bell Tel. Transcontinental System. He was the father of radio broadcasting and all radio broadcasting apparatus is built upon his audion inventions. Since 1919 he has devoted his attention to the phonofilm, photographing sound waves directly on motion picture film. Pres. of the De Forest Phonofilms, Inc.; Dir. of De Forest Phonofilms, Ltd., London. 1921, Awarded Cross of Legion of Honor by the French Govt. in appreciation rendered by audion to the Allies during the war. 1922 awarded Elliott Cresson medal by Franklin Inst. 1922 awarded medal by Inst. of Radio Engineers. Fellow American Inst. of Elec. Engineers. Fellow Inst. of Radio Engineers. Mem of N. Y. Elec. Society; mem. of Nat Geog. Soc.

Dr. Lee De Forest.

DE FOREST COILS—A type of coil named after *Lee De Forest* (q.v.). The wires are wound in a diamond shaped pattern in such manner that the turns of one layer cross the turns of the preceding layer at an angle, thus making the *distributed capacity* (q.v.) very small compared with the capacity of other types of coils. (See *Honeycomb Coils.*)

DEGREES, ELECTRICAL—In an alternating current curve each complete oscillation or cycle is shown as two loops, one above the line representing time, and the other below it, the entire cycle being considered as 360 *electrical degrees.* It is possible in this manner to show relations between current and voltage phases, amplitude at any point, etc. (See *Alternating Current*, also *Sine Wave.*)

The idea of electrical degrees comes originally from the *elementary electric generator*, with two-slip-rings taking off the alternating current, and having only two poles. During one-half revolution of the armature turns, the current increases from zero to a maximum and then back to zero again. During the other half revolution the current increases in an opposite direction from a zero value to maximum and then back to zero again. The curve shown in the figure represents the current variation during the complete revolution, a crosssection of the rotating armature wire being shown at A. The angles indicated on the line BC are laid off to any convenient scale, and the intersection of the ordinates from these points with horizontals drawn from the circle as shown, give the points on the curve.

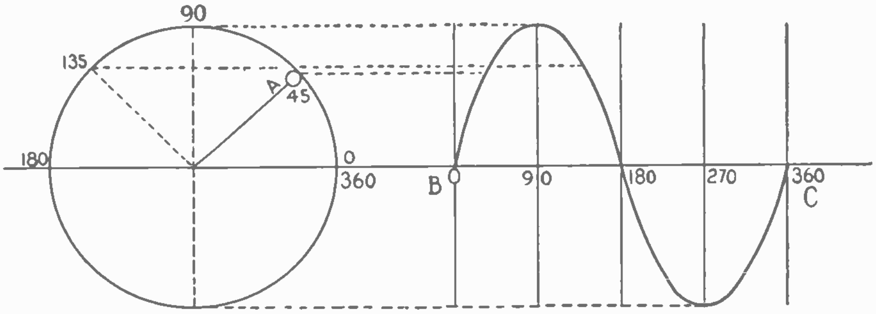

Curve representing current variation during one revolution of the armature of a generator.

The same ideas hold when the generator has more than two poles; in this case a full 360 electrical degrees is taken to represent the armature motion from across two adjacent pairs of poles, so that a complete cycle of voltage or current is generated. In a four pole machine, therefore, a full revolution of the armature could be represented by 2 x 360 or 720 electrical degrees. This corresponds to 2 complete cycles. Or, looking at it the other way, 360 electrical degrees would represent a half revolution of the armature wire.

DEIONIZATION—The action of returning a mass of gas that has become *ionized* (breaking up of the atoms of gas into positive and negative ions being the elements to which and to whose motion, under the action of electric forces, is supposed to be due their electric conductivity) to its original state. In the case of a *spark gap* (q.v.) the ionization creates a conductive path which remains after the actual spark has passed, thus preventing sharply defined oscillations. There are various methods of deionization, including cooling, absorption, air-blast or magnetic field, all tending to deionize or place the spark gap in its previous condition to prevent an arc from following the discharge. This action is also known as *quenching.* (See *Quenched Gap.*)

DELTA—The fourth letter in the Greek alphabet. The capital letter delta (Δ) is used as a symbol for delta connection of *three phase alternating current generators or transformers.* The small or lower case delta (δ) is used as a symbol for *Logarithmic decrement.* (See *Decrement, also Delta Connection.*)

DELTA CONNECTION—Δ In a three phase *alternating current generator,* where three coils are mounted symmetrically around a shaft rotating in a *magnetic field,* the connection or grouping of the windings is called *delta* after the Greek letter which the grouping somewhat resembles. The illustration shows delta grouping in a three phase system. The sum of the instantaneous Electromotive Forces of the three coils is zero, or in other words the sum of the Electromotive Forces of any two coils is equal and opposite to that of the other coil, the *line voltage* thus being equal to the *phase voltage.* (See *Alternator, Polyphase.*)

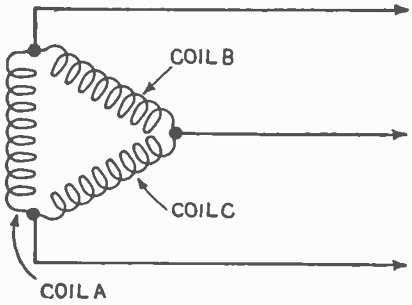

Delta grouping in a three phase system.

DEMAGNETIZATION—The act of returning a body to an unmagnetized state or of reducing the degree of *magnetization.* In the case of phones used in radio reception the permanent magnets are often weakened or *demagnetized* by connecting them in the *electron-tube circuits* in the wrong way. (See *Magnetization,* also *Telephone Receiver.*)

DENSITY, CURRENT—See *Current Density.*

DENSITY, FLUX—See *Flux Density.*

DENSITY OF ELECTROSTATIC CHARGE—The *electrostatic charge* per unit area. (See *Electrostic Charge,* also *Condenser.*)

DEPOLARIZATION—In an electric cell hydrogen bubbles form on the surface of the positive electrode and unless some arrangement is made to counteract this effect, the usefulness of the cell will soon be impaired due to the insulating action caused by this film of gas, known as *polarization.* Usually some means of oxidizing is employed to act upon the hydrogen as fast as it is produced on the positive electrode, thus preventing or reducing the effect of *polarization.* (See *Cell,* also *Polarization,* and *Depolarizer.*)

DEPOLARIZER—The oxidizing agent or other means used in a cell to counteract the effect of polarization. Bichromate of potash or peroxide of manganese is commonly used for this purpose. (See *Dry Battery,* also *Polarization.*)

DETECTOR—A device for converting oscillating currents of high frequency (radio waves) into a form suitable for operating a telephone receiver or sensitive measuring instrument. It is often referred to as a *rectifier,* because it serves to change the incoming currents from *alternating* to *pulsating direct currents. Detectors* vary in type and efficiency, ranging from the now obsolete *coherer,* to the modern sensitive *regenerative vacuum tube.* The different types of detectors are taken up under their various headings. (See *Crystal Detector, Vacuum Tubes, Electrolytic Detector.*)

DETECTOR CIRCUIT—That part of the circuit in a radio receiver which contains the *detector* (q.v.). The *detector circuit* may be closely coupled to the balance of the receiving circuit by means of a tuning coil or it may be inductively coupled by using a variocoupler or loose coupler. (See *Coupling*).

DETECTOR, DAMPED WAVE—Any special type of detector, such as a crystal, etc., used for reception of *damped waves.* (q.v.). (See *Tikker,* also *Heterodyne,* and *Vacuum Tube.*)

DETECTOR, VACUUM TUBE—The *vacuum tube* used as a *detector* of *radio frequency* (*high frequency*) *oscillations.* A vacuum tube may be used to change or rectify the high frequency alternating currents received at the aerial to *pulsating direct currents* (q.v.) capable of operating a telephone receiver or recording device. Formerly vacuum tubes were made for an individual purpose, such as *radio frequency amplifier, audio frequency amplifier, or detector,* the amplifier tubes being highly exhausted and the detector tube having a low vacuum and requiring critical filament adjustment. Of late tubes have been developed to such an extent that no special tube is necessary for the detector circuit, all being adaptable for almost any purpose. Fig. 1 shows the ordinary circuit for a vacuum tube detector. This method is not much used now since the advent of regeneration, except where stages of radio frequency amplification are used ahead of the detector to amplify the incoming signals. In this case incoming signals are impressed on the grid "G" and the slight variations of potential at this point control comparatively large currents at the plate "P," actuating the telephone "T." (See *Vacuum Tube, Theory of Detector Action.*)

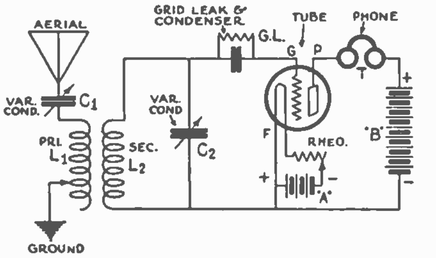

Fig 1. Non-regenerative vacuum tube circuit.

Used in this manner the tube is a considerable improvement over the crystal detector in point of sensitivity. However, by means of regeneration, the sensitivity may be increased many times. Fig. 2 shows a simple *regenerative circuit.* Here the unrectified currents passed through to the *plate* of the vacuum tube are fed back by means of the coil L2 and by being impressed again on the *grid,* further amplification can be attained. (For complete explanation see *Regenerative, Theory of Operation of Vacuum Tube, Detector, Pulsating Direct Current,* also *Crystal Detector.*)

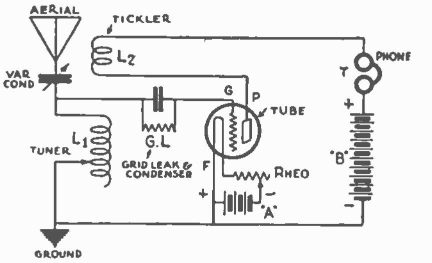

Fig. 2. Simple regenerative circuit.

DETUNING—The opposite of *tuning* (q.v.) The process of varying the effective inductance or capacity or both, to throw the radio receiver out of resonance with the particular signals to which it is tuned. This may be done to decrease the volume of the signals or it may be employed to eliminate or reduce interference from some undesired signals. For example, a station operating with a wavelength of 400 meters might interfere with reception of a station operating on 360 meters, providing it has sufficient power, or the receiver is not particularly selective. If the radio receiver is detuned slightly below 360 meters, the desired signals will still come in, though at lesser volume, but the interfering signals will be lost entirely. Tuning is the act of *producing* resonance and detuning is the act of *destroying* resonance. In the case of a *heterodyne receiver* (q.v.) the heterodyne circuit is slightly detuned from the incoming oscillations, thus producing a difference in frequency or *beat frequency* (q.v.). (See *Resonance,* also *Dead-space.*)

DIAGRAM—A system of lines drawn to represent the circuit or connections for radio receivers or transmitters or associated apparatus, etc. Diagrams may be in any of several forms, the two common methods being known as schematic and perspective. In the schematic form, symbols are generally used to represent the various pieces of apparatus, whereas in the perspective form, drawings of the apparatus are used. Fig. 1 illustrates a typical schematic diagram of a simple crystal receiver. Here the tuning coils, de-

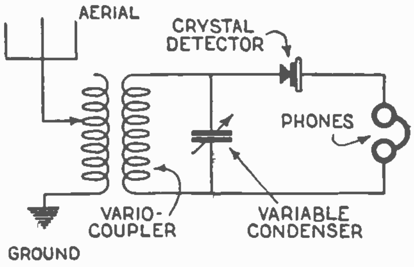

Fig. 1. Schematic diagram of a simple crystal receiver.

tector, phones and so on, are shown in the form of symbols. Fig. 2 shows the same circuit in perspective form, with the tuning coil and other parts actually pictured. Diagrams may be used to show connections for any electrical or radio circuits or for individual parts of instruments. (See *Hookup*.)

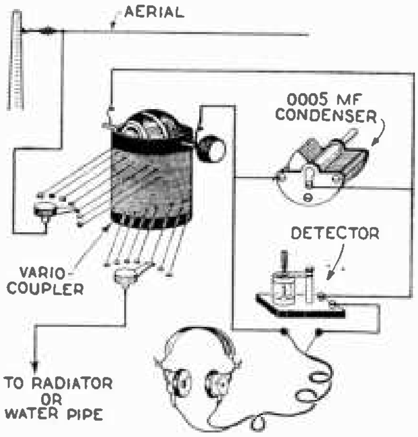

Fig. 2. Showing in perspective form the circuit given in Fig. 1.

DIAL—Devices used to control the movement of condensers or other moving parts of radio apparatus. Most dials are made of a composition such as Bakelite or hard rubber; some are made of metal. The periphery is generally spaced off in numbers from 0 to 100 or 180, for purposes of keeping a record of the positions at which the various stations are received. Dials may be arranged to have the readings refer to definite values by means of *calibration* (q.v.). (See *Dial Vernier*. See *Log*.)

DIAL, VERNIER—A term incorrectly applied to a dial arranged with slow-motion linkage in order to permit very fine adjustment. The illustrations show two types of vernier dials. Vernier dials are generally used on *variable condensers* and occasionally on other instruments, where close adjustment is necessary. They are particularly useful in the case of a *Super-Heterodyne receiver*, the oscillator control being critical, and the wavelength control as well, in certain cases. (See *Dial*, also *Critical*.)

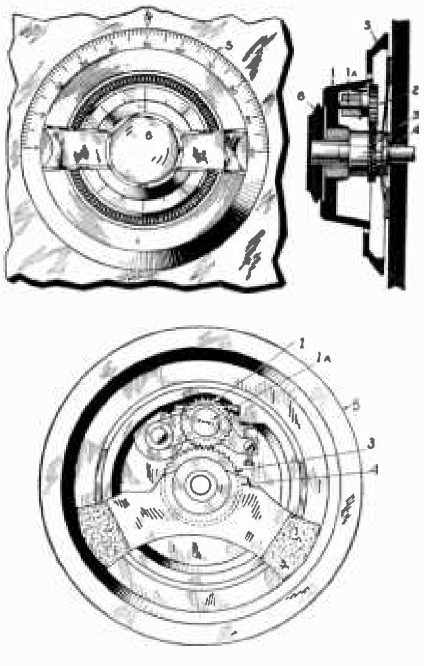

This dial utilizes the slip tooth principle. 5 is the main dial. 6 turns the slip gears (3) and (4), which cause 2 to turn. 2 is mounted on the shaft (1A) fastened in the main dial (1). 5 is the scale, fastened to the condenser shaft.

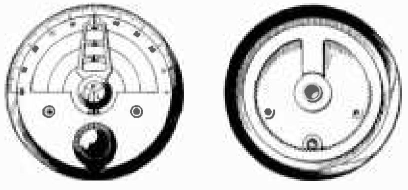

A simple vernier dial using a pinion and internal rack.

DIAMAGNETIC MATERIAL—Substances not readily susceptible to *magnetism*. If a substance does not pass magnetic lines of force, but rather is repelled by a magnetic pole, it is said to be *diamagnetic;* its permeability, or ability to permit magnetism to pass through it, is considered negative or less than unity (air). Diamagnetic substances are apparently repelled from the poles of a magnet. If placed in a magnetic field they have a tendency to diminish the *magnetic induction* (q.v.). (See *Paramagnetic*, also *Permeability*.)

DIAPHRAGM—A thin disk, generally of soft iron used in *telephone receivers* and *microphones* (q.v.) to either produce or detect electrical pulsations by means of vibrations caused by an *electromagnet*. Variations in the current through the electromagnet cause corresponding vibrations in the diaphragm. While the common type of diaphragm is made of soft iron, there are numerous other materials used. Certain alloys have higher permeability, and therefore are more responsive to the attraction of the magnet. In one type of telephone receiver that is particularly sensitive, there are two diaphragms—one that is actuated directly by the electromagnet under the influence of a varying current and another of thin mica, coupled to it by a fine wire. Another type employs a small iron armature suspended on a pivot, and attached to a mica diaphragm by means of a delicate lever, at right angles to the armature. The vibrations of the armature in the magnetic field of the magnets are communicated to the diaphragm by means of this lever. (See *Baldwin Receiver*.) Another type is of metal with fine corrugations and still another is a mica or parchment disk with a piece of soft iron in the center. (See *Microphone*, *Telephone Receivers*.)

DIAPHRAGM, CARBON—A phone diaphragm in the form of a thin carbon disc. (See *Diaphragm*.)

DIELECTRIC—This term is rather broadly used in electrical work to indicate a non-conductor or *insulator*. A dielectric may be a solid, liquid or gas and is generally employed to separate two conducting surfaces. The most common instance of the application of a *dielectric* is in the case of a condenser. Here the plates are separated by a dielectric material, the nature of this material depending on the particular type and purpose of condenser. A *variable condenser* for receiving purposes generally employs air as the dielectric. That is, the two sets of plates are kept from actual contact while one is being rotated, by an air space. In *fixed condensers* the dielectric is usually waxed paper, mica, glass, oil or compressed air, depending on the purpose for which the condenser is to be used and the dielectric strength necessary. While we refer to a dielectric as a non-conductor or insulator, there is actually no dividing line between conductors and insulators, practically any dielectric permitting passage of a certain amount of current under the proper conditions. If a substance is a very good insulator it is said to have good dielectric properties, while one that is a poor insulator and permits leakage of current is a correspondingly poor dielectric. (See *Insulator*, *Conductor*, *Resistance*, also *Dielectric Coefficient and Constant*.)

DIELECTRIC ABSORPTION—The tendency of the dielectric in a condenser to apparently absorb a certain amount of the power applied to it. When a condenser is charged by a direct current source such as a battery, the instantaneous charge is often followed by a small charge or flow of current that steadily decreases as it flows into the condenser. This additional charge is in effect absorbed by the dielectric of the condenser. The reverse effect is obtained when the condenser is discharged, the instantaneous discharge being followed by a steadily decreasing additional discharge. It will be apparent, then, that if the condenser is charged from an *alternating current* source, such as by *radio frequency currents*, the dielectric will show a tendency to withhold a certain part of the charge, and as the charging and discharging goes on rapidly due to the rapid reversals of the current (see *High Frequency Alternating Current*), the dielectric will continue to hold back this portion of the power as long as the charging and discharging process persists.

This loss, due to the absorption in the dielectric, must not be confused with the losses due to faulty dielectric materials, wherein a certain portion of the current is allowed to leak away by conductance through or on the surface of the insulator. *Dielectric absorption* is often referred to as *dielectric viscosity* or *hysteresis* because of its similarity to viscosity in liquids. Viscosity is the property possessed by liquids to resist deformation. The usual method of measuring viscosities is by measuring the time taken by a known volume of the liquid, at a known temperature, in flowing through an aperture of known form and dimensions under a known pressure. Thus tested, water will flow rapidly, while cylinder oil is very sluggish, and hence is said to possess great viscosity. (See *Condenser*, also *Dielectric*.)

DIELECTRIC COEFFICIENT and CONSTANT—The specific *inductive capacity* of a dielectric. Generally speaking, its properties to act as a *dielectric*. In the *Centimeter Gram Second system* (q.v.) the *inductive capacity* and the *dielectric constant* are numerically equal, the constant being the dielectric value of the material as compared with air at ordinary pressure taken as the standard (1). The table of constants for the more important dielectric materials follows.

It will be seen that glass, oils and mica have the highest values, for which reason they are widely used. (See *Specific Inductive Capacity, Dielectric*, also *C. G. S.*)

TABLE OF DIELECTRIC CONSTANTS

Dielectric	*Constant*
Air at ordinary Pressure (Taken as the Standard)	1.000
Manila Paper	1.50
Celluloid	1.555
Parrafine (clear)	1.68 to 2.32
Beeswax	1.86
Parrafine Wax	1.9936 to 2.32
Parrafined Paper	3.65
Hard Rubber (Enonite)	2.05 to 3.15
India Rubber (pure)	2.22 to 2.50
Gutta Percha	2.46 to 4.20
Shellac	2.74 to 3.60
Olive Oil	3.00 to 3.16
Glass (Low Frequency value)	3.25 to 4.00
Glass (High Frequency value)	4.21
Mica (Pure Sheet)	4.00 to 8.00
Porcelain	4.38
Castor Oil	4.80
Flint Glass, very light	6.57
Flint Glass, light	6.85
Flint Glass, very dense	7.40
Flint Glass, double extra dense	10.10

DIELECTRIC CONSTANTS—See *Table of Constants* under *"Dielectric Coefficient* and *Constants."*

DIELECTRIC HYSTERESIS—When the *electric field* in a dielectric material is varied rapidly, as when the condenser is charged by *high frequency currents*, heat may be generated. This is due to *dielectric hysteresis which is* synonymous with *dielectric absorption* (q.v.)

DIELECTRIC STRENGTH—When an electric field is established in a dielectric, that is to say, a charge is applied to a condenser, and the field attains a certain intensity, the dielectric ceases to be an insulator and becomes, in effect, a conductor. This condition generally is accompanied by a puncturing of the dielectric material. When the charge applied to a condenser is too high the spark will burn through the dielectric and the condenser is then said to have broken down. In some cases the condenser may be permanently ruined by this action, and in others, the dielectric may be conductive at a certain voltage, while remaining effective as long as the voltage is held below the safety point. The particular voltage (critical voltage) or field intensity at which the breakdown of the dielectric occurs is called the dielectric strength of the material. (See *Dielectric*, also *Table of Dielectric Strength.*)

DIFFERENCE FREQUENCY—The frequency of oscillations produced by superposing oscillations of one frequency on oscillations of a different frequency. In Super-heterodyne receivers the action is based on the production of a difference or beat frequency. If the incoming oscillations from the antenna are combined with oscillations of a different frequency produced locally in the receiver, a new series of oscillations will be produced, these oscillations having a frequency numerically equal to the difference in frequency between the other two sets of oscillations. (See *Beat Frequency*, also *Super-Heterodyne.*)

DIODE—A thermionic vacuum tube having only two *electrodes*, namely *filament* and *plate*. The original vacuum tubes were known as *Fleming Valves* and contained a hot filament and cold plate but no grid as found in the three element vacuum tube now in general use. Several types of *diode* are still in use as detectors, their chief value lying in the fact that they require no careful adjustments as in the case of a *crystal detector.* (See *Triode, Vacuum Tube*, also *Fleming Valve.*)

DIPLEX RECEPTION OR TRANSMISSION — The simultaneous reception or transmission of two series of signals by or from a single operating station. The systems are so arranged that two messages may be sent or received at the same time without interfering with each other. (See *Duplex Signalling.*)

DIRECT CONDUCTIVE CIRCUIT—Any circuit having a direct conductive path and not depending on capacity or electromagnetic coupling. A metallic, conducting circuit.

DIRECT COUPLING—The coupling or relation between two or more coils or circuits, wherein the connection is *metallic*. In the illustration Fig. 1 is shown a two slide tuning coil. In this case A-B is the primary or aerial circuit and C-D the secondary or detector circuit. The two circuits are joined together by metallic connection.

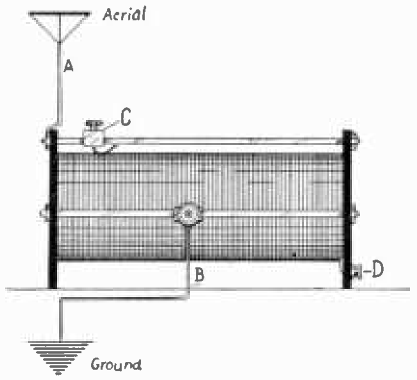

Fig. 1. Double slide tuner which affords direct coupling between the primary and secondary circuits.

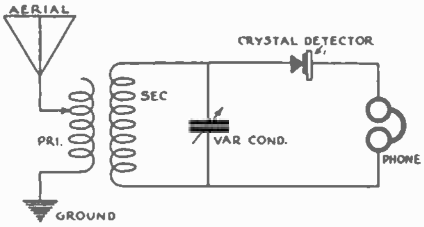

Fig. 2. Circuit in which primary and secondary are inductively coupled.

The illustration Fig. 2 shows essentially the same circuit, except that the primary and secondary are two separate coils. Here the two coils or circuits are joined by *inductive* coupling.

DIRECT COUPLED—Circuit in which primary and secondary circuits are metallically connected. Generally permits little selectivity. For this reason, *inductive coupling* is more generally used to permit close tuning or selectivity (q.v.). (See *Coupling, Inductive Coupling.*)

DIRECT CURRENT—See *Current, Direct.*

DIRECT READING—A term applied to various measuring or recording instruments which are so arranged as to show values directly without requiring mathematical reduction. *Voltmeters, Ammeters* and such instruments, used widely for showing the *voltage, current*, etc., of *batteries* or *cells*, are direct reading, showing the values directly in *volts* or *amperes*. A *decremeter* or *wavemeter* very often shows a reading that requires application of a formula to produce the desired quantity. (See *Calibration.*)

DIRECTION FINDER—See *Radio Compass.*

DIRECTIONAL—The effect of an aerial wherein waves are transmitted better, or entirely, in a certain direction, or received from a certain direction, depending upon the type and direction of aerial. (See *Transmitting Aerial*, also *Receiving Aerial.*)

DIRECTIONAL EFFECT OF ANTENNA—An effect of an antenna (aerial) for reception or transmission wherein signals from a certain direction are more readily received than from another direction, or in the case of transmission, the range being greater in a certain direction.

DISCHARGE—Generally speaking, a comparatively sudden passage of electricity. The term however, as applied to a *storage battery* will mean merely the effect of releasing the *electrical charge* stored up in it, in this case the discharge not being necessarily rapid. A *condenser* has the ability to hold charges of electrical energy and to release them suddenly when the proper contact is made. The term "discharge" as applied to a *transmitter* indicates the passage, usually in the form of a spark or succession of sparks, of the electrical energy across a gap between *electrodes*. The illustration shows spark discharge from spark coil. (See *Spark Discharge, Storage Battery*, also *Discharger.*)

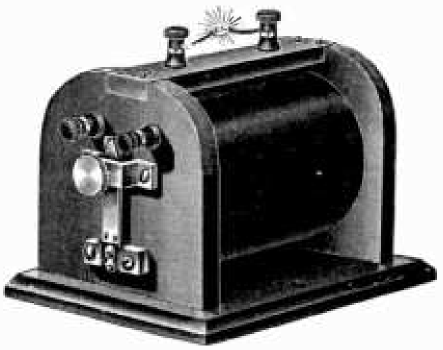

Discharge from a spark coil.

DISCHARGER—Any device allowing a path for an electrical discharge. The term may be considered roughly a synonym for *spark gap*. Usually two or more *electrodes*, either stationary or rotary, spaced a short distance apart to permit the released electrical energy to bridge the space and thus complete an *oscillatory circuit*. (See *Disc Discharger Rotary Gap.*)

DISC DISCHARGER—A form of *discharger* employing one or two rotating discs carrying the sparking surfaces. With the advent of modern undamped transmission methods, these dischargers are gradually passing out. (See *Synchronous Discharger.*)

DISPLACEMENT CURRENT—A current which flows for a short interval in an insulating material or *dielectric* when an *electromotive force* is impressed across it, or when the intensity of the electromotive force impressed across it is increased or decreased. This *displacement current* will flow only when the impressed electromotive force is altered in intensity. After the initial current due to the sudden change in the electromotive force being impressed across the material, the dielectric material or insulating material will remain in a state of strain as long as the *charging force* persists without further change in intensity, and no further displacement current will flow. (See *Current*, also *Dielectric.*)

DISPLACEMENT, PHASE—See *Phase Displacement*, also *Phase Angle.*

DISRUPTIVE VOLTAGE—The voltage sufficient to disrupt or break down a sample of *dielectric* material under given conditions. Known also as *breakdown potential.* For example, if a *condenser* will stand an impressed potential or voltage of 500 volts without injury, but will be punctured by a potential of 1000 volts minimum, the disruptive voltage is said to be 1000. (See *Break Down Potential.*)

DISSOCIATION THEORY—A theory advanced by Arrhenius in 1887, explaining *electrolytic conduction* by the assumption that a substance in solution is dissociated or separated into *positive and negative ions*, these ions carrying their respective charges in opposite directions. (See *Electrolytic Action.*)

DISSONANCE—"Discord; disagreement"—Webster. The antonym of *resonance.* A term broadly applied in radio to indicate lack of resonance, or coordination of signals or impulses. When an alternating current is superimposed on another alternating current of different frequency, the resulting lack of resonance—*dissonance*—is known as a *beat.* (See *Distortion.*)

DISTORTION—Lack of purity or faithfulness in the reproduction of a vibration or series of vibrations. The most common application of the term in broadcast reception is in the case of reproduction by a loud-speaker. It will very often be found that music or speech is not perfectly reproduced, due to any of a variety of causes. Vacuum tubes may themselves cause distortion; too high potential applied to the plate of a tube may result in distortion; many transformers used in audio frequency amplifier circuits may distort notes of certain frequencies, or it may be due to *self-oscillation* or *regeneration* in the receiving set. This effect is also caused by inefficient transmitting (broadcasting) apparatus, or by lack of careful adjustment. The transformation of speech or music into electrical impulses and its subsequent transmission and reception are attended by many difficulties.

The control of wave form in broadcasting is known as *modulation.* (q.v.). If the transformation of speech or music into electrical impulses and its propagation into space is accomplished without materially changing or distorting the wave form from the original voice or music vibrations, the problem of accurate reproduction is entirely dependent on the receiving apparatus. Distortion may often be traced to the diaphragm of the loud-speaker, certain types being more efficient in this respect than others. (See *Loud-speakers.*) If the ordinary type of disc diaphragm is used, it may become bent and thus cause inequalities in the vibrations, producing distortion of the music or speech being reproduced. (See *Wave Form, Modulation,* also *Amplifier.*)

DISTRESS SIGNAL—At the International Radio Telegraph convention held at Berlin, Germany, in July, 1908, the call letters C Q D, established by the Marconi Co. in February, 1904, as the official distress call, were superseded by the letters S O S as the Marine distress signal.

S O S is the International distress call for ships and airships requiring assistance. The letters have no particular significance, being chosen mainly for their distinctive sound. In the International Morse code, S O S is composed of three dots, three dashes and three dots, thus: . . . — — — . . . an unusual combination which permits easy recognition among other messages and calls. When distress signals are heard, the nearest government transmitting station generally sends out immediate notification to all broadcasting stations in the vicinity to suspend operations until the ship has been located and assistance rendered.

DISTRIBUTED CAPACITY—The condenser effect in a coil of wire. Any coil of wire possesses *inherent capacity* to a certain extent, depending on its particular shape, size, etc. In the case of a coil of wire wound on a cylindrical form as indicated in the illustration Fig. 1, the passage of current through the coil sets up a *local magnetic field*, but also an *electrostatic field*, the latter being in the form of lines of force perpendicular to the conductor. Thus in effect, each two adjacent turns of wire act as a miniature condenser (Fig. 2).

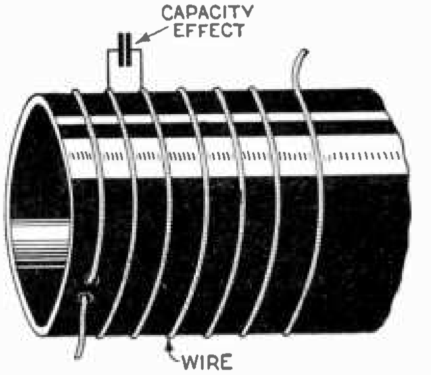

Fig. 1. Coil showing effect of distributed capacity between winding.

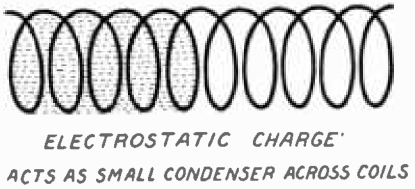

Fig. 2.

There are numerous methods and forms for winding coils in order to do away as much as possible with this effect. (See *Low-Loss Coils.*)

DISTRIBUTED INDUCTANCE—In a long cable or any great length of wire used for electrical purposes, additional inductance (wire) distributed throughout the entire length to compensate for the inherent capacity of the line. (See *Distributed Capacity.*)

DOT—The short signal in the International Morse Code. (See *Code.*)

DOUBLE AMPLIFICATION CIRCUIT—Also known as *dual amplification circuit.* The arrangement whereby a vacuum tube or several tubes may be made to do double duty, acting as both *radio frequency* and *audio frequency amplifiers.* (See *Reflex.*)

DOUBLE COTTON COVERED WIRE—Abbreviation D. C. C.—Cotton covered copper wire widely used in radio work for coils and all forms of inductances. The insulating covering is composed of two distinct layers of cotton, wound on in opposite directions to prevent loosening. Such wire is obtainable in a variety of sizes according to the standard gauges. (See *Wire Gauge.*)

DOUBLE FREQUENCY OSCILLATIONS—Sometimes referred to as a *double humped wave.* An irregular wave resulting from two frequencies, generally due to too close coupling of the *oscillation transformer.* When the coupling between the open and *closed oscillatory circuits* is too tight, the *open circuit* very often oscillates at two frequencies, resulting in the radiation of an irregular or *impure wave.* The illustration Fig. 1 shows in graphic form, such an irregular wave. The two peaks of resonance explain the term double humped wave. Fig. 2 shows another curve illustrating two frequencies.

In this case however, the amplitude of the second wave is considerably less than that of the first or main wave. According to U. S. Government regulations, if the amplitude of the lesser wave is not more than 0.1 of the amplitude of the main wave, it is said to be pure. In other words, where there are two frequencies, but the strength of one less than one tenth that of the other, it is considered negligible as it cannot be expected to cause any appreciable interference. (See *Resonance, Amplitude, Decrement,* etc.)

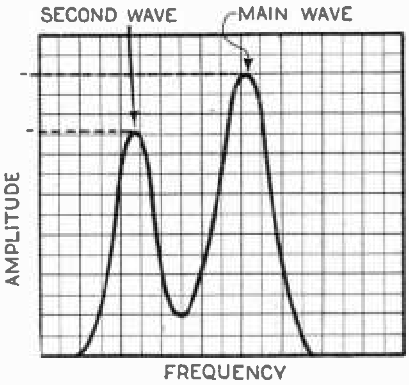

Fig. 1. Resonance curve showing second wave of interfering nature.

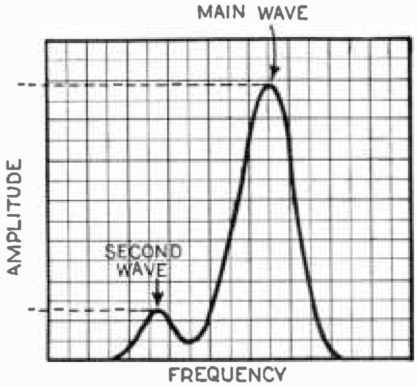

Fig. 2. Resonance curve showing second wave of negligible proportion.

DOUBLE GRID TUBE—A vacuum tube having two distinct *grid* members in addition to the usual *filament* and *plate* elements. Such a tube is used in circuits where the customary "B" battery is not employed, the extra grid acting as a booster for the electron flow, the "A" battery furnishing a small positive potential to the plate. (See *Solodyne,* also *Theory of Vacuum Tubes.*)

DOUBLE MODULATION—Successive modulation of a *radio frequency alternating current* at two lower frequencies. The *intermediate frequency* is usually above *audio frequency* and the lowest frequency is customarily within audio limits or a combination of audio frequencies, as in the case of *radio* *telephony* (broadcast). (See *Modulation.*)

DOUBLE POLE SWITCH—A switch used in electrical practice and in radio installations having two poles or connections, thus permitting both sides of a circuit to be opened or closed simultaneously.

DOUBLE RANGE METERS—Meters used for electrical measurements, arranged to read to two scales. As an example, the voltmeter shown has two scale readings, one in fine degrees for a maximum of 7.5 volts, the other a coarser scale to a maximum of 150 volts. Three binding post connectors are furnished, one being a common positive for both scales. A meter for almost any measurement might be made to have double range, although such instruments are generally confined to measurement of current in amperes and pressure in volts—i. e., ammeters and voltmeters. (See *Meter.*)

Photo by Courtesy of Weston Electrical Instrument Co.

Double range meter, one range to 7½ volts, higher range to 140 volts maximum.

DOUBLE SLIDE TUNER—A tuning coil provided with two sliding contacts, generally used in crystal receivers.

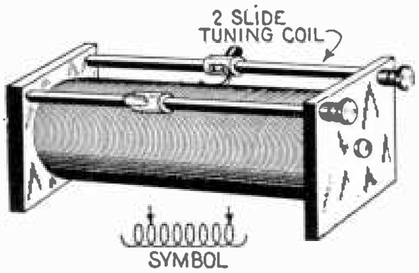

Two-slide tuner.

The illustration shows a common type of two slide tuner. (See *Coupling Crystal Receiver*, also *Tuning.*)

DOUBLE THROW SWITCH—A switch so arranged that a circuit or a certain instrument is connected in either of two different positions by throwing the switch lever. The illustration shows a single pole double throw switch. (See *Switch.*)

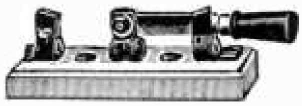

Double throw single pole switch.

DOWN LEAD—More commonly known as *lead-in.* The wire running from the aerial to the receiving or transmitting apparatus, whereby the signal energy is either collected from, or fed to, the aerial. The illustration shows a typical down lead or lead-in for a receiving set. Obviously the chief difference between such a lead for receiving purposes and for connecting a transmitter to the antenna is the matter of insulation. The comparatively feeble impulses received or collected by the aerial when receiving require only ordinary care, whereas the high voltages used in transmitting require careful insulation and much heavier apparatus all around. (*For more complete details regarding insulation, etc., see Aerial.*)

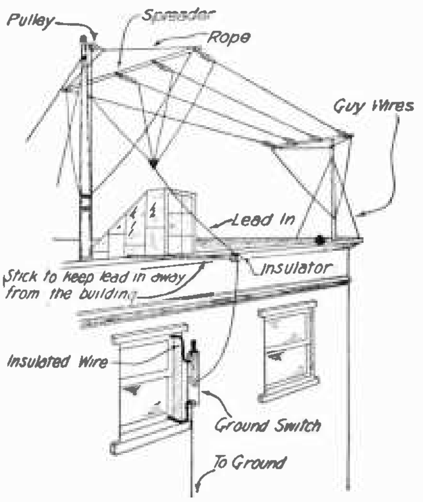

A typical 4-wire aerial showing lead in and method of installing lightning switch.

DRIFT, AVERAGE ELECTRON—The assumed rate of flow or drift of *electrons* under average or specific conditions. (See *Electronic Flow.* Also *Current, Assumed Direction of Flow.*)

DRIVER—A term broadly used to denote any system used to produce *oscillations* (vibrations) of a local nature. More specifically any means of producing oscillatory currents as used to test or make measurements in radio circuits. A buzzer connected inductively to a circuit to produce oscillations in that circuit for the purpose of making measurements of *capacity, wavelength,* etc. Another means might be a circuit involving a vacuum tube and the necessary apparatus to make it oscillate at a given or variable frequency. The term may be used to signify the tube circuit used to produce local oscillations in the case of a *super-heterodyne* (q.v.). Here the driver produces a series of oscillations of a frequency different from the incoming oscillations, the difference between the two being known as the beat frequency. (See *Heterodyne, Local Oscillations, Buzzer, Excitor.*)

DRIVER CIRCUIT—The circuit of the apparatus used to produce oscillations for purposes of test or measurement, or, in the case of a super-heterodyne, to produce a beat effect. (See *Driver.*)

DRUM ARMATURE—A form of armature winding in the approximate shape of a drum. (See *Armature.*)

DRY BATTERY—A battery or group of cells not employing a liquid electrolyte, the cell being filled with a mixture of carbon, manganese dioxide and sawdust (or other absorbent) saturated with a solution of sal ammoniac, these forming a paste, as distinguished from Wet Battery. (See *Battery,* also *Cell* and *B Battery.*)

DRY CELL—See *Dry Battery.*

D. S. C.—The abbreviation for "Double Silk Covered" as applied to copper wire having two distinct layers of silk, wound generally in opposite directions. This wire is furnished in standard sizes according to the various wire gauges. (See *Brown & Sharpe Gauge.*)

DUAL AMPLIFICATION—The process of obtaining both *radio frequency* and *audio frequency amplification* from the same tube instead of using two separate tubes. (See *Double Amplification Circuit,* also *Reflex.*)

DUBILIER, WILLIAM—President and technical director; born, New York, July 25, 1888. Educated New York Schools, Technical Inst. and Cooper Union; Chief Engineer of Continental Wireless Tel. & Tel. Co.; 1910 President of Com. Wireless Tel. Co.; at present Technical Director of Dubilier Condenser and Radio Corp. of New York, The Dubilier Condenser Company Ltd. of London, The Deutsche Dubilier Kondensator Gesellschaft in Berlin, La Protection Electrique Capart-Dubilier in Paris. Has obtained over three hundred patents and applications of electrical devices, which have been purchased or licensed by many companies. Member American Institute Electrical Engineers, Inst. of Radio Engineers, Honorary Member of Societe Academique D'Historie Internationale. Member of Royal Society of Arts.

DUCON—Trade name of a device which can be fitted to any electric light socket and made to serve in place of the usual outdoor aerial. This attachment consists essentially of two condensers, so arranged as to prevent the passage of any direct current or low frequency alternating current from the lighting main to the receiving set, but at the same time to permit passage of the incoming radio signals, the electric light line acting, therefore, as an aerial. (See *Adapter, Aerial.*)

DUDDELL SINGING ARC—Also called *Musical arc.* An *arc* actuated by a source of direct current through a *resistance* and shunted by a *condenser* and *inductance* in series. An *oscillating current* is thus produced in the condenser circuit, the result being a singing note corresponding in pitch to the frequency of the oscillations in the condenser circuit. The schematic arrangement is shown by the illustration. (See *Arc Generator.*)

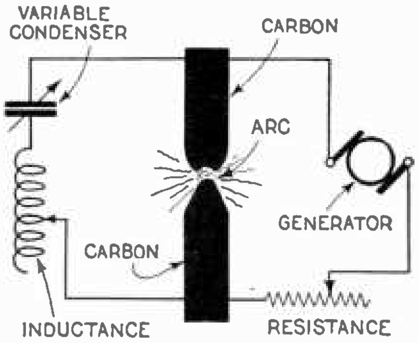

Schematic arrangement for operation of Duddell singing arc.

DULL EMITTER—The English term for vacuum tubes having *thoriated filaments* and operating with low current consumption. While the term is generally used to denote the tubes requiring only one or two dry cells to operate, it may apply as well to storage battery tubes where the current requirements are low. (See *Filament, Thoriated.*)

DUOLATERAL COILS—An alternate term for *honey-comb* coils; inductances wound in diamond-shaped layers to reduce the distributed capacity effect. (See *Honey-Comb Coils.*)

DUPLEX SIGNALLING or DUPLEX TELEGRAPHY OR TELEPHONY—The simultaneous transmission of signals or telephony in both directions between two stations. (See *Diplex Reception or Transmission*, also *Telephony.*)

DUST CORE—A form of core for certain types of transformers, employing iron filings or dust in place of the customary iron wire or laminations. Iron dust is used occasionally where simplicity of construction is desired, the assembly of such a core being obviously much easier than with wire or laminated types. (See *Core.*)

DX—The popular term for long distance; referring to the transmission to or reception from distant points of radio signals or broadcasting.

DYNAMIC CHARACTERISTIC—The curves obtained by impressing an alternating E. M. F. (electromotive force) on the grid of a vacuum tube, as distinguished from the curves obtained by application of a steady direct potential. When alternating potentials such as incoming signals are impressed on the grid of a tube, the curve resulting is apt to take an entirely different form from that of the curve showing *static characteristics* (q.v.). (See *Vacuum Tube characteristics.*)

DYNAMIC CONDENSER—A term occasionally applied to a *synchronous motor* used to improve power factor (in alternating current circuits, the ratio of the electric power in *watts* to the apparent power in *volt amperes* is known as the power factor). Used in this manner the motor has the effect of advancing the current phase in a manner similar to that of a condenser. (See *Angle of Lead or Lag.*)

DYNAMIC ELECTRICITY—A term sometimes used for electric currents to distinguish them from *static electricity* (q.v.).

DYNAMO—A machine for converting mechanical energy into electrical energy. Is also known as *generator*, but usually confined to machines for generating *direct* current. The action of a dynamo is based on the production of an *electromotive force* in a conductor moving in a magnetic field. Fundamentally a dynamo is a machine that generates alternating current, but instead of collecting this current by means of rings, which would cause alternating current to flow in the external circuit, a device known as an *armature* is employed in such a manner that the current collected by brushes and delivered to the external circuit is direct—that is, it flows more or less steadily in one direction. Thus the kind of current delivered for use depends on the method used to collect the currents generated. (For more complete explanation of the production of alternating current see *Alternating Current, Theory of Production.*) In the case of a dynamo, the essential parts are the *field magnets*, the *armature* and the *commutator*. Now in the case of an *alternator*, each end of the loop or armature is connected by a brush bearing against a collector ring and the end of each loop is always connected to the same brush. This *results in an external current that changes its direction with each reversal of direction of the induced electromotive force.* If the external current is to be direct it is necessary to have some means of collecting the current through one brush at the positive instant and the other at the negative instant. This is done by means of an armature, which is in effect a switching arrangement, so designed that it will reverse the connections of the external circuit at the instant of each reversal of current in the armature. The illustration shows a *shunt* wound dynamo. MM are the field magnets, FF the field windings, BB the brushes, A the armature and C the commutator. (See *Alternator, Armature, Generator.*)

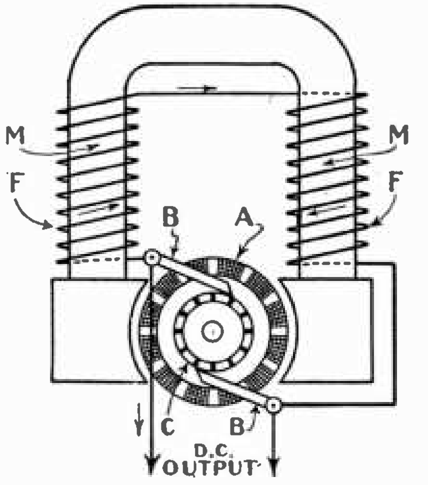

Shunt wound dynamo.

DYNAMOMETER—A delicate and accurate instrument for the measurement of currents and voltages or both. Such instruments depend upon the action of a circuit carrying current upon another carrying the same current. Essentially it comprises two coils, one fixed and the other movable. This system of measurement is used in a *wattmeter*, where it is necessary to measure the instantaneous current and voltage, i.e., power. (See *Wattmeter*, also *Electro Dynamometer.*)

DYNAMOTOR—A direct current machine which acts either as a *motor* or *dynamo*. It has an armature with two separate windings and two separate commutators, one at each end of the armature. Either winding may be used as a motor and the other as the generator winding. Such a machine is used to convert high voltage direct current into low voltage direct current or vice versa, thus performing the same function with direct current as a power transformer performs for alternating current. (See *Converter.*)

DYNATRON—A form of vacuum tube generally used for producing oscillations as in radio telephony, wherein the phenomena of "*secondary electron emission*" is used. In this form of tube the electrons traveling at high speed from the hot filament or cathode are made to collide with a metallic surface. The collision of these electrons on the surface of the interposed element has the effect, under proper conditions, of jarring other electrons out of the metal. The secondary emission thus obtained depends upon the speed of the original electrons which collide with the metal surface. Normally, these secondary electrons will immediately re-enter the surface from which they were emitted, but if another electrode of higher potential is in the vicinity they will travel toward it in the same manner as the electrons are attracted to the plate of an ordinary three element tube. A standard tube can be connected as shown in the illustration to act as a dynatron. Here the filament F emits electrons, some of which pass through the grid G and collide with the plate P. This collision with the plate P may jar loose from the surface additional electrons, which normally would immediately re-enter P. However, as the grid G is held at a higher potential (higher positive voltage) than P, the electrons will be drawn toward it. A tube arranged in this manner may be used for practically any of the purposes of the standard tube, such as regenerative detector, detector of continuous waves or as a generator of high frequency oscillations, but it has not been shown to have any advantages in this respect, being used mainly as an *oscillator* (q.v.). (See *Electron Emission, Vacuum Tube*, etc.)

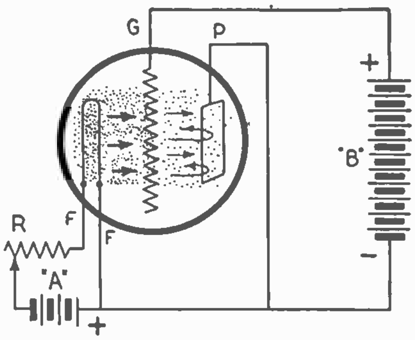

Showing how a standard vacuum tube can be connected to act as a dynatron.

DYNE—The unit of force in the absolute or CGS system of units. It is defined as the force which, acting on a mass of one gram for one second, will impart to the mass a velocity of one centimeter per second. (See *CGS System*, also *Force.*)

E

E—Common symbol for *Electromotive force* (q.v.).

EAR CUSHION—Pads or cushions of soft rubber used in conjunction with head-phones to prevent unpleasant pressure against the ears and also to exclude outside sounds when listening in.

EARTH—An alternate term for *ground*, where the earth or any metallic connection thereto is used as a return in transmission or reception of *electromagnetic waves*. (See *Ground.*)

EARTH CURRENTS—See *Ground Currents.*

EARTH, DEAD—See *Ground.*

EARTH DETECTOR—See *Ground Detector.*

EBONITE—See *Insulating Materials.*

EBURIN—An insulating compound used for *strain insulators* (q.v.). (See *Insulating Materials.*)

EDDY CURRENT LOSS—The portion

of the total loss in electrical apparatus due to *Eddy Currents* (q.v.).

EDDY CURRENTS—Currents induced in the mass of a solid *conductor* due to the action of a varying *magnetic field.* The most common example of eddy currents will be found in the case of a *dynamo* or *generator*. Here the useful current generated in the *armature* is produced by the motion of the armature in the magnetic field. At the same time currents are generated in the iron core due to its motion in the magnetic field. As all currents generated in an electrical machine are produced at the expense of a certain amount of energy, and since the Eddy currents cannot be gathered or put to useful account in this case, they represent a loss. This loss usually is apparent in the form of heat. If the core is made up of solid metal it will be obvious that a good closed path is offered for eddy currents produced. Now if the core is composed of a number of sheets or *laminations*, insulated by thin layers of paper or by an insulating scale (see *Transformer Steel*) the path for the eddy currents is partially broken and the effect is reduced. Eddy currents are produced in the core of a transformer due to the same action—the variation of the magnetic field in which the core is located. In order to reduce the effect, laminated cores are employed as in the case of many generating machines. Eddy current loss is often given as the loss in watts per pound of core material at 10,000 *gausses* (q.v.) and 60 cycles for a sheet 0.0141 inch or 0.0358 centimeter thick. While eddy currents are usually productive of losses they are put to useful account in certain types of instruments. For example, the eddy currents produced in the metal frame of a moving coil meter may be employed to *damp* or retard the free *oscillations* of the needle. (See *Damping of Instruments.*) Eddy currents are also used in a form of speed indicator where the reaction of eddy currents created in a moving disk are made to deflect a pivoted, spring controlled, magnetic needle. (See *Core Loss, Hysteresis,* also *Foucault Currents.*)

EDDY CURRENT COEFFICIENT—The coefficient or numerical multiplier, generally termed K_1, used in calculations of eddy current loss. Its numerical value depends upon the *specific resistance* (q.v.) of the iron used in the core of the machine in question, the character or *wave shape* of the induced voltage, the distribution of *magnetic flux* (q.v.) in the core material, the degree of insulation between sheets, where *laminations* are used, and upon the *flux distribution* (q.v.) due to the shape of the magnetic circuit. Its value will also vary according to the units used in the computation.

EDDY CURRENT LOSS, FORMULA FOR CALCULATING—A common formula for calculation of eddy current loss is the following: $P_\epsilon = \epsilon V(xfB)^2$. Here the loss is expressed in watts as P_ϵ, V is the volume of core metal in cubic centimeters, x the thickness of the sheets in centimeters, f the *frequency* in cycles per second, B the *flux density* (q.v.) in *gausses* and ϵ the eddy current coefficient.

EDGE EFFECT—The effect on the capacity of a condenser due to the curving of the lines of stress at the edge of the plates. This effect is not very pronounced where there is considerable *dielectric* surface (insulating material) extending beyond the edge of the plates. (See *Distributed Capacity.*)

EDISON BATTERY—A number of Edison cells grouped together in one case to supply various currents and voltages as in radio. The illustration shows an Edison "B" battery, used to supply potential to the plates of vacuum tubes. (See *"A" Battery, "B" Battery, Plate, Filament,* also *Storage Battery.*)

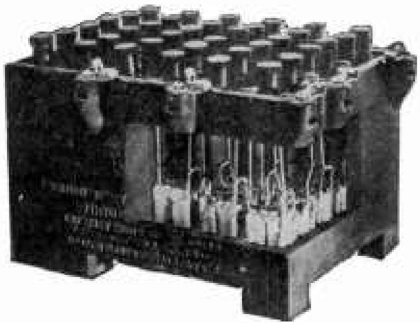

Photo by Courtesy of Edison Storage Battery Co.

Edison "B" battery with portion of case removed to show interior construction.

EDISON CELL—A storage cell employing electrodes of nickeled steel and a solution of potassium hydrate for the electrolyte as distinguished from the usual type of storage cell using lead plates and dilute sulphuric acid as the electrolyte. The chief point of superiority of this type of cell is its comparative ruggedness, due partially to steel construction and also its freedom from the usual ills of cells employing acids and lead plates. The positive electrode in the Edison cell is made by tamping nickel hydrate and alternate layers of pure flake nickel into perforated steel tubes under pressure. The negative electrode is made by pressing iron oxide into flat perforated steel pockets. A number of these pockets are forced into a steel grid to form the complete negative plate. (Note: The *negative* electrode connection on the outside of the cell is referred to as the *positive* pole because it is positive in its relation to the external circuit. Similarly the *positive* electrode connection will be known as the *negative* pole.) The potassium hydrate or hydroxide which takes the place of the acid solution used in the lead plate type of cell acts as a preservative of the steel elements. This naturally means greater life; in fact, it is not subject to the chemical deterioration of other cells. The illustration shows a cell with part of the sides cut away to show the interior construction. (See *Edison Battery.*)

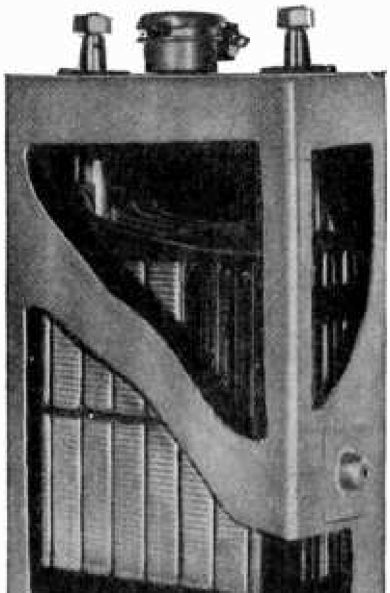

Photo by Courtesy of Edison Storage Battery Co.

Edison cell with part of case cut away to show how constructed.

EDISON, THOMAS ALVA—Born 1847. An inventor famous for his experiments in applied electricity. He began life with newspaper work which he soon abandoned for telegraphy, making many original inventions in duplex systems of operation. After a varied experience in that line he came to New York in 1871, where his talents were recognized and he had opportunity to profitably develop his ideas. The duplex telegraph was made a success the following year, and two years later the quadruplex; and thereupon he began manufacturing on a large scale for the Western Union Telegraph Co. In 1876 he gave up his factory, and established his experimental station at Menlo Park, N. J., where for several years he worked upon the problem of the incandescent electric light, exhibiting a successful bamboo filament lamp in Paris in 1881. He invented the phonograph in 1878. He superintended the construction of the first incandescent lighting station in New York in 1882. Moving his laboratory to Orange, N. J., he established there a large plant for electrical experiment and invention and as a result of his labors there he has taken out 400 patents. Among his inventions may be further named: a type of dynamo, a microphone, the chemical electrical meter, an electric pen, the mimeograph, the magnetic ore separator, dead beat galvanometer, the electric torpedo, a telephone transmitter, and a storage battery. His chief fame rests with his development of the telegraph, his invention of the incandescent lamp and the phonograph.

Thomas A. Edison

EDISON EFFECT—The blackening of the inner surface of an electric light bulb during use. Edison noticed that after the lamp had been burned for some time a coating of black formed on the inner surface and increased in density, becoming finally almost opaque. Edison's experiments, followed by those of *Professor Fleming* (q.v.) resulted in the development of the *Fleming valve* (q.v.)., it having been determined that the black coating was due to the discharge of *electrons* from the hot *filament*. (See *Vacuum Tube, Electron Emission*, etc.)

EFFICIENCY—A very flexible term in electrical practice. Generally speaking, the ratio of useful output of any piece of apparatus to total input. Efficiency is customarily expressed in terms of percentage. Thus, broadly, a piece of apparatus may produce as useful output 80 per cent of the total input. (See *Power Factor*.)

EFFECTIVE ELECTROMOTIVE FORCE—The square root of the mean square of the full alternating current wave. Usually abbreviated R. M. S. (root—mean—square). The effective value of the E. M. F. (electromotive force) is thus taken as the square root of the mean of the squares of the *instantaneous values* over a complete period. (See *Instantaneous Values*.)

EGG INSULATOR—A name applied to a certain type of *strain insulator* due to its egg-like form. The illustration shows such an insulator. The two connections to the insulators are so arranged that the wires will still be looped together in the event that the insulator breaks. (See *Insulator*, also *Insulation of Aerial*.)

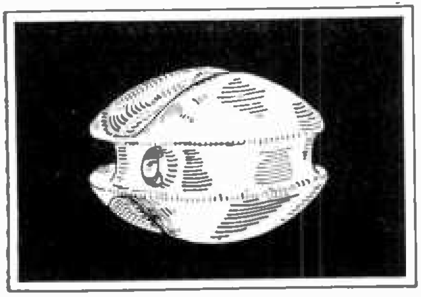

An egg insulator.

ELASTICITY—A property of matter which permits it to resist any change in shape or bulk and permits it after such a change to return to its original state. This property of materials by virtue of which they are enabled to return to their original state after the force causing the distortion has been removed is a very important one in all branches of engineering. Its chief application in radio is in the matter of aerials and aerial guy wires. While the average aerial installation does not represent any serious problems in stress, the large aerial systems of commercial stations must be carefully designed in this respect. For example, a guy wire of one material may have more elasticity than one of similar size in another material. The amount which the wire will stretch under a certain load may be determined by use of a table giving the elasticity of various metals. (See *Electric Elasticity*.)

ELECTRIC ABSORPTION—An effect in *condensers*. That quality of a condenser by means of which it absorbs or "soaks up" a charge of electricity, and conversely, retains part of the charge when the condenser is momentarily discharged. This effect is more apparent in cases where a solid *dielectric* material is used. In fact, *air condensers* (q.v.) are said to have little electric absorption. The reason for this is that mica or other solid dielectrics are strained to a greater extent than air dielectric after the condenser has received a charge. (See *Soaking In*, also *Dielectric Absorption*.)

ELECTRIC DISPLACEMENT—A term proposed by Maxwell to denote a quantity expressing the state or condition of a dielectric in an electric field in accordance with the supposition that such a field is created by the transfer of positive electricity to one end of the field of force and negative electricity to the other end of the so-called *tubes of force* (q.v.). (See *Displacement Current*, also *Electrostatic Induction*, and *Flux Density, Electrostatic*.)

ELECTRIC ELASTICITY—In a dielectric material, the property which permits it to arrest passage of a *displacement current* (q.v.) due to electric stress. It is equivalent to the electric stress divided by the electric strain. (See *Dielectric*, also *Displacement Current*.)

ELECTRIC FIELD—The area surrounding an electrified body in which the electrical influence of that body can be measured or noticed. Also any area or region in which there is an electric force either steady or varying in intensity. For example, the *dielectric* or insulating surface of a *condenser* (material separating the plates) contains an electric field. Similarly, an electric field always surrounds a wire or other conductor carrying electric current. (See *Field, Magnetic; Field, Electromagnetic; Field, Electrostatic;* also *Flux*.)

ELECTRIC INDUCTION—The transfer of an electric state from a charged or electrified body to a non-electrified body without electrical contact. (See *Induction*.)

ELECTRICITY—The term given to an invisible form of energy. Electricity is not tangible to the human senses, but its effects can readily be detected. The phenomenon of electricity is regarded as due to the separation and independent movement of constituent parts of atoms—known as *electrons*. While electricity is really an intangible thing so far as the human perceptions are concerned, it is known to science not only through its effects or manifestations of energy, but has actually been reduced to quantities. The mass of electricity has been determined, as has also the *rate of flow*, etc. We have units which are used for the basis of all calculations in electricity. Thus, *voltage* is regarded as the propelling force; the quantity of electricity is given in *coulombs*, the *current* or rate of flow of electricity is given in *amperes*, etc. Each term for the various phases of electricity is based on a certain unit which represents definite values under given conditions. The universally accepted theory in chemistry is that all matter is made up of molecules, which in turn are composed of one or more atoms. These atoms now are regarded as comprising minute solar systems made up of positive and negative electrical particles. (For more complete explanation of the *electron theory* see *Electron Theory*; see also *Magnetism, Current, Voltage*.)

ELECTRO-CHEMICAL CONDENSER—See *Electrolytic Condenser*.

ELECTRO-CHEMICAL EQUIVALENT—The amount of a substance liberated by *electrolytic action* (electrolysis) by the passage of one *coulomb* (q.v.) of electric current. It represents the weight in grams of each element of an *electrolyte* (q.v.) which is deposited by one coulomb of electricity. It has been determined that one coulomb will liberate .00001035 grams of hydrogen, which is therefore used as the electro-chemical equivalent of hydrogen. Electro-chemical equivalents are much used in applying the law of electrolysis. This law states that the amount (weight) of an *ion* liberated at an *electrode* each second during *electrolysis* is equal to the strength of the *current* (amperes) multiplied by the electro-chemical equivalent. The corresponding equivalents of other elements compared to hydrogen may be determined by multiplying the figure .00001035 by the atomic weight of the other element in grams and dividing the result by the *valency* (q.v.) of the element in question. (See *Electrolysis*, also *Electrolyte*.)

ELECTRO-CHEMICAL SERIES—A table of metals arranged in accordance with the *potential* they produce in a particular *electrolytic* solution.

ELECTRO-CHEMISTRY — The science dealing with chemical changes due to electricity or effected by means of electricity. (See *Electrolysis*, also *Cell*, and *Storage Batteries*.)

ELECTRODE — The conductor through which electric current enters or leaves an electrical device. That is to say the pole or terminal of the current carrying conductors separated by a medium through which electric current can flow from one to the other. The most common usage of the term is in referring to the *positive* and *negative* poles (q.v.) of a *primary* or *storage cell*. In the case of a *dry cell*, the two electrodes are the carbon rod or block in the center and the zinc container in most cases. In radio it is used to designate the *filament* and *plate* elements of a vacuum tube. The filament is termed the *cathode*, or *negative electrode* and the plate is the *positive electrode* or *anode*. The illustration shows an ordinary *dry cell* or *primary cell* with the two electrodes designated as positive and negative. (See *Anode, Cathode, Battery, Electrolysis, Positive Electrode* and *Negative Electrode*.)

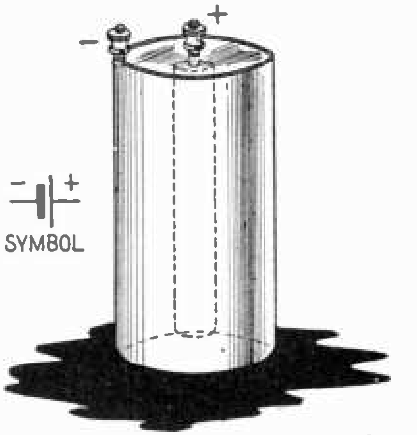

Dry cell showing positive and negative electrodes.

ELECTRODYNAMIC—Term used to denote electric currents or the forces exerted by one current upon another.

ELECTRO-DYNAMICS — A branch of the science of electricity dealing with electricity in motion; the study of the force exerted by electric currents upon each other. This branch of electricity was first accorded serious attention when the French scientist, *A. M. Ampere* (q.v.) announced the result of his investigations in collaboration with D. F. J. Arago. Ampere stated in 1820 that *parallel conductors* through which electric currents were flowing in the same direction had a tendency to be attracted to each other and to be repelled when the currents were flowing in opposite directions. The subject is discussed under its various separate headings. (See *Ampere, Electricity, Magnetism.*)

ELECTRO-DYNAMOMETER—A device for measuring the *current* and *voltage* in a circuit; that is, the actual power in a circuit. It is a form of *wattmeter* (q.v.) designed to indicate the power in *watts*. It is of particular importance in measuring the power in *alternating current* circuits. In the case of *direct current*, the watts in the circuit are the product of the volts multiplied by the amperes, the result being the actual power. In the case of alternating current, however, the product of the volts and amperes is not necessarily, in fact, seldom, the true power or watts. Here we have the problem of true watts and apparent watts, or volt-amperes. If a voltmeter and an ammeter are connected in a circuit and read separately, the readings indicate the maximum voltage and the maximum current or amperage. The product of these two values is the power in watts in a direct current circuit. If the current is an alternating one, the true watts or actual power will be the product of the *instantaneous values* (q.v.) of voltage and current. Under certain conditions in a circuit carrying alternating currents, the current may lag behind or precede the voltage. In other words, the maximum value of current is not reached at the same moment as the maximum value of the voltage. In this case the true watts can only be determined by obtaining the values of volts and amperes at the same instant, i.e., instantaneous values. The electro-dynamometer shown in the illustration gives the instantaneous values, or rather their product, and thus indicates the power or true watts. Here the heavy lines represent the current or moving coil and the light lines the stationary or voltage coil. (See *Dynamometer*, also *Wattmeter.*)

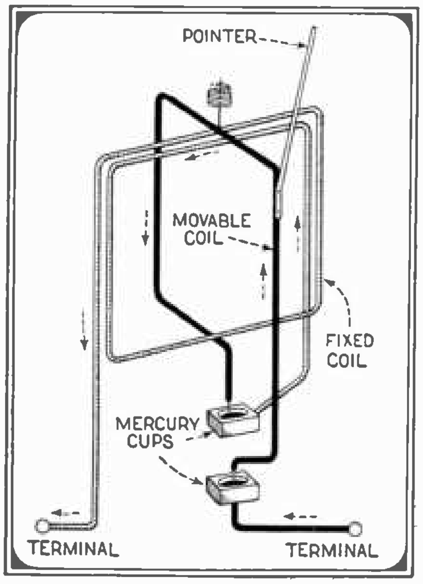

Electro-dynamometer with the current or moving coil shown by heavy lines and the stationary or voltage coil by light lines.

ELECTROLINES—A term advanced by *Professor J. A. Fleming* (q.v.) to designate the lines of electric force radiating from an electron. These lines are pictured as analogous to long straight wires extending in all directions from the center of a small sphere. (See *Electron*, also *Electromagnetic Waves.*)

ELECTROLYSIS—The decomposition of a comopund substance, generally a liquid, into its component parts by the action of an electric current passing through it. While electrolysis is an undesired effect in certain phases of electrical power work, causing composition of grounded structures, its application to radio is mainly in the case of cells for producing an *electromotive force* or to furnish current for lighting tubes. The constituent elements of a chemical compound such as used in a wet cell or any device using an electrolyte for the production of electric current, are known as *ions* (q.v.). These ions, which are separated during electrolysis, are of two kinds. The electro-positive ions are known as *cations*, and the electro-negative ions are known as *anions*. The electro-positive ions appear at the cathode or negative electrode during electrolysis and therefore are of positive origin. The electro-negative ions appear at the positive electrode and are therefore of negative origin. This is easily understood when the ions are considered as moving from the positive electrode or to the negative. Thus the ions appearing at the negative electrode must have come from the positive electrode and are therefore considered electro-positive in nature. The illustration shows the action of the ions in a typical *primary cell* (q.v.). The ions are considered as carrying the current through the electrolyte—each ion carrying a fixed charge of electricity of positive or negative nature. An ion is capable of carrying only a certain fixed charge of electricity and therefore any increase in the current will necessarily be accomplished by an increase in the number of ions. (See *Primary Cell, Electrolyte, Electrolytic Detector, Interrupter* and *Condenser.*)

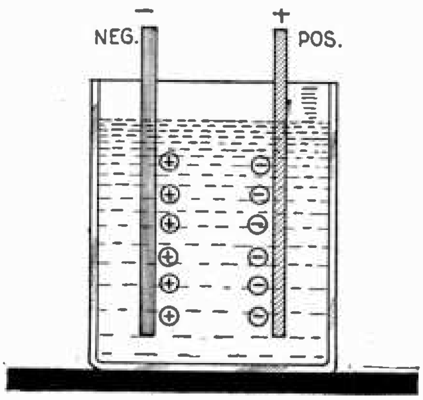

Illustration showing the action of ions in a primary cell.

ELECTROLYTE—The liquid decomposed during electrolysis, as the exciting fluid or solution in a wet cell or primary cell. The liquid used in any primary, secondary or electrolytic cell. (See *Cell*, also *Storage Battery.*)

ELECTROLYTIC ACTION—The decomposition of a chemical compound (electrolyte) into its constituent elements—ions, by the passage of an electric current through it. (See *Electrolysis, Ions*, also *Cell.*)

ELECTROLYTIC BATTERY CHARGER—A device for converting alternating current such as supplied for house lighting purposes, to direct current for the purpose of charging storage batteries. Essentially one or more cells containing an electrolyte and two metal electrodes, generally lead and aluminum, so arranged that current can only pass in one direction, thus making it uni-directional and suitable for charging batteries. Such chargers are arranged to rectify either one-half of the alternating current cycle, or both halves of the wave. (See *Charger, Storage Battery, Electrolytic Rectifier*, also *Electrolytic Cell.*).

ELECTROLYTIC CELL—An arrangement of two *electrodes* in an electrolyte through which electric current can be passed to produce electrolysis. The electrolyte is decomposed or separated into ions by the action of the current, and at each of the electrodes one of the substances of which the electrolyte is composed accumulates. Such a cell is often used to measure electric current, as the amount of a substance deposited on the electrode depends directly on the amount of current passed through the electrolyte. (See *Electrolysis.*)

ELECTROLYTIC CONDENSER—A form of condenser making use of two electrodes placed in an electrolyte and used for alternating current circuits of high voltage and low frequency, low voltage and high frequency or high voltage and high frequency. The action of such an arrangement as a condenser depends on the fact that upon the passage of certain forms of alternating currents through it, a gas is formed around the electrodes, causing *polarization* (q.v.) and opposing the current. This polarization creates a very good insulating medium and a condenser of such nature can be made to have very high capacity. An electrolytic condenser may be used to prevent any sudden surge of pressure due to lightning or other causes. The polarization is instantaneous when any surge of voltage takes place and thus forms an automatic safety device which operates instantly and returns to its normal characteristics as a condenser as soon as the surge has passed. (See *Capacity*, also *Condenser.*)

ELECTROLYTIC CONDUCTION—The action in an electrolyte wherein the ions that have been separated carry positive and negative electric charges in opposite directions to the electrodes. (See *Electrolysis.*)

ELECTROLYTIC DETECTOR A device which converts high frequency currents into direct current pulsations capable of operating a telephone receiver. It is an extremely sensitive detector of electromagnetic waves (radio signals). The illustration shows a simple form as originally used in place of the customary crystal detector. A fine platinum wire dips into a

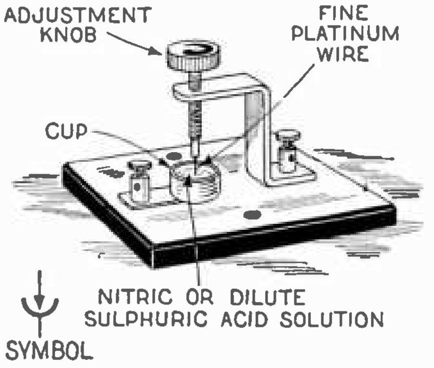

Electrolytic detector originally used in place of a crystal detector for radio reception.

cup of acid solution, generally either nitric or dilute sulphuric. The fine wire is lowered until it just touches the acid solution and must be very carefully adjusted as it has a tendency to curl up, its diameter being usually but one five thousandth of an inch or less. The high frequency currents can pass in only one direction through the detector which gives the rectifying action and makes the signals audible in the head phones. There are several theories regarding the action of an electrolytic detector, the most probable one being that the contact of the wire on the surface of the acid solution gives the effect of a very small *electrolytic condenser.* (q.v.) When the high frequency currents are passed through the detector there is a polarization action which causes high resistance to current in one direction but not in the other. (See *Detector,* also *Polarization.*)

ELECTROLYTIC INTERRUPTER—A device using electrodes immersed in an electrolyte for the purpose of interrupting or breaking up the flow of current. It is used mainly for induction coils for producing high voltages. The common form comprises a lead plate as a cathode or negative element and a platinum wire extending beyond a glass or porcelain tube in which it is tightly imbedded, both immersed in an electrolyte of dilute sulphuric acid. When current passes through, bubbles form on the platinum electrode causing temporary stoppage of the current. The bubbles are automatically dissipated and the current flows again, this operation occurring at rapid intervals and giving as many as 1000 interruptions per second. (See *Interrupter,* also *Induction Coil.*)

ELECTROLYTIC RECTIFIER—A device for changing alternating current into direct current by use of electrodes placed in an electrolytic solution. The action of such electrolytic cells or rectifiers is to permit current to pass only in one direction, thus suppressing one half of each cycle and producing a direct current. The illustration shows one of the customary arrangements for electrolytic rectification, being an electrolytic cell with two electrodes, one being aluminum and the other a lead electrode. One or more lamps are placed in series or shunt in the circuit to permit the proper current flow. The positive pole of a storage battery is connected to the aluminum plate as shown and the lead electrode connects to the other side of the alternating current line and back to the negative side of the battery. At one instant current flows in the normal direction through the battery, but when the current reverses and flows in the opposite direction through the electrolytic cell, oxygen is given off at the aluminum electrode and forms a thin film of aluminum oxide which acts as an insulator and prevents the passage of the current in that direction. Thus with one cell one half of each alternating current cycle is stopped or cut off when the aluminum electrode is made positive and the result is an intermittent direct current suitable for charging batteries. Such rectifiers are also used for supplying current to the filaments of vacuum tubes, or voltage to the plate members of the tubes, particularly for transmitting. (See *Electrolytic Rectifier Full Wave,* also *Balkite Battery Charger.*)

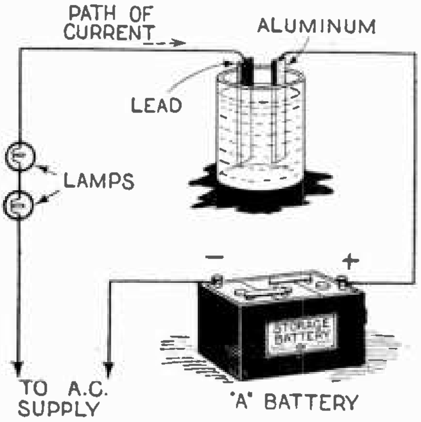

Customary method of electrolytic rectification.

ELECTROLYTIC RECTIFIER FULL WAVE—This system is much used for charging storage batteries from an alternating current house lighting supply. Four cells are used in this instance. The electrolyte may be a solution of hydrolchloric or sulphuric acid, or a saturated solution of sodium phosphate or bicarbonate of soda. Each cell has two electrodes, one iron or aluminum and the other lead. The positive pole of the storage battery is connected to the lead electrodes of two cells and the negative pole is connected to the aluminum electrodes of two cells. By this means both halves of the alternating current cycle are rectified to direct, pulsating current. This is known as *full wave rectification* (q.v.) (See *Charger, Storage Battery, Balkite Battery Charger,* also *Rectifier.*)

ELECTRO-MAGNET—A magnet using a curved or straight piece of soft iron or other magnetic material so arranged as to become magnetized by the passage of electric current through a coil of wire surrounding it. The illustration shows one of the commonest forms of electro-magnet. This type is used in a bell or buzzer to draw down the vibrator bar and break the circuit. (For explanation of action see *Buzzer.*) When the current is withdrawn from an electro-magnet, the magnetic force or attraction ceases. Electro-magnets are used for a variety of purposes in radio, such as for test buzzers, head phones, relays, etc. (See *Magnet,* also *Magnetic Lines of Force.*)

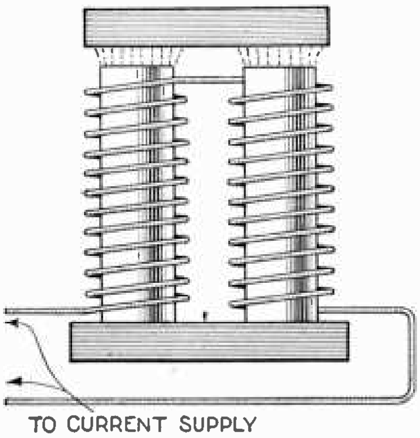

One of the common forms of electro-magnets.

ELECTROMAGNETIC CONTROL—The control of various switches and other apparatus by means of *electromagnets,* usually through an auxiliary contact at a distant point. (See *Remote Control,* also *Electromagnet.*)

ELECTROMAGNETIC FIELD—The space understood to be filled with *electromagnetic lines of force.* (See *Flux,* also *Lines of Force.*)

ELECTROMAGNETIC FLUX—The distribution of the lines of force through an electromagnetic field. (See *Flux,* also *Lines of Force.*)

ELECTROMAGNETIC IMPULSE—An impulse or electromagnetic disturbance conveyed to the ether from a conductor carrying high frequency currents. Another term for the electromagnetic waves involved in radio. (See *Electromagnetic Waves.*)

ELECTROMAGNETIC INDUCTION — The production of electric currents through a change in a *magnetic field* (q.v.). Its production is due to a change in the number of lines of force which are linked with a conducting circuit. The simplest example of this means of producing current is shown in the illustration. Here a conductor is arranged to be moved in the immediate vicinity of a magnet, causing a change in the number of lines of force that link the conductor and creating an induced current. This current will persist only so long as the conductor continues to be moved in the magnetic field. That is to say when there is no longer any change in the magnetic lines of force linking the conductor, there will be no current induced. The discovery of electromagnetic induction and formulation of its laws was primarily due to the scientist *Faraday* (q.v.). Faraday's law governing electromagnetic induction follows: "The induced electromotive force around any circuit is the rate of the decrease of the total flux of magnetic induction through the circuit." (See *Induction, Flux,* and *Lines of Force.*)

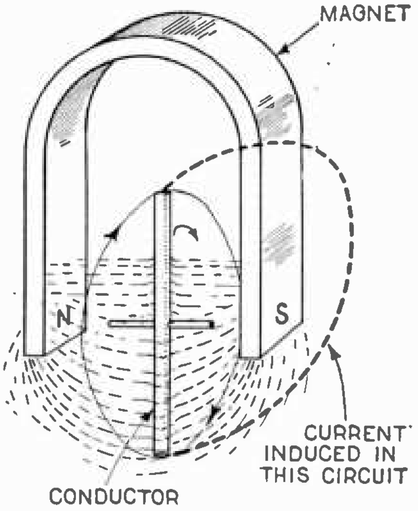

Simple method of producing electric currents by causing changes in the magnetic field.

ELECTROMAGNETIC INSTRUMENTS —Electrical measuring instruments such as voltmeters, ammeters, etc., which depend for their action on the exertion of electromotive forces upon iron armatures. (See *Ammeter, Thermal;* also *Voltmeter.*)

ELECTROMAGNETIC MICROPHONE —A type of microphone in which the sound waves due to speech or music cause vibrations of a suspended coil in a magnetic field, inducing in the coil minute waves of electromotive force corresponding in form to the original sound waves. A form of microphone used in radiophone (broadcasting) transmission when purity of sound is an important consideration. Such a microphone requires that the induced currents be heavily amplified. (See *Microphone*, also *Speech Amplifier.*)

ELECTROMAGNETIC RADIATION—The propagation or radiation of electromagnetic waves through the ether. (See *Electromagnetic Waves*, also *Ether.*)

ELECTROMAGNETIC ATTRACTION—The attraction existing between unlike poles of electromagnets. This attraction is identical with that of two unlike poles of ordinary magnets. (See *Magnetic Attraction and Repulsion.*)

ELECTROMAGNETIC REPULSION—The tendency of two like poles of electromagnets to act against or repel each other. (See *Magnetic Attraction and Repulsion.*)

ELECTROMAGNETIC SWITCH—A switch for breaking electrical circuits by the action of an electromagnet. (See *Electromagnet*, also *Remote Control.*)

ELECTROMAGNETIC THEORY OF LIGHT—The theory and consequent scientific proof that radio waves and light waves are practically identical in nature. Light waves travel at a tremendous speed through the ether, approximately 186,000 miles per second, which is also understood to be the velocity of electromagnetic or radio waves. (See *Electromagnetic Waves.*)

ELECTROMAGNETIC UNITS—The fundamental or absolute units employed in electricity as a basis for calculations and on which the practical units are based. The effects of electricity, and hence electricity itself, can be measured in two ways. The first is the measurement of the quantity of electricity by the force it exerts upon another stationary quantity of electricity. This is known as the *electrostatic system.* The second is by measuring the force which a given quantity of electricity exerts upon a *magnetic pole* when flowing through a nearby conductor. This is referred to as the electromagnetic system. Fundamental or absolute units are derived from the fundamental units of *length, mass* and *time.* These units are the *centimeter of length, the gramme of mass* and the *second of time.* The term centimeter-gramme-second, referring to this system, is abbreviated C.G.S. system. The practical units of electricity are known as *volts, amperes*, etc., and are based on the absolute units. (See *Electrostatic, Units, C.G.S. System*, also *Ampere, Erg, Dyne* and *Coulomb.*)

ELECTROMAGNETIC WAVES—The term applied to the energy radiated from an *antenna* in radio transmission. So called because it is understood to travel in the form of waves. One-half the energy of such a wave is electrical and the other half electromagnetic. *Electromagnetic waves* travel with the speed of light waves or approximately 186,000 miles per second. These waves are of extremely high frequency—that is, their direction changes many thousand times each second. This will be better understood by means of the illustration showing the arrangement of the electric and *magnetic fields* around a radiating antenna. Here the lines represent the electric field and the shaded portion the magnetic field. The lower loops or "feet" of these waves are assumed to pass over the surface of the earth which is a conductor. Thus when electromagnetic waves are being radiated from an antenna, circular distributions of electric charges are in the earth's surface around the antenna, these charges being alternately negative and positive, the bands of charge spreading out or radiating from the antenna in the manner shown in the illustration. These waves are subject to various influences which have a tendency to refract or reflect the waves much in the manner of light waves.

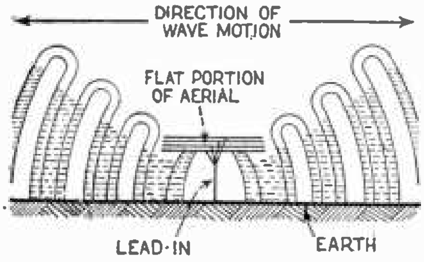

Illustration showing the arrangement of the electric and magnetic field around a radiating antenna.

As these waves move out from the antenna they gradually increase in height, although the distance between the "feet," i.e., from one wave to another, remains the same. This distance is known as the wave-length. (See *Reception of Electromagnetic Waves*, also *Wave-length, Radiation Transmission.*)

ELECTROMETER—An instrument used for measuring potential differences (voltages) and, under certain conditions, power in *watts.* The operation of the instrument depends upon the attraction or repulsion of electrostatic charges and it may be used for either direct or alternating current measurements. (See *Voltmeter, Electrostatic.*)

ELECTROMETRY—The practice or science of electrical measurement. (See *Meter, Voltmeter.*)

ELECTROMOTIVE FORCE—Symbol E or e—abbreviation E.M.F. The force or electrical pressure which starts and maintains a current of electricity through a conductor. Electromotive force is commonly referred to in terms of volts and is abbreviated E.M.F. The electromotive force in electricity is analogous to the difference of level which produces pressure in pipes carrying water and forces the water to flow. Electromotive force is produced by a difference of potential, and this in turn sets up the flow of current. The illustration *Fig. 1A* shows a simple primary cell or electrolytic cell, and *Fig. 1B* is the analogous action of two tanks holding water, connected by a pipe and so arranged that the water level in one tank is higher than in the other, thus having a difference of level and creating pressure which makes the water flow. The water in tank A being at a higher level than in the other, the pressure causes the water to seek its common level and it flowed into tank B through the pipe. This flow continues as long as there is difference of level. In the case of the cell *Fig. 1A* there are two electrodes, one of which is at a higher potential than the other. That is to say, one is in a state of higher electrification than the other, causing a difference of potential and producing an electromotive force. This force is in the direction from the point of greater potential to point of lesser potential as in the case of the water, and causes a flow of current in that direction. Here the flow is from B to A in the electrolytic solution and from A to B externally through the conductor C. Thus as long as there exists a difference of potential, current will flow in an attempt to equalize the potentials. (This statement holds true only under ideal conditions. See *Polarization.*) This flow of current in the illustration *Fig. 1A* is according to *Ohm's law* (q.v.), i.e., a difference of potential of one volt will cause one ampere to flow through the external circuit when the resistance of the circuit is one ohm. (See *Ampere, Ohm, Volt, Resistance.*)

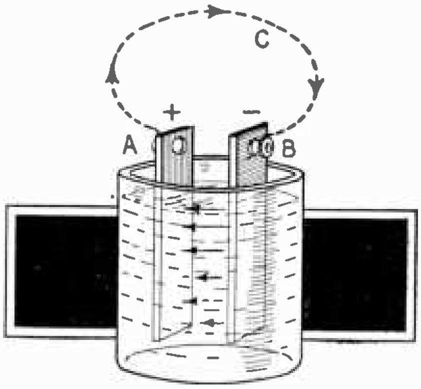

Fig. 1-A. A simple primary or electrolytic cell.

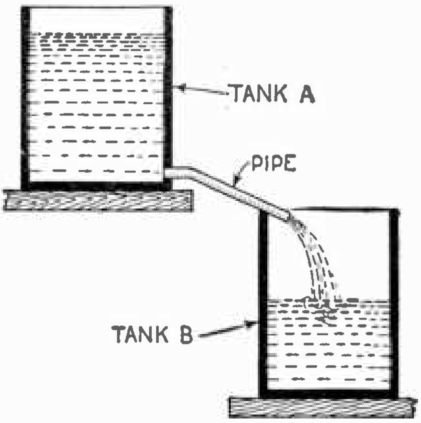

Fig. 1-B. Water in tank "A" seeks its level by discharging into tank "B" in a manner analogous to the flow of electric current where a potential difference exists.

ELECTRON—The smallest negative charge of electricity known. An English investigator, Dr. Johnstone Stoney, is credited with having named the electron, although it was so named by him in anticipation of its being isolated and measured. The actual proof of the earler assumptions was due to Sir J. J. Thomson and others, who actually measured, and in all effect weighed, the minute particles of negative electricity. (See *Electron Theory.*)

ELECTRON FLOW—The theory of movement of free electrons along a conductor through which an electric current is flowing. (See *Electron*, also *Current, Assumed Direction of Flow.*)

ELECTRON EMISSION—The emission or discharge of electrons or minute particles of negative electricity from a heated body, particularly in a vacuum. The emission of electrons in a vacuum tube is the basis of operation of all vacuum tubes as used in radio. (See *Electron Theory*, also *Vacuum Tube, Theory of Operation.*)

ELECTRON THEORY—The assumption that the atoms of all substances contain one or more particles of negative electricity, which is of course paramount to saying that all matter is electricity. These particles are known as electrons, and while the assumption is referred to as a theory, it has been rather definitely proven by numerous experiments, and electrons have actually been measured. Under the electron theory it is assumed that the charge carried by a single electron is the smallest possible charge which can exist in nature and that no charge of electricity exists or can be produced which is not an integral multiple of this charge. (Note: This is merely a mathematical way of stating that any charge of electricity is equivalent to the combined charges of a certain number of electrons.) Probably the most accurate conclusion regarding the charge carried by an electron is due to experiments carried on by Prof. A. E. Millikan of the University of California. His measurements state this charge of a single electron as being 4.77×10^{-10} electrostatic unit, which is 1.59×10^{-19} of a *coulomb* (q.v.). While the mass or weight of an electron is said to vary with its velocity, assuming a velocity of less than one-tenth that of light, the mass will be 8.9×10^{-28} grams. This is approximately one seventeen hundredth of the weight or mass of an atom of hydrogen. For purposes of calculation the electron is assumed to be in the shape of a sphere. The radius has been estimated at approximately 10^{-13} centimeter. When it is considered that a centimeter is somewhat less than half an inch it is readily seen that an electron is an infinitesimal affair.

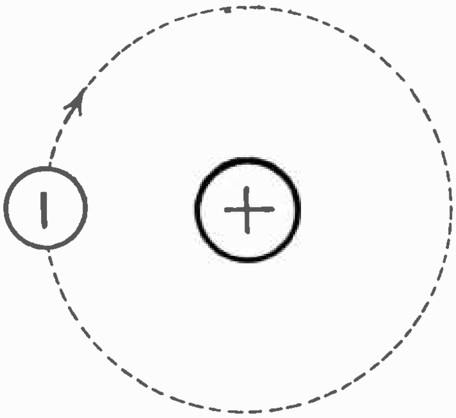

Illustration showing the electron, or negative particle, revolving around the proton or positive nucleus.

The figures as to velocity of electrons vary somewhat with different authorities, but it is probably about 100 kilometers (approximately 62½ miles) per second at a temperature of zero degrees centigrade, the speed increasing as the temperature rises. Electrons may be understood as comprising minute solar systems of their own. In the illustration the *electron* or negative particle is shown revolving around the *proton* (q.v.) or positive nucleus in the manner of a planet revolving around the sun. This represents a normal atom of hydrogen, the simplest of all atoms, which, as shown, is composed of a single positive nucleus and a single electron revolving around it. In the case of an atom of helium there is a doubly charged positive nucleus, or proton, and two negative electrons revolving around it. Other atoms have as many as 92 electrons.

All vacuum tubes used in radio operate on the basis of the emission of electrons from a heated filament within the tube. These electrons are being emitted from the hot filament at all times during operation of the tube. When the incoming signals are such as to charge the grid negatively they are repelled and held back to the filament, thus offering no path for the signals. When the incoming signals have reversed their polarity, as they do thousands of times each second, the grid will be positively charged and the electrons will be aided in their travel to the plate, thus forming a conductive path for the currents. This, of course, results in the elimination of one-half of each cycle of the incoming currents or oscillations. That is to say, they are rectified to direct pulsating currents. (See *Amplifier, Vacuum Tube, Theory of Operation* and *Space Charge.*)

ELECTRON RELAY—A term occasionally applied to a three element tube (vacuum tube). This alternate term is obviously derived from the fact that the vacuum tube makes use of the path of electrons and in effect acts as a relay. (See *Electron.*)

ELECTRON TUBE—Any tube using a high vacuum and depending for operation on the emission of electrons from a heated *cathode.* (See *Electron,* also *Vacuum Tube.*)

ELECTRONEGATIVE—Having the property of being attracted to the positive pole of an electrolytic cell. The electronegative component. (See *Anion,* also *Cathion.*)

ELECTROPOSITIVE—Having the property of being attracted to the negative pole of an electrolytic cell during electrolysis. The electropositive component. (See *Anion,* also *Cathion.*)

ELECTROSCOPE—A device for detecting minute differences of potential (voltages). Such a machine will not measure the voltage but merely indicate its presence. The simplest form of electroscope is shown in the illustration. Here a ball B, composed of the pith of some wood, preferably elder due to its extreme lightness, is suspended by a thin silk thread from an insulated stand. The object under investigation is moved near the pith ball, which will be repelled or attracted by the object if the object is in a state of electrification; but will remain motionless if no difference of potential exists. (See *Voltmeter, Potential, Difference of.*)

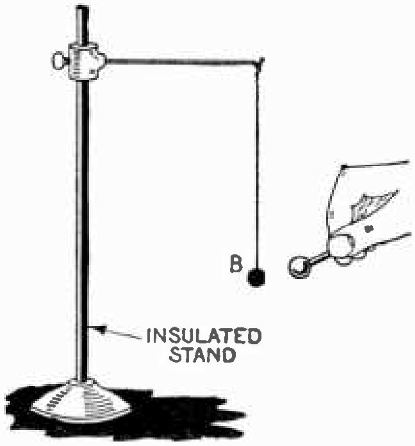

Simplest form of electroscope.

ELECTROSE—An insulating composition much used in electrical power work and for radio purposes. It has very high compressive strength, is hard and tough without being brittle, and is moisture, water and oil proof. Electrose is used extensively for antenna insulation. (See *Insulator*).

ELECTROSTATIC—The term used to designate EMF's (electromotive forces) or potentials, charges, etc., and the resultant forces. That is to say, electricity at rest or stationary in a body in order to distinguish it from the forces due to *electro-dynamics* (q.v.).

ELECTROSTATIC CAPACITY—The amount of electricity which a condenser or similar body is capable of storing. If a condenser is capable of receiving and storing one *coulomb* of electricity under a potential of one *volt* it is said to have an electrostatic capacity of one *farad,* which is the unit of electrostatic capacity. The electrostatic capacity of a condenser depends on several factors, mainly the size of the plates and the nature and thickness of the dielectric material (insulating sheets). (See *Capacity,* also *Condenser.*)

ELECTROSTATIC CHARGE—The presence of electricity in the form of electrostatic lines of force in a body such as a charged conductor. (See *Charge.*)

ELECTROSTATIC COUPLING—The association or coupling of circuits one to another by means of electrostatic action, i.e., condensers. Any capacitive association of one circuit with another. (See *Coupling.*)

ELECTROSTATIC FIELD—A region wherein forces are present and exerted due to the presence of electric charges. It is the area immediately surrounding a charged body. (See *Field, Electromagnetic,* also *Field, Electrostatic*).

ELECTROSTATIC FLUX—Lines of force making up an electrostatic field. (See *Field, Flux,* also *Flux, Electrostatic*).

ELECTROSTATIC FLUX DENSITY—The density or number of electrostatic lines of force in a unit area of an electrostatic field. (See *Electrostatic Field,* also *Flux Density.*)

ELECTROSTATIC FORCE—The force due to interaction of electric charges. (See *Electrostatic Field,* also *Electrostatic Flux Density.*)

ELECTROSTATIC INDUCTION—The production of a charge in a body due to the presence of an opposite charge in a nearby conductor, or synonymous with displacement in the case of a condenser. Electrostatic influence or induction is the basis on which condensers operate. In the illustration *Fig.* 1 is a simple circuit containing a source of potential, which in this case is a battery. As the circuit is closed through the key a momentary flow of current takes place. This flow of current is accompanied by a condition of electric strain in the insulating material of the condenser. The electrically strained dielectric or insulating material exerts a back pressure against the pressure or force of the battery, and when this back pressure is equal to the original battery pressure the flow ceases. The strain in this dielectric is known as *electric displacement* and this displacement can be transmitted or passed on to neighboring parts of the medium and thence into space. Under the proper conditions this displacement and the magnetic field which accompanies it may be detached from the circuit and move independently in space. This is the fundamental process in the radiation of *electromagnetic waves* (q.v.). Now, in *Fig.* 2, A and B represent two conductors, both insulated. If conduc-

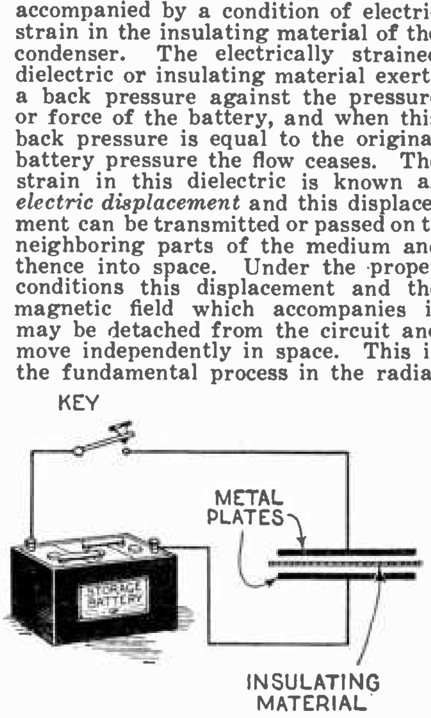

Fig. 1. A simple circuit for illustrating electrostatic induction.

tor A has a positive charge and is brought close to conductor B, although not in actual contact with it, the result is a displacement of electric charge which results in the near side of conductor B becoming negative, while the opposite side assumes the same sign (positive) as conductor A. Actually B has received no charge, but has been placed in a state of strain due to electrostatic induction between it and the conductor A. Due to the negative charge induced on B, the positive potential difference of A has been decreased and in order to maintain the constant potential it must be given a further supply of electricity. In this manner it is actually possible to store the electricity by electrostatic induction between two conductors or plates, by reason of the strain in the insulating medium—air in the case of the two conductors—which in turn passes the displacement on to conductor B. (See *Conductor, Capacity, Electrostatic,* also *Induction.*)

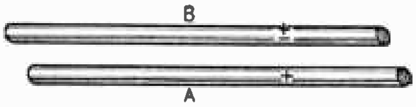

Fig. 2. Illustration of displacement of electric charge due to proximity of charged conductors.

ELECTROSTATIC LEAKAGE—The gradual loss through leakage, of the electrostatic charge in a body, due to the fact that perfect insulation is impossible to obtain. (See *Electrostatic Induction.*)

ELECTROSTATIC LINES OF FORCE—The lines of force traversing an electrostatic field. That is to say the *electrostatic flux* (q.v.).

ELECTROSTATIC STRAIN—The strain or state of stress in which a body is placed when in an electrostatic field. In a condenser, the strain experienced by the *dielectric* (q.v.) when the condenser is in a state of charge. (See *Condenser,* also *Capacity, Electrostatic*).

ELECTROSTATIC SYSTEM—The electrostatic system of C.G.S. units. One of the absolute unit systems in electrical practice. (See *C.G.S.*, also *Electrostatic Units* and *Electromagnetic Units.*)

ELECTROSTATIC UNITS—A system of measurement used in electrical engineering based on the unit of electric quantity under the *C.G.S. system* (q.v.). This unit quantity of electricity is the quantity of electricity which, placed on a small sphere will repel an equal quantity of electricity of the same sign (both negative or both positive) on a similar sphere with a force of one *dyne.* (q.v.) when the centers of the respective spheres are one centimeter apart. With this definition it is assumed that the spheres are in a vacuum, otherwise the distance between centers would vary according to the dielectric or insulating medium. (See *C.G.S., Electromagnetic Units,* also *Conversion Factors.*)

ELECTROSTATIC VOLTMETER — A meter for measuring voltage, depending for its action on the electrostatic action between a fixed and a moving plate connected respectively to the two poles or terminals between which the source of potential is connected. (See *Electrostatic Force,* also *Voltmeter.*)

ELECTROSTATIC WATTMETER — A type of wattmeter depending on the electrostatic action between two sets of fixed and moving plates. (See *Electrostatic Force,* also *Wattmeter.*)

ELECTRO-THERMAL METER — The term sometimes applied to a *hot-wire ammeter* (q.v.) or voltmeter operating on the principle that the heat generated by the passage of an electric current through a constant resistance will be equal to the square of the current. (See *Thermal Ammeter.*)

ELEMENT—A term primarily applied in chemistry to the substances which are not subject to decomposition by electric analysis (see *Electrolysis*). These elements are hydrogen, helium, etc. The term is also used quite generally to denote components such as the electrodes in a primary cell or the plate, grid and filament members of a vacuum tube. (See *Cell,* also *Vacuum Tube.*)

EMANATE—"To flow forth or proceed, as from a source."—Webster. Thus, emanating, issuing from, as the emanation of electrons from a hot filament in a vacuum tube. The throwing off of electrons. (See *Electron Emission.*)

E.M.F. — The abbreviation for electromotive force. The practical unit of E.M.F. is the volt. (See *Electromotive Force,* also *Volt.*)

EMISSION—The act of sending forth. Used chiefly to designate the emission of electrons from the filament of a vacuum tube as used in radio. The flow of these electrons between filament and plate is controlled by the grid. (See *Electron Emission,* also *Vacuum Tube, Theory of Operation.*)

EMISSION CURRENT—In a vacuum tube for transmitting it is the product of the tube current multiplied by the plate voltage and is generally given in watts. Generally speaking, the power for which the tube is designed. (See *Transmitting Tube, Power Rating,* also *Vacuum Tubes.*)

EMISSIVITY OF FILAMENT — The ability of a filament to emit electrons when heated by an electric current. (See *Electron,* also *Vacuum Tube.*)

EMPIRE CLOTH OR PAPER—A closely woven cambric coated with two or more films of an oxidized oil, or a tough paper treated with some oxidized oil. Empire products serve as insulating materials for transformers, condensers and various other devices requiring thin insulating sheets. In a transformer for relatively high voltages, insulated paper is used between the layers of windings and in certain types of fixed condensers it is used as a dielectric material. The cloth has a higher breakdown strength; that is, it will stand higher voltages than paper. (See *Breakdown Potential,* also *Dielectric.*)

ENAMELED WIRE — Wire, generally copper, covered with a coating of enamel as an insulating covering and to protect it from moisture and corrosion. (See *Magnet Wire.*)

ENDODYNE—A term occasionally used to describe reception of radio signals by means of locally generated oscillations such as in the case of the *heterodyne* (q.v.), but wherein these oscillations are created by a circuit or system that is part of the receiving circuit and not a separate arrangement. (See *Self Heterodyne.*)

ENERGY—The capacity for doing work in an electrical sense. Energy is something furnished to a body by means of work done on it. Thus, if a ball is thrown, the operation of throwing is the work, and the ball itself in traveling performs work, the means of doing this work being the energy.

There are two kinds of energy, *potential,* as in the case of electricity, and *kinetic,* the energy of a body due to motion. Kinetic energy belongs properly in the realms of mechanics and so will be passed over here with only this brief mention.

When we speak of a difference of potential in a primary cell, it means in effect that the body or cell has potential energy. That is to say, the energy is present although not active. If, now, a wire is connected across the two terminals of the cell, the difference of potential will cause electric current to flow from one pole to the other. Actually, in a cell of this type, we have chemical energy stored up. When a circuit is completed, the stored chemical energy is converted into electrical energy or creates the difference of potential, which in turn causes current to flow and electrical *work* is thus done. (See *Potential,* also *Voltage* and *Generator.*)

EPSTEIN HYSTERESIS TESTER — A device for measuring *hysteresis* and *Eddy current loss* in a sample of sheet iron by using an alternating current wattmeter. The device is used for determining the properties of samples of iron, particularly those to be used for transformers. (See *Eddy Currents,* also *Transformer, Hysteresis* and *Wattmeter.*)

EQUIVALENT RESISTANCE — The value which a resistance would be required to have in order to permit the same quantity of current to flow with the same applied electromotive force (voltage) as in the case of a piece of apparatus wherein there are other factors besides pure resistance to determine the amount of current. (See *Impedance,* also *Reactance.*)

ERG—The fundamental unit of mechanical work under the *C.G.S. system* (q.v.). This is defined as the work done by a force of one *dyne* (q.v.) when the body producing the force and the body on which the force acts are one centimeter apart. (See *Joule,* also *Practical Units.*)

E.S.U. — Abbreviation sometimes used for *electrostatic units* (q.v.). (See *Electromagnetic Units.*)

ETHER—or **AETHER**—The term used to denote the supposed or hypothetical medium by means of which electromagnetic and light waves are propagated. (See *Ether Waves.*)

ETHER WAVES—Term sometimes applied to electromagnetic or radio waves on account of the assumption that they travel through the supposed medium ether. The assumption of an all-pervading medium such as ether has been the basis of the major part of the extensive research work carried on with light and electromagnetic waves for many years. Various experiments have shown beyond reasonable doubt that there is some such medium, capable of conducting or rather permitting passage of light waves and radio waves. So accurate has this work been that it has been proven beyond question that light waves travel through the ether with the tremendous velocity of 186,000 miles per second (approximately). It has also been determined that radio waves travel at this same speed through the ether. It has further been proven that radio waves all travel at the same identical velocity, regardless of their length, form and other factors. (See *Electromagnetic Waves,* also *Theory of Propagation of Electromagnetic Waves.*)

EXCITER—A small direct current dynamo used for the purpose of exciting the field magnets of an alternating current generator. Such a device is generally a part of the main machine and mounted on the main shaft, but it may also be a separate unit. The purpose is to furnish current to the field magnets (electromagnets) to create a magnetic flux. (See *Alternator*, also *Generator.*)

EXPLORING COIL—A coil of insulated wire wound on a rotating form as an integral part of a direction finder or *goniometer*. (See *Direction Finder*, also *Goniometer.*)

F

F—The abbreviation for *dielectric field intensity* (see *Field*, also *Dielectric*), also used as the symbol for the unit of *magneto-motive force*, the *gilbert* (q.v.). The small or lower case f is sometimes used to denote the *farad*, the practical unit of *electrostatic capacity* (q.v.) and also as a symbol for *frequency* (q.v.).

FADING—The tendency of radio signals to decrease in volume or amplitude, or to fade. The strength of a signal is seldom the same at different hours of the day or night, generally being stronger during the dark hours of the night. This has been attributed to a variety of causes, the most plausible being that the air is more free of disturbances during the night and thus permits more ready passage of the waves through the ether. The difference between the strength of signals in the daytime and at night is sometimes very pronounced, but the phenomenon of fading is usually noticed only when distant signals are being received. It has been fairly definitely established that fading of signals is due to the reflection of the transmitted waves. (See *Wave Distortion*, also *Reflection and Refraction.*)

FALL OF POTENTIAL—See *Potential Drop.*

FAN AERIAL—An aerial, usually for transmission, constructed in the shape of a fan. (See *Aerial.*)

FAN CONNECTOR—A small triangular or fan-shaped piece of metal arranged with several holes along one edge, with set screws for fastening in each the several wires of the lead-in of a multiwire aerial. The lower apex is fitted with some sort of connecting device to permit a single wire to be led in to the receiver. (See *Lead-In.*)

Illustration by Courtesy of Radio Specialty Co., Inc.

Fan Connector—The leads from each wire of a multi-wire aerial go to the wide portion and are held by the set screws: the lead to the set goes to the small end.

FARAD — The unit of *electrostatic capacity*. It is the practical unit of capacity and is defined as the capacity of a condenser that is capable of being charged to a potential difference of one volt by a charge of one ampere for one second; i.e., one *coulomb* (q.v.). "Farad" is a contraction of the name of the distinguished English scientist *Michael Faraday* (q.v.). Inasmuch as the farad is too large a unit for practical purposes, the *micro-farad*, or one millionth of a farad, is commonly used. (See *Electrostatic Capacity*, also *Condenser.*)

FARADAY, MICHAEL. Born 1791, died 1867. An English scientist famous for his discoveries in chemistry, electricity and magnetism. He first produced the rotation of the magnetic needle around the electric current in 1821, based upon Oersted's discovery of electromagnetism in 1820; he discovered electromagnetic induction (1831), a principle upon which is founded the development of dynamo machinery; specific inductive capacity (1838); magnetic polarization of light (1845); diamagnetism (1846). He was a brilliant experimenter, and contributed greatly to the knowledge upon which is based present-day practice of electricity.

Michael Faraday.

FEEBLY DAMPED—In spark transmission, signals that are only slightly damped. Where each succeeding oscillation in a train of oscillations is of slightly lesser *amplitude* (q.v.) than the one preceding it, the signals are said to be slightly damped to distinguish them from signals that are highly damped and thus die out quickly. (See *Damping*, also *Damped Waves.*)

FEED—In electrical parlance, to furnish with a current of electricity. (See *Feed-Back.*)

FEED-BACK—A term applied in radio to the coupling of one circuit to another, whereby a portion of the current present in one circuit is fed back or returned to the other circuit. The common use of the term is in reference to regeneration, where a certain portion of the radio frequency current present in the plate circuit of a detector is returned to the grid circuit of the tube by some form of coupling between the two circuits. It may also apply to the tendency of high-frequency currents to become transferred from one circuit to another due to undesired coupling between two or more circuits. The illustration shows a simple feed-back or regenerative circuit. The incoming signal energy is transferred from the primary of the coupler to the secondary by means of inductive coupling and tuned by the condenser C. A grid condenser and leak GC are placed in the grid circuit in the customary manner. A feed-back or reaction coil T is placed in the plate circuit and inductively coupled to the main coil. A portion of the radio frequency current in the plate circuit or output circuit is transferred back to the grid circuit by the coupling effect between the coil T and the secondary of the coupler. This serves in effect to reduce the resistance of the grid circuit and produces more current in the output or plate circuit, thus acting as a form of amplifier. (See *Regeneration, Tickler.*)

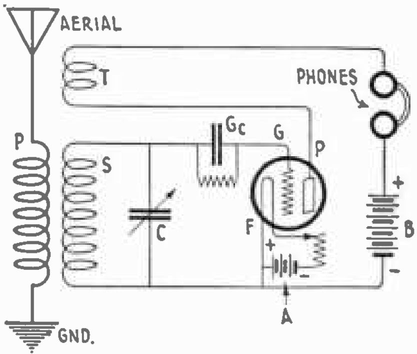

Regenerative circuit in which the coil "T" returns, or "feeds back" to the grid circuit a portion of the radio frequency current in the plate circuit, resulting in increased amplification.

FEED-BACK COIL—An inductance coil placed in the plate circuit of a regenerative receiver and arranged in inductive relation to the grid coil (secondary of tuner) for the purpose of effecting feed-back or regeneration. (See *Feed-Back*, also *Regeneration.*)

FEED-BACK COUPLING—Any process by which the plate circuit of a detector is coupled to the grid circuit to allow a portion of the output current to be transferred back to the grid circuit to produce regeneration. (See *Feed-Back*, also *Regeneration.*)

FEED-BACK EFFECT — The effect of feeding back of currents from one circuit to another as in the case of a *regenerative circuit* (q.v.). Generally the expression "feed-back effect" refers to an undesired feed-back. This may be the action of radio frequency components of the currents present in one part of a circuit being fed back to another part of the circuit due to the coupling effect of various coils or other devices used in the circuit. (See *Feed-Back*, also *Regeneration.*)

FEED-VOLTAGE MODULATION — A process of modulating or varying the amplitude of a radio frequency alternating current to correspond to any wave form of speech vibrations as in radio broadcasting. The system involves the introduction of additional power into the circuit of the radio frequency generator until the desired wave form variations are obtained. This is done particularly in the case of a vacuum tube transmitter, in which case the voltage supplied to the plate of the tube is altered (increased or decreased) to accomplish the purpose. (See *Modulation.*)

FERRO-MAGNETIC MODULATOR—A device for modulating or varying the amplitude of radio frequency current to correspond in wave form to the speech or music vibrations in broadcasting. The system makes use of what is known as the *hysteretic energy absorption* of iron, or in some cases utilizes the variation of inductance of iron-core coils. (See *Modulation*, also *Ferro-magnetic Substances.*)

FERRO-MAGNETIC SUBSTANCES — Elements or compounds such as iron, nickel, cobalt and others which are strongly attracted by magnetic fields of force. Such substances vary widely in their magnetic properties when subjected to temperature changes. Soft iron loses its magnetization almost immediately on removal of the magnetizing source, and for this reason it is used as a pole piece in an electromagnet. (See *Electromagnet*, also *Magnetism* and *Magnetic Properties.*)

FESSENDEN, REGINALD AUBREY. —Canadian-American Radio expert. Born at Milton, Canada, October 6, 1866, and educated at New York and Port Hope, Ontario. Fessenden became inspecting engineer for the Edison Company, New York, and afterwards professor of physics and electrical engineering at Western University, 1892. Professor Fessenden is the author of a well-known system of wireless, and below are briefly described some of the patents bearing his name.

In 1906 and 1907 Fessenden invented a number of microphone transmitters which carried heavy currents for long periods, and also a heavy current telephone relay which allowed the controlling of heavy currents by means of small currents originating in an ordinary microphone circuit or coming from a telephone line. One of these transmitters was called by Fessenden a trough transmitter. It consisted of a soapstone annulus to which were clamped two plates having platinum-iridium electrodes. Through a hole in the center of one plate passed a rod attached at one end to a diaphragm, and at the other to a platinum-iridium spade. The two outside electrodes were water-jacketed. This form of transmitter required no adjusting, all that was necessary being to place about a teaspoonful of carbon granules in the center space. It was able to carry as much as 15 amperes continuously without articulation falling off, and had the advantage that it never packed. By a combination of the trough transmitter and a differential magnetic relay, Fessenden produced a transmitting relay for magnifying very feeble currents. An amplification of fifteen times is possible without any loss of distinctness. Fessenden is also responsible for a duplex system of Radio telephony, and the *heterodyne* method of reception is due to him. Fessenden has written largely on radio subjects, and is one of the leading authorities on both transmission and reception.

Reginald Aubrey Fessenden.

FIBER—or **FIBRE**—A term used for a variety of substances, composed of fine slender thread-like materials, particularly a composition of vulcanized or compressed paper. It is much used in radio as an insulating material for panels and connection blocks. It is not as efficient as insulating medium as bakelite or some forms of rubber composition, but is durable, cheap and comparatively easy to drill. (See *Bakelite.*)

FIELD—The space occupied by electric or magnetic lines of force. (See *Electric Field.*)

FIELD DENSITY—The strength of an electric or magnetic field. It is measured by the number of electric or magnetic lines of force contained in a given cross-sectional area. (See *Density. Flux*, also *Electric Field* and *Field, Magnetic.*)

FIELD, ELECTRIC — The space traversed or occupied by electric lines of force. (See *Electric Field.*)

FIELD, ELECTROMAGNETIC — The total area traversed by electromagnetic lines of force. The region in which electromagnetic lines of force are exerted. (See *Field*, also *Electrostatic Force.*)

FIELD, ELECTROSTATIC—The region in which electrostatic forces are present or exerted. The total area traversed by electrostatic lines of force. (See *Field*, also *Electrostatic Flux.*)

FIELD MAGNET — An *electromagnet*, usually a core of soft iron surrounded by a winding of insulated wire, employed in generating machines to produce a strong magnetic field. As the *armature* of a *generator* revolves the conducting surfaces pass through the *magnetic field* created by the *field magnets.* This field may be merely the ordinary magnetic force due to permanent magnets, but is usually due to the powerful magnetic lines of force resulting from passage of electric current through electromagnets. (See *Alternator*, also *Alternating Current, Theory of Production* and *Electromagnet.*)

FIELD, MAGNETIC—The region through which magnetic flux passes. Usually refers to the space (air) through which the magnetic flux lines pass as distinguished from the activity in the iron path tself. (See *Field*, also *Magnetic Flux.*)

FIELD REGULATOR or **RHEOSTAT**—Any device for varying the magnetic strength of the field magnets in a dynamo or electric motor, by which variations certain changes may be accomplished in the operation or performance of the machine. A field rheostat is often used in an electric motor to vary the strength of the field and thus either increase or decrease the speed of the machine without the necessity of changing the applied voltage. If the strength of the field (magnetic) is reduced by insertion of resistance, the armature or rotating part speeds up in an attempt to generate the same *counter-electromotive force* (q.v.), which means greater speed. On the other hand, if the resistance is removed—that is, the strength of the field increased—the fact that the armature is running through a stronger field reduces the speed of rotation, the same counter-electromotive force being present. The use of a field regulator or rheostat in connection with a generator is for the purpose of varying the output voltage without the necessity of altering the speed of the machine. The illustration shows the connection of a field rheostat or regulator in the circuit of the *shunt wound generator* (q.v.). (See *Generator, Rheostat, Resistance.*)

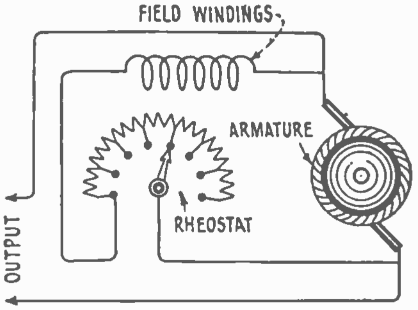

Illustration showing method of connecting a field rheostat or regulator in the circuit of a shunt wound generator.

FILAMENT — A fine wire, one of the elements of a vacuum tube, which is heated to throw off electrons. The term is also used to refer to the fine wire used in an incandescent lamp. The filament of a vacuum tube is generally made of tungsten and is supported by glass pillars within the tube in such manner as to be insulated from the other elements. Filaments are made to be heated with various voltages from 1.1 to 6 volts in most receiving tubes, but much higher in the case of high power transmitting tubes. The filament may consume considerable current as in the old styles of six-volt tubes which consumed one ampere. The modern tubes have thoriated filaments, tungsten treated with thorium, which permits maximum efficiency with a minimum of current consumption. When the filament is heated in the vacuum tube, by having the proper amount of voltage applied, electrons are thrown off and fill the space between filament and grid and grid and plate as shown in the illustration. (See *Electron Theory*, also *Vacuum Tube, Theory of Operation.*)

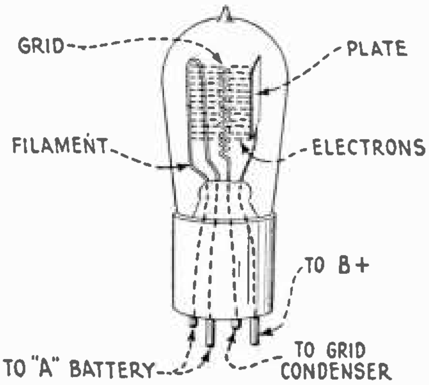

When an "A" battery is connected to the filament terminals, heating the filament, electrons are thrown off and fill the space between filament, grid and plate.

FILAMENT BATTERY—The battery or cell used in lighting the filament of a vacuum tube. These range from a *dry battery* (q.v.) having a voltage of one and a half volts, to storage batteries of six volts. Filament batteries are connected across the filament of a vacuum tube in the manner shown in the illustration. The rheostat is inserted to permit the supply of current to the filament to be varied and thus change the operation of the tube. (See *Filament*, also *Vacuum Tube, Theory of Operation.*)

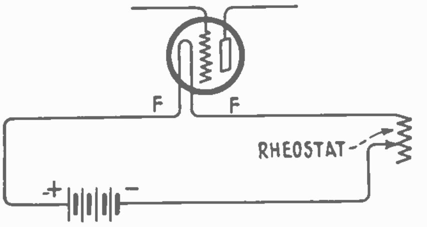

Method of connecting "A" battery in series with rheostat to regulate filament of vacuum tube.

FILAMENT CONSUMPTION — The amount of current consumed by the filament of a vacuum tube. The current requirements of vacuum tubes are always stated by the manufacturer, either in amperes or fractional amperes, or in the case of extremely low consumption, in thousandths of an ampere (milliamperes). Some typical examples are the UV201A and C201A types, which require one-quarter ampere at five volts; the WD11 and 12 and C11 and 12 types, which require one-quarter ampere at approximately 1.1 volts, and the UV200 and C200 tubes, which require one ampere at six volts. Of the so-called standard types, the UV199 and C199 tubes consume the least current, requiring but sixty milliamperes or .06 ampere, with from three to four and a half volts. The statement that a certain tube consumes one-quarter ampere indicates that in operation one-quarter ampere is consumed each hour, this current being supplied by dry cells or storage batteries known as the *"A" battery* (q.v.).

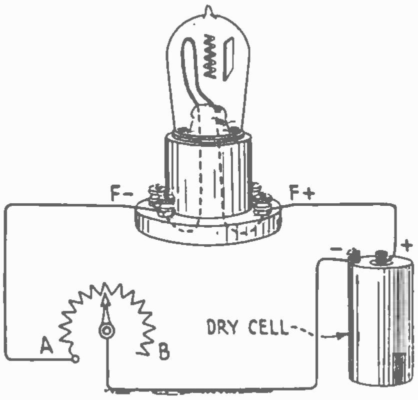

The current produced by the dry cell battery, flowing through the rheostat and filament, is consumed by the filament, at a rate depending upon the type of tube.

It will be apparent from the above that the more current consumed by the filament the greater will be the drain on the dry cells or storage battery. For this reason, and also to lengthen the life of the filaments, it is well to operate the tubes at the lowest efficient brilliancy. The brilliancy of the filament may be controlled by adjustment of the filament rheostat. In the illustration the usual "A" battery, which in this case is a dry cell, is shown connected to the filament of a vacuum tube with a rheostat in series between one filament connection and the negative terminal of the dry cell. Now as the arm of the rheostat is moved from A toward B resistance is introduced into the circuit between the filament and dry cell. As this resistance increases the current in the circuit of the filament will decrease according to *Ohm's Law* (q.v.); i.e., Current = Voltage ÷ Resistance, the current in amperes or fractional amperes and the resistance in ohms. (See *Ampere, Current, Filament Resistance or Rheostat* and *Vacuum Tube.*)

FILAMENT CURRENT — The current used to light the filament of a vacuum tube. Different tubes require different amounts of current; that is, one tube may consume a small fraction of an ampere and another may require a full ampere. Of the tubes most commonly used, the UV199 type uses the least current, this tube requiring only .06 of an ampere each hour. The WD11 and 12 types require one-quarter ampere per hour and also the UV201A, while the now obsolete UV200 and UV201 required a full ampere. The 199 type operates with filament voltage of 3 volts, although it is often used with three dry cells in series which deliver 4½ volts; the WD11 and 12 require 1.1 volts each, which is usually furnished by a dry cell, while the UV201A type is operated with a six-volt storage battery. (See *Vacuum Tubes, Types of*, also *Vacuum Tube, Theory of Operation.*)

FILAMENT, FORMULA FOR ELECTRON FLOW—The extent of electron emission—that is, the number of electrons—emitted from a hot filament in a vacuum tube is subject to certain definite laws. In ordinary practice the actual number of electrons are not referred to, the ability of the filament to emit electrons being measured in fractional amperes of output current. Thus a tungsten filament having a diameter of .0125 centimeters operating with a power dissipation of 3.1 watts per centimeter of length has been found to emit a current of .03 amperes. This result was obtained with a filament having a clean surface and with no absorbed gas on its surface. The formula for electron flow most commonly used is as follows:

$$I_B = AT^{1/2}\,\epsilon - \frac{b}{\tau}$$

I_B is the electron current in milliamperes per square centimeter of filament surface, T is the absolute temperature and A and b depend on the metal used for the filament.

FILAMENT RESISTANCE or RHEOSTAT—A device placed in the filament circuit of a vacuum tube, connected between the filament battery and either filament lead on the tube socket for the purpose of reducing or controlling the filament voltage. The illustration Fig. 1 shows one type of filament rheostat which is used to permit variation of the filament current. Fig. 2 shows the manner of connecting such resistance in the filament circuit of a vacuum tube. Fig. 3 illustrates a type of filament control known as an automatic rheostat. In this form of filament resistance a fine wire is placed in a glass tube filled with an inert gas. The wire has the property of increasing its resistance rapidly with any increase in the current passing through it. It is thus possible to arrange such a device to pass only a fixed amount of current for any type of vacuum tube. The device is thus not only an automatic rheostat, but also a safety valve which prevents excessive current reaching the filament and injuring it. These controls are coming into widespread use due to these features and the fact that they reduce the number of controls necessary to operate a receiver. Fig. 4 shows a resistance which is used in series with a rheostat having small resistance in order to enable its use with tubes requiring less current to operate. (See *Filament*, also *Vacuum Tube.*)

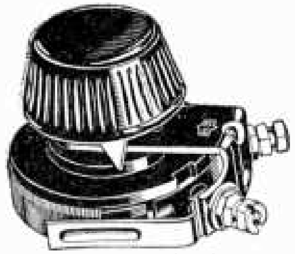

Illustration by Courtesy of The Cutler-Hammer Mfg. Co.

Fig. 1. A type of rheostat used to control the flow of current to the filament of a vacuum tube.

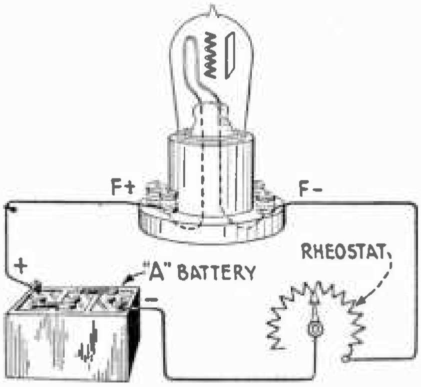

Fig. 2. Showing method of connecting a rheostat in "A" battery circuit of a vacuum tube.

Illustration by Courtesy of the Radiall Co.

Fig. 3. A type of automatic rheostat.

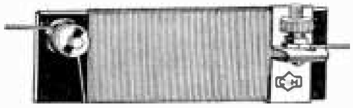

Illustration by Courtesy of The Cutler-Hammer Mfg. Co.

Fig. 4. Resistance used in series with a rheostat having a small resistance in order to enable its use with tubes requiring less current.

FILAMENT, THORIATED—A filament or heating element of a vacuum tube in which a certain amount of *thorium* (q.v.) has been incorporated. The introduction of this rare metal into the filament in small quantities gives it the property of emitting electrons at a relatively low temperature. From a practical standpoint this permits a vacuum tube to be operated with a low *filament consumption* (q.v.). A common example of the use of thorium in filaments is that of the UV201A or C201A vacuum tubes, which operate with a filament voltage of 5 volts and consume but one-quarter ampere of current, giving essentially the same results as the former type UV201 and C201 tubes which consumed one ampere. The value of thorium used in this manner is readily appreciated, the current consumption being materially lessened and the life of the tube conceivably lengthened. (See *Thorium Treatment of Filaments*, also *Vacuum Tubes, Type UV201A and C201A.*)

FILTER—In radio, a device arranged to permit passage through an electrical circuit of currents of certain frequency or *frequencies* (q.v.) while excluding others of different frequencies, or any device for producing smooth direct current from a generating source such as a *direct curent dynamo*, or for smoothing out *rectified direct current.* Filters of one sort or another are much used where rectified direct curent or direct current from a generator is to be used for operation of tubes in radio transmission or reception. In the case of rectified direct current, that is the current derived from any device for changing alternating to direct current, there is always present a tendency on the part of the current to surge or ripple. In other words, it is not smooth and of constant voltage, or there may be present a generator hum which will seriously interfere with proper operation of a tube. The illustration shows the conventional method of filtering rectified direct current for use as plate supply with vacuum tubes in receiving sets. This system may also be used in connection with the output of a direct current generator or dynamo. A filter circuit may be arranged to pass a band of frequencies (see *Band Pass Filter*) or it may be made to pass either high or low frequency current. (*See Filter, High Pass;* also *Filter, Low Pass.*)

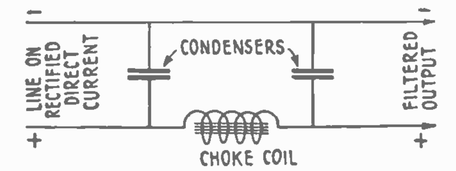

Conventional method of filtering rectified direct current for use as plate supply of a vacuum tube.

FILTER, BAND PASS—See *Band Pass Filter.*

FILTER CONDENSER—The condenser, either fixed or variable, forming a part of a circuit for filtering. (See *Filter*, also *Acceptor, Wave Trap* and *Rejector.*)

FILTER, HIGH PASS—A filter circuit arranged to permit current above a certain frequency to pass freely while blocking out or attenuating all frequencies below it. (See *Filter*, also *Filter, Low Pass.*)

FILTER, LOW PASS—A filter circuit arranged to permit currents below a certain frequency to pass freely while blocking currents of higher frequency. (See *Band of Frequencies*, also *Filter, High Pass.*)

FIXED COILS—Inductance coils as used in radio circuits, having no mechanical means of variation, such as in a *Vario-coupler or movable coil.* (q.v.) The majority of tuned radio frequency receivers employ fixed coils, shunted by variable condensers. These are known as *tuned radio frequency transformers.* (q.v.) (See *Coupler, Loose*, also *Coupler, Vario.*)

FIXED CONDENSER — A condenser having a certain definite or fixed value of capacity. Any condenser that is not variable. (See *Condenser*, also *Condenser, Variable.*)

FIXED RESISTANCE—Any resistance device having a fixed or non-variable value. (See *Rheostat*, also *Potentiometer.*)

FIXED DISCHARGER—A fixed *spark gap.* (q.v.) A spark gap used in transmission of radiotelegraphic signals in which the electrodes or sparking surfaces are not moved during operation. (See *Disc Discharger.*)

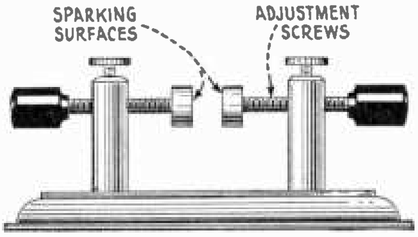

A type of fixed spark gap.

FLAME MICROPHONE—A type of microphone for radio telephone transmission, wherein the height of a coal gas flame is varied by and according to the speech fluctuations and the resulting pulsations made audible by means of a selenium cell and phones. (See *Microphone*, also *Selenium Cell.*)

FLASH SIGNALLING— See *Code*, also *Morse Light.*

FLAT-TOP AERIAL or ANTENNA—An aerial, generally composed of a number of wires placed adjacent to each other equally spaced on a spreader, or hoop, and suspended parallel with the earth. A flat-top aerial may be of the kind known as the *T type aerial* (q.v.) or in the form of an inverted L. (See *Aerial.*)

FLAT TUNING—A term sometimes applied where the *resonance curve* (q.v.) has a flat top, that is where the peak of resonance is not sharply defined. The effect is known as *broad tuning.* (q.v.) (See *Resonance*, also *Tuning.*)

FLEMING, JOHN AMBROSE. British radio expert. Born in Lancaster, November 29, 1849, he was educated at University College, London, the Royal School of Mines, and St. John's College, Cambridge.

From 1873-74 Fleming was demonstrator at the Royal College of Chemistry, and in 1877 he began work under Clerk-Maxwell at the Cavendish Laboratory, Cambridge. There he carried out a series of experimental researches on the British Association Standards of Electrical Resistance. In 1881 Dr. Fleming was appointed the first professor of mathematics and physics at University College, Nottingham, but the following year he joined the Edison Electric Light Company, and on the amalgamation of the Edison and Swan Companies, he was appointed advising electrician, a post he held for twenty years. In 1885 Fleming was appointed to the newly founded professorship of electrical engineering at University College, London, and he was entirely responsible for the design and equipment of the new electrical and engineering laboratories which were opened in 1893. In 1892 he was elected Fellow of the Royal Society.

Dr. Fleming has been very closely associated for many years with radio telegraphy and telephony, and has since 1899 acted as the scientific advisor to Marconi's Wireless Telegraph Company. He is well known as the inventor of the thermionic two electrode vacuum tube, or glow-lamp detector, which was one of the greatest steps forward in radio telephony.

Dr. Fleming has written a very large number of papers and books on wireless telegraphy and telephony, and his "Principles of Electric Wave Telegraphy" is the standard treatise on the subject. He has also published an "Elementary Manual of Radio Telegraphy and Radiotelephony"; "The Wonders of Wireless Telegraphy"; "The Thermionic Valve and Its Development"; "A Pocket Book for Wireless Telegraphists," and other books. As a lecturer at the Royal Society of Arts he has obtained a world-wide reputation for his lectures on electric oscillations and electric waves, Hertzian wave telegraphy, high-frequency measurements, etc. He has given many lectures on wireless at the Royal Institution.

John Ambrose Fleming.

Fleming was awarded the Hughes gold medal of the Royal Society in 1910 as an acknowledgement of the value of his work in electrical science and engineering, and he has twice been awarded the Institution Premium of the Institution of Electrical Engineers, the highest award for communicated papers, and he is an Albert Gold Medalist of the Royal Society of Arts awarded to him for the pioneer invention of the thermionic valve.

FLEMING'S RULE—A rule for determining the direction of the *induced current* (q.v.) in a circuit. Hold the right

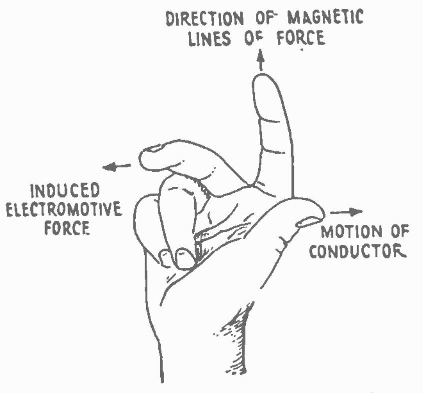
DIRECTION OF MAGNETIC LINES OF FORCE
INDUCED ELECTROMOTIVE FORCE
MOTION OF CONDUCTOR

Method of using the right hand in determining the direction of the induced current.

hand as shown in the illustration with the thumb, forefinger and middle finger as nearly as possible at right angles to each other. If the thumb is then pointing in the direction of motion of the conductor and the forefinger points along the direction of the magnetic lines of force, the middle finger will then point in the direction of the induced electromotive force. (See *Induced Current*, also *Electromotive Force*.)

FLEMING VALVE—The forerunner of the modern vacuum tube. It was invented in 1904 by Professor J. A. Fleming and was for a short time used as a detector in place of the customary crystal or coherer types. The value of the discovery of the Fleming valve lies chiefly in its influence on the course of radio development, as in itself it was not nearly as efficient as a good crystal detector. As a result of the discovery of the Fleming valve, however, many leading scientists bent their efforts toward the discovery of better methods of detecting radio waves, culminating in the three element vacuum tube designed by Dr. Lee De Forest.

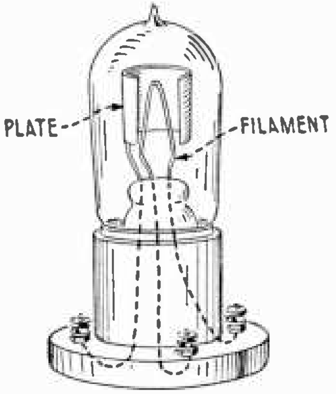

Fig. 1. The Fleming valve, which contained but two elements; i. e., filament and plate. It was the forerunner of the modern vacuum tube.

The Fleming valve consists essentially of an incandescent lamp using a filament of carbon, tungsten or tantalum which can be rendered incandescent by an electric current. The filament is sealed in a glass tube which is highly evacuated, that is from which practically all air has been excluded. Surrounding the filament, but separated from it is a metal cylinder, both ends of the filament and the plate being attached to outside connections by means of leads sealed in the tube. The illustration, Fig. 1, shows the general arrangement of such a tube. The operation of the tube is due to the emission of electrons from the heated filament. It was demonstrated by several investigators that a filament made of certain metals has the property of throwing off minute particles of negative electricity or electrons when heated to a relatively high temperature. It was also found (see *Electron*) that these electrons are thrown off in greater profusion when the filament is placed in high vacuum. It was also found that the escape of negative electrons from the filament would in time leave it positively charged, this charge having an attraction for the negative electrons and preventing their escape from the filament. Hence it was necessary to apply a charge of negative electricity to the filament. The electrons were now found to escape in great number, but had the effect of forming a *space charge* (q.v.) of negative electricity around the filament, which in time served to repel the escape of any more electrons. To counteract this tendency a strong positive charge was applied to the metal plate surrounding the filament. This made the plate positive in respect to the electrons and served to attract them away from the filament. The stream of electrons thus affords a path for current to flow from filament to plate, and the action of the tube as a detector of high frequency waves depends on this principle. If the tube is connected as indicated in the illustration, Fig. 2, it will serve to rectify the incoming signals, that is change them from high frequency alternating currents, to pulsating direct current. The action is as follows: The aerial is connected to the secondary circuit by means of a standard coupler. The coils, marked L2, are the primary and secondary of this coupler; C1 and C2 are variable condensers, F the filament of the Fleming valve, B the battery for lighting the filament of the valve and also to supply voltage to the plate, T a pair of headphones, P a 400 ohm potentiometer and R a standard filament rheostat. The incoming signals in the form of high frequency currents, change direction many times each second. Thus, at one instant the plate will have a positive charge while the next instant it will have a negative charge as a result of the changes in direction of the currents impressed across the tube from the antenna circuit. While the plate is positive, electrons will be attracted away from the filament, which at that instant will be negative with respect to the plate. The current thus flows for an instant, stopping when the impressed current changes its direction and makes the plate negative. In this manner we have a flow of current from filament to plate in one direction in the form of surges appearing at every other alternation or change of current. The diaphragms of the phones will respond to these uni-directional surges and will give out sounds almost identical with the original signals transmitted. The tube thus serves as a rectifier of the incoming signals and is known as a detector of high frequency waves, or oscillations. The Fleming valve principle is used in rectifying tubes for changing alternating current to direct current for charging batteries, its use as a detector for radio reception having died out almost entirely. (See *Charger*, *Storage Battery*, *Detector*, *Tungar*, *Rectigon*, also *Vacuum Tube*.)

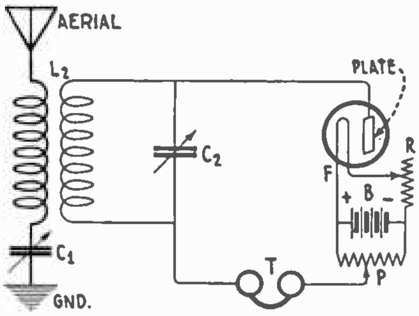

Fig. 2. Method of connecting a Fleming valve in a radio receiving circuit.

FLEWELLING CIRCUIT—A modification of the Armstrong Super-regenerative circuit, designed by the American experimenter, Flewelling. In this system, the tendency toward *self-oscillation* (q.v.) is retarded by the use of fixed condensers which are arranged to have a *damping* (q.v.) effect on the grid of the detector tube, automatically controlling the oscillations of the tube. The circuit controls the self-oscillation of the tube without the necessity for the separate control tube used in the Armstrong circuit. The illustration shows the original Flewelling circuit. It is essentially a standard regenerative circuit with certain necessary modifications.

The coils L1 and L2 give magnetic coupling between the plate circuit and the grid circuit. C1 is a variable condenser for tuning and condensers C2, C3 and C4 are of the fixed type and afford a path of relatively low *impedance* (q.v) for the high frequency currents. C5 is the customary grid blocking condenser across which a grid leak is shunted. The usual telephone receivers T, the plate battery B and the filament battery A are included in the circuit in the customary manner. The coils L1 and L2 may be two honey-comb coils having respectively 50 and 75 turns. The tuning condenser C1 should be .0005 mfd. for the broadcast band of wave-lengths and the loop may be almost any form.

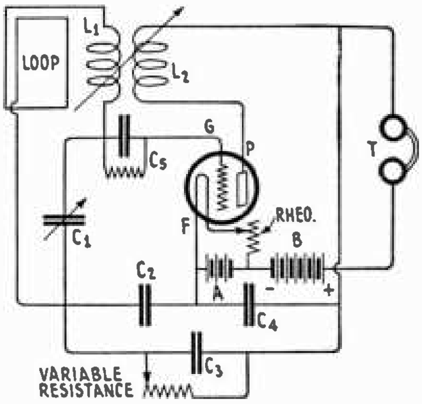

The Flewelling circuit.

In operation three forms of current pass through the plate circuit including the coil L2, namely, the direct current from the plate battery B, the unidirectional surges of rectified current due to the action of the tube as a rectifier and the high frequency currents transferred through the medium of the magnetic coupling of the two coils L1 and L2. The battery current is present only in the plate circuit, the condensers C2, C3 and C4 blocking the direct current from the grid circuit. These condensers also block the rectified currents but afford two paths for the high frequency currents or oscillations. The condensers C2, C3 and C4 have definite values, respectively .005, .005 and .006 micro-farads, to allow a by-pass through the grid to the filament, thus causing accumulation of electrons on the grid and stopping oscillation of the tube. The charge then leaks off by means of the grid leak connected across C5, the tube then building up oscillations again and continuing the process during operation. The high resistance connected across C3 is variable and permits control of the by-pass of high frequency current of the grid. Careful adjustment is necessary to obtain maximum signal strength without distortion. (See *Armstrong Circuits*, *Regenerative Circuit*, also *Super-Regenerative Circuit*.)

FLEX—Term used in Great Britain and occasionally in this country to refer to a number of small wires twisted or stranded together to form a single flexible conductor. (See *Flexible Wire*.)

FLEXIBLE LEAD—Any conductor, particularly the connections from rotary parts of couplers or condensers, in which the conductor is made up of a number of fine wires in such a way that the necessary bending in operation of the unit will not snap the wire as would be the case with a solid conductor. (*See Pig-tail*, also *Flexible Wire.*)

FLEXIBLE WIRE—Any conductor composed of a number of fine wires to permit constant handling and bending without breaking. (See *Flexible Lead.*)

FLOAT, BATTERY — Term sometimes applied to the bulb or float in a hydrometer for testing the condition of storage batteries. (See *Hydrometer.*)

FLOW OF CURRENT—An expression used to denote the passage of electric current along a conductor. (See *Current, Direction of Flow* (*Assumed*), also *Current.*)

FLOW OF MAGNETIC FLUX—The passage of lines of magnetic force through any magnetic circuit. (See *Flux*, also *Field, Magnetic.*)

FLUCTUATING CURRENT—An electric current that does not have steady value, rising and falling in strength or pressure. A current that is not constant. (See *Constant Current*, also *Voltage, Constant.*)

FLUORESCENT SCREEN—A screen or surface coated with fluorescent matter, used to exhibit shadows cast by rays from various types of tubes. Its application in radio is chiefly in the *Cathode Ray Tube.* (q.v.)

FLUX—A very broad term in physics. The three more common examples of flux are: (a) to denote *electrostatic or magnetic flux* (q.v.); (b) a substance used in soldering for the purpose of making the solder flow readily and making it adhere to the parts being united, and (c) in metal smelting, a mineral—customarily chalk or limestone—added to the charge in the furnace to absorb mineral impurities and dispose of them in the form of slag. *Flux* is also used to refer to the amount of heat or light flowing or passing through a given distance or area in a given time. (See *Electrostatic Flux, Flux Density*, also *Flux, Magnetic.*)

FLUX DENSITY—The number of lines of force per square centimeter of sectional area of a magnetic path. The unit of magnetic flux in the C.G.S. (Centimeter Gram Second) or electromagnetic system of units is the Maxwell. It is generally referred to as a "line." As a magnetic flux is merely the effect of magneto-motive force, we can take the rule that the magnetomotive force is 1.257 x ampere turns as the rule for determining the intensity or density of magnetic flux. (See *Flux; Field, Magnetic*, also *Ampere Turns* and *Maxwells.*)

FLUX DISTRIBUTION—The distribution of lines of force in a magnetic or electrostatic field of force. (See *Field*, also *Flux.*)

FLUX, ELECTROSTATIC — Another term for electric displacement as in the case of a condenser. The density of electrostatic flux refers to the intensity of an *electrostatic field.* (q.v.) (See *Flux Density, Field, Electrostatic*, also *Electric Displacement.*)

FLUX LEAKAGE—Any loss or dissipation of magnetic or electrostatic flux. (See *Flux, Electrostatic*, also *Flux, Magnetic.*)

FLUX LINES—The lines of force in a field of force in which electrostatic or electromagnetic force is exerted. An alternate term for *lines of force.* (q.v.)

FLUX, MAGNETIC—The term broadly used to denote the grouping of magnetic lines of force in a *magnetic field.* (q.v.). The most common example of magnetic flux is in the case of an ordinary bar magnet as shown in the illustration Fig. 1. Here the lines of force or magnetic flux create a magnetic field indicated by the dotted lines. The number of lines of force in a field of unit area is used as a basis for determining the density of the flux. (See *Flux Density.*) This effect or magnetic flux field is also present when a current of electricity is allowed to circulate around a coil of wire, the larger the current and the greater the number of turns, the more powerful the flux created. The illustration Fig. 2 shows a coil of wire through which current is assumed to be passing, the dotted lines indicating the flux field. (See *Choke Coil, Transformer, Magnet*, also *Flux, Electrostatic* and *Magneto-motive Force.*)

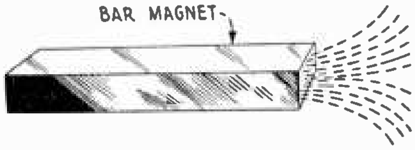

Fig. 1. Bar magnet, showing how the lines of force, or magnetic flux, create a magnetic field as indicated by dotted lines.

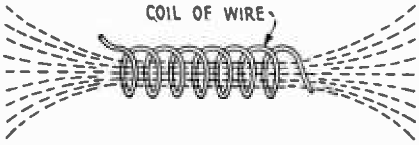

Fig. 2. Coil of wire through which current is assumed to be passing, the dotted lines indicating the flux field.

FLUX MAGNETISM — The magnetic flux, or total number of lines of force passing through a magnetic circuit. It is equal to the product of the magnetic density (see *Flux Density*) and the cross sectional area of the magnetic path. It is also referred to as magnetic flow. (See *Magnetic Flow.*)

FLUXMETER—A device for measuring flux density. The Grassot fluxmeter uses an *exploring coil* (q.v.) in conjunction with a special form of moving coil galvanometer. The coil is suspended on a silk fiber and its movements actuate a pointer across a scale. (See *Flux, Magnetic.*)

FLUX, SOLDERING—See *Soldering Flux, Flux.*

FOCUS or FOCI—The point or points at which converging lines meet. The magnetic foci are the points on the earth's surface in the vicinity of the magnetic poles, where the magnetic force of the earth is the greatest. Essentially a mathematical term.

FORCE—An influence exerted upon a body so as to produce a change or a tendency to change its state of rest or motion. This is mechanical force. In electricity we have its electrical analogy—*electromotive force.* This is the force which causes motion of electricity and sustains the electric currents against the resistance of a circuit. (See *Electromotive Force.*)

FORCE, MAGNETO-MOTIVE—The force which produces magnetic flux as in the case of a magnet moving through a magnetic circuit. The abbreviation is M. M. F. and its unit is the *gilbert.* (q.v.) (See *Magneto-motive Force.*)

FORCED OSCILLATIONS — Electrical vibrations or high frequency oscillations which are imposed upon an *oscillatory circuit* without being in tune or resonance with the natural period of vibration of the circuit. The opposite of a free oscillation in which case the circuit oscillates freely on its own account at the same frequency as the imposed or incoming oscillations or one of the *harmonics* (q.v.) of this frequency. Free and forced oscillations may be present in the same oscillatory circuit, the forced oscillations being set up during the first stages of induction of free oscillations. (See *Free Oscillations*, also *Oscillatory Circuit.*)

FORM FACTOR—The ratio of the *root-mean-square value* (q.v.) to the algebraic mean ordinate taken over one alternation (half cycle), starting with zero value of current. It is the ratio of the r.m.s. value of an alternating electromotive force, current, etc., to its true mean value. (See *Wave Analysis.*)

FORMULA—A principle or rule expressed in mathematical terms or chemical symbols. Formulae are much used in radio to state a problem as, for example, a formula for calculation of wave-length, which is written: Wave-length=$1884\sqrt{LC}$. This states the problem or rule for the determination of wave-length. Here we have certain values represented by symbols. The capacity is represented by C, the inductance by L, the wave-length in meters being the product of the square root of the product of these two multiplied by the factor 1884. (See *Measurement of Wave-length.*)

FOSTER BRIDGE—A form of *Wheatstone Bridge* (q.v) of the slide wire variety designed to permit comparison of two low value resistances that are nearly equal. It was devised by Professor Carey Foster and is used to determine the difference between two resistances as distinguished from the usual bridge for measuring the resistance of a single unit. (See *Wheatstone Bridge.*)

FOUCAULT CURRENT—Another name for *eddy currents*, the *electro-magnetic currents* due to the circulation of induced current in the mass of a conductor. Foucault first proved that these currents were productive of losses of energy as they cannot be utilized except in certain forms of meters. (See *Damping of Instruments.*) Alternating currents tend to set up eddy currents in metal masses near them. Thus in the case of a transformer, eddy currents are produced in the iron core due to the induced currents circulating through its mass. These currents are wasteful and means are therefore used for reducing them. The core of a transformer may be made of numerous flat pieces called laminations, which serve to prevent the free movement of eddy currents in the core. (See *Eddy Currents*, also *Transformer* and *Core Losses.*)

FOUR ELEMENT or FOUR ELECTRODE TUBE—A vacuum tube having four elements instead of the usual three. The standard tube has three elements, namely, grid, filament and

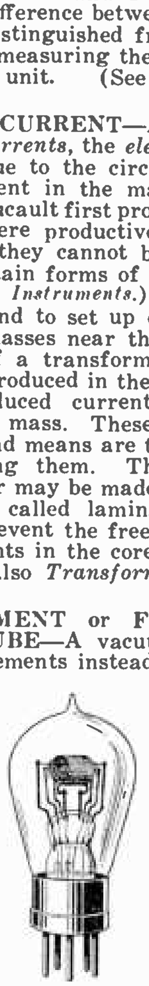

A four element vacuum tube.

plate. The four element tube has an additional grid element. The illustration shows one form of tube having a double grid arrangement. Such a tube can be used for high and low frequency amplification and as a detector simultaneously or it may be used in conjunction with the Solodyne system, which does not use the customary B battery to supply potential to the plate element. (For more complete explanation of the action of a four element tube in the latter manner see *Solodyne*). (See also *Vacuum Tube.*)

FOURTH CIRCUIT—A method of controlling oscillation of a regenerative detector used in the *Cockaday Circuit.* (q.v.)

FRAME AERIAL—An alternate name for *loop aerial,* a type of aerial consisting of a number of turns of wire wound on an insulated frame and used in place of the regular outdoor aerial with sensitive receiving sets. (See *Loop Aerial.*)

FREE ALTERNATING CURRENT—Term applied to alternating current assumed to be produced by the application of a single impulse of electricity to an *oscillatory circuit.* (q.v.) See *Free Oscillations.*)

FREE ELECTRONS—Name given to the outer electrons of an atom, which become detached under various chemical transformations and are assumed to participate in electrical effects or phenomena. Free electrons are so named to distinguish them from the bound electrons of the proton or positive nucleus of an atom. (See *Electron,* also *Proton.*)

FREE MAGNETISM—The portion of magnetism which does not follow the magnetic path through the metal of a magnetized body, but leaves the surface of the magnet. Sometimes referred to as *surface magnetism.* (See *Magnetism.*)

FREE OSCILLATIONS—Oscillations having the natural frequency of the circuit in which they occur. (See *Forced Oscillations.*)

FREQUENCIES—The plural of the term denoting the number of complete cycles per second in an alternating current. In referring to vibrations, particularly audible or so-called *voice frequencies,* the term is used to imply all frequencies that are normally considered audible. (See *Frequency.*)

FREQUENCY—A term adapted from academic English to use in electrical practice. Thus, the number of times a certain action is repeated per unit of time. In the case of an *alternating current* (q.v.) the current reverses its direction of flow at certain definite intervals. If the reversing process takes place 120 times per second—that is 60 reversals from negative to positive and vice versa—the frequency of the current is said to be 60 cycles. In ordinary commercial practice the majority of alternating current generators produce an alternating current or E.M.F. having a frequency from 25 cycles per second to 60 cycles per second. In radio transmission the alternating current or E.M.F. (Electro Motive Force) has an infinitely higher frequency—the reversals taking place many thousands of times each second. (See *Alternating Current, Theory of Production,* also *Alternation, Cycle* and *Periodicity.*)

FREQUENCY, AUDIO—In radio, a frequency within the limits at which audible sounds are produced. It is generally considered as a frequency less than 10,000 cycles per second, although some authorities refer to frequencies up to 16,000 or 20,000 cycles as being within the limits of audibility. For ordinary practice, the human ear is capable of hearing sounds only up to about 2,500 cycles per second. In receiving sets the audio frequency currents appear only in the circuits carrying the rectified currents, although some types of super-heterodyne receiver use an intermediate frequency as low as 30,000 cycles per second, these signals then being rectified and made audible in the telephone receivers. (See *Frequency Circuit, Audio,* also *Frequency, High.*)

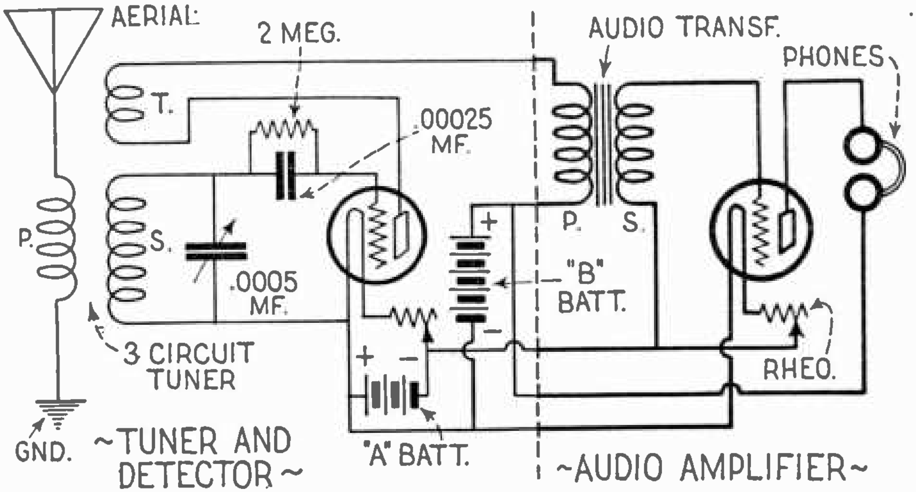

In the diagram above, the heavy lines to the right of the dotted line indicate the audio-frequency circuit of a radio receiver.

FREQUENCY, ALTERNATOR OF—The frequency of an alternating current dynamo or alternator is expressed in cycles per second. With each complete revolution of the *armature* a complete *cycle* or double *alternation* of current is generated. (See *Alternator.*) It is therefore a simple matter to determine the frequency of alternating current generated when the number of revolutions per second and the number of field poles or magnets is known. The frequency will be $f=\frac{N \times S}{2}$, N being the number of field poles and S the number of revolutions per second. The divisor 2 is used to convert the result in *alternations* per second to cycles per second, one cycle representing two alternations. (See *Alternating Current, Theory of Production of,* also *Frequency.*)

FREQUENCY BAND—A term used to denote all the frequencies lying within a certain minimum and maximum limit. Thus, in broadcasting, the frequency band will be the frequencies corresponding to the wave-lengths within the broadcast band. This will be generally between 550,000 and 1,500,000 cycles, or a wave-length band of approximately 200 to 550 meters. (See *Band of Frequencies.*)

FREQUENCY CARRIER—In radio broadcast transmission, the radio frequency (high frequency) wave with which the audio frequency speech variations are superposed. (See *Broadcasting, General Treatise on Methods,* also *Carrier Wave.*)

FREQUENCY CHANGER—A device for converting currents at a certain frequency into currents of a higher or lower frequency, without change of phase or voltage. A frequency converter. Generally an *alternator* driven by a *synchronous motor.* The term is also applied erroneously to certain forms of super-heterodynes. (See *Ultradyne.*)

FREQUENCY CIRCUIT, AUDIO—Any circuit in a radio receiver or transmitter which carries currents within the arbitrary limits of audibility. In the diagram shown herewith, the heavy lines are the connections of the audio frequency circuit. In this part of the circuit only currents of audio frequencies are present. (See *High Frequency Component of Plate.*)

FREQUENCY CIRCUIT, RADIO—The circuits in a receiver in which only high or radio frequency currents are supposed to be present. (See *Feed-Back, Audio Frequency,* also illustration in connection with *Frequency Circuit, Audio.*)

FREQUENCY, FUNDAMENTAL—The frequency of the fundamental wave. (See *Fundamental Frequency.*)

FREQUENCY, GROUP—The frequency of the separate trains of waves in radio communication. (See *Group Frequency,* also *Wave Train.*)

FREQUENCY, HIGH—Frequencies over certain definite limits, generally from 10,000 to 20,000 cycles per second in radio usage. Commonly used as synonymous with radio frequency. It is the opposite of low or audio frequency. (See *Frequency,* also *Audio Frequency, Limits of Audibility.*)

FREQUENCY, LOW—Frequencies below certain limits, generally 10,000 or 20,000 cycles per second. (See *Frequency; Limits of Audibility.*)

FREQUENCY METER—A device for indicating the frequency of an alternating current in cycles per second. The most simple form uses a system of tuned, vibrating reeds. The illustration shows a frequency meter of the

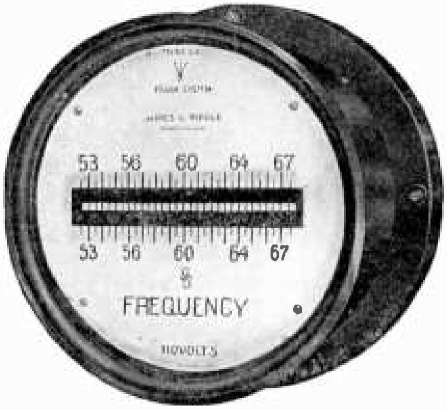

Illustration by Courtesy of Jas. G. Biddle.
A vibrating reed type of frequency meter.

reed type. These reeds may be arranged to respond to various frequencies in a similar manner to the reeds in various musical instruments. These reeds are acted upon by an alternating current magnet. When the magnet is acted upon by an alternating current sent through the windings, the reed nearest to that frequency will vibrate in sympathy with the vibration due to the alternating magnetic effect on the magnet. Such meters and also those of the *induction type* are used for comparatively low frequencies such as occur in commercial lighting systems. For measurement of the high frequencies used in radio communication somewhat different systems are used.

FREQUENCY, RADIO—An alternating current or electromotive force above certain definite limits, generally above 10,000 or 20,000 cycles per second. The line of demarcation between *audio frequency* and *radio frequency* is not very sharply defined, as, in radio usage frequencies that are actually inaudible to the human ear may yet be referred to as audio frequencies. The term is used to differentiate between the rectified or audio currents and the high frequency oscillations such as are impressed on the antenna in receiving or transmitting. (See *Audio Frequency,* also *Frequency, Low.*)

FREQUENCY, UNIT OF—Frequency of alternating currents is expressed in cycles per second and is generally written f. (See *Frequency,* also *Cycle.*)

FREQUENCY, VARIATION OF POWER FACTOR WITH—When a condenser is charged and *dielectric absorption* (q.v.) takes place, it is always accompanied by a loss of power which is apparent in the form of heat produced in the condenser. This loss of power varies with the frequency of the current impressed on the condenser and it is therefore said that the *power factor* (q.v.) varies with the frequency. (See *Condenser, also Dielectric Absorption.*)

FULL LOAD CURRENT—The current delivered from any electrical source, such as a dynamo or generator operating at its maximum load. The maximum current obtainable from an electric source.

FULL WAVE RECTIFICATION — In rectifying alternating current, the process of rectifying both the positive and negative alternations or half cycles. The majority of rectifiers for charging storage batteries rectify only a half wave, or half cycle, and thus do not utilize the full alternating current cycle. (See *Rectifier, Full Wave,* also *Charger, Storage Battery.*)

FULLER CELL—A type of primary cell formerly much used in telephone, telegraph or any intermittent work. (See *Cell, Open Circuit.*) A zinc electrode cast in the shape of a cone is at the end of a soft copper wire and rests at the bottom of a porous cup immersed in a dilute solution of sulphuric acid. The carbon electrode is in an outer jar in a bichromate solution. (See *Cell.*)

FUNDAMENTAL FREQUENCY — The frequency or number of cycles per second of the fundamental wave. (See *Frequency,* also *Fundamental Wave.*)

FUNDAMENTAL UNITS—Units based on the fundamental quantities of length, mass and time. Under the *C. G. S. system* the unit of length is the centimeter; of mass, the gram; and time, the second. The absolute system of units. (See *Centimeter-Gram-Second.*)

FUNDAMENTAL WAVE—In an alternating current, the component of the wave form that is a pure sine wave of the principal frequency. Upon this principal frequency the various higher frequencies (harmonics) may be superposed. A fundamental wave when combined with other frequencies which are a whole multiple of it produce the harmonious note. These terms are derived from acoustics in which the fundamental is the pure sound wave or pure note. (See *Sine Wave of Alternating Current or EMF.*)

FUNDAMENTAL WAVELENGTH — Term used to denote the wave-length at which the inductance and capacity of a coil or antenna would be in resonance. In referring to antennas the fundamental wave-length is the wave-length corresponding to the frequency at which free oscillations will occur. This is often referred to as the natural wave-length and the corresponding frequency is referred to as the natural frequency of the antenna.

For practical purposes the fundamental wave-length of an aerial (antenna) may be computed from the dimensions. Where no loading inductance is used, a single wire flat top antenna (T or inverted L type) will have a fundamental wave-length approximately four times the distance from the ground connection to the extreme end of flat or horizontal portion as measured in meters. Thus, if the antenna is of the L type with a top one hundred feet long and a lead-in and ground connection sixty feet long, the fundamental wave-length will be about $4\times(100+60)\times1.1$, the result being given in feet. Now as wave-length is always stated in meters it will be necessary to divide this result by 3.28 which is the rough conversion factor from feet to meters. The multiplier 1.1 adds ten per cent to the figures for necessary correction of length. If the antenna is of the T type, the length of the flat portion would be considered as half of the total horizontal length or 50 feet instead of 100 feet. The illustration shows method of measuring length. (See *Wavelength Calculation for Antennae.*)

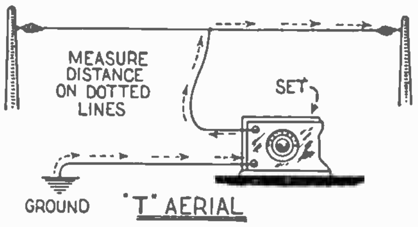

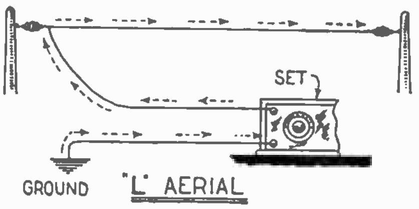

FUSE—A device used in electrical circuits to prevent the passage of excessive currents. Generally a strip of fusible metal, commonly lead with a small percentage of tin, enclosed in a fiber or insulated and fireproof container and inserted in the circuit. When the temperature, due to excessive current, reaches a certain limit, the fuse "blows" causing a break in the circuit. It is thus possible to limit the current flowing in a circuit to a definite maximum in order to protect instruments from excessive charges. The most common use is in commercial lighting and heating systems where the current on any line is generally limited to ten or twenty amperes for the protection of the lines leading in to the building.

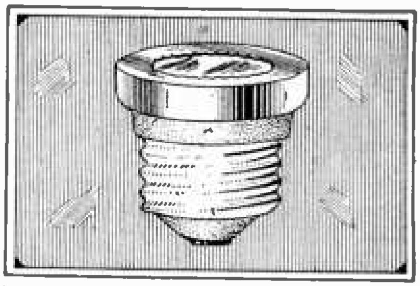

Fig. 1. A common type of fuse used in lighting, heating and power circuits.

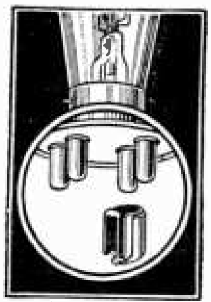

Illustration by Courtesy of Radio Equipment Co.

Fig. 2. A fuse used to protect the filament of a vacuum tube from damage by excessive voltage.

G

GALENA—A natural crystal sulphide of lead. Chemical symbol PbS. It is an ore of lead, grey in color and showing brilliant metallic lustre when fractured in cubical form. Galena crystals are used either alone or in conjunction with other minerals as *crystal rectifiers* (*detectors*) in radio reception. When used in combination it is generally in conjunction with *graphite* or *zincite.* Used alone, the contact may be a fine wire or *cat whisker.* (q.v.) (See *Crystal Detector.*)

GALVANIC BATTERY—The term, now obsolete, formerly applied to a primary battery or group of primary cells.

GALVANI, LUIGI. Born 1737, died 1798.—An Italian physician and physiologist noted as the discoverer of galvanic or current electricity. While dissecting the legs of a frog, they came by accident into contact with dissimilar metals which caused muscular action in them. Galvani thought that he had discovered electricity in animal matter. Volta attributed the action to the metallic contact and thereupon constructed his voltaic pile, the forerunner of the primary cell.

GALVANOMETER—An extremely sensitive instrument used for detecting and measuring electric currents and also for indicating the direction of the current in a circuit. In its simplest form it consists of a small iron needle pivoted in the center of a hollow coil of wire. The needle is controlled by the earth's field and points north and south when at rest. When an electric current is passed through the galvanometer winding the needle is deflected over a graduated scale the extent of the deflection depending on the force of the current. The instrument is arranged in such a manner that the needle will be deflected in one direction when the current flows through the windings in a certain direction, and when the current is reversed the needle will be deflected in the opposite direction. Thus when the connections are known the direction of the current is accurately indicated as well as its strength. The galvanometer is one of the most sensitive current measuring devices, the deflection of the needle being due to the electro-magnetic field created by the passage of electric current through the coil of wire. (See *Ammeter* or *Ampere Meter*, also *Galvanometer, D'Arsonval* and *Galvanometer, Mirror.*)

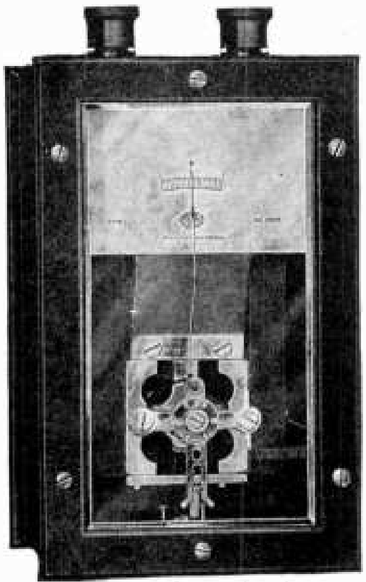

Illustration by Courtesy of the Roller-Smith Co.

A galvanometer which is used for measuring electric currents, and also for indicating the direction of flow of current in a circuit.

GALVANOMETER CONSTANT—Generally speaking, a number used to change the galvanometer reading to ordinary units. The instrument measures extremely small currents and this number times a certain function of the galvanometer reading gives the value in customary units. Technically, the strength of the field produced at the center of the coil of a galvanometer by the unit of current to be used. (See *Constant.*)

GALVANOMETER, D'ARSONVAL—A form of galvanometer having a dead beat indicator (see *Dead Beat Instrument*) in which the indicating coil is suspended in the field of a powerful horseshoe magnet. It is one of the most sensitive forms of galvanometer. The name was derived from that of the designer, A. D'Arsonval. (See *Galvanometer.*)

GALVANOMETER, MIRROR—A type of galvanometer, the moving coil or indicator of which carries a mirror which reflects a beam of light from a fixed lamp. It is so arranged that the lamp is provided with a lens and the beam is so reflected that a visible spot of light is seen on a scale and deflected when the moving element (carrying the mirror) swings with the passage of current. In the beam of light is a sharp image of fine wire stretched in front of the lamp. This type of galvanometer may also be arranged so that the reflection of a fixed scale may be observed in the mirror by means of a small telescope arrangement. (See *Galvanometer.*)

GAMMA—Γ—The third letter in the Greek alphabet. (See *Gamma Type Antenna.*) The lower case *Gamma* γ is used as a symbol for *conductivity.* (q.v.)

GAMMA TYPE ANTENNA—An alternate name for the inverted L type of antenna, so called because the third letter in the Greek alphabet, Γ, resembles an inverted "L," and a single wire antenna with lead-in from one end is in the shape of an inverted L. (See *T Type Aerial.*)

GAP—See *Spark Gap*, also *Lightning Gap.*

GAP, AIR—Term applied to the gap between sections of an iron core as in a transformer for the purpose of preventing *saturation.* (q.v.) Also applied to speak dischargers where a space is provided between the electrodes across which the spark jumps in operation. The term is used as well in referring to the radial depth of the space between surfaces of the rotor and stator in generating machines or motors. (See *Spark Gap*, also *Rotary Gap.*)

GAP ARRESTER—A form of *lightning arrester* (q.v.) in which small air gaps are provided between non-arcing metal electrodes. (See *Arrester.*)

GAP, MICROMETER—A small gap designed to protect apparatus from excessive potentials. (See *Safety Gap.*)

GASEOUS—Anything of the nature of a gas, having no natural physical boundaries but accommodating itself to the shape of the container, expanding to the dimensions of any space in which it may be confined.

GASEOUS TUBE—A term sometimes applied to a vacuum tube in which a certain amount of gas is present. Such tubes under ordinary conditions are not efficient as amplifiers in vacuum tube circuits, the minute gas atoms presenting a more or less effective obstacle to the free passage of the electrons. (See *Vacuum Tube, Theory of Operation.*)

GAS RECTIFIER—The term applied to a vacuum tube which is only partially exhausted of its residual gas. They are also known as *soft* tubes and are generally used as detectors although the tendency now is toward *hard* or highly evacuated tubes due to their adaptability to all purposes in radio receivers whereas the soft tubes do not function well as *amplifiers.* (See *Vacuum Tube.*)

GASSING—When the charging current is continued after a storage battery has taken a normal charge the free liberation of oxygen and hydrogen takes place. This is manifest in the form of gaseous emanation or liberation of bubbles from the electrolyte. The same effect may be noted when a battery is charged at higher rate than the normal charging rate stated by the manufacturer, and under certain conditions may occur at the later stages of charging, even when the battery is not fully charged. In a storage battery that is in good condition it is a fairly accurate indication of a full charge. (See *Charging*, also *Rate of Charging.*)

GAUGE, WIRE—The various systems of measurement used for denoting the physical dimensions of wire. The standard gauge in the United States is known as the B & S or Brown and Sharpe gauge. (See *Brown & Sharpe Guage.*)

GAUSS—The term adopted by the American Institute of Electrical Engineers as applying to the *magnetic flux density.* (q.v.) It is written B or $\mathcal{B}$ and is also used in conjunction with the *Gilbert* to denote magnetizing force, one gauss being one gilbert per Centimeter. It is incorrectly used to apply to *magneto motive force.* (See *Centimeter Gram Second or C. G. S.*)

GENERATOR—Any machine for producing electric energy. Usually a device for converting mechanical power into electrical energy. It may apply to a vacuum tube used to generate oscillations as in transmission or for *heterodyne action.* (See *Heterodyne*, also *Vacuum Tube Generator.*)

GERNSBACK, HUGO—Born 1884. Educated Ecole Industrielle, Luxembourg; Technikum, Bingen, Germany. President, The Experimenter Publishing Company, Publishers of Radio News, Science and Invention, Amazing Stories and Radio Internacional. President of The Consrad Company, Publishers of Radio Review, Radio Listeners' Guide and Call Book, and Money Making. Editor of Radio News, Science and Invention, Amazing Stories. Author of "Radio for All," "Wireless Telephone," "Ralph 124C-41+."

Hugo Gernsback.

GERNSBACK, SIDNEY—Born 1876. Educated College of Luxembourg, Lycee de Nancy, France. Vice-President and Treasurer of The Experimenter Publishing Company, Publishers of Radio News, Science and Invention, Amazing Stories, and Radio Internacional. Vice-President and Treasurer of the Consrad Company, Publishers of Radio Review, Radio Listeners' Guide and Call Book, and Money Making. Editor of Radio Review, Radio Listeners' Guide and Call Book, and Money Making. Author of “A Thousand and One Formulas,” “Wireless Course in Twenty Lessons,” “Experimental Electricity Course,” “Radio Encyclopedia.”

Sidney Gernsback.

GILBERT—Symbol *F*, is the unit of magneto-motive force adopted by the American Institute of Electrical Engineers. (See *Magneto-motive Force.*) It is equivalent to $4\pi/10$ or .7958 ampere turns. (See *Ampere Turns.*)

GLASS PLATE CONDENSER—A condenser, generally of the heavy duty fixed type, consisting of alternate layers of metal or foil and thin glass sheets as the *dielectric.* Glass plate condensers are mainly used for transmitting, glass having a very high dielectric constant. Such condensers may be immersed in oil to stand higher potentials without breaking down. (See *Condenser,* also *Dielectric Coefficient and Constant.*)

GLASS SILENCER—A glass tube having wooden ends with holes just permitting discharge electrodes or rods to pass, used in conjunction with spark coils and gaps to enclose the sparking surfaces and thus dampen the noise of the sparking. (See *Spark Gap.*)

GLOW DISCHARGE—See *Corona.*

GLIDING THEORY—A theory regarding the propagation of electromagnetic (radio) waves, which assumes the waves as following the curvature of the earth and not being subject to reflection due to objects in their path or by any reflecting surface above the earth. This theory has been discarded; due to observations of phenomena of transmission it now is generally assumed that there must be a reflecting surface high in the atmosphere. However, it is considered probable that the waves do not extend very high as a general rule, the exceptions being the cases of extreme distance transmission in certain directions. In the upper atmosphere, due to the rarification of the atmosphere *ionization* (q.v.) is very pronounced and the air may become a fairly good conductor. This conducting region or layer may impose severe obstacles to the electromagnetic waves and thus limit them to a certain restricted zone. In the majority of cases this ionized region may thus be regarded as damping or absorbing the waves, and in others it may conceivably intensify or concentrate them in one direction, producing instances of transmission over extreme distances with low power. For more complete discussion along these lines see *Heaviside Layer,* also *Refraction and Reflection of Electromagnetic Waves.*

GOLD LEAF ELECTROSCOPE—A very sensitive device for determining the presence and nature of an electric charge. It comprises, essentially, two strips of gold foil joined at one end by a conductor and suspended in a glass jar. When an electric charge is present the two strips of foil have a tendency to repel one another and thus, by diverging, indicate the presence of the charge. (See *Galvanometer.*)

GOLDSCHMIDT ALTERNATOR—One of the first practical *alternators* for generating *high frequency currents* suitable for *continuous wave* radio transmission. It was designed by *Rudolph Goldschmidt* in 1912 and was used mainly in Europe. This machine operated on a frequency step-up principle. A single phase alternating current was first produced, the frequency being about ten thousand cycles. This was about the highest possible frequency obtainable with an ordinary type of alternator. By means of a *cascade* effect, obtained by an arrangement of tuned circuits connected to stator and rotor of the alternator, a much higher frequency could be impressed on the antenna to radiate continuous waves. The machine had a number of disadvantages which limited its commercial value, chiefly the fact that the air gap between rotor and stator was necessarily very small—a difficult matter with a bulky machine, the rotor of which weighed several tons. (See *Alternator, Continuous Waves,* also *Goldschmidt, Rudolf.*)

GOLDSCHMIDT, RUDOLF—Born March 19, 1876, at Neu-Buckow, Mecklenburg, Germany. After finishing his education at Wiemar Municipal School, he studied engineering at Charlottenburg and Darmstadt Technical High School. In Darmstadt he obtained his degree as electrical engineer in January, 1898, and then became assistant to Professor Kittler. In 1900 he obtained the college and travelling scholarship, which enabled him to visit engineering works in Belgium, England and France. Later in the same year he was appointed engineer in the laboratory of the Allgemeine Elektrizitaets Gesellschaft in Berlin. In 1901-2 he occupied the position of chief laboratory engineer and designer to Kolben & Co., Ltd. in Prague. He came to England in connection with the Willesden Electricity Supply Station and was later appointed chief engineer to Messrs. Crompton & Co., of Chelmsford. In 1905 he joined the Westinghouse Company in Manchester. After private preparation he passed the German abitur-examination and obtained the degree of Dr. Eng. In 1907 he returned to Germany as lecturer at Darmstadt Technical College. Here he practiced as a consulting engineer, and also pursued the development of several inventions, chiefly occupying himself with the invention and design of high-frequency alternators for wireless telegraphy. In 1911 he became manager of the “Hochfrequenz-Machinen Aktiengesellschaft fur drahtlose Telegraphie” in Berlin, a company formed for the utilization of his inventions in wireless telegraphy. In this position he established two large wireless stations at Eilvesen, Province of Hanover, and Tuckerton, New Jersey, for wireless communication between Germany and America.

GOLDSMITH, ALFRED NORTON—Elec. and Radio Engr.; b. New York, N. Y., Sept. 15, 1887; B.S., Coll. of the City of New York, 1907; Ph.D., Columbia Univ., 1911; cons. radio expert, U. S. Dept. of Justice, 1912; cons. radio engr., Atlantic Communication Co., 1914; cons. engr., Gen. Elec. Co., 1915-17; dir. of research, Marconi Wireless Telegraph Co. of America, 1917-19; associated professor of electrical engineering, College of City of New York, since 1924; dir. of research dept., Radio Corp'n of America, 1919-1922; Chief Broadcast Engineer, Radio Corp'n of America, since 1922; editor, Proceedings of the Institute of Radio Engrs., since 1912. Made investigations in simplex and duplex radio telegraphy and telephony, transmission of canal rays, precision measurements in radio eng. Author: “Radio Telephony” (Wireless Press), 1918; “Radio Measurements,” “Radio Frequency Changers” (Proc. Inst. Radio Engrs.), 1915; “World Communication” (Jour. A. I. E. E.), 1921. Technical director, U. S. Signal Corps School of Communication, 1917-18; U. S. Naval Radio School, 1917-18. Fellow, A. I. E. E., I. R. E., American Association for Advancement of Science. Hon. mem., Radio Club of America, Radio Society of Great Britain, Am. Physical Soc.

Dr. Alfred N. Goldsmith.

GONIOMETER—A device used in radio reception to determine the direction from which the electromagnetic waves (signals) are coming. The term is also used at times to designate any system of directional transmission or reception. The device was originally designed as an aid to navigation.

The original goniometer was designed by Messrs. Bellini and Tosi (see *Bellini, Dr. Ettore,* also *Tosi, A.*), but numerous changes have since been made, notably by Marconi and his associates, to adapt it to the requirements of ships at sea. The form most used is based on the original Bellini-Tosi machine and is essentially a special form of *oscillation transformer* used in conjunction with a special aerial or loop. As an example of the use of a goniometer in connection with ships at sea, suppose a ship is in distress at a point about two hundred miles off the eastern coast. Perhaps the ship's radio apparatus is working, but for some reason the mariner does not know the ship's position. The radio signals are then received, let us say, by three different land stations at widely separated points. The receiving stations use their goniometers and determine more or less accurately the direction from which the signals are coming. The three stations compare notes on the respective directions and any one of them can readily plot out lines running on a chart from each station in the directions indicated by the goniometers. Then by simple triangulation the location of the ship can be figured, enabling other ships in the vicinity to be instructed accordingly and go to their rescue. There are, of course, many other uses that occur in everyday practice. Goniometers are also much used by radio inspectors to locate stations that are interfering, using an unlicensed transmitter, or are otherwise not complying with the regulations.

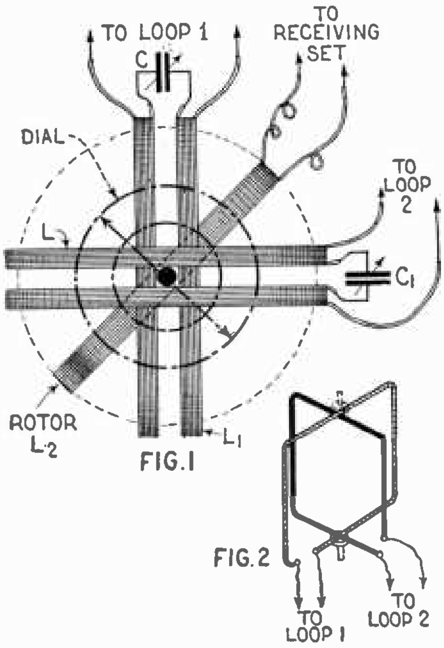

Fig. 1 shows the general arrangement of a goniometer and Fig. 2 shows the loops to which coils L and L-1 are connected.

The illustration Fig. 1 shows the general arrangement of the instrument. Here, coils L and L1 are the main windings of the goniometer. They are wound in two sections with condensers C and C1 inserted between the sections of the two windings for purposes of tuning. Coil L2 is the *exploring coil.* This is connected to the receiving apparatus, which is the customary detector circuit. Coils L and L1 are connected to the two loop aerials shown in Fig. 2. Here it will appear that the loops are placed at right angles, their relation in this respect being variable. In operation the coils L and L1 have minimum *mutual induction*, as they are placed at right angles. The exploring coil L2 can be rotated through a complete circle and the pointer indicates the angle with relation to the fixed coils. When one of the two loops points in the direction of a transmitting station it will pick up the signals, the other being at that moment inoperative, as it is at right angles to the first, as previously explained. If the exploring coil L2 is now turned until maximum signal strength is obtained, the pointer will be in the direction of the plane of the loop. If the advancing wave strikes the two loops simultaneously at any angle, a resulting magnetic field will be set up in the coils L and L1 of the goniometer, the field having a direction at right angles to the direction from which the waves are advancing. Now when the exploring coil L2 is turned to allow maximum signal strength due to this field (permitting maximum inductive transfer), the pointer will be in the direction of the transmitting station. The direction indicated by the pointer is, of course, not the geographical direction, but this can readily be determined by means of a compass. It also appears from the foregoing description that the pointer will indicate two directions, which, however, will be on the same geographic plane. In other words, the pointer may be, for example, exactly north and south.

GRAMME ARMATURE—A form of ring type *armature* (q.v.) used in generators, named for its originator. (See *Generator.*)

GRANULAR CARBON—The carbon grains used in a telephone transmitter to vary the resistance between the electrodes. The highest grade is made by carbonizing pure anthracite coal in an electric furnace, after which it is screened and graded. (See *Microphone, Carbon.*)

GRANULAR COHERER—A form of *coherer* (q.v.) used in the early days of wireless. It employed carbon granules between two electrodes, the whole enclosed in a glass tube. (See *Detector.*)

GRAPH—The presentation in diagrammatic form of facts, formulae, etc., particularly in the electrical and mechanical fields. Graphs are much used in radio to give graphic illustrations of various functions, such as the relation of currents and voltages. For this purpose ruled and squared paper is used. One of the most common and useful forms of graph is known as a characteristic curve of a vacuum tube. These curves may show the relation of changing values, such, for instance, as the relation of grid voltage to plate current, indicating certain of the operating characteristics of the vacuum tube. (See *Characteristic Curve*, also *Sine Curve.*)

GRAPHITE RESISTANCE—A form of resistance composed of graphite, used in any of several ways in radio circuits. The most common usage is in the form of a *grid leak* (q.v.). Here it may take the form of a thin sheet of paper coated with graphite or a glass tube filled with graphite and some binding medium. Such resistances are also used in *resistance coupled amplifiers* (q.v.), in which case they are interposed between the plate and "B" battery and couple the successive stages of amplification. (See *Resistance*, also *Resistance Coupled Amplifier.*)

GRASSOT FLUXMETER—An instrument used to measure the strength of magnetic fields and the regions around magnetic poles. It is of particular value in measuring the relative strengths of various magnets. The instrument comprises a very sensitive galvanometer of the moving coil type with a special exploring coil. This coil is introduced into the magnetic field to be measured and readings of the proportionate strength of the field are obtained directly from the scale of the galvanometer, which is appropriately divided. The field strength is obtained by dividing the deflection value in *Maxwell turns* (q.v.) by the area of the search or exploring coil in square centimeters and the number of turns of the coil. (See *Galvanometer*, also *Flux Density.*)

GRAVITY BATTERY—A combination of gravity cells. A group of cells, each cell having two different electrolytes which remain separated without use of a diaphragm due to their difference in specific gravity. (See *Gravity Cell.*)

GRAVITY CELL—A type of *primary cell* (q.v.) in which the specific gravity of one electrolyte is less than the other, which causes this lighter electrolyte to float on top of the heavier fluid. One of the typical examples of gravity cells is the *Daniell's Cell* (q.v.). (See *Secondary Cell.*)

GRID—Generally, any system of wires placed parallel to each other and in the same plane, as a mesh. Specifically in radio usage, the fine network of wires interposed between the filament and plate elements in a vacuum tube. It may be in the form of a flat network of wires or in the form of an oval cylindrical mesh completely surrounding the filament and in turn surrounded by a similarly shaped plate of metal. The illustration shows the form taken by a typical grid member in a modern type of vacuum tube. Here the plate is not shown in order to permit the grid member to show more clearly.

In receiving circuits the grid acts as a valve to control the flow of electrons from hot filament to cold plate. When signals are impressed on the aerial they are tuned by the customary agencies and passed to the *grid circuit* (q.v.). As the incoming signals are in the form of *high frequency oscillations* (q.v.) changing their polarity (direction) many thousands of times each second, it will be apparent that the grid is altered rapidly from positive to negative and back. As the electrons flowing from the filament are negative, the grid will thus have a tendency to retard or increase the flow. That is to say, at the instant the grid is positive it will attract more electrons, but when it is momentarily negative the like charges will repel each other and less electrons will flow, some of them being held back to the filament. (See *Grid Battery, Grid Bias*, also *Vacuum Tube, Theory of Operation of.*)

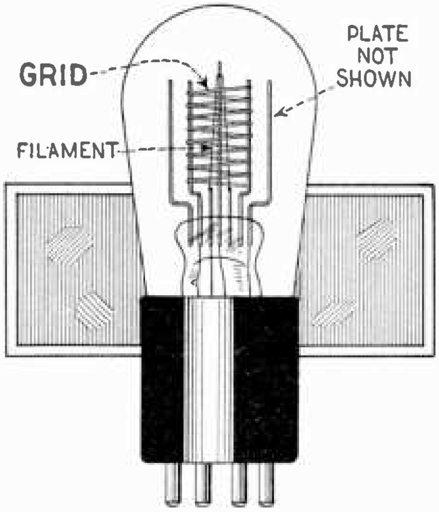

Showing the form taken by a typical grid in a modern vacuum tube.

GRID BATTERY—A battery of small cells connected in the grid circuit of a vacuum tube to place a negative or positive bias or charge on that element. The illustration Fig. 1 shows a small four and one-half volt battery inserted in such a manner that a negative potential is impressed on the grid. In this case the *characteristic curve* (q.v.) of the tube is assumed to indicate that more difference in current occurs between points B and C than between A and B in Fig. 2. For this reason a negative potential is applied as shown in Fig. I. (See *Grid Bias.*)

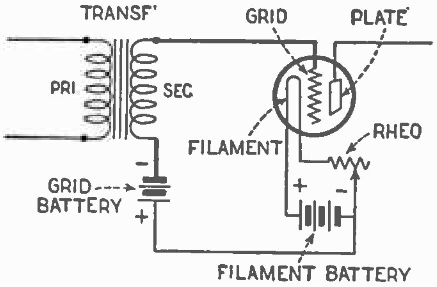

Fig. 1. Method of a 4½ volt battery to impress a negative potential on the grid of a vacuum tube.

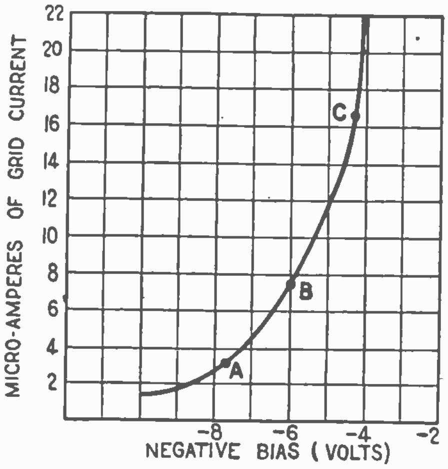

Fig. 2. Characteristic curve of a vacuum tube indicating more difference in current occurs between "B" and "C," than between "A" and "B."

GRID BIAS—A potential of a few volts, generally from four to six, applied to the grid of a vacuum tube to influence its operation by making it more or less negative. The grid bias is usually negative and determines the point of the *characteristic curve* at which the tube will operate. In a sensitive receiver, and particularly where a tube is used as an *amplifier* (q.v.), it is essential to obtain as great a change of grid current as possible. (Note: The greater the change of grid current, the greater the change in plate current and hence the more powerful will be the output.) By applying a negative potential on the grid it is possible to hold it at the point of maximum response. (See *Grid Battery,* also *Vacuum Tube, Theory of Operation of.*)

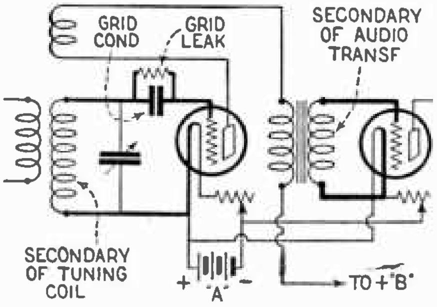

Diagram in which heavy lines at left show the grid circuit of the detector, while those on right indicate audio amplifier grid circuit.

GRID CIRCUIT—In a receiver employing vacuum tubes, the part of the circuit containing the grid of the tube or tubes, but generally referring to the tuning circuit or that part of the system which contains the tuning elements. In the illustration is shown a standard arrangement of a regenerative detector and one stage of audio-frequency amplification. The heavy lines at the left of the diagram represent the tuning circuit or detector *grid circuit,* while the section with heavy lines at the right is the grid circuit of the amplifier tube. The grid circuit of a detector tube may generally be distinguished by means of the grid leak and condenser (so marked in the illustration), which controls the incoming signals and permits the tube to function as a rectifier. (See *Detector, Grid Control,* also *Grid Condenser* and *Grid Leak.*)

GRID CONDENSER—A small fixed condenser, generally from .00025 to .0005 microfarad capacity, inserted in the circuit of a detector tube between the tuning coil and the grid member of the tube. The illustration shows the manner of connecting the grid condenser in a conventional detector circuit. This condenser insulates the grid from the filament by breaking the path F to G and permits the tube to act as a rectifier or detector. A resistance (called *grid leak*) is usually placed across this condenser to allow the accumulated charges on the grid to leak off. (See *Grid Leak,* also *Grid Control* and *Vacuum Tube, Theory of Operation of.*)

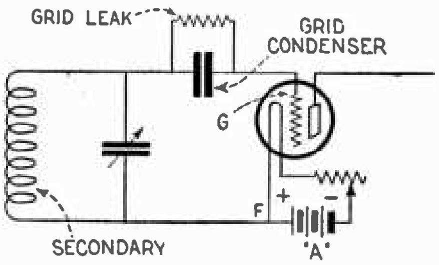

Showing how grid condenser is connected in a detector circuit.

GRID CONTROL—The general term used to designate the various devices and associated connections placed in or allied with the grid circuit of a vacuum tube for the purpose of controlling to a greater or lesser extent the potential of the grid, or to tune or control the incoming oscillations. In the case of a vacuum tube used as a rectifier (detector) the grid condenser and grid leak are placed in series with it. (Note: the grid leak is often connected from the grid directly to the filament instead of across the grid condenser, the operation, however, being approximately the same in either case.) Here the controls permit the tube to function as a detector of the incoming oscillations. Where the tube is used as a radio-frequency (high-frequency) amplifier, the radio-frequency transformer or other amplifying unit controls the grid of the following tube by influencing the fluctuating potential differences impressed on it. (See *Amplifier, Radio-Frequency.*) The most important example of a grid control is by using some means to control the potential applied to the grid, such, for example, as a grid bias or grid battery. (See *Grid Bias, Grid Battery,* also *Vacuum Tube, Theory of Operation of.*)

GRID CURRENT—The current present in the grid circuit of a vacuum tube. This is generally very small, perhaps of the order of fifteen to twenty *micro-amperes* (q.v.). (See *Plate Current,* also *Grid Bias.*)

GRID LEAK—A high resistance placed in the grid circuit of a vacuum tube to permit the electrons forming the grid current to leak off after each charge, thus preventing their accumulating on the grid in such numbers as to stop the flow from the filament. The electrons being negative, a surplus of them on the grid member of the tube would act as a barrier to the free flow of other electrons from the filament through the grid to the plate, thus diminishing the plate current toward zero. It will be understood that a heavy negative charge on the grid will repel a like or negative charge, in this case the negative electrons being thrown off by the hot filament.

The grid leak may have a value ranging from about 250,000 ohms to several million ohms. The value of a grid leak is generally stated in *megohms*—one megohm being one million ohms. Thus a grid leak of 250,000 ohms resistance would be referred to as a ¼-megohm leak and, similarly, one having a resistance of two million ohms would be referred to as a two-megohm grid leak.

Such resistances are furnished in many forms, the most common being a strip of paper impregnated with graphite or some similar high resistance preparation, placed in a glass tube and sealed to prevent moisture from affecting the value. The illustration, Fig. 1, shows a tubular type of grid leak.

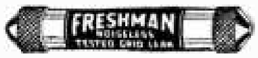

Courtesy of Chas. Freshman Co.

Fig. 1.—A tubular type of grid leak.

For experimental work a fairly satisfactory grid leak can be made by the simple expedient of drawing a line on a piece of cardboard with a soft pencil—the graphite mark acting as the resistance. At each end of the cardboard is placed a binding post in contact with the graphite, each binding post being connected to opposite sides of the grid condensers. Variable grid leaks, the resistance of which can be changed at will, are also furnished in an infinite variety of types, two of which are shown in Fig. 2. (See *Variable Grid Leaks,* also *Grid Control* and *Resistance.*)

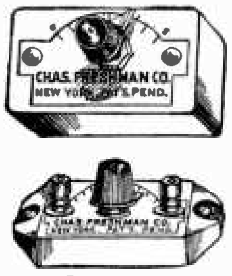

Courtesy of Chas. Freshman Co.

Fig. 2.—Two types of variable grid leaks with condenser incorporated.

GRID POTENTIAL, MODULATION OF—In the transmission of *undamped waves* (q.v.) generated by a vacuum tube, the process of varying or modulating the potential of the grid element of the tube with respect to the filament. In this method an alternating potential may be applied to the grid circuit, the frequency of this varying potential being within the audio band—possibly about 500 cycles—supplied by a generator of that freqnency. (See *Modulation.*)

GRID POTENTIOMETER—A potentiometer—a variable resistance unit—Fig. 1, used in the grid circuit of a vacuum tube for the purpose of controlling the potential applied to the grid. The conventional method is to place the potentiometer across the "A" battery with the center or variable contact arm connected to the grid of the tube in place of the usual connection to either positive or negative filament lead. This is shown in the illustration, Fig. 2, Here the customary circuit is shown for the "A" battery to light the filament of the tube. Pri. and Sec. are respectively primary and secondary of the transformer (either radio-frequency or audio-frequency) and C the center arm of the potentiometer. The center arm may be moved either to the right or left, permitting either negative or positive potential to be applied to the grid. This system is much used in amplifying tube circuits to control the action of the tube as an amplifier. Another method would be to connect one of the ends of the potentiometer to the negative lead of the "A" battery and the center arm to the grid leak. In this way the potentiometer would be acting as a pure variable resistance, the potential applied to the grid being always negative, but susceptible of variation within the limits of the ressitance and the voltage of the "A" battery. (See *Grid Bias, Grid Battery,* also *Grid Control.*)

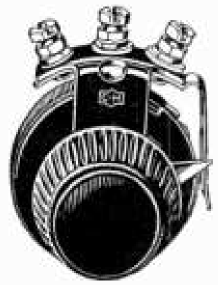

Courtesy of the Cutler-Hammer Co.

Fig. 1.—A type of potentiometer.

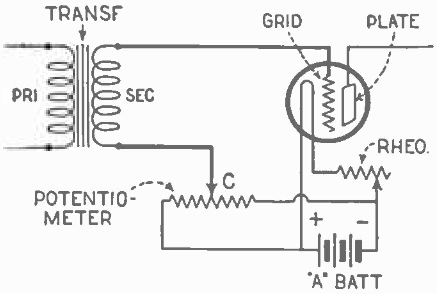

Fig. 2.—Method of connecting potentiometer in grid circuit.

GRINDER—Term occasionally applied to one of the various forms of atmospheric disturbance known under the general heading of *static.* This form of static is more prevalent in warm weather and does not appear to have its origin in local electric storms. Its effect is to produce rattling sounds of no definite pitch. (See *Strays,* also *Static.*)

GROUND—The earth considered as an electric conductor. In radio it is considered as the return circuit for electromagnetic waves. (See *Ground Connection,* also *Theory of Propagation of Electromagnetic Waves.*)

GROUND CLAMP—A strip of metal, generally copper, having some sort of arrangement for fastening rigidly to a water or steam pipe to form the ground connection for a radio receiver. The illustration shows a conventional form of clamp. (See *Ground Connection.*)

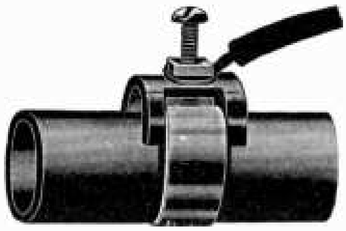

A ground clamp for fastening the ground lead to a water or steam pipe.

GROUND CONNECTION—In radio reception or transmission, the connection between the earth and the apparatus whereby the currents set up by the electromagnetic waves are returned to the earth as one side of the assumed circuit. The ground or earth connection from a radio receiver or transmitter should be as short as possible and preferably a heavy wire as shown in the illustration. Such a connection for receiving purposes may be one or more heavy metal plates sunk in the ground and connected by a heavy wire to the apparatus. The more common method is to connect the set to a convenient water or steam pipe, such pipes being almost invariably well grounded, especially the former. (See *Earth,* also *Ground* and *Transmission.*)

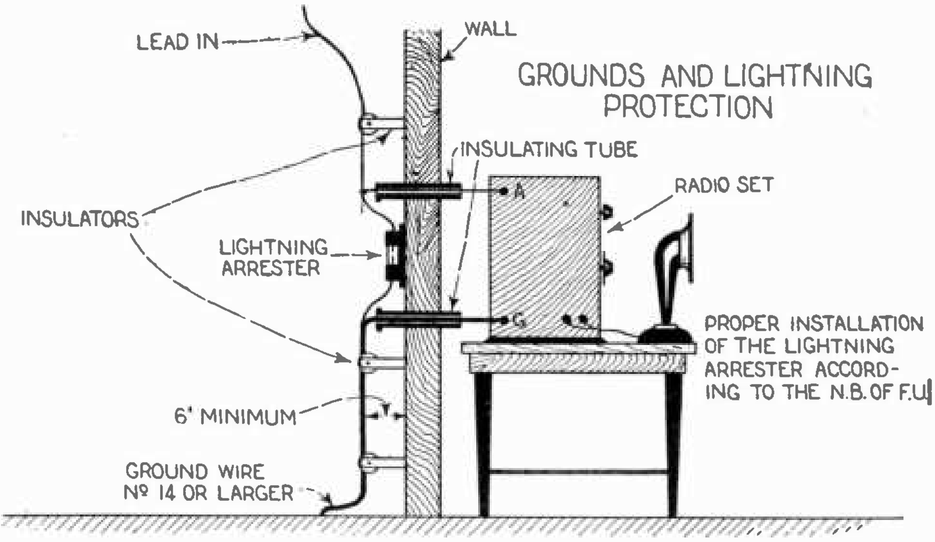
LEAD IN
WALL
GROUNDS AND LIGHTNING PROTECTION
INSULATING TUBE
RADIO SET
INSULATORS
LIGHTNING ARRESTER
A
G
PROPER INSTALLATION OF THE LIGHTNING ARRESTER ACCORDING TO THE N.B.OF F.U
6' MINIMUM
GROUND WIRE Nº 14 OR LARGER

Illustration showing by means of a heavy line, the ground connection of a radio receiver.

GROUND CURRENTS—Electric currents present in the ground due to the grounding of commercial electric machines. These currents very often prove a disturbing factor in radio reception. (See *Ground.*)

GROUND DETECTOR—A device for indicating an accidental ground on a power transmission line or for any conducting systems for electric currents. (See *Detector,* also *Ground.*)

GROUND RESISTANCE—See *Resistance of Ground Connection.*

GROUND SWITCH—A switch, generally of rugged construction, connected to the aerial, ground and apparatus in such manner that at one position the aerial is connected to the set, which in turn is connected to the ground, and in the other position the aerial being connected directly to the ground to protect the set from excessive static when not in use. (See *Lightning Switch,* also *Aerial Switch, Down Lead.*)

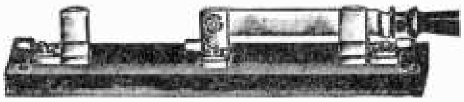

Switch having a break distance of not less than 4 inches.

GROUND WIRE—The wire connecting a receiving or transmitting set to the earth. The underwriters specify not less than No. 14 B & S gauge copper wire. (See *Ground Connection.*)

GROUNDED ROTOR—See *Rotor, Grounded Connection.*

GROUP FREQUENCY—The number of separate trains or groups of waves per second in a damped or undamped system of radio transmission. This frequency is to be distinguished from the frequency of the individual waves. Thus in a spark transmitter, for example, the condenser in the transmitting circuit will be momentarily charged and discharged. The discharges will occur in the form of trains of oscillations, each oscillation having a certain definite frequency, but the trains or groups occurring at definite intervals, the frequency of these intervals being called the group frequency. (See *Frequency,* also *Spark Transmission* and *Undamped Waves.*)

GUARD LAMP—An incandescent lamp having a straight filament, sometimes shunted across the armature and field leads of a rotary converter used in radio transmission. Its purpose is to protect the windings from injury due to the inducing of oscillatory surges from the high-frequency circuits. (See *Rotary Converter.*)

GUTTA PERCHA—An insulating material derived from the milky secretion of certain trees, used as an insulation for submarine cables on account of its ability to resist the action of salt water. It is used to a lesser extent as a general insulating medium for wires, but owing to its scarcity and limited field of usefulness, rubber or rubber compounds are more widely used. (See *Insulating Materials.*)

GUY WIRE—Wires used to brace aerial masts for receiving or transmitting stations. Such guys are generally heavy galvanized wires attached to the mast at various points and extending radially downward and anchored to surrounding rigid objects or by means of some form of anchor in the ground. (See *Aerial.*)

H

H—The chemical symbol for *hydrogen.* In electrical and radio usage it is the symbol for *magnetizing force* or *magnetic field intensity*, adopted as such by the International Electrotechnical Commission and by the American Institute of Electrical Engineers. It has also been adopted by the I. E. C. as the symbol for *Henry* (q.v.). (See *Magnetic Field.*)

HALF WAVE RECTIFICATION—The process of changing alternating currents to pulsating direct currents wherein only one-half of the full alternating current wave or cycle is rectified, the other portion being blocked by the action of the rectifier. In the illustration is shown a typical half wave rectification process. The upper graph shows the sine wave representing an alternating current, such as supplied from an ordinary alternating current house lighting system. Here A-B represents one complete cycle or two alternations, one positive and the other negative. In the lower graph is shown the form taken by the current after passing through a half wave rectifier. It now appears as a pulsating or intermittent direct current, only the upper or positive alternations having passed through the rectifier. The shaded portions below the line represent the negative or unused alternations. (See *Full Wave Rectification, Sine Wave of Alternating Current or E M F*, also *Charger, Storage Battery.*)

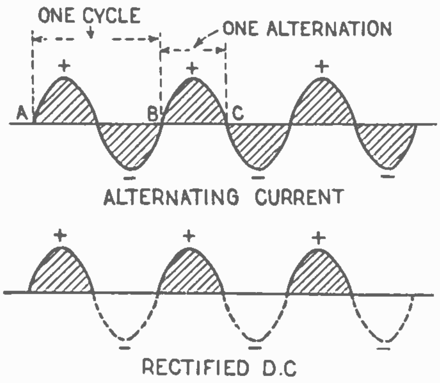

The upper graph represents an alternating current, and the lower graph shows the form taken by the current after passing through a half wave rectifier.

HALF WAVE RECTIFIER—Any system for rectifying or changing alternating currents to pulsating direct currents, wherein only half of the full alternating current wave is used. Essentially such a device is a uni-directional conductor, permitting current to pass in one direction only and submerging or blotting out the surges in the opposite direction. A typical example of a half wave rectifier is the crystal detector used in simple receiving circuits. Here the nature of the crystal is such that the currents are permitted to pass only in one direction. In half wave rectifiers used for charging storage batteries the effect is much the same. Where tubes such as the Tungar or Rectigon are employed, one tube is more commonly used, allowing the current to flow in one direction but submerging half of each cycle. (See *Rectifier, Full Wave*, also *Rectifier and Charger, Storage Battery.*)

HAMMER BREAK—A magnetically operated device for making and breaking contact in an electrical circuit. Hammer breaks are essentially the same as a vibrator or buzzer break but are made to withstand heavy currents without undue arcing, and hence are much slower in action. (See *Buzzer*, also *Spark Coil.*)

HAMMOND, JOHN HAYS, JR.—Inventor. Born San Francisco, Calif., April 13, 1888. B.S., Sheffield Scientific School (Yale), 1910; Sc.D., George Washington University. Inventor of type of torpedo for coast defense, controlled by wireless energy from coast fortifications, which was recommended to Congress for exclusive purchase by U. S. by Board of Ordnance and Fortifications, U. S. A.; invented system of automobile torpedo firing; aluminothermic incendiary projectiles employed by Allied armies in World War; radio system of control of ships, employed on U. S. S. Iowa for target practice; a system of coastal patrol by aeroplane; a system of selective radio telegraphy, adopted by the U. S. Navy, U. S. Signal Corps, and U. S Army; system of aerial coast surveying adopted by the Bartlett Expedition for Polar exploration. President Radio Engineering Company of New York; director Hammond Radio Research Labs.; director Radio Corp. of America; consulting engineer to Radio Corp. of America and General Electric Co. Has licensed the American Tel. & Tel. Co., and the Radio Corp. of America under his electrical patents in U. S. and S. A. for commercial purposes only, U. S. Gov't retaining option on military uses. Has applied for over 224 patents in U. S. and Europe relating to Radio Telegraphy and Telephony and Wirelessly controlled torpedoes and various improvements in pipe organ mechanisms. Formerly member Advisory Board, U. S. Naval Board of Inventors, member Advisory Com. Langley Aerodynamic Lab. of Smithsonian Inst., co-operating with Third Naval Dist.; member Conf. Com. on National Preparedness (Sub.-Com. on Finance); Gov. Aero Club America (mem. Aerodynamic, Tech., Pub. Safety and Map and Landing Places Coms.); v.-p. Am. Soc. Aeronautic Engineers; mem. Inst. Radio Engineers (ex-treas., etc.), Am. Inst. E. E., Royal Soc. Arts (London); asso. mem. Am. Soc. M. E.; hon. mem. Nat. Inst. of Inventors, Harvard Aeronaut. Soc.; Fellow Am. Geog. Soc.; U. S. del. Radio-Telegraphic Com., London, 1912.

John Hays Hammond, Jr.

HAND CAPACITY—See *Body Capacity.*

HARD DRAWN WIRE—In radio usage generally confined to hard drawn copper wire, i. e., copper wire that has been hardened by repeated drawing without being annealed. It is used chiefly for aerials. (See *Aerial.*)

HARD TUBE—A vacuum tube is said to be hard when it has been evacuated of all gases to a high degree. The opposite is a soft tube, one in which the vacuum is not of a very high order and more or less gas particles are present. Hard tubes are used for amplifiers, both audio and radio frequency, while the soft tubes are used mainly as detectors or oscillators. Hard tubes may be used with high plate voltages, whereas soft tubes require comparatively low plate voltages, the tubes having a tendency to turn blue (see *Blue Glow*) under the influence of high plate potentials. (See *Detector*, also *Amplifier.*)

HARMONIC CURRENT—An alternating current which can be represented by a simple harmonic or sine curve. It may be considered as an alternating current, the waves of which are sinusoidal or uniform. (See *Simple Harmonic*, also *Sine Curve.*)

HARMONIC MOTION—A term signifying any periodic oscillatory motion, as, for example, the vibration or beating of a pendulum. The illustration shows the manner of indicating simple harmonic motion. The distance A-B represents 360 degrees of a circle, or, if it is assumed to represent a sine wave of alternating current, it will represent one complete cycle. Curves representing harmonic motion are used extensively in radio engineering to illustrate, in graphic form, the oscillations of an alternating current. (See *Sine Wave*, also *Simple Harmonic Vibration* and *Oscillations.*)

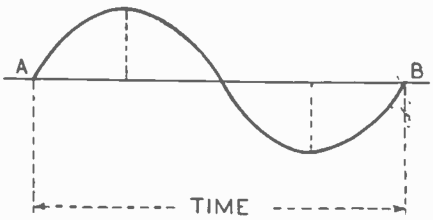

Illustration showing method of indicating simple harmonic motion.

HARMONICS—Oscillations to which a circuit will respond in addition to the basic or fundamental oscillations. The effect is similar to that in sound, where, for instance, the primary or fundamental tone struck by an instrument may be accompanied by notes or tones of a higher pitch. In sound these harmonics are principally the third, fifth, seventh and the octave. In radio, a circuit may respond to oscillations of a frequency either higher or lower than the fundamental frequency, as, for example, one third, one fifth or one seventh, or, again, double or treble the original frequency. As a concrete example, with a fundamental wave-length of 240 meters, the harmonics might be

40 meters or 80 meters. The corresponding frequencies would be approximately 1,250,000 cycles, 7,500,000 cycles and 3,750,000 cycles. (See *Harmonic Suppressor, Side Bands,* also *Harmonic Motion, Oscillations, Wave-length.*)

HARMONIC SUPPRESSOR—A device used in radio transmission and in laboratory work, for the purpose of eliminating undesired frequencies (wavelengths). Essentially such a device is a *filter* (q.v.) so arranged as to suppress or eliminate certain frequencies or *harmonics* (q.v.). In radio broadcasting it is particularly essential that all undesired frequencies or *harmonics* be suppressed. Basically the system involves a special coil shunted by a condenser and tuned to the frequency of the interfering wave, the main tuning circuit at the same time being adjusted to the desired frequency. (See *Broadcasting, General Treatise on Methods,* also *Transmission* and *Radio Telephony.*)

HAZELTINE NEUTRODYNE RECEIVER—The radio receiver designed by Professor Hazeltine, employing the Neutrodyne circuit. Essentially these are five-tube receivers employing two stages of tuned and neutralized radio-frequency amplification with the conventional detector and two audio-frequency stages. In addition to the five-tube Neutrodyne receiver there are numerous types involving the Neutrodyne principle in conjunction with *reflex systems.* (See *Neutrodyne.*)

HEAD TELEPHONE OR HEADPHONES—Telephone receivers arranged with a band to fit over the operator's head, thus leaving him free use of both hands. Head-phones are used in commercial radio work, the operator being free to handle a key or the controls of the receiver or transmitter. They are also used in broadcast reception where the strength of signals is not sufficient to permit the use of a loud speaker, and also in some cases to tune in distant signals before attaching a loud speaker. (See *Watch Case Receivers.*)

Illustration by courtesy of C. Brandes, Inc.
A type of headphone used extensively in radio reception.

HEAVISIDE, DR. OLIVER. — Born in London on May 13, 1850, Dr. Heaviside was engaged in telegraph work for a few years, but after 1874 he lived in retirement studying Clerk-Maxwell's theory and applying it to telegraph and wireless problems. Elected a Fellow of the Royal Society in 1891, he was also Faraday Medallist of the Institution of Electrical Engineers and Hon. Ph.D. of Gottingen University. The name of Dr. Heaviside is primarily associated with the theory, propounded by himself, of the existence of a permanently ionized layer in the upper atmosphere capable of deflecting electro-magnetic waves and thus permitting wireless communication round the earth.

HEAVISIDE LAYER—An assumed layer or stratum of the atmosphere supposed to exist from fifty to one hundred miles above the earth's surface according to the Heaviside Theory. Heaviside's theory supposes this stratum to consist of heavily ionized gas which is conductive and acts as a gigantic reflector of electromagnetic waves (radio waves). It is this layer which is supposed to reflect the electric waves and force them to follow the curvature of the earth. This theory partially explains—if true—why radio signals are transmitted over great distances instead of being radiated off into space and lost. It has been accorded the support of many eminent scientists. Elihu Thompson has propounded an alternative theory which assumes the waves as attached to the earth and gliding along over the surface. (See *Gliding Theory.*)

HEAVISIDE LAYER THEORY—The theory of an upper ionized stratum from fifty to one hundred miles above the earth's surface, which is assumed to act as a reflector of radio waves and thus force them to follow the curvature of the earth. None of the theories thus far advanced has been proven, but it has been definitely established that these waves do follow the curvature of the earth. (See *Heaviside Layer,* also *Gliding Theory.*)

HEDGEHOG TRANSFORMER—A type of audio transformer so called because of its peculiar construction. The core is cylindrical in shape and consists of a great number of soft iron wires around which is a bobbin holding the primary and secondary windings. After the windings are completed the wires forming the core are bent back over the form and held in place, completely surrounding the windings. (See *Audio-Frequency,* also *Transformer.*)

HEIGHT OF AERIAL—The elevation of an aerial system as used in radio transmission or reception. The effective height is roughly considered as the height above the earth or surrounding objects. The range of a station having constant power is dependent upon the height of the aerial, other factors of radiation being unchanged. In other words, increase of height increases the range within certain limits. (See *Radiation from Antenna,* also *Transmission.*)

HEISING, RAYMOND A.—Electrical Engineer, born Albert Lea, Minn., August 10, 1888. Educated University of N. D., Elec. Engineer, 1912; University of Wisc., M.S., 1914 Grad. Work in Physics, Research and Development of Radio Tel. since 1914; Trans-Atlantic test in 1915; for Army and Navy use during War, aeroplane, chaser, intership communication; since War continued research and development ship to shore use, trans-Atlantic test, 1923. Author of several technical papers; numerous inventions widely used in present broadcasting stations.

HEISING MODULATION—A system of modulating radio currents for radio transmission, such as broadcasting, developed by *R. A. Heising.* The fundamental principle is shown by the illustration. Here the customary antenna and ground are connected with a high-frequency alternator, A; the field of the alternator, FW, is connected to the grid return and filament of the vacuum tube circuit, and through the "B" battery BB to the plate P of the tube. The microphone circuit comprises a microphone M connected to the primary of a transformer T through the battery MB. The secondary of the transformer is connected to the grid G of the vacuum tube and the filament F. R is the usual rheostat to control the filament of the tube and AB is the filament lighting battery. In operation the speech or sound vibrations are impressed on the microphone M, thus varying the resistance of the current flowing from the battery MB and through the transformer T. This serves to vary the potential of the grid G, and as this potential changes, the current in the plate circuit will also be varied, the same characteristics being maintained. As the field winding FW is in the plate circuit, it is evident that the antenna current will be modulated in accordance with the variations of the speech or sounds introduced at M. (See *Grid Control,* also *Modulation* and *Broadcasting, General Treatise on Methods.*)

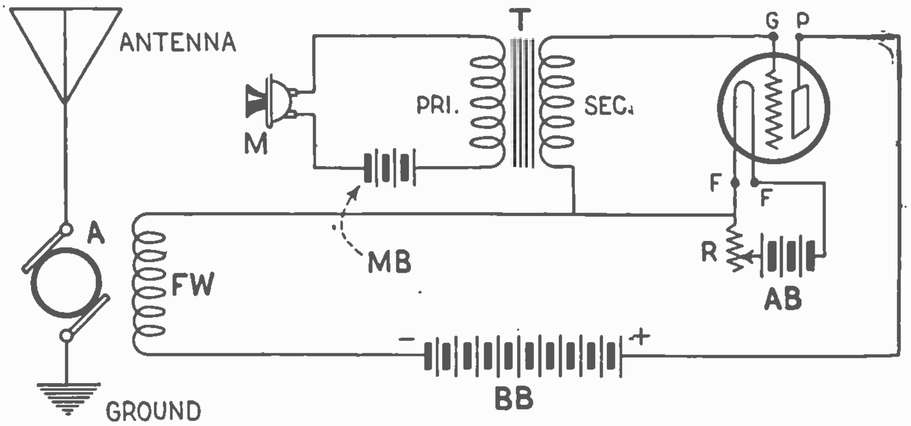

Circuit illustrating the fundamental principle of the Heising system of modulation.

HELIOGRAPH—An instrument involving a movable mirror, by means of which the sun's rays are reflected and made use of as a signalling device. In practice it is usually arranged with a blind or shade to cover the mirror at rapid intervals by pressing on a key device. In this manner the flashes may be made at definite intervals, as in the case of the Morse light. (See *Code,* also *Morse Light.*)

HELIX—The shape assumed by a line or wire wound in a single layer around a cylinder. Generally speaking, a *solenoid* (q.v.). A device in the form of a helix was formally used in transmitting circuits to vary the inductance, or, practically speaking, the wave-length. It usually was a number of turns of heavy copper or aluminum wire wound on an insulating form. Clips were provided to permit a variation in the contact point and hence any necessary change in the value of the inductance. (See *Oscillation Transformer.*)

HENRY, JOSEPH—Born in 1797, died in 1878. An American Physicist, noted for his researches in electromagnetism. He developed the electromagnet which had been invented by Sturgeon in England, so that it became an instrument of far greater power than before. In 1831 he employed a mile of fine copper wire with an electromagnet, causing the current to attract the armature and strike a bell, thereby establishing the principle employed in modern telegraph practice. He was made a professor at Princeton in 1832 and during his experimenting there he devised an arrangement of batteries and electromagnets embodying the principle of the telegraph relay which made possible long distance transmission. He was the first to observe magnetic self-induction, and performed important investigations in oscillating electric discharges (1842), and other electrical phenomena. In 1846 he was chosen secretary of the Smithsonian Inst. at Washington, an office which he held until his death. As chairman of the U. S. Lighthouse Board he made important tests in marine signals and lights.

Joseph Henry.

In meteorology, terrestrial magnetism and acoustics he carried on important researches. Henry enjoyed an international reputation and is acknowledged to be one of America's greatest scientists.

HENRY—The unit of inductance, named after the American physicist, Joseph Henry, who first produced high-frequency oscillations in 1832. A coil of wire is said to have an inductance value of one henry when a current passing through it changes at the rate of one ampere per second and an electromotive force of one volt is produced. As the henry is too large a unit for practical purposes, the thousandth part of a henry, known as the *milli-henry*, or the millionth part, known as the *micro-henry*, are more commonly used. Another much-used term is the centimeter, which is the thousandth part of a micro-henry. The abbreviation for henry is H, and for micro-henry μH. (See *Centimeter Gram Second*, *Centimeter*, also *Inductance.*)

HERTZ, HEINRICH RUDOLF—Born 1857, died 1894. A German physicist noted as the discoverer of electromagnetic waves (1888) which form the basis of wireless telegraphy. He found that waves produced by the spark of a simple device called the oscillator could be detected by a loop or square of wire known as the resonator, and he was able to show the reflection, refraction, diffraction and poralization of the waves. These remarkable discoveries demonstrated the practical possibilities of radio-telegraphy. He also made valuable experiments with electric phenomena in vacuum tubes.

Heinrich Rudolf Hertz.

HERTZIAN WAVES—Term occasionally applied to radio or electromagnetic waves, they being named after Dr. Heinrich Rudolf Hertz owing to his extensive researches into electromagnetic waves. (See *Hertz*, also *Clerk-Maxwell*, and *Ether Theory.*)

HETERODYNE—The term originally applied by Fessenden to a method of reception of electromagnetic waves, wherein the principle of *Beats* (q.v.) was employed. It is derived from the words "hetero," meaning other or different, and "dyne," meaning power.

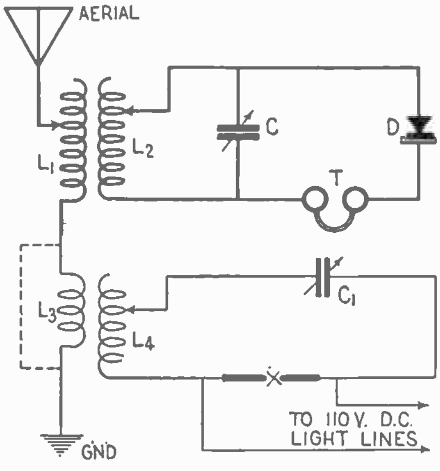

Fig. 1.—The original form of heterodyne receiver.

In its original conception by Fessenden, before the widespread use of vacuum tubes, it was used in conjunction with crystal receivers for the reception of *undamped waves* (q.v.). The heterodyne principle involves production of a local series of oscillations which are tuned to a frequency differing from the frequency of the incoming oscillations from the antenna. The difference between the two is known as the *beat frequency*. (See *Beats*, also *Heterodyne Receiver.*)

HETERODYNE RECEIVER—A receiver for electromagnetitc (radio) waves employing the *heterodyne* principle. The illustration Fig. 1 shows the original form of heterodyne receiver, wherein the local oscillations were produced by means of a direct current arc gap and associated circuits, and combined through a coupling medium, with the oscillations in the antenna circuit. Here the heavy lines represent a conventional crystal receiver circuit. L1 and L2 are respectively primary and secondary of a vario-coupler, C is a variable tuning condenser, D the crystal detector and T the head phones. If the dotted line is assumed as a connection to the ground, this will constitute a complete receiver. For the heterodyne purpose, however, a special coil L3 is employed in the ground circuit. Coupled to this is the secondary L4. C1 is another tuning condenser and A is an arc gap fed by 110 volts direct current. In operation the circuit represented by the light lines acts as a generator of undamped oscillations, these oscillations being supplied to the antenna circuit and the frequency varied by means of the condenser C1 and the coil L4. If the receiving tuner is tuned to a certain wave-length corresponding to incoming signals, and the generator system tuned to a different frequency, the difference between the two frequencies will be a *beat* frequency which will be present in the detector circuit. This new series of oscillations will be rectified by the detector D and will be audible in the head phones, T. The action of this system is shown by the series of curves given in Fig. 2. Here the incoming oscillations (signals) from the antenna are shown in graph 1 at the top of the illustration. These oscillations are of continuous amplitude or *undamped* (q.c.). Graph 2 shows the local oscillations. If we assume that the distance from A to B represents the time, it will be apparent that there are more oscillations per unit of time in the case of the incoming signals—

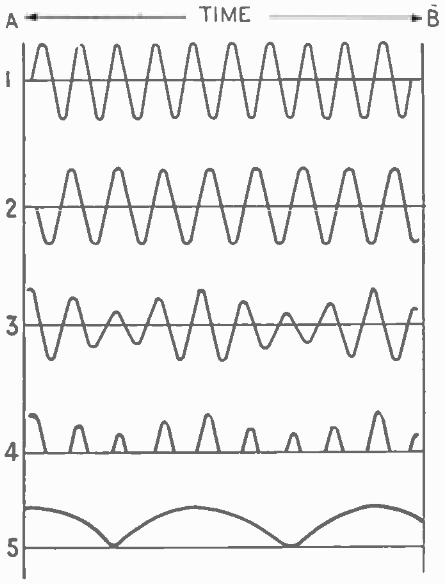

Series of graphs showing the action of a heterodyne receiver.

hence the second graph indicates that the local oscillations have a lower frequency. Graph 3 represents the *beat* resulting from interaction of the two upper series of oscillations and the frequency corresponds to the numerical difference. When the two upper groups of oscillations are opposed, the beat current is zero, and when they are in phase, or assisting each other, the beat current is maximum. When rectified by the crystal detector action the beats take the form shown in graph 4, the lower half of each cycle having been submerged. The telephone current of a periodic nature is indicated by the bottom graph, the original signals being represented in the phones by a series of contractions of the diaphragm. (See *Heterodyne*, also *Beats* and *Super Heterodyne.*)

HETERODYNE, SELF—See *Self-Heterodyne.*

HETERODYNE WAVE-METER—A wave-meter—device for measuring the wave-length of coils, circuits or incoming and outgoing signals—which utilized the heterodyne principle. Briefly the action is as follows: It has already been explained under Heterodyne Receiver that when the local oscillations and the incoming oscillations are opposed, the beat current is zero. It follows, then, that with a pair of phones connected in the circuit, the beat signals will be audible only when the two frequencies are different. Now if a suitable means is present for varying the frequency of the local oscillations, in practice a combination of coil and condenser, the local circuit can be tuned to a point where the two frequencies concur or are similar and no beat is audible. As the frequency is proportional to the capacity and inductance in a circuit, and as wave-length can easily be figured when the frequency of the local generating system is known, the wave-length of the incoming signals must be that of the local circuit when the beat is zero. The condenser dial on the heterodyne wave-meter can thus be calibrated directly in meters and a nearly true reading of the wave-length of incoming signals obtained. (See *Beats*, also *Heterodyne* and *Wave-meter.*)

HIGH FREQUENCY—A very general term applied to alternating electric currents. As a rule, alternating currents having frequencies above one thousand *cycles* (q.v.) per second are referred to as high frequency currents. When the frequency is extremely high, of the order of hundreds of thousands of cycles, the current is usually referred to as high frequency, oscillatory current. There is no definite level or line of distinction between high and low frequency currents. In radio practice, the audio frequency currents are generally considered below twenty thousand cycles per second and known as low frequency currents. The radio frequency or frequencies above twenty thousand are considered as the high frequency currents, although here again there is no definite level or line of demarcation between the two areas. (See *Audio Frequency*, also *Radio Frequency* and *Frequency.*)

HIGH FREQUENCY ALTERNATING CURRENT—See *High Frequency Current.*

HIGH FREQUENCY ALTERNATOR—An alternating current machine or alternator, designed to produce currents having high frequencies, generally above one thousand cycles per second. Such machines are used in high power radio telegraph transmission (not broadcasting) producing undamped or continuous waves. (See *Alternator*, also *Alexanderson Alternator.*)

HIGH FREQUENCY AMPLIFICATION—The process of stepping up or amplifying currents of high frequency. In radio circuits, the high frequency currents are known as radio frequency oscillations. In radio circuits the audio frequency (low frequency) currents and the high or radio frequency currents are usually in different parts of a circuit. When we consider alternating currents having a frequency above, say, twenty thousand cycles per second as radio frequency, and those below that limit as audio frequency, it is easy to consider the radio frequency currents as existing in all stages between the aerial and the detector. Beyond the detector, the alternating currents having become uni-directional surges, we can consider these currents as being audio frequency currents.

Thus we have a definite dividing line in any radio receiver using but one detector, the detector or rectifier marking the transition from radio frequency to audio frequency. Thus, any means of amplifying the signals before they reach the detector will be known as high frequency amplification. (See *Audio Frequency*, also *Amplifier.*)

HIGH FREQUENCY BUZZER—A *buzzer* (q.v.) designed to produce a very high note. The term "high frequency" in this case merely indicates that the pitch or frequency (vibration) is high in comparison with the ordinary buzzer, the sounds of course being actually of audible frequency. The general scheme of such a buzzer is about that of the ordinary type as used in conjunction with electric bells or signal systems, the armature, however, being very thin and thus permitting rapid make and break for the current. Such buzzers are used extensively in connection with crystal receivers for the purpose of testing and also for code practice. (See *Buzzer* also *High Frequency* and *Armature.*)

HIGH FREQUENCY COMPONENT OF PLATE—The high frequency currents present in the plate circuit of a vacuum tube receiver, as distinguished from the audio or low frequency currents. (See *Feed-Back*, also *Plate Component, High Frequency*, and *Regeneration.*)

H.F.C.—Abbreviation used abroad and occasionally in the United States for high frequency current. The more common term is radio frequency current and no abbreviation is used as a rule. (See *Radio Frequency Current*, also *High Frequency.*)

HIGH FREQUENCY CURRENT—Alternating currents having frequencies above certain limits, as, for example, the radio frequency currents employed in radio transmission, which are usually of a frequency above one hundred thousand cycles per second. (See *High Frequency*, also *Low Frequency* and *Radio Frequency.*)

HIGH FREQUENCY OSCILLATION—The term oscillation is generally applied to an alternating current of very high frequency. Thus radio frequency currents, which may run into hundreds of thousands of cycles per second, are called high frequency oscillations. (See *Oscillation.*)

HIGH FREQUENCY RESISTANCE—Resistance to the passage of alternating currents of high frequency. Said of wires, coils, circuits and apparati. The resistance of a certain conductor or part of an electrical circuit, as in radio, may be comparatively higher in the case of high frequency currents than to alternating currents of a low frequency. This is due to the "*skin effect*" (q.v.) it being a phenomenon of radio frequency (high frequency) currents that they have a tendency to concentrate near the surface of the conductor. For this reason, small wires are more apt to have high radio frequency resistance, as obviously the surface on which the currents can travel is restricted. (See *High Frequency.* also *Skin Effect* and *Resistance.*)

HIGH PASS FILTER—An arrangement of electric circuits designed to permit ready passage of high frequency alternating currents while at the same time presenting high resistance or impedance to alternating currents of low frequency, generally within certain definite limits. (See *Filter* also *Filter, High Pass.*)

HIGH PRESSURE—A term occasionally used to denote high potentials or voltages. (See *High Tension*, also *High Potential.*)

HIGH POTENTIAL—One of the several terms used to indicate high voltage. (See *High Potential Battery*, also *High Voltage.*)

HIGH POTENTIAL BATTERY—Another term for "B" battery, the battery used to supply potential to the plate of a vacuum tube. The term "high potential" is used here, only in a comparative sense, the voltage of the battery being high in proportion to that of the filament or "A" battery. High potential batteries for radio usage are generally furnished in blocks of 22½ or 45 volts. (See *High Voltage, High Tension* and *"B" Battery.*)

HIGH RESISTANCE—The property of any circuit whereby it offers considerable opposition to the flow of electric currents. There is no very distinct line of demarcation between high and low resistance. One of the common applications of the term is in the case of a *grid leak* or grid resistance. Here the resistance may be upward of a half million ohms, running into several million ohms in certain cases. An example of low resistance elements is that of a *rheostat* for controlling the filament current. Here the resistance is relatively low, rarely more than thirty ohms. The average resistance of the magnet windings in headphones (as used in radio) is one thousand ohms. The term "high resistance" is also used in a comparative sense in respect to various instruments or parts of radio circuits. Here the resistance in most cases should be as low as possible, and therefore a resistance of even a few ohms may be referred to as *high*. While resistance to direct currents is generally the feature of resistance coils and units, the opposition to currents of an oscillatory nature (high frequency alternating currents) may be of the utmost importance, and it is often the case that a circuit has little direct current resistance but very high resistance or *impedance* (q.v.) to high frequency currents. (See *Resistance, Impedance*, also *Skin Effect.*)

HIGH RESISTANCE JOINT—Any connection between two wires or conductors which offers resistance to the passage of current due to defective contact or imperfect connection through solder or any other medium. When bare wires are joined they should be carefully scraped and soldered. Too much solder is a bad feature, as the resistance of solder is naturally far higher than that of the copper wire. (See *Corrosion*, also *Soldered Joints.*)

HIGH RESISTANCE TELEPHONES—Headphones or telephones used in radio reception, having high resistance windings. In the ordinary telephone receiver used with the commercial telephone systems, the windings of the telephone magnets have a relatively low resistance, generally less than one hundred ohms. Such phones would not be sufficiently sensitive for radio usage, the average resistance in this case being about one thousand ohms. The resistance here refers to the ohmic resistance of the many turns of wire wound around the magnets. This resistance is not a required feature, the factor of importance being the number of turns. Inasmuch as space is a factor to be considered, it is necessary to use very fine wire in order to permit use of the thousands of turns essential. The resistance of this fine wire is of course high. Where very high resistance windings are used—and this indicates small wire—it is sometimes advisable to use a by-pass for the heavy plate potentials to protect the windings. (See *Ampere Turns*, also *Output Transformer* and *High Resistance.*)

HIGH SPEED RECEPTION AND TRANSMISSION—The reception or transmission of radio messages by automatic means in order to obtain much greater speed than would be possible under the manual system or hand opereating and recording. The general practice is to record the messages to be transmitted on a tape by means of a punching machine. They are then sent automatically through a high speed automatic transmitter and received automatically at the same high speed by a machine which punches the messages or records them on some form of tape or sheet. The more common of these methods is the *Wheatstone Transmitter.* (q.v.).

HIGH TENSION—One of the numerous terms used to denote high voltage. "B" batteries as used in radio reception are generally known as high tension batteries to distinguish them from the low voltage storage battery or dry cells used to light the filaments. Abbreviation H. T. mostly used in magazines and books printed in England.

HIGH TENSION BATTERY—Term applied to "B" batteries used in radio reception to supply potential to the plates of the vacuum tubes. These batteries usually are in blocks of twenty-two and a half or forty-five volts. (See *"B" Battery.*)

HIGH TENSION CIRCUIT—Term occasionally applied to the part of a radio circuit in which high potentials are present, as in the plate circuit of a vacuum tube. Here the "B" battery is in the circuit. (See *High Tension*, also *"B" Battery.*)

HIGH TENSION INSULATOR—Insulators of large dimension and rugged design, possessing unusual insulating properties in order to withstand high voltages. As a rule such insulators are made of high grade porcelain and arranged with annular ridges or "petticoats" to give greater surface and reduce the possibility of surface leakage. The illustration shows a form of high tension or high potential insulator. (See *Insulator*, *High Tension*, also *Petticoat Insulators.*)

A high potential insulator.

HIGH VACUUM—A term applied to vacuum tubes when they have been highly exhausted of air. That is, an almost perfect vacuum. When a vacuum tube has a very good vacuum—has been highly evacuated of gases—it is generally called a "hard" tube. A high vacuum is essential in tubes to be used as amplifiers or for transmission. (See *Vacuum Tubes, Theory of Operation.*)

HIGH VOLTAGE—A very indefinite term, used in ordinary electrical practice to denote voltages above 600, but applied in radio practice to much lower potentials, as, for example, a "B" battery, which is generally referred to as a high voltage battery in spite of being generally below 100 volts. The term used in this sense is one of comparison with the low voltage of a storage or other type of "A" battery. (See *Voltagee*, also *Potential.*)

HIGHLY DAMPED WAVES—An expression applied to waves or trains of waves wherein the amplitude or strength of each succeeding oscillation or train of oscillations is greatly decreased. If the progressive decrease in amplitude or strength is very gradual, that is, if the waves die out slowly, they are said to be feebly damped, but if they die out rapidly they are referred to as being highly damped. (See *Damped Waves*, also *Decrement* and *Amplitude.*)

HOGAN, JOHN V. L.—American radio expert. Born at Philadelphia, Pa. He was educated at Sheffield Scientific School, Yale University, where he made a special study of physics and mathematics. In 1906 he became assistant to Dr. Lee De Forest, and in 1909 he joined the National Electric Signalling Co. In 1914 he was appointed chief research engineer to the International Radiotelegraphic Company. Hogan has written many articles on radio and is a past president of the Institute of Radio Engineers, member of the American Institute of Electrical Engineers, of the American Association for the Advancement of Science, of the Radio Club of America and other societies.

John V. L. Hogan.

HOMOPOLAR—Synonymous with unipolar. Having but one pole. Said of a dynamo having its conductor or armature rotating around a single pole of a field magnet. (See *Multi-Polar.*)

HONEYCOMB COIL—A coil used in radio, having a peculiar construction, somewhat similar to a honeycomb. The illustration shows a typical honeycomb coil used as an inductance in receiving circuits. The object of this staggered winding system is to reduce the self-capacity between wires to a minimum without reducing the inductive value. The chief value of such a winding method is that it permits the retaining of high inductance value in a limited space, without undue *high frequency resistancee* (q.v.). (See *Inductance, Distributed Capacity* and *High Frequency Resistance.*)

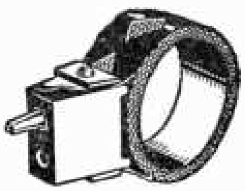

Honeycomb coil used in radio reception.

HOOK UP—The general term in the United States to designate the diagram for any complete radio receiver or transmitter, or for any specific part of such a circuit. Thus, when we refer to a hook-up for Radio receiver, we mean a diagram, either schematic (using symbols for the parts) or perspective (showing pictures of the parts) which gives the various parts necessary and the manner of connecting them to each other to form the complete circuit. Similarly we may use the term "hook-up" to describe a sketch or circuit for connecting various instruments such as *ammeters*, or *voltmeters*, *wavemeters* and the like. (See *Circuit*, also *Diagram.*)

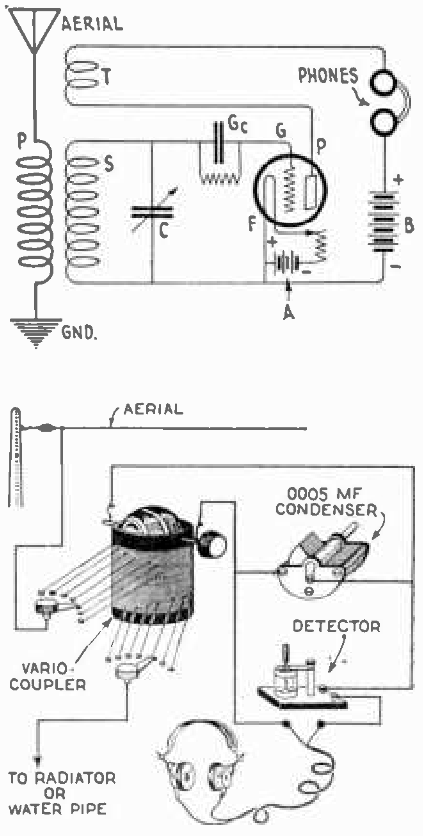

The upper diagram is in schematic form and shows a single tube regenerative receiver, and the lower diagram is in perspective form and shows a crystal receiving set.

HOOPER, STANFORD C.—American Naval wireless expert. Born August 16, 1884, at Colton, California, and educated at San Bernardino, California. He joined the Southern Pacific Company as a telegraph operator. In 1901 he entered the Naval Academy at Annapolis, and entering the Navy was advanced to Commander in 1918. From 1910-11 he was instructor in electrical engineering, physics and chemistry at the U. S. Naval Academy, and in 1912-13 Fleet Radio Officer of the United States Atlantic Fleet. During the great war he was responsible for the supply of wireless instruments, etc., for the American Navy, and he was also responsible for the construction of many of the larger American wireless stations including those at Annapolis, San Diego and Pearl Harbor. Hooper was one of the chief men concerned with the radio-compass system now used in the American Navy.

HORIZONTAL AERIALS—Aerials—overhead wire systems—for the reception or transmission of radio signals, wherein the major portion of the aerial is horizontal—parallel to the earth's surface. Typical examples are the "T" and inverted "L" aerials. In the former type, the flat portion suspended horizontally between two supporting structures and a lead-in wire is attached to the center and thence to the apparatus. In the latter type, the lead-in is attached to either end, giving a shape somewhat similar to an inverted L, from which the name is derived. The more common term is "flat top." This term is now somewhat ambiguous, as single wire aerials are used more or less extensively for receiving purposes, whereas formerly a system of several horizontal wires was employed. (See *Flat Top Aerial or Antenna*, also *Aerial T Type* and *Inverted Γ Aerial.*)

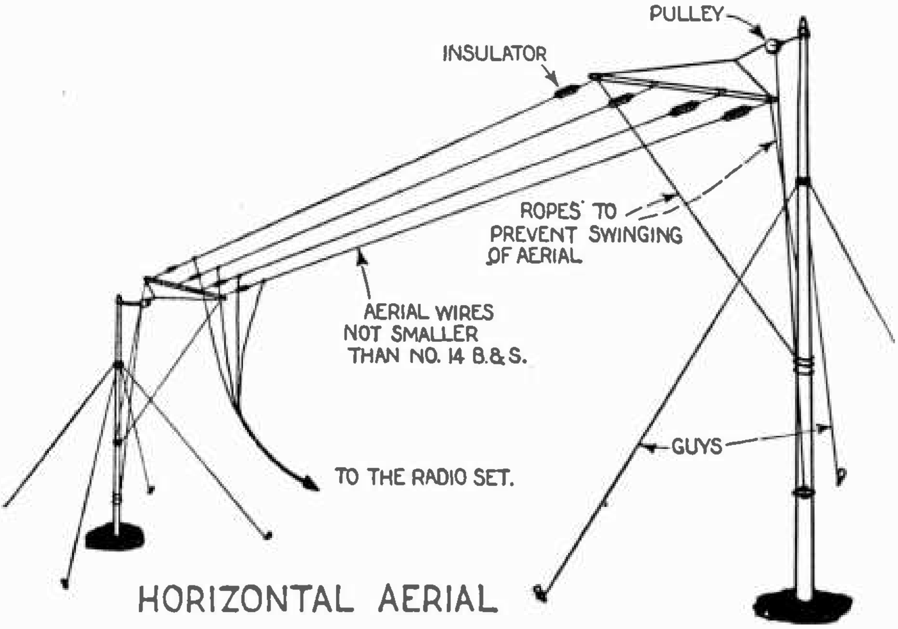
PULLEY
INSULATOR
ROPES TO PREVENT SWINGING OF AERIAL
AERIAL WIRES NOT SMALLER THAN NO. 14 B.&S.
GUYS
TO THE RADIO SET.
HORIZONTAL AERIAL

A horizontal aerial of the inverted "L" type.

HORN, LOUDSPEAKER—A device, generally in the shape of a horn, used in conjunction with a heavy duty telephone receiver or loudspeaker unit, to reproduce loudly the signals or music from a radio receiver. In operation, the horn has attached to it at its small end a unit or telephone receiver. When the diaphragm of the unit is actuated by the signals, speech or music, the column of air in the open tube is displaced, creating audible sounds which can be heard at a distance or throughout a room. Assuming other conditions equal, the volume will be more or less in proportion to the size of the norn, or more specifically to the amount of air displaced by the signals. When a telephone receiver is used without horn, the opening above the diaphragm being unconfined, only a small column of air is displaced and the volume will not be great under ordinary circumstances. When a horn is attached, however, the sound being confined will displace a relatively greater amount of air—depending mainly on the shape and size of the horn—and the volume will be measurably greater. (See *Loudspeaker*, also *Acoustics* and *Loudspeaker Units.*)

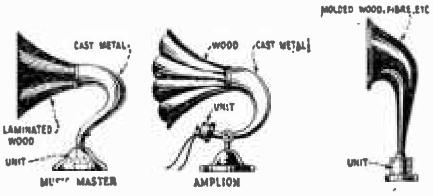

Three types of horn loudspeaker.

HORSE POWER—The unit of power used in the United States and Great Britain. It is defined as the power required to lift 550 pounds to a height of one foot in one second, or similarly, 33,000 pounds to that height in a minute. The term "foot-pounds" is derived from this value. Thus, one horsepower is the power required to perform 550 foot-pounds of work per second of time. (See *Horse Power, Electrical.*)

HORSE POWER, ELECTRICAL—The unit of electrical work similar to the mechanical horse power. It is actually mechanical horse power expressed in *watts* (q.v.). One electrical horse power is equivalent to 746 watts. In the *centimeter gram second* (q.v.) system of units, the unit of work is the *erg.* This unit is too small for ordinary usage and the customary unit is the watt, which is equal to 10^7 ergs per second. The unit used in electrical power work is still larger; it is the *kilowatt* or one thousand watts. (See *Units*, also *Watts.*)

HORSESHOE MAGNET—A *magnet* (q.v.) in the shape of a horseshoe or letter U, both poles being brought nearer together so that they may act on the armature or "keeper." Generally a magnetized steel bar which retains its magnetism for a long period. The illustration, Fig. 1, shows a permanent horseshoe magnet. Such magnets are also used with external power, then being termed *electromagnets.* In this case the magnet is made of soft iron. The illustration, Fig. 2, shows a horseshoe *electromagnet.* Here the magnetic attraction is supplied by application of an external electric force to the windings WW, and the magnet being of soft iron, loses its magnetic attraction almost immediately upon withdrawal of the electric current, magnetizing force. (See *Telephone Receiver*, also *Electromagnet.*)

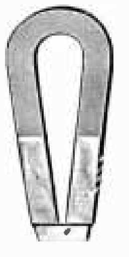

Fig. 1. A permanent horseshoe magnet.

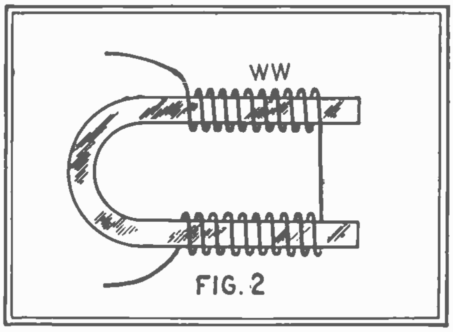

Fig. 2. A horseshoe electromagnet.

HOT-WIRE AMMETER—An ammeter, or instrument for measuring electric current amperes, which depends for its action on the expansion of a fine wire under the influences of the heat produced in it, by the passage of the electric current to be measured. In the illustration the current to be measured is passed through the fine resistance wire A-B. P is a pointer moving over a graduated scale, and attached at its other end to a coiled spring S, which in turn is attached to the wire A-B by the silk thread T. The tendency of the spring is to move the pointer to the right, but owing to the pull of the thread in the normal position of the wire A-B, the pointer rests at zero at the left of the scale. As current is passed through the wire A-B, it heats and expands, thus permitting more or less slack to occur in the thread T and allowing the spring S to move the pointer over the scale toward the right, As soon as the current is withdrawn, the wire cools and contracts to its former size, thus again restoring the tension to the thread and returning the pointer to zero. Hot Wire ammeters are used extensively for measurement of currents at *radio frequency*, that is, extremely high frequency as in radio transmission. The average ammeter using a coil of fine wire would not be adaptable to this use, as the high voltage, high frequency current would burn out the coil. In addition to this, the length of these windings would affect the circuit and probably throw it out of *resonance* (q.v.). (See *Ammeter* or *Ampere Meter, Frequency, Resonance,* also *Instrument Shunts.*)

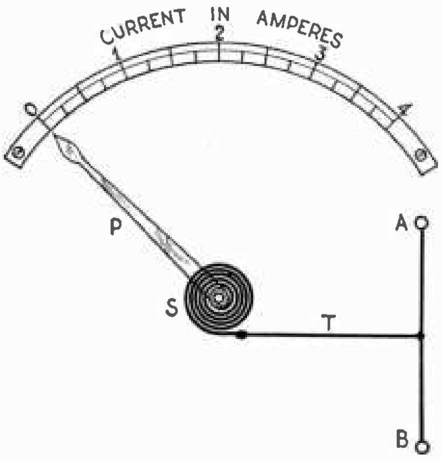

Illustration showing operation of a hot wire ammeter.

HOT WIRE INSTRUMENTS—Current measuring instruments which utilize the tendency of a fine wire to sag or expand under the heat caused by passage of an electric current through it. Such instruments may be used for direct or alternating current. (See *Hot Wire Ammeter.*)

HOT WIRE OSCILLOGRAPH — See *Oscillograph.*

HOT WIRE TELEPHONE—See *Thermal Telephone.*

HOT WIRE WATTMETER—An instrument used to measure combined volts and amperes, i. e., watts in a circuit. It comprises essentially a mirror, which is deflected according to the difference in expansion of two fine wires. (See *Hot Wire Instruments,* also *Wattmeter.*)

HOUR METER, AMPERE—An instrument used extensively in connection with storage batteries, to measure the input or output in amperes per hour. (*Ammeter* or *Ampere Meter.*)

HOUSE MAINS—A term much used in connection with storage battery charger and battery elimination devices. The term implies the main circuit of any standard house lighting system, either direct or alternating current. Thus in the case of a rectifier for charging storage batteries from alternating current, a lead is provided for connecting to the house mains, offering a means for introducing the required current into the device. (See *Current Direct,* also *Alternating Current* and *Chargers, Storage Battery.*)

HOWLING—The general term used to imply any undesired sounds produced in or by a radio receiver. Howling is usually caused by *feed-back effects* (q.v.). The most common example of howling is in the case of a regenerative receiver. Here the detector tube may have a tendency to generate its own oscillations As long as these oscillations are equal in frequency to any incoming signals there will be no howling, but should these local oscillations differ in frequency from the incoming series of signals, the result will be a howl resulting from the *beats* (q.v.) established owing to the difference between the two sets of oscillations. This is essentially the principle of *heterodyne* or "beat" reception, but in the case of a heterodyne, the effect is utilized, whereas in the case of a regenerative receiver it is undesirable. Interaction between various parts or different parts of a circuit may often lead to howling. (See *Feed-Back,* also *Regeneration, Reflex, Beats* and *Re-Radiation.*)

H. P.—Abbreviation sometimes used in radio for *high potential* or *high pressure.* In mechanics it is the standard abbreviation for *horse power.*

HOZIER-BROWN DETECTOR—A form of electrolytic detector, which, however, does not use a liquid electrolyte. It consists of a small pellet of lead oxide held more or less rigidly between two electrodes, one platinum and the other lead. (See *Detector,* also *Electrolytic Detector.*)

HYDROGEN—A gaseous chemical element. The symbol is H and its specific gravity is taken as 1, that of air being 14.4. It is colorless, tasteless and odorless and is the lightest known body. Hydrogen is a constituent of water, which is composed of one atom of oxygen and two atoms of hydrogen according to the chemical formula for water, H_2O. Hydrogen is used in radio transmission in connection with *arc generators.* (q.v.) (See *Poulsen Arc.*)

HYDROMETER—An instrument used to measure the *specific gravity* (q.v.) of a liquid. In radio it has an extensive use in connection with storage batteries. Various types of hydrometers are used to indicate the specific gravity of the electrolyte of a storage battery and hence in a manner, to indicate the state of charge of the battery. The illustration shows the standard form of hydrometer using a weighted float. In the illustration the float is shown inside the barrel of the instrument. In operation the bulb B is compressed and the tube T inserted in the open vent of a storage battery. When pressure on the bulb is released some of the electrolyte is drawn up inside the barrel and the float rises more or less within the barrel in proportion to the specific gravity or density of the electrolyte. With the hydrometer held in the manner shown, that is, upright, the numbers just at the level of the liquid in the barrel indicate the specific gravity. Thus, if the reading is 1250° or higher the battery is considered in fairly good operating condition —full charge being 1280° to 1300°. If the reading is below 1200° it indicates the need of charging, these levels not being arbitrary, a discharged condition being indicated by a reading of 1150° or lower. The use of a hydrometer is generally advisable and the reading accurate under ordinary conditions. If, however, water has just been placed in the battery, the reading may not be accurate for several hours due to the fact that the water and acid may not mix immediately. (See *Storage Battery Tests,* also *Electrolyte.*)

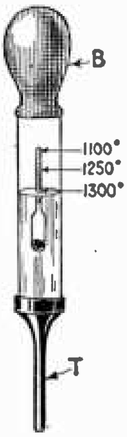

Hydrometer used to test the specific gravity of liquids.

HYGROSCOPIC EFFECT — The effect in any substance which permit it to absorb moisture. In the case of forms or tubes for radio coils and inductances this feature is particularly undesirable, as obviously, moisture lowers the dielectric value of the form and may cause it to warp. (See *Bakelite,* also *Tubing.*)

HYSTERESIS—The tendency of magnetization to lag behind the magnetizing force as, for instance, in an iron core in a *transformer* (q.v.). When the core of such a transformer is undergoing rapid reversals of magnetism, there is often an expenditure of energy which is not useful and which is converted into heat, thus representing a loss in power. This effect is more noticeable with certain qualities of iron. In other words, when rapid reversals of magnetism take place in a poor quality of iron core it may display a certain sluggishness. This is actually the lagging of the magnetic flux behind the magnetizing force producing it. This loss of energy is assumed due to the work required to alter the position of the molecules of the iron composing the core and the less energy expended in this manner the less the hysteretic losses. In designing transformers or armature cores, hysteresis must be taken into consideration and the iron must be of high quality to hold these losses at a minimum. (See *Eddy Currents,* also *Core, Transformer.*)

HYSTERESIS COEFFICIENT — The amount of energy wasted or lost in the process of magnetizing and de-magnetizing a unit volume of magnetic material is referred to as the hysteresis coefficient of that material. (See *Hysteresis,* also *Coefficient.*)

HYSTERESIS LOOP—Graphic illustration of losses due to hysteresis in any sample of core material. Loops are plotted in graphic curves of cycles of magnetization. (See *Curve,* also *Hysteresis.*)

HYSTERESIS LOSSES — The energy wasted or lost through hysteresis. (See *Hysteresis.*)

HYSTERETIC LAG — The lagging of magnetization due to the effect of hysteresis. (See *Hysteresis,* also *Magnetization.*)

I

I—The abbreviation adopted by the *International Electrotechnical Commission* (q.v.) and American Institute of Electrical Engineers, to denote current in amperes. The small or lower case (i) may also be used where the capital letter is not advisable. (See *Current.*)

I ARMATURE—An armature used in generating machines, having a core resembling the letter I in shape. (See *Armature,* also *Core.*)

I.C.W.—The customary abbreviation for "interrupted continuous waves." This is a system of radio transmission wherein the waves are modulated at a constant low frequency. (See *Interrupted Continuous Waves,* also *Modulation.*)

IDLE CURRENT—A name often applied to the component of an alternating current which contributes nothing to the power. This is also known as the *wattless component* (q.v.) or the *wattless current* (q.v.). It represents the component which, being in quadrature with the electromotive force in the circuit, thus has no active value—adds nothing to the total power. (See *Current,* also *Electromotive Force* and *Alternating Current.*)

I.E.C.—The abbreviation for *International Electrotechnical Commission,* the body which suggested the international electrical and magnetic abbreviations and symbols. (See *Symbols,* also *Units.*)

IMPACT OR SHOCK EXCITATION—The method of producing *free oscillations* (q.v.) in a circuit by means of an exciting current of short duration compared to that of the resultant oscillations. The term is a general one and broadly covers the various methods of generating oscillations in which an oscillatory circuit is thrown into electrical vibration at its natural or fundamental frequency by means of an electromotive shock or exciting force of short duration. Also covers the term *Impulse Excitation.* (See *Oscillations,* also *Fundamental Frequency.*)

IMPACT TRANSMITTER—Apparatus for transmitting radio waves wherein the transfer of energy from the primary or exciting circuit to the oscillatory circuit takes place during one pulse or vibration of the exciting current. (See *Impact or Shock Excitation,* also *Transmitter.*)

IMPEDANCE—The total resistance to the flow of alternating current in a circuit. Impedance combines the ohmic resistance and the apparent resistance due to self-induction. When a direct electromotive force is impressed on a circuit, the current flowing in that circuit depends directly on the ohmic resistance of the circuit. The formula is according to Ohm's law $I = \frac{E}{R}$, I representing the current in amperes, R the resistance in ohms and E the electromotive force in volts. Now if this direct electromotive force is replaced by an alternating EMF (electromotive force) and the ohmic resistance remains the same, the total resistance of the circuit to the alternating EMF will nevertheless be greater than in the case of direct current. Now the ohmic resistance is still the same but we have added to it the apparent resistance due to the self-induction of the circuit. In other words the inductance of the circuit has a marked effect on the current flowing in the circuit. What actually occurs is that the self-induction sets up a *counter-electromotive force* (q.v.) which continues to act against the impressed EMF as long as the flow of current persists. The result is a reduction in the current flowing in the circuit, the EMF having been, in effect, reduced, and hence according to the law above, $I = \frac{E}{R}$, the current is less. This added resistance, due to self-induction, is referred to as *reactance* and the flow of current in an alternating current circuit is thus governed by the ohmic resistance and the reactance. The application of the term "apparent resistance" to describe the added resistance due to self-induction, is a concession to the fact that it is not actually resistance in the proper sense of the word. In a direct current circuit, the resistance actually wastes or dissipates a certain amount of energy. In the alternating current circuit, the ohmic resistance has the same effect. The reactance, however, does not dissipate energy. It merely reduces the effective EMF by means of the counter EMF and thus necessitates the application of a higher impressed voltage in order to maintain the same flow of current.

This is one of the major advantages of alternating over direct current. Alternating current can be transformed readily and can also be readily controlled by means of a reactance or choke coil, which by creating a counter EMF limits the current flowing in the circuit without dissipating any great amount of energy. (See *Alternating Current, Current Direct, Choke Coil,* and *Reactance.*)

IMPEDANCE COIL—A coil of wire or inductance coil, generally wound on an iron core, used in alternating current circuits to limit the current flowing in the circuit. In action such a coil serves through the medium of its self-induction, to create a counter electromotive force which acts against the impressed EMF (electromotive force) and thus reduces the current in the circuit according to ohm's law that the current is equivalent to the EMF in volts, divided by the resistance in ohms. An impedance coil is also known as a choke coil although the two may have somewhat different meanings. For example, a choke coil might be used in either direct or alternating current circuits, whereas an impedance coil as its name indicates, offers impedance, an effect only present in alternating current circuits. The simplest formula for impedance of a coil is the following: Impedance $= \sqrt{R^2+P^2L^2}$. Here it is assumed that the current takes the form of a *sine wave* (q.v.), then R is the resistance of the coil, L the coefficient of self-induction in henries and p is 2 π times the number of alternations per second. (See *Choke Coil, Alternating Current, Current Direct, Reactance, Impedance,* also *Impedance Coupled Amplifier.*)

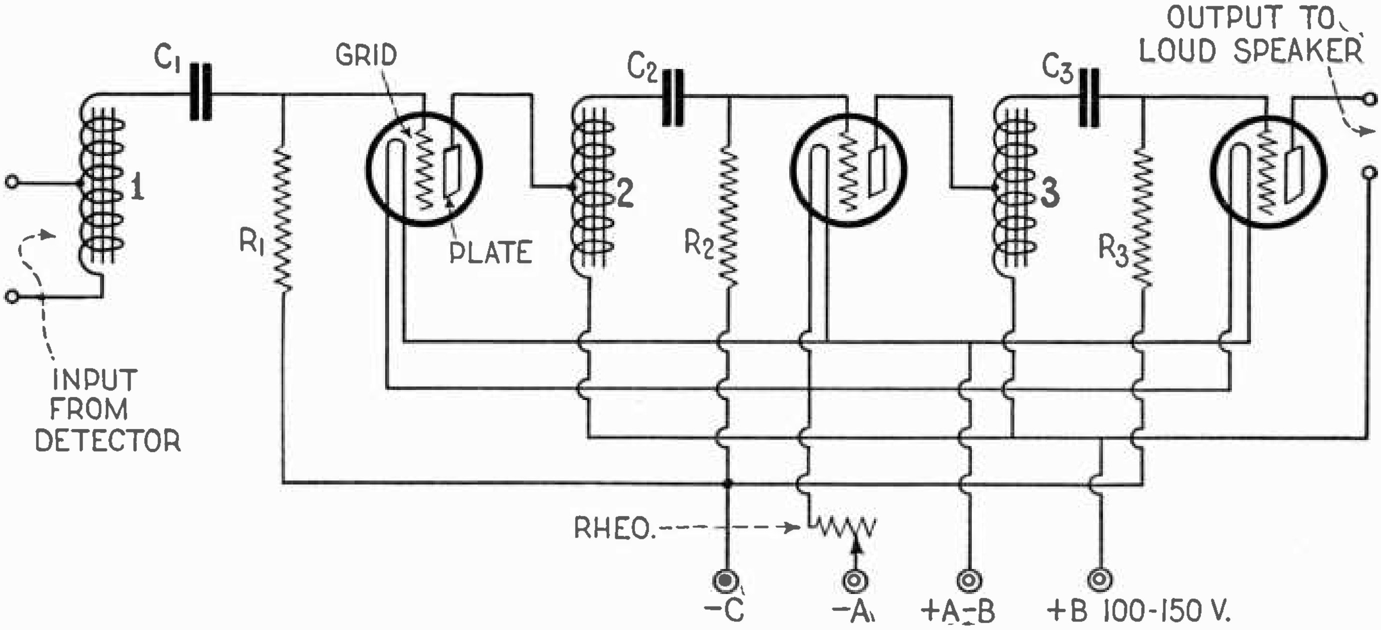

Schematic diagram of an impedance coupled audio frequency amplifier. (*See Impedance Coupled Amplifier.*)

IMPEDANCE COUPLED AMPLIFIER—A system of amplifying radio signals, voice or music, whereby the successive stages are coupled or joined by means of impedance coils. This system is much used in audio amplifiers, the amplification being much more uniform over the necessary range of frequencies than in the case of the ordinary audio frequency transformer. This is also known as choke coil coupling or choke coil amplification. The illustration on the preceding page shows a typical impedance coupled amplifier for audio frequency currents. It will be noted that three stages are used. This is necessary owing to the lower amplification per stage. Using this system, then, three stages are necessary in place of the customary two stages of transformer coupled amplification. In the illustration, which is a schematic diagram of the circuit for such an amplifier, the impedances 1, 2 and 3 are in the form of *auto transformers.* (q.v.) R1, R2 and R3 are grid leaks ranging in value between one quarter megohm and one and one-half megohms. C1, C2 and C3 are stopping condensers of about 1 micro-farad capacity each. The variations in potential are transferred successively from stage to stage, the amplification factor being roughly the sum of the individual factors of the tubes. In other words the ratio of the windings of the impedances is one to one, or unity, and therefore no voltage amplification takes place. The resistances are the usual grid leaks and the condensers C1, C2 and C3 are to block the high plate voltages from the grids of the succeeding tubes. (See *Tuned Impedance,* also *Amplifier* and *Vacuum Tube Amplifiers, Theory of.*)

IMPEDANCE FACTOR—The ratio of the impedance to the ohmic resistance in an alternating current circuit. (See *Impedance,* also *Resistance.*)

IMPEDANCES, MATCHED—Arrangement of an amplifying circuit and the choice of an audio transformer or loudspeaker with due regard for the impedance of the circuit. Thus if the impedance of a certain amplifier tube is 10,000 ohms, and the resistance of a loudspeaker unit being used is 2,500 ohms, then for maximum response when used with an *output transformer* (q.v.) the impedance of the primary of the transformer should be equivalent to the total or 12,500 ohms. (See *Matched Impedances.*)

IMPREGNATED INSULATING MATERIALS—Cloth or paper impregnated with an oily or resinous substance to render it moisture proof and improve the insulating properties. Cardboard is sometimes treated in this manner and used for tubing to wind inductance coils. (See *Empire Cloth or Paper,* also *Insulating Property.*)

IMPRESSED ELECTROMOTIVE FORCE—The term used to distinguish the EMF (electromotive force) impressed on or applied to a certain part of an electrical circuit from the counter or back-electromotive force due to constants of the circuit. The term is sometimes used where a direct current is applied in a circuit, but it should be confined to use where a counter EMF (electromotive force) is present. (See *Counter Electromotive Force,* also *Choke or Choking* and *Reactance.*)

IMPRESSED FIELD—An electric or magnetic field of force impressed on or applied to a body or region (area) for the purpose of producing other fields of force. (See *Field.*)

IMPULSE—Electrically, an electromotive force which produces an impulsive rush or discharge of electricity. The term is used to distinguish between this form and an electromotive force producing a steady flow of current. (See *Current,* also *Electromotive Force.*)

IMPULSE E.M.F.—The more specific term designating an EMF (electromotive force) which produces an impulsive rush or discharge of electricity. It is defiined technically, as having a maximum value which is large compared with its average value, the average value being taken over a time equivalent to the *time-constant* (q.v.) of the circuit on which the EMF is impressed. (See *Current, also Electromotive Force.*)

IMPULSE EXCITATION—Method of exciting or starting oscillations in a circuit, such as the aerial circuit of a transmitter, by a sudden surge of impulse EMF (electromotive force) rather than by application of oscillations of the frequency of the circuit.

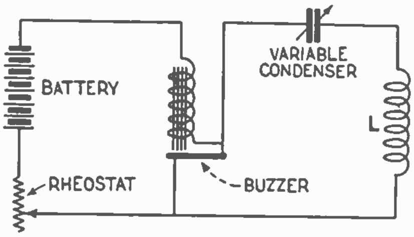

Connections of buzzer for generating damped oscillations suitable for testing by impulse excitation.

This system of impulse or shock excitation may be used for transmission and also has application in connection with testing or comparing of wavemeters. The illustration shows a buzzer connected with the necessary components to form a source of high frequency oscillations. This system is capable of producing through coupling from the coil L, currents by shock excitation. These currents can be used in a variety of ways for testing purposes. In operation the condenser is charged to the voltage of the battery at the instant the buzzer contact is open and when the contact closes, discharges through the coil L thus furnishing a series of oscilations to surge through the coil and by inductive coupling to any other coil placed in proximity to it. (See *Buzzer Exciter.*)

INCANDESCENCE—The glowing of any substance when brought to a white heat. The term is used in radio to signify the condition of the heated cathode or filament in a vacuum tube. In order to produce *electrons* (q.v.) it is necessary to cause the filament to become white hot, the condition being known as incandescence. (See *Filament.*)

INCONDUCTIVITY—A little used term signifying the condition of a substance by virtue of which it will not act as a conductor for electric currents. (See *Conductivity, Specific.*)

INDEPENDENT UNIT—A unit of a quantity, such as length, mass or time, chosen arbitrarily and without having any relation to any other unit. (*Centimeter Gram Second, or C. G. S.,* also *Fundamental Units.*)

INDICATING INSTRUMENTS—Any instrument or device which indicates values, as for example, a *voltmeter* or *ammeter.* (q.v.) (See *Instruments, Measuring.*)

INDIRECT CURRENT—Term for alternating current—now obsolete. (See *Alternating Current.*)

INDUCED CURRENT—The current in a circuit due to an induced EMF (electromotive force). The current produced by an EMF induced in a circuit due to a change in the magnetic flux linking the circuit with another. (See *Induced EMF,* also *Induction.*)

INDUCED E.M.F.—The electromotive force induced in a conductor or body. The phenomena of induced EMF (electromotive force) is the basis of operation of generating machines. Here a conductor is rotated or moved in a field of force, the action causing a change in the flux lines linking the circuits or the conductor. An EMF may also be induced by placing a coil in inductive relation to another coil through which alternating current is passing. Here the transfer is caused by induction. (See *Induction.*)

INDUCTANCE—The broad term used to designate any wire wound coil as used in radio practice for tuning units, inductive couplers and any variety of purpose. Actually the term is derived from self-induction or self-inductance, a property of a coil of wire by virtue of which it sets up a counter EMF (electromotive force) which serves to oppose the original EMF passing through it. When a certain circuit has inductance, it is to say that a number of turns of wire possessing self-induction are present in the circuit. The greater the inductance in a circuit the greater its opposition to the passage of an alternating current. The unit of inductance is the Henry. This is defined as the self-induction of a circuit or coil when the induced EMF is one volt with the inducing current varying at the rate of one ampere per second. (See *Alternating Current, Henry,* also *Inductance Coils.*)

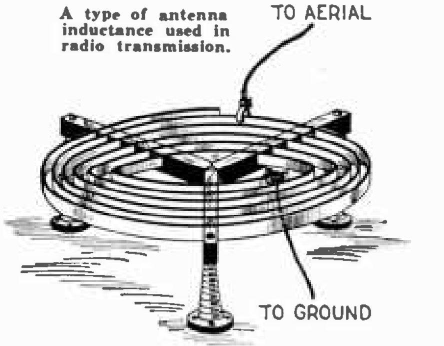
A type of antenna inductance used in radio transmission.

INDUCTANCE, ANTENNA—Inductance (coil) placed in antenna circuit of transmitting set for purpose of increasing wavelength of the open or antenna circuit to any desired level. It is generally referred to as an antenna loading inductance or loading coil. In high power transmitting stations it is very often the case that the natural period or fundamental wavelength of the antenna is not as great as the intended operating wavelength. In this case a loading inductance is used to bring it up to the desired value. In some instances the fundamental wavelength is too high, in which case a condenser is placed in series to reduce the wavelength and then a few turns of inductance placed in series with the condenser and antenna for the purpose of tuning. Such inductances must be of a very rugged nature when high power is used, generally consisting of a comparatively few turns of heavy copper ribbon mounted on high tension insulators. The illustration shows a common form of antenna inductance used in radio transmission. (See *Loading Inductance,* also *Tuned Antenna.*)

INDUCTANCE COILS—Any coil of wire wound in such a manner as to possess the property of self-induction. Such a coil may be in almost any conceivable form, providing the turns are separated or insulated from each other. Inductance coils are a very necessary part of radio circuits, being used in conjunction with condensers to obtain resonance or desired wave-length. Such coils can also be arranged as separate circuits and placed in inductive relation to each other, thus permitting energy present in one to be transferred through induction to the other. The most common example of inductance in a receiving set is that of a tuning coil or vario-coupler. The illustration shows several forms of inductances known as tuning inductances. (See under separate headings as follows: *Inductance, Loading, Tuning, Tapped,* also *Inductance, Units of.*

Honeycomb type inductance coils.

Low-loss air-wound inductance coil made especially for tuned radio frequency work.

INDUCTANCE, DISTRIBUTED—Added inductance distributed throughout the length of a cable or telephone line to compensate for the capacity of the line. (See *Distributed Inductance.*)

INDUCTANCE, LOADING—Literally an inductance to load a circuit. The term is used to designate any inductance coil or unit used in series with the antenna for the purpose of increasing the wave-length of the circuit. (See *Inductance, Antenna.*)

INDUCTANCE, MEASUREMENT OF—Any process or means of determining the inductance value of a coil or circuit. The usual method is by formula. (Fully described under *Measurement of Inductance.*)

INDUCTANCE, MUTUAL—Symbol M. The number of linkages of flux lines or the flux that is common or mutual to two circuits which are adjacent to each other or inductively coupled. It is generally defined as the amount of the *flux* (q.v.) which is common to two linked circuits per unit of current flowing in the inducing circuit. In the illustration circuit A contains an inductance L_1 which is being traversed by an alternating electromotive force. Circuit B contains merely an inductance L_2 and a measuring device, M. Now, as the alternating E.M.F. passes through coil L_1 it sets up a field of force, or rather flux lines as shown. These lines link with the coil L_2 and an alternating E.M.F. is set up. Now, suppose that the inducing current flowing in circuit A is changing at the rate of one ampere per second and an E.M.F. of one volt is induced in the circuit B, or rather in coil L_2. We then say that the mutual inductance of the two circuits will be one henry. This is the same unit used for self-inductance and in fact, mutual inductance and self-inductance are virtually the same thing.

The effect of mutual inductance must be taken into consideration when measuring the fundamental wave-length, capacity or inductance of an antenna. That is to say, where a number of parallel wires are used, the mutual inductance between them must be taken into consideration as well as other values or constants. (See *Self-Inductance, Mutual Inductance, Induction, Coupling* and *Induced E.M.F.*)

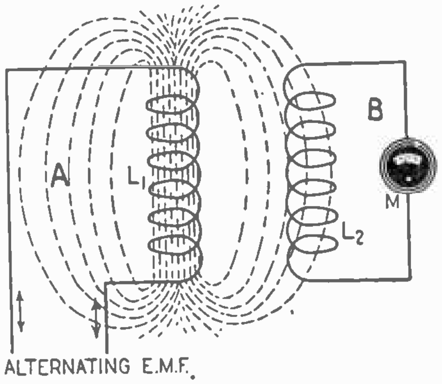

Diagram which illustrates the mutual inductance between coil L_1 in circuit A and coil L_2 in circuit B.

INDUCTANCE, REACTANCE—Symbol x—The *reactance* present in a coil possessing self-inductance. The self-inductance in such a coil opposes the flow of alternating current due to the back *electromotive force* set up as a result of the changing magnetic field of the current. *Inductance reactance* (or *inductive reactance*) is the term used to define this opposition. *Inductance reactance* acts similarly to *resistance* (q.v.) in that it opposes the flow of an electric current. It is measured in *ohms* (q.v.), just as *resistance* is measured. It should be noted, however, that *inductance reactance* only opposes the flow of alternating currents, whereas resistance opposes the flow of both direct and alternating currents. Furthermore, resistance in offering opposition to current flow consumes power, while *inductance reactance* merely takes up energy from the circuit momentarily while current is increasing and thereafter gives it back while the current is decreasing.

The combined opposition offered to the flow of alternating current by the resistance and the inductance reactance of a circuit is called the *impedance* (q.v.) of that circuit. The voltage drop due to resistance is in *phase* (q.v.) with the current. The voltage drop due to *inductance reactance* is 90 degrees in advance of the current.

Expressed as a formula, $x = \sqrt{Z^2 - R^2}$, where x is the inductance reactance, Z is the impedance and R is the resistance, all measured in ohms.

Knowing the *frequency* (q.v.) of the alternating current and the inductance of the circuit through which this current flows, it is possible to calculate the *inductance reactance* by using the formula $x = 2\pi fL$, where π is a constant equal to 3.1416, f is the frequency in *cycles* per second, and L is the *inductance* of the circuit in *henry's* (q.v.). This formula is sometimes abbreviated by substituting the symbol ω (omega) for $2\pi f$. From the formula it can be seen that the *inductance reactance* increases directly both with the inductance of the circuit and with the frequency of the current flowing. (See *Current,* also *Electromotive Force, Resistance, Reactance.*)

INDUCTANCE, RECEIVING—Broadly, any form of inductance used in receiving circuits. The term thus includes all types of coils for tuning purposes, such as couplers, tuned transformers, variometers, etc. The illustration shows a common form of tuning inductance. Here the coil L is an inductance in the aerial circuit and L_1 the coil in the secondary circuit. Here the energy from the aerial traverses coil L and by induction is transferred to coil L_1. (See *Tuning Coil, Vario-Coupler, Receiving Circuit.*)

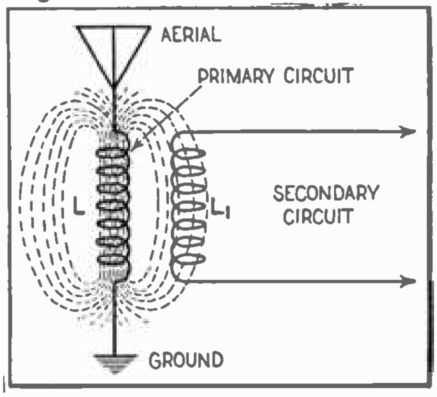

A form of tuning inductance sometimes used in radio receiving sets.

INDUCTANCE, SELF—Symbol L—An alternate term with self-induction for the *inductance* (q.v.) due to the field of force produced by the circuit itself. It is virtually the same as mutual inductance. For example, in a coil of wire being traversed by alternating currents the adjacent turns may act upon each other on the principle of mutual induction between two separate adjacent circuits. The induced currents are the result of self-induction. (See *Self-Induction,* also *Mutual Inductance* and *Induction.*)

INDUCTANCE, TAPPED—Any type of inductance having taps for connecting at different points on the coil in order to vary the inductance value. The illustration shows a form of tapped inductance. Here the taps at different portions of the coil are connected to switch points and the switch lever permits instant change to any of the available values. (See *Inductance Coils,* also *Tuning Coil.*)

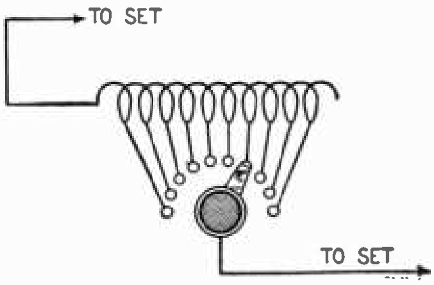

A tapped inductance which permits inductance to be added or taken out of the circuit by rotating the lever switch.

INDUCTANCE, UNITS OF—The measuring units for inductance. The practical unit is the henry (H), the value of inductance which, if current varies at the rate of one ampere each second, results in an *induced electromotive force* (q.v.) of one volt. When we speak of an induced E.M.F. of one volt, the volt is used as a practical unit. Under the electromagnetic system of units, the absolute unit of induced E.M.F. is known as the abvolt. This unit is very small, 100,000,000 of them being required to equal one volt. We thus say that the volt of induced E.M.F. is equal to $1x10^8$ abvolts. The electromagnetic unit of E.M.F., or the abvolt, is defined as the E.M.F. induced in a conductor cutting one line of magnetic force each second. As we have shown that one volt is equal to $1x10^8$ abvolts of induced E.M.F., and since one ampere is equal to $1x10^{-1}$ abamps of current (see *Units, Electromagnetic*) then the unit of inductance or one henry $= \dfrac{1x10^8}{1x10^{-1}}$ or) $1x10^9$ abhenrys. The unit of inductance (the henry) being too large for practical purposes, the *milli-henry* (one-thousandth henry) or the *micro-henry* (one-millionth henry) are commonly used. Occasionally the *centimeter* is used as a measure of inductance, being the thousandth part of a micro-henry or a *milli-micro-henry*. The milli-henry is written as $1x10^{-3}$ henry, the micro-henry as $1x10^{-6}$ henry, and the centimeter as $1x10^{-9}$ henry. (See *Induced Current, Induced E.M.F., Induction, Electromagnetic,* also *Units.*)

INDUCTANCE, VARIABLE—Any coil possessing self-inductance and being arranged in such a manner that the value of the inductance can be changed at will. The simplest form of variable inductance is the sliding contact tuner. This is merely a coil of wire of any predetermined value (maximum) mounted and arranged with a sliding contact moving over the wires, which are bared of insulation at a certain point to permit contact. By using the sliding contact as one connection and either end of the coil as the other lead, the effective value of the inductance in the circuit may be varied within the limits of the coil itself. Other forms of variable inductance include the variometer, vario-coupler and various modifications. The illustration shows the two more common forms of variable inductance. (See *Inductance,* and *Inductance Coils.*)

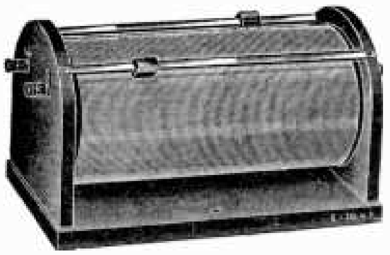

A sliding contact tuner formerly used in radio receiving sets.

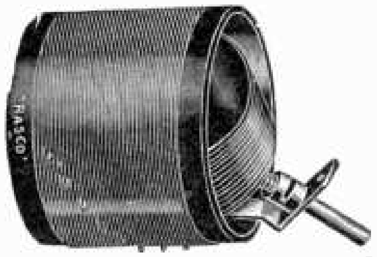

Illustration by courtesy of Radio Specialty Co.
Variable inductance known as a variocoupler.

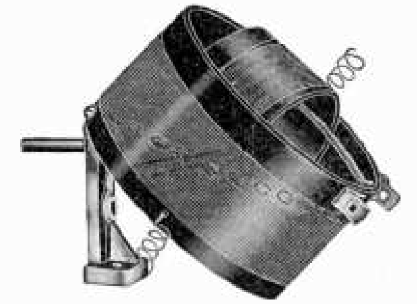

Illustration by courtesy of Radio Specialty Co.
Another type of variocoupler.

INDUCTANCES IN PARALLEL—Coils possessing the property of self-induction or self-inductance, connected in parallel or shunt to each other. The illustration shows two standard coils connected in a circuit in this manner. The effective inductance is less than that of either coil alone. (See *Inductance Coils.*)

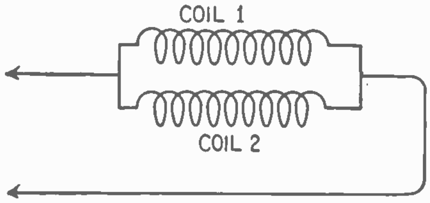

Diagram showing connection of two inductance coils in parallel in a circuit.

INDUCTANCES IN SERIES—Coils possessing the property of self-induction, connected in series in a circuit. The illustration shows two coils connected in series in a circuit. Here the effective inductance value is the sum of the separate inductances. (See *Inductance Coils.*)

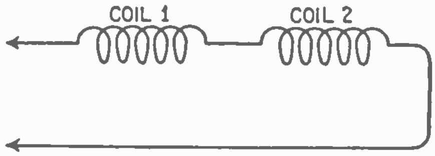

Diagram showing two inductance coils in series in a circuit.

INDUCTION—The influence exerted by one field of force (electromagnetic or electrostatic) on another or on a conductor. Whenever a magnetic field interlinked with an electric circuit is changed, an *electromotive force* (q.v.) is induced in that circuit. The magnetic field is considered to originate from the heart of the conductor, spreading out concentrically from the conductor as the current increases and shrinking back into the center of the conductor as the current decreases. In either case, the conductor is cut by the magnetic field, and in accordance with the *law of electromagnetic induction* (q.v.) an electromotive force is induced in the circuit proportional in amount to the rate of change of the magnetic field and acting in a direction which would, by producing a current, tend to oppose the change.

A bothersome type of electromagnetic induction sometimes occurring in radio takes place when a radio aerial is run parallel to a high-tension line or even a trolley line. The varying field of force about the high-tension line sets up, by induction, a corresponding field in the parallel aerial, thus causing a disturbing hum in the loud speaker. Such a condition can usually be eliminated by running the aerial at right angles to the power line. By doing this, induction between the two conductors is reduced to a minimum.

Electrostatic induction is a re-arrangement of an electrostatic field whereby a body in the neighborhood of a charged body will have an opposite charge induced on it. An *induction machine* (q.v.) known as the Wimshurst utilizes the principle of electrostatic induction.

INDUCTION BRIDGE—Also called inductance bridge. A balance arranged similarly to a *Wheatstone Bridge* (q.v.), used in induction measurement. (See *Induction, Measurement of Inductance,* also *Induction Density.*)

INDUCTION COIL—A form of transformer, having a primary winding arranged with a means of making the direct current applied to it, pulsate. Essentially a transformer with open magnetic circuit, carrying a pulsating direct current which induces a high voltage, alternating current in the secondary by means of the step-up effect of the windings. The illustration shows a typical form of induction coil and the associated circuits for producing high voltages. Here, there are two windings, a primary and a secondary, over and insulated from a core, composed of a number of soft iron wires. The primary is arranged with a key, K, and a battery, B, to supply direct current to the circuit. When the key is depressed the current flows through the primary circuit, including the contact C and the vibrator V. The core becomes *saturated* (q.v.) or magnetized and attracts the armature or spring V. As this spring is drawn toward the core it breaks the contact between the spring and the contact post. This, of course, breaks the circuit from the bat-

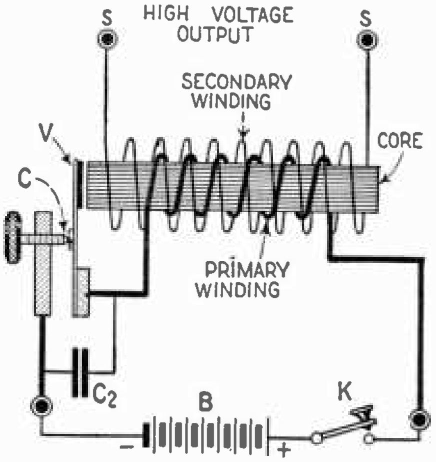

A typical form of induction coil showing diagram of connections, vibrator, core, battery, condenser and key.

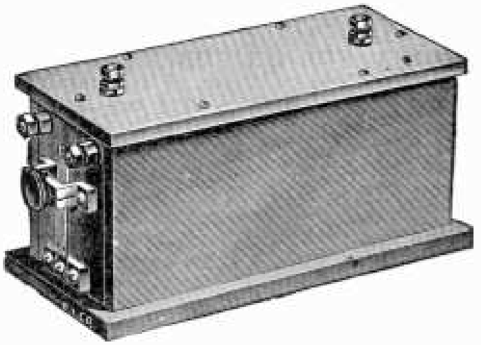

A form of spark coil.

tery and the flow of current stops, resulting in the collapse of the flux through the core. When the magnetic attraction ceases the spring goes back into normal position and the contact is once more complete, allowing current to flow through the windings. This make-and-break process is more or less rapid and is repeated continuously. By induction, an alternating E.M.F. (electromotive force) is created in the secondary windings, and as these windings have many more turns than the primary winding the induced E.M.F. will be proportionately greater. If the breaks occur at a rate of from thirty to one hundred per second, the key can be used to interrupt the primary current and produce at the secondary terminals a high voltage suitable for spark transmission. (See *Induction, Transformer, Spark Coil.*)

INDUCTION COIL, PRIMARY—The simplest form of induction coil. It consists of an iron core wound with a relatively large amount of low resistance copper wire, which is well insulated. The iron core increases the effect of self-induction and the size or length of the core and the number of turns of wire determine the efficiency of the coil. Such coils are used extensively for low tension ignition. (See *Induction Coil.*)

INDUCTION COIL, SECONDARY—A type of induction coil employing a primary and secondary winding. The primary is wound over an iron core and the secondary is generally wound over the primary. (See *Induction Coil.*)

INDUCTION DENSITY—A term occasionally used to refer to the density of flux in a magnetic or electromagnetic field of force. It is synonymous with flux density, under which heading it will be found more fully defined. (See *Induction, Electromagnetic,* also *Gauss,* and *Induction, Magnetic.*)

INDUCTION, ELECTROMAGNETIC—The production or *inducing* of electric currents in a conductor when it is moved in a magnetic field of force in such manner as to cut magnetic lines of force. The simplest example of this phenomena is the Faraday experiment in which he discovered that a wire with its ends joined, when moved rapidly in the field of a magnet, would result in current being induced in the wire or conductor. All dynamo electric machinery is based upon this principle of electromagnetic induction, as are also alternating current transformers and various other electrical devices.

In the illustration is shown the theory of production of current by electromagnetic induction. Fig. 1 shows a coil of wire placed in a uniform magnetic field of force between two magnets. Here the coil is assumed to have been moved from A to B in the same plane. This is known as a *motion of translation,* and the field being uniform at all points the same number of lines pass through the coil at both points. Therefore no current is induced in the coil and a sensitive instrument placed across the leads, C, C, would show no current present. In Fig. 2 the same coil is shown at A in its original position. Here, however, it is rotated, and when in position B it is apparent that less lines of force pass through it. We have thus fulfilled the conditions for inducing current, and if the instrument is placed across the leads, C, C, the presence of electric current will be indicated. If the motion of rotation is continued the number of lines of force passing through the coil will continue to increase and decrease and current will be produced in the coil as long as the motion persists. From the above it will appear that current can be induced either by rotating the coil around any axis in its plane, or by tilting it in its motion across the field. The coil of wire in this case is known as an *inductor* because the current is induced in it. It is also referred to as a *conductor* when describing the production of alternating currents.

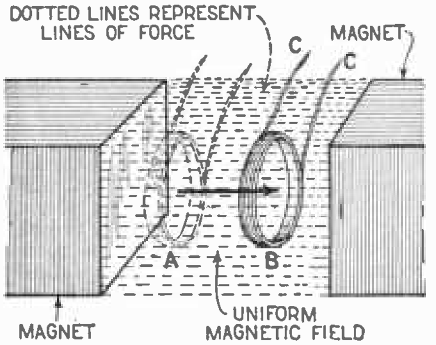

Fig. 1. Coil of wire in a uniform field moving parallel to direction of the field.

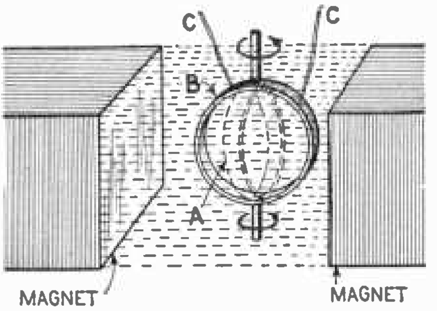

Fig. 2. Same coil, showing cutting of field due to the rotation of the coil.

Now it was stated that current is induced in the inductor whether the lines of force through it are *increased* or *decreased* by its motion through the field. We can reduce the action to two simple rules. First, the relative motion between an inductor and a magnetic field must be such that the number of lines passing through the inductor are altered, i.e., increased or decreased. Second, the electromotive force induced in the inductor circuit will be proportional to the rate of increase or decrease in the number of magnetic lines embraced by the inductor circuit. We know from the *Lenz Law* (see *Alternating Current, Theory of Production of*) that current will flow in the inductor in a direction dependent upon whether the lines of force through it are being increased or decreased, and as the inductor is being rotated continuously the current will change direction with each change in the number of lines. (See *Alternating Current, Alternator, Electromotive Force, Fleming's Rule, Laws of Electromagnetic Induction, Lenz's Law.*)

INDUCTION, ELECTROSTATIC—The production of an electric charge in a body due to the presence of a nearby conductor having an opposite charge. The effect is known as a displacement of electric charge. (See *Electrostatic Induction,* also *Induction.*)

INDUCTION INSTRUMENTS—Measuring instruments used in electrical and radio practice, which make use of the principle of electromagnetic induction. Such instruments are used for measuring alternating currents, the torque or actuating force for the indicator being produced by the effect of a rotating field on a metal disc or drum which is not connected to the circuit carrying the currents. The instruments operate on the same basic principle as the *induction motor* (q.v.).

INDUCTION MACHINE—A machine wherein the primary and secondary windings rotate with respect to each other. Examples of induction machines are induction motors, certain types of frequency converters and induction generators. An entirely different kind of induction machine is one which generates low potentials, utilizing the principle of *electrostatic induction* (q.v.). The most common form of this type is known as the Wimshurst (q.v.).

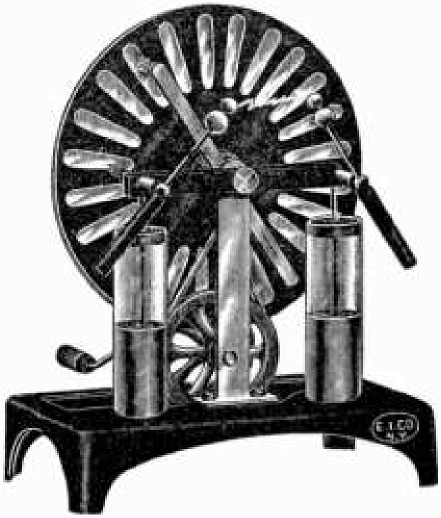

A common type of induction machine known as the Wimshurst.

INDUCTION, MAGNETIC—The communication of magnetism to a metal by the presence of a magnet without any actual contact. The effect of the presence of a magnet in the neighborhood of a piece of magnetic material is to induce magnetic poles in it. (See *Magnetism, Magnetic Poles, Induction,* also *Induction, Electromagnetic.*)

INDUCTION MOTOR—An alternating current motor in which the input current is passed through the field coils only, the armature not being connected to the external circuit. Here the armature is rotated by currents induced by the varying field set up through the field coils. The induction motor is the most widely used of the alternating current types, maintaining a practically constant speed from no load to full load and being simple and rugged in construction. This type of motor has two windings, the primary and secondary, and two members called stator and rotor. The primary winding is usually on the stator and the secondary winding, which is therefore customarily on the rotating member, is not connected electrically to the primary. Current to operate the rotor is induced in the secondary windings by the magnetic action of the current circulating in the primary windings. (See *Alternating Current,* also *Induction, Electromagnetic.*)

INDUCTION, MUTUAL—The interference or mutual effect between two electric or magnetic fields, due entirely to their proximity and without electrical contact. The mutual induction or electromagnetic influence of one circuit upon another is measured by the coefficient of mutual inductance or induction. (See *Mutual Induction,* also *Mutual Induction Coefficient,* and *Mutual Inductance.*)

INDUCTION REACTANCE—The value in ohms of the inductance in an electrical circuit as distinguished from the capacity or capacitive reactance. It is the part of the total impedance in an alternating current circuit that is due to the presence of inductance in the circuit. (See *Inductance Reactance, Ohm, Impedance.*)

INDUCTION REGULATOR—A system of regulating the voltage in alternating current circuits, employing a choke coil with removable core, or some form of variable ratio transformer which will assist or oppose the current by readjustment of the ratio. (See *Choke Coil, Reactance,* and *Variable Ratio Transformer.*)

INDUCTION SCREEN—A metal screen or shield placed between two electrified or magnetic bodies to reduce the effect of induction between them. A common example of this in radio usage is the placing of a sheet of metal between two coils, the sheet being generally grounded, or the provision of a metal *shield* (q.v.) entirely around each coil. This reduces the interaction of the fields of the coils, which leads to losses or howling, especially in *tuned radio frequency amplifiers* (q.v.). (See *Shielding.*)

INDUCTION UNIT—The *henry,* defined as the induction in a circuit when the electromotive force induced in that circuit is one *international volt* (see *volt*) while the inducing current varies at the rate of one ampere each second. (See *Henry,* also *Self-Induction.*)

INDUCTION VOLTMETER—See *Induction Wattmeter,* also *Induction Instruments.*

INDUCTION WATTMETER—An instrument for measuring the power delivered to a circuit. The induction type wattmeter operates on a principle used in the induction motor, a magnetic body attached to a pointer being rotated by a changing magnetic field. (See *Induction Motor,* also *Wattmeter.*)

INDUCTIVE CAPACITY—Generally referred to as the specific inductive capacity in rating the dielectric qualities of a substance. It is the quality of any dielectric substance by reason of which it permits electrostatic induction through it. The term is more or less synonymous with *Dielectric Constant,* the latter, however, being generally given in terms of a numerical value representing its comparative specific inductive capacity with some material taken as a standard. (See *Dielectric Coefficient and Constant,* also *Specific Inductive Capacity.*)

INDUCTIVE CIRCUIT—An electrical circuit possessing inductance but not capacity. (See *Inductance,* also *Capacity.*)

INDUCTIVE CONNECTION—A connection between two circuits entirely through the property of induction and without any metallic contact. The two coils, primary and secondary, of a vario-coupler as used in radio receiving circuits, are said to have inductive connection where the windings are separate. Thus, the primary coil will be in one electrical circuit and the secondary coil will receive energy from it by means of induction. (See *Induction, Coupling, Vario-Coupler.*)

INDUCTIVE COUPLER—The inductances or coils placed in inductive relation or inductively connected with each other for the purpose of transferring energy by electromagnetic coupling. Any device consisting of a primary and secondary inductance and used to transfer energy from one circuit to another by means of induction. (See *Coupler, Induction, Vario-Coupler.*)

INDUCTIVE COUPLING—The coupling or connection between two circuits by means of electromagnetic induction.

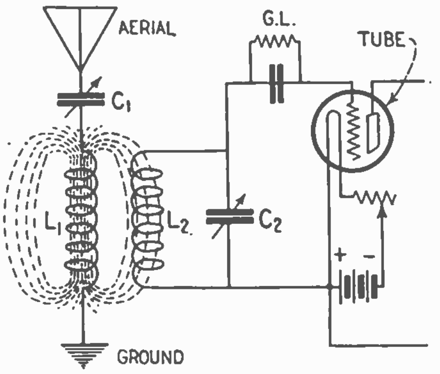

The utilization of a vario-coupler to inductively couple the aerial system with the rest of a receiving circuit.

The most common example of inductive coupling is in radio receiving circuits where coupling coils are used to associate the primary or aerial circuit with the secondary circuit. In the illustration L_1 is the primary of a vario-coupler and L_2 the secondary. The primary is inserted between the ground and aerial and the secondary is connected to the grid and filament of the detector tube. The variable condenser C_1 tunes the primary circuit, the primary coil also being variable by means of taps if desired. The variable condensed C_2 tunes the secondary inductance and the condenser and grid leak GL are the customary elements in the detector grid circuit. As the high frequency currents traverse the primary circuit, passing through coil L_1, lines of force are set up as shown and some of these lines embrace the secondary coil L_2, inducing similar currents in the secondary or vacuum tube circuit. The balance of the circuit is not shown for obvious reasons. This type of coupling between circuits is much more selective than the straight coupling or conductive coupling. (See *Coupling,* also *Conductive Coupling.*)

INDUCTIVE DISTURBANCE—The disturbing effects sometimes experienced in radio broadcast or telegraph reception due to induction from nearby power or telephone lines. This sometimes occurs when an aerial is strung close to and parallel with a line carrying high voltages, and occasionally is experienced with telephone wires near the receiver. (See *Induction.*)

INDUCTIVE DROP IN VOLTAGE—The drop or diminution of voltage in an alternating current circuit due to the presence of inductance. Inductance in a circuit carrying alternating current tends by reason of the self-induction to set up a counter-electromotive force which opposes the original E.M.F. (electromotive force), thus reducing the effective voltage in the line. (See *Inductance,* also *Inductance Reactance.*)

INDUCTIVE EMF—The electromotive force due to induction. That is to say, the electromotive force induced in a circuit. (See *Induction, Induced EMF.*)

INDUCTIVE REACTANCE—The reactance or portion of the total impedance in a circuit, due to the presence of self-inductance. (See *Inductance Reactance,* also *Impedance.*)

INDUCTIVE RESISTANCE—A resistance element possessing inductance, as for example a resistance coil, which also has the property of self-inductance. (See *Resistance,* also *Inductance.*)

INDUCTIVE RISE—The rise in voltage noted in transformers or other alternating current apparatus due to the presence of a *leading current* (q.v.). (See *Inductance,* also *Transformer* and *Angle of Lead or Lag.*)

INDUCTIVITY—Another term used alternately with dielectric constant and conductive capacity, referring to the dielectric properties of materials. (See *Dielectric Coefficient or Constant,* also *Inductive Capacity.*)

INDUCTOR—The coils or conductors in which current is induced through electromagnetic induction, as in an induction motor. Any conductor in which current is induced due to a change in magnetic flux may be called an inductor. (See *Induction, Inductor Alternator,* and *Induction Motor.*)

INDUCTOR ALTERNATOR—An alternator for producing high frequency currents. It employs the *armature* and *field windings* on projections inside the *stator,* the rotor consisting of a drum carrying a magnetic material. By using a *rotor* of solid steel, very high speeds and consequently high frequency currents are obtainable. The *Alexanderson Alternator* (q.v.) is an inductor type alternator. (See *Alternator,* also *Inductor.*)

INDUCTOMETER—A device or instrument used to measure self or mutual inductance. The basic principle depends on the relation of two coils, a primary and secondary, one of which can be moved in its relation to the other, the inductance, either self or mutual, being registered on a scale calibrated in units of inductance. (See *Inductance, Mutual,* also *Inductance, Self.*)

INERTIA—In physics, the property of a body which tends to keep it in a state of rest or to resist any change of motion. (See *Inertia, Electric.*)

INERTIA, ELECTRIC—Term occasionally applied in place of *self-inductance* (q.v.). When a current flows in a circuit possessing inductance and the circuit is broken, current continues to flow for a short time. It thus appears that in this instance the resistance alone will not stop the flow of current on the instant of breaking the circuit, an interval of time being required as in the act of bringing to rest a moving material. (See *Inertia, Inductance, Self,* also *Resistance.*)

INERTIA, ELECTROMAGNETIC—The same as electric inertia, being, in effect, the energy required to start or stop a a current in a circuit possessing self-inductance. (See *Inertia.*)

INFERRED ZERO—A term used in connection with certain instruments having extremely high sensitivity, where the zero is removed from the scale. In a galvanometer, for example, the zero position on the scale may be merely assumed, an electrical or mechanical force being applied to bring the zero off the scale entirely, only a part of the full range of the full scale being utilized. (See *Galvanometer.*)

INFLUENCE—A broad term signifying action at a distance, as by electrostatic induction, without any actual physical contact. It is considered as the effect of a charged body on a neutral body or conductor coming within its field. (See *Field*, also *Induction.*)

INFRA-RED RAYS—The heat waves or rays which lie between light waves and electromagnetic (radio) waves. The infra-red lies at the extreme low frequency end of the invisible spectrum, the wave-length being extremely high and the frequency conversely low. (See *Wave*, also *Ether Waves.*)

INFRA-SATURATION PART OF CHARACTERISTIC—In a characteristic curve of a vacuum tube, this is the portion over which the tube is operable, or the operating portion of the curve. (See *Characteristic Curve*, also *Vacuum Tube, Theory of Operation.*)

INITIAL MAGNETIZATION — The initial stages of magnetization as distinguished from the effect after saturation has been approached or reached. (See *Saturation, Magnetic.*)

INPUT—Generally speaking, the energy absorbed by a machine as distinguished from the output energy. (See *Input Voltage*, etc.)

INPUT CIRCUIT OF VACUUM TUBE—Vacuum tube circuits are generally divided into three separate classes or circuits. The first is the circuit including the filament, i.e., the "A" battery circuit through the filament. The second is the circuit between filament and grid and the third is the circuit between filament and plate. The filament grid circuit is called the input circuit because the incoming oscillations are impressed on this circuit. The filament plate circuit is known as the output because the rectified or amplified currents are taken from across the plate and filament. The illustration shows both input and output circuits, the former in heavy lines and the latter in light lines. The input circuit of a tube is also known as the *grid circuit.* (See *Grid, Filament, Plate,* also *Vacuum Tube, Theory of Operation.*)

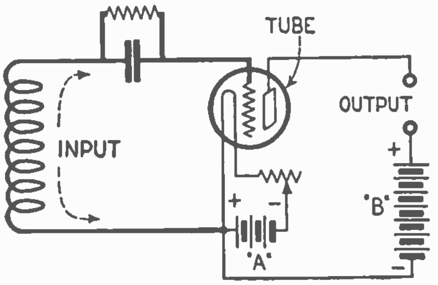

Input and output circuits of a vacuum tube receiver. Heavy lines show the former, light lines the latter.

INPUT VOLTAGE — The voltage impressed on a circuit or machine as distinguished from the output voltage. (See *Input.*)

INSTANTANEOUS CURRENT — The value of current in an alternating current circuit measured at any instant, as distinguished from the average current value. (See *Instantaneous Values.*)

INSTANTANEOUS VALUES—The actual value of an alternating current, E.M.F. (electromotive force), etc., measured at any instant. This value may be anything from zero to maximum or peak It must be remembered that in an alternating current the current and voltage are both variable factors, rising to a maximum in one direction, then falling to a minimum, and then rising to a maximum in the other direction. Thus, a wattmeter measures the actual value of current and voltage at each instant and the instantaneous torque at each instant is proportional to the instantaneous power. The average torque acting on the moving element during each cycle of the current and voltage is then proportional to the average power. If we measure an alternating current with an ammeter, and then measure the voltage with a volt-meter, the product of the two will be the apparent power, (apparent watts). It should be noted that these instruments measure effective and not instantaneous values. In order to ascertain the true *watts* (q.v.), it is necessary to multiply the apparent watts by the power factor. A wattmeter measures true watts. (See *Effective Electromotive Force, Power Factor,* also *Wattmeter.*)

INSTANTANEOUS VOLTAGE OR PRESSURE—The voltage or pressure in an alternating current circuit measured at any instant, as distinguished from the average value of E.M.F. (electromotive force). (See *Instantaneous Values.*)

INSTRUMENTS, MEASURING — In electricity or radio, any device or instrument for recording or indicating values, such as an ammeter for indicating the current in amperes, or a voltmeter to indicate the electromotive force in volts. (See *Ammeter* or *Ampere Meter, Voltmeter, Wattmeter.*)

INSTRUMENT SHUNTS—In electrical measuring instruments, a branch conductor, generally of relatively heavy wire or metal strip, joining the meter circuit at two points and forming a parallel circuit, thus dividing the current and protecting the delicate windings of the instrument. (See *Shunt,* also *Ammeter,* or *Ampere Meter, Hot Wire Ammeter.*)

INSULATE — To provide a protective covering or shield for a part or conductor, preventing flow of electricity from it or the passage of electricity through it. A conductor of electric currents is generally insulated to prevent the flow of the electricity through contact with other parts or conductors except where contact is intended to be made. (See *Insulated Wire.*)

INSULATED WIRE — Wire, generally copper, covered with insulating material to prevent leakage of the electric current. Standard insulated wire of the heavier gauges is usually covered with a layer of rubber and an outer covering of braided cotton. (See *Lead-In,* also *Insulation.*)

INSULATING COMPOUND—Liquid or easily liquifiable mixture of some insulating material, used in radio to coat cardboard tubes and to cover any open point of contact in an otherwise insulated conductor or part. Coils are very often coated with shellac or some other liquid and quick-drying substance. This serves to hold the windings in place and also acts as a protective covering. A very valuable application of an insulating liquid is shown in the illustration. Here an inductance is wound on a form of stiff paper or cardboard tube and after being completed is coated with one or two treatments of collodion or similar liquid. (Shellac introduces losses through the medium of *distributed capacity.*) The form is then cut away and the coil remains self-supporting as shown. This method produces a coil having the least possible dielectric and distributed capacity loss for a cylindrical shape. (See *Dielectric, Inductance Coil, Low Loss Coils, Collodion.*)

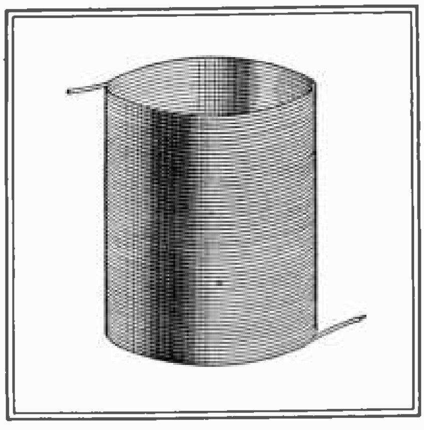

Air core coil made rigid by use of insulating compound.

INSULATING MATERIALS—Non-conducting substances, for example, glass, hard rubber or porcelain, which do not conduct electric currents. Strictly speaking, there are no perfect insulating materials. The most perfect insulators are certain gases, although at low pressure these may act as very good conductors. The more common insulating materials in radio are the following: mica, glass, bakelite and hard rubber, the insulating values being in that order. When we speak of the insulating property of a material we are actually referring to its resistance, a good insulator offering infinite resistance to the passage of electric currents and poor ones being relatively good conductors. The term *specific resistance* is used extensively in connection with the insulating properties of materials. This term is defined as the resistance of a piece of material of unit length and unit cross-sectional area at a given temperature, the value being given in ohms. (See *Dielectric Coefficient and Constant, Table of,* also *Ohms, Resistance* and *Insulate.*)

INSULATING PROPERTY—The ability of any material to *insulate* electric currents. Accepting mica as an insulator, we say that its insulating properties are superior to glass or hard rubber. (See *Insulating Materials,* also *Dielectric Coefficient and Constant* and *Insulation of Aerial.*)

INSULATION—Broadly any material or gas used to insulate electric currents. In radio, insulation plays a most important part, providing protective coverings or separators between conductors of the electric current and between the conductors and the ground. Insulation is really a material or gas having extremely high specific resistance; the better the insulating property, the higher the resistance to passage of electricity. There is no clearly defined line of demarcation between an insulator and a conductor, a poor insulator being to some extent a conductor. (See *Bakelite, Conductor, Dielectric Coefficient and Constant, Insulations of Aerial, Panel.*)

INSULATION OF AERIAL—The means used in connection with an aerial for reception or transmission of radio waves, to prevent leakage or ground-

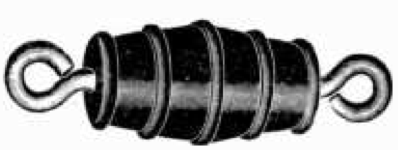

Aerial Insulator.

ing of the currents. The insulation of an aerial for reception purposes should be given very careful consideration. The wires may be suspended by means of small glass or porcelain insulators and the lead-in rubber covered wire, entering the building through a *lead-in insulator* (q.v.). The aerial for transmission must be accorded specially careful attention as far as insulation is concerned. The bare wires must be suspended with heavy duty insulators, capable of withstanding high voltages without *"breaking down."* The lead wire must be conducted into the building through a high-tension insulating tube or other lead-in device to prevent leakage of the current to the ground. (See *Aerial, Insulation of Aerial* and *High Tension Insulators,* also *Lead-In Insulator.*)

INSULATION RESISTANCE—The actual resistance in ohms, generally megohms, of the insulation of any conductor or piece of apparatus. If we assume two conductors, carrying separate currents and crossed so that the only protection would be the actual insulation of the wires, the insulation resistance would be the actual resistance in megohms from one path to the other. Conceivably a small current might leak from one conductor to the other, this current, of course, depending on the resistance of the insulation. The term "insulation resistance" is more easily understood when it is considered that an insulator or insulating material is merely a very poor conductor, or one having extremely high resistance to electric currents, there being actually no perfect insulator. (See *Insulation, Insulated Wire,* also *Conductor* and *Resistance.*)

INSULATION TESTER—An instrument or set of meters for testing the resistance of an insulator in fractional ohms, generally megohms. (See *Insulation Resistance.*)

INSULATOR—Any substance having insulating properties. *Insulators* are used to keep electric currents flowing in predetermined paths. It is necessary to use insulators on all electric circuits in order to prevent *grounds, leakage* and *short circuits.* Insulators are of the utmost importance in radio. At the transmitting end, where high voltages as well as high frequencies are used, the insulators must be specially designed to prevent leakage. A typical installation of insulators in use on the aerial system of a modern broadcasting station utilizes a 12-inch pyrex insulator at each end of each wire between the spreaders. A standard porcelain strain insulator is used every 14 feet on each guy wire. At each end of each guy wire a standard 24-inch porcelain insulator is used.

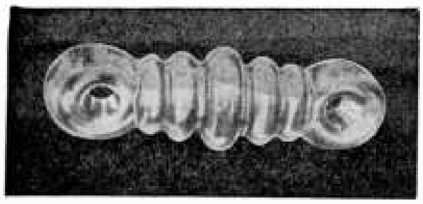

Illustration by courtesy of Corning Glass Works.
A typical small antenna insulator, showing corrugated surface.

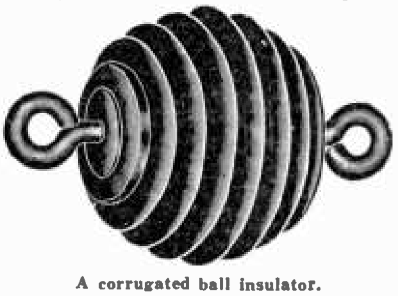

A corrugated ball insulator.

Insulators used on receiving antennas are made of porcelain, pyrex, rubber and composition. These are usually about 4 inches in length. Sometimes porcelain tubes are used as insulators to bring a *lead-in* (q.v.) into a building, but these have been generally superseded by flat *window lead-ins.*

INSULATOR, LEAD IN—The insulating tube through which a lead wire from a radio aerial enters a building, used to prevent grounding or leakage of the current by contact with another conducting surface. (See *Lead-In Insulator.*)

A flat window lead-in which permits entrance of lead-in into building without the necessity of boring holes.

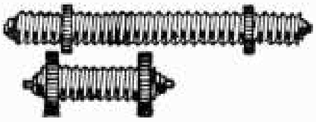

Lead-in insulators of conventional type.

INTEGRATING DETECTOR—The name sometimes given to any detector which yields a response proportional to the total energy received from a spark train rather than to the maximum value of the current or voltage in the train.

INTEGRATING INSTRUMENTS—Meters so designed as to record the total quantity of electricity (ampere hours) or the total energy (watt-hours) passing in a circuit in a given time. (See *Ammeter* or *Ampere Meter,* also *Integrating Wattmeter.*)

INTEGRATING WATTMETER—An instrument which measures and records the total amount of electrical energy being consumed in a circuit. These instruments are also known as *recording watthour meters* (q.v.), *integrating watthour meters,* and sometimes are simply referred to as *watthour meters* (q.v.).

Broadly speaking the *integrating wattmeter* is a type of small motor whose rotations are counted by means of a worm on the armature shaft engaging a set of cogs working a counter. The construction is such that the average *torque* (q.v.) exerted by the motor is proportional to the average power taken by the load. In other words, the motor rotates at a speed directly proportional to the energy being expended. For direct current work, *integrating wattmeters* are either of the commutator type, such as the Thompson, Westinghouse and Duncan meters or of the no-commutator type, such as the Sangamo. The commutator type instruments operate on the *dynamometer* (qv.) principle. Meters of the no-commutator type are essentially motors whose armatures always cut the same direction flux. That is to say, they are essentially *homopolar* (q.v.) motors. The principle of operation of these meters is the law of motor action which states that a conductor, free to move and carrying a current whose direction of flow is at right angles to a fixed magnetic field, will be moved out of this field, and in a direction at right angles both to the direction of current flow and to the field. For alternating current work, *integrating wattmeters* are of the *induction* (q.v.) type. The principle of operation of these meters is identical to that of *induction motors* (q.v) having shunt and series windings stationary and so related and located as to produce a rotating field acting upon a closed rotable secondary. In the induction meter the secondary consists of a light aluminum disc. These meters may be either of *single phase* (q.v.) or *polyphase* (q.v.) construction. (See *Electro-dynamometer.*)

INTEGRATOR—A device which automatically, by means of clockwork, adds up and records on a dial items of calculation or measurement. The system used in a recording or integrating wattmeter. (See *Integrating Wattmeter.*)

INTENSIFICATION—Broadly an intensifying or increasing of the density of electric current, but specifically in radio it refers to the tendency of radio signals to increase in intensity under certain conditions. The phenomena of intensification is naturally closely related to fading, as obviously the intensification process must necessarily follow or be followed by a fading process. The tendency toward intensification is more pronounced toward night and generally on the shorter wavelengths, it being noted that quite often signals from a distant station will be increased in volume or intensity as much as fifty per cent or more between the late afternoon and late evening. While the periodic *fading* of signals and the subsequent increase in volume over regular periods, usually of a few minutes' duration, may be considered as *fading* and *intensification,* it is preferable to think of fading as one operation together with the subsequent increase in volume, and the term *intensification* as referring strictly to the change to greater volume that takes place toward night. (See *Fading,* also *Phenomena of Electric Wave Propagation.*)

INTENSITY OF ELECTRIC CURRENT—A term more or less in use to imply the strength of an electric current in amperes, the symbol I for current being derived from it. (See *Ampere, Current Density,* also *Intensity of Field.*)

INTENSITY OF FIELD—The intensity or strength of the force exerted in a magnetic field. It is measured by its action or effect on a unit pole placed at any point in the field, the intensity of the field, of course, varying at different points. (See *Flux Density, Intensity of Magnetic Flux.*)

INTENSITY OF MAGNETIC FLUX—The strength of the force exerted in a magnetic field. (See *Intensity of Field,* also *Flux Density.*)

INTENSITY OF MAGNETIZATION—The density of force per unit cross-sectional area of a magnetized material. The extent of the magnetizing effect on a material placed in a magnetic field of force. This applies only to material in a magnetic circuit and should not be confused with the lines of force due to a field arising from the presence of neighboring currents or magnets. (See *Magnetization,* also *Intensity of Magnetic Flux.*)

INTERFERENCE—The interruption or interference with desired electromagnetic waves by undesired or extraneous waves. The interference of undesired broadcast or radio-telegraph signals with the desired signals. Also, broadly, the detrimental effect of power lines and other electric circuits on radio communication. Interference is encountered chiefly in receiving sets having poor *selectivity* (q.v.). In some instances interference may be due to the fact that another transmitter is operating on the desired wave-length, and being sufficiently close as to make it practically impossible to obtain clear signals from the desired station. Ordinarily, and where the effect is due to lack of sharp resonance or tuning, certain changes can be made in the receiver to sharpen the tuning, or if the disturbance is due to the transmitting stations and the receiver is normally sharp, various circuits can be employed to filter out the undesired signals. (See *Interference Preventer*, also *Wave Trap.*)

INTERFERENCE ELIMINATOR—A device or system as shown in the illustration for reducing or preventing reception of undesired signals in radio work. Practically all *interference eliminators* make use of a separately tuned inductance across the aerial and ground and in parallel with the receiving set. The inductance is usually tuned by means of a variable condenser which allows it to be tuned to the wave-length of the interfering signal. The interference can thus be by-passed to the ground. *Interference eliminators* of the kind described are generally known as *wave traps* (q.v.). A special form of interference due to atmospheric and electrical disturbances is known as *static* (q.v.). (See *Static Eliminator.*)

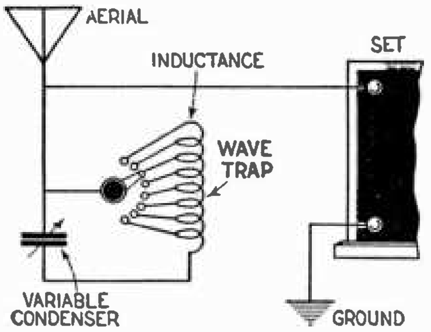

A wave trap or interference eliminator which dispenses with the ground connection.

INTERFERENCE PREVENTERS—Any arrangements either within the receiving set, or external to the set, for limiting or preventing interference. In addition to *interference eliminators* (q.v.) and static *eliminators* (q.v.) various other methods are utilized in radio to prevent interference. Thus the arrangement of a receiving set to give sharp tuning and great selectivity, results in the prevention of much interference. A method of preventing interference which has been used, is the utilization of filter circuits which can be tuned to pass any desired frequency and exclude all others. The use of the *loop aerial* (q.v.) results in great selectivity. The loop has a very pronounced directional effect and this in itself acts to narrow the field of stations offering possible interference. For example, if stations of very nearly the same wave length are in opposite or different directions and are sending at the same time, the loop permits selection of the station wanted, with no interference from the others.

INTERMEDIATE CIRCUIT—A closed circuit consisting of two inductance coils shunted by a variable condenser. One coil acts as primary to the secondary coil of the detector circuit and the other coil acts as secondary to the primary of the aerial circuit. A circuit such as this is sometimes referred to as a *multiple tuner intermediate* circuit and may be used as an interference eliminator.

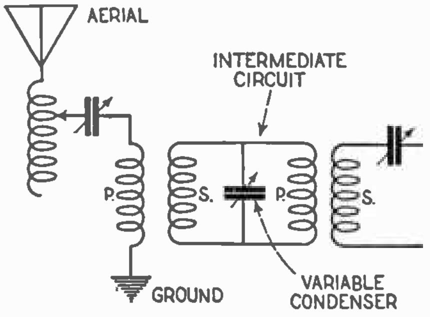

A multiple tuner intermediate circuit.

INTERMEDIATE FREQUENCY—This is a frequency higher in number of oscillations than *audio frequency* (q.v.) but lower than *radio frequency* (q.v.). It is generally around 30,000 cycles or 10,000 meters wave-length. In the *super-heterodyne* (q.v.) the incoming signal is converted from the broadcast frequency to the so-called *intermediate frequency* at which it can be amplified by means of fixed winding transformers to almost any extent desired. The method of converting the incoming signal in a *super-heterodyne* from radio frequency to *intermediate frequency* is by the *beat* method. In this case an oscillating circuit controlled by means of a variable condenser so as to obtain any desired frequency, is brought into interference with the first detector and produces a *beat note* (q.v.) for the *intermediate frequency* which has all the characteristics of the original signal. In this way, no matter what the frequency of the intercepted signal may be, the intermediate frequency amplification is always carried on at a fixed wave-length, thus making for high efficiency. (See *Beat Frequency; Beats.*)

INTERMEMEDIATE FREQUENCY AMPLIFIER—An arrangement of tubes and transformers for stepping up the current, used in the *super-heterodyne* (q.v.) circuit. The intermediate frequency amplifier is located between the first detector and the second detector. (See *Intermediate Frequency.*)

INTERMEDIATE TRANSFORMER—The name given to the transformers used in the *intermediate frequency amplifier* (q.v.) of the *super-heterodyne* circuit. *Intermediate transformers* are often referred to as *intermediate frequency transformers*. They are *radio frequency transformers* (q.v.) but have an iron core and instead of tuning to a maximum of 550 or 600 meters they are designed to cover wave-lengths of 10,000 meters or higher. (See *Intermediate Frequency.*)

INTERMITTENT CURRENT—A current which flows irregularly or which is interrupted at intervals, or without continuity. The current obtained from a *magneto generator* (q.v.) is an interrupted or *intermittent current.* Such a current may be either alternating or direct. A magneto generator constructed to give direct current is equipped with a two segment commutator having a stationary brush in contact with it. (See *Pulsating Current.*)

INTERNAL RESISTANCE—Resistance within an electric source. In the case of a primary cell such as a dry battery, the internal resistance causes an interference with the flow of current. This varies in amount with the construction and materials of the battery.

INTERNATIONAL MORSE CODE—This is also known as the *Continental code.* (See *Code.*)

INTERPOLE—A small magnetic pole placed between the main field poles of an electric generator for the purpose of obtaining better commutation. Interpole windings are connected in series with the armature winding and the load. The action of the interpole is exactly analogous to the shifting of the brushes, but when interpoles are used, brush shifting is dispensed with and the magnetic flux is shifted instead.

INTERRUPTED CONTINUOUS WAVES—abbreviation I.C.W.—These waves are obtained by the modulation at audio frequency, during signalling, of an otherwise *continuous wave* (q.v.). (See *Continuous Waves Key Modulated*, also *Continuous Waves Modulated at Audio Frequency.*)

INTERRUPTED WAVES—abbreviation I.W.—Interrupted waves are waves produced by modulation at audio frequency, of otherwise continuous waves. (See *Interrupted Continuous Waves.*)

INTERRUPTER—A combination of an electro-magnet with a vibrating armature which carries a contact point on a piece of spring steel fastened to the armature. When the electro-magnet attracts the armature the contact points are drawn apart thus interrupting the circuit. This action is only momentary for the spring again brings the circuit back to its original condition. The most important use of the interrupter in radio is as a *rectifying valve* (q.v.) in mechanical chargers. A typical vibrator panel is shown in the illustration. The armature, A, is the moving element. This carries at its outermost end a heavy Tungsten contact, B, which is caused to vibrate in synchronism with the supply current, by the actuating coil, C. When the current from the transformer is of the correct polarity and value the actuating coil attracts the armature, closing the charging circuit to the battery through the fixed contact, D. The contact end of the armature, A, vibrates within the air gap of the pole shoe, G.

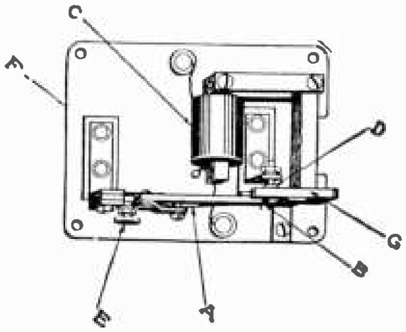

The vibrator panel of an efficient type of mechanical charger.

A common use of the *interrupter* is in the ordinary electric bell. In this apparatus, the electromagnet pulls the armature carrying the spring and contact point, breaking contact and opening the circuit. This immediately de-energizes the magnet since its coils are in series with the circuit. The spring tension pulls the armature back to its original position thus again closing the contact. As soon as the contact is closed, the electromagnet is again energized and it again attracts the armature, opening the circuit. This action keeps on indefinitely as long as a difference of potential is maintained at the terminals of the bell.

The *interrupter* is used in the *induction coil* (q.v.) for breaking up the direct current into a series of impulses, thus producing an intermittent current. (See *Circuit Breaker.*)

INTERRUPTER, ELECTROLYTIC—A jar containing diluted sulphuric acid as the electrolyte, a large sheet of lead as one electrode and a platinum needle point introduced through a glass tube as the other electrode. If these electrodes are connected through an inductance to a source of current supply, the current in the circuit will be rapidly interrupted. *Electrolytic interrupters* are suitable for radio transmitters using small spark coils and are also used in X-ray work. (See *Electrolytic Interrupter.*)

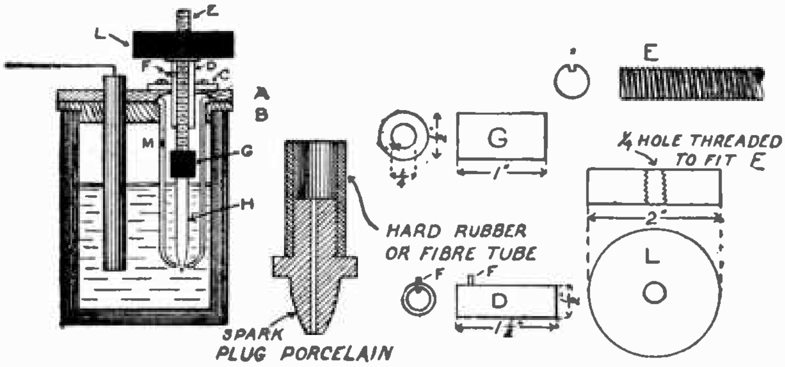

A simple type of Wehnelt interrupter. The positive pole of the direct current supply is connected to the brass part C, while the negative pole is attached to the lead rod. L is a fibre handle; E is a threaded rod, which is constructed to fit the threaded handle, but slotted so as to engage the pin F and prevent it from turning; H is a copper or platinum rod; D is a brass tube; M. is the overflow hole in the fibre tube, into the lower end of which a portion of an old spark plug porcelain part has been driven; finally C is a small brass block as shown. This is the adjustable type of Wehnelt interrupter, since the fibre handle can be operated to allow more or less surface to be exposed to the acidulated water.

INVERTED "L" AERIAL—A flat top aerial (antenna) in which the down lead is tapped off one end of its horizontal span. (See *Length of Aerials, Horizontal Aerial.*)

INVERSE DUPLEX CIRCUIT—A circuit in which the vacuum tubes are utilized both for radio frequency and for audio frequency amplification as in the *reflex* (q.v.) circuit with this important differentiation that the first radio frequency tube is utilized as the second audio frequency tube and the second radio frequency tube is used as the first audio frequency tube. In the ordinary reflex circuit the first radio frequency tube is also the first audio tube and the second radio frequency tube is the second audio tube. The *inverse duplex* circuit has the advantage over the reflex in that the tubes which are reflexed operate more efficiently since the relatively weak radio currents pass through the same tube as the stronger audio currents and vice-versa. The *inverse duplex circuit* is the invention of David Grimes.

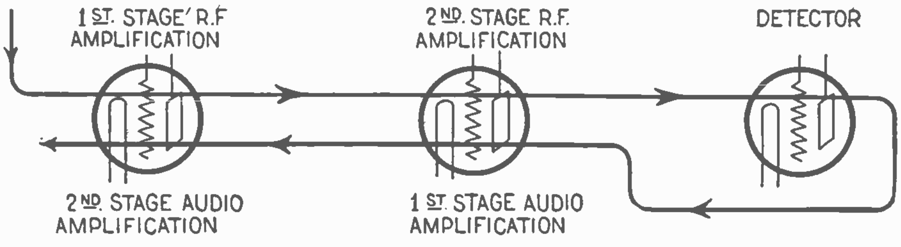
1ST. STAGE R.F AMPLIFICATION
2ND. STAGE R.F. AMPLIFICATION
DETECTOR
2ND. STAGE AUDIO AMPLIFICATION
1ST. STAGE AUDIO AMPLIFICATION

Diagram showing method of reflexing in an inverse duplex circuit.

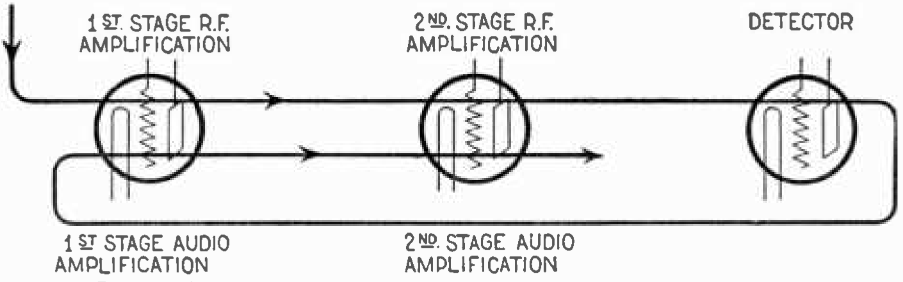
1ST. STAGE R.F. AMPLIFICATION
2ND. STAGE R.F. AMPLIFICATION
DETECTOR
1ST STAGE AUDIO AMPLIFICATION
2ND. STAGE AUDIO AMPLIFICATION

Diagram showing the current path and reflexing in an ordinary reflex circuit.

ION—An atom of matter carrying an *electron* or an *atom* deprived of electrons. *Monad-ion* refers to a unit charge, *dyad-ion, divalent* carries two units, the *triad-ion* carries three unit charges. The positive *ion* is an *atom* minus an *electron.* A negative *ion* is an *atom* plus an *electron.* (See *Anion,* also *Cathion, Electron, Electron Theory.*)

IONIZATION—sometimes spelled **IONISATION**—The splitting up of *molecules* into *ions* (q.v.). *Ionization* may apply either to a chemical compound or to a gas. When a electric current is passed through an *electrolyte* (q.v.) the constituents of the electrolyte which are liberated or deposited at the *electrodes* (q.v.) are called *ions* and this process is called *ionization.* Gases at or near atmospheric pressure are normally good insulators. They are said to undergo *ionization* when they are made conductive, that is when they allow the flow of electricity. A gas may be made conductive by passing *X-rays* (q.v.) through it, by the action of ultra-violet light, by subjecting it to radium rays, by passing it close to a flame or hot piece of metal or by forming in it a brush discharge or *corona* (q.v.). Gases which have been made conductive carry current because of their dissociation into *ions,* which carry positive or negative charges of electricity. (See *Electrolysis, Electrolytic Action,* also *Valve Detectors.*)

I.R. DROP.—The voltage drop due to the resistance in a current-carrying conductor. It is directly proportional to the current flowing in the circuit multiplied by the resistance of that circuit.

I^2R LOSS—The power loss in any circuit due to the resistance offered to current flow. It is proportional to the square of the current flowing in the circuit multiplied by the resistance of the circuit. Whenever current flows there is an I^2R loss. This is measured in *watts* (q.v.) or *kilowatts* (q.v.).

IRON PYRITES—Formula FeS_2—This is a mineral, a disulphide of iron, found in large quantities. It was formerly used in radio as a crystal detector mineral but has been almost entirely superseded by vacuum tube detectors. *Iron pyrites* is flaked with shiny spots which have the appearance of gold, and for this reason it is sometimes called "fools gold." *Iron pyrites* is very hard and brittle. The present commercial use of *iron pyrites* is in the production of sulphuric acid. Ferrous Sulphide, FeS, has also been employed in the past as a radio detector. This substance is chemically formed by fusing equivalent quantities of sulphur and iron together. The resultant substance has a smooth surface. (See *Crystal Detector.*)

ISOCHRONOUS—A term applied to two or more oscillatory circuits, meaning that they have the same *natural frequency* (q.v.). Stated in another way, two or more radio frequency circuits are isochronous when they are in electrical *resonance* (q.v.). Circuits are said to be in resonance when they have the same *oscillation constant* (q.v.). The *oscillation constant* of a circuit is equal to the square root of the product of the inductance multiplied by the capacity of that circuit. Circuits having equal oscillation constants will have the same discharge frequency.

ISODYNAMIC LINES—Lines on a magnetic map connecting all parts of the earth where the magnetic intensity is the same. In other words, these lines pass through points of equal horizontal component of the earths magnetic field.

ISOGONIC LINES—Lines on a magnetic map passing through points of equal *declination.* Lines connecting all points of equal variation of the magnetic meridian from the geographic meridian.

ISOTHERMS—Lines on a meteorological chart joining all points having the same temperature.

ISOTROPIC CONDUCTIVITY—Equal conductivity in every direction. A conductor is said to be *isotropic* when it offers the same resistance to the flow of an electric current in every direction through its mass.

J

JACK—An arrangement of spring terminals as shown in the illustrations which can be connected together or separated thus closing or opening circuits through the insertion of a *plug* which also forms a part of the circuit. The *plug* (q.v.) has a tip and a sleeve at one end (that is at the end inserted into the jack) and at the other end it has the two terminals for connecting the apparatus to be cut into the circuit. As an example, the terminals of the plug may be connected to a *loud speaker* (q.v.). The jack in this case is of the *open circuit* variety. That is to say, the springs of the jack are separated when not in use and the circuit is normally open. When the plug is inserted, the tip of the plug makes contact with one spring and the sleeve makes contact with the other thus closing the circuit through the *loud speaker*. In cases where the plug is used to open a circuit, that is to make the *jack* springs break contact with each other, the *jack* is called a *closed circuit jack* since its spring terminals are normally making contact. In cases where the jack is used not only to put the loud speaker in the circuit, but also to act as a filament switch, it is known as an *automatic filament control jack.* Jacks which are used to control a single circuit only, are referred to as *single circuit jacks.* Those controlling two or more circuits simultaneously are known as *multiple circuit jacks. Jacks* are usually mounted on a panel so that the plug can be inserted from the outside of the panel while the spring terminals are concealed behind the panel.

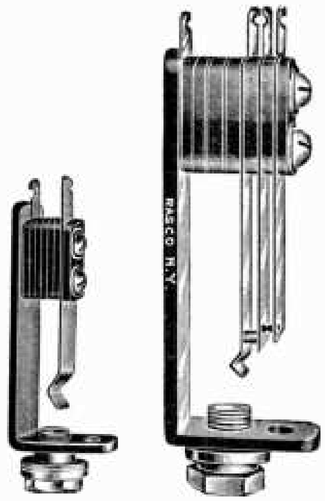

Illustration by courtesy of Radio Specialty Co.

Illustration at left shows one-spring open circuit jack. At the right is a three-spring jack.

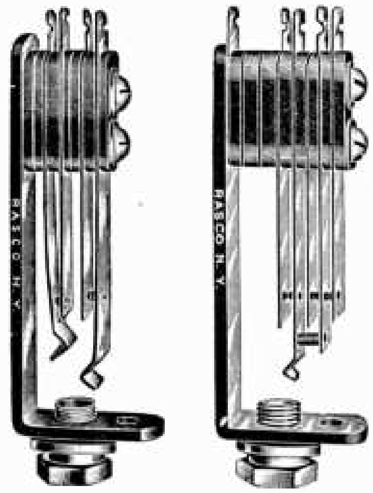

Illustration by courtesy of Radio Specialty Co.

The jack shown at the left is a four-spring double circuit jack. At the right is a five-spring automatic jack.

JAMMING—An expression denoting interference at a receiving station caused by a transmitting station not desired to be heard. More specifically, such interference, prevents the reception of signals, *jamming* the receiving station by drowning out with stronger signals or noises the station desired.

JANET, PAUL—Born in Paris, January 10, 1863. Educated at Lycee Louis-le-Grand and at the high school. He was Professor of Physics at the University of Paris and University of Grenoble. Professor Janet was the author of several important works, and first to make the experiments in electrical resonance successful. The results of these experiments are applied in present day wave meters.

JAR—A unit of electrostatic capacity. It is mainly used in the British Naval service. One *jar* equals 1000 *centimeters* which equals .0011 *microfarads* (q.v.).

JIGGER—A term generally accepted to denote an *oscillation transformer* (q.v.). Used for transforming trains of oscillations from one circuit to another. Dr. J. Erskine Murray refers to high frequency oscillations as *Jigs. Jiggers* are also known as *magnetic coupling transformers* (q.v.). They are also sometimes referred to as *auto-jiggers.*

JOULE—symbol J—The practical unit of electrical energy. It is equal to 10^7 *ergs* (q.v.). The *joule* is also the unit of electrical work. It is equal to the work done or the heat generated by a *watt second* (q.v.); or an *ampere* (q.v.) flowing for a second through a resistance of one *ohm* (q.v.). Expressed another way the *joule* is the work done by one *coulomb* (q.v.) flowing under the pressure of one *volt* (q.v.). One *joule* is equal to .73732 foot pounds, or .24 calories (q.v.).

JOULE EFFECT, or **JOULEAN HEATTING**—This occurs when an electric current traverses a resisting conductor. The rate of generation of heat in the conductor by a steady current equals the square of the current multiplied by the resistance.

JOULES EQUIVALENT—The amount of energy equal to a heat unit.

JOULES LAW—The heat produced in an electric circuit is directly proportional to the square of the current, to the resistance of the conductor and to the time of current flow.

K

KANALSTRAHLEN—A German word meaning Rays of Canal. It has been found that by perforating the cathode of a tube producing *cathode rays* (q.v.), faint luminous streaks will come through these perforations in a direction opposite to that of the *cathode* rays. These are called *kanalstrahlen.* They communicate a positive charge to an insulated conductor. In terms of the *electron theory* (q.v.) they are positively charged *ions* (q.v.) The cathode particle is an electron traveling in one direction, and the kanalstrahlen particle traveling in the opposite direction is what remains of the atom which has lost an electron.

KATHODE—This is a negative *electrode* (q.v.). (See *Cathode,* also *Electrolysis.*)

KATHODE OF CELL—The positive pole of a cell. This is indicated by the plus (+) sign. (See *cell.*)

KATHODE RAYS—The stream of *electrons* (q.v.) thrown off by the *kathode* of a vacuum tube. These produce a glow when they strike the walls forming the tube. They can be deflected by a magnet. (See *Cathode Rays.*)

KATION—The charged particles or *ions* (q.v.) moving in the direction of the *cathode* (q.v.). (See *Cathion.*)

KEEPER—The piece of iron used to close a magnetic circuit to protect it from external disturbances. The term *keeper* is generally used to refer to such a piece of iron when used with a permanent magnet such as a horseshoe magnet. In the case of the piece of iron which closes the magnetic circuit of an electromagnet, this is more often referred to as an *armature* (q.v.). The *keeper* may be used with a straight bar magnet also. In this case, two bar magnets are laid side by side with unlike poles adjacent and two keepers are used, one at each end. The use of a keeper avoids the demagnetizing effect of *leakage lines* (q.v.). (See *Horseshoe Magnet.*)

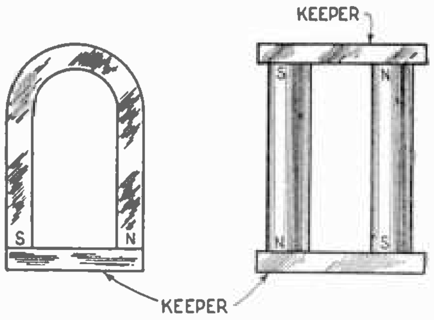

At left is shown keeper used with a horseshoe magnet. At right two keepers are shown used with two bar magnets.

KELVIN, LORD (William Thomson) was born at Belfast in 1824; he entered Glasgow University as a student at 10 years of age; graduated at Cambridge in 1845. When only 22 years of age he was called to occupy the chair of natural philosophy in the University of Glasgow, a chair which he adorned by his genius for fifty years. He was for 40 years or more regarded as the acknowledged leader of British science on its physical and mathematical side. His great inventions in telegraphy, his magnetic compass and sea sounding apparatus brought him fame and fortune. He was knighted in 1866; and in 1892 was raised to the peerage as Baron Kelvin of Largs; he died on December 17, 1907.

Lord Kelvin (William Thomson)

KELVIN'S BRIDGE—A method of measuring low resistance in which the voltage drops, produced by the same current in the resistance under test and in a standard low resistance slide wire are balanced against each other. This *bridge* is a modification of the *Wheatstone Bridge* (q.v.) and was designed by Lord Kelvin to eliminate the errors introduced when measuring resistances much less than one *ohm* (q.v.). In the case of the *Wheatstone Bridge,* such resistances could not be measured accurately on account of the errors produced by the terminal and contact resistances.

KELVIN'S ELECTROSTATIC VOLTMETER—A voltmeter used for measuring high, and in some cases low, alternating current voltages. It is constructed on the principle of an air condenser. One type of electrostatic voltmeter is a high potential instrument having the needle made of a thin aluminum plate suspended vertically on delicate knife edges, with a pointer extending from the upper part to a scale. Two quadrant plates, metallically connected together, are placed on either side of the needle and parallel to its face. These serve as one terminal of the circuit to be measured. The needle acts as the opposite terminal. When there is a difference of potential between the needle and the plates, the needle is deflected out of its neutral position. The value of the scale indications can be changed by hanging calibrated weights on the bottom of the needle.

KENOTRON—A type of *vacuum tube rectifier* (q.v.) in which the vacuum is extremely high and the discharge is carried almost entirely by electrons, not by gas ions. In these tubes, the plate current is always less than that actually emitted by the hot filament. *Kenotrons* are made sufficiently large to rectify several kilowatts.

KENOTRON
MODEL U.V.
PATENTED
Licensed for Amateur
Illustration by courtesy of Radio Corp. of America
A kenotron tube.

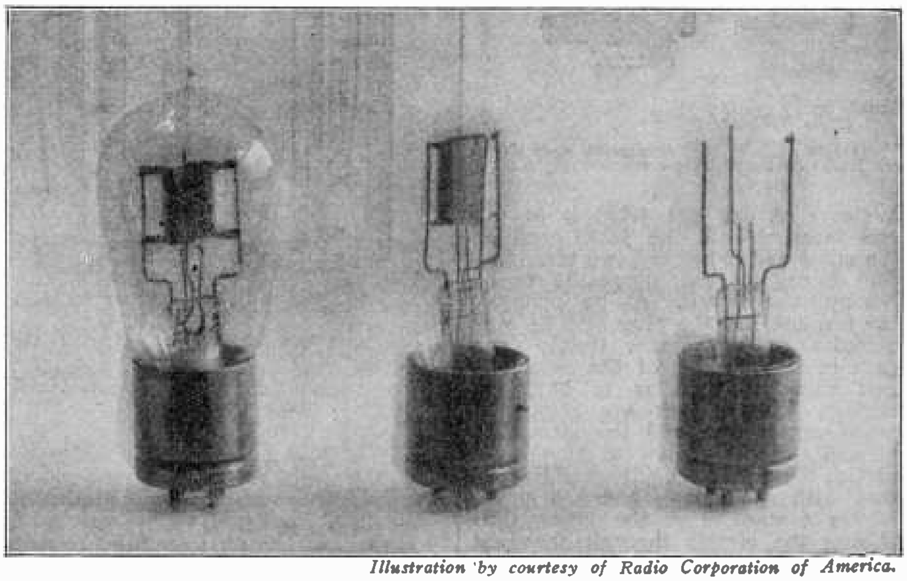
Illustration by courtesy of Radio Corporation of America.
Stages in the construction of a kenotron.

KENNELLY, A. E.—Anglo-American radio expert. Born at Colaba, Bombay, December 17, 1861, he was educated in England, Belgium, France and Italy. In 1875 he became a telegraph operator in the employ of the Eastern Telegraph Company, and in 1886 became the principal electrical assistant to Thomas Alva Edison in the laboratories at Orange, N. J., a post he held until 1892.

He was engineer-in-chief with E. J. Houston, of the Thomson-Houston Company, for the laying of the cables from Vera Cruz to Campeche, 1902. Since 1902 he has been Professor of Electrical Engineering at Harvard University, and since 1914 professor of the same subject at Massachusetts Institute of Technology.

Kennelly has written a large number of books on electricity and radio, and is the author of one of the standard elementary text-books on same. He has written a very large number of papers on radio, and he is an authority on alternating currents.

He is past president of the American Institute of Electrical Engineers, was president in 1916 of the Institute of Radio Engineers, and vice-president of the International Electrical Congresses, held in Paris and Turin. He is a member of many scientific societies, and has received many honorary degrees. In 1921 he was appointed a delegate to the Interallied Radio Technical Committee in Paris.

KEY—A form of switch for conveniently and quickly opening and closing a transmitting circuit, in the act of transmitting signals. An *operating,* or *telegraph,* so-called *Morse key* is a form of *tapping key* designed especially for the rapid sending of signals by the dot and dash code. A tapping key has one contact carried by a springy strip of metal. In signalling, this key is depressed by the fingers, thus closing the circuit by bringing the contact together with a fixed contact. A standard type of *transmitting key* (q.v.) such as used in sending wireless messages is shown in the illustration. These keys must necessarily be made with larger contacts and of more solid construction than those used to interrupt the smaller currents used in wire telegraphy.

A standard transmitting key.

KEY, ANTI-CAPACITY—This is in reality a form of switch. (See *Switch.*)

KEY, HIGH-SPEED—A transmitting key usually operated from side to side instead of downwards, which makes

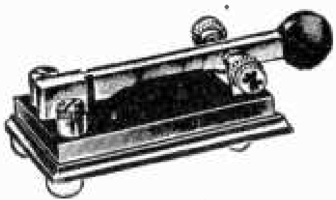
High speed key.

contact both on the right and on the left side. Such a key requires only half the travel of the ordinary transmitting key and thus facilitates high-speed manual transmission of code signals.

KEY SWITCH—A type of switch adapted for operation by a removable portion in the form of a handle or key. In radio work, *key switches* are sometimes used to lock a receiving set so that it cannot be operated unless the filament circuit is closed by the insertion and turning of the key. Key switches are generally operated by a hollow slotted key, although there are other types on the market. The slot in the key fits a projection or pin on the spindle or shaft of the switch which carries the movable contact. When the key is inserted in the socket and turned, it turns the shaft and brings the movable contact against a fixed contact. *Key switches* are also used for turning on or off electric lights and for locking or unlocking the ignition systems of automobiles.

KILO—Symbol K—A prefix placed before the name of a unit to indicate the multiple 1000 or 10^3. In cases where standard units are too small for convenient use, it is customary to utilize this prefix as, for example, in stating the voltage of a high tension transmission line. This would often be given as 10 kilovolts (K.V.) rather than as 10,000 volts. In like manner, in referring to large amounts of power, it would be preferable to give the amounts in kilowatts rather than in watts. Thus the better way to express 10,000,000 watts would be to divide by 1,000, thus converting to 10,000 kilowatts.

KILOAMPERE-TURN—One kiloampere-turn is equivalent to 1,000 ampere-turns. In referring to the magnetic circuits of electrical machinery it is usual to specify the magnetomotive-forces in kiloampere-turns rather than in ampere-turns. (See *ampere-turns*, also *magnetomotive-force.*)

KILOCYCLE—1,000 *cycles*—Two immediately succeeding half waves of an alternating current constitute a *cycle* (q.v.). The number of *cycles* per second is called the *frequency* (q.v.). In house-lighting and similar circuits, where a comparatively low frequency is used, it is usual to refer to the *frequency* in cycles per second. However, in radio circuits, where currents as high as 300,000,000 cycles per second are used, it is preferable to divide by 1,000, thus converting to *kilocycles*. If the wave-length in meters is known, it is a simple matter to determine the frequency in *kilocycles*. This is calculated as follows: The speed of light is 186,000 miles per second and experiments have shown that this is practically the same as the velocity of electromagnetic or radio waves. Expressed in meters this is equal to 300,000,000 (appx.) meters per second. Assuming that a transmitting station is using a wave-length of 500 meters, the frequency will be obtained in cycles per second by dividing the velocity of the waves by the

KILOCYCLES TO METERS, OR METERS TO KILOCYCLES

Meters	Kilocycles	Meters	Kilocycles	Meters	Kilocycles	Meters	Kilocycles	Meters	Kilocycles	Meters	Kilocycles	Meters	Kilocycles
10	29980	720	416.4	1430	209.7	2140	140.1	2850	105.2	4120	72.77	6400	46.85
20	14990	730	410.7	1440	208.2	2150	139.5	2860	104.8	4140	72.42	6450	46.48
30	9994	740	405.2	1450	206.8	2160	138.8	2870	104.5	4160	72.07	6500	46.13
40	7496	750	399.8	1460	205.4	2170	138.1	2880	104.1	4180	71.73	6550	45.77
50	5996	760	394.5	1470	204.0	2180	137.5	2890	103.7	4200	71.39	6600	45.43
60	4997	770	389.4	1480	202.6	2190	136.9	2900	103.4	4220	71.05	6650	45.09
70	4283	780	384.4	1490	201.2	2200	136.3	2910	103.0	4240	70.71	6700	44.75
80	3748	790	379.5	1500	199.9	2210	135.7	2920	102.7	4260	70.38	6750	44.42
90	3331	800	374.8	1510	198.6	2220	135.1	2930	102.3	4280	70.05	6800	55.09
100	2998	810	370.2	1520	197.2	2230	134.4	2940	102.0	4300	69.73	6850	43.77
110	2726	820	365.6	1530	196.0	2240	133.8	2950	101.6	4320	69.40	6900	43.45
120	2499	830	361.2	1540	194.7	2250	133.3	2960	101.3	4340	69.08	6950	43.14
130	2306	840	356.9	1550	193.4	2260	132.7	2970	100.9	4360	68.77	7000	42.83
140	2142	850	362.7	1560	192.2	2270	132.1	2980	100.6	4380	68.45	7050	42.53
150	1999	860	348.6	1570	191.0	2280	131.5	2990	100.3	4400	68.14	7100	42.23
160	1874	870	344.6	1580	189.9	2290	130.9	3000	99.94	4420	67.83	7150	41.93
170	1764	880	340.7	1590	188.6	2300	130.4	3020	99.28	4440	67.53	7200	41.64
180	1666	890	336.9	1600	187.4	2310	129.8	3040	98.62	4460	67.22	7250	41.35
190	1578	900	333.1	1610	186.2	2320	129.2	3060	97.98	4480	66.91	7300	41.07
200	1499	910	329.5	1620	185.1	2330	128.7	3080	97.34	4500	66.63	7350	40.79
210	1428	920	325.9	1630	183.9	2340	128.1	3100	96.72	4520	66.33	7400	40.52
220	1363	930	322.4	1640	182.8	2350	127.6	3120	96.10	4540	66.04	7450	40.24
230	1304	940	319.0	1650	181.7	2360	127.0	3140	95.48	4560	65.75	7500	39.98
240	1249	950	315.6	1660	180.6	2370	126.5	3160	94.88	4580	65.46	7550	39.71
250	1199	960	312.3	1670	179.5	2380	126.0	3180	94.28	4600	65.18	7600	39.45
260	1153	970	309.1	1680	178.5	2390	125.4	3200	93.69	4620	64.90	7650	39.19
270	1110	980	305.9	1690	177.4	2400	124.9	3220	93.11	4640	64.62	7700	38.94
280	1071	990	302.8	1700	176.4	2410	124.4	3240	92.54	4660	64.34	7750	38.69
290	1034	1000	299.8	1710	175.3	2420	123.9	3260	91.97	4680	64.06	7800	38.44
300	999.4	1010	296.9	1720	174.3	2430	123.4	3280	91.41	4700	63.79	7850	38.19
310	967.2	1020	293.9	1730	173.3	2440	122.9	3300	90.86	4720	63.52	7900	37.95
320	936.9	1030	291.1	1740	172.3	2450	122.4	3320	90.31	4740	63.25	7950	37.71
330	908.6	1040	288.3	1750	171.3	2460	121.9	3340	89.77	4760	62.99	8000	37.48
340	881.8	1050	285.5	1760	170.4	2470	121.4	3360	89.23	4780	62.72	8050	37.25
350	856.6	1060	282.8	1770	169.4	2480	120.9	3380	88.70	4800	62.46	8100	37.02
360	832.8	1070	280.2	1780	168.4	2490	120.4	3400	88 18	4820	62.20	8150	36.79
370	810.3	1080	277.6	1790	167.5	2500	119.9	3420	87.67	4840	61.95	8200	36.56
380	789.0	1090	275.1	1800	166.6	2510	119.5	3440	87.16	4860	61.69	8250	36.34
390	768.8	1100	272.6	1810	165.6	2520	119.0	3460	86.65	4880	61.44	8300	36.12
400	749.6	1110	270.1	1820	164.7	2530	118.5	3480	86.16	4900	61.19	8350	35.91
410	731.3	1120	267.7	1830	163.8	2540	118.0	3500	85.66	4920	60.94	8400	35.69
420	713.9	1130	265.3	1840	162.9	2550	117.6	3520	85.18	4940	60.69	8450	35.48
430	697.3	1140	263.0	1850	162.1	2560	117.1	3540	84.70	4960	60.45	8500	35.27
440	681.4	1150	260.7	1860	161.2	2570	116.7	3560	84.22	4980	60.20	8550	35.07
450	666.3	1160	258.5	1870	160.3	2580	116.2	3580	83.75	5000	59.96	8600	34.86
460	651.8	1170	256.3	1880	159.5	2590	115.8	3600	83.28	5050	59.37	8650	34.66
470	637.9	1180	254.1	1890	158.6	2600	115.3	3620	82.82	5100	58.79	8700	34.46
480	624.6	1190	252.0	1900	157.8	2610	114.9	3640	82.37	5150	59.22	8750	34.27
490	611.9	1200	249.9	1910	157.0	2620	114.4	3660	81.92	5200	57.66	8800	34.07
500	599.6	1210	247.8	1920	156.2	2630	114.0	3680	81.47	5250	47.11	8850	33.88
510	587.9	1220	245.8	1930	155.3	2640	113.6	3700	81.03	5300	56.75	8900	33.69
520	576.6	1230	243.8	1940	154.5	2650	113.1	3720	80.60	5350	57.11	8950	33.50
530	565.7	1240	241.8	1950	153.8	2660	112.7	3740	80.17	5400	55.52	9000	33.31
540	555.2	1250	239.9	1960	153.0	2670	112.3	3760	79.74	5450	55.01	9050	33.13
550	545.1	1260	238.0	1970	152.2	2680	111.9	3780	79.32	5500	54.51	9100	32.95
560	535.4	1270	236.1	1980	151.4	2690	111.5	3800	78.90	5550	54.02	9150	32.77
570	526.0	1280	234.2	1990	150.7	2700	111.0	3820	78.49	5600	53.54	9200	32.59
580	516.9	1290	232.4	2000	149.9	2710	110.6	3840	78.08	5650	53.07	9250	32.41
590	508.2	1300	230.6	2010	149.2	2720	110.2	3860	77.67	5700	52.60	9300	32.24
600	499.7	1310	228.9	2020	148.4	2730	109.8	3880	77.27	5750	52.14	9350	32.07
610	491.5	1320	227.1	2030	147.7	2740	109.4	3900	76.88	5800	51.69	9400	31.90
620	483.6	1330	225.4	2040	147.0	2750	109.0	3920	76.49	5850	51.25	9450	31.73
630	475.9	1340	223.7	2050	146.3	2760	108.6	3940	76.10	5900	50.82	9500	31.56
640	468.5	1350	222.1	2060	145.5	2770	108.2	3960	75.71	5950	50.39	9550	31.39
650	461.3	1360	220.4	2070	144.8	2780	107.8	3980	75.33	6000	49.97	9600	31.23
660	454.3	1370	218.8	2080	144.1	2790	107.5	4000	74.96	6050	49.56	9650	31.07
670	447.5	1380	217.3	2090	143.5	2800	107.1	4020	74.58	6100	49.15	9700	30.91
680	440.9	1390	215.7	2100	142.3	2810	106.7	4040	74.21	6150	48.75	9750	30.75
690	434.5	1400	214.2	2110	142.1	2820	106.3	4060	73.85	6200	48.36	9800	30.59
700	428.3	1410	212.6	2120	141.4	2830	105.9	4080	73.49	6250	47.97	9850	30.44
710	422.3	1420	211.1	2130	140.8	2840	105.6	4100	73.13	6300	47.59	9900	30.28
										6350	47.22	9950	30.13

length of a single wave, which in this case would be equal to 300,000,000 (appx.) divided by 500, equalling 600,000 cycles per second or (reduced to kilocycles by dividing by 1,000) 600 kilocycles. On the preceding page is given a chart for converting kilocycles to meters, or meters to kilocycles.

KILOLINE—This is equal to 1,000 *lines of force*, or 1,000 *maxwells* (q.v.). The *maxwell* or the *line of force* is the unit used to measure *magnetic flux*. (See *Line of Force, Maxwell, Magnetic Flux.*)

KILOVOLT — abbreviated K.V. — The kilovolt is equal to 1,000 volts.

KILOVOLT-AMPERE — Abbreviated K.V.A. or kv-a.—The product of the effective volts across the terminals of a circuit by the effective amperes flowing in that circuit, divided by 1,000. Alternators are rated in *kilovolt-amperes* because their load determines the power factor. Thus with a zero *power factor* (q.v.) the *true power* (q.v.) in kilowatts would equal zero even if the alternator were delivering maximum amperes and maximum volts. At 50% power factor the kilowatt capacity would be half that at 100% power factor. On the other hand the kilovolt-ampere capacity remains the same no matter what the power factor.

KILOWATT — abbreviation K.W.—The kilowatt is a unit of electrical power having the value of 1,000 watts. One kilowatt equals 1,000 watts. To reduce the power expressed in kilowatts to watts, it is necessary to multiply the number of kilowatts by 1,000. Most electrical formulas use watts rather than kilowatts. Therefore, in solving such formulas, where the power is given in kilowatts, it is necessary to multiply this by 1,000 before using it in the formula. For example, suppose the power expended in an electrical circuit carrying 100 amperes is 20 kilowatts. The voltage will equal the power in watts divided by the current in amperes. This is $\frac{20 \times 1,000}{100} = 200$ volts. Motors are often rated in kilowatts instead of horsepower. For practical calculations, the horsepower rating of a motor is equivalent to four-thirds of its kilowatt rating. (See *watts.*)

KILOWATT-HOUR—The energy expended when work is done for one hour at the rate of one kilowatt. It is possible to calculate the energy expended in an electrical circuit over a certain period of time, knowing the current in amperes and the resistance of the circuit. Using the following formula the energy expended will be obtained in kilowatt-hours:

$$\frac{\text{Work performed or energy expended} = \text{Current}^2 \times \text{Resistance}^* \times \text{Time}^\dagger}{3,600,000}$$

$= \text{kilowatt-hours.}$

KIMURA SHUNKICHI—Japanese radio authority. Kimura was born in 1866, and educated at the Scientific College of the Tokyo Imperial University, his special study being physics, which he continued to study at Harvard and Yale Universities, United States.

In 1901 Kimura entered the Japanese Navy and first began to study wireless telegraphy, particularly for naval use. In 1906 he was appointed the Japanese delegate to the International Wireless Telegraph Conference, Berlin, and after his retirement from the Navy, 1912, he became the director of the Nippon Radio Telegraph and Telephone Company. Kimura has written largely on radio subjects for various scientific journals.

KINETICS—That branch of dynamics dealing with forces that produce or change motion. Energy is usually classified either as *potential* or *kinetic.* When energy is available for the production of work it is referred to as potential energy. Energy at work is *kinetic energy* (q.v.). The term *kinetic* refers to motion.

KINETIC ENERGY—The work a body is able to do by reason of its motion. The unit of energy is the *erg.* This unit is also used to measure work. (See *Erg.*) It should be noted that *kinetic energy* is stored in the magnetic field. It requires more energy to start a current in a circuit having great inductance than in one having small inductance. This greater amount of energy is stored in the magnetic field and is returned to the circuit when the current decreases to zero.

KINRAIDY SPARK GAP—A form of *quenched spark gap* (q.v.) consisting of two water-cooled flat terminals close together and between which discharge occurs.

KIRCHOFF'S LAWS—These are two in number. First Law: The sum of the currents flowing to a point, in any electrical circuit, is equal to the sum of the currents flowing away from that point. Second Law: In any closed electrical circuit, the sum of the impressed *electromotive forces* (q.v.) will equal the sum of the *voltage drops* (q.v.). This statement requires modification, in so far as "addition" of voltages is concerned. Voltages are added provided that they are in the same direction, but must be subtracted if in opposite directions. These two laws depend on *Ohm's Law* (q.v.) and constitute a further application of that law to more complicated electrical circuits. In addition to simple electrical circuits, conductors may be connected in various complicated networks, all of which come under the heading of *divided circuits*. By means of *Kirchoff's Laws*, the current in any part of a divided circuit may be found, if the resistances of the various parts, and the electromotive forces are given.

KITES AS AERIAL SUPPORTS—During the early experiments with wireless, kites were successfully employed by Marconi as aerial supports.

KLEIN, RENE HENRI—Anglo-French radio expert. Born at Soults-sous-Foret, France, in 1880, he first became interested in wireless telegraphy in 1908. He was the founder and first honorary secretary of the Wireless Society of London, 1913-20, and became a vice-president of the society. He was one of the inventors with H. L. McMichael of the synthetic galena crystal Radiocite, and has written many articles on wireless subjects.

KNIFE SWITCH—A switch in which the movable arm wedges in between two parallel spring clips. Knife switches may be either of the *single pole* (s.p.) or *double pole* (d.p.) types and also may either be *single throw* (s.t.) or *double throw* (d.t.). Knife switches are often referred to as *Lever* switches They are used on direct current circuits up to 250 volts and on alternating current circuits up to 500 volts (See *Ground Switch; Double Throw Switch.*)

KLEMENCIC THERMAL JUNCTION—A device consisting of two thin sheets of brass, designated in the illustration as A, A. These are 4 inches by 11.8 inches and are placed about 1 inch apart. Soldered to one plate is a fine platinum wire and to the other plate is a fine platinum-nickel wire. These two wires are connected together at J to form a *thermal junction* (q.v.) and are then led off at right angles and are soldered at their other ends to the leads R, R, of a sensitive *galvanometer* (q.v.). This resonance system is fixed at the focal line of a suitable cylindrical metallic reflector. When electric waves, with the electric force parallel to A A fall on this receiver, electric oscillations between A and A produce heating at the junction J which is the point of contact between the two dissimilar metals. In consequence, the heat developed gives rise to a thermoelectromotive force at the junction and consequently to a current in the galvanometer. This instrument can be used to detect and measure electric waves.

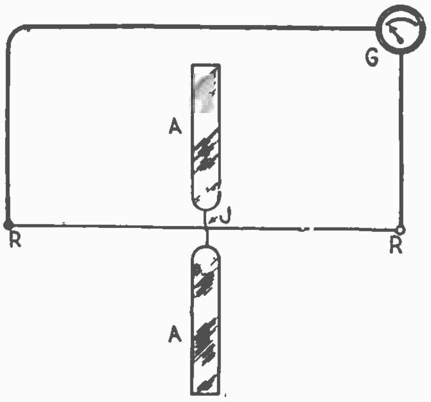

Schematic diagram showing Klemencic thermal junction.

KOEPSEL PERMEAMETER—An instrument used for measuring the *permeability* (q.v.) of an iron or steel bar. In its principle of operation, this device is a movable coil milli-voltmeter in which the permanent magnet is replaced by an electromagnet. The sample to be tested forms a part of this electromagnet. The flux produced in the sample and in the pole pieces of the instrument is measured by a movable coil similar to those used in direct current voltmeters. The coil is supplied with energizing current from a small dry battery. Deflections are indicated by a pointer on a dial, the scale being calibrated directly in magnetic densities.

KOHLRAUSCH'S LAW—This law states that the rate of motion of each atom in a solution undergoing electrolysis for a given liquid is independent of the element with which it may have been in combination.

KOLSTER, FREDERICK A.—American radio authority. Born at Geneva, Switzerland, January 13, 1883, he was educated at Cambridge, Mass., and Harvard University. From 1902-08 he was assistant to J. S. Stone, the radio expert, and from 1909-12 to Dr. Lee de Forest. He was appointed chief of the Radio Section of the Bureau of Standards, a post he held until 1921, when he became research engineer to the Federal Telegraph Company.

Kolster has written a large number of articles on radio subjects, including those on the "Effects of Distributed Capacity in Coils," "Reinforced Harmonics in Radio Transmission," etc. He is the inventor of the Kolster decremeter, of a radio compass, of directional radio systems, etc., and is a member of the American Institute of Radio-Engineers.

---

*Ohms. †Seconds.

KOOMANS, NICHOLAS—Dutch radio authority. Born at Delft, 1879, he was educated there as a mechanical and electrical engineer. In 1908 he received his doctorate for his dissertation "Regarding the Influence of Self-Induction in Telephone Conducting Wires." He was one of the founders of and editor of the "Monthly Review of Telephony and Telegraphy," and was also one of the founders of the Dutch Society for Radio Telegraphy. Koomans is a member of the International Electrotechnical Commission and professor in physics and theoretical electrical engineering at the school of the Dutch Post and Telegraph Administration.

KORDA AIR CONDENSER—A variable *condenser* (q.v.) using air as the *dielectric* (q.v.) consisting of two sets of semi-circular plates, one set being connected together and fixed in position and the other set being also connected together but capable of being rotated about a central axis. Rotating the movable plates brings a greater or less area of the two sets of plates into interlapping positions, thus permitting the capacity of the condenser to be varied. This condenser was patented by Korda in Germany in 1893.

KORN, ARTHUR—German radio expert. Born at Breslau, Germany, 1870, Korn was educated at Leipzig and Paris and became professor of physics at the University of Munich, 1903-8, and afterwards professor at the Polytechnical High School, Charlottenburg, Berlin.

Korn is well known for his experiments on and the invention of a method of transmitting photographs by telegraphy, the first photograph being telegraphed from Munich to Berlin in 1907 by his methods. He is the inventor of a system of telautography and wireless phototelegraphy, and the author of a number of standard books on electricity.

L

L—The symbol of *inductance* (q.v.) or the *coefficient* of *self-induction.* In general, the co-efficient *L* by which the rate of change of current in any circuit must be multiplied to obtain the value of the self-induced *electromotive force* (q.v.) is called the *coefficient* of *self-induction* or simply the *inductance.* (See *Self-induction, coefficient.*)

LAG—One *alternating current* quantity reaching its maximum at a later time than a second alternating current quantity. As an example,—the current is said to *lag* behind the impressed voltage when maximum current occurs after maximum voltage. In other words, the current and voltage in such an alternating current circuit are out of *phase* (q.v.). In the above example, since the current *lags* behind the voltage, the voltage is said to lead the current. *Inductance* (q.v.) in an alternating current circuit will cause the current to *lag* behind the impressed voltage. *Capacity* (q.v.) in such a circuit will cause the current to *lead* (q.v.) the impressed voltage. The word *lag* may be applied to magnetic quantities as well as electrical quantities. Thus, one magnetic flux may *lag* behind another as in the case of an induction motor where the induced *magnetic flux* (q.v.) *lags* behind the impressed magnetic flux. Another type of magnetic *lag* is the retardation of magnetic effects behind their causes, due to *hysteresis* (q.v.) in the magnetized substance. (See *Lead,* also *Leading Current.*)

LAMINATED—Composed of a number of thin plates, one on top of another. Examples of laminated construction are *laminated brushes, laminated cores laminated conductors,* etc.

LAMINATED CORE—An arrangement of soft iron plates or stampings forming the core of a transformer, or of an armature of a motor or generator. Cores are laminated in this way, principally to reduce *eddy current* (q.v.) losses. In the case of an armature core, this must be laminated parallel to the direction of rotation (in other words, perpendicular to the axis of rotation). Laminations are insulated from each other by means of japan, varnish, shellac or simply rust.

LAMP BANK—A bank of incandescent lamps arranged so that they can be connected in series or parallel. *Lamp banks* are used in laboratories in conjunction with electrical machinery under test. In this case, the lamps act as the *load,* the more lamps being connected in *parallel,* the greater the amount of current being drawn. *Lamp banks* are sometimes connected in *series* in a circuit for the purpose of reducing voltage. An example in this connection is the use of a lamp or bank of lamps in charging a 6-volt storage battery from a 110-volt direct current source. Lamp banks are sometimes called *lamp panels* or *lamp batteries.*

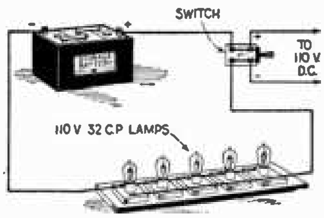

Use of lamp bank in charging storage battery.

LANGMUIR, IRVING—American radio authority. Born at Brooklyn, New York, January 31st, 1881, he was educated at the School of Mines, Columbia University, and was for some time assistant to Professor Nernst at the University of Gottingen. In 1906-9 he was instructor in chemistry at the Stevens Institute of Technology, and in the latter year he became a research assistant to the General Electric Company, at Schenectady. He has carried out a series of brilliant researches on apparatus used in wireless telegraphy and telephony, on electron discharge apparatus, atomic and molecular structure, etc.

Langmuir has paid particular attention to high-vacuum valves, and is the inventor of the Langmuir valve. To him is due the discovery that by treating the tungstic oxide used in the construction of tungsten filaments with certain compounds of thorium the filament becomes thoriated tungsten and the electronic emission is enormously increased. These filaments are the ones used in dull emitter valves, the electrons being given off at comparatively low temperatures Langmuir is the author of many papers to scientific and technical journals, including those on pure electron discharge, thermionic currents in high vacua, etc. He is also well known for his mercury condensation pump for high vacua.

LAP WINDING—An armature winding in which the opposite ends of each coil are connected to adjoining commutator segments. As a result the windings lap back and upon themselves forming loops. *Lap winding* is also called *loop winding, parallel winding* and *multiple winding.* In the elementary form of *lap winding* the method of connection to the commutator bars is as follows: One side of a coil is connected to a commutator bar, and the other side of the coil, which is located 180 electrical degrees (the distance from the center of a north pole to the center of the adjacent south pole) away, has its end connection brought back and soldered to the next adjacent commutator segment. This method of overlapping the ends is continued all the way around the armature until all the slots have been filled and the circuit has been closed on itself. The coils must all be symmetrical. This means that if there is a forward throw of a certain num-

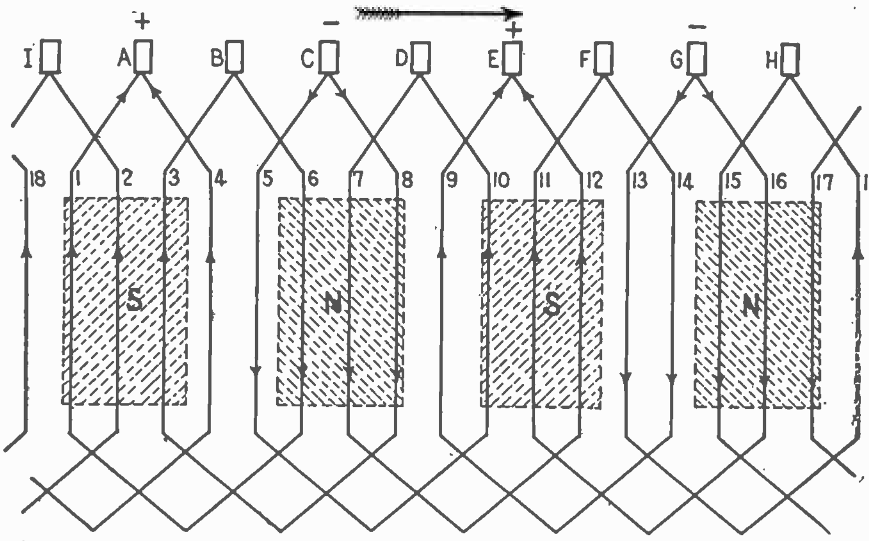

A typical lap winding shown in developed view. In the winding shown each inductor is connected at the rear of the armature to one five slots away. Thus #3 is connected to 8, 5 to 10, etc. After having made one complete "element"—as, for example, A-1-6-B—the winding forms a second "element," B-3-8-C, which laps over the first. This is continued around the armature until the winding closes on itself.

ber of slots on the first coil, all the other coils must have the same forward throw. In like manner all the return throws must be the same. In the *lap wound* generator there are as many paths for the current as there are poles. Hence the name *parallel* or *multiple winding*. This necessitates the use of as many sets of brushes as there are poles. An increased number of paths in parallel means increased current capacity. In the case of a motor, the greater the number of armature paths between brushes, the greater will be the speed, other conditions remaining equal.

LAW OF ELECTROMAGNETIC INDUCTION—Also known as the "*Law of Generator Action*" and as "*Faraday's Law.*" If there is relative motion between an electrical conductor and a magnetic field, such that the conductor cuts *flux*, an *electromotive force* (q.v.) will be induced in the conductor. The amount of the induced electromotive force will depend upon the strength of the magnetic field, upon the speed of cutting, upon the length of the conductor cutting the field and upon the direction of motion between the conductor and the field. Thus the stronger the magnetic field, the greater the induced electromotive force. The induced electromotive force will also increase with increased speed of cutting and with increased conductor length. When the conductor cuts the flux at right angles, the induced electromotive force is maximum. If the flux is cut in any other direction, the induced electromotive force will be less. The induced electromotive force decreases from maximum to zero as the angle of cutting decreases from 90 degrees to zero degrees. If a conductor is of such length and moves in such a direction, and at such a speed, as to cut 100,000,000 lines of force in one second, one *volt* (q.v.) will be induced in it. Expressed as an equation the induced electromotive force (in volts) will equal the flux cut (in *lines* or *maxwells*) divided by the product of the time (in seconds) times 100,000,000. This equation refers to the average value of the electromotive force induced in a single conductor. If a number of conductors are used the same electromotive force will be induced in each. If the conductors are so connected that their electromotive forces add up, the total average electromotive force can be found by multiplying the above equation by the number of conductors.

LATOUR, MARIUS—French radio expert. Born in October, 1875, he was educated at the University of Paris and the Ecole Superieure d'Eléctricité, Paris. He was for many years consulting engineer to the General Electric Company, and he specialized in the construction of high-frequency machines. Latour is the designer of the so-called S. F. R. alternator, in which there are a smaller number of stator slots than usual.

During the World War Latour was engaged in research work under General Ferrie, and he invented a system of elimination of the interference produced in telephone lines by neighboring high-tension power lines which has been installed throughout the greater part of France. Latour is the inventor of the now widely used system of high-frequency multiplex telegraphy and telephony, using the three-electrode valve for generation and reception. He is a member of many scientific societies, including the Institute of Radio Engineers, and of the American Institute of Electrical Engineers. He has written many articles on radio for scientific and technical journals, and is the inventor of a great many important radio developments covering practically the whole range of the art.

LEAD—Pronounced LEED—One alternating current quantity reaching its maximum at an earlier time than a second alternating current quantity. Thus in an alternating current circuit, the current is said to *lead* the voltage if maximum current occurs before maximum voltage. If the current reaches its maximum 90 *electrical degrees* ahead of the voltage, current and voltage are said to be in *quadrature* (q.v.) with current *leading* and voltage *lagging*. (See *Degrees, Electrical*, also *Leading Current* and *Lag.*)

LEAD SULPHATE—Formula $PbSO_4$—A white insoluble salt which is found in native form in the ground and which can be formed artificially by adding sulphuric acid to a soluble lead salt. The active elements of a lead *storage battery* (q.v.) consists of (PbO_2) *lead peroxide* on the positive plate and (Pb) *spongy lead* on the negative plate, with dilute *sulphuric acid* (H_2SO_4) as the *electrolyte* (q.v.). The final result, on discharge of such a battery is the formation of *lead sulphate* on both the positive and negative plates. This is a part of the process of producing current. After a normal discharge, the sulphate is finely crystalline and of such nature that it is easily reduced by the current flowing through the battery on charge. If charging is neglected and the battery is allowed to stand in a discharged state, the sulphate gradually ceases to be crystalline. It fills the pores of the plates and finally makes the active material hard and dense, thus lengthening the time required for a charge. A battery in this condition is said to be *sulphated*. If sulphating proceeds too far, it will result in *buckling* of *plates* (q.v.), loss of active material, greatly decreased capacity, lower efficiency and an increased internal resistance. A moderately sulphated battery can be restored by means of a prolonged charging at a moderate rate. In a badly sulphated battery, a number of successive charges and discharges may be necessary.

LEAD-IN—The wire leading from the aerial to the radio apparatus. *Lead-in* wire, as used for receiving sets, is usually rubber covered wire although bare antenna wire can be used provided it is properly insulated from the walls of the building. A poorly insulated lead-in will greatly reduce the efficiency of the finest antenna. (See *Down Lead.*)

LEAD-IN INSULATOR—Any form of insulator used for passing aerial lead-in through window, wall or roof. These insulators are usually made of porcelain, electrose or insulated copper strips. (See *Insulator, lead-in*, also *Window Lead-in.*)

LEADING CURRENT—An alternating current which reaches its maximum value at an earlier time than the voltage. When an inductance is inserted in an alternating current circuit, the self-induced *electromotive force* (q.v.) will combine with the *impressed* electromotive force and the resultant of the two will be the active electromotive force which causes the current flow. The current will always be exactly in *phase* (q.v.) with and proportional to the resultant electromotive force. The induced electromotive force is 90 degrees behind the current and therefore behind the resultant electromotive force, which is in phase with the current. The impressed electromotive force will lead the current and will be higher than the resultant electromotive force by an amount which at each instant is equal to the counter-electromotive force of the inductance. When a condenser is inserted in an alternating current circuit the current will charge the condenser. As the amount of current stored in the condenser increases, the electromotive force of the condenser increases also until the impressed and the condenser electromotive forces are equal. The condenser electromotive force, being a counter pressure, the current flow will cease when the two electromotive forces balance. The current being zero at this point and the condenser electromotive force a maximum, it follows that the condenser electromotive force is 90 degrees in advance of the current and hence of the resultant electromotive force which is in phase with the current. The impressed electromotive force lags behind the current which in this case is a *leading current* and moreover the impressed electromotive force is greater than the resultant electromotive force. It follows that if either capacity or inductance be placed in an alternating current circuit, the phase of the current with respect to the impressed electromotive force will change and the current flow will be reduced. Capacity sets up an electromotive force 90 degrees in advance of current flow, while inductance sets up an electromotive force of 90 degrees behind current flow. These two effects tend to neutralize each other when the proper amount of inductance and capacity are connected in series in a circuit and when they are just equal, no electromotive force other than the impressed is left to act on the circuit, the resultant and impressed electromotive forces are identical and there is no phase displacement. This is said to be a *resonant* (q.v.) circuit. (See *Lead, Lag*, also *Inductive Rise.*)

LEADS—Wires conveying or "leading" current from one point to another in a circuit. The conductors in any system of electrical distribution are referred to as *leads*. Sometimes the term *lead* is applied to the flexible wire or cable conveying current to or from a piece of apparatus.

LEAK—See *Grid Leak*.

LEAKAGE LINES—Also termed *leakage flux*—That part of the total magnetic flux in a magnet, transformer, generator, etc., which fails to complete the closed magnetic circuit and hence is dissipated or wasted. (See *Keeper.*)

LEAKAGE, SURFACE—The flow of electricity across the surface of insulators. Its amount depends on the material and preparation of the surface and the humidity of the atmosphere and is recorded by measuring the surface *resistivity* (q.v.). Several methods are employed to reduce *surface leakage*. Porcelain *insulators* for outdoor use are often of the *petticoat* type so that a portion of the surface of the insulator remains dry during a rain or snowstorm. Another method very generally used in radio work to reduce surface leakage, is the use of corrugated surfaces on insulators. This method increases the length and hence the resistance of the surface across which leakage is likely to take place.

LENGTH OF AERIAL—The distance from the aerial binding post at the instrument to the farthest point of the antenna system, measured to any of its extremities. Only one component wire of a system is taken into consideration even if two, three or four are used. In a *T aerial* only half of the horizontal span must be added to the length of the down lead. In the *umbrella* type, the length of one radial lead is added to the down lead.

LENZ'S LAW—An induced current always tends to stop the motion which produces it. Stated in another way, whenever current flows in a conductor, due to electromagnetic induction, the action between the magnetic field which is being cut and the field around the conductor is always such as to tend to oppose further motion of the conductor. The following is still another way of stating *Lenz's Law*: The field due to the induced current always acts to prevent a change of the original field.

LEYDEN JAR—A condenser consisting of a glass jar with an outer and an inner coating of tinfoil covering the bottom and the sides nearly to the neck.

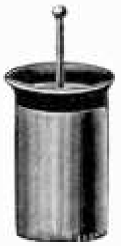

A small leyden jar.

A brass rod terminating in an external knob passes through a wooden stopper and is connected to the inner coating by a loose chain. This condenser received its name from Leyden, the city in Holland in which it was invented.

L. F. C.—Abbreviation for *Low Frequency Current*. (See *Frequency, Low*, also *Frequency*.)

LIGHT—Radiant energy or visible *radiation* which affects the eye so that objects become visible. Light consists of *ether* vibrations varying between 400 billions per second producing waves 271 ten-millionths of an inch in length and giving red light, up to 750 billions per second producing waves of 165 ten-millionths in length, giving violet light. Faster or slower vibrations produce waves not visible to the human eye. Shorter waves are called *ultraviolet rays*, while those immediately above the red rays are known as *infra-red rays*. The speed of light waves is identical with that of other electromagnetic waves and electricity, and is 186,000 miles or 300,000,000 meters per second.

LIGHT LINE AERIAL—A device which acts as an aerial for a radio receiving set and which utilizes the electric lighting system by means of a plug to which a condenser is attached. The plug is inserted in any lighting socket and the aerial connection of the receiving set it attached by means of a flexible lead to one terminal of the condenser. (See *Adapter, Aerial*.)

LIGHTNING ARRESTER—(Also called *lightning protector*.) A device for protecting radio or other electrical apparatus from lightning or from atmospheric discharges. Radio *lightning arresters* are provided with two terminals, one being connected with the aerial and the other directly to a ground. Between these terminals is a medium which offers practically no resistance (or technically speaking, which offers low *impedance* (q.v.)) to the flow of very high frequency currents such as lightning discharges. However, this medium acts as an insulator to the normal operating currents. Hence, during the ordinary reception of radio signals, the currents flow past the *arrester* to the receiving instrument but they are unable to pass through the insulating medium of the *ligtning arrester* to the ground. When there is a lightning discharge, the arrester offers the shortest and most direct path to the ground and hence the dangerous current never reaches the apparatus but proceeds directly to the ground through the *lightning arrester*. The more popular designs of lightning arresters are of three types: vacuum, non-air-gap and air-gap. The vacuum type is the more sensitive of the three, due to its having a low functioning voltage. In a usual form, it consists of two terminals in a sealed glass chamber from which air has been partially exhausted. This type of arrester can be discharged repeatedly without impairing its efficiency. The non-air-gap type of arrester involves the use of a carborundum stone, with a special clay binder. This type of arrester has a rectifying action similar to that used in a carborundum detector. The air-gap type of *lightning arrester* depends for its efficiency on a closely spaced air-gap which offers a low break-down voltage. The Underwriters requirements call for the use of an approved lightning protective device on all radio receiving sets where outside aerials are used. Further requirements are that such arresters shall operate on 500 volts or less. (See *Arrester*.)

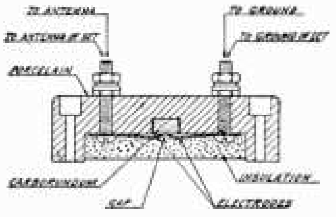

A carborundum type lightning arrester consisting of a block of carborundum held between two pieces of high resistance carbon.

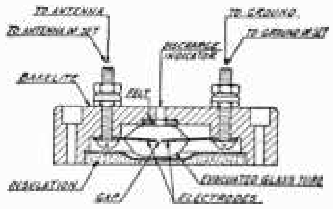

Vacuum type of lightning arrester for indoor work.

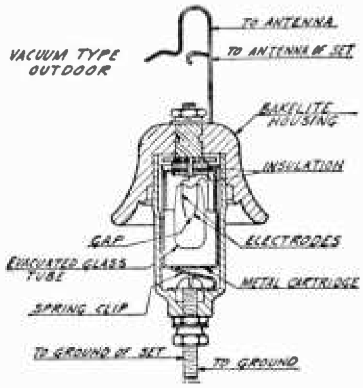

Sectional view of vacuum type lightning arrester used for outdoor work.

Outdoor vacuum type lightning arrester shown above in sectional view.

LIGHTNING GAP—This refers to the *air gap* or *vacuum gap* in a *lightning arrester* (q.v.). This gap is connected across the aerial and ground leads to the receiving set. When the potential in the aerial rises above a certain value (usually 500 volts or less) a spark jumps the gap, discharging the excessive energy to the ground without damage to the radio apparatus. (See *Gap Arrester*.)

LIGHTNING SWITCH—A single pole, double throw knife *switch* (q.v.) used for grounding the antenna during a storm, or in some cases at all times when the apparatus is not in use. The upper terminal of the switch is usually connected to the aerial, the central terminal to which the switch arm is connected, leads to the radio set and the lower terminal is connected to the ground. In most installations, the *lightning switch* is mounted outside the building in which the radio apparatus is located. (See *Ground Switch*.)

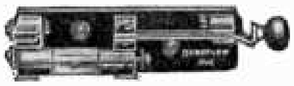

Lightning switch combined with vacuum type lightning arrester.

LIKE CHARGES—Bodies which carry like *electrostatic charge* (q.v.) exert a repulsion against each other.

LIKE POLES—Like magnetic poles repel one another. This is one of the *laws of magnetism*. (See *Magnetism, Laws of*.)

LIMITING VOLTAGE OF BRUSH DISCHARGE—The potential at which a brush discharge passes into an arc or spark. Brush discharges and arcs may be reduced in radio transmitting apparatus by eliminating sharp points or edges on conductors and by coating the edges of metal plates with paraffin. (See *Brush Discharge*, also *Corona*.)

LIMITING VOLTAGE OF GLOW DISCHARGE—The potential at which the glow gives place to the *brush discharge* (q.v.) in a high tension line or in a radio transmitting aerial (q.v.). (See *Corona*.)

LIMITS OF AUDIBILITY—The limiting *frequencies* (q.v.) at which audible sounds are produced. These range from 500 to 1000 cycles per second up to as high as 10,000 cycles per second. *Audio frequencies* (q.v.) are sometimes defined as those conveniently heard in the telephone. (See *Frequency, High*, also *Frequency, Low*.)

LINE—A conducting wire between terminals in a system of electric communication or distribution. The positive wire is sometimes specifically referred to as the *line* and the return wire as *earth*. In many cases the earth or ground is used in the place of a return wire.

LINE OF FORCE—A theoretical conception used to map out a magnetic field and to show the direction along which the magnetism acts. *Lines of force* are conceived as forming closed loops and the path they traverse is called the *magnetic circuit* (q.v.) The positive direction of the *lines of force* is arbitrarily taken as that in which the north-seeking end of a compass moves. *Lines of force* are represented in diagrams by curves so drawn that the direction of the tangent at any point is the direction of the *magnetic force* (q.v.) at that point. The closeness of the lines is used to indicate the intensity of the force. The *line of force* is also used to indicate and map out an *electrostatic* (q.v) field in the same way as it is used for the magnetic field. The *line of force* is also called the *line of flux* or the *line of induction*. This latter term is used especially where the line of force is being considered in respect to its electrical effect. In some cases it is simply referred to as a *line*. Thus a line (*of force*, or *of flux*, or *of induction*) is an electromagnetic unit of magnetic flux and is equal to 10^{-8} *webers* (q.v.). In the practical system of measurements, the *maxwell* (q.v.) equal to one *line*, is the unit of *magnetic flux* (q.v.). When 100,000,000 *maxwells* are cut by a single conductor, in one second, one *volt* (q.v.) will be induced. (See *Law of Electromagnetic Induction*, also *Electromagnetic field.*)

LINEAR OSCILLATOR—See *Oscillator, Linear*.

LINE RADIO—Also known as *Wired Wireless*—A system of communciation devised by Major General George O. Squier which permits the transmission of electromagnetic waves along a wire or conductor. Instead of the waves spreading out in all directions, as in ordinary broadcasting, the waves travel along a predetermined path, viz: the conductor. It is possible to transmit either code or speech by *wired wireless*. Vacuum tubes are used as oscillators at the transmitting end and a specially designed vacuum tube set is used as the receiver. When code transmission is desired, a telegraph key is inserted in the circuit to intermittently stop and start the oscillating waves. This produces dots and dashes and in this way it is possible to send code using undamped radio frequency waves. When it is desired to telephone, the key is omitted and the radio frequency or *carrier waves* (q.v.) are produced continuously except when they are varied or modulated in amplitude corresponding to the voice of the speaker at the transmitting end. The wave which passes from the transmitting end of the wire to the receiving end is of radio frequency and is called the *carrier wave*. In a way, this wave is analogous to the wire which carries the voice current in wire telephony. *Line radio* can be transmitted over ordinary telephone or telegraph lines at the same time as these lines are being used for their ordinary purposes and without any interference whatsoever. High-tension lines are also used for transmission of *wired wireless*.

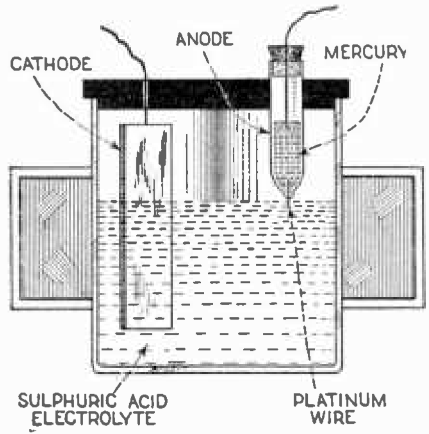

A liquid barreter.

LIQUID BARRETER—An electrolytic detector which utilizes a very fine platinum wire dipped to a small depth in nitric or sulphuric acid. These detectors were formerly used to some extent in receiving wireless code messages. The electrolytic detector of the *barreter* type was invented and named by *Fessenden* (q.v.). The illustration shows a liquid barreter. In the glass jar is a platinum glass-anode, lead cathode and sulphuric acid electrolyte.

LITZENDRAHT—abbreviated Litz—A German word, meaning braided wire. A high frequency conductor built up of a number of very fine conducting strands. The individual strands are enameled and the conductor is usually so constructed that each strand comes to the surface or very near the surface at regular intervals. In the most effective form of conductor the strands are braided so as to form a woven tube. The strands are usually covered by a single braided silk wrapper. The use of *litz* tends to reduce *skin effect* (q.v.). This type of wire has been found to be especially effective when used in the construction of *loop antennas* (q.v.).

LOADSTONE—also spelled *Lode-stone*—Iron ore which shows magnetic properties. The name loadstone is derived from a word meaning leading stone and refers to the property of a compass of pointing to the north magnetic pole. The loadstone is also referred to as a *natural magnet*. (See *Magnetite*, also *Magnet.*)

LOADING—Adding capacity or inductance to a circuit. *Inductance coils* (q.v.) called *loading coils* are used to increase the inductance and condensers are used to increase the capacity. A loading coil may be used to increase the *wave length* (q.v.) of an antenna by placing it in series between the *lead-in* (q.v.) and the receiving set. In the case of a transmitting antenna, the loading coil may be used to tune the aerial for *resonance* (q.v.) at a desired frequency. An example of the use of loading, is that of the addition of a loading coil to a receiving set for the purpose of enabling the set to receive higher wave lengths.

LOADING INDUCTANCE—An inductance coil used to increase the wave length of a circuit. This term usually refers to an inductance coil used in connection with a radio receiving set for increasing its receptive wave length range. Sometimes the term *loading coil* is applied to a unit consisting of an inductance coil, mounted beneath a panel in a small cabinet. On the outside of the panel is a contact arm and a number of studs. These latter lead to taps taken from the inductance coil. By moving the contact arm from one stud to another it is possible to vary the loading of the circuit as desired. A loading coil connected in series with the aerial will increase the natural wave length of the aerial. (See *Loading*, also *Inductance, Antenna.*)

LOCAL ACTION—Electrochemical action within a cell, usually at the positive plate, which does not add to the useful output. In primary cells, zinc is the material usually employed for the positive plate. This zinc practically always contains impurities such as carbon, iron, etc. When the zinc is immersed in the electrolyte, there is a difference of potential between the zinc and the impurities and small local currents flow from the zinc through the electrolyte to the impurities and back to the zinc again. These currents serve no useful purpose and moreover use up the zinc and the electrolyte. *Local action* can be practically eliminated by coating the surface of the zinc with mercury. This is called *amalgamation*. In secondary cells or *accumulators* (q.v.) local action is due to the action which takes place between the lead of the positive and the coating of lead sulphate.

LOCAL OSCILLATIONS—Oscillations set up in the vicinity of a receiving set for test purposes or within the set, as in the *super-heterodyne*, to produce oscillations differing in frequency from the incoming oscillations so as to establish a beat frequency. *Local oscillations* may be produced by means of a buzzer. This is a method frequently used in testing. The use of such oscillations enables a manufacturer of radio receiving sets to test these sets for selectivity, etc., without waiting for actual broadcasting. In the *super-heterodyne* (q.v.) local oscillations are produced by a suitably arranged vacuum tube. (See *Driver.*)

LOCUS—The path of a movement graphically represented by a curve. In mathematical language, the *locus* of an equation is the curve that contains all the points whose co-ordinates satisfy the equation. The circle is defined as the *locus* of all points in a plane equidistant from a fixed point in the plane called the center.

LODGE, SIR OLIVER—British Physicist and radio pioneer. Born at Penkhull, near Stoke-upon-Trent, June 12th, 1851, Oliver Lodge went to the Newport Grammar School at the age of eight, but at fourteen was taken into business to help his father. He worked in the evenings for the intermediate examination in science at the University of London, and eventually took first-class honors in physics. In 1872 he went to University College, London, and in 1877 took his D.Sc. in electricity and became demonstrator, and afterwards assistant professor in physics at University College.

Sir Oliver Lodge.

In 1881 he was appointed first professor of physics at Liverpool, and in 1887 he was made a Fellow at the Royal Society. In 1888 he received the honorary L.L.D. of the University of St. Andrews, the first of a long series of such distinctions. In 1900 he was appointed the first principal of the University of Birmingham, a post from which he retired in 1920. In 1902 he was knighted. Sir Oliver Lodge is one of the foremost physicists of his time, and his brilliant series of researches into electrical phenomena, which he has made his special study have had a profound effect in all branches of that subject. To him, indeed radio owes as much as to any man, for he dealt with fundamentals when the whole subject of electromagnetic radiation was in its infancy. He was led to study the surging and oscillating character of an electrical discharge along wires as a result of his investigations as to the best means of protection against lightning.

One very remarkable experiment of his is now well known under the name of Lodge's resonating jar, now used as a measurer of wave length. In these experiments Lodge was really dealing with the electromagnetic waves discovered by Hertz in 1888. Lodge early recognized the vast importance of the discovery by Hertz, which so brilliantly confirmed the theories of Clerk-Maxwell, and he threw himself into work of investigating these ways with an energy which made him the foremost British physicist on the subject at the time. All this was pioneer work before Marconi and others came along and built upon the solid foundations which Lodge had prepared.

It was during this early period that Lodge discovered his coherer, and with this detector devised the first practical wireless telegraph, sending signals over a distance of several hundred yeards. In May, 1897, Lodge took out his fundamental patent for tuning—a patent that was successfully upheld in the law courts, and extended for seven years by Lord Parker. The patent was acquired by the Marconi Company in 1911. Sir Oliver Lodge has been a pioneer in many other directions. His theories and experiments in connection with the ether, of which he is the leading exponent, are well known, and he has carried out a number of interesting researches on the passage of electricity through liquids. Sir Oliver was the first to devise a simple and direct experiment to show the speed and direction of travel of the constituents of a salt in solution when broken up by a current of electricity. For his researches on electric waves and the passage of light through a moving medium, Sir Oliver Lodge received in 1898 the Rumford Medal of the Royal Society, one of the highest honors the society bestows.

Sir Oliver Lodge has won a well-deserved reputation for the clearness and simplicity with which he is able to explain many of the difficult and fundamental truths of physics, and this clarity of expression is well shown in the many articles he has written.

LODGE DISC DETECTOR—An imperfect contact detector consisting of a steel disc rotating with its sharp knife edge just touching some mercury having a film of oil on it. When an oscillation passes through the circuit, momentary cohesion sets in between the disc and the mercury, but this is immediately broken by the revolving motion of the disc. In this way, oil re-insulates the disc at the completion of each passage of oscillations. Using this type of detector, code signals can be recorded with great precision on a *Morse* tape using a syphon recorder. The *disc detector* is in reality a specialized form of *coherer* (q.v.).

LOG—A list of dial readings of a radio receiving set with corresponding call letters of the broadcasting stations received at these readings. A radio set is said to *log* when the same stations always come in at exactly the same dial readings. Thus on a three dial set, if WEAF comes in at 61, 60 and 60 and always comes in at these identical readings using the same aerial, the set *logs*. All present day sets are constructed so that stations can be accurately logged. Sometimes the term log is used to refer to a graph or chart on which dial settings are plotted against wave lengths of stations recorded. A graph is a diagrammatic representation by means of lines, of the relation between several quantities. Figure 1, gives a simple illustration of this. Suppose it is desired to obtain a general idea of how the radio business varies from year to year or from season to season. On a piece of paper, ruled both vertically and horizontally, as in Figure 1, the months and years are marked at the bottom horizontally. The vertical lines then represent the months and the bottom horizontal line is called the *axis of time*. These vertical lines are known technically as *ordinates*. The vertical line to the extreme left is called the axis of ordinates and after it has been marked as shown in Figure 1, to represent dollars, it may be referred to as the *axis of dollars*. The horizontal lines are known as *abscissas*, the bottom horizontal line being the *axis of abscissas*. Assume that a certain radio concern did $10,000 worth of business in the month of January, 1924. Looking upward along the vertical line representing this month, until the horizontal line representing $10,000 is reached, a point is put at the intersection of these two lines. Suppose that in the following month, February, 1924, the firm does $12,000 worth of business. The point is found as before by following the line February, 1924, up to the line indicating $12,000. Each horizontal line in the figure represents $1,000, so that this point should be found two lines higher than the one marked for $10,000. This can then be continued indefinitely, plotting the amount of business for each month. Assume that in the summer there is a lull in business. The decline starts, say, in the month of May. The business done amounted to $8,000. By glancing at the chart it is possible to ascertain a number of facts which stand out much clearer than where columns of figures are given. Thus the seasonal character of the business is apparent at a glance. It can also be seen readily that the business was better in 1925 than in 1924, etc. Radio logs are plotted in exactly the same way. Each point must be obtained individually in order for the curve to be absolutely accurate, although in many cases a few well chosen points are sufficiently satisfactory. In Figure 2, the dial setting is marked along the bottom of the chart. The scale is here chosen at one division on the chart equal to 5 divisions of the dial. If preferred the scale could be chosen so that one division on the chart would equal one division on the dial as shown on chart on the following page, which is a typical log of an actual set. Proceeding, it is first necessary to find out the limits of the chart. The limits of the horizontal scale are already known, i.e., zero to 100 on the dial. Looking over the list of broadcasting stations, it is found that they all have wave lengths which fall, roughly between 200 and 600 meters. Therefore it is not necessary to make the vertical scale include wave lengths outside of this range. Consequently the vertical scale is started at 200. Counting upward, two divisions on the chart are taken, for convenience, as equal to 50 meters. Having done this, actual readings are next taken on the radio set. Suppose that the first station tuned in is WEAF and that this is tuned in at 85 on the dial. This station has a wave length of 492 meters. Looking along the bottom of the chart for 85, the vertical line is then followed until the position of 492 meters is reached. This is slightly below the 500 line on the chart, Figure 2, and is indicated by the letter A. Next suppose that WOO, having a wave length of 509 meters is tuned in at 90 on the dial. This, then gives the point marked B. Then another station is tuned in, indicated by the point C,

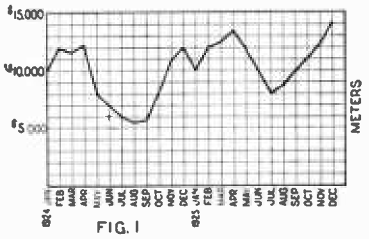

FIG. I

Log showing variation of a business.

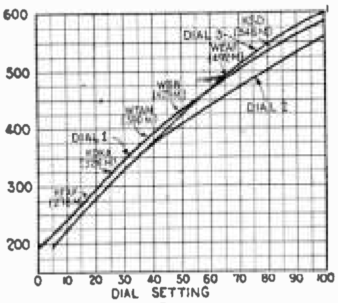

FIG.3

A three-dial radio log using only one chart.

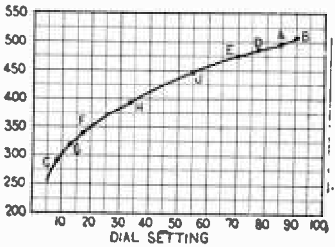

FIG.2.

A radio log.

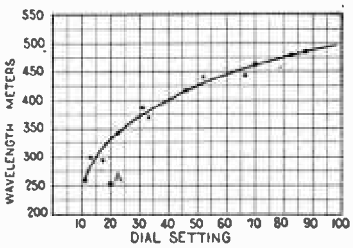

FIG.4

Radio log showing curve drawn so as to touch as many points as possible.

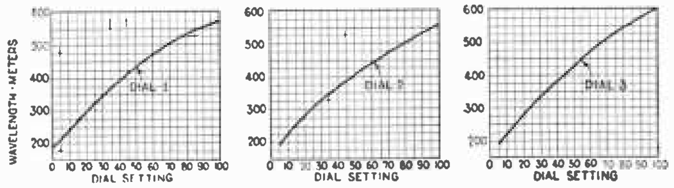

Fig. 5—A three-dial radio log using three separate charts.

Various stations are thus located by points on the chart until enough points are obtained to draw the curve. If the points do not give a smooth curve, it is necessary to approximate a curve, that is to draw it so as to pass through as many of the points as possible. This has been done in Figure 4.

Sometimes points are obtained that do not seem to have any meaning, such as the point A in Fig. 4. This point may have been subject to inaccuracies in reading the dial, or perhaps the operator did not read the wave length correctly from the list of stations. There is always the possibility, also, that he did not plot the point properly. He has plotted the point A as 251 meters at a setting of 20 when it might have been 335 meters at 20 on the dial, in which case it would have fallen directly upon the curve. If none of these difficulties can be found with the point A it can be discarded, however, because it is an absurd point, as the 12 other points, which determine the curve, will testify. Generally it is not good practise to discard points, but the experimenter must learn to use his good judgment.

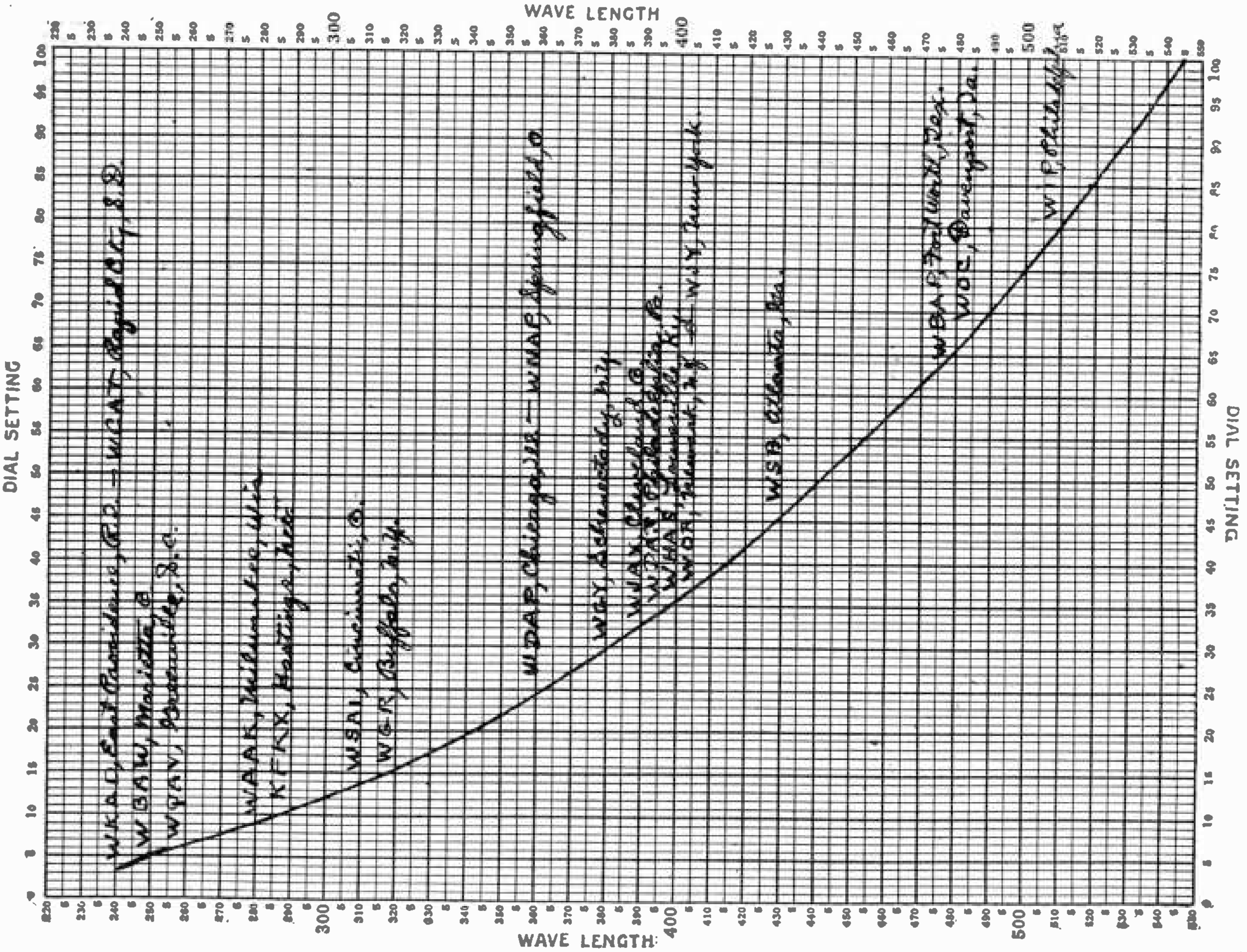

An example of a station log chart.

In single-control sets, where only one dial is employed it is only necessary to have one of these radio log curves. In receivers, such as the Neutrodyne or any two-stage radio frequency amplifier, where three dials are employed, it is generally necessary to log each dial separately. The reason for this is, that although these sets are supposed to tune to stations with identical dial settings, many of them do not do so, a difference as great as 10 divisions on the dial often being found. A three dial log is shown in Fig. 5. Separate charts may be used for each dial, or, for convenience, as has been done in Fig. 3, the three charts may be combined into one. If the difference between the dial settings is very small, say only a division or two, one curve will be sufficient.

Many things can be learned from studying the curves obtained. Take for instance, Fig. 2. This is practically a straight line at the higher wave lengths, changing from about 500 meters to about 400 in 55 divisions of the dial. In the remaining 35 divisions of the dial the wave length changes from 400 to 250. In the first case there was a range of 100 meters for 55 divisions and in the latter case a range of 150 meters in 35 divisions. This is because the curve bends downward. At the low dial settings a small change in the dial reading means a relatively large change in wave length, while at the upper end of the dial the same change in dial reading will mean a relatively small change in wave length. The reason why the stations are so crowded at present, at the small end of the dial, is then very apparent, i.e.:

To illustrate, a station operating on a 200 meter wave corresponds to a frequency of 1500 kilocycles (1 kilocycle equals 1000 cycles), and a station on 200 meters corresponds to a frequency of 1,363.5 kilocycles. In other words the difference in wave length is only 20 meters whereas the difference in kilocycles in 136.5. These two wave bands are at the low end of the present broadcast wave band. Now at the high end of the band we find a wave length of 600 meters corresponds to 500 kilocycles, and a wave of 580 meters corresponds to 517.2 kilocycles, or a difference of but 17.2 kilocycles for the 20 meter difference in wave length. At the lower end of the broadcast band between 200 and 220 meters we found a difference of 20 meters equalled 136.5 kilocycles. This indicates that as the wave length becomes lower the difference in frequency becomes very high.

To bring this most strongly to the mind, go down to the extremely low wave of 5 meters. Here the frequency is found to be 60,000 kilocycles. Going up to 25 meters, this corresponds to 12,000 kilocycles, so that for this difference of 20 meters there is the tremendous difference of 48,000 kilocycles.

The accompanying illustration, Fig. 6 will explain the matter very clearly. In changing the dial setting from 100 to, let us say, 90, the wave length is changed roughly from 600 to 580 meters. This is a change from 500 to 517 kilocycles. On the other end of the dial from zero to about 10 we have a change of wave length of roughly 200 to 220 meters, an approx-mate change of from 1500 to 1363 kilocycles. In the first case on the upper part of the dial we have only made a change of 17 kilocycles in the frequency. On the lower end of the dial we have made a change of 137 kilocycles.

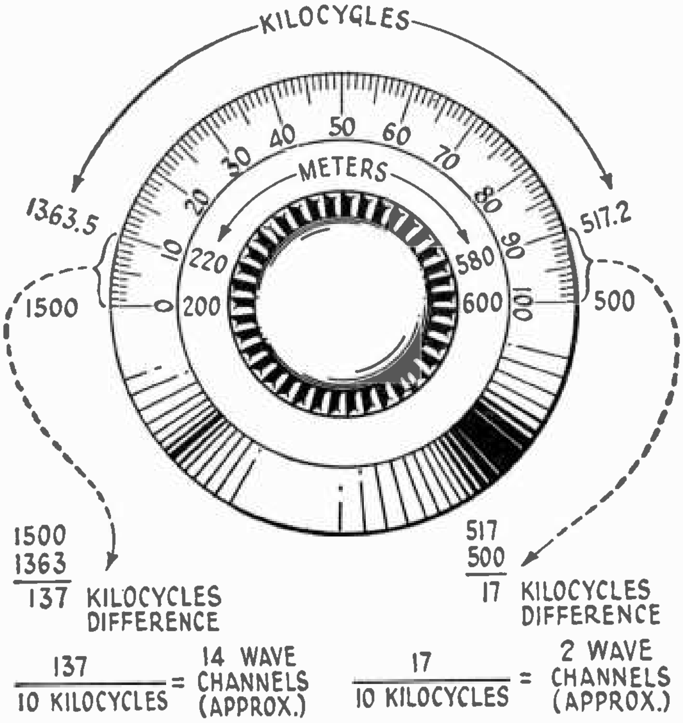
KILOCYCLES
METERS
0
10
20
30
40
50
60
70
80
90
100
1363.5
517.2
1500
500
200
220
580
600
1500
1363
137 KILOCYCLES DIFFERENCE
517
500
17 KILOCYCLES DIFFERENCE
137 / 10 KILOCYCLES = 14 WAVE CHANNELS (APPROX.)
17 / 10 KILOCYCLES = 2 WAVE CHANNELS (APPROX.)

Fig. 6—This illustration shows clearly why stations are crowded on the dial at the lower wave lengths.

According to the frequency allocation as made by the Government, which places broadcast frequencies 10 kilocycles apart, on the upper 10 divisions of the dial there is room for approximately 2 frequency channels whereas on the lower 10 divisions of the dial there is room for roughly 14 frequency channels. This shows very clearly that we would have to be able to separate 14 stations on the lower 10 divisions whereas we have only 2 to separate on the upper 10 divisions of the dial.

So it will be understood why turning the dials but a degree or two at the low setting causes a relatively large change in frequency, tuning in perhaps several stations within one degree of the dial, while at the higher settings, due to the small difference in frequency, as stated, on the higher waves, it may be necessary to turn the dials two or three degrees on either side of the peak of the broadcast wave before it is tuned out. (See *Dial.*)

LOGARITHM—abbreviation LOG—The power to which a number (defined as a base) must be raised in order to equal a given number. As an example, given the number 100, assume that it is desired to determine the *logarithm* of this number to the base 10. (When 10 is used as a base, the *logarithms* obtained are said to belong to the *common* or *Briggs system.*) Then the exponent or power which will raise 10 to the value of 100 is known as the *logarithm* of 100. Obviously, 10^2 equals 100, therefore 2 is the logarithm of 100. *Logarithms* are used in the processes of multiplication, division, raising to powers and taking roots. The whole number, or integer part of a logarithm is called its *characteristic* and the fractional part, is referred to as its *mantissa*. The characteristic of a number greater than unity is positive and is one less than the number of figures to the left of the decimal. The characteristic of a number less than unity is negative and is one greater than the number of ciphers between the decimal and the first significant figure. A *table of logarithms* gives the mantissas only, the characteristic being determined as indicated above.

Logarithms are also used which have as their base the number 2.718 +. These are called *natural* or *Naperian logs*. The base number is usually represented by the symbol ε (epsilon). The natural logarithm of any number may be found from a table of common logarithms by dividing the common log by the logarithm of 2.718 or simply by multiplying the common log by 2.30259. The symbol l_n is often used for the natural logarithm and the abbreviation *log* without a subscript is usually used for the logarithm to the base 10. (See *Naperian Log*, also *Mantissa*, and *Log Table.*)

LOGARITHMIC CURVE—A curve which has no definite minimum in a decreasing curve, or no definite maximum in an increasing curve. A *logarithmic curve* may be further defined as a curve whose ordinates increase arithmetically while its abscissas increase geometrically. A series of damped electrical oscillations diminish according to a logarithmic law and hence a curve drawn tangent to the successive maximum displacements will be a *logarithmic curve.*

LOGARITHMIC DAMPING—See *Damping.*

LOGARITHMIC DECREMENT—The *Naperian logarithm* (q.v.) of the ratio of any maximum to the next following maximum, with the current in the same direction in a decaying wave train. In other words the logarithm of the ratio of two maxima, one cycle apart. The rapidity with which the oscillations decay depends not only on the resistance of the circuit, but on the inductance as well. The greater the resistance and the smaller the inductance, the more rapid is the damping and the rate at which the oscillations increase. If the resistance, inductance and capacity of the circuit have fixed values, each successive maximum of current is the same fraction of the preceding maximum as the latter is of the maximum immediately preceding it. Thus, if the second maximum is 75 percent of the first maximum, the third will be 75 percent of the second, etc. The numerical value of the rate of decrease in this case is .75 and the natural logarithm of .75 is the *logarithmic decrement* or simply the *decrement.* (See *Decrement*, also *Damped Waves.*)

LOG TABLE—A tabulated arrangement of logarithms which gives the *mantissas* (q.v.) only. The numbers whose logarithms are to be looked up are arranged for convenience in consecutive order. (See *Logarithm.*)

LOMBARDI, DR. LUIGI—Italian radio authority. Dr. Lombardi was born at Dronero, August 21, 1876. He graduated from the Royal Engineering School at Turin, Italy. He was Professor of Electricity at the Zurich Polytechnical School, Zurich, Switzerland, and also at the Royal Polytechnic School, Naples, Italy. Dr. Lombardi published many scientific papers including a study of condensers for transmission and he also invented a special high-tension electrical condenser.

LONGITUDE—The distance east or west of a meridian passing through Greenwich, England, measured in degrees, minutes and seconds.

LOOP ANTENNA—*Loop Aerial*—Also called *Coil Aerial* or simply *Loop*—A receiving or transmitting *aerial* (q.v.) consisting essentially of one or more turns of wire connected to the radio set so as to form a closed circuit. The *loop* acts as a simple inductance coil. The ordinary radio aerial acts as a huge condenser with the aerial system as one plate of the condenser and the *ground* or *counterpoise* (q.v.) as the other. Electromagnetic waves reaching the aerial set up an alternating electromotive force between the wires forming the upper plate of the condenser and the ground or lower plate of the condenser. This action takes place through *electrostatic induction* (q.v.). In the *loop aerial, electromagnetic induction* (q.v.) sets up an induced electromotive force thus causing alternating current to flow to the detecting apparatus. While coil antennas may be used for transmitting, their use in this connection is rather limited. On the other hand, the use of the loop for radio receiving is widespread. The greater the area enclosed

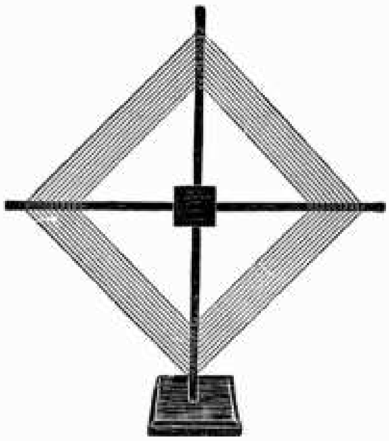

A typical loop antenna.

by a loop the greater the amount of energy it will receive. Loops more than two feet in diameter are usually constructed so as to be collapsible. A common type of loop shown in the illustration is 36 inches high and 28 inches wide. It consists of twelve turns of wire wound in the same vertical plane and held in place by polished hard rubber strips. It has a heavy cast base permitting rotation of the loop as desired. This loop folds up so as to fit into a four inch tube. In another type of loop, sometimes called the *box type*, the wires are wound so as to lie in the same horizontal plane instead of the vertical plane. The use of *Litzendraht* (q.v.) for winding loops has been found to add to their efficiency. Loops are commonly used on *super-heterodyne* (q.v.) sets, *reflex* (q.v.) sets and in some *tuned radio frequency* (q.v.) sets. These latter generally require an extra stage of amplification, although tuned radio frequency sets having only five tubes (detector, two stages of radio frequency and two stages of audio frequency) will operate on a loop and give especially good results when used with the new power tube. As a general rule, it may be stated that any set which will give good results with a loop, will give better results, insofar as distance is concerned, with an outside aerial. For this reason loop operating sets are quite generally constructed with extra binding posts for connecting aerial and ground. The loop is connected directly into the grid circuit, while the aerial and ground are connected to the circuit through an inductance such as an *oscillo-coupler* or antenna coupler. Loops may be used with satisfactory results inside of buildings. Where a compact, portable and selective aerial is desired, the loop is especially useful. The use of the loop, results in a marked directional effect and also has a tendency to reduce *static* (q.v.) and other forms of *interference* (q.v.). Loops are employed in wireless direction finders or *radio compasses* (q.v.) and are also used as aerials in submarine and aeroplane work. Where the loop is used as an aeroplane antenna it may be wound on the wings of the plane or it may be wound on a small frame and placed in the rear of the plane. (See *Coil Antenna*, also *Frame Aerial.*)

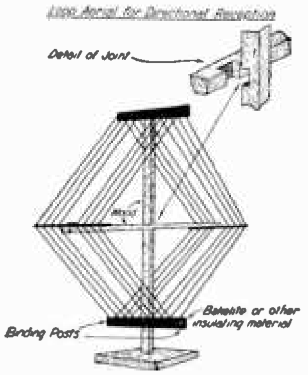
Loop Aerial for Directional Reception
Detail of Joint
Bakelite or other insulating material
Binding Posts

A box-type loop.

LOOPS—The points of greatest amplitude in a wave train. Sometimes called *antinodes*. The term *loop* is also used in describing the electrostatic field spreading out from a charged aerial. Each time an aerial is charged (for example by an induction coil) and discharged across a spark gap a group of gradually decaying oscillations is produced in the aerial called a *train* of oscillations. The interval of time which elapses between a flow of the electricity in the aerial in one direction, and a succeeding similar movement is called the *periodic time* and the reciprocal of this is called the frequency. For example, if the periodic time is one-millionth of a second, the frequency is one million. The diagram shows the manner in which *semi-loops* of electric force represented by the dotted lines move away from the aerial which is represented by the short black line.

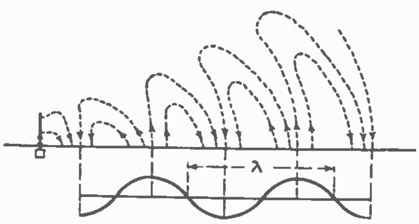

Semi-loops of electric force represented by the dotted lines moving away from the aerial represented by the short black line.

LORENZ COIL—A form of low loss inductance coil, having an air core and of cylindrical shape, being wound in a single layer, with turns slightly kinked to make the coil self-supporting. (See *Coil, Inductance Coils*, also *Low Loss Coils.*)

LOUD SPEAKER—A sound producing device for converting audio frequency currents into sounds loud enough to be heard by an assemblage of people. Nearly all *loud speakers* utilize an electromagnet in order to directly or indirectly cause a diaphragm to vibrate.

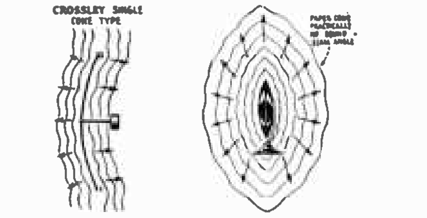

At the left is shown a cone-type speaker having a single diaphragm; at the right one with a double diaphragm.

The audio frequency currents flow through the windings of the electromagnet and the vibrations of the diaphragm correspond to the minute changes and fluctuations of the current thus accurately reproducing sounds. In loud speakers which utilize a metallic diaphragm, the electromagnet directly actuates the diaphragm. In other types, the diaphragm may be of mica, as in the Baldwin unit type, or of parchment or wood as in the cone type speakers. In these speakers, the electromagnet actuates an armature which in turn actuates the diaphragm through a lever action. In certain loud speakers, the electromagnet is of very powerful type and has a very small coil of fine wire attached to a diaphragm but not touching the poles of the magnet. The small coil is connected to the radio set or amplifier and the fluctuating audio frequency currents pass through it. These currents force the coil out of the strong magnetic field in the air gap. In moving, it moves the diaphragm to which it is attached. Speakers of this nature are sometimes called electrodynamic speakers. They are also referred to as power speakers. Such speakers give plenty of volume and in certain makes are reasonably good in tonal quality but they have the disadvantage that the electromagnet must be energized from a source of current such as the storage battery. A loud speaker has been developed in England called the Crystavox for use with crystal sets. This consists essentially of an electromagnetically operated speaker connected with a microphonic relay. The relay is actuated by a storage battery. Its function is to magnify the incoming signals from the crystal set, passing them along with increased strength to the electromagnet of the loud speaker. Still another type of speaker which has been used to a limited extent abroad, dispenses entirely with the electromagnet as the actuating source of diaphragm vibration. In this type, electrostatic principles are involved. An agate cylindrical drum, over which a band of copper or similar metal is wound, is revolved at a uniform rate by a small motor or by clockwork. To one end of the band a diaphragm is fastened mechanically. The other end of the band is secured by a fairly stiff spring. Current from the receiving set is led to the cylinder by means of a brush contact and as the current fluctuates the attraction of the band to the cylinder fluctuates with it and the drag on the diaphragm due to the rotation is increased and decreased. The use of the motor constitutes the most serious disadvantage in this type of loud speaker.

Having discussed the various *types* of loud speakers, from the standpoint of principles of operation, the more important of these will now be considered from the constructional standpoint. The early forms of directly actuated metallic diaphragm, electromagnetic loud speakers consisted essentially of a watch-case telephone receiver, or a pair of receivers attached to a horn. The purpose of the horn was to concentrate and redistribute the sounds caused by the vibration of the diaphragm. One of the first improvements was to modify the shape of the horn so as to make it conform more nearly with correct acoustical principles. Another improvement was to introduce an adjustment by means of which the air gap could be varied. In some cases this was accomplished by moving the magnet closer to the diaphragm and in others by moving the diaphragm itself. This adjustment provided a means of getting greater sensitivity on distant stations (by drawing diaphragm closer to the magnet) and of preventing sticking of diaphragm or rattling, on local stations. The ordinary metallic diaphragm has *one* frequency to which it will respond most strongly. This is called its *resonant* frequency. Special speakers

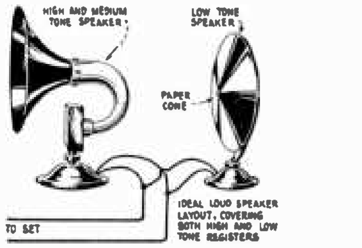
HIGH AND MEDIUM TONE SPEAKER
LOW TONE SPEAKER
PAPER CONE
TO SET
IDEAL LOUD SPEAKER LAYOUT, COVERING BOTH HIGH AND LOW TONE REGISTERS

A high and a low pitched speaker connected in parallel for simultaneous operation.

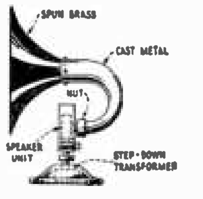

A loud speaker of the horn type having curved cast metal neck and spun brass bell.

called *tuned receivers* have been designed in England which permit the resonant frequency to be varied. In one type, the variations of the magnetic field operate a vibrating reed attached to a non-magnetic diaphragm. The resonant frequency can be varied by adjusting the position of the vibrating reed with a set screw. Another method of getting rid of resonant points in diaphragms has been to use corrugated instead of flat diaphragms. In some of these diaphragms the center is left flat, the remainder of the diaphragm being corrugated with concentric circles either evenly spaced or spaced in section at radii bearing a ratio to each other corresponding to prime numbers. In some cases conical diaphragms with corrugated edges are used. The material as well as the shape of the horn have been found to materially affect the sound reproduction. Horns are constructed of metal, fibre, composition, hard rubber, wood, etc. In general the horn should be of a material which will not give out metallic sounds. The tone arm which carries the speaker unit at one end and the horn at the other is usually constructed of cast metal such as a lead alloy or aluminum. The requirement here is that it must be of an acoustically "dead" material. In some cases the speaker unit and sound chamber are concealed in an artistically designed cabinet thus forming a *cabinet speaker*. Loud speaker units are also made especially for use in conjunction with the sound chamber of a phonograph. These are referred to as *phonograph attachments*. They are ordinarily attached to the phonograph tone arm being put in place of the phonograph reproducer. Some phonographs are now made with provision for attachment of the speaker unit without removal of the reproducer. There is also an attachment on the market for use in connection with a phonograph which has provision for conveniently changing over from the reproducer to the speaker unit or vice versa. Cone speakers may be of the single cone variety or of the double cone style. In these speakers a metal rod extends from the magnet armature and is rigidly fastened to the center of the cone. The most common type of cone speaker has a parchment diaphragm. This seems to give excellent reproduction especially for tones in the lower register. Its points of disadvantage are that it is affected to some extent by moisture and temperature change and furthermore seems to bring out static interference more clearly than some other types. Another type of cone speaker uses a teak wood diaphragm. Some of these speakers have a small fastening screw at the point of attachment between the rod and the diaphragm. This is provided so that tension can be removed from the diaphragm during shipment and also if the speaker is to remain in a room where change in temperature may take place. A cylindrical parchment speaker which utilizes the same principle as the cone speaker, has recently been placed on the market. Its chief point of difference is that the parchment is not under tension as in the case of the cone speaker. Large volume loud speakers are provided with power amplifiers which are often an integral part of the speaker itself. One of the most recent loud speakers is contained in a large cabinet and contains apparatus for utilizing ordinary lighting circuits for operating the set and speaker. The power amplifier unit is also contained within the speaker cabinet and the design of the speaker is such that it can handle any volume without distortion. The rectifier is arranged to supply not only the necessary voltages for the power amplifier but also plate voltage for practically any type of receiving set. (See *Horn, Loudspeaker,* also *Distortion.*)

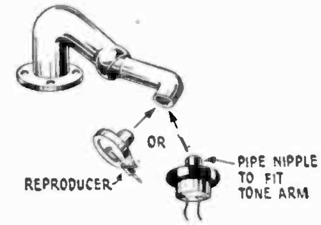

A loud speaker unit used as a phonograph attachment.

LOUDSPEAKER UNIT—The sound reproducing unit attached to the tone arm of the horn type speaker or to the diaphragm of the cone type speaker. Various types of units are described under the heading *Loudspeakers.*

LOW FREQUENCY—See *Low Frequency Current.*

LOW FREQUENCY CURRENT—An alternating current of less than 10,000 cycles. In radio work the alternating currents used are often classified as *low* frequency currents and *high* frequency currents. While there is no fixed value separating high and low frequencies, 10,000 cycles is generally taken as the dividing line. Thus audio frequency currents are considered as low frequencies and radio frequency currents are high frequencies. The term *low frequency* current as applied to power and lighting circuits refers to circuits having frequencies of 25, 60, etc., cycles per second. (See *Current, High Frequency.*)

LOW FREQUENCY IRON CORE INDUCTANCE—A variable inductance having an open-ended iron wire core. This inductance is used with spark transmitters to place the primary transformer circuit in resonance with the alternating current frequency.

LOW LOSS COILS—*Inductance coils* (q.v.) having low resistance and specially wound so as to reduce to a minimum condenser effect and hence dielectric losses. The term low loss has been used rather loosely but for practical purposes coil losses should be taken as referring to self-capacity, often called *distributed capacity* (q.v.) ohmic resistance, *skin effect* (q.v.) in the wire due to the effects of frequency and leakage and absorption in or through the insulating materials upon which the coil is built or with which the wire is covered. While it is impossible to obtain a coil having pure inductance only, it is feasable to design the coil so that the various losses will be reduced to a minimum. Often, it has been found that a design which will greatly reduce one type of loss will result in an increase of another. The illustrations show various types of low loss coils together with curves showing comparative efficiency tests. Figure 1 shows the results of an investigation regarding the distributed

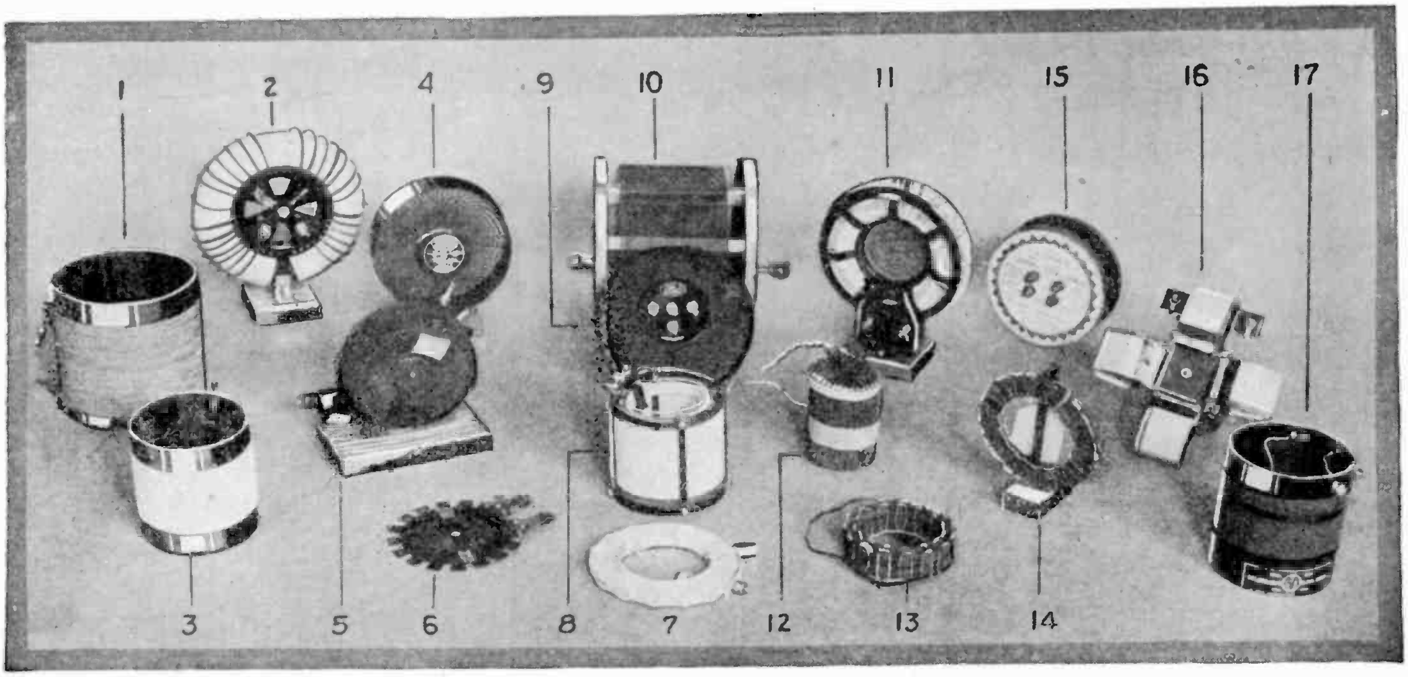

Various types of low loss coils including: 1, Bell wire coil; 2, Naxon; 3, Marco; 4, All-American; 5, Orbit; 6, Turney; 7, Sickles; 8, Aerocoil; 9, Erla; 10, Wavemeter coil; 11, Thorola; 12, Pathe; 13, Freshman; 14, Coast Coil; 15, Summitt; 16, Quadroformer; 17, Bruno.

capacity of various coils. This was obtained by making measurements of the inductance of the coils. If the self-capacity of the coil is very small, the self-inductance will not change perceptibly as the frequency changes. On the other hand, if the self-capacity amounts to anything, the curve showing the relation between inductance and frequency will bend upward as the frequency increases, and the greater the capacity of the coil, the more sharply will the curve bend. The curves shown in Figure 1 do not give any idea of what value the self-capacity may have, but simply show whether the inductance varies or not, which is one method of determining how harmful is the effect of the distributed capacity. The more sharply the curve bends upward the less desirable is the coil. In general, the smaller the size of the wire (for sizes smaller than about No. 16 B & S) the higher the resistance. The larger the wire and the thinner the insulation, the greater is the skin effect. The greatest inductance is obtained for a given amount of wire when the coil has a true cylindrical shape. The skin-effect in multi-layer coils is much greater than in single-layer coils. The effect of coil capacity and absorption or leakage in insulation is small compared with the skin-effect except in multi-layer coils, where the distributed capacity may become very great. To keep the physical size of toroidal coils within practical limits the diameter of the turns must be relatively small, so that many more turns are required to obtain a given inductance.

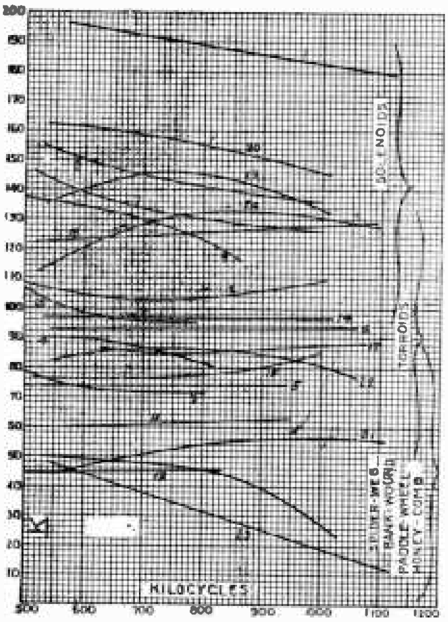

The efficiency of a coil is indicated by the height of its curves on this graph.

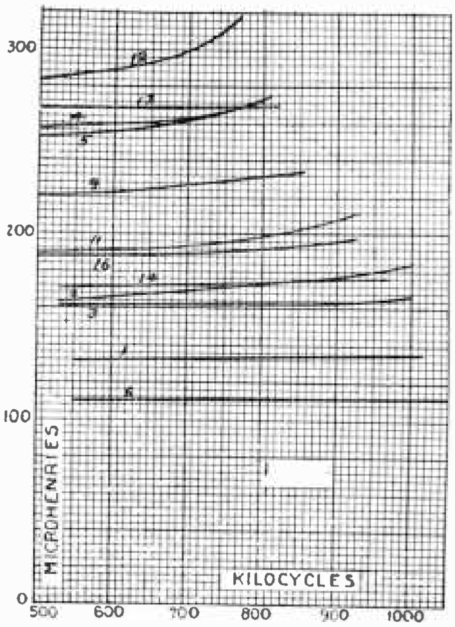

Fig. 1.—The curvature of the graphs indicate the presence of distributed capacity in the coils.

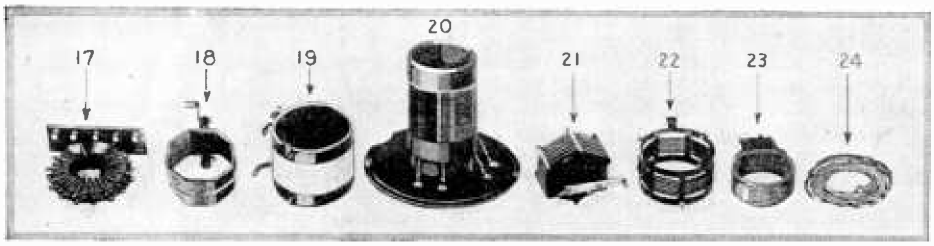

Additional examples of low loss coils: 17, Marwol; 18, Eastern Coil; 19, Workrite; 20, Walbert; 21, Andrews; 22, Bremer Tully; 23, Cotocoil; 24, Kresge Lorenz coil.

LOW TENSION — abbreviation L. T. — A comparatively low voltage. A circuit supplied by a low voltage battery such as the filament circuit of a vacuum tube is called a low tension circuit. In electrical engineering work, any circuit of 600 volts or under is considered to be a low tension circuit.

LUGS—Metal projections used for permanently connecting conductors with electrical apparatus. In one type of lug, the metal is bent over at one end to form a tube into which the conductor can be soldered, while the other end is flat and has a hole drilled to receive any desired terminal.

Typical lugs.

M

MAGNET—A piece of iron or steel having the power to attract other pieces of iron and to attract or repel other magnets. Natural magnets are known as *Loadstones* (q.v.). (See *Electromagnet*, also *Permanent Magnet* and *Temporary Magnets*.)

MAGNET WIRE — Insulated wire for the construction of magnet coils. In radio extremely fine wire is used to wind the coils of the electromagnets used in headsets and loud speaker units. The most common insulation used in radio work for magnet wire is enamel. Other types of magnet wire utilize a single cotton covering, a double cotton covering or a silk covering. (See *Enameled Wire*.)

MAGNETIC ATTRACTION AND REPULSION—The mechanical force tending to draw together two magnetic poles of unlike polarity, or to force apart two magnetic poles of like polarity.

MAGNETIC BLOWOUT—An electromagnet with its pole pieces placed so that the direction of the magnetic field will be at right angles to the movement of the ions in an electric arc. The coil of this magnet is energized in some cases in series with the line and acts to expel or blow out the arc. In the series arrangement, the blow-out action is stronger, the larger the current to be interrupted. For small shunted arcs, permanent magnets are often used. The operating principle of the blowout coil depends upon the fact that in a magnetic field, an electric current tends to move across the lines of force or in other words to cut them. In the illustration, the pole pieces are shown so placed that the general direction of the lines of force is at a right angle to the flow of the ions in the arc. The electromagnetic action causes the arc to move outside the field. This increases the length of the arc, thus rapidly breaking it. Magnetic arcs are used in connection with *keys* (q.v.) used for interrupting a large current, in certain types of *lightning arresters* (q.v.) and also in motor controllers..

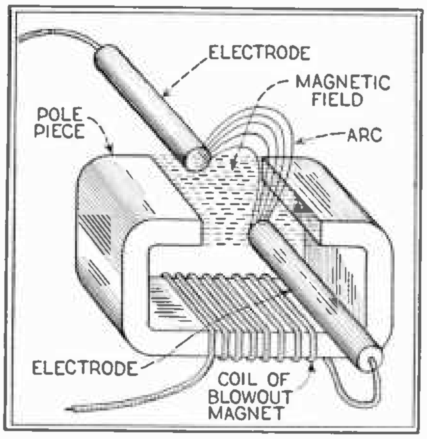

A magnetic blowout.

MAGNETIC CIRCUIT — The path traversed by magnetic lines of force. A magnetic circuit may be entirely through iron, as in the case of a closed core transformer or partly through iron and partly through air as in the case of a motor or generator. (See *Line of Force*.)

MAGNETIC COUPLING—See *Coupling*.

MAGNETIC COUPLING TRANSFORMERS—See *Coupling Transformer*, also *Jigger*.

MAGNETIC FIELD — The whole space over which a magnet exerts influence. The space traversed by the magnetic flux. Wherever magnetism is present the space in the immediate vicinity is called a magnetic field. (See *Electromagnetic Induction.*)

MAGNETIC FLOW—The passage of magnetic lines through a magnetic circuit. This term is sometimes used in the place of magnetic flux. (See *Flux Magnetism.*)

MAGNETIC FLUX—Symbol ϕ—The total magnetism present at any cross-section perpendicular to the lines of force. Magnetic flux is analagous to current flow. That is to say, it is convenient to consider a flow of magnetism in the same way as electricity is considered to flow. The unit of magnetic flux in the c.g.s. (centimeter gram second system) electromagnetic system is the *maxwell* (q.v.) or simply the *line.* The unit of magnetic flux in the practical system is also the *maxwell.* (See *Line of Force, Alternator,* also *Flux, Magnetism.*)

MAGNETIC FLUX DENSITY—The number of lines of force per unit cross-sectional area of a magnetic path. The *gauss* (q.v.) is the unit of magnetic flux density. It is equal to one *line* (q.v.) per square centimeter.

MAGNETIC FORCE—The force with which a magnet attracts or repels any piece of iron or steel. The intensity of a magnetic field at any point as measured by the force which it would exert upon a unit magnetic pole if placed at that point.

MAGNETIC FORCE DUE TO CURRENT—A magnetizing force due to current flow which can be determined mathematically for any point in the surrounding medium. Consider any closed stream line of electric current surrounded by a medium having uniform magnetic properties. It can be demonstrated that each elementary length of the current stream line, dl, contributes to the magnetizing force H, at any point T an amount dH =

$$\frac{(\mathrm{I}\sin\Theta)\,dl}{r^2}$$

where I is the current flowing along this stream line, r is the distance from T to dl, and Θ (theta) is the angle between r and dl. The direction of dH is perpendicular to the plane of r and dl. The total magnetizing force at T is the vector sum of dH for all the elementary lengths into which the current stream line is divided.

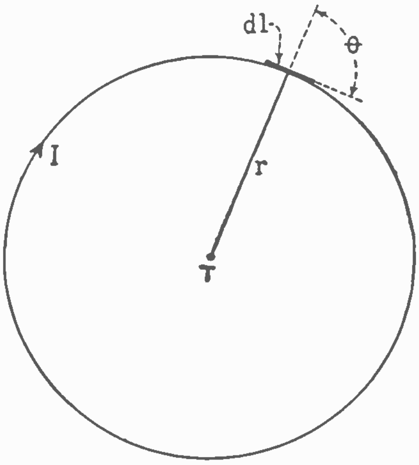

Diagram illustrating method of determining magnetic force due to current.

MAGNETIC HYSTERESIS—The opposition which minute particles of a magnetic material offer to being magnetized or to having their magnetization changed. It is sometimes referred to as molecular friction or as *magnetic lag.* (See *Hysteresis.*)

MAGNETIC INDUCTION—See *Induction,* also *Diamagnetic Material.*

MAGNETIC INDUCTIVE CAPACITY—symbol μ (mu)—Also termed Magnetic Inductivity—Names sometimes used instead of *Permeability* (q.v.). The ratio between the number of lines of force per unit area passing through a magnetizable substance and the magnetizing force which produces them. Stated another way, it is the ratio of flux density to magnetizing force. It can be considered as a measure of the ease with which magnetism passes through any substance. The magnetic inductive capacity of air is taken as unity. The term magnetic inductive capacity is analogous to *specific inductive capacity* (q.v.) which is the ratio of the dielectric constant of a material to that of a vacuum.

MAGNETIC INTERRUPTER—A device energized by an electromagnet for interrupting a circuit. (See *Interrupter.*)

MAGNETIC KEY—An electromagnetic adjunct to a wireless transmitting key for reducing or eliminating sparking due to the self-induction of the heavy currents which are interrupted. A form of magnetic key is shown in the illustration. Underneath the lever arm of the key is placed a spring to which an iron armature C and a platinum contact A_1 are attached. If the key and hence the spring are pressed down contact A_1 touches contact A_2 and the path of the current is completed. The current passes not only through the key, but also through the winding of the electromagnet. If the key is released so as to move upward, the contact A_1 still remains touching A_2, as the magnetic action due to the current continues to hold down the armature C and hence the spring with the contact A_1. Not until the alternating current reaches a zero value is the armature released. This releases A_1 from A_2, at an instant of zero current and hence there is no spark.

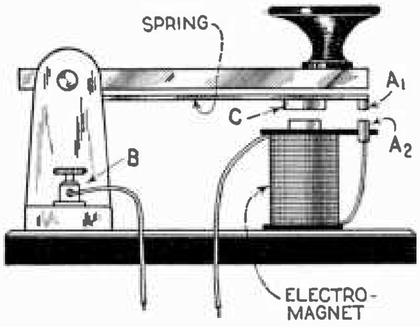

A magnetic key.

MAGNETIC LEAKAGE—Magnetic flux which serves no useful purpose. There are no substances which will prevent the passage of magnetic flux. This fact introduces a certain amount of difficulty into practical magnetic circuit calculations, since it is impossible to confine all the flux into a predetermined magnetic path. Since there is no magnetic insulator there will always be a certain amount of *leakage flux.* However this can be reduced to a minimum by proper design of the magnetic circuit. Magnetic Leakage is also referred to as *Stray Flux.*

MAGNETIC LINES OF FORCE—See *Line of Force,* also *Electromagnet.*

MAGNETIC MOMENT—The product of the strength of pole of a magnet and its virtual length. The torque exerted by a magnetic field upon a magnet depends upon the magnetic moment.

MAGNETIC NEEDLE—A light, thin steel magnet mounted on a pivot or suspended so as to be free to move. It takes a position pointing to the magnetic north.

MAGNETIC POLES—Points on a magnet where the lines of force leaving or entering the iron are concentrated. In a magnetic needle, that end which tends to point north is called a north or positive pole; that end which tends to point south is called a south or negative pole. A unit magnetic pole is defined as one which, when placed at a distance of one centimeter from a like pole of equal strength, will repel it with the force of one dyne. The north magnetic pole of the earth is situated in latitude 70 North and longitude 97 West. The south pole is located at latitude 70 South and longitude 102 East. It should be noted that the magnetic poles do not coincide with the geographical poles. (See *Induction, Magnetic.*)

MAGNETIC PROPERTIES—Materials are classified according to whether or not they offer a good path for the magnetic lines of force. Substances of comparatively low *reluctance* (q.v.) are termed *magnetic.* They are also often referred to as *ferro-magnetic.* They include iron, steel, nickel, cobalt, manganese, chromium, magnetite, certain alloys of copper, manganese and aluminum (called Heusler Alloys), and certain other oxides. The magnetic laws governing these substances can best be set forth by means of magnetization curves. Substances having high reluctance are termed *non-magnetic.* For commercial purposes, air is the most important non-magnetic substance. However, practically all substances, with the exception of the ferro-magnetic ones, follow the same magnetic laws as air. While the reluctance of a non-magnetic substance is constant, the reluctance of magnetic substances varies, as the amount of *flux* (q.v.) in the substance changes. Although air and all other substances, with the exception of the ferro-magnetic ones are classed as non-magnetic substances, and are poor conductors of magnetic flux, when compared to iron, etc., nevertheless they are by no means magnetic insulators. There are no substances which will absolutely prevent the passage of magnetic flux. (See *Ferro-Magnetic Substances.*)

MAGNETIC PYRITES—See *Pyrites.*

MAGNETIC STORM—Sudden and irregular variations of the earths magnetic field. Magnetic storms are sometimes coincident with the appearance of sun spots. These storms often disrupt telegraph service and also cause much interference with radio communication.

MAGNETISM—A phenomena exhibited by certain materials which is characterized by the attraction or repulsion of soft iron or of conductors carrying electricity. Magnetism is one of the manifestations of electricity, since a current flow is always associated or linked with a *magnetic field* (q.v.). (See *Free Magnetism, Magnet,* also *Magnetic Properties.*)

MAGNETISM, LAWS OF—Like magnetic poles repel one another; unlike magnetic poles attract one another. The force exerted between two magnetic poles is proportional to the product of their strengths and is inversely proportional to the square of the distance between them. (See *Like Poles.*)

MAGNETITE — symbol Fe_3O_4 — A mineral composed mainly of magnetic oxide of iron. When found in magnetized condition it is known as a *loadstone* (q.v.) or a natural magnet. It is a black brittle substance having a specific gravity of 5.2.

MAGNETIZATION—(also spelled *Magnetisation*)—The process of communicating magnetism. The act of rendering iron, steel, etc., magnetic. According to Ewing's theory, a substance which is magnetized has its molecules rearranged so that they assume symmetrical positions, each molecule lying in line with or parallel with its neighbor. In this rearranged position each molecule adds its separate magnetic force to every other one and the cumulative effect of this is that the substance exhibits the property of a magnet. (See *Demagnetization.*)

MAGNETO GENERATOR—A small generator consisting essentially of an armature rotating between permanent magnets, usually of the horse-shoe type. Magnetos are mainly used for motor ignition, medical coils, bell ringing and for firing high explosives. (See *Intermittent Current.*)

MAGNETOMOTIVE FORCE OR DIFFERENCE OF MAGNETIC POTENTIAL—abbreviation m.m.f.—The force causing magnetic flow. *Magnetomotive force* is the magnetic cause, while *flux* (q.v.) is the magnetic effect, just as in electrical terms, *electromotive force* (q.v.) is the cause and current is the effect. The c.g.s. (centimeter gram second) unit of magnetomotive force is the gilbert (q.v.). The unit ordinarily employed in practical calculations of the magnetic circuit is the *ampere-turn* (q.v.). (One ampere-turn is equal to 1.2566 gilberts.) An ampere-turn consists of a current of one ampere flowing through one complete turn of a conductor. The expression "turn" means that the conductor is wound in a solenoid. That is to say, it is wrapped around a core which may consist of air or of any other substance. The same magnetic field will result from ten ampere-turns consisting of ten amperes flowing through one turn, as from ten ampere-turns consisting of one ampere flowing through ten turns. (See *Force, Magnetomotive.*)

MAGNIFIER, OR RELAY—An apparatus for intensifying the variations of a current. The term *magnifier* has been incorrectly used for *amplifier* (q.v.). One type of *magnifier* used in radio is the microphonic relay (see loud speaker) by means of which the minute currents in a crystal set can be magnified so as to actuate a loud speaker. In the ordinary telephone repeater, a microphone is acted upon by the receiver of the first line circuit so as to introduce a local circuit which acts inductiveley on the second line wire. Relays are also used in wire telegraphy to boost weak line currents by means of strong local batteries. (See *Relay.*)

MANTISSA — The decimal part of a *logarithm* as differentiated from the integral part or characteristic. For example in the logarithm 1.23578, 1 is the characteristic and the decimal .23578 is the mantissa. (See *Logarithms.*)

MARBLE—The name given to any limestone sufficiently compact to permit being polished. Pure marble is white, the presence of oxides giving the different colors. Marble is used extensively for switchboard work and when so used must be free from metallic veins. Where circuits carrying 1000 volts or more are used, the switchboard should be saturated with an insulating varnish and baked. Inasmuch as the marble shows oil spots, it is sometimes stained black and given a finish called a *marine finish*.

MARCHANT, EDGAR WALFORD. British radio expert. Born in 1876, he was educated at University School, Hastings, and Central Technical College and London University. In 1897 he was appointed superintendent of the laboratory and workshops of Lord Blythswood at Renfrew, where he carried out many experiments in wireless telegraphy. He became chief assistant at Finsbury Technical College, 1900, under Professor Silvanus P. Thompson, and the following year was appointed lecturer in electro-technic at University College, Liverpool. In 1903 he was appointed the first professor of electrical engineering at the university.

Marchant was for some time closely associated with W. Duddell in developing the oscillograph, and the two read a joint paper on the study of the electric arc by the aid of oscillographs to the Institution of Electrical Engineers, of which institution Marchant became vice-president. He is also a vice-president of the Radio Society of Great Britain. He has written many papers on radio, and contributed an article to the proceedings of the Royal Society on the magnetic behaviour of iron under oscillatory discharge of a condenser.

MARCONI FILINGS COHERER—Also called Marconi Coherer—A detector of electric waves consisting of fine nickel and silver filings contained in a suitable glass tube between two electrodes consisting of two silver plugs. These two plugs are connected to platinum wires brought out through the sealed ends of the tube, which is carefully exhausted of its air. The metallic filings in a loose condition have feeble conductivity for the current from a single dry cell. If a very weak electric oscillation is passed through the coherer, the filings cling together or cohere and hence conduct electricity better and they will then pass sufficient current from a single dry cell to operate a telegraphic relay. To bring these filings back to a non-conductive condition, the tube is tapped by an electromagnetic tapper, similar to an electric bell with the gong removed. The Marconi coherer is now obsolete. (See *Coherer.*)

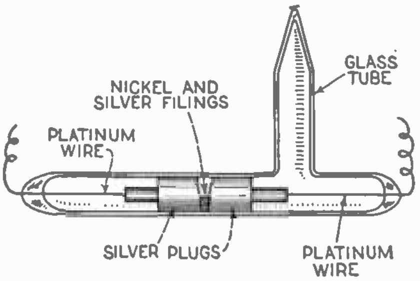

The Marconi filings coherer.

MARCONI, GUGLIELMO—Italian-Irish radio pioneer. Guglielmo Marconi was born at Bologna, Italy, April 25, 1874, and was educated at the Leghorn Technical School, under Professor Rosa. From his earliest years he was keenly interested in communication by means of Hertzian waves, and began his brilliantly successful series of experiments in June, 1895. These first experiments were carried out on his father's estate near Bologna. Marconi soon found that the Hertzian form of resonator gave only feeble signals at a distance, and he substituted a vertical wire, with the result that in 1895 he was able to transmit signals to a distance of one and a half miles. About the same time he improved the Branly coherer, which he was using as a detector, and invented an electric tapping device to decohere the filings.

This early apparatus of Marconi, from which has sprung the far-flung wireless chain of to-day, consisted of a coherer, a relay, a decoherer, and a Morse printing instrument, all working from storage batteries. Between the coherer and the relay Marconi interposed choke coils, and this had a very marked effect on the receptivity of his set. By close attention to the details of his system he was able to carry on his signals at greater ranges than had up to this time been accomplished by other experimenters.

The transmitting apparatus used by Marconi in these early efforts consisted of a large spark gap to which the aerial and earth wires were connected. The high-tension current for the spark was provided through an induction coil from batteries. The spark gap consisted of a ball discharger comprising four brass balls. The two middle balls were separated by a small space filled with vaseline oil, the actual spark jumping from the two end balls to the middle ones and through the vaseline, producing a high-frequency spark.

Marconi came to England in 1896, where he took out the first patent ever granted for a practical system of wireless telegraphy. He made a number of experiments at Westbourne Park, and demonstrated his system before Sir William Preece and other high officials at the Post Office. In the following year he increased the range of his set to nine miles using a 29 in. spark coil and kites to raise the vertical aerials. In July, 1897, in demonstrating before the Italian Government, he covered 12 miles between warships, and he began to instal a number of his sets for lighthouses, and the success of his experiments led to stations being erected for the corporation of Trinity House.

Guglielmo Marconi.

In 1899, the first proof of the advantages of wireless over other forms of communications came with the saving of the lives on board the ship R. F. Matthews, which ran into the East

Goodwin lightship. The latter was equipped with one of Marconi's transmitting sets and was able to communicate with the South Foreland lighthouse and summon assistance.

The genius of Marconi lies not so much in his inventions, as in his far-sightedness in being the first man to realize the immense commercial possibilities of radio, and to make the best use of all the scientific effort of his time to further the object he had in view.

Marconi has received innumerable honors from all countries. He holds the honorary degrees of universities all over the world, and the freedom of a large number of cities. In 1909 with Professor Braun, he was awarded the Nobel Prize for Physics, and in 1914 he was given the Honorary Knighthood of the Grand Cross of the Victorian Order. He has been awarded the Albert Medal of the Royal Society of Arts, the Gold Medal of the Institute of Radio Engineers of New York, the John Fritz Gold Medal for the invention of wireless telegraphy, and the Franklin Gold Medal of the Franklin Institute. He was the Italian delegate to the Peace Conference, and signed on behalf of Italy. He is chairman of the board of directors of the Marconi Company.

MARINE TYPE CHARGING PANEL—A switchboard equipped with resistance coils, carbon filament pilot lights, voltmeter and necessary switches. Manipulation of the switches allows the battery to be placed on charge or on the line. By throwing a switch on this board, it is also possible to work the wireless transmitter either from the ship's generator or from the storage battery.

MASS—The quantity of matter which a body contains. Physical quantities such as force, velocity, etc., are expressed in terms of length, *mass*, and time. The unit of length is the centimeter; of mass, the gram; and of time, the second. Hence this system of measurement is known as the C.G.S. (centimeter gram second) system. The gram is equal to 15.432 grains and represents the mass or quantity of a cubic centimeter of water at 4 degrees centigrade.

MASTER-OSCILLATOR SYSTEM—A radio telephone transmitting circuit in which a small extra *audion* (q.v.) is used to supply the power to excite the grid circuit. This extra oscillator or exciter is referred to as the master-oscillator. The master-oscillator system has several advantages over the self-excited system. Changes in antenna constants, such as might be caused by a swaying aerial or *lead-in* (q.v.) do not affect the transmission where the former system is used. In addition the master-oscillator system is more convenient to work with and the adjustment for maximum output for different wave lengths and antenna resistances is more easily made.

MASTER-OSCILLATOR SYSTEM AMPLIFIER—In the master-oscillator system, the power tube acts as an amplifier of the power supplied by the small extra exciting, or master-oscillator. The exciter must develop sufficient power to supply the losses in its own oscillating circuit and those of the grid circuit of the power tube.

MASTS, ERECTION OF—High powered stations generally use latticed steel masts. In some cases tubular metal masts in telescoping sections are used. Wooden masts have been used to quite an extent. For portable outfits, these are made in sections which can be fitted together like a fishing rod. Guy ropes or wires are necessary in practically every instance. In erecting a mast, it is usual not to make the installation too rigid, but rather to allow some freedom of movement so that the mast may sway slightly in the wind. Where a one-wire aerial is being erected this may be fastened to any support available. In marine construction, the ship's masts are used to support the aerial, the wires being stretched between the two booms or spreaders. from which halyards run to the masts. (See *Aerial.*)

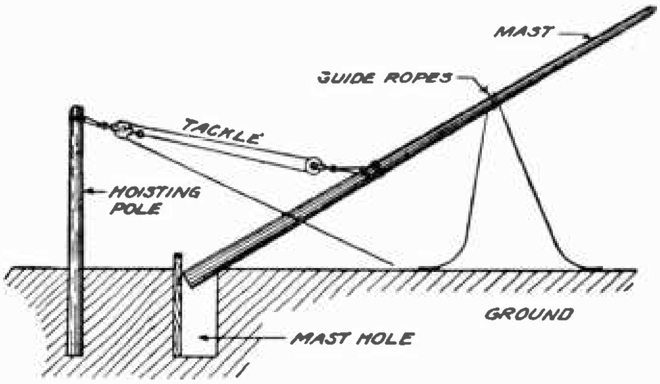

A method of erecting a tall mast using tackle and hoisting pole.

MATCHED IMPEDANCES—Circuits balanced or matched so that the total impedance of one circuit will equal the impedance of another circuit with which it is coupled. (See *Impedances, Matched.*)

MAXWELL—symbol φ—The unit of *magnetic flux* (q.v.). It is also known as the *line* (q.v.). One maxwell is equal to one line of force. (See *Flux Density*, also *Kiloline.*)

MAXWELL'S CORKSCREW RULE—If the direction of travel of a right-handed corkscrew represents the direction of current in a straight conductor, the direction of rotation of the corkscrew will represent the direction of the magnetic lines of force.

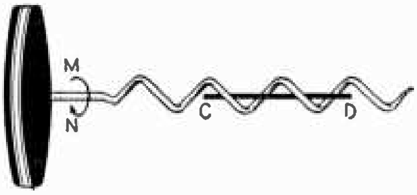

Illustration of Maxwell's corkscrew rule. When current flows from C to D, magnetic lines encircle the wire in the direction of the curved arrow MN.

MAXWELL'S LAWS—The magnetic fields surrounding parallel currents in the same direction will react upon each other in such a way that the conductors will tend to move together. In the case of parallel currents in opposite directions, the fields between the two conductors will be in the same direction and will merge together, thus tending to push the conductors apart. The magnetic fields around two conductors at an angle will react upon each other in such a way as to tend to bring the conductors parallel, and with the currents flowing in the same direction.

MAXWELL'S LAW OF MOTOR ACTION—A conductor carrying a current whose direction of flow is at right angles to a magnetic field, will tend to move out of this field.

MAXWELL'S RULE—Every portion of a circuit is acted upon by a force urging it in such a direction as to make it enclose the greatest possible number of lines of force.

MAXWELL TURNS—The deflection value in a ballistic galvanometer or *fluxmeter* (q.v.). This depends upon the number of turns in the exploring coil and the instantaneous value of the flux in maxwells linked with this coil. (See *Grassot Fluxmeter.*)

McLACHLAN, NORMAN W.—British radio authority. Born at Longtown, Cumberland in 1888, and educated at Carlisle Grammar School and the George Watson and the Heriot Watt Colleges, Edinburgh, and Liverpool University. He was appointed lecturer in Engineering and Mathematics at Newcastle-on-Tyne in 1909. In 1913 McLachlan was appointed supervisor of classes in engineering subjects in the Liverpool Technical Institutes, and after the Great War, during which he was engaged in aeronautical research and the study of anti-submarine devices, he made a special study of magnetos at the National Physical Laboratory, Teddington. Appointed research engineer to the Marconi Company, he is the author of many papers on wireless and electrical subjects in the journal of the Institution of Electrical Engineers and other scientific journals, including those on Characteristic Curves of a Poulsen Arc., the Magnetic Behaviour of Iron at very High Frequencies; and Theory of Iron-cored High Frequency Transformers. He has taken out a number of radio patents.

MEASUREMENT OF INDUCTANCE—Some methods of measuring inductance (also applicable to the measurement of *capacity*) are various bridge methods; by measuring the voltage, current, power, and frequency; and by the wave meter method. Connections in one bridge method for measuring inductance are similar to those in the ordinary *Wheatstone Bridge* (q.v.). A coil having a known inductance is balanced against the inductance to be measured. A device called a *Secohmmeter* is in use to increase the sensibility of bridge measurement of inductance. The Secohmmeter serves the purpose of making an alternating current to use in measurements of self-induction and of commutating such portion of this current as flows in the galvanometer circuit to a direct current. When a source of alternating current of known frequency is available the following method is convenient. Place the inductance across the alternating current mains, measure the power absorbed, the current flowing, the voltage across the unknown inductance and the frequency of the circuit. Then the inductance L is equal to the reactance, x, divided by

2 π times the frequency. Expressed as a formula

$$L = \frac{x}{2\pi f} = \frac{\sqrt{\left(\frac{E}{I}\right)^2 - \left(\frac{\text{Power}}{I}\right)^2}}{2\pi f}$$

The inductance L should be expressed in henries. This method gives results of fair accuracy but is not adapted to measurements of low values of inductance. Small inductances are more usually measured by bridge methods. In the wave meter method, a condenser is discharged through a spark gap in a circuit containing a known capacity and by means of a wave meter measuring the wave length. The expression for the number of oscillations in a closed circuit containing both inductance and capacity is given by the formula

$$n = \frac{1}{2\pi\sqrt{CL}}$$

and the wave length is equal to the velocity of propogation divided by the number of waves and calling λ the wave length we have

$$\lambda = V \text{ times } 2\pi\sqrt{CL}$$

In this last equation everything is known but the inductance which can be calculated. (See *Inductance*, Measurement of.)

MEASUREMENT OF WAVE LENGTH—The usual method of measuring an unknown wave length is by means of a wavemeter. Other apparatus required in this connection is a hot wire ammeter indicator in a coupled circuit, or crystal detector and telephones. Also an antenna excited oscillating circuit supplied by electron tube. A wavemeter placed in inductive coupling with a coil or antenna carrying radio current will show decided increase of current in its own coil and condenser when it is tuned to resonance with the source. The wave length is read directly from the wavemeter setting for resonance or from a calibration curve. A receiving set can be used to measure the wave lengths of received waves if it is first standardized in terms of wave lengths. The operator listens in through the head set attached to the receiving set. As the wavemeter condenser knob is turned the loudest sound is heard when the wavemeter circuit is tuned to the same wave length as that for which the receiving set is adjusted. The wave length is then read from the wavemeter scale or calibration curve. Continuing in this manner, the receiving circuit can be calibrated as a wavemeter by setting it at many different adjustments and reading the wave lengths at resonance each time. The wave length of the waves emitted by an antenna, when no added inductance or capacity is inserted in the antenna circuit, is known as its fundamental wave length. For a single vertical wire grounded antenna the fundamental wave length is slightly greater than four times the length of the wire. The approximate constant often used is 4 2 and this applies also to "L" or "T" type aerials with vertical lead in wire, the total length being measured from the transmitting apparatus up the lead-in wire and over to the end of the flat top. It is of course, more accurate to measure the wave length radiated from an antenna directly by use of a wavemeter. The wavemeter coil needs merely to be brought near the antenna or lead-in wire and the condenser of the wavemeter adjusted to give maximum current in the wavemeter indicator. The wave length corresponding to the wavemeter setting is then the length of the waves radiated by the antenna. The fundamental wave length of the antenna may be determined by gradually decreasing the number of turns in the *loading inductance* (q.v.), measuring the wave length for each setting of the loading coil and plotting a curve showing the wave length corresponding to the various numbers of turns of the loading coil. The fundamental, then, is the wave length corresponding to zero turns and is found at the point where the extension of the curve cuts the wave length axis. (See *Measurement of Inductance*, also *Formula*.)

MECHANICAL MIXTURE—A mixture of two or more elements which leaves them essentially the same. This is sometimes incorrectly called a mechanical compound. Sugar and sand mixed together give an example of a mechanical mixture. A *chemical compound* is formed when two or more substances combine chemically to form another substance having different characteristics. Thus iron filings and sulphur can be combined to form a third substance unlike either of the original two.

MEG OR MEGA—symbol M—The prefix placed before the name of a unit to denote one million times (10^6) that unit, as for example *megohm* (q.v.) *mega-volt*, *megerg*, etc.

MEGGER—An instrument for measuring insulation resistance or in general any resistance by the direct deflection of a pointer on a scale. The "Evershed" type is made in two forms (1) the *megger* insulation testing set and (2) the bridge-megger testing set which combines the functions of type (1) with those of a *Wheatstone Bridge* (q.v.). Both types consist of an ohmmeter of the moving coil type combined in one box with a hand driven generator for providing the necessary testing voltage and current. When the handle of the generator is not being turned, the pointer is entirely free and will rest anywhere on the scale. The general principles underlying the construction and operation of the megger are shown in the illustration. As can be seen, the megger is a combination of a *magneto generator* (q.v.) shown at the right, with an ohmmeter shown at the left, in a magnetic circuit common to both and consisting of two pairs of field poles braced by strong bar magnets, NS, NS, and forming two bi-polar field magnets in series. In the right hand one, and rotated by a folding handle and spur gearing D, is the armature of the generator, with its brush gear B, and terminal bars marked + and —. In the left hand field is the current coil A, pressure coil P, and compensating coil C of the ohmmeter, connected to resistances Q and R, a *guard plate* G, and the only two external terminals L and E (marked Line and Earth on the actual instrument). The bridge-megger testing set is available for use both as an insulation testing set and as a specialized Wheatstone Bridge. It differs from the ordinary megger in outward appearance only by the addition of two pairs of terminals and of two switches, one of which is the *ratio* switch of the Wheatstone Bridge and the other one is a two-way change over-switch. This latter when set to "Megger" prepares the instrument for measuring insulation resistance or high metallic resistance by coupling the two windings of the constant voltage generator in series and making the two front terminals, marked Line and Earth, the only two available for connection. When the change-over switch is set to "Bridge" the instrument is ready for Wheatstone Bridge work, the two windings of the generator now being connected in parallel in order to increase the current obtainable from it. In this case the ohmmeter part is changed into a galvanometer for the bridge and the arms of the bridge are switched into their appropriate places in the circuit and connected to the terminals marked R and X at the end.

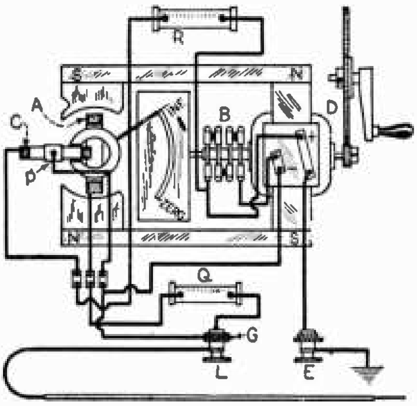

A diagrammatic illustration of a megger.

MEGOHM—One million ohms. This unit is usually employed in the measurement of insulation resistance. It is also used in measuring *grid leaks* (q.v.), etc. See *Ohm*, also *Begohm.*)

MEGOHMITE, also spelled Megohmit—An insulating material similar to micanite, obtainable in several forms, one of which is known as *hard megohmite* and is made of thin sheets of mica built up of shellac. *Flexible megohmite* consists of sheets of mica stuck together by vegetable adhesives. Flexible megohmite covered with Japanese paper is called mica paper while flexible megohmite covered with linen, is called *mica linen.*

MEISSNER, ALEXANDER.—Austrian radio expert. Born at Vienna, 1883, he was educated at the Technical High School and University of Vienna. He joined the Telefunken Company in Berlin, 1907, and is one of the leading authorities in Germany on radio. To him is due a large number of wireless inventions, including the Telefunken compass, musical quenched sparks, interference preventers, the direct current cathode valve relay for the reception of Morse Code, etc. In 1913 he invented the generation of oscillations by means of the three-electrode valve. He holds many patents, including spark gap for impulse excitation, quenched-spark signalling in conjunction with G. Von Arco, etc., and he is the author of many papers for scientific journals and societies on the subject.

MERCURY ARC RECTIFIER (MERCURY VAPOR CONVERTER)—A device for changing alternating current to direct current, which depends for its action upon the unidirectional conductivity of the mercury arc in an exhausted vessel. An idea of the principle of operation of the mercury arc rectifier may be obtained from the illustration. Assume an instant when the terminal H of the supply transformer is positive, the anode A is then positive and the arc is free to flow between A and B, B being the mercury cathode. Following the direction of the arrows still further, the current passes through the lead J, through the reactance coil E and back to the negative terminal G on the transformer. A little later when the impressed electromotive force falls below a value sufficient to maintain the arc against the counter electromotive force of the arc and the lead. the reactance E which heretofore has been charging, now discharges, the discharge current being in the same direction as formerly. This serves to maintain the arc in the rectifier until the electromotive force of the supply has passed through zero, reverses and builds up to such a value as to cause A′ to have a sufficiently positive value to start an arc between it and the mercury cathode B. The discharge circuit of the reactance coil E is now through the arc A′B, instead of through its former circuit. Consequently the arc A′B is now supplied with current, partly from the transformer and partly from the reactance coil E. The new circuit from the transformer is indicated by the arrows enclosed in circles. The amount of reactance inserted in the circuit reduces the pulsations of the direct current sufficiently for all ordinary commercial purposes. The mercury rectifier is used in small sizes for charging storage batteries from alternating current circuits and in the larger sizes for operating direct current arc lamps from alternating circuits and also in electric traction work in the place of rotary converters in the substations or even on the train itself.

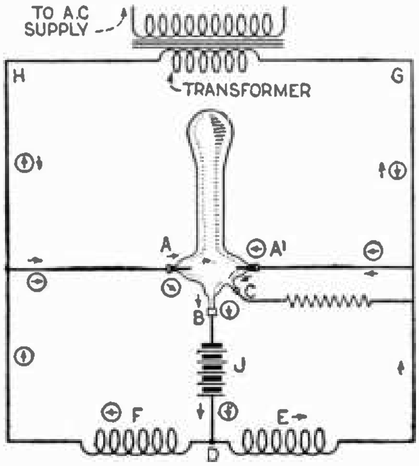

Illustration showing action of mercury arc rectifier.

MERCURY INTERRUPTER (MERCURY JET INTERRUPTER)—A contact breaker in which contact is made between moving metallic contacts and a jet of mercury provided by a small centrifugal pump driven by the same motor as the contacts. When the primary winding of an induction coil is supplied with a continuous current it is necessary to interrupt the current in order that a secondary voltage shall be induced in the secondary winding. The mercury interrupter offers a satisfactory method of accomplishing this. In one type of mercury interrupter a stream of mercury is forced from a jet against a metal plate. The stream of mercury is interrupted by a toothed wheel of insulating material so that the electrical circuit is periodically made and broken between the mercury and the metal plate. In some forms of mercury interrupters, the jet itself revolves, the mercury impinging against a fixed metal plate. The mercury interrupter is used in operating X-ray bulbs from an induction coil where the source is direct current. A form of mercury interrupter is shown in the illustration. A is a tooth-shaped wheel made of insulating material. This wheel is driven by a small motor and the number of revolutions can be varied within wide limits. A small centrifugal pump forces the mercury upon the jet B, while the revolving wheel A interrupts the contact of the mercury with the plate C. The wheel A can be raised or lowered, thereby enabling the duration of contact of the mercury with the plate to be varied without altering the number of interruptions in a given time.

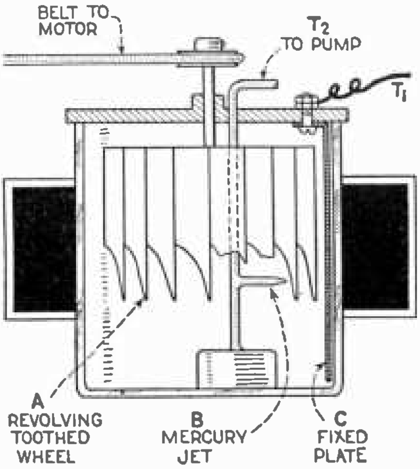

A type of mercury interrupter.

MESH GROUPING—A method of connecting coils in a polyphase circuit whereby they form a closed circuit and have the line wires attached to the points of junction between the coils. The term mesh is used interchangeably with delta. (See *Delta Connection.*)

METER—An instrument used to make measurements. (See *ammeter* or *amperemeter, double range meters, integrating wattmeter, voltmeter, wavemeter, wattmeter.)*

METER (*Measure*)—A unit of length in the *metric system* (q.v.). One meter is equal to 100 centimeters or 39.37 inches. It is equal to the length of a standard platinum bar, kept in Paris, and representing approximately a ten-millionth part of a quadrant of the earth's meridian measured from the equator to the pole through Paris. Wireless waves are measured in meters. The length of a wave is the distance, usually measured in meters between two points in the successive waves where the disturbance is at a maximum or at a minimum or between any two points of equal disturbance. There is a fixed relationship between the length of a single wave, the frequency of oscillation and the velocity of the wave. Thus if 3,000,000 waves pass a given point in one second, each wave must be 100 meters long, since the velocity of radio waves is equal to 300,000,000 meters per second.

METER-AMPERES—The product of the antenna current in amperes at the point of maximum current and the antenna effective height in meters, for any radio transmitting station. It constitutes a factor for indicating the radiating strength of radio transmitting stations.

METRIC SYSTEM—A system of measurement in which the meter is the fundamental unit of length and in which all the units both fundamental and derived are divided decimally and higher units are formed in multiples of ten. (See *Centimeter Gram Second.*)

MFD—Abbreviation for *Microfarad* (q.v.)

MICA—An anhydrous silicate of aluminum and potash or sodium. Mica has very high insulating qualities and can withstand high temperatures. Mica obtained in the natural state is separated into laminations which are sorted and graded. These are then cemented together to form plate or flexible mica of any desired thickness or purity. Mica is used in radio work as the dielectric of small fixed condensers. It is also used as the sound producing diaphragm in certain types of head sets and *loud speakers* (q.v.).

MICABOND CLOTH—India sheet mica faced on one side with muslin and on the other side with Japanese insulating paper bonded together with a special binder. It is used for insulating field coils and transformer coils.

MICABOND PAPER—India sheet mica faced on one side with Japanese insulating paper of one grade and on the other side with paper of a different grade.

MICABOND PLATE—Mica sheets bonded together with orange shellac. Various grades of plate are made some of which are used for magnet spools, others for insulating of commutator segments, while still others are used in electric irons.

MICANITE—A form of reconstructed mica. Canadian amber mica and white India mica are used.

MICRO—A prefix placed before the name of a unit to denote one-millionth part of that unit, as for example *microfarad* (q.v.).

MICRO-AMPERE—One-millionth of an ampere. A unit used in measuring extremely small currents. (See *Ampere.)*

MICROFARAD—abbreviation Mfd.—A unit of electrical capacity equal to one-millionth of a farad. Since the farad is too large for most practical measurements, the microfarad is generally used. (See *Farad,* also *Jar.)*

MICROHM—A unit of resistance equal to one-millionth of an *ohm* (q.v.).

MICROMETER—An instrument for making minute measurements. Usually controlled by an accurate screw of fine pitch.

MICROMETER SPARK GAP—A minute adjustable spark gap of about four-hundredths of an inch placed in the aerial circuit of a multiple tuner to allow heavy charges to pass readily to earth by sparking across its points.

Courtesy of Western Electric Co.
An example of the microphone used in modern broadcasting.

MICROPHONE—A sound magnifier. The ordinary form of microphone consists essentially of a diaphragm set in vibration by the sound waves and causing by its motion variation in the resistance of a mass of loosely packed carbon granules. This variation in resistance causes corresponding variations in the current through the instrument so that at every instant the form of the current corresponds to the variations in sound. This type of microphone is often referred to as a *Carbon Microphone*. The telephone transmitter is a form of microphone, and the microphone (nicknamed *Mike*) used in radio broadcasting is merely a larger and more sensitive variation of the same instrument. (See *Electro-magnetic Microphone.*)

MICROPHONE, CARBON—See *Microphone*, also *Granular Carbon*.

MICROPHONE TRANSMITTER—See *Microphone*.

MICROPHONIC JOINT—A loose contact between two solid conductors, so mobile that feeble vibrations vary its resistance.

MICROPHONIC RELAY—An instrument for magnifying the variations of a given current by passing it through an electromagnet that acts on the diaphragm or reed of a microphone. The resistance variations are repeated in the current produced by a constant voltage in a circuit in series with the *microphone* (q.v.). In general a microphonic relay can be described as a microphone combined with a telephone so that a message transmitted over the telephone is repeated by the microphone over another line.

MIL—A unit of length equal to one-thousandth part of an inch.

MIL-FOOT—A wire one foot long having a diameter of one mil. Unit resistance may be measured in *ohms* (q.v.) per mil-foot. This refers to a conductor having a cross-sectional area of a *circular mil* (q.v.) and a length of one foot.

MILLI—The prefix placed before the name of a unit to indicate one-thousandth part of that unit, as for example *milli-ampere* (q.v.).

MILLI-AMPERE—One-thousandth of an *ampere* (q.v.). Sometimes abbreviated to *milli-amp*.

MILLI-MICRO—A prefix denoting one thousand millions. (10^{-9}). The term *Billi* is an alternate prefix.

MILLI-VOLT—One-thousandth of a volt.

MILS, CIRCULAR—See *Circular Mils*.

MIRROR GALVANOMETER—A galvanometer for measuring very small currents which utilizes a reflected beam of light for a pointer. In one type of mirror galvanometer a fixed nearly circular magnet is used and a coil of wire is suspended in its field. When the current flows through the coil it tends to turn the coil so as to set its axis in the direction of the magnetic force. This movement can be observed by means of a beam of light reflected from a mirror fixed to the coil. (See *Galvanometer.*)

MINERALLAC—An insulating compound made in several different grades all of which are impervious to water and show no alkali or acid reactions.

M.M.F.—Abbreviation for *Magnetomotive force* (q.v.).

MODULATED CURRENTS—Currents which have their amplitude periodically varied as in the case of the carrier current used in radio telephony which is varied or modulated in accordance with the vibrations of a microphone. If a sound wave actuates a microphone, its inward and outward displacement, varying the resistance in the aerial circuit, results in a high frequency current in the aerial of variable amplitude, called a modulated high frequency current. The illustration shows an extremely simplified diagram of the apparatus in a broadcasting station. The initial impulse that starts a radio signal towards the receiving set is furnished by the voice of the speaker or singer or by some other sound originating in the studio. This sound strikes the microphone, the operating element of which is a thin metal disk or diaphragm. The sound waves cause this diaphragm to vibrate. In doing so, the disk causes changes to take place in the electrical circuit. Pulsations of current are set up, the pulsations corresponding in strength to the variations of the spoken word, music or other sound as the case may be. These fluctuations are quite weak when they first come from the microphone and therefore they must be strengthened. They go through what is known as a voice amplifier which utilizes vacuum tubes to make the current stronger. The next piece of apparatus to consider is the *oscillator* (q.v.). This consists of one or more vacuum tubes, connected in a circuit of such a type that the tubes, when lighted and properly furnished with a high voltage direct current, will generate another current, alternating in character, which is said to be oscillating at radio frequency. Upon this current so generated, the voice current is impressed or super-imposed. This process is called *modulation* (q.v.). The current from the oscillator is known as the *carrier wave* (q.v.) current and when voice currents are impressed upon it, it is known as modulated radio frequency current. This current flows in the aerial and there sets up radio waves.

MODULATION—Variation of current or wave form to conform to sound waves or to other predetermined forms. Various forms of modulation are dot and dash modulation, chopper modulation, buzzer modulation and *speech modulation* (q.v.). (See *Distortion, Double Modulation*, also *Modulation Frequency Ratio.*)

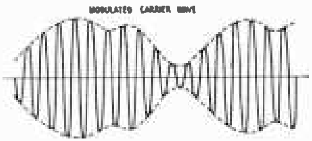

Graph of a modulated carrier wave.

MODULATION, DOUBLE—See *Double Modulation*.

MODULATION FREQUENCY RATIO—The ratio of modulation frequency to wave frequency. An alternating current is said to be *modulated* when the amplitude of its oscillations is varied periodically. The frequency at which the variations occur, that is to say the *modulation frequency* is necessarily less than the frequency of the alternating current which is being modulated. The nature of the variations may assume practically any form. As examples, there are dot and dash modulation, chopper modulation, buzzer modulation and *speech modulation* (q.v.).

MOLECULE—The smallest group of *atoms* of an element or compound which can exist by themselves. A familiar analegy states that if a drop of water could be magnified to the size of the earth, its component molecules would be the size of base-balls.

MOLYBDENITE—symbol MoS_2—A lead gray sulphide of Molybdenum. It is used as a rectifier detector in contact with copper.

MOLYBDENUM—Symbol Mo—A silver white metal found in nature as *molybdenite* (q.v.), a sulphide of molybdenum. When pure this metal is ductile and can be forged. It is used chiefly as an alloy to produce extremely hard steel. Molybdenite is used as a *crystal* (q.v.) in the detection of radio waves.

MORSE LIGHT—A search light used to signal code messages by means of intermittent flashes. (See *Code.*)

MORSE INKER—An instrument for recording Morse code signals in the form of dots and dashes on a moving paper band. In one form of Morse recorder, the attraction of the armature raises an inked wheel against the paper ribbon and thereby prints upon it the dots and dashes of the message.

MOSCICKI CONDENSER—A condenser in which the dielectric thickens out at the edges. It is essentially a form of *Leyden Jar* (q.v.) composed of a glass tube, especially thickened at the neck where the dielectric stress is greatest, coated inside and out with metal foil. The design of this condenser is such as to minimize *brush discharge* (q.v.) or *corona* (q.v.) effect at its edges.

MOTOR—A device for transforming electrical energy into mechanical energy. Motors may be designed for operation on direct current, on alternating current or on either direct or alternating current. These latter are known as *universal* motors. Direct current motors are classified as shunt, series and compound. This classification is arrived at according to the manner in which the field windings are connected. In the shunt motor the field winding is connected in parallel with the brushes (that is with the armature winding). In the series motor the field winding is connected in series with the armature and the line. Compound motors have a shunt winding in parallel with the brushes and a series winding in series with the armature and the line. Compound motors may be further classified as differentially compounded and cumulatively compounded. A differentially compounded motor is one in which the series field winding is so connected as to oppose the shunt field. A cumulatively compounded motor is one in which the series field aids the shunt field. The speed of the shunt motor is nearly constant, falling slightly as the load increases. The speed of the series motor decreases as the load increases. The differential motor may be constructed so that an absolutely constant speed will be maintained at all loads. The speed of the cumulative motor decreases rapidly as the load is increased. Alternating current motors are of the series type, the induction type, commutator type, synchronous type and repulsion type. There are a number of special classifications of alternating current motors such as single phase induction, polyphase induction, repulsion induction, etc. An ordinary direct current series motor used on alternating current will operate but at a low power factor, with a large eddy current loss and with violent sparking at the commutator. In order to adapt the series motor for universal use, it must be specially designed, the field structure being *laminated* (q.v.) in order to avoid *eddy currents* (q.v.) and the field coils having only a few turns to avoid too great *self-inductance* (q.v.) and consequently low *power factor* (q.v.). Sparking at the brushes can be avoided to a certain extent by designing the motor for a small field flux and with but a few turns in series in each armature coil.

MOTOR-ALTERNATOR—A *motor generator* (q.v.) set consisting of an *alternator* (q.v.) driven by a direct or alternating current motor. Such a set may be used for converting from direct to alternating current, or for frequency changing.

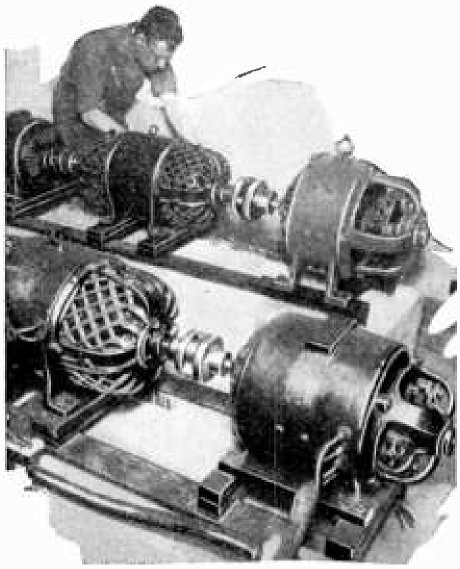

Motor Alternator set installed at Station WRNY.

MOTOR-GENERATOR—A generator or generators driven by an electric motor. Motor generators are used for converting alternating current to direct current or vice versa, for changing from one voltage or from one frequency to another or to obtain a variable from a fixed voltage.

MOVABLE COIL—Inductance coils which can be moved or rotated so as to vary the inductance. Examples of movable coils used in radio are vario couplers, loose couplers, etc. (See *Coupler, Loose; Coupler, Vario,* also *Fixed Coils.*)

MULLARD, S. R.—British radio authority. Born in 1884, he was educated as an electrical engineer and was ultimately appointed head of the research laboratory of Edison & Swan, Ltd., where he developed the Pointolite arc lamp. During the World War, he served in the Royal Air Force and became head of the wireless section laboratory at the Imperial College of Science for the Air Ministry. He carried out a series of researches on vacuum tubes and in 1920 founded the Mullard Radio Valve Company.

MULTIPLE—As applied to an electric circuit, a divided circuit. Two conductors are said to be connected in multiple or *parallel* (q.v.) when their two ends are joined together.

MULTIPLE SERIES—See *Series Parallel.*

MULTIPLIER—A resistance placed in series with a voltmeter for limiting and controlling the amount of current flowing through the voltmeter windings. This resistance may be either internal (enclosed by the voltmeter case) or external.

MULTI-POLAR—Having more than two poles. Usually a generator or motor whose field magnet has more than two poles. Small machines may be made with two poles, *bi-polar* (q.v.), but all others have a number of poles, that is to say they are *multi-polar.* The use of multi-poles results in a lighter machine and incidentally there is a saving of field copper and a reduction in sparking. If there are too many poles, however, there will be excessive *leakage flux* between the poles. (*See Homo-polar.*)

MUSICAL SPARK—A spark giving a regular distinct musical note. This may be produced by a high speed rotary discharger, a *quenched gap* (q.v.) or an arc.

MUSICAL SPARK SIGNALS—Signals in which the sparks occur at regular intervals of time and fast enough to give a musical note. Usually the spark rate is between 100 to 1200 discharges per second.

MUTE ANTENNA—A local circuit or resistor used in testing transmitting apparatus. The mute antenna in this case is substituted in the place of the actual antenna. It is also referred to as a *dummy aerial, phantom antenna, mock antenna,* or *artificial antenna.*

MUTUAL CAPACITY—The capacity effect of one conductor upon another one in the same electrostatic field. The mutual capacity is not the same as the capacity of the two wires regarded as the two plates of a condenser, one positively charged while the other is charged negatively. It really represents a decrease in the capacity of one of the wires with respect to earth caused by the presence of the electrostatic field of the other. The total capacity of the two wires is diminished by the overlapping of the two individual fields. (See *Capacity Measurement of Antenna.*)

MUTUAL INDUCTANCE—Symbol M—The magnetic flux that is common or mutual to two inductively coupled circuits. The mutual inductance of two circuits is the change in the interlinkage of flux that takes place in one circuit for a change of unit current in the other. This is also called the *coefficient of mutual induction* or the *mutual induction coefficient.* The units of mutual inductance are the same as those of *self-inductance.* (See *Inductance, Mutual; Self-Inductance, Induction, Coupling, Induced E.M.F., Induction, Mutual; Coefficient,* and *Mutual Induction Coefficient.*)

MUTUAL INDUCTION—The interaction between two current-carrying coils, not having direct metallic connection. The reactance of a coil such as shown in the illustration at A_1 is reduced by the proximity of another coil, A_2 if the circuit of the latter contains no external source of electromotive force, such as an alternator. The reason for this lessening in the primary reactance is that the flux excited by the first coil induces a current in the secondary coil and this current opposes the action of the primary current. As a result the total *magnetomotive force* (q.v.) the *flux* (q.v.) and consequently the *counter-electromotive force* (q.v) induced in the first coil are reduced. The effect of the secondary coil depends upon its distance from the primary coil and upon the amounts of the resistance R_2 and reactance X_2. If both coils are mounted on the same iron core, the action is greatly increased. In some instances mutual induction is desirable, as in the case of coupled radio circuits, or even in the ordinary transformer. In other cases, mutual inductance may be undesired as in the case where a radio antenna parallels a power line. (See *Induction, Mutual; Mutual Inductance,* also *Mutual Induction Coefficient.*)

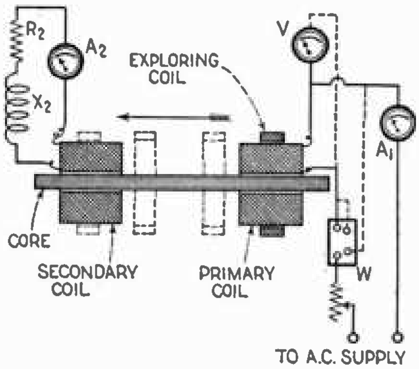

Diagram illustrating mutual inductance. The reactance of the coil A_1 is reduced by the proximity of coil A_2, since the circuit of A_2 contains no source of electromotive force. The effect of the secondary coil depends not only upon its distance from the primary coil but also upon R_2 and X_2. The presence of the iron core greatly increases the action.

MUTUAL INDUCTION COEFFICIENT—Another name for *mutual induction* (q.v.) That coefficient by which the rate of change of the current in a circuit must be multiplied to give the electromotive force induced in an adjacent circuit. From the principle of conservation of energy, it can be shown that the mutual inductance of a circuit A with respect to a second circuit B equals the mutual inductance of B with respect to A. (See *Mutual Inductance, Inductance, Mutual; Self-Inductance, Induction, Coefficient,* also *Coupling.*)

N

NAGYAGITE—A mineral containing lead, gold, antimony, sulphur, and tellurium. It receives its name from the location of its discovery near Nagyag in Transylvania. It is also known as black or leaf tellurium, and has been used as a crystal rectifier in conjunction with zincite.

NALLY, Edward Julian—American radio pioneer. Born April 11th, 1859, in Philadelphia, he joined the Western Union Telegraph Company in St. Louis and worked his way up in the telegraph service until in 1913, he was appointed vice-president and general manager of the Marconi Wireless Telegraph Company of America. Nally was one of the first to see the possibilities of wireless, and under his control the first commercial wireless communication was established between the United States and Japan, in 1914, and in 1920, he founded the first commercial wireless service between the United States and Great Britain, and afterwards to other countries. He was appointed the first president of the

Radio Corporation of America, and is a member of many scientific and other societies.

Edward Julian Nally.

NAME PLATE—A metal plate affixed to a radio set or other apparatus, giving the maker's name, trade name of apparatus, serial number or other details concerning the machine.

NAPERIAN BASE—The base of the so-called natural system of *logarithms* (q.v.). It is usually represented by the symbol ϵ (epsilon) and is numerically equal to 2.718 appx. (See *Damped Waves.*)

NAPERIAN LOG—See *Naperian Logarithms.*

NAPERIAN LOGARITHMS—Also called *Natural Logarithms.* Logarithms to the base ϵ. Logarithms constitute a tabular system of numbers, by which the operation of multiplication can be performed by addition, division by subtraction, involution by a single multiplication, and evolution by a single division. John Napier, Laird of Merchiston, Scotland, is generally regarded as the inventor of logarithms. The logarithms set forth by Napier were those of trigonometric functions. Later on Napier's logarithms were adapted to positive integers, using as a base the number 2.718 appx. and these logarithms are now called natural or Naperian logarithms. (See *Decrement,* also *Logarithms.*)

NATIONAL ELECTRIC CODE—A uniform code of rules, based upon the requirements of fire underwriters, for the electric wiring and electric installations in buildings. Unless these rules are compiled with fire insurance will not be issued or if issued previously will be voided. The National Electric Code contains a section covering special requirements for radio installations. The code is revised annually to meet new conditions and requirements. Copies of the code can be obtained from local insurance agents or from the National Board of Fire Underwriters at New York or Chicago.

NATURAL ELECTRIC WAVES—Wireless waves due to natural causes such as lightning discharges. When received in the radio apparatus, they are called *strays* (q.v.), *static* (q.v.), atmospherics, *X's* (q.v.), and various other names. There are many different causes for these stray waves. Some produce a grinding noise in the loud speaker, others a hissing noise often associated with snow or rain. Nearby lightning produces a sharp snap. Another type of strays cause loud crashes in the speaker.

NATURAL MAGNET—A substance which possesses in a natural state the properties of a magnet. Such a substance is the *lodestone,* a magnetic oxide of iron also called *magnetite* (q.v.). (See *Loadstone.*)

NATURAL FREQUENCY—The frequency at which free oscillations occur in a circuit. If L, C, and R are respectively the inductance, capacity and resistance of a circuit, then free oscillations in the circuit will have the frequency

$$f = \frac{1}{2\pi}\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}$$

This is the natural or *fundamental frequency* of the circuit. The natural frequency of an aerial is the frequency corresponding to the natural wave length. (See *Natural Wave Length,* also *Measurement of Wave Length.*)

NATURAL RECTIFIER—A mineral which possesses the property of conducting an electric current in one direction only. (See *Crystal,* also *Crystal Detector.*)

NATURAL WAVE-LENGTH—Length of the wave emitted by an aerial when no added inductance or capacity is inserted in the aerial circuit. This is also known as the *fundamental wave length* (q.v.). By putting *inductance coils* (q.v.) (*loading inductance*) (q.v.) in the aerial circuit, longer waves may be radiated, while condensers put in series with the antenna enable it to produce shorter waves than the aerial would ordinarily radiate. The use of a series condenser is avoided wherever possible since it has the effect of decreasing the total capacity of the aerial circuit and thereby decreasing the amount of power which can be given to the antenna. The addition of some inductance has a beneficial effect, since the decrement of the aerial is thereby lessened and a sharper wave results. (See *Measurement of Wave Length; Wave Length.*)

NAUTICAL MILE—abbreviation *Naut*—A marine unit of distance. Equivalent to one minute of longitude at the equator. This unit is also sometimes referred to as a *Telegraph Naut.* It is equal to 1.1528 statute miles (the statute mile being equal to 5,280 feet). The term *nautical mile* is used especially in submarine cable work. and also in matters pertaining to navigation. The term *naut* is differentiated from *knot* in that the former refers to a distance whereas the latter refers to a rate or unit of speed.

NEGATIVE CARRIER*—An electron combined with neutral gas molecules. When the velocity of the electrons in a vacuum tube is less than the value necessary to cause ionization by collision, the electrons attract the neutral gas molecules and so form heavy negative carriers. The ease with which this formation of negative carriers takes place depends on the nature of the gas. Such gases as argon and mercury vapor do not readily form negative carriers, while hydrogen and oxygen combine with electrons more easily. The effect of this negative carrier formation is to counteract the reduction of the negative space charge occasioned by the heavy positive ions formed by collision ionization. The positive ions are ions of the gas from which one or more electrons have been removed. The ions therefore have very nearly the same weight as the gas atoms. The negative carriers, on the other hand, may consist of an atom or molecule to which has been attached an electron. It is also possible that the attraction between an electron and the neutral gas molecules can result in the formation of clusters consisting of more than one molecule held together by the electron. These negative carriers, therefore, move as slowly as, and often more slowly than, the positive ions and consequently have a relatively great effect in counteracting the tendency of the positive ions to reduce the negative space charge of the electrons. (See *Electron Theory,* also *Ion.*)

NEGATIVE CORPUSCLE—The natural unit of negative electricity. The *electron* (q.v.). It is more usual to refer to the negative corpuscle simply as the *corpuscle.* This was the name given by Sir J. J. Thomson to the carriers of electricity shot off from the cathodes in vacuum tubes. The corpuscle was found to have a charge equal to the electron and a mass which is $\frac{1}{1845}$ of that of the hydrogen atom.

NEGATIVE ELECTRODE—In a *primary cell* (q.v.) the *cathode* (q.v.) which is the carbon, copper, etc., electrode, is the *negative electrode,* while the *pole* of this electrode is the positive pole, since it is positive in relation to the external circuit. In a *secondary* or *storage cell,* the spongy lead plate, which is the anode during discharge, is called the *negative electrode* and its pole the *negative pole.*

NEGATIVE ION—An atom, which is the smallest particle of an element capable of existing, plus an *electron* (q.v.) (See *Anion, Cathion, Electron Theory,* also *Ion.*)

NEGATIVE POLE—The south-seeking end of a magnet. In a generator, the terminal into which the current returns from the external circuit. In a storage cell, the terminal of the negative plate. In a primary cell, the pole of the positive plate.

NEGATIVE RESISTANCE—A current path within a vacuum tube or an arc in which current decreases as voltage increases. The emission of electrons from cold electrodes under the impact of electrons (known as *secondary electron emission* (q.v.) or *delta rays*) results in a negative resistance or *falling characteristic.* The presence of secondary electrons can be shown by referring to the illustration. The plate P is kept at a constant positive potential with respect to the filament F by means of the battery B. When there is no difference of potential between the filament and the grid G, the cur-

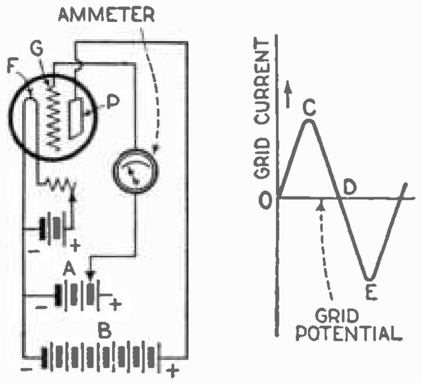

At the left is a hook-up for demonstrating negative resistance. The curve at the right shows the decrease of grid current although grid potential is increased.

(*Van der Bijl—The Thermionic Vacuum Tube.)

rent in the circuit FGA is very small, since practically all the electrons emitted from the filament are drawn through the openings of the grid and thrown on to the plate. If the potential of the grid (positive with respect to the filament) is increased, the current to the grid at first increases, as shown by the part OC in the accompanying curve. When the grid potential reaches a certain value, the current as indicated by the ammeter begins to decrease, drops down to zero at D and then reverses its direction of flow. The reason for this is that while the difference of potential between the filament and the grid is small, the electrons that strike the grid enter it, but as the positive grid potential is increased the electrons on striking the grid emit *secondary electrons* from it and these are drawn to the plate which is maintained by the battery B at a positive potential, with respect to the grid. The net current as shown by the ammeter is the sum of the electrons entering the grid and those leaving it. When the velocity with which the electrons strike the grid increases beyond a certain value, one primary electron can knock out more than one secondary electron from the grid and the current in the filament-grid circuit reverses. When the positive grid potential is increased to such an extent that the grid becomes positive with respect to the plate, the secondary electrons are no longer drawn away to the plate, but are driven back to the grid so that the reversed current in the grid circuit again decreases as shown at E on the curve, and finally assumes the original direction. Hence, considering the current as indicated by the ammeter and the voltage between filament and grid, it can be seen that over the portion of the curve CDE, the current decreases as the voltage increases. Therefore CDE represents a *negative resistance* characteristic. It has been found that a device which possesses a negative resistance can function as an amplifier and also as a generator of continuous oscillations.

NEGATRON—An English vacuum tube containing a filament, a grid and two plates. This tube was devised by J. Scott-Taggart, who used it for obtaining *negative resistance* (q.v.) characteristics. The two plates are fixed one on each side of the filament. The accompanying illustration shows a typical negatron. The bulb is tubular in shape and the four electrodes are plainly shown. The grid, in this tube, is a metal rod.

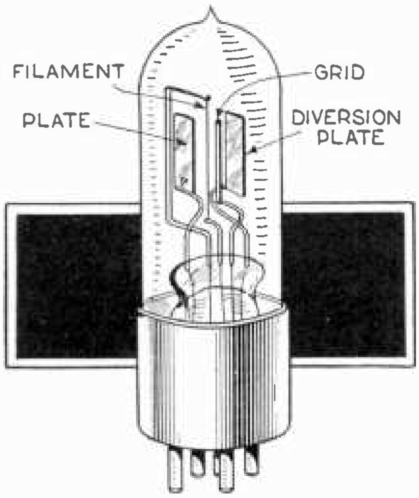

A negatron tube, showing the two plate construction.

NEON LAMP—An incandescent lamp in which a reddish light is produced by the incandescence of neon gas at low pressure. This lamp has very low current consumption and has been applied to pilot lights, signs, etc.

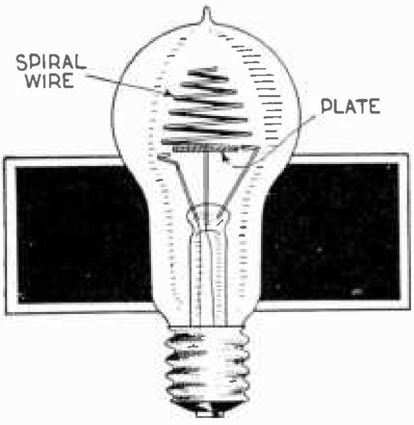

Neon Lamp

NEON TUBE—A vacuum tube containing the gas neon at low pressure used in the Fleming *Cynometer* (q.v.) etc. In this instrument, which is used to measure the frequency of electric waves, wave lengths, etc., a capacity formed by one brass tube sliding over another can be varied simultaneously with an inductance consisting of an air core solenoid and arranged so that the point at which resonance takes place is indicated by the glowing of the *neon tube* and can be read from a calibrated scale.

NERNST LAMP—An incandescent lamp in which the incandescent body consists of a strip of material composed of a mixture of oxides of metal such as zirconium, magnesium and other refractory oxides. The incandescent portion of the lamp is called the *glower* and is conductive only at a high temperature. Consequently it has to be heated by an auxiliary heating coil. The heating coil is automatically cut out of the circuit when the current starts to flow through the glower.

NEUTRAL BODY—A body void of electrification. A charged body which loses its electrification is said to be *discharged* or *neutralized.* All matter is capable of electrification.

NEUTRAL LINE OF MAGNET—The middle portion between the two magnetic poles of a bar magnet, where there is no manifestation of magnetism.

NEUTRAL WIRE—A middle wire, in a *three wire system* (q.v.) of power distribution, which is kept at a potential mid-way between the positive and the negative mains. In the three-phase system of alternating current distribution, the conductor connected to the neutral point of a "star" or "Y" connected alternator is also known as the neutral wire. The neutral wire is here also known as the *common return.*

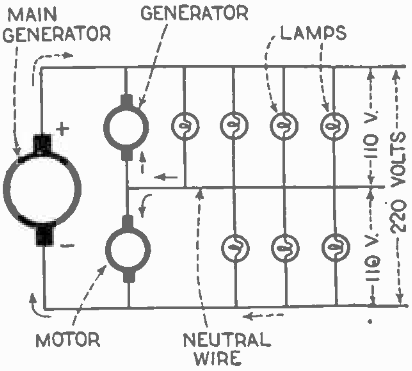

A three wire system with unbalanced load, showing neutral wire.

NEUTRODYNE—See *Neutrodyne Circuit,* also *Hazeltine Neutrodyne Receiver.*

NEUTRODYNE CIRCUIT—A circuit used in radio reception in which radio frequency amplification is used for neutralizing the effects of capacity of the tube and its socket. Figure 1 shows a portion of a receiving circuit. In this diagram the tube capacity is represented by a condenser shown in dotted lines. By studying this circuit, it can be seen that a closed oscillatory circuit exists, made up of the inductances and capacities in the plate and grid circuits. Just as radio frequency currents can pass through condensers in a receiving set, so they can pass through or across the vacuum tube from the plate to the grid, which forms a condenser. This condenser effect is the reason why radio frequency amplifiers oscillate. A difference of potential occurs in the plate side of the coil in the plate circuit. Instead of this being handed on to the next tube for additional amplification, a portion of this potential leaks through the capacity of the tube and its socket and affects the grid in such a way that trouble is experienced with self-oscillation. In the neutrodyne circuit, the

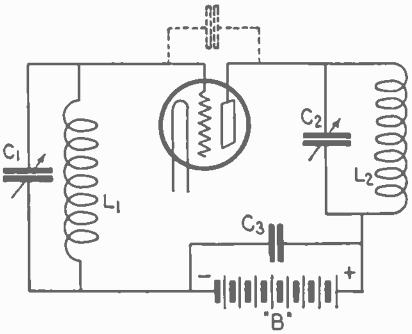

Fig. 1. A portion of a receiving circuit. Dotted line shows tube capacity represented by an equivalent condenser.

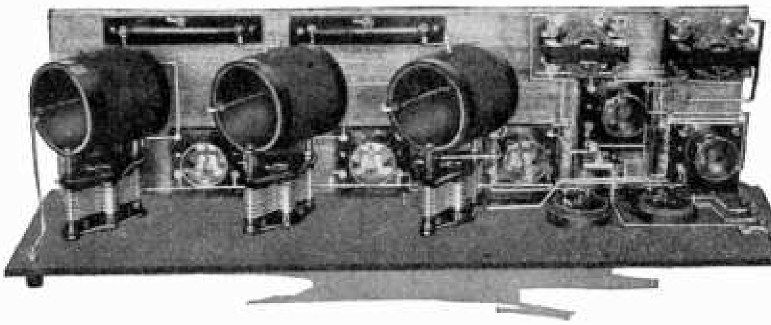

Fig. 2. A typical neutrodyne set.

inter-element capacity of the amplifier vacuum tubes is neutralized by means of special condensers called *neutrodons*. The capacity of these is very low, being approximately equal to the internal capacity of a vacuum tube. By reason of this equality, any tendency of a large amount of radio frequency current to pass back through the tube through the grid is defeated and instead is neutralized by the combination of the neutralizing capacities, the inter-element capacity of the vacuum tubes and the secondary windings of the tuned radio frequency transformers. This effect is in reality a bucking one, since the current is made to take two paths. Each neutralizing condenser must be adjusted so that its capacity will equal that of the vacuum tube it is connected with. Since it is impossible for a neutrodyne circuit to oscillate, there is no radiation of energy from the set, with consequent disturbance of other receiving sets in the vicinity.

The neutrodyne circuit was invented by Professor L. A. Hazeltine. The typical neutrodyne receiving set uses five tubes, employing two stages of tuned and neutralized radio frequency amplification, with detector and two audio frequency stages. (See *Hazeltine Neutrodyne Receiver.*)

NICHROME WIRE—A nickel-chromium-steel alloy wire, which is able to withstand a bright red heat in the atmosphere without oxidizing. Nichrome is practically non-corrosive. It has a very high melting point, about 1550 degrees Centigrade and is used extensively in electrical heating appliances and resistance elements.

NIGHT EFFECT—Changes in the strength of radio signals noticeable especially at night. This effect is also known as *fading* (q.v.). Fading does not generally occur in the immediate vicinity of the transmitting station. It seems to happen more often on wave lengths below 400 meters. A certain station is being received with normal intensity, when suddenly the sounds become very faint and in a few minutes they may again become normal or may even become stronger than usual. In some cases the variations from strong signals to weak take place every few minutes while in other cases there may be an interval of several hours. It has been observed that fading takes place over land more often than in radio transmission at sea.

NOBLE, Sir William—British engineer. Born in 1861 and educated at Gordon's College, Aberdeen, he thereafter entered the Aberdeen telegraph office. In 1893 he was appointed engineer for the northeast area of Scotland, rising rapidly in the service until he became chief engineer in 1919. He retired from the Post Office in 1922, and became a director of the General Electric Company and of the British Broadcasting Company. Sir William Noble has written many articles on telegraphy and telephony and is considered to be an authority on these subjects.

NODES—In a wave form, such as shown in the illustration which represents an alternating current sine wave, the zero points. Thus in the alternating current sine wave, the points of zero current or potential, at A, C, E, and G, are nodes, and the points B, D, and F are the anti-nodes or *loops* (q.v.).

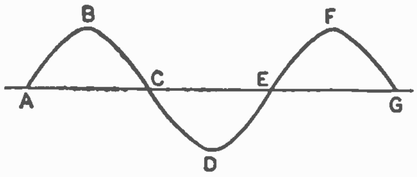

Sine wave illustrating nodes and anti-nodes.

NODEN VALVE—An *electrolytic rectifier* (q.v.) which allows only current in one direction to pass through it. This rectifier utilizes an aluminum rod or cylinder as a cathode and the anode may be of iron, lead or of carbon. Ammonium phosphate is used as the electrolyte. The principal of operation of this rectifier depends upon the fact that the current in the positive direction is stopped at the aluminum plate through the formation of an insulating crust of aluminum phosphate and aluminum oxide. These two substances present an extremely high resistance to the passage of the current in the positive direction, but the compounds dissolve again upon reversal of the current. By suitably combining two or more cells, both the half waves of an alternating current can be rectified. In Fig. 1 is shown the principle upon which the Noden Valve acts, and Fig. 2 shows a hook-up for connecting four cells to obtain full wave rectification. (See *Electrolytic Rectifier Full Wave*, also *Rectifier.*)

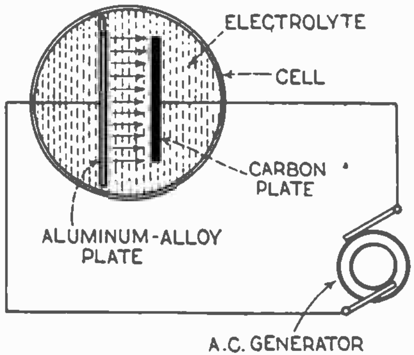

Fig. 1. Diagrammatic illustration of the principle of the Noden valve.

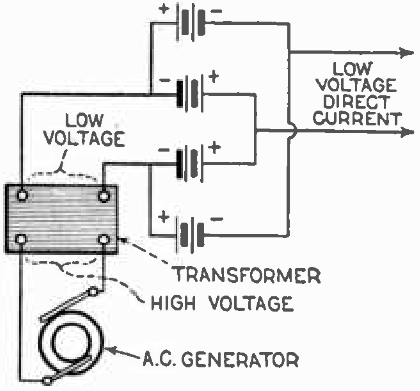

Fig. 2. A hook-up for connecting four Noden cells to obtain full wave rectification.

NON-CONDUCTOR—See *Insulator.*

NON-INDUCTIVE COIL—A coil having negligible inductance. Such a coil may be wound by doubling the wire upon itself and then winding the two parallel halves side by side.

NON-OSCILLATORY—Free from oscillations or vibrations. Without periodicity, that is to say, *aperiodic* (q.v.). A current which is uniform in direction and flow and free from oscillations is said to be a non-oscillatory current. A non-oscillatory circuit is one in which an impressed potential will produce current that gradually diminishes in amplitude without reversing its direction of flow. This is also called an *aperiodic circuit* (q.v.). Such a circuit can have no natural period of oscillation.

NON-PERIODIC—See *Aperiodic*, also *non-oscillatory.*

NON-RADIATIVE ANTENNA—A combination of capacity and inductance equivalent to that of an aerial, used to test transmitting apparatus without radiating waves. (See *Mute Antenna.*)

NORTH MAGNETIC POLE—A point on the earth at a latitude 70 degrees North, longitude 97 West. The north magnetic pole does not coincide with the north (geographic) pole. (See *Magnetic Poles.*)

NORTH POLE—See *Magnetic Poles*, also *North Magnetic Pole.*

NO-VOLT RELEASE—An electro-magnetically controlled device, inserted in the field circuit of a motor so that it holds the handle of the motor starter in place on the last stud, but if the supply current fails, the magnet becomes de-magnetized and allows the spring attached to the starter handle to pull the handle back to the "off" position. This prevents damage to the windings of the motor, in case the current comes on again, since it is then necessary to start up the motor in the usual way, using the starter and gradually increasing the voltage by cutting out the limiting resistances provided in the starter. (See *Overload Release.*)

NULL METHOD—An electrical method of testing in which adjustments are made so that zero deflection is obtained in a galvanometer, as for example when using a *Wheatstone Bridge* (q.v.). This is sometimes referred to as the *zero method.* (Null is also a German word meaning zero.)

O

OCTAHEDRITE—Chemical Symbol TiO_2—Titanium dioxide, an eight sided crystal which has been used as a crystal detector for detecting and rectifying radio currents. (See *Crystal.*)

OERSTED—The unit of *reluctance* (q.v.). The reluctance or in other words, the magnetic resistance offered by a cubic centimeter of vacuum. Unit *magneto-motive-force* (q.v.) will generate a unit of *magnetic flux* (q.v.) through unit reluctance. The oersted was named after the Danish physicist, H. C. Oersted.

OHM—Symbol Ω (Omega)—The unit of electrical resistance. It is the ratio of unit *electromotive force* (q.v.) to unit current. The ohm is the basis of the entire electrical system of units and the volt and the ampere are defined in relation to it. Very elaborate apparatus is required in order to determine the ohm absolutely. However, resistances can be compared easily and once the standard is determined it is a simple matter to make duplicates. In accordance with an act of Congress, the ohm was adopted as a legal unit and defined as follows: The unit of resistance shall be what is known as the international ohm, which is substantially equal to one thousand million units of resistance of the centimeter-second-gram system of electromagnetic units and is represented by the resistance offered to an unvarying electric current by a column of mercury at the temperature of melting ice, 14.4521 grams in mass, of a constant cross-sectional area and of the length of 106.3 centimeters. Stated in English units, the ohm is specified as the resistance offered to an unvarying electric current by a column of mercury at the temperature of melting ice (32 degrees Fahrenheit), 0.5050 of an ounce in mass (approximately 3/100ths of a pound), of a constant cross-sectional area and of a length of 3.487 feet.

OHMMETER—An instrument used to measure resistance directly in ohms. It is particularly adapted for measuring high resistances, although certain types of ohmmeters are made for measuring lower resistances. In one form of ohmmeter, the moving system is deflected by forces due to currents in two coils at right angles to one another. These carry currents in one case proportional to the current through the conductor under test, and in the other case proportional to the potential drop across it. This is in effect, a special form of *galvanometer* (q.v.) and when combined with a hand driven generator gives the conventional *megger* (q.v.) The bridge type ohmmeter depends on the *Wheatstone Bridge* (q.v.) principle. In this instrument, however, the rheostat resistance remains constant and the bridge arms are formed by a wire, called a *slide wire*, resistances being varied by means of a moving contact as shown in the illustration. This type of instrument is often referred to as a *slide wire ohmmeter*. While a straight slide wire is used in some types, in others the wire is wound on an insulating cylinder. In the Weston direct-reading ohmmeter, there is a permanent magnet and a moving coil having two windings and an adjustable magnetic shunt. Provisions are made for connecting an unknown resistance and an auxiliary battery in the circuit of the instrument. A checking plug is provided for checking the accuracy of the instrument before use. When the key is closed, the battery current is divided between two windings. The magnetic force exerted on the one winding, which is in series with a fixed resistance, tends to deflect the pointer up the scale, while the force exerted on the other winding tends to return it to zero. With the correct battery voltage the magnetic shunt may be adjusted to bring the pointer to full scale position. When the plug is shifted to the low range position, a resistance equal to that of the low range is removed from the circuit whose magnetic force tends to deflect the pointer towards the zero position. By connecting a resistance equal to that of the low range across the "unknown" binding posts, the previous condition is restored and the pointer will go back to the full scale position. For values of resistance lower than this the opposing force will be greater and hence the coil will be brought back to a lower position on the scale. At zero resistance across the "unknown" binding posts, the magnetic forces of the two windings will balance each other, thus causing the coil to stand at zero on the scale. The Vawter indicating ohmmeter is of the galvanometer type and utilizes two coils carried by a shaft and moving in the field of a permanent magnet. A specially shaped core is used to give the desired scale characteristics. By making connections to the coils through spirals of negligible torque, the position of the pointer does not depend upon the value of the current nor on the strength of the permanent magnet.

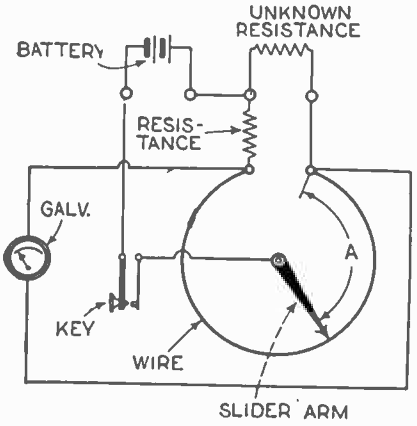

Diagram of connections of a slide wire ohmmeter.

OHMIC COUPLING—See *Resistance Coupling.*

OHMIC DROP—The fall in potential which occurs when an electric current flows through a resistance. (See *IR Drop.*)

OHMIC LOSS—Power or energy loss due to the resistance which an electrical circuit offers to the flow of current. (See *I^2R Loss.*)

OHM'S LAW—Current flowing in a conductor will increase directly with increase of voltage and will decrease directly with increase of resistance. In other words, voltage is the cause, while current is the effect. The effect is directly dependent upon the cause and inversely dependent upon the opposition offered to the cause. Ohm's law is the fundamental law of electrical engineering. Ohm's law can be stated as an equation as follows:

$$\text{Current (in amperes)} = \frac{\text{Voltage (in volts)}}{\text{Resistance (in ohms)}}$$

or

$$\text{Voltage} = \text{Current} \times \text{Resistance}$$

$$\text{Resistance} = \frac{\text{Voltage}}{\text{Current}}$$

The above three equations all mean the same thing and serve to express in various forms, the idea set forth above. Ohm's law can be applied to every conductor, provided that the conductor stays at a constant temperature, that there is no internal electromotive force in the conductor and that the distribution of the stream lines in the conductor remains unchanged. Ohm's law is easy to understand and easy to apply. Thus in a particular circuit, a certain voltage causes a certain current to flow. Double the voltage will cause double the current to flow, if the resistance is the same. If the resistance is doubled, however, the same voltage will result in only half the current flow. Double voltage and double resistance results in unchanged current flow. Ohm's law applies to a portion of an electrical circuit as well as to the entire circuit. Care must be taken when applying the law to only part of the circuit, to consider only the resistance, current and voltage of that part.

OMNIBUS BAR—A conducting bar of copper, mounted in back of a switchboard, or panel and serving as a common connector for two or more pieces of apparatus. This term is more usually shortened to *bus bar* (q.v.)

OMNIGRAPH—An intrument which sends code (dots and dashes) mechanically. It is usually operated by clockwork and is connected with a buzzer. It employs metal disc records each of which has a series of notches which correspond to the dots and dashes of the telegraph code. The omnigraph is used by beginners in learning the code and is also used to test the speed of reception of applicants for radio operators' licenses.

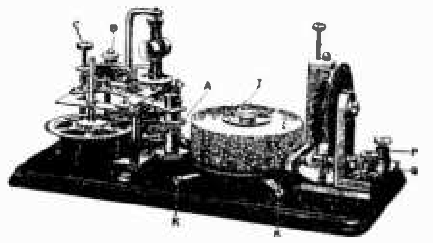

The omnigraph—an instrument which sends code mechanically.

ONDOGRAPH—An instrument for autographically recording the wave forms of varying currents, especially currents which are alternating rapidly. This form of curve tracer was invented by Hospitalier.

ONDOMETER—Another name for a *wave meter* (q.v.)

OPEN ANTENNA—According to the report of the Institute of Radio Engineers standardization committee, this is a condenser antenna, that is to say an antenna consisting of two capacity areas. Antennas may be subdivided into two general classes, those which act principally as condensers, usually simply called antennas, and those which act principally as inductances. These latter are *loop antennas* (q.v.), *coil antennas* (q.v.), etc. When electric waves reach a condenser antenna, they set up an alternating electromotive force between the wires forming the upper plate of the condenser and the ground or the other lower plate of the condenser. As a result of this alternating electromotive force, alternating current will flow in the antenna system. (See *Aerial.*)

OPEN CIRCUIT—An electrical circuit which does not offer a complete path for the flow of electricity. There may be one or more breaks in the path. An open circuit is often referred to as a *broken circuit*. Open circuits often cause trouble in radio apparatus. They may be due to the burning out of a winding, to a connection coming off a terminal, to a soldered connection breaking loose or to similar causes. The method of testing for an open circuit is to connect a battery and a headset in series with the circuit in question. If there is no click in the headset, there is evidently a break in the circuit.

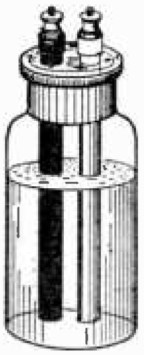

An open circuit primary cell.

OPEN CIRCUIT PRIMARY CELL—A *primary cell* (q.v.) designed for intermittent use and normally kept on open circuit. Such cells usually have a *depolarizer* (q.v.) which acts slowly. In certain cases, no depolarizer whatsoever is employed. Open circuit cells are designed for use during short intervals of time only and must stand for long periods on open circuit, during which the hydrogen is gradually taken away from the negative plate. The operation of the open circuit cell will not be satisfactory if the rest periods are not long enough or are not at frequent intervals.

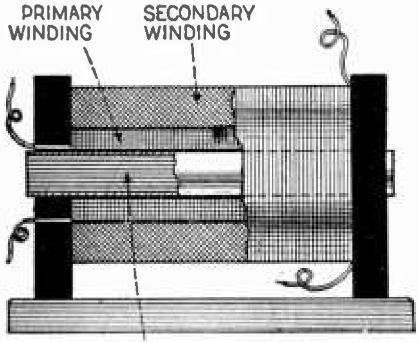

An open-core transformer.

OPEN-CORE TRANSFORMER—A transformer in which the (useful) magnetic circuit is partly through iron and partly through air, instead of entirely through iron as in the *closed-core transformer* (q.v.). It is sometimes referred to as a *polar* transformer. In one type of open-core transformer the arrangement is similar to that of an *induction coil* (q.v.), a hollow iron core, filled with soft iron wires being employed to concentrate the magnetic field.

OPPOSING ELECTROMOTIVE-FORCE—See *Back Electromotive-force.*

OPTOPHONE—An instrument utilizing a telephone receiver controlled by the variations of resistance of a *selenium cell* (q.v.) by means of which the blind are enabled to read ordinary type, through the sense of hearing.

ORDINATE—A mathematical term used in radio in making *logs* (q.v.) or other curves. Defined in mathematical language, the ordinate is that one of the coordinates of a point which is drawn parallel to a line (called the axis of ordinates) to the point from the other axis (called the axis of abscissas) or from the plane of the other axis of coordinates, assumed as a basis of reference. This definition will be clearer if reference is made to the accompanying illustration. The dotted line drawn from the point A to the axis of abscissas is the ordinate of the point A.

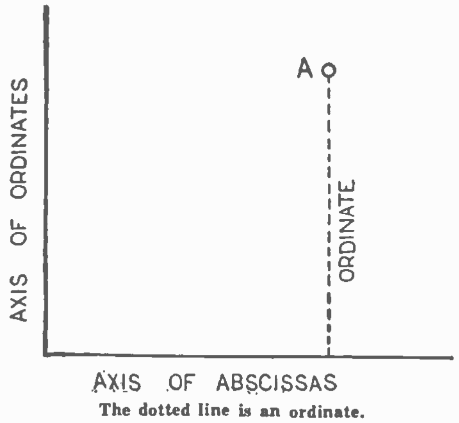

The dotted line is an ordinate.

OSCILLATING CURRENT—An alternating current of high frequency, consisting of a succession of waves of constant length. In some cases, the term oscillating current refers particularly to a current where the amplitude is decreasing in constant proportion due to damping. (See *Oscillation, Oscillation Frequency,* also *Logarithmic Decrement.*)

OSCILLATING IMPULSE—*An oscillating current* set up for the purpose of producing electric waves.

OSCILLATING PERIOD—The time of one complete *oscillation* (q.v.).

OSCILLATION—Periodical surging first in one direction and then in the other. Electrical oscillations are the surgings backward and forward of current periodically, as, for example, in the case of a condenser being discharged in an oscillating circuit. The oscillating action in a discharging condenser is as follows: In the case of a condenser, positively charged, upon being discharged through an inductance, the charge rushes out of the condenser, the current increasing in strength until it reaches a maximum. Then, due to self-induction the current prolongs itself, and it continues until its energy is changed back into an electric charge of an opposite sign, current flowing into the condenser which is then charged negatively. The condenser again discharges giving a maximum current in the opposite direction. the current thereafter flowing back into the condenser with the same direction charge as at first. This action constitutes one complete swing or oscillation and the current continues to oscillate, each time becoming weaker and weaker as the energy is dissipated into heat and also partially into electric waves. It is important to note that the propagation of electric waves could not be accomplished without electrical oscillations. Thus at the transmitting station, the electrical oscillations are converted into electric waves and at the receiving station, the waves are re-converted back into oscillations.

OSCILLATION CONSTANT—The square root of the product of the capacity and the inductance of a circuit. The *natural period* of a circuit is determined by its inductance and capacity in the same manner as the natural period of a steel spring clamped at one end in a vise is determined by its mass and elasticity. Where a simple aerial is used it is an easy matter to calculate the length of a wave radiated from it, provided the oscillation constant has been found. The velocity of electric waves is approximately 3 x 10^{10}cm. per second, and hence the length of the wave radiated is about 60 times the oscillation constant where the wave length is measured in meters. (See *Aerial Tuning Condenser,* also *Isochronous.*)

OSCILLATION FREQUENCY—The number of complete oscillations or cycles flowing in a circuit per second. The oscillation frequency of a circuit varies inversely as the square root of the product of the inductance and capacity. The frequency is inversely proportional to the wave length, hence the addition of either *capacity* (q.v.) or *inductance* (q.v.) to the oscillatory circuit will result in an increase of *wave length* (q.v.). (See *Frequency.*)

OSCILLATION POINT—That current value, known as the *critical current* (q.v.), at which the vacuum tube or tubes in a radio receiving set start to oscillate.

OSCILLATION TRANSFORMER—A high-frequency air-core transformer used for transferring electrical oscillations from one circuit to another. Either one or both the windings may be tuned to the frequency of the oscillations. The transformer used for coupling the oscillator of a transmitting set to the aerial is an example of an oscillation transformer. (See *Coupling, Inductive Coupler,* also *Transformer.*)

OSCILLATION VALVE—See *Oscillator,* also *Vacuum Tube.*

OSCILLATOR—A device or apparatus for producing electrical *oscillations.* The *vacuum tube* (q.v.) under correct conditions may be used to produce sustained oscillations of a definite amplitude and frequency. In order for a vacuum tube to act as an oscillation generator it must be capable of amplifying, part of its output energy must be returned to the input, an oscillation circuit must be combined with the tube, which possesses inductance, capacity and resistances of values such that the tube will oscillate with the desired frequency and finally the tube must have certain characteristics which when combined with the constants of the oscillatory circuit, will determine the amplitude of the oscillations. Oscillations may also be produced by means of a high frequency alternator and by a direct current generator and electric arc. (See *Alternator,* also *Alexanderson.*)

OSCILLATOR, HERTZIAN—Two insulated rods, with axes in line, with their nearer ends forming a spark gap and their outer ends bearing plates or balls to give the system the required capacity. The overall length is usually less than 6½ ft. and the whole constitutes an open oscillator. A Hertzian oscillator is sometimes called a *dipole.* When the rods are brought near each other but not in actual contact and the conductors are connected to the terminals of an *induction coil* (q.v.) a succession of sparks can be caused to jump the gap and these set up electric waves in the *ether* (q.v.).

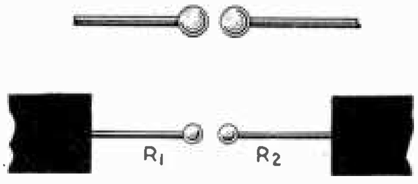

Typical Hertzian Oscillators.

OSCILLATOR, LINEAR—The straight wires connected on either side of the spark gap of an induction coil, traversed by oscillations. Linear oscillators are of the open type. As used by Hertz, they consisted of two rods, having a spark gap between their adjusted ends, while their outer ends terminated in plates or balls in order to increase the capacity of the system.

OSCILLATOR, RIGHI—In this type of oscillator, the spark gap is established between two metal spheres placed between two smaller spheres. (See *Oscillator,* also *Oscillator, Hertzian.*)

OSCILLATOR, THERMIONIC—A vacuum tube used to generate continuous oscillations. (See *Oscillator.*)

OSCILLATORY CIRCUIT—A circuit possessing *inductance* (q.v.) and *capacity* (q.v.), the electrical constants being such that an oscillatory current can be set up. The accompanying figure shows a simple oscillatory circuit, consisting of a condenser, an inductance and a spark gap. When a source of high potential is connected across the spark gap and the latter is properly adjusted, the condenser is charged and there is a spark across the gap. Provided that there is not too much resistance in this circuit, there will be set up an *oscillating current* (q.v.) of high frequency, decreasing in amplitude according to a fixed ratio. Such a current can be used to set up electric waves. Oscillatory circuits are of two general classes, closed as shown in the illustration and open, as in the case of an aerial system. An aerial which has no tuning condenser or inductance in its circuit still has distributed inductance and capacity, whereas an aerial having a tuning inductance and a condenser in its circuit, has a combination of concentrated and distributed inductance and capacity.

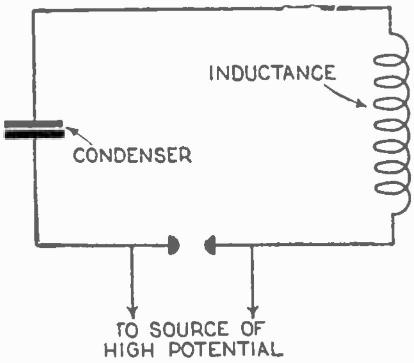

A simple oscillatory circuit.

OSCILLATORY CURRENT—See *Oscillating current.*

OSCILLION—Name which has been applied to the *vacuum tube* (q.v.) especially where used as an oscillation generator.

OSCILLOGRAM — A picture, graph or record of a wave form usually obtained by means of an *oscillograph* (q.v.).

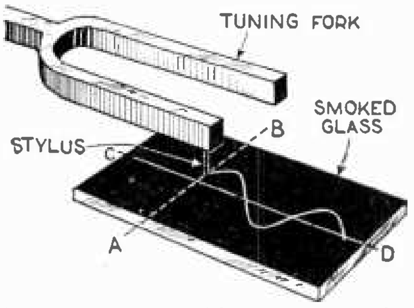

An oscillogram obtained by means of a tuning fork.

The illustration shows an oscillogram obtained on smoked glass by means of a tuning fork. (See *Oscillagraph.)*

OSCILLOGRAPH — An instrument for recording photographically or showing visually the wave form of alternating currents or of other rapidly changing currents or voltages. The essential parts of such an instrument are a moving coil mirror galvanometer, a rotating or vibrating mirror and a moving photographic plate or film. Figure 1 illustrates the principle of the oscillograph and shows the curves it produces. The oscillograph was invented by M. A. Blondel in 1893. By means of the oscillograph the wave form of a current may be shown as a curve and its characteristics may thus be examined and studied. In order to form the curve, it is necessary to have simultaneous motion in two directions. The vibrator of the oscillograph which consists of a delicate coil of wire, with mirror attached, is made extremely light so that the moving system will have as little inertia as possible. The coil is placed in a powerful magnetic field and when an oscillating current is passed through it, it oscillates in synchronism with the current. A spot of light focused on the mirror is made to fall on a screen at the predetermined distance from the vibrator. This spot of light traces out a straight line on the screen. Substituting a photographic film for the screen and arranging the film so that it will move in a vertical path across the moving beam of light (which is assumed to be moving from right to left across the screen) a curve is traced on the film. The commercial form of oscillographs are provided with apparatus whereby the curve may be examined without the necessity of photographing it. In these the beam of light is reflected from a rotating or vibrating mirror in such a way that it receives a horizontal and an up and down motion and is focused onto a curved plate of glass. Tracing paper may be stretched over the glass so that the image of the spot of light traces a visible curve on it. The curve being repeated very rapidly appears to be standing still and hence it is possible to trace over it with a pencil or even to place a piece of photographic paper on the curved glass and make a record of the curve in this way. The motion picture camera has been applied to the oscillograph and is useful where longer records are required. It should be noted that oscillographs which utilize a moving coil, such as the Duddell or Blondin types. will work over a band of frequencies up to a maximum of about 300 cycles per second. Inasmuch as the oscillations common in radio work are of much higher frequency than 300 cycles, it has been found necessary to use an oscillograph whose moving parts are without appreciable weight. The use of electrons as moving parts has solved this problem. A Frenchman, Alexander Dufour has applied the Braun tube to the construction of an oscillograph which has no appreciable inertia and which is capable of operating at a frequency up to a million cycles per second. This device uses only minute amounts of energy in its operation and therefore does not disturb the original circuit. It is able to register both voltage and current simultaneously and is so arranged that a single impulse is sufficient for a photographic impression. The Dufour oscillograph consists essentially of two glass tubes fitted by means of a ground joint into a bronze chamber. The upper glass tube carries a cathode and an anode and the other tube has one pair of deflecting plates for electrostatic deflection of the electron stream. Two sets of coils perpendicular to each other are used for magnetic deflection and these are placed outside the tube and slightly below the deflecting plates. In order to photograph the oscillations, a drum is used which is provided with a film magazine allowing six films to be taken in succession. When visual examinations of the oscillations are to be made, a fluorescent screen is turned up into position covering the opening into the interior of the drum thus preventing the films from being exposed when using the screen. After the films are put into the drum, this is placed within the bronze chamber and locked into position. An air-tight door is used to close the opening and the film changing mechanism is operated by means of external controls. Glass windows on either side of the bronze chamber allow the fluorescent screen to be seen. The accompanying sketch shown at Figure 2, gives a diagrammatic representation of the operation of the Dufour oscillograph.

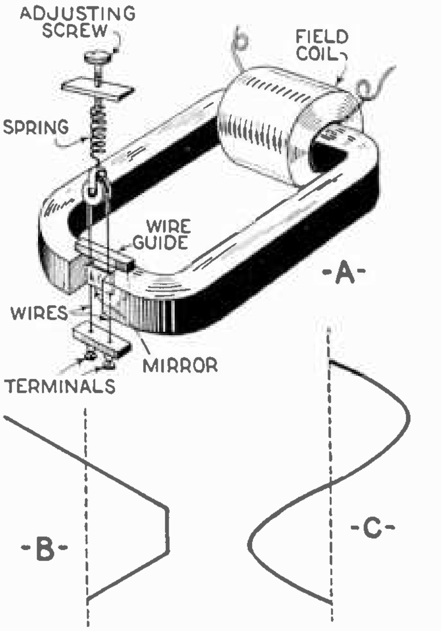

Fig. 1. The principle of operation of the oscillograph is shown at A. At B are shown diagonal and straight paths traced by the beam of light falling on a moving photographic plate. The type of curve which an oscillograph traces for alternating currents is shown at C.

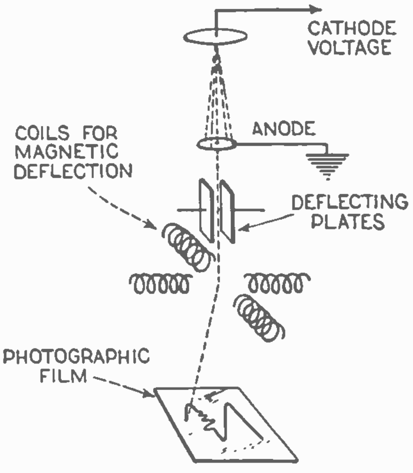

Fig. 2. A diagrammatic representation of the Dufour oscillograph.

OSMOSIS, ELECTRIC—When an electric current is passed through an electrolyte having an anode on one side of a porous diaphragm and a cathode on the other side, there is a tendency for the liquid to pass through the diaphragm towards the cathode raising the level of the liquid on the cathode side. This phenomenon is known as electric osmosis.

OTOPHONE—A device for the hard of hearing, which utilizes an extremely sensitive microphone connected to a two-stage vacuum tube amplifier. The instrument uses dry cell tubes and the batteries are contained in the case with the microphone. A small telephone receiver completes the outfit.

OUDIN RESONATOR — A device used for obtaining high-frequency brush discharges. Discharges of this nature are used in the investigation of resonance effects and also in medical work. The accompanying diagram explains the principle of the Oudin resonator.

B D is an uninsulated coil of copper wire wound on an insulating core with each turn separated from the next adjacent one. At B, the helix is connected to an induction coil and it is tapped by means of a sliding contact at C. The sliding contact is connected through a condenser to the induction coil H, with a spark gap G in parallel in the circuit as shown. The circuit BGFC is a closed oscillatory circuit receiving its excitation from the induction coil H. The circuit C D E is an open circuit in contact with the oscillatory circuit. By moving the sliding contact, C, a point is reached where a brush discharge appears at E, thus indicating that electric oscillations are being set up in the open circuit. (See *Brush Discharge, Corona,* also *Oscillator.*)

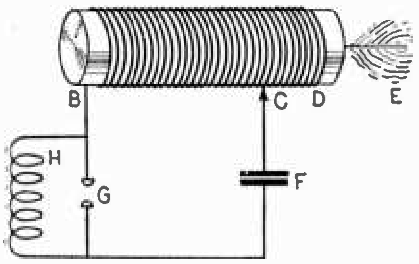

Connections of the Oudin resonator.

OUTPUT—The useful energy given out by a machine, circuit, vacuum tube or other device. (See *Input; Output, Transformer; Output Circuit of Vacuum Tube.*)

OUTPUT CIRCUIT OF VACUUM TUBE—The filament-plate circuit as differentiated from the filament-grid or input circuit. (See *Input Circuit of Vacuum Tube.*)

OUTPUT, TRANSFORMER—The product of the voltage at the secondary terminals of a transformer by the current flowing in the secondary winding. The output of any transformer is always less than the input by an amount equal to the losses. The *efficiency* of a transformer is the ratio of its useful output to the total *input.* (q.v.)

OVERLOAD RELEASE—An electromagnetic device which is used to cause a motor starter handle to return to the off position when the current exceeds a pre-determined overload value.

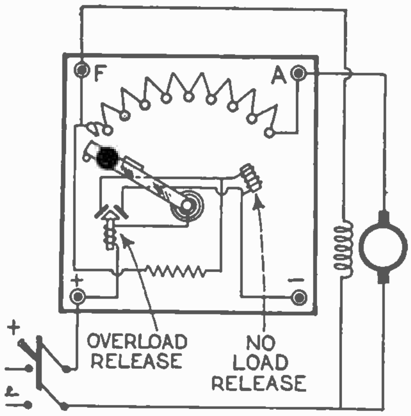

Overload Release.

The overload magnet is usually arranged so that too strong a current will result in short-circuiting the no-volt release thus switching off the driving current. One form of overload release consists of a magnet in series with the line and the armature. This magnet is arranged to attract a pivoted iron keeper whenever excessive current flows. The pivoted arm, when attracted by the magnet, makes connection between two contacts, themselves connected so as to short circuit the holding magnet when the circuit is closed. Therefore, when more than the allowable current is drawn from the line, the overload magnet attracts its pivoted keeper which closes the circuit short-circuiting the holding magnet. The contact arm is then released by the holding magnet and the arm is pulled back to the off position, thus cutting the motor off the line. Another form of overload release utilizes an electromagnet to attract a pivoted arm, arranged so that this will break the circuit between the motor and the line. This results in operating the no-voltage release. (See *No-Volt Release.*)

OVERTONE CURRENTS—Alternating currents which have harmonic frequencies of a higher value than the fundamental or first harmonics.

OVERTONES—Higher harmonics associated with the first harmonic. This term particularly refers to sound waves. In various musical instruments, different overtones are brought out and this results in each instrument having its own tone characteristic.

P

PACKING—(referring to a *microphone*). The tendency of carbon granules in a microphone to settle or pack. In this position they are no longer sensitive and do not readily respond to the vibrations of the diaphragm. Packing may be caused by irregularity in the sizes of the carbon granules. It may be remedied by gently tapping the microphone.

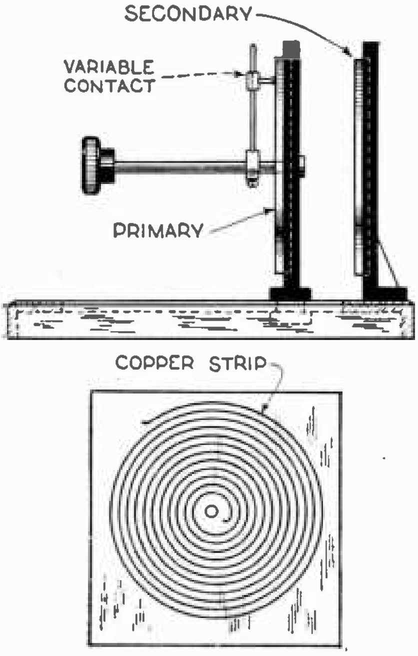

Front and side view of a pancake coil.

PANCAKE COIL—A flat inductance coil. Coils of this type are used in radio receiving sets as radio frequency transformers, also as antenna couplers and oscillator couplers. They may be *basket wound* (q.v.), spider wound, etc. In wireless transmitting apparatus, aerial tuning inductances, loading inductances, and oscillation transformers (illustrated) are frequently of the pancake coil type. (See *Inductance, Antenna.*)

PANEL—A board of insulating material carrying the controlling or measuring devices of an electric circuit. The panel for controlling the charging of storage batteries is called a *charging panel.* A panel carrying generator controls is referred to as a *generator panel.* Where only a single panel is used this is usually called the *switchboard.* The usual switchboard comprises a number of panels. Switchboard panels are commonly made of marble. The usual varieties available are white Italian, pink or grey Tennessee and blue Vermont marbles. Plain slate is sometimes used for panels where the voltage is not too high. When the panels do not require a finish, soapstone is sometimes used. The front of a radio receiving set on which the dials and rheostat controls, etc., are mounted is also called a panel. These panels are usually made of Bakelite, hard rubber, composition or similar insulating substances, although a number of sets are being made with metal panels. The horizontal panel on which the sockets and transformers are usually mounted is known as the *sub-panel* (q.v.).

PANNILL, Charles Jackson—American radio pioneer. Born in Petersburg, Virginia, May 13, 1879. He entered the American Navy, 1898. In 1902 he took a post under Professor R. A. Fessenden and carried out a series of wireless experiments at Hampton Roads, later inaugurating wireless communication between New York and Philadelphia. Pannill was the first man to install wireless on the battleships of the U. S. Navy. Afterwards he carried out a series of communication experiments between various parts of the United States and erected a number of wireless stations. He joined the Marconi Wireless Telegraph Company in 1912 and became a radio adviser to the United States Government, 1914, and assistant director of Naval Communications, 1916. Pannill is a fellow of the Institute of Radio Engineers and a member of the Washington Society of Engineers.

PAPER CONDENSER—A fixed condenser, usually made of tin foil and utilizing a dielectric of paraffin-waxed paper. The 1 microfarad by-pass condenser is a typical example of a paper condenser. (See *Condenser, By-pass.*)

PARALLEL—See *Parallel Connection.*

PARALLEL CONNECTION—Two or more parts of an electrical circuit so connected that the current divides between them. This is also known as a *multiple* (q.v.) connection, *divided circuit,* or *shunt* circuit. Where a number of parts of a circuit are connected in parallel, the same potential is impressed on each part. The current flowing in each branch will depend upon the impressed voltage and the resistance of each branch. (See *Circuit, Parallel.*)

PARALLEL RESISTANCES—Resistances connected so that their terminals have the same difference of potential between them. The greater the number of resistances in parallel, the less will be the total resistance of the circuit. If there are three resistances in parallel of ½, ⅓ and ⅕ ohm, respectively, then the total resistance is equal to the reciprocal (one divided by a number) of the sum of the separate reciprocals. In this case the sum of the reciprocals is 2 plus 3 plus 5 equal to 10 and the reciprocal of this is equal to one-tenth, so that the total resistance of this parallel circuit is one-tenth of an ohm. It can be seen that the total resistance of any circuit having a number of resistances in parallel is less than the resistance of any *one* of the branches. (See *Circuits Parallel*, also *Parallel Connection.*)

PARALLEL RESONANCE—When a concentrated capacity and a concentrated inductance are connected in parallel between terminals to which an alternating electromotive force is applied, and the inductance or capacity or frequency is varied, the condition of parallel resonance exists when the current supplied by the source is a minimum. It should be noted that in series resonance, the total current supplied by the source flows through both the inductance and the capacity whereas in parallel resonance the current supplied by the source is the vectorial sum of the two currents, one flowing through the capacity and one through the inductance.

PARAMAGNETIC—Having a permeability greater than unity. Magnetic as opposed to diamagnetic. The term *paramagnetic* is used in some cases to apply to substances which have only a slightly greater permeability than air such as liquid oxygen, the rare metal erbium, etc. If used in this sense it does not include *ferro-magnetic substances* (q.v.). (See *Diamagnetic material.*)

PARTIAL—An acoustical term denoting any one of the natural vibrations of which a body is capable. An electrical *oscillator* (q.v.) possessing distributed capacity and inductance has, theoretically, an infinite number of possible frequencies, each of which is called a *partial*. The lowest frequency is called *the fundamental.* (See *Harmonics, Harmonic Current, Overtones,* also *Overtone Currents.*)

PARTITION INSULATOR—An ebonite tube having a metallic rod running through its center with wing nuts at each end. It is used for continuing a circuit through a wall or partition.

P.D.—Abbreviation for *potential difference.* (See *Potential, Difference of.*)

PELTIER EFFECT—The change in temperature, either heating or cooling, at the junction of two unlike metals, depending upon whether an electric current originating from an external electromotive force, is sent through the junction in one direction or the other. The Peltier effect is the opposite of the *thermo-electric* effect.

PERCENTAGE COUPLING—The coefficient of coupling between the primary and the secondary of an oscillation transformer in a wireless transmitting system. When the primary winding is placed close to the secondary, the coupling is said to be *close* or *tight.* When the two windings are apart the coupling is said to be *loose.* With the ordinary spark gap in the closed circuit, if the primary and the secondary windings are closely coupled a broad wave will be radiated from the aerial system. If they are loosely coupled, a *sharp* wave will be radiated.

PERIKON DETECTOR—A crystal rectifier utilizing a piece of zincite in firm contact with a piece of chalcopyrite. (See *Combination Detector,* also *Crystal Detector.*)

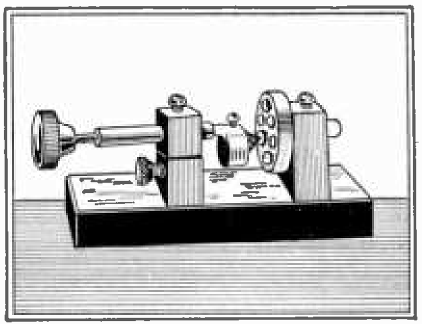

A Perikon detector.

PERIOD—The time required for one complete *cycle.* (q.v.) Referring to an *alternating current* (q.v.) it is the time required for the current or electromotive force to pass through the various values from zero to a positive maximum back to zero again, then to a negative maximum and finally to zero.

PERIODIC—A vibration is said to be periodic if all the phenomena are repeated regularly at equal intervals of time. This time is called *periodic time* or the *period.*

PERIODICITY—A synonym for frequency, or for angular velocity.

PERMANENT MAGNET—Hardened steel which when magnetized retains its magnetism after the removal of the magnetizing force. The usual method of making permanent magnets is to place the steel in a powerful electromagnetic field. They can also be made by stroking with another magnet. Permanent magnets are used in all electromagnetically operated headsets and *loud speakers.* (q.v.) (See *Magnet, Electromagnet,* also *Temporary Magnets.*)

PERMEABILITY—The *permeance* (q.v.) existing between the opposite faces of a cube of the substance each side of which is one centimeter in length. Since the permeability of air is assumed to be unity, the permeability of any substance is the ratio of the flux that passes through it to the flux that would exist in air if the magnetomotive and flux path remained unchanged.

PERMEANCE—That property of a magnetic circuit which allows the flow of magnetic flux. Its reciprocal is *reluctance* (q.v.) which is the property of a magnetic circuit by which it resists the flow of magnetic flux.

PERMITTANCE—A synonym for *capacity* (q.v.) or *capacitance* (q.v.). This property is possessed by every electrical circuit, but does not manifest itself unless there is a change in voltage. The greater the permittance of a circuit, the greater will be the opposition offered to a change in voltage.

PERMITTIVITY—See *Dielectric Coefficient and Constant, Inductive Capacity,* also *Inductivity.*

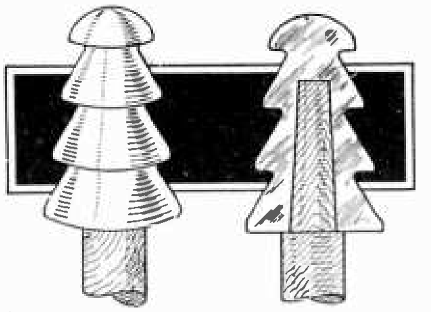

A petticoat insulator. Shown in cross-section at the right.

PETTICOAT INSULATOR—An insulator having two or more flanges or petticoats arranged one on top of the other for the purpose of reducing leakage and preventing accumulation of moisture. Petticoat insulators are usually made of porcelain. Insulators used on low tension lines are also of the petticoat type, in such cases being constructed of glass. (See *High Tension Insulator,* also *Insulator.*)

PHANTOM AERIAL—See *Mute antenna.*

PHASE—A particular state in a regularly recurring cycle of changes. In simple harmonic motion, uniform circular motion, or in periodic changes of any magnitude varying according to a simple harmonic law (as for example an alternating current) the point or stage in the period to which the rotation, oscillation or variation has advanced, considered in its relation to the position or instant of starting or to some other standard position. This relationship is usually expressed in angular measure. Defined in electrical terms, phase is the distance, usually measured as an angle, of the base of any *ordinate* (q.v.) of an alternating wave from any chosen point on the time axis. This distance represents the phase of the ordinate with respect to the point. In any sinusoidal alternating current quantity, the phase at any instant may be shown by the corresponding position of a *vector* (q.v.) revolving about a point with such an angular velocity that its projection at each instant upon a chosen reference line is proportional to the value of the quantity at that instant. As an example, consider a conductor rotating in a magnetic field. The induced electromotive force, in one complete revolution, passes through the following values. It starts at zero and rises to a maximum positive value, then falls to zero and rises to a maximum positive value in the reverse direction, finally falling to zero again. The transitions through these various values constitute a *cycle* (q.v.). The time of this transition is called a *period,* the number of cycles per second being the *frequency* (q.v.). The angle at any instant through which the conductor has rotated since the electromotive force changed its sign from negative to positive is the *phase angle.* This is ordinarily designated by the Greek letter Φ (phi). If ω (omega) is the angular velocity of the rotating conductor and f the frequency, then $\omega = 2\pi f$ (since there are 2π radians in a complete revolution). If the time in seconds which has elapsed since the conductor last passed the point of zero electromotive force be represented by t, then the phase Φ at that particular instant is equal to $\omega t = 2\pi ft$. Periodic variations are conveniently expressed by sine curves and the relative phases are indicated by the relative positions of the *nodes* (q.v.) and *loops* (q.v.).

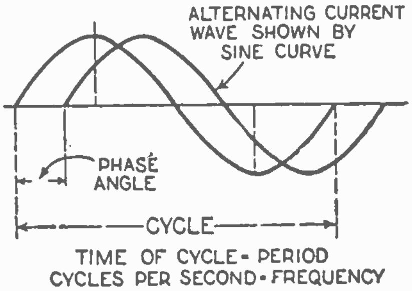
ALTERNATING CURRENT WAVE SHOWN BY SINE CURVE
PHASE ANGLE
CYCLE
TIME OF CYCLE = PERIOD
CYCLES PER SECOND = FREQUENCY

Phase displacement, cycle, etc., illustrated by means of sine curves.

Magnitudes having maximum and minimum values occurring simultaneously are said to be *in phase.* If they do not occur simultaneously, they are *out of phase.* When the corresponding maximum or minimum values of two sinusoidal alternating quantities of the same frequency occur at different instances, the two quantities are said to differ in phase by the angle between their nearest negative maxima or their nearest positive maxima or any other nearest corresponding values. The quantity whose maximum value occurs first is said to *lead* (q.v.) the other. The quantity whose maximum value occurs later is said to *lag* (q.v.). Where one quantity lags behind another by 180 degrees, the values are said to be of *opposite phase.* An alternating current may be *single phase* or *polyphase.* Another term for polyphase is *multiphase.* In a three-phase alternating current system the three separate currents differ in phase from each other by 120 degrees. In a polyphase system, if the various phases each carry or supply equal current, the system is said to be *balanced.*

PHASE ANGLE—See *Phase.*

PHASE DISPLACEMENT—The difference in phase or the phase angle between two alternating quantities, such as voltage and current, of the same frequency. Such quantities do not pass through their respective maximum positive quantities or through their zero values at the same time. The instantaneous values of any sine-shaped wave can be represented as being proportional to the sine of an angle, a complete cycle being equal to 360 degrees. When two quantities are out of phase, the respective angles corresponding to each quantity at any given instant are unequal, the difference between the two angles being called the *phase displacement* or the *phase difference.* In the accompanying illustration, the current and voltage relationship in an alternating current circuit are shown by the two sine curves. As can be seen, the current and voltage are out of phase, the phase displacement being equal to the angle Φ, in this case 90 degrees ($\frac{\pi}{2}$ radians). The current reaches its maximum value at a later time than the voltage. In this case the current *lags* behind the voltage by the angle Φ, the voltage leading the current by this angle. Since the angle of lag of the current is 90 degrees, the current and voltage are said to be in *quadrature.* In an alternating current circuit having pure inductance and no resistance (a theoretical circuit, since all electrical conductors possess resistance), the current would lag by 90 degrees as in the example given. (See *Lag, Lead, Leading Current,* also *Phase.*)

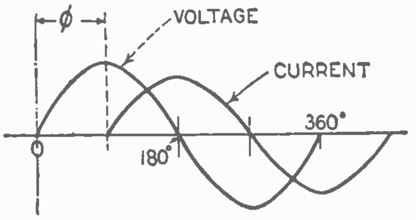

Sine curves of current and voltage, showing difference in phase.

PHENOMENA OF ELECTRIC WAVE PROPAGATION—Electric waves are created by electric *oscillations* (q.v.). They are often referred to as *radio waves.* Electric waves, light waves and radiated heat waves are included in the general term *electromagnetic waves* (q.v.). The velocity of electric waves is the same as that of light waves, 300,000,000 meters per second or 186,000 miles per second (appx.). Electric waves used for radio work have frequencies from about 10,000 to 3,000,000 cycles per second. Since the velocity of electric waves is known, it is possible to calculate the length of a wave if the frequency is known. Thus if the electric wave has a frequency of 10,000 cycles per second, the wavelength is found by dividing the velocity in meters per second by the frequency in cycles per second, which gives in this case a wavelength of 30,000 meters. Electric waves are in reality a combination of electromagnetic and electrostatic disturbances in space. The displacements are at right angles to the motion of the wave train. The electric field and the magnetic field are at right angles and travel together. They spread out or expand from the point of disturbance just as water waves spread out from a point where a stone strikes the surface of the water. Electric waves are propagated from an aerial when certain forms of alternating current flow in the aerial. Since the frequency of the propagated wave corresponds to the frequency of the current in the aerial, high frequency current must flow in the aerial to produce the high frequency electric waves used in radio. Such current may be produced by a high-frequency generator, by a direct current generator and an electric arc, or by means of vacuum tubes. The latter offer the simplest means of producing the rapidly reversing currents. Electric waves may be classified as *continuous* and *discontinuous.* Examples of the former are the waves produced by a high frequency alternator, an oscillating vacuum tube or a Poulsen arc. Examples of the latter are the waves produced by condenser discharges in a spark circuit. In these the amplitude of the waves diminishes in each wave train and the waves are referred to as *damped* waves. The high frequency current used to produce continuous waves used in radio telephony is *modulated* in accordance with the vibrations of a *microphone* (q.v.). As a result modulated waves are sent out from the aerial and these correspond in wave form to the variations of the sound waves. The sensitive apparatus at the distant receiving set picks up or intercepts the modulated electric waves, reconverts them to high frequency electric currents which are rectified into direct currents capable of actuating a 'phone and usually amplified so that they have energy enough to operate a loud speaker. (See *Oscillator, Carrier Wave,* also *Modulated Currents.*)

PHILLIPS, Raymond—British radio expert. Born October 6, 1879, he was educated at Edgebaston, Birmingham Schorne College, Buckinghamshire, and Windsor High School. For three years he was engaged on railway constructions and repairs, with special application to electric railways. In 1902 he invented a system of automatic train control and a number of appliances for electric railways which were widely adopted. In 1905 he took up the study of wireless and specialized in the control of mechanisms at a distance by electric waves. In 1910 he patented a system for controlling airships by radio and by means of a working model demonstrated how an airship could be thus controlled. During the World War he was inspector of ordnance machinery and in 1921 was appointed a member of the Inter-Allied Commission of Control in Germany. He is the author of many articles on the control of machines by electric waves.

PHONOGRAPHIC RECORDING—This refers to the amplifying of code signal currents by means of relays in tandem until the sounds produced in a telephone receiver are loud enough to cut a distinct record on the wax of a phonograph. The signals can be received at high speed and thereafter read at low speed. Instead of a phonograph a telegraphone may be used.

PHOTO-ELECTRIC EFFECT—A change in the electrical conductivity of a gaseous or solid substance when radiations of certain wavelengths come in contact with them. This refers particularly to rays of the visible spectrum and ultra-violet and infra-red rays.

PHOTOGRAPHIC RECORDER—A device for recording high-speed wireless messages by recording the deflections of a sensitive galvanometer on a moving photographic film. (See *Auto-Receiver.*)

PICKARD, Greenleaf Whittier—American radio expert. He was born at Portland, Maine, on February 14, 1877, and was educated at Westbrook Seminary, Lawrence Scientific School, Harvard and Massachusetts Institute of Technology. He made a special study of wireless telegraphy and telephony and has taken out many United States and foreign patents for radio inventions. He is noted for his pioneer work in radio telephony. Pickard began radio work in 1899 at Blue Hill observatory, Milton, Massachusetts, under a grant from the Smithsonian Institute. He was on the engineering staff of the American Telephone and Telegraph Company from 1902 to 1906. Later he was connected with the Wireless Specialty Apparatus Company as consulting engineer. He has an extensive practice as a patent expert in radio patent litigation. Pickard is a fellow of the American Institute of Electrical Engineers, a Member of the American Chemical Society and also is a Member of the Society of Chemical Industry.

Greenleaf Whittier Pickard.

PIG-TAIL—A flexible braided or stranded copper conductor used to carry the current to the rotor of a condenser or a coupler. Pig-tails are also used on the carbon brushes of motors and generators. (See *Flexible Lead.*)

PITCH — Frequency of vibration of a sound. A shrill note is said to be of high pitch, a bass note of low pitch. The pitch of an armature winding refers to the distance from the center of a winding to the center of the next winding.

PLAIN AERIAL—An aerial having the transmitting or receiving circuit connected directly to it without the utilization of inductive coupling or any intermediate tuned circuit. An example of a plain aerial is one used to transmit code, having a spark gap in series with the aerial and the ground.

PLATE — The anode or positive electrode of a vacuum tube. The vacuum tube plate is usually made of nickel. The shape of the plate varies according to the type and construction of the tube. In some tubes the plate is tubular; in others flat. When the cathode or filament is heated, it emits electrons which pass to the grid or through the grid to the plate, depending upon the grid charge. In power tubes (transmitting) it is necessary to use some cooling device to conduct away the heat dissipated at the plate. In some cases a liquid bath or a blast of air is used or the plate may be constructed so that it can be cooled by water circulation. Other methods of preventing overheating of these tubes are increasing the area of the plate as for example increasing the diameter of a tubular plate, or blackening the surface of the plate to increase its heat emissivity.

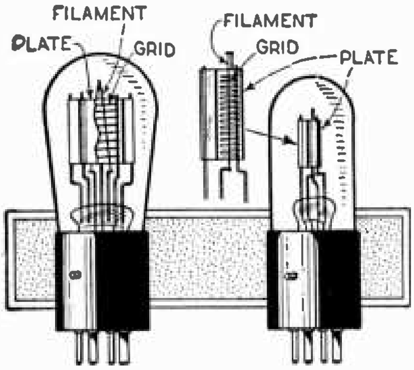

Vacuum tubes, showing construction of plates.

PLATE BATTERY—A battery, usually referred to as the *"B" Battery* having its positive terminal connected to the plate of the three-electrode vacuum tube and its negative terminal leading to the filament. The purpose of this battery is to keep the plate voltage positive with respect to the filament so that the electrons will be attracted sufficiently to the plate. Plate batteries are usually dry cells although there are a number of storage plate or "B" batteries on the market, both of the acid and the alkaline types. It is possible to dispense with the plate battery by using a *"B" eliminator*, which is connected to the ordinary house-lighting circuit. (See *"B" Battery.*)

PLATE CIRCUIT — The circuit connected to the *plate* (q.v.) or anode of a three-electrode vacuum tube.

PLATE COMPONENT, HIGH FREQUENCY—The radio frequency current flowing in the *plate circuit* (q.v.) of a vacuum tube. (See *High Frequency Component of Plate.*)

PLATE CONSUMPTION — This refers to the current consumed in energizing the plate of a vacuum tube. Increasing the grid voltage tends to increase the plate consumption. An increasing plate voltage also increases the plate current, but tends to decrease the relatively small grid current. (See *Consumption, Current;* also *Consumption, Plate.*)

PLATE CONTROL — The vacuum tube starts oscillating at a plate voltage which will depend on the adjustments of the circuit. When the plate voltage is raised, the oscillation current becomes greater, increasing in direct proportion.

PLATE CURRENT—The current in the *plate circuit* (q.v.) of the three-electrode vacuum tube. The current passing between the plate and the heated cathode or filament.

PLATE IMPEDANCE — The total resistance offered to alternating current between the filament and the plate in a vacuum tube. It is often referred to as the *tube impedance.* The impedance is independent of the alternating current input voltage.

PLATE VOLTAGE—The voltage of the plate of a vacuum tube above the filament. The potential difference between the plate and the filament. This term is usually used to refer to the voltage of the *"B" battery* (q.v.) connected to the plate, as for example, if the 22½-volt terminal of the "B" battery is connected to the plate of the detector tube, the plate voltage is said to be 22½ volts. Soft tubes, formerly used as detectors, operated best at this voltage. Most of the amplifier tubes now used as detectors operate best with 45 volts on the plate, although these tubes vary individually, some operating better with slightly increased voltage and others with a voltage less than 45. The plate voltage generally used on the amplifiers is 90 volts. The new power reception tubes use plate voltages up to 135 volts.

PLIODYNATRON — A special vacuum tube, designed by A. W. Hull, having two grids instead of one. This tube utilizes the normal amplifying property of the vacuum tube to obtain a very high voltage amplification, as high as 1,000 fold. When such high amplification is obtained, however, the operation is unstable and extremely careful adjustment is necessary.

PLIOTRON—A name given to a highly evacuated vacuum tube used for transmitting purposes. Large pliotrons are capable of developing as high as 500 watts output since they are so highly exhausted that several thousand volts may be applied to the plate and also since they can carry a plate current of nearly one-half an ampere.

PLUG — A connecting device used extensively in radio and telephone work for making connection or "plugging in" *loud speakers* (q.v.), head sets, power supply, microphone transmitters, etc. The plug has an insulating handle and the metal contact tip is designed to fit into a standard *jack* (q.v.). A flexible wire fits into terminals in the plug, the other end of the wire being connected to the apparatus such as the loud speaker, etc. The plug connection terminals may be of the conventional small binding post type or they may consist of internal spring clips.

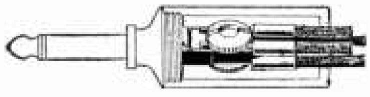

Cross-section of plug.

PLUG-IN COIL — An interchangeable coil, often of the honey-comb type, which can be plugged into a socket, thus permitting it to be quickly and conveniently put into or taken out of a circuit. Plug-in coils are used in some cases in receiving sets designed for operation on widely varying wave bands.

PLUG-IN TRANSFORMER—*A plug-in coil* (q.v.) specifically used as a radio frequency transformer.

POLARITY — In magnets the differentiation between *North Pole* and *South Pole.* In electrical devices or circuits, the difference between *positive* and *negative poles.* It is often necessary to determine which is the positive pole or terminal of a direct current circuit, and which the negative. There are various methods of indicating polarity. A voltmeter or an ammeter will show the direction of current flow. An extremely simple method of determining polarity is to place the two terminals in a glass jar containing salt water or some other electrolyte. A much greater number of bubbles will be observed to collect around the terminal of negative polarity. A piece of used blue print paper can be utilized to indicate polarity. It is simply necessary to moisten the paper and place the terminals on it, at a distance of about ¼ of an inch apart. The paper will be whitened at the negative terminal.

POLARIZATION — In a primary cell, the current liberates hydrogen from the electrolyte in passing through it. The hydrogen is carried to the negative plate with the current and collects there. It thus increases the internal resistance of the cell. This makes it harder for the current to pass through it and as a result the electromotive force available at the terminals is materially lowered. In addition there is a slight difference of potential between the negative plate and the hydrogen and this acts in opposition to the electromotive force of the cell. This further reduces the effective voltage of the cell. The entire effect is known as polarization. Polarization may be carried to such an extent that current can no longer flow from the cell. It may be counteracted or prevented by the use of *depolarizers* (q.v.). There are three different kinds of depolarizers. Chemical depolarizers are placed in the cell near the cathode. Substances such as manganese dioxide, potassium bichromate, nitric acid or cuprous chloride are used. When the hydrogen reaches the depolarizer it combines with it to form a new compound. Electrochemical depolarizers depend for their action upon the use of a substance which will liberate a metal such as copper at the cathode instead of hydrogen. Thus with no hydrogen present there is no polarization. Of course it is necessary to use the proper electrodes and the correct electrolyte to obtain this result. Mechanical depolarizers merely blow or sweep the hydrogen away from the cathode. Being the least effective of the three methods, the latter is seldom used. (See *Local Action.*)

POLARIZATION, ELECTRIC—The state of a dielectric when subjected to electrostatic forces. Electric polarization is synonymous with electric displacement.

POLARIZED RELAY — A relay in which a magnetized swinging arm is placed between poles of two electromagnets. When current passes, one pole must change, so that the arm is attracted by one and repelled by the other. A polarized relay is much more sensitive to weak currents than the ordinary relay.

POLES—See *Magnetic Poles.*

POLYPHASE—Referring to an alternating current system in which the circuits are divided up into two or more branches having their currents displaced in phase from each other. The term *multiphase* has the same meaning as polyphase. (See *Phase.*)

POLYPHASE ALTERNATOR—An alternating current generator wound in such a way that it supplies two or more currents which differ in *phase* (q.v.) from each other by definite phase angles. The usual form of polyphase alternator is the three-phase alternator.

PORCELAIN CLEAT—See *Cleat.*

POSITIVE ELECTRODE—In a primary cell, the zinc plate or *anode* (q.v.) is the *positive electrode,* while the *pole* of this electrode is the *negative pole* (q.v.) because it is negative in relation to the external circuit. In a storage cell, the lead peroxide plate, which is the cathode during discharge, is called the *positive electrode* and its pole the *positive pole.*

POSITIVE TERMINAL POLE—The pole or terminal out of which current is considered to flow. This is more or less of a handy convention, but is contrary to the present day knowledge of electrical flow inasmuch as electrons move from negative to positive.

POTENTIAL—See *Potential, Electric.*

POTENTIAL, DIFFERENCE OF—abbreviation P.D.—Difference of electrical level. Theoretically the potential difference between two points in an electric field is measured by the work done by the electric field on a unit point charge of positive electricity moved from one point to another without disturbing the field. Potential difference is measured in volts. It may refer to any two points in a circuit, even though there is no source of electromotive force between these points. When applied to circuit between two points not possessing a source of electromotive force, the potential difference between the two points under consideration is equal to the product of the current flowing times the resistance of the circuit between the two points. In the case of an open circuit, such as the terminals of a cell, the potential difference is equal to the electromotive force. (See *Potential, Electromotive force,* also *Volt.*)

POTENTIAL DROP—The fall in voltage in a circuit due to the resistance of the conductor. (See *Potential, Electric;* also *Potential, Difference of.*)

POTENTIAL, ELECTRIC—Electrical pressure, or the degree of *electromotive force* (q.v.). Potential is electrical level. Electrical potential can be compared with the "head" of water. Thus it is customary to measure the height or head of water above sea level, which is taken as an arbitrary base. In a somewhat similar manner, the potential of the earth is taken as zero and all potential is measured from this as an arbitrary base. With this understood, it is no longer necessary to refer to the degree of electromotive force, but simply to the electromotive force or the potential. See *Potential, Difference of,* also *Volt.*)

POTENTIAL ENERGY—Energy of position. When energy is available for the production of work, it is referred to as potential energy. Energy at work is *kinetic energy* (q.v.).

POTENTIAL RECTIFIER—A *crystal* (q.v.) or other form of rectifier which requires an initial current to pass through it to become sensitive.

POTENTIOMETER (Potential Divider)—A resistance connected across a battery or other source of electromotive force and equipped with a slider so that the voltage on another circuit may be varied from zero to the full potential of the battery. Where a gas content detector vacuum tube is used, the purpose of the potentiometer is to permit a close adjustment of the *plate battery* (q.v.) potential. The potentiometer is shunted around the filament heating or "A" battery, its resistance being so high that its power consumption is practically negligible. In tuned radio frequency sets, a potentiometer is used to control the tube oscillation. The advantage of being able to make a system oscillate is obvious, as the greatest amplification is obtained when the circuits are operated just at the point before self-oscillation starts. This condition is easily found in operation, since the instant self-oscillation starts, a whistle is heard in the phone or loud speaker when tuning to any station it is desired to receive. The use of a potentiometer to control oscillations in a tuned radio frequency circuit is illustrated in the diagram shown at Fig. 1. Potentiometers used in radio apparatus are usually made either from high resistance wire or from carbon granules. In electrical engineering work, the term potentiometer means, in general, an arrangement of circuits for measuring potential difference.

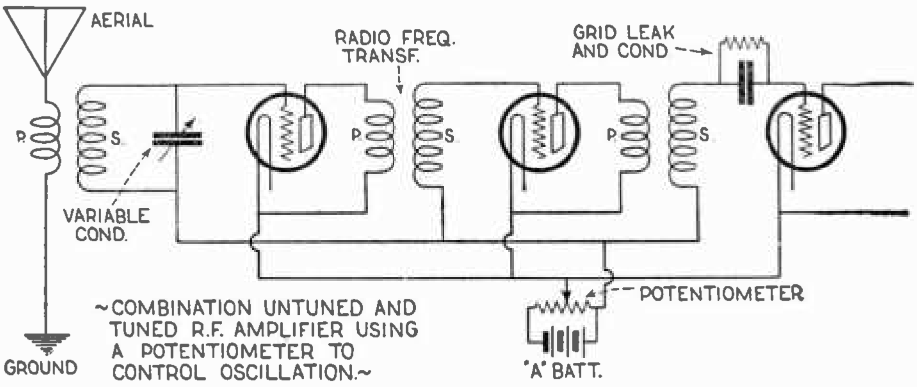
AERIAL
RADIO FREQ. TRANSF.
GRID LEAK AND COND
P
S
P
S
P
S
VARIABLE COND.
~COMBINATION UNTUNED AND TUNED R.F. AMPLIFIER USING A POTENTIOMETER TO CONTROL OSCILLATION.~
GROUND
POTENTIOMETER
'A' BATT.

Fig. 1. Tuned radio frequency circuit showing use of potentiometer to control oscillations.

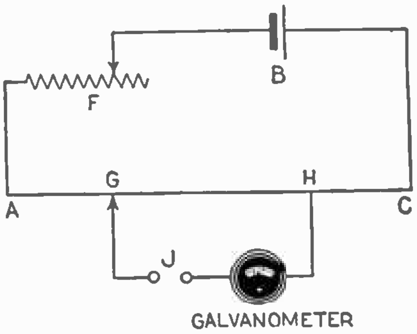

Fig. 2. A simple form of potentiometer used for electrical measurements.

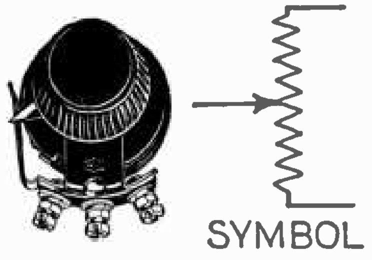

Illustration of a radio potentiometer, showing symbol at the right.

Using certain accessories, laboratory potentiometers can be used to measure voltages over all ranges, and by applying's *Ohm's law* (q.v.) the measurements obtained can be used to determine a wide range of current values.

The simplest form of potentiometer used for electrical measurements is shown in Fig. 2. AC is a resistance in which a constant current from the battery B is flowing. F is a regulating resistance used to compensate for the variations of the electromotive force of the battery. The current in AC is checked for constancy by noting that the drop in potential between the two points selected in it is equal to the electromotive force of a standard cell. The standard cell is connected in the circuit GHJ at J and the regulating resistance is adjusted until the sensitive galvanometer shows no deflection. With the assumption that AC has a uniform resistance throughout its length, and the current in it remaining constant, it is obvious that any other voltage not greater than the drop between A and C can be measured by connecting it in the circuit at J and shifting the points G and H until the galvanometer again comes to a balance.

A direct reading scale may be placed between A and C. For ordinary potentiometer measurements, the drop between A and C is made about 1½ volts as this is approximately the electromotive force of a standard Clark cell.

POULSEN ARC—An arc usually burned in hydrogen between a rotating carbon cathode and a water-cooled copper anode in a magnetic field. This arc is connected to an oscillating circuit, and due to its instability it maintains continuous oscillations which are utilized for the production of continuous waves for radio telegraphy and telephony.

POULSEN, Valdemar—Danish radio pioneer. Born in Copenhagen, November 23, 1869; he was educated at Copenhagen University and in 1893 he joined the Copenhagen Telephone Company. In 1903 he invented his famous system of arc transmission. Poulsen is a fellow of the Danish Society of Sciences. In 1900 he was awarded the Grand Prix at Paris.

POUNDAL — A unit of force. That force which acting on a mass of one pound gives it a velocity of one foot per second. It is equal to the weight of a pound mass divided by the acceleration of gravity; equal to 13,825 dynes or approximately half an ounce.

POWER—The rate of doing work. A certain amount of work may be done in a day or in a week. In the former case, more power is expended than in the latter case. Mechanical power is measured in *horse power* (q.v.). One horse power is equivalent to 33,000 foot-pounds per minute. Since power is the rate of doing work, the expression for power must contain a time element and a work element. Work is a measure of a force exerted through a distance. Hence the foot-pound measures work. Electrical power is the product of voltage times amperage. It is measured in *watts* (q.v.). (See *Kilowatt.*)

POWER AMPLIFIER—See *Power Amplification.*

POWER AMPLIFICATION—Amplification by means of apparatus consisting of a combination of power vacuum tubes, designed to carry larger currents than the ordinary tubes and specially designed transformers. Power amplifiers provide additional audio amplification and are used in connection with radio receiving sets where great volume is required as in the case of apparatus used for public address systems and for use with power operated loud speakers. (See *Amplifier.*)

POWER AMPLIFICATION COEFFICIENT—The ratio of the plate output power to the grid input power of a vacuum tube. This value may be as high as 10,000. The output power of a vacuum tube comes from the plate battery. When the grid is operated at low or negative voltage, the input power needed to produce the changes in grid voltage may be made very small compared with the output power of the tube.

POWER, APPARENT—The product of the voltage and the current in an alternating current circuit expressed in volt-amperes. This expression does not take into account any difference in *phase* (q.v.) between the voltage and the current. Thus in an extreme case (theoretical) where an alternating current flows in a purely inductive circuit, the current will lag behind the voltage by a phase angle of 90 degrees. In such a circuit, the *true power* (q.v.) would be zero, since the cosine of the phase angle between the current and the voltage would be zero, whereas the apparent power might be quite high and a heavy current might be flowing in the circuit. Apparent power is ordinarily expressed in *kilovolt-amperes* (q.v.). (See *Power Factor.*)

POWER FACTOR — The ratio of the *true power* (watts) in an alternating current circuit to the *apparent power* (volt-amperes). Where the wave form of the alternating current is sinusoidal, the power factor is equal to the cosine of the angle between the voltage and the current. Expressed as an equation, power P is equal to the product of the current I and the component of the impressed electromotive force which is in phase with the current, $E \cos \Phi$, where Φ is the angle of phase displacement between the current and the voltage ($P = IE \cos \Phi$). Expressed in another way, the power factor is the fractional amount which, if multiplied by the effective values of the current and the voltage, will give the value of the power expended. (See *Effective Electromotive force, Power, Apparent,* also *Power, True; Kilovolt-ampere, Kilowatt.*)

POWER, TRUE — The product of the *apparent power* and the *power factor* (q.v.) in an alternating current circuit. A wattmeter will measure true power, inasmuch as it takes into account the *phase displacement* (q.v.) between the current and the voltage. A voltmeter and an ammeter cannot be used to measure power, however, since the product of the two readings only gives the apparent power.

POWER TUBE—See *Vacuum Tube, Power.*

PRACTICAL UNITS—In the practical system of electrical units, the ohm, the ampere and the volt are the fundamental units. While the practical system is based originally on the Centimeter Gram Second systems, it is in no way dependent upon these systems for practical use. When used with the fundamental units of length and time, all other electrical as well as magnetic units can be derived from the fundamental units of the practical system. At the recommendation of the Chicago International Electrical Congress in 1893, the fundamental electrical units of the practical system were adopted as legal units in the United States by an act of Congress, July 12, 1894.

P R B—International Radiotelegraphic Convention abbreviation. Followed by question mark it means, "Do you wish to communicate by means of the International Signal Code?" Without the question mark it means, "I wish to communicate by means of the International Signal Code." (See *Abbreviations, International Radio Telegraphic Conventional.*)

PRESSURE—See *Potential, Electric.*

PRESSURE, ELECTRICAL—The term is synonymous with *voltage* (q.v.) or *electromotive* force (q.v.). Electrical pressure is the cause of current flow. It is measured in *volts* (q.v.). The flow of electricity along a conductor is analogous to the flow of water through a pipe. In order to make water flow, a pump is often used to force the water or in other words to apply pressure. Similarly, electrical pressure must be applied in order for an electric current to flow.

PRIMARY CELL—A device for transforming chemical energy into electrical energy. It consists of two different conductors placed in a liquid which acts chemically on one conductor, or more on one conductor than on the other, thus keeping a difference of potential between them. While the cell is being discharged either one or both of the conductors is used up and the liquid is also consumed. In many types of primary cells these can be replaced. The liquid is called an *electrolyte* (q.v.). In some cases an acid is used while in others a salt solution is the *electrolyte.* (q.v.) (See *Cell, Cell Secondary; Cell, Theory of Primary; Electrolysis, Local Action, Polarization,* also *Storage Cells.*)

PRIMARY CIRCUIT—A circuit supplying current to another which is called the secondary circuit. The two circuits are usually coupled by means of a transformer, the primary winding being in the primary circuit and the secondary winding being in the secondary circuit.

PRIMARY CURRENT — See *Current, Primary.*

PRIMARY ELECTRONS — The main stream of electrons given off by the hot filament of a vacuum tube, as differentiated from *secondary electrons* (q.v.) which may be detached from the plate by the impact of the primary electrons. (See *Negative Resistance.*)

PRIMARY TUNING INDUCTANCE—A variable inductance in the primary circuit of a radio transmitter. It is also referred to as the *aerial tuning inductance.* The aerial inductance coil may be continuously variable utilizing a sliding contact which bears on an edgewise copper strip or it may be of the drum type, in this case being made of stranded copper cable. (See *Inductance Coils,* also *Pancake Coils.*)

PRINCE, Charles Edmond — British radio expert. Born in 1874 at Cape Town and educated at Clifton and Faraday House, he joined the Marconi Wireless Telegraph Company in 1907, where he was concerned chiefly in research work on wireless telephony. In 1909 he carried out experiments in Italy and Switzerland with the first Marconi Field Station and made a number of valuable improvements in the Bellini-Tosi direction finder. In the World war, he joined the Royal Flying Corps and in 1915 he installed the first radio telephone on aircraft.

PRINTER, MORSE—See *Morse Inker.*

PROTON—A unit of positive electricity. The smallest quantity of positive electricity capable of existing in a free state. A minute particle or quantity of positive electricity carried by the nucleus of a hydrogen atom. See *Electron Theory,* also *Free Electrons.*

PULSATING CURRENT — A periodic current, that is to say a current passing through the successive equal cycles of values, the average value of which is not zero. A pulsating current is the sum of an alternating and a direct current. A current which varies regularly in magnitude is said to be pulsating while if it remains the same in magnitude it is called continuous. (See *Continuous Direct Current,* also *Current, Pulsating Direct.*)

PULSATING DIRECT CURRENT—See *Pulsating Current.*

PUPIN, Michael — American radio authority. Born in Hungary, October 4th, 1858, he came to the United States at the age of 10, and was educated at Columbia University. He also studied at Cambridge University, England. In 1891 he was appointed professor of mathematical physics at Columbia University, where he carried out an important series of researches in radio telegraphy and telephony. He is the holder of a large number of patents relating to telephony and radio, has done valuable research work on reso-

Michael Pupin.

nance and has developed radio circuits of extreme selectivity.

PUSH-PULL AMPLIFICATION—A method of connection in a vacuum tube receiving set, for supplying more power to the loud speaker than that ordinarily obtainable from a one or two stage audio amplifier. In the push-pull amplifier two tubes are used in the final stage of the amplifier, connected in such a way that they are used alternately on the two halves of each audio frequency cycle. A push-pull amplification circuit is shown in Figure 1. The first tube at the left gives the first stage of amplification and the other two in combination give the second stage. An ordinary audio frequency amplifier transformer is connected to the grid and filament of the first tube. Special transformers are usually used for the other two transformers. These have connections brought out from the center point of one of the windings as shown. The transformer between the two stages has a split secondary while the last transformer has a split primary. In cases where a tapped wound transformer is not available, two transformers can be used instead. In this case the two primary windings are connected in series and the two secondary windings are also connected in series. Figure 2 shows the method of connecting two transformers for this purpose, a connection being made to the mid-point between the secondary windings. It is essential that the two windings in series be connected so that they are continuous in the same direction around the transformer cores. In theory, the push-pull amplifier uses one tube during one half of the cycle and the other tube during the other half of the cycle. While the terminal of the secondary transformer winding leading to the grid of the tube is at a positive potential with respect to the middle of the winding, the tube is effective. During the remaining half of the cycle when the terminal is negative with respect to the tap, the tube is not operating. The action however is such that one tube is effective at all times. Although ordinary amplifier tubes can be used in the push-pull amplifier circuit it is desirable to use special power tubes which have a high amplification coefficient.

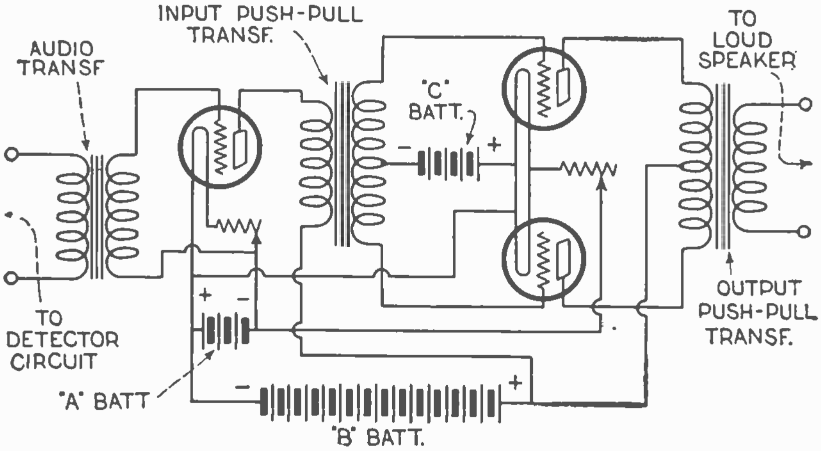

Fig. 1. Diagram of a push-pull amplification circuit.

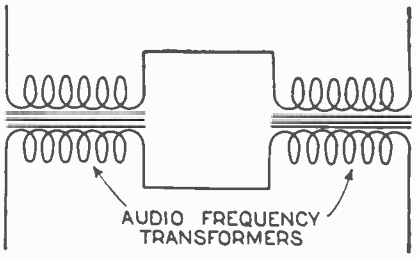

Fig. 2. Method of connecting two transformers where a tapped wound tranformer is not available.

PYRITES—Compounds of iron or copper and sulphur. Three of these are *Ferrous Sulphide* (FeS), *Bisulphide* (FeS_2), and an intermediate known as *Magnetic Pyrites* (Fe_3S_4). Pyrites crystals have been used in radio as detector crystals. (See *Crystal*, also *Crystal Detector.*)

PYRON DETECTOR—A *crystal detector* (q.v.) having an iron pyrites crystal and a copper or other metallic contact point.

Q

Q SIGNALS—International Radiotelegraphic Convention abbreviations. When the signal is followed by a question mark it asks the question. Otherwise the answer is indicated. A list of these abbreviations and their meanings follow:

Abbrev'n	*Question*	*Answer*
QRA	What ship or coast station is that?	This is . . . .
QRB	What is your distance?	My distance is . . . .
QRC	What is your true bearing?	My true bearing is . . . degrees.
QRD	Where are you bound for?	I am bound for . . .
QRF	Where are you bound from?	I am bound from . . . . .
QRG	What line do you belong to?	I belong to the . . . Line.
QRH	What is your wave length in meters?	My wave length is . . . . meters.
QRHH	What tune shall I adjust for?	Adjust to receive on tune . . . .
QRJ	How many words have you to send?	I have . . . . words to send.
QRK	How do you receive me?	I am receiving well.
QRL	Are you receiving badly? Shall I send 20 . . . — . for adjustment?	I am receiving badly. Please send 20 . . . — for adjustment.
QRLL	Request permission to test . . . minutes.	Permission to test granted.
QRM	Are you being interfered with?	I am being interfered with.
QRN	Are the atmospherics strong?	Atmospherics are very strong.
QRO	Shall I increase power?	Increase power.
QRP	Shall I decrease power?	Decrease power.
QRQ	Shall I send faster?	Send faster.
QRS	Shall I send slower?	Send slower.
QRT	Shall I stop sending?	Stop sending.
QRU		I have nothing to transmit. I have nothing for you.
QRV	Are you ready?	I am ready. All right now.
QRW	Are you busy?	I am busy (or, I am busy with . . . . Please do not interfere).
QRX	Shall I stand by?	Stand by. I will call you when required.
QRY	When will be my turn?	Your turn will be No. . . . .
QRZ	Are my signals weak?	Your signals are weak.
QSA	Are my signals strong?	Your signals are strong.
QSB	Is my tone bad? Is my spark bad?	The tone is bad. The spark is bad.
QSC	Is my spacing bad?	Your spacing is bad.
QSD	What is your time?	My time is . . . .
QSF	Is transmission to be in alternate order or in series?	Transmission will be in alternate order.
QSG		Transmission will be in a series of 5 messages.
QSH		Transmission will be in a series of 10 messages.
QSJ	What rate shall I collect for?	Collect . . . .

Abbrev'n	*Question*	*Answer*
QSK	Is the last radiogram cancelled?	The last radiogram is cancelled.
QSL	Did you get my receipt?	Please acknowledge.
QSM	What is your true course?	My true course is . . . . . degrees.
QSN	Are you in communication with land?	I am not in communication with land.
QSO	Are you in communication with any ship or station (or: with . . . .)?	I am in communication with . . . . (through . . . . .
QSP	Shall I inform . . . . that you are calling him?	Inform . . . . that I am calling him.
QSQ	Is . . . . . calling me?	You are being called by . . . .
QSR	Will you forward the radiogram?	I will forward the radiogram.
QSS	Are my signals fading?	Your signals are fading.
QST	Have you received the general call?	General call to all stations.
QSU	Please call me when you have finished (or: at . . o'clock).	Will call when I have finished.
QSV	Is public correspondence being handled?	Public correspondence is being handled. Please do not interfere.
QSW	Shall I increase my spark frequency?	Increase your spark frequency.
QSX	Shall I decrease my spark frequency?	Decrease your spark frequency.
QSY	Shall I send on a wave length of . . . . meters.	Let us change to a wave length of . . . . meters.
QSZ		Send each word twice. I have difficulty in receiving you.
QTA		Repeat the last radiogram.
QTC	Have you anything to transmit?	I have something to transmit. I have one or more radiograms for . . . . . .
QTE	What is my true bearing?	Your true bearing is . . . . degrees from . . . . . .
QTF	What is my true position?	Your true position is . . latitude . . . . . .longitude.

QUADRANT—A term formerly in use as the unit of *inductance* (q.v.), now called the *henry* (q.v.). Inductance has the same dimensions as length and this unit is equivalent to 10^9 cm., this being approximately the length of one quadrant of the earth.

QUADRANT ELECTROMETER—A sensitive *electrometer* (q.v.) which in its calibrated form is known as an *electrostatic voltmeter* (q.v.). In the usual type, the "needle" consists of a thin aluminum plate suspended on delicate knife edges, with a pointer extending from the upper part to a scale. Two pairs of quadrant plates, having opposite faces metallically connected together, are placed on either side of the movable plate and parallel to its face. Each pair is insulated from the other pair. When the pairs of quadrants are subjected to a difference of potential, the needle, if highly charged, will be attracted into one pair or repelled out of the other. (See *Kelvin's Electrostatic Voltmeter.*)

QUADRATURE—The differing in *phase* (q.v.) of two alternating current quantities by 90 degrees. For example, when the current lags behind the impressed voltage by 90 degrees, as in the case of pure inductance in a circuit, the current and the voltage are said to be in quadrature. (See *Lag, Lead,* also *Leading Current.*)

QUANTITY OF ELECTRICITY—When referring to static electricity, it is the amount of the charge. Referring to electricity in motion, or an electric current, it is the amount of current flowing in a given time. Quantity of electricity is measured in *coulombs* (q.v.). The coulomb is the quantity of electricity which has passed when a current of one ampere flows for one second. The symbol which is usually used to signify quantity of electricity is Q.

QUANTOMETER—A type of ballistic galvanometer in which the swing of the movable system is proportional to the quantity of electricity passing through the instrument. A *galvanometer* (q.v.) of this type is used for magnetic testing as in the case of the *fluxmeter* (q.v.).

QUARTER PHASE—A two-phase alternating current system. An alternating current system having two currents differing in phase by 90 degrees or in other words by one-quarter of a cycle.

QUARTZ LAMP—A mercury vapor lamp which utilizes a tube of quartz instead of a glass container. Increased current densities can then be employed, since higher temperatures can be used with quartz than with glass.

QUARTZ OSCILLATOR—The piezo-electric quartz plate may be used as a *resonator* (q.v.) or as an *oscillator* (q.v.). When used as an oscillator it may be used either as a *master oscillator* controlling the frequency of a radio transmitting station's output or as a frequency indicator. All three of these uses are applicable to the work of maintaining station frequencies constant. The applicability of the piezo oscillator as a frequency indicator in broadcasting stations has been tested carefully by the United States Bureau of Standards, and found to be useful and satisfactory. This method gives a means of maintaining an extremely accurate check of the transmitting station frequency. (See *Master-Oscillator System,* also *Frequency Meter.*)

QUENCHED GAP—See *Quenched Spark Gap.*

QUENCHED SPARK — See *Quenched Spark Gap.*

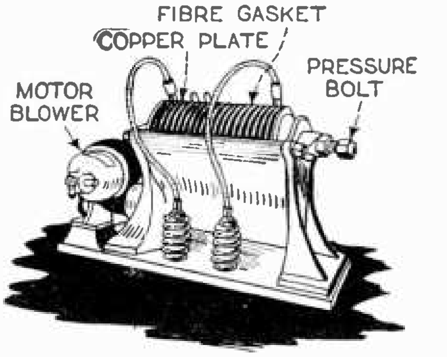

A typical quenched spark plug.

QUENCHED SPARK GAP—A spark discharger in a wireless telegraph transmitting circuit in which the spark is deionized to prevent the gap remaining conductive after the spark has passed. In this way only a few sharply defined oscillations are permitted to occur. The particular advantages of the quenched spark gap is that it enables the aerial to oscillate at its own frequency regardless of the period of the spark system. In one form of construction, a number of heavy copper plates are used, separated by micanite, fiber insulating washers or other insulating material. These are placed in an iron rack and compressed by means of a pressure bolt. The inside edge of the washer rests upon a groove which is cut in each plate. This makes it impossible for the spark to discharge over the edge of the washer. In order to make the discharge surface airtight, the washers are specially treated and as a result this helps in quenching the primary oscillations and makes the discharge noiseless. A small motor driven blower is mounted at the base of the spark discharger for the purpose of forcing air through the base of the rack against the cooling flanges of the plates. (See *Deionization, Musical Spark, Kinraidy Spark Gap,* also *Quenched Spark Transmitter.*

QUENCHED SPARK TRANSMITTER—A radio telegraph transmitter which employs a quenched spark gap. The quenched spark transmitter is not used for very high powers as it is not satisfactory for such use. In this case, a form of disc discharger is usually used. The quenched spark transmitter, however, has no moving parts and is noiseless in operation. In addition it gives large values of antenna current because closer coupling is possible. (See *Quenched Spark Gap.*)

QUICK BREAK SWITCH—A switch provided with auxiliary contacts which are released after the main contacts. The auxiliary contacts are separated rapidly, regardless of the speed with which the switch handle is pulled. This action is obtained by the tension of a spring connected between the main switch blade and the auxiliary blade, as shown in the illustration.

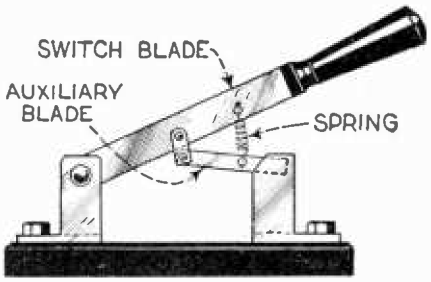

Quick Break Switch. Note the auxiliary spring-controlled blade.

QUIESCENT AERIAL—In a radio telephone transmitter, an aerial operated in such a manner that the carrier wave is suppressed when transmission is not actually taking place.

R

"R"—Symbol for *resistance* (q.v.)

RADIANT ENERGY—Energy transmitted through space by means of vibrations of the ether, such as electromagnetic energy, light or heat. (See *Actinic Ray.*)

RADIANT HEAT—Infra-red *radiation* (q.v.) as for example radiation from substances at temperatures insufficient to cause visible light rays.

RADIATING CIRCUIT—An electrical circuit which emits energy in the form of ether waves. In a transmitting system, the aerial circuit. Any circuit which radiates energy is a radiating circuit.

RADIATION—The transference of energy in waves through space which is not necessarily occupied by matter. Transference of energy through *ether* (q.v.). Electric or electromagnetic radiation refers to the radiation by means of electromagnetic induction. The radiation field is transmitted by wave motion. Various types of electromagnetic waves may be sent out by a radiating system. In one system the waves are known as *damped waves* (q.v.), in another *undamped waves* (q.v.) are used.

RADIATION EFFICIENCY—The radiation efficiency of an antenna at a given wave length is the ratio of the power radiated to the total power delivered to the antenna (resistance).

RADIATION FROM ANTENNA—In the case of a simple vertical antenna, electromagnetic waves are radiated in semi-loops since the lower end of the aerial is grounded. Fig. 1 gives an idea of the way in which the intertwined electric and magnetic fields radiate from the antenna. It should be noted that the electric and the magnetic fields are in planes which are at right angles to each other. The electric lines are shown in elevation while the magnetic lines appear in plan, with the wave form common to both being shown between. At any single point in space, the lines separate and come together like a bellows. After the wave has travelled a certain distance, the lines can be regarded as sections of planes and the electric lines can be represented by the diagram shown in Fig. 2. The magnetic lines can be shown by the same figure, provided this is considered to have been rotated about its horizontal axis through an angle of 90 degrees. Regardless of whether the aerial is of the open or closed coil type, the radiation will take place in the same general way, *loops* (q.v.) being formed and detached as described. (See *Electromagnetic Waves,* also *Height of Aerial.*)

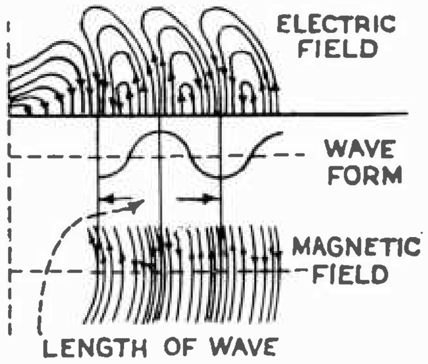

Fig. 1. Electric and magnetic fields about a vertical wire aerial.

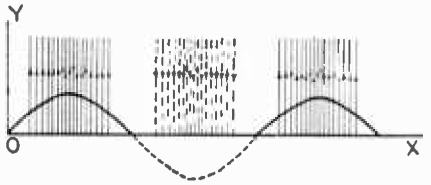

Fig. 2. Electric lines of force shown after the wave has travelled some distance from the aerial.

RADIATION RESISTANCE—The ratio of the total power radiated by an antenna to the square of the effective current at the point of maximum current. The property of an antenna by means of which radiation of energy takes place. Some of the power supplied to a circuit, carrying radio currents, is radiated in the form of electric waves. The radiation is a measure of the useful work obtained from the circuit. The greater the power radiated in comparison with the power dissipated in the circuit itself, the better will be the transmission. At any frequency, the power radiated is proportional to the square of the current flowing, hence the radiative effect may be regarded as causing artificially, an increase in the effective resistance of the circuit. This resistance increase is known as "radiation resistance" and is directly proportional to the square of the frequency or inversely proportional to the square of the wave length. (See *Antenna Resistance.*)

RADIATION TRANSMISSION—See *Radiation from Antenna.*

RADIO—The transmission of intelligence by means of electromagnetic waves through the ether. Radio falls under two general classifications; *radio telephony* (q.v.) and *radio telegraphy* (q.v.). (See *Radio Channel, Radio Communication, Radio Control,* also *Radio Frequencies.*)

RADIO-ACTIVITY—The emanation of energy by certain substances, such as radium, caused by a continuous detachment of electrons from the substance.

RADIOCAST—A term sometimes used instead of *Broadcast* (q.v.). To transmit speech, music or other sounds by *radio telephony* (q.v.).

RADIO CHANNEL—A band of wave lengths or frequencies of a width sufficient to permit of its use for radio communication without the radiation of subsidiary waves of more than a certain intensity at wave lengths or frequencies outside of such a band.

RADIO COMMUNICATION—The transmission of intelligible signals by means of electromagnetic waves. The signals may be in the form of code as in radio telegraphy or in the form of speech or music as in radio telephony. They may be transmitted through the *ether* (q.v.) without the use of intervening conductors as in ordinary radio, or along conductors as in *wired wireless* (q.v.).

RADIO COMPASS—A simple form of radio *direction finder* especially used for marine work. The radio compass depends for its principle of operation upon the directional effect of a coil antenna. Hence a coil antenna mounted on a vertical axis so that it can be rotated freely, is an essential part of the radio direction finder. By rotating the coil while a particular station is transmitting, it is possible to fix the line of direction of the station. The coil may be mounted so that the maximum signal will be received when the transmitting station lies in the plane of the coil. The device may also be set for minimum signal reception in which case the transmitting station lies in a direction perpendicular to the face of the coil. It is possible to obtain a much sharper determination of direction by setting the coil on the minimum position and hence this method is generally used for direction finder work. (See *Goniometer.*)

RADIO CONTROL—The control of motor driven devices such as marine vessels, trains, torpedoes, automobiles, etc., by means of special radio receiving apparatus designed to respond to radio impulses. E. P. Glavin's radio controlled automobile is an example of the radio controlled vehicle. *John Hays Hammond, Jr.* (q.v.) has invented radio controlled torpedoes, radio controlled ships and also radio controlled air craft.

R.F.—Abbreviation for radio frequency.

RADIO FREQUENCIES—Frequencies higher than those corresponding to normally audible sound waves. (See *Frequency, Radio; Frequency, High; Audio Frequency;* also *Frequency, Low.*)

RADIO FREQUENCY COILS—*Inductance coils* (q.v.) used to carry radio frequency currents or to couple radio frequency circuits. Radio frequency *transformers* (q.v.) may be of either iron core or air core types although the air core transformers are more generally used. (See *Low Loss Coils,* also *"D" Coils.*)

RADIO FREQUENCY CURRENT—An alternating current having a frequency of from 20,000 to 2,000,000 cycles per second. There is no hard and fast rule regarding the upper and lower limits, since much higher or much lower frequencies may still be considered as radio frequency currents. (*See High Frequency Current.*)

RADIO FREQUENCY SELECTIVITY—The radio frequency selectivity of a simple element of a receiving system is the ratio of resonant response (in terms of effective voltage or current measured at the indicator) to the nonresonant response when the radio frequency portions of the elements of that system are detuned by one percent of the resonant frequency. In this case, a simple element refers to a combination of an inductance, a capacitance and optionally a resistance.

RADIO FREQUENCY SIGNALS—Currents flowing in a radio receiving circuit up to the point of rectification. The currents flowing in the antenna circuit are radio frequency currents corresponding to the currents flowing in the transmitting aerial. (See *Amplifier, Radio Frequency.*)

RADIO GONIOMETER—See *Goniometer.*

RADIOGRAM—A message sent in code by radio. A radio telegraphic message.

RADIOPHARE—A radiotelegraphic "lighthouse" intended to aid navigation by emitting characteristic signals. By estimating the bearings of two charted radiophares, the mariner may determine the position of his ship.

RADIO TELEGRAPHY—The transmission of messages by means of electromagnetic waves using dot and dash code signals. A transmitter furnishes a source of high frequency alternating current which oscillates rapidly in the antenna system thus sending out electromagnetic waves. These are intercepted at the receiving station by the receiving aerial. At the receiving aerial, the electromagnetic waves are changed back into alternating current which is rectified by a detector in the receiving apparatus, then amplified and made audible in a *head telephone* (q.v.) or *loud speaker* (q.v.). Of course, a means of interrupting the current to produce dots and dashes must be present in the transmitter and suitable tuning devices are used in both the transmitter and the receiver.

RADIO TELEPHONY—The transmission of speech, music or other sounds by means of electromagnetic waves. In the standard systems of radio telephony continuous waves are used of a much higher frequency than those of audible sounds. The amplitude of these is modulated by the connection of a microphone in the transmitting circuit. The receiving apparatus is similar to that employed in radio telegraphy and is arranged so that the head set or loud speaker responds to the modulations of the waves and reproduces the sounds as originally produced at the transmitter.

RADIO TRANSMISSION—The sending of signals, speech, music, or other sounds by means of electromagnetic waves. The transmission of intelligence through the ether. (See *Radio Communication, Radio Control, Radio Telegraphy,* also *Radio Telephony.*)

RAT-TAIL—A single wire, or a group of wires used to connect the main part of an aerial with the *lead-in* (q.v.) wire.

RATIO—Relative values of quantities of the same kind, or number of times one quantity is contained in the other.

RATIO OF TRANSFORMATION—The ratio between the primary and secondary windings of a transformer. This determines the ratio of the primary to the secondary voltages. Thus if a transformer has a ratio of transformation of 3 to 1, then the secondary will have three times as many turns as the primary and the secondary voltage will be, approximately triple that of the primary voltage.

RATIO OF UNITS—The ratio of the unit pole in the electromagnetic system to the unit pole in the electrostatic system, or of the unit charge in the electromagnetic system to the unit charge in the electrostatic system. Its numerical value is approximately 3 x 10^{10}.

RAYLEIGH, John William Strutt—British physicist. Born at Langford Green, Essex, November 12th, 1842. He was educated at Trinity College, Cambridge. In 1879 he was appointed Cavendish professor of experimental physics in succession to Clerk-Maxwell, a post he held until 1884. In 1887 he became professor of natural philosophy at the Royal Institution and was appointed president of the Royal Society in 1905. In 1908 he was appointed Chancellor of Cambridge University, and died June 30th, 1919. Rayleigh was considered by many to have been one of the most brilliant experimental physicists of the nineteenth century. He not only had a remarkable power of mathematical analysis, but was equally skillful in carrying out the experimental proofs of his theoretical researches. Many of his experiments were made with home-built apparatus. He threw a fresh light on nearly every branch of physics, from the theory of gases to wave theories and electric and magnetic problems. Due to the care with which he carried out his experiments, he was the first to detect the presence of neon in the atmosphere, this being the forerunner of the discovery of a number of inert gases in the atmosphere. The Rayleigh balance for absolute current measurement is as well known as it is important. Lord Rayleigh was one of the original members of the Order of Merit. In 1882 he was awarded the Royal Medal, in 1899 the Copley Medal of the Royal Society and in 1904 the Nobel Prize for physics.

REACTANCE—That property of an electric circuit, aside from its ohmic resistance, which tends to oppose the flow of an alternating current. Mathematically, reactance is defined as the square root of the difference between the square of the *impedance* (q.v.) and the square of the effective resistance of a given portion of an electric circuit. Expressed as an equation reactance $x = \sqrt{z^2 - r^2}$ where z is the impedance and r the resistance of the part of the circuit under consideration. Reactance due to capacity in a circuit is capacity or condensive reactance. The reactance of a condenser of capacity C to an alternating current (of sine wave form) of frequency f is expressed by the equation $X_c = \frac{1}{2\pi fC}$. Reactance due to inductance in a circuit is *inductive reactance* (q.v.). The reactance of an inductance coil of inductance L to an alternating current (of sine wave form) of frequency f is expressed by the equation $X_L = 2\pi fL$. Reactance is measured in *ohms.*

REACTANCE COIL—REACTION COIL—Terms used in England to describe the *tickler coil* (q.v.) used in regenerative sets to couple back part of the energy in the plate or output circuit to the input circuit.

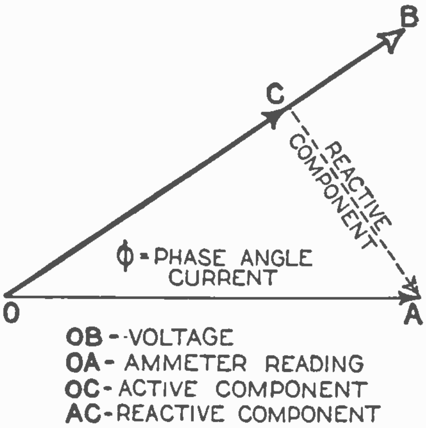

The current represented by vector OA may be considered as divided into two components, OC in phase with the voltage, and CA the reactive component.

REACTIVE COMPONENT—The component of an alternating current used to overcome the reactance of the circuit. This component acts at an angle of 90 degrees from the voltage. It is also known as *wattless current* (q.v.).

REACTIVE CURRENT—See *Reactive Component.*

REACTIVE DROP—The portion of the voltage drop in an alternating current circuit, due to the reactance.

REAL POWER—See *True Power.*

RECEIVER—A term rather loosely applied to cover the entire radio receiving apparatus within a station. Also sometimes used as a synonym for *Detector* (q.v.)

RECEIVING AERIAL—A wire or group of wires suspended at a suitable height and connected to a radio receiving set. The purpose of the receiving aerial is to intercept the electromagnetic waves. It should be placed well above surrounding buildings if possible, and should be well insulated from the ground. (See *Aerial,* also *Directional.*)

RECEIVING CIRCUIT—Any electrical circuit used in connection with the reception of radio. The more important vacuum tube receiving circuits are the *super-heterodyne* (q.v.), the *neutrodyne* (q.v.), the *reflex circuit* (q.v.), the *regenerative circuit* (q.v.) and the *tuned frequency circuit* (q.v.). There are various modifications and combinations of these circuits.

RECEPTACLE—A form of socket into which a plug is inserted for making electrical connections.

RECONSTRUCTED MICA—See *Micanite.*

RECORDER—A telegraphic instrument which automatically records messages sent over a telegraph system.

RECORDING WATTHOUR METER—An instrument for measuring and recording the total amount of electrical energy being consumed in a circuit. (See *Integrating Wattmeter.*)

RECTIFICATION—The changing of alternating current to direct current. In the usual radio receiving circuit rectification is accomplished by means of the vacuum tube, although the ordinary crystal detector rectifies the alternating current also. Before a storage battery can be charged from an alternating current source, the current must first be changed to direct current. Hence the battery charger designed to work on alternating current must be equipped with a rectification device, or rectifier (q.v.).

RECTIFIED CURRENT—An *alternating current* (q.v.) which has been changed to uni-directional or direct current either by preventing alternate waves from flowing or by reversing their direction. (See *Rectified Signals, Rectification,* also *Rectifier.*)

RECTIFIED SIGNALS—Radio signals which have passed through a detector, such as a vacuum tube or a crystal. Rectification of the oscillations results in audible or audio frequency currents which can be heard in a telephone receiver. (See *Rectification, Rectifier, Rectifying Tube,* also *Rectifying Action of Crystal Detector.*)

RECTIFIER—An apparatus for converting alternating current to unidirectional or direct current. Examples of rectifiers are certain crystals used as crystal detectors, the commutators of direct current motors and generators, vacuum tubes, electrolytic devices used for charging batteries, etc. (See *Rectified Signal, Rectification, Rectifying Tube, Rectifying Action of Crystal Detector, Rectified Current, Rectifier, Full Wave, Full Wave Rectification,* also *Half Wave Rectifier, Electrolytic Rectifier, Mercury Arc Rectifier, Charger, Storage Battery; Chemical Rectifier; Tungar; Rectigon.*)

RECTIFIER, FULL WAVE—A device for converting alternating current to direct current, in which both the positive and negative alternations are utilized. This differs from the half-wave rectifier which rectifies by suppressing alternate waves. (See *Full*

Wave Rectification, also *Half Wave Rectifier*.)

RECTIFYING ACTION OF CRYSTAL DETECTOR—The crystal detector has the property of allowing radio frequency alternating currents to flow through it in one direction only. In other words the alternating current is rectified. (See *Crystal Detector-Theory of Operation, Crystal Detector, Crystal Rectifier, Rectified Signals, also Rectified Current.*)

RECTIFYING TUBE—A vacuum tube used to convert alternating current into direct current. Rectification can be accomplished by a two-electrode tube, having filament and plate. In this case the electron flow is from the hot filament to the plate. A two-electrode vacuum tube made especially for rectifying purposes is known as the *kenotron* (q.v.). Another recent type of rectifying tube, known as the *helium tube* (trade name, *Raytheon*) utilizes a different operating principle, known as the "short path" principle, whereby the rarefied gas acts as an insulator between points which are in close proximity. This is in apparent contradiction of the observed phenomenon that the smaller the distance between two points the more readily a spark will jump between them due to the ionization of the gas. However, if the distance is small enough and a suitable gas is used at a low enough pressure, an electron may encounter no gas molecules in its path between the points and there will be no ionization by collision. Consequently, the inert gas helium, may be made to act as a perfect insulator at low pressures. Furthermore, it has been found that when a larger electrode of a gas conduction tube is negative, there is a greater current flow than when the smaller electrode is negative and that the smaller the positive relative to the negative electrode, the smaller is the back current and hence the more complete is the rectification. Figure 1 shows a typical helium tube, while Figure 2 illustrates its construction. The two small positive electrodes *A*, *A*, are carried through two small glass tubes imbedded in a lava insulating block *L* so as to project very slightly into a lava cup *C* whose wells constitute the negative electrode, being connected to the negative terminal through the base. The diameters of the small wires and the diameters and positions of the holes whereby they enter the cup are so proportioned as to give the necessary short path to the negative electrode. The cup *C* contains helium gas at such low pressure as to prevent gaseous conduction. This gives an insulation of great reliability and long life and makes possible extremely small anode surfaces which reduce the back current to a negligible quantity.

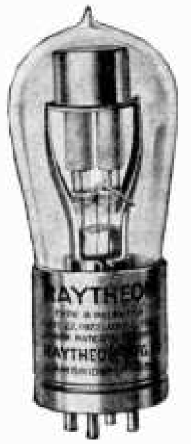

Fig. 1. A Raytheon tube.

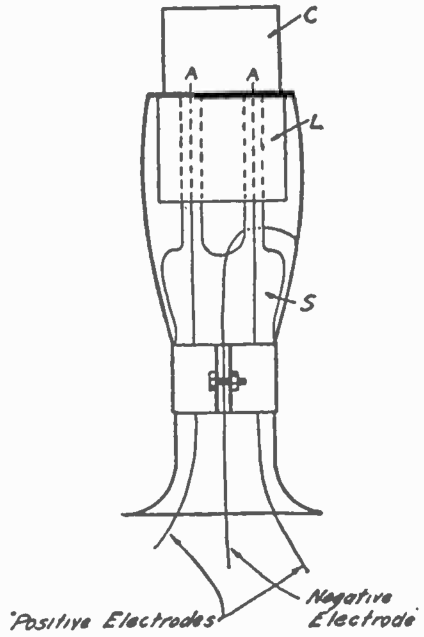

Fig. 2. Construction of the Raytheon tube.

RECTIFYING VALVE—This usually refers to a *rectifying tube* (q.v.) although it may refer to an *electrolytic rectifier* (q.v.)

RECTIGON—The trade name of a storage battery charger which utilizes the vacuum tube rectifier principle. (See *Charger, Storage Battery*, also *Rectifying Tube.*)

RED MAGNETISM—An obsolete term for the lines of force coming from the north pole of a magnet.

REDUCE WAVE-LENGTH OF AERIAL—The wave-length of an antenna may be reduced by the use of series condensers. However, this method should be avoided where possible, since it results in a loss of power. Shortening the aerial is a preferable method of reducing wave-length.

REFLECTING GALVANOMETER—See *Mirror Galvanometer.*

REFLECTION — Radio waves are reflected and also refracted when they pass into a region of different dielectric constant. A perfect conducting sheet would reflect all of the wave energy striking it. A conductor parallel to the direction of motion of a wave serves to guide the direction of the wave. The reflection and other modifications of radio waves serves to explain *fading* (q.v.) and other irregularities of broadcast signals. It has been found that broadcast signals will have most uniform intensity when long waves are used and during the daytime. Moreover, transmission will be better over a uniform conducting surface, such as the ocean. It is assumed that the space through which a radio wave travels is bounded by the surface of the earth and by an upper region which during the day is a fairly good conductor. The transmitted waves are guided by the two conducting mediums but in addition are assisted materially by reflection especially during the time that the air is partially ionized by the radiation of the sun.

REFLECTION COEFFICIENT—(Of a surface of discontinuity between two media.) The ratio of the reflected field intensity near the surface to the incident field intensity near the surface.

REFLEX—See *Reflex Circuit.*

REFLEX CIRCUIT—A radio receiving circuit invented by Marius Latour in which the vacuum tubes are made to perform double duties as both radio and audio frequency amplifiers. The incoming radio frequency current is amplified at radio frequency, rectified by a detector (which may be a crystal detector) and then amplified at audio frequency, using the same tube. A typical four tube reflex set uses the second and third tubes as amplifiers simultaneously for both the radio and the audio frequency amplification. As a result the four tubes are made to produce the effect of six. Reflexing may be accomplished in a number of different ways. In some cases all the tubes are made to work twice. In other cases, only a part of the tubes are used for dual amplification. A variation of the reflex circuit, is the *Inverse Duplex Circuit* (q.v.). In this system the incoming signal reaches the first tube where it is amplified at radio frequencies and passes to the second tube, where it is again amplified at radio frequencies. From the second radio frequency tube, it is passed to the detector tube, where it is converted to audio frequency. The current then flows back to the second radio frequency tube which is now used as an audio frequency amplifier. Thence it flows to the first radio frequency tube which acts as the second

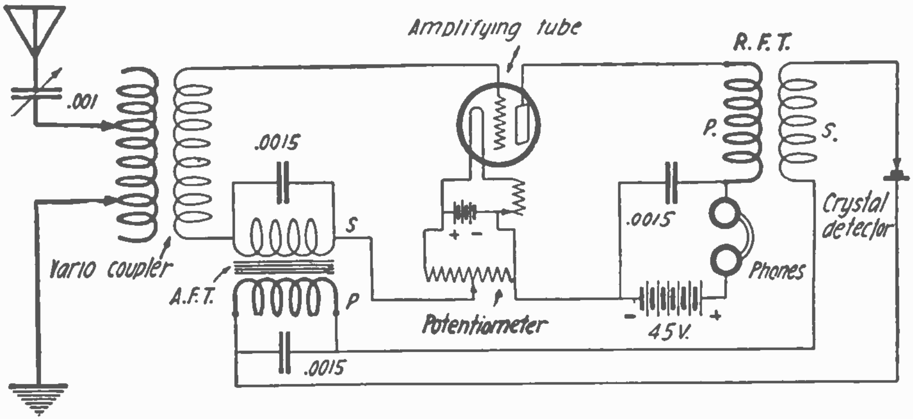

A typical reflex circuit using one tube and a crystal detector.

stage of audio frequency amplification.

REFRACTION—The refraction of electric waves is analogous to the refraction of light. The electric (radio) waves are refracted when they pass into a region of different dielectric constant. That is to say, their direction is slightly altered. (See *Fading, Gliding Theory,* also *Reflection.*)

REFRACTION AND REFLECTION OF ELECTROMAGNETIC WAVES—See *Reflection,* also *Refraction.*

REGENERATION—The amplifying properties of the three-electrode vacuum tube can be used to obtain what is termed *regeneration.* Since it is possible to have greater output energy than input energy, part of the output may be returned to the input side, thus resulting in amplification of energy or in regeneration. By *feeding back* the correct amount of energy and in the right phase relation a constant reamplification may be obtained, and the tube can be made to act as a generator of sustained oscillations. (See *Armstrong Circuits, Feed-Back, Feed-Back Coil, Feed-Back Coupling, Feed-Back Effect,* also *Regenerative Circuit.*)

REGENERATIVE—See *Regenerative Circuit.*

REGENERATIVE CIRCUIT—A method of connecting up vaccum tubes for radio reception in which the amplified variations of the plate circuit are superimposed inductively on the grid or imput circuit thus producing a reinforcing effect which results in increased sensitivity. (See *Armstrong Circuits.*)

REGENERATIVE COUPLING—A radio receiving system designed to increase amplification in a vacuum tube, by coupling the plate or output circuit back to the grid or input circuit thus producing regeneration. (See *Feed-Back, Feed-Back Coil, Feed-Back Coupling, Feed-Back Effect,* also *Regenerative Circuit.*)

REINARTZ CIRCUIT—A selective regenerative circuit devised by John L. Reinartz in which a *spider-web* inductance coil is used, the coil acting as primary, secondary and tickler. Capacity *feed-back* (q.v.) is accomplished by means of a variable condenser. In the improved Reinartz a variocoupler is used, connected as a variometer, in addition to the condenser feed-back.

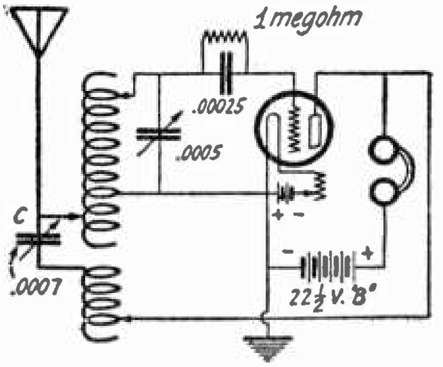

Diagram showing Reinartz circuit

REJECTOR—See *Rejector Circuit.*

REJECTOR CIRCUIT—A circuit consisting of a combination of inductance and capacity for filtering out or preventing the passage of currents of certain frequencies. This circuit is used in connection with a radio receiving set to obtain greater selectivity. The inductance and the capacity are arranged in parallel to form a resonant circuit. The constants of this circuit are chosen so that at the particular frequencies to be filtered, the capacity current will exactly equal the inductive current, thus giving zero resultant current. The primary application of the rejector circuit is as an *interference eliminator* (q.v.) or *wave trap* (q.v.). Figure 1 shows the application of the rejector circuit applied to a radio receiving set. The circuit shown is a combination *acceptor-rejector* circuit. The rejector circuit is shown in heavy lines. The inductance and the capacity are connected in parallel with each other and with the acceptor circuit and this latter is inductively coupled to the receiving set circuit. Both the acceptor and the rejector circuits are tuned to the wave length of the receiving aerial. By using more than one rejector and acceptor circuit increased selectivity can be obtained. A rejector circuit used to filter out undesired spark signals is shown in Figure 2. A small inductance of about ten turns is employed in this circuit and this is inductively coupled to a closed tuned circuit and wired in series with the aerial, and the aerial terminal of the set. If a spider-web coil is used in the closed circuit, a suitable value for broadcast wave lengths is a 50 turn coil, placing the aerial coil in such a position as to get closest coupling. The variable condenser should have a capacity of .0005 mfd. (See *Acceptor.*)

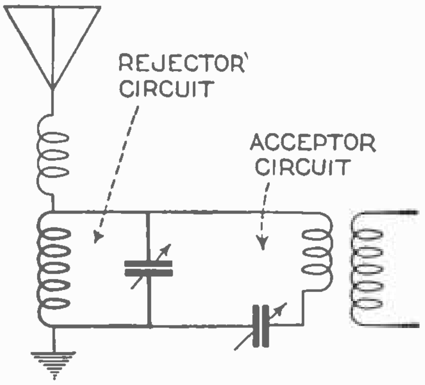

Fig. 1. An acceptor-rejector circuit used as a wave trap.

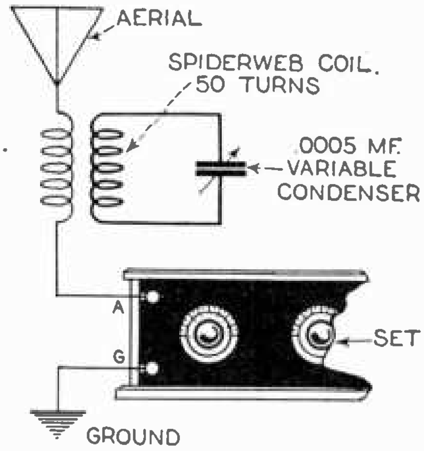

Fig. 2. A rejector circuit designed to filter out undesired spark signals.

RELAY—An electromagnetic device by means of which contacts in one circuit are operated by a change in conditions in the same circuit or in one or more associated circuits. (See *Magnifier.*)

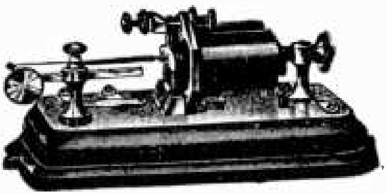

A telegraphic relay.

RELUCTANCE—Magnetic resistance. The reluctance of a magnetic circuit is analogous to electrical resistance. All magnetic circuits offer reluctance to flux. The name *rel* is widely used as the unit of reluctance. The rel is the reluctance of a magnetic circuit which will have one line of force in it, upon application of one *ampere-turn.* The rel can also be specified as the reluctance of a cylinder of air, or of any other non-magnetic material, one square inch in cross-sectional area, and 3.19 inches long. The flux through a magnetic circuit is equal to the *magnetomotive force* (q.v.) divided by the reluctance. (See *Magnetic Properties.*)

RELUCTIVITY—The *reluctance* (q.v.) per unit length and per unit cross-section. The reluctance per inch^3 or per cm^3. Reluctivity may be defined as specific magnetic resistance. It is the reciprocal of *permeability.* (q.v.).

REMANENCE—The term applied to the magnetism remaining in a magnet after the magnetizing force is removed.

REMOTE CONTROL—A control system whereby a radio transmitting or receiving apparatus may be operated at any desired distance from the apparatus. The chief advantage of remote control lies in the fact that the high tension apparatus and the tall aerial masts may be located in outlying districts whereas the studio can be located at any desired, convenient point. In addition it is possible to broadcast concerts, theatrical entertainments, prize fights, etc., direct from the scenes of action by utilizing remote control.

REPEATING COIL—A transformer of unity ratio used in telephone practice.

RE-RADIATION—The relaying of a radio wave through the action of a receiver employing regeneration. This phenomenon is most commonly observed in receivers of the *super-heterodyne* type.

RESIDUAL CHARGE—The residuary electricity retained by a condenser after discharge. This is also known as *electric residue* or *soakage.* After the condenser is discharged in the ordinary way, the residual charge permits a second discharge, smaller than the first. (See *Soaking in.*)

RESIDUAL MAGNETISM—The magnetism retained by a piece of iron, after the magnetizing force is withdrawn. After a mass of iron has become magnetized, some of the magnetism will remain, even though the magnetizing force has been taken away. The amount of residual magnetism depends on the quality of the iron. Pure wrought iron, in general, retains very little residual magnetism, whereas wrought iron which has gone through a hardening process, or which contains a large percentage of impurities and also cast iron possess a much larger amount of residual magnetism.

RESINOUS ELECTRICITY—An obsolete term for negative charges of electricity, adopted because of the fact that resinous bodies become negatively charged by friction. (See *Vitreous.*)

RESISTANCE—Symbol Ω (Omega)—Every conductor of electricity offers opposition to the flow of electricity. Resistance is a property of the conductor itself. It increases with increased length of conductor, decreases with increased cross-sectional area, and is greater or less depending upon the material of the conductor. Resistance is measured in *ohms.* (q.v.). The ohm is equal to 10^9 C.G.S. electromagnetic units.

RESISTANCE BOX—A box containing a number of wire coils of known resistances connected in various combinations, by means of switches or plugs, to produce any desired total resistance.

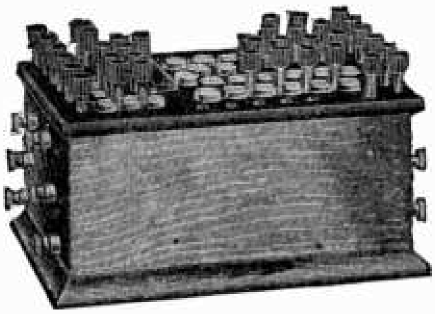

A resistance box.

RESISTANCE COUPLED AMPLIFIER—A type of *amplifier* (q.v.) which utilizes high resistances for interstage coupling. Resistance coupled amplification is generally used only for audio frequency amplification although it may be used for radio frequencies above 1000 meters. The chief advantage of resistance coupling consists in the purity of tone obtainable and the absence of distortion. In addition radio receivers of this type are comparatively inexpensive to build. It should be noted, however, that with resistance coupled amplification, the amplification per tube is much less than with transformer coupled amplification and furthermore the drain on the "B" batteries is also much greater.

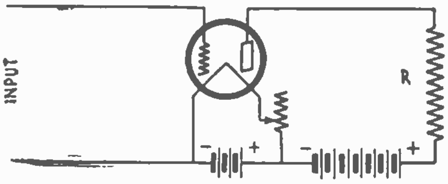

Fig. 1. The introduction of the non-inductive resistance in the plate circuit of the tube is the first step for explaining the principles involved.

Of course, where a "B" eliminator is used "B" battery consumption is not a consideration. The resistances used for resistance coupling are non-inductive and may have resistances of from .1 megohm to 2 megohms.

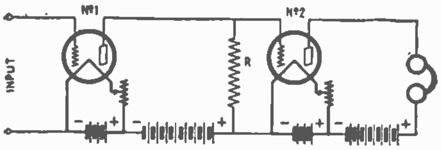

Fig. 2. In this circuit there must be two sets of batteries, but even then the grid of tube No. 2 will have a positive bias, which is undesirable.

Figure 1 shows an elementary circuit utilizing a non-inductive resistance in the plate circuit of the tube. When a signal is impressed upon the grid circuit of this amplifying tube, there is a corresponding voltage set up in the plate circuit. The resistance R in the plate circuit causes a voltage drop to take place in two places in the tube.

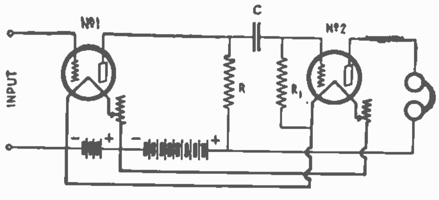

Fig. 3. The resistance coupled amplifier circuit with the introduction of a grid leak and condenser. The reasons for this are explained in the accompanying text.

The total resistance of the plate circuit is equal to the internal resistance of the tube added to the resistance R, and if the "B" battery has 90 volts and if R is equivalent to the tube resistance at that moment the voltage drop across the tube will be 45 volts with the same drop across R. Theoretically it is possible to vary the voltage drop across R from zero to almost 90 degrees. Discounting the small resistances of the "A" and "B" batteries, a changing voltage will occur across the filament and plate when the grid potential varies, and also across the resistance R. Naturally these last variations depend upon the tube resistance and upon the value of the resistance R. These variations will be largest when R has a greater resistance than that of the tube. The resistance or impedance of the tube is constantly in a state of change and can assume very high values. For this reason the value of R should be equal to the higher values of the tube impedance in order to maintain maximum effectiveness. As used with ordinary amplifier tubes, a value of 100,000 ohms has been found to give satisfactory results. The next step is to apply the voltage variations across R to the grid circuit of the next tube. In Figure 2, the resistance R lies in the grid circuit of tube 2 as well as in the plate circuit of tube 1. Hence the magnified changes in tube 1 are transferred from R to the grid and filament of tube 2, thus the latter tube will be affected directly by the output circuit of the first tube and thus there will be amplification in both tubes thereby causing a resultant magnified signal to be heard.

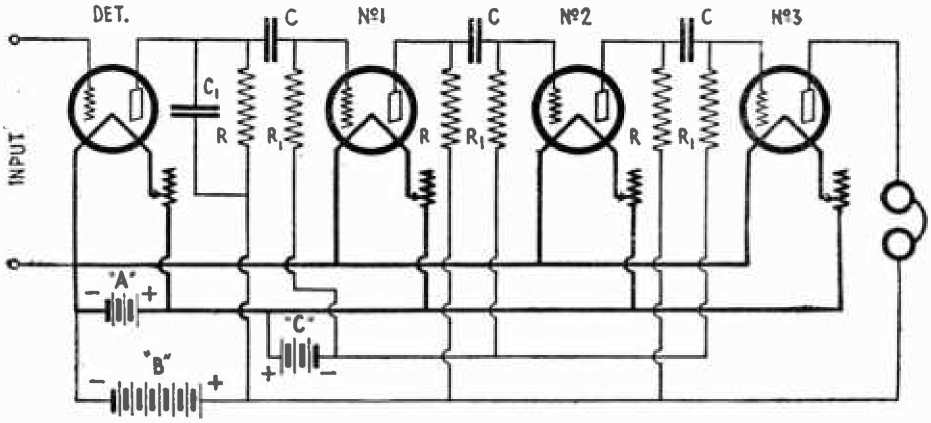

Fig. 4. Diagram showing three stages of resistance coupled amplification.

However, in such a hook-up, the same "A" and "B" battery cannot be used since this would mean that there would be a positive charge on the grid of tube No. 2, which would then be unable to function. Naturally, it would be impracticable to use a circuit having so many batteries, so a means must be found whereby it is possible to have a power supply consisting of but one filament battery and one plate battery for any given number of tubes in an amplifier. In order to do this the grids of the tubes must be separated electrically from the high voltage of the plates of the preceding tubes, but only as far as the steady voltage is concerned, while the fluctuating voltage of these plates must be impressed on the grids of the following tubes. This is done by introducing condensers between the two resistances. These effectively block the direct currents while they afford easy passage to alternating currents. Figure 3 shows the introduction of the condenser between the plate resistance and the grid resistance. Inspection of this diagram shows that although tube No. 1 remains the same, the plate instead of being connected directly to the grid of tube No. 2 is separated from the latter by means of the condenser C. If voltage fluctuations now occur across the resistance R, they will also be effective across the grid and filament of tube No. 2, because R exerts its influence between plate and filament of tube No. 1. As the condenser, C, conducts the variations very easily, the voltage fluctuations against R are then effective across the filament and grid of tube No. 2. It is necessary that tube No. 2 be provided with a grid leak, R_1 for the accumulated negative charges to leak off the grid to the filament. This grid leak at the same time serves the purpose of a by-pass for voltages, occurring due to possible leaks in the condenser C, which would tend to make the grid positive, thus allowing a grid current to flow and stopping the perfect action of the tube as an amplifier. Since the grid leak R_1 is a direct shunt across the grid and filament of tube No. 2, it must have a relatively high value in order to retain the highest possible amplification. R_1 should have a value lying between 500,000 and 2,000,000 ohms. If this resistance is made variable, it will be possible to control the resulting amplification completely. The value of the condenser C, must be chosen so that its resistance to the flow of alternating current will be as low as possible. Figure 4 shows three stages of resistance coupled amplification with a "C" battery used to give proper grid bias. The value of the resistance R, in the plate circuit of the tubes is 100,000 ohms and the value of the resistance in the grid circuits is 1,000,000 ohms (1 megohm). The condenser C has in each case a capacity of 0.05 mfd. The value of the "C" battery is about 4½ volts, but this should be determined by experiment for best results.

RESISTANCE COUPLING—RESISTIVE COUPLING—The transference of electrical energy from one circuit to another, by means of resistance common to both. (See *Resistance Coupled Amplifier.*)

RESISTANCE, HIGH FREQUENCY—The increased resistance of a conductor to high frequency currents, due to concentration of current at the surface of the conductor. This phenomenon is known as *skin effect* (q.v.) Skin effect becomes more pronounced, the greater the frequency of the impressed electromotive force. Skin effect also increases with increased cross-section of the conductor, with increased conductivity of the conductor and with increased magnetic *permeability* (q.v.). As a result of skin effect, the actual resistance of a conductor to alternating currents is greater than to direct currents.

RESISTANCE OF GROUND CONNECTION—The connecting wire leading from the radio apparatus to the ground should be of low resistance. If possible, the ground should be made by burying a copper plate in moist earth. The usual way to obtain a ground is to fasten a copper strap or *ground clamp* (q.v.) around a water or a steam pipe.

RESISTANCE (RADIATION)—This is the ratio of the total energy radiated (per second) by the antenna to the square of the root-mean-square current at a potential node (generally the ground connection).

RESISTANCE, RADIO FREQUENCY—The ratio of the heat produced per second, in watts, to the square of the root-mean-square current (radio frequency), in amperes, in a conductor.

RESISTIVITY—See *Resistivity of a Material*, also *Resistivity, Surface.*

RESISTIVITY OF A MATERIAL—SPECIFIC RESISTANCE—Symbol ρ (rho)—The resistance in ohms of unit length and unit cross-section of the material under consideration. Resistivity may be measured in ohms per inch cube or per centimeter cube, but more commonly it is measured in ohms per *mil-foot* (q.v.). The material of a conductor has an important bearing upon its resistance. Thus a unit length and unit cross-section of aluminum has about one and one-half times the resistance of copper having the same dimensions. Platinum has about six times the resistance.

RESISTIVITY, SURFACE—Referring to an insulating substance, surface resistivity is the resistance between two opposite edges of a surface film, 1 centimeter square. Surface resistivity depends to a considerable extent on humidity.

RESONANCE—A condition which exists in an electrical circuit where the frequency of the applied electromotive force is the same as the natural frequency of the circuit. Resonance is the cumulative effect produced by a periodic force in a circuit, so adjusted in frequency that the effect can attain the highest value. The effect measured is usually the current and the circuit when adjusted for maximum current, is said to be in resonance with the force, or with the circuit from which the force comes. The effect of resonance is to bring the current into phase with the electromotive force. Resonance in a circuit has the effect of bringing the reactance due to the inductance (inductive reactance) in exact opposition to the reactance due to capacity (capacity reactance). The total reactance is therefore zero and therefore the impedance is equal to the resistance. When two or more circuits have the same natural frequency they are said to be in resonance. Resonance of a circuit to a given exciting alternating electromotive force is that condition due to variation of the inductance or capacity in which the resultant effective current (or voltage) in that circuit is maximum. Instead of varying the inductance and the capacity, the frequency of the exciting field may be varied. The condition of resonance is determined by the frequency at which the current or voltage is maximum.

RESONANCE CURVE—A curve showing the relationship between the current induced in an oscillatory circuit and the frequency of the inducing current. In the case of two oscillating circuits placed near each other so that oscillations in one circuit will induce oscillations in the other, the induced oscillations will be weak or strong depending whether or not the second circuit is in tune with the first one. The capacity and the inductance of the second circuit can be varied until it is exactly in tune with the first at which time maximum current will flow in it. The resonance curve is obtained by measuring the current in the second circuit for various noted oscillation constants, and plotting the curve with the ordinates representing the current and the abscissae the corresponding natural frequency of the second circuit. The diagram shows a typical resonance curve. The peak of the curve is the point where the two circuits are in tune, or in other words, the point of resonance.

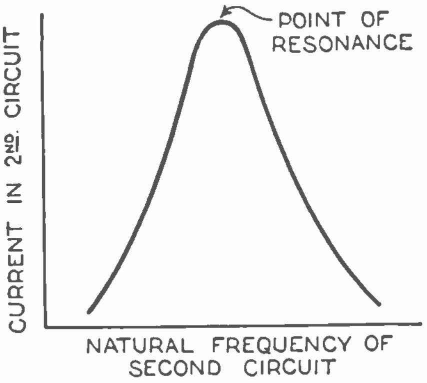

A typical resonance curve.

RESONANCE INTENSIFIER—An automatic receiver used for the reception of high-speed radio telegraph signals.

RESONANCE TRANSFORMER—Any loose-coupled tuning inductance having a primary and secondary each with a variable condenser in the circuit. Tuning the secondary circuit brings it in resonance with the primary, thus enabling signals to be heard with greatest volume.

RESONANT—See *Resonance.*

RESONATOR—A device for detecting by resonance, oscillations produced by an oscillator. This term is also used to refer to the sound box employed with the telegraph sounder.

RESONATOR, ACOUSTIC—A vessel containing a volume of air which can easily be set into vibration by sound waves or by the oscillation of a solid object. Typical forms are cylindrical or spherical and the ear is applied to a small tube opening into the air chamber.

RETARDATION COIL—A reactance coil used in a circuit for the purpose of selectively reacting on currents which vary at different rates.

RHEOSTAT—A variable resistance used to control current flow. Rheostats may be made of resistance wire or of carbon or for special purposes liquids may be used. In radio work, rheostats are connected in series with the filaments of vacuum tubes and the "A" battery in order to control the amount of current flowing in the filaments.

A rheostat constructed of resistance wire.

R. M. S.—Abbreviation for *Root Mean Square* (q.v.)

ROBERTS, Joseph Harrison Thomson—Born in 1890, he was educated at Liverpool and Cambridge Universities. He carried out many physical researches at the Cavendish Laboratory, Cambridge, and did a great deal of work on thermionics and other radio problems. He is a doctor of science and a Fellow of the Institute of Physics. He has several radio patents to his credit and is an expert in electrical, optical and acoustical sciences.

ROBINSON, James—British radio inventor. Born in 1884, he was educated at Durham and Gottingen Universities. He was appointed lecturer in physics at Durham University, 1906-07, lecturer in physics Sheffield University, 1910-12, and lecturer in physics at the East London College, London University, 1912-15. Robinson is known chiefly for his direction finding apparatus. He is a member of the Radio Society of Great Britain.

ROBINSON, Samuel S.—Rear-Admiral, U.S.N. He was born May 10th, 1867, and was educated at the United States Naval Academy, from which he graduated. He wrote the Manual of Wireless Telegraphy for Naval Electricians, and was in charge of the Bureau of Equipment, Navy Department.

ROOT MEAN SQUARE—The square root of the mean value of the square of an alternating current or voltage. An alternating current whose root-mean-square value is one ampere has the same heating effect as a direct current of one ampere. Root mean square is also referred to as *effective* or *virtual value.* The root mean square is derived as follows. In electrical measurements, it is necessary to compare direct and alternating currents. Moreover, the same units of measurement, the *ampere* (q.v.) and the *volt* (q.v.) are used in referring to both direct and alternating currents. However, in a direct current, a certain voltage will produce a steady current, whereas in an alternating current, both the voltage and the current are constantly varying both in magnitude and direction. The common basis of measurement is found in the value of the alternating current which will produce the same amount of heat as a certain direct current. Heating effect, according to *Joules Law* (q.v.) is directly proportional to the square of the current. Hence, in an alternating current varying according to a sine law, it is necessary to square each ordinate, find the mean or average value of these squares, and then extract the square root of this mean. Stated in concise terms, it is necessary to find the root mean square. An alternating current having a *maximum* value of one ampere would have a root mean square value of .707 amp. (See *Form Factor.*)

ROOT-MEAN-SQUARE VALUE—See *Root-Mean-Square.*

ROTARY CONDENSER—A synchronous motor used on a power line to advance the *phase* (q.v.) of the current. As a result there is an improvement in the *power factor* (q.v.). The action of the synchronous motor is similar to that of a condenser, whence the name rotary condenser.

ROTARY CONVERTER—A device for converting alternating current into direct current or vice versa. The rotary converter is also known as a *synchronous converter.* It is built like a direct current generator, except that it has certain commutator segments or conductors in the armature, connected to slip rings. Two, three, four, or six rings may be used. When the armature is rotated there is an alternating voltage between the rings. If the converter is driven by a motor or by an engine, it may be used as a *double current generator,* direct current being obtained from the brushes and alternating current from the slip rings (collector rings). When the collector rings are supplied with alternating current, the converter runs as a synchronous motor and direct current can be obtained from the brushes on the commutator. It should be understood that there is only one set of windings on the rotary converter which act simultaneously as alternating current motor and direct current generator. A converter which is used to convert direct current to alternating current is called an *inverted converter.* In the *single-phase* converter, there are two slip rings connected to the windings by as many equally spaced taps as there are poles. In the three-phase converter, three rings are used, with three equally spaced taps for each pair of poles. The single phase rotary converter must be brought up to synchronous speed by external power as it will not start itself. The polyphase converter will start itself from the alternating current side, but since it draws a very great lagging current, it should be started whenever possible from the direct current side. (See *Converter.*)

ROTARY CURRENT—A term infrequently used to denote *polyphase* (q.v.) current.

ROTARY DISCHARGER—A rotating disc discharger used in wireless transmitters in which the sparks occur between metal studs mounted on a motor driven disc. (See *Disc Discharger.*)

ROTARY FIELD—A magnetic field rotating in space. It may be due to rotating windings or stationary windings properly spaced through which polyphase currents are flowing.

ROTARY GAP—A form of rotating spark gap utilizing two or more rotating discs carrying studs between which the sparks pass. Spark gaps of this type were formerly used in wireless telegraph transmitters. (See *Discharger,* also *Disc Discharger.*)

ROTARY HYSTERESIS—Hysteresis (q.v.) produced by a rotating field as distinguished from that produced by a changing field.

ROTARY SPARK GAP—See *Rotary Discharger.*

ROTARY TICKER—A *ticker* (q.v.) in which the interruptions are produced by the chattering of a springy brush on a rapidly rotating metal wheel.

ROTOR—The rotating part of any electrical apparatus. It may refer to the revolving portion of a *motor* (q.v.) or *generator* (q.v.) or to the rotable part of a variable *condenser* (q.v.). The rotor in a direct current motor or generator is the armature. In an alternator or an alternating current motor, the rotor may be either the armature or the field.

ROTOR, GROUNDED CONNECTION—This refers to the grounding of the rotors of variable condensers in a radio receiving set. This is often done to eliminate hand capacity. Sometimes the rotors are connected to the shielding, which is grounded.

ROUGH CALIBRATION—Meters such as voltmeters, ammeters, etc., are often calibrated by comparison with standard instruments. In case such instruments are not available, it may be necessary to use less accurate methods. As an example of rough calibration, an experimental wave meter could be calibrated roughly by tuning in on various broadcasting stations whose wave lengths are known. (See *Calibration.*)

RUHMKORFF COIL—See *Induction Coil.*

RUTHERFORD, Sir Ernest—British physicist. Born at Nelson, New Zealand, he was educated at Canterbury College, Christchurch, New Zealand University, and Trinity College, Cambridge. From 1898 to 1907 he held the position of MacDonald professor of physics at Manchester University. In the latter year he was appointed Cavendish professor of physics and director of the Cavendish laboratory at Cambridge. Rutherford is an extremely brilliant physicist and has carried out a series of researches on the ultimate composition of matter. Many of these experiments have confirmed the researches of J. J. Thomson and others. Some of Sir Rutherford's experiments have revolutionized the theories of matter. He has received many honors for his brilliant researches including the Nobel prize for chemistry awarded to him in 1908. He has written extensively on radio activity, his books including "Radioactivity," 1904; "Radioactive Transformations," 1906; "Radioactive Substances and their Radiations," 1913; as well as many papers for technical journals. Sir Ernest was knighted in 1914.

S

SAFETY GAP—A spark gap which is connected in parallel with a condenser or other apparatus, and which protects the apparatus from excessive voltage. When the voltage exceeds a predetermined limit, the gap breaks down, thus passing the current around the apparatus. (See *Gap, Micrometer.*)

SAL AMMONIAC—A term commonly used for ammonium chloride. This substance is a white crystal, very soluble in water, and is used as an electrolyte in primary cells and also in many electroplating processes. (See *Cell.*)

SATURATED—The magnetic condition of a substance when increase of magnetizing force produces no further increase in flux density. A point is reached in every magnetic material, where further increase of magnetic intensity will not produce any increase of flux density. At this point, the material is said to be saturated. Cast iron becomes saturated at a lower flux density than steel or wrought iron.

SATURATION—See *Saturated.*

SATURATION, MAGNETIC—See *Saturated.*

SECONDARY CELL—A unit consisting of a positive and a negative electrode immersed in an *electrolyte* (q.v.) which becomes active only after the passage of an electric current. The electrodes are of metal or of a metallic compound. An electric current, flowing into a secondary cell causes certain chemical actions to take place whereby electrical energy is transformed into chemical energy. This process is termed "charging." After the cell is charged, the reverse action takes place and chemical energy is transformed into electrical energy. In this case, the cell is able to deliver an electric current to an outside circuit. This process is termed "discharging." The secondary cell is also known as an *accumulator* or *storage cell.* However, the cell does not store electricity. It merely holds or stores the chemical energy for immediate transformation into electrical energy. It is usual practice for several positive plates to be connected in multiple to form one positive electrode and for several negative plates to be connected in parallel to form a single negative electrode. The *storage battery* consists of two or more storage cells. The operation of the storage cell depends upon the fact that certain metals immersed in an electrolyte may differ in potential. In the secondary cell, no electrical action takes place until after charging. Thereafter, chemical action maintains a difference of potential until the stored energy has been retransformed. While the negative electrode is used up to a certain extent during the process of discharging, it is replaced electrochemically during the charging process, due to the flow of current through the electrolyte in the opposite direction to that of discharge. The two principal types of storage cells are the lead cell and the nickle-iron-alkaline (or Edison Cell). In lead cells, the negative electrode is made of spongy lead and the positive electrode is of lead peroxide. The electrolyte is dilute sulphuric acid. The following chemical action takes place within the lead cell. When the cell is charged, hydrogen is released at the negative plate, and oxygen at the positive plate. The oxygen combines with the lead of the positive plate to form lead peroxide while the hydrogen reduces the oxide in the negative plate. When the cell is discharged, both electrodes are changed to lead sulphate and the liberated hydrogen combines with the oxygen to form water. Care must be taken not to allow the cell to discharge too far as in this case, too great an amount of lead sulphate may form and this may cause buckling of the plates or may make it very difficult to recharge the cell. The indications of a completely charged cell are increased gassing of electrolyte together with a stop in the rising of the voltage and the specific gravity. In the Edison-type secondary cell, the active materials are of nickel and iron. Nickel hydrate forms the active material of the positive plate, while iron oxide forms the active material of the negative plate. The electrolyte is a 21 per cent. solution of potassium hydrate combined with a small amount of lithium hydrate. The chief advantage of the *Edison cell* (q.v.) is that it can be left idle indefinitely in a fully charged, semi-charged, or fully discharged state without any harm to the cell. In addition it is lighter and smaller than the same capacity lead cell. The unit used for measuring the capacity of storage cells is the *ampere-hour* (q.v.). This is based on the rate of discharge normally used in measurement, the eight-hour rate. As an example, an 80 ampere-hour storage battery will furnish a continuous current of 10 amperes for 8 hours. If the battery is discharged at a more rapid rate, it will not have the rated capacity.

SECONDARY CURRENT—The current flowing in the secondary windings of inductive apparatus. Current induced in a secondary winding due to the action of *primary current.* (See *Current, Primary.*)

SECONDARY ELECTRON EMISSION—See *Secondary Electrons,* also *Secondary Emission.*

SECONDARY ELECTRONS—Electrons emitted by the plate of a vacuum tube when struck by the electron stream from the filament. (See *Negative Resistance,* also *Secondary Emission.*)

SECONDARY EMISSION—Electron emission in which the exciting agency is bombardment of the emitting material by electrons. The emission of electrons from cold electrodes under the impact of other electrons. (See *Negative Resistance.*)

SELECTIVE—See *Selectivity.*

SELECTIVITY—The sharpness with which a radio receiving set can be tuned to a particular wavelength. A selective set is one which tunes sharply, whereas a non-selective set tunes broadly. Too long an aerial will often result in making a set tune broadly. A short aerial makes the receiving set more selective but reduces the volume.

SELECTOR—A tuning device connected in a radio receiving circuit for the purpose of eliminating undesired stations. (See *Wave Trap.*)

SELENIUM CELL—A cell constructed of a thin layer of the metal selenium. This metal is extremely sensitive to light. In its crystalline form, it is a conductor of electricity, its conductivity varying with the amount of light allowed to fall on it. As a result it is possible to regulate the flow of current in a circuit by varying the light. The selenium cell thus offers a means of transmitting pictures by wire or radio. Selenium cells are also being used in experimental devices for the transmission of motion pictures and for obtaining *television* (q.v.).

SELF-CAPACITY—The capacity of a circuit or a portion of a circuit. Every electrical circuit possesses inherent capacity, which varies in degree, depending upon the shape, size, etc., of the circuit. Self-capacity is also referred to as *distributed capacity* (q.v.). In general self-capacity is detrimental, especially in radio circuits, since it introduces losses. By using special forms of windings, the self-capacity of coils can be greatly reduced. (See *Low Loss Coils.*)

SELF-EXCITED GENERATOR—A generator which depends upon its own current for its field excitation. Such a generator must build up its magnetic field and hence its voltage, gradually. In order to get a start, it depends upon the *residual magnetism* (q.v.) present in the iron of the poles. There are three types of self-excited generators, the *shunt generator* (q.v.), the *series generator* (q.v.) and the *compound generator.* (See *Compound Wound.*)

SELF-HETERODYNE—A system of reception of continuous wave signals by the production of audio frequency beats, through the use of a device which is both a radio frequency generator and a detector of the audio frequency beat currents produced.

SELF-INDUCTANCES—See *Inductance, Self.*

SELF-INDUCTION—Inductance due to the field produced by the circuit itself. Changes in current are always accompanied by self-induction. When the current is decreasing, the electromotive force of self-induction is in the same direction, thus tending to prevent the decrease. When the current is increasing, the electromotive force of self-induction tends to prevent the increase. (*See Inductance, Self; Mutual Inductance,* also *Induction.*)

SELF-INDUCTION COEFFICIENT—The ratio of the number of lines of induction produced by a current flowing in the circuit divided by the current in the circuit. The coefficient of self-induction of a conductor of finite cross-section may be defined as the ratio of twice the energy of the magnetic field produced by the current flowing to the square of the current. The following factors affect the coefficient of self-induction of a circuit: the cross-section and shape of the conductor, the size and shape of the circuit, the permeabilities of the surrounding medium and of the conductor itself. Where the skin effect is large, the frequency of the current and the specific resistance of the conductor will also affect the coefficient of self-induction.

SELF-OSCILLATION—The tendency for the vacuum tubes in a radio receiving set to act as generators of radio frequency oscillations. (See *Body Capacity,* also *Bolitho Circuit.*)

SENSITIVE SPOTS—Points on the mineral of a crystal detector which show a greater response to incoming oscillations than others. Dust, grease, or dampness reduce the sensitivity of these spots and therefore the crystal should be properly protected.

SEPARATELY EXCITED GENERATOR—A generator in which current for producing the magnetic field is obtained from some source external to the generator itself. This source is either another generator or else a storage battery. Separately excited generators are generally used only for special work such as battery charging, electroplating and testing.

SERIES—See Series Connection.

SERIES CONNECTION—When two or more parts of an electrical circuit are so connected that the same current flows through them they are said to be in *series* and the circuit is called a *series circuit,* the connection being a *series connection.* In order to connect primary or other cells in series, it is necessary to connect the positive terminal of one to the negative terminal of another, etc. In this case, the total voltage of the battery equals the sum of the separate voltages.

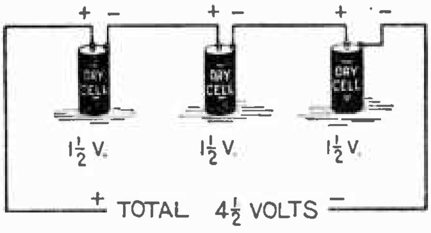

The three dry cells are here shown connected in series.

SERIES GENERATOR—A generator in which the exciting winding is connected in series with the armature and the load. This type of generator is of small commercial importance being used mainly for certain constant current systems.

SERIES MOTOR—A commutator-type motor having the field and armature windings connected in series. Ordinarily, the series motor is a direct current motor, but with minor modifications, such as lamination of the field as well as armature, this motor can be used for alternating current also. With the applied voltage constant, the greater the load, the less will be the speed of a series motor. As the load is lessened, the motor will speed up, and if the load is entirely removed, it will "run away." The *torque* (q.v.) characteristic is the most important feature of a series motor. Instead of varying directly with the current as in the shunt motor, the torque increases with the square of the current.

Hence, if the current is doubled, the torque will be increased four times, while if the current is quadrupled, the torque will be increased sixteen times. For the same current, the torque of the series motor is higher than that of the shunt and in addition, the series motor has a much better starting torque.

SERIES-PARALLEL—This refers to a circuit which is a combination of a *series circuit* and a *parallel circuit.* The term *series-parallel* is generally used to indicate a series circuit having certain minor branches in parallel; the term *parallel-series* denotes a parallel circuit having certain minor branches in series. The illustration shows a series parallel circuit having two dry cells in parallel and three dry cells in series. The push button and the bell are also in series. (See *Parallel Connection, Series Connection,* also *"A" Battery.*)

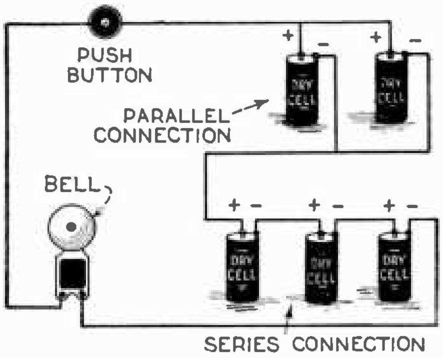

A series-parallel circuit. The three dry cells are connected in series with each other and with the bell and the push button. They are also in series with the other two dry cells, these latter being in parallel with each other.

SERIES RESONANCE—When a single lumped capacity and a single lumped inductance are connected in series between terminals to which an alternating electromotive force is applied, and the inductance or capacity or frequency is varied, the condition of series resonance exists, when the current is a maximum. (See *Resonance,* also *Resonance Curve.*)

SHARPNESS OF RESONANCE—A quantity expressing the fractional change of current in a simple series circuit for a given fractional change in either capacitative or inductive reactance.

SHARPNESS OF TUNING—See *Selectivity.*

SHARP WAVE—A wave in which the energy radiated is confined to a single frequency of oscillation. Tight coupling of the oscillation transformer tends to give a broad wave, whereas loose coupling gives a sharp wave. (See *Decrément.*)

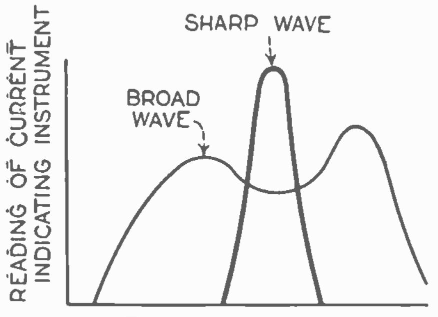

Graph showing difference between sharp and broad waves.

SHIELDED TRANSFORMER—An audio or a radio frequency transformer surrounded by a metal casing. Shielding of radio frequency transformers prevents interference from external fields and also keeps body capacity from affecting the tuning. Audio frequency transformers are shielded mainly to prevent interaction between the flux from the transformer and other parts of the radio set.

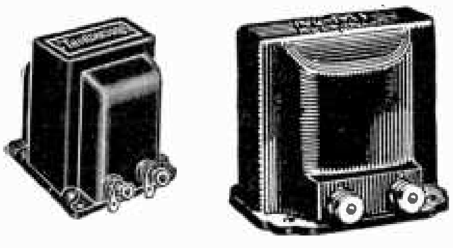

Typical shielded transformers.

SHIELDING—Metallic screening placed around radio apparatus to prevent interference from stray electromagnetic or electrostatic fields. In radio sets, the inside surface of the panel is often shielded by means of a metallic lining, thus eliminating body capacity. It is usual to ground the shielding. In sets having more than two stages of radio frequency amplification complete shielding is essential. (*See Induction Screen.*)

S.H.M.—Abbreviation for *Simple Harmonic Motion* (q.v.).

SHOCK EXCITATION—See *Impact Excitation.*

SHORT—See *Short Circuit.*

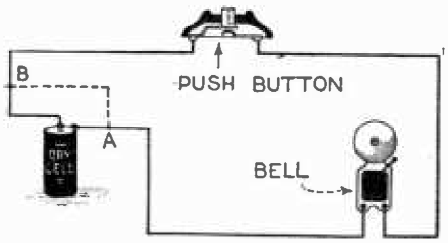

The dotted line between A and B represents a conductor short-circuiting the dry cell.

SHORT CIRCUIT—A short circuit refers to a condition of an electrical circuit whereby an external path of low resistance is connected across the source of electrical energy. A short circuit is accompanied by an increase in current flow, although this current is no longer usefully employed. Apparatus may be protected from short circuits by means of *circuit breakers* (q.v.) or *fuses* (q.v.).

SHORTENING CONDENSER, or ANTENNA TUNING CONDENSER—A condenser connected in series with the antenna and the ground, for the purpose of diminishing the natural wave length.

SHUNT WOUND GENERATOR—A generator in which the field winding is shunted across the armature terminals. The resistance of the shunt winding is made high enough so that only a small percentage of the load current flows through it.

SHUNT WOUND MOTOR—A motor in which the field winding is connected in parallel with the brushes (and the armature winding). The speed of a shunt motor is practically constant, but falls to a slight extent as the load is increased. The drop in speed may be only a few per cent between no load and full load, for large motors, while for small ones it may be as high as 15 per cent. As the load on a shunt motor is increased, the torque which the motor develops will increase and the armature current will also increase in almost direct proportion.

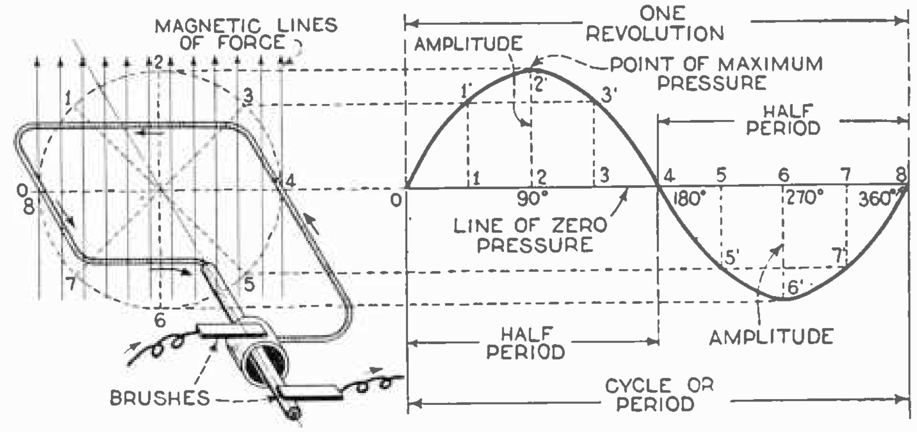

A sine curve is shown at the right. At the left a coil is shown rotating in a magnetic field. As the coil cuts the magnetic lines it generates the varying electromotive force represented by the sine curve.

S.I.C.—Abbreviation for *Specific Inductive Capacity* (q.v.).

SIDE BANDS—A group or band of frequencies formed by the interaction of the *Carrier Wave* (q.v.) and the modulations above and below the main carrier wave frequency. (See *Harmonics.*)

SIGNAL, STRAY RADIO—See *Strays.*

SILICON—A metalloid occurring widely in nature. It is a component of most of the rocks forming the earth's crust. Silicon crystals are used as detectors in radio work. (See *Crystal*, also *Silicon Detector.*)

SILICON DETECTOR—A *crystal detector* (q.v.) using silicon as its crystal. The silicon is used in a crystalline state and a *cat-whisker* (q.v.) is generally used to locate the ***sensitive spots*** (q.v.) and make contact.

SIMPLE HARMONIC—Referring to a *simple harmonic motion* (q.v.). If a point moves at a uniform rate around a circle and the point be projected in a straight line in the plane of the circle, the back and forth motion of the projected point is referred to as ***simple harmonic motion.*** True simple harmonic motions are the motions of a pendulum, a vibrating tuning fork, particles of water in a wave, etc. Other motions which are not precisely simple harmonic motions may be treated as resulting from several such motions. Examples of these are motions occurring in alternating currents in light waves, in electric waves and in sound waves.

SIMPLE HARMONIC MOTION—Abbreviation S.H.M.—The motion executed by the foot of a perpendicular let fall on the diameter of a circle from a particle moving with uniform velocity in that circle. The piston rod of a steam engine, turning a crank uniformly approximates a simple harmonic motion.

SIMPLE HARMONIC VIBRATIONS—The vibrations of a body such as a tuning fork which follow the law of ***simple harmonic motion*** (q.v.). (See *Angular Velocity*, also *Simple Harmonic.*)

SINE CHARACTERISTIC OF ALTERNATING CURRENT—See *Sine Curve*, also *Sine Wave.*

SINE CURVE—A curve whose perpendicular at any point is proportional to the sine of the angle corresponding to that point. The sine curve is used to represent the changes in direction and strength of an alternating current or voltage. The coil illustrated at the left in the accompanying diagram generates a sine wave current as it revolves through the magnetic field shown by the arrows. When the coil is moving parallel to the magnetic flux, no electromotive force is induced and hence current flow is zero. As the coil is rotated in a clockwise direction, it cuts the flux at a greater and greater angle, until the maximum is reached when the flux is cut at a right angle. This gives a gradually increasing electromotive force (and hence current) until maximum is reached, then when the angle at which flux is cut decreases, the electromotive force decreases until the coil again moves parallel to the flux, at which point electromotive force is again zero. Then when the coil starts to cut the flux in the opposite direction, electromotive force is reversed, increasing to a maximum and then decreasing to zero thus completing the ***cycle*** (q.v.). Current or electromotive force in one direction (positive) is represented by the height of the ordinates of the sine curve above the line, while current in the reverse direction is represented by the height of the ordinates below the line. The horizontal distances (abscissae) represent time generally measured in phase angles (position of coil at any particular instant). (See *Alternating Current,* also *Curve of Sines.*)

SINE WAVE—As referring to an alternating current, a sine wave or a simple alternating current, is a current whose wave shape is sinusoidal. The wave shape is the shape of the curve obtained when the instantaneous values of the current are plotted against time in rectangular co-ordinates. The wave shape is independent of the frequency and of the scale to which the curve is plotted.

SINE WAVE OF ALTERNATING CURRENT OR E.M.F.—See *Sine Wave.*

SINGING ARC—See *Duddell Singing Arc.*

SINGING SPARK—A spark used in a wireless telegraph transmitter which produces a singing note in the receiving head set. A spark which occurs at regular intervals and at an audible frequency such as that obtained in a quenched spark system of wireless telegraphy.

SINGLE CIRCUIT TUNER—A regenerative receiving circuit in which the antenna and grid circuits are conductively coupled. *Feedback* (q.v.) is obtained by placing the plate circuit in inductive relation with the primary.

SINGLE-PHASE—This term is used to refer to an alternating current, an alternating current system of distribution, or to alternating current apparatus. A single phase system or circuit is one energized by a single alternating electromotive force. A single phase circuit is usually supplied through two wires. The currents in these two wires counted outwardly from the source, differ in phase by one-half a cycle (180 degrees). In the single phase generator the voltage per phase is the same as the voltage between the lines and the current per phase is the same as the current per line. Since there is a great saving of copper in the use of three phase transmission of power, three phase generators are more frequently used than single phase machines. (See *Alternator.*)

SINOIDAL VIBRATION—See *Simple Harmonic Motion.*

SIPHON RECORDER—A moving coil galvanometer originally used for recording messages sent over long submarine cables. The received currents are passed to a coil suspended in a strong magnetic field. This coil carries a fine glass syphon which discharges ink onto a moving paper tape. In cable signaling, dots and dashes are usually indicated by opposite deflections which result in a wavy line being produced on the tape.

SKINDERVIKEN TRANSMITTER BUTTON—A small carbon grain microphone (q.v.) button less than an inch in diameter and about half an inch high containing a polished metallic button affixed to a mica diaphragm and a surrounding case of brass, the space between the case and the polished button being partially filled with a good grade of carbon granules. The button is very sensitive to sound waves when it is attached to any form of a diaphragm, thus making it applicable wherever a sensitive telephone transmitter should be used. Fig. 1 shows the external appearance, full size, of the Skinderviken button. This transmitter button has a number of applications in radio work and can be used in connection with speech amplifiers, telephone transmitters, detectaphones, etc. It can be made to transmit in any position, either horizontal or vertical. It is small enough to be readily concealed. Fig. 2 shows how radio music can be re-transmitted to distant points and incidentally amplified at the same time. The microphone button is shown fastened to a cone and mounted in front of the loud speaker. The leads from the button are connected in series with several dry cells and another loud speaker at a distant point. Fig. 3 shows how the transmitter button can be attached to an ordinary receiver to produce the same effect. In this case also, the leads are connected in series with several dry cells and with a low resistance phone at the distant end. In Fig. 4 the transmitter button is shown in use as a grid leak in a regenerative circuit. Where it is desired to get greater sensitivity and volume for either radio telephone transmitting or for ordinary phone transmission the system illustrated in Fig. 5 will be found highly suitable.

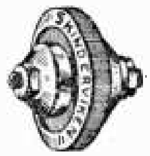

Fig. 1. Skinderviken Transmitter Button.

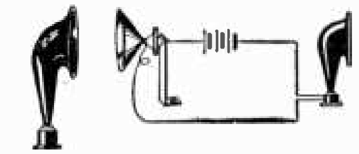

Fig. 2. Retransmitting radio programs.

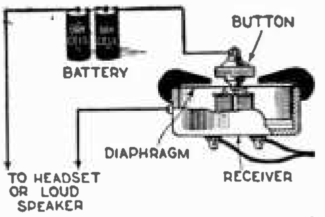

Fig. 3. Retransmission from ear phone.

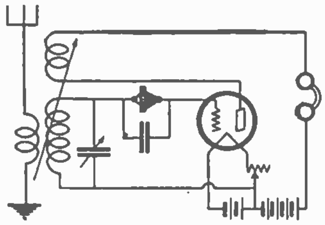

Fig. 4. Button used as grid leak.

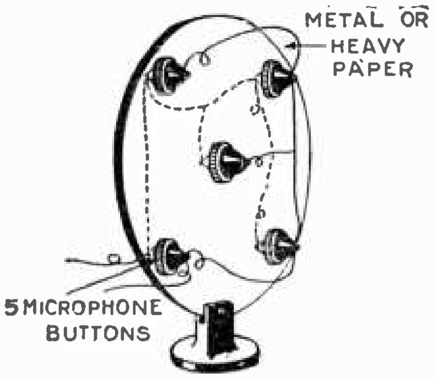

Fig. 5. Combining buttons to form sensitive microphone.

Fig. 6 shows the construction of a hand microphone for transmitting sets or for inter-departmental telephone circuits. A very simple circuit closer is mounted in the handle of the microphone which in this particular case is simply an ordinary piece of wood hollowed out to receive the switch and the wires. This wooden handle is then fastened to the rubber case of a telephone receiver. The coils and magnets in the telephone receiver are removed, a small hole is drilled or punched in the diaphragm, and the transmitter button is fastened as indicated in the diagram.

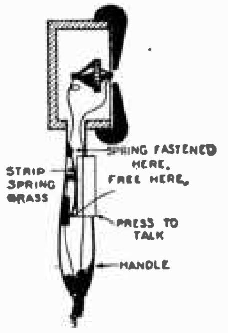

Fig. 6. Construction of hand microphone.

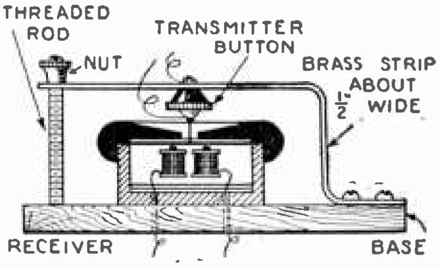

Fig. 7. Method of fastening a button to ear 'phone.

Another method of fastening a transmitter button to a telephone receiver is illustrated in the diagram in Fig. 7. Here the transmitter button is mounted on a brass strip which is bent so that the center of the button will rest upon the diaphragm of the telephone receiver fixed to the base. A threaded rod and nut regulate the pressure of the button against the diaphragm. In the event that it is not desired to make any changes to a pair of telephone receivers, the transmitter button can be fastened in the shell of an old receiver, B in Fig. 8, and the head phone may be rested on the shell. For

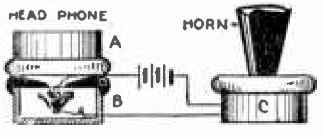

Fig. 8. A simple form of amplifier.

use with a Baldwin receiver, a microphone transmitter button is fastened to it in the following manner: The diaphragm of the Baldwin receiver is first removed, and a thin rod is soldered to the lever, as indicated in Fig. 9. The transmitter button is then soldered to the rod as shown. Another way of making a loud speaker set from an ordinary one-tube receiving set is illustrated in Fig. 10. Where a phonograph is available, the hook-up illustrated may be employed, placing the reproducer of the phonograph on the receiver, which should be placed in a suitable position on the turntable. Inasmuch as the movement of the diaphragm of the seventy-five ohm receiver is an up and down one, the reproducer of the phonograph should be turned as it would be for "hill and dale" records. A tiny drop of solder should be dropped upon the diaphragm of the seventy-five ohm receiver and a small nick placed in the center of the solder for receiving the needle of the phonograph reproducer.

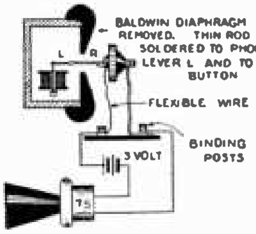
BALDWIN DIAPHRAGM REMOVED. THIN ROD R SOLDERED TO PHONE LEVER L AND TO BUTTON
L
R
FLEXIBLE WIRE
BINDING POSTS
3 VOLT
75

Fig. 9. Button attached to Baldwin 'phone.

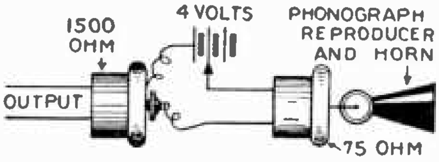
1500 OHM
4 VOLTS
PHONOGRAPH REPRODUCER AND HORN
OUTPUT
75 OHM

Fig. 10. Loud speaker operation from 1-tube set.

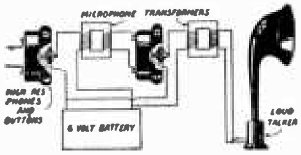
MICROPHONE TRANSFORMERS
HIGH RES PHONES AND BUTTONS
6 VOLT BATTERY
LOUD TALKER

Fig. 11. An effective radio amplifier.

Very often it is desired to further amplify the music or sounds picked up by a transmitter button. For this purpose, microphone transformers, or as they are better known, modulation or telephone induction transformers, are employed. In Fig. 11 this instrument is illustrated, together with the circuit for its use. Here the incoming radio signals are made to operate the microphone fastened to the receiver. The current which the microphone passes then goes through the primary of the microphone transformer, the secondary of which is connected to another receiver similarly fitted with another transmitter button. The output of the second microphone transformer is then connected to a loud speaker. A diagram for using but one stage of amplification through transformers is illustrated in Fig. 12. For the better class of work, however, it is desirable to use a standard vacuum tube amplifier as illustrated in Fig. 13. Here the microphone button is placed in series with the primary of a modulation transformer, and in series with several dry cells or flash light batteries. The secondary of the modulation transformer then connects to the grid and filament, the same as it would in a standard vacuum tube amplifying circuit. The music from the output is, of course, very loud.

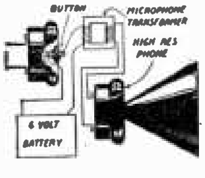
BUTTON
MICROPHONE TRANSFORMER
HIGH RES PHONE
6 VOLT BATTERY

Fig. 12. A single stage amplifier.

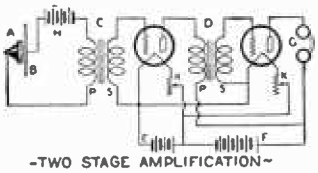
A
H
B
C
D
P
S
P
S
G
K
E
F
-TWO STAGE AMPLIFICATION~

Fig. 13. A 2-stage vacuum tube amplifier.

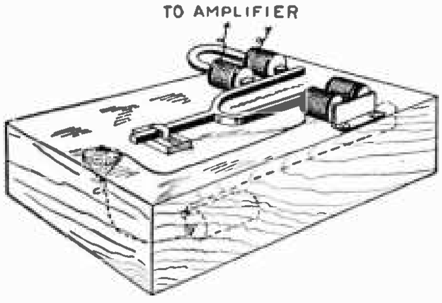
TO AMPLIFIER

Fig. 14. A tuning fork oscillator.

A very splendid oscillator is illustrated in Fig. 14. Here a tuning fork of standard frequency is mounted on a suitable holder in a cigar box. Two low resistance coils, about five to seventy-five ohms, are then placed under one prong of the tuning fork. These are connected in series with the transmitter button fastened in the cigar box, which serves as a base for mounting the tuning fork. The batteries may be concealed in the cigar box. Thus when the tuning fork is struck, it will continue to vibrate as long as the current is applied to the microphone. To use this as an oscillator, a pair of magnets from a telephone receiver and the permanent magnets are mounted near the other vibrating prong. The current constantly being changed by means of the vibrating bar of steel produces a standard frequency, which when amplified can be used for making radio measurements of different kinds. Fig. 15 shows the transmitter button connected in a radio circuit for use as a detector of radio waves. Fig. 16 shows a method of making a code practice device. An audio frequency howl is produced which will enable an instructor to teach an entire classroom the radio code. Sometimes the key, instead of interrupting the circuit, merely shunts a small portion of the current passing through the transmitter button, and for this purpose the key would have to be placed directly across the transmitter button and in series with a small resistance. Fig. 17 shows how the transmitter button may be used in Heising modulation. Fig. 18 indicates its adaptation to the variation of grid voltage of the oscillator tube and Fig. 19 shows how the transmitter button may be hooked up to modulate the output current. These last three figures indicate the use of transmitter buttons in vacuum tube transmitters.

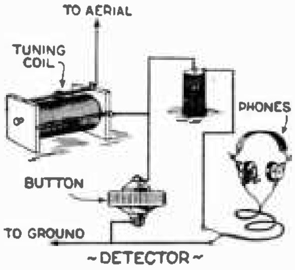
TO AERIAL
TUNING COIL
PHONES
BUTTON
TO GROUND
~DETECTOR~

Fig. 15. Transmitter button used as radio detector.

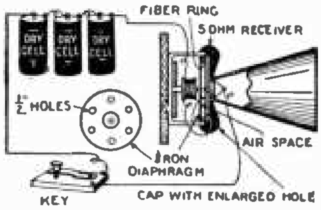
FIBER RING
5 OHM RECEIVER
DRY CELL
DRY CELL
DRY CELL
½" HOLES
AIR SPACE
IRON DIAPHRAGM
KEY
CAP WITH ENLARGED HOLE

Fig. 16. A code practice machine.

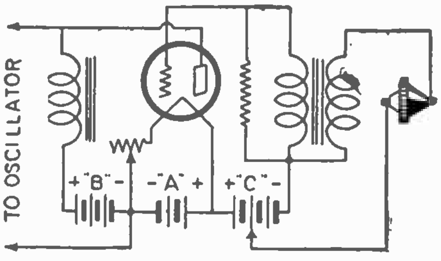
TO OSCILLATOR
+"B"-
-"A"+
+"C"-

Fig. 17. Use of button in Heising modulation system.

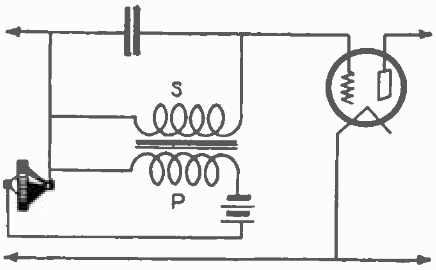

Fig. 18. Method of varying grid voltage of oscillator tube.

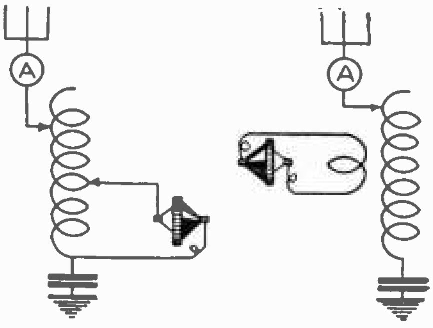

Fig. 19. Method of modulating output current.

SKIN EFFECT — The name given to the crowding of alternating or oscillatory currents into the surface layers of a solid conductor. This phenomenon increases with increased frequency. As a result the resistance of solid conductors is much higher at high than at low frequencies. (See *High Frequency*, also *High Frequency Resistance.*)

SLABY-ARCO-SYSTEM — A system of wireless telegraphy which used a quenched spark gap. This system is now obsolete.

SLABY-ARCO-VACUUM COHERER — A filings coherer enclosed in a sealed glass tube from which the air has been extracted.

A Slaby-Arco vacuum coherer.

SLIP RINGS — Metallic rings, separated by insulators, from which collecting brushes pick up current generated by an alternator. Slip rings take the place of the commutator used on the direct current generator.

SLOTS—Channels cut in the armature core discs and into which the windings fit.

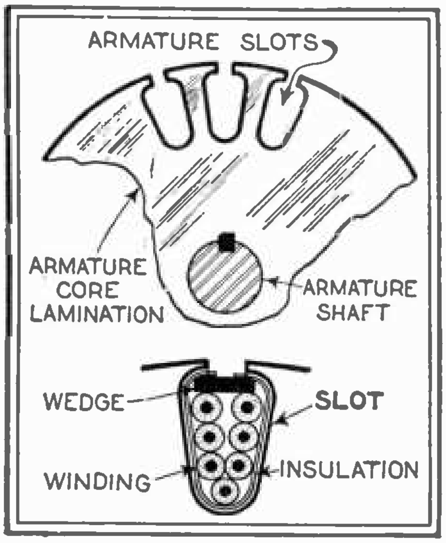

Details of armature slot.

SLUDGING OF TRANSFORMER OIL —The thickening of the oil used for cooling in large transformers. A muddy deposit is formed due to oxidation. The presence of metallic copper in contact with the oil is considered to hasten this action catalytically.

SOAKING-IN—An increase in a condenser charge above its initial value, which takes place gradually if the potential is kept applied. Soaking-in takes place due to a change in the *dielectric* (q.v.).

SOAKING-OUT — A gradual continuation of the discharge of a condenser after the first rapid discharge. In this case, the dielectric is returning to its original state.

SOCKET—A receptacle for holding and making electrical connections with a vacuum tube, an incandescent lamp, or a connection plug. Vacuum tube sockets are made of bakelite, hard rubber, porcelain, pyrex and other suitable *dielectrics* (q.v.). They are usually molded. One type of socket is composed of a shell of insulating material, or in some cases of metal. Four contact blades extend within the base of the socket, for making contact with the prongs of the tube. These blades may be of strong phosphor bronze. Their outer ends are attached to binding posts. It is important that the base supporting the binding posts and contacts be of good insulating material such as bakelite or hard rubber. Modern sockets are of the *spring* or *shock absorbing* type. This construction protects the tube and does away with microphonic sounds. One type of spring socket consists of two molded bakelite parts, the outer one square with a binding post at each corner and the inner one circular with holes for the tube prongs and steel spring contacts fastened underneath the holes. The circular molding is suspended within the square one by means of two rubber pads. Other types of shock absorbing sockets use spring suspension instead of rubber.

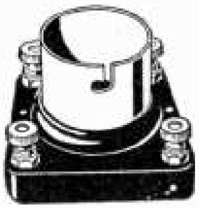

A metal shell socket.

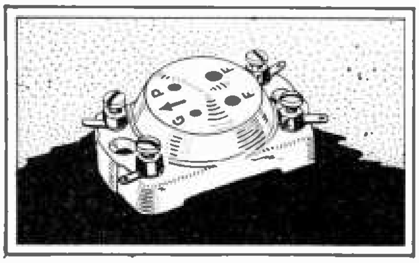

Shock absorbing or spring-type socket.

SOFT IRON INSTRUMENTS — Ammeters, voltmeters and other measuring instruments which depend upon the magnetic force between a current carrying coil and a movable core of soft iron to produce the deflection of the needle or pointer. The movement of the iron may be controlled by a spring or by gravity. These instruments may be used for either direct or alternating current measurements. They are not as accurate as instruments of the dynamometer type.

SOFT TUBE — A vacuum tube which contains a gas content instead of being highly evacuated. Soft tubes were formerly recommended for use as detector tubes but at present have been superseded for this use by the ordinary amplifying tube.

SOLDER—An alloy of lead and tin used for making good electrical connections.

SOLDERED JOINT — A connection between two metallic surfaces, cemented together by means of solder. Where the joint is between two electrical conductors, it must be made both mechanically and electrically secure.

SOLDERING — The process of joining together two or more conductors or other metallic surfaces by means of solder, so as to give a good electrical or mechanical connection, or both. After the surfaces to be united are carefully cleaned, a flux is applied to dissolve the oxides which occur on the surface of the parts to be joined with solder.

SOLDERING FLUX—A substance used to dissolve the oxides on the surfaces of the metals to be soldered. When these oxides are dissolved, the flux enables the solder to enter the minute pores of the metal surface, effectually sealing it against the penetration of oxygen. Fluxes range in character from very strong acids to very mild acid bearing substances. For radio use, it is necessary to use a flux which is non-corrosive and which in its use will leave a residual matter that will have no tendency to collect moisture, dust, or other foreign material. Rosin has been found to be very suitable as a flux in radio soldering. (See *Flux, Soldering.*)

SOLENOID — A coil of wire of helical form used as an electromagnet, generally to attract a movable iron plunger, which is drawn within the solenoid when the current flows through the windings. The solenoid never has a fixed core. A coil wound on a cylinder so as to give uniform magnetic field within it when traversed by a current is often called a *straight solenoid.*

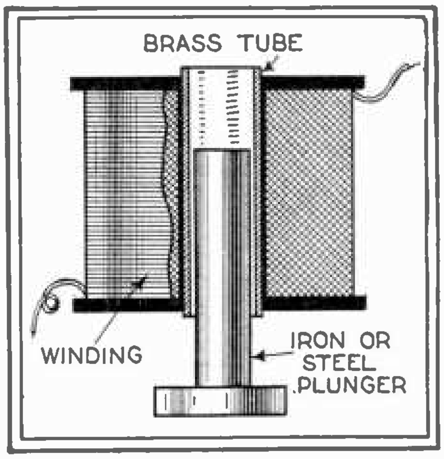

A typical solenoid.

SOLODYNE—A radio circuit which dispenses with high-tension or "B" batteries and which utilizes a double grid vacuum tube. The solodyne principle is also known as the *Unidyne.* The theory of the solodyne is as follows: The small plate current due to the electron emission from the lighted filament passes through the tickler coil which feeds back to the main grid circuit in the usual way, the primary of a step-up transformer, through the telephone receivers, and then back again to the filament of the tube. The electron stream passing from the filament to the plate inside the tube must pass the two grids. The first of these, which may be referred to as the additional grid, is primarily made positive by connecting it directly to the positive terminal of the "A" battery. This in itself tends to assist the electron stream, to reduce the resistance of the vacuum of the tube. But the additional grid is assisted in its work by having impressed upon it the stepped up voltage from the plate circuit, due to the transformer, the secondary of which is in direct connection with the additional grid. Thus a building up process is introduced, every possible electron primarily due to the electron emission of the filament of the tube being made use of and ultimately passing through the telephone receivers to be reproduced in the form of audible signals. The main grid functions in the usual manner, except that this too, can be made to help the additional grid as well, by giving it a strong positive bias. A typical Solodyne circuit is shown in Figure 1. In this circuit L and L^1 may be honeycomb coils or a variocoupler. In Figure 2 is shown the Cowper Solodyne circuit which utilizes standard tubes. In this diagram, the numbers printed near each coil indicate the size of honeycomb or duo-lateral coils to use in each circuit. (See *Four Element Tube.*)

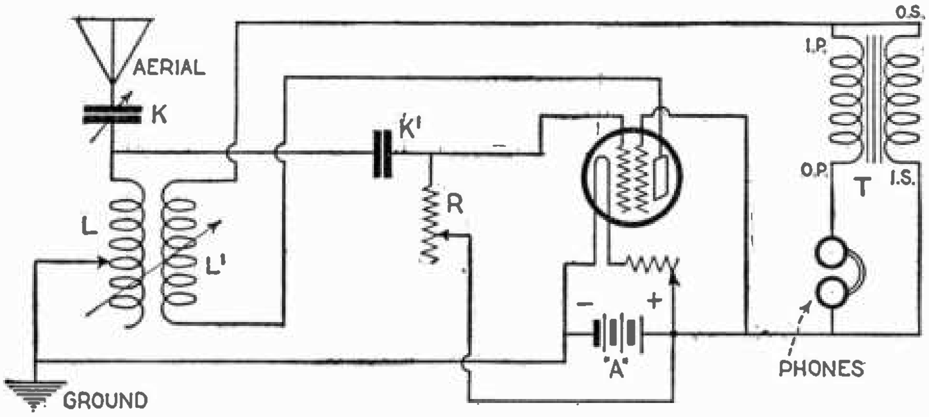

A Solodyne circuit. L and L^1 may be honeycomb coils or a vario-coupler.

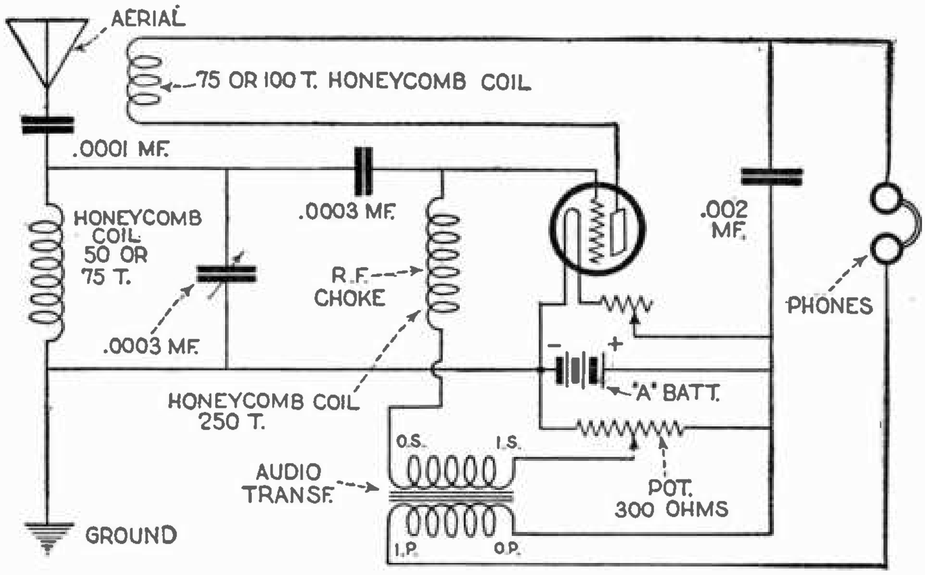

The Cowper circuit uses no "B" battery and is quite stable in operation.

SOUNDER—An instrument used for receiving telegraph signals which utilizes the attraction of an armature by an electromagnet to make sounds as the armature hits against stops at the beginning and end of each current impulse. The sounds thus produced form the characteristic dots and dashes of the telegraph code and can be easily read by the experienced telegraph operator.

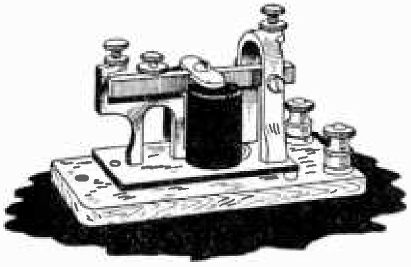

Telegraph sounder.

SOUND WAVE—A wave of alternate condensation and rarefication through an elastic body such as air, water, etc. Sound waves travel at the rate of 1090 feet per second in air.

SOUTH MAGNETIC POLE—Situated in Latitude 70 South and Longitude 102 East. The South Magnetic Pole does not coincide with the geographic South Pole.

SOUTH POLE—When referring to a magnet, the pole which tends to point to the south when the magnet is freely suspended.

SOUTH SEEKING POLE—See *South Pole.*

SPACE CHARGE—The difference between the number of electrons and positive ions in unit volume, multiplied by the charge per ion. This is also called the volume density of electrification. Let us consider the distribution of electrons in a vacuum tube, between the hot filament and the plate. An electron close to the surface of the plate is attracted to the plate by two forces, attraction from the plate and repulsion from the electrons located between it and the filament. An electron close to the surface of the filament, however, is repelled by the electrons between the plate and itself, although it receives some attraction from the plate. It may either move towards the plate or go back to the filament. This will depend upon whether the plate voltage is high enough to result in a force of attraction sufficiently great to overcome the repulsive force of the space charge.

SPACE CURRENT—Current which flows between the cathode and the anode in a vacuum tube. Space current is the result of the motion of electrons through space. It does not follow the same laws which govern the flow of current in a metallic circuit, deviating from Ohm's law to a certain extent, depending upon the relation between the impedances of the tube and the metallic portion of the circuit and also upon the nature of the discharge.

SPAGHETTI—A varnish-impregnated cloth tubing used to insulate bare conductors. (See *Cambric, Varnished.*)

SPARK—See *Spark Discharge.*

SPARK COIL—An *induction coil* (q.v.) used to produce *spark discharges* (q.v.).

SPARK DISCHARGE—A passage of electrical current between two conductors across a previously non-conducting space. The discharge is accompanied by light, heat and sound. It may take place through a liquid, a solid or a gas but whenever it occurs, it means that there has been a breaking down of the dielectric stress. A spark passing through a solid will puncture it. The passage of a spark through a liquid does not affect the dielectric properties of the liquid. On the other hand, a spark passing through a gas, usually ionizes the gas for a period of time after the passage of the discharge and as a result a continuous arc may follow the first spark. An *oscillating current* (q.v.) may produce a rapid train of sparks which persists until the oscillations are damped out. (See *Sparking.*)

SPARK DISCHARGER—A form of *spark gap* (q.v.) such as a *rotary discharger* (q.v.) or a fixed gap, such as a *quenched spark gap* (q.v.). (See *Discharger*, also *Disc Discharger.*)

SPARK GAP—A break in an oscillating circuit which acts as an automatic safety valve to the condenser. Since the air between the gap has a high resistance, the condenser cannot discharge until the potential is sufficient to break down its insulation, thus permitting only heavy discharges to take place. (See *Quenched Spark Gap, Fixed Discharger*, also *Safety Gap.*)

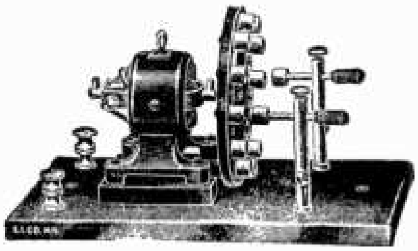

A rotary spark gap.

SPARKING—The production of spark discharges, especially when an inductive circuit is broken. Sparking at the brushes of a motor or generator is caused by armature reaction and self-induction. Sparkless commutation can be obtained by shifting brushes or by the use of *interpoles* (q.v.). Sparking is injurious to contact surfaces and should be avoided if possible. Sparking between contacts of induction coil vibrators is usually reduced by placing a condenser across the gap.

SPARK MICROMETER—A graduated adjustable spark gap for determining sparking distances for various voltages, etc. This device also permits the approximate measurement of high voltages provided the sparking distances for the particular type of terminals, are known.

SPARK RATE, SPARK FREQUENCY, or GROUP FREQUENCY—The number of sparks per second occurring in a spark system of wireless telegraph transmission. It refers to the *Group Frequency* of the wave train rather than to the waves themselves.

SPARK RECORDER—An instrument for recording telegraph signals, in which sparks from an induction coil pass through and mark a paper tape carried on a drum which turns under a metallic pointer. The use of a spark recorder dispenses with the use of ink, thus doing away with the friction of the pen on the paper.

SPARK TRANSMISSION—A wireless telegraph transmitting system which uses a succession of spark discharges in an oscillating circuit, to produce oscillations. These traverse an aerial system and a series of short trains of damped waves are emitted.

SPECIFIC GRAVITY—abbreviation S.G.—The weight of a body compared with that of another, having equal bulk, considered as a standard. The standard for liquids and solids is water, while hydrogen or air is the standard for gases. The specific gravity test of the radio storage battery is an important means of testing the condition of the battery. A *hydrometer* (q.v.) is used for this purpose.

SPECIFIC INDUCTIVE CAPACITY—symbol k—abbreviation S. I. C.—A measure of the degree to which a body permits electrostatic induction through it. The ratio between the capacities of two condensers, one employing the material under consideration as the dielectric and the other using an air or vacuum dielectric. The Specific Inductive Capacity of a material is the *inductivity* (q.v.) of that material relative to that of air. (See *Inductive Capacity* also *Dielectric Coefficient or Constant*.)

SPECIFIC RESISTANCE—The resistance in ohms, of unit length and unit cross-section of a conductor. (See *Resistivity of a Material*, also *Resistivity, Surface*.)

SPECTRUM—An image formed by rays of light or other radiant energy in which the parts are arranged according to their refrangibility or wave length. The image may be visible or invisible. The arrangement is such that all parts of the same wave length fall together, while those of different wave lengths are separated from each other forming a regular series. The spectrum produced by the light of the sun passing through a triangular glass prism and falling on a screen is one of the most common forms. The various colors, since they are unequally refracted, are spread out into a band, showing the seven rainbow colors. The red is at one end (that of the least deviation), while the violet is at the other end.

SPEECH AMPLIFIER—An audio frequency amplifier designed especially for public address systems, where addresses are to be heard over a comparatively wide area. By means of speech amplifiers, it is possible for one speaker to address an unlimited number of listeners.

SPEECH MODULATION—The modulation of radio frequency currents, as utilized in radio telephony. The production of speech modulated waves calls for a source of undamped waves and a method of causing variations in the current output of this source which will accurately correspond to the vibrations of the voice. The source of undamped waves may be a high-frequency alternator or an electron tube generator, the latter being in most common use. The radio frequency antenna current may be varied by inserting a speech-controlled variable resistance, such as a microphone, in the antenna circuit at the transmitter. The microphone may be put in the direct current power supply of the generating system, so that the radio frequency output of the system will be varied as the power input is varied. (See *Modulation*, also *Modulation Frequency Ratio*.)

SPEECH VIBRATIONS—The wave corresponding to a given sound. There are various methods of obtaining graphically, the picture of the wave form of any particular sound. One method is to make a phonograph record of the sound and then as the sound is reproduced from the record, to magnify the movement of the needle, using a lever to trace the form of the waves.

SPIDER WEB COIL—A form of inductance coil in which the wires are wound on a frame consisting of radiating arms similar to the spokes of a wheel. The wire is wound in successive turns in and out around the arms, starting from the center, until it reaches the ends of the arms. An odd number of turns are used so that alternate turns will follow the same wave; that is to say, adjacent turns will be on opposite sides and separated by an arm. Coils of this type have comparatively low distributed capacity and in addition have the advantage of being extremely compact. Figure 1 illustrates the frame on which the wire is wound. This is known as the *spider*. A spider cut from $\frac{1}{16}$th inch bakelite, as shown in Figure 2, can also be used, although, owing to a smaller number of divisions, the latter does not have such a large inductance as the former. In winding, the wire is fastened around one of the pegs and is then taken in and out around the pegs, as shown in Figure 3. When starting the second turn, the wire will be around the pegs the opposite way to the first turn, and all successive odd-number turns will be on the same side as No. 1, while all successive even number turns will be on the same side as No. 2. The finished coil is shown in Figure 4. Spider web coils are also known as *basket wound* (q.v.), *basket woven*, or *stagger wound* coils.

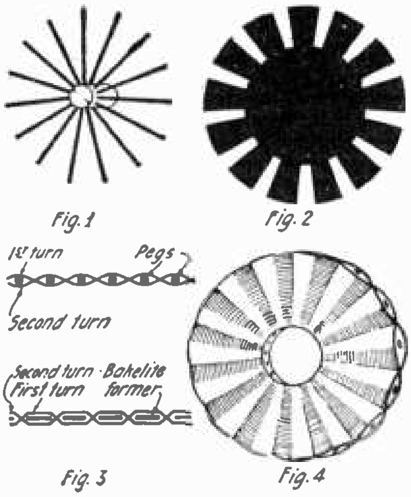

Fig. 1 Fig. 2

Fig. 3 Fig. 4

Details of spider web coil construction.

SPLICE—A method of joining two or more conductors by interweaving or entwining the strands, in a similar manner to that of splicing a rope.

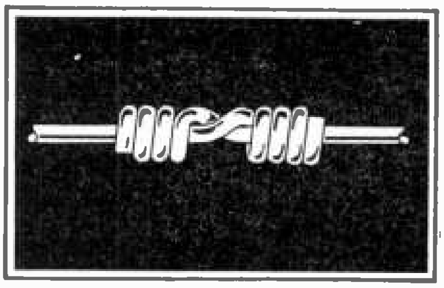

A strong splice.

SPREADER—A spar or pole used on an aerial, where two or more wires are used, to keep these wires properly spread out and parallel to each other.

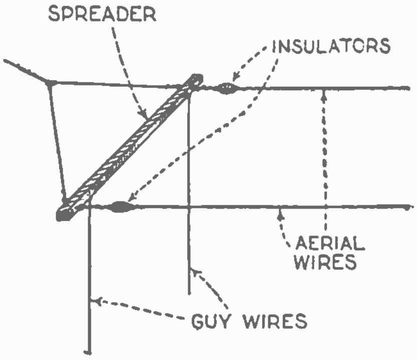

Aerial showing use of spreader.

SQUARE MIL—A square mil is the area of a square, one mil, or 0.001 inch on each side. (See *Circular, Mil*.)

SQUIER, MAJOR-GENERAL SIR GEORGE OWEN—American radio authority. He was educated at the Johns Hopkins University and became a research student under Professor Rowland and Sir William Preece at the British General Post Office. In 1904 he published his famous paper on the absorption of electromagnetic waves by living vegetable organisms and showed how trees could be used for the reception of radio messages. In 1911 he read a paper on multiple telephony before the American Institute of Electrical Engineers. In 1912 he was awarded the Elliot Cresson Gold Medal for his researches in multiplex telephony and in 1919 the Franklin Medal of the Franklin Institute. Major-General Squier was awarded

Major-General George Owen Squier.

the K. C. M. G. for distinguished services during the World War. He is a member of the National Academy of Sciences and the International Electrotechnical Commission. He is the inventor of *Line Radio* (q.v.) which is also known as *Wired Wireless, Wire-Radio Telephony, Guided Wave Telephony*, and *Carrier Current Telephony.*

SQUIRREL CAGE INDUCTION MOTOR—A type of induction motor having a rotor consisting of copper bars connected to rings at each end so as to form a short-circuited system. There are no windings in this type of rotor nor are there any external connections through slip rings. These motors are used for constant speed work, where starting is necessary only at infrequent intervals. The squirrel cage motor draws a large starting current, but has a relatively small starting torque. It is possible, however, by properly designing the rotor so as to have enough resistance, to use small motors of this type for loads requiring frequent starting, rapid acceleration and high starting torque.

SQUIRREL CAGE ROTOR—See *Squirrel Cage Induction Motor.*

STAR GROUPING—A method of connecting up polyphase apparatus or circuits. One end of each phase is connected to a common point, usually called the neutral point. This method of connection is called a "Y" grouping, in the case of a three-phase system.

STATIC—An irregular disturbing noise, heard in the radio loud speaker or head set due to atmospheric discharges, lightning and similar phenomena. A common form of static produces an intermittent crashing sound. Other types cause grinding noises or hissing sounds. Nearby lightning is usually accompanied by a sharp snap in the loud speaker. Various methods have been devised for eliminating or reducing static, although it cannot be said that any of these has been entirely successful. The combination of coil and ground aerials results in the reduction of static; also the use of *beat reception*, using continuous waves as in certain methods of wireless telegraphy. Static is also referred to under various other names, such as *strays, atmospherics, X's*, etc.

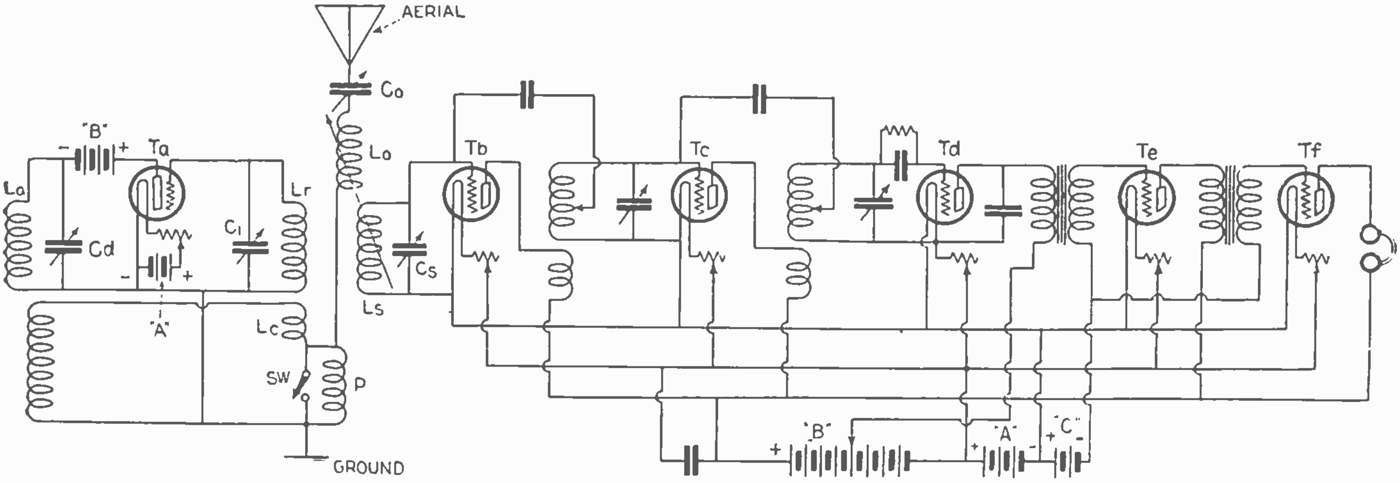

Arrangements of a static reducing device employing repeater circuits.

STATIC CHARACTERISTIC—The curves obtained by the use of a steady direct-current potential for showing the performance, efficiency, etc., of a vacuum tube. The static characteristics of the tube itself are obtained when the resistance of the external circuit is very small in comparison to the plate resistance of the tube. In cases where the external circuit contains reactance, the true characteristic could not be obtained using direct current. In this case, an alternating electromotive force is impressed on the grid and the curves obtained are known as *dynamic characteristics* (q.v.).

STATIC ELECTRICITY—Electricity which is stored in a circuit, manifesting itself in the form of charges at high potential. Electricity, as produced by frictional or influence machines.

STATIC ELIMINATORS—A device for eliminating or reducing the effect of *static* (q.v.). Various circuits have been proposed for accomplishing static elimination. Some of the most recent and successful of these are the McCaa anti-static circuits described in the July, 1925, issue of RADIO REVIEW, on page 63. There are two types of McCaa circuits, one applicable to radio telegraphy and the other to radio telephony. Both have for their object the reduction of signal static ratio. The wiring diagram shown is that of a standard five-tube neutrodyne receiver plus a McCaa anti-static device of the receiver type. In this diagram Ta is the repeater tube, Tb and Tc are the two radio-frequency amplifier tubes, Td is the detector tube and Te and Tf are the two audio-frequency amplifier tubes. The circuit from tube Tb on is unchanged and is the same as any standard neutrodyne receiver. The input circuit of tube Tb, however, and also the antenna circuit, have been altered to make possible the installation of the anti-static device.

For the neutroformer which usually couples the first tube of a neutrodyne to the antenna circuit, the coils Lo, Ls and Lp have been substituted. These coils may either be homemade coils or honeycomb coils. If honeycomb coils are selected each may have fifty turns. The coupling between Ls and Lp should be fixed so that the coils are about an inch apart and some arrangement should be made for changing the coupling between Lo and Ls. If homemade coils are used each coil should have approximately sixty turns of No. 22 d. c. c. wire on a three-inch diameter tube and they should be coupled the same as the honeycomb coils.

The coils Lr and Lc are also coupled and may be either 50 and 10 turn honeycomb coils, respectively, or 60 and 10 turn homemade coils wound on a three-inch diameter tube with No. 22 d. c. c. wire. The coupling between these coils is not variable. After the proper value of coupling has been found by experiment it may be made permanent.

Coils Ld and Lq are also shown coupled in the diagram. They may be 50 and 75 turn honeycomb coils, respectively, or 60 and 90 turn homemade coils wound on a three-inch diameter tube with No. 22 d. c. c. wire. The coupling between these coils is also fixed after the proper value has been determined by experiment.

The proper values for the various capacities used in this circuit are as follows: Co., .001 mfd. (43 plate); Cs, .000035 mfd. (17 plate); Cr., .00035 mfd. (17 plate); Cd, .00035 mfd. (17 plate).

In building this set it is advisable to place both the anti-static device and the receiver in the same cabinet. This cabinet must be shielded and the shield should be connected to the ground. It is also advisable to shield the anti-static device from the receiver proper. The batteries should also be shielded either by placing them inside the cabinet or in metal boxes which are connected to the ground. The wires from the batteries to the set may be effectively shielded by using BX cable, the outside covering of which should also be connected to the ground. For the phone wires special phone cords which are shielded with a flexible copper braid are desirable.

STATOR—The stationary plates of a rotary condenser. The fixed portion of a motor or a generator, which carries a winding. In induction motors, that portion of the motor which is fixed in position and which carries the winding connected to the external circuit. (See *Rotor*, also *Alternator.*)

STATOR PLATES—The plates of a condenser which remain stationary in position as differentiated from the plates which rotate. (See *Shielding*, also *Stator.*)

STOPPING CONDENSER—A condenser which opposes or prevents the flow of direct current, but which allows alternating current to flow in a circuit.

STORAGE BATTERY—See *Secondary Cell.*

STORAGE BATTERY TESTS—The testing instrument most commonly used in storage battery work is the *hydrometer* (q.v.). The purpose of the hydrometer is to determine the state of the charge of the battery, as indicated by the *relative* specific gravity of the electrolyte. The specific gravity of a fully charged battery is not a fixed value but depends entirely upon the standard specific gravity employed by its manufacturer which may be anywhere from 1.250 to 1.300. The user should always ascertain the correct specific gravity for the make of battery before attempting to test it with a hydrometer. A voltmeter may be employed as a means of testing a storage battery provided it is of the proper type and is used correctly. It must be accurate with sufficient length of scale to read tenths of a volt clearly. Voltmeters are obtainable which may be used for both the "A" and "B" batteries whether dry or storage. It is not possible to test the storage "A" battery with cheap types of so-called pocket voltmeters. It must be kept in mind that a voltmeter reading on a storage "A" battery is of value only when the battery is either charging or discharging, except when the battery is practically dead. A voltmeter reading taken when the battery is not connected to the circuit, either charging or discharging, may be misleading. If a reading is taken with the filaments lighted and the voltmeter indicates a full six volts, the battery is charged. Under the same conditions, if the voltmeter lags appreciably below six volts the battery should be charged in order to keep it in the best condition. When it falls to 5.4 volts the battery is discharged and cannot possibly operate the set with any degree of satisfaction. When the battery is connected to the charger and the voltmeter shows from 7.5 to 7.8 volts the battery is fully charged. It is possible to test roughly without instruments. If neither a hydrometer nor a voltmeter is handy it is possible to recognize the necessity for charging by the fact that the amplifying rheostats have to be moved forward of the normal working position, indicating low voltage and that the battery must be charged. When the battery is on the charging line, being charged at the ordinary rate of the commercial charger, and all of the cells of the battery are bubbling, this is an indication that the battery is charged. An interesting meter has recently been placed on the market which should prove of considerable use to owners of storage batteries. It is an ammeter of special design which is to be connected in series with the storage battery on charge. There is a third terminal on back of the instrument by means of which the same instrument can be used to measure the current consumption of the filaments. By pressing a small button in the center of the meter, the needle swings over to the right-hand portion of the scale and indicates whether the battery is fully charged, half charged or low.

STORAGE CELL—See *Secondary Cell.*

STRAIN INSULATOR — An insulator used under tension, as for example the insulators to which the guy wires of an aerial are attached. The simplest form of strain insulator consists of a cylinder or ball having an eye-bolt at each end. Where the aerial guy is fastened to a strain insulator, the purpose of the insulator is to prevent the guy from having a natural wave length nearly the same as the wave length of the antenna. A common form of strain insulator is of nearly spherical form and is so grooved as to carry the two wires firmly, without permitting them to come into contact.

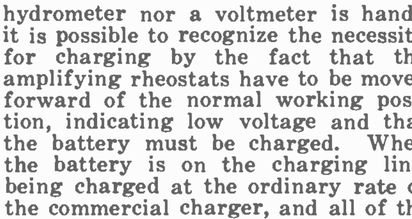
A strain insulator.

STRANDED WIRE — A wire made up of a number of smaller wires, twisted or braided together. An uninsulated, large stranded wire is generally referred to as a bare cable. Stranded wires are of four general classes; (1) wire braid, (2) bunched wire, (3) rope-lay cables, (4) concentric-lay cables. Wire braid is used to afford protection to the insulation of various types of cable. In flat form, it is used as a flexible lead. Bunched wire is a type of cable also referred to as a cord. This form of cable is very flexible, since the individual wires are small. Rope-lay cables are made up of a central core of stranded wire, having one or more layers of stranded wire wound helically about it. Concentric-lay cables are formed in the same way as the rope-lay type, except that the core and layers are of individual solid wires instead of stranded. Stranded wire is very often used in radio work as aerial wire. It has the advantage of being more flexible than solid wire and it also has a lower resistance at high frequencies because of *skin effect* (q.v.). A greater cross-sectional area is available in the stranded conductor than in the solid conductor of the same weight, for carrying current. If used for radio frequency currents, however it is essential that the individual strands be enameled if the lower resistance is to be attained.

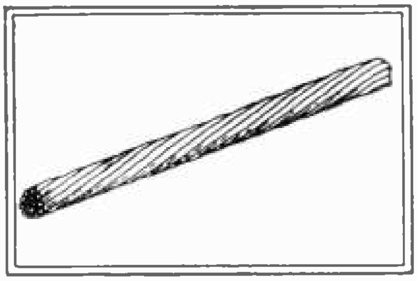
Stranded wire.

STRAY CURRENTS—Currents induced by stray magnetic fields such as eddy currents. Such currents always result in an energy loss. In electric railway systems, the currents returning through the earth, through piping, etc., rather than through the path provided are known as stray currents.

STRAY FLUX—Magnetic flux which is not usefully employed. The *leakage flux* or *leakage lines* (q.v.) which stray from the closed magnetic path provided in a transformer or other electromagnetic or magnetic apparatus. (See *Magnetic Leakage.*)

STRAYS—Atmospheric disturbances which manifest themselves as noises in the radio loud speaker or head set. (See *Static.*)

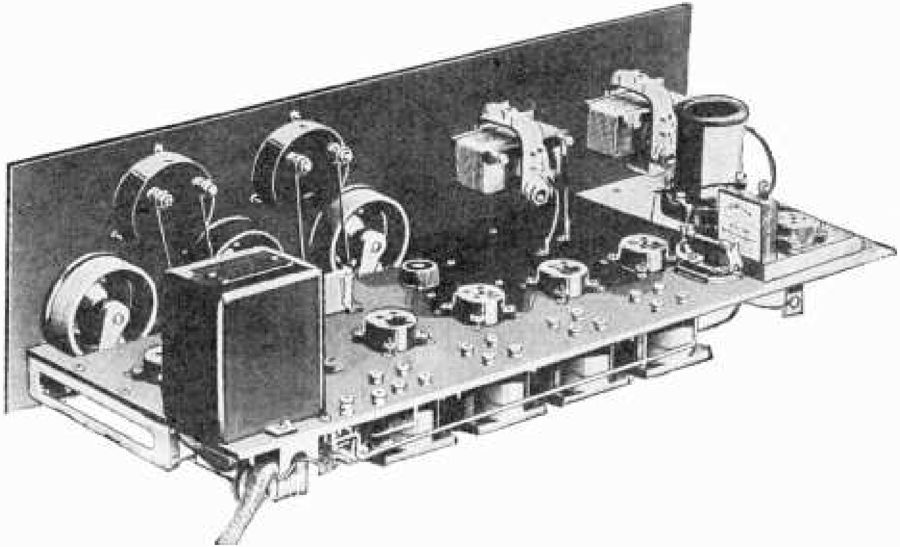
This is the rear view of the Callies Super. Note particularly the long wave transformers at the rear of the set and how they support the sub-panel.

SUB-PANEL—A secondary panel in a radio set, mounted at an angle with the main panel. The sub-panel usually carries the transformers, sockets, grid condenser and grid leak. Sub-panels are made of bakelite or composition, hard rubber, etc. (See *Panel.*)

SUPER-HETERODYNE—See *Super-Heterodyne Receiver.*

SUPER-HETERODYNE RECEIVER—A circuit used in radio reception in which the wave lengths of the incoming signals are increased to several thousand meters, by the aid of a local source of oscillations. It is a recognized fact that radio signals at "radio frequencies," that is, the original signals as they are impressed on the receiving set from the aerial, must be amplified or built up in some way to operate the detector. However, signals at low wave lengths such as in use for broadcasting, cannot be amplified very efficiently at radio frequencies. By changing the waves from the ordinary broadcast band between about 200 and 550 meters to wave lengths of from 4,500 to 10,000, it is possible to obtain more complete amplification. As a result the receiver is more sensitive and greater distance can be received. In the super-heterodyne receiver, a *difference frequency*, termed a *heterodyne note*, is created. This is done by an arrangement for generating oscillations locally. When the incoming oscillations are combined or superimposed with the local oscillations in such a manner that there is a difference in frequency, this difference will be in the form of a new set of oscillations. If the receiver is arranged in such a form that the new set of oscillations has a comparatively low frequency, the wave length will be correspondingly high. In a standard super-heterodyne receiver the action can be imagined as follows: the signal energy

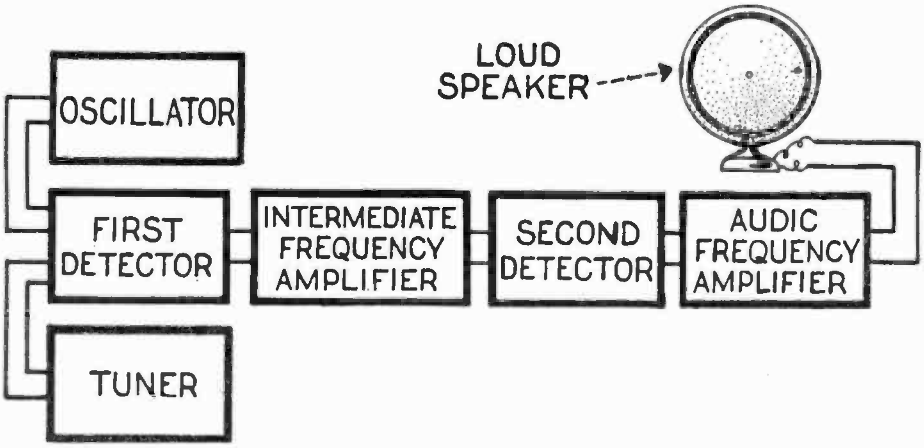
OSCILLATOR
FIRST DETECTOR
TUNER
INTERMEDIATE FREQUENCY AMPLIFIER
SECOND DETECTOR
AUDIC FREQUENCY AMPLIFIER
LOUD SPEAKER

Fig. 1.—This shows the general arrangement of the units of a typical super-heterodyne receiver.

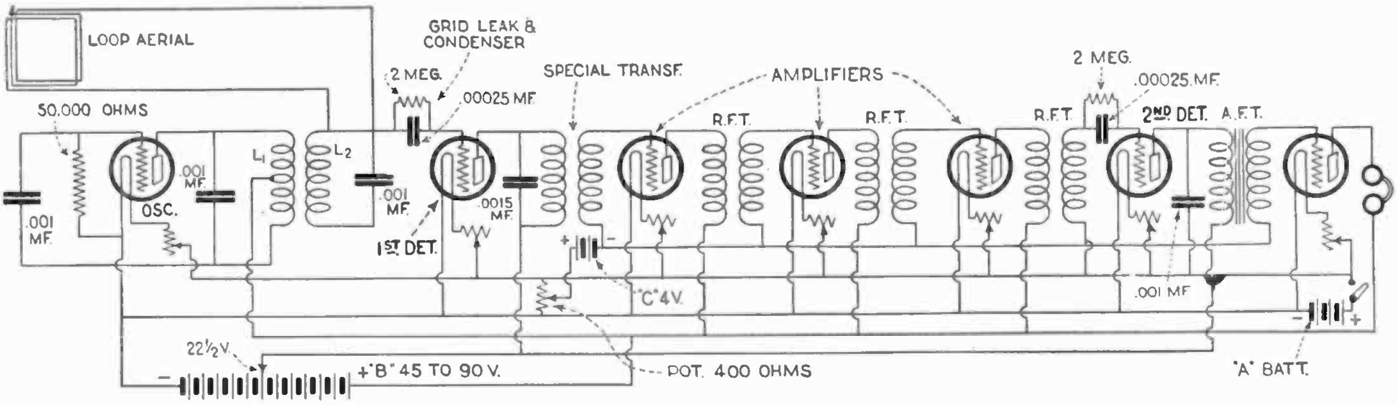
LOOP AERIAL
GRID LEAK & CONDENSER
2 MEG.
.00025 MF
SPECIAL TRANSF.
AMPLIFIERS
2 MEG.
.00025 MF.
50,000 OHMS
R.F.T.
R.F.T.
R.F.T.
2ND DET.
A.F.T.
L1
L2
.001 MF
OSC.
.001 MF
.001 MF
.0015 MF
1ST DET.
"C" 4V.
.001 MF
22½V.
+"B" 45 TO 90 V.
POT. 400 OHMS
"A" BATT.

Fig. 2.—Wiring diagram of a standard 7-tube super-heterodyne.

from a certain transmitting station is gathered on an antenna or loop aerial. These signals are tuned in by means of a condenser in the case of a loop aerial, or by a regular tuning device such as a coupler and condenser in the case of an outside antenna. The energy thus obtained is passed through to the grid member of a vacuum tube, which is known as the first detector tube. (In some cases the energy is first amplified by a radio frequency amplifier, consisting essentially of a vacuum tube and a radio frequency transformer). At this point in the circuit, between the tuning device and the first detector, the apparatus for producing the heterodyne action is located. This part of the outfit consists of a vacuum tube with coupled coils in the grid and plate circuits, with a variable condenser (the heterodyne condenser) connected across them, having a suitable value for obtaining any desired local frequency. In this way, the difference between the incoming oscillations and the locally generated oscillations may be adjusted to a prearranged value or frequency. In inductive relation to the grid and plate coils, another coil is arranged, referred to as a *pick-up* coil. The energy obtained from the first detector tube signals having a frequency equal to the difference between that of the incoming signals and the local oscillations, is sent to a *filter transformer* or *tuned filter*. The construction of this filter and the number of windings used is of the utmost importance, as it determines the frequency of the new set of oscillations and will permit the passage of no other frequency than the desired band. This is a very essential feature, as it is necessary to pass only the desired frequency in order to permit maximum transfer of energy. The energy at this predetermined frequency is then passed to the *intermediate frequency amplifier* (q.v.). After being amplified or built up by the intermediate frequency amplifier consisting of tubes and transformers, the energy is then passed to the second detector tube and rectified or changed from radio frequency. The audio frequency signals are now amplified by one or two audio frequency amplifiers consisting of amplifying tubes and transformers and the resulting signals, speech or music

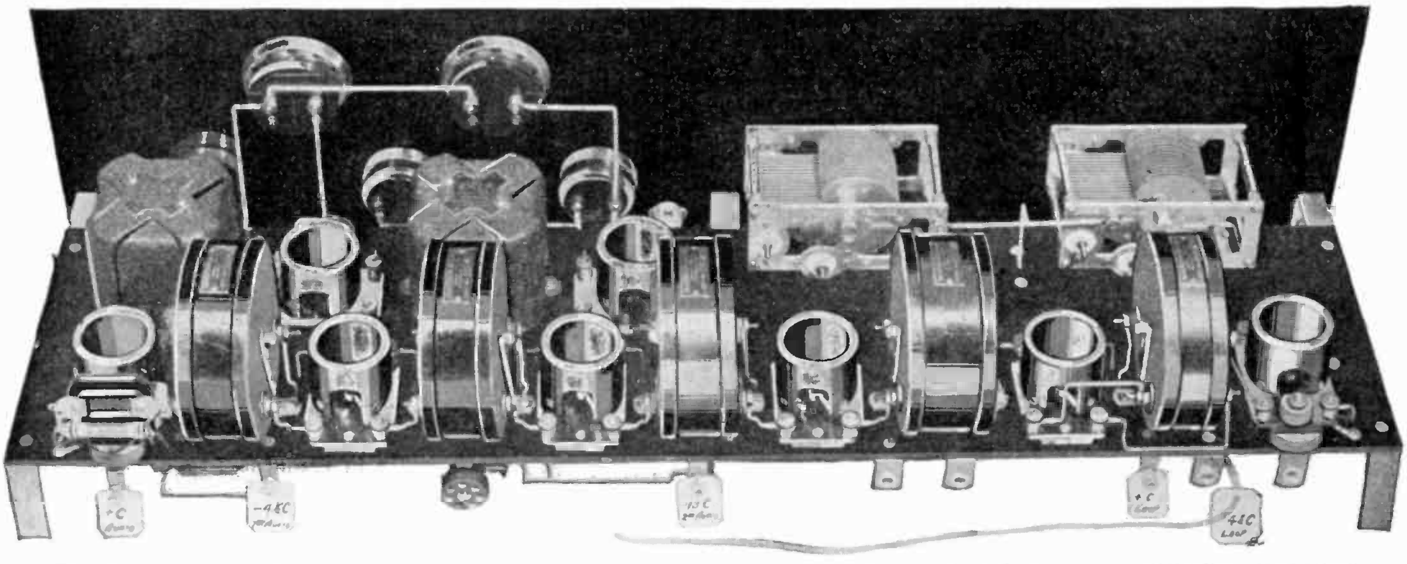

The photograph above shows an eight-tube super-heterodyne set of extra fine quality. Each of the heteroformers are shielded in nickel-plated metal shields. The two variable condensers for tuning the aerial and oscillator circuits should be of the straight line frequency type. The two audio frequency transformers are extra large concert type, the first stage of audio being all that is necessary for average reception on a loud speaker.

are used to operate the loud speaker. Figure 1 shows the general arrangement of the various sections or units of a typical Super-Heterodyne receiver. To go a little more fully into the action of the receiver, let us suppose that a certain station is broadcasting on a wave length of 400 meters. Every wave length is equal to a certain frequency.

That is to say, every wave has a certain number of vibrations per second. The frequency of a wave 400 meters in length will be about 750,000 cycles or 750 kilocycles. By adjusting the wave length condenser, the aerial or loop circuit is placed in resonance with this particular wave length or frequency. This energy is passed through the oscillator pick-up coil and the local oscillations super-imposed on it.

The oscillator condenser, let us say, has been adjusted to permit the oscillator circuit to produce oscillations of a frequency such that the difference between the incoming oscillations and the local ones will be in resonance with the windings of the tuned filter. For instance let us suppose that the tuned filter is arranged for tuning to 10,000 meters, which corresponds to 30,000 cycles.

Now the oscillator circuit must be tuned either 30,000 cycles above or below the incoming wave frequency. In other words, if the incoming wave frequency is as stated—750,000 meters, the oscillator circuit will be tuned to either 780,000 or 720,000 cycles so that the difference between the two frequencies will be 30,000 cycles. No matter what the frequency of the incoming oscillations, the local oscillating circuit will be tuned in a manner to produce the definite difference of 30,000 cycles. These signals will then be passed to the intermediate frequency amplifier and built up before being passed to the second detector tube for rectification.

After this, the signals that are now of audio frequency can be further amplified or built up by the audio frequency amplifier. The two operations that are of the utmost importance in receiving with the super-heterodyne are first, the tuning of the incoming signals and second, the adjustment of the oscillator circuit to produce the difference between two frequencies according to a definite plan. Figure 2 shows the hook-up of a typical Super-Heterodyne receiver.

SUPER-REGENERATION—See *Super-Regenerative Circuit.*

SUPER-REGENERATIVE CIRCUIT—A radio receiving circuit which permits reception under conditions of operation, which in an ordinary regenerative set would result in howling. In the regenerative set, which is tightly coupled, the signal strength increases to a maximum, and thereafter a still further increase in regeneration will result in greatly increased signal strength for a minute fraction of time, followed by violent oscillations. In the super-regenerative circuit, the signals are received at the point of very high amplification by suitably controlling the set. The great advantage of the super-regenerative circuit is the enormous amplification obtainable. Another advantage is the fact that radio code signals are not amplified to the same extent as radio telephony. The super-regenerative circuit was devised by Major Armstrong. The *Flewelling Circuit* (q.v.) is a modification of the Armstrong circuit. (See *Armstrong Circuits.*)

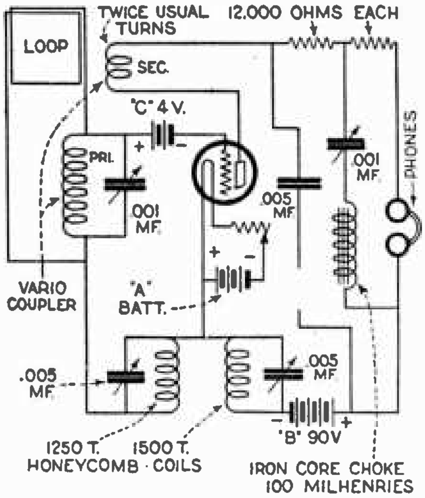

Flivver Armstrong Super-Regenerative circuit.

SURFACE MAGNETISM—A synonym for *Free Magnetism* (q.v.).

SUSTAINED WAVES—Continuous or undamped waves. Waves in which similar current cycles follow each other continuously instead of being broken up into groups like *damped waves* (q.v.). The sustained waves may be interrupted by means of a sending key and thus used for sending radio telegraph messages or they may be modulated and thus utilized in radio telephony. Sustained waves are obtained by the use of high frequency alternators, arc condensers, or thermionic electron tubes (vacuum tubes). The use of sustained waves has made radio telephony practicable. Sustained waves permit sharper tuning, require less energy at the transmitter than in the case of undamped waves, and permit the use of "beat" reception and other sensitive methods of reception. (See *Continuous Waves.*)

S.W.G.—Abbreviation for Standard Wire Gauge. The full name of this wire gage is the *British Standard Wire Gauge*. It is also referred to as the *New British Standard*, the *Imperial Wire Gauge*, and the *English Legal Standard.* It is the legally adopted standard of Great Britain. The following table shows a comparison between the American Wire Gauge (B. & S. Gauge and the Standard Wire Gauge:

Gauge No.	*American Wire Gauge Dia. in Mils*	*Standard Wire Gauge Dia. in Mils*
10	102	128
11	91	116
12	81	104
13	72	92
14	64	80
15	57	72
16	51	64
17	45	56
18	40	48
19	36	40
20	32	36
21	28.5	32
22	25.3	28
23	22.6	24
24	20.1	22
25	17.9	20
26	15.9	18
27	14.2	16.4
28	12.6	14.8
29	11.3	13.6
30	10.0	12.4
31	8.9	11.6
32	8.0	10.8
33	7.1	10
34	6.3	9.2
35	5.6	8.4
36	5.0	7.6
37	4.5	6.8
38	4.0	6.0
39	3.5	5.2
40	3.1	4.8

SWITCH—A device for conveniently making or breaking an electrical circuit. Switches must be designed to carry their rated current without overheating or undue voltage drop, to handle overloads, to prevent arcs on being opened and to properly insulate live parts when switch is open. The most common form of switch is the *knife switch* (q.v.). A switch which opens or closes but one circuit, that is operating in only a single position, is called a *single throw switch.* One operating when thrown in either of two positions is called a *double throw switch* (q.v.). A switch which controls only one side of a circuit is a *single pole switch;* both sides of a circuit, a *double pole switch.* The abbreviation for a single pole, single throw switch is S. P. S. T. The abbreviation for a double pole, double throw switch is D. P. D. T., etc. There are various types of switches ranging from *oil switches* used to handle enormous currents to the small push button or snap switches for turning on or off electric lights. Among the switches used in radio work may be mentioned the *ground switch* (q.v.), *aerial switch* (q.v.), *quick break switch* (q.v.) and *anti-capacity switch.* This latter usually consists of a small handle with a cam attached to its other end which serves to press together or release spring contacts. The electrostatic capacity between the springs is low, due to the construction used. The *plug* (q.v.) and *jack* (q.v.) constitute a form of switch extensively used in radio equipment. The *push-pull switch,* operating as its name implies, is often used to control the filament lighting circuits of radio receiving sets.

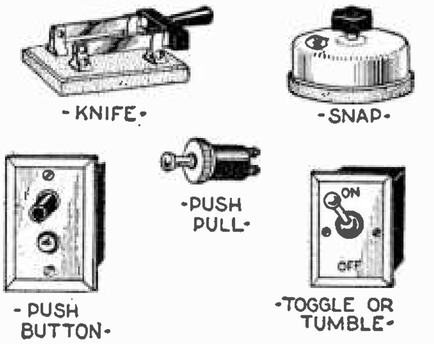

Various types of switches.

SWITCHBOARD—In its broadest sense, this term is applied to any collection of control, operating and measuring apparatus mounted on a panel or panels for the purpose of starting, stopping or otherwise controlling an electrical installation. In a small electrical plant, all control and switch gear may be mounted on a single structure, which is referred to as the switchboard. Switchboards are used at the radio broadcasting station to carry the various meters, control rheostats, switches, etc. (See *Panel.*)

SWITCHGEAR—Apparatus used in connection with the control of an electrical installation, such as switches, rheostats, starters, circuit breakers, etc.

SYMBOLS—Conventional signs used in radio hook-ups to represent various electrical and radio apparatus. The use of symbols has been standardized and results in clearer and more easily drawn diagrams. A table of symbols is given below:

Table of Radio Symbols

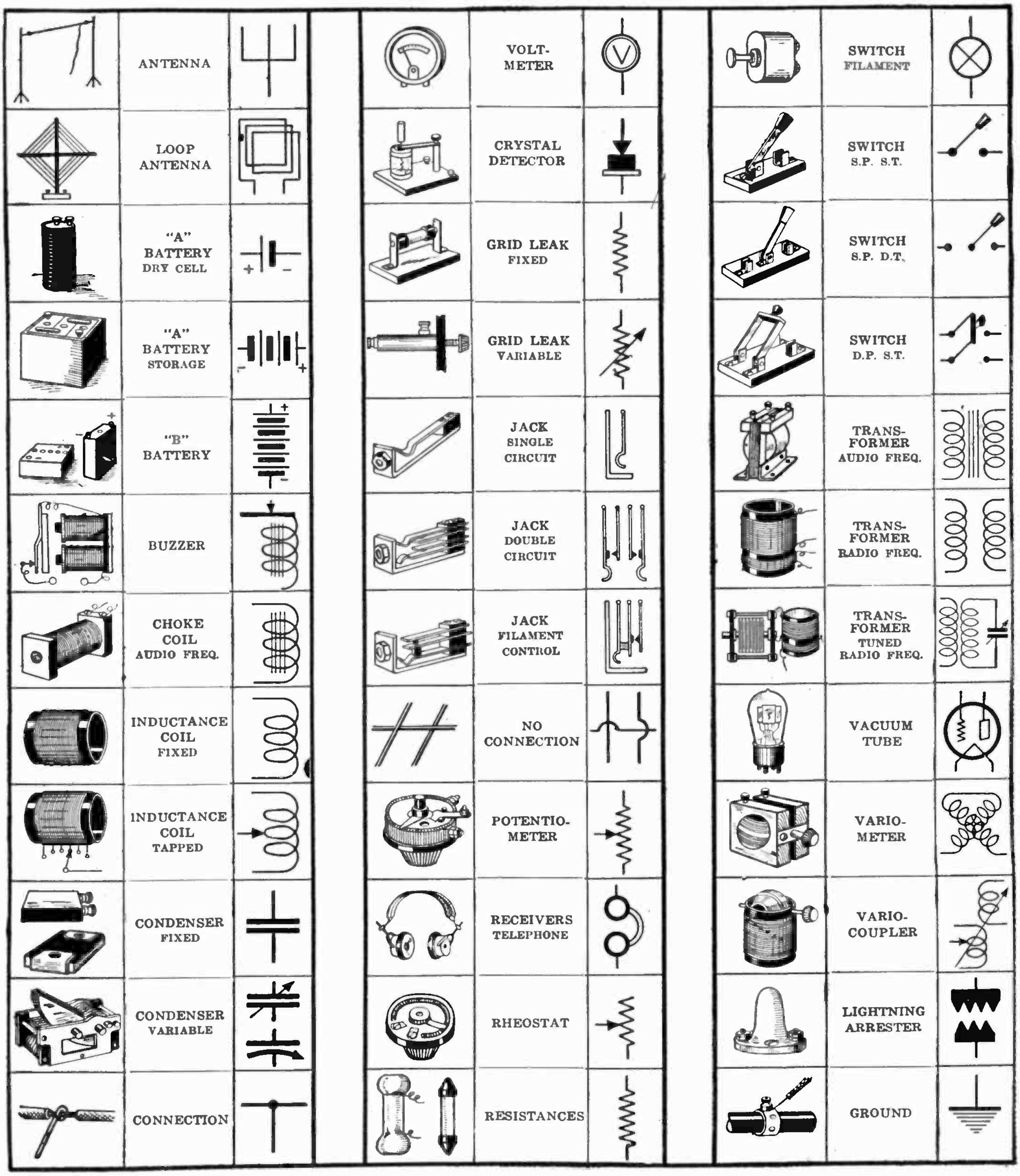

ANTENNA	VOLT-METER	SWITCH FILAMENT
LOOP ANTENNA	CRYSTAL DETECTOR	SWITCH S.P. S.T.
"A" BATTERY DRY CELL	GRID LEAK FIXED	SWITCH S.P. D.T.
"A" BATTERY STORAGE	GRID LEAK VARIABLE	SWITCH D.P. S.T.
"B" BATTERY	JACK SINGLE CIRCUIT	TRANS-FORMER AUDIO FREQ.
BUZZER	JACK DOUBLE CIRCUIT	TRANS-FORMER RADIO FREQ.
CHOKE COIL AUDIO FREQ.	JACK FILAMENT CONTROL	TRANS-FORMER TUNED RADIO FREQ.
INDUCTANCE COIL FIXED	NO CONNECTION	VACUUM TUBE
INDUCTANCE COIL TAPPED	POTENTIO-METER	VARIO-METER
CONDENSER FIXED	RECEIVERS TELEPHONE	VARIO-COUPLER
CONDENSER VARIABLE	RHEOSTAT	LIGHTNING ARRESTER
CONNECTION	RESISTANCES	GROUND

SYNCHRONOUS—As applied to alternating currents, these are said to be synchronous when of the same *frequency* (q.v.) and exactly in *phase* (q.v.).

SYNCHRONOUS DISCHARGER—A rotary gap discharger driven by a synchronous motor from the same line supplying the transformer in a radio telegraph transmitting system or a rotary gap mounted on the shaft of the alternator supplying the 500 cycle or higher frequency current. (See *Disc Discharger.*)

SYNTONY—The relationship between two oscillating currents when one resonates to the waves produced by the oscillations of the other.

T

TABLE OF DIELECTRIC STRENGTH—A table showing the voltages at which certain thicknesses of various materials will break down or puncture. The following table has been obtained from Pender's Handbook for Electrical Engineers:

MATERIAL	DIELECTRIC STRENGTH Specimen thickness mm.	Kv. per mm. (a)
Ambroin	0.84	6.0
Asbestos paper	1.2	4.2
Asphalt (Byerlyte)	3.6	14.0
Bakelite, C-1		up to 27.5
Bakelite, wood molding mixture		17.7 to 21.6
Bakelite, asbestos molding mixture		up to 9.8
Bakelite, Continental	3.2	15.7
Bakelite, Micarta-213		up to 31.4
Bakelite, Micarta-21D		5.9
Bakelite, Micarta-21H		15.7
Bakelite-Dilecto-X	3.2	25.6
Bakelite-Dilecto-XX	3.2	25.6
Celluloid (clear)	0.25	12 to 28
Celluloid (colored)	0.25	10.2 to 18.9
Condensite (molded)	5.7	19.7
Condensite (celoron)	5.7	29.5
Conite	0.13	15.7
Copal	3.0	3.2
Empire cloth, canvas	.41	28.9
Empire cloth, linen	.15	54.0
Empire cloth, muslin	.38	48.0
Empire cloth, silk	.15	48.0
Faturan	3.	10
Fiber, vulcanized, including hard fiber all colors	0.79	8.9 to 16.7
	3.2	4.9 to 10.8
	6.4	3.9 to 8.9
	12.7	3 to 5.9
Galalith (white)		6 to 8.5
Glass (ordinary)		8 to 9
Hermit		2 to 3
Jute (impregnated)	6	1.2
Lava		3 to 10
Litholite	4.5	4.5
Marble		6.5
Mica	6	21 to 28
Micabond, plate	1.6	37.5
Micabond, flexible	1.6	23.1
Micanite, plate	1.6	37.5
Micanite, flexible	1.6	23.1
Minerallac		39.4
Paper	0.13	8.7
Paraffin (parawax)		11.5
Porcelain	20	8
Pressboard (oiled)	0.25	39.3
Pressboard (oiled)	1.58	29.2
Pressboard (oiled)	3.17	21.1
Pressboard (varnished)	0.25	26.3
Pressboard (varnished)	1.58	15.5
Pressboard (varnished)	3.17	9.5
Presspahn		5.2 to 9.3
Redmanol (molded)	5.1	11.8 to 18.5
Redmanol (laminated)	0.8	41 to 51
Rubber (hard)	0.5	70
Slate	10.3	1.3
Vulcabeston	1.9	31.5 to 7.1
Wood (maple), paraffined	15.2	4.6

(a) To obtain volts per mil, multiply kilovolts per millimeter by 25.4.

"T" ANTENNA—A flat top antenna in which the down lead is taken from the center of the horizontal portion. (See *Aerial*, also *Flat Top Aerial.*)

TAILS—Name given to small iron wires forming core of *induction coils* (q.v.).

TANGENT GALVANOMETER—A type of galvanometer in which the strength of the currents which pass through the coil is proportional to the tangent of the angle of deflection of the magnetic needle in the center of the coil. In construction, the tangent galvanometer consists essentially of a small magnetic needle suspended in the center of a large circular coil made of the few turns of insulated wire. Attached to the needle is a light aluminum pointer which allows the deflections to be read on a horizontal scale. The coil is placed in the magnetic meridian with the needle and the coil in the same plane. When the current passes through the coil, the pointer is deflected and the angle of deflection recorded. Knowing the constant of the galvanometer, it is possible to calculate the current in absolute units or in amperes by multiplying the constant by the tangent of the angle recorded.

TANGENT OF AN ANGLE—In a right angle triangle having the given angle as one of its angles, the tangent of the angle under consideration is equal to the ratio of the side opposite the angle to the base of the triangle.

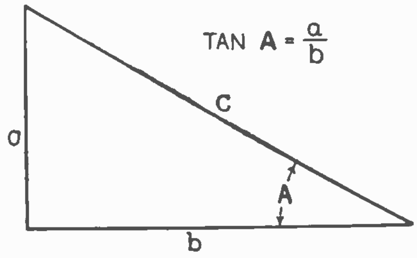

The tangent of angle A is the ratio of side a to side b.

TAP-TAPPING—The connection made to an intermediate point in a winding thus permitting the number of turns in the circuit to be varied at will. Variometers, variocouplers and other coils are made with taps, also transformers, armatures of rotary converters, etc.

TAPPED—See *Tap*, also *Tapped Inductance.*

TAPPER—An electromagnetic device, similar to an electric bell, but with gong removed, arranged so that a hammer gently taps a coherer thus decohering filings after the passage of incoming oscillations. This device was formerly used as a detector of wireless telegraph signals.

TAPPING-BACK—The application of a light blow to a filings coherer to decohere the filings. (See *Tapper.*)

TELEFUNKEN SYSTEM—A wireless telegraph system employing a *quenched spark gap* (q.v.) and giving a singing spark of rather high note. This system has been developed and used by the Germans.

TELEGRAPH—Any system of transmitting intelligence from one point to another over a distance. The word was at one time applied to visual systems such as wig-wagging and semaphore, but at the present is used only to refer to electrical systems. The one exception to this rule is the use of the term to refer to signals sent either mechanically or electrically from the bridge of a ship to the engine room.

TELEGRAPHONE—A magnetic phonograph in which speech currents or signal currents are recorded as permanent magnetic impressions on a moving steel wire. Instruments such as this have been used for recording telephone conversations and also in high speed automatic wireless telegraph transmission. (See *High Speed Reception and Transmission.*)

TELEGRAPH NAUT—A nautical mile of 2,029 yards or 1.1528 statute miles. (See *Nautical Mile.*)

TELEGRAPHY, RADIO—See *Radio Telegraphy.*

TELEPHONE CONDENSER—In radio work, this is a fixed condenser shunted across the head set or loud speaker. Used in this connection, the condenser is also referred to as a by-pass condenser since it offers a path of low resistance to the radio frequency currents. By-pass condensers are usually made of tin foil with a mica or paper dielectric. Condensers used for ordinary telephone work are also referred to as telephone condensers.

TELEPHONE RECEIVER—An electromagnetic instrument by means of which variations of current are caused to reproduce sound waves corresponding to words spoken into a transmitter at a distant point. The essential features of the telephone receiver are an electromagnet, a permanent magnet and a diaphragm held in a suitable case. The diaphragm is supported by the rim of the case, at a short distance from the face of the magnet and it is attracted by the magnet when it is energized by the current flowing in the windings. The steady magnetic flux of the permanent magnet draws the diaphragm to the pole pieces, leaving a small gap between, so that the diaphragm is held under tension. The currents passing through the electromagnet are modulated in accordance with the transmitted speech. These change the tension of the diaphragm due to the change in magnetic pull and as a result the diaphragm vibrates giving forth speech or music. The telephone receivers used in radio work are of the *watch case* type. (See *Head Telephone or Head-Phone.*)

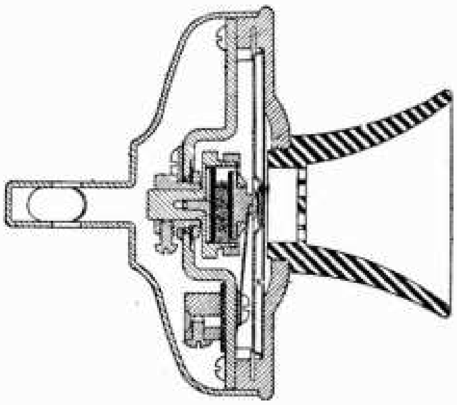

A cross sectional view of a telephone transmitter showing details of construction.

TELEPHONE TRANSMITTER—A sound-wave operated or vibration-operated device designed to produce electromagnetic waves or vibrations which correspond to the sound waves or vibrations actuating it. The standard transmitter used in ordinary telephone work is known as the "solid-back" transmitter. A gong-shaped back supports the transmitter and carries all the parts. A mouth piece of hard rubber is screwed into the front. Just to the rear of the front is placed an aluminum diaphragm which fits into a receptacle cut out for it. An insulated cushion seat of rubber is placed above and below the diaphragm. The diaphragm is held securely by damping springs having soft rubber cushions at their tips, so that it can only assume forced vibrations. An auxiliary diaphragm of mica is used which is fastened to the front electrode of the transmitter button. Both the front and the rear electrodes are made of carbon discs, between which carbon granules are placed in a cylindrical chamber lined with varnished paper. A rigidly connected pin communicates the vibrations of the diaphragm to the front electrode. The movement of the electrode varies the pressure on the carbon granules, thus giving the required variation of resistance which in turn varies the current. The *microphone* (q.v.) used in radio broadcasting is a modification of the telephone transmitter.

TELEPHONE, TUNED—See *Tuned Telephone.*

TELEPHONY, RADIO—See *Radio Telephony.*

TELEVISION—Electric vision at a distance. The transmission by wire or by radio of vision—that is to say the ability to see objects at any distance by changing the light waves to electric currents or to electromagnetic waves, transmitting these to the distant receiving point, and then retransforming them back to the original light waves. Television is still in the experimental stage. Of course, the sending of pictures by radio has been accomplished by Capt. Ranger, in this country, Belin in France, Karolus in Germany, Baird in England, and several others. Both Belin and Baird have been working on the problem of television and are reported to be nearing success. In this country, Mr. C. Francis Jenkins, of Washington, D. C., has constructed a non-commercial (that is in its present stage) machine which will send motion pictures by radio. This is practically the last word in the present development of television. It is possible to broadcast a film picture with this device at the rate of sixteen pictures per second showing the moving outlines as silhouettes, but without the intermediate tones. The method by which a photograph is transmitted by radio in the Jenkins system may be illustrated by the example of a coin covered by a thin sheet of paper on which serried pencil lines are drawn. The design of the coin appears in lines varying in intensity. The variations in the lines may be transformed into pulsing electric current by means of a light sensitive cell passing over the lines in question. Place the lines end to end and the current can modulate a carrier wave. The forming of the continuous line presented a difficulty which was solved in the following way. The inventor proceeded along the line that a prism bends a ray of light and that an image will appear upon a photographic lens placed before a picture. From this it follows that a prism placed near a lens will cause the image to be displaced sideways, the extent of the displacement being dependent on the angle of the prism. If the prism could change its angle, the result would be motion on the part of the image. The prism used is circular, with rapid change of form across the prismatic section. Around the circumference of the disc, the prism changes in thickness according to a graduated scale. Revolving the disc causes the image to move in a straight line. When the end of the prism passes the lens, there is instantaneous snapping back of the picture and the journey is repeated. If a small hole be made in the screen, the light from a given point on the picture will pass through and can be made to operate a light sensitive cell at the back of the screen. When the image is made to move down over the screen by the prismatic ring, the aperture will admit light of varying intensity, according to the light value along a line crossing the picture. This means that the light falling on the photo cell as the image travels will represent a line across the picture from top to bottom. This leads to the creation of the successive lines that make up the picture as a whole. The image is moved along from side to side by means of a second prismatic ring working along at right angles to the first, but at much lower speed. This second prism makes but one revolution during the time that the first makes one hundred, producing the hundred lines assumed to be necessary for making the picture. The cell used may be selenium or thalium oxide. The former requires twenty minutes to cover the picture, the latter six. It is obvious that to reproduce the picture at the other end the process must be reversed, the lines being replaced side by side by means of a pair of prismatic discs similar to those at the transmitting end. As noted above, the main difference between sending photographs and motion pictures is in the necessary speeding up. For this purpose, Mr. Jenkins uses a potassium light-sensitive bulb, the action of which is much more rapid than that of either selenium or thalium oxide. The cell is composed of a bulb with its inner surface coated with metallic potassium. The bulb has a high vacuum. The device has two electrodes, one being a wire through the stem, ending in a loop in the center of the bulb; the other passing through the side of the bulb and connecting with the potassium deposit. Absence of the potassium from a small part of the surface forms a "window" opposite the last named contact. Application of a positive potential to the central electrode, attended by connection with the potassium coating, permits no current to flow as long as no light enters the bulb. The admission of light frees electrons from the potassium, in numbers proportionate to the strength of the light. In this way a current flows through the bulb when light enters the "window." In this cell, there is no appreciable lag due to the amazing speed of the electrons.

The cylinder is covered with plane mirrors, revolving downward on the side toward the lantern. One ray' at a time from 1/25,000 of the area of the image, passes through the opening in C. The fixed mirrors D and F send it back to the mirror G, opposite B on the cylinder, and it is finally reflected against H in a position corresponding exactly to the portion of the image from which it was first taken. The effect of continuous vision is produced.

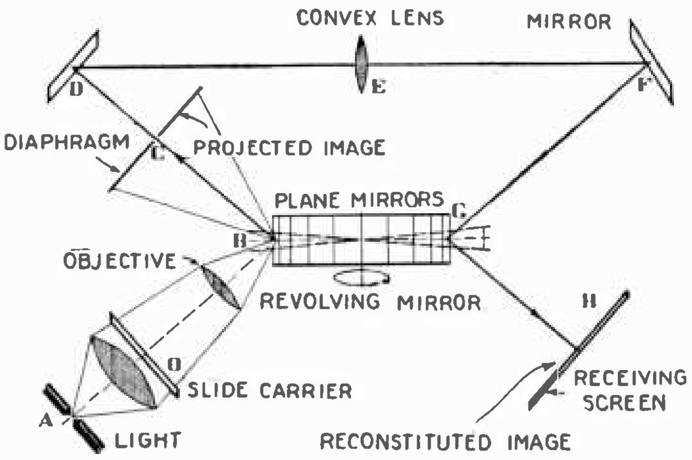

The sending station has a motion picture projector, a small machine used to cut up the image, a light-sensitive cell, and a transmitter. The receiving station has a radio receiving set, a machine for reassembling the picture, and a screen for the reproduction of the image.

The lamp required for the receiving station must be able to pass swiftly from darkness to extreme brilliancy, and the ordinary filament lamp could not attain the tremendous speed involved. The lamp used by Mr. Jenkins is the creation of Professor McFarlane Moore, and is capable of handling complete cycles at the rate of 75,000 to the second. This lamp has a glass bulb containing two concentric cylinders, the larger almost filling the bulb, and the smaller almost filling the larger. The inner cylinder has a small axial hole, drilled almost to the bottom and forming a deep cup. The principle of this lamp is much the same as that of the Amrad "S" valve.

It is a truism that the fundamental principle of cinema pictures is that the human eye continues to see an image

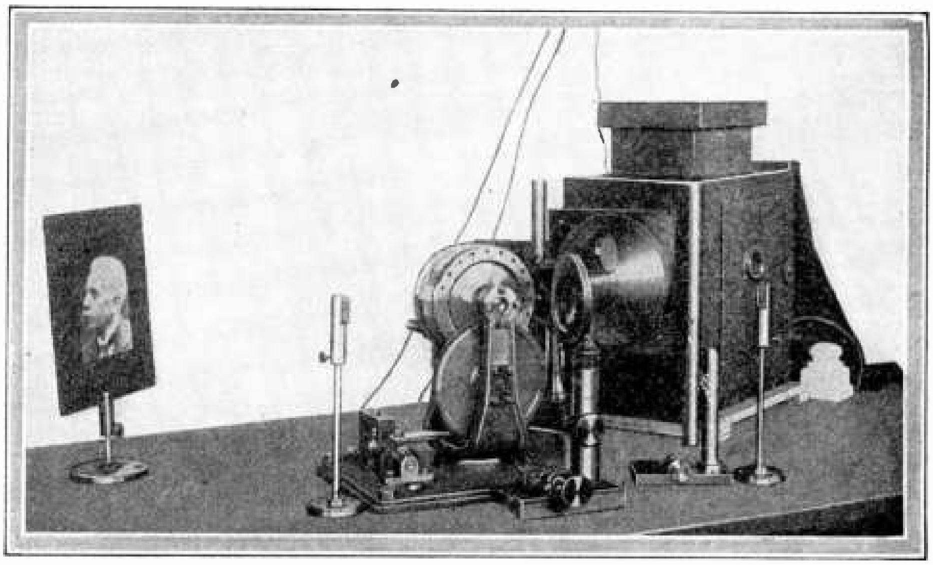

This apparatus is diagrammed above: the projecting lantern at the right; the drum of mirrors, center; and the adjusting stands for diaphragm and fixed mirrors in front. At the left a continuous image appears on the screen, although only 1/25,000 of it is actually projected at any instant.

after that image itself has disappeared. In the case of an electric spark lasting one ten thousandth of a second, the image lasts for at least one-sixteenth of a second as visualized by the eye. Mr. Jenkins assumes that if 10,000 dots of light and shade flash successively on a screen with sufficient speed, the eye will see the picture to which the dots belong even though but one dot is on the screen at a time. He, therefore, aimed at flashing the 10,000 dots on the screen within the space of one-sixteenth of a second.

Theoretically, this feat might be accomplished by halving the slow disc revolving rate to 960 revolutions per minute, keeping the faster disc at a rate exactly one hundred times greater. But the glass disc would be shattered by the centrifugal force before the attainment of a speed of 96,000 revolutions per minute.

Mr. Jenkins gets over this difficulty by the use of forty-eight lenses attached to a large aluminum disc rotating in front of the prismatic disc, thus giving the effect of a slow-moving lens. Each of these forty-eight lenses makes a line across the plate, which thus becomes the screen of the motion picture apparatus. This makes it possible to reduce the speed to one forty-eighth of that which would otherwise be required, that is, to one of 2,000 revolutions per minute.

At the receiving station, the modulated wave is received by a set but slightly different from many in common use. The last amplifier tube is one of 5 watts. By the use of the McFarlane Moore light already described, the modulations in the plate circuit of the last valve are transformed into light, and the rays are distributed over the screen by prismatic and lens discs similar to those used at the sending station. In actual practice, the picture appears on a screen about 6 in. by 8 in. More powerful lamps would allow of larger pictures.

The process described is for sending photographic pictures. But it can be adapted for the sending of direct views without the intervention of photography.

All that is done in this case, is to remove the projector and focus the lens so that it will throw the view as an image.

So far as radio features of the transmission are concerned, there is little to be explained. At the receiving station the circuit resembles the superheterodyne. The picture is at 75 kilocycles frequency and does not interfere with broadcasting.

The carrier wave may be modulated with voice frequency, so that the voice can travel with the picture, which enables vocal explanations to be given simultaneously with the flashing of the film or actual scene being transmitted.

TEMPORARY MAGNETS—Magnets which lose their magnetic properties as soon as the magnetizing force is removed. Soft iron displays the properties of a temporary magnet and hence is ideal for use as the core of electromagnets. (See *Magnet*, *Electromagnet*, also *Permanent Magnet*.)

TERMINAL—A binding post or other fitting attached to the electrodes, ends of windings or other parts of electrical apparatus so as to permit the external circuit to be connected to the apparatus.

TESLA COIL—An oscillation transformer for producing high potential discharges from oscillations of low potential. The *Tesla Coil* or *Tesla Transformer*, as it is sometimes called, is somewhat similar to the ordinary transformer, although it is much more heavily insulated and has the ends of the secondary connected to a condenser which discharges across a spark gap, thus increasing the rapidity of the oscillations, which then pass into a secondary induction coil. This second coil has no iron core. The Tesla Coil consists essentially of a primary winding having a relatively small number of turns of heavy wire and a secondary having a large number of turns of fine wire. An air, glass, or ebonite dielectric separates the primary and the secondary. Due to the sufficient thickness of the dielectric, there is no possibility of a direct discharge between the two windings. Both the primary winding and the secondary winding consist of but one layer each. The primary winding forms part of an ordinary oscillation discharge circuit. A high frequency electromotive force is induced in the secondary winding, due to the high frequency oscillating currents which flow in the primary winding thus setting up an oscillating magnetic field. Due to the high ratio between the windings of the primary and the secondary, very high voltages are set up in the secondary, in some cases as high as a million volts, and hence a very powerful brush discharge can be obtained.

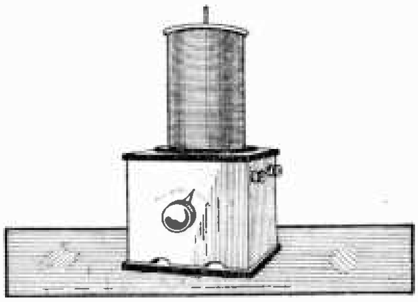

Tesla coil.

TESLA, NIKOLA—American electrical and radio expert. Tesla was born at Smiljan, in Jugo-Slavia, 1857, and was educated at Graz and Prague Universities. He entered the Austrian Telegraph Service and in 1884 he came to the United States where he became an assistant to Edison. He specialized in the study of high frequency and high potential alternating currents. Tesla is noted for his many revolutionary inventions which include polyphase alternating current systems, the rotating field alternating current motor, arc lighting, the *Tesla transformer* (q.v.), and radio apparatus of many kinds including *radio control* (q.v.) devices.

Nikola Tesla.

TESTING CIRCUIT (BUZZER)—A circuit in which a buzzer is used for the purpose of testing or of locating faults in radio apparatus. A buzzer connected in series with a dry cell can be used for locating open circuits and also for testing for short circuits. A buzzer connected in series with a battery and an inductance coil, so as to form a closed circuit and having the inductance coil placed so as to be coupled with the antenna coil of a crystal detector receiving set, can be used to determine the most sensitive point of contact of the crystal. A *tuned buzzer tester* is sometimes used in connection with radio direction finder systems.

THEORY OF CURRENT FLOW—While it is assumed for convenience, that an electric current flows along a conductor from the positive terminal of the source and back to the negative terminal, modern theory is that the electrons actually travel in the opposite direction. Whenever there is a flow of electricity in a conductor, it is believed that extremely small particles of electricity, called *electrons* (q.v.), pass along this conductor. Matter is made up of atoms, which in turn are composed of electrons. The electrons are minute particles of negative electricity revolving around a positive particle. All electric current flow is based upon the motion of electrons within a conductor.

THEORY OF DETECTOR ACTION (OF VACUUM TUBE)—The vacuum tube used as a detector performs a two-fold purpose—it rectifies the current so that it is capable of actuating a telephone receiver and in addition it amplifies or strengthens the current. The rectification takes place due to the fact that the electron flow is only possible from the filament to the plate, but not in the reverse direction. The radio frequency currents applied to the grid regulate the number of electrons passing from the filament to the plate, the variation being exactly in accordance with the variations of the radio frequency current. (See *Detector, Vacuum Tube*.)

THEORY OF OPERATION OF VACUUM TUBES—The operation of the three-electrode vacuum tube depends upon the emission of electrons from the hot filament to the plate and the control of the flow of these electrons by means of the potential variations applied to the grid. The vacuum tube consists of an evacuated bulb containing a tungsten filament treated with thorium, an anode or plate, which may be in the form of a cylinder or plate, and a grid consisting of a wire grating inserted between the filament and the plate. When the filament is heated it gives forth electrons which are attracted to the plate under the influence of a difference of potential by which the plate is maintained positive with respect to the filament. The grid is an automatic control, regulating the flow of electrons from filament to plate. Potential variations applied to the grid serve to vary the electron flow between the filament and the plate. The circuit in which the vacuum tube is used may be divided into two parts, the input circuit connecting the filament to the grid through the secondary of a transformer or other method or applying potential changes to the grid, and the output circuit between filament and plate. The vacuum tube may be used as a detector or rectifier, as an oscillator or generator of high frequency oscillations, as an amplifier, as an electrostatic voltmeter, as a voltage and current regulator, a power limiting device, etc.

THEORY OF PROPAGATION OF ELECTROMAGNETIC WAVES—Waves are due to vibrations transmitted through a continuous elastic medium. Electromagnetic waves are due to vibrations caused by an oscillating electric charge in a circuit (the transmitting aerial circuit). The current flowing in the aerial is oscillating (surging back and forth) at an enormously high frequency and in addition to its induction field, the aerial has associated with it a *radiation field* which spreads out in all directions from the aerial. The strength of the radiation field decreases in inverse proportion to the distance from the aerial. At any given point the strength is directly proportional to the frequency. The transmission of electromagnetic waves is considered as due to the motion of the lines of force. There is a displacement of electricity along these lines against the elastic force of the medium, termed for convenience the *ether* (q.v.). Displaced electricity continuously tends to go back to its position of rest under the action of the elastic forces. There is pressure at right angles to the lines of force in addition to the tension along them. The pressures may be considered as due to the repulsion between the displaced charges of the same sign in neighboring lines of force.

THEORY OF RECEPTION OF ELECTROMAGNETIC WAVES—As electromagnetic waves cut across a receiving aerial, the electric field intensity along the aerial alternates in value. There is an alternating potential between the aerial and the ground which gives rise to the flow of an alternating current. Another explanation is based on the principle of induction. The magnetic and electric fields, in cutting the aerial, induce electromotive forces which cause current to flow. This is radio frequency current and it is passed to the grid of the detector tube and rectified as explained under *Theory of Detector Action* (*of Vacuum Tube*). (See *Electromagnetic Waves.*)

THEORY OF VACUUM TUBE OPERATION—See *Theory of Operation of Vacuum Tubes.*

THERMAL—Pertaining to heat. There are various electrical and radio devices which depend upon heating or thermal properties. The *thermal ammeter* (q.v.) is described under the heading *Hot-Wire Ammeter.* The *thermal detector* is taken up under *Barreter* and *Liquid Barreter.*

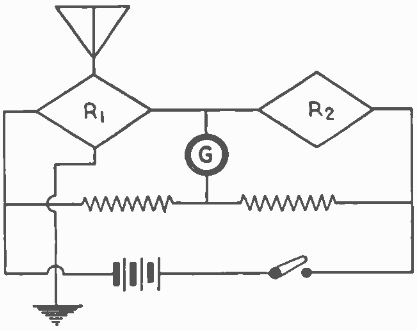

A thermal detector. By means of the change of resistance of a conductor, due to heating, electro-magnetic waves are detected. R_1 and R_2 are two rectangles of very fine iron wire.

THERMAL AMMETER—A current meter, in which the current or a fixed proportion of the current in question, passes along a fine wire. This heats the wire, causing it to expand or sag. This deflects a pointer or mirror across a calibrated scale. The thermal ammeter can be used for measuring direct or alternating currents. (See *Electrothermal Meter,* also *Hot-Wire Ammeter.*)

THERMAL DETECTOR—A radio *detector* (q.v.) which depends for its operation upon the heating of a fine wire by the passage of electrical oscillations. (See *Barreter.*)

THERMAL JUNCTION—**THERMO-ELECTRIC JUNCTION**—The contact *point* or joining point of the two dissimilar metals of a *thermo-electric couple* (q.v.). (See *Klemencic Thermal Junction.*)

THERMAL TELEPHONE—A telephone receiver in which the movements of the diaphragm are regulated by the variations of expansion of a wire heated by the telephone currents thus reproducing sound waves. In another type of thermal telephone, the diaphragm is dispensed with, sound waves being reproduced directly through the expansion and contraction of air in contact with the heated wire. This device is also known as a *thermal receiver* or a *hot-wire telephone.*

THERMIONIC EMISSION—The emission of a stream of negative electrons from a heated filament (cathode) in a vacuum tube.

THERMO-ELECTRIC COUPLE—**THERMO-COUPLE**—A pair of dissimilar metal pieces placed in contact and connected by electrical conductors in a closed circuit. When the junction of the two dissimilar metals is heated, an electric current will flow.

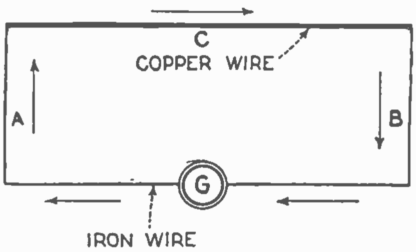

A current is caused to flow by heating the junction of two dissimilar metals. If the end at B is heated the direction of flow will be as shown by the arrows.

THERMO-ELECTRIC CURRENT—A current which is caused to flow due to an electromotive force set up at the junction of two dissimilar metals when these metals are at different temperatures. (See *Peltier Effect.*)

THERMOPILE—A device for magnifying the effect of the thermo-couple by connecting a number of these in series and exposing one set of alternate junctions to the heat, thus causing the electromotive forces to add up.

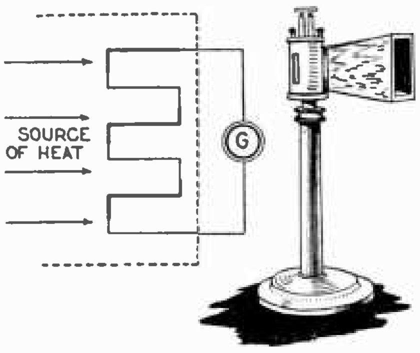

At the left is shown method of joining thermopile junctions to multiply the effect. At the right is a sensitive thermopile.

THOMSON EFFECT—The phenomenon of the appearance or disappearance of heat when a current flows from a cold towards a hot part of a conductor. If copper is unequally heated, heat is liberated at a point when the current and the heat flow in the same direction and is absorbed when they flow in opposite directions. The Thomson effect is measured by a quantity termed "the specific heat of electricity."

THORIUM—A rare metal which in certain compounds in radio-active. This metal has a specific gravity of eleven and an atomic weight of 232.5. Thorium is used for coating Welsbach mantles and is also used in connection with the coating of the tungsten filaments used in vacuum tubes. (See *Filament, Thoriated,* also *Thorium Treatment of Filaments.*)

THORIUM TREATMENT OF FILAMENTS—A compound of thorium, thoria (ThO_2) is used extensively in the manufacture of filaments for vacuum tubes. Sometimes thorium nitrate is used for this purpose. The filament so treated is termed thoriated tungsten. Since the rate of electron emission from a thoriated tungsten filament at a given temperature is several thousand times greater than in the case of the ordinary tungsten filament, it is possible to use the former at a much lower temperature. (See *Filament, Thoriated.*)

THREE CIRCUIT TUNER—A *regenerative circuit* (q.v.) in which the primary, the secondary or grid circuit, and the plate circuit are capable of being tuned in resonance with each other.

THREE-ELECTRODE THERMIONIC TUBE—An evacuated bulb or vessel containing a combination of a heated cathode, a relatively cold anode and a third electrode for controlling the current flowing between the other two electrodes. (See *Triode,* also *Vacuum Tube.*)

THREE-ELECTRODE TUBE—The three element thermionic vacuum tube having a filament or cathode, a plate or anode, and a grid. This tube can be used with minor modifications as a *rectifier* (q.v.) or *detector* (q.v.) of radio signals, an *amplifier* (q.v.), an *oscillation generator* (*oscillator* (q.v.)), an *electrostatic voltmeter* (q.v.), as a *power limiting device,* etc.

THREE WIRE SYSTEM—In direct current work, a system of transmitting electrical energy, which utilizes three wires, one of which is maintained at a potential, midway between the other two. A part of the load is connected between one wire and the neutral and part between the other wire and the neutral. As a result the neutral wire only carries current equal to the difference in the loads. The main advantages of this system are a saving in copper and availability of double voltage. Thus 110 volts may be required for lights, whereas 220 may be wanted for electric motors. In alternating current work reference is sometimes made to a two-phase three wire system. In this system there are two single phase currents which differ in phase by 90 degrees. Each current has a separate outgoing wire, but unites in a common return wire. (See *Neutral Wire.*)

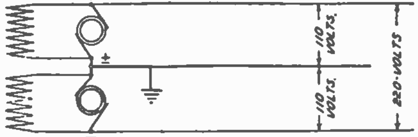

THE EDISON THREE WIRE SYSTEM

TICKER or TIKKER—A form of interrupter or an equivalent apparatus, used in sustained wave telegraphy. It is really a commutator interrupter. In a commercial type illustrated, a disk is mounted on a motor shaft. The disk has a number of teeth filled in between with fibre or other insulating material. The radio frequency currents flow from one brush to the other through the disk which interrupts them at the rate of from 300 to 1000 times per second. A charge is built up in the condenser C_a by resonance with the aerial system and this discharges into the telephone condenser C_b at regular intervals. The telephone condenser then discharges through the head set, thus producing a single sound corresponding to the charge accumulated. Since the tikker discharges the condenser, C_a at different places on the cycle of the incoming oscillations, the note produced is not uniform and is hard to read through atmospheric interference. (See *Chopper.*)

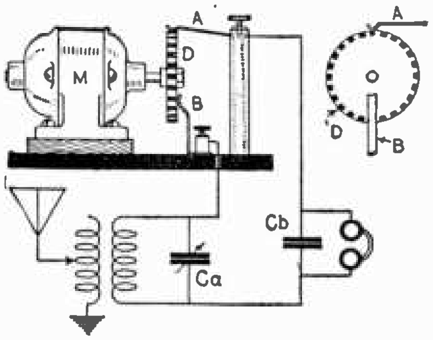

A simple Poulsen ticker. This is used to interrupt circuits of a receiving tuner at a uniform rate per second.

TICKLER—TICKLER COIL—A coil placed in the plate circuit of a regenerative receiver and coupled inductively in the grid coil in order to obtain regeneration. (See *Feed-Back, Feed-Back Coil, Feed-Back Coupling, Regenerative Circuit, Regenerative Coupling,* also *Regeneration.*)

TIME CONSTANT—The ratio, in an electric circuit, of the inductance in henries to the resistance in ohms. It is the time, in seconds, required for a current to attain 63.2 per cent of its ultimate value as given by Ohm's Law. (See *Impulse E.M.F.*)

TONE WHEEL—An apparatus used in the reception of wireless telegraph signals in continuous wave systems, utilizing a wheel in which the conducting teeth are separated by insulating material. A brush is used to make contact with the wheel. The tone wheel is similar in principle to the *tikker* (q.v.) except that it converts the incoming oscillations to audio frequency currents, thus doing away with the non-uniform note of the tikker. In addition to this the tone wheel utilizes more of the incoming energy than does the tikker. The tone wheel is run at such a speed that the frequency of the interruptions differs slightly from that of the incoming oscillations thus giving *beats* (q.v.) which produce audio frequency currents which can be heard in the telephone receiver. Since the tone wheel mechanically converts radio frequency currents into audio frequency currents it is sometimes called a *frequency transformer.*

TORQUE—The moment of a system of forces tending to produce rotation. It is usually measured in foot-pounds. Torque may also be defined as that which produces or tends to produce torsion. Thus if a force acts at a tangent to the periphery of a circle, thus producing a turning moment about the center of the circle, the product of the force and the radius of the circle is called the torque.

TRAIN RADIOPHONE—Communication by means of radio between moving trains and ordinary land-line telephone systems. Such a system is in operation commercially in Germany on trains operating between Berlin and Hamburg. The voice frequency currents from the telephone are impressed on a carrier wave by a method similar to the *Heising Modulation* (q.v.) system. Each train has a special wave length assigned to it. The modulated wave is emitted through a four-wire antenna, strung along the roofs of two coaches. The nearby telegraph wires pick it up and carry it to the receiving sets, located in stations along the line. The ground for the system is obtained through contact of the car wheels with the rails. In reverse order, the voice frequencies of messages from the telephone at the other end are used to modulate a carrier wave (using a different frequency) at the nearest train-telephone switchboard. They are then transmitted over the wires and inductively to the train. An arrangement is provided to prevent interference between the sending and the receiving sets.

TRANSFORMER—A type of induction apparatus, usually stationary, having primary and secondary windings, ordinarily insulated from each other. These are wound about an iron core. The function of the transformer is to change the electrical energy either increasing or decreasing the voltage, current, frequency or phase. Transformers can only be used in connection with alternating currents. Transformers for changing voltage have a different number of turns in the primary than in the secondary. In the step-up transformer, the primary usually consists of a small number of turns of comparatively heavy wire, while the secondary consists of a larger number of turns of fine wire. The ratio of the primary and secondary windings determines the ratio between the input and the output voltages; thus a 10 to 1 step-up transformer could be used to step a primary voltage of ten to a secondary voltage of 100. There are various methods of classifying transformers, namely, according to their operating characteristics, according to their method of cooling, or to their construction. Transformers which are meant to give a constant voltage on the secondary side are known as *constant potential transformers.* Those meant to give constant current are called *constant current transformers.* Transformers may be air cooled, water cooled or oil cooled. In some cases a natural air draft is used as in transformers of the smaller types. Larger transformers may be designed with special ducts through which blowers force a current of air. Oil cooled transformers may have the cores and windings immersed in oil, or the oil may be circulated through external coils. Water cooled transformers are immersed in oil, but have pipes carrying running water submerged in the oil. According to construction, transformers are of the core type and the shell type. The core type transformer is used more for high voltage, low capacity work, while the shell type is used for low voltages and high capacities. Transformers find a wide application in radio apparatus. *Audio transformers* are used to couple audio frequency circuits as for example in the case of an audio frequency amplifier. Audio transformers have an iron core and may be *shielded* or *unshielded. Radio frequency* transformers are used to couple radio frequency circuits. They are usually of air core type, although in special instances iron cores are used. Transformers are also used in alternating current chargers, in "A" and "B" battery eliminators and in power sets which operate directly from alternating current houselighting circuits. (See *Charger, Storage Battery, Audio Amplifier, Coupling Transformer, Jigger, Intermediate Transformer, Inductive Coupler, Oscillation Transformer, Air Core Transformer, Core Transfomer, Ratio of Transformation,* also *Shielded Transformer.*)

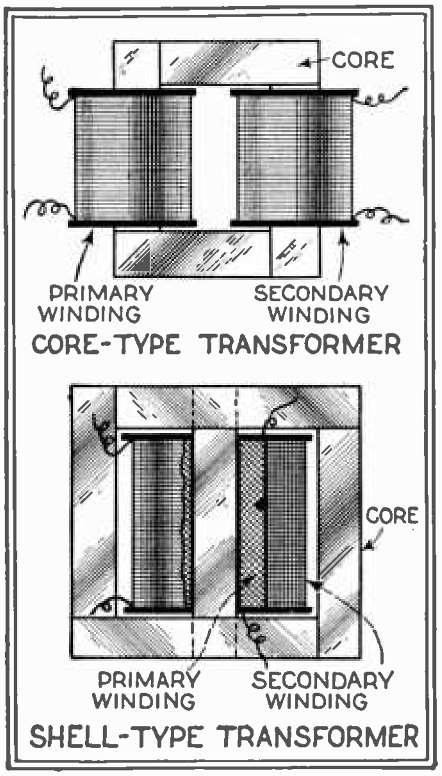

Two standard types of transformers.

TRANSFORMER COUPLING—A method of transferring the electrical energy of an alternating current from one circuit to another inductively by means of a transformer, one of whose windings is connected in one circuit, the other winding being connected in the second circuit. Radio frequency circuits are usually coupled by *radio frequency transformers, audio frequency transformers* being used to couple audio circuits. (See *Coupling,* also *Amplifier, Intermediate Frequency.*)

TRANSFORMER STEEL—A special steel usually containing a high silicon content used for stamping out transformer laminations. (See *Eddy Currents.*)

TRANSFORMER, TUNED—See *Resonance Transformer.*

TRANSMISSION—As applied to radio communication, this is the sending of signals, speech or music from one point to another by means of electromagnetic waves. (See *Radio Telephony, Radio Telegraphy, Theory of Propagation of Electromagnetic Waves.*)

TRANSMISSION OF PHOTOGRAPHS BY RADIO—Various methods have been devised and are now in use for the transmission of photographs by radio. Among these may be mentioned the systems of *Belin* (q.v.), *Baird*, and *Jenkins*. The principles underlying the Jenkins system are explained under the heading of *Television*. Using the system developed by Capt. R. H. Ranger, photographs were transmitted by radio from Honolulu to New York, a distance of 5,136 miles. Recently commercial picture transmission service has been inaugurated between New York and London using the Ranger apparatus. Two distinct methods have been applied for analyzing the picture in the process of transmission. One arrangement consists of producing an image as a non-conducting deposit upon a metal foil which is traversed by a stylus, while the other method makes use of an opaque image deposited upon a transparent film which is traversed by a beam of light, the light interruptions being recorded by a light sensitive cell. The Ranger system makes use of this latter method. The image is photographically recorded upon a celluloid sheet large enough to accommodate easily a picture of half-plate size. In the case of sketches and written messages the image for transmission is made directly by writing upon a piece of transparent film with a dense black ink. The image is then secured to the face of a glass cylinder, and by means of a lamp, focussing lens and reflecting prism, a narrow pencil of light is passed through the film. The cylinder is mounted on a carrier which is caused to be moved backward and forward so that the beam of light is concentrated in turn on all parts of the picture. A rotary motion is applied, as well as the transverse motion, the cylinder being given a slight rotation when it completes each transverse motion. The beam of light is passed through a special photo-electric cell. This consists of a spherical globe, coated on the inside with potassium hydroxide, which is very sensitive to light. The coating is connected to the grid of a vacuum tube, while an "electron collector" near the center of the tube is joined to the plate of the first amplifier. When no light is falling on the deposit on the inner surface of the globe, the grid acquires a negative charge, stopping the flow of electrons between filament and plate, and hence no current flows in the external circuit. The ray of light, however, causes an electron stream to flow between the coating and the collector, and since the coating is connected to the grid, the electron flow constitutes a discharged circuit, so that the grid becomes less negative. The first amplifying tube is a direct current potential amplifier, and is resistance coupled. The grid and plate connections of the amplifier are connected across a condenser which becomes discharged with the fall in the grid to plate resistance of the valve brought about by the grid potential fluctuations. A charging circuit is connected to the condenser and is controlled by a valve, the grid circuit of which operates by variations of the potential across the condenser. The charging current is fed through the plate circuit of this valve, in which a relay is connected, which working through other mechanical relays in cascades, controls the radio transmitter. Wave trains from the transmitting station after detection and amplification, are applied to the picture recorder. The recording mechanism, in order that it may be sensitive to exceedingly small currents, comprises, a small moving coil, in a magnetic field created by three electromagnets. The coil of wire, in moving in the field, as the received fluctuations are applied through its windings, operates a stylus while travelling across the surface of the paper. The stylus traverses the paper in perfect synchrony with the carriage of the transmitter, the paper being lifted each time the stylus completes a forward and backward movement across the paper. A small flashing neon lamp is used to indicate the correct speed adjustment of the driving motor.

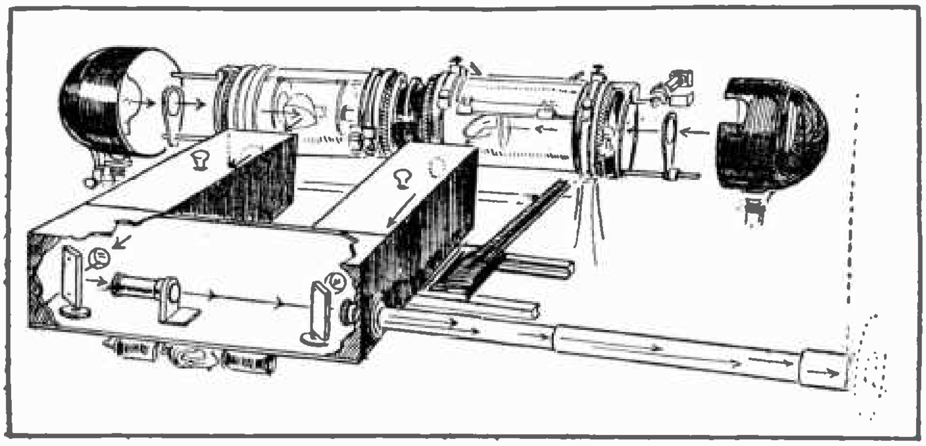

A pencil of light traverses the picture which is attached to the glass drums and is analyzed by a slow rotating action as well as a backwards and forwards movement of the carrier.

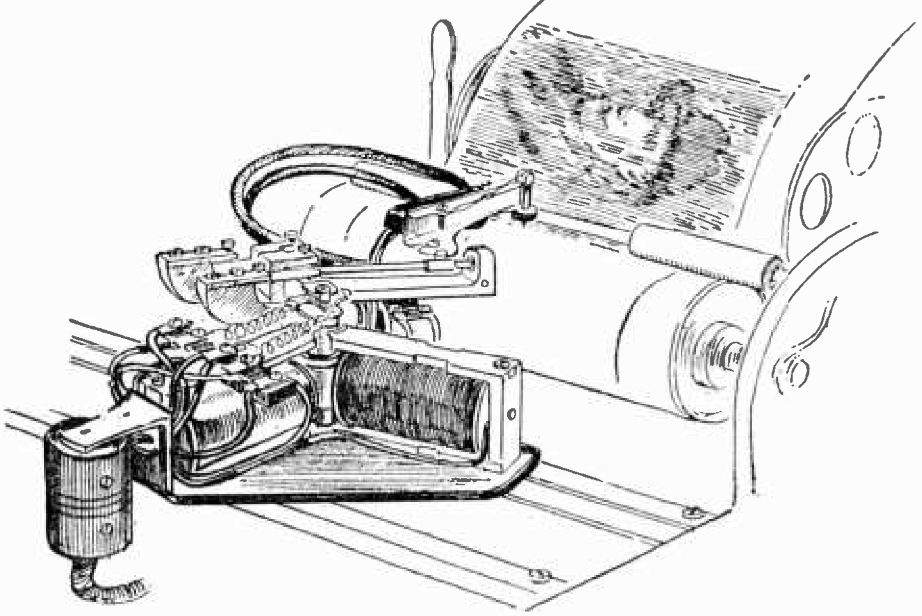

The recording mechanism of the receiver. Three electromagnets produce the magnetic field in which a moving coil controls the stylus.

TRANSMITTER—An apparatus for sending out electrical messages. As applied to radio telegraphy or radio telephony, the transmitter refers to the entire sending apparatus. The term transmitter is often used to refer to a *telephone transmitter* (q.v.). (See *Automatic Transmitter*, also *Wheatstone Transmitter.*)

TRANSMITTING AERIAL—A wire, or more usually a group of wires, suspended at a suitable height and connected to a radio transmitting set. The purpose of the aerial is to facilitate the radiation of the electromagnetic waves generated by the high frequency oscillating current which flows in the aerial. (See *Directional*, also *Receiving Aerial.*)

TRANSMITTING JIGGER—An *oscillation transformer* (q.v.) having a variable secondary, permitting of various degrees of coupling, by adjustment, between the two circuits. (See *Jigger*).

TRANSMITTING KEY—A telegraph key, used in the sending of radio code messages. This key must be of rather heavy construction since it handles larger currents than those used in ordinary wire telegraphy. Pressing the key, closes the circuit, and by holding the key down for a longer or a shorter period, the dots and dashes of the continental code are reproduced. (See *Code, Key, Key High Speed.*)

TRANSMITTING TUBE, POWER RATING—The useful power output from an oscillating tube is the power expended by the oscillating current in the resistance of the output circuit. The power input to the tube, not counting that expended in heating the filament, is the product of the plate supply voltage and the average plate current during an oscillation.

TREMBLER—A springy metallic blade carrying a soft iron armature. The spring makes contact with a fixed contact point, but when current passes through the electromagnet, the armature is attracted and the contact is broken. The spring then resumes its normal position, re-establishing the contact and the same process is repeated again and again. (See *Interrupter.*)

TRIGGER BATTERY—A term (seldom used) denoting a small battery inserted in the grid circuit to give the grid its initial charge when a tube is being used for radio transmission. In this connection, the battery is used to replace the potentiometer. (See *Grid Battery,* also *Grid Bias.*)

TRIODE—A name used to designate the three-electrode type of *thermionic tube.* Other names sometimes applied to the same device are *audion* (q.v.), *audiotron, aerotron, electron relay* (q.v.), *electron tube* (q.v.), *pliotron* (q.v.), *thermionic valve, oscillion* (q.v.), *vacuum tube* (q.v.), etc. (See *Three-Electrode Thermionic Tube.*)

TRUE POWER—The product of the *apparent power* and the *power factor* (q.v.) in an alternating current circuit. (See *Power, True,* also *Power, Apparent.*)

"T" TYPE AERIAL—See *"T" Antenna.*

TUBE OF FORCE—A theoretical conception, similar to *Line of Force* (q.v.) used in mathematical considerations of electrostatic or electromagnetic fields.

TUBE RECTIFIER—A vacuum tube or gas filled tube used to change or rectify alternating current to direct current. The principle of the vacuum tube rectifier is the fact that the electrons will flow from the hot filament to the plate, but not in the reverse direction. Tubes are made for half-wave or full-wave rectification. (See *Rectifying Tube.*)

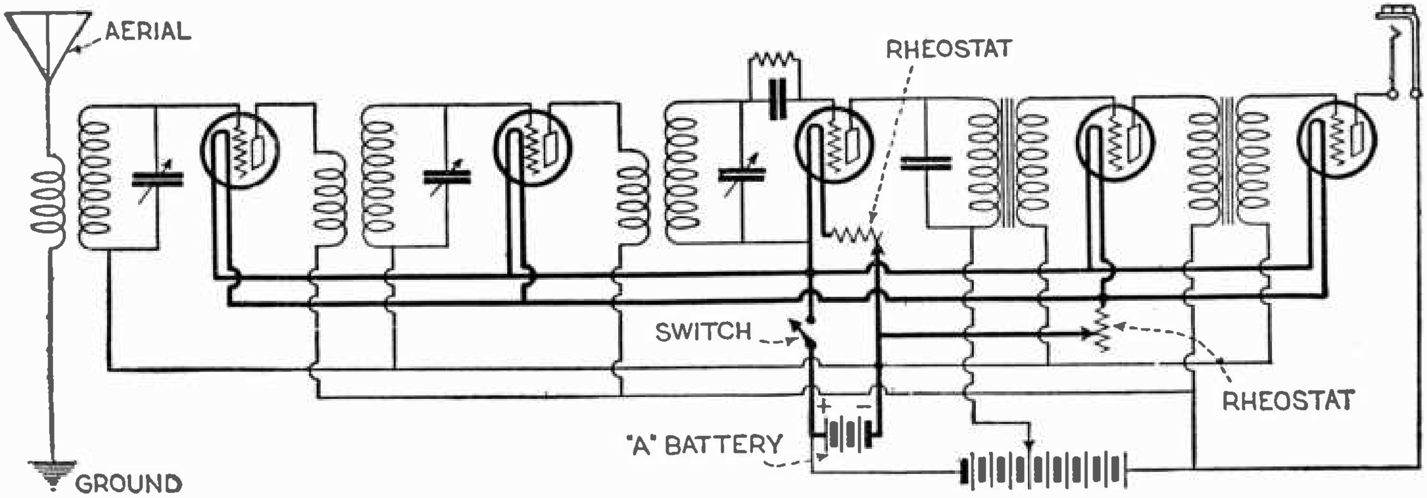

A standard five tube tuned radio frequency circuit.

TUBE, THREE ELEMENT—See *Triode,* also *Three-Electrode Thermionic Tube.*

TUBE, TWO ELEMENT—See *Fleming Valve.*

TUBING—In general, this refers to the bakelite or hard rubber forms used to wind inductance coils on. Another type of radio tubing is defined under *Spaghetti.*

TUDOR ACCUMULATOR—A storage battery having special positive plates formed by the *Plante* process and negative plates of the *Faure* or pasted type. The Tudor positive is of the cast-lead type. This plate has no central web and is made by casting pure soft lead in a mold. Casting is advantageous since it permits the distribution of the metal in the plate without limitations in the manufacturing process. After the plate is removed from the mold, its surface consists of a large number of short vertical ribs which run entirely through the plate. These are bound together by short transverse ribs for the purpose of giving strength to the plate. This gives a large surface and since there is no central web, the electrolyte is enabled to circulate through the plate thus working the active material uniformly.

TUNED ANTENNA—A transmitting aerial, having a loading inductance or a variable condenser or both in series with it, for the purpose of changing the antenna wave length.

TUNED IMPEDANCE—This refers especially to a circuit, or to a method of radio frequency amplification in which a tuned inductance connected in parallel with a variable condenser is used as the method of coupling rather than transformer or resistance coupling. (See *Impedance Coupled Amplifier.*)

TUNED RADIO FREQUENCY—See *Tuned Radio Frequency Circuit.*

TUNED RADIO FREQUENCY AMPLIFIER—A combination of a *radio frequency transformer,* a variable condenser and an amplifying vacuum tube in a radio receiving circuit for the purpose of increasing the intensity or amplifying the radio frequency signals.

TUNED RADIO FREQUENCY CIRCUIT—A radio receiving circuit employing radio frequency amplification in which the radio frequency amplifier circuits may be tuned to the desired wave lengths by varying the inductance or the capacity, or both although the usual method of tuning is by means of a variable condenser in parallel with the secondary of the radio frequency transformer. A typical five-tube tuned radio frequency circuit is shown in the accompanying illustration. This consists of two stages of tuned radio frequency amplification, a detector and two stages of audio frequency amplification.

TUNED RADIO FREQUENCY TRANSFORMERS—An air core *transformer* (q.v.) shunted by a variable condenser used in a *tuned radio frequency circuit* (q.v.). (See *Fixed Coils,* also *Low Loss Coils.*)

TUNED TELEPHONE—One in which the diaphragm is adjusted to vibrate at the same frequency as the current impulses to be indicated. In addition the telephone circuit may have its inductance and capacitance chosen to have the same electrical frequency, and a tuned acoustic resonator may intervene between the diaphragm and the ear.

TUNED TRANSFORMER—See *Resonance Transformer.*

TUNGAR—A trade name for a two-electrode vacuum tube especially designed for rectifying purposes. Tungar tubes are made in two ampere and five ampere capacities. They are used in *chargers* (q.v.) and also in radio sets operating directly from the lighting socket without the use of batteries. (See *Rectifying Tube.*)

A typical rectifier tube of the Tungar type. The screw-base allows it to be inserted in a standard lamp socket from which current is obtained for the filament. Connection is made to the graphite disc by means of a Fahnestock connector which clips on to the wire post projecting from the other end of the tube.

TUNING—The process of regulating the inductance and the capacity of a radio circuit in order to be in unison with a desired wave length.

TUNING, BROAD—Reception of a signal over a wide range, that is on a number of wave lengths, rather than on a single one. A receiving set which tunes broadly cannot be used where the broadcasting stations are transmitting on wave lengths near to one another. Too long an aerial may result in a receiving set tuning broadly. (See *Selectivity.*)

TUNING COIL—A coil of wire so arranged that each turn is electrically insulated from its neighbor and with a device by which contact may be made to bring any desired number of turns into the circuit so as to increase or decrease the inductance of a radio circuit so as to tune it to the desired wave length. A variable *inductance* (q.v.) used for tuning an oscillatory circuit. (See *Lengthening Coil.*)

T.Y.K. ARC—A transmitting arc used for radio telephony invented by W. Torikata, E. Yokoyama and M. Kitamura of Japan. This arc uses magnetite and brass electrodes instead of copper-carbon electrodes. A circuit diagram of the T.Y.K. system is shown in the illustration. The materials used for the electrodes are such that a high resistance film forms on their surfaces. This requires a temporary high voltage to start the discharge. This is accomplished as follows: An armature, A, is attached to one of the electrodes. The two electrodes are in contact, and a steady current is caused to flow in the circuit, through the spark induction coil B. This attracts the armature A, drawing the electrodes apart. The break of the spark-coil current at its interrupter induces a high electromotive force which, acting through the coil and the spark gap in series, breaks down the film on the gap electrodes. The power supplied to the gap is 500 volts and 0.2 ampere. A condenser of approximately .05 mfd. is used in the primary oscillating circuit.

T.Y.K. sets have been used only in small units, having a range of 30 or 40 miles.

Several unusual features appear in the circuit shown in the accompanying illustration, which gives the circuit diagram of the T.Y.K. arc.

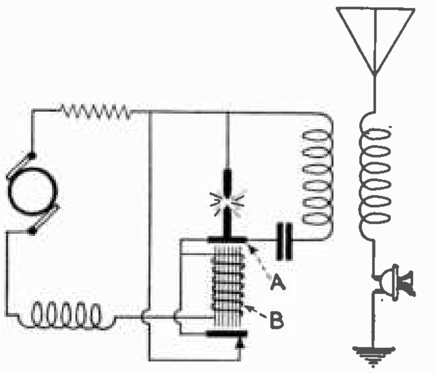

Circuit diagram of the T. Y. K. arc.

U

ULTRA-AUDION—A name sometimes applied to the vacuum tube used in connection with *beat* reception for supplying local oscillations. In general, however, the name ultra-audion applies to a circuit for radio reception. (See *Ultra-Audion Circuit.*)

ULTRA-AUDION CIRCUIT—A type of circuit used for long wave radio reception which uses a form of regeneration, without calling for the introduction of auxiliary equipment in the circuit. A typical ultra-audion circuit is illustrated. It will be noticed that while one terminal of the secondary goes to the grid, the other terminal, which usually goes to the filament, in this case goes to the plate. Thus there is a direct plate-filament circuit which does not involve the plate and filament batteries in the usual manner.

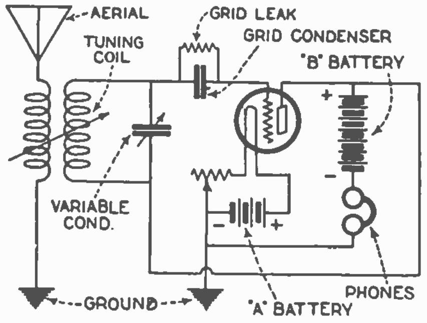

Diagram of an ultra-audion regenerative circuit.

ULTRADYNE—A modification of the *super-heterodyne circuit* (q.v.), which uses the modulation method to produce beats. In this method the incoming signal is caused to modulate the oscillations produced locally in a similar way to that in which the speech or music modulate the carrier wave of a broadcast station. It is claimed that this method is simpler than the ordinary super-heterodyne and also more sensitive to weak signals.

The circuit of the improved model L-2 Ultradyne, incorporating regeneration is shown below. The long wave intermediate transformers are accurately tuned to a wave length of 3000 meters. Referring to the diagram, it will be noted that a special form of oscillator, comprising a grid and plate coil, are used together with an oscillator tuning condenser connected across the grid coil. The first tube, which is usually the detector in the standard super-heterodyne circuit, is known as the modulator tube. (See *Super-Heterodyne Receiver.*)

UMBRELLA ANTENNA—An antenna, the conductors of which form elements of a cone with the apex at the top to which the *down lead* (q.v.) is connected.

UNDAMPED—See *Undamped Alternating Current,* also *Undamped Waves.*

UNDAMPED ALTERNATING CURRENT—Periodic current (i.e., current passing through successive equal cycles of values) whose average value is zero.

UNDAMPED OSCILLATIONS—Oscillations which are sustained. *Undamped alternating current* (q.v.). Oscillations such as are generated by a *vacuum tube oscillator* or by an *arc generator.*

UNDAMPED WAVES—Continuous waves. Waves in which similar current cycles follow each other continuously instead of being broken up into groups. (See *Sustained Waves, Decadent Wave,* also *Damped Waves.*)

UNITS—Specified amounts of physical quantities used as a basis of measurement. (See *Practical Units.*)

UNITS, ELECTROMAGNETIC—See *Electromagnetic Units.*

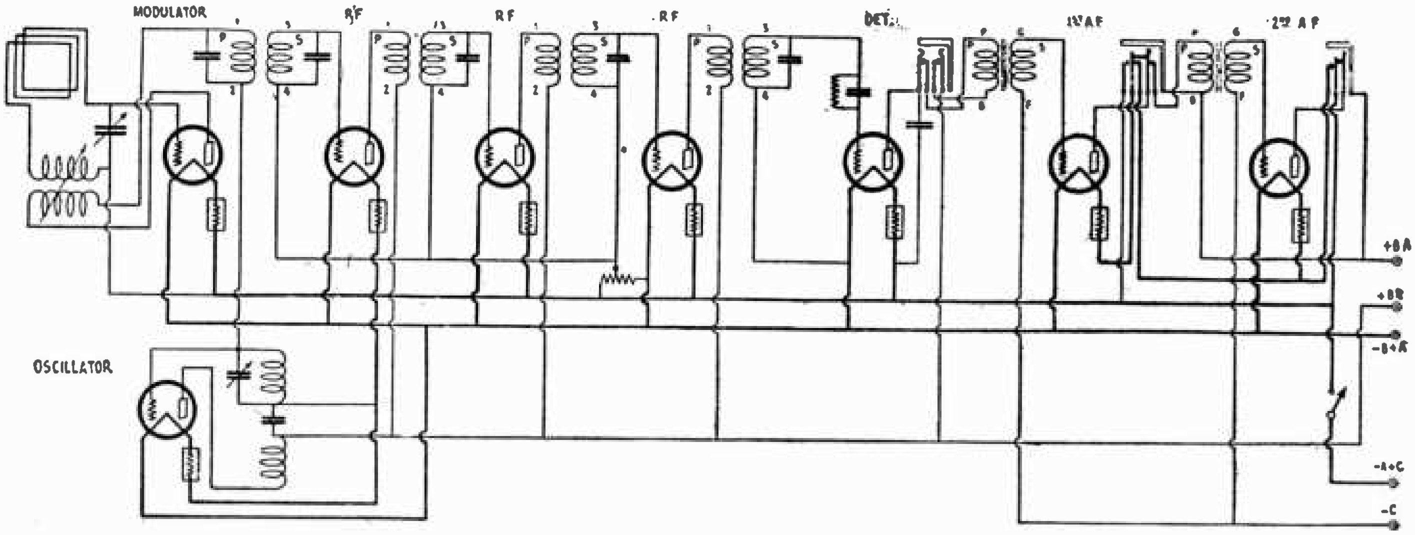

This shows the hook-up of the Ultradyne super-het. This set is not radically different from the average super-heterodyne. The heteroformers are of the tuned air core type and the filter is shown at the input side of the circuit. Aerial and ground as well as loop may be used with this set and instead of the first detector the inventor uses a tube known as the "modulator."

V

VACUUM TUBE—An evacuated bulb, or one containing a rare gas and having three elements, filament, plate and grid. (See *Three-Electrode Thermionic Tube, Triode,* also *Dynatron.*)

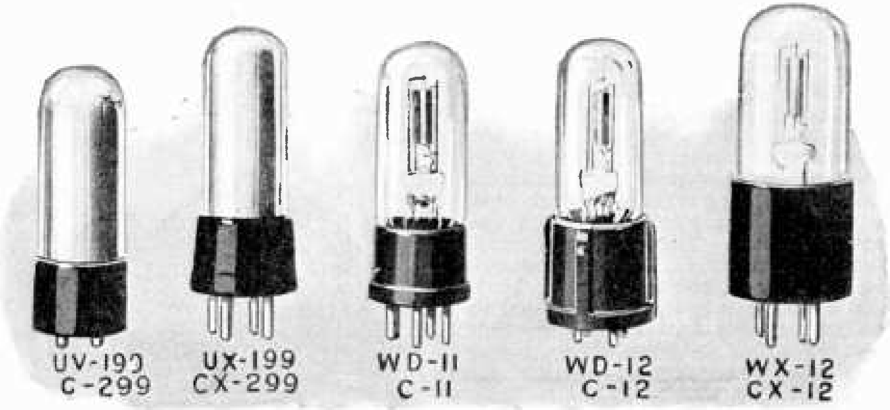

A group of dry-cell vacuum tubes. The first two differ only as to the type of base; both are intended to be operated from a 4½-volt "A" battery, that is, three dry-cells connected in series. The filaments draw very little current, as can be determined from the chart. The last three tubes also have identical characteristics, but are designed to fit different types of sockets. These tubes operate from a single 1½-volt dry cell and draw approximately ¼ ampere of current. The filaments are oxide-coated and very rugged.

VACUUM TUBE AMPLIFIER, THEORY OF—The vacuum tube may be used as an amplifier or magnifier because of the fact that under proper conditions, a small variation of the electromotive force applied between the filament and the grid produces a very large variation in the current flowing in the plate circuit. It is usual to use several amplifying tubes together, suitably coupling the plate of one tube to the grid of the next, thus increasing the amplification to the desired value.

VACUUM TUBE CHARACTERISTICS Performance curves of vacuum tubes. These are obtained experimentally by keeping the filament current constant and applying various known voltages between the plate and the filament and between the grid and the filament and reading the resultant currents that flow to the plate and the grid. The various values obtained are then plotted on cross-section or graph paper. Such curves are convenient for determining the operating characteristics of a particular vacuum tube under consideration. (See *Characteristic Curve.*)

VACUUM TUBE GENERATOR—The vacuum tube can be used to produce sustained oscillations without the necessity of supplying potential variations to the grid from an external source. This can be accomplished as a direct result of the amplification properties of the vacuum tube, since the energy of the output circuit is greater than the energy of the input circuit and part of this output energy may be returned to the input to produce constant reamplification.

VACUUM TUBE, POWER—A tube of special design, made for use in the last audio stage of a radio receiving set, which will handle greater current than the ordinary amplifying tube, thus resulting in more volume and at the same time clearer and more satisfactory tone reproduction. Power tubes now on the market require higher plate voltages than conventional tubes, ranging from 135 volts upwards. The amount of grid bias or "C" battery ranges from approximately 4½ volts to 45 volts, depending upon the type of power tube. Tubes suitable for developing considerable power, in radio transmission are also referred to as power tubes. By increasing the surface area and hence the filament emission or by increasing the degree of evacuation of the bulb, thus allowing a higher plate voltage, the output of the tube may be increased. A form of transmitting power tube is described under the heading *Pliotron.* The illustration below shows one of the new power tubes used for reception.

LAST AUDIO STAGE
UX-171
CX-371

A new storage-battery type power amplifier tube, for use in the last stage audio amplifier only. This tube takes the same filament voltage and filament current as the No. 112 tube but employs a "B" voltage of 180 and 40½-volt "C" battery. Special means should be provided for handling its heavy output.

VACUUM TUBE RECTIFIER—A vacuum tube designed or used to change alternating current to direct current. (See *Rectifying Tube, Kenotron,* also *Tube Rectifier.*)

VACUUM TUBE, THEORY OF OPERATION—See *Theory of Operation of Vacuum Tubes.*

CHART OF VACUUM TUBE TYPES AND CHARACTERISTICS

TYPE	USE	"A" battery volts (Supply)	Filament Terminal Volts	"A" Battery Current Amperes	"B" Battery Volts Det.	Amp.	Negative "C" Battery Volts	Voltage Amplification Factor	Output Resistance Ohms
UV. UX-199 C. CX-299	Det. or Amp.	4.5	3	.06	45	90	4.5	6.25	15,000
UV. UX-200 C. CX-300	Det. only	6	5	1.0	16 to 22½				
UX-200-A CX300-A	Det. only	6	5	.25	Max. 45				
UV. UX.201-A C. CX. 201-A	Det. or Amp.	6	5	.25	45	90 to 135	4.5 9.0	8 8	12,000 11,000
UX-120 CX-220	Pow. Amp. (Lst. stg. only)	4.5	3	.125		135	22.5	3.3	6,600
UX-112 CX-112	Det. or Amp.	6	5	.5	22½ to 45	135	9.0	Var.	Var.
UX-171 CX-371	Pow. Amp. (Lst. stg. only)	6	5	.5		180	40.5		
UX-210 CX-310	Pow. Amp. Oscillator	6	6	1.1		90 to 425	4.5 to 35	Var.	Var.
WD-11 C-11	Det. or Amp.	1.5	1.1	.25	22½	90	4.5	5.6	14,000
WD. WX-12 WX. CX-12	Det. or Amp.	1.5	1.1	.25	22½	90	4.5	5.6	14,000
3VB-199 3VBX-199	Det. or Amp.	4.5	3	.06	20	80	4.5	6.0	
3V-A 3VAX	Det. or Amp.	4.5	3	.12	20	90	4.5	6.5	
5V-A 5VAX	Det. or Amp.	6	5	.25	20	100	4.5 to 9.0	9.4	9,400
5VC 5VX	Pow. Amp. or Det.	6	5	.5	22½	90 to 157½	6 to 10.5	8.6	5,900
99 99X	Det. or Amp.	4.5	3	.06	22½	90 to 150	3 to 12		
O1A O1X	Det. or Amp.	6	5	.25	22½	90 to 150	3 to 12		
MU-20	Audio Amp.	6	6	.25		90 to 150	4.5 to 10.5	20	40,000
MU-6	Pow. Amp. (Lst. stg. only)	6	6	.5		90 to 150	4.5 to 10.5	6	5,000
B-6	Det. only	6	5	.25	16 to 22½				
A	Det. or Amp.	6	5	.25	20	120	4.5 to 9		
B C	Det. or Amp.	4.5	3	.06	20	80	4.5		
E	Pow. Amp. (Lst. stg. only)	4.5	3	.125		135	22.5		
F	Pow. Amp. (Lst. stg. only)	6	5	.5		90 to 180	4.5 to 9		
G	Audio Amp.	6	5	.25		90 to 180	4.5 to 9		
DC—	Det. or Amp.	4.5	3	.06	45	90	4.5	6.3	16,500
DC—	Det. or Amp.	6	5	.25	45	90	4.5	8.5	10,000
DC—	Pow. Amp. (Lst. stg. only)	4.5	3	.125		112 to 135	13 to 22.5	3.3	6,300
DC—	Pow. Amp. (Lst. stg. only)	6	5	.5		90 to 157.5	6 to 10.5	8.0	8,500

VACUUM TUBES, TYPES OF—Vacuum tubes are of various types and designs depending upon the work which they must do. They may be divided into two general classes, receiving tubes and transmitting tubes.

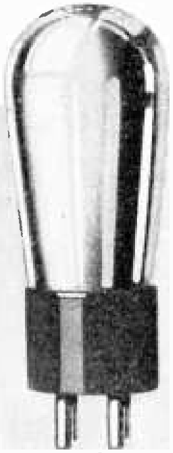

A new type of detector tube which has greater sensitivity than former types, due to the use of an alkaline vapor. It can be used as a detector in any type of receiving circuit.

Tubes for receiving purposes are designed either for dry cell operation or else are of the storage battery type.

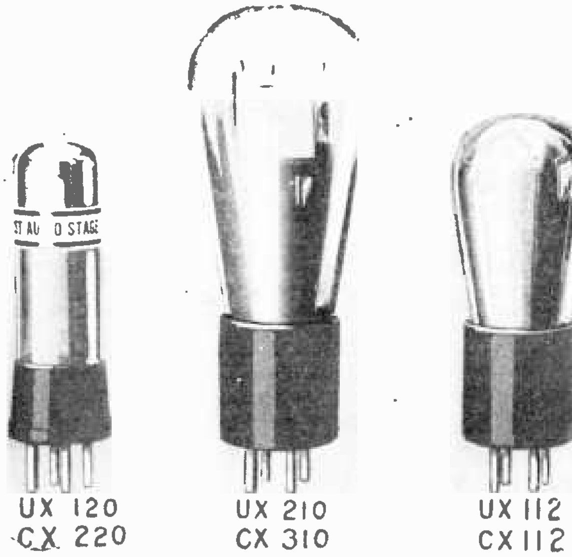

These power amplifier tubes are designed for use in the last stage of an audio frequency amplifier. On the left is a dry cell tube which requires a 4½-volt "A" battery. The center tube employs exceptionally high "B" battery voltage, and is usually employed with an amplifier that operates from the house current. This tube can also be used for transmitting purposes. The right-hand tube is of the storage battery type and has a very rugged filament. It makes an excellent detector tube as well as a power amplifier.

They may be further classified as detector tubes, amplifying tubes or power tubes. There are many special types of tubes, such as *kenotrons* (q.v.), *pliotrons* (q.v.), etc.

VALENCY—The property possessed by elements in combining with or replacing other elements in a certain definite proportion. Also referred to as *valence*. (See *Electro-Chemical Equivalent.*)

VALVE—This generally refers to a vacuum tube used as a detector or rectifier. The name *valve* is used in many foreign countries to refer to the thermionic vacuum tube, as used for any purpose.

VALVE DETECTOR—See *Valve.*

VARIABLE CONDENSER—A condenser having rotable or movable metal plates and an air dielectric. Both the stationary and rotating plates may be made of brass, copper or aluminum. Bakelite, fibre or composition are used to insulate the stationary from the movable plates. Turning the movable plates either increases or decreases the capacity of the condenser. Variable condensers are primarily used in radio for *tuning* (q.v.). They are constructed so as to have straight line frequency characteristics either by undercutting a portion of the movable plates or else by making these of a variable thickness.

VARIABLE GRID LEAK—A variable high resistance unit designed to be placed in the grid circuit of a vacuum tube used as a detector. Certain types of variable grid leaks may have their resistance varied from ¼ megohm to 10 or more megohms by merely tightening a thumb-screw. (See *Grid Leak*, also *Adjustable Grid Leak.*)

VARIABLE RATIO TRANSFORMER—A transformer in which the ratio between the primary and the secondary windings can be varied, usually by means of suitably located taps.

VARIABLE RESISTANCE—Both the *rheostat* and the potentiometer used in radio work, are variable resistances. Variable resistances are usually of high resistance wire type, with a sliding contact so that the resistance can be easily altered.

VARIO COUPLER—A tuner formerly used in radio receiving sets, having a primary and a secondary coil inductively coupled, with the secondary arranged so as to be rotable within the primary. The amount of coupling depends upon the variation of the angle between the axes of the two coils.

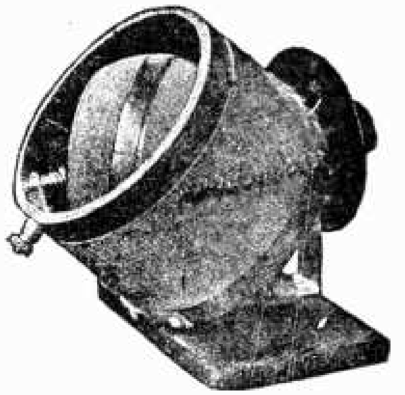

A standard form of vario coupler.

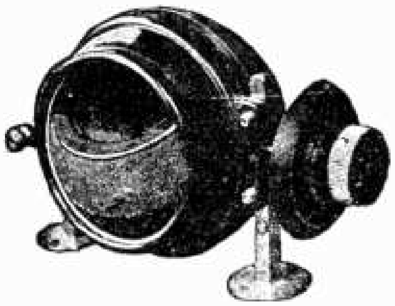

A variometer.

VARIOMETER—A tuner, similar to the vario-coupler, except that the primary and secondary coils are electrically connected.

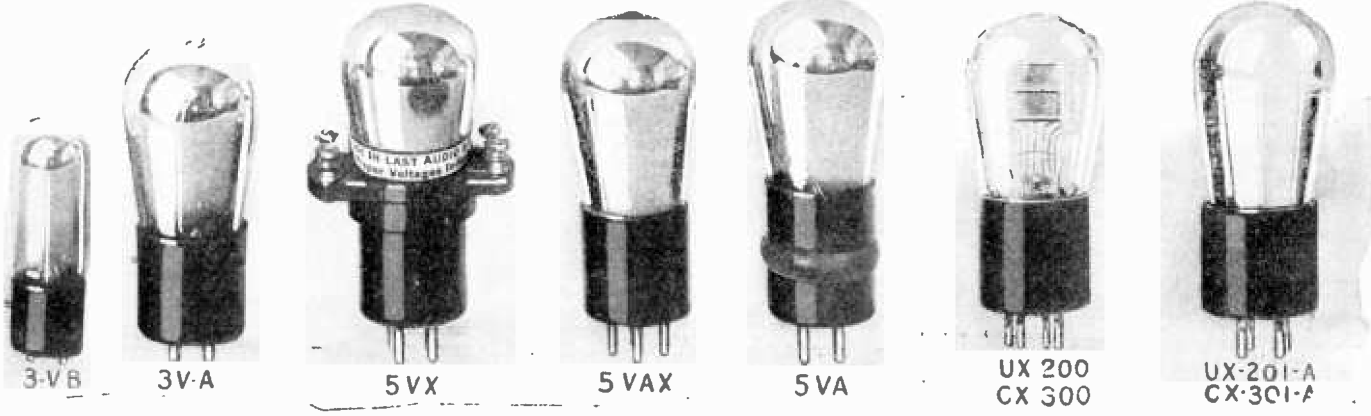

The characteristics of these vacuum tubes are given in the vacuum tube chart on page 187. The first tube is of the dry-cell type, requiring a 4½-volt "A" battery for lighting the filament. 3V-A is a standard 5-volt tube. 5VX is a power amplifier which can be used in the last stage of any audio-frequency amplifier. There is no necessity of changing any of the wiring in the set for the addition of more "B" battery, as binding posts are included on the tube itself. 5VA has a sponge rubber ring included as a part of the base which tends to absorb all vibrations which might otherwise cause the tube to become noisy. The next type of tube is for use as a detector only. The last is the well-known 201-A type, which can be employed as a detector or an amplifier.

VECTOR—A graphical illustration used in mathematical calculations, consisting of a line with an arrow-head at one end, used to show by means of its length, direction and the angle between itself and another vector or vectors, the magnitude, direction and phase angle of alternating current quantities.

VELOCITY—The distance passed through in a certain time. Velocity is measured in feet per second, miles per hour, etc.

VIBRATION—A to and fro motion. An oscillating or swinging motion.

VITREOUS—Consisting of or pertaining to glass. (See *Resinous Electricity.*)

VOLT—The practical unit of *electromotive force* (q.v.). The volt is a measure of the electromotive force which will cause a current of one ampere to flow through a resistance of one ohm. It is equal to 10^8 absolute electromagnetic units (Abvolts).

VOLTAGE—The electrical pressure or the electromotive force between two points in an electrical circuit, measured in volts.

VOLTAGE, CONSTANT—A steady unvarying electrical pressure or electromotive force, as differentiated from a pulsating or fluctuating voltage.

VOLTAGE DROP—The fall in potential caused by the resistance of the conductor through which the current is flowing. The longer the conductor, the greater will be the voltage drop. The greater the cross-section of the conductor, the less will be the voltage drop.

VOLT-AMPERES—The product of volts as measured with a voltmeter, by the current as shown on an ammeter, in an alternating current circuit, gives the apparent power, or in other words the apparent watts. (See *Instantaneous Values.*)

VOLTMETER—An instrument for measuring voltage drop or potential difference. Most of the voltmeters used in radio work operate on the moving coil principle. (See *Galvanometer.*)

A plug-in type of voltmeter.

Panel-type of voltmeter.

VOLTMETER CALIBRATION—A method of correctly marking the voltage readings on the scale of a voltmeter or of checking up the voltmeter readings to make sure that they are accurate. The usual method of calibration is to check the voltmeter reading by comparison with a standard voltmeter. (See *Calibration.*)

V.T.—Abbreviation for *vacuum tube.*

VULCANITE — (EBONITE) — A hard black substance produced by vulcanizing rubber with about 25 per cent sulphur. It is readily polished and has excellent insulating properties.

VULCANIZED RUBBER—Pure rubber, mixed with five per cent sulphur and baked at a temperature of about 150 degrees Centigrade. Vulcanized rubber must not be allowed to come into contact with copper, which is chemically attacked by the sulphur.

W

WALL TUBE—See *Partition Insulator.*

WATCHCASE RECEIVER—A telephone receiver having a shape somewhat similar to a watch. Watch case receivers are compact in construction and are the usual type used in the radio telephone *head set.*

WATT—The unit of electric power. One *Joule* (q.v.) per second. To find power in a direct current circuit, multiply voltage by amperage. The *Kilowatt* (q.v.) equals one thousand watts. Seven hundred and forty-six watts are equivalent to one electrical horsepower.

WATT HOUR—The commercial unit of electrical work. The work done in one hour by a current of one ampere flowing between two points of a conductor having a difference of potential of one volt.

WATT-HOUR METER—An integrating meter which measures energy in watt-hours or in kilowatt-hours. (See *Integrating Wattmeter.*)

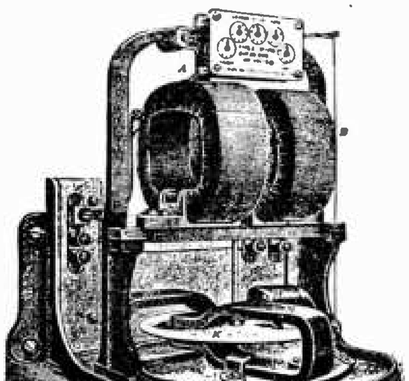

A direct current watt-hour meter.

WATT, INTERNATIONAL—symbol W—The energy expended per second by an unvarying electric current of one International ampere under an electrical pressure of one International volt. IW = 1 Joule per second = 10^7 ergs per second.

WATTLESS COMPONENT—WATTLESS CURRENT—The component of an alternating current which is in quadrature with the voltage. (See *Reactive Component.*)

WATTMETER—An instrument designed to measure the power being expended in an electrical circuit. The most common form of direct current wattmeter is of the dynamometer type. This has two coils, one fixed and one movable, one coil being connected so as to exert a force proportional to the current flowing in the circuit, and the other coil exerting a force proportional to the electromotive force. The *induction* type of wattmeter is in general use for alternating current circuits. (See *Electro-Dynamometer, Dynamometer,* also *Induction Wattmeter.*)

WATT SECOND—A unit of electrical energy, representing the energy expended by one watt flowing for a second. The watt second is the same as the *Joule* (q.v.) and is usually applied to the measurement of heat developed by an electric current.

WAVE—A periodic alternation of an alternating current. As applied to an electric or electromagnetic disturbance, an electric wave is an undulatory movement of the ether, radiated from conductors carrying electrical oscillations.

WAVE ANALYSIS—A study of the wave form of an alternating current or other type of wave. The wave form of commercial alternating current used for lighting purposes approximates a *sine curve* (q.v.). A convenient instrument for the analysis and study of alternating current wave form is known as the *oscillograph* (q.v.).

WAVE ANTENNA—A horizontal antenna, the physical length of which is approximately equal to the length of signaling waves to be received, and which is so used as to be strongly directional.

WAVE DISTORTION—Alteration in the wave form of a wave after it has traversed a considerable distance. (See *Distortion.*)

WAVE FORM—Referring to an alternator, the "shape" of the curve of the current generated.

WAVE FREQUENCY—See *Frequency.*

WAVE LENGTH—symbol λ (lambda)—The distance between two successive *antinodes* (q.v.) in the same direction. In referring to the wave length of electrical oscillations in a circuit, this means the length of the waves in free space that would have a frequency corresponding to the given oscillations. Electromagnetic waves used in radio work have frequencies of from 10,000 to 3,000,000 cycles per second. Since these waves have a velocity of approximately 300,000,000 meters per second, it is possible to calculate the length of a wave by dividing the velocity by the frequency. In other words meters per second divided by cycles per second gives length of the wave in meters. Thus a wave having a frequency of 100 kilocycles (100,000 cycles) will have a wave length of 300,000,000 divided by 100,000 or 3,000 meters.

WAVE LENGTH ALLOCATIONS—In order to prevent radio transmitting stations from interfering with each other, or creating interference at a receiving station, various wave length bands have been allotted to each type of service. The following table shows the present short wave assignments and the service for which they are used:

Wave-length in Meters	*Service*
109–105	Relay broadcasting only.
105–85.7	Public toll service, Government mobile, and point-to-point communication by electric power supply utilities, and point-to-point and multiple address message service by press organizations only.
85.7 –75.0	Amateur, Army mobile, naval aircraft, and naval vessels working aircraft only.
75.0 –66.3	Public toll service, mobile Government point-to-point and point-to-point public utilities.
66.3 –60.0	Relay broadcasting only.
60.0 –54.5	Public toll service only.
54.5 –52.6	Relay broadcasting only.
52.6 –42.8	Point-to-point only.
42.8 –37.5	Amateur and Army mobile only.
37.5 –33.1	Public toll service, mobile, Government point-to-point, and point-to-point public service utilities.
33.1 –30.0	Relay broadcasting only.
30.0 –27.3	Public toll service only.
27.3 –26.3	Relay broadcasting only.
26.3 –21.4	Public service, mobile, and Government point-to-point.
21.4 –18.7	Amateur only.
18.7 –16.6	Public toll service, mobile and Government point-to-point.
16.6 – 5.35	Experimental.
5.35 – 4.69	Amateur.
4.69 – 0.7496	Experimental.
0.7496– 0.7477	Amateur.

Radio broadcast stations work within the wave length band of 200 to 545 meters at intervals of 10 kilocycles. 600 meters is reserved for distress signals. The higher wave lengths are allocated to various marine, government, university, aircraft and other uses.

WAVE LENGTH CALCULATION FOR ANTENNA—For the ordinary vertical wire grounded antenna, the fundamental wave length is slightly greater than four times the length of the wire. A constant suggested is 4.2 and applies approximately also to flat top antennae having vertical lead ins. For other calculations see *Fundamental Wave Length.*

WAVE LENGTH, FUNDAMENTAL—See *Fundamental Wave Length.*

WAVE LENGTH, NATURAL—In a loaded antenna (that is with series inductance or capacity) the natural wave length corresponds to the lowest free oscillation. (See *Natural Wave Length.*)

WAVE LENGTH OF STATIONARY S. H. M. WAVES—On a straight wire, the smallest distance between two points where the disturbance is of the same amplitude and phase; or, since consecutive loops (q.v.) are in opposite phases, the wave length is double the distance between consecutive loops or consecutive *nodes* (q.v.). (See *S.H.M.*)

WAVEMETER—A radio instrument for measuring frequency. A calibrated resonator of variable frequency, capable of easy excitation by the oscillation under test, combined with a means of indicating the attainment of resonance. The calibration may be given directly in wave lengths or frequencies or both.

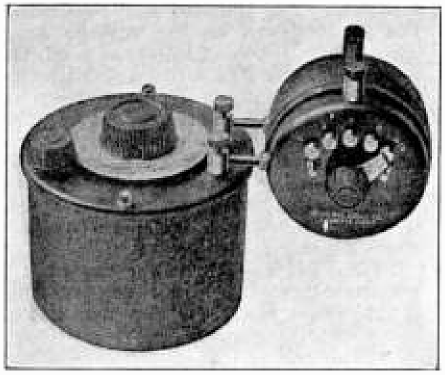

Photo by courtesy of General Radio Co.

The selectivity of the radio set may be greatly increased by the use of this wavemeter as a radio filter.

WAVEMETER CALIBRATION—Wavemeters may be calibrated by comparison with a standard wavemeter. An accurate method of calibrating a wavemeter, known as the *Three-Way Method* is shown in the accompanying diagram. In this illustration A is the standard wavemeter, B the wavemeter to be calibrated, and C is a condenser and coil of size permitting resonance with A. To calibrate, the buzzer should be put in operation, varying the capacity of the condenser C^1 until L^1C^1 is in resonance with L^2C^2 as noted in the head set. The wavemeter B should then be put in inductive relation to C and the condenser C^2 varied until the wavemeter A resonates with C. The wave length of B is now the same as A. This procedure should be carried out over the whole range of the condenser C^3.

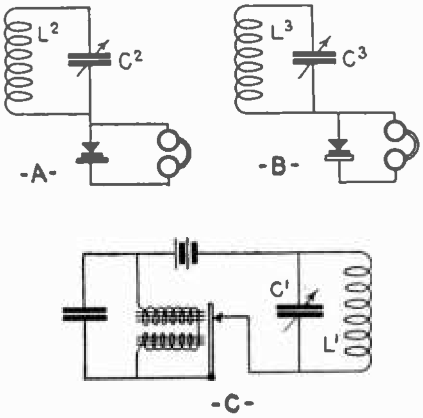

Diagram showing three-way method of wavemeter calibration.

WAVE MOTION—A disturbance of the equilibrium of a medium or body, extended or propagated from point to point with a continuous motion, each particle vibrating only on each side of its position of equilibrium, while each phase of vibration moves onward. Examples of wave motion are the propagation of electromagnetic waves, or the waves on the surface of a body of water.

WAVES, CONTINUOUS—See *Continuous Waves.*

WAVES, CONTINUOUS, KEY MODULATED—See *Continuous Waves, Key Modulated.*

WAVES, CONTINUOUS, MODULATED AT AUDIO FREQUENCY—See *Continuous Waves at Audio Frequency.*

WAVES, ELECTROMAGNETIC—Periodic electromagnetic disturbances progressive through space. (See *Wave.*)

WAVES, SUSTAINED—See *Sustained Waves*, also *Continuous Waves.*

WAVE TRAIN—A series of waves. An example of a wave train is the series of disturbances set up by each spark in a spark system of radio telegraphy.

WAVE TRAP—A device consisting of a variable condenser and coil, usually connected in the antenna circuit of a receiving set, to eliminate interference from undesired stations. (See *Filter, Filter Condenser, Acceptor*, also *Rejector.*)

WAVE-WOUND—A class of drum armature in which connections produce a "stepping forward" in a zig-zag wavy line all the time. Also called *Series* and *Two Circuit Windings.*

WEBER—The *Practical Unit* (q.v.) of magnetic flux. The engineers' unit is the "line" which is also the electromagnetic unit. 1 Weber is equal to 108 electromagnetic units (or "lines"). The Weber is produced by a current of one ampere flowing through a circuit having one henry of inductance.

WEHNELT BREAK—An *electrolytic interrupter* (q.v.).

WEIGHT—The force with which the earth attracts a body. A body varies in weight according to its distance from the centre of the earth, although its mass remains constant.

WESTON CELL—The present standard cell. It consists of mercury with a paste of mercurous and cadmium sulphates which form the cathode. The Weston cell has an anode of 12.5 per cent cadmium amalgan in an electrolyte of saturated solution of cadmium sulphate. It has a constant electromotive force of 1.0125 volts at 20 degrees Centigrade. (See *Electromotive Force*, also *Cadmium.*)

WHEATSTONE BRIDGE—An instrument for measuring resistance in which the current from a battery divides into two parallel circuits, each

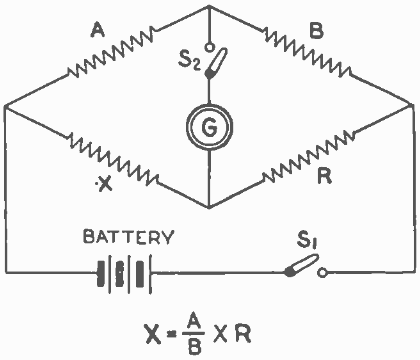

Diagram illustrating principle of the Wheatstone Bridge.

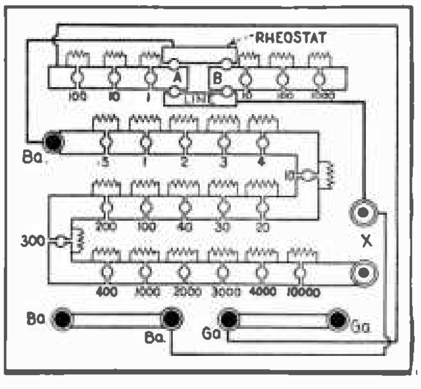

Post-Office type of Wheatstone Bridge.

divided into two "arms" capable of being adjusted so that there is no difference of potential between the dividing points on both sides. The point of zero difference of potential is indicated by the absence of deflection of a galvanometer placed across the dividing points. The ratio of the resistance of the two parts of the arms is then the same, so that if the resistance of three of the arms is known, that of the fourth arm (i.e. the resistance to be measured) can be determined.

WHEATSTONE TRANSMITTER—An apparatus for delivering telegraphic currents to a line at high speed. Control is effected by aid of a moving perforated paper tape, prepared according to a code.

WHIPPING—The binding of string or small wire round the end of a rope or multiple wire to prevent the ends from fraying out.

WIMSHURST—A type of Induction or Influence Machine consisting, in a simple form, of two discs of insulating material revolving in opposite directions. These carry a number of equal sectors which form combined inductors and carriers. There are usually two collectors and two pair of brushes. (See *Induction Machine.*)

WINDOW LEAD-IN—A form of insulator for passing an aerial *lead-in* through a window. In some cases porcelain tubes are used as window lead-in's. A common form of window lead-in consists of a flat copper conductor covered with woven cotton insulation.

WIRED WIRELESS—A method of transmitting radio messages, in which the waves are guided by wires instead of radiating freely through the ether. (See *Line Radio.*)

WIRE GAUGE—A device for measuring the diameter of a round wire in accordance with a predetermined scale. The gauge commonly used in this country is the Brown and Sharpe (B. & S.) or American gauge.

WIRELESS TELEGRAPHY—Also known as **RADIO TELEGRAPHY** (q.v.).—A system of telegraphy, utilizing electromagnetic waves set up by oscillating currents as the means of transmission. The waves generated at the transmitting station spread out from the *aerial* (q.v.) in all directions. They are intercepted by the receiving aerial, detected by a rectifying device such as a crystal or vacuum tube, amplified and delivered to a head set or loud speaker. A key is used at the transmitter to control the duration of the waves so that they will correspond to the dots and dashes of the *continental code* and these dots and dashes are reproduced in the phones at the receiving station.

WOLLASTON WIRE—Platinum wire, drawn extremely fine by the process of coating fine platinum wire with ductile material, drawing the whole down together, and (if required) removing the coating by dissolving it off.

WORK—The action of a force upon a body to overcome resistance. Work is measured by the product of the force exerted and the distance moved. Work may be measured by *ergs* (q.v.), *foot-pounds, horsepower*, etc.

WOOD'S METAL—A soft metallic alloy having a low melting point. This alloy consists of two parts of lead, one part of tin, four parts of bismuth, and one part of cadmium, all by weight. This metal has been used extensively for setting detector crystals in their cups, so as to make a good electrical contact.

X, Y, Z

X—Symbol for *Reactance* (q.v.).

"X" RAYS—Also called *Roentgen Rays.* Electric waves of much higher frequency than light, produced by the striking of *cathode rays* upon a solid substance. These rays are not deflected by electric or magnetic fields and possess the property of penetrating solid substances which could not be penetrated by ordinary light rays. The use of the "X" Rays permit photographs to be taken of bones or other parts of the human body which could not be seen in any other way.

X'S—Disturbances of an erratic nature heard from time to time in the 'phones or loud speaker of a radio receiving set, due to storms, electric disturbances, etc. (See *Atmospherics, Natural Electric Waves, Static,* also *Strays.*)

YAGI SPARK GAP—A form of spark gap originated by H. Yagi. It is a quenched spark gap, the electrodes of which are aluminum and brass. The gap functions in an atmosphere of coal gas.

"Y" GROUPING—See *Star Grouping.*

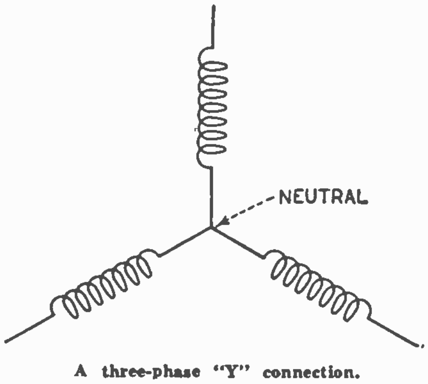

A three-phase "Y" connection.

YOKE—A piece of soft iron used in certain forms of electromagnets to yoke two parallel cores together magnetically. The cores of the ordinary electric bell are fastened together at one end by means of a yoke.

YOKOJAMA, EITARO—Japanese radio expert. He was born in 1883 and was educated at the Engineering College of the Tokyo Imperial University. While at college, he specialized in radio. He was appointed to the Electro-technical Laboratory of the Japanese Ministry of Communications, to carry out research work in wireless telegraphy and telephony. He was one of the inventors of the T.Y.K. oscillation gaps for radio telephony, for which he received many distinctions. In 1910 he was appointed head of the Radio Section. Yokojama, who is one of the most brilliant Japanese radio experts, is a member of the Institute of Radio Engineers and of many other scientific societies.

ZEEMAN EFFECT—Doubling of the spectrum lines of light sources when placed in a strong magnetic field.

ZENNECK, J.—German wireless expert. He was born April 15th, 1871, at Wurtemberg and was educated at Tuebingen. In 1895 he was appointed assistant in the Physical Institute in Strassburg, a post he held until 1899, when he carried out a series of tests in radio telegraphy in the North Sea. In 1905 he was appointed assistant professor of physics at the Institute of Technology, Brunswick and was appointed professor at Munich in 1913. Professor Zenneck has written a number of authoritative books on wireless and also a large number of articles on electro-magnetic oscillations.

ZERO BEAT RECEPTION—The detection of continuous modulated waves using a local source of high frequency current having a frequency equal to that of the incoming wave. When this continuous voltage is impressed on the incoming modulated voltage, (both of the same frequency) the output of the detector contains a current of audio frequency similar in character to the modulated current at the transmitter.

ZERO METHOD—A method of measurement in which various adjustments are made until the current flowing through a galvanometer is reduced to zero, as in a *Wheatstone Bridge* (q.v.). (See *Null Method.*)

ZINC—A metallic element. Its chemical symbol is Zn. In color, zinc is bluish-gray. It is practically non-corrosive in the atmosphere, is capable of taking a high polish, is unaffected by water, but is soluble in nitric acid and in soda and potash solutions. Pure zinc is attacked very slowly by sulphuric acid, but this feature is one of the greatest in the application of zinc in radio work and in electrical work generally.

Zinc is one of the most important components of most dry cells. It forms the negative terminal in most "B" batteries. Zinc in rod form is used in most forms of wet cells.

ZINCITE—An oxide of zinc. Zincite crystals can be distinguished by their red color, often broken up by orange-yellow streaks. This crystal, in combination with several other crystals makes a very excellent detector for radio purposes. The *perikon detector* (q.v.) is a combination of zincite and chalcopyrites. Zincite is also used in combination with bornite, galena, copper, iron pyrites, tellurium and silicon.

ZIRCONIUM—One of the metallic elements. Its chemical symbol is Zr, and its atomic weight is 90.6. Zirconium is an iron-gray powder in one form, or it may be made to crystallize. The crystals look like antimony, are very brittle and extremely hard, being capable of scratching glass and rubies. Zirconium resembles thorium in many of its chemical properties. For the control of the vacuum in high vacuum tubes, a small quantity of thorium or zinconium is included in the tube. These metals combine with hydrogen, oxygen, nitrogen, etc., to form compounds of very low vapor pressure.

SUPPLEMENT

CLASSIFIED INDEX TO S. GERNSBACK'S RADIO ENCYCLOPEDIA

In the following index various co-related subjects have been gathered together for the convenience of the reader or student, under appropriate headings. Thus, under the heading of "Circuits," every circuit contained in the Encyclopedia is listed. Other important subjects are similarly treated. Where the reader is interested in but a single item such as, for example, "Bolitho Circuit," this can be found more conveniently by referring to the Encyclopedia itself, which, of course, is arranged in alphabetical order. All separate and unrelated headings or subjects should be looked up in this manner, that is by referring directly to the Encyclopedia, as these are not listed in the index.

A

AERIAL, (6) Page
Balancing 21
Circuit 9
Fan 65
Frame 71
Height of 79
Horizontal 83
Insulation of 91
Insulator 92
Inverted "L" 94
Length of 100
Light Line 101
Plain 125
Quiescent 129
Receiving 131
Switch 9
Transmitting 154
Tuning Condenser 9
Wire 10

ALTERNATING CURRENT, (11)
Theory of Production 12
Undamped 156

ALTERNATOR, (12)
Alexanderson 11
Definition 12
Goldschmidt 74
High Frequency 81
Inductor 90
Laminated Core 99
Polyphase 126
Single Phase 139
Synchronous Motor 148

AMPLIFICATION
Audio Frequency 19
Cascade 32
Constant 14
Factor 14
Impedance Coupled 86
Radio Frequency 15
Resistance Coupled 134
Voltage 14

AMPLIFIER, (14)
Action of Vacuum Tube 15
Audio 19
Intermediate Frequency 15, 93
Radio Frequency 15
Resistance Coupled 134

ANTENNA, (6)
Aeroplane 10
Cage 29
Capacity Measurement 31
Gamma type 73

Page
Inductance 80
Non Radiative 118
Resistance 16
Tuned 155
Tuning Condenser 138
Umbrella 156
Wave 159
Wave Length Calculation for 160

ARC
Characteristic 16
Duddell Singing 56
Generator 16
Generator System 16
Magnetic Blowout for 108
Poulsen 126

B

BATTERIES, (22)
"A" 3
Accumulator 136
"B" 22, 125
"C" 32
"C" Curve 32
Cell 33
Charger, Balkite 21
Chloride 35
Daniell's Cell 50
Electrolytic Cell 60
Filament 67
Float 70
Galvanic 72
Gravity 75
Grid 75
High Potential 81
High Tension 82
Of Alternators 22
Parallel Connection 3
Plate 125
Primary Cell 127
Secondary Cell 33, 136
Series Connection 3
Series-Parallel Connection 3
Storage 3, 136
Testing Instruments 22
Theory of Primary Cell 34
Trigger 155

BIOGRAPHIES
Alexanderson, Ernst F. W. 10
Appleton, Edward Victor 16
Arco, Graf George Von. 17
Armstrong, Edw. H. 17
Austin, Louis Winslo 19
Belin, Edward 23
Bell, Alexander Graham 23

Page
Bellini, Dr. Ettore 23
Blondel, Andree 23
Blondlot, Professor Prosper Reno 23
Branly, Edward 24
Braun, Ferdinand 25
Brown, Sidney George 27
Carpentier, Jules 31
Clerk-Maxwell, James 36
Cohen, Louis 39
Crookes, Sir William 45
De Forest 51
Dubilier, William 56
Edison, Thomas Alva 58
Faraday, Michael 65
Fessenden, Reginald Aubrey 66
Fleming, John Ambrose 68
Gernsback, Hugo 73
Gernsback, Sidney 73
Goldschmidt, Rudolph 74
Goldsmith, Dr. Alfred N. 74
Hammond, John Hays, Jr. 78
Heaviside, Dr. Oliver 79
Heising, Raymond A. 79
Henry, Joseph 80
Hertz, Heinrich Rudolph 80
Hogan, John V. L. 82
Hooper, Stanford C. 83
Janet, Paul 95
Kelvin, Lord William Thompson 96
Kimura, Shunkichi 98
Kolster, Frederick A. 98
Koomans, Nicholas 99
Korn, Arthur 99
Langmuir, Irving 99
Latour, Marius 100
Lodge, Sir Oliver 102
Lombardi, Dr. Luigi 105
Marchant, Edgar Walford 110
Marconi, Guglielmo 110
McLachlan, Norman W. 111
Meissner, Alexander 112
Mullard, S. R. 115
Nally, Edward Julian 115
Noble, Sir William 118
Pannill, Charles Jack 122
Phillips, Raymond 124
Pickard, Greenleaf ttier 124
Poulsen, Valdemar 126
Prince, Charles Ed d 127
Pupin, Michael 127
Rayleigh, John W am Strutt 131
Roberts, Joseph rrison Thompson 135
Robinson, Jame 135
Robinson, Sam S. 135
Rutherford, S Ernest 136
Squier, Majo eneral George Owen 143
Tesla, Nikol 151
Yokojama, aro 161
Zenneck, 161

C

CAPACITY, (30) Page
Calculation 29
Amt. of 34
Circuit 30
Condenser 40
Condenser Curves 40
Condensers in Parallel 30, 41
Condensers in Series 31, 41
Conversion Factor 30
Definition 30
Distributed 55
Earth 30
Electrostatic 63
Farad 65
Frequency Factor 30
Ground 30
Measurement of Antenna 31
Mutual 115
Resistance 31
Switch 31

CIRCUITS
"A" Battery 3
Armstrong Circuit (Original) 18
Armstrong 3-Circuit Regenerative Receiver 18
Armstrong Super-Regenerative 147
Audio Amplifier 19
Audio Frequency 71
Autoplex 19
Balanced 20
Balanced Metallic 20
"B" Battery 22
Bolitho 24
Breaker 36
Broken 36
Buzzer 29
Buzzer Exciter 29
Cascade N. F. and detector 32
Closed Circuit 37
Cockaday 37
Cockaday, with push-pull amplifier 37
Coupled circuits 44
Crystal 31, 46
Definition 36
Detector 52
Direct Coupled 54
Fourth 71
Heising 79
Heterodyne 79, 80
Impedance Coupled 85
Interference Eliminator 93
Inverse Duplex 94
Neutrodyne 117
Open 36
Parallel 36
Receiving 131
Reflex 132
Regenerative 65
Reinartz 133
Resistance Coupled 134
Series 3
Series Parallel 3
Short 138
Single Stage Radio Frequency amplifier 15
Super-Heterodyne 145, 146
Super-Regenerative, Armstrong 147
Three-Wire 117
Tuned Radio Frequency 155
Two-stage Audio Amplifier 19
T. Y. K. Arc 156
Ultradyne 156

Page
Vacuum Tube, Non-Regenerative 52
Vacuum Tube, Regenerative 52

COHERER, (39)
Autocoherer 19
Branley 24
Definition 39
Granular 75
Lodge Disc Detector 108
Marconi Filings 110
Slaby-Arco-Vacuum 141

CONDENSER, (40)—See also Capacity
Aerial Tuning 9, 138
Air 10
Balanced Rotary 21
Billi 23
By-Pass 40
Compressed Air 39
Curves 40
Definition 40
Electrolytic 60
Filter 68
Fixed 41
Grid 41, 76
In Parallel 41
In Series 41
Korda Air 99
Leyden Jar 101
Mica 41
Moscicki 114
Paper 122
Phone 41
Rotary 136
Shortening 138
Stopping 144
Telephone 149
Variable 41

COILS (39)
Antenna 39
Astatic 18
Basket 21
Choke 35
Cockaday 38
"D" 50
Damping 50
Definition 39
Duolateral 56
Exploring 65
Feed-Back 65
Fixed 68
Honey-Comb 82
Impedance 85
Inductance 87
Induction 88
Jigger 95
Loop Antenna 105
Lorenz 106
Low Loss 107
Movable 115
Pancake 122
Plug-In 125
Primary Induction 89
Radio Frequency 130
Reactance 131
Reaction 131
Retardation 135
Repeating 133
Spark 142
Spider Web 143
Tesla 151
Tickler 153
Tuning 155

Page
Vario Coupler 44, 158
Variometer 158

COUPLING, (44)
Coefficient 44
Close 37
Conductive 41
Coupled Circuits 44
Degree of 45
Definition 44
Direct 54
Electrostatic 63
Feed Back 65
Inductive 90
Regenerative 133
Transformer 45, 153

CRYSTALS
Bornite 24
Carborundum 31
Cat-Whisker for 32
Cerusite 34
Chalcopyrite 34
Definition 45
Detector 45
Galena 72
Iron Pyrites 94
Molybdenite 114
Nagyagite 115
Quartz Oscillator 129
Receiver 46
Rectifier 46
Silicon 138
Theory of Operation 46
Zincite 161

CURRENT, (46)
Alternating 11, 47
Anode 16
Assumed Direction of Flow 18
Audio frequency 19
Carrying Capacity of Wire 47
C. G. S. Unit of 34
Constant 42
Consumption 42
Continuous Direct 42
Convection 42
Critical 45
Decay of 50
Definition 46
Density 47
Direct 47
Direction of Flow (Assumed) 47
Displacement 54
Eddy 58
Electron Theory 62
Emission 64
Flow of 70
Fluctuating 70
Foucault 70
Free Alternating 71
Full Load 72
Grid 76
Ground 77
Harmonic 78
High Frequency 48, 81
High Tension 48
Idle 85
Intermittent 98
Leading 100
Low Frequency 48, 107
Modulated 114
Nature of Electric 49

Page
Oscillating 120
Oscillatory 120
Overtone 122
Primary 48
Production of Electric 48
Pulsating 127
Pulsating Direct 48
Radio Frequency 130
Reactive 131
Rectified 131
Rotary 136
Secondary 48, 137
Space 142
Stray 145
Theory of Alternating 11
Theory of Electric 49
Theory of Flow 151
Undamped Alternating 56

D

DETECTOR, (52)
Balanced 20
Barreter 21
Carborundum 31
Circuit 52
Coherer 39
Combination 39
Crystal 45
Crystal, Theory of Operation 46
Cymoscope 49
Damped Wave 52
Disc, Lodge 103
Electrolytic 60
Hozier-Brown 84
Liquid Barreter 102
Lodge Disc 103
Marconi Filing Coherer 110
Perikon 123
Silicon 138
Theory of Detector Action 151
Thermal 152
Ticker 153
Tikker 153
Tone Wheel 153
Vacuum Tube 52, 157, 158
Valve 158

DIELECTRIC, (53)
Absorption 53
Coefficient and Constant 53
Frequency 54
Constants, Table of 54
Hysteresis 54
Strength 54

DISCHARGE
Brush 27
Corona 43
Oscillatory 120
Oudin Resonator for 121
Spark 142

E

ELECTRIC
Absorption 59
Adhesion 6

Page
Displacement 59
Elasticity 59
Field 59, 66
Induction 59
Electricity 59
Osmosis 121
Potential 126

ELECTROLYTIC
Action 60
Battery Charger 60
Cell 60
Condenser 60
Conduction 60
Decomposition 51
Detector 60
Interrupter 61, 94
Rectifier 61
Rectifier, Full Wave 61

ELECTROMAGNETIC
Attraction 62
Control 61
Field 61, 66
Flux 61
Impulse 61
Induction 61, 89
Inertia 90
Instruments 61
Microphone 61
Radiation 62
Repulsion 62
Switch 62
Theory of Light 62
Units 62
Waves 62, 160

ELECTROMOTIVE FORCE, (62)
Active Pressure 6
Back 20
Breakdown Potential 25
C. E. M. F. 34
Counter 44
Current, Resistance Relationships (Ohms Law) 119
Definition 62
Induced 86
Kirchoff's Laws 98
Potential, Electric 126
Pressure, Electrical 127
Volt 159
Voltage 159
Voltage, Constant 159
Voltage Drop 159

ELECTRON (62)
Flow 62
Emission 62
Relay 63
Theory 62
Tube 63

ELECTROSTATIC, (63)
Capacity 63
Charge 63
Coupling 63
Field 63, 66
Flux 63, 70
Flux Density 63
Force 63
Induction 63, 89
Leakage 64
Line of Force 64
Strain 64

Page
Systems 64
Units 64
Voltmeter 64
Wattmeter 64

F

FEED BACK, (65)
Armstrong's Circuits 17
Coil 65
Coupling 65
Definition 65
Effect 65
Regeneration 133
Regenerative Circuit 133
Reinartz Circuit 133
Tickler 153

FIELD, (66)
Density 66
Electric 66
Electromagnetic 66
Electrostatic 66
Magnet 66
Magnetic 66
Regulator 66

FILAMENT, (66)
Battery 67
Consumption 67
Current 67
Formula for Electron Flow 67
Rheostat 67
Resistance 67
Thoriated 67

FLUX, (70)
Density 70
Distribution 70
Electrostatic 70
Leakage 70
Lines 70
Magnetic 70
Magnetism 70

FREQUENCY, (71)
Alternator of 71
Audio 19, 71
Band 71
Carrier 71
Changer 71
Circuit, Audio 71
Circuit, Radio 71
Cycle 49
Definition 71
Fundamental 71
Group 71, 77
High 71
Low 71
Meter 71
Periodicy 123
Radio 72
Unit of 72
Variation of Power Factor with 72

G

GALVANOMETER, (73)
Constant 73
D'Arsonval 73
Mirror 73, 114

GRID, (75) Page
- Battery 75
- Bias 76
- Circuit 76
- Condenser 41, 76
- Control 76
- Current 76
- Leak 6, 76
- Potential, Modulation of 76
- Potentiometer 77

GROUND, (77)
- Clamp 77
- Connection 77
- Currents 77
- Detector 77
- Switch 77
- Wire 77

H

HETERODYNE, (80)
- Beats 22
- Dissonance 55
- Driver 56
- Local Oscillations 102
- Receiver 80
- Self- 137
- Super-Heterodyne Receiver 145
- Wave Meter 81

HIGH FREQUENCY, (71, 81)
- Alternator 81
- Amplification 81
- Buzzer 81
- Choke 35
- Current 48, 81
- Component of Plate 81
- Oscillation 81
- Resistance 81, 135

I

IMPEDANCE, (85) Page
- Coil 85
- Coupled Amplifier 86
- Factor 86
- Matched 86
- Tuned 155

INDUCTANCE, (86)
- Antenna 86
- Coils 87
- Distributed 87
- Loading 87
- Measurement of 87
- Mutual 87
- Parallel 88
- Reactance 87
- Receiving 87
- Self 87
- Series 88
- Tapped 87
- Units of 88
- Variable 88

INDUCTANCE COIL
- Basket Wound 21
- Choke 35
- Definition 87
- Duolateral 56
- Honey-Comb 82
- Lorenz 106
- Spider Web 143
- Tuning 155
- Vario Coupler 44, 158
- Variometer 158

INDUCTION, (88)
- Bridge 88
- Coil 88
- Coil, Primary 89
- Coil, Secondary 89
- Density 89
- Electromagnetic 89
- Electrostatic 89
- Instruments 89
- Machine 89
- Magnetic 89
- Motor 89
- Motor, Squirrel Cage 144
- Mutual 89
- Reactance 90
- Regulator 90
- Screen 90
- Unit 90
- Wattmeter 90

INDUCTIVE
- Capacity 90
- Circuit 90
- Connection 90
- Coupler 90
- Coupling 90
- Drop in Voltage 90
- E. M. F. 90
- Reactance 90
- Rise 90
- Resistance 90

INSULATING MATERIALS, (91)
- Bakelite 20
- Compound 91
- Definition 91
- Dielectric Constants 5
- Ebonite 57, 159
- Eburin 57
- Electrose 63
- Impregnated 86
- List of 149
- Mica, ... 113
- Micabond Cloth 113
- Micabond Paper 113
- Micabond Plate 113
- Micanite 113
- Property 91
- Table of Dielectric Strength 149
- Vulcanite 159
- Vulcanized Rubber 159

K

KEY, (96) Page
- Anti-Capacity 96
- Definition 96
- High-Speed 96
- Magnetic 109
- Modulated Continuous Waves 42
- Switch 97
- Transmitting 154

M

MAGNET, (108) Page
- Artificial 18
- Compound 40
- Definition 108
- Electro- 61
- Horseshoe 83
- Natural 116
- Negative Pole 116
- North Pole 118
- Permanent 123
- South Pole 142
- Temporary 151
- Wire 108

MAGNETIC
- Attraction and Repulsion 108
- Blowout 108
- Circuit 108
- Field 66, 108
- Flow 109
- Flux 70, 109
- Flux Density 109
- Force 109
- Force Due to Current 109
- Hysteresis 109
- Induction 109
- Inductive Capacity 109
- Interrupter 109
- Key 109
- Leakage 109
- Needle 109
- Poles 109
- Properties 109
- Storm 109

O

OSCILLATIONS
- Back 20
- Constant 120
- Damped 49
- Definition 120
- Forced 70
- Free 71
- Frequency 120
- High Frequency 81
- Impact Excitation 85
- Local 102
- Logarithmic Decrement 105
- Master Oscillator System 111
- Oscillion 121
- Oscillogram 121
- Oscillograph 121
- Point 120
- Poulsen Arc 126
- Resonator 135
- Tesla Coil 151
- Transformer 45, 120
- Undamped 156
- Valve 120

OSCILLATOR, (120)
- Hertzian 120
- Linear 120
- Magnetic Blowout for 108
- Master 111
- Negatron 117

Page

- Righi ... 120
- Quartz ... 129
- Thermionic ... 120

P

PLATE, (125) Page

- Battery ... 125
- Circuit ... 125
- Component, High Frequency ... 125
- Consumption ... 42, 125
- Control ... 125
- Current ... 125
- Impedance ... 125
- Voltage ... 125

R

RADIATION, (130)

- Definition ... 130
- Efficiency ... 130
- Electromagnetic ... 62
- From Antenna ... 130
- Gliding Through ... 74
- Heaviside Layer ... 79
- Resistance ... 130, 135
- Theory of Propagation of Electromagnetic Waves ... 152
- Theory of Reception of Electromagnetic Waves ... 152

RADIO, (130)

- Activity ... 130
- Channel ... 130
- Communication ... 130
- Compass ... 130
- Control ... 130
- Frequency ... 72
- Telegraphy ... 130, 161
- Telephony ... 130
- Transmission ... 130

RADIO FREQUENCY, (72)

- Ammeter ... 14
- Amplifier ... 15
- Circuit ... 71
- Coils ... 130
- Current ... 130
- Resistance ... 135
- Selectivity ... 130
- Signals ... 130

REACTANCE, (131)

- Capacitative ... 30
- Coil ... 131
- Inductance ... 87

RECTIFIER, (131)

- Electrolytic ... 61
- Full Wave ... 131
- Half Wave ... 78
- Potential ... 126

RESISTANCE, (134)

- Antenna ... 16
- Box ... 134
- Contact ... 42
- Coupled Amplifier ... 134

Page

- Coupling ... 134
- Definition ... 134
- Field Regulator ... 66
- Graphite ... 75
- High Frequency ... 81, 135
- Litzendraht ... 102
- Of Ground Connection ... 135
- Ohm ... 118
- Ohmic Drop ... 119
- Ohmic Loss ... 119
- Ohm's Law ... 119
- Radiation ... 135
- Radio Frequency ... 135
- Variable ... 158

RESONANCE, (135)

- Curve ... 135
- Intensifies ... 135
- Parallel ... 123
- Series ... 137
- Transformer ... 135

S

SWITCH, (147) Page

- Aerial ... 9
- Board ... 147
- Capacity ... 31
- Dead End ... 50
- Double Throw ... 56
- Electromagnetic ... 62
- Gear ... 148
- Ground ... 77
- Jack ... 95
- Key ... 96
- Key, Anti-Capacity ... 96
- Knife ... 98
- Lightning ... 101
- Plug ... 125
- Points ... 42
- Quick Break ... 129

T

THEORY OF CURRENT FLOW, (151)

- Detector Action (of Vacuum Tube) ... 151
- Electric Current ... 49
- Magnetism ... 14
- Operation of Crystal Detector ... 46
- Operation of Vacuum Tubes ... 151
- Primary Cell ... 34
- Production of Alternating Current ... 12
- Propagation of Electromagnetic Waves ... 152
- Reception of Electromagnetic Waves ... 152
- Vacuum Tube Amplifier ... 157

TRANSFORMER, (153)

- Air Core ... 10
- Auto ... 20
- Closed Core ... 37
- Coupling ... 45, 153
- Hedgehog ... 79
- Intermediate ... 93
- Open Core ... 120
- Oscillation ... 45, 120
- Output ... 122
- Plug-In ... 125

Page

- Resonance ... 135
- Shielded ... 138
- Steel ... 153
- Tuned Radio Frequency ... 155

TRANSMISSION

- Aerial, Cage ... 8
- Automatic ... 19
- Broadcasting, General Treatise on Methods. 25
- Carrier Wave ... 31
- Chopper ... 31
- Fading ... 65
- Fan Aerial ... 65
- Feed-Voltage Modulation ... 65
- Ferro-Magnetic Modulator ... 66
- Flame Microphone ... 68
- Frequency Carrier ... 71
- Heising Modulation ... 74
- Key ... 96, 154
- Key, High Speed ... 96
- Jamming ... 95
- Jigger ... 95, 154
- Microphone ... 113
- Modulation ... 114
- Of Photographs by Radio ... 154
- Phenomena of Electric Wave Propagation.. 124
- Quartz Oscillator ... 129
- Queenland Gap Transmitter ... 129
- Radio ... 131
- Radio Telephony ... 131
- Radio Telegraphy ... 130
- Remote Control ... 133
- Spark Coil ... 142
- Spark Rate ... 142
- Speech Modulation ... 143
- Synchronous Discharger ... 148
- Telefunken System ... 149
- Television ... 150
- Tesla Coil ... 151
- Theory of Propagation of Electromagnetic Waves ... 152
- T. Y. K. Arc ... 156

TUNING, (155), See also Coil; Condenser

- Broad ... 155
- Coil ... 155
- Condenser, Aerial ... 9
- Inductance Coils ... 87
- Inductance, Variable ... 88
- Interference Eliminator ... 93
- Interference Preventers ... 93
- Rejector Circuit ... 133
- Selectivity ... 137
- Sharp ... 137
- Static Eliminators ... 144
- Syntony ... 148
- Wave Trap ... 160

U

UNITS, (156)

- Ampere ... 14
- Ampere-Hour ... 14
- Ampere, International ... 14
- Ampere Turns ... 14
- Ampere Volt ... 14
- B.A. Ampere ... 20
- Beg Ohm ... 23
- B. O. T. ... 24
- B. T. U. ... 27

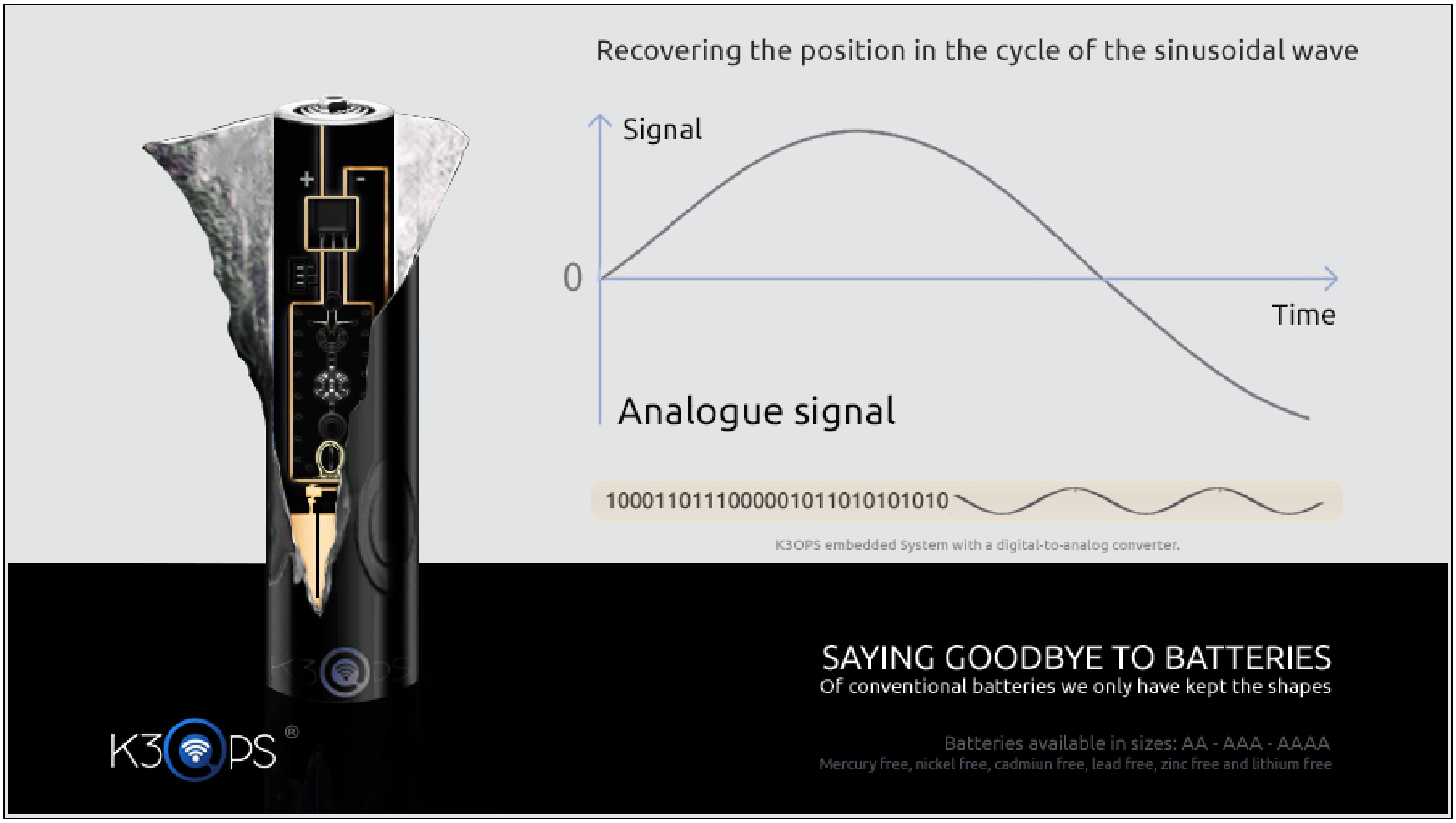
Recovering the position in the cycle of the sinusoidal wave
Signal
0
Time
Analogue signal
1000110111000001011010101010
K3OPS embedded System with a digital-to-analog converter.
SAYING GOODBYE TO BATTERIES
Of conventional batteries we only have kept the shapes
Batteries available in sizes: AA - AAA - AAAA
Mercury free, nickel free, cadmiun free, lead free, zinc free and lithium free
K3OPS®

An electromagnetic field uses the photon as an elementary particle to transmit force. It combines:

- **A magnetic field** *force resulting from the movement of loads µT.*
- **An electric field** *force created by the attraction of repulsion loads, measured in volts per meter - V/m.*

With an energy determined according to the speed of light, the RF are by far the best medium to transmit any kind of information.

The multiplication of wireless communications systems in our environment ensures sufficient microwave leakages to harvest from the ambiant and enough energy to convert into DC electricity. Electromagnetic fields are everywhere and since they carry energy, they became the best candidate to deliver an endless source of renewable energy.

$\vec{B}$ is the magnetic induction expressed in T refered to **Nikola Tesla**, "Father of Free Energy", which is at the origin of the electromagnetism.

Using meta-materials combined with nanotechnology has deeply increased the performance and miniaturization of rectennas embedded in K3OPS system. Our products operate autonomously, offering an endless supply of green energy in a respectful and environment-friendly approach.

principles and beautiful sentences. At that point in time I would have simply been Xin, a child like any other...

But one day I dreamed. I woke up far away, somewhere else, in another past. I decided to change my destiny and even if I was supposed to become a mathematician, I eventually decided to create, because already as a child my heart was chasing the stars.

Of the hundreds of directions shown to me after graduating, only one captured my attention: a single goal... Build the impossible for a safer world. So, over the past 3 years, our real challenge to overcome for all RF Energy Harvesting technics was to optimize electricity conversion. The massive proliferation of wireless telecommunication systems since the past two decades brought a saturation of the electromagnetic fields with a constant growth of 15% every year in our environments. As a result, this situation reversed the base problem that makes today Harvesting RF Energy a game changer. The key was the Power Management System.

We are far beyond the conversion constraint and performance by controlling "RF-interferences", by harvesting different frequencies *from near and far*, by using Metamaterials combined with nanotechnologies. We dramatically have improved power conversion efficiency and reduced the size of our Energy Harvesting systems embedded in all K3OPS' products.

Thanks to Nikola Tesla, my inspiring mentor, K3OPS' products have reached by far their objectives in terms of converting and performance, offering an endless efficient source of green energy, reliable in an environmentally friendly approach.

Xin WEI Co-Founder of K3OPS technology with Alexandre Despallieres

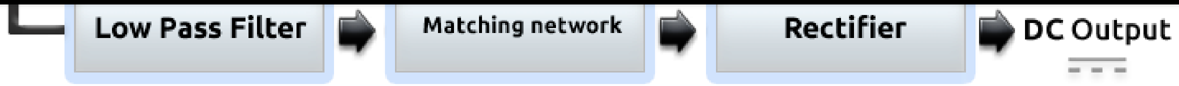
Low Pass Filter
Matching network
Rectifier
DC Output

The Rectenna was invented in 1964 by William C. Brown, patented in 1969. It is a rectifying antenna used to convert microwave energy into DC. A simple Rectenna consists of a dipole antenna with an RF diode connected across the dipole elements. The diode rectifies the AC current induced in the antenna to produce DC power.

RF Energy Harvesting Battery
K3OPS

K3OPS batteries

K3-Case

K3OPS Copyright © 2018 Alexandre Despallieres and Xin Wei - All rights reserved.

Terms of use | Privacy Policy | Press | Login

The Formation of Natural Gypsum Megacrystals in Naica (Mexico)

Supplementary Information deposited into the GSA Data Repository.

Juan Manuel García-Ruiz[1], Roberto Villasuso[2], Carles Ayora3, Angels Canals[4], Fermín Otálora[1]

[1] Instituto Andaluz de Ciencias de la Tierra. Consejo Superior de Investigaciones
Científicas-Universidad de Granada, Campus Fuentenueva, E-18002, Granada, Spain
[2] Compañía Peñoles, Unidad Naica, Naica, Chihuahua, Mexico.
[3] Institut de Ciències de la Terra Jaume Almera, Consejo Superior de Investigaciones Científicas, Lluis Solé Sabarís s/n, E-08028 Barcelona, Spain
[4] Dept. de Cristal·lografia, Mineralogia i Dipòsits Minerals, Universitat de Barcelona, Martí Franquès s/n, E-08028 Barcelona, Spain

1. Supplementary Figures and legends

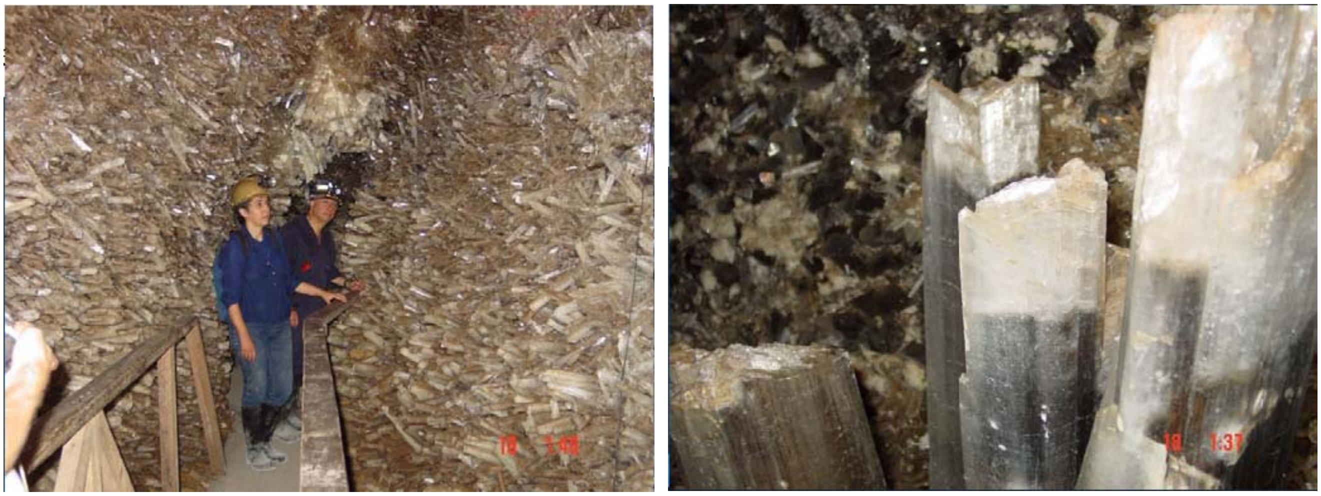

Figure DR1: *The Cueva de las Espadas* (Cave of the Swords) discovered in 1910 is located at level −120 m close to the Montaña fault (see Fig. 1). It is a long corridor (shown on the left) which walls are entirely covered by crystals from centimeter to almost a meter in size (Degoutin, 1912; Foshag, 1927). The cave is close to the natural phreatic level and this explains the light grey and black color of the crystals due to clay as well as iron and manganese oxides solid inclusions (right). These solid inclusions are in most cases aligned on growth fronts, suggesting the occurrence of episodic avalanches of strong oxidation and a growing environment from muddy waters containing colloidal clays.

2. Supplementary methods

2.1. Gypsum Crystal Habit

The habit of gypsum crystals from the Cave of the Crystals has been studied by single crystal X-ray diffraction and by laser reflection goniometry. X-ray diffraction experiments on cleavage plates were performed to get the first inputs on crystal orientation and morphology as well as the unit cell parameters required for the geometrical calculations of inter-facial and inter-edge angles. Our laboratory diffractometer, a Bruker Proteum X8 equipped with a Microstar generator, and a KappaCCD goniometer and detector was used for these X-ray diffraction studies. The unit cell parameters of the crystals investigated were a=5.684Å, b=15.189Å, c=6.282Å, β=114.10º; the face parallel to the main cleavage is, as expected, {0 1 0} and the elongation of the crystals was identified to be [0 0 1].

Interfacial angles (Fig. DR2b) were measured using a straight edge and a protractor for macroscopic forms and a simple light goniometer for microscopic faces (Fig. DR3). These studies indicate that the crystals are prisms bounded by {0 1 0} and {1 k 0} faces and caped by {-1 1 1} faces (Fig. DR2a). The {1 k 0} faces are striated and composed of alternating bundles of {1 2 0}, {1 4 0} and {1 6 0} faces making and overall orientation close to {1 4 0} (Fig. DR4 and table DR1).

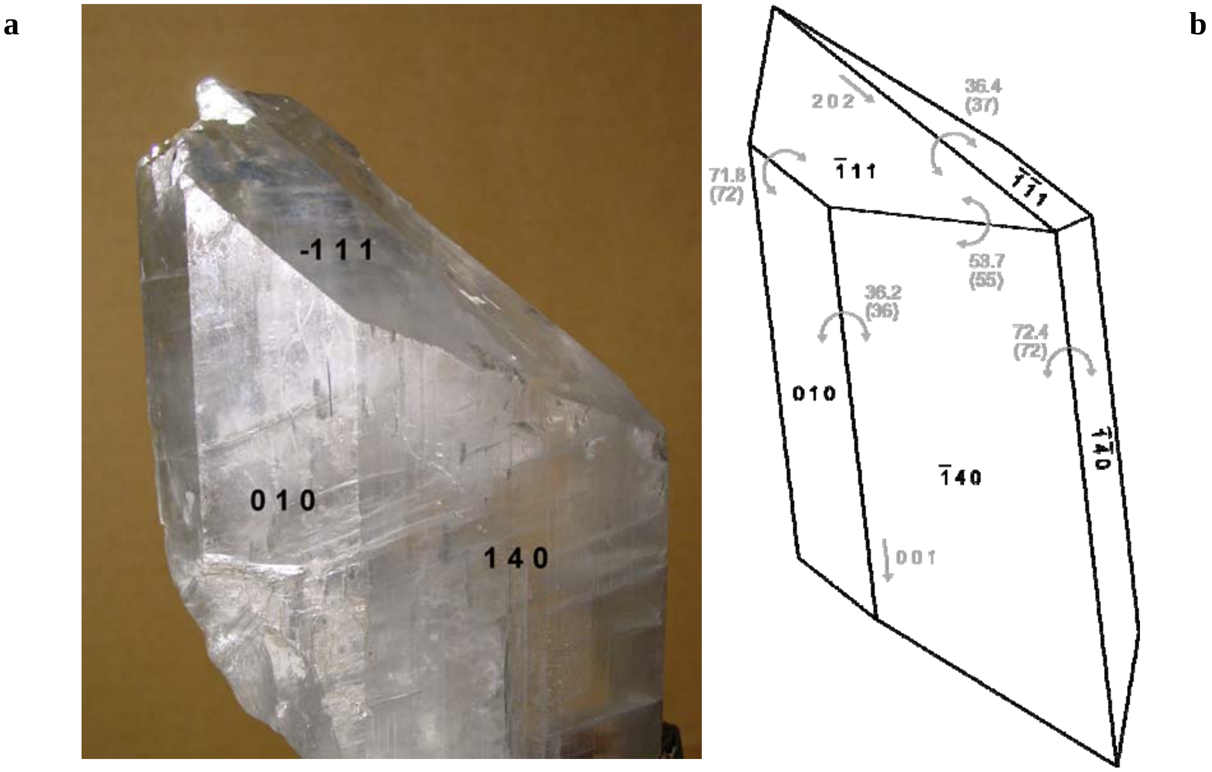

Figure DR2: a) Large blocky gypsum crystal from the *Cueva de los Cristales*. The crystal size in the vertical direction is 12 cm. b) Sketch of the crystal habit showing the indices of the faces, the main zone directions (arrows) and the main interfacial angles (curved double arrows). The numbers next to the double arrows show the values computed from the unit cell followed by the measured angles in parentheses.

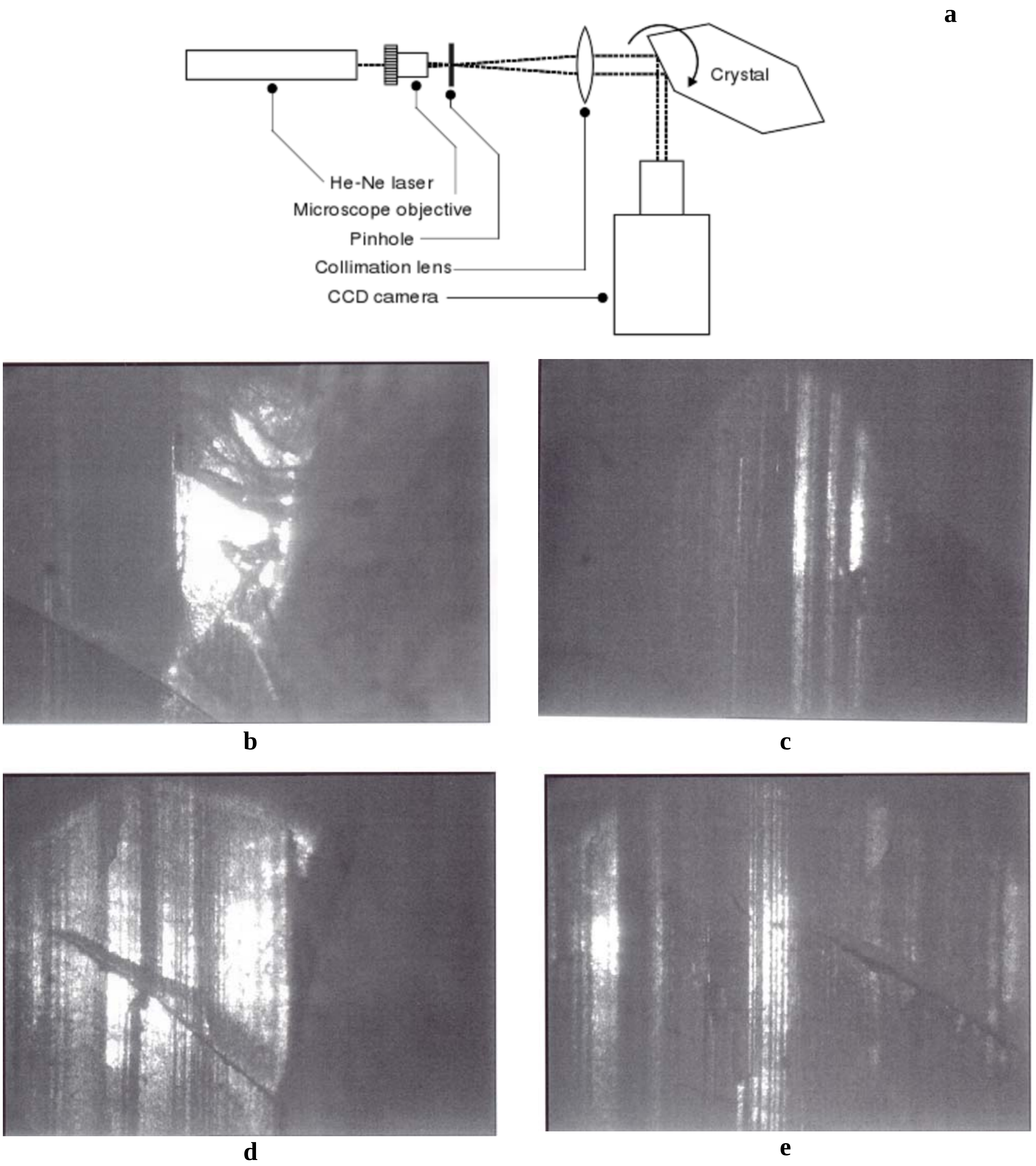

a
Crystal
He-Ne laser
Microscope objective
Pinhole
Collimation lens
CCD camera
b
c
d
e

Figure DR3: a) Experimental setup of the laser goniometer used for the measurement of angles between mirror reflections belonging to the [001] zone. b) Reflection on the (0 1 0) face as seen from the CCD camera. c-e) Mirror reflections from, respectively, (1 2 0), (1 4 0) and (1 6 0). The angles at which these pictures were collected are shown in Table DR1 (third column)

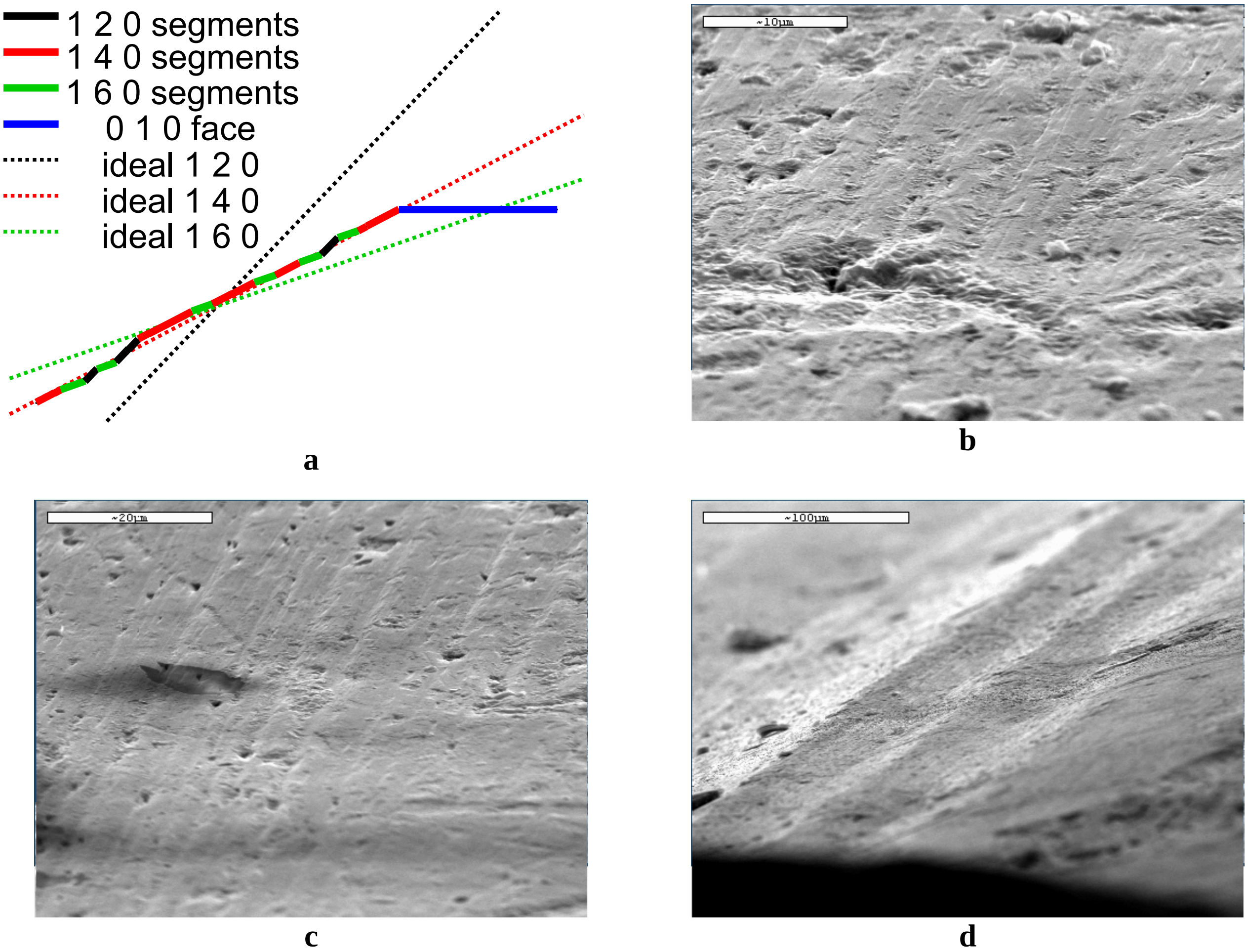
1 2 0 segments
1 4 0 segments
1 6 0 segments
0 1 0 face
ideal 1 2 0
ideal 1 4 0
ideal 1 6 0
a
~10µm
b
~20µm
c
~100µm
d

Figure DR4. Structure of the {1k0} surface. a) Sketch of the face segments distribution and the corresponding angles. b-d) Scanning electron microscope views of the surface showing the variable face segments at three different magnifications..

Table DR1: Calculated and measured angles between 010 and 1k0. The measured angles can be assigned to k = 2, 4, 6

Face	Angle (calculated)	Angle (measured)
010	0.00	0
110	71.14	-
120	55.66	51
130	44.30	-
140	36.20	34
150	30.35	-
160	26.01	25

Face	Angle (calculated)	Angle (measured)
170	22.69	-

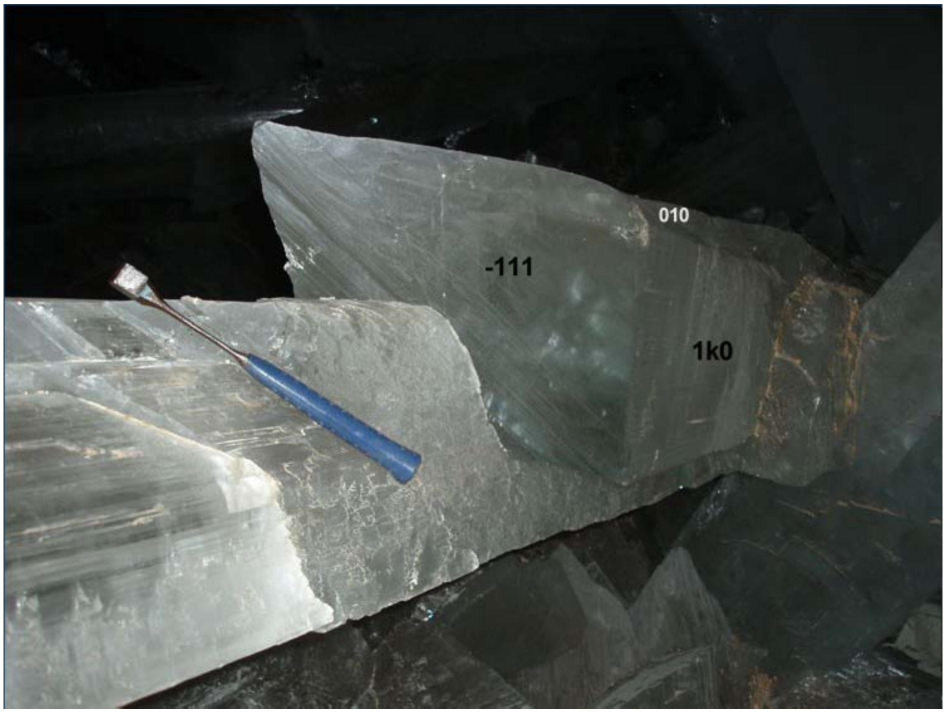

Figure DR5: Top view of a crystal beam showing the terminal faces {-111}: Notice that the large development of the prism {1k0} provokes very small pinacoidal {010} faces.

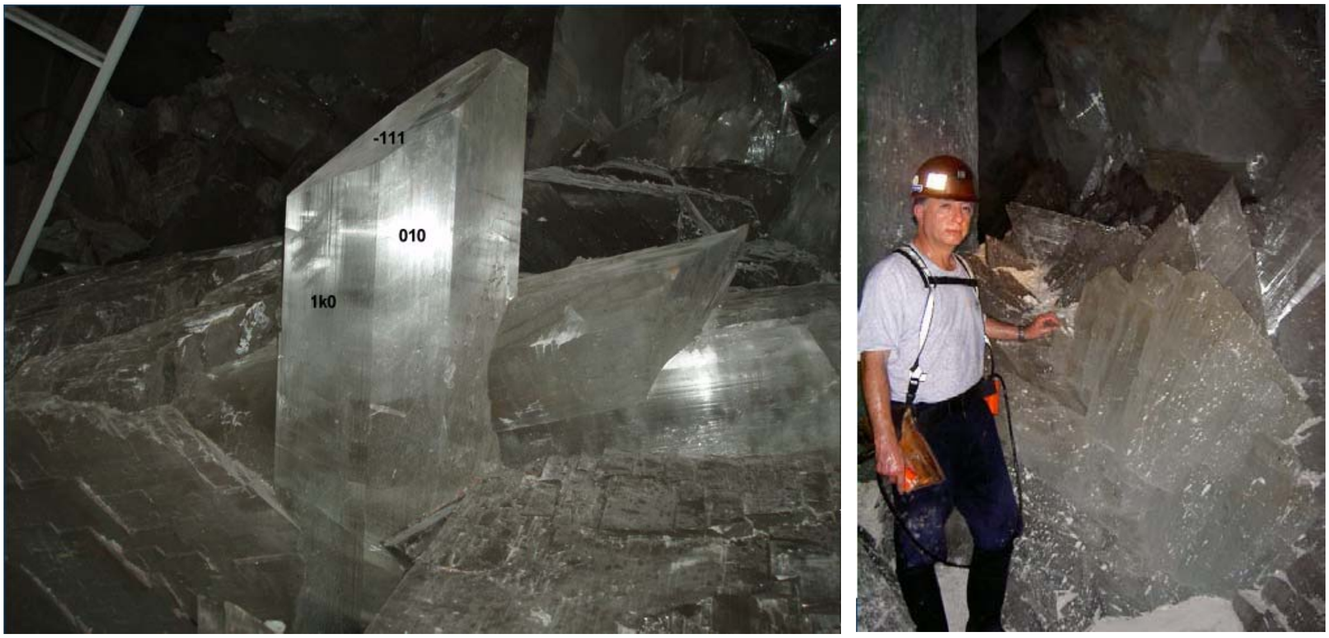

Figure DR6: Two examples of blocky crystals. The unusual blocky habit, the development of the more acute prismatic faces {140} and {160} instead of the usual {120} (Simon and Bienfait, 1965) and the alignment of the crystals with the edge [010] vertical, give them a distinctive sharktooth appearance.

2.2. Fluid inclusions

Fluid inclusion petrography and microthermometry were performed in less than 0,1m long gypsum cleavage fragment from a gypsum sample of the Cave of Crystals, where clear primary fluid inclusions related to growth bands were observed. The large fragment of the crystal from Cave of the Crystals was first sawed using an oil saw. Inclusions sizes range from 40 to 600µm and most of them display a idiomorphic shape. Microthermometry of fluid inclusions was carried out on a Linkam THMS 600 heating and freezing stage. All measurements were made using a 10× objective lens. Calibration of the stage was checked using pure CO_2 and H_2O inclusions. Although heating rates seem not to have a crucial role for regular fluid inclusions microthermometric results, as we work with large inclusions and thick gypsum slices we checked different heating rates and fragment sizes. We concluded that in our case the heating rate must be as slow as 0.1 ºC/minute near the final melting temperature, and the slices must be thinner than 1mm.

Primary fluid inclusions from gypsum growth bands were in most cases monophasic (liquid phase) at room temperature (Figure DR7). Less than 5% of them, however, also show a gas bubble. This suggests that inclusions were formed from a liquid phase at higher temperatures, but the gas phase bubble did not usually nucleate because the temperature gap to the room temperature is short. In order to increase the cooling interval, gypsum fragments were located and maintained at 2 ºC for few weeks and as expected, about 15% of inclusions showed a gas bubble after the cooling. Only those inclusions which were biphasic at room temperature were selected for cryogenic studies (33 fluid inclusions). Eutectic temperatures were not clearly observed in any run. Melting ice temperatures show a maximum at -0.1 with a dispersion of the values down to -1.1 ºC (Fig. DR10). This range of temperatures indicates a low salinity conventionally expressed in the $NaCl-H_2O$ system: 0.17 mass % NaCl with values increasing up to 1.9 mass % NaCl (Bodnar, 1993; Bakker, 2003). The dispersion of melting ice temperatures within an inclusion band is of the same order of magnitude of the variations along the whole set of samples and, therefore, no significant salinity variation with crystal growth can be established.

Heating experiments were performed in 31 fluid inclusions which were biphasic at room temperature. To avoid possible damage, the crystal fragments used in freezing experiments were discarded for heating. Homogenization took place to liquid phase, at temperatures showing a normal distribution with a maximum around 54±7 ºC (Table DR2). From the geological setting no correction due to confinement pressure is required, so this temperature range can be considered the actual temperature of crystal growth.

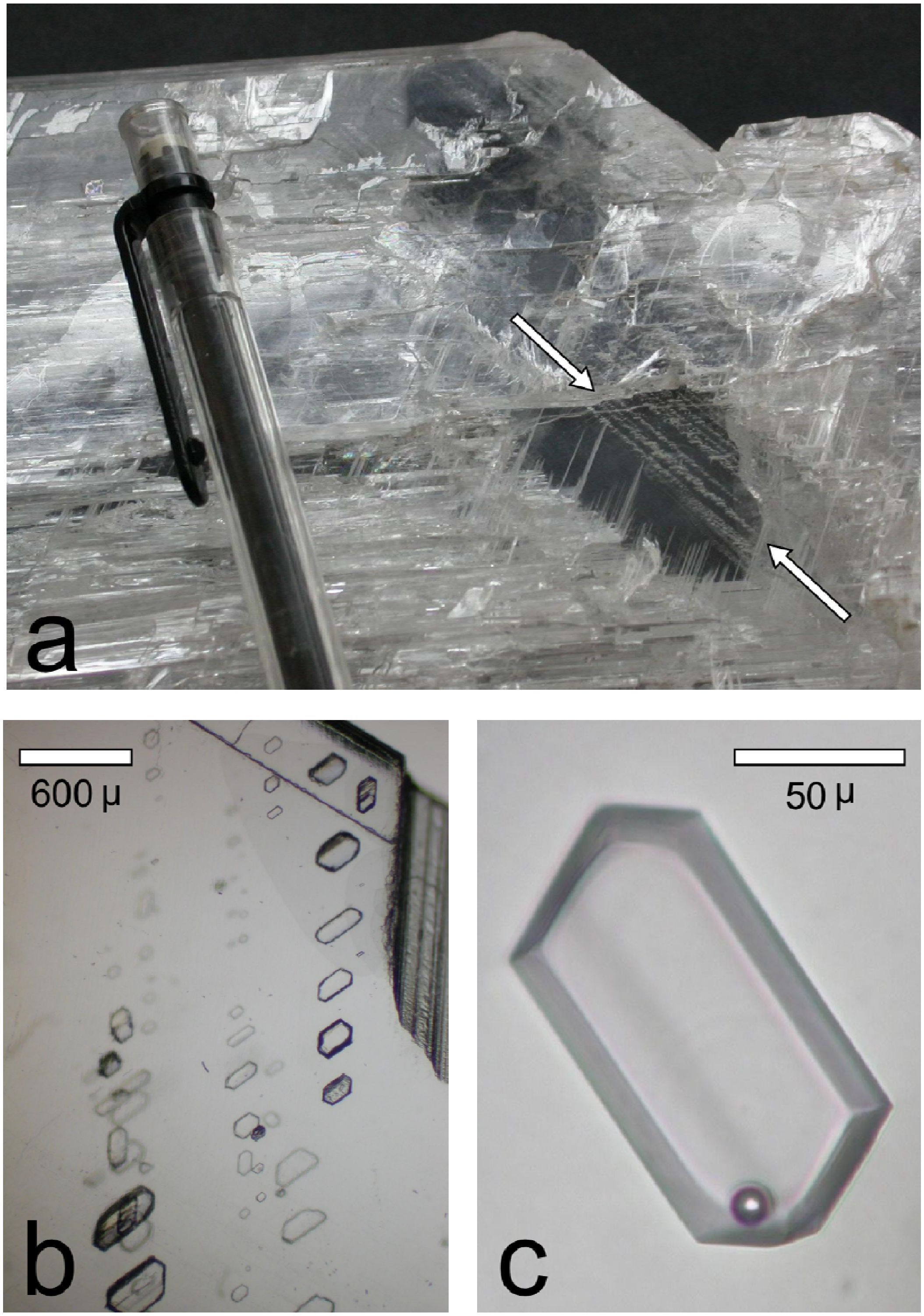

Figure DR7: a) Gypsum crystal form the Cave of Crystals showing growth bands with primary fluid inclusions (arrows). b) A detail of these inclusions along growth fronts. c) Detailed view of one of the biphasic inclusions used in the study.

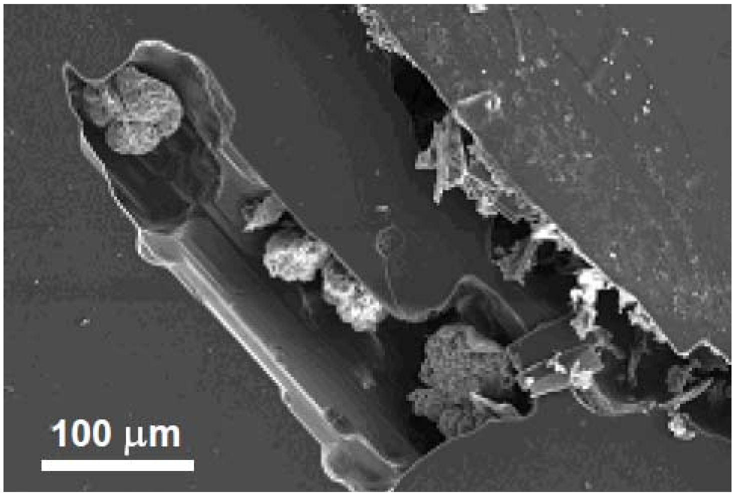

Figure DR8: Spherulitic solid inclusions trapped within fluid inclusions in gypsum crystals from the Cave of Crystals.

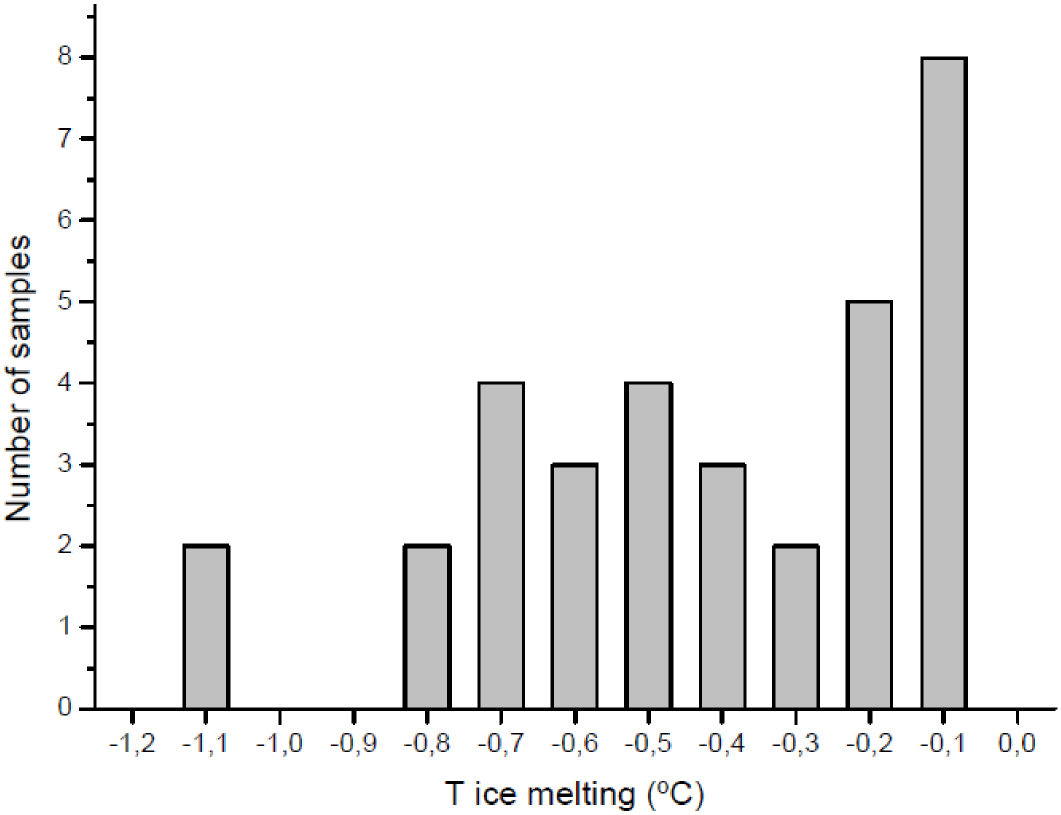

Figure DR9: a) Ice melting temperatures from 33 selected fluid inclusions which were biphasic at room temperature. The range of obtained temperatures (-0.1 ºC to -1.1 ºC, with a maximum at -0.1 ºC) indicates a low salinity of 0.31 mass % NaCl (for -0.1 ºC) up to 1.9 mass % NaCl (for -1.1 ºC).

Table DR2: Measured values of the homogenisation temperatures

54,1	53,5	54,3	50,5	52,6	52,1	54,5	54,5
54,0	52,7	48,3	56,4	60,0	58,3	61,5	57,8
52,0	52,0	51,4	46,7	46,0	46,0	38,1	49,1
51,6	56,2	54,2	52,3	53,6	56,2	48,1	

Solid inclusions have a spherulitic texture and are located within the fluid inclusions (Fig. DR8). The spherulites have a black core and a red halo under transmitted light microscopy, and SEM-EDS microanalysis shown that they are made of Mn, Fe, and Pb, and minor amounts of Si, Al, Na, Zn, and Cu. No X-ray diffraction pattern could be obtained from this spherulitic material. Similar spherulites are found in abundance in the Cave of Swords, where they produce the dark grey color of the gypsum crystals. The XRD pattern of this material shows two peaks (4.91 Å and 2.95 Å) corresponding to Mn_5O_8 compound (J17-0962 card).

2.3. Isotopic analysis

Several samples of anhydrite and gypsum from different locations in the mine were collected for stable isotope analysis. A representative sample of 0.1g of the mineral was first dissolved in 50mL of miliQ water. The dissolved sulfate was then precipitated as $BaSO_4$ by the addition of $BaCl_2 \cdot 2H_2O$ in acidic media (pH<2) to prevent $BaCO_3$ precipitation. The $\delta^{34}S$ was determined from $BaSO_4$ with an elemental analyzer coupled in continuous flow with a δC mass spectrometer. The $\delta^{18}O$ was measured in duplicate by on-line pyrolisis coupled with a δC mass spectrometer. The data were reported in the usual $\delta^{34}S$ CDT notation for the sulfur, and $\delta^{18}O$ SMOW for the oxygen. The overall reproducibility for the samples and laboratory internal standards was better than ±0.2‰ for sulfur and ±0.5‰ for oxygen.

The sulfur and oxygen isotopic composition of 17 sulfates (eight gypsum crystals, eight anhydrite crystals and one from water collected at level 530 are shown in Table DR3. Two $\delta^{34}S$ and $\delta^{18}O$ values of gypsum from Cave of the Crystals show no variation, besides the habit of the crystals (beams or blocky), with a $\delta^{34}S$ of 16.9‰ and the $\delta^{18}O$ of 17.6‰. One sample from the Cave of the Queen (sample 22) and another collected close to the Cave of the Crystals (sample 28) have similar isotopic composition. Gypsum samples from the Cave of the Swords (level 120) have slightly lower $\delta^{18}O$ values, 1‰ less than samples from the Caves of the Crystals and the Queen (level 290). The isotopic composition of the mineralizing solution is estimated to be around $\delta^{34}S$=15.3±0.2‰, with a gypsum-solution enrichment factor of 1.65‰ (value at 25°C) for sulfur (Thode and Monster, 1965), and $\delta^{18}O$=14.3‰, for a gypsum-solution enrichment factor of 3.3‰ (value at 55°C) for oxygen (Pearson and Rightmire, 1980). These values are perfectly consistent with the isotope composition of the present day water collected at level 530 (Fig. 3 and Table DR3). As seen in the same figure, the isotopic values of sulfate dissolved in the water could be explained by mixture of sulfates from the dissolution of anhydrite from different locations of the mine, those from the ore deposit area (samples 23, 24, 25 and

34) and those from upper levels in the mine (samples 4 and 21).

Table DR3: Isotopic composition of Naica samples.

sample	Mineral	location	$\delta^{34}S$	$\delta^{18}O$
9	Gyp.	Cave of the Swords	17,3	16,5
8	Gyp.	Cave of the Swords	17,6	16,7
15	Gyp.(beam)	Cave of the Crystals	17	17,4
31	Gyp.(blocky)	Cave of the Crystals	16,8	17,8
4	Anh.	close to Cave of the Crystals	16	13,7
28	Gyp.	close to Cave of the Crystals	16,3	17,8
21	Anh.	Cave of the Queen	16,7	13,0
22	Gyp.	Cave of the Queen	16,8	17,7
23a	Anh.	Ore body	14,4	16,0
24a	Anh.	Ore body	14,9	15,6
25a	Anh.	Ore body	13,8	15,7
25b	Gyp.	Ore body	17	17,9
34a	Anh.	Ore body	12,9	14,3
34b	Gyp.	Ore body	16,5	17,0
35	Anh.	level 590	13,8	14,7
36	Anh.	level 700	14,6	15,6
26	Water	level 530	14,6	14,4

2.4. Hydrochemistry

Two samples of groundwater from the levels 640 and 530 have been analyzed for major dissolved components. Cations and anions were analyzed by conventional Inducted Coupled Plasma with Atomic Emission Spectrometry (ICP-AES) and High Performance Liquid Cromatrography (HPLC), respectively, with an standard error lower than 5% in all cases. The δD and $\delta^{18}O$ values for the same two water samples were determined by H_2 and CO_2 equilibrium respectively and Isotope Ratio Mass Spectrometry (IRMS). The data were reported in the usual $\delta^{18}O$ and δD SMOW. The sample from level 530 has been also analyzed for tritium counting with a precision of ±0.31 TU (1 TU = 10^{-18} $^{3}H/^{1}H$). These data are summarized in Table DR4.

Table DR4: Chemical analyses (mg/L) of water samples from Naica mine. Legend : 290-1976, means sample collected at level 290 m in 1976. * (asterisk), measured at 20°C. Analyses 290-1976 and 430-1976 are after Dames and More (1977). The saturation index of some main minerals has been calculated with the code PHREEQC (Parkhusrt and Appelo, 2003).

	290-1976	**390-1976**	**430-1976**	**640-2002**	**530-2003**
Temperature (°C)	54	53	54	52	53
pH *	7.2	7.4	7.4	7.2	7.1
Alkalinity ($CaCO_3$)*	160	170	170	130	100
Cl (mg/L)	24	21	21	22	28
SO_4	2130	1960	2140	1922	2033
F	2.5	2.6	2.1	2.8	2.5
Ca	640	660	690	585	643
Mg	130	120	110	124	121
Sr	4.5	3.9	3.9	11	11.4
Na	120	125	110	89	96
K	11	11	11	14	13
SiO_2	75	110	50	33	31
Al	1.1	2.3	2.3	<0.1	<0.1
Mn	0.22	0.27	0.26	0.02	0.03
Fe	2.5	0.82	0.7	<0.1	<0.1
$\delta^{18}O$ (per mil)	-7.5	-	-7.8	-7.8	-7.5
$\delta^{2}H$ (per mil)	-58.7	-	-56.9	-57.0	-57.5
^{3}H (UT)	-	-	-	-	<0.31

Saturation Index

Calcite	0.41	0.64	0.68	0.28	0.10
Dolomite	0.54	0.94	0.96	0.31	-0.11
Anhydrite	-0.01	-0.04	0.02	-0.09	-0.03
Gypsum	0.02	0.01	0.05	-0.03	0.01
Fluorite	-0.13	-0.06	-0.24	-0.02	-0.11
Chalcedony	0.35	0.52	0.17	0.01	-0.03
Celestite	-0.35	-0.44	-0.42	0.02	0.02

3. Supplementary Discussion

3.2 Gypsum Nucleation

Solubility curves for gypsum and anhydrite as a function of temperature and salinity have been computed using version 2.12.1-669 (Linux) of the PREEQC software of the U.S. Geological Survey for three different pH values (6.0, 7.0 and 8.0) and three different salinity values. Calculations identified as "salinity 1" included a concentration of Na, Cl and Mg equal to the one measured in present-day waters collected in the Naica mine (100 mg/L Na, 24 mg/L Cl and 120 mg/L Mg). "salinity 2" conditions include twice this concentration and, finally, "salinity 0" calculations were made starting with pure water. In contrast with the negligible effect of pH, the salinity is a major control in the solubility of both gypsum and anhydrite. Fig. 4a of the paper shows these solubility curves.

Supersaturation values were calculated as $\sigma(T) = m_a(T)/m_g(T)$ where m_a and m_g are the equilibrium concentration of Ca in a saturated solution in equilibrium at temperature T with anhydrite and gypsum respectively as calculated by PHREEQC. This means that $\sigma(T)$ is the supersaturation with respect to gypsum of a solution obtained by equilibrium dissolution of anhydrite at temperature T. The computed supersaturation values as a function of T are shown in figure DR10.

The curves of induction time for nucleation were in turn computed from these supersaturation curves using equation

$$J = A exp - \frac{16\,\pi\gamma^3 v^2}{3k^3 T^3 ln^2 S}$$

discussed in the article. Data from three different papers have been analyzed to collect the empirical data to parameterize this equation (Lancia *et al.*, 1999; Prisciandaro *et al.*, 2001; He *et al.,* 1994). These articles show a considerable scattering in the values of activation energy, which is a critical parameter for the nucleation and can have a very large impact (several orders of magnitude) in the value of the induction time. Fortunately, owing to the extremely step slope of the induction time curve, these changes can only produce shifts of a few degrees in temperature for reasonable time values in the range "one day" to "some thousand of years". Consequently, the inherent uncertainty in these empiric data does not compromise our results and the only expected influence is a shift of (at most) 5 ºC in the lower boundary of the temperature window proposed. The parameters used for the curves shown in the article (Fig. 4b) were obtained from Lancia *et al.* (1999) because this article contains data for supersaturation values smaller than the remaining references, making more accurate the extrapolation needed for the very low supersaturation values involved in the crystallization of the Naica gypsum crystals.

3.3 Evidence of phase transition anhydrite-gypsum in Naica

Several mineral samples of gypsum, anhydrite and enclosing limestone were collected from the Caves of Crystals, Swords and Queen and from the surroundings. The mineralogy of the samples was investigated by transmitted light microscopy, Powder X-ray diffraction method (XRD), and by Scanning Electron Microscopy and microanalysis (SEM-EDS). In Figure DR11 evidence of a transformation anhydrite-gypsum at microscopic level is shown.

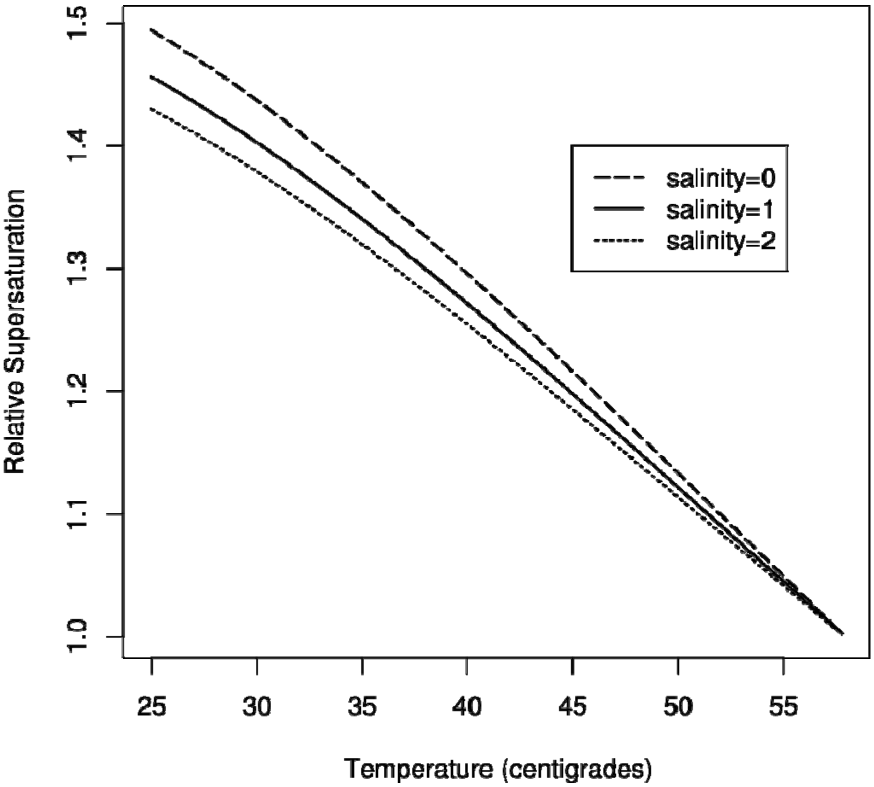

Figure DR10: Temperature dependence of supersaturation with respect to gypsum of a solution obtained by equilibrium dissolution of anhydrite.

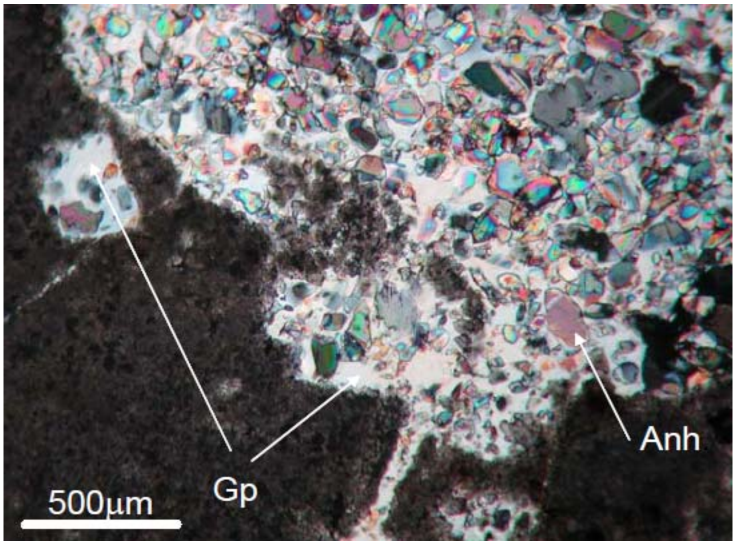

Figure DR11: Poikiloblastic gypsum crystals growing inside an anhydrite matrix. Sample collected close to the Cave of Crystals, Naica mine, level 290.

4. References cited in the supplementary materials

Bakker, R. J., 2003, Package FLUIDS 1. Computer programs for analysis of fluid inclusion data and for modeling bulk fluid properties: Chemical Geology, v. 194, p. 3-23.

Bodnar, R. J., 1993, Revised equation and table for determining the freezing point depression of H_2O-NaCl solutions. Geochimica Cosmochimica Acta, v. 57, p. 683-684.

Butler, G.P., 1973, Strontium geochemistry of modern and ancient calcium sulphate minerals, in Purser, B.H., ed., The Persian Gulf: Springer-Verlag, Berlin Heidelberg New York, p. 423-452.

Dames and More, 1977, Reporte de Estudio hidrogeológico de Naica. Compañía Fresnillo S.A. de C.V. Internal report.

Degoutin, N., 1912, Les grottes a cristaux de gypse de Naica (Mexique): Soc. Cient. Ant. Alz. Rev. 32, 35-38.

Foshag, W., 1927, The Selenite Caves of Naica, Mexico: American Mineralogist, v. 12, p. 252-256.

He, S., Oddo, J. E. and Tomson. M. B., 1994, The nucleation kinetics of calcium sulphate dehydrate in NaCl solutions up to 6 m and 90 ºC: Journal of Colloid Interface Science, v. 162, p. 297-303.

Lancia, A., Musmarra, D. and Prisciandaro, M., 1999, Measuring Induction Period for Calcium Sulfate Dihydrate Precipitation: AIChE Journal 45, 390-397.

Parkhurst D. L. and Appelo C.A.J., 2003, PHREEQC- A hydrogeochemical transport model. http://wwwbrr.cr.usgs.gov/projects/GWC_coupled/phreeqc/index.html

Pearson F.J. Jr. and Rightmire, C.T., 1980, Sulphur and oxygen isotopes in aqueous sulphur compounds., in Fritz, P. and Fontes, J. Ch. (Eds.), Handbook of Environmental Isotope Geochemistry. Elsevier.

Prisciandaro, M., Lancia, A., and Musmarra, D., 2001, Gypsum nucleation into sodium chloride solutions: AIChE Journal, v. 47, 929-934.

Simon, B., and Bienfait, M., 1965, Structure et mécanisme de croissance du gypse: Acta Crystallographica, v. 19, p. 750-756.

Thode H.G., and Monster, J., 1965, in Fluids in Subsurface Environment, AAPG Memoir nº4. 367-377.

5. Acknowledgments for supplementary materials

We acknowledge María del Carmen Jimenez de Haro (Instituto de Ciencia de Materiales de Sevilla) for operation of the scanning electron microscope.

BACK TO G1EXG RADIO PAGE

EPAD MOSFET - 'near zero' and 'zero' threshold devices (e.g. the ALD110900)

I will add more details to this page later on but for the moment it gives links to important and inspiring sites.

ARTICLES
1) High Sensitivity Crystal sets, Technical Topics, Pat Hawker, Rad Com, p. 77-78, May 2007.
2) A novel kind of 'crystal set' radio, Giles Read, Rad Com, p. 60-61, June 2007.
3) also see Pat Hawker, Technical Topics, Rad. Com. January 200, page 53-57.
4) High Sensitivity Crystal Set by By Bob Culter (N7FKI)
5) Next-generation Zero-Threshold Voltage EPAD™ design enables circuits with greater operating range in low voltage supply environments

LINKS ABOUT THE CHIPS / DEVICES
1) Advanced Linear Devices web site
2) EPAD MOSFET (ALD web site)
3) ALD110800 and ALD110900 page (ALD web site)
4) ALD110802 and ALD110802 page (ALD web site)

ORDERING THE CHIPS / DEVICES
1) mouser.com (USA)

'CRYSTAL SET' RADIO'S - DIFFERENT DETECTORS

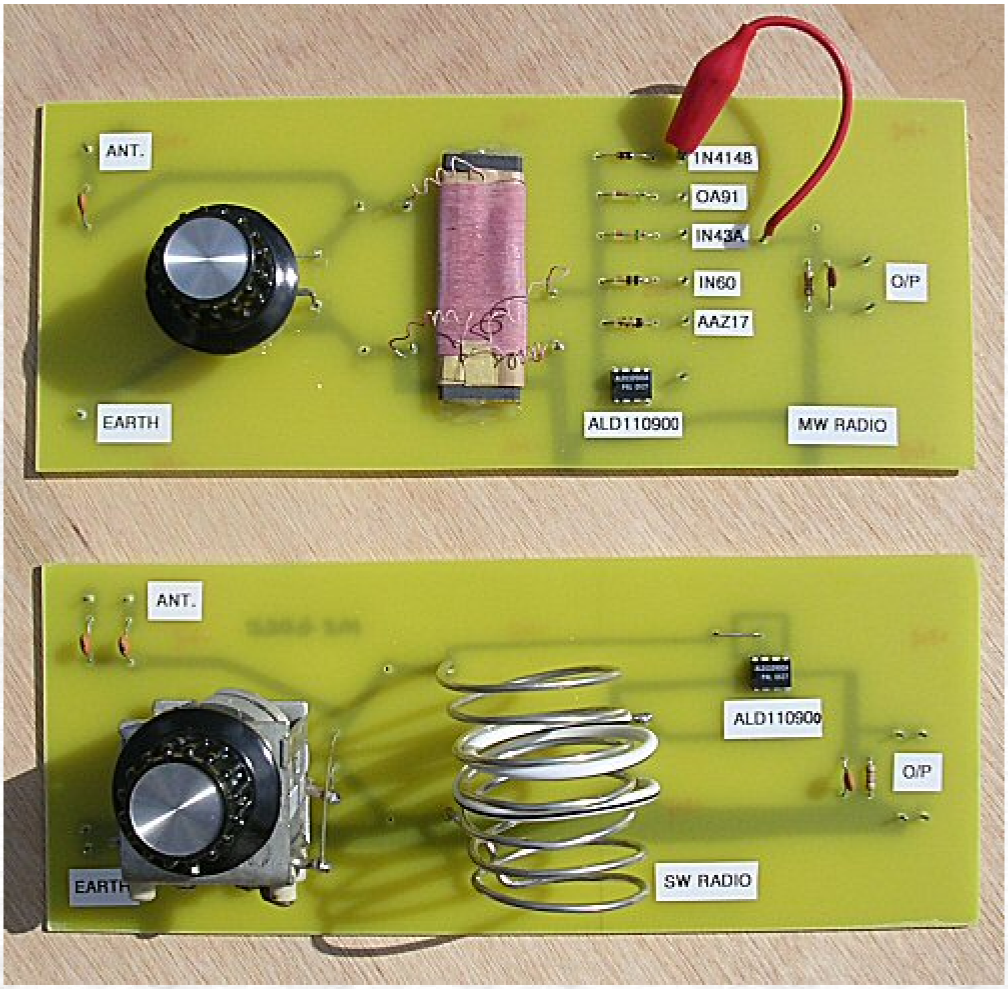

**Top: the medium wave (MW) test radio with a selection of 'detectors' (selected by the croc clip lead)
Bottom: the short wave (SW) version with a ALD110900 detector
(on each circuit board the connections are: antenna - top left, Earth - bottom left, audio output (heaphones / amp) - far right.)**

RADIO AND RADIO WAVES

The simplest 'radio' can be a piece of wire attached to the input of an amplifier. Why does this pick up radio signals? Well in principle it should not be able to pick up anything but in reality poor solder joints in the amplifier circuitry, as well as point contact effects between the ends of the anenna wire going into the input socket of the amp as well as other effects mean that you do often hear radio signals.

Radio waves are electromagnetic waves as they pass through a metal they induce small voltages into it. Any piece of metal e.g. metal fram specs, tape measures, metal window frames, a piece of wire etc. will act as an antenna and have tiny voltages induced into them. These voltages will be due to natural radio waves (from Space as well as the Sun and Earth), radio stations, satellites, mobile phones, garage door remote controls, microwave ovens .... the list goes on and on.

In the case of long, medium and short wave radio stations the amplitude of the radio wave signal is modified by the music, voice or program - we call it Amplitude modulation AM. Here the strength of the radio waves varies as the tones, loudness and pitch of the program vary. As a result the voltages induced in the metal object antennas distant from the radio wave stations also vary in accord this program information or modulation.

To actually hear the programs on the radio waves you cant actually take this tiny signal and listen to it directly (say with headphones or an amplifier). This is because the signal is a very high frequency signal out of the range of human hearing. To get the audible information - the program - 'off' the radio signal you need a device known as a 'detector'. This is usually a diode but lots of other things can act as inefficent detectors for example a mineral called galena, coke (burnt coal) ... as well as rusty screws and bad solder joints which is how the ampilfier mentioned above apparently picks up radio signals.

In the early days of radio crystals of galina were used as diodes and so these radios became known as 'crystal set radios'. They did not have any amplifiers and did not require any battery to work. They got all their power from that induced into the (of very long) antenna by the radio signal(s). The radio detector is an extremely important part of the radio and its proper function determines how well the radio works.

The radio signal is an AC signal. A speaker or headphones can not resonate at such high frequencies as radio waves and if you try to wire in the RF to the headphones the diaphram just averages the power. As half the time the RF is positive and half the time negative, in an AC signal, the average is zero and so you dont hear anything. The diode or detector is a device called a diode which only conducts electricity in one direction. So it only allows the positive, or negative half (one half but but not both) of the AC signal to go through (depending on which way it is wired). After the diode the average is now no longer zero its actually a changing signal dependant on the modulation of the radio wave - which is the information we want to hear. As this is at audio frequencies, and as the diaphram can move at audio frequencies, you hear the information on the detected RF signal - the music, voice or what ever.

Why should some detectors (diodes) be 'better' than others? The answer is not that some are better at 'magnifying' the signals but rather that each diode requires a certain threshold voltage in order to start conducting (one way). Signals below this level will not get passed through and a signal only slightly larger will therefore only pass through weekly (the diode unfortunatly absorbing the majority of the signal). As a result for large signals you wont expect to see much difference between different detectors but for weak signals there can be a great deal of difference. In an ideal world the diode would start to (forward) conduct at just above zero volts and this ideal detector (which does not conduct at all in reverse) would produce the greatest signal possible for the particular strength signal applied to it.

Recently I was playing around with OA91 germanium diodes for a mobile phone detector. I made a simple 10cm per side square loop and wired it directly to an LED. I was hoping that if I brought the mobile near to the loop and texted, or made a phone call, the RF should be picked up by the loop (which is almost resonant) and the LED should light - it didnt work! But putting an OA91 in series I did get it to work. So why did I need the OA91 diode when the LED is itself a diode?

It can't be due to the threshold voltage in this case as I am still using the LED and I did get it to work with the same signal! So I guess it because the capacitance of the LED is very high compared to the point-contact germanium diode. If we extend this thinking to the crystal set radio I guess the germanium diodes have very little capacitance and so dont allow any of the 'wrong' half of the AC signal through and so act as a more perfect detector. An LED would have a high capacitance and so let some of the wrong half of the AC signal through which on averaging by headphones would be nearer zero (see above) - hence less sound.

A low threshold (turn-on) voltage and a low self capacitance are crucial things in a good detector. Point contact germanium diodes having a very small area (a point of contact) have low capacitance and the germanium semiconductor-metal junction of the contact has a low turn on voltage. EPAD Mosfets also have low capacitance, high input resistance (low loading of the signal) and low turn on voltage.

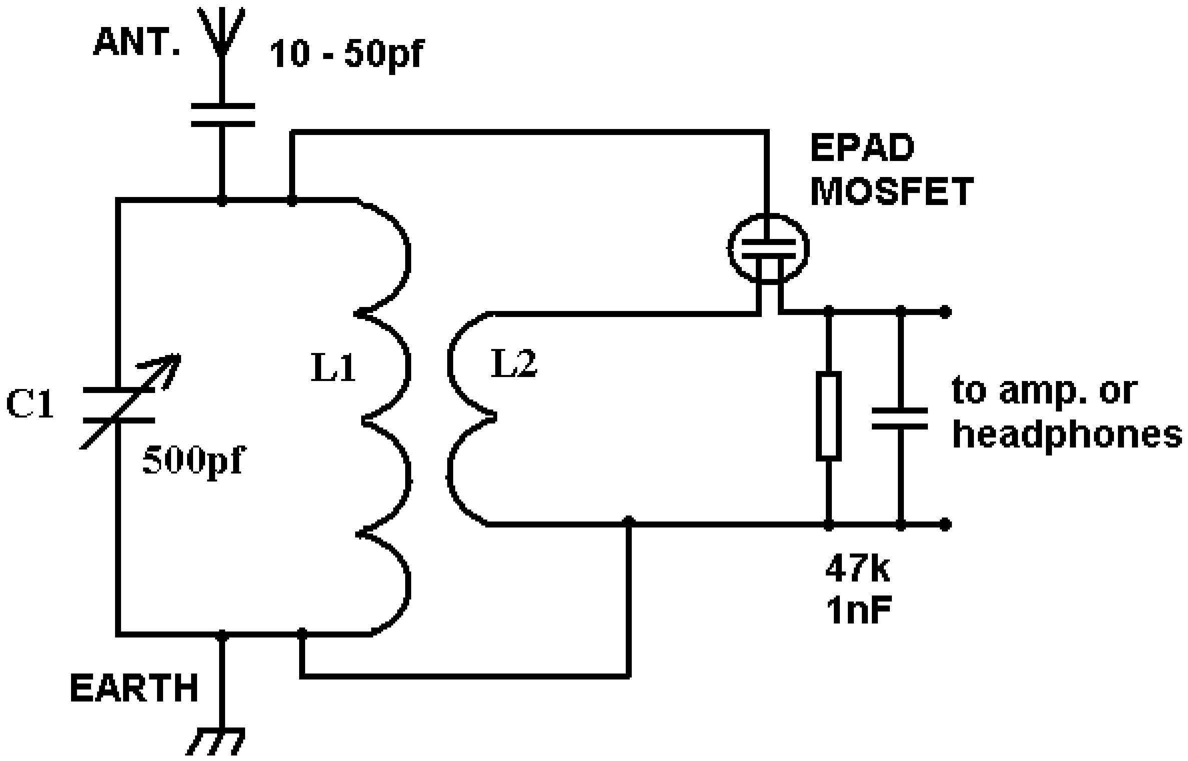

MEDIUM WAVE TEST CIRCUIT (top in photo)
As a simple test circuit to compare detectors I made up a simple 'crystal set' radio from a ferrite rod coil from an old radio, tuning capacitor an array of diodes (crystals) and a simple RF filter composed of a 47k resistor and 1nF capacitor. The coil and capacitor formed the resonant circuit which was fed via a 10 - 50pf capacitor by the antenna (long wire). This capacitor helps to reduce the effect of the antennas own capacitance and inductance from modifying the resonant frequency of the tuned circuit. Ideally this should be as small a capacitance as you can get away with but too small a value will reduce the radios sensitivity ....

An earth was connected to the ground connection of the circuit. Instead of taking the antenna end of the resonant circuit to a diode detector (as many simple crystal sets do) I used a ca. 10 turn coupling coil around the coil. This was done to limit the loading on the resonant circuit by the detector. It should help keep the tuning sharp and also help to keep the overall (band) spread of tuning as large as possible.

Because of the reduction in turns (ca. 100 turns : 10 turns, i.e. 10:1) the coupling coil reduces the signal that is available to drive the detector. For example if I use a standard silicon diode (e.g. a 1N4148) which requires about 0.6V to conduct then we might need ca. 10 x 0.6 = 6V of RF to be developed in the resonant circuit! A high Q resonant circuit might do this with a long antenna coupled to it but without such a generious antenna such a set-up wont be very sensitive.

RESULTS
I tried an array of 'crystal set' type diodes to campare them. Typical rough results are shown in the table below. This was with an Earth and about 3-4m of antenna wire randomly strung near to ground level (so its a pretty bad set up really - a good test).

The EPAD MOSFET data sheets described how the gate and drain of one of the MOSFETS can be joined together to form the anode of a diode the source then forms the cathode. I tried this arangement with the ALD110900 in order to make up a 'crystal set' diode from this MOSFET but it didnt perform well.

Then I tried the alternative arrangement as described in the references above. In these articles they dont use the MOSFETS as diodes but connect the gate directly to the resonant circuit. As the input resistance is extremely high (c.a. 1E12 ohms) and the input capacitance is very low (c.a 3pf) it does not load the tuned circuit by any appreciable amount. The source and drain are then wired between the coupling coil and filter just as the usual diode detector is in a crystal set radio. In this configuration the MOSFET is working more as a gate voltage controlled resistor rather than a diode i.e 'off' for half an RF cycle and 'on' for the rest - this lets the envelope through and you hear the music or sound. One article describes the arrangement as a synronious detector. As there are two MOSFETS in one package the two are wired in parallel i.e. gate to gate, drain to drain (there is a common source) etc.

DETECTOR TYPE	AF voltage produced after filtering
1N4148 silicon	no signal
OA91 germanium	1mV
1N43A germanium	1mV
1N60 germanium	1mV
AAZ17 germanium	1.5mV
ALD110900 EPAD MOSFET wired as diode	<1mV
ALD110900 EPAD MOSFET wired as sync. detc.	10mV

Typical results of the AF voltages produced by the detector (for use by headphones or amplifier) for an Earth connection, small (ca. 3-4m antenna wire near ground level) and the prototype 'crystal set' radio.

For the simple set-up described above there was not enough signal to get the silicon diode to work at all (although a much longer wire did provide a very small signal). A simple listening test and measurement of the AF prodiced by each diode showed that their was very little difference between the various germanium diodes. The MOSFET simply wired as a diode (gate wired directly the drain) was a poor detector - better than a silicon but not nearly so good as any of the germanium diodes. Finally the EPAD MOSFET wired as a syncronious detector worked really very well apparently producing ca. 10 times the signal (in other words it has much less loss than a germanium diode) for this particular set up and station than the germanium diodes.

As the syncronious arrangement was so succesful it got me thinking about this circuit. As the gate goes straight in on the main resonant circuit where the voltages, in a good Q circuit, might well be relatively high I thought I would try a standard FET (2N3819) and MOSFET (VN10KM) instead - and they worked! (they are a lot cheaper than an EPAD MOSFET chip and easier to get). Some FETS have a larger input capacitance so it helped to put a 10pf capacitor in series to the gate (a 10M ohm from gate to earth might reduce 50Hz pick up).

They were not as good as the EPAD MOSFET (although I only tried a single device) but the arrangment was still better than a simple 'crystal set' diode. For weak signals the EPAD MOSFET should perform better as it is able to keep going while the FET and the standard MOSFET will start to trail off as the resonant circuit voltage drops below the relatively higher gate voltage required for these devices. However it was interesting to see that it seems to be possible to get a more efficent set-up for a 'crystal' radio using a standard FET or MOSFET than with the standard germanium diode detector.

By the way I recently re-tried the 1N4148 diode but wired a 1.5V AA battery with a 100k resistor across it to forward bias it - and it worked ! It seemed to work better than a OA91!

SHORT WAVE 'CRYSTAL SET RADIO (bottom in photo)

The bottom circuit shown in the photo above is my own short wave version of the 'crystal set' radio but based on the references above. It's the same circuit as the MW radio but has a reduction drive tuning capacitor and different coil. I used about 7 turns (ca. 4cm diameter) for the resonant circuit and a couple of turns of insulated wire (white) for the coupling coil. I guess the radio tunes from about 3-8 MHz - I havent tried a sig gen on it yet.

I made this in the summer months (not the best time for listening on this part of the spectrum). Using an Earth and few meters of wire for antenna I heard 2 or 3 stations from around the world during the day while at night there were at least 7 including the voice of Russia world service, Radio Thailand as well as German and French speaking stations.

BACK TO G1EXG RADIO PAGE

THE CREATIVE SCIENCE CENTRE

Dr Jonathan Hare, The University of Sussex
Brighton, East Sussex. BN1 9QJ

home | diary | whats on | CSC summary | latest news

TARINGA! Posts Shouts Comunidades Juegos Tops INGRESAR REGISTRARTE

Free Resume Template
Free Resume Samples
Free Resume Formatting
Free Resume Builder
My Perfect Resume

2 años Taringa! › Hazlo tú mismo › Construí tu propia radio Galena
Me gusta 587 mil Seguir

A favoritos
Ir a comentarios
Compartir vía mail
0
T!
7

Construí tu propia radio Galena

Construcción radio Galena

Si te gustan las radio, este proyecto es muy educativo y entretenido, con la “Radio Galena” podrás recepcionar frecuencia de AM y es la base de la radio recepción, además es una radio que no requiere energía proveniente pilas, baterías ni fuente de alimentación. Ya que funciona gracias a la energía que propaga las emisoras, aparte se presta para experimentar mucho, solo busca en google “radio galena” encontraras muchísimas variantes.
Como punta pie puedes leer un poco más sobre el tema aquí http://100ciaencasa.blogspot.com.ar/2014/01/radio-sin-pilas.html y aquí http://es.wikipedia.org/wiki/Radio_a_galena

El circuito es por demás sencillo

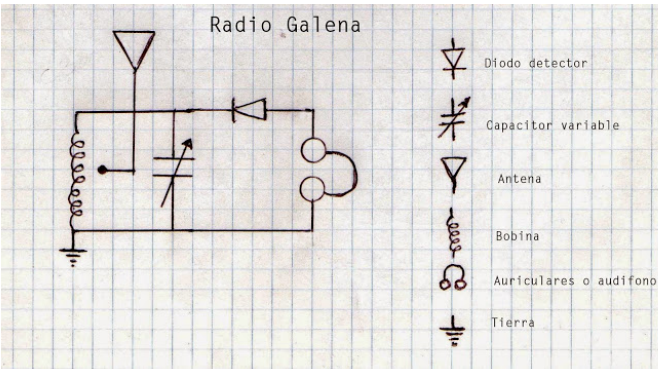

Lista de materiales
1-Diodo de germanio aquí usamos el 1N60 (podrás usar otros siempre y cuando sea detector)
1-auricular de alta impedancia (son un tanto difícil de conseguir, pero puedes usar otro de teléfonos viejos o a la salida puedes conectarle un pequeño amplificador) (no te pierdas próximamente el proyecto Amplificador portátil de bajo coste con LM386)
1-Tandem o capacitor variable, (se lo puede comprar o conseguirse de una radio vieja)
Unos metros de cable para bobinado o cable magneto, yo recicle el

dagonett
EXPERTO
72 SEGUIDORES
5 POSTS
2K PUNTOS

Recomendados
Ver más

construir una radio de galena

Cinco canales de Youtube para aprender experimentos caseros

5 Experimentos Caseros... ¡Listos Para Hacer!

Experimentos caseros

bobinado de un motor quemado cuyo cable esmaltado es de .045 milímetro (casi ½ milímetro de espesor)
1-base de madera para hacer el montaje medidas 13x13 cm aprox.
1-tubo de pvc de 4 cm (1,5 pulgadas) por 9 cm de largo (se lo consigue en ferreterías es el que se usa para desagüe)
1-una varilla o alambre de unos 15 cm. aproximadamente (tips se puede usar un rayo de bicicleta)
Como siempre necesitaras estaño, soldador, alicates, algunos tornillos o clavitos, algún pegamento

Comenzar aquí

3 pasos sencillos

1) Haga clic en "Comenzar aquí"

2) Descargue desde nuestro sitio web

3) Obtenga el conversor de archivos gratis

EasyPDFCombine™

Diodo de germanio 1N60

Audífonos alta impedancia

Base madera de 13x13 cm aprox

Cable esmaltado para bobinado o cable magneto

Tandem o capacitor variable y bobina montado en base.

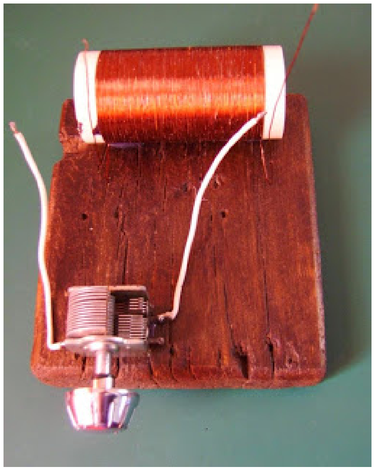

Vallamos a la construcción
Lo primero que vamos hacer es la bobina de nuestra radio, para ello usamos el tubo de pvc fijamos en un extremo del cable esmaltado dejando un excedente de unos 10 cm. para hacer las conexiones luego, podemos usar una cinta adhesiva o como en mi caso una gotita de cianocrilato. Empezar el bobinado en el sentido que quieras (en este caso sentido de las agujas del reloj), solo ten en cuenta que las espiras estén bien prolijas y lo más juntas posibles sin que se encimen unas con otras, realizar 130 espiras y asegurar el otro extremo dejando también 10 cm. para la posterior conexión.

Después montamos según diagrama citado arriba, recuerda que el cable esmaltado de la bobina tienes que pelarlo ayudándote con una lija fina para sacar el esmalte sino el mismo no hace contacto ni se suelda ya que la película esmaltada es aislante.

Fotos de galena terminada

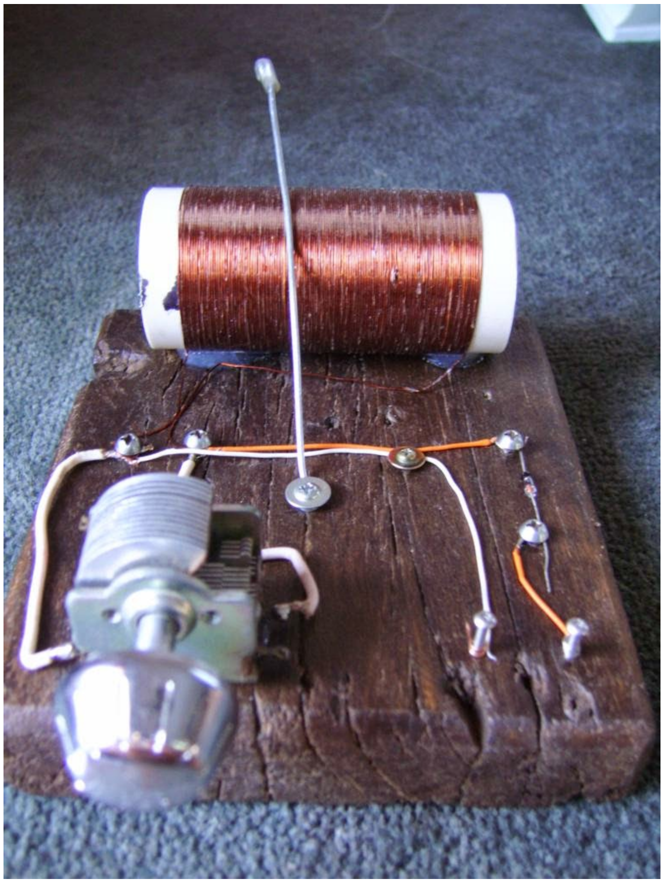

Para la recepción:
Conectamos el audífono de alta impedancia o en su defecto el mini amplificador
A tener en cuenta
vamos a lijar una sección del esmalte a lo largo de la bobina (ver figura) , la antena esta conectada a una varilla, "recepción gruesa" por así llamarlo con ella limitaremos la cantidad de espira de nuestra bobina para poder sintonizar varias estaciones de radio a modo de derivaciones, si todo funciona bien tendremos sintonía quizás algo mezclada y de no muy buena calidad, entonces hacemos el ajuste fino moviendo el tándem o capacitor variable.
Una de las cosas más importante y tener en cuenta en este tipo de radio es la antena y puesta a tierra, la antena se aconseja unos 10 metros como mínimo, pero no te asustes, puedes usar como en mi caso un alambre extendido como lo que usan para colgar la ropa y la puesta a tierra una reja que esté haciendo contacto con la tierra o cañería metálica como la de una bomba de agua vieja

Detalle de conexiones.

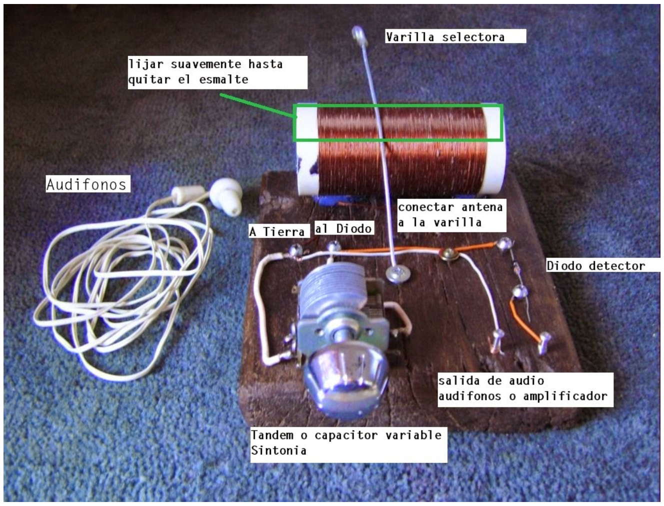
Varilla selectora
lijar suavemente hasta quitar el esmalte
Audifonos
conectar antena a la varilla
A Tierra
al Diodo
Diodo detector
salida de audio audifonos o amplificador
Tandem o capacitor variable Sintonia

Nota: se puede experimentar con el tamaño del núcleo de la bobina como así también la cantidad de espiras y espesor del cable esmaltado de la bobina o cable magneto, yo he realizado para este proyecto unas 5 bobinas diferentes aunque todas recepcionaban alguna emisora, pero el que mejor me resulto es el que aquí describo. También puedes experimentar con largo de antena y todo los cambios que se nos ocurran, este es un buen proyecto para la experimentación, también como detector en vez del diodo de germanio, se puede usar la piedra galena que es la que le da el nombre al modelo (un tanto difícil de conseguir), con un transistor conectando solo dos terminales, hasta como hicieron en la segunda guerra mundial con una hoja de afeitar y una punta de un lápiz, es cuestión de que investigues un poco y que experimentes. En mi caso llego a funcionar sin el capacitor variable y con recepción muy buena de las estaciones potentes.
Bueno espero te animes ya que es muy gratificante cuando sale funcionando.
Si quieres puedes mandarme o dejarme un comentario con las fotos de tu Galena.

Saludos y hasta pronto

Los invito a pasar por mi blog mi blog:
http://cachivacheselectronicos.blogspot.com.ar/

1 | Antivirus Gratis 2018
Mejor Antivirus Gratis de 2018. Comparar y Encontrar el Mejor Antivirus Gratuito mejorantivirus10.com

2 | Abogado de DUI
Abogados Expertos En Casos De “Manejar Bajo Influencia” abogadocross.com

Compartir Twittear

Fuentes de Información - Construí tu propia radio Galena

El contenido del post es de mi autoría, y/o, es un recopilación de distintas fuentes.

Tags: electronica | radio | experimentos caseros | reciclado | galena

Dar puntos

1 2 3 4 5 6 7 8 9

"New Age" Crystal Radio

A project by Luciano Loria

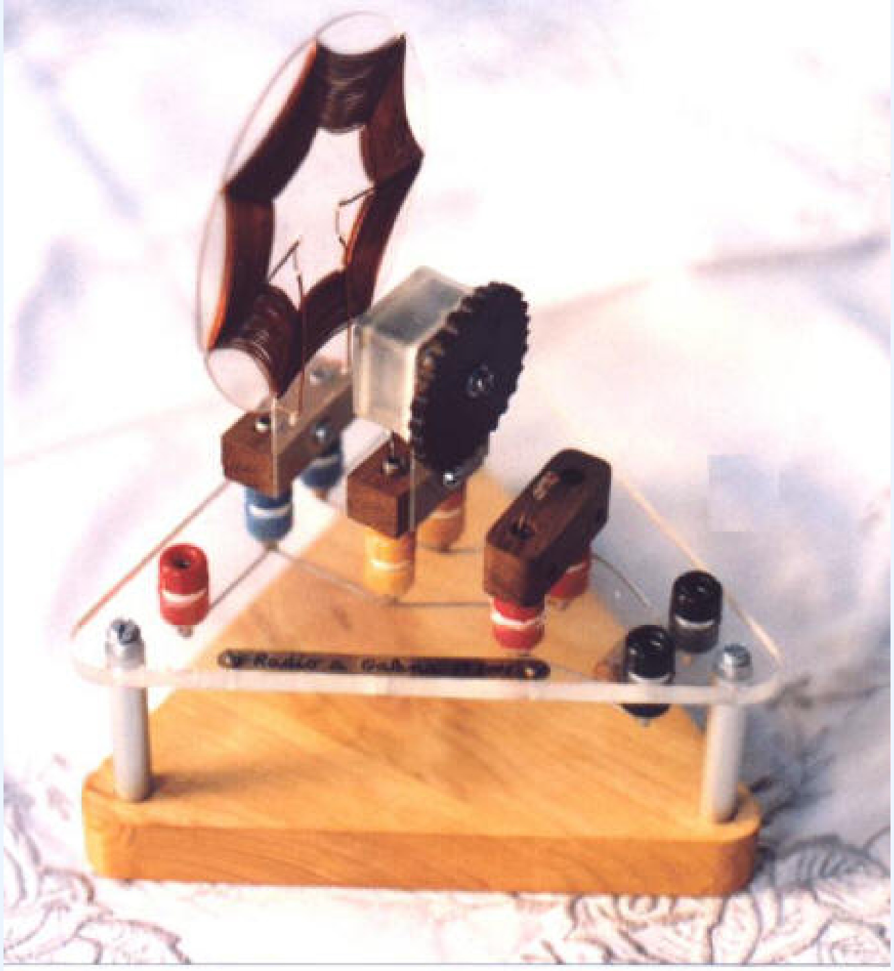

Following is proposed the construction of the most classic and simple crystal receiver, the principle scheme is, in fact, reported on all books or manuals of radio-technical and physics. In this realization the look was taken in order to expose the appliance, in the living room window, recalling, certainly, the interest of those who observe it. Clearly, the radio is fully functional: just provide adequate antenna and ground and a high impedance headset and listening to the local station is guaranteed. The construction is quite simple and affordable for anyone with good will and a minimum of simple hobbyist equipment. First of all, obtain the following material:

- 1 variable capacitor, also for recovery, transistor radio type (any capacity value);
- 1 germanium diode (any type);
- 1 capacitor of 2.2 or 3.3 or 4.7 nF;
- 10 4 mm diameter bushings of various colors;
- 6 bananas diameter 4 mm;
- 1 wooden strip 10 x 10 mm, length 8-10 cm approximately;

- 1 wooden cutout, 2 cm thick, from 15-20 x 15-20 cm on the side;
- 1 cutout of plexiglass, thickness 4-5 mm, from 15-20 x 15-20 cm on the side;
- 1 aluminum or brass tube, diameter 6-8 mm, length about 15 cm;
- 3 wood screws 3mm diameter, about 45-60 mm long;
- 1 piece of wire for connections in tinned or silvered copper with a diameter of about 0.8-1 mm.
- For the basket-shaped coil see separate **description** .

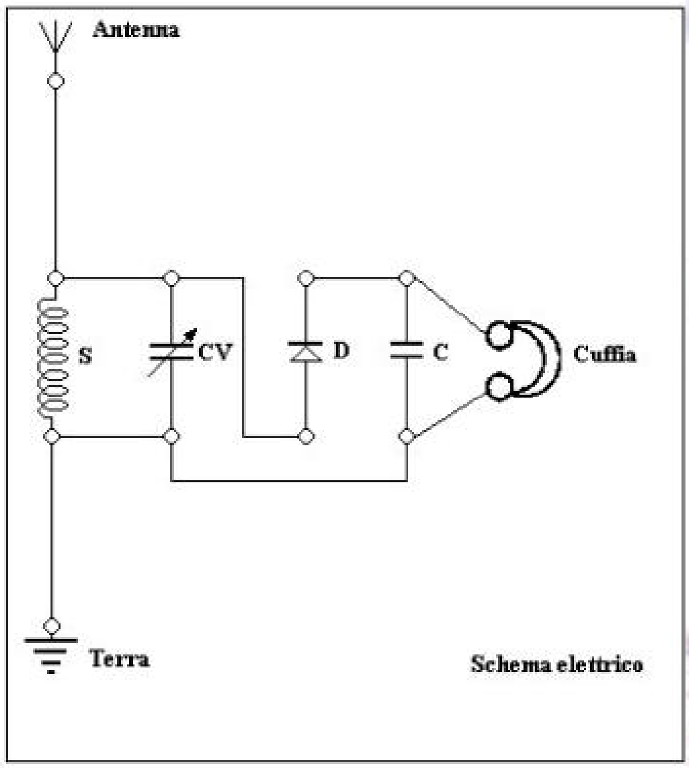

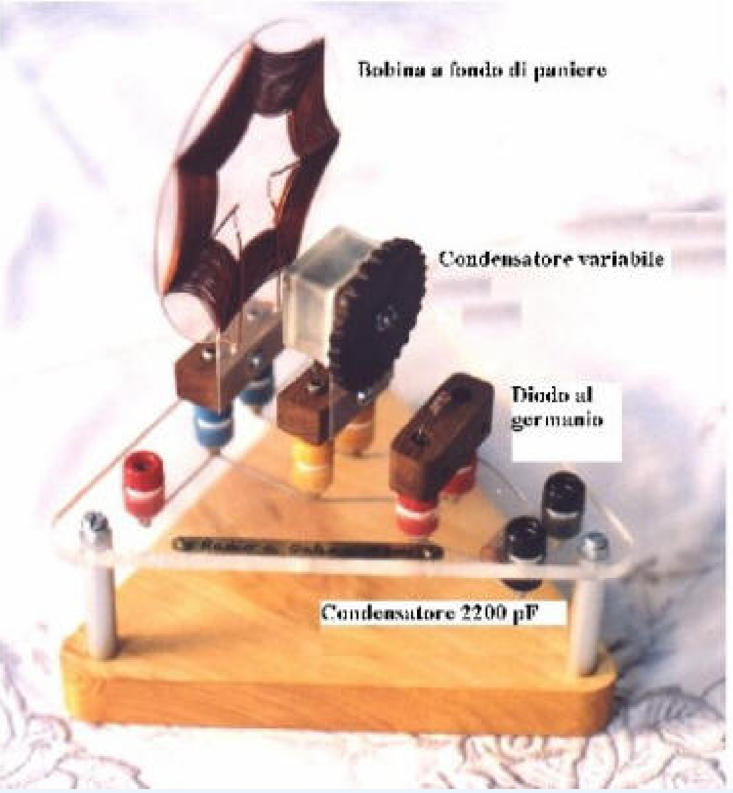

Construction:

Cut out from the wood and plexiglass an equilateral triangle of about 15 cm on each side and smooth the edges; on the plexiglass triangle, about 3 cm from the vertex, three 3 mm diameter holes for the passage of the screws and 10 8 mm holes to fix the bushes. From the diagram and from the photo it is easy to establish the position of these holes, the interaxis between the bushings where they will be inserted: the coil, the variable capacitor, the diode and the bonnet, is 19 mm (standard pitch of the normal plugs and sockets 10 A). Fit the bushings in their place and, following the diagram, make the few connections with the tinned or silver wire, the only element that will be mounted in a fixed position is the capacitor from 2.2-4.7 nF, to be welded directly on the headphone output bushings. Divide the aluminum or brass tube into three parts, making pieces of about 4 cm; connect the top plate in plexiglass with the wooden base using the wood screws, using the three tubes as spacers. Then, using the wooden strip, make the wooden spikes, 19 mm pitch, of support for coil, CV and diode; the bananas will provide the necessary brass thorns. Fix the coil and the variable capacitor to their supports by means of screws and connect the terminals to the plugs. The diode must be fixed directly without any other precautions. Insert in their place, the tuning coil, the variable capacitor, the germanium diode; connect antenna and ground,

luciano.loria@gmail.com

Back to the Crystal Receivers Page

Construction and Operation of a Simple Homemade Radio Receiving Outfit

The 1922 Bureau of Standards publication, *Construction and Operation of a Simple Homemade Radio Receiving Outfit* [1], is perhaps the best-known of a series of publications on radio intended for the general public at a time when the embryonic radio industry in the U.S. was undergoing exponential growth.

While there were a number of earlier experiments with radio broadcasts to the general public, most historians consider the late fall of 1920 to be the beginning of radio broadcasting for entertainment purposes. Pittsburgh, PA, station KDKA, owned by Westinghouse, received its license from the Department of Commerce just in time to broadcast the Harding-Cox presidential election returns. In today's world where instant global communications are commonplace, it is difficult to appreciate the excitement that this event generated.

News of the new development spread rapidly, and interest in radio soared. By the end of 1921, new broadcasting stations were springing up all over the country. Radios were selling faster than companies could manufacture them. The demand for information on this new technology was almost insatiable. The Radio Section of the Bureau of Standards provided measurement know-how to the burgeoning radio industry as well as general information on the new technology to the public. Letters to the Bureau seeking information on radio technology began as a trickle, and then soon became a flood. Answering them became a burden.

Circular 120, published in April 1922, began: "Frequent inquiries are received at the Bureau of Standards for information regarding the construction of a simple receiving set which any person can construct in the home from materials which can be easily secured. This publication has been prepared to meet these inquiries." The circular also noted that the Bureau was cooperating with the Bureau of Markets and Crop Estimates of the Department of Agriculture. In the early days of radio, broadcasts to farmers of crop price reports were an important public service.

In 1921 factory-made radios were very expensive. Many of them cost more than $2000 in today's dollars, and less affluent families could not afford to have one. When the Bureau published this circular (which sold for only five cents) on how to build a simple crystal radio at home, it was an instant success. Thousands of orders

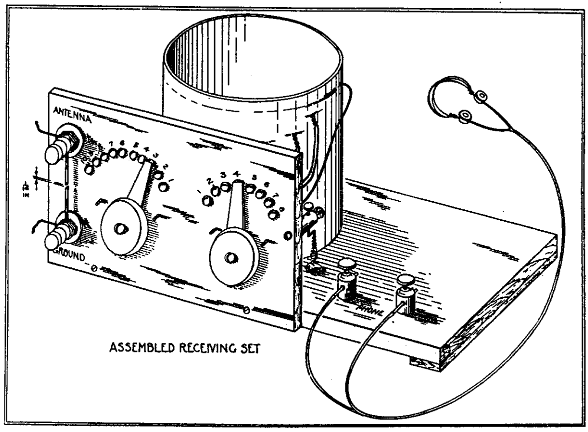

Fig 1. The crystal radio described in Circular 120.

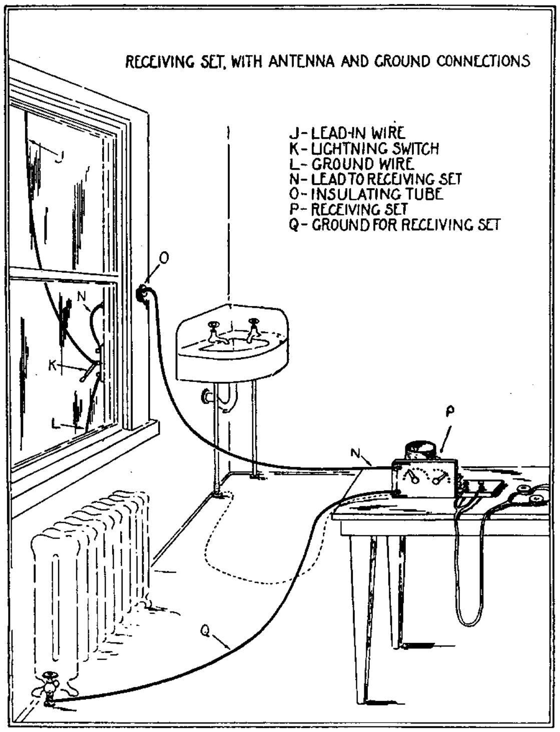

Fig. 2. Diagram showing the antenna and ground connections to the radio.

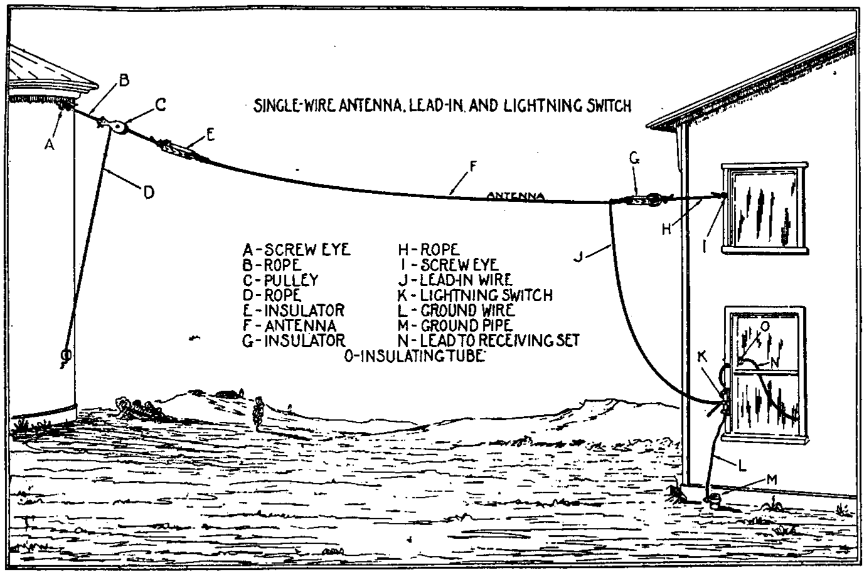

Fig. 3. Details of the outdoor antenna recommended for use with the crystal radio.

flowed in from every state. The circular showed how almost any family having a family member handy with simple tools could make a radio. The circular states "Satisfactory results have been obtained from sets constructed according to these instructions by persons having no previous experience with radio." The authors stated that the cost of the materials for the radio and associated headphones and antenna system should not exceed $10.

Newspapers around the country urged readers interested in radio to order a copy, and in many cases they reprinted the entire article. Even today, antique radio collectors searching attics and basements often come across dusty home-built crystal radio sets that were obviously constructed using the Bureau's plans. No doubt the remarkably rapid growth of the radio industry in the U.S. during the 1920s (analogous to the growth in Internet-related businesses today) was aided considerably by information provided to the public by the Bureau, such as this circular and the other items noted in the bibliography [2-6].

Circular 120 proved such a success that the Bureau published a number of other circulars on radio technology intended for the general public, as noted in the bibliography. All were popular. In some cases, commercial radio manufacturers copied the circuits proposed by the Bureau. During the same era, other radio publications of the Bureau captured public attention as well, though perhaps in smaller quantities than Circular 120. *Radio Instruments and Measurements* [7], Circular 74, published first in 1918 with a second edition published in 1924, became a best seller in the post-WW I era. Its principal authors were most likely J. H. Dellinger, J. M. Miller, F. W. Grover, and G.C. Southworth. The authors of Circular 120 are not stated, but some of those same individuals likely contributed to it. J. Howard Dellinger's story is particularly interesting. He came to the Bureau in 1907 as a laboratory assistant at $900 per year, and retired as Chief of the Central Radio Propagation Laboratory in 1948. He headed the Radio Section for much of its existence. Appendix D of Reference [9] is a biography of Dellinger.

The Principles Underlying Radio Communication [8], Signal Corps Radio Communication Pamphlet No. 40, published in 1919 and issued in a revised edition in 1922, is another interesting example of a popular radio book. Written by the Bureau of Standards for training Army Signal Corps officers, it was an expanded and updated version of Circular 74, noted above, but was also widely used by college students and others studying radio. In spite of the "pamphlet" designation, it was actually a 600-page book. Thomas Edison said ". . . This is the greatest book on this subject that I have ever read, and I want to congratulate you and your Bureau on its production."

While the general public was interested in radio information for the lay person, the fledgling radio industry needed in-depth technical assistance from the Bureau, which it got. The Bureau helped radio manufacturers make better measurements for quantities such as resistance, capacitance, voltage, and frequency. Results of Bureau research on antennas and radio wave propagation, direction finders, vacuum tube characteristics, and testing were eagerly utilized by major companies such as General Electric, Westinghouse, Western Electric, and RCA, as well as by a myriad of small radio companies no longer in existence. At the same time that the Bureau was providing information to people constructing radios on their kitchen tables, Radio Section staff members were presenting technical papers on radio at conferences and in journals such as the *Proceedings of the Institute of Radio Engineers.* A good example of the latter is the classic work by John Miller, who discovered and explained "the Miller effect"—a feedback effect in triode vacuum tubes that limited their ability to amplify at radio frequencies [9].

Circular 120 described in detail not only the crystal radio itself but also how to construct an antenna and ground system. To minimize the cost, the circular suggested winding the tuning coil on a discarded cylindrical oatmeal box. For years afterwards, home experimenters used oatmeal boxes as coil forms for homemade radios. The crystal radio did not require batteries, but it did require the user to purchase a commercially made set of headphones (or telephone receivers as they were called in those days), since that accessory was not suitable for home construction.

One can just imagine the thrill of a farm family that had never heard a radio broadcast before trying the new set for the first time, and one of the family members crying out, "Come listen! I am hearing something!" The electronics revolution was underway and the Bureau played a significant part in accelerating the new technology.

Prepared by Brian Belanger.

Bibliography

[1] *Construction and Operation of a Simple Homemade Radio Receiving Outfit,* Circular of the Bureau of Standards, No. 120, U.S. Government Printing Office, Washington, DC, April 24, 1922.

[2] *Construction and Operation of a Two-Circuit Radio Receiving Equipment with Crystal Detector,* Circular of the Bureau of Standards, No. 121, U.S. Government Printing Office, Washington, DC, July 17, 1922.

[3] *Sources of Elementary Radio Information,* Circular of the Bureau of Standards, No. 122, U.S. Government Printing Office, Washington, DC, June 22, 1922. (Updated to second edition, Sept. 12, 1923.)

[4] *Description and Operation of an Electron-tube Detector Unit for Simple Radio Receiving Outfits,* Circular of the Bureau of Standards, No. 133, U.S. Government Printing Office, Washington, DC, November 10, 1922.

[5] *Auxiliary Condensers and Loading Coil Used with Simple Homemade Radio Receiving Outfits,* Circular of the Bureau of Standards, No. 137, U.S. Government Printing Office, Washington, DC, Feb. 23, 1923.

[6] *Description and Operation of an Audio-Frequency Amplifier Unit for Simple Radio Receiving Outfits,* Circular of the Bureau of Standards, No. 141, U.S. Government Printing Office, Washington, DC, March 24, 1923.

[7] *Radio Instruments and Measurements,* Circular of the Bureau of Standards C74, U.S. Government Printing Office, Washington, DC, 1918. (Second edition published in 1924.)

[8] *The Principles Underlying Radio Communication,* Signal Corps Radio Communication Pamphlet No. 40, U.S. Government Printing Office, Washington, DC (1919). (Written by Bureau staff members.)

[9] W. F. Snyder and C. L. Bragaw, *Achievement in Radio: Seventy Years of Radio Science, Technology, Standards, and Measurement at the National Bureau of Standards,* NBS Special Publication 555, National Bureau of Standards, Boulder, CO (1986).

21 SIMPLE TRANSISTOR RADIOS YOU CAN BUILD

—From Crystal Sets to Superhets

By R. H. Warring

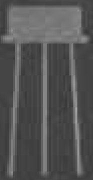

21 SIMPLE TRANSISTOR RADIOS YOU CAN BUILD

—From Crystal Sets to Superhets

FIRST EDITION

FIRST PRINTING—NOVEMBER 1975
SECOND PRINTING—JUNE 1977

Copyright © R.H. Warring, 1975

Printed in the United States
of America

21 Simple Transistor Radios You Can Build was originally published in Great Britain in 1975, as *Radio Construction for Amateurs*. Printed by permission of Sir Isaac Pitman & Sons Limited.

Reproduction or publication of the content in any manner, without express permission of the publisher, is prohibited. No liability is assumed with respect to the use of the information herein.

Hardbound Edition: International Standard Book No. 0-8306-5790-8

Paperbound Edition: International Standard Book No. 0-8306-4790-2

Library of Congress Card Number: 75-31457

CONTENTS

1 Radio made easy! 11
2 The language of radio 23
3 Crystal sets 27
4 More about tuned circuits 37
5 Amplifiers 47
6 The output stage 55
7 Transistors, bias and stabilization 61
8 TRF receivers 67
9 Regenerative and reflex receivers 71
10 Superhets 83
11 Components 93
12 Circuit construction 111
13 FETs and ICs 125
Appendix 133
Index 139

WORKING CIRCUIT DESIGNS

1 Simple crystal set 31
2 Crystal set using transistor 34
3 Transistor detector/amplifier 35
4 Double-diode crystal set 36
5 Crystal set with one stage of amplification 48
6 Crystal set with two-stage amplifier 49
7 Two-transistor amplifier for loudspeaker output 50
8 Practical two-stage amplifier 51
9 Two-stage amplifier with volume control 52
10 Push-pull output with direct coupling 56
11 Push-pull output with transformer coupling 57
12 Basic TRF receiver 68
13 Three-transistor TRF receiver 69
14 Single-transistor receiver with pre-amplifier 72
15 Single-transistor regenerative receiver 72
16 Simple regenerative receiver 73
17 Three-stage reflex receiver 76
18 Three-stage reflex receiver with choke-capacitance coupling 77
19 Single-transistor reflex receiver 78
20 Four-stage reflex receiver 79
21 'Mighty Atom' single-transistor receiver 81
22 Three-transistor regen receiver 86
23 Superhet receiver 89
24 Mullard design for six-transistor superhet 90
25 Mullard miniature superhet receiver 76
26 FET receiver 126
27 4-W IC amplifier 129
28 TRF broadcast receiver 129
29 Wireless microphone 131
30 Marine-band converter 131

PUBLISHER'S NOTE

Some of the circuits in this book have been adapted from H. Ness's *Handbook of Tested Transistor Circuits* (Babani Press), and the tables of transistor characteristics are based on data from Electrovalue Ltd., to which grateful acknowledgement is made.

1 RADIO MADE EASY!

Everyone knows what a radio is, but many people find it difficult to understand how a radio receiver works. Reading technical books on the subject is not always a help. The language of radio, and the understanding of technicalities, can prove difficult or even impossible without some previous background knowledge of the subject. So we will start right from scratch—in plain language!

The 'sending' of sound is easy to understand. When someone speaks he (or she) sends out a sort of pressure wave from the mouth. This travels through the air at approximately 335 metres per second, or 1,100 feet per second (the speed of sound). When a sound wave reaches the ear of a listener (anyone who happens to be in the way of it), the pressure wave will set parts in the ear vibrating, the 'signal' being transmitted to the brain and 'received' as an impression of sound—Fig. 1.1.

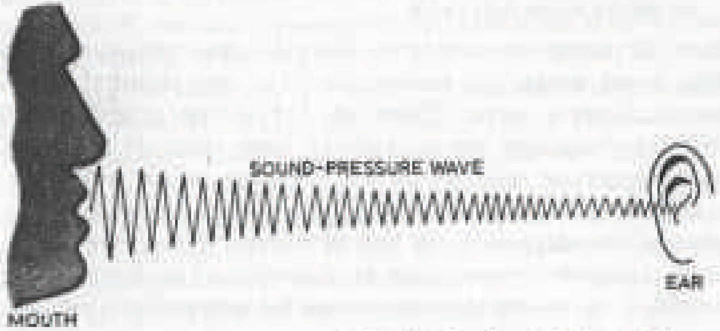

Fig. 1.1 Speech is transmitted directly through the air by sound-pressure waves.

Sound-pressure waves have another characteristic: their strength decreases quite rapidly with distance. Doubling the distance between 'sender' and 'receiver' would result in the level of sound received being reduced by one *quarter* (not one half).

The only effective way to increase 'distance of sending' or *range* is to increase the strength of the original sound, by shouting, for instance, although this will not greatly increase the range. (Remember that the strength of the sound-pressure waves decrease in proportion to distance *times* distance, or (distance)^2.) The other method is to increase the strength of the original sound by using a megaphone or public-address system. Both are devices which boost sound, known as *amplifiers*. A simple megaphone is a mechanical amplifier, while a public-address system is an electronic sound amplifier, based on a *microphone*, an *amplifier circuit*, and a *loudspeaker*–Fig. 1.2.

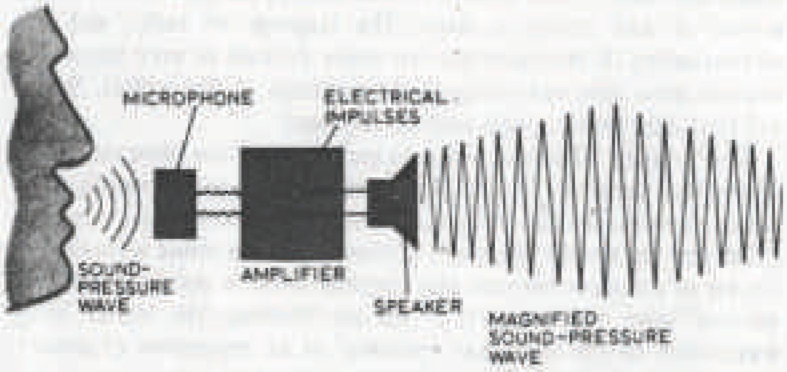

Fig. 1.2 To extend the range of sound transmitted, some form of amplifying device can be used.

Here the microphone acts rather like the human ear, being vibrated by the sound waves, but turning this effect into electrical impulses rather than nerve pulses. These are fed to the amplifier, which considerably magnifies the strength of these electrical signals. The stronger signals are then fed to a loudspeaker, which works like a microphone in reverse. It is fed with electrical impulses to set a diaphragm vibrating, which in turn generates a sound-wave 'output'. Because the incoming signal (into the microphone) has been magnified or amplified, the sound wave issuing from the loudspeaker is very much stronger than the original 'input' sound.

Such a system is still limited in range. Words spoken into a public-address system may be heard, under favourable conditions, as far away as, perhaps, one mile–but certainly not much further. Also, wind blowing in opposition to the sound will considerably reduce the range. So to increase range still further, some other method of transmitting the sound must be used–Fig. 1.3.

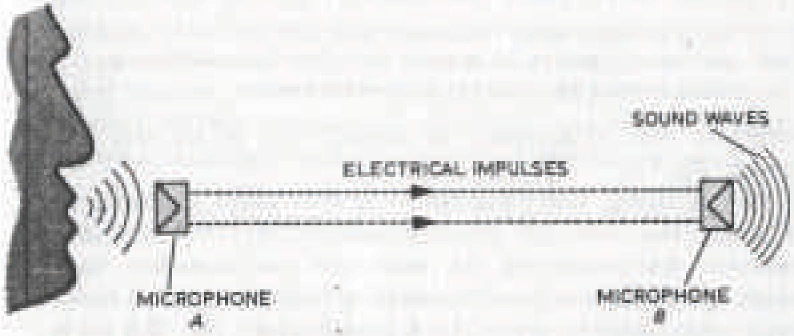

Fig. 1.3 To transmit sound over longer distance, wires are used to connect two microphones. In a simple system, the microphones can also work as speakers. In the usual telephone system, a separate microphone and speaker (earpiece) are used at each end.

Again the original sound is directed into a microphone, which is connected by two wires to another microphone. Each microphone can act either as a microphone or speaker. Microphone *A* turns sound into electrical impulses, which are transmitted along wires to microphone *B* which is vibrated to turn these impulses into the original sound; in other words, it is working as a speaker. If sound is now introduced into microphone *B*, electrical impulses will be transmitted along the wires back to *A*, which now acts as a speaker.

The particular advantage of a 'wire' system is that once the sound is turned into electrical impulses, these impulses can be transmitted over quite long distances without greatly diminishing in power. There will be a reduction in power with increasing length of wire, because all wires or conductors of electricity offer some *resistance* to the passage of electric currents. The greater the wire length, the greater the resistance, and thus the greater the loss of electrical energy. This will show up in the 'received' sound, being much weaker than the original.

There are ways of compensating for this; for example, the inclusion of a battery in the circuit, to produce a more 'responsive' microphone circuit (i.e. make the electrical impulses, generated by the microphone,

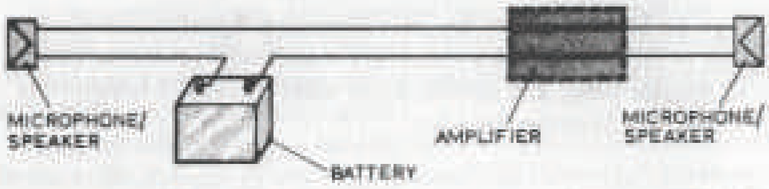

Fig. 1.4 Better sound transmission over wires is given by added power (via a battery). An amplifier can also be introduced into the circuit to boost the strength of the electrical impulses.

stronger). Also, if necessary, an amplifier can be included in the circuit – Fig. 1.4.

Before leaving this particular set-up it is worth explaining that, although two wires are shown connecting the two microphone/speakers, the system will also work with *one* wire. But, since a single-wire connection cannot complete a 'loop' for electrical currents, some other 'common' connection is needed at each end. This can be a connection to *earth*, and provides the 'return path' for the circuit – Fig. 1.5. There are disadvantages in such a system used for telephonic communication (which we have been describing), but the use of earth as a 'common' connection is very useful in radio work, as will be seen later. The earth may be 'real' – i.e. a connection to a physical earth; or 'virtual' – i.e. connection to a common line in the radio circuit with no external earth connection.

Telephonic transmission is possible over very long distances, but it does need interconnecting wires and one or more 'switching centres' in order to be able to connect one 'sender' to more than one 'receiver', or alternative 'receivers'.

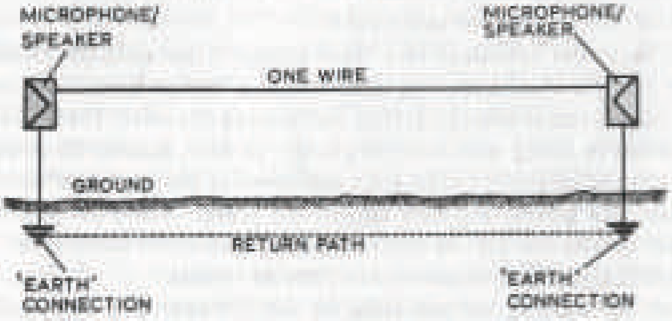

Fig. 1.5 Simple telephone systems can operate over a single wire with 'earth return'.

Radio is wire-less transmission (hence the original name, *wireless*): electrical impulses generated in a microphone are sent directly through the air – and *any number* of receivers can *tune in* to the same signal at the same time. The 'sender' becomes a *broadcast station*, each of the listeners using a radio receiver to hear the broadcast signals.

But electrical impulses, generated by sound waves reaching a microphone, cannot be sent directly. In the absence of connecting wires in the microphone, nothing will happen. The trick is to turn the electrical impulses generated in the microphone into a type of electrical signal which *will* travel through the atmosphere (and even into space); involving turning the electrical impulses into *radio-frequency waves*.

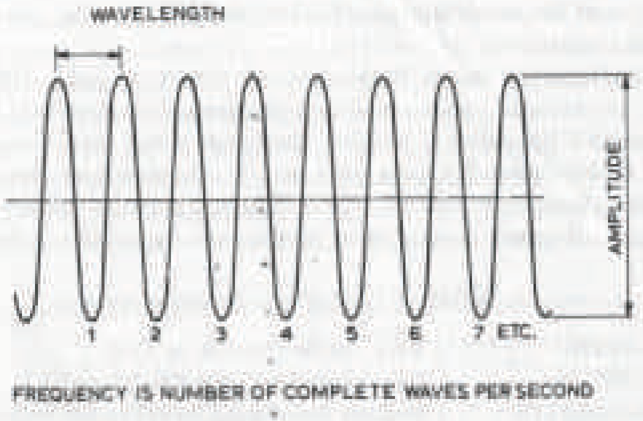

Fig. 1.6 A waveform is defined by its wavelength – or now, preferably frequency – and amplitude. Amplitude is a measure of the 'power' or strength of the wave.

A wave is simply a way of expressing, in physical form, a transmission involving energy. (A simple wave is shown in Fig. 1.6.) The characteristics of the wave are defined by its *wavelength*, or distance from crest to crest, and by its *amplitude* or 'swing' on either side of the centre-line along which the wave is moving.

Instead of wavelength, however, it is more convenient to speak of *frequency*, the number of complete waveforms produced per second. Wavelength, frequency and the *speed* of the wave are simply related as follows:

$$\text{wavelength} = \frac{\text{wave speed}}{\text{frequency}}$$

Wavelength (where used) is normally quoted in *metres*; frequency is now quoted in *Hertz* (abbreviated Hz).

All the sound and electrical waves we have considered so far have been at *audio* (or 'hearing') frequency (abbreviation AF). The original sound waves (which are an audio frequency) are converted directly into electrical waves, and back again into sound waves, all at the same frequency. The complete range of audio frequencies is from about 30 Hz (a very low, deep note) to around 16,000 Hz. The ear cannot respond to frequencies outside this range, so they are not heard. Neither, as we have noted, are any of these audio-frequency waves directly transmissible through the air over any distance, mainly because their *amplitude* decreases rapidly with increasing distance. *Amplitude* is a measure of the power or 'strength' of the wave (think of sea waves for a simple comparison).

Radio-frequency waves travel with the *speed of light*, 300,000 metres per second. Because of their high speed, their amplitude is far less modified by distance. Another distinction is that radio-frequency waves (abbreviation RF) are sent out in all directions from the source–a characteristic of all electro-magnetic radiation (unless constrained or 'beamed' in a position or direction by suitable reflecting devices).

The various radio waves are divided into categories, namely:

10–30 kHz	Very low frequencies, or VLF.
30–300 kHz	Low frequencies, or LF (long wave).
300–3,000 kHz	Medium frequencies, or MF (medium wave).
3–30 MHz	High frequencies, or HF (short wave).
30–300 MHz	Very high frequencies, or VHF.
300–3,000 MHz	Ultra-high frequencies, or UHF.
3,000–30,000 MHz	Super-high frequency, or SHF.

The LF range corresponds to the 'long wavelength' band; the MF range covers the 'medium wavelength'; and the HF covers the 'short waveband'.

An immediate problem arises. Radio waves (RF) cannot be heard, so how can they be made to transmit sound? The answer is surprisingly simple: the radio-broadcast station transmits an RF signal corresponding to the particular RF frequency allotted to it–Fig. 1.7. This is called the 'carrier', and since it will normally be well above 30,000 Hz it cannot be heard (except possibly some side-effects in the form of a little 'noise').

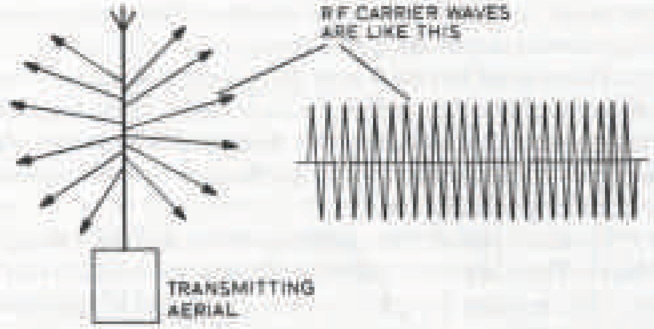

Fig. 1.7 Form of a steady RF carrier wave which is transmitted at a specific frequency.

To transmit sound–speech or music–a microphone is used to turn sound waves into audio-frequency (AF) electrical waves, just as in the previous systems described. Alternatively, these AF electrical waves can be obtained direct from a disc or tape recording, etc.

This 'sound' signal is superimposed on the carrier wave being transmitted. Instead of having a constant waveform (constant amplitude, as in Fig. 1.7), the combination of the AF ('sound' wave) with the carrier results in a wave something like that shown in Fig. 1.8. The 'top'

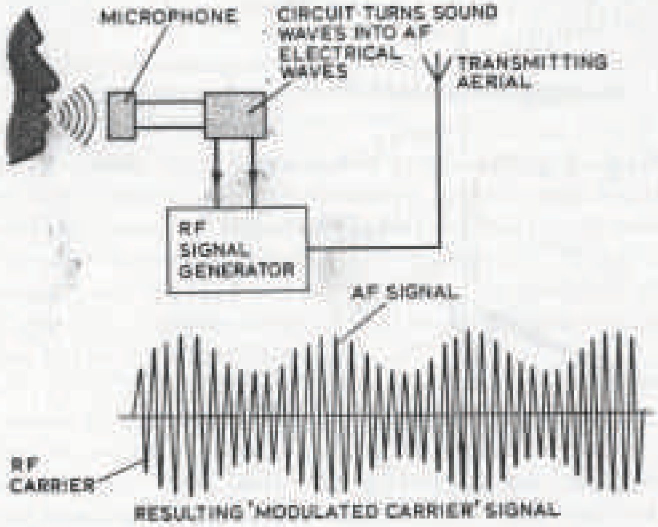
MICROPHONE
CIRCUIT TURNS SOUND WAVES INTO AF ELECTRICAL WAVES
TRANSMITTING AERIAL
RF SIGNAL GENERATOR
AF SIGNAL
RF CARRIER
RESULTING 'MODULATED CARRIER' SIGNAL

Fig. 1.8 The effect of superimposing an audio-frequency (AF) signal on the RF carrier is a modulated carrier-wave transmission.

and 'bottom' of the carrier wave now assume the form of the 'sound' wave, which is thus carried along with the carrier.

This combination is known as a *modulated wave*. The 'sound' or AF wave has modified the form of the RF carrier wave, or modulated it. It is also easy to see that what has actually been modified or modulated is the *amplitude* of the carrier wave; so this form of radio transmission is called *amplitude-modulation* or AM.

This is the only type of radio transmission we shall be dealing with in this book. There are others, notably *frequency modulation* or FM, used for VHF working (which is outside the scope of reception by simple radio receivers).

No problems, therefore, in getting AF signals 'on the air'. In fact, the air is literally full of such signals almost every hour of the day, radiated from different broadcast-transmitting stations.

The next step is to be able to pick them out of the air, so to speak; and, in particular, to pick out or *tune in* to the one particular station we are trying to listen to.

A single length of wire will 'pick up' virtually all AM signals present. This is because each signal represents varying amounts of electrical energy passing across the wire when, by a phenomenon known as

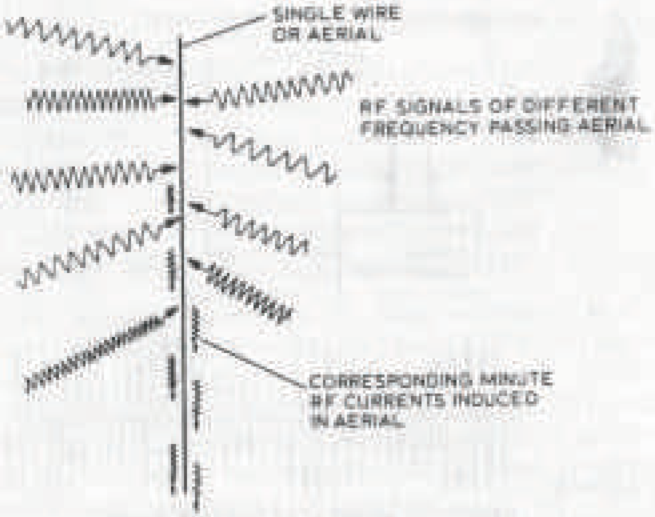
SINGLE WIRE OR AERIAL
RF SIGNALS OF DIFFERENT FREQUENCY PASSING AERIAL
CORRESPONDING MINUTE RF CURRENTS INDUCED IN AERIAL

Fig. 1.9 A simple wire aerial is a very weak 'receiver' of radio signals (by a process known as induction).

induction, a part of that varying energy will appear in the wire as an *induced current*. This current will vary with the same *frequency*, and duplicate the same *modulation*, as the original signals—Fig. 1.9.

The single length of wire, or *aerial* as it can now be called, is thus probably being affected by dozens of different RF signals, each inducing its own 'current image' in the wire; all these currents will be extremely minute—far too small to measure. So the aerial is only the starting point for a practical receiver; another example of 'electronic trickery' is required, namely a *tuned circuit* connected to the bottom of the aerial—Fig. 1.10.

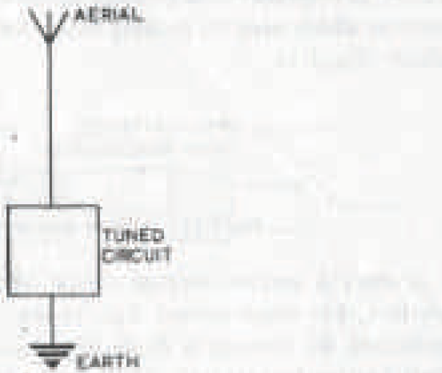

Fig. 1.10 The addition of a tuned circuit to an aerial produces amplification of signals induced in the aerial at one frequency only (the resonant frequency of the tuned circuit).

A tuned circuit normally comprises a *coil* (which is far more efficient than a straight wire for picking up induced currents), and a *capacitor* (which acts as a stop or a break in a circuit for direct current, d.c., but passes alternating current (a.c).

A capacitor also has the important property, when combined with a coil, of generating a greatly magnified induced current at the particular frequency known as *resonant frequency*.

This is a very important factor in the behaviour of a *tuned circuit*. Radio signals at that frequency, present in the aerial, will be greatly magnified or amplified, while other signals at different radio frequencies will not. Thus, apart from picking up and magnifying a particular broadcast frequency, the tuned circuit virtually rejects all the other signal currents present in the aerial because they are so minute.

Since the resonant frequency of the tuned circuit depends on the value of the capacitor and coil *inductance*, a suitable combination of values can be calculated to give maximum response (i.e. 'resonate') at any particular broadcast frequency. However, this would largely restrict the scope of the receiver; therefore, one of the tuned-circuit components is made *variable* in value, which means that the tuned circuit can now be adjusted to *tune in* to a whole range of separate broadcast frequencies.

One more step is needed to complete a basic receiver. The tuned circuit, when adjusted for resonance, is picking up the original carrier plus sound signals, or modulated RF signals. As this signal cannot be fed directly to a speaker (not being at audio frequency) it is fed through a *detector*, which need be nothing more than a simple component called a *diode*–Fig. 1.11.

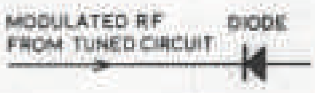

Fig. 1.11 A diode acts as a detector.

A *diode* is another stop-go device, which passes a.c. or d.c. in one direction, but stops current flow in the other. The effect of passing modulated RF through a diode is to 'chop off' the 'bottom' of the signal, leaving two separate components; one being the 'top half' of the original carrier, the other a varying d.c. component, following exactly the variations of the original AF signal used to modulate the carrier–Fig. 1.12.

The output from a diode detector can, in fact, be fed direct to a suitable speaker, which will respond to the varying d.c. signal, turning this into audible sound in a similar way as telephonic transmission–Fig. 1.13. The RF signal 'top half' present will not affect the speaker,

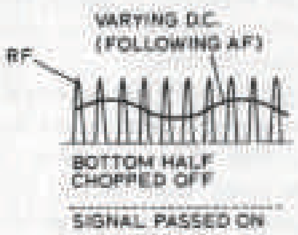

Fig. 1.12 The form of the signal passed by a diode detector–a mixture of 'chopped' RF and a d.c. component varying in AF

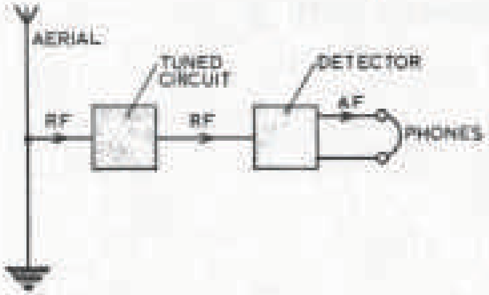

Fig. 1.13 The basic stages which make up a simple receiver.

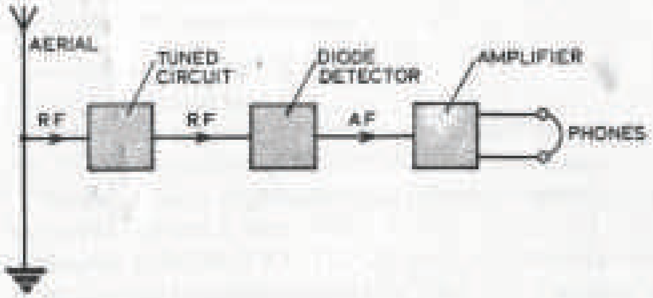

Fig. 1.14 More output power can be obtained from a receiver by interposing one (or more) stage(s) of amplification after detection.

although it may be desirable to introduce some further components to smooth out possible interference with the sound signal required.

Such a basic circuit relies only on the *tuned circuit* for the amount of amplification produced–and thus the volume of sound heard in the speaker. Surprisingly, this can be all that is necessary to hear a limited number of stations quite clearly on a simple crystal set, using an earpiece as a speaker. To produce better volume, however, it is a relatively straightforward matter to introduce one or more stages of *amplification* after the detector–Fig. 1.14.

2 THE LANGUAGE OF RADIO

The following abbreviations are normally used to designate *components*, particularly on circuit drawings.

A Aerial; or current value in amps.
E Earth; or potential (e.m.f.).
R Resistor (the different resistors in a circuit being numbered R1, R2, R3 etc., for individual identification); or resistance.
VR Variable resistor or potentiometer (although sometimes the 'V' is dropped and just R used).
C Capacitor (numbered C1, C2, C3 etc., in a circuit). Variable capacitors are normally designated C, *not* VC. Also symbol for capacitance.
CT Trimmer capacitor.
D Diode.
L Coil (or inductance).
RFC Radio-frequency choke.
T Transformer.
TR Transistor, although there are alternatives used, e.g. Tr, VT and T; also sometimes Q. If T is used for transistors, then TR is used for transformers.
FET Field-effect transistor.
I Current.
IC Integrated circuit (module).
PC Printed circuit or printed-circuit board (PC board).
S or SW Switch.
J Jack.
V Voltage or e.m.f.
Z Impedance.

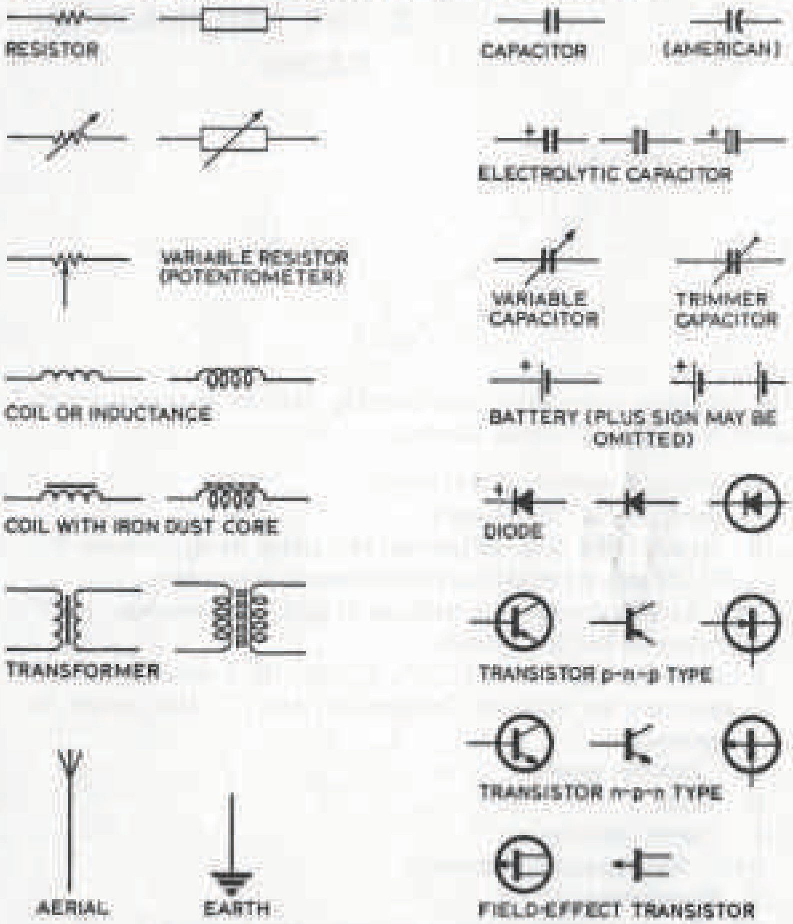
RESISTOR
CAPACITOR
(AMERICAN)
ELECTROLYTIC CAPACITOR
VARIABLE RESISTOR (POTENTIOMETER)
VARIABLE CAPACITOR
TRIMMER CAPACITOR
COIL OR INDUCTANCE
BATTERY (PLUS SIGN MAY BE OMITTED)
COIL WITH IRON DUST CORE
DIODE
TRANSFORMER
TRANSISTOR p-n-p TYPE
TRANSISTOR n-p-n TYPE
AERIAL
EARTH
FIELD-EFFECT TRANSISTOR

Fig. 2.1 Standard symbols used in circuit diagrams with alternatives in most cases. These cover the symbols most likely to be found on British, European and American circuit diagrams.

Components on circuit diagrams are also designated, and easily identified, by symbols, and here again there are some variations to be found—see Fig. 2.1.

Basic quantities are measured in units as follows:

Resistance	in ohms (Ω).
Capacitance	in farads (F).
Inductance	in henrys (H).

Impedance	in ohms.
Potential difference or e.m.f.	in volts (V).
Current	in amperes (amps).
Power	in watts (W).

Numerical values of these units are often too large or too small to be expressed conveniently, when the following *prefixes* are used (representing multiples or sub-multiples).

Prefix	Symbol	Factor by which unit is multiplied	Example
mega	M	1,000,000	10 MHz (10 megahertz) = 10,000,000 Hz
kilo	k	1,000	10 kΩ = 10,000 ohms
milli	m	1/1000 or 0.001	2 mA (2 milliamps) = 0.002 amps
micro	μ	1/1,000,000 or 0.000001	6 μV (6 microvolts) = 0.000006 volts
nano	n	0.000000001	not widely used as a prefix in basic electronics
pico	p	0.000000000001	220 pF (220 picofarads) = 0.000000000220 farads

*Note: the unit name is often omitted when the inference is obvious. Thus for a 10-kΩ resistor, the value would be quoted as 10 k.

As a simple guide, mega (M) and kilo (k) are most usually found describing high values of *radio frequency* and *resistors*.

Milli (m), and to a lesser extent micro (μ), are most commonly associated with very low values of *current* or *voltage*, and practical values of *inductances*.

Practical values of *capacitors* are invariably quoted in microfarads (μF) or picofarads (pF).

Note: components are described in detail in Chapter 11.

3 CRYSTAL SETS

The simplest type of radio receiver consists of an aerial, a tuned circuit and a detector–plus headphones to listen to the signals received by the tuned circuit. It is generally called a *crystal set* because it works on the same principle as the very earliest radio sets, using a lead-galena crystal and a 'cat's whisker' as a detector. The only real difference is that the crystal and cat's whisker have been replaced by the much more efficient point-contact *diode*.

A simple *tuned circuit* can be made by winding a coil of wire of specific diameter and length (number of turns) to match the capacitance range of a variable capacitor. The combination of coil (inductance) and capacitance can then be adjusted by turning the variable capacitor spindle to be resonant, or tune in to various broadcast frequencies. The signal which is tuned in is then passed to the diode which blocks the RF content, and rectifies and passes on the AF content to the headphones, which transform these signals into audible sounds.

The tuned circuit can be combined with the aerial by winding the coil on a *ferrite rod*. This increases the inductance of the coil (reducing the length of wire required), whilst the rod itself also acts as an aerial to pick up the radio signals. This is the type of aerial employed in most domestic radio receivers. However, with very simple sets working with very weak signals, an external aerial may also be required.

The performance of all simple, low-powered receivers is, in fact, very much dependent on the efficiency of the aerial and tuned circuit, which can be regarded as the critical part of the system. Performance can perhaps be improved by experimenting with different aerial/tuned-circuit systems, and by the addition of further circuitry, as explained in later chapters.

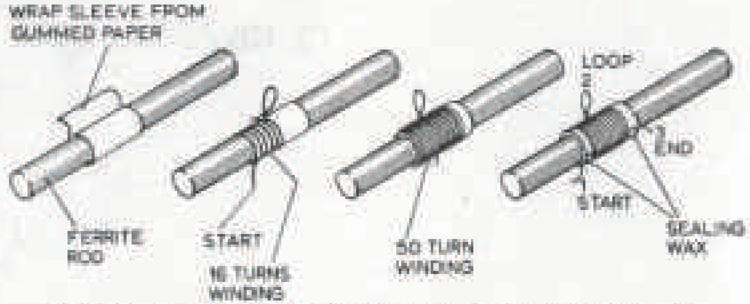

Fig. 3.1 Aerial tuning coil wound from 38-s.w.g. wire on a $\frac{3}{8}$-in. diameter ferrite rod.

Fig. 3.1 details the construction of a suitable coil for a tuned circuit, matched to a ferrite rod of $\frac{3}{8}$ in. diameter and about 4 in. long (see Chapter 4 for alternative coil windings on different rod sizes). If a rod is purchased longer than 4 in., it can be used as it is, or cut down by marking around with a file and then breaking off the surplus length.

Cut seven 1-in. lengths of gumstrip. Moisten one and wrap around the rod gummed side up. Now add about another half a dozen wrappings of similar length over the first, this time with gummed side *down* to form a reasonably rigid tube. Make sure that the paper tube is a *sliding* fit on the ferrite rod and leave to dry thoroughly (preferably removed from the rod so that it cannot become stuck to it).

When the paper tube is quite dry it should be rigid, when the coil windings can be applied. The wire to be used is 38-s.w.g. enamelled copper wire, the number 38 referring to the actual diameter of the wire according to the standard wire gauge (s.w.g.).*

Starting about $\frac{1}{4}$ in. in from one end of the paper tube, wind the wire carefully round the tube, with each turn tight against the one before it, until sixteen full turns have been completed. Then make a loop in the wire, as shown, and carry on winding, with succeeding turns touching, until fifty turns in all have been completed. The two loose ends of the coil (the start and finish) can be secured with a dab of sealing wax whilst the projecting loop can be twisted together (e.g. by putting a pencil through the loop and twisting up). Cut off the loop, leaving about $\frac{1}{2}$ in. protruding from the main coil, bare the wire ends and solder together. This forms point 2 on the coil; the start is point 1, and the end point 3—see Fig. 3.1. It will be easy to remember these

---

* American wire gauge (AWG) is about two sizes smaller than the British standard wire gauge (s.w.g.). For example, 38 s.w.g. is approximately the same as 36 AWG.

without marking since the loop or tapping point (2) comes much closer one end (1) than the other (3).

Cut a panel of *Paxolin sheet* to about the size shown in Fig. 3.2, using a hacksaw.* On this secure a *tag strip*, as shown, and drill a hole to mount a miniature or small-size 500 pF *variable capacitor*.

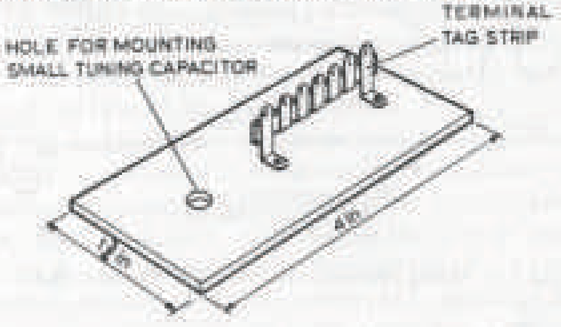

Fig. 3.2 Paxolin panel and tag strip for crystal set circuits.

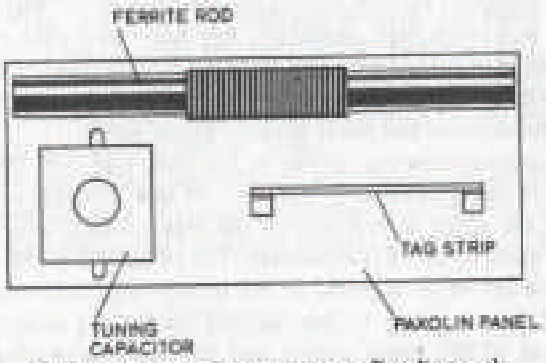

Fig. 3.3 Layout of components on Paxolin panel.

The aerial coil is then mounted on the panel as shown in Fig. 3.3, gluing the coil on to the Paxolin with two or three dabs of sealing wax, or some other suitable adhesive. *Note:* the ferrite rod must be free to slide in the paper tube for 'tuning' adjustments.

Virtually any miniature germanium or silicon diode will be suitable for the detector. Recommended types, which are readily available, are 1N34 and 1N914.

Earphones must be of *high-impedance* type, which need not be expensive to buy, and the higher the impedance the better the reception.

---

* A phenolic or fiberglass sheet may also be used.

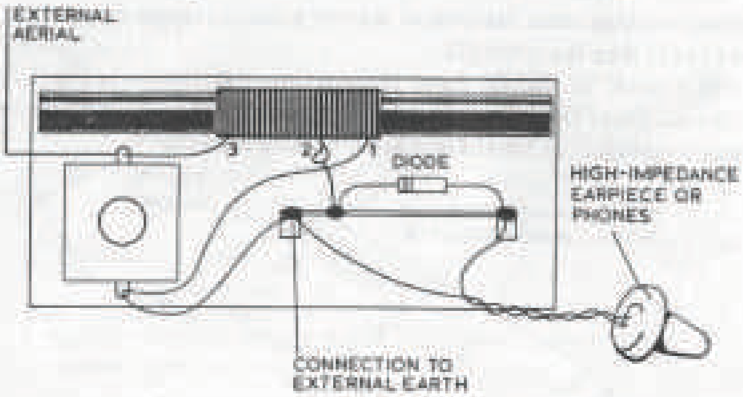
EXTERNAL AERIAL
3
2
1
DIODE
HIGH-IMPEDANCE EARPIECE OR PHONES
CONNECTION TO EXTERNAL EARTH

Fig. 3.4 Wiring connections to complete the crystal set.

Alternatively a deaf-aid type earpiece can be used; this will not give the same volume or quality of reproduction as headphones, but is a less expensive component. This should preferably be of the *high-impedance magnetic* type, with high sensitivity. Any high-impedance earpiece will suffice, but if of crystal type will require a resistor, connected across it, to complete the circuit. This will reduce the amount of current flowing through the earpiece and lower the strength of signal.

Wiring connections are shown in Fig. 3.4. End 3 of the aerial coil (the end of the 50-turn coil) connects to one terminal of the tuning capacitor, the aerial or 'hot' end of the tuned circuit, and the point to which an external aerial is connected. The other end of the coil (end 1) connects to the other terminal of the tuning capacitor, from which an additional wire is taken to the first tag on the tag strip. This is the 'earthy' end of the tuned circuit, and the point to which an external earth is connected. Leave plenty of slack wire between the coil and tuning capacitor.

The other connections are then as follows:

(i) Tapping point of the coil wire bared and connected to the second tag.
(ii) The diode also soldered to this same tag, and to any other free tag.
(iii) Headphone (or earpiece) leads to the 'earthing' tag, and to the 'free' tag to which the above diode has been connected. All connections should be made with *soldered* joints.

The wiring-up can be checked against the *circuit diagram* shown in Fig. 3.5 (ignoring the components shown with broken lines). The set should now be 'working'.

In areas of strong signal strength, no external aerial or earth connections should be necessary. Performance will, however, be improved in any area by attaching an aerial wire (which can be any thin wire, e.g. using the same wire as for the coil winding), of up to 160 feet. The longer the aerial the better the reception, provided it is led away from the receiver to as *high* a point as possible.

An earth connection may further improve aerial performance; by this we mean a connection to some conductor positively in contact with the ground (preferably buried), an excellent example being a metal water-pipe. Thus, if an earth connection is found to be necessary (or you want to try one to see how performance is affected), connect a wire from the 'earthing' tag on the receiver to a convenient water-pipe.

This question of obtaining a *good* aerial and earth is a most important one in areas of poor signal strength. Linking up to a television aerial is often a good plan, since TV aerials are also usually mounted as high as possible. If bare wire is used, it is also important that the upper (free) end of the aerial is not made fast to something which could produce an earth connection (e.g. a damp tree), or at least is suitably insulated from such a support. String is not an efficient insulator; that, too, can conduct when wet.

Quite good results are often obtained by using the springs of a bed as an aerial, in which case an earth connection is usually not necessary. Sometimes, too, when other attempts to yield a good signal strength in

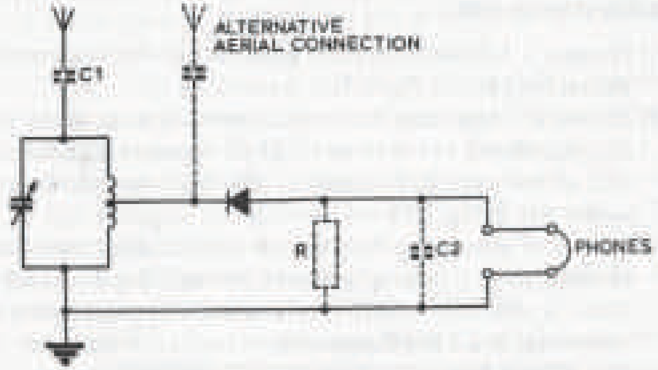

Fig. 3.5 *Circuit 1.* Circuit diagram of basic crystal set.

the aerial have failed, connecting the *aerial* side of the tuning coil to a *good earth* (a water-pipe) can produce better results, the normal earth connection being left off.

Tuning

The receiver is adjusted as follows. Turn the tuning capacitor to fully close the vanes, then open about half a turn on the spindle or knob. (If you are using a trimmer as a tuning capacitor, screw right down and then open half a turn.)

The tuning coil should now be slid up and down the ferrite rod (the coil leads were left a fairly long time to give the necessary freedom of movement) until BBC Radio 3 is heard.*

It may be necessary to slightly alter the adjustment of the variable condenser to tune in to this programme. Also, because of the ferrite-rod aerial the set will be directional, that is, the signal strength received will depend to some extent on the direction in which the aerial rod is pointing, so position the set to pick up the maximum volume.

Having established the best position of the tuning coil on the rod to receive BBC Radio 3, fix permanently with a dab of sealing wax. You should then find it possible to tune in to further stations by altering the setting of the variable capacitor–e.g. typically Radio 4 in about the middle of the capacitor travel and Radio 1 towards the other end.

Any reception you get will almost certainly be very weak and (unless you live close to a broadcast station) you can really feel satisfied if you get any station at all at audible strength. But it is surprising how, sometimes, even quite distant stations can be heard. Also you can often improve the reception and listening strength by quite simple modifications. Try these in order:

(i) Connect a 1,000-pF capacitor across the headphone (earpiece) connections (C2 in Fig. 3.5).

(ii) Instead of connecting the external aerial directly to the tuned circuit, connect one lead to a 220-pF capacitor, and the other end of the capacitor circuit to the 'hot' end of the tuned circuit (C1 in Fig. 3.5).

(iii) Instead of connecting the external aerial to the tuned circuit, connect to the tapping point of the coil (tag to which the diode is also connected). Try a direct connection, and also connecting in a 220-pF capacitor.

---

* Tune to an AM broadcast station near the lower end of the dial—close to 550 kHz.

(iv) Try connecting a 1.2-k resistor (or higher value) across the phone connection (R in Fig. 3.5). You may be using the wrong type of phones or earpiece, which do not provide a proper load or complete the circuit.

If there is a complete lack of response, check for faulty wiring-up. A more likely cause, however, is lack of an external aerial or earth connection in an area where these are strictly necessary for adequate reception; or an inefficient aerial (too short) or poor earth connection (bad electrical contact to a good earth point, or connection to a bad earth point).

Another possible cause of apparent failure may be too much outside noise entering the ear so that it is impossible to detect the very weak radio signal as it is being tuned in. Headphones are better than a single deaf-aid type of earpiece in this respect but, in any case, a really quiet room is virtually essential for initial setting up and tuning adjustments. Also, if your adjustment of the tuning control is too coarse, you may completely miss the setting for the station you are looking for, without realizing it.

Reception will also tend to vary with weather conditions. Some days it may be so poor that what was normally a strong station is hardly heard at all. The simple basic receiver has many limitations but, since it costs very little to construct and nothing at all to operate, this must be regarded as inevitable.

Providing you can hear something–even if too weak a signal to distinguish properly–you can certainly improve the performance of your basic set by further experimentation with tuned circuits (see Chapter 4) and/or the addition of amplification to the circuit. You can also try other types of basic crystal set, as described in the following projects.

Circuit 2 (Fig. 3.6) is identical to *Circuit 1* except that, instead of a diode, a transistor is used as a detector. Only two of the transistor leads are connected–the emitter (*e*) connection to the tapping part of the coil, and the base (*b*) to the 'earthy' end of the circuit. The collector lead of the transistor is ignored (bend it out of the way so that it cannot accidentally short out the other leads).

You can try almost any type of low-cost AF transistor; recommended types are OC42 2N370, 2N2925, 2N5088.

With the addition of two more components, Circuit 2 can be modified to work the transistor both as a detector and an amplifier, to give stronger signals through the headphones. Using the same transistor type

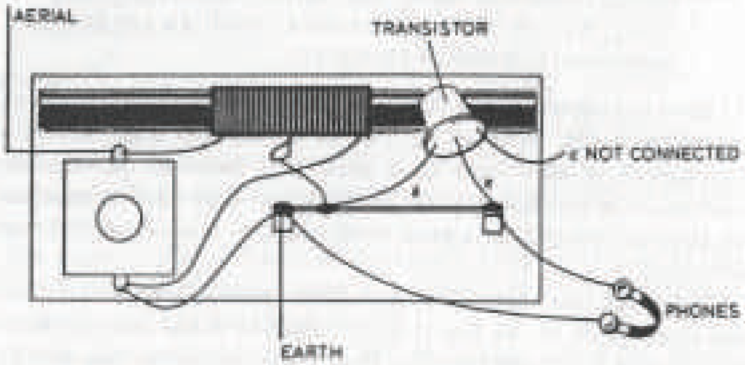
AERIAL
TRANSISTOR
e NOT CONNECTED
b
c
PHONES
EARTH

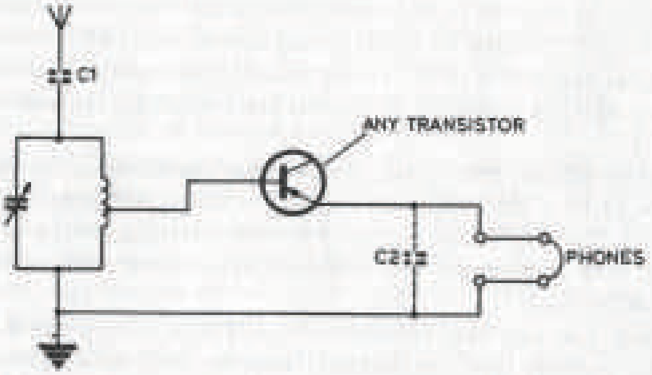
C1
ANY TRANSISTOR
C2
PHONES

Fig. 3.6 *Circuit 2.* Crystal set using transistor instead of a diode.

as above (or near-equivalent) resistor R should be 15 k and capacitor C3 1 μF or higher.

This time a battery is also required to supply power for the transistor to work as an amplifier. This can be from 1.5 V up to 9 V. Remember the rules for polarity of connection: those shown in Fig. 3.7 are for a p-n-p transistor; an n-p-n transistor would need the battery connected the opposite way round. Battery polarity also affects the connections of capacitor C3 (if an electrolytic or polarized type).

Experiment further by trying the effect of using additional capacitors in the circuit, e.g. C1, 220 pF; C2, 0.001 μF (try other values as well); C4, 0.001 μF (try other values as well).

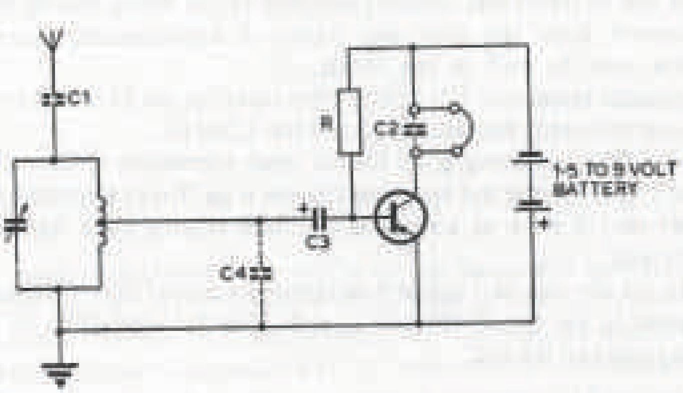
C1
R
C2
1·5 TO 9 VOLT BATTERY
C3
C4

Fig. 3.7 *Circuit 3.* Crystal set with amplification.

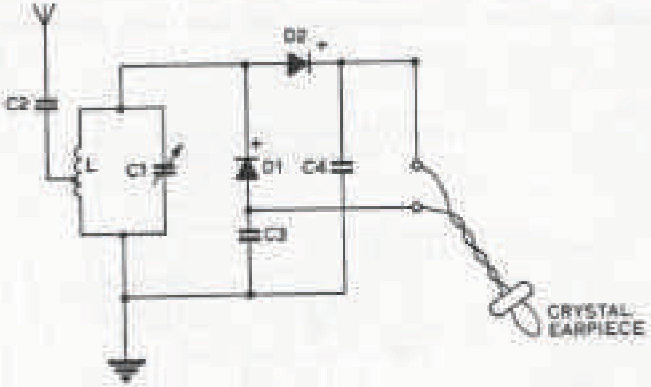
D2
C2
L
C1
D1
C4
C3
CRYSTAL EARPIECE

Fig. 3.8 *Circuit 4.* Double diode crystal set.

This circuit incorporates 'voltage doubling' to improve the signal volume and should give better performance than a single diode circuit. L and C1 are the usual tuned circuit, but an unusual feature is that the *aerial* is connected to the coil tapping point. Any type of germanium diodes can be used (they should preferably be the same), making sure to connect them the right way round. A high-impedance *crystal* earpiece must be used in this circuit.

Capacitor values are: C2, 220 pF (this capacitor can be omitted–try with and without in the circuit); C3 and C4, 1,000 pF.

The optimum tapping point for the aerial connection to the coil L is best found by trial and error (see Chapter 4 on 'Tuned Circuits'), but the set should work at a nominal one-third tapping point from the 'earthy' end.

This set can also be tried with conventional tuned-circuit coupling–i.e. aerial to the top of the coil L, and diode D2 connection to the tapping point on the coil.

4 MORE ABOUT TUNED CIRCUITS

The basic combination of a coil of certain *inductance* connected in parallel with a *variable capacitor* is the form of tuned circuit used in most radio receivers, from simple crystal sets upwards. The *tuning range* of such a circuit is determined by the respective values of inductance and capacitance used. At the same time, the *resistance* of the coil can have a significant effect on the *selectivity* of the circuit, so that a particular broadcast frequency is tuned into sharply, with other nearby frequencies rejected.

In the simple tuned circuit used for the crystal set of Chapter 3, the aerial is connected to the upper or 'hot' end of the coil, the other end of the coil being earthed; the diode is connected to a tapping point on the coil. Performance may well be improved by adjusting both the aerial-connecting point and the diode-connecting point, as shown in Fig. 4.1.

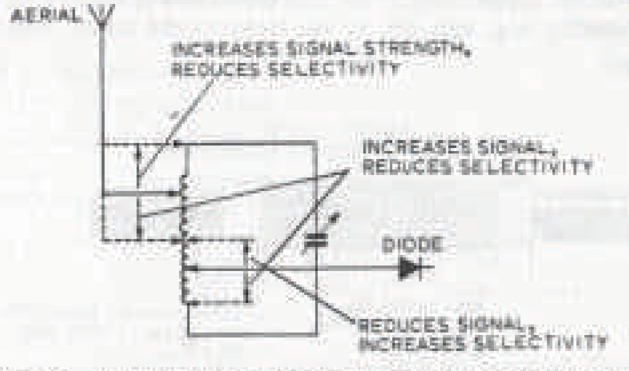

Fig. 4.1 The effect of varying the tapping points on the aerial coil.

Connecting the aerial to the 'hot' end should give the loudest signal, but the *selectivity* of the circuit may be poor because of the coil *resistance* involved in the actual circuit to earth. Connecting the aerial to a tapping point lower down the coil (towards the 'earthy' end) should improve selectivity, but reduce signal strength. Experimentation will produce the optimum aerial-connecting point. There will also be an optimum tapping point for the diode.

Finding optimum tapping points for the aerial and diode connections is a bit tricky on a close-wound coil since each tapping point tried has to be bared by scraping off the enamel insulation, with the risk of producing shorted turns. The same *effect* can be provided by removing, or adding, turns at each end of the coil, using the original tapping point for the diode connection–e.g. see Fig. 4.2. Unless the *total* number of turns on the coil remain the same, however, the inductance of the complete coil will be altered, and thus the tuning range of the tuned circuit (see later). Then the *diode* tapping point will be shifted.

For this type of experiment an air-cored coil of relatively large diameter is easier to use, wound on a card or paper tube to which two $\frac{1}{4}$ in.^{2} lengths of hardwood or balsa strip have been cemented. The windings can be bared by scraping or sandpapering over this raised section, and individual wires carefully separated to avoid shorting–Fig. 4.3. Optimum tapping points can be established by pushing bared ends on the aerial- and diode-connecting wires into the raised section of the coil, soldering in place very carefully once they have been found. This should not result in shorting out more than a few turns at these points; which will not greatly modify the coil inductance if allowed for in the initial winding (e.g. add five or six turns to the design number of windings).

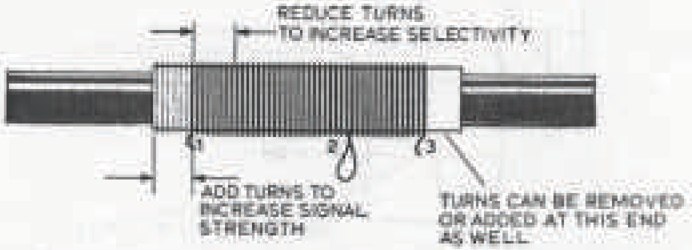
REDUCE TURNS
TO INCREASE SELECTIVITY
1
2
3
ADD TURNS TO INCREASE SIGNAL STRENGTH
TURNS CAN BE REMOVED OR ADDED AT THIS END AS WELL

Fig. 4.2 The effective tapping point on a simple coil can also be varied by adding or removing turns on each end of the winding.

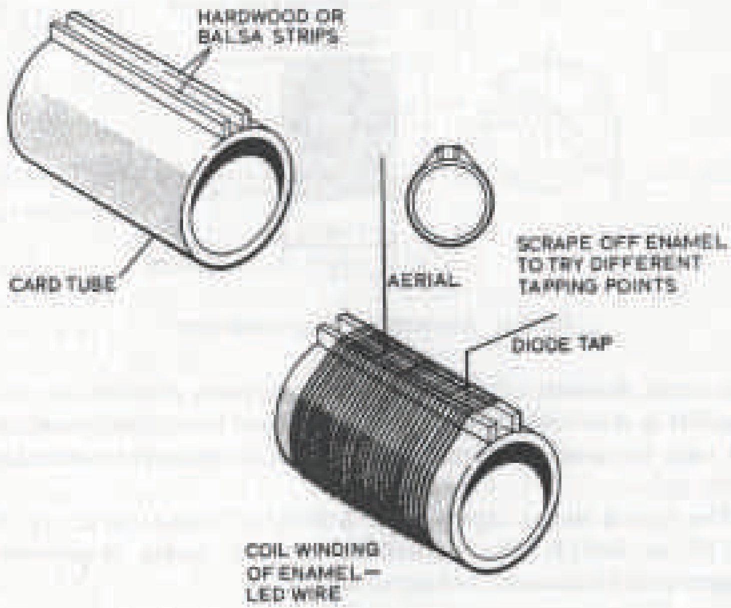
HARDWOOD OR BALSA STRIPS
CARD TUBE
AERIAL
SCRAPE OFF ENAMEL TO TRY DIFFERENT TAPPING POINTS
DIODE TAP
COIL WINDING OF ENAMEL-LED WIRE

Fig. 4.3 Construction of an air-cored aerial coil.

In air-cored coils, the inductance can be calculated with reasonable accuracy from the diameter and number of turns (see also Fig. 4.4):

$$\text{L, inductance, microhenries} = \frac{R^2 \times N^2}{9R + 10l}$$

where R = coil radius in inches,
l = coil length in inches,
N = number of turns.

The more usual form of this formula is:

$$N = \sqrt{\left(\frac{(9R + 10l) \times L}{R^2}\right)}$$

(A worked-out example using this formula is given later.) This is based on the turns of the coil being close-wound, in enamelled copper wire.

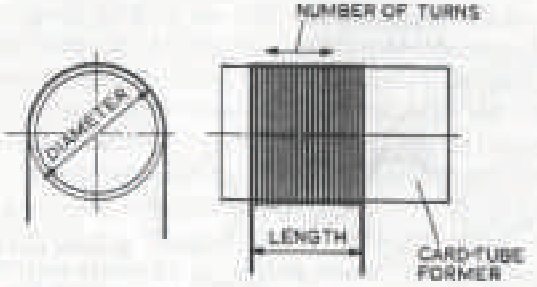

Fig. 4.4 Air-cored coil design parameters.

The actual diameter of the wire is not significant, provided the coil diameter is reasonably large (e.g. greater than 1 inch). The formula is not valid for smaller-diameter air-cored coils, or for coils wound on a ferrite rod.

The typical tuning capacitor has a rating of (about) 50 pF up to 350 pF or 500 pF. The relationship between tuning or *resonant frequency* and inductance and capacitance is:

$$\text{resonant frequency } (f) = \frac{1}{6.28\sqrt{(LC)}} \times 10^6 = \frac{0.16}{\sqrt{(LC)}} \times 10^6$$

where L = inductance in microhenries (μH),
C = capacitance in picofarads (pF).

This is more conveniently written as a solution for the *product* of inductance and capacitance required, i.e.

$$LC = \frac{0.025}{f^2} \times 10^{12}$$

For the medium waveband the range of frequencies to be covered is 500 to 1,500 kHz.* Thus, to encompass this range, the products of LC required are:

at 500 kHz

$$LC = \frac{0.025}{(500)^2} \times 10^{12}$$

$$= 0.001 \times 10^8 \text{ (approx.)}$$

---

* In the U.S.A. the AM broadcast band is from 540 to 1600 kHz.

at 1,500 kHz

$$LC = \frac{0.025}{(1{,}500)^2} \times 10^{12}$$

$$= 0.0001 \times 10^8 \text{ (approx.)}$$

At one end of the capacitor tuning range where $C = 50$ pF, inductance L required is:

$$\frac{0.001 \times 10^8}{\cancel{50}\ 500}$$

$$= 200\ \mu\text{H}$$

At the other end of the capacitor tuning range where $C = 500$ pF, inductance L required is:

$$\frac{0.0001 \times 10^8}{\cancel{500}\ 50}$$

$$= 200\ \mu\text{H}$$

Using a 1-in. coil diameter (0.5-in. radius) and a coil length of 1.5 in.

$$N = \sqrt{\left(\frac{(4.5 + 15) \times 200}{0.5^2}\right)}$$

$$= \sqrt{15{,}600}$$

$$= 125$$

In other words, a 125-turn air-cored coil of this size should be about right for covering the medium waveband, matching a 50–500 pF variable capacitor. Tapping point (or diode connection) would nominally be one-third of the coil length from the 'earthy' end.

Air-cored coils can also be designed in this way for long-wave coverage (more turns); or short-wave reception (less turns), which is not likely to be so good because of the low efficiency of the coils. Also, at higher frequencies the resistance of the *capacitor* is also significant, and so simple calculation of coil size is no longer valid.

The efficiency of a tuning coil can be expressed in terms of the magnification of the original signal received, when tuned to resonance, and decreases with increasing coil resistance. The magnification produced in a resonant circuit is referred to as the 'Q'; a high-Q coil is thus desirable for maximum performance and selectivity, and is essentially a low-resistance one. For this reason larger diameter air-cored coils, using thicker wire, are more effective than smaller diameter coils wound from thinner wire (since wire resistance increases with decreasing diameter).

Equally, the inductance of a wound coil can be increased by winding it over a magnetic-iron core, such as a ferrite rod. Much smaller diameter coils can be used, requiring less turns, to produce the inductance required—both reducing the length of wire and its resistance.

Because the actual inductance produced is dependent on the size and characteristics of the core material, windings for ferrite-rod aerials cannot be calculated, and are based on empirical results, or optimum windings established by trial and error.

Common sizes of ferrite rod available are:

¼-in. diameter	length 3½ in.
5/16-in. diameter	lengths 4, 5, 6 and 8 in.
⅜-in. diameter	lengths 4, 5, 6, 8 and 10 in.

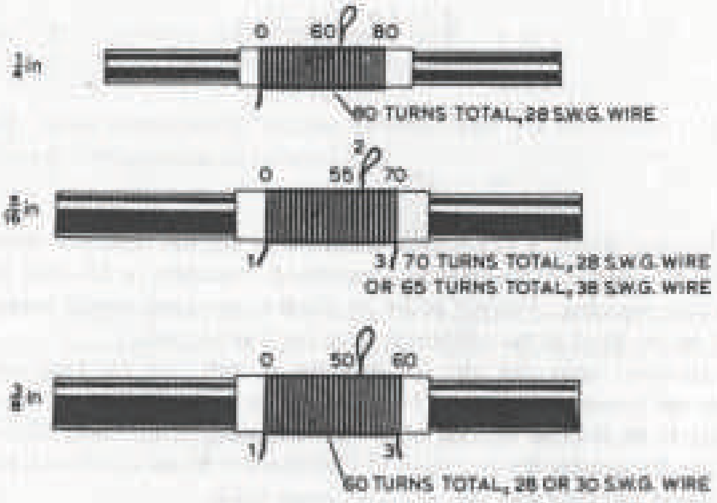
0
60
80
¼in
80 TURNS TOTAL, 28 S.W.G. WIRE
2
0
55
70
5/16in
1
3
70 TURNS TOTAL, 28 S.W.G. WIRE
OR 65 TURNS TOTAL, 38 S.W.G. WIRE
0
50
60
⅜in
1
3
60 TURNS TOTAL, 28 OR 30 S.W.G. WIRE

Fig. 4.5 Medium-wave aerial coil windings on three sizes of ferrite rod.

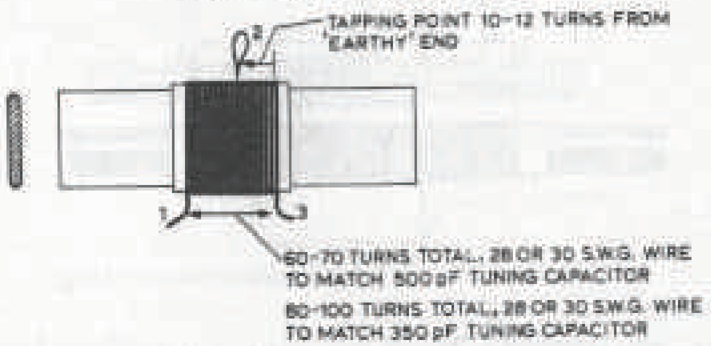
TAPPING POINT 10–12 TURNS FROM 'EARTHY' END
2
1
3
60–70 TURNS TOTAL, 28 OR 30 S.W.G. WIRE TO MATCH 500 pF TUNING CAPACITOR
80–100 TURNS TOTAL, 28 OR 30 S.W.G. WIRE TO MATCH 350 pF TUNING CAPACITOR

Fig. 4.6 Medium-wave winding for a ferrite slab.

A long length enables maximum adjustment of coil position (a way of 'trimming' the aerial) and, also, additional coils to be incorporated on the same rod, if the circuit needs them (e.g. for multi-waveband coils). But it is the diameter which really determines the winding specification.

Recommendations for medium-wave coil windings for these three standard ferrite-rod diameters are given in Fig. 4.5, all windings being enamelled copper wire. Normal tapping points are one-third from the 'earthy' end. It is a simple matter to experiment with different windings—e.g. adding more turns to extend the *wavelength* upwards (i.e. lower frequencies), or reducing the number of turns to extend the *wavelength* coverage downwards (i.e. for higher-frequency transmissions).

Ferrite-core material is also available in slab form, which is often easier to accommodate in a small receiver case. A typical ferrite-slab size is 2¾ in. long by ¾ in. wide, with a thickness of ⅛ in. or less. A typical winding specification is shown in Fig. 4.6.

Generally, the long waveband should be covered by *doubling* the number of turns used for the medium waveband. *Reducing* the number of turns shifts the coverage out of the medium waveband down into the trawler band (80 m) and 'top-band' (160 m). For short-wave coverage very few turns indeed will be needed—perhaps less than half a dozen—and the optimum number can only be determined by careful experiment.

Separate coils for the various wavebands to be covered can be accommodated on the same ferrite rod and connected to a common tuning capacitor via a waveband switch—Fig. 4.7. This is the usual

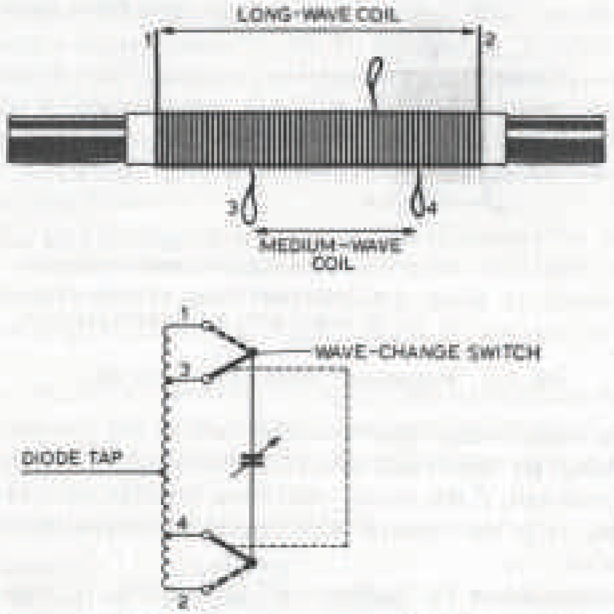
LONG-WAVE COIL
1
2
3
4
MEDIUM-WAVE COIL
1
3
WAVE-CHANGE SWITCH
DIODE TAP
4
2

Fig. 4.7 Long- and medium-wave aerial coil connected to wave-change switch.

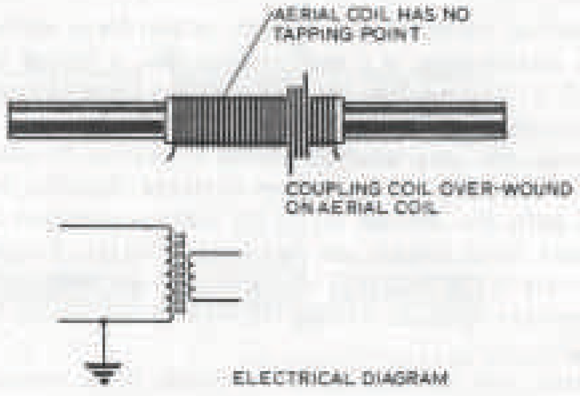
AERIAL COIL HAS NO TAPPING POINT
COUPLING COIL OVER-WOUND ON AERIAL COIL
ELECTRICAL DIAGRAM

Fig. 4.8 Inductively-coupled aerial coil.

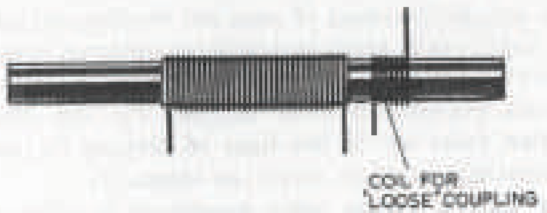
COIL FOR 'LOOSE' COUPLING

Fig. 4.9 'Loose' inductive coupling with separate coupling coil.

procedure with an elaborate receiver circuit and domestic radios, but is seldom satisfactory with the simpler circuits because of the tuned-circuit characteristics (i.e. the need for meticulous adjustment to get optimum results for any single waveband coverage).

All the coil constructions so far described are based on single windings (covering a specific waveband) with a tapping point for the output (diode connection in the case of simple receivers). The alternative form is to provide *inductive coupling* for the output in the form of an overwinding, which can comprise five or more turns of the same gauge wire used for the main winding–Fig. 4.8. The optimum number of turns for the overwinding depends on the degree of coupling required, which in turn is influenced by the characteristics of the following circuit. The position of the coupling coil can also affect the performance, although usually it is best positioned over the 'earthy' end

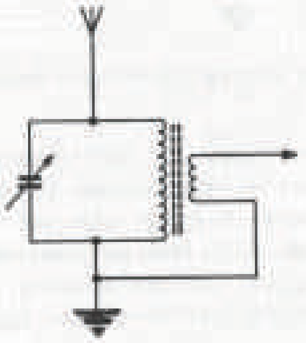

Fig. 4.10 Transformer coupling between tuned circuit and next stage of receiver.

of the main coil. Both number of turns and coupling-coil position are subjects for experimental adjustment to obtain optimum results.

Sometimes a coupling coil is wound on the ferrite rod, separate from the aerial coil, providing a 'loose' coupling–Fig. 4.9. Regenerative receivers often make use of this form of coupling for introducing 'feedback' into the front of the circuit (see Chapter 9).

Inductive coupling is also called *transformer* coupling, since the principle of passing current generated in one coil into another coil is the same. With transformer coupling (Fig. 4.10) the aerial coil has no tapping point, and is an alternative to direct coupling, where the next stage of the circuit is connected directly to a tapping point on the coil.

There is a third method of coupling which may be used between stages in a radio circuit–*capacity coupling*. Here, connection between the stages (e.g. to the tapping point on the aerial coil) is made through a capacitor, which isolates the two stages as far as d.c. flow is concerned (only a.c. will pass through the capacitor)–Fig. 4.11.

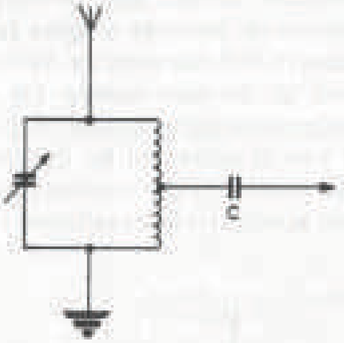

Fig. 4.11 Capacity coupling between tuned circuit and next stage of receiver.

Again the optimum value for the capacitor depends on the characteristics of the circuits being coupled. To couple an aerial coil to the next stage in a receiver, optimum values can vary from 200 pF to 10 μF. In general, though, a high value gives best results.

Direct, transformer, and capacity coupling are alternatives for connecting any two stages in receivers–not just the tuned circuit to the next stage.

5 AMPLIFIERS

It is a characteristic of a transistor that it works as an *amplifier* of signals. In the most widely used mode of connection of transistors for such duties with input and output circuits both connected to the emitter (common-emitter mode), the degree of amplification or current gain is called the 'beta' (β) of the transistor. This is given as h_{FE} (or static forward current ratio) in transistor characteristics data (see also Chapter 7).

A basic transistor amplifier circuit is very simple–and is identical for a p-n-p or n-p-n transistor, except for the battery polarity–see Fig. 5.1. Virtually any low-power AF transistor can be used in this circuit. The *bias resistor* (R) must have a value providing a collector current not exceeding the maximum specified rating for the transistor used, the actual current flowing in the collector circuit also being influenced by the voltage of the battery.

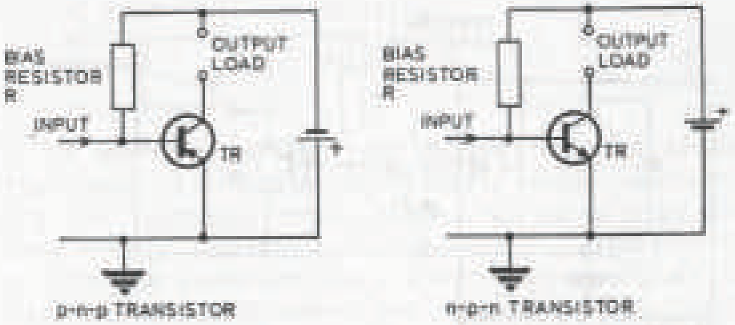

Fig. 5.1 Basic amplifier circuits with simple current biasing via a single resistor.

Knowing the transistor characteristics, a suitable value for R can be calculated as follows:

$$R = \text{gain} \times \frac{\text{battery voltage}}{\text{collector current}}$$
$$= h_{FE} \times \frac{\text{battery voltage}}{I_c}$$

where I_c is equal to, or preferably less than, the specification figure for I_c max.

Assuming that an OC72 transistor is being used, and the battery voltage is 9 V, specified figures are:

$$I_c \text{ max} = 250 \text{ mA}$$
$$h_{FE} = 30 \text{ to } 90$$

Taking the maximum gain, and 2.5 mA as a 'safe' working figure for the collector current:

$$R = 90 \times \frac{9}{0.0025}$$
$$= 324\text{k}\Omega$$

A suitable (preferred-value) resistor would thus be 330k.

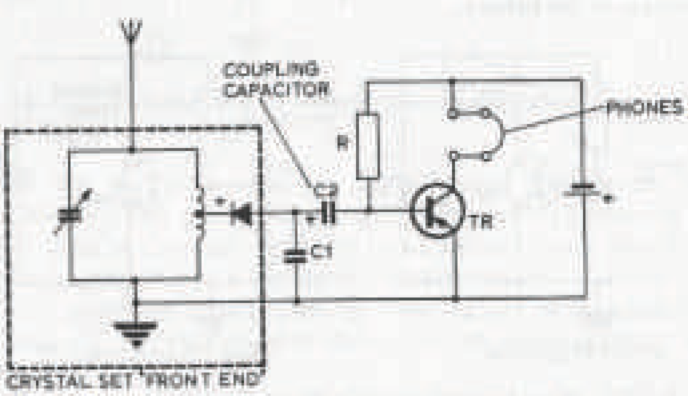

Fig. 5.2 *Circuit 5.* Crystal set with single stage of amplification. Any ~~RF~~ transistor can be used for TR.
AF

Fig. 5.2 shows the complete circuit for a simple AF amplifier of this type, coupled to the 'front end' of a basic crystal set. The detector (output) is coupled to the amplifier via capacitor C2, a suitable value for which would be 10 μF or higher (e.g. 20, 25, 30, 40 or 50 μF); capacitor C1 (0.001 μF) may not be necessary. The output load in the collector circuit is formed by high-impedance phones.

Note that the *polarity* of the battery used to power this circuit is important (as far as the transistor connection is concerned); also the connections to the *diode* and C2 (which will normally need to be an electrolytic type to provide the high-capacity value required). With an n-p-n transistor the battery polarity would be reversed; and also the diode and electrolytic capacitor connections.

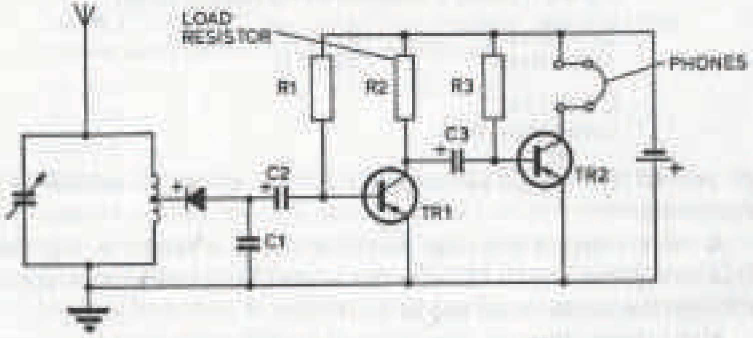

Fig. 5.3 *Circuit 6.* Crystal set with two stages of amplification. Component values matching OC71, OC72 or near equivalent for TR1 and TR2:

R1, 470 k	C1, 0.001 μF
R2, 4.7 k	C2, 8 or 10 μF
R3, 470 k	C3, 8 or 10 μF

Battery voltage 6 to 9 V.

The same type of circuit can be used to provide additional amplification, if required, simply by adding another amplifier stage (Fig. 5.3). The second amplifier stage can be identical to the first or based on a higher-power transistor taking the higher output present and providing even greater gain. The load in the output (collector) of the first amplifier stage is provided by a resistor (R2) (which should be about the same value as the phone resistance, 3.3 to 4.7 k), and the two stages coupled by a capacitor (C3). Value of R3 depends on the second transistor used, and may be anything from 47 Ω to 470 kΩ; capacitors

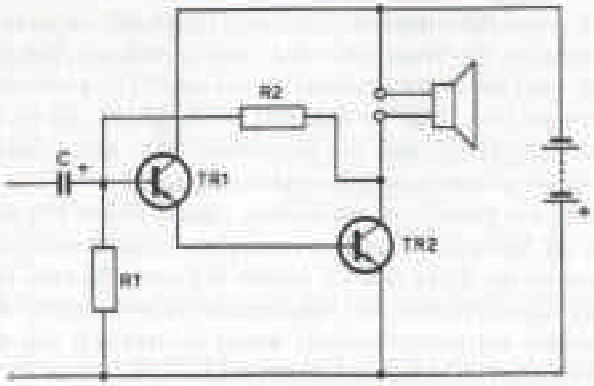

Fig. 5.4 *Circuit 7.* Amplifier for loudspeaker output

TR1, OC72 R1, 6.8 k
TR2, AD140 R2, 33 k
C, 8 or 10 μF
Loudspeaker, 80 Ω

C2 and C3 can be 10 μF or larger (20 μF up to 50 or 100 μF), electrolytic.

A more compact two-stage amplifier circuit is shown in Fig. 5.4. This may prove capable of operating a small 80 Ω loudspeaker direct, although the current drain will be quite high.

Alternatively, three or four stages of amplification using low-power transistors, following a basic 'front end' crystal set, should provide enough power to drive a small loudspeaker at the final output, through a suitable step-down transformer, to provide an impedance match (see Chapter 6).

Simple amplifier circuits of this type have the important limitation that performance of the transistor(s) will tend to vary with temperature. There is also the possibility of 'thermal runaway' developing, which can destroy the transistor(s), because as the external temperature increases the collector current also tends to increase, which in turn causes a further increase in junction temperature, so the effect is cumulative and goes from bad to worse, even to the point of ruining the transistor completely. It is possible to overcome this trouble by arranging for the circuit to be self-biasing or d.c. stabilized so that a constant operating collector current is provided, regardless of transistor type or temperature variations. In other words, the working point of the collector circuit is stabilized.

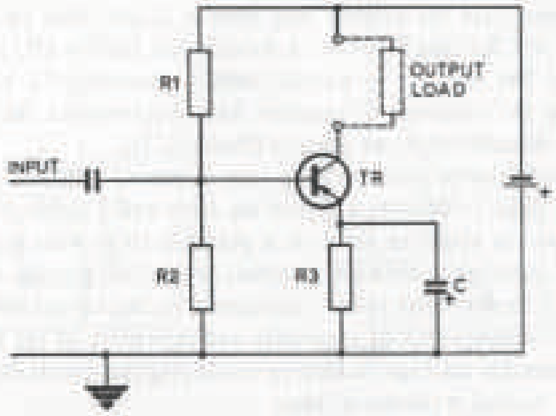

Fig. 5.5 Stabilized bias circuit for transistor amplifier stage. Typical values for low/medium-power transistors

R1, 22 k C, 8 or 10 μF
R2, 10 k
R3, 1 k

A further advantage of a stabilized circuit is that it makes the performance of the amplifier less dependent on the characteristics of individual transistors, which can differ appreciably even for the same type. Capacity coupling between stages should be used since this makes each stage independent as regards working, rather than inter-dependent.

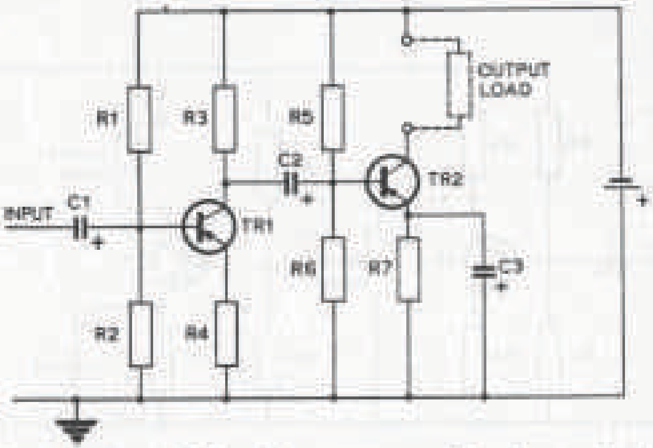

Fig. 5.6 *Circuit 8.* Practical two-stage amplifier using two low/medium-power AF transistors. (See Chapter 7 for determination of matching component values.)

To achieve this the original bias resistor is split into two separate values R1 and R2 (see Fig. 5.5). A further bias resistor (R3) is applied directly to the emitter, in parallel with a capacitor (C) to act as a by-pass for AF currents. (Transistor bias requirements, and data on transistor characteristics, are given in Chapter 7.)

A receiver design incorporating two stages of amplification with stabilized circuits following a crystal set front end is shown in Fig. 5.6. This, in fact, is about as far as it is practical to go with such a basic circuit as, although additional complete stages will provide more gain and greater final output power, deficiency in the circuit will also be aggravated–notably lack of sensitivity and selectivity in the front end. Such deficiencies are best tackled by modifying the circuit design from the simple 'crystal set detector' basis.

The amplifier stage(s) also provides a convenient point to insert a *volume control* into the receiver circuit. This takes the form of a potentiometer which can replace one of the bias resistors, or be placed in series with the flow of the current after the first stage of amplification–Fig. 5.7. This will introduce minimum distortion over the volume control range.

On low-power receivers which give satisfactory reception without an external aerial, a volume control can be an unnecessary refinement. Since the ferrite aerial has directional characteristics, merely turning the set one way and the other will produce changes in signal strength.

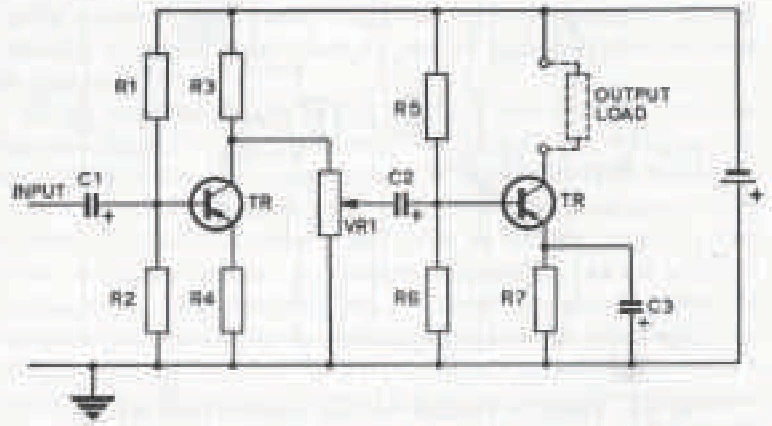

Fig. 5.7 *Circuit 9.* Two-stage amplifier with volume control.

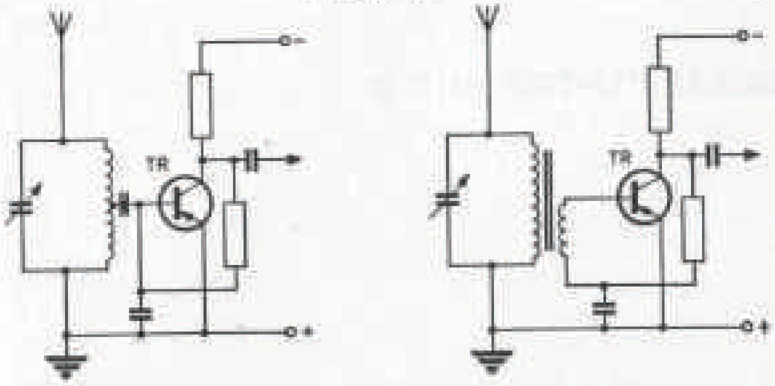

Fig. 5.8 Basic pre-amplifier circuits. The transistor (TR) must be an RF type; p-n-p transistors shown in these diagrams. If n-p-n transistors are used, the battery polarity must be reversed.

Pre-amplifiers

A pre-amplifier is designed to amplify signals between the tuned circuit and detector stages. The main difference is that the transistor(s) selected must be an RF type since it is handling RF signals at this stage. Otherwise the basic circuit is similar to that of an AF amplifier–see Fig. 5.8.

6 THE OUTPUT STAGE

By a suitable choice of transistors, the amplifier stages can be used to power a loudspeaker direct, using a minimum of components. A typical circuit of this type was shown in Fig. 5.4, using a high-gain AF transistor for TR1, and a power transistor for TR2. The main disadvantage of such a set-up is the relatively high current drawn by the circuit.

Single transistor outputs work with 'Class A' operation, which means that the values of bias and signal voltage applied to the transistor ensure that collector current always flows. Fig. 6.1 shows a basic Class A output circuit incorporating transformer coupling to a loudspeaker.

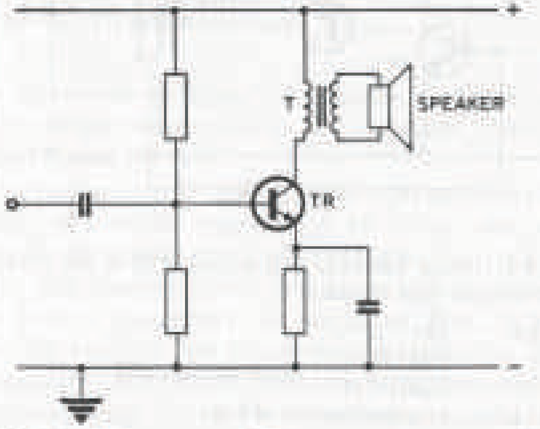

Fig. 6.1 Basic Class A output. TR is output transistor working as an amplifier. Loudspeaker impedance is matched to output load requirements via the step-down transformer T.

A more economic way of producing satisfactory output power is to employ a single transistor *driver* working a complementary pair of transistors (an n-p-n and a p-n-p selected with matched characteristics) in push–pull configuration. The output power obtained from a pair of transistors in push–pull is considerably more than double the power obtained from a single transistor of the same type. With 'Class B' operation the transistors are biased to nearly cut-off, so that only a marginal current flows under 'quiet' conditions. Push–pull outputs may, however, also be designed for 'Class AB' operation, with rather higher current drains.

Basically, distortion is lowest with Class A operation, whilst Class B operation provides the lowest current drain but introduces the possibility of crossover distortion being present, which can be overcome by applying a slight forward bias to each transistor. Class AB offers a compromise between the two.

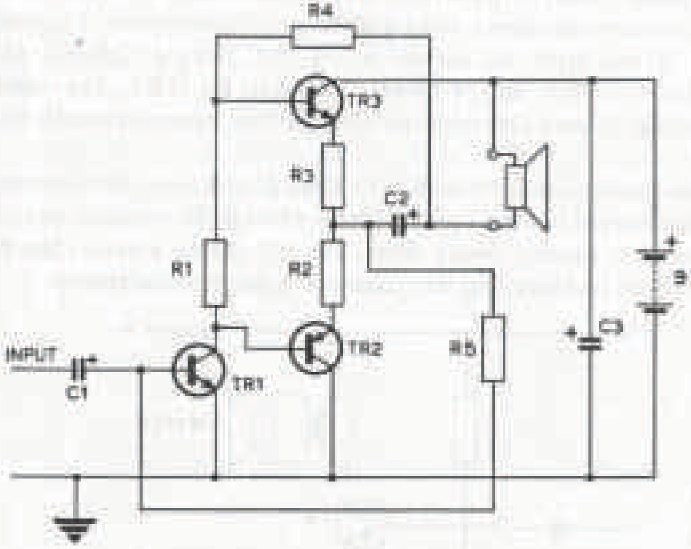

Fig. 6.2 *Circuit 10.* Push–pull output. TR1 is the driver, TR2, TR3 the push–pull amplifier.

R1, 68
R2, 1
R3, 1
R4, 1 k
R5, 56 k
C1, 5 μF
C2, 250 μF
C3, 50 μF

TR1, OC72
TR2, OC81 } complementary pair
TR3, AC127 }
Loudspeaker, 3 Ω or 8 Ω

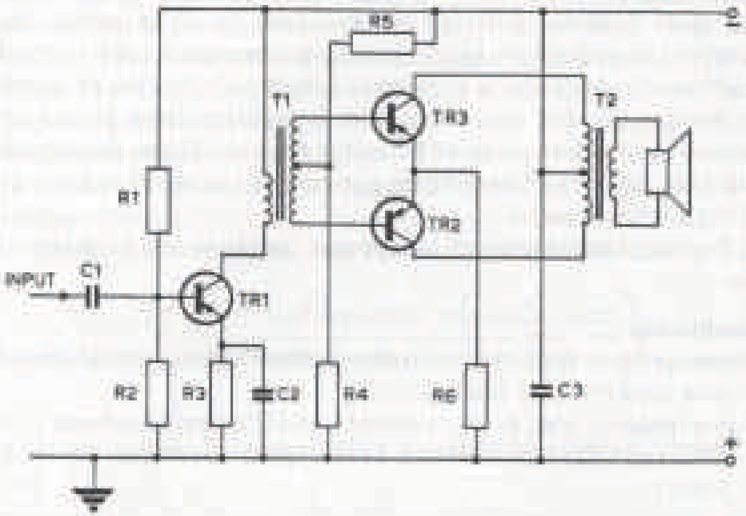

Fig. 6.3 *Circuit 11.* Push–pull output with transformer coupling. TR1 is the driver; TR2, TR3, AD140 or AD149. Typical component values:

R1, R2 to match TR1 (depending on type used)

R3, 1 k
R4, 100
R5, 4.7 k
R6, 10

C1, C2, 50 μF
C3, 50 μF or 80 μF
T1, coupling transformer
T2, output transformer

Two types of basic push–pull output circuits are shown in Figs. 6.2 and 6.3, one with direct coupling and the other using transformer coupling, both interstage (between driver and push–pull input) and to the loudspeaker. The coupling transformer can provide additional amplification; an output transformer, on the other hand, is invariably a step-down type to adjust the loudspeaker impedance to the required output impedance. Both types of circuits have their advantages and disadvantages, although for simpler receivers all-transistor circuits are usually preferred.

Much also depends on the requirements of the receiver. Thus to operate a small loudspeaker successfully an *audio power output* of 5 mW or better is required (higher still for larger speakers, of course). At the other extreme, about 10 μW represents, for most people, the threshold of audibility in high-impedance phones; and 0.1 mW a normal minimum for comfortable listening and ready identification of sounds in headphones. For easy listening with high-impedance phones, an audio output power of up to 0.5 mW is desirable. Higher signal levels will tend to 'swamp' headphones but can, of course, be reduced by fitting a volume control.

Typical characteristics of headphones, earpieces and loudspeakers are:

Headphones

High-impedance type, d.c. resistance 2,000–4,000 Ω; typical impedance 10,000 Ω (at 1 kHz).

Low-impedance type, (i) d.c. resistance 15 Ω, typical impedance 80 Ω (at 1 kHz); (ii) d.c. resistance 80 Ω, typical impedance 120 Ω (at 1 kHz).

Earpieces

High-impedance type, d.c. resistance 2,000 Ω; typical impedance 7,500 Ω (at 1 kHz).

Low-impedance type, (i) d.c. resistance 4 Ω, typical impedance 15 Ω (at 1 kHz); (ii) d.c. resistance 14 Ω, typical impedance 60 Ω (at 1 kHz); (iii) d.c. resistance 60 Ω, typical impedance 250 Ω (at 1 kHz).

Loudspeakers

Typical d.c. resistance 3 Ω; typical impedance 8–16 Ω.

It will be obvious from a study of these figures that low-impedance phones, a low-impedance earpiece, or a loudspeaker will be a mis-match

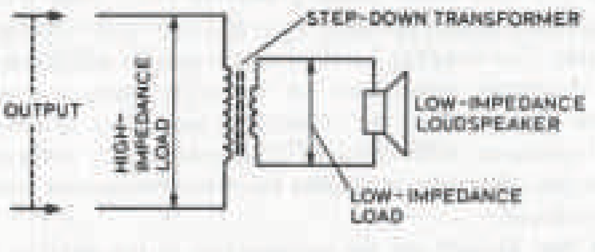

Fig. 6.4 Load balancing via output transformer.

for coupling to an output requiring a *high-impedance load* (as in the case of most of the simple all-transistor output circuits).

To employ low-impedance phones, earpiece or a loudspeaker with an output requiring a high-impedance load, a matching step-down transformer (output transformer) must be used. The primary of the transformer then provides the required output load, indirectly coupled to the secondary to which is connected the low-impedance phones or speaker—Fig. 6.4.

The turns ratio required from the transformer is easily calculated as:

$$\sqrt{\left(\frac{\text{output load impedance required in ohms}}{\text{phone or speaker impedance in ohms}}\right)}$$

Some typical transformer ratios and their suitability for matching are:

Ratio	Listening device	Equivalent output load impedance (Ω)
41:1	8-Ω Speaker	15,000
35:1	4-Ω earpiece	5,000
30:1	8-Ω Speaker	7,000
18:1	14-Ω earpiece	5,000
14:1	15-Ω earpiece	2,000
9:1	60-Ω earpiece	5,000
5:1	80-Ω headphones	2,000

For other required output load impedances, the turns ratio can be calculated from the formula.

Tone controls

The output voltage after amplification will only be a faithful reproduction of the original input if the amplifier produces the same gain for all signals, whatever their frequency and complexity. This is seldom the case with simple circuitry, so that the balance of the original sound or speech is upset and distortion results. If the gain is inadequate at low frequencies, the sound reproduced will tend to be tinny or harsh; conversely, if the gain is inadequate at the higher frequencies the sound is subject to booming.

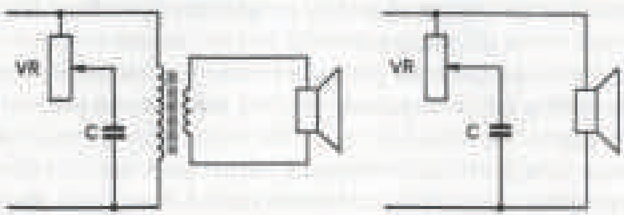

Fig. 6.5 Simple tone controls.

An adequate measure of tone control can be realized by connecting a variable resistance (potentiometer) and fixed capacitor in series across the primary terminals of the output transformer; or directly across the speaker, as shown in Fig. 6.5.

This, in effect, forms another tuned circuit, the resonant frequency of which can be altered by adjusting the setting of the potentiometer to favour or raise the treble or bass in the output signal by, effectively, additional amplification. The tone control on a domestic receiver is usually of this form—the knob controlling a potentiometer connected in series with a fixed capacitor across the output transformer. Typical values used are: potentiometer, 0–5 k; fixed capacitor, 0.1 μF.

7 TRANSISTORS, BIAS AND STABILIZATION

The two conventional bias arrangements for transistors are shown in Figs. 7.1 and 7.2, both having advantages and disadvantages. Current bias is the simpler of the two, since it requires only one resistor which determines the values of the base current I_B and the operating or emitter current I_E, viz:

$$I_B = \frac{V_S - V_{BE}}{R}$$

$$I_E = \frac{V_S - V_{BE}}{R}(1 + h_{FE})$$

V_{BE}, the base-emitter voltage, is of the order of 0.1 to 0.2 V for germanium transistors, 0.6 to 0.7 V for silicon transistors. In normal practice there will be negligible change in current with variations of V_{BE}, but will be dependent on the spreads of h_{FE} (the small signal

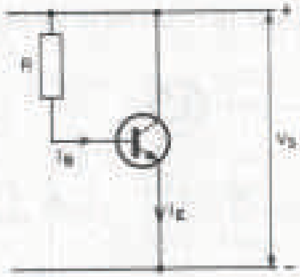

Fig. 7.1 Simple current bias circuit.

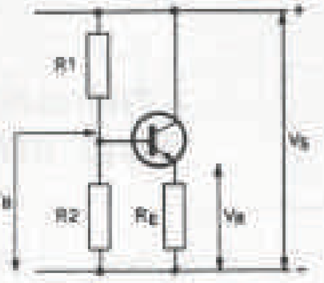

Fig. 7.2 Conventional voltage bias circuit.

forward current ratio of the transistor). Thus simple current biasing is most suitable for transistors which have only narrow h_{FE} spreads.

Voltage biasing with emitter feedback yields the following voltage relationships (see also Fig. 7.2):

$$V_B = \frac{R2 \, . \, V_S}{R1 + R2}$$

$$V_E = \frac{R2 \, . \, V_S}{R1 + R2} - V_{BE}$$

The emitter current is therefore:

$$I_E = \frac{R2 \, . \, V_S}{R_E(R1 + R2)} - \frac{V_{BE}}{R_E}$$

For the effect of spreads to be negligible the voltage V_E must be large compared with changes in V_{BE}. Also, R_E must be large if changes in emitter current I_E, due to variations in the supply voltage V_S, are to

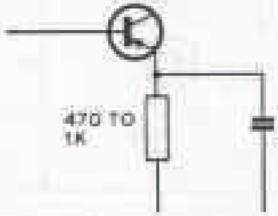

Fig. 7.3 Stabilization, with common component values used.

be kept small. In practical circuits a voltage drop (V_E) of 1 V with germanium transistors and 3 V with silicon transistors is aimed at.

The collector-to-base leakage current is a major factor contributing to the shift in operating point of a transistor, since this is highly sensitive to temperature. Thus bias stabilization is normally desirable with germanium transistors, the usual arrangement being as shown in Fig. 7.3. Silicon transistors have negligible leakage current, even over their full working temperature range, and stabilization is not necessary, except in some cases where other characteristics developed in the particular circuit may make it a prudent addition.

Working characteristics of a number of readily available transistors are summarized in the following tables. Data on other types are available from the manufacturers; also in many suppliers' catalogues.

Germanium Transistors

p-n-p low power

	OC41	OC42	OC70	OC72	OC75	OC81	OC81D
V_{cbo} max.	16	16	30	32	30	32	32
V_{ceo} max.	15	15	30	32	30	32	16
V_{eb} max.	12	12	10	10	10	10	10
I_c max. mA	150	125	50	250	100	500	250
P_t max. mW	84	83	125	125	150	240	200
h_{FE} (h_{fe})	20–80	40 min.	20–40	30–90	60–120	50–250	20 min.
@I_c mA	50	50	0.5	80	1	50	2
I_{cbo} max. µA	30	30	5	15	10	10	10
@V_{cb}	15	15	15	10	4.5	10	10
V_{ce} sat. max.	0.2	0.2	0.21	0.25	0.21		
@I_c/I_b mA	50/3	50/1.5	9/0.5	125/12.5	9/0.5		
f_T min. MHz	3	5.5	1 typ (fhfb)	1 typ	1 typ (fhfb)	1 typ (fhfb)	1 typ (fhfb)

Small, medium current switching, AF amplifiers n-p-n

	2N1302	2N1304	2N1306	2N1308	ASY28	ASY29	NKT713
V_{cbo} max.	25	25	25	25	30	25	30
V_{ceo} max.	25	20	15	15	15	15	30
V_{eb} max.	25	25	25	25	20	20	15
I_c max. mA	200	200	200	200	200	200	500
I_c peak mA	300	300	300	300	300	300	
P_t max. mW	150	150	150	150	150	150	150
h_{FE}	20 min.	40–200	60–300	80 min.	30–80	50–150	50–150
@ I_c mA	10	10	10	10	20	20	50
I_{cbo} max. µA	6	6	6	6	35	35	15
@ V_{cb}	25	25	25	25	30	25	10
V_{ce} sat. max.	0.2	0.2	0.2	0.2	0.25	0.25	0.15
@ I_c/I_b mA	10/0.5	10/0.25	10/0.17	50/2	50/2	50/1.25	50/5
f_T typ. MHz	10	15	20	30	14	20	1 (hfb)

Audio frequency amplifiers, low power output, n-p-n

	NKT773	AC127	AC176	AC176K	AC187K
V_{cbo} max.	15	32	32	32	25
V_{ce} max.	15	32	20	20	15
V_{eb} max.	5	10	5	5	10
I_c max. mA	300	500	300	300	1000
P_t max.	150	200	700	1000	1000
h_{FE}	50–250	25–143	52–180	52–180	100–500
@ I_c mA	50	500	500	500	500
I_{cbo} max. µA	15	10	100	100	15
@ V_{cb}	10	0.5	25	25	10
V_{ce} sat. max.		1.0	0.6	0.6	
@ I_c/I_b mA		500/50	500/10	500/10	
f_T min. MHz		1.5	1.0	10	1.0

High power output, p-n-p

	NKT403	NKT404	NKT405	AD140	AD142	AD149	AD150
V_{cbo} max.	80	60	60	55	80	50	32
V_{ceo} max.	32	32	32		50	30	30
V_{eb} max.	40	20	20	10	10	20	10
I_c max. A	10	10	5	3	10	3.5	3.5
P_t (@°C) W	50(25)	50(25)	50(25)	36(37)	30(55)	22(50)	22(50)
h_{FE}	50–150	50–150	100–200	30–100	30–170	30–100	30–100
@ I_c A	1	1	1	1	1	1	1
I_{cbo} max. µA	150	150	150	100	5000	350	1000
@ V_{cb}	30	30	30	14	80	14	30
V_{ce} sat. max.	0.42	0.42	0.42		0.3		
@ I_c/I_b A	1/0.1	1/0.1	1/0.1		5/0.25		
f_T typ. MHz	0.35	0.35	0.35		0.45	0.5	0.45

Silicon transistors

Audio frequency amplifiers, small signal, general purpose, n-p-n

	2N3708	2N3709	2N3710	2N3711	2N3904	2N4124
V_{cbo} max.	30	30	30	30	60	30
V_{ceo} max.	30	30	30	30	40	25
V_{eb} max.	6	6	6	6	5	5
I_c max. mA	30	30	30	30	100	200
P_t max. mW	250	250	250	250	310	310
h_{FE}	45–660	45–165	90–330	180–660	100–300	120–360
@ I_c mA	1	1	1	1	10	2
I_{cbo} max. µA	0.1	0.1	0.1	0.1	0.1	0.05
V_{ce} sat. max.	1	1	1	1	0.2	0.3
@ I_c/I_bmA	10/0.5	10/0.5	10/0.5	10/0.5	10/1	50/5
f_T typ. MHz					200	300 min.

Audio frequency amplifiers, low level, low noise, *n-p-n*

	2N2484	2N2924	2N2925	2N2926	2N4286	2N5088
V_{cbo} max.	60	25	25	18	30	35
V_{ceo} max.	60	25	25	18	25	30
V_{eb} max.	6	5	5	5	6	3
I_c max. mA	50	100	100	100	100	50
P_t max. mW	360	200	200	200	250	310
h_{FE}	100–500	150–300	235–470	35–470	100 min.	300–900
@I_c mA	0.01	2	2	2	0.01	0.1
I_{cbo} max. μA	0.01	0.5	0.5	0.5	0.05	0.05
Noise factor dB	3 max.	2.5 typ.	2.5 typ.	2.5 typ.		3 max.
f_T typ. MHz	60 min.	120	120	120	280	50 min.

8 TRF RECEIVERS

The tuned radio frequency (TRF) receiver is based on 'double tuning'–one before and one immediately after the pre-amplifier or RF amplifier, as shown in the block diagram of Fig. 8.1. The two tuned circuits are literally identical and, in fact, are usually 'ganged' together, using twin variable condensers and adjusted by a single control knob (although the respective circuits are wired to the separate electrical parts of the two-part condenser–see Fig. 8.2).

To provide *complete* matching of these two tuned circuits, the two separate variable condensers are individually adjusted with the aid of smaller 'trimming' capacitors mounted in parallel with them. Similarly, the inductances of the two tuning coils completing the respective tuned circuits can be matched for identical performance by winding on coil formers with iron-dust cores. In this manner small differences in the actual wiring up of the two tuned circuits, and in effective component values, can be balanced out for optimum performance. This, in fact, is a necessary process in setting up or *aligning* the complete receiver for best performance, and can be done quite satisfactorily on a trial-and-error basis.

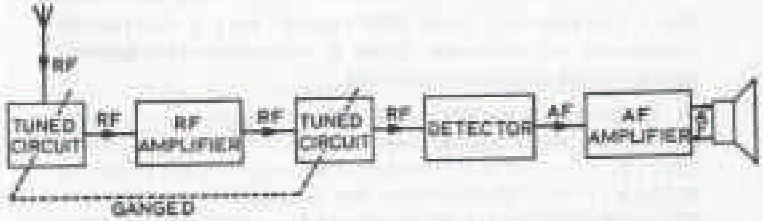

Fig. 8.1 Block diagram of a tuned radio frequency (TRF) receiver.

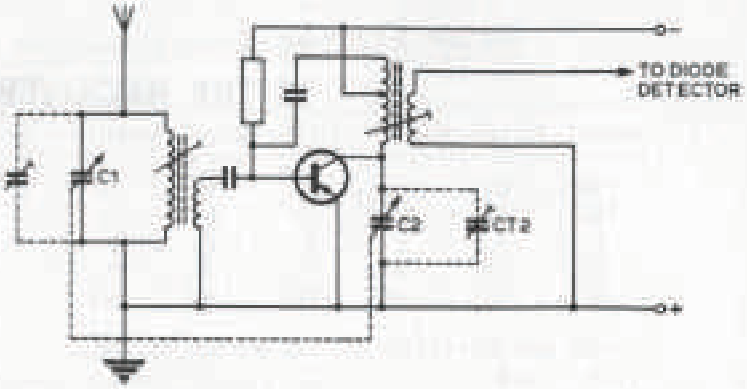

Fig. 8.2 Double-tuned circuit with RF amplifier – the 'front end' of a TRF receiver.

A simple one-transistor TRF receiver circuit is shown in Fig. 8.3. Alignment difficulties are minimized by employing a proprietary coil L1, L2, L3 for TRF working, eliminating the need for ganged tuning capacitors with separate trimmers. It is an interesting circuit to experiment with and compare the performance against that obtained from a basic crystal set with one stage of amplification added. It should be superior, but like the crystal set will require an efficient external aerial and good earth connections before satisfactory listening volume is obtained in most areas.

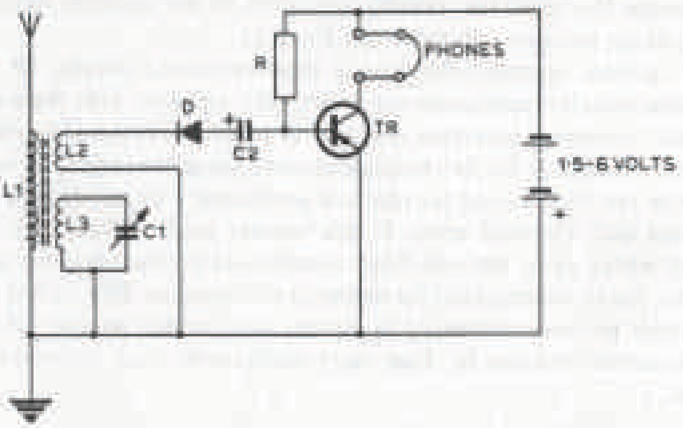

Fig. 8.3 *Circuit 12.* Basic TRF receiver using a minimum of components. An interesting circuit to experiment with against a crystal set with transistor amplifier(s).

L1, L2, L3, proprietary TRF coil (Teletron type HAX)
C1, 500 pF variable capacitor (Jackson)
C2, 2 μF
R, 220 k
TR, OC70, OC71, or any AF transistor (value of R to be adjusted to match higher power transistors)

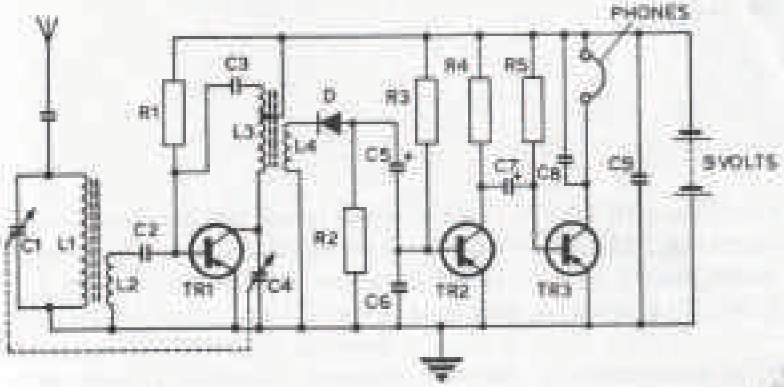

Fig. 8.4 *Circuit 13.* Three-transistor TRF receiver. Using higher power AF transistors (TR2 and TR3), this design could operate a loudspeaker instead of phones.

C1, C4, ganged tuning condenser (Jackson, 500 pF)
C1, 0.01 μF
C3, 15 pF
C5, 8 μF electrolytic
C6, 0.01 μF
C7, 8 μF electrolytic
C8, 0.005 μF
C9, 100 μF, 12 volt working

R1, 1 MΩ
R2, 10 kΩ
R3, 1 MΩ
R4, 4.7 kΩ
R5, 1 MΩ
TR1, OC45
TR2, OC71
TR3, OC71

L1, L2, standard ferrite aerial windings (see Chapter 4)
L3, 50 turn coil on ferrite rod with centre tapping
L4, 10 turns
D, any crystal diode

L3, L4 can be wound exactly as L1, L2, except that the 50-turn winding has a centre tap. A short length of ferrite rod should be used for the core, L3, L4 being wound on a paper sleeve so that the rod can be slid in and out for adjustment purposes.

Important. Coil L3, L4 will need screening from L1, L2 and associated components. This can be done with thin aluminium sheet, suitably fitted around L3, L4.

Performance can be improved by using a higher power transistor for amplification, and/or adding a further stage or two of amplification. This follows along the same lines as for other receivers—see Chapter 5.

A rather more elaborate TRF circuit, but again using low-power transistors, is shown in Fig. 8.4. This uses a standard ferrite rod aerial with inductive coupling (see Chapter 4), and a ganged tuning capacitor for C1, C4.

9 REGENERATIVE AND REFLEX RECEIVERS

In the crystal set, and similar simple receivers, a weak RF signal in the aerial is fed directly to the detector, and the extracted AF signal amplified to the required level to operate headphones or a loudspeaker. However, no amount of AF amplification *after* detection is effective if the original RF strength is too small.

An obvious method of improving this state of affairs would be to amplify the RF signal picked up by the aerial and tuned circuit *before* passing to the detector. This is quite a practical arrangement, the immediate advantages being an improvement in both *sensitivity* and *selectivity*, and also *quality* of reproduction. Improved sensitivity means that weaker (more distant) stations can now be received at greater volume; better selectivity means that sharper tuning can be obtained, a necessity to eliminate interference from adjacent stations now more readily picked up.

Quality is improved because there is less demand for AF amplification, while the output power can readily be boosted to operate a loudspeaker without putting any particularly high demands on either the efficiency of the aerial system or the AF amplifier stages.

An elementary circuit of this type is shown in Fig. 9.1, using a single transistor as a pre-amplifier, followed by a diode detector. This is easy to build and should work quite well in most areas, although still almost certainly requiring a good external aerial and earth connections for the best results, and must be used with an *inductively coupled* tuned circuit.

There is, however, a method of considerably improving the performance of this set by incorporating *regeneration*. This is done by taking part of the amplified signal and feeding it back into the tuned

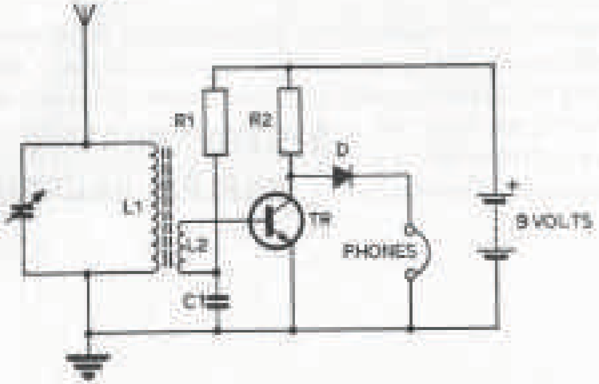

Fig. 9.1 *Circuit 14.* Single-transistor receiver with pre-amplifier.
TR, 2N2926 (or equivalent). Matching component values:

C1, 0.01 μF
R1, 1 MΩ
R2, 2.2 k
L1, L2, inductively coupled aerial coils
D, any crystal diode

circuit, where it will be amplified again, resulting in a further improvement in selectivity and sensitivity. Care should be taken that not too much signal is fed back, causing the circuit to become unstable and oscillate or 'squeal'.

This modification is shown in Fig. 9.2. One length of insulated wire

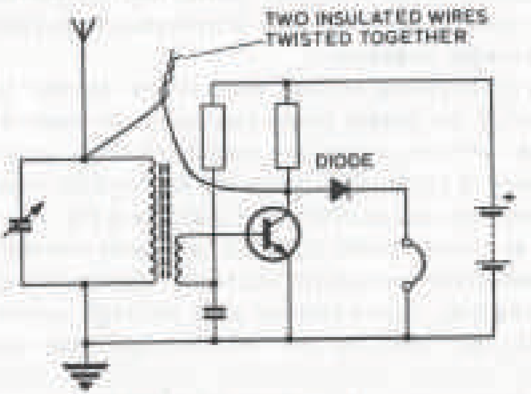

Fig. 9.2 *Circuit 15.* The design of Fig. 9.1 converted to regenerative working. The twisted wires can be replaced by a 10-pF or 20-pF trimmer capacitor.

is connected to the front end of the diode, another to the top end of the tuned circuit. The free ends of these two wires are then twisted together over a length of about half an inch.

This, in effect, provides a capacitive coupling for the feedback signal where the actual amount of feedback can be varied by twisting the wires more tightly together, or untwisting part of their length. The optimum amount of 'twist' can be determined by experiment, when the final performance should be considerably better than that of the original circuit. However, the *position* of the twisted wire coupling relative to other components is critical; any slight disturbance can considerably modify the performance or cause the receiver to oscillate.

For a more stable arrangement, a 10-pF or 20-pF trimmer capacitor can be used instead of twisted wires to provide the feedback coupling, adjusted for optimum results.

If the set does not work with 'regeneration', or the results are poorer than the original circuit, then the feedback is probably 'negative' rather

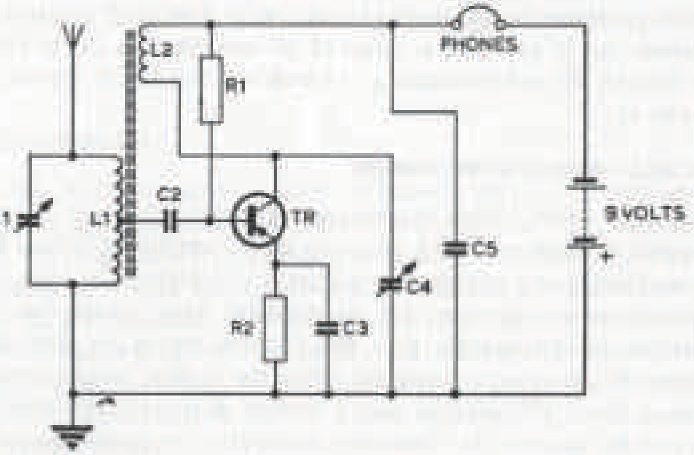

Fig. 9.3 *Circuit 16.* Simple regenerative receiver with inductively coupled feedback

TR, OC42 (or equivalent)
C1, 500-pF tuning capacitor
C2, 220 pF to 0.1 μF (find best value by experiment)
C3, 2 μF to 8 μF
C4, 500-pF variable capacitor
C5, 10 μF
R1, 1 M to 4.7 M (find best value by experiment)
R2, 3.3 k

than 'positive'. Reversal of the coupling-coil connections should correct this.

An even better arrangement is to use *inductive* rather than capacitive coupling for the feedback, which can be done by winding a feedback coil on the same ferrite rod as the aerial coil. A suitable feedback coil is made by winding 8 turns of 38-s.w.g. wire on a similar paper sleeve to the tuning coil, so that the finished coil can be slid up and down the ferrite rod. The degree of coupling can then be adjusted by sliding this feedback coil up and down the rod to arrive at a position which gives optimum results.

The circuit is also modified somewhat, as shown in Fig. 9.3, to include a variable capacitor connected to the transistor collector to control the amount of reaction or regenerative feedback. If the set does not oscillate with this variable capacitor adjusted to *minimum* capacity (i.e. vanes fully closed) then the feedback coil must be reversed on the ferrite rod (alternatively reverse the coil connections to the rest of the circuit).

This particular circuit can also be used as the 'front end' of a receiver to which one or two further stages of AF amplification can be added (see Chapter 5); and terminating in a push–pull output if desired (see Chapter 6).

The super-regenerative receiver

The limit to which ordinary regenerative amplification can be carried is the point at which oscillation starts, so the RF amplification (and thus the sensitivity) of a regenerative receiver is limited by this factor. The super-regenerative receiver is a development which overcomes this limitation by introducing into the detector circuit an alternating voltage, of a frequency somewhat above the audible range (typically between 20 to 100 kHz), in such a manner as to vary the operating point of the detector. In effect, this interruption or *quench* frequency switches the detector (originally adjusted to be near the point of oscillation) in and out of operation. The time taken for the RF oscillations to build up to their peak will be proportional to the original modulated signal (i.e. the RF signal picked up by the aerial), and by arranging for these to be quenched before they can reach their peak, extremely high amplification is possible–equivalent to a gain of over one million in a single stage in certain cases.

On this basis, the circuit is ideal for increasing the sensitivity of a receiver in a relatively simple manner. However, both the design requirements and adjustment are somewhat critical, but more advanced and detailed treatment of these is beyond the scope of this book. The super-regenerative receiver, too, is not particularly suited for the reception of normal broadcast frequencies. The circuit requires a ratio of signal frequency to quench frequency of the order of 1,000:1 or greater, which means that for a broadcast signal of 500 kHz the quench frequency must be of the order of 500–1,500 Hz, or in the audio band. A further limitation is that, as they are continually oscillating, they will cause interference on other receivers in the neighbourhood–although this effect can be minimized by suitable circuit design.

A further variation, which is even less suitable for lower radio frequencies but again is capable of giving excellent results with very high frequencies, is the self-quenching super-regenerative receiver. In this case, as the name implies, the super-regenerative detector supplies its own quench frequency. The frequency of these quench oscillations depends on the feedback and the 'time constant' of the circuit, the time between each burst of oscillation varying as the input signal varies. The action of the oscillations is said to be a blocking or 'squegging' effect.

Reflex receivers

A reflex circuit works on a rather different principle. Here the *audio frequency* signal following the detector is fed back into a pre-amplifier stage, the complete circuit shown in block diagram form in Fig. 9.4 also providing RF and AF amplification. It looks a more complicated circuit involving a considerable number of components but, in fact, a *single* transistor can be made to perform the work of RF and AF amplifier.

Fig. 9.5 shows a basic single-transistor reflex circuit which is fairly non-critical and should be easy to operate satisfactorily; this is actually a three-stage circuit.

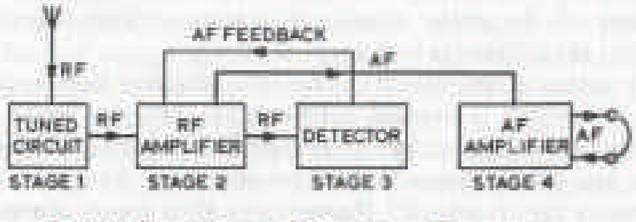

Fig. 9.4 Block diagram of a four-stage reflex receiver.

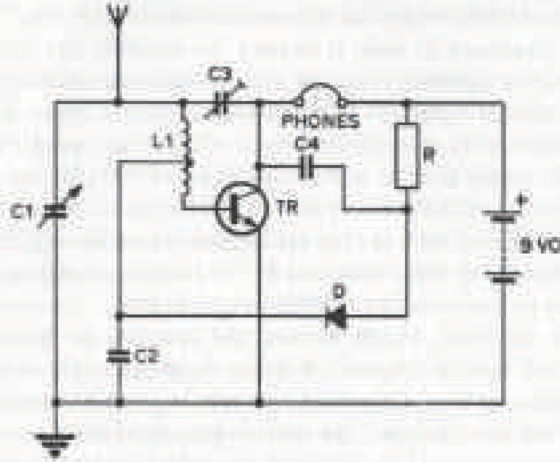

Fig. 9.5 *Circuit 17.* Single-transistor three-stage reflex receiver.

L1, standard aerial coil with tapping point
C1, 500-pF variable capacitor
C2, 0.05 µF
C3, 10 pF or 20 pF
C4, 1,000 pF
R, 1 M
TR, 2N2926

A rather more conventional design is shown in Fig. 9.6; the 'front end' of this circuit comprises the tuned circuit feeding an RF amplifier stage. In this type of circuit the tuning coil must be inductively coupled. The output from the RF amplifier employs what is known as *choke capacitance* coupling, a radio-frequency choke coil (RFC) being used to ensure an adequate load impedance for RF currents without having a d.c. voltage drop (*i.e.* offering high resistance to RF but low resistance to d.c.). The use of this choke coil also means that the gain of the amplifier varies with the frequency of the input, because the reactance of the choke varies with frequency. This effect is not desirable, but is tolerable in this type of circuit.

The output of amplified RF is fed from this stage into the detector stage. A feedback is arranged as follows: the AF signal passed by the detector, plus a little residual RF appearing at this point, is passed back to the base of the transistor in the previous stage, the AF being little affected by the capacitor C7. However, any RF is heavily attenuated or reduced, thus the transistor is made to perform the additional function of an AF amplifier. By choosing a suitable value for C7 a low reactance to RF is assured, but a high reactance to AF. For example, a 0.01 µF capacitor has a reactance of 30 Ω at 500 kHz, increasing to 3,000 Ω at 5 kHz. If 5 kHz represents the *upper* limit of AF signals likely to be encountered, reactance will become proportionately higher at lower audio frequencies. Conversely, for radio frequencies above 500 kHz the reactance will decrease proportionately.

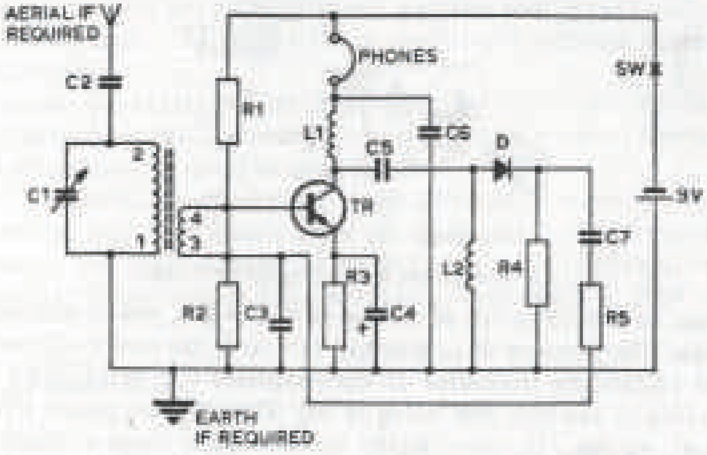

Fig. 9.6 *Circuit 18.* Single transistor three-stage reflex receiver with choke capacitance coupling.

C1, 100-500 pF	R1, 22 kΩ
C2, 220 pF	R2, 4.7 kΩ
C3, 0.01 µF	R3, 3.3 kΩ
C4, 32 µF/1.5 V	R4, 22 kΩ
C5, 47 pF	R5, 1 kΩ
C6, 0.005 µF	D, diode
C7, 10 µF/3 V	TR, OC42

L1 and L2, RFC (choke coils) (see text)

Specification for the RFC is 1 or 1.5 millihenries (mH). A suitable coil can be made by winding on a standard ¼-in. diameter coil former, using 38-s.w.g. enamelled wire. Cut two ¾-in. diameter cheeks from thin Paxolin or similar insulating material to fit together on the coil former, spacing them ⅛ in. apart–Fig. 9.7. The space between the cheeks is then completely filled with a winding of 38-s.w.g. insulated wire, winding on one layer with turns adjacent, then a second layer over the top, and so on until the coil has built up to the same diameter as the cheeks. The coil windings can be held with a wrapping of cellulose tape and the ends

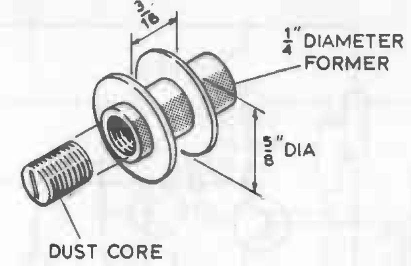

Fig. 9.7 Coil former for winding the RFC coils.

should be secured to one of the cheeks with sealing wax or something similar. The provision of an iron-dust core to fit the centre of the coil core enables the inductance of the completed coil to be varied by screwing in and out, thus taking up any differences introduced in the actual winding; in other words, the dust core forms a trimming adjustment.

All the other components used are familiar, standard types. High-impedance headphones, or a high-impedance magnetic type earpiece must be used with this circuit.

The same circuit can be modified for regenerative working as well (regenerative-reflex receiver), simply by connecting a 10-pF trimmer

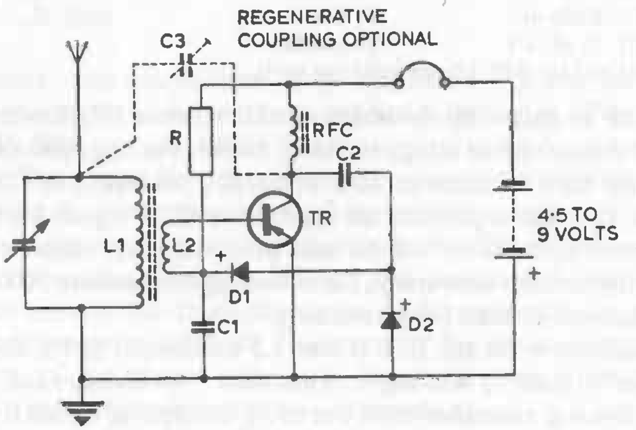

Fig. 9.8 *Circuit 19.* Single-transistor reflex receiver using two diodes.

C1, 0.01 µF	TR, OC44 (or equivalent)
C2, 220 pF	
C3, 20 pF	
R, 100 k	

capacitor between the collector of the transistor and the 'hot' end of the tuned circuit. This capacitor is adjusted to give optimum regenerative coupling.

In areas where radio reception is generally good, either circuit (reflex or regenerative-reflex) if assembled correctly should give good listening strength without the use of an external aerial.

It is possible to simplify the single-transistor three-stage reflex receiver by using two diodes in the circuit instead of one. A typical circuit of this type is shown in Fig. 9.8. Again the circuit may be converted to regenerative-reflex working by connecting a 10-pF or

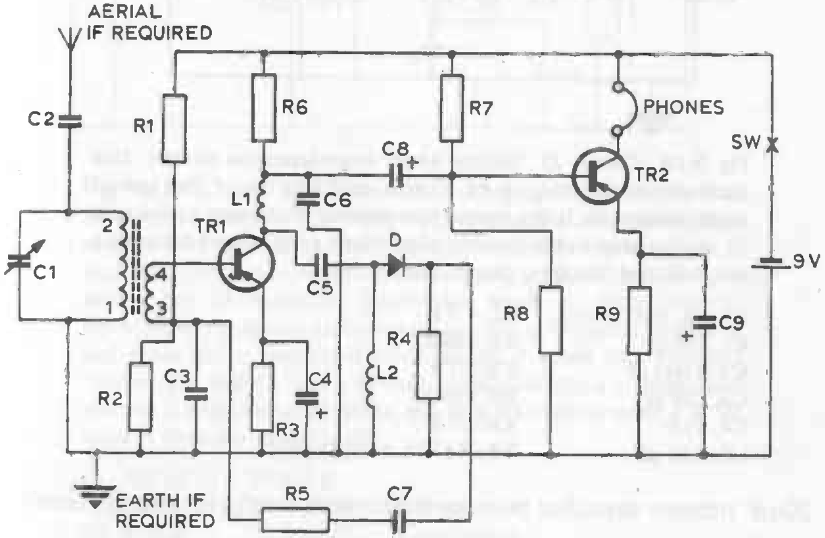

Fig. 9.9 *Circuit 20.* Four-stage reflex receiver. This is an extended version of the design given in Fig. 9.6–the same basic circuit with the addition of an AF amplifier stage. Using low-power transistors, output power is matched to high-impedance phones.

C1, 500 pF	D, any crystal diode	R1, 22 k
C2, 220 pF	L1, L2, choke coils	R2, 4.7 k
C3, 0.01 µF	(see text)	R3, 3.3 k
C4, 32 µF	TR1, OC44, OC45, XA102,	R4, 22 k
C5, 47 pF	or near equivalent	R5, 1 k
C6, 0.005 µF	TR2, OC71	R6, 4.7 k
C7, 10 µF		R7, 22 k
C8, 8 µF		R8, 10 k
C9, 8 µF		R9, 4.7 k

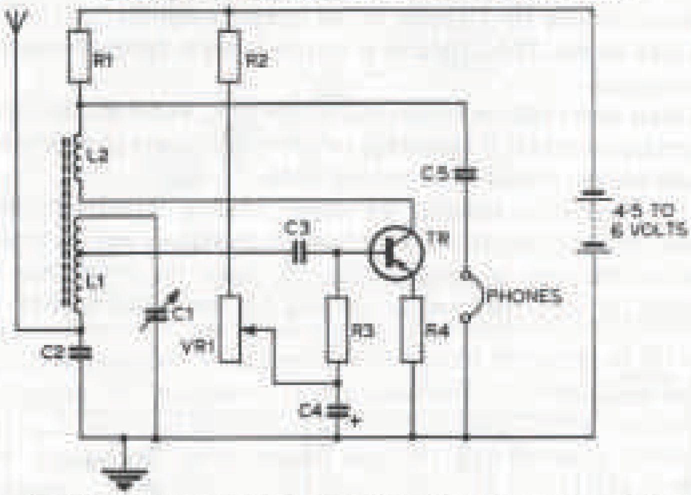

Fig. 9.10 *Circuit 21.* 'Mighty Atom' single-transistor receiver. This particular circuit design is by Weyrad, matching one of their special regen aerial coils. It is a regenerative receiver which uses a minimum of components and is capable of excellent performance for such a simple circuit. Matching phones are 50–100 Ω.

L1, L2, Weyrad coil
C1, 350 pF
C2, 0.001 μF
C3, 220 pF
C4, 8 μF
C5, 0.01 μF

R1, 4.7 k
R2, 100 k
R3, 4.7 k
R4, 100 k
VR1, 5 k
TR, XA103, or equivalent

20-pF trimmer capacitor between the transistor collector and the tuned circuit.

The output from the reflex stage may be fed directly into headphones (three-stage circuit); or may be followed by a stage of normal AF amplification (four-stage circuit–see Fig. 9.9); or further stages of amplification (five-stage circuit), etc, all the additional stages being AF amplifier stages.

As with the simpler circuits, however, there is a practical limit to the number of amplification stages which can usefully be added. The basic three-stage circuit (with no AF amplifier) should give audible headphone reception on a number of different stations; the four-stage circuit should give good headphone reception with local stations very loud; or equally be capable of giving good volume with a loudspeaker, depending on choice of transistors for the amplifier stages.

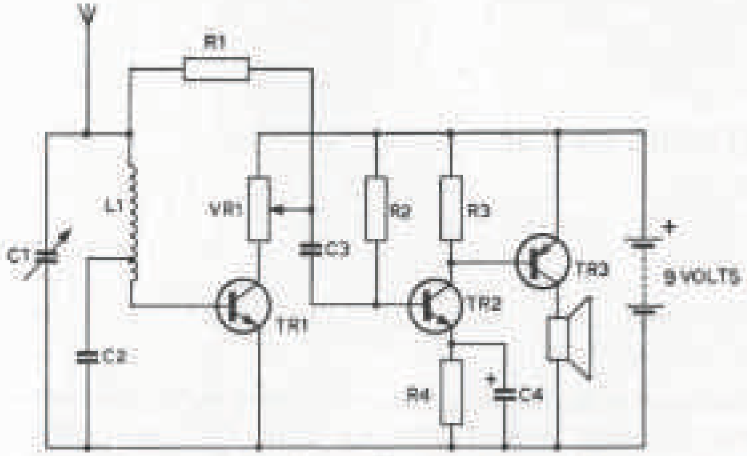

Fig. 9.11 *Circuit 22.* Three-transistor regenerative receiver with loudspeaker output. This design of regenerative receiver employs resistive feedback rather than capacity feedback. The advantage is that the regeneration, once adjusted, should remain at an optimum setting for all broadcast frequencies tuned in. With capacity regeneration, optimum adjustment is usually to a specific frequency and needs to be readjusted when the set is tuned to a different station. The coil (L1) is a 90-turn winding of 32-s.w.g. enamelled wire on a $\frac{3}{8}$-in. diameter ferrite rod with the tapping point 8 to 10 turns in from the 'earthy' end.

C1, 350 pF
C2, 0.05 μF
C3, 0.05 μF
C4, 25 μF
R1, 390 k
R2, 2.2 M
R3, 100
R4, 330

VR1, 5 k
TR1, 2N2926, or equivalent
TR2, 2N2926, or equivalent
TR3, AC128, or equivalent
Loudspeaker, 3 to 8 Ω

10 SUPERHETS

The superheterodyne or *superhet* receiver works on a more complicated principle than other types and involves more stages and more components. The RF signal from the tuned circuit is combined with another HF signal generated by a *local oscillator* in the receiver itself in a *mixer* or *convertor* stage. This stage then has an output which is a 'beat' frequency, equal to the *difference* between the two signals, this 'difference' frequency being called the *intermediate frequency* (IF).

The IF output from the mixer is then amplified (in one or more stages) before being passed to the second detector stage, where the modulated IF is converted to AF and subjected to further stages of amplification. The second HF signal is generated within the receiver itself by the HF or local oscillator, applied to the mixer. A block

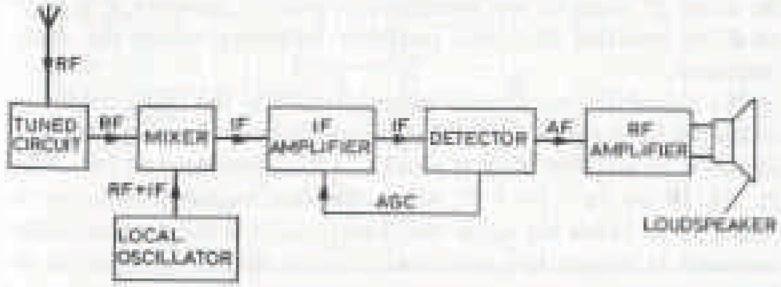

Fig. 10.1 Block diagram of a superhet circuit. Feedback of a proportion of the signal after detection is usually necessary to reduce gain on strong signals and prevent distortion through overloading. This is known as automatic gain control (AGC).

diagram is shown in Fig. 10.1, where it will be noticed that RF amplification can again be applied after the tuned circuit, if required.

The principal advantages offered by the superhet are that the gain and selectivity obtained from the IF amplifier do not depend on the frequency of the signal as they do with 'straight' receivers. Also the IF can be made lower than the signal frequency, resulting in higher stage gain and a narrower response curve. Again, too, since the IF amplifier is dealing with a constant frequency it can dispense with variable capacitors, reducing the risk of unwanted feedback and generally making the receiver less critical in design. The main disadvantages are the higher cost of such a circuit and the necessity of achieving proper alignment.

Most commercial domestic receivers are of the superhet type because of the superior performance offered and, although rather more complex, the type is equally suited to amateur construction.

The usual IF employed is between 450 and 475 kHz (455 kHz is more or less standard in this country) and the IF amplifier is tuned to the specific IF employed; this IF amplifier thus employs one or more stages of RF amplification with fixed tuning. A little thought will show that to 'feed' the IF amplifier correctly it will be necessary to vary the local oscillator in step with the tuned circuit. Thus if the receiver is designed to tune over a range of from, say, 800 kHz to 1,800 kHz and the IF is 470 kHz, then the oscillator frequency must range from 800 + 470 = 1,270 kHz to 1,800 + 470 = 2,270 kHz. Equally, to produce the same IF response the local oscillator could be set to 800 – 470 = 330 kHz to 1,800 – 470 = 1,330 kHz. The result would be the same as far as the IF amplifier was concerned; in practice, however, it is more usual to establish the local oscillator frequency above the signal frequency.

The converter thus actually receives two RFs: the signal frequency which can be designated Fs and the local oscillator frequency Fo. The following IF amplifier stage is tuned to a fixed frequency of Fo – Fs = Fi, the IF (or Fs – Fo = Fi, where the local oscillator frequency is chosen to be below the signal frequency); and it is the purpose of the converter to supply this particular IF as its output by a process of frequency conversion.

In fact, any oscillator HF will cause IF response at two signal frequencies, one equal to Fo – Fs and the other equal to Fo + Fs. One will be the real signal required, e.g. the correct value of Fs for the station selected, whilst the other is an undesirable *image* signal.

To clarify this by a numerical example, suppose the desired signal frequency is 800 kHz and the IF 450 kHz with the local oscillator set above the input signal frequency. To produce the necessary IF output at 800 kHz signal input the local oscillator will have to be set for 800 + 450 = 1,250 kHz. Exactly the same IF output will be produced, however, if a 1,250 + 450 = 1,700 kHz input signal is received, since the result of mixing will again be a frequency of 1,700 – 450 = 1,250 kHz. An essential feature of satisfactory working, therefore, is good *selectivity* in order to reduce the response to image signals to negligible effects.

The signal-to-image ratio, or image ratio as it is usually called, depends on the selectivity of the RF tuned circuits preceding the mixer or converter. At the same time the higher the IF the higher the image ratio, since raising the IF increases the frequency separation between signal and image and places the latter farther away from the peak of the resonance curves of the signal-frequency circuits.

The less experienced amateur–and certainly the beginner–is usually reluctant to tackle superhet construction because of the many complexities and large number of components involved, particularly in alignment, to get the circuit working satisfactorily. Probably the best approach, until experience is gained with superhet behaviour, is to build from kits, which include pre-aligned IF transformers. However, it is

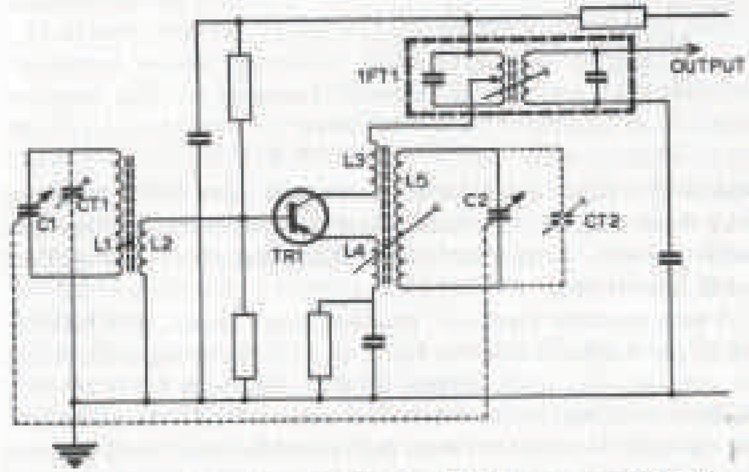

Fig. 10.2 Typical superhet 'front end', featuring conventional tuned circuit, local oscillator and mixer, and also one stage of IF amplification.

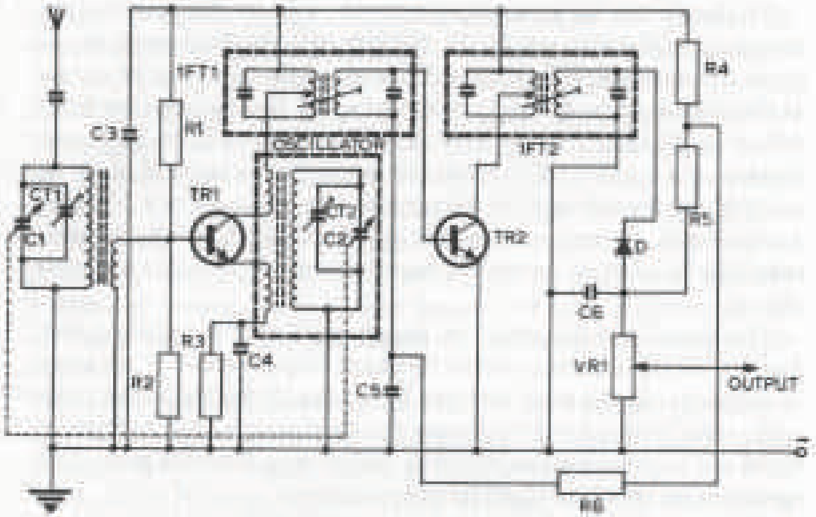

Fig. 10.3 *Circuit 23.* Complete superhet circuit up to output stage–i.e. to be followed by any suitable output circuit. This design uses n-p-n transistors.

TR1, 2N2924
TR2, 2N2924
R1, 330 k
R2, 2.7 k or higher
R3, 2.2 k
R4, 100 k
R5, 39 k
R6, 33 k

VR1, 5-k potentiometer
(volume control)
C1, C2, ganged tuning capacitor
C3, 5 or 8 μF
C4, 0.01 μF
C5, 5 μF
C6, 0.01 μF
1FT1 and 1FT2, miniature
IF transformers
0A70 or equivalent diode

possible to produce a working superhet circuit based on two transistors and a diode detector, with suitably selected components for the critical circuit elements–notably the ganged tuning capacitors, local oscillator and IF transformers.

A basic superhet 'front end' including tuned circuit, local oscillator and IF transformer is shown in Fig. 10.2. L1 is the tuning coil and L2 is the coupling coil (each wound on a ferrite rod); L5 is the local oscillator coil, inductively coupled to transistor TR1 via L3 and L4. The resonant frequencies of these two circuits are tuned simultaneously by the twin-ganged variable capacitors C1 and C2. In a practical circuit each variable capacitor would have its own trimmer capacitor, CT1 and CT2 respectively, for fine trimming.

Transistor TR1 is the mixer or *autodyne frequency changer,* the IF signal being selected by the IF transformer 1FT1.

The following sections, fed from 1FT1, use TR1 as an IF amplifier, and also provide the AF stage to feed the diode detector via a second IF transformer 1FT2. The whole of this circuit is shown in Fig. 10.3. An additional stage of IF amplification could be included, and also a more sophisticated output stage. Common practice is to follow a single stage of AF amplification by a push–pull output, with the driver transistor selected for high gain, and the output load formed by the primary of a phase-splitting transformer. However, keeping the number of components involved to a minimum and using only two IF transformers, considerably simplifies both construction and alignment.

Correct tracking of the aerial- and oscillator-tuned circuits to maintain the constant IF difference can be obtained by using a tuning capacitor with specially shaped vanes for the oscillator circuit; or by a conventional padded capacitor. Stray capacitance between the aerial and oscillator sections of the tuning capacitor can form a path for unwanted feedback between the two sections, so a screen is usually placed between them to prevent this. These effects can also arise between the wires connecting to the wavechange switch, so these must be kept as short as possible (or if completely incorporated on a printed circuit, the circuit elements kept short and well spaced to eliminate capacity effects).

In all cases, regardless of the type of oscillator and mixer, the first IF transformer receives a mixture of local oscillator frequency, signal frequency and the sum and difference of the two. It is therefore necessary to use a capacitor across the primary to produce an LC circuit to resonate at the IF. This capacitor is of fixed value (matched to the inductance of the coil), when any adjustment required for alignment is provided by an iron-dust core in the coil enabling the inductance to be varied over the necessary range. Once the overall LC value has been adjusted to a resonant frequency corresponding to the IF the tuning remains fixed.

Output from this stage is an AF modulated HF signal at the IF, which has also been subject to amplification.

Alignment with simple circuits should be quite straightforward and can be done without the aid of a signal generator. Particular care, however, must be taken not to overtighten the tuning slugs in miniature coils as these are easily jammed, resulting in permanent damage to the coil since it will be virtually impossible to remove the core without first removing the coil from the printed circuit; and equally almost impossible to remove the coil without damage.

To align without a signal generator, the set is tuned to the mid-way point on the long waveband and the core of L4 adjusted to receive BBC Radio 2. The tuning control should then be adjusted to the higher wavelength end of the medium waveband and the cores of 1FT1 and 1FT2 adjusted for maximum output. Further adjustment of volume can then be made by sliding the aerial coil along the ferrite slab, and once the best position is found the sleeve should be fixed with wax or cellulose tape. The receiver should then be tuned to 300 m approximately and capacitor C1 or its trimmer CT1 adjusted until an identifiable station is received and then brought up to maximum volume by adjustment of capacitor C2 or its trimmer CT2. If necessary, the alignment stages should be repeated for the medium waveband until all stations are received at good signal strength.

Using a signal generator, alignment is even simpler and more positive. The basic procedure should be:

(i) Injected signal, 600 kHz, adjust tuning to correspond to the end of the (medium) waveband.

(ii) Injected signal, 470 kHz, adjust 1FT1 and 1FT2 cores for maximum output.

(iii) Injected signal, 600 kHz, adjust coil position on ferrite rod for best results.

(iv) Injected signal, 1,450 kHz, adjust trimmers CT1 and CT2 for maximum output.

Mullard 6-transistor receiver

This design is for a simple portable receiver using alloy-junction transistors and a single 9-V battery with transformerless push–pull output. The circuit is suitable for making as a medium-size portable receiver with a Ferroxcube rod aerial, or in miniature size if preferred. The standard IF of 470 kHz is used with the local oscillator frequency above the signal frequency. One OC44 and two OC45 transistors are used in the mixer and IF stages. The detector is a germanium diode type OA70. The audio stages comprise one OC71 transistor driving a matched pair of OC72 transistors.

The OC44 operates as a self-oscillating mixer, with RF signals from the aerial coupling coil fed into the base of the OC44, leading to the

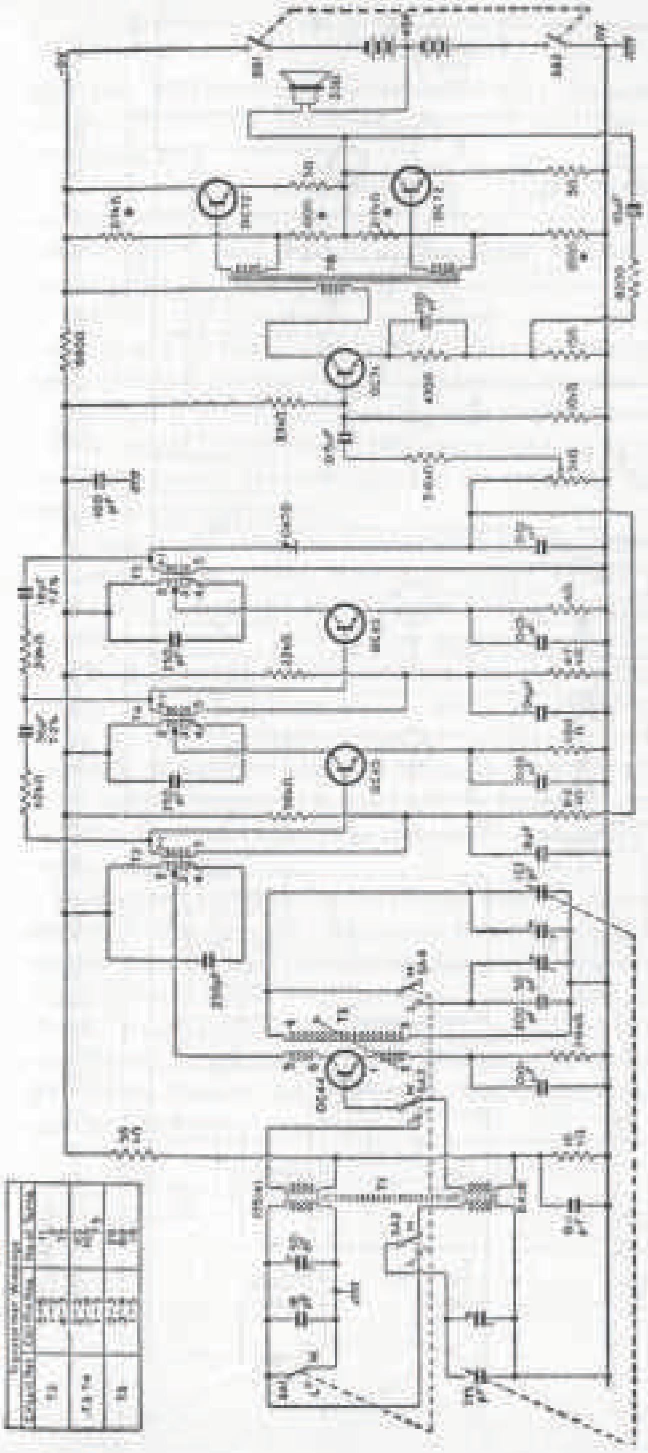

Fig. 10.4 *Circuit 24.* Mullard design for six-transistor superhet. Tolerance of resistors marked with an asterisk should be 5 per cent, others 10 per cent.

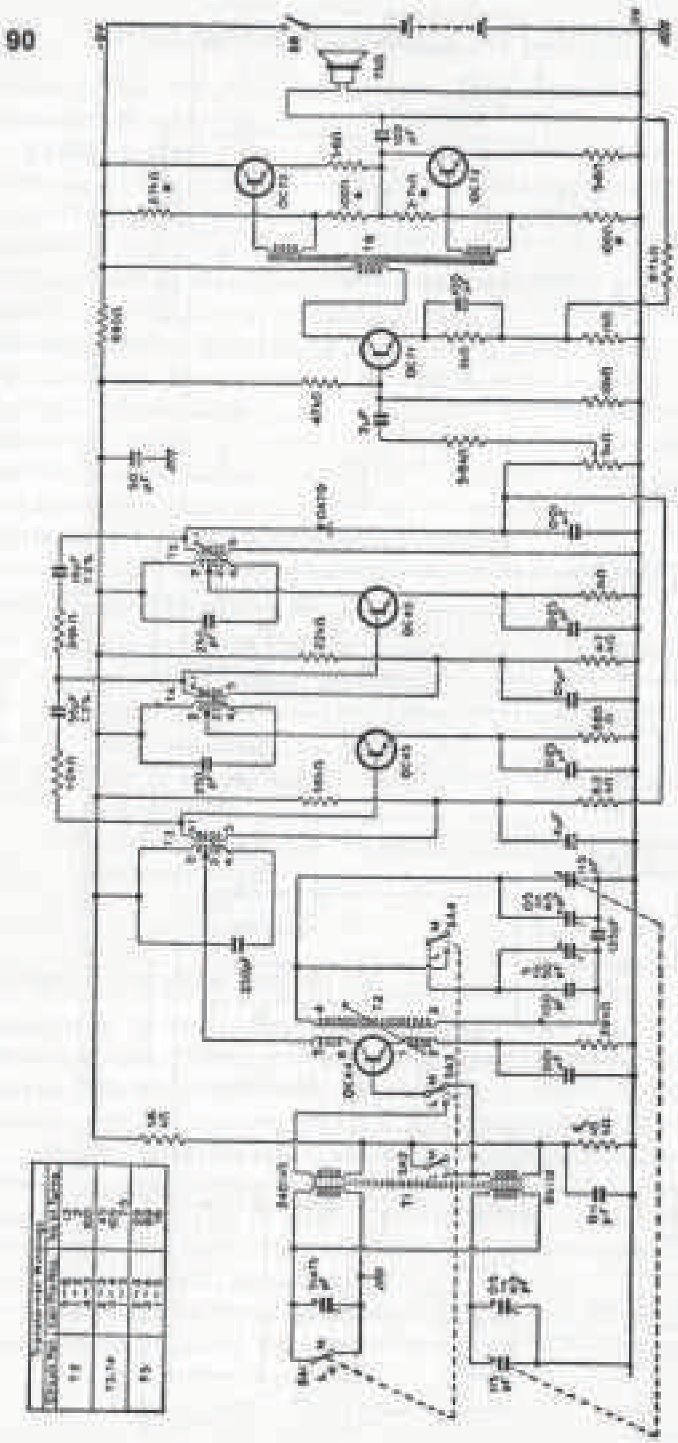

Fig. 10.5 Circuit 25. Mullard design miniature superhet. Tolerance of resistors marked with an asterisk should be 5 per cent; others 10 per cent.

generation of local oscillation feedback from the collector to the emitter. The IF is selected at the collector of the OC44 by the first IF transformer T3.

The IF amplifier comprises two OC45 common-emitter circuits operating unilateralized, with the choice of bandwidth compromising between quality and selectivity. The third IF transformer T5 is connected to the OA70 diode detector, with the d.c. output fed back to the first IF transformer to provide automatic gain control. Double-tuned IF transformers are recommended for optimum performance as regards frequency response and image rejection.

The Class B transformerless output stage requires a loudspeaker with a 35-Ω speech coil to provide the correct load for an output of 200 mW, with negative feedback applied to the emitter of the OC71 driver from the loudspeaker terminal.

All component values are shown on the circuit diagram (Fig. 10.4). The value of the tuning capacitance is not critical, but must be sufficient to provide the desired frequency coverage. The aerial section has a capacitance of 175 pF and the oscillator section a capacitance of 125 pF. A screen should be used between the oscillator and aerial sections of the tuning capacitor (i.e. the tuning capacitor so specified) to be sure of eliminating undesired feedback in the circuit. The possibility of feedback will be at maximum when the receiver is tuned to its highest frequency, and will also be increased when the tuning capacitor has a low value (e.g. as typically the case with a miniature tuning capacitor).

A miniaturized version of the original circuit is shown in Fig. 10.5, again with all component values marked. Here the output is reduced to 100 mW driving a 75-Ω miniature speaker. Current consumption of this circuit is reduced from 9 mA zero signal, 20 mA average, to 7 mA and 13 mA, respectively. This miniaturized circuit permits of further modification by substituting a Class A output with transformer drive to the speaker, reducing the number of transistors required to five, but increasing the battery consumption.

11 COMPONENTS

The following descriptions are intended to familiarize readers with the various basic components used in radio construction, as well as being a useful reference when selecting components for making a particular circuit.

Resistors

The first thing to remember about resistors is that their *size* is no indication of their actual resistance value; rather it is a rough indication of the *power rating*. Transistor circuits operate with low power ratings and so small (physical) resistors are commonly employed.

Actual *sizes* of resistors range from 4 mm long by 1.2 mm diameter (with a $\frac{1}{20}$-W power rating), up to about 50 mm long by 6 mm diameter (with a 7-W power rating); and even larger in the old-fashioned type of carbon rod resistors. A comparison of modern resistor sizes is given in Fig. 11.1; this diagram is also a useful method of identifying the various *types* of resistors by their shape.

There are five main *types* of resistors, classified by their construction.

(i) *Moulded-carbon type*, literally a rod made from a mixture of carbon and binder fired into a rigid form. This rod is usually protected with a lacquer coating, or a paper or ceramic sleeve. These are the old-fashioned type which are still available (usually as 'surplus' stock) and are suitable for any non-critical circuit. Ratings are usually $\frac{1}{4}$ W, $\frac{1}{2}$ W and 1 W.

(ii) *High-stability carbon resistors*, much smaller and compact, made from carbon film. These are miniature and sub-miniature in size and are

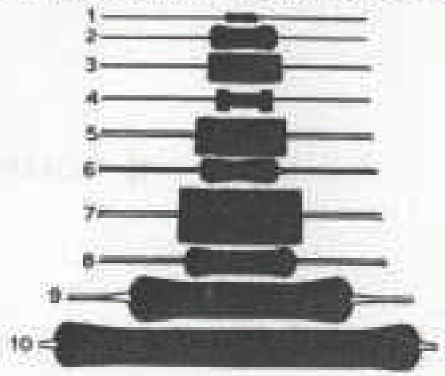

Fig. 11.1 Modern resistors.

1, $\frac{1}{20}$ W carbon-film	6, $\frac{1}{2}$ W metal-oxide
2, $\frac{1}{8}$ W carbon-film	7, 1 W carbon-film
3, $\frac{1}{4}$ W carbon-film	8, 1 W wire-wound
4, $\frac{1}{10}$ W metal-film	9, 3 W wire-wound
5, $\frac{1}{3}$ W carbon-film	10, 7 W wire-wound

the preferred type for general use because of their good stability, low noise, and quite low cost. Ratings are usually $\frac{1}{20}$ W, $\frac{1}{8}$ W, $\frac{1}{4}$ W, $\frac{1}{2}$ W, $\frac{3}{4}$ W and 1 W.

(iii) *Metal-film resistors*, which are made by depositing a film of nickel-chromium on a high-grade ceramic body and then cutting a helical track through the film to produce the required resistance values. End caps are fitted to carry the leads and the resistor body is protected by a lacquer coating. These are much more stable than carbon resistors but cost about four times as much. Ratings are from $\frac{1}{10}$ W upwards.

(iv) *Metal-oxide resistors*, made by depositing a film of tin oxide on a special glass rod, the whole being subsequently covered with a heat-resistant coating. Stability is again very high and this type is virtually proof against damage through overheating (e.g. when soldering) and is also unaffected by dampness. The usual rating is $\frac{1}{2}$ W.

(v) *Wire-wound types*, generally only required for special circuits where very low resistance values are required and/or very high currents have to be carried. Typical ratings may range from 1 to 5 W for a 0.5-Ω resistor to 25 W or more for higher resistance values.

Regardless of size, shape and type, the usual method of marking a resistor *value* is by coloured bands on the body, read in order 1, 2, 3 from the band nearest one end–Fig. 11.2. The equivalent value is determined by:

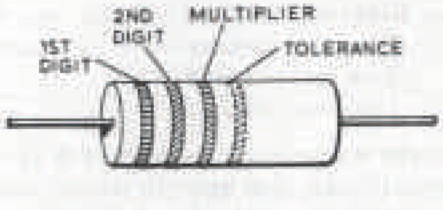

Fig. 11.2 Standard resistor colour coding.

Colour	1 gives first figure of resistance value	2 gives second figure of resistance value	3 gives number of noughts to put after first two figures
Black	0	0	None
Brown	1	1	0
Red	2	2	00
Orange	3	3	000
Yellow	4	4	0000
Green	5	5	00000
Blue	6	6	000000
Violet	7	7	0000000
Grey	8	8	00000000
White	9	9	000000000

Example: resistor colour code read as Brown, Blue, Orange.

	Brown	Blue	Orange
value read as	1	6	000

i.e. 16,000 Ω or 16 kΩ (kilohms).

On some old-type resistors the three colours are applied in a different manner–a coloured body, coloured tip and coloured spot on the body. The code is then read in the order body, tip, spot.

Many resistors have an additional (fourth) coloured band, which expresses the *tolerance* on the resistor value. The equivalent tolerances so indicated are:

Silver band	10 per cent tolerance
Gold band	5 per cent tolerance
Red band	2 per cent tolerance
Brown band	1 per cent tolerance

For most circuits resistors with a gold band (5 per cent tolerance) are the recommended choice; silver band (10 per cent tolerance) are good enough for general use. Finer tolerances (red or brown band) are not necessary, except for critical circuits. No fourth band implies a tolerance of 20 per cent.

Other variations which may be found on colour coding are:

a salmon pink band indicating a high-stability type;
a *double ring* of the first colour band, indicating a wire-wound resistor.

Some types of modern resistors are not colour coded but the values are marked directly on to them in numbers and letters. *Numbers* indicate numerical values and *letters* indicate multipliers, viz:

R = × 1
K = × 1,000
M = × 1,000,000

Tolerances are also given by a second letter, viz:

M = 20 per cent
K = 10 per cent
J = 5 per cent
H = 2.5 per cent
G = 2 per cent
F = 1 per cent

The first letter ('multiplier') comes in a position which determines the decimal point. Thus:

rating *5R* designates 5 × 1 = 5 Ω
4K7 designates 4 × 1,000 and 0.7 or 4.7 kΩ
82K designates 82 × 1,000 or 82 kΩ

The tolerance letter comes at the end, e.g.

4K7J means a 4.7-kΩ resistor with a 5 per cent tolerance

Actual resistor *values* are also normally based on *preferred numbers* and not simple arithmetical steps. The reason for this is that preferred number steps give approximately constant *percentage* change in

resistance from one value to the next. The usual resistance values obtainable are thus:

Ω	kΩ	MΩ
10*	1.0	1.0
12	1.2	1.2
15	1.5	1.5
18	1.8	1.8
22	2.2	2.2
27	2.7	2.7
33	3.3	3.3
39	3.9	3.9
47	4.7	4.7
56	5.6	5.6
68	6.8	6.8
82	8.2	8.2
100	↓	
120		
150		
180		
220		
270		
330		
390		
470		
560		
680		
820		

*Typical values for *wire-wound* resistors below 10 Ω are: 0.33, 0.5, 1, 1.5, 2, 3, 3.3, 3.9, 4, 4.7, 5, 5.6, 6, 6.8 and 8.

Variable resistors (potentiometers)

Variable resistors, or *potentiometers*, are normally constructed from a resistive element, formed into a 270° arc, with a wiper arm connected to it, and turned by a central spindle. There are three external terminal tags; the outer ones connect to the ends of the resistance element and the centre tag to the wiper—see Fig. 11.3.

The resistance element may be a carbon track or a winding of resistance wire (wire-wound potentiometers). Carbon-track potentiometers are cheaper than wire-wound types and are suitable for most general circuit applications, at low power levels, e.g. ¼ to ½ W for low

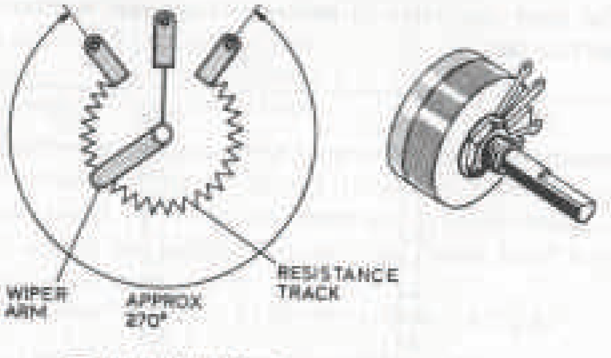

Fig. 11.3 Variable resistor or potentiometer.

resistance values, reducing with higher resistance values. Wire-wound potentiometers have a higher power rating–usually of the order of 1 to 3 W continuous for the whole track; they are also made in lower resistance values obtainable with carbon-track potentiometers.

The usual resistance values available in each type are:

Carbon-track: 100, 220, 470, 1 k, 2.2 k, 4.7 k, 10 k, 22 k, 47 k, 100 k, 220 k, 470 k, 1 M, 2.2 M and 4.7 M.

Wire-wound: 10, 22, 47, 100, 220, 470, 1 k, 2.2 k, 4.7 k, 10 k, 22 k, and 47 k.

The general appearance of typical carbon-track and wire-wound potentiometers is shown in Fig. 11.4; they are also made with a sliding

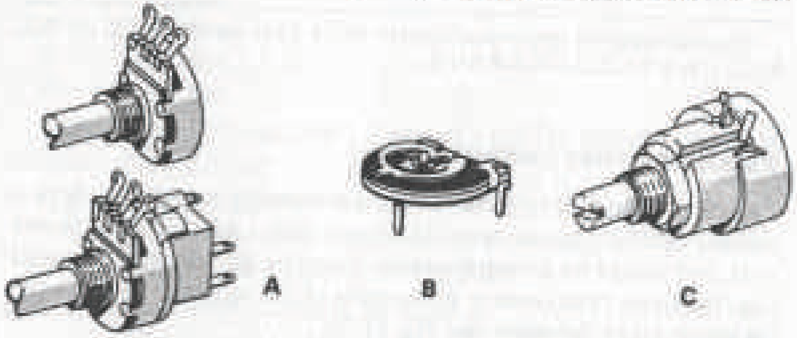

Fig. 11.4 Potentiometers
A, Typical small carbon-track, about 20 mm diameter;
B, Open-type carbon-track linear potentiometer, about 15 mm diameter;
C, Typical wire-wound potentiometer, diameter 25 to 35 mm.

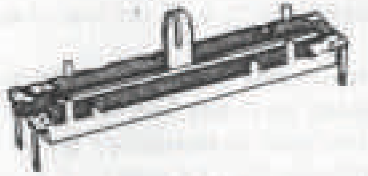

Fig. 11.5 Modern slider-type carbon-track potentiometer.

rather than a rotary movement. The old type consisted of a wire winding on a suitable cylindrical former, traversed by a sliding contact assembly. The modern miniature slider-potentiometer is much more compact and fully enclosed, except for the slot in which the slider arm travels–Fig. 11.5. The resistance track may be carbon, or wire-wound. Slider-type variable resistors have come into favour for volume controls on modern radio designs.

Potentiometers are also described as linear or non-linear, this referring to the manner in which resistance values vary with wiper movement and *not* whether the potentiometer is a slider or rotary-action type. With a *linear* potentiometer the resistance change is directly proportional to the actual movement, *e.g.* a 50 per cent

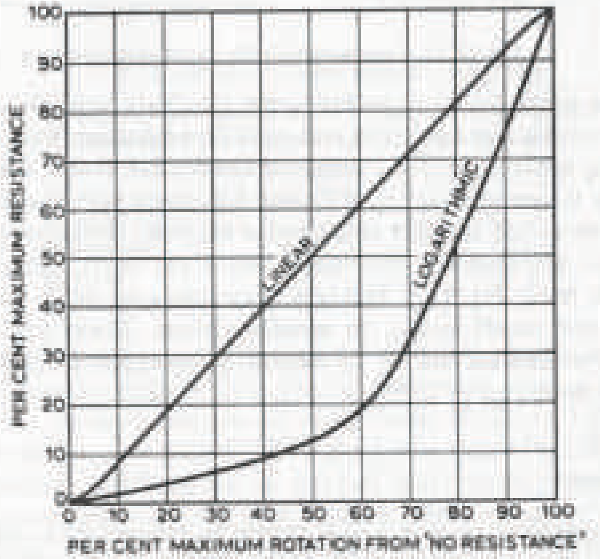

Fig. 11.6 Linear and logarithmic potentiometer characteristics.

movement will correspond to a 50 per cent change in resistance—see Fig. 11.6.

Slide-type potentiometers normally automatically provide linear characteristics. Rotary types may have linear or *logarithmic* characteristics, the latter producing an increasingly greater change in resistance with spindle rotation (see Fig. 11.6 again).

Potentiometers may also be designed with characteristics intermediate between linear and logarithmic, e.g. semi-log or linear-tapered; and also with inverse characteristics, e.g. anti-log. Either linear or logarithmic potentiometers are suitable for most volume control duties etc., although logarithmic types are generally preferred.

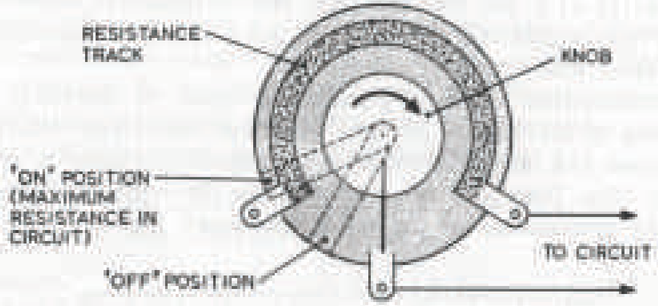

Fig. 11.7 Potentiometer with on-off switching.

Wiring connections to a potentiometer are simple, remembering that the two outside tags connect to each end of the resistance track and the centre tag to the wiper. For a simple volume control, connection is thus normally to one end tag and the centre tag. Some potentiometers are made with an 'off' location at one end of the wiper movement (usually maximum anti-clockwise movement) when the wiper runs off the resistance track—Fig. 11.7. This type can combine the duties of volume control and on-off switch, in suitable circuits. Power supplies are usually switched on and off by a switch mechanically coupled to the volume control.

Capacitors

Modern capacitors come in a wide variety of different types, sizes and constructions, many having overlapping values. The older *paper tubular* type, constructed of silver foil interleaved with mica sheets, have largely

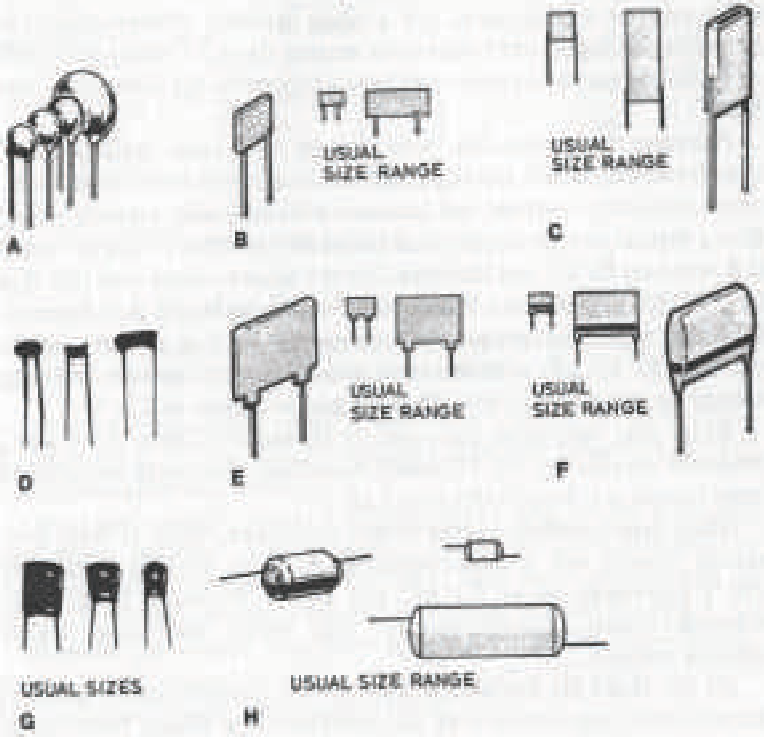

Fig. 11.8 Examples of modern miniature capacitors

A, Ceramic disc 1,000 pF to 0.1 μF (approx. actual size);
B, Ceramic plate, sizes from 4 × 2.5 mm to 12 × 5 mm;
C, Ceramic plate tubular;
D, Tubular ceramic, typical sizes;
E, Silvered mica;
F, Polyester;
G, Mylar;
H, Polystyrene.

been replaced by miniaturized ones, although they continue to be used in the capacity 0.05 to 2 μF (e.g. in preference to an electrolytic capacitor). Working voltages are of the order of 200 to 1,000 V for paper types, or up to 10,000 V with plastic-impregnated types.

Miniature capacitors are of ceramic, polystyrene, silver mica or mylar construction—see Fig. 11.8. *Ceramic capacitors* may be of 0.1 μF (but with working voltages generally substantially reduced above 5,000 pF capacity). *Tubular ceramic capacitors* have a similar value range and are usually rated for a mean working voltage of 350 V. *Silvered mica capacitors* with values ranging from 2.2 pF up to 10,000 pF (0.01 μF) are rather larger and have a higher voltage rating (typically 500 V).

Polyester capacitors are produced in two main configurations, cylindrical with offset end leads, and slab shape with radial leads for PC board mounting. Physical size increases with increasing capacity rating: thus a typical size for a cylindrical polyester capacitor of 220 pF or less is 8 mm long by 3.5 mm diameter. For the highest values with this type (e.g. 22,000 pF) size can be up to 32 mm long by 12 mm diameter. Polyester capacitors are usually constructed for high-capacity values, e.g. 0.01 to 2.2 μF, with size again increasing with capacity. Working voltage for these is normally 125 V for smaller values, or 250 V.

Mylar film capacitors are similarly constructed, with radial leads, produced specifically for PC board mounting. The range of values is more limited, e.g. from 0.001 to 0.2 μF.

Other types include *polycarbonate capacitors*, made in both rectangular ('box') and tubular configurations, with working voltages of 250 V for values up to 0.1 μF, and 100 V for values up to 1 μF, although special constructions for high values can accept higher working voltages.

All the above are *non-polarized capacitors*, meaning that it does not matter which way round they are connected in a circuit. Other types are *polarized* and must always be connected in a specific manner ('plus to plus'), otherwise they will be degraded and eventually destroyed.

The best known type of polarized capacitor is the *electrolytic capacitor*, originally based on an aluminium rod or foil plate in contact with a paste electrolyte. The capacitor is then formed by passing a direct current through it during manufacture, depositing a thin film of aluminium oxide on the aluminium surface by electrochemical action. This type of construction allows quite high values of capacitance to be produced in a relatively compact envelope, e.g. up to 1,000 μF. *Etched foil* electrolytic capacitors permit even greater capacitance values to be achieved for the same size, and so make it possible to produce miniature types with high capacity values. In the latter case the working voltage is generally reduced.

Modern electrolytic capacitors are usually of etched foil type, or the more recently developed tantalum construction, which are produced in both conventional cylindrical configuration with axial leads and *tantalum bead* form–see Fig. 11.9. Both types have low voltage ratings,

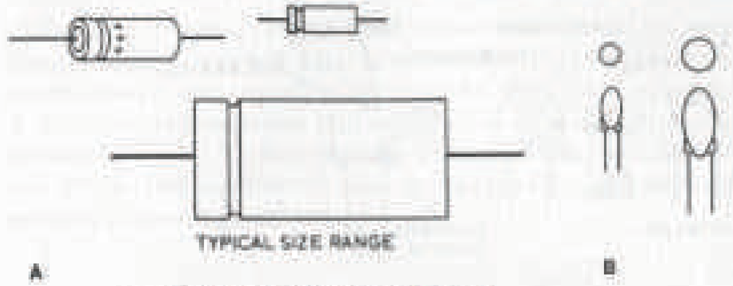

Fig. 11.9 Electrolytic capacitors

A, Metal-cased cylindrical with axial leads;
B, Tantalum bead, typical size range.

decreasing with increasing capacity value, e.g. from 35 to 40 V for 0.1 μF capacity; to 3 V at 100 μF capacity. *Tantalum* capacitors are also made in non-polarized form.

Values are normally marked directly on the body of the capacitor; some types, however, may be colour coded, the code adopted not necessarily being universal (it can vary from one manufacturer to another and from country to country in the case of older products).

Polyester slab-shaped capacitors are normally colour coded instead of marked, the colours following the same code as for resistors. Up to five bands can be used, the sequence being:

1st colour gives first significant digit in pF
2nd colour gives second significant digit in pF
3rd colour gives the multiplier
4th colour gives the *tolerance*, e.g. white 10 per cent, blue (sometimes black) 20 per cent
5th colour gives the voltage, red 250, yellow 400 V

The fourth and fifth colour bands may be missing; if only the fifth is missing, this implies a working voltage of 30 V.

Colour coding by rings may also be found on moulded-paper and some tubular ceramic capacitors. Some disc and slab-type capacitors may be colour coded by dots–see Fig. 11.10.

Variable capacitors

Variable capacitors are used for *tuning* or *trimming* to enable the capacitance in a critical circuit to be adjusted. There are physical

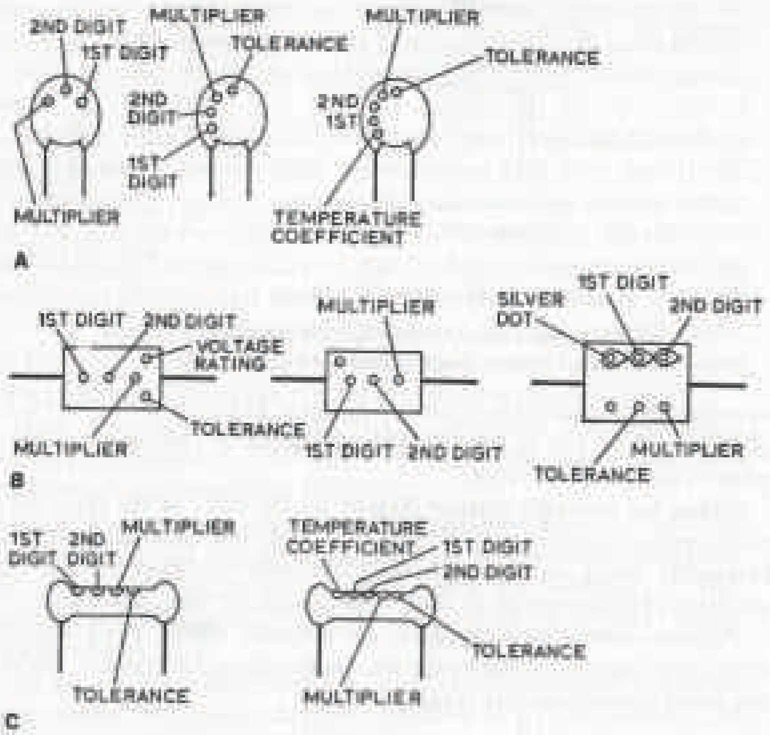
2ND DIGIT
1ST DIGIT
MULTIPLIER
MULTIPLIER
TOLERANCE
2ND DIGIT
1ST DIGIT
MULTIPLIER
TOLERANCE
2ND
1ST
TEMPERATURE COEFFICIENT
A
1ST DIGIT
2ND DIGIT
VOLTAGE RATING
MULTIPLIER
TOLERANCE
MULTIPLIER
1ST DIGIT
2ND DIGIT
SILVER DOT
1ST DIGIT
2ND DIGIT
MULTIPLIER
TOLERANCE
B
1ST DIGIT
2ND DIGIT
MULTIPLIER
TOLERANCE
TEMPERATURE COEFFICIENT
1ST DIGIT
2ND DIGIT
TOLERANCE
MULTIPLIER
C

Fig. 11.10 Some capacitor colour-dot codes

A, Disc ceramic;
B, Moulded paper;
C, Phenolic and ceramic.

differences between a tuning capacitor and a trimming capacitor, both in size and construction, but both work on the same principle and could, in fact, be used for either duty.

Tuning capacitors are commonly of the air-dielectric type, consisting of two sets of intermeshing metal vanes, one set fixed and the other rotatable by about 180 degrees by means of a spindle. When fully closed (i.e. the spindle is turned so that the rotating vanes are fully in mesh with the fixed vanes) capacitance is a minimum, and when fully opened (i.e. turned to full movement the opposite way) a maximum.

The maximum capacity of air-dielectric variable capacitors of this configuration ranges from about 10 pF to 1,000 pF. Physical size may range from 'miniature', e.g. about 25 mm square and 13 mm thick, up to 50 × 60 × 80 mm, or even larger, but is no indication of the working capacity range. It is largely a matter of convenience and compactness to use miniature tuning capacitors for transistor circuits and these may be of the solid-dielectric rather than air-dielectric type. Typical forms are illustrated in Fig. 11.11, showing single and ganged condensers, the latter used where two separate circuits are required to be tuned simultaneously from a single control.

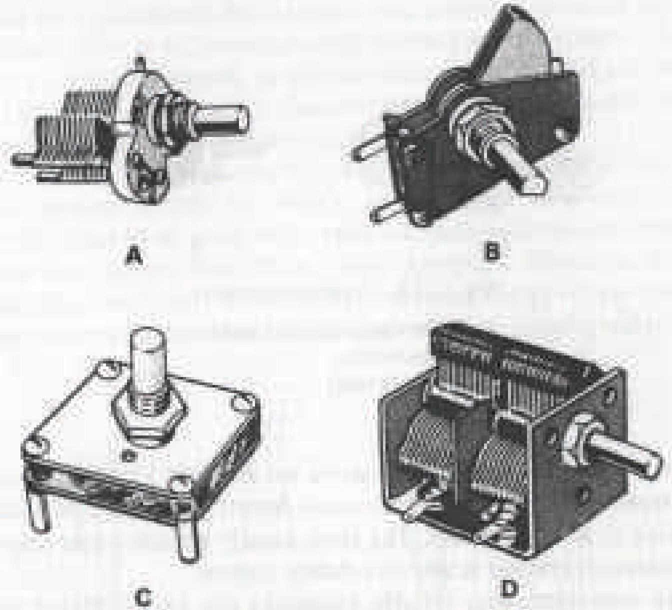

Fig. 11.11 Tuning capacitors

A, Air-dielectric;
B, Solid-dielectric;
C, Solid-dielectric, miniature;
D, Two-gang air-dielectric.

Trimming capacitors are usually based on a mica-dielectric and consist of two or more plates separated by very thin sheets of mica. Capacity is reduced by turning a central screw to adjust the pressure between plates and mica. They are designed primarily to be adjusted to a required capacity value and then left set in this position but can, however, be used as a tuning capacitor where their small physical size is an advantage (e.g. in a sub-miniature receiver); they are also cheaper than tuning capacitors.

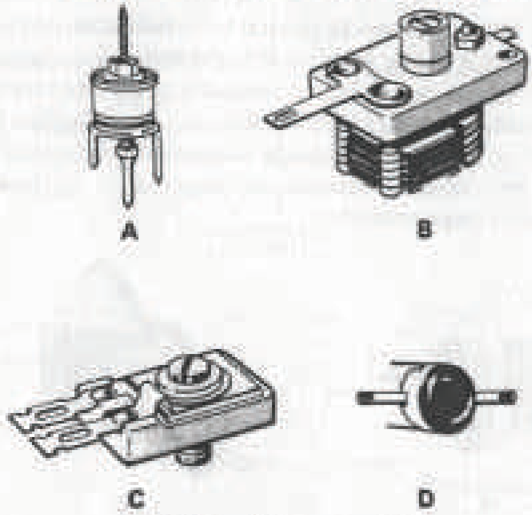

Fig. 11.12 Trimmer capacitors

A, Beehive type (air-dielectric);
B, Air-dielectric;
C, Mica-dielectric;
D, Ceramic.

Some forms of trimming capacitors are shown in Fig. 11.12. Not all such types have a mica-dielectric, some designs being based on a plastic dielectric or even air-spaced. The latter usually provide better response if a trimming capacitor is used as a *tuning* control.

Like potentiometers, variable capacitors can have different working characteristics, i.e. linear, logarithmic etc., depending primarily on the shape of the plates. These characteristics are usually:

(i) *Linear capacitance*, where each degree of spindle rotation produces an equal change in capacitance.
(ii) *Even frequency*, where each degree of spindle rotation produces an equal change in *frequency* in a tuned circuit.
(iii) *Square law*, where the change in capacitance is proportional to the *square* of the angle of movement.
(iv) *Logarithmic*, where each degree of spindle movement produces a constant *percentage* change in frequency.

Linear-frequency characteristics are usually preferred for a radio receiver tuning capacitor.

Transistors

Transistors are classified generally under two main types depending on the material used in their construction, germanium or silicon. Germanium transistors are the earlier type, but performance has been considerably improved by adoption of alloy-junction or alloy-diffusion methods of manufacture; silicon-junction transistors are normally made by the planar process. Where special characteristics are required (e.g. RF and IF applications), the manufacturing process is modified slightly to produce what is called an epitaxial (silicon-planar) transistor. There is also a third type known as a field-effect transistor (FET) in which both the construction and characteristics differ appreciably from germanium and silicon-junction types.

In general, germanium transistors are well suited to low-voltage AF circuits because of their low voltage losses and also because they can easily be matched in p-n-p/n-p-n pairs for push–pull output circuits; they are also cheaper than silicon-planar transistors. Silicon transistors, on the other hand, have lower leakage losses and higher voltage ratings, making them more suitable for higher output audio circuits, and for use

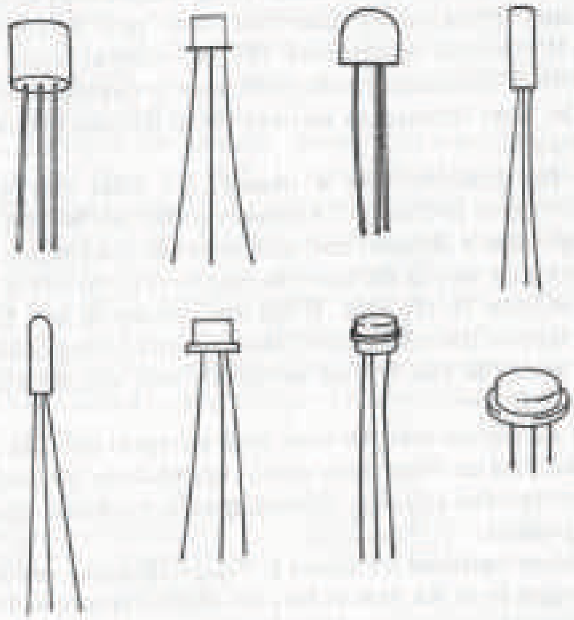

Fig. 11.13 Some transistor outline shapes.

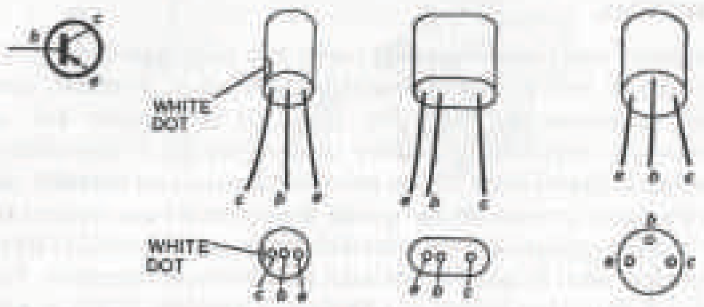

Fig. 11.14 Common transistor lead configurations.

at higher frequencies. They have the further advantage that they can operate successfully at temperatures up to 150°C or more, whereas the maximum working temperature for germanium transistors is normally below 100°C.

Transistors are basically three-element devices, the connections to the elements being in the form of three wires emerging from the bottom of the transistor, the whole being encapsulated in a metal or plastic case. Transistor outlines and sizes vary enormously (see Fig. 11.13), but the arrangement of the emergent wires normally follows one of the standard patterns to assist in identification—see Fig. 11.14. The three connections involved are to the *base* (*b*), *emitter* (*e*) and *collector* (*c*).

When the transistor base is circular, the wires may emerge in triangular form or in line. In the former case the two bottom points of the triangle lie in a diameter line, and viewed in this position from the bottom the apex wire is the base, the emitter is to the left of the base, and the collector to the right. If the wires emerge in line, a white or coloured spot on the can or body marks the end corresponding to the collector lead. The base is then the middle one, and the emitter the other end lead.

With a rectangular base the wires again emerge in line. The collector lead is identified as being more widely spaced from the middle wire (base) than the other end wire, although specific types may depart from this configuration.

Some other variations are shown in Fig. 11.15. Some may have *four* wires emerging from the base in line; the additional wire is the middle one connected to an internal shield (*s*). This does not form part of any connection to a circuit and is either ignored or connected externally to

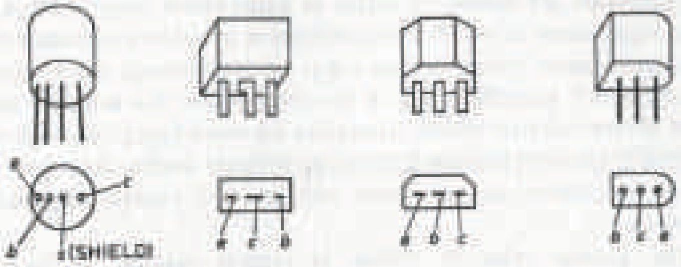

Fig. 11.15 Some other types of transistor shapes and lead configurations.

the case. Other types are designed specifically to plug into PC boards and have rigid terminal *strips* rather than wires—again connections can vary, so it is important to verify them for a particular type.

Power transistors are of more rugged construction and generally have different connections again. Some may have three wire connections, others just two terminal strips—see Fig. 11.16. In the latter case the collector connection is absent, the collector being connected internally to the case. Emitter and base leads are normally marked 'E' and 'B' respectively. The *collector* connection is then normally made to one of the screws mounting the transistor (or to an external terminal tag suitably connected to the case).

Power transistors are normally mounted on a *heat sink* to assist in dissipating heat from the case. It is usually necessary (except in a grounded collector circuit) to electrically insulate the transistor from this heat sink; at the same time this insulator must be a good conductor of heat. Suitable insulators are readily available to make heat sinks, consisting of a simple gasket or washer of mica, anodised aluminium or fabric. Fabric washers require greasing with silicon grease to make them heat-conducting. Some types of power transistors have tubular bodies when a finned cooling clip may be used to dissipate heat, instead of a heat sink.

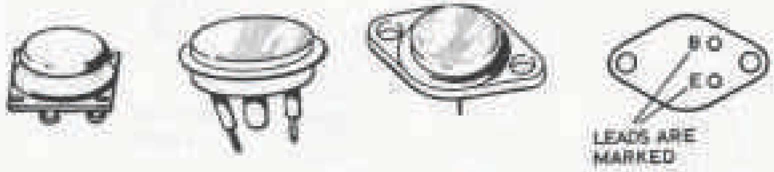

Fig. 11.16 Power transistors.

Transistors are further classified as being either p-n-p or n-p-n, the main significance of which is the difference in polarity of connections. Most germanium transistors are p-n-p types (although n-p-n ones are made for AF amplifiers and as the complement for matched pairs); most general-purpose silicon transistors are n-p-n types, although again p-n-p ones are also produced. If in the slightest doubt, the actual type should be checked as connection with reversed polarity can ruin the transistor.

The general rules to follow as regards polarity of transistor connections are:

with a p-N-p transistor the *collector* must be *Negative*;
with a n-P-n transistor the *collector* must be *Positive*.

12 CIRCUIT CONSTRUCTION

The general rule for designing a circuit layout is to follow, as far as possible, the same disposition and relative positioning of the individual components as on the circuit diagram. This should eliminate over-long wires and avoid unnecessary crossing over of leads.

Tag strips form a convenient method of assembling simple circuits involving a relatively small number of components. Two tag strips, mounted an inch or so apart on a Paxolin panel, enable individual components to be soldered between pairs of tags with common connecting leads where necessary—Fig. 12.1. Alternatively, ready-mounted tag strips can be purchased as *tag boards*. A component continuing a circuit from a 'common' point has its second lead soldered to a spare tag, which can become a 'common' connecting point for another component, and so on.

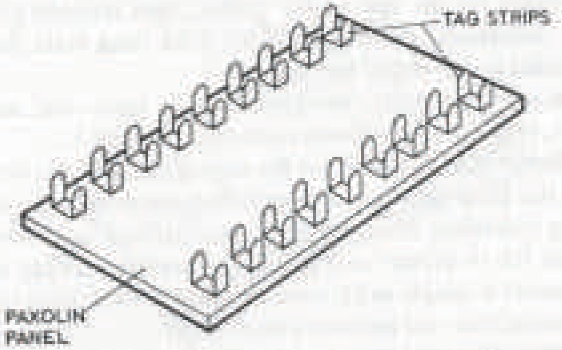

Fig. 12.1 Tag boards can be made by mounting tag strips on a Paxolin panel, or bought ready-made.

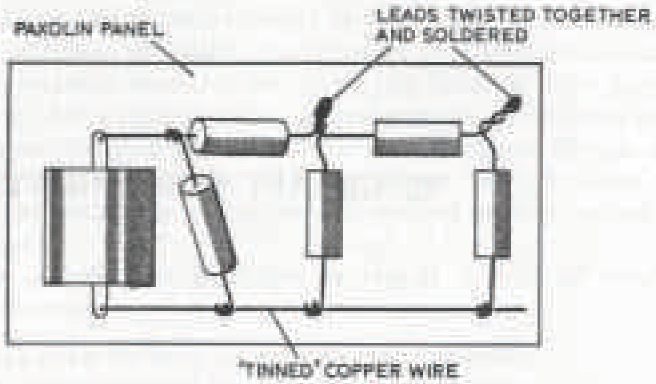

Fig. 12.2 Suitable method of circuit construction involving small number of components.

This technique was illustrated in the construction of simple crystal sets in Chapter 3. It is adaptable to more complex circuits, particularly for experimental work, but is generally untidy and unnecessarily bulky for general use. The greater the number of components, the more the circuit assembly tends to expand lengthwise, with lots of wasted space.

Direct assembly can be an attractive alternative where the components are laid out in order on a Paxolin panel. Connections are made by twisting appropriate leads together, then soldering (cutting off surplus length of leads where necessary). Larger common connections can be made to a length of 'tinned' copper wire of about 16 s.w.g. laid flat on the panel. One or two major components, such as aerial coils and the tuning capacitor, can be secured to the panel, forming an anchor point for the rest of the circuit. Main terminal point, e.g. for battery connectors, can be formed by 6-BA brass bolts fitted through holes drilled at one end of the panel.

Constructed correctly, the result can be quite neat and relatively compact, with the whole circuit quite rigid–Fig. 12.2.

'Christmas tree' layout is for the experimental or test circuit only; it follows the same principle of connecting components by twisting and soldering individual leads together, but without any attempt to lay them out flat or mount on a panel. The result invariably looks pretty messy, but it is simple and it works–provided leads do not accidentally touch one another and produce a short-circuit.

Breadboard construction is probably the best for the more serious experimenter, and can also be used for permanent circuits where overall

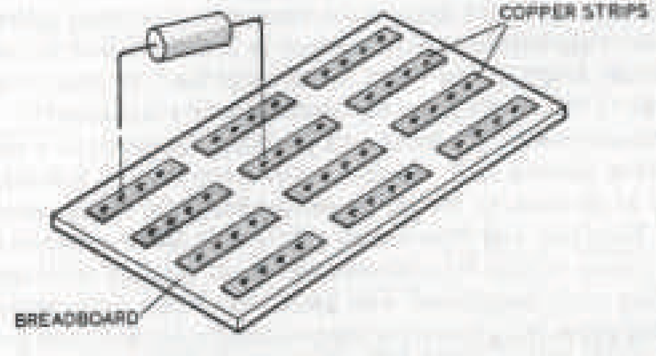

Fig. 12.3 'Breadboard' panel with pre-drilled contact strips. In some cases the copper is in long strips, sections being isolated by cutting through, as required.

size is not important. It is similar to tag-strip assembly except that the terminal points are laid out in the form of strips on a flat panel. Each strip has a number of holes into which individual component leads can be inserted for common connection, the other lead(s) connecting across to further separate strips as required to continue the circuit–Fig. 12.3. Some proprietary breadboards have small spring clips fitted under individual holes both to provide positive contact and hold the component in position without soldering–Fig. 12.4. In others the strips

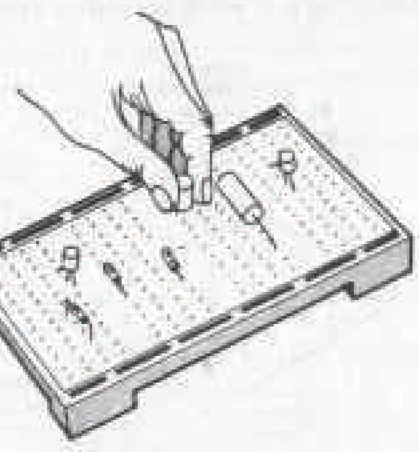

Fig. 12.4 Modern breadboard unit with spring location for components pushed in position.

are simply copper with drilled holes, component leads being soldered in position; strip lengths in this case may be in separate sections, or run the whole length of the panel, in the latter case the strip being cut through to isolate each group of connection points as required.

Pegboard assembly is based on a perforated baseboard of a suitable insulating material with terminal pillars which can be inserted into holes, as required, to provide connecting points for components—Fig. 12.5. The pillars may have a screw top for solderless assembly, or be of plain copper or brass for soldered connections. A similar technique can be allied to a 'breadboard' with prefabricated connecting strips, the terminal pillars being inserted in the pre-drilled holes, and strips isolated as necessary.

Proprietary constructional panels of these two types are designed for both amateur and general use and are widely favoured by many constructors who find designing their own component layout a tricky problem. A feature of many of these is that they can be used over and over again. The best advice regarding their suitability is to experiment and find by experience which appeals most to the individual builder, or is found easiest to use.

An alternative 'home-made' system, which would be regarded as crude by most professionals, is to cut a panel of ply or even hard balsa for the base, draw on the position of the components (as near to scale as possible) and clearly mark the common connection points required. A copper nail is then driven into the base at each end of these points, forming terminal posts to which component leads can be soldered—Fig. 12.6. Longer common leads can again be a length of tinned copper wire soldered between two end nails.

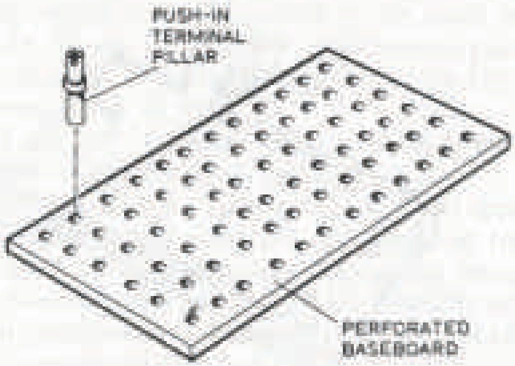

Fig. 12.5 'Pegboard' construction using a perforated baseboard and push-in terminal pillars.

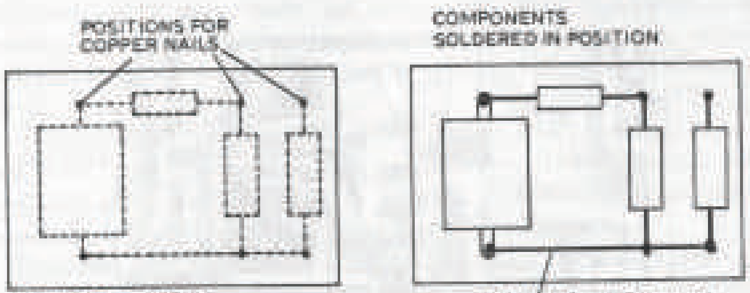

Fig. 12.6 'Copper nail' technique—again suitable for small circuits.

A refinement of this is to use a Paxolin panel, and screw or bolt small individual tags to main and common connecting points; the longer common connections are formed from tinned copper wire soldered between two tags; individual common connecting points are formed by separate tags—Fig. 12.7.

The best technique, particularly where miniaturization of the circuit is required, is printed-circuit assembly—Fig. 12.8. Printed-circuit stock is invariably purchased in panels which can be cut to the overall size required, one face of which is coated with copper foil and the actual circuit pattern produced by etching. This involves transferring a drawing of the circuit on to the copper, coating all copper areas which are to remain with a suitable resist, and then immersing the panel in an acid bath to etch or dissolve away the remaining unwanted copper. The

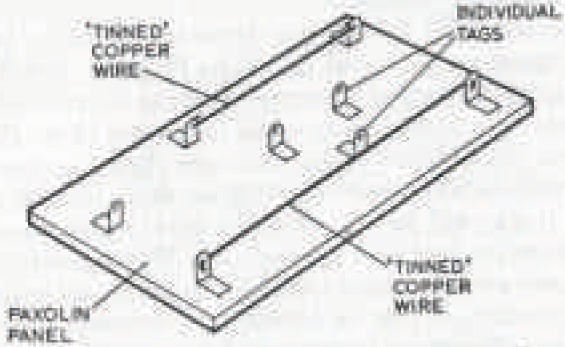

Fig. 12.7 Home-fabricated tag- and bus-bar panel.

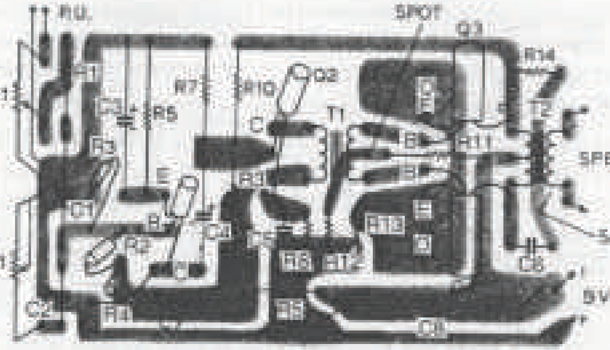

Fig. 12.8 Printed-circuit drawing layout for a push–pull amplifier.

resist is then removed with a solvent, when the PC panel can be prepared for component assembly by drilling, etc.

All this work is well within the scope of the amateur enthusiast, either working from a PC plan or designing a suitable printed pattern equivalent to a theoretical circuit design, although the latter can become quite an involved business. Alternatively, many standard radio designs etc., are available as kits or in component form for home assembly, including a complete PC, ready drilled as necessary; this latter feature saves a lot of time and effort, being a point of design on which the less experienced constructor is likely to go wrong. One of the basic essentials of PC design, in fact, is complete familiarity with component sizes so that mounting holes or holes for leads properly match the components to be accommodated, all arranged in logical order, both physically and electrically. Unlike ordinary wiring-up, PC conductors cannot be crossed over each other.

The important thing to remember is that the PC pattern has to be etched 'in reverse' on the copper side of the PC board. Simply make a tracing of the original drawing of the component layout and connections and turn this over to transfer on to the copper side of the PC board.

The layout should follow basically the same physical arrangement as the theoretical circuit. Where this diagram itself includes crossing conductors it may well be of considerable advantage to discover if it can be replanned so that these are eliminated. If this proves physically impossible, and a solution cannot be found by altering the arrangement of the components, it may be necessary to terminate certain connections on the PC lands and interconnect to cross other lands with an

insulated jumper wire. Whilst this may be considered bad design, it is perfectly suitable for one-off or amateur construction. Bear in mind, too, that components themselves can be used for bridging over adjacent conductors.

The current-carrying capacity of conductors varies with foil thickness but, since the current values in transistor receiver circuits are invariably low, conductor size is unlikely to be critical, although recommended minimum figures should be adhered to.

Minimum recommended conductor width is $\frac{1}{16}$ in., with at least $\frac{1}{32}$ in. clear spacing between adjacent conductors to reduce the possibility of accidental shorts or 'bridging' between them when soldering. The drawing should also allow for a minimum spacing of at least $\frac{1}{32}$ in. between the outside conductor and the edge of the PC board.

Where holes have to be drilled to take component leads, the diameter should closely match the lead size, e.g. with a typical resistor lead of 0.028 in. (22 s.w.g.) the corresponding hole diameter should be at least $\frac{1}{32}$ in. or No. 67 drill. Sufficient area of copper 'land' should be provided around each hole for a minimum width of $\frac{1}{16}$ in. (Fig. 12.9). The holes should be correctly spaced to match component leads, allowing for a 'finger bend' at each end or down the side of the component, according to whether horizontal or vertical assembly is to be used (Fig. 12.10). The former is to be preferred except where a minimum size panel is aimed at, when vertical mounting of resistors etc., will occupy minimum base area. At no point, however, should spacing between adjacent holes be less than twice the laminate thickness (*i.e.* normally not less than $\frac{1}{8}$ in. on standard laminate).

Other points to watch are that where conductors join at an acute angle they should be faired in with a generous fillet, increasing the area

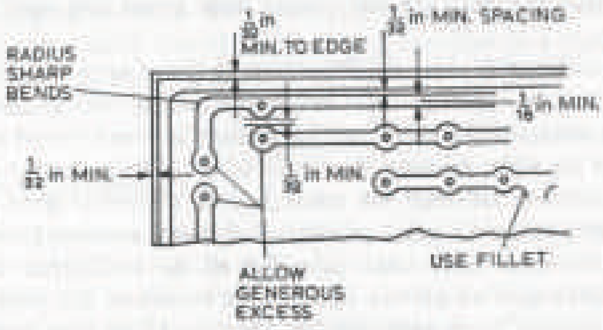

Fig. 12.9 Recommended proportions for conductors, spacing etc.

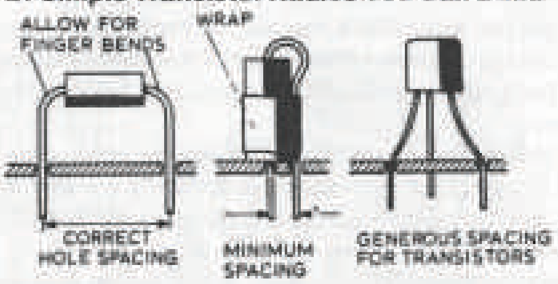

Fig. 12.10 Mounting of components on a printed-circuit board. Vertical mounting saves space.

bonded at the joint and making the copper far less liable to be lifted. Also, do not leave unnecessarily wide or large areas of copper as conductors as these may be subject to excessive heating and expansion, with the result that the copper tends to lift from the base. Either reduce the outline of such areas or relieve the surface area with slots to be etched away. This is not so important on low-voltage circuits, but on mains-operated ones no copper area of more than about one square inch should be left 'solid'.

The first step, having prepared a PC drawing, is to cut the laminate to the required overall size using a fine-tooth saw. The copper surface has probably become greasy and dirty through handling, and so should now be thoroughly cleaned by washing with a detergent and rubbing dry with a clean cloth. If the copper is discoloured through corrosion, use a domestic abrasive cleaner to bring it up bright clean.

A test of cleanliness is to hold the panel under a tap, copper side up, and allow water to run on to it. If the water flows smoothly over the whole area, the surface is clean and grease-free; if isolated patches of copper stay dry they are still coated with grease and need further cleaning.

Once completely clean the PC pattern is drawn or traced on to the copper. Cellulose paint or resist ink should then be used to paint in all the land areas, using a ruling pen for straight lines and a small brush to complete the wider sections; the whole of the pattern may be painted on, if preferred, although the result may be somewhat more ragged. Avoid using too much paint or resist, as this may overrun the outlines but, at the same time, make sure that all the land areas are fully covered. The painted pattern should then be left to dry, which may take an hour or more with cellulose paint or 10 to 15 minutes with resist inks.

The solution normally used for etching is ferric chloride mixed with a little hydrochloric acid, or straight dilute nitric acid. The former is generally preferred since it does not 'gas' as much as the acid alone, but either will be equally effective. This solution is poured into a suitable shallow container, and the laminate immersed in it. Rate of etching will depend on the temperature of the solution and also the degree of agitation. Thus, at a temperature of 10°C a ferric-chloride bath will etch the copper at a rate of about 1 thou. in 20 minutes—or nearly an hour to etch 3 thou. foil; at 20°C the rate of etching is increased to about 10 minutes per thou.; and at 35°C is almost twice as fast again. The rate can also be increased by gentle agitation, moving the board gently backwards and forwards in the bath, or gently rocking the container to swill the etching solution from end to end.

Etching should be allowed to proceed until all traces of copper have disappeared from the surface. The board can then be removed and rinsed under running water to remove any traces of etchant. The paint or resist ink covering the lands is then removed either with a solvent (e.g. cellulose thinners, for cellulose paint) or a cleaner (resist inks). After this, again wash and dry the board; then it is ready for drilling.

Three basic rules must be observed when drilling the PC panels:

(i) Always use a sharp drill (preferably a new one, or one which has been re-sharpened prior to use).
(ii) Always drill from the copper side (i.e. copper face up).
(iii) Always use a backing of hard material underneath so that the drill point will not tear out a section of the laminate when it breaks through.

Drilling may be done with a hand or electric drill, the latter being far less tiring to use when there are a large number of holes to be made, although the small size of drill required may lead to a high breakage rate unless special care is taken.

Notes on soldering

Electric irons are invariably used for electronic assemblies together with resin-cored solder (i.e. the solder is in the form of a hollow wire with the core filled with resin flux). Flame-heated irons are not satisfactory for they are usually too large and cumbersome, and also do not permit good temperature control. Equally, one should always use an electrical (cored) solder when no separate flux is required; an acid-type flux should *never* be used on electrical work as this will inevitably produce corrosion.

The basic rules for good soldering are extremely simple, although they are frequently ignored. They may be summarized as:

(i) The iron should be of the right size and type.
(ii) The iron should be hot enough to melt the solder freely.
(iii) The tip of the iron must be kept 'tinned' and clean.
(iv) The work surfaces to be soldered must be clean and grease-free.

The correct size of iron is important for if too small it will rapidly lose heat when applied to the work, and if too large may overheat adjacent components or be awkward to apply. For PC assemblies a $\frac{1}{8}$-in. bit diameter is about right for general work–not too large, but large enough to retain enough heat for more or less continuous work. It may be awkward to use when soldering up a miniature panel where the lands are close together, when a $\frac{3}{32}$-in. or even a $\frac{1}{16}$-in. bit may be preferred. This smaller size will, however, usually lose so much heat in completing a single joint that it has to be left to heat up again before it can be applied to the second joint.

Iron size is also specified by wattage, but this is more of a nominal rating than anything else, for irons by different manufacturers but of the same stated wattage can have a considerable difference in performance, e.g. differences in heating-up times and in bit temperature achieved (and variations in the latter can be as high as 100°C). To avoid the possibility of over-heating, an iron with a rating of more than 50 W should never be used for PC assemblies.

Iron voltage must be matched to the mains voltage available and it is usual to operate an electric iron on the middle voltage of the range, e.g. for a 240-V mains a 230/250-V iron would be correct. An iron should never be operated on a mains voltage below the lowest figure of its rating as in such a case it will not develop its correct bit temperature.

Typically, a good electric iron will achieve a saturation temperature of approximately 375°C. The time taken to reach this temperature will vary with the size and design, but very roughly should be of the order of wattage divided by four (in seconds), e.g. in the case of a 30-W iron the bit should reach its maximum or saturation temperature in about $30 \div 4 = 7\frac{1}{2}$ seconds. It should also take about one-third of this time (e.g. watts divided by 12) to reach a satisfactory bit temperature for

soldering (250°C), although this can readily be judged by trying the solder on the bit to see if it runs freely.

The *minimum* bit temperature required for satisfactory soldering is 40°C above the melting point of the solder used, which will vary with the composition of the solder.

Whether the temperature of the bit is satisfactory or not can readily be judged by the time it takes to complete a joint. This should be not more than three to four seconds, and the resulting joint should be bright clean, with the solder completely 'wetting' both surfaces. If the joint takes longer to make and/or the solder has a pasty or dull appearance, the iron is not hot enough. If the solder is reluctant to take or flow over the joint, or collects in blobs rather than spreading out, then the surfaces are dirty.

Before attempting assembly on a PC board the panel should always be cleaned so that the lands are bright all over with no dull spots, and subsequently should not be finger-marked by handling. An ordinary domestic powder cleaner is as good as anything for cleaning the PC lands, used wet or dry, the panel then being rinsed under running water to remove any traces of abrasive, and dried on a clean rag.

Component leads are normally 'tinned' and therefore in a suitable state for soldering. Almost certainly, however, the 'tinned' surface will have become dirty or partly corroded during handling and storage and it is generally recommended to clean leads immediately prior to assembly and soldering in position. A scrap of fine emery paper is excellent for this, simply wrapped around and pulled along the length of the lead, taking care not to impose excess mechanical strain; similarly with tags etc. Time taken in cleaning leads etc., is usually time saved, for one can then be sure of a satisfactory soldered joint at the first attempt.

Then there is the question of heat damage to consider. Provided the joint is completed quickly–i.e. within three or four seconds of application of the iron–no component is likely to suffer heat damage. This even applies to transistors which are normally rated to withstand continuous application of a soldering bit no closer than $\frac{1}{4}$ in. from the base for a period of up to ten seconds without damage. Thus if all goes well and soldered joints are completed quickly and neatly, it is seldom necessary to worry about heat damage.

It is still, however, commonly recommended that a 'heat sink' should always be used on each lead of a transistor when soldering in position. A heat sink, basically, is a mass of conducting material which will absorb heat from the iron rather than let the full heat flow up the lead. The jaws of a pair of pliers gripping the transistor lead form a suitable heat sink; or, if these are not convenient to use, a crocodile clip can serve the same purpose (especially if the jaws are filed flat to provide maximum surface contact with the lead).

The most likely cause of heat damage is re-working a joint which has not been made properly, or trying to unsolder a lead which has been wrongly positioned. This may entail leaving the iron in contact with the joint far longer than the 'safe' three or four seconds, when damage can result to the component. Excessive heat applied locally in this manner will also tend to 'lift' the copper land away from the base material on a PC panel.

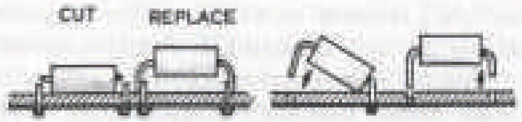

Fig. 12.11 The safest way to remove a faulty component is usually to cut it off.

Removing a component which has been mounted on a PC can, in fact, be a tricky process. If it is a faulty component which is to be replaced the *safest* way to go about such a job is to cut it off, as shown in Fig. 12.11, leaving stub lengths of the original lead protruding from the plain side of the PC panel. The new component can then be soldered in place *to* these stub leads. Alternatively, having cut off the component, each lead can be removed in turn by laying the tip of the iron on the solder joint (PC side) and withdrawing the lead with pliers as soon as the solder has melted. Then, before it has had time to set again, blow surplus solder out of the hole (Fig. 12.12).

To remove a component intact usually means working on one lead at a time, lifting first one end of the component and then the other. This is not always possible when the component is mounted on rigid leads or tags, generally demanding working on pairs of tags at a time and gradually levering the component upwards. This usually entails considerable risk of overheating the component and of 'lifting' the PC lands if too much heat is applied to the tag side.

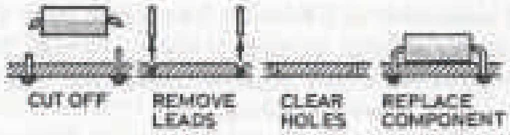

Fig. 12.12 Remaining leads can then be removed, ready to fit new component.

A trick which often helps in removing a rigidly mounted component, such as an IF transformer, is to heat the tags and remove the solder with a suitable stiff-bristled brush. This is repeated on each tag until the component is sufficiently loosened to be prised free. It is important, however, to avoid splashing excess solder on to other parts of the PC board.

Unless obviously faulty, transistors or diodes should not be removed from a PC board once assembled, as the possibility of permanent damage resulting is high. If imperative that they should be removed (e.g. if they have been connected up wrongly), a heat shunt should always be used (e.g. a pair of thin-nosed pliers gripping the lead between the transistor body and the PC board).

Where the board itself is damaged—e.g. foil has come unstuck through excessive local heating—loosened lands can usually be stuck back again by carefully heating the foil with the soldering iron and then pressing in place on the board until the adhesive has reset. If not, the loose land may have to be stuck down with an application of a general-purpose domestic or epoxy adhesive.

Mechanical damage to PC boards can be repaired by bridging breaks or cracks in the copper lands with 'jumper' wires soldered in place. These can be plain 'tinned' copper wire, but if a longer jumper is required which passes over adjacent lands, either insulated wire should be used or the wire length fitted with insulating sleeving. All repaired areas should then be coated with clear lacquer for protection.

Where a PC board is quite extensively damaged, or needs to have a number of components changed, it is usually more satisfactory in the long run to scrap the original board and start again with a new one; and preferably new components, rather than to try to salvage all the original ones, as these may be damaged in the process of removal from the old board.

13 FETs AND ICs

Field-effect transistors (FETs) have a construction, and characteristics, that differ from conventional junction transistors. Their behaviour, in fact, is more like that of a triode valve, although retaining the characteristic low-power requirements of a transistor.

The electrodes of an FET are designated gate, source and drain with the following approximate equivalents:

FET	Transistor	Triode valve
Gate	Base	Grid
Source	Emitter	Cathode
Drain	Collector	Anode

The standard symbol for an FET differs from that of a conventional transistor, with the gate carrying an arrow, which points outwards for a *P*-channel FET and inwards for an *N*-channel FET–Fig. 13.1. The general rule is that the polarity of the *drain* supply is negative for a *P*-channel FET and positive for an *N*-channel FET.

A simple regenerative receiver based on two *P*-channel FETs is shown in Fig. 13.2. This is adjusted for optimum working by the

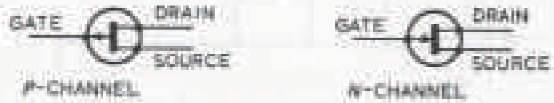

Fig. 13.1 Symbols for the two basic types of field-effect transistors (FETs).

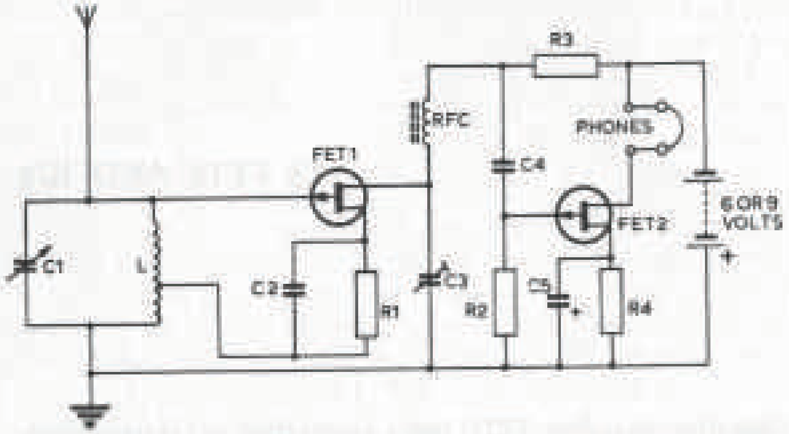

Fig. 13.2 *Circuit 26.* Simple regenerative receiver using *P*-channel FETs.

FET1, FET2, 2N3819, 2N5457, or similar

C1, 0–500 pF
C2, 8 or 10 µF
C3, 50 pF
C4, 0.01 µF
C5, 8 or 10 µF

R1, 22 k
R2, 470
R3, 22 k
R4, 10 k
RFC, 10 mH
L, aerial coil

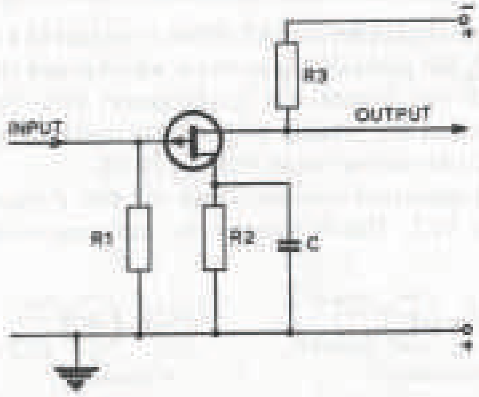

Fig. 13.3 Basic AF amplifier circuit using a *P*-channel FET. R1 is bias resistor; R2 and C stabilizing load; R3 drain resistor.

trimmer capacitor C3. L and C1 are a conventional aerial tuned circuit with tapping point.

Fig. 13.3 shows a typical arrangement for an FET amplifier stage. This could be added on to the receiver circuit to provide amplification which should be sufficient to power a small loudspeaker.

Integrated circuits

Integrated circuits (ICs) are one of the latest products of modern electronic technology–small integrated circuit modules in which a complete circuit duplicates the performance of perhaps a dozen or more active devices (e.g. transistors and diodes) and a similar number of passive components (e.g. resistors and capacitors) in a silica 'chip' only a few millimetres square.

Depending on their design function, ICs may be used as a complete stage or logic element in a circuit; or, more usually, associated with a number of conventional external components to complete the circuit. Fig. 13.4, for example, shows the equivalent circuit diagram of one IC (Siemens TAA 861), and Fig. 13.5 the incorporation of that particular IC in a 4-W amplifier.

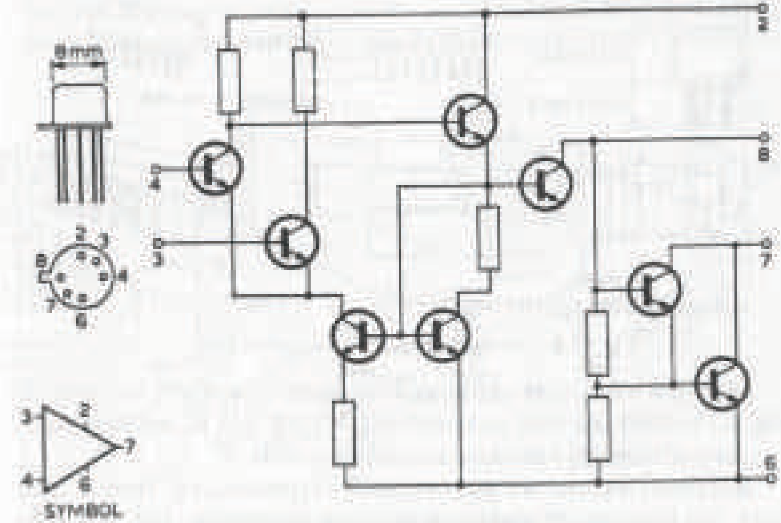

Fig. 13.4 Siemens TAA 861 integrated circuit together with diagram of equivalent circuit–i.e. embracing eight transistors and six resistors.

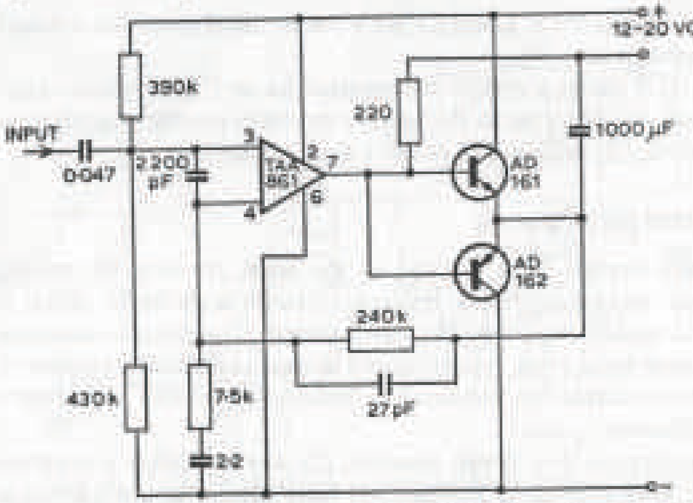
12–20 VOLTS
390k
220
1000 µF
INPUT
0·047
2200 pF
TAA 861
3
2
7
4
6
AD 161
AD 162
240k
430k
7·5k
27 µF
2·2

Fig. 13.5 *Circuit 27.* Siemens TAA 861 'chip' used in a complete circuit with external components. This circuit is actually that of a 4-W amplifier, working off 12–20 V.

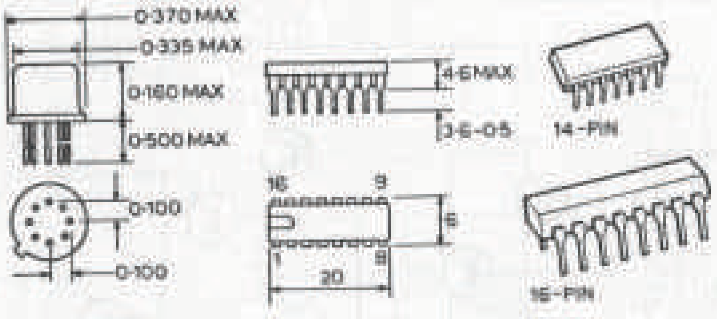
0·370 MAX
0·335 MAX
0·160 MAX
0·500 MAX
0·100
0·100
4·6MAX
3·6–0·5
14–PIN
16
9
1
8
6
20
16–PIN

DIMENSIONS IN MILLIMETRES

Fig. 13.6 Example of outlines of typical ICs.

Examples of outline appearance of some typical ICs are given in Fig. 13.6. They are best mounted by plugging into IC holders, which are normally designed for mounting on PC boards.

Integrated circuits are not necessarily expensive, e.g. from a little over 50p upwards, depending on type and complexity, and they offer an interesting new field of experiment for the amateur constructor. Manufacturers of ICs normally provide suggested applications for their various types.

Low Frequency TRF Receiver

Because the LM172 is a broadband functional module, it may be used to amplify and detect signals below 2 MHz directly, without the more complex frequency conversion of superheterodyne receivers. In the AM Broadcast Band (550-1650 kHz), the strip has sufficient sensitivity to operate alone in urban reception areas, since AGC action is useful down to about 50 microvolts at Pin 2. With additional gain either preceding the module, or inserted between Pins 1 and 3, it may also be useful in monitoring Loran (1.8-2.0 MHz), or the numerous directional and informational channels below 550 kHz.

While the complete TRF (Tuned Radio Frequency) broadcast receiver of Fig. 13.7 has relatively poor selectivity, because only a single, low "Q" tuned circuit is used in the entire receiver, it serves to illustrate the straightforward design possible in TRF construction. More sophisticated designs might use multisection tuning, ahead of the strip. The prototype was constructed using very inexpensive imported "transistor-radio" components.

A ferrite loopstick antenna, L1, resonates with a small, polyethylene dielectric tuning capacitor within the broadcast band. The LM172 performs its gain function just as it would in an IF application, but in this case, directly drives a class A power amplifier. Since the DC output voltage at Pin 6 is relatively constant (from 2.1 to about

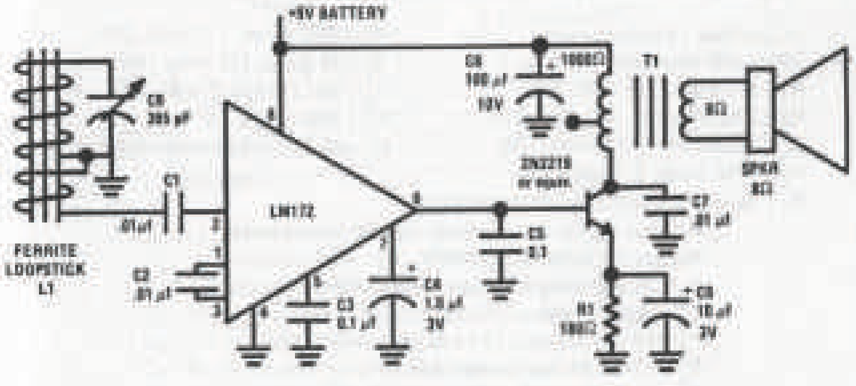
+9V BATTERY
C9
365 pF
C1
.01 µF
FERRITE LOOPSTICK L1
C2
.01 µF
LM172
2
1
3
4
5
6
7
8
C3
0.1 µF
C4
1.0 µF
3V
C5
0.1
C8
100 µF
10V
1000Ω
T1
8Ω
SPKR
8Ω
2N2219
or equiv.
C7
.01 µF
R1
100Ω
C6
10 µF
3V

L1 – Ferrite Loopstick – Philmore FF15 (packaged as set of 3 sizes)
C9 – Sub-miniature variable capacitor – Philmore 1945G – 365 pF max.
T1 – Midget Audio Transformer, 1000Ω:8Ω – Archer 273-1380 (Radio Shack, Inc.)
SPKR – 2" PM Speaker, 8Ω, 0.1 watt – Philmore T520

Fig. 13.7 TRF Broadcast Receiver

2.4 volts as a function of AGC). It is used to bias the class A stage directly, eliminating a number of components. C7 and C8 are needed to prevent regenerative audio oscillations with weak batteries. Total receiver drain from the 9-volt supply is 10 mA, of which only 1.9 mA is used in the LM172; the rest is needed for the audio amplifier.

A volume control was not provided in the prototype, as volume was excellent with the small (2" diameter) speaker used, and AGC was so effective that no perceptible difference in stations was heard. Volume control is possible by inserting a potentiometer between the emitter of the audio output transistor and R1.

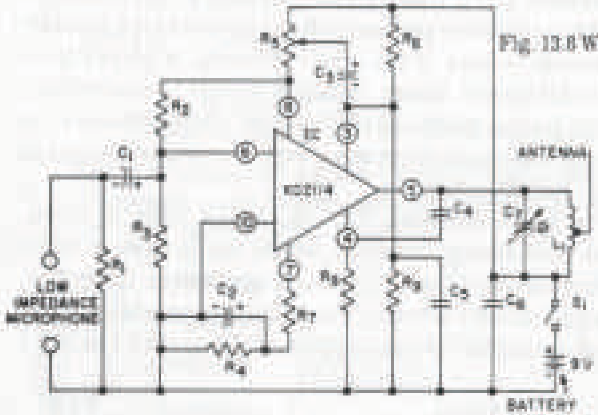

Fig. 13.8 Wireless microphone.

Microphone : low-impedance,
200 to 600 ohms, RCA HK-99
R_1 : 270 ohms, 1/2 watt, 10%
R_2 : 150,000 ohms, 1/2 watt, 10%
R_3 : 22,000 ohms, 1/2 watt, 10%
R_4 : 1200 ohms, 1/2 watt, 10%
R_5 : 10,000 ohms, trimmer potentiometer
R_6 : 8200 ohms, 1/2 watt, 10%
R_7 : 68 ohms, 1/2 watt, 10%
R_8 : 330 ohms, 1/2 watt, 10%
R_9 : 6800 ohms, 1/2 watt, 10%
S_1 : switch, single-pole single-throw
slide type
Antenna : 3/16-inch-diameter rod, 2-1/2 inches long
Battery : 9 volts; RCA VS323
C_1 : 25 microfarads, 6 volts, electrolytic
C_2 : 200 microfarads, 6 volts, electrolytic
C_3 : 20 microfarads, 10 volts, electrolytic
C_4 : 15 picofarads, silver mica
C_5 : 1000 picofarads, 25 volts or greater, ceramic
C_6 : 0.1 microfarad, 25 volts or greater, ceramic
C_7 : 3 to 30 picofarads, trimmer type,
IC : integrated circuit; RCA KD2114
L_1 : 6 turns No. 12 wire; coil I.D. 5/16 inch, length 3/4 inch

Wireless Microphone

This wireless microphone (Fig. 13.8) transmits on any desired FM broadcast-band frequency from 88 to 108 MHZ (tuning controlled by C7). The range of the wireless microphone is 50 to 150 feet depending on the location of the receiver and its antenna. Circuit features include small size, high quality, and low current drain; the antenna length has been limited to keep radiation within FCC wireless-microphone regulations. The low-impedance microphone to be used with this circuit is the RCA HK-99.

Marine-Band Converter

This marine-band converter (Fig. 13.9) converts any standard AM broadcast receiver to one capable of receiving transmission in the 2 to 3-MHz marine band. The output is connected to the AM receiver antenna terminals. If the receiver does not have antenna terminals the output is connected to 10 turns of No. 32 wire wound over the loopstick antenna in the receiver.

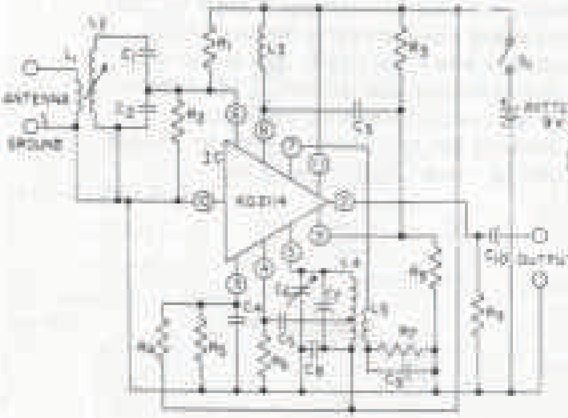

Fig. 13.9 Marine-band converter.

Battery : 9 volts; RCA VS323
C_1 : 180 picofarads, 25 volts or greater, silver mica
C_2 : 680 picofarads, 25 volts or greater, silver mica
C_3 : 470 picofarads, 25 volts or greater, ceramic
C_4 C_5 C_{10} : 0.01 microfarad, 25 volts or greater, ceramic
C_6 : 50 picofarads, variable; Hammarlund HF-50 or equivalent
C_7 : 56 picofarads, 25 volts or greater, silver mica
C_8 C_9 : 0.05 microfarad, 25 volts or greater, ceramic
IC : integrated circuit; RCA KD2114
L_1 : 10 turns No. 32 enameled wire wound on the same form
as L_2 and close to L_2

L_2 : Miller No. 4206 or equivalent

L_3 : rf choke, 2.5 millihenries; Miller No. 70F253A1 or equivalent

L_4 : 40 turns of B and W coil No. 3012 or equivalent, see Note

L_5 : 9 turns of B and W coil No. 3012 coil stock (or equivalent) spaced 1 turn from positive end of L_4; see Note

R_1 : 62,000 ohms, 1/2 watt, 10%
R_2 : 4700 ohms, 1/2 watt, 10%
R_3 : 150,000 ohms, 1/2 watt, 10%
R_4 : 47,000 ohms, 1/2 watt, 10%
R_5 : 22,000 ohms, 1/2 watt, 10%
R_6 : 2200 ohms, 1/2 watt, 10%
R_7 : 1000 ohms, 1/2 watt, 10%
R_8 : 100,000 ohms, 1/2 watt, 10%
R_9 : 470 ohms, 1/2 watt, 10%
S_1 : switch; single-pole single-throw

Note : L_4 and L_5 are made from coil type B and W No. 3012 (or equivalent) coil stock cut to 50 turns. Allow enough extra wire at the outside ends to form leads. The 50-turn coil is cut at turn 40-1/2; this forms a 40-1/2-turn coil (L_4), and a 9-1/2-turn coil (L_5). The extra 1/2 turn on each coil is lifted (or unrolled) and used as a lead. L_4 is tapped at turn 31.

Appendix

ELECTRONICS SYMBOLS

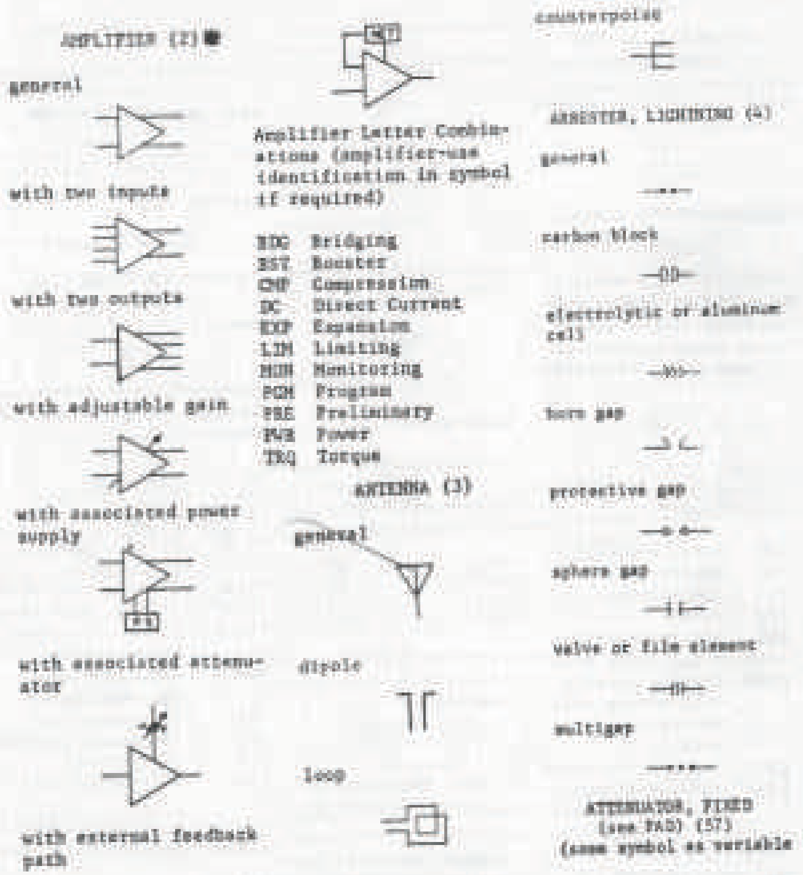
AMPLIFIER (2) ●
general
with two inputs
with two outputs
with adjustable gain
with associated power supply
with associated attenuator
with external feedback path
Amplifier Letter Combinations (amplifier-use identification in symbol if required)
BDG Bridging
BST Booster
CMP Compression
DC Direct Current
EXP Expansion
LIM Limiting
MON Monitoring
PGM Program
PRE Preliminary
PWR Power
TRQ Torque
ANTENNA (3)
general
dipole
loop
counterpoise
ARRESTER, LIGHTNING (4)
general
carbon block
electrolytic or aluminum cell
horn gap
protective gap
sphere gap
valve or film element
multigap
ATTENUATOR, FIXED (see PAD) (57)
(same symbol as variable

● NUMBER IN PARENTHESES INDICATES LOCATION OF SYMBOL IN MIL-STD PUBLICATION attenuator, without variability)

ATTENUATOR, VARIABLE (5)

balanced

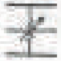

unbalanced

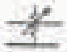

AUDIBLE SIGNALING DEVICE (6)

bell, electrical; ringer telephone

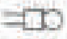

buzzer

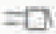

horn, electrical; loudspeaker; siren; underwater sound hydrophone, projector or transducer

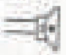

Horn, Letter Combinations (if required)

*HN	Horn, electrical
*HW	Howler
*LS	Loudspeaker
*SN	Siren
‡EM	Electromagnetic with moving coil
‡EMN	Electromagnetic with moving coil and neutralizing winding
‡MG	Magnetic armature
‡PM	Permanent magnet with moving coil

identification replaces (*) asterisk and (‡) dagger)

sounder, telegraph

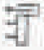

BATTERY (7)

generalized direct current source; one cell

multicell

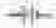

CAPACITOR (8)

general

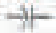

polarized

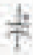

adjustable or variable

continuously adjustable or variable differential

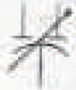

phase-shifter

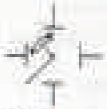

split-stator

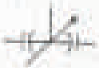

feed-through

CELL, PHOTOSENSITIVE (Semiconductor) (9)

asymmetrical photoconductive transducer

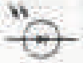

symmetrical photoconductive transducer

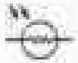

photovoltaic transducer; solar cell

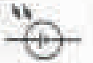

CIRCUIT BREAKER (11)

general

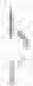

with magnetic overload

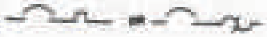

drawout type

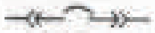

CIRCUIT ELEMENT (12)

general

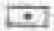

Circuit Element Letter Combinations (replaces (*) asterisk)

EG	Equalizer
FAX	Facsimile set
FL	Filter
FL-BE	Filter, band elimination
FL-BP	Filter, band pass
FL-HP	Filter, high pass
FL-LP	Filter, low pass
PS	Power supply
RG	Recording unit
RU	Reproducing unit
DIAL	Telephone dial
TEL	Telephone station
TPR	Teleprinter
TTY	Teletypewriter

Additional Letter Combinations (symbols preferred)

AR	Amplifier
AT	Attenuator
C	Capacitor
CB	Circuit breaker
HS	Handset
I	Indicating or switch board lamp
L	Inductor
J	Jack
LS	Loudspeaker
MIC	Microphone
OSC	Oscillator
PAD	Pad
P	Plug
HT	Receiver, headset
K	Relay
R	Resistor
S	Switch or key switch
T	Transformer
WR	Wall receptacle

CLUTCH; BRAKE (14)

disengaged when operating means is de-energized

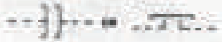

engaged when operating means is de-energized

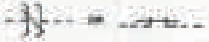

COIL, RELAY and OPERATING (16)

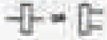

semicircular dot indicates inner end of wiring

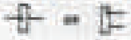

CONNECTOR (18)

assembly, movable or stationary portion; jack, plug, or receptacle

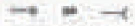

jack or receptacle

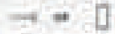

plug

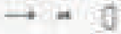

separable connectors

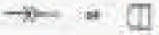

two-conductor switchboard jack

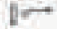

two-conductor switchboard plug

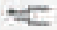

jacks normalized through one way

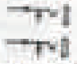

jacks normalized through both ways

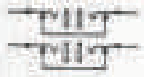

2-conductor nonpolarized, female contacts

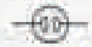

2-conductor polarized, male contacts

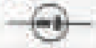

waveguide flange

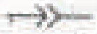

plain, rectangular

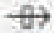

choke, rectangular

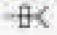

engaged 4-conductor; the plug has 1 male and 3 female contacts, individual contact designations shown

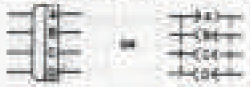

coaxial, outside conductor shown carried through

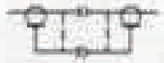

coaxial, center conductor shown carried through; outside conductor not carried through

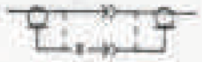

mated choke flanges in rectangular waveguide

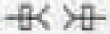

COUNTER, ELECTROMAGNETIC; MESSAGE REGISTER (26)

general

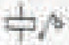

with a make contact

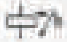

COUPLER, DIRECTIONAL (27) (common coaxial/waveguide usage)

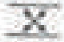

(common coaxial/waveguide usage)

E-plane aperture-coupling 30-decibel transmission loss

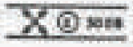

COUPLING (28)

by loop from coaxial to circular waveguide, direct-current grounds connected

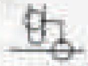

CRYSTAL, PIEZO-ELECTRIC (62)

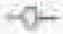

DELAY LINE (31)

general

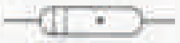

tapped delay

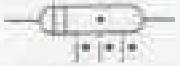

bifilar slow-wave structure (commonly used in traveling-wave tubes)

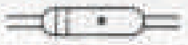

(length of delay indication replaces (*) asterisk)

DETECTOR, PRIMARY; MEASURING TRANSDUCER (30)

(see HALL GENERATOR and THERMAL CONVERTER)

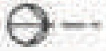

DISCONTINUITY (33)

(common coaxial/waveguide usage)

equivalent series element, general

magnetic core

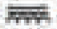

tapped

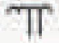

adjustable, continuously adjustable

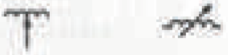

KEY, TELEGRAPH (43)

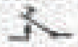

LAMP (44)

ballast lamp; ballast tube

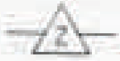

capacitive reactance

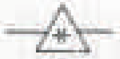

inductive reactance

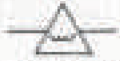

inductance-capacitance circuit, infinite reactance at resonance

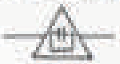

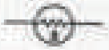

lamp, fluorescent, 2 and 4 terminal

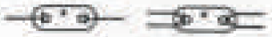

lamp, glow; neon lamp
a-c

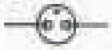

d-c

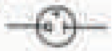

lamp, incandescent

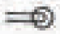

indicating lamp; switchboard lamp
(see VISUAL SIGNALING DEVICE)

LOGIC (see 806B and Y32-14) (including some duplicate symbols; left and right-hand symbols are not mixed)

AND function

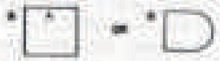

OR function

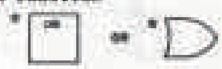

EXCLUSIVE-OR function

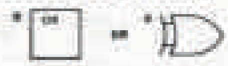

((*) input side of logic symbols in general)

condition indicators

state (logic negation)

a Logic Negation output becomes 1-state if and only if the input is not 1-state

an AND func. where outp
is low if and only if a
inputs are high

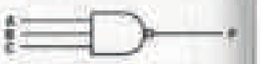

electric inverter

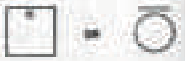

(elec. invtr. output becomes 1-state if and on
if the input is 1-state
(elec. invtr. output is more pos. if and only i
input is less pos.)

level (relative)

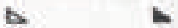

1-state is less + 1-state is more +

(symbol is a rt. triang
pointing in direction o
flow)

an AND func. with inpu
1-states at more pos. level and output 1-stat
at less pos. level

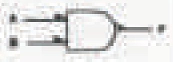

single shot (one output)

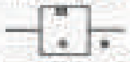

(waveform data replaces inside/outside (*))

schmitt trigger, wavefor
and two outputs

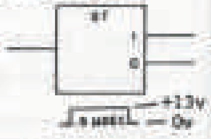

flip-flop, complementar

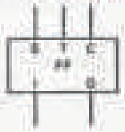

flip-flop, latch

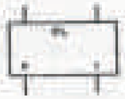

register

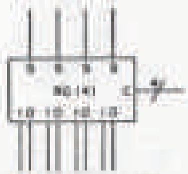

(binary register denoting four flip-flops and bits)

amplifier (see AMPLIFIER)

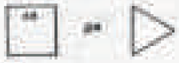

stereo

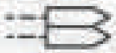

RECTIFIER (65)

semiconductor diode; metallic rectifier; electrolytic rectifier; asymmetrical varistor

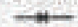

mercury-pool tube power rectifier

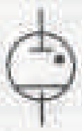

fullwave bridge-type

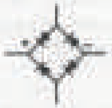

RESISTOR (68)

general

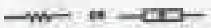

tapped

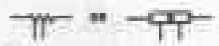

heating

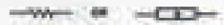

symmetrical varistor resistor, voltage sensitive (silicon carbide, etc.)

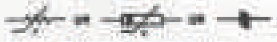

(identification marks replace (*) asterisk)
with adjustable contact

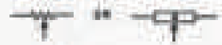

adjustable or continuously adjustable (variable)

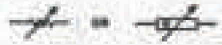

(identification replaces (*) asterisk)

SEMICONDUCTOR DEVICE (73)

(Two Terminal, diode)

semiconductor diode; rectifier

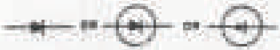

capacitive diode (also Varicap, Varactor, reactance diode, parametric diode)

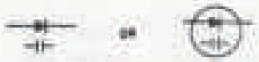

breakdown diode, unidirectional (also backward diode, avalanche diode, voltage regulator diode, Zener diode, voltage reference diode)

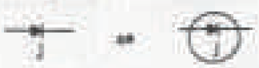

breakdown diode, bidirectional and backward diode (also bipolar voltage limiter)

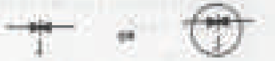

tunnel diode (also Esaki diode)

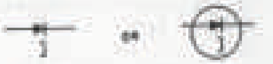

temperature-dependent diode

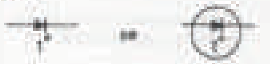

photodiode (also solar cell)

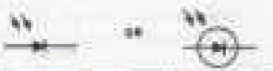

semiconductor diode, PNPN switch (also Shockley diode, four-layer diode and SCR)

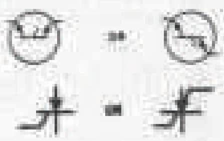

(Multi-Terminal, transistor, etc.)

PNP transistor

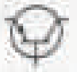

NPN transistor

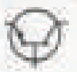

unijunction transistor, N-type base

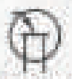

unijunction transistor, P-type base

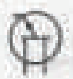

field-effect transistor, N-type base

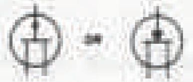

field-effect transistor, P-type base

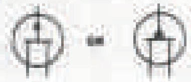

semiconductor triode, PNPN-type switch

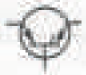

semiconductor triode, NPNP-type switch

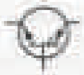

NPN transistor, transverse-biased base

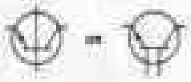

PNIP transistor, ohmic connection to the intrinsic region

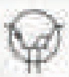

NPIN transistor, ohmic connection to the intrinsic region

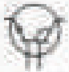

PNIN transistor, ohmic connection to the intrinsic region

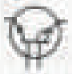

NPIP transistor, ohmic connection to the intrinsic region

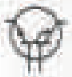

SQUIB (75)

explosive

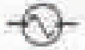

igniter

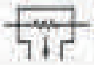

sensing link; fusible link operated

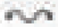

SWITCH (76)

push button, circuit closing (make)

push button, circuit opening (break)

nonlocking; momentary circuit closing (make)

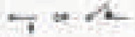

nonlocking; momentary circuit opening (break)

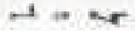

transfer

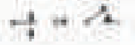

locking, circuit closing (make)

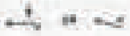

locking, circuit opening (break)

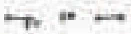

transfer, 3-position

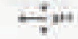

wafer

(example shown: 3-pole 3-circuit with 2 non-shorting and 1 shorting moving contacts)

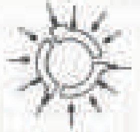

safety interlock, circuit opening and closing

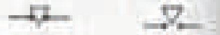

2-pole field-discharge knife, with terminals and discharge resistor

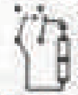

(identification replaces (*) asterisk)

TRANSFORMER (86)

general

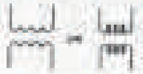

magnetic-core

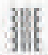

one winding with adjustable inductance

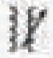

separately adjustable inductance

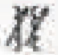

adjustable mutual inductor, constant-current

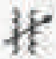

INDEX

A

Abbreviations 23
Abrasive cleaner 118
Acid-type flux 130
Active devices 136
Aerial 18, 31
  trimming 43
  tuning coil 28
AF amplifier 49, 53
AF transistor 34
Air-core coils 34
Air dielectric 104
Alignment 67, 87
Amperes (amps) 25
Amplification 21, 23, 35, 49
Amplifier 12, 47, 75
  FET 120
  IF 84
Amplitude 15
  modulation (AM) 18
Antennas (*See aerials*)
Anti-log 100
Audio frequency 16, 75
Audio power output 58
Autodyne frequency changer 87
Automatic gain control (AGC) 70

B

Base 108
Basic amplifier 47
Basic crystal set 31
Basic quantities 24
Beehive capacitors 106
Bias circuits 61
Bias resistor 47
Bit temperature 121
Blocking 75
Breadboard construction 112
Breadboard units 113
Broadcast station 15
Bus bar 115

C

Capacitance 24
Capacitive coupling 73
Capacitor 19, 100
  ceramic 106
  ceramic-disc 101
  ceramic-plate 101
  characteristics 106
  colour codes 103
Capacity coupling 106
Carbon-track potentiometers 97, 98
Cellulose paint 118
Choke-capacitance coupling 76
Christmas-tree layout 112
Class A output 55, 91
Class AB operation 56
Class B operation 56
Class B output 91
Coil 19
  calculations 29
  former 78
  tapping points 28
Collector 108, 109
Colour codes, resistors, 95, 96
  capacitors 103
Component mounting, printed circuits 118
Component suppliers 133
Components 94
Conductor spacing, printed circuits 117
Connections, potentiometer 100
Convertor 83, 85
Copper 'land' 117
Coupling 35, 45
  coil 45
Crossover distortion 57
Crystal sets 27
  basic 31
  double-diode 35
Current bias 61

D

Damaged PC boards 123
Deaf-aid earpiece 29
Detector 20, 29
Diode 20, 35, 49
Directional aerial 32
Domestic receiver 27, 45, 60, 84
Drain 125
Drilling PC panels 119

E

Earphones 29
Earpiece characteristics 58
Earth 14, 31
return 14
Efficiency, tuning coils 41
Electrode (FET) 125
Electrolytic capacitors 102, 103
Emitter 108
current 62
Epitaxial transistors 107
Etchant 119
Etched-foil capacitors 102
Etching 119

F

Farads 24
Feedback 33
Ferric chloride 119
Ferrite rod 37, 42
Ferrite slab 43
Field-effect transistors (FET) 107, 125
Four-stage reflex 79, 80
Frequency 19
modulation (FM) 18

G

Gain 36
Ganged controls 67
Gate 125
Germanium diode 29
Germanium transistors 63, 105

H

Headphone characteristics 58
Heat damage 121
Heat sink 108, 121
Henries 24
Hertz 16
High-impedance phones 29, 58, 78
High-stability resistors 94
Holders, ICs 128

I

IC holders 128
IF amplifier 84, 91
IF frequencies 84
IF transformer 85
Induced current 19
Inductance 24, 37, 42
Induction 19
Inductive coupling 45, 69, 70, 74
Injected signal 88
Integrated circuit (IC) 129
Intermediate frequency 70

J

Jumper 117

L

Linear frequency 106
Linear potentiometer 99
Linear taper 100
Load balancing 58
Load impedances 59
Local oscillator 83, 85
Logarithmic potentiometer 99
Long-wave coil 44
Long wavelength 16
Loose coupling 46
Loudspeakers 12, 58
output 50, 82
Low impedance phones 59

M

Magnetic earphones 30
Medium-wave 42, 43, 44
Medium wavelength 16
Megaphone 12
Metal-film resistors 94
Metal-oxide resistors 94
Mica dielectric 105
Microphone 12, 17
Miniature capacitors 101
Mixer 81, 83, 87, 88
Modulation 17, 19
Moulded carbon resistors 94
Mylar capacitors 101

N

N-channel 125
Nitric acid 119
Non-polarized capacitors 102
N-p-n transistors 64, 105

O

Ohms 24
Oscillation 72

Oscillator 87
Outline shapes, transistors 107, 109
ICs 128
Output 55
transformers 58, 59

P

P-channel 125
Paper tubular capacitors 100
Passive components 127
Paxolin 29, 77
PC board 109, 116
PC pattern 118
Peg board 114
Plastic dielectric 106
P-n-p transistors 63, 110
Polarity of transistors 110
Polarized capacitors 102
Polycarbonate capacitors 102
Polystyrene capacitors 101
Potentiometer 60, 97
Power rating 94
Power transistors 109
Pre-aligned transformers 85
Pre-amplifiers 53
Preferred numbers 96, 97
Prefixes 24
Pressure waves 11
Printed circuits 115, 117
mounting components 118
Proprietary breadboards 113
Public-address system 12
Push-pull outputs 37, 57, 74

Q

Q 42
Quantities 24
Quench 24
frequency 75

R

Radio 1 11
Radio 2 12
Radio 3 13, 32
Radio 4 14, 22
Radio-frequency choke (RFC) 76
Radio-frequency waves 15
Radio waves 16
Rate of etching 119
Reactance 77
Reception 33
Reflex receivers 71, 75
Regeneration 71, 73
Regenerative receivers 71
Removing components 122
Resin-cored solder 119
Resist 118
Resistance 13, 24, 37
Resistors 94
colour code 94, 95
tolerances 95, 96
Resonant frequency 19
RF carrier 17
RF transistors 53
RFC coil 78

S

Saturation temperature 120
Selectivity 37, 71, 83
Semi-log 100
Sensitivity 71, 74, 75
Shortwave band 18
Signal generator 88
Silica 'chip' 128
Silicon diode 29
Silicon planar transistors 107
Silicon transistors 64, 107
Silvered mica capacitors 101
Simple receiver 21
Sizes of resistors 94
Sizes of soldering irons 120
Slider-type potentiometer 99, 100
Soldering 119
iron sizes 120
Solid dielectric 105
Sound waves 11, 17
Source 125
Speech 11
Speed of light 16
Speed of radio waves 15
Square law 106
'Squeal' 72
Squegging 75
Stabilization 62
Stabilized circuit 51
Stages, amplifiers 49
Standard IF 84
Standard symbols 24
Strength of sound waves 12
Superhet front end 85, 86
Superhet receiver 83
Suppliers 133
Symbols 24, 25

T

Tag boards 111
Tag strip 29, 111
Tantalum bead capacitors 102
Tantalum capacitors 103
Tapping points 36

Telephone, simple 14
Terminal strips 109
Three-stage reflex 79
Time constant 75
Timing 121
Tone control 59
'Top-band' 43
Transformer coupling 45, 46
Transformer drive 91
Transformer ratios 48
Transformerless output 91
Transistor amplifier 47
Transistor calculations 69
Transistor connections 96
Transistor crystal set 34
Transistor lead configurations 108
Transistor crystal se 34
Transistor lead configurations 108
Transistor polarity 110
Transistors 34, 61, 107
Trawler band 42
TRF receiver 67
Trimmer capacitors 105, 106
Tubular ceramic capacitors 101
Tuned circuit 18, 21, 27
Tuning 32
  capacitors 104
  range 37
TV aerial 31
Two-stage amplifier 51
Types of resistors 94, 95

V

Variable capacitor 29, 37, 104
Variable resistors 97
VHF 18
Voltage bias 62
Voltage doubling 36
Voltage, soldering irons 120
Volts 25
Volume control 52

W

Wattage, soldering irons 120
Watts 25
Wavelength 15
Wire aerial 18
Wire-wound potentiometers 97
Wire-wound resistors 94
Wiring connections, potentiometers 100
  transistors 108

22 RADIO AND RECEIVER PROJECTS

FOR THE EVIL GENIUS

AM 540 680 900 1300 1600 KHz
FM 88 94 100 105 108 MHz

- Teaches how to read a schematic, to solder, to use tools, and construction techniques
- Discusses all types of radios and receivers
- Covers both early radio models and contemporary options

TAB ELECTRONICS

TOM PETRUZZELLIS

22 Radio Receiver Projects for the Evil Genius

Evil Genius Series

Bionics for the Evil Genius: 25 Build-it-Yourself Projects

Electronic Circuits for the Evil Genius: 57 Lessons with Projects

Electronic Gadgets for the Evil Genius: 28 Build-it-Yourself Projects

Electronic Games for the Evil Genius

Electronic Sensors for the Evil Genius: 54 Electrifying Projects

50 Awesome Auto Projects for the Evil Genius

50 Model Rocket Projects for the Evil Genius

Mechatronics for the Evil Genius: 25 Build-it-Yourself Projects

MORE Electronic Gadgets for the Evil Genius: 40 NEW Build-it-Yourself Projects

101 Spy Gadgets for the Evil Genius

123 PIC® Microcontroller Experiments for the Evil Genius

123 Robotics Experiments for the Evil Genius

PC Mods for the Evil Genius

Solar Energy Projects for the Evil Genius

25 Home Automation Projects for the Evil Genius

51 High-Tech Practical Jokes for the Evil Genius

22 Radio Receiver Projects for the Evil Genius

TOM PETRUZZELLIS

New York Chicago San Francisco Lisbon London Madrid
Mexico City Milan New Delhi San Juan Seoul
Singapore Sydney Toronto

Copyright © 2008 by The McGraw-Hill Companies, Inc. All rights reserved. Manufactured in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

0-07-159475-2

The material in this eBook also appears in the print version of this title: 0-07-148929-0.

All trademarks are trademarks of their respective owners. Rather than put a trademark symbol after every occurrence of a trademarked name, we use names in an editorial fashion only, and to the benefit of the trademark owner, with no intention of infringement of the trademark. Where such designations appear in this book, they have been printed with initial caps.

McGraw-Hill eBooks are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. For more information, please contact George Hoare, Special Sales, at george_hoare@mcgraw-hill.com or (212) 904-4069.

TERMS OF USE

This is a copyrighted work and The McGraw-Hill Companies, Inc. ("McGraw-Hill") and its licensors reserve all rights in and to the work. Use of this work is subject to these terms. Except as permitted under the Copyright Act of 1976 and the right to store and retrieve one copy of the work, you may not decompile, disassemble, reverse engineer, reproduce, modify, create derivative works based upon, transmit, distribute, disseminate, sell, publish or sublicense the work or any part of it without McGraw-Hill's prior consent. You may use the work for your own noncommercial and personal use; any other use of the work is strictly prohibited. Your right to use the work may be terminated if you fail to comply with these terms.

THE WORK IS PROVIDED "AS IS." McGRAW-HILL AND ITS LICENSORS MAKE NO GUARANTEES OR WARRANTIES AS TO THE ACCURACY, ADEQUACY OR COMPLETENESS OF OR RESULTS TO BE OBTAINED FROM USING THE WORK, INCLUDING ANY INFORMATION THAT CAN BE ACCESSED THROUGH THE WORK VIA HYPERLINK OR OTHERWISE, AND EXPRESSLY DISCLAIM ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. McGraw-Hill and its licensors do not warrant or guarantee that the functions contained in the work will meet your requirements or that its operation will be uninterrupted or error free. Neither McGraw-Hill nor its licensors shall be liable to you or anyone else for any inaccuracy, error or omission, regardless of cause, in the work or for any damages resulting therefrom. McGraw-Hill has no responsibility for the content of any information accessed through the work. Under no circumstances shall McGraw-Hill and/or its licensors be liable for any indirect, incidental, special, punitive, consequential or similar damages that result from the use of or inability to use the work, even if any of them has been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause arises in contract, tort or otherwise.

DOI: 10.1036/0071489290

About the Author

Thomas Petruzzellis is an electronics engineer currently working at the geophysical laboratory at the State University of New York, Binghamton. Also an instructor at Binghamton, with 30 years' experience in electronics, he is a veteran author who has written extensively for industry publications, including *Electronics Now*, *Modern Electronics*, *QST, Microcomputer Journal*, and *Nuts & Volts*. Tom wrote five previous books, including an earlier volume in this series, *Electronic Sensors for the Evil Genius*. He is also the author of *Create Your Own Electronics Workshop*; *STAMP 2 Communications and Control Projects*; *Optoelectronics, Fiber Optics, and Laser Cookbook*; *Alarm, Sensor, and Security Circuit Cookbook*, all from McGraw-Hill. He lives in Vestal, New York.

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

This page intentionally left blank

Acknowledgments

I would like to thank the following people and companies listed below for their help in making this book possible. I would also like to thank senior editor Judy Bass and all the folks at McGraw-Hill publications who had a part in making this book possible. We hope the book will inspire both radio and electronics enthusiasts to build and enjoy the radio projects in this book.

Richard Flagg/RF Associates

Wes Greenman/University of Florida

Charles Higgins/Tennessee State University

Fat Quarters Software

Radio-Sky Publishing

Ramsey Electronics

Vectronics, Inc

Russell Clift

Todd Gale

Eric Vogel

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

This page intentionally left blank

For more information about this title, click here

Contents

Acknowledgments vii

Introduction xi

1 Radio Background and History 1

2 Identifying Components and Reading Schematics 12

3 Electronic Parts Installation and Soldering 25

4 AM, FM, and Shortwave Crystal Radio Projects 39

5 TRF AM Radio Receiver 49

6 Solid-State FM Broadcast Receiver 59

7 Doerle Single Tube Super-Regenerative Radio Receiver 70

8 IC Shortwave Radio Receiver 81

9 80/40 Meter Code Practice Receiver 94

10 WWV 10 MHz “Time-Code” Receiver 104

11 VHF Public Service Monitor (Action-Band) Receiver 116

12 6 & 2-Meter Band Amateur Radio Receiver 127

13 Active and Passive Aircraft Band Receivers 140

14 VLF or Very Low Frequency Radio Receiver 153

15 Induction Loop Receiving System 165

16 Lightning Storm Monitor 175

17 Ambient Power Receiver 186

18 Earth Field Magnetometer Project 192

19 Sudden Ionospheric Disturbance (SIDs) Receiver 203

20 Aurora Monitor Project 212

21 Ultra-Low Frequency (ULF) Receiver 224

22 Jupiter Radio Telescope Receiver 233

23 Weather Satellite Receiver 246

24 Analog to Digital Converters (ADCs) 262

Appendix: Electronic Parts Suppliers 273

Index 277

This page intentionally left blank

Introduction

22 Radio Receiver Projects for the Evil Genius was created to inspire readers both young and old to build and enjoy radio and receiver projects, and perhaps propel interested experimenters into a career in radio, electronics or research. This book is for people who are interested in radio and electronics and those who enjoy building and experimenting as well as those who enjoy research.

Radio encompasses many different avenues for enthusiasts to explore, from simple crystal radios to sophisticated radio telescopes. This book is an attempt to show electronics and radio enthusiasts that there is a whole new world "out there" to explore.

Chapter 1 will present the history and background and elements of radio, such as modulation techniques, etc. Chapter 2 will help the newcomers to electronics, identifying components and how to look and understand schematics vs. pictorial diagrams. Next, Chapter 3 will show the readers how to install electronic components onto circuit boards and how to correctly solder before embarking on their new radio building adventure.

We will start our adventure with the simple "lowly" crystal radio in Chapter 4. Generally crystal radios are only thought of as simple AM radios which can only pickup local broadcast stations. But did you know that you can build crystal radios which can pickup long-distance stations as well as FM and shortwave broadcasts from around the world? You will learn how to build an AM, FM and shortwave crystal radio, in this chapter.

In Chapter 5, you will learn how AM radio is broadcast, from a radio station to a receiver in your home, and how to build your own TRF or Tuned Radio Frequency AM radio receiver. In Chapter 6, we will discover how FM radio works and how to build an FM radio with an SCA output for commercial free radio broadcasts.

Chapter 7 will present the exciting world of shortwave radio. Shortwave radio listening has a large following and encompasses an entire hobby in itself. You will be able to hear shortwave stations from around the world, including China, Russia, Italy, on your new shortwave broadcast receiver. Old time radio buffs will be interested in the single tube Doerle super-regenerative shortwave radio.

If you are interested in a portable shortwave receiver that you could take on a camping trip, then you may want to construct the multi-band integrated circuit shortwave radio receiver described in Chapter 8.

If you are interested in Amateur Radio or are thinking of learning Morse code or want to increase your code speed, you may want to consider building this 80 and 40-meter code receiver. This small lightweight portable receiver can be built in a small enclosure and taken on camping trips, etc.

In Chapter 10, you will learn how to build and use a WWW time code receiver, which can be used to pick up time signal broadcast from the National Institute of Standards and Technology (NIST) or

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

National Atomic Time Clock in Boulder CO. Time signal broadcasts present geophysical and propagation forecasts as well as marine and sea conditions. They will also help you set the time on your best chronometer.

With the VHF Public service receiver featured in Chapter 11, you will discover the high frequency "action bands" which cover the police, fire, taxis, highway departments and marine frequencies. You will be able to listen-in to all the exciting communications in your hometown.

The 6-meter and 2-meter dual VHF Amateur Radio receiver in Chapter 12 will permit to discover the interesting hobby of Amateur Radio. The 6-meter and 2-meter ham radio bands are two of the most popular VHF bands for technician class licensees. You may discover that you might just want to get you own ham radio license and talk to ham radio operators through local VHF repeaters or to the rest of the world.

Why not build an aircraft radio and listen-in to airline pilots talking from 747s to the control tower many miles away. You could also build the passive Air-band radio which you use to listen-in to your pilot during your own flight. Passive aircraft radios will not interfere with airborne radio so they are permitted on airplanes, without restriction. Check out these two receivers in Chapter 13.

Chapter 14 will also show you how to build an induction communication system, which will allow you to broadcast a signal around home or office using a loop of wire, to a special induction receiver. The induction loop broadcast system is a great aid to the hearing impaired, since it can broadcast to hearing aids as well.

The VLF, or "whistler" radio in Chapter 15, will pickup very very low frequency radio waves from around the world. You will be able to listen to low frequency beacon stations, submarine transmissions and "whistlers" or the radio waves created from electrical storms on the other side of the globe. This project is great for research projects where you can record and later analyze your results by feeding your recorded signal into a sound card running an FFT program. Use your computer to record and analyze these interesting signals. There are many free programs available over the Internet. An FFT audio analyzer program can display the audio spectrum and show you where the signals plot out in respect to frequency.

If you are interested in weather, then you will appreciate the Lightening to Storm Receiver in Chapter 16, which will permit you to "see" the approaching storm berfore it actually arrives. This receiver will permit you to have advanced warning up to 50 miles or more away; it will warn you well in advance of an electrical storm, so you can disconnect any outdoor antennas.

The Ambient Power Module receiver project illustrated in Chapter 17, will allow to you pickup a broad spectrum of radio waves which get converted to DC power, and which can be used to power low current circuits around your home or office. This is a great project for experimentation and research. You can use it to charge cell phones, emergency lights, etc.

Our magnetometer project shown in Chapter 18 can be used to see the diurnal or daily changes in the Earth's magnetic field, and you can record the result to a data-logger or recording multi-meter.

If you are an avid amateur radio operator or shortwave listener, you many want to build a SIDs receiver shown in Chapter 19. A SIDs receiver can be used to determine when radio signals and/or propagation is disturbed by solar storms. This receiver will quickly alert you to unfavorable radio conditions. You can collect the receiver to your personal computer's sound card and use the data recorded to correlate radio propagation against storm conditions.

The Aurora receiver project in Chapter 20 will alert you, with both sound and meter display, when the Earth's magnetic increases just before an Aurora display is about to take place. UFO and Alien contact buffs can use this receiver to know when UFOs are close by.

For those interested in more earthly research projects, why not build your own ULF or ultra low frequently receiver, shown in Chapter 21, which can be utilized for detecting low frequency wave generated by earthquakes and fault lines. With this receiver you will be able to conduct your own research projects on monitoring the pulse of the Earth. You can connect your ELF receiver to a data-logger and record the signals over time to correlate your research with that of others.

You can explore the heavens by constructing your own radio telescope to monitor the radio signals generated from the planet Jupiter. This radio receiver, illustrated in Chapter 22, will pick up radio signals which indicate electrical and or magnetic storms on the Jovian planet. The Radio Jupiter receiver can be coupled to your personal computer, and can be used for a research project to record and analyze these radio storm signals.

Why not construct your own weather satellite receiving station, shown in Chapter 23. This receiver will allow you to receive APT polar satellites broadcasting while passing overhead. You can display the satellite weather maps on the computer's screen or save them later to show friends and relatives.

Chapter 24 discusses different analog to digital converters which you can use to collect and record data from the different receiver projects.

We hope you will find the *22 Radio Projects for the Evil Genius* a fun and thought-provoking book, that will find a permanent place on your electronics or radio bookshelf. Enjoy!

This page intentionally left blank

22 Radio Receiver Projects for the Evil Genius

This page intentionally left blank

Chapter 1

Radio Background and History

Electromagnetic energy encompasses an extremely wide frequency range. Radio frequency energy, both natural radio energy created by lightning and planetary storms as well as radio frequencies generated by man for communications, entertainment, radar, and television are the topic of this chapter. Radio frequency energy, or RF energy, covers the frequency range from the low end of the radio spectrum, around l0 to 25 kHz, which is used by high-power Navy stations that communicate with submerged nuclear submarines, through the familiar AM broadcast band from 550 to 1600 kHz. Next on the radio frequency spectrum are the shortwave bands from 2000 kHz to 30,000 kHz. The next band of frequencies are the very high frequency television channels covering 54 to 2l6 MHz, through the very popular frequency modulation FM band from 88 to l08 MHz. Following the FM broadcast band are aircraft frequencies on up through UHF television channels and then up through the radar frequency band of 1000 to 1500 MHz, and extending through approximately 300 gHz. See frequency spectrum chart in Figure 1-1. The radio frequency spectrum actually extends almost up to the lower limit of visible light frequencies.

Radio history

One of the more fascinating applications of electricity is in the generation of invisible ripples of energy called radio waves. Following Hans Oersted's accidental discovery of electromagnetism, it was realized that electricity and magnetism were related to each other. When an electric current was passed through a conductor, a magnetic field was generated perpendicular to the axis of flow. Likewise, if a conductor was exposed to a change in magnetic flux perpendicular to the conductor, a voltage was produced along the length of that conductor.

Joseph Henry, a Princeton University professor, and Michael Faraday, a British physicist, experimented separately with electromagnets in the early 1800s. They each arrived at the same observation: the theory that a current in one wire can produce a current in another wire, even at a distance. This phenomenon is called electromagnetic induction, or just induction. That is, one wire carrying a current induces a current in a second wire. So far, scientists knew that electricity and magnetism always seemed to affect each other at right angles. However, a major discovery lay hidden just beneath this seemingly simple concept of related perpendicularity, and its unveiling was one of the pivotal moments in modern science.

The man responsible for the next conceptual revolution was the Scottish physicist James Clerk Maxwell (1831–1879), who "unified" the study of electricity and magnetism in four relatively tidy equations. In essence, what he discovered was that electric and magnetic fields were intrinsically related to one another, with or without the presence of a conductive path for electrons to flow. Stated more formally, Maxwell's discovery was this: a changing electric field produces a perpendicular magnetic field, and a changing magnetic field produces a perpendicular electric field. All of this can take place in open space, the alternating electric and magnetic fields supporting each other as they travel through space at the speed of light. This dynamic structure of electric and magnetic fields propagating through space is better known as an electromagnetic wave.

Later, Heinrich Hertz, a German physicist, who is honored by our replacing the expression "cycles per second" with hertz (Hz), proved Maxwell's theory between the years 1886 and l888. Shortly after that, in 1892, Eouard Branly, a French physicist, invented a device that could receive radio waves (as we know them today) and could cause them to ring an electric bell.

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

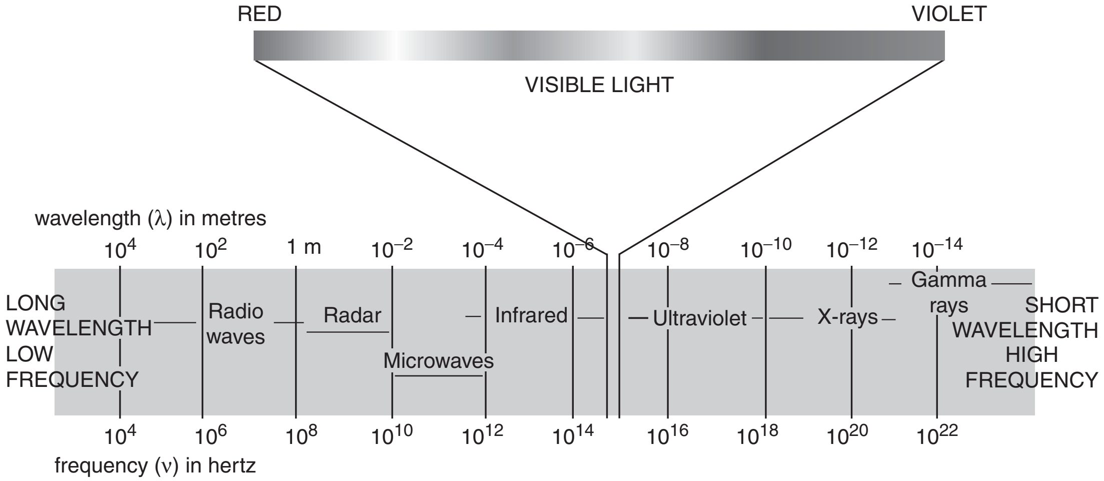

Figure 1-1 *Electromagnetic spectrum*

Note that at the time all the research being conducted in what was to become radio and later radio-electronics, was done by physicists.

In 1895, the father of modem radio, Guglielmo Marconi, of Italy, put all this together and developed the first wireless telegraph and was the first to commercially put radio into ships. The wire telegraph had already been in commercial use for a number of years in Europe. The potential of radio was finally realized through one of the most memorable events in history. With the sinking of the Titanic in 1912, communications between operators on the sinking ship and nearby vessels, and communications to shore stations listing the survivors brought radio to the public in a big way.

AM radio broadcasting began on November 2, 1920. Four pioneers: announcer Leo Rosenberg, engineer William Thomas, telephone line operator John Frazier and standby R.S. McClelland, made their way to a makeshift studio—actually a shack atop the Westinghouse "K" Building in East Pittsburgh—flipped a switch and began reporting election returns in the Harding vs. Cox Presidential race. At that moment, KDKA became the pioneer broadcasting station of the world.

Radio spread like wildfire to the homes of everyone in America in the 1920s. In a few short years, over 75 manufacturers began selling radio sets. Fledgling manufacturers literally came out of garages over-night. Many young radio enthusiasts rushed out to buy parts and radio kits which soon became available.

Radio experimenters discovered that an amplitude-modulated wave consists of a carrier and two identical sidebands which are both above and below the carrier wave. The Navy conducted experiments in which they attempted to pass one sideband and attenuate the other. These experiments indicated that one sideband contained all the necessary information for voice transmission, and these discoveries paved the way for development of the concept of single-sideband or SSB transmission and reception.

In 1923, a patent was granted to John R. Carson on his idea to suppress the carrier and one sideband. In that year the first trans-Atlantic radio telephone demonstration used SSB with pilot carrier on a frequency of 52 kc. Single sideband was used because of limited power capacity of the equipment and the narrow bandwidths of efficient antennas for those frequencies. By 1927, trans-Atlantic SSB radiotelephony was open for public service. In the following years, the use of SSB was limited to low-frequency and wire applications. Early developments in FM transmission suggested that this new mode might prove to be the ultimate in voice communication. The resulting slow development of SSB technology precluded practical SSB transmission and reception at high frequencies. Amateur radio SSB activity followed very much the same pattern. It wasn't until 1948 that amateurs began seriously experimenting with SSB, likely delayed by the wartime blackouts.

The breakthroughs in the war years, and those following the war, were important to the development of HF-SSB communication. Continued advances in technology made SSB the dominant mode of HF radio communication.

The radio-frequency spectrum, once thought to be adequate for all needs, has become very crowded. As the world's technical sophistication progresses, the requirements for rapid and dependable radio communications increase. The competition for available radio spectrum space has increased dramatically. Research and development in modern radio systems has moved to digital compression and narrow bandwidth with highly developed modulation schemes and satellite transmission.

The inventor most responsible for the modern day advances in radio systems was Edwin H. Armstrong. He was responsible for the Regenerative circuit in 1912, the Superheterodyne radio circuit in 1918, the Superregenerative radio circuit design in 1922 and the complete FM radio system in 1933. His inventions and developments form the backbone of radio communications as we know it today. The majority of all radio sets sold are FM radios, all microwave relay links are FM, and FM is the accepted system in all space communications. Unfortunately, Armstrong committed suicide while still embittered in patent lawsuits: later vindicated, his widow received a windfall.

Sony introduced their first transistorized radio in 1960, small enough to fit in a vest pocket, and able to be powered by a small battery. It was durable, because there were no tubes to burn out. Over the next 20 years, transistors displaced tubes almost completely except for very high power, or very high frequency, uses. In the 1970s; LORAN became the standard for radio navigation system, and soon, the US Navy experimented with satellite navigation. Then in 1987, the GPS constellation of satellites was launched and navigation by radio in the sky had a new dimension. Amateur radio operators began experimenting with digital techniques and started to send pictures, faxes and teletype via the personal computer through radio. By the late 1990s, digital transmissions began to be applied to radio broadcasting.

Types of radio waves

There are many kinds of natural radiative energy composed of electromagnetic waves. Even light is electromagnetic in nature. So are shortwaves, X-rays and "gamma" ray radiation. The only difference between these kinds of electromagnetic radiation is the frequency of their oscillation (alternation of the electric and magnetic fields back and forth in polarity). By using a source of AC voltage and a device called an antenna, we can create electromagnetic waves.

It was discovered that high frequency electromagnetic currents in an antenna wire, which in turn result in a high frequency electromagnetic field around the antenna, will result in electromagnetic radiation which will move away from the antenna into free space at the velocity of light, which is approximately 300,000,000 meters per second.

In radio transmission, a radiating antenna is used to convert a time-varying electric current into an electromagnetic wave, which freely propagates through a nonconducting medium such as air or space. An antenna is nothing more than a device built to produce a dispersing electric or magnetic field. An electromagnetic wave, with its electric and magnetic components, is shown in Figure 1-2.

When attached to a source of radio frequency signal generator, or transmitter, an antenna acts as a transmitting device, converting AC voltage and current into electromagnetic wave energy. Antennas also have the ability to intercept electromagnetic waves and convert their energy into AC voltage and current. In this mode, an antenna acts as a receiving device.

Radio frequencies spectrum

Radio frequency energy is generated by man for communications, entertainment, radar, television,

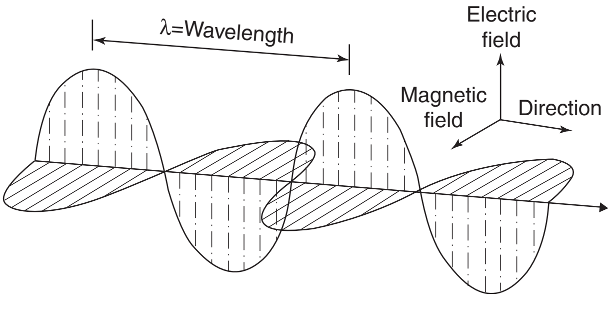

Figure 1-2 *Magnetic vs. electric wave*

navigation, etc. This radio frequency or (RF) energy covers quite a large range of radio frequencies from the low end of the radio spectrum from l0 to 25 kHz, which is the domain occupied by the high-power Navy stations that communicate with submerged nuclear submarines: these frequencies are called Very Low Frequency waves or VLF. Above the VLF frequencies are the medium wave frequencies or (MW), i.e. the AM radio broadcast band from 550 to 1600 kHz. The shortwave bands or High Frequency or (HF) bands cover from 2000 kHz to 30,000 kHz and make use of multiple reflections from the ionosphere which surrounds the Earth, in order to propagate the signals to all parts of the Earth. The Very High Frequencies or VHF bands begin around 30 MHz; these lower VHF frequencies are called low-band VHF. Mid-band VHF frequencies begin around 50 MHz which cover the lowest TV channel 2. Low-band television channels 2 through 13 cover the 54 to 2l6 MHz range. The popular frequency modulation or (FM) broadcast band covers the range from 88 to l08 MHz, which is followed by low-band Air-band frequencies from 118 to 136 MHz. High-band VHF frequencies around 144 are reserved for amateur radio, public service around 150 MHz, with marine frequencies around 156 MHz. UHF frequencies begin around 300 MHz and go up through the radar frequency band of 1000 to 1500 MHz, and extending through approximately 300 gHz. Television channels 14 through 70 are placed between 470 and 800 MHz. American cell phone carriers have cell phone communications around 850 MHz. Geosynchronous weather satellites signals are placed around 1.6 GHz, and PCS phone devices are centered around 1.8 GHz. The Super-high frequency (SHF) bands range from 3 to 30 GHz, with C-band microwave frequencies around 3.8 GHz, then X-band, from 7.25 to 8.4 GHz, followed by the KA and KU-band microwave bands.

Table 1-1 illustrates the division of radio frequencies. The radio frequency spectrum extends almost up to the lower limit of visible light frequencies, with just the infrared frequencies lying in between it and visible light. The radio frequency spectrum is a finite resource which must be used and shared with many people and agencies around the world, so cooperation is very important.

So how does a radio work? As previously mentioned, radio waves are part of a general class of waves known as electromagnetic waves. In essence, they are electrical and magnetic energy which travels through space in the form of a wave. They are different from sound waves (which are pressure waves that travel through air or water, as an example) or ocean waves (similar to sound waves in water, but much lower in frequency and are much larger). The wave part is similar, but the energy involved is electrical and magnetic, not mechanical.

Electromagnetic waves show up as many things. At certain frequencies, they show up as radio waves. At much higher frequencies, we call them infrared light. Still higher frequencies make up the spectrum known as visible light. This goes on up into ultraviolet light, and X-rays, things that radio engineers rarely have to worry about. For our discussions, we'll leave light to the physicists, and concentrate on radio waves.

Radio waves have two important characteristics that change. One is the amplitude, or strength of the wave. This is similar to how high the waves are coming into shore from the ocean. The bigger wave has a higher amplitude. The other thing is frequency. Frequency is how often the wave occurs at any point. The faster the wave repeats itself, the higher the frequency. Frequency is measured by the number of times in a second that the wave repeats itself. Old timers remember when frequency was described in units of cycles per second. In more recent times we have taken to using the simplified term of hertz (named after the guy who discovered radio waves). Metric prefixes are often used, so that 1000 hertz is a kilohertz, one million hertz is a megahertz, and so on.

A typical radio transmitter, for example, takes an audio input signal, such as voice or music and amplifies it. The amplified audio is in turn sent to a modulator and an RF exciter which comprises the radio frequency transmitter. The exciter in the transmitter generates a main carrier wave. The RF signal from the exciter is further amplified by a power amplifier and then the RF signal is sent out to an antenna which radiates the signal into the sky and out into the ionosphere. Depending upon the type of transmitter used the modulation technique can be either AM, FM, SSB signal sideband, CW, or digital modulation, etc.

AM modulation

Amplitude modulation (AM) is a technique used in electronic communication, most commonly for transmitting information via a carrier wave wirelessly.

Table 1-1

Radio frequency spectrum chart

Frequency range			
Extremely Low Frequency (ELF)	0	to	3 kHz
Very Low Frequency (VLF)	3 kHz	to	30 kHz
Radio Navigation & maritime/aeronautical mobile	9 kHz	to	540 kHz
Low Frequency (LF)	30 kHz	to	300 kHz
Medium Frequency (MF)	300 kHz	to	3000 kHz
AM Radio Broadcast	540 kHz	to	1630 kHz
Travellers Information Service	1610 kHz		
High Frequency (HF)	3 MHz	to	30 MHz
Shortwave Broadcast Radio	5.95 MHz	to	26.1 MHz
Very High Frequency (VHF)	30 MHz	to	300 MHz
Low Band: TV Band 1 - Channels 2-6	54 MHz	to	88 MHz
Mid Band: FM Radio Broadcast	88 MHz	to	174 MHz
High Band: TV Band 2 - Channels 7-13	174 MHz	to	216 MHz
Super Band (mobile/fixed radio & TV)	216 MHz	to	600 MHz
Ultra-High Frequency (UHF)	300 MHz	to	3000 MHz
Channels 14-70	470 MHz	to	806 MHz
L-band:	500 MHz	to	1500 MHz
Personal Communications Services (PCS)	1850 MHz	to	1990 MHz
Unlicensed PCS Devices	1910 MHz	to	1930 MHz
Superhigh Frequencies (SHF)			
(Microwave)	3 GHz	to	30.0 GHz
C-band	3600 MHz	to	7025 MHz
X-band	7.25 GHz	to	8.4 GHz
Ku-band	10.7 GHz	to	14.5 GHz
Ka-band	17.3 GHz	to	31.0 GHz
Extremely High Frequencies (EHF)			
(Millimeter Wave Signals)	30.0 GHz	to	300 GHz
Additional Fixed Satellite	38.6 GHz	to	275 GHz
Infrared Radiation	300 GHz	to	430 THz
Visible Light	430 THz	to	750 THz
Ultraviolet Radiation	1.62 PHz	to	30 PHz
X-Rays	0.30 PHz	to	30 EHz
Gamma Rays	0.30 EHz	to	3000 EHz

It works by varying the strength of the transmitted signal in relation to the information being sent.

In the mid-1870s, a form of amplitude modulation was the first method to successfully produce quality audio over telephone lines. Beginning in the early 1900s, it was also the original method used for audio radio transmissions, and remains in use by some forms of radio communication—“AM” is often used to refer to the medium-wave broadcast band (see AM Radio–Chapter 5).

Amplitude modulation (AM) is a type of modulation technique used in communication. It works by varying

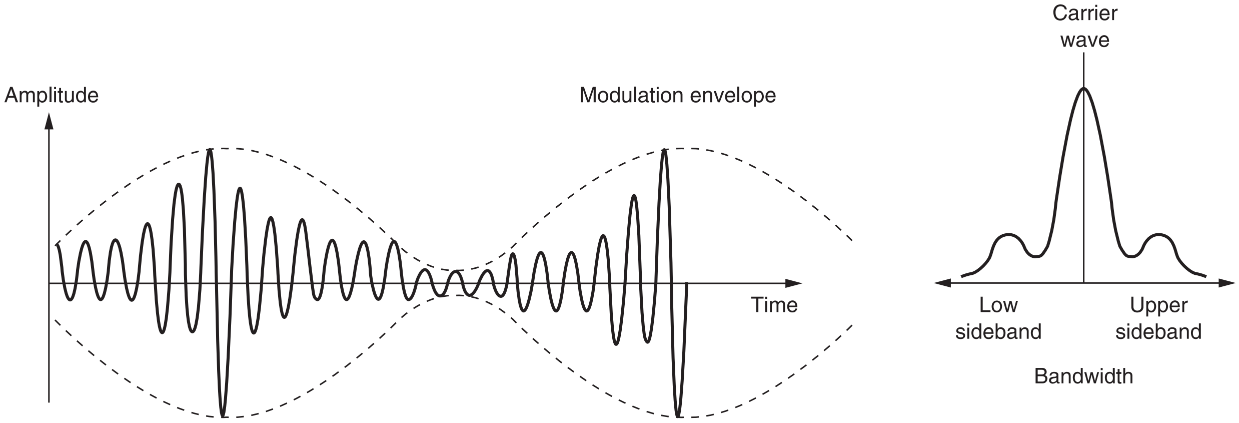

Figure 1-3 *Amplitude modulation waveform*

the strength of the transmitted signal in relation to the information being sent, for example, changes in the signal strength can be used to reflect sounds being reproduced in the speaker. This type of modulation technique creates two sidebands with the carrier wave signal placed in the center between the two sidebands. The transmission bandwidth of AM is twice the signal's original (baseband) bandwidth—since both the positive and negative sidebands are 'copied' up to the carrier frequency, but only the positive sideband is present originally. See Figure 1-3. Thus, double-sideband AM (DSB-AM) is spectrally inefficient. The power consumption of AM reveals that DSB-AM with its carrier has an efficiency of about 33% which is too efficient. The benefit of this system is that receivers are cheaper to produce. The forms of AM with suppressed carriers are found to be 100% power efficient, since no power is wasted on the carrier signal which conveys no information. Amplitude modulation is used primarily in the medium wave band or AM radio band which covers 520 to 1710 kHz. AM modulation is also used by shortwave broadcasters in the SW bands from between 5 MHz and 24 MHz, and in the aircraft band which covers 188 to 136 MHz.

FM modulation

Frequency modulation (FM) is a form of modulation which represents information as variations in the instantaneous frequency of a carrier wave. Contrast this with amplitude modulation, in which the amplitude of the carrier is varied while its frequency remains constant. In analog applications, the carrier frequency is varied in direct proportion to changes in the amplitude of an input signal. Digital data can be represented by shifting the carrier frequency among a set of discrete values, a technique known as frequency-shift keying. The diagram in Figure 1-4, illustrates the FM modulation scheme, the RF frequency is varied with the sound input rather than the amplitude.

FM is commonly used at VHF radio frequencies for high-fidelity broadcasts of music and speech, as in FM broadcasting. Normal (analog) TV sound is also broadcast using FM. A narrowband form is used for voice communications in commercial and amateur radio settings. The type of FM used in broadcast is generally called wide-FM, or W-FM. In two-way radio, narrowband narrow-FM (N-FM) is used to conserve bandwidth. In addition, it is used to send signals into space.

Wideband FM (W-FM) requires a wider bandwidth than amplitude modulation by an equivalent modulating signal, but this also makes the signal more robust against noise and interference. Frequency modulation is also more robust against simple signal amplitude fading phenomena. As a result, FM was chosen as the modulation standard for high frequency, high fidelity radio transmission: hence the term "FM radio." FM broadcasting uses a well-known part of the VHF band between 88 and 108 MHz in the USA.

FM receivers inherently exhibit a phenomenon called capture, where the tuner is able to clearly receive the stronger of two stations being broadcast on

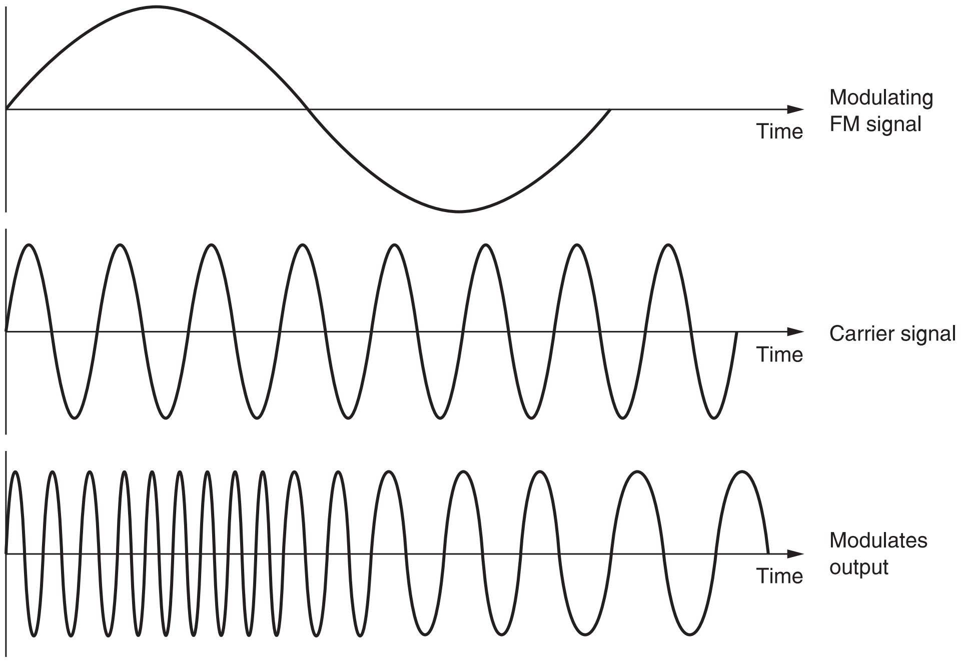

Figure 1-4 *FM modulation waveform*

the same frequency. Problematically, however, frequency drift or lack of selectivity may cause one station or signal to be suddenly overtaken by another on an adjacent channel. Frequency drift typically constituted a problem on very old or inexpensive receivers, while inadequate selectivity may plague any tuner. Frequency modulation is used on the FM broadcast band between 88 and 108 MHz as well as in the VHF and UHF bands for both public service and amateur radio operators.

Single sideband (SSB) modulation

Single sideband modulation (SSB) is a refinement upon amplitude modulation, which was designed to be more efficient in its use of electrical power and spectrum bandwidth. Single sideband modulation avoids this bandwidth doubling, and the power wasted on a carrier, but the cost of some added complexity.

The balanced modulator is the most popular method of producing a single sideband modulated signal. The balanced modulator provides the “sidebands” of energy that exist on either side of the carrier frequency but eliminates the RF carrier, see Figure 1-5. The carrier is removed because it is the sidebands that provide the actual meaningful content of material, within the modulation envelope. In order to make SSB even more efficient, one of these two sidebands is removed by a bandpass. So the intelligence is preserved with SSB and it becomes a more efficient use of radio spectrum energy. It provides almost 9 Decibels (dBs) of signal gain over an amplitude modulated signal that includes an RF “carrier” of the same power level! As the final RF amplification is now concentrated in a single sideband, the effective power output is greater than in normal AM (the carrier and redundant sideband account for well over half of the power output of an AM transmitter). Though SSB uses substantially less bandwidth and power, it cannot be demodulated by a simple envelope detector like standard AM.

SSB was pioneered by telephone companies in the 1930s for use over long-distance lines, as part of a technique known as frequency-division multiplexing (FDM). This enabled many voice channels to be sent down a single physical circuit. The use of SSB meant that the channels could be spaced (usually) just 4000 Hz apart, while offering a speech bandwidth of nominally 300–3400 Hz. Amateur radio operators began to experiment with the method seriously after World War II. It has become a *de facto* standard for long-distance voice radio transmissions since then.

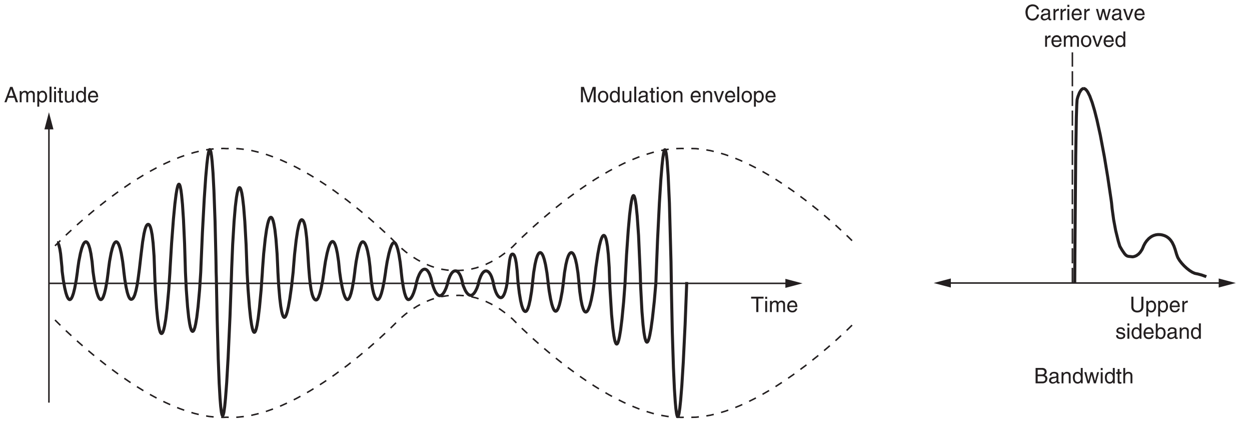

Figure 1-5 *Single sideband modulation waveform*

Single Sideband Suppressed Carrier (SSB-SC) modulation was the basis for all long-distance telephone communications up until the last decade. It was called "L carrier." It consisted of groups of telephone conversations modulated on upper and/or lower sidebands of contiguous suppressed carriers. The groupings and sideband orientations (USB, LSB) supported hundreds and thousands of individual telephone conversations. Single sideband communications are used by amateur radio operators and government, and utility stations primarily in the shortwave bands for long-distance communications.

Shortwave radio

Shortwave radio operates between the frequencies of 1.80 MHz and 30 MHz and came to be referred to as such in the early days of radio because the wavelengths associated with this frequency range were shorter than those commonly in use at that time. An alternate name is HF or high frequency radio. Short wavelengths are associated with high frequencies because there is an inverse relationship between frequency and wavelength. Shortwave frequencies are capable of reaching the other side of the Earth, because these waves can be refracted by the ionosphere, by a phenomenon known as Skywave propagation. High-frequency propagation is dependent upon a number of different factors, such as season of the year, solar conditions, including the number of sunspots, solar flares, and overall solar activity. Solar flares can prevent the ionosphere from reflecting or refracting radio waves.

Another factor which determines radio propagation is the time of the day; this is due to a particular transient atmosphere ionized layer forming only during day when atoms are broken up into ions by sun photons. This layer is responsible for partial or total absorption of particular frequencies. During the day, higher shortwave frequencies (i.e., above 10 MHz) can travel longer distances than lower ones; at night, this property is reversed.

Different types of modulation techniques are used on the shortwave frequencies in addition to AM and FM. AM, amplitude modulation, is generally used for shortwave broadcasting, and some aeronautical communications, while Narrow-band frequency modulation (NFM) is used at the higher HF frequencies. Single sideband or (SSB), is used for long-range communications by ships and aircraft, for voice transmissions by amateur radio operators. CW, Continuous Carrier Wave or (CW), is used for Morse code communications. Various other types of digital communications such as radioteletype, fax, digital, SSTV and other systems require special hardware and software to decode. A new broadcasting technique called Digital Radio Mondiale or (DRM) is a digital modulation scheme used on bands below 30 MHz.

Shortwave listening

Many hobbyists listen to shortwave broadcasters and for some listeners the goal is to hear as many stations from as many countries as possible (DXing); others listen to specialized shortwave utility, or "UTE," transmissions such as maritime, naval, aviation, or military signals. Others focus on intelligence signals. Many, though, tune the shortwave bands for the program content of shortwave broadcast stations, aimed to a general audience (such as the Voice of America, BBC World Service, Radio Australia, etc.). Some even listen to two-way communications by amateur radio operators. Nowadays, as the Internet evolves, the hobbyist can listen to shortwave signals via remotely controlled shortwave receivers around the world, even without owning a shortwave radio (see for example http://www.dxtuners.com). Alternatively, many international broadcasters (such as the BBC) offer live streaming audio on their web-sites. Table 1-2, lists some of the popular shortwave broadcast bands.

Shortwave listeners, or SWLs, can obtain QSL cards from broadcasters, utility stations or amateur radio operators as trophies of the hobby. Some stations even give out special certificates, pennants, stickers and other tokens and promotional materials to shortwave listeners.

Major users of the shortwave radio bands include domestic broadcasting in countries with a widely dispersed population with few long-wave, medium-wave, or FM stations serving them. International broadcasting stations beamed radio broadcasts to foreign audiences.

Table 1-2
Shortwave broadcast chart

Band	*Megahertz Band (MHz)*	*Kilohertz (KHz)*
120 Meter	2.3–2.5 MHz	2300–2500 KHz
90 Meter	3.2–3.40 MHZ	3200–3400 KHz
75 Meter	3.90–4.00 MHZ	3900–4000 KHz
60 Meter	4.750–5.060 MHz	4750–5060 KHz
49 Meter	5.950–6.20 MHz	5950–6200 KHz
41 Meter	7.10–7.60 MHz	7100–7600 KHz
31 Meter	9.20–9.90 MHz	9500–9900 KHz
25 Meter	11.60–12.200 MHz	11600–12200 KHz
22 Meter	13.570–13.870 MHZ	13570–13870 KHz
19 Meter	15.10–15.800 MHz	15100–15800 KHz
16 Meter	17.480–17.900 MHz	17480–17900 KHz
13 Meter	21.450–21.850 MHz	21450–21850 KHz
11 Meter	25.60–26.100 MHz	25600–26100 KHz

Speciality political, religious, and conspiracy theory radio networks, individual commercial and non-commercial paid broadcasts for the north American and other markets. Utility stations transmitting messages not intended for a general public, such as aircraft flying between continents, encoded or ciphered diplomatic messages, weather reporting, or ships at sea. Amateur radio operators have rights to use many frequencies in the shortwave bands; you can hear their communications using different modulation techniques and even obtain a license to communicate in these bands yourself. Contact the Amateur Radio Relay League for more information. Table 1-3 illustrates the amateur radio frequencies and how they are divided between the different license classes. On the shortwave band you will also encounter time signal stations and number stations, thought to be spy stations operating on the shortwave bands.

Types of receivers

A radio signal is transmitted through the ionosphere and is picked up by the antenna in your radio receiver. The antenna is fed to an RF amplifier and usually an intermediate amplifier or IF amplifier and then on to a detector of some sort depending upon the type of receiver you are using. From the detector, the resultant audio signal is amplified and sent to a loudspeaker for listening. Figure 1-6 illustrates a block diagram of a typical AM radio receiver. The antenna is sent to the RF amplifier. The mixer is fed by both the local oscillator and the RF amplifier. The signal from the mixer is sent to a bandpass filter and then on to the first IF amplifier. The first IF amplifier is next sent to the detector and then on to the final audio amplifier stage which drives the speaker. The illustration depicted in Figure 1-7 shows a typical FM receiver block diagram. The antenna feeds the RF amplifier stage. Both the RF amplifier and local oscillator are fed into the mixer. The signal from the mixer is next sent to the IF amplifier stage. From the IF amplifier stage the signal is next sent to the FM demodulator section, which feeds the signal to the voltage amplifier and then the signal is fed to the final audio amplifier stage and on to the speaker. Note the feedback path between, i.e. the AGC or automatic frequency control from the FM demodulator back to

Table 1-3

US amateur radio bands

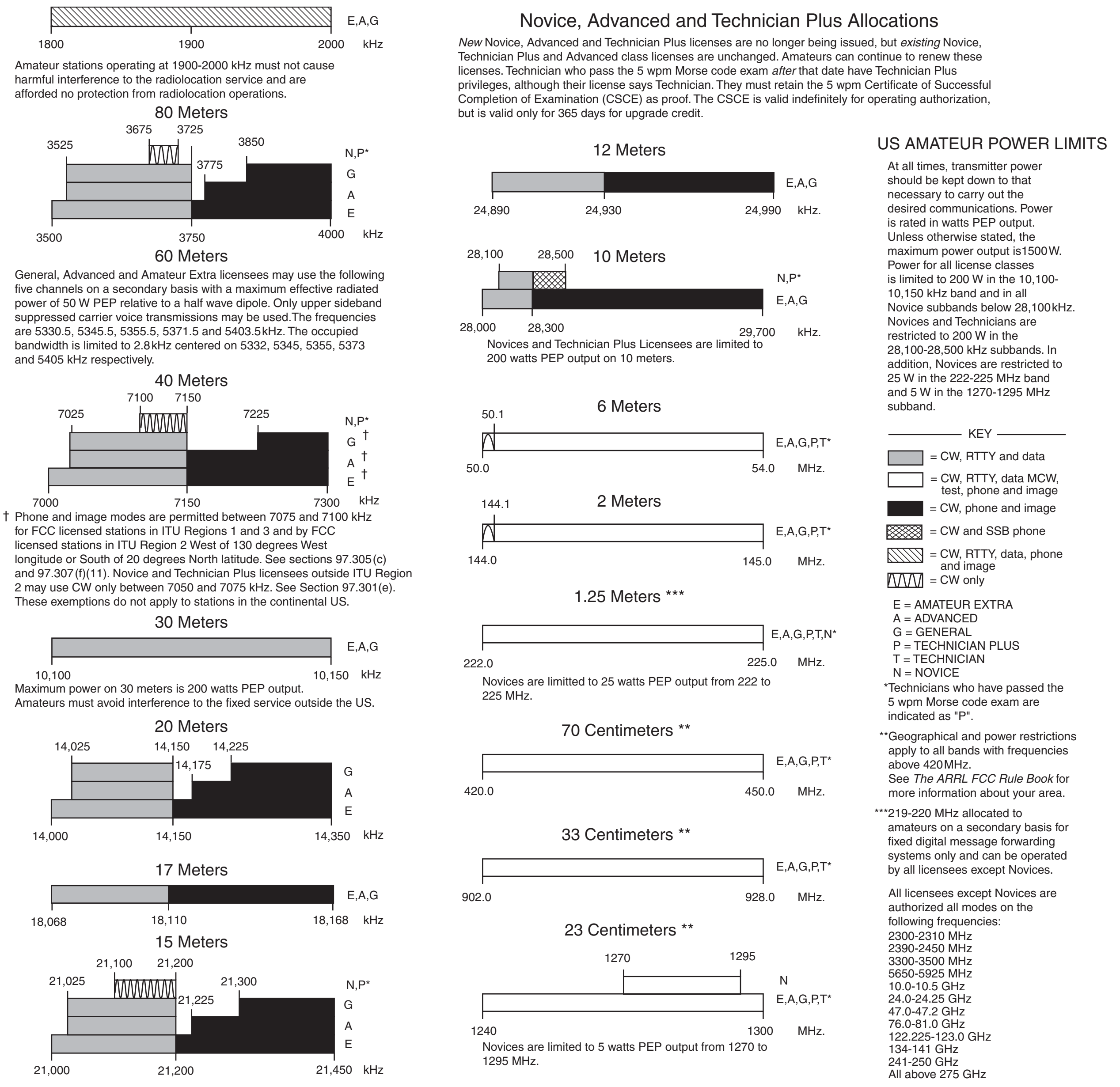
E,A,G
1800
1900
2000
kHz
80 Meters
3675
3725
3525
3850
3775
N,P*
G
A
E
3500
3750
4000
kHz
60 Meters
40 Meters
7100
7150
7025
7225
N,P*
G †
A †
E †
7000
7150
7300
kHz
30 Meters
E,A,G
10,100
10,150
kHz
20 Meters
14,025
14,150
14,225
14,175
G
A
E
14,000
14,150
14,350
kHz
17 Meters
E,A,G
18,068
18,110
18,168
kHz
15 Meters
21,100
21,200
21,025
21,300
21,225
N,P*
G
A
E
21,000
21,200
21,450
kHz
12 Meters
E,A,G
24,890
24,930
24,990
kHz.
28,100
28,500
10 Meters
N,P*
E,A,G
28,000
28,300
29,700
kHz.
6 Meters
50.1
E,A,G,P,T*
50.0
54.0
MHz.
2 Meters
144.1
E,A,G,P,T*
144.0
145.0
MHz.
1.25 Meters ***
E,A,G,P,T,N*
222.0
225.0
MHz.
70 Centimeters **
E,A,G,P,T*
420.0
450.0
MHz.
33 Centimeters **
E,A,G,P,T*
902.0
928.0
MHz.
23 Centimeters **
1270
1295
N
E,A,G,P,T*
1240
1300
MHz.
KEY
▒ = CW, RTTY and data
☐ = CW, RTTY, data MCW, test, phone and image
■ = CW, phone and image
▩ = CW and SSB phone
▨ = CW, RTTY, data, phone and image
= CW only

Amateur stations operating at 1900-2000 kHz must not cause harmful interference to the radiolocation service and are afforded no protection from radiolocation operations.

General, Advanced and Amateur Extra licensees may use the following five channels on a secondary basis with a maximum effective radiated power of 50 W PEP relative to a half wave dipole. Only upper sideband suppressed carrier voice transmissions may be used. The frequencies are 5330.5, 5345.5, 5355.5, 5371.5 and 5403.5 kHz. The occupied bandwidth is limited to 2.8 kHz centered on 5332, 5345, 5355, 5373 and 5405 kHz respectively.

† Phone and image modes are permitted between 7075 and 7100 kHz for FCC licensed stations in ITU Regions 1 and 3 and by FCC licensed stations in ITU Region 2 West of 130 degrees West longitude or South of 20 degrees North latitude. See sections 97.305 (c) and 97.307 (f)(11). Novice and Technician Plus licensees outside ITU Region 2 may use CW only between 7050 and 7075 kHz. See Section 97.301(e). These exemptions do not apply to stations in the continental US.

Maximum power on 30 meters is 200 watts PEP output. Amateurs must avoid interference to the fixed service outside the US.

Novice, Advanced and Technician Plus Allocations

New Novice, Advanced and Technician Plus licenses are no longer being issued, but *existing* Novice, Technician Plus and Advanced class licenses are unchanged. Amateurs can continue to renew these licenses. Technician who pass the 5 wpm Morse code exam *after* that date have Technician Plus privileges, although their license says Technician. They must retain the 5 wpm Certificate of Successful Completion of Examination (CSCE) as proof. The CSCE is valid indefinitely for operating authorization, but is valid only for 365 days for upgrade credit.

Novices and Technician Plus Licensees are limited to 200 watts PEP output on 10 meters.

Novices are limitted to 25 watts PEP output from 222 to 225 MHz.

Novices are limited to 5 watts PEP output from 1270 to 1295 MHz.

US AMATEUR POWER LIMITS

At all times, transmitter power should be kept down to that necessary to carry out the desired communications. Power is rated in watts PEP output. Unless otherwise stated, the maximum power output is1500 W. Power for all license classes is limited to 200 W in the 10,100-10,150 kHz band and in all Novice subbands below 28,100 kHz. Novices and Technicians are restricted to 200 W in the 28,100-28,500 kHz subbands. In addition, Novices are restricted to 25 W in the 222-225 MHz band and 5 W in the 1270-1295 MHz subband.

E = AMATEUR EXTRA
A = ADVANCED
G = GENERAL
P = TECHNICIAN PLUS
T = TECHNICIAN
N = NOVICE

*Technicians who have passed the 5 wpm Morse code exam are indicated as "P".

**Geographical and power restrictions apply to all bands with frequencies above 420 MHz. See *The ARRL FCC Rule Book* for more information about your area.

***219-220 MHz allocated to amateurs on a secondary basis for fixed digital message forwarding systems only and can be operated by all licensees except Novices.

All licensees except Novices are authorized all modes on the following frequencies:
2300-2310 MHz
2390-2450 MHz
3300-3500 MHz
5650-5925 MHz
10.0-10.5 GHz
24.0-24.25 GHz
47.0-47.2 GHz
76.0-81.0 GHz
122.225-123.0 GHz
134-141 GHz
241-250 GHz
All above 275 GHz

the local oscillator. Finally, the shortwave radio block diagram is illustrated in the diagram in Figure 1-8. The antenna line is fed to the RF amplifier section. Both the local oscillator and the RF amplifier are fed to a filter section, which is in turn sent to the IF amplifier section. The output signal from the IF amplifier section is next sent to the product detector. A BFO or beat frequency oscillator signal is sent to the product detector, this is what permits SSB reception. The signal from the product detector is next sent to the audio amplifier and then on to the speaker. The receivers shown are the most common types of receivers. There are in fact many different variations in receiver designs including receivers made to receive special digital signals, which we will not discuss here.

Next, we will move our discussion to identifying electronics components and reading schematics and learning how to solder before we forge ahead and begin building some fun radio receiver projects.

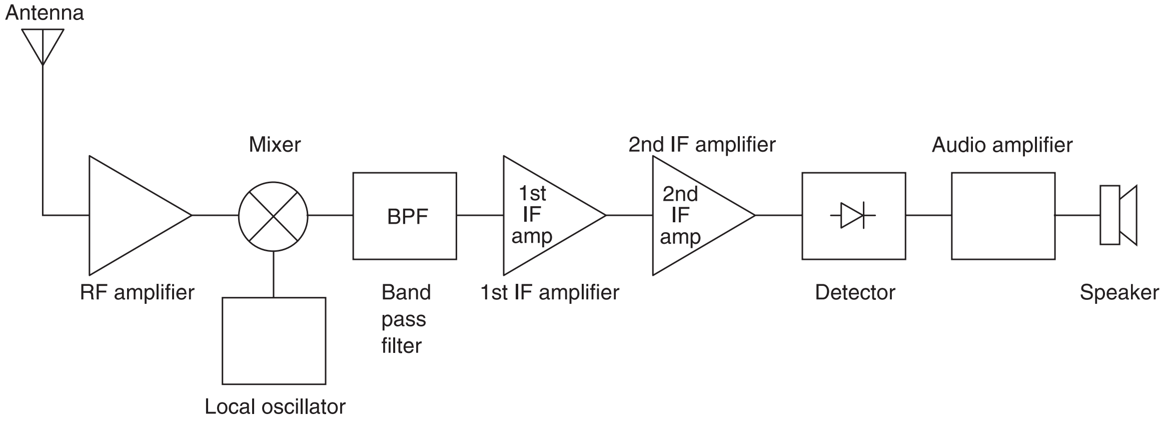
Antenna
Mixer
2nd IF amplifier
Audio amplifier
1st
IF
amp
2nd
IF
amp
BPF
RF amplifier
Band
pass
filter
1st IF amplifier
Detector
Speaker
Local oscillator

Figure 1-6 *AM radio block diagram*

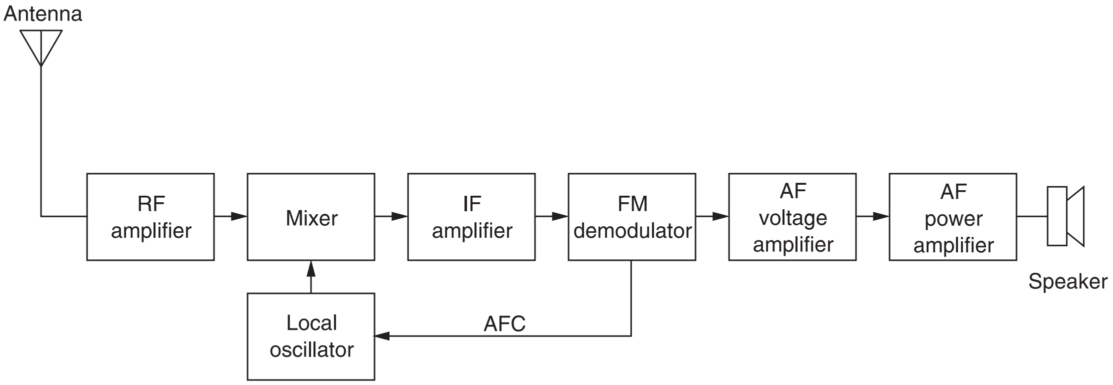
Antenna
RF
amplifier
Mixer
IF
amplifier
FM
demodulator
AF
voltage
amplifier
AF
power
amplifier
Speaker
Local
oscillator
AFC

Figure 1-7 *FM radio block diagram*

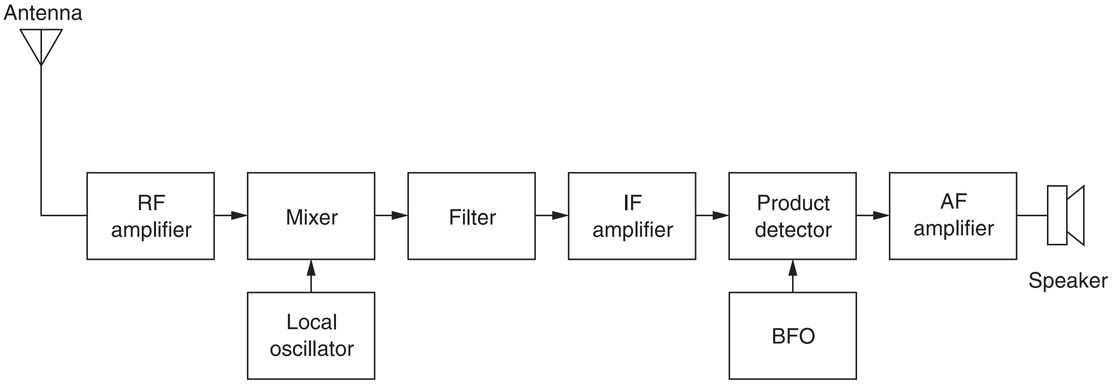
Antenna
RF
amplifier
Mixer
Filter
IF
amplifier
Product
detector
AF
amplifier
Speaker
Local
oscillator
BFO

Figure 1-8 *SSB shortwave receiver block diagram*

Chapter 2

Identifying Components and Reading Schematics

Identifying electronic components

If you are a beginner to electronics or radio, you may want to take a few minutes to learn a little about identifying electronic components, reading schematics, and installing electronic components on a circuit board. You will also learn how to solder, in order to make long-lasting and reliable solder joints.

Electronic circuits comprise electronic components such as resistors and capacitors, diodes, semiconductors and LEDs, etc. Each component has a specific purpose that it accomplishes in a particular circuit. In order to understand and construct electronic circuits it is necessary to be familiar with the different types of components, and how they are used. You should also know how to read resistor and capacitor color codes, recognize physical components and their representative diagrams and pin-outs. You will also want to know the difference between a schematic and a pictorial diagram. First, we will discuss the actual components and their functions and then move on to reading schematics, then we will help you to learn how to insert the components into the circuit board. In the next chapter we will discuss how to solder the components to the circuit board.

The diagrams shown in Figures 2-1, 2-2 and 2-3 illustrate many of the electronic components that we will be using in the projects presented in this book.

Types of resistors

Resistors are used to regulate the amount of current flowing in a circuit. The higher the resistor's value or resistance, the less current flows and conversely a lower resistor value will permit more current to flow in a circuit. Resistors are measured in ohms (Ω) and are identified by color bands on the resistor body. The first band at one end is the resistor's first digit, the second color band is the resistor's second digit and the third band is the resistors's multiplier value. A fourth color band on a resistor represents the resistor's tolerance value. A silver band denotes a 10% tolerance resistor, while a gold band denotes a 5% resistor tolerance. No fourth band denotes that a resistor has a 20% tolerance. As an example, a resistor with a brown, black, and red band will represent the digit (1), the digit (0), with a multiplier value of (00) or 1000, so the resistor will have a value of 1k or 1000 ohms. There are a number of different styles and sizes of resistor. Small resistors can be carbon, thin film or metal. Larger resistors are made to dissipate more power and they generally have an element wound from wire.

A potentiometer (or pot) is basically a variable resistor, generally having three terminals and fitted with a rotary control shaft which varies the resistance as it is rotated. A metal wiping contact rests against a circular carbon or wire wound resistance track. As the wiper arm moves about the circular resistance, the resistance to the output terminals changes. Potentiometers are

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

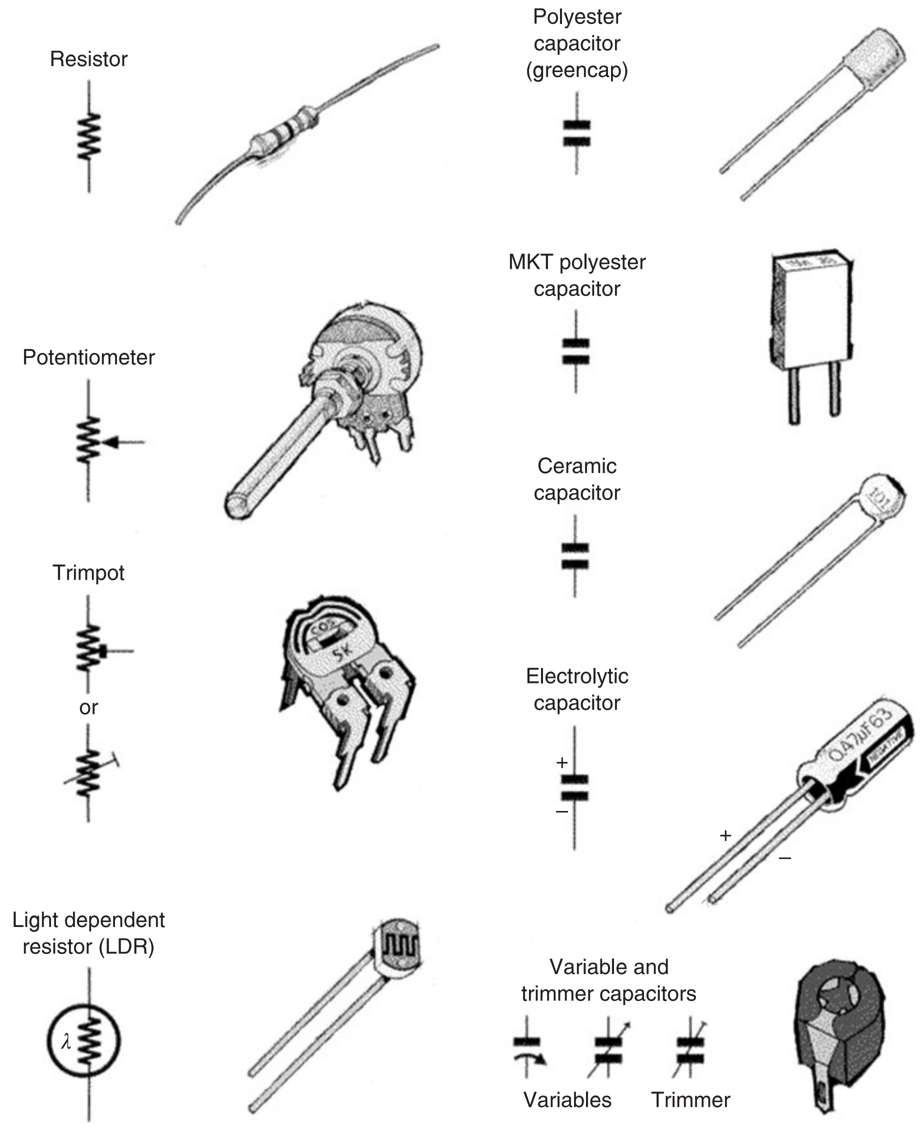
Resistor
Potentiometer
Trimpot
or
Light dependent resistor (LDR)
λ
5K
Polyester capacitor (greencap)
MKT polyester capacitor
Ceramic capacitor
101
Electrolytic capacitor
+
–
0.47μF63
+
–
Variable and trimmer capacitors
Variables
Trimmer

Figure 2-1 *Electronic components 1*

commonly used as volume controls in amplifiers and radio receivers.

A trimpot is a special type of potentiometer which, while variable, is intended to be adjusted once or only occasionally. For this reason a control shaft is not provided but a small slot is provided in the center of the control arm. Trimpots are generally used on printed circuit boards.

A light-dependent resistor (LDR) is a special type of resistor that varies its resistance value according to the amount of light falling on it. When it is in the dark, an LDR will typically have a very high resistance, i.e. millions of ohms. When light falls on the LDR the resistance drops to a few hundred ohms.

Types of capacitors

Capacitors block DC current while allowing varying or AC current signals to pass. They are commonly used for coupling signals from part of a circuit to another part of a circuit, they are also used in timing circuits. There are a number of different types of capacitor as described below.

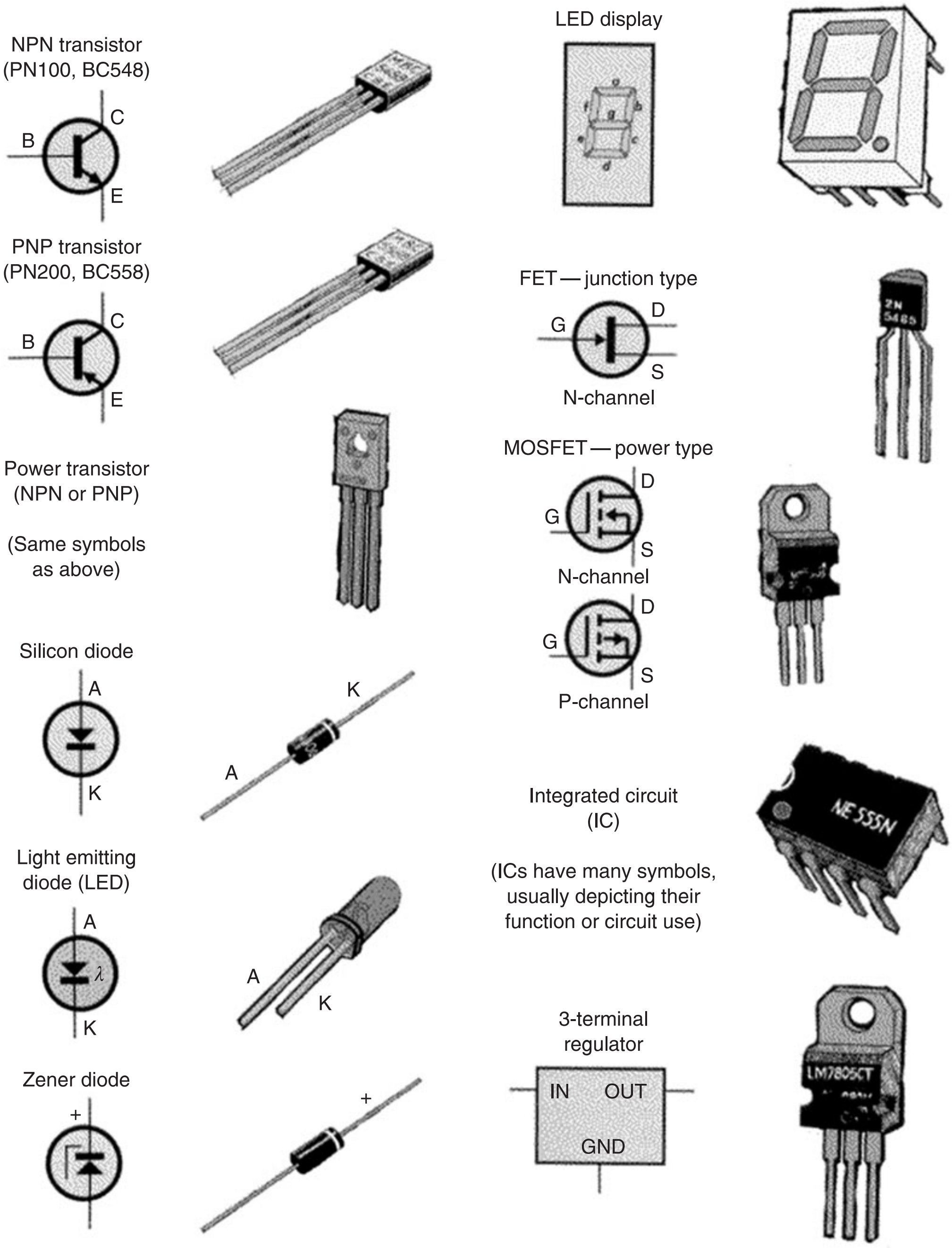
NPN transistor
(PN100, BC548)
B
C
E
PNP transistor
(PN200, BC558)
B
C
E
Power transistor
(NPN or PNP)
(Same symbols
as above)
Silicon diode
A
K
A
K
Light emitting
diode (LED)
A
λ
K
A
K
Zener diode
+
+
LED display
a
b
c
d
e
f
g
FET — junction type
G
D
S
N-channel
2N
5485
MOSFET — power type
G
D
S
N-channel
G
D
S
P-channel
Integrated circuit
(IC)
(ICs have many symbols,
usually depicting their
function or circuit use)
NE555N
3-terminal
regulator
IN
OUT
GND
LM7805CT

Figure 2-2 *Electronic components 2*

Polyester capacitors use polyester plastic film as their insulating dielectric. Some polyester capacitors are called greencaps because they are coated with a green or brown color coating on the outside of the component. Their values are specified in microfarads or (µF), nanofarads, (nF), or picofarads (pF) and range from 1 nF up to about 10 µF. These capacitors do not have polarity and have fixed values.

MKT capacitors are another type of capacitor, but they are rectangular or (block) in shape and are usually yellow in color. One of the major advantages of these capacitors is a more standardized lead spacing, making them more useful for PC board projects. The components can generally be substituted for polyester types.

Ceramic capacitors use a tiny disk of ceramic or porcelain material in their construction for a dielectric

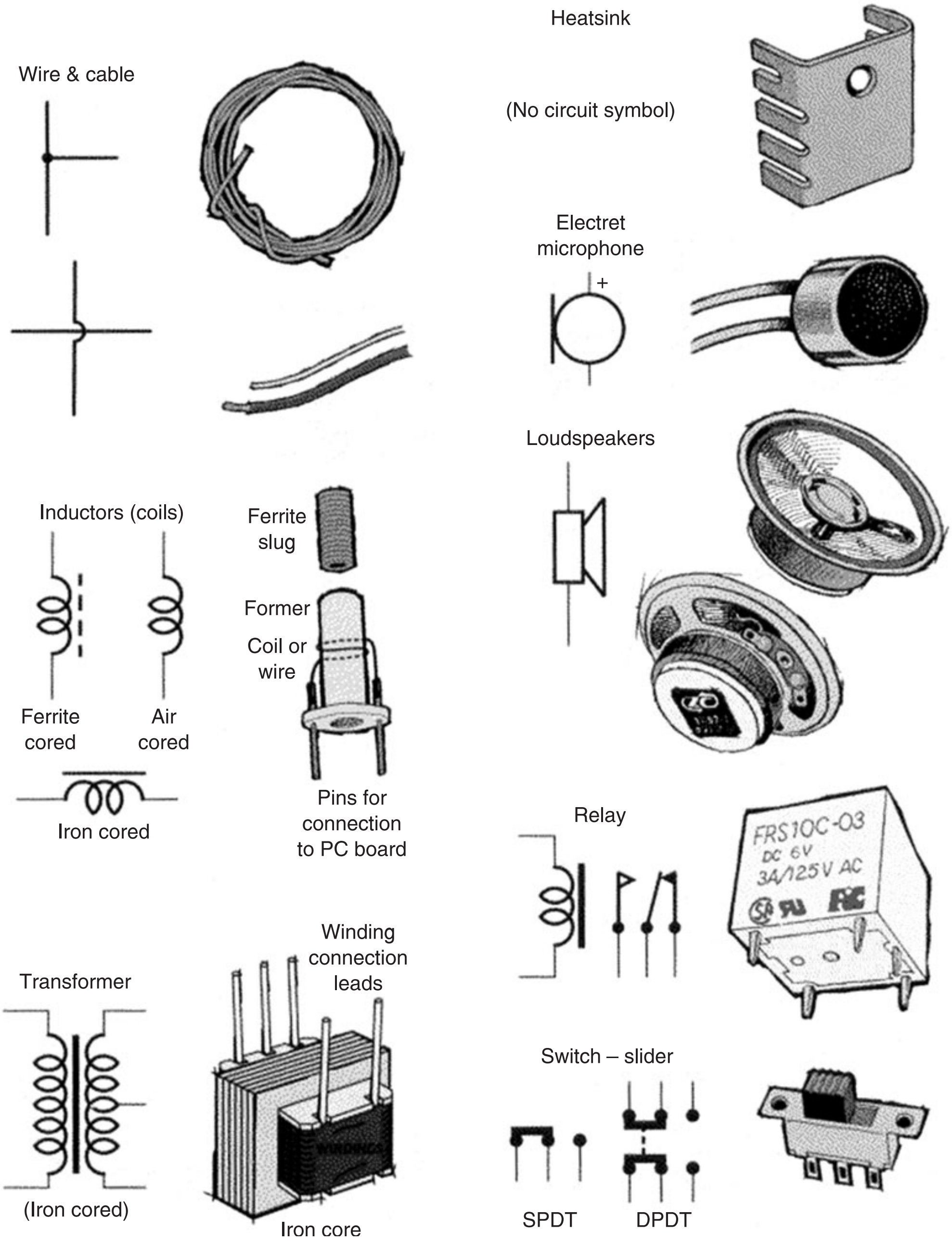
Wire & cable
Inductors (coils)
Ferrite slug
Former
Coil or wire
Ferrite cored
Air cored
Iron cored
Pins for connection to PC board
Transformer
(Iron cored)
Winding connection leads
Iron core
Heatsink
(No circuit symbol)
Electret microphone
+
Loudspeakers
Relay
FRS10C-03
DC 6V
3A/125V AC
Switch – slider
SPDT
DPDT

Figure 2-3 *Electronic components 3*

and they range in value from 1 pF up to 2.2 µF. Those with values above 1 nF are often made with multiple layers of metal electrodes and dielectric, to allow higher capacitance values in smaller bodies. These capacitors are usually called 'multilayer monolithics' and distinguished from lower value disk ceramic types. Ceramic capacitors are often used in RF radio circuits and filter circuits.

Electrolytic capacitors use very thin film of metal oxide as their dielectric, which allows them to provide a large amount of capacitance in a very small volume. They range in value from 100 nF up to hundreds and thousands of microfarads (µF). They are commonly used to filter power supply circuits, coupling audio circuits and in timing circuits. Electrolytic capacitors have polarity and must be installed with respect to these polarity marking. The capacitor will have either a white or black band denoting polarity with a plus (+) or minus (–) marking next to the color band.

Variable capacitors are used in circuits for (trimming) or adjustment, i.e. for setting a frequency. A variable capacitor has one set of fixed plates and one set of plates which can be moved by turning a knob. The dielectric between the plates is usually a thin plastic film. Most variable capacitors have low values up to a few tens of picofarads (pF) and a few hundreds of microfarads for larger variable capacitors.

Diodes

A diode is a semiconductor device which can pass current in one direction only. In order for current to flow the anode (A) must be positive with respect to the cathode (K). In this condition, the diode is said to be forward biased and a voltage drop of about .6 volt appears across its terminals. If the anode is less than .6 volt positive with respect to the cathode, negligible current will flow and the diode behaves as an open circuit.

Types of transistors

Transistors are semiconductor devices that can be used either as electronic switches or to amplify signals. They have three leads, called the Collector, Base, and Emitter. A small current flowing between base and emitter (junction) causes a much larger current to flow between the emitter and collector (junction). There a two basic types of transistors, PNP and NPN styles.

Field Effect Transistors, or FETs, are a different type of transistor, which usually still have three terminals but work in a different way. Here the control element is the “gate” rather than the base, and it is the “gate” voltage which controls the current flowing in the “channel” between the other terminals—the “source” and the “drain.” Like ordinary transistors FETs can be used either as electronic switches or as amplifiers; they also come in P-channel and N-channel types, and are available in small signal types as well as power FETs.

Power transistors are usually larger than the smaller signal type transistors. Power transistors are capable of handling larger currents and voltages. Often metal tabs and heatsinks are used to remove excess heat from the part. These devices are usually bolted to the chassis and are used for amplifying RF or audio energy.

Integrated circuits

Integrated circuits, or ICs, contain all, or most, of the components necessary for a particular circuit function, in one package. Integrated circuits contain as few as 10 transistors or many millions of transistors, plus many resistors, diodes and other components. There are many shapes, styles and sizes of integrated circuits: in this book we will use the dual-in-line style IC, either 8, 14 or 16 pin devices.

Three-terminal regulators are special types of integrated circuits, which supply a regulated or constant and accurate voltage from output regardless (within limits) of the voltage applied to input. They are most often used in power supplies. Most regulators are designed to give specific output voltages, i.e. a ‘LM7805” regulator provides a 5 volt output, but some IC regulators can provide adjustable output based on an external potentiometer which can vary the output voltage.

Heatsinks

Many electronic components generate heat when they are operating. Generally heatsinks are used on semiconductors like transistors to remove heat. Overheating can damage a particular component or the entire circuit. The heatsink cools the transistor and ensures a long circuit life by removing the excess heat from the circuit area.

Light-emitting diodes

Light-emitting diodes, or LEDs, are special diodes which have a plastic translucent body (usually clear, red, yellow, green or blue in color) and a small semiconductor element which emits light when the diode passes a small current. Unlike an incandescent lamp, an LED does not need to get hot to produce light. LEDs must always be forward biased to operate. Special LEDs can also produce infrared light.

LED displays consist of a number of LEDs together in a single package. The most common type has seven elongated LEDs arranged in an "8" pattern. By choosing which combinations of LEDs are lit, any number of digits from "0" through "9" can be displayed. Most of these "7-segment" displays also contain another small round LED which is used as a decimal point.

Types of inductors

Inductors or "coils" are basically a length of wire, wound into a cylindrical spiral (or layers of spirals) in order to increase their inductance. Inductance is the ability to store energy in a magnetic field. Many coils are wound on a former of insulating material, which may also have connection pins to act as the coil's terminals. The former may also be internally threaded to accept a small core or "slug" of ferrite, which can be adjusted in position relative to the coil itself to vary the coil inductance.

A transformer consists of a number of coils of windings of wire wound on a common former, which is also inside a core of iron alloy, ferrite of other magnetic material. When an alternating current is passed through one of the windings (primary), it produces an alternating magnetic field in the core and this in turn induces AC voltages in the other (secondary) windings. The voltages produced in the other winding depend on the number of turns in those windings, compared with the turns in the primary winding. If a secondary winding has fewer turns than the primary, it will produce a lower voltage, and be called a step-down transformer. If the secondary winding has more windings than the primary, then the transformer will produce a higher voltage and it will be a step-up transformer. Transformers can be used to change the voltage levels of AC power and they are available in many different sizes and power handling capabilities.

Microphones

A microphone converts audible sound waves into electrical signals which can be then amplified. In an electret microphone, the sound waves vibrate a circular diaphragm made from very thin plastic material which has a permanent charge in it. Metal films coated on each side form a capacitor, which produces a very small AC voltage when the diaphragm vibrates. All electret microphones also contains FET which amplifies the very small AC signals. To power an FET amplifier, the microphone must be supplied with a small DC voltage.

Loudspeakers

A loudspeaker converts electrical signals into sound waves that we can hear. It has two terminals which go to a voice coil, attached to a circular cone made of either cardboard or thin plastic. When electrical signals are applied to the voice coil, its creates a varying magnetic field from a permanent magnet at the back of the speaker. As a result the cone vibrates in sympathy with the applied signal to produce sound waves.

Relays

Many electronic components are not capable of switching higher currents or voltages, so a device called a relay is used. A relay has a coil which forms an electromagnet, attracting a steel "armature" which itself pushes on one or more sets of switching contacts. When a current is passed through the coil to energize it, the moving contacts disconnect from one set of contacts to another, and when the coil is de-energized the contacts go back to their original position. In most cases, a relay needs a diode across the coil to prevent damage to the semiconductor driving the coil.

Switches

A switch is a device with one or more sets of switching contacts, which are used to control the flow of current in a circuit. The switch allows the contacts to be controlled by a physical actuator of some kind—such as a press-button toggle lever, rotary or knob, etc. As the name denotes, this type of switch has an actuator bar which slides back and forth between the various contact positions. In a single-pole, double throw, or "SPDT" slider switch, a moving contacte links the center contact to either of the two end contacts. In contrast, a double-pole double throw (DPDT) slider switch has two of these sets of contacts, with their moving contacts operating in tandem when the slider is actuated.

Wire

A wire is simply a length of metal conductor, usually made from copper since its conductivity is good, which means its resistance is low. When there is a risk of a wire touching another wire and causing a short, the copper wire is insulated or covered with a plastic coating which acts as an insulating material. Plain copper wire is not usually used since it will quickly oxidize or tarnish in the presence of air. A thin metal alloy coating is often applied to the copper wire; usually an alloy of tin or lead is used.

Single or multi-strand wire is covered in colored PVC plastic insulation and is used quite often in electronic applications to connect circuits or components together. This wire is often called “hooku” wire. On a circuit diagram, a solid dot indicates that the wires or PC board tracks are connected together or joined, while a “loop-over” indicates that they are not joined and must be insulated. A number of insulated wires enclosed in an outer jacket is called an electrical cable. Some electrical cables can have many insulated wires in them.

Semiconductor substitution

There are often times when building an electronic circuit, it is difficult or impossible to find or locate the original transistor or integrated circuit. There are a number of circuits shown in this book which feature transistors, SCRs, UJTs, and FETs that are specified but cannot easily be found. Where possible many of these foreign components are converted to substitute values, either with a direct replacement or close substitution. Many foreign parts can be easily converted directly to a commonly used transistor or component. Occasionally an outdated component has no direct common replacement, so the closest specifications of that component are attempted. In some instances we have specified replacement components with substitution components from the NTE brand or replacements. Most of the components for the projects used in this book are quite common and easily located or substituted without difficulty.

When substituting components in the circuit, make sure that the pin-outs match the original components. Sometimes, for example, a transistor may have bottom view drawing, while the substituted value may have a drawing with a top view. Also be sure to check the pin-outs or the original components versus the replacement. As an example, some transistors will have EBC versus ECB pin-outs, so be sure to look closely at possible differences which may occur.

Reading electronic schematics

The heart of all radio communication devices, both transmitters and receivers, all revolve around some type of oscillator. In this section we will take a look at what is perhaps the most important part of any receiver, and that is the oscillator. Communication transmitters, receivers, frequency standards and synthesizers all use some type of oscillators circuits. Transmitters need oscillators for their exciters, while receivers most often use local oscillators to mix signals. In this section you will see how specific electronics components are utilized to form oscillator circuits. Let’s examine a few of the more common types of oscillator designs and their building considerations. In this section, you will also tell the difference between a schematic diagram and a pictorial diagram. A schematic diagram illustrates the electronic symbols and how the components connect to one another, it is the circuit blue-print and pretty much universal among electronic enthusiasts. A pictorial diagram, on the other hand, is a “picture” of how the components might appear in an actual circuit on a circuit board of one type or another. Take a close look at the difference between the two types of diagram, and they will help you later when building actual circuits. Our first type of oscillator shown below is The Hartley.

Hartley oscillator

The Hartley RF oscillator, illustrated in Figure 2-4, is centered around the commonly available 2N4416A FET transistor. This general purpose VFO oscillator operates around 5100 kHz. The frequency determining components are L1, C1, a 10 pF trimmer and capacitors C2, C3, C4, C5 and C6. Note capacitor C6 is a 10-100 pF variable trimmer type. Capacitor C7 is to reduce the loading on the tuned circuit components. Its value can be small but be able to provide sufficient drive to the succeeding buffer amplifier stage. You can experiment using a small viable capacitor trimmer, such as a 5-25 pF.

The other components, such as the two resistors, silicon diode at D1 are standard types, nothing particularly special. The Zener diode at D2 is a 6.2 volt type. Capacitor C8 can be selected to give higher/lower output to the buffer amplifier. Smaller C6 values give lower output and conversely higher values give larger output.

In order to get the circuit to work properly, you need to have an inductive reactance for L1 of around about 180 ohms. At 5 MHz this works out at about 5.7 uH. The important consideration, is that the feedback point from the source of the JFET connects to about 25% of the windings of L1 from the ground end. An air cored inductor is shown in the diagram. It could be, for example, 18-19 turns of #20 gauge wire on a 25.4 mm (1″) diameter form spread evenly over a length of about 25.4 mm (1″). The tap would be at about 4½ turns. Alternatively, with degraded performance, you could use a T50-6 toroid and wind say 37 turns of #24 wire (5.48 µH) tapping at 9 turns.

So to have the oscillator operate at around 5 MHz, we know the LC is 1013 and if L is say 5.7 µH then total C for resonance (just like LC Filters eh!) is about 177 pF. We want to be able to tune from 5000 to 5100 kHz, a tuning ratio of 1.02, which means a capacitance ratio of 1.04 (min to max).

Colpitts oscillators

Colpitts oscillators are similar to the shunt fed Hartley oscillator circuit except the Colpitts oscillator, instead of having a tapped inductor, utilizes two series capacitors in its LC circuit. With the Colpitts oscillator the connection between these two capacitors is used as the center tap for the circuit. A Colpitts oscillator circuit is shown at Figure 2-5, and you will see some similarities with the Hartley oscillator.

The simplest Colpitts oscillator to construct and get running is the “series tuned” version, more often referred to as the “Clapp Oscillator.” Because there is no load on the inductor, a high “Q” circuit results with a high L/C ratio and of course much less circulating current. This aids drift reduction. Because larger inductances are required, stray inductances do not have as much impact as perhaps in other circuits.

The total capacitive reactance of the parallel combination of capacitors depicted as series tuning below the inductor in a series tuned Colpitts oscillator or “Clapp oscillator” should have a total reactance of

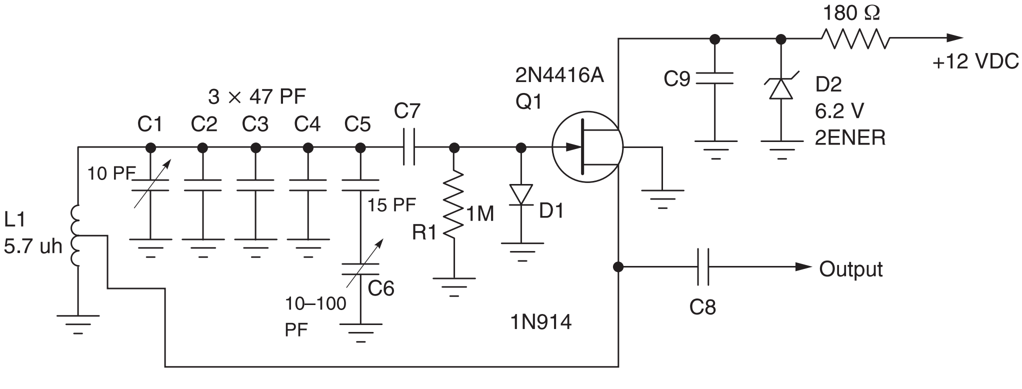

Figure 2-4 *Hartley oscillator circuit*

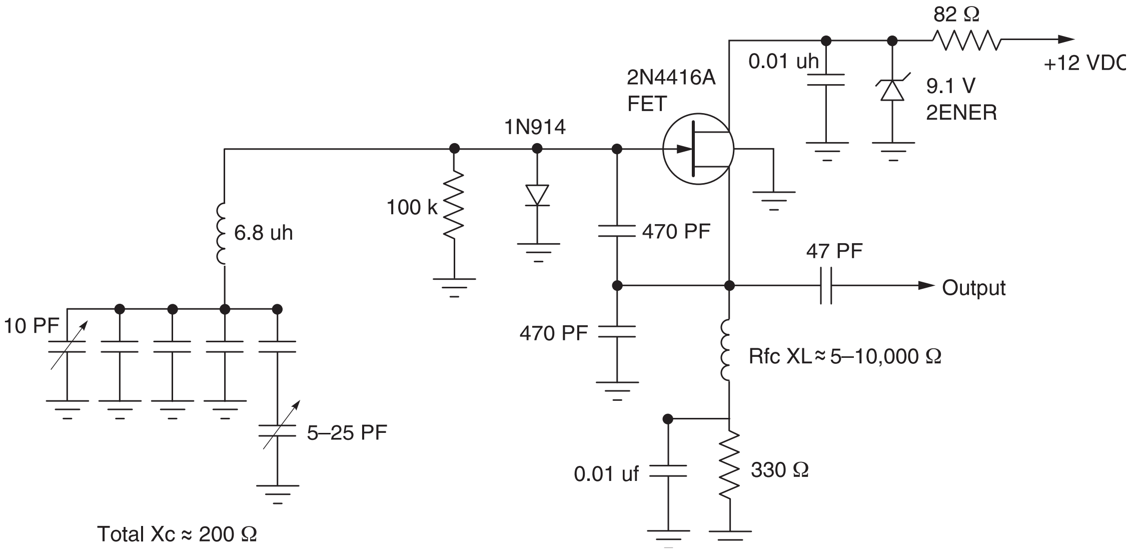

Figure 2-5 *Series tuned Colpitts oscillator circuit*

around 200 ohms. Not all capacitors may be required in your particular application. Effectively all the capacitors are in series in a Colpitts oscillator, i.e. they appear as parallel connected but their actual values are in fact in series.

Ideally, your frequency determining components L1 and the parallel capacitors should be in a grounded metal shield. The FET used in the Colpitts oscillator is the readily available 2N4416A. Note, the metal FET case is connected to the circuit ground. The output from the Colpitts oscillator is through output capacitor 47 pF; this should be the smallest of values possible, consistent with continued reliable operation into the next buffer amplifier stage.

Crystal oscillators

Crystal oscillators are oscillators where the primary frequency determining element is a quartz crystal. Because of the inherent characteristics of the quartz crystal the crystal oscillator may be held to extreme accuracy of frequency stability.

Crystal oscillators are, usually, fixed frequency oscillators where stability and accuracy are the primary considerations. For example, it is almost impossible to design a stable and accurate LC oscillator for the upper HF and higher frequencies without resorting to some sort of crystal control. Hence the reason for crystal oscillators.

The crystal oscillator depicted at Figure 2-6 is a typical example of an RF or radio frequency crystal oscillators which may be used for exciters or RF converters. Transistor Q1, can be any transistor which will operate up to 150 MHz, such as a 2N2222A.

The turns ratio on the tuned circuit depicts an anticipated nominal load of 50 ohms. This allows a theoretical 2.5k ohms on the collector, so typically a 7:1 turns ratio for T1 would be used. Use the: L * C = 25330 / $F0^2$ formula for determining L and C in the tuned circuits of crystal oscillator. Personally I would make L a reactance of around 250 ohms. In this case, I'd make C1 a smaller trimmer capacitor. The crystal at X1 could be an overtone type crystal for the crystal, selecting a L * C for the odd particular multiple of overtone wanted in your crystal oscillators. Typically the output of the crystal oscillator would be followed by a buffer circuit. A pictorial diagram of the crystal controlled oscillator circuit is shown in Figure 2-7. Note that the components in this diagram are illustrated as component blocks as they might actually look placed on the circuit, rather than as electronic symbols in a circuit diagram or schematic.

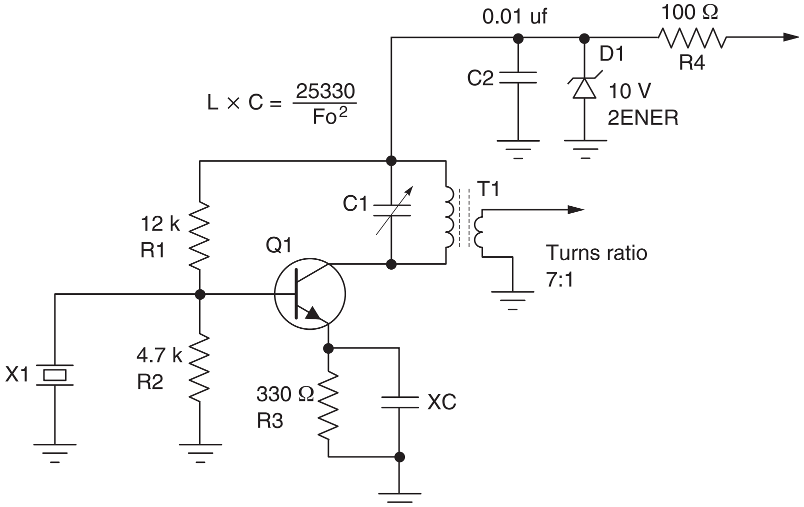
0.01 uf
100 Ω
D1
R4
C2
10 V
2ENER
L × C = 25330/Fo²
T1
C1
12 k
R1
Q1
Turns ratio
7:1
X1
4.7 k
R2
330 Ω
R3
XC

Figure 2-6 *Crystal oscillator circuit*

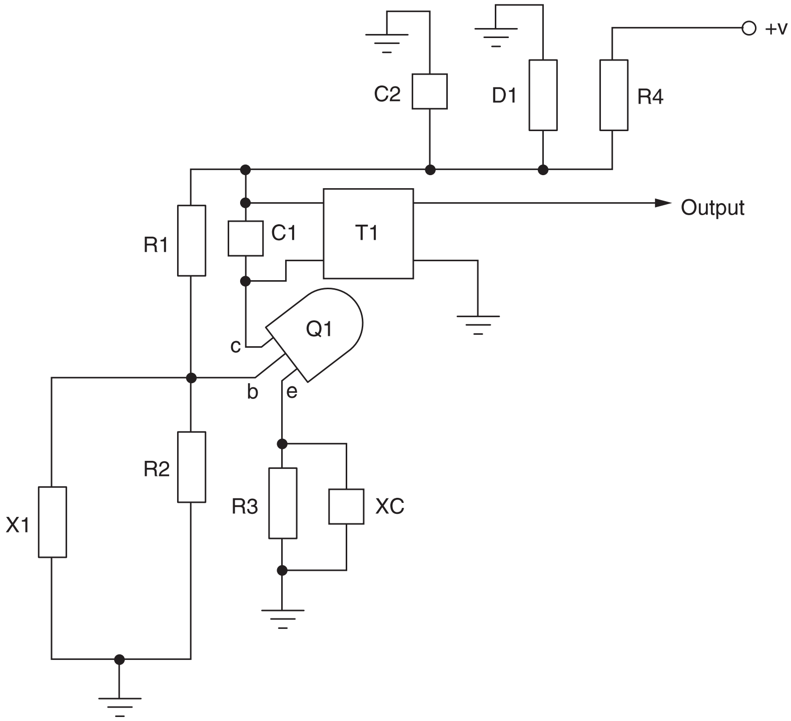
+v
C2
D1
R4
Output
R1
C1
T1
Q1
c
b
e
R2
R3
XC
X1

Figure 2-7 *Crystal oscillator pictorial diagram*

Ceramic resonator VFO oscillator

The ceramic resonator VFO, or variable frequency oscillator circuit shown in Figure 2-8, illustrates a 7MH oscillator with a variable crystal oscillator (VXO). The VXO oscillator is extremely stable, but allows only a small variation in frequency, as compared with a conventional VFO. In contrast, a VFO with an LC resonant circuit can be tuned over a range of several hundred kHz, but its frequency stability will depend upon its construction, and is never as good as a crystal oscillator. The use of a ceramic resonator as a frequency determining component fulfills both requirements. The VXO oscillator is very stable yet it can vary the frequency, so the oscillator can be tuned. The range of a VXO oscillator circuit is not as wide as an LC oscillator but it offers a tuning range of 35 kHz with good frequency stability. The somewhat unusual resonant LC circuit at the collector of Q1 has two functions. It improves the shape of the output signal and also compensates for the amplitude drop starting at approx. 7020 kHz. Transistor Q1 is a readily obtainable 2N3904 and the ceramic resonator is a Murata SFE 7.02 M2C type or equivalent. Inductor L1 consists of two coils on a T50-2 powered iron toroid. The primary coil is 8-turns, while the secondary coil is 2-turns.

The diagram shown in Figure 2-9 depicts the ceramic resonator as a pictorial diagram, where the components look as a components block that might be placed in a circuit rather than an actual schematic diagram.

Voltage controlled oscillator (VCO)

A voltage controlled oscillator, or VCO, is an oscillator where the main variable tuning capacitor is a varactor diode. The voltage controlled oscillator is tuned across the “band” by a well regulated Dc voltage applied to the varactor diode, which varies the net capacitance applied to the tuned circuit. The voltage controlled oscillator, shown in Figure 2-10, illustrates a VCO which operates in the amateur radio band between 1.8 and 2.0 MHz.

Buying quality variable capacitors today is often quite expensive, so VCOs are an extremely attractive alternative. As an alternative, all you need is an extremely stable and clean source of Dc power, a varactor diode and a high quality potentiometer—usually a 10 turn type. Note that circuit “Q” tends to be somewhat degraded by using varactor diodes instead of variable capacitors.

When a reverse voltage is applied to a diode, it exhibits the characteristics of a capacitor. Altering the voltage alters the capacitance. Common diodes such as

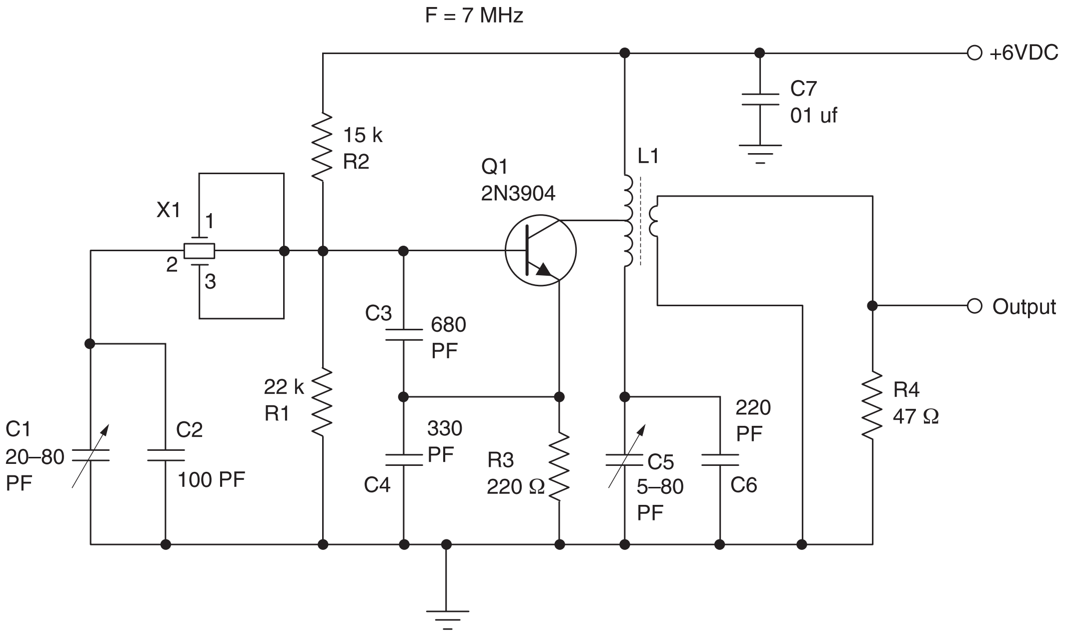

Figure 2-8 *Ceramic resonator VFO circuit*

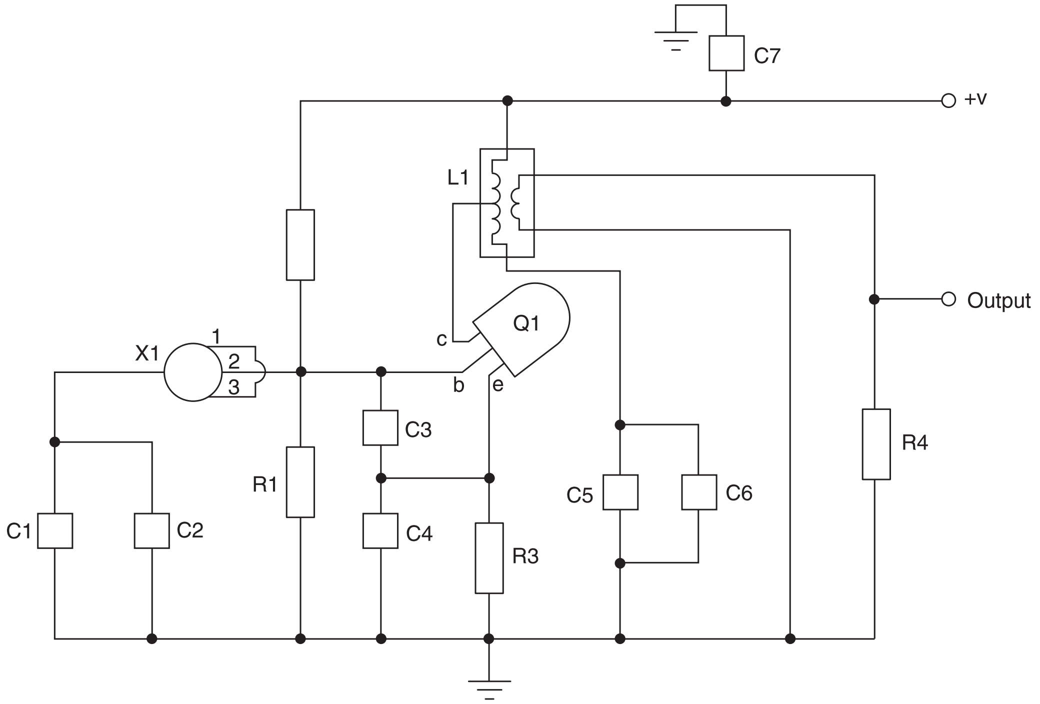
C7
+V
L1
Output
Q1
c
1
X1
2
3
b
e
C3
R4
R1
C5
C6
C1
C2
C4
R3

Figure 2-9 *Ceramic resonator VFO pictorial circuit diagram*

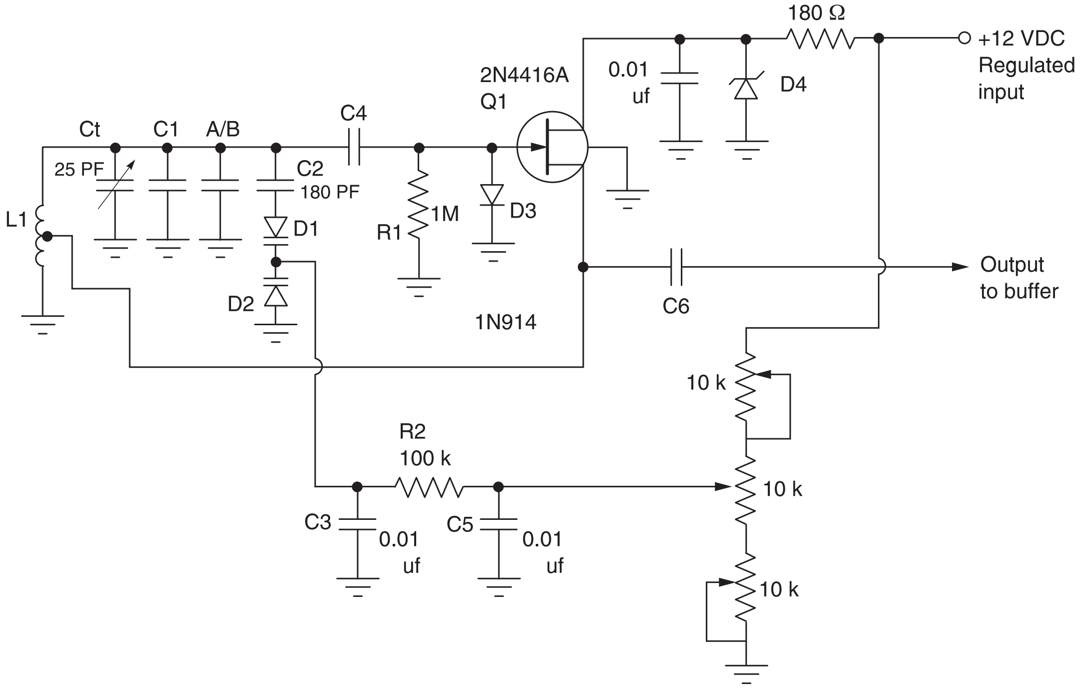
180 Ω
+12 VDC
Regulated
input
0.01
uf
D4
2N4416A
Q1
C4
Ct
C1
A/B
25 PF
C2
180 PF
L1
D1
1M
D3
R1
Output
to buffer
D2
C6
1N914
10 k
R2
100 k
10 k
C3
C5
0.01
uf
0.01
uf
10 k

Figure 2-10 *Voltage controlled oscillator (VCO) circuit*

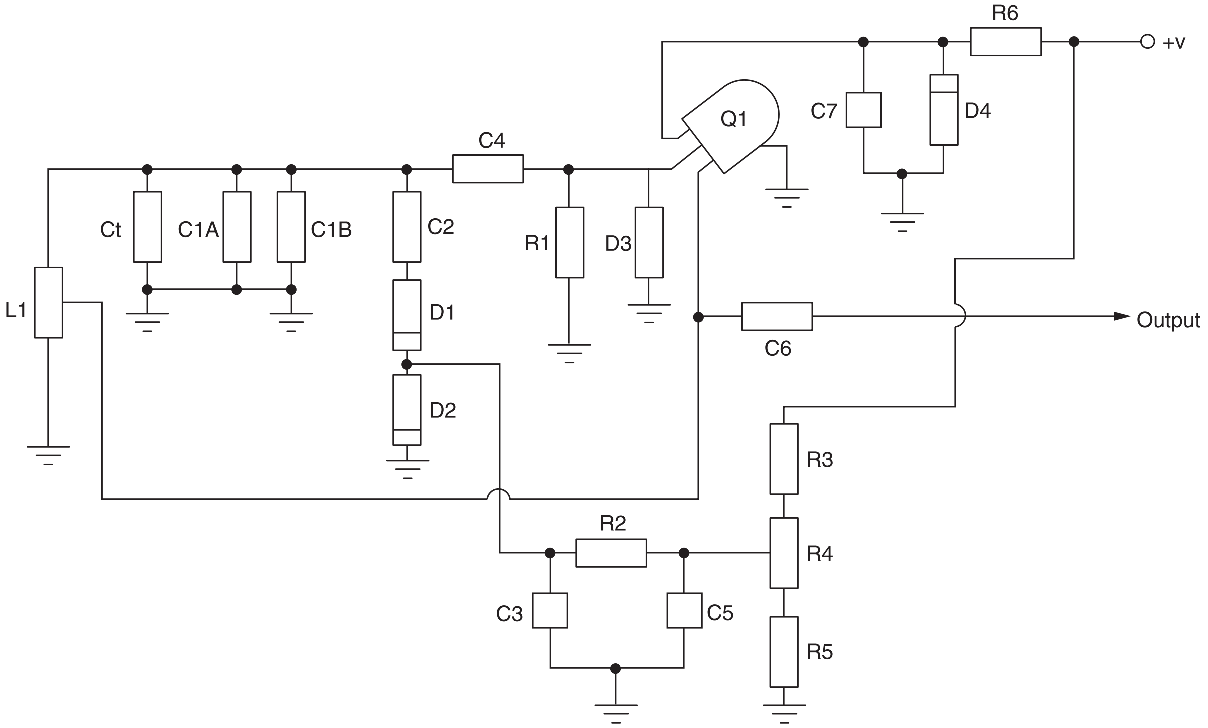

Figure 2-11 *Voltage controlled oscillator (VCO) pictorial diagram*

1N914 and 1N4004 can be used, but more commonly a special diode known as a varactor diode is specifically manufactured for VCO use, such as Motorola's MVAM115.

The variable capacitor in this circuit is replaced by a varactor diode as a tuning diode, in series with capacitor C2. Note, the two, diodes back-to-back in series. Although this in effect divides total varactor diode capacitance by two, it eliminates RF energy present in the tank circuit driving a single diode into conduction on peaks which will increase the bias voltage; this also gives rise to harmonics.

Inductor L is wound on a toroid, an Amidon T50-2 type which would require about 55 turns of #26 wire, or even the T68-2 type requiring about 51 turns of #24 wire. Both gauges mentioned are those which will conveniently fit around the core. The inductor should have a value of about 15 µH. Capacitor C2 should have around180 pF, and the total capacitance of C1 a/b should be 390 pF. C4 is used to reduce the loading of the tuned circuit components. Its value can be small but be able to provide sufficient drive to the succeeding buffer amplifier stage. You can experiment using a small viable capacitor trimmer, such as a 5-25 pF.

The 10K potentiometer is a 10 turn "quality" potentiometer for tuning; the upper and lower trim pots (set and forget) allow you to adjust the voltage range of your choice that your tuning potentiometer will see. Use "quality" trimpots for these potentiometers, since they will affect the stability of the tuning of the circuit. The 100K resistor and the two 0.1 µF capacitors are further filtering. The diagram in Figure 2-11 depicts a pictorial diagram of the voltage controlled oscillator circuit. The circuit is shown as the components might appear rather than as symbols in a schematic diagram. The pictorial diagram often reflects an actual layout or parts placement and how the parts may appear on a circuit board.

Now that you can identify electronics components and you understand how they function in an electronics circuit, as well how electronics circuits are drawn, we can move on to the how-to-soldering section. In the next chapter, you will learn how to install components and how to solder the components onto a circuit board.

Chapter 3

Electronic parts installation and soldering

Before we discuss mounting the components to the printed circuit board we should take a few minutes to discuss the resistor and capacitor codes which are used to help identify and install these components. Table 3-1 lists the resistor color code information. Resistors generally have three or four color bands which help identify the resistor values. The color band will start from one edge of the resistor body; this is the first color code which represents the first digit, and the second color band depicts the second digit value, while the third color band is the resistor's multiplier value. The fourth color band represents the resistor tolerance value. If there is no fourth band then the resistor has a 20% tolerance; if the fourth band is silver, then the resistor has a 10% tolerance value, and if the fourth band is gold then the resistor has a 5% tolerance value. Therefore, if you have a resistor with the first band having a brown color, the second band with a black color and a third band with an orange color and a fourth with gold, then one (1) times zero (0) multiplied by a value of (1000), so the resistor will have a 10,000 ohm or 10k ohm value with a 5% tolerance.

Now let's move to identifying various capacitors. When identifying and installing the ceramic, mylar or poly capacitors, you will need to refer to Table 3-2, which lists capacitor (code) values vs. actual capacitor values. Most large capacitors, such as electrolytic types, will have their actual value printed on the body of the capacitor. But often with small capacitors they are just too small to have the actual capacitor values printed on them. An abbreviated code was devised to mark the capacitor, as shown in the table. A code marking of (104) would therefore denote a capacitor with a value of .1 µF (microfarads) or 100 nF (nanofarads). Once you understand how these component codes work, you can easily identify the resistors and capacitor and proceed to mount them on the printed circuit board.

Before we begin building actual electronic radio receiver circuits, you will need to understand how to prepare the electronic components before installing them onto the circuit board. First, you will need to know how to "dress" or prepare resistor leads, which is shown in Figure 3-1. Resistor leads are first bent to fit the PC board component holes, then the resistor is mounted to the circuit board. Figure 3-2 depicts how capacitor leads are bent to fit the PC board holes. Once, the capacitor leads are prepared, the capacitor can be installed onto the circuit board. The diagram shown in Figure 3-3 illustrates how to prepare diode leads before mounting them to the PB board. Finally, Figure 3-4 depicts transistor lead preparation, while Figure 3-5 shows how integrated circuit leads are prepared for installation.

Learning how to solder

Everyone working in electronics needs to know how to solder well. Before you begin working on a circuit, carefully read this chapter on soldering. In this section you will learn how to make good solder joints when soldering point-to-point wiring connections as well as for PC board soldering connections.

In all electronics work, the wiring connections must be absolutely secure. A loose connection in a radio results in noise, scratching sounds, or no sound at all. In a TV, poor connections can disrupt the sound or picture. The safe operation of airplanes and the lives of astronauts in flight depend on secure electronics connections.

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

Table 3-1

Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

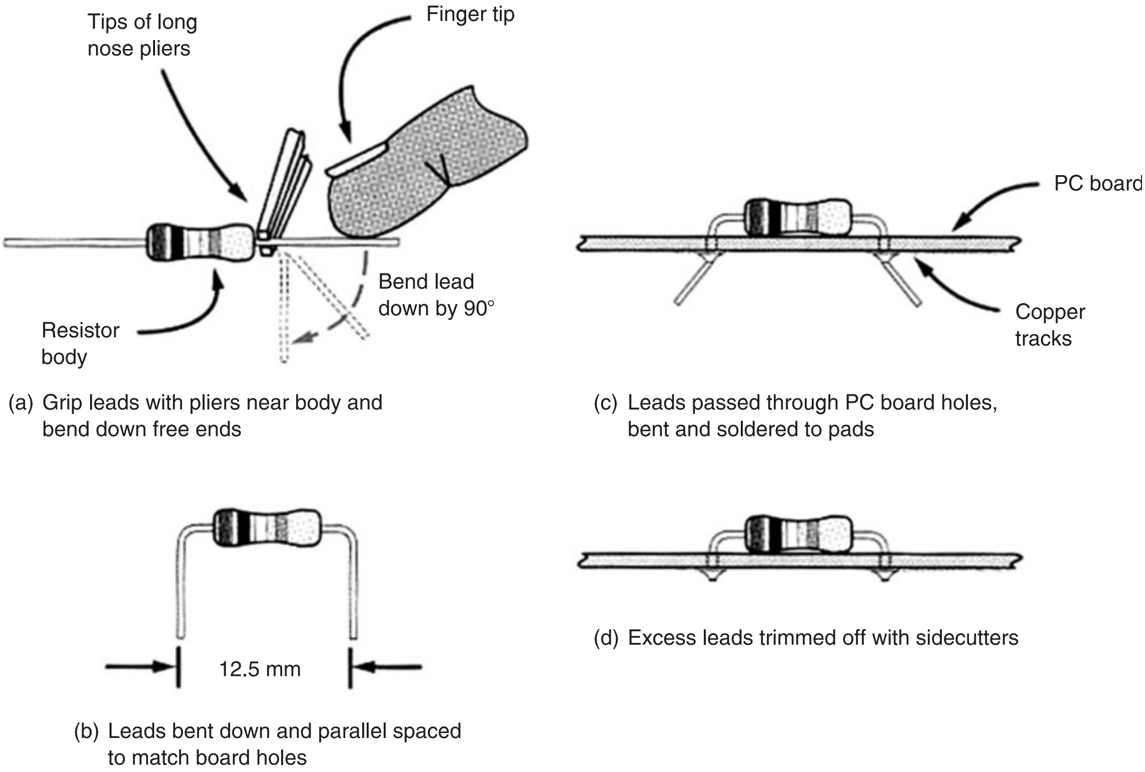
Tips of long nose pliers
Finger tip
Bend lead down by 90°
Resistor body
PC board
Copper tracks
12.5 mm

(a) Grip leads with pliers near body and bend down free ends

(c) Leads passed through PC board holes, bent and soldered to pads

(d) Excess leads trimmed off with sidecutters

(b) Leads bent down and parallel spaced to match board holes

Figure 3-1 *Preparing resistor leads for horizontal mounting*

Table 3-2

Capacitor code identification information

This table is designed to provide the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Most capacitors actually have the numeric values stamped on them; however, some are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

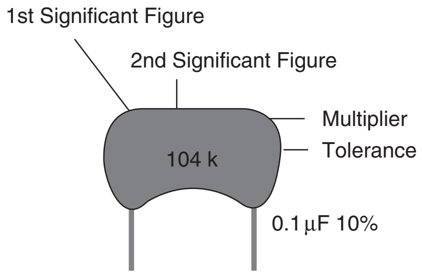

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

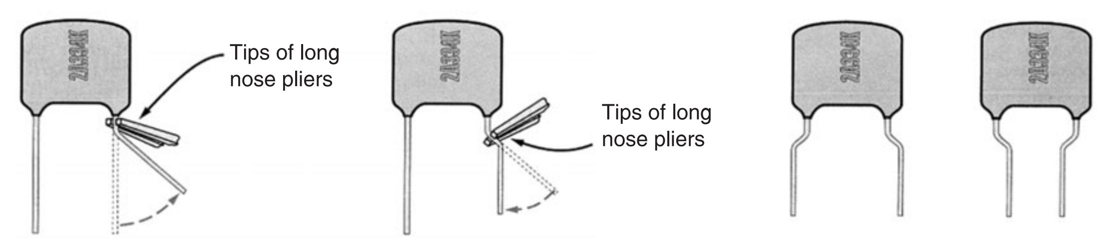
Tips of long nose pliers
Tips of long nose pliers

(a) Grip leads with pliers near body and bend out or in by 45°

(b) Grip now about 2 mm down and bend down by 45°, parallel again

(c) Both leads cranked in (or out) if necessary to match PC board hole spacing

Figure 3-2 *"Dressing" capacitor component leads*

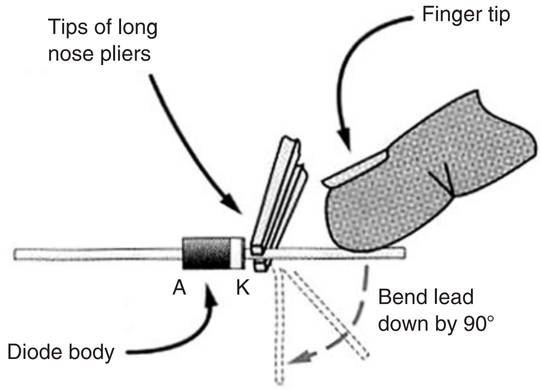
Tips of long nose pliers
Finger tip
A
K
Bend lead down by 90°
Diode body

(a) Grip leads with pliers near body and bend down free ends

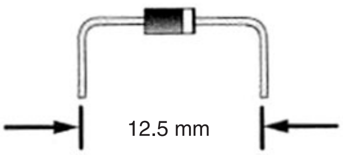

(b) Leads bent down and parallel, spaced to match board holes

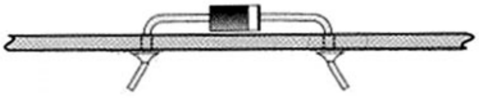

(c) Leads passed through PC board holes, bent and soldered to pads

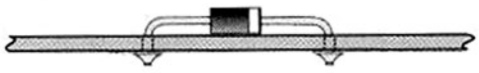

(d) Excess leads trimmed off with sidecutters

Figure 3-3 *Preparing diode leads for horizontal mounting*

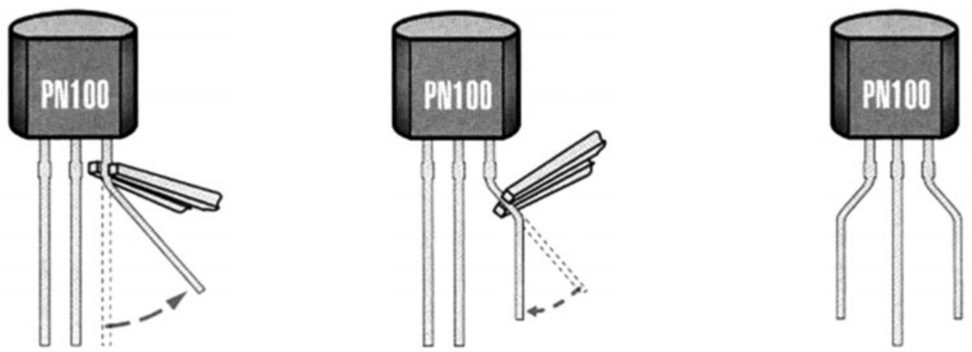
PN100
PN100
PN100

(a) Grip outer lead near body and bend out by 45°

(b) Grip about 2 mm down and bend down by 45°, parallel again

(c) Both leads cranked to match 2.5 mm PC board hole spacing

Figure 3-4 *"Dressing" TO-92 transistor leads*

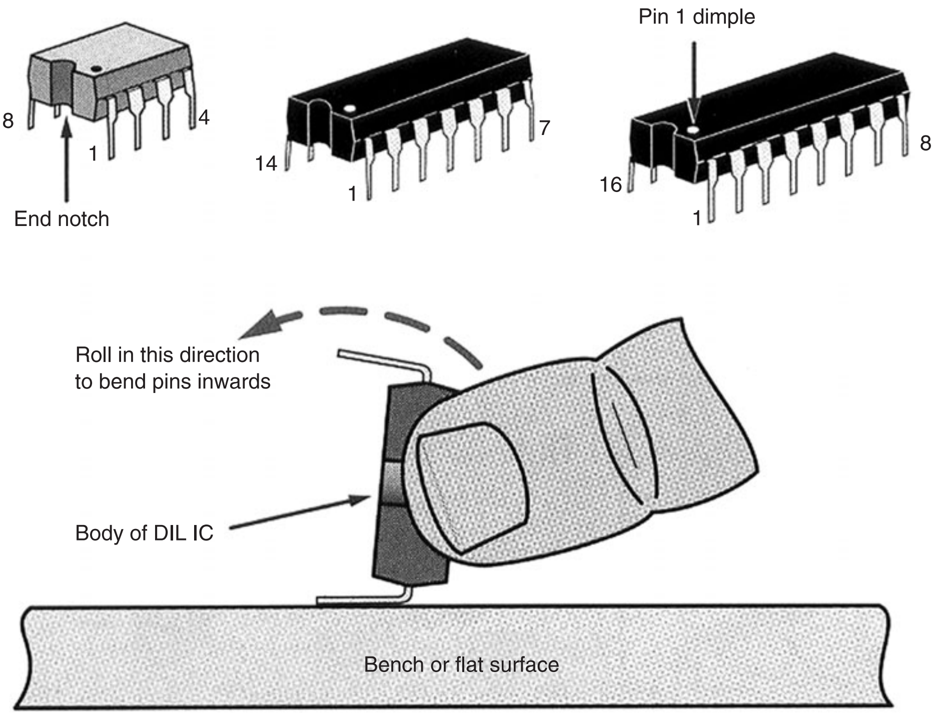

Figure 3-5 *DIL IC packages and preparing them for insertion*

Soldering joins two pieces of metal, such as electrical wires, by melting them together with another metal to form a strong, chemical bond. Done correctly, it unites the metals so that electrically they act as one piece of metal. Soldering is not just gluing metals together. Soldering is tricky and intimidating in practice, but easy to understand in theory. Basic supplies include a soldering iron, which is a prong of metal that heats to a specific temperature through electricity, like a regular iron. The solder, or soldering wire, often an alloy of aluminum and lead, needs a lower melting point than the metal you're joining. Finally, you need a cleaning resin called flux that ensures the joining pieces are incredibly clean. Flux removes all the oxides on the surface of the metal that would interfere with the molecular bonding, allowing the solder to flow into the joint smoothly. You also need two things to solder together.

The first step in soldering is cleaning the surfaces, initially with sandpaper or steel wool and then by melting flux onto the parts. Sometimes, flux is part of the alloy of the soldering wire, in an easy-to-use mixture. Then, the pieces are both heated above the melting point of the solder (but below their own melting point) with the soldering iron. When touched to the joint, this precise heating causes the solder to "flow" to the place of highest temperature and makes a chemical bond. The solder shouldn't drip or blob, but spread smoothly, coating the entire joint. When the solder cools, you should have a clean, sturdy connection.

Many people use soldering in their field, from electrical engineering and plumbing to jewelry and crafts. In a delicate procedure, a special material, called solder, flows over two pre-heated pieces and attaches them through a process similar to welding or brazing. Various metals can be soldered together, such as gold and sterling silver in jewelry, brass in watches and clocks, copper in water pipes, or iron in leaded glass stained windows. All these metals have different melting points, and therefore use different solder. Some "soft" solder, with a low melting point, is perfect for wiring a circuit board. Other "hard" solder, such as for making a bracelet, needs a torch rather than a soldering iron to get a hot enough temperature. Electrical engineers and hobbyists alike can benefit from learning the art and science of soldering.

Solder

The best solder for electronics work is 60/40 rosin-core solder. It is made of 60% tin and 40% lead. This mixture melts at a lower temperature than either lead or tin alone. It makes soldering easy and provides good connections. The rosin keeps the joint clean as it is being soldered. The heat of the iron often causes a tarnish or oxide to form on the surface. The rosin dissolves the tarnish to make the solder cling tightly. Solders have different melting points, depending on the ratio of tin to lead. Tin melts at 450°F and lead at 621°F. Solder made from 63% tin and 37% lead melts at 361°F, the lowest melting point for a tin and lead mixture. Called 63-37 (or eutectic), this type of solder also provides the most rapid solid-to-liquid transition and the best stress resistance. Solders made with different lead/tin ratios have a plastic state at some temperatures. If the solder is deformed while it is in the plastic state, the deformation remains when the solder freezes into the solid state. Any stress or motion applied to "plastic solder" causes a poor solder joint.

The 60-40 solder has the best wetting qualities. Wetting is the ability to spread rapidly and bond materials uniformly; 60-40 solder also has a low melting point. These factors make it the most commonly used solder in electronics.

Some connections that carry high current can't be made with ordinary tin-lead solder because the heat generated by the current would melt the solder. Automotive starter brushes and transmitter tank circuits are two examples. Silver-bearing solders have higher melting points, and so prevent this problem. High-temperature silver alloys become liquid in the 1100°F to l200°F range, and a silver-manganese (85-15) alloy requires almost 1800°F.

Because silver dissolves easily in tin, tin bearing solders can leach silver plating from components. This problem can greatly reduced by partially saturating the tin in the solder with silver or by eliminating the tin. Tin-silver or tin-lead-silver alloys become liquid at temperatures from 430°F for 96.5-3.5 (tin-silver), to 588°F for 1.0-97.5-1.5 (tin-lead-silver). A 15.080.0-5.0 alloy of lead-indium-silver melts at 314°F.

Never use acid-core solder for electrical work. It should be used only for plumbing or chassis work. For circuit construction, only use fluxes or solder-flux combinations that are labeled for electronic soldering.

The rosin or the acid is a flux. Flux removes oxide by suspending it in solution and floating it to the top. Flux is not a cleaning agent! Always clean the work before soldering. Flux is not a part of a soldered connection—it merely aids the soldering process. After soldering, remove any remaining flux. Rosin flux can be removed with isopropyl or denatured alcohol. A cotton swab is a good tool for applying the alcohol and scrubbing the excess flux away. Commercial flux-removal sprays are available at most electronic-part distributors.

The soldering iron

Soldering is used in nearly every phase of electronic construction so you'll need soldering tools. A soldering tool must be hot enough to do the job and lightweight enough for agility and comfort. A temperature controlled iron works well, although the cost is not justified for occasional projects. Get an iron with a small conical or chisel tip. Soldering is not like gluing; solder does more than bind metal together and provide an electrically conductive path between them. Soldered metals and the solder combine to form an alloy.

You may need an assortment of soldering irons to do a wide variety of soldering tasks. They range in size from a small 25-watt iron for delicate printed-circuit work to larger 100 to 300-watt sizes used to solder large surfaces. If you could only afford a single soldering tool when initially setting up your electronics workbench than, an inexpensive to moderately priced pencil-type soldering iron with between 25 and 40-watt capacity is the best for PC board electronics work. A 100-watt soldering gun is overkill for printed-circuit work, since it often gets too hot, cooking solder into a brittle mess or damaging small parts of a circuit. Soldering guns are best used for point-to-point soldering jobs, for large mass soldering joints or large components. Small "pencil" butane torches are also available, with optional soldering-iron tips. A small butane torch is available from the Solder-It Company. Butane soldering irons are ideal for field service problems and will allow you to solder where there is no 110 volt power source. This company also sells a soldering kit that contains paste solders (in syringes) for electronics, pot metal and plumbing. See the Appendix for the address information.

Keep soldering tools in good condition by keeping the tips well tinned with solder. Do not run them at full temperature for long periods when not in use. After each period of use, remove the tip and clean off any scale that may have accumulated. Clean an oxidized tip by dipping the hot tip in sal ammoniac (ammonium chloride) and then wiping it clean with a rag. Sal ammoniac is somewhat corrosive, so if you don't wipe the tip thoroughly, it can contaminate electronic soldering. You can also purchase a small jar of "Tip Tinner," a soldering iron tip dresser, from your local Radio Shack store. Place the tip of the soldering iron into the "Tip Tinner" after every few solder joints.

If a copper tip becomes pitted, file it smooth and bright and then tin it immediately with solder. Modern soldering iron tips are nickel or iron clad and should not be filed. The secret of good soldering is to use the right amount of heat. Many people who will have not soldered before use too little heat dabbing at the joint to be soldered and making little solder blobs that cause unintended short circuits. Always use caution when soldering. A hot soldering iron can burn your hand badly or ruin a tabletop. It's a good idea to buy or make a soldering iron holder.

Soldering station

Often when building or repairing a circuit, your soldering iron is kept switched "on" for unnecessarily long periods, consuming energy and allowing the soldering iron tip to burn and develop a buildup of oxide. Using this soldering-iron temperature controller, you will avoid destroying sensitive components when soldering.

Buying a lower wattage iron may solve some of the problems, but new problems arise when you want to solder some heavy-duty component, setting the stage for creating a "cold" connections. If you've ever tried to troubleshoot some instrument in which a cold solder joint was at the root of the problem, you know how difficult such defects are to locate. Therefore, the best way to satisfy all your needs is to buy a temperature controller electronics workbench.

A soldering station usually consists of a temperature controlled soldering iron with an adjustable heat or temperature control and a soldering iron holder and cleaning pad. If you are serious about your electronics hobby or if you have been involved with electronics building and repair for any length of time, you will eventually want to invest in a soldering station at some point in time. There are real low cost soldering station for hobbyists, for under $30, but it makes more sense to purchase a moderately price soldering station such as the quality Weller series. A typical soldering station is shown in Figure 3-6.

Soldering gun

An electronics workbench would not be complete without a soldering gun. Soldering guns are useful for soldering large components to terminal strips, or splicing wires together or when putting connectors on coax cable. There are many instances where more heat is needed than a soldering iron can supply. For example, a large connector mass cannot be heated with a small soldering iron, so you would never be able to "tin" a connector with a small wattage soldering iron. A soldering gun is a heavy duty soldering device which does in fact look like a gun. Numerous tips are available for a soldering gun and they are easily replaceable using two small nuts on the side arm of the soldering gun. Soldering guns are available in two main heat ranges.

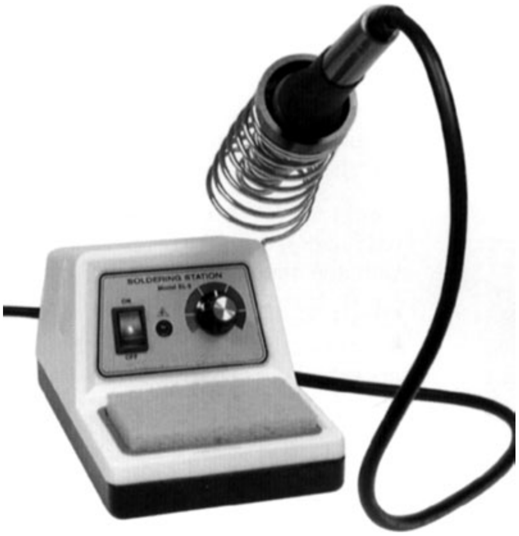

Figure 3-6 *Temperature controlled soldering station*

Most soldering guns have a two-step "trigger" switch which enable you to select two heat ranges for different soldering jobs. The most common soldering gun provides both a 100 watt setting when the "trigger" switch is pressed to its first setting, and as the "trigger" switch is advanced to the next step, the soldering gun will provide 150 watts, when more heat is needed. A larger or heavy-duty soldering gun is also available, but a little harder to locate is the 200 to 250 watt solder gun. The first "trigger" switch position provides 200 watts, while the second switch position provides the 250 watt heat setting. When splicing wires together either using the "Western Union" or parallel splice or the end splice, a soldering gun should be used especially if the wire gauge is below size 22 ga. Otherwise the solder may not melt properly and the connections may reflect a "cold" solder joint and therefore a poor or noisy splice. Soldering wires to binding post connections should be performed with a soldering gun to ensure proper heating to the connection. Most larger connectors should be soldered or pre-tinned using a soldering gun for even solder flow.

Preparing the soldering iron

If your iron is new, read the instructions about preparing it for use. If there are no instructions, use the following procedure. It should be hot enough to melt solder applied to its tip quickly (half a second when dry, instantly when wet with solder). Apply a little solder directly to the tip so that the surface is shiny. This process is called "tinning" the tool. The solder coating helps conduct heat from the tip to the joint face, the tip is in contact with one side of the joint. If you can place the tip on the underside of the joint, do so. With the tool below the joint, convection helps transfer heat to the joint. Place the solder against the joint directly opposite the soldering tool. It should melt within a second for normal PC connections, within two seconds for most other connections. If it takes longer to melt, there is not enough heat for the job at hand. Keep the tool against the joint until the solder flows freely throughout the joint. When it flows freely, solder tends to form concave shapes between the conductors. With insufficient heat solder does not flow freely; it forms convex shapes-blobs. Once solder shape changes from convex to concave, remove the tool from the joint. Let the joint cool without movement at room temperature. It usually takes no more than a few seconds. If the joint is moved before it is cool, it may take on a dull, satin look that is characteristic of a "cold" solder joint. Reheat cold joints until the solder flows freely and hold them still until cool. When the iron is set aside, or if it loses its shiny appearance, wipe away any dirt with a wet cloth or sponge. If it remains dull after cleaning, tin it again.

Overheating a transistor or diode while soldering can cause permanent damage. Use a small heatsink when you solder transistors, diodes or components with plastic parts that can melt. Grip the component lead with a pair of pliers up close to the unit so that the heat is conducted away. You will need to be careful, since it is easy to damage delicate component leads. A small alligator clip also makes a good heatsink to dissipate from the component.

Mechanical stress can damage components, too. Mount components so there is no appreciable mechanical strain on the leads.

Soldering to the pins of coil forms male cable plugs can be difficult. Use suitable small twist drill to clean the inside of the pin and then tin it with resin-core solder. While it is still liquid, clear the surplus solder from each pin with a whipping motion or by blowing through the pin from the inside of the form or plug. Watch out for flying hot solder, you can get severe burns. Next, file the nickel tip. Then insert the wire and solder it. After soldering, remove excess solder with a file, if necessary. When soldering to the pins of plastic coil-forms, hold the pin to be soldered with a pair of heavy pliers to form a heatsink. Do not allow the pin to overheat; it will loosen and become misaligned.

Preparing work for soldering

If you use old junk parts, be sure to completely clean all wires or surfaces before applying solder. Remove all enamel, dirt, scale, or oxidation by sanding or scraping the parts down to bare metal. Use fine sandpaper or emery paper to clean flat surfaces or wire. (Note, no amount of cleaning will allow you to solder to aluminum. When making a connection to a sheet of aluminum, you must connect the wire by a solder lug or a screw.)

When preparing wires, remove the insulation with wire strippers or a pocketknife. If using a knife, do not cut straight into the insulation; you might nick the wire and weaken it. Instead, hold the knife as if you were sharpening a pencil, taking care not to nick the wire as you remove the insulation. For enameled wire, use the back of the knife blade to scrape the wire until it is clean and bright. Next, tin the clean end of the wire. Now, hold the heated soldering-iron tip against the under surface of the wire and place the end of the rosin-core solder against the upper surface. As the solder melts, it flows on the clean end of the wire. Hold the hot tip of the soldering iron against the under surface of the tinned wire and remove the excess solder by letting it flow down on the tip. When properly tinned, the exposed surface of the wire should be covered with a thin, even coating of solder.

How to solder

The two key factors in quality soldering are time and temperature. Generally, rapid heating is desired, although most unsuccessful solder jobs fail because insufficient heat has been applied. Be careful; if heat is applied too long, the components or PC board can be damaged, the flux may be used up and surface oxidation can become a problem. The soldering-iron tips should be hot enough to readily melt the solder without burning, charring or discoloring components, PC boards or wires. Usually, a tip temperature about 100°F above the solder melting point is about right for mounting components on PC boards. Also, use solder that is sized appropriately for the job. As the cross-section of the solder decreases, so does the amount of heat required to melt it. Diameters from 0.025 to 0.040 inches are good for nearly all circuit wiring.

Always use a good quality multi-core solder. A standard 60% tin, 40% lead alloy solder with cores of non-corrosive flux will be found easiest to use. The flux contained in the longitudinal cores of multi-core solder is a chemical designed to clean the surfaces to be joined of deposited oxides, and to exclude air during the soldering process, which would otherwise prevent these metals coming together. Consequently, don't expect to be able to complete a joint by using the application of the tip of the iron loaded with molten solder alone, as this usually will not work. Having said that, there is a process called tinning where conductors are first coated in fresh, new solder prior to joining by a hot iron.

Solder comes in gauges like wire. The two most common types of solder, are 18 ga., used for general work, and the thinner 22 ga., used for fine work on printed circuit boards.

A well-soldered joint depends on

1. Soldering with a clean, well-tinned tip.
2. Cleaning the wires or parts to be soldered.
3. Making a good mechanical joint before soldering.
4. Allowing the joint to get hot enough before applying solder.
5. Allowing the solder to set before handling or moving soldered parts.

Making a good mechanical joint

Unless you are creating a temporary joint, the next step is to make a good mechanical connection between the parts to be soldered. For instance, wrap the wire carefully and tightly around a soldering terminal or soldering lug, as shown in Figure 3-7. Bend wire and make connections with long-nosed pliers. When connecting two wires together, make a tight splice before soldering. Once you have made a good mechanical contact, you are ready for the actual soldering.

The next step is to apply the soldering iron to the connection, soldering the connection as shown.

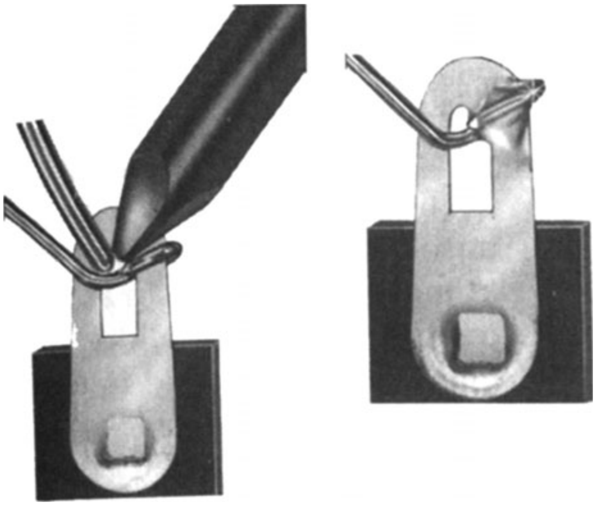

Figure 3-7 *Wire to lug soldering joint*

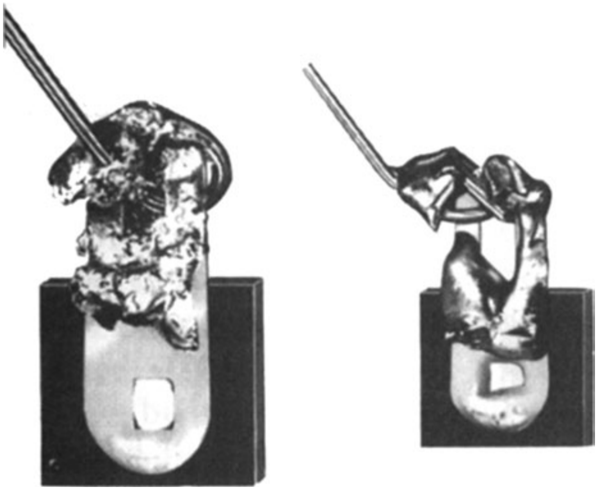

Figure 3-8 *Inferior wire to lug solder joint*

In soldering a wire splice, hold the iron below the splice and apply solder to the top of the splice. If the tip of the iron has a bit of melted solder on the side held against the splice, heat is transferred more readily to the splice and the soldering is done more easily. Don`t try to solder by applying solder to the joint and then pressing down on it with the iron. Be sure not to disturb the soldered joint until the solder has set. It may take a few seconds for the solder to set, depending upon the amount of solder used in making the joint. Now take a good look at the joint. It should have a shiny, smooth appearance—not pitted or grainy. If it does have a pitted, granular appearance as seen in Figure 3-8. reheat the joint, scrape off the solder, and clean the connection. Then start over again. After the solder is well set, pull on the wire to see if it is a good, tight connection. If you find that you made a poor soldering job don't get upset, be thankful you found it and do it over again. A quick reference solder check list is shown in the listing in Table 3-3.

Soldering printed circuit boards

Most electronic devices use one or more printed circuit (PC) boards. A PC board is a thin sheet of fiberglass or phenolic resin that has a pattern of foil conductors "printed" on it. You insert component leads into holes in the board and solder the leads to the foil pattern. This method of assembly is widely used and you will probably encounter it if you choose to build from a kit. Printed circuit boards make assembly easy. First, insert component leads through the correct holes in the circuit board. Mount parts tightly against the circuit board unless otherwise directed. After inserting a lead into the board, bend it slightly outward to hold the part in place.

When the iron is hot, apply some solder to the flattened working end at the end of the bit, and wipe it on a piece of damp cloth or sponge so that the solder forms a thin film on the bit. This is tinning the bit.

Table 3-3
Soldering check list

1. Prepare the joint. Clean all surfaces and conductors thoroughly with fine steel wool. First, clean the circuit traces, then clean the component leads.
2. Prepare the soldering iron or gun. The soldering device should be hot enough to melt solder applied to the tip. Apply a small amount of solder directly to the tip, so that the surface is shiny.
3. Place the tip in contact with one side of the joint, if possible place the tip below the joint.
4. Place the solder against the joint directly opposite the soldering tool. The solder should melt within two seconds; if it takes longer use a larger iron.
5. Keep the soldering tool against the joint until the solder flows freely throughout the joint. When it flows freely the joint should form a concave shape; insufficient heat will form a convex shape.
6. Let the joint cool without any movement, the joint should cool and set-up with in a few seconds. If the joint is moved before it cools the joint will look dull instead of shiny and you will likely have a cold solder joint. Re-heat the joint and begin anew.
7. Once the iron is set aside, or if it loses its shiny appearance, wipe away any dirt with a wet cloth or wet sponge. When the iron is clean the tip should look clean and shiny. After cleaning the tip apply some solder.

Melt a little more solder on to the tip of the soldering iron, and put the tip so it contacts both parts of the joint. It is the molten solder on the tip of the iron that allows the heat to flow quickly from the iron into both parts of the joint. If the iron has the right amount of solder on it and is positioned correctly, then the two parts to be joined will reach the solder's melting temperature in a couple of seconds. Now apply the end of the solder to the point where both parts of the joint and the soldering iron are all touching one another. The solder will melt immediately and flow around all the parts that are at, or over, the melting part temperature. After a few seconds remove the iron from the joint. Make sure that no parts of the joint move after the soldering iron is removed until the solder is completely hard. This can take quite a few seconds with large joints. If the joint is disturbed during this cooling period it may become seriously weakened.

The most important point in soldering is that both parts of the joint to be made must be at the same temperature. The solder will flow evenly and make a good electrical and mechanical joint only if both parts of the joint are at an equal high temperature. Even though it appears that there is a metal-to-metal contact in a joint to be made, very often there exists a film of oxide on the surface that insulates the two parts. For this reason it is no good applying the soldering iron tip to one half of the joint only and expecting this to heat the other half of the joint as well.

It is important to use the right amount of solder, both on the iron and on the joint. Too little solder on the iron will result in poor heat transfer to the joint, too much and you will suffer from the solder forming strings as the iron is removed, causing splashes and bridges to other contacts. Too little solder applied to the joint will give the joint a half finished appearance: a good bond where the soldering iron has been, and no solder at all on the other part of the joint.

The hard cold solder on a properly made joint should have a smooth shiny appearance and if the wire is pulled it should not pull out of the joint. In a properly made joint the solder will bond the components very strongly indeed, since the process of soldering is similar to brazing, and to a lesser degree welding, in that the solder actually forms a molecular bond with the surfaces of the joint. Remember it is much more difficult to correct a poorly made joint than it is to make the joint properly in the first place. Anyone can learn to solder, it just takes practice.

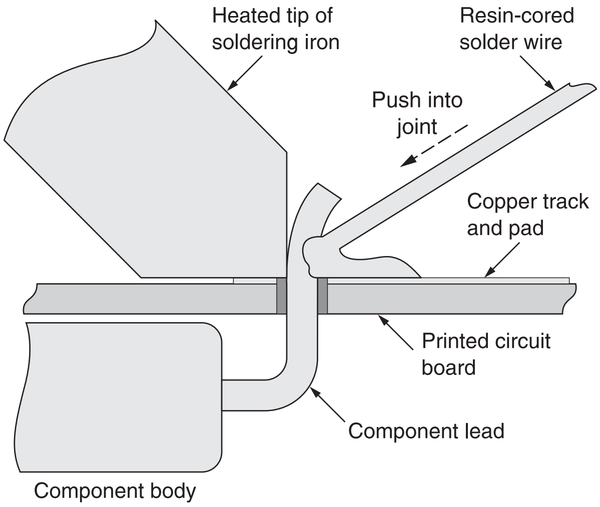

Figure 3-9 *Proper technique for soldering components on PC board*

The diagram in Figure 3-9 shows how to solder a component lead to a PC board pad. The tip of the soldering iron heats both the lead and the copper pad, so the end of the solder wire melts when it's pushed into the contact. The diagram illustrated in Figure 3-10 shows how a good solder joint is obtained. Notice that it has a smooth and shiny "fillet" of solder metal, bonding all around to both the component lead and the copper pad of the PC board. This joint provides a reliable electrical connection.

Try to make the solder joint as quickly as possible because the longer you take, the higher the risk that the component itself and the printed circuit board pad and track will overheat and be damaged. But don't work so quickly that you cannot make a good solder joint.

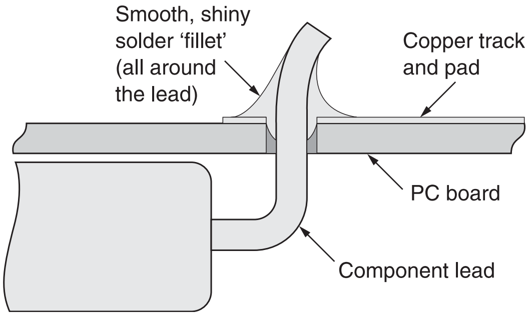

Figure 3-10 *Excellent solder joint*

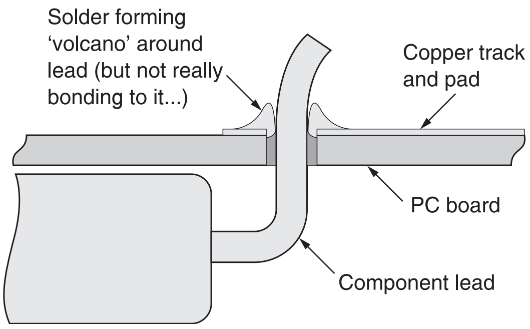
Solder forming 'volcano' around lead (but not really bonding to it...)
Copper track and pad
PC board
Component lead

Figure 3-11 *Inferior solder joint: Example 1*

Having to solder the joint over again always increases the risk of applying too much heat to the PCB.

As the solder solidifies, take a careful look at the joint you have made, to make sure there's a smooth and fairly shiny metal "fillet" around it. This should be broadly concave in shape, showing that the solder has formed a good bond to both metal surfaces. If it has a rough and dull surface or just forms a "ball" on the component lead, or a "volcano" on the PCB pad with the lead emerging from the crater, you have a "dry joint." If your solder joint looks like the picture shown in Figure 3-11, you will have to re-solder the joint over again. Figure 3-12 shows another type of dry solder joint which would have to be re-soldered. These types of "dry" solder joints if now redone will cause the circuit to be unreliable and intermittent.

For projects that use one or more integrated circuits, with their leads closely-spaced pins, you may find it easier to use a finer gauge solder, i.e. less than 1 mm in diameter. This reduces the risk of applying too much solder to each joint, and accidentaly forming "bridges" between pads to PC "tracks."

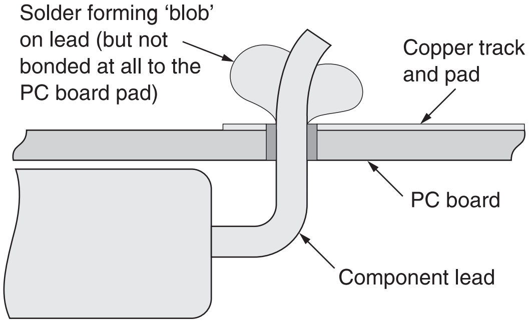
Solder forming 'blob' on lead (but not bonded at all to the PC board pad)
Copper track and pad
PC board
Component lead

Figure 3-12 *Inferior solder joint: Example 2*

The finished connection should be smooth and bright. Reheat any cloudy or grainy-looking connections. Finally, clip off the excess wire length, as shown in Figure 3-13.

Occasionally a solder "bridge" will form between two adjacent foil conductors. You must remove these bridges; otherwise a short circuit will exist between the two conductors. Remove a solder bridge by heating the bridge and quickly wiping away the melted solder with a soft cloth. Often you will find a hole on the board plugged by solder from a previous connection. Clear the hole by heating the solder while pushing a component lead through the hole from the other side of the board. Good soldering is a skill that is learnt by practice.

How to un-solder

In order to remove components, you need to learn the art of de-soldering. You might accidentally make a wrong connection or have to move a component that you put in an incorrect location. Take great care while un-soldering to avoid breaking or destroying good parts. The leads on components such as resistors or transistors and the lugs on other parts may sometimes break off when you are un-soldering a good, tight joint. To avoid heat damage, you must use as much care in un-soldering delicate parts as you do in soldering them. There are three basic ways of un-soldering. The first method is to heat the joint and "flick" the wet solder off. The second method is to use a metal wick or braid to remove the melted solder. This braid is available at most electronics

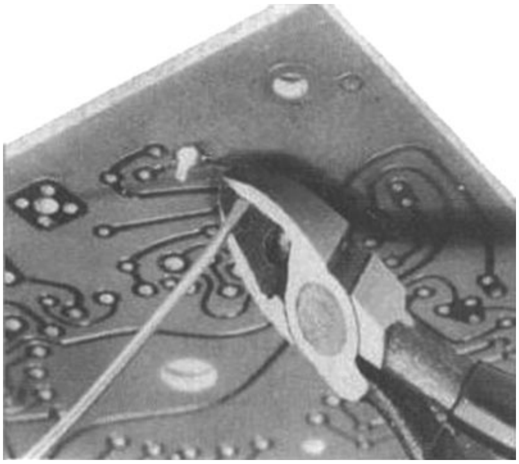

Figure 3-13 *Trimming excess component leads*

parts stores; use commercially made wicking material (braid). Place the braid against the joint that you want to un-solder. Use the heated soldering iron to gently press the braid against the joint. As the solder melts, it is pulled into the braid. By repeating this process, you can remove virtually all the solder from the joint. Then re-heat the joint and lift off the component leads. Another useful tool is an air-suction solder remover.

Most electronics parts stores have these devices. Before using a de-solder squeeze bulb, use your soldering iron to heat the joint you want to un-solder until the solder melts. Then squeeze the bulb to create a vacuum inside. Touch the tip of the bulb against the melted solder. Release the bulb to suck up the molten solder. Repeat the process until you've removed most of the solder from the joint. Then re-heat the joint and gently pry off the wires. This third method is easy, and is the preferred method, since it is fast and clean. You can use a vacuum device to suck up molten solder. There are many new styles of solder vacuum devices on the market that are much better than the older squeeze bulb types. The new vacuum de-soldering tools are about 8 to 12 inches long with a hollow Teflon tip. You draw the vacuum with a push handle and set it. As you re-heat the solder around the component to be removed, you push a button on the device to suck the solder into the chamber of the de-soldering tool.

De-soldering station

A de-soldering station is a very useful addition to an electronics workshop or workbench, but in many cases they just cost too much for most hobbyists. De-soldering stations are often used in production environments or as re-work stations, when production changes warrant changes to many circuit boards in production. Some repair shops used de-soldering stations to quickly and efficiently remove components.

Another useful de-soldering tool is one made specifically for removing integrated circuits. The specially designed de-soldering tip is made the same size as the integrated circuit, so that all IC pins can be de-soldered at once. This tool is often combined with a vacuum suction device to remove the solder as all the IC pins are heated. The IC de-soldering tips are made in various sizes; there are 8 pin, 14 pin and 16 pin versions, which are used to uniformly de-solder all IC pins quickly and evenly, so as not to destroy the circuit board. The specialized soldering tips are often used in conjunction with vacuum systems to remove the solder at the same time.

Remember these things when un-soldering:

1. Be sure there is a little melted solder on the tip of your iron so that the joint will heat quickly.
2. Work quickly and carefully to avoid heat damage to parts. Use long-nosed pliers to hold the leads of components just as you did while soldering.
3. When loosening a wire lead, be careful not to bend the lug or tie point. Use pieces of wire or some old radio parts and wire. Practice until you can solder joints that are smooth, shiny, and tight. Then practice un-soldering connections until you are satisfied that you can do them quickly and without breaking wires or lugs.

Caring for your soldering iron

To get the best service from your soldering iron, keep it cleaned and well tinned. Keep a damp cloth on the bench as you work. Before soldering a connection, wipe the tip of the iron across the cloth, then touch some fresh solder to the tip. The tip will eventually become worn or pitted. You can repair minor wear by filing the tip back into shape. Be sure to tin the tip immediately after filing it. If the tip is badly worn or pitted, replace it. Replacement tips can be found at most electronics parts stores. Remember that oxidation develops more rapidly when the iron is hot. Therefore, do not keep the iron heated for long periods unless you are using it. Do not try to cool an iron rapidly with ice or water. If you do, the heating element may be damaged and need to be replaced or water might get into the barrel and cause rust. Take care of your soldering iron and it will give you many years of useful service.

Remember, soldering equipment gets hot! Be careful. Treat a soldering burn as you would any other. Handling lead or breathing soldering fumes is also hazardous. Observe these precautions to protect yourself and others!

Ventilation

Properly ventilate the work area where you will be soldering. If you can smell fumes, you are breathing them. Often when building a new circuit or repairing a “vintage” circuit you may be soldering continuously for a few hours at a time. This can mean you will be breathing solder fumes for many hours and the fumes can cause you to get dizzy or lightheaded. This is dangerous because you could fall down and possibly hurt yourself in the process. Many people highly allergic are also allergic to the smell of solder fumes. Solder fumes can cause sensitive people to get sinus infections. So ventilating solder fumes is an important subject. There are a few different ways to handle this problem. One method is to purchase a small fan unit housed with a carbon filter which sucks the solder fumes into the carbon filter to eliminate them. This is the most simple method of reducing or eliminating solder fumes from the immediate area. If there is a window near your soldering area, be sure to open the window to reduce the exposure to solder fumes. Another method of reducing or eliminating solder fumes is to buy or build a solder smoke removal system. You can purchase one of these systems but they tend to be quite expensive. You can create your own solder smoke removal system by locating or purchasing an 8 to 10 foot piece of 2 inch diameter flexible hose, similar to your vacuum cleaner hose. At the solder station end of the hose, you can affix the hose to a wooden stand in front of your work area. The other end of the hose is funneled into a small square “muffin” type fan placed near a window. Be sure to wash your hands after soldering, especially before handling food, since solder contains lead; also try to minimize direct contact with flux and flux solvents.

Chapter 4

AM, FM, and Shortwave Crystal Radio Projects

Parts list

Simple Crystal Radio

L1 150 turns of 24 ga. enameled wire on 4″ × 2″ dia tube

C1 365 pF tuning capacitor (old radio)

D1 1N34 germanium diode (not a silicon diode)

C2 .005 µF, 35 volt disk capacitor

H1 high impedance headphones (crystal or 2000k type)

ANT 80–100 foot long wire antenna

GND cold water pipe or ground rod

Misc wood block, coil tube, hardware, screws, clips, washers, Krylon spray, wire, etc.

FM Crystal Radio Set

R1 47k ohm ¼ watt, 5% resistor

C1 82 pF capacitor

C2 80 pF air variable capacitor

C3 18 pF capacitor

D1 1N34 or rock crystal diode

L1 5 turns AWG #18 bare copper or silver wire, 12 mm inside diameter, tapped at 2.5 turns

H1 high impedance headphones (crystal type of 2k ohm type)

Ant antenna- 7″ of #18 bare copper wire

Misc wood block, wire, hardware, clips, etc.

Supersensitive AM/Shortwave Crystal Radio

L1 11 turns, 22 ga. enameled wire on 2⅞″ cardboard form

L2 (AM) 54 turns, 22 ga. enameled wire on 3″ air-core form

Taps are brought out at 40T, 27T, 15T and 6T

L2 (SW) 15 turns, 22 ga. enameled wire on 3″ air-core form

Taps are brought out at 12T, 9T, 6T and 3T

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

D1	germanium diode or galena crystal w/”cat’s whisker”
C1	500 pF turning capacitor
C2	.001 µF, 35 volt disk capacitor
S1	single-pole – 5-position rotary switch
S2	single-pole - 6-position rotary switch
H1	crystal headphone or 2k headphones (Baldwin, Brandes)
ANT	80–100 foot long wire antenna
GND	cold water pipe or ground rod
Misc	wood block, hardware, wire, clips, screws, Krylon, etc.

A crystal radio is the simplest form of AM (amplitude modulation) receiver ever invented. It has great potential for experimentation and usually requires no source of power for its operation other than the radio signal itself, and costs little. Most people do not realize that crystal radios can be built to pickup shortwave as well as FM radio signals. Did you know that you can build very sensitive and selective crystal radios far better than most commercial AM radios! Ever thought of building a crystal radio? Building a crystal radio will give you immense satisfaction, and the results are sure to please.

Crystal sets date back to the earliest days of wireless (pre-W.W.I) and an enormous variety of circuit designs have been produced over the years. Their popularity has been variable as developments in other more elaborate forms of reception have taken place. However, fascination with this crystal radio design, building and experimentation, is still very strong today with national and international organizations and clubs offering competitions.

So how does a crystal radio work? An AM transmitter sends out its broadcast in the form of an electromagnetic wave that radiates from its transmitting antenna. The AM transmitter sends out a fixed frequency carrier wave. When sound is present in the program material, the strength, or amplitude, of these waves is made to vary in response to the audio content of the program. The resulting wave is called an amplitude modulated wave. See Figure 4-1.

The purpose of a simple AM or crystal radio receiver is to pick up these AM waves and extract the audio signal so that it can be heard by the listener. It does this by a process called “detection.” The detection process utilizes a device called a detector which effectively strips off either the upper half or lower half of the AM wave. It only remains to filter out the carrier wave to leave the audio signal and the job is done! The following diagram shows how the detector and filter work together to “recover” the original audio signal. The strength of the recovered audio signal is small but, given the right conditions, it can be large enough to drive a pair of headphones; all without any power source other than the signal itself.

The diagram shown in Figure 4-2 illustrates a basic crystal radio. In its simplest form, a crystal radio is composed of just our main components. A good antenna, a detector, a filter capacitor and pair of headphones. The basic diagram shows the input section at (A), the rectifier section at (B), the filter section at (C) and finally the headphones at section (D).

The antenna consists of a length of wire suspended above the ground, while the ground or earth connection could be a metal spike driven into the ground. When a

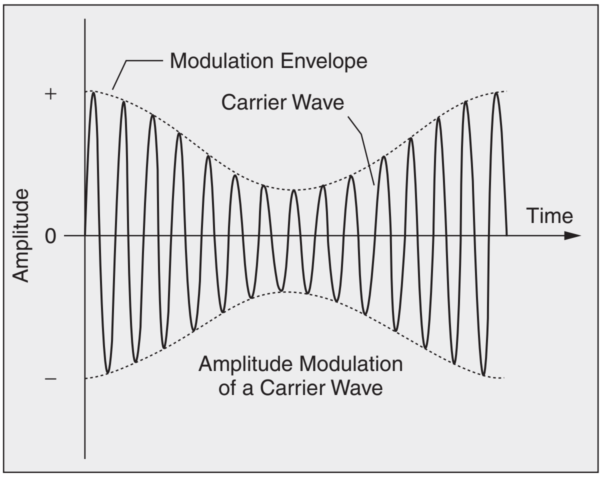

Figure 4-1 *Amplitude modulation*

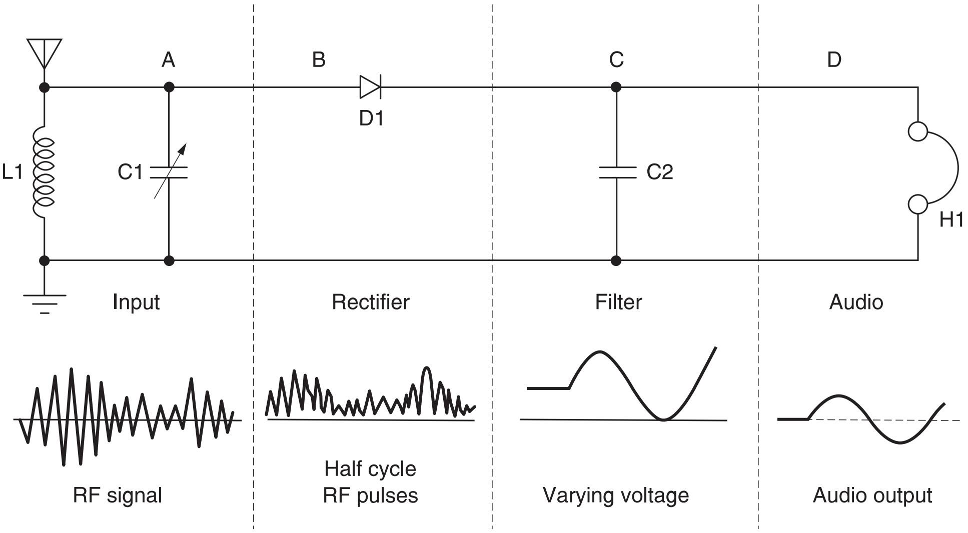

Figure 4-2 *Basic crystal radio*

capacitor is connected in parallel with an inductor and an alternating voltage applied across the combination, alternating current will flow. The amount of current that flows depends upon the frequency of the applied voltage. At a particular frequency, called the resonant frequency, almost no current flows. For frequencies above or below the resonant frequency, significant current will flow. In Figure 4-2, capacitor C1 plus the aerial capacitance forms the capacitor of our tuned circuit, while L1 forms the inductor. At the resonant frequency of the tuned circuit, almost no current flows to earth through L1 or C1, leaving virtually all of it free to flow through the detector. Alternating currents at broadcast frequencies either above or below the resonant frequency will tend to flow through C1 or L1 to earth. Thus we have added selectivity to the receiver. Notice that C1 is adjustable, as signified by the arrow. By varying C1 we can tune the receiver to select specific broadcast frequencies.

The main requirement for a detector is that it should act as a non-return "valve" or one-way "switch" to the alternating currents of the AM wave. If you could look at the alternating current in a circuit you would see it flowing back and forth, first one way and then the other. A detector placed in such a circuit allows the alternating current to flow easily in one direction but not in the other. Certain naturally occurring minerals were found to have this property and these became some of the earliest forms of detectors. One such mineral, the crystalline form of galena (lead sulphide), was found to be particularly good at detection, and became very popular for building crystal radios. The detector consists of galena in a small cup or tin. A small coil of wire, with one free end was called the "cat's whisker" which was used to make contact with the crystal.

The remaining uni-directional current is filtered by the combined effect of the headphone impedance and the capacitor at C2 and only the desired audio signal current passes on through to the headphones at H1. One drawback with the very simple circuit shown is that it has little selectivity. That is, it picks up all AM broadcast signals with similar efficiency and often close together. So as time passed more advanced, more sensitive and more selective receivers were born.

Let's construct a simple crystal radio. In order to build the crystal radio shown in Figure 4-3, you will need to secure some tools and supplies. First you will need to find a clean well lit work bench or table to spread out all your tools, charts, diagrams and components. You will also need to secure a soldering iron, some 60/40 rosin core solder and a small jar of "Tip Tinner," a soldering iron tip cleaner/dresser, obtainable from your local Radio Shack store. You will also want to locate some small tools such a pair of end-cutters, a pair of needle-nose pliers, a magnifying glass and a set of Phillips and flat-blade screwdrivers for this project. Grab the crystal radio schematic, see Figure 4-2, and the resistor and capacitor identification charts, and place them in front of you, so we can get started building the project. The resistor identification chart in Table 4-1 will help you

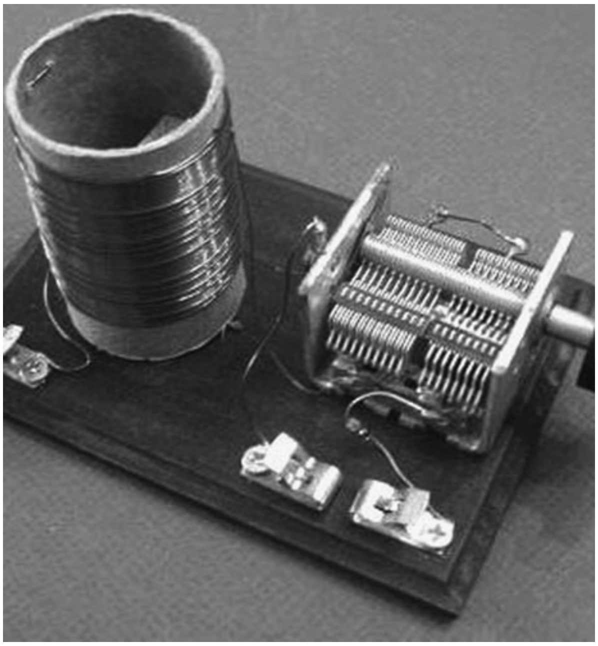

Figure 4-3 *AM crystal radio*

identify resistors needed for the some of the projects in this chapter. Each resistor will have three or four colored bands on the body of the resistor. These colors identify the resistor's value. The color bands start at one end of the resistor, the first colored band is the first digit of the resistor's value, while the second resistor represents the resistor's second digit of the value. The third colored band is the resistor's multiplier value. For example, a resistor whose first color band is yellow represents a digit four (4), while a second color band which is violet is seven (7). If the third color band is orange, then the multiplier would be (000) or one thousand, so the value would be 47,000 or 47k ohms. Resistors usually have a fourth colored band which represents the resistor's tolerance value. A silver band denotes a 10% tolerance value, while a gold band denotes a 5% tolerance resistor. No color band represents a tolerance value of 20%.

Our crystal radio projects in this chapter all utilize capacitors of some sort. Capacitors usually are available in two major classifications, polarized and non-polarized. The project in this chapter will utilize only non-polarized capacitors. Non-polarized capacitors can often be physically very small and not have their actual value printed on the body of the capacitor, so a chart was developed to help identify a capacitor using a three-digit code. See Table 4-2.

As this project will not require a circuit board, we will build the crystal radio on a 6″ × 6″ × ¾″ wood block base. You can elect to use an additional piece of wood at a right angle to the base as a front panel to secure the tuning capacitor as shown in Figure 4-3.

Table 4-1
Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

Table 4-2

Capacitance code information

This table provides the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Some capacitors actually have the numeric values stamped on them; however, many are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

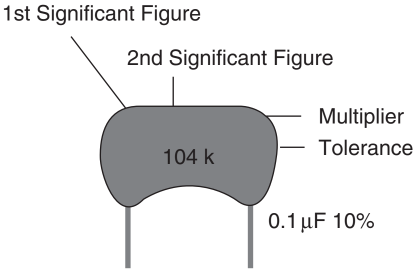

CSGNetwork.Com 6/4/92

Value	Type	Code	Value	Type	Code
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.001 5µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

Next, you will need to locate a plastic or cardboard tube approximately 3½ to 4″ long by 2″ in diameter on which to wind your coil. Punch a hole at both ends of the tube about ½″ from the actual end of the tube. Start winding the coil at one end of the tube using #24 ga. enameled wire, leave about 6″ of wire free at the first hole to allow connection to the circuit. Wind the coil with each winding touching the previous winding but not overlapping the prior winding. Wind about 150 turns of wire across the length of the tube. After winding about 150 turns take the free end of the remaining wire and put it through the last punched hole, leave about 6″ of wire so you can connect the coil to the circuit. When you are finished winding the coil, you can varnish the coil winding or just spray Krylon around the coil to keep the winding close to the coil, so they won't spread out or come loose. Fasten the coil to a 6″ × 6″ wood block; you can mount the coil upright as shown in Figure 4-3, or you can lay the coil down near one edge of the wood block and secure it to the block with some brackets or clips. You will have to use a knife and scrape off the enameled insulation of the two coil wires before you can fasten or solder them to the tuning capacitor. Now fasten the tuning capacitor to the wood block near the coil assembly, and wire the coil in parallel with the tuning capacitor. Next locate a germanium diode; make sure you use only a germanium diode and not the more common silicon diode. Remember to observe the correct polarity when connecting the diode in the circuit, the black to white colored band is the cathode lead. The cathode end of the diode points to the headphone in the circuit diagram. The anode side of the diode connects to one end of the "tank" or coil/capacitor and the cathode side of the diode connects to capacitor C2 and one end of the high impedance headphones. The free end of the headphones connects to the remaining end of the capacitor at C2 which connects to the free end of the "tank" circuit. You can connect the headphones via a terminal strip, a Fahnestock clip or headphone jack mounted on the wood block. Crystal radios generally require high impedance headphones, so you must use either a small crystal headphone or a pair of older high impedance (2000) ohm headphones; these are still available through EBAY or from Antique Electronics.

You could elect to build the radio using a galena crystal, with a "cat's whisker" instead of the germanium diode to give the radio an "old-time" look if desired: see the Appendix for Antique Radio Supply.

Connect an antenna and a ground wire to rod in the ground or an electrical ground in your house and your crystal radio will be ready to test out. The ground wire can be attached to a metal cold water pipe, or to a metal rod stuck a couple of feet into the earth. Do not attach it to a line carrying gas or electricity. The crystal radio will operate forever without a power supply, but you do need a good antenna to make the radio work well. The simplest outdoor antenna for receiving signals is a horizontal wire with the lead-in wire to the set connected on one end. An insulated wire 50 to 100 feet long placed as high and clear of surrounding objects as possible works best: see Figure 4-4. Both ends of the antenna should have insulators and rigid support. Tree branches are not good because they sway in the wind. If you use a tree trunk, have slack in the wire so it can move without pulling the wire. Keep the antenna wire far away from chimneys, metal roofs, gutters, drain pipes, telephone, and especially power lines. The antenna and lead-in should be one continuous piece of wire. Use stranded copper wire, with only the lead-in extending beyond the insulator. A lightning arrestor and ground wire should also be used to avoid a direct hit by lightning.

A crystal radio will generally be able to detect radio signals within 25 to 35 miles, during the day and often much further at night. A simple crystal radio is not very selective, if there are many local AM radio stations that are close in frequency, then you may have some trouble tuning them in, and you may have to build a more selective radio, such as the All-Wave model below. The crystal radio project above is a great project for demonstrating the principles of AM radio and the fun and excitement of building a crystal radio or "free-power" radio.

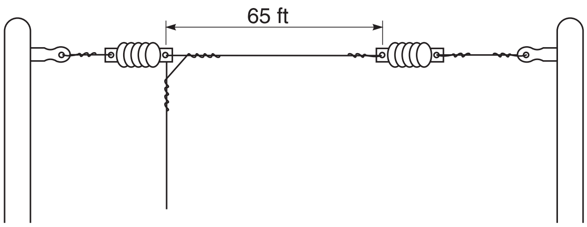

Figure 4-4 *Basic long-wire antenna*

FM radio crystal radio

Believe it or not, you can actually create a simple FM radio crystal radio. An FM crystal receiver is depicted in Figure 4-5. The key to understanding how this simple circuit can decode FM signals is to understand what happens when an FM signal is coupled to a tuned circuit. A good tuned circuit (one with a high "Q factor") will attenuate signals that are not near its resonant frequency. If you apply an FM signal to a good tuned circuit, as the frequency of the FM signal moves away from the resonant frequency of the tuned circuit, the tuned circuit will attenuate that signal. As the frequency of the FM signal moves toward the resonant frequency of the tuned circuit, the tuned circuit will let more of the signal pass through. Thus, a tuned circuit will impose amplitude changes on an FM radio signal that match the frequency changes. A crystal diode is sensitive to amplitude changes, so it converts the amplitude changes into a signal that the listener can hear through headphones.

Using a tuned circuit to induce amplitude changes onto an FM signal, then converting the amplitude changes to sound frequencies, is called slope detection. "Slope" refers to the slope of the tuned circuits' attenuation curve. If you have an AM receiver you can listen to an FM signal by using the *slope* of the tuned circuits in the receiver. This *slope tuning* method gives you none of the real advantages of FM. For instance, FM is noted for its immunity to static. An AM receiver used to slope detect will hear static. The only real advantage to slope detection is that is can be cheap and convenient. When tuning in an FM signal with an AM receiver, the clearest sound occurs when the FM center frequency is offset from the center frequency of the receiver's tuned circuits. If the FM signal is at the receiver's center frequency, then frequency deviations both up and down produce downward amplitude changes. When the FM signal is off center frequency, up and down frequency changes create up and down amplitude changes in the receiver.

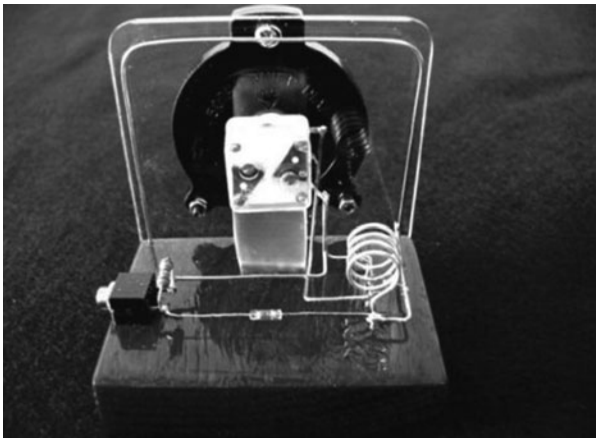

Figure 4-5 *FM crystal radio*

The FM crystal receiver circuit, shown in Figure 4-6, looks almost identical to a classic AM crystal circuit; but look closely, there are a few major changes. The FM crystal radio is relatively easy to build and can be built on a small block of wood as shown. The components values for the "tank" or resonant circuit were reduced to resonate at higher frequencies in the FM band. This was done by experimenting with smaller and smaller coils and capacitors values. The antenna is also much reduced in size (from that of AM crystal radio) to resonate at higher frequencies. The antenna is actually a 7″ long bare copper wire. The coil at L1 is just four turns #18 copper or silver wire, 12 mm inside diameter, tapped at 2.5 turns, The coil was wrapped around a "magic marker," then slipped off and expanded. As you can see, this is far different than the 150 turn coil for the AM crystal radio. The detector diode, once again, is a 1N34 germanium diode and not a silicon diode, or you could use a galena crystal as a detector if you desire. Capacitor C3 is 18 pF, but is not critical and can be 10 to 50 pF. Resistor R1 is a 47k ohm, ¼ watt, 5% carbon composition type. The air variable capacitor, used in the prototype, was an 80 pF variable type and it had two trimmers in it which should be adjusted for best reception.

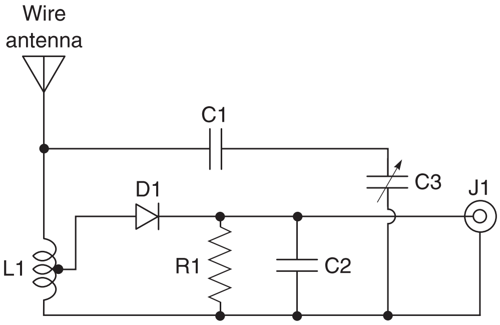

Figure 4-6 *Simple FM crystal radio*

Commonly available vernier dial and knob will fit the tuning capacitor and will work nicely. When building the FM crystal radio it is important to keep the components physically close together and the wiring short as possible. You can build the FM crystal radio in an open style format on a wood or Lucite plastic sheet. You can make your receiver a work of art, and impress your friends with an FM crystal radio built from scratch. Note that you will have to use high impedance headphones for this receiver. A detected FM signal is converted to AM due to an effect called slope detection that modulates amplitude.

This FM crystal radio works best near the transmitter, reception seems fine up to 15 miles away. Secondly, the sound level is not very loud; a quiet room is best for listening. One must be willing to move the set around to find a location for the best reception of signals. If desired, you could connect the audio output of the FM set directly to an audio amplifier's fed to a small speaker.

Supersensitive AM/Shortwave crystal radio

There are several characteristics that make the difference between a good and a great receiver. One of these characteristics is sensitivity. Great receivers can pick up weak signals and amplify them into a useful output. A second characteristic is selectivity. Great receivers can separate signals so that only one is heard at a time. The third and final thing that separates good receivers from great receivers is fidelity. Great receivers not only pick up weak signals but can reproduce a signal without distortion. The last project in this chapter is the All-Wave, AM/Shortwave crystal radio. This All-Wave design takes the crystal radio concept a little further and provides a more sensitive and more selective radio for you to build and enjoy.

The All-Wave Crystal Radio is a supersensitive and selective AM and shortwave receiver; see Figure 4-7. The design of the All-Wave radio has been around since the early 1930s and is still a very efficient receiver design that is fun to build and use. The receiver will tune from below the AM broadcast band to around the 40 meter band, so you will be able to pickup international shortwave reception quite well. You will be surprised how well it performs with a good pair of high impedance headphones.

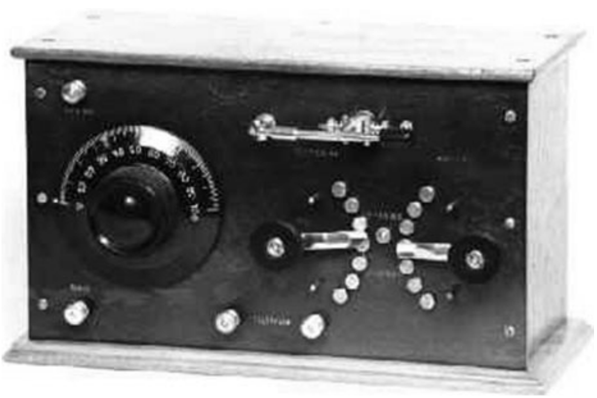

Figure 4-7 *Shortwave crystal radio*

The key to this efficient receiver's design lies with the two coils, L1 and L2, which are shown in the schematic at Figure 4-8. Coil L2 is a multi-tapped plug-in coil based on the band selected. For the AM radio band, L2 is the 54 turn multi-tapped coil, with tapped from the ground end at 6T, 15T, 27T and at 40T. The shortwave version of the L2 coil is a 15 turn multi-tapped coil with tapped at 3T, 6T, 9T, and 12 turns. Depending upon the band chosen, you simply plug in the AM or shortwave coil. The AM and shortwave coils (L2) were wound on a 3″ cardboard tube, but you could use a plastic form instead. The untuned primary coil (L1) was wound a slightly smaller 2 ⅞ form, so that it can be slipped inside of the L2 coil at the ground side of the coil assembly. The interchangeable coils (L2) were mounted on a non-metallic carrier, a piece of phenolic board with five pin-plugs secured to the carrier board. The coil form for L2 was glued to the phenolic carrier board which forms the plug-in coil assembly. The wire tapes are brought out to the five pin-plugs on the carrier board, so that the coil (L2) assembly can be plugged into sockets on the radio's wood base plate, see Figure 4-9. This will allow either the AM coil or the shortwave coil to be plugged into the crystal receiver. You could elect to build a fixed coil version of the crystal radio for either AM or shortwave reception if you desire, making the receiver a bit easier to build.

The detector for the All-Wave Crystal Radio can be either a 1N34 germanium diode or you could elect to

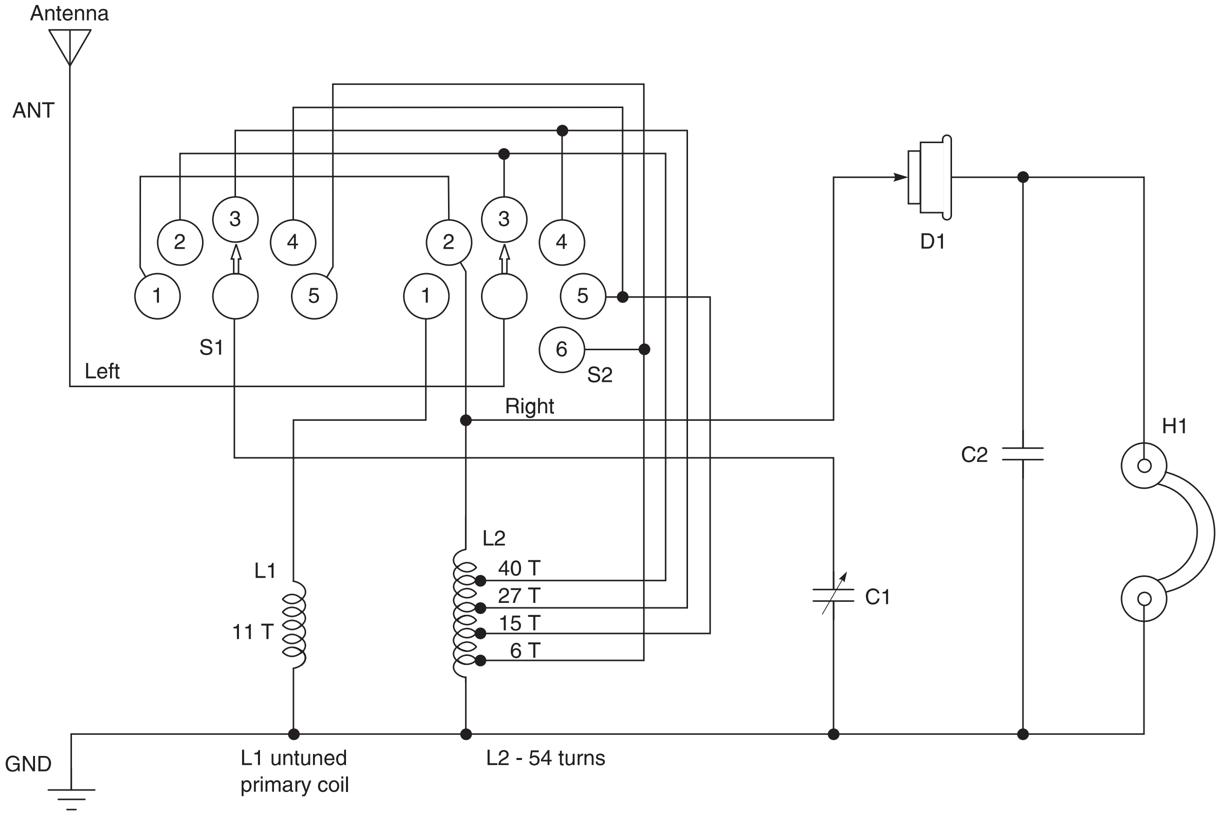
Antenna
ANT
1
2
3
4
5
S1
Left
1
2
3
4
5
6
S2
Right
D1
H1
C2
C1
L1
11 T
L2
40 T
27 T
15 T
6 T
GND
L1 untuned
primary coil
L2 - 54 turns

Figure 4-8 *Sensitive/selective AM/shortwave crystal radio. Courtesy of Midnight Science*

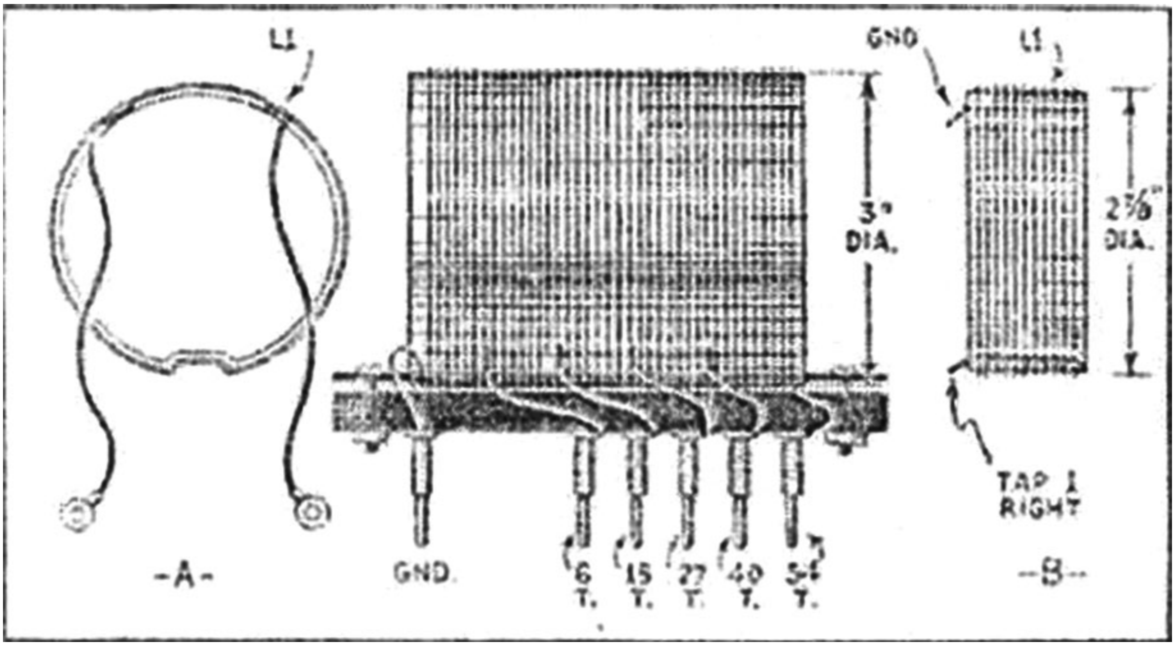
L1
GND
L1
3" DIA.
2⅝" DIA.
TAP 1 RIGHT
-A-
GND.
6 T.
15 T.
27 T.
40 T.
54 T.
-B-

Figure 4-9 *Shortwave coil assembly*

build the radio more like the "original" version by using a galena crystal with a "cat's whisker." The galena crystal and "cat's whisker" are still available, see Appendix. Be careful to install the diode or galena crystal with the proper orientation.

The original All-Wave receiver used a 500 pF tuning capacitor at C1, but you could also use a 365 pF type which are easier to locate from sources such as Antique Radio Supply. Mount the tuning capacitor to the wood block base using standoff and screws. The ceramic disk capacitor at C2 acts as an audio filter, ahead of the high impedance headphones.

The All-Wave receiver needs to be used with either a crystal earphone or a pair of dual high impedance (2K) headphones such as the old Baldwin or Brandes headphones which are still available via EBAY or from Antique Radio Supply, see Appendix.

Switches S1 and S2 are used to select the taps between the antenna and the detector. The two switches are labeled **Right** and **Left** as shown in the diagram. The same station may come in on several taps, but use the one which places the station lowest on the turning capacitor setting for the loudest signals. Move the **Right** (S2) forward or clock-wise, one to three taps before advancing the **Left** switch (S1). The efficiency of this radio seems to lie partly in the low loss coils, but mostly in the one basic change in the circuit, in which the detector is connected permanently to the last turn of L2 towards ground.

The All-Wave Crystal radio is just about completed and you now have only to connect a good antenna and ground for the receiver to spring-to-life! The All-Wave Crystal Radio should be connected to a good ground such as a cold water pipe or 6 to 8 foot copper ground rod pounded into the ground. The All-Wave receiver will perform best with an 80 to 100 foot insulated long-wire antenna placed up-in-the-air outdoors away from metal objects. An alternate antenna might be a 50 to 60 foot insulated wire strung around your attic, away from any metal objects. You will be amazed at how well this receiver will work with a good antenna and a good pair of high impedance headphones.

In the event that your new All-Wave Crystal Radio fails to work, you will have to re-check a few details. First check the diode and how it was oriented in the circuit, you may have placed it backward and this will prevent the receiver from functioning. You will also need to check the coil connections and placement of L1 with respect to coil L2. There is no polarity regarding coil L1, but its placement at the ground end of L2 is important in order for the receiver to work properly. Also check the connections to L2, you may have reversed two end wires of the coil, and this will prevent the radio from working as well. Try the receiver once more, hopefully all of the connections are right now and the receiver should work fine. Reconnect up your ground and antenna, put on your high impedance headphones and you'll be ready to listen to your new All-Wave receiver. You may want to have a "log" book close to your new radio to log down the far away stations you will receive with this radio. During nighttime hours, you will be able to receive stations much further away, and when the shortwave coil is plugged in you will hear international shortwave stations from all around the world on your new All-Wave receiver. Have fun with your new All-Wave crystal radio.

Chapter 5

TRF AM Radio Receiver

Parts list

TRF AM Radio Project

R1 120k ohm, ¼ watt, 5% resistor

R2 2.7k ohm, ¼ watt, 5% resistor

R3 10k ohm potentiometer (volume)

R4 10k ohm potentiometer (trim)

R5 820 ohm, ¼ watt, 5% resistor

R6,R9,R13 10k ohm, ¼ watt, 5% resistor

R7,R8 22k ohm, ¼ watt, 5% resistor

R10 150k ohm, ¼ watt, 5% resistor

R11,R12 4.7k ohm, ¼ watt, 5% resistor

C1 10 nF, 35 volt polyester capacitor

C2 500 pF tuning capacitor (variable)

C3 100 nF, 35 volt polyester capacitor

C4,C5 2.2 µF, 35 volt electrolytic capacitor

C6 33 µF, 35 volt electrolytic capacitor

C7 220 µF, 35 volt electrolytic capacitor

C8 2200 µF, 35 volt electrolytic capacitor

L1 220 µH coil; see text (Ocean State-LA-540)

D1,D2,D3,D4,D5,D6 1N4148 silicon diodes

Q1 BC108B-NTE123A transistor

Q2 BC109C-NTE123AP transistor

Q3 BC179C-NTE234 transistor

U1 ZN414 or MK484 - TRF Radio IC

U2 LM741 Op-amp IC

S1 SPST power switch

B1 9-volt transistor radio battery

SPK 8-ohm speaker or headphone

Misc PC board, chassis, battery holder, battery clip, wire, hardware

AM radio is the name given for the type of radio broadcasting used in the medium wave band in all parts of the world and in the long wave band in certain areas, notably Europe. Medium wave frequencies in the US range from 540 to 1710 kHz with 10 kHz spacing. The letters AM stand for amplitude modulation,

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

the technique by which speech and music are impressed on the radio waves. The sound waves from the studio are converted into electrical signals by the microphones, and these electrical signals are used to vary the amplitude, or the distance from 'peak' to 'trough' of the main constant frequency carrier wave generated at the transmitting station. As a result, the signal radiated by the station varies in amplitude in step with the variations in the loudness and pitch of the sound in the studio. In the receiver, the amplitude variations of the radio signal are separated from the carrier wave by a process known as demodulation, and after further amplification are fed to the loudspeaker.

Amplitude modulation is also used by international broadcasting stations operating in the shortwave band, and for some professional communications such as air traffic control, ship-to-shore traffic, and some mobile radio services.

AM radio stations in the medium wave or AM radio band can provide a service over an area the size of a town or of a whole continent, depending upon the power of the transmitter which can vary from less than a hundred watts to hundreds of kilowatts. Long wave AM radio stations cover even larger areas. Normally, all radio waves travel only in straight lines, which would limit a radio station to an area of only about 40 miles (65 km) in diameter—as far as the horizon. Medium and long wave stations can actually cover much larger areas because their signals are bent round the curvature of the Earth by the ionosphere (an electrically charged layer in the upper atmosphere which can reflect radio waves). In the US, where AM radio stations are operated on a commercial basis, coverage is generally and deliberately limited to one city and its environs. Major cities such as New York are served by large numbers of stations, each radiating a different program. AM radio stations in the US generally operate with omni-direction patterns during the day. In the evening many AM broadcasters switch to directional antenna beam patterns, and their broadcasts go much further due to ionospheric bending. This is why you can hear AM broadcast stations more distant in the evening.

AM radio first began with the experimental broadcast in 1906 by Reginald Fessenden, and was used for small-scale voice and music broadcasts up until World War I. The great increase in the use of AM radio came the following decade. The first licensed commercial radio services began on AM in the 1920s (the first licensed American radio station was started by Frank Conrad: KDKA in Pittsburgh, Pennsylvania). Radio programming boomed during the "Golden Age of Radio" (1920s–1950s). Dramas, comedy and all other forms of entertainment were produced, as well as broadcasts of news and music.

The quality of reproduction on AM radio is not as good as FM radio because of limitations in the amplitude modulation technique itself and because of overcrowding and subsequent interference in the medium wave band. Interference is particularly serious at night because the nature of the ionosphere alters after sunset, so that it can then reflect medium wave signals back to Earth a long way beyond their normal service area. Here, they interfere with other stations, producing whistles and other unwanted noises.

The AM method is still used for most popular broadcasting. Although the alternative FM radio system gives better quality, it is more difficult to receive, and is less suitable than AM for use in inexpensive transistor portables and car radios.

The most important concept in any communications system is the ability to send information from one place to another via a transmitter to a receiver, see Figure 5-1. This means you have to find a way to impress the audio signal information on the radio wave in such a way that it can be recovered at the other end. This process is known as modulation. In order to modulate a radio wave, you have to change either or both of the two basic characteristics of the wave: the amplitude or the frequency.

If you change the amplitude, or strength, of the signal in a way corresponding to the information you are trying to send, you are using a modulation system called amplitude modulation, or AM. The earliest means of radio communications was by Morse code, and the code key would turn the transmitter on and off. The amplitude went from nothing to full power whenever the key was pressed, a basic form of AM.

Modern AM transmitters vary the signal level smoothly in direct proportion to the sound they are transmitting. Positive peaks of the sound produce maximum radio energy, and negative peaks of the sound produce minimum energy.

The main disadvantage of AM is that most natural and man made radio noise is AM in nature, and

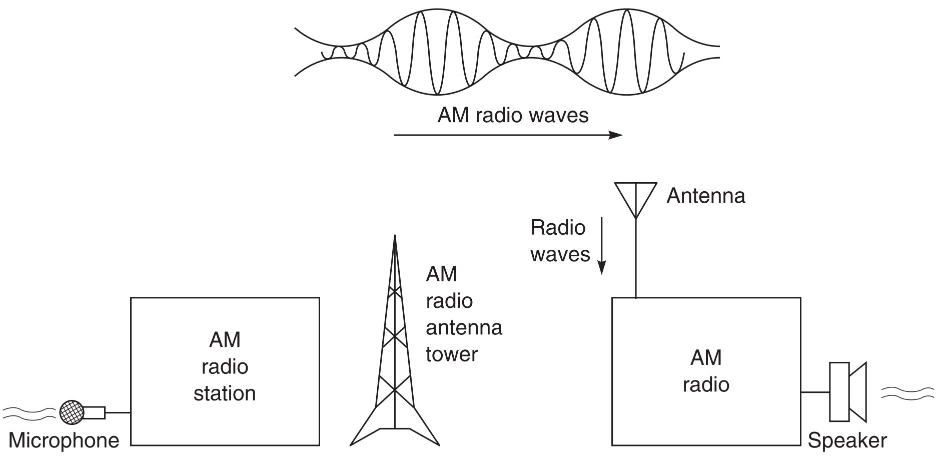

Figure 5-1 *AM broadcast system block diagram*

AM receivers have no means of rejecting that noise. Also, weak signals are (because of their lower amplitude) quieter than strong ones, which requires the receiver to have circuits to compensate for the signal level differences.

An AM radio station takes a low-frequency audio signal such as people talking or music (2500 Hz) and mixes it or superimposes on a high-frequency carrier wave (540 kHz), see Figure 5-2. It sends this radio signal out into the air as electromagnetic waves using a tall antenna. The electromagnetic wave travels through the air, going in all directions. At home, your AM radio takes this incoming signal, in a process called detection. With amplitude modulation it is simple. All you need to do is to rectify the signal, through a rectifier which removes the bottom half of every cycle. Essentially you are throwing away the high-frequency or carrier part, but keeping the low-frequency or audio part. At the output of the detector the original audio information is sent to the speaker in your radio. You then "hear" whatever was broadcast at the radio station.

The AM radio project that we are going to build in this chapter is called a TRF or tuned radio frequency type AM radio, see Figure 5-3. The TRF AM Radio design was at the forefront of radio design in the early 1920s. It utilized a tuned front end which allowed only RF energy of interest to be amplified by the RF amplifier, the signal was then passed on to the detector

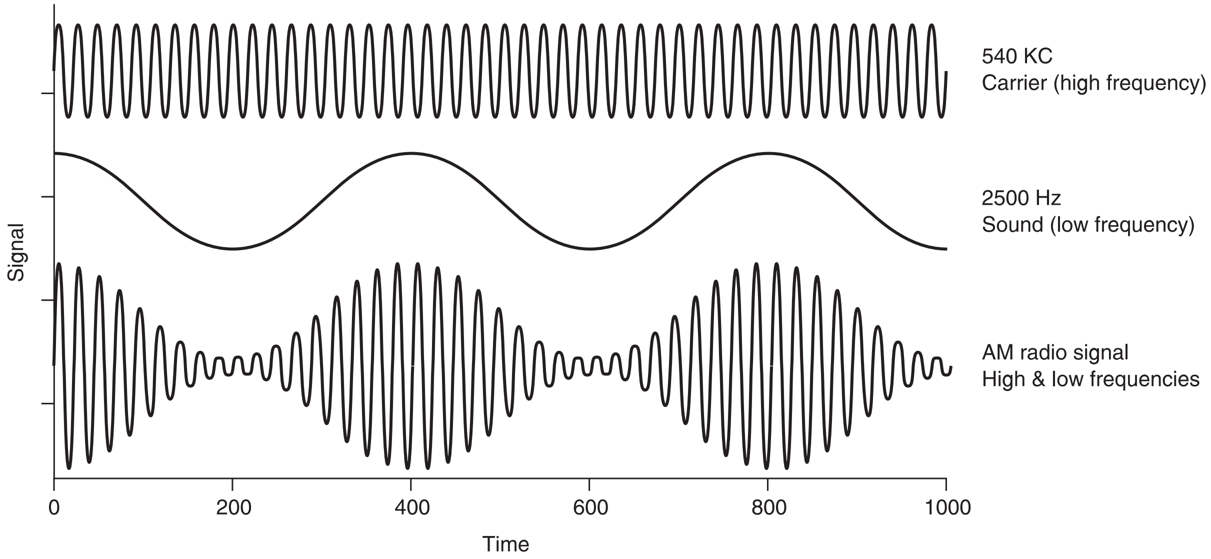

Figure 5-2 *AM carrier and modulation signals*

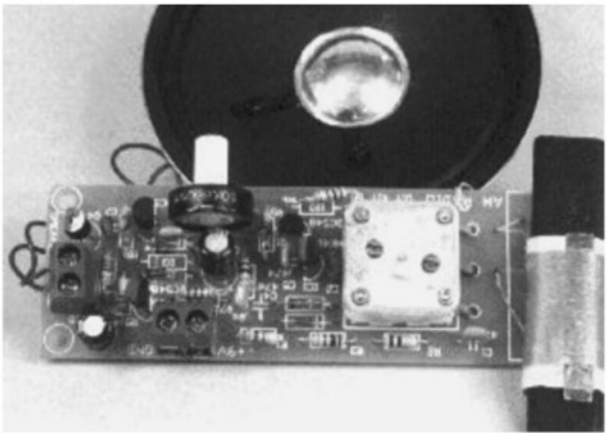

Figure 5-3 *AM TRF receiver*

followed by an audio amplifier to drive headphones. Early TRF radios had up to three controls for tuning the receiver, but later all three controls were consolidated into "tuning" control. Our modern day TRF receiver shown in the block diagram in Figure 5-4 illustrates the three major components of the AM TRF receiver. The first block consists of the tuned RF front-end resonant coil/capacitor, the second block consists of the MK484 IC, an integrated circuit RF amplifier and detector, and finally the third block, which is an audio amplifier.

The TRF AM radio project was designed around the popular MK484 TRF integrated circuit module, shown in Figure 5-5. This receiver covers the medium wave band or (AM) radio band from approximately 540 to 1700 kHz. The TRF front end consists of two components, the ferrite aerial coil and the tuning capacitor and the MK484 TRF IC. The important word is tuned, since this AM radio is called a TRF, or tuned RF radio All the AM signals reaching the radio are very weak. Only that signal which matches the TRF frequency is magnified by resonance so that it stands out at a very much higher level of signal strength. The original ZN414 IC, has now been replaced by the MK484, which is considerably easier to locate. The integrated circuit is a 3-pin, tuned radio frequency circuit, and incorporates several RF stages, automatic gain control and an AM detector. The receiver front-end is easily overloaded and the operating voltage of the IC is somewhat critical to achieve good results, but all-in-all the receiver is a great little radio for AM reception. The TRF radios were some of the first radios on the scene when the dawn of radio appeared in the early 1920s.

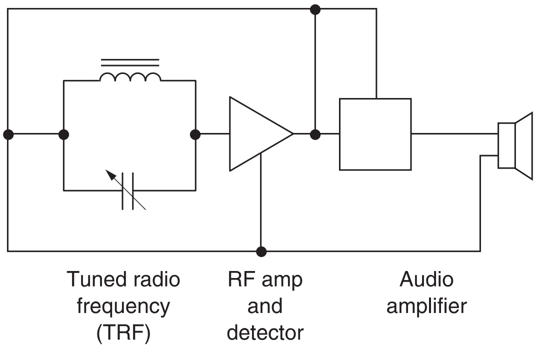

Figure 5-4 *TRF AM radio block diagram*

The front-end of the TRF receiver as mentioned consists of a tuned circuit composed of the ferrite inductor, a tuning capacitor and the TRF radio chip. A small square plastic 0-200 pF tuning capacitor was used for this radio project, but you could locate any suitable tuning capacitor within the range and it will work equally well. Often these small tuning capacitors will have additional on-board trimmer capacitors, but we used only the main tuning capacitor in the center of the device.

The second critical component of the TRF receiver is the coil at L1, which is wound on a ferrite bar. First, let's begin winding the coil at L1. The coil needs approximately 65 turns of #28 gauge enameled Litz wire on a 2½″ long ferrite rod. The windings should measure approximately 25 mm wide. Start winding the coil, close to one end, by securing the enameled wire by a piece of insulating tape. Leave approximately 100 mm for connection. The quality of reception is dependent upon the care taken in winding the antenna coil.

The signal reception will vary, dependent upon location, and is better away from large buildings that can distort the signal. Wind the wire tightly around the ferrite rod, taking care to keep the coil together and not to overlap the windings. To help hold the windings in place a piece of double-sided tape can be placed on the ferrite rod. Before soldering the antenna coil to the circuit board, you will need to scrape off the enamel coating from the ends of the wire. The two components L1 and C2 are connected in parallel to form an LC network. The Litz wire coil and the ferrite bar give the LC network a high "Q" or quality factor. This is important to the selectivity of the receiver, that is the ability to tune into one radio station at a time.

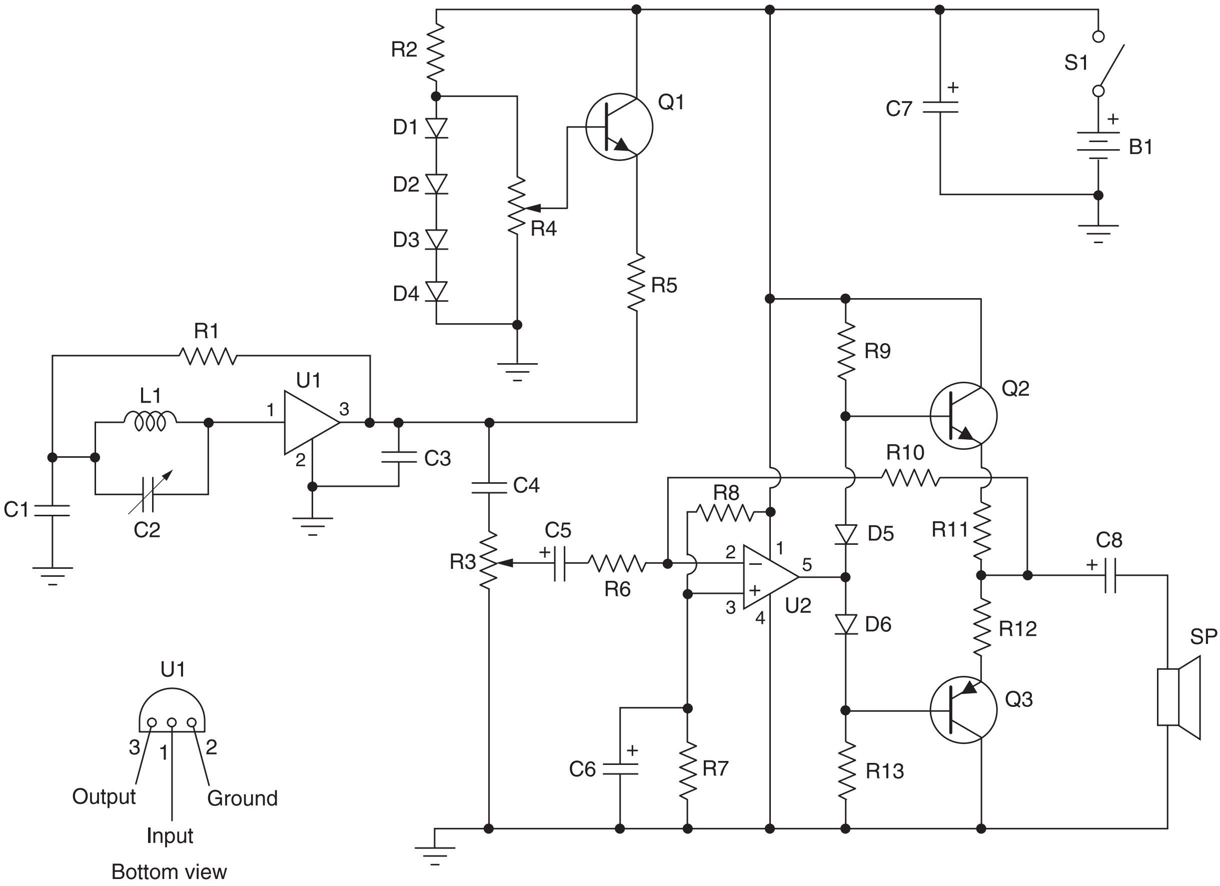

Figure 5-5 *TRF AM receiver broadcast band receiver*

The coil winding and the ferrite bar act as an efficient wire antenna, as no additional antenna is required. The only disadvantage is that the tuned circuit is directional and depends on the orientation of the bar with respect to the originating broadcast station. The coil winding and the ferrite bar act as an efficient wire antenna. No additional external antenna is needed. The only disadvantage is that the tuned circuit is directional. Signal strength depends on the orientation of the bar with respect to the origin of the signal. The coil and tuning capacitor may be taken from an old AM radio to save time, if desired.

The third main component of the TRF radio front-end is the TRF integrated circuit. The MK484 is a Japanese copy of the original ZN414. It contains an RF amplifier, active detector and automatic gain control (AGC to improve sensitivity) all in a 3-pin package. The input impedance is typically 4 M ohm. It operates over a range of 150 kHz to 3 MHz. DC supply of 1.1 V to 1.8 V and 0.3 mA current drain makes it ideal for battery operation. The output is typically 40–60 mV of audio signal. It is important to realize that the radio IC does not create any sounds by itself. It can only take the RF signal provided from the TRF circuit, amplify it, separate the audio signal from the RF (called detection) and pass the audio signal on to be amplified. It has no selective or rejection components contained in it. (This is in contrast to superhetrodyne receivers.)

In this circuit, a small voltage regulator is built around the BC108B transistor, four 1N4148 diodes, the 2.7k and 10k preset resistor and the 820 ohm resistor. The 10k pot acts as a selectivity control for the whole receiver, controlling the operating voltage for the ZN414 (or MK484). If you live in an area that is permeated with strong radio signals, then the voltage may need to be decreased. I found optimum performance with a supply of around 1.2 volts.

The audio amplifier is built around an inverting 741 op-amp amplifying circuit. Extra current boost is provided using the BC109C/BC179 complementary transistor pair to drive an 8 ohm loudspeaker. The voltage gain of the complete audio amplifier is around 15. The audio output of the complete receiver is really quite good and free from distortion. I used a small wooden enclosure and the complete tuning assembly from an old radio.

Before we begin the AM radio project, you will want to locate a well lit and well ventilated work space. You will want a large work table or work bench to spread out all you components, diagrams and tools. Locate a small pencil tipped 27 to 33 watt soldering iron and a roll of 60/40, 22 ga. tin/lead rosin core solder. If you don't have a jar of "Tip Tinner," a soldering iron tip conditioner/dresser, then you will want to go to a local Radio Shack store and purchase one. Place all the component parts in front of you along with the necessary project diagrams, so it will be easy to work with everything in front of you. Grab a few small tools, such as a pair of end-cutters, a pair of needle-nose pliers, a pair of tweezers, a magnifying glass, a small flat-blade, and small Phillips screwdriver. Once everything is right in front of you we can begin building the AM radio project. Warm up your soldering iron and we will begin. It is a good idea to use a circuit board for this project, since circuit boards provide a more reliable method used to build circuits: you could elect to use prototype construction or point-to-point wiring on a perf-board as an alternative.

Refer to the chart in resistor color chart in Table 5-1 and the capacitor code chart in Table 5-2. Resistors used in this project will be mostly carbon composition type resistors and potentiometers. Each resistor will have three or four color bands on the body of the resistor. The first band closest to one edge will be the first color which represents the first value. For example, a brown color band will signify the first digit as a one (1). The second color band represents the second digit value. So if the second color is red, then the value of the second digit would be two (2). The third color band is the multiplier value. So, if the third color is orange then the multiplier value would be (000), so the resistor value would be 12,000 or 12k ohms. Often there is a fourth color band which represents the tolerance value. No fourth band denotes a resistor with a 20% tolerance, a silver band denotes a 10% tolerance value, and a gold band represents a 5% tolerance value. Now refer the capacitor code chart. Many small capacitors are too small to have their full value marked on the body of the capacitor, so a three-digit code is used to represent the value.

Let's begin installing the project resistors now; find and identify a couple of resistors at a time. Once you are

Table 5-1
Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

Table 5-2

Capacitance code information

This table provides the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Some capacitors actually have the numeric values stamped on them; however, many are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

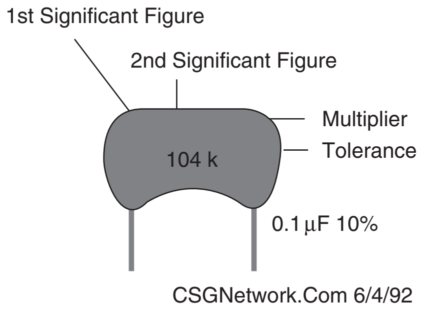

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

sure you have correctly identified each resistor, and you know where each resistor is placed on the PC board, you can go ahead and install them on the circuit board. Solder the resistors in place and then trim the excess component leads with your end-cutters. Cut the excess component leads flush to the edge of the circuit board. Install the remaining resistors, solder them in place and remember to trim the excess leads from the board.

Next we will move on to installing the capacitors. This project uses two types of capacitors, small mica or polyester types and electrolytic types. The smaller capacitors may not have their actual value printed on them, but might have a three-digit code on them. Refer to the capacitor code chart to help identify each capacitor before installing them on the PC board. Slowly identify each capacitor and place it in the correct location on the circuit board. Solder these small capacitors in their respective locations after referring to the schematic and parts layout diagrams. This project also utilizes larger electrolytic capacitors. Electrolytic capacitors have polarity and this must be observed if the circuit is to work correctly. Take a look at one of the larger electrolytic capacitors and you will notice that it may have a black or white color band or a plus (+) or minus (–) marking on one side of the capacitor body. When placing these types of capacitors on the circuit board, be sure you have oriented the capacitor correctly. Install the electrolytic capacitors on the circuit board and solder them in place; remember to trim the excess component lead with your end-cutter.

Let's move on to installing the silicon diodes. The AM radio project uses six silicon diodes, and they too have polarity which must be observed. Each diode will have either a white or black band at one end of the diode's body. This band denotes the cathode lead of the particular diode. Refer to the parts layout and schematic diagram when installing the six resistors. Place the diodes on the circuit board and solder them in place on the board. Don't forget to cut the extra lead lengths.

Locate the semiconductor pin-out diagram shown in Figure 5-6. Now, go ahead and identify all three of the transistors for the project and install the transistors on the circuit board. There are two NPN types and one PNP type transistors in the AM radio project. Transistors usually have three leads, a Base lead, a Collector lead and an Emitter lead. The symbol for a transistor is a vertical line which is the Base lead and two lines running at 45°angles with respect to the Base lead. The angled lead with the arrow pointing towards or away from the Base lead is the Emitter lead and the remaining lead is the Collector lead. Be sure you can identify each of the transistor leads before placing them on the circuit board. Solder the transistors in place on

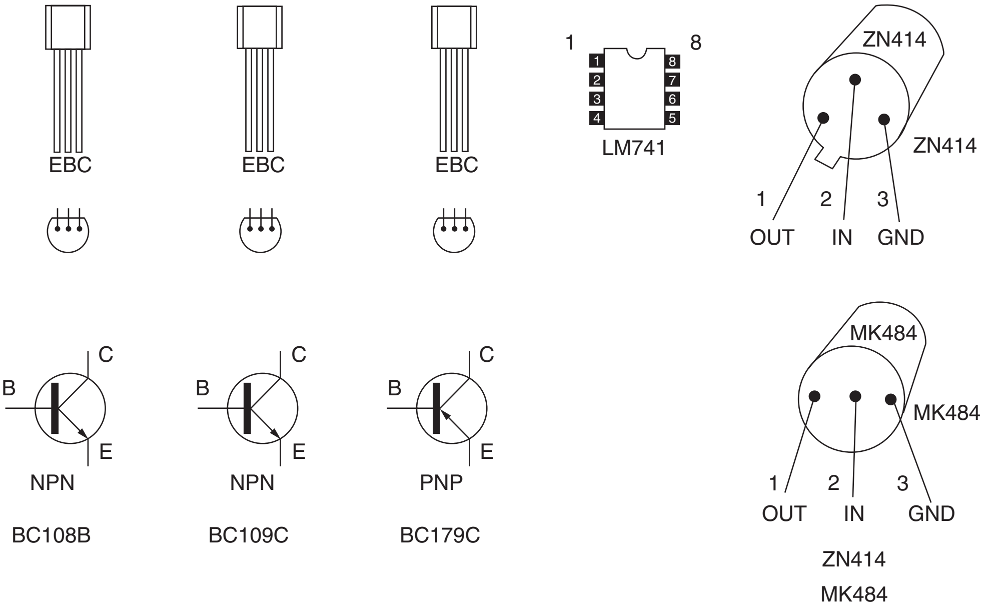

Figure 5-6 *Semiconductor pin-outs*

the circuit board; when finished remember to trim the excess component leads.

The ZN414 or MK484 TRF module looks very much like a common three-lead transistor. Carefully refer to the pin-out diagram for the MK484 before installing, in order to prevent damaging it upon power-up. One lead is the input lead from the tuned circuit. One lead is the ground lead and the remaining lead is the output lead which is directed to the actual circuit.

The AM radio project uses a single dual in-line integrated circuit, an LM471. We elected to use an IC socket as insurance against a possible circuit failure at some point in time. Integrated circuit sockets are inexpensive and make life easier in the event of a circuit problem, since it is much easier to unplug an IC rather than attempting to unsolder a multi-pin integrated circuit. Go ahead and solder-in a good quality IC socket for U2. Take a look at U2 and you will notice that at one end of the plastic package you will find a cut-out, notch or indented circle. Pin 1 of the IC will be just to the left of the notch or cut-out. Be sure to align pin of the IC to pin 1 of the IC socket, pin 1 of U2 is the null offset pin which is not used in this circuit. Note that pin 2 of U2 is the inverting input which is connected to the junction of R6 and R8.

Next locate the ferrite inductor coil or L1; you will notice that it is in parallel with the tuning capacitor C2. The parallel tuned circuit is placed between capacitor C1 and the input of U1, the MK484 TRF IC. These components are best placed at one end of the PC board if possible, or can be placed off the PC board on the front of your enclosure. Now, locate the volume control, the 10k ohm potentiometer; it is a chassis mounted control which can be mounted on the front panel of your enclosure.

Let’s wrap up the project. First you will want to solder the battery clip; note that the black wire of the battery clip will go to the circuit board minus or ground bus and the red battery clip wire will go in series with one post on the on-off switch at S1, the remaining post of the SPST toggle switch will connect to the Collector of Q1. Finally locate two 4″ pieces of #22 ga. insulated hookup wire. Bare both ends of the two wires, solder one end of each wire to the speaker leads. Now take one free wire end and solder it to the common ground of the circuit and then solder the remaining free wire end to the minus end of capacitor C8.

This completes the construction of the AM radio project. Let’s take a well deserved rest and when we return we will check the circuit board for possible “cold” solder joints and possible “short” circuits. Pick up the circuit board with the foil side facing upwards toward you. Look closely at the circuit board, first we will inspect the circuit board for possible “cold” solder joints. The solder joints should all look clean, shiny and bright. If any of the solder joints look dull, dark or “blobby” then you should remove the solder from that joint and re-solder the joint over again. Now pick up the circuit board once again and we will look for possible “short” circuits. “Short” circuits are usually caused from one of two reasons. The first cause of “short” circuits are loose solder “blobs” or solder balls which can sometimes stick to the circuit board. The second cause for “short” circuits is to have “stray” component leads stick to the foil side of the board. Rosin core solder residue can often be sticky and often solder “blobs” or “stray” component leads can bridge across circuit traces causing a “short” circuit. Look over the board and remove anything that looks like it might be bridging any circuit traces. Once the circuit inspection has occurred, you can prepare the circuit for its first test.

Connect up a fresh 9 volt battery to the AM radio circuit. Once the battery is connected to the circuit, you can “flip” the on-off switch to the on position and you should begin to hear some “hiss” from the speaker. If you do not hear anything, turn up the volume control to mid-position. At this point you should begin to hear some sounds from the speaker. Move the tuning capacitor control across the AM radio band and if all goes well you should begin to hear some local stations. Move the ferrite core coil or the circuit board around in a 360° pattern to see where the station is loudest. Hopefully the radio works fine without any further adjustments.

In the event that the AM radio circuit does not work at all, you will have to remove the battery from the circuit and inspect the circuit board one more time, but this time you will be looking for components placed incorrectly. Remember earlier, we talked about certain components have polarity such as diodes, electrolytic capacitors, transistors, etc. Now, you will need to inspect the PC board closely to see if you might have placed a component incorrectly. Its easy to make a mistake and insert a capacitor or diode incorrectly; it happens to professionals too! Another possible cause for failure is that you may have placed an IC incorrectly in the socket! Finally, it's easy to incorrectly install one of the transistors. Look at each transistor carefully, you may want to refer to the manufacturer's specification and/or pin-out information sheets along with the circuit layout and schematic diagram to confirm transistor pin-outs. Once you have inspected the circuit board for misplaced components and have corrected any mistakes, you can reapply power to the AM radio circuit and check out the circuit once again. Have fun listening to your new AM radio, you will have the great satisfaction of listening to the radio that you just built.

Chapter 6

Solid-State FM Broadcast Receiver

Parts list

FM Radio Receiver

R1,R15 4.7k ohm, ¼w, 5% resistor (yellow-violet-red)

R2,R18 1K resistor, ¼w, 5% resistor (brown-black-red)

R3,R16,R17 270 ohm, ¼w, 5% resistor (red-violet-brown)

R4,R11 47k ohm, ¼w, 5% resistor (yellow-violet-orange)

R5,R7,R8,R9,R14 10K, ¼w, 5% resistors (brown-black-orange)

R6 470 ohm, ¼w, 5% resistor (yellow-violet-brown)

R10 2.2k ohm, ¼w, 5% resistor (red-red-red)

R12,R13 10k ohm, potentiometers

R19 2 ohm, ¼w, 5% resistor (red-black-gold)

C1 4.7 or 5 pF, 35v disk capacitor (marked 5 or 5K)

C2,C3,C7,C8,C16,C18, .01 µF, 35v disk capacitor (marked .01 or 103 or 10 nF)

C4,5,15,27 .001 µF, 35v disk capacitor (marked .001 or 102 or 1 nF)

C6,C14 100 pF, 35v disk capacitor (marked 100, 101, or 101K)

C9,C17,C22,C26 4.7 or 10 µF, 35v electrolytic capacitor

C10,C13,C19 100 to 220 µF, 35v electrolytic capacitor

C11,C12 22 pF, 35v disk capacitor (marked 22)

C20,C21,C23,C24,C28,C29 .01 µF, 35v disk capacitor (marked .01 or 103 or 10 nF)

C25 .1 µF, 35v disk capacitor (marked .1 or 104)

FL1 10.7 MHz ceramic filter; Digi-Key Electronics

L1 slug tuned plastic molded coil

L2 10.7 MHz shielded can style IF transformer (421F122)

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

D1 varactor diode, 33 pF

D2 Zener diode, 6.2 volt

U1 ULN2111 or MC1357 IC

U2 LM386 IC

U3 SA602 IC

Q1,Q3 2N3904 NPN transistor

Q2 2SC2498 or 2SC2570A NPN transistor

B1 9 volt transistor radio battery

Misc PC board, battery holder, battery clip, RCA jacks

⅛″ earphone jack, wire, antenna

The FM broadcast receiver will permit you to listen to your favorite radio shows as well as allow you to tune outside of the standard FM broadcast band to “listen-in” to FM room bugs, FM mikes and other interesting hidden transmitters. The FM receiver could also be connected to an SCA adapter, so that you may listen to ‘MUSAK’ radio show used in stores and restaurants. The FM radio project is just plain fun and will teach you about FM radio theory.

The FM broadcast receiver demonstrates the basics of FM radio theory in one compact, simple circuit! Separate circuits are “sections” of the receiver, making it easy to visualize every function. The latest in IC circuitry is used in the FM receiver, which allows high performance.

The FM receiver project is designed to receive wide-band FM transmissions typical of FM broadcast stations, wireless mikes, and telephone transmitters or “bugs,” rather than Narrow Band FM transmissions (NBFM) as used on police or public service radios.

Why not build your own FM Broadcast receiver like the one shown in Figure 6-1. Not only will you learn how FM receiver works but you will be able to build and utilize your new FM radio. The FM receiver is great for listening to FM broadcasts and news as well as listening-in to commercial SCA—free music, stock quotes, sports, etc. The receiver has an SCA or sub-carrier output which can be fed into an SCA converter to allow this reception of these signals.

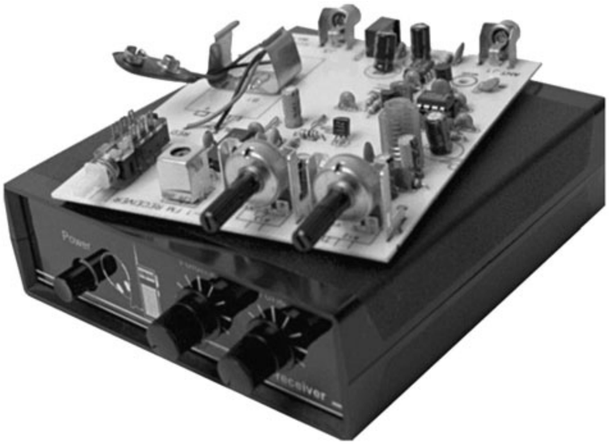
Figure 6-1 *FM receiver*

The receiver tunes from 70 to 110 MHz, ideal for tuning in out-of-band FM wireless mikes, “bugs” and other “hidden” transmitters. The FM receiver also features plenty of audio output, and operates from a standard 9 volt battery for truly portable operation.

FM modulation

Frequency modulation (FM) is a form of modulation which represents information as variations in the instantaneous frequency of a carrier wave. Contrast this with amplitude modulation, in which the amplitude of the carrier is varied while its frequency remains constant. In analog applications, the carrier frequency is varied in direct proportion to changes in the amplitude of an input signal. Digital data can be represented by shifting the carrier frequency among a set of discrete values, a technique known as frequency-shift keying. The diagram in Figure 6-2 illustrates the FM modulation scheme, the RF frequency is varied with the sound input rather than the amplitude.

FM is commonly used at VHF radio frequencies for high-fidelity broadcasts of music and speech, as in FM broadcasting. Normal (analog) TV sound is also broadcast using FM. A narrowband form is used for voice communications in commercial and amateur radio settings. The type of FM used in broadcast is generally

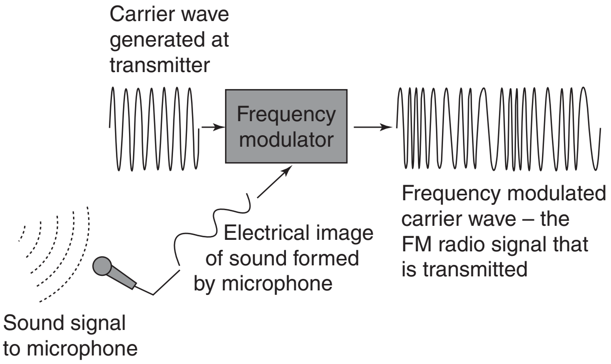
Carrier wave generated at transmitter
Frequency modulator
Frequency modulated carrier wave – the FM radio signal that is transmitted
Electrical image of sound formed by microphone
Sound signal to microphone

Figure 6-2 *FM modulation*

called wide-FM, or W-FM. In two-way radio, narrowband narrow-fm (N-FM) is used to conserve bandwidth. In addition, it is used to send signals into space.

Wideband FM (W-FM) requires a wider bandwidth than amplitude modulation by an equivalent modulating signal, but this also makes the signal more robust against noise and interference. Frequency modulation is also more robust against simple signal amplitude fading phenomena. As a result, FM was chosen as the modulation standard for high frequency, high fidelity radio transmission: hence the term "FM radio." FM broadcasting uses a well-known part of the VHF band between 88 and 108 MHz in the USA.

FM receivers inherently exhibit a phenomenon called capture, where the tuner is able to clearly receive the stronger of two stations being broadcast on the same frequency. Problematically, however, frequency drift or lack of selectivity may cause one station or signal to be suddenly overtaken by another on an adjacent channel. Frequency drift typically constituted a problem on very old or inexpensive receivers, while inadequate selectivity may plague any tuner. Frequency modulation is used on the FM broadcast band between 88 and 108 MHz as well as in the VHF and UHF bands for both public service and amateur radio operators.

FM receiver circuit description

The FM receiver block diagram in Figure 6-3 illustrates the fundamentals of FM reception. The schematic diagram of the FM broadcast receiver is depicted in Figure 6-4. FM broadcast signals captured by the whip antenna or from the input jack are applied to the RF amplifier Q2, a high gain, low noise microwave style transistor. Input signals are amplified about 100 times or equivalent to 20 dB. After being boosted, these input signals are routed to the SA602 mixer-oscillator chip for conversion down to the 10.7 MHz Intermediate Frequency (IF). The SA602 is a very popular chip used in a great variety of receivers from cellular telephones to satellite receivers. The reason for its popularity is that it contains most of a receiver's "front end" circuitry on a single, easy-to-use chip. Internal to the SA602 is an oscillator tuned by an external tank circuit; C11, 12, L1 and varactor diode D1. A varactor diode acts like a voltage variable capacitor and the voltage applied across it comes from tuning pot R12. This IF signal is an exact reproduction of the desired signal, but at 10.7 MHz. We then band-pass filter it with ceramic filter FL-1.

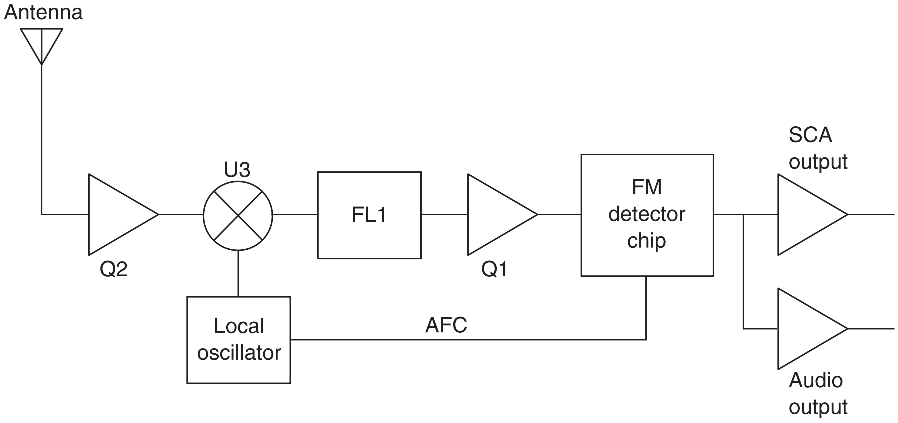
Antenna
U3
Q2
FL1
Q1
FM detector chip
SCA output
Audio output
Local oscillator
AFC

Figure 6-3 *FM broadcast receiver block diagram*

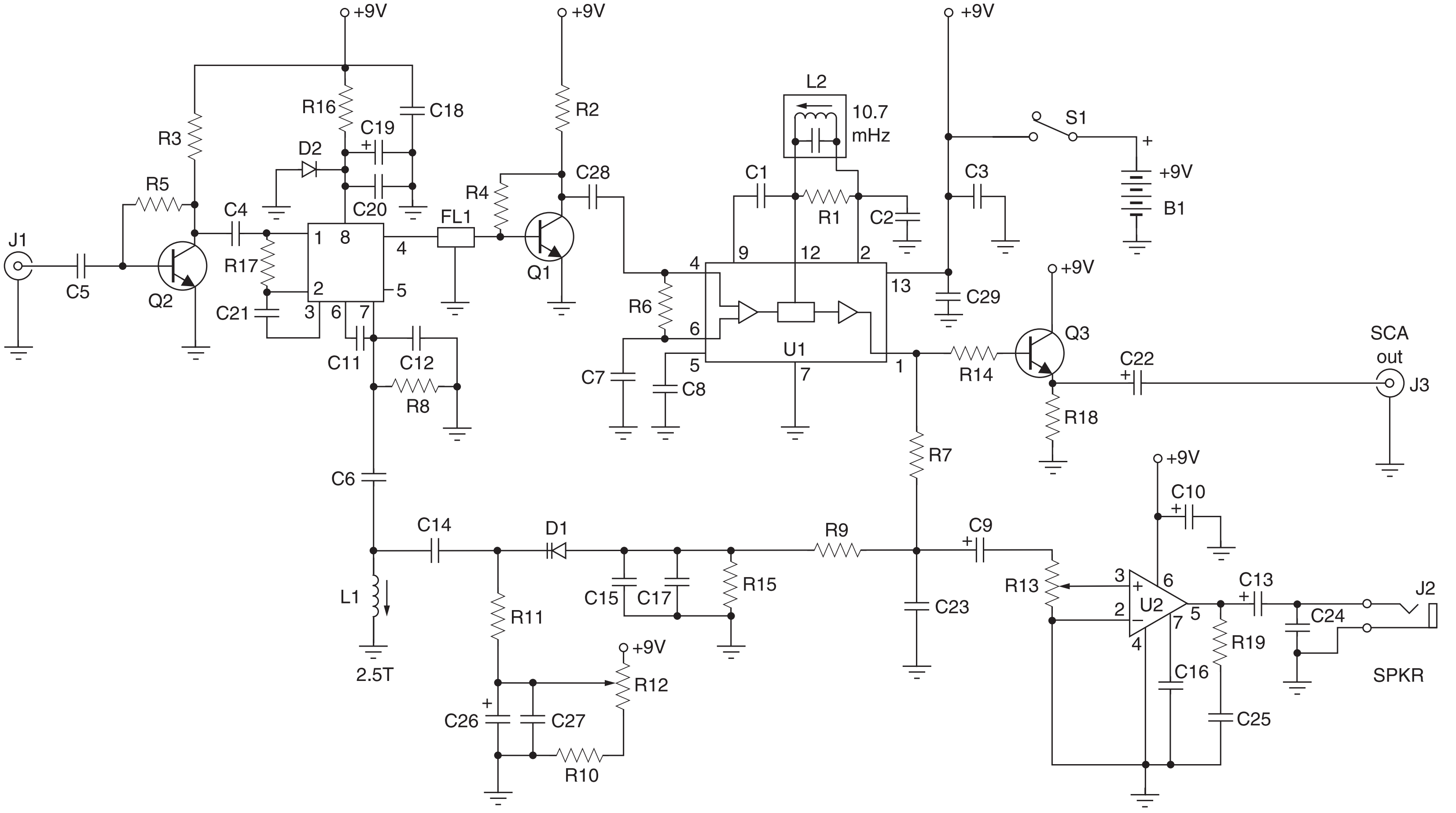
+9V
+9V
+9V
R16
C18
R2
L2
10.7
mHz
S1
+
R3
C19
+
D2
C28
C1
C3
+9V
R5
C20
R4
R1
C2
B1
J1
C4
1
8
FL1
4
9
12
2
R17
13
C29
+9V
C5
Q2
2
5
Q1
R6
6
Q3
C21
3
6
7
SCA
out
C11
C12
U1
C22
+
J3
5
1
C7
C8
7
R14
R8
R18
R7
C6
+9V
C10
+
C14
D1
R9
C9
+
C15
C17
R15
R13
3
+
6
L1
C13
+
J2
R11
2
U2
5
C23
−
7
C24
2.5T
4
R19
+9V
C16
R12
SPKR
+
C26
C27
C25
R10

Figure 6-4 *FM broadcast receiver. Courtesy of Ramsey Electronics*

The ceramic filter allows only the desired 10.7 MHz signal to pass, rejecting any adjacent channel signals. Transistor Q1 amplifies the IF signal before it is applied to the detector chip U1.

One of the primary advantages of FM is that the program information (modulation) is contained in carrier frequency variations and not in amplitude changes like AM, which would be susceptible to noise and interference. This means that we can limit or clip the signal, thus removing any noise and not affecting the desired information. The first section of U1 performs this limiting function, providing high gain and clipping any amplitude (noise) variations. FM is detected in the second section using a process known as quadrature detection. We'll not get into this too deeply because it requires a knowledge of trigonometry identities—let's just say it extracts the frequency modulated program information. Capacitor C1 and coil assembly L2 comprise a phase shift network that is required by the quadrature detector. Demodulated signals are amplified and sent out through pin 1 to the volume control, AFC circuitry and the SCA buffer amp. The AFC circuitry is used to properly keep the receiver locked on to the tuned-in radio station. Resistor R9 and capacitor C17 filter out any audio voltage variations so that only major frequency changes will shift the tuning voltage to lock the frequency.

Audio signals from the volume control are amplified to speaker levels by U2, an LM-386 audio amplifier chip. Transistor Q3 provides a buffer stage to external SCA devices you may connect up to the receiver. SCA units decode (demodulate, actually) "hidden" audio and data services transmitted over standard FM broadcast channels. These services include stock quotes, sports scores, talk shows and even background music—commercial free!

Circuit assembly

Look for a clean well lit area in which to assembly your shortwave receiver. You will need a 27 to 30 watt pencil tip soldering iron, some 60/40 tin/lead rosin core solder. You will also need a few hand tools including a pair of end-cutters, needle-nose pliers, a few screwdrivers and a pair of tweezers to pick up small parts. You will want to locate a small jar of "Tip Tinner" from your local Radio Shack store. "Tip Tinner" is a soldering iron tip cleaner/dresser which is invaluable when soldering circuit board. Locate the schematic and layout diagrams and we will begin to construct your new shortwave receiver. Place the project circuit board in front of you and we will begin building the receiver.

Locate all of the resistors for the project and place them in front of you. Each resistor has a color code which describes its value. Check the color code against the parts list and schematic before installing them onto the circuit board. Refer to the resistor color chart in Table 6-1. Place about four or five resistors on the PC board and place these resistors in their respective locations carefully making sure each on is in the correct location. Let your soldering iron tip heat both the component lead wire and PC board trace enough so that the wire itself and the foil trace both become hot enough together to melt a bit of solder so that it flows smoothly from the component leads to the PC board trace. Now, go ahead and solder each of the resistors onto the circuit board. Trim off the excess component leads with and small pair of end-cutters. Pick up and identify the next grouping of resistors and install them on the circuit board, follow up by soldering this next batch of resistors to the circuit board, remember to trim the extra lead lengths form the PC board.

Once all the resistors have been installed on to the circuit board, we can move on to identifying all of the capacitors for the project. Note that there will be different looking capacitors for this project, some will be small disks, while others will be larger body devices with minus and plus marking on them. These larger capacitors with polarity marking are electrolytic capacitors and they have polarity which must be observed for the circuit to work correctly. Look through the parts pile and locate the capacitors and place them in front of you. You will notice a number of small disk capacitors, we will install these first. Look closely at each of the disk capacitors at their markings. Sometimes the capacitors are very small and they may not have their actual value printed on them but will instead have some sort of code. Refer to the parts list for these codes or the capacitor code chart Table 6-2. Locate four or five of these small disk capacitors, identify them and install them on the circuit board, while referring to the schematic and layout diagrams.

Table 6-1

Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

After inserting the disk capacitors on to the circuit board, you can go ahead and solder them to the circuit board. Remember to cut the excess component leads. Next, move on and install another grouping of small disk capacitors, then solder them in place on the printed circuit board. Cut the extra component leads after soldering in the capacitors. When you are finished installing the small capacitors we will move on to the larger electrolytic types.

Next, we will identify and install the electrolytic capacitors. Electrolytic capacitors each have polarity markings on them in the form of a black band with either a plus or minus marking, which denotes the polarity of the device. You must pay particular attention to these markings in order for the FM receiver to work properly. Refer to the schematic and parts layout diagrams when installing C9, C17, C22 and C26, which are all 4.7 to 10 µF electrolytic capacitors. Install these four capacitors and solder them to the circuit board. Remember to trim the excess component leads. Cut the excess leads flush to the edge of the PC board with a pair of small end-cutters. Now locate and install electrolytic capacitors C10, C13 and C19, these are all 100 to 200 µF electrolytic types. After placing them on the PC board in their proper location and with respect to polarity marking you can go ahead and solder them to the PC board.

At this time, locate diodes D1, the varactor “tuning” diode and D2, a Zener diode from the parts pile. Install D1, varactor diode. It has a black body with green and red bands. Observe the correct positioning of the banded end by following the silkscreen. The green band is positioned as shown on the silkscreen. Now, install D2, the Zener diode. It has a gray painted glass body with a black band. Orient the band as shown on the layout diagram. A Zener diode acts as a voltage regulator, keeping the voltage across it constant. This keeps frequency drift to a minimum. Once both of the diodes have been placed on the PC board you can solder them in place and then trim the extra lead lengths flush to the edge of the circuit board.

Next, we are going to locate and install the two inductors; see Table 6-3. Locate the slug tuned plastic molded coil, which is coil L1. Once you have located it, you can place it on the PC; remember to check the schematic and parts layout diagram for placement details. Now, locate L2, the 10.7 MHz shielded metal quadrature coil. Notice that it will only go into the PC board one way. Place L2 on the board and now solder both L1 and L2 to the PC board. Trim the excess component leads from the board.

At this time, you can locate FL1, the 10.7 MHz ceramic IF filter. This part looks like a disk capacitor

Table 6-2
Capacitance code information

This table provides the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Some capacitors actually have the numeric values stamped on them; however, many are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

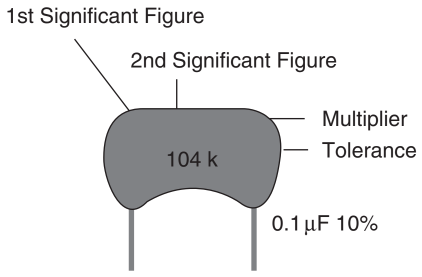

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

Table 6-3
Inductor coil winding information

L1	2½ turn coil #22 ga. insulated magnet wire on slug tuned coil form. Unwind original coil and rewind. 2½ turns onto Series 48 adjustable coil form 48A Series Coil (48A117MPC) Circuit Specialists.
L2	10.7 MHz shielded can style IF transformer (red) Part # (421F122) Mouser Electronics.

but has three leads. Amazingly, its function is equal to a small handful of inductors and capacitors hooked up in a bandpass filter configuration. For those techno-types, its bandwidth is 280 kHz centered at 10.7 MHz with an insertion loss of only 6 dB. Go ahead and solder FL1 in now.

Let's locate the transistors for the receiver now! Next, refer to the semiconductor pin-out diagram shown in Figure 6-5. Transistors are three semiconductors which generally have three leads: Base lead, a Collector lead and an Emitter lead. The transistor symbol shows that the Base lead is a vertical line. Both the Collector and Emitter leads are diagonal lines which point to the Base lead. The Emitter lead has a small arrow which points toward or away from the Base lead. Try and locate transistor Q1, a 2N3904 type. Watch placement of its flat side, refer to the parts layout diagram and the schematic for proper placement. Install Q2, 2SC2498 or 2SC2570A NPN transistor. Transistors are VERY critical of correct orientation of their leads. Be sure to pay close attention to the position of the flat side of the transistor. Finally, install Q3, 2N3904 transistor. Be sure to observe flat side orientation of the transistors when mounting them. Now go ahead and solder the transistors in place on the PC board.

Let's take a short break and when we return we will install the integrated circuits and the final components. Before installing the integrated circuits, please consider using integrated circuit sockets as an insurance policy against having to try and un-solder a 14 pin IC from the circuit board which is a difficult task even for the seasoned electronics technician. Integrated circuits DO fail and if one ever does, you can simply unplug it and replace it with a new one. Refer to the schematic and parts layout diagrams when installing the integrated circuits. Solder in the IC sockets in their respective locations and then we will install the actual ICs.

Integrated circuit packages will generally have a cut-out, notch or indented circle at the top end of the IC package. The cut-out or notch will usually be at the top center while if there is an indented circle it will be at the top left of the IC package. Pin 1 of the IC will always be left of the cut-out or notch, so you can easily identify the pin-outs. Using your schematic and parts

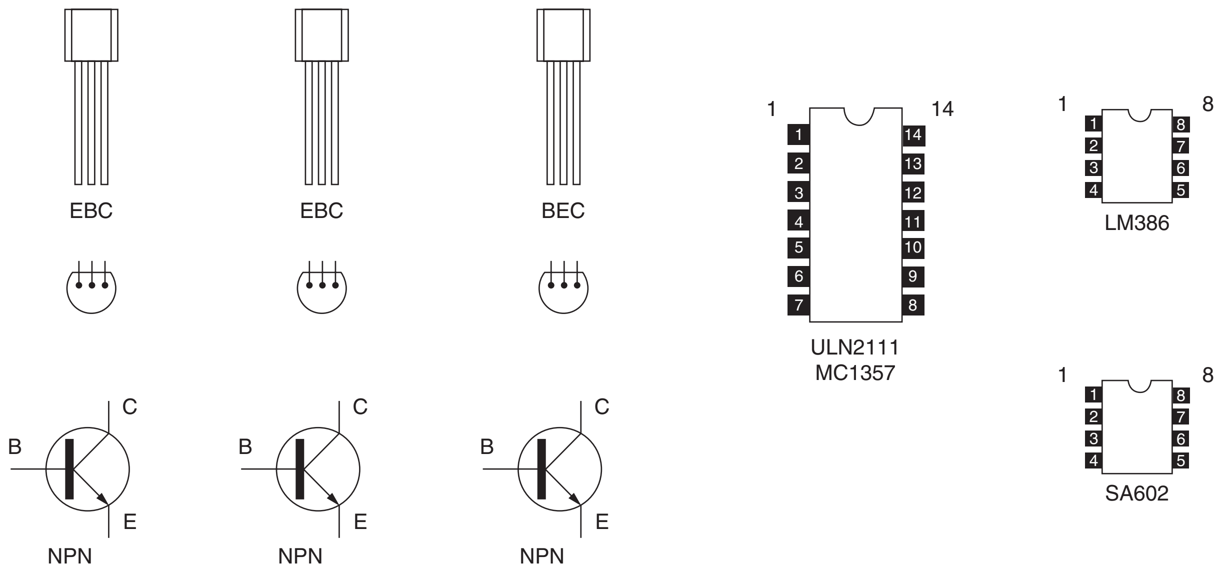

Figure 6-5 *Semiconductor pin-outs*

layout diagrams observe the correct orientation of the integrated circuits before inserting them into the sockets. First locate and install U1, the ULN2111 FM demodulator chip. Since this chip has 14 pins, be especially careful not to bend any pins under as you insert it into the PC board. When you solder the chip, verify that you are soldering all 14 pins. Next, let's install U2, the LM386 audio amplifier IC chip. The LM386 is a very popular part, and with good reason. Tucked inside that little plastic case are about 20 transistors that take a feeble audio signal and amplify it up to room-filling volume! Finally, go ahead and install U3, the SA602 IC mixer-oscillator. The marked end of the IC (band or dot) must face the edge of the PC board. If you wish, install an 8-pin DIP socket, still remembering to orient and install the SA602 correctly. Remember to firmly press the integrated circuits into their sockets.

The receiver project has two potentiometers, R12 is the "tuning" control, while R13 is the volume control. Locate the two 10k ohm potentiometers and install them, being sure to refer to the schematic and parts layout diagrams. Note that both potentiometers are on the front panel of the PC board. Install potentiometer R12, the tuning control. Insert it into the PC board so it seats firmly and is straight and even. Solder the three pins and the two mounting tabs. Next, install potentiometer R13, the volume control in the same manner.

Now locate S1, the on-off push button switch. Solder all six pins firmly. We've installed some "landmark" parts so far just to get us used to soldering our PC board. Finally, we are going to place the panel jacks on the rear portion of the circuit board. Locate and install J1, the RCA antenna jack, solder all four points. Now locate and install J2, the speaker jack, solder all three pins. At last, install J3, the SCA output jack, solder all four points.

There are only a few more components to go before we finish our FM receiver project. Install the 9 volt battery clamp. Use a scrap component lead wire looped through the holes in the clamp and into the holes in the PC board. Solder the wire to the clamp but don't use too much solder as this will prevent you from seating the battery fully into the clamp. Now, go ahead and install the 9 volt battery snap connector. Make sure you put the right color wire into the proper hole on the PC board. At long last, you are all done and ready to listen to tunes from your radio.

Before we apply power to the FM receiver for the first time we should make one last inspection of our work. First you will want to make sure that you have installed the diodes, electrolytic capacitors and the transistors with the correct orientation. Failure to do so may result in a damaged circuit, when power is first applied. Don't forget to check the orientation of the integrated circuits for proper location and orientation. Now we will inspect the PC board for "cold" solder joints. Pick up the PC board with the foil side facing upwards towards you. Take a look at all of the solder connections and make sure that they all look clean, smooth and shiny. If you find a solder joint that looks dirty, dark, dull or "blobby" then un-solder the connection, remove the solder and re-solder the connection once again until it looks good. Finally, we are going to inspect the PC board for possible "short" circuits. Often "cut" components leads can form "bridges" which can "short" out the circuit board traces, thus causing power to flow into the wrong places, and possibly damage the circuit. Also look carefully for any solder blobs that can form "bridges" between circuit traces. Now that you have inspected the PC circuit board, we can move on to testing and aligning your new receiver.

Testing and alignment

A hex head alignment tool is required to tune up your FM receiver. If you do not already have a set of plastic or nylon coil alignment tools and expect to build other radio or transmitter type kits, such tools are well worth having and can be found at any electronics store, including Radio Shack. While a hex style Allen wrench will fit the coil slug, the metal itself will drastically change the coil's inductance while it is in the slug. For this reason you should not use an Allen wrench unless you don't mind the trial and error process of turning the slug, withdrawing the tool, checking tuning, reinserting the tool, re-tuning, checking and so on.

You can fabricate your own non-metallic tuning tool with some patience, a bit of sandpaper and a piece of plastic rod. Try a large wooden matchstick, kebab skewer, or small plastic crochet needle. Carefully sand down the end to a hex shaped form and check for proper fit within the coil slug. It is better to have the tool fit loosely rather than too tight, since a tight tool will cause stress on the delicate slug.

Connect up a fresh 9 volt battery (preferably heavy duty alkaline type), a speaker or earphone, a short length of wire (about two feet) to serve as antenna, or the telescopic whip if the optional matching case and knob set was purchased. Now, set the volume and the tuning controls to the center of their range, then press the power switch "ON," you should hear some rushing noise from your speaker. Next, rotate the volume control and see if it does its job, then slowly turn L2's slug back and forth for maximum noise. This noise should be similar to what you would hear when a regular FM radio is tuned between stations. L2 may be tuned with a regular blade style screwdriver.

Try turning the tuning control, and see if you can receive any stations, but don't be alarmed if nothing is received since we haven't adjusted the oscillator coil yet. The oscillator coil, L1, is adjustable to receive from 70 to 110 MHz. We have to adjust it to cover the frequency range of interest. For alignment purposes we'll work with the understanding that the standard FM 88-108 MHz band is what we wish to receive.

Using an insulated alignment tool, slowly rotate L1's slug gently so that the slug is even with the top of the coil form. (Do not force the slug; if it binds, rotate the slug back and gently "rock it" to-and-fro as you would a thread forming tap. Slugs are by nature very brittle and will crack easily.) Now carefully rotate the slug into the coil form seven turns. The slug will be in about ¼″. Rotate the tune control until you hear an FM broadcast signal. Adjust L2 for best sound. Install R9, 10k ohm (brown, black, orange). R9 adds in AFC voltage to stabilize against drift. Without it your radio may "wander." After a signal is received, rotate the tune control to receive other stations. If the stations seem to "jump away" as you try to tune one in, the coil L1 is adjusted incorrectly and should be re-tuned. The reason for this odd phenomenon is that the AFC circuit is pushing the signal away rather than pulling it in. This is caused by having the oscillator 10.7 MHz below the desired signal rather than 10.7 MHz above it. A similar example of this is closing your eyes, hearing a sound, but not being able to determine if it is in front of, or behind you. Readjusting L1 will correct this situation. Make sure the slug is about three turns from the top of the coil form.

Calibrating the receiver is easy. First tune in a station known to operate at the low end (88 MHz) of the FM broadcast band. Rotate the tune control almost fully counter-clockwise to where that station should logically be located on the dial. Turn L1's slug until that same station is received. The tune control should now tune in the entire FM broadcast band. An easy way of making your own dial is to paste a small rectangle of paper behind the tuning dial knob. Then as you tune in different frequencies, pencil in on the paper the received frequency.

Additional receiver considerations

Your FM receiver was designed for clean, undistorted audio from average FM broadcast stations, while keeping battery current drain low. You may significantly increase the FM radio's audio output by adding a single 10 uF electrolytic capacitor across pins 1 and 8 on the LM-386 audio amplifier IC. The "+" side should connect to pin 1. An easy way of adding this capacitor is to simply "tack" solder it right on the IC pins on the solder side of the PC board. A 50-cent of speaker, lying naked on your workbench, will not give you a fair test of your FM receiver audio quality or volume. Speaker quality and well-designed enclosures have their clear purposes! A reasonable speaker in a box delivers a pleasant listening volume as well as significantly improved fidelity.

Out-of-band operation

Although many people use their receiver for comfortable FM broadcast reception, some James Bond types will want to use it to receive clandestine transmissions from phone bugs, room transmitters, etc. It's a simple matter to re-tune coil L1 to receive the frequency band of interest. First make sure you verify that your receiver receives the FM broadcast band correctly before re-tuning.

SCA output

Many FM broadcast stations now transmit sub-carriers on their signals. Sub-carriers are signals that are modulated on the carrier, just like the normal audio signals, except that they are too high in frequency to be heard. Normal audio signals range in frequency from 20 to 20,000 hertz. Most sub-carriers start at 56,000 hertz (56 kHz). These sub-carriers are themselves modulated, sometimes with audio signals, such as background music, but more and more these days with various forms of data. Some of the information carried on these sub-carrier data services include stock quotes, weather reports, news, sports and even paging signals. There is no limit to the variety of data that can be sent on a sub-carrier signal, and broadcasters are finding new things to send all the time!

The SCA output is a baseband demodulated sub-carrier signal output before any low pass filtering or de-emphasis. It provides a handy connection point for add-on SCA adapters such as the Ramsey SCA-1 kit. You can also connect this output to a receiver capable of receiving 30 to 75 kHz FM signals, or a PLL demodulator you've built yourself. Check out the data sheet from Signetics concerning the NE-565 PLL chip; an SCA demodulator is featured.

Chapter 7

Doerle Single Tube Super-Regenerative Radio Receiver

Parts list

Doerle Receiver – Original #19 Dual Triode

R1 2.2 megohm ½w, 10% resistor

R2 330k ohm ½w, 10% resistor

R3 47k ohm, ½w, 10% resistor

R4 100k ohm, potentiometer panel mount

R5 25 ohm reostat

C1 5–18 pF antenna trimmer capacitor

C2 140 pF main tuning capacitor

C3 .001 mF, 200 volt disk capacitor

C4 100 pF, 200 volt disk capacitor

C5 .1 mF, 200 volt disk capacitor

C6 .005 mF, 200 volt disk capacitor

L1 2.5 mH choke coil - Digi-Key Electronics

L2 “tickler” coil (see text)

L3 main tuning coil (see text)

T1 #19 triode tube – 2.5 volt filaments

B1 Type A filament batteries (6 volts)

B2 Type B battery 90 volt battery pack

HP 200 ohm high impedance headphones

Misc chassis, tuning dial, wire, sockets, standoffs, screws, etc.

Doerle Receiver – 6SN7 Version

R1 2.2 megohm ½w, 10% resistor

R2 1000 ohm, ½w, 10% resistor

R3 470k ohm potentiometer panel mount

R4 100k ohm, ½w 10% resistor

R5 100k ohm potentiometer

C1 8–18 pF trimmer capacitor

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

C2 100 pF main tuning capacitor

C3 50 pF variable tuning control (fine tune)

C4,C5 470 pF, 200 volt disk capacitor

C6 100 pF, 200 volt disk capacitor

C7 .01 mF, 200 volt disk capacitor

C8 .22 mF, 200 volt capacitor

L1 2.5 mH choke coil Digi-Key Electronics

L2 "tickler" coil (see text)

L3 main tuning coil (see text)

T1 6SN7 triode tube

B1 type A battery - 6.3 volt power supply or batteries

B2 type B battery - 150 volt power supply or batteries

HP 200 ohm high impedance headphones

Misc chassis, tuning dial, wire, standoffs, screws, etc.

Old Time Radio lovers will enjoy building this single tube regenerative AM/Shortwave receiver called the Doerle Radio, shown in Figures 7-1 and 7-2. It is a classic radio from around 1934, which was offered as a radio kit that a radio enthusiast could build at home. This very sensitive receiver is a fun-to-build radio which can be powered from batteries as they did in the "old days" or you could construct an AC power supply for it. The original Doerle radio receiver was offered as a kit for $3.00. The early Doerle receiver used a #19 dual triode receiving tube, one section of the tube acted as a regenerative detector and the second section of the tube acted as an audio amplifier. Doerle perfected this receiver design in 1933 when the new #19 tube became readily available.

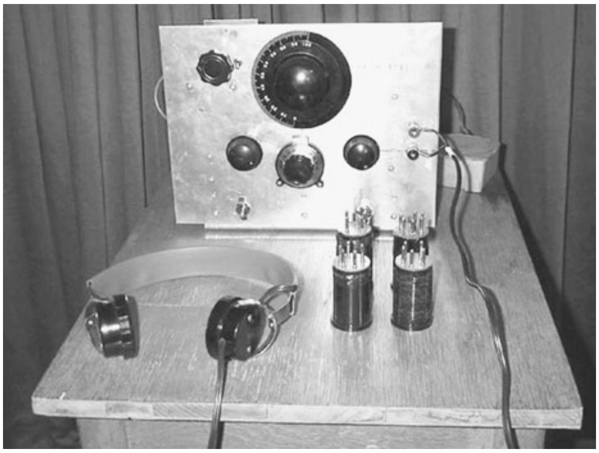

Figure 7-1 *Doerle receiver*

In this chapter we will build a copy of the original Doerle radio and we will also show you how to build the Doerle radio receiver using a more modern tube such as the 6SL7, which may be a lot easier to locate these days. If you are a "purist" you may want to try and locate the original #19 triode tube, which is still actually available. You can build either version, which you can construct to look just like the original radio looked back in the 1930s.

This radio is a study in simplicity, the low parts count kept the original cost down. A single sheet of metal, probably steel, painted in black wrinkle finish became a chassis. The single #19 tube powered by batteries eliminated unnecessary tubes and power supplies. The result is a simple, high-performance shortwave radio that beginners could afford to build and operate during the depths of the Great Depression.

First let's examine the circuit diagram, shown in Figure 7-3 for the original Doerle radio. The antenna is

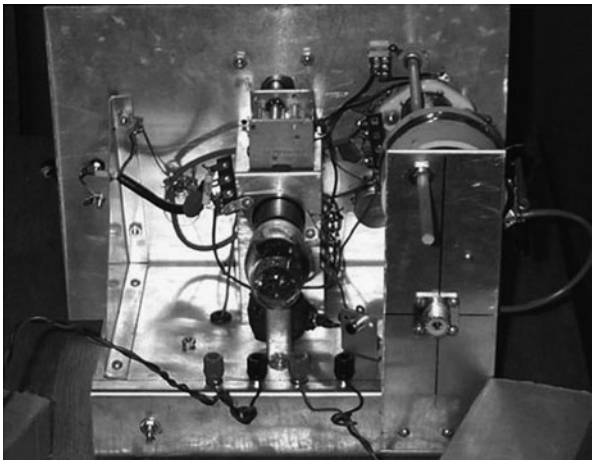

Figure 7-2 *Doerle receiver – inside view*

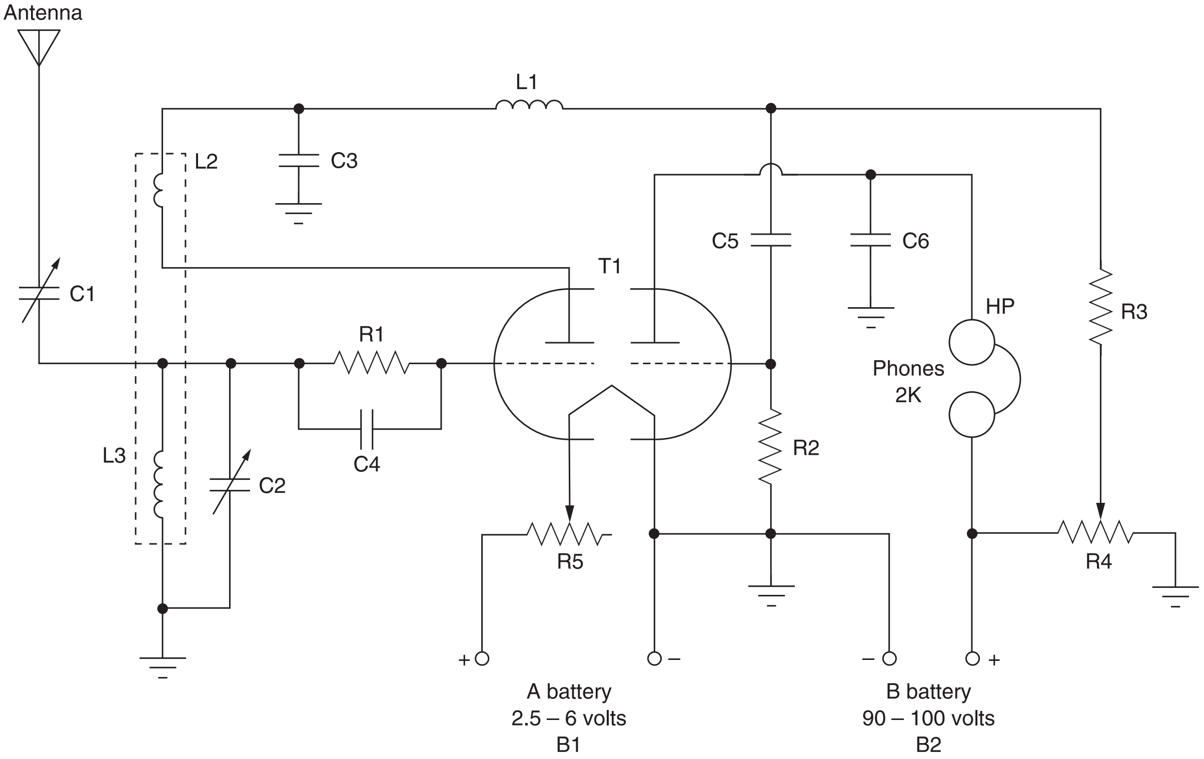

Figure 7-3 *Original #19 dual triode receiver*

coupled through a small 15 pfd screwdriver adjusted variable capacitor to the main tuning coil wound on a four pin form. A 10 pfd variable capacitor across it gives a fairly good trade-off between slow tuning rate and wide tuning range. The signal is fed through a grid leak configuration to one of the grids of the #19 tube. From the plate, the signal goes to the tickler wound on the same four pin coil form. The top side of the tickler coil connects to a 500 pfd capacitor and a 2.1 mH radio frequency choke (RF). The choke allows DC plate voltage into the tickler but prevents the radio frequency (RF) alternating current from getting into the rest of the circuit. The 500 pfd capacitor takes the RF to ground, thereby allowing RF voltage to build up on the tickler winding. Without this 500 pfd capacitor the receiver would probably not regenerate properly.

Variations in the DC voltage delivered to the detector plate feed through a .1 mfd capacitor to the second grid. A 250K, ½ watt resistor pulls the grid to DC ground, to provide the necessary zero bias for this class B amplifier tube, yet the resistance is high enough so that the AC avoids the ground and readily flows into the second grid.

Voltage on the detector plate is changed by a 50K resistor attached to the wiper of a 50K regeneration potentiometer that is connected from the B+ battery to ground. Plate current in the detector flows through these two resistors. Variations in current creates a varying voltage drop across this pair. The varying voltage is the audio signal that is applied to grid of the audio triode.

The B+ voltage source is fed through headphones to the second plate to complete the audio amplifier part of the circuit. The sound that is heard in the headphones comes from the variations in the audio plate current. A .006 mfd capacitor on the plate is large enough to take any stray RF to ground, but is not so large to take audio frequency (AF) voltage to ground. This capacitor aids receiver stability and audio quality.

A 20 ohm rheostat in series with one of the filament legs allows adjustment of the voltage. A rheostat is a low resistance, high wattage potentiometer. In the 1930s, you would have used two No. 6 1.5 volt dry cells (they were about 6″ tall and 2¾″ diameter) as your filament power supply. The rheostat would cut the voltage down to two volts needed by the tube's filament. As the voltage provided by the batteries dropped with use, you would adjust the rheostat to maintain the proper voltage on the filament. In other words, the rheostat allowed you to continue using the filament batteries as their voltage dropped to the point of total failure. Since the filament current for a #19 triode tube is just over 250 mA, so a 20 ohm resistor would allow you to use filament batteries up to 6 volts. And quite frequently that 6 volts was the battery borrowed from the family car.

Building the original Doerle receiver

The first step in building this or any project is getting the parts. In fact, this may be the most difficult step. When it comes to building a replica of an old radio like this, you just cannot go out and buy everything you need. You may have to adapt and use whatever you have on hand or can scavenge. For a project of this type you may want to build the circuit using point-to-point wiring between terminal and components rather than a circuit board to give the radio an "old-time" look. You are probably going to need both a soldering iron and a soldering gun to build this radio. You will also want to consult with the resistor color code chart in Table 7-1 and the capacitor code chart in Table 7-2 in order to identify the correct components.

Building this old radio can be lots of fun, but locating all of these old components can be difficult, and in fact you won't be able to locate the exact same items. You'll get something close. You'll have to mix and match whatever you can find. Most of these components are available at flea markets if you visit enough of them. Many of the components are available new from dealers, such as Antique Radio Supply, see Appendix. But often you'll have to pay substantially higher yet fair prices. The used #19 tube was purchased from a flea market dealer for $5. Yet, that same tube can be obtained brand new from Antique Electronic Supply for about $6, see Appendix. Again, you may have to scrounge, make do, adapt, experiment, and make a few mistakes. But the results are well worth it.

One idea is to build a prototype receiver with whatever components you can find, no matter how new or old, and get it working. Then, over the next few years, visit flea markets, antique radio meets, check out dealers' catalogs, and accumulate the older, more authentic parts. After you accumulate some of those rare, old, beautiful parts from the 20% and 30%, rebuild the receiver. And then a few years later rebuild it again. And every time you rebuild it, add older and more authentic parts, until you get an exact copy of the #19 Doerle receiver. The radio will become more and more of an authentic replica of a receiver from the early days of radio. With patience, practice, and plenty of enjoyment, you can build a fine radio.

Looking at the illustration of the Doerle-1 version, we see in the center a large, slow-motion dial drive

Table 7-1
Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

Table 7-2

Capacitance code information

This table provides the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Some capacitors actually have the numeric values stamped on them; however, many are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

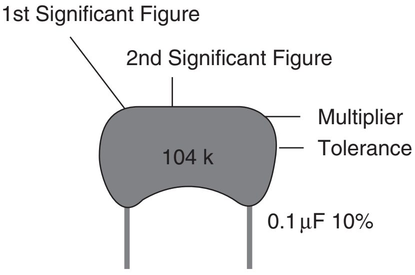

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 μF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 μF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 μF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 μF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 μF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 μF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 μF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 μF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

Table 7-3
Short-wave band listening frequencies

Shortwave Band	*Shortwave Band*
120 meters = 2300–2495 kHz	25 meters = 11.500–12.160 MHz
90 meters = 3200–3400 kHz	22 meters = 13.570–13.870 MHz
75 meters = 3900–4000 kHz	19 meters = 15.030–15.800 MHz
60 meters = 4750–5060 kHz	17 meters = 17.480–17.900 MHz
49 meters = 5730–6295 kHz	16 meters = 18.900–19.020 MHz
41 meters = 6890–6990 kHz	13 meters = 21.450–21.750 MHz
41 meters = 7100–7600 kHz	11 meters = 25.670–26.100 MHz
31 meters = 9250–9990 kHz	

friction-driven by a small knob at the bottom of the set. This allows fine tuning of stations. To the upper left are the antenna and ground terminals. Just to the right in a hole in the front panel is exposed the screw which adjusts the antenna coupling capacitor. Below left is the filament rheostat knob. At the upper right are the headphone terminals with the regeneration control located at the lower right. Although the receiver as shown in the original drawings may appear complex, it really is not. This is one of simplest receivers that you could build.

The original chassis for the Doerle receiver was an L-shaped piece of sheet metal, most likely steel, and painted with a black wrinkle finish. While the front panel was perfectly rectangular, the base was trapezoidal. Dimples were embossed into the base plate which served as feet.

For our replica Doerle receiver, you can start with an 11″ × 9″ sheet of .063″ (1⁄16″) aluminum to construct the chassis. The most complicated mechanical problems that you may encounter in building the receiver is the mounting #19 dual triode tube up on standoffs, mounting the variable capacitors and building the coils on the coil-form, see Figure 7-4.

The original Doerle model used a simple single bearing variable capacitor. We elected to use a heavy duty Hammarlund tuning capacitor; it is of significantly higher quality and is much heavier, so you will have to support the rear end of the capacitor with a bracket of some type. Once a Z-bracket as fabricated and attached to the breadboard, the prototype dial-drive-capacitor assembly was solid, yet turned smoothly.

Since a solid, rigid mount for the tuning capacitor is essential if you are to achieve frequency stability and eliminate microphonics, you must get this part of the assembly correct. And since your components will need their own custom mounting hardware, you'll need to experiment until you get it right. That's why the breadboard prototype is so useful. Take your time.

In continuing the capacitor assembly, I used a pair of ½″ long 6-32 machine screws in the top and bottom pair of holes on the drive mechanism. In the left-right pair of holes, I used a pair of 1½″ long 6-32 machine screws. The extra length became mounting studs to which an aluminum cross-arm was attached. In the cross-arm was drilled a ⅜″ hole to accommodate the mounting flange of the capacitor.

After the capacitor holes were drilled, the burrs removed, and edges of the panel were rounded and smoothed, I roughed up the surface with fine sandpaper.

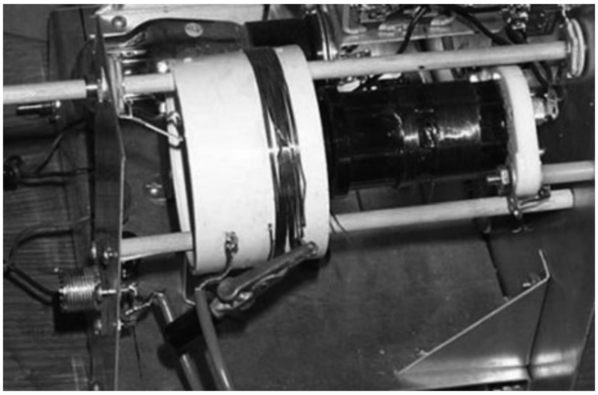

Figure 7-4 *Doerle coil assembly*

Next, the panel was bent into an L shape. You can use an old sheet metal brake, or you can could do the same by clamping the metal between a couple of 1 by 2's or lengths of angle iron held tightly together in a vise or with C-clamps. Since only a single, simple bend is required, you might get by clamping the sheet metal to a table top and by bending the sheet over the edge.

In order to obtain smooth tuning with a large range, you should locate some sort of planetary drive mechanism in front of the main tuning capacitor, such as a used National Velvet Vernier dial drive, which can be removed from an old radio. The first step was to dismantle it into the drive mechanism, the knob, and the dial plate. To mount the actual planetary drive mechanism, I had to cut a ⅝″ hole in the plate with a socket punch. Next, the drive collar was slipped through the ⅝″ hole. With a marker I marked the location of the four mounting holes.

In moving the capacitor and dial drive to the finished receiver, I chose to position the angle mounting bracket vertically as shown in the photos. This made fabrication and installation of a new base bracket much easier, and allowed more clearance for the tube and coil socket as well as for the terminal strip below the dial drive.

When the components are "test" mounted, you can move from the prototype stage to the finished L-shaped panel, if you are happy with the final layout. Then you can spray on a thick coat of fast drying metal primer. Once dry, I wet-sanded it under a stream of water until smooth. Then came the black wrinkle paint. If you wish to try and duplicate the original, you will want to try and create the wrinkle finish. If you are not that concerned about having the radio look exactly like the original radio, you can just paint the chassis with a single coat of black paint. Black Krylon paint is available in a spray can in your local home improvement store. For best results in creating a wrinkle finish, you must put down a heavy layer of paint. It must be heavy so that it can wrinkle up as it dries. If the coat is too light, or if you try to apply a second light coat over the existing layer, you won't get the best results. It might even be shiny where it is supposed to be flat black.

It is recommended that you paint one surface at a time and let it dry. Then reposition the chassis so that another surface is horizontal and then paint that. Of course, it will take four times as long to paint the radio than if you get in a hurry and paint the whole thing at once, but the results will be more predictable. You can bake it in an electric kitchen oven at the lowest possible heat. Nasty smelling solvents are baked out of the paint, so your wife or mother may not be too happy with you!

After mounting the tuning capacitor, the only other assembly of any complexity is the coil socket and antenna trimmer mount. Mounting the other components was fairly easily done. You can use machine screws, spacers of wood, plastic, porcelain, and metal. You can slice wood spacers from hardwood planks, but you'll need a table saw to do a good job. Plastic spacers can be cut from plastic tubing, but just be sure to get the rigid stuff used by plumbers for water supply lines. Porcelain spacers are something, which can be scavenged at a local hamfest or radio flea market. A small trimmer capacitor is mounted on porcelain standoffs in order to insulate both terminals from the chassis. Between the standoffs is a drilled ⅜″ hole in the front panel to allow screwdriver adjustment. That sounds simple enough. But behind the trimmer is the socket that accepts the four-pin lube socket coil form. What I used here were a couple of 2″ 6-32 machine screws cut to the appropriate length with heavy clippers. They were then fed through the coil socket mounting holes and through brake line spacers into the porcelain standoffs. Two soldering lugs were sandwiched between each metal spacer and porcelain standoff. One lug on each side was soldered to a capacitor lead. The other lug on each side became connector for the trimmer. Since the metal spacer pressed tightly against the capacitor lead, it was imperative that the four pin socket be made of insulating material. Here I used an old phenolic wafer socket. A metal socket would short out the capacitor. It sounds complicated but it should be fairly easily understood by studying the photographs.

Wiring was done with some modern plastic covered hookup wire, and with some genuinely old cloth-covered hookup wire salvaged from a piece of old tube gear. Watch for old tube gear with long lengths of cloth-covered wiring at flea markets. New old stock cloth covered wire is around but it is not always easy to find. Sometimes you can look for old wire while scavenging at hamfests.

The coils used in the original circuit were 1½″ dia. by 3″ tall phenolic coil forms. You can use phenolic bases from four prong tubes for the coil forms and they can be purchased from Antique Electronic Supply. The low frequency or main “short-wave” coil is close-wound with about 30 turns of #26 wire. You wind the “tickler” coil, which is 9 or10 turns of #24 ga. wire below the tuning coil, and both coils were covered with clear fingernail polish. The coil covers approximately 2.5 to 4.5 MHz. The exact number of turns you’ll need is going to depend on the tuning capacitor you use and the stray capacitance that results from your particular component layout, among other things. You’ll have to experiment to get the right number of turns.

The original Dorele receiver had plug-in coils for different radio bands. In order to construct an AM radio coil, you will have to wind another main coil along with another “tickler” coil on a second coil form. The main coil will need about 40 turns of #26 ga. wire, with a 9–10 turn “tickler” winding.

Battery terminals can also be purchased at a radio flea market. Using a table saw you could elect to slice a ¼″ strip from an oak 1×6. That gave me a mounting strip of ¾″ × 5 ½″ × 1⁄41″ thick, just right for this radio.

Firing up your original Doerle receiver

Now your #19 Doerle radio receiver is bolted together and wired up. It’s time to try it out. First, check the wiring, then check it again. It’s easy to make stupid wiring errors, sometimes after checking the wiring once or twice! Have a friend look over the circuit before applying power to the circuit for the first time. The coils used in the original drawing look like 1½″ dia. by 3″ tall coil forms. You can use phenolic bases from four prong tubes. They’re not from old tubes but from Antique Electronic Supply.

The low frequency coil is close-wound with about 30 turns of #26 wire. You can wind about 10 turns of tickler below the tuning coil, and cover both coils with clear fingernail polish. The coil covers approximately 2.5 to 4.5 MHz. The exact number of turns you’ll need is going to depend on the tuning capacitor you use and the stray capacitance that results from your particular component layout among other things. You’ll have to experiment to get the right number of turns.

The original ten turns of tickler turned out to be too many, so you may have to improvise to get it right. You’ll discover this problem very quickly if, when you advance the regeneration control, the receiver goes into a deafening howl. You may need to remove one turn of tickler coil at a time until regeneration starts up smoothly near the end of the control clockwise rotation. You should adjust the tickler with the tuning capacitor open, that is, at the receiver’s highest frequency. Here, more regeneration is necessary than at lower frequencies. And you must adjust the antenna trimmer control for the minimum capacity that gives excellent sensitivity. Adjusting the antenna trimmer “couples” or “de-couples” the connection between the receiver and the antenna, and can have a pronounced effect on regeneration adjustment and overall sensitivity.

The #19 tube filaments can be powered by alkaline flashlight batteries. Alkaline batteries should be used rather than cheaper zinc-carbon types since alkaline can deliver higher currents, and because they’ll last much longer. Alkaline batteries will give many more hours of use for your money. You can hook the filament batteries up to the set with the rheostat set to maximum resistance. You can then rotate the rheostat, with a voltmeter connected to the filament leads, in order to allow the meter to show two volts. Then put a reference dot on the panel with a marker pen. You can do something similar to the finished machine. As the batteries age, their voltage will drop making it necessary to adjust the control to maintain the two volts on the filaments.

If you do not want to use flashlight batteries, you may consider nickel-cadmium batteries. They’re inexpensive, each cell delivers about 1.2 volts, so you will need two, three or even four in series, which will deliver as much as 4.8 volts. And when they run down, you can recharge them. But before you go to that trouble: try flashlight batteries. You may be surprised by how long they last. As for a 90 volt B+ battery, you will most likely not be able to find them and if you do they will be expensive. Newark Electronics still had them, the last time I checked their catalog. It is easy enough to build a 90 volt “stack” of batteries. You can build a battery “pack” by buying ten 9 volt transistor radio batteries, and taping them together with electrical tape. When you do this, alternate their terminals so that you can easily wire the batteries in series.

The tube manuals lists the filament current at 260 mA. A voltmeter will show that the filament battery voltage drops about .2 volt over 15 to 20 minutes of operation. The batteries will last quite a long time. As for B+, a "stack" of 9 volt transistor radio batteries produced about 97 volts. With the receiver fired up and properly adjusted, the current draw was only 4 mA, significantly less than what a transistor radio battery is capable of delivering. After several minutes, the B+ battery dropped to 95.9 volts, so you can expect to get long life out of this battery pack. Several changes of filament batteries will be needed before the B+ battery dies.

So how does a Doerle radio receiver perform? For such a simple radio, it performs very well. Just a quick run through the band allowed me to copy amateur CW and marine Morse code. The foreign shortwave broadcasters came blasting-in.

Improvements? Yes, I can think of some. First, a volume control would be useful. I suspect that radio stations today run far more power than they did in the 30s. Therefore, the signals in the headphones are so loud as to be painful. A volume control would be useful to cut some signals down to size. I find the tuning a little too fast. Tuning a single sideband signal in is touchy and difficult in some cases. This could be fixed by going to five pin coils forms where the tuning capacitor could be attached to a tap in the tuning coil. Such a modification, however, changes the Doerle kit which was built for shortwave listeners, not amateurs. If you're a ham, you'll probably want to modify the design so that a ham band is spread out along the whole dial. And the last change that I would make is to improve the antenna trimmer adjustment. Adjusting the trimmer with a screwdriver can be painful after awhile. You don't do it very often, but when necessary it's not a convenient thing.

The 6SL7 Doerle receiver

In 1934, there was only one dual triode available for battery operation, the #19 dual triode! But the 1930s were years of explosive growth in the electronics industry. Just months after the #19 triode was released new metal tubes appeared, and in another ten years, seven and nine pin miniature tubes became available. These new series of tubes offered a variety of new dual triode tubes, many of which were better suited for the Doerle receiver circuit: see the 6SL7 Doerle receiver schematic shown in Figure 7-5.

If you elect to build this modified version of the Doerle, you will need to locate a 6SL7. The 6SL7 is a glass octal high-μ or (hi-gain) twin triode. The μ (mu) of a tube is the amplification factor, or voltage gain. Although the value of what is considered "high μ" has increased over the years, today "high" means a μ of from 60 to 100 or more. That means that the maximum voltage gain possible is 60 to 100, but usually it is significantly less.

High voltage gain tubes usually don't develop very much power. High power tubes usually have lower voltage gain. You can get both in the same tube. The #19 is an audio power tube with a relatively low μ. The 6SL7, on the other hand, has a μ of 70, but delivers little power. If you build both the #19 and 6SL7 versions, you'll immediately hear the increased headphone volume of the latter tube. Headphones don't need much power, so high voltage gain tubes works well, although more gain is usually better. A medium-μ tube like the #19 triode will give great performance. The 6SL7 was very common, and is still available. It is a higher-power, lower-voltage-gain tube with the same pin-out arrangement as the 6SL7.

The only major change introduced into this version of the Doerle is the mechanical layout. In this version, the prototype used an old National ACN dial drive, found at a local hamfest or flea-market, for a couple of dollars. I mounted it on an aluminum face place wrapped around a piece of 1 × 6 oak. The actual planetary drive is identical to that under the phenolic dial plates of the Velvet Vernier used in the model 19 Doerle. The ACN is much larger, 5″ × 7″ overall, and allows the user to mark a paper scale so that the dial becomes direct reading.

The antenna coupling trimmer was replaced by an 8-18 pF air variable stripped from an old radio. A custom bracket was snipped out of aluminum sheet and quickly bent up. By mounting the bracket on the wood base away from other meta components the capacitor electrically "floated" as needed. A knob allows the operator to reach around the side and easily adjust coupling when necessary. But you can take a similar small capacitor and cannibalize it by removing one or more plates to reduce the capacity. Just try get something less than 50 pfd for antenna coupling.

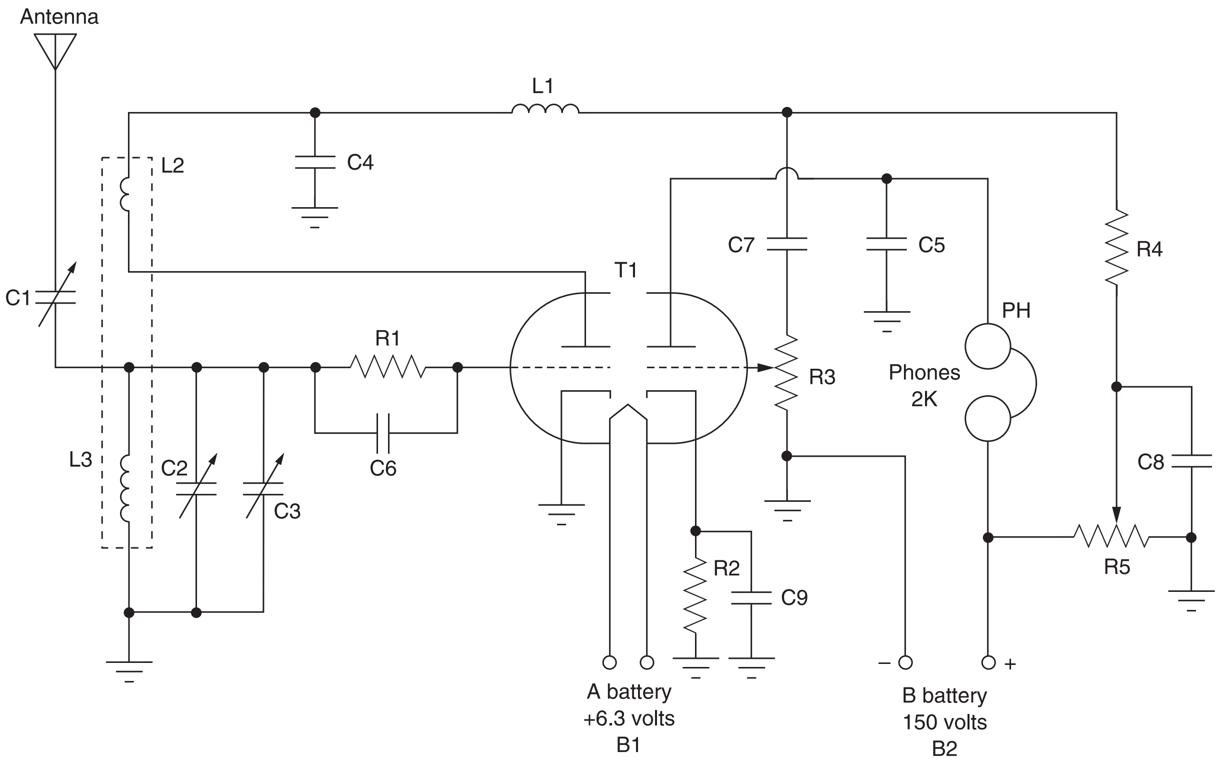

Figure 7-5 *New 6SL7 Doerle received*

A half megohm potentiometer replaced the 330 kilohm grid resistor used in the audio side of the original receiver, by connecting the panel-mounted pot to the rest of the circuit with miniature coaxial cable to reduce AC hum. In the original circuit, the audio grid was biased as before, but now has control over the volume. With a high gain tube like the 6SL7, it is important to reduce the audio output so that you don't destroy your hearing. A one kilohm resistor and a 10 mfd capacitor in parallel are inserted in the cathode lead of the audio triode to provide at least some negative bias to help the tube amplify with less distortion. You may want to increase or decrease the value of the resistor to achieve the necessary grid voltage specified in the tube manual. The voltage you measure across the cathode resistor is your grid bias.

If you compare the original circuit with this version, you'll find that almost all the same components can be used. The few differences are minor. And, again, don't be afraid to substitute. Just about anything within reason will work.

The biggest problem with this version of the Doerle receiver is that you will likely need a power supply, since you have to supply 6.3 volts AC into the filaments. Remember the original Doerle radio had 2.5 volt filaments, and you'll need about 150 volts for the plates on the 6SL7.

Does the 6SL7 version of the Doerle radio work? It sure does! You can use the same coils used in the original version of the Doerle radio for the 6SL7 version of the radio. When powering this version of the Doerle receiver be sure to set the volume control all the way down before applying power, since the audio output is much greater than that of the original radio. Put on a pair of headphones and turn on the 150 volts B+. Remember, you must use high impedance type headphones for this type of radio, not 8 ohm types. The Internet is a good source of the old 2000 ohm high impedance headphones. Bringing up the regeneration and audio controls brought in stations immediately. Many stations appeared across the dial, and they were surprisingly easy to tune in.

One important change in this circuit is the use of two capacitors in the main tuning LC circuit. In the circuit shown, you'll see a 50 pF fixed capacitor is parallel with a 100 pF variable capacitor. Since capacitance in parallel adds, the combination of the two essentially creates a virtual variable capacitor with a range of 50 to 150 pF. The net effect is to slow down the rate of tuning. The amount of band that can be covered by a single coil is reduced, but makes tuning far less “touchy.” This is important when tuning a sideband signal because very slight changes in tuning are needed to clear up the signal. You might want to replace the 50 pF with a 20–150 pF trimmer or something similar. By adjusting the trimmer, you can change the “speed” at which signals are tuned in. More capacitance reduces the speed, but it will also decrease the band of frequencies covered by the coil being used. You might want to replace the 50 pF fixed capacitor with a 140 pF for use as a band-set capacitor and a 35 pF for use as a band-spread capacitor.

Both should be equipped with slow-motion dial drives. You can use just about any dial drive that turns up at a flea market or in a radio parts catalog. You can mix and match components. Many times you’ll need more mechanical skills than you will need electronic skills for building old radios, so dig through your junk box, visit a hamfest and have fun building an old radio!

Chapter 8

IC Shortwave Radio Receiver

Parts list

Shortwave Receiver Project

R1,R5,R20 100k ohm, ¼w 5% resistor [brown-black-yellow]

R2,R4 22k ohm, ¼w, 5% resistor

R3,R7,R9,R10 1k ohm, ¼w 5% resistor [brown-black-red]

R13,R17,R18,R24 1k ohm, ¼w 5% resistor [brown-black-red]

R6,R8,R11,R19,R23 10k ohm, ¼w 5% resistor [brown-black-orange]

R12 1M ohm, ¼w 5% resistor [brown-black-green]

R14,R21 10K potentiometers, 103, or 10nF

R15 2 ohm, ¼w 5% resistor [red-black-gold]

R16,R22,R25 270 ohm, ¼w 5% resistor [red-violet-brown]

R26 100 ohm, ¼w 5% resistor [yellow-violet-orange]

C1,C3,C20,C21,C25 .001 µF, 35 vdc disk capacitor

C2 330 pF, 35 vdc mylar capacitor

C4,C6,C8,C28 .01 µF disk capacitors, 35 vdc (marked .01, 103, or 10nF)

C5,C17 100 pF disk capacitors, 35 vdc (marked 100, 101, or 100K)

C7,C10,C12,C18,C19,C29 1 µF ceramic disk capacitor, 35 vdc (marked .1 or 104)

C9,C24 1 µF, 35 vdc electrolytic capacitor

C11,C22,C23,C27 10 µF, 35 vdc electrolytic capacitor

C13,C14 220 µF, 35 vdc electrolytic capacitor

C15 (see text)

C16 (see text)

C26 1000 µF, 35 vdc electrolytic capacitor

C30 10 pF, 35 vdc mylar capacitor

L1,L2 1 µH molded type inductor (brown-black-gold)-Digi-Key

L3 12 µH molded type inductor (brown-red-black) -Digi-Key

T1,T2 shielded can inductor (42IF-103) Mouser Electronics

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

T3 shielded can inductor (42IF-123) Mouser Electronics

D1 1N270 diode, glass bead style

D2 Red LED (RSSI)

D3 Varactor diode, transistor style with 2-leads (MVAM108)

Q1,Q2,Q3 2N3904 NPN transistors

U1 LM358 8-pin DIP IC

U2 LM386 8-pin DIP IC

S1 DPDT PC-mount pushbutton switch

J1 RCA-type PC-mount jack

J2 subminiature $^1/_8$ inch PC-mount jack

B1 9 volt transistor radio battery

Misc PC board, battery holder, battery clip, wire, hardware, speaker

Have you ever wanted to listen-in into the fascinating world of radio? The shortwave radio project will allow you to listen to international stations such as the BBC, Radio Canada, Radio Moscow, Radio Sweden. Using just a few feet of wire as you antenna you will be able to reach out and hear the world. Shortwave listening is fun for folks of all ages from 8 to 80 years of age.

The shortwave radio project shown in Figure 8-1 covers the shortwave bands from 4 to 10 MHz. The receiver uses varactor diode for smooth tuning across the band. The superheterodyne shortwave receiver provides both great sensitivity and selectivity. The tuning control as well as the volume and RF gain controls are all located on the front panel for easy operation. The shortwave receiver is powered from a common 9 volt transistor radio battery, so the receiver can be taken virtually anywhere. The receiver would make an ideal scout, school or club project.

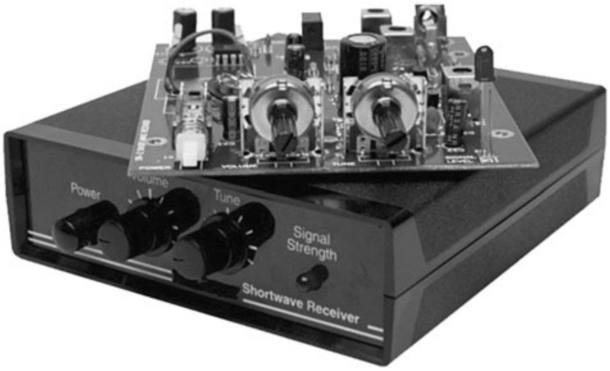

Figure 8-1 *Shortwave radio*

You'll easily tune into broadcasts from many other countries as well. As you become more and more familiar with the world of shortwave broadcasting, you'll be deciding on your own favorite band. You will hear a variety of other "interesting" sounds, but just remember that this receiver is designed for AM only. If a Morse code signal really sounds "good," it is because it is being transmitted in AM tone-modulated form, or perhaps the signal is so close to an AM broadcast carrier that the carrier acts as a "beat-frequency-oscillator" (BFO). Even though this receiver will permit you to tune through several different ham radio bands, the signals are not likely to be intelligible. Reception of CW and SSB signals on an AM receiver requires a BFO. This is not a complicated feature, but it is not a feature of this receiver. Our companion receivers designed for the ham bands will let you tune into these SSB and CW broadcasts.

Shortwave listening as a hobby in itself

Many people worldwide enjoy listening to shortwave broadcasts of all kinds, and they keep written records of what they hear. Almost every nation on Earth has some sort of shortwave broadcast service, though many are much more challenging to tune than the powerful signals of Radio Moscow and the BBC. In addition, these "SWLs" (Shortwave Listeners) listen to ham operators, government and commercial stations and even clandestine operations. Some shortwave listeners enjoy collecting QSL cards from stations which they have logged. Shortwave listening is, for some, a step toward getting a ham radio license. For others, it is a great hobby in itself.

The Shortwave Receiver is a good introductory receiver for this hobby. After you decide exactly what kinds of listening are of the most interest to you, you'll be in a better position to choose a more elaborate receiver.

To learn more about this SWL hobby, look for a copy of *Popular Communications* at newsstands. An inexpensive and interesting general introduction to all kinds of radio listening is the book, *Shortwave Listening Guide* by William Barden, Jr. (1987; Radio Shack Catalog Number 62-1084). To learn more about the hobby of ham radio, write to ARRL (American Radio Relay League), 225 Main Street, Newington, CT 06111.

Circuit description

Take a moment and examine the shortwave receiver block diagram in Figure 8-2. The simplified signal flow of the block diagram shows the basic sections of the receiver. The corresponding components are noted under each main block and can be cross-referenced to the schematic.

The main shortwave receiver diagram is illustrated in Figure 8-3. The receiver circuit begins at the antenna input at J1. RF signals (Fc = carrier frequency) from the antenna are applied to the RF input and filtering allowing only the signals of interest to pass through. The high pass filter helps eliminate unwanted signals picked up by the antenna, so improving the overall reception quality of the radio. After the input signal is filtered, it moves to the mixer stage. Notice on the diagram that there are two inputs to the mixer. We have discussed one of these input signals coming into the mixer but not the other. Now, look down at the local oscillator block. The local oscillator (LO for short) acts as your tuning control for what frequencies you can receive by generating a signal on the board close in value to that which will be used by the mixer. There is a direct relationship between the generated frequency of the local oscillator (LO) and the exact receive frequency (Fc) you want to listen to. This will become clear when we finish discussing the block diagram. The LO section is a Colpitts oscillator that takes advantage of smooth varactor diode tuning. The varactor (D3) forms an L/C (inductor/capacitor) tank circuit with T3. Increasing the voltage on the varactor diode with R21 (tuning pot) increases the capacitance of D3, thus increasing the frequency output of the LO section.

Now that we know the two signals coming into the mixer stage, both the Fc (receive carrier frequency) and the LO (generated local oscillator), we can better cover its operation. The mixer takes these input signals and performs a few very basic operations. The technical explanation of how the mixer combines these

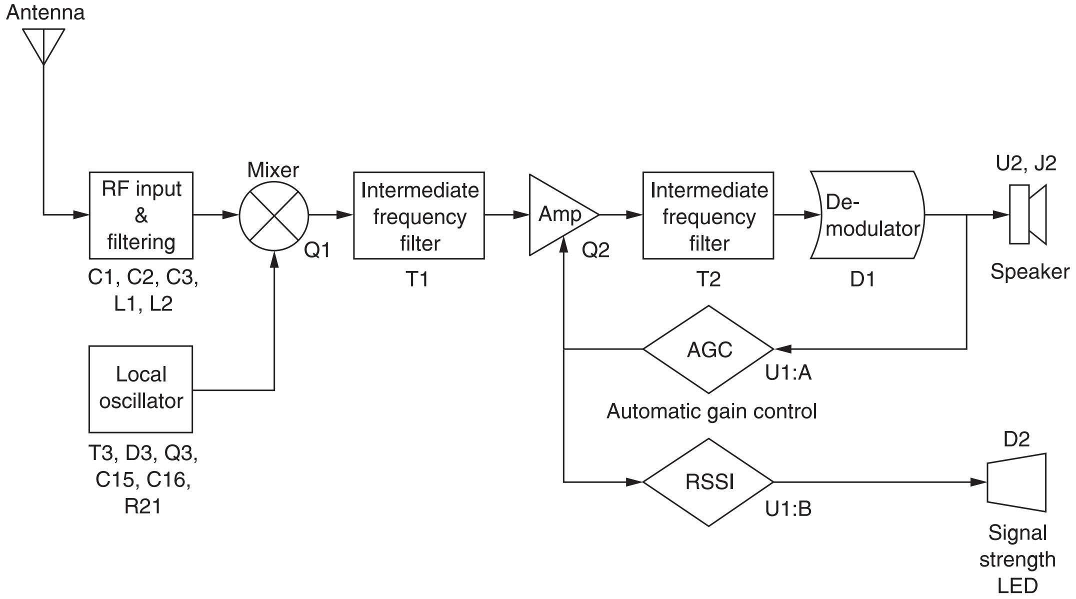

Figure 8-2 *Shortwave receiver block diagram*

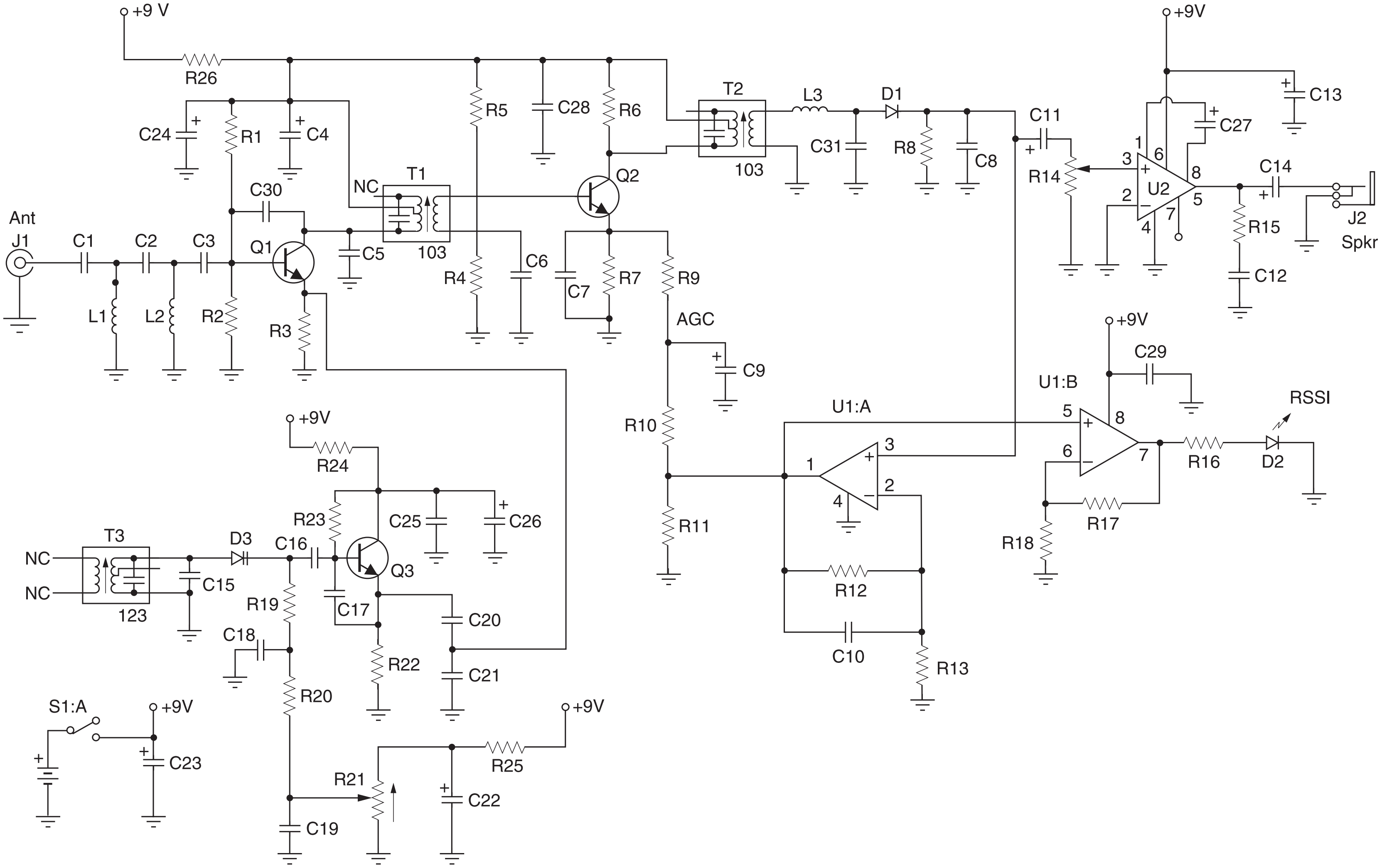
+9 V
R26
C24
R1
C4
R5
C28
R6
T2
L3
D1
+9V
C13
C27
C31
R8
C8
103
C11
R14
U2
C14
NC
T1
Q2
C30
Ant
J1
C1
C2
C3
Q1
C5
103
C6
C7
R7
R9
R4
L1
L2
R2
R3
AGC
C9
R15
J2
Spkr
C12
+9V
C29
U1:B
RSSI
U1:A
R10
R16
D2
R17
R18
R11
R12
C10
R13
+9V
R24
R23
C25
C26
T3
D3
C16
Q3
NC
NC
C15
123
R19
C17
C20
C18
R22
C21
R20
S1:A
+9V
C23
+9V
R21
R25
C22
C19

Figure 8-3 *Shortwave receiver schematic. Courtesy of Ramsey Electronics*

signals through Fourier Series is interesting but very drawn out. The point of using the block diagram, however, is to simplify matters. Therefore, the function of the mixer is to obtain the "product" and the "sum" of the input signals. This means you take the input signals and merely "add" and "subtract" their values to determine what you get on the output. The general formulas to use are quite simply, "Fc + LO = Mixer Output" and "Fc – LO = Mixer Output."

The output of the mixer stage unfortunately has other signals besides the ones we want. This brings us to the next stage, Intermediate Frequency Filtering. An Intermediate Frequency (IF) is a signal somewhere between the RF signal Fc we started with and the final audio message we are trying to get. The desired IF we are dealing with is a fixed number, such as 455 kHz. As long as the proper relationship between Fc and the LO exists, the IF value will stay constant. Due to the fact that the IF frequency stays the same all the time, the Intermediate Frequency Filter can be very narrow. The filter will remove any other signals coming from the mixer that are not in the proper pass band and yield a clean signal for further processing.

The next stage is an amplifier with an adjustable gain feedback loop. The gain control is dependent on the amount of signal being received (this Automatic Gain Control "AGC" will be covered in a moment). The amplifier boosts the signal level of the incoming IF and gives us a stronger signal to work with. After the amplifier stage is another Intermediate Frequency Filter. This helps remove any unwanted residual signals still present and cleans up the amplified IF for a high quality signal. At this point the audio signal we are trying to obtain is riding on the IF signal.

The demodulator circuit finally extracts the message from the IF section through a process called envelope detection. Now that we have our message back in the audio realm, it is directed through the audio circuitry to the speaker output. The demodulated audio branches off before the audio circuitry and is used to perform some useful functions. The RSSI LED, Received Signal Strength Indicator Light Emitting Diode, gives us a general signal level feature. The stronger the signal we receive, the brighter the LED will glow. This is great for help pulling in those weak transglobal transmitter signals. The gain of the RSSI circuit is controlled by the value the feedback resistor R17. If you are commonly using a small whip antenna and listening to weak signals, you may need to increase the value of R17 to customize the response indication of D2, the signal strength LED.

The final stage, and the real secret to the sensitivity of the receiver, is the Automatic Gain Control (AGC). The gain control looks at the amount of signal level present at the output of the demodulator and varies the amount of gain the amplifier has accordingly. If there is a strong signal coming through the demodulator the AGC circuit lowers the gain of the amplifier. If the received signal coming through the demodulator is very weak, the AGC circuit increases the gain of the amplifier (Q2) allowing us to receive signals from around the world and listen to them with clarity.

Receiver assembly

Look for a clean well lit area in which to assemble your shortwave receiver. You will need a 27 to 30 watt pencil tip soldering iron, some 60/40 tin/lead rosin core solder, a few hand tools including a pair of end-cutters, a needle-nose pliers, a few screwdrivers and a pair of tweezers to pick up small parts. Locate your schematic and layout diagrams and we will begin to construct your new shortwave receiver. Let your soldering iron tip heat both the component lead wire and PC board trace enough so that the wire itself AND the foil trace BOTH become hot enough TOGETHER to melt a bit of solder so that it flows smoothly from the pin to the PC board trace. Locate the printed circuit board and all the components and we will start.

Before we begin constructing the shortwave receiver, refer to the resistor color code chart in Table 8-1 and the capacitor code chart in Table 8-2. Each resistor will have three or four color bands on them. The first color band will start close to one edge of the resistor body. The first color band will represent the first digit of the resistor value. The second color band will represent the second digit value, while the third color band illustrates

Table 8-1
Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

the resistor multiplier value. The fourth band, if there is one, will represent the resistor's tolerance value. A silver band is a 10% tolerance, while a gold band depicts a 5% tolerance value. No fourth color band notes a resistor with a 20% tolerance. Locate all of the resistors for the project and place them in front of you. Each resistor has a color code which describes its value. Check the color code against the parts list and schematic before installing them onto the circuit board. Place about four or five resistors on the PC board and place these resistors in their respective locations carefully making sure each one is in the correct location. Next solder each of the resistors onto the circuit board. Trim off the excess component leads with a small pair of end-cutters. Pick up and identify the next grouping of resistors and install them on the circuit board, follow up by soldering this next batch of resistors to the circuit board, remember to trim the extra lead lengths form the PC board.

Once all the resistors have been installed onto the circuit board we can move on to identifying all of the capacitors for the project. Note that there will be different-looking capacitors for this project, some will be small disks, while others will be larger body devices with minus and plus marking on them. These larger capacitors with polarity marking are electrolytic capacitors and they have polarity which must be observed for the circuit to work correctly. Look through the parts pile and locate the capacitors and place them in front of you. You will notice a number of small disk capacitors, we will install these first. Look closely at each of the disk capacitors at their markings. Sometimes the capacitors are very small and they may not have their actual value printed on them but will instead have some sort of code. Refer to the parts list for these codes. Locate four or five of these small disk capacitors, identify them and install them on the circuit board, while referring to the schematic and layout diagrams. After inserting the disk capacitors onto the circuit board, you can go ahead and solder them to the circuit board. Remember to cut the excess component leads. Next move on and install another grouping of small disk capacitors, then solder them in place on the printed circuit board. Cut the extra component leads after soldering in the capacitors. When you are finished installing the small capacitors we will move on to the larger electrolytic types.

Capacitors C9 and C24 are 1.0 µF electrolytic capacitors and capacitors, while C11, C22, C23, and C27 are 10 µF electrolytic capacitors. Capacitors C13 and C14 are 220 µF, while C26 is a 1000 µF electrolytic capacitor. When installing these capacitors you must orient them with respect to their polarity. Each electrolytic will have either a plus or minus marking

Table 8-2

Capacitance code information

This table provides the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Some capacitors actually have the numeric values stamped on them; however, many are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

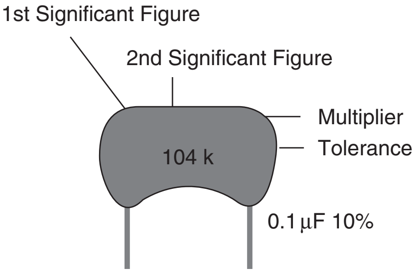

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

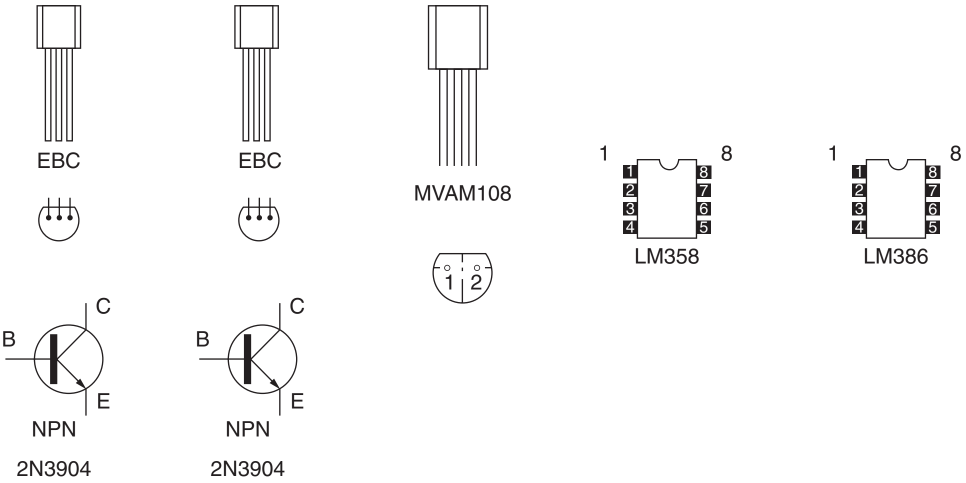

Figure 8-4 *Semiconductor pin-outs*

on them. You must look at the schematic and layout diagram and observe the polarity marking on the diagram match up with the markings on the capacitor when installing them to ensure that the circuit will work. Failure to observe polarity marking may result in damage to the circuit when it is first powered up. Identify these electrolytic capacitors and install on the PC board. Solder them in place and then cut the extra lead lengths flush to the edge of the circuit board.

Take a short break and locate the semiconductor pin-out diagram, shown in Figure 8-4, which illustrates the output pins of the transistors and integrated circuits in the receiver. Next, let's move on to identifying and installing the two diodes; note that diode D2 is the varactor or tuning diode and D1 is a glass bead style diode, while D3 is the red LED RSSI indicator. Diodes have marking on them that denote polarity and it must be observed for the circuit to work correctly. You will notice that a diode will have a small black or white colored band at one end of the diode body. This band denotes the diode's cathode. You will remember that a diode's symbol looks like a triangle pointing to a line. Well, the triangle is the anode and the line corresponds to the black or white band on the diode. Refer to the schematic when installing the two diodes and make sure you have installed them in the correct PC locations before soldering them in place. The RSSI LED is mounted on the front of the PC board near the potentiometers.

Now locate the three NPN transistors and we will install them on the PC board. Remember that transistors each will most likely have three leads, a Base lead, a Collector lead and an Emitter lead. The symbol for a transistor is a vertical line, with two other leads, at 45° angles, pointing to this vertical line. One of these angled leads will have a small arrow pointing towards or away from the vertical line, this lead is the Emitter lead. The vertical lead is the Base lead and the remaining lead is the Collector. Locate and identify these four transistors and install them on the PC board. Be sure you can identify each lead and that each lead goes to the correct PC hole on the board; refer to the layout diagram and schematic before soldering the transistors to the printed circuit board. Don't forget to trim the excess lead from the transistors after the transistors have been soldered to the board.

Now, it's time for a little "destruction"! (If you jumped ahead and installed T3, we've got bad news for you.) Before this, two shielded transformers can be installed, and the internal capacitors need to be removed. Looking at the underside of these two transformers, you'll see a tubular part, probably white with a brown band, somewhat like a resistor. These are brittle and easily crushed with any sharp object that can be pressed against them with mild force (small nail, small screwdriver). You'll find these capacitors will easily disintegrate into particles. DO NOT crush the capacitors in the other two inductors. After crushing the capacitor, install T3. Next locate the remaining inductor cans, i.e. T1 and T2 and install them on the PC board; remember orientation of the inductor cans is important. Now locate the small "resistor like" inductor, i.e. L1, L2 and L3. Solder the inductors to the circuit board and remember to trim the excess leads.

Finally we are going to install the integrated circuits to the circuit board. Consider installing IC sockets, instead of mounting the IC's directly to the PC board. Integrated circuit sockets are good insurance in the event of a possible circuit failure at some later date. It is much easier to unplug an IC rather than trying to unsolder one from a PC board. Mount three IC sockets to the PC boards at the respective locations of U1 and U2. Before installing the ICs we will need to make sure that you can identify the pin-outs on each of the ICs. Have a look at the eight pin integrated circuits, note that at the top of each of them you will see a notch in the center of the device or a small indented circle at the left side at the top of the package. The IC pin to the left of the notch or indented circle is pin one (1). Refer to the schematic and layout diagram to observe where pin one (1) is for each of the integrated circuits before installing them to the PC board. Once you are sure of the pin-outs for each of the ICs and where they go on the PC board, you can insert them into the proper IC socket. If there are any jumpers on the circuit board, you can go ahead and install them now.

Well, we have made quite a bit of progress on building your shortwave receiver. Take a short break and when we return we will mount the remaining final components. At this time look through the remaining parts and look for the two chassis mounted potentiometers R1 and R2. Place the potentiometes at the front of the PC board and solder them in place. Now, locate and install the power on-off switch at S1, and solder it in place on the printed circuit board.

Finally locate and install J1, an RCA chassis jack and J2, the ⅛″ phone jack onto the PC board. Locate the battery clip and solder it to the PC board, where you see the plus and minus markings. The red battery clip wire will go the plus marking on the board, while the black wire will go to the minus marking on the board. Install the battery clamp. Position battery and holder so as not to cover nearby PC board mounting holes. Use the method for securing the clamp that is most convenient for you, such as a (1) wire looped through clamp and PC board holes and soldered, (2) small screws, (3) double-faced adhesive strips or (4) hot-melt glue.

Your shortwave broadcast receiver is now almost finished. Now, you will need to decide what frequency range you would like your receiver to tune, and then select the values for C15 and C16 from the chart, shown in Table 8-3. A suggested configuration would be C15 = 47 pF and C16 = 100 pF. This will give you complete coverage from 6 to 11 MHz by adjusting T3 in or out. This is a very active section of the shortwave band and will provide you with hours of listening enjoyment any time of the day. Select and install C15 and C16.

Table 8-3
Band selection components

C16 = .001	**T3 Slug In**	**T3 Slug out**
C15	6.5–13 MHz	9–18.3 MHz
C15	5.1–7.9 MHz	7.1–10.7 MHz
C15	4.3–5.3 MHz	6–8.3 MHz
C16 = 100 pF	**T3 Slug in**	**T3 Slug out**
C15	8.4–14.2 MHz	11.5–19.5 MHz
C15	6–8.1 MHz	8.5–11 MHz
C15	4.9–6.3 MHz	6.7–8.4 MHz

Finally, you will need to locate an enclosure for the receiver, and then install the circuit board into the enclosure. Place the knobs on the potentiometer shafts, locate a 9 volt transistor radio battery and connect it to the receiver battery clips. Connect up an antenna and you are ready to test out your new receiver.

Before turning on your receiver, double-check that you have correctly oriented all three of the ICs, that you have correctly oriented the transistors, that you have installed the diodes correctly, and that you have installed the electrolytic capacitors correctly.

Finally we will inspect the circuit board for "cold" solder joints and possible "short" circuits. Pickup the PC board with the foil side of the board facing you and inspect the solder connections. You will want to make sure that all of the solder connections look clean, smooth and shiny. If any of the solder joints look dull, dirty or "blobby" then you should un-solder that particular joint and re-solder it over again so that it looks clean, smooth and shiny. Next, let's inspect the CP board for any possible "short" circuits. Often "stray" cut component leads will "bridge" across between circuit traces and "short" out a PC trace. Another possible source of "short" circuits is a solder blob which can "bridge" across a circuit trace. Carefully inspect the PC board for any solder blobs or "stray" wires before connecting up a 9 volt battery.

Initial testing and adjustment

Now, connect up a speaker or headphones, connect up some sort of antenna and install a fresh 9-volt battery and we can begin testing the receiver. We have come to the "moment-of-truth," go ahead and turn the receiver power switch to the "on" position. After adjusting the volume to a pleasant level, you should hear some shortwave stations by turning the Tune Control, no matter how any of the adjustable coils happen to be set. While listening to any kind of station, whether broadcast or Teletype, etc., use a small screwdriver to adjust the slugs in transformers T1 and T2 for the best-sounding reception. The Tuning Control covers varying segments of the bands selected by adjustment of oscillator coil T3. Adjustment of T3 anywhere between the full In position to the full Out position will give the user full range between the minimum and maximum frequency coverage set by C15 and C16.

Both T1 and T2 must be adjusted with a non-metallic alignment tool such as is used in radio-TV service. If you do not have one, a suitable tool can be made by patiently sanding a screwdriver-like blade on the end of a wooden match stick, kebab skewer or small plastic crochet needle. Again, please be aware that a metal screwdriver blade will drastically increase the coil inductance and make adjustment quite difficult. T1 and T2 are simply adjusted for strongest reception of any signal range that is tuned in.

If you do not have any sort of testing or frequency reference equipment whatsoever, the easiest way to begin using the receiver, is to set the Tune Control at its midpoint, slowly tune T3 with your alignment tool as though it were a tuning dial. Stop when you come into the middle of a group or cluster of foreign broadcast stations. Try tuning around these stations with the Tune Control. If you like what you hear, readjust both T1 and T2 for best reception. Eventually, you will get a clue as to what general frequency band you are hearing, because many stations periodically announce their frequencies, particularly at sign-on and sign-off times.

If you like precision, use a frequency counter or calibrated receiver to find the SR2's strong oscillator signal, remembering that there is a 455 kHz IF difference (above or below) between the local oscillator frequency and the broadcast signal you are hearing.

Shortwave antenna ideas

The type of antenna you'll want to use for your shortwave receiver depends on the degree of interest you have in shortwave listening, on whether you are limited to an indoor or balcony antenna. If you cannot use an outdoor antenna, you can make yourself a whip antenna from telescoping radio antenna from your local Radio Shack store. A better antenna would be to throw a 10 to 20 foot piece of insulated wire out the window, secured to a pole or tree away from your window. You could also place a length of wire in your attic to act as an antenna. An even better antenna would be a dipole antenna "cut" for the desired radio band of most interest. In order to construct your own dipole antenna, you can refer to the diagram depicted in Figure 8-5. Simply use the formula: 468 divided by the frequency in Megahertz for example. If you wanted to construct a dipole for 7.410 MHz, you would divide 468 by 7.410. The results would be 63.157 feet, so one-half of that number is 31.5 feet. Each leg of the dipole antenna would then be 31.5 feet. If you do construct your own dipole antenna, you should keep the antenna away from metal objects and use insulated wire and be sure to keep

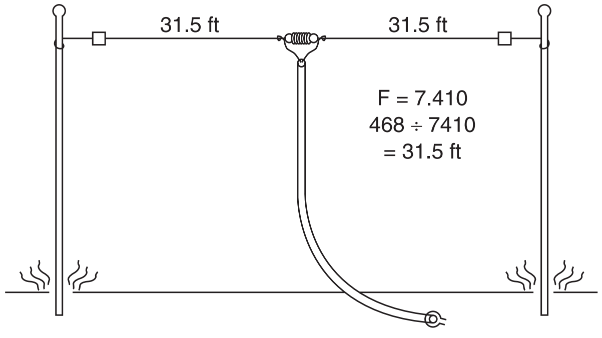

Figure 8-5 *Shortwave dipole antenna*

away from all power lines. The lead-in wire for your dipole antenna to your new receiver can be a length of RG8X min-coaxial cable.

Shortwave listening

Shortwave listening is a fun hobby for young and old alike. Once you have your antenna all set up, it is time to begin listening to your new radio. The chart in Table 8-4 illustrates the most common shortwave bands and their shortwave frequency in Megahertz vs. their wavelength in meters. Many times the shortwave bands are referred to in meters, such as the 40 or 80 meter band, so this chart will help you see the relationship between the two. As a shortwave radio listener, you will want to know what International broadcasters are broadcasting when you have time to listen to your new radio. Most people have spare time in the evening for their hobby. So you will want to know which bands are best for night time use, and who is broadcasting at that time. Refer to the chart in Table 8-5, which illustrates the broadcaster and their frequencies, etc. Usually the best shortwave bands, for evening use are the lower frequency bands such as the 40, 41 and 49 meter bands. Use this chart or check on the Internet for the latest shortwave listings. Have fun!

Table 8-4
Popular shortwave bands

Frequency vs. Meters

Broadcast Frequency – kHz/MHz	*Frequency in Meters*
2300–2495 kHz	120 Meters
3200–3400 kHz	90 Meters
3900–4000 kHz	75 Meters
4750–5060 kHz	60 Meters
5850–6200 kHz	49 Meters
7100–7350 kHz	41 Meters
9400–9900 kHz	31 Meters
11600–12050 kHz	25 Meters
13570–13800 kHz	22 Meters
15100–15800 kHz	19 Meters
17480–17900 kHz	16 Meters
18900–19020 kHz	15 Meters
21450–21850 kHz	13 Meters
25600–26100 kHz	11 Meters

Troubleshooting tips

If you experience difficulty, think of your receiver as consisting of several sections or stages: oscillator-mixer, IF audio and AGC, and final audio output (LM386). The first step in case of problems is to make sure that the tunable oscillator is working, which can be done by listening for its signal on another receiver. After the oscillator circuit is confirmed as working, standard signal tracing procedures should isolate any problem, which will be either an incorrectly installed part, a defective part or a bad solder connection. Correct orientation or polarity of all diodes, transistors, electrolytic capacitors and ICs is essential.

Problem: Strong shortwave broadcast audible throughout tuning range.

Solution: This can occur if your antenna is “too good” or if the RF Gain control is turned up too high. The high sensitivity of the front end is designed for simple antennas, with most reception quite satisfactory with only 25 feet or so of wire.

Problem: Local AM radio station audible throughout tuning range.

Solution: Whether this will even happen depends on how close you are to a local AM station. It is very important that all component leads be as short as possible, since just a bit of wire can help D1 and the several stages of audio amplification give you a free, unwanted “classic crystal radio.” A grounded metal case for the SR2 is one possible solution.

Table 8-5

Evening shortwave broadcast listings

English shortwave broadcasts to the Americas sorted by time:

1700–1800 NA	U.S.A.	WYFR	13695, 17555
1700–1800 NA, LA	Japan	R. Japan	9535
1700–1900 Am	U.S.A.	WHRI	9840, 15285, 15650
1700–2200 NA	U.S.A.	WBCQ	18910
1705–1905 Am	Canada	R. Canada Int'l	9610
1705–1905 NA	Canada	R. Canada Int'l	9800 drm
1800–1900 NA	U.S.A.	WYFR	13695, 17535, 17555
1800–1900 NA, Su-F	U.S.A.	WBCQ	7415
1830–1845 NA, Eu	Israel	Israel B/C Auth.	6985, 7545, 9345
1900–2000 NA	U.S.A.	WYFR	13695, 17535, 17555
1900–2000 LA	U.S.A.	WYFR	6085
1900–2000 Am	U.S.A.	WHRI	9840, 13760, 15285
1900–2100 NA	U.S.A.	KJES	15385
1900–2100 NA, Sa-Su	Netherlands	R. Netherlands	15315bo, 15525bo, 17725ca
1900–2200 NA	U.S.A.	WWCR4	9975
1900–2400 NA	U.S.A.	WBCQ	7415
2000–2025 NA, Eu, Af	Israel	Israel B/C Auth.	6280, 7545, 15640
2000–2100 Am	U.S.A.	WHRI	11765, 15285
2000–2100 LA	U.S.A.	WYFR	17575
2000–2200 NA	U.S.A.	WYFR	17535, 17555
2000–2400 NA	Costa Rica Univ.	Network	13750
2030–2130 NA, Eu	Cuba	R. Havana Cuba	9505, 11760
2045–2130 Am	Vatican City	Vatican R.	9800 drm
2100–2127 Eu, As, NA	Czech Republic	R. Prague	5930, 9430
2100–2130 NA, Sa	Canada	CBCNQ	9625
2100–2200 NA, Eu, Af	U.S.A.	WEWN	6890, 15785
2100–2200 NA	Japan	R. Japan	17825
2100–2200 LA	United Kingdom	BBC	9480sc, 11675gr
2100–2200 Am, Sa-Su	U.S.A.	WRMI	7385
2100–2200 Am	U.S.A.	WHRI	9660, 11765
2100–2300 NA	U.S.A.	WWCR1	7465
2100–2400 Am, M-F	U.S.A.	WRMI	7385
2130–2156 NA, Eu	Romania	R. Romania Int'l	6055, 6115, 7145, 9755
2130–2200 NA, M-Sa	Canada	CBCNQ	9625
2130–2200 NA	Netherlands	R. Netherlands	9800ca drm
2200–2257 NA	Netherlands	R. Netherlands	15425fg drm
2200–2300 NA, M-F	U.S.A.	WBCQ	5110
2200–2300 LA	United Kingdom	BBC	5975fg, 9480sc
2200–2300 NA	Canada	R. Canada Int'l	9800 drm
2200–2300 Am	U.S.A.	WHRI	7490, 9660

Table 8-5

Evening shortwave broadcast listings—cont'd

English shortwave broadcasts to the Americas sorted by time:

2200–2300 NA, M-F	U.S.A.	WBCQ	18910
2200–2300 NA, Sa-Su	U.S.A.	WWRB	3185
2200–2400 LA	Guyana	V. of Guyana	3290
2200–2400 NA, Su-F	Canada	CBCNQ	9625
2200–2400 Am	Anguilla	Wld Univ Network	6090
2200–2400 NA	U.S.A.	WWCR3	5070
2200–2400 NA	U.S.A.	WWCR4	9985
2200–2400 NA, Eu, Af	U.S.A.	WEWN	7560, 9975
2200–2400 NA	U.S.A.	WWRB Overcomer	6890
2230–2257 NA, Af	Czech Republic	R. Prague	5930, 9435
2300–2330 NA	Germany	Deutsche Welle	9800ca drm
2300–2350 NA	Turkey	V. of Turkey	5960
2300–2356 NA, Eu	Romania	R. Romania Int'l	6055, 6115, 7105, 9610
2300–2400 NA	Egypt	R. Cairo	11885
2300–2400 NA	China	China R. Int'l	6040, 11970
2300–2400 NA	U.S.A.	WBCQ	5110
2300–2400 Am	U.S.A.	WINB	9265
2300–2400 NA	Cuba	R. Havana Cuba	9550
2300–2400 NA, LA	China	China R. Int'l	5990ca
2300–2400 NA	U.S.A.	WHRA	5850
2300–2400 Am	U.S.A.	WHRI	7315, 7490
2300–2400 LA	U.S.A.	WYFR	15170, 15400
2300–2400 NA	U.S.A.	WWCR1	3215
2300–2400 NA	U.S.A.	WWRB	3185, 5050
2300–2400 LA, Su-F	U.S.A.	WWRB	5745
2330–2357 NA	Czech Republic	R. Prague	5930, 7345
2330–2359 NA	Sweden	R. Sweden	9800ca drm
2330–2400 NA	Lithuania	R. Vilnius	7325
2330–2400 Am, Su	U.S.A.	WRMI	9955
2335–2400 LA, Su-M	Austria	Osterreich 1	9870
2343–2358 LA, Tu-Sa	Austria	Osterreich 1	9870

Relay site codes: ae-United Arab Emirates, al-Albania, an-Antigua, ar-Armenia, as-Ascension, au-Austria, bo-Bonaire Neth. Antilles, bu-Bulgaria, ca-Sackville Canada, ch-China, cl-Chile, cu-Cuba, cy-Cyprus, de-Delano USA, fg-French Guiana, fr-France, ga-Gabon, ge-Germany, gr-Greenville USA, it-Italy, ja-Japan, ka-Kazakhstan, ko-South Korea, la-Latvia, li-Lithuania, ma-Madagascar, ml-Mali, mo-Moldova, ne-Netherlands, om-Oman, po-Sines Portugal, ru-Russia, rw-Rwanda, sc-Cypress Creek SC, se-Seychelles, si-Singapore, sl-Sri Lanka, so-Slovakia, sp-Spain, sw-Sweden, ta-Taiwan, th-Thailand, uk-United Kingdom, uz-Uzbekistan, wb-WBCQ USA, wr-WRMI USA, wy-WYFR USA, za-South Africa.

Notes – Days of week: Su-Sunday, M-Monday, Tu-Tuesday, W-Wednesday, Th-Thursday, F-Friday, Sa-Saturday, exW-except Wednesday.
Target areas: Af-Africa, Am-America, As-Asia, Eu-Europe, LA-Latin America, ME-Middle East, NA-North America, Oc-Oceania/Australia; Other: alt-alternate frequency, drm-Digital Radio Mondiale, occ-occasional use, se-Special English.

Courtesy of Daniel Sampson; www.primetimeshortwave.com

Chapter 9

80/40 Meter Code Practice Receiver

Parts list

80/40 Meter Direct Conversion Receiver Specifications

Sensitivity 3-uV or better

Modes CW, RTTY (FSK), SSB, AM

Audio output 50 mW typical

Frequency coverage Portions of the 80, 75, and 40 meter bands

Power Source 9 volt transistor battery, alkaline preferred

PC Board 4.000″ × 4.700″

80/40 Meter Direct Conversion Receiver

R2 15 ohm, $^{1}/_{4}$w, 5% resistor (brown, green, black)

R3,R8 100 ohm, $^{1}/_{4}$w, 5% resistor (brown, black, brown)

R1 1000 ohm variable potentiometer

C1 8.2 pF ceramic disk (8.2)

C18 100 pF ceramic trimmer capacitor

C19 15-50 pF variable capacitor (main tuning) - see text

C2,C6,C7 .1 μF monolithic (.1 or 104)

C3 see chart below

C4 not used

C5 see chart below

C8,C9,C10 .1 μF monolithic (.1 or 104)

C20 .01 μF ceramic disk (.01 or 103)

C12,C13,C14 1 μF, 35 volt electrolytic capacitor

C16,C17 470 μF, 35 volt electrolytic capacitor

C11,C15 10 μF, 35 volt electrolytic capacitor

U1 SA602 RF Mixer IC, 8-pin DIP

U2 78L05 voltage regulator

U3 LM386 audio amplifier, 8-pin DIP

S1 Power On-off switch

J1 RCA antenna jack

J2 $^{1}/_{8}$″ headphone jack

B1 9 volt transistor radio battery

Misc PC board, IC sockets, chassis, knobs, wire, battery

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

holder, battery clip, hardware, etc.

80 Meter Band Tuning components

C3,C5 470 monolithic capacitor (471)

L1 6.8 µH variable inductor, shielded

L2 3.3 µH molded choke (orange-orange-gold)

L3 33 µH molded choke (orange-orange-black)

40 Meter Band Turning Components

C3 470 pF monolithic (471)

C5 680 pF monolithic (681)

L1 1.5 µH variable inductor, shielded

L2 1 µH molded choke (brown-black-gold)

L3 10 µH molded choke (brown-black-black)

* Inductors available from Digi-Key Electronics, see Appendix.

Whether you're a novice or an "old-timer," you'll be pleasantly surprised at how this dual band Direct Conversion Receiver performs! The receiver can be constructed to receive either the 80 meter or 40 meter amateur radio bands. You will be able to receive AM, CW and SSB, signals from hams around the world.

Far-away signals are easily copied; and the sensitivity of the 80/40 Meter Direct Conversion Receiver rivals more expensive receivers. This receiver is ideal for practicing and increasing your CW or code efficiency or just having fun listening to shortwave broadcasts or ham radio operators, at home or away. Build the receiver for your favorite band, tuning options allow you to customize the tuning range from full-band to 100 kHz or less. Powered by a common 9 volt transistor battery, it's always ready for action on vacation and camping trips!

Figure 9-1 *80/40 meter code receiver*

The 80/40 Meter Direct Conversion Receiver shown in Figure 9-1, can easily be constructed by a radio neophyte in a few hours with the easy-to-follow instructions.

Circuit description

The circuit diagram shown in Figure 9-2 illustrates the schematic for the 80/40 Meter Direct Conversion Receiver. A dipole or suitable antenna is fed directly to input jack J1, which is coupled to potentiometer R1, which serves as a continuously variable RF attenuator and also serves as the receiver's gain control. Using the RF gain control to set volume is advantageous as it also reduces strong in-band signals that could otherwise overload the receiver front end. Series inductors L2 and L3, and trimmer capacitor C18, provide front-end bandpass filtering, and an impedance match to 50 to 75-ohm antenna systems. IC U1 is an NE602/612 mixer and Local Oscillator (LO) in an 8-pin dip package. The mixer section is an active Gilbert cell design for good conversion gain and low noise figure. The LO section uses heavy capacitive loading to minimize frequency drift. Tuning is capacitive, using a modern Hi-Q miniature plastic variable. The local oscillator (LO) stability is enhanced by a 78L05 voltage regulator.

Since this is a direct conversion receiver, the LO is tuned to the carrier frequency of Single Sideband Signals, or to a difference of 300 to 800 Hz on CW

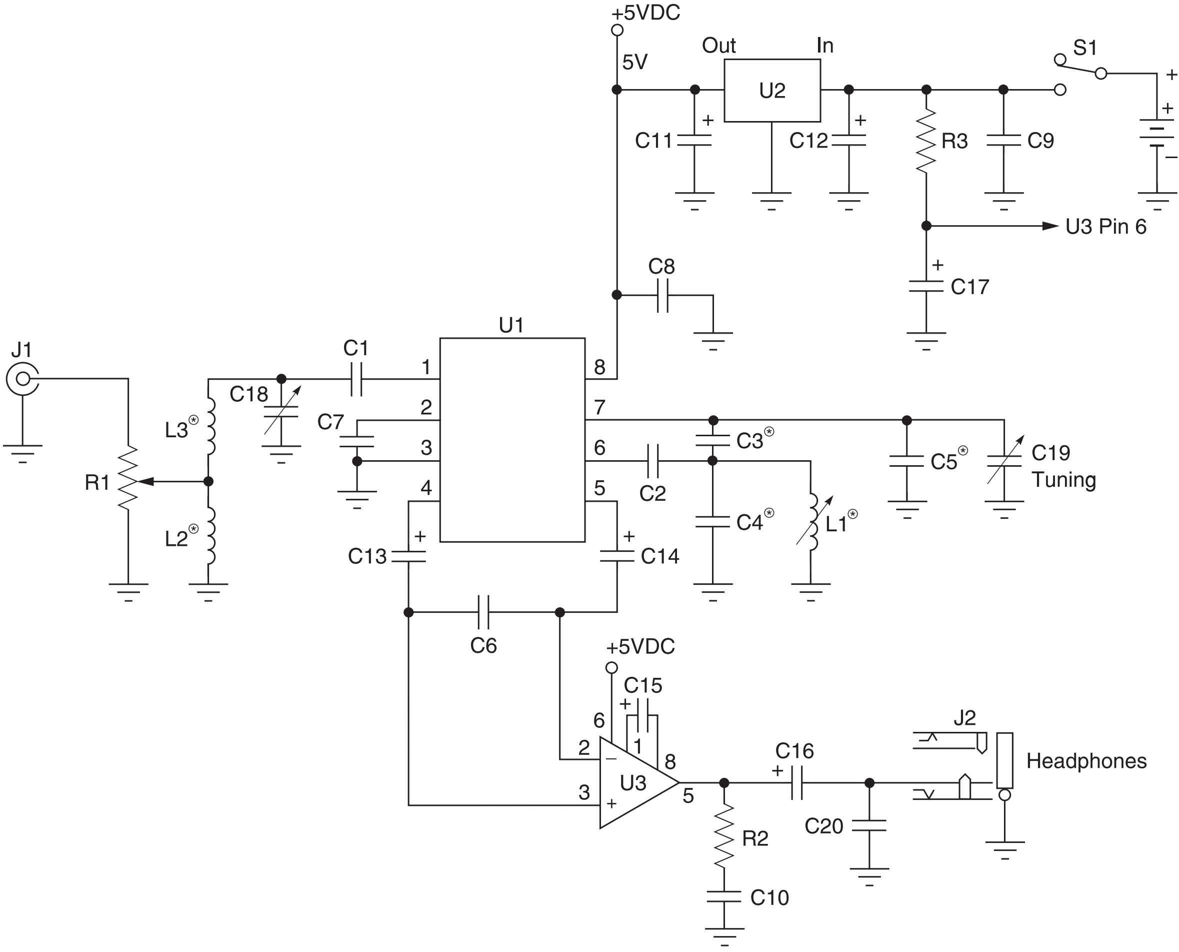

Figure 9-2 *80/40 meter direct conversion amateur radio receiver. Courtesy of Vectronics*

signals, to produce an aural output that is differentially coupled to a LM386 audio IC. The audio amplifier IC is coupled to the headphone jack at J2 by capacitor C16. The receiver circuit is power from a single 9 volt transistor radio battery.

Let's get started! Before we begin building the 80/40 meter receiver, you will need to locate a clean well lit and well ventilated work area. A large table or workbench would be a suitable work surface for your project. Next you will want to locate a small 27 to 33 watt pencil tipped soldering iron, as well as a roll of #22 gauge 60/40 tin/lead solder and a small jar of "Tip Tinner," a soldering iron tip cleaner/dresser, from your local Radio Shack store. You will also want to secure a few hand tools for the project, such as a pair of small end-cutters, a pair of tweezers and a pair of needle-nose pliers. Locate a small Phillips and a small flat-blade screwdriver, as well as a magnifying glass to round out your tool list. Grab the schematic, parts layout diagram as well as the resistor and capacitor identifier charts and we will begin our project. Place all the project components on the table in front of you. The 80/40 direct conversion radio is an RF or radio frequency project and it is best constructed on a printed circuit board with large ground plane areas covering the board for the best RF grounding techniques. Once you have all the parts and PC board in front of you, heat up the soldering iron and we'll get started!

First, find your resistor identifier chart in Table 9-1, which will help you select the resistors from the parts pile. Resistors used in this project are mostly small ¼ watt carbon composition type resistors, which have colored bands along the resistor body. The first color band should be closest to one end of the resistor body. This first color band represents the first digit of the resistor value. For example, resistor R2 has four color bands, the first one is a brown band followed by a green band followed by a black band. The fourth band is gold.

Table 9-1

Resistor color code chart

Color Band	1st Digit	2nd Digit	Multiplier	Tolerance
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

The first band is brown, which denotes a digit one (1), the second band is green, which represents a digit five (5), and the third band is black, which represents a multiplier of zero (0), so the resistor value is 15 ohms with a tolerance value of 5%. Identify the remaining resistors for the project and we can begin "populating" the PC board. Place a few resistors on the board at one time, so as not to confuse the process. Make sure that you place the correct resistor into the correct PC location before soldering it in place. Once you solder a few resistors in place on the PC board you can use your end-cutter to trim the excess component leads. Cut the excess leads flush to the edge of the circuit board. Then place a few more resistors on the PC board and solder them on to the board. Remember to trim the component leads as necessary.

Next we will locate the capacitors for the 80/40 meter receiver. Capacitors are listed as electrolytic and non-electrolytic types. The non-electrolytic types are generally smaller in value and size as compared with the electrolytic types. Non-electrolytic capacitors, in fact, can be so small that their actual value cannot be printed on them, so a special chart was devised as shown in Table 9-2. The chart illustrates the three digit codes which are often used to represent capacitors. For example, a .001 µF capacitor will have (102), while a .01 µF capacitor will have (103) marked on it to represent its true value. Use the chart to identify the small capacitors in the project. Non-electrolytics have no polarity markings on them so they can be installed in either direction on the PC board.

Electrolytic capacitors are usually larger in size and value and they will have a white or black band on the side of the capacitor body or a plus (+) or minus (−) marking on the body of the capacitor near the leads. These markings are polarity markings and that indicate the direction in which the capacitor must be mounted on the PC board. Failure to observe polarity when installing the capacitor may result in damage to the capacitor or to the circuit itself, so pay particular attention to capacitor polarity when placing the capacitors on the board. Note that the receiver can be built for either the 80 meter ham band or the 40 meter ham band. You will need to look at the parts list when deciding which band you want to receive. The inductors L1, L2 and L3 as well as capacitors C3, C4 and C5 determine the band selection.

Let's go ahead and place some of the non-electrolytic capacitors on the PC board. Identify a few small capacitors at a time and place them on the PC board and solder them in place. Trim the component leads after. Go ahead and install the remaining non-electrolytic capacitors after choosing your desired band. Install the capacitors and solder them in place, remember to cut

Table 9-2

Capacitance code information

This table provides the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Some capacitors actually have the numeric values stamped on them; however, many are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

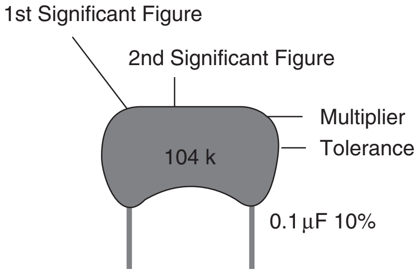

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

the extra leads flush to the edge of the circuit board. Now, you can go ahead and install the electrolytic capacitors onto the circuit board. Go ahead and solder them in place and trim the leads as necessary.

This receiver project uses a number of small inductors. These small inductors will generally have color bands on them to help identify them. Molded chokes appear, at first glance, to be similar to resistors in both shape and band marking. However, a closer look will enable you to differentiate between the two—chokes are generally larger in diameter and fatter at the ends than resistors. When doing your inventory, separate out any chokes and consult the parts list for specific color-code information. Note, that inductor L1 is an adjustable slug tuned type. Remember that specific chokes are used for each band: see parts list before mounting the chokes. Chokes do not have polarity so they can be mounted in either direction on the PC board.

The 80/40 meter receiver utilizes two integrated circuits and a regulator IC. Take a look at the diagram shown in Figure 9-3, which illustrates the semiconductor pin-outs. When constructing the project it is best to use IC sockets as an insurance against a possible circuit failure down-the-road. Its much easier to unplug an IC rather than trying to un-solder it from the PC board. IC sockets will have a notch or cut-out at one end of the plastic socket. Pin one (1) of the IC socket will be just to the left of the notch or cut-out. Note that pin 1 of U1 connects to C1, while pin one (1) of U2 connects to C15. When inserting the IC into its respective socket make sure you align pin one (1) of the IC with pin one (1) the socket. Failure to install the IC properly could result in damage to the IC as well as to the circuit when power is first applied.

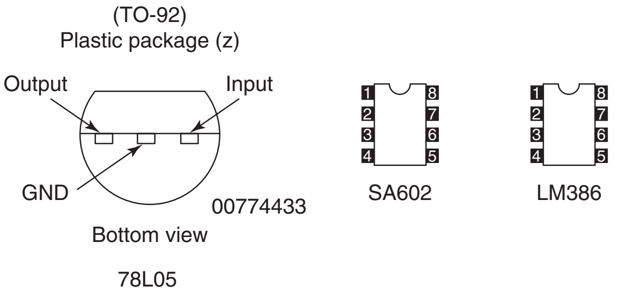

Figure 9-3 *Semiconductor pin-outs*

Let's finish the circuit board by mounting the volume control and the adjustable capacitors. The volume control potentiometer at R1 is a right angle PC board mounted type, which is placed at the edge of the PC board as is the main tuning capacitor. Capacitor C18, which is connected at the junction of L3 and C1, is an 8.2 pF trimmer type; go ahead and solder it to the circuit board using two pieces of bare #22 ga. stiff single conductor wire.

Locate main-tuning capacitor C19, now locate the mounting location for main tuning capacitor C19. The tuning capacitor was mounted on its side using double-sided sticky tape. Remove the protective cover from the double-sided tape, and firmly press the body of capacitor C19 to mount it to the PC board. Firmly press the double-sided tape over the silk-screened outline for the body of C19. You may decide to mount the tuning capacitor in a different way. At the rear of the capacitor: locate the four internal trimmer capacitors, and using a small jeweler's screwdriver or alignment tool, fully open (unmesh) all four trimmers. Note: tuning shaft faces front of board. Bend the two rotor lugs so they are parallel to the front face of the capacitor as shown above. Connect the two rotor lugs to the PC board ground points as shown using scraps of lead wire trimmed from resistors as jumper wires. Cut a 6″ length of 24-AWG insulated hook-up wire in half. Remove about ¼″ of the insulation from each of the cut ends. Solder the jumpers to the capacitor rotor lugs, and to the ground foil run on the bottom of the PC board. Since the tuning capacitor has four sections, you can increase the tuning ranges by paralleling different sections to give a greater tuning range: see Table 9-3. If you used just the 140 pF section, your tuning range would be 190 kHz, but if you combined the 140 pF section with the 40 pF section your tuning range would become 180 pF and so forth. Capacitor jumpers: 180 pF = use 140-pF and 40-pF sections paralleled; 222 pF = use 140-pF and 82-pF sections paralleled; 262 pF = use 140-pF, 82-pF and a 40-pF sections paralleled; 302 pF = use all four capacitor sections in parallel.

The 80/40 meter code practice receiver can be built for either of your favorite ham bands. First you will need to decide which band you are interested in. Then you will have to select the proper components from the chart in Table 9-4, for the band of interest. Note that you will have to select three capacitors and three coils for each band.

Table 9-3
Band tuning range

80 Meter Tuning	
302 pF	350 kHz
262 pF	310 kHz
222 pF	270 kHz
180 pF	230 kHz
140 pF	190 kHz
82 pF	110 kHz
40 pF	55 kHz
40 Meter Tuning Range	
302 pF	N/A
262 pF	N/A
222 pF	N/A
180 pF	310 kHz
140 pF	250 kHz
82 pF	150 kHz
40 pF	75 kHz

Let's take a few minutes for a short well-deserved break, and when we return we will look over the circuit board for any possible "cold" solder joints or "short" circuits. Pick up the circuit board with the foil side facing upwards toward you. First, we will examine the circuit board for "cold" solder joints. Take a look at all of the solder joints on the board, they should look clean, bright and shiny. If you see any solder joints which look dark, dull or "blobby," then you should remove the solder with a solder-sucker and then reapply more solder and re-solder the joint all over again so that it looks good. Next, we will look the board over for any possible "short" circuits which are often caused by "stray" component leads cut from the board or from small solder blobs which may have bridged between circuit traces. Rosin core solder can often leave a sticky residue which will collect component leads and solder blobs, so look the PC board over carefully for any foreign objects on the foil side of the board.

Table 9-4
80/40 Band change components

40-Meter Components	*80-Meter Components*
C3 - 470 pF capacitor	C3 - 470 pF capacitor
C4 - not used	C4 - not used
C5 - 680 pF capacitor	C5 - 470 pF capacitor
L1 - 1.5 µH molded coil	L1 - 6.8 µH molded coil
L2 - 1 µH molded coil	L2 - 3.3 µH molded coil
L3 - 10 µH molded coil	L3 - 33 µH molded coil

The 80/40 meter receiver radio prototype was mounted in a metal chassis box on plastic standoffs. The PC board was aligned in the enclosure so that the volume control, and main tuning control and power switch were all mounted at one side of the case. Alight the PC board along the edge of the chassis box, so that you drill holes in the chassis for the volume and tuning controls. The ⅛″ switched headphone jack at J2 was mounted on the front of the chassis, while the antenna input jack at J1 was mounted on the rear of the case. The 9 volt transistor radio battery was mounted in a battery holder which was fastened to the bottom panel of the chassis box.

Your new 80/40 meter direct conversion receiver project is almost completed. Apply the battery clip to the 9 volt battery. Connect up an antenna to the antenna jack, a dipole antenna cut for the band of interest would be the best type of antenna. For test purposes, you could connect a long piece of wire to the antenna jack for a temporary antenna. Plug in an 8 ohm headphone and then, turn on the power switch to the "On" position. Now, adjust the volume control to the mid-position and then turn the main tuning control. If all-goes-well, you should begin to hear some radio stations. You may want to test the radio in the early morning or evening when these two bands are most active for best results. Hopefully you will hear some stations while you tune across the band.

If for some reason, your receiver is DEAD, then you will want to turn off the receiver and remove the battery clip and do another inspection of the circuit board. This inspection is best done by another pair of eyes, someone who is skilled at electronics might be your best choice for inspection. You will want to make sure that you installed the electrolytic capacitors correctly with respect to their polarity. You will also want to make sure that the integrated circuits are aligned correctly in their respective sockets: it's easy to make a mistake. Check to make sure that the regulator IC at U2 was correctly installed with the input of the regulator going to capacitor C12, and the output of the regulator going to capacitor C11. Finally, make sure that you have the battery clip polarity wired correctly. Once you have examined the circuit board once more, you can go ahead and reconnect the battery and antenna and try the receiver again to see if it works now.

If your receiver is now working correctly, we can move on to the alignment of your new receiver. There are two primary methods of aligning the receiver's local oscillator.

Receiver alignment method 1: using an additional receiver

Place your new 80/40 meter receiver next to an amateur or general coverage receiver that covers the band of interest. Connect a short 3 or 4 foot wire antenna to the receiver, route the antenna wire so it passes near the direct conversion receiver PC board. Next, select a tuning tool that properly fits coil L1. The 20 meter kit requires a hex type alignment tool. The other kits require a small blade type tool that matches the slot in the coil core, or use a small jeweler's screwdriver. Now, set coil L1 tuning core to the top of the coil form. At this time, tune the receiver to the lowest operating frequency by turning the tuning capacitor shaft fully CCW (counter clockwise). Next, go ahead and set the monitor receiver to the lowest frequency you wish to receive on the direct conversion receiver. The receiver should be set to a wide filter bandwidth, and if it has a BFO, it should be on. Install a fresh 9 volt alkaline battery to the receiver, and turn it on by depressing the push-action power switch.

Finally, carefully adjust the core in coil L1 until it is heard sweeping across the monitor receiver frequency. Set the coil for zero beat, or strongest S-meter reading on the receiver. Note that metal alignment tools will cause some frequency shift, and interact with the tuning when removed or inserted into the coil. You will have to adjust the coil in small increments, and observe the point where the proper frequency is reached when the tool is removed.

Receiver alignment method 2: using a signal generator

First, set the signal generator to the direct conversion receiver's lowest operating frequency. Set the generator output level for a 30-µV CW signal. Next, connect the generator RF output to the VEC direct conversion antenna jack (J1) using coax cable and the appropriate mating connectors or adapters. Now, connect a speaker or headphones to jack J2 on the receiver. At this point, set the receiver gain fully CW (clockwise) and the tuning knob fully CCW (counter clockwise). Next, select a tuning tool that properly fits coil L1. The 20 meter kit requires a hex type alignment tool. The other kits require a small blade type tool that matches the slot in the coil core. Or, a jeweler's screwdriver may be used. Go ahead and set coil L1 tuning core to the top of the coil form. Now, connect a fresh 9 volt alkaline battery to the receiver, and turn it on using the push-action power switch. Finally, carefully adjust the core in coil L1 until the generator signal is heard. Set the coil for zero beat. Note that metal alignment tools will cause some frequency shift, and interact with the tuning when removed or inserted into the coil. You will have to adjust the coil in small increments, and observe the point where the proper frequency is reached when the tool is removed.

Final alignment

Once the tuning range is set, the RF input stage must be peaked for best sensitivity. This can be done using a signal generator, or on-the-air signals. Find an alignment tool or jeweler's screwdriver that fits trimmer capacitor C18. Use one of the following three methods to peak C18.

1. Using a signal generator, peak C18 for best sensitivity. Reduce the signal generator output, or reduce the R1 Gain control setting, to prevent overload as C18 is brought into resonance.
2. Using off-air signals, adjust C18 until signals are loudest. Reduce the R1 Gain control as C18 is

brought into resonance to prevent overload. Verify that C18 is properly peaked while monitoring a known on-air signal.

3. With an outside antenna connected, peak C18 for maximum atmospheric noise in the headphones or speaker. Verify that C18 is properly peaked while monitoring a known on-air signal.

Antenna considerations

With your new receiver completed and aligned, you can attach a dipole antenna for the chosen band of operation. Direct conversion receivers need good antennas. Using a random length of wire will give poor results. Half-wave dipoles are very effective antennas and are simple to construct. For low-power or QRP work, or for receive only applications, low-cost RG-58 52 ohm coax is fine. The formula for dipole antennas: 234/(frequency in MHz) = ¼-wavelength in feet will give the length of wire needed for each leg of the dipole. For example, if you wanted to build a 40-meter dipole for the CW portion of the band you would divide 234 by 7.040 MHz and the result would be 33.238 feet per leg on either side of the center insulator: see Figure 9-4.

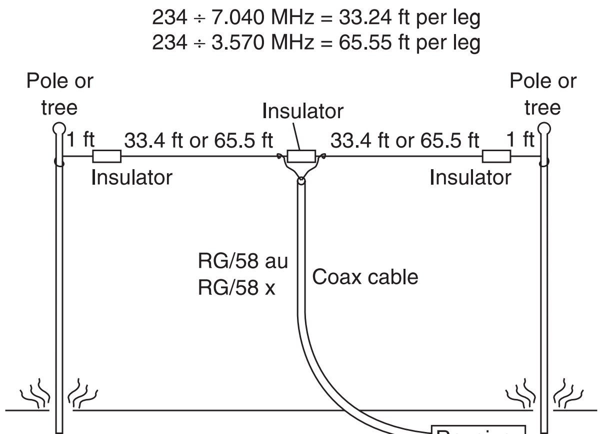

Figure 9-4 *80/40 meter dipole antenna*

Tuning notes

While trying to tune CW stations, you should tune across the band until a CW signal is heard. Set the gain control at R1 for a comfortable listening level. Then, carefully adjust the tuning control until the "beat note," or tone, sounds best for you. Most experienced CW operators prefer a lower beat note of about 300 or 400 Hz.

When trying to tune-in single-side band (SSB) stations, you may need a little practice to tune-in the signal properly. Direct conversion receivers offer no upper or lower sideband rejection, so you will have to slowly tune across the SSB signal until the voices become natural sounding. Use the gain control to prevent overload and to set a comfortable listening level.

In order to tune-in AM stations, you will need to realize that, unlike SSB signals, AM signals have a carrier frequency. That means you will hear a beat note as you tune across an AM signal. Tune the signal so the beat note becomes lower in frequency and stop when "zero beat" (beat note disappears) is reached.

Possible overload problems

An unfortunate fact of life is that low-power amateur signals are surrounded by high-power SW broadcast signals, so there may be a possibility of overload problems. In order to compensate for this possible problem, you may need to lower the gain control. Simply lowering the gain setting will resolve many overload problems. Another approach to an overload problem is to use a **preselector**. A preselector provides protection from strong out-of-band signals. They are most effective when the interfering signal is far from the operating frequency.

A resonant antenna favors signals in the band it was cut for, so when faced with a severe overload problem, you elect to use a resonant antenna. The final approach to solving an overload problem is to change the direction the dipole faces; this may lower unwanted pickup of undesired SW signals. Lowering the antenna will change its angle of pickup; a low antenna may favor local stations over the low-angle of distant broadcast signals. A shorter antenna can also be tried; it will be less efficient and signals will be weaker.

Table 9-5

W1AW on-air code practice sessions

Pacific	*Mtn*	*Central*	*East*	*Mon*	*Tues*	*Weds*	*Thur*	*Fri*
6 am	7 am	8 am	9 am		fast code	slow code	fast code	slow code
7-9 am	8-10 am	9-11 am	10-noon	visiting	operator	time	————	————
10 am-12:45 pm	11 am-1:45 pm	noon-2:45 pm	1-3:45 pm					
1 pm	2 pm	3 pm	4 pm	fast	slow	fast	slow	fast
2 pm	3 pm	4 pm	5 pm	————	code	bulletin	————	————
3 pm	4 pm	5 pm	6 pm	————	teleprint	bulletin	————	————
4 pm	5 pm	6 pm	7 pm	slow	fast	slow	fast	slow
5 pm	6 pm	7 pm	8 pm	————	code	bulletin	————	————
6 pm	7 pm	8 pm	9 pm	————	teleprint	bulletin	————	————
6:45 pm	7:45	8:45	9:45	————	voice	bulletin	————	————
7 pm	8 pm	9 pm	10 pm	fast	slow	fast	slow	fast
8 pm	9 pm	10 pm	11 pm	————	code	bulletin	————	————

Frequencies are 1.8175, 3.5815, 7.0475, 14.0475, 18.0975, 21.0675, 28.0675 and 147.555 MHz.

Slow Code = practice sent at 5, 7½, 10, 13 and 15 words per minute (wpm).

Fast Code = practice sent at 35, 30, 25, 20, 15, 13 and 10 wpm.

The chart shown in Table 9-5 illustrates the W1AW code practice sessions for Ham Radio operators or those interested in obtaining a Ham Radio license. The 80-meter frequency is 3.5815 MHz and the 40-meter frequency is 7.0475. The Amateur Radio Relay League (ARRL) offers these code practice sessions every day to help amateur radio enthusiasts increase their code speed. The ARRL is a great resource for those interested in Amateur Radio and offers many books, magazines and free literature in including videos to help you understand what Amateur Radio is all about. Take a few minutes and explore their web-site.

Chapter 10

WWV 10 MHz "Time-Code" Receiver

Parts list

WWV Time Receiver

R1 22k ohm, $\frac{1}{4}$w, 5% resistor

R2,R6,R7,R10 100 ohm, $\frac{1}{4}$w, 5% resistor

R16,R18,R30 100 ohms, $\frac{1}{4}$w, 5% resistor

R3,R21 1 megohm, $\frac{1}{4}$w, 5% resistor

R4 3k ohm, $\frac{1}{4}$w, 5% resistor

R5,R11,R12 2k ohm, $\frac{1}{4}$w, 5% resistor

R8,R13 47k ohm, $\frac{1}{4}$w, 5% resistor

R9 220k ohm, $\frac{1}{4}$w, 5% resistor

R14,R24,R29 100k ohm, $\frac{1}{4}$w, 5% resistor

R19,R22 5.6k ohm, $\frac{1}{4}$w, 5% resistor

R20 27k ohm, $\frac{1}{4}$w, 5% resistor

R17,R23,R28 4.7k ohm, $\frac{1}{4}$w, 5% resistor

R31,R33,R34 4.7k ohm, $\frac{1}{4}$w, 5% resistor

R25,R26,R27 10k ohm, $\frac{1}{4}$w, 5% resistor

R32 12k ohm, $\frac{1}{4}$w, 5% resistor

R35,R36 22 ohm, $\frac{1}{4}$w, 5% resistor

R37 2.7k ohm, $\frac{1}{4}$w, 5% resistor

C1,C2,C6,C7,C15 .001 μF, 35v ceramic disk capacitor

C3,C17,C19,C22 .01 μF, 35v ceramic disk capacitor

C4,C5,C8,C9,C14 .1 μF, 35v ceramic disk capacitor

C16,C23,C28,C31 .1 μF, 35v ceramic disk capacitor

C10,C11,C12,C13 20 pF, 35v plastic capacitor

C18 1000 pF, 35v plastic capacitor

C20,C21 220 pF, 35v plastic capacitor

C24,C27 1.0 μF, 35v electrolytic capacitor

C25,C32 22 μF, 35v electrolytic capacitor

C26 2.2 μF, 35v electrolytic capacitor

C29,C30 390 pF, 35v plastic capacitor

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

C33 100 µF, 35v electrolytic capacitor

C34 2200 µF, 35v electrolytic capacitor

C35 10 µF, 35v electrolytic capacitor

C36,C37 1000 µF, 35v electrolytic capacitor

C38 330 µF, 35v electrolytic capacitor

C39 470 µF, 35 volt electrolytic capacitor

CV1 10-40 pF trimmer capacitor

CV2 10-70 pF trimmer capacitor

VR1 10k ohm RF gain potentiometer

VR2 1M ohm Sensitivity potentiometer

VR3 10k ohm volume control potentiometer

L1 4.1 µH coil - 27 turns on T-28-2 ferrite core using 22 AWG wire tap 2-turns from ground side

L2,L4 12 turns on FT37-43 ferrite toroid core

L3 3.0 µH - 25 turns #26 AWG wire on T50-2 ferrite core

L5 12 turns on FT37-43 ferrite toroid core

L6 1000 µH coil

D1 1N5711 silicon Schottky diode NTE 112

D2 optional LED - power "on" indicator

Y1,Y2,Y3 10 MHz crystal

(F) 3-ferrite beads - (see front-end)- run wire through beads

Q1,Q2,Q3 Dual Gate MOSFET 40673 or ECG222

Q4,Q5 nJFET MPF102

Q6,Q7 2N3904 NPN transistor

Q8 2N3906 PNP transistor

U1 Signetics NE5532 Op-Amp

J1 RCA jack - antenna

J2 Coaxial power jack

SPK 8-ohm 3" speaker

S1 SPST power switch

Misc PC board, wire, hardware, sockets, etc.

Does anybody really know what time it is? The WWV receiver, shown in Figures 10-1 and 10-2, is designed to tune the National Institute of Standards and Technology (NIST) atomic clock for moment by moment time accuracy! This is a great receiver project that actually serves a dedicated purpose. Put one in the boat or on a desk and you'll always know what time it is!

The WWV receiver is quite sensitive, and with a good antenna you should be able to receive time signals most of the time! The receiver features easy tuning and small size, it can be packed up and taken anywhere and runs on a 12 volt DC power supply, or a 12 volt battery for portable operation. An earphone jack is provided for headset operation so you can use it night or day.

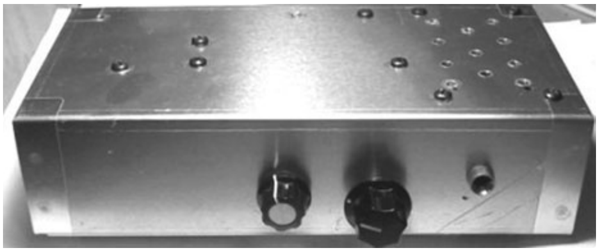

Figure 10-1 *WWV time code received*

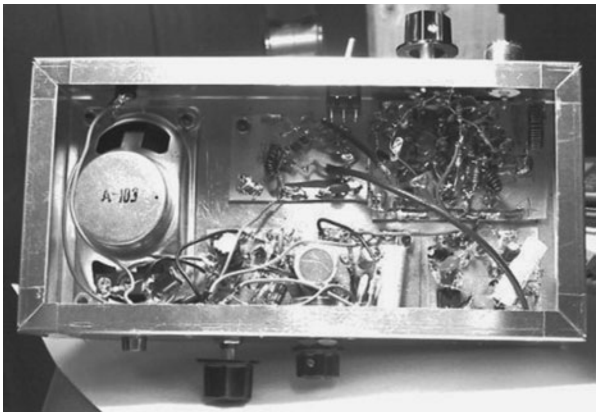

Figure 10-2 *WWV time code received – inside view*

People have always been fascinated with the measurement and properties of time. What happened to that last second that you just spent reading this sentence? Perhaps this is the reason that so many of us have tuned in over the years to the National Institute of Standards and Technology (NIST) "clock" operating at a transmitting frequency of 10 MHz. The cesium standard NIST-7 has an accuracy of 5 × 10-15 seconds. Now that's accurate! In fact, the definition of a "second" is the duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom.

We set our clocks, watches, and chronometers to the all-important "beep" of the radio signal that denotes the correct time to the second. In addition; avid listeners of the NIST clock are also treated to the following (from the NIST web site).

The OMEGA Navigation System status reports are voice announcements broadcast on WWV at 16 minutes after the hour. The OMEGA Navigation System consists of eight radio stations transmitting in the 10- to 14-kHz frequency band. These stations serve as international aids to navigation. The status reports are updated as necessary by the U.S. Coast Guard. Current geophysical alerts (Geoalerts) are broadcast in voice from WWV at 18 minutes after the hour. The messages are less than 45 s in length and are updated every 3 hours (typically at 0000, 0300, 0600, 0900, 1200, 1500, 1800, and 2100 UTC). Hourly updates are made when necessary.

Marine storm warnings are broadcast for the marine areas that the United States has warning responsibility for under international agreement. The storm warning information is provided by the National Weather Service. Storm warnings for the Atlantic and eastern North Pacific are broadcast by voice on WWV at 8, 9, and 10 minutes after the hour. An additional segment (at 11 minutes after the hour on WWV) is used occasionally if there are unusually widespread storm conditions. The brief voice messages warn mariners of storm threats present in their areas. The storm warnings are based on the most recent forecasts. Updated forecasts are issued by the National Weather Service at 0500, 1100, 1700, and 2300 UTC for WWV. Since March 1990 the U.S. Coast Guard has sponsored two voice announcements per hour on WWV giving current status information about the GPS satellites and related operations. The 45-s announcements begin at 14 and 15 minutes after each hour.

WWV radiates 10,000 watts on 5, 10, and 15 MHz. The WWV antennas are half-wave dipoles that radiate omnidirectional patterns. The station uses double sideband amplitude modulation. The modulation level is 50 percent for the steady tones, 25 percent for the BCD time code, 100 percent for the seconds pulses and the minute and hour markers, and 75 percent for the voice announcements. That's a lot of information being broadcast, and you can easily pickup these signals with the high frequency 10 MHz Time Receiver shown in this project. The WWV Time receiver has excellent sensitivity, selectivity, and dynamic range. (from the NIST web site http://www.nist.gov/)

Circuit description

The 10 MHz WWV Receiver is shown in the block diagram of Figure 10-3 and in the main schematic diagram in Figure 10-4. The WWV Time Code receiver is an AM TRF or Tuned Radio Frequency or TRF type receiver. This receiver project is designed specifically for 10 MHz time signal operation. The block diagram of the WWV receiver illustrates the four major sections of the receiver. The receiver's front-end, as shown in the schematic diagram, features two sensitive dual-gate MOSFETs in the front-end circuit with a single-pole filter. An RF gain control is shown at VR1 which adjusts the input sensitivity, via a 10k ohm potentiometer. Note inductor L1 is 4.1 µH coil, which consists of 27 turns of 22 ga. AWG wire wound on a

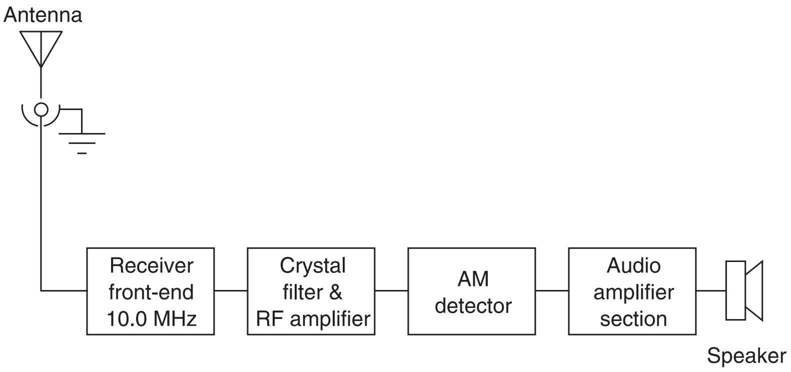

Figure 10-3 *10 MHz WWV receiver block diagram*

T-68 ferrite core. Inductors L2 and L4 consist of 12 turns of 22 ga. AWG wire on an FT37-43 ferrite core, and finally L3 consists of 25 turns of #26 ga. AWG wire on a T50-2 ferrite core. For even greater selectivity (but greater insertion loss), you can consider moving the L1 tap to 1 turn from ground. Fine tuning the WWV receiver is accomplished by adjusting CV1 and CV2 trimmer capacitors. The front-end MOSFETs are extremely sensitive to static electricity, so be very careful in handling them. Note the two ferrite beads, one is placed in series with the Drain lead of Q1, while the second ferrite bead is in series with the Drain of Q2. These ferrite beads are important, so be sure to use them. The receiver front-end, at J1, should be connected to a dipole antenna "cut" for 10 MHz. TRF receivers are especially sensitive and are subject to overloading a BCB or Broadcast Band Interference, so internal lead lengths should be kept short between components and the front-end should be shielded and the receiver placed in a metal enclosure for best results.

The output impedance of Q2 is 2000 ohms, which nicely matches the input impedance of the Cohn crystal filter, in the crystal filter section of the receiver. This filter utilizes three 10 MHz crystals. Choose 10 MHz crystals that are marked for a 20 pF or 32 pF load capacitance if possible. Using a 10 MHz crystal oscillator, find three that are closest to one another in frequency. You may substitute 2.2k ohm resistors instead of the specified 2k with a slight penalty in pass band shape. The crystal filter section utilizes a single dual-gate MOSFET transistor at Q3. Coil L5 is 12 turns on an FT37-43 ferrite core. The output of the Cohn crystal filter is next fed to the input of the AM detector stage.

The AM detector has three distinct advantages; it has high bandwidth, low distortion and incredible (and variable) sensitivity. The variable bias control allows the listener to adjust the bias to maintain detected audio fidelity even when the RF signal is weak. The detector uses a Schottky UHF mixer diode at D1. Increasing the diode bias from zero volts to maximum causes three things to happen: increased sensitivity, increased audio high frequency response, and a slight increase in receiver noise.

The AM detector section matches the 2000 ohm impedance of the crystal filter stage quite nicely. A sensitivity control at VR2 is a 1 megohm potentiometer, which feeds directly to L6, a 1000 µH coil. Coil L6 is connected directly to a Schottky diode at D1. The resultant signal is next sent through a 5.6k ohm resistor and on to a 1 µF capacitor which feeds the nJFET at Q4. The Source lead of Q1 is connected to the Drain lead of Q5, and the Source lead of Q5 is connected to ground via a 5.6k ohm resistor. Note the two 22 µF electrolytic capacitors on each of the Drain leads of the nJFETs.

When the WWV RF signal is weak, turning the bias off may result in the detected WWV signal disappearing. Increasing the bias will bring the WWV signal back in. You run the bias control pot about half way and, of course, higher as WWV fades out. The fidelity that the bias adds even when the WWV signal is strong is quite pleasing to the ear. For the 1 µF and 2.2 µF capacitors, I used polyester film types which sounded better than electrolytic capacitors.

The audio amplifier section features low noise audio amplifier. Distortion is very low as long as the input is

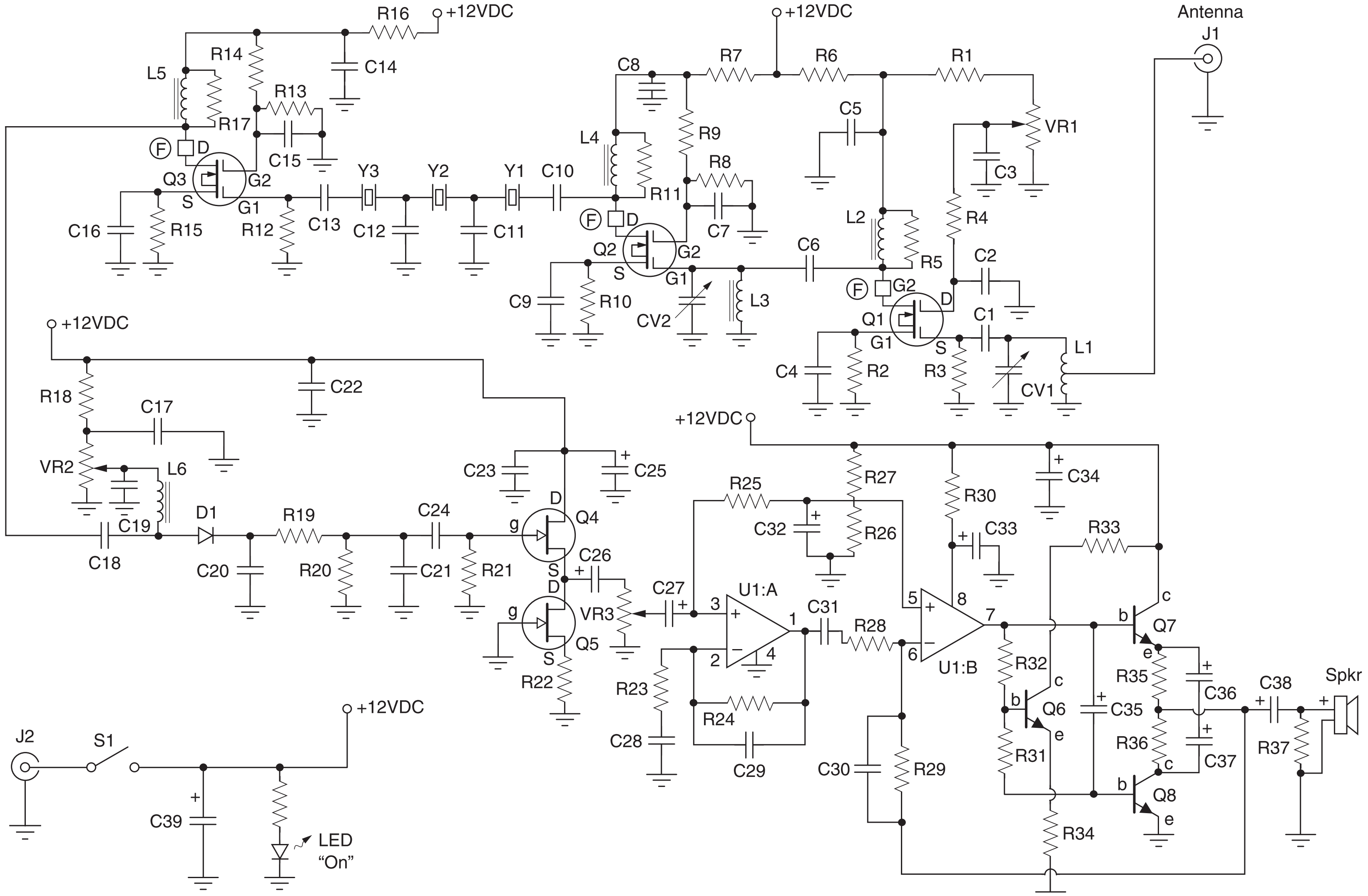
R16
+12VDC
R14
C14
L5
R13
R17
C15
F
D
Q3
G2
S
G1
Y3
Y2
Y1
C10
C16
R15
R12
C13
C12
C11
C8
R7
+12VDC
R6
R1
Antenna
J1
C5
VR1
R9
L4
R8
R11
C3
C7
F
D
Q2
G2
S
G1
L2
R4
R5
C6
C2
C9
R10
CV2
L3
F
G2
D
Q1
C1
L1
G1
S
C4
R2
R3
CV1
+12VDC
C22
R18
C17
VR2
L6
C19
C18
D1
R19
C24
C20
R20
C21
R21
C23
C25
D
g
Q4
S
C26
D
g
VR3
Q5
S
R22
+12VDC
R25
R27
C34
R30
C32
R26
C33
R33
C27
U1:A
3
2
1
4
C31
R28
5
8
7
6
U1:B
c
b
Q7
e
R23
R24
C28
C29
R32
c
b
Q6
e
C35
R35
C36
C38
Spkr
R36
C37
R37
C30
R29
R31
c
b
Q8
e
R34
+12VDC
J2
S1
C39
LED
"On"

Figure 10-4 *WWV time code receiver schematic. Courtesy of VC7BFO – Todd Gale*

not over driven. The audio amplifier section consists of a dual NE5532 op-amp and three transistors. Transistors Q6 and Q7 are NPN types, while transistor Q7 is a PNP type. The input signal from the detector is fed directly to the volume control at VR3, a 10k ohm potentiometer. The input signal is first sent to the non-inverting stage of the first op-amp at pin 3. The output of the first op-amp section at pin 1 is next sent to the second stage of the op-amp through a .1 µF capacitor in series with a 4.7k ohm resistor to pin of U1:b. The output of U1, at pin 7 is then fed to the output transistor stage. The output of the transistor pair is two 1000 µF capacitors which coupled to a 330 µF capacitor which feeds the 8 ohm speaker.

The entire WWV receiver is powered from a 12 volt "wall-wart" power or can be powered from a 12 volt battery for portable operation.

Receiver assembly

You are probably quite anxious to begin the assembly of the receiver project; it is best to follow the assembly steps in order. Try to avoid the urge to "jump ahead" installing components. Locate your layout and schematic diagram. Place all the components in front of you and locate a 27-30 watt pencil tip soldering iron. Make sure you have a large work table in order to spread out all the parts and diagrams. A printed circuit board is highly recommended for this project, so you should have your circuit board in front of you oriented correctly, in order to begin populating the PC board. Remember that the components will be mounted on the "component" side of the circuit board and soldered on the "solder" side of the circuit board, the side with the printed circuit traces. Locate a spool of 60/40 tin/lead rosin core solder for this project.

First we will begin by locating all of the resistors for the project. Most all of the resistors are ¼ watt carbon composition type resistors except for the potentiometer. Before we begin building the Time receiver, take a moment to refer to the resistor color chart shown in Table 10-1. Note that each resistor will have three or four color bands beginning at one end of the resistor body. The first color code represents the resistor's first digit value, while the second color band denotes the resistor's second digit value. The third color band represents the multiplier value of the resistor and the fourth band denotes the tolerance value of the resistor. If the fourth color band is silver then the resistor has a tolerance of 10%, while a gold band denotes a 5% tolerance. If there is no fourth color band then the resistor has a tolerance of 20%. The first color band of a 10,000 ohm resistor, for example,

Table 10-1
Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

would be brown with a value of (1). The second color band would be black with a value of (0). The third color band would be orange with a value of (000). Read the color codes on the resistors and compare them to the parts list and layout diagram so that you place the correct resistor in the correct hole on the printed circuit board. You can install the resistors in groups of four or five. Solder the first group of resistors onto the circuit board, then cut the excess component leads flush to the edge of the PC board using a pair of small end-cutters. Then move on to the next group of resistors until all of the resistors have been installed.

Next we will move on to installing the capacitors. Most all of the capacitors will have some sort of marking on them, so compare the value or marking on the component with the parts list before installing the components on the PC board. Note that capacitors C24, C25, C26, C27 are all electrolytic capacitors; these capacitors will have additional polarity markings on them and they must be installed with respect to these polarity markings in order for the circuit to work properly. Each of these capacitors will have a white or black band with either a plus or minus marking on one side or edge of the component. You must orient the capacitor correctly, so you should refer to the layout diagram and schematic diagram when installing these parts. Also note that capacitors C32 through C38 are also all electrolytic types. All the other capacitors have no polarity marking so they can be installed in either direction. Now, you will want to refer to the capacitor code chart shown in Table 10-2, which illustrates the codes used for small size capacitors. Often very small capacitors will not have their actual value printed on them, because there may not be enough room, so a three-digit code was developed. Look at the chart and compare the numbers on the chart with the actual capacitors to help identify them. You can go ahead and install all of the capacitors at this time. It is best to install the capacitors in groups of four or five at a time, then solder them on to the circuit board before moving on to the next group of capacitors; this helps keep some order to the building process. It is also advisable to check off each of the components against the parts list, so you can keep tabs on your progress and make sure that all of the parts are accounted for. Solder a group of four or five components, cut the excess leads and then move on to the next group of components. This will keep things in an orderly fashion.

Next, go ahead and locate and install the Schottky diode at D1. Note that the diode has a polarity marking on the glass enclosure. At one end of the diode body you should see a black or white band; this denotes the cathode side of the diode. Look at your schematic diagram to observe the direction of the cathode band in the circuit before you install the diodes.

The WWV receiver incorporates six inductors; five of the inductors are wound on ferrite cores, while L6 is a ready-made 1000 µH choke coil. Refer to the chart in Table 10-3 for details on winding each of the coils. When winding the coil you will have to scrape the insulation off each of the wire ends in order to solder the coil to the circuit board. You will need to secure the coil to the circuit board using a couple of "dabs" of rubber cement. Once the coils have been secured to the PC board, you can solder the coil leads to the circuit board, and then remove the excess lead lengths.

Locate the three crystals, Y1, Y2 and Y3 and place them on the PC board. Now solder the crystals to the PC board and then trim the excess leads flush to the edge of the circuit board.

Before we attempt to install the semiconductors, take a quick look at the semiconductor pin-out diagram shown in Figure 10-5. Now locate the MOSFETs Q1, Q2 and Q3. These devices are very static sensitive and must be handled very carefully using an anti-static wrist-band to avoid damage before they are soldered into the circuit. Transistor sockets are highly recommended for all of the transistors, see the pin-out diagram for the MOSFETs and make sure you can identify each of the leads before placing the MOSFETs on to the circuit board. The MOSFETs have two Gate leads, G1 and G2, as well as a Source lead (S) and a Drain lead (D). Once you are sure of each of the leads, you can go ahead and install the MOSFETS onto the circuit board. The AM detector portion of the receiver utilizes two nJFETS at Q4 and Q5, these semiconductors are also static sensitive and must be handled with extreme care to avoid damage. The nJFETS have three leads, like most conventional transistors, but they have a Gate, Source and Drain lead,

Table 10-2
Capacitance code information

This table provides the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Some capacitors actually have the numeric values stamped on them; however, many are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

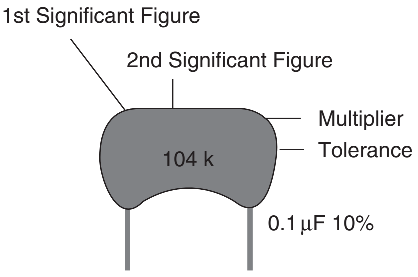

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

Table 10-3

Inductor coil winding information

Coil	*Value*	*Ferrite Core*	*Description*
L1	4.1 µH	T-28 toroid (*)	27 turns #22 AWG wire on ferrite core
L2		FT-37-43 (**)	12 turns #22 AWG wire on ferrite core
L3	3.0 µH	T50-2 toroid *	25 turns #26 AWG wire on ferrite core
L4		FT37-43 (**)	12 turns #22 AWG wire on ferrite core
L5		FT37-43 (**)	12 turns #22 AWG wire on ferrite core
L6	1000 µH		pre-wound choke coil; Digi-Key

* Toroids can be obtained from Circuit Specialists; see Appendix.

** FT-37-43 toroid available from RP electronics; see Appendix.

unlike conventional transistors, with a Base, Emitter and Collector. Once the nJFETS have been identified, they can be installed on the circuit board. Finally, the audio amplifier section utilizes three conventional transistors. Transistors Q6 and Q7 are NPN types, while Q8 is a PNP transistor. Transistors have three "legs" and must be oriented correctly. Notice that the part contains a "flat" side with the writing imprinted on it. Be sure to follow the parts diagram for correct placement. To install, slide the center legs through the circuit board and push the component as close to the board as possible without "straining" the leads. Identify the transistor pin-outs and then go ahead and install the transistors onto the circuit board. Remember to trim the

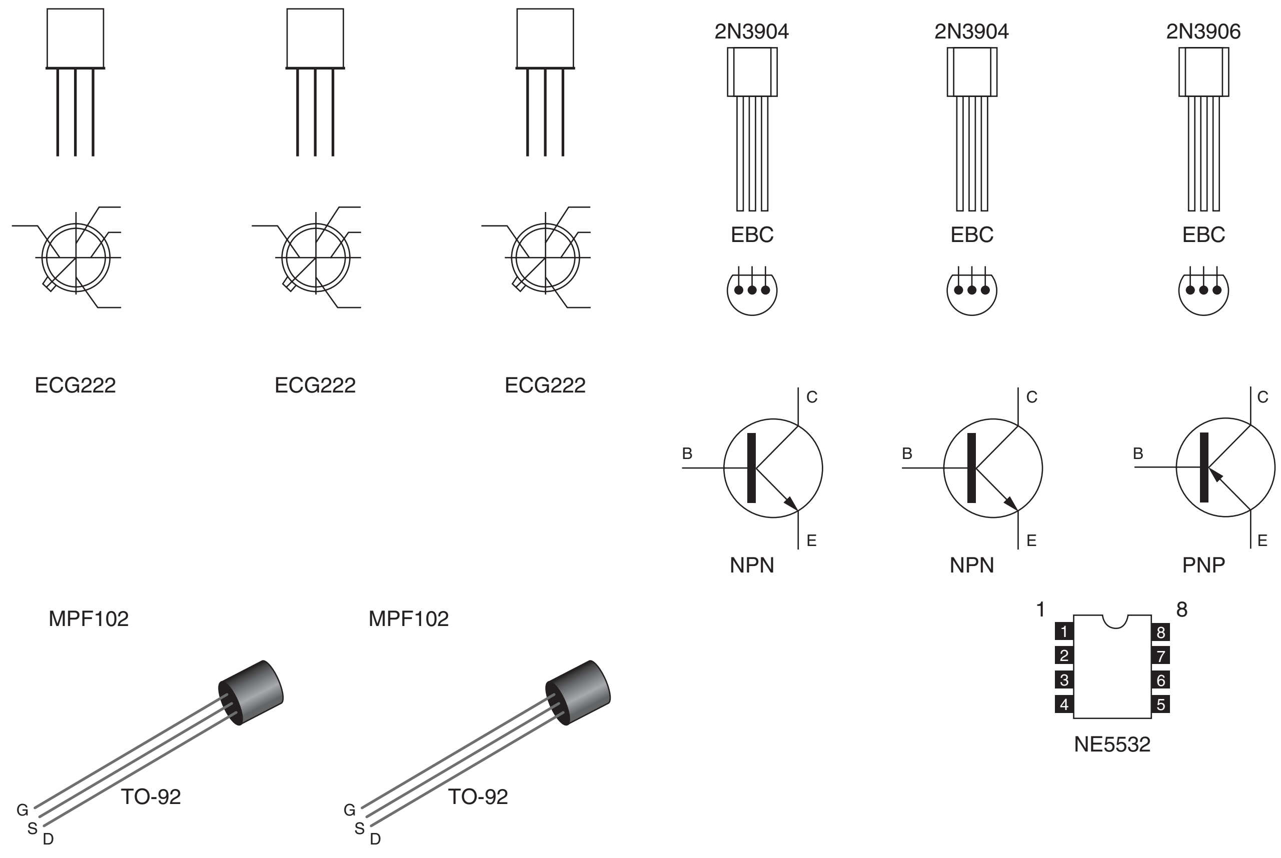

Figure 10-5 *Semiconductor pin-outs*

excess transistor leads after all of the transistors have been installed on the PC board.

It is a good idea to use an IC socket for each of the integrated circuits in the event of a component failure at a possible later date.

Leads length from the board

Now that your soldering skills are at their peak, we'll install the integrated circuit. An integrated circuit socket is highly recommended in the event of a possible circuit failure. It is much easier to simply unplug a "burnt" IC, rather than trying to un-solder a defective IC from the circuit board. Most hobbyists have considerable trouble trying to un-solder multi-pin integrated circuits from a PC board successfully, so it is easier to install a socket from the beginning. Most IC sockets will have a notch or cut-out and one end of the plastic package. Pin 1 will be just to the left of the notch or cut-out. Make sure you orient the socket, so that pin 1 will be close to C29, C31 and R22. When installing the IC into the socket make sure you align pin of the IC with pin 1 of the IC socket.

That's a lot of parts! Take a short break and rest your weary eyes. Before we start up again, we will check your solder joints for "short" circuits and "cold" solder joints. First, pick up the circuit board with the foil side of the board facing upwards toward you. Take a close look at the solder joints in front of you. The solder joints should all look clean, bright and shiny. If any of the solder joints look dark, dull or "blobby," then you need to remove the solder from that particular joint, re-apply solder and re-solder the joint all over again. Next, we will inspect the circuit board for possible "short" circuits. Look closely for small solder "blobs" that often bridge circuit traces. Also look carefully for any "stray" components leads which may have stuck to the foil side of the PC board. Sometimes sticky rosin from solder can leave a residue that will "hold" pieces of wire on the board, often causing "short" circuits between circuit traces.

Identify and install the three potentiometer; they can now be mounted on the metal enclosure that you chose to use for the receiver. Make up three or four inches leads for each of the potentiometer leads, so you can connect the potentiometer to the circuit board. All three potentiometers can be mounted on the front panel of the receiver enclosure. Using a length of speaker wire, connect the speaker to the circuit board. The speaker can be mounted on the top or front of the enclosure.

If you desire you can customize your new receiver with an LED power indicator. Install a red LED on the front panel. Being a diode, this component is polarized and must be oriented correctly. Examine the LED and notice how one lead is longer than the other. The longer of the two leads is the anode, or (+) connection. Most diodes also have a flat side molded in the component body. This corresponds to the cathode or (–) side of the part. This flat should face in the direction of the band marking of the diode. Leave the diode's leads as long as possible as this component will also mount to the front panel as a power indicator. You can connect up the LED across the power input of the circuit's +12 volt lead and ground.

You can also customize the receiver by installing a closed circuit earphone jack if desired, so that the jack will bypass the speaker when the headphones are plugged into the jack.

Moving to the rear of the enclosure, install antenna connector J1, you can use an RCA jack or a BNC connector as desired. Next, go ahead and mount a coaxial power plug on the rear panel of the chassis and then solder the connectors to the circuit board using a length of insulated wire. Now you will need to decide how you will power your new WWV receiver. You elect to power the receiver from a 12 volt battery or re-chargeable power pack or from a 12 volt DC "wall-wart" power supply.

Congratulations, you have successfully completed your WWV receiver; now let's move on to final testing of the receiver!

Initial receiver testing

This project was designed to be pretty much "plug and play," with minimal testing upon completion. Turn all three potentiometers to their mid-point position. Connect up you power supply or battery and depress the power switch. The power on LED should illuminate. Connect up your outdoor antenna via connector J1.

Now increase the RF gain control, more clockwise. You can also adjust the volume control for more gain as well. You should begin to hear the "beep" signal from the WWV Colorado transmitting station. You may have to adjust VR2 until the signal is sound best in the speaker. If the signal is too strong, you may have to adjust the RF gain and volume control as needed.

In the event that you do not hear any sounds from your new receiver, you will have to turn off the receiver and remove the battery and do a close inspection of the circuit. The most common reasons for circuit problems is due to placing the wrong component in a particular circuit board location or possible reverse placement of a component. Take a look at the Time-Code Receiver circuit board very carefully. Sometimes it is better to get a pair of "fresh eyes" or have a friend take a look at your circuit to find a possible error. Look at each of the resistors and make sure you have placed the correct resistor in each resistor location; read the color code carefully. Check the placement of the diodes and electrolytic capacitors to make sure you have installed them with respect to their polarity marking. Next, take a close look to make sure that you have installed the transistors correctly, looking closely at the pin-outs with respect to the manufacturer's pin-out diagram. Finally, examine the circuit to make sure that the integrated circuit has been oriented properly in its IC socket, since this is a common cause for error. Once you have fully inspected the circuit board and located your error, you can re-connect the battery and switch on the receiver to test it once again.

Antenna considerations

In order to obtain the best reception on your new WWV Time-Code receiver you will want to construct an optimum antenna. Using a random length of wire will give poor results. A very efficient method to use with the WWV Time-Code receiver is a dipole antenna "cut" for the proper frequency of 10 MHz. Half-wave dipoles are very effective antennas and are simple to construct. For low-power work or receiver use, you can use low-cost RG-58 52 ohm coax from the dipole antenna to your new receiver. The formula for dipole antennas: 234/(frequency in MHz) = ¼-wavelength in feet will give the length of wire needed for each leg of the dipole. For example, if you wanted to build a 10 MHz dipole for your Time-Code receiver, you would divide 234 by 10 MHz and the result would be 23.4 feet per leg on either side of the center insulator. The diagram in Figure 10-6 illustrates a dipole antenna which can be used with the Time-Code receiver. Depending upon your location, you will want to orient the antenna for best reception. The antenna should be placed broad-side or parallel towards Colorado for best reception.

Using your new WWV receiver

You will have fun learning about other interesting things on WWV besides the time function feature, such as accurate tone frequencies, geophysical alerts (solar activity reports), marine storm warnings, the global positioning system (GPS) and Omega navigation system status reports.

Table 10-4 shows the broadcast schedule for WWV in Fort Collins, CO. A similar schedule is available for WWVH in Hawaii. This diagram is extracted from a National Institute of Standards and Technology (NIST) publication. If you want to order a copy, it is available for a modest charge from the Government Printing Office, Washington DC 20402. Write and ask for a current list of available publications. The two features of most interest are precise frequency of the carrier signal, to calibrate radio and test equipment, and precise time information given each minute as a voice transmission. But there are many other interesting parts of their transmissions.

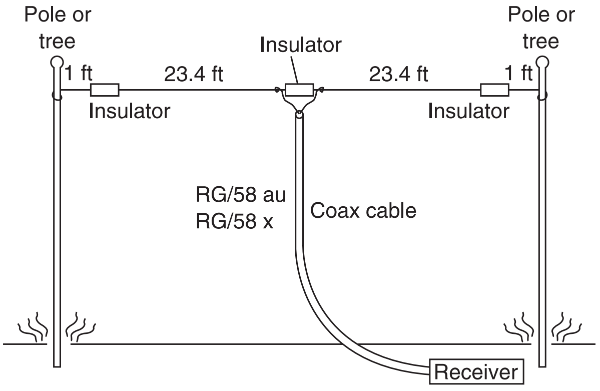

Figure 10-6 *MHz dipole antenna*

Table 10-4

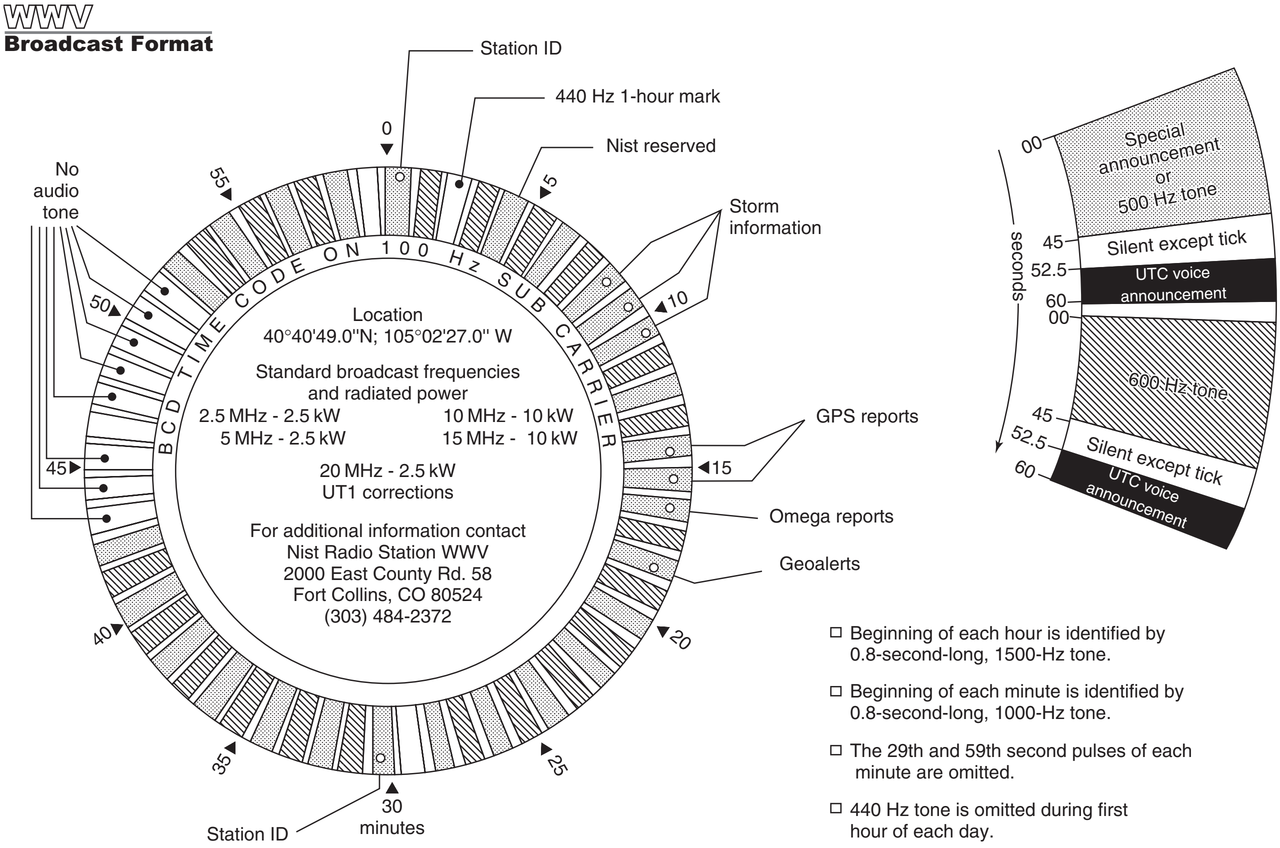
WWV
Broadcast Format
Station ID
440 Hz 1-hour mark
Nist reserved
Storm information
GPS reports
Omega reports
Geoalerts
No audio tone
Station ID
30 minutes
0
5
10
15
20
25
35
40
45
50
55
BCD TIME CODE ON 100 Hz SUB CARRIER
Location
40°40'49.0"N; 105°02'27.0" W
Standard broadcast frequencies and radiated power
2.5 MHz - 2.5 kW
5 MHz - 2.5 kW
10 MHz - 10 kW
15 MHz - 10 kW
20 MHz - 2.5 kW
UT1 corrections
For additional information contact
Nist Radio Station WWV
2000 East County Rd. 58
Fort Collins, CO 80524
(303) 484-2372
seconds
00
45
52.5
60
00
45
52.5
60
Special announcement or 500 Hz tone
Silent except tick
UTC voice announcement
600 Hz tone
Silent except tick
UTC voice announcement

☐ Beginning of each hour is identified by 0.8-second-long, 1500-Hz tone.

☐ Beginning of each minute is identified by 0.8-second-long, 1000-Hz tone.

☐ The 29th and 59th second pulses of each minute are omitted.

☐ 440 Hz tone is omitted during first hour of each day.

There are various tones which can be used to check audio equipment or musical instruments and they can be captured to detect the start of each hour or each minute. At certain intervals, voice announcements are made for the benefit of other government agencies.
The predominant ones are as follows:

Marine storm warnings are prepared by the National Weather Service and broadcast for areas of the Atlantic and Pacific Oceans which are of interest to the US Government.

Global Positioning System status announcements are prepared by the Coast Guard to give current status information about the GPS satellites.

Omega Navigation System reports are prepared by the Coast Guard to give the status of the 8 Omega transmitting stations in the 10–l4 kHz frequency range. These serve as navigation aids.

Geophysical alerts are prepared by the Space Environment Service Center of NOAA. They are broadcast on the 18th minute of the hour and give information of interest to amateur radio operators, as well to as various scientific organizations regarding solar activity, geomagnetic fields, solar flares, and other geophysical statistics. This propagation information can help you decide when the DX will be good.

- Beginning of each hour is identified by 0.8 second-long, 1500 Hz tone.
- Beginning of each minute is identified by 0.8 second-long, 1000 Hz tone.
- The 29th and 59th second pulses of each minute are omitted.
- 440 Hz tone is omitted during the first hour of each day.

You will have fun learning about other interesting things on WWV at certain times besides just the time function feature. Features such as accurate tone frequencies, geophysical alerts (solar activity reports), marine storm warnings, and global positioning system (GPS) and Omega navigation system status reports.

Chapter 11

VHF Public Service Monitor (Action-Band) Receiver

Parts list

VHF Public Service Monitor Receiver:

R1,R2,R3 10k ohm potentiometer

R4,R8,R13,R21 47k ohm, $^{1}/_{4}$w, 5% resistor [yellow-violet-orange]

R5 470 ohm, $^{1}/_{4}$w, 5% resistor [yellow-violet-brown]

R6,R9,R20 270 ohm, $^{1}/_{4}$w, 5% resistor [red-violet-brown]

R7,R17,R19 10k ohm, $^{1}/_{4}$w, 5% resistor [brown-black-orange]

R10 1 megohm, $^{1}/_{4}$w, 5% resistor [brown-black-green]

R11,R12 1k ohm, $^{1}/_{4}$w, 5% resistor [brown-black-red]

R14 33k ohm, $^{1}/_{4}$w, 5% resistor [orange-orange-orange]

R15 470k ohm, $^{1}/_{4}$w, 5% resistor [yellow-violet-yellow]

R16 100k ohm, $^{1}/_{4}$w, 5% resistor [brown-black-yellow]

R18 18k ohm, $^{1}/_{4}$w, 5% resistor [brown-gray-orange]

R22 2 ohm, $^{1}/_{4}$w, 5% resistor [red-black-gold]

C1,C6,C22 100 pF, 35 vdc disk capacitor [marked 100, 101, or 101K]

C2,C5 10 pF, 35 vdc disk capacitor

C3,C4 56 pF, 35 vdc disk capacitor

C7,C8,C11,C20 .001 µF, 35 vdc disk capacitor [.001 or 102 or 1 nF]

C9,C10 15 pF, 35 vdc disk capacitor

C13,C16,C17,C21 .01 µF, 35 vdc disk capacitor [marked .01 or 103 or 10 nF]

C12 8.2 pF, 35 vdc disk capacitor

C14,C32,C33,C35 100 to 220 µF, 35 vdc electrolytic capacitor

C15,C30,C36 4.7 µF, 35 vdc electrolytic capacitor

C18 220 pF, 35 vdc disk capacitor [marked 220 or 221]

C19 22 pF, 35 vdc disk capacitor

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

C23,C24,C25 .001 μF, 35 vdc disk capacitor [.001 or 102 or 1 nF]

C26,C28,C29,C31 .01 μF, 35 vdc disk capacitor [marked .01 or 103 or 10 nF]

C34 .1 μF, 35 vdc disk capacitor [marked .1 or 104]

C27 1.0 μF, 35 vdc electrolytic capacitor

L1,L2 2-turn coil 5/32” OD dia - #24 ga. Insulated magnet wire

L3 slug-tuned $3\frac{1}{2}$ turn coil, see Table 11-3

L4 450 kHz shielded quadrature coil, see Table 11-3

D1 Varactor diode, BB505 [orange body marked BB505]

D2 1N4148 signal diode

Q1 2SC2498 or 2SC2570A transistor

Q2,Q3,Q4 2N3904 transistor

U1 SA602 8-pin IC

U2 MC3359 18-pin FM receiver IC

U3 LM386 8-pin audio amplifier IC

Y1 10.24 MHz Crystal

FL1 10.7 MHz ceramic filter [brown, molded, 3 leads]

FL2 450 kHz ceramic filter [black, square]

S1 DPDT push switch

J1 PC mount RCA jack

J2 PC mount subminiature speaker jack

B1 9 volt transistor radio battery

ANT telescoping antenna

Misc PC board, battery holder, battery clip, speaker, etc.

Keep an ear on the local action from your easy chair with the VHF Public Service Receiver project, shown in Figure 11-1. The VHF Public Service Radio will allow you to tune the VHF high-band Police and Fire band, marine band, weather band, as well as railroad frequencies. The receiver tunes any 5 MHz portion of the 136 to 175 MHz band. Excellent performance, less than 1 uV sensitivity. The receiver features: tuned input, low noise pre-amp stage with true dual-conversion superhet design with 2 pole Ceramic High Intermediate Filter or I-F and 6 pole ceramic low I-F filters! The receiver also features front panel volume, squelch, and tuning controls. This is a great inexpensive receiver to put on your desk at home or to “watch” local radio activity while you are away from home on vacation or business trip.

Circuit description

The heart of the VHF Public Service Receiver shown in the block diagram of Figure 11-2 is the MC3359

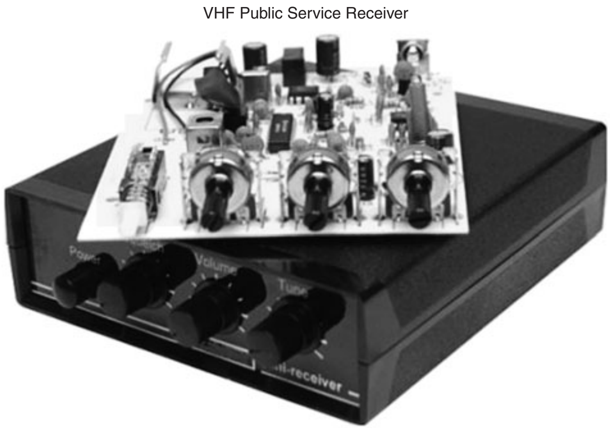

Figure 11-1 *VHF public service receiver*

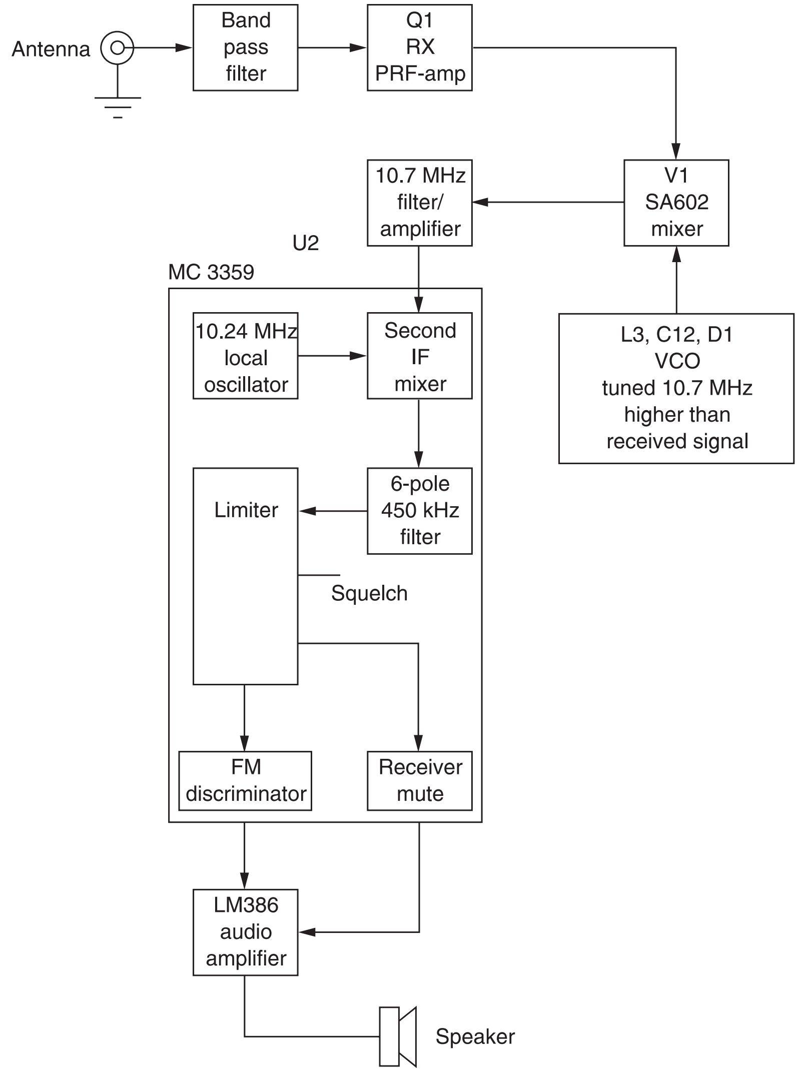
Antenna
Band pass filter
Q1 RX PRF-amp
10.7 MHz filter/ amplifier
V1 SA602 mixer
U2
MC 3359
10.24 MHz local oscillator
Second IF mixer
L3, C12, D1 VCO tuned 10.7 MHz higher than received signal
6-pole 450 kHz filter
Limiter
Squelch
FM discriminator
Receiver mute
LM386 audio amplifier
Speaker

Figure 11-2 *VHF monitor receiver block diagram*

integrated circuit. The integrated circuit is a complete high gain low power FM IF sub-system which includes an oscillator, a mixer, a limiting amplifier, AFC, quadrature discriminator, op-amp, squelch, scan control and mute switch all combined in a single IC. The MC3359 was designed to detect narrow-band FM signals using a 455 KHz ceramic filter.

Let's get into more details of the VHF Public Service Receiver by examining the schematic diagram shown in Figure 11-3. VHF signals from the antenna are amplified through the tuned input circuit (L1, C3 and L2, C4) by Q1, a microwave bipolar transistor. Q1's output is fed to the input of the SA602 IC, an efficient single-package (8-pin DIP) mixer-product detector-oscillator. The tuneable oscillator section of the SA602 is aligned to operate at 10.7 MHz higher than the signal fed and amplified by Q1. For example, to receive 144-148 MHz signals, the oscillator must tune 154.7 to 158.7 MHz in order for the SA602's mixing capability to produce a steady 10.7 MHz output signal to the rest of the circuit. The oscillator frequency is determined by L3 and its associated capacitors, and varied by the varactor tuning network using D1 and varied by R1.

The output from pin 4 of the SA602 passes through a ceramic 10.7 MHz filter, amplified by transistor Q2 and applied to input pin 18 of U2. Q3 provides AFC (Automatic Frequency Control) by keeping the local oscillator of U1 from drifting away from an incoming signal. This is accomplished by tuning the varactor circuit in the direction opposite the drift.

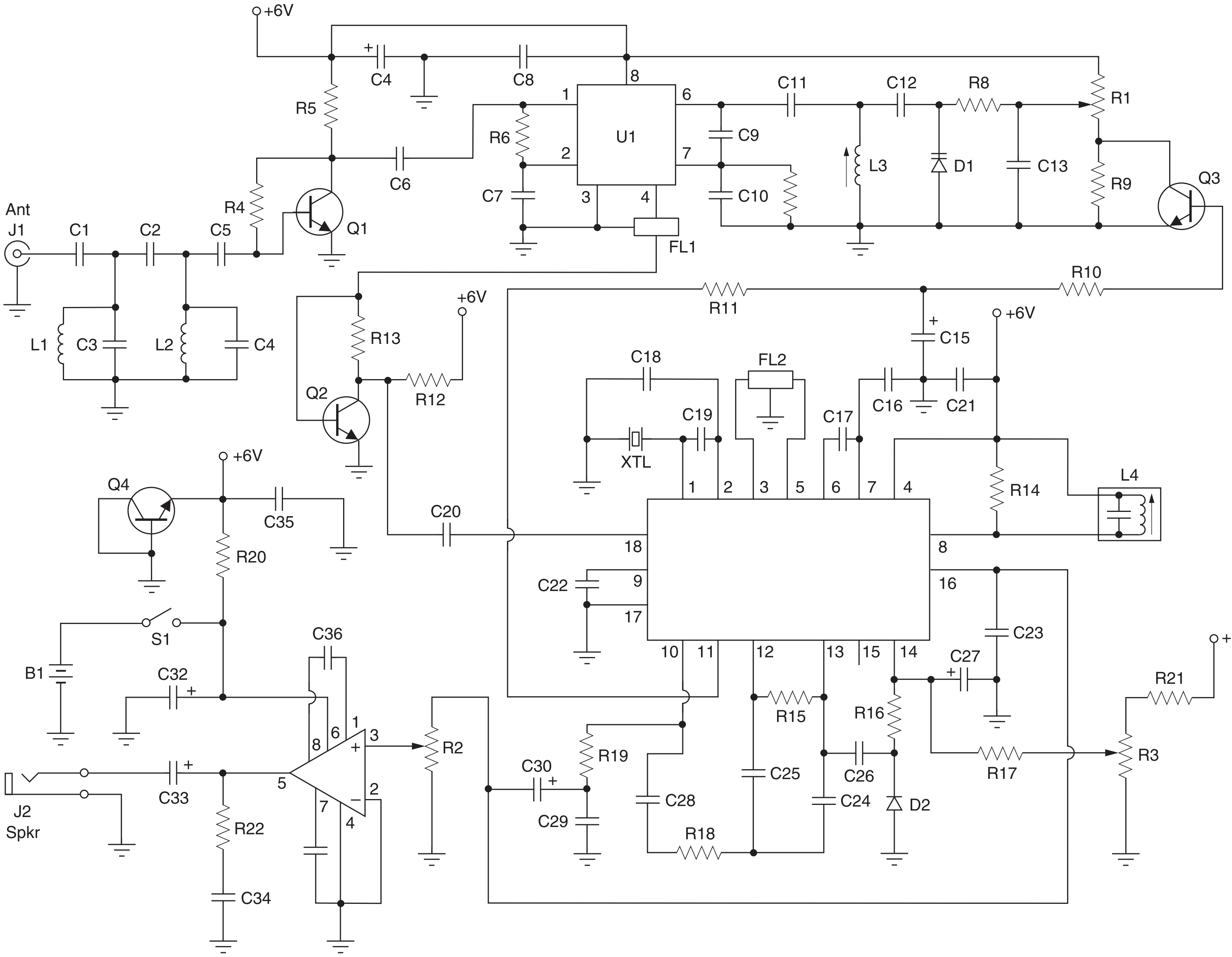
+6V
C4
C8
8
1
6
C11
C12
R8
R1
R5
R6
U1
C9
2
7
C6
C7
3
4
C10
L3
D1
C13
R9
Q3
Ant
J1
C1
C2
C5
R4
Q1
FL1
R10
R11
L1
C3
L2
C4
R13
+6V
+6V
C15
R12
C18
FL2
C16
C21
Q2
C19
C17
XTL
1
2
3
5
6
7
4
R14
L4
+6V
Q4
C35
C20
18
8
R20
C22
9
16
17
S1
C36
10
11
12
13
15
14
C23
C27
+6V
B1
C32
R15
R16
R21
1
8
6
3
R2
R19
C30
C26
R17
R3
5
2
C28
C25
C24
D2
J2
Spkr
C33
R22
7
4
C29
R18
C34

Figure 11-3 *VHF public service block diagram*

The MC3359 IC has an internal oscillator controlled by the 10.24 MHz crystal. The 10.24 MHz signal is mixed with the 10.7 MHz input from U1 to convert down to the low IF of 450 kHz. The 450 kHz IF is filtered by FL2 and then amplified by a limiting amplifier in U2. Audio demodulation takes place in the quadrature detector, with L4 adjusting the detector. The audio modulation is detected in U2, and the background noise is used to control the squelch. U3 amplifies the audio output from pin 10 of U2 to a practical level for speaker or headphone operation. Pin 16 of U2 will ground the input of U3 when the squelch is closed. L4, a 450 kHz IF coil, permits adjustment of the 90 ° voltage-current phasing ("quadrature") of FL2's output to the FM detector demodulator (pin 8 of the MC3359).

Circuit assembly

Look for a clean well lit area in which to assembly your shortwave receiver. You will need a 27 to 30 watt pencil tip soldering iron, some 60/40 tin/lead rosin core solder, a few hand tools, including a pair of end-cutters, a needle nose pliers, a few screwdrivers and a pair of tweezers to pick up small parts. Locate your schematic and layout diagrams and we will begin to construct your new shortwave receiver. Let your soldering iron tip heat both the component lead wire and PC board trace enough so that the wire itself AND the foil trace BOTH become hot enough TOGETHER to melt a bit of solder so that it flows smoothly from the pin to the PC board trace. Locate the printed circuit board and all the components and we will start.

Locate all of the resistors for the project and place them in front of you. Each resistor has a color code which describes its value. The first color band at one edge of the resistor body represents the resistor's first digit value, while the second color band denotes the second digit of the resistor value. The third color band represents the multiplier value of the resistor. A fourth color band is the tolerance value. A silver band denotes a 10% tolerance, while a gold band denotes a 5% tolerance. The absence of a fourth color band denotes a 20% tolerance value. Check the color code against the parts list and schematic before installing them onto the circuit board. Refer to the resistor color chart in Table 11-1. Place about four or five resistors on the PC board and place these resistors in their respective locations carefully making sure each one is in the correct location. Next, solder each of the resistors onto the circuit board. Trim off the excess component leads with a small pair of end-cutters. Pick up and identify the next grouping of resistors and install them on the circuit board, follow up by soldering this next batch of resistors

Table 11-1
Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

to the circuit board; remember to trim the extra lead lengths form the PC board.

Once all the resistors have been installed onto the circuit board, we can move on to identifying all of the capacitors for the project. Note that there will be different looking capacitors for this project, some will be small disks, while others will be larger body devices with minus and plus marking on them. These larger capacitors with polarity marking are electrolytic capacitors and they have polarity which must be observed for the circuit to work correctly. Look through the parts pile and locate the capacitors and place them in front of you. You will notice a number of small disk capacitors, we will install these first. Look closely at each of the disk capacitors at their markings. Sometimes the capacitors are very small and they may not have their actual value printed on them but will instead have some sort of code marked on them. Refer to the parts list for these codes or the capacitor code chart in Table 11-2. Locate four or five of these small disk capacitors, identify them and install them on the circuit board, while referring to the schematic and layout diagrams. After inserting the disk capacitors onto the circuit board, you can go ahead and solder them to the circuit board. Remember to cut the excess component leads. Next, move on and install another grouping of small disk capacitors, then solder them in place on the printed circuit board. Cut the extra component leads after soldering in the capacitors. When you are finished installing the small capacitors we will move on to the larger electrolytic types.

Capacitors C14, C32, C33 and C35 are all 100 uF to 200 µF electrolytic types. These capacitors will have either a plus or minus marking and/or black band denoting polarity. You must orient these capacitors with respect to their polarity markings. Check both the layout diagram and the schematic diagram for polarity references. Note also that capacitors C15, C30 and C36 are also electrolytic capacitors with a 4.7 µF value. These too must be oriented correctly for the circuit to operate correctly when power is applied. It is possible to actually damage the circuit upon power-up if these capacitors have been installed incorrectly, so pay careful attention when mounting these capacitors. Install the electrolytic capacitors in groups of two or three. Once they have been placed on the circuit board, you can move on to soldering them in place. Remember to trim the excess component leads from the PC board after soldering.

Next we will move to installing the two diodes. Diode D1 is the varactor "tuning" diode while D2 is a 1N4148 silicon diode. When mounting the diodes be careful to observe polarity markings on the diodes. Usually each diode will have a black or white band at one end of the diode body. This color band denotes the cathode end of the diode. Refer to the schematic diagram and layout diagrams in order to see how the diodes should be mounted with respect to polarity considerations.

Before we go ahead and install the semiconductors, take a look at the semiconductor pin-out diagram in Figure 11-4 which will help you orient the components. Now locate the transistors from the components. Transistor Q1 is a 2SC2498 or 2SC2507A, while transistors Q2, Q3 and Q4 are 2N3904 NPN types. Transistors have three leads, a Base lead, a Collector lead and an Emitter lead. The symbol for the Base is a vertical line, while the Collector and Emitter are slanted lines pointing to the Base lead. The Emitter lead is shown with a small arrow pointing towards or away from the Base lead. Refer to the schematic and layout diagrams to determine the orientation of the transistors. Clearly identify Q1, the 2SC2498 or 2SC2570A transistor. Do not confuse it with the other transistors supplied. Position Q1 as shown on the PC board layout, with the flat side facing to the right, toward the middle of the board. Press the transistor snugly into the PC board so that only a minimum amount of wire lead is exposed above the board. In soldering, don't be afraid to use enough heat to make good clean connections.

Once the transistors are in place on the PC board, you can go ahead and solder them to the foil side of the PC board. Don't forget to trim the extra component leads.

Now we are going to install FL1, the ceramic filter. This component looks like a capacitor with three leads and may be installed either way. Next go ahead and locate FL2, the 450 kHz filter. Its three leads are delicate and fit in only one way. At this time you can locate Y1, the 10.24 MHz crystal. No special procedure is required. Simply press the crystal firmly into its holes as far as it will go. Now go ahead and solder in the crystal and ceramic filters, remove the excess leads if any.

Table 11-2

Capacitance code information

This table provides the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Some capacitors actually have the numeric values stamped on them; however, many are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

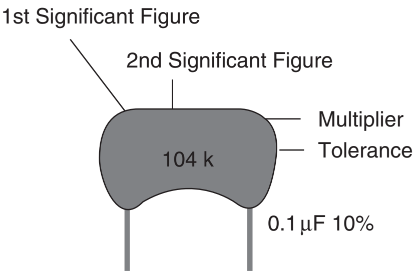

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

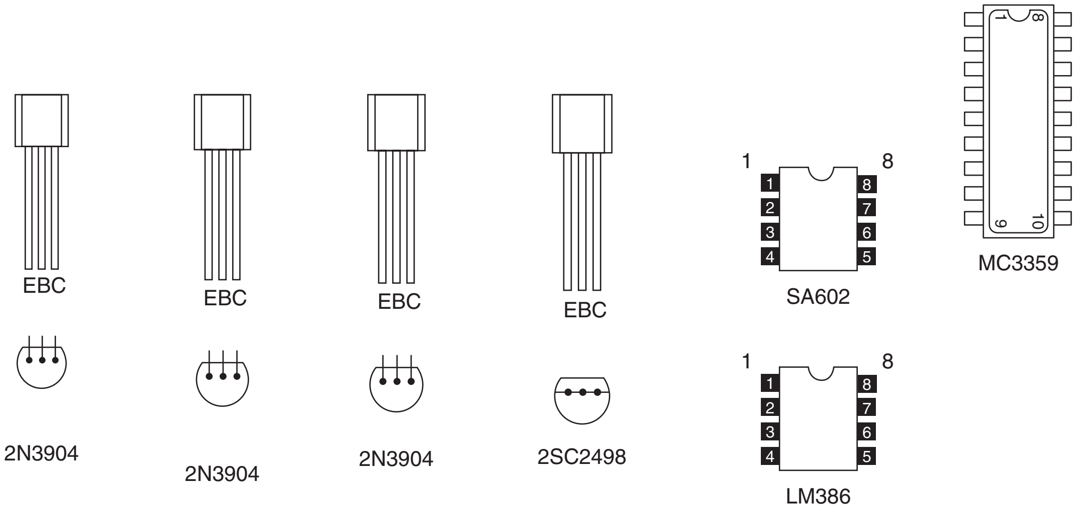

Figure 11-4 *Semiconductor pin-outs*

Look through the remaining components and locate the inductors L1 through L4. Coils L1 and L2 are small 1 and ½ turn coils. Go ahead and install coils L1 and L2, they are both the same. Now, locate and install L3, the slug tuned coil. It is important that the body of this coil be snug against the PC board for mechanical stability, which directly affects the frequency stability of the receiver. Make good, solid solder connections, when installing the above coils. Finally locate coil L4, the 455 kHz quadrature coil, soldering the two pins and the two mounting tabs. Remove any excess lead lengths if any; see Table 11-3 for coil winding information.

Let's move on to installing the SA602 IC, you may wish to use an 8-pin DIP socket rather than soldering the IC directly to the board. Reasons for doing this might include the peace of mind of beginners afraid of damaging the IC, or experienced builders testing the merits of the SA602's cousin, the SE602. However, please be aware that we have seen more service problems with improper socket insertion than from soldering heat damage to ICs. Even if this is your first IC, don't be afraid to use enough heat to make eight clean connections, but DO be sure to correctly orient the end marked by a band, dot, or notch. Before soldering, make sure that the IC or socket is perfectly flat against the top of the PC board. Install U1, the SA602, in accord with the preceding directions.

Now, install MC3359 IC, U2. As in the case of the SA602 IC, you could choose to install an 18-pin IC socket rather than soldering the IC directly. Re-read the discussion of IC sockets offered for the installation of U1. Larger ICs such as the MC3359 require considerably more care in socket insertion. Notice that the end of the IC marked by a band, dot, or notch must be oriented correctly as shown on the parts layout diagram. Solder each of the 18 connections carefully.

Table 11-3

Coil winding information

L1,L2	2-turn coil - #24 ga. Insulated magnet wire - 5/32″ OD dia. Air-core coil.
L3	3 ½ turn coil ferrite slug-tuned coil, unwind original coil from form and rewind 3½ turns of #24 ga. insulated magnet wire on form - use series 49A coil Circuit Specialists (49A127MPC).
L4	450 kHz shielded quadrature coil TOKO America Coils (7MC-81282).

Make sure you have good lighting as well as good technique to make sure that no solder "bridges" flow between the connections.

Finally locate and install U3, the LM386 IC audio amplifier integrated circuit. Once again use of an IC socket is highly recommended. Solder the IC socket to the PC board and insert the LM386 paying particular attention to the small notch or indented circle on the top of the IC body. Pin 1 is to the left of the notch, cut-out or indented circle.

Take a short well-deserved break and when we return, we will install the remaining components on the PC board and finish up building the receiver. Now locate power switch S1. Press S1 firmly into its six holes and solder all six pins. The switch fits the board only one way. Next, install J1, the RCA-style antenna jack and solder all four points. Now, install J2, the subminiature headphone jack. Solder all three points. Be gentle and patient in inserting, so as not to damage the solder tabs.

Locate all three potentiometers and we will install them now. Potentiometer controls R1, R2 and R3 are all mounted in a row next to switch S1. Insert the three PC mount potentiometers into their positions. Check that the controls are pressed in firmly and straight against the top of the board. Solder the three center pins and then the two larger mechanical mounting tabs. Use enough solder for a solid connection.

The circuit board may require a few jumpers; "cut" component leads make good jumpers. Insert the jumpers and solder into place, remember to trim the excess leads. We are now down to the home-stretch. Install the battery snap terminal in the two holes below S1, making sure that the positive (red) lead is inserted into the (+) hole on the PC board.

Finally, go ahead and mount the battery bracket, it may be attached in a variety of ways. A wire jumper can be passed through the two holes on the PC board, then both ends soldered on the underside of the board. You may prefer to use very small screws or even a double-sided mounting adhesive strip or glue. In using such methods, make sure that the battery will not block the PC board's mounting hole in its vicinity.

Wiring of the PC board is now complete. If you have the patience, we suggest a short break. Then take a magnifying lens and a bright light and examine all your solder joints, touching up any connection which appears less than perfect. Make sure all excess leads have been trimmed, and that one is not bent back flat against the board, possibly causing a short. Brush the solder side of the board with a stiff brush to make sure that no loose wire trimmings or solder drippings are lodged between connections. Connect up a suitable antenna and connect up the 9 volt battery to the receiver and you are ready to try out the receiver for the first time.

After power is first applied and you have connected up your antenna, you will need to adjust the gain control to midrange. Next, you should adjust the squelch control just below the threshold point, so that you can hear the normal background hiss. Now begin turning the tuning control and you should soon come across a public service transmission from a local police or fire broadcast.

In the event that you do not hear any hiss coming from the speaker or headphones, or you do not hear any stations after about 20 to 30 minutes, then you may have to troubleshoot the receiver for any errors that may have occurred during the building process. First, disconnect the battery and antenna and un-mount the receiver circuit board from the enclosure. Have a friend assist you in inspecting the circuit board; a different set of eyeballs can often help to find errors quickly.

The most common causes for circuit failure are resistors placed in the wrong locations, the second possible cause for errors are improper installation of diodes and electrolytic capacitors. Make sure that you re-check the orientation of these components one more time and refer to the schematic, and parts layout diagrams. Another possible cause for mistake is the improper installation of transistors and integrated circuits. Be sure to recheck the orientation of transistors. Each transistor has three leads, an Emitter, a Collector and a Base lead. Closely inspect and compare the schematic diagram along with the parts layout diagram and manufacturer's specifications sheets to ensure that you have oriented the transistors correctly. Lastly, integrated circuits are often placed backwards in their sockets. After you have checked all the above tips, make sure that the battery is connected correctly with the black battery or minus lead (–) connected to the circuit ground and the plus (+) or red battery lead connected to the power input regulator of the circuit. Reconnect the battery and the antenna and then switch the receiver to the "On" position and retest the receiver once again to make sure that it is now working.

Testing, alignment and adjustment

In order to prepare the VHF FM Receiver for reliable monitoring operation, you will need a few basic tools, such as useful VHF signal source, a hexagonal, non-metallic coil slug alignment tool for L3, and a small screwdriver to adjust L4. To align your new receiver, you will need to locate some sort of a signal source. This signal can be from your test bench equipment or from a "live" source such as a local repeater. The "live" signals are best for fine-tuning, but they also have that VHF communications character of being fast and to the point. Good test bench signal sources include your own 2-meter transceiver, if you are licensed to use it, a signal generator or grid-dip oscillator. If you do not have a steady test signal source for the band within 135–175 MHz of primary interest to you, and still wish to verify immediately the proper operation of your receiver, a good and steady VHF signal source will be your local NOAA Weather Service station, broadcasting around 160 MHz. Once you have verified reliable reception of your regional NOAA station, adjustment to your desired listening range will become easier.

If you do not already have a set of plastic or nylon coil alignment tools and do expect to try further ham radio or electronic hobby projects, such tools are worth having and can be found inexpensively at any electronics store including Radio Shack. While a metal hex key wrench can fit the coil slug, the metal itself would damage the coil inductance drastically and therefore should not be used. With patience and sandpaper, a useable tool might be formed from a wood or plastic rod.

The alignment procedure should begin as follows: first make sure the power switch is off. Next, connect the antenna, earphone or speaker and battery, then turn the slug of L3 until it is even with the top of the coil form. Then turn the slug of L3 clockwise seven turns back into the form.

Now, turn the slug of L4 until it is flush with the top of the coil and then turn it two turns back into the coil. Next turn all three controls to the left, fully counterclockwise. Now, you can turn the power switch to "On." Now turn R2 (volume control) until you hear some noise. Finally, adjust L4 for maximum noise from the speaker.

Further alignment now consists of adjusting the oscillator coil L3 to permit the tuning control (R1) to cover the 5 MHz segment between 135 and 175 MHz of primary interest to you. Assuming you wish to adjust for the 2-meter 144–148 MHz Amateur Band, adjust L3 until you hear your intended test signal. If you are a beginner with no license or other equipment, any ham operator with a 2-meter transceiver should be willing to give you the test signal and extra help that you need. Your new VHF receiver is very sensitive, so operate the transceiver on low power on a simplex frequency from a distance of at least across the room. An 8″ piece of wire will be a sufficient receiving antenna for such tests. If you don't know any hams, visit a friendly two-way radio service center to get close to the test signal you need!

Receiver sensitivity

Your FM receiver features sensitivity under 1 uv. Radio hams constantly marvel at how an FR146 displayed at hamfests tunes in dozens of hand-held QSOs on the premises without an antenna connected! You can expect to monitor local repeater and simplex transmissions easily, using a simple ground-plane style antenna.

Customization

For many people, a pilot lamp to indicate "power on" is more than a nice touch. They expect it and depend on it. Adding a simple LED power-on indicator to your VHF receiver is easy. All you need is the LED itself and a small 1k to 2.2k resistor. Study the PC traces between the positive battery supply wire and the on-off switch. The unused connectors on top of your switch are an ideal point to get the + DC voltage needed for the anode (longer lead) of the LED. Plan where and how you wish to install the LED in your enclosure.

Locating the LED immediately above the on-off switch is logical and ideal. The simplest way to make a good installation is to drill a neat hole just slightly smaller than the diameter of the LED. Then, enlarge the hole a little bit at a time just enough to let the LED be pressed in and held firmly. The resistor may be connected to either the anode or cathode of the LED, but the anode MUST go to + DC, with the cathode connected to the nearest common ground point.

Antenna considerations

For local VHF reception, the VHF receiver will operate fine just using a whip antenna. If you wish to receive stations which are further away than just the local stations, you may want to install some sort of an outdoor antenna such a ground plane or Discone antenna for VHF operation. Since most of the transmissions in this band are of vertical polarization, vertical antennas, ground plane and Discone antennas will work best in this band. Various VHF scanner type antennas from your local Radio Shack store will work fine. The chart illustrated in Table 11-4 shows the utilization of the VHF Public Service band. Have fun and enjoy you new VHF "action" band receiver.

Table 11-4

VHF high-band frequency band utilization

150.995 MHz to 151.130 MHz	Highway maintenance
151.145 MHz to 151.475 MHz	Forestry conservation
152.030 MHz to 152.240 MHz	Mobile telephone
152.270 MHz to 152.450 MHz	Taxi services
152.510 MHz to 152.810 MHz	Mobile telephone
153.410 MHz to 154.115 MHz	Utilities/light/power/water and local gov't
153.830/153.890/154.010/154.070 MHz	Fire
154.130 MHz to 154.445 MHz	Fire
154.650 MHz to 155.700 MHz	State & Local police
155.715 MHz to 156.255 MHz	Local Gov't / police/ highway maintenance
156.275 MHz to 157.425 MHz	Marine Frequencies
157.530 MHz to 158.265 MHz	Taxi & Mobile telephones
158.280 MHz to 158.460 MHz	Utilities water/power/industrial
158.730 MHz to 159.210 MHz	Highway Maintenance & police
159.225 MHz to 159.465 MHz	Forestry
159.495 MHz to 160.200 MHz	Motor carrier trucks
160.215 MHz to 161.565 MHz	Railroad
161.640 MHz to 161.760 MHz	Broadcast pickups & Studio to Xmiter Links
161.800 MHz to 162.00 MHz	Maritime Shore Stations
162.400 MHz to 162.550 MHz	NOAA weather broadcasts
162.250 MHz to 173.500 MHz	Gov't sonobuoys
169.00 MHz to 172.000 MHz	Low power wireless mikes
173.2037 MHz to 173.396 MHz	Fixed/mobile industrial & remote control
173.400 MHz to 173.5000 MHz	Remote control & telemetry

Chapter 12

6 & 2-Meter Dual-Band Amateur Radio Receiver

Parts list

6N2 Meter Amateur Radio Receiver

R1 68 k, ¼w, 5% resistor

R2 47 k, ¼w, 5% resistor

R3, R7 3.3k, ¼w, 5% resistor

R5 20 k 5% ¼W

R8 8.2k, 5% ¼W

R9 Jumper wire (use a discarded resistor lead)

R10, R12 1k, ¼w, 5% resistor

R11 10 k, ¼w, 5% resistor

R15 10 ohm, ¼w, 5% resistor

VR1, VR3 100 k potentiometer (chassis mount)

VR2 5k, 20% linear potentiometer (chassis mount)

C1 120 pF, 25v capacitor (121)

C2 .005 µF, 35v 20% capacitor (502)

C3,C4,C9 .1 µF +80, −20% 50v capacitor (104)

C5 3.9 pF, 50v disk capacitor

C6 33 pF, 50v, 10% disk capacitor

C7 470 pF, 50v, 10% 50 v disk

C8 7 pF, 25v, 20% disk capacitor

C10 1000 pF, 50v, 10% disk capacitor (102)

C11,C13,C16 .01 µF, 50 v, +80, −20% capacitor (103)

C17 51 pF, 25v, 10% capacitor (51K)

C12 68 pF, 25v, 10% capacitor (68K)

C14, C21 4.7 µF, 50v Radial capacitor

C15 220 µF, 16v Radial capacitor

C18,C19,C20 .1 µF, 50v, +80, −20% capacitor (104)

C22 .047 µF, 50v, +80, −20% capacitor (473)

D1, D2 Diode 1N914 silicon diode

U1 IC MC3362P FM Receiver IC

U2 IC LM386N-1 Audio Amp IC

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

U3	IC 78L05 Regulator +5v 78L05
L1	Coil (4.5 turns coated wire- #24 ga.) -air-core
L2, L3	Coil (on form with core)
L4	Coil (1.5 turns coated wire- #24 ga.) -air-core
L5	.64 mH Coil (yellow) I-F Transformer (42IF301)
F1	455 kHz Ceramic Filter
F2	10.7 MHz Ceramic Filter
Y1	10.245 MHz Crystal
SW1, SW2	Switch Slide DPDT
B1	9 volt transistor radio battery
SPKR	8 ohm speaker
J1	RCA jack - antenna
Misc	PC Board, battery holder, battery clip,

The 6N2 receiver is a dual conversion, dual band amateur radio receiver; it combines the excitement of two very active amateur radio bands in one small receiver, see Figure 12-1. The 6N2 receiver will allow radio enthusiasts, both the young and old, to get a glimpse of the active world of VHF amateur radio. This receiver combines both the 2-meter and the 6-meter bands which are popular with radio enthusiasts around the country. The 2-meter band is probably the most active ham radio band especially for newcomers with a technician class license. Two-meter band coverage is usually quite good in most parts of the country. The 2-meter band features many ham radio repeaters, and in some have wide-area long-range repeaters in most US states and in many Canadian provinces. These ham radio repeaters allow amateur radio operators to communicate over long distances. The 6-meter band has often been called the “magic band.” This amateur band is usually only used for fairly local coverage but when the band is “open” you can often hear hams on the opposite side of the US or in Cuba or the Caribbean, which is really exciting!

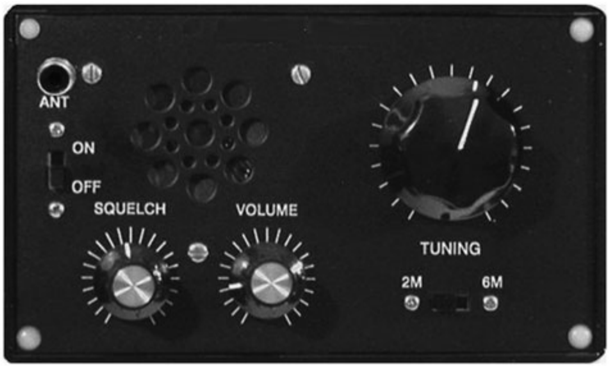

Figure 12-1 *6N2 meter amateur radio receiver*

You can listen-in to these exciting two ham radio bands and hear all the action, day or night with the 6N2 portable receiver. You may just decide the action sounds like something you would like to be part of, and you can easily get your own ham radio “ticket” or license. Contact the Amateur Radio Relay League (ARRL) at http://www.arrl.org; they have many books, publications as well as introductory videos. They are a great resource in locating local ham radio clubs, where you can make friends and learn a lot about ham radio and communication.

The 6N2 radio is known as a narrow band FM VHF receiver. FM of course is a type of modulation and is denoted as Frequency Modulation. Frequency modulation means the data or voice changes the frequency of the radio wave. The other popular forms of modulation are AM (Amplitude Modulation) and PM (Phase Modulation). We also mentioned narrow band, which notes that the selectivity of the entire system is limited to only enough frequencies to pass voice or low frequency data. A commercial broadcast FM receiver would have a bandwidth large enough to pass music and high frequency data transmissions.

The 6N2 receiver is a dual conversion, two-band radio receiver. In a dual conversion receiver, the original radio frequency is converted first to a 10.7 MHz intermediate frequency (I-F) and amplified in (Block 4). The 10.7 MHz signal is then converted to 455 kHz and amplified in (Block 3). Noise is removed and the

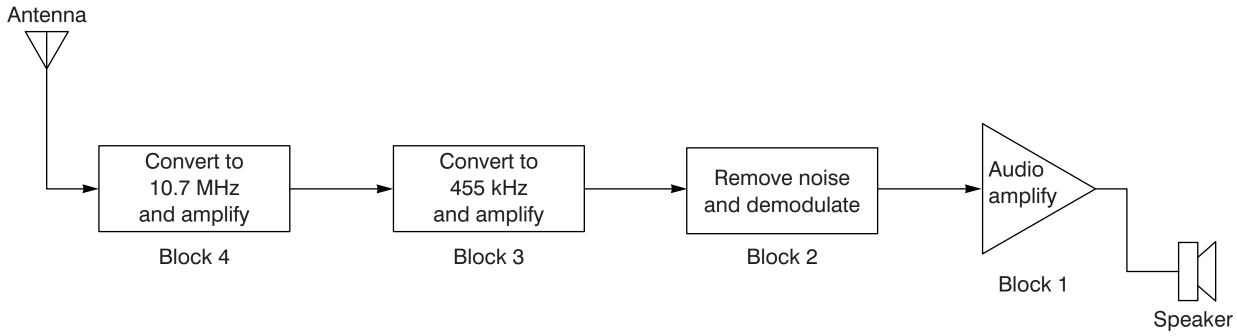
Antenna
Convert to 10.7 MHz and amplify
Block 4
Convert to 455 kHz and amplify
Block 3
Remove noise and demodulate
Block 2
Audio amplify
Block 1
Speaker

Figure 12-2 *6⁄2 meter amateur radio receiver block diagram*

modulated signal is recovered from the 455 kHz I-F signal in (Block 2). The audio signal is amplified to drive a speaker in (Block 1), see Figure 12-2. The 6N2 receiver specifications are illustrated in Table 12-1.

So you ask, why dual conversion? To answer this question, take a look at the diagram shown in Figure 12-3, and it may become more clear. In order to obtain the desired intermediate frequency (I-F) signal by mixing the local oscillator with the desired radio frequency, an unwanted output may result due to a transmission spaced one intermediate frequency or (I-F) on the opposite side of the oscillator (image). By using a large I-F frequency, this image is moved further out of the band of desired frequencies. The second conversion provides the selectivity to filter out the desired narrow band transmission.

The 6N2 receiver operates on two of the most popular VHF amateur radio bands. The 6N2 VHF amateur radio receiver receives both the 6-meter ham band and the 2-meter ham bands, and you can select the band with a band-switch.

To help you visualize the relationship between frequency and wavelength, take a look at the diagram shown in Figure 12-4. If the speed of a wave (meters per second) is divided by the number of waves that pass a given point (cycles per second), the seconds cancel and you obtain the wavelength λ = meters per cycle. The speed of radio waves is approximately 300,000,000 meters per second. If the frequency is 50 to 54 MHz, the wavelength becomes 300,000,000/50,000,000 or 6 meters. So for the 2-meter band, which is approximately 150 MHz, the wavelength becomes 300,000,000/150,000,000 or 2 meters. The actual bands are: 2-meter band which covers 144 MHz to 148 MHz, and the 6-meter band which covers 50 MHz to 54 MHz.

The 6N2 dual band VHF receiver is composed of four major blocks, i.e. blocks 1 through 4. It is a bit easier to follow how the receiver works if we break it down into smaller defined blocks, see Figure 12-5. The heart of the receiver is the integrated circuit at U1, an almost complete FM receiver on-a-chip. The FM receiver chip only requires timing components, a crystal and RF tuning components, a filter. The output of the FM radio chip is sent to the audio amplifier chip at U2. Power for the receiver is regulated via the regulator chip at U3.

Table 12-1
The 6N2 amateur radio specifications

2-meter band covers 144 MHz to 148 MHz
6-meter band covers 50 MHz to 54 MHz

- Single supply voltage (4-12 v)
- Idle current - 4 milliamps
- Inputs referenced to ground
- Input resistance - 50 k
- Self-centering output voltage
- Total harmonic distortion less than 0.2%
- Output power with 9 volt supply voltage
- Voltage gain with 10 µF from pin 1 to 8 - 200 or 46 dB
- Voltage gain with pins 1 and 8 open - 20 or 26 dB
- Bandwidth with pins 1 and 8 open - 300 kHz

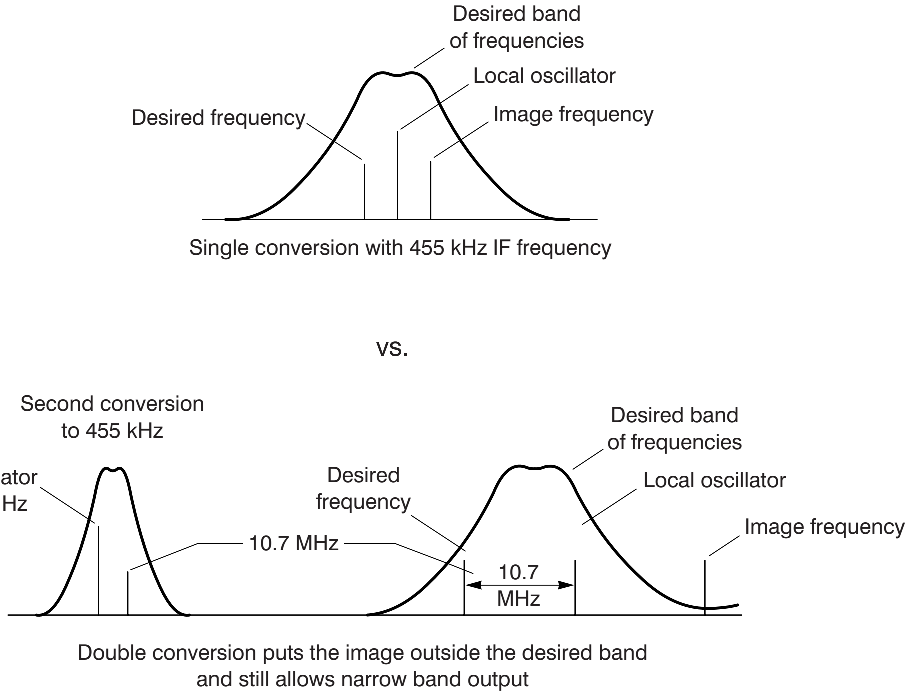
Desired band
of frequencies
Local oscillator
Desired frequency
Image frequency
Single conversion with 455 kHz IF frequency
vs.
Second conversion
to 455 kHz
Crystal oscillator
at 10.245 MHz
10.7 MHz
Desired
frequency
Desired band
of frequencies
Local oscillator
Image frequency
10.7
MHz
Double conversion puts the image outside the desired band
and still allows narrow band output

Figure 12-3 *Single vs. double conversion receiver*

The receiver is relatively straightforward and can therefore be constructed in stages, one through four. We will begin with Stage 1, which can be built and tested separately. You can progress step by step and build and test each stage before moving on to the next stage. Let's begin with Block 1, the audio amplifier stage.

The audio amplifier – block 1

The audio section in this receiver is amplified by using an integrated circuit audio power amplifier. The output

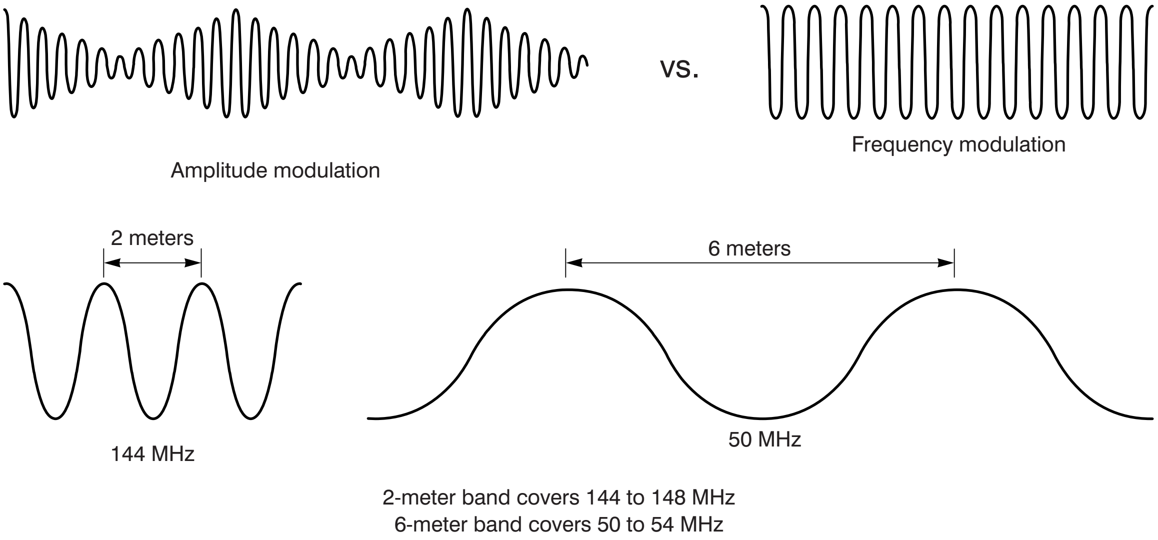
vs.
Amplitude modulation
Frequency modulation
2 meters
144 MHz
6 meters
50 MHz
2-meter band covers 144 to 148 MHz
6-meter band covers 50 to 54 MHz

Figure 12-4 *Frequency coverage diagram*

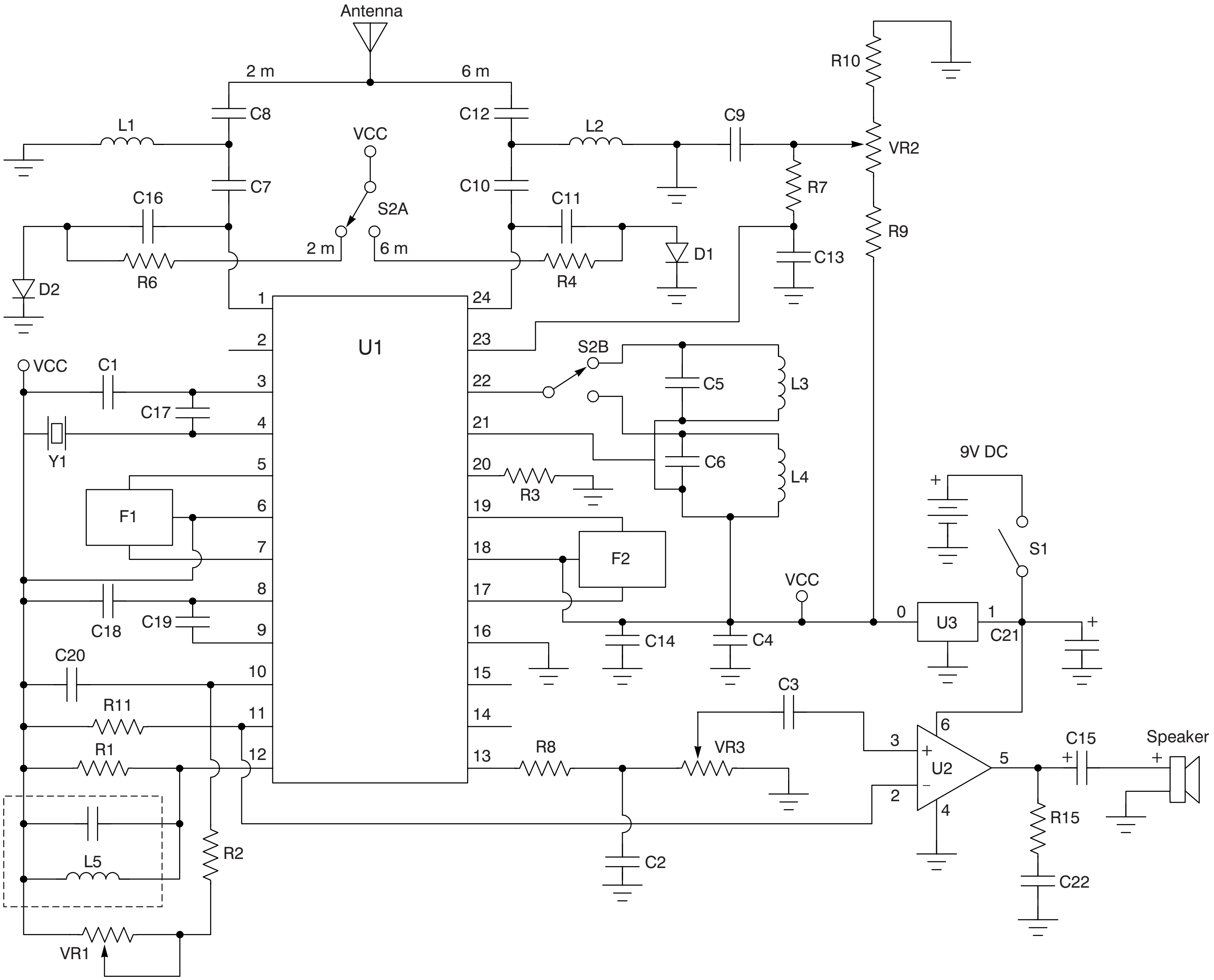
Antenna
2 m
6 m
C8
C12
L1
L2
C9
R10
VR2
VCC
C7
C10
S2A
R7
C16
C11
R9
2 m
6 m
D1
C13
D2
R6
R4
1
24
U1
2
23
S2B
VCC
C1
3
22
C5
L3
C17
4
21
Y1
5
20
C6
L4
R3
F1
6
19
9V DC
7
18
F2
S1
8
17
VCC
C18
C19
9
16
0
U3
1
C14
C4
C21
C20
10
15
C3
R11
11
14
6
R1
12
13
R8
VR3
3
C15
Speaker
U2
5
2
4
L5
R2
C2
R15
C22
VR1

Figure 12-5 *VHF public service receiver schematic. Courtesy of Elenco Electronics*

impedance of the amplifier is low enough to drive an 8 ohm speaker directly. The coupling capacitor value is picked to pass audio signals down to 100 cycles by matching the reactance of the capacitor with the speaker impedance.

Due to the high input resistance of the amplifier (50 k), the audio coupling capacitor C3 can be as small as 0.1 µF. The equivalent resistance at the junction of R8 and VR3 is approximately 6.6 k (the parallel combination of R8, VR3 and the 50 k input impedance of the LM-386). The capacitor C2 and this equivalent resistance sets the 3 dB corner used to attenuate any IF voltage at pin 13. A simple RC filter attenuates at a rate of 6 dB per octave (an octave is the same as doubling the frequency). By using 6.6 k as the equivalent resistance and 0.005 µF as the capacitance, we get a 3 dB corner at approximately 4.8 kHz. To get to 455 kHz, you must double 4.8 kHz approximately 6.6 times. This equates to a reduction of the IF voltage at the R8 - VR3 junction of 39.6 dB (6 dB per octave times 6.6 octaves), or 95 times.

Recovering the data – block 2

In working with FM receivers, there are some terms that need to be defined. First, let's determine the term **deviation** as the frequency swing of the incoming FM signal. When no modulation is present, the incoming signal is a fixed frequency **carrier wave (Fc)**. **Positive deviation (Fp)** is the increase in Fc due to modulation, and **negative deviation (Fn)** is the decrease in Fc due to the modulation. The detector must be linear over the maximum total deviation (Fp max. - Fn max.) produced by the transmitter or distortion will occur. Before the second I-F signal reaches the detector, it is applied to a **limiting amplifier** inside the integrated circuit. A limiting amplifier is designed to remove any amplitude variations in the signal by amplifying the signal well beyond the limit of the amplifier. The frequency variations (FM) are not affected by the limiter. The limiter removes the affects of fading (driving through a tunnel) and impulse noise (lightning and ignition). These affects produce significant unwanted amplitude variations in the received signal, but minor frequency variations. Noise immunity is one of the great advantages of frequency modulation over amplitude modulation. The remaining signal contains only the frequency modulation.

The 6N2 receiver uses a quadrature detector to demodulate the FM signal. After the noise is removed by the limiter, the signal is internally connected to the quadrature detector. A parallel tuned circuit must be connected from pin 12 to the power supply. This circuit produces the 90° phase needed by the quadrature detector. A resistor shunting this coil sets the peak separation of the detector. If the value of the resistor is lowered, it will increase the linearity, but decrease the amplitude of the recovered audio.

The quadrature detector combines two phases of the I-F signal that are 90° apart, or in quadrature, to recover the modulation. The shifted signal is used to gate the non-shifted signal in a synchronous detector. When the frequency increases above the carrier frequency (positive deviation), the phase shift increases causing a decrease in the width of the gated impulse output. In a similar manner, when negative deviation occurs, the gated impulse output will widen. The gated output is then filtered to remove the pulses and recover the modulating signal.

455 kHz conversion – blocks 3 & 4

The first local oscillator should be set at 133.3 MHz to 137.3 MHz for the 2-meter band and 39.3 MHz to 43.3 MHz for the 6-meter band. This oscillator is free-running varactor-tuned. The first mixer converts the RF input to an I-F frequency of 10.7 MHz. This I-F signal is then filtered through a ceramic filter and fed into the second mixer. If the oscillator of the second mixer is precisely set at 10.245 MHz, it will produce an output containing the sum (20.945 MHz) and the difference (0.455 MHz or 455 kHz) frequencies. This 455 kHz signal is then sent to the limiter, detector, and audio circuits. The mixers are doubly balanced to reduce spurious (unwanted) responses. The first and second mixers have conversion gains of 18 to 22 dB respectively.

Conversion gain is the increase in a signal after the signal has been converted to a new frequency and amplified. For both converters, the mixers are designed to allow the use of pre-tuned ceramic filters. After the first mixer, a 10.7 MHz ceramic bypass filter is used. This eliminates the need for special test equipment for aligning I-F circuits. The ceramic filter also has a better aging and temperature characteristic than conventional LC tuned circuits.

Assembling the 6N2 receiver

Before constructing your new 6N2 receiver, you will need to locate a clean well lit, well ventilated work space. Locate a small pencil tipped 27 to 33 watt soldering iron. You should also locate a spool of 22 ga. rosin core solder, a small container of “Tip Tinner” from your local Radio Shack store. “Tip Tinner” helps to clean and dress the soldering iron tip. Place the schematics and parts layout diagram in front of you along with all of the project components. A few small tools would be helpful to constructing your project, so find a small pair of end-cutters, a pair of tweezers, a magnifying glass, a small Phillips and flat-blade screwdrivers, and we can begin. When building a project with integrated circuits it is wise to consider installing integrated circuit sockets on the PC board, as a form of insurance in the event of a possible circuit failure somewhere down the road. It is much more easy to simply unplug a damaged IC and simply replace it by plugging in a new one.

Before we get some momentum on the project, you may also want to refer to the chart in Table 12-2. This table lists the color codes for the resistors, and will greatly aide you in constructing the receiver. Go ahead and locate resistor R1, a 68 k resistor, its color code is (blue-gray-orange). Now locate resistor R2, a 47 k ohm resistor with (yellow-violet-orange) color code. Place these resistors on the circuit board in their respective locations and solder them in place. Now you can identify the remaining resistors and place them on the circuit board and solder them in their proper locations. Use your end-cutters to trim the excess component leads. Cut the extra component leads flush to the edge of the circuit board.

Now locate and install capacitors; first locate and identify capacitors C1 and C2. Capacitor C1 is a small disk capacitor marked (121) or120 pF. Small capacitors often do not have their actual values printed on them but use a three-digit code to represent their value. Refer to the chart in Table 12-3 to help you identify the small capacitors. Now look for capacitor C2, its value

Table 12-2
Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

Table 12-3

Capacitance code information

This table is designed to provide the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Some capacitors actually have the numeric values stamped on them; however, some are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

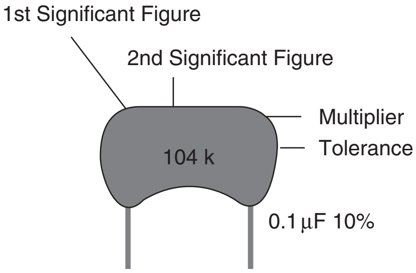

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

is .005 µF, but may be marked (502). Go ahead and install these capacitors in their respective locations on the circuit board. Remember to trim the excess component lead lengths with your end-cutters. Next, locate the remaining small capacitors and install them on the circuit board.

Next, we are going to locate and install the electrolytic capacitors. Electrolytic capacitors are generally larger in size and higher in value than the small value capacitors and usually they are vertical or horizontally mounted. Remember, electrolytic capacitors have polarity and must be installed with respect to this polarity if the circuit is going to work properly. The capacitor's value will be printed on the body of the capacitor along with a white or black polarity marking, either a plus or minus marking. Pay attention to this plus or minus marking and align it with the proper holes on the PC board, you will have to refer to the schematic and or parts layout diagram when placing the electrolytic capacitors. Locate and install the remaining electrolytic capacitors onto the circuit board. Remember to trim the excess electrolytic capacitor leads after they are soldered in place.

In this project there are two silicon diodes. Diodes are generally rectangular in shape like a resistor but they are often clear glass or painted black. You will see a black or white band at one edge of the diode body, this marking denotes the diode's polarity. The colored band is the diode's cathode lead. When installing the diode make sure that you observe the polarity marking by referring to the parts layout and the schematic diagram. Polarity is important and must be observed in order for the circuit to work properly.

The 6N2 receiver has two crystal filters, a 455 kHz one and a 10.7 MHz one. These devices are small three-lead devices and look much like a small capacitor with three leads. Make sure that you can identify them properly before installing them on the circuit board. The 455 kHz unit connects to U1 at pins 5, 6 and 7, while the 10.7 MHz crystal filter is connected to pins 17, 18 and 19. The receiver also employs a single 10.24 MHz crystal which is connected between pins 4 and 6. This is a two-lead device and is usually a small metal can.

Now locate the air-wound coils L1 and L4. Coil L1 is 4½ turns of wire, it will be placed on the right edge of the PC board, and coil L4, a 1½ turn air-wound coil, will be mounted at the bottom center of the board. Coils L2 and L3 are 9 turn coils wound on a coil form with a ferrite tunable slug in the center of the form. Coil L5 is a 0.64 mH coil in a small metal can with five leads. Coil L5 is a tunable slug tuned coil with a yellow ferrite core, see Table 12-4 for coil winding details.

The 6N2 receiver utilizes three integrated circuits. Before we go ahead and install the integrated circuits, let's take a brief look at the semiconductor pin-out diagram shown in Figure 12-6. The main receiver chip U1 is the Motorola MC3362, a 24 pin dual conversion FM receiver on-a-chip which contains oscillators, mixers, quadrature discriminator, and meter drive/carrier detect circuitry all in one chip. The audio amplifier at U2 is an LM386 chip, while U3 is 5 volt regulator, an LM78L05. When installing the integrated circuits, you will need to identify the pin-outs of each of the integrated circuits. It is advisable to install IC sockets prior to installing the ICs on the circuit board, to avoid the rare event of circuit failure at a late date. It is much easier to simply unplug an IC and insert another without un-soldering a 24 pin chip from the circuit board. Integrated circuits will generally have a small cut-out at one end of the IC package or a small indented circle at one end of the chip. Pin one (1) of the integrated circuit will be to the left of these markings. Be sure to orient the IC before placing into its proper socket. Note that pin 1 of U1 will connect to the junction of capacitors C7 and C16.

Table 12-4

Coil Winding Information

L1	4.5 turns insulated magnet wire - #22 ga. Air-core coil wound on ¼″ form then remove coil.
L2, L3	8½ turns #24 ga. insulated magnet wire coil-wound on 49A Series form - ferrite slug adjustable (49A127MPC) Circuit Specialists. Remove original coil from form and rewind 8½ turns.
L4	1.5 turns insulated magnet wire - #22 ga. Air-core wound on ¼″ form then remove coil.
L5	.64 mH Coil (yellow) 455 kHz I-F Transformer Circuit Specialists - (42IF301).

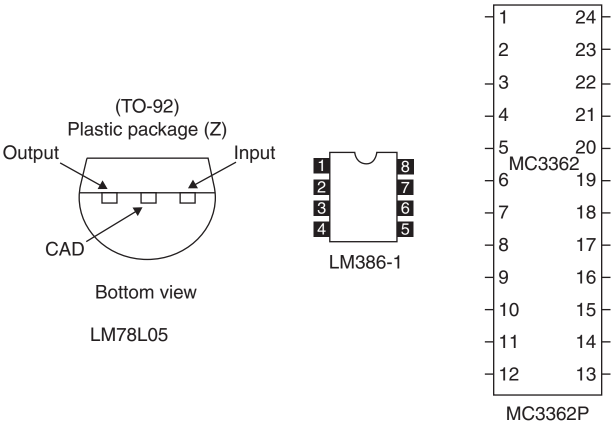

Figure 12-6 *Semiconductor pin-outs*

Finishing the PC board assembly

Now cut three 3″ pieces of wire from the roll of 22 gauge solid wire and strip ¼″ of insulation off each end. Using these wires, solder the 5k potentiometer (VR2) to the PC board. Be sure to mechanically attach each wire. Note that VR3 is the tuning control potentiometer, the center lead is connected to the PC board hole marked (D), while the left-most pot terminal gets connected to the hole marked (N). The final potentiometer wire on the right side of the potentiometer goes to the hole marked (P).

Finally, connect up a short piece of coax to the terminals on the PC board marked ANT and GND. The center wire of the coax goes to the terminal marked antenna (ANT) and shield wire from the coax goes to the terminal marked ground or (GND). The other free end of the coax goes to RCA antenna jack which gets mounted on the front panel of the chassis box.

After completing the circuit board, take a short break and when we return, we will examine the circuit board for possible "cold" solder joints and "short" circuits. Pick up the circuit board with the foil side of the circuit board facing upwards toward you. First, we will look at the circuit board for possible "cold" solder joints. Take a close look at the solder joints, they should all look clean, shiny and bright. If any of the solder joints look dark, dull or "blobby," then you should un-solder that particular joint and clean it and then re-solder the joint so that it looks clean, shiny and bright. Next we will examine the circuit board for possible "short" circuits.

The front panel assembly

The speaker was mounted on the receiver's front panel. Using the three black 4-40 screws, 4-40 nuts and #4 internal tooth washers, mount the three small right-angle brackets to the speaker mounting holes. The short side of the brackets should be against the panel. Use a blunt tool to bend the brackets over the speaker. Using the hardware shown, mount the phono jack. Using the hardware shown, mount the On-Off switch, 2M/6M switch, squelch pot, volume pot, and tuning pot to the front panel.

Orient the speaker terminals, so they don't interfere with the nearby battery mounting. Solder the loose end of the shielded cable to the phono jack, center conductor to the phono jack terminal and shield to the GND terminal. Peel the backing off one side of the ½″ wide, double-sided tape and stick it to the battery clamp. Peel the backing off the other side of the double-sided tape and stick it to the inside bottom of the plastic case. With the power switch OFF, connect a 9V alkaline battery to the battery snap and put the battery into the battery clamp. When the final test and alignment (next section of manual) is completed, place the front plate on the plastic case and insert the four nylon plugs into the holes in each corner of the plate.

Connecting the front panel parts to the PC board

To wire the speaker, pot and switch, cut the indicated length of wire from the roll of 22 gauge solid wire and strip ¼″ of insulation off each end. Next, go ahead and mount the phono jack at J1. Finally mount the On/Off switch, the 2 meter/6 meter band switch, the squelch potentiometer, the volume potentiometer, and tuning potentiometer to the front panel as shown in Figure 12–7. For the jumper wire, cut 2″ of 22 gauge solid wire, strip ¼″ of insulation off each end and solder to the points indicated. Attach two 4″ long pieces of insulated wire to the speaker terminals and solder them. Now solder two 4″ pieces of insulated wire to the On-Off switch, use the center terminal and one of the outside terminals on the switch. Next attach three 3″ long insulated wires to the volume control on VR3.

Next locate the Squelch control (VR1) and solder two 3″ long insulated wires to the potentiometer. Take the two free wires and connect them to the PC board. The center terminal from the potentiometer goes to the PC board hole marked (Q), while the other potentiometer lead goes to the hole marked (R).

Take the free ends of the two speaker wires and solder them to the circuit board at their respective terminal strip on the bottom left side of the PC board. Now take the two free ends of the On-Off switch leads and solder them to the terminals on the left side of the PC board. Next take the three potentiometer leads and solder them to the PC board. The center terminal of VR3 goes to the terminal marked (D) on the circuit board. The left-most potentiometer terminal goes to the hole marked (C) on the PC board, while the right-most potentiometer terminal goes to the hole marked (E) on the circuit board. Finally locate the 9 volt snap battery terminal stip and connect it to the battery terminals on the far left marked BATT.

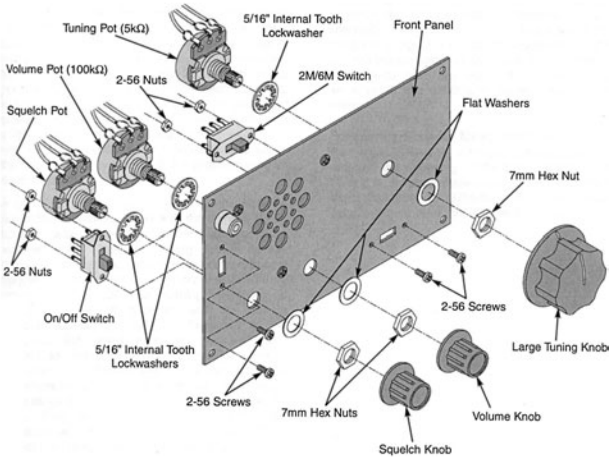

Figure 12-7 *6N2 meters*

Now that your 6N2 receiver has been completed, it is time to test out your receiver to see if it is going to work. Connect up the 9 volt battery clip to the 9 volt battery. Connect an antenna to the antenna terminal. For testing purposes you can simply connect a 19″ piece of wire to the antenna terminals, or if you have an outdoor antenna you can connect it up to the receiver. Position the band switch to the 2-meter band position. Now adjust the volume control to the mid-way position, and adjust the squelch control counterclockwise so that you just break the squelch and you should hear the receiver noise coming from the speaker. If all is well, you should hear a hissing sound or receiver noise. Adjust the tuning control around the band, back and forth, and you should eventually hear a ham radio operator calling another friend.

In the event that you hear no receiver noise and the receiver appears “dead,” you will have to disconnect the antenna and battery clip and examine the receiver’s circuit board for any errors which might have occurred during construction. Some of the most common causes for construction errors include installing diodes and electrolytic capacitors backwards in the circuit; remember that these components have polarity and it must be observed in order for the circuit to work properly. Another common problem is having resistors installed in the wrong place on the circuit board. Finally, when installing semiconductors such as transistors and FETs, you must refer to the manufacturer’s pin-out diagram and the parts layout diagram along with the schematic when constructing the circuit. Often times transistors are installed incorrectly; also check the orientation of all integrated circuits to make sure that you have oriented them correctly with respect to their marking and pin numbers. Hopefully you have found your error by now and are ready to reassemble the circuit in order to test once again.

Reconnect the antenna and battery leads and turn the On-Off switch to the “On” position. Turn up the volume control to the midway position, and set the BAND switch to the 2-meter position. The 2-meter band will generally be much more active than the 6-meter band unless there is a 6-meter band "opening." Now, you will need to adjust the squelch control. The correct adjustment of the squelch control is to advance the control until the receiver noise is silenced. When the receiver is tuned across the band and a signal is found, the squelch will release and the transmission will be heard. As you continue tuning away from this signal, the squelch should again come into effect and "squelch" the receiver noise. Remember that the function of the squelch control is to eliminate static when there is no signal present. This control sets the level of passable signals. The squelch control works by raising the voltage to the (–) input to the audio amplifier and thus disabling the amplifier output. Turn the squelch control fully counterclockwise. The audio should produce a hissing sound. Turning the squelch control clockwise should quiet the audio. Check the leads going to the control.

Now that the receiver seems to be functioning, you will need to calibrate the receiver using an RF frequency generator. You will need to borrow an RF signal generator from a friend or local radio enthusiast in order to calibrate your receiver for best operation.

Frequency measurements

Attach a two foot or longer length of wire to the male RCA plug and connect it to the radio antenna jack. Install a fresh battery and turn the power ON. Make sure that the squelch control is fully counterclockwise and the volume control is set at mid-position. A "rushing" noise should be heard from the speaker when no signal is present, which is normal. Slowly tune the radio on each band and listen for activity. Reception will only be possible if someone is transmitting. If you know someone with a 2M or 6M transmitter, you should test your receiver with known frequencies of transmission and adjust the high frequency oscillator for proper tuning of desired bands.

A calibrated RF generator may also be used to adjust the oscillators. Start with the 2-meter band. Loosely couple the output of the RF generator to the receiver input. Turn the tuning control on the receiver to the full counterclockwise position (lowest frequency). With the squelch control tuned fully counterclockwise (OFF), a "rushing" noise will be heard. Tune the RF generator until the receiver noise disappears and note the frequency on the generator dial. Next, rotate the receiver tuning control to the full clockwise position (higher frequencies). Repeat the above procedure to determine the upper frequency of your tuning range. If the range is too low, decrease L4 by spreading the turns. If the range is too high, substitute the 39 pF capacitor with the 33 pF capacitor in C6. Then, repeat both procedures to determine the upper and lower frequency limits. With a little perseverance, you should arrive at a range covering the 2-meter band (144 to 148 MHz). Set the generator at different points (144, 145, 146, etc.) and tune the radio for quieting at each setting. Mark the dial with the appropriate values. Using the plastic alignment tool to adjust L3, follow the same procedure for the 6-meter band (50 to 54 MHz). **Note:** Before the receiver is properly adjusted, you may receive a broadcast from an FM station or TV station. These signals will be distorted because they are wide band FM transmissions and the 6N2 is a narrow band amateur receiver.

If a frequency counter is available, it can be used to measure the frequency of the local oscillator at pin 20 of U1. On the 2-meter band, the counter should read 133.3 MHz to 137 MHz, and on the 6-meter band, it should read 39.3 MHz to 43.3 MHz.

Antenna considerations

For local VHF reception, the VHF receiver will operate fine just using a whip antenna. If you wish to receive stations which are further away than just the local stations, you may want to install some sort of outdoor antenna such as a ground plane or discone antenna for VHF operation. Since most of the transmissions in this band are of vertical polarization, vertical antennas, ground plane and discone antennas will work very well in this band. Various VHF scanner type antennas from your local Radio Shack store will work fine as well.

Connect up a whip antenna or an outdoor Discone antenna and you can now begin to listen-in to the exciting world of VHF ham radio communication. The charts in Table 12-5 illustrate the band plan for the two popular ham radio VHF bands. Note the SSB satellite and long-range portion of the 2-meter band from 144.1 MHz to 144.3 MHz and the 6-meter DX windows around 51.1 and 52.05 MHz.

Table 12-5

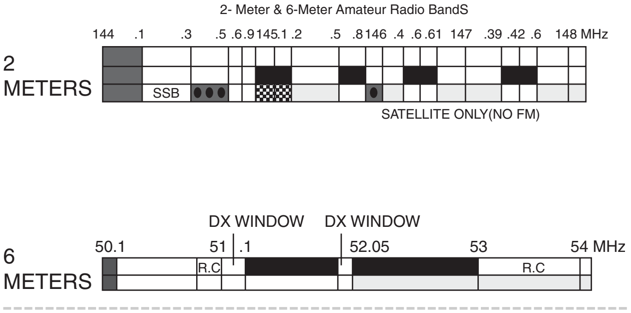
2- Meter & 6-Meter Amateur Radio BandS
144 .1 .3 .5 .6 .9 145.1 .2 .5 .8 146 .4 .6 .61 147 .39 .42 .6 148 MHz
2 METERS
SSB
SATELLITE ONLY(NO FM)
DX WINDOW
DX WINDOW
50.1
51
.1
52.05
53
54 MHz
6 METERS
R.C
R.C

Chapter 13

Active and Passive Aircraft Band Receivers

Parts list

Active Aircraft Receiver

R1,R2,R3	10k potentiometers
R5,7,11,18,25,27	1k ohm [brown-black-red]
R6,R28	270 ohm [red-violet-brown]
R8,12,17,23	10k ohm [brown-black-orange]
R26	22k ohm [red-red-orange]
R13,R22	33k ohm [orange-orange-orange]
R4,9,15,16,20,21,24	47k ohm [yellow-violet-orange]
R19	100k ohm [brown-black-yellow]
R10,R14	1 megohm [brown-black-green]
C3,C5	3.9 pF ceramic capacitor
C11	10 pF ceramic capacitor
C12,C14	27 pF ceramic capacitor
C2,C4,C6	82 pF ceramic capacitor
C1,7,8,13,16	.001 µF disk capacitor (marked.001 or 102)
C9,17,19,20,28,30	.01 µF disk capacitor (marked .01 or 103 or 10 nf)
C23,C24	.1 µF disk capacitor (marked .1 or 104)
C22	.47 µF electrolytic capacitor
C10,15,21,25,26,31	4.7 to 10 µF electrolytic capacitor
C18,27,29	100 to 220 µF electrolytic capacitor
Q1	2SC2498 or 2SC2570A NPN UHF transistor
Q2	2N3904 NPN transistor
U1	SA602 8 pin IC
U2	MC1350 8 pin IC
U3	LM324 14 pin IC
U4	LM386 8 pin IC
D1	BB505 varactor diode (marked BB505)
D2	1N270 germanium diode
D3	1N4148 silicon diode
L1,L3,L5	See Table 13-3
L2,L4	See Table 13-3
L6	See Table 13-3
L7	See Table 13-3
FL1	10.7 MHz ceramic filter - (Murata/ Digi-Key)

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

S1 PC mount DPDT switch

J1 PC mount RCA jack

J2 subminiature phone jack

B1 9 volt battery

Misc PC board, battery clip, battery holder

*** Ramsey Electronics Inc.

Passive Aircraft Monitor

R1,R2 10 megohm, $^1/_4$ watt, 5% resistor

R3 100k ohm potentiometer - volume (chassis mount)

R4,R5 3.3k ohm, $^1/_4$ watt, 5% resistor

R6 100 k ohm, $^1/_4$ watt, 5% resistor

R7 100 ohm $^1/_4$ watt, 5% resistor

C1 1.5 pF, 35 volt mica/poly capacitor

C2 5-18 pF trimmer capacitor (tuning)

C3 .01 µF, 35 volt ceramic disk capacitor

C4 330 pF, 35 volt polyester capacitor

C5,C8 100 µF, 35 volt electrolytic capacitor

C6,C7 1 µF, 35 volt electrolytic capacitor

D1 1N34 germanium diode

L1 .15 µH choke coil - (Digi-Key)

U1 LM358 dual op-amp

S1 SPST toggle switch

B1 9 volt transistor radio battery

PH high-impedance (2 k ohm) headphone

Misc battery clip, battery holder, wire, enclosure, etc.

Eavesdroping in on the pilots as they trace their way across the sky can be very fascinating and exciting listening for aircraft enthusiasts of all ages. An aircraft receiver will allow you to listen-in to high flying commercial aircraft from jumbos to pipers, as well as control towers passing instructions to the airplanes overhead. It will provide many hours of interesting listening.

In this chapter, we will present two different types of aircraft receivers: a passive aircraft receiver and an active aircraft receiver. A passive aircraft receiver is a more simple receiver with no local oscillator to interfere with the on-board aircraft electronics. A passive aircraft receiver is defined as a tuned detector with an amplifier which can be carried aboard an airplane in order to listen-in to the pilot while you are on a flight. You will be able to listen-in to the pilot in flight as well as the control tower when in fairly close range. This type of receiver does not interfere with the airplane's electronics and in most instances you will be permitted to use this type of receiver on-board your aircraft. The active receiver is a more sensitive active receiver which can be used at home to listen-in to high-flying aircraft as well as control towers over longer distances. Listening to aircraft communications can be very interesting but very cryptic, with lots of bursts of information passing in a short period of time.

Active aircraft band receiver

Ideal for arm-chair pilots, student pilots and the seasoned pilots, this sensitive receiver will pick up planes up to 100 miles away. The active aircraft band receiver tunes the aircraft band between 118 and 136 MHz and will allow you to listen to control towers, centers, and planes en-route to their destination. The active aircraft band receiver, shown in Figure 13-1 has good sensitivity, image rejection, signal-to-noise ratio and stability. It is designed for casual "listening in" both

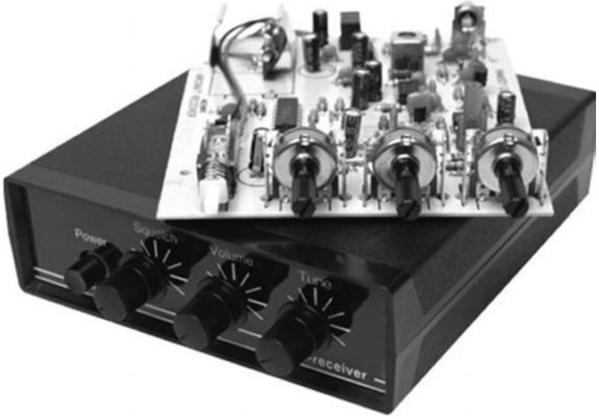

Figure 13-1 *Active aircraft band receiver*

ground and air communication, for both commercial airlines and general aviation, and it will provide you with many years of easy sky-monitoring enjoyment.

This project is especially good for people with an interest in learning more about aviation and electronics and radio. The active aircraft receiver can be built by folks of all ages and skill levels.

What you can hear

A basic fact about the VHF Aviation Band which even licensed pilots can overlook or forget is that communications are in the AM mode, not FM, as in the case of the FM broadcast band immediately below it, and the VHF public service and ham bands immediately above it. No matter where you live, you will be able to receive at least the airborne side of many air traffic communications. If you know where to tune, you'll hear any aircraft you can see, PLUS planes a hundred miles away and more, since VHF signals travel "line of sight". An airliner at 35,000 feet altitude in the next state is still line of sight to your antenna. Similarly, whatever ground stations you may hear are also determined by this "line of sight" character of VHF communication. If there are no major obstacles between your antenna and an airport (tall buildings, hills, etc.), you'll be able to hear both sides of many kinds of aviation communication. Be prepared for them to be fast and to the point, and for the same airplane to move to several different frequencies in the span of a few minutes! The most common types of ground services with which pilots communicate is the control tower, ground control, clearance delivery and ATIS or "Automatic Terminal Information System."

At most metropolitan airports, a pilot communicates with the FAA on a frequency called "Clearance Delivery" to obtain approval or clearance of the intended flight plan. This communication is done before contacting ground control for taxi instructions. From the control tower, ground movements on ramps and taxiways are handled on the "Ground Control" frequency, while runway and in-flight maneuvers near the airport (takeoffs, local traffic patterns, final approaches and landings) are on the "Control Tower" frequency. ATIS, or "Automatic Terminal Information System" is a repeated broadcast about basic weather information, runways in use, and any special information such as closed taxiways or runways. Such a broadcast offers an excellent steady signal source for initial adjustment of your receiver. If you are close enough to the airport to receive ATIS, you will hear the Approach Control and the Departure Control transmissions. These air traffic radar controllers coordinate all flight operations in the vicinity of busy metropolitan airport areas, and are called the ATC Center.

When you hear a pilot talking with "Jacksonville Center" or "Indianapolis Center," you know the aircraft is really enroute on a flight rather than just leaving or just approaching a destination. A pilot will be in touch with several different "Regional Centers" during a cross-country flight. These smaller centers are called "Unicom" centers.

Airports without control towers rely on the local "Unicom" frequency dedicated only to advisory communications between pilots and ground personnel such as fuel service operators. The people on the ground can advise the pilot on the status of incoming or outgoing aircraft, but the pilot remains responsible for landing and takeoff decisions. Typical Unicom frequencies are 122.8 and 123.0 MHz.

The FAA's network of Flight Service Stations keeps track of flight plans, provides weather briefings and other services to pilots. Some advisory radio communication takes place between pilots and a regional "FSS." If there is an FSS in your local area, but no airport control towers, the FSS radio frequency will stay interesting.

Fast talking pilot and controllers

Aviation communication is brief, but it is clear and full of meaning. Usually, pilots repeat back exactly what they hear from a controller so that both know that the message or instructions were correctly interpreted. If you are listening in, it is hard to track everything said from a cockpit, particularly in big city areas. Just to taxi, take off and fly a few miles, a pilot may talk with six or eight different air traffic control operations, all on different frequencies, all within a few minutes! Here are the meanings of a few typical communications:

"Miami center, delta 545 heavy out of three-zero for two-five."

Delta Flight 545 acknowledges Miami Center's clearance to descend from 30,000 feet to 25,000 feet altitude. The word "heavy" means that the plane is a jumbo jet such as 747, DC-10, etc.

"Seneca 432 lima cleared to outer marker. contact Tower 118.7."

The local Approach Control is saying that the Piper Seneca with the N-number (tail number) ending in "432L" is cleared to continue flying an instrument approach to the outer marker (a precision radio beacon located near the airport) and should immediately call the airport radio control tower at 118.7 MHz. This message also implies that the controller does not expect to talk again with that aircraft.

"Cessna 723, squawk 6750, climb and maintain five thousand."

A controller is telling the Cessna pilot to set the airplane's radar transponder to code 6750, climb to and fly level at an altitude of 5000 feet.

"United 330, traffic at 9 o'clock, 4 miles, altitude unknown."

The controller alerts United Airlines flight #330 of radar contact with some other aircraft off to the pilot's left at a 9 o'clock position. Since the unknown plane's altitude is also unknown, both controller and pilot realize that it is a smaller private plane not equipped with altitude-reporting equipment.

Active receiver circuit description

Now, take a look at the schematic diagram shown in Figure 13-2. The aircraft band antenna is first coupled through C1 to a three-section tuned LC filter input network. The aircraft band signals ranging from 118 to 135 MHz signals are amplified by VHF transistor Q1 and fed to the input of U1, the SA602 mixer-oscillator. Inductor L6 and its associated capacitor network establish the LO (local oscillator) frequency at 10.7 MHz higher than the incoming 118–135 MHz signals. The local oscillator frequency may be tuned across about 15 MHz by the varactor tuning network formed by diode D1 and resistor R1. The 10.7 MHz difference between the local oscillator and the received signal is fed through the 10.7 MHz ceramic filter FL1, and then amplified by Q2 and the signal is then applied to U2, the MC1350 IF amplifier IC with AGC input. The 10.7 MHz IF is peaked by inductor L7, and the AM audio signal is then demodulated by diode D2 and fed through the four op-amps of U3, the LM324, where volume control, AGC output, audio filtering and squelch functions are managed. The LM386 (U4) is the audio amplifier, and is capable of driving simple communications speakers to excellent volume levels.

Circuit construction

The aircraft band receiver is best constructed on a printed circuit board for best results. Place your schematic diagram in front of you, heat up your 27 to 30 watt pencil tipped soldering iron and prepare to insert

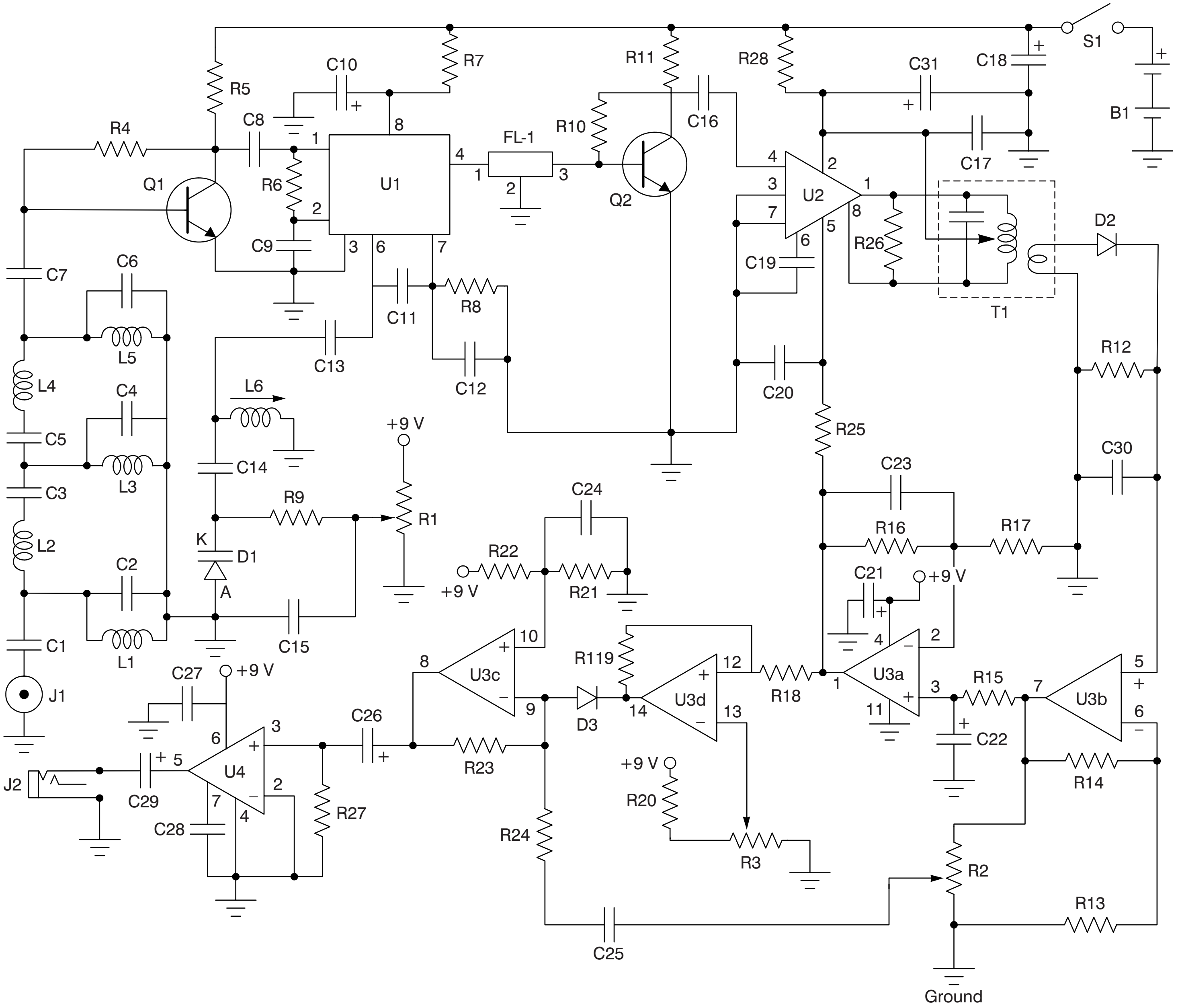
R5
C10
R7
R11
R28
C31
C18
S1
B1
R4
C8
R6
Q1
U1
C9
FL-1
R10
Q2
C16
U2
R26
C17
C19
T1
D2
C7
C6
L5
C11
R8
C13
C12
C20
R12
L4
C4
L6
C5
L3
C14
+9 V
R25
C30
C23
C3
R9
R1
C24
R16
R17
L2
K
D1
A
C2
R22
+9 V
R21
C21
+9 V
L1
C1
C15
U3c
R119
U3d
R18
U3a
R15
U3b
J1
C27
+9 V
D3
C22
C26
R14
J2
U4
C29
C28
R27
R23
+9 V
R20
R24
R3
R2
R13
C25
Ground

Figure 13-2 *Active aircraft receiver schematic. Courtesy of Ramsey Electronics*

Table 13-1
Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

the components into the circuit board. You can fabricate your own circuit board from scratch if you are an experienced builder or you could elect to build the receiver from the kit supplier, Ramsey Electronics, listed in the Appendix. We will build the receiver in sections and will begin first with the antenna and input pre-amplifier section.

Before we go ahead and build the aircraft receiver, let's take a few moments to look at Table 13-1, which illustrates the resistor color chart. Resistors have three or four color bands, which begin at one edge of the resistor body. The first color band represents the resistor's first digit value, while the second color band denotes the second digit of the resistor value. The third color band represents the multiplier value of the resistor. The fourth color band denotes the tolerance value of the resistor. A silver band represents a 10% tolerance, while a gold band notes a 5% tolerance value. The absence of a fourth color band states that the resistor has a 20% tolerance value. Therefore, a 10,000 ohm or 10 k ohm resistor would have a brown band (1) a black band (0) and a multiplier of (000), and an orange band.

Go ahead and install resistors R4, R5 and R6, then solder them in place on the PC board. Then cut the excess leads from the PC board with a pair of end-cutters. Cut the excess leads flush to the PC board.

Next we are going to install the capacitors for the project. Capacitors are described as electrolytic and non-electrolytic types. Electrolytic types are large sized and larger in value than non-electrolytic types. Electrolytic type capacitors have polarity, i.e. both a positive and negative terminal, often with color band and/or (+) or minus (−) marking on them. These electrolytic capacitors must be installed with respect to these polarity markings if you expect the circuit to work correctly. Take your time installing them and refer regularly to both the schematic and parts layout diagrams. Non-electrolytic capacitors can sometimes be very small in size, and often their actual value will NOT be printed on them. A three-digit code is used to help identify these capacitors; refer to Table 13-2. For example, a .001 µF capacitor would be marked with (102), while a .01 µF capacitor would be marked with (103).

Go ahead and install capacitors C1 through C8 into their respective PC holes on the PC board, being sure to observe polarity on the capacitors before installing them on the PC board. Most electrolytics will have a black band or a plus or minus marking next to a pin. Solder the capacitors to the PC board, then remove the excess leads. Next install inductors L1, L2, L3 and L4 into

Table 13-2

Capacitance code information

This table provides the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Some capacitors actually have the numeric values stamped on them; however, many are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

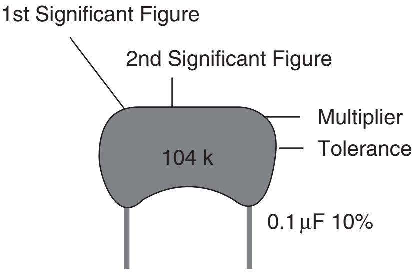

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

Table 13-3
Coil winding information

L1,L3,L5	1.5 turns #28 ga. magnet wire, 5 mm inner diameter air core - about 1 mm spacing (2-turns works fine)
L2,L4	.33 µH, molded inductor (Digi-Key M9R33-ND)
L6	.1 µH, 3.5 turns, slug tuned coil (Digi-Key TK2816-ND)
L7	10.7 MHz shielded I-F transformer (Mouser 42FIF122) or Circuit Specialists

their respective holes, followed by L5; see Table 13-3 for winding instructions. When you have inserted the inductors, you can solder them in place and then cut the excess component leads flush to the edge of the circuit board.

Before you go ahead and install the transistors and integrated circuits, refer to the semiconductor pin-out diagram shown in Figure 13-3, which will help you orient the components. Now locate and install transistor Q1, being sure you have the correct component before installing it on the PC board. Be sure you can identify all three leads as to their correct pin-outs before inserting it into the circuit board to avoid damage to the circuit once power is later applied. Solder Q1 to the board and then remove the excess leads.

Next we are going to move on to installing the Local Oscillator components. First install resistors R1, R7, R8, and R9 followed by resistors R10, R11, R12 and then resistors R25, R26 and R28. Solder the resistors into their respective locations and be sure to remove the excess lead lengths. Next we will install capacitors C9, C10, C11 and C12 followed by capacitors C13, C16, C17, C19 and C20 followed by C30 and C31. Be sure to observe the polarity of the capacitors when installing them on the board to prevent circuit damage when power is first applied. Install the capacitors into their respective location on the PC board. Once inserted, solder the capacitors to the circuit board and then remove the excess leads.

Now locate diode D1 and D2, and note that the cathode or black band side of the diode points to the battery. Diode D1 is the varactor "tuning" diode. Go ahead and install these diodes, then solder them to the board. Remove the excess leads. Locate transistor Q2 and install it on the PC, once you have identified all the leads. Solder Q2 to the board, and then remove the excess leads with your end-cutter. Now locate and install inductor L6, the 3.5-turn slug tuned coil and L7, the 10.7 I-F transformer. Remember to trim the extra lead lengths as needed. Locate the FL1, the 10.7 MHz ceramic filter, it will have three leads and is usually a small square component. Solder it in place and remove the excess leads. Finally install the 8-pin SA602 local oscillator-mixer at U1. It is advisable to install an IC socket into place in the proper location and then insert the IC into the socket. This is good insurance in the event of a possible component failure at some later point. When inserting the IC into the socket be sure to first identify which pin is #1. Usually each IC will have a small circle or cut-out on the top or top left side of the IC. Usually the pin is just to the right of this indented circle.

Next we will install the LM324 IC circuit components. First locate resistors R13, R14, R15, R16, R17, R18 and R19 followed by resistors R20, R21, R22, R23 and R24. Install these resistors in their respective locations and then solder them in place on the

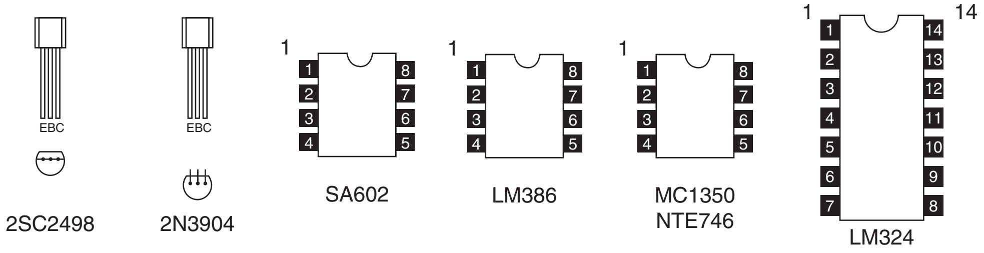

Figure 13-3 *Semiconductor pin-outs*

PC board. Remember to cut the extra leads. Now locate diode D3, remember to observe the polarity when installing D3. Solder the diode in place, then remove the excess lead lengths. Now locate and install capacitors C18, C21, C22 followed by capacitors C23, C24 and C25. Remember electrolytic capacitors have polarity markings and they must be observed for the circuit to operate correctly. You will see either a plus or minus marking at one end of the capacitor. Once installed, solder the capacitors in place and then remove the extra leads by cutting them flush to the circuit board. Finally install an integrated circuit socket for U3 and then place U3 into the socket, but be sure to install it correctly, observing the indented circle as near pin 1 on the IC.

Finally we are going to install the audio amplifier components. First locate resistor R27 and install it on the PC board. Next solder it in place and cut the extra leads. Now, locate and install capacitors C26, C27, C28, and C29. Remember C26 and C27 are electrolytic types, so polarity must be observed before installing them. Finally install an IC socket for U4 the LM386 amplifier IC. When installing U4 be sure to observe the correct orientation before inserting it into the socket to prevent damage when power is applied.

Now you finish up the circuit by installing the 9 volt battery holder, the 9 volt battery clip, followed by switch S1. Once these components have been installed you can locate and install jacks J1 and J2 onto the circuit board.

If you haven't already installed the potentiometers, R1, R2 and R3, go ahead and install them and solder the controls firmly in place. Be sure the mounting tabs are completely soldered for a good mechanical connection.

Let's take a short break and when we return we will inspect the PC board for poor solder joints and "short" circuits before applying power to the circuit board. Pick up the circuit board and bring the foil side of the board up towards you to inspect the board. First, we will look for "cold" or poor solder joints on the PC board. Take a look at each of the solder joints. You want to make sure each of the solder joints appears to look clean, smooth and shiny. If you see any solder joints that look dull or "blobby" then you will need to remove the solder from the joint with a solder-sucker or wick and then re-solder the joint, so that it appears to be clean, shiny and smooth.

Next we will inspect the PC board for "short" circuits. Often "stray" or "cut" component leads will "bridge" across or between circuit traces on the PC board. These possible "shorts" can cause damage to the circuit when power is first applied, so it is best to remove any "bridging" leads or solder blobs or "shorts" before power is sent to the circuit board. Look over the board carefully for any signs of "bridging" wires or blobs of solder. Once you are comfortable with your inspection, you can lay the circuit board back down on the table.

Enclosure recommendations

If your first goal is economy and rugged portability, you will find that the circuit board can be nicely mounted in a standard VHS videotape storage box, which also gives room for a speaker or earphone storage, and even a roll of antenna wire. The controls are easily mounted at one end of such a box. It may be necessary to cut away the molded posts which secure the tape cassette itself. These storage boxes come in several styles, so pick one which truly looks practical as a project enclosure.

The most economical metal enclosure nicely suited for Ramsey PC board kits is the Radio Shack 270-253A. This metal utility cabinet can accommodate one or two different receiver boards plus a speaker, with room for various refinements you might want to add, such as a bigger battery pack, fine tuning control, and so forth. When you have selected your enclosure, you can place the circuit board into the case and mount it. Be sure to allow the extension of the squelch control, the volume control and the tuning control on the front panel of the chassis. The speaker can be mounted on the top of the case, and the earphone and antenna jack can be mounted on the rear of the receiver.

Congratulations, you have successfully completed your active aircraft band receiver. Now you are ready to test and adjust the receiver before setting down and enjoying it. Now, you can attach a speaker or pair of headphones to jack J2. Next, attach a whip antenna to jack J1, and then connect a 9 volt battery to the battery clip leads and we can go ahead and test the active aircraft receiver. Now, turn the On-Off switch to the "ON" position. Be sure to turn down the squelch control to below the threshold point, so you can hear a hissing sound coming from the speaker or headphone. Now, turn up the volume control to about midway. Take a few minutes to tune up and down the band with the main tuning control. If after about 15 minutes, you do not hear any transmissions on your radio or if you do not hear any hissing sounds from the speaker, then you know there could be a problem with your receiver.

In the event of a problem with your new aircraft band receiver, you will have to disconnect the speaker, antenna and battery and visually inspect the circuit board for any errors. Often a second or new pair of eyes is very helpful in locating problems on circuit boards. The most likely causes of circuit failure are misplaced components or components such as diodes and electrolytic capacitors installed backwards. The second most likely cause for problems are semiconductors, such as transistors installed backwards. Be sure to re-check the manufacturer data sheets for proper pin-out information when installing the transistors. Finally, check the orientation of the integrated circuits in their respective sockets; sometimes the ICs can be put in backwards in their sockets causing the circuit to malfunction. Once you have fully inspected and hopefully located your problem, you can reconnect the battery, headphones and antenna and then retest the receiver once again.

Antenna considerations

An antenna for your new aircraft receiver can be as simple as a 21″ piece of wire, an extendable whip antenna or a roof-mounted ground-plane aviation antenna. Most folks near an airport will get plenty of in-the-air action from a wire or whip antenna, but if you're more than a few miles away, a decent roof-mount antenna is the way to go. Radio Shack sells an ideal antenna designed for scanners which covers the aircraft band nicely and it costs around $30. A low cost TV antenna works well, even better if rotated 90 degrees (remember aircraft antennas are vertically polarized).

Adjustment and alignment

Alignment of the air-band receiver consists of simply adjusting L6 for the desired tuning range and peaking the IF transformer (L7). If you are using a signal generator, frequency counter or other VHF receiver for calibration, remember that you want to set the local oscillator frequency 10.7 MHz higher than the desired signal or range to be received. Adjustment of the L6 oscillator coil MUST be made with a non-metallic alignment tool. The use of a metal tool of any kind will detune the coil drastically, making alignment almost impossible. Also, if you're receiving FM broadcast stations, you have the slug tuned too far down in the coil form. Turn it until it is higher in the form and try again. One other thing that you can do to improve the operation of your kit is to spread out the three 2-turn coils at the antenna input, L1, L3, and L5. Stretching these out will give you greater sensitivity.

Once you know you are receiving aircraft or airport transmissions, adjust the IF transformer (L7) for best reception. Typically, L7 is adjusted 2–3 turns from the top of the shield can. If you don't have any signal reference equipment at home, and are not yet hearing airplanes, your best bet is to pack up your AR1C and needed tools and head for the nearest airport! If there is no control tower, don't hesitate to visit a general aviation service center on the airport grounds. If you've never done this before, you will probably find it to be a fun and interesting experience. Ask which are the most active frequencies and adjust L6 and your front panel tuning control until you hear the action. A ground service operator or private pilot may be willing to give you a brief test transmission on the 122.8 Unicom frequency. Remember, also, that if your airport has ATIS transmissions, you can get a steady test signal as soon as you are line-of-sight with its antenna.

The aircraft receiver does not produce a loud hiss when no signal is being received (unlike an FM receiver or expensive AM receiver); this is due to the somewhat limited amount of IF or Intermediate Frequency gain. Increasing the IF gain would produce a hiss and marginally better sensitivity (about a microvolt) but also require much more alignment, AGC circuitry, and builder ability—far beyond the intention of this kit.

Tuning the aircraft receiver

With the varactor tuning control capable of going across 10–15 MHz, and with pilots and controllers talking so briefly, you will need to get used to tuning your receiver! You'll find that ANY knob gives smoother tuning than the bare control shaft, and that a "vernier" dial will make the procedure even easier—but at the expense of being able to check up and down the band quickly, which you might like to do if you're tracking the same airplane. The air-band receiver is designed to let you explore the entire communications section of the aviation band. If you become really interested only in being able to check a certain frequency such as a nearby FAA control tower, or Unicom, it will be helpful to mark that spot on a dial template such as is reproduced for your convenience below.

"VOR" or "OMNI" transmissions

While driving around the countryside you may have come across a tall white "cone" structure near your airport or in the middle of a big farming field; it is useful for you to know that these are VHF navigational aids operating in the 118–135 MHz frequency range, just below the air-ground communications range, sending a steady signal which may be helpful in initial alignment. If you tune in such a signal on your receiver, remember that you will have to increase the local oscillator frequency later in order to listen to air traffic communications.

Passive aircraft receiver

The passive air-band receiver is a type of "crystal" radio which contains no local oscillator which might interferes with the on-board aircraft's sensitive electronics. The passive aircraft receiver broadly tunes from 118 MHz to 136 MHz, and was designed to listen-in to in-flight communications between your pilot and the control tower. The passive aircraft receiver is shown in Figure 13-4. The passive aircraft receiver can be built small enough to place inside your vest pocket and it operates from an ordinary 9 volt battery.

The passive aircraft receiver is basically an amplified type of "crystal radio" designed to receive AM aircraft transmissions. The "passive" design uses no oscillators or other RF circuitry capable of interfering with aircraft communications. This receiver utilizes a coil/capacitor tuned "front-end," which feeds an RF signal to a detector diode. The tuning capacitor may be any small variable with a range from about 5 pF to about 15 or 20 pF. The 0.15 µH inductor may be a molded choke or a few turns wound with a small diameter. Experiment with the coil to get the desired tuning range. The aircraft frequencies are directly above the FM band so a proper inductor will tune FM stations with the capacitor set near maximum capacity. (The FM stations will sound distorted since they are being slope detected.) A 1N34 germanium detector diode or a Schottky diode like the 1N5711 or HP2835 should be used as the detector diode in this receiver circuit. The 10 megohm resistors provide a small diode bias current for better detector efficiency.

The LM358 dual op-amp amplifier draws under 1 ma so the battery life should be quite long. Potentiometer R3 is used to adjust the gain to the second stage of the dual op-amp. The second op-amp stage drives a 100 ohm resistor in series with a 100 µF capacitor. You must use a high impedance "crystal" headphone, since the op-amp output will not drive a speaker directly. A speaker amplifier may be added to drive a speaker or low-z earphone, if desired, but the power consumption will increase sharply, as will the size of the receiver.

The passive aircraft receiver is powered from an ordinary 9 volt transistor radio battery. The entire aircraft band receiver can be built inside a small plastic box. Simply mount an SPST slide switch on the side of the case to apply power to the circuit. If you elect, you can install a ⅛″ phone jack for the "crystal" headphone.

Construction

The passive air-band receiver can be constructed on a regular circuit board or it could be built on a

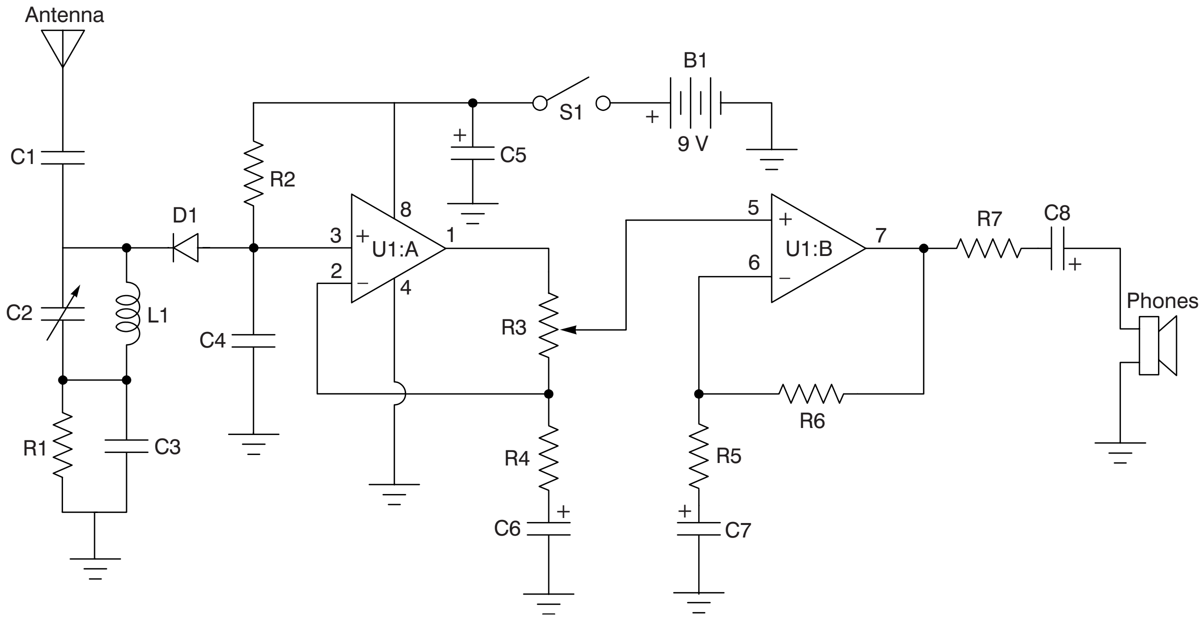

Figure 13-4 *Passive aircraft receiver schematic*

"quick-board" which contains ready-made circuit traces, these are available at Radio Shack stores.

You may want to lay out the circuit so that the tuning control is at one edge of the board, which will make things a little easier when installing the circuit in an enclosure. Building the circuit is pretty straightforward and should only take a few hours to build. Refer to Table 13-1 for identifying the resistors and Table 13-2 for identifying the small capacitors in the circuit. Remember it is very important to install the diode correctly with respect to its polarity in order for the circuit to work properly. The passive air-band receiver also contains a few electrolytic capacitors, and you will remember that these devices have polarity and it must be observed for the circuit to work. The aircraft receiver uses a dual op-amp to amplify the signals and you will have to ensure that it is installed correctly. An integrated circuit is highly recommended for the op-amp. In the event of a circuit failure at some later date, it is much easier to simply unplug an IC rather than trying to unsolder one from a PC board to replace it with a new one. When installing the IC into its socket, you will have to make sure that you orient it correctly. An IC will typically have a notch or small cut-out at one end of the IC case. Pin one (1) will be to the left of the cut-out of notch when viewed from the top of the IC. In this project pin 1 of the IC connects to potentiometer R3, while pin 8 of the IC is connected to the power source. Note that pin 4 of the IC is connected to the circuit ground.

You can elect to use a small speaker in this circuit, but in order to make a miniature receiver you may want to simply use a small ⅛″ mini jack and just plug into a pair of small earphones. The circuit is powered from a standard 9 volt battery; remember the black or minus (−) battery clip is connected to ground, while the plus (+) or red battery clip lead is connected to capacitor C5 at the plus (+) end.

You will want to have the tuning capacitor C2 near the edge of the circuit board, so that you can allow the tuning knob to protrude through the case so you can easily tune the receiver. Once the receiver has been completed, you should take a few minutes rest.

After the break, you should inspect the foil side of the circuit board for any "cold" solder joints. Cold solder joints often look dark, dull or "blobby." If you see any solder joints that look this way, you should re-solder the connections. You should also take a moment to examine the circuit board for any "short-circuits" which may be the result of a "stray" component lead shorting between circuit traces.

You can make yourself an antenna from a piece of stiff 20 ga. wire as a whip antenna to capacitor C1, or install a miniature jack in order to plug in a whip antenna. The antenna can be a 6″ piece of #20 ga. stiff solid copper wire or a small telescoping aluminum antenna connected to capacitor C1. Radio Shack has a good selection of whip antennas. You could elect to solder the wire antenna directly to C1 and just bring the antenna out the top of the enclosure, or you could install a small ½″ mini jack for the antenna on the rear of the enclosure. The selectivity is reduce as the antenna length is increased so best performance is achieved with the shortest acceptable antenna. Try increasing the 1.8 pF capacitor value when using very short antennas and decreasing it for long antennas. In order to "tune" the passive aircraft receiver you can either leave a small hole on the side of the plastic enclosure to adjust the frequency using a plastic tuning tool, or you need to find a way to attach a plastic knob onto the tuning screw; you will have to drill a larger hole in the case to bring the knob outside the case for easy tuning.

To test the passive aircraft receiver, park near your local airport. Turn on the passive air-band receiver, attach your wire antenna or pull up your telescoping antenna and tune the receiver. You should try to listen when you see an aircraft take-off or land or when you see an aircraft taxiing along the runway. If you can get close to the tower in a small airport, that should ensure that you will be able to test your receiver adequately. As you see a plane arrive or leave, "tune" around the band and you should hear either the pilot or the control tower. If you plan on taking the passive aircraft receiver on-board an aircraft to listen-in to the pilot during your flight, you should ask the pilot or stewardess if it is permitted aboard your aircraft. Generally you are given permission without a problem. Enjoy building and using your new aircraft receivers!

Chapter 14

VLF or Very Low Frequency Radio Receiver

Parts list

VLF Whistler Receiver

R1,R4 10k ohm ¼w, 5% resistor

R2,R3 22k ohm ¼w, 5% resistor

R5 6.2 megohm, ¼w 5% resistor

R6 1 meg ohm potentiometer (trimpot) - single turn

R7 10 megohm, ¼w 5% resistor

R8 1k ohm, ¼w 5% resistor

R9 1k ohm, ¼w 5% resistor

R10 3.3k ohm, ¼w 5% resistor

R11,R12 820 ohm, ¼w 5% resistor

R13 10k ohm potentiometer (panel)

R14,R17 10 ohm, ¼w, 5% resistor

R15,R16 2.2k ohm, ¼w, 5% resistor

C1,C3 47 pF, 35 volt mica capacitor

C2 100 pF, 35 volt mica capacitor

C4,C5,C6 3.3 nF, 35 volt mylar capacitor

C7,C15 .01 μF, 35 volt ceramic disk capacitor

C8 27 pF, 35 volt mica capacitor

C9 1 μF, 35 volt electrolytic capacitor

C10 .18 μF, 35 volt tantalum capacitor

C11 .12 μF, 35 volt tantalum capacitor

C12 1.8 μF, 35 volt tantalum capacitor

C13 .68 μF, 35 volt tantalum capacitor

C14 .22 μF, 35 volt tantalum capacitor

C16 .068 μF, 35 volt tantalum capacitor

C17,C18 .22 μF, 35 volt tantalum capacitor

C19 .05 μF, 35 volt ceramic disk capacitor

C20,C22,C24 100 μF, 35 volt electrolytic capacitor

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

C21,C23 .1 µF,35 volt ceramic disk capacitor
L1 120 mH coil (Mouser electronics)
L2 150 mH coil (Mouser electronics)
L3 18 mH coil (Mouser electronics)
L4 56 mH coil (Mouser electronics)
Q1a,Q1b U401 Siliconix Dual matched N-Channel FETs
Q2 MPSA56 transistor
U1 LM386 op-amp
S1 DPDT toggle switch (trap)
S2 DPDT toggle switch (high-pass)
S3 DPDT toggle switch (low-pass)
B1 9 volt transistor radio battery
J1,J2 ⅛ mini phone jack w/switch
J3 SO-238 UHF chassis jack
P1 PL259 UHF plug
P2 binding post (ground)
Misc PC board, IC socket, solder, wire, antenna, battery holder
Battery clip, hardware, chassis

"Natural Radio" is a name sometimes used for describing radio noise with natural origins, mainly due to but not confined to lightning. At first suggestion, you may think that listening to lightning crackles would be dull or uninteresting, but electromagnetic radiation from lightning can travel great distances and undergo strange modifications along the way. The frequencies of the original pulse can be spread out in time (a process called "dispersion") since the higher frequencies travel a little faster than the lower. The result is that the short impulse from a lightning strike in South America can sound like a chirp in Texas. Slower sweeping tones are called "whistlers" and they are a bit of a mystery. The energy from a lightning bolt streams out into space into a region called the "magnetosphere," magnetized plasma created by the interaction of solar wind with the Earth's magnetic field. The lightning pulse is reflected or "ducted" back down to Earth after a very long trip during which time the frequencies are spread out by a dispersion-like process. Short whistlers might be due to dispersion, but some whistlers last five seconds so ordinary dispersion is probably inadequate an explanation. A radio wave can travel a million miles in five seconds, so to accumulate that much difference in arrival times, the signal would have to travel hundreds of millions of miles, assuming a pretty steep dispersion curve. More likely, the whistler is an emission from the magnetosphere triggered by the lightning pulse. When conditions are just right, numerous lightning strikes combine with numerous reflections to give an eerie chorus that sounds a bit like a flock of geese.

Most people have never heard of Natural Radio produced by several of natures's processes, including lightning storms and aurora, aided by events occurring on the Sun. The majority of Earth's natural radio emissions occur in the extremely-low-frequency and very-low-frequency (ELF/VLF) radio spectrum. Whistlers, one of the more frequent natural radio emissions to be heard, are just one of many natural radio "sounds" the Earth produces at all times in one form or another, and these signals have caught the interest and fascination of a small but growing number of hobby listeners and professional researchers for the past four decades.

Whistlers are amazing sounding bursts of ELF/VLF radio energy initiated by lightning strikes which "fall" in pitch. A whistler, as heard in the audio output from a VLF or "whistler" receiver, generally falls lower in pitch, from as high as the middle-to-upper frequency range of our hearing downward to a low pitch of a couple hundred cycles-per-second (Hz). Measured in frequency terms, a whistler can begin at over 10,000 Hz and fall to less than 200 Hz, though the majority are heard from 6000 down to 500 Hz. Whistlers can tell scientists a great deal of the space environment between the Sun and the Earth and also about Earth's magnetosphere.

The causes of whistlers are generally well known today though not yet completely understood. What is clear is that whistlers owe their existence to lightning storms. Lightning stroke energy happens at all electromagnetic frequencies simultaneously. The Earth is literally bathed in lightning-stroke radio energy from an estimated 1500 to 2000 lightning storms in progress at any given time, triggering over a million lightning strikes daily. The total energy output of lightning storms far exceeds the combined power output of all man-made radio signals and electric power generated from power plants.

Whistlers also owe their existence to Earth's magnetic field (magnetosphere), which surrounds the planet like an enormous glove, and also to the Sun. Streaming from the Sun is the Solar Wind, which consists of energy and charged particles, called ions. And so, the combination of the Sun's Solar Wind, the Earth's magnetosphere surrounding the entire planet, and lightning storms all interact to create the intriguing sounds and great varieties of whistlers.

How do whistlers occur from this combination of natural solar-terrestrial forces? Some of the radio energy bursts from lightning strokes travel into space beyond Earth's ionosphere layers and into the magnetosphere, where they follow approximately the lines-of-force of the Earth's magnetic field to the opposite polar hemisphere along "ducts" formed by ions streaming toward Earth from the Sun's Solar Wind. Solar-Wind ions get trapped in and aligned with Earth's magnetic field. As the lightning energy travels along a field-aligned duct, its radio frequencies become spread out (dispersed) in a similar fashion to light shining into a glass prism. The higher radio frequencies arrive before the lower frequencies, resulting in a downward falling tone of varying purity.

A whistler will often be heard many thousands of miles from its initiating lightning stroke—and in the opposite polar hemisphere! Lightning storms in British Columbia and Alaska may produce whistlers that are heard in New Zealand. Likewise, lightning storms in eastern North America may produce whistlers that are heard in southern Argentina or even Antarctica.
Even more remarkably, whistler energy can also be "bounced back" through the magnetosphere near (or not-so-near) the lightning storm from which it was born!

Whistlers are descending tones. Their duration can range from a fraction of a second to several seconds, with the rate of frequency shift steadily decreasing as the frequency decreases. A whistler's note may be pure, sounding almost as if it was produced using a laboratory audio signal generator; other whistlers are more diffuse, sounding like a breathy "swoosh" or composed of multiple tones. On occasion, some whistlers produce echoes or long progressions of echoes, known as "echo trains," that can continue for many minutes.

Whistlers can occur in any season. Some types are more likely to be heard in summer than winter and vice-versa. Statistically, the odds are especially high for good whistler activity between mid-March and mid-April.

Whistlers and the sounds of the dawn chorus are not heard equally well everywhere in the world. Reception of these are poor in equatorial regions, with best reception at geomagnetic latitudes above 50°. Fortunately the continental United States and Canada are well positioned for reception of whistlers and other signals of natural radio.

Considered by many listeners to be the "Music of Earth", whistlers are among the accidental discoveries of science. In the late 19th century, European long-distance telegraph and telephone operators were the first people to hear whistlers. The long telegraph wires often picked up the snapping and crackling of lightning storms, which was mixed with the Morse code "buzzes" or voice audio from the sending station. Sometimes, the telephone operators also heard strange whistling tones in the background. They were attributed to problems in the wires and connections of the telegraph system and disregarded.

The first written report of this phenomenon dates back to 1886 in Austria, when whistlers were heard on a 22-km (14 mile) telephone wire without amplification. The German scientist, H. Barkhausen, was eavesdropping on Allied telephone conversations during World War I. In order to pick up the telephone conversations, he inserted two metal probes in the ground some distance apart and connected these to the input of a sensitive audio amplifier. He was surprised to hear whistling tones which lasted for one or two seconds and which glided from a high frequency in the audio range to a lower frequency where they disappeared into the amplifier or background noise level. On occasion, the whistlers were so numerous and loud that he could not detect any type of telephone.

Later, three researchers from the Marconi Company in England reported in 1928 on the work they had done in whistler research. Eckersley, Smith, and Tremellen established a positive correlation between whistler occurrence and solar activity, and they found that whistlers frequently occur in groups preceded by a loud click. The time between the click and the whistlers was about three seconds. They concluded that there must have been two paths of propagation involved; one for the click which preceded the whistler, and a second, much longer path, over which the whistler and its echoes propagated.

Eckersley was able to show in 1931 that Earth's magnetic field permits a suitably polarized wave to pass completely through the ionosphere. This was in accordance with the magneto-ionic theory. Tremellen had on one occasion observed that during a summer thunderstorm at night every visible flash was followed by a whistler. This served as the first definite evidence of the relationship between lightning discharges and whistlers—the click being the lightning discharged which caused the whistler. As the Marconi workers continued their research, they also discover a new type of atmospheric which sounded somewhat like the warbling of birds. Since the sound tended to occur most frequently at dawn, they gave it the name *dawn chorus*.

L.R.O. Storey of Cambridge University began an intensive study of whistlers in 1951. He confirmed Eckersley's law which showed that most whistlers originate in ordinary lightning discharges, and found the path of propagation of the whistler to be along the lines of the Earth's magnetic field.

Other interesting sounds include "tweeks," which have been described by one listener as a cross between chirping birds and a hundred little men hitting iron bars with hammers. The ephemeral "dawn chorus," is a cacophony of sounds that resembles nothing else on Earth.

Tweeks are abrupt descending notes that resemble "pings." They are usually heard at night during the winter and early spring.

The dawn chorus has been variously described as sounding like birds at sunrise, a swamp full of frogs, or seals barking. In fact, it varies constantly. "Hooks," "risers," and "hiss" are all part of the dawn chorus, but sometimes these sounds are also heard as solitary events. A hook starts out like a whistler but then abruptly turns into a rising tone. Risers increase in frequency from beginning to end. And hiss sounds just like its name. Appropriately enough, dawn chorus is best heard around dawn, but it may occur at any time of the day or night. The dawn chorus is most likely to be heard during and shortly after geomagnetic storms.

A whistler receiver

You cannot easily buy a radio capable of tuning 1 to 10 kHz, but it is possible to build a sensitive receiver to permit you to hear the sounds of natural radio. We will explore what causes these sounds and see how you can study them yourself.

A whistler receiver, even though whistlers and related emissions occur at acoustical frequencies, is a receiver of radio signals. To hear whistlers, you must intercept their electromagnetic energy with an antenna and transform it to the mechanical vibrations to which our ears respond.

The "classic" whistler receiving system consists of an antenna for signal collection, an amplifier to boost the signal level, and headphones or a speaker to transform the signal to sound waves. (A magnetic tape recorder can be substituted for the headphones or speaker.) A whistler receiver can simply be an audio amplifier connected to an antenna system. However, powerful man-made interference immediately above and below the frequencies of interest tend to seriously overload receivers of this sort and make reception of whistlers and related phenomena difficult.

To overcome these problems, a good whistler receiver circuit includes a circuit called a low-pass filter that attenuates all signals above 7 kHz. This greatly reduces interference from such sources. To escape interference from AC power lines and other forms of non-natural radio noise, the receiver is designed for portable operation "in the field" away from such interference sources.

A sensitive dual FET whistler receiver is shown in the photo in Figure 14-1, and in the schematic diagram in Figure 14-2. The antenna is first fed to an RF filter

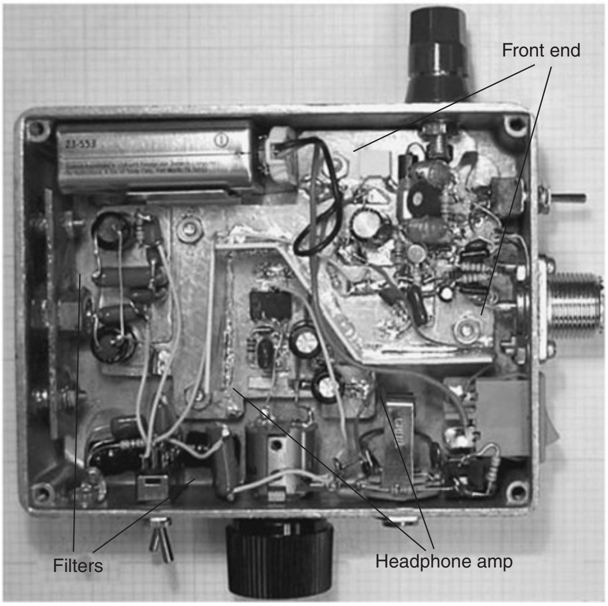

Figure 14-1 *VLF whistler receiver*

which is composed of resistors R1 through R4 and capacitors C1 through C3. This filter only allows the desired frequencies to be sent to the next section of the receiver. The RF filter is next fed to the 60 Hz filter or "trap" which filters out 60 Hz power line frequencies before the RF signals are amplified. The 60 Hz filter can be switched in and out of the circuit with switch S1:a. The dual matched FET transistor Q1 along with transistor Q2 are used to amplify the RF signals of the desired band of frequencies. The amplified signals are then fed to the input of a set of selectable filters which can be switched in and out of the circuit. Toggle switch S2 is used to switch in the high-pass filter circuit, while switch S3 is used to switch-in the low-pass filter. You can select either or both filters if desired.

The output of the low-pass filter circuit is coupled to the final audio amplifier stage via capacitor C17. Potentiometer R12 controls the audio level entering the LM386 audio amplifier at U1; also note that R12 also contains a switch S5, which is used to supply power to the LM386 audio amplifier. The output of the LM386 is coupled to an audio output jack via capacitor C20. Note that capacitor C18 is coupled to a recorder output jack to allow a chart recorder or analog to digital converter to be used to monitor and record the output from the whistler receiver. The whistler receiver is powered from a 9 volt battery for both stationary and field receiving applications.

Building a whistler receiver

Before we start building the Whistler receiver, you will need to locate a clean, well lit and well ventilated area. A workbench or large table top to spread out all you diagrams, components and tools for the project. First, you will need a small pencil tipped 27 to 33 watt soldering iron, a spool of 22 ga. 60/40 tin/lead solder, and a small jar of "Tip Tinner." "Tip Tinner" is a soldering iron tip cleaner/dresser and is available at your local Radio Shack store. You will also want to locate a few small tools such as a pair of end-cutters, a pair of needle-nose pliers, a pair of tweezers and a magnifying glass. Look for a small flat-blade screwdriver as well as a small Phillips screwdriver to round out your tools for the project. Next locate all the components for the Whistler project as well as all of the diagrams including the schematic, parts layout diagram and the necessary charts and tables and we will begin.

In order to construct a stable well performing radio receiver, it is recommended practice to build the Whistler receiver on a printed circuit board, with large ground plane areas. You will need to use good RF building practices, such as keeping leads as short as possible, clean solder joints, etc. With your circuit board in hand we will begin building the circuit.

Take a look at the schematic diagram and you will notice that there are dotted or dashed lines. These dashed lines represent shielded compartments on the circuit board. These shielded areas are used to separate different portions of the circuit from one another. You can create a shielded area by soldering a vertical piece of scrap circuit board material onto the circuit board ground plane areas on the circuit board. You could also modularize the circuit building three small circuit boards, one for the RF filter and "trap" circuit, one for the high-pass and low-pass filter, and a third board for the audio amplifier and power circuit. Small metal shield covers could be built using scrap circuit board material.

First, you will need to locate the resistors for the project, and you will also want to refer to the resistor color code chart in Table 14-1, which will help you to identify each of the resistors. Resistors generally have three or four color bands which are used to denote the

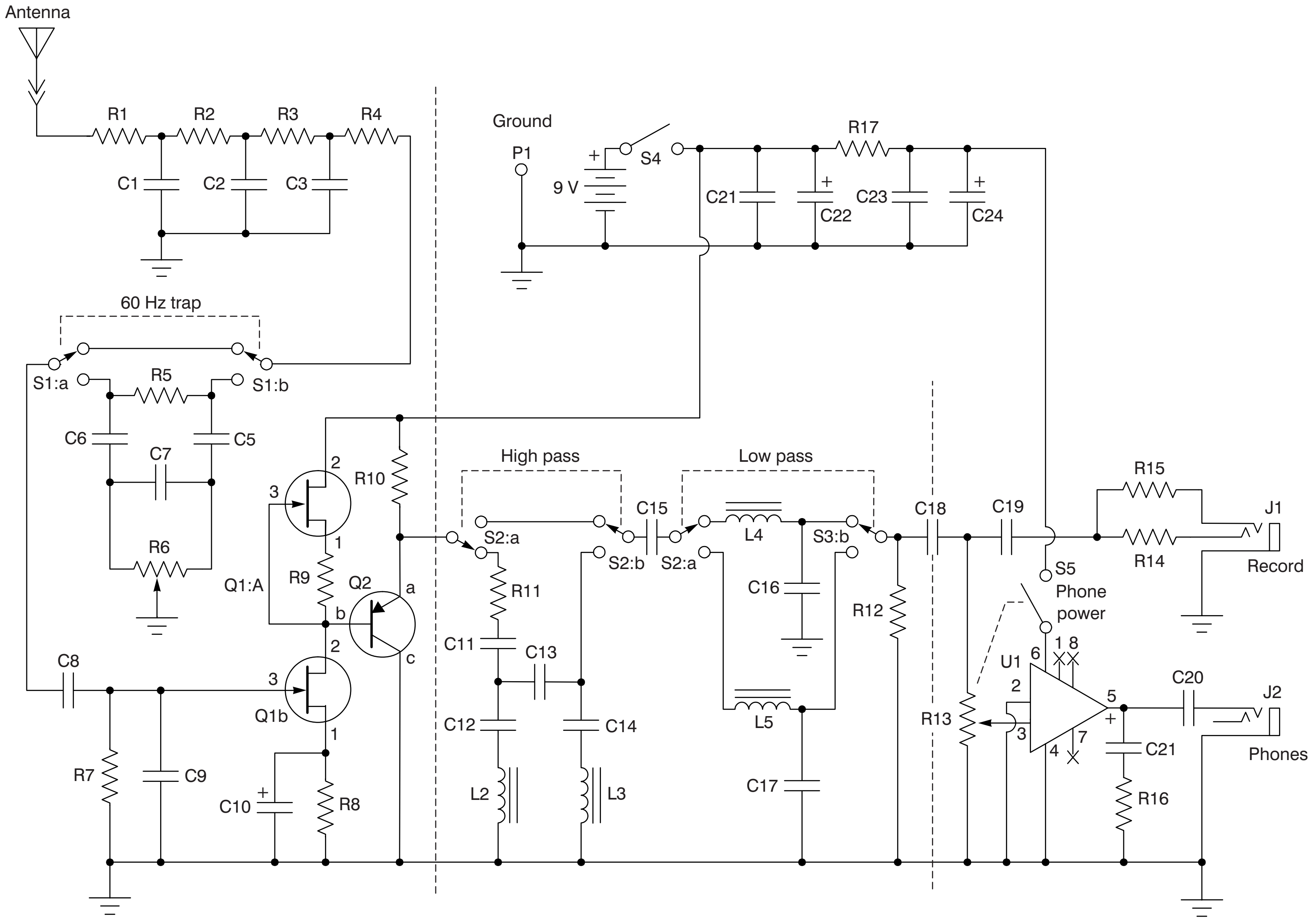
Antenna
R1
R2
R3
R4
C1
C2
C3
Ground
P1
+
9 V
S4
R17
C21
+
C22
C23
+
C24
60 Hz trap
S1:a
S1:b
R5
C6
C5
C7
R6
2
3
1
R10
Q1:A
R9
Q2
a
b
c
2
3
1
Q1b
C8
R7
C9
+
C10
R8
High pass
S2:a
S2:b
C15
R11
C11
C13
C12
C14
L2
L3
Low pass
S2:a
L4
S3:b
C16
L5
C17
R12
C18
C19
S5
Phone
power
R13
U1
6
1 8
2
3
4
7
5
+
R15
R14
J1
Record
C20
J2
C21
R16
Phones

Figure 14-2 *FET VLF receiver schematic. Courtesy of Eric Vobel*

Table 14-1

Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

resistor's value. The first color band will be closest to one edge of the resistor. The first color band represents the first digit of the value. Resistor R1, for example, is a 10k ohm resistor. Its first color band would be brown (1), the second color band would be black (0) and the third color band which denotes the resistor's multiplier value would be orange or (000), so 10,000 or ten thousand ohms. A fourth color band represents the resistor's tolerance value. The absence of a fourth color band denotes a 20% tolerance, while a silver band represents a 10% tolerance and a gold denotes a 5% tolerance resistor. Use the color code chart to identify resistors R1, R2 and R3. When you are confident that you know the color codes and can identify these resistors, you can go ahead and solder these resistors onto the circuit board. Use your end-cutters to cut the excess component leads flush to the edge of the circuit board. Now locate another grouping of resistors and install them to the circuit board, then solder them in place on the PC board. Remember to trim the extra component lead lengths from the PC board. Install the remaining resistors onto the circuit board, until all have been mounted to the board, solder them in place and trim the leads as necessary.

Next, we are going to install the capacitors for the Whistler receiver. Capacitors come in two major types, i.e. electrolytic and non-electrolytic. Generally the non-electrolytic capacitors are the small types such as ceramic disks, mylar or polyester types. These small capacitors may have their actual value printed on them, or if the capacitors are very small may instead have a three-digit code printed on them. Refer to the chart in Table 14-2, which will help you identify these capacitors. A .001 µF capacitor, for example, would have (102) marked on the body of the device to denote its value of .001 µF. Electrolytic capacitors are generally the larger capacitors with polarity marking on them. You will find either a black or white color band on them or a plus (+) or minus (–) marking on them. Electrolytic capacitors have polarity, so this means that you have to orient them correctly while installing them in order for the circuit to work properly. Reversing electrolytic capacitors can often cause the circuit to malfunction, and it can destroy the component and/or the circuit itself, so be careful when installing them. Refer to the schematic and the parts layout diagram when installing electrolytic capacitors.

Let's begin installing the capacitors now. Try and separate the electrolytic capacitors from the other types of capacitors. Locate a few of the non-electrolytic capacitors, identify them and install a few of the capacitors on the circuit board. Go ahead and solder these capacitors onto the PC board, once you are sure you have correctly identified them. Trim the excess

Table 14-2
Capacitance code information

This table provides the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Some capacitors actually have the numeric values stamped on them; however, many are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

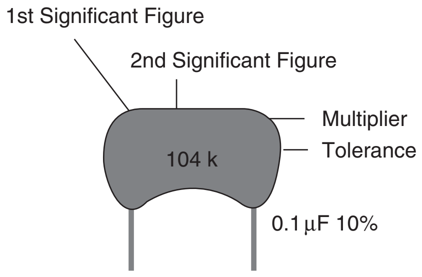

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 Pf	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

component leads from the PC board. Install a few more non-electrolytic capacitors and solder them to the board, remember to cut the extra leads from the board. Install the remaining non-electrolytic capacitors to the circuit board and solder them in place.

Next, locate the electrolytic capacitors and we will install them on the PC board. Refer to the schematic and parts layout diagrams to make sure you can orient the capacitors correctly before soldering them. Place a few electrolytic capacitors on the board and solder them in, trim the excess leads as necessary. Add the remaining electrolytic capacitors on to the PC board and solder them in place.

The Whistler receiver also has a number of coils in the circuit, most of them are low cost and readily available from Mouser Electronics, see Appendix. Identify the coils for the project and place them in their respective locations on the PC board. Now, solder them in place and remember to trim the excess component leads flush to the edge of the circuit board.

Looking at the circuit you will notice three transistors, more correctly one dual FETs and one conventional transistor. Before we go ahead and install the semiconductors, take a quick look at the semiconductor pin-out diagram shown in Figure 14-3. The FETs are very sensitive to static electricity damage, so be extremely careful when handling them, use a grounded anti-static wrist-band. The FETs have three leads much like a regular transistor but their pin-outs are a bit different. An FET will usually have a Drain, Source and Gate lead; see pin-out diagram for each particular FET, when installing them on the PC board. The conventional transistor at Q2 is a PNP type which has a Base, Collector and Emitter lead, once again refer to the transistor pin-out diagram, the schematic and parts layout diagrams when installing the transistor and FET. Installing the FET and transistor incorrectly will prevent the circuit from working correctly and may damage the component as well, so be careful when installing the FET and transistor.

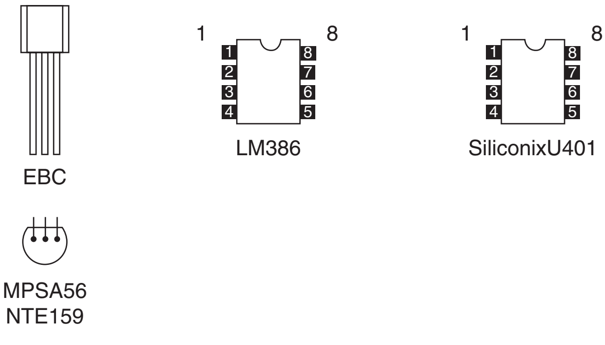

Figure 14-3 *Semiconductor pin-outs*

The receiver has a single integrated circuit at U1, an LM386 audio amplifier IC. As an insurance policy against a possible circuit failure in the present or possible future date, it is recommended that you use an integrated circuit socket. Integrated circuit sockets will usually have a notch or cut-out at one end of the socket; this is supposed to represent pin 1 of the IC socket. You will have to align the socket correctly so that pin 2 of the IC socket connects to the ground bus. Note that pin 1 is not used in this application. When installing the IC into the socket, you will note that the IC itself will have either a small indented circle, a cut-out or a notch at one end of the plastic IC body. Pin 1 of the IC is just to the left of the notch. Insert the IC into the socket making sure pin of the IC is aligned with pin of the socket.

Take a short break and when we return we will check over the printed circuit board for “cold” solder joints and possible “short” circuits. Pick up the PC board with the foil side facing upwards toward you. First we will inspect the PC board for possible “cold” solder joints. Take a look at all of the solder joints. All the solder joints should appear to look clean, shiny and bright, with well formed solder connections. If any of the solder joints appear dull, dark or “blobby,” then you should un-solder and remove the solder from the joint and re-solder the joint all over again, so that it looks good. Next, we will inspect the PC board for possible “short” circuits. “Short” circuits can be caused by small solder balls or blobs which stick to the underside of the board. Many times rosin core solder leaves a sticky residue on the PC board. Sometimes “stray” component leads will adhere to the underside of the board and often they can form a bridge across the PC traces on the PC board causing a “short” circuit, which can damage the circuit upon power-up, so be careful to remove any component leads or solder blobs from the underside of the board.

The VLF Whistler receiver was housed in a metal chassis box to prevent “stray” RF from affecting the circuit operation. The main circuit board can be mounted to the bottom of the enclosure with 1/8″

plastic standoffs. The 60 Hz "trap" filter switch, the high-pass filter and low-pass filter switches, along with the power on-off switch, were mounted on the front panel of the chassis box. The headphone power amplifier power switch, the headphone jack and volume control at R13 were also mounted on the front panel of the chassis box. The SO-238 antenna jack, the 1/8″ record jack and the ground binding post were all mounted on the rear panel of the chassis box. A 9 volt battery holder was used to secure the battery to the left rear corner of the chassis box. Once the circuit has been mounted in the selected chassis, you can move on to applying power and testing the VLF receiver. Connect a fresh 9 volt battery to the VLF receiver, connect a pair of headphones and a whip antenna to the antenna terminals. Turn the power switch to the "ON" position and you should hear a hissing sound from the headphones. If the receiver is totally quiet and you hear nothing, you have a problem! First, you will need to disconnect the battery, the headphones and the antenna, and you will need to carefully examine the circuit board. The most common cause for circuit failure is improper installation of electrolytic capacitors and diodes. Remember that these components have polarity and it must be observed for the circuit to work correctly. Look over the circuit very carefully. It is often advisable to have a second pair of eyes. Have a knowledgeable electronics friend help you examine the circuit board for mistakes. Another cause for possible errors are the incorrect placement of resistors. Often colors are mistaken and resistors are installed in the wrong location. Another possible problem can be the installation of the semiconductors, namely in this circuit the transistors or FETs. Check the manufacturer's pin-out diagrams for the particular device and carefully re-examine the installation. Hopefully at this point you have found your error and corrected it and you can then re-connect the battery, antenna and headphones and re-test the VLF receiver to make sure it is now working.

Using your VLF receiver

The whistler receiver works best out doors in open areas, with a vertical antenna of roughly 3 to 12 feet in height. It can also be used in wooded or obstructed areas with wire antennas between roughly 50 to 200 feet in length connected instead of the vertical antenna. Wire antennas should be insulated and supported off the ground as high as you can place them. The performance of a wire antenna improves the more you can get it in the clear and the more vertically it can be positioned. If you do use a wire antenna, you may have to reduce the value of the input resistor R1 for best results.

The whistler receiver also needs a ground connection, also called a counterpoise. A simple but often effective ground can be provided by your body in contact with the ground terminal (BP2). I have used an improvised finger ring attached to a wire connected to the ground terminal. Other places to make a ground connection include car bodies, wire fences, or other large metal objects not connected in any way to an AC power circuit. Another approach is to use short sections of copper tubing as ground stakes which can be driven into the ground and connected to your receiver. Try several different grounding schemes to see which one provides the quietest reception. Don't be too surprised, however, to discover that you wind up having to still make body contact with the ground terminal for quietest reception!

The prototype receiver, with a whip antenna, was mounted on a piece of wood approximately 2″ × 60″. You can install the antenna at one end of the long rectangular piece of wood (it can be mounted on a bracket or just taped on) and the receiver can be mounted at the other end of the piece of wood, about one foot down from the antenna using duct tape, two-sided foam tape or a mounting bracket. A length of wire was run from the receiver's ground terminal down the wood and taped in place at several places. Now I had a handy walking stick that was easy to carry and use! The base can be stuck into any convenient hole, jammed into a rock pile, clamped onto a fence or vehicle, or even held by hand.

For the best chance of hearing whistlers and related phenomena, you will have to put some distance between the whistler receiver and the AC power grid. The bare minimum distance is about one quarter mile from any AC power line; the more distance you can put between the receiver and AC power lines, the better. You'll need to do some exploration to discover quiet sites near you. Once you arrive at a potential listening site, set up your receiver, put in a ground connection, and see what happens. What you will hear depends to some extent on the time of day, but you can always expect some sharp, crackling static. The intensity and volume of this static will depend on propagation conditions and where thunderstorms are in relation to your location. Chances are that you'll also hear some power line hum, but hopefully it won't be very loud. If your listening site is too noisy, you need to try another listening site.

If your ears have good high frequency response, you will probably hear a continuing sequence of one-second tones. These are from the OMEGA radio-navigation system. OMEGA is transmitted from 10 to 14 kHz, and OMEGA transmitters are very powerful. You may also hear anything that can produce an electrostatic discharge, particularly if the humidity is low. These miscellaneous noise sources can include wind noise, the buzz of insects flying near the antenna, dry leaves or grass moving, and even the electrostatic charges that build up on your clothing. Passing vehicles will often produce noise from their electrical and ignition systems. And if you have a digital watch, keep it away from the antenna or you'll be listening to it instead of whistlers!

Anything you hear other than whistlers and related phenomena or OMEGA signals are likely to be the result of extraneous signals "overloading" your receiver. Burbling sounds (possibly mixed with OMEGA tones) are caused by military signals in the 15 to 30 kHz range, perhaps overloading the receiver. A "ticking" sound at a 10 Hz rate is from Loran-C radio-navigation signals at 100 kHz. In general, don't be surprised if you experience "overloading" from transmitters operating on any frequency if they are within sight of the location where you're using your whistler receiver! Table 14-3 illustrates some very low frequency stations around the world that broadcast low frequency signals, these signals are located at the upper edge of the VLF receiver's range.

Table 14-3

Very low frequency radio stations

Site	*ID*	*Frequency kHz*	*Power KW*
Cutler, ME	NAA	24.0	1000
Jim Creek, WA	NLK	24.8	250
Lualualei, HI	NPM	21.4	566
LaMoure, ND	NML4	25.2	
Aquada, Puerto Rico	NAU	40.75	100
Keflavik, Iceland	NRK	37.5	100
Niscemi, Italy		39.9	25
Harold E. Holt, Australia	NWC	19.8	1000
Rhauderfehn, Germany		18.5	500
Rosnay, France	HWU	15.1	400
St. Assie, France	FTA	16.8	23
Bombay, India		15.1	
Tavolara, Italy	ICV	20.27	43
Ebino Huyshu, Japan		23.4	
Noviken, Norway	JXN	16.4	45
Arkhanghelsk, Russia	UGE	19.4	150* input pwr
Batumi, Russia	UVA	14.6	100* input pwr
Kaliningrad, Russia	UGKZ	30.3	100* input pwr
Matotchkinchar, Russia	UFQE	18.1	100* input pwr
Vladivostok, Russia	UIK	15.0	100* input pwr
Anthorn, United Kingdom	GQD	.019	42

Time of listening

Most of the VLF signals, such as clicks, pops, and tweeks, can be heard almost at any time of the day or night. Whistler occurrence is a function of both thunderstorm occurrence and propagation conditions through the ionosphere. The rate is higher at night than during the day because of the diurnal variation in the D-region absorption, which is highest when the Sun is above the horizon. At night, the D-layer does not exist since it is ionized by solar ultraviolet rays. The whistler rate, or the number of whistlers heard per minute, has a marked dependence on sunspot activity, the rate increasing with the sunspot number. Whistlers are heard only when sufficient ionization exists along the path to guide the waves along the Earth's magnetic field. This ionization is assumed to be supplied from the Sun during a solar. To hear a whistler or other signals such as *dawn chorus*, you will no doubt have to get away from the power line and industrial noise. The listening time for these signals will be from near local midnight to early morning hours.

Observing and recording hints

Remember the best location to observe VLF signals is out in the countryside, in an area where few cars or trucks are likely to pass. You will want to get as far away from power lines as possible for optimum results. The lower the power line noise is, the stronger the VLF emission signals will be.

Some whistlers are related phenomena and so short-lived and impossible to predict that you might want to consider using a tape recorder to record the signals in the field unattended. Take a portable, battery-operated, cassette tape recorder with you and use a long audio cable to run to the tape recorder from the audio output of the VLF receiver. Any tape recorder with an external microphone input can be used. Best results are obtained with a recorder which has no ALC or automatic level control. The VLF receiver shown has an output jack which can be coupled to a recorder input for field recording. You may have to place the tape recorder away from the receiver to avoid picking up motor noise.

After returning home with your recording, if your receiver was out in the field, you can "play" your recording back into the sound card of your personal computer using an FFT or DSP sound recording program to analyze the recording. For those who are going to use their whistler receiver as a stationary receiver at home, you could use your PC to record on a continuous basis. Audio recording and FFT programs are available on the Internet for free. Good luck and happy exploring!

Additional notes

You can view spectrograms of whistlers as well as hear excellent recordings of many more whistlers than we used in generating the audio clip for this article, courtesy of NASA: http://image.gsfc.nasa.gov/poetry/inspire/advanced.html

A free Fast Fourier Transform (FFT) Waveform Analyzer program and an Audio Waveform Generator program can be downloaded for use on your own computer at:
FFT discussion:
http://www.dataq.com/applicat/articles/an11.htm

Some US Government agencies doing whistler research are:
NASA's Godard Space Flight center invites participation in a program called "INSPIRE" (an acronym for Interactive NASA Space Physics Ionosphere Radio Experiments) http://image.gsfc/poetry/inspire

NOAA: http://ngdc.noaa.gov/stp/stp.html
http://www.ngdc.noaa.gov/stp/stp.htm
gopher://proton.sec.noaa.gov

Some of the university labs involved with whistlers are:
University of Alaska Geological Institute
University of Iowa: http://www-w.physics.uiowa.edu/~/jsp/polar/sounds/whistlers.html
Stanford University: http://www.stanford.edu/~vlf/Science/Science.html
[An excellent site, but I could only access it via the "Google" search engine]
UCLA: http://www.physics.edu/plasma-exp/Research/TransportWhistlers
Kyoto University, Kyoto, Japan

Chapter 15

Induction Loop Receiving System

Parts list

Induction Loop Transmitter I

TX Mono amplifier or receiver - 5-10 watts output

R1 10 ohm 10 watt 10% resistor

L1 loop coil around room (see text)

Induction Loop Transmitter II

R1 10k ohm potentiometer chassis mount

R2 220 ohm ¼ watt resistor, 5% (red-red-brown)

R3 2.2k ohm ¼ watt, 5% resistor (red-red-gold)

R4 1 ohm ½ watt, 5% resistor (brown-black-gold)

R5 10 ohm 10 watt, 10% resistor

C1 10 µf, 25 volt electrolytic capacitor

C2,C5 100 nF, 35 volt poly capacitor

C3 100 µF, 35 volt electrolytic capacitor

C4 470 µF, 35 volt electrolytic capacitor

C6 2200 µF, 35 volt electrolytic capacitor

U1 TDA2002/3 6 to 10 watt audio amplifier module

HS heatsink for U1

L1 loop coil (see text)

Misc PC board, wire, hardware, chassis, etc.

Induction Loop Receiver

R1 1 megohm ¼ watt, 5% resistor

R2 10k ohm ¼ watt, 5% resistor

R3 1000 ohm potentiometer (tone)

R4 470 ohm ¼ watt, 5% resistor

R5 10k ohm potentiometer (volume)

R6 10 ohm, ¼ watt, 5% resistor

C1 39 nF, 35 volt mylar capacitor

C2 1 µF, 35 volt electrolytic capacitor

C3,C7 .1 µF, 35 volt disk capacitor

C4 4.7 nF, 35 volt poly capacitor

C5 10 µF, 35 volt electrolytic capacitor

C6 220 µF, 35 volt electrolytic capacitor

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

Q1	BC548 transistor
U1	LM386 audio amplifier IC
L1	800 turns on a 9 × 70 mm ferrite rod (see text)
S1	SPST on-off switch
B1	9volt transistor radio battery
PH	8 ohm headphones
Misc	PC board, wire, ⅛″ phone jacks, battery clip, case

Induction loop technology has been around for many years, for use with the hearing impaired. Induction loop communication systems were originally designed to provide assistance in the school classrooms for the deaf, but today induction loop communications technology is now used for many other applications, such as museum displays, theater communications, mine and cave communication systems, and signal tracing. Induction loop systems have been installed in churches, public halls and auditoriums, schools, lecture halls, cinemas, service counter windows, drive-thru order and pick up windows, information kiosks, offices, airports, train stations, parks, tour and guide buses, automobiles, boats, riding academies, and homes.

Induction loop amplification operates on the basic principle of electronics called magnetic induction. When an electrical current is amplified and passes through a loop of wire, a magnetic field is generated around the wire. The field varies in direct proportion to the strength and frequency of the signal being transmitted. When another wire is placed close to the field of the first wire, an identical electrical current is induced within it. The second current can be amplified and converted into an exact duplicate of the original sound signal.

In technical language, a current is said to be “induced” in the second wire. Hence the term “induction.” In an induction loop system, the induction loop coil in the transmitter sends an electromagnetic signal, which is “induced” to the receiver’s induction loop coil in the receiving unit. An Induction Loop System consists of an amplifier and a loop, the amplifier is connected to a sound source such as a TV, a radio, a public address system or a dedicated microphone. It then amplifies this sound signal and sends it out, in the form of an alternating current, through the loop. The loop itself consists of insulated wire, one turn of which is placed around the perimeter of the room in a simple loop system. When the alternating current from the amplifier flows through the loop, a magnetic field is created within the room. The magnetic field “induces” the signal into the Induction loop coil in the induction receiver unit.

In the portable induction receiver, the induction coil picks up the induced signal from the transmitting loop coil. The fluctuations in the magnetic field are converted into alternating currents once more. These are in turn amplified and converted by the induction loop receiver into sound. The magnetic field within the loop area is strong enough to allow the person with the hearing aid or Induction Loop Receiver to move around freely in the room and still receive the sound at a comfortable listening level.

Many modern day hearing aids incorporate a telecoil into the hearing device. So if a hearing aid user switches to the T or MT position on their hearing aid, the telecoil in the hearing aid picks up the induced signal from an induction loop transmitter and amplifies the signal for the hearing aid user to hear. This is a convenient feature for hearing aid users, in classrooms, churches, and museums with induction coil transmitters.

The listener can receive an audio signal without being physically or electrically connected to the loop, so in effect induction loop transmission is a form of wireless telephony. The listener receives the sound in one of two ways; either through the use of the T-coil or telephone switch on a hearing aid instrument or through an induction receiver that can work with earphone, which we will discuss in our project.

The basic induction loop system consists of a “transmitter” section and a “receiver” section, as shown in Figure 15-1. The “transmitter” section consists of a microphone or signal source, an audio amplifier to boost the signal, and a large loop or wire around the room or listening area. The “receiver” section consists of a smaller loop coil, an audio amplifier which boost the incoming audio and a pair of headphones.

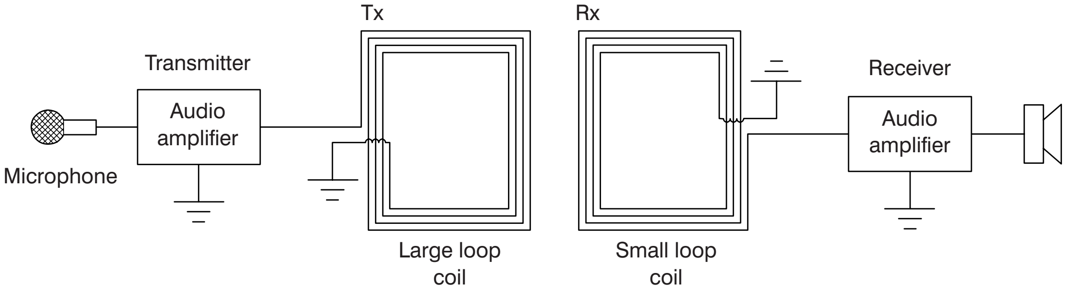

Figure 15-1 *Induction loop communications system block diagram*

"Transmitter" loop unit

You can easily build your own induction loop communication system. First we will begin by describing the "transmitter" section of the system, as shown in Figure 15-2. In its simplest form an induction loop "transmitter" can consist of a mono or stereo Hi-Fi receiver, a resistor and a large loop coil placed around the perimeter of a house, room, or your easy chair. If you have a stereo receiver with a speaker output A-B switch, you can select the (B) output for your induction loop system. You would place an 8 ohm, 10 watt resistor across the output of the left channel's output and then place an 8 ohm, 10 watt resistor in series with a large loop coil as shown. When you wanted to use the induction loop system, you would switch to the (B) speaker output; it's that simple.

If you wanted to construct your own amplifier specifically for an induction loop audio system, you could utilize the 6 to 10 watt audio amplifier shown in Figure 15-3. In order to drive an induction loop, you would need an amplifier with at least about 5 watts of power or more to have a workable system. The amplifier shown in this diagram would work just fine for driving a large loop coil. An audio input signal is fed into the audio input section at the point marked (A) on the schematic. Your audio source could be adjusted via the potentiometer at R1. Note, that you may need an additional pre-amp ahead of the amplifier if you use a low output microphone. You can feed a radio or TV signal directly into the potentiometer at R1 to drive the induction loop "transmitter." The audio amplifier module shown is an 8 watt National Semiconductor LM2002 module, but you could also use a higher power module such as their LM2005. The output of the audio amplifier is wired in series with an 8 to 10 ohm, 10 watt

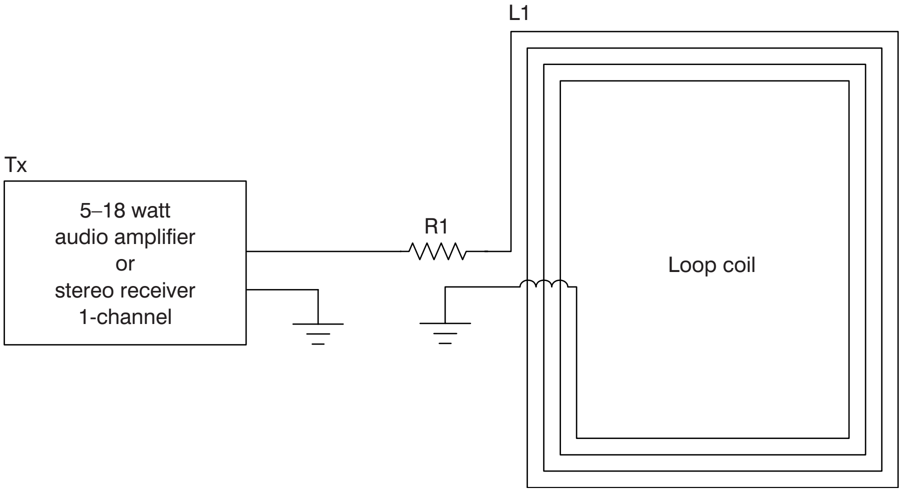

Figure 15-2 *Induction loop transmitter I*

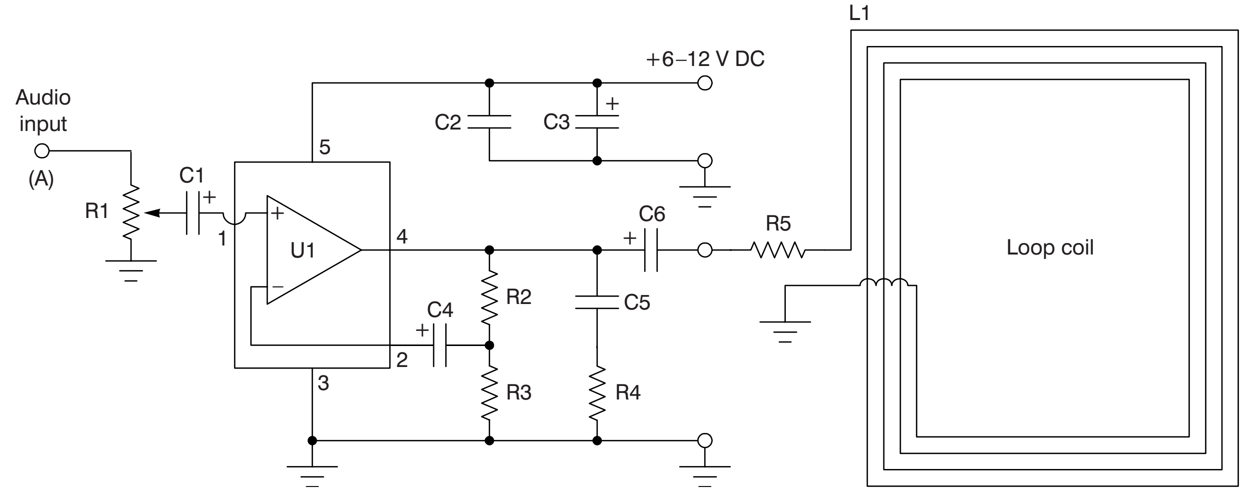

Figure 15-3 *Induction loop transmitter II*

resistor in order to drive the coil. This resistor also protects the output of the audio amplifier.

The inductive loop is a winding of wire through which an audio signal passes, which creates a field within which the audio signal can be detected by appropriate equipment, either a telecoil in a hearing aid or an induction loop receiver. The loop coil is basically a simple 4 or 5 turn loop placed around the room fed from an audio amplifier. The loop then converts the sound to a magnetic field that can be received by a second loop coil and amplified. Note that more turns in the transmission coil result in a stronger transmitted signal. The loop should have a DC impedance of at least 8 ohms, or you can add a resistor in series with the coil to reach 8 ohms.

Pick a path around the room on the ceiling, or under the carpet. A room measuring 15′ × 20′ will have a perimeter of 70′, which will require a minimum of four complete turns of 22 gauge wire, six turns of 20 gauge wire, or nine turns of 18 gauge wire. Pick a path around the room near the ceiling, and wind (or hang) your "coil," or loop, starting and ending at the amplifier. If the back of your amplifier says something like "8 ohms minimum," double the required wire length (load). Buy 120% of the wire you'll need to ensure that you can complete the loop. You should add a power resistor to ensure an appropriate load and reduce wire length, but the amplifier output will be divided between the loop and the resistor, proportionately, which means less radiated energy. This solution requires a lot of energy so try not to compromise. If required, get 4 ohms at 10 watt minimum. There's another and simpler approach: I have a coil built into my favorite chair. It's only 18″ in diameter, but includes 120′ of 28 gauge wire (0.071 ohms per foot). That's about 25 turns of that relatively thin wire.

Another approach to constructing a large loop is to take a long length of 25 pair telephone cable 'Inside Wire' that's long enough to go around the perimeter of a room. Then get a 66 block (or some other kind of punch down block) and punch down the wires so they connect in a multiple turn loop. In this way you can add or subtract turns as you wish if you need that later. You could tack this "telephone" cable to the ceiling of the basement underneath the room you choose.

Connect either end of the loop coil wire to the plus (+) or red terminal and the other end to the minus (−) black terminal, on the audio amplifier's speaker terminals. Most amplifiers or stereo receivers have SPEAKER A and B switches; put your coil circuit on the unused output (B). Remember to place an 8 to 10 ohm, 10 watt resistor in series with the speaker output and the coil loop windings.

"Receiver" loop unit

The induction loop receiver, shown in the photo Figure 15-4 and the schematic diagram in Figure 15-5, is very sensitive and can be the basis for a monaural induction loop wireless headphone system. The induction coil receiver is used inside the area of transmission loop. The receiver coil at L1 consists of 800 turns of #26 ga. enamel wire wound on a 9 × 70 mm ferrite core.

Another approach to constructing a small loop coil for your inductive loop receiving coil is to obtain a "telephone pickup coil" if available or a suitable coil from some other device. The coil in the prototype was salvaged from a surplus 24 volt relay. Actually, two relays were needed since the first was destroyed in the attempt to remove the surrounding metal so that a single solenoid remained. Epoxy putty was used to secure the thin wires and the whole operation was a bit of a challenge. A reed relay coil will give reduced sensitivity but would be much easier to use. The experimentally inclined might try increasing the inductance of a reed relay by replacing the reed switch with soft iron. Avoid shielded inductors or inductors with iron pole pieces designed to concentrate the magnetic field in a small area or confine it completely (as in a relay or transformer) unless you can remove the iron. The resulting coil should be a simple solenoid like wire wrapped around a nail. Don't try to wind your own—it takes too many turns. Evaluate several coils simply by listening. Coils with too little inductance will sound "tinny" with poor low frequency response, and other coils will sound muffled, especially larger iron core coils. This prototype was tested with a large 100 mH air core coil with superb results but the 2″ diameter was just too big for this application. The induction receiver develops a 1–2 mV signal across its internal coil from the fluctuating magnetic field. Transistor Q1 provides pre-amplification to drive a headphone amplifier. The tone at R3 control varies the amount of gain in the high frequencies. Capacitors C1 and C2 block AC hum (C2 also blocks DC to the amplifier input). C3 functions as a low-pass filter to prevent oscillation. The output of the volume control at R5 is fed to the input of an LM 386 audio amplifier IC. In this configuration the amplifier has a gain of 200. Capacitor C7 and R6 form the output "boost" circuit which drives the output to the speaker through C6. The output of the LM385 at pin is used to drive an 8 ohm headphone. The induction loop receiver circuit can be powered from a 9 volt DC power supplies or batteries.

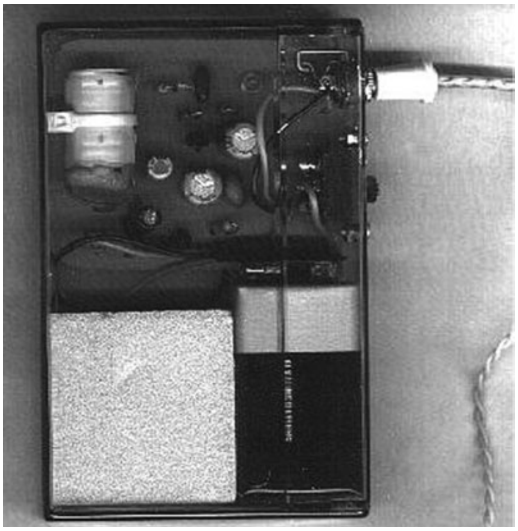
Figure 15-4 *Induction receiver*

Are you ready to begin constructing the induction receiver circuit? First you will need to locate a clean, well lit work table or work bench. You will want a large table so you can spread out all the components, the diagrams and tools that you will need to build the project. Next you will need to locate a small pencil tipped 27 to 33 watt soldering iron and some 60/40 tin/lead rosin core solder for the project. Try and locate a small container of "Tip Tinner," a soldering iron tip cleaner/dresser which conditions the soldering iron tip; this is available from your local Radio Shack store. You should also locate a few small hand tools such as a pair of end-cutters, a pair of needle-nose pliers, a pair of tweezers, a small flat-blade and a Phillips screwdriver and a magnifying glass.

The induction loop receiver prototype was constructed on a small glass-epoxy circuit board which measured 2¼″ × 3″. A circuit board makes for a more reliable and professional circuit. Place all the diagrams in front of you on the work table along with all of the components for the induction loop receiver and we will begin building the project.

The induction loop receiver project contains about four resistors and two potentiometers. Locate the resistors using the color chart in Table 15-1, this will help you identify each of the resistors. Place the resistors into their respective locations on the circuit board, then solder them in place on the PC board. Locate your end-cutters and trim the excess component

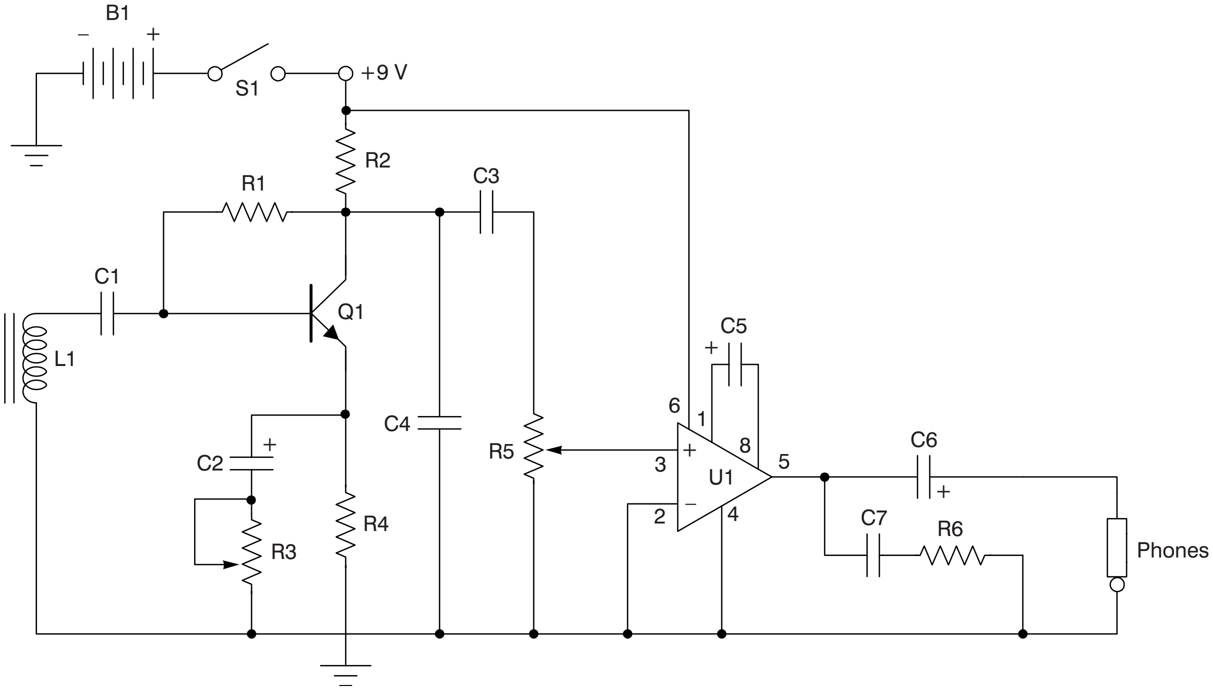

Figure 15-5 *Induction loop receiver schematic*

leads, by cutting them flush to the edge of the circuit board. Potentiometer R3 is a CP board mounted type, you can go ahead and install it on the board and then solder it in place; you can remove any excess leads if necessary.

Next we are going to move to installing the capacitors for the project. There a two distinct types of capacitors—small ceramic, mylar or polyester capacitors and larger electrolytic types. The small capacitors will most likely have a three-digit code stamped on them and not their actual value, since often the capacitor body is very small. Refer to the chart shown in Table 15-2, which illustrates the capacitor codes. You can search through the parts pile and try and identify the capacitors against the code chart and place the correct value into its respective location on the circuit board. After all of the small capacitors have been placed on the PC board, you can solder them in place. Remember to cut the excess leads from the components. Now, locate the larger electrolytic capacitors and place them in front of you and determine where they will go on the circuit board. Once you are certain where the capacitors should be placed on the board, you should take another look to make sure that you have installed the capacitors correctly with respect to its polarity. If an electrolytic capacitor is not installed correctly the circuit will not work properly and may be damaged upon power-up. So you will want to install them correctly the first time through.

Note, there are a number of capacitors in this project, and in looking over the schematic, you will notice that a few of the capacitors will have polarity marking on them. These capacitors are electrolytic types and you must observe the polarity when installing them on the circuit board in order for the circuit to work properly. Electrolytic capacitors will have either a white or black band at one side of the capacitor body or they may be marked with a plus (+) or minus (–) marking on the capacitor body and this denotes the polarity. Refer to the schematic and parts layout diagrams when installing electrolytic capacitors in order to install them correctly with respect to their polarity markings.

Before we go ahead and install the semiconductors, take a brief look at the diagram depicted in Figure 15-6. The semiconductor pin-out diagrams will help you orient the semiconductors when building the induction receiver. The induction loop receiver also incorporates one transistor at Q1 which is connected to R1/R2 and R4 and C2. Transistors generally have three leads protruding from the bottom of the body. A Base lead,

Table 15-1

Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

a Collector lead and and an Emitter lead. Referring to the schematic you will note that the vertical line on the transistor symbol is the Base lead, while the Collector lead is connected to the junction of R1 and R2. The Emitter lead has the small arrow pointing downwards which connects to the junction of C2 and R4. Refer to the actual transistor pin-out diagram to see where each lead exits the transistor case. Generally small plastic TO-92 transistor case bodies will have all three leads in-a-row with the Emitter at one end, the Base lead in the center, and the Collector lead at the opposite end of the device.

The induction loop receiver has one integrated circuit at U1, this is an LM386 audio amplifier chip, which has eight leads in a dual in-line package. Prior to installing the IC, you should install a good quality IC socket. An IC socket is good insurance in the event of a possible circuit failure at some point in time. Since most people cannot successfully un-solder an integrated circuit without damaging the PC board, an IC socket makes good sense. Place the IC socket on the circuit board and solder it in place on the PC board. Locate the LM386 and take a look at the chip. You will note that the IC package will have either a cut-out, a notch or small indented circle at one end of the IC package. Pin 1 of the IC will be just to the left of the notch or cut-out. Orient the IC correctly into the socket, while referring to the schematic and layout diagrams. Place the IC into the socket.

The receiver loop coil at L1 can be either the 800 turn coil wound on a plastic form, an old "high" impedance relay, or one of the black plastic suction cup-telephone pickup devices. Solder the coil in place between the circuit ground and the capacitor at C1. You could also elect to use a second 1/8″ phone jack to allow the coil to be plugged into the circuit using a 6″ to 8″ length of coax cable.

The volume control used for the prototype was a chassis mounted miniature 10k ohm potentiometer. We used a mini 1/8″ PC style phone jack on the circuit board at the output of the audio amplifier at C6; this will allow you to plug in a headphone for listening.

Take a short well deserved rest and when we return we will check the circuit board for possible "short" circuits and "cold" solder joints. Pickup the circuit board with the foil side of the circuit board facing upwards toward you. Look the circuit board over carefully to see if there are any "cold" solder joints. The solder joints should look clean, smooth and shiny. If any of the solder joints look dull, or "blobby," then you

Table 15-2

Capacitance code information

This table provides the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Some capacitors actually have the numeric values stamped on them; however, many are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

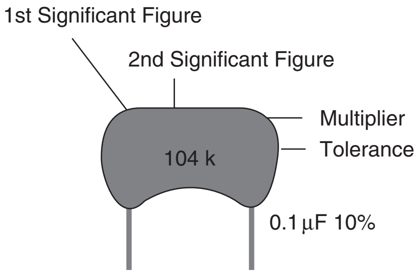

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

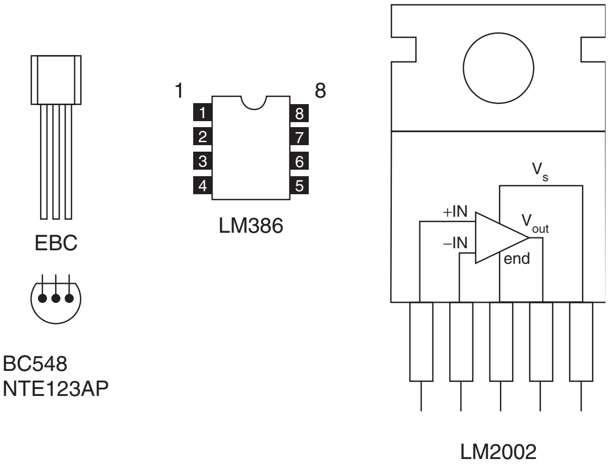

Figure 15-6 *Semiconductor pin-outs*

should remove the solder from the joint and re-solder the connection point over again until it looks good. Next we will inspect the PC board for any possible “short” circuits which may be the result of “stray” component leads which may have stuck to the foil side of the PC board. These “stray” components leads often “stick” to the board by solder residue. Solder blobs are often a cause of “short” circuits, so look the board over carefully for any solder blobs as well. Once the circuit board has been inspected, we can move on to installing the circuit board into a case or enclosure.

We mounted the induction loop receiver in a 3″ × 6″ by 1½″ metal chassis box. We placed one ⅛″ jack at the rear of the case for coil L1, then we placed the second ¼″ phone jack, power on-off switch and volume control on the front side of the chassis box. Locate a 9 volt battery clip and solder the black or minus (−) lead to the circuit ground. Next you will want to wire the red or positive (+) battery clip in series with the On-Off switch to the two power junction points on the PC board, one at R2 and the other at pin 6 of the LM386 IC. Once the circuit is inside the enclosure, you can begin testing the induction loop system. First you will want to make sure that the “transmitter” portion of the system is up and running. If you elected to use an existing stereo receiver or amplifier, make sure that the large coil is connected through a series resistor to the speaker terminals of the amplifier. Set the receiver or amplifier to MONO. If you elected to build your own audio amplifier to drive the large loop, again connect up the coil via an 8 ohm 10 watt series resistor. Once the “transmitter” is ready, with your music or audio source connected, we can move on to testing the induction loop receiver.

Switch the induction receiver to the “On” position. When you first turn on the unit you will probably hear a lot of buzzing from the wiring in the room. Rotate the receiver in a horizontal plane to find a “null” where the hum is minimal. If you can get a reasonable null, you should be able to hear the source material from the “transmitter” loop or at least some hum coming from the earphone. If your “transmitter” loop is “broadcasting” then you should hear the source music or speech on the headphone of your induction loop receiver. If all goes well, both your transmitter and receiver section will be working and you will be hearing the source material in your remote headphones.

In the event that the induction receiver does not work when first power-up, you will have to remove the battery and carefully examine the circuit board for any errors. The most common cause for failure are improper placement of resistors, electrolytic capacitors and diodes installed backwards, and semiconductors such as transistors installed incorrectly. Have a knowledgeable electronics friend provide a second pair-of-eyes to help you examine the circuit for errors. It is very easy to miss a problem since, as the builder, you will continue to see the same circuit, the same way over and over. Once the error has been found and corrected, you can re-connect the battery and test the circuit once again.

As mentioned earlier, you can use the system for hard of hearing people, for museum displays and demonstrations, for theater personnel, late night TV listening, etc. You could wind a transmitter loop around your easy chair and use the system to “broadcast” TV sound to your earphones for late night listening. This is a great Boy Scout project for helping old folks who live in nursing homes to better enjoy TV, radio and music.

You can also use the induction loop receiver by itself without the transmitter loop, to trace power wires behind a wall or ceiling by listening for a sharp increase in hum as the coil passes near the wire. Make sure that current is flowing in the wires to be traced by turning on a lamp or other appliance. Other wires can be traced if they are carrying alternating current in the audio range or a signal generator can be connected to produce the current. Connect the generator to the wire to be traced and connect the generator's "ground" to the house wiring ground. Also ground the far end of the wire you are tracing so that current flows in the wire. This ground connection can also just be a temporary wire laying on the floor running from the generator ground to the far end of the wire you wish to trace. Have fun using your new induction loop receiver!

Chapter 16

Lightning Storm Monitor

Parts list

Lightning Storm Monitor

R1 47k ohm ¼w, 5% resistor

R2,R4,R7 10k ohm ¼w, 5% resistor

R3 5.6k ohm ¼w, 5% resistor

R5 470k ohm ¼w, 5% resistor

R6,R10,R11 22k ohm ¼w, 5% resistor

R8 390 ohm ¼w, 5% resistor

R9 4000megohm or greater (see text)

R12 220k ohm ¼w, 5% resistor

R13 1 megohm ¼w, 5% resistor

R14 100k ohm ¼w, 5% resistor

R15 10 ohm

VR1 500k ohm potentiometer (chassis mount)

VR2,VR9 1 megohm potentiometer (trimmer)

VR3 3k ohm potentiometer (trimmer)

VR4,VR5,VR6 10k ohm potentiometer (chassis mount)

VR7 25k ohm potentiometer(trimmer)

VR8 10 megohm potentiometer (chassis mount)

VR10 10 megohm potentiometer (trimmer)

C1,C9 2.2 pF, 50 volt mica capacitor

C2,C10 180 pF, 50 volt mica capacitor

C3,C4,C6,C8,C11,C12,C22, C23 10µF, 50 volt electrolytic capacitor

C5 .01µF, 50 volt disk capacitor

C7,C21 100 pF, 50 volt mica capacitor

C13 .22µF, 50 volt disk capacitor

C14,C15,C16,C19,C24 .1µF, 50 volt disk capacitor

C17 .47µF, 50 volt disk capacitor

C18 220µF, 50 volt electrolytic capacitor

C20 470µF, 50 volt electrolytic capacitor

D1,D2 1N914 silicon diode

L1 coil optional - 3000 turns on 2′ × 6″ PVC form

Q1 J176 P-Channel JFET

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

U1 LF355 (JFET)

U2,U3,U5 LM741N op-amp

U4 LM380 audio amplifier

S1 SPDT switch (Hi/Lo)

S2 SPST switch (load) ceramic high voltage type

M1 500 ua meter (RM1200 ohm)

M2 250-0-250 FS meter (RM450 ohm)

J1 ⅛″ switches jack

J2 RCA jack - signal output jack for data collection

J3 SO-239 - Antenna-2 input jack

J4 Antenna-1 - RCA jack for whip or outdoor Ant.

CT Capacitive Hat (see text)

Misc IC and transistor sockets, wire, PC board.

Multi-voltage Power Supply

R1 5 k potentiometer - (trimmer)

R2 240 ohm, ¼watt, 5% resistor

C1,C4,C7 4700µF, 50 volt electrolytic capacitor

C2,C5,C8 .1 µF, 50 volt ceramic disk capacitor

C6,C6,C9 10 µF, 50 volt electrolytic capacitor

BR1 4-diode bridge rectifier – 3 amp, 100 volt

U1 LM317 - (+) voltage adjustable regulator

U2 LM7805 - (+) voltage regulator

U3 LM7905 - (–) voltage regulator

T1 24-VCT transformer 2-amp-center tapped

S1 SPST toggle switch - On-Off

F1 2-amp fast blow fuse

Misc PC board, wire, connectors, etc.

Intense electrical storms created by unstable weather conditions frequently occur all around the world. At any given point in time there is likely to be an electrical storm occurring somewhere in the world. Most people would really appreciate having an advance warning of an approaching electrical storm, especially hikers, boaters or backyard party goers. Why not construct your own lightning monitor to obtain an advance warning of an approaching storm before it arrives, thus giving you ample time to seek shelter from the elements. The lightning detector will give you ample warning of an approaching electrical storm before you see and hear the lightning. The sensitive lightning detector project will alert you to an oncoming storm over 50 miles away, giving you time to take cover or to go inside to safety.

During a lightning strike, is the Earth considered positive or negative, is a question often asked. In an electrical storm, the storm cloud is charged like a giant capacitor. The upper portion of the cloud is positive and the lower portion is negative. Like all capacitors, an electrical field gradient exists between the upper positive and lower negative regions. The strength or intensity of the electric field is directly related to the amount of charge build-up in the cloud. This cloud charging is created by a colliding water droplets.

As the collisions continue and the charges at the top and bottom of the cloud increase, the electric field becomes more intense—so intense, in fact, that the electrons at the Earth’s surface are repelled deeper into the Earth by the strong negative charge at the lower portion of the cloud. This repulsion of electron causes the Earth’s surface to acquire a strong, positive charge.

The strong electric field also causes the air around the cloud to break down and become ionized (a plasma).

A point is reached (usually when the gradient exceeds tens of thousands of volts per inch) where the ionized air begins to act like a conductor. At this point, the ground sends out feelers to the cloud, searching for a path of least resistance. Once that path is established, the cloud-to-earth capacitor discharges in a bright flash of lightning.

Because there is an enormous amount of current in a lightning strike, there's also an enormous amount of heat (in fact, a bolt of lightning is hotter than the surface of the sun). The air around the strike becomes super heated, hot enough in fact that the air immediately close to the strike actually explodes. The explosion creates a sound wave that we call thunder. A typical high energy lightning strike is shown in Figure 16-1.

Cloud-to-ground strikes are not the only form of lightning though. There are also ground-to-cloud (usually originating from a tail structure) and cloud-to-cloud strikes. These strikes are further defined into normal lightning, sheet lightning, heat lightning, ball lightning, red sprite, blue jet, and others that are lesser defined. For more information on lightning, see: http://science.howstuffworks.com/lightning.htm.

Satellites are more often used to follow lightning strikes around the world but haven't advanced to the point where they can accurately map local areas. There are two major types of sensors commonly used—magnetic direction finders and VHF Interferometery.

The National Lightning Detection (NLDN), which is operated by Global Atmospherics, Inc., (GAI) in Tucson, AZ, is a network of more than 130 magnetic direction finders that covers the entire USA—more than twice the coverage of existing weather radar networks.

Figure 16-1 *Lightning*

Each direction finder determines the location of a lightning discharge using triangulation and is capable of detecting cloud-to-ground lightning flashes at distances of up to 250 miles and more. Processed information is transmitted to the Network Control Center where it's displayed in the form of a grid map showing lightning across the USA: (www.lightningstorm.com/tux/jsp/gpg/lex1/mapdisplay_free.jsp).

Recently, NASA has improved the resolution of the system by adding acoustical measurements to the mix. Although the flash and resulting thunder occur at essentially the same time, light travels at 186,000 miles per second, whereas sound travels at the relative snail pace of one-fifth of a mile in the same time. Thus, the flash—if not obscured by clouds—is seen before the thunder is heard. By counting the seconds between the flash and the thunder and dividing by 5, an estimate of the distance to the strike (in miles) can be made.

In the NASA lightning sensor, a low frequency receiver detects the lightning strike. The leading edge of the electric-field pulse is used to start a timer and the leading edge of the thunder pulse is used to stop the timer. A microcontroller in each receiver transmits the time measured to a processing station, where the times are converted to distances that are used to compute the location of the lightning strike to within l2 inches. However, the NASA sensors have to be located within a 30 mile radius of the strike to be accurate.

If you are interested in monitoring static levels around you in stormy weather, and the ability to have some advance warning of an impending storm, then this circuit will be of great interest to you especially if you are involved in the radio hobby and have antennas high in the air. The Lightning Storm Monitor has proven to be very reliable and useful for advanced storm warning activities.

The Lightning Monitor circuit is shown in the schematic in Figure 16-2. The Lightning Monitor is actually two receivers in one, a spheric monitor and a static electricity monitor. The static monitor portion of the circuit begins at the outdoor Antenna-2, a capacitive hat type antenna which is fed with coax from the antenna to the receiver. The gate of the P JFET is tied

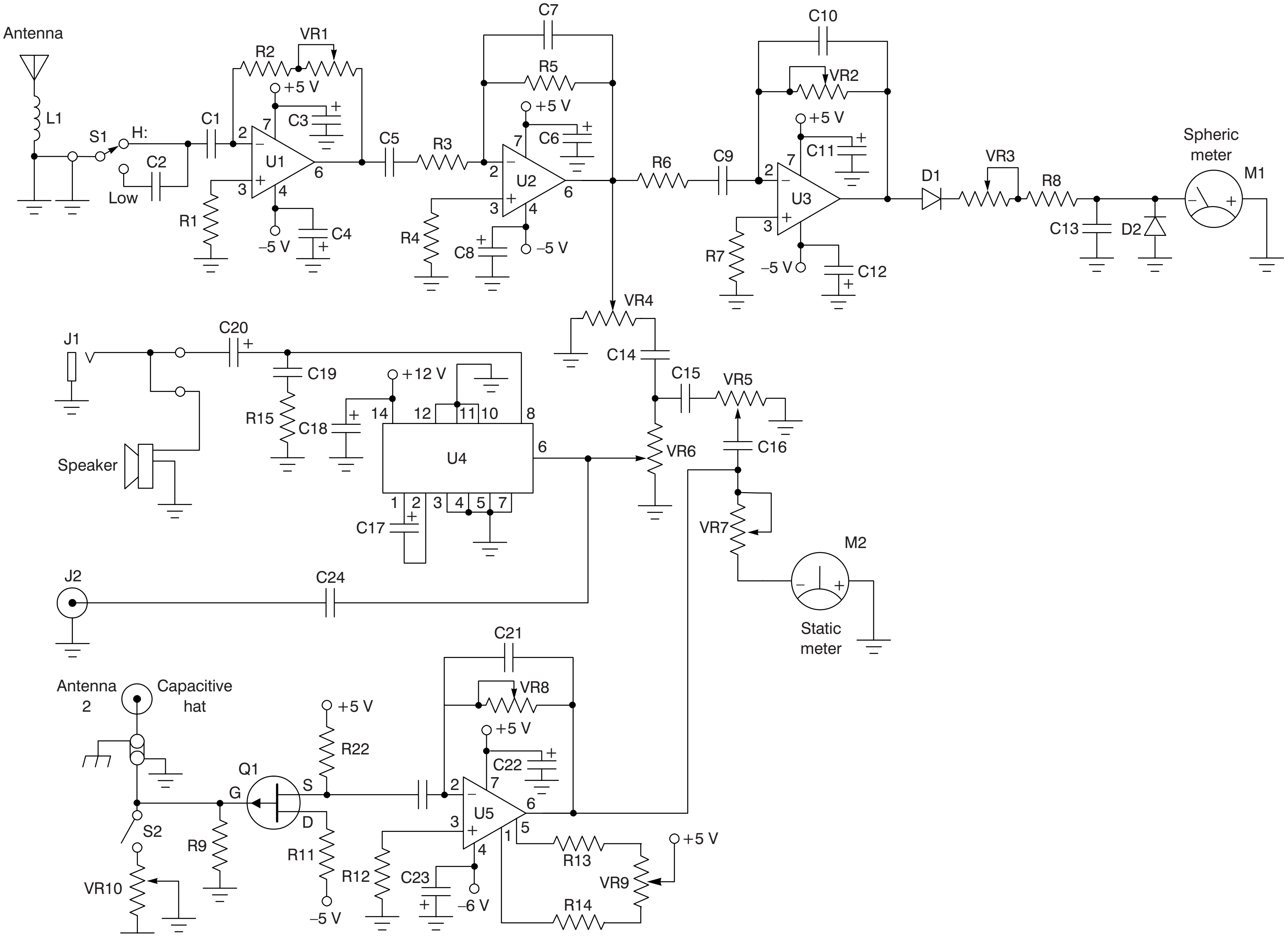
Antenna
L1
S1
H:
Low
C1
C2
R1
R2
VR1
+5 V
C3
U1
–5 V
C4
C5
R3
C7
R5
U2
C6
R4
C8
R6
C9
C10
VR2
U3
C11
R7
C12
D1
VR3
R8
C13
D2
Spheric meter
M1
VR4
C14
C15
VR5
C16
VR6
VR7
M2
Static meter
J1
C20
C19
R15
C18
+12 V
U4
C17
Speaker
J2
C24
C21
VR8
Antenna 2
Capacitive hat
Q1
G
S
D
R22
R9
S2
VR10
R11
–5 V
U5
C22
R12
C23
–6 V
R13
VR9
R14

Figure 16-2 *Lightning monitor. Courtesy of Russell Clit*

to ground since a dual polarity supply is used, through a very high resistance. In the original prototype, I experimented with an 11,000 megohm resistor. These high value resistors are hard to find; you may notice that it is a moot point if there is too much leakage in your circuit. In that light, it is best to standoff the gate and use a high quality new coax to your outside antenna, usually a capacitive hat type. You must also come up with a mount design which keeps rain from the point on the mount where current could flow to ground, or you will notice that your meter becomes very insensitive as it starts raining.

The prototype used a 22″ shaft from an amateur radio antenna in a mobile antenna mount, with two nuts on top to hold the capacitive hat and a plastic umbrella to protect the mount from moisture. Likewise the coax connection must be well sealed from moisture, an SO-239 type connectors at the antenna and on the chassis for the meter inside the house. As far as the resistance load goes, if you cannot locate a high value for R9, you can fabricate a high value type if need be, for the strong field load. A 10 meg potentiometer which could be switched out of circuit was used in the prototype. For the switch at S2, you could use a ceramic high voltage type to minimize leakage. The FET type is not critical; in the prototype a J176 came from All Electronics, also the meter and the 10 meg potentiometer came from them as well.

Through experimentation it was found that controlling the offset of an op-amp gave a very sensitive way to control meter balance, far superior to trying to shift the meter itself. If you cannot locate a zero center meter you might ground the end, or put it on a trimmer tap where the ends of the pot are on the plus and minus of the supply, say a 5 or 10 K pot. If you can find it, a 250-0-250 µA meter is suggested. In the maximum gain mode you can notice field gradient changes in clear weather from morning to night, as well as detect storms more than a state away. One problem you will notice with this circuit, especially at high gain settings, is that you have to re-zero the meter often as the polarity changes overhead in stormy weather.

The higher the gain of the op-amp or the higher the input impedance of the FET gate, the more this becomes a problem, which is why a provision was made to lower the impedance of the gate circuit and reduce the op-amp gain in strong fields. Note the provision for audio to be taken from the op-amp, and mixed with separate levels for the static and RF signals, as well as a master volume.

In the Spheric, or RF section of the Lightning Monitor, the whip antenna at Antenna-1 was fed to a coil consisting of a 4′ foot long 6″ inch diameter PVC pipe, the coil consists of about 3000 turns of insulated wire on it. You will find that just a straight whip antenna works fine, so it was found that the large coil is not mandatory. The original thought for using the coil at the antenna was to have a lot of signal at low frequencies with a high Q inductor, to give more signal while trying to reduce the overall gain of the circuit, to minimize 60 cycle hum. On that note, long whips or especially long wires are not very desirable here, so stay with a 5′ to 6′ vertical whip antenna.

The input signal is amplified with a JFET input op-amp at U1, with selectable input range through a small value capacitor to allow high gain to be achieved with minimum 60 cycle hum. A LM741 op-amp gives gain to provide audio signal and then another LM741 is used to drive a 500 µA single ended meter for a reading of RF levels. You may find it easier to make the series resistance control for the meter to be a front mounted control as well as the gain control for the meter driver, i.e. the 741 op-amp, as it give much more flexibility under differing weather conditions. This meter is very helpful in determining the number of lightning strokes per unit time in severe weather.

The audio output from the Spheric monitor section is derived form an LM380 IC circuit; you will notice interaction between controls if you build it as shown. A buffer and mixer circuit might be helpful here, but in the original design we decided to keep the parts count low. Also, an equalizer circuit for the signal out is a very good addition, so you can tailor the output response and reduce unwanted signals such as line power noise. A signal output is provided so the output can be fed to a digital multi-meter with PC interface, if desired to log readings of the Lightning Monitor over time.

Building the lightning monitor

Prior to constructing the Lightning Monitor, you will want to locate a well lit and well ventilated work area. A large clean work bench or work table would be suitable for building the project. Grab a 27 to 33 watt pencil tipped soldering iron, a roll of 22 gauge 60/40 tin/lead rosin core solder. You should also have a small jar of "Tip Tinner," a soldering iron tip dresser which is available from your local Radio Shack store. You'll probably want to gather a few tools, such as a pair of end-cutters, a small pair of needle-nose pliers, a pair of tweezers, a magnifying glass, a small Phillips screwdriver as well as a small flat-blade screwdriver. Place your schematic diagram along with your circuit layout diagram and various charts and tables in front of you on your work table. Place your project components and circuit board in front of you and we can begin constructing the Lightning Monitor. A printed circuit is highly recommended for building this project rather than trying to use perf-board construction to permit more reliable operation from your circuit. Warm up your soldering iron and we will begin construction!

Let's begin building the monitor by first identifying the project resistors. Locate the resistors from the parts pile. Refer to the chart in Table 16-1, which illustrates the resistor color codes and their values. Resistor R1, for example, is a 47k ohm ¼ watt 5% resistor. Try and locate it from the parts pile. The first color band at one edge of the resistor body should be yellow, which is the number (4). The second color band should be violet or number (7) on the chart. The third color band is the multiplier and should be orange in color and represents a multiplier value of 1000, so 47 times 1000 is 47,000 ohms or 47k ohms. The fourth color band on resistors represents the resistor's tolerance. A silver band denotes a 10% tolerance, a gold band represents a 5% tolerance resistor and no color band denotes a 20% tolerance resistor. Once you have found R1, you can try and locate resistor R2 and so on. You can place the resistors on the circuit board in groups of four or so in their respective locations on the circuit board. Once you have the first four resistors placed on the PC board, you can go ahead and solder them in place on the board. After the resistors have been soldered onto the circuit board, you can cut the excess component leads flush to the edge of the circuit board using your end-cutters. Now, you can locate the next batch of four resistors and place them in their respective locations on the circuit board, then go ahead and solder them in place. Remember to trim the excess component leads. Finish up installing the resistors and then we will move on to installing the two silicon diodes.

There are two silicon diodes in this project, and they are located in the meter range and display circuit. Diodes have two leads, a cathode lead and an anode lead. Take a look at the resistors and you will notice that there will be either a black or white color band at one edge of the diode's body. The color band denotes the cathode lead of the diode. It is important that you install

Table 16-1
Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000(M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

the diode correctly with respect to this polarity. One you have installed the two diodes in their proper place, you can solder them in place on the circuit board. Don't forget to cut the extra component lead lengths.

Next, we will try and locate the capacitors for the Lightning Monitor project. This project uses two types of capacitors, small ceramic disk types and the larger electrolytic types which have polarity. The small capacitors come in many different styles and shapes and sometimes they are so small that there is not enough room to print their actual value on the component, so a three-digit code is used. Refer to the chart shown in Table 16-2 when installing the small capacitors. For example, a .1 μF capacitor would have a code value of (104). The Lightning Monitor also utilizes larger value electrolytic capacitors, these capacitors are usually large and have polarity marking on them, such as a black or white color band or a plus (+) or minus (−) marking on them. These polarity markings must be observed when installing the capacitors on the circuit board. Failure to observe the correct polarity could result in either a damaged component or a damaged circuit when first power-up.

Look through your component parts pile and let's try and identify some capacitors. Let's begin with some of the small types such as C1 and C2. Capacitor C1 may be marked 2.2 or 2n2, while Capacitor C2 will most likely be labeled 181. Try and locate the lower value capacitor from the parts stock and place about four components at a time on the PC board in their respective locations. After placing four or so of the first lot of capacitors, you can go ahead and solder them in place. Remember to trim the excess component leads from the board. Take the next grouping of four capacitors and install them on the PC board while referring to the schematic and parts layout diagrams. Go ahead and solder the next group of four capacitors and trim the excess lead lengths. Install the remaining smaller capacitors and then we will move on to installing the larger value electrolytic capacitors. As mentioned, electrolytic capacitors will have polarity markings on them and they must be installed correctly in order for the circuit to work properly. Take your time when identifying the electrolytic capacitors. Observe their polarity marking and then make sure you orient them correctly and in their proper location on the PC board. Refer to the schematic and parts layout diagram. Install a few electrolytic capacitors at one time, and them solder them in place, follow up by removing the excess lead lengths. Install the remaining electrolytic capacitors and solder them in place. Use your end-cutters to remove the extra lead lengths by cutting them flush to the edge of the circuit board.

The Lightning Monitor uses a single FET transistor at Q1, which is connected to antenna-2. The P channel J-FET at Q1 is a J176 FET, which must be handled with care using an anti-static wrist band to avoid damage. This component is very sensitive to static electricity and can be easily damaged in handling. The three leads of the transistor Q1 are all in a single row with the Drain lead at one end and the Gate lead in the center and the Source lead at the opposite end of the plastic case of the transistor. Refer to Figure 16-3, which illustrates the semiconductor pin-outs.

Before we install the integrated circuits in this project, we would highly recommend using IC sockets. Integrated circuit sockets are good insurance in the event of a possible circuit failure at some possible distant point in time. It is much easier to simply unplug a defective IC, rather than trying to un-solder a 14-pin IC from the circuit board. Integrated circuit sockets are inexpensive and easy to locate. There a number of integrated circuits in this project, from 8-pin to 14-pin dual in-line types. In order to install the integrated circuits correctly, you must observe the correct orientation. Failure to install the integrated circuits correctly can cause serious damage to the IC and to the circuit itself, so you must get it right the first time! Each plastic IC package will have either a cut-out, a notch or small indented circle at one end of the plastic package. Looking just to the left of the notch or cut-out you will find pin 1 of the IC. You must align pin 1 of the IC to pin 1 of the IC socket when installing the chip on the PC board. Take a look at U1, this is an 8-pin dual in-line IC. Since we will not be using pin 1 on this IC, you will have to make sure that pin 2 or the minus (−) input is connected or aligns with C1 and R2, while pin 3 of U1 is connected to resistor R1 and so on. Now, go ahead and insert the integrated circuit into their respective sockets.

There are a number of potentiometers in this project, some of which are chassis mounted controls, while other potentiometers will be circuit board mounted types. Refer to the parts list below and Table 16-3. You can locate the circuit board trimmer type potentiometer and install them at their respective locations on the

Table 16-2

Capacitor code information

This table is designed to provide the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Most capacitors actually have the numeric values stamped on them; however, some are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

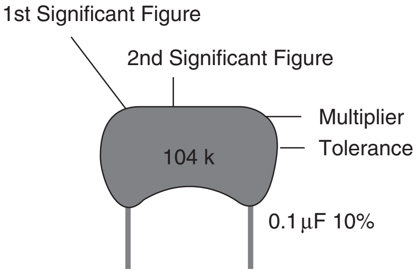

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

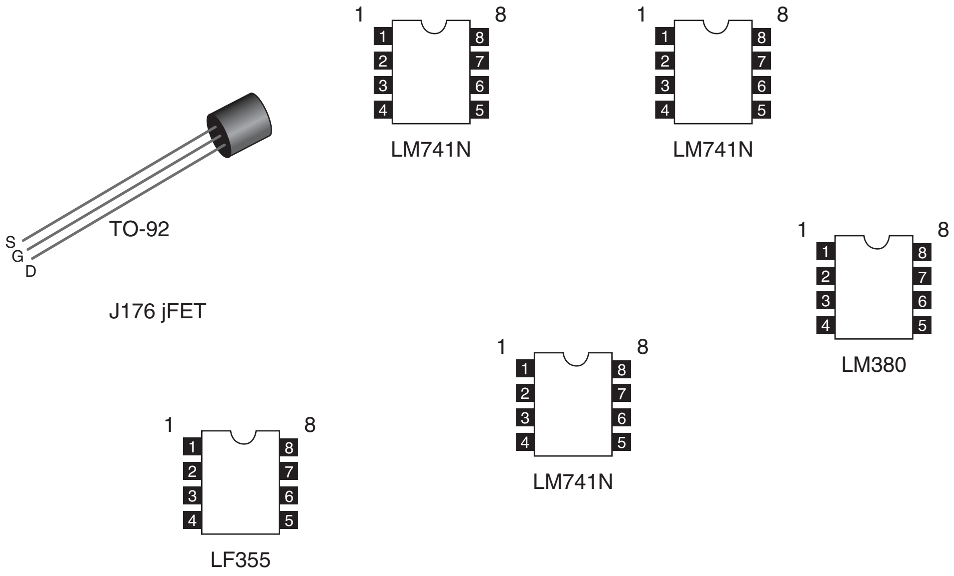

Figure 16-3 *Semiconductor pin-outs*

circuit board. Solder them in place and then trim the excess leads if necessary.

Let's take a well deserved break and when we return we will inspect the circuit board for possible "short" circuits and "cold" solder joints. Pick up the Lightning Monitor circuit board with the foil side of the board facing upwards towards you. First we will inspect the PC board for possible "cold" solder joints. Take a look at the solder joints, they should all look clean, shiny and bright. If you see any solder joints that appear dull, dark or "blobby," then you should remove the solder from the joint and re-solder the joint all over again until it looks good. Once you are confident that all looks well, we can move on to inspecting the PC board for possible "short" circuits. There are two main causes for "short" circuits, small round solder blobs laying across a PC "trace" can cause a circuit to "short-out." The second cause of "short" circuits is a "stray" component lead. Often when cutting excess component leads, they will "stick" to the circuit board from rosin core solder residue. Make sure that there are no "stray" leads bridging between the circuit board traces. Once you are confident that there are no "shorts," we can move on to finishing the project.

Now, let's move on to mounting the larger components and hardware on the chassis box. The chassis mounted

Table 16-3
Potentiometer functions

Potentiometer	*Function*	*Type*
VR1	Spheric Gain Control	500k chassis mount
VR2	Meter Gain Control	1 megohm trimmer
VR3	Meter Range Control	3k trimmer
VR4	Spheric Volume Control	10k chassis mount
VR5	Static Volume Control	10k chassis mount
VR6	Main Volume Control	10k chassis mount
VR7	Static Range Control	25k trimmer
VR8	Static Gain Control	10M chassis mount
VR9	Static Meter Balance Control	1 megohm trimmer
VR10	Load Control	10 megohm trimmer

potentiometers were all mounted on the front of the metal chassis used to house the project, and wired to the circuit board with 20 ga. insulated wire. The two toggle switches, S1 and S2, were also mounted on the front panel of the chassis and were wired to the circuit board using 22 ga. insulated wire.

The Lightning Monitor utilizes two analog meters, one for each monitoring circuit. The meter for the Spheric circuit at M1, was a 500 µA meter, while the Static meter circuit uses a 250-0-250 µA center reading meter at M2. Both of these meters were found surplus and were mounted on the front panel of the metal chassis.

The speaker was mounted on the rear panel of the chassis and wired to the circuit board using two 6″ lengths or #22 ga. insulated wire. A ⅛″ switched mini audio panel jack for headphone use was mounted on the front panel. An RCA jack was provided for antenna-1; you can use a whip antenna or an outdoor whip antenna. An SO-239 antenna connector was mounted at the rear panel to allow the Capacitive Hat antenna-2 to connect to the circuit board, via a small piece of coax cable. The Capacitive Hat antenna is shown in Figure 16-4, it can be built with an antenna bracket mounted on a pole outdoors. An SO-239 jack was mounted on the small antenna bracket. A threaded shaft was silver soldered to the center terminal of the connector with a small rain guard on it. At the top of the threaded shaft a metal coffee can lid is held in place with two nuts. The antenna can be placed outdoor on a pole near the house with a length of coax cable leading from the antenna to the antenna jack on the Lightning Monitor.

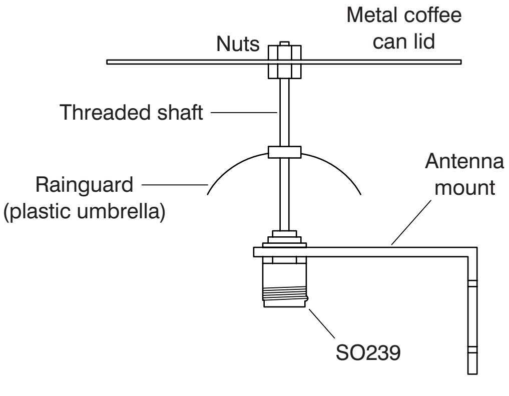

Figure 16-4 *Antenna-2 – Capacitive hat antenna*

If you wish to experiment, you can try an indoor 5′ whip without the coil and see how the circuit performs. You could also experiment with an outdoor whip antenna without the coil and you could build the coil and place it in series with an outdoor whip antenna if desired.

The Lightning Monitor needs a power supply which can provide a 12 volt source for the audio amplifier circuit as well as a plus and minus power supply for the op-amps. The diagram in Figure 16-5 illustrates a 12 volt and dual plus/minus 5 volt power supply which can be used to power the Lightning Monitor. The prototype Lightning Monitor incorporated the power supply in the same chassis as the actual monitor. You will need to locate a 12 volt, 250 mA center-tapped transformer ahead of the rectifier circuit. The rectifier circuit is fed directly to the regulator integrated circuits as shown. You will need to connect the dual voltage power supply to the Lightning Monitor. The plus voltage is connected to pins 7 on both the LM741 and the LF355 op-amp as well as to the J-FET at Q1. The negative voltage supply is connected to pin 4 of both the LM741 and the LF355 op-amp, as well as to the JFET at Q1. The 12 volt power supply is used to power the LM380 audio amplifier.

Once the power supply and Lightning Monitor have been completed, you connect the whip antenna-1 and the Capacitive Hat antenna-2 and prepare for testing the Lightning Monitor. To best test the circuit you will need to wait for a thunderstorm. However, you can do some testing of the circuit. First, apply power to the Lightning Monitor circuit, turn the Off-On switch to the "On" position. Make sure both antennas are connected. Turn up the main volume and the Static Volume controls past the ½ position and adjust VR7 to past ½. You will need to generate some static electricity by rubbing a lucite rod with some animal fur to create a static charge. Bring the lucite rod near the Capacitive Hat antenna and you should see the Static Meter at M2 swing into action. This will give you a good indication that the Static portion of the meter is working. You may be able to test the Spheric Meter portion of the circuit if you have an RF broadband noise generator.

If the Lightning Monitor doesn't seem to work at all or appears DEAD, then you will want to turn "Off" the circuit and remove the power and inspect the circuit board once again. While inspecting the circuit board

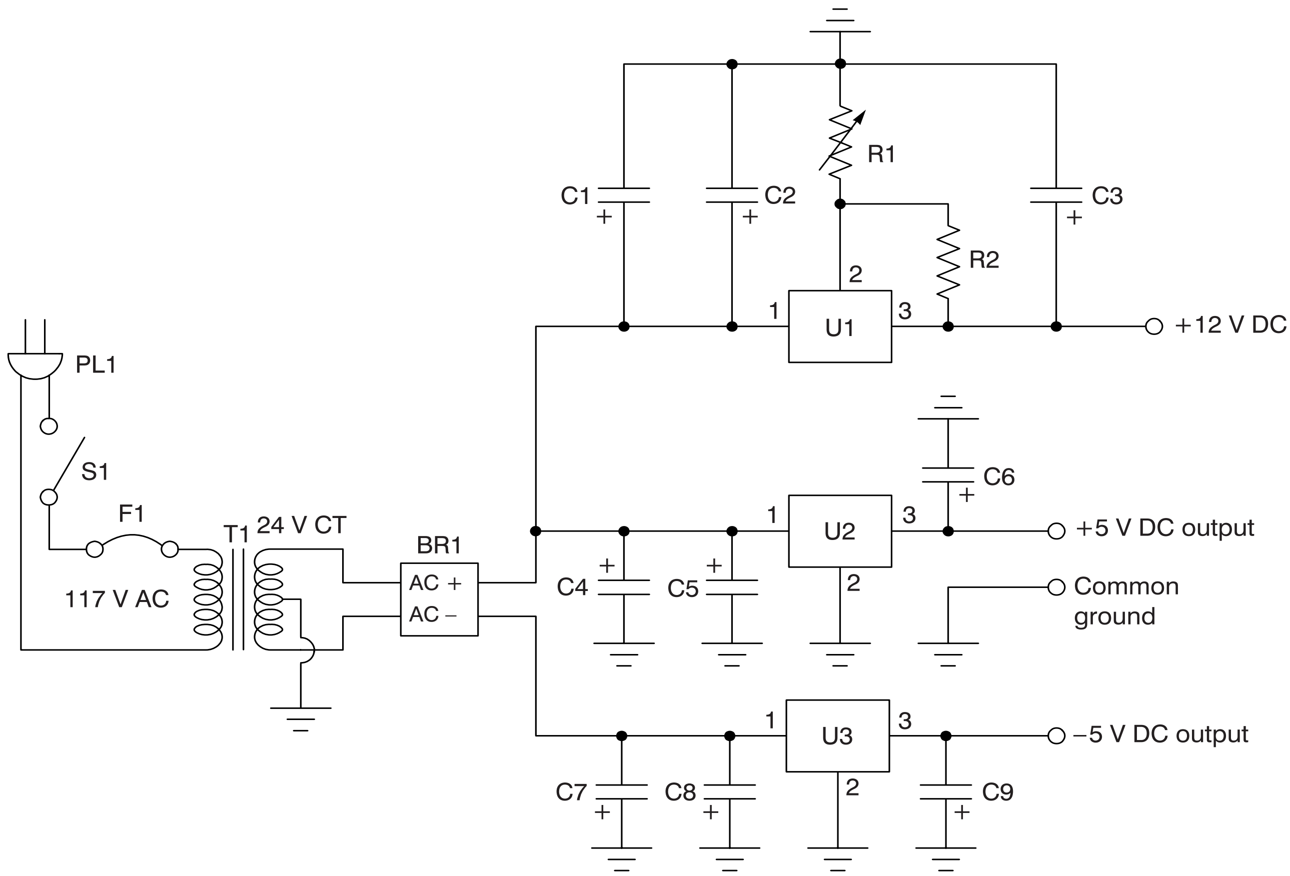
PL1
S1
F1
117 V AC
T1
24 V CT
BR1
AC +
AC −
C1
C2
R1
R2
C3
U1
1
2
3
+12 V DC
C6
U2
+5 V DC output
C4
C5
Common ground
U3
−5 V DC output
C7
C8
C9
+

Figure 16-5 *Multi-voltage power supply*

you will be looking to make sure that the electrolytic capacitors have been installed correctly, to be sure that the diodes and the JFET have been installed correctly as well. Have a knowledgeable electronic hobbyist inspect the circuit, since it is easy to overlook something and not catch it the second time. Make sure that your power supplies are connected to the correct points and that the op-amps have been installed correctly. Once you are confident of your inspection and hopefully you found your error, you can re-apply power and re-test the circuit again. Good luck and be safe when experimenting with outdoor antennas, keeping them away from power lines.

Chapter 17

Ambient Power Receiver

Parts list

Original Ambient Power Module I

C2,C3 .047µF, 50 volt capacitor

C1,C4 40µF, 50 volt electrolytic capacitor

D1,D2 1N34 diodes

D3,D4 1N34 diodes

ANT 50+ foot broadband monopole antenna

GND Earth ground stake

Misc PC board, wire, hardware, enclosure

Ambient Power Module II

D1,D2,D3,D4,D5,D6 1N60 germanium diodes

D7,D8,D9,D10,D11,D12 1N60 germanium diodes

C1,C2,C3,C4,C5,C6 470µF, 50 volt electrolytic capacitors

C7,C8,C9,C10,C11,C12 470µF, 50 volt electrolytic capacitors

L1 480 turns #22 wire on 3″ plastic tube

ANT 50+ foot broadband monopole antenna

GND earth ground stake

Misc printed circuit board, wire, hardware, case

In 1979, Joe Tate began experimenting with methods of turning radio energy in the air into usable electric power. He first developed a clock which drew its power from an antenna that was just a long piece of wire stretched out horizontally about 20 feet above the ground. The power supply for the clock worked something like an old-style crystal radio, except that it did not have a tuning circuit. Because of this, the “Crystal Clock” was able to absorb a wide spectrum of radio noise from the antenna and yield electric power. The power supply was able to deliver much more current than was developed in a crystal radio, although its output was just still just a few millivolts.

Joe began recording the power supply’s output over a long period of time to determine its dependability. After all, the device relied completely on whatever stray signals happened to be in the air. Using an Atari computer, the output of the clock’s power supply was measured continuously and recorded on floppy disk. This was done by feeding the unregulated voltage output directly into the computer’s joystick port.

Joe began calling this power supply the “Ambient Power Module” (APM) because it extracted power from ambient background radio noise. This small circuit, when connected to antenna and ground, used the potential difference between air and ground to generate a small direct current continuously. Joe’s original claims stated that the APM could produce up to 36 volt/9 watts. Joe applied for a US Patent and was later granted patent #4,628,299. As he studied the recorded data, mild fluctuations were noted in a daily cycle. The patterns were consistent over long periods of time, though they differed in different locations. Aside from that, the APM looked like a very dependable source of power. In the spring of 1984, he discovered an interesting correlation between voltage fluctuations from the APM and a local earthquake event; this point will be discussed later.

The APM operates, as a broadband receiver of energy in the lower portion of the radio spectrum, receiving

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

Ambient power module

Figure 17-1 *Ambient power module*

most of its power from below 1 MHz. The basic circuit may be combined with a variety of voltage regulation schemes.

The original Ambient Power Module is shown in Figure 17-1 and the schematic diagram circuit in Figure 17-2 is used to convert the ambient RF energy to a direct voltage which can be used and handled by data processing equipment. This design makes use of doubler, splitter and rectifier circuit configurations. Designated as an ambient power module (APM), the circuit is connected to an antenna, preferably a 50 foot or greater broadband monopole antenna. The antenna is connected to the APM circuit via two capacitors C2 and C3, each being in series with the antenna input line for coupling and each having a value of 0.047 microfarad. Taking the left or negative side of the circuit first, it comprises two rectifiers (diodes) D3 and D4 (1N34 type) and a filter capacitor C4 (40 microfarads). Rectifier D2 is connected in parallel to the signal path and rectifier D1 is connected in series, in the well-known voltage multiplier arrangement. Capacitor C1 is connected in parallel across the output of the APM to smooth the rectified output. The right or positive side of the circuit is similar, except for the polarity of the diodes.

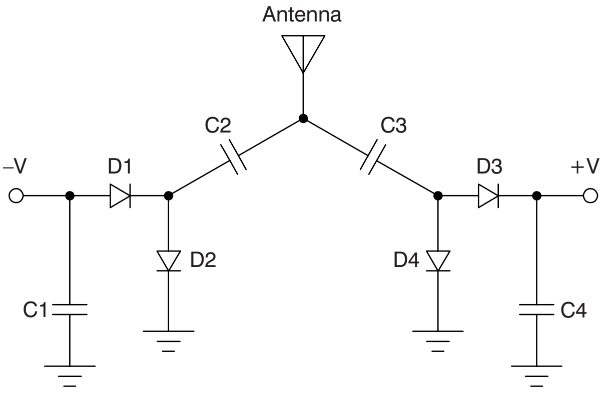

Figure 17-2 *Ambient power module 1*

In operation, an RF voltage is developed across the antenna; this voltage is voltage multiplied by the two rectifiers on each side of the circuit. The resultant voltage output is smoothed or filtered by capacitors C2 and C4. The circuit is very symmetrical with the plus side mirroring the minus side of the circuit. The output of the APM is ideal for low voltage, high impedance devices, like digital clocks, calculators, radios and smoke alarms, which are the most likely applications for this power source.

One possible application for the APM is to charge small NiCad batteries which provide effective voltage regulation as well as convenient electrical storage. Charging lead acid batteries is not practical because their internal leakage is too high for the APM to keep up with. Similarly, this system will not provide enough power for incandescent lights except in areas of very high radio noise. The APM could also be used to power small electronic devices with CMOS circuitry, like clocks and calculators. Smoke alarms and low voltage LEDs, and small emergency power/lighting circuits can also be powered by the APM. Use your imagination to devise an application for the APM.

Building the APM circuit

The builder has a choice of wiring techniques which may be used to construct the module. It may be hand wired onto a terminal strip, laid out on a bread board, experiment board, or printed circuit. We chose to build the original circuit on a small printed circuit board. Locate a clean well lit work table to build the APM. Find a small 27 to 33 watt pencil-tipped soldering iron, as well as some “Tip Tinner.” “Tip Tinner” is a soldering iron tip cleaner/dressing compound which helps to clean the soldering tip after use and can be obtained at your local Radio Shack store. Try and locate a pair of end-cutters, a pair of small needle-nose pliers and a set of screwdrivers. Grab your schematic diagram and look closely at it, before we begin construction.

The APM power receiver consists mainly of diodes and capacitors and you will recall that both of these type of components have polarity which must be observed for the circuit to function correctly. The diodes used in the APM circuit are germanium diodes and NOT silicon diodes which have a voltage drop. Diodes will have either a white or black band at one side of the diode's body. This colored band denotes the cathode end of the diode. When installing the diodes on the PC board make sure you observe the correct placement of all the diodes.

The APM circuit utilizes various capacitors, both electrolytic and non-electrolytic types. Electrolytic capacitors are often larger size capacitors with larger values. Electrolytic capacitors have polarity and may have either a black or white color bands with plus or minus marking at one end of the capacitor body. When installing the electrolytic capacitors make sure that you orient the capacitors correctly with respect to the circuit diagram. The chart in Table 17-1 may help you to identify the small value capacitors. In many instances, small value/size capacitors often use a three-digit code to represent a capacitor value.

Once all the components have been placed on the PC board, you can take a short break and when we return we will inspect the circuit board for possible "short" circuits and "cold" solder joints. Pick up the circuit board with the foil side facing upwards toward you. Look over the board carefully and inspect the solder connections. The solder joints should all look clean, smooth and shiny. If any of the solder joints look dull, dirty or "blobby" then remove the solder from the joint and re-solder the connection over again so that it looks good.

Next we will inspect the circuit board for any "short" circuit which could result from solder blobs or remaining component leads left from the building process. A component lead resting across a PC trace can result in a "short" circuit which could cause the circuit to fail.

Once you are satisfied that your circuit looks good, you can move on to testing the APM receiver. First you will need to attach a good antenna ground connection to the APM receiver module.

Antenna requirements

The antenna needs to be of sufficient size to supply the APM with enough RF current to cause conduction in the germanium diodes and charge the ground coupling capacitors. It has been found that a long horizontal antenna wire works best. A fifty foot monopole, or long-wire antenna will work well, but a small antenna should work, and of course you can experiment with your own antenna ideas. The antenna will work best when elevated to 20–30 feet. Lower elevations will work, but a longer wire may be necessary for best results.

In most locations, possible supporting structures already exist. The wire may be stretched between the top of a building and some nearby tree or telephone pole, see Figure 17-3. If live wires are present on the building or pole, care should be taken to keep your antenna and body well clear of these hazards. To mount the wire, standard commercial insulators may be used as well as homemade devices. Plastic pipe makes an excellent antenna insulator. Synthetic rope also works very well, and has the advantage of being secured simply by tying a knot. It is convenient to mount a pulley at some elevated point so the antenna wire may be pulled up to it using the rope which doubles as an insulator.

Grounding the APM

The APM requires a good ground circuit in order to operate efficiently. Usually a good ground can be established by connecting a wire to the water or gas pipes of a building. Solder or screw the wire to the APM ground terminal. In buildings with plastic pipes or joints, some other hookup must be used. A metal rod or pipe may be driven into the ground in a shady location where the earth usually is damper. Special copper coated steel rods are made for grounds, which have the advantage of good bonding to copper wire. A ground of this type is usually found within the electrical system of most buildings. Conduit is a convenient ground provided that the conduit is properly grounded. This may be checked with an ohmmeter by testing continuity between the conduit and system ground (ground rod). Just as with the antenna, keep the ground

Table 17-1
Capacitance code information

This table provides the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Most capacitors actually have the numeric values stamped on them; however, some are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

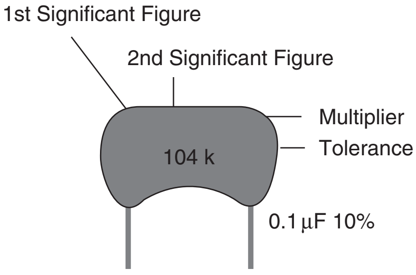

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

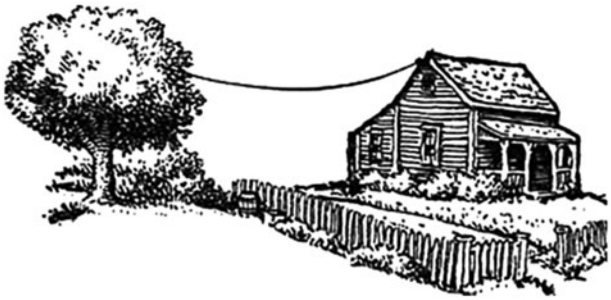

Figure 17-3 *Horizontal long-wire antenna*

wire away form the hot wires. The APM's ground wire may pass through conduit with other wires but should only be installed by qualified personnel. Grounding in extremely dry ground can be enhanced by burying some salt around the rod. The salt will increase the conductivity of the soil and also help retain water content around the ground rod. Some researchers have found that it is much better to get the largest surface area of metal in contact with the ground as possible, and have tried digging a hole and planting aluminum foil sheets in the ground.

Once you have tested the operation of the APM, you can disconnect the antenna and decide how you wish to enclose the APM circuit. You may want to experiment with the housing, since researchers have discovered that the germanium diodes are sensitive to sunlight and the voltage output is increased when the circuit is in sunlight. If you want to experiment along these lines, then you may wish to enclose the circuit in a clear plastic box. You can drill holes for two terminals posts for the plus and minus output connections, an RCA jack for the antenna input connection, a terminal for the ground connection, and your APM is complete and ready for operation.

The APM-II

The original APM is a great circuit for experimentation. It has been found that the germanium diodes are also sensitive to light, so if you house the APM in a clear plastic box and have it exposed to sunlight it will produce more output. Have fun experimenting with the APM receiver and you may find the perfect application for it, such as powering another receiver or night light, clock, emergency radio/lighting, etc.

Over time experimenters have improved upon the original APM design and claim increased output from the original design, so the reader is free to experiment and build both versions of the APM if desired. The second APM receiver, shown in Figure 17-4, illustrates the newer version of the APM with increased output voltage and current. The APM-II has quite a few more stages and has many more diodes and capacitors. You will note that the APM-II also uses germanium diodes which have more output than trying to use silicon diodes. Also note the higher capacitance values in the second design which were used to increase the output considerably. When constructing the APM-II be sure to pay careful attention to mounting the diode and capacitors; since there are many more components and they all have polarity marking on them, it is much more possible to make a mistake in orientation. Take your time when constructing the circuit. The APM-II also uses an input coil, which is said to improve the output. Some experimenters claim that the antenna can be eliminated and that the antenna connection can be connected to another separate ground, but this has not been verified. The APM-II is also sensitive to light which can increase the output if used outside. Some users claim that the power produced by the APM also seems to improve near bridges, ships and anything containing a great deal of metal.

Over time many scientists and researchers have reported bright flashes in the sky during strong earthquakes, as well as computer breakdowns during severe tremors. Scientists have long suspected that seismic activity is associated with a variety of electrical effects. Researchers have been taking a careful look at this link, with an eye toward using it to predict earthquakes.

Joseph Tate and William Daily of Lawrence Livermore National Laboratory in Livermore, CA. created a system of radio wave monitors distributed along California's San Andreas fault using the APM design. Tate and Daily recorded two kinds of changes in atmospheric radio waves prior to earthquakes that occurred between 1983 and 1986. Joe was able to witness and record the output of the APM during the earthquakes that followed, since he continuously monitored the output of the APM on a computer, so he could go back and look at the output on any given day.

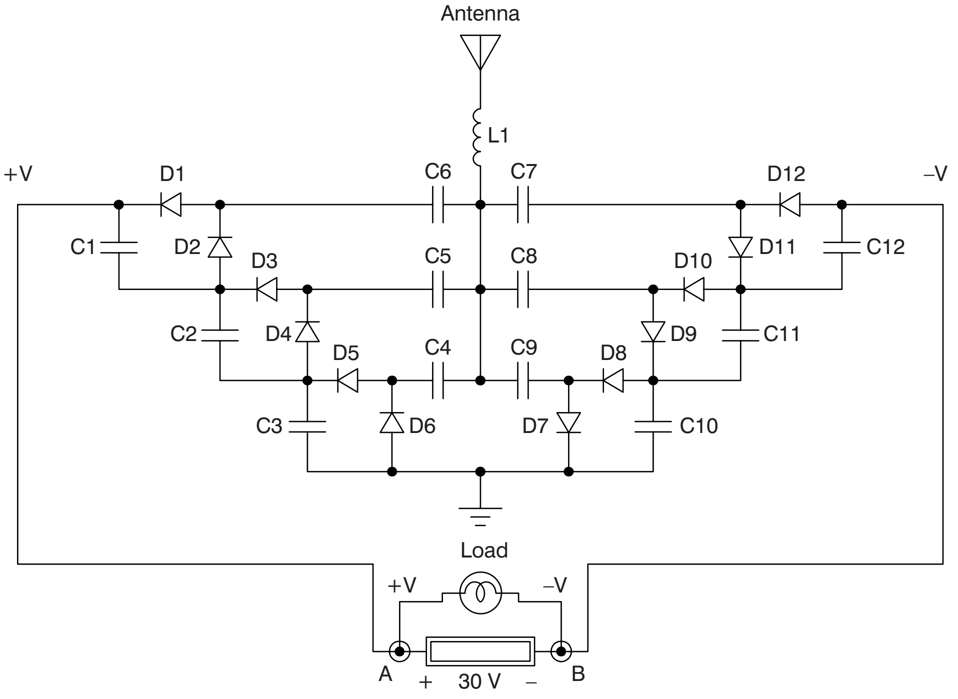

Figure 17-4 *Ambient power module II*

The most common change is a drop in the radio signals that normally pervade the air as a result of lightning and human sources such as car ignition systems and electric power grids. This reduction typically occurs one to six days before an earthquake and can last for many hours. For example, a magnitude 6.2 earthquake that shook Hollister, CA, in April 1984 was preceded six days earlier by a 24-hour drop in radio signals being monitored 30 miles from the quake's epicenter. Tate and Daily have found that the larger the earthquake, the longer the time between the radio wave depression and the quake.

Laboratory studies have shown that the electrical conductivity of rocks increases as they are stressed. Based on this and their electrical modeling of the ground, Tate and Daily think the increased conductivity of stressed rocks near the fault causes more radio waves to be absorbed by the ground rather than their traveling through the air. They also plan to test a possible link between radio wave drops and the emission of radon gas, which itself is thought to be a quake precursor. The radon may ionize the air, making it temporarily more absorptive than the detector antenna.

The researchers have also found, in addition to these drops, another pre-quake phenomenon in which short pulses of increased radio wave activity are emitted. For example, five days before the magnitude 6.5 earthquake hit palm Springs, CA, in July 1986, a station 15 miles from the epicenter detected a rise in radio signals. This sort of emission is consistent with laboratory work showing that cracking rocks release electromagnetic signals.

Tate says that in their first attempts at predicting earthquakes, they did not miss a single event, so he is optimistic about using this technique for short-term forecasting of San Andreas quakes.

If you live in an earthquake prone area such as California, you could use you APM not only for small low voltage emergency power/lighting system but you could connect the APM to an analog to digital converter card in a personal computer and record the fluctuations during earthquakes. The APM will provide many avenues for the builder and/or researcher. Disconnect your APM during thunderstorm activity to avoid damage to the APM and other household devices. Have fun and be safe.

Chapter 18

Earth Field Magnetometer Project

Parts list

Earth Field Magnetometer

R1,R3,R4,R5 4.7 k ohm, $\frac{1}{4}$ w, 5% resistor

R2 100 ohms, $\frac{1}{4}$ w, 5% resistor

R6 10k ohm $\frac{1}{4}$ w, 5% resistor

C1,C2 15 pF, 35 volt capacitor

C3,C4 47 nF, 35 volt electrolytic capacitor

C5 10 μF, 35 volt capacitor

C6,C8 1 μF, 35 volt capacitor

C7 22 μF, 35 volt electrolytic capacitor

C9 10 μF, 35 volt Tantalum capacitor

D1,D2 1N914 silicon diode

D3 1N4001 silicon diode

X1 10 MHz crystal

U1 SCL006A - magnetometer output IC

U2 AD-557 - digital to analog converter chip

U3,U4 LM7805, 5 volt regulator

J1,J2 wire jumpers

S1 4-position rotary gain switch

S2 normally open pushbutton switch

S3 SPST toggle power switch

Logger HOBO H08-002-02 (ONSET)

Misc PC board, wire, IC sockets, chassis, etc.

Optional Field Magnetometer Power Supply

C1 470 μF, 50 volt electrolytic capacitor

C2 1 μf, 50 volt ceramic capacitor

C3 15 μf, 50 volt electrolytic capacitor

BR1 diode bridge rectifier 100v/2-amp

T1 24 volt center tapped transformer -.5 A

U1 LM7809 IC regulator 9 volts

S1 SPST power switch

PL1 115 VAC power plug

MISC PC board, wire, hardware, etc.

1. Speake & Co. Ltd. (Elvicta Estate, Crickhowell, Powys. Wales, UK; http://www.speakesensors.com).

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

2. Fat Quarters Software (24774 Shoshonee Drive, Murieta, CA 92562; Tel: 909-698-7950; Fax: 909-698-7913; Web: http://www.fatquarterssoftware.com)
3. ONSET Computers, PO Box 3450 Pocasset, MA 02559-3450 (1-800-564-4377) http://www.onsetcomp.com

Magnetic fields are all around us. The Earth itself produces a magnetic field, which is why compasses work. Anytime an electrical current flows in a conductor, a magnetic field is generated. That is why transformers, inductors, and radio antennas work. There are several different devices that could be used to sense a magnetic field. One of the most familiar to electronics hobbyists is the Hall-effect device. However, in this chapter we'll take a look at a magnetic sensor that is as easy to use, but is more sensitive, more linear, and more temperature stable than typical Hall-effect devices. And just like Hall-effect devices, it can be used to make a variety of instruments, including magnetometers and gradiometers.

For those unfamiliar with them, magnetometers are used in a variety of applications in science and engineering. One high-tech magnetometer is used by Navy aircraft to locate submarines. Radio scientists use magnetometers to monitor solar activity. Earth scientists use magnetometers to study diurnal changes in the Earth's magnetic field. Archeologists use magnetometers to locate buried artifacts, while marine archeologists and treasure hunters use the devices to locate sunken wrecks and sunken treasure.

The Earth Field Magnetometer that we will be exploring uses a "flux-gate magnetic sensor." The device, in essence, is basically an over-driven magnetic-core transformer which the "transducible" event is the saturation of the magnetic material. These devices can be made very small and compact, yet will still provide reasonable accuracy.

In its most simple form the flux-gate magnetic sensor consists of a nickel-iron rod used as a core, wound, with two coils. One coil is used as the excitation coil, while the other is used as the output or sensing coil. The excitation coil is driven with a square-wave signal with an amplitude high enough to saturate the core. The current in the output coil will increase in a linear manner so long as the core is not saturated. But when the saturation point is reached, the inductance of the coil drops and the current rises to a level limited only by the coil's other circuit resistances.

If the simple flux-gate sensor was in a magnetically pure environment, then the field produced by the excitation coil would be the end of the story. But there are magnetic fields all around us, and these either add to or subtract from the magnetic field in the core of the flux-gate sensor. Magnetic field lines along the axis of the core have the most effect on the total magnetic field inside the core. As a result of the external magnetic fields, the saturation condition occurs either earlier or later than would occur if we were only dealing with the magnetic field of interest. Whether the saturation occurs early or later depends on whether the external field opposes or reinforces the intended field.

Flux-gate magnetometer

The Earth Field Magnetometer project is illustrated in Figure 18-1 and the remotely located FM-3 flux-gate sensor is depicted in Figure 18-2.

The heart of the magnetometer is of course the flux-gate sensor. A compact low-cost line of flux-gate sensors are manufactured by Speake & Co. Ltd. and distributed in the United States by Fat Quarters Software. The FGM-3 sensor is the device used in

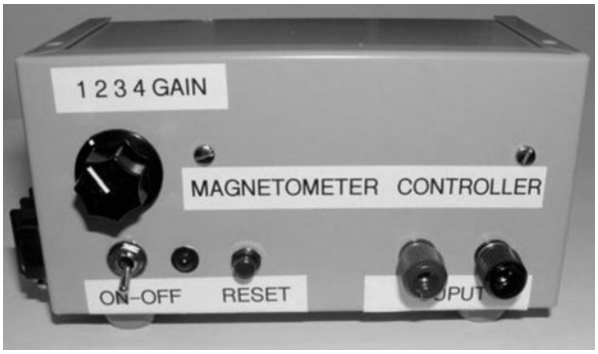

Figure 18-1 *Magnetometer*

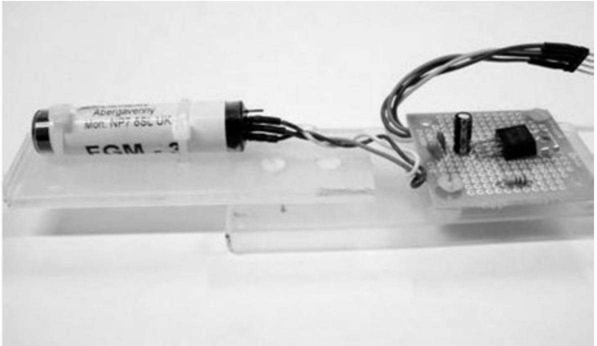

Figure 18-2 *FGM-3 Magnetometer sensor*

this project. It is a 62 mm-long by 16 mm-diameter (2.44 by 0.63 inch) device. Like all the devices in the line, it converts the magnetic field strength to a signal with a proportional frequency. The FGM-3 sensor has only three leads: Red: +5 VDC (power), Black: 0 volts (ground), and White: output signal (a square-wave whose frequency varies with the applied field). The output signal is a train of TTL compatible pulses with a period that ranges from 8 to 25 mS, or a frequency range of 40 to 125 kHz. The detection sensitivity of the FGM-3 device is +/– 0.5 oersted (+50 utesla). That range covers the Earth's magnetic field, making it possible to use the sensor in Earth-field magnetometers. Using two or three sensors in conjunction with each other provides functions such as compass orientation, three-dimensional orientation measurement systems, and three-dimensional gimbaled devices such as virtual-reality helmet display devices. It can also be used in applications such as ferrous metal detectors, underwater shipwreck finders, and in factories as conveyer-belt sensors or counters. There are a host of other applications where a small change in a magnetic field needs to be detected.

The FGM series also includes two other devices, the FGM-2 and the FGM-3h. The FGM-2 is an orthogonal sensor that has two FGM-1 devices on a circular platform at right angles to one another. That orthogonal arrangement permits easier implementation of orientation measurement, compass, and other applications. The FGM-3h is the same size and shape as the FGM-3, but is about 2.5 times more sensitive. Its output frequency changes approximately 2 to 3 Hz per gamma of field change, with a dynamic range of +/– 0.15 oersted (about one-third the Earth's magnetic field strength).

The output signal in all the devices in the FGM series is +5 volt (TTL-compatible) pulse whose period is directly proportional to the applied magnetic field strength. This relationship makes the frequency of the output signal directly proportional to the magnetic field strength. The period varies typically from 8.5 μs to 25 μs or a frequency of about 120 kHz to 50 kHz. For the FGM-3 the linearity is about 5.5% over its +/– 0.5 oersted range.

The response pattern of the FGM-x series sensors is shown in Figure 18-3. It is a "figure-8" pattern that has major lobes (maxima) along the axis of the sensor, and nulls (minima) at right angles to the sensor axis. This pattern suggests that for any given situation there is a preferred direction for sensor alignment. The long axis of the sensor should be pointed towards the target source. When calibrating or aligning sensor circuits, it is common practice to align the sensor along the east-west direction in order to minimize the effects of the Earth's magnetic field.

The flux-gate sensor is coupled directly to a special interface chip, manufactured by Speake. This special SCL006A integrated circuit is shown in the main schematic diagram shown in Figure 18-4. It provides the circuitry needed to perform the Earth field magnetometry sensing. It integrates field fluctuations in one-second intervals, producing very sensitive output variations in response to small field variations. The magnetometer is of special interest to people doing radio-propagation studies, and those who want to monitor for solar flares. It also works as a laboratory magnetometer for various purposes. The SCL006A is housed in an 18-pin DIP IC package and is shown at U1. The FGM-3

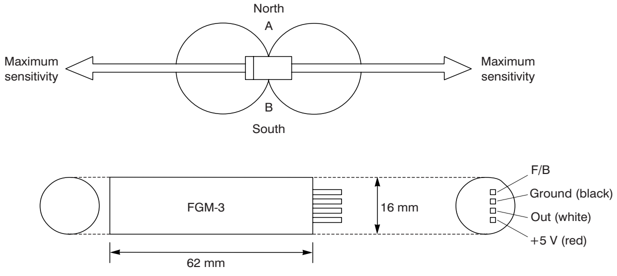

Figure 18-3 *FGM-3 Magnetometer sensor*

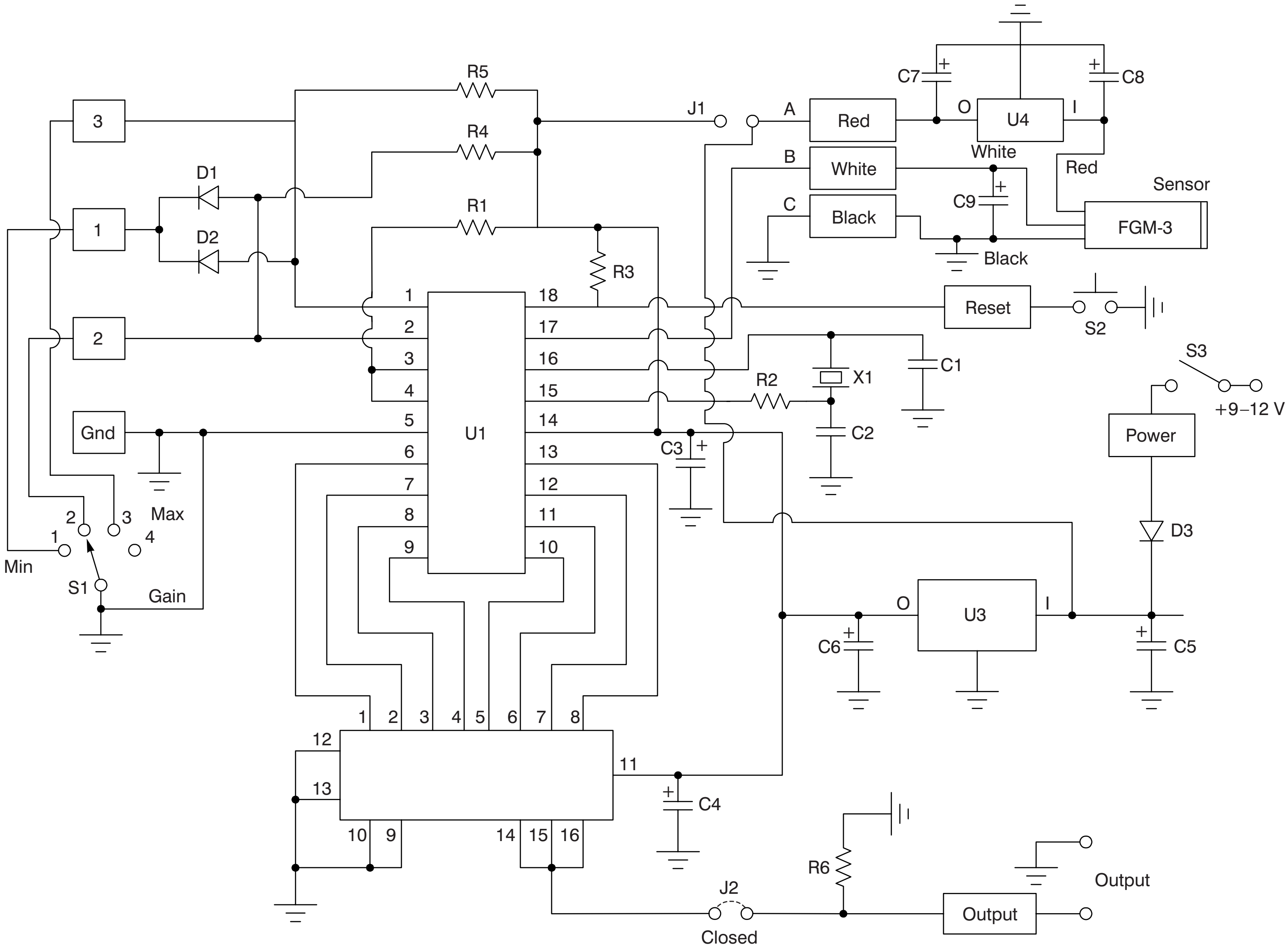
R5
R4
R1
R3
R2
R6
D1
D2
D3
J1
J2
Closed
A
B
C
Red
White
Black
U1
U3
U4
O
I
C1
C2
C3
C4
C5
C6
C7
C8
C9
X1
Sensor
FGM-3
Reset
S1
S2
S3
Power
+9–12 V
Output
Gnd
Min
Max
Gain
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

Figure 18-4 *Main magnetometer circuit*

flux-gate sensor is coupled to the input pin at pin 17 of U1. The reference oscillator of the SCL0006A is controlled by components X1, C1, C2 and R2. A sensitivity switch, S1, provides four positions, each with a different overall sensitivity range. Switch S2 is used to reset the SCL0006A, when applied for about two seconds.

The second op-amp at U2, is an Analog Devices AD557 D/A converter. The magnetometer circuit is designed so that it could be run from a 9 to 15 volt battery for use in the field. Power supplied from the battery is controlled by power switch S3 which supplies 9 volts to the regulator. For stationary research, the magnetometer should be powered from an 110 volt AC to 9–15 volt DC power supply. The regulator U3 drops the 9 volt input supply power down to 5-volts to power the magneto-meter circuit. The output signal is a DC voltage that can be monitored by a strip-chart or X-Y paper recorder, a voltmeter or the signal can be fed into a computer using an A-to-D converter. In the prototype we used the output to drive a portable mini data-logger such as the ONSET corporation HOBO H08-002-02. You will want to select the 2.5 volt model logger. You could also elect to use a digital multi-meter with RS-232 output which can be fed to a computer's serial port for data-logging.

Construction

The Earth-field magnetometer project was built on two printed circuit boards. The main circuit board is shown in Figure 18-5, was constructed on a 3½″ × 4½″ PC board, while the remotely located FGM-3 sensor was paired with a second 5-volt regulator shown in Figure 18-6. Regulator U4 was mounted on a 2″ × 3″ PC board, which was attached to a sheet of flexible plastic using plastic screws. The FGM-3 sensor probe was then mounted on the 2″ × 4½″ plastic sheet. The plastic sheet was used to mount the probe, ahead of the regulator in order to give the flux-gate probe some isolation from surrounding metal. The flux-gate probe is mounted length-wise and parallel to the length of the PVC pipe. The FGM-3 sensor and remote regulator were then housed in a 2½″ diameter × 12″ long PVC pipe, which is separated by a length of three-conductor cable. Note that you could also elect to purchase a flux-magnetometer kit including PC board, the FGM-3

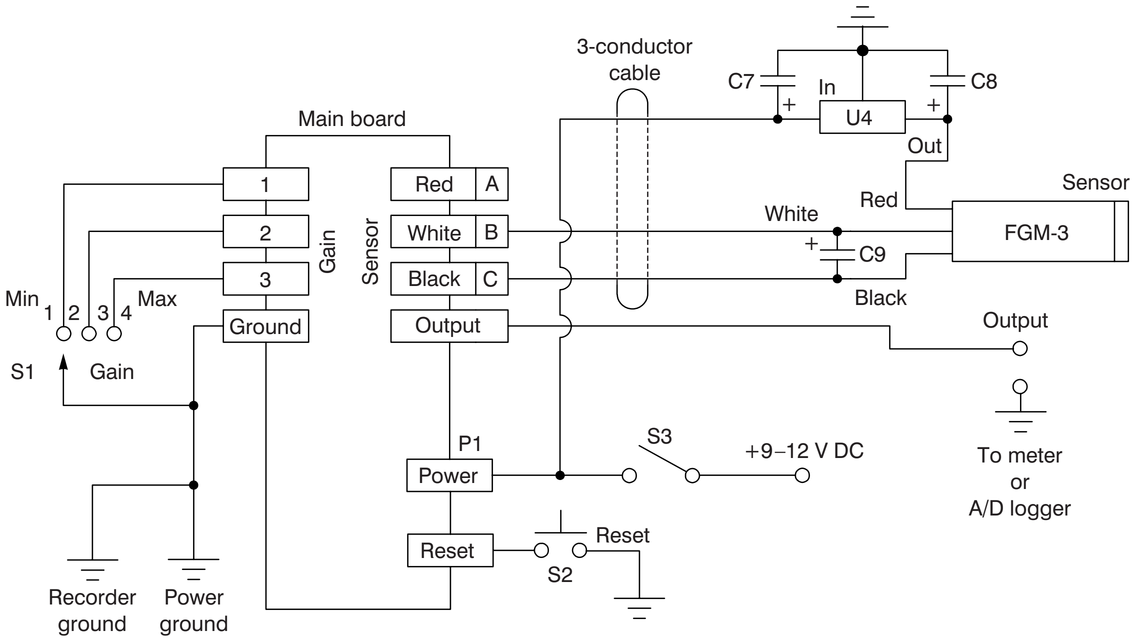

Figure 18-5 *Magnetometer connection diagram*

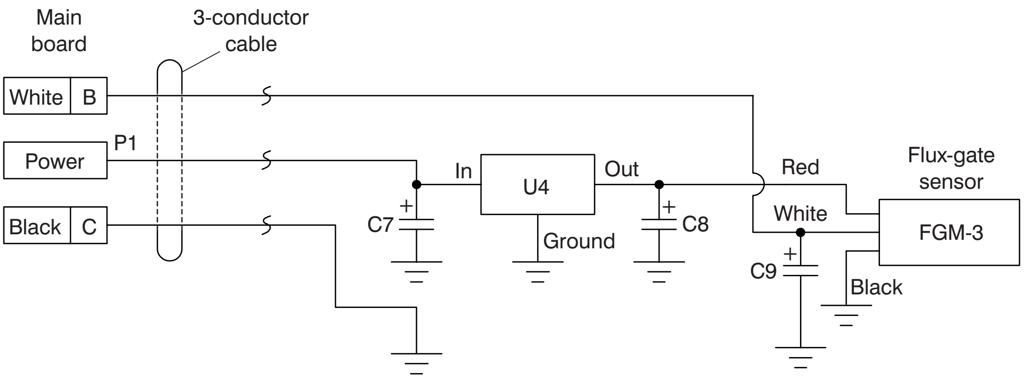

Figure 18-6 *Flux-gate sensor board*

sensor, ICs, and most other parts, but less the gain switch, can be from Fat Quarters Software for about $80.00.

Are you ready to begin constructing the flux-gate magnetometer? Before we begin you will want to secure a well lit and well ventilated work area, as well as a bench or large table for assembling your magnetometer project. Locate a small pencil-tipped 27 to 33 watt soldering iron, as well as a roll of 22 ga. 60/40 tin/lead rosin core solder. Go to your local Radio Shack store and purchase a small jar of “Tip Tinner”; this is a soldering iron tip cleaner/dresser compound, which works wonders. You will also need a few hand tools for the project. You will need a pair of small end-cutters, a pair of needle-nose pliers, a pair of tweezers, and a magnifying glass. A small flat-blade screwdriver and a small Phillips screwdriver should round out your tool requirements.

While you could build the magnetometer project on a perf-board or you could use point-point wiring it is recommended that you build the circuit on a printed circuit board for most reliable circuit operation and long life. Prepare your circuit board, and then we will be able to go ahead and build the magnetometer receiver project.

Locate all of the project diagram, such as the schematic, parts layout diagrams and the component pin-out sheets as well as the resistor and capacitor identification tables and we will be almost ready to begin. Place all of the magnetometer project components on the work bench in front of you, warm up your soldering iron and let’s get going.

First, refer to the chart in Table 18-1, which illustrates the resistor color code. The resistor for this project will have three or four color bands on them. The first color band is closest to one end of the resistor body. The first color band represents the first digit of the resistor’s value, the second color represents the second digit of the resistor value, the third color band represents the resistor’s multiplier value, and the fourth band denotes the resistor’s tolerance value. A silver fourth band denotes a 10% tolerance value, while a gold band represents a 5% tolerance value; no fourth color band denotes a resistor with 20% tolerance. Resistors R1, R3, R4 and R5, for example, are 4.7k ohm types. The first color band is yellow, which represents a four (4), while the second color band is violet or seven (7), and the third color band is red denoting a multiplier of (100), so you see (4) (7) times (100) equals 4700 or 4.7k ohms. Identify all of the 4.7k ohm resistors and place them on the circuit board at their correct location. Next, solder these resistors to the circuit board. Then follow-up by cutting the excess component leads flush to the edge of the circuit board with your end-cutters. Now locate R2 and R6 and install them in their proper locations, then solder them in place on the PC board. Remember to trim the excess resistor leads with your end-cutter.

Next we will move to identifying and installing the capacitors for the project. Capacitors are available in two major categories—electrolytic and non-electrolytic types. Electrolytic capacitors are generally both larger in size and larger in value than non-electrolytic types. Electrolytic capacitors have polarity and hence, they will have polarity marking on them to identify

Table 18-1
Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

their leads. You will usually find a black or white colored band or a plus (+) or minus (–) marking near one of the capacitor leads; this denotes the polarity. You will have to orient an electrolytic capacitor onto the circuit board with respect to the capacitor's polarity. When you mount electrolytic capacitors on the circuit board, you will have to refer to both the circuit schematic and the parts layout diagrams in order to install the capacitor properly. Failure to install electrolytic capacitors correctly may result in damage to the capacitor itself and perhaps damage to the circuit as well. Non-electrolytic capacitors will generally be smaller in value and physical size. Sometimes the non-electrolytic capacitors will have their value printed on them, and at other times the capacitor may be physically very small when a three-digit code is used to identify the capacitor. The chart in Table 18-2, helps to identify capacitors based on this code. Use the chart to help identify the capacitors. You will want to make sure that you install the correct value capacitor into the right location on the circuit board in order for the circuit to work correctly. For example, a capacitor marked with (102) denotes a value of .001 µF, while a capacitor with a marking of (103) denotes a capacitor value of .01 µF.

Look through the component pile and identify the capacitors and what type they are and then refer to the parts layout diagram and capacitor code chart before installing them on the circuit board. Go ahead and place a few of the capacitors at their respective locations on the PC board and solder them in place. Remember to trim the extra component lead length with your end-cutters.

The magnetometer main circuit board utilizes three silicon diodes. Diodes have polarity and they must be installed correctly for the circuit to operate correctly. Each diode will have either a black or white colored band at one end of the diode body. The colored band represents the diode's cathode lead, the opposite lead is the anode. Remember the diode's electronic symbol is an arrow or triangle pointing to a vertical line. The vertical line is the cathode of the diode. Refer to the schematic and parts layout diagram when mounting the diodes onto the circuit board. Place the diodes on the circuit board and solder them in place. Trim the excess leads as necessary.

The magnetometer project employs a 10 MHz crystal which is connected to pin 16 of U1. The crystal has no polarity so it can be placed in either direction on the board. Handle the crystal carefully so as to not break off the leads from the crystal body.

Before we begin placing the semiconductors on the PC board, you will want to refer to the semiconductor

Table 18-2

Capacitance code information

This table provides the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Most capacitors actually have the numeric values stamped on them; however, some are color coded and some have alphanumeric codes. The capacitor’s first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

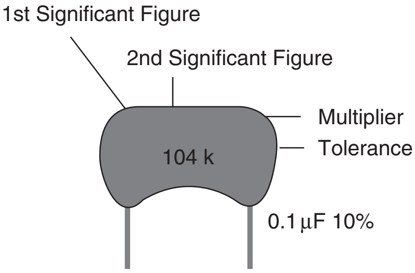

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

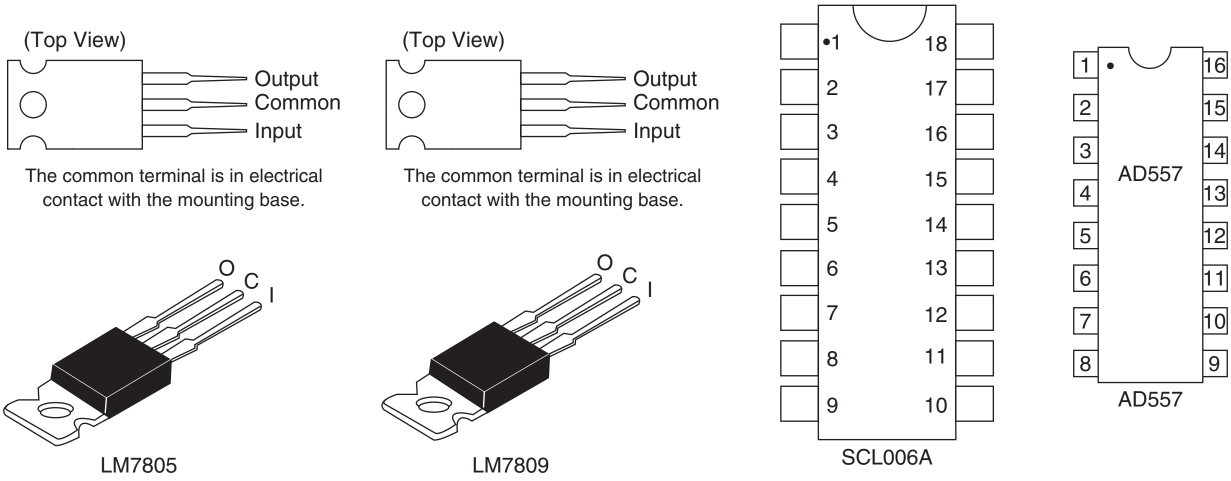
(Top View)
Output
Common
Input
The common terminal is in electrical contact with the mounting base.
O
C
I
LM7805
(Top View)
Output
Common
Input
The common terminal is in electrical contact with the mounting base.
O
C
I
LM7809
•1
2
3
4
5
6
7
8
9
18
17
16
15
14
13
12
11
10
SCL006A
1
2
3
4
5
6
7
8
16
15
14
13
12
11
10
9
AD557
AD557

Figure 18-7 *Semiconductor pin-outs*

pin-out diagram shown in Figure 8-7, which will assist you in orienting the ICs. The magnetometer project utilizes two main integrated circuits which are rather expensive, so we recommend using IC sockets. As a good insurance policy locate two IC sockets and solder them in place for U1 and U2. If an IC fails at some later date 'down the road', you can simply unplug the defective one and replace it with a new one. Note that the IC sockets will usually have a notch or cut-out at one end of the plastic package. The lead just to the left of the notch is pin. You will have to make sure that pin one (1) of each socket is connected to the correct components; refer to the schematic and parts layout diagrams. Note pin one (1) of U1 goes to the anode of diodes D1 and D2, while pin one (1) of U2 is connected to pin six (6) of U1. When installing the ICs into their respective sockets, you will have to make sure pin one (1) of the IC is placed into pin one (1) of the socket to prevent damage to the integrated circuits.

It is a good time to take a short, well deserved rest and when we return we will examine the main circuit board for any possible "cold" solder joints or "short" circuits. Pick up the main circuit board with the foil side facing upwards towards you. Look over the main circuit board, we will be looking at the solder joints for possible "cold" solder joints. The solder joints should all look clean, bright and shiny, if you see any solder joints that look dull, dark or "blobby," you will need to un-solder the joint and remove the solder with a solder-sucker and then re-solder the connection, so that the new solder joint looks good. Next we will examine the PC board for possible "short" circuits. Often "short" circuits can be caused by "stray" component leads or solder "blobs" which may have stuck to the underside of the PC board from solder residue. Look for solder blobs or solder bridges which may be "shorting" or bridging between circuit traces on the PC board. With the PC board free from any defect, we can continue on and prepare for connecting the sensor and power-up the circuit.

The main magnetometer board was mounted on standoffs in a metal chassis box, as seen in Figure 18-8. The on-off power switch and sensitivity rotary switch were mounted on the front panel of the chassis box, along with the meter binding posts. A 3-position screw terminal block or 3-circuit connector can be mounted on the side of the chassis box to allow the main circuit to connect to the remote sensor board.

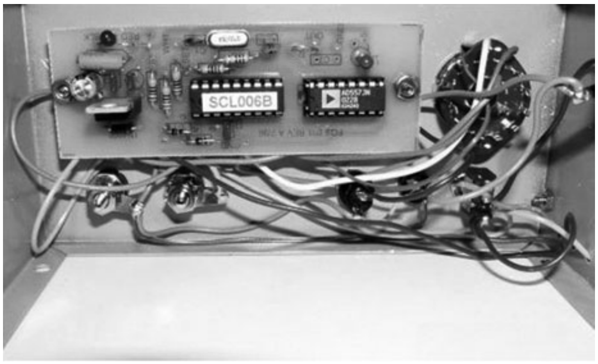

Figure 18-8 *Magnetometer – inside view*

Now refer to the sensor power supply circuit diagram shown in Figure 18-6. This diagram illustrates the remote flux-gate sensor and power regulator PC board. The sensor and regulator were placed on a small circuit board, and in a water proof PVC cylinder so it can be buried below the frost-line at about 4 feet underground. The sensor should be placed well away from your house, buried level in the ground with the ends of the sensor facing East-West. The remote sensor is powered from the main circuit board. You will want to place the remote sensor and power supply in a non-metallic PVC case and waterproof the case so water cannot leak inside the case. The sensor will need to be facing the outside of the PVC cylinder away from the other electronics inside the package. Use no metal fasteners to secure the circuit to the PVC enclosure. You will have to make provisions to allow the three conductor to exit the opposite end of the PVC case away from the sensor end.

The main magnetometer circuit board was placed in a metal chassis which measures 6″ × 3″ × 2″ as shown in the photo. The power switch, power LED, gain switch and reset switch along with the output binding posts, were mounted on the front of the chassis box. A four position terminal strip was mounted along the side of the chassis box to allow the remote sensor to be connected to the main circuit. A coaxial power jack was also mounted on the side of the chassis box to allow a 12 volt "wall wart" power supply to plug-in and power the circuit.

If you wish to construct your own power supply for the magnetometer project, refer to the diagram in Figure 18-9, which illustrates a 12 volt power supply from line-voltage. A 24 volt center tapped transformer feeds AC to the 4-diode bridge rectifier package at BR1. The DC output from the rectifier is filtered via C1 and C2 and then sent to a 9 volt IC regulator chip at U1. The output of the regulator chip is filtered at its output. This optional power supply will provide 9 volts which can be used to power the magnetometer circuit, if you choose not to use a 9 to 12 "wall wart" power supply.

The main circuit board was mounted on standoffs inside the bottom of the chassis box. Once you have mounted the main circuit board and the controls, you can locate a length of 20 to 24 gauge, three or four conductor cables and connect the sensor board to the main circuit board. Connect a digital voltmeter to the output terminal binding posts and set the meter to read 5 volts to 10 volts. Connect up your power supply to the main circuit board and we can now test the magnetometer to see if it is working. Turn the power switch to on, and set the meter range or gain switch to maximum sensitivity. With the sensor away from the main circuit board and chassis, bring a small magnet within rage of the sensor to say about 2 feet away. After about eight or ten seconds, you should see the effects of bringing the magnet near to the sensor. The reading will NOT be instantaneous, but you should see the meter read up to 2 volts or so. Move the magnet away and you should see the meter respond in 8 or 10 seconds. If all is well you will have a working portable flux-gate magnetometer, which can be used for many applications such as monitoring diurnal magnetic Earth variations or solar storms. You can even use the magnetometer as a UFO detector or driveway monitor.

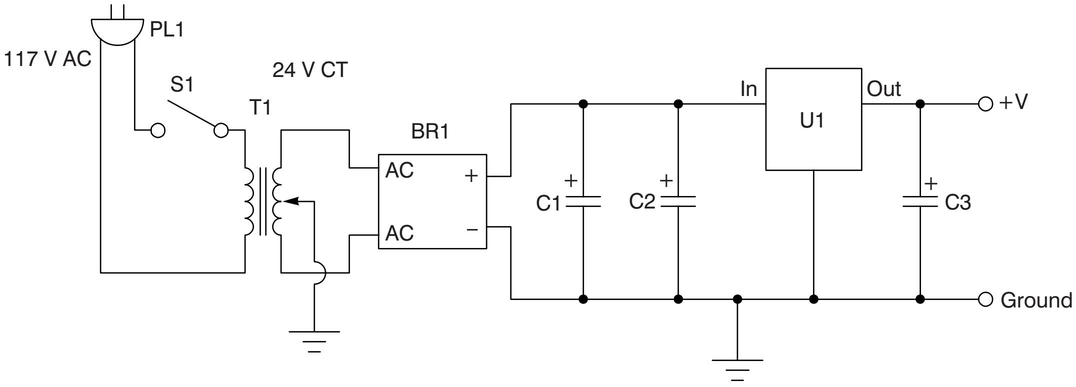

Figure 18-9 *Magnetometer power supply*

If your magnetometer did not “spring-to-life” then you will have to turn off the circuit and remove the power supply, and look the circuit over carefully to make sure you didn’t install a component incorrectly. Grab your schematic, parts layout diagrams and charts and let’s take a closer look at the circuit board. Remember we talked about certain components, like capacitors and diodes having polarity. Make sure that you installed the electrolytic capacitors and diodes with respect to their proper polarity; this is common cause for error. If need be, have another person, such as a parent or skilled technician or engineer or knowledgeable friend, look over the circuit for you. The same person looking over the same circuit numerous times often prevents you from finding your errors; you need a fresh look from another person to catch mistakes. Another cause for error is placing the integrated circuits in the wrong socket or in the wrong direction in the socket. After looking the board over carefully, you can re-apply the sensor and power supply connections and test the circuit once again, hopefully it will work fine this time around.

Once you know the magnetometer and sensor circuits are working properly, you can make provisions for mounting the sensor in a water-tight PVC case. The remote flux-gate sensor probe was mounted in a 2½″ × 12″ length of PVC pipe. Circuit board guides were epoxied to the inside of the PVC pipe to allow the circuit board to be slid into the PVC pipe. One 2½″ end cap was cleaned and cemented to the outside probe end of the PVC pipe. A ½″ hole was drilled and tapped in the center of the second PVC end cap. Try to use as little metal as possible when mounting the sensor in the PVC pipe, use brass or plastic screws where possible. A brass nipple fitting cemented in the tapped hole allows the three conductor cables to exit the probe assembly in the PVC pipe. Before cementing the final end cap to the PVC pipe you will want to test the Earth-Field magnetometer before burying the sensor probe assembly. Since the sensor is NOT an expensive temperature referenced or compensated type sensor it is subject to extreme temperature variations, so if you live in the northern climate zones, it will be necessary to mount the sensor underground, below the frost-line, to prevent temperature swings from affecting the reading of the meter. In most frost-prone areas, you will need to mount the water-proof sensor assembly about 4′ to 4½′ underground. The sensor should be mounted as far away from metal objects and traffic as possible to ensure the best results. The sensor package should be mounted so that it will face North-South as shown in Figure 18-3.

The output of the magnetometer can be connected to a logging digital multi-meter, so the imported data can be later analyzed by your personal computer. You could also use an inexpensive portable data-logger such as the HOBO series from Onset Computers, or you could use an analog to digital converter card in your PC to collect the data sample over time. If your meter is set to the most sensitive position, and everything is in working order, you should easily be able to monitor the Earth’s diurnal variations twice a day. You should also be able to detect strong magnetic storms on the Sun, with your new magnetometer.

Chapter 19

Sudden Ionospheric Disturbance (SIDs) Receiver

Parts list

SIDs Receiver

R1,R2,R3,R4	3.3 k ohm ¼ w resistor
R5	10 k ohm potentiometer (chassis mount)
R6	1 k ohm ¼ w resistor
R7	100k potentiometer (PC mount)
R8	100 ohm ¼ w resistor
R9	10 k ohm ¼ w resistor
R10	470 k ohm ¼ w resistor
R11	56 k ohm ¼ w resistor
R12	22 k ohm ¼ w resistor
R13	5 k ohm potentiometer (PC mount)
C1	100 pF, 35 v ceramic capacitor
C2	1500 pF, 35 v mylar capacitor
C3	.001 µF, 35 v ceramic capacitor
C4	1 µF, 35 v tantalum capacitor
C5	10 µF 35 v tantalum capacitor
C6,C7	10 µF, 35 v electrolytic capacitor
C8,C9	0.1 µF, 35 v ceramic capacitor
C10,C11	1000 µF, 35 v electrolytic capacitor
D1,D2	1N914 silicon diode
D3,D4	1N34 germanium diode
U1	LM353 op-amp (National Semiconductor)
U2	RC4136 op-amp (Texas Instruments)
U3	79L09 - 5 volt regulator (– volts)
U4	78L09 - 5 volt regulator (+ volts)
T1	600 to 600 ohm inter-stage/ matching transformer
L1	Loop antenna (see text)
S1	DPST toggle switch (on-off)
F1	.50 amp fast-blow fuse
BR1	bridge rectifier - 50 v DC/2 amp
PL1	117 VAC plug
Misc	PC board, IC sockets, wire, solder, solder lugs, chassis hardware, screws, standoffs, etc.

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

A sudden ionospheric disturbance (SID) is an abnormally high plasma density in the ionosphere caused by an occasional sudden solar flare, which often interrupts or interferes with telecommunications systems. The SID results in a sudden increase in radio-wave absorption that is most severe in the upper medium-frequency (MF) and lower high-frequency (HF) ranges.

When a solar flare occurs on the Sun, a blast of ultraviolet and X-ray radiation hits the day-side of the Earth after 8 minutes. This high energy radiation is absorbed by atmospheric particles raising them to excited states and knocking electrons free in the process of photo-ionization. The low altitude ionospheric layers (D region and E region) immediately increase in density over the entire day-side. The Earth's ionosphere reacts to the intense X-ray and ultraviolet radiation released during a solar flare and often produces shortwave radio fadeout on the day-side of the Earth as the result of enhanced X-rays from a solar flare.

Shortwave radio waves (in the HF range) are absorbed by the increased particles in the low altitude ionosphere causing a complete blackout of radio communications. This is called a Short Wave Fadeout. These fadeouts last for a few minutes to a few hours and are most severe in the equatorial regions where the Sun is most directly overhead. The ionospheric disturbance enhances long wave (VLF) radio propagation. SIDs are observed and recorded by monitoring the signal strength of a distant VLF transmitter. As the fadeouts occur, reception of the station strength varies thus creating a fluctuating voltage output at the receiver which can be recorded and observed using a computer with a chart recorder program.

You can build and investigate the phenomena of SIDs with a special receiver and a low cost data-logger setup. You can not only observe when solar flares are occurring but you can collect and analyze the data and display it on your computer. The SIDs receiver is a great opportunity to observe first-hand when a solar event occurs, but you can use it to predict when radio blackouts will affect radio propagation, which is extremely useful for amateur radio operators.

VLF signal propagation

Why do VLF signals strengthen at night instead of getting weaker? If propagation is basically via the waveguide effect, why doesn't the signal drop DOWN at night when the waveguide disappears with the D layer? Is there some kind of reduced absorption at night? If so, where is it taking place and why? Also, what accounts for the big fluctuations in signal strength at night, apparently more or less at random?

The strength of the received signal depends on the effective reflection coefficient of the region from which the radio wave reflects in its multi-hop path between Earth and the ionosphere. In daytime the reflecting region is lower, the air density is higher, and the free electron density is controlled strongly by the solar radiation, etc. At night time the reflecting region is higher, the air density is lower, and the free electron density is controlled by variable ambient conditions as well as variable influences from electron "precipitation" from above, etc.

At noon the electron density is about 10 electrons/cm^3 at an altitude of 40 km, 100 electrons/cm^3 at an altitude of 60 km, 1000 electrons/cm^3 at an altitude of 80 km and 10,000 electrons/cm^3 at an altitude of 85 km.

At night these figures become 10 electrons electrons/cm^3 at 85 km, 100 electrons/cm^3 at 88 km and 1000 electrons/cm^3 at 95 km and then remains somewhat the same up to at least 140 km. At night the electron density in the lower part of the D region pretty much disappears. At 40 km the electron collision frequency is about 1,000,000,000 collisions per second whereas at 80 km the collision frequency drops to 1,000,000 collision per second. The reflection coefficient depends on (among other things) the number density of free electrons, the collision frequency, and the frequency of the radio signal. It is found by a mathematical integration throughout the entire D-region and of course the result depends on what time of the 24 hour day one performs the integration.

We can think of the E-Layer propagating the signal at night. Then the prominent sunrise pattern we see is a shift from E-Layer propagation back to D-Layer as the sun rises and forms the daytime D-Layer.

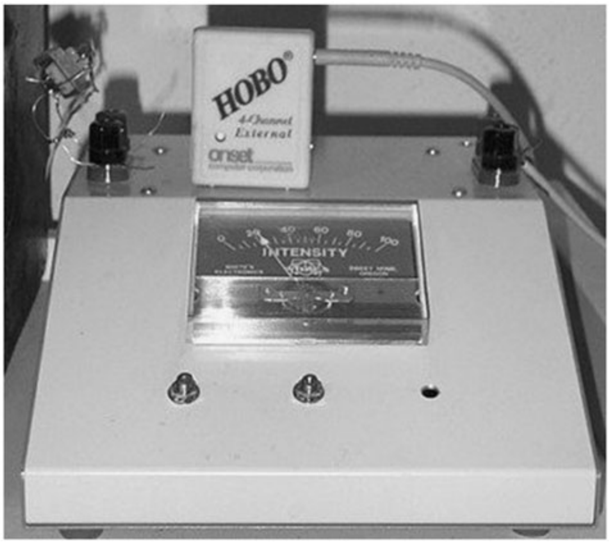
HOBO
4-Channel
External
onset
computer corporation
INTENSITY

Figure 19-1 *Sudden ionospheric disturbance receiver*

The sunset pattern is the reverse. An interesting feature of waveguide mode propagation was that the signal was split into two components which can form an interference pattern.

You can build your own Sudden Ionospheric Disturbances (SIDs) receiver, shown in Figure 19-1, and begin your own investigation of solar flares and their effects on radio propagation. The SIDs receiver is a simple VLF receiver designed to be used with a loop antenna which can be placed either inside or outside. The receiver monitors the strong VLF signal from the US Navy NAA 24 kHz transmitter in Cutler Maine.

The Sudden Ionospheric Disturbance (SID) receiver is a modified Stokes' Gyrator tuned VLF receiver based on the circuit from Communications Quarterly Spring 1994 pp. 24–26. The main circuit diagram shown is Figure 19-2, it illustrates SIDs receiver which consists of two integrated circuits, a Texas Instruments RC4136 and a National LF353 operation amplifier. The SIDs receiver is fed from a loop antenna at L1. The antenna is coupled to the SIDs receiver via a miniature 600 to 600 ohm matching transformer at T1. The output from the secondary of T1 is fed to two protection diodes at D1 and D2 at the front end of the receiver. The output of the diode network is then coupled to a 100 pF capacitor at C1. Capacitor C2 is connected from the

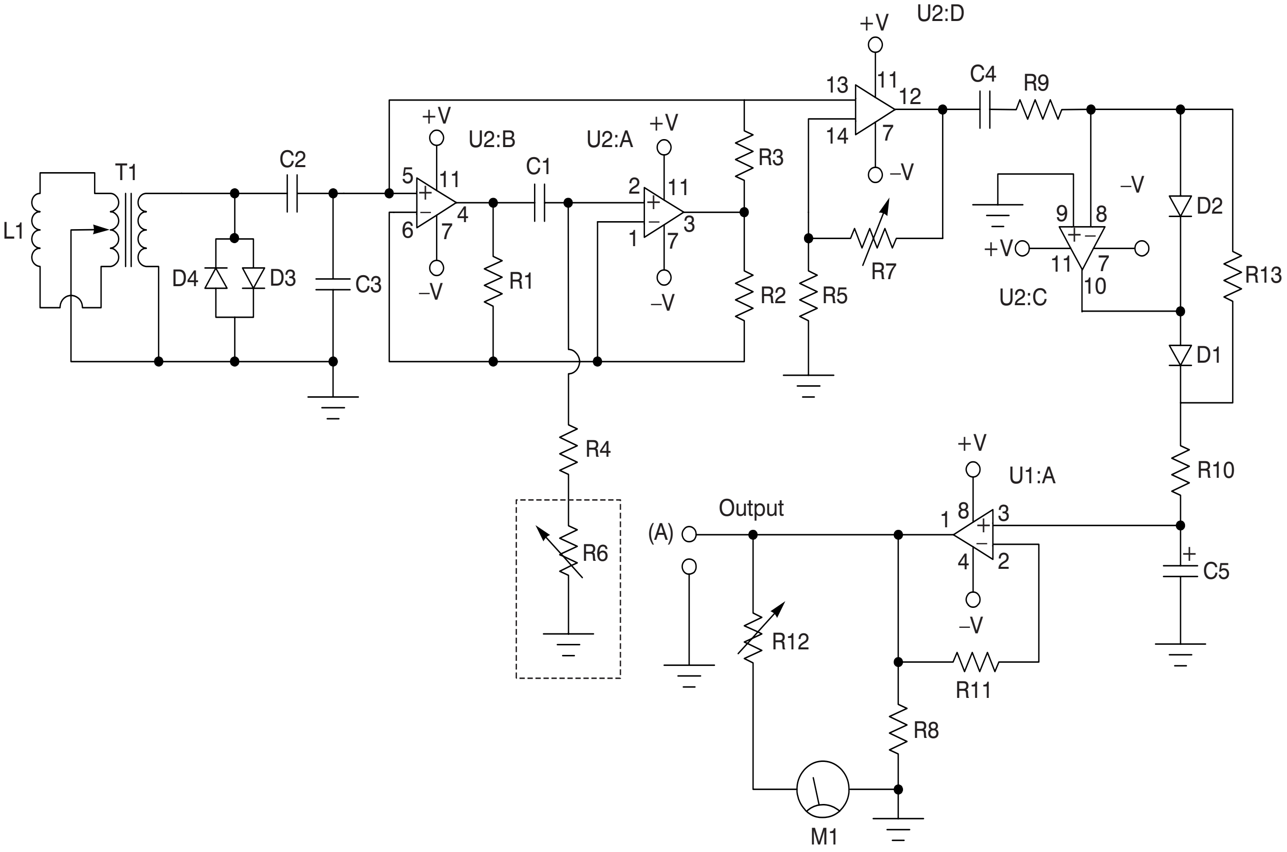
L1
T1
D4
D3
C2
C3
+V
U2:B
5
11
6
4
7
–V
R1
C1
U2:A
+V
2
11
1
3
7
–V
R3
R2
R4
R6
+V
U2:D
13
11
12
14
7
–V
R7
R5
C4
R9
–V
9
8
+V
11
7
10
U2:C
D2
R13
D1
R10
C5
+V
U1:A
8
1
3
4
2
–V
Output
(A)
R12
R11
R8
M1

Figure 19-2 *Sudden ionospheric disturbance receiver*

output of C1 to ground at the input of the first op-amp stage U2:B. Integrated circuit sections U2:A and U2:B form an amplifier/filter and tuning. The "tuning" is controlled via R6, which is placed in a shielded enclosure to prevent circuit oscillation. The potentiometer is kept separate in a small shielded box formed by some scrap pieces of thin circuit board soldered together. The output from U2:A is next fed to IC:D and U2:C which forms an amplifier/integrator section. A final buffer amplifier section at U1 is used to drive the 0-1 mA meter at the output. The output of the SIDs receiver at (A) can be coupled a low cost data-logger in order to record signals over time.

The SIDs receiver was built on a circuit board circuit board, for most reliable operation, but other types of RF building techniques could be used. The circuit board is available from FAR Circuits—see Appendix. Before we begin building the circuit take a few minutes to locate a clean well lit work table or work bench. You will want to locate a small 27 to 33 watt pencil-tipped soldering iron, a roll of 60/40 tin/lead rosin core solder, and a small jar of "Tip Tinner," a soldering iron tip cleaner/dresser. "Tip Tinner" is available at a local Radio Shack store. You will also need some small hand tools, such as a small flat-blade screwdriver, a Phillips screwdriver, a pair of small needle-nose pliers, a pair of tweezers, a magnifying glass and a pair of end-cutters.

Once you have gathered all the tools, locate all the circuit diagrams, such as the schematics, the parts layout diagram, the resistor and capacitor code charts, etc. Place all the component parts, along with all your tools and diagrams, in front of you, warm up your solder iron and we will start building the SIDs receiver project.

First we will identify the resistors for the project. Refer to the chart shown in Table 19-1, which will help you to identify the resistor color codes. Resistors usually have three or four color bands on the body of the resistor which denotes the resistor value. The color bands begin at one end of the resistor body close to the edge. The first color band on resistor R1 is orange, this denotes the digit three (3). The second color band is also orange, which is the digit number three (3), and the third color band is the resistor's multiplier value which is red, and this suggests a multiplier value of 100. So the resistor value for R1 is (3) times (3) times (100) or 3.3k or 3300 ohms. The tolerance value of a resistor is noted in the four color band. A fourth color band which is silver is a 10% tolerance resistor, while a gold fourth band denotes a 5% tolerance value. If there is no fourth color band then the resistor will have a 20% tolerance value. Once you have identified resistor R1, you will note that resistors R1, R2, R3 and R4 all have the same color codes, therefore they all have the same value. Check your parts layout diagram and the schematic to

Table 19-1
Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

see where these resistors are to be placed on the circuit board. Once you verify where the resistors are placed, you can install all four resistors on the circuit board and solder them in place on the board. Now locate your end-cutter and trim the excess component leads. Cut the extra component lead lengths flush to the edge of the circuit board. Now go ahead and identify another group of three or four resistors and place them on the circuit board. Once you are sure of their respective locations you can solder the next grouping of resistors onto the circuit board, remember to trim the excess component leads from the board. Place all the remaining resistors on the circuit board and solder them all in place, trim the extra leads as necessary.

Now we will move on to installing the capacitors for the SIDs receiver project. Capacitors come in two major flavors, non-polarized and polarized types. First, we will talk about non-polarized capacitors. Usually non-polarized capacitors are smaller value types which could take form as ceramic disks, mica, tantalum or polyester types. Often the capacitors are small in size and frequently the capacitor values are printed as three-digit codes rather than their actual values, since this code system takes less space to print. We have included a chart in Table 19-2, which lists the actual capacitor values and their respective codes. For example, a capacitor with a value of .001 µF will have a code designation of 102, a .01 µF capacitor will have 103, and a 0.1 µF capacitor will have a code of 104 marked on it. The SIDs project has a number of non-polarized type capacitors which you will need to install. Take a look through the component pile and see if you can identify the small non-polar capacitors before attempting to install them on the PC board.

Polarized capacitors are called electrolytic capacitors, and they are often larger in physical size. Electrolytic capacitors will always have some sort of polarity marking on them. You will see either a black or white band or a plus (+) or minus (–) marking somewhere on the capacitor body. It is very important that you install these types of capacitors with respect to their polarity in order for the circuit to work properly and to not be damaged upon power-up. You will need to refer to the schematic and parts layout diagrams when installing electrolytic capacitors to make sure that you have installed them correctly. The plus (+) marking on the capacitor will point to the most positive portion of the circuit.

Let's go ahead and identify the non-polarized capacitors first. Check the capacitor code chart and make sure that you can identify each of the non-polarized capacitors. In groups of two or three capacitors install them on the PC board. When you have determined that you have installed the correct values into their respective location on the circuit board, you can solder the capacitors in place on the board. Remember to trim the excess capacitor leads from the circuit board. Once you have the first grouping of capacitors installed you can move on to installing another group of non-polarized capacitors, until you have installed all of the non-polarized capacitors.

Next, find the electrolytic capacitors from the parts pile and we will install them in their proper locations. Refer now to the schematic and parts layout diagrams so that you can orient the capacitors correctly when mounting them on the PC board. Install two or three electrolytic capacitors and then solder them in place on the PC board. Remember to cut the extra lead lengths flush to the PC board. Identify and install the remaining electrolytic capacitors on the board and solder them in place. Trim the component leads as necessary.

The SIDs receiver employs two silicon diodes at the "front-end" and two germanium diodes near the final output stage. Diodes have polarity so they must be installed in the proper orientation for the circuit to function properly. Diodes will generally have either a white or black band on the side of the diode's body. The colored band denotes the cathode of the diode. Remember that a symbol for a diode is a triangle pointing to a vertical line. The vertical line is the cathode of the diode. Make sure you install the proper diodes in their proper locations with respect to polarity: you will need to refer to the schematic and parts layout diagrams.

The SIDs receiver uses two integrated circuits at U1 and U2. Integrated circuit sockets are highly recommended as an insurance policy against any possible failure in the distant future. Before we go ahead and install the semiconductors, take a quick look at the semiconductor pin-out diagram shown in Figure 19-3. The pin-out diagram will assist you install the integrated circuits. IC sockets are inexpensive and will greatly be appreciated in the event of a circuit failure, since most people cannot un-solder a 14 or 16 pin integrated circuit

Table 19-2

Capacitance code information

This table provides the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Most capacitors actually have the numeric values stamped on them; however, some are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

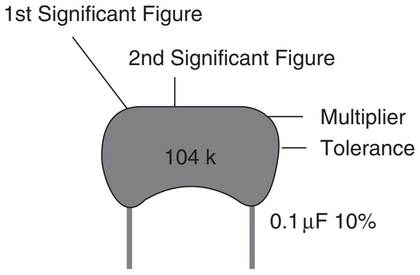

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

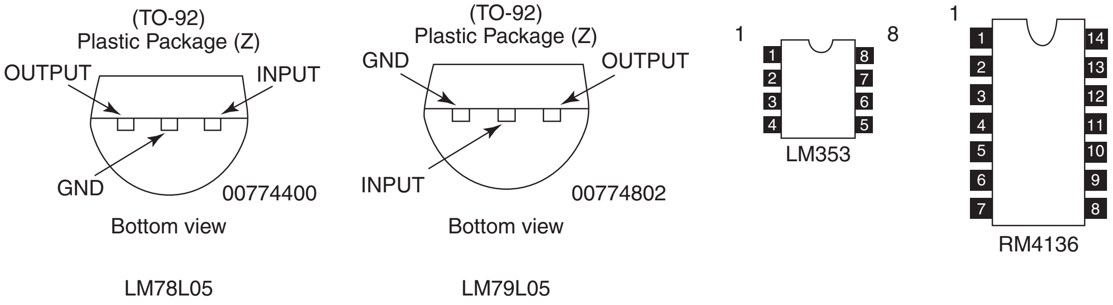

Figure 19-3 *Semiconductor pin-outs*

without damage to the circuit board. Integrated circuits must be installed correctly for the circuit to work properly. Looking at the IC, you will notice a small notch or cut-out or small indented circle at one end of the IC body. Pin one (1) of the IC will be just to the left of this cut-out or notch. The IC socket will often appear the same way with a notch or cut-out with pin one (1) to the left of the notch or cut-out. You must orient pin of the IC with pin 1 of the IC socket, and you must them be sure to orient pin of the IC socket to match up with the proper components on the circuit board. For example, pin one (1) of U1 is connected to the output of the circuit at the junction of R8 and R11. Note that pin 8 is the plus power supply input, while pin 4 is the ground on U1. Pin one (1) of U2 corresponds to the minus (–) INPUT #1 at U2:A, which is connected to the junction of R1 and R2.

A small metal compartment should be placed around the main tuning potentiometer R6 in order to shielded this control. This can be done with the use of some scrap circuit board material which can be soldered together to form a small box. You will have to use a soldering gun or a higher temperature soldering iron, in order to solder large areas of circuit board material.

The output of the SIDs receiver is fed to the 0-1 mA analog meter at M1. The output of the SIDs monitor can also be fed to the input of an analog to digital converter place in a personal computer, for recording over long lengths of time.

Since the SIDs receiver utilizes integrated circuits which require the use of a dual power supply providing both a plus and minus voltage to the circuit, and since the circuit is meant to be left on for long periods of time the use of a dual voltage AC power supply is recommended. The diagram shown in Figure 19-4 illustrates a simple dual plus and minus voltage power supply which can be used to power the SIDs receiver. A 9 volt 500 mA center tapped transformer was used to drive a bridge rectifier, which provides both a plus and minus voltage that is sent to a 9 volt plus regulator. An LM78L09 regulator and the minus leg of the bridge is sent to an LM79L09 minus 9 volt regulator. The power supply could be built on perf-board or on a printed circuit board if desired. The power supply and SIDs receiver could be mounted alongside one another in the same chassis.

Take a short break and when we return we will inspect the circuit board for any possible “cold” solder joints or “short” circuits. Pick up the circuit board so that the foil side of the PC board faces upwards toward you. Look over the foil side of the board very carefully. The solder joints should all look clean, bright and shiny. If you see any solder joints that look dull, dark, dirty or “blobby” then you should remove the solder and re-solder the joint so that it looks clean, bright and shiny. Next we are going to inspect the PC board for any possible “short” circuits. Sometimes “cut” component leads will “stick” to the board from solder residue left on the board. These “stray” wires can often cause a “short” between the solder traces. Also look closely for any solder “blob” or “solder balls,” which may have stuck to the underside of the PC board.

The SIDs receiver and power supply can both be installed in a metal chassis box enclosure. The prototype receiver was mounted in a sloping cabinet as in Figure 19-1. The power supply was mounted alongside

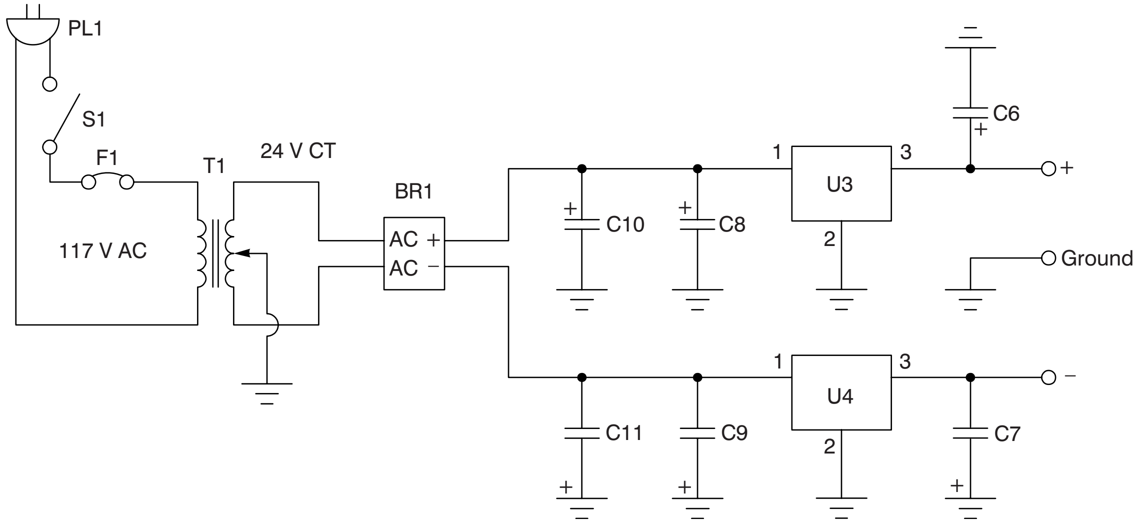

Figure 19-4 *Dual voltage power supply*

the SIDs receiver. The tuning control as well as the power on-off switch and meter were all mounted on the front panel of the enclosure. The power cord was led out of the rear panel through a stain-relief. A power fuse was installed inside a chassis mounted fuse holder on the rear of the panel. A dual binding post/banana chassis jack was mounted on the top of the chassis to connect the antenna loop to the SIDs receiver. An RCA jack was placed on the rear of the chassis to allow the receiver to connect to an analog to digital converter.

The SIDs receiver requires a rather large loop antenna, which can be placed inside your attic or placed outside away from large metal structures or aluminum siding, in a sheltered area. The coil at L1 is a diamond shaped loop antenna as shown in Figure 19-5. The loop consisting of 50 turns of solid #24 enameled or plastic coated wire, on a PVC pipe square or diamond form. The loop has an enclosed area of about 9 square feet. Locate four 2½″ diameter × 24″ long pieces of PVC pipe and form a square or diamond shape coil form as shown. A PVC elbow was placed on top and both sides of the loop form as shown, while a PVC "T" was used at the bottom of the loop form to allow winding of the coil around the four protrusions at each leg of the loop. The four hole center hub holds the loop form together. The main vertical PVC pipe was secured into a flange which was fastened to a wooden block at the base of the loop antenna. The lead-in from the loop antenna should form a loose twisted pair. This balanced lead-in is routed as far away from metal as possible to the 600 to 600 ohm matching transformer at T1. This transformer arrangement reduces 60 Hz hum interference from entering the SIDs receiver.

The output of the SIDs receiver can be coupled to a solid state data-logger from the output of the receiver at point (A) at the receiver output terminals. In order to save and later view and correlate your recorded data you will need to acquire some form of data-logger. There are three good options for saving your data. The first option is the Onset Computer HOBO series of data-loggers. Onset offers a number of different models from 8-bit to 12-bit models which are reasonably priced from around $60.00. The data-loggers are powered by a small button battery which will last for a long time. Onset also offers low cost software for up/downloading information to the data-logger. Starting times and dates can be preset as well as voltage parameters and timing between samples. These are a great alternative. Another data-logger option is the new introductory "starter package" from Dataq DI-194RS 10-bit resolution PC data-logger kit for $24.95. This data-logger option is a real bargain, providing both hardware and software and it can get you started recording data in just a short time. The company offers many other data recorder options as well. Check their website for details http://www.dataq.com.

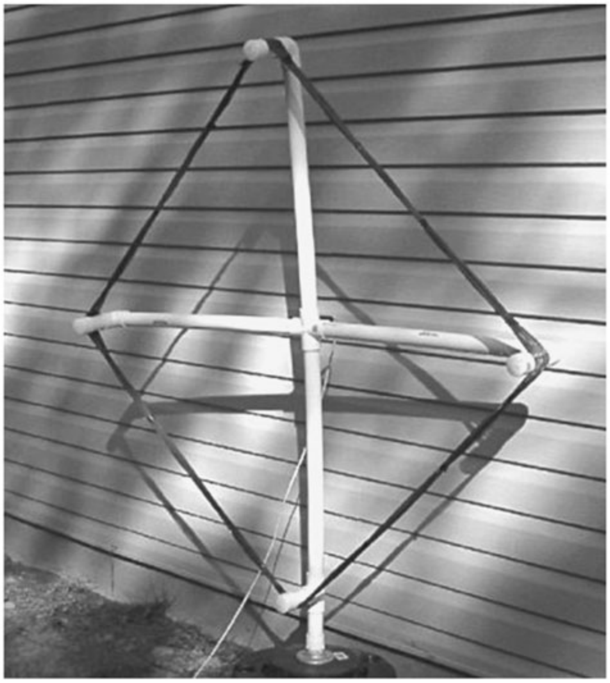

Figure 19-5 *Loop antenna*

In normal operation, the SIDs receiver would be permanently tuned to a powerful VLF station such as (NAA), the powerful US Navy 24 kHz VLF transmitter station in the state of Maine, and your data-logger or A/D converter would constantly monitor the signal from NAA to record propagation changes on your personal computer for study. Once the receiver has been completed, you can attach the antenna, power supply and data recorder. Connect the output of the data recorder to your personal computer and you are ready to begin recording. Now turn on the receiver and you should see a signal on your computer chart recorder display. If you remove the antenna from the circuit the output should go to zero or base-line. If your SIDs receiver does not appear to work, then you will need to disconnect the power supply and data recorder and antenna and inspect the circuit board for possible errors. The most common cause for circuit failure after building a new circuit is the improper installation of components, such as electrolytic capacitors, diodes and semiconductors such as transistors or integrated circuits. You may also want to check the placement of resistors on the circuit board. Ask a knowledgeable electronics enthusiast to help you inspect the circuit board for errors. Having completed the inspection, and hopefully an error was found, you can reconnect the antenna and power supply and then re-test the receiver.

SIDs research opportunities

You can join the foremost group involved with SIDs research. The AAVSO or American Association of Variable Start Observers SID Program consists of solar observers who monitor very low frequency (VLF) radio stations for sudden enhancements of their signals. Earth's ionosphere reacts to the intense X-ray and ultraviolet radiation released during a solar flare. The ionospheric disturbance enhances VLF radio propagation. By monitoring the signal strength of a distant VLF transmitter, sudden ionospheric disturbances (SIDs) are recorded and indicate a recent solar flare event. All SID monitoring stations are home built by the observers. The SID station operates unattended until the end of each month. Recordings are then analyzed for the beginning, end, and duration of SID events. A simple A/D converter design for specific use with the VLF receivers is available by contacting the chairman. SID observers submit strip-charts or computer plots to the SID coordinator for visual inspection at the end of each month. Many observers analyze their own stripcharts and computer plots. Analyzed results are submitted via e-mail to the SID Analyst for correlation with other observers' results. The final SID report combines individual observers' reports with the SID Coordinator visual analysis. SID event results are sent monthly to the National Geophysical Data Center (NGDC) for publication in the Solar-Geophysical Data Report where they are accessed by researchers worldwide. The reduced SIDs data and particularly interesting plots are reproduced in the monthly AAVSO Solar Bulletin mailed to all contributing members.

Chapter 20

Aurora Monitor Project

Parts list

Aurora Monitor: Sensor Head Unit

R1,R2	1.5 megohm ¼ watt, 5% resistor
R3	5.6 megohm, ¼ watt, 5% resistor
R4	100 k potentiometer PCB trimmer
R5	10 megohm ¼ watt, 5% resistor
C1,C2,C3	.1 µF, 35 volt disk capacitor
D1	1N914 silicon diode
LM4250	low noise/high gain op-amp
L1	sensing coil, see text
B1,B2	1.5 volt “C” cell battery
J1	RCA input jack
Misc	PC board, coil form, wire, coax, etc.

Control/Display Unit

R6	18 k ohms, ¼ watt, 5% resistor
R7	12k ohms, ¼ watt, 5% resistor
R8,R9,R12,R20	10k ohms, ¼ watt, 5% resistor
R10	30k ohms, ¼ watt, 5% resistor
R11	50k ohms, panel-mount potentiometer
R13	200k ohms, ¼ watt, 5% resistor
R14	3 megohm, ¼ watt, 5% resistor
R15	1 megohm, PCB trimmer potentiometer
R16	1 megohm, ¼ watt, 5% resistor
R17	1 megohm, panel-mount potentiometer
R18,R19	5k ohm panel-mount potentiometer
R21	1k ohm, ¼ watt, 5% resistor
R22	100 ohms, ½ W, 5% resistor
R23	4.7k ohms, ¼ watt, 5% resistor
C4	0.005 µF, 25 volt ceramic disk capacitor
C5,C7,C9,C10	0.1 µF, 25 volt ceramic disk capacitor
C6	0.02 µF, 25 volt ceramic disk capacitor
C8	4.7 µF, 25 volt aluminum electrolytic capacitor
C11	.010 µF, 25 volt disk capacitor (optional)
C12	1 µF, 35 volt

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

	aluminum electrolytic capacitor
D1,D2,D3,D4	1N914 silicon diodes
Q1	2N2907 PNP transistor
Q2,Q3	2N394, NPN transistor
U1	LM4250 low noise op-amp (National Semiconductor),
U2	LTC1062 (Linear Technology)
U3	LM201A dual op-amp (National Semiconductor)
U4	LM747 op-amp (National Semiconductor)
S1	DPDT toggle power switch
S2,S4	SPST toggle switches
S3	3-position rotary switch, break before make
M1	100-micro-ampere moving-coil panel-mounted meter
SPKR1	8 ohm speaker, 2″ dia.
B1,B2,B3,B4	1.5v (AA) or (C) cells (+) plus voltage supply
B5,B6,B7,B8	1.5v (AA) or (C) cells (-) minus voltage supply
J2,J3	RCA jack
Misc	PC board, chassis, wire, coax, hardware, standoffs, RTV compound, washers, etc.

The Aurora Borealis is one of nature's most spectacular nighttime displays. Shimmering curtains of green, white, and even red light dance in the northern skies. Visible effects of charged particles from the Sun raining down on the Earth's ionosphere, northern lights or auroras, are visible in the northern night sky during high sunspot activity. The Aurora Australis, the southern hemisphere's counterpart of the Aurora Borealis, can be seen at night by looking toward the south pole.

These displays of undulating light are formed when flares from the Sun's surface (sunspots) launch showers of high-energy ionized particles and X-rays into space. Mostly electrons, the showers stream out from the sun and are attracted by the Earth's magnetosphere, an invisible magnetic field around the Earth.

Shaped like a pumpkin, the magnetosphere terminates at both magnetic poles but is many miles thick above the equator. Dimples at both poles form "sinks" that funnel the particles toward the poles where they ionize the gas in the ionosphere. Those collisions induce the gases to emit their characteristic light wavelengths—as in neon signs and fluorescent lamps.

The charged particle bombardment of the magnetosphere initially compresses it, temporarily increasing the strength of the Earth's geomagnetic field. The aurora monitor described here is sensitive enough to detect changes in the field caused by those "magnetic storms." Thus it can indirectly sense sunspots and predict the presence of auroras in the night sky.

The monitor also senses changes or anomalies in the magnetic field caused by large metal objects such as cars or trucks moving near the monitor. This permits it to act as an intrusion detection monitor able to detect the approach of vehicles at night in restricted areas. The monitor can also detect the presence of permanent magnets (such as those in speakers), and stray fields from AC-power lines.

Early warning of auroras will both permit you to observe them in the night sky or use them for boosting the range of your amateur radio transmissions. Auroras and their accompanying magnetic storms generally block or scramble the lower radio frequencies, but the higher frequencies can overcome this interference. Radio amateurs aim their antennas north during those storms, thus taking advantage of the phenomena to reach other hams on the opposite side of the Earth that

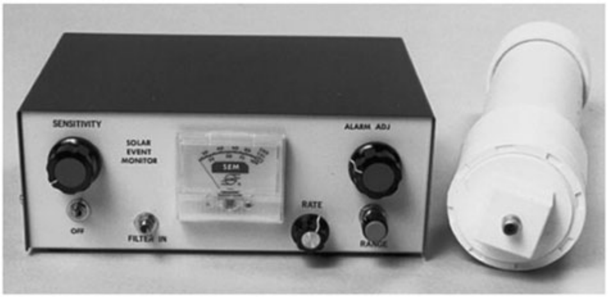

Figure 20-1 *Aurora monitor*

could not be contacted during periods of low sunspot activity.

In addition to scrambling low-frequency radio communications, the magnetic storms caused by auroras can induce large currents in power transmission lines. Those currents can cause overload, plunging large regions of the country into darkness. Auroras and related magnetic storms are quite common during the decreasing parts of the 11-year sunspot cycle such as the period we are now in.

The Aurora Monitor consists of two main components, a sensing head and the control/display unit which are connected by a coaxial cable. The sensing head contains a sensing coil, a DC nano amplifier capable of current amplification of 500, and a separate power pack, all enclosed in the tubular case shown on the right side of Figure 20-1. Figure 20-2 illustrates the interior of the Aurora Monitor prototype. The Aurora Monitor is sensitive to a pulse of one ampere at distance of one meter, which corresponds to one milligauss. (The Earth's magnetic field is about 0.5 ga.)

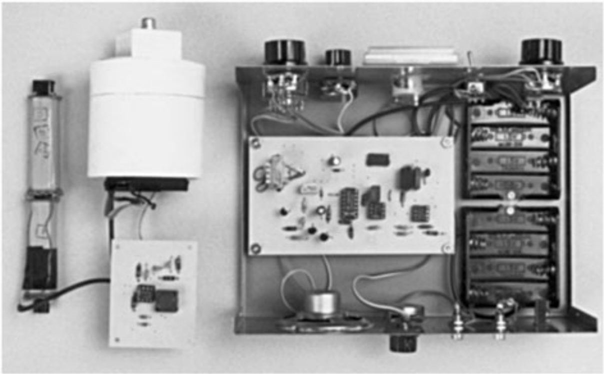

Figure 20-2 *Aurora monitor console and sensor*

The control/display unit contains an active filter, additional amplification circuitry, a moving-coil ammeter, and an audio oscillator with speaker. Figure 20-1 shows the front-panel controls and indicators of the monitor: moving-coil ammeter, SENSITIVITY, RATE, and ALARM ADJUST potentiometers as well as power (OFF), filter-in and RANGE switches.

The sensor head unit

The sensor head schematic circuit is illustrated in Figure 20-3. Gain is provided by U1, a National Semiconductor LM4250 programmable operational amplifier. It is protected from over voltage and transients by diode D1 and D2, and its overall gain is set by resistors R1 and R2. The output of U1 is driven to zero or balanced by network R3 and R4. Its output should remain at zero as long as no changes occur in the ambient magnetic field.

Bypass capacitors C2 and C3 are placed across the positive and negative power supply. Power for the sensing amplifier is obtained from two "C" cells. The sensor circuit draws very little current, so it can be left on at all times, thus eliminating the power switch.

The control/display unit

The control/display circuit schematic is depicted in the schematic in Figure 20-4. It provides an additional gain of 200 over that of the sensor circuit. The control/display circuit includes an adjustable low-pass notch filter, U2, a Linear Technology LTC 1062. By adjusting the clock frequency of the filter with resistor RI1 and the capacitors C4, C5 and C6, the filter cancels interference frequencies and noise in the 2 Hz to 10 kHz band. The notch filter can also screen out 60 Hz noise. Switch S2 inserts or removes the filter. U3, an LM201A general purpose op-amp, filters out the clock noise generated within the filter chip. The output of U3 is fed into the non-inverting input of U4-A, half of a dual 747 general purpose op-amp. The overall gain of U4-A is adjusted by resistor R16. Trimmer potentiometer R15

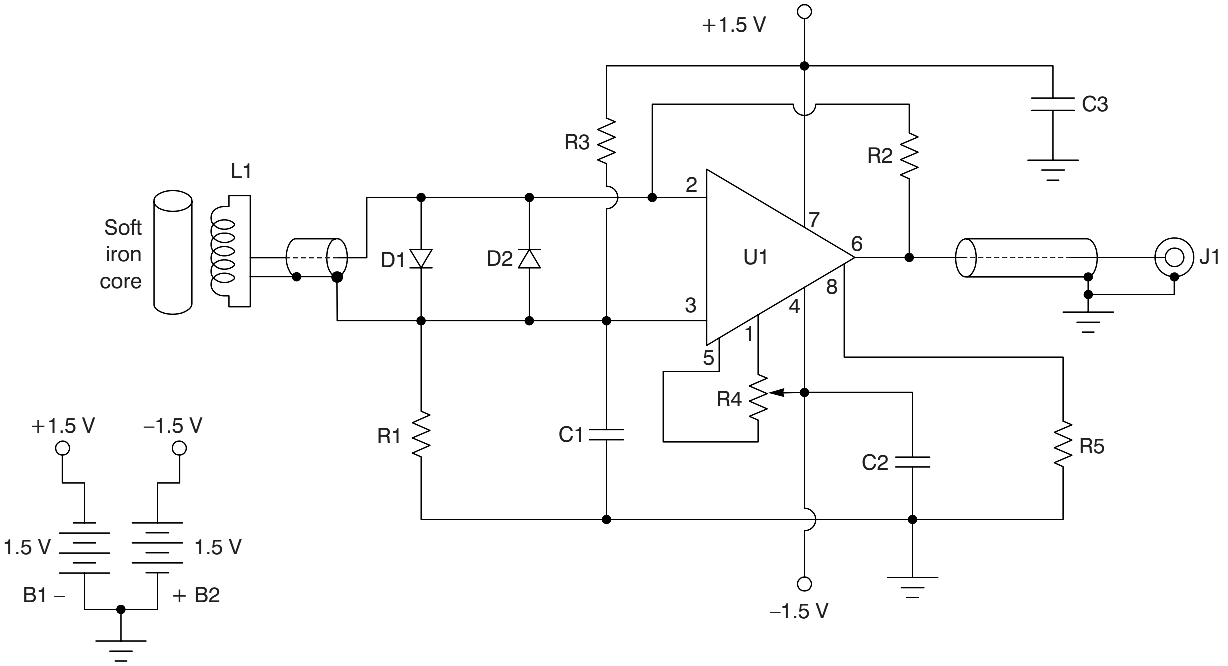

Figure 20-3 *Aurora monitor sensor head*

adjusts the offset or balance. PULSE SHAPE potentiometer R17 and capacitor C5 form an optional pulse-shaping network for coupling the Aurora Monitor to a chart recorder or an analog-to-digital conversion board of a personal computer.

The output of U4-A is coupled to a voltage follower at U4-B. The output of U4-B is divided into two channels. One channel is fed through 5000-ohm SENSITIVITY potentiometer R18, which adjusts the output level of the signal fed to the 100-microampere panel meter MI. Movement of the meter's needle shows changes in the local magnetic field. Potentiometer R18 also adjusts the output signal that can be fed to a chart recorder for data-logging.

The other channel is fed through ALARM ADJUST 5000-ohm potentiometer, R19, which sets the threshold or set-point for the reflex oscillator circuit that follows it. The oscillator consists of transistors Q1, Q2, and Q3 and associated components. Speaker SPKR1 gives an audible indication of changes in the local magnetic field. The network of diode D3 and aluminum electrolytic capacitor C8 performs additional filtering for the input signal to the reflex oscillator section.

Transistor Q1 controls the audible alarm by clamping the negative voltage returning through the ground path. When a magnetic event occurs, the speaker emits an audible alarm, and the meter gives a visual indication of a changing magnetic field. The adjustment of ALARM ADJUST potentiometer R24 can remove distortion from the sound of the speaker.

The author's prototype control/display unit is powered by eight AA cells: four cells provide positive voltage and four cells provide negative voltage. As an alternative, the monitor can be powered by rechargeable nickel-cadmium cells.

Construction

In order to begin constructing the Aurora Monitor, you will need to secure a clean will lit work table or work bench so you can spread out all the tools, project components, charts and diagrams. You will need to locate a small 27 to 33 watt pencil-tipped soldering iron, a small roll of 60/40 rosin core solder, a small jar of "Tip Tinner," soldering iron tip cleaner/dresser, from your local Radio Shack store. A few small hand tools

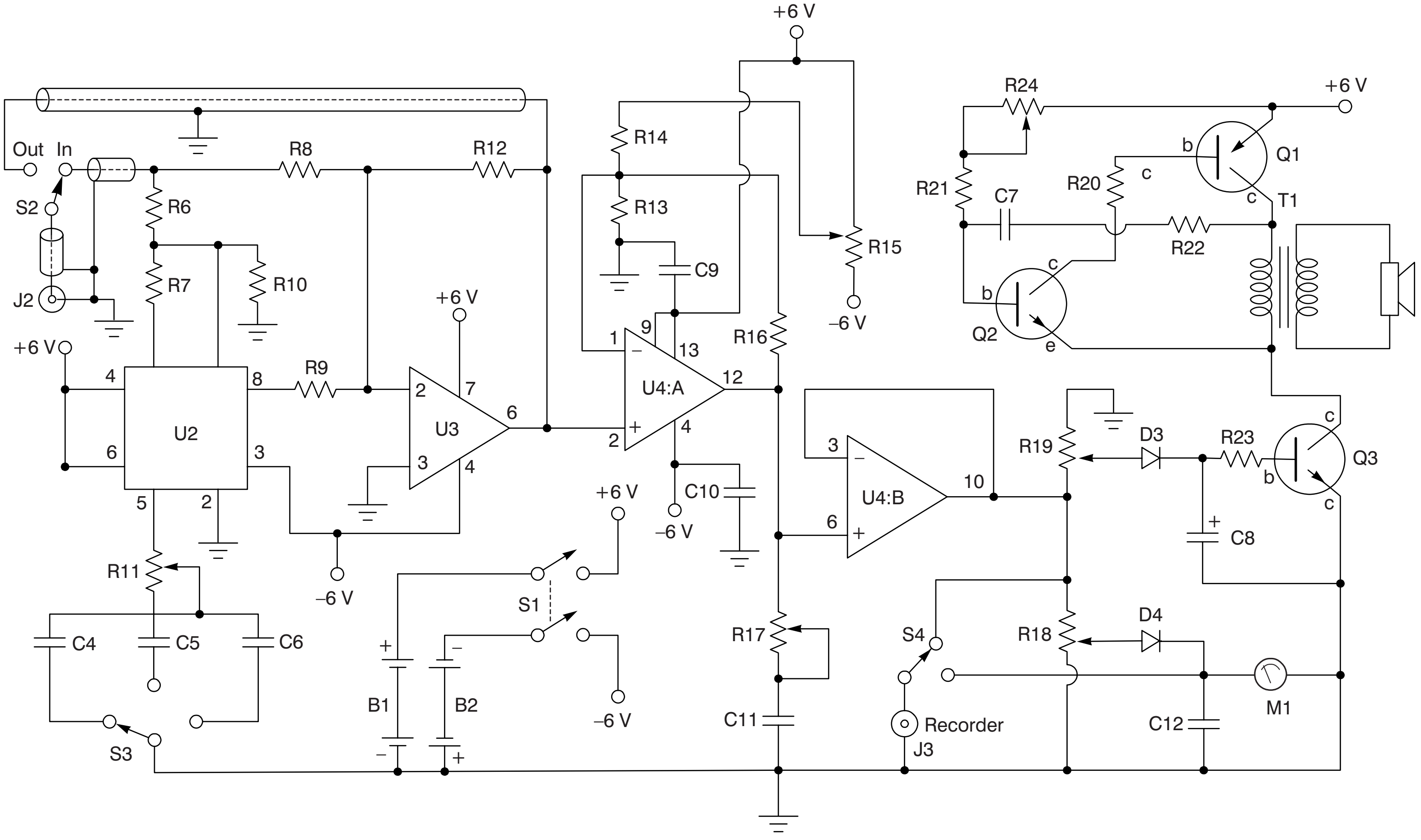
+6 V
R24
+6 V
Out
In
R8
R12
R14
b
Q1
S2
R6
R21
C7
R20
c
c
T1
R13
R15
R22
R7
R10
C9
J2
+6 V
b
c
–6 V
R16
Q2
e
1
9
13
+6 V
R9
4
8
2
7
U4:A
12
U2
U3
6
2
+
4
c
6
3
3
10
R19
D3
R23
Q3
b
4
C10
U4:B
c
5
2
R11
–6 V
+6 V
–6 V
6
+
C8
–6 V
S1
D4
C4
C5
C6
R17
S4
R18
+
B1
B2
–6 V
C11
Recorder
J3
C12
M1
S3
+

Figure 20-4 *Aurora monitor display/control unit*

are also required to construct the project. You should locate a small pair or end-cutters, a pair of needle-nose pliers, a pair of tweezers, a magnifying glass a small flat-blade screwdriver and a small Phillips screwdriver.

The Aurora Monitor project was constructed on two circuit boards, one for the sensor head unit and a second PC board for the control display unit. We will begin constructing the sensor head unit first, see the schematic in Figure 20-3. The sensor head board measured 2 $\frac{1}{16}$ by 2 $\frac{9}{16}$, it was housed along with the sensor head coil in a suitable PVC pipe housing.

Place the project components in front of you on the table, locate the resistor and capacitor identifier charts in Tables 20-1 and 20-2, heat up your soldering iron and we will begin. First refer to the resistor identifier chart, it illustrates the resistor color codes and their representative values. Most resistors will have three or four color codes on the body of the resistor. The color codes begin at one end of the resistor's body. The first color band denotes the first digit of the resistor value, while the second color band represents the second digit of the resistor code. The third color band on the resistor denotes the multiplier value of the resistor, and the fourth color code represents the tolerance of the resistor. A silver band notes a 10% tolerance value, while a gold band denotes a 5% resistor, and the absence of a fourth band implies a 20% tolerance for the resistor. For example, resistor R1 has a 1.5 megohm value. The first color band would be brown or one (1), the second color band would be green or five (5), and the third color or multiplier would be green or (100,000). Go ahead and locate resistors R1 and R2 which are used in the sensor head unit and place them on the sensor head PC board in their respective locations. Solder them in place and then trim the excess component leads with your end-cutter. Cut the excess leads flush to the edge of the circuit board. Next locate the remaining resistors on the sensor head unit and populate the board. Solder the resistors to the board and trim the excess leads.

Next we are going to install the capacitors on the sensor head unit of the Aurora monitor. Capacitors are listed as electrolytic or non-electrolytic types. Non-electrolytic types are usually small capacitors, often they are so small that their actual value is not printed on the body of the capacitor but a three-digit code is used, see Table 20-2. For example, a .01 µF capacitor would be marked (103) while a .1 µF capacitor would be marked (104). Locate capacitors C1, C2 and C3 which are .1 µf values. Place them on the sensor head PC board and solder them in place on the board, remember to trim the excess component leads with your end-cutters. There are no electrolytic type capacitors on the sensor head unit but there are some on the control/display PC board. Electrolytic capacitors are usually larger in size and value, and usually they will have a polarity marking of some kind on the body

Table 20-1
Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

Table 20-2

Capacitor code identification information

This table is designed to provide the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Most capacitors actually have the numeric values stamped on them; however, some are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

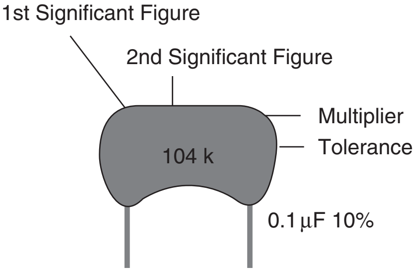

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

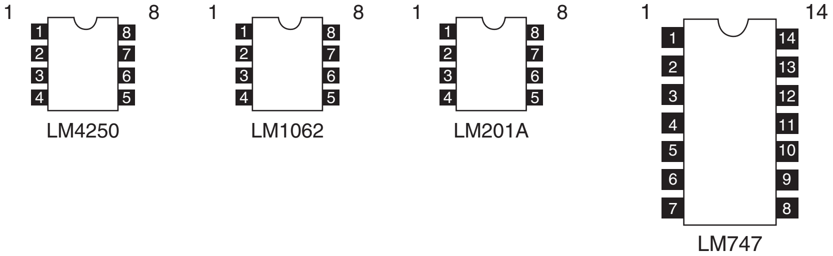

Figure 20-5 *Semiconductor pin-outs*

of the capacitor. You will find either a plus (+) or minus (−) marking near one of the leads on the capacitor body. Both the sensor head and control/display board utilize silicon diodes. Diodes also have polarity, so they must be installed correctly in order for the circuit to work properly. Look closely at a diode and you will see either a white or black color band near one end of the diode body. The colored band indicates the cathode end of the diode. Go ahead and install D1 on the sensor head board.

The sensor head and the display boards both have integrated circuits on them, so before we begin installing the ICs, take a quick look at the pin-out diagram depicted in Figure 20-5, which will help you orient the integrated circuits. Integrated circuits are highly recommended as an insurance against possible circuit failure at some later date. It is much easier to simply unplug an IC and place a new one in the socket rather than trying to un-solder a 14 or 16 pin IC from a circuit board. Integrated circuit sockets will usually have a notch or cut-out at one end of the plastic IC socket. To the immediate left of the notch or cut-out you will see pin one (1) of the socket. Pin (1) of U1 is connected to one end of potentiometer R4. The IC package itself will also have a notch or cut-out at the top end of the IC. When placing the IC into its respective socket make sure you align pin one (1) of the IC with pin one (1) of the socket. The integrated circuits must be inserted into the socket properly aligned or the circuit will not work and may damage the components. Go ahead and place U1 into its socket on the sensor head PC board.

Sensor coil assembly

The sensor coil at L1 detects changes in the local magnetic field. You will need to wind approximately 50,000 turns of 28 AWG magnet wire over a soft iron core ½″ in diameter and 10″. (The iron core concentrates the flux lines by offering a lower reluctance path than air.)

Wind the fine insulated magnet wire carefully on the iron core to avoid kinks and breakage. Tape the ends of the winding temporarily to the core and carefully solder hook-up wire at each end to form permanent terminals. The terminals can be secured to the core with room-temperature vulcanizing (RTV) adhesive to relieve any strains that might develop in the fine magnet wire.

In the prototype, the sensing circuit board, coil, and battery pack, consisting of two “C” cells were housed in a case made from standard diameter PVC water pipe cut to a length that will accommodate all of those elements as shown in Figures 20-6.

The covers for the sensing head housing are PCV end-caps that press-fit over the 2⅜″ outside diameter of the pipe. The upper cap is a simple cup, but the lower cap is a sleeve with a threaded insert at its end. Drill a hole in the square base of the threaded insert for jack J1 and fasten it with a ring nut. Then close the cover on the empty pipe and drill two pilot holes 180° apart in the sleeve for self-tapping screws to clamp the cap in position after the sensor head is assembled.

Cut about a 6″ length of RG-174/U coaxial cable, strip both ends and solder the inner conductor of one end to the jack terminal and its shield to the jack lug. Solder the inner conductor and shield of the other end to the sensor circuit board.

Attach the coil to the sensor circuit board with about a 6″ length of RG-174/U coaxial cable as shown. Connect the inner conductor to one terminal and the shield to the other and solder both in position. The prototype includes a twin “C” cell holder that, with cells in position, has a maximum width dimension of

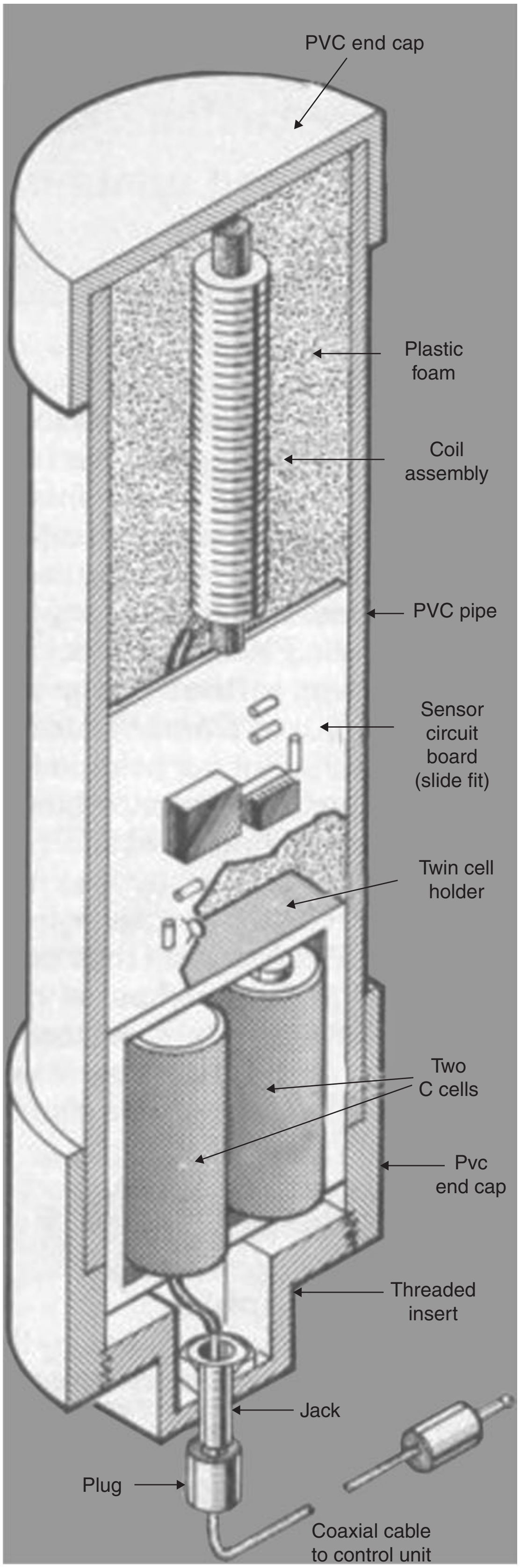

Figure 20-6 *Sensor*

less than 2″ permitting it to slide easily into the sensor head pipe section.

Cut three lengths of hook-up wire to extend the lengths of the positive, negative and ground leads of the battery pack. Solder one end of those wires to the holder leads and the other ends to the sensor circuit board.

Bond the upper cap to the PVC pipe section with PVC cement. Then position U1 in its socket, making certain that pin 1 is located correctly. Finally, insert the "C" cells in the holder. Wrap a layer of polyurethane foam around coil L1 to center it within the pipe section and insert both in the pipe section with the cover at the end. Then assemble the completed circuit board and battery pack to make sure that all of these parts will fit in the pipe section. Set the assembly aside until you are ready to perform the checkout steps.

Control/display unit

The control/display prototype circuit board measured 2⅞″ × 5″ to fit inside the instrument case selected. The schematic for the control/display board is shown in Figure 20-4. The dimensions of the circuit boards can be modified for packaging in the instrument and sensor head cases of your choice. Place the control/display circuit board in front of you, along with all its respective components, and we will begin constructing the control/display board.

First we will begin by placing the resistors on the PC board; locate R6, an 18k ohm resistor, and R7 a 12k ohm resistor, and place them on the board and solder them in place in their respective locations on the board. Next locate and install resistors R8, R9, R12, and R20; these are all 10k ohm resistors, solder them in place and cut the excess component leads with your end-cutter. Locate and install the remaining resistors, solder them in place and trim the excess component leads afterward.

Next, you can locate the capacitors C4, a .005 µF disk capacitor, and it may be marked (502). Install C4 on the PC board and solder it in place, remember to cut the excess lead lengths. Now, locate capacitors C5, C7, C9 and C10; these are all .1 µF capacitors and may be marked (104). Install these capacitors and solder them in place and then trim the extra leads. Now, locate C8, a 4.7 µF electrolytic capacitor. Capacitor C8 has polarity, so it must be installed with respect to these marking. Look at the schematic, and you will see that the plus (+) side of C8 is connected to the junction of resistor R23 and diode D3. Install C8, and then solder it place, trim the excess leads as necessary. Go ahead and install the remaining capacitors and solder them in place on the PC board.

The control/display section of the Aurora Monitor employs two silicon diodes, and they must be installed with respect to their polarity. The colored band denotes the cathode end of the diode. The cathode end of D3 is fed to R23, while the cathode end of D4 is sent to C12 and M1.

Transistors are generally three-lead devices, with a Base lead, a Collector lead, and an Emitter lead. The Base lead is depicted as a straight vertical line with two arrows pointing to it. The Emitter lead will have a small arrow pointing towards or away from it. When the arrow points toward the Base lead the transistor is a PNP type. If the arrow is pointing away from the Base lead then the transistor is an NPN type. Transistor Q1 is a PNP type, while Q2 and Q3 are NPN types. Identify the transistor leads and install them in their respective locations on the PC board, trim the leads as necessary.

The control/display portion of the Aurora Monitor utilizes three integrated circuits. Integrated circuits are highly recommended in the event of a circuit failure at a later date. Be sure to install the ICs correctly: pin one (1) of U2 is connected to R7, while pin one (1) of U3 is not used, pin two (2) is jointed with R9 and R8 and R12. Integrated circuit U4 has pin one (1) connected to the junction of R13 and R14. Failure to install the integrated circuits correctly may result in damage to the IC or to other components as well.

Finally you will need to connect the small audio transformer to the circuit. The higher impedance (1K) end of the transformer is connected to the junction of Q1 and R22, and the other 1k lead is connected to the collector of Q3. The secondary or 8 ohm winding of T1 is sent to the small 8 ohm speaker. The transformer can be mounted on the circuit board or on the bottom of the chassis box. The 100 uA panel meter is wired between the Emitter of Q3 and ground.

The control/display assembly is housed in a standard 7″ × 5″ × 3″ aluminum electronics instrument case with a removable cover. Drill the holes in the front panel of the case for power ON switch S1, FILTER IN switch S2, RANGE switch S3 and (if used) optional recorder switch S4. Also drill the holes for SENSITIVITY, ALARM ADJUST, and RATE potentiometers (R18, R19, and R11 respectively). Cut out the hole for mounting meter M1. Note: If you want to use the monitor strictly as a security monitoring system the meter can be omitted.

Drill the holes in the back panel for jacks J2 and J3, and PULSE SHAPE potentiometer R17. Drill a series of holes in a circular pattern for the speaker SPKR1. Drill holes in the bottom of the case for mounting the control/display circuit board and two quad “C” cell holders. Remember the circuit uses both a plus (+) 6 volt supply and a minus (–) 6 volt supply. We chose to power the circuit from batteries, but you could use a dual 6 volt power supply to power the control/display unit.

Assemble the panel-mounted switches, jacks, potentiometers, meter and speaker to the front and back panels of the case as shown in Figure 20-2. (You might prefer to bond the face of the speaker to the inside of the back panel with RTV adhesive rather than bolts and nuts.)

Determine the length of speaker wires needed to permit assembly of the circuit board in the instrument case, cut the wires to length and solder them to the board. Cut and solder all leads from panel-mounted components to the circuit board, allowing sufficient lengths to permit unimpeded circuit board assembly, yet not allowing too much slack.

Cut about a 6″ length of RG-174/U coaxial cable to connect INPUT jack J2 and the FILTER IN Switch S2 and two other lengths about 6″ long to connect S2 to the FILTER-IN and FILTER-OUT pads on the control/display board. Strip all cable ends and solder and trim all connections.

Assemble the control/display board to the base of the case with screws and ½″ insulating standoffs, lock washers and nuts. Bolt the two quad AA cell holders to the base of the case. Position the ICs in their sockets on the control-display board, making sure that all pin 1’s are in their correct positions. Insert the eight AA cells in the spring-loaded holders, observing the correct polarity.

Cut a length of RG-174/U coaxial cable to the length that will suit your installation. Strip the wire ends and attach phono plugs PL1 and PL2 to cable ends and solder or crimp them in position.

Before applying power to the Aurora Monitor, you may want to inspect the two circuit boards for “cold” solder joints and “short” circuits. Look carefully at the foil side of the sensor head unit and examine the solder joints. The solder joints should all look clean, bright and shiny. If any of the solder joints look dark, dull, or “blobby” then remove the solder from the joint and re-solder the connection. Next examine the control/display board for “cold” solder joints. Now, we will inspect the two boards for possible “short” circuits. “Short” circuits are usually caused by “stray” component leads which get stuck to the board from the sticky residue left from rosin core solder or from small solder balls which may “short” across the circuit traces. Examine both circuit boards for “stray” leads and solder balls and remove them if you see any. Now, you are ready to apply power to the circuits in order to “test” the circuits.

Test and checkout

Test the sensing circuit first. Connect a general purpose oscilloscope or multi-meter to the output of U1. Position a permanent magnet near coil L1 and the oscilloscope display should show a pronounced pulse. If a multi-meter is used, its readout should jump.

Next, move L1 away from the magnet and the reading on the multi-meter should fall to zero. If the reading does not go to zero, adjust trimmer potentiometer R4 in the sensor circuit. When the sensing head is adjusted and working correctly, close the lower cap and fasten it with two non-magnetic stainless steel self-tapping screws to complete the assembly. Connect one end of the plug-terminated coaxial cable to the sensor head jack 1 and the other end to the control/ display unit jack J2.

Turn on the power toggle switch of the control/display unit and turn SENSITIVITY potentiometer R18 full clockwise. The meter should remain at zero. Rotate ALARM ADJUST potentiometer R19 clockwise until the alarm just begins to sound. Then turn it back slightly so that no sound is heard to obtain the maximum sensitivity setting. Next, turn potentiometer R19 full clockwise. There should be no sound from the speaker.

Position a small magnet or piece of metal next to coil L1 in the sensor head and the unit should now be activated: the speaker should emit sound, and the meter should read full scale. If everything checks out, you can now start observing magnetic field disturbances or anomalies.

In a quiet magnetic environment it might be necessary to adjust potentiometer R4 in the sensor head to the threshold of the meter movement. This fine adjustment eliminates any small dead zones in sensitivity. Test the instrument’s ability to detect the Earth’s magnetic field by rotating the sensing head with short, quick, snapping motions in a counter-clockwise direction. The meter movement should jump off scale.

As rotation is continued, a direction will be found where the meter will have its lowest response. This nulling point is the north-south direction. Any objects containing permanent magnets such as speakers or meters that are brought into close proximity to the Aurora Monitor’s sensing head coil L1 will affect the accuracy of the instrument’s readings.

In the event that the Aurora Monitor does not work upon power-up, you will need to disconnect the power supply and sensor head unit and inspect the circuit board for possible errors which might have occurred while building the Aurora Monitor. The most common cause for error when constructing the circuit is the incorrect installation of the electrolytic capacitors, the diodes and semiconductors such as transistors and integrated circuits. You may have installed one of these components backwards, so you will need to carefully inspect the circuit board. A second pair of eyes could be very helpful in locating potential errors of this sort. Have a knowledgeable electronic enthusiast help you inspect the circuit for errors. Another cause for error is mis-locating resistors. You may have placed the wrong value of resistor in a particular location, this would cause the circuit to malfunction. After carefully inspecting the Aurora Monitor circuit boards, you may quickly find your error and correct it. Once the error has been located, you reconnect the sensor head unit, and then re-apply power and then you can test the Aurora Monitor once again.

The Aurora Monitor can be connected to a chart recorder or it can pass signals to a personal computer with an analog-to-digital converter board. The recorder or PC can collect data for the study of magnetic fields, magnetic storms, and sunspot activity over long periods of time for further analysis. A set of high-impedance (greater than 1 kilohm) headphones can be plugged into jack J3 if you wish to “hear” the changes in magnetic fields.

If you want a permanent installation, mount the sensing head assembly so that it is directed away from any large metal obstructions, oriented on a north-south axis, and pointed slightly upwards. It’s a good idea to fasten it to a heavy wooden post to prevent wind-induced vibrations.

Chapter 21

Ultra-Low Frequency (ULF) Receiver

Parts list

ULF Receiver

R1,R4 50k ohm trim potentiometer

R2 50k ohm - 20 turn pot

R3 100k ohm trim pot

R5,R6,R10 10 k ohm, ¼w, 5% resistor

R7,R8,R9,R12,R21 33 k ohm, ¼w, 5% resistor

R11,R16 1 k ohm, ¼w, 5% resistor

R13,R20 22 k ohm, ¼w, 5% resistor

R14,R15,R18,R19 47 k ohm, ¼w, 5% resistor

R17 50 k ohm pot

R22 2.2k ohm, ¼w, 5% resistor

C1,C2,C3,C4,C5,C6 .1 µF, 35 volt disk capacitor

C11,C12,C13,C14 .1 µF, 35 volt disk capacitor

C8,C9 1 µF tantalum, 35 volt electrolytic capacitor

C10 .01 µF, 35 volt disk capacitor

C7 .05 µF, 35 volt disk capacitor

D1 Zener diode

U1,U2 LTC1063 IC

U3,U4 TL074 op-amp IC

U5 INA114 Differential amplifier IC

J1,J2,J3 3-terminal screw terminal strips

S1 SPST power switch

F1 1-amp fast-blow fuse

Misc PC board, chassis, wire, hardware, power cord, etc.

ULF Receiver Power Supply

C1,C4 2200 µF, 35 volt electrolytic capacitor

C2,C5 .33 µF, 35 volt ceramic capacitor

C3,C6 .1 µF, 35 volt ceramic capacitor

U1 LM7805 - fixed positive voltage regulator

U2 LM7905 - fixed negative voltage regulator

D1,D2,D3,D4 50 volt, 2 amp silicon diode

T1 110v-12 v center tapped transformer - 2amp

Misc circuit board, wire, terminals, plug, fuse, etc.

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

If you are interested in monitoring the ultra-low frequency spectrum from .001 Hz to 30 Hz, the ULF receiver will allow you to do some serious research into earthquake monitor and Earth field studies. This project is intended for curious individuals who want to participate in the research and experimentation of the study of extremely low frequency signals which travel through the interior of our Earth. This project will allow you to construct a ULF receive station to be used to collect data for later analysis and compare it with others who belong to the ELFRAD group and those researchers involved with the Public Seismic Network.

The ELFRAD low frequency group coordinates a research group which studies the low frequency spectrum and reports to one another the research that individual members do and makes available the information for comparison through their central website. In order to achieve their goal, it is necessary to have most of the equipment and construction parameters standardized as much as possible, so the ELFRAD Group will furnish software and support.

A complete standardized ELFRAD monitoring station consists of a ULF receiver, an antenna array, and interface card and a personal computer, and in this project we will discuss how to put a receive station together in your home workshop.

The Ultra-Low Frequency Receiver allows the desired ultra-low frequency signals to be recovered from the Earth. The signals are directed to a conditioning circuit which not only protects the receiver equipment from transients caused by lightning and other sources but attenuates the ambient 60 or 50 Hz frequency caused by commercial power, with the receiver's power-line notch line filter. The signals then pass through a circuit which removes any ground loop problems which may occur at your computer interface. The low frequencies of interest are then passed through a series of filters to remove most of the noise and unwanted signals above 25 Hz. This desired band of signals is then coupled to the analog-to-digital interface in your personal computer to be recorded on the hard drive. The interface used in the prototype accepted up to eight channels of data samples and was able to collect 200 samples per second, record them and then display them on the computer screen in either real time or compressed time format.

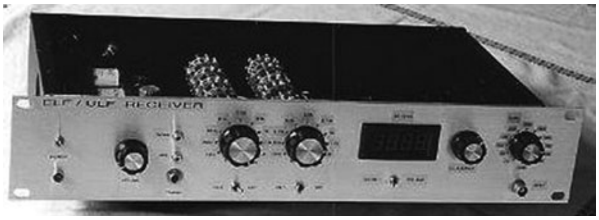

Figure 21-1 *ULF receiver*

At any time it is possible to take any period of data previously recorded, generate a file and display the data for analysis. You are also able to post filter the data, with band pass, notch, low pass, and high pass capability.

The Ultra Low Frequency Receiver project is shown in Figure 21-1. This ULF receiver circuit was designed as an all purpose ULF and ELF receiver with a frequency range from DC to 30 hertz. The signals from both leads of the non-polarized antenna array are connected to J3, as seen in the schematic diagram in Figure 21-2 and 21-3, a two sheet diagram The antenna input signal travels to pins 2 and 3 on U5, which is a high impedance differential instrumentation amplifier. This chip's gain is controlled by a single variable resistor at R3.

The combined signals then travel from pin 6 of U5 on to U3:A which is a buffer amplifier, followed by a 2 pole low pass filter at U3:B and U3:C and signal amplifier at U3:D. From pin 14 on U3, the partially filtered signal travels on to integrated circuit U2 and U1 which are 5 pole Butterworth low pass filters in cascade. The LTC1063 ICs at U1 and U2 are clock controlled and the frequency is set at 3000 hertz by R9. This configures the two LTC1063 chips to have a cut-off frequency of about 30 hertz. The clock ratio is 1/100.

The integrated circuit at U4 consists of a X5 gain amplifier at U4:D, a buffer at U4:C and another 2 pole analog low pass filter at U4:B and U4:A. Trim potentiometer R17 is the final gain control for the desired output voltage at U4:A. IC section U4B is a 2 pole low pass filter which removes any residual clocking pulses from the signal which may have been caused by U1 and U2. The amplified output signal at J1 can be routed to your analog-to-digital converter card in your PC. Power is supplied to the ULF receiver at the terminals of J2. The output from the ULF receiver can be routed to an analog-to-digital

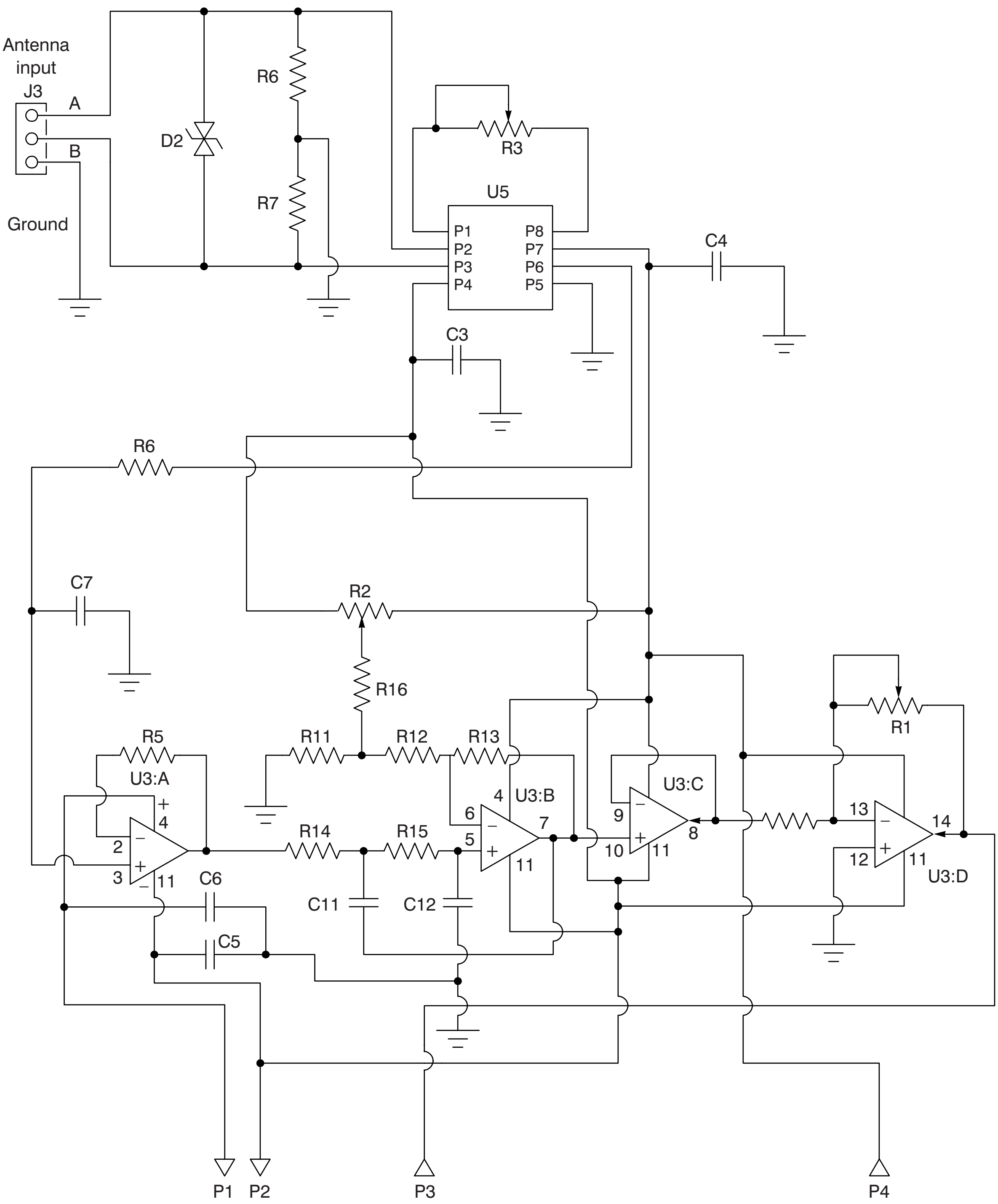
Antenna input
J3
A
B
Ground
D2
R6
R7
R3
U5
P1
P2
P3
P4
P8
P7
P6
P5
C4
C3
R6
C7
R2
R16
R5
U3:A
R11
R12
R13
U3:B
U3:C
R1
U3:D
R14
R15
C6
C5
C11
C12
2
3
4
11
5
6
7
8
9
10
12
13
14
P1
P2
P3
P4

Figure 21-2 *ULF receiver schematic – Part I*

converter card in a personal computer for recording and displaying the receiver's output in real time or for later viewing.

The ULF receiver circuit is powered from a dual plus/minus 5 volt power supply, shown in the schematic at Figure 21-4. The power supply will furnish plus and minus 5 volts DC for all of the op-amps in the receiver. A 110 to 12 volt center tapped transformer is fed directly to a four diode bridge rectifier. The resultant DC signal is then coupled to two 2200 µF filter capacitors at C1 and C4. The plus voltage output from C1 is then fed to the input of an LM7805. The output of U1 is then sent to a .1 µF capacitor at the positive output terminals. The minus voltage output from C4 is sent to a

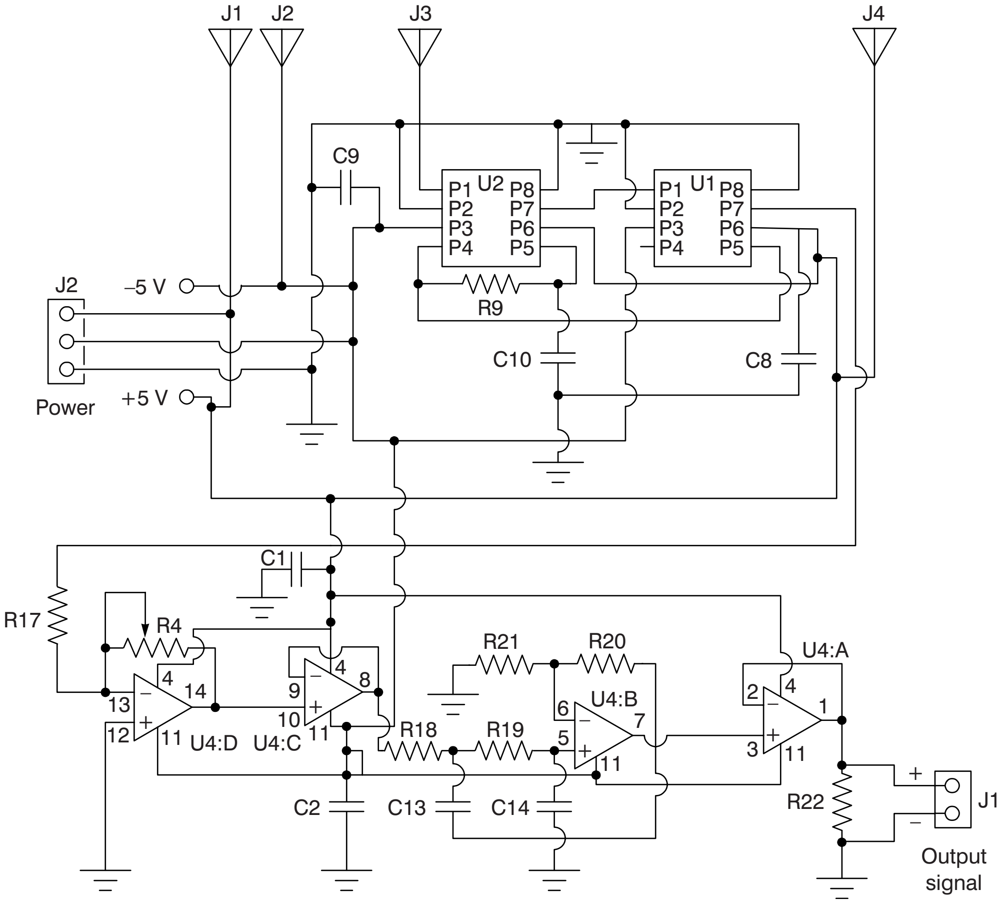
J1
J2
J3
J4
C9
U2
P1
P2
P3
P4
P8
P7
P6
P5
U1
J2
–5 V
R9
C10
C8
Power
+5 V
C1
R17
R4
4
14
13
12
11
U4:D
U4:C
9
10
4
8
11
R18
R21
R20
R19
6
5
U4:B
7
11
U4:A
2
4
1
3
11
C2
C13
C14
R22
J1
Output
signal

Figure 21-3 *ULF receiver schematic – Part II*

negative voltage regulator at U2, an LM7905. The output from U2 is then sent to capacitor C6, at the minus voltage output terminals.

Let's get started building the ULF receiver! Before we begin building the ULF receiver, you will need to locate a clean well lit and well ventilated work area. A large table or workbench would be a suitable work surface for your project. Next you will want to locate a small 27 to 33 watt pencil-tipped soldering iron, as well as a roll of #22 ga. 60/40 tin/lead solder and a small jar

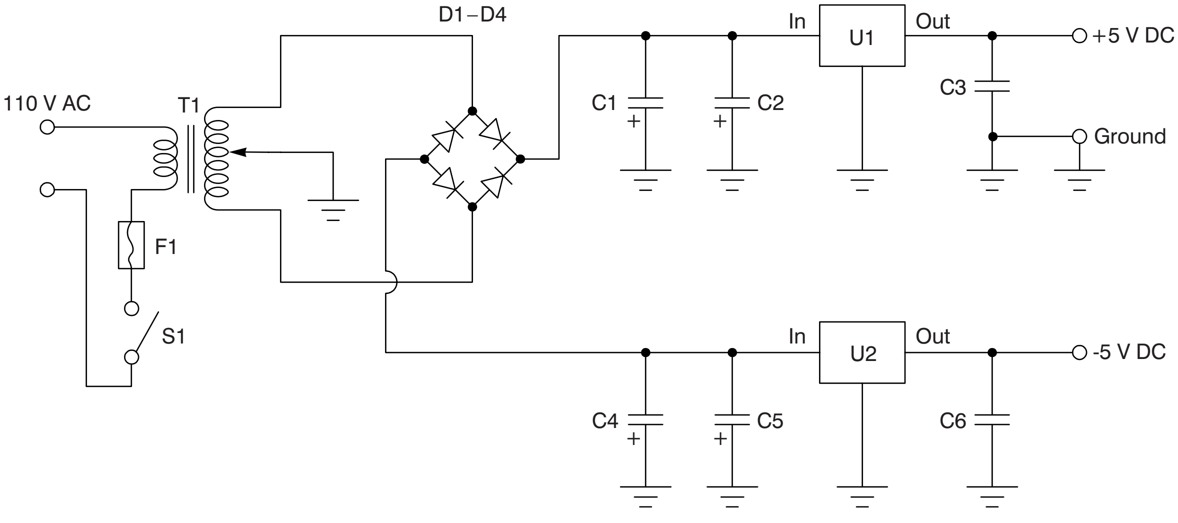
D1–D4
110 V AC
T1
F1
S1
C1
C2
In
U1
Out
+5 V DC
C3
Ground
C4
C5
In
U2
Out
-5 V DC
C6

Figure 21-4 *ULF receiver power supply*

of "Tip Tinner," a soldering iron tip cleaner/dresser from your local Radio Shack store. You will also want to secure a few hand tools for the project, such as a pair of small end-cutters, a pair of tweezers and a pair of needle-nose pliers. Locate a small Phillips and a small flat-blade screwdriver, as well as a magnifying glass to round out your tool list. Grab the schematic, parts layout diagram as well as the resistor and capacitor identifier charts and we will begin our project. Place all the project components on the table in front of you. The ultra-low frequency receiver is an RF or radio frequency project and it is best constructed on a printed circuit board with large ground plane areas covering the board for the best RF grounding techniques. Once you have all the parts and PC board in front of you, heat up the soldering iron and we'll get started!

First, find your resistor identifier chart in Table 21-1, which will help you select the resistors from the parts pile. Resistors used in this project are mostly small ¼ watt carbon composition type resistors, which have colored bands along the resistor body. The first color band should be closest to one end of the resistor body. This first color band represents the first digit of the resistor value. For example, resistor R5, R6 and R10 has four color bands, the first one is a brown band followed by a black band followed by an orange band. The fourth band is gold. The first band is brown which denotes a digit one (1), the second band is black which represents a digit five (0) and the third band is orange, which represents a multiplier of zero (000), so the resistor value is 10,000 or 10k ohms with a tolerance value of 5%. Identify the remaining resistors for the project and we can begin "populating" the PC board. Place a few resistors on the board at one time, so as not to confuse the process. Make sure that you place the correct resistor into the correct PC location before soldering it in place. Once you solder a few resistors in place on the PC board, you can use your end-cutter to trim the excess component leads. Cut the excess leads flush to the edge of the circuit board. Then place a few more resistors on the PC board and solder them on to the board. Remember to trim the component leads as necessary.

Next we will locate the capacitors for the ULF receiver. Capacitors are listed as electrolytic and non-electrolytic types. The non-electrolytic types are generally smaller in value and size as compared with the electrolytic types. Non-electrolytic capacitors in fact can be so small that their actual value cannot be printed on them, so a special chart was devised as shown in Table 21-2. The chart illustrates the three-digit codes which are often used to represent capacitors. For example a .001 µF capacitor will have (102), while a .01 µF capacitor will have (103) marked on it to represent its true value. Use the chart to identify the

Table 21-1

Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

small capacitors in the project. Non-electrolytic have no polarity markings on them so they can be installed in either direction on the PC board.

Electrolytic capacitors are usually larger in size and value and they will have a white or black band on the side of the capacitor body or a plus (+) or minus (–) marking on the body of the capacitor near the leads. These markings are polarity markings and that indicate the direction in which the capacitor must be mounted on the PC board. Failure to observe polarity when installing the capacitor may result in damage to the capacitor or to the circuit itself, so pay particular attention to capacitor polarity when placing the capacitors on the board.

Let's go ahead and place some of the non-electrolytic capacitors on the PC board. Identify a few small capacitors at a time and place them on the PC board and solder them in place. Trim the component leads after. Go ahead and install the remaining non-electrolytic capacitors after choosing your desired band.
Install the capacitors and solder them in place, remember to cut the extra leads flush to the edge of the circuit board. Now, you can go ahead and install the electrolytic capacitors onto the circuit board. Go ahead and solder them in place and trim the leads as necessary.

The ULF receiver utilizes five integrated circuits and two regulators integrated circuits. Locate the semiconductor pin-out diagram illustrated in Figure 21-5, which will help you install the ICs. When constructing the project it is best to use IC sockets as an insurance against a possible circuit failure down-the-road; it's much easier to unplug an IC rather than trying to un-solder an IC from the PC board. IC sockets will have a notch or cut-out at one end of the plastic socket. Pin one (1) of the IC socket will be just to the left of the notch or cut-out. Note, pin 1 U1 connects to pin 7 of U2, while pin 1 of U2 connects to the output of U3:D. Now take a look at U3, pin 1 of U3 is connected to R14, while pin 1 of U4 is fed to the output through R22. Finally, pin 1 of U5 is connected to potentiometer R3. Now place each of the integrated circuits into their respective sockets.

A protection Zener diode is placed across the antenna input pins between resistors R6 and R7. Note that all diodes have polarity, including Zener diodes. Diodes will have either a black or white colored band at one end of the resistor body. The lead closest to the colored band is the cathode lead of the diode.

The main ULF schematic diagram is broken into two diagrams, since the schematic is rather large. You have most likely noted that along the bottom of diagram (A), you will see designations of P1, P2, P3 and P4. These are not connectors as such but merely to help separate the two diagrams for illustration purposes. The receiver was constructed on a single printed circuit board. In the original prototype, the antenna input shown at J3 was a screw terminal strip, but this could be changed to an SO-238 antenna jack mounted on the chassis enclosure. The power connection was made at a second screw terminal jack at J2, which had three terminal posts for plus/minus and ground connections. The output of the ULF receiver was connected to a third screw terminal jack at J3, this could be changed to an RCA jack if desired.

The ULF receiver contains a number of variable resistors or potentiometers in the circuit. Potentiometer R1, R3 are PC board trimmer types, while R2, which is a 20 turn precision chassis mounted potentiometer, and R4 is a conventional chassis mounted linear taper potentiometer.

The ULF receiver is now almost completed, so take a short well deserved break, and when we return we will inspect the circuit board for any possible "cold" solder joints and "short" circuits before we apply power to the circuit. Pick up the PC board with the foil side of the board facing upwards toward you. Examine the foil side of the board carefully, take a look at the solder joints. All of the solder joints should look clean, bright, shiny and smooth; if any of the solder joints look dark, dull or "blobby" then you should remove the solder from the joint with a solder-sucker and then completely re-solder the joint all over again so that the solder joints all look good. Next we will examine the PC board for any possible "short" circuits, which may be caused from "stray" component leads which were cut from the board. Another possible cause of short circuits are solder blobs which may have stuck to the foil side of the board. Rosin core solder often leaves a sticky residue which can cause leads and solder blobs to stick to the board and possibly "bridge" across circuit traces.

Next we will need to build the dual voltage power supply circuit board. The power supply was built on the circuit board as well. The transformer was mounted off

Table 21-2
Capacitance code information

This table provides the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Some capacitors actually have the numeric values stamped on them; however, many are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

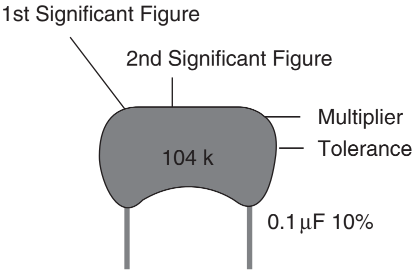

CSGNetwork.Com 6/4/92

Value	Type	Code	Value	Type	Code
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

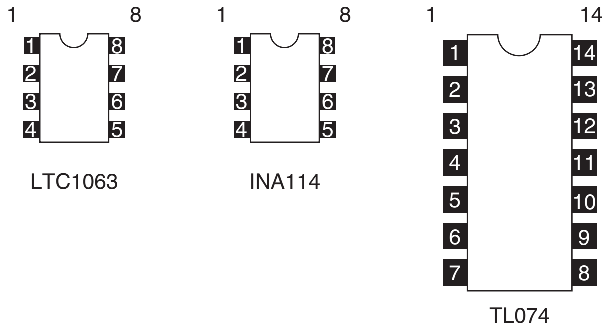

Figure 21-5 *Semiconductor pin-outs*

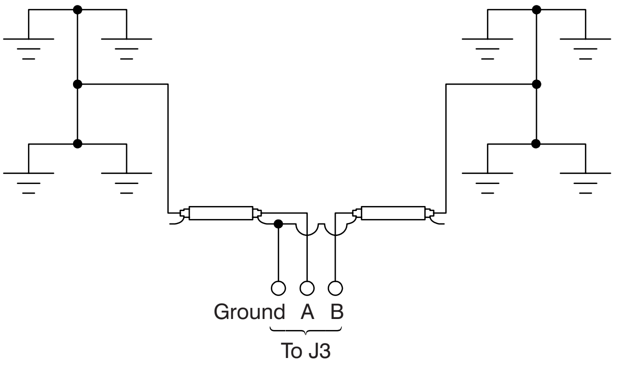

Figure 21-6 *ULF receiver antenna circuit*

the circuit board and on the bottom of the chassis. The diode bridge was placed on the circuit board, paying careful attention to the polarity of the diodes. You could elect to individual diodes or one of those four diode bridge rectifiers all in one package. If you choose the second option again you must observe the polarity of the plastic package. The AC input section will be marked with two wavy lines, the AC side has no polarity. The DC side of the bridge will have a plus (+) and a minus (–) marking. When installing the six capacitors be sure to observe polarity since four of them are electrolytic types. Note that there are two power supplies and the polarity placement orientation is different for the two power supply sections. The LM7805 is the plus voltage regulator, while the LM7905 regulator is the minus voltage regulator; don't mix them up, and take your time to install them at the correct locations to avoid damage to the circuit. A fuse was placed in series with the power switch to the transformer input leads. A power line-cord with a 110 volt plug was used to supply power to the transformer. Once the power supply circuit has been completed, you can carefully inspect the circuit board for any "cold" solder joints and "short" circuits before applying power for the first time.

The Ultra-Low Frequency Receiver requires an antenna "site" for proper operation. This is not the usual conventional type of antenna that you encounter with most receivers, since this is a special type of receiver project. The array is designed to be built completely underground. Space for construction is necessary and may be anywhere from 100 to 1000 linear feet, see Figure 21-6.

Due to the fact that we will be dealing with extremely long wavelengths on the order of miles rather than feet, you are going to need a little space. As long as you have a strip of property at least one 100 or more feet in length you will be able to detect the signals. There are several items of interest in determining your location. You will be building two arrays widely separated from each other. The further apart the better. Also if you have a choice, a damp location is better than dry rocky soil.

The first step in constructing your antenna/detector is to determine the magnetic orientation. Take an ordinary pocket compass and locate magnetic North. The line of sight between the two points should be aligned with North-South. The next step is to obtain some copper clad ground rods at least eight feet in length, put on some gloves and start driving one into the ground. We found that a standard metal fence post driver purchased from the local farm store was ideal for the job. It helps also to take a pair of post diggers and dig out a hole about 18″ deep. Then drive the rod down into the ground below the surface of the hole leaving enough room to make a connection.

Next measure out a distance as far as possible, at least 30 meters (98 ft) in a North or South direction and repeat the procedure above, driving a copper clad eight foot ground rod into the soil. Then connect the two widely spaced rods together using a length of wire. Obtain an ohmmeter, and open the wire conductor and measure the resistance. Make a note of the reading, then reverse the ohmmeter leads and measure again. Then take this reading, average it with the other reading and you will have a close indication of the DC resistance between the two probes (ground rods). Very likely, unless you have excellent soil conductivity in your area, the reading will be several thousand ohms. The point is to lower this reading as much as possible.

The lower the better. A workable resistance is 50 to 500 ohms but the ideal resistance would be around 5 to 10 ohms. In most cases, however, this would be unobtainable unless you used many ground rods or copper strapping at each end.

Keep adding rods at each probe site, spaced 20 feet apart from one another connected together with #6 copper bare wire, until the resistance gets down to between 25 and 500 ohms. Keep in mind two items. The further apart the probe sites, the lower the resistance, and the more rods driven into the ground, the lower the resistance. We have found that on the average, five grounds at each end of the site are sufficient to obtain this reading.

Another option rather than using ground rods, in rocky or extremely hard soil, you may use a ditch trencher and bury #6 or #4 copper wire at least 24 inches deep. The copper wire should be placed in an "X" fashion with a length of at least 25 feet. The lead cable will then be connected to the center point of the "X."

Now obtain a length of CATV lead in wire such as RG-11 or RG-59, from your local electronic parts store. A good source to try is the local Cable TV company. Sometimes they have lengths of scrap lying around their warehouse and will be happy to get rid of it. Connect the center conductor only (not the outside shield) to each antenna array using copper clamps. Take each piece of co-ax to the location of your receiver. Then install a suitable connector to the two ends. These two cables will be the input leads to the receiver. The shield should be connected to earth ground using a single copper clad ground rod. Do not use the power ground and make sure the ground connection is at least 20 feet from power ground.

With the ULF receiver, power supply and antenna constructed and ready to go, we just have one more detail left to complete our ELFRAD ULF monitoring station. The most expensive components will be the computer and analog-to-digital interface device. A PC type of computer is recommended using Windows 98, NT, Win2000, or XP software will work nicely. Since most of everyone owns a computer or laptop these days, the only thing left to be purchased is the A/D interface. The software for the A/D converter is available from the supplier of the A/D interface listed below. You will need at least a two channel A/D board, since it is recommended that you also record a short wave radio tuned to WWV, which keeps computer synched to within plus or minus 2 milliseconds. Also if desired, a GPS type of time standard is available. The A/D converter is available at http://seismicnet.com/ serialatod.html.

For more information on Earthquake monitoring you may finds these sites helpful.

ELFRAD Research Group

The low frequency radio research group, which provides support to interested researchers.
Helps researchers in correlating and comparing data.

http://www.elfrad.com/

Public Seismic Network

A seismic research group dedicated to earthquake monitoring.
Helps researchers in correlating and comparing data.

http://psn.quake.net/

Chapter 22

Jupiter Radio Telescope Receiver

Parts list

Jupiter Radio Telescope

R1	68 ohm
R2	294 ohm
R3	17.4 ohm
R4	294 ohm
R5	100 ohm
R6	2.2k ohm
R7	10k ohm linear potentiometer
R8	2.2k ohm
R9, R19	100k ohm
R10	220 ohm
R11	1.5k ohm
R12,R20,R21,R27	1k ohm
R13,R18	27k ohm
R14	100k ohm
R15	10k ohm potentiometer/ switch
R16	10k ohm
R17	1.5k ohm
R22,R23	2 ohm
R24	1 ohm
R25	220 ohm
R26	47 ohm
R28,R29,R30,R31,R32	10 ohm
C1	39 pF, 35v disk ceramic
C2	4-40 pF, variable capacitor
C3	56 pF, 35v disk ceramic
C4	22 pF, 35v disk ceramic
C5,C8,C11,C14	.01 mF, 35v dipped ceramic
C6	4-40 pF, variable capacitor
C7	not used
C9,C12,C13	47 pF, 35v disk ceramic
C10	270 pF, 35v disk ceramic
C15	10 pF, disk ceramic
C16,C24,C25	10 mF, 35 vdc, electrolytic
C17,C18,C21,C23,C26C29	.1 mF, 35v dipped ceramic
C19	1 mF, 35v metal polyester
C20,C22	0.068 mF, 35v 5% metal film
C27	10 mF, 35 vdc, tantalum, stripe,
C28	220 pF, 35v disk ceramic
C30,C31,C33	10 mF, 35 vdc, electrolytic
C32	330 mF, 35 vdc, electrolytic
C34,C35,C36	0.1 mF, 35v dipped ceramic
C37	10 pF, 35v disk ceramic

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

C38	10 mF, 35 vdc, electrolytic
C39	100 mF, 35 vdc, electrolytic
C40,C41,C42,C43	0.1 mF, 35v dipped ceramic
C44	10 mF, 35 vdc, electrolytic
D1	1N4001
D2,D3	1N914
LED	Red LED
VD1	MV209, varactor diode
ZD1	1N753, 6.2v, Zener diode, 400 mw
ZD2	1N5231, 5.1v, Zener diode, 500mw
L1	0.47 mH, (gold, yellow, violet, silver)
L2	1 mH, (brown, gold, black, silver)
L3	3.9 mH, (orange, gold, white, gold)
L4,L5	1.5 mH, adjustable inductor
L6,L7	82 mH, fixed inductor
Q1	J-310, junction field effect (JFET)
Q2	2N-3904, bipolar, NPN
Q3	2N-3906, bipolar, PNP
U1	SA602AN, mixer/oscillator IC
U2,U3	LM387 or NTE824 - audio pre-amplifier IC
OSC1	20 MHz crystal oscillator module
J1	Power jack, 2.1 mm
J2	F female chassis connector
J3,J4	3.5 mm stereo jack, open ckt
Misc	PC Board, enclosure, knobs, solder lugs, wire

Jupiter is the fifth planet from the Sun and is the largest one in the solar system. If Jupiter were hollow, more than one thousand Earths could fit inside. It also contains more matter than all of the other planets combined. It has a mass of 1.9 × 1027 kg and is 142,800 kilometers (88,736 miles) across the equator. Jupiter possesses 28 known satellites, four of which—Callisto, Europa, Ganymede and Io—were observed by Galileo as long ago as 1610. Another 12 satellites have been recently discovered and given provisional designators until they are officially confirmed and named. There is a ring system, but it is very faint and is totally invisible from the Earth. (The rings were discovered in 1979 by Voyager 1.) The atmosphere is very deep, perhaps comprising the whole planet, and is somewhat like the Sun. It is composed mainly of hydrogen and helium, with small amounts of methane, ammonia, water vapor and other compounds.

Colorful latitudinal bands, atmospheric clouds and storms illustrate Jupiter's dynamic weather systems, see Figure 22-1. The cloud patterns change within hours or days. The Great Red Spot is a complex storm moving in a counter-clockwise direction. At the outer edge, material appears to rotate in four to six days; near the center, motions are small and nearly random in direction. An array of other smaller storms and eddies can be found throughout the banded clouds.

Auroral emissions, similar to Earth's northern lights, were observed in the polar regions of Jupiter. The auroral emissions appear to be related to material from Io that spirals along magnetic field lines to fall into Jupiter's atmosphere. Cloud-top lightning bolts, similar to super-bolts in Earth's high atmosphere, were also observed.

The Jupiter Radio Telescope is a special shortwave receiver which will pick up radio signals from the planet Jupiter and also from the Sun, and it is a very exciting project, see Figure 22-2. This project will allow

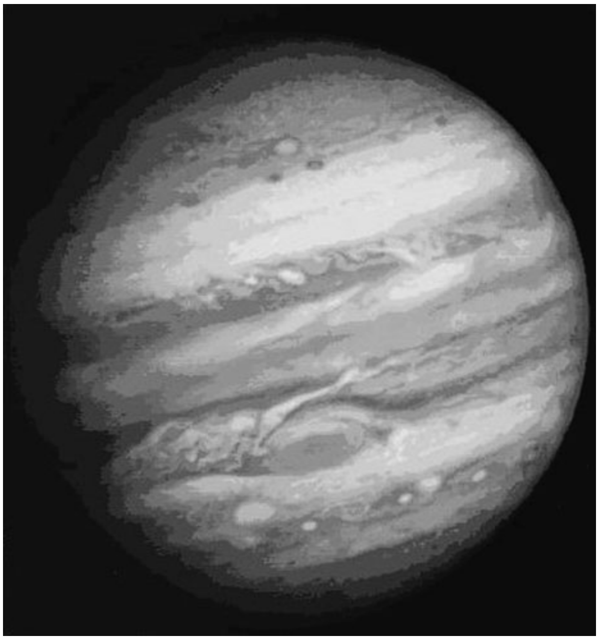

Figure 22-1 *Jupiter*

you to do amateur research into radio astronomy in your own back yard.

Radio signals from Jupiter are very weak—they produce less than a millionth of a volt at the antenna terminals of the receiver. These weak radio frequency (RF) signals must be amplified by the receiver and converted to audio signals of sufficient strength to drive headphones or a loudspeaker. The receiver also serves as a narrow filter, tuned to a specific frequency to hear Jupiter while at the same time blocking out strong Earth based radio stations on other frequencies. The receiver and its accompanying antenna are designed to operate over a narrow range of shortwave frequencies centered on 20.1 MHz (megahertz). This frequency range is optimum for hearing Jupiter signals.

The block diagram shown in Figure 22-3 illustrates the various components of the Jupiter Radio Telescope. First shown the antenna intercepts weak electromagnetic waves which have traveled some 500 million miles from Jupiter to the Earth. When these electromagnetic waves strike the wire antenna, a tiny RF voltage is developed at the antenna terminals. Signals from the antenna are delivered to the antenna terminals of the receiver by a coaxial transmission line. Next the signals are delivered to the RF Bandpass Filter and Preamplifier, they are filtered to reject strong out-of-band interference and are then amplified using a junction field effect transistor (JFET). This transistor and its associated circuitry provide additional filtering and amplify incoming signals by a factor of 10. The receiver input circuit is designed to efficiently transfer power from the antenna

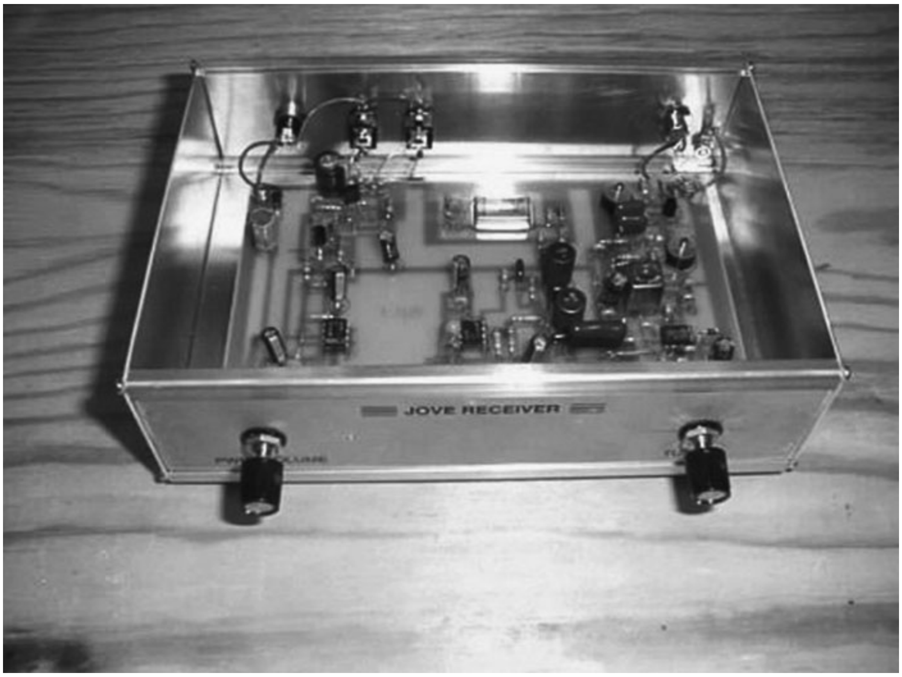

Figure 22-2 *Radio Jupiter receiver project*

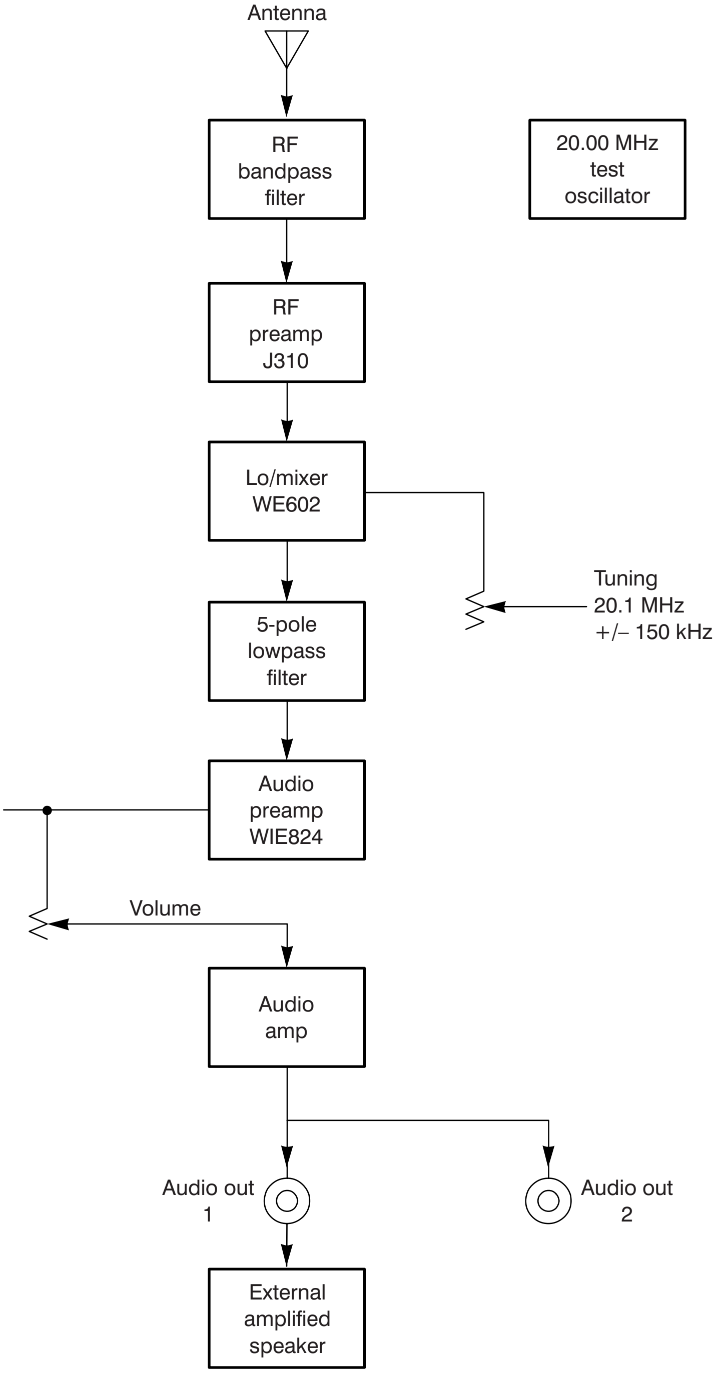
Antenna
RF bandpass filter
20.00 MHz test oscillator
RF preamp J310
Lo/mixer WE602
Tuning 20.1 MHz +/– 150 kHz
5-pole lowpass filter
Audio preamp WIE824
Volume
Audio amp
Audio out 1
Audio out 2
External amplified speaker

Figure 22-3 *Jupiter radio block diagram*

to the receiver while developing a minimum of noise within the receiver itself.

Next the signals move to the local oscillator (LO) and mixer to perform the important task of converting the desired radio frequency signals down to the range of audio frequencies. The local oscillator generates a sinusoidal voltage wave form at a frequency in the vicinity of 20.1 MHz. The exact frequency is set by the front panel tuning control. Both the amplified RF signal from the antenna and the LO frequency are fed into the mixer. The mixer develops a new signal which is the arithmetic difference between the LO and the incoming signal frequency. Suppose the desired signal is at 20.101 MHz and the LO is tuned to 20.100 MHz.

The difference in frequency is therefore 20.101–20.100 =.001 MHz, which is the audio frequency of 1 kilohertz. If a signal were at 20.110 MHz, it would be converted to an audio frequency of 10 kHz. Since the RF signal is converted directly to audio, the radio is known as a direct conversion receiver.

To eliminate interfering stations at nearby frequencies, a low pass filter is next used, which is like a window, a few kilohertz wide through which Jupiter signals can enter. When listening for Jupiter or the Sun, the radio will be tuned to find a "clear channel." Since frequencies more than a few kilohertz away from the center frequency may contain interfering signals, these higher frequencies must be eliminated. This is the purpose of the low pass filter following the mixer. It passes low (audio) frequencies up to about 3.5 kHz and attenuates higher frequencies.

Finally the low pass filter feeds an audio amplifier stage which takes the very weak audio signal from the mixer and amplifies it enough to drive headphones directly, or to drive an external amplified speaker assembly.

Circuit diagram

We have already seen a block diagram of the Jupiter radio receiver which shows the radio as a group of functional blocks connected together. While this type of diagram does not show individual components like resistors and capacitors, it is useful in understanding signal flow and the various functions performed within the radio. The next level of detail is the schematic diagram. A schematic is used to represent the wiring connections between all of the components which make up a circuit. The schematic diagram uses symbols for each of the different components rather than pictures of what the components actually look like.

A schematic diagram of the complete receiver is seen in Figure 22-4. On this schematic, the part types are numbered sequentially. For example, inductors are denoted L1 through L7, and resistors are denoted R1 through R31. Signal flow as shown in the schematic is as follows. The signal from the antenna connector (J2) is coupled to a resonant circuit (bandpass filter L1, C2, C3) and then to the J-310 transistor (Q1), where it is amplified. The output of the J-310 goes through another resonant filter (L3, C6) before being applied to the resonant input circuit (L4, C9, C10) of the SA602 integrated circuit (U1), which serves as the local oscillator and mixer. The center frequency of the local oscillator is set by inductor L5 and adjusted by the tuning control R7. The audio output from U1 passes through the lowpass audio filter (L6, L7, C20, C21, and C22). The audio signal is next amplified by U2 (an NTE824) before going to the volume control R15. The final audio amplifier stages comprise U3 (another NTE824), and the output transistors Q2 (2N-3904) and Q3 (2N-3906). After the receiver has been assembled, the variable capacitors C2 and C6 and variable inductors L4 and L5 will be adjusted to tune the receiver for operation at 20.1 MHz.

PC board assembly

Before we begin building the circuit, take a few minutes to locate a clean well lit work table or work bench. You will want to locate a small 27 to 33 watt pencil-tipped soldering iron, a roll of 60/40 tin/lead rosin core solder, and a small jar of "Tip Tinner," a soldering iron tip cleaner/dresser. "Tip Tinner" is available at a local Radio Shack store. You will also need some small hand tools, such as a small flat-blade screwdriver, a Phillips screwdriver, a pair of small needle-nose pliers, a pair of tweezers, a magnifying glass and pair of end-cutters.

Once you have gathered all the tools, locate all the circuit diagrams, such as the schematics, the parts layout diagram, the resistor and capacitor code charts, etc. Place all the component parts along with all your tools and diagrams in front of you, warm up your solder iron and we will start building the Radio Jupiter receiver project.

First we will identify the resistors for the project. Refer to the chart shown in Table 22-1, which will help you to identify the resistor color codes. Resistors usually have three or four color bands on the body of the resistor which denotes the resistor value. The color bands begin at one end of the resistor body close to the edge. The first color band on resistor R1 is blue, this denotes the digit six (6). The second color band is

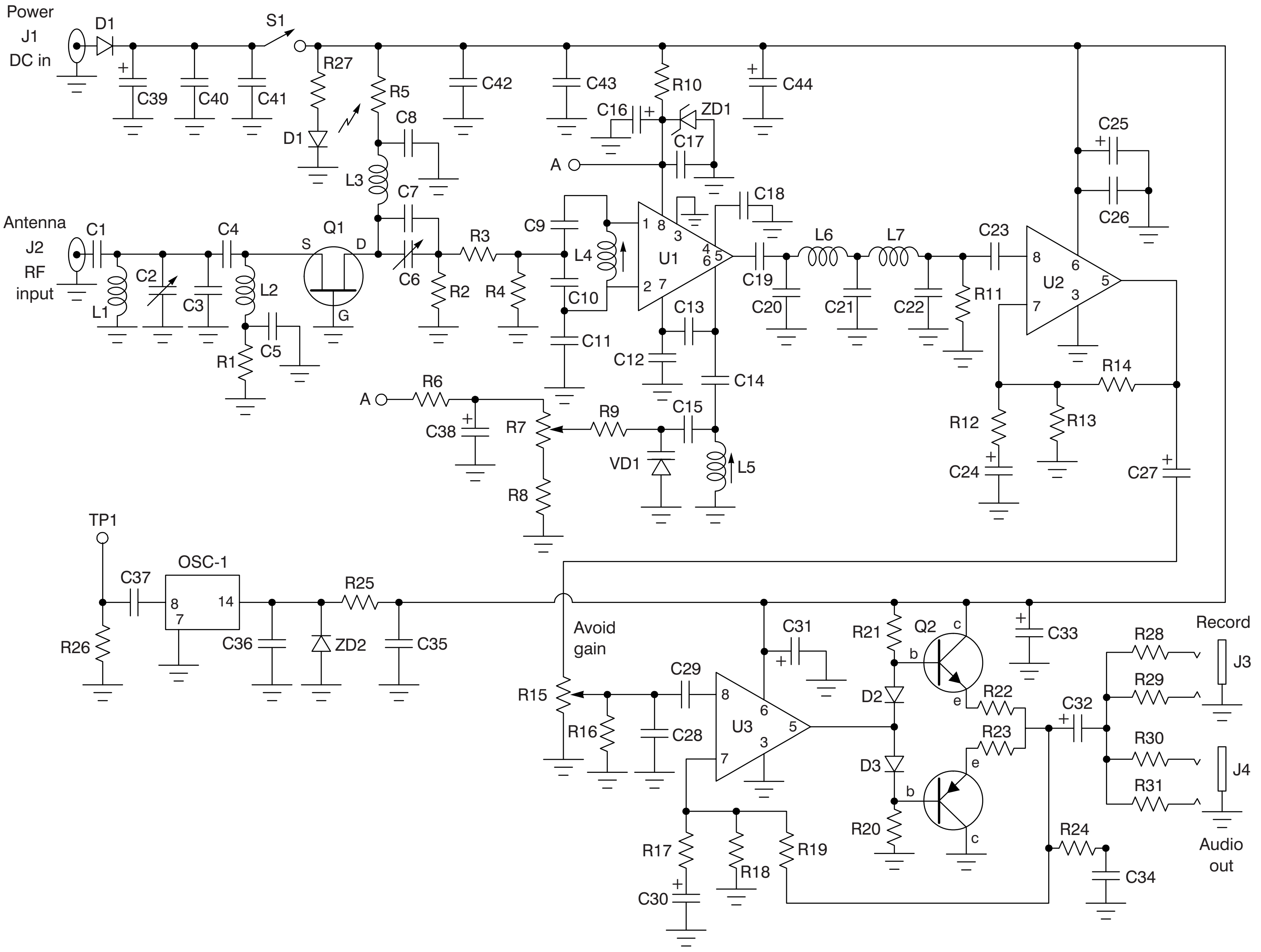
Power
J1
DC in
D1
S1
R27
D1
R5
C8
L3
C7
C6
C39
C40
C41
C42
C43
C16
R10
ZD1
C17
C44
A
C18
C25
C26
Antenna
J2
RF
input
C1
L1
C2
C3
C4
L2
R1
C5
Q1
S
D
G
R2
R3
R4
C9
L4
C10
C11
U1
1
8
3
2
7
4
6
5
C12
C13
C14
C19
L6
L7
C20
C21
C22
C23
R11
U2
8
6
5
7
3
R14
R12
R13
C24
C27
A
R6
C38
R7
R8
R9
C15
VD1
L5
TP1
C37
OSC-1
8
7
14
R26
C36
ZD2
R25
C35
Avoid
gain
R15
R16
C28
C29
U3
8
6
5
3
7
R17
C30
R18
R19
C31
R21
Q2
b
c
e
D2
D3
R20
b
e
c
R22
R23
C33
C32
R24
C34
R28
R29
R30
R31
Record
J3
J4
Audio
out

Figure 22-4 *Radio Jupiter receiver schematic*

Table 22-1
Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

gray, which is the digit number eight (8) and the third color band is the resistor's multiplier value which is black and this suggests a multiplier value of 1. So the resistor value for R1 is (6) times (8) times (1) or 68 ohms. The tolerance value of a resistor is noted in the fourth color band. A fourth color band which is silver is a 10% tolerance resistor, while a gold fourth band denotes a 5% tolerance value. If there is no fourth color band then the resistor will have a 20% tolerance value. Next, locate R2, R3 and R4, then check your parts layout diagram and the schematic to see where these resistors are to be placed on the circuit board. Once you verify where the resistors are placed, you can install these resistors on the circuit board and solder them in place on the board. Now locate your end-cutter and trim the excess component leads. Cut the extra component lead lengths flush to the edge of the circuit board. Now go ahead and identify another group of three or four resistors and place them on the circuit board. Once you are sure of their respective locations you can solder the next grouping of resistors onto the circuit board, remember to trim the excess component leads from the board. Place all the remaining resistors on the circuit board and solder them all in place, trim the extra leads as necessary.

Now we will move on to installing the capacitors for the Radio Jupiter receiver project. Capacitors come in two major flavors, non-polarized and polarized types. First, we will talk about non-polarized capacitors. Usually non-polarized capacitors are smaller value types which could take form as ceramic disks, mica, tantalum or polyester types. Often the capacitors are small in size and frequently the capacitor values are printed as three-digit codes rather than their actual values, since this code system takes less space to print. We have included a chart in Table 22-2, which lists the actual capacitor values and their respective codes. For example, a capacitor with a value of .001 µF will have a code designation of 102, a .01 µF capacitor will have 103 and a 0.1 µF capacitor will have a code of 104 marked on it. The Radio Jupiter project has a number of non-polarized type capacitors which you will need to install. Take a look through the component pile and see if you can identify the small non-polar capacitors before attempting to install them on the PC board.

Polarized capacitors are called electrolytic capacitors, they are often larger in physical size. Electrolytic capacitors will always have some sort of polarity marking on them. You will see either a black or white band or a plus (+) or minus (−) marking somewhere on the capacitor body. It is very important that you install these types of capacitors with respect to their polarity in order for the circuit to work properly and to not be damaged upon power-up. You will need to refer to the

Table 22-2

Capacitance code information

This table provides the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Some capacitors actually have the numeric values stamped on them; however, many are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

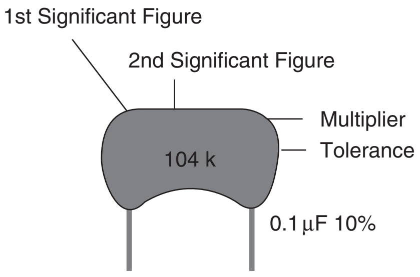

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 μF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 μF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 μF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 μF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 μF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 μF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 μF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 μF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

schematic and parts layout diagrams when installing electrolytic capacitors to make sure that you have installed them correctly. The plus (+) marking on the capacitor will point to the most positive portion of the circuit.

Let's go ahead and identify the non-polarized capacitors first. Check the capacitor code chart and make sure that you can identify each of the non-polarized capacitors. In groups of two or three capacitors install them on the PC board. When you have determined that you have the installed the correct values into their respective location on the circuit board, you can solder the capacitors in place on the board. Remember to trim the excess capacitor leads from the circuit board. Once you have the first grouping of capacitors installed you can move on to installing another group of non-polarized capacitors, until you have installed all of the non-polarized capacitors.

Next, find the electrolytic capacitors from the parts pile and we will install them in their proper locations. Refer now to the schematic and parts layout diagrams so that you can orient the capacitors correctly when mounting them on the PC board. Install two or three electrolytic capacitors and then solder them in place on the PC board. Remember to cut the extra lead lengths flush to the PC board. Identify and install the remaining electrolytic capacitors on the board and solder them in place. Trim the component leads as necessary.

In order to avoid damaging the receiver, it is important to install the capacitors, diodes as well as the semiconductors with respect to proper polarity marking on each of the components. Electrolytic capacitors have either a plus or minus marking on them and must be installed correctly by observing the polarity of the capacitor with respect to the PC board.

Next, we will install the semiconductors. Locate the semiconductor pin-out diagram shown in Figure 22-5, which will help you install the semiconductor devices properly. There are a number of diodes in the Radio Jupiter project, including Zener and varactor diodes and they all have polarity. In order for the circuit to work properly, these devices must be installed with respect to their polarity. Diodes will generally have a white or black colored band at one edge of the diode body. The band usually denotes the cathode end of the diode. Transistors generally have three leads—a Base, Collector and Emitter; pay particular attention to these leads when installing the transistors. Integrated circuits always have some type of markings which indicate their orientation. You will find either a plastic cut-out on the top of the integrated circuit package or an indented circle next to pin 1 of the IC. Be sure to use these markings when installing the integrated circuits. It is a good idea to incorporate integrated circuit sockets as a low cost insurance policy, in the event of a circuit failure, if it ever happens. It is much easier to remove an old dead IC, and simply insert a new one rather than trying to undo many pins and risk damage to the circuit board.

There are a number of coils in this project, and although there is no polarity to the coils, you will need

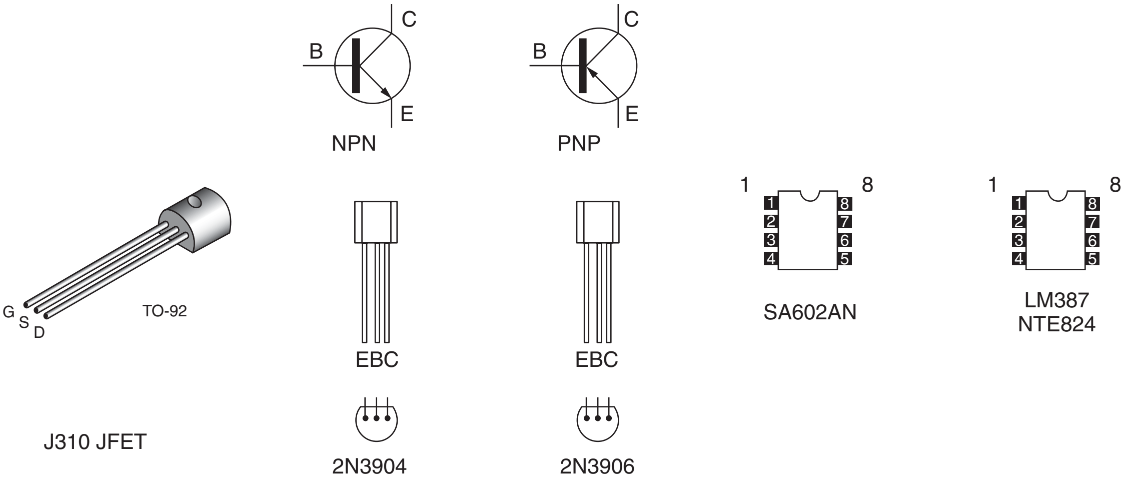

Figure 22-5 *Semiconductor pin-outs*

to pay attention to the values of the coils as well as which ones are adjustable and which are fixed before installing them on the PC board. Note, on the schematic there are two points marked (A), one is at R6 and the other at C16, both of these points are connected together. Note, there is a test point marked (TP1) at the OSC-1 module which can be used to ensure that the oscillator module is working correctly.

Power is brought to the circuit via the coaxial power input jack at J1. The antenna connection is shown at J2. This connector can be either a BNC or an "F" type chassis mounted connector. The receiver circuit has two audio output jacks both at J3 and J4. Two circuits of mini ⅛″ stereo jacks were used in the prototype.

Once the printed circuit has been assembled, you will need to recheck the solder joints to make sure they are smooth and that there are no cold solder joints. Also observe the PC circuit lines to ensure that there are no bridges or "shorts" formed by cut component leads. Once the circuit has been checked, you are ready to install the circuit board in the chassis.

A metal enclosure was chosen, since this is a sensitive RF receiver and you will want to keep outside interference from affecting the circuit operation, see Figure 22-2. You will have to wire the four jacks J1 through J4 to the circuit board. We elected to mount the power and antenna jacks on the rear panel of the chassis box. One audio jack was mounted on the front panel, while the second audio jack was mounted on the rear chassis panel. The power switch S1 was wired to the PC and mounted on the front panel of the receiver along with the power indicating LED. The tuning control R7 and volume control R15 were both mounted on the front panel of the chassis box.

Once you are satisfied that your circuit board is wired correctly. You are ready to apply power to the circuit and begin testing the circuit for proper operation.

Testing and alignment

The receiver requires 12 volts DC (vdc) which may be obtained from a well regulated power supply or from a battery. Current drain is approximately 60 milliamps (ma). The power cable supplied with the kit has a female power plug on one end and stripped leads on the other. Notice that the power cable has a black stripe, or tracer, along one of the wires. This is the wire that is connected to the center conductor of the plug and must be connected to the (+) side of the power source. The Radio Shack RS 23-007 (Eveready) 12 volt battery or equivalent is suitable.

Next, turn the Jupiter receiver's power switch to OFF. Connect either headphones or an amplified speaker (Radio Shack 277-1008C or equivalent) to the receiver audio output (J3 or J4). These jacks accept 3.5 mm (⅛″) monaural or stereo plugs.

If you are using a Radio Shack amplified speaker, turn it ON and adjust the volume control *on the speaker* up about ⅛ turn. If you are using headphones, hold them several inches from your ear as there may be a loud whistle due to the internal test oscillator. Turn the receiver ON. The LED should light. Set the volume control to the 12 o'clock position. Allow the receiver to "warm-up" for several minutes.

Set the TUNING control to the 10 o'clock position. *Carefully* adjust inductor L5 with the white tuning stick until a loud low frequency tone is heard in the speaker (set volume control as desired). *Caution*: Do not screw down the inductor slugs too far, as the ferrite material could crack. By adjusting L5 to hear the tone, you are tuning the receiver to 20.00 MHz. The signal which you hear is generated in OSC1, a crystal controlled test oscillator built into the receiver. Once L5 has been set, DO NOT readjust it during the remainder of the alignment procedure. (When the receiver tunes 20.00 MHz with the nob set to the 10 o'clock position it will tune 20.1 MHz with the knob centered on the 12 o'clock position.) The following steps involve adjusting variable capacitors (C2 and C6) and a variable inductor (L4) to obtain the maximum signal strength (loudest tone) at the audio output. For some, it is difficult to discern slight changes in the strength of an audio tone simply by ear. For this reason three different methods are described, each using a form of test instrument. In the event that no test equipment is available, then a fourth method—simply relying on the ear—is possible. In all cases, adjust the receiver tuning knob so that the audio tone is in the range of about 500 to 2000 Hz. Use procedure A as a method to use in tuning up your receiver. Procedures B-D are optional tuning methods.

If no test equipment is available, then simply tune by ear for the loudest audio signal. Listen to the tone and carefully adjust the tuning knob to keep the pitch constant. If the pitch changes during the alignment, it indicates that the receiver has drifted off frequency. As you make adjustments, the signal will get louder. Reduce the receiver volume control as necessary to keep the tone from sounding distorted or clipped.

To test the receiver on the air, simply connect the antenna. For best performance, use a 50 ohm antenna designed to operate in the frequency range of 19.9–20.2 MHz. At certain times of the day, you should be able to hear WWV or WWVH on 20.000 MHz. These are standard time and frequency stations, located in Colorado and Hawaii, which broadcast the time of day as well as other information related to propagation and solar-terrestrial conditions. The Jupiter Radio Telescope direct conversion receiver design does not allow clear reception of amplitude modulated (AM) stations like WWV, so the voice will probably be garbled, unless you tune very precisely. The receiver does work well on single sideband (SSB) signals and code (CW).

In the event that the Jupiter receiver does not appear to work, you will need to disconnect the power supply and antenna and examine the circuit board for errors. The most common cause for building errors are the incorrect installation of the electrolytic capacitors and diodes which may have been installed backwards, since the components have polarity which must be observed. Another possible cause for errors are the incorrect placement of resistors. Often color codes are read wrong and the wrong resistor may be installed at a particular location. Finally one of the most common causes for errors is the incorrect installation of transistors and integrated circuits. Make sure that you can identify the manufacturer's pin-out diagrams and pay particular attention to if the pin-outs are for top or bottom views. Have a knowledgeable electronics enthusiast help you scan the circuit for possible mistakes; a second pair of eyes always helps to find errors you may have missed as the builder. Once you have located and fixed the problem, you can apply power and re-test the circuit for proper operation.

The jupiter radio telescope antenna

The antenna intercepts weak electromagnetic waves that have traveled some 500 million miles from Jupiter to the Earth or 93 million miles from the Sun. When these electromagnetic waves strike the wire antenna, a tiny radio frequency (RF) voltage is developed at the antenna terminals. A basic dipole antenna is shown in Figure 22-6. The recommended antenna for the Jupiter Radio Telescope is the dual dipole shown in Figure 22-7. Signals from each single dipole antenna are brought together with a power combiner via two pieces of coaxial cable. The output of the power combiner is delivered to the receiver by another section of coaxial transmission line. The antenna system requires a fair-sized area for setup: minimum requirements are a 25 × 35 ft. flat area that has soil suitable for putting stakes into the ground. As the antenna system is sensitive to noise it is best not to set it up near any high tension power lines or close to buildings. Also for safety reasons, please keep the antenna away from power lines during construction and operation.
The best locations are in rural settings where the interference is minor. Since many of the observations occur at night it is wise to practice setting up the antenna during the day to make sure the site is safe and easily accessible.

If you want to locate your receiver indoors and your antenna is outside or on the rooftop, and your antenna lead will not reach, you must use a longer

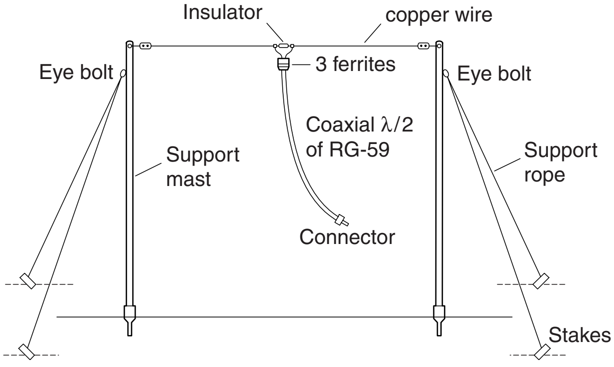

Figure 22-6 *Basic dipole antenna diagram. Courtesy Wes Greenman, Charles Higgins*

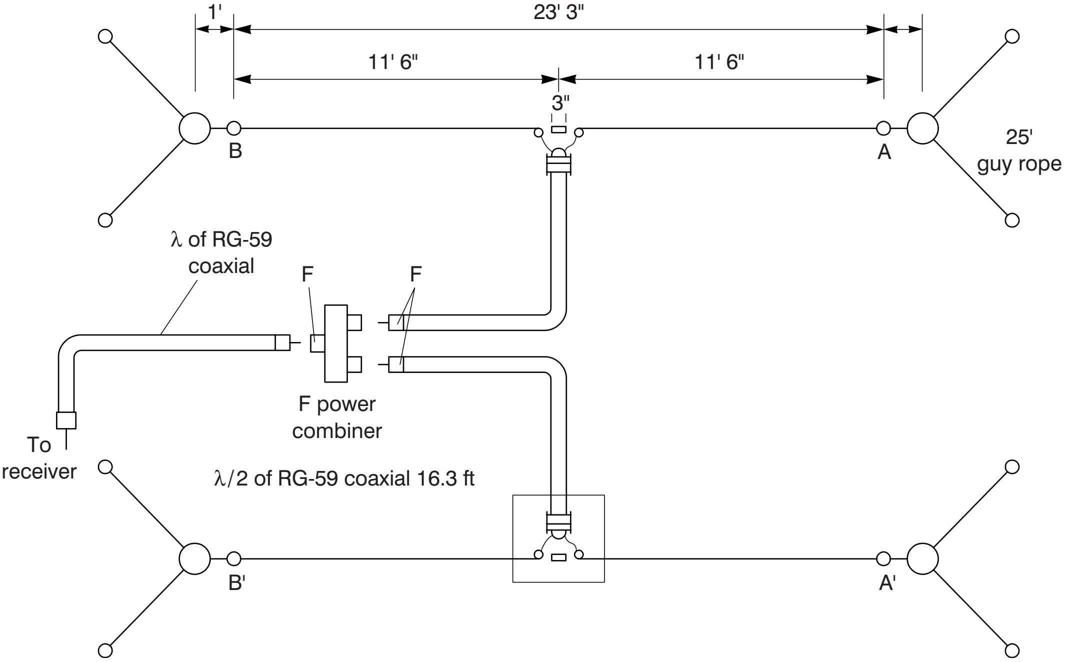

Figure 22-7 *Antenna combiner diagram. Courtesy Wes Greenman, Charles Higgins*

coax cable. Do not use just any length of coax for the combiner assembly. The cable going from the power combiner harness assembly to the receiver should be a multiple half wavelength long. The cable supplied with the kit is one wavelength—that is 9.85 meters long (taking into account the 66% velocity factor of the RG-59/U cable). The maximum recommended cable length is 5 wavelengths. There are many different manufacturers and qualities of coaxial cable. The 75 ohm cable supplied with the kit is manufactured by Belden and has a solid center conductor and velocity factor of 66%. Radio Shack does not carry RG-59/U cable but they do have RG- 6 and the higher grade RG-6QS (quad shield), which is also 75 ohm cable. Both of these cables have a velocity factor of 78%. One wavelength at 20.1 MHz in RG-6 cable is 11.64 meters. If you are going to put in a longer feedline we recommend that you completely replace the existing one wavelength piece—rather than splicing another length of cable onto the end.

Observing tips—jupiter's radio emissions

Most Jovian radio storms, received with the Jupiter radio telescope, will likely not be very strong. You will need to maximize your chances of receiving the Jovian storms by picking the best opportunities, selecting a good observing site, and using good observing techniques.

During the daytime, the Earth's Ionosphere blocks Jupiter's radio emissions, so therefore you will have to confine your investigations to night time listening. Radio noise from electrical sources, such as power lines, fluorescent lights, computers, and motors, can mask the signals from Jovian storms.

If possible, try to select an observing sight away from buildings and power lines, also try to use portable power supplies such as a battery power supply. Note that Jovian radio storms may be short in duration, so persistence and luck are needed to capture Jupiter's radio signals.

Night time observing from a temporary field setup can be dangerous, so PLAN CAREFULLY!! Get permission in advance to use the site, set up before dark, and use caution when walking around the site at night. The Jupiter Radio Telescope has two audio outputs, one channel can be used for ordinary listening and the second output jack can be used for recording. You can elect to use a portable audio tape recorder, however, note that nearly all common portable tape recorders use an automatic recording level control circuit (ALC). The ALC can mask the variations in the Jovian L bursts.

Use a tape recorder that allows you to SWITCH OFF the ALC. Keep the ALC turned off.

Another alternative for recording Jovian storms is the use of a computer or laptop to record the signals from your Jupiter radio telescope. Using a laptop computer to record the Jovian storms has the distinct advantage that you can analyze the audio recording with special spectral software which can help you make sense of the complex signals received.

Note you could also elect to use an audio tape recorder to record the Jovian signals in the field and later when you return home play the tape recording into your laptop or personal desktop computers. There are many audio analyzing software programs out there on the Internet, many are free. Many programs allow FTT and frequency division analysis.

Chapter 23

Weather Satellite Receiver

Parts list

Weather Satellite Receiver

R1 22 ohm, ¼watt, 5% resistor

R2 150 k ohm, ¼watt, 5% resistor

R3 110 k ohm, ¼watt, 5% resistor

R4,R10,R11 22 k ohm, ¼watt, 5% resistor

R5 47 k ohm, ¼watt, 5% resistor

R6 1.8 k ohm, ¼watt, 5% resistor

R7 360 ohm, ¼watt, 5% resistor

R8 100 ohm, ¼watt, 5% resistor

R9 470k ohm, ¼watt, 5% resistor

R12,R13,R16,R27 1 k ohm, ¼watt, 5% resistor

R14 240 ohm, ¼watt, 5% resistor

R15 1.2 k ohm, ¼watt, 5% resistor

R17 5.6 k ohm, ¼watt, 5% resistor

R18 39 k ohm, ¼watt, 5% resistor

R19,R21 10 k ohm, ¼watt, 5% resistor

R20,R31 2.2 k ohm, ¼watt, 5% resistor

R22 100 k ohm, ¼watt, 5% resistor

R23 220 ohm, ¼watt, 5% resistor

R24 390 k ohm, ¼watt, 5% resistor

R25 10 ohm, ¼watt, 5% resistor

R26 1.5 k ohm, ¼watt, 5% resistor

R28 4.7 k ohm, ¼watt, 5% resistor

R29 300 ohm, ¼watt, 5% resistor

R30 47 ohm, 14watt, 5% resistor

VR1 50 k linear pot, 16mm PC board mount

VR2 25 k linear pot, 16mm PC board mount

VR3 50k log pot, 16mm PC board mount

VR4,VR5 50 k 10-turn trimpots, PC board mount

C1,C2,C3,C5 2.2 nF, 16v, disk ceramic capacitor

C6,C7,C8,C12 2.2 nF, 16v, disk ceramic capacitor

C4,C10 3-10 pF trimcaps

C9 1 nF, 16v, disk ceramic capacitor

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

C11,C34,C35 10µF, 16v tantalum capacitor

C13,C27 10pF, 16v, NPO ceramic capacitor

C14,C19,C21,C23,C25 10nF, 16v, mono ceramic capacitor

C15,C16 15pF, 16v, NPO ceramic capacitor

C17 10µF, 35v Tantalum capacitor

C18a,C18b,C20,C22 100nF, 16v, mono ceramic capacitor

C24,C26,C28,C32 100nF, 16v, mono ceramic capacitor

C29 390pF, 16v, NPO ceramic capacitor

C30 4,7nF,16v, metal polyester capacitor

C31 1nF, 16v, metal polyester capacitor

C33 470nF, 16v, metal polyester capacitor

C36 330µF, 16v, electrolytic capacitor

C37 47nF, 16v, metal polyester capacitor

C38 22nF,16v, metal polyester capacitor

C39 2200µF, 25v, electrolytic capacitor

C40 470µF, 35v, electrolytic capacitor

RFC1 4.3µH slug tuned coil – see Table 23-3

RFC2 10µH slug tuned coil – prewound

RFC3 68µH slug tuned coil – prewound

L1 see Table 23-3

L2 see Table 23-3

L3 see Table 23-3

L4 see Table 23-3

D2 5.1V 400mW Zener diode

D3,D4 3mm red LED

D1 ZMV833ATA varicap

X1,X2 5.5MHz ceramic filters, Murata SFTRD5M50AF00-B0

Q1 BF998 dual-gate MOSFET

Q2 PN100 NPN transistor

U1 SA605D mixer/IF amplifier/FM detector

U2 TL072 dual op-amp

U3 LM386 audio amp

U4 78L05 +5V regulator

J1,J2 BNC type jacks

J3 ⅛, 3.5mm stereo headphone jack

J4 1 2.5mm concentric power socket, PC-mount

S1 SPDT miniature toggle switch

Misc PC board (double-sided, not plated through)

TO-220 heatsink,

1 coil former, 4.83mm OD with F16 ferrite slug

1 F29 ferrite bead (for RFC1)

1 short length of 0.8mm tinned copper wire for L1

1 length of 0.8mm enameled copper wire for L2, L3

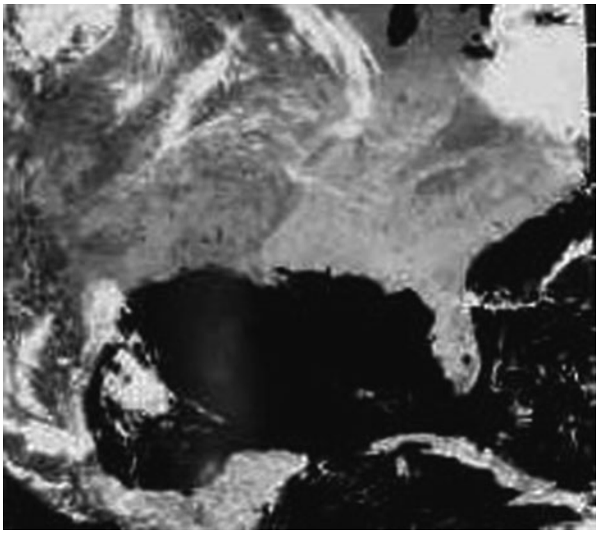

Figure 23-1 *Satellite view – USA*

The days of guessing the weather by looking at the clouds overhead have just ended. Now you can look at the clouds from above! This satellite receiver/decoding project will allow you to receive pictures from satellites 600 km overhead, see Figure 23-1. A typical NOAA satellite can cover nearly one-sixteenth of the Earth in a single pass! Living in New York, for example, we can clearly capture images from mid-Hudson bay (where there was still ice in late spring), all the way down past Cuba, as well as spanning from Wisconsin to far out in the Atlantic Ocean. The clarity of the image was enough to see the individual Finger Lakes (in New York), and shadows on the underside of thunderstorms. With the weather satellite receiver and a personal computer, you can easily receive and decode weather satellite pictures from overhead satellites.

The US National Oceanographic and Atmospheric Administration (NOAA) operates both GEO and LEO polar satellites. Here we concentrate primarily on the polar orbiters. The NOAA polar satellites orbit at 850 km and pass within view of all areas on Earth at least twice day. The satellite carries a number of instruments including cameras for both visible and infrared light.

The cameras scan back and forth at right angles to the ground path, like a broom sweeping side-to-side as you walk forward, taking picture strips that cover an area 3000 km wide. The satellite thus makes a continuous picture as if it were a tape reeling out from an endless roll. The image, however, is not recorded on the satellite. Each image strip is immediately broadcast to the ground at a frequency just above 137 MHz. The satellite will be in range for up to 12 minutes as the satellite passes from horizon to horizon.

The broadcast uses the APT (Automatic Picture Transmission) analog format for the imagery. (A digital format High Resolution Picture Transmission (HRPT) signal is also transmitted but it is more difficult to receive and decode and we did not try to receive it in this project.) Currently, there are six operational NOAA polar orbiters but usually only two or three have the APT activated at a given time. (In this report, only images from NOAA-14 and NOAA-15 were received.) See NOAA satellite status.

The Russian METEOR series of satellites also broadcast with the APT format. In addition, the remote sensing SICH and OKEAN radar satellites occasionally broadcast images but only over Russia (scanners in nearby countries have picked up some of these images).

The APT system

The APT satellites give about 12 good images a day, wherever you are in the world. The resolution is 4 km per pixel and there are visible and infrared sensors. This is a simple system where the satellite is in low Earth orbit and very easily received at 137 MHz. The two frequencies in main use are 136.50 for NOAA 12 and 15 and 137.62 MHz for NOAA 14. A simple crossed dipole antenna can be used, or for marine use a stainless steel quadrifilar helix antenna. These simple omni-directional antennas do not need moving or tracking to receive the satellite. The satellite is so strong that when is about 20° above the horizon perfect results will be obtained. The APT signal is easily recognizable by its modulated frequency (FM) and carrier tonality ranging between 1500 and 2500 Hz. New satellites have a 2400 Hz carrier like this NOAA 14 APT signal.

In APT mode, satellites transmit at the rate of 120 LPM (lines per minute) alternating two channels, one for the visible image, the other for the infrared image. This is through the decoding software that you select one or another channel.

Russian weather satellites only transmit infrared images at 120 LPM. The sound of their transmission is thus slightly different from the other weather satellites like on this record of a RESURS 01-N4 signal with an IOC of 382. On their side, METEOR 3-05 uses a higher carrier closer to 2500 Hz with an IOC of 382 as well.

The WX satellite receiver project will allow you to receive weather satellite transmissions on the VHF band, where most of the polar-orbiting satellites are located. You will recognize these transmissions on the news when you see the time lapse of the clouds darting across the countryside. The weather man in this case has taken multiple images on the computer, aligned and pieced them together, and then run through one image after the other. It is possible to do this same thing with this kit and the proper software.

The way in which a weather satellite works is fairly simple. Just think of your office fax machine as an example. The satellites circle the Earth going north to south back to north again almost directly over the poles, which is why they call it a polar orbit. This means that the satellite will cover every location on the Earth at least twice per day. With a good antenna, and partly because of overlap of consecutive orbits, you can conceivably receive the same satellite up to six times a day! Notice though that the image received from polar orbits will be upside down on every other pass.

The satellite retrieves the data in a linear fashion, one line at a time using a *scanning radiometer*. The scanning radiometer transmits the equivalent of a single television horizontal line as the satellite circles the earth. The system uses a series of optics and a motor driven rotating mirror system to receive a very narrow line of the image of the Earth. Each line is received at a right angle to the satellite's orbital track, so as the satellite circles the Earth, a line is received from west to east or east to west depending on the orbit of the satellite. The total image is received from north to south or south to north, depending on the orbit also, and this motion is what relays the equivalent of the vertical scan in a television. You can continue receiving this satellite as long as it is within the line of sight.

Since all of the receivable satellites are similar, we will describe the ones you will most commonly receive. The NOAA/TIROS satellites, during the first half of the transmission, send visible light data to the receiver at the same time they are taking in the view. Meanwhile during the same part of the scan, they are recording the infrared view. During the second half of the scan, while the sensors are facing away from the Earth, it sends the infrared data. The user then sees the data as two images side by side, on the left the visible light data is seen, and on the right, infrared data is seen. In between the images are synchronization pulses that help computers to align the individual lines precisely.

These particular satellites continuously transmit an FM signal modulated with a 2400 Hz tone. This tone is very precise in frequency so the image seen is aligned properly. The 2400 Hz tone is AM modulated with the intensity of the current view of the Earth. The brighter or colder the point on the Earth, the higher in amplitude the 2400 Hz signal is.

The receiver demodulates the FM signal and retrieves the 2400 Hz tone. The detector board in the computer will then find the peak amplitude of each wave of the 2400 Hz tone, and each peak, upper and lower, now represents a single pixel on the screen. For the NOAA/TIROS satellites, each horizontal line represents 2400 pixels, since the incoming frequency is 2400 Hz, and the scanning radiometer rotates twice per second. The full 12 minute pass of a NOAA satellite requires approximately 3.5 MB (3.5 million 8 bit pixels) of storage! This is much more data (pixels) than can be seen on a super VGA screen at any one time.

Many people mistakenly think that they can easily utilize a conventional VHF scanner as an APT satellite picture receiver. Most scanners and communications receivers usually only provide a choice of two bandwidth settings for VHF FM reception: "narrow" and "wide." The narrow setting gives a bandwidth of ±15 kHz or less, which is fine for NBFM reception, and the wide bandwidth setting usually gives a bandwidth of about ±100 kHz, so this is the setting that must be used. Unfortunately, this "wide" bandwidth is actually too wide for weather satellite signals and, as a result, the demodulated audio level is relatively low. At the same time, the wider reception bandwidth allows more noise through, so the signal-to-noise ratio can become quite poor. Only a few scanners have the correct bandwidth and these currently include the AOR5000 and Icom IC-PCR1000. The second problem with all scanners,

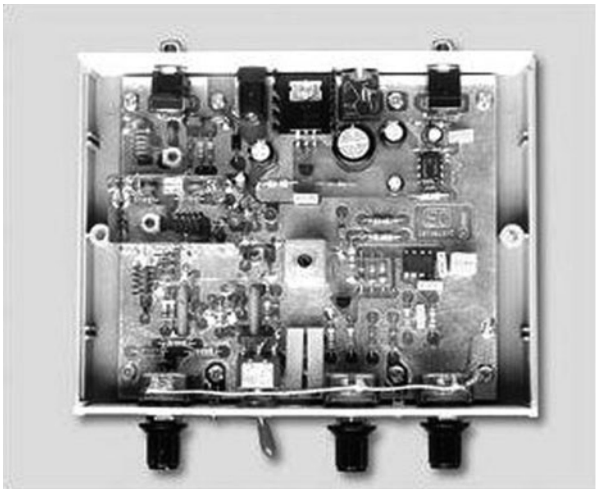

Figure 23-2 *Weather satellite receiver*

including the ones mentioned, is that a weather satellite system needs a very good performance receiver, one that provides high sensitivity, good signal to noise and high immunity to other adjacent transmissions. So, you may get a scanner to work, but it will not provide the optimum results.

Ideally, in order to obtain the best results you really need to have a receiver with a ±30 kHz bandwidth. This type of specialized VHF receiver is available but fairly pricey. Hence the motivation for developing the low-cost weather satellite receiver described here.

The weather satellite receiver project in this chapter features good sensitivity, at about 0.7V for 12 dB of quieting. At the same time, the effective bandwidth is approximately ±35 kHz, which is quite suitable for weather satellite reception. The satellite receiver is built into a very compact plastic instrument box, as shown in Figure 23-2. All of the circuitry is mounted on a double-sided PC board, so it's quite easy to build. It has switch tuning between two preset frequency channels, for ease of use. There are RF Gain, Audio Muting and Audio Gain controls and the receiver can drive a small monitor speaker or headphones, as well as providing a line level signal to feed into your PC for recording and decoding.

The heart of the receiver is a Philips SA605D IC high-performance low-power FM mixer and IF system on a single chip. As you can see from the block diagram of Figure 23-3, it contains a local oscillator transistor and balanced mixer, plus a high-gain IF amplifier and IF limiter, a received signal strength (RSSI) detector, an

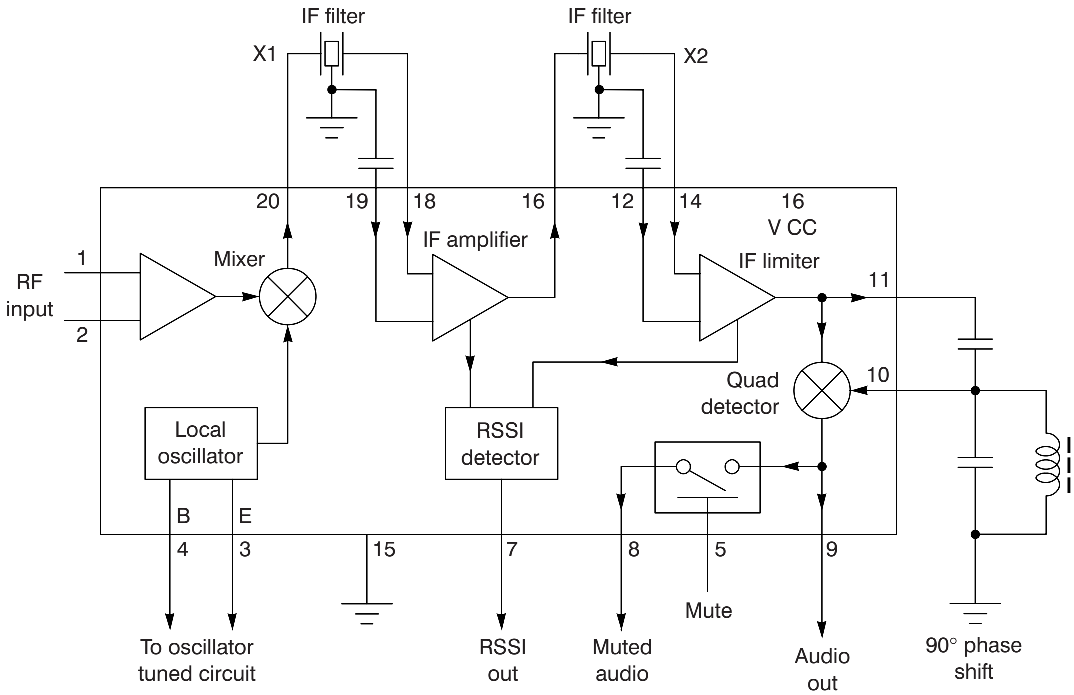
IF filter
X1
IF filter
X2
20
19
18
16
12
14
16
V CC
Mixer
IF amplifier
IF limiter
1
RF
input
2
11
Quad
detector
10
Local
oscillator
RSSI
detector
B
E
4
3
15
7
8
5
9
Mute
To oscillator
tuned circuit
RSSI
out
Muted
audio
Audio
out
90° phase
shift

Figure 23-3 *Weather satellite receiver block diagram*

FM quadrature detector and finally an audio muting circuit. The local oscillator transistor can operate at frequencies up to about 500 MHz in an LC circuit, or up to 150 MHz with a suitable crystal. The mixer can operate up to 500 MHz as well, while the IF amplifier and limiter can operate up to about 25 MHz with a combined gain of about 90 dB.

The main receiver diagram illustrated in Figure 23-4 depicts the complete weather satellite receiver circuit details. In this receiver project, we're using the SA605D in a fairly conventional single-conversion superhet configuration, with the IF amplifier and limiter working at 5.5 MHz. This allows us to take advantage of high selectivity 5.5 MHz TV sound IF ceramic filters to provide most of our bandwidth shaping. The two filters in question are X1 and X2, which are both Murata SFT5.5MA devices.

As shown in the receiver schematic, X1 is connected between the mixer output and the IF amplifier input, while X2 is connected between the IF amplifier output and the limiter input. The resistors connected to the filter inputs and outputs are mainly for impedance matching, while the 10 nF capacitors are for DC blocking. The 90° phase shift required for U1's quadrature FM detector is provided by coil L4 and its parallel 390 pF capacitor, which are tuned to 5.5 MHz.

The local oscillator transistor inside U11 is connected in a Colpitts circuit. This includes coil L3, together with the two 15 pF capacitors (which provide the emitter tap) and a 10 pF capacitor in series with the Varicap diode D1.

The Varicap diode D1 is the receiver's tuning capacitor. Its tuning voltage for each of the two channels is set by 10-turn trimpots VR4 and VR5, with switch S1 selecting between them. We can tune the receiver simply by changing the local oscillator frequency because we only need to tune over a relatively small range (i.e. 137.3–137.85 MHz maximum), which is within the selectivity curve of the "front end" tuned circuits.

The satellite receiver's front end uses a BF998 dual-gate MOSFET (Q1) connected in a standard cascode amplifier configuration. The incoming VHF signals are fed into a tap (for impedance matching) on antenna coil L1, which is tuned to about 137.55 MHz using trimmer capacitor VC1. The signal from the top of this tuned circuit is then fed directly to gate 1 of Q1, while gate 2 is bypassed to ground but also fed with an adjustable DC voltage via VR1 for RF gain control. The amplified VHF signal on Q1's drain is fed to pin 1 of U1 via a 1 nF coupling capacitor. Additional RF selectivity is provided by coil L2 and trimmer capacitor VC2, which are again tuned to about 137.55 MHz. The 100 ohm resistor and 10 µH RF choke form an untuned high-impedance load for Q1.

Notice that as well as being coupled to the tap on L1 via a 2.2 nF capacitor, the antenna input is also connected to the +12V supply line via RFC1 and a series 22 ohm resistor. As you may have guessed, these components are there to provide "phantom" DC power for the masthead pre-amplifier. At the output end of U1, we take the demodulated APT signals from the "muted audio" output at pin 8. This allows us to take advantage of the SA605's built-in muting circuit, which works by using comparator stage U2b to compare U1's RSSI output from pin 7 (proportional to the logarithm of signal strength) with an adjustable DC control voltage from muting pot VR2. When the RSSI voltage rises above the voltage from VR2, U2b's output switches high and this is fed to pin 5 of U1 via a 2.2 k ohm series resistor to un-mute the audio. D2, a 5.1V Zener diode, limits the swing on pin 5 of U1 to less than 6V.

Transistor Q2 and LED1 form a simple signal strength indicator. This also uses the RSSI output from U1. In operation, the voltage across the 390 k ohm resistor and 100 nF capacitor rises from about +0.26V under no-signal conditions to about +5V with a very strong input signal. So with Q2 connected as an emitter follower and LED1 in its emitter load, the LED current and brightness are made to vary quite usefully with signal strength.

The demodulated APT signal from pin 8 of U1 is first fed through op-amp U2a, which is configured as an active low-pass filter. This has a turnover frequency of 5 kHz and is used for final de-emphasis and noise reduction. From there, the signal is fed to audio gain control VR3 and then to audio amplifier stage U3. This is a standard LM386 audio amplifier IC, configured for a gain of about 40 times. Its output is fed to both the monitor speaker socket and to a line output socket for connection to your PC's sound card.

Most of the receiver's circuitry operates from +12V, with the exception of U1 which needs +6V. As a result,

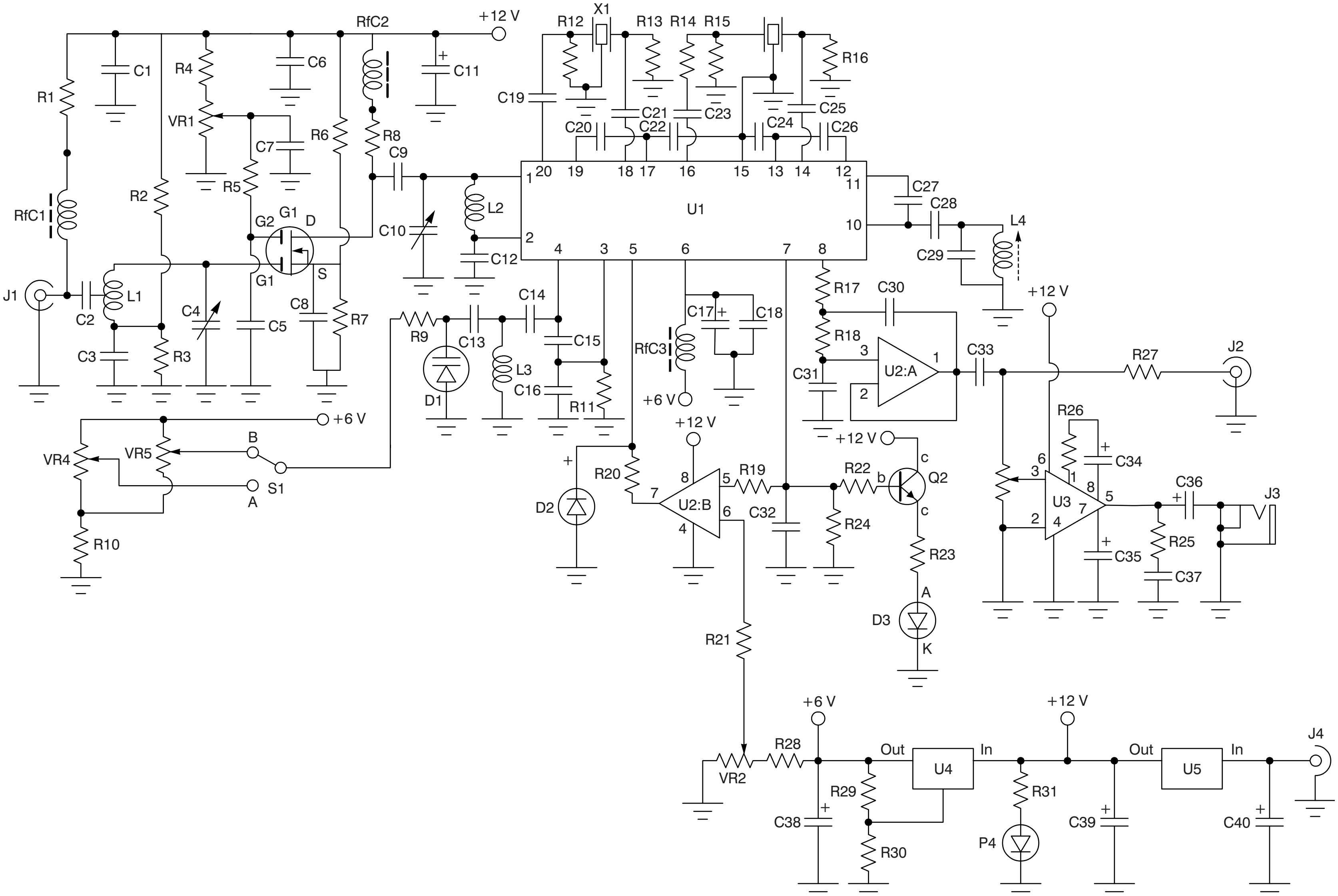
RfC2
+12 V
R12
X1
R13
R14
R15
R16
C1
C6
C11
R4
VR1
R1
C19
C21
C23
C25
C20
C22
C24
C26
C7
R6
R8
C9
R5
R2
RfC1
G2
G1
D
S
L2
C10
C12
U1
1
2
3
4
5
6
7
8
10
11
12
13
14
15
16
17
18
19
20
C27
C28
C29
L4
J1
C2
L1
C3
R3
C4
C5
C8
R7
R9
D1
C13
L3
C14
C15
C16
R11
C17
C18
RfC3
+6 V
R17
C30
R18
C31
U2:A
C33
R27
J2
VR4
VR5
B
S1
A
R10
D2
R20
U2:B
R19
C32
R22
R24
Q2
b
c
R23
D3
K
R26
C34
U3
C35
R25
C36
C37
J3
R21
VR2
R28
U4
Out
In
R29
R30
C38
R31
P4
C39
U5
C40
J4

Figure 23-4 *Weather satellite receiver schematic. Courtesy of Silicon Chip Magazine*

the power supply circuitry includes U5 to provide a regulated and smoothed +12 V supply from an external supply such as a 14.5–18 V "wall-wart." This is followed by 5V regulator U4 which has its output "jacked up" using 300 ohm and 47 ohm resistors to provide close to +6 V for the SA605D (U1). Note that if you want to run the receiver from a 12 V battery, this can be quite easily done by replacing REG1 with a wire link. In addition, the 2200 µF capacitor should be replaced with a 16 V Zener diode (ZD2) for over-voltage protection.

Construction

Before beginning the project make sure that you have a clean well lit and well ventilated work space for constructing your satellite receiver. Locate all the diagrams and place them in front of you on the table or work bench. Locate a 27 to 30 watt pencil-tip soldering iron, along with some "Tip Tinner," which is used to prepare and clean your soldering iron tip. Locate a roll of 60/40 tin/lead rosin core solder as well as a few small hand tools such as a pair of small end-cutters, a needle-nose pliers, a pair of tweezers, and some small screwdrivers.

Construction of the weather satellite receiver is straightforward with virtually all of the parts mounted on a small PC board. The board is double-sided but the top copper pattern is used mainly as a ground-plane. This means that the board doesn't need to have plated-through holes but there are quite a few component leads which do have to be soldered on both sides of the board. There are also a number of "pin-through" or "hole-through" wires which have to be soldered to both sides of the PC board around U1. These connect the Earth patterns on both sides of the board and ensure that this high-gain chip operates in a stable manner.

The various input and output connectors are mounted along the rear edge of the circuit board, while the controls and indicator LEDs mount along the front edge. The only component not actually mounted on the board is S1, the channel select toggle switch. This mounts on the front panel, with its three connection lugs wired to PC board terminal pins directly underneath using very short lengths of insulated hookup wire.

Start the assembly by mounting these three terminal pins first, they are the only pins used in the receiver project. Next, mount the four connectors CON1-CON4 along the rear edge, followed by the resistors. Table 23-1 shows the resistor color codes but it's also a good idea to check each value using a digital multi-meter before

Table 23-1
Resistor color code chart

Color Band	*1st Digit*	*2nd Digit*	*Multiplier*	*Tolerance*
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

soldering it in position. All resistors are mounted to the top of the PC board but note that some of them have one lead soldered to the top copper as well as the bottom copper. Once the resistors are in place, you can use excess resistor leads to allow connection from one side of the board to the other side, since this project uses a two-sided circuit board. Each "pin-through or through-hole" is fitted by simply passing a wire through the hole in the board, then soldering it on both sides and trimming off the excess lead lengths.

The small ceramic capacitors can now all be installed on the left-hand side of the board. Note that some of these also have their "cold" leads soldered on both sides of the board, as indicated by the red dots. Once they're in, install the MKT capacitors and the electrolytic capacitors, making sure that the latter are all correctly orientated. Use the chart shown in Table 23-2 to help locate the small capacitors from the parts pile. Often small capacitors do not have their actual value printed on them but instead use a three-digit code to represent the actual value.

Finally go ahead and mount the trimmer capacitors VC1 and VC2. These should be mounted so that their adjustment rotors are connected to earth (this makes it much easier to align the receiver later). It's simply a matter of orientating them on the board.

Next, we will move on to installing the RF chokes. RFC2 and RFC3 are both pre-wound (10 µH and 68 µH respectively) but RFC1 needs to be wound on an F29 ferrite bead. It's very easy to wind though, because it requires only three turns of 0.25 mm enameled copper wire.

Winding coils

At this point in time, it's a good idea to wind and mount the remaining coils. Table 23-3 gives the winding details of the coils. As shown, L1–L3 are air-cored types, each consisting of five turns of 0.8 mm diameter wire wound on a 5 mm mandrel. Note, however, that L1 is wound using tinned copper wire, while L2 and L3 are both wound using enameled copper wire. Don't forget to scrape off the enamel at each end, so they can be soldered to the board pads.

Coils L1–L3 should all be mounted so that their turns are about 2 mm above the board. After you've fitted L1, don't forget to fit its "tap" connection lead as well. This can be made from a resistor lead off-cut, since it's very short. It connects to a point ⅓ of a turn up from the "cold" (earth) end of the coil, i.e. just above half-way up the side of the first turn.

The final coil to wind is quad detector coil L4. Unlike the others, this is wound on a 4.83 mm OD former with a base and a copper shield can. It's wound from 20 turns of 0.25 mm enameled copper wire and tuned with an F16 ferrite slug. Coil formers with ferrite slugs are difficult to locate, check the Appendix. You could also experiment with a small "junk" IF transformer can. You could try and unwind the existing coil and re-winding your own coil over the plastic former. You must make sure that there is a ferrite slug in the center of the coil former. Once L4 is wound, and installed onto the board, cover it with its shield can. You can next mount the two ceramic filters X1 and X2. These devices can be fitted either way around but make sure that their pins are pushed through the board holes as far as they'll go before you solder them underneath.

The next step is to construct and install the small metal shield around coil L2, C10 and most of the components in the drain circuit of Q1. This shield is U-shaped and measures 20 mm high, with the front and back "arms" 36 mm long and the side section 20 mm long. The bottom edges of all three sides are soldered to the board's ground-plane in a number of places, to hold it firmly in position and to ensure it stays at earth potential.

The two 10-turn trimpots (VR4 and VR5) can now be soldered in position at the front-center of the board. They can then be followed by the three main control pots, which are all 16 mm diameter types. Trim each pot's spindle length to about 9 mm before fitting it and make sure you fit each one in its correct position as they are all different. In particular, note that VR1 and VR3 both have a value of 50 k ohm but VR1 is a linear pot while VR3 is a log type.

After installing the potentiometers, it's a good idea to connect their metal shield cans together and then run a lead to the board's top copper to earth them. This is done using a length of tinned copper wire, with a short length of insulated hookup wire then connecting them to the board copper at front right.

Table 23-2
Capacitance code information

This table provides the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Some capacitors actually have the numeric values stamped on them; however, many are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

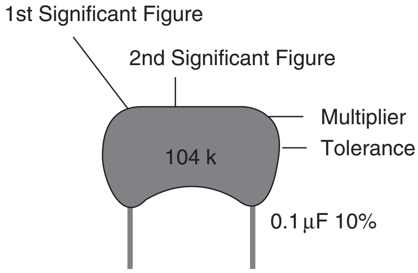

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

Table 23-3

Coil winding details

L1	5 turns of 0.8 mm (20 ga.) tinned copper wire, 7 mm long (wound on 5 mm diameter mandrel). Tap at 0.3 turns. Air core, mounted 2 mm above board.
L2	5 turns of 0.8 mm (20 ga.) enameled copper wire, 9 mm long (wound on 5 mm diameter mandrel). Air core, mounted 2 mm above board.
L3	5 turns of 0.8 mm (20 ga.) enameled copper wire, 8 mm long (wound on 5 mm diameter mandrel). Air core, mounted 2 mm above board.
L4	20 turns of 0.25 mm (30 ga.) enameled copper wire, tight wound on 4.83 mm OD former with F16 tuning slug and shield can.
RFC1	3 turns of 0.25 mm (30 ga.) enameled copper wire wound on an F29 ferrite bead.

Mounting the semiconductors

Before we go ahead and install the semiconductors, let's take a brief look at the semiconductor pin-out diagram illustrated in Figure 23-5, which will assist you in orienting the integrated circuits. Now you can go ahead and mount the semiconductors. Begin by installing the Varicap diode at D1, and the 5.1V Zener diode at D2 and transistor Q2. Next, install regulator U5 and its associated heatsink. These parts are secured to the board using a 6 mm-long M3 screw, nut and lock-washer. Note that U5's center pin should be soldered on both sides of the board but take care not to touch either of the two large adjacent electrolytic capacitors with the barrel of your soldering iron.

Go ahead and mount regulator U4, followed by U22 (TL072) and U3 (LM386). Note that pin 4 on both these devices should be soldered to the copper on the top of the board as well as the bottom. The two D3 and D4 are both mounted horizontally, so that they later protrude through matching 3 mm holes in the front panel. Note that they are both fitted with their cathode leads towards the left. Bend their leads down through 90° about 5 mm from the LED bodies, then solder them in position so that the axis of each LED is 5 mm above the board.

The final components to be mounted on the PC board include the Dual-Gate MOSFET at Q1 and the SA605D IC at (U1), U2 and transistor Q2. These components should be handled very carefully, using an anti-static wrist-band to avoid static charges building up on the components. The two integrated circuits should be mounted in IC sockets. IC sockets are an insurance policy against circuit failure at a later date. It is much easier to simply unplug a faulty IC, rather than trying to un-solder 14 or 16 pins from the circuit. Most people cannot un-solder 16 pins without damaging a circuit board. When mounting the IC sockets, observe a small cut-out at one edge of the socket. The cut-out will identify the top of the socket. Usually pin of the IC will be located to the left of the cut-out. Solder the sockets in their respective locations. Finally insert the ICs into their respective sockets, noting pin 1 is to the left of the small indented circle or cut-out on the top of the IC package. When handling the MOSFET be extremely

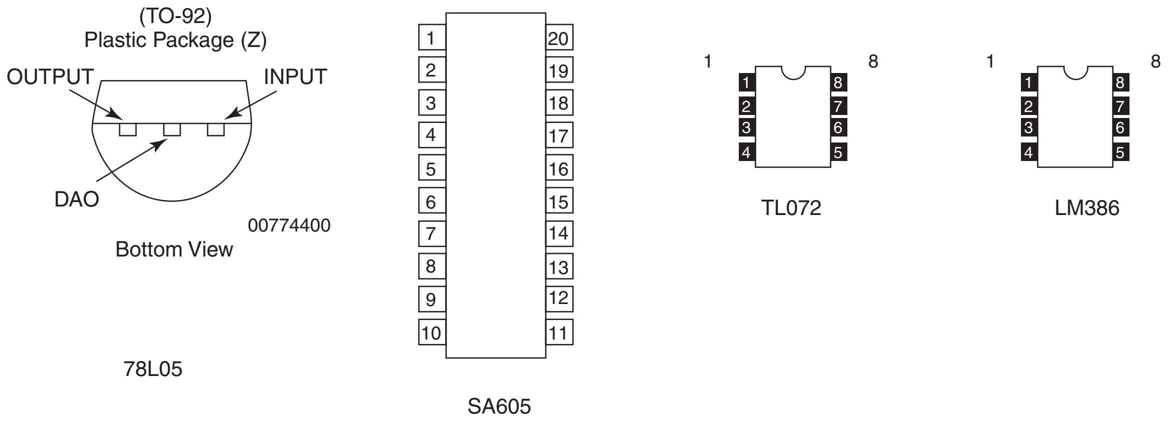

Figure 23-5 *Semiconductor pin-outs*

careful to avoid damage from static electricity. Do not move around when installing these static sensitive components and be sure to use an anti-static wrist band to avoid damage.

The completed PC board is housed in a suitable low-profile plastic instrument case.

First, you can mount switch S1 on the front panel and connect three 30 mm lengths of insulated hookup wire to the three pins on the PC board. That done, the front panel can be mated with the PC board by positioning it on the three pot ferrules and doing up the nuts. The three leads from the PC board pins can then be soldered to the switch lugs.

The rear panel is not attached to the board assembly. Instead, it simply slips over connectors CON1 and CON2 and is then slid into the rear slot when the assembly is fitted into the bottom half of the case. Now go ahead and mount the antenna connector onto the rear panel. Finally, the completed assembly is fastened in place using eight 6 mm-long 4-gauge self-tapping screws.

After completing the circuit board assembly, you should take a short rest and when we return we will inspect the circuit board for any possible "short" circuits or "cold" solder joints. Pick up the circuit board with the foil side of the circuit facing upwards toward you. First inspect the circuit board for possible "cold" solder joints. All of the solder joints should look clean, bright and shiny. If any of the solder joints look dark, dull or "blobby," then you should un-solder that particular joint, remove the solder then re-apply some solder and re-solder the joint all over again. After inspecting all of the solder joints, we can move on to looking for "short" circuits. Often sticky rosin core solder will leave a residue on a circuit board. When excess component leads are cut from the board during assembly, many times the "cut" leads will stick to the circuit board and they can "bridge" across the circuit pads and cause a "short" circuit which could damage the circuit when first powered-up. Another potential cause for a "short" circuit is a small solder ball which may have stuck to the circuit board, this too can cause a "short" circuit.

Once you have inspected the circuit board to your satisfaction, you can connect up the power supply leads to the power input jack and you will be ready to test the weather satellite receiver.

Circuit test

First, turn all three front-panel pots fully anticlockwise, connect up your antenna to the antenna input connector, and then apply power from a suitable 14–18V DC "wall-wart" power supply or a 12V battery. Check that the green power LED immediately begins glowing.

If it does, check the voltage on REG1's output lead (i.e. the right-hand lead) with your multi-meter, it should read close to +12V with respect to ground. Similarly, the voltage at REG2's (right-hand) output pin should measure very close to +6V.

If you now plug an 8 ohm speaker into CON3 terminals, and then turn up audio gain control VR3, you should hear a small amount of hiss and noise. When you turn up the RF gain control VR1 as well, this noise should increase a little further but LED1 shouldn't begin glowing except only very faintly when VR1 is turned fully clockwise.

Now turn VR1 fully anticlockwise again and use your multi-meter to measure the DC voltage at the top of the 390k resistor located just behind transistor Q2 (i.e. to the right of 5.1V Zener diode). The voltage across this resistor should be less than 0.30V and preferably about 0.26V. If it's any higher than 0.30V, the IF amplifier in IC1 may be unstable. Assuming that your receiver has passed all these tests, it should be working correctly and is now ready for alignment.

Receiver alignment

The final alignment of the weather satellite receiver is best done with a satellite signal. However, you need to give it a basic alignment first so that you can at least find the signal from a satellite when it's within range. For the basic alignment, you'll need access to a frequency counter capable of measuring up to 150 MHz and an RF signal generator which can be set to give an output at 137.50 MHz and at 137.62 MHz. It should be able to provide either unmodulated (CW) output or frequency modulation, with a modulating frequency of 2.4 kHz and a deviation of ±25 kHz or thereabouts. If the generator can't be accurately set to the above frequencies, you'll need to use the frequency counter to help set its frequency. You'll also need your multi-meter during the "tuning-up" process, to monitor received signal level.

The first step is to set the local oscillator frequencies for the two reception channels. This is done by adjusting trimpots VR4 and VR5 respectively, while measuring the oscillator's frequency with the frequency counter. The oscillator signal is coupled to the counter via a "sniffer" coil which is connected to the end of a coaxial cable. The other end of this cable is then connected to the counter's input. Note that there is no direct physical connection between the oscillator coil and the counter's sniffer coil. Instead, the sniffer coil is placed about 9 mm in front of oscillator coil (L3) and roughly on-axis (i.e. just in front of the 10 pF capacitor).

The sniffer coil can be made by winding four turns of 0.8 mm enameled copper wire on a 5 mm drill shank. Its ends can then be soldered to a BNC socket which is then connected to the end of the counter input cable. This arrangement picks up enough oscillator energy to give reliable counter readings, without needing to be any closer to L3 (to avoid "pulling" the frequency).

Assuming you want to receive the NOAA satellites, set the oscillator frequency for channel A to 132.0 MHz (using VR4), and the frequency for channel B to 132.12 MHz (using VR5). These correspond to reception frequencies of 137.5 MHz for NOAAs 12 and 15 and 137.62 MHz for NOAA 17. If you want to try for other satellites, you'll need to find out their APT frequency and set the oscillator frequency to 5.5 MHz below that figure instead, and adjust the tuning controls.

Once the oscillator frequencies have been set, the next step is to peak up the RF stage tuned circuits. This is done by setting your RF signal generator to produce an unmodulated (CW) signal at 137.5 MHz, initially with a level of about 30 mV. That done, connect the generator's output to the antenna input of the receiver, using a series DC blocking capacitor if the generator doesn't have one (so that the generator doesn't short out the +12V phantom power for the masthead amplifier).

Next, connect your multi-meter (set to the 5V DC range) across the 390k ohm resistor just behind Q2 and make sure switch S1 is set to the channel A position. Now turn up RF gain control VR1 to about midway and use an alignment tool or a very small screwdriver to adjust trimmer cap C10 until you find a peak in the voltage reading on a multi-meter. If you can't find a peak, you may have to pull the turns on coil L2 slightly further apart to reduce its inductance. Once the peak is found, adjust VC2 carefully to maximize the meter reading (the multi-meter is reading the RSSI voltage from IC1, so it's essentially showing the received signal strength).

When you're certain that the L2/C10 circuit is tuned to 137.5 MHz, check the actual voltage reading of the meter. If it's more than 2.5V, reduce the output level from the RF generator until the multi-meter reading drops to about 2.0V. Now you're ready to peak the receiver's input tuned circuit—i.e. L1 and VC1. This is done in exactly the same way as for L2 and C10. Just adjust C4 slowly until the multi-meter indicates a peak and then carefully set C4 for the maximum peak reading. If you can't find another peak, you may need to pull the turns of L1 slightly further apart as before.

The final alignment step is to set the slug in quadrature detector coil L4 to the correct position for optimum FM demodulation of the 5.5 MHz IF signals. This is done by first switching the signal generator so that it's still producing a 137.5 MHz signal but this should now be frequency modulated—preferably with a 2.4 kHz tone and a deviation of about ±25 kHz.

Now, connect an 8 ohm speaker to the receiver's speaker socket (CON3) and turn up the audio gain control (VR3) to about the 10 o'clock position. You may not be able to hear the 2.4 kHz modulating signal at this stage but in any case, slowly and carefully adjust the slug in L4 using a non-magnetic alignment tool. Sooner or later you'll start to hear the 2.4 kHz tone and you should also be able to tune the coil for maximum audio level and minimum distortion and noise.

Once this has been done, the basic alignment of your weather satellite receiver is finished and it's ready for final alignment using the signals from a weather satellite. But before you'll be able to do this, you'll need to build or buy a suitable antenna and perhaps a small masthead pre-amplifier to boost the signal if necessary. The best antenna for APT weather satellite reception is either a turnstile or quadrifillar antenna.

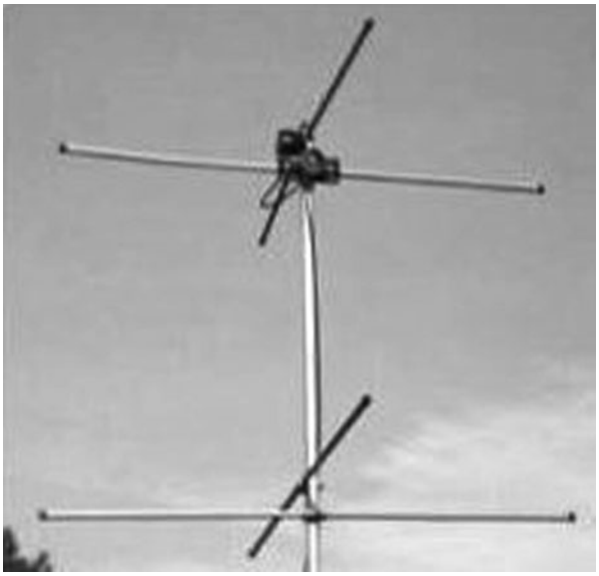

Figure 23-6 *VHF turnstile antenna*

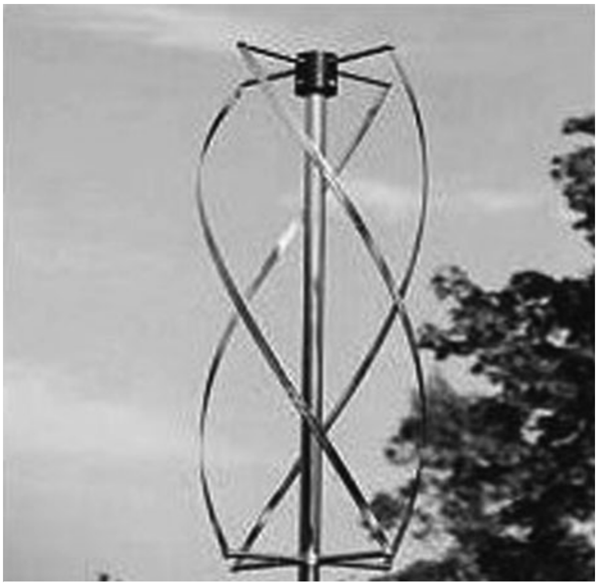

Figure 23-7 *Quadrifillar antenna*

APT antenna considerations

The crossed dipole or turnstile antenna is shown in Figure 23-6; it is recommended for nearly all APT weather satellite applications. It is made from aluminum and is strong and durable. It can be fixed onto a vertical mast and should be positioned as high as possible. The satellite signal will not go through trees or buildings and to get the best possible continuous signal it needs to be mounted onto a mast as high as possible. On the end or top of a building is preferred although, for test purposes, it can simply be stuck into the grass at ground level.

Another good choice for APT satellite reception is the quadrifilar helix antenna, see Figure 23-7. It offers slightly better performance than the crossed dipole but it is considerably more costly.

The success of weather satellite reception is very much dependent on having a good antenna. The ARRL *Weather Satellite Handbook*, is a good source of information on building antennas and low-cost antenna pre-amplifiers. You can either build a good antenna or readily purchase one from one of the vendors listed in the Appendix. Good quality, low-loss 50. coax should be used between the antenna and the satellite receiver, since satellite signals are generally weak. The coax should be plugged into the BNC jack on the rear of the receiver. If you are using large diameter coax, it may be necessary to use a UHF-BNC adapter. Another alternative is to use a short length of smaller diameter, more flexible, coax between the main cable and the receiver. That cable can have a BNC plug on one end and a UHF plug and double female adapter on the other. If you use a 137 MHz pre-amplifier, you can overcome some cable losses by installing the pre-amp right up at the antenna. This establishes a low noise figure before going through the loss of the cable. If you find you do need a 137 MHz mast-head pre-amplifier, you can build one or buy one from Hamtronics or Ramsey Electronics. We do not recommend trying to feed B+ to power the pre-amp through the coax as some enthusiasts do. It doesn't pay to try to save the small cost of lightweight hook-up wire and de-grade the performance of the pre-amp in the process.
Run a separate wire pair to power the pre-amplifier if needed.

Receiver operation

Operation of the receiver is fairly obvious, but we will comment on several features of interest. Normally, when used to drive a computer's sound card, you will want to listen on a speaker to hear what the signal sounds like. The VOLUME control can be adjusted to whatever level is comfortable for listening and proper for the computer's sound card. This is especially true during the initial acquisition period. However, most times you probably want the ability to mute the speaker; so installing a switch in line with the receiver output to the speaker is handy.

To adjust the SQUELCH control, turn it clockwise just a little beyond the point where the squelch closes with no signal (just noise). The squelch circuit has adjustable sensitivity—the further clockwise you set the control, the stronger the signal must be to open the squelch. This is handy to set how strong you want the satellite signal to be before the tape recorder turns on. Be careful not to set it too far clockwise, though, because the squelch might not open even with a strong signal. The weather satellite receiver will allow you to select two different channels. Channel selection is done with switch S1. Use the selector switch to choose channel A or B, then use potentiometer VR4 or VR5 to "tune-in" the desired satellite frequency. The satellite receiver's audio is fed to the speaker output at J3 and also to a low-level output at J2 which can be fed to either a tape recorder or decoder card in a personal computer.

Setting up you APT receive station

In order to set up your own weather satellite receiving station you will have to acquire a few more items, as seen in Figure 23-8. In addition to the satellite receiver and antenna, you will need a personal computer, some sort of interface and a software package to display the satellite picture and a satellite tracking program if desired. A tracking program is not required, but it is fun to be able to know in advance when the satellite is overhead.

How to see satellite weather pictures on your personal computer, you must follow the steps below:

1. Track the desired satellite to find when it will appear and how long it will be in view (i.e. how high in elevation it will rise). (See NASA's online J-Track program for an example of NOAA satellite tracking.)
2. Tune the receiver to the right frequency, e.g. 137.620 MHz for NOAA-14 and 137.500 for NOAA-15 and pick off the AM signal that carries the image data. The 2400 Hz AM sub-carrier signal will produce a distinctive "tick-tock" sound from the speakers.

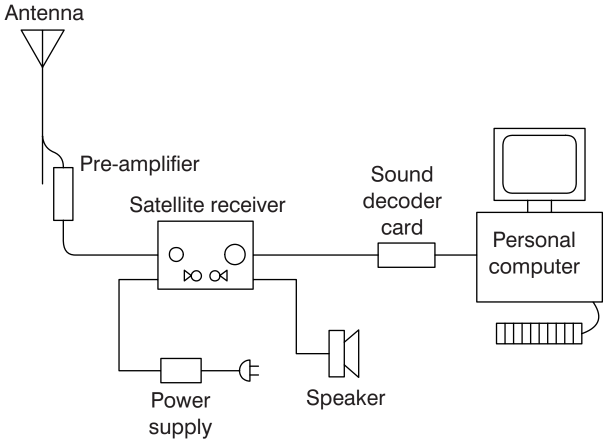

Figure 23-8 *Weather satellite receiving system diagram*

3. Receive the signal for as much of the satellite pass as possible.
4. Decode the signal into the image. (Depending on the system, this is done in real-time or the signal is saved to file that is decoded later.) Most new computers (i.e. running at 800 MHz to 3 GHz with WIN2000 or XP software) with a good sound card can be used to decode the software directly from the satellite receiver audio.

With your 137-138 MHz turnstile or quadifillar satellite antenna connected to your weather satellite receiver, and your WX receiver connected to your sound card in your PC, via a shielded audio cable, you will be able to get stunning color images directly from weather satellites.

In order to decode an APT weather satellite signal from your WX satellite receiver, you will need to load some WX decoding software into your personal computer.

There are number of choices for APT satellite decoding software, some of which are free for the download, such as WXtoImg which is a fully automated APT and WEFAX weather satellite (wxsat) decoder. The software supports recording, decoding, editing, and viewing on all versions of Windows, Linux, and

Table 23-4

137 MHz Weather satellite receiver frequencies

Country	*Satellite*	*Format*	*Frequency*
USA	NOAA - 12	APT	137.500 MHz
USA	NOAA - 14	APT	137.620 MHz
USA	NOAA - 15	APT	137.500 MHz
USA	NOAA - 17	APT	137.620 MHz
USA - 2008 launch	NOAA - 18	APT	137.9125 MHz
Russia	METEOR	APT	137.300 MHz
Russia	METEOR	APT	137.400 MHz
Russia	METEOR	APT	137.850 MHz
Russia	OKEAN	APT	137.400 MHz

MacOS X. The WXtoImg software supports real-time decoding, map overlays, advanced color enhancements, 3-D images, animations, multi-pass images, projection transformation (e.g. Mercator), text overlays, automated web page creation, temperature display, GPS interfacing, and control for many weather satellite receivers, communications receivers, and scanners. Check the download site at: http://www.wxtoimg.com/downloads/.

Another good source for the WXSAT satellite decoding software is the HF-FAX web-site located at http://www.hffax.de/html/hauptteil_wxsat.htm. Once your software is loaded and your system check completed, you can refer to the chart in Table 23-4, for the most popular satellite frequencies.

Additional APT weather satellite source information

137 MHz quadrifillar antenna construction info: http://www.askrlc.co.uk/

137 MHz turnstile antenna construction info: http://www.qslnet.de/member/pa1pj/turnstile.pdf

TimeStep – http://www.time-step.com

Hamtronics – http://www.hamtronics.com/

Ramsey Electronics – http://www.ramseyelectronics.com/

Dartcom – http://www.dartcom.co.uk

Swagur Enterprises, USA www.swagur.com

Vanguard Electronic Labs, USA www.cyberpad.com

Chapter 24

Analog To Digital Converters (ADCs)

Parts list

ADC - 1

C1	.1 μF, 35 volt capacitor
C2	10 μF, 35 volt electrolytic capacitor
C3	5 μF, 35 volt electrolytic capacitor
U1	LM78L05 5 volt regulator IC
U2	MAX187 A/D converter IC

ADC - 2

R1-R8	100 k ohm, 1/4 watt, 5% resistor
C1	1 μF, 35 volt electrolytic capacitor
C2,C5	.1 μF, 35 volt disk capacitor
C3	47 μF, 35 volt electrolytic capacitor
C4	4.7 μF, 35 volt electrolytic capacitor
D1	1N4001 silicon diode
U1	MAX186 A/D converter IC (Maxim Semiconductor)
U2	LM7805 – 5 volt regulator IC
P1	DB-25 – 25-pin parallel port connector
Misc	PC board, wire, enclosure, terminal strip, etc.

A number of radio receiver projects in this book are based on collecting data from receivers rather than listening to sound coming from a radio. In this chapter we will discuss four different approaches to buying and/or building analog to digital converters which you can use to collect or save data for analyzing at a later time. By using an analog to digital converter, you can collect, save and later playback and analyze signals that were recorded over days or over weeks and then, using special software, you can analyze the signals to determine trends of specific signals. We will present two different A/D or analog to digital converter projects which you can build with inexpensive integrated circuits and connect to your personal computer's parallel port. Next, we will show you how you can purchase a low cost commercial analog to digital converter with software and finally we will describe a high quality A/D system which is used by members of the Public Seismic Network to record earthquake signals around the country. The PSN A/D converter is a true 16-bit multi-channel assembled A/D system which connects to your personal computer's serial port. These four A/D systems are great choices for collecting radio signals over an extended period of time. Examine each of these A/D solutions and see which one is best for your particular applications. Most of the A/D collections systems described below come with free software for recording data.

ADC – 1: single channel analog to digital converter

This single channel 12-bit analog to digital converter, shown in Figure 24-1, can be built with the minimum of

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

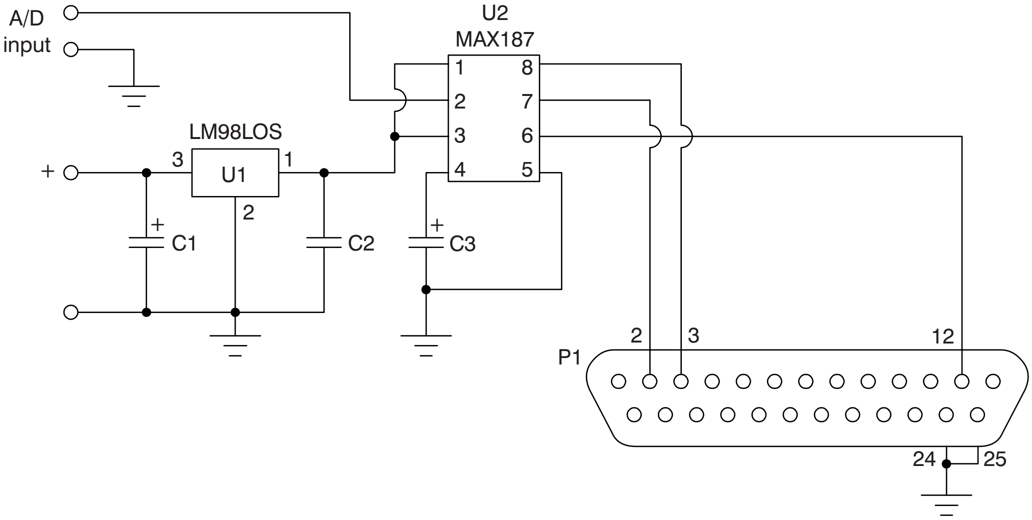

Figure 24-1 *Single channel A/D converter – ADC 1. Courtesy of Radio-Sky Publishing*

components and the software for it is free. The single channel A/D converter is built around the $14 low cost MAX187 A/D converter from MAXIM Semiconductors. The circuit consists of the A/D converter chip, a regulator chip, and a few capacitors; it can't get more simple than this! The single channel A/D converter readily connects to your computer's parallel port through a BD-25 connector. The analog input voltage can be from zero (0) to 4.095 volts as a maximum signal level. A 9 volt battery or small 9 to 12 volt DC "wall-wart" power supply can be used to power the A/D circuit, which can be housed in a small plastic case.

The A/D input from your signal source is fed to pin 2 of the MAX187 chip. The 5 volt regulator's output powers the MAX187 chip via pin 1. The MAX817 outputs to your personal computer are on pins 6, 7 and 8. Only three capacitors are needed for this A/D, two of which are used in the regulated power supply section of the circuit.

You will need a small 30 watt soldering iron, some 60/40 rosin core solder for building this project. A small jar of "Tip Tinner" from your local Radio Shack is used to clean and dress the soldering tip. Locate a few small hand tools, including a pair of small end-cutters and a needle-nose pliers and a flat-blade and Phillips screwdriver and we will begin building the A/D converter.

Heat up your soldering iron and let's get building. The ADC-1 circuit can be built on a small perf-board or direct prototyping board available at Radio Shack or you could make your own circuit board for the project. The circuit only utilizes capacitors, both polar and non-polar types. Non-polarized capacitors are usually smaller capacitor and sometimes they do not have their actual value printed on them. Often a three-digit code is used to represent their value. For example, a . 001 µF capacitor will have (102), while a .01 µF capacitor will have a (103) marked on it. This project also uses large polarized or electrolytic capacitors, these capacitors have polarity markings on them which must be observed if you want the circuit to work properly. Use the chart in Table 24-2 to help identify the project capacitors. Identify all of the capacitors before installing them on the circuit board. Once you have identified all of the capacitors, you can go ahead and install them on the circuit board. Trim the excess component leads flush to the edge of the circuit board with your end-cutters.

Take a quick look at the semiconductor pin-out diagram shown in Figure 24-2. This diagram will help you install the integrated circuits in the A/D projects. An integrated circuit socket for this project is highly recommended, as an insurance from a possible circuit failure if it ever happens. It is much easier to simply replace the IC rather than trying to un-solder the IC from the circuit board. The IC socket will have a cut-out or notch at one end of the package. Solder the IC socket onto the PC board. Pin one (1) will be assigned to the pin just to the left of the notch or cut-out. Pin one (1) of the IC socket is connected to the

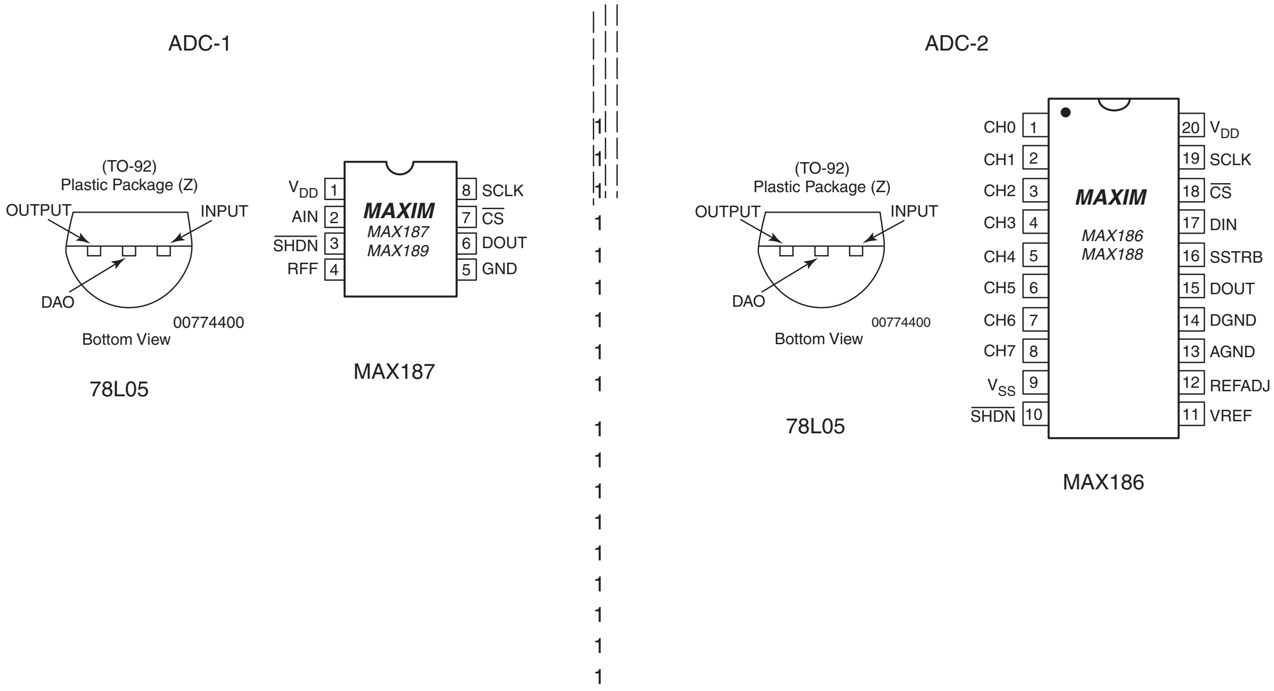
ADC-1
(TO-92)
Plastic Package (Z)
OUTPUT
INPUT
DAO
00774400
Bottom View
78L05
VDD 1
AIN 2
SHDN 3
RFF 4
MAXIM
MAX187
MAX189
8 SCLK
7 CS
6 DOUT
5 GND
MAX187
ADC-2
(TO-92)
Plastic Package (Z)
OUTPUT
INPUT
DAO
00774400
Bottom View
78L05
CH0 1
CH1 2
CH2 3
CH3 4
CH4 5
CH5 6
CH6 7
CH7 8
VSS 9
SHDN 10
MAXIM
MAX186
MAX188
20 VDD
19 SCLK
18 CS
17 DIN
16 SSTRB
15 DOUT
14 DGND
13 AGND
12 REFADJ
11 VREF
MAX186

Figure 24-2 *Semiconductor pin-outs*

output of the regulator U1. When inserting the IC into its socket pay particular attention to the orientation. Pin one (1) of the IC must be inserted into pin one (1) of the IC socket. Note that there will be a notch or cut-out or small indented circle on the top of the IC package. Pin one (1) of the IC will be just to the left of the cut-out or notch.

One neat way to package the units is inside a two way parallel port manual switch box. These can be bought for about $10 at computer shows. One of the DB25 connectors is removed from the back of the switch and replaced with a power and a analog input jack. Leads 2, 3, 12, 24 and 25 from the removed DB25 are soldered directly to the appropriate places on the A to D board. The switch can then be used to pass through to the printer when the A to D is not in use. You could also elect to mount the A/D converter in a small plastic box if desired with a 25 pin connector on the rear of the enclosure.

Before applying power to the A/D circuit, take a moment to inspect the foil side of the PC board. Pick up the PC board with the foil side of the board facing upwards towards you. We are going to examine the PC board for possible "cold" solder joints and "short" circuits before applying power for the first time. Look carefully at the solder joints, they should all look clean, shiny and bright. If any of the solder joints look dark, dull or "blobby" then remove the solder from the joint and then re-solder the joint all over again. Next, we will examine the PC board for possible "short" circuits which are often caused by the sticky rosin residue left on the circuit board from the solder. Often "cut" components or small solder balls can adhere to the foil side of the board and "short" between the circuit traces. Once you have examined the PC board carefully, you can go ahead and power-up the circuit for the first time. Install the free A/D collection software, either Data Collect Lite or Radio-Sky Pipe from the Radio-Sky web-site, or see Appendix, and start collecting data from one of your new receivers. In the event the circuit does not work, check over the installation of the electrolytic capacitors and the polarity of the power supply connections. The MAX87 chip should be available from Digi-Key Corporation, see Appendix.

ADC – 2: 8 channel analog to digital converter

An 8 channel 12-bit analog to digital converter (ADC) project is shown in Figure 24-3, and connects to the parallel port of your personal computer. This second low cost analog to digital converter circuit is one you can readily build yourself in a few hours from scratch. This A/D converter works with free software from a company known as Radio-Sky, which specializes in products for Radio Astronomy. Their free software, which works with the low cost ADC, is called Radio-Sky Pipe. The ADC is based on a single low cost integrated circuit from MAXIM Semiconductors. Pins 1 through 8 on the MAX186 A/D converted chip correspond to Channels 1 through 8. A 100 kilo-ohm resistor is placed across each input channel pin and is connected to ground. The A/D converter circuit connects to a personal computer via the parallel port. You will need to choose the appropriate parallel port address which is selected in the free Radio-Sky Pipe software. Note that all of the .1 µF (microfarad) capacitors are non-polarized types, typically disk ceramic capacitors. They should be rated for at least 15 volts. The A/D circuit may be powered by a battery or any well filtered external DC power supply. (You may order the MAX186 IC from Radio-Sky Publishing, or from Digi-Key or Mouser Electronics, see Appendix.)

The A/D circuit can be built on a small breadboard or on a printed circuit board. A circuit board and partial kit of parts (no case, connectors, or power supply) is available from Radio-Sky. Before we begin building the single channel A/D converter you will need to secure a clean well lit work area. You will need to locate a small 27 to 33 watt pencil-tipped soldering iron, a roll of 60/40 rosin core solder and small jar of "Tip Tinner," a soldering iron tip cleaner/dresser. Collect together a few small hand tools such as a pair of end-cutters, a pair of needle-nose pliers, a flat-blade and a Phillips screw-driver and a magnifying glass. Grab the schematic, parts layout diagram and the resistor and capacitor identifications charts below, and we will begin.

This project employs eight 100,000 ohm resistors at the input to each of the A/D channels. Take a look at the resistor identification chart in Table 24-1. Each resistor will have three or four color bands on the body of the resistor, which start at one end of the resistor body.

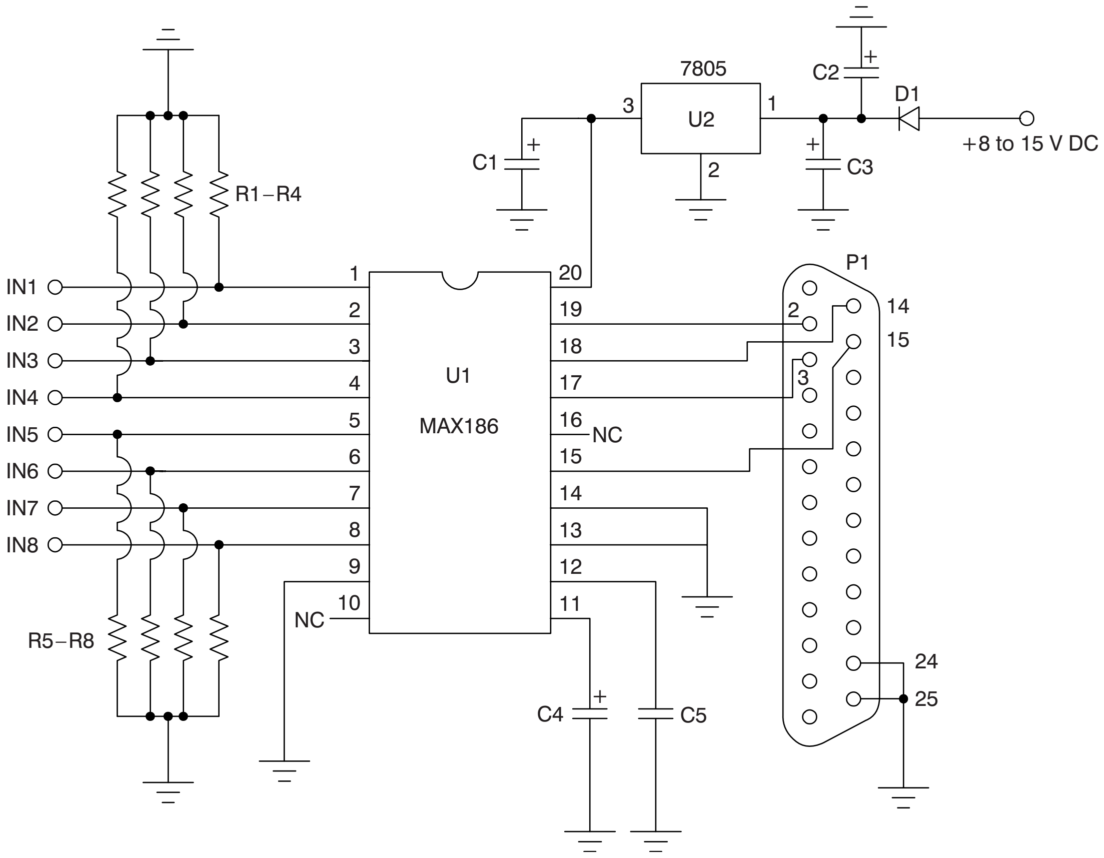

Figure 24-3 *8 channel A/D converter – ADC 2. Courtesy of Radio-Sky Publishing*

Table 24-1

Resistor color code chart

Color Band	1st Digit	2nd Digit	Multiplier	Tolerance
Black	0	0	1	
Brown	1	1	10	1%
Red	2	2	100	2%
Orange	3	3	1,000 (K)	3%
Yellow	4	4	10,000	4%
Green	5	5	100,000	
Blue	6	6	1,000,000 (M)	
Violet	7	7	10,000,000	
Gray	8	8	100,000,000	
White	9	9	1,000,000,000	
Gold			0.1	5%
Silver			0.01	10%
No color				20%

The first color band will represent the first digit of the resistor code, the second color band denotes the second digit of the code, and finally the third color code represents the resistor's multiplier value. Resistors will generally have a tolerance code which is the fourth color band. The absence of a fourth color band indicates a 20% tolerance value, while a silver color band denotes a 10% tolerance and a gold band represents a 5% tolerance resistor. A 100k ohm resistor's first band should be brown (1), its second color band is black (0) and its third band will be yellow (0000), so (1)(0) times (0000) = 100,000 ohms or 100k ohms. Place the eight resistors on the circuit board and solder them in place. Next, trim the excess component leads flush to the edge of the PC board with your end-cutters.

Capacitors are generally identified as polar and non-polar types. The polarized capacitors are also called electrolytic capacitors and they have polarity, as each of the two leads are labeled as plus (+) or minus (–) and this polarity must be observed when installing these capacitors in order for the circuit to work correctly. Non-polarized capacitors are usually small in physical size and in their actual value. Take a look at the capacitor identifier chart in Table 24-2, it illustrates the capacitor codes used to identify smaller sized capacitors. Many times small capacitors do not have room to print their entire details on the body of the capacitor, so a three-digit code is utilized to label them. For example, a .001 µF is labeled (102), while a .01 µF capacitor is labeled as (103). Identify all of the capacitors in the ADC-2 project and separate them into piles polarized vs. non-polarized and place them in front of you. Once identified you can go ahead and install the capacitors on the PC board in their respective locations. Pay particular attention to installing the electrolytic capacitors correctly. Once you have placed the capacitors on the board, you can solder them in place, then cut the extra component leads flush to the edge of the board.

The analog to digital converter contains a single silicon diode at D1. Remember that diodes have polarity and they must be mounted correctly for the circuit to work. Diodes will generally have an anode and cathode lead, they will have a white or black band closer to one end of the diode. The end closest to the colored band is the cathode lead, look at the schematic or part layout diagram and then install the diode.

This circuit employs two integrated circuits, the main A/D converter chip and a regulator chip. An integrated circuit socket for U1 is highly recommended, so in the event of a possible IC failure, you can simply plug-in a new chip without having to un-solder 20 pins or more from the foil side of the PC board. Integrated circuit sockets will usually have a notch or cut-out at one end of the package. The pin just to the left of the notch or

Table 24-2

Capacitance code information

This table provides the value of alphanumeric coded ceramic, mylar and mica capacitors in general. They come in many sizes, shapes, values and ratings; many different manufacturers worldwide produce them and not all play by the same rules. Some capacitors actually have the numeric values stamped on them; however, many are color coded and some have alphanumeric codes. The capacitor's first and second significant number IDs are the first and second values, followed by the multiplier number code, followed by the percentage tolerance letter code. Usually the first two digits of the code represent the significant part of the value, while the third digit, called the multiplier, corresponds to the number of zeros to be added to the first two digits.

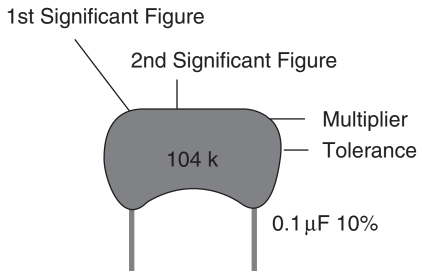

CSGNetwork.Com 6/4/92

Value	*Type*	*Code*	*Value*	*Type*	*Code*
1.5 pF	Ceramic		1,000 pF /.001 µF	Ceramic / Mylar	102
3.3 pF	Ceramic		1,500 pF /.0015 µF	Ceramic / Mylar	152
10 pF	Ceramic		2,000 pF /.002 µF	Ceramic / Mylar	202
15 pF	Ceramic		2,200 pF /.0022 µF	Ceramic / Mylar	222
20 pF	Ceramic		4,700 pF /.0047 µF	Ceramic / Mylar	472
30 pF	Ceramic		5,000 pF /.005 µF	Ceramic / Mylar	502
33 pF	Ceramic		5,600 pF /.0056 µF	Ceramic / Mylar	562
47 pF	Ceramic		6,800 pF /.0068 µF	Ceramic / Mylar	682
56 pF	Ceramic		.01	Ceramic / Mylar	103
68 pF	Ceramic		.015	Mylar	
75 pF	Ceramic		.02	Mylar	203
82 pF	Ceramic		.022	Mylar	223
91 pF	Ceramic		.033	Mylar	333
100 pF	Ceramic	101	.047	Mylar	473
120 pF	Ceramic	121	.05	Mylar	503
130 pF	Ceramic	131	.056	Mylar	563
150 pF	Ceramic	151	.068	Mylar	683
180 pF	Ceramic	181	.1	Mylar	104
220 pF	Ceramic	221	.2	Mylar	204
330 pF	Ceramic	331	.22	Mylar	224
470 pF	Ceramic	471	.33	Mylar	334
560 pF	Ceramic	561	.47	Mylar	474
680 pF	Ceramic	681	.56	Mylar	564
750 pF	Ceramic	751	1	Mylar	105
820 pF	Ceramic	821	2	Mylar	205

cut-out will be designated as pin 1. Be sure to align pin one (1) of the IC socket with IN1 and pin twenty (20) with the output of the regulator at U2. The actual IC itself will also have a notch or cut-out and pin one (1) will also be just to the left of the notch. Make sure that when you place the IC into its socket that you align pin one (1) of the IC with pin one (1) of the socket. Failure to observe the correct orientation of the IC will result in damage to the IC and perhaps damage to the circuit when power is first applied.

Once the circuit board is built, you can wire a DB-25 connector to the A/D chip and ground and your analog to digital converter will be just about finished. Find a suitable small plastic or metal enclosure for the circuit board. The 10 inputs can be wired to a 10-position screw terminal strip or you could use the second 25 pin connector as the input connector to the circuit. Use whatever is convenient.

Before applying power to the circuit for the first time, you will need to inspect the circuit board for possible "cold" solder joints and "short" circuits. Pick up the circuit board, with the foil side facing upwards towards you. Examine the solder joints carefully. Each of the solder joints should look clean, shiny and bright. If any of the solder joints look dark, dull or "blobby", then remove the solder from the joint and then re-solder the joint over again so that it looks good. Next, we will inspect the board for possible "short" circuits. Many times sticky residue from the rosin core solder will cause "cut" component leads and small solder balls to adhere to the foil side of the board, thus "shorting-out" circuit traces. Remove anything that looks foreign.

Now you can connect the A/D circuit to the parallel port of your computer and apply power to the circuit. You can power the A/D converter from a small DC "wall-wart" power supply. Install the free Radio-Sky software and you are ready to begin collecting data. The A/D circuit will work in the Standard version of Radio-Sky Pipe, but only Channel 1 will be available. The Pro-Upgrade allows you to use any number of channels up to 8. The channels used must be sequential. Analog inputs should be restricted to the 0 to 4.095V range.

If the circuit doesn't work on power-up, disconnect power and re-examine the circuit board to make sure that you didn't make an error when constructing the circuit. Most common errors are installing the capacitor and diodes incorrectly. Also take note of the IC installation if the circuit does not initially work.

ADC – 3: Dataq DI-194RS/ DI-154RS starter kits

The Dataq DI-194RS is a low cost ready-made 10-bit ADC starter kit available for the low price of $24.00, while the **DI154RS** is a 12-bit ADC system priced at $39.00, see Figure 24-4. This A/D system has many features and includes a copy of WinDaq/Lite chart recorder software. This is a great budget ADC system for readers who want to begin collecting data immediately. These ADCs would go well with any of the receiver projects in this book.

4-Channel, 10-Bit, data acquisition starter kit - DI194RS

4-Channel, 12-Bit, data acquisition starter kit - DI154RS

- Compact Data Acquisition Starter Kit
- 4-Channel, 10-Bit, ±10V ADC
- Provided with a Serial Port Interface Cable

The WinDaq Resource
DATAQ INSTRUMENTS
DI-194RS
DATAQ INSTRUMENTS

Figure 24-4 *Dataq data-logger*

- Two Digital Inputs for Remote Start/Stop and Remote Event Marker Control
- Includes WinDaq/Lite Chart Recorder Software, WinDaq Waveform Browser Playback and Analysis Software, and Documentation
- Provided Active X Controls allow you to program the DI-194 from any Windows programming environment
- Kids - Get a free Data Acquisition Starter Kit for your Science Fair Project.

The DI-194RS and DI-154RS, DATAQ's four-channel data acquisition starter kits, are a low-cost way to experience WINDAQ software, now with more features and capabilities than ever before. With these starter kits, you can digitize virtually any transducer's analog output signal and record it to your PC's hard disk. At the same time, view the transducer's output on your PCs monitor in a triggered sweep (oscilloscope-like) or scrolling (chart recorder-like) display format. The starter kits provide a taste of the exceptional power and speed possible with WINDAQ software. They provide 12-bit (DI-154RS) and 10-bit (DI-194RS) measurement accuracy, a ±10V analog measurement range, up to 240 samples/second throughput, and four analog input channels. A CD demonstrates WINDAQ Waveform Browser (WWB), our playback and analysis software, but to get a hands-on illustration of the data recording and display capabilities of WINDAQ/Lite, you need a DI-154RS or DI-194RS starter kit. When connected to your PC's serial port, these starter kits allow you to record, display, and analyze data using your own signals. The kits ship with WINDAQ/Lite (Recording Software and Playback and Analysis software). Data acquisition rates up to 240 samples per second are supported for Windows 2000 and XP.

The DI-154RS and DI-194RS derive their power directly from the RS-232 serial port to which it is connected. So there are no batteries to replace or external power supplies to connect. The DI-154RS and DI-194RS starter kits are equipped with two digital inputs for remote start/stop and remote event marker control. Both starter kits are provided with an ActiveX Control Library that allows you to program the starter kits from any Windows programming environment.

The included WINDAQ/Lite data acquisition software offers real time display and disk streaming for the Windows environment. Their real time display can operate in a smooth scroll or triggered sweep mode of operation, and can be scaled into any unit of measure. Event markers with comments allow you to annotate your data acquisition session with descriptive information as you're recording to disk. Raise your productivity to new heights with WINDAQ's unique multi-tasking feature. Record waveform data to disk in the background while running any combination of programs in the foreground—even WINDAQ Waveform Browser playback software to review and analyze the waveform data as it's being stored! The DI-194RS Module with the WINDAQ Resource CD, screwdriver, and cable—all included with each starter kit. DI-154RS module case is identical.

DI-194RS Starter Kit:

http://www.dataq.com/products/startkit/di194rs.htm

DI-154RS Starter Kit:

http://www.dataq.com/products/startkit/di154rs.htm

ADC – 4: PSN true 16-bit serial output analog to digital converter

The Public Seismic Network analog to digital board is a true 16-bit analog to digital converter for serious researchers who want high speed and high accuracy and excellent resolution. This ADC is used by the Public Seismic Network, a group of dedicated researchers based in California who collect earthquake events. Many researchers world-wide use this equipment to collect data. PSN ADC's are used to collect seismic, magnetic, radio and weather information. A free data collection software package is also available. The PSN board costs about $180, see Figure 24-5.

True 16-bit PSN A/D board

16-Bit ADC board features:

- Low Cost Analog to Digital Card for Seismic and Other Low Frequency Data Recording
- True 16-Bit Resolution with Oversampling (more information)

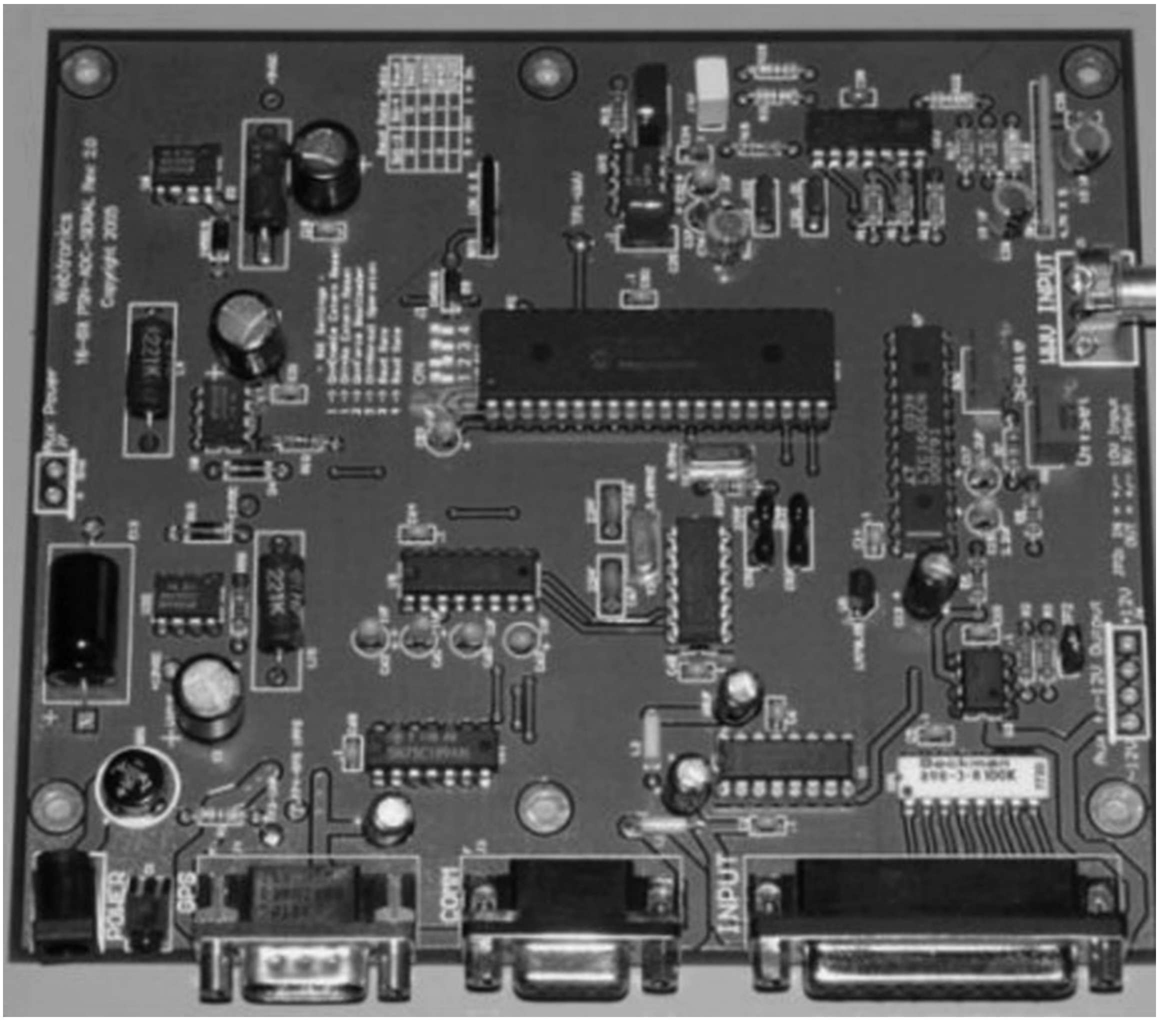

Figure 24-5 *16-bit analog to digital converter*

- Standalone ADC Board with RS-232 Serial Output
- 8 Single-Ended Analog Input Channels
- Compatible with WinSDR Datalogging Software
- Time Accuracy < 3 Millisecond with GPS Timing System
- Uses GPS, WWV, WWVB and Other Time References for Accurate Time Keeping
- Low power Requirement: < 1 Watt without Timing Reference; < 1.5 Watts with GPS Timing
- Uses Low Power High Speed PIC18F452 Microprocessor

Specifications

- Analog Inputs:
    - Channels: 8 Single-Ended Inputs
    - Resolution: 16-Bit Successive Approximation
    - Accuracy: 16 Bits
    - Dynamic Range: 96 dB
    - Input Range: ± 5 or ±10 Volts, Jumper Selectable
    - A/D Converter Chip: Linear Technology LTC1605CN
    - Input Impedance: 4.7k or 100k (Note 1)
- Sample Rates:
    - 200, 100, 50, 20 and 10 Samples Per Second
- WWV Time Correction Option:
    - Tone Detector and Bandpass Filter to Detect 800 ms Top of the Minute Tone Mark
    - ±20 Millisecond Time Accuracy when Locked to WWV Time Signal

- GPS Support:
    - Supports these GPS Receivers: Garmin GPS 16/18 Sensor or the Motorola ONCORE series of GPS receivers
    - ±3 Millisecond Time Accuracy when Receiver is Locked to GPS Time
- Board Size:
    - 6.500 × 5.750 inches or 16.51 × 14.605 centimeters
- Power Supply Requirements:
    - 16 to 28 VDC or 11 to 18 VAC at 100 MA or more
- PSN-ADC-SERIAL Version II Board without WWV Timing Circuitry and Power Supply - Cost: $185.00
- WWV Timing Option - Cost: $25.00
- Garmin GPS 18 Timing System - Cost: $98.00
- Power Supply Options:
    - 110 VAC 60 Hz Wall Mounted Power Supply - Cost: $10.00
    - Universal 110 VAC to 220 VAC 50 / 60 Hz Power Supply with one Adapter Plug - Cost: $25.00
- Six foot 9-pin Male / Female RS-232 Cable - Cost: $5.00
- Enclosures for the PSN-ADC-SERIAL Board.
- Plastic Box with Metal Front/Rear Panels - Cost: $15.00
  Box Size: L: 8.0 W: 7.8 H: 2.5 inches. This box can be used to hold the A/D and Amp/Filter boards. Note: This is just the box, you will need to drill any front/rear panel and mounting holes.

True 16-bit performance

The Analog to Digital converter chip on the PSN-ADC-SERIAL board has an internal noise level of ± 3 counts. With the use of over-sampling, the internal noise level of the chip can be averaged out to produce a true 16-bit (96 dB) dynamic range. The amount of oversampling depends on the sample rate. The number of samples that are averaged together at 200 samples per second is 16 and at sample rate below 200 sps the number of samples averaged together is 32.

Board differences

There are two versions of this ADC board. The boards are identical except for the Analog Input connector and location on the board. The board with the DB25 right angle connector is for customers who only need the ADC board. This version allows access to all of the connectors at the edge of the board. The board with the 26 pin ribbon connector header can be used with the PSN-ADC-EQAMP Amplifier/Filter board with mating 26 pin header connector. This arrangement produces a module that can be mounted in an off-the-self box.

Software

WinSDR (Download) can be used to record data using the PSN-ADC-SERIAL Version I or II ADC boards. This program is included with the purchase of the board. This Windows DLL Driver (pdf version) can be used by programmers who would like to write their own data-logger program or add the A/D board to an existing program. The DLL Driver zip file includes an example, written in VC++ 6.0, showing how to use the PSN A/D Board DLL file.

Earthworm, a powerful seismic software package that can run under various operating systems, can now receive data using the PSN-ADC-SERIAL board or WinSDR. See this link for more information. A device driver for the Windows version of Seislog is now available. Contact me for more information.

To order a board contact Larry at: 1.650.365.7162

16-Bit ADC system board: http://psn.quake.net/serialatod.html#16-Bit%20Performance

WinSDR software: http://psn.quake.net/winsdr/

This page intentionally left blank

Appendix

Electronic Parts Suppliers

We have listed some electronic parts suppliers. We cannot guarantee any particular level of service from these companies and supply this list only as a convenience to our readers.

Allegro Electronic Systems - Semiconductor manufacturer
3 Mine Mountain Rd. Cornwall
Bridge, CT. 06754
Phone: 1-203-672-0123
http://www.allegromicro.com/

AmaSeis Seismic Software - Free seismic software available from: Alan Jones
www.binghamton.edu/faculty/jones/AmaSeis.html

Amidon Associates Inc. - Toroidal cores
240 Briggs Ave
Costa Mesa, CA 92626
Phone: 714-850-4660 1-800-898-1883
http://www.amidoncorp.com/
Email: sales@amidoncorp.com

Analog Devices Inc. - Now selling their parts to individual semiconductors on the Internet
3 Technology Way
Norwood, MA - 02062
http://www.analog.com/

Centerpointe Electronics, Inc. - I-F can type inductors/transformers
5241 Lincoln Ave. Unit #A-7
Cypress, CA 90630
Email: cpcares@sbcglobal.net
http://www.cpcares.com/

Circuit Specialist Inc. - Good parts selection, test equipment, kits
220 South Country Club Drive #2
Mesa, AZ 85210
Phone: (800) 528-1417 +1-480-464-2485
http://www.cir.com/

C & S Sales - Some kits and test equipment
Wheeling, IL 60090
Phone: 1-800-292-7711
http://www.cs-sales.com/
sales@cs-sales.com

Dan's Small Parts and Kits - Low tech site, but nice parts selection
PO Box 3634
Missoula, Montana. 59806-3634
Phone: 406 258 2782
http://www.danssmallpartsandkits.net/

DATAQ Instruments, Inc. - A/D data-loggers
241 Springside Drive
Akron, OH 44333
Phone: 330-668-1444
http://www.dataq.com/
E-mail: info@dataq.com

Debco Electronics Inc. - Sometimes has some good bargains
4025 Edwards Road
Cincinnati, Ohio, USA, 45209
Phone: (513) 531-4499
http://www.debco.com/

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

Digi-Key Corporation - Extensive selection of electronic parts on-line
PO Box 677
Thief River Falls, Minnesota 56701
http://www.digikey.com/

Downeast Microwave Inc - Microwave components and kits for amateurs
954 Rt. 519 Frenchtown
NJ 08825 USA
Phone: 908-996-3584 (Voice)
http://www.downeastmicrowave.com/

Edmund Scientific Company - Science items, parts and kits
101 E. Gloucester Pike
Barrington, New Jersey 08007
Phone: 800-728-6999
http://scientificsonline.com/

Electronic Goldmine - Many unique surplus parts
9322 N. 94th Way
Suite 104
Scottsdale, AZ 85258
http://www.goldmine-elec.com/

Electronix Express - Parts and kits
365 Blair Road
Avenel, New Jersey 07001
Phone: 1-800-972-2225
http://www.elexp.com/

Electus Distribution - Ferrite coil formers
100 Silverwater Road, Silverwater
Sydney, Australia
http://www.electusdistribution.com.au/productResults.asp?form=CAT&CATID=34&SUBCATID=83

Elenco Electronics Inc. - Radio and electronic kits
150 Carpenter Avenue
Wheeling, IL 60090 USA
http://www.elenco.com
email: sales@elenco.com
(847) 541-3800
(800) 533-2441

Fat Quarters Software - FM-1, FM-3 magnetometer sensors
24774 Shoshonee Dr.
Murieta, CA 92562
Phone: 909-698-7950
http://www.fatquarterssoftware.com

FAR Circuits - Circuit boards for many projects.
18N640 Field Court
Dundee, Illinois 60118
Phone: (847) 836-9148 Voice/Fax
http://www.farcircuits.net/

Freescale Semiconductor - Semiconductor manufacturer
7700 West Parmer Lane
Mailstop PL-02
Austin, TX 78729
http://www.freescale.com/

Hamtronics - Radio kits and parts
65 Moul Road
Hilton, NY 14468-9535
http://www.hamtronics.com/

Kits and Parts - Semiconductors and Toroids
http://partsandkits.com/

Linear Technology Corporate Headquarters - Manufacturer of high quality op-amps
1630 McCarthy Blvd.
Milpitas, CA 95035-7417
Phone: 408-432-1900
http://www.linear.com/index.jsp

Loadstone Pacific - Ferrite coil formers, inductors
4769 East Wesley Drive
Anaheim, CA 92807
Phone: 800-694-8089 714-970-0900
http://www.lodestonepacific.com/coilforms

Maplin Electronics - AM Radio chip ZN414/MK484
Maplin Electronics Ltd
National Distribution Center
Valley Road
Wombwell
Barnsley
South Yorkshire
S73 0BS
www.maplin.co.uk

Maxim Integrated Circuits - IC manufacturer - great documentation samples
120 San Gabriel Drive
Sunnyvale, CA 94086
Phone: 408-737-7600
http://www.maxim-ic.com/

Mincircuits - RF parts, MMICs, mixers, etc.
PO Box 350166,
Brooklyn, NY 11235 U.S.A.
Phone: (718) 934-4500
http://www.mini-circuits.com/

Mouser Electronics - Extensive electronics components, no minimum order
1000 North Main Street
Mansfield, TX 76063-1514
http://www.mouser.com/

National Semiconductor - Semiconductor manufacturer
2900 Semiconductor Dr.
PO Box 58090
Santa Clara, California
USA 95052-8090
Phone: (408) 721-5000
http://www.national.com/

Newark Electronics - Extensive inventory of electronic parts
4801 North Ravenswood Avenue
Chicago, Illinois 60640-4496
Phone: 1-800-342-1445
http://www.newark.com/

Ocean State Electronics - AM Radio chip ZN414/MK484
PO Box 1458
6 Industrial Drive
Westerly, RI 02891
Phone: 1-401-596-3080
http://www.oselectronics.com

Onset Computers - HOBO Data Loggers
PO Box 3450
Pocasset, MA 02559
Phone: 800-564-4377
www.onsetcomp.com

Public Seismic Network - 16-bit Analog to Digital converter
Larry Cochrane
http://psn.quake.net/serialatod.html
email: lcochrane@webtronicsc.oom

Radio Jove Project - Radio Jupiter project
c/o Dr. James Thieman
Code 633
NASA's GSFC
Greenbelt, MD 20771
Phone: 1-301-286-9790

Radio Shack - The Internet version of the omnipresent store
Riverfront Campus World Headquarters
300 Radio Shack Circle
Fort Worth, TX 76102-1964
Phone: (817) 415-3700
http://www.radioshack.com/home/index.jsp

Radio-Sky Publishing - Amateur radio astronomy, parts and publications
PMB 242
PO Box 7063
Ocean View, Hawaii 96737
http://www.radiosky.com/

Ramsey Electronics - Electronic kit supplier
590 Fishers Station Dr.
Victor, NY 14564 USA
Phone: 800-446-2295
http://www.ramseyelectronics.com/

RF Parts Company - Small RF parts
435 South Pacific Street
San Marcos, California 92078
http://www.rfparts.com/

RP Electronics
2060 Rosser Ave.
Burnaby B.C.
Canada V5C 5Y
Phone: (604) 738-6722
http://www.rpelectronics.com

Speake & Co Limited - Manufacturer of FM1, FM-3 magnetometer Elvicta Estate
Crickhowell
Powys NP8 1DF
United Kingdom

Texas Instruments Inc. - Semiconductor manufacturer
12500 TI Boulevard
Dallas, TX 75243-4136
Phone: 800-336-5236
http://www.ti.com/

Toko America, Inc.
1250 Feehanville Drive
Mt. Prospect, IL 60056
Phone: (847) 297-0070
1-800-PIK-TOKO
http://www.tokoam.com/

Vectronics - Radio kits
300 Industrial Park Road
Starkville, MS 39759
http://www.vectronics.com/

Vishay Intertechnology, Inc. - Dual matched N-Channel FETs
63 Lincoln Highway
Malvern, PA 19355-2143
http://www.vishay.com/

Interesting Web-sites

Amateur Radio Relay League - Ham radio organization
225 Main Street
Newington, CT, 06111-1494 USA
Phone: 860-594-0200
email: http://www.arrl.org/

ELFRAD Group - Low frequency research group
114 Artist Ln
Statesville, NC 28677
http://www.elfrad.com/
email: cplyler@elfrad.com

Radio Waves Below 22 kHz -Interesting low frequency web-site
http://www.vlf.it/

Index

Headings in *italics* indicate projects. Page numbers in *italics* indicate illustrations.

3-terminal regulators *14*, 16
6N2-meter Band Amateur Radio receiver 127–39
  6N2 meters *137*
  antenna 138–39
  appearance *128*
  band plan *139*
  coil winding information *135*
  diagram (block) *129*
  diagram (schematic) *131*
  frequency coverage (diagram) *130*
  receiver specifications *129*
  semiconductor pin-outs *136*
  single vs. dual conversion *130*
80/40 Meter Direct Conversion Receiver 94–103
  antenna 102, *102*
  appearance *95*
  band change components *100*
  band tuning range *100*
  diagram (schematic) *96*
  overload problems 102
  semiconductor pin-outs *99*
  tuning 102
  W1AW code practice sessions *103*

A

Aircraft Band Receiver (active) 140–50
  antenna 149
  appearance *142*
  audible communications 141–43
  coil winding information *147*
  diagram (schematic) *144*
  semiconductor pin-outs *147*
  tuning 150
Aircraft Band Receiver (passive) 150–52
  antenna 151–52
  diagram (schematic) *151*
aircraft frequencies 150
AM (amplitude) modulation 4–6
  wave-form *6*
AM radio broadcasting 1, 2, 4, 49–51
  carrier and modulation signals *51*
  diagram (block) *51*
AM radio receiver 49–58
AM/Shortwave Crystal Radio 39–40, 46–48
  appearance *46*
  diagram (block) *47*
  shortwave coil assembly *47*
AM Simple Crystal Radio 39–44
  amplitude modulation *40*
AM Simple Crystal Radio (Continued)
  antenna 40, 44, *44*
  appearance *42*
  coil winding 44
  detector 41
  diagram *41*
amateur radio receiver 127–39
  bands *10*
Amateur Radio Relay Ligue (ARRL) 103, 128
Ambient Power Receiver 186–91
  antenna 188, *190*
  appearance *187*
  diagram (schematic) module I *187*
  diagram (schematic) module II *191*
  earthquakes 191
  grounding 188, 190
amplifier 9–10
amplitude modulation *see* AM
Analog to Digital Converters (ADCs) 262–71
  16-bit converter 269–71
  16-bit converter (appearance) *270*
  8 channel converter 264–68
  8 channel converter diagram *265*
  Dataq kit (appearance) *268*
  Dataq kits 268–69
  semiconductor pin-outs *264*
  single channel converter 262–64
  single channel converter diagram *263*
  software 271
antennas 3, 9–10
  APM-II 190–91
  basic long-wire *44*
  capacitive hat *184*
  capacitive hat antenna 177, 179
  dipole *90*, *102*, *114*, 243, *243*
  Discone 126, 138
  dual-dipole *244*
  ground rods 231–32
  horizontal long-wire 188, *190*
  loop *210*
  loop antenna *211*
  power supply 184
  power supply schematic *185*
  quadrifillar 259, *259*, 261
  VHF scanner 138, 149
  VHF turnstile 259, *259*, 261
  whip 126, 138, 149, 151–52
Armstrong, Edwin H. 3
ARRL (Amateur Radio Relay Ligue) 103, 128
atomic clock 105–06

Copyright © 2008 by The McGraw-Hill Companies, Inc. Click here for terms of use.

aurora borealis 213–14
Aurora Monitor 212–23
  appearance *214*
  coil winding information 219–20
  control/display schematic circuit *216*
  control/display unit 214–15
  semiconductor pin-outs *219*
  sensor *220*
  sensor head schematic circuit *215*
  sensor head unit 214

B

Barkhausen, H. 155
Branly, Edouard 1
bugs (FM) 60, 68–69

C

cable *15*
capacitance code 25, *27*
capacitors 13–16, 42
  ceramic *13*, 14–15
  code chart *27*
  disk 63–64
  electrolytic *13*, 15, 64
  identification 12, 25
  MKT polyester *13*, 14
  non-electrolytic 97
  polarization 42
  polyester (greencap) *13*, 14
  preparing leads *28*
  trimmer *13*
  variable *13*, 16
  variable (alternative) 22–24
circuits
  integrated *14*
  printed 34–36
Clapp Oscillator *see* Colpitts oscillator
coils 44 *see also* inductors
  soldering to the pins 32
color code chart *26*
Colpitts oscillator *19*
  circuit *20*
Conrad, Frank 50
converters, analog to digital 262–71
cristal oscillator 20
  circuit *21*
  pictorial diagram *21*
Cristal radio projects 39–48
  AM/Shortwave Crystal Radio *39–40*
  AM Simple Crystal Radio *39*
  FM Crystal Radio *39*
cycles per second *see* Hertz

D

Daily, William 190–91
dawn chorus 156
de-solder 36–37
detector 41
deviation 132
diagrams, reading 18–24
Digital Radio Mondiale (DRM) 8
digital transmission 3
diodes 16
  germanium 44
  LEDs (light-emitting diodes) *14*
  preparing leads *28*
  silicon *14*, 44
  zener *14*
disk capacitors 63–64
Doerle Receiver - 6SL7 *70–71*, 78–80
  6SL7 tube 78
  diagram (schematic) *79*
  history 78
Doerle Receiver - Original 70–78
  appearance *71*
  coil winding *75*, *76–77*
  diagram (schematic) *72*
  history 71
  original parts 73
  short-wave band frequencies *75*
double-sideband (DSB) AM 6

E

Earth Field Magnetometer 192–202
  appearance *193*, *200*
  connection diagram *196*
  diagram (schematic) *195*
  flux-gate sensor *194*
  flux-gate sensor board *197*
  flux-gate sensor diagram (block) *194*
  magnetic fields 193
  power supply schematic *201*
  semiconductor pin-outs *200*
earthquakes, predicting 191
Eckersley, Smith, Tremellen 156
electret microphone *15*
electric cables 18
electromagnetism 1–4
electronic components 12–24, *13*, *14*, *15*
  installation 25–38
  preparing leads 25, *26*, *28*, *29*
ELFRAD group 225, 232
extremely low frequency (ELF) *see* ELFRAD

F

Faraday, Michael 1
ferrite bead (slug) 17
Fessenden, Reginald 50
FET *see* transistors
firefighters VHF 117
flux 29, 30
FM broadcast receiver 59–69
FM Crystal Radio 39, 45–46
  appearance *45*
  circuit *45*

FM (frequency modulation) 6–7, *61*
  radios 3
  wave-form *7*
FM radio 3
Frazier, John 2
frequence modulation *see* FM
frequencies (spectrum) *2*, 4
frequency drift 61

G

galena crystal 41, 44
geophysical alerts 115
GPS (Global Positioning System) 3, 115
greencaps *13*, 14
grounding 188, 190

H

Ham Radio licence 9, 103, 128
Hartley oscillator 19, *19*
headphones 44
health and safety 38
heatsinks *15*, 16
Henry, Joseph 1
Hertz, Heinrich 1
High Frequency (HF) bands 4
history, radio 1–3

I

IC (integrated circuits) 16
  3-terminal regulators 16
IC Shortwave Radio Receiver 81–93
  appearance *82*
  diagram (block) *83*
  diagram (schematic) *84*
  evening broadcast listings *92–93*
  frequency range selection *89*
  popular shortwave bands *91*
  semiconductor pin-outs *88*
  shortwave antennas 90–91
  shortwave listening (SWL) 82–83, 91–93
Induction Loop Receiving System 165–74
  appearance *169*
  coil winding information 171
  diagram (block) *167*
  induction loop technology 166
  receiver 168–74
  receiver schematic *170*
  semiconductor pin-outs *173*
  transmitter 167–68
  transmitter I schematic *167*
  transmitter II schematic *168*
  use 173–74
induction loop technology 166
inductors (coils) *15*, 16–17
integrated circuits (IC) *14*, 16
  preparing leads *29*
ionospheric disturbance receiver 203–11

J

Jupiter 234, *235*
Jupiter Radio Telescope 233–45
  antenna 243–44, *243*, *244*
  appearance *235*
  diagram (block) *236*
  diagram (schematic) *238*
  planet 234, *235*
  radio storms 244–45
  semiconductor pin-outs *241*

L

LDR (light-dependent resistor) 13, *13*
leads, preparing 25
LEDs (light-emitting diodes) *14*, 16–17
  power indicator light 125–26
licenses 9, 103, 128
  US amateur radio bands *10*
light-dependent resistor (LDR) 13, *13*
light-emitting diodes *see* LEDs
lightning 176–77, *177*
Lightning Storm Monitor 175–85
  antenna 177, 179, 184
  diagram (schematic) *178*
  lightning 176–77, *177*
  semiconductor pin-outs *183*
listening-in
  aircraft communication 141–43
  public service VHF 117
loop antenna 210, *211*
loudspeakers *15*, 17

M

magnetic
  fields 193
  induction 166
  storms 213–14
magnetometers 193
Marconi, Guglielmo 2
marine storm warnings 115
marine VHF 117
Maxwell, James Clerk 1
McClelland, R.S. 2
medium wave (MV) *see* AM
microphones *15*, 17
MKT capacitor *13*, 14
MOSFETs *see* transistors

N

narrow-band FM transmission 60–61
National Institute of Standards and Technology (NIST) 105–06
Natural Radio *see* whistlers
NPN *see* transistors

O

ohm (Ω) 12
OMEGA Navigation System 106, 115
oscillation 3
oscillators 9–10, 18–24

P

parts suppliers 272–75
PNP *see* transistors
police VHF 117
potentiometer (pot) 12–13, *13*
power indicator LED 125–26
power transistors *14*, 16
printed circuit (PC) boards 34–36
public service VHF 116–26, *126*

R

radar frequency band 1
radio frequency (RF) spectrum 1, 3–4
  chart *5*
radio storms (Jupiter) 244–45
radio telescope 233–45
radio waves 4, *6*, *7*
railroad VHF 117
receivers 9–10
  diagrams *11*
relays *15*, 17
resistors 12–13, *13*, 42
  codes 25
  color code chart *26*
  light-dependent 13, *13*
  preparing leads *26*
  tolerance 12
  variable 12
Rosenberg, Leo 2
rosin flux 30

S

satellite navigation *see* GPS
schematics, electronic 18–24
seismic network, public (PSN) 232
semiconductors 16 *see also* capacitors; diodes; resistors; transistors
  substitution 18
SHF (Super High Frequency) 4
shortwave
  bands 1, 91
  broadcast chart *9*, 92–93
  fadeout 204
shortwave listening (SWL) 8–9, 82–83, 91–93
shortwave radio 8
  receiver 81–93
silicon diodes *14*
single-sideband *see* SSB
sniffer coil 258
solder 30, 33
soldering 25, 29–38
  check list soldering 34
  check list un-soldering 37
  flux 29
  guns 30, 31–32
  irons 30–31, 37
  joint quality *35*, *36*
  mechanical joints 33–34, *33*, *34*
  overheating 32
  preparation (iron) 32
  preparation (material) 32–33
  printed circuit (PC) boards 34–36
  remove solder 36–37
  solder 30, 33
  station 31, *31*
  temperature 33
  trimming leads *36*
  ventilation 38
  to wire 32–33
  wire gauges 33
Solid-state FM Broadcast Receiver 59–69
  diagram (block) *61*
  diagram (schematic) *62*
  inductor coil winding information *66*
  semiconductor pin-outs *66*
speakers *15*, 17
SSB technology 2–3
  modulation 7–8
  wave-form *8*
Storey, L.R.O. 156
storms, electrical 176–77
sub-carriers 69
Sudden Ionospheric Disturbance Receiver 203–11
  antenna 210, *211*
  appearance *205*
  diagram (schematic) *205*
  power supply schematic *210*
  semiconductor pin-outs *209*
  short wave fadeout 204
  use 211
  VLF signal propagation 204–05
Super High Frequency (SHF) 4
switches 17–18
  electronic 16
  slider *15*

T

Tate, Joseph 186, 190–91
temperature (soldering) 33, 35
Thomas, William 2
thunderstorms 176–77
time code receiver 104–15
tinning 33, 34, 41
tools and supply 41
  hexagonal plastic tool 67–68
transformers *15*
transistorized radio 3
transistors 16
  FET *14*
  MOSFET *14*

transistors *(Continued)*
  NPN *14*
  PNP *14*
  power transistors *14*
  preparing leads *28*
TRF AM Radio Receiver 49–58
  appearance *52*
  diagram (block) *52*
  diagram (schematic) *53*
  semiconductor pin-outs *56*
trimpot *see* potentiometer (pot)
tube #19 71–73

U

UHF (Ultra High Frequency) 4
ultra-low frequency spectrum 225
Ultra-Low Frequency (ULF) Receiver 224–32
  antenna 231–32
  antenna circuit *231*
  appearance *225*
  diagram I (schematic) *226*
  diagram II (schematic) *227*
  ELFRAD group 225, 232
  power supply schematic *227*
  semiconductor pin-outs *231*
  ULF-spectrum 225
un-solder 36–37
  check list 37

V

Variable Frequency Oscillator (VFO) *see* VFO
VCO (Voltage Controlled Oscillator) 22–24, *23*
  circuit *23*
  pictorial diagram *24*
ventilation 38
Very High Frequency *see* VHF
Very Low Frequency *see* VLF
VFO (Variable Frequency Oscillator) 22
  circuit *22*
  pictorial diagram *23*
VHF Public Service Receiver 116–26
  antenna 126
  appearance *117*
  coil winding information *123*
  diagram (block) *118*
  diagram (schematic) *119*
  power LED 125
  semiconductor pin-outs *123*
  VHF public service frequency bands *126*
VHF (Very High Frequency) 1, 4
VLF (Very Low Frequency) 4
VLF Very Low Frequency Radio Receiver 153–64
  appearance *157*
  dawn chorus 156
  diagram (schematic) *158*
  recording 164
  semiconductor pin-outs *161*
  times of listening 164
  using the receiver 162–63
  VLF radio stations *163*
  whistlers 154–56
voltage controlled oscillator *see* VCO
volume control *see* potentiometer (pot)

W

W1AW on air code practice sessions *103*
Weather Satellite Receiver 246–61
  antenna 259, *259*, 261
  appearance *250*
  coil winding information 254, *256*, 258
  diagram (block) *250*
  diagram (schematic) *252*
  frequencies *261*
  operation 259–60
  satellite view (USA) *248*
  semiconductor pin-outs *256*
  set-up 260–61
  set-up diagram (block) *260*
weather VHF 117
whistler receiver 156–64
whistlers 154–56
wide-band FM transmission 60–61
wire *15*, 18
  soldering to 32–33
WWV broadcast schedule *115*
WWV Time Code Receiver 104–16
  antenna 114, *114*
  appearance (inside) *106*
  appearance (outside) *105*
  atomic clock 105–06
  coil winding information *112*
  diagram (block) *107*
  diagram (schematic) *108*
  LED power indicator 113
  NIST (Nat.Inst.Standards and Technology) 106
  semiconductor pin-outs *112*
  WWV broadcast schedule *115*

Z

zener diodes *14*, 64

Everything a *just* *slightly wicked* electronics hobbyist could want!

101
SPY GADGETS
FOR THE EVIL GENIUS
TAB
BRAD GRAHAM AND KATHY McGOWAN

101 Spy Gadgets for the Evil Genius
Brad Graham & Kathy McGowan
0-07-146894-3

50
MODEL ROCKET
PROJECTS
FOR THE EVIL GENIUS
TAB
GAVIN D. J. HARPER

50 Model Rocket Projects for the Evil Genius
Gavin Harper
0-07-146984-2

ELECTRONIC
SENSORS
FOR THE EVIL GENIUS
54 ELECTRIFYING PROJECTS
TAB
TOM PETRUZZELLIS

Electronic Sensors for the Evil Genius
Thomas Petruzzellis
0-07-147036-0

MORE
ELECTRONIC
GADGETS
FOR THE EVIL GENIUS
40 NEW BUILD-IT-YOURSELF PROJECTS
TAB
ROBERT IANNINI

MORE Electronic Gadgets for the Evil Genius
Robert Iannini
0-07-145905-7

50 AWESOME
AUTO
PROJECTS
FOR THE EVIL GENIUS
TAB
GAVIN HARPER

50 Awesome Auto Projects for the Evil Genius
Gavin Harper
0-07-145823-9

BIONICS
FOR THE EVIL GENIUS
25 BUILD-IT-YOURSELF PROJECTS
TAB
NEWTON C. BRAGA

Bionics for the Evil Genius
Newton Braga
0-07-145925-1

MECHATRONICS
FOR THE EVIL GENIUS
25 BUILD-IT-YOURSELF PROJECTS
TAB
NEWTON C. BRAGA

Mechatronics for the Evil Genius
Newton Braga
0-07-145759-3

123 PIC
MICROCONTROLLER
EXPERIMENTS
FOR THE EVIL GENIUS
TAB
MYKE PREDKO

123 PIC Microcontroller Experiments for the Evil Genius
Myke Predko
0-07-145142-0

Other Books Available:

Electronics Circuits for the Evil Genius
Dave Cutcher
0-07-144881-0

Electronic Gadgets for the Evil Genius
Robert Iannini
0-07-142609-4

123 Robotics Experiments for the Evil Genius
Myke Predko
0-07-141358-8

TAB ROBOTICS

Learn more. McGraw Hill Do more.

A Simple Radio Receiver

< Return to Part IA Linear Circuits & Devices page ...

The diagrams, images and screen shots below detail the development of a simple modular radio receiver based entirely on circuits and devices studied during the Part IA course on Linear Circuits and Devices.

This receiver may not take the market by storm! However, we hope it will help illustrate how the circuits and principles we are studying all contribute to the art of electronic circuit design.

Contents

- **A little history**
- **Modern crystal sets**
- **Development of Crystal Receiver**
- Resonant LC circuit
- Diode Detector
- **Development of Improved Receiver**
- Audio Amplifier
- Radio Frequency (RF) Amplifier
- Battery Lifetime
- Output Amplifier
- **The Complete Receiver**
- Photo Gallery
- Circuit Diagrams
- **Notes and Acknowledgments**

A little history

The first radio receivers were crystal sets, and became available in the 1920's with the opening of Marconi's first broadcast station in Chelmsford.

A crystal set does not have a battery. It runs completely from the energy extracted from radio waves it picks up from the antenna. A resonant LC (or tuned) circuit coupled to a large aerial or antenna was used. Many amateur experimenters constructed crystal sets, often with the tuner inductor coil wound on a tubular box or a drinking glass. At this time the semiconductor diode had not been invented, so extracting the audible modulation signal from the transmission relied on the non-linear electrical properties of the 'crystal', typically a piece of coke or galena. In early sets a "cat�s whiskers" - a fine piece of wire - was adjusted by trial and error to make a suitable contact with the crystal.

There were many limitations to the crystal set: it needed a big aerial (antenna), an earth connection, the clumsy cat�s whisker, and the weak signal could only be listened to by one person at a time with headphones. Very quickly the crystal set began to be replaced by valve radios with loudspeakers, powered by batteries.

In World War II, crystal sets were used by prisoners of war in prison camps to listen to news from home. Much ingenuity went into improvising the necessary components.

See **this presentation** for further details of historic crystal set receivers.

Back to top

Modern crystal sets

Nowadays people only use crystal sets as a hobby. There are various different kits available on the market.

Dave's Crystal Radio Set Page

http://www.crystalradio.net/

Back to top

Development of Crystal Receiver

The design shown here is a little more complex than strictly necessary, but some of the adaptations incorporated make it easier to develop the design, adding amplifiers and other stages as we meet them in the course.

Resonant LC circuit

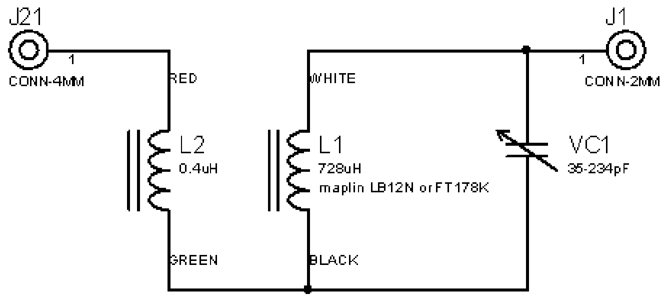

Circuit (a) alongside shows the LC resonant circuit comprising L1 and VC1 used to select or tune the required frequency and station. The inductance used is 729 uH, and VC1 can be varied over a range 30 to 234 pF. That means the range of resonant frequencies is about 400 to 1075 kHz, which includes a good part of the Medium Wave band of frequencies.

L1 is a coil of wire wound on a ferrite rod. Ferrite is a material with a high magnetic permeability - that simply means that that to achieve a given inductance, a smaller number of turns is required than if the coil had an air core. That keeps its winding resistance low, and enhances the Q, which in turn means the resonant circuit has a narrower resonance peak, and is better at rejecting unwanted frequencies.

VC1 is a variable capacitor, in which the capacitance is varied by rotating a knob, which controls the extent to which a set of small parallel plates are enmeshed. The capacitance can never be reduced to zero, of course, owing to the residual capacitances due to its structure. To these figures must be added any stray or parasitic capacitances arising from the connecting wires, circuit board, and other elements. These can never be eliminated, but careful construction aims to keep them down to a few tens of pF.

The signal from the antenna (perhaps a few tens of microvolts, or hundreds for a nearby transmitting station) is introduced to the LC circuit either through a small capacitance, or, as in this case, by means of a second coil L2 wound on top of L1, with its other end connected to earth. This behaves like a *transformer* - currents flowing in L2 generate a changing magnetic flux which cuts L1 and induces an emf in it.

Back to top

Diode Detector

Transmissions on Medium Wave frequencies are by *Amplitude Modulation*. The amplitude of the transmitted radio signal or *carrier* is modulated - made to rise and fall - by another signal of audible frequency - for example, speech or music. The receiver must extract or *detect* this audio signal by separating it from the carrier. One way of doinjg this is by means of a diode detector.

The circuit below represents the detector. The key component is D1, type OA47, a pn junction diode.

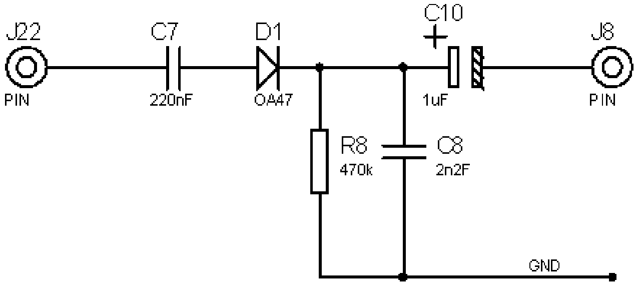

The signal is introduced via capacitor C7 (note that this component is not strictly necessary for the basic crystal receiver, but is included for ease of connection to later additions to the design). It blocks the flow of any DC current into the detector. Its value is chosen so that it has a low reactance at the (medium wave) frequencies of interest, which therefore pass without hindrance.

The pn junction diode D1 allows the flow of conventional current from left to right only. Thus, the waveform observed at the right of D1 would comprise positive-going parts of the incoming sinusoidal signal; negative going half-cycles are effectively removed. In this way, information can be extracted from the transmitted signal, whose amplitude is *modulated* by a modulating signal (typically an audio frequency derived from speech, music etc.

See **this presentation** **(UP in 2005)** for further information on how the detector works.

The combination of R8 and C8 removes any remnants of the carrier wave, leaving only the lower frequency, modulation signal. This can be applied to a sensitive earphone. More realistically, an amplifier may be used to boost the signal to a level where it is capable of driving a loudspeaker.

Back to top

Development of Improved Receiver

Audio Amplifier

The audio signal from the detector is at best of the order of a few mV. This is just enough to be audible if applied to a sensitive earphone.

For comfortable listening with headphones an audio signal of about 1 volt amplitude is desirable, so that an amplifier with a gain of about 100 times is needed.

This specification is comparable with the Design Example in Lecture 14. The circuit below has been prepared for entry to the pSpice simulator so that its performance can be predicted. It is almost identical to the one discussed in the lecture. The only difference is that pSpice does not readily accommodate volume controls (implemented in the case study using a variable potential divider, or potentiometer); instead, a fixed potential divider (using resistors RV1 and RV2) is shown; the resistors are chosen so that 90% of the input signal VIN is applied to the amplifier. The values chosen for the coupling and bypass capacitors are based on a determination of their reactance at the lowest frequency of interest, typically 60 Hz.

Function	**Value**	**Reactance at 60 Hz**	**Notes**
Coupling capacitor	1μF	3000 Ω	Much smaller than RG1 and RG2
Bypass capacitor	100 μF	30 Ω	Much smaller than RS1 and RS2

Note that even with such an amplifier, the performance will only be satisfactory with strong signals from a nearby transmitter. This point is considered in the section below.

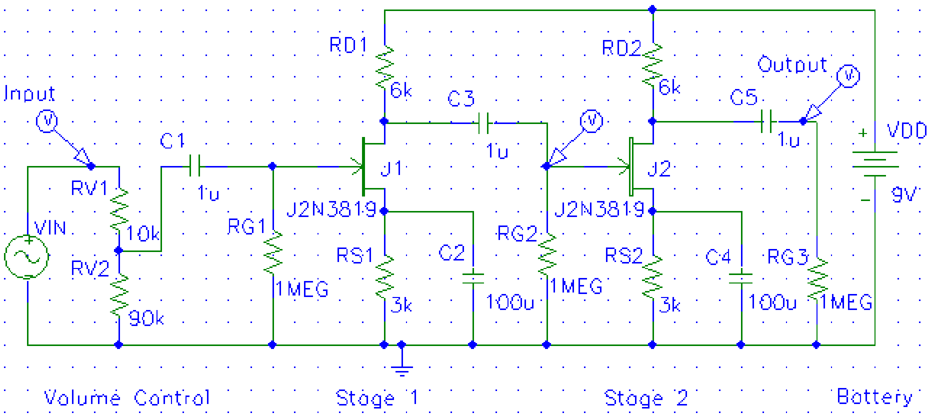

The (V) markers in the circuit above indicate the three points in the circuit at which the output is plotted:

- Input signal
- Output from Stage 1
- Output from Stage 2

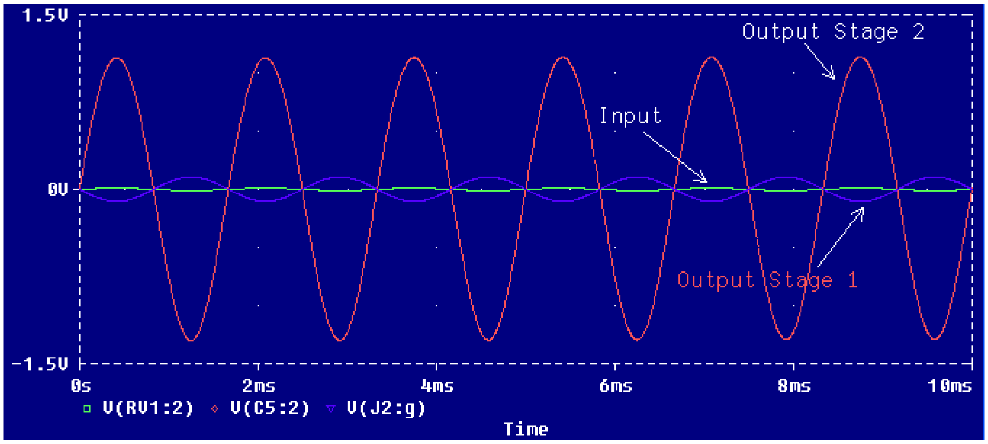

The output plots the signals at the designated mark points as a function of time:

- Input signal (V(RV1:2), green trace) is a sine wave at 600 Hz with amplitude 10 mV
- Output from Stage 1 (V(J2:g), blue trace) has amplitude 100 mV approximately, and is inverted with respect to the input
- Output from Stage 2 (V(C5:2), red trace) has amplitude about 1.2 V. It has the same polarity as the input.

The overall gain is thus predicted to be about x120, about -11 for each of the stages. This is enough to allow comfortable listening with earphones for reasonably strong signals.

pSpice can also be programmed to tabulate or plot circuit currents - for example, the current flowing between Drain and Source in either JFET, or the total current being supplied by the battery. These figures are useful in helping us get some idea of battery lifetime. A typical plot of battery current is shown below, and it can be seen that the current consumption averages 1.5 mA.

In the practical implementation of the receiver, the JFETs used in the audio amplifier turned out to have rather low mutual conductance, g_m, of about 2.5 mS. The manufacturer's data-sheet indicates that for a random sample, g_m may lie between 2 and 6.5 mS - a very large spread! The simulator uses an average value of about 4 mS. Since the voltage gain is directly proportional to g_m, the measured voltage gain was less than predicted, and was about -5 for each stage.

[Back to top](#)

Radio Frequency (RF) Amplifier

For the design to work well with weak signals from more remote transmitting stations, or with a smaller antenna, additional amplification is needed.

Since the detector works more efficiently with a larger signal applied, it makes sense to amplify the signal before it is applied to the detector. At this point, the signals being amplified are at the carrier, or Radio Frequency, so this kind of amplifier is called an RF amplifier. The resultant block diagram is shown below.

The 2N3819 JFET will amplify over a wide range of signal frequencies, up to some hundreds of MHz, whereas in our design study the carrier frequency is approximately 1MHz (550 kHz to 1.6 MHz in fact for the Medium Wave band). As a result, a circuit very similar to that used for the audio amplifier may be used for RF amplification. The main difference is that smaller coupling and bypass capacitors can be used since the frequency of operation is about 100 times as high as for the audio case; and capacitive reactance is proportional to 1/f. The circuit used is shown below, together with the predicted response (determined by pSpice) to a 600 kHz signal of 100 microvolts amplitude.

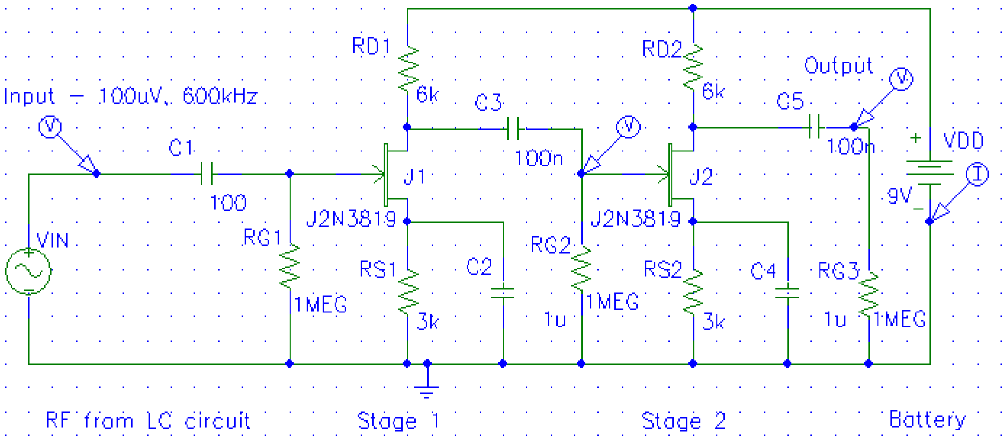

As before, the (V) and (I) markers in the circuit above indicate points in the circuit at which the output is plotted:

- RF input signal from resonant LC circuit
- Output from Stage 1
- Output from Stage 2
- Current flowing into the negative terminal of the battery (i.e. net current consumption)

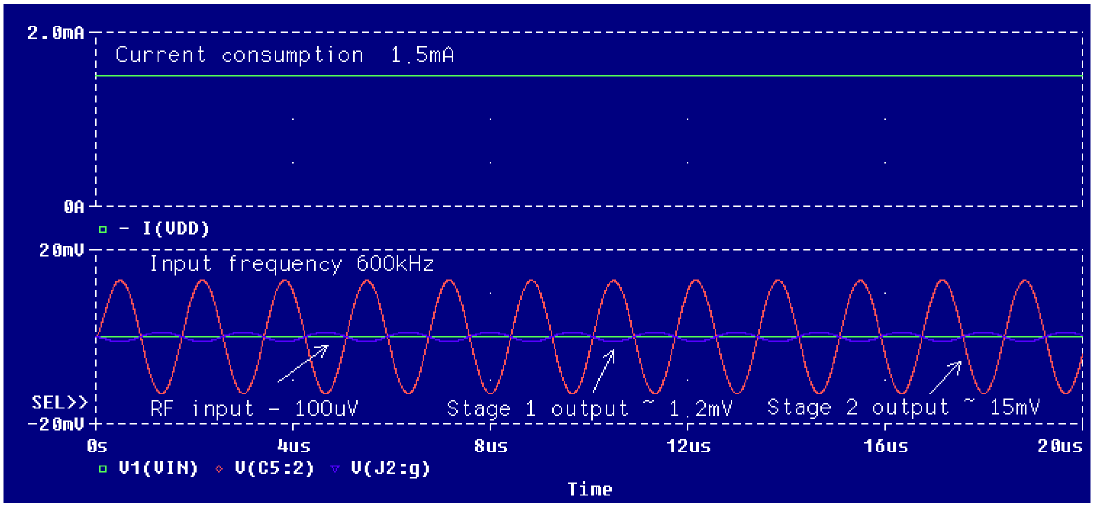

With the same voltage scales in use, the input (amplitude 100 μV, shown in green) appears almost flat in comparison with the two output signals. The gain achieved by each stage is approximately 12.5, so the overall gain of the two-stage amplifier is about 150.

If we were to tabulate the performance over the full range of radio frequencies, we would expect to find some variation. That is one disadvantage of this simple approach to receiver design. So far as possible, modern design styles avoid dependence on frequency by transforming the incoming carrier frequencies to a more convenient, fixed frequency (called the *intermediate frequency*), in a process called *mixing* or *heterodyning*, to allow the amplifiers to be better optimised for good performance.

As with the audio amplifier, the JFETs used in the receiver had lower g_m than assumed by pSpice, so the measured gain was rather lower than predicted.

[Back to top](#)

Battery Lifetime

The capacity of a typical 'budget' PP3 9-volt battery is about 400 mAh (milliampere-hours), which means it can (theoretically) supply a current of 1mA for 400 hours.

In the 'experiment' shown in Lecture 14-15, a 6-volt 0.5A cycle lamp connected to a fresh PP3 burnt for about 40 minutes. This suggests that the battery could provide approximately 40/60 x 0.5 Ah, or about 330 mAh.

In practice, batteries can provide more mAh when discharged slowly than if overloaded, as in this case.

Battery current is required by the RF- and audio amplifiers in the receiver design described.

The consumption figures are (approximately):

- RF Amplifier - 0.75 mA (first stage) and 0.75 mA (second stage) - see the plot above
- Audio Amplifier - 0.75 mA (first stage) and 0.75 mA (second stage)

On this basis, if used in our receiver circuit (based on two RF and two audio stages), the total consumption is about 3mA, and the battery lifetime would be approximately 130 hours. Not bad! But still the design is suitable for headphone reception only.

[Back to top](#)

Output Amplifier

A loudspeaker capable of being heard from a range of a few feet needs to be energised with about 1 watt of electrical power - typically about 4 volts and a current of about 500 mA (both peak figures). Although the JFET stage could possibly give sufficient voltage gain to satisfy this need, it could never provide enough current to drive a 'speaker directly.

For this, we need an additional amplifier with limited voltage gain, but high **current gain**. JFETs are not the best choice for this purpose, and you will meet circuits much better suited to these applications in Part IB. The design demonstrated in the lectures is fitted with an output amplifier based on a proprietary chip, the LM386. This is in many ways like an operational amplifier (to be covered in lectures 17-18), but specially designed to be able to deliver substantial current. This completes the development of our modular radio - the final block diagram is shown below.

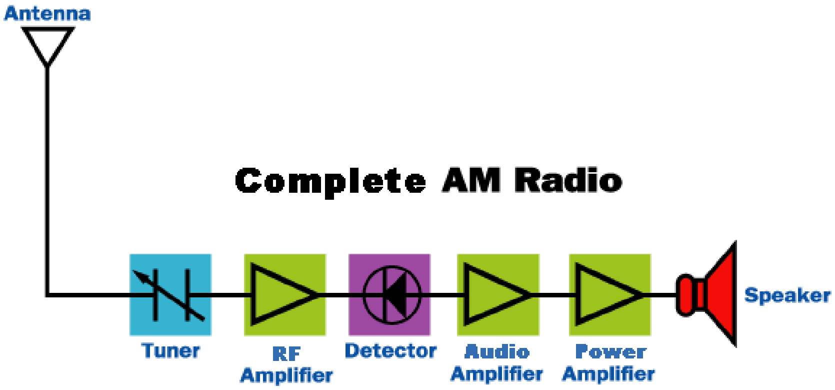

With such an output amplifier fitted, the current consumption increases substantially. In fact, the current draw depends on the setting of the volume control, since this controls the audio output, i.e., how much electrical energy is directed to the loudspeaker. With the volume set high, a battery lifetime of only an hour or two must be expected. This is consistent with observations of commercial radio designs.

[Back to top](#)

The Complete Receiver

Photo Gallery

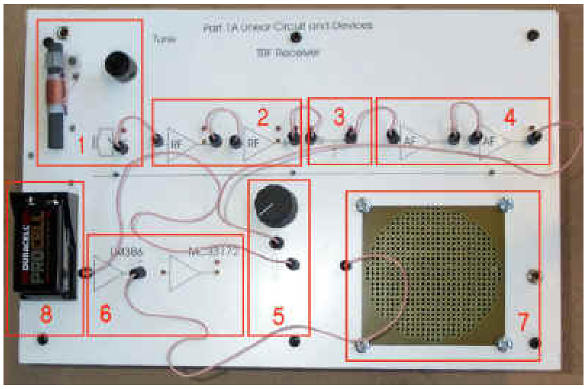
Tune
Part 1A Linear Circuit and Devices
TRF Receiver
1
2
3
4
RF
RF
AF
AF
LM386
MC33172
8
6
5
7

Click [here](#) or on the thumbnail image to see a larger version of the picture with a key identifying the functional blocks

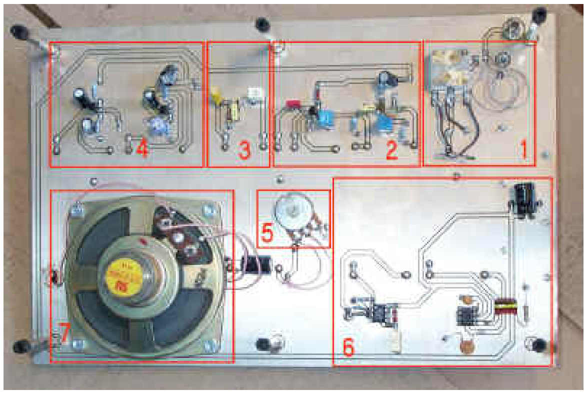

Rear view of receiver - click on thumbnail to see the original full-size picture and a key to the functional sections (coming soon)

Two stage radio-frequency amplifier

Diode detector

Two stage audio frequency amplifier

Output stages (op-amp and LM386)

Loudspeaker

Volume control

[Back to top](#)

Circuit Diagrams

Audio amplifier (two stages)

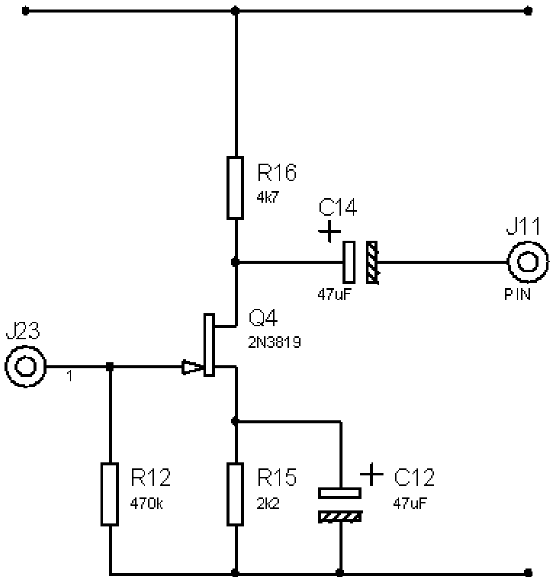
R16
4k7
C14
47uF
J11
PIN
Q4
2N3819
J23
1
R12
470k
R15
2k2
C12
47uF

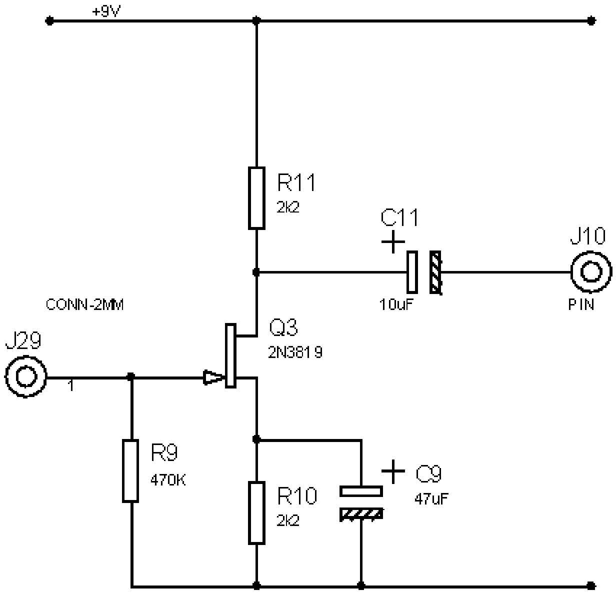
+9V
R11
2k2
C11
10uF
J10
PIN
CONN-2MM
Q3
2N3819
J29
1
R9
470K
R10
2k2
C9
47uF

Back to top

Volume control (potential divider)

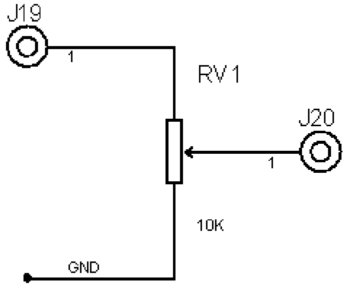
J19
1
RV1
J20
1
10K
GND

A Radio Frequency Amplifier (two stages)

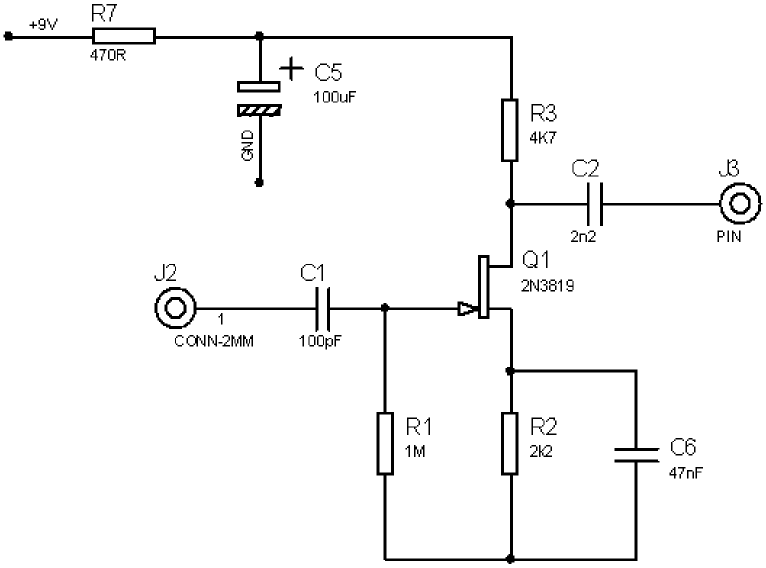
+9V
R7
470R
C5
100uF
GND
R3
4K7
C2
2n2
J3
PIN
Q1
2N3819
J2
1
CONN-2MM
C1
100pF
R1
1M
R2
2k2
C6
47nF

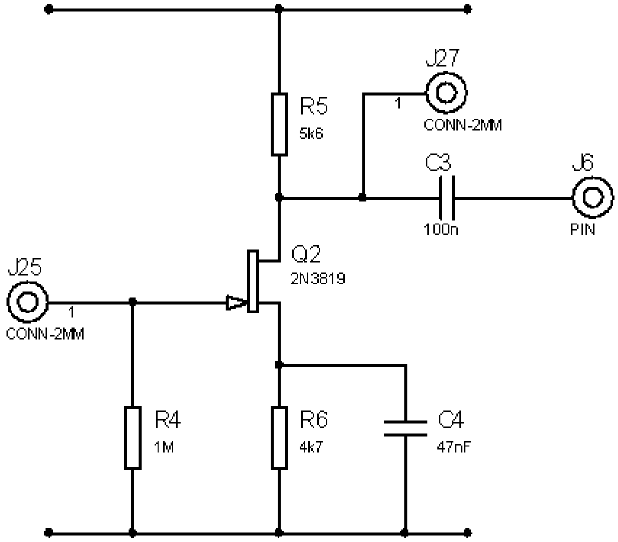
J27
1
CONN-2MM
R5
5k6
C3
100n
J6
PIN
Q2
2N3819
J25
1
CONN-2MM
R4
1M
R6
4k7
C4
47nF

The two RF stages are essentially identical and are very similar in structure to the audio amplifier. The network comprising R7 and C5 is called a *decoupling* circuit. It prevents signals being coupled to other stages by way of the power supply leads.

With the higher operating frequency, smaller coupling and bypass capacitors (C1, C2, C3, C4, C6) may be used than for the audio amplifier without adverse effect on performance.

Back to top

Output Audio Amplifier (based on operational amplifier)

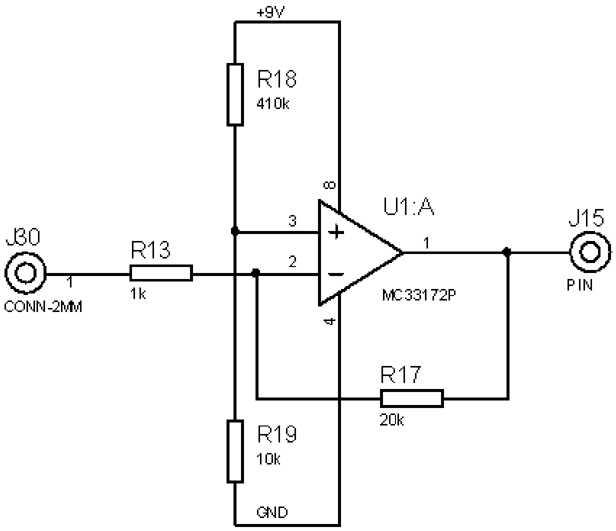
+9V
R18
410k
8
3
2
U1:A
+
−
1
J15
PIN
J30
1
CONN-2MM
R13
1k
MC33172P
4
R17
20k
R19
10k
GND

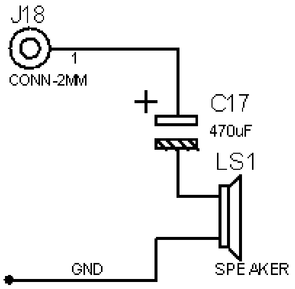
J18
1
CONN-2MM
+
C17
470uF
LS1
GND
SPEAKER

MC33172 is a standard op-amp. The circuit is set to give a voltage gain of -20, which is more than enough. The output produced by this circuit is quite limited because the device can only provide a current of about 15 mA maximum at its output terminal, simply not enough to drive a loudspeaker effectively.

The loudspeaker.
The large capacitor C17 is to prevent a substantial DC current flowing through the speaker, which could cause long-term damage.

Back to top

Audio power amplifier (based on LM386 chip) capable of driving a loudspeaker

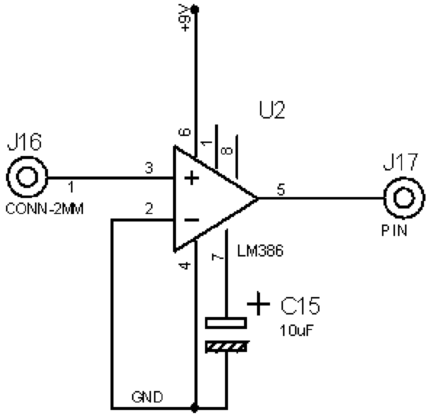
+9V
U2
6
1
8
J16
3
+
1
CONN-2MM
2
−
5
J17
PIN
LM386
4
7
+
C15
10uF
GND

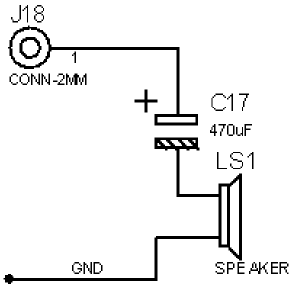
J18
1
CONN-2MM
+
C17
470uF
LS1
GND
SPEAKER

LM386 circuit

Loudspeaker

Back to top

Notes and Acknowledgments

Virtually all the circuitry discussed above is abstracted from the lecture notes for the course. This includes:

- LC resonant circuit (Lecture 12)
- RF amplifier (two stages, based on JFET - Lecture 14)
- Audio amplifier (two stages, based on JFET - Lecture 14)
- Volume control (potentiometer - Lecture 14)
- Audio amplifier (based on an op-amp - Lecture 17).

The only items not strictly part of the IA curriculum are the use of the p-n junction diode as a detector, and the power amplifier (used to drive a loudspeaker) based on the LM386 IC. You will meet the latter in Part IB.

With thanks to CUED Electronic Development Group for assistance in fabricating the design.

Back to top

David Holburn Last updated: 18:35 on 18th July 2008

all about
CRYSTAL SETS

by
Charles Green

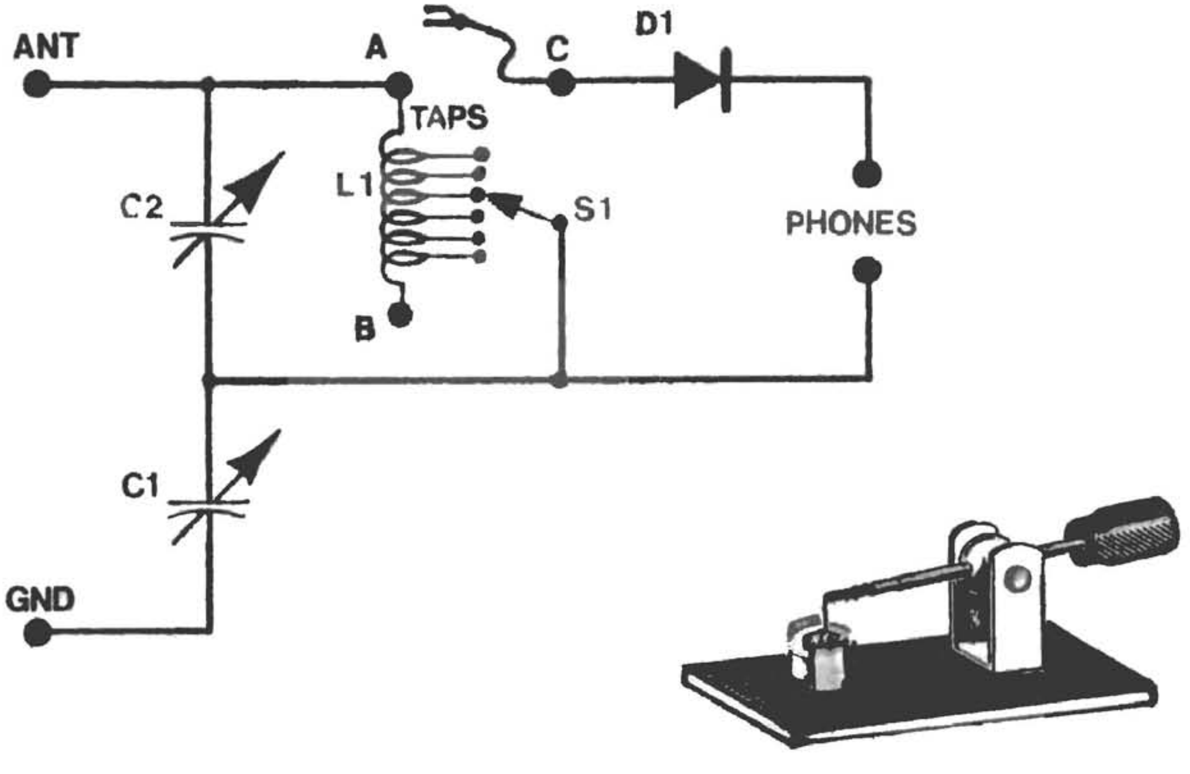
ANT
A
C
D1
TAPS
L1
C2
S1
PHONES
B
C1
GND

How-To-Build Simple Crystal Set Radios

621.3841
G795A
c.1

allabout books no. 101

all about
CRYSTAL SETS

by

Charles Green

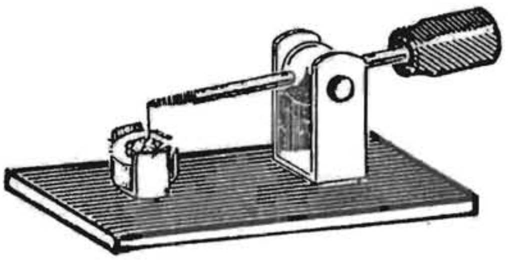

How - To - Build Simple Crystal Set Radios

FIRST EDITION

All rights reserved, including the
right of reproduction in whole or
in part in any form.

Copyright © 1984 by Charles Green

Published by Allabout Books
P.O. Box 22366
San Diego, CA 92192-2366

ISBN 0-930003-00-4

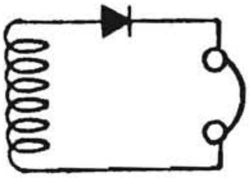

PREFACE

In the beginning of this century, when radio was called "Wireless", most of the early experimenters used a crystal set to receive radio signals.

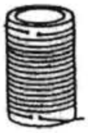

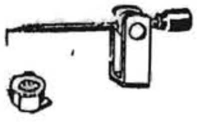

The crystal set is still the simplest type of radio receiver. A catwhisker (made from a thin wire) touching a sensitive spot on a piece of galena or silicon mineral crystal and a tuning coil, form the basic parts of this radio.

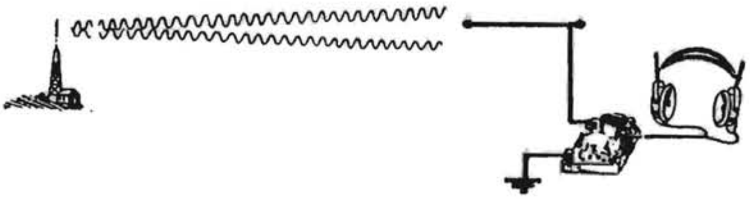

Radio signals from a nearby station are picked up via antenna and ground connections to the crystal set. The galena or silicon crystal detects the signals and the resultant audio is heard with headphones. No battery or electrical power is required.

This book tells you about crystal sets and shows you how to build them.

Charles Green

CONTENTS

1

THE CRYSTAL DETECTOR....................................4
- How-it-Works, 4
- Detection of Radio Waves, 6
- Tuning the Crystal Set, 7
- Capacitor, 8
- RF Coil, 8
- Tuned Circuit, 9
- Antenna and Ground, 10
- Headphones, 11

2

SIMPLE RECEIVER PROJECTS................................12
- Simple Crystal Set, 12
- Slider Tuned Crystal Set, 17
- Ferrite Coil Crystal Set, 23
- Spiderweb Coil Crystal Set, 27

3

ADVANCED RECEIVER PROJECTS..............................32
- Dual-Tuned Crystal Set, 32
- Selective Crystal Set, 37
- Carborundum Crystal Set, 43

4

COMPONENT PROJECTS......................................48
- Book Tuning Capacitor, 48
- Rotary Switch, 53

5

WORKSHOP..54
- Soldering, 54
- Project Construction Tips, 55
- Mounting Minerals, 56

INDEX...57

1

THE CRYSTAL DETECTOR

A crystal detector of radio waves can be either a traditional catwhisker and natural mineral crystal, or a modern semiconductor diode.

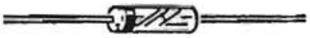

SEMICONDUCTOR DIODE

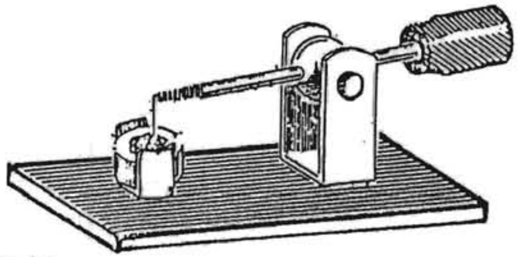

CATWHISKER DETECTOR

Certain natural minerals such as galena (lead sulphide) and silicon have electrically sensitive surface spots. In the beginning of this century, experimenters discovered that electrical current would flow into the sensitive spot on a mineral crystal from a catwhisker, but would not pass readily (high resistance) from the crystal to the catwhisker. This unique feature was used to detect (rectify) radio waves and extract the audio modulation.

HOW-IT-WORKS

There are two general types of crystal detctors;

- . Point contact diode
- . Junction diode

Both types use semiconductor material such as silicon, germanium, or galena. The basic semiconductor junction diode is made of two portions of the material "doped" in manufacturing by adding controlled amounts of chemicals. The doped portions have opposite electrical properties:

"P" TYPE, with an excess of positive electrical charges (holes).

"N" TYPE, with an excess of negative electrical charges (electrons).

the crystal detector

Junction Diode

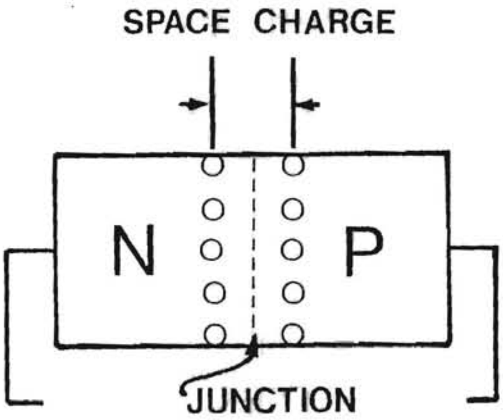

The boundary between the two doped portions is called a PN junction. The area on both sides of the PN junction is called the space charge region. When a positive voltage is applied to the "P" material (and negative voltage to the "N" material), the space charge region becomes narrow and electron current flows through the PN junction.

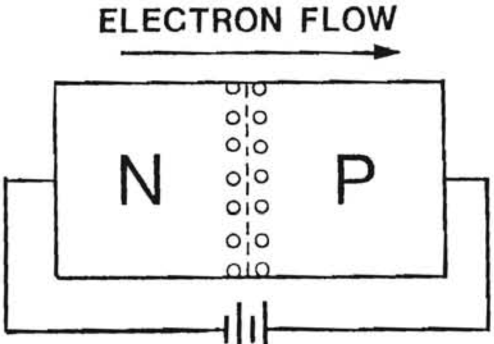

FORWARD BIAS

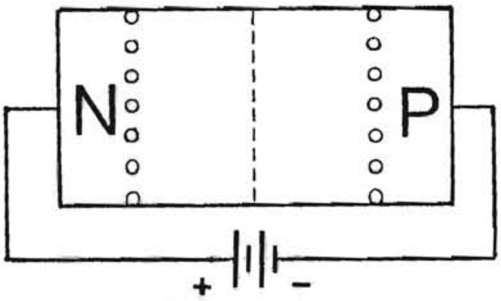

REVERSE BIAS

When the voltage polarities are reversed, negative to "P" and positive to "N", the space charge region widens (increasing the effective resistance of the PN junction) and the electron flow is very small (back current). If an ac voltage is applied to the PN junction, it will act as a rectifier (or radio wave detector).

Point Contact Diode

The point contact diode uses a "catwhisker" of thin wire contacting the surface of doped "N" semiconductor material. The pointed end of the wire produces a high intensity electric field in the area of the semiconductor surrounding the wire. When an electric current flows from the wire to the semiconductor material, a small region of "P" material forms around the area of the point contact. Thus a PN junction is formed which operates in the same way as the junction diode.

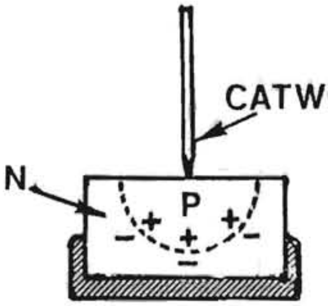

The sensitive spots of galena and silicon crystals used in traditional catwhisker detectors have naturally doped areas of semiconductor material, and operate in the same way as the point contact detector.

DETECTION OF RADIO WAVES

At the transmitting radio station, a microphone converts sound waves into electrical audio signals. The audio signals then modulate the transmitted radio waves.

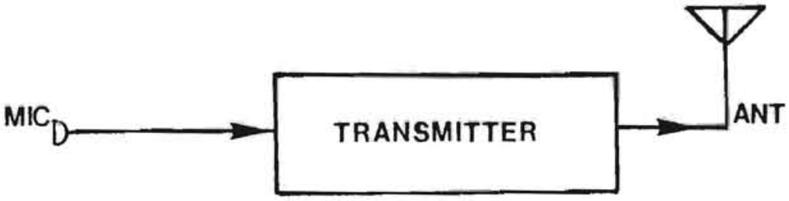

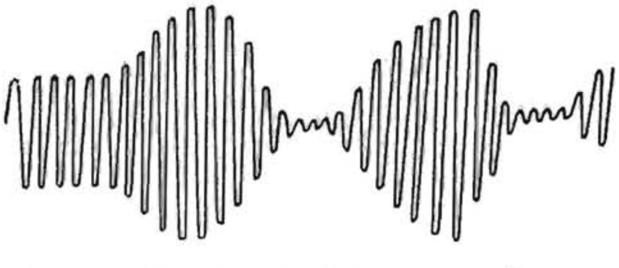

MODULATED RADIO WAVES ⟶

the crystal detector

The transmitted radio waves are intercepted by the receiving antenna and small rf currents are induced in the antenna wire. The rf currents (high frequency ac) flow through the detector to ground. As previously explained, the crystal detector will pass current in only one direction, therefore, pulsating dc voltages (detected audio signals) are developed by the crystal equivalent to the audio modulation of the transmitted radio waves.

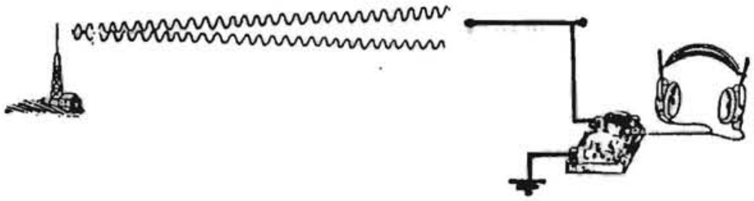

The detected audio signals flow through the headphone coils and affect the intensity of a magnetic field. This causes the steel discs of the headphones to vibrate and produce audible sound waves.

TUNING THE CRYSTAL SET

The antenna intercepts all radio waves in the area, and the signals will be heard jumbled up in the earphones. A nearby radio station may be received with the other stations heard in the background.

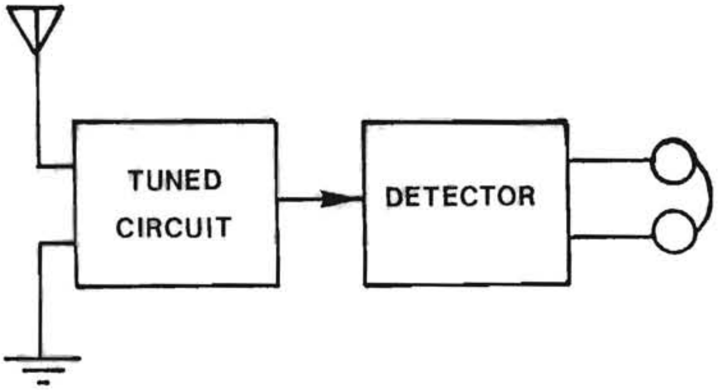

In order to receive a radio station properly, the crystal set must be tuned to the frequency of the transmitter. A tuned circuit composed of an rf coil (inductance) and a capacitor, is connected between the antenna/ground circuit and the detector.

Capacitor

The modern capacitor is evolved from the old Leyden jar. It consists of two or more metal sections separated by a dielectric (air, paper, or plastic). The capacitor stores energy in the form of an electrostatic charge in its dielectric. A variable capacitor (used in a tuned circuit) has movable metal sections that adjust the amount of capacity.

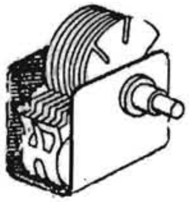

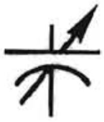

A capacitor will allow pulsating dc current or ac current to flow through it, but will block steady dc current. A capacitor has an oppositon or "reactance" to ac that varies with frequency. The higher the frequency, the lower the reactance of a given value of capacity. Capacity is measured in micro-farads (uF) or micro-micro-farads (picofarads) (uuF or pF).

RF Coil

An rf coil is made of insulated copper wire wound on a plastic or cardboard tube. When electric current flows through the coil, a magnetic field is created. This magnetic field resists any change in electric current flow. This opposition or "reactance" to ac varies with frequency. The higher the frequency, the higher the reactance of a given value of rf coil inductance. Inductance is measured in millihenrys (mH) and microhenrys (uH).

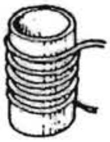

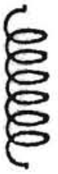

Tuned Circuit

A tuned circuit is composed of a capacitor and an rf coil. As previously described, the reactances of a capacitor and an rf coil change with frequency (but in opposite ways). At a particular frequency, the two reactances will cancel each other out. This condition is called "resonance". The impedance of a parallel tuned circuit is at its highest at resonance. Radio signals at the resonant frequency will be greatly magnified, while signals at other frequencies will be shunted to ground (rejected).

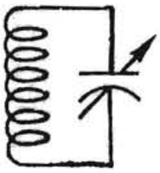

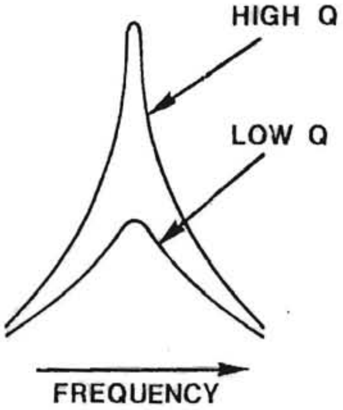

The resonant frequency of a tuned circuit can be changed by adjusting the capacitance (using a variable capacitor), or by varying the inductance (using a slider or taps on the rf coil). A ferrite core can also be used to change the inductance of a coil. The selectivity of a tuned circuit is dependent upon the quality (Q) of the components and the inductance /capacity ratio (L/C ratio). A high-Q tuned circuit is more selective than a low-Q circuit. A tuned circuit with a high L/C ratio (large inductance and small capacity) is more selective than a low L/C ratio circuit.

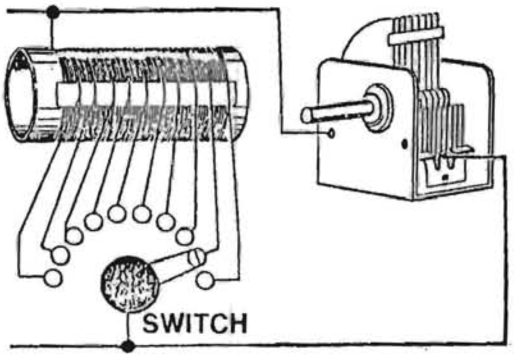

ANTENNA AND GROUND

A simple crystal set depends upon the strength of the radio waves received by the antenna to produce audio in the earphones. If you live in a strong signal area (a nearby radio transmitter), a crystal set can produce loud audio in earphones with only an inside antenna. But in lower signal areas, a good outside antenna with a ground connection is required.

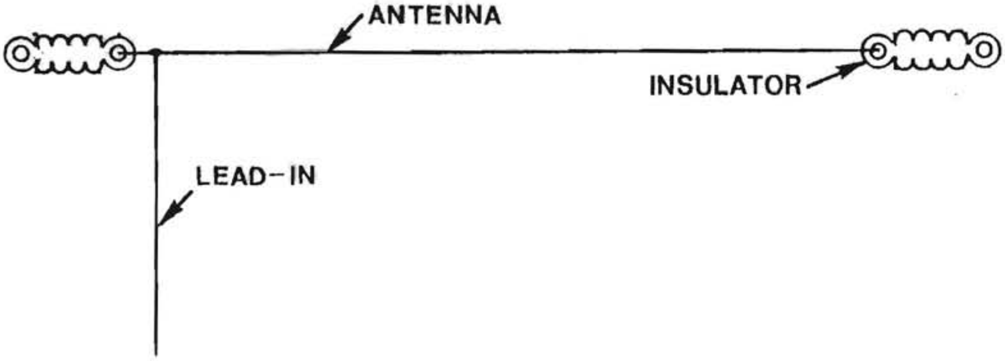

If you are fortunate enough to be able to put up an outside antenna, the basic rule of thumb is the higher, the better reception. Check your local electronics store for an antenna kit. You can construct an antenna using a glass or plastic insulator at each end of a length of copper wire (insulated or bare). Make the antenna as long as possible. It does not have to be in a straight line to work.

CAUTION

ANTENNA INSTALLATION
NEAR ELECTRICAL POWER
LINES IS DANGEROUS.

Make sure that your antenna is far enough away from any power line, so that it will not touch a wire if it falls.

Connect a length of insulated wire between the end of the antenna and your crystal set antenna terminal. This antenna lead-in wire can be fed into your house through a window.

If you can not erect an outside antenna, run a length of insulated wire as long and high as possible through your house or apartment. It will work, but not as well as an outside antenna because of the shielding affect of the house walls.

Connect a ground lead to a cold water pipe. Make sure that it is a metal pipe. Or, you can buy a ground rod from your local electrical supply store and drive it into the ground near your house. Then connect your crystal set ground lead to the rod. For safety install a lightening arrestor between the antenna lead and ground (see the instructions with the device).

HEADPHONES

Headphones (also called earphones, or headset) are required to hear the received audio from a crystal set. The standard headphones are made with dual coils of fine wire connected in series in each phone. Horseshoe magnets are built into each phone case to provide a magnetic field. A disk of thin sheet iron (diaphragm) is mounted near the top of the coils. The phones are connected in series with a set of flexible leads terminated in metal "phone tips".

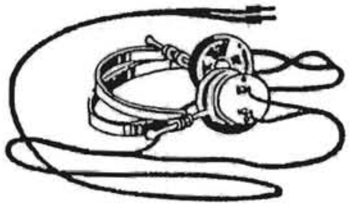

Detected audio signals flow through the coils and affect the intensity of the magnetic field from the horseshoe magnet. This changes the magnetic attraction of the diaphragm, and causes it to vibrate and produce audible sounds.

The total resistance of the series headphone coils is used as a measure of audio sensitivity. A resistance of 2000 ohms is sufficient. More sensitive headphones have a resistance of over 5000 ohms and are used to hear very weak signals.

The 2000 ohm headphones can be obtained at electronic parts mail order houses.

2

SIMPLE RECEIVER PROJECTS

SIMPLE CRYSTAL SET

Look at the schematic diagram. The crystal set uses a tunable ferrite coil L1, and a crystal diode D1. This type of tuning coil is called a "loopstick". The ferrite core is screwed into the coil form and increases the inductance of L1, thus tuning the receiver to a lower frequency.

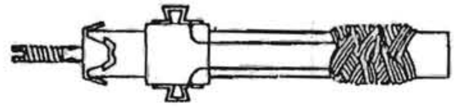

CONSTRUCTION

As shown in the illustration, the crystal set components are mounted on a 5 1/2-inch by 2 1/2-inch by 3/4-inch wood base. Layout and install the crystal set parts in the locations shown. The loopstick L1, is mounted with a metal strip supplied with it. Push the snap-in end of L1 into the large hole of the metal strip. Then bend 1/2-inch of the other end of the strip to form a bracket and mount it on the base with a wood screw. If the metal strip is not available, make one out of sheet metal approximately 2-inches long by 1/2-inch wide. Then drill a hole to fit L1 at one end, and a small hole to fit a wood screw at the other end.

Mount the four fahnestock clips and solder lugs with wood screws in the locations shown in the illustration.

WIRING

Use hookup wire to connect the crystal set components as shown in the schematic and illustration. The loopstick L1 may have three terminals. One of the terminals is a tap, which is not used in this circuit. Make sure that you connect only to the terminals at the start and end of the loopstick coil.

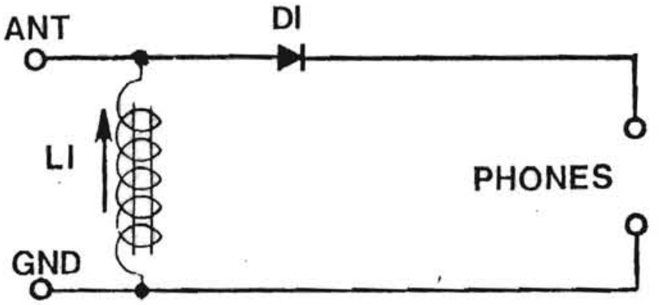
ANT
DI
LI
PHONES
GND

SIMPLE CRYSTAL SET

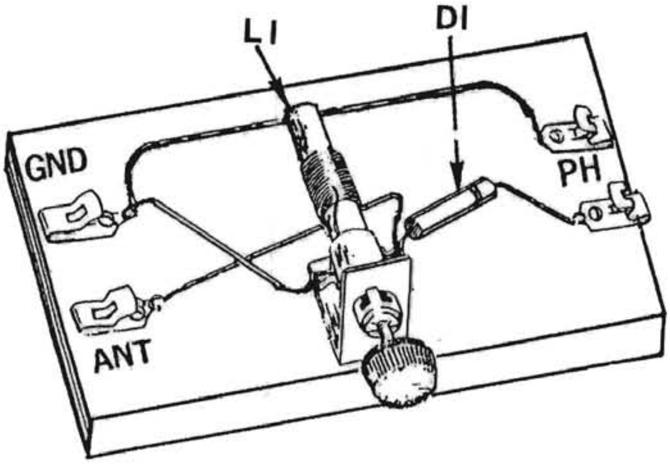
LI
DI
GND
PH
ANT

List of Materials

D1 - Germanium diode
(1N34A or equiv.)
L1 - Loopstick
(see text)
4 - Fahnestock clips
BASE - 5 1/2 X 2 1/2 X
3/4-inch wood section
PHONES - 2000-ohm headphones
MISC. - Solder lugs, wood screws, knob and bracket to fit L1, hook-up wire.

Install a small knob on the end of the L1 adjustment screw. Make sure that the screw turns freely with the knob. Position the 1N34A diode as shown before soldering the leads.

Experiment 1

Connect the antenna, ground, and 2000-ohm headphones to the crystal set terminals. Rotate the L1 knob counter-clockwise as far as it will go. Do not turn past this point, or you may break the ferrite rod inside the coil form. The ferrite is now outside of the coil winding. This is the highest frequency point of the tuning range of L1.

Slowly turn the L1 knob clockwise and listen for received radio stations. When the L1 knob is turned fully clockwise, the ferrite will be completly inside the coil (increasing the inductance) and it will be at the lowest frequency point of the tuning range.

Experiment 2

Connect capacitors, with values from 50 mmf to 300 mmf, in parallel with L1. Then tune L1 and note how the frequency range has changed.

AMPLIFIER CONNECTION

The simple Crystal Set (and the other receiver projects) can be connected to an amplifier in place of earphones. This amplifier can be a single channel input of your home stereo.

Parts values are not critical; C1 is approximately .005 mfd, and the volume control R1 can be any value from 1000 ohms to 10,000 ohms. Use shielded wire between R1 and the amplifier input jack.

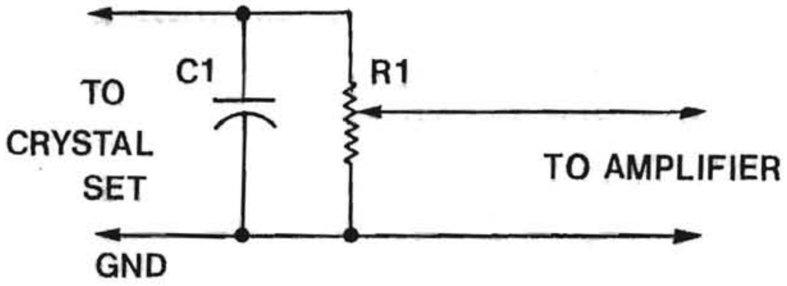
C1
R1
TO
CRYSTAL
SET
GND
TO AMPLIFIER

Amplifier Connection

Schematic Symbols

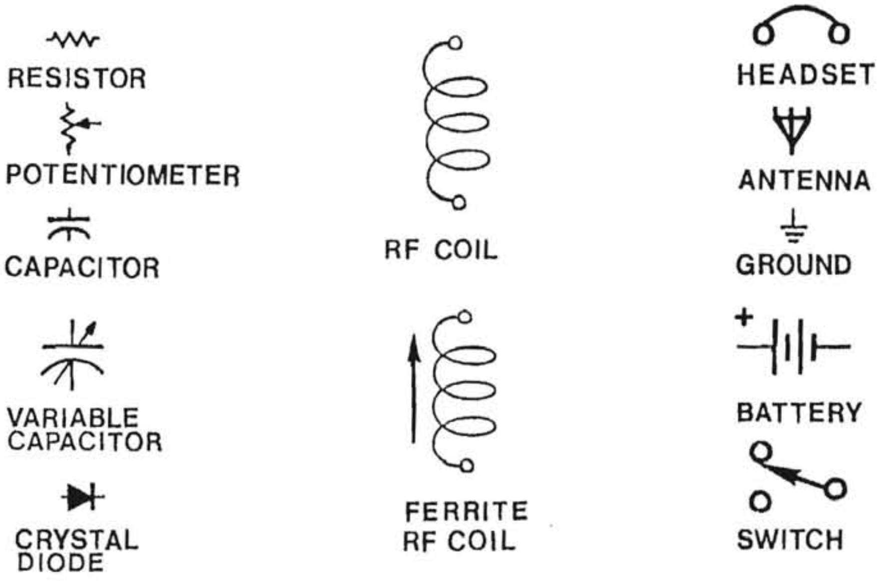
RESISTOR
POTENTIOMETER
CAPACITOR
VARIABLE CAPACITOR
CRYSTAL DIODE
RF COIL
FERRITE RF COIL
HEADSET
ANTENNA
GROUND
+
BATTERY
SWITCH

List of Materials

D1 - Germanium diode, type 1N34A, 1N60, 1N82A, or equiv.
L1 - Slider-tuned coil (see text)
4 - Fahnestock clips (terminals)
BASE - 6 x 3 x 1/2-inch section of wood (see text)
PHONES - 2000 ohms headphones
MISC. - Coil form (see text), No. 24 enameled copper wire, 3/32 inch dia. brass rod, brass ball slider (see text), solder lugs, two small brackets, hook-up wire, wood screws, machine screws and nuts.

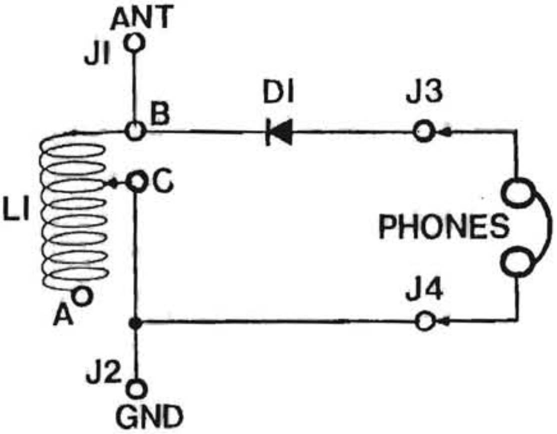
ANT
J1
B
D1
J3
C
L1
PHONES
J4
A
J2
GND

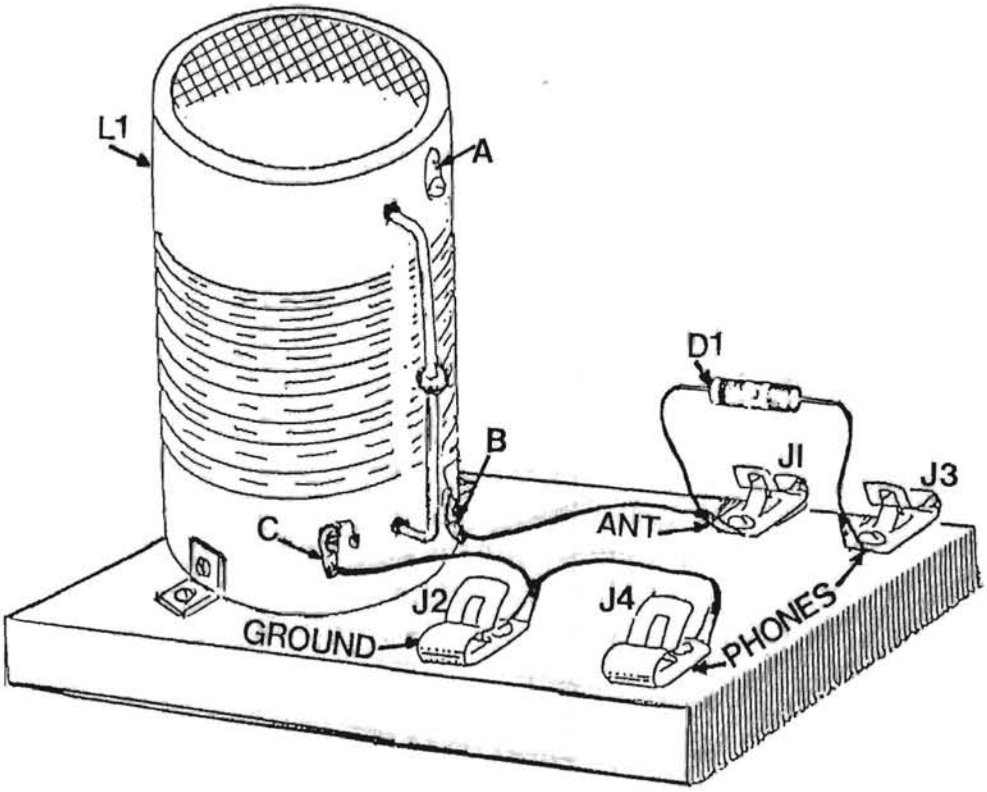
L1
A
D1
B
J1
J3
C
ANT
J2
J4
GROUND
PHONES

SLIDER TUNED CRYSTAL SET

SLIDER TUNED CRYSTAL SET

In the early days of radio, slider tuned coils were used in crystal sets. This method of tuning the inductance uses an adjustable metal contact (slider) that moves over the coil. The slider makes contact with individual coil wires to tune in the radio stations.

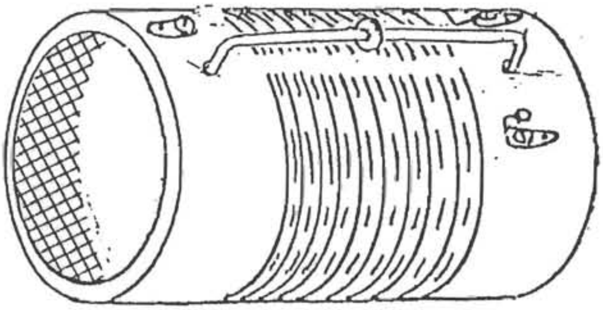

The enamel insulation of the coil wires in the slider path is removed for good electrical contact.

You can experiment with a slider tuned coil by building this crystal set. As shown in the illustration, the coil is mounted upright on a 6-inch by 3 1/2-inch by 1/2-inch wood base. A semiconductor diode is used for a detector.

TUNING COIL CONSTRUCTION

Look at the drawing of the tuning coil L1 construction details on the next page. The tuning coil is wound on a cardboard or plastic mailing tube section, 2 5/8-inches in diameter and 5-inches long. Make sure that you select a rigid, thick-walled tube for a coil form. The tube used in our model has a 3/16-inch wall. The coil form must be rigid to hold the slider ball and rod assembly.

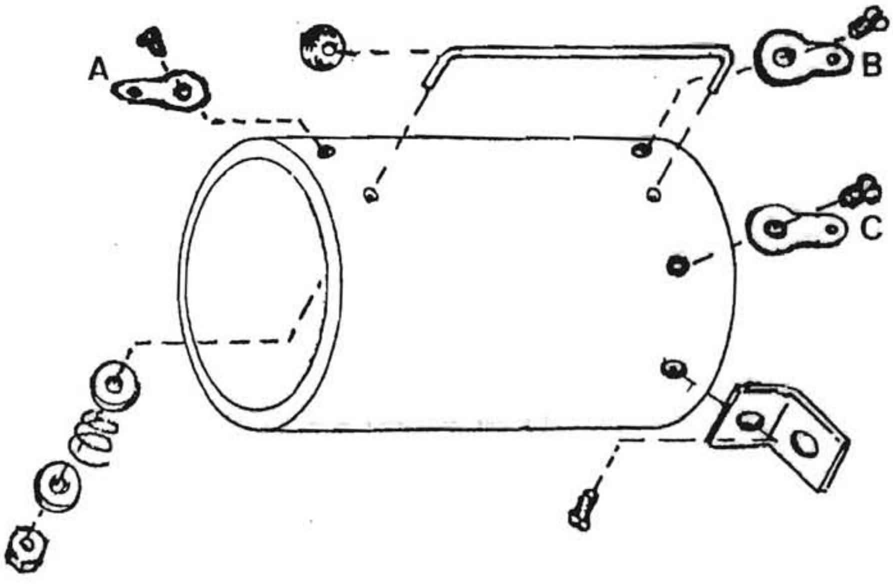

Before winding the coil, drill all the holes in the form as shown in the drawing, and then install the solder lugs. Position the solder lug ends toward the coil form edges.

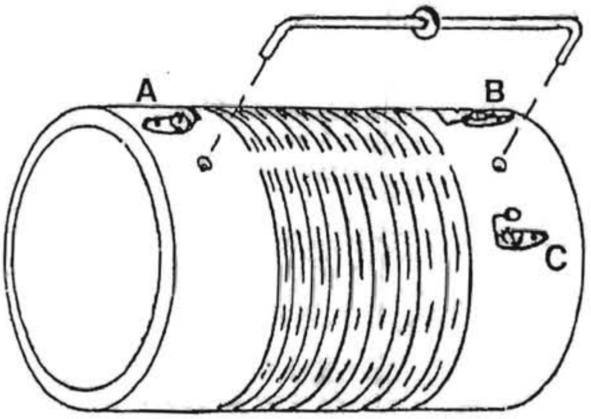

The coil is wound with No. 24 enameled copper wire. Remove the enamel from the wire end with sandpaper and then solder the wire to the lug. Start winding the coil, keeping the turns as tight and close together as possible. Wind the wire to the front of the bottom solder lug B. Remove the enamel and solder the wire end to the lug.

A brass ball approximately 3/8-inches in diameter is used as the moving contact slider. This brass ball can be found at your local electrical supply store. It is a type of ball that is used as a top nut for small table lamps. Drill a hole through the ball so that it will slide over a length of 3/32-inch diameter brass rod. Bend the brass rod at each end (after placing the ball in the center of the rod), to fit through the holes in the coil form as shown in the drawing. Cut the rod 1-inch down from the bend at each end.

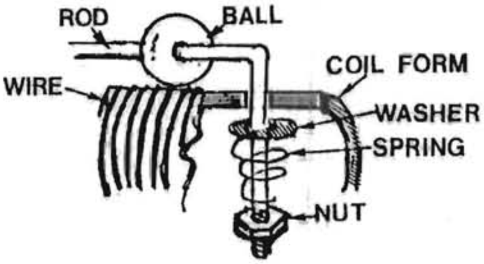

Insert the brass rod with the slider ball into the coil form holes. Then place a small washer and spring on one end of the rod (inside the coil form). Solder a nut on the end of the rod to hold the spring in place. Do the same to the other end of the rod. After soldering, slide the ball over the wires and check that it slides smoothly. Use sandpaper to remove the enamel insulation of the wires in the path of the ball slider. Connect a wire between the soldered nut on the bottom end of the brass rod and the solder lug near the rod.

BASEBOARD

All of the crystal set components are mounted on a wood baseboard as shown in the drawings. Cut the baseboard to size; a 6-inch by 3 1/2-inch section of 1/2-inch thick wood (maple is used in our model). Sand the edges after cutting. It is not necessary to paint or stain the baseboard.

Mount the four terminals and solder lugs as shown in the drawing. Label the four terminals with either rub-on lettering or cemented paper tags. Install two brackets on opposite sides of the bottom of the coil, and mount it on the baseboard with the lugs toward the terminals.

WIRING

Bare No.22 wire was used to wire our model, but any kind of hookup wire can be used. Connect the wires between the solder lugs as shown in the drawing, except do not connect the antenna terminal lug wire to the coil terminal at this time. Connect the crystal diode between the solder lugs as shown.

Experiment 1

This experiment will show how an untuned crystal detector functions.

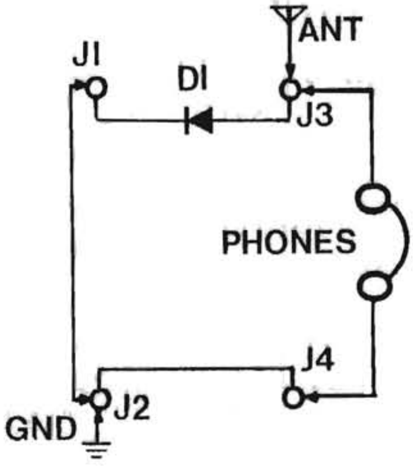

Connect the untuned detector circuit as shown in the drawing, and then hook-up your antenna and ground leads. Connect 2000-ohm headphones to the circuit and listen to the detected radio waves.

You will probably hear a strong local radio station, with other, weaker radio station transmissions all jumbled together.

Experiment 2

Reconnect the baseboard wires to the circuit of the Slider-Tuned Crystal Set as shown in the drawing. Listen to the detected radio waves as you tune the coil by sliding the brass ball over the sanded wire path. You should be able to receive the radio stations much better than the untuned detector circuit of Experiment 1.

LOADING COIL

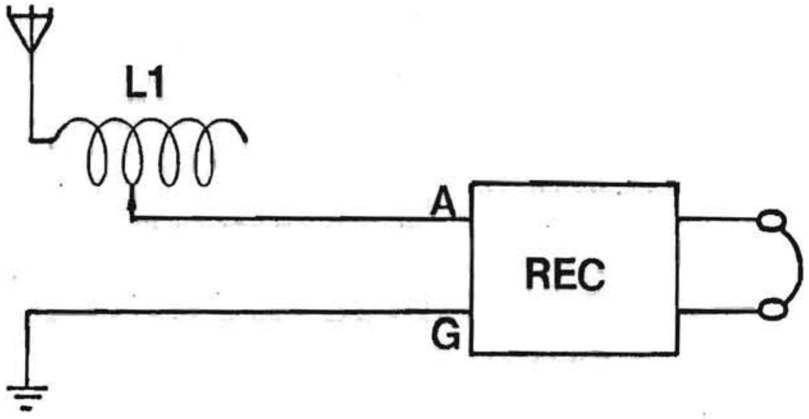

An additional slider tuned coil, L1, can be used as a Loading Coil to increase the electrical length of your antenna. This may allow better reception of the lower broadcast band frequencies. Connect L1 in series with the antenna lead to the receiver, and tune it for best reception.

WAVE TRAP

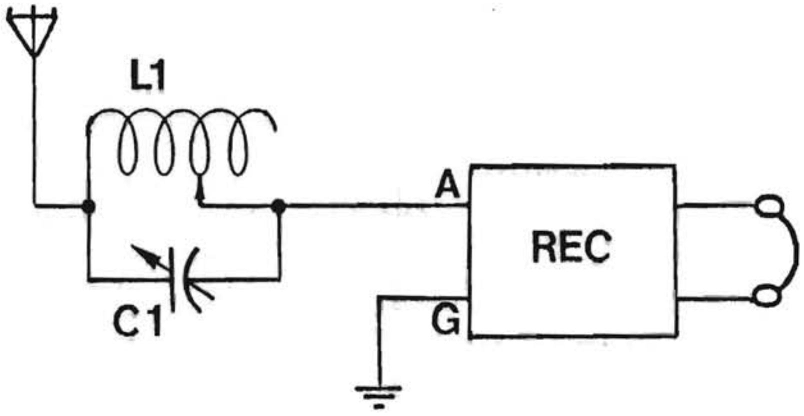

L1 can also be used in a Wave Trap circuit to tune out a strong local signal, and allow easier tuning of weaker stations. C1 is a 365 mmf variable capacitor connected in parallel with L1. Adjust L1 and C1 so that the interfering station is at minimum.

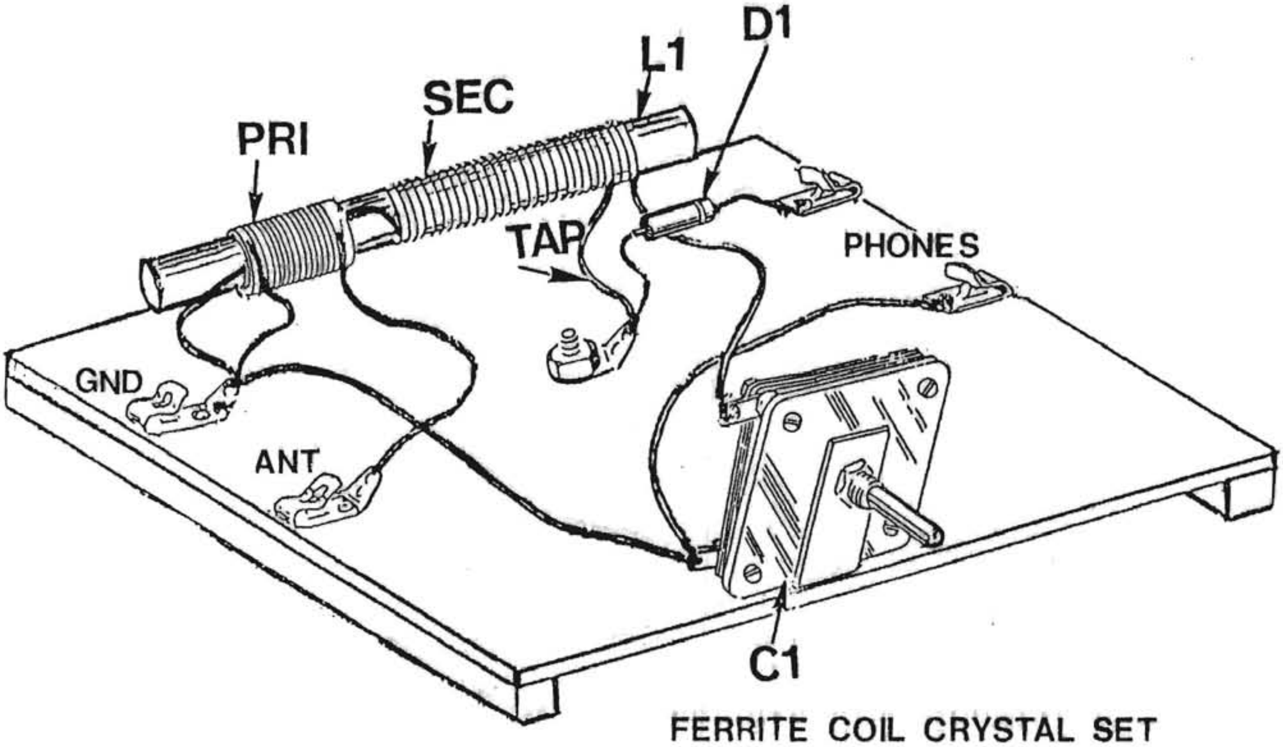
D1
L1
SEC
PRI
TAP
PHONES
GND
ANT
C1

FERRITE COIL CRYSTAL SET

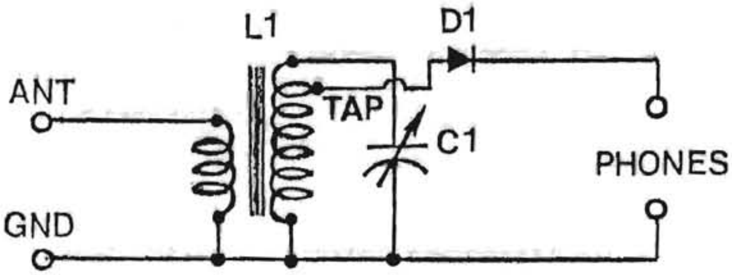
L1
D1
ANT
TAP
C1
PHONES
GND

List of Materials

C1 - 365 mmf variable capacitor.
D1 - Germanium diode (1N34A or equiv.).
L1 - Ferrite coil (see text).
4 - Fahnestock clips.
2 - wood sections; 6 X 1/2-inch square (see text).
BASE - 8 1/2 X 6-inch sheet plastic (see text).
PHONES - 2000-ohm headphones
MISC. - plastic tube section for L1 primary winding (see text), solder lugs, machine screws and nuts, wood screws, bracket for C1 (see text), knob for C1, hookup wire.

FERRITE COIL CRYSTAL SET

This crystal set has a large ferrite core coil L1 of a type that is normally used for a built-in antenna in portable radios. This type of coil has a very high "Q" and may be more selective than an air-core coil. As you can see in the schematic, a primary coil is wound around the ferrite coil. The original winding is used as the secondary.

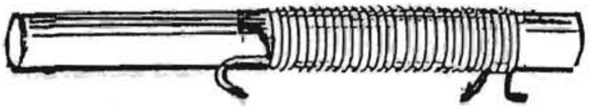

The primary coil (connected to the antenna and ground) can be moved close to, or away from, the secondary (original winding connected to D1). This variable coupling is used to control the selectivity of the crystal set.

PRIMARY COIL

The primary coil is wound on a 1-inch section of plastic tube that can fit loosely over the ferrite rod. A 1/2-inch diameter plastic tube is used on the receiver shown in the illustration. Wind 12-turns of hookup wire evenly over the tube and twist the ends together to prevent movement. Leave 4-inch leads from the primary coil winding. Slide the primary coil over the center of L1.

BASEBOARD

The crystal set components are mounted on a 8 1/2-inch by 6-inch section of sheet plastic. The thickness of the plastic is not critical, but it must be rigid enough to hold the parts without bending. A 6-inch long 1/2-inch square wood section is installed under each end of the baseboard. A section of wood can be used for the baseboard with wood screws to mount the parts, in place of machine screws and nuts.

The model shown uses a plastic variable tuning capacitor, but an air type capacitor can be used for C1 instead. Make a metal bracket approximately 1-inch wide by 1 1/4-inch high and a 1/2-inch right angle mounting foot from sheet aluminum. Drill a hole in the top of the bracket to fit C1 and two holes in the mounting foot to fit small machine screws. Locate the components on the baseboard as shown in the illustration and then drill mounting holes. If you have a cardboard or plastic mounting supplied with ferrite coil L1, you can mount it on the baseboard. If not, L1 can be taped loosely on top of the baseboard. Th ferrite coil L1 in our crystal set was obtained from an old radio. Install the components on the board with solder lugs, machine screws, and nuts.

WIRING

Connect the parts with hookup wire as shown in the illustration. Use the schematic diagram as a guide. Make sure that the ground lead from L1 secondary passes inside the primary coil. Also check that the primary coil slides freely over the L1 winding. Install a knob on the C1 shaft, and mark the terminal designations on the baseboard.

Experiment 1

Hookup the antenna, ground and 2000-ohm headphones to the crystal set terminals. Slide the primary coil over the center of the L1 winding. Tune C1 for radio stations. Then slide the primary coil towards the ground end of L1 while tuning C1. Note the difference in selectivity.

Experiment 2

Change the number of turns of the primary winding and repeat Experiment 1.

Operating Tips

After operating a crystal set for some time, you may notice a difference between day, evening, and night reception of radio stations. This is due to the two ways that radio signals travel between the transmitting antenna and a receiving antenna.

(a) The ground wave; which moves along the surface of the earth.
(b) The sky wave; which travels upward and is reflected by the ionosphere.

The reception of broadcast stations is dependent upon the time of day. During daylight hours, stations are received primarily by ground wave transmission. At night, the broadcast stations will be received via sky wave transmission. Also, some radio stations change to a lower power at night. Reception in the evening may be more difficult as the fading ground wave is interfered with by reflected waves. But later, it may be possible to hear far-away stations transmitted by sky wave.

Your geographical location to radio stations is very important for reception, as a crystal set receiver does not amplifiy signals. It is entirely dependent upon the strength of the radio waves. If you are in a deep valley surrounded by high mountains, reception will be difficult. If you live in a flat plain, with no obstacles to radio transmission, reception may be very good.

An outside antenna is necessary for for best reception. If possible, try erecting an extra antenna at a right angle to the original one, and switch it in to provide a directional effect for increased signal strength. Keeping a log of received stations for a week may be helpful in determining the radio reception conditions in your area.

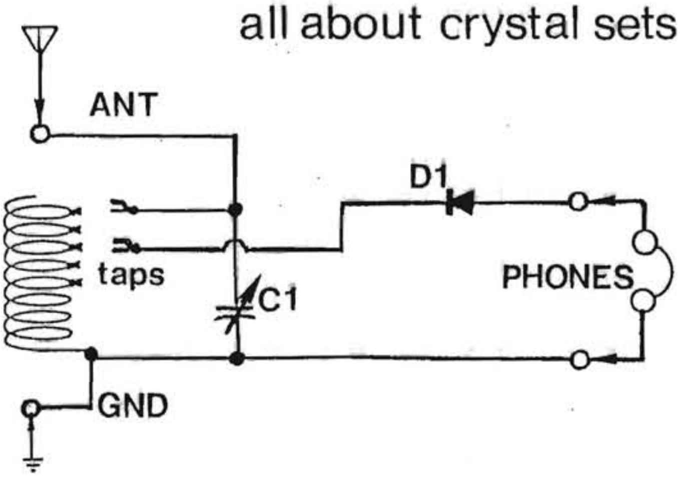
ANT
D1
L1
taps
C1
PHONES
GND

List of Materials

C1 - 365 mmf variable capacitor.
D1 - Catwhisker crystal detector (or 1N34A or equiv. diode).
L1 - Spiderweb coil (see text).
4 - Fahnestock clips.
2 - small alligator clips
BASE - 8 X 6 X 1/2-inch section of wood (see text).
MTG. POST - 5 x 3/4 x 3/8-inch wood section (see text).
PHONES - 2000 ohms headphones.
MISC. - Coil form (see text), No. 22 enameled copper wire, solder lugs, hook-up wire, wood screws, machine screws, and nuts.

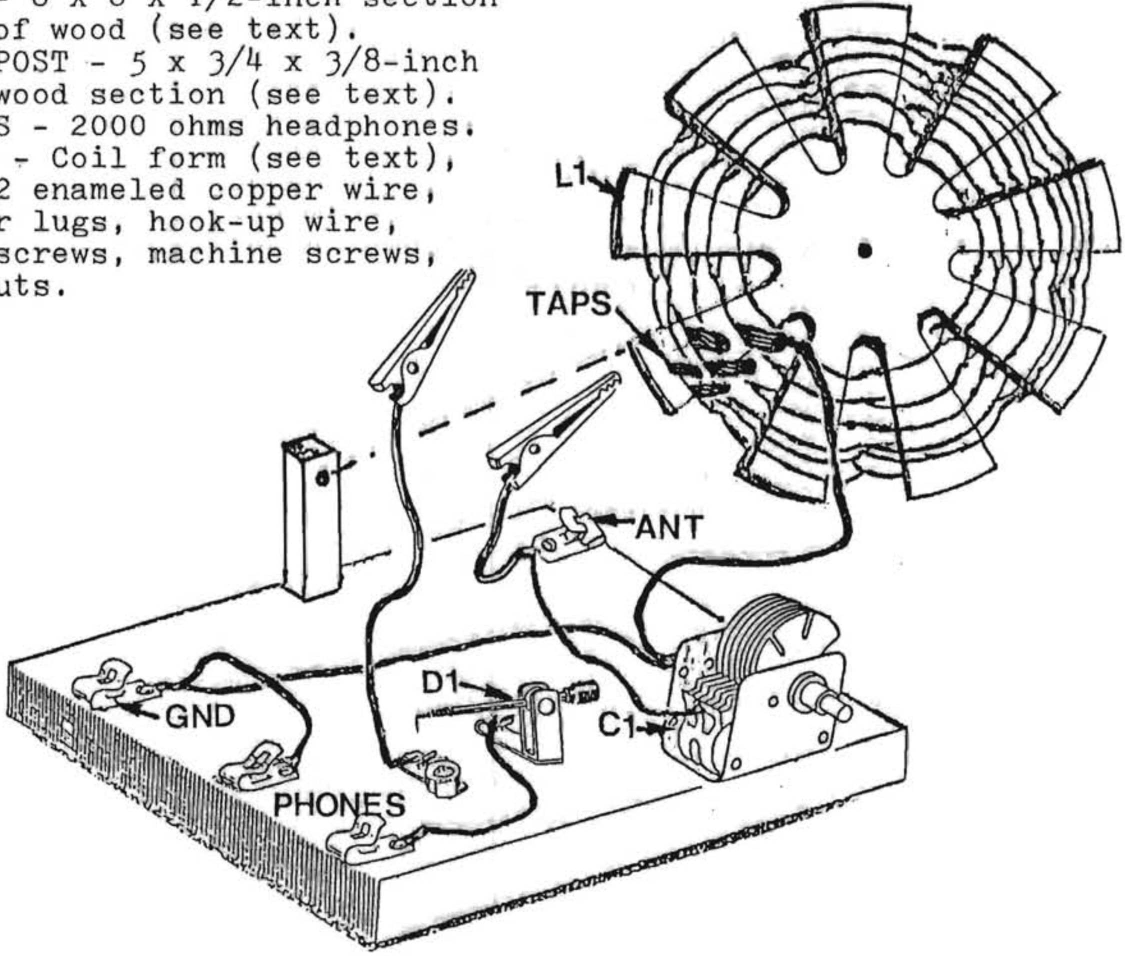
L1
TAPS
ANT
D1
C1
GND
PHONES

SPIDERWEB COIL CRYSTAL SET

SPIDERWEB COIL CRYSTAL SET

The spiderweb coil is a tuning coil that is wound on a flat wheel form that has an uneven amount of spokes. This type of coil was also used as a loop antenna in early radios.

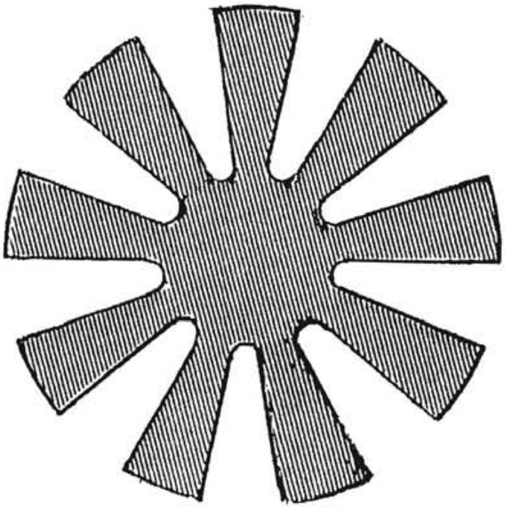

Our receiver uses a tapped spiderweb coil that is tuned by a variable capacitor. As shown in the illustration, the spiderweb coil is mounted at the rear of a 8-inch by 6-inch by 3/4-inch wood base. A catwhisker crystal detector and a variable capacitor are mounted near the front of the wood base.

SPIDERWEB

The spiderweb coil is wound on a 6-inch diameter flat form that has nine vanes. Make a tracing of the spiderweb coil form template on the next page. Then use carbon paper to transfer the tracing to a section of 1/8-inch fibreboard or sheet plastic. Cut the coil form out with a jeweler's saw, and drill a center hole to fit a small wood screw.

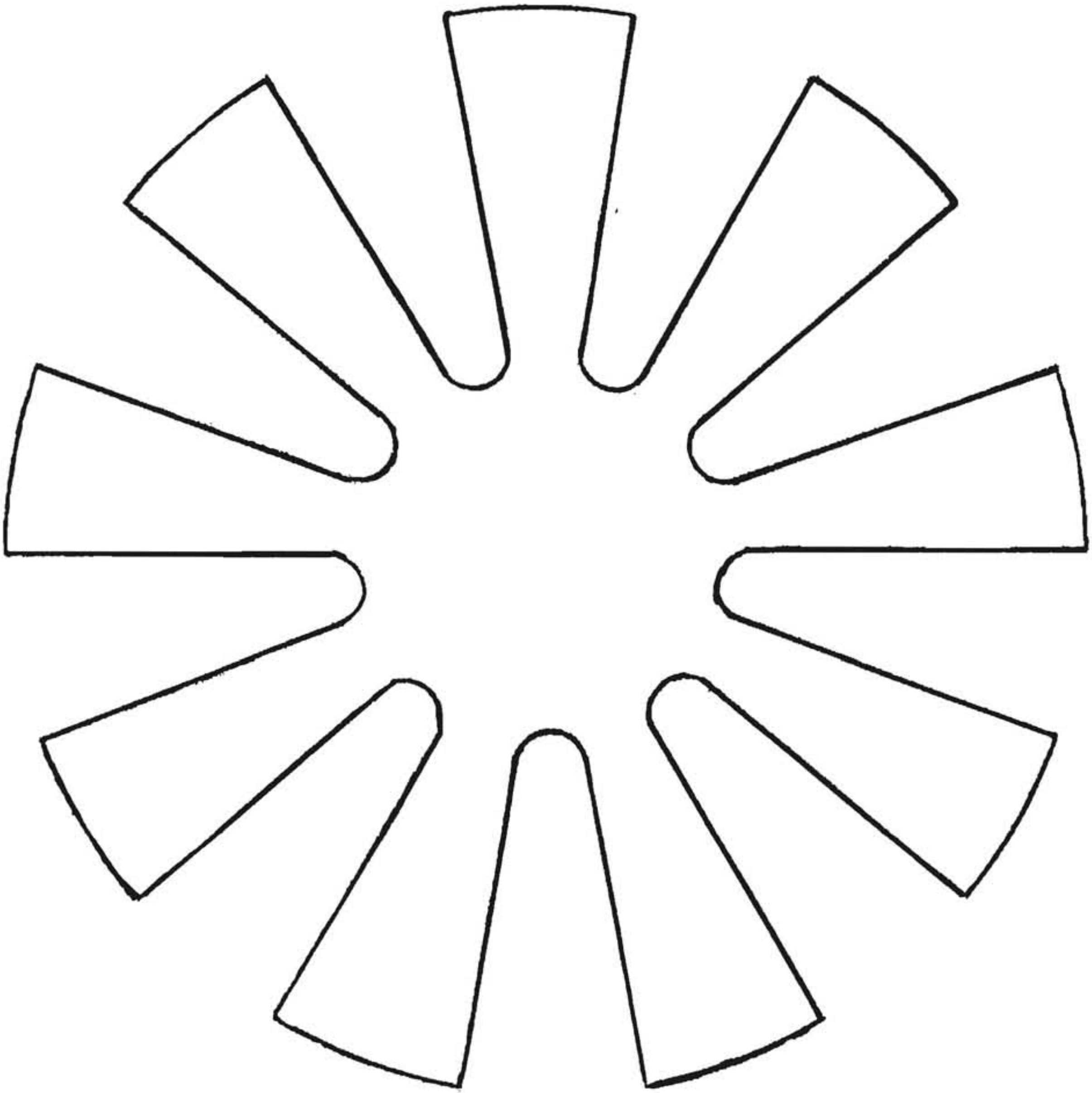

Spiderweb Coil Form Template

Wind the coil with No. 22 enameled copper wire, starting from the center and winding outward. Leave about 5-inches of the starting end of the wire free. It will be used as a connecting lead. Feed the wire behind a vane and over the front of the next vane as you wind. Wind the wire closely and evenly on the form. It is not necessary to count the turns. Wind the wire to about 1/4-inch from the end of a form vane. Then drill a small hole near the vane end, and bend a solder lug through it. Solder the wire end to the lug.

On one of the vanes, count down five wires and then carefully lift it outward. Remove about 1/4-inch of the enamel with a section of sandpaper and then bend the end of a solder lug aound the bared wire. Solder the lug to the wire, then cut off remaining portion of the lug (away from the bentover end). Repeat this every five turns until you have a total of six lug taps on the coil. Stagger the lugs away from each other to keep them from shorting.

BASEBOARD

As shown in the illustration, the Spiderweb Coil Crystal Set parts are mounted on a wood baseboard. Cut the 8-inch by 6-inch baseboard from a section of 1/2-inch thick wood. White pine was used in our model. Sand the surface of the wood after cutting.

Make the spiderweb coil mounting post by cutting out a 5-inch by 3/4-inch by 3/8-inch wood section. Then cut a notch in the center of the rear of the baseboard to fit the wood section. Install the wood section in the notch with small wood screws.

Mount the parts on the baseboard as shown in the illustration with small wood screws. If the mounting holes of the tuning capacitor C1 are on the bottom, drill matching holes in the baseboard and use flathead machine screws to hold the capacitor.

Install solder lugs with the four terminals, and mount the spiderweb coil to the mounting post with a wood screw through a center hole in the coil form. The coil should be positioned with the lugs facing forward (on the left side of the coil form).

WIRING

Connect the crystal set components with hookup wire as shown in the illustration. Use 6-inch lengths of insulated stranded wire to the two small alligator clips.

Experiment 1

Connect antenna, ground, and 2000-ohm headphones to the crystal set. Then connect the D1 alligater clip lead to the antenna terminal. Set the tuning capacitor C1 to minimum capacity (rotor blades all the way out). Connect the antenna clip lead to the tap on L1 near the center of the form.

Adjust the crystal detector D1 catwhisker to a sensitive spot and tune C1 for a radio station. It may be easier to tune for stations if you temporarily connect a crystal diode (1N34A or equiv.) in place of the catwhisker crystal detector.

After a station is heard, you can disconnect the diode and probe for a sensistive spot on the crystal with the catwhisker. Change the antenna clip lead to other L1 taps and continue tuning C1 for radio stations.

Experiment 2

Disconnect the D1 clip lead and reconnect it to a tap on L1 close to the center of the coil. This will load L1 more lightly than if the detector was connected to the top of the coil. Repeat the tuning procedure of Experiment 1 and note the difference in selectivity and volume of the received radio stations. Then tune in a radio station and move the D1 clip lead to other taps on L1. There will be a particular tap point for optimum selectivity and volume for each received station.

STATION LOG

DATE	TIME	DIAL	STATION	NOTES

3 ADVANCED RECEIVER PROJECTS

DUAL TUNED CRYSTAL SET

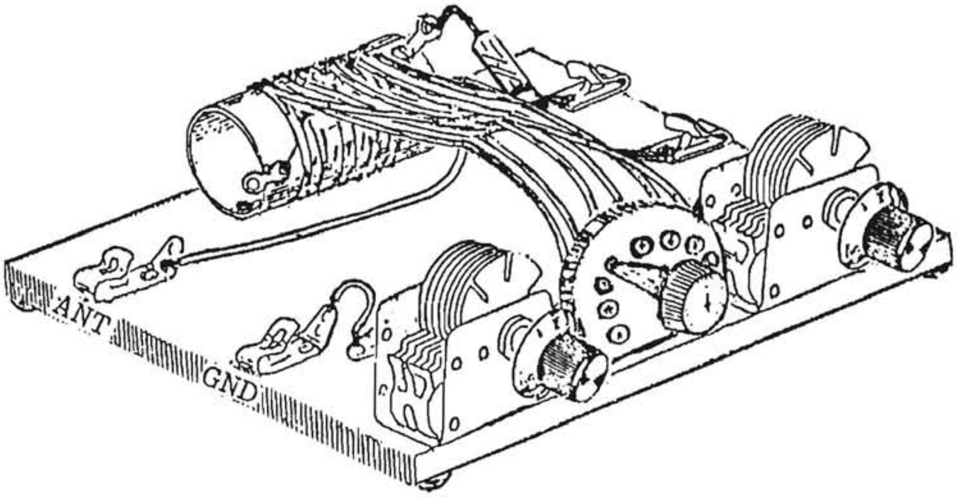

This crystal set circuit uses two variable capacitors C1 and C2. As you can see in the schematic drawing, C1 tunes the ground circuit, and C2 tunes a tapped coil L1. The taps of L1 are connected via the rotary switch S1.

TUNING COIL

Look at the drawing of the coil form construction details on page 34. The tuning coil L1 is wound on a section of 1/8-inch wall plastic tube, 2-inches in diameter, and approximately 4-inches long.

Drill holes in the form and install three solder lugs with machine screws and nuts. Position the solder lug ends towards the coil form ends as shown.

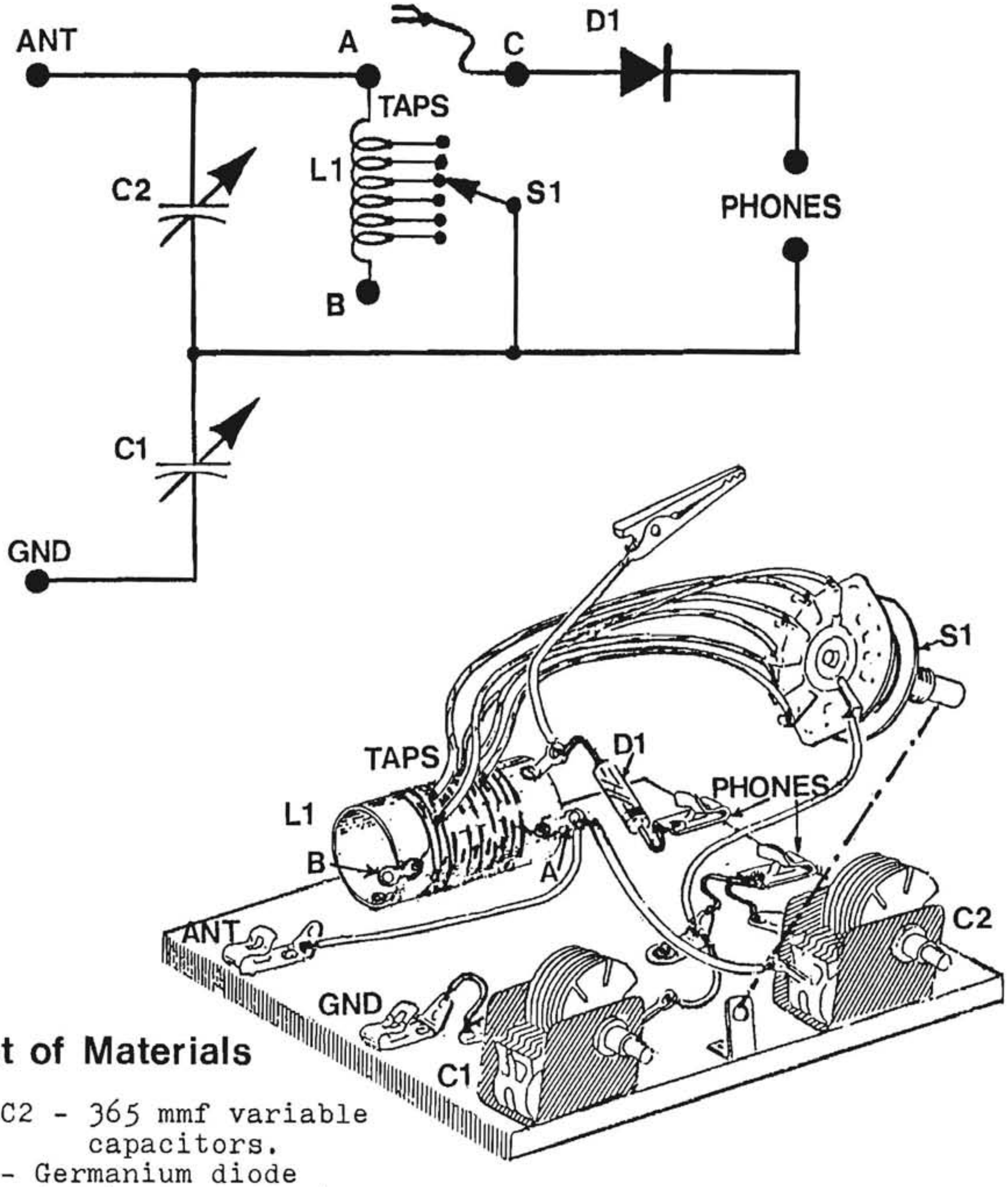
ANT
A
C
D1
TAPS
L1
S1
PHONES
C2
B
C1
GND
S1
TAPS
D1
PHONES
L1
B
A
ANT
C2
GND
C1

List of Materials

C1,C2 - 365 mmf variable capacitors.
D1 - Germanium diode (1N34A or equiv.).
L1 - Tuning coil (see text).
S1 - 1-pole, 10-position rotary switch (see text).
4 - Fahnestock clips.
1 - Small alligator clip.
BASEBOARD - 5 x 8 x 3/8-inch wood section (see text).
PHONES - 2000-ohm headphones.
MISC. - Coil form (see text), No. 24 enameled copper wire, solder lugs, hook-up wire, wood screws and nuts, knobs, bracket for S1.

DUAL TUNED CRYSTAL SET

The coil L1 is wound with No. 24 enameled copper wire, and is tapped every five turns for a total winding of 75 turns.

Drill a small hole in the coil form approximately 1 1/4-inch from the "A" lug. Remove the enamel from the end of the No. 24 wire and then feed the wire end through the hole. Solder the wire to lug "A".

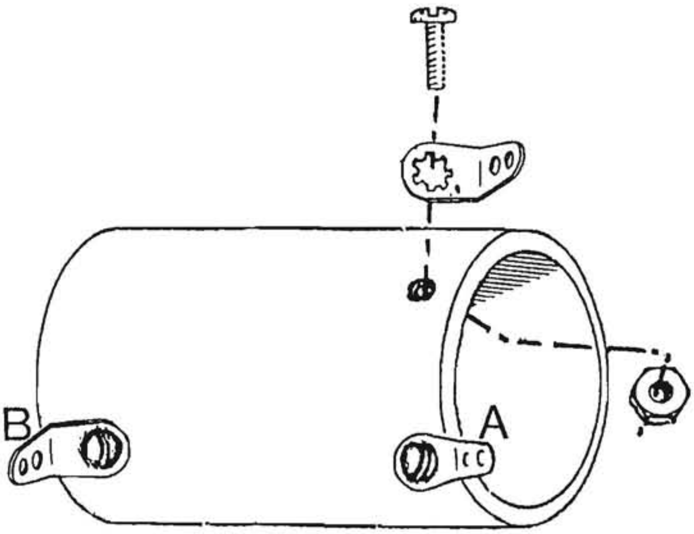

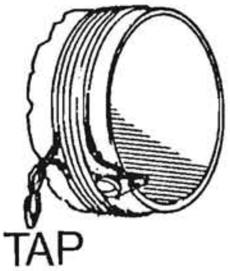

Wind five closely spaced tight turns around the coil form. Then bend back and twist together 1/2-inch of the wire to make a tap. Position the tap outward and then continue winding the coil, keeping the turns as tight and close together as possible. Make a tap every five turns and then drill a small hole in the coil form after the 75th turn. Feed the wire through the hole, remove enamel, and solder the end to lug "B".

BASEBOARD

The Dual-Tuned Crystal Set is built on a wood baseboard as shown in the illustration. Cut the 5-inch by 8-inch baseboard from a section of 1/4-inch thick plywood. Sand the edges and surface of the baseboard to remove any splinters. Drill holes in the baseboard and install the parts with small machine screws and nuts in the locations shown in the illustration. Drill mounting holes at each end of the tuning coil L1 and position it on the rear of the baseboard with the coil taps toward the front. The illustration shows a commercial rotary switch, but a wooden rotary switch can be used for S1 instead. (See page 53 for construction details.)

Install the four fahnestock clips (used as antenna, ground and headphone terminals) with solder lugs. Mount an additional solder lug on the baseboard midway between S1 and L1.

WIRING

Connect the crystal set parts with hookup wire as shown in the pictorial illustration and the schematic diagram. Use the center solder lug as a junction point for leads from C1, C2 and S1 as shown in the illustration. Connect the leads from S1 to the taps on L1 in numerical sequence; every 5 turns for 7 taps. Then connect the last three leads every 10 taps. Insert the lead ends into the taps and hold them with a drop of solder. This will make it easier to move leads to other L1 taps. Use a 5-inch length of stranded hookup wire for the alligator clip lead connected to D1 at solder lug C.

Experiment 1

Connect 2000-ohm headphones, antenna and ground leads, to the crystal set terminals. Then connect the D1 clip lead to the L1 "A" solder lug.

Set S1 to each tap position in turn and tune C1 and C2 for received radio stations. If all of the local stations can not be heard, experiment with different connections to the L1 taps. More turns may have to be added to L1 (with additional five-turn taps) to receive stations at the lower frequency end of the broadcast band.

Experiment 2

Remove the D1 clip lead from "A" and then connect it to the "B" solder lug on L1. Repeat the tuning method of Experiment 1 and note if the crystal set is more selective. This is possible, because L1 is now connected as an rf transformer. You can also connect the D1 clip lead to each of the S1 taps and note if there has been any improvement in reception.

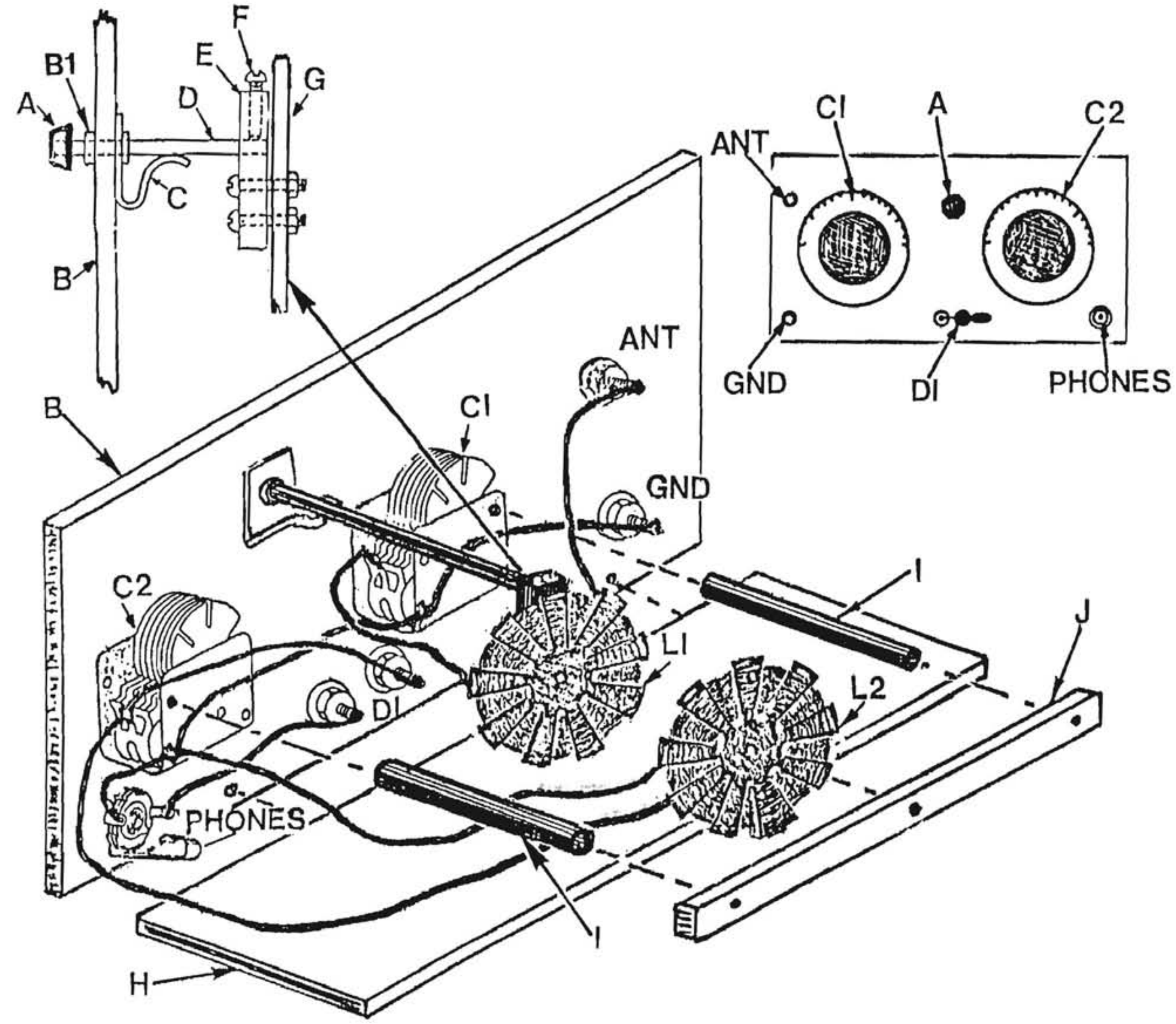
F
B1
E
G
A
D
C
B
CI
A
C2
ANT
GND
DI
PHONES
ANT
B
CI
GND
C2
I
J
LI
L2
DI
PHONES
I
H

SELECTIVE CRYSTAL SET

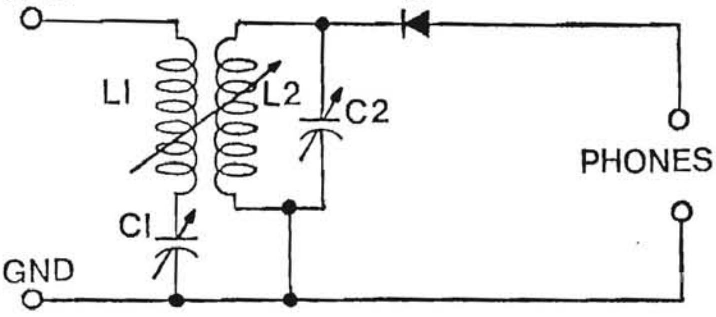
ANT
DI
LI
L2
C2
PHONES
CI
GND

SELECTIVE CRYSTAL SET

This receiver uses two spiderweb coils in a two-circuit tuner for good selectivity and sensitivity. As you can see in the schematic diagram, the antenna-ground rf input circuit is tuned by L1-C1 for maximum signal gain. L1 acts as the primary winding of an rf transformer, with L2-C2 as the tuned secondary. Coupling between L1 and L2 is variable for best selectivity. L2 is connected to a catwhisker crystal detector and the audio signals are fed through the phone jack to 2000-ohm headphones.

As shown in the illustration, the crystal set parts are mounted on the rear of a black plastic panel. This type of layout is similar to the type of construction of the old-time crystal sets.

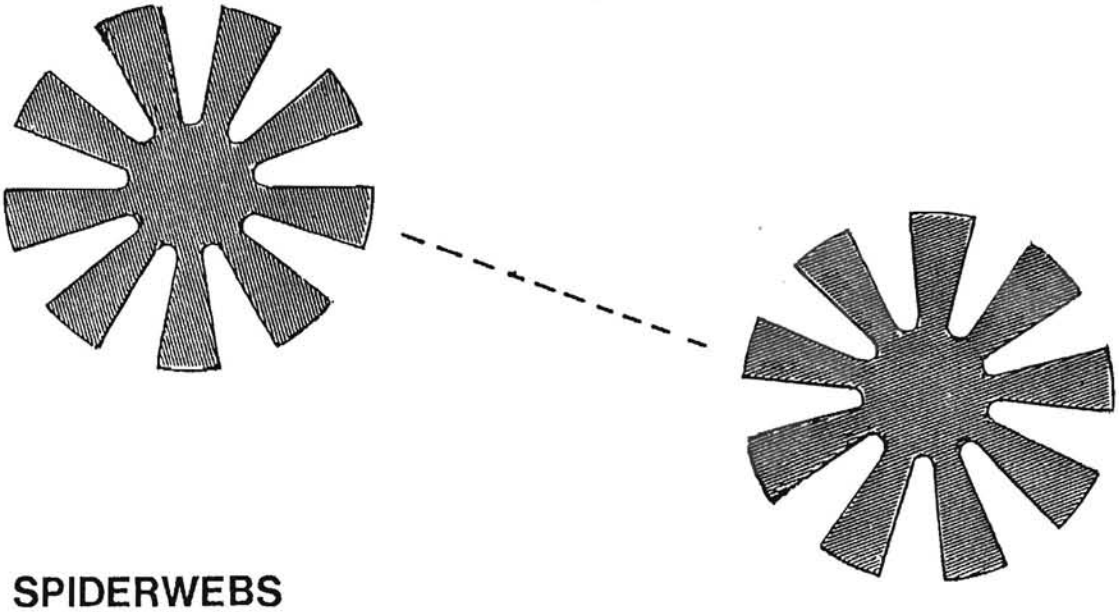

SPIDERWEBS

Both L1 and L2 spiderweb coils are made exactly the same way. L1 and L2 are wound on 17-vane coil forms 3 5/8-inches in diameter. The coil forms in our receiver are made from 1/16-inch sheet plastic of the same type used for printed circuit boards (but without the copper foil). Any kind of rigid sheet plastic should be suitable for the coil forms.

Make a tracing of the spiderweb coil template. Then use carbon paper to transfer the tracing on a section of 1/16-inch thick sheet plastic. Cut out the vanes with a jeweler's saw. Drill two holes as shown, and install the two solder lugs.

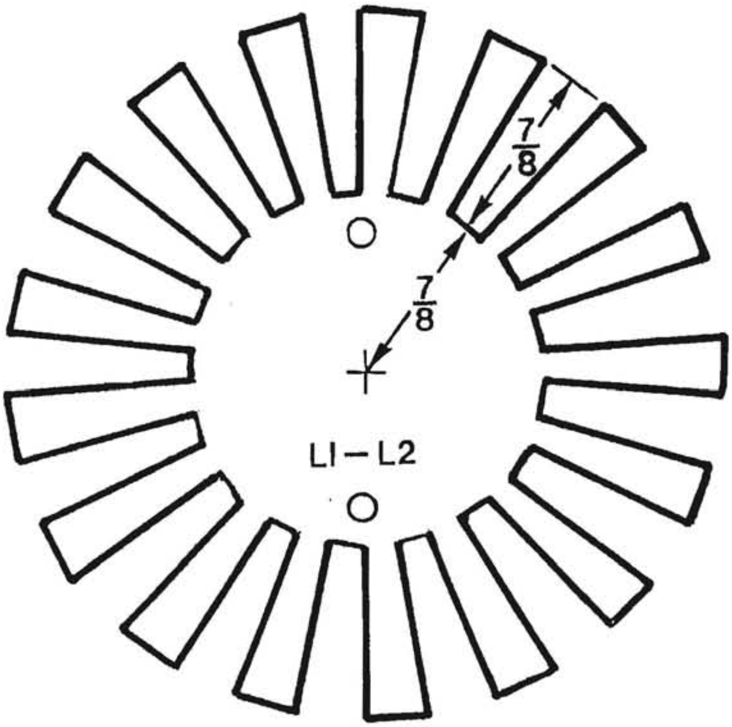

COIL FORM TEMPLATE FOR L1–L2

Wind as much No.22 enameled wire as possible on each coil form. Solder the wire to a solder lug before winding a coil. Then wind the wire over a vane and under the next vane. Wind the wire closely and evenly on the form. It is not necessary to count the turns. Solder the wire end to the remaining solder lug

CONSTRUCTION

The panel used in our receiver is a 7 3/4-inch by 4 1/2-inch section of 1/4-inch thick black acrylic plastic. The size is not critical, but it should be large enough to mount all of the parts as shown in the illustration.

Layout and drill the mounting holes of the components on the panel. Make sure that C1 and C2 are positioned so that the edges of the tuning knob dials are more than 1/2-inch apart. Install the catwhisker crystal detector assembly, antenna and ground terminals, and the phone jack in the locations shown.

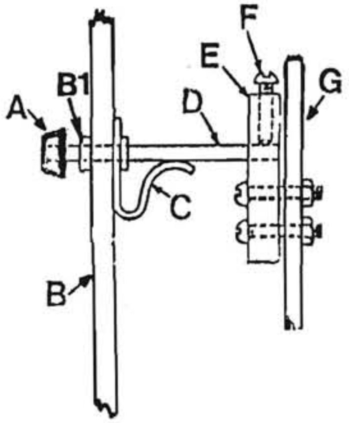

As you can see in the illustration, L1 is installed on the end of a variable coupling metal rod. This 1/4-inch diameter rod is 3-inches long with a flat side. Drill and tap a small hole close to the end of the rod. The rod used in our model came from a discarded volume control. Cut a 2-inch long by 1/2-inch wide by 3/8-inch thick wood section and drill holes to fit the rod, two mounting screws, and a hole for the rod set screw. Locate and mount the wood section in the center of a spiderweb coil L1 so that the rod will be 3/4-inch down from the edge of the top vane.

Hold L1 behind the panel and using the free end of the rod, locate the panel hole. Position L1 midway between C1 and C2, and make sure that the top edge of the coil is below the top of the panel. Drill a hole to fit the rod bushing in the panel hole location. The panel bushing in our model was removed from a discarded volume control.

Drill a hole to fit the bushing in one end of a 2-inch by 1/2-inch section of sheet aluminum. Bend the other end of the section to fit against the flat side of the rod. Install the metal section and the bushing in the panel hole.

Insert the rod into the bushing and make sure that the metal section end fits close to the flat side to prevent rotation of the rod. Install a knob on the rod end and make sure that the rod can be pushed in and out easily. If necessary, put a small amount of grease on the flat side of the rod.

Install the remaining spiderweb coil L2 on a wood section that is mounted on the rear of the panel with two spacers. The spacers should be long enough for L2 to be located 1/4-inch behind L1 with the rod pushed all the way in. The 5 1/2-inch by 1-inch by 1/4-inch wood section in our model is supported with screws installed through 1 1/2-inch metal spacers into tapped holes in the rear of C1 and C2. The screws and bushings can be made longer and mount into panel holes in your receiver.

Install a wood section on the bottom rear of the panel to keep the receiver upright. A 5-inch by 3-inch section of 1/4-inch thick wood is used in our receiver.

SET WIRING

Connect the receiver components with hookup wire as shown in the illustration and schematic diagram. Use flexible stranded wire for connections to L1. Make sure that the L1 leads do not touch any component as the variable coupling rod is moved in and out.

Install knobs on C1, C2, and the variable coupling rod. Then place the crystal in the detector assembly.

OPERATION

Connect antenna and ground leads to the receiver terminals. Plug in a set of 2000-ohm headphones into the panel phone jack. For easier initial operation of the receiver, temporarily connect a germanium diode (1N34A or equivalent) across the D1 connecting lugs. Make sure that the catwhisker is not touching the crystal at this time.

Push the variable coupling rod half way in and tune C1 and C2 for a station. Adjust the variable coupling rod and retune C1-C2 for best reception. Check operation over the entire broadcast band. It may be necessary to add or subtract turns of wire on L1-L2 for complete coverage. After checking operation with the germanium diode, remove it and try using the catwhisker crystal detector.

List of Materials

C1,C2 - 365 mmf variable capacitor.
D1 - Catwhisker crystal detector assembly.
L1,L2 - Spiderweb coils (see text).
2 - Terminals.
1 - Phone jack.
PHONES - 2000-ohm headphones.
MISC.- Plastic sheet for spiderweb coils, solder lugs, machine screws and nuts, wood screws, hookup wire.
CONSTRUCTION MATERIALS (See Text).
A - knob
B - Panel
B1- Bushing
C - Aluminum section
D - Rod
E - Wood section
F - Setscrew
G - L1 coil form
H - Wood section
I - Spacers (2)
J - Wood section

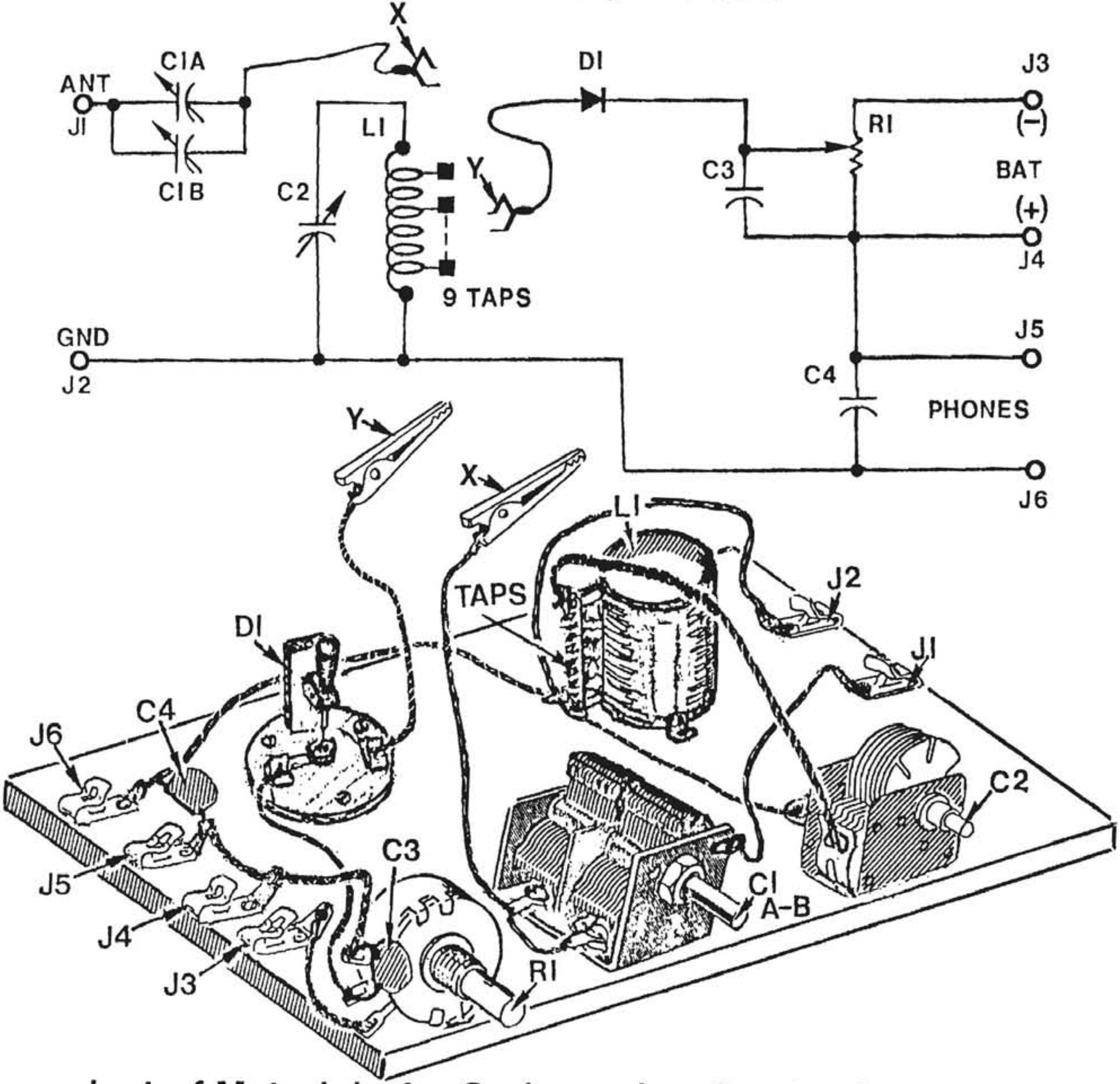
X
C1A
ANT
J1
C1B
C2
L1
Y
9 TAPS
D1
C3
R1
J3
(−)
BAT
(+)
J4
GND
J2
J5
C4
PHONES
J6
Y
X
L1
TAPS
J2
J1
D1
C4
J6
C2
C3
C1
A-B
J5
J4
J3
R1

List of Materials for Carborundum Crystal Set

C1A-C1B - Dual 365 mmf variable capacitor.
C2 - 365 mmf variable capacitor.
C3 - 0.1 mf capacitor.
C4 - 1000 mmf capacitor.
D1 - Carborundum crystal and catwhisker detector assembly (see text).
J1 to J6 - Fahnestock clips.
L1 - Tuning coil (see text).
R1 - 5,000-ohm potentiometer (linear taper).
BAT - 6-volt battery
PHONES - 2000-ohm headphones.
MISC. Coil form (see text), No.24 enameled copper wire, 1/2-inch long spacers, perfboard strip (see text), push-in clips, solder lugs, knobs, hook-up wire, 1N34A diode (or equiv.), wood for base and D1 detector assembly (see text), wood screws, machine screws and nuts, alligator clips X-Y, brackets for L1 and R1.

CARBORUNDUM CRYSTAL SET

Materials other than galena, silicon or germanium have the property of detection. In the beginning of radio, carborundum (the name given to a compound of silicon carbide) was also used as a detector. The required heavy catwhisker pressure made it suitable for the early radio stations on ships, as it did not fall out of adjustment. It would also handle the strong signal energy from nearby spark transmitters without burning out. But what makes this type of crystal detector different, is the necessity for a bias battery.

A crystal detector has a high current flow with voltage applied so that it conducts in the forward direction (catwhisker to crystal), and a very low current flow in the reverse direction. The amount of current flow in the forward direction depends upon the characteristics of the crystal material and the applied forward voltage.

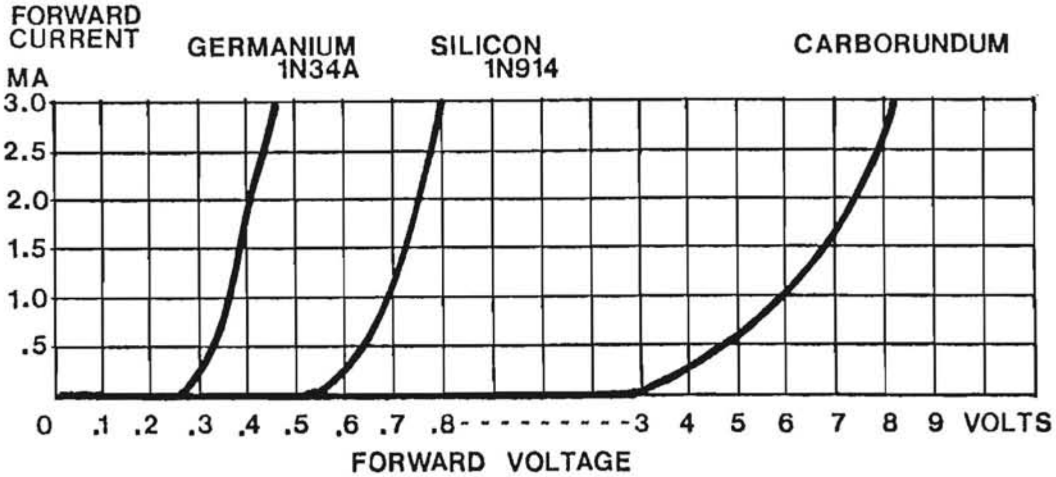

TYPICAL CRYSTAL DIODE CURRENT

The graph compares carborundum with germanium and silicon crystal detectors. Germanium minimum conduction voltage is about 0.3 V. Silicon is 0.6 V, and carborundum is 3 V. The high carborundum minimum conduction voltage is the reason a bias battery is used to move the threshold down so that weak radio signal voltages can be detected.

TUNING COIL

As shown in the illustration, the tuning coil L1 is wound on a cardboard (or plastic) mailing tube section 2-inches in diameter and 2 3/4-inches long.

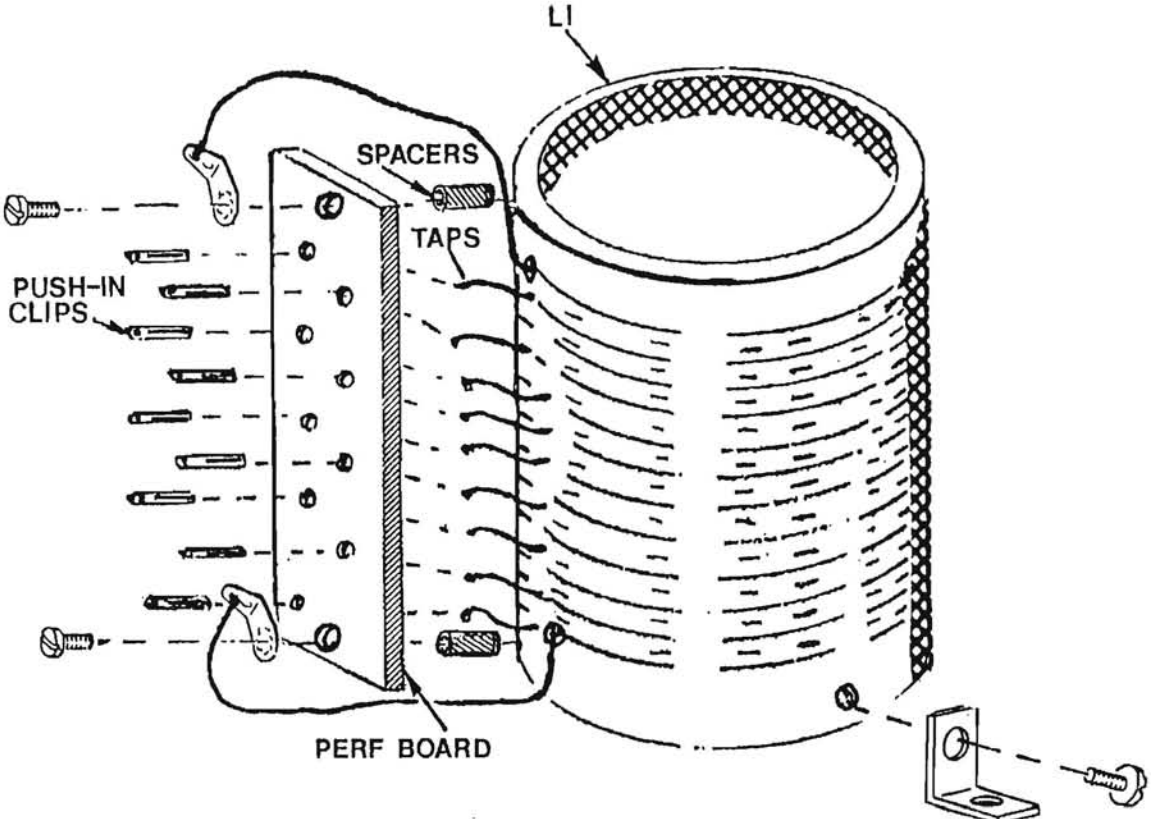

TUNING COIL ASSEMBLY

Locate and drill a hole 1/4-inch from each coil form end for machine screws to fit the two spacers. Also drill holes for two mounting brackets at one end. Then drill a small hole near each spacer hole to feed the wire on the form before you start winding.

Wind the coil with No. 24 enameled copper wire. Tap the coil every 10 turns for a total of 9 taps. An easy way to make the taps is to twist the wire together for a half inch and position the free end out. Make all the coil taps in the area between the two spacer holes. Thread the free end of the wire through the small hole at the end of the form near the spacer hole. Make sure that all of the turns are tightly wound and close together. Use a section of sandpaper to remove the enamel from the tap wire ends. Then tin the ends.

Mount 9 push-in clips in a 1/2-inch by 2 1/4-inch perf board section. Locate and drill holes on the board and install it on the coil form with two solder lugs, 1/2-inch spacers, machine screws, and nuts. Solder the coil taps to the ends of the push-in clips. Connect the coil leads to the solder lugs and mount the two 1/2-inch brackets on the coil form.

DETECTOR

The crystal detector assembly is vertical, (as shown in the illustration) instead of horizontal as in most catwhisker detectors. Begin construction by cutting the 2 1/2-inch diameter base and 2-inch by 4-inch vertical section from 1/4-inch thick wood.

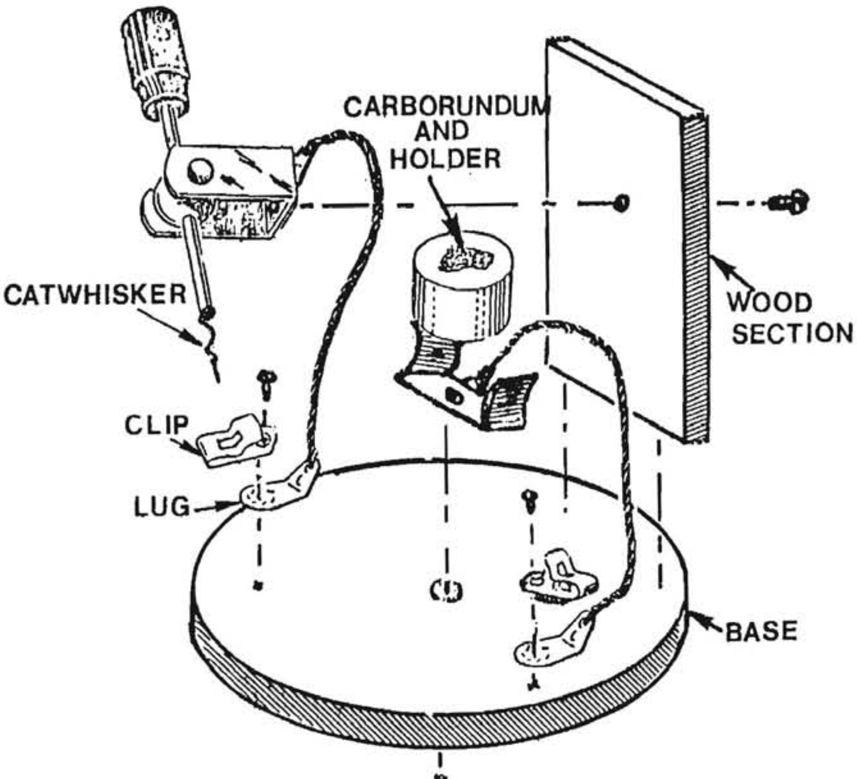

DETECTOR ASSEMBLY

Locate and install the two fahnestock clips on the base with solder lugs and small wood screws. Mount the vertical section on the edge of the base with wood screws (or cement). Locate and install the catwhisker holder with a solder lug, machine screw and nut near the top of the vertical section.

Temporarily position the crystal holder on the base so that the catwhisker will touch the center of the crystal. Locate and mount the crystal holder in this position with a solder lug and machine screw and nut. Connect the detector assembly with hookup wire as shown in the illustration. Solder a short section of coiled steel spring wire to the end of the positioning rod to serve as a catwhisker. Cut the end of the wire diagonally to give it a sharp point.

BASEBOARD

The receiver parts are installed on a 8-inch by 7-inch by 3/4-inch wood base. Wherever poossible, small wood screws are used to hold the components on the base.

Begin construction by laying out and installing the parts in the positions shown in the illustration. It will be necessary to make a metal mounting bracket to fit R1. C1A-B and C2 are mounted with machine screws in the tuning capacitor frames through countersunk holes in the base. Install solder lugs in the rear frame of both tuning capacitors. Also mount solder lugs with all of the terminals. Install knobs on the tuning capacitors and R1.

WIRING

Use hookup wire to connect the crystal set parts as shown in the schematic and illustration. Connect 5-inch leads of flexible stranded insulated wire to each alligator clip. Solder one lead to the stator lugs of C1A-B and connect the other lead to D1. Connect C3 and C4 as shown and keep their leads as short and direct as possible.

Experiment 1

Connect your antenna to J1, ground to J2, and 2000-ohm headphones to J5 and J6. Temporarily connect a germanium diode (1N34A or equiv.) in place of the carborundum crystal detector D1. The diode will make it easier to check out the receiver.

Do not attach the battery
to J3-J4 at this time, or
the diode may be damaged.

Connect **X** C1A-B clip lead to the top tap of L1, and **Y** D1 clip lead to the second tap from the top. Tune in a radio station with C2 and then adjust C1A-B for best volume. Change the position of the clip leads and retune C1A-B and C2 for best reception. Check operation of the receiver over the entire broadcast band.

Experiment 2

Tune the receiver to a radio station and then disconnect the germanium diode. Do not disturb the receiver tuning. Install a carborundum crystal into the D1 assembly. Set R1 to mid range, and connect a 6-volt battery to the receiver; negative lead to J3, positive lead to J4.

Adjust the D1 catwhisker for a sensitive spot on the carborundum crystal. Then set R1 for best volume of received signal. Retune C1A-B and C2 and change the clip leads on the L1 taps as necessary for reception of radio stations.

Note

Carborundum that will work as a detector may be hard to find. Grinding wheels are not suitable as a source, because the carborundum is usually in fine particles mixed with other material. Best bet is to contact collectors of antique radio parts. But you can use this receiver to experiment with minerals for semiconductor operation. See the Workshop Section for information on mounting mineral sections for detector experiments.

4 COMPONENT PROJECTS

BOOK TUNING CAPACITOR

During the roaring 20's,the Crosly Radio Corporation used a type of tuning capacitor that had two plates hinged together on one edge. The dielectric was a sheet of mica between the two plates. This type of tuning capacitor was called a Book Condenser.

The tuning capacitor shown in the illustration is similar to the Crosly Book Condenser. It uses aluminum foil plates cemented to fixed and movable wood sections, and a plastic sheet dielectric (cut from thin transparent plastic notebook protector material).

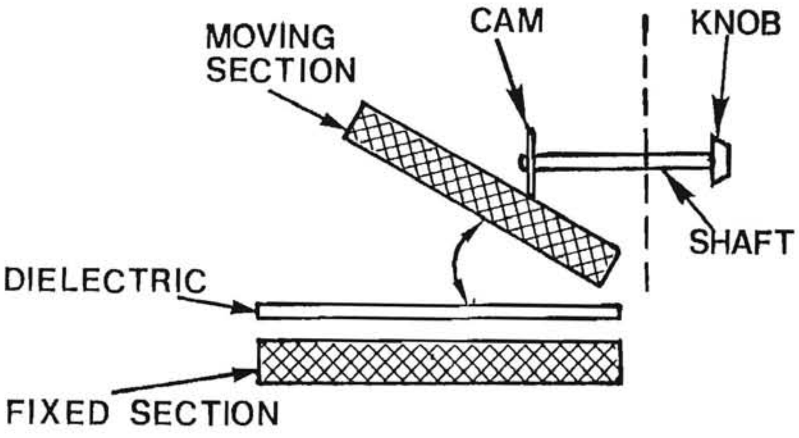

Adjusting the tuning knob causes a metal cam to rotate and control the movement of a spring loaded wood section. This varies the capacity between the two foil plates. As the plates are brought closer together, the capacity increases. As they are separated, the capacity decreases.

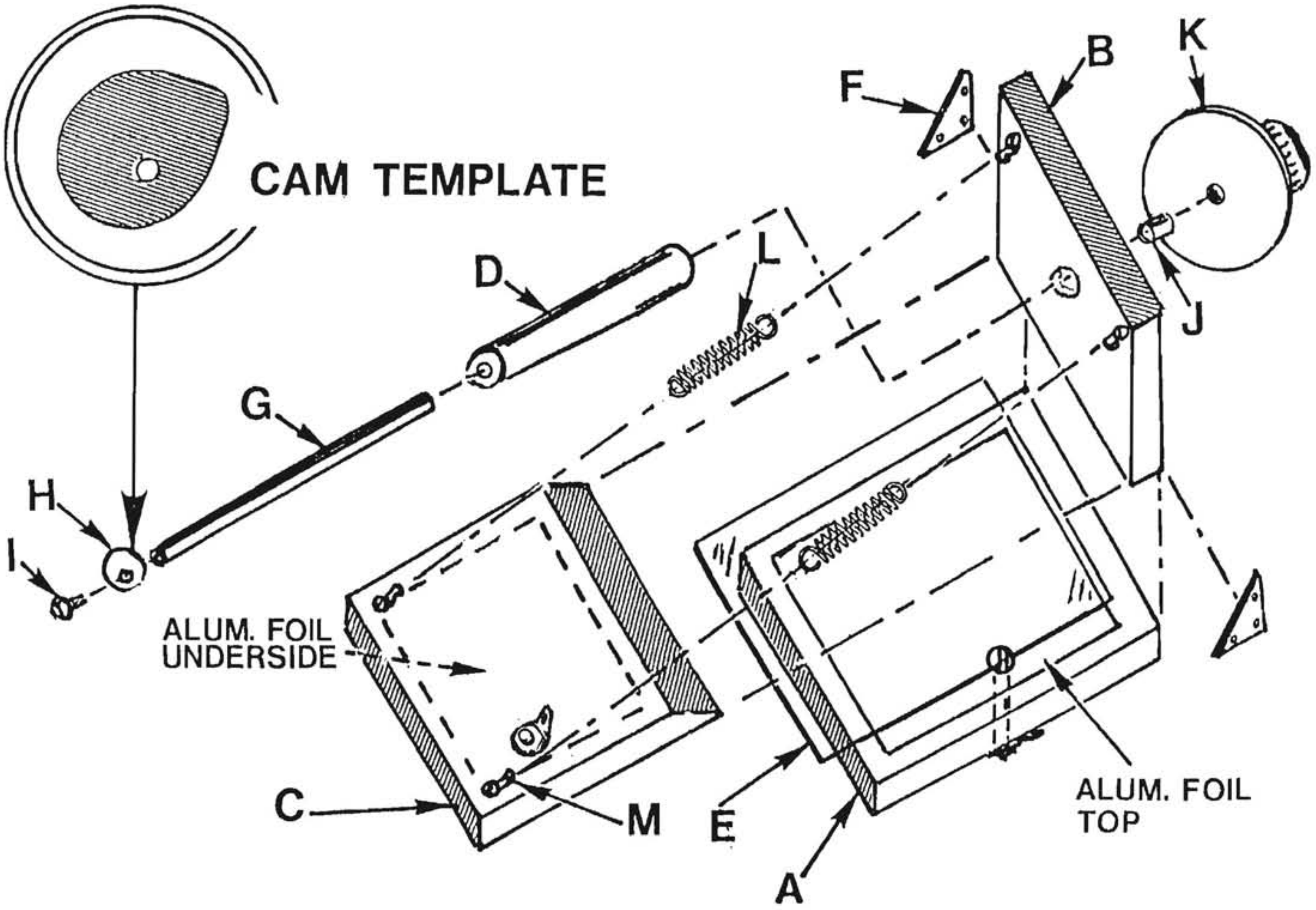

BOOK TUNING CAPACITOR

List of Materials

A - 3 x 3 1/2 X 3/8-inch wood section
B - 3 x 1 3/4 x 3/8-inch wood section
C - 3 x 3 1/8 x 3/8-inch wood section (one end beveled 45 degrees - see text).
D - 3/4 - inch diameter wood dowel, 2 1/4 inches long.
E - 3 1/2 x 3 -inch sheet plastic (See text).
F - 2 triangular 1-inch metal sections
G - 1/4-inch metal rod, 3 -inches long.
H - Metal cam (see text).
I - Machine screw (see text)
J - Bushing 3/8-inch long (to fit rod G).
K - Tuning knob (to fit rod G).
L - 2 springs 1/4-inch diameter by 1 1/4-inch long.
M - 4 solder lugs.
MISC. Small wood screws, nails, aluminum foil, machine screws, nuts, solder lugs and washers.

CONSTRUCTION

Start construction by cutting the wood sections from 3/8-thick wood to the sizes shown in the drawing. Section C has a 45 degree bevel at one end. This acts as a pivot point for the movable section. Drill and countersink a hole to fit a flathead machine screw in the places shown on sections A and C. Do not install the screws at this time.

Cement aluminum foil on the top of section A and on the bottom of section C. Then cut and remove a 1/2-inch strip of foil from the front edge of section A. Cut a 1/8-inch strip from the remaining edges. Also cut away a 1/8-inch strip from all the edges of the foil on section C.

Carefully slit the foil on top of the machine screw hole on section A and press the foil gently inside the hole. Install the flathead machine screw into the hole, making sure that it is in contact with the foil. Fasten the screw with a washer and nut. Make sure that the screw does not turn and rip the foil. Install a solder lug and nut on the free end of the screw. Repeat this operation with section C.

Cut the sheet plastic E to size and place it on top of section A. Then place section B on top of the plastic sheet and install it on the end of section A with wood screws. Install the two metal sections F with small nails.

Saw the metal rod G to size and tap one end to fit a machine screw. Cut the dowel D to size and center drill it to fit the rod G. Cut a flat on one side of the dowel.

Make a tracing of the cam H template and cement it on a section of heavy gauge sheet aluminumm. Cut the cam to size and drill a hole to fit the rod machine screw in the location shown. Install the cam with a lock washer on rod G. Insert the free end of the rod into dowel D. Place section C on top of the plastic sheet E on section A. Then temporarily position dowel D with rod G in the center of section B. Rotate the cam so that its long end touches the top of section C. Mark the placement of the dowel on section B. Drill a hole in section B to fit dowel D. Cement the dowel in place.

Locate and install solder lugs with wood screws on sections A and C as shown in the drawing. Bend over the ends of the solder lugs and install two springs L in the lugs.

Mount tuning knob K and bushing J on the free end of rod G. Adjust the knob and check that section C moves in and out with the cam rotation. changing the capacity between the foil plates.

You may have to experiment with the cam diameter for best operation of section C. File the cam edges in small increments only and recheck the operation often. Use very light springs to keep section C from pressing against the cam and moving it.

Reverse the cam on rod G to alter the direction of capacity (clockwise or counterclockwise). The capacitor can be installed vertically on a receiver project with a small metal bracket mounted on the bottom side of section A.

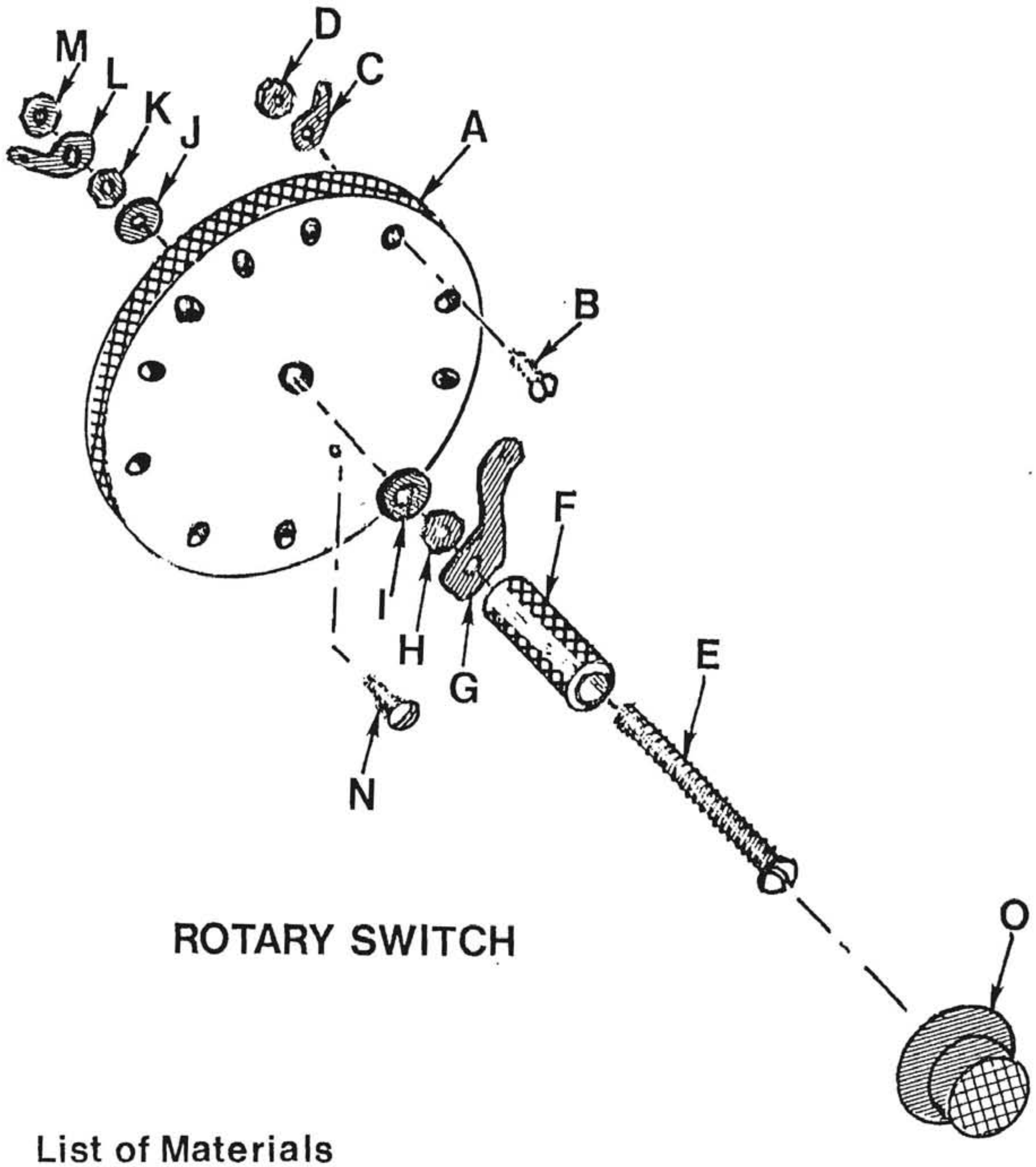

ROTARY SWITCH

List of Materials

A - 2 1/4 diameter x 3/8-inch wood section.
B - Flat head machine screws.
C,L - Solder lugs.
D - Nuts to fit B.
E - Machine screw 1 1/4-inches long.
F - 1/4 diameter x 1/2-inch long bushing (to fit E).
G - Switch arm 1 x 1/4-inch (sheet brass, see text).
H,K,M - Nuts to fit E.
I,J - Washers to fit B.
N - Small woodscrew.

ROTARY SWITCH

In the early days, radio experimenters could not buy parts that are readily available today. They had to build their radio components from wood sections and common hardware.The rotary switch ilustrated is similar to a type constructed for use with tapped coils of crystal sets.

The rotary switch shown uses flat head screws for contact points that are mounted on a wood section. The switch arm is cut from sheet brass and is installed on a machine screw.

CONSTRUCTION

Refer to the drawing of the rotary switch. Begin construction by cutting the 2 1/4-inch diameter section A from 3/8-inch thick wood. Locate, drill,and countersink the holes for the flathead machine screws B. The rotary switch shown has 10 machine screws, but you can space the hole locations for the number of switch contact points desired. Allow a space between the starting and ending contact points for the woodscrew N. This woodscrew is used as a stop for the rotary arm G. Locate and drill a center hole for the machine screw E. Install the screws B in section A with solder lugs C and nuts D. Install the woodscrew N.

Measure and cut the rotary arm G from a section of sheet brass. The arm should be long enough to reach the contacts. Make sure that it is not too wide, or it may short out to a nearby contact. Make a dent or "dimple" on the contact end, and drill a hole for screw E on the other end. Bend the rotary arm as shown in the drawing. Install screw E into bushing F, and arm G, and tighten the assembly with nut H. Place washer I on the screw and install screw E into the center hole of A. loosely fasten the screw with washer J and nut K. Then rotate the bushing F and arm G. Adjust the tightness of nut K as required to allow rotation of G with only a slight "drag". Install solder lug L on screw E with nut M, and bend the lug towards the screw end. Install knob O on bushing F, and recheck operation of the arm G. It should touch all of the contacts B and stop at screw N.

The rotary switch is used as shown in the Dual-Tuned Crystal Set Construction Project. Make sure that extra lead length is allowed for connection to the solder lug L. Bend the lead into a "pig-tail," so that it will not break as the solder lug L is rotated.

5

WORKSHOP

The construction projects in this book can be built with common hand tools found in an average home workshop.

Useful Hand Tools

Safety glasses or face shield
Steel rule
Square
Hack saw
Wood saw
Screwdrivers
Pliers
Wirecutters
Tin snips
Center punch
Files
Hammer
Soldering Iron

Wear safety glasses or face shield while working. Observe common sense safety precautions.

SOLDERING

Make sure that the part to be soldered is clean. Use only rosin core solder. Prepare the soldering iron tip as follows:

File each surface of the tip to remove oxidation. Plug in the iron and let it heat until you can melt solder on the tip (before the copper tip turns blue or bronze color). Melt enough solder until all sides of the tip are coated. Then wipe off excess solder with a damp sponge or cloth.

Make sure that the bare wire lead to be soldered is free of oxide. It is best to tin it first by melting a bit of rosin core solder on the lead end. Crimp the lead in the part hole to keep it from moving during soldering. Hold the iron tip under the part until it is hot enough to melt solder. Apply only enough rosin core solder to cover the surface of the part. Remove the iron. Do not move the part or wire while it is cooling.

PROJECT CONSTRUCTION TIPS

Tape a sheet of graph paper over a project wood section. Mark locations of the parts on the paper. Center punch mounting hole centers through the paper. Use a compass to draw circles the same size as the holes. Then drill holes through the paper.

Lettering on the projects can be done with a labelmaker that prints on strips of adhesive coated plastic. You can also use decals or dry transfer lettering sheets.

Small metal brackets can be made from heavy gauge sheet aluminum. Layout and drill the mounting holes and then cut it to size with a tin snips. Form the bracket by using a heavy mallet and a vise.

Tuning Capacitors, Knobs, and Dials

Old radios are a good source of components. The radios generally use two tuning capacitor sections ganged together. The receivers previously described will operate with one of the capacitor sections (365 mmf). You can cut out cardboard discs and cement them to radio knobs and use them as dials with the tuning capacitors. Mark the dials with dry transfer lettering, or black drawing ink.

Panels and Cabinets

The layout and construction of crystal sets is not critical. They can be fitted into many different types of wood cabinets. You can use almost any type of non-metallic box, but metal boxes are not recommended as sheet metal close to the tuning coil may lower its "Q". Black sheet plastic used for the front panel with white dry transfer lettering will give your receiver project an antique look.

MOUNTING MINERALS FOR CRYSTAL DETECTORS

Minerals (galena, etc) for experimentation as crystal detectors, will work best when mounted in a metallic base. For temporary mounting (while testing), wrap aluminum foil tightly around the sides and bottom of the mineral. Then place it into a sheet metal ring or cup that will fit into the detector assembly.

After testing, the mineral can be permanently mounted in a metal slug. Drill a shallow hole of a diameter equal to the detector assembly crystal in a block of wood. Fill the hole with melted solder. While the solder is still molten, carefully place your mineral section into the center. Hold the mineral in place with tweezers or long nose pliers until the solder cools. Then pry the metal slug out of the wood. Be sure that the top of the mineral section is free of solder.

INDEX

Advanced receiver projects,32
Amplifier connection,14,15
Antenna lead-in,10
Antenna,10

Book tuning capacitor,48
Boundary,5
Bias,5

Capacitor,8
Capacity,8
Carborundum crystal set,43
Carborundum,43
Catwhisker detector,4,6
Catwhisker,6
Component projects,48
Construction tips,55
Crystal detector assembly,45
Crystal detector,4

Detected audio signals,7
Detection,4,6
Diode electron flow,5
Dual spiderweb coils,37
Dual-tuned crystal set,32

Earphones,11
Enamel insulation,17
Electron current flow,5

Ferrite coil crystal set,23
Ferrite coil,12
Ferrite rod,23

Galena,4
Germanium,4
Ground lead,10
Ground wave,25
Ground,10

Hand tools,54
Headphones,7,11
Headset,11

Junction diode,4,5

Knobs and dials,55

Lettering,55
Loading coil,21
Loopstick,12

Modulated radio waves,6
Mounting minerals,56

N type material,4
Natural crystal,4

Operating tips,25

P type material,4
PN junction,5
Panels and cabinets,5
Parallel tuned circuit,9
Point contact diode,4,6
Primary coil,23

Reactance,8
Receiving radio waves,7
Resonant frequency,9
Rf coil,8
Rotary switch project,53

Schematic symbols,15
Secondary coil,23
Selective crystal set,37
Selectivity,9
Semiconductor diode,4
Silicon,4
Simple crystal set,12
Simple receivers,12
Sky wave,25
Slider tuned coil,17
Slider tuned set,17
Soldering,54
Space charge region,5
Spiderweb coil set,27
Station log,31

Tapped coil,32,44
Tapped spiderweb,27
Transmitted waves,6,7
Transmitting station,6
Tuned circuit,7,9
Tuning crystal set,7

Untuned detector,20

Variable capacitor,8
Variable coupling,39

Wave trap,21
Workshop,54

About the Author

Charles Green has been involved in the design and construction of electronic equipment for industry and the U.S. Government for 30 years. Much of his experience has been as a technical writer preparing operation and maintenance manuals. His continuing interest is in constructing radios and he has had over a hundred articles published in hobbyist magazines over the past two decades.

Building an AM radio

Objectives

By the end of this lab, the students should be able to:

- describe the function of the variable resistor, variable capacitor in the radio circuit
- describe how variable capacitors are connected in circuits
- describe how variable resistors are connected in circuits
- describe the “hidden workings” of a electronics prototyping board
- construct a AM radio (or any circuit) using a project board, components, jumper wires and a diagram

Introduction

Amplitude Modulated (AM) radio signals carry an audio signal (electrical information that can be converted to sound). The audio information is encoded on the radio wave by changing the amplitude of a high frequency radio wave. The base frequency of the electromagnetic radio wave (called the carrier wave) is much higher than the frequency of the audio information it carries (550kHz to 1600kHz versus 20Hz to 20kHz). The diagram at right shows how the encoding works. The amplitude of the carrier wave is not constant. The fast variation is the carrier wave. The relatively slow variations in the amplitude create an envelope that shows the frequencies in the audio signal. The diagram is somewhat distorted to show the principle. The amplitude of a typical radio signal varies much more slowly than the picture suggests. (How many periods of a 600kHz carrier would have to be shown in order to show half of a period of a 100 Hz sound?)

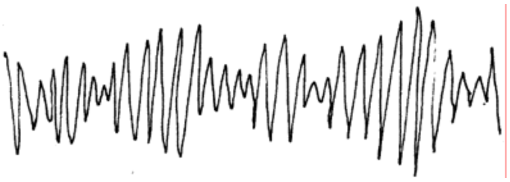

Figure 1: Amplitude modulated carrier signal

An extremely basic AM radio, consisting of a detector (a diode in the circuit at the right) and an antenna, is shown at right. The diode only allows current to flow in the direction indicated by its symbol. This circuit detects AM signals and converts them to audio signals.

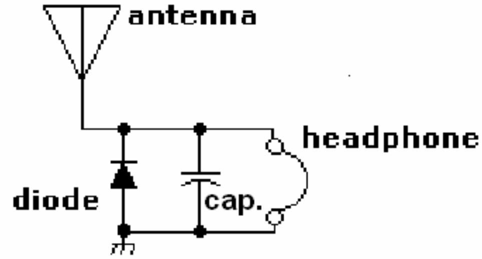

Figure 2: Simple AM detector circuit

The second circuit at the right is a (slightly) more sophisticated version. The design includes a tuning section. By setting the resonant frequency of the circuit to the frequency of the desired carrier wave, the user selects one carrier frequency for conversion.

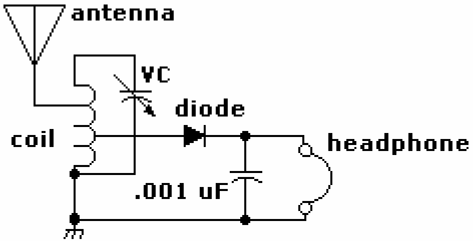

Figure 3: AM detector circuit with tuner block

The radio you will build (see the diagrams on the last page of the handout) includes an AM detector (a transistor in this case), a low pass filter (to remove the carrier wave from the audio signal) and an amplifier.

Web Resources

The web has some good resources on AM radios. Googling “crystal radio” yields quite a bit of info on how to build a radio. One such site is http://sci-toys.com/scitoys/scitoys/radio/radio.html. The web also has some good explanations of how the AM detector works. Google “AM detector” or “envelope detector”. One such resource is http://www.st-andrews.ac.uk/~www_pa/Scots_Guide/RadCom/part9/page2.html.

Building an AM radio

Instructions

1. Some of the AM radio is already made. A battery holder, coil, variable capacitor, potentiometer (a.k.a. variable resistor), project board and earphone jack have been mounted to a small piece of plywood.
    - Identify which component is which.
    - Use a multimeter to figure out how the potentiometer works. (Hint: Use the ohm-meter setting on the multimeter).
2. Prototype board is used for making and testing temporary circuits. Prototype boards have internal connections. The internal connections allow you to build a circuit with relatively few wires needed and to avoid soldering). Obtain an un-mounted project board. Inspect it. Figure out the pattern of internal connections. You can do this by looking at the back of an un-mounted project board or by using an ohm-meter.
    - Which holes will be at the same potential (provided small currents are flowing)?
    - Which holes can be held at different potentials? In other words, which sets of holes are electrically isolated from each other?
3. Obtain a kit of components for your radio. Inspect them.
    - Identify the different kinds of components: resistors, capacitors, transistors, and integrated chip.
    - *Hooking up asymmetric components incorrectly will lead to unintended operation* (more often than not unintended operation means *self-destruction*). Identify all non-symmetric components. Note: Some “two-legged” components, including many capacitors, are asymmetric.
4. Use the diagrams on the next page to construct the circuit. You may note slight discrepancies between the parts in your kit and the parts provided. In particular, the resistance values called for might not match the contents of your kit. In general, you can use resistance values that are within a factor of two or so and still get the circuit to work. *Pay attention to the proper connection of asymmetric components* (including capacitors)! *Leave the battery out* until you have visually double-checked the circuit against the diagram.
5. Put in the battery and test the radio. In order to get a signal, you may need to …
    - tune the radio. (Which circuit element does this?)
    - turn up the volume on the earphone (Which circuit element does this?)
    - rotate the entire radio. (Why might the orientation matter? Which element needs to be aligned? What does it need to be aligned with? Why?)

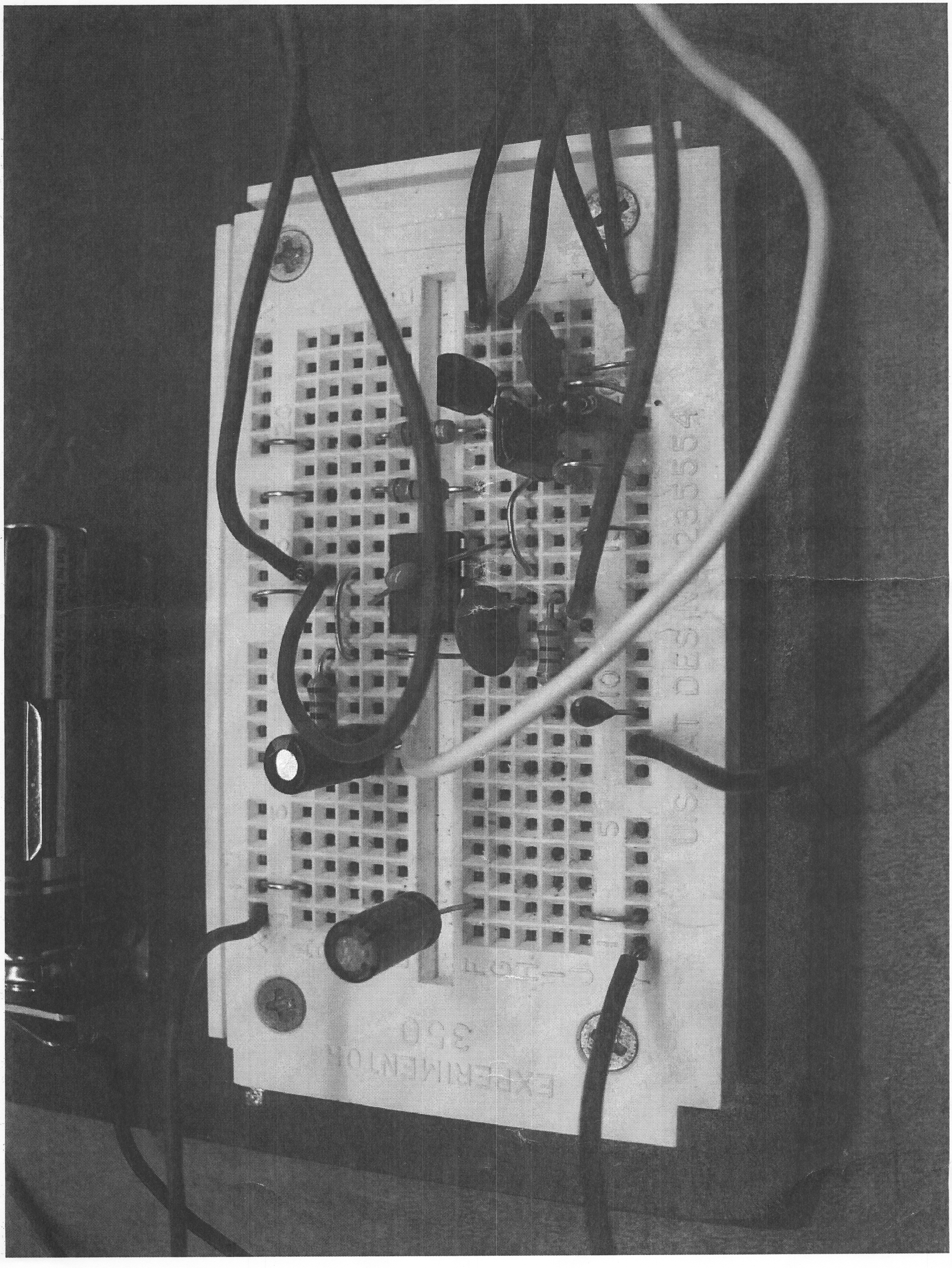
EXPERIMENTOR 350

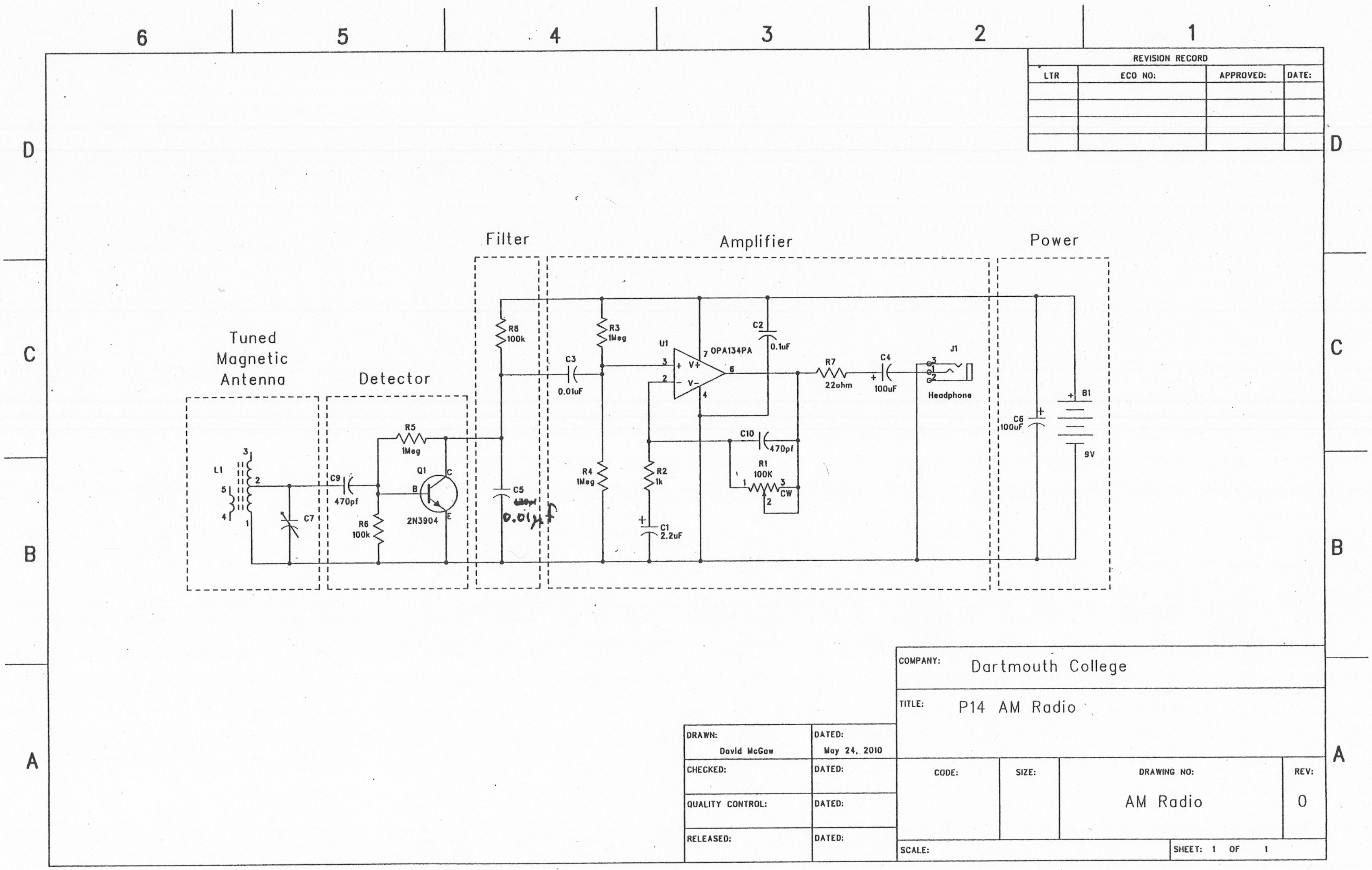
6
5
4
3
2
1
D
C
B
A
Tuned
Magnetic
Antenna
Detector
Filter
Amplifier
Power
L1
3
2
5
4
1
C7
C9
470pf
R5
1Meg
Q1
B
C
E
2N3904
R6
100k
R8
100k
C5
0.01μf
C3
0.01uF
R3
1Meg
R4
1Meg
U1
OPA134PA
7
3
2
4
6
V+
V-
+
-
C2
0.1uF
R2
1k
C1
2.2uF
C10
470pf
R1
100K
1
2
3
CW
R7
22ohm
C4
100uF
J1
Headphone
C6
100uF
B1
9V

REVISION RECORD

LTR	ECO NO:	APPROVED:	DATE:

DRAWN: David McGaw	DATED: May 24, 2010
CHECKED:	DATED:
QUALITY CONTROL:	DATED:
RELEASED:	DATED:

COMPANY: Dartmouth College

TITLE: P14 AM Radio

CODE:	SIZE:	DRAWING NO:	REV:
		AM Radio	0

SCALE:

SHEET: 1 OF 1

Radio and Electronics Cookbook

Radio and Electronics Cookbook

Edited by
Dr George Brown, CEng, FIEE, M5ACN

Newnes
OXFORD AUCKLAND BOSTON JOHANNESBURG MELBOURNE NEW DELHI

Newnes
An imprint of Butterworth-Heinemann
Linacre House, Jordan Hill, Oxford OX2 8DP
225 Wildwood Avenue, Woburn, MA 01801-2041
A division of Reed Educational and Professional Publishing Ltd

A member of the Reed Elsevier plc group

First published 2001

© Radio Society of Great Britain 2001

All rights reserved. No part of this publication may be reproduced in any material form (including photocopying or storing in any medium by electronic means and whether or not transiently or incidentally to some other use of this publication) without the written permission of the copyright holder except in accordance with the provisions of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London, England W1P 0LP. Applications for the copyright holder's written permission to reproduce any part of this publication should be addressed to the publishers

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN 0 7506 5214 4

RSGB
Lambda House
Cranborne Road
Potters Bar
Herts
EN6 3JE

Composition by Genesis Typesetting, Laser Quay, Rochester, Kent
Printed and bound in Great Britain

PLANT A TREE
BTCV
British Trust for Conservation Volunteers
FOR EVERY TITLE THAT WE PUBLISH, BUTTERWORTH-HEINEMANN WILL PAY FOR BTCV TO PLANT AND CARE FOR A TREE.

Contents

Preface ix

1. A medium-wave receiver 1
2. An audio-frequency amplifier 4
3. A medium-wave receiver using a ferrite-rod aerial 9
4. A simple electronic organ 12
5. Experiments with the NE555 timer 17
6. A simple metronome 21
7. What is a resistor? 24
8. Waves – Part 1 27
9. A beat-frequency oscillator 31
10. What is a capacitor? 34
11. Waves – Part 2 38
12. An LED flasher 41
13. Waves – Part 3 44
14. Choosing a switch 46
15. An aerial tuning unit for a receiver 49
16. A simple 2 m receiver preamplifier 52
17. Receiving aerials for amateur radio 54
18. The Colt 80 m receiver – Part 1 58
19. A crystal radio receiver 62
20. The varactor (or varicap) diode 64
21. A portable radio for medium waves 65
22. The Colt 80 m receiver – Part 2 70

23. A simple transistor tester 73

24. An introduction to transmitters 77

25. The Colt 80 m receiver – Part 3 81

26. A two-way Morse practice system 88

27. The Colt 80 m receiver – Part 4 91

28. A simple crystal set 95

29. A crystal calibrator 100

30. A simple short-wave receiver – Part 1 104

31. A fruit-powered medium-wave radio 106

32. A capacitance bridge 109

33. A simple short-wave receiver – Part 2 113

34. A basic continuity tester 117

35. A charger for NiCad batteries 119

36. An 80 metre crystal-controlled CW transmitter 123

37. A solar-powered MW radio 129

38. A receiver for the 7 MHz amateur band 133

39. Diodes for protection 137

40. An RF signal probe 140

41. An RF changeover circuit 142

42. A low-light indicator 146

43. A J-pole aerial for 50 MHz 149

44. Measuring light intensity – the photometer 153

45. A 70 cm Quad loop aerial 156

46. A UHF field strength meter 160

47. Christmas tree LEDs 162

48. An audio signal injector 166

49. Standing waves 168

50. A standing-wave indicator for HF 170

51. A moisture meter 174

52. Simple aerials 177

53. A breadboard 80 cm CW transmitter 182

54. A 7-element low-pass filter for transmitters 186
55. Radio-frequency mixing explained 189
56. A voltage monitor for a 12 V power supply 192
57. A 1750 Hz toneburst for repeater access 196
58. A circuit for flashing LEDs 201
59. Digital logic circuits 205
60. A resistive SWR indicator 210
61. An audio filter for CW 213
62. An electronic die 215
63. The absorption wavemeter 222
64. An HF absorption wavemeter 224
65. A vertical aerial for 70 cm 228
66. A UHF corner reflector aerial 230
67. A switched dummy load 234
68. A simple Morse oscillator 238
69. A bipolar transistor tester 240
70. The ‘Yearling’ 20 m receiver 245
71. Adding the 80 metre band to the Yearling receiver 251
72. How the Yearling works 255
73. A field strength meter 258
74. Preselector for a short-wave receiver 261
75. An audible continuity tester 265
76. An experimental 70 cm rhombic aerial 268
77. Water level alarm 272
78. A delta loop for 20 metres 275
79. A simple desk microphone 279
80. Morse oscillator 284
81. A simple 6 m beam 287
82. An integrated circuit amplifier 291
83. A novice ATU 293
84. CW QRP transmitter for 80 metres 297

85. An audio booster for your hand-held 303
86. A grid dip oscillator 306
87. A CW transmitter for 160 to 20 metres 312
88. Matching the end-fed random-wire aerial 315

Preface

Although we are surrounded by sophisticated computerised gadgets these days, there is still a fascination in putting together a few resistors, capacitors and the odd transistor to make a simple electronic circuit. It is really surprising how a handful of components can perform a useful function, and the satisfaction of having built it yourself is incalcuable.

This book aims to provide a wide variety of radio and electronic projects, from something that will take a few minutes to a more ambitious weekend's worth. Various construction techniques are described, the simplest requiring no more than a small screwdriver, the most complex involving printed circuit boards.

Originally published by the Radio Society of Great Britain, the projects were all chosen to be useful and straightforward, with the emphasis on practicality. In most cases the workings of the circuit are described, and the projects are backed up by small tutorials on the components and concepts employed. In the 21st century it may seem strange that few of the published circuits use integrated circuits (chips). This is intentional as it is much easier to understand how the circuit works when using discrete components.

Anyone buying the *Radio and Electronics Cookbook* will find that it will lead to hours of enjoyment, some very useful and entertaining gadgets, and increased knowledge of how and why electronics circuits work, and a great sense of satisfaction. Beware, electronic construction is addictive!

WARNING: This book contains construction details of transmitters. It is illegal to operate a transmitter without the appropriate licence. Information on how to obtain an Amateur Radio Licence can be obtained from the Radiocommunications Agency, tel. 020 7211 0160.

1 A medium-wave receiver

Introduction

Let us start off with something that is really quite simple and yet is capable of producing a sense of real satisfaction when complete – a real medium-wave (MW) radio receiver! It proves that receivers *can* be simple and, at the same time, be useful and enjoyable to make. To minimise the confusion to absolute beginners, *no* circuit diagram is given, only the constructional details. The circuits will come later, when you have become accustomed to the building process. In the true amateur spirit of ingenuity and inventiveness, the circuit is built on a terminal strip, the coil is wound on a toilet roll tube (as amateur MW coils have been for 100 years!), and the receiver is mounted on a piece of wood.

Putting it together

Start by mounting the components on the terminal strip as shown in **Figure 1**, carefully checking the position and value of each one. The three capacitors are all the same, and so present no problem. They (and the resistors) may be connected either way round, unlike the two semiconductors (see later). The resistors are coded by means of coloured bands. You can refer to Chapter 7 if you have difficulty remembering the colours and their values.

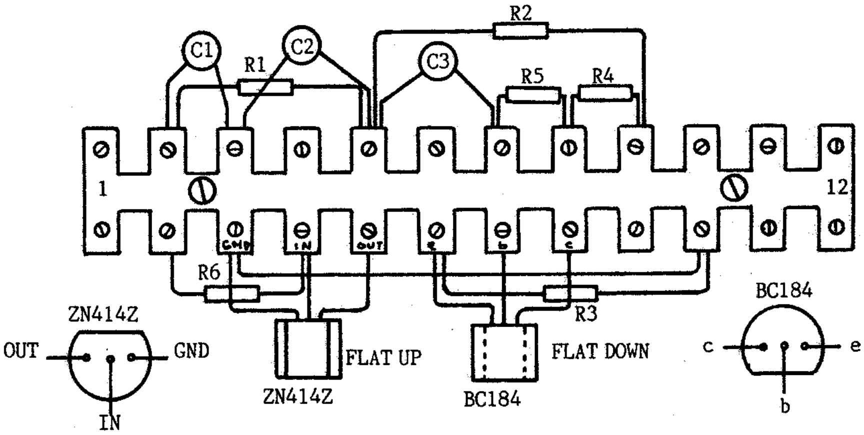

Figure 1 Terminal strip – position of components

1.	Brown, Black, Yellow	100 000 ohms	(R1, R5, R6)
2.	Green, Blue, Brown	560 ohms	(R2)
3.	Red, Violet, Brown	270 ohms	(R3)
4.	Brown, Black, Orange	10 000 ohms	(R4)

The integrated circuit (the ZN414Z) and the transistor (the BC184) *must* be connected correctly. Check Figure 1 carefully before fitting each device.

Now wind the coil. Most tubes are about 42 mm diameter and 110 mm long. Don't worry if your tube is slightly different; it shouldn't matter. Make two holes, about 3 mm apart, about 40 mm from one end, as shown in **Figure 2**. Loop your enamelled wire into one hole and out of the other, and draw about 100 mm through; loop this 100 mm through again, thus anchoring the wire firmly. Now wind on 80 turns, keeping the wire tight and the turns close together but not overlapping. After your 80th turn, make another two holes and anchor the wire in the same way as before. Again, leave about 100 mm free after anchoring. Using another piece of enamelled wire (with 100 mm ends as before), loop one end through the same two holes which contain the end anchor of the last winding, wind two turns and anchor the end of this short winding using another pair of holes. Figure 2 shows the layout.

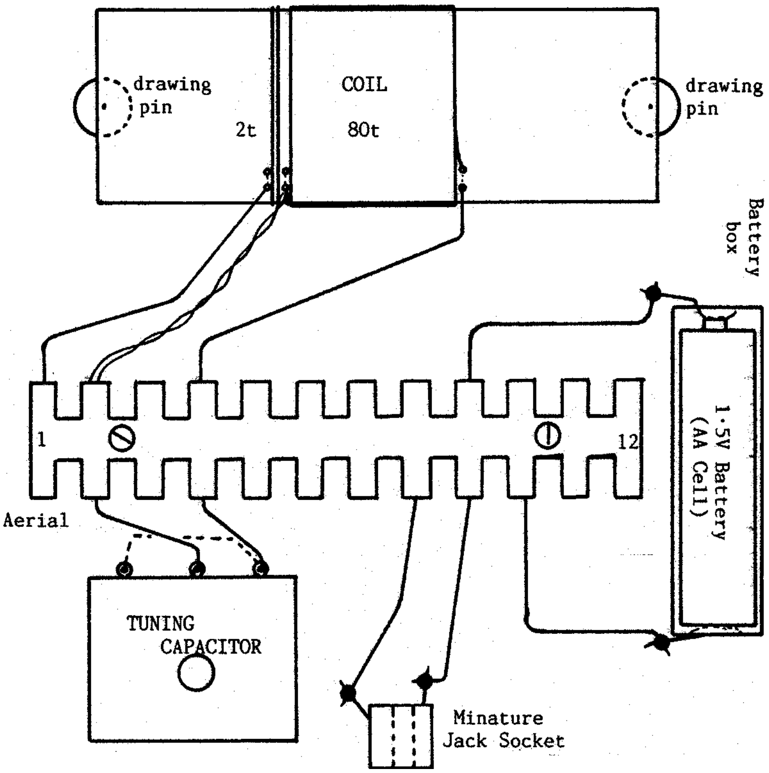

Figure 2 The layout of the parts on the wooden base

With some glass paper, remove the enamel from the ends of both pieces of wire which go through the same holes (i.e. the bottom of the large coil and the top of the small coil), then twist these bare ends together. Remove the enamel from the remaining ends of the coil. The coil is now finished!

The baseboard can be any piece of wood about 150 mm square. Fix the coil near the back edge using drawing pins and connect the wires from the coils to the terminal strip as shown in Figure 2. Using short pieces of PVC-insulated wire (and with assistance if you have never soldered before), solder one piece across the two outer tags of the variable capacitor, shown by the dotted line in Figure 2, and then two longer pieces to the centre tag and one outside tag. Connect these to the terminal strip. Then solder two more insulated wires on to the jack socket (into which you will plug your crystal earpiece), the other ends going to the terminal strip. The last two wires (one must be red) need to be soldered on to the battery box, their other ends going to the terminal strip also. Make sure the red wire goes to the *positive* terminal on the battery, and is connected to terminal 9. The other connection to the battery goes to terminal 10.

Attach the terminal strip to the baseboard with small screws or double-sided sticky tape. The other parts can be mounted the same way.

Listening is done ideally with the recommended crystal earpiece. Don't be tempted to use your Walkman earpieces; they are not the same and will not perform anything like as well. The receiver should work without an extra aerial, but one can be attached to terminal 1 if necessary. A long piece of wire mounted as high as possible is ideal. The *Audio-frequency Amplifier* project will enable you to use a loudspeaker with your receiver, using the signal from the jack socket. No circuit modifications will be needed!

Parts list

Resistors: all 0.25 watt, 5% tolerance	
R1, R5, R6	10 kilohms (kΩ)
R2	560 ohms
R3	270 ohms
R4	10 kilohms (kΩ)
Capacitors	
C1, C2, C3	100 nanofarads (nF)
	500 picofarads (pF)
Semiconductors	
ZN414Z, BC184	

Additional items
- 12-way 2 A terminal strip
- 22 metres of 28 SWG enamelled copper wire
- A few short pieces of coloured PVC-insulated wire
- Crystal earpiece
- 3.5 mm jack socket
- 1.5 V AA-size battery and box
- Toilet roll tube
- Double-sided sticky tape or selection of screws

Tools required

Small screwdriver, soldering iron.

2 An audio-frequency amplifier

Introduction

This simple amplifier can be built by anyone who is able to solder reasonably well. It doesn't require any setting up and, provided our instructions are followed exactly, will work very well. The circuit diagram is included for the benefit of our more advanced readers, but it is not needed in the construction process. Please practise your soldering before you start, and don't use a printed circuit board (PCB) until you are confident that your soldering is up to scratch.

The amplifier can be used with other projects; it will provide plenty of sound from the MW Radio or from the Morse Sounder projects. It will usually be built into other pieces of equipment, so a box is not supplied with the kit. There is no reason why it shouldn't be put into a box and used as a general-purpose amplifier to help test other projects.

The components

Before you start, you should check that you have *all* the components to hand. A list and some helpful hints are given below.

1. PCB. The plain side is the *component side* and the soldered side is the *track side*. **Figure 1** shows the track side full size. Make the PCB from the pattern given in Figure 1. Otherwise, build the circuit on a matrix board.
2. Three resistors. Locate the gold or silver band around the resistor, and turn the resistor until this band is to the right. There are three coloured bands at the left-hand end of the resistor. Find the resistor whose colours are YELLOW, VIOLET, RED, and look at the resistor colour code chart which you will find in Chapter 7. From this, you will see that YELLOW indicates the value 4, VIOLET the value 7, and RED the value 2. The first two colours represent real numbers, and the last value is the number of zeros (noughts) which go *after* the two numbers. So, the value is 47 with two zeros, i.e. 4700 ohms. In this way, the resistor coloured BROWN, GREY, BROWN has a value of 180 ohms, and the last one, BROWN, RED, GREEN, has a value of 1 200 000 ohms. The ohm (often written as the Greek letter omega (Ω)) is the unit of resistance. If you do not yet feel confident in identifying resistors by their colours, use the Resistor Colour Codes given in Chapter 7.
3. Four capacitors. The two small 'beads' are tantalum capacitors and will be marked 4.7 μF or 4μ7, with a '+' above one lead. A tubular capacitor with wires coming from each end should be marked 220 μF, with one end marked '+' or '–'. This is called an *axial* capacitor because the wires lie on the axis of the cylinder. This is in contrast to the final capacitor, where both wires emerge from the same end. This is a *radial* capacitor, and will be marked 47 μF. Again, one lead will be marked '+' or '–'. Capacitors marked like this are said to be *polarised*, and it is vital that these are placed on the PCB the right way round, so take notice of those signs!
4. Two diodes. These are tiny glass cylinders with a band around one end, and may be marked 1N4148; this is their type number. Like polarised capacitors, they *must* be put on the PCB the correct way round!

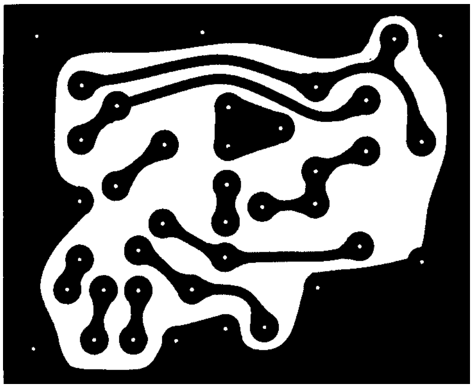

Figure 1 The toil pattern of the PCB – looking from the track side

5. Three transistors. One should be a BC548 (or a BC182), the other two should be BC558 (or BC212).
6. One volume control with internal switch.
7. One loudspeaker. This is quite fragile – don't let anything press against the cone.
8. One PP3 battery clip with red and black leads.

Putting it together

Lay the PCB on a flat, clean surface with the track side downwards. It is always useful to compare the layout with the circuit diagram, given here in **Figure 3**. Although you can't see it, the D-i-Y Radio sign should be at the top. Compare the hole positions with those shown in **Figure 2**. Bend the resistor wires at right angles to their bodies so that they fit cleanly into the holes in the PCB. Push each resistor towards the board so that it lies flat on the board. Then supporting each one, turn the board over and splay out the wires *just* enough to prevent the resistor falling out. Then, solder each wire to its pad on the PCB, and cut off the excess wire. When you have more confidence, you can cut of the excess wire *before* soldering; it often makes a tidier joint.

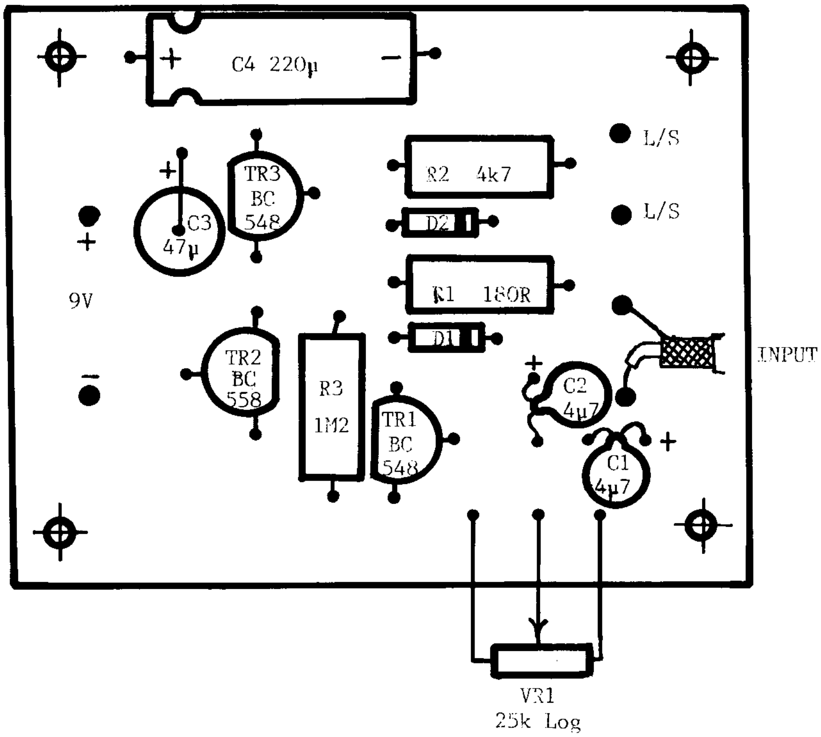

Figure 2 Positions of the components on the printed circuit board (PCB)

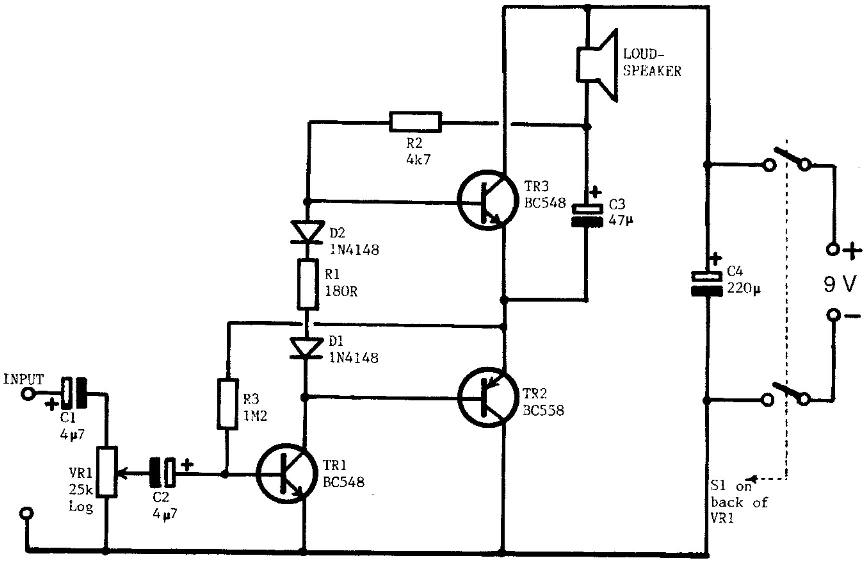
LOUD-
SPEAKER
R2
4k7
TR3
BC548
C3
47µ
D2
1N4148
R1
180R
C4
220µ
+
9 V
−
D1
1N4148
INPUT
C1
4µ7
R3
1M2
TR2
BC558
VR1
25k
Log
C2
4µ7
TR1
BC548
S1 on
back of
VR1

Figure 3 The amplifier's circuit diagram

Now fit the four capacitors. Each must be connected the right way round, so look at each component, match it up with the diagram of Figure 2, bend its wires carefully and repeat the soldering process you performed with the resistors, making sure that the components are close to the board and not up on stilts! Fit the two diodes the correct way round, and solder then as quickly as you can – they don't like to be fried!

Mount the transistors about 5 mm above the PCB. Make sure the correct transistors are in the correct places, and that the flats on the bodies match up with those shown in Figure 2.

Mount the volume control so that the spindle comes out from the front of the board. Use a piece of red insulated wire to the pad marked + on the PCB, and a black piece to the pad marked –, and solder these to the tags on the back of the control, as shown in **Figure 4**. Connect the two leads from the battery clip to the other tags on the switch; Figure 4 will help you. Finally, use two pieces of insulated wire about 100 mm long, twisted together, to connect the loudspeaker to the PCB.

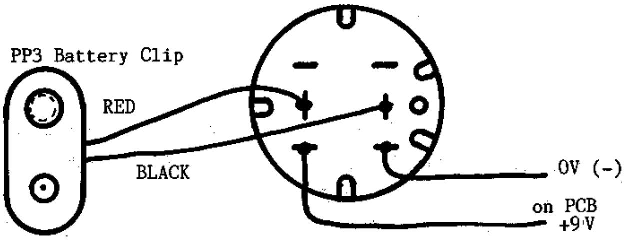
PP3 Battery Clip
RED
BLACK
0V (-)
on PCB
+9V

Figure 4 Connections to switch on back of VR1

Box clever!

If you wish to put the amplifier into a box, there is no problem; almost any box that is big enough will do. All that is needed is one hole big enough to accept the bush of the volume control; the PCB will be supported by the volume control. The prototype was not fitted into a box, but mounted on an odd piece of aluminium, bent into an L-shape and screwed on to a wooden base. The loudspeaker was mounted on the aluminium panel by two small pieces of aluminium with 3 mm holes drilled in them, which acted as clips around the edge of the speaker. Drill a few holes in the panel in the position of the speaker to let the sound get out!

Your input signal can be connected to the amplifier with two short pieces of wire, but if the connection needs to be long, use screened cable, with the braid connected as shown in Figure 2.

If you decide to use a different loudspeaker, make sure that its impedance (the resistance value marked on the back of the magnet) is at least 35 ohms. Anything lower may damage TR2 and TR3, and will certainly run down your battery very quickly. You will be surprised at the uses you can find for this little amplifier!

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1	180 ohms (Ω)
R2	4.7 kilohms (kΩ)
R3	1.2 megohms (MΩ)
VR1	25 kilohms (kΩ) log with DPST switch

Capacitors: all rated at 25 V minimum

C1, C2	4.7 microfarads (μF)
C3	47 microfarads (μF)
C4	220 microfarads (μF)

Semiconductors

TR1, TR3	BC548 npn
TR2	BC558 pnp
D1, D2	1N4148

Additional items
- PCB
- Speaker >35 ohms
- PP3 battery clip and battery

3 A medium-wave receiver using a ferrite-rod aerial

Introduction

This design came from the Norfolk Amateur Radio Club, and enables you to build a simple Amplitude Modulation (AM) receiver for frequencies between 600 kHz and 1600 kHz. It should take you around 2 hours to build, and can be used with Walkman-type earpieces. **Figure 1** shows the circuit diagram.

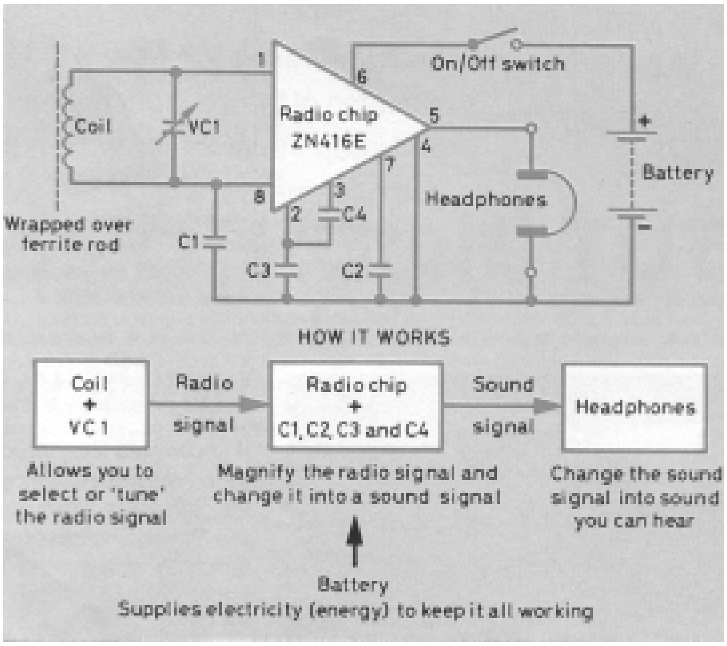
On/Off switch
Coil
VC1
Radio chip
ZN416E
1
2
3
4
5
6
7
8
C4
Battery
+
−
Headphones
Wrapped over
ferrite rod
C1
C3
C2
HOW IT WORKS
Coil
+
VC 1
Radio
signal
Radio chip
+
C1, C2, C3 and C4
Sound
signal
Headphones
Allows you to
select or 'tune'
the radio signal
Magnify the radio signal and
change it into a sound signal
Change the sound
signal into sound
you can hear
Battery
Supplies electricity (energy) to keep it all working

Figure 1 Circuit and block diagrams of the radio

Description

The whole circuit is built on a 50 mm by 50 mm printed circuit board (PCB) designed to fit on the inside of the lid of a plastic box, and is stuck there using sticky pads, the shaft of the variable capacitor going through a hole in

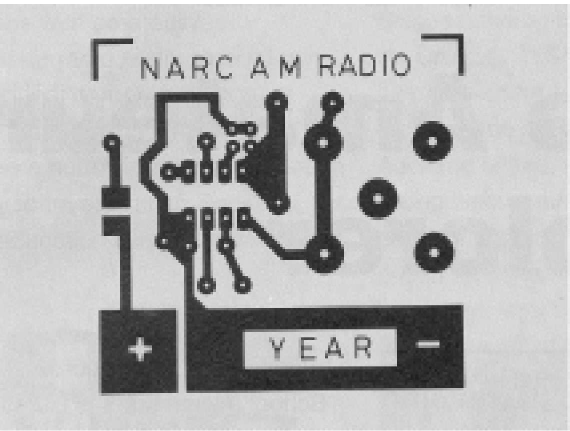

Figure 2a The PCB, solder side

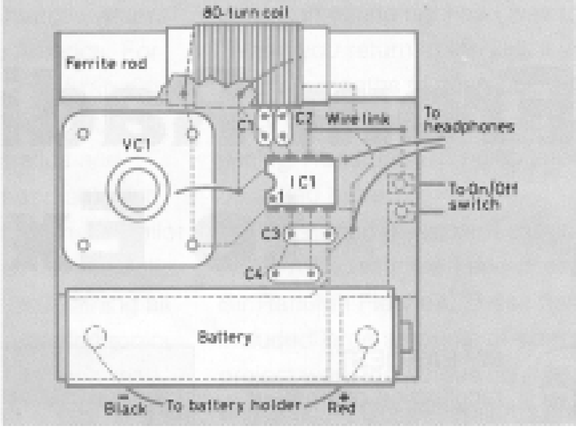

Figure 2b The PCB, component side

the lid. Only two pairs of leads are soldered to the board – one pair goes to the 1.5 V battery in its holder, and the other to the earphone socket. **Figures 2a** and **2b** show the printed circuit and the component layout double size for clarity. You are not obliged to build the circuit on a PCB.

Building it

1. **Check and identify components**. Tick the parts list.
2. **Carefully unwind the wire.** Use paper to make an insulating tube (called a 'former') around the centre of the ferrite rod and secure it with Sellotape. Now, close-wind *all* the wire (leave no gaps between adjacent turns) around the paper former. Secure the winding with more Sellotape, leaving 50 mm of wire free at each end for connection to the circuit. See **Figure 3a.**
3. **Solder in VC1.**
4. **Solder in the integrated circuit holder.** There is a notch in one end of the holder; this should face VC1. Solder also the wire link and the capacitors. Be careful to avoid solder 'bridges' between adjacent tracks on the PCB.
5. **Solder the battery leads.** These must be connected properly – the red battery lead to the + (positive) area and the black lead to the – (negative) area.
6. **Strip bare 1 cm of insulation from the ends of two wires**. Solder them between the PCB and the headphone socket (see **Figure 3b**). Use the end tabs on the socket. Using another pair of insulated wires connect the ON/OFF switch to the PCB tabs shown in Figure 2b.

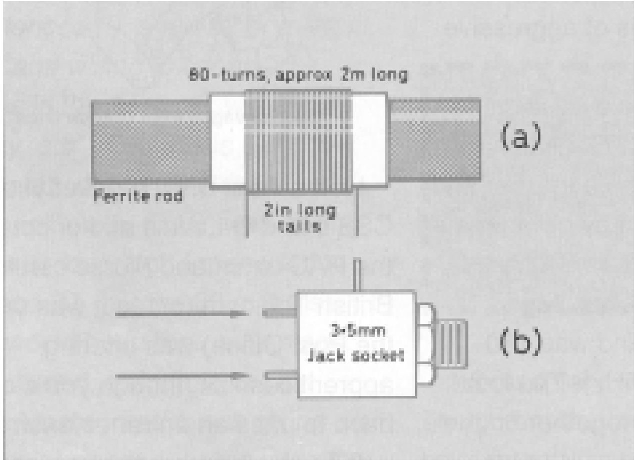

Figure 3 Details of coil and headphone socket

7. **Fix the elastic band.** This goes through the holes at the top of the PCB, with the ferrite rod being slipped through the two end loops. (Note: although the coating on the copper wire is designed to melt away during soldering, it is quite common for difficulty to be experienced in obtaining a good soldered joint; to be on the safe side, remove the coating *before* soldering (with a small piece of sandpaper).) Carefully place the wire ends of the coil through the PCB just above VC1, and solder on the track side.
8. **Fit IC1 into its holder.** This should be done carefully, making sure that *all* the pins are located above their respective clips *before* applying any pressure! Make sure also that the notch on the IC (as shown in Figure 2b) matches the notch in the holder, and faces VC1.
9. **Put battery in its holder.** Listen for some noise in the headphones as VC1 is rotated. Make sure the headphone plug is fully inserted into its socket.
10. **Fix the working board to the lid.** Use the sticky pads and apply *gentle* pressure. Fit the tuning knob, the ON/OFF switch and the earphone socket.
11. **Test again.** If all is still working, fit the lid screws and admire your completed radio!

Parts list

Capacitors	
C1, C2	0.01 microfarad (μF)
C3, C4	0.1 microfarad (μF)
VC1	500 picofarads (pF)
Semiconductor	
IC1	ZN416E

Additional items
- Plastic box (recommended size 76 × 64 × 50 mm internal)
- 8-pin DIL socket for IC1
- Printed circuit board
- Tuning knob for VC1
- Wire link for PCB
- 2 m of 30 SWG copper wire, self-fluxing
- Piece of paper 25 × 50 mm, to make the coil former
- Ferrite rod 70 mm long by 10 mm diameter, approximately
- Battery, AA size 1.5 V, with holder and attached wires
- Miniature earphone socket (3.5 mm stereo jack)
- ON/OFF switch (push-button SPST latched or slide switch)
- 4 off 100 mm insulated connecting wires, for jack socket and ON/OFF switch
- Pair Walkman-type earphones
- Elastic band, to attach ferrite rod to PCB
- 4 off sticky pads for securing PCB to box lid

Kits

Ready-made PCBs may be available from Alan J. Wright, G0KRU, Hewett School, Cecil Road, Norwich NR1 2PL.

4 A simple electronic organ

Introduction

This project has nothing to do with radio but, let's admit it, *any* electronics project is good experience! Why not build this little organ – it will keep the children amused at least! It uses the popular NE555 integrated circuit, which contains a circuit which will periodically switch the voltage on the output pin between the supply voltage and zero. Just how frequently this switching occurs depends upon the components external to the integrated circuit. If this switching occurs several hundred or thousand times a second, the change in voltage produced will generate a musical note when connected to a small loudspeaker. The circuit is shown in **Figure 1**.

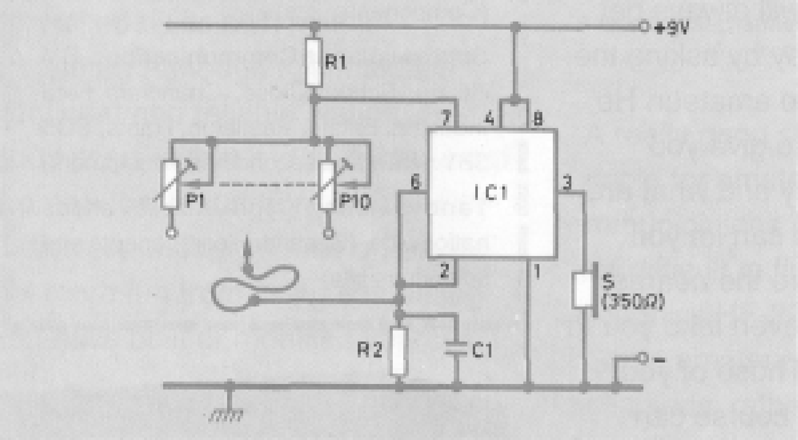
+9V
R1
7
4
8
IC1
6
3
P1
P10
2
1
S
(350Ω)
R2
C1
-

Figure 1 Circuit diagram

Putting it together

(a) Using a PCB. The job is very simple. The placement of components on the unsoldered side of the board is shown in **Figure 2** and the design on the copper track is illustrated in **Figure 3**. Put each component, in turn, on the board, making sure that it lies flat on the board with its tags or wires going cleanly through the holes provided for it; then, solder the wires to the board, cropping them before or after the soldering,

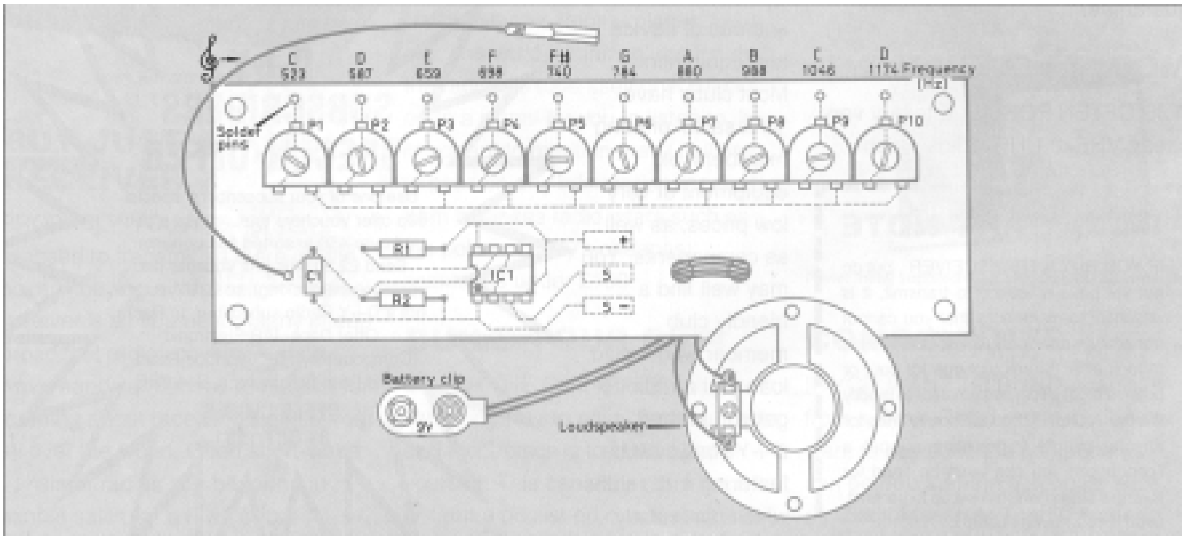
C
523
D
587
E
659
F
698
F#
740
G
784
A
880
B
988
C
1046
D
1174 Frequency (Hz)
Solder pins
P1
P2
P3
P4
P5
P6
P7
P8
P9
P10
R1
R2
C1
IC1
+
S
S -
Battery clip
9V
Loudspeaker

Figure 2 Position of components on the printed circuit board (PCB)

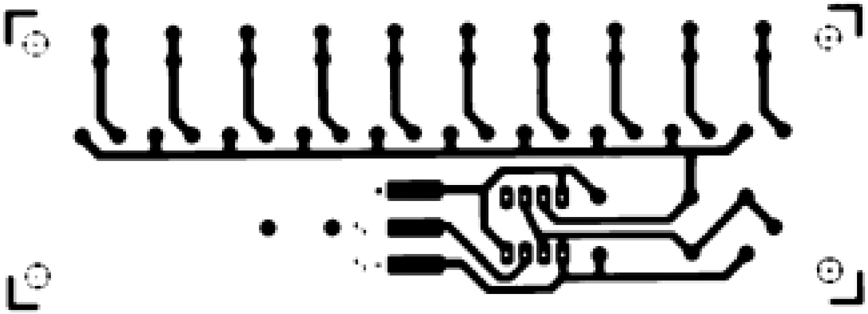

Figure 3 The connections

depending on your preference. If you choose to use a holder for your integrated circuit (highly recommended if your soldering is less than perfect), make sure that the end with a notch in it faces R1 and R2, as shown in Figure 2. Solder the two leads to the speaker to the tabs marked S (either way round), having looped them through the two holes to the right of the tabs in Figure 2. Looping them through the holes acts as a strain relief, ensuring that the soldered joints are not subjected to pulling and bending as you move the wires about. Do the same with the battery leads, the red lead going to the + tab and the negative lead to the – tab (which also has one speaker lead already attached to it). **Figure 4** shows this in detail. Treat the loudspeaker with care – the cone is quite fragile and must not be touched.

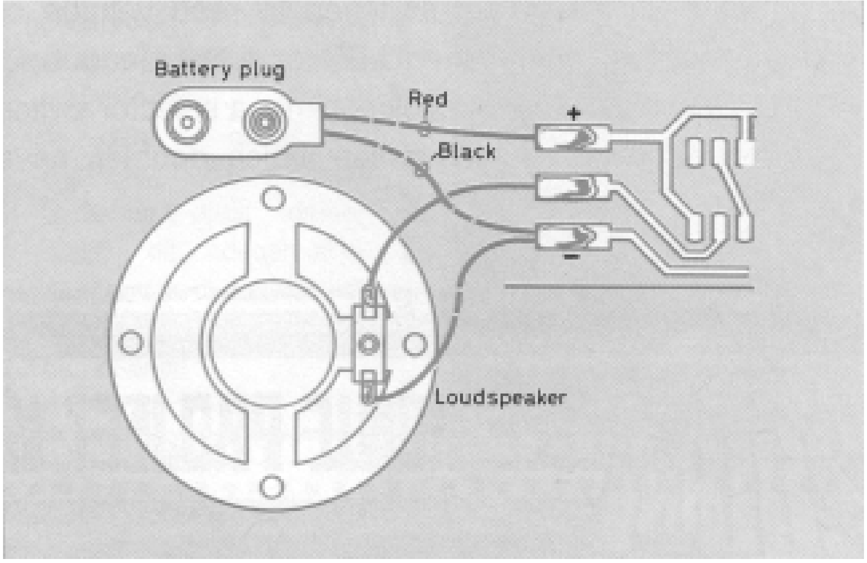

Figure 4 Battery plug and loudspeaker connections

(b) Without a PCB. This is more difficult, and you may need to enlist some help. Using some matrix board (such as Veroboard) is probably the best way of replacing the PCB. You could arrange your circuit in exactly the same way as in the PCB in Figure 3, using wires to replace the copper track.

A simple ‘keyboard’

The keyboard is a row of solder pins along the rear edge of the PCB, one for each note covering the range shown in Figure 2. A flying lead with a small spade on it is provided to touch any of the pins in turn, producing any one of ten different notes.

Testing

Check first that each component is in the correct place. When inserting the NE555 chip, first make sure that the end carrying the notch lies over the end of the holder with the notch; then, make sure each pin of the chip lies directly above the hole into which it fits, before pressing *gently* to insert the chip into the socket. Make sure the battery connections are correct, and insert the battery into the clip. Nothing should happen, except for a click from the loudspeaker; touching the spade on any of the pins should produce a coarse note from the speaker. If nothing happens, check everything again; don’t *assume* that wires go where you think they go!

After you get the first note, all the others should work, too, but they will sound off-tune at first. The organ needs tuning up by adjusting the 10 preset variable resistors P1 to P10. The approximate frequency to which each note should be tuned is given in Figure 2; if you can beg, borrow or steal a frequency counter, setting up is easy. If you have a piano, the organ can be tuned by comparison of the notes with those on the piano. The frequencies are given in Hertz (abbreviation Hz), and represent the number of times the IC switches on and off every second. If the sound coming from the loudspeaker is too loud or very distorted, then try putting an 330 Ω resistor (colour code orange, orange, brown) in series with the loudspeaker. This is done by taking the resistor and cutting its leads to about 5 mm; then, disconnect one speaker lead from the tab on the PCB (it doesn’t matter which). Solder one end of the resistor to the vacated speaker tab, and the free speaker lead to the other end of the resistor. This will limit the volume of sound from the speaker, and lengthen the life of your battery. If it is still too loud, try a resistor of a larger value, or use a smaller resistor to make it louder.

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1	2.7 kilohms (kΩ)
R2	1 Megohm (MΩ)
P1–P5	Preset resistor 100 kilohms (kΩ)
P6, P7	Preset resistor 50 kilohms (kΩ)
P8	Preset resistor 25 kilohms (kΩ)
P9, P10	Preset resistor 10 kilohms (kΩ)

Capacitor

C1	100 nanofarads (nF) or 0.1 microfarad (μF)

Integrated circuit

IC1	NE555 timer chip

Additional items

S	Loudspeaker >60 ohms (Maplin)
	1 off battery clip (for PP3 battery)
	1 off spade terminal
	12 off solder pins ‘Veropins’
	3 off 10 cm lengths of ‘hook-up’ wire

(This article is based on projects originally designed by Radio Scouting, Netherlands.)

5 Experiments with the NE555 timer

Introduction

Several of the projects in this book use the NE555 timer, an integrated circuit which is at the heart of many circuits whose processes are determined by time intervals. **Figure 1** shows the circuit diagram of an audio oscillator using the 555. The timing voltages (governing the frequency of oscillation) are produced by R1, R2 and C2; a voltage appears at pin 3 which ‘switches’ at this frequency between zero and a voltage close to the supply voltage, which in this case can be anywhere between 6 V and 14 V. The output current, when applied through R3 to a small loudspeaker, produces an audible tone, provided that there is a DC path between the two test leads.

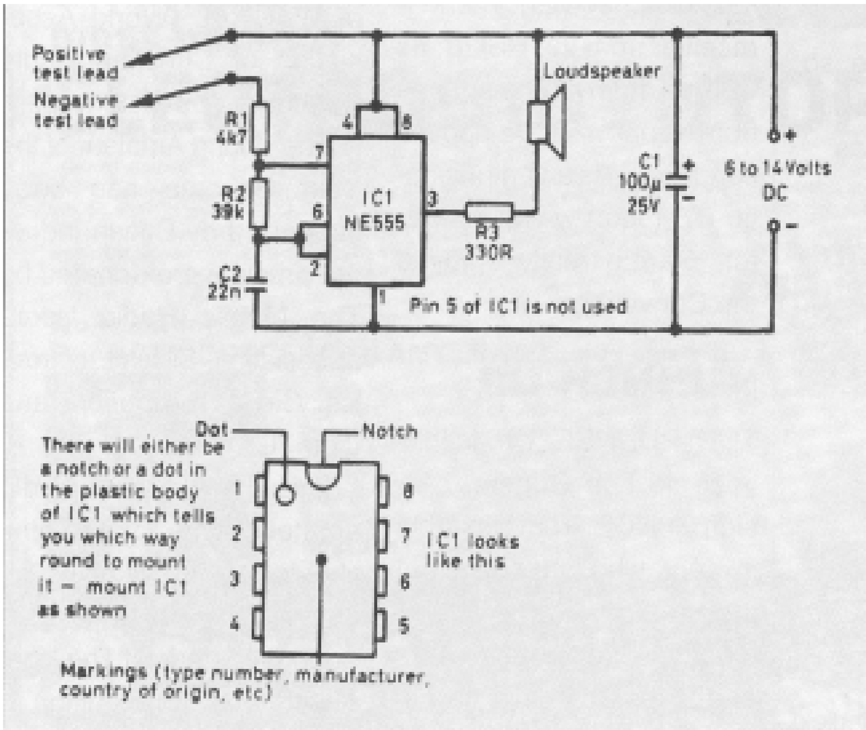
Positive test lead
Negative test lead
R1 4k7
R2 39k
C2 22n
4
8
7
6
2
1
3
IC1 NE555
R3 330R
Loudspeaker
C1 100µ 25V
+
−
6 to 14 Volts DC
+
−
Pin 5 of IC1 is not used
Dot
Notch
There will either be a notch or a dot in the plastic body of IC1 which tells you which way round to mount it – mount IC1 as shown
1
2
3
4
8
7
6
5
IC1 looks like this
Markings (type number, manufacturer, country of origin, etc)

Figure 1 The circuit diagram

Construction

The simplest way to mount the components is on a piece of matrix board (Veroboard), available from any of the good suppliers. The prototype of this circuit used the type of board with copper strips along the underside; these strips are used like the copper tracks on a PCB, to join components together. Firstly, cut the four strips between the positions of the pins of the IC socket, as shown in **Figure 2**. You can buy a tool for this purpose, but a small twist drill (about 3 mm diameter) is just as good. Turn it between your fingers – if you use a drill you will end up with holes right through the board! Then solder in the IC socket (with the notch in the position shown), followed by the four links made with single-conductor insulated wire. Put in each component as shown, ensuring that C1 (an *electrolytic* or *polarised* capacitor) is connected correctly. When all the components have been soldered in, take the 555 chip and lay it on its socket, with its own notch lying above that of the holder. Then, making sure that each pin lies *directly* above its corresponding socket, press down *gently* on the chip, with the board supported on a flat surface.

Testing

Connect the circuit to a battery or small power supply, ensuring that the positive and negative leads are the right way round. Always use red and black leads here, then you are less likely to get it wrong! Switch on. Nothing

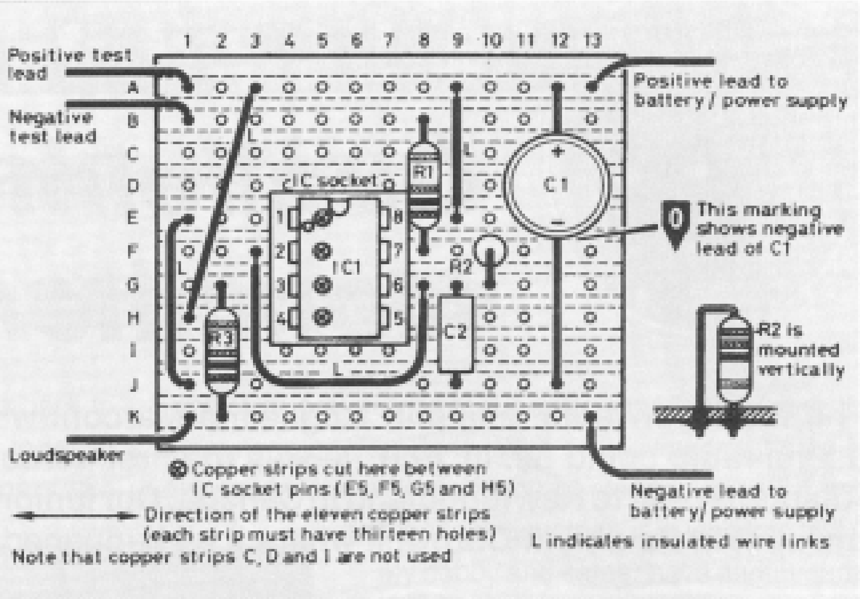
1 2 3 4 5 6 7 8 9 10 11 12 13
A B C D E F G H I J K
Positive test lead
Negative test lead
Loudspeaker
IC socket
1 2 3 4 5 6 7 8
IC1
R1
R2
R3
C1
C2
+
–
L
Positive lead to battery / power supply
This marking shows negative lead of C1
R2 is mounted vertically
Negative lead to battery/ power supply
Copper strips cut here between IC socket pins (E5, F5, G5 and H5)
Direction of the eleven copper strips (each strip must have thirteen holes)
Note that copper strips C, D and I are not used
L indicates insulated wire links

Figure 2 Veroboard layout. If you can read a circuit diagram, the project can be built using other methods

should happen until you short together the two test leads, when there should be a note from the loudspeaker. If this doesn't happen, switch off, disconnect the circuit and check your wiring and soldering. Is it exactly like Figure 2? Are the soldered joints round and shiny? If any are dull, then 'sweat' them briefly with a hot soldering iron until the solder runs, remove the iron, and check that they are as shiny as the rest. Check that there are no solder 'bridges' between adjacent tracks by holding your board up to a strong light. Then, reconnect, switch on and touch the test leads together. All should now work!

Uses of your circuit

1. As a Morse practice oscillator. Simply connect the two test leads to your key and, each time the key is pressed, you should hear a note from the speaker. The frequency of the note may be altered by putting a resistor in series with the key. To do this, remove one test lead from the key and select a resistor; connect one end of the resistor to the free test lead and the other to the empty terminal on your key. Selecting the value of resistor that you need will be a useful experiment in itself.
2. As a continuity tester. You can check fuses and lamp bulbs by connecting them across your test leads. If the speaker remains silent, the fuse or bulb has blown.
3. To indicate changes of resistance. Hold the ends of the test leads in each hand; you should hear a low note, because of the high resistance of your body. Squeeze the ends harder, and the frequency of the note should rise, because you are now making better contact. Repeat this with damp hands and the frequencies will be higher still.
4. As a thermometer. Connect the test leads to a thermistor (a device whose resistance changes with temperature) and warm it with a hair-dryer, or even in your hands, and you will hear the pitch changing with the temperature of the thermistor. A suitable 'bead' thermistor is available from Maplin (order code FX21).
5. As a diode tester. Use any diode, and connect the negative test lead to the end of the diode marked with the ring. This is the cathode of the diode. The other end, the anode, should be connected to the positive test lead, and a note should be heard from the speaker. This does not necessarily mean that the diode is working – yet. Reverse the connections and *nothing* should be heard. If this is the case, the diode *is* working.
6. As a light meter. Use a photoconductive cell (a device whose resistance changes with light intensity) connected between the test leads. A note should be heard. Shading the device with your hand will increase its resistance and the note should decrease in frequency. A suitable device is the ORP12 cell from Maplin (order code HB10).

There are many more applications. **Do not connect the test leads to other circuits that are switched on.** Your circuit, or the circuit you are connecting it to could be damaged. Think of a *passive* device or circuit (i.e. one *not* requiring a power supply or battery) where changes of resistance occur, and you have found another application!

Parts list

Resistors: all 0.25 watt carbon film types

R1	4.7 kilohms (kΩ) – yellow, violet, red
R2	39 kilohms (kΩ) – orange, white, orange
R3	330 ohms (Ω) – orange, orange, brown

Capacitors

C1	100 microfarads (μF) 25 V radial electrolytic
C2	22 nanofarads (nF) or 0.022 microfarad (μF) polyester type with 10 mm lead spacing

Integrated circuit

IC1	NE555

Additional items

- Miniature loudspeaker (preferably 35, 40 or 80 ohm)
- 8-pin DIL socket for IC1
- 0.1 inch Veroboard (‘stripboard’), size 11 strips by 13 holes
- PVC-covered stranded wire for test leads, loudspeaker and battery connections
- PVC-covered solid wire for links on the board
- A power source of between 6 V and 14 V, such as a 9 V battery (PP3)

Component sources

Cirkit
Tandy – many high street shops

6 A simple metronome

Introduction

A metronome is a device used by musicians to indicate the tempo of a piece of music. Until electronics came on the scene, this 'beating of time' was achieved in much the same way as a clock keeps time, i.e. with a pendulum device, the clicking of the escapement indicating the beats of the music.

Those of you who have already built the Morse Key and Buzzer from the designs in this book, will recognise the circuit of this metronome – it is exactly the same as was used to produce the note of the buzzer. This circuit is shown in **Figure 1**.

The circuit

Three components determine the speed at which the circuit oscillates – the speaker (LS), the resistors (VR1 + R1) and the capacitor (C1). VR1 is a variable resistor, so that the speed at which the oscillator operates can be varied. Compared with the component values of the Morse Buzzer (which operated at around 800 Hz), these components now give an oscillation frequency of around 1.25 Hz, which is far too low to be heard as a note. What we *do* hear, however, is a series of clicks, as the voltage across the speaker changes quickly from 0 to 9 V and back again.

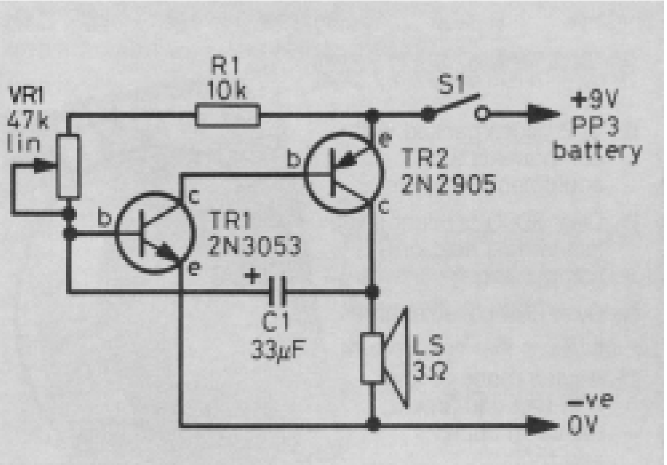

Figure 1 The metronome circuit is rather like the Morse oscillator

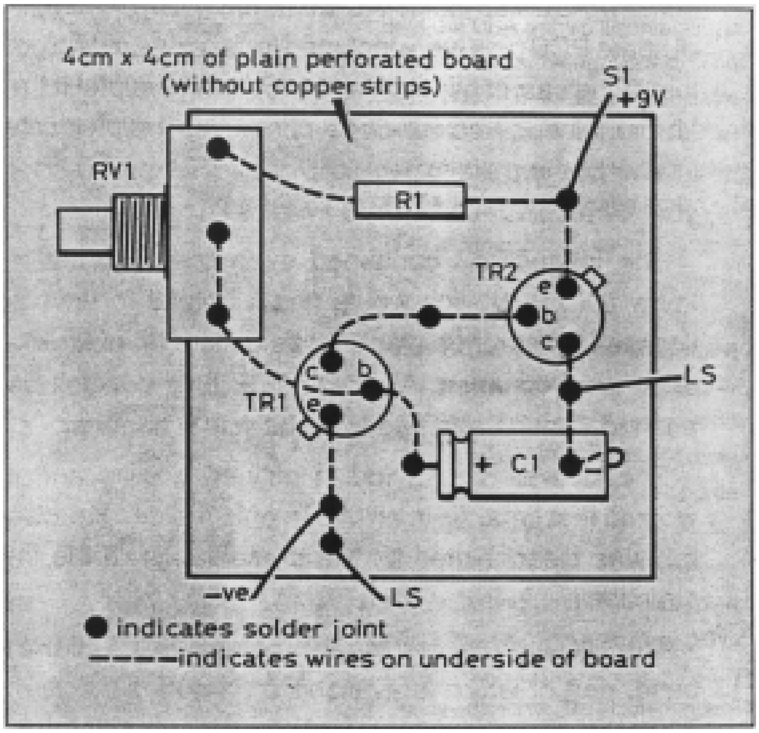
4cm x 4cm of plain perforated board
(without copper strips)
S1
+9V
RV1
R1
TR2
e
b
c
LS
c
b
TR1
e
+ C1
−ve
LS
● indicates solder joint
-----indicates wires on underside of board

Figure 2 The component wires are pushed through holes in the circuit board and joined together underneath

Variation of speed could be achieved by varying resistance or capacitance. However, as you may already know, variable capacitors have values in the picofarad range, not the tens of microfarads used here, so it is very simple to employ a variable resistor (potentiometer) to control the oscillator. You could use a multi-way switch to switch in one of several capacitors, as well as having the variable resistor, but this was found to be an unnecessary complication. This design operates between about 100 clicks per minute and 200 clicks per minute.

Making the prototype

A single piece of plain matrix board (no copper strips) measuring about 40 × 40 mm is sufficient to hold all the components except the potentiometer and switch (see later). The case can be plastic or aluminium, and one measuring 65 × 100 × 50 mm is about right. Make sure there are holes in the case beside the speaker cone to let the sound out, and larger holes for the potentiometer and switch. If a potentiometer is used with a combined ON/OFF switch, then the extra hole for the switch is not necessary! It is advisable to construct the circuit *before* putting it in the box, so that it can be tested to ensure that everything is working. If it is, then you can exercise your ingenuity in mounting the speaker, battery and board inside the box. A final test can be made before starting the calibration process.

Calibration

There is no ‘easy’ way to do this. The frequencies involved are too low to be measured with the average frequency counter, so you will need to resort to using a stopwatch and counting the number of clicks per minute.

Parts list

Resistors: 0.25 watt, 5% tolerance

R1	10 kilohms (kΩ)
VR1	47 kilohms (kΩ) linear potentiometer

Capacitor

C1	33 microfarads (μF) electrolytic

Transistors

TR1	2N3053	npn
TR2	2N2905	pnp

Additional items

S1	SPST	ON/OFF switch
LS	3 ohms (Ω)	loudspeaker
	Knob with pointer for VR1	
	PP3 battery and connector	
	Aluminium case, 65 × 100 × 50 mm	
	Matrix board (plain), 40 × 40 mm	

7 What is a resistor?

Introduction

Materials that carry electricity easily are called **conductors**. They include all metals and salt water, for example. We use wire as a conductor, and the ease with which it passes an electric current depends upon the material, its thickness and its length. Silver (symbol Ag), gold (Au), copper (Cu) and aluminium (Al) are the best metallic conductors. Most wires are made of copper, although the best conductor, weight for weight, is aluminium.

Materials that *don't* carry current (or, at least, do so very badly) are called **insulators**, and they include dry wood, rubber, plastic and glass among their number. Wires are often coated with a layer of insulator to prevent adjacent wires touching and causing an accident.

Resistors

If there wasn't such a thing as resistance, the subject of electronics wouldn't exist; only infinite currents would flow and voltages wouldn't exist either! We need to reduce the flow of current if we are to make current do something useful for us. Components that resist the flow of current are called **resistors**, and they are said to have a **resistance** which is measured in ohms (Ω), named after Georg Ohm, who formulated the law (also named after him) by which the voltage and current through a conductor are related. His law gave rise to the formula *everyone* remembers:

$$I = \frac{V}{R},$$

where I is the current flowing, measured in amps,
V is the voltage across the conductor, and
R is the resistance of the conductor, measured in ohms.

From this equation, you can see that, for a constant value of voltage, V, if the resistance goes up, the current will go down, and vice versa. The circuit symbols for resistors are shown in **Figure 1**. You will find the upper symbol in older magazines; it is still preferred by many engineers. The lower symbol is the prevalent standard symbol.

Resistors are made in several ways, the cheapest using carbon; another type is usually made from a ceramic cylinder (used only as a support) on which

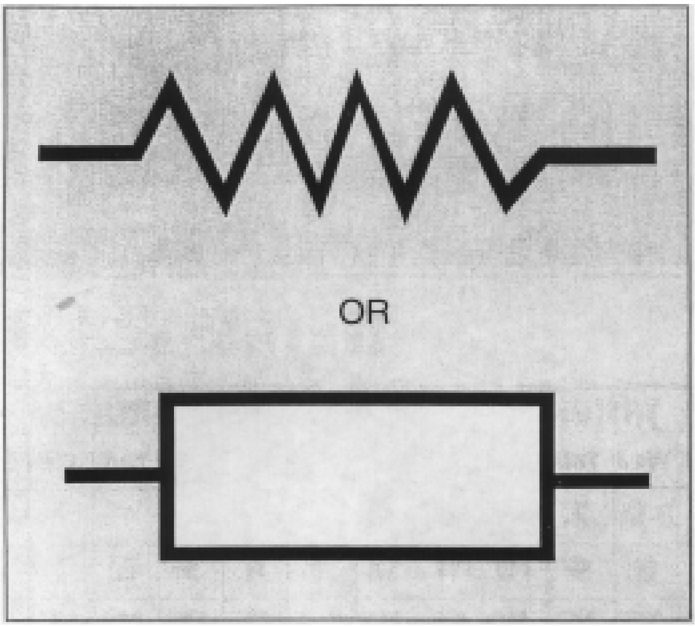

Figure 1 Circuit symbols for resistors

is placed a very thin film of metal – the thinner the film, the greater the resistance. All resistors are coated with a thin film of insulation, for the same reason we discussed earlier.

The colour code

Each resistor has coloured bands on it which enable us to see what value of resistance it has. There are normally three (but sometimes four) at one end, and a single one at the other (see **Figure 2**). The colours indicate figures, according to the list below.

Colour	*Value*	*Colour*	*Value*
Black	0	Green	5
Brown	1	Blue	6
Red	2	Violet	7
Orange	3	Grey	8
Yellow	4	White	9

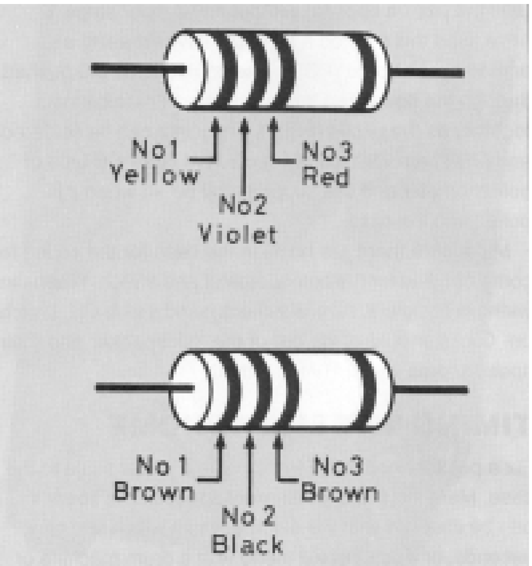

Figure 2 Some examples of resistor colour codes; top 4700 Ω (4.7 kΩ) and bottom 100 Ω

Using the colour codes is easy, once you see the logic behind it. Hold the resistor so that the single band is towards the right. The three colours on the left are read in the normal order from left to right. The first two bands *always* indicate numbers; the third band gives the number of zeros to add to the right of these two numbers. So, looking at the top resistor in Figure 2, yellow, violet, red means 4, 7, and two zeros, giving 4700 ohms! Looking at the lower resistor, brown, black, brown means 1, 0, and one zero, giving 100 ohms.

Remembering the order of the colours may be difficult at first. The colours from red to violet are the colours of the rainbow, in order, so if you know those, you're almost there! Around those colours are black and brown below the red, and grey and white above the violet, which you can imagine as getting brighter from black to white (well, almost!). It won't be long before you don't need to remember them at all.

The isolated band on the right-hand side is not part of the resistor's value; it indicates its *tolerance*, i.e. how close it might be to the indicated value. A brown band indicates ±1%, a red band ±2%, a gold band ±5% and a silver band ±10%. For example, a resistor marked as being 100 ohms with a ±5% tolerance will have an *actual* value somewhere between 95 ohms and 105 ohms.

8 Waves – Part 1

Introduction

Waves are responsible for most of the processes in life where energy is transferred from one place to another. Heat and light energy from the sun, for example, come to us as *electromagnetic waves*. Sound travels through the air as a wave; it is not the same sort of wave as light or heat, but it obeys many of the same properties. Damage is caused to coastal margins by the waves of the sea – again, another type of wave, but still obeying many of the same properties. Radio waves are of the same type as heat and light waves, as are gamma rays, X-rays, ultra-violet waves and infra-red waves. So, once we begin to understand what radio waves do, we are also learning about a huge chunk of physics at the same time! All these waves are part of the *electromagnetic spectrum*. The word ‘spectrum’ simply means a ‘range’, so what we have is a range of electromagnetic waves – that’s all!

Sensing things

Light waves are invisible, but our eyes can detect the *effect* they have on different materials because the waves produce an effect on the retina of the human eye which the brain can interpret. We cannot see heat waves either, but we can feel the *effect* they have on our skin. Gamma rays and X-rays are also invisible, but their detrimental *effects* on human tissue are well known. It is not surprising, then, that we cannot see radio waves. We cannot sense them, either, until we produce a device upon which they have an *effect*. That device is a **radio receiver**; it is able to process certain characteristics of radio waves, and make these characteristics audible by generating sound waves from the loudspeaker or headphones. Other characteristics of the same waves may be turned into light as a TV picture on a cathode-ray tube, or as a fax image on a sheet of paper.

Visible waves

Let’s start our description with some waves that we can actually see! When a small stone is thrown into a pond, we see circular water waves *radiating* from the point where the stone fell into the water, as **Figure 1** shows. (Notice that we use the word *radiating*, even with water waves; it is not a radio term, but one which describes any motion where the *radius* of a circle is increasing. In this case, it is the radius of the circular waves which is

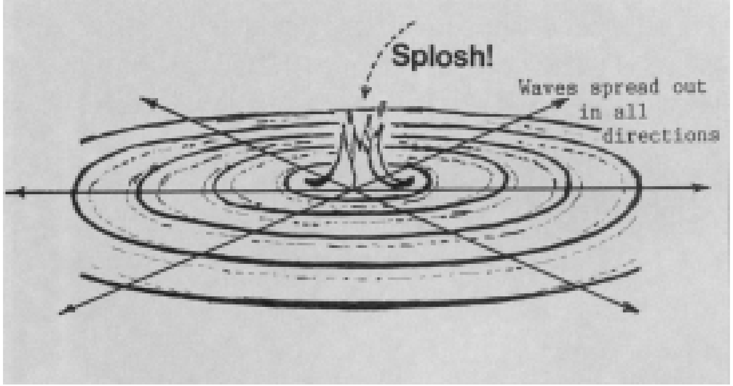
Splosh!
Waves spread out in all directions

Figure 1 The waves move out from the point where the stone landed

increasing.) If you were to watch the water waves down at the water level, perpendicular to the direction in which they are travelling, you would see something like the illustration in **Figure 2**.

The horizontal line represents the water level before the wave started, and the vertical line represents the direction in which the water is displaced at any instant.

All waves are described in the same way

'Freezing' the motion of the water in this way allows us to define two very important characteristics of a wave, characteristics which we talk about every day – *wavelength* and *amplitude*. The wavelength of a wave is simply the distance (measured in metres) from any point on one wave to the same point on the adjacent wave. Look at the diagram and you will see what is meant by 'the same point on the adjacent wave'. The amplitude of a wave

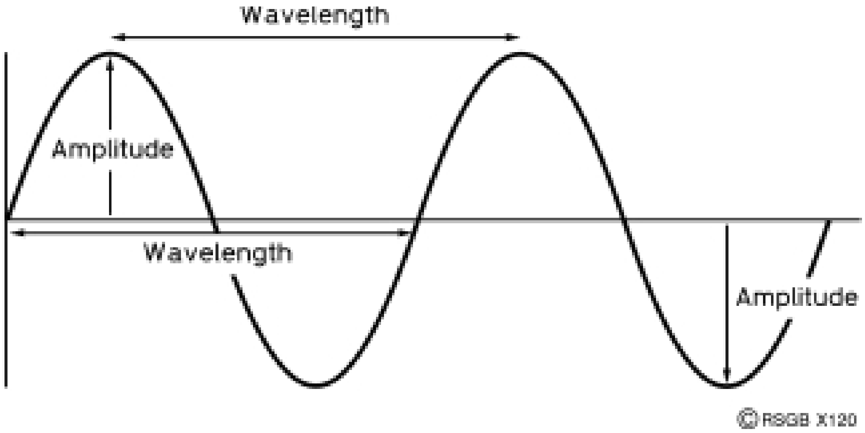

Figure 2 A water wave, viewed in cross-section

is always measured from the centre (undisturbed) position of a wave to the peak (or the trough) of the wave. Both these positions are shown, the arrow indicating that the measurement is taken *from* the centre *to* the peak or trough. The amplitude of a wave is defined as the maximum displacement of the wave from the centre position – the direction (up or down) of that displacement does not matter. Waves of greater amplitude carry more energy with them.

If we now 'unfreeze' the wave, we will see it travel from left to right (or right to left, depending on where we are looking). The speed at which it moves is called its *velocity*, and is measured in metres per second.

Another useful word is *propagate*; it means *travel*. We talk about radio waves propagating from a transmitter to a receiver. This velocity of propagation (for electromagnetic waves) is very fast indeed – they will cover 300 million metres in one second. This is virtually incomprehensible, so think of a radio wave travelling around the earth – it can travel 7½ times round the earth in one second! We use the symbol *c* for the velocity of radio waves (which is the same as the velocity of light, of course – all electromagnetic waves travel at this speed through air and space).

The last thing we need to know about the wave is its *frequency*. Imagine a cork floating on the water in the path of the wave; it will bob up and down. If we were able to count the number of times it went through its highest position in one second, then that number would be its frequency. Any periodic motion like this is said to go through one *cycle* each time one complete wave passes a point (in this case, our cork). We are thus counting the number of cycles per second of the cork's motion. The unit of frequency is thus 'cycles per second'; this unit is now named after Hertz, a radio pioneer, and is abbreviated to Hz.

Our description of the wave is now quite simple – we need only four quantities:

(a) **Frequency** symbol f – unit, hertz (Hz)
(b) **Wavelength** symbol λ (Greek letter lambda, pronounced 'lamb-da') – unit, metre (m)
(c) **Amplitude** symbol a – unit depends on application
(d) **Velocity** symbol c – unit, metres per second (m/s or ms^{-1})

The basic formula

Whatever may happen to a wave while it travels through different media (vacuum, air, brick, wood, etc.), one thing and **only** one thing remains constant – its frequency. Its wavelength, amplitude *and* velocity may change, but its frequency never does. Three of the four characteristics already identified are connected by the simple relationship

$$c = f \times \lambda.$$

Remember that *c* is constant if the wave travels in air or in a vacuum. This means that waves having higher frequencies (*f* large) must have smaller wavelengths (λ small) and vice versa. You can imagine frequency and wavelengths being on opposite ends of a see-saw!

Divisions of units

Because the frequencies of radio waves are so high (despite them having the lowest frequencies in the electromagnetic spectrum!) we have a problem with writing them down. Do you write in your log book that you have just heard a station on 14 100 000 Hz? Of course not, you write it as 14.1 MHz, knowing that the prefix mega (M) means 'one million'. The prefixes which you need to know (when applied to frequency) are:

kHz	kilohertz	meaning	1000 Hz
MHz	megahertz	meaning	1 000 000 Hz
GHz	gigahertz	meaning	1 000 000 000 Hz.

Notice that the 'k' in kilohertz is a lower case letter. It is **incorrect** to write it as an upper case letter. 'K' is a computer-related prefix meaning **not** 1 000 but 1 024!

When we come on to discuss heat and light waves, we will use wavelengths rather than frequencies, because of the see-saw effect – as the frequencies get larger and larger, the wavelengths get smaller, and hence are numbers which are more manageable, both to talk about and to write down!

Bands

Gamma rays, X-rays, ultra-violet waves, light waves, infra-red waves are all part of the electromagnetic spectrum, but we divide them up because they have different properties. This is why we divide up our radio frequencies into different bands. The radio waves of top-band signals (around 2 MHz) have completely different properties compared with those in the 20 metre band, so we are dividing up the *radio spectrum* in the same way – by property.

9 A beat-frequency oscillator

Introduction

Many readers will know that, although they have a short-wave radio which covers at least one of the amateur bands (e.g. 7 MHz or 14 MHz), they are unable to listen to SSB or Morse signals. This is because the receiver lacks a Beat-Frequency Oscillator (BFO). We need the ‘carrier’ frequency of a BFO to replace the carrier that has been removed from the signal at the transmitter. When listening to Morse signals, the BFO signal ‘beats’ with the incoming signal to produce a note in the loudspeaker. If you are a musician, you will be familiar with the method of using ‘beats’ to tune one musical instrument from another; in the BFO, the beat frequency produced is the tone signal you hear.

In the more complex amateur radio receiver, a BFO is incorporated as part of the whole system. In our model, it is an external circuit that sits alongside your radio. The circuit diagram is shown in **Figure 1**.

Construction

Built on a small piece of matrix board about 80 × 50 mm, the circuit can be fitted inside a small plastic box. For once, we *don’t* want to screen the circuit to prevent it interfering with other equipment; we *want* it to interact

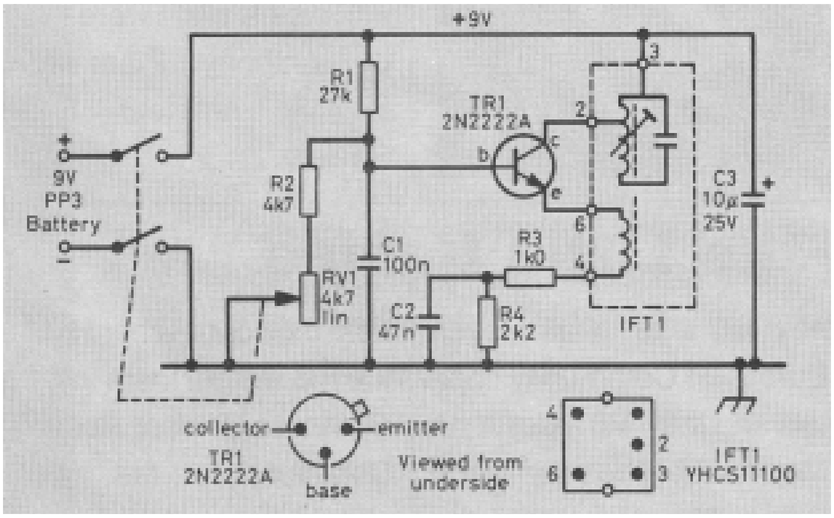
+9V
R1
27k
TR1
2N2222A
b
c
e
2
3
6
4
C3
10µ
25V
+
9V
PP3
Battery
R2
4k7
C1
100n
R3
1k0
RV1
4k7
lin
C2
47n
R4
2k2
IFT1
collector
emitter
TR1
2N2222A
base
Viewed from
underside
4
2
6
3
IFT1
YHCS11100

Figure 1 Circuit diagram of the BFO

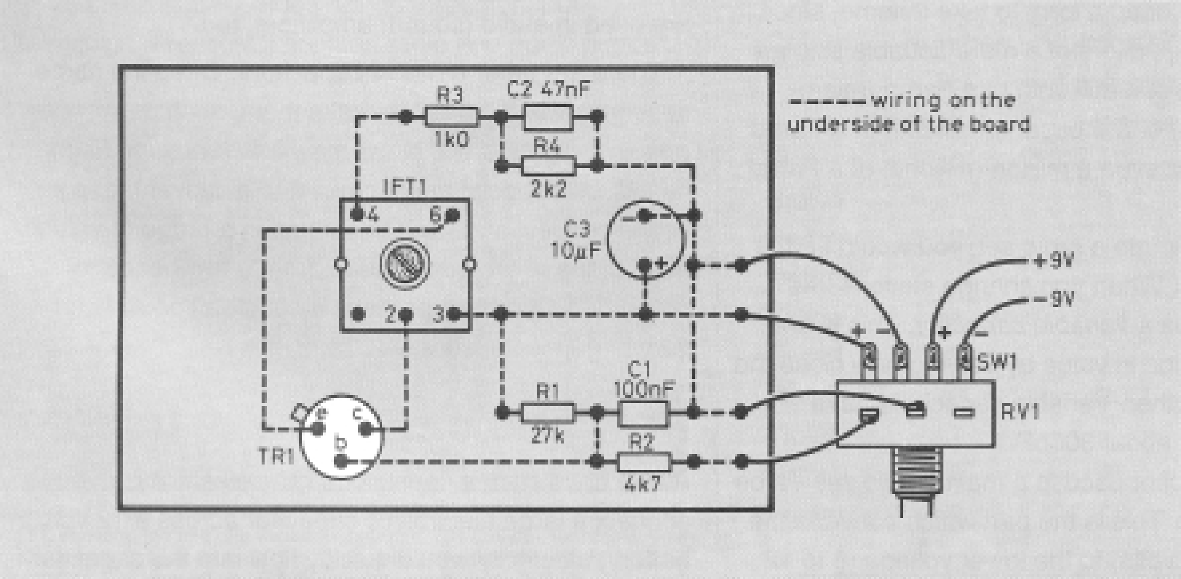

Figure 2 Matrix board layout shown from the component side. Adjust IFT1 carefully for the best results

with our receiver! This is why we use a plastic box. Maplin Electronic Components supply a suitable box, complete with the matrix board to fit inside (order code YU46).

Look at **Figure 2** carefully before you start to build the circuit, so that you can position the components correctly. Firstly, inspect the intermediate frequency (IF) transformer, IFT1, and remove its metal screening can very carefully. Again, this is to allow some signal to escape from our circuit and enter our radio. Having done this, solder the components, using the matrix board as a support. Underneath the board, the components are linked by single-conductor, insulated wire. Take particular care with the polarity of the electrolytic capacitor, C3, and the connections to the transistor, TR1 (see Figure 1).

The variable resistor, VR1, has a switch mounted behind the control itself, and the insulated leads to it from the battery should be about 10 cm long. Connect these before fitting VR1 into the case, so that the BFO can be calibrated (adjusted) correctly.

Calibration

After a final check that all the components have been fitted and soldered correctly, connect the battery, switch on, and hold the transistor between your fingers, to check that it is not getting hot. Place the circuit close to your receiver, and set RV1 to mid-position. Tune your receiver to find an amateur SSB transmission; the frequencies listed below will help you in knowing where to look. It may sound very strange, but don't worry. *Slowly* turn the core of IFT1 with a small, non-metallic screwdriver or with the correct 'trimming tool'. The core into which the blade fits is very fragile, so attempt this process with care. When the speech sounds as natural as you can get it, leave the core at this position, and use RV1 to make the speech sound natural.

Using the BFO

For best results, you may have to move the BFO nearer or further away from your radio. At the lower end of most bands (for instance just above 7.000 MHz or 14.000 MHz) you should be able to resolve Morse code (CW) signals. If you find that the BFO signal is a little weak, solder a 15 cm length of insulated wire to pin 2 of IFT1, and place it alongside your radio. This should improve signal intelligibility. When you are happy with the performance, switch off, drill a 10.5 mm hole in the box and fit RV1, followed by the matrix board assembly. Screw the base to the box, fit the knob, and the BFO is complete!

Where to listen

Band	**Frequencies (MHz)**
15 m	21.000–21.450
17 m	18.068–18.168
20 m	14.000–14.350
30 m	10.100–10.150
40 m	7.000–7.100
80 m	3.500–3.800

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1	27 kilohms (kΩ)
R2	4.7 kilohms (kΩ)
R3	1 kilohm (kΩ)
R4	2.2 kilohms (kΩ)
VR1	4.7 kilohms (kΩ), linear, with DPST switch

Capacitors

C1	100 nanofarads (nF) or 0.1 microfarad (μF), ceramic
C2	47 nanofarads (nF) or 0.047 microfarad (μF), ceramic
C3	10 microfarads (μF), 25 V radial, electrolytic

Additional items

TR1	2N2222A npn
IFT1	Toko type YHCS11100
Box	plastic, approximate size 100 × 70 × 45 mm
Board	matrix, to fit inside the box
Connector	for PP3 battery
Knob	for RV1

10 What is a capacitor?

Introduction

A capacitor is a device that stores energy in the form of electricity. Much less energy than a battery, and for a shorter time, however. The simplest form of capacitor takes the form of two flat metal sheets separated by air; connections are made to each plate, as **Figure 1** shows. If you imagine a pair of these plates, 30 cm square and separated from each other in air by 1 mm, the capacitance of this device would be almost exactly 80 picofarads (pF), i.e. 80 million-millionths of the unit of capacitance, the farad. Now this is quite a small value, you will agree, and it comes about because the farad is such a large unit. Nevertheless, as you will probably know, we may have capacitors of value 10 000 microfarads (μF) in our radio equipment, and they can be smaller than your little finger, so they are obviously not made the same way!

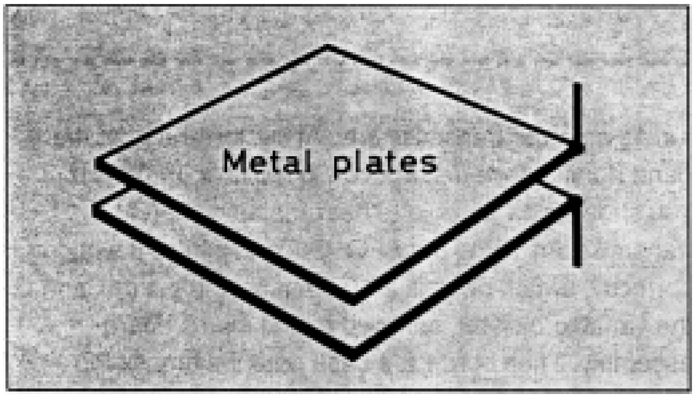

Figure 1

Large and small

We cannot go into the manufacture of capacitors here – after all, we are users of the devices, not the designers! First of all, beware of incorrect statements; the words 'capacitor' and 'capacitance' are *not* the same. For example, a large capacitor would be a description of one the size of a toilet roll. It need not have as large a capacitance as one the size of your little finger. A 'large capacitor' is one which is physically *big*, a 'large capacitance' refers to a capacitor which can store a larger amount of energy when a certain voltage is applied between its plates. The capacitors in a mains power supply are usually big **and** have large capacitances. High-power RF amplifiers may have large capacitors with small capacitances!

Electrolytics . . .

Electrolytic capacitors usually have capacitances of 1 μF or above. They differ from other capacitors in that they **must** be connected the right way round (i.e. they have positive and negative connections, just like a battery). They may explode if the connections are reversed!

. . . and the others

Other capacitors may be connected either way round, despite their names. We have *polystyrene*, *ceramic*, *silver-mica* and *tantalum*. Each has its own advantages and disadvantages, and the parts list for a project will always tell you which type is best.

Storing energy

If you were to connect a large capacitance across a 12 V power supply, nothing would appear to happen. Removing the capacitor from the supply and connecting it to a voltmeter would show that the capacitor has 12 V between its ends. This shows that, while the capacitor was connected to the supply, energy flowed from the supply into the capacitor. We say that the capacitor was *charged up* by the supply. If you are using an *analogue* voltmeter (i.e. one with a meter and pointer), you will notice that the indicated voltage slowly drops until, eventually, there is no voltage across the capacitor. This is because the capacitor has *discharged* its energy into the voltmeter. If you had used a smaller capacitance, the same would happen, except that the voltage would drop to zero more quickly – the capacitor stores a smaller amount of energy because its capacitance is smaller. Capacitors behave like other things in life – a small car can move more quickly than a large bus – a small piccolo emits a higher note than a flute – the voltages in a circuit containing a small capacitance will change more quickly than those in a circuit with large capacitance.

Varying the capacitance

Some capacitors are capable of having their capacitance changed manually; these are called *variable capacitors*. They work like the basic capacitor of Figure 1. Imagine moving the top plate of the pair a little to one side; the capacitance is determined, not just by the size of the two plates, but by their *area of overlap*. As this decreases, so does the capacitance. Such devices are limited in their capacitance, about 500 pF being the maximum value.

AC and DC

Because the plates of a capacitor do not touch each other, a direct current (DC) cannot pass between them. However, an alternating voltage on one plate can induce an identical alternating voltage on the opposite plate, and thus a capacitor appears to pass an alternating signal, even though currents as such, do not pass between the plates. This property of passing AC and not DC is very important, and a capacitor used in this way is called a *DC blocking capacitor* or, simply, a *blocking capacitor.* A blocking capacitor can be used at the same time, to couple a signal from one circuit to the next; here it would be known as a *coupling capacitor. Decoupling capacitors* are to be found where the capacitor is employed to remove an AC signal while retaining a DC component.

Finally . . .

Unlike resistors, the manufacture of capacitors renders them susceptible to excess voltage, so if you find a capacitor labelled 10 μF 16 V, it means that operating it above 16 V may fatally damage the device (and the circuit around it). This voltage is called the *working voltage* of the capacitor; on some electrolytics, you may find it expressed as *volts working* (i.e. 8 μF 450 V WKG).

Many smaller capacitors have their properties marked on them in a colour code, like resistors. **Figure 2** shows these codes, and their meaning, and the table below summarises the values of the colours.

Table 1

Colour	*Value*	*Voltage (tantalum capacitor)*	*Voltage (polyester capacitor)*
Black	0	10	–
Brown	1	–	100
Red	2	–	250
Orange	3	–	–
Yellow	4	6.3	400
Green	5	16	–
Blue	6	20	–
Violet	7	–	–
Grey	8	25	–
White	9	3	–

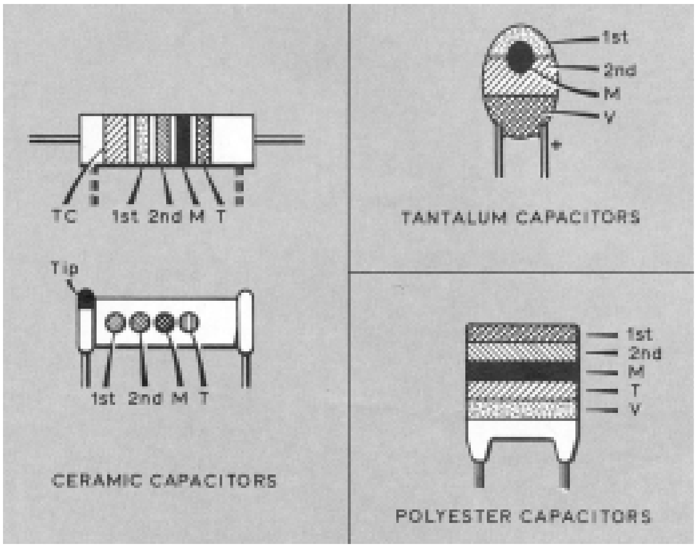
TC
1st
2nd
M
T
1st
2nd
M
V
TANTALUM CAPACITORS
Tip
1st
2nd
M
T
CERAMIC CAPACITORS
1st
2nd
M
T
V
POLYESTER CAPACITORS

Figure 2 Some capacitors have coloured bands or stripes, rather like resistors. The colour code, which is the same as the resistor code, is shown in Table 1. The band shown on the chart as '1st' is the first number of the capacitor's value in pico-Farads, '2nd' is the second number and 'M' is the Multiplier or number of noughts. For example, a capacitor reads from the top: Brown, Black, Yellow, Black, Red. Its value is One, then Nought, then Four more noughts = 100 000 pF (also referred to as 0.1 μF or 100 nF). Its tolerance (Black) is 20% and the working voltage (Red) is 250 V. The 'V' means the maximum working voltage. The band marked 'T' shows the tolerance, just like resistors, and the one marked 'TC' is only used on special capacitors designed to change their value with temperature

11 Waves – Part 2

Introduction

We left Part 1 with the concept that radio waves are divided up into bands which have different properties. Not *all* the properties are different, though. We need to discuss several wave properties, so we will start with what happens to waves as they propagate over long distances.

Getting weaker

Imagine a torch battery connected to a bulb by wires about 1 metre long. The bulb lights normally. If we now take the bulb 100 metres away from the battery and wire it up, we would expect the bulb to be somewhat dimmer, which is exactly what would happen. It happens because of the resistance of the wires – the wires do not form a perfect conductor. A similar situation occurs with radio waves.

All waves suffer from *attenuation* – they get weaker the further they travel. In cases of extreme attenuation, we need to apply some *amplification* before the attenuated wave can be used in a receiver.

Carrying information

When we speak over the telephone, the range of frequencies in our voices extends from very low frequencies up to about 15 or 20 kHz. In audio terms, this is quite a large *bandwidth* (meaning a wide band of frequencies). For communications purposes, however, most of this bandwidth is not needed, and in the telephone system (and in our transceivers), this is cut down so that it extends from about 200 Hz to 3 kHz, a reduction in bandwidth from 20 kHz to about 3 kHz. A bandwidth of 3 kHz has been found to be sufficient to convey speech intelligibly which, after all, is just what we need!

The radio waves coming from an amateur transmitter convey our speech signals over long distances. By themselves, the speech signals do not travel very far, so they have to be combined with a radio signal that *will* travel long distances. This extra signal is called the *carrier wave* (or just the *carrier*), because its job is to *carry* the speech signals along with it! The process of combining the speech (or Morse code) with the carrier is called *modulation*.

Wider and wider

Sending Morse code is achieved simply by switching (or *keying*) the carrier on and off. The bandwidth of the transmitted signal is only about 100 Hz. Speech, with its reduced bandwidth of 3 kHz, will produce a *single-sideband* (SSB) transmitted signal with a bandwidth of 3 kHz. If the same speech signal were used to produce an *amplitude-modulated* (AM) signal from the transmitter, it would have a bandwidth of about 6 kHz. Perhaps you can now understand why the bandwidths needed to produce hi-fi broadcasts need to be so large. TV signals need bandwidths running into tens of megahertz!

Waves need aerials

Radio waves are produced whenever changing currents flow through a wire, and when that wire is made in such a way as to maximise the *radiation* from the wire, it is called an *aerial* or *antenna*. The same piece of wire will receive radiation from other aerials; an aerial will transmit and receive. This is an important property of the aerial: when a current flows through it, electromagnetic waves are launched into the air; when electromagnetic waves in the air encounter the aerial, currents are produced in it.

From the simplest transmitting aerial, waves travel in all directions, like the waves on the pond that we considered in Part 1. They will travel a long way through air and space before they become too week to be received. They won't travel very far into the earth, however! The earth will *reflect* some of the wave and will *absorb* the rest. That portion of the wave which is reflected will again travel through air and space until it is totally attenuated.

Look at **Figure 1**; A represents a radio transmitter, with B1, B2 and B3 being receiving stations. The two arrows pointing 'downwards' from A represent two of the waves from A which just graze the earth's surface. Waves above

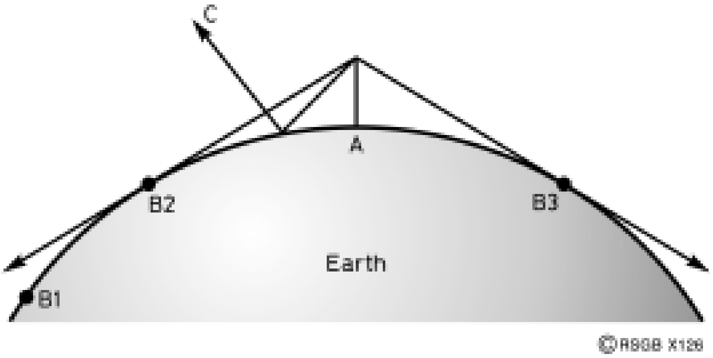

Figure 1 Stations between B2 and B3 receive the *ground wave*

these will travel on into space; waves below them will either be absorbed and reflected by the earth or received by aerials. B1 will not receive any signals from A, because it is below A's horizon. B2 and B3 can receive A's signals because they are *just* on A's horizon. Any stations between B2 and B3 will also receive A's signals, which are known as *ground-wave signals*. The wave at C represents one which is reflected by the ground and travels into space.

This description begs the question of how signals are received from stations well beyond the ground-wave range.

Mirrors in space

Suppose that there was something, out in space, that would reflect radio waves. Waves from A that travel out into space could be reflected off it and return to earth, enabling stations such as B1 to receive A's signals. The situation just described is illustrated in **Figure 2**.

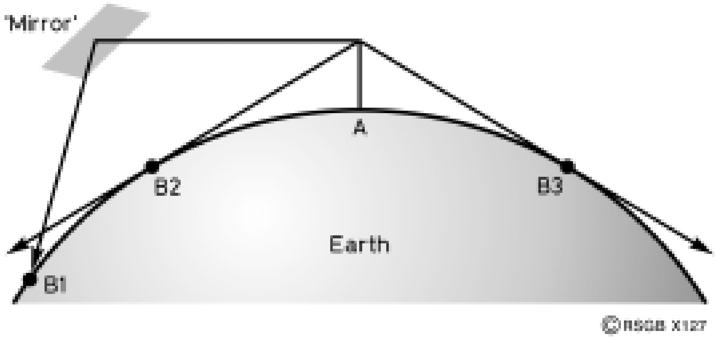

Figure 2 Reception of A's signals at B1 using the *sky wave*

Just such a mirror in the sky really *does* exist. It is not man-made, of course! The earth is surrounded by the atmosphere, a mixture of many gases, such as nitrogen and oxygen. The energy contained in the radiation from the sun is more than sufficient to *ionise* these gases, thus making them into electrical conductors. When a gas is ionised, some of its electrons are physically stripped out of the atoms and are free to move about, just as the electrons of a metal do in a wire. Consequently, we can regard this part of the atmosphere (the part illuminated by the sun) as acting like a sheet of metal, which *reflects* radio signals! It is not a perfect reflector, but is sufficient to produce long-range (DX) radio propagation via the *sky wave* under the right conditions. (A more down-to-earth example of ionised gases conducting electricity can be found in the fluorescent tube and the neon sign – many gases glow when they are continuously ionised.)

This conducting region at the extremity of the atmosphere is called the *ionosphere*, and it exists in layers between 60 km and 700 km above the earth's surface. When the ionosphere is sufficiently ionised, it glows; this is the natural phenomenon known as the *aurora borealis*, or the *northern lights*.

The property of the ionosphere that enables radio waves to be reflected does not act in a uniform way; it is very selective about which waves it reflects, and which waves go straight through it and into outer space. In general, it reflects only those waves with frequencies below about 30 MHz – the HF bands!

In Part 3 we will look at families of waves.

12 An LED flasher

Introduction

The LM3909 is an integrated circuit (IC) which will flash a light-emitting diode (LED). Using only two extra components and a battery, the circuit is cheap and has a very low current drain from a 1.5 V cell. The circuit can be used as a novelty flasher, an indicator for a dummy alarm bell box, or it could be attached to a torch so that it could be found easily in the dark! The simple circuit is shown in **Figure 1**.

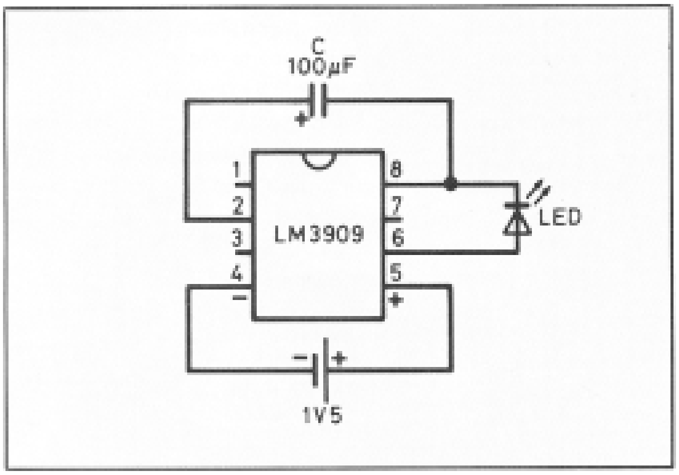

Figure 1 Circuit diagram of the LED Flasher. Pins 1, 3 and 7 of the IC are not used

Assembly

The circuit can be built on a small piece of Veroboard (the piece shown in **Figure 2** measures 15 holes by 10 strips). Using such a board, follow these instructions.

1. Depending on how far away you want the LED from the circuit board, solder a length of insulated wire to each lead of the LED. Use different colours of insulation – say, red and black, connecting the *red* lead to the *anode* (a) lead (the longer one) of the LED, and the black one to the *cathode* (k). Figure 2 shows these leads.
2. Cut the copper tracks as shown in Figure 2, using a 3 mm (⅛ inch) diameter drill, rotated between thumb and forefinger, or use the proper tool. Make *absolutely* sure that the tracks are completely broken!
3. Fit the IC holder in the correct position, using the cut tracks as guides, and make sure the small notch is facing towards the top of the board. Solder the pins to the copper tracks.
4. Mount the capacitor, positive end to the left, so that the positive lead is soldered to track F, which connects it to pin 2 of the IC; the negative lead is soldered to the right-hand side of track E, this being connected to pin 8 of the IC.
5. Solder on the battery leads, positive to the right, and the extended LED leads, positive downwards.

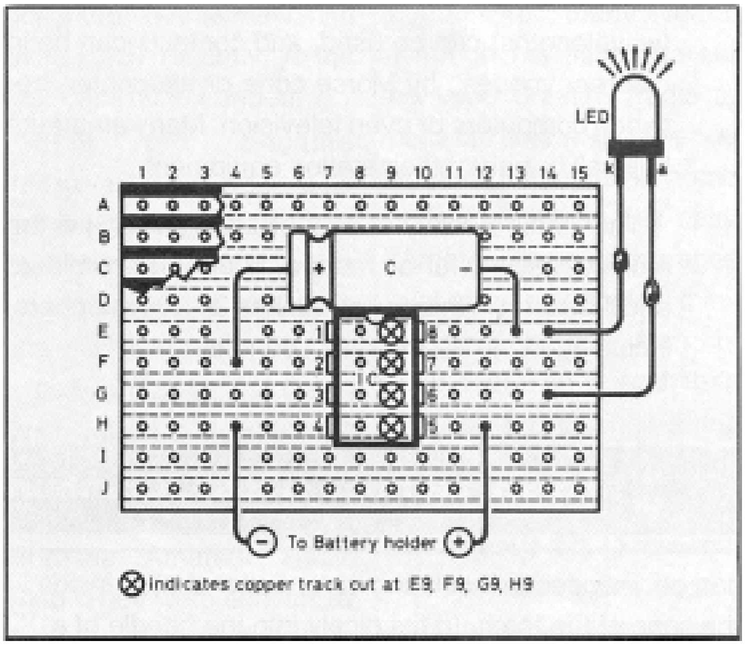

Figure 2 Board layout viewed from the component side. The tracks are cut under the board where shown

6. Check the circuit, and hold up the board to a bright light and look carefully for solder bridges between the tracks and pieces of copper swarf which may have escaped your inspection in 2 above! Remove whatever you find.
7. When all seems well, put the IC into the socket, ensuring that the notch or dot on the upper surface of the IC lines up with the notch on the holder. Line up each pin on the IC with the hole below it before pressing *gently* on the IC with the board supported on a firm surface.
8. Connect the battery; the LED should start to flash. The circuit is complete and working!

If you prefer, the whole circuit (battery included) can be mounted in a small plastic box, with the LED mounted on a clip and protruding through the panel. There are many other possibilities, and it is up to you to find an application for your own use.

Parts list

		Maplin code
LM3909	Integrated circuit	WQ39N
IC socket	8-pin DIL	BL17T
LED	5 mm diameter	WL27E
100 microfarad (μF)	Electrolytic capacitor (10 V)	FB48C
Battery holder	For AA-size cell	YR59P
Battery	1.5 V AA cell	

Small piece of Veroboard (15 holes by 10 strips)
Small plastic box (if required)
LED clip (if required)
Two lengths of coloured, insulated wire for LED (as required)

Availability

All parts can be obtained from Maplin Electronics Ltd.

13 Waves – Part 3

Introduction

When we talk about the spectrum being divided up into bands, this is just for our convenience; there are no natural divisions although, as we have seen, some of the properties of different bands really *are* different! Let's look at **Figure 1**, and see how the frequencies are divided up.

The divisions

- **Very low frequencies** (VLF) cover the range from a few kilohertz up to 30 kHz. Very long-range communication is possible, but at *very* small bandwidths. It is used for special purposes.
- **Long waves** (LW) are used for medium-distance commercial broadcasting and have frequencies from 30 kHz to 300 kHz.
- **Medium waves** (MW) are used for commercial broadcasting, and use frequencies from 300 kHz to about 1.5 MHz (1500 kHz). Typical range is about 200 km.
- **Short waves** (SW) encompass both the low-frequency (LF) and high-frequency (HF) amateur radio bands. There are nine narrow amateur bands in the SW spectrum between 1.8 MHz and 30 MHz. Some of these bands give round-the-world communication.
- **Very high frequencies** (VHF) span the range between 30 MHz and 300 MHz. Relatively short-range communication is possible. They were once used for broadcast TV before it moved to UHF. There are now three

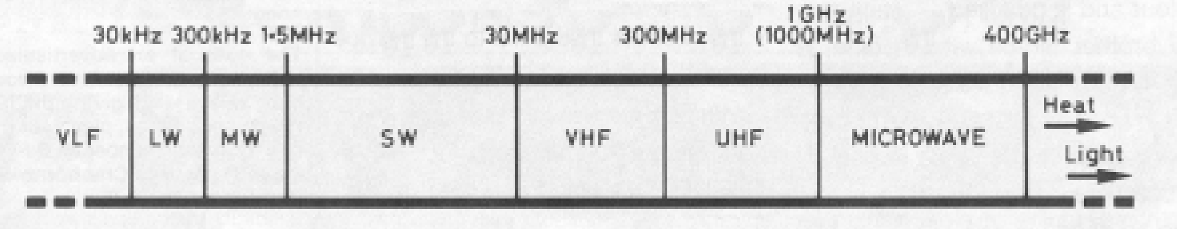

Figure 1 Diagram of radio frequency spectrum

amateur bands here – 6 m, 4 m, and 2 m. Repeaters are used to extend the usable range of mobile stations. VHF waves are not usually reflected by the ionosphere, but when they are, ranges of several thousand kilometres are possible. Weather affects these waves on a regular basis, however. In addition to amateur users, the VHF part of the spectrum is also used by the police, the fire and ambulance services, weather satellites, and many others,

- **Ultra-high frequencies**, sometimes called *centimetre waves*, cover the range from 300 MHz to 1000 MHz (or 1 GHz). The only amateur band in this range is the 70 cm band, and we share it with radar, TV and cellular telephone users as well.
- **Microwaves** begin at 1 GHz and extend to about 400 GHz. They are *never* reflected by the ionosphere, are partially attenuated by buildings, and are reflected from aircraft and cars. Microwave absorption in the atmosphere is quite significant, and rain and fog can attenuate microwaves quite heavily.
- **Heat, light** . . . Above 400 GHz we run into the infra-red bands and on into the visible light and ultra-violet bands. We generally take 400 GHz as being the limit of what we class as radio waves.

Bandwidth again

Complex signals need more bandwidth than simple signals. Even if it were possible, we would not be able to transmit a single television channel in the whole of the MW broadcast band! When TV used part of the VHF band, only five channels were possible in the range from 45 MHz to 68 MHz. By moving TV to the UHF band, we now have 47 channels between 470 MHz and 855 MHz!

It’s your choice!

The number of permutations you have amongst all the modes and all the bands is enormous! Only you can decide what you are interested in and what you want to learn about. That is the attraction of amateur radio!

14 Choosing a switch

Simple, but be careful!

A switch is the simplest electronic component, but care is needed in choosing the correct one for the job. Here is a list of the main characteristics of a switch, which will help you to select what you want.

1. **Rating**. This gives the maximum voltage and current that a switch can handle. For example, a 250 V 1.5 A switch will switch mains voltage at a current not exceeding 1.5 A. If the current is greater than 1.5 A, the switch may get hot and fail. If the voltage is too great, the switch may arc each time it is switched off, thus wearing away the contacts.
2. **Number of poles**. A single switch can control many circuits; the number of poles tells you *how many* different circuits it can handle. See **Figure 1**.
3. **Number of throws**. This tells you the number of positions each pole can have. This is best illustrated in Figure 1. The simplest ON/OFF switch is a Single-Pole, Single-Throw (SPST) switch. A Single-Pole,

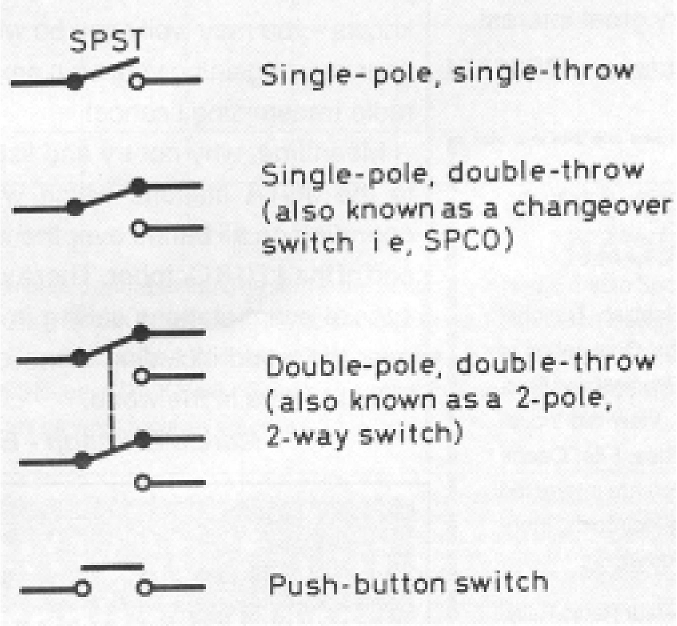
SPST
Single-pole, single-throw
Single-pole, double-throw
(also known as a changeover
switch i e, SPCO)
Double-pole, double-throw
(also known as a 2-pole,
2-way switch)
Push-button switch

Figure 1 Some circuit symbols for different types of switch

Double-Throw (SPDT) switch may also be used as an ON/OFF switch, but is used mainly to change between two parts of a circuit, and is commonly known as a Single-Pole Change-Over (SPCO) switch. Two or more SPCO switches can be operated at once; Figure 1 shows an example. The two switches are said to be *ganged*. See also 'Types of switch' below.

4. **Number of ways.** When a switch has more than one throw, we tend to use the word 'ways' instead. This means that if a switch has one pole and six throws, we would normally call it a '1-pole 6-way' switch. Such switches tend to be rotary switches, as are described below.

Types of switch

(a) **Push-button**. These are found on calculators, telephones, electronic games and most equipment with a digital display.

(b) **Rotary**. These are switches controlled by a knob, and are turned instead of moved up and down. **Figure 2** shows the rear of such a switch and its circuit symbol. The commoner types of rotary switch are: 1-pole, 12-way; 2-pole, 6-way; 3-pole, 4-way; 4-pole, 3-way; 6-pole, 2-way. All these switches have 12 click positions, as you may have guessed, but each one comes with an adjustable end-stop so that you can set the correct number of ways according to the contacts on the switch.

(c) **Slide**. This switch is common on the cheaper types of radio, mainly as an ON/OFF or band-changing switch. They are not very rugged, but are small and cheap to produce. Very small types are manufactured for use on PCBs.

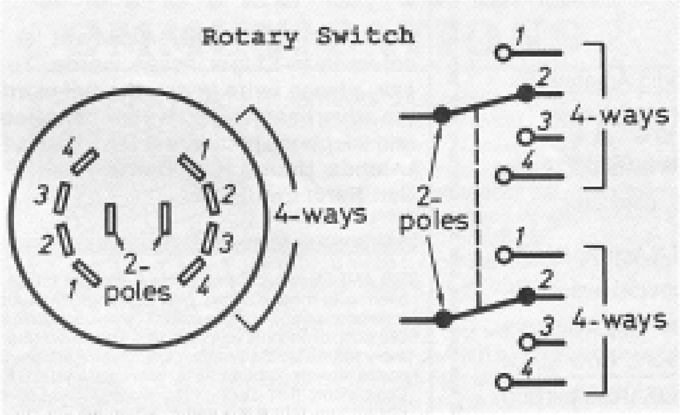

Figure 2 The rotary switch can select several different circuits at once

(d) **Toggle**. Available as two-way types or three-way, with a centre-off position. Can be manufactured to handle very high currents.

(e) **Micro-switch**. These are devices which are usually operated indirectly, such as when the cover is removed from a high-voltage power supply, or when the door of a fridge is opened. The 'micro' part of the name doesn't refer to the size of the switch, but to the *small* movement that is required to activate it.

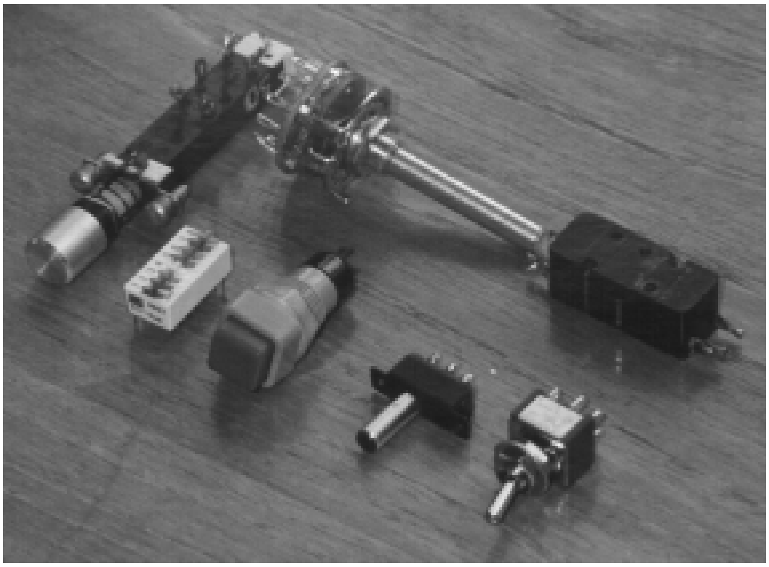

15 An aerial tuning unit for a receiver

Introduction

Any length of wire will act as an antenna (or aerial) but, to get the best results from a transceiver or receiver, an aerial tuning unit (ATU) is required. This matches the *impedance* of your aerial to the impedance at the aerial socket of your radio. Impedance is like resistance, and is measured in ohms, but it is used for alternating currents, and hence is common in audio and RF engineering. Most receivers have an impedance at the aerial socket of about 50 ohms (Ω); aerial impedance, on the other hand, can be anywhere between 20 Ω to over 1000 Ω, depending on its length and its height above ground.

Fortunately, you don't *need* to be able to calculate your aerial's impedance; all you need is a device that will perform the matching operation for aerials with a large range of impedances, and this is exactly what is described here! The subject of *matching* is a complex one, but all you need to know is that most signals will become clearer, and that there will be less noise and interference. Stations will become louder, so you will probably be able to reduce the setting of your RF gain control (always a good thing to do).

This design of ATU covers all amateur and broadcast bands from 10 m (28 MHz) to 80 m (3.5 MHz), and is very easy to build. The circuit is shown in **Figure 1**.

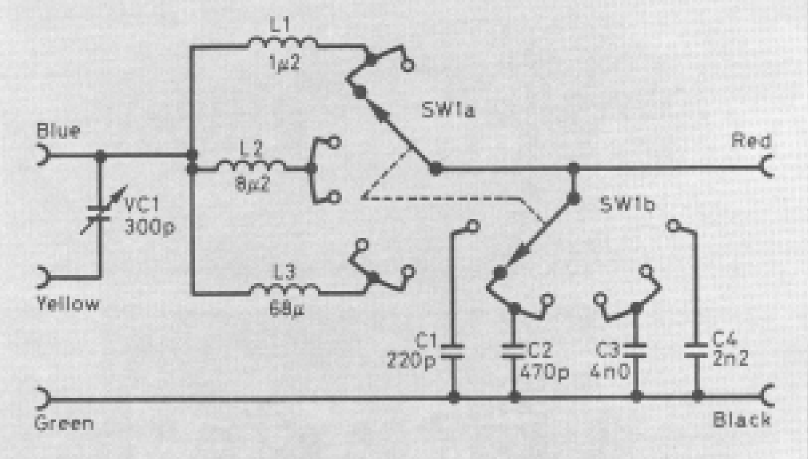

Figure 1 Circuit diagram of the Antenna Tuning Unit, showing the use of a 2-pole, 6-way rotary switch to select inductors (L) and capacitors (C)

Construction

1. Firstly, you will need a simple plastic case in which to house the ATU. The size should be approximately 85 × 145 × 50 mm.
2. Start by drilling two 10.5 mm holes in the front of the case; these are for the 6-way switch and the tuning capacitor.
3. Drill three 8 mm diameter holes in the left-hand side of the box, for the three sockets, coloured blue, yellow and green.
4. On the right-hand side of the box, drill two 8 mm diameter holes for the red and black sockets.
5. Now, fit the 6-way switch (SW1), the tuning capacitor (VC1), and all the sockets to the case. Check that the vanes of the capacitor rotate smoothly when the shaft is turned.
6. Wire up the inductors (coils). Figure 1 and the wiring diagram of **Figure 2**, will help with this. As you can see, each one side of each coil is connected to two switch connections, the other end going to VC1.
7. Solder in the fixed capacitors. One end of each goes directly to the ground socket (black), and the other end goes to the switch.
8. Solder a wire between the green and black sockets. The output from the ATU comes from the red socket, and this is connected to the two tags in the centre of the switch, as Figure. 2 shows clearly.
9. Finally, connect the blue and yellow sockets to the tuning capacitor, and the ATU construction is complete.

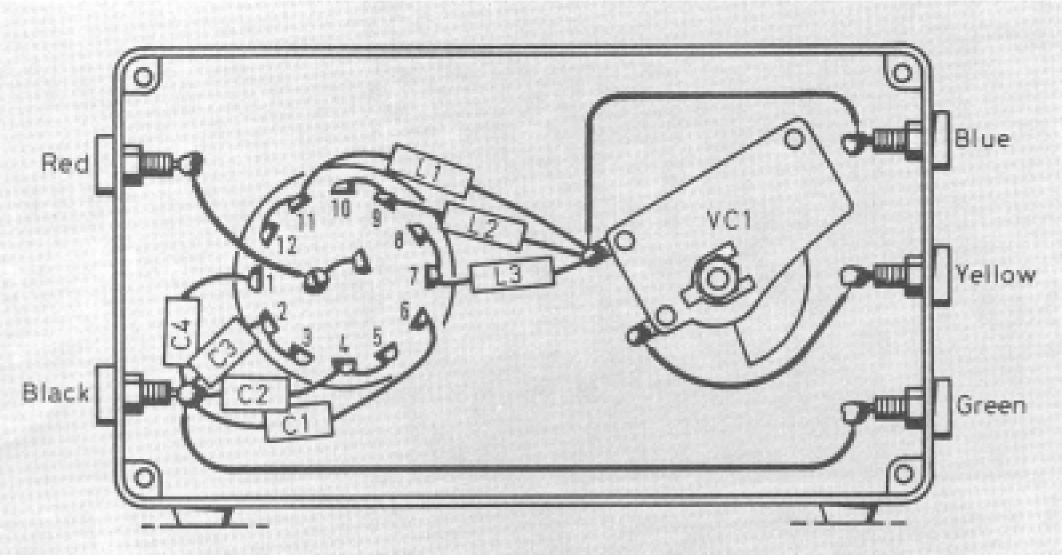

Figure 2 The internal view of the case shows the main tuning capacitor, VC1. This is a solid dielectric type, which has adjustable brass plates

In practice . . .

Figures 3 and **4** show two different ways of connecting your aerial to your ATU. In each case you will need to select each switch position in turn, and rotate the tuning capacitor through its full range while listening to a station. You should find that one switch position enables VC1 to produce a peak in the signal strength in the loudspeaker. At this point, your aerial and receiver are said to be *matched*. Stations in the same band will probably peak with VC1 at the same setting of SW1, but different bands will almost certainly require different positions of SW1.

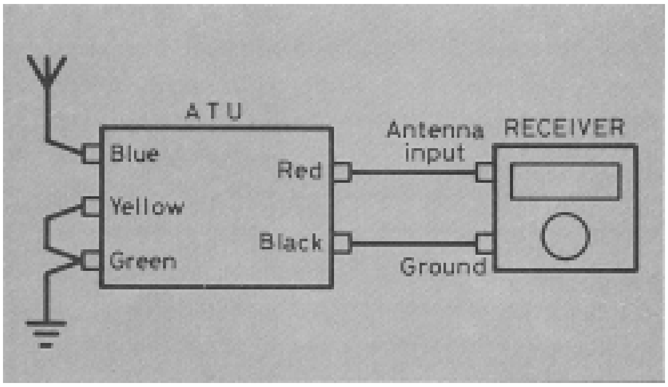
ATU
Blue
Yellow
Green
Red
Black
Antenna input
RECEIVER
Ground

Figure 3 You may find that parallel tuning gives best results with your antenna

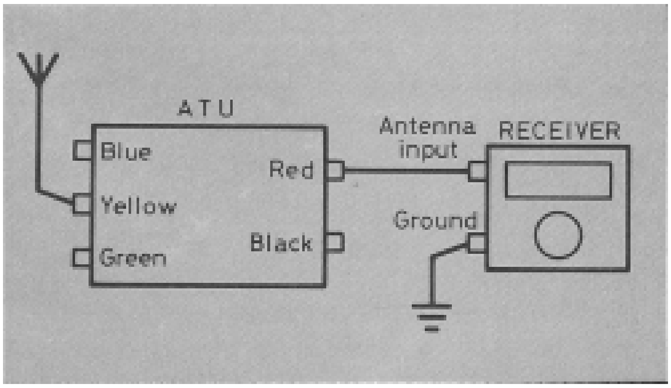
ATU
Blue
Yellow
Green
Red
Black
Antenna input
RECEIVER
Ground

Figure 4 In other cases, series tuning could be the most effective arrangement

Parts list

Capacitors (all rated at 16 V or more)	
C1	220 picofarads (pF) polystyrene
C2	470 picofarads (pF) polystyrene
C3	1000 picofarads (pF) or 1 nanofarad (nF) polystyrene
C4	2200 picofarads (pF) or 2.2 nanofarads (nF) polystyrene
Inductors	
L1	1.2 microhenries (μH)
L2	8.2 microhenries (μH)
L3	68 microhenries (μH)
Switch	
SW1	2-pole 6-way rotary
Sockets	
	4 mm type, one each of red, black, yellow, blue, green
Additional items	
	Plastic or metal case, e.g. Maplin type YU54
	Two large knobs for SW1 and VC1

16 A simple 2 m receiver preamplifier

Introduction

Designed specifically to complement the modified air-band portable (also described in this series), this can be used with some success on many receivers suffering from 'deafness' on VHF.

The circuit

An RF preamplifier is a device which improves the input signal to an existing receiver, enabling it to work more effectively. Because of the noise which is added to the signal by the preamplifier, *very* weak signals may not be usefully enhanced, but stronger signals will be improved considerably.

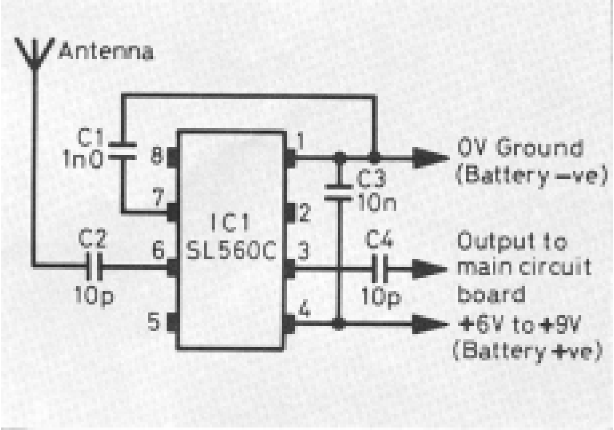

Figure 1 The integrated circuit is mounted upside down. Make sure you identify pin 1 which has a small dot next to it

This little circuit (shown in **Figure 1**) uses a GEC/Plessey integrated circuit type SL560C. With the addition of four capacitors, it is used between your aerial and the aerial input of the radio's PCB.

Putting it together

1. Use a small piece of prototype (matrix) board about 25 mm square. Use an 8-pin DIL socket for the integrated circuit (don't risk soldering the chip – it is seldom a risk worth taking). Figure 1 shows the connections to the socket, *looking from underneath*.
2. Make special note of the pin numbers, so that you know how to put the chip into the socket when you have finished. The positive and negative connections to the circuit are taken from the main PCB *after* the ON/OFF switch – so that the switch operates the preamplifier, too.
3. Unsolder the lead to the radio's telescopic aerial and connect it to the free end of C4, as shown. Then solder a short lead between the telescopic aerial and C2.
4. Sometimes it is possible to cajole your little preamplifier into the radio's plastic case, provided there is room and that you make sure that none of the soldered joints on your little board touch any of the metal inside the case when you replace the back and screw it on again.
5. If there is not enough room inside, then put the preamplifier into its own box, with battery and switch, and its own aerial. Keep the lead from the preamplifier to the aerial connection of the radio as short as possible – perhaps using screened cable.

Because your preamplifier is untuned, you will find not only that it helps with reception on 2 metres, but also that reception on the FM broadcast band is improved!

Parts list

Integrated circuit

IC1	GEC/Plessey SL560C

Capacitors

C1	1 nanofarad (nF) ceramic
C2, C4	10 picofarads (pF) ceramic
C3	10 nanofarads (nF) ceramic

Additional item

Prototype broad approx. 25 × 25 mm

17 Receiving aerials for amateur radio

Introduction

For any radio receiver to work well, it must have some form of antenna, or aerial. In almost all domestic transistor radios, the aerial is built into the set, either as a ferrite rod (which looks like a rod of dark grey metal) on which are wound coils of wire, or as a chromium-plated telescopic metal rod. Some radios have both forms of aerial, using the ferrite rod aerial for long waves (LW) and medium waves (MW) and the metal rod for very high-frequency (VHF) stations using frequency modulation.

There's broadcast reception . . .

No aerial is perfect, and these two types are *far* from perfect! As in most mass-produced equipment, *they serve their purpose*, which is not critical, and they allow the radio to be carried around easily, because they are not big and bulky. The broadcast stations are very powerful, usually quite close, and the circuits in the radio are quite sensitive, so the need for large aerials disappears.

. . . and there's amateur DX

Here we have an application which, in comparison with the broadcast situation, couldn't be much more different. The stations do not use much power, they may be half the world away, and the requirement for good aerials and very sensitive receivers is paramount. The aerials must be large and they must be as high as possible, away from buildings and trees, which cause reduction in signal strength, and away from man-made sources of interference.

You may not have thought about this but, in general, the larger an aerial becomes, the longer the cable (or 'feeder') must be in order to reach your shack and the receiver inside it. Cables reduce the received signal, so what your aerial gains by being large, the feeder (if you're not careful) will lose!

A simple aerial . . .

We now know that a 'good' aerial is essential. But what is a 'good' aerial? It depends on your purse, your property and your enthusiasm! One of the simplest (and, incidentally, one which is *not* subject to the cable loss problem discussed above) is the **Long Wire**, shown in **Figure 1**. This is, quite literally, a long piece of wire going from a chimney stack to a tree or pole

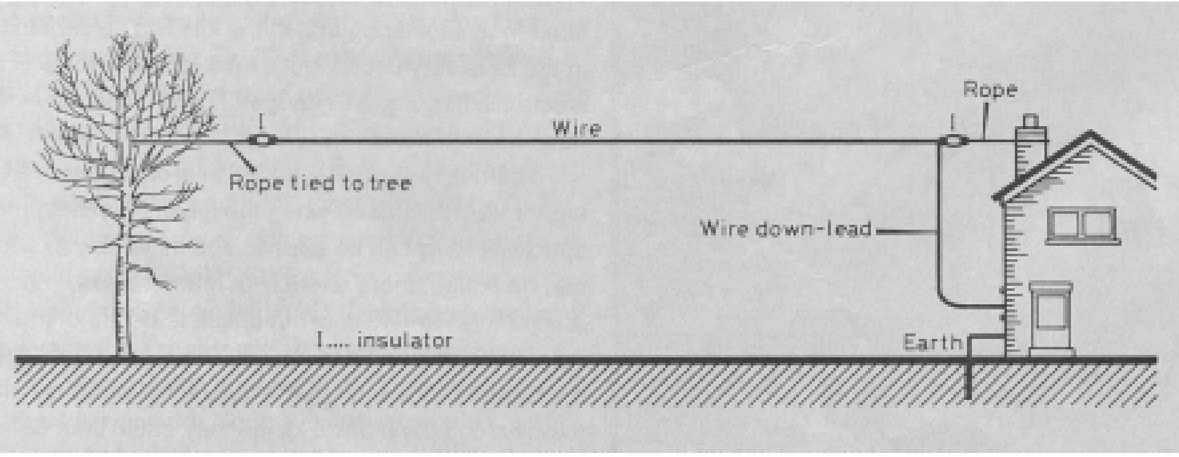

Figure 1 A tree at the bottom of the garden can provide a useful antenna support. Insulators can usually be obtained from Tandy stores or TV aerial suppliers

at the end of the garden. The longer it is, the better. Notice that the wire itself is *not* used to loop around the tree or chimney. Rope is used for both, and is secured to the end of the wire using an insulator of the ‘dog bone’ or ‘egg’ variety, to be found in profusion at rallies.

The wire is brought into the house through a window into the shack where the receiver is situated. The long wire aerial is best used with an aerial tuning unit (ATU), which is also described in this book. One of the advantages of the long wire is that it can be used on several frequency bands.

. . . and more complicated ones

Next up the ladder of complexity is the **dipole** (meaning two poles, or two elements). One form of dipole for lower frequencies would take the same basic form as the long wire, except that the feed to the receiver is taken, not from the end, but from the centre. The wire is essentially cut into two halves, and the two ends at the centre are connected to one end of a coaxial cable, which is then taken to the receiver. A smaller form of this is shown in **Figure 2**, which is conveniently mounted in a house loft. Aerials should always be mounted outside for best results, but will work when mounted inside, and the loft space is the logical situation.

The total length of the dipole should ideally be one-half of the wavelength of operation – hence the term ‘half-wave dipole’. For example, a dipole for use on 20 m should be about 10 m long. Dipoles are thus ‘single-band’ aerials. They can be modified for use on several bands, and then become known as ‘trapped dipoles’, having coils and capacitors at certain points along their lengths.

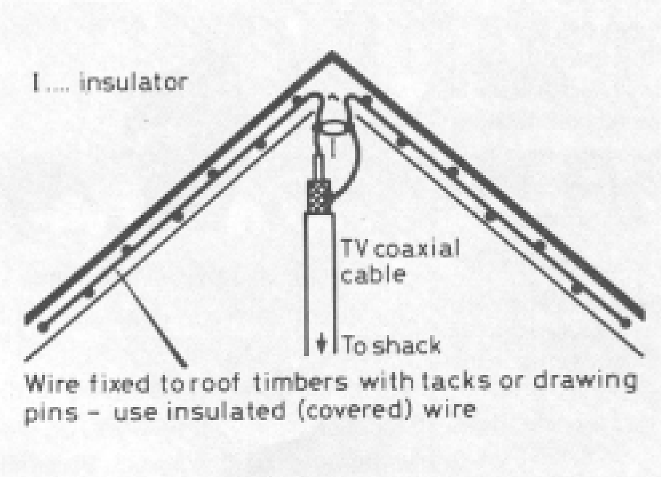
I.... insulator
I
TV coaxial cable
To shack
Wire fixed to roof timbers with tacks or drawing pins – use insulated (covered) wire

Figure 2 An indoor antenna should be mounted as high as possible

Indoor aerials

For flat- or apartment-dwellers, an indoor aerial is often the only solution, in which case the dipole of Figure 2 may be of great use. Other popular types include the **loop aerial**, of which several types are shown in **Figure 3**. The shape of the loop is of secondary importance, but it should be as large as possible.

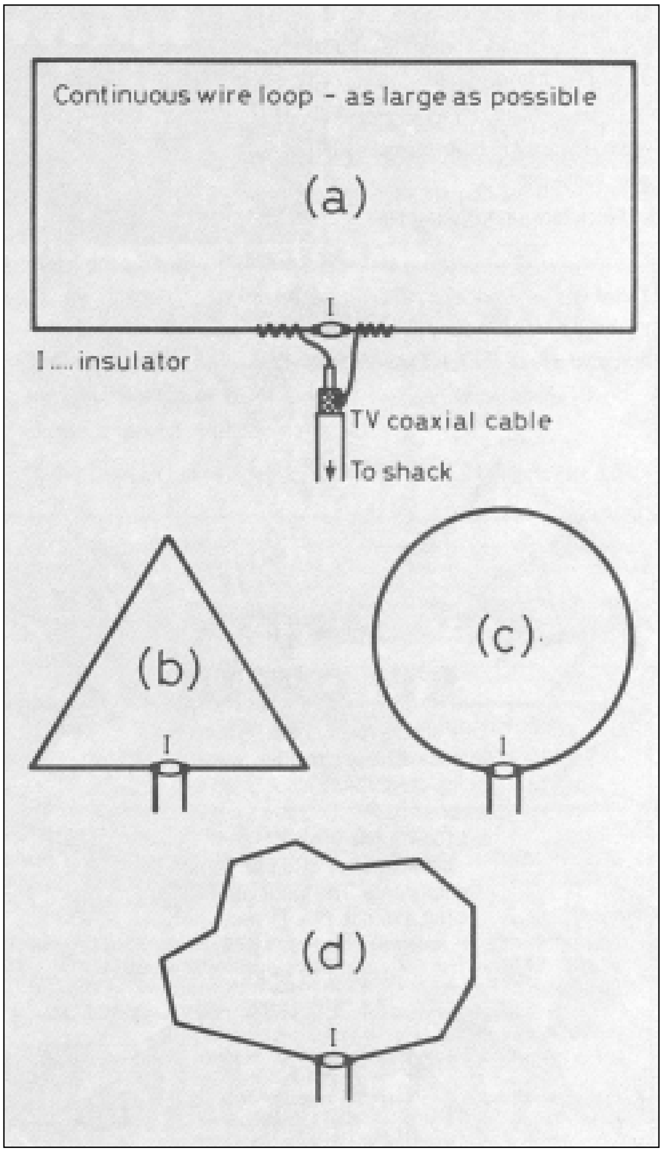
Continuous wire loop – as large as possible
(a)
I
I.... insulator
TV coaxial cable
To shack
(b)
I
(c)
I
(d)
I

Figure 3 Try different shapes of loop antenna to see which works best

Ingenuity is called for when considering indoor aerials. A dipole can be mounted along a picture rail or pelmet. A vertical loop of aluminium foil can be stuck to an inside wall and covered with wallpaper! The designs are limited only by your imagination, and will give you hours of listening pleasure

Outdoor aerials

If you want to venture beyond the simple long-wire discussed earlier, then the erection of outdoor aerials can be a bit tricky; they can be more than a one-man job, and can involve seeking planning permission. Get some help advice from someone who has done it before – what about your local radio club? Plenty of designs for aerials appear in *RadCom*, so you should never be short of ideas!

18 The Colt 80 m receiver – Part 1

Introduction

This is the first of four parts detailing the design and construction of a simple radio capable of good reception of amateur radio signals on the 80 metre band. If you can solder and have some basic hand tools, at the end of Part 4 you will have a working receiver, in which you can take pride, and start some serious listening on 80 m! For testing, all you will need is another receiver and a multimeter.

Description

The radio will be built in three modules, or stages, as illustrated in **Figure 1**. Each of these is built on a printed-circuit board (PCB) or matrix board and can be tested in its own right. The case will need some holes drilled, and care will be needed to mount the PCBs in the case.

Get to it!

Surprisingly enough, when you are building a radio section by section (which is *always* the best way), it is easiest to start at the output and work

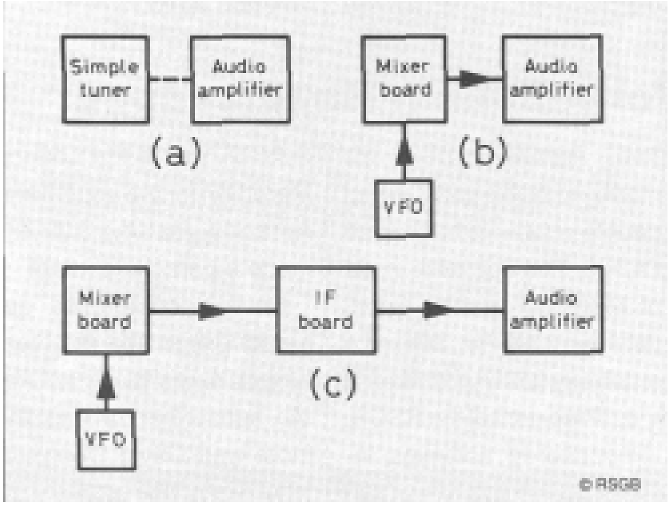
Simple tuner
Audio amplifier
(a)
Mixer board
Audio amplifier
(b)
VFO
Mixer board
IF board
Audio amplifier
(c)
VFO
© RSGB

Figure 1 Build your Colt and watch it grow. A simple crystal set (a) becomes a direct conversion receiver (b) and finally an 80 metre amateur band superhet (c)

towards the input, and you can test what you have done stage by stage. You will see what this means as you progress with your construction.

Every receiver needs some audio frequency (AF) amplification to make the sound signals big enough for you to hear. The circuit for the AF amplifier is shown in **Figure 2**. It uses the TDA7052 integrated circuit (IC) plus a handful of extra components. R1 in conjunction with C2 and C3 *decouple* the battery supply, preventing any audio signals getting through to it and affecting other parts of the radio, when they are connected. C1 acts to prevent high frequencies (above the 3 kHz bandwidth) going into the amplifier input. A volume control, VR1, is connected across the amplifier input, so that the amplifier can accept signal inputs over a wide range. **Figure 3c** shows the connections, which are made with screened cable. The

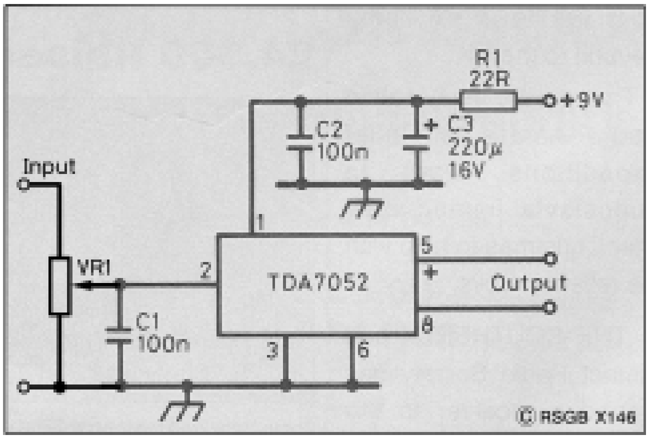
R1
22R
+9V
C2
100n
C3
220µ
16V
Input
1
VR1
2
TDA7052
5
+
Output
8
C1
100n
3
6
© RSGB X146

Figure 2 The Philips TDA7052 integrated circuit (IC) used in the audio amplifier needs very few extra components. It has a signal voltage gain of 100 times and the output is suitable for a loudspeaker or headphones

centre conductor of the left lead goes to the point marked 'input' in Figure 2, the braid being connected to the amplifier earth (0 V) tag. The right lead to VR1 goes to whatever signal source you have for testing – see later. VR1 is *not* mounted on the PCB.

The circuit is constructed on a small PCB or matrix board. Make sure that the electrolytic capacitor, C3, is soldered into the board the correct way – its positive and negative connections are shown in **Figure 3b** for reference. If you are **at all** concerned about soldering the IC into the board, enlist some help, *or*, obtain an 8-pin DIL socket, which you can solder in and then carefully insert the chip into the socket, making sure that is the correct way round. The markings on the chip are shown in **Figure 3a**.

The output leads from the amplifier go to a plastic 6.3 mm (¼ inch) mono jack socket, so that *neither* output lead is connected to the metal case. The amplifier will drive a pair of headphones or a small 8 Ω loudspeaker.

The battery leads must be the right way round also; the battery itself can be a PP3 or PP9, or you can use a small DC power supply.

Testing

First, check that all the components are in the right places, that your soldering is good, and that you have headphones or a loudspeaker connected. Set VR1 about halfway along its travel. Connect the battery.

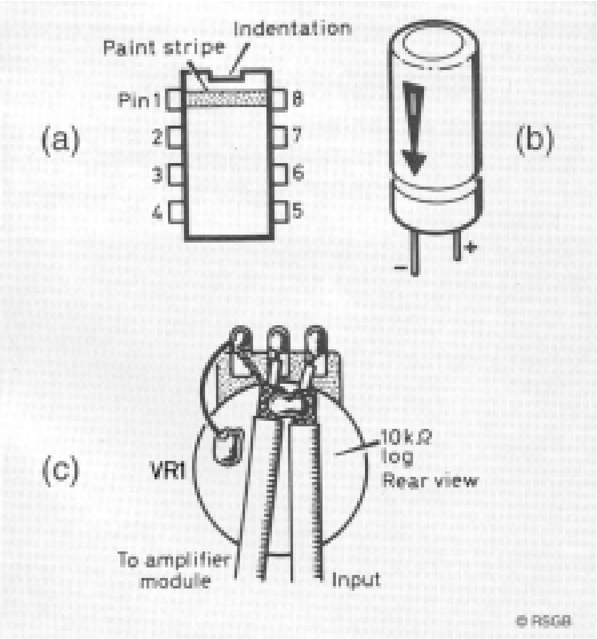

Figure 3 It is important to check the component connections carefully. The diagram shows (a) top view of the IC, (b) electrolytic capacitor and (c) volume control (VR1) connected across the input of the amplifier

A slight hissing noise should be heard; touching the input lead to the amplifier (the centre of the three connections on VR1) should produce a loud buzz. Touching the shaft at the same time will make the buzz quieter. This is the quickest way of confirming that your amplifier *seems* to be working. The only real test is to give it something meaningful to amplify! See the design of our *Crystal Radio Receiver* for full details.

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1	22 ohms (Ω)
VR1	10 kilohms (kΩ) log

Capacitors

C1, C2	0.1 microfarad (μF)
C3	220 microfarads (μF) electrolytic 16 V

Integrated circuit

IC1	TDA7052 audio amplifier

Additional item

PCB	(see below)

Component suppliers:
Maplin

The next part . . .

The metal case will be marked out ready to receive the completed modules.

19 A crystal radio receiver

Introduction

Here's a quick project to fill in a winter's evening! It was originally designed as a piece of test gear for the *Colt* receiver, but can be successfully used by anyone as a first radio project. Another use for it would be as test gear for any audio amplifier project requiring an audio input for test purposes; this is ideal, as it does not require any power supply!

Details

The initial rough-and-ready test for the audio amplifier of the *Colt* receiver will have whetted your appetite; you will want to prove more conclusively that your amplifier works, and in a way that others in the household will appreciate. Buzzing noises are not convincing in this respect! You are going to put together a simple crystal set – the simplest type of radio that there is – and use it as a signal injector for your amplifier. In this way, you build a real medium-wave (MW) receiver which drives a loudspeaker, as an intermediate product of the construction of an 80 metre amateur band receiver!

The circuit diagram is shown in **Figure 1**. It has a 60-turn coil mounted on a small piece of paper or card wrapped round a ferrite rod. The coil has a connection made to its centre-tap (the middle turn of the coil). The tuning capacitor, VC1, is the most expensive part of the circuit but don't worry, it will be used in the final design of the receiver also! Connect the diode, D1, from the centre-tap to the input to the potentiometer of the amplifier circuit of the *Colt*. Be careful to connect the aerial to the *vanes* of VC1, and not to its *frame*, or you will experience some strange effects when you are tuning. If your amplifier is working correctly, you should be able to receive local stations on medium-waves quite well.

If you think the crystal set will be of use to you in the future as a signal injector, all you will need will be another variable capacitor! If you do not intend to use the amplifier, a small crystal earpiece will allow you to listen. Walkman-type headphones will *not* work!

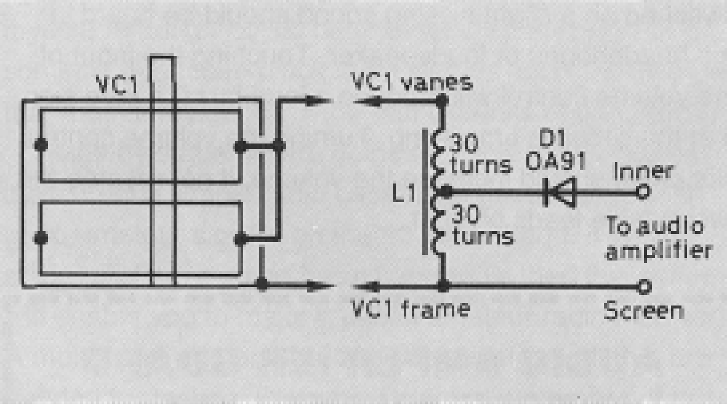
VC1
VC1 vanes
30 turns
D1 OA91
Inner
L1
30 turns
To audio amplifier
VC1 frame
Screen

Figure 1 Both sets of moving vanes are joined as shown. L1 is wound on a ferrite rod with 32 SWG wire and centre-tapped. A single winding (no tap) can be used, joining D1 to the top of the winding

Parts list

Variable capacitor	
VC1	125 + 125 picofarads (pF) twin-gang
Ferrite rod	
About 100 mm to 140 mm long	
Diode	
D1	OA91 germanium diode
Additional items	
Wire	2 m of enamelled copper wire, between 23 SWG and 32 SWG
Earpiece	High-impedance crystal type (only needed if you are not using the amplifier)

20 The varactor (or varicap) diode

Introduction

Many of the circuits for receivers and transmitters presented in this series rely upon the variable capacitor as a means of tuning. Another method of varying capacitance (without any moving parts) is provided by the *varactor diode*, sometimes called a varicap diode. This is a component which changes its capacitance as the voltage across it is varied.

The details

Figure 1 shows how a varactor diode might be connected to demonstrate its operation. Its symbol is that of an ordinary diode, with a capacitor symbol next to it. A variable voltage is applied across it in such a way that the diode is *reverse-biased*. This means that virtually no current passes through it – the positive voltage is applied to the cathode. Varactors are cheaper than variable capacitors, and they are tiny in comparison, very suitable for today's miniature circuits. If A and B were connected across the tuning coil in a simple receiver (with a series capacitor to block the DC from the battery reaching the coil), the tuning operation would be accomplished by turning the knob on the 10 kilohm potentiometer.

Varactors are available with different values, from less than 20 picofarad (pF) for VHF applications to 500 pF for medium-wave radios. They are

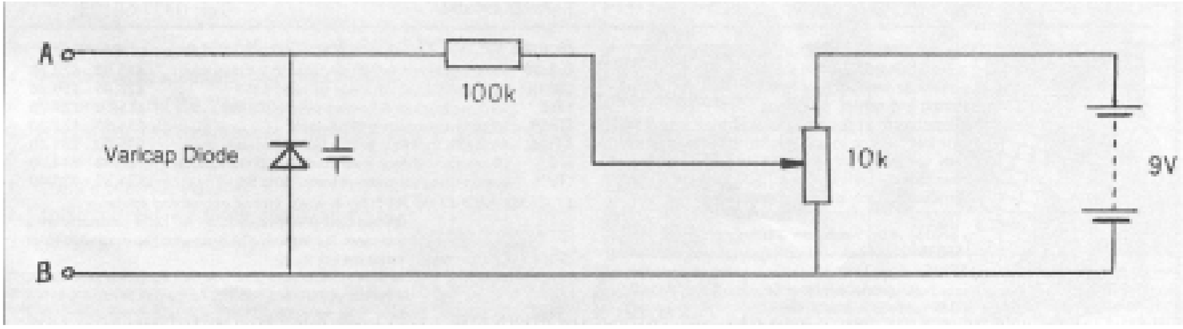

Figure 1 The capacitance of the varicap diode (between A and B) increases as the voltage is reduced, using the variable resistor

tuned usually by voltages between 2 V and 9 V. For a real application of varactors, you should consult the circuit diagram of the *Yearling* 20 metre receiver, elsewhere in this book.

In some circuit designs, several circuits are all tuned to the same frequency in order to improve the overall *selectivity* (the ability of the circuit to reject signals very close in frequency to the wanted signal). Special dual- and triple-varactors are available for circuits like this. Having been made at the same time from the same materials makes their individual characteristics virtually identical. Like all other diodes, they must be correctly wired into the circuit – their *polarity* is important.

Changes in temperature will cause the capacitance to change which, if it were part of an oscillator circuit, would cause the oscillator to drift – you would have to keep retuning the radio! This can be corrected by using a special integrated circuit called a *phase-locked loop* (PLL). Modern TV sets and satellite receivers use varactors and PLLs in this way.

Some useful varactor types

Type No.	*Tuning range*		*Description*
	pF/V	*pF/V*	
BB204B	42/2.0	15/12	Dual VHF
BB212	560/0.5	22/8	AM tuning
KV1235	450/2.0	30/8.5	Triple AM
KV1236	450/2.0	30/8.5	Dual AM
MV1404	120/2.0	9/10.0	HF tuning

21 A portable radio for medium waves

Introduction

The ZN415E integrated circuit (IC) can be used to make a very efficient AM portable MW broadcast radio with a built-in loudspeaker. Here's how!

The circuit

Figure 1 shows the circuit diagram of the portable radio. It's not as complicated as it may appear, especially after you have got started. L1 is a coil of wire mounted on a ferrite rod, acting as an aerial; VC1 is a variable capacitor which works, with L1, to tune in different stations. IC1 contains circuits of its own which boost the selected signal and it includes a *detector* which extracts the audio signal from the incoming RF signal. Earphones could be connected to the output of IC1 (between pins 4 and 5), but the output would not be powerful enough to drive a loudspeaker.

More sound

This is where IC2, an LM386 comes in. This is a small audio *power* amplifier which produces audio signals with enough power to drive a small

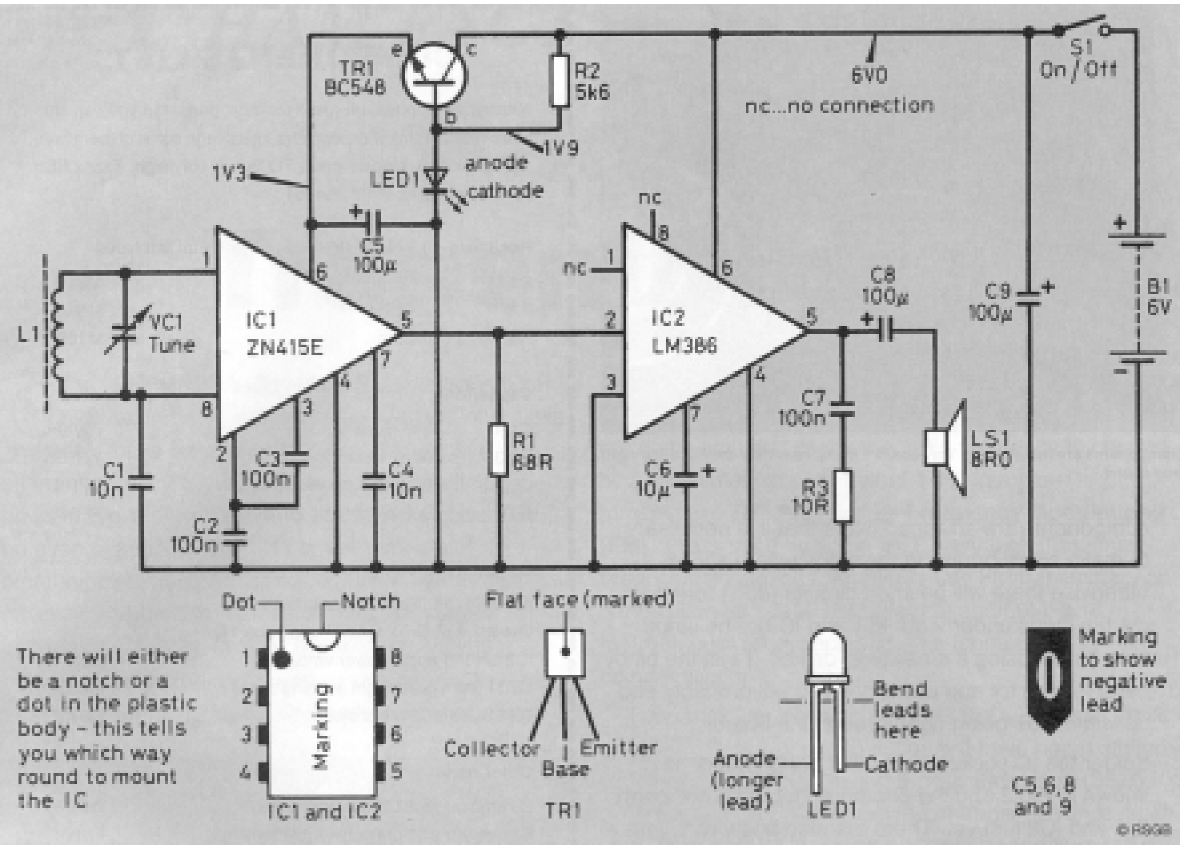

TR1
BC548
e
c
b
R2
5k6
6V0
nc...no connection
S1
On / Off
1V9
1V3
LED1
anode
cathode
C5
100µ
nc
nc
IC1
ZN415E
IC2
LM386
C8
100µ
C9
100µ
B1
6V
L1
VC1
Tune
R1
68R
C7
100n
LS1
8R0
C1
10n
C3
100n
C4
10n
C6
10µ
R3
10R
C2
100n
Dot
Notch
Flat face (marked)
There will either be a notch or a dot in the plastic body – this tells you which way round to mount the IC
Marking
IC1 and IC2
Collector
Base
Emitter
TR1
Bend leads here
Anode (longer lead)
Cathode
LED1
Marking to show negative lead
C5, 6, 8 and 9
© RSGB

Figure 1 The circuit diagram of our easy-to-build portable radio. Take care to mount the ICs and LED the correct way round

loudspeaker, LS1. The radio uses a 6 volt battery, which is made by connecting four 1.5 volt AA cells in series (4 × 1.5 V = 6 V) using a battery holder designed for this purpose. Although a 6 V supply is ideal for IC2, it is far too great for IC1, which needs only about 1.3 V. This lower voltage is provided from the 6 V supply by TR1 (an npn transistor), R2 and LED1 (a light-emitting diode). When current passes through an LED (see the description of the LED in this series) a reasonably constant voltage of 1.9 V appears between the anode and the cathode. Because of the voltage (0.6 V) that *always* exists between the base and emitter of a working transistor, the voltage on the emitter is about 1.9 V – 0.6 V = 1.3 V, and this is used as the power supply for IC1.

To keep the radio as simple as possible, no volume control has been fitted. Instead, you can use the directional properties of the ferrite rod aerial (see the information on ferrites in this book) to reduce the volume by rotating the set about a vertical axis using the handle provided.

Putting it all together

1. Start by covering the ferrite rod with Sellotape, or alternatively wrap a piece of paper tightly around it, and secure it with Sellotape. Then, with at least 2 metres of 24 SWG enamelled copper wire, wind 75 turns tightly around the rod. To be safe, leave about 50 mm of wire at the ends of the coil, then wrap the whole coil with Sellotape to hold the turns in place, leaving only the ends free. Then, using a small piece of sandpaper, remove the enamel from the last centimetre of each end of the coil.
2. Most of the components are mounted on a piece of Veroboard (the type with parallel copper strips on one side). The piece used on the prototype measured 32 holes by 10 strips, as **Figure 2** shows. Before you start fitting components, cut the copper strips as shown. It is easier to do it now than when the board is littered with components! The strips may be cut with a 3 mm (⅛ inch) twist drill rotated between thumb and forefinger. Resist the temptation to use a hand drill – the idea is just to cut the copper, **not** to drill right through the board!
3. Solder the IC sockets and the other components on the board as shown in Figure 2. Make sure that the IC sockets are fitted with their notches towards the *top* of the board, as viewed in Figure 2. Do not insert the chips yet. Always keep the wires left over from cropping resistors and capacitors, they will come in handy at times like this: make the wire links that are clearly shown in Figure 2. Connect the electrolytic capacitors (C5, C6, C8 and C9), the transistor and the LED the correct way round; then check it again when you have done it!
4. Finally, solder lengths of stranded insulated wire to act as ‘flying leads’ for future connection to L1, VC1, LS1, S1 and the battery connector.
5. Apart from the battery holder, everything is mounted on the case *lid*. This makes assembly and testing much easier, and eases fault-finding if

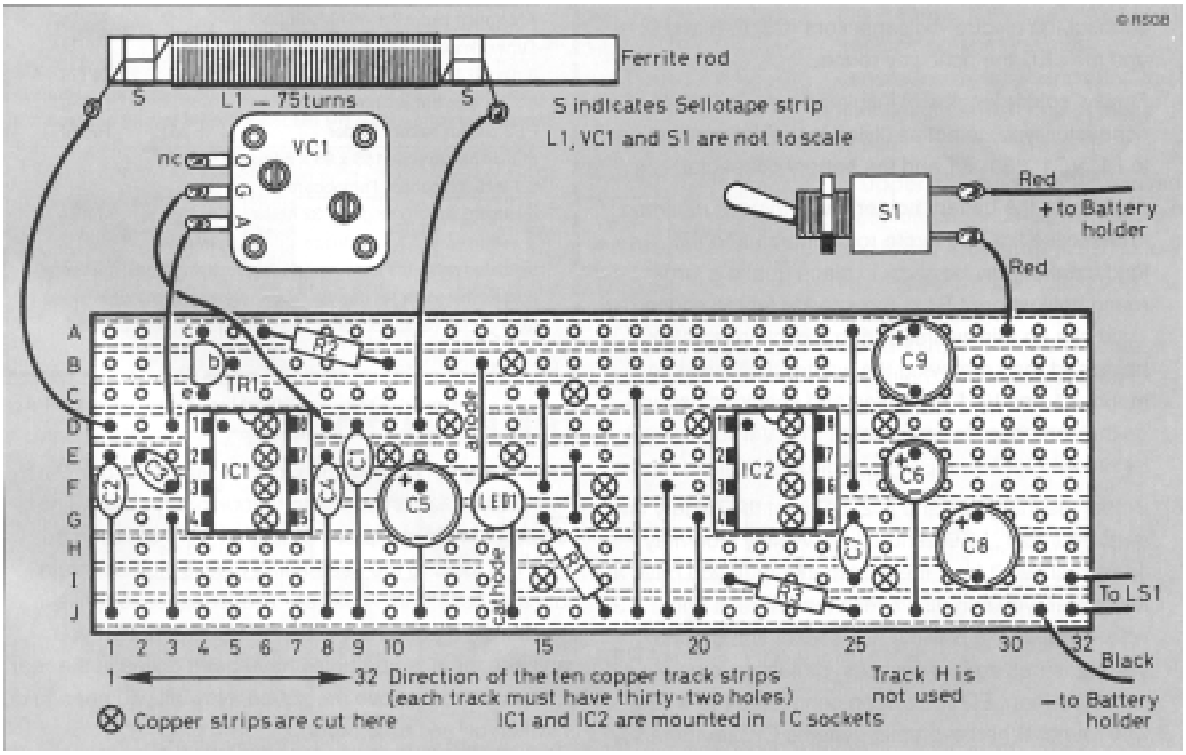
© R508
Ferrite rod
S
L1 – 75 turns
S
S indicates Sellotape strip
L1, VC1 and S1 are not to scale
VC1
nc
O
G
A
S1
Red
+ to Battery holder
Red
A
B
C
D
E
F
G
H
I
J
c
b
e
TR1
R2
C9
C2
C3
IC1
C4
C1
C5
anode
LED1
cathode
R1
IC2
C6
C7
R3
C8
To LS1
1 2 3 4 5 6 7 8 9 10 15 20 25 30 32
1 ← → 32 Direction of the ten copper track strips (each track must have thirty-two holes)
Track H is not used
Black
– to Battery holder
⊗ Copper strips are cut here
IC1 and IC2 are mounted in IC sockets

Figure 2 Veroboard layout for the Portable Radio. Make sure that all the wire links are included

the need arises! At the speaker position, make one large hole or a series of small holes to let the sound out. The ferrite rod may be stuck to the lid, as may the loudspeaker. Drill holes of the correct size to fit the particular types of variable capacitor (VC1) and switch (S1) that you are using. The Veroboard may be held in position by Blu-Tack or double-sided sticky tape.

6. Before inserting IC1 and IC2, connect the battery and switch S1 on. The LED should glow dimly (you may have to shield it with your hand in order to see it). If you have a test meter, check that there is about 1.3 V between pins 6 and 4 of IC1. If the reading is around 6 V or there is no glow, you may have connected the LED the wrong way round! When everything seems normal, switch off and disconnect the battery. Insert IC1 and IC2, making sure that the pins are straight and lie immediately above their corresponding holes in the sockets, and that the notches line up with the notches in the holders. Then push *gently* downwards on each IC in turn until the chip is firmly seated in its socket.
7. Switch on! By rotating the tuning capacitor, VC1, you should now be able to tune in many stations, rotating the radio to give you some volume control.

Final touches . . .

The handle was made with part of an old leather belt, secured to the case with ‘number plate’ nuts and bolts from Halfords. The loudspeaker grille is the lid from a *pot-pourri* container, and some extra holes were drilled in the back to improve the sound. See what you can find to finish off your radio!

Parts list

Maplin order codes are given for most of the parts, but you should get used to using the ‘beg, borrow or steal’ technique, or to use your ever-expanding junk box.

		Maplin code
Resistors: all 0.25 watt, 5% tolerance		
R1	68 ohms (Ω)	M68R
R2	5.6 kilohms (kΩ)	M5K6
R3	10 ohms (Ω)	M10R
Capacitors		
C1, C4	10 nanofarads (nF) or 0.01 microfarad (μF) ceramic	BX00A
C2, C3, C7	100 nanofarads (nF) or 0.1 microfarad (μF) ceramic	YR75S
C5, C8, C9	100 microfarads (μF) electrolytic, at least 10 V	FF10L
C6	10 microfarads (μF) electrolytic, at least 25 V	FF04E
Semiconductors		
IC1	ZN415E radio chip	
IC2	LM386 audio power amplifier	UJ37S
LED1	3 mm green LED	WL33L
TR1	BC548 npn transistor	QB73Q
Additional items		
LS1	Miniature 8 ohm loudspeaker	WB08J
S1	Miniature SPST toggle switch	FH97F
Ferrite rod	Length approx. 100 mm	YG20W
	24 SWG enamelled copper wire	BL28F

Additional items (*continued*)

VC1	Tuning capacitor 140 to 300 picofarads (pF)	FT78K
	Tuning knob	FK41U
	8-pin DIL IC sockets (two required)	BL17T
	4 × AA-size battery holder (long)	HF94C
	PP3-type clip for battery holder	HF28F
	Plastic box approx. 158 × 95 × 54 mm	LH51F
	0.1 inch Veroboard, min. size 32 holes × 10 strips	JP46A

Plus

Stranded insulated conductor for flying leads
Multicore solder
Materials for handle and speaker grille
Double-sided sticky tape or Blu-Tack
Sellotape
Glue
Four AA-size 1.5 V batteries

22 The Colt 80 m receiver – Part 2

Introduction

In Part 1 we constructed the audio amplifier module for the system and tested it in a very simple way. If you did what was suggested and built the simple Crystal Set to use as a signal source, you will know just how well the amplifier works.

The case

Metal cases for the project are available from Maplin, telephone 01702 554 161 (code XB67). From the photograph on p. 72 you can see the way the components are mounted. The audio amplifier is seen at the top right of the base, to the right of the tuning capacitor VC1. The next in this series will deal with the variable-frequency oscillator (VFO) and VC1. The current part deals with preparing the case to receive the components.

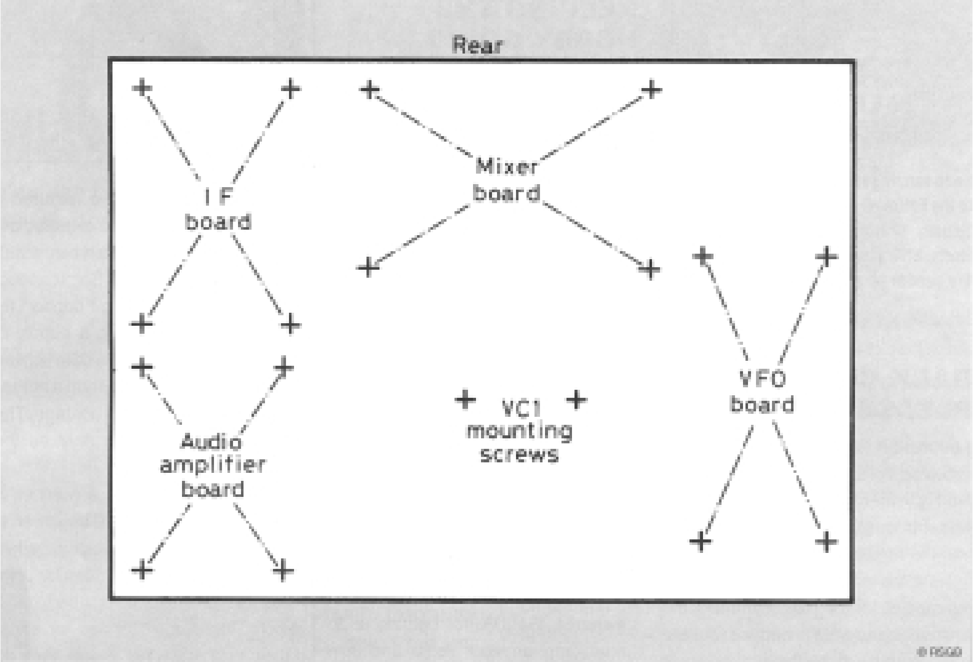
Rear
I F board
Mixer board
Audio amplifier board
VC1 mounting screws
VFO board
© RSGB

Figure 1 Fixing holes for each module are best measured from each printed circuit board or matrix board

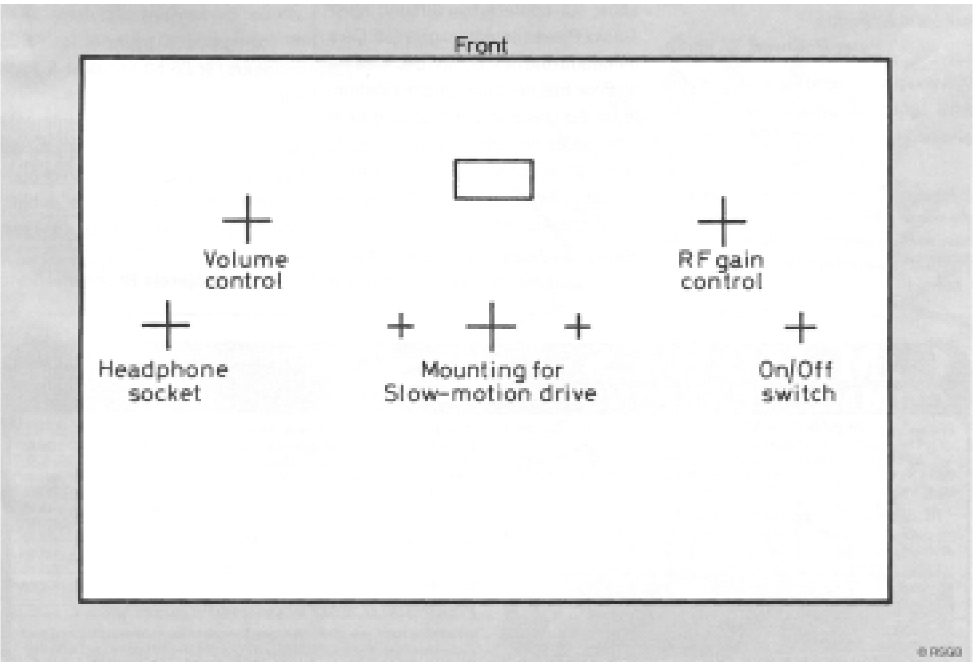
Front
Volume control
RF gain control
Headphone socket
Mounting for Slow-motion drive
On/Off switch
© RSGB

Figure 2 Position the slow motion drive to allow viewing of the tuning dial

Figures 1 and **2** show the markings for preparing the front panel and base. All the circuit boards and the tuning capacitor are mounted on the base using 10 mm stand-off pillars with 6BA bolts. The board locations are shown in Figure 1. The front panel control positions are shown in Figure 2, together with the small rectangular hole for viewing the tuning dial.

The best way to mark out the holes for the boards is to lie the boards flat on the base (before you've started soldering the components in) and marking the base through the holes in the boards. This minimises the scope for errors!

A *reduction drive* is used between the tuning knob and the capacitor shaft. This is simply a gear mechanism that slows down the capacitor shaft by a factor of six compared with the tuning knob, and makes tuning very much easier. The recommended variable capacitor also has a pulley wheel mounted on the shaft. Glued to this wheel will be a scale marked with frequency and is visible through the rectangular hole in the front panel.

The next part . . .

The variable-frequency oscillator and mixer will be added to the project.

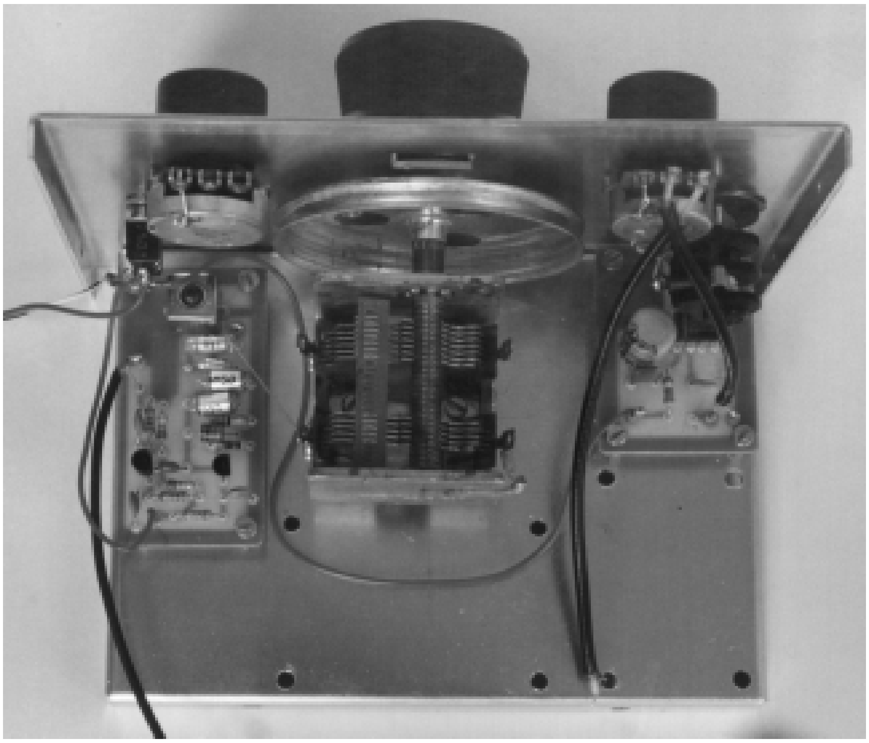

23 A simple transistor tester

Introduction

Although transistors aren't used as much as they were before integrated circuits came along, a transistor tester is still a useful piece of test equipment to have around the shack. This design is about the simplest possible and will produce an indication of whether a transistor is giving any current gain; this does not necessarily mean that the transistor is perfect but that it is working to a certain extent. This tester will **not** test field-effect transistors (FETs). If you buy a bag of transistors at a rally, this tester is useful for giving a yes/no indication of which ones go straight in the bin and which are kept for further use.

How it works

Figure 1 shows the simple wiring circuit. In order to explain the working of the circuit, a circuit diagram is shown in **Figure 2**, with the npn transistor, TR1, under test shown as part of the circuit. The pnp/npn selector switch, SW1, is omitted for clarity.

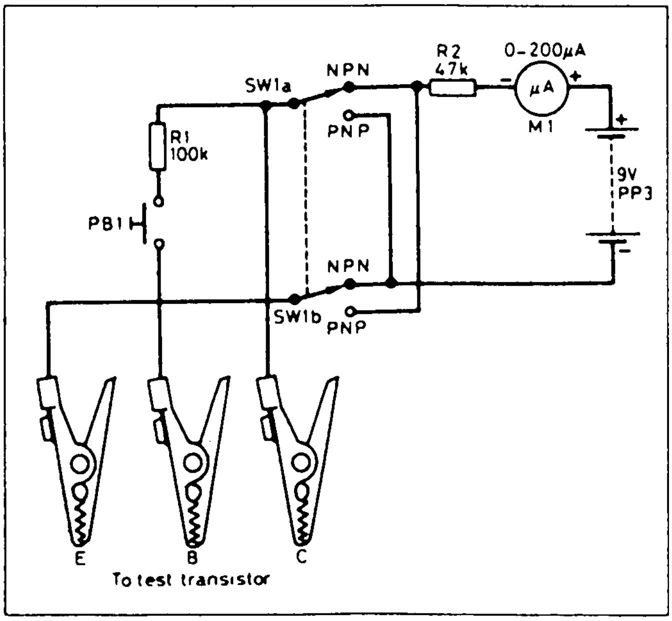

Figure 1 Circuit diagram of the transister tester

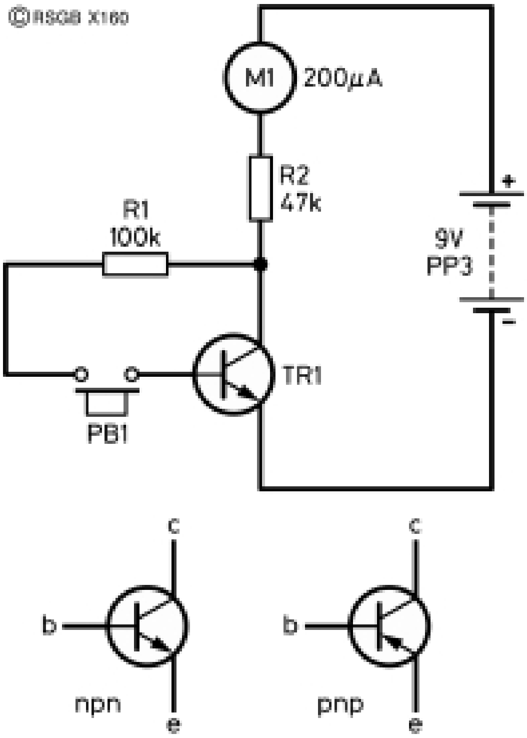

Figure 2 Testing an npn transistor

Any current that flows through the transistor, TR1, must flow from the battery, through the meter, M1, and through the protective resistor, R2. R2 prevents excessive current flowing through the meter and damaging it. Even if there is a short-circuit between emitter and collector, the maximum current that will flow is given by the simple equation

$$I = \frac{V}{R},$$

where I is the current flowing in amps,
V is the battery voltage, and
R is the total circuit resistance in ohms.

Putting in the correct values, gives

$$I = \frac{9}{47\ 000} = 0.000191 \text{ A, or 191 microamps } (\mu\text{A}).$$

The resistance of the meter, M1, will cut this down a little more, but it is within the indicating range (200 µA) of the meter.

As shown, with the push-button switch, PB1, *open*, a good transistor will not draw any current from the battery, and M1 will thus remain at zero. When PB1 is *pressed*, a *very small* current is injected into the base of the transistor. If the transistor is working, it will produce a much larger current between the collector and emitter, and this current will also flow through M1 and R2, giving a significant reading on M1, showing that all appears to be well. If an appreciable current flows when PB1 is *open*, then your transistor is suspect.

Don't be put off by the apparently complicated switch, SW1. It is there to allow the other type of transistor, the pnp type, to be tested. All it does is reverse the battery connections, so that the emitter goes to the *negative* battery terminal for testing an npn transistor, and to the positive terminal for a pnp type!

Most transistors in common use are of the npn type, which is why Figure 2 shows the testing of an npn type. The connections to the different transistor *encapsulations* (shapes) are given in any good component catalogue. Avoid the trial-and-error method to discover the connections to a transistor. This is unscientific, and can be very frustrating, particularly if the transistor is faulty in the first place!

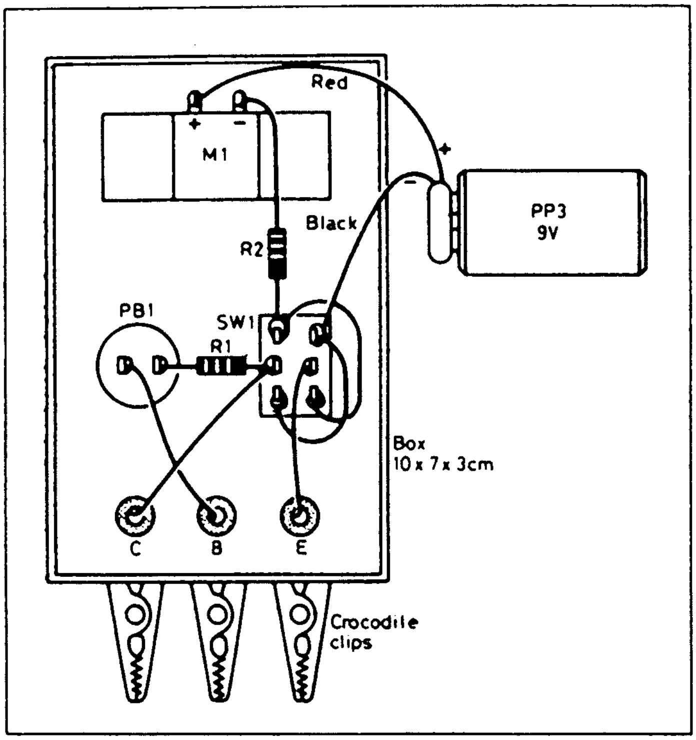

Figure 3 This shows how to wire up the tester

Construction

The unit can be built into a small plastic box, the components being soldered directly to the fixed terminals of the meter and the two switches; no circuit board is necessary! Use different colours of wire for the three test leads, and make sure you know which is which! Check the connections against the wiring diagram of **Figure 3**. When you are confident that all is correct, connect the battery and make sure there is no reading on the meter when nothing is connected to the crocodile clips!

Find any transistor for which you know the connections and the type (npn or pnp). Set the npn/pnp switch accordingly. Connect the three clips, making sure that they do not touch each other. A small reading on the meter at this stage means the transistor is suspect; a large reading means it is not working and should be thrown away! Press PB1 and watch the meter; a reading greater than half of full-scale indicates a good transistor. If it is less than half, you may have a transistor with 'low gain'; it may be usable for non-critical applications, but if you are in any doubt – throw it away!

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1	100 kilohms (kΩ)
R2	47 kilohms (kΩ)

Additional items

PB1	Push-button switch
SW1	DPDT (double-pole double-throw) switch
M1	Micro-ammeter – *not* less than 200 μA full-scale deflection
	Crocodile clips (3 needed)
	PP3 battery and connector
	Plastic box about 10 × 7 × 3 cm

Common types of transistor

BC108, BC109, 2N2369A

These are small-signal npn types, used in audio amplifiers. They are in metal cases (called TO18) and have a tab next to the emitter lead. Common types have a B or C suffix (e.g. BC109C). The C suffix indicates a higher current gain than those with a B suffix. The 2N2369A is specially designed for radio use at high frequencies.

2N3703, BC212L, BCY71
These are pnp transistors, and so must be used with the collector and base *negative* with respect to the emitter. These three types are used in small amplifiers and audio oscillators. The first two have plastic encapsulations (TO92), while the BCY71 has a metal case (TO18).

BFY50, BFY51, BFY52
For slightly higher powers, these are ideal. They have been used in novice transmitters up to 600 milliwatts (mW). The TO5 case is a scaled-up version of the TO18 case. They are all npn types.

2N3055, 2N3773, TIP35C
These are high-power transistors in bigger encapsulations. The thick metal TO3 case of the 2N3055 and 2N3773 is designed to bolt to a heat sink, a large piece of metal which conducts the heat into the air more rapidly than the transistor itself can. The TIP35C is made of plastic but has a thick metal tab by which it, too, can be bolted to a heat sink.

24 An introduction to transmitters

Introduction

We usually think of a transmitter as being a 'black box'. However, that is the form a transmitter takes for our use on the amateur bands. Many electrical circuits are transmitters, even though transmitting may not be their primary function!

Anything that emits electromagnetic energy at *any* frequency is a transmitter, from radio at the low-frequency end of the spectrum, to gamma rays at the high-frequency end. We all know that a magnet will attract certain metals and that a comb rubbed on your coat sleeve will pick up small pieces of paper. The former is an example of the effect of a magnetic field, the latter of an electric field. Electromagnetic fields are combinations of both types of field, and are produced whenever an electromagnetic wave is transmitted.

What frequency?

Many everyday objects have a natural frequency of oscillation. This is called their *resonant frequency*. A wine glass will ring when struck gently; an empty wine bottle will sound if you blow across the top; a guitar string will vibrate when plucked. These are all examples of resonance, and the resonant frequencies will **not** change unless the objects themselves are changed physically in some way. These are resonances in *sound*; we are primarily interested in electrical resonances.

The basic electrical resonant circuit is the combination of an inductor (coil) and a capacitor, as shown in **Figure 1**.

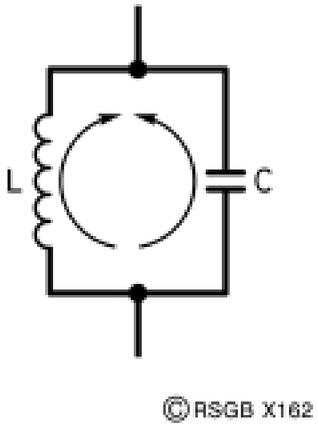

Figure 1 The basic electrical resonator. The energy in the circuit alternates between the inductor and the capacitor

A pulse of energy applied to this *tuned circuit* will make it *ring* (oscillate) at its resonant frequency. The energy in the circuit transfers between the inductor and the capacitor every cycle of the oscillation. Just like the wine glass, its oscillation dies away because it is losing some energy to its surroundings – it is transmitting! The frequency of the resonance depends on the values of *L* and *C*.

Keeping it going

If we want to keep the circuit oscillating, rather than having it die away, we must supply the circuit with just enough energy to replace the energy lost both by radiation and by losses in the circuit itself. Because of this, you will find in all oscillator circuits, a transistor, valve or FET working with the tuned circuit to provide this extra energy.

As it stands, of course, even with its transistor, our oscillator will not radiate very far. Connecting an aerial to it, and a Morse key to interrupt the power supply, it would become a very low-power CW transmitter. Add a couple more transistors to form a radio-frequency (RF) amplifier, and you have the basis of a simple low-power (QRP) transmitter.

Resonant circuits can also be made using quartz crystals; these work at the crystal frequency only, and this is marked on the crystal case.

A tiny spark transmitter

This is a simple piece of test gear that will increase your knowledge and understanding of resonance. You can use it to estimate the resonant frequency of most of the inductor/capacitor (LC) tuned circuits that you build. The circuit is shown in **Figure 2**. It operates around a relay. Any relay that operates from a 6 V to 9 V source and has contacts which are normally closed (i.e. closed when the battery is not connected). Fit the relay, a toggle switch and the battery in a metal box, and connected up as shown in the diagram. Some foam rubber inside the box may help to reduce the escaping noise of the relay. A small hole in the side of the box enables the 2-turn loop to emerge. This should be about 40 mm diameter, made with insulated wire. Switch on; there should be a loud buzzing noise from the relay. If not, you have probably chosen the wrong contacts on the relay!

When it is working, bring the loop close to the aerial of a radio – it should produce a loud noise from the speaker!

How it works

When you switch on, current flows through the relay contacts and through the relay coil. The relay operates and opens the contacts, causing the relay to 'drop out'. When it does, the circuit is completed again and the contacts are opened, and the cycle repeats. Each time the relay contacts open, there is a small spark between them, causing very rapid current surges through the wire loop. This makes the loop transmit RF energy, very briefly. In the early days of radio, this type of circuit was known as a *spark transmitter.*

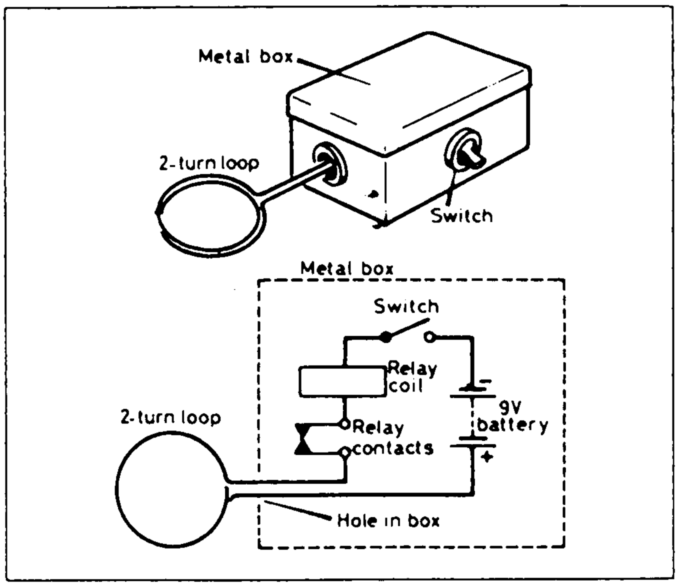

Figure 2 The current path is interrupted when relay is energised as shown above

Make a tuned circuit

Use a discarded toilet-roll centre, and wind about 10 turns of enamelled copper wire round it, keeping each turn close to the next. Scrape off the enamel for about 1 cm at each end, and solder a 100 picofarad (pF) capacitor (or a variable capacitor of about the same value) between the ends. The resonant frequency should be about 10 MHz. If you have used fewer turns or a smaller capacitor, the frequency will be higher.

Measuring the resonant frequency

Set up the buzzer as shown in **Figure 3**, with the loop around one end of your coil. Then make a similar loop, solder it to the end of a piece of coaxial cable going to the aerial socket of a calibrated receiver. Set the buzzer going, tune the receiver around 10 MHz, and search for the maximum noise level from the speaker. When you have found it, move the two loops as far away as possible from the main coil. This is called reducing the *coupling* between the coils, and it may result in a slightly different, but more accurate, resonant frequency.

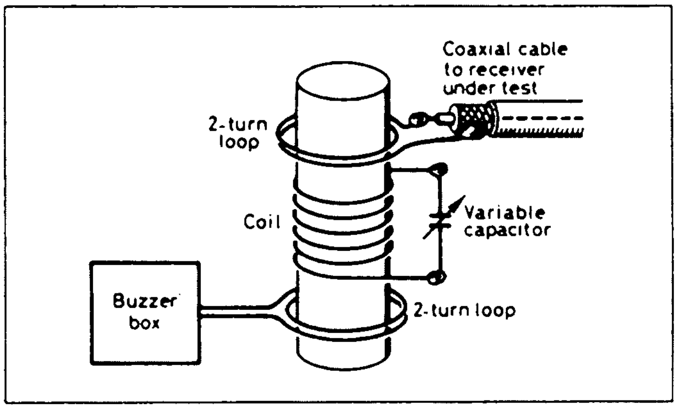

Figure 3 Experiment with the spacing between the loop and the tuned circuit

Parts list

Any small relay which operates between 6 V and 9 V
Metal box – do not use a plastic box!
9 volt battery and connector
On/off (SPST) toggle switch
Plastic foam, as required

25 The Colt 80 m receiver – Part 3

Introduction

In Part 2 we marked out the case ready for installation of the modules when they are completed. In this part, the building of the variable-frequency oscillator (VFO) and mixer will be described. This will produce a type of receiver known as *direct-conversion*, because it converts the radio-frequency (RF) signal directly into an audio-frequency (AF) signal which we can hear in a loudspeaker after amplification. A block diagram of the system is shown in **Figure 1**. The modifications needed to make a full superheterodyne receiver will be left until later.

The direct-conversion receiver covers the 80 metre amateur band and will receive both Morse (CW) and speech (SSB) signals. The audio amplifier was covered in Part 1, so your *Colt* is rapidly taking shape! By the time your construction has reached the end of this part, you will have a receiver ready to use, even if the project is not yet complete!

The direct conversion process

Like most things in radio, the principles of direct conversion are not difficult. From the aerial, the signal we want to hear is selected by the *tuned filter*, which rejects the signals we don't want. The signal then enters the

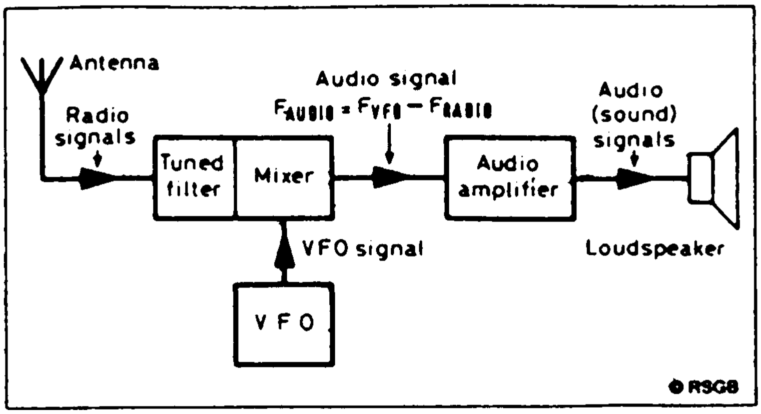
Antenna
Radio signals
Tuned filter
Mixer
Audio signal
$F_{AUDIO} = F_{VFO} - F_{RADIO}$
Audio amplifier
Audio (sound) signals
Loudspeaker
VFO signal
V F O
© RSGB

Figure 1 Stages of direct-conversion receiver

mixer, along with the signal from the *VFO*. The VFO produces a sine wave whose frequency can be varied across the whole of the 80 metre amateur band (3.5–3.8 MHz), by turning the knob on the tuning capacitor, VC1. The mixer produces, at its output, *two* signals; one signal is at the frequency of the **sum** of the signal and VFO frequencies, the other is at the **difference** of the two frequencies. It is the latter that we want. Let's look at the numbers involved. If the signal is at 3650 kHz and the VFO is at 3651 kHz, then the sum frequency is 7301 kHz, and the difference frequency is 1 kHz. If we feed the output of the mixer into our audio amplifier, the 7301 kHz signal is automatically removed (it is far too high to be considered an audio signal!) and the resulting 1 kHz signal is amplified and fed to the speaker, producing a note which we can hear!

Building the VFO

Figure 2 shows the circuit of the VFO. It is a tried and tested circuit, and should work first time. It uses a *field-effect transistor* (FET) for TR1, the oscillator itself. RFC is a *radio-frequency choke*, a coil of wire which will pass a direct current (DC) but which will prevent radio-frequency (RF) signals getting through.

All VFOs have a tuned circuit which, in this case, is formed by the coil L1 and the capacitors C1 and VC1. The frequency will also be affected to some extent by C2, C3, C4 and C5. Transistor TR2 is an *emitter follower*, a stage which gives no voltage gain but provides a good *buffer* stage, isolating the VFO from the effects of the stages that follow it. When building a VFO, the

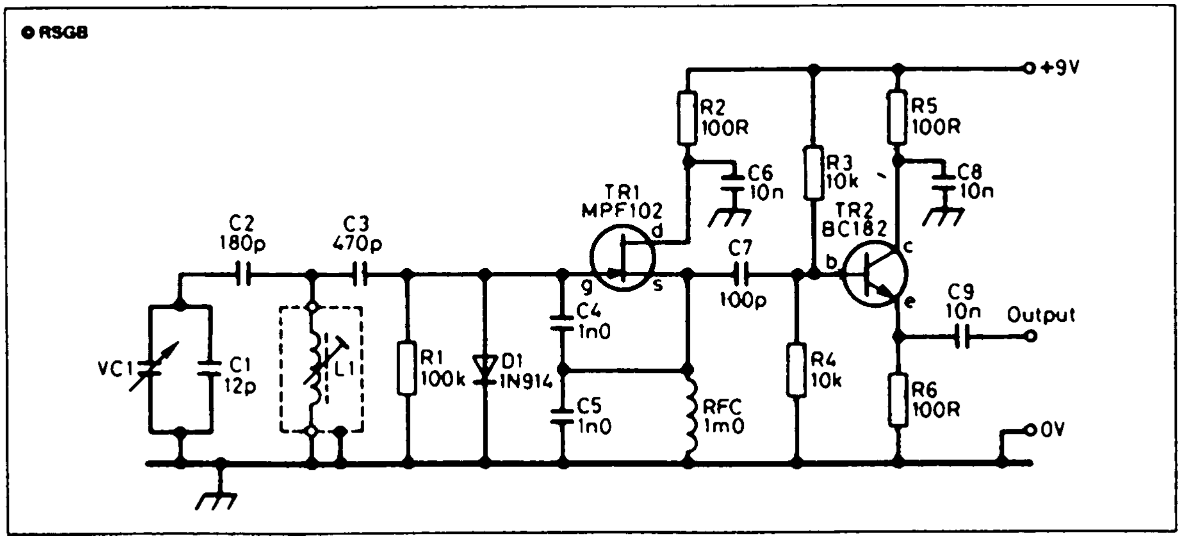

Figure 2 The variable-frequency oscillator uses a field-effect transistor (FET)

parts must be securely mounted. If components move, so does the frequency! At worst, the oscillation will become unstable and the VFO will be useless. Keep the component leads as short as possible – this improves their mechanical stability as well as their electrical stability!

Mount the parts on the printed-circuit board (PCB) or matrix board and, when completed, the VFO should look like the one shown on the left in the photograph on p. 72. Make sure that TR1, TR2 and D1 are the right way round.

On completion, check the component positions then mount it in the case as shown in the photograph, to the left of the tuning capacitor when viewed from the rear. The VFO coil, L1, will need some adjustment, but that will have to wait until the mixer is built. Connections to the other boards are made with screened cable.

The mixer board

So far, we have an audio-frequency (AF) amplifier and a VFO; the addition of a mixer board gives us a complete direct-conversion receiver for 80 m. The mixer circuit diagram is shown in **Figure 3**. Let's follow the signal path.

- From the aerial, the RF signal goes to the gain control potentiometer, RV1. This reduces very strong signals, to prevent them overloading the mixer.
- To select the required band of signals, a *bandpass* filter is made up of RF transformers T1 and T2; these are tuned by C1 and C3, and are coupled together by C2.

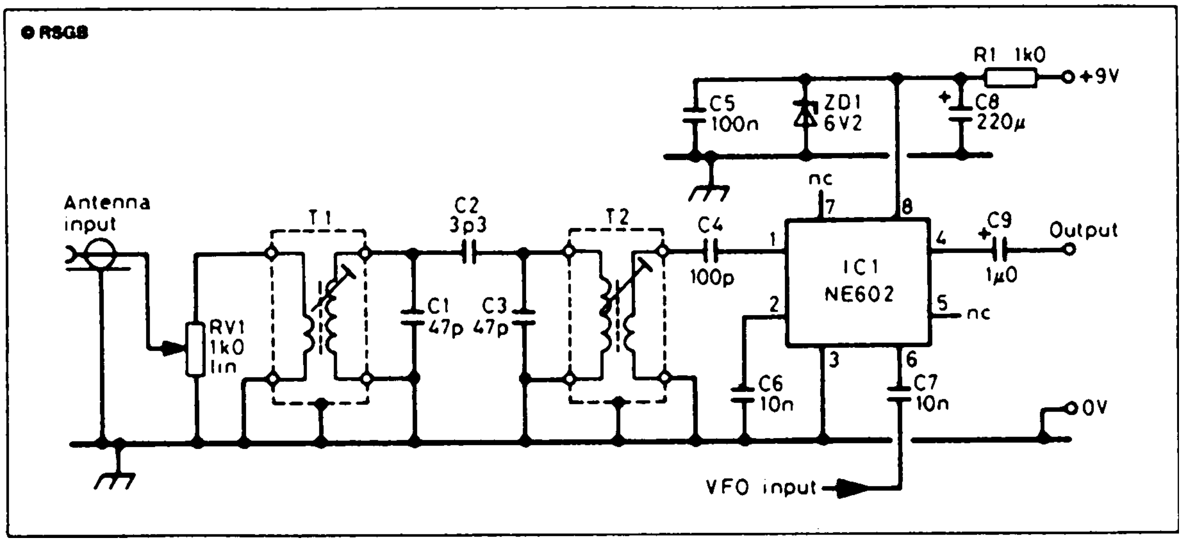

Figure 3 The mixer board has a bandpass filter and stabilised supply

- After the filter, the signal is coupled into the integrated circuit (IC) mixer type NE602, by the capacitor C4. Capacitors C5 and C8 *decouple* the supply line, to prevent unwanted signals on the supply from disturbing the VFO operation. The use of the term *decouple* is exactly the opposite of *couple*; when two circuits are *coupled* together, the signal passes from one to the other; when two circuits are *decoupled*, signals **cannot** pass from one to the other.
- The NE602 works with a 6 V supply; it is produced here from the 9 V supply by the *Zener diode* ZD1 and the resistor R1. ZD1 operates at 6.2 V, and gives a steady output for the mixer. The audio output from the mixer appears at pin 4 or IC1. This is taken to the audio amplifier board via C9 and the volume control (see Part 1).

Care must still be taken to insert some components the right way round. These are the electrolytic capacitors, C8 and C9, the Zener diode, ZD1, and the integrated circuit, IC1. Check all component positions and make sure all your soldered joints are bright and shiny.

Putting it together

The interboard wiring, shown in **Figure 4**, uses screened cable; ideally, this should be thin coaxial cable, but screened microphone cable is suitable. The diagram shows how the two controls, the RF Gain and Volume, are connected to the boards. The leads marked '+9 V' are all connected to the battery supply via a miniature on/off toggle switch. Double check all connections before connecting the battery.

Setting the VFO

Very little adjustment is needed to get the receiver going. Firstly, the VFO must be adjusted to cover the required band, in this case 3.5–3.8 MHz.

If you have a frequency counter, connect it to the output of the VFO. If you haven't, read this part anyway so you understand the process, then another means of setting the VFO will be given especially for you! Rotate VC1 until the vanes are fully meshed. Very carefully, adjust the core of the VFO coil (L1) with a plastic trimming tool so that the frequency approaches and settles at 3.500 MHz. When you rotate VC1, the frequency should increase to **at least** 3.800 MHz at the far end of its travel.

In the absence of a frequency counter, borrow a communications receiver, set it for SSB reception (USB or LSB) on exactly 3.500 MHz. Set VC1 with the vanes fully meshed and turn the core of L1 in both directions until you hear a whistle in the communications receiver. Rotate the core so that the whistle reduces in frequency. It will eventually fade out at around 200 Hz;

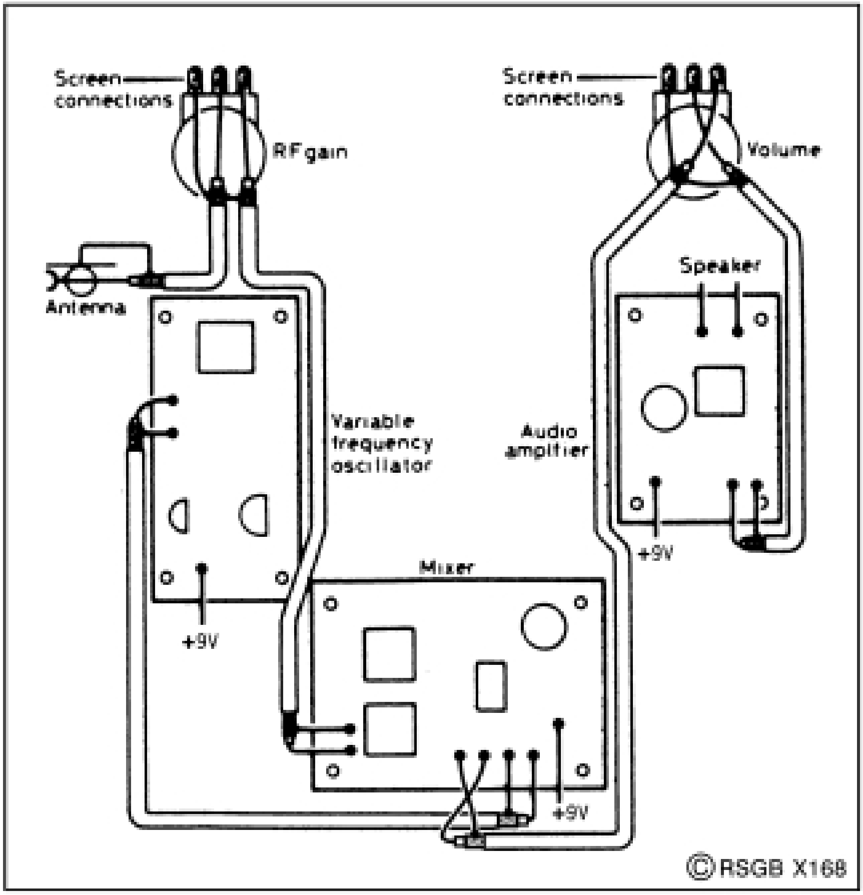
Screen connections
RF gain
Antenna
Variable frequency oscillator
+9V
Screen connections
Volume
Speaker
Audio amplifier
+9V
Mixer
+9V
©RSGB X168

Figure 4 Make sure that interconnections between the boards are correct, including cable screens

turn the core a little further and then leave it at that position. Rotate VC1 right to the other end of its travel, and search for the whistle with the tuning knob of the communications receiver. Check that the VFO frequency is at least 3.800 MHz.

Then, whichever method you are using for frequency measurement, mark the dial with frequency steps of 50 kHz. Setting and calibration are finished!

Setting the mixer

Again, there are various ways of doing this. If you have a signal generator, inject a signal at a frequency within the 80 m band, and adjust the cores of T1 and T2 sequentially for maximum output.

If you haven't a signal generator, connect an aerial to the mixer input, set the RF Gain to maximum (fully clockwise), and find a consistent signal. Adjust the volume control to a comfortable level. Rotating the core of T2 with your trimming tool, maximise the output. Then do the same with T1, although this will have much less effect. Find another station, and check that the positions of the cores aren't too different for a maximum signal.

You may find that your receiver benefits from the insertion of an aerial tuning unit (ATU) between the aerial and the input, to compensate for the impedance of your aerial not being 50 Ω. A design for such an ATU is presented in another part of this series. If the signals are still weak, connect the ATU to the junction of C1 and C2 via a 100 pF capacitor.

Try listening!

Remember that 80 m is a variable band. During daylight hours, your will hear Morse signals at the lower end of the band, and some British and closer continental stations between 3.7 and 3.8 MHz. In the evenings, stations up to 1000 miles away should be heard. Look for Novices around 3.7 MHz!

Parts list – VFO board

Resistors: all 0.25 watt, 5% tolerance		
R1	100 kilohms (kΩ)	
R2, R5, R6	100 ohms (Ω)	
R3, R4	10 kilohms (kΩ)	
RV1	1 kilohm (kΩ) linear	
Capacitors		
C1	12 picofarads (pF) polystyrene	
C2	100 picofarads (pF) polystyrene	
C3	470 picofarads (pF) polystyrene	
C4, C5	1 nanofarad (nF) polystyrene	
C6, C8, C9	10 nanofarads (nF) polystyrene	
C7	100 picofarads (pF) min. ceramic	
VC1	140 + 140 picofarads (pF) variable	
Semiconductors		
TR1	MPF102	FET
TR2	BC182	npn
D1	1N914	silicon
Inductors		
L1	Toko KANK3334	
RFC	1 mH	RF choke

Parts list – mixer board

Resistors: 0.25 watt, 5% tolerance

R1	1 kilohm (kΩ)
VR1	1 kilohm (kΩ) potentiometer (linear)

Capacitors

C1, C3	47 picofarads (pF) min. ceramic
C2	3.3 picofarads (pF) min. ceramic
C4	100 picofarads (pF) min. ceramic
C5	100 nanofarads (nF) min. ceramic
C6, C7	10 nanofarads (nF) min. ceramic
C8	220 microfarads (μF) electrolytic 16 V
C9	1 microfarad (μF) electrolytic 16 V

Integrated circuit

IC1	Philips NE602 or NE602A

Additional items

T1/T2	Toko KANK3333
	On/off switch
	Miniature toggle switch

The next part . . .

The IF amplifier and the Beat-Frequency Oscillator will be added to convert the *Colt* into a superheterodyne receiver.

26 A two-way Morse practice system

Introduction

It is said that the pleasure of sending and receiving Morse code more than compensates for the learning effort needed. Transmitters for Morse code can be much simpler than those needed for any of the speech modes.

Many people choose to learn sending and receiving on their own. It can be much more fun if you have someone to learn with you, and it is for this reason that the following project arose. It comprises just one small circuit, and you build an identical circuit for each of the people who want to learn with you. All the individual circuits are connected with their output leads in *parallel*, as the two-way circuit of **Figure 1** shows.

A simple circuit

This diagram shows two identical circuits, as would be used if two people wanted to learn together. The circuit centres around our old friend the NE555 integrated circuit (IC). As it is connected here, it works as an *astable multivibrator*, a daunting name for what is essentially an oscillator. Each circuit is self-contained, having its own battery, Morse key, sounder and

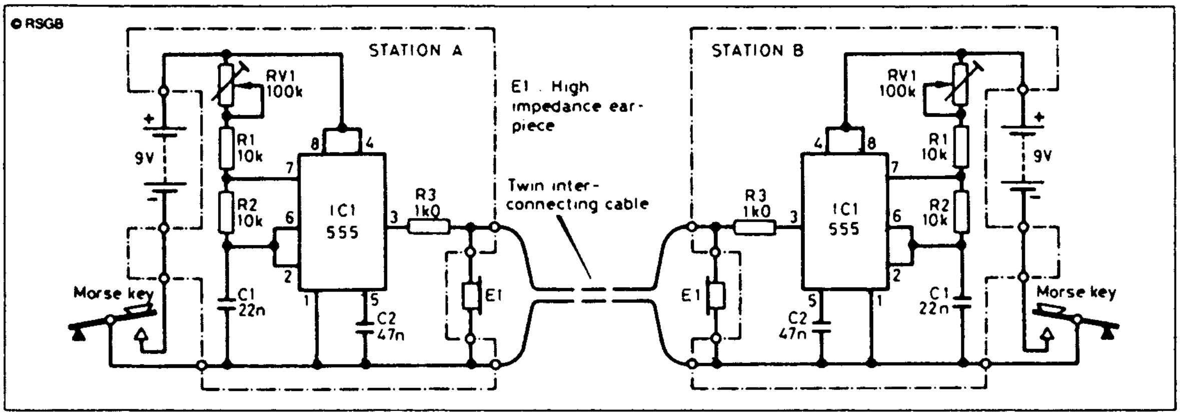
© RSGB
STATION A
STATION B
RV1 100k
R1 10k
R2 10k
9V
IC1 555
R3 1k0
C1 22n
C2 47n
E1
Morse key
E1 . High impedance ear-piece
Twin inter-connecting cable

Figure 1 The circuit diagram shows two stations, but you could connect several more if needed

plastic box. The Morse key makes and breaks the power supply to the circuit, thus turning on and off the note emitted by the earpiece. The frequency of the note is controlled by the variable resistor RV1, resistors R1 and R2, and the capacitor C1. The range given by RV1 is approximately from 500 Hz to 2200 Hz.

R3 protects the output of the IC against accidental shorting either to the 0 V or to the 9 V supply rails. For best results, a crystal earpiece or high-impedance headphones should be used. You could try Walkman-type headphones, but the volume may be too low for you.

The units may be interconnected with thin twin cable (the sort used for wiring doorbells) and can be comfortably separated by 25 m, more than enough to communicate between rooms, or even with a neighbour!

In use

When an operator presses his key, the note is heard by himself and by everyone else who is connected. If more than one key is pressed at the same time, two or more notes are heard by everyone! It helps if each person adjusts his own potentiometer, RV1, to give a note which is different from the others. When one operator stops sending, the other can start immediately, without pausing to change from transmit to receive. In practice, this technique is known as *full break-in*.

The current drawn by the circuit is only 10 milli-amps (mA) when the key is down, meaning that the life of a typical PP3 battery will be virtually its shelf-life! No switch is needed, because the key acts as the switch. The current drain will be even less if the NE555 is replaced by an ICM7555 IC. It has exactly the same pin connections as the NE555, so no circuit modifications are needed.

The circuit board

The circuit is constructed easily on a small piece of Veroboard of the copper-strip type. **Figure 2** shows the layout of the prototype, the board measuring 18 holes by 10 strips. **Be aware that there is no row 'I' when you are transferring mental images of where the parts are to real positions on the board!** Break the copper tracks where shown with a 3 mm (⅛ inch) twist drill, rotated carefully between thumb and forefinger. Hold the board up to a bright light to make sure that the tracks are completely broken and that there are no fragments of copper swarf shorting adjacent tracks. Then, solder in the wire links shown in Figure 2, followed by the resistors, RV1, and the capacitors, C1 and C2. Solder the battery connector leads to tracks A and G, G being positive. Use insulated wire to connect from the board to the jack sockets used as connectors for the key and the earphones. If you use

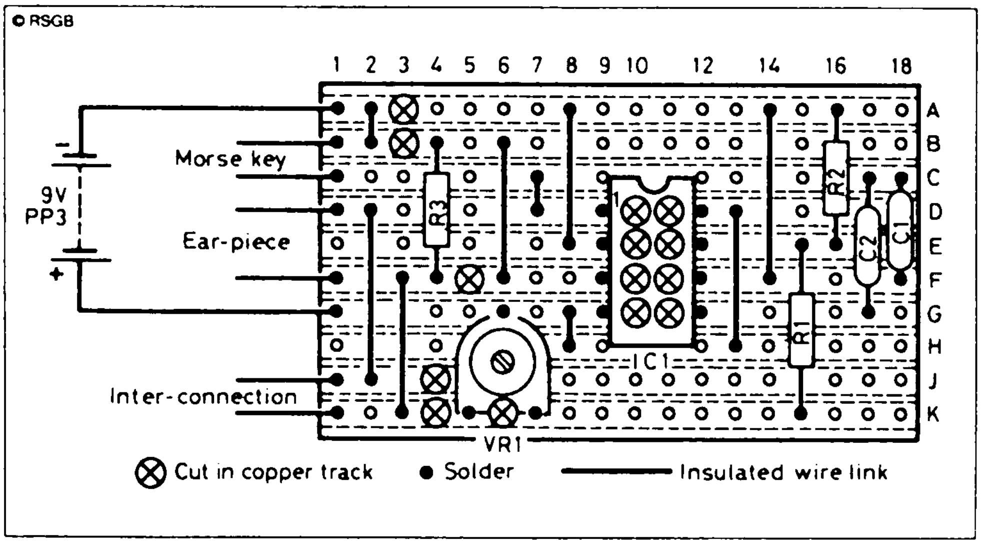
© RSGB
1 2 3 4 5 6 7 8 9 10 12 14 16 18
A B C D E F G H J K
9V
PP3
-
+
Morse key
Ear-piece
Inter-connection
R1
R2
R3
C1
C2
IC1
1
VR1
Cut in copper track
Solder
Insulated wire link

Figure 2 The Morse Duet is an easy project to build on stripboard (Veroboard)

different sized jack sockets (i.e. 2.5 mm and 3.5 mm) you won't plug the earphones into the key socket! A two-screw terminal strip is useful to connect the cable runs between operators.

The key

The circuit is designed to be used with a *straight key* (one that moves up and down). It is a requirement of the UK Morse test that a straight key is used, so it is very sensible to learn sending with a straight key before you try anything more complex! You will find many to choose from at rallies.

Parts list

For each board, you will need:		**Maplin order codes**
Resistors: all 0.25 watt, 5% tolerance		
R1, R2	10 kilohms (kΩ)	M10K
R3	1 kilohm (kΩ)	M1K0
RV1	100 kilohm (kΩ) min. preset (horizontal mounting)	UH06G
Capacitors		
C1	22 nanofarads (nF) ceramic, 25 V	WX78K
C2	47 nanofarads (nF) ceramic, 25 V	RA47B

Integrated circuit	
IC1 NE555	QH66W
Additional items	
8-pin DIL socket	BL17T
Jack plug and socket 2.5 mm	HF76H, HF78K
Jack plug and socket 3.5 mm	HF80B, HF82D
Battery type PP3	
Battery connector	HF28F
Plastic box with lid 114 × 76 × 38 mm	LH14Q
Veroboard 0.1 inch with copper strips 18 holes × 10 strips	JP46A
Terminal strip (two section)	FE78K
Crystal earpiece	LB25C
Morse key	

27 The Colt 80 m receiver – Part 4

Introduction

In Parts 1 to 3 of this series, the design of an 80 metre direct-conversion receiver has been described. In this final part, we are going to change the circuit to operate as a superheterodyne receiver, or *superhet*. Most radio receivers are superhets, and they change the incoming signal to another frequency, known as the *intermediate frequency*, or IF, before producing an audio signal. The use of a superhet in a good receiver is mandated by the requirements for good *sensitivity* and *selectivity.*

Sensitivity and selectivity

Figure 1 shows the block diagram of the receiver. If you look closely, you will see that it is two similar circuits, one after the other. The incoming

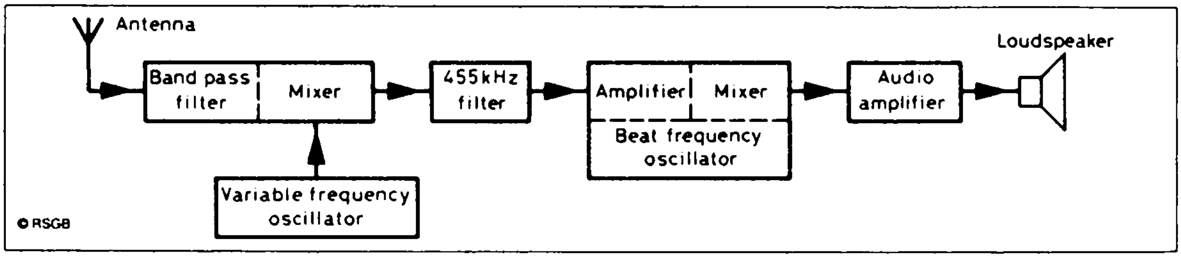

Figure 1 This is how the different stages of the Colt go together to make a superhet receiver

signal is filtered and fed to a mixer where it is combined (mixed) with the signal from the variable-frequency oscillator (VFO). This oscillator operates at a frequency which is 455 kHz higher than the incoming signal from the aerial, and the mixer output is therefore at a frequency of 455 kHz. If you are not sure about this, please refer to the section 'The direct conversion process' in Part 3 of this project. This new frequency is called the *intermediate frequency*, or IF. This frequency doesn't change; the tuning is accomplished by the VFO, and the mixer output is *always* at 455 kHz. The extraction of the audio signal from the IF signal is identical with the direct conversion process which is used in your existing receiver.

This may seem a long-winded way of doing things, but it has its advantages. A receiver must have a good *sensitivity*, or gain, so that it can receive very weak signals. In very general terms, it is easier to handle low-frequency signals than it is to handle high-frequency signals. We are changing our signal frequency from around 3.6 MHz down to 0.455 MHz (455 kHz), which is much lower and can be filtered and amplified relatively easily. A second advantage is that it is easier to provide gain at a fixed frequency than at a variable frequency. Remember that the IF is fixed, and providing gain is, again, relatively simple.

Our receiver also needs good *selectivity*, the ability to separate (or select) one station from another very close to it in frequency. This requirement is significantly simplified by the fact that the IF is fixed, and a good filter in the IF circuits can do wonders for the rejection of adjacent-frequency stations! Several stages of IF amplification and filtering are possible in more adventurous designs.

The filtered signal at 455 kHz passes to another mixer which has an associated oscillator, usually called a *beat frequency oscillator* (BFO). When receiving CW (Morse) signals, the BFO is usually tuned about 1 kHz above or below the IF (i.e. at 454 or 456 kHz) to produce a 1 kHz beat note as the audio signal. This signal is then amplified and fed to a loudspeaker or headphones.

The circuit

Figure 2 shows the circuit of the IF section shown in the photograph on p. 94; this section is added to the existing circuit to make it a superhet. The existing mixer board (described in Part 3) is used as the first mixer. Between it and the audio amplifier is connected the new IF board.

The signal output from the first mixer (at 455 kHz IF) is fed into a *crystal filter*, the most expensive part in the whole receiver. It provides the selectivity which makes the Colt such a good receiver. Another NE602 mixer/oscillator chip follows the filter. The oscillator section is controlled by the tuned circuit in T1, the frequency of which can be altered by rotating the core inside the coil. Once set, it remains fixed.

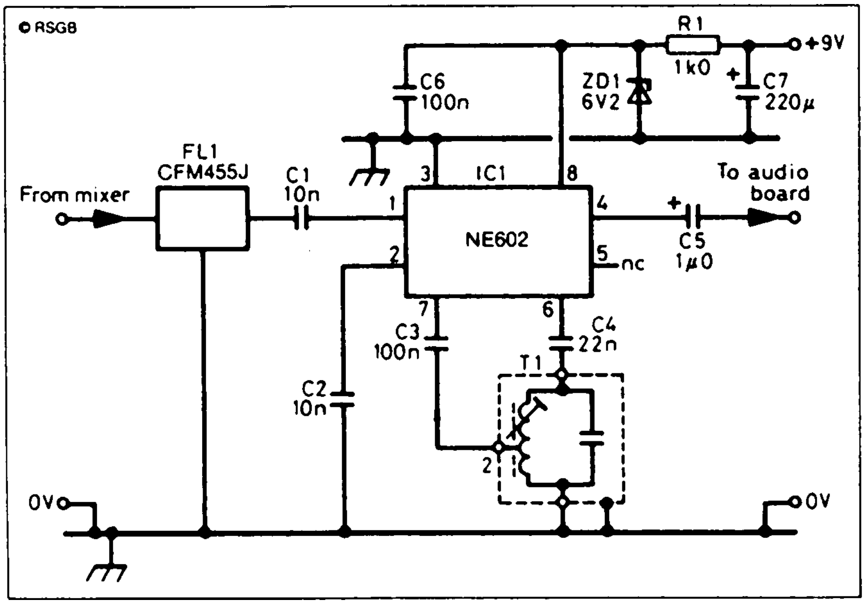

Figure 2 The IF board is connected between the mixer and audio amplifier

At this stage you should have built the first mixer and the audio amplifier, and proved that they *both* work by using the circuit as a direct-conversion receiver. When you have finished constructing and checking the IF board, you will need to add it to your existing circuits.

Adjusting and testing

There are three pairs of connections to the IF board – the 9 V supply leads, the input leads and the output leads. The IF board should be mounted on the metal baseplate, along with the other boards. **In making the following connections, make absolutely sure that the braid of each piece of coaxial**

cable is soldered to the correct connection on each board. The same applies to the polarity of the 9 V leads. Disconnect the screened lead which presently goes from the first mixer board to the AF amplifier at the amplifier end. Connect it instead to the *input* of the IF board. Connect a new piece of screened cable from the IF output to the vacated AF input. Now connect the 9 V supply leads to the same points as the supply leads from the other boards. Check these newly made connections.

You should have confidence at this point that things should be right. After all, you *have* tested the direct-conversion process and you *know* it works. All that you are now testing is the IF board. This is the attraction of building a receiver in modules, working from the speaker backwards, and testing as you go!

The VFO and BFO need to be correctly adjusted. Mesh the vanes of the tuning capacitor and, using the same frequency measurement method as you did originally, set the VFO frequency to 3.955 MHz (which is 3.500 MHz + 0.455 MHz, if you hadn't guessed!). The BFO can be set using a frequency counter, but it is just as good to set it by listening to SSB or CW signals. A high-pitched hissing sound should be heard in the speaker. As you rotate the core, the pitch should reduce, go through a minimum, then increase again. Set the core at the minimum pitch position. You may want to readjust the two cores a little as your listening skills improve but, once you are happy, they will never need to be altered again!

In conclusion . . .

You should now have built a superhet receiver capable of excellent results. It uses the same type of circuit as that found in far more advanced receivers.

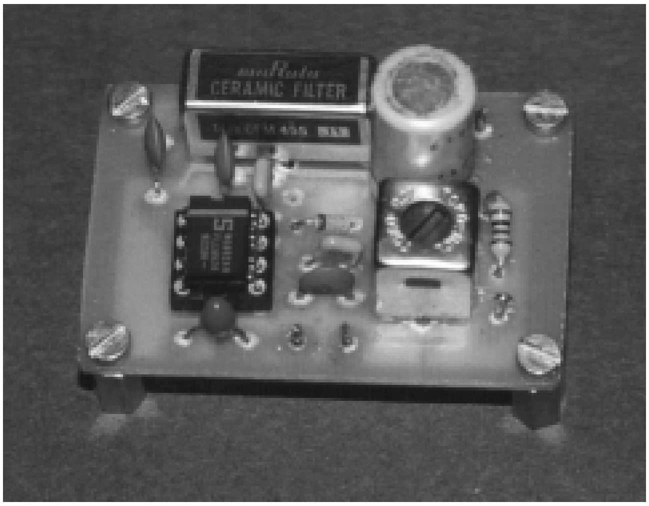

The superhet is far more sensitive than the direct-conversion type, and can weed out those elusive DX stations. You may have found that a station will appear at two places on the dial of the direct-conversion receiver; you will have no such problem with the superhet.

Parts list

Resistor

R1	1 kilohm (kΩ), 0.25 watt, 5% tolerance

Capacitors

C1, C2	10 nanofarads (nF) min. ceramic
C3, C6	100 nanofarads (nF) min. ceramic
C4	22 nanofarads (nF) min. ceramic
C5	1 microfarad (μF) 16 V electrolytic
C7	220 microfarad (μF) 16 V electrolytic

Integrated circuit

IC1	NE602

Additional items

ZD1	Zener, 6.2 V 0.5 watt
FL1	Crystal filter, Murata CFM455J
T1	Tuned inductor, Toko YHCS11100AC

28 A simple crystal set

Introduction

There are many designs of crystal set – they are all ‘simple’, and even the most seasoned radio amateur will build one of these every so often because it is something that never ceases to amaze! Although using modern components, the design is a period piece!

Design and construction

The circuit diagram of **Figure 1** shows the simplicity of the circuit. The prototype was constructed on a wooden baseboard with an aluminium front panel, although this could be made of wood if you prefer.

Winding the coil always causes the most groans but when finished, gives the most satisfaction. The construction of the coil is detailed in **Figure 2**. A discarded toilet roll centre is an ideal *former* on which to wind the coil. The coil will take up about 7 cm of the length of the tube, so assuming the tube is about 11 cm long, the ends of the coil will be about 2 cm in from each end of the tube.

The aerial coil, the *primary winding* of this radio-frequency (RF) transformer, is made from about 30 turns of 26 SWG enamelled copper wire, wound on the matchstick ribs. These ribs should be about 2 cm long, and are glued on **top** of the end of the secondary coil *after* the secondary coil has been wound. The *secondary winding* has 140 turns with *taps* every 10 turns for 70 turns. This coil covers most of the length of the cardboard tube. To make your coil-winding easier, here are some tips for completing the coil and still having some hair left at the end!

- Support the reel of wire on a dowel rod or pencil, clamped in a vice or held rigid in a vertical position by some other means. Don't leave it to trail around on the floor.
- Have a small piece of sandpaper handy to remove the enamel from the copper wire, starting with the end coming from the reel. Remove about 1 cm only, until the shiny copper is visible all round the end of the wire.

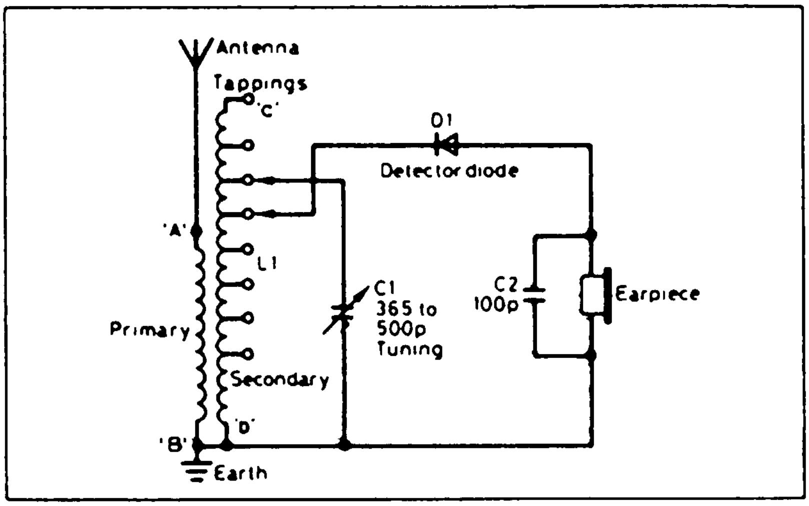

Figure 1 Circuit diagram of the crystal set

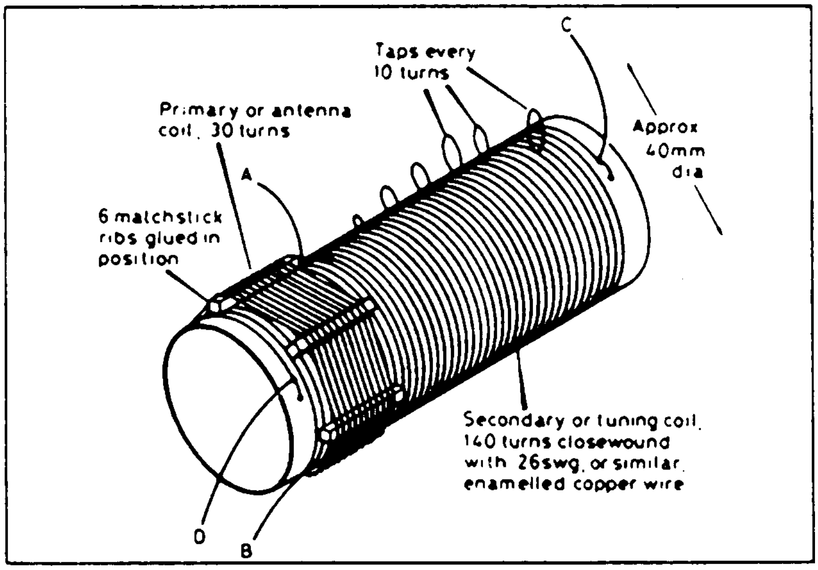
C
Taps every 10 turns
Primary or antenna coil, 30 turns
A
Approx 40mm dia
6 matchstick ribs glued in position
Secondary or tuning coil, 140 turns closewound with 26swg or similar enamelled copper wire
D
B

Figure 2 The coil is mounted on a cardboard tube as shown. Exact size is not too important. Try different taps for best results

- This part is the most difficult of all. Seasoned coil winders may have their own favourite ways of doing this, but for first-timers, play it safe and do it this way! Have a second bobbin (or a piece of stout card) ready, on which you will need to wind the wire loosely while you prepare the taps. Then, starting from the free end of the wire, measure 125 cm and, at this point, remove about 2 cm of the enamel with the sandpaper. Bend the wire firmly back on itself in the middle of the bared area, and *tin* (cover with solder) the exposed copper. This is your first tap. Then repeat this six more times at 125 cm intervals, thus giving you seven taps; wrap the wire as you go, on to the second bobbin or piece of card. After the seven taps, you are half-way down your coil so, to begin winding it in earnest, wind the wire carefully from the second bobbin or card **back** on to the original reel of wire.
- Looking at Figure 2, make two small holes about 3 mm apart, about 2 cm down from the top of the tube, where wire C will be entering. Poke the end of the wire into one hole, then bring it out again through the other; leave about 10 cm of wire on the free end. Loop this wire in and out once more, thus anchoring the wire firmly.
- Then, begin winding; keep the wire tight and make adjacent turns touch neatly. As you pass each tap, make sure it sticks outwards while avoiding flexing the wire too much; the wire is inherently weak at each tap.
- When you reach the last 10 cm of wire, stop. Put two more holes, like the first, beside your stopping point, and anchor the end, D, of the coil in the same way as you did with end C. The secondary winding is complete. Take a break *after* you have completed the next step!

- Prepare the six matchsticks for supporting the primary coil and glue these every 60 degrees around the end of the coil at D. Remove the enamel from the free end of the new copper wire on your reel and, using your favourite quick-drying cement, put a drop over the wire on one of the matchsticks, leaving 10 cm free as before. If you are using 26 SWG enamelled copper wire, its diameter is 0.46 mm, so 30 turns should occupy 14–15 mm of the 20 mm matchstick length. This will allow you to position the glued point such that the coil will be roughly in the centre of the matchsticks. When the glue has set, wind the 30 turns and anchor this end of the coil in the same way. Then remove the enamel at the end of the 10 cm lead.

Assembly

The coil former can be screwed to the baseboard at each end, *before* the front panel is screwed to the baseboard! As the photograph on the next page shows, the tuning capacitor is mounted in the centre of the front panel. Five solder tags (or drawing pins as a last resort) should be screwed into the baseboard. The two to the left of the photograph are the connections to the earphones, the two along the rear edge have the diode, D1, soldered to them. One end of the diode (and, for once, you can connect the diode either way round!) is connected to one earphone tag. The other end has a flying lead of about 10 cm attached to it and terminated in a crocodile clip which is used to connect to one of the taps on the coil. The fifth tag secures the end, A, of the primary coil, to which you will attach the aerial. The earth tag on the tuning capacitor, C1, serves as an anchor point for the ground connections to B and D and to the other earphone tag.

Use

The longest piece of wire you have available to use as an aerial should be connected to the aerial tag just mentioned. If you can connect the earth tag of the tuning capacitor to a good electrical earth, this will help also. You should be able to hear something as you turn the tuning knob. Try adjusting the tapping connections on the coil – change only one at a time, or you will never find the optimum positions. You now have a fully operational crystal set!

Simple it may be, but this circuit illustrates some important principles which are used even in the most expensive receivers. Firstly, the coil, L1, and the capacitor, C1, form a *tuned circuit* which *resonates* at the frequency of the station you have tuned in. This *selects* the signal you want to hear. The *modulation* on this signal is removed by the *detector*, D1, and fed to the earphones, which act as *transducers*, turning the electrical energy into sound energy which you can hear. It doesn't need a battery or other power supply either!

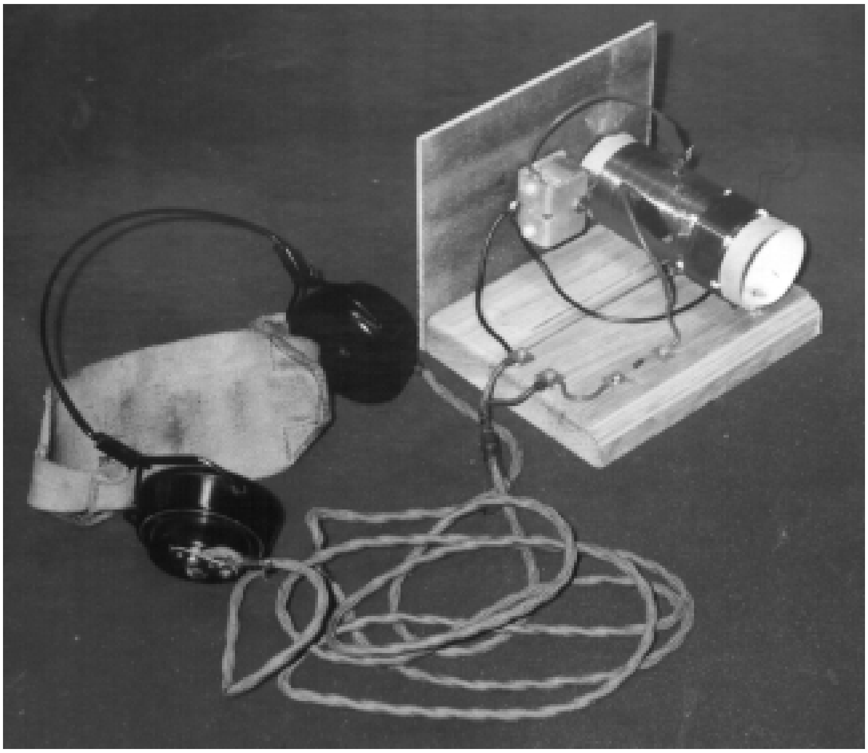

Parts list

Capacitors

C1	Variable capacitor of between 200 and 500 picofarads (pF) maximum
C2	100 picofarads (pF) min. ceramic

Additional items

D1	Germanium, type OA47 or OA90
	High impedance crystal earphone (not Walkman type)
	Tuning knob
	Wooden base, approx. 110 mm square
	Aluminium (or wood) front panel, approx. 80 × 100 mm
	Cardboard tube (toilet roll centre), about 110 mm long, 40 mm diameter
	Reel of 26 SWG enamelled copper wire for both windings of L1

29 A crystal calibrator

Introduction

A crystal calibrator is a device that produces oscillations of a precise frequency which are rich in *harmonics*. A harmonic is an integer multiple of the fundamental frequency; for example, if our fundamental frequency was 100 kHz (as it is in this circuit), there would be harmonics of this frequency at 200 kHz, 300 kHz, 400 kHz, and so on. These harmonics can be used, when picked up by any receiver, to calibrate that receiver, as the harmonics are quite accurate in frequency (see later for an assessment of accuracy). However, a gap of 100 kHz between harmonics is rather wide for most purposes, so we reduce this frequency to 25 kHz, so that the harmonics are then 25 kHz apart, thus producing a much more useful set of *marker points*. A calibrator such as this is often called a *crystal marker*, producing these marker points from 25 kHz to beyond 30 MHz.

The circuit

The circuit diagram is shown in **Figure 1**. The circuit around TR1 is the fundamental oscillator, and its frequency is controlled by the quartz crystal, X1. Even crystal oscillators are not 100% accurate, and the small trimmer capacitor, TC1, is able to 'pull' the frequency to one which is nearer the

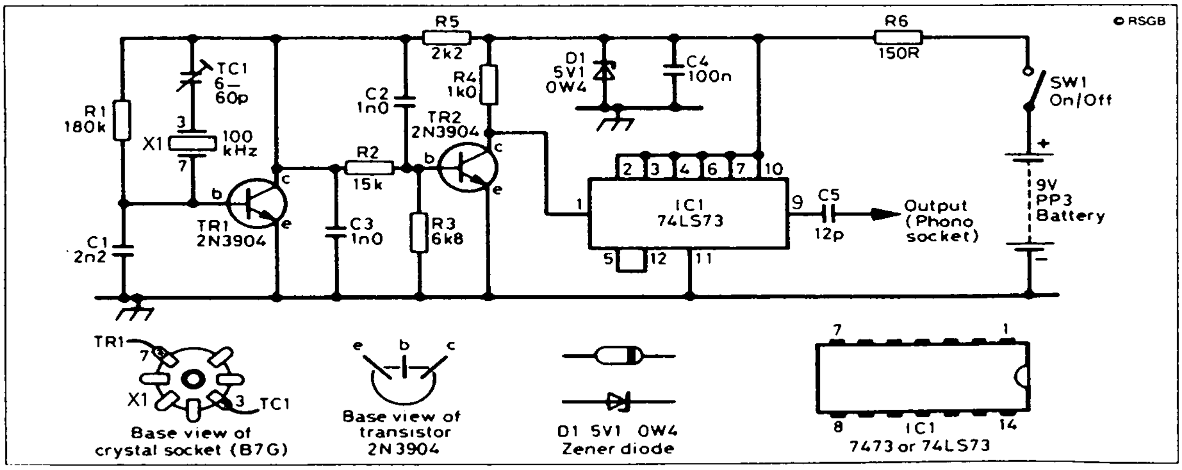
R5
2k2
R6
150R
© RSGB
D1
5V1
0W4
C4
100n
R4
1k0
C2
1n0
TR2
2N3904
TC1
6–
60p
R1
180k
X1
3
100
kHz
7
R2
15k
b
c
e
TR1
2N3904
C3
1n0
R3
6k8
C1
2n2
SW1
On/Off
2
3
4
6
7
10
1
IC1
74LS73
9
5
12
11
C5
12p
Output
(Phono
socket)
+
9V
PP3
Battery
−
TR1
7
X1
3
TC1
Base view of
crystal socket (B7G)
e
b
c
Base view of
transistor
2N3904
D1 5V1 0W4
Zener diode
7
1
8
IC1
14
7473 or 74LS73

Figure 1 Circuit diagram of the calibrator which uses easily obtainable components

correct one. TR2 is a *buffer* stage, which isolates the oscillator from the rest of the circuit. It acts as a switch, and applies a good signal (switching between 0 and 5 V) to the input of IC1.

IC1 is an integrated circuit which can be connected to do several things. It is connected here to divide the incoming frequency by a factor of 4, producing on pin 9 an output frequency of 25 kHz.

The combination of R6, C4 and D1 produces a supply of 5 V for IC1; it would be damaged if the battery voltage were applied to it. You will no doubt be ready to assemble the circuit, so here is some information for you.

Construction

If possible, always build a circuit in individual stages, which you can test as you go along. It is not always easy in small projects like this, but even the crystal calibrator can be split into two for construction and testing.

It is an ideal project for assembly on Veroboard of the copper-strip type; the prototype layout is shown in **Figure 2**, on a board measuring 11 holes by 24 strips. First of all, remove the copper strip at the locations shown, using a 3 mm drill rotated between the fingers. Hold the board up to the light to ensure that there are no pieces of copper swarf bridging adjacent strips, and that the copper is completely removed where it should be! Assemble the circuit from left to right, but do not wire up anything around IC1. Be aware that the diagram shows the board *from the component side*. Although you

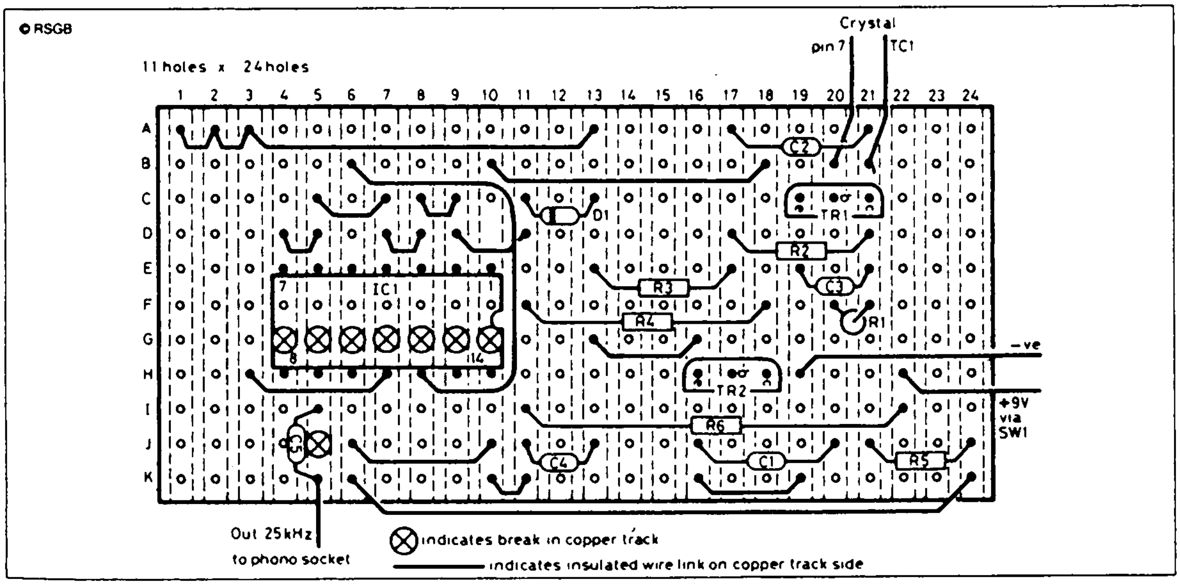
© RSGB
Crystal
pin 7
TC1
11 holes x 24 holes
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
A B C D E F G H I J K
IC1
7
8
14
D1
TR1
TR2
R1
R2
R3
R4
R5
R6
C1
C2
C3
C4
−ve
+9V via SW1
Out 25kHz to phono socket
indicates break in copper track
indicates insulated wire link on copper track side

Figure 2 Stripboard, such as Veroboard, should be cut as shown to approx. 62 × 28 mm

can use virtually any kind of crystal, you will need to mount it firmly. If you use a small one, it can be soldered directly in position on the board. With one in a valve envelope, you will need a B7G valveholder. Be prepared to mount the circuit in a **metal** container; the prototype used an (empty) tin of tuna! The valveholder mounts well on any metal case.

Disconnect the aerial of your receiver, and replace it with about 30 cm of wire laid near your oscillator circuit. Connect the battery and switch on. Tune the receiver around until you can hear a whistle on SSB. Rotate the tuning knob to reduce the frequency of the whistle and, as the whistle becomes too low to hear, you have reached one of the calibration markers, and your frequency will be an integer multiple of 100 kHz. If you are already using a calibrated receiver, you will be able to verify this. Going up or down in frequency should locate another marker 100 kHz away, and so on right through the receiver's tuning range. If the signals coming from the circuit are very weak, switch off the calibrator and connect another 30 cm piece of insulated wire to the collector of TR2, and lay it close to (but *not* touching) the wire from the aerial connector. Switch on an try again; you should not now have a problem! If you have access to a multimeter, check that the voltage at the collector of TR1 is very close to 5 V. If it is close to 9 V, you have connected the diode, D1, the wrong way round!

Having verified that the oscillator is working, you can now wire up the integrated circuit socket, being careful to put the notched end of the socket pointing towards R4, the collector resistor for TR2.

Check your connections around IC1, and when you are satisfied that they are correct, line up IC1 with its socket, making sure that the notched ends are together, and press down *gently* to insert the IC into its socket. Insert the 30 cm piece of insulated wire into the output socket, connect the battery and switch on. You should still hear whistles in your receiver, but now they should be 25 kHz apart, rather than 100 kHz apart.

All that now remains to be done is to mount the circuit rigidly inside whatever casing your have chosen. Make sure that none of the connections under the board touch the metal case, and secure the valveholder, on/off switch and output connector to the case. You now have a completed crystal calibrator.

Calibration

The simplest way to calibrate your circuit is with a frequency counter. Most clubs will have one of these and, if not, will know someone who has! Connect it to the collector of TR2, where the frequency should be 100 kHz. Do *not* connect it to any part of the circuit around TR1, or you may alter the frequency you are trying to measure! If the frequency is not exactly 100.000 kHz, rotate TC1 until it is (or is as close as you can get it). Now your calibrator is as accurate as the counter with which you have calibrated it.

Accuracy

Despite your best efforts at calibration, by whatever means, your crystal will *never* have a constant frequency. Such a thing is a scientific impossibility. It is usual to express the accuracy of a crystal in parts-per-million (ppm), and it is governed by many things, principally its temperature. You will be *very* fortunate if your circuit maintains an accuracy better than about ±10 ppm. Expressed in figures, it means that the true frequency can be anywhere between 99.999 kHz and 100.001 kHz, i.e. within 1 Hz of the correct frequency.

Although this may seem more than adequate, it is as well to remember that the accuracy degrades as the frequency increases. At 1 MHz the error will be ±10 Hz and at 10 MHz it will be ±100 Hz. At 30 MHz it will be 300 Hz. Even so, this should be acceptable for most non-critical applications.

Parts list

Resistors: all 0.25 watt, 5% tolerance		**Maplin code**
R1	180 kilohms (kΩ)	M180K
R2	15 kilohms (kΩ)	M15K
R3	6.8 kilohms (kΩ)	M6K8
R4	1 kilohm (kΩ)	M1K0
R5	2 kilohms (kΩ)	M2K0
R6	150 ohms (Ω)	M150R
Capacitors		
C1, C2, C3	1 nanofarad (nF) or 1000 picofarads (pF)	WX68Y
C4	100 nanofarads or 0.1 microfarad (μF)	YR75S
C5	12 picofarads (pF)	WX45Y
TC1	60 picofarads (pF) trimmer	WL72P
Semiconductors		
TR1, TR2	2N3904 npn	QR28F
IC1	7473 or 74LS73N dual JK flipflop	YF30H
D1	5.1V 500 mW Zener	QH07H
Additional items		
100 kHz crystal		
B7G valveholder (found at most rallies)		
14-pin DIL IC socket		BL18U
Phono socket		YW06G
PP3 battery connector		HF28F
PP3 battery		
On/off switch		FH00A
Veroboard		JP47B
Metal case as required		

30 A simple short-wave receiver – Part 1

Introduction

No apologies for yet another design of short-wave radio. This is the beauty of our hobby – there is always something new to try. Not all circuits operate in the same way; not all circuits work equally well; sometimes a simple design suits the operator better than a complicated design. So here is another one for your consideration; it's a good project for a novice, *and* a good one-evening project for someone more experienced.

A basic description

Figure 1 gives the block diagram of the receiver, which employs a *regenerative detector*, one of the earliest techniques by which excellent *selectivity* (the ability to separate two stations very close in frequency) could be combined with good *sensitivity* (the ability to pick up very weak stations) using a very simple circuit. The receiver falls into the category known as *tuned radio frequency* (TRF), meaning that the whole circuit (prior to the extraction of the modulation from the carrier) operates at the incoming radio frequency. In other words, it is *not* a superhet.

As you can see, the regenerative detector has what is called a *feedback loop*, which feeds a small amount of the output signal back to the input. You have heard the effects of feedback with a public address system, when the microphone gain is too great – everything becomes very loud and then bursts into a deafening squeal! This is exactly what the feedback does here, except that it is carefully controlled, thus providing both gain and selectivity. The resulting audio is then amplified for use with headphones.

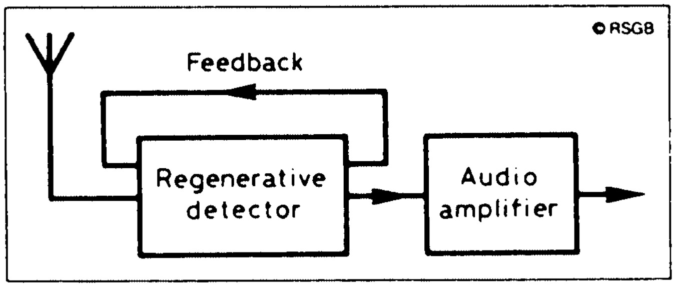

Figure 1 Just two simple stages make up the circuit

The circuit

The circuit of **Figure 2** shows the complete system. TR1 is an untuned *field-effect transistor* (FET) stage, and is used to match the aerial to the next stage. Occasionally, a regenerative detector produces unwanted signals, and TR1 also prevents them from reaching the aerial and being transmitted! The smaller of the two windings on T1 (the *primary* winding) will match a low-impedance aerial, the capacitor input matching a high-impedance aerial.

Ignore TR2 for the moment – the next stage in the signal path is TR3, the detector, using another FET. The tuned circuit is formed by T2, VC1 and VC2 (remember TR1 is untuned). The primary winding on T2 taking the output from TR1. The reason for having two variable tuning capacitors, one large, the other small, is that the large capacitor is the main tuning capacitor, while the small one is used as a *bandspread* control (for very fine tuning). The tuned signals are then detected (converted to audio frequencies) by TR3.

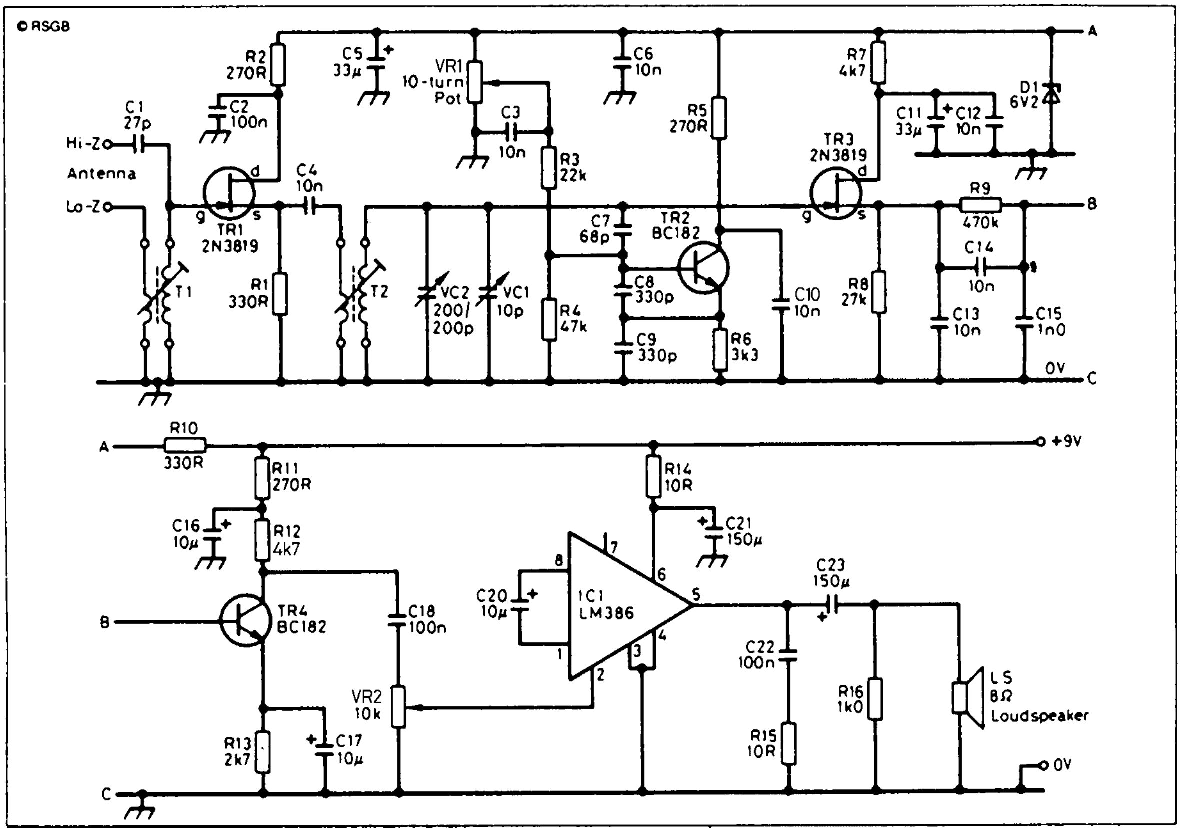
©RSGB
R2
270R
C5
33µ
VR1
10-turn
Pot
C6
10n
R7
4k7
A
D1
6V2
C2
100n
C1
27p
Hi-Z
Antenna
Lo-Z
C3
10n
R5
270R
C11
33µ
C12
10n
TR3
2N3819
C4
10n
R3
22k
d
g
s
R9
470k
B
TR1
2N3819
C7
68p
TR2
BC182
C14
10n
T1
R1
330R
T2
VC2
200/
200p
VC1
10p
R4
47k
C8
330p
C9
330p
R6
3k3
C10
10n
R8
27k
C13
10n
C15
1n0
0V
C
R10
330R
+9V
R11
270R
R14
10R
C16
10µ
R12
4k7
C21
150µ
C23
150µ
C20
10µ
IC1
LM386
8
7
6
5
4
3
2
1
TR4
BC182
C18
100n
C22
100n
VR2
10k
R16
1k0
LS
8Ω
Loudspeaker
R13
2k7
C17
10µ
R15
10R
0V

Figure 2 The circuit diagram showing the 10-turn potentiometer VR1. This is the regeneration control for the circuit

TR2 is a *Q-multiplier* stage. It is a Colpitts-type oscillator which uses C8 and C9 to give feedback and produce the oscillation. C7 couples the tuned circuit (and the input to TR3) to TR2. The tuned circuit thus controls the frequency of oscillation of TR2 and the signal passed on to TR3.

The secret of the ease of operation of this receiver is VR2, a 10-turn potentiometer (or *helipot*). The resistance wire is wound in the form of a helix, giving a much greater wire length than in a normal potentiometer, and the shaft must be turned ten times to cover the whole length. Helipots are very useful when very fine adjustments have to be made. Here, VR2 sets the regeneration (or *reaction*) level, depending on the type of signal you are receiving, as will be discussed in Part 2.

TR4 provides the first stage of audio amplification and, after the volume control, VR2, the audio amplifier integrated circuit, IC1, will drive a small loudspeaker or headphones.

In Part 2, the construction will be discussed, together with the choice of aerial, the parts list, and advice on using the receiver.

31 A fruit-powered medium-wave radio

Introduction

This is a one-evening project that will result in a working medium-wave (MW) radio, and will also teach you a little about the way electricity can be generated from the right metals and a little (safe) acid. All you need are three lemons or other citrus fruit, three pieces of copper and three pieces of zinc (or galvanised metal) for your power supply.

Construction

Figure 1 shows the circuit and **Figure 2** its layout on a simple ‘plug-in’ prototype board. The six pieces of metal are connected as shown, to wire the three lemons in series; use ordinary wire between each lemon and the next. If you have a meter to measure the total voltage, it should be about 1.8 V. Use a standard ferrite rod, and wind on it about 40 turns of single-conductor PVC-insulated wire.

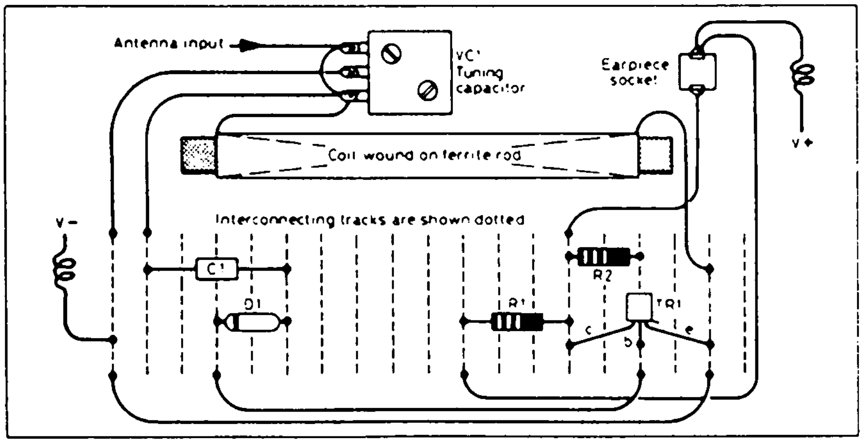
Antenna input
VC1
Tuning capacitor
Earpiece socket
V+
Coil wound on ferrite rod
Interconnecting tracks are shown dotted
V–
C1
D1
R1
R2
TR1
c
b
e

Figure 1 Most parts are plugged into the board as shown – soldering is not required

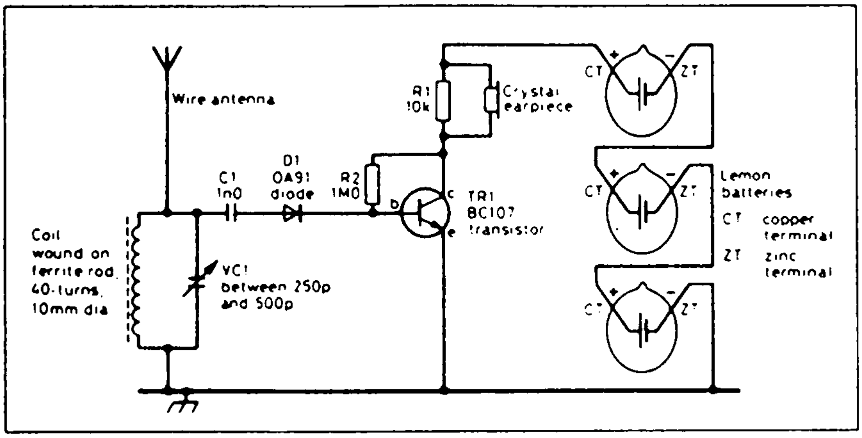
Wire antenna
Coil wound on ferrite rod, 40-turns, 10mm dia
C1
1n0
D1
OA91
diode
R2
1M0
VC1
between 250p and 500p
b
c
e
TR1
BC107
transistor
R1
10k
Crystal earpiece
+
–
CT
ZT
Lemon batteries
CT copper terminal
ZT zinc terminal

Figure 2 Three lemons power the radio, which gives good results for such a simple circuit

Wire up the circuit on the board as illustrated in Figure 2. Soldering is not required with this type of board – just plug in the components and the wires. Only one transistor is needed. The tuning capacitor, VC1, selects the station you want to hear, and D1 helps to remove the carrier from the RF signal. The resulting audio signal is fed to TR1, a small transistor audio amplifier, which makes the signal big enough to drive a crystal earpiece comfortably. Walkman-type earphones will not work, so invest in a crystal earpiece which you can use in several other projects, too! If you use a smaller capacitor than that specified for VC1, you will need more turns on the aerial coil.

Tests on the prototype indicated that the radio will run for about a week on three lemons!

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1	1 megohm (MΩ)
R2	10 kilohm (kΩ)

Capacitors

C1	1 nanofarad (nF) min. ceramic
VC1	250–500 picofarad (pF) variable

Semiconductors

TR1	BC107 npn (or BC108, BC109C)
D1	OA90, OA91 germanium (not silicon)

Coil

L1	2 metres of single-conductor insulated wire on a standard ferrite rod

Additional items

- Plug-in prototype board, e.g. Maplin YR84F
- Wire aerial at least 3 m long
- Crystal earpiece
- Three juicy fruits

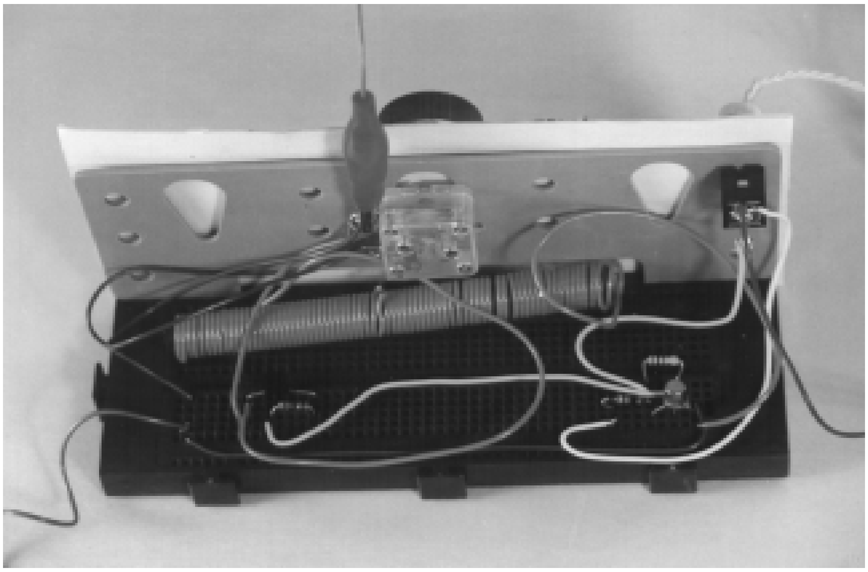

32 A capacitance bridge

Introduction

It is always valuable to have a means of measuring capacitors amongst your test gear. Universal LCR bridges (i.e. systems that can measure inductance, capacitance and resistance) can be found at most rallies, and you can work your way up to owning one of these. In the meantime, a piece of home-made equipment gives you experience as well as resulting in a useful measuring instrument.

The circuit described here is called a *capacitance bridge*, because it balances the effects of one resistor/capacitor pair against another; if one capacitor has an unknown value, then the other can be calculated. The basic bridge circuit is shown in **Figure 1**. To avoid having calculations to perform, this instrument will be calibrated by using capacitors of known values. The bridge is a useful way of performing measurements, because a knob is turned until there is a null in the signal from an earpiece or loudspeaker. The ear is very precise in being able to perceive nulls, which makes the bridge easy to use and reasonably accurate. At the null, the bridge is said to be *balanced*.

How does it work?

Figure 2 is the circuit diagram for this capacitance bridge. Transistors TR1 and TR2 form an oscillator. This is the audio oscillator shown in Figure 1, and produces an alternating voltage which is fed to the bridge. RV1 (in the collector lead of TR2) replaces both R1 and R2 in Figure 1 – that part of

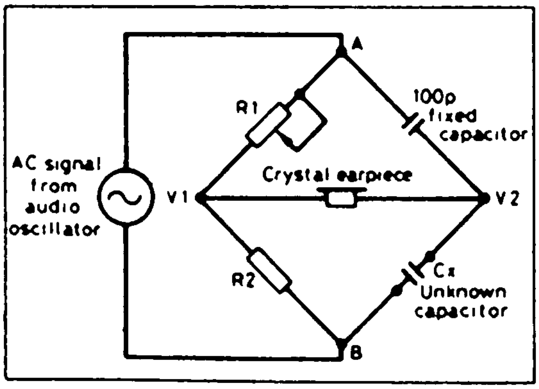

Figure 1 Simplified circuit of a capacitance bridge. R1 is adjusted for minimum sound

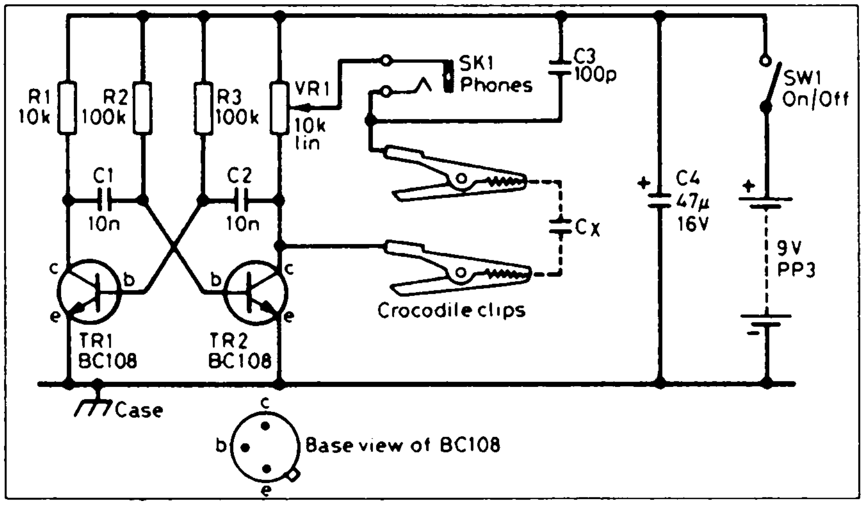
R1
10k
R2
100k
R3
100k
VR1
10k
lin
SK1
Phones
C3
100p
SW1
On/Off
C1
10n
C2
10n
C4
47µ
16V
Cx
9V
PP3
c
b
e
TR1
BC108
TR2
BC108
Crocodile clips
Case
Base view of BC108

Figure 2 Transistors TR1 and TR2 give an audio signal which is adjusted by variable resistor RV1

RV1 *above* the wiper represents R1, while that part *below* the wiper represents R2. The voltage on one side of the earpiece is determined by the ratio of these values, and is adjusted by rotating RV1. The voltage on the other side of the earpiece is determined by the ratio of C3 to Cx, where Cx is the unknown-valued capacitor. When these two voltages are the same, the bridge is balanced and there is no current through the earpiece.

Figure 3 shows the layout of the components on the matrix board of the prototype. It measures 10 holes by 14 holes, and is of the plain type, i.e. no copper strips. All earth connections are taken to a single solder tag. When

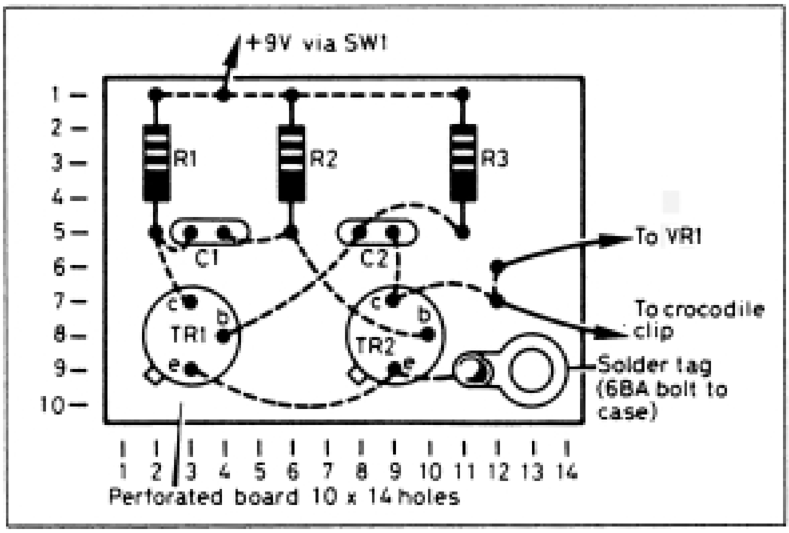
+9V via SW1
1
2
3
4
5
6
7
8
9
10
R1
R2
R3
C1
C2
To VR1
To crocodile clip
TR1
TR2
c
b
e
Solder tag (6BA bolt to case)
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Perforated board 10 x 14 holes

Figure 3 Components are soldered together on a small prototype board

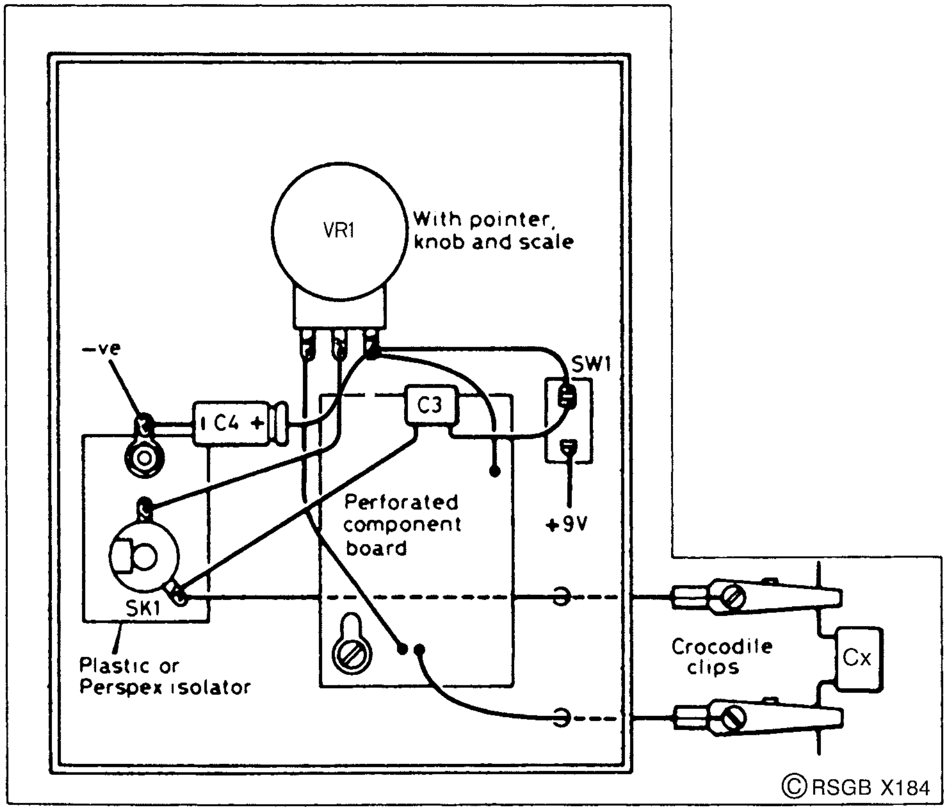
VR1
With pointer, knob and scale
−ve
C4 +
C3
SW1
Perforated component board
+9V
SK1
Plastic or Perspex isolator
Crocodile clips
Cx
©RSGB X184

Figure 4 Layout of the parts inside an aluminium box. The earpiece socket is insulated from the case

mounting the board inside a metal box, a long screw is used with extra nuts, to earth the board to the case while providing a stand-off, thus preventing any unwanted short-circuits between the board connections and the case. Be aware that a *crystal* earpiece must be used; low-impedance headphones of the Walkman type are not satisfactory. Solder two flying leads as shown in **Figure 4**, about 15 cm long, terminated in small crocodile clips for attaching to the unknown capacitor.

Calibration

After checking the circuit carefully, attach the battery and switch on; a buzzing noise should be heard in the earpiece. This is the first sign that everything should be OK. If there is no buzz, switch off and recheck the connections.

You will now need a range of close-tolerance (1%) silver-mica capacitors covering the range 10 to 1000 picofarads (pF). Arrange the capacitors in ascending order and connect them to your bridge in sequence. After having prepared a neat piece of card or paper mounted behind the knob on the front panel, mark the dial at the positions of the nulls for all the capacitors. You have now calibrated your capacitance bridge. If you are more likely to want to measure larger capacitors, replace C3 by a 1 nanofarad (nF) capacitor, and the bridge will measure up to 10 nF approximately.

Parts list

Resistors: all 0.25 watt, 5% tolerance	
R1	10 kilohms (kΩ)
R2, R3	100 kilohms (kΩ)
RV1	10 kilohms (kΩ) linear
Capacitors	
C1, C2	10 nanofarads (nF) ceramic
C3	100 picofarads (pF) silver mica or polystyrene
C4	47 microfarads (μF) 16 V electrolytic
Semiconductors	
TR1, TR2	BC108 npn
Additional items	
SW1	SPST on/off switch
	Battery connector, PP3 type
	Earpiece, crystal type
SK1	3.5 mm jack socket for earpiece
	Battery, 9 V PP3
	Matrix board, approx. 10 holes by 14 holes
	Aluminium case, approx. 10 × 8 × 5 cm

33 A simple short-wave receiver – Part 2

Introduction

In Part 1, the design of this receiver was discussed in some detail. Now we are going to put it all together and see how it works. The receiver is laid out on a printed-circuit board (PCB) or on a matrix board.

Construction

The layout of the components is shown in **Figure 1**. Identify each part separately, insert it into the holes in the board and solder carefully. Long leads may be cropped before or after soldering, depending on your skill and preference. All electrolytic capacitors, T1, T2 and IC1, must be connected correctly. The front-panel controls are connected to the PCB terminals shown in Figure 1.

The layout of the controls and the placing of the board inside the case are matters of personal preference. The size of the prototype front panel is shown in **Figure 2**. The prototype had a slow-motion drive fitted to VC2, the main tuning capacitor. This required the capacitor to be fitted on its own small panel. The bandspread control does not need any form of

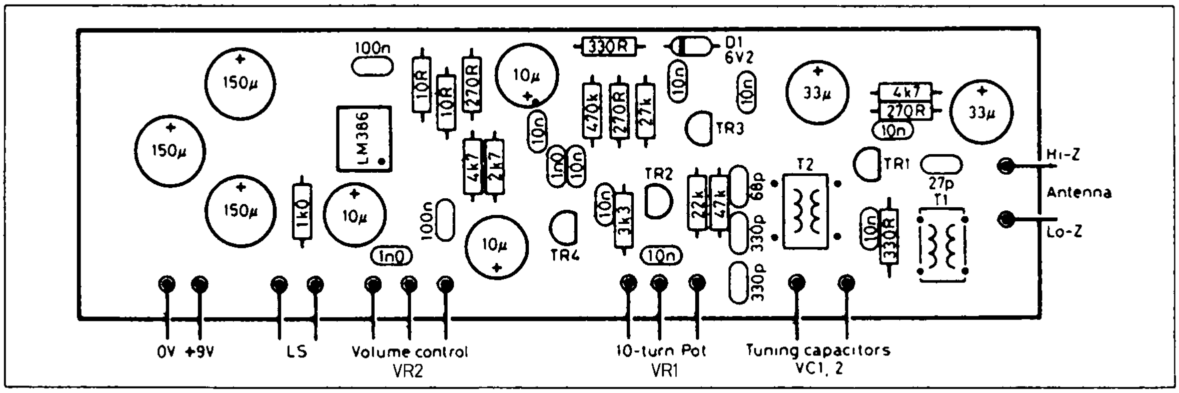
100n
330R
D1
6V2
150μ
150μ
150μ
10R
10R
270R
10μ
470k
270R
27k
10n
10n
33μ
4k7
270R
10n
33μ
LM386
4k7
2k7
10n
1n0
10n
TR3
TR2
T2
TR1
27p
T1
Hi-Z
Antenna
Lo-Z
1k0
10μ
100n
10μ
1n0
TR4
10n
3k3
10n
22k
47k
68p
330p
330p
10n
330R
0V
+9V
LS
Volume control
VR2
10-turn Pot
VR1
Tuning capacitors
VC1, 2

Figure 1 Layout of the receiver circuit board

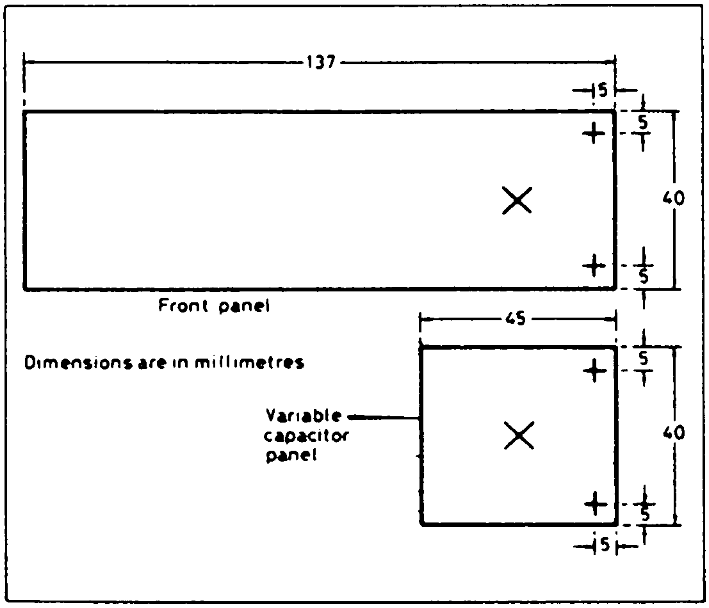

Figure 2 The case can be metal or made from PCB material

slow-motion drive. A tuning dial may be fitted behind the tuning knob on VC2 if required. The volume control, VR2, has its own on/off switch for the battery, which can be mounted behind the back panel, on which are mounted the aerial input sockets and the headphone/speaker socket. Check your soldering carefully, together with all the wiring between the board, the potentiometers and variable capacitors. When you are confident that everything is perfect, connect the battery and switch on.

The first tests

If you have a good aerial and are using an aerial tuning unit (ATU), use the low-impedance input. Excellent results are possible, though, with about 3 metres of wire connected to the high-impedance input. Set the volume control to give a gentle hiss in the headphones or speaker. Advance the reaction control to give a definite hiss. As you tune in an AM station, the hiss will change to a whistle; back off the reaction until the oscillation *just* stops. The receiver is now correctly set for AM reception.

On the amateur bands, the stations will be SSB or CW, and the reaction needs to be set *just above* the oscillation point, and will need slight adjustment from time to time for different qualities of signal. Juggling with the volume, reaction and tuning is part of the pleasure of using regenerative receivers!

In action

Practice is needed for best results. The regenerative receiver is renowned for its versatility in being able to be set up *exactly* for all types and strengths of signal. The basic receiver tunes from 6.5 MHz to 11 MHz approximately; this includes two amateur bands at 7.0 MHz and 10.1 MHz and two broadcast bands.

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1, R10	330 ohms (Ω)
R2, R5, R11	270 ohms (Ω)
R3	22 kilohms (kΩ)
R4	47 kilohms (kΩ)
R6	3.3 kilohms (kΩ)
R7, R12	4.7 kilohms (kΩ)
R8	27 kilohms (kΩ)
R9	470 kilohms (kΩ)
R13	2.7 kilohms (kΩ)
R14, R15	10 ohms (Ω)
R16	1 kilohm (kΩ)
VR1	10 kilohm (kΩ) linear 10-turn potentiometer
VR2	10 kilohm (kΩ) log potentiometer

Capacitors

C1	27 picofarads (pF)
C2, C18, C22	100 nanofarads (nF)
C3, C4, C6, C10, C13, C14	10 nanofarads (nF)
C5, C11	33 microfarads (μF) 16 V electrolytic
C7	68 picofarads (pF)
C8, C9	330 picofarads (pF)
C15	1 nanofarad (nF)
C16, C17, C20	10 microfarads (μF) 16 V electrolytic
C21, C23	150 microfarads (μF) 16 V electrolytic
VC1	10 picofarads (pF) variable
VC2	200 picofarads (pF) variable

Semiconductors

TR1, TR3	2N3819
TR2, TR4	BC182
IC1	LM386
D1	6.2 V 0.5 W Zener

Inductors

T1	3 turns primary, 15 turns secondary, wound on 2-hole ferrite bead, with 28 SWG wire
T2	2 turns primary, 17 turns secondary, wound on a T68-2 toroidal former, with 28 SWG wire

Additional items
- Printed-circuit board
- Battery connector
- PP3 battery
- 8-pin DIL socket
- 8 ohm speaker or headphones (Walkman type)
- Case to suit

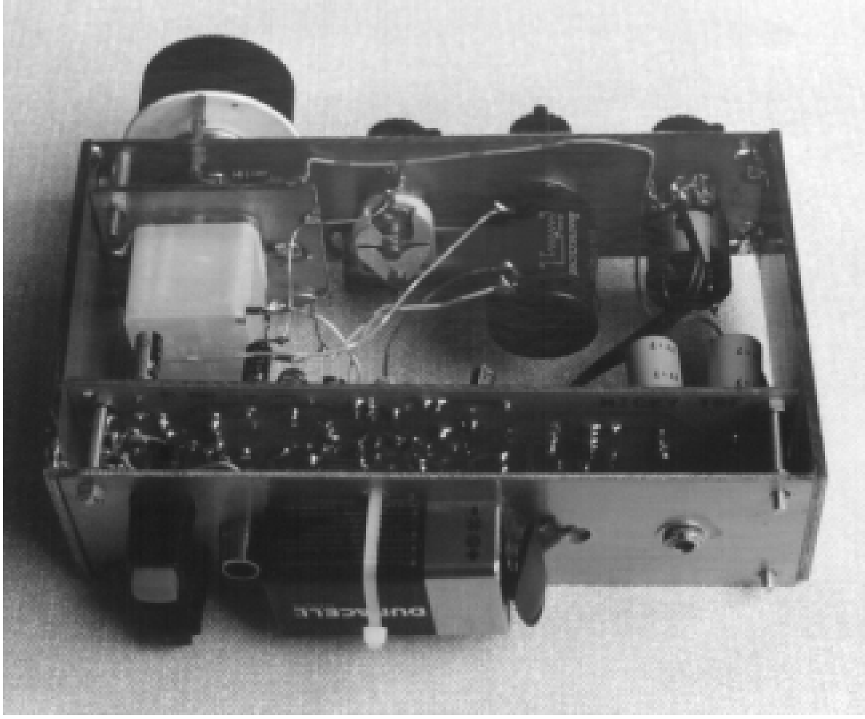

34 A basic continuity tester

Introduction

This little device can be built on a plug-in breadboard, and is ideal for testing fuses and cables, as well as doubling as a signal source for testing amplifiers, etc. It can even be used for testing npn transistors by replacing either of the transistors in the circuit, and seeing if the circuit still works!

Simple and quick to build

Using a plug-in breadboard, this circuit is so simple it could almost be built when you need it, and then dismantled again! If you want the circuit to make a different sound, the components to change are R2, R3, C1 and C2. Always make sure that R2 = R3 and C1 = C2, or the sound may be excessively ‘edgy’ and lacking in volume.

Basically, the circuit is an oscillator which drives a little loudspeaker directly. The fuse or cable you are testing is connected between the crocodile clips. If there is a current path between the two clips, the current also flows through the circuit, thus operating the oscillator and producing a sound from the loudspeaker.

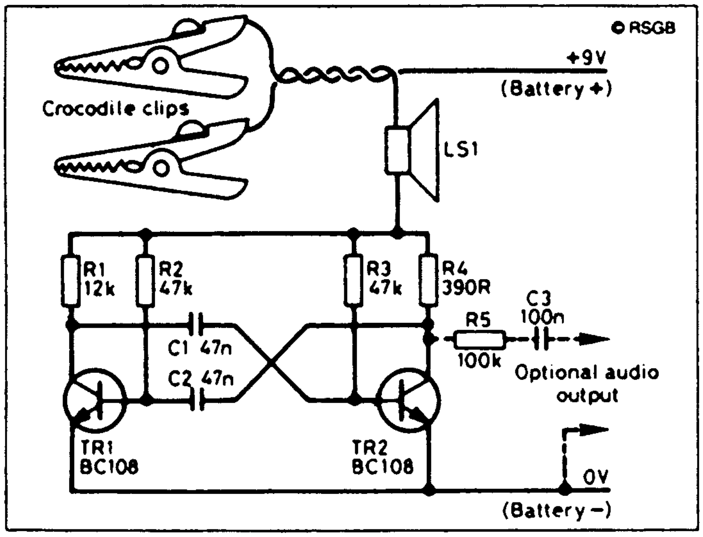
© RSGB
+9V
(Battery +)
Crocodile clips
LS1
R1
12k
R2
47k
R3
47k
R4
390R
C3
100n
R5
100k
C1 47n
C2 47n
Optional audio
output
TR1
BC108
TR2
BC108
0V
(Battery −)

Figure 1 The circuit diagram of the continuity tester

If you want to use the circuit to produce a signal to test an amplifier, for instance, connect the croc clip which comes directly from the battery to the junction of R1, R2, R3 and R4. This supplies current to the circuit while, at the same time, bypassing the loudspeaker. Use the two components R5 and C3 between the oscillator and your amplifier. The loudspeaker is bypassed because you will want to listen to the output from your amplifier, *not* from the oscillator!

Remember, NEVER test any equipment which is still connected to the mains electricity supply. Avoid testing anything which is switched on and has its own power supply. You may damage both your tester ***and*** the circuit you are 'testing'!

Parts list

Resistors: all 0.25 watt, 5% tolerance	
R1	12 kilohms (kΩ)
R2, R3	47 kilohms (kΩ)
R4	390 ohms (Ω)
R5	100 kilohms (kΩ) optional
Capacitors	
C1, C2	47 nanofarads (nF) ceramic
C3	100 nanofarads (nF), 0.1 microfarads (μF) ceramic optional
Semiconductors	
TR1, TR2	BC108 npn
Additional items	
LS1	Loudspeaker (8 Ω)
	2 crocodile clips
	Prototype circuit board
	PP3 battery connector
	PP3 battery

35 A charger for NiCad batteries

Introduction

The cost of replacing dry batteries can be alleviated to a great extent by the use of rechargeable batteries such as those containing nickel and cadmium (called *NiCads*). They are more expensive than the batteries they replace, but they can be charged hundreds of times and thus prove cheaper in the long run.

NiCads produce only 1.2 V per cell, compared with the 1.5 V of ordinary cells, so before you rush out and buy lots of these, please make sure that the equipment on which you plan to use them will work on the reduced voltage! For example, if you are using four cells to give you a 6 V supply, NiCads will give you only 4.8 V, which is quite a reduction. Using six cells to replace a 9 V battery will give you only 7.2 V. Not all equipment is happy with these reductions!

However, we all use them when we can, and they save substantial amounts of money. Here is the circuit of a charger to keep them in prime condition.

Charging NiCads – the ampère-hour

NiCads require charging at constant current, which means that connecting one across a normal power supply (constant *voltage*) is useless and can destroy it. They need pampering to the extent of needing a long charge (around 16 hours) at a rate dependent upon the *capacity* of the battery. By the capacity of a battery, we mean how much energy it can store. You will probably know that energy is measured in joules. For the purposes of storing energy in batteries, the joule is *not* the ideal unit, so we use one that is! This unit is the ampère-hour (Ah), and must be interpreted with some realism. For example, if the battery is rated at 2 Ah, it will deliver a current of 0.5 amp for 4 hours, or 0.25 amp for 8 hours. Provided the current is not too high, the product of the current (in amps) and the time for which it will flow (in hours) before the battery is flat will always be around 2 Ah. A workable 'rule of thumb' for calculating the charging current is that its value should be around one-tenth of the numerical value of the capacity; so, for our 2 Ah battery, a charging current of around 200 mA (2 ÷ 10 = 0.2 A or 200 mA) would be used.

Constant voltage to constant current

Many integrated circuit (IC) chips are available for use as voltage regulators, i.e. they supply a constant voltage. Most of these can be persuaded to become constant current supplies with one external resistor!

The voltage regulator IC usually has only three connections – 'input', 'output' and 'common'. It is designed (in the case of the LM7805) to produce a constant 5 V output between the 'output' and the 'common' connections, at currents of up to 1 A. If a resistor is connected between these, the IC will maintain 5 V across it. If you look at **Figure 1**, you will see the circuit performing this conversion.

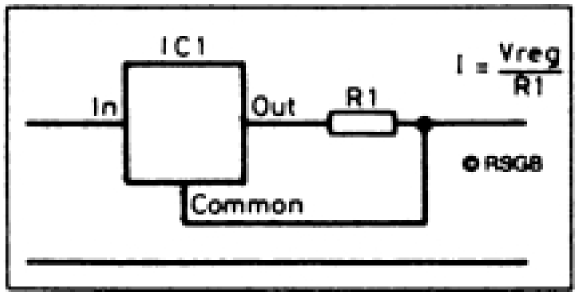

Figure 1 Voltage regulator arranged to produce a constant current

For the previous example, we derived a charging current of 200 mA, so we now need to calculate the value of resistor that will produce this current. Using the equation which is derived from Ohm's law:

$$I = \frac{V}{R}, \text{ from which } R = \frac{V}{I},$$

where R = R1, the resistance in ohms that we are calculating,
V is the voltage across R1 (5 V), and
I is the current flowing (200 mA).

So,

$$\text{R1} = \frac{5}{0.2} = 25\,\Omega.$$

25 ohms is not a 'common' or 'preferred' resistor value, so we must choose the next *largest* value, which is 27 Ω. This reduces the current, but only slightly – it is now 185 mA. When calculating resistor values in power supply circuits, we must *always* check the power that they dissipate and make sure we specify and fit a suitable resistor.

Power (in watts) is the product of the voltage across and the current through a device, so in this case it is given by:

$$\text{Power} = V \times I = 5 \times 0.185 = 0.925\,\text{W}.$$

Rather than use a 1 watt resistor operating very near its limit, it is safer to use a 2 watt resistor operating well within its limits.

Looking again at Figure 1, we now have a constant-current source producing 185 mA, when R1 is a 27 Ω, 2 W resistor. For use with NiCads requiring charging currents other than 200 mA, you will need to repeat the two equations above, using a new value for *I*.

The full circuit and its assembly

This is shown in **Figure 2**, and can be broken down into two parts. The first is the AC to DC conversion produced by the mains transformer, T1, the bridge rectifier, BR1, and the smoothing capacitor, C1. The second is the constant-current section already discussed.

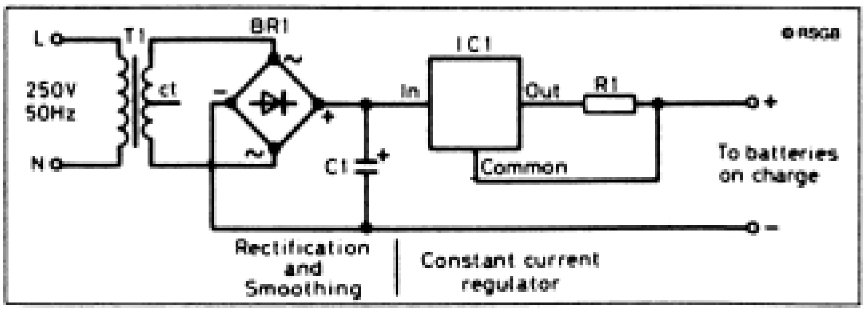

Figure 2 The complete circuit including mains components

The prototype was assembled on matrix board measuring 18 holes by 12 strips, although, as **Figure 3** shows, this is much larger than is strictly necessary. No strip cutting is needed, but make sure that IC1 and C1 are inserted correctly.

Warning! Before you attempt to wire up the transformer and the bridge rectifier, be aware that you will eventually be connecting the circuit to the mains supply, so there are three possibilities for you: (1) get a qualified friend to supervise your completion and testing of the circuit; (2) get your qualified friend to complete and test the circuit for you; (3) replace the transformer and bridge rectifier with a mains adapter.

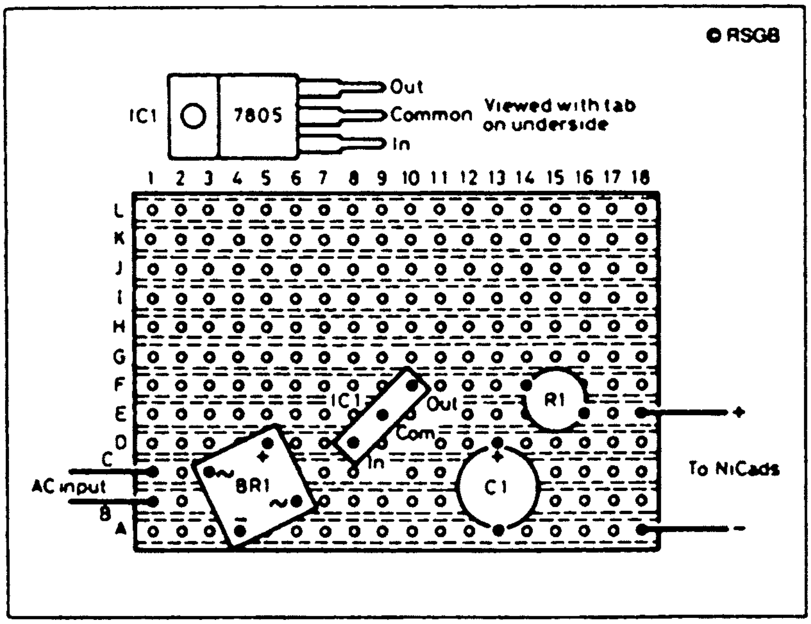

Figure 3 The charger can be built on Veroboard

If you decide to use the mains adapter, its output is connected directly across C1, because T1 and BR1 are now no longer needed. Make sure the polarity (positive and negative) is correct and that the adapter output is set to 12 V.

A quick test

If you have built the mains version, make sure all connections are correct, and that there are no soldered joints which will touch other parts of the circuit. The box must be securely closed before tests begin. The RSGB cannot be held responsible for damage to equipment or batteries! The version using the mains adapter need not be closed during tests.

Switch on. With nothing connected to the output, the unit should run cold. If this is not the case, switch off and recheck your circuit. If all is well, switch on again and connect a DC multimeter (on the current range) across the output. It should indicate only a slight difference from the calculated value of 185 mA. You can now charge your NiCads!

Other charging currents can be set by having different values for R1, perhaps selectable by a rotary switch. Remember to make sure that the values of both resistance *and* power dissipation are correct, and don't exceed the 250 mA rating of the transformer (or the 1 amp rating of the IC if you are using a bigger transformer).

Parts list

		Maplin code
Resistor		
R1	As required – see text	
Capacitor		
C1	1000 microfarads (μF) 35 V electrolytic	FF18U
Semiconductors		
IC1	LM7805 5 V, 1 A regulator	CR14Q
BR1	W005 50 V, 1 A full-wave rectifier	QL37S
Transformer		
T1	9–0–9 V, 250 mA sub-min. transformer	YN15R
Additional items		
Case		FG41U
Veroboard, 18 holes by 12 strips		
Plug to suit NiCads		
Double-sided sticky tape as required		
Insulated wire for battery connections		

36 An 80 metre crystal-controlled CW transmitter

Introduction

A simple transmitter is ideally suited to anyone venturing into our marvellous hobby for the first time. If you are put off by the complexities and prices of the 'black boxes' then this is the transmitter you've been looking for!

The circuit

The complete circuit is shown in **Figure 1**. It is a single-transistor crystal oscillator which is *keyed* (switched on and off) by the Morse key in the +12 V supply rail. The circuit first appeared in *QST* (the American equivalent of *RadCom*) in 1982, and has since appeared in a modified form in the *ARRL Handbook*. The circuit can produce about 2 watts on 80 m, and can be constructed on a piece of plain matrix board; the prototype board measured 7 cm by 3 cm, and its layout is shown in **Figure 2**.

Note that a bare copper wire runs along the bottom edge of the board to act as an earth wire for the relevant components. The only 'difficult' part of the construction is the winding and wiring of T1. The main winding is 38 turns of 26 SWG enamelled copper wire, and there are two link windings of four turns each. Make sure that all windings are wound the correct way round the toroid – Figure 1 shows this and should be studied carefully. If the windings do not have the correct *sense* (i.e. a clockwise coil has been wound anticlockwise, or vice versa), or have been connected incorrectly to the rest of the circuit, the oscillator will not work!

When putting components on the board, wire in the crystal socket *without the crystal in it*. Crystals do not like to be subjected to the horrors of a soldering iron, so keep your crystal to one side during the construction process!

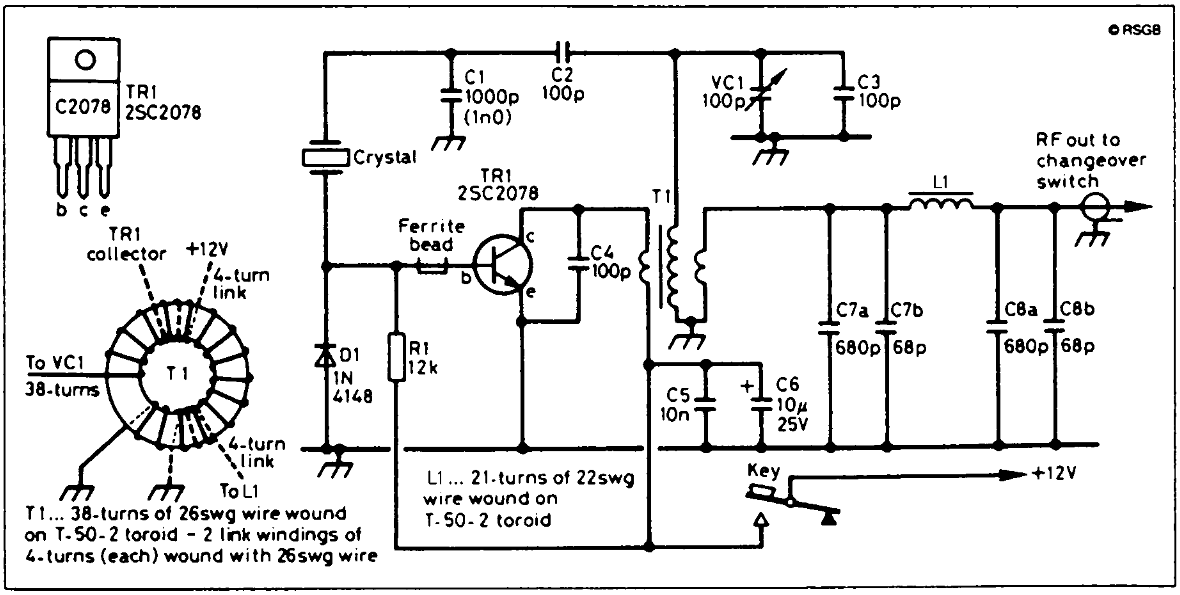
© RSGB
C2078
TR1
2SC2078
b c e
TR1
collector
+12V
4-turn
link
To VC1
38-turns
T1
4-turn
link
To L1
T1... 38-turns of 26swg wire wound
on T-50-2 toroid – 2 link windings of
4-turns (each) wound with 26swg wire
C1
1000p
(1n0)
C2
100p
Crystal
TR1
2SC2078
Ferrite
bead
b
c
e
C4
100p
D1
1N
4148
R1
12k
L1 ... 21-turns of 22swg
wire wound on
T-50-2 toroid
T1
VC1
100p
C3
100p
C5
10n
+
C6
10µ
25V
Key
+12V
L1
C7a
680p
C7b
68p
C8a
680p
C8b
68p
RF out to
changeover
switch

Figure 1 The simplest form of transmitter is a keyed crystal oscillator. Note that L1, C7 and C8 make up a *low-pass filter* which reduces unwanted *harmonics* (outputs at the transmitter frequency multiplied by 2, 3, 4, etc.)

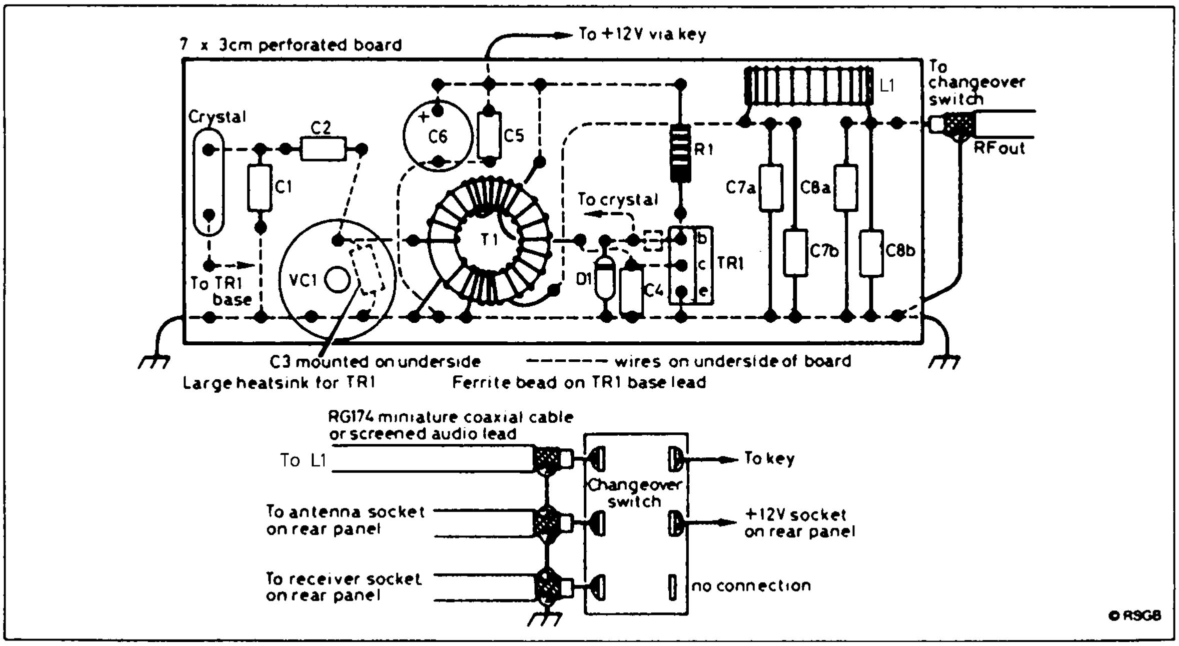
7 x 3cm perforated board
To +12V via key
Crystal
C2
C6
C5
R1
L1
To changeover switch
RFout
C1
To crystal
C7a
C8a
T1
TR1
b
c
e
C7b
C8b
To TR1 base
VC1
D1
C4
C3 mounted on underside
wires on underside of board
Large heatsink for TR1
Ferrite bead on TR1 base lead
RG174 miniature coaxial cable or screened audio lead
To L1
To key
Changeover switch
To antenna socket on rear panel
+12V socket on rear panel
To receiver socket on rear panel
no connection
© RSGB

Figure 2 Component layout is straightforward on a 'matrix board' (it has holes but no copper strips). The dotted lines are the connecting wires. The lower part shows the connections for the transmit/receive switch

The output filter, which comprises C7, C8 and L1, is a *low-pass* filter, which helps reduce any *harmonics* present in your signal. Harmonics are integral multiples of your transmitter frequency, so if you are transmitting on a frequency, *f*, harmonics will be present at frequencies 2*f*, 3*f*, 4*f*, . . . and so on. L2 is another inductance using 22 SWG enamelled copper on a ferrite toroid. The changeover switch is external to the transmitter board, and is used to switch your aerial between the transmitter and the receiver; its wiring is shown in Figure 2.

Use a dummy load

A dummy load enables you to test your circuit without actually transmitting a signal. If you haven't such a thing already, it is easy to construct one to use with this transmitter. Don't use it for transmitters of more than 2 watts output, though. Use two 100 ohm, 1 watt resistors, connected in parallel across the end of a short piece of coaxial cable, terminated in a BNC, PL259 or N-type free plug, as shown in **Figure 3**. Plug this into the aerial socket on your transmitter, plug in your crystal and connect the transmitter to a 12 V supply. Have another receiver switched on and tuned to the crystal frequency. Although the radiation from your dummy load is minimal, it will

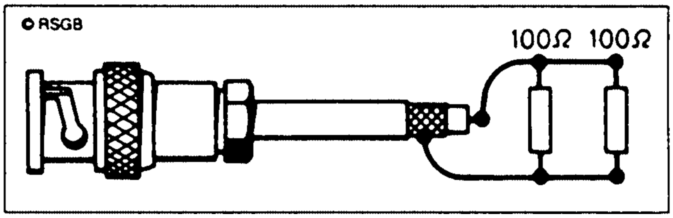

Figure 3 A simple 2 W dummy load can be made from two 100 Ω (ohm) resistors in parallel. The plug should match the socket on your transmitter

be enough to be picked up by a receiver in the same room. Send dashes with the Morse key, and adjust VC1 until the received note is *clean*. It should not sound rough, or have a *chirp* (change its frequency during a dash or dot). Avoid tuning for maximum power; this is seldom the correct setting!

You will need to put your completed transmitter in a metal box, using sockets for the power supply, aerial, receiver and Morse key. The sockets can be chosen to match your existing equipment.

Figures 4–7 are taken from the RSGB book *Practical Antennas for Novices*, and may give you some ideas on the type of aerial to be used with your transmitter.

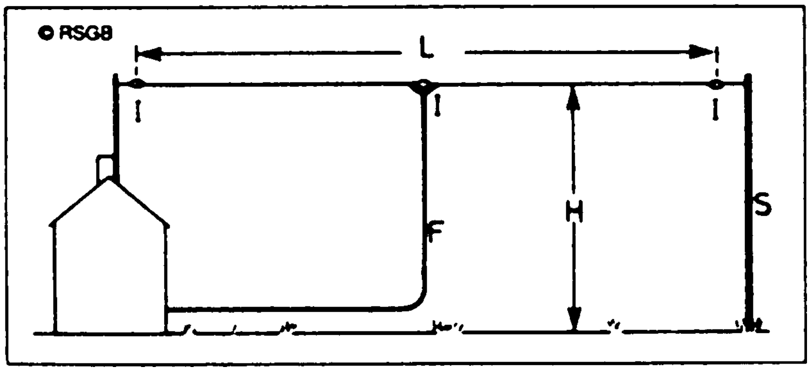

Figure 4 A simple dipole can be very effective. For the 3.5 MHz band, length L is 40 metres and height H should be as large as possible. The far support S can be a tree, pole or building. Insulators I may be home made from strong plastic and the feeder F should be 50 Ω (ohm) coax cable

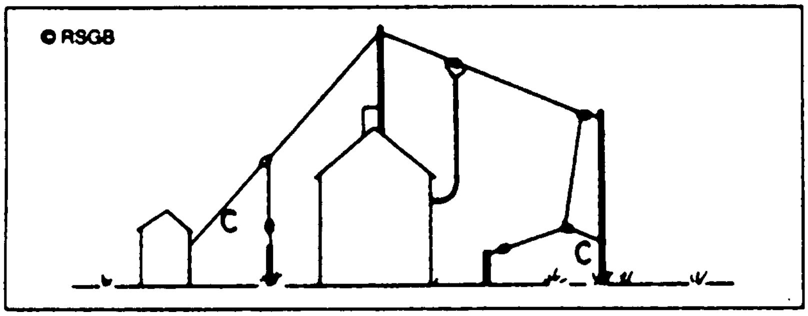

Figure 5 Your signal is radiated mostly from the centre of the dipole so the ends can droop or even be bent but the length may need shortening by a few centimetres because the ground and the bends will detune the dipole. Cords C are best made from strong plastic rope from a sailing or camping shop

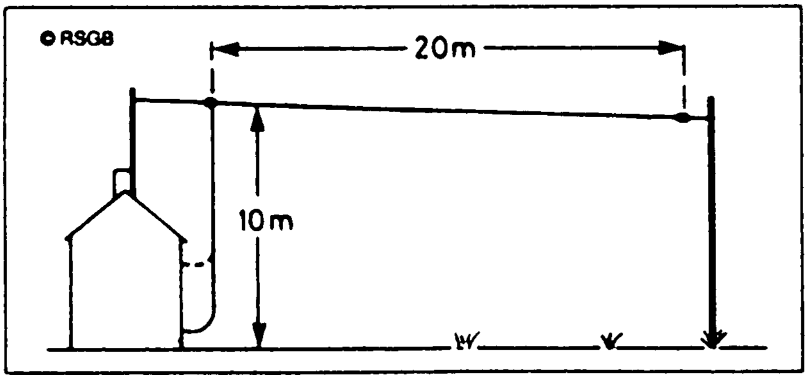
© RSGB
20m
10m

Figure 6 An 'inverted-L' takes up less space than a dipole and doesn't need coax cable. Like the dipole, the end can droop or be bent to save space as in this case most of the radiation comes from the area around the top of the vertical part

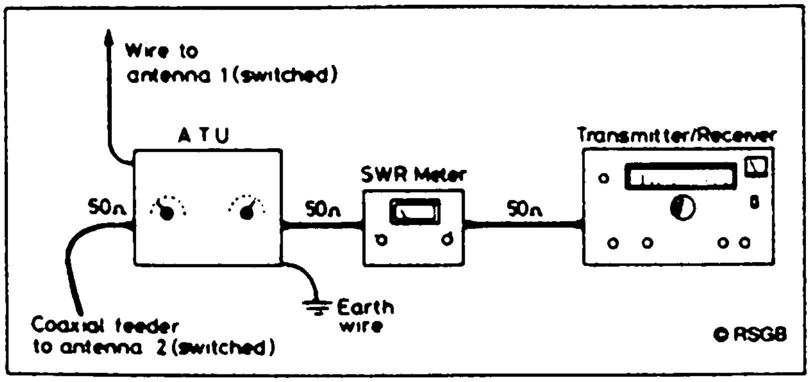
Wire to antenna 1 (switched)
ATU
50Ω
50Ω
SWR Meter
50Ω
Transmitter/Receiver
Earth wire
Coaxial feeder to antenna 2 (switched)
© RSGB

Figure 7 Almost any length of wire more than 10 m or so long will work (though it will work better the longer and higher it is) but an Aerial System Tuning Unit (ASTU or ATU) will be needed

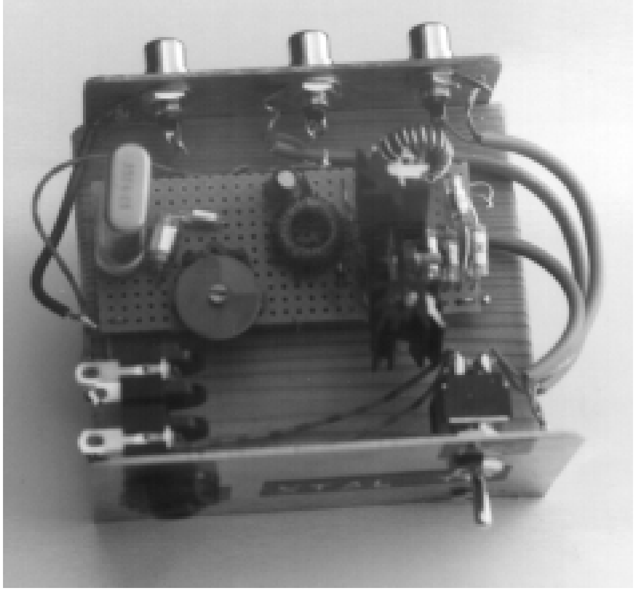

Parts list

Resistor

R1	12 kilohms (kΩ), 0.25 watt, 5% tolerance

Capacitors

C1	1000 picofarads (pF) polystyrene
C2	100 picofarads (pF) polystyrene
C7a, C8a	680 picofarads (pF) polystyrene
C7b, C8b	68 picofarads (pF) polystyrene
C3, C4	100 picofarads (pF) ceramic
C5	10 nanofarads (nF) polyester
C6	10 microfarads (μF) electrolytic 25 V
VC1	100 picofarads (pF) trimmer

Semiconductors

D1	1N4148
TR1	2SC2078 (see sources list)

Inductors

T1	38 turns 26 SWG enamelled copper on T–50–2 toroid, with two link windings of four turns
L1	21 turns 22 SWG enamelled copper on T–50–2 toroid

Additional items

Ferrite bead
Crystal (e.g. 3.579 MHz) and holder
Metal box
Socket for Morse key.
This must be totally isolated from the metal of the box, as *both* connections can be at +12 V.
Sockets for 12 V supply, aerial and receiver
Switch – DPDT
Heat sink for TR1
RG174 miniature coaxial cable for signal leads (see Figure 2)

Component sources

Special components

2SC2078	Cricklewood Electronics Ltd, 40 Cricklewood Broadway, London, NW2 3ET.

37 A solar-powered MW radio

Introduction

What could be more ecologically friendly than a radio powered by the sun's energy? This design is quite standard, and if you have built any of the other medium-wave radios in this series, then this one should present few problems.

The solar panel

The solar panel is to the solar cell as the battery is to the cell; in other words a solar panel is several solar cells connected in series. The solar panel quoted for this radio will generate about 9 V at a current of around 30 mA on a sunny day. The circuit will operate on a supply of around 2 V, so bright sunshine is *not* necessary for satisfactory operation. The volume will be less, of course.

The circuit

The radio uses the ZN415E integrated circuit (IC), connected as shown in **Figure 1**, the complete circuit diagram. The signal is tuned in by the combination of L1 and VC1. L1 is made by winding about 35 turns of 24 SWG enamelled copper wire on a ferrite rod. A standard ferrite rod of 10 cm length and 1 cm diameter is used.

Signals selected by the tuned circuit are passed to IC1, which amplifies the signals and removes the audio component, which is then amplified further by IC2 for driving a small loudspeaker. The removal of the audio component is the process we call *detection*. In addition to this, IC1 provides *automatic gain control* (AGC), which helps to keep the audio signal constant, even when the incoming RF signal may vary due to fading.

The prototype board

Veroboard (also known as matrix board or stripboard) is ideal for the construction of the radio. The layout is shown in **Figure 2**. The board size is 11 strips by 30 holes (**please note that there is no row 'I', so take care with your counting!**). Using a 3 mm (⅛ inch) twist drill rotated between

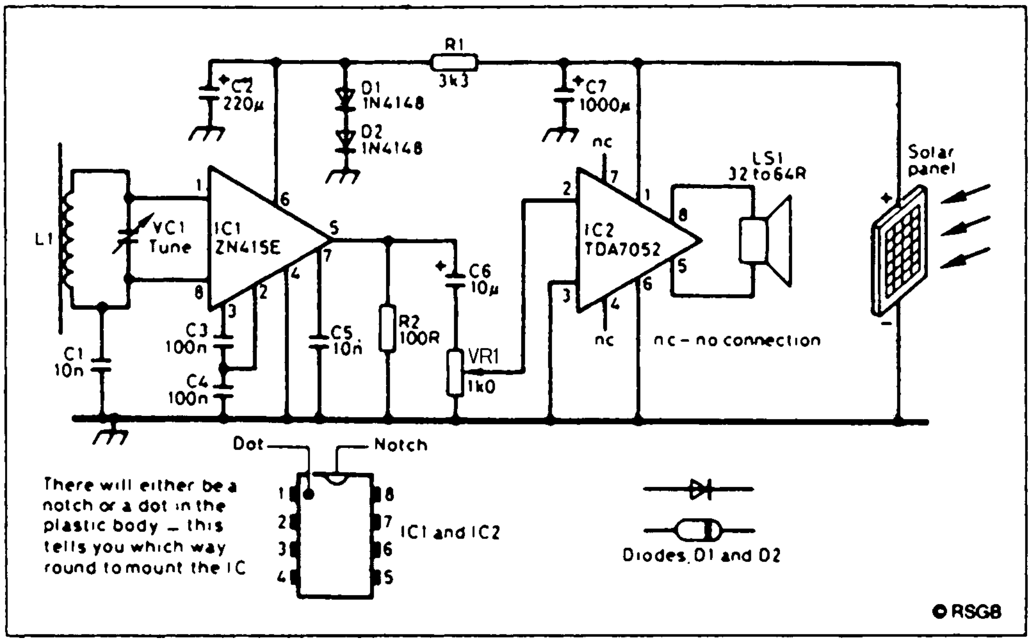
R1
3k3
C2
220µ
D1
1N4148
D2
1N4148
C7
1000µ
nc
LS1
32 to 64R
Solar panel
L1
VC1
Tune
IC1
ZN415E
IC2
TDA7052
C6
10µ
C1
10n
C3
100n
C4
100n
C5
10n
R2
100R
VR1
1k0
nc
nc – no connection
Dot
Notch
There will either be a notch or a dot in the plastic body – this tells you which way round to mount the IC
IC1 and IC2
Diodes D1 and D2
© RSGB

Figure 1 The radio uses just two integrated circuits (chips) and operates at any voltage from 2 to 9 V

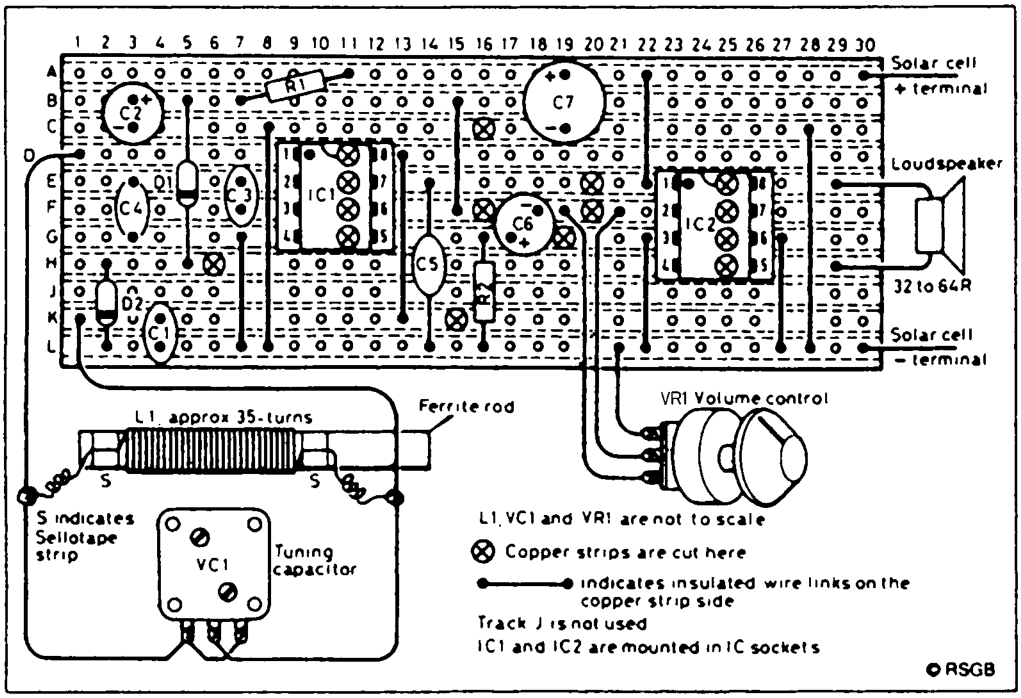
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
A
B
C
D
E
F
G
H
J
K
L
R1
C2
C7
D1
C4
C3
IC1
IC2
C6
C5
R2
D2
C1
Solar cell + terminal
Loudspeaker
32 to 64R
Solar cell – terminal
Ferrite rod
L1 approx 35-turns
S
S
VR1 Volume control
S indicates Sellotape strip
VC1
Tuning capacitor
L1, VC1 and VR1 are not to scale
⊗ Copper strips are cut here
●——● indicates insulated wire links on the copper strip side
Track J is not used
IC1 and IC2 are mounted in IC sockets
© RSGB

Figure 2 Follow the layout carefully, making sure that all connections are neatly soldered

your thumb and forefinger, break the strips at the points marked with a cross inside a circle. Hold the board up to the light to ensure that the tracks *have* been broken and that there is no copper swarf shorting tracks together.

Firstly, solder in the 8-pin DIL sockets for the ICs, making sure that the notches in the sockets are facing upwards, as shown on the diagram. Then, solder in the wire links, resistors, capacitors and diodes, making sure that the electrolytic capacitors (C2, C6 and C7) and the diodes (D1 and D2) are connected the correct way round. Use different colours of wire for the connections to the volume control, VR1, to avoid incorrect connections. Note the wiring of the tuning capacitor (VC1) shown in Figure 2; a two-section type is used, and both sections are wired in parallel to give twice the capacitance of a single section.

There is no on/off switch – just turn the volume down when you are finished using the set! The solar panel can be mounted parallel to the top of the case, or angled to receive the maximum energy from the sun, as shown in the photograph. You could have a battery available as a standby source to use the radio after dark; any battery of between 6 V and 9 V will do. Wire it with a simple changeover switch, so you can switch between solar and battery power. Ask a friend for help with this if you are not sure how to do it.

You may need to adjust the number of turns on L1 to get the best results, but it should be possible to receive at least five stations at good volume with your sun-powered radio!

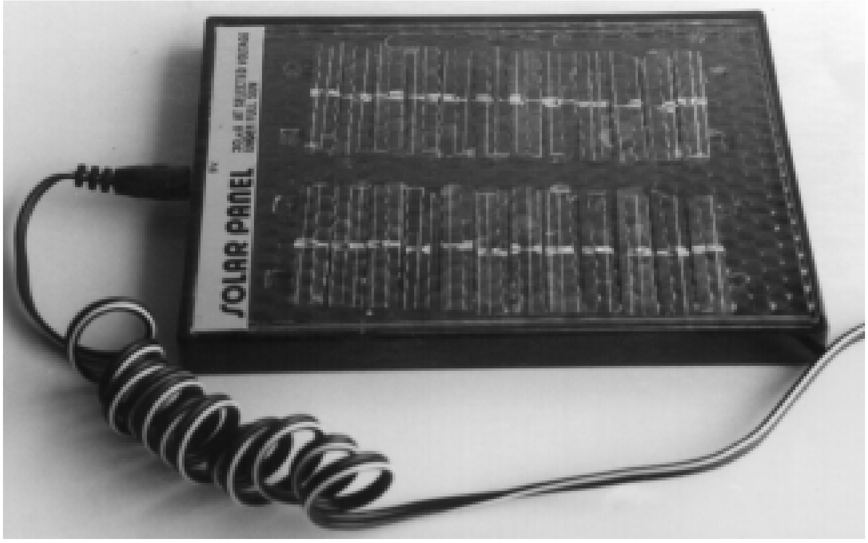

Parts list

		Maplin code
Resistors: 0.25 watts, 5% tolerance		
R1	3.3 kilohms (kΩ)	M3K3
R2	100 ohms (Ω)	M100R
VR1	1 kilohm (kΩ) log	FW21X
Capacitors		
C1, C5	10 nanofarads (nF) ceramic	BX00A
C2	220 microfarads (μF) electrolytic (10 V)	FB60Q
C3, C4	100 nanofarads ceramic	YR75S
C6	10 microfarads (μF) electrolytic (10 V)	FB22Y
C7	1000 microfarads (μF) electrolytic (10 V)	FB81C
VC1	140–300 picofarads (pF)	FT78K
Semiconductors		
IC1	ZN415E (or ZN416E)	UR70M
IC2	TDA7052	UK79L
D1, D2	1N4148	QL80B
Solar panel	9 V at 50 mA	RK23A
Additional items		
LS1	32–64 ohm miniature loudspeaker	YT28F
	Ferrite rod	YG20W
	24 SWG enamelled copper wire	BL28F
	Plastic box, approx. 220 × 140 × 70 mm	YN39N
	Veroboard, cut to size	JP47B
	8-pin DIL sockets, 2 required	BL17T
	Knobs, 2 required	FD67X
	Material for speaker grille	
	Connecting wire	

38 A receiver for the 7 MHz amateur band

Introduction

Listening on the 40 metre band (from 7.0 to 7.1 MHz) can be very rewarding – it is a popular haunt for HF Special Event stations, and at night there are signals to be heard from all over Europe. This receiver is designed purely for the 40 m band, and is ideal for those who have built the simpler receivers and are looking for something a little more challenging. The more experienced constructor may prefer to build this on prototype board.

The circuit and its construction

Figure 1 shows the circuit diagram. The receiver will work well with headphones or loudspeaker. Walkman-type headphones and speakers are ideal for use here.

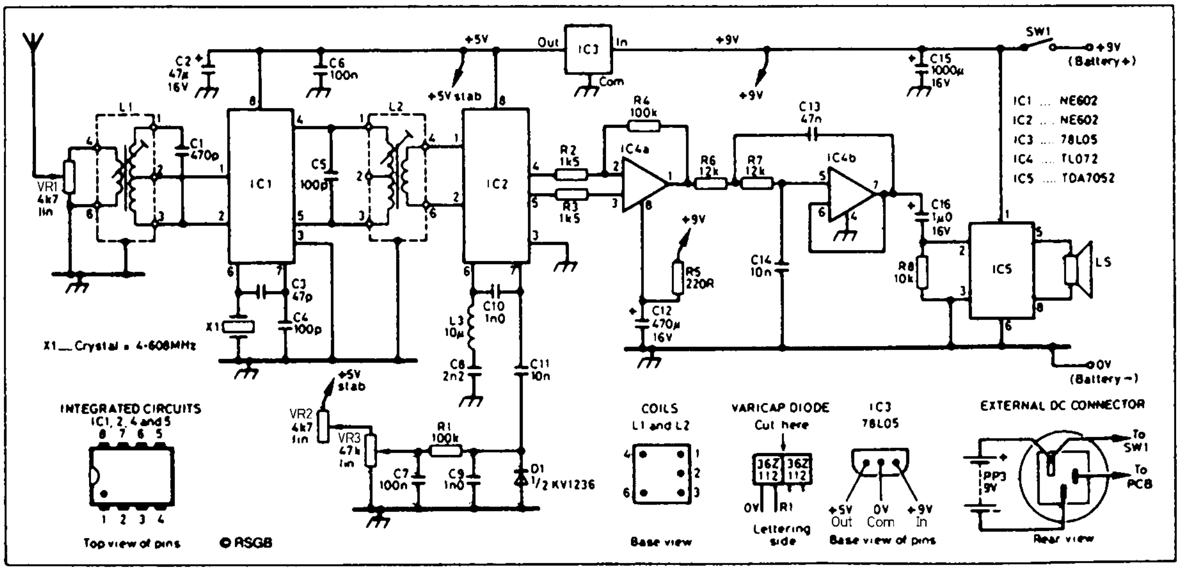
C2 47μ 16V
C6 100n
+5V
Out
IC3
In
Com
+9V
SW1
+9V (Battery+)
C15 1000μ 16V
+5V stab
+9V
L1
L2
C1 470p
IC1
C5 100p
IC2
VR1 4k7 lin
R2 1k5
R3 1k5
R4 100k
IC4a
R6 12k
R7 12k
C13 47n
IC4b
C16 1μ0 16V
R8 10k
IC5
LS
R5 220R
C14 10n
C12 470μ 16V
0V (Battery −)
C3 47p
C4 100p
X1
X1 — Crystal = 4·608MHz
C10 1n0
L3 10μ
C8 2n2
C11 10n
+5V stab
VR2 4k7 lin
VR3 47k lin
R1 100k
C7 100n
C9 1n0
D1 1/2 KV1236
IC1 ... NE602
IC2 ... NE602
IC3 ... 78L05
IC4 ... TL072
IC5 ... TDA7052
INTEGRATED CIRCUITS IC1, 2, 4 and 5
8 7 6 5
1 2 3 4
Top view of pins
© RSGB
COILS L1 and L2
Base view
VARICAP DIODE
Cut here
36Z 112
0V
R1
Lettering side
IC3 78L05
+5V Out
0V Com
+9V In
Base view of pins
EXTERNAL DC CONNECTOR
PP3 9V
To SW1
To PCB
Rear view

Figure 1 The receiver gives good performance on the 7 MHz amateur band as well as being simple to construct and align

Signals arriving at the aerial are coupled into IC1 via gain control VR1, which also functions as the on/off switch. Tuning is provided by varying the voltage on the *varactor diode* (or varicap), D1. VR3 is the main tuning control, and VR2 is the *bandspread* (fine tuning) control. The varactor diode is supplied as a dual device, which must be cut down the middle *carefully* with a sharp knife; with the lettering upwards, the ground lead (0 V) is on the left-hand side, as Figure 1 illustrates.

Solder in the IC sockets first, followed by the coils. After this come the links, resistors, capacitors and varactor diode. Ensure that IC3, the voltage regulator, is wired correctly, and check the polarity of the electrolytics. The crystal, X1, is very fragile, so take extra care with it. The wiring of the three controls is shown in **Figure 2**.

Before putting the ICs in their sockets, connect up the battery and check the following voltages with the negative voltmeter lead connected to the negative terminal of the battery:

Pin 8	IC1	5 V
Pin 8	IC2	5 V
Pin 8	IC4	9 V
Pin 1	IC5	9 V

When all these have been found to be correct, switch off and put the ICs carefully into their sockets. Use wire of different-coloured insulation to wire up the front-panel controls.

The case can be a small plastic box of size 22 cm by 15 cm by 8 cm, with three 10.5 mm holes drilled in the front and two 8 mm holes in the side for the aerial and earth connections. On one side are a 6 mm hole for the speaker socket and an 11 mm hole for the optional external power supply.

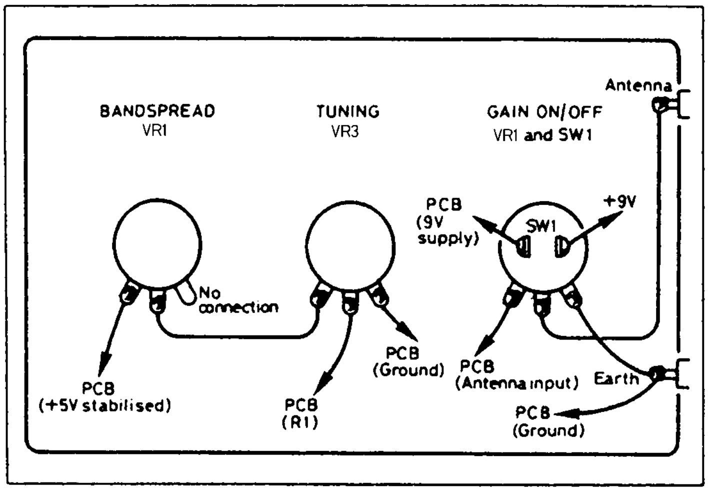

Figure 2 Rear view of the variable resistors. Check the connections carefully to make sure the wires are fitted correctly

Testing and tuning

The aerial for the receiver should be between 30 and 70 feet of wire, mounted as high as you can make it, away from trees and buildings if possible. Connect the battery and switch on. Adjust L1 and L2 for the best results. Tune slowly with VR3; you should find CW stations at the lower end of the band (anticlockwise) and SSB stations at the upper end (clockwise). You may find that it is easiest to make these adjustments *before* mounting the board in the case with double-sided sticky tape or pads. If you are planning to use an external DC supply, make sure it is a safety approved stabilised 9 V type, and **disconnect the battery before you use such a supply!**

If you suspect that the tuning doesn't quite cover the lower CW end of the band, try increasing C9 to 1200 pF. If it is the upper SSB end which is missing, decrease C9 to 820 pF.

It is always advisable to use an aerial tuning unit (ATU) between your aerial and the receiver. A suitable design of ATU is included as a project in this book.

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1, R4	100 kilohms (kΩ)
R2, R3	1.5 kilohms (kΩ)
R5	220 ohms (Ω)
R6, R7	12 kilohms (kΩ)
R8	10 kilohms (kΩ)
VR1	4.7 kilohms (kΩ) linear, with SPST switch
VR2	4.7 kilohms (kΩ) linear
VR3	47 kilohms (kΩ) linear

Capacitors: all rated 16 V or more

C1	470 picofarads (pF) polystyrene 5%
C2	47 microfarads (μF) electrolytic
C3	47 picofarads (pF) polystyrene 5%
C4, C5	100 picofarads (pF) polystyrene 5%
C6, C7	100 nanofarads (nF) ceramic
C8	2.2 nanofarads (nF) polystyrene 5%
C9, C10	1 nanofarad (nF) polystyrene 5%
C11, C14	10 nanofarads (nF) ceramic
C12	470 microfarads (μF) electrolytic
C13	47 nanofarads (nF) ceramic
C15	1000 microfarads (μF) electrolytic
C16	1 microfarad (μF) electrolytic

Inductors	
L1	Toko KANK3335R
L2	Toko KANK3333R
L3	10 µH 5%, e.g. Toko 283AS–100
Semiconductors	
IC1, IC2	NE602 or NE602A
IC3	78L05 5 V, 100 mA
IC4	TL072
IC5	TDA7052
Additional items	
D1	Toko KV1236 cut into two sections (one half used)
X1	4.608 MHz (available from Cirkit)
	3 × silver knobs, one with pointer
	Plastic case approx. 22 × 15 × 8 cm
	Speaker 8–32 Ω, or headphones
	4 × 8-pin DIL sockets for IC1, IC2, IC4, IC5
	2 × 4 mm sockets (red and black) for aerial and earth
	3.5 mm chassis-mounting jack socket for speaker
	DC power socket for external supply (if required)
	Prototype board

Kits

A complete kit is available from JAB Electronic Components.

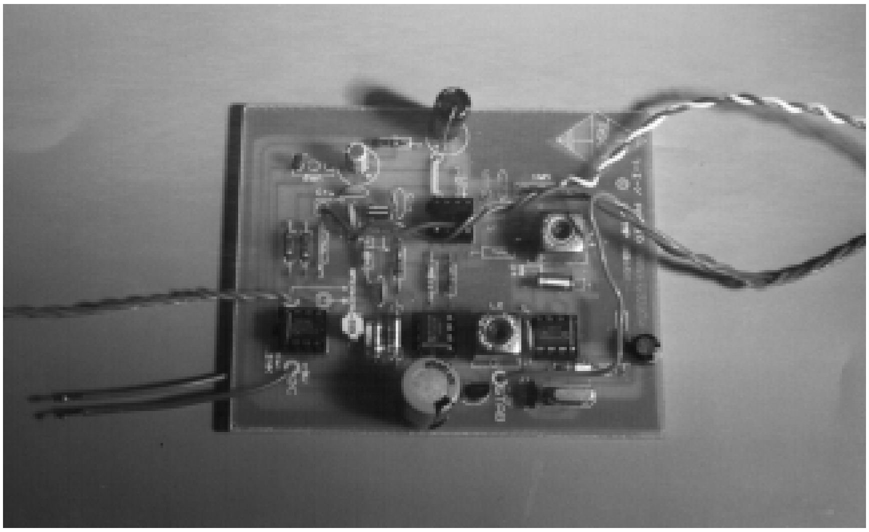

39 Diodes for protection

Introduction

Many semiconductor devices can be destroyed in an instant if their supply is reversed. With the use of batteries as power sources, it is quite common for the battery to be connected ‘the wrong way round’, and the scope for damaging equipment is significant and very real. Diodes can protect equipment in several ways, and you may do worse than to consider one of these approaches to protect your next expensive project.

Choose wisely!

All the diode circuits given here are so simple as to invite calamity. The circuits are not foolproof but, with a little care, will work first time. Remember that a semiconductor diode has a forward voltage drop of between 0.5 V and 0.7 V, depending on its type and the current flowing. This will be mentioned later.

The first thing you need to do is to insert a good multimeter *in series* with the circuit you want to protect; set the range to Amps DC, and switch on. Check that the circuit works properly. Then, decrease the current range on the meter until a good reading is indicated. Make a note of this current, as it is the normal running current of your circuit.

To choose a diode, you must consult the catalogues or data sheets and find one where the quoted *maximum forward current* exceeds the current you have measured; preferably it should be at least twice your measured current. Secondly, the diode will have a *peak inverse voltage* (PIV); this is the maximum voltage it can withstand when the cathode is made positive with respect to the anode, i.e when it is *reverse-biased* and not conducting. This voltage must be greater than your battery voltage, again by a factor of about 2. Except in the case of the bridge rectifier (see later), these criteria will enable the selection of a suitable diode to be made easily.

The series diode

The simplest and most obvious way to protect equipment is to insert a diode in the positive supply lead, as shown in **Figure 1**, with the diode passing current *only* when the supply is of the correct polarity. Because of the 0.6 V that exists across the diode, your equipment will normally operate on a

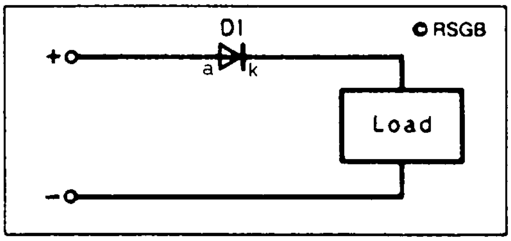

Figure 1 Series diode protection

slightly lower voltage. If you imagine reversing the supply to this circuit, you will see that the negative terminal of the battery is connected to the anode of the diode; the diode becomes reverse-biased and will not conduct any current, thus protecting the load. Your equipment will not operate when the battery connections are reversed.

The parallel diode

The circuit of **Figure 2** overcomes this voltage drop. It places the diode in parallel with the load (your equipment) but in a normally reverse-biased condition so that it draws no current when the battery is correctly connected. Reverse the battery connections, however, and a very large current will flow through the diode, thus blowing the fuse! For this technique to work successfully, the current drawn by the diode when the battery connections are reversed *must* be much greater than the maximum current drawn by the equipment, in order to blow the fuse. This is usually not a problem, however.

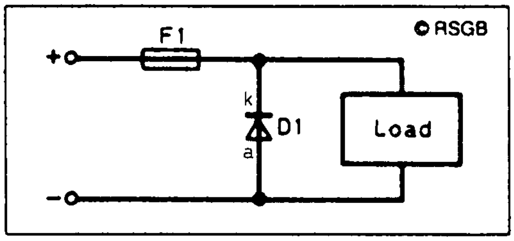

Figure 2 Parallel diode and fuse

The diode bridge

For sheer elegance, the circuit of **Figure 3** takes the biscuit! It uses four diodes connected in the form known as a *bridge rectifier*. Such rectifiers exist, and do not have to be made up from four discrete diodes.

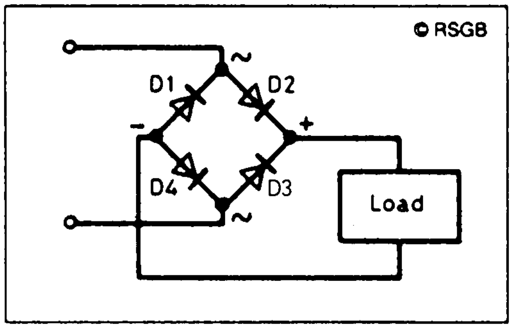

Figure 3 The diode bridge

Follow the current round the circuit from the supply, assuming initially that the top wire is positive and the bottom wire negative. It flows from the positive supply

(a) through D2
(b) through the load (top to bottom)
(c) through D4

and back to the negative of the supply.

Now assume that the bottom supply lead is positive and the top lead negative. The current flows from the positive supply

(a) through D3
(b) through the load (top to bottom)
(c) through D1

and back to the negative of the supply.

So, whichever way round the battery is connected, the current will *always* flow the same way through your equipment!

The circuit does have a drawback, however. Whichever way round the battery is connected, there are always **two** diodes conducting the current at any time. In the first case it is D2 and D4; in the second case, D1 and D3. This means that there is a total voltage drop of about 1.2 V. If your equipment can tolerate that reduction in voltage, then you will not have a problem.

Decoupling

Whenever the supply rail to a piece of equipment, or even to an individual stage of a circuit, is broken for the insertion of a device that will drop voltage, strange things can happen. This is because the supply for any circuit is assumed to have a low resistance to DC and a low *impedance* to AC. (Impedance is the AC equivalent of DC resistance.) These two are not the same, and the insertion of a diode or diodes is certain to make a big difference to them both. To overcome any irregularities in the operation of the circuit you are protecting in any of the ways described previously, the protected supply needs to be *decoupled* with a capacitor.

If the circuit handles audio frequencies only, placing an electrolytic capacitor directly across the load in all three circuits should solve the problem. The parallel circuit of Figure 2 is less at risk than the other circuits. The size of the capacitor will be determined by the current taken by the circuit, and may need to be chosen within the range 100 µF to 10 000 µF, with a working voltage greater than the supply voltage.

If the circuit is mainly handling RF currents, placing a capacitor of 0.01µF across the load should prevent any problems. A second capacitor, also across the load, of between 10 µF and 100 µF may be needed. Again, the parallel circuit of Figure 2 is less at risk than the other two.

Don't be afraid to experiment, but confine your experimenting (at first) to small equipment and low currents, until you get a 'feel' for the technique.

40 An RF signal probe

Introduction

A radio-frequency (RF) diode probe is a simple device which, when used with a conventional multimeter, enables the measurement of RF voltages in a circuit. When constructed, this will be one of the most useful pieces of test equipment for the experimenter who revels in the construction of transmitters and receivers.

The circuit

Figure 1 shows the simple circuit diagram. It is *almost* the same as a common diode rectifier circuit, but with a simple change to make it more sensitive. The circuit is known as a *voltage-doubler*, and is often found in high-voltage supplies, with beefier capacitors and diodes, of course! Because we are dealing with high frequencies and smaller voltages, the diodes and capacitors can be physically very small. The diode circuit of D1 and D2 *rectifies* or *detects* the RF from the probe, and any remaining AC is removed (short-circuited to ground) by C2. This produces, at the output, a constant voltage proportional to the peak-to-peak RF voltage present at the input; the output voltage is fed to an ordinary multimeter (on a voltage range).

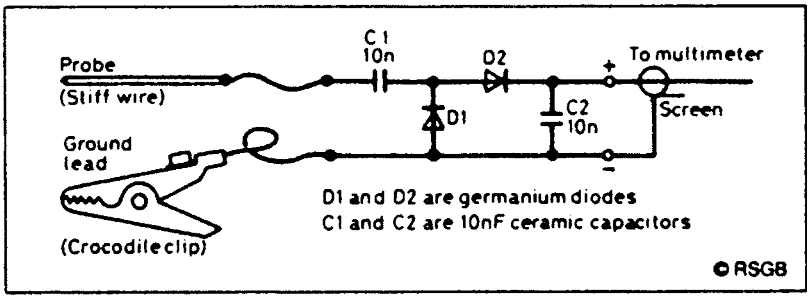
C1
10n
Probe
(Stiff wire)
D2
To multimeter
+
C2
10n
D1
Screen
Ground
lead
−
D1 and D2 are germanium diodes
C1 and C2 are 10nF ceramic capacitors
(Crocodile clip)
© RSGB

Figure 1 The diode probe can be made from four electronic components

Construction

Although the circuit layout is not critical, a description of the prototype is given here for your information. **Figure 2** has the details. The components are soldered to square pieces of copper-clad printed-circuit board (PCB) glued to a larger piece of PCB. The larger piece serves as a ground connection. Use a stiff copper wire as the probe, and an insulated flexible wire with a crocodile clip to connect the probe to the ground of the circuit under test.

Cut a piece of plain PCB, 30 mm by 45 mm and another of 15 mm by 45 mm. Cut the smaller piece into three measuring about 15 mm square. Stick the three small pieces to the larger piece, ensuring that there are small gaps between each, as Figure 2 shows. Solder the components in place. The connection to the multimeter should be thin coaxial cable or screened microphone cable. If you use unscreened cable, there may be RF pickup here which can lead to false readings.

Simple to use

Using the probe is simple. Connect it to the multimeter and set the meter to around 10 V DC – you may need to reduce this, depending upon the magnitude of the RF voltage you are trying to measure. Hold the probe by

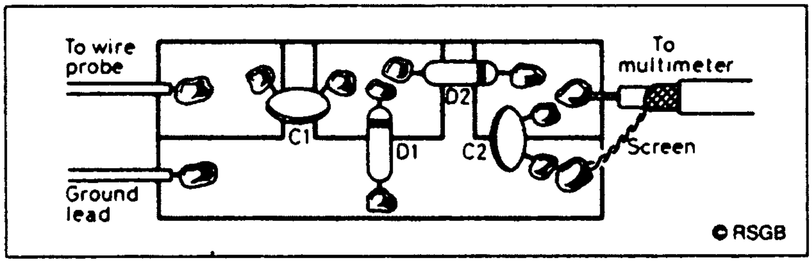
To wire
probe
To
multimeter
D2
C1
D1
C2
Screen
Ground
lead
© RSGB

Figure 2 The components are mounted on a copper-clad PCB

its base, being careful *not* to touch any of the components. Connect the croc clip to a ground point on the circuit and touch the probe on the circuit point to be tested. If an RF signal is present, there should be a reading on the multimeter. That's all there is to it!

Parts list

Capacitors	
C1, C2	10 nanofarads (nF), or 0.01 microfarad (μF) ceramic
Semiconductors	
D1, D2	Any *germanium* signal diode, e.g. OA91 or AA119

Additional items
- Stiff copper wire
- Insulated flexible wire
- Copper-clad PCB
- Crocodile clip
- Coaxial or screened cable

Source

All components are available from Maplin.

41 An RF changeover circuit

Introduction

The simplest entry route to operating an HF station is to buy or build a receiver, and to add a simple CW (Morse) transmitter. In theory, this sounds so simple, yet in practice there is one significant hurdle to be overcome. How can they both share the same aerial? A manually operated changeover switch is the obvious solution, but it takes time to perform the switching operation, by which time someone else has squeezed in before you and contacted the DX station! This small circuit accomplishes the changeover automatically, as soon as the transmitter is keyed.

How it's done

The circuit is shown in **Figure 1**, and uses the signal from the transmitter to operate a relay. A relay is a switch which is operated electrically in the following way. A coil of wire with an iron core is used as an electromagnet. When current flows through the coil, it produces a magnetic field which, in turn, is used to pull a set of switch contacts. These contacts will be used to switch the aerial from transmitter to receiver and vice versa. The relay used here has a double-pole changeover (or DPDT, or DPCO) switch. If you are confused by the different types of switch, look at the basic descriptions elsewhere in this book. The *pole* of a switch is the part that doesn't move; in Figure 1, one pole is connected to L1 and the other to the aerial output. In a circuit diagram, the switch contacts are always shown in their *normal* state, i.e. when *no* current flows in the relay coil. The changeover switch is normally in the receive position, so we say the receive switch is *normally closed*, and the transmit switch is *normally open*. We choose to have the circuit in the receive position normally, because everyone spends a lot more time receiving than transmitting. The relay is energised only when you are transmitting.

There are three RF sockets – one for the transmitter, one for the aerial, and one for the receiver. When the transmitter is not in use, there is a direct connection (via the Rx normally closed contacts of the relay) between the aerial and the receiver. The transmitter (which is not keyed at this time) is connected to a 50 ohm dummy load, R1.

When the transmitter is activated by pressing the Morse key, an RF signal appears at the transmitter socket. The wire carrying the signal to the relay

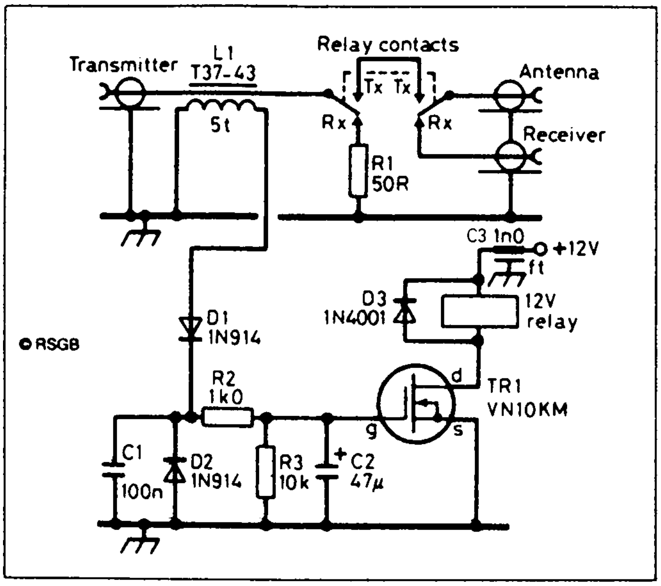

Figure 1 RF activated relay

Tx contacts passes through a toroid (a ferrite coil former, looking like a black Polo mint), which has a five-turn coil wrapped around it. This five-turn coil picks up the signal, which is then detected by D1 and D2, and converted to a steady voltage. C1 removes any remaining RF and C2 provides *hang*. Without it, the circuit would detect the gaps between every dot and dash being sent, and switch the aerial over very rapidly and very frequently! This is *not* what we want. We need the relay to remain in the transmit condition at normal Morse keying speeds, and then return to receive when the sending is complete. The combination of C2 and R3 achieves this.

The voltage appearing across C2 is fed to a VMOS field-effect transistor (FET), TR1, which acts as an electronic switch. When a voltage appears across C2, the FET switches on and passes a current through the relay coil, changing the contacts from receive to transmit. The diode, D3, across the relay winding protects the FET from being damaged by the large reverse-voltage spike which occurs across the coil when the FET switches off.

Construction

This is very simple – a circuit board is not required. The prototype was built into an old 50 g tobacco tin. All ground leads are soldered directly to the tin, thus supporting the components automatically. Phono sockets were used for the aerial, receiver and transmitter; you could use whatever connectors matched the rest of your station. The 12 V supply is fed through the tin using a 1000 pF feed-through capacitor, C3.

The 'dummy load', R1, must be able to withstand the RF output power of the transmitter. For novice use, 4 watts is adequate. Two suggestions for making up R2 from standard resistors are shown in **Figure 2**. Six 330 ohm, 1 watt resistors in parallel have a combined resistance of 330/6 = 55 Ω at 6 W. Two 100 ohm, 2 watt resistors in parallel have a combined resistance of 100/2 = 50 Ω at 4 W.

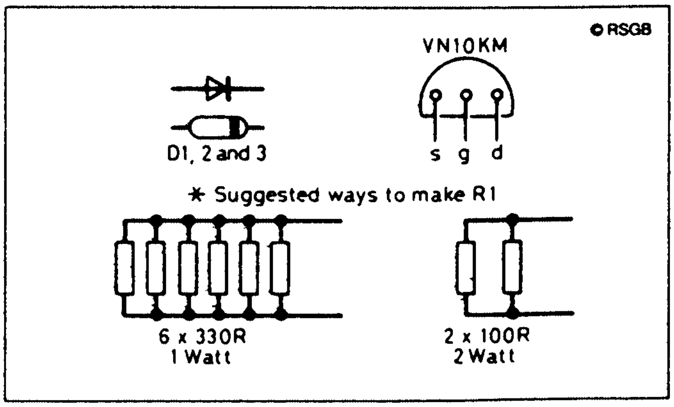

Figure 2 Component details

Make sure all the contacts to the relay are correct; also ensure that TR1, C2, D1, D2 and D3 are correctly wired.

Winding the toroid, L1, is very simple. It is made from PVC-insulated hook-up wire. Each time the wire passes through the centre of the toroid counts as one turn. The wire from the transmit socket to the relay simply passes through the centre! Use thin 50 Ω coax for the leads to the three sockets.

What happens to the receiver?

On transmit, the receiver is not connected to the aerial, but is only the separation of the switch contacts (about 0.5 mm!) away from the transmit lead. The receiver *will* pick up the transmitted signal, and there is a possibility that this signal will be enough to damage the receiver's sensitive input circuits. This can be prevented with the simple addition shown in **Figure 3**. Two diodes are connected back to back across the receiver socket. These act as a *limiter*, reducing the amount of signal that can enter the receiver. Solder these directly between the receiver socket and the tin.

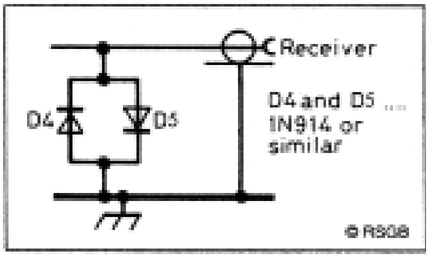

Figure 3 Receiver input protection

Warning

The wiring **must** be thoroughly checked. If the relay switch connections are wrong, there is a **great probability** that the full RF power will be applied to your receiver! You can check these connections as follows. With the 12 V supply connected, check that the relay energises (clicking sound heard) when the FET drain (d) is shorted to earth. Check the aerial and receiver connections with an ohmmeter or continuity tester. Disconnect the 12 V supply, and check that there is continuity between the centre pins of the aerial and receiver sockets. Reconnect the 12 V supply, short the FET drain to earth again (this will not damage the FET) and check that there is continuity between the centre pins of the aerial and transmitter sockets. If these tests show correct operation, disconnect the drain shorting wire (*most* important!) and your aerial auto-changeover is ready for use! Avoid using the device without the lid fitted.

Parts list

Resistors: 0.25 watt, 5% tolerance, unless otherwise stated

R1	50 ohms (Ω) – see text and Figure 2
R2	1 kilohm (kΩ)
R3	10 kilohms (kΩ)
Capacitors	
C1	100 nanofarads (nF), 0.1 microfarad (μF)
C2	47 microfarads (μF) 15 V electrolytic
C3	1000 picofarads (pF) feedthrough
Semiconductors	
D1, D2, D4, D5	1N914
D3	1N4001
TR1	VN10KM
Additional items	
L1	T37–43 toroid
	Relay 12 V DPDT (DPCO) relay
	Three sockets to suit (phono, SO239, etc.)
	Thin coax cable and hook-up wire

42 A low-light indicator

Introduction

This is a simple one-evening project that can be built for the pure fun of it, or to use as the basis of a more complex project to switch your shack lights on when it gets dark! In its prototype form, it simply flashes an LED when the ambient light level drops to a preset point.

Operation

The heart of the circuit shown in **Figure 1** is a *photo-conductive cell*, also called a *light-dependent resistor* (LDR), a device whose resistance changes according to the amount of light falling on it. In bright light, the resistance is low (about 1 kΩ), whereas in the dark, its resistance is very high (up to 10 MΩ). The cell is made from a semiconducting material known as cadmium sulphide (CdS), and is enclosed in a small plastic container. The semiconductor is laid on a flat insulating surface in the form of a small flat ribbon. The ribbon construction gives a good area of surface for a given

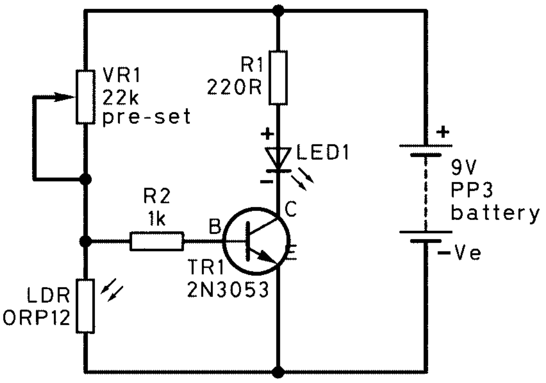
VR1
22k
pre-set
R1
220R
+
LED1
−
+
9V
PP3
battery
−Ve
R2
1k
B
C
E
TR1
2N3053
LDR
ORP12

Figure 1 Circuit diagram

©RSGB DY107

length of ribbon, and the length of the ribbon is maximised by laying it out in a zig-zag pattern, as can be seen diagrammatically in **Figure 2**. In the dark, CdS is an insulator; when light falls on it, electrons are released inside the CdS, making it conduct. The more light there is, the more electrons there are, and the resistance falls.

In this circuit, the LDR is connected across the 9 V supply in series with a variable resistor, VR1. In this arrangement, the voltage that exists across the LDR will be determined by the light level. As the light intensity increases, the resistance of the LDR falls, dropping a smaller voltage across it. The

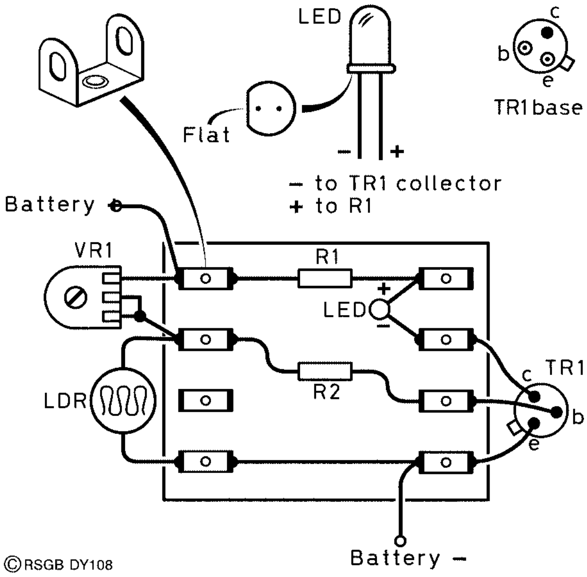
LED
Flat
− +
− to TR1 collector
+ to R1
c
b
e
TR1 base
Battery +
VR1
R1
+
LED
−
LDR
R2
c TR1
b
e
Battery −

Figure 2 Component layout on tagstrip

©RSGB DY108

reverse happens when the light intensity falls – the voltage across the LDR increases. This voltage is used to drive an npn transistor, TR1, connected as an electronic switch. As the voltage on the base (B) rises, it will reach a point where TR1 will suddenly start to pass current, just as an ordinary switch does when pressed. The current flowing through TR1 also flows through a flashing light-emitting diode (LED), D1 and its series resistor, R1. D1 can be a steadily glowing type, if preferred. If you want to make the circuit switch the LED on at a different ambient light level, adjust VR1.

Construction

The prototype was made on an eight-tag tagboard (Figure 2). The resistors R1 and R2 can be laid on the tagboard for soldering, the rest of the components lying above or to the side of the board. Check the circuit after you have soldered everything on, then connect the battery. The LED should flash if you put your hand over the LDR, and VR1 can be adjusted to vary the point at which the LED lights. The project can be housed in a small plastic box with holes provided for the LDR and LED and an on/off switch if you want one.

Parts list

Resistors: all 0.25 ohm, 5% tolerance

R1	220 ohms (Ω)
R2	1 kilohm (kΩ)
VR1	22 kilohms (kΩ) linear pre-set

Semiconductors

LDR	ORP12
TR1	2N3053 npn
LED1	Flashing LED

Additional items
Plastic box
Tagboard with 8 tags
PP3 battery connector

43 A J-pole aerial for 50 MHz

Introduction

For FM communication (i.e. voice and data) on the VHF bands, a vertical aerial is used to give all-round (non-directional) coverage. This is a half-wave aerial which can be fed at the end, thus removing the principal problem with the conventional vertical centre-fed half-wave dipole, which is that the feeder should leave the dipole at right angles. This is no problem when the dipole is horizontal, but can be difficult for the vertical dipole.

Basic facts

A feeder must be connected to an aerial at a point where the *impedance* (AC ‘resistance’) of the aerial closely matches that of the feeder. The difference between the two impedances gives rise to the voltage standing-wave ratio (VSWR), which is unity only when the two impedances are the same. With 50 Ω feeders, the feed point of a half-wave aerial is at the centre, where the aerial impedance is around the same value. At the end of a half-wave aerial, the impedance is high, so it is *not* a suitable point to connect a 50 Ω feeder.

Connection at this point can be effected using an RF transformer. RF transformers act in the same way as ordinary transformers, except that they are much smaller, and usually comprise wires of particular lengths adjacent to each other. **Figure 1** is a good starting point. It shows the aerial in its diagrammatic form. Notice that the aerial is in the form of an elongated letter ‘J’; this shape gives rise to its nickname – the *J-pole*. The quarter-wave RF transformer is the lower ‘U’ section below the half-wave element. At the bottom of the U section, the impedance is zero (this may become clearer later) and at the top of the U section it is high, thus matching the aerial impedance. The coaxial feeder cable is connected part-way up the U section, where the impedance is around 50 Ω.

The practicalities

Figure 2 shows the aerial as constructed. As it is about 4.5 metres high, it may be too high for the average house loft, but is ideal for mounting outside, supported by a non-metallic pole or hung from a tree branch. The upper half-wave section is made from 1.5 mm insulated copper wire, as used in domestic mains wiring. The quarter-wave transformer below the half-wave section is made from 300 Ω balanced line (‘ribbon cable’). The

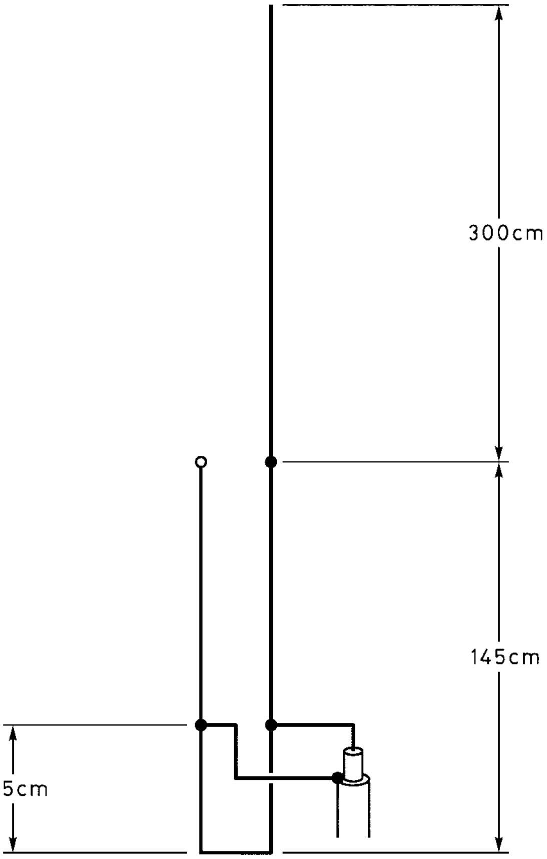

Figure 1 Overall dimensions of the 6 metre J-antenna

wires at the bottom of the transformer section are stripped of their insulation, twisted and soldered. At the upper end, only one wire of the balanced line is soldered to the bottom of the half-wave section. The other wire of the pair is not connected and is left insulated.

At the feed point of the transformer, the insulation needs to be carefully stripped from the balanced line. You will need a standing-wave meter (VSWR meter) in the coaxial line between your transmitter and the aerial, and you will need to adjust the position of the feed point. 45 cm from the bottom was the best point on the prototype, but this position is dependent upon the immediate surroundings of the aerial, and *must* be done when the aerial is in its final operating position. **Warning: *Never* make adjustments to the feed point when the transmitter is on. Make a VSWR measurement, switch off, move the feed point, switch on again, make another measurement, and so on.** You will need to aim for the lowest VSWR you can – certainly better than 2:1. Having found the best position, wrap all the exposed wires with self-amalgamating tape, to seal them against the ingress of moisture.

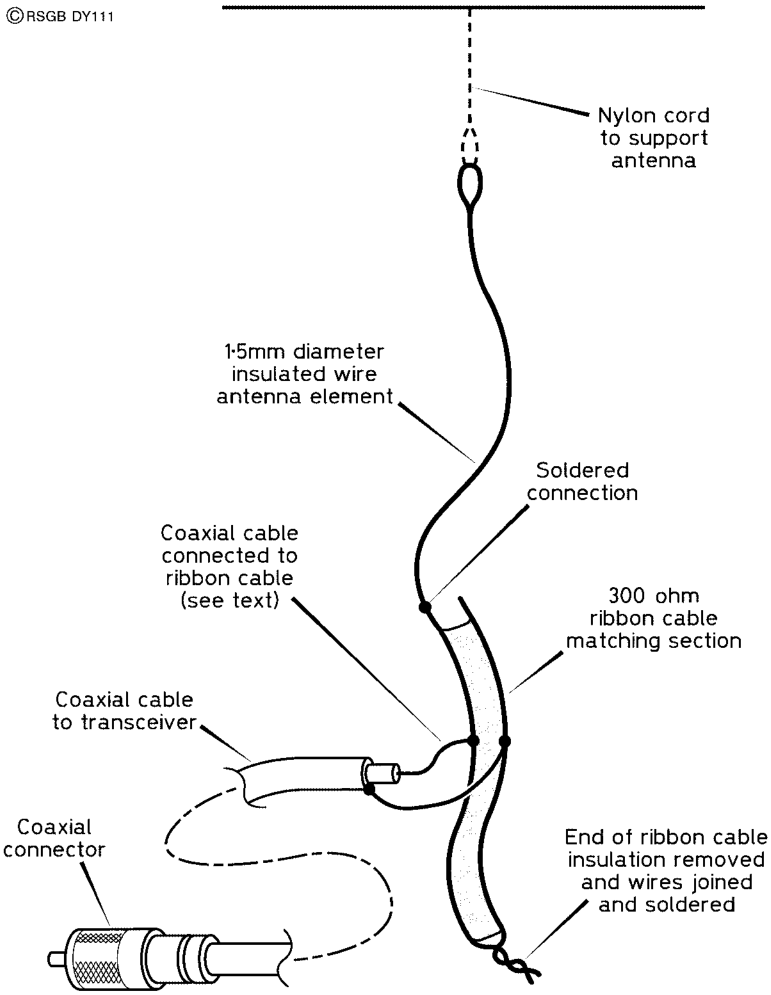
©RSGB DY111
Nylon cord to support antenna
1·5mm diameter insulated wire antenna element
Soldered connection
Coaxial cable connected to ribbon cable (see text)
300 ohm ribbon cable matching section
Coaxial cable to transceiver
Coaxial connector
End of ribbon cable insulation removed and wires joined and soldered

Figure 2 Construction details

How it performs

Figure 3 shows a computer prediction of how the J-pole radiates. It is called a *polar diagram*, and shows the distribution of your transmitted power when viewed 'from the end of your garden'. Most of your signal is sent at a fairly small angle to the horizontal; very little signal goes upwards, which is a good thing, of course. This also shows why the J-pole (or any other vertical aerial) should not be called 'omnidirectional', which means it radiates in all directions. It is omnidirectional *only* in the horizontal plane.

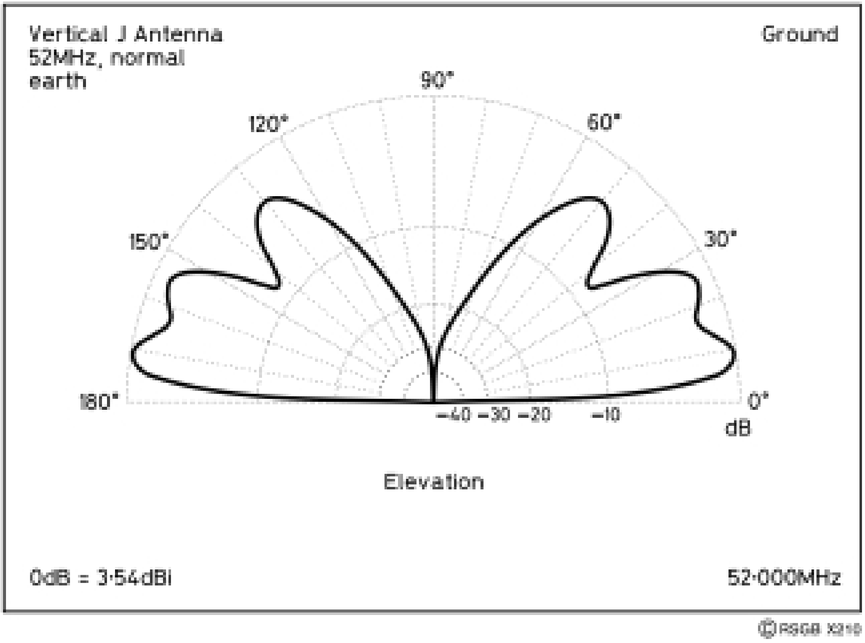

Figure 3 This polar diagram computed for the J-pole

Safety

Where you mount your aerial is a matter of personal preference and the restrictions of height and space, but the following safety rules *must* be applied.

1. Never fix an aerial where it may come into contact with power lines or telephone lines.
2. When climbing a ladder to put up an aerial outside, make sure the ladder is safe and that it is secured.
3. Don't do this alone. Preferably have someone with you. If this is not possible, make sure someone knows where you are.

Parts list

3.00 metres	1.5 mm insulated copper wire
1.50 metres	300 ohm balanced line
As required	50 ohm coaxial cable
As required	Self-amalgamating tape

44 Measuring light intensity – the photometer

Introduction

Before the days of automatic 'point-and-shoot' cameras, a photographer would use a light meter or *photometer* to measure the light level, then manually convert this reading into shutter speed and lens aperture settings to ensure a correctly exposed negative. Modern cameras have quite sophisticated photometers, which control the shutter speed and iris settings automatically.

A short explanation

A simple photometer circuit is shown in **Figure 1**, and is based on a device called a *light-dependent resistor*, or LDR. As its name tells us, its resistance depends upon the amount of light falling on it. In bright light, the resistance is relatively low (about 1 kΩ), whereas in the dark, its resistance is very high (up to 10 MΩ). The cell is made from a semiconducting substance known as cadmium sulphide (CdS), and is enclosed in a small plastic container. The semiconductor is laid on a flat insulating surface in the form of a small flat ribbon. The ribbon construction gives a good area of surface for a given length of ribbon, and the length of the ribbon is maximised by laying it out in a zig-zag pattern. In the dark, CdS is an insulator; when light falls on it, electrons are released inside the CdS, making it conduct. The more light there is, the more electrons there are, and the resistance falls.

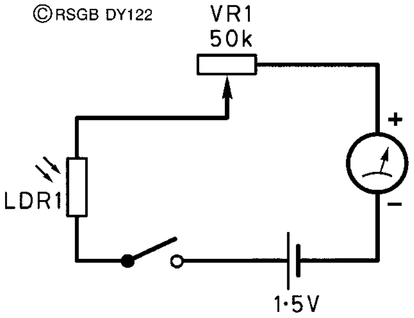

Figure 1 Circuit diagram of photometer

The photometer circuit

The circuit is simply a series connection of four things – the battery, the LDR, a variable resistor and a meter. A switch is also provided. The combination of the LDR resistance and that of VR1 determines the current flowing, which is indicated on the meter. Altering the resistance of VR1 sets the *sensitivity* of the photometer – you may want full-scale deflection of the meter needle for a bright light, or for a dim light.

It is simple to make, and a plug-in type matrix board is ideal to test the circuit, so that you can decide if you want to make a permanent version. Connect all the components in series; the only change you may want to make is the connection to VR1. If you find that the sensitivity control seems to work 'backwards', simply unsolder the wire from the end tag of VR1 and solder it to the opposite end tag. Problem solved!

In use

As soon as you connect up the battery, you will probably have a meter reading because of the daylight falling on the LDR. Shading it with your hand should reduce the reading. If the meter needle is hard over against the end-stop, turn VR1 until it indicates about half-scale. The LDR is very sensitive, and will read zero only in a dark room. If you put on a torch to see what the meter reading is, the LDR will detect the torch light, and will give a reading!

Here is a simple project where you can use the photometer in an experiment which has an analogy in radio. Draw a circle on a large (A3 or bigger) sheet of paper and divide it up into 30-degree sectors, as shown in **Figure 2**. Draw a series of smaller circles which divide the maximum radius into five. Look at the figure if you're not sure about this. Bring the LDR away from the circuit by using two long, flexible wires. The experiment must be performed in a darkened room (preferably in total darkness). Prepare a table with two columns, the left-hand one headed 'Angle (degrees)' and the right-hand one 'Meter reading'. Fill in the left-hand column 0, 30, 60 . . . and so on up to 360°.

Place the torch in the position shown, with its lens at the centre of the circle and pointing along the 0° line; switch it on. Place the LDR facing the torch and adjust VR1 until you have full-scale deflection of the meter needle. Suppose the meter indicates 10 units at this point. Enter this into your table in the 0° row. Keeping the torch the same distance from the circle centre, and pointing at it, move the LDR round all 30° positions and record the meter readings. Switch off the torch and take the sheet of paper and your tabulated results into daylight!

Lay the large sheet of paper on a table with your results beside it; then, at each 30° interval, plot the point along the radius corresponding to the meter

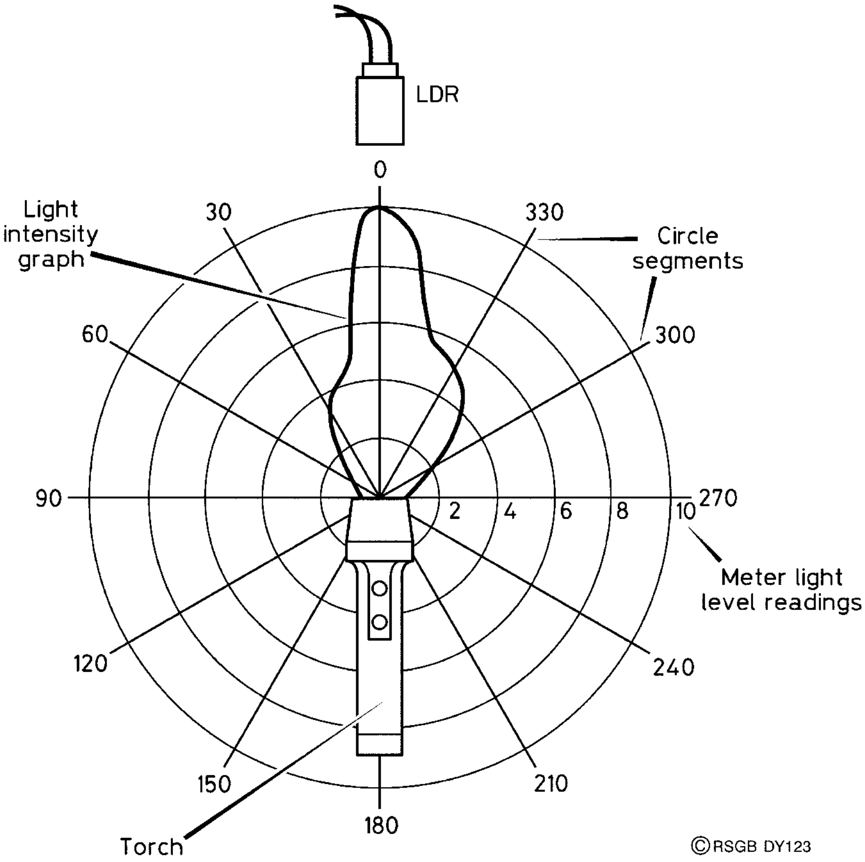

Figure 2 Torch light intensity pattern plotted at RSGB HQ

reading. Then, join up all the points and you have what is called a *polar diagram* of light intensity. The use of the word 'polar' implies that the readings have been taken in a circle and plotted that way.

Light waves and radio waves are both examples of *electromagnetic radiation*. The torch is designed to 'beam' its radiation in a particular direction, just like an aerial does. Hence the use of the word 'beam' for a directional aerial. If a similar polar diagram is drawn for a Yagi-type aerial, it will show the same general characteristics as does Figure 2, namely a main direction (or 'lobe') where most of the energy is concentrated, with evidence of *sidelobes*, indicated by 'lumps' on the otherwise smooth main lobe.

Also in this book you will find a project for the construction of a UHF field-strength meter, which you could use to carry out a measurement of the polar diagram of a UHF aerial.

Parts list

Resistor

VR1	50 kilohms (kΩ) linear

Semiconductor

LDR1	ORP12

Additional items

Meter, 50 or 100 μA
Battery clip for single AA cell
AA cell

45 A 70 cm Quad loop aerial

Introduction

This is a description of how to make an aerial for the 70 cm band which has gain compared with the 'rubber duck' or the dipole aerial. It can easily be dismantled and reassembled, making it ideal for contest use.

The principles

Most aerials which comprise several similar *elements* arranged along a *boom* are variations of the design originated by Yagi and Uda, and which takes the name (for historical reasons) of the former, and is know as the Yagi array. Let us suppose we have a Yagi aerial beaming left to right in front of us. The elements get progressively shorter from the left (look at almost any TV aerial to see this). All the elements on a Yagi aerial are classified as follows:

- The *reflector* – the leftmost element as we look at the array. It is the longest. Next to it is:
- The *driven element* – this is the element connected to the feeder, which in turn runs down to the transceiver. It is slightly shorter than the reflector.
- All the elements beyond the driven element are called *parasitic elements*, or *directors*. They are shorter than the driven element and usually get progressively shorter as we progress to the right along the boom. The directors are mainly responsible for the *directivity* (or *beamwidth*) of the array.

The progression from a simple dipole (a driven element in isolation) to a Yagi array is simple, but is nevertheless important. To make an aerial of two elements, a reflector (*not* a director) is added to the driven element. For three or more elements, directors are added to the two-element design. Adding more and more directors soon becomes impractical, the reduction in beamwidth (such as it is) does not warrant the extra expense, weight and wind resistance that is incurred.

Instead of using linear (straight) elements, as in the generic Yagi, this design uses loops. Designs using squares of wire instead of loops are known as *Quad* aerials, and HF designs require large X-shape frames to support the large squares of wire. At 70 cm, however, the use of wire loops is easier, and they are self-supporting.

Construction

This is quite simple. Any type of material (metal, plastic, wood) can be used for the *boom* (the support for the elements) and for the mast. The elements are made from 14 SWG enamelled copper wire. 16 SWG hard-drawn aerial wire, which is not enamelled, has also been used with success. Thinner wire might result in a rather ‘floppy’ aerial! The separate parts of the aerial are held together with *jubilee clips* (hose clips).

The driven element is secured to the boom with a jubilee clip and a three-connector plastic connector block as shown in detail in **Figure 1**, and in the photograph. Cut the wire for the driven element 70 mm longer than the 700 mm indicated in Figure 1. Then, using sandpaper, remove the enamel from one end to a distance of 20 mm, and from the other to a distance of 50 mm. After forming the loop of the driven element, bend both stripped ends through 90°, and insert them into the first to holes of the plastic connector block (Figure 1). Do not tighten the screws yet. Push both ends into the block as far as they will go, then bend the 50 mm end back on itself and pull the ends back through the connector block so that the end you have just bent goes into the third hole in the block. Now tighten the screws in the block and in the jubilee clip.

Each director and the reflector should be made 40 mm longer than the circumferences shown in Figure 1. Strip the enamel, as before, from the last 20 mm at each end. Form the wire into the loop, slip the ends under the jubilee clip (Figure 1) and tighten it. You may find that it helps to solder the stripped ends together before securing the jubilee clip.

The boom is fixed to the mast using jubilee clips and wire, as shown in Figure 1. Solder the feeder cable to the driven element, with the braid soldered to the end of the driven element which is connected to the boom (this applies to metal booms; with plastic or wooden designs, the feeder connections are not critical).

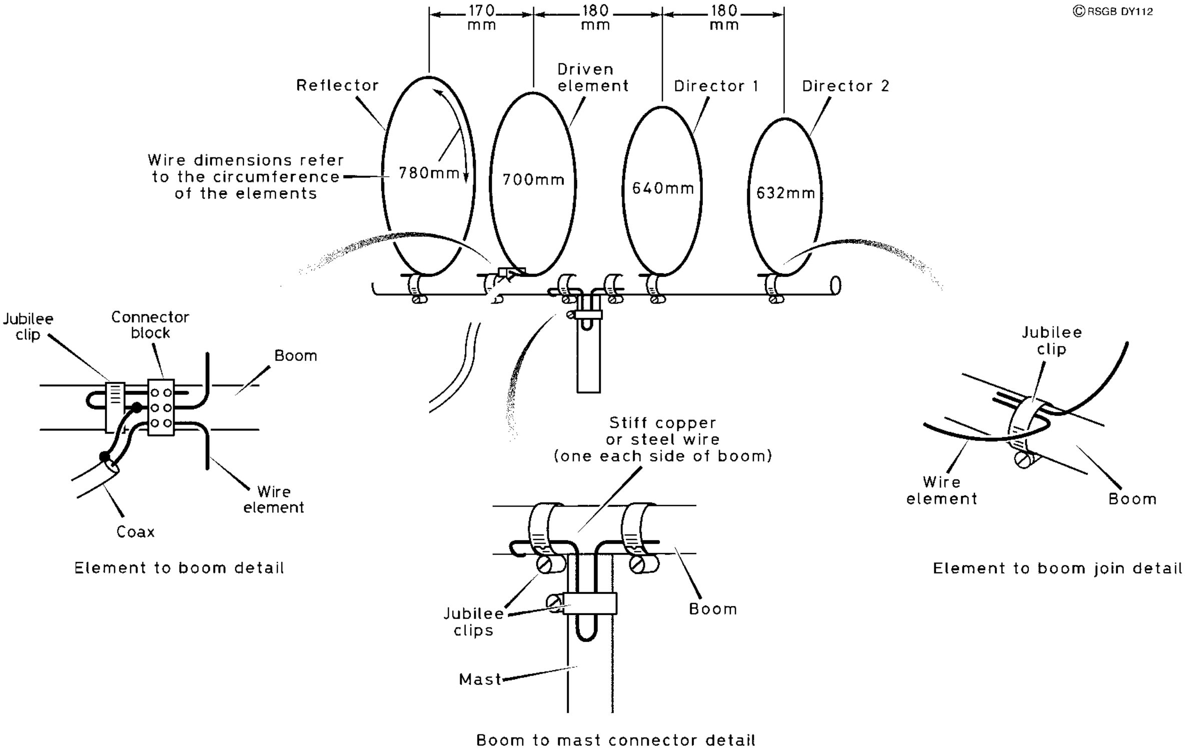
©RSGB DY112
170 mm
180 mm
180 mm
Reflector
Driven element
Director 1
Director 2
Wire dimensions refer to the circumference of the elements
780mm
700mm
640mm
632mm
Jubilee clip
Connector block
Boom
Wire element
Coax
Element to boom detail
Stiff copper or steel wire (one each side of boom)
Jubilee clips
Boom
Mast
Boom to mast connector detail
Jubilee clip
Wire element
Boom
Element to boom join detail

Figure 1 70 cm 4 element quad construction, with detail of how hose clamps are used in the construction

You will need a plug on the shack end of your feeder to suit the transceiver, aerial tuning unit (ATU) or the standing-wave meter (SWM) you are using. Always connect these in the order: transceiver – SWM – ATU – aerial.

Testing

Always test aerials outside and away from buildings (if possible!). This avoids getting misleading results.

Use a rubber duck, or whatever aerial you usually use, and tune around to find a repeater or beacon signal which is consistent. Note the reading on the S-meter. Then, connect your new Quad loop. Rotate it to give the strongest S-meter reading (don't forget it is directional). Verify that the meter reading varies as you rotate it. How does the S-meter reading compare with the original reading? Much depends on the siting of your original aerial; if you are comparing your Quad loop at ground level with a vertical on the chimney, you wouldn't expect your new aerial, even with its gain, to outperform a vertical which is well elevated!

Now a test on transmit is called for. The use of an SWR meter is essential here. Find a clear frequency and check that it *really is clear* before announcing your presence and carrying out the test. A reading of 1:1 is excellent, but any value less than about 1.8:1 would be acceptable. You can measure the directivity of your aerial using the field strength meter, also described in this series.

Materials

4 metres of 14 SWG enamelled copper wire
Material for boom and mast
15 amp connector block
7 Jubilee clips

The enamelled copper wire is available from AA&A Ltd, Sycamore House, Northwood, Wem, Shropshire SY4 5NN. Everything else is available from most hardware stores.

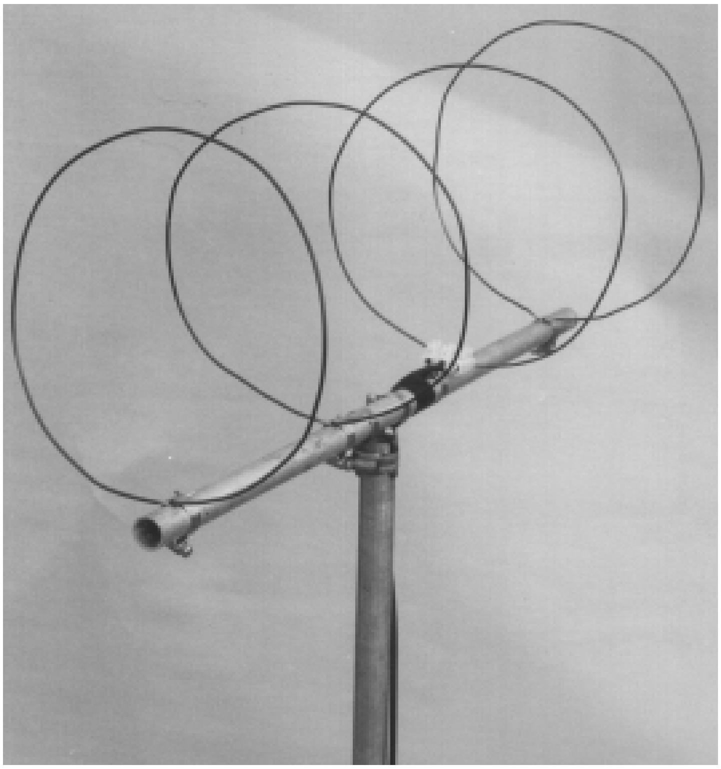

46 A UHF field strength meter

Introduction

It is always interesting, and often useful, to know just *where* the radiation is going from your aerial. How much of your radiated energy is going in the general direction of the station you are in contact with, and how much is being effectively wasted? Some of these questions can be answered with the use of a *field strength meter*. A field strength meter is simply a receiver, stripped down to its bare essentials, such that it responds only to the magnitude of the carrier. The use of a field strength meter assumes that the aerial under test is radiating a continuous carrier. Don't forget to find a clear frequency and identify your transmissions at least every quarter-hour, in order to comply with the terms of your licence.

Description

Two types of field strength meter are shown in **Figure 1**. You will recognise both circuits (Figure 1b particularly) as being types of 'crystal set' with a meter replacing the headphones. Figure 1a is a broad-band HF design (there is no tuning provision) and Figure 1b is tuned in the same way as the crystal set; with a loop of wire as an aerial, it will perform well in the VHF/UHF range.

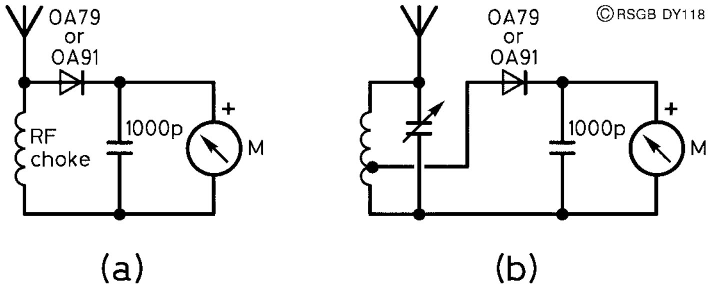

Figure 1 Construction of the UHF field strength meter

A tuned field strength meter also doubles as an *absorption wavemeter* if it is well calibrated. Such devices are useful for detecting transmitter harmonics also. This design is intended for use in the UHF band, so it will have to be sensitive around 432 MHz.

Construction

The field strength meter comprises a loop of wire, 600 mm long, which acts both as aerial and as the tuning inductance, a diode, a capacitor, a connector block, a meter and a length of twin wire. All the components, with the exception of the meter, are fixed to a pole with a jubilee clip, as shown in **Figure 2**. The meter should have a sensitivity of between 50 µA and 100 µA, or a multimeter can be used. The multimeter is more flexible, as you can select different current ranges, giving you a range of sensitivities.

Using the field strength meter

Connect your meter to the ends of the twin wire from the pole. Place a hand-held transceiver about 2 metres away, and press Transmit. If there is no meter reading switch off the transmitter and check the wiring. If the needle attempts to go negative, simply reverse the wires to the meter. If the reading is too high, either move the transmitter further away, or increase the current range on the multimeter. Try changing the orientation of the transmitter aerial, and note how the signal varies.

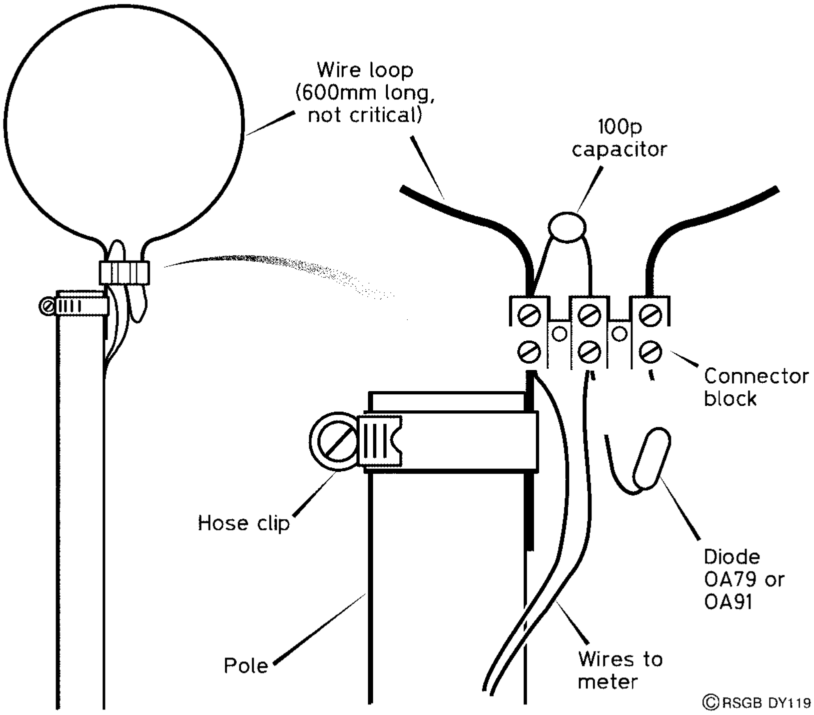
Wire loop
(600mm long,
not critical)
100p
capacitor
Connector
block
Hose clip
Diode
OA79 or
OA91
Pole
Wires to
meter
©RSGB DY119

Figure 2 Assembly of the field strength meter

To measure the *polar diagram* (a graph of the signal strength against the angle between the aerial boom and the field strength meter, plotted on polar axes) of a beam aerial, mount the loop as far away from the aerial as you can, preferably at the same height, and rotate the aerial, say 15 degrees at a time, and record the signal strength, until the aerial has been turned through 360 degrees. If you need help plotting the graph, enlist the help of a friend who has done it before, or one who knows about polar graphs! If you have already built the Photometer project, you will have measured and plotted the polar diagram of the light intensity from a torch. Now you will see the great similarity!

Parts list

Capacitor	100 picofarads (pF)
Diode	Germanium, OA79 or OA91
Connector block	10 A, 3-way
Wire	600 mm of 16 SWG enamelled copper
	Length of twin cable for meter connection
Clip	Jubilee (hose) clip

47 Christmas tree LEDs

Introduction

A novelty ideal for the festive season, this circuit causes one LED at a time to light up around a small cardboard Christmas tree.

Warning

This circuit uses members of the integrated circuit family known as CMOS (complementary metal-oxide semiconductor). These use very little current and can be completed destroyed if they come into contact with the magnitudes of static electricity that most of us carry about when we walk on carpets and wear rubber shoes. You will never know if this wanton destruction has happened – all you will discover is that your circuit doesn't work and that you have tested *everything*. To avoid this problem do the following things:

1. **Before you open the little packet in which each IC is supplied, touch something which you *know* to be earthed – the metalwork of any equipment which is mains-earthed, for example. Then open the packet.**
2. **Let the IC fall gently on the bench – don't pick it out with your fingers. Touch your earthed metalwork again. Pick up the IC and insert it gently into its holder.**
3. **Repeat the process with the second IC.**

The circuit is safe from destruction while it is connected to the battery. However, when the battery is removed, the same care should be exercised with its handling, because there are no supply decoupling capacitors across each IC.

Description

Figure 1 shows the layout, and how the wires are connected from the circuit board to the LEDs around the tree. The circuit, shown in **Figure 2**, is quite complicated, so you need to be confident in your logical approach to circuit-building before you attempt this one! It uses two common integrated circuits (ICs). IC1 is simply an oscillator which provides timing pulses for IC2, which 'counts' up to a maximum of 10. The outputs of the counter are indicated by light-emitting diodes (LEDs); red, orange, yellow and green are common colours which you can use.

The circuit is built on a single piece of Veroboard measuring 29 holes by 12 strips. **Be aware that there is no strip labelled 'I', so don't make mistakes in your counting!**

First of all, cut the tracks in the positions shown in Figure 1 using a 3 mm ($\frac{1}{8}$ inch) twist drill rotated between thumb and forefinger. Then, solder in the IC sockets, the notches facing row M. Wire up and solder in the links and then the Veropins for the connections to the LEDs and battery. Having done this, the resistors and the capacitor should be fitted. When wiring the LEDs, each cathode (the lead adjacent to the 'flat' on the LED encapsulation) is connected to R4, and the anodes go to separate pins on IC2.

Testing

Hold up your circuit board to the light and check carefully for solder bridges between adjacent tracks. Then check again that the wiring is correct. Place the ICs in their holders, with the notched ends lining up with those on the holders. Connect the battery, and the LEDs should illuminate in sequence.

If you have no success, you are now wishing you had checked the circuit more carefully! Learn something from your mistake and it will not have

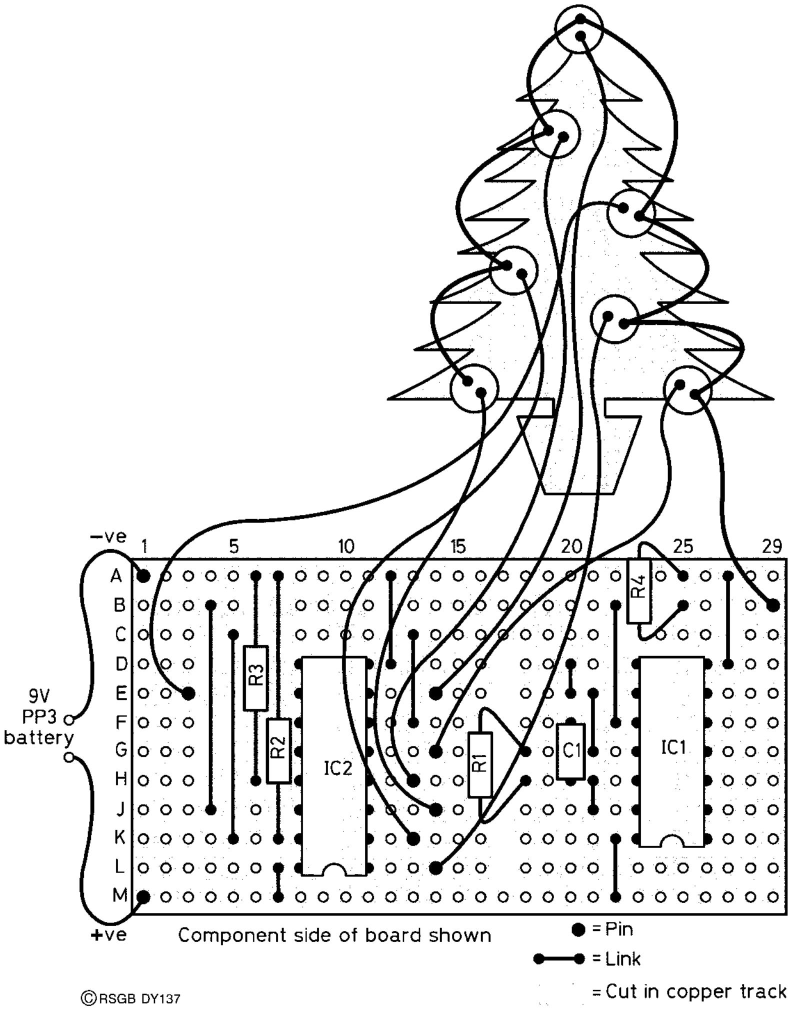
−ve
1
5
10
15
20
25
29
A
B
C
D
E
F
G
H
J
K
L
M
9V
PP3
battery
+ve
R3
R2
IC2
R1
C1
IC1
R4
Component side of board shown
● = Pin
●—● = Link
= Cut in copper track
©RSGB DY137

Figure 1 Christmas tree, component layout

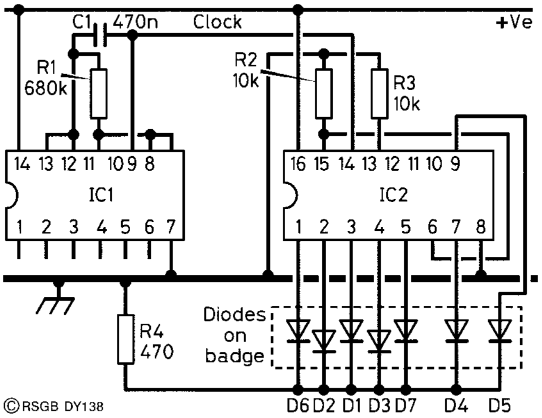
C1
470n
Clock
+Ve
R1
680k
R2
10k
R3
10k
14 13 12 11 10 9 8
IC1
1 2 3 4 5 6 7
16 15 14 13 12 11 10 9
IC2
1 2 3 4 5 6 7 8
R4
470
Diodes
on
badge
D6 D2 D1 D3 D7 D4 D5
©RSGB DY138

Figure 2 Christmas tree, circuit diagram

been in vain! The first thing to check is that the LEDs are the correct way round. If that is OK, then check the positions where the tracks are broken (intentionally!). After that, are the wire links in the right places – had you forgotten there is no row 'I'? Check *all* the wiring, then check *again* for solder bridges, and switch on again. One of these tests should have revealed a fault. If it *still* refuses to work, perhaps one (or both) of your ICs were damaged by static electricity, despite your precautions – or did you choose to ignore them?

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1	680 kilohms (kΩ)
R2, R3	10 kilohms (kΩ)
R4	470 ohms (Ω)

Capacitor

C1	0.47 microfarad (μF) min. metallised polyester film

Semiconductors

IC1	4011
IC2	4017

Additional Items

D1–D7	5 mm LEDs in choice of colours
	14-pin DIL socket for IC1
	16-pin DIL socket for IC2
	PP3 battery clip and battery
	Veroboard, Veropins
	Insulated wire for links

48 An audio signal injector

Introduction

An audio signal injector is a device used to test audio frequency circuits. It is simply an oscillator running at a frequency in the audio range, so that when its output is fed to the input of an amplifier, it will produce a sound in the loudspeaker if the amplifier is working. The oscillation is so rich in harmonics that the signal can also be heard (sounding rather different) when injected into an RF circuit.

The design

The circuit is shown in **Figure 1**, and is a basic *astable multivibrator*, a free-running oscillator producing a roughly rectangular-wave output. The two transistors, TR1 and TR2, operating as switches, switch on and off alternately at a frequency around 500 Hz. The prototype was constructed on plain matrix board (no copper strips), as illustrated in **Figure 2**.

Both transistors are type BC108, which are only a few pence each new, and can be found at almost any rally. You can add an on/off switch, or simply disconnect the battery when you are finished using it. To make the unit in one piece, the battery can be taped to the board, as the diagram shows.

The probe itself is made from a short piece of stiff insulated wire, soldered to a tag on the board; an earth lead is also soldered to the board, and terminated in a crocodile clip to attach to the ground lead of the equipment under test.

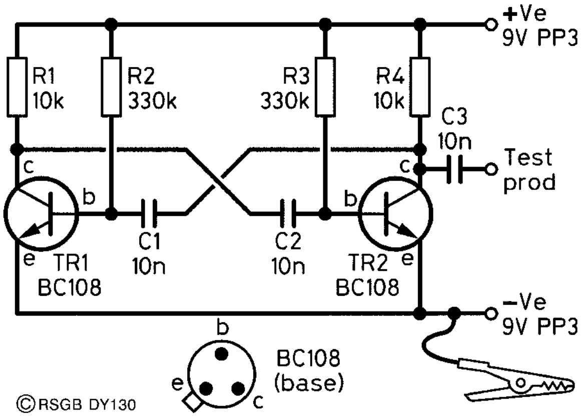

Figure 1 Circuit diagram of the signal injector

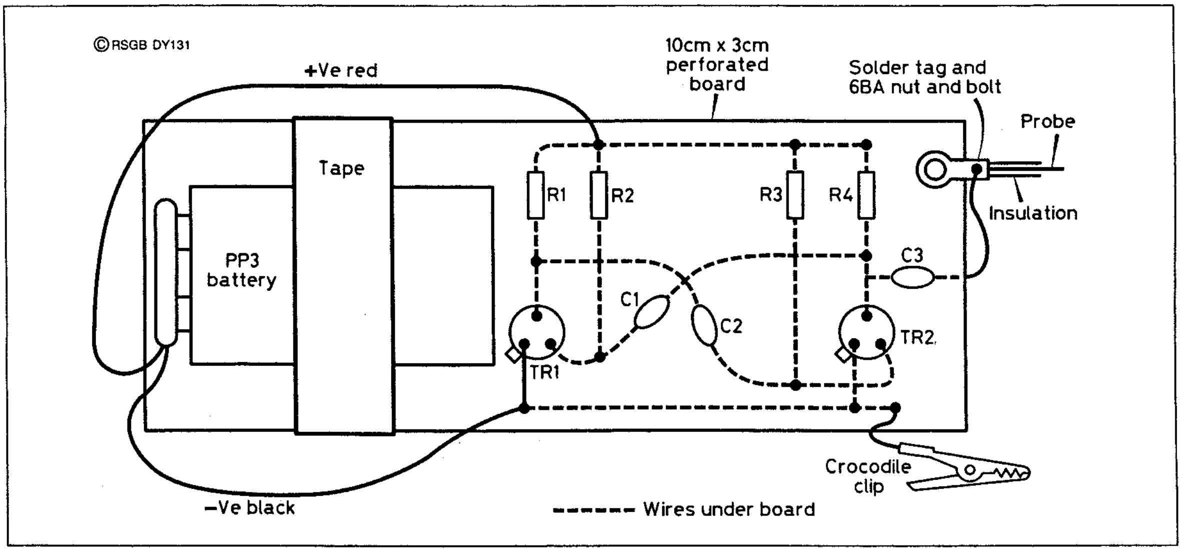
©RSGB DY131
+Ve red
10cm x 3cm perforated board
Solder tag and 6BA nut and bolt
Probe
Tape
PP3 battery
R1
R2
R3
R4
Insulation
C3
C1
C2
TR1
TR2
Crocodile clip
−Ve black
------ Wires under board

Figure 2 Component layout of the signal injector

Using it

After checking the wiring of the circuit, connect the battery. Find a transistor radio that is known to work. Carefully remove the case, switch on and apply the probe to the centre tag of the volume control. A *very* loud buzz should be heard from the speaker, indicating that the audio circuits of the radio are working. Using the injector to fault-find equipment you have made yourself is rather more instructive and rewarding, because you *know* where to inject the signal, and you should *know* what to expect when you do.

Warning

Do *not* work on any equipment connected to the AC mains. Work only on battery-powered circuits, for your own safety.

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1, R4	10 kilohms (kΩ)
R2, R3	330 kilohms (kΩ)

Capacitors

C1, C2, C3	10 nanofarads (nF) or 0.01 microfarad (μF) ceramic

Semiconductors

TR1, TR2	BC108

Additional items
- PP3 battery and connector
- Matrix board 10 cm by 3 cm
- 6BA solder tags
- Thick insulated wire for probe
- Crocodile clip

49 Standing waves

Introduction

Everyone has, or *should* have, a standing-wave ratio (SWR) meter as part of his/her array of test gear. Most people know how to use it, but what does it really do?

Before attempting to answer that question, we need to look at some aerial fundamentals. An aerial is a *transducer*, the word meaning 'to lead across'. We use it whenever one form of energy is converted into another form. A bulb is a transducer; it converts electrical energy in the filament to radiated heat and light energy. **Figure 1** shows the situation.

An aerial, or antenna, is also a transducer; it converts radio-frequency (RF) energy in the feeder (or *transmission line*, to give it its proper name) into radiated electromagnetic energy, in the manner shown in **Figure 2**. The aerial has resistance, just like the bulb filament, and if the filament had zero resistance, there would be no radiated energy. In the same way, if the aerial had no resistance, there would be no radiation from it.

Transmission lines should convey RF energy from transmitter to aerial with the minimum power loss. A common form of transmission line is the *coaxial*

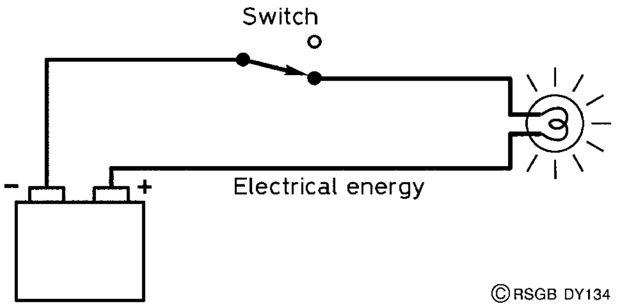

Figure 1 The light bulb converts electrical energy to electromagnetic light energy

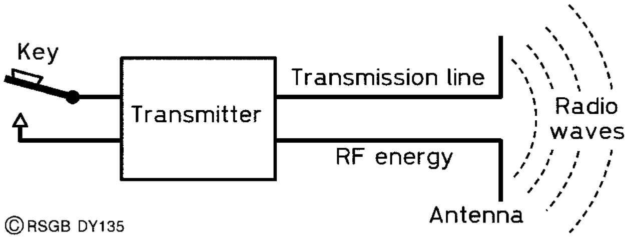

Figure 2 The antenna converts RF current to electromagnetic radio waves

cable. As you might expect, any feeder cable has a DC resistance caused by the resistance of the copper wire from which it is made. It also has an AC resistance, caused by the capacitance between the centre conductor and the braid, and by the inductance of the cable itself. This means that the feeder has an *impedance*, which is constant for the particular type of cable. This is what we call the *characteristic impedance* and, for the cables used in most amateur radio applications, it is 50 Ω.

If this impedance can be made the same as that of the aerial (it is already the same as the impedance of the transmitter output), then the transfer of energy will occur with minimum loss. If the aerial and cable impedances are *not* the same, then there is a *mismatch*, which causes some of the RF energy to be reflected back towards the transmitter. We now have a situation where RF energy is flowing along the cable from the transmitter to the aerial (the *forward* wave) and, at the same time, flowing from the aerial to the transmitter (the *reflected* wave). The two waves interact along the cable and form a stationary pattern of voltage and current. The pattern is known as a *standing wave*, and can be visualised from the waveforms in **Figure 3**. The ratio of the maximum voltage to the minimum voltage on a given wave defines the *voltage standing-wave ratio* (VSWR), or just *standing-wave ratio* (SWR) for short.

The SWR meter is easy to use. It is positioned in the feeder between the transmitter and the aerial tuning unit (ATU) if there is one, or between the transmitter at the aerial otherwise. Most meters have a single meter and a four-position switch. The transmitter is first keyed and, with the switch in

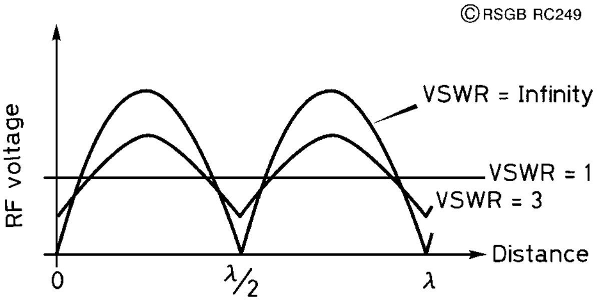

Figure 3 Standing waves on a transmission line

the ‘calibrate’ position, the sensitivity control is adjusted to give full-scale deflection on the meter. The switch is then changed to ‘forward’ to read the forward power, to ‘reflected’ to read the reflected power, and to ‘SWR’ to read the value of the standing-wave ratio. When there is no reflection (see Figure 3), the meter should read 1:1 or, simply, 1.

Most SWR meters remain in the feeder line while the transmitter is operating, so the condition of the aerial and feeder can be constantly monitored. Problems with the aerial (such as water entering the feeder at its junction with the driven element) are immediately shown up. Without the use of the SWR meter, the situation would slowly deteriorate over several months and you would be left wondering why so few stations were answering your calls!

50 A standing-wave indicator for HF

Introduction

The standing-wave ratio (SWR) meter shows how well the aerial system, including the feeder, is *matched* to the output of the transmitter. This design does not *measure* SWR, but it gives an indication of when the SWR is minimum for a given system of aerial and feeder. The design is usable on the HF bands from 1.8 to 28 MHz, and can be used at 50 MHz with reduced sensitivity.

How it works

There are two types of wave in any feeder: the *forward* wave, which travels from the transmitter to the aerial; the *reflected* wave, which travels back to the transmitter from the aerial. The presence of a reflected wave is evidence that some of your transmitted power is not being radiated, but is being returned to the transmitter to be lost as excess heat. When aerial and feeder are perfectly matched, there is *no* reflected wave, and all the power from the transmitter is radiated.

Referring to the circuit of **Figure 1**, a tiny fraction of the signal is removed by the transformer, T1, and by the capacitors, VC1 and C1. It is then detected by the germanium diodes, D1 and D2, and any residual RF removed by the capacitors, C2 and C3. The currents through the diode and meter (depending on the position of switch, S1) represent the forward and reflected signals. VR1 acts as a sensitivity control for the meter.

It pays to shop around for a suitable meter. Surplus types from tape recorders and hi-fi equipment are usually ideal for this purpose. A new one would cost several pounds. The more sensitive the meter, the more sensitive your indicator will be. Meter sensitivity is given by the current required to give full-scale deflection (FSD) of the pointer. One with an FSD of between 50 and 200 micro-amps (μA) is suitable for this circuit. The higher the FSD, the less sensitive the circuit.

Construction

The meter circuit and the sampling transformer (see **Figure 2**) are built and mounted on Veroboard of the copper-strip variety. It simplifies construction

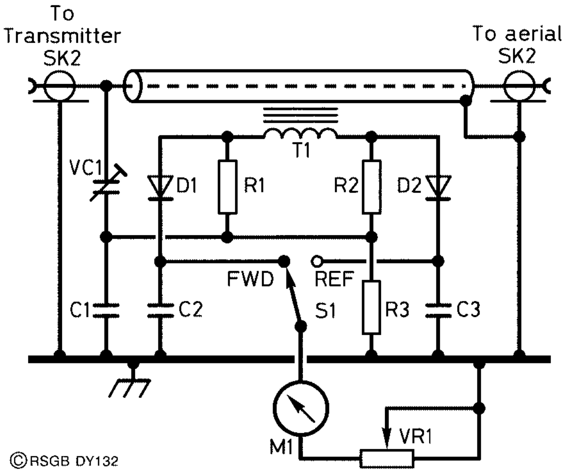

Figure 1 Circuit diagram of the SWR meter

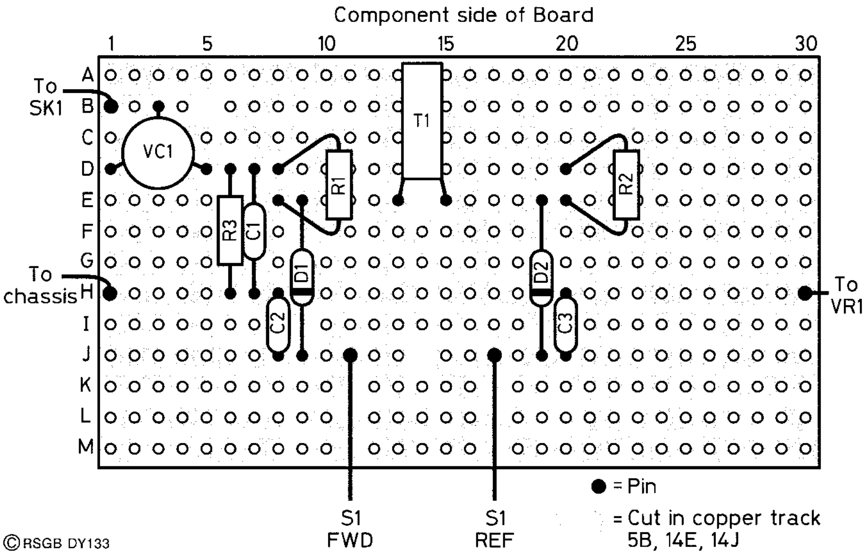

Figure 2 Component layout of the SWR meter

but reduces the operational range of the meter to below 30 MHz because of the capacitive coupling between strips. The board has 13 strips by 30 holes, although you can reduce this if you have a smaller case.

Firstly, cut the tracks at the three points shown. Then insert and solder Veropins for connections to the external components, the switch, variable resistor and the meter. Solder in the components starting with the resistors and followed by the capacitors and the diodes, ensuring that the diodes are inserted correctly.

Now you have to wind the transformer, T1, on a small toroidal ferrite core. Wind the secondary with 15 turns of 36 SWG enamelled copper wire, spaced evenly over about two-thirds of the former. The turns should not overlap, and considerable care must be taken; the wire is very thin, will kink easily and will break if you apply too much tension. The 'primary' is an 8 cm length of 50 Ω coaxial cable which passes through the toroid on its way between the input and output connectors. The braid of the cable is connected to the case at only *one* of the connectors (see Figure 1); this prevents the screen and the metal case between the two sockets forming a single, shorted turn.

The ends of the secondary winding must be carefully stripped of their enamel with sandpaper, before attaching the toroid to the board with cotton or nylon fishing line. On no account must wire be used for this!

Solder the secondary connections of T1 to the board and thread through the coaxial cable ready for soldering to the connectors.

The case used was an aluminium box (Maplin LF02C), but any suitable metal box could be used. Aluminium is preferable, as it is easily drilled with simple tools. Use standoff insulators to mount the board in the case. Once this has been done, the leads from the board to the chassis-mounted components can be soldered. So can the coaxial cable passing through the toroid. Make the lead from the input socket to VC1 as short as possible.

Setting up

You will need a 50 Ω dummy load and a transmitter to set up your indicator. Connect the transmitter to SK1 and the dummy load to SK2. Set the toggle switch, S1, to *forward* and the sensitivity control, VR1, to mid-travel. Switch on the transmitter, and set VR1 for maximum meter deflection. Switch to *reflected* and adjust VC1 until the reading is minimum (ideally zero). This completes the setting up!

Using the indicator

For setting up an aerial, connect your circuit between the transmitter and the cable leading to the aerial. With S1 in the *forward* position, key the transmitter and adjust VR1 for maximum reading on the meter. Switch to *reflected*, and then adjust your ATU to give minimum reflected power. If your adjustments are to be made to the aerial itself, to give minimum reflected power, you must make a note of the reflected reading, switch off the transmitter, make a change to the aerial, key the transmitter, and note whether the reflected power is greater or less than before. Then, make more changes to the aerial. **Never adjust your aerial with the transmitter on**. Make your adjustments on an unused frequency, and do it as quickly as possible, thus avoiding (or minimising) interference to other stations.

Parts list

Resistors: all 0.25 watt, carbon 5% tolerance
(or Maplin 0.6 watt metal film)

R1, R2	27 ohms (Ω)
R3	2.2 kilohms (kΩ)
VR1	10 kilohms (kΩ) linear

Capacitors

C1	220 picofarad (pF) disc ceramic 50 VDC
C2, C3	0.1 microfarad (μF) disc ceramic 50 VDC
VC1	20 picofarad (pF) trimmer

Semiconductors

D1, D2	OA91 germanium

Additional items

S1	Single-pole changeover (SPDT or SPCO)
SK1, SK2	Coaxial sockets to suit station standards
	Veroboard – 13 strips by 30 holes
	Veropins (7 off)
	Amidon FT 50–43 ferrite toroid
	Meter, less than 200 µA FSD
	36 SWG enamelled copper wire
	Short length of UR43 or RG58 coaxial cable
	Insulated stranded wire
	Aluminium box
	Standoff insulators for mounting the board
	Knob for the sensitivity control

51 A moisture meter

Introduction

Dry rot (*Merulius Lacrymans*) can strike havoc in buildings, causing the timbers to decay and crumble to dust – hence the term *dry rot*. Wood is attacked only if its moisture content rises above 20%.

Construction

The circuit of the moisture meter is shown in **Figure 1**. The two probes touch the wood, and the current that flows between them depends on the moisture content of the wood. If the moisture is sufficiently high, the current, after amplification, will be enough to light the LED.

The meter can be made on a piece of plain matrix board (no copper strips), as **Figure 2** shows. The board is big enough (10 cm by 2.5 cm) to accommodate the PP3 battery, taped on. No case is needed, unless you want

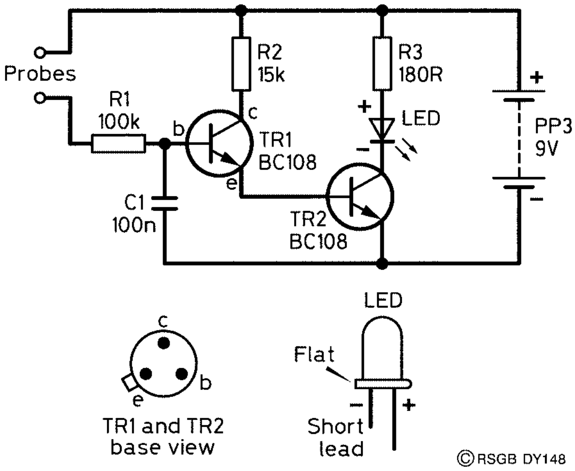
Probes
R1
100k
b
c
e
R2
15k
TR1
BC108
C1
100n
R3
180R
+
LED
–
TR2
BC108
+
PP3
9V
–
c
e
b
TR1 and TR2
base view
LED
Flat
–
+
Short
lead

©RSGB DY148

Figure 1 Circuit diagram of a moisture meter

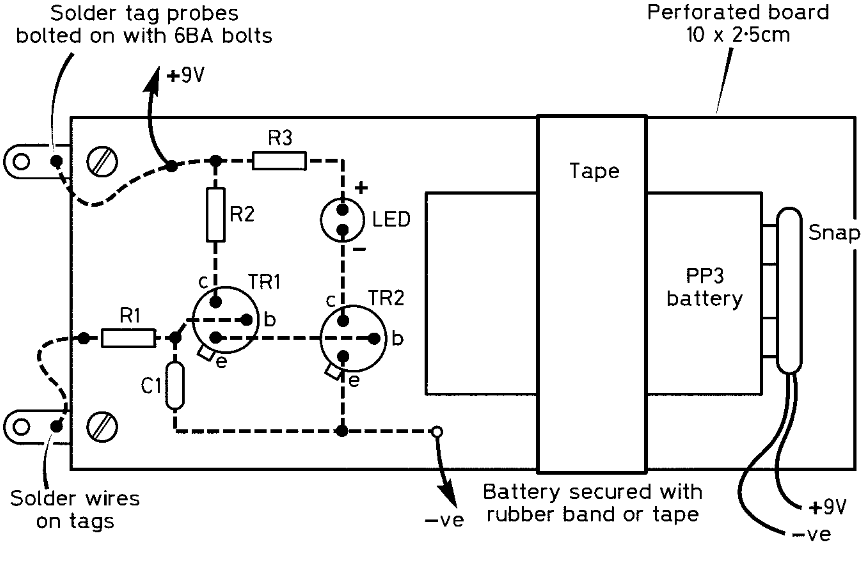
Solder tag probes
bolted on with 6BA bolts
+9V
R3
R2
+
LED
–
c
TR1
b
e
c
TR2
b
e
R1
C1
Tape
PP3
battery
Snap
Perforated board
10 x 2·5cm
Solder wires
on tags
–ve
Battery secured with
rubber band or tape
+9V
–ve

©RSGB DY149

Figure 2 Moisture meter, component layout

to leave the meter in a damp location for a protracted period! Make sure that the transistors and LED are mounted correctly. In Figure 2, the connections as shown to TR1 and TR2 are illustrated as if the transistors were transparent. An on/off switch is not really necessary, as only a very small current flows when the probes do not touch anything. Use solder tags, screwed to the board, to act as probes.

After the assembly is completed, check your circuit one final time, and then connect the battery. Nothing should happen at first. If you lick your forefinger and hold it across the probes, the LED should light.

Using it

The prototype was compared with a commercial moisture meter, and the LED lit when the moisture was around 20%. This was quite fortuitous, as the point at which the LED lights depends both on the separation of the probes and on the gain of the two transistors.

In addition to searching for dry rot, the instrument may be used to monitor the moisture in the soil of household plants. In this case, probes made of 16 SWG copper (*not* enamelled) should be soldered on to the two tags, and should penetrate the soil to a depth of several centimetres, and R1 may require adjusting so that the LED extinguishes if the soil is too dry, and lights if the soil is sufficiently moist. If you wanted to leave the meter with the probes in the soil, an on/off switch *would* be necessary.

Parts list

Resistors: all 0.25 watt, 5% tolerance	
R1	100 kilohms (kΩ)
R2	15 kilohms (kΩ)
R3	180 ohms (Ω)
Capacitor	
C1	0.1 microfarad (μF) polyester
Semiconductors	
TR1, TR2	BC108
LED	Any shape or colour will do

Additional items
  PP3 battery and connector
  Solder tags (2 off) for probes
  Matrix board 10 cm by 2.5 cm

52 Simple aerials

Introduction

The performance of *any* receiver or transceiver, no matter how expensive it is, is limited by the aerial that feeds it. Two of the most frequently asked questions are:

- Which is the best sort of aerial to use?
- Where is the best place to locate an amateur radio aerial?

To answer these questions, you must ask yourself what sort of operation you want to do. Are you interested in local, chatty contacts on the lower bands or VHF, or are you more disposed towards long-distance (DX) contacts, and on what band?

A house with a moderately sized garden is assumed in the diagrams here, to illustrate the configurations of some simple aerials. You would not need all these aerials festooned around your house, because one or two would be sufficient for your needs. The problems incurred by properties with more restricted space will be covered later.

VHF aerials

For VHF operation, the aerial should be mounted as high as possible, either on a mast or on a chimney. For all-round coverage on FM and the local repeaters, a vertical colinear is a good choice. For SSB and CW DX operation, a horizontal rotatable beam is needed. If satellite working is envisaged, you will need to contemplate mounting an elevator on top of your rotator, so that your beam can point in *any* direction, including vertically upwards! An advantage of satellite working is that the aerials do not *necessarily* have to be up in the air, provided you have a relatively uncluttered site. Your rotator and elevator can be at ground level, which is good!

If the VHF aerial is mounted on the chimney, use a double mounting bracket, particularly if you have a beam and rotator. Keep the TV, broadcast FM and amateur aerials as far apart as possible, and keeping the feeders separated is also a good plan.

The dipole aerial

One of the simplest types of aerial for single-band operation is the half-wave dipole. (The name ‘dipole’ simply means ‘two poles’ or ‘two elements’, and

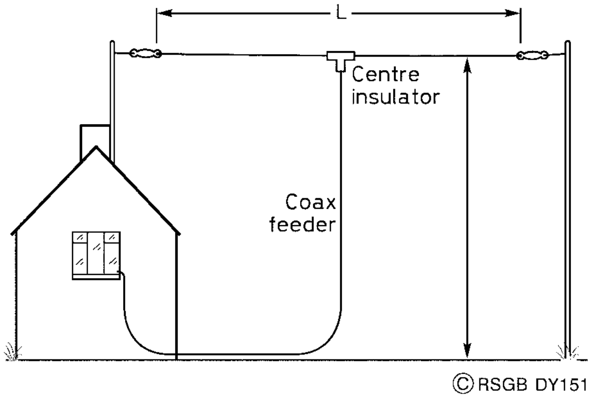

Figure 1 Layout for a dipole aerial

in this case the total length of the dipole is approximately half a wavelength at the operating frequency.) It is usually fed in the centre by coaxial cable as shown in **Figure 1**. The length of the dipole for the lowest frequency in each band is shown in **Table 1**. Normally, the length of the aerial will be 'trimmed' to be tuned to the centre frequency of the part of the band in which you will operate. This is done using the data in the right-hand column of Table 1. As an example, suppose you wanted your aerial to be resonant at 3.7 MHz. The table gives an overall dipole length of 42.86 m for 3.5 MHz. To resonate the aerial 200 kHz *higher*, then this length must be shortened by 2 × 0.595 m = 1.190 m. Your dipole would thus be 41.67 m long. Remember to allow extra wire for fixing the dipole ends to the insulators.

Table 1 Dipole lengths for lowest frequency of each band and the length to be trimmed from each to raise the resonant frequency by 100 kHz

Band (MHz)	*Dipole length (m)*	*Trim each end (mm/10 kHz)*
1.8	83.33	2190
3.5	42.86	595
7	21.43	150
10	14.85	70
14	10.71	35
18	8.33	20
21	7.14	15
24	6.03	12
28	5.36	10
50	3.00	6

On the lower-frequency bands, the lengths become rather large. In this case, you can 'bend' your dipole, as illustrated in **Figure 2**. The length of wire required to give an acceptable value of SWR (less than 2:1 on transmit) may need to be different from the calculated value, so be prepared to experiment!

Dipoles are single-band aerials, although they will often work acceptably on the third harmonic of their design frequency: a 7 MHz dipole often operates reasonably well on 21 MHz. It is possible to operate several dipoles in parallel, as **Figure 3** shows. Interaction between the elements can occur if the spacing between them is less than about 10 cm. A multi-band dipole, as shown in Figure 3, has the elements separated with plastic spacers, and drooping ends to produce maximum spacing between the elements' ends.

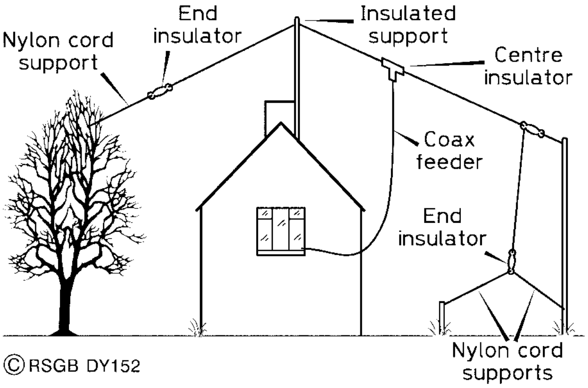
End insulator
Nylon cord support
Insulated support
Centre insulator
Coax feeder
End insulator
Nylon cord supports
©RSGB DY152

Figure 2 Possible layout for a dipole aerial in a confined space

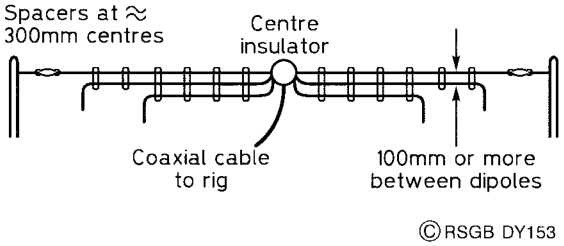
Spacers at ≈ 300mm centres
Centre insulator
Coaxial cable to rig
100mm or more between dipoles
©RSGB DY153

Figure 3 Multi-band dipole aerial

The long-wire aerial

This aerial is simple, cheap, easy to erect, and suits most houses and gardens, as **Figure 4** shows. Using an aerial tuning unit (ATU), an end-fed long wire can function on several bands when used with a set of radials or a counterpoise. **Figure 5** illustrates the setup. The length of the aerial will determine the bands which will be covered.

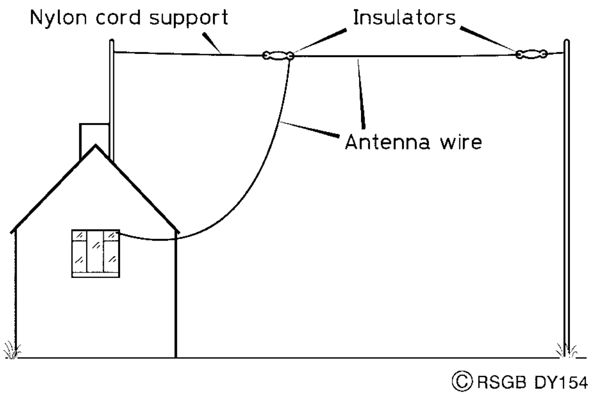
Nylon cord support
Insulators
Antenna wire
©RSGB DY154

Figure 4 Long-wire or inverted-L aerial

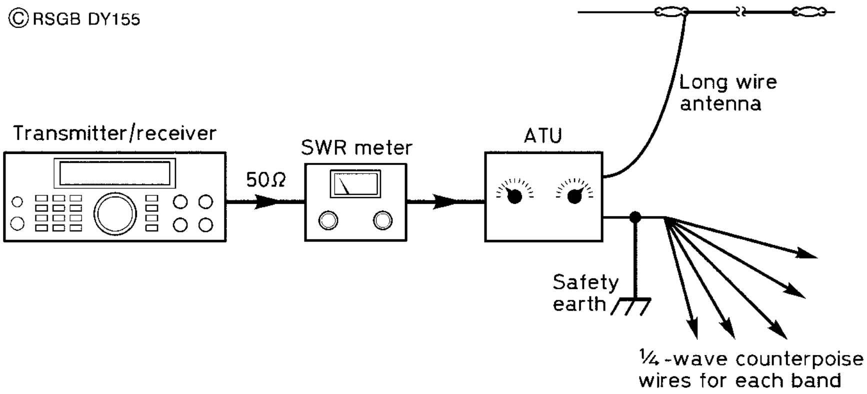
©RSGB DY155
Long wire antenna
Transmitter/receiver
SWR meter
ATU
50Ω
Safety earth
¼-wave counterpoise wires for each band

Figure 5 How to connect a radio to a long-wire aerial

A wire length of 10.5 m will work on the 40, 30, 17, 15 and 12 m bands.

A wire length of 15.5 m will work (with an ATU) on the 80, 40, 20 and 12 m bands and possibly (depending on your ATU) on the 17 and 15 m bands.

A wire length of 26.5 m will operate on all bands, but may be difficult to load on 10 m.

The wire lengths given here may need some adjustment because of the geometry of your particular house and garden. For receive-only purposes, the lengths are far less critical.

In general, you cannot get a good radio-frequency (RF) earth from a first-floor (or higher) shack. Unless a good RF earth exists within a small fraction of a wavelength of the transceiver, an *artificial ground* comprising a single $\frac{\lambda}{4}$ radial or counterpoise will be needed. You will need one counterpoise for each band you intend to use, and the wire can be concealed around the skirting-board of the shack, or under the carpet. Make sure that the free end

Table 2 Lengths of elements for vertical antennas, radials for verticals and counter-poises for end-fed long wire antennas

Band (MHz)	*Element length (m)*
1.8	39.66
3.5	20.40
7	10.20
10	7.14
14	5.1
18	3.96
21	3.4
24	2.95
28	2.55

of each counterpoise is well insulated; this point can carry a *very high voltage* when you transmit; anyone coming into contact with this can suffer very severe RF burns. Counterpoise lengths can be read from **Table 2**.

The vertical aerial

The single-band vertical aerial is sometimes used by DX operators because it has a low angle of radiation, which favours long-distance propagation. However, it must be sited clear of obstructions and must have a *good* counterpoise or radial system. Illustrations of the vertical aerial are shown, and the lengths of the vertical and radial sections are given in Table 2. The centre of the coaxial feeder is connected to the vertical section, and the braid to the counterpoise or radial system, which is made up of four or more wires buried just below the surface and joined together near the base of the aerial.

Cable entry to the house

Bringing coaxial cable into the house by an open window must be regarded as a temporary measure. Wooden window frames can be drilled, one hole for each feeder. Make the holes slope downwards from inside to out to prevent rain entering, and treat these with wood preservative. Leads from long-wire and inverted-L aerials should be kept separate from other cables.

Alternatively, a plastic pipe large enough to take all your feeders could be fitted into the brickwork (again, sloping downwards towards the outside). You may want to let a friendly builder do this for you.

53 A breadboard 80 m CW transmitter

Introduction

In the early days of radio, many circuits were built on a wooden baseboard, the parts being screwed down on the board. This was called breadboard construction, because it was a breadboard that was frequently commandeered for the process! Wives have always been generous in this respect, it appears!

This circuit was originally designed by GM3OXX, and became known as the *Oner*, because it was built on a circuit board one inch square! The circuit appeared in the G QRP Club journal *Sprat* and, since that time, many hundreds of *Oner* circuits have been built and used on the air. It is a well-proven circuit.

The transmitter has no tuned circuits in the power amplifier (PA) and thus has a rather high *harmonic* content. It **must** be used with the ***low-pass filter*** described elsewhere in this book. Without the low-pass filter, interference will be caused to other stations.

Simple aerial changeover switching is provided, which allows this circuit to be used with any of the 80 m receivers, such as the *Colt*, described in this book. It can also be used with any kit or commercial receiver for the 80 metre band.

The circuit

The transmitter circuit is shown in **Figure 1**. TR1 is a crystal oscillator, the frequency of which is controlled by crystal X1. A small trimmer capacitor, TC1, is added to allow the frequency of X1 to be varied by a small amount. If adjustment of this trimmer is made possible from the front panel, it is useful to adjust the transmit frequency to avoid other stations already on the crystal frequency. The collector load resistor, R2, of the oscillator transistor, TR1, determines the power output; a value of 3.3 kΩ seems to work well in producing an output of 3 watts.

TR1 is directly coupled to TR3, a VMOS transistor (a type of field-effect transistor (FET)). This acts as the power amplifier (PA) stage. TR3 should

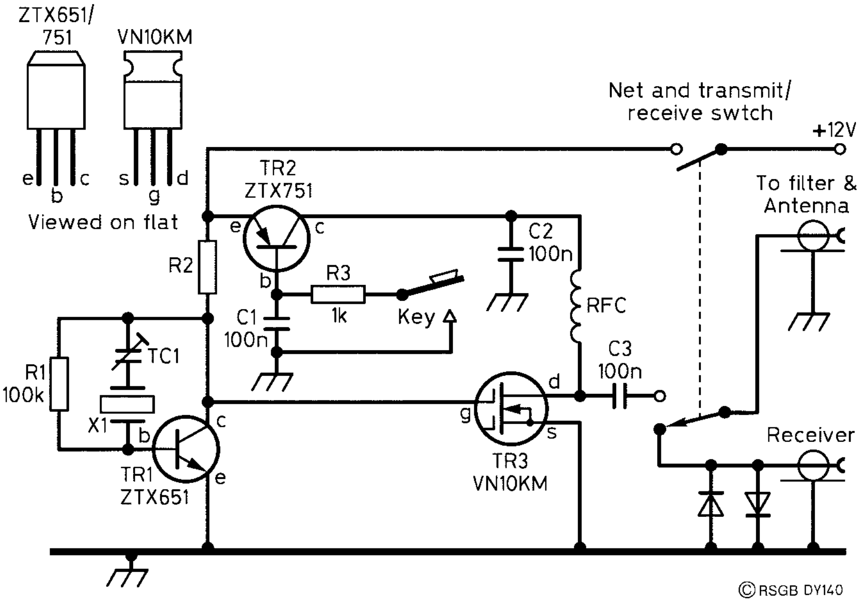
ZTX651/
751
VN10KM
e b c
s g d
Viewed on flat
Net and transmit/
receive swtch
+12V
TR2
ZTX751
To filter &
Antenna
e
c
b
C2
100n
R2
R3
1k
Key
RFC
C1
100n
TC1
R1
100k
X1
C3
100n
d
g
s
TR3
VN10KM
TR1
ZTX651
Receiver
©RSGB DY140

Figure 1 Circuit diagram of the breadboard transmitter

give about 3 W output, which is then coupled to the output by C3. The radio-frequency choke (RFC) providing the drain load of TR3 is simply a few turns of wire on a ferrite bead.

TR2 is an interesting addition to the circuit. It is used as a switch to 'key' the PA, TR3. The transmitter *could* be built without TR2, just placing the Morse key between the top of the RFC and the 12 volt supply. Adding TR2 is helpful, because it means that one side of the Morse key can be grounded (always a good thing), and some degree of *shaping* of the output RF waveform is provided by R3 and C1. This makes the transmission sound a little better and reduces the possibility of spurious frequencies being generated and transmitted. TR2 is a pnp transistor; note that it is the *emitter* of this transistor which is connected to the positive side of the supply.

Some form of changeover switching is needed for the aerial. A double-pole changeover toggle switch can be used. See the chapter on switches, later on in this book. One pole is used to switch the aerial between transmitter and receiver; the other pole is connected in the 12 volt supply line, and is labelled RECEIVE/TRANSMIT – NETTING. Its use will be described later. In this simple circuit, the PA cannot work when the key is open, because the key breaks its supply (via the RFC). When the key closes, TR2 switches on and applies the 12 volt supply to the top of the RFC. C2 is a *decoupling* capacitor, which prevents any residual RF signals at the top of the RFC reaching TR2.

Building

The prototype was put together as follows. Take a piece of plain printed-circuit board (PCB) measuring 5 cm by 4 cm. Then, with a *new*, *sharp blade* in a junior hacksaw, draw the blade horizontally across the surface of the copper in order to make a pattern of 6 squares along the 5 cm side and 5 squares along the 4 cm side. No more pressure should be applied than is necessary to cut through the copper! All the parts will be soldered on these pads in a form of surface-mount construction. To do this, each active pad (i.e. one that is going to have a component soldered to it) needs to be *tinned*. This means coating the pad's surface with solder, and is carried out as follows. Place the hot tip of the soldering iron on to the pad, and hold it there for a second or so. Then, with the tip still in place, touch the end of your reel of solder *on the pad*, not the tip of the iron. The solder should flow evenly all over the pad, and you can remove the iron. The solder should solidify in a rounded, shiny blob! This provides a good surface for making soldered joints.

To join component leads to the pads, cut each lead about 1 cm long, and then bend the last 2 mm at right angles to the rest of the lead. As you did before, tin the 2 mm length of each lead. Place the tinned portion on to the pad, and place the tip of your iron *on the pad*, close to the lead. The solder on the pad and on the lead will melt and run together; remove the iron and *hold the component still* until the solder solidifies. When the joint has cooled, give the lead a gentle tug to make sure you have a good joint (a good *mechanical* joint is usually a good *electrical* joint, too!). Each transistor straddles three pads, so the centre lead will need to be shorter than the other two. Take care here to get the lead lengths right – if you do, you will be surprised how much more firmly the transistor is held than if you just botched the lead lengths by bending them to fit! Make sure the connections to the transistors are correct.

Winding the RFC is quite simple. Seven (or more) turns of thin (32 SWG) enamelled copper wire are threaded through a small ferrite bead. This requires care, because the bead is small and the wire is thin. Trim the ends to within about 1 cm of the bead, remove the enamel carefully with sandpaper and tin the bare ends, prior to soldering the choke to the board.

After completion of the wiring, check the circuit against Figure 1. Bread-boarding a circuit like this has its advantages, but it can have disadvantages, too. One of these disadvantages is that it can make circuit checking difficult. For a simple circuit like this, it is not too bad! Check that no solder has run between the pads. Plug in your crystal for the 80 m band, and connect up the 12 V power supply. Do not connect the Morse key yet and do not switch on the power.

Testing and operating

Clear your workbench of all metallic objects, slivers of copper, bits of wire, etc., switch on your power supply. With an external receiver, listen on and around your crystal's frequency for a signal. Remember that the oscillator runs all the time and, because you haven't yet connected the Morse key, your receiver is close enough to pick up the signal from the oscillator. This confirms that your oscillator is running. Switch off.

Connect the station aerial to the transmitter's aerial socket, and the receiver to the transmitter's receiver socket. Connect the Morse key, put the Receive/ Transmit–Netting switch in the receive position and switch on. You should be able to hear stations in the normal way. Now put the switch in the Transmit–Netting position. Signals in the receiver should almost disappear, as the circuit has disconnected the receiver's aerial.

Tune the receiver until you can hear your own crystal oscillator signal. This is known as *netting*, tuning your receiver and transmitter to the same frequency. Pressing the key will now transmit your signals when the switch is in the Transmit–Netting position; switch back to the receive position to listen for stations answering your call. As soon as you are happy that your circuit is functioning properly, **you must build the low-pass filter circuit** before using the transmitter regularly.

Parts list

Resistors: all 0.25 watt, 5% tolerance	
R1	100 kilohms (kΩ)
R2	See text
R3	1 kilohm (kΩ)
Capacitors	
C1, C2, C3	100 nanofarads (nF), or 0.1 microfarad (μF)
TC1	3–60 picofarads (pF) trimmer
Semiconductors	
TR1	ZTX651
TR2	ZTX751
TR3	VN10KM
Additional items	
RFC	7 turns of 32 SWG enamelled copper on a ferrite bead
Switch	Double-pole changeover (DPDT or DPCO) toggle
Crystal	For 80 m band
Crystal holder	HC25 type
Sockets	According to station fittings

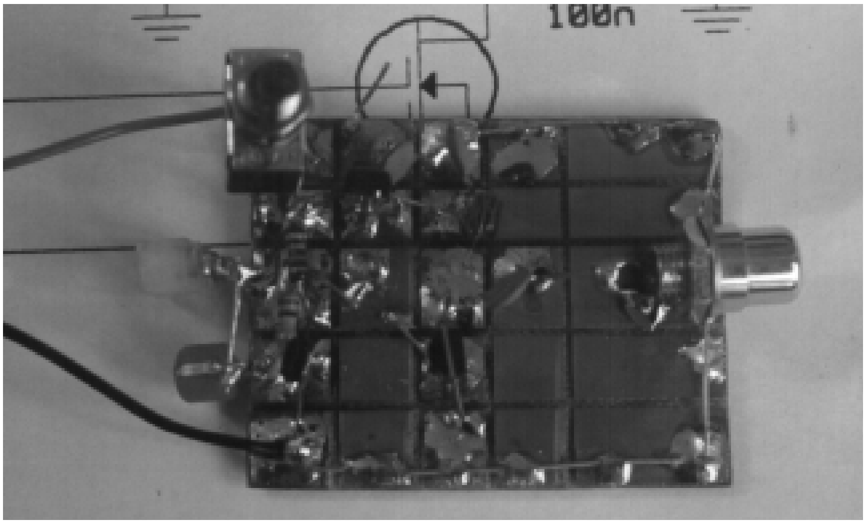

54 A 7-element low-pass filter for transmitters

Introduction

The simpler the transmitter, the more likely it is to radiate *harmonics* of its fundamental frequency. Harmonics are integral multiples of the frequency on which the transmitter is *designed* to operate. If you think you are transmitting on a frequency, *f*, for instance, you will also be radiating the harmonics of *2f*, *3f*, *4f*, . . . and so on. This results in your signals being heard on *several frequencies*, spread over a very wide frequency range. You are also contravening the terms of your licence. To avoid this, it is always advisable to use a *low-pass* filter between your transmitter and aerial. This is a filter which will *pass* your signal frequency, *f*, and all frequencies *below* it, but will *not pass* (attenuate) frequencies above *f* to any significant extent. The value of *f* is known as the *cutoff frequency* of the filter.

A design of 7-element low-pass filter

A 7-element low-pass filter (LPF) is so called because it has seven components, as the circuit diagram of **Figure 1** shows. Filters containing any odd number of elements are possible: a 3-element filter would

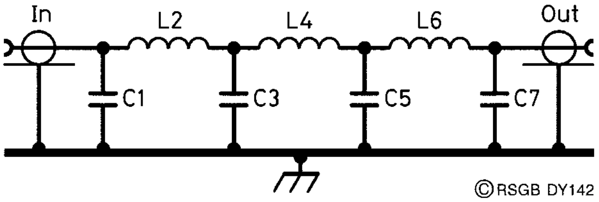

Figure 1 Circuit diagram of filter

comprise C1, L2 and C3 only, and is sometimes called a pi-network because the element disposition resembles the Greek letter pi (π); a 5-element filter would comprise C1, L2, C3, L4 and C5 only, and so on. In general, the more elements the filter has, the more effectively it attenuates signals above *f*.

The circuit of Figure 1 is designed to have an input and an output impedance of 50 Ω, which means that it can be placed in the aerial feed of any common transmitter. Filter design is a very complex business, and is best left to the experts. One such expert is W3NQN, who produced a number of computer designs of LPF using commonly available (*preferred value*) capacitors, and aimed specifically for use on amateur frequencies. The results of this work are condensed into **Table 1**. The inductors are wound on standard toroidal cores, and their details are included in the table.

Table 1 Filter component values for each brand

Band metres	*C1,7 pF*	*C3,5 pF*	*L2,6 turns*	*L4 turns*	*Core type*	*Wire SWG*
80	470	1200	25	27	T37-2	28
40	270	680	19	21	T37-2	26
30	270	560	19	20	T37-6	26
20	180	390	16	17	T37-6	24
15	82	220	12	14	T37-6	24
10	56	150	10	11	T37-6	22

Making the filter

The filter was made originally as an adjunct to the *Breadboard 80 m CW transmitter*, which you will also find in this book. It uses the same constructional technique, based on a single piece of plain, copper-clad

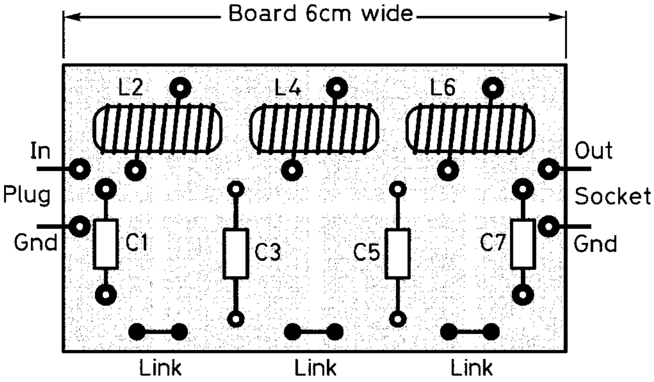

Figure 2 Layout and construction of filter

©RSGB DY141

PCB, with 'pads' created by using a sharp blade in a junior hacksaw. The cuts in the copper are shown in the layout diagram of **Figure 2**. There are two ways of mounting the components: the first way is to drill small holes in each pad, as shown in Figure 2, and mount the components through the holes in the normal PCB manner; the second way is to solder the components directly to the pads, in the way that was described for the *Breadboard transmitter.*

Winding the inductors is quite simple. All you need to remember is that each time the wire passes through the core counts as one turn. Cut off the spare wire at each end of each coil to about 1 cm, scrape off the enamel with sandpaper, and tin the exposed copper. See the transmitter description if you are unsure of how to do this. Note the wire links between each of the lower pads, forming a solid 'ground' for the elements. The prototype had a plug and socket on the ends, to match the transmitter and aerial terminations.

The type of capacitor used in the design is not critical; the polystyrene type works well.

55 Radio-frequency mixing explained

Introduction

Mixers find widespread use in electronic circuitry. Many of the projects in this book, together with every TV set and radio in the home, contain mixer circuits – a good indication of their usefulness.

Confused?

Audio mixers (as used in recording studios and radio broadcast stations) are used to **add** or ‘balance’ the signals from various sources such as microphones, CD players, etc. These have nothing whatsoever to do with radio-frequency (RF) mixers, and should never be confused with them.

RF mixers and beat frequencies

Instead of **adding** signals (as in the audio mixer), the RF mixer **multiples them together.** As you might expect, this has an entirely different effect. The two signals entering the mixer *beat* or *heterodyne* with each other to produce *signals on other frequencies*. One example of this occurs in sound, when two musical notes of almost the same frequency are heard together. Instead of hearing two separate notes, the listener hears one note whose *intensity* (*loudness*) appears to increase and decrease. This intensity variation is called a *beat*, and its frequency is equal to the difference in the frequencies of the two original notes. The technique is used by musicians to tune their instruments. If one note is known to be a correct frequency, the other can be tuned to it by making the beat frequency as close to zero as is possible.

Multiplying together

The process of mixing presupposes that we have a device which will automatically multiply two signals together. Fortunately, this is easy; so easy, in fact, that it often occurs when we do not want it! Multiplying is achieved by any device which is *non-linear*; this means a device whose output is not a constant factor larger than its input, something that can be achieved by many electronic devices and circuits.

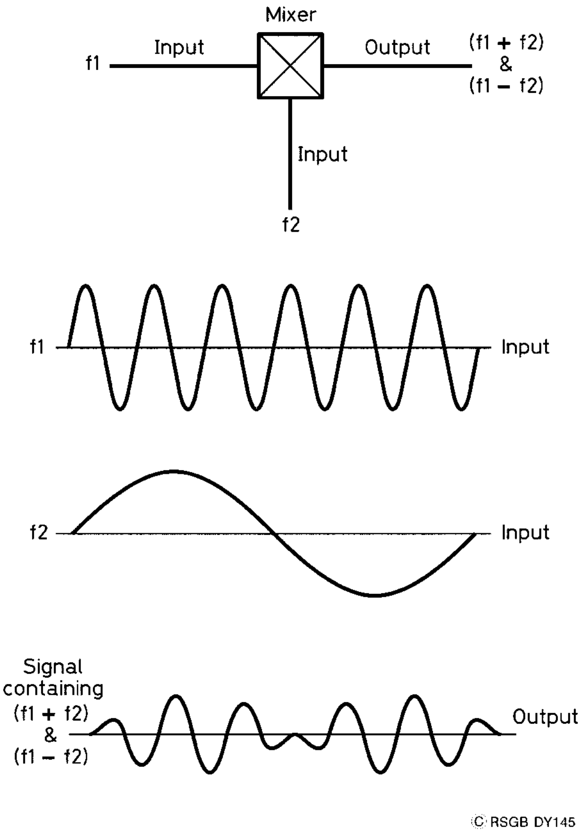

Figure 1 The effect of multiplying (or mixing) two signals together

Let us look now at what a mixer does in concrete terms. Suppose two signals, of frequencies *f*1 and *f*2 go into our mixer. These signals are shown in **Figure 1**. Putting numbers in, to make the situation clearer, suppose *f*1 is 1.000 MHz and *f*2 is 160 kHz. The beat frequency is the *difference* of these: 1.000 MHz – 0.160 MHz = 0.840 MHz, or 840 kHz. A mixer *also* produces an output at the *sum* of these frequencies; in this case the new frequency would be 1.000 MHz + 0.160 MHz = 1.160 MHz.

Suppose you fed the output of your mixer, operating with these input frequencies, into a receiver and tuned around to find what frequencies were present. You would find two signals, one at 840 kHz and one at 1.160 MHz, showing that the two 'new' frequencies were very real!

In addition to drawing out the waveform of the resultant signal, as in Figure 1, we can draw the inputs and outputs on a frequency axis, to form a *spectrum* of the signal components. This is done in **Figure 2**. The top two diagrams show the input signals at *f*1 and *f*2. The bottom diagram shows the output signals in relation to the input signals. Depending on the type of mixer used, one or both of the input signals would be removed.

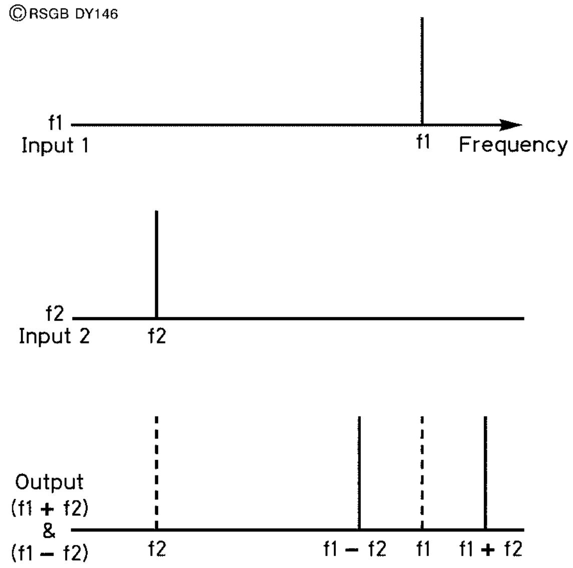
©RSGB DY146
f1
Input 1
f1
Frequency
f2
Input 2
f2
Output
(f1 + f2)
&
(f1 – f2)
f2
f1 – f2
f1
f1 + f2

Figure 2 The result of multiplying or mixing two signals together as seen on a spectrum analyser

A mixer in every radio

Basically, a mixer is used to change a signal from one frequency to another, something it does without altering the characteristics of the incoming signal. If the incoming signal is amplitude modulated (AM), then the frequency-changed signal would be AM also. The same applies to FM, SSB, CW, and all other modulation forms you can think of. This explains why mixers are often called *frequency changers*.

Frequency changing is the key process in the type of radio known as a *superheterodyne* (or *superhet*). By mixing the incoming signal with a variable-frequency *local oscillator* as **Figure 3**, shows, the signal can be converted to the fixed frequency of a filter and amplifier. This is useful

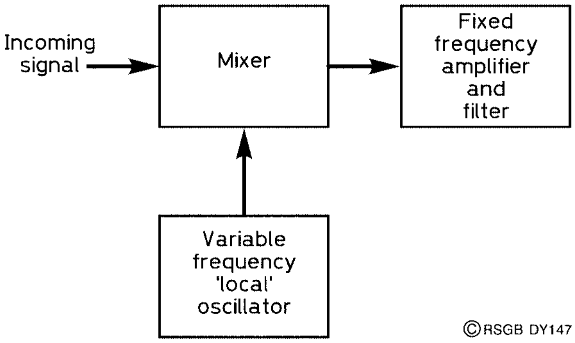
Incoming signal
Mixer
Fixed frequency amplifier and filter
Variable frequency 'local' oscillator
©RSGB DY147

Figure 3 The basic idea of a superhet receiver

because it is easier to make a very high-quality filter and amplifier at a single fixed frequency, than at a variable frequency.

All TV receivers and virtually all radio receivers (and transmitters) use mixers. Both the *Yearling* and *Colt* receivers (see the relevant projects) are superhets, and use mixers.

56 A voltage monitor for a 12 V power supply

Introduction

If for any reason, the stabilisation of your main 12 volt power supply unit (PSU) breaks down, it is possible that a voltage much higher then the nominal 13.8 V will be applied to your precious equipment. If you would like to know the instant that this occurred, and hence be able to switch things off before it was too late, then this circuit is what you need. It will give audible *and* visual indications if the voltage rises above 14.4 V, and a visual indication only if the voltage is reduced.

The circuit

The circuit uses three ICs and is shown in **Figure 1**. The circuit is powered by the PSU whose output is being monitored, and the circuit's immunity to supply line variations is secured by the 6 volt regulator, IC1. The heart of the circuit is IC2, an LM3914; it is a *bargraph driver*, which operates ten LEDs in a display resembling a thermometer – the string of lit LEDs increases in length as the voltage on pin 5 increases.

The input voltage range on pin 5 is 1.2 V maximum, making each LED correspond to one-tenth of this, which is 0.12 V, the *step size*. R1 and R2 act as a voltage divider, so that voltages of up to the maximum of 14.4 V may be applied to R1 without exceeding 1.2 V at pin 5. R3 sets the brightness of the LEDs and R4 determines the step size.

IC3 is an *opto-isolator*, a device containing an LED and a phototransistor in one package. This enables the piezoelectric sounder to operate without affecting the operation of the bargraph driver. The input to IC3 is provided by the voltage on D8, so that if any of the LEDs at or above D8 are lit, the

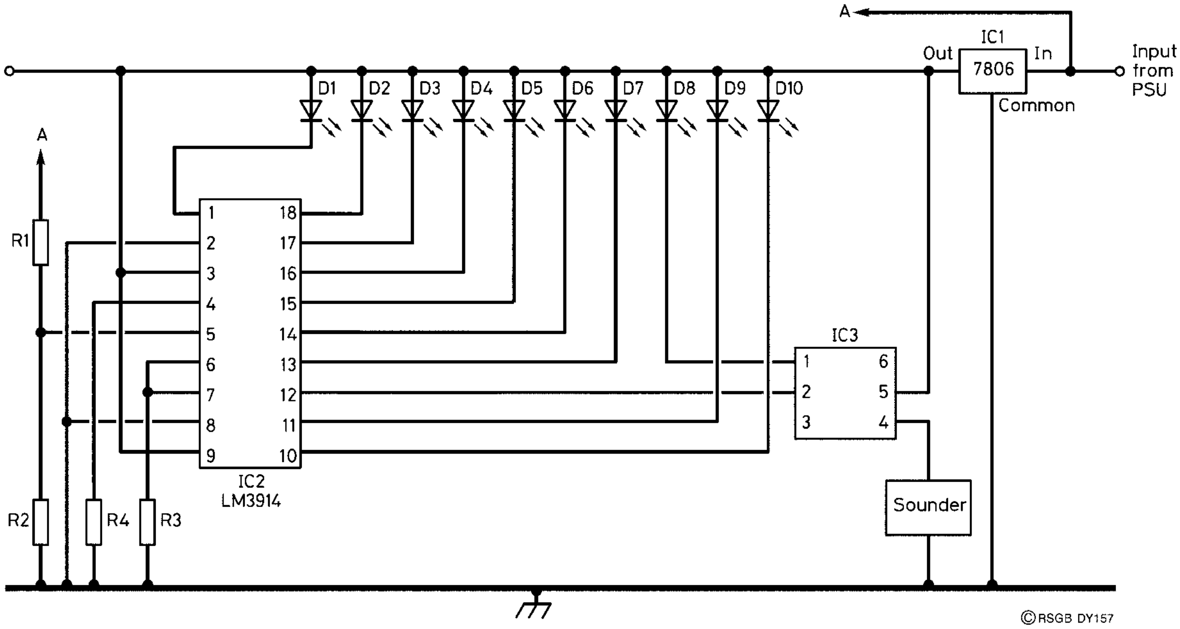

Figure 1 PSU monitor, circuit diagram

sounder will operate, indicating overvoltage. Despite the fact that some of your other equipment might be damaged by this overvoltage, the monitor circuit itself is unaffected.

Construction

The prototype was built on two pieces of Veroboard of the copper-strip variety. The main (circuit) board measured 15 strips by 25 holes, and the display (LED) board measured 4 strips by 30 holes. Track cuts are necessary in each board. The correct places are shown for the main board in **Figure 2**, but there is sufficient flexibility in the layout of the display board for a prescriptive layout not to be needed. All the anodes of the LEDs are connected to the same strip, which makes things comparatively simple!

For the main board, insert the Veropins and the wire links first, and solder them to the copper strips. Then fit the IC holders and the resistors. Fit the IC holders with their notches towards the top of the board. When fitting the voltage regulator, IC1, note than the centre lead (the 'common' lead in Figure 1) does *not* go to position A3; the track should be cut at A3, and the common lead soldered at B3. This is indicated in Figure 2.

Wire up the LED board with its ten LEDs and 11 connecting wires, each about 10 cm long. This length depends on how far away from the main

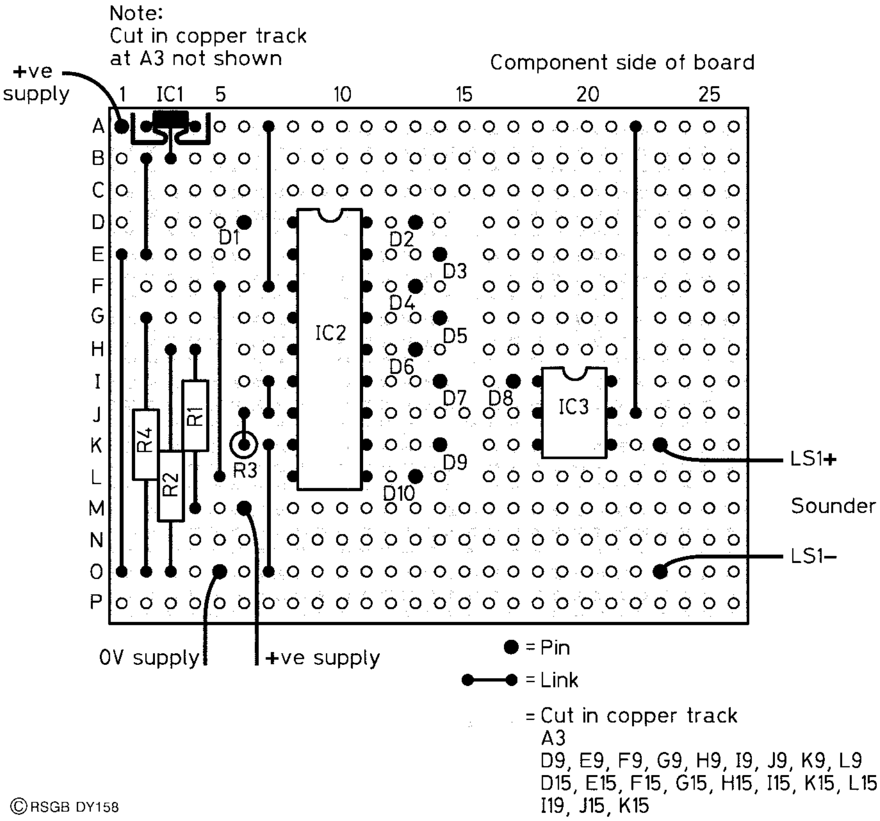
Note:
Cut in copper track
at A3 not shown
+ve supply
Component side of board
1
IC1
5
10
15
20
25
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
D1
D2
D3
D4
D5
D6
D7
D8
D9
D10
IC2
IC3
R1
R2
R3
R4
LS1+
Sounder
LS1−
0V supply
+ve supply
● = Pin
●—● = Link
= Cut in copper track
A3
D9, E9, F9, G9, H9, I9, J9, K9, L9
D15, E15, F15, G15, H15, I15, K15, L15
I19, J15, K15
©RSGB DY158

Figure 2 PSU monitor, PCB layout

board you are planning to mount the display. The LEDs should have different colours: three orange, four green and three red, to indicate ‘low’, ‘medium’ and ‘high’ voltage.

Now solder the other ends of these leads to the Veropins on the main board, making sure that the order is correct. Finally, connect the piezoelectric sounder to the main board; the polarity is important, so make sure the red lead goes to pin 4 of IC3 and the black lead to the ground rail. If you were careful to fit the IC holders with their notches in the correct positions, match these up with the notches on the ICs before pushing home the ICs *gently*. Check your circuit for solder bridges and unwanted pieces of copper swarf before screwing the small heat sink to the voltage regulator, IC1.

Testing

(a) **With a variable-voltage PSU**. Set the PSU for 10 V, connect the circuit and switch on. Increase the voltage slowly, and check that each LED lights up after the one before it. As the voltage exceeds about 14.4 V, the first red LED should light and the sounder should operate. If an LED does not illuminate, you should immediately suspect either a dry joint or an incorrect LED polarity. The voltage at which the first red LED lights can be adjusted by varying R1; increase R1 if the LED comes on too early; decrease R1 if the LED comes on too late.

(b) **Without a variable-voltage PSU**. For your 'variable supply', you can use several AA-type 1.5 V cells (or 1.2 V NiCad cells) in series. The voltages produced by a range of cells is shown in **Table 1** – because of the lower voltage of NiCad cells, more of them are needed to produce a given voltage.

Once the operation of the circuit has been checked, it can be fitted into a plastic or metal box. Only two connections are needed for the PSU. You may want to drill some holes in the case to increase the apparent loudness of the sounder.

Table 1 Test voltages available from batteries in series

No. of batteries	*Voltage, nicads*	*Voltage, dry cells*
7	8.4	10.5
8	9.6	12
9	10.8	13.5
10	12	15
11	13.2	–
12	14.4	–
13	15.6	–

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1	11 kilohms (kΩ) – see text
R2	1 kilohm (kΩ)
R3	1.2 kilohms (kΩ)
R4	18 kilohms (kΩ)

Semiconductors

IC1	L7806
IC2	LM3914
IC3	Opto-isolator – Maplin code WL35Q
D1–D3	3 mm LED, orange
D4–D7	3 mm LED, green
D8–D10	3 mm LED, red

Additional items

LS1	Piezoelectric sounder, wire leads
	6-pin DIL socket for IC3
	18-pin DIL socket for IC2
	Veroboard – two pieces for main and display boards, see text for sizes
	Veropins
	Heat sink for IC1
	Single-core insulated wire for links
	Insulated stranded wire for interconnecting the boards
	Case as required

57 A 1750 Hz toneburst for repeater access

Introduction

Repeaters across the UK and much of Europe need an access tone to switch the transmitter from standby ready for use. Commonly, this is a 1750 Hz tone of duration no greater than half a second. Although many UK repeaters may now be accessed using the *continuous tone-coded squelch system* (CTCSS – see the *RSGB Yearbook*), you may wish to access a repeater whose CTCSS frequency you don't know; in this case, using the universal 1750 Hz tone will gain you access. Commercial transceivers are usually fitted with an automatic toneburst, but if you are using a home-made design, then you may want to incorporate this little circuit.

Warning

This circuit uses a member of the integrated circuit family known as CMOS (complementary metal-oxide semiconductor). These use very little current

and can be completed destroyed if they come into contact with the magnitudes of static electricity that most of us carry about when we walk on carpets and wear rubber shoes. You will never know if this wanton destruction has happened – all you will discover is that your circuit doesn't work and that you have tested *everything*. To avoid this problem do the following things:

1. **Before you open the little packet in which the IC is supplied, touch something which you *know* to be earthed – the metalwork of any equipment which is mains earthed, for example. Then open the packet.**
2. **Let the IC fall gently on the bench – don't pick it out with your fingers. Touch your earthed metalwork again. Pick up the IC and insert it gently into its holder.**

The circuit is safe from destruction while it is connected to the battery.

Circuit description

The circuit is shown in **Figure 1**. The tone is generated by an integrated circuit oscillator (IC1), whose frequency is controlled by a *ceramic resonator*, XL1. Its frequency is very high, and is divided down to the 1750 Hz needed by the same chip. The ceramic resonator is designed to operate at 455 kHz, the intermediate frequency of many receivers. Because all divider circuits use powers of 2, we need the oscillator to run at 448 kHz so that when it is divided by 256 ($256 = 2^8$), we end up with 1750 Hz. Try it on your calculator:

$$\frac{448\ 000}{256} = 1750.$$

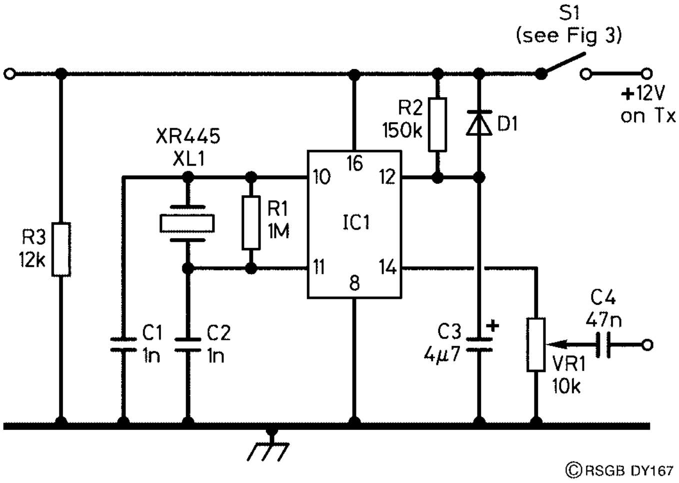

Figure 1 Toneburst module, circuit diagram

We use C1, C2 and R1 to *pull* the frequency of the oscillator away from 455 kHz to 448 kHz. The divider chain has eight *counters* in it, and each counter divides the frequency of the signal it sees by two, giving the final division of 256.

This counting process can be stopped at any time by taking the voltage on the *reset* pin (pin 12) up to the supply voltage. When the circuit is switched on by closing S1, pin 12 is at 0 V because C3 is discharged. The oscillator runs, producing the output frequency of 1750 Hz. As time progresses, C3 charges up through R2 and the voltage on pin 12 rises. When this has risen sufficiently, and in a time determined by the values of C3 and R2, the counter resets and stays in the reset state; no division takes place and there is no output. The duration of the toneburst is thus governed by C3 and R2.

When S1 is opened, the circuit is switched off, and C3 is discharged through D1 and R3, ready for the next toneburst. If you have used a repeater, you will know that a toneburst is needed **only** to activate a repeater in the standby condition; it is not needed once a contact has been established.

VR1 adjusts the amplitude of the tone fed to the microphone, and C4 prevents any voltage that may be present on your microphone connector from damaging the integrated circuit.

Construction

The prototype circuit was built on Veroboard of the copper-strip variety, measuring 20 holes by 14 strips. The layout is shown in **Figure 2**. Make the track cuts first, and check that there are no slivers of copper wedged between adjacent tracks. Then, solder in the IC socket (with the notched end

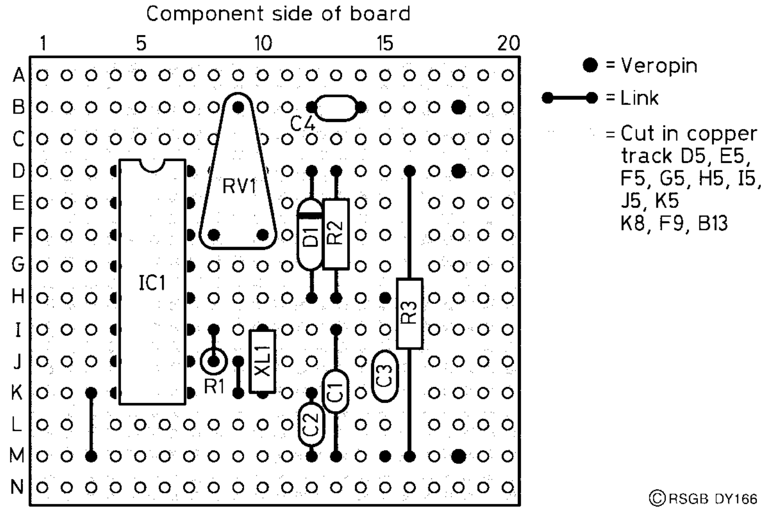

Figure 2 Veroboard component layout

facing towards track A), the wire links and the three Veropins. Having done this, solder in the resistors, capacitors and diode, making sure that D1 and C3 are the right way round! Using your best soldering technique, solder in the ceramic resonator quickly, to prevent heat damage. Recheck your circuit, check for solder splashes and bridges, and then gently insert IC1 into its socket, matching up its notch with that of the socket.

Testing

Set VR1 to half-way and connect a crystal earpiece to the output; apply power to the circuit. You should hear the tone, lasting for about half a second. If there is no tone, disconnect your circuit from the power supply, and check for dry joints in the vicinity of pin 12. Is the diode, D1, the correct way round? Is C3 the correct way round? Did you choose to ignore the CMOS safety precautions given earlier?

Once the circuit is working, you need to decide how you are going to connect it to your transmitter. Two options are shown in **Figure 3**. If you have access to a point in your transmitter circuit that has between 9 V and 12 V positive on it during transmit, you can use this to power your circuit. As the toneburst is needed only for repeaters, the switch, S1, disconnects it when not needed, as shown in Figure 3a. If you want the circuit to be self-powered, then a 9 volt PP3 battery may be used; Figure 3b shows this configuration.

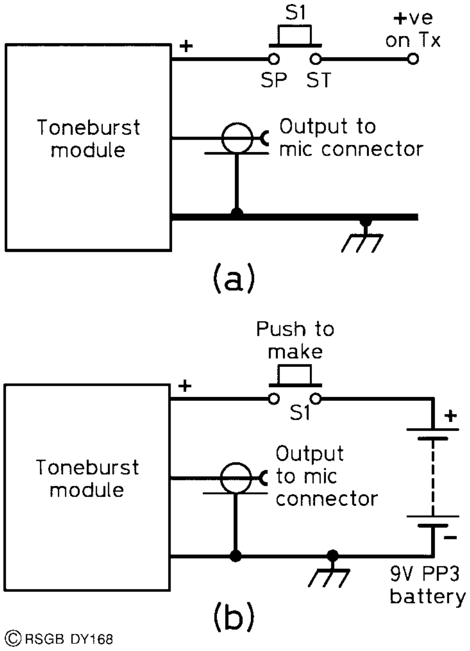

Figure 3 Toneburst module: two alternative switching arrangements

The output from the circuit board is fed directly into the microphone socket, in parallel with the microphone itself; use thin coaxial or screened cable for this lead, or you may induce hum into the microphone circuit and suffer from RF breakthrough into the audio circuits. To adjust the setting of VR1, start with it at the zero output position and connect a dummy load to your transmitter. Slowly, increase the output while monitoring your transmitted signal on another nearby receiver. Make sure you do **not** increase the output so far that the signal sounds distorted. If you would prefer that the tone was on continuously while you made this adjustment, simply connect a wire across C3 remembering, of course, to remove it as soon as you have completed the test!

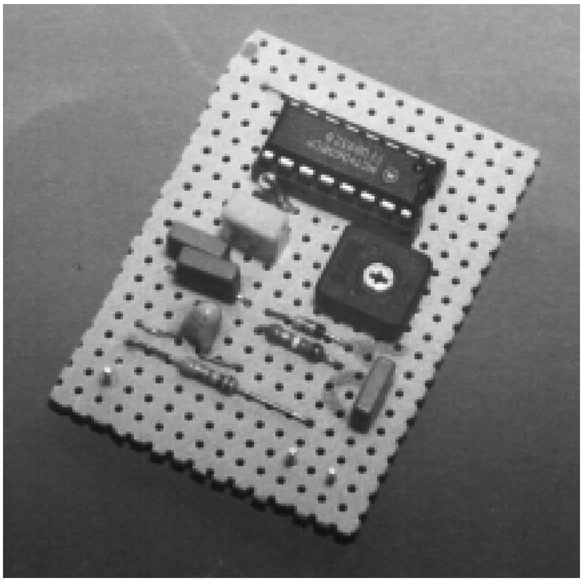

Parts list

Resistors: all 0.25 watt, 10% tolerance (or better)

R1	1 megohm (MΩ)
R2	150 kilohms (kΩ)
R3	12 kilohms (kΩ)
VR1	10 kilohms (kΩ) horizontal preset

Capacitors: all 16 V WKG or higher

C1, C2	1 nanofarad (nF) disc ceramic
C3	4.7 microfarad (μF) tantalum bead
C4	47 nanofarads (nF) disc ceramic

Semiconductors

IC1	4060
D1	1N4148

Additional items

XL1	XR455
	Veroboard (see text for size)
	Veropins (3)
S1	Switch (momentary action push-to-make SPST)
	16-pin DIL socket for IC1
	Single-core insulated wire for links
	Coaxial or screened cable for microphone connection

58 A circuit for flashing LEDs

Introduction

There are many occasions when one's attention needs drawing to the fact that something important has happened. A single red light coming on is seldom sufficient to attract attention, particularly if it is surrounded by other lights and indicators. The eye is known to be very sensitive to *changes* in its peripheral vision; such changes can be brought about by movement or by differences in light level – a flashing light, for example. So, a circuit that flashes a single LED or a pair of LEDs finds plenty of uses in the amateur station.

Warning

This circuit uses a member of the integrated circuit family known as CMOS (complementary metal-oxide semiconductor). These use very little current and can be completed destroyed if they come into contact with the magnitudes of static electricity that most of us carry about when we walk on carpets and wear rubber shoes. You will never know if this wanton

destruction has happened – all you will discover is that your circuit doesn't work and that you have tested *everything*. To avoid this problem do the following things:

1. **Before you open the little packet in which the IC is supplied, touch something which you *know* to be earthed – the metalwork of any equipment which is mains earthed, for example. Then open the packet.**
2. **Let the IC fall gently on the bench – don't pick it out with your fingers. Touch your earthed metalwork again. Pick up the IC and insert it gently into its holder.**

The circuit is safe from destruction while it is connected to the battery. However, when the battery is removed, the same care should be exercised with its handling, because there is no supply decoupling capacitor across the IC.

Basic description

LEDs can be made to flash (switch on and off) by driving them from sources that switch on and off. Such a source is an *astable multivibrator*. If you have built or read about *A basic continuity tester*, elsewhere in this book, you will have come across such a beast before. That circuit used an astable multivibrator made from two transistors. This new circuit achieves the same behaviour from a single integrated circuit, the CMOS 4011. To give it its full description, the 4011 is a *quad 2-input NAND gate*. Quite a mouthful, but all it means is that inside the chip are four NAND gates, each with two inputs.

A NAND gate needs a positive voltage (known as a *logic 1*) on *both* inputs in order to produce zero volts (known as *logic 0*) at the output. Two NAND gates can be connected, as are A and B in **Figure 1**, to make our astable multivibrator. The combination of A and B has been described as the most perverse circuit in electronics; as soon as the output goes to logic 1, the circuit decides that it would prefer to have a logic 0 there, and switches over. With logic 0 at the output, the circuit now prefers to have logic 1 there, and so it goes on! We are going to use this continuous switching backwards and forwards to flash two LEDs. The rate at which A and B 'change their minds' is the frequency at which our LEDs will flash, and is controlled by the charging and discharging times of C1 through R2 and C2 through R1. As the values of R1 and R2 are the same, and those of C1 and C2 are the same, the ON and OFF states of the circuit are the same.

Gates C and D do not contribute to the flashing action; they act as *buffers* to isolate the LEDs from the multivibrator circuit itself. You will find in electronics that an oscillator is seldom used to drive another device *directly*; there is usually a buffer between it and the stage it drives.

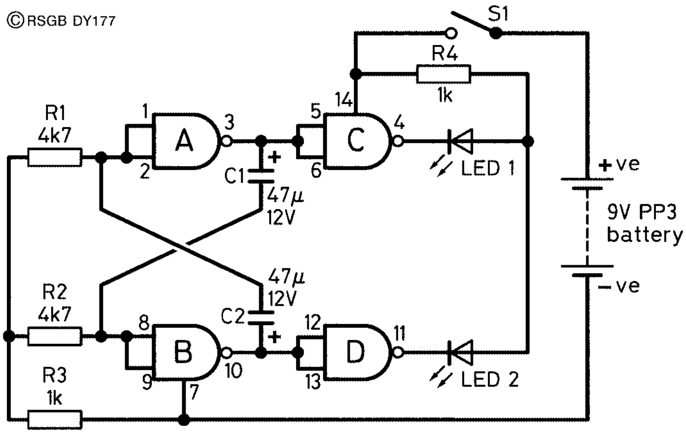

Figure 1 Flashing LEDs, circuit diagram

A characteristic of all multivibrators, astable or not, is that they have two outputs. In this case, those outputs are at pins 3 and 10, which are then buffered and appear at pins 4 and 11, respectively. When one output is at logic 1, the other is at logic 0, and vice versa. This means that LED1 is off when LED2 is on, and LED1 is on when LED2 is off, the two states switching backwards and forwards at the frequency of the oscillator.

Construction

Read the warning at the beginning of this article again. It is not intended to scare you off from building this, but is a genuine piece of advice which can save you time and irritation when all your labours result in a circuit that doesn't work! That extra bit of care can make all the difference!

Veroboard (the copper strip type) is used for the layout, shown in **Figure 2**. It measures 20 holes by 12 strips. **Be aware that there is no row 'I' in the layout, so don't miscount when you are placing components on the board!**

Firstly, cut the tracks using a 3 mm ($\frac{1}{8}$ inch) twist drill held between thumb and forefinger; check that there are no slivers of copper bridging any of the tracks, and that the tracks have been completely cut by the drill. Solder in the components carefully. Leave the IC in its carrier for the time being, and solder in the IC socket, with the notched end facing row A. On completion, check the circuit carefully. If you are happy that it is correct, follow the instructions given earlier and fit IC1 into its socket, matching up the two notches. Connect the battery and switch on. The two LEDs should flash on and off alternately. If only one LED flashes, you have probably connected the other one the wrong way round. Switch off, check and correct if necessary. If neither LED flashes, you must have a significant error in your circuit, which will need checking *again*! Or did you choose to ignore the handling precautions for the CMOS chip?

Figure 2 Flashing LEDs, component layout

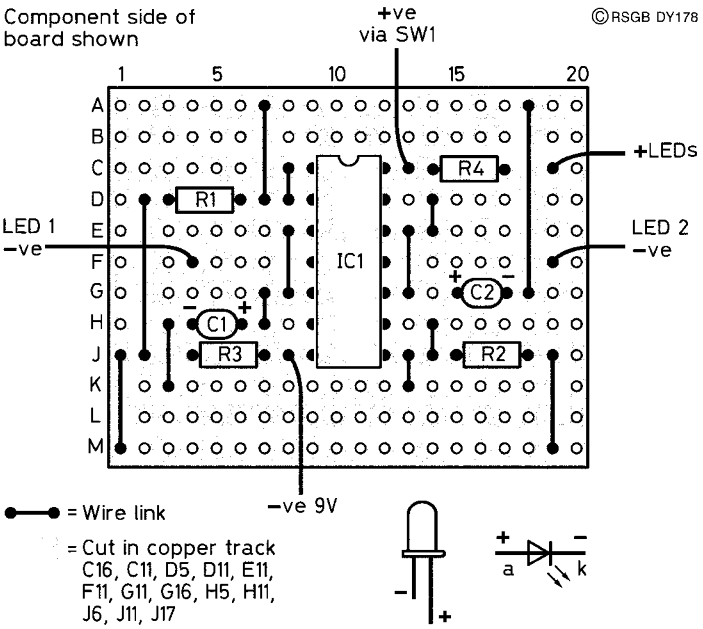
Component side of board shown
+ve via SW1
©RSGB DY178
1
5
10
15
20
A
B
C
D
E
F
G
H
J
K
L
M
R1
R4
IC1
C1
C2
R3
R2
+LEDs
LED 1 −ve
LED 2 −ve
−ve 9V
= Wire link
= Cut in copper track C16, C11, D5, D11, E11, F11, G11, G16, H5, H11, J6, J11, J17
−
+
a
k

The board can be mounted near to the point where you want your flashing LEDs to be seen, although long leads to the LEDs are acceptable. The LEDs can be different colours – it's all up to you now!

Parts list

Resistors: all 0.25 watt, 10% tolerance or better	
R1, R2	4700 ohms (Ω)
R3, R4	1 kilohm (kΩ)
Capacitors	
C1, C2	47 microfarads (μF) 12 V WKG
Semiconductors	
IC1	4011
Additional items	
LED1, LED2	Any size of LED, any colour
S1	SPST on/off
	Plastic box if needed, 8.5 by 5 by 2.5 cm

Source

Components are available from Maplin.

59 Digital logic circuits

Introduction

Logic circuits form the backbone of even the most advanced computer, yet their basic operation can be demonstrated by a couple of switches, a battery and a bulb.

Logic using switches

Everyone reading this article will look at **Figure 1** and know immediately how it works and be able to write down something like 'When switch A and switch B are closed, the light will come on'. Without knowing it, you have written down a logic statement involving the so-called **AND** operation; the light comes on only when switches A **AND** B are ON. Below the circuit in Figure 1, is a table showing the only possible positions of the two switches and the state of the bulb for each position. This is called a *truth table*, and is frequently used in logic analysis.

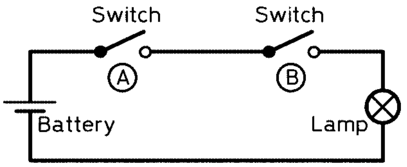

A	B	Lamp
Off	Off	Off
On	Off	Off
Off	On	Off
On	On	On

©RSGB DY182

Figure 1 Switches and lamp AND gate

Figure 2 shows a different circuit. Here, the two switches are in parallel rather than in series, as was the case in Figure 1. Again, if you analyse the circuit in words, you would say that the light will be on when switch A **OR** switch B is ON. This is an example of the **OR** operation, and its truth table is shown in Figure 2. The statement above is not complete, however; can you see why? The truth table will show you. The light comes on if A is ON, **OR** if B is ON, ***OR*** if A **AND** B are both ON. That third condition is easy to miss, but don't worry about it!

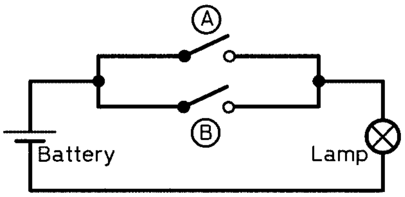

A	B	Lamp
Off	Off	Off
Off	On	On
On	Off	On
On	On	On

©RSGB DY184

Figure 2 Switches and lamp OR gate

Believe it or not, some very complicated logic is possible (in theory) using switches and lights, but it is highly impractical and would be very slow. This is where electronic logic circuits come in.

Switches with no moving parts

You may have come across projects in this book where a statement is made such as '. . . the transistor is being used as a switch . . .'. Transistors *can* be used as switches, as were thermionic valves in the world's first programmable computer *Colossus*, at Bletchley Park. However, technology has moved on from valves, through transistors to *logic gates*, combinations of electronic switches designed specifically to perform logic functions.

These act on voltage levels as their inputs and produce changes in voltage levels as their outputs. A positive voltage is called *logic 1*, and corresponds to a switch being ON in our previous descriptions; a zero voltage is called *logic 0*, and corresponds to a switch being OFF. The output from a logic gate (normally labelled Q) is also logic 1 or logic 0, corresponding to our light being ON or OFF, respectively, in our switch analogy.

Many logic devices operate from a stabilised 5 V supply, and this determines the *ideal* voltages corresponding to the two logic states:

logic 0 ≡ 0 V,

logic 1 ≡ 5 V.

The world isn't an ideal place, so the *real* voltage ranges used by the logic gates are:

logic 0 ≡ 0.0 to 0.4 V,

logic 1 ≡ 3.0 to 5.0 V.

The **AND** circuit of Figure 1 is now called an **AND gate**, and requires logic 1 inputs on A **AND** B to produce a logic 1 at the output. **Figure 3** shows this. The truth table is identical with that of Figure 1 – logic 1 replaces ON and logic 0 replaces OFF. Now compare Figure 2 with **Figure 4** – circuits and truth tables for the **OR** function. Again, we have exact similarity.

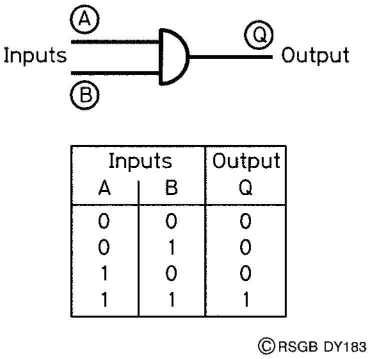

Inputs		Output
A	B	Q
0	0	0
0	1	0
1	0	0
1	1	1

©RSGB DY183

Figure 3 Electronic AND gate

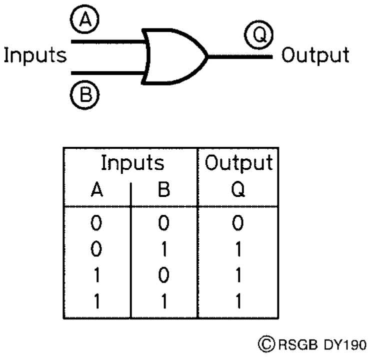

Inputs		Output
A	B	Q
0	0	0
0	1	1
1	0	1
1	1	1

©RSGB DY190

Figure 4 Electronic OR gate

There is another very common logic gate, which performs the **NOT** function. It is easy to understand. Just ask yourself the question 'What is **NOT** logic 0?', and the answer is obviously 'logic 1'. Similarly, logic 0 is **NOT** logic 1. A **NOT** gate simply changes the logic state of the input; it is also known (because of this behaviour) as an *inverter.* Its symbol and truth table can be found in **Fig 5**. Note the little circle on the output of the gate in Figure 5. In logic circuits, this symbol always implies inversion, or the presence of a **NOT** gate. Keep an eye open for it!

So far, the logic functions we have discussed have all been words which we use in everyday language, which has made the electronic interpretation of them relatively easy. Now we must introduce a function for which there is *no* analogy in normal speech – the **NAND** function. This means a

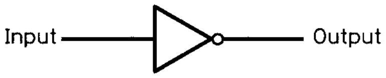

Input	Output
1	0
0	1

©RSGB DY185

Figure 5 Electronic NOT gate

combination of an **AND** gate and a **NOT** gate, and the sharp-eyed reader will have spotted the little circle added on to the normal **AND** gate symbol in **Figure 6**!

To make things easier to understand, the truth table in Figure 6 has four columns, not three as in previous tables. The third column is the standard **AND** output – compare it with the third column in Figure 3. That is the output from the **AND** gate *before* it encounters the little circle that inverts it, so the final output from the **NAND** gate is the output of the **AND** gate, inverted! The third and fourth columns are the inverse of each other.

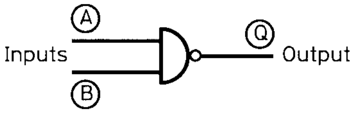

Inputs		and	Output
A	B		Q
0	0	0	1
0	1	0	1
1	0	0	1
1	1	1	0

©RSGB DY186

Figure 6 Electronic NAND gate

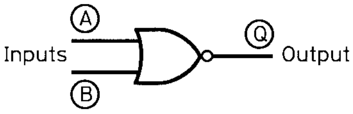

Inputs		or	Output
A	B		Q
0	0	0	1
0	1	1	0
1	0	1	0
1	1	1	0

©RSGB DY187

Figure 7 Electronic NOR gate

We can add the inverting operation to the output of an **OR** gate also, producing a **NOR** gate! The symbol and its truth table are shown in **Figure 7**. Again, notice that the final output is that of the ordinary **OR** gate, inverted!

These new functions of **NAND** and **NOR** are used more than the **AND** and **OR** functions because it makes other circuits easier to design using combinations of these gates.

A taste of Boolean algebra

The design of circuits using combinations of logic gates usually begins with a little mathematics, where the functions to be implemented are analysed. The mathematics used is surprisingly simple, and is a slightly changed version of ordinary algebra called *Boolean algebra*, which allows manipulation of logic functions to be made. Normal algebra has operations in it such as addition, subtraction and multiplication and division. The mathematician Boole found that the logical **AND** operation could be handled by the algebraic operation of *multiplication* (symbols × or •), and the **OR** operation by the algebraic operation of *addition* (symbol +). The **NOT** operation involved a new symbol, that of a bar over the input being inverted, such as $\overline{A}$.

So, our five basic logic operators can now be written in a mathematical form:

AND $Q = A \times B$

OR $Q = A + B$

NOT $Q = \overline{A}$

NAND $Q = \overline{A \times B}$

NOR $Q = \overline{A + B}$.

Using logic operations in a mathematical form enables the most complex logic to be designed, simplified and converted into circuit diagram form in a very efficient and rapid way. There are more logic operators than the five we have considered here but, in general, they can all be broken down into combinations of these gates alone!

60 A resistive SWR indicator

Introduction

When a transmitter produces some output power, we want to make sure that as much as possible of this power is radiated by the aerial. This often requires the use of an *aerial tuning unit* (ATU), which matches the aerial impedance to that of the transmitter.

How do you know when this matching has been achieved? The most usual way is to use a *standing-wave ratio* (SWR) indicator. If the impedance of the aerial does not match that of the transmitter output, some of your transmitter power (also known as the *forward* power) is reflected back along the aerial feeder and back into the transmitter, where it causes excess heating. The forward and reflected waves interact along the feeder to produce a wave whose position remains constant, and which is therefore called a *standing wave* or a *stationary wave*. An SWR meter simply indicates forward power and reflected power, and adjustments are made to your ATU until the reflected power is as small as possible (ideally zero, of course). If there is no reflected power then, by a process of elimination, all your forward power is reaching the aerial!

Sampling the RF

Whatever type of SWR indicator you use, it must use some sort of sampling circuit to pick up the forward and reflected waves. The project *A standing-wave indicator for HF*, elsewhere in this book, uses a toroidal transformer to separate the readings for the forward and reverse waves. This design differs in that it measures the voltages across resistors through which the RF current is passing. Its advantages are:

(a) it uses cheap parts – four resistors, two capacitors and a diode, together with a rotary switch, a surplus meter, a preset potentiometer and two sockets;
(b) within this SWR indicator, there is always a resistive path for the RF current from the transmitter, formed by R1, R2 and R3; this can prevent damage to simpler home-made transmitters, which may be damaged during adjustment of the ATU when using more conventional SWR indicators.

The only disadvantage of this form of indicator is that it must be switched out of circuit once the ATU has been adjusted for a particular band.

Construction

The SWR indicator is very simple to build, as most of the components can be mounted on the back of the 3-way rotary switch. This is shown in **Figure 1**. The switch is a 4-pole, 3-way rotary type, of which only two poles are used.

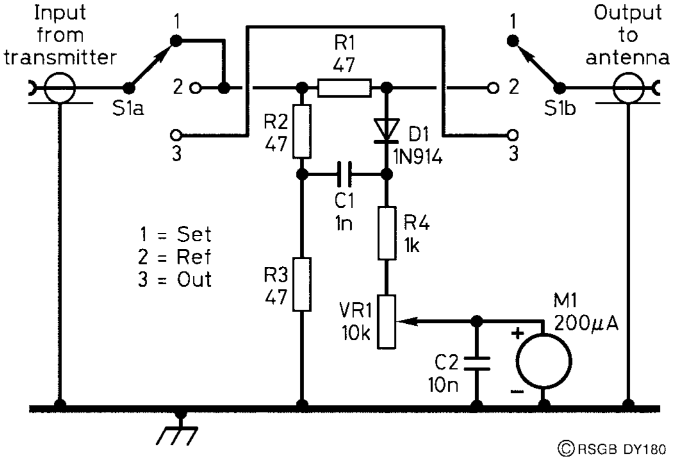

Figure 1 Resistive SWR meter, circuit diagram

Because the other switch contacts are not being used, they can be employed as support tags for other components. The ground wires are all soldered on to the metal frame of the switch. If your switch frame is of all-plastic construction, then a 12 SWG copper wire run around the switch will make a good earth connection to the metal case for the leads shown in Figure 1. The preset potentiometer used to control the sensitivity of the circuit can be mounted directly on the meter tag.

Resistors R1, R2 and R3 handle the RF power during the tuning-up process. If you have them, use 1 watt resistors; otherwise, you can use two 100 ohm half-watt resistors in parallel for each of R1, R2 and R3. The meter, M1, can be any DC type of sensitivity around 200 μA.

In use

First, find a clear frequency, and without the indicator in circuit, check that the frequency really *is* clear by asking and identifying yourself. If it is, connect the indicator between the transmitter and the ATU which, in turn, is connected to your aerial. Turn SW1 to the SET position and key the transmitter. Adjust VR1 until the meter reads full scale. Switch off the transmitter. Turn SW1 to the REF position and key the transmitter again. Adjust the ATU until the lowest reading is obtained on the meter. Switch off the transmitter. For the chosen frequency, you have adjusted your ATU for minimum reflected power and hence the lowest SWR. You will need to repeat the process when you change bands, and possibly when you change frequency within the same band. Switching SW1 to the OUT position, you are ready to transmit. You may have noticed that it is good practice to switch the transmitter off when operating SW1. Get into that habit!

Parts list

Resistors	
R1, R2, R3	47 ohm (Ω) 1 watt (or 2 × 100 Ω, ½ watt, see text)
R4	1 kilohm (kΩ) 0.25-watt
VR1	10 kilohms (kΩ) preset
Capacitors	
C1	1 nanofarad (nF)
C2	10 nanofarads (nF)
Semiconductors	
D1	1N914 or similar
Additional items	
SW1	4-pole 3-way rotary switch, of which only 2 poles are used
M1	200 μA DC
	Case

61 An audio filter for CW

Introduction

This is a simple passive circuit (it has no power supply) that adds some audio selectivity for Morse code reception and also includes a very simple noise limiter that gives a visual indication of when noise spikes are being removed!

The circuit

Figure 1 shows the complete circuit. The tuned circuit of C1 and L1 resonates very close to 800 Hz, so initially you will have to tune a signal in carefully until it sounds loudest – you will soon be able to do this without thinking. The two LEDs connected back to back across the signal path act as a noise limiter, reducing the amplitudes of static crashes and noise from car ignition systems, etc. The LEDs blink when they conduct – this is not necessary to the operation of the circuit, but adds a little colour to your listening! The noise limiter *does* make listening more comfortable, though.

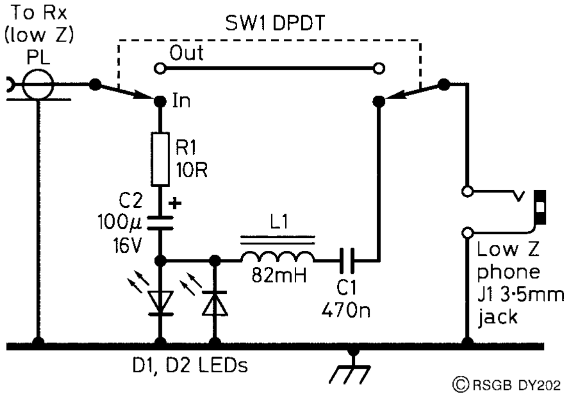

Figure 1 CW filter, circuit diagram

Construction

The circuit layout is shown in **Figure 2**. Point-to-point wiring is used, with a small tag-strip being the only item used for the extra support of C2 and R1. The LEDs and SW1 support the other components. An aluminium box (**Figure 3**) is used to make the circuit tidy and usable.

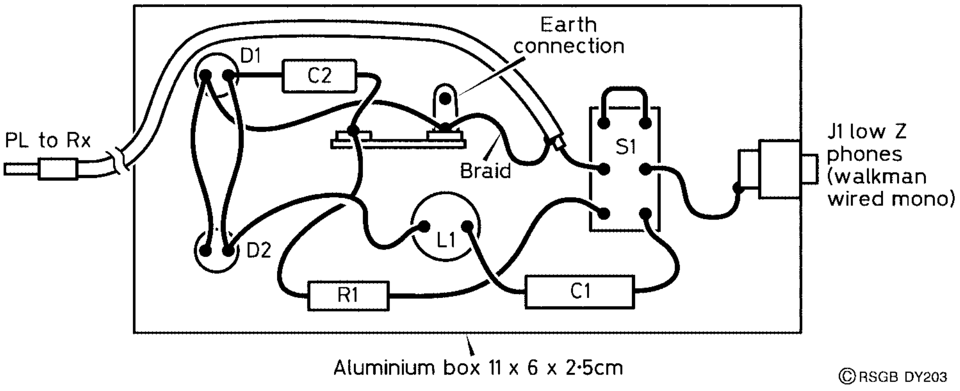
Earth connection
D1
C2
PL to Rx
S1
Braid
J1 low Z phones (walkman wired mono)
D2
L1
R1
C1
Aluminium box 11 x 6 x 2·5cm
©RSGB DY203

Figure 2 CW filter, component layout

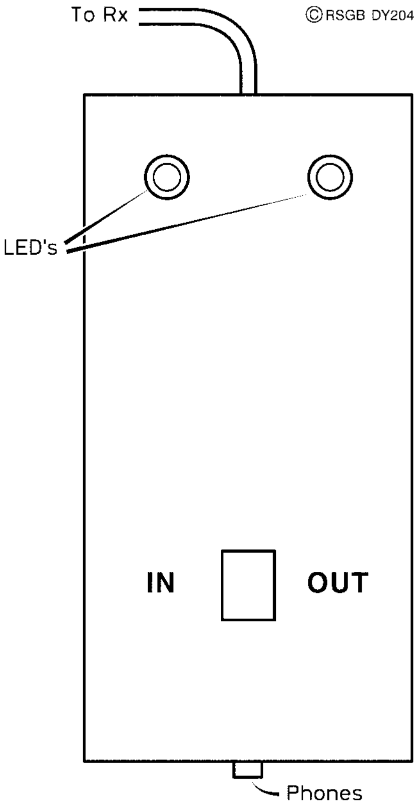
To Rx
©RSGB DY204
LED's
IN
OUT
Phones

Figure 3 CW filter, front panel

You must use low-impedance headphones for the circuit to perform properly. Plug the completed unit into the headphone socket of your receiver and adjust the volume so that the LEDs are just *not* blinking on normal audio. The unit should be switched out of circuit for speech listening; this is the purpose of the toggle switch, SW1.

Parts list

Resistor	
R1	10 ohms (Ω), 0.25 watt, 10% tolerance
Capacitors	
C1	470 nanofarads (nF)
C2	100 microfarads (μF) electrolytic, 16 V WKG
Inductor	
L1	82 millihenries (mH)
Semiconductors	
D1, D2	LEDs
Additional items	
SW1	DPDT
J1	3.5 mm jack socket
PL	Jack plug to suit receiver
	Aluminium box, approx. 11 by 6 by 2.5 cm

Source

Components are available from Maplin.

62 An electronic die

Introduction

Throwing a die is a venerable way of generating a 'random' number between 1 and 6. A die is easy to lose, so here is an electronic die-throwing circuit which brings the technique up to date and serves as another application of logic circuits. If you need a reminder of the basics of logic, refer to *Digital logic circuits*.

Warning

This circuit uses members of the integrated circuit family known as CMOS (complementary metal-oxide semiconductor). These use very little current and can be completed destroyed if they come into contact with the

magnitudes of static electricity that most of us carry about when we walk on carpets and wear rubber shoes. You will never know if this wanton destruction has happened – all you will discover is that your circuit doesn't work and that you have tested *everything*. To avoid this problem do the following things:

1. **Before you open the little packet in which the IC is supplied, touch something which you *know* to be earthed – the metalwork of any equipment which is mains-earthed, for example. Then open the packet.**
2. **Let the IC fall gently on the bench – don't pick it out with your fingers. Touch your earthed metalwork again. Pick up the IC and insert it gently into its holder.**

The circuit is safe from damage while it is connected to the battery.

Description

The circuit is shown in **Figure 1**. IC1a and IC1b form an *astable multivibrator*, similar to that used in 'An LED Flasher' also in this book. It runs constantly, and its output is fed to IC2 via a single NAND gate, IC1c. The other input of IC1c is held at 0 V by R3 and C3. Whenever one input of a NAND gate is logic 0 (or 0 V), there can be *no output* from the gate, irrespective of what is happening at other inputs. So, despite the fact that the oscillator (or *clock*) is running all the time, its square-wave output never

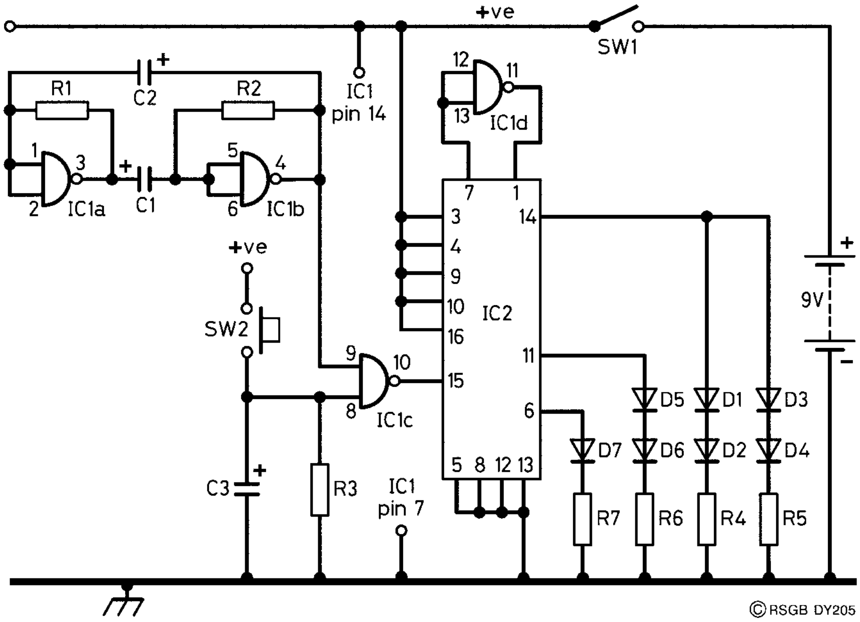
+ve
SW1
R1
C2
R2
IC1 pin 14
12
11
13
IC1d
1
3
IC1a
2
C1
5
4
6
IC1b
7
1
3
14
4
9
10
16
IC2
+ve
SW2
9
10
15
8
IC1c
11
6
D5
D1
D3
D7
D6
D2
D4
5 8 12 13
C3
R3
IC1 pin 7
R7
R6
R4
R5
9V
©RSGB DY205

Figure 1 Electronic dice, circuit diagram

reaches IC2 until the 'other' input of IC1c is brought away from logic 0. This happens when SW2 is pressed. Pressing SW2 momentarily, as is normally done, charges up C3 to 9 V, putting a logic 1 on the second input of IC1c and allowing the clock signal through to IC2. As the switch is released, C3 begins to discharge through R3, gradually lowering the voltage on pin 8 of IC1c. As this voltage crosses 4.5 V (half the supply voltage), IC1c then treats this input as logic 0, which cuts off the clock signal from IC2 again. Thus, when SW2 is momentarily pressed, the clock signal is fed to IC2 for a short interval of time, before being blocked again.

IC2 is a binary counter, and the outputs that concern us are from pins 14, 11 and 6. Pin 14 is the *most significant bit* and pin 6 the *least significant bit*. See the panel for an explanation of what happens here.

To understand these outputs, you should be able to count in binary, using three bits. At the start of the counting process, all the bits have zero values, i.e. 000. The left-most bit is called the *most significant bit*, and the right-most bit is called the *least significant bit*. (When we write numbers normally, a number 1 in the left-hand position represents one hundred, 1 in the middle column represents ten, while the number 1 in the right-hand position means one. One hundred is a more significant number than one, hence the nomenclature.)

The three bits of our binary number have values, from left to right, of 4 ($=2^2$), 2 ($=2^1$) and 1 ($=2^0$). This means that a number 1 in the left-hand column signifies the normal number four, in the middle column it would represent two, and in the right-hand column, it would represent one. This should help you understand the patterns of bits which emerge as the clock waveform is counted, as the next paragraph explains.

As each cycle of the clock enters IC2, it increments its internal counter and the values of that counter are shown by the states of pins 14, 11 and 6. After the first pulse, these three pins would have states corresponding to 001; after the second, 010; after the third, 011; after the fourth, 100; after the fifth, 101; after the sixth, 110. This sequence of 3-bit numbers represents a binary count from 1 to 6 in 'normal' parlance. These six states are used to illuminate the conventional pattern of dots on a die, using LEDs.

As the clock cycles are counted, the LEDs flicker as the die is 'rolled'. Resistors R4 to R7 are used to limit the current through the LEDs. When the counter stops, it can be in any of the positions shown in **Table 1**. Because of

Table 1 Dice number relative to counter output

Step	*Pins*			*Dice number*
	14	*11*	*6*	
1	0	0	1	4
2	0	1	0	2
3	0	1	1	6
4	1	0	0	1
5	1	0	1	5
6	1	1	0	3

the way in which the dots are grouped on the faces of a die, the wiring of the LEDs is simpler than it might otherwise be. You can see from Figure 1 that there are really only three sets of connections to the LEDs – one from each bit of the counter output. When pin 14 is at logic 1, four LEDs are lit, corresponding to the number four. For the number five, pins 14 and 6 will be at logic 1. For the number one, only pin 6 is at logic 1. For three, pins 11 and 6 are at logic 1, for six, pins 14 and 11 are at logic 1, and for two, only pin 11 is at logic 1. These conditions are summarised in Table 1.

Construction

Two pieces of Veroboard (of the strip type) are needed – one for the main circuit (**Figure 2**) and the other for the display LEDs (**Figure 3**). The display board is easier to build, so we will do that first. It measures 20 holes by 18 strips.

Cut the tracks as shown in Figure 3, using a 3 mm (1⁄inch) twist drill rotated between thumb and forefinger. Insert and solder the Veropins, resistors and the wire links. Then solder the LEDs in place, making sure their polarities are correct. The LEDs are mounted proud of the board (**Figure 4**, which will help when you come to fix the board into a case.

The main board measures 32 holes by 18 strips. As before, remove the tracks in the places shown in Figure 2. Note that track 'L' is *not* broken under the position for IC2. Solder in the IC holders and the links. Solder in R1 and R2 vertically, and R3 horizontally. Then solder in C1, C2 and C3; these all being electrolytics, check their polarities. Solder in the Veropins and prepare the connections to those components not on the main board, using stranded, insulated wire. Check the circuit carefully, and read the handling precautions given earlier before carefully inserting the two ICs.

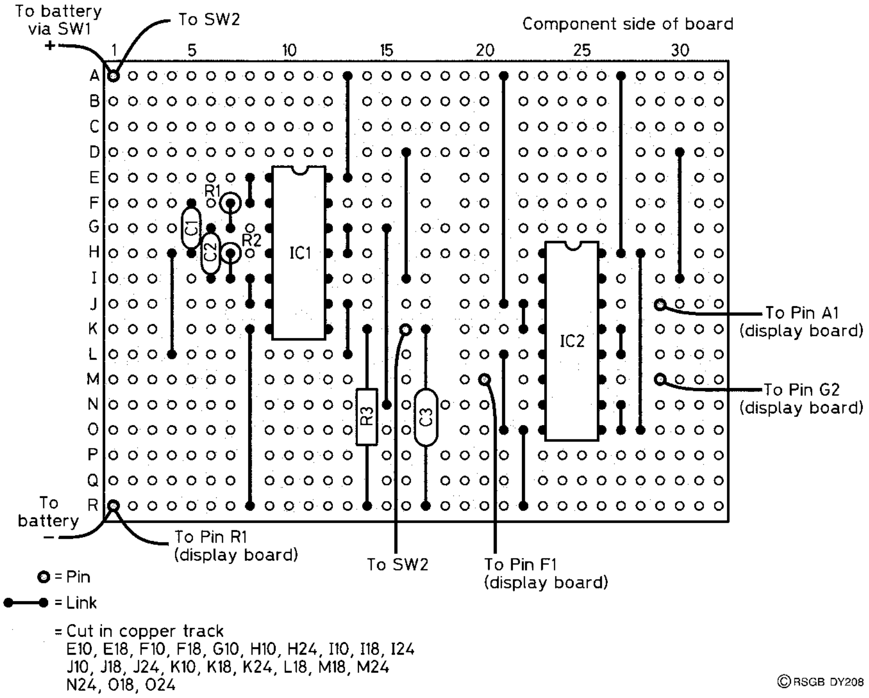
To battery via SW1
+
To SW2
Component side of board
1
5
10
15
20
25
30
A B C D E F G H I J K L M N O P Q R
R1
R2
C1
C2
IC1
IC2
R3
C3
To Pin A1 (display board)
To Pin G2 (display board)
To battery –
To Pin R1 (display board)
To SW2
To Pin F1 (display board)
= Pin
= Link
= Cut in copper track
E10, E18, F10, F18, G10, H10, H24, I10, I18, I24
J10, J18, J24, K10, K18, K24, L18, M18, M24
N24, O18, O24
©RSGB DY208

Figure 2 Main board, component layout

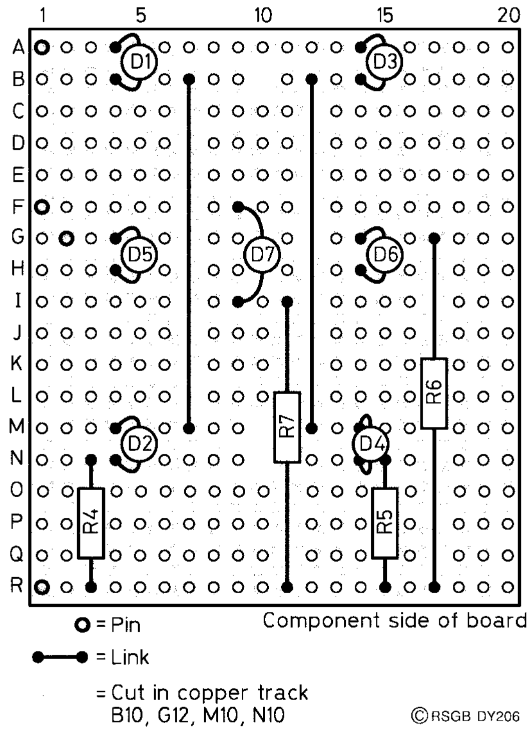
1
5
10
15
20
A B C D E F G H I J K L M N O P Q R
D1
D3
D5
D7
D6
D2
D4
R4
R5
R6
R7
= Pin
Component side of board
= Link
= Cut in copper track
B10, G12, M10, N10
©RSGB DY206

Figure 3 Board for the six LEDs

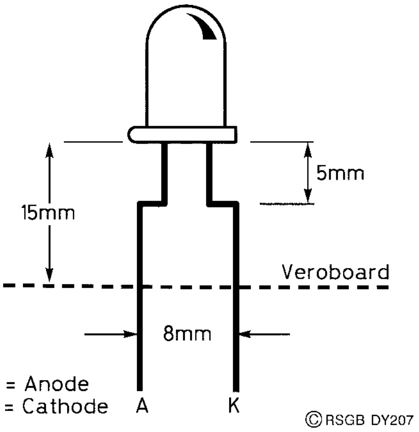

Figure 4 D7 lead bending

Testing and use

When the circuit is switched on, it is likely that the centre LED only will be lit. Press SW2, the 'throw' switch. All the LEDs should flicker, even after the switch is released, but only for a second or so. Then a standard pattern should show, signifying numbers between one and six. If this does not happen, switch off and check your circuit again. Are there any missing links? Are there any obvious dry joints (usually dull instead of shiny)? Can you see any solder bridges between tracks? Did you choose to ignore the handling precautions? If none of these results in a working circuit, try some fault-finding.

- Disconnect the lead to D7 on the main board (marked 'to pin F1' on Figure 2). Touch the lead on to the positive supply rail. D7 should light. Now transfer this lead on to pin 4 of IC1. If it flickers, then the oscillator is running (which it should).
- Transfer this lead to pin 10 of IC2. D7 should flicker when SW2 is pressed, but should be on permanently when it is released. This means that the correct signals are reaching IC2 from IC1.
- If one of the chains of LEDs (i.e. D1 and D2, D3 and D4, D5 and D6) does not light, you may have one *or both* LEDs the wrong way round.

When the die is working, you may care to experiment with the values of R1 and R2, but you should keep their values the same, i.e. R1 = R2. The larger the values of these, the more slowly the die will appear to 'roll'.

Finishing touches

On completion of the project, you will want to mount it in a smart case; any plastic box is suitable for this, with the display board mounted so that the

LEDs protrude through the top and are fixed to it using LED clips. The way in which you mount both boards to the case is entirely up to you.

Parts list

Resistors: all 0.25 watt, 10% tolerance, or better

R1, R2	5600 ohms (Ω)
R3	15 kilohms (kΩ)
R4–R7	470 ohms (Ω)

Capacitors

C1, C2	1 microfarad (μF) electrolytic, 16 V WKG
C3	68 microfarads (μF) electrolytic, 16 V WKG

Semiconductors

IC1	4011
IC2	4029
D1–D7	LEDs, any size and colour

Additional items

SW1	On-off switch SPST toggle
SW1	Push-to-make, non-latching
	Veroboard, 2 pieces, see text for sizes
	Veropins, 10
	Stranded insulated wire for general wiring
	Single-core insulated wire for links
	PP3 battery clip
	PP3 battery
	Plastic box to suit
	LED mounting clips (7)
	Means of mounting boards to box

Source

Components are available from Maplin.

63 The absorption wavemeter

Introduction

The purpose of the absorption wavemeter is to check that a transmitter is radiating within the correct waveband, and to detect any spurious *harmonic* emissions. A harmonic frequency is an integral multiple of the carrier frequency – e.g. if the carrier is at a frequency *f*, the harmonics are at frequencies 2*f*, 3*f*, 4*f* . . . and so on.

It can also be employed as a relative field-strength indicator, being used in experimentation with aerials. What it *cannot* do is to measure the transmitted frequency accurately; all it can do is to place the signal within a particular band, say the 7 MHz or the 18 MHz band (or their harmonics).

How it works

You may recognise the circuit of **Figure 1** as a crystal set with a meter replacing the headphones. It is simply the parallel combination of an inductance with a variable capacitance, which constitutes a resonant circuit with a variable frequency. A dial on the variable capacitor can be roughly calibrated with frequency. The fixed capacitor across the meter helps to improve the meter reading by making it read the peak value of the carrier wave rather than the average value.

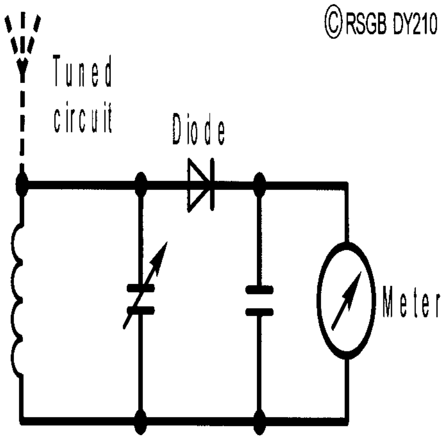

Figure 1 Absorption wavemeter circuit

When the wavemeter is situated in the vicinity of a relatively strong radio-frequency (RF) field, the inductor absorbs a small amount of energy from that field. If the circuit is resonant at the same frequency as that of the transmitter, an RF voltage is produced across the coil and is proportional to the strength of the RF field. The RF voltage is *rectified*, or *detected* (turned from AC into DC) by the diode and the second capacitor, and is displayed by the meter.

The tuned circuit

The formula for calculating the resonant frequency of a coil and capacitor is relatively simple. It is

$$f = \frac{1}{2\pi\sqrt{LC}},$$

where f is the resonant frequency (Hz),
L is the inductance (H), and
C is the capacitance (F).

If this formula puts you off, there is this slightly easier version:

$$f = \frac{159}{\sqrt{LC}},$$

where f is the frequency in **MHz**,
L is the inductance in **microhenries** (**μH**), and
C is the capacitance in **picofarads** (**pF**).

When a variable capacitor is specified, it is usual to quote its maximum capacitance, i.e. the value when the vanes are fully meshed. When the value of the capacitor is small, it is usual to quote its minimum capacitance also, the value when the vanes are fully open. If you use a variable capacitor of 10–100 pF with a 100 μH inductor, you will use the value of 10 pF to calculate the maximum resonant frequency, and the value of 100 pF to calculate the minimum resonant frequency, giving a tuning range of 1.6 MHz to 5 MHz. Check it yourself!

64 An HF absorption wavemeter

Introduction

An absorption wavemeter is basically a simple tunable detector circuit, such as would be used in a crystal set, but with the headphones replaced by a meter, in order to indicate the strength of the received signal.

The circuit

Figure 1 shows the circuit of the wavemeter. It is designed to cover all the HF bands from 1.8 MHz to 28 MHz in four switchable ranges. There is no built-in method of amplification, so a sensitive meter is needed in order to indicate sufficiently using the available absorbed energy. The meter does not need calibrating, it serves to produce only an indication of the absorbed power, *not* its absolute value.

An absorption wavemeter does not have outstanding selectivity, mainly because of the loading effect of the meter. This problem is ameliorated here (but only slightly) by the inclusion of R1 in series with the meter. It gives

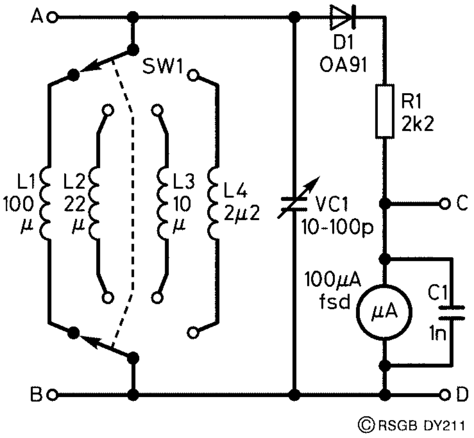

©RSGB DY211

Figure 1 HF absorption wavemeter, circuit diagram

improved selectivity at the expense of a slight loss of sensitivity. The appearance of capacitor C1 across the meter is discussed in *The Absorption Wavemeter*, in the previous section.

Coils

These can be purchased (see the parts list) or hand-wound on short lengths of PVC water pipe or conduit, and the coil-winding details are given in **Table 1**. In theory, the 100 pF tuning capacitor would cover all the HF bands with only three coils, omitting the 22 μH coil. Sample calculations are explained in the previous article. However, its inclusion avoids the common problem of having some bands cramped at the extreme edges of the tuning scales.

Table 1 Close-wound coil values

L (μH)	No of turns	Span (mm)	SWG
PVC FORMER OF OUTSIDE DIAMETER 25 mm			
2.2	9	9	20
10	24	24	20
22	44	44	20
100	110	60	26
PVC FORMER OF OUTSIDE DIAMETER 20 mm			
2.2	10	10	20
2.2	9	5	26
10	32	32	20
10	24	12	26
22	63	62	20
22	41	21	26
100	134	70	26
PVC FORMER OF OUTSIDE DIAMETER 19 mm			
2.2	11	11	20
2.2	9	5	26
10	35	34	20
10	25	13	26
22	68	67	20
22	44	22	26
100	157	79	26

Construction

Hand-wound coils, using 19–25 mm PVC formers, have single-layer windings, with their ends secured by threading the wires through two small holes drilled in the formers at each end of each winding.

You may have wondered about the use of the two-pole switch SW1, as all the lower ends of the coils are grounded anyway. The answer lies in the ease of construction. Each coil can be mounted on the tags of the switch for support. If you are using the commercial coils, they are small enough to be accommodated on a miniature rotary switch; a larger rotary switch is required for hand-wound coils. The specification of a 6-way switch allows for the addition of extra frequency ranges at little extra cost. Rotary switches such as the 2-pole, 6-way variety have an adjustable end-stop which allows the number of ways to be set anywhere between 2-ways and 6-ways. Remove the nuts from the switch, and a ‘washer’ will fall out. It is not a washer, really, and has a little piece of metal bent over at right angles to the washer. This fits into one of 12 holes in the body of the switch, and prevents the shaft from turning through 360° and selecting the number of ways.

Calibration

This process needs a commercial amateur-bands transceiver, if it is to be done quickly and accurately.

- Connect the transceiver to a dummy load, and connect a piece of flexible wire from point A in Figure 1 and wrap two or three turns around the cable between the transceiver and the dummy load.
- Turn the transceiver power level control to minimum, and set the frequency at 1.81 MHz and the mode to FM or AM. Switch to transmit.
- With the wavemeter set to its lowest frequency range (i.e. 1.8–3.5 MHz), rotate the tuning capacitor until maximum deflection is obtained from the meter needle. Write the frequency on the inner ring of the wavemeter scale at that point.
- Switch off and retune the transmitter to 3.5 MHz. Repeat the above process.
- Repeat, using switch range 2 for 3.5, 7 and 10 MHz, range 3 for 7, 10 and 14 MHz, and range 4 for 14, 18, 21, 24 and 28 MHz.

An alternative source of calibration signal could be an HF signal generator, with a single-turn loop of wire at the remote end of its cable, and the wire from the wavemeter brought close to the loop.

Remember that the marks on the scale represent frequency *bands*, not precise frequencies.

Extending the range

If you would like to experiment with increasing the range to the lower VHF band, then 50 MHz should be achievable with an additional self-supporting coil (no former). Try three turns of 20 SWG enamelled copper wire of 25 mm ID (internal diameter). Wind it on some 25 mm PVC pipe, as before, then slide out the pipe! The same result could be achieved with four turns of 19 or 20 mm ID, or 12 turns of 7.1 mm ID.

Parts list

Resistor	
R1	2200 ohms (Ω) 0.25 watt, 10% tolerance
Capacitors	
C1	1 nanofarad (nF) min. ceramic
VC1	100 picofarad (pF) variable
Inductors	
L1	100 microhenries (μH)
L2	22 microhenries (μH)
L3	10 microhenries (μH)
L4	2.2 microhenries (μH)
Semiconductors	
D1	Germanium OA91 or 1N4148
Additional items	
SW1	Rotary 2-pole, 6-way
	Moving-coil meter, 50 or 100 μA FSD (any meter of this sensitivity will do)
	Plastic box, approx. 150 by 80 by 50 mm
	Connecting wire, coloured

65 A vertical aerial for 70 cm

Introduction

If the range of your hand-held transceiver is very limited, what you need is a vertical aerial mounted outside. This design is a half-wave dipole fed at the end instead of the middle. Because the impedance of the dipole is high at its ends (and low in the centre), a matching circuit is needed so that this high impedance can be matched to that of the low-impedance coaxial cable. All that is needed is a coil, which increases the *electrical length* of the aerial to ⅝-wavelength. (Note that the *electrical* length (i.e. the length as it appears to an RF signal) is not necessarily the same as the *actual* length.)

Construction

The aerial element and the coil are made from a single piece of 1.5 mm welding (brazing) rod, and the dimensions are given in **Figure 1**. Wind the coil around a 4 mm rod or the shank of a twist drill. The lower end is filed to a point and then soldered into the centre conductor of a 4-hole panel-mounting BNC socket. (Try to obtain a good-quality BNC socket with PTFE insulation – the insulation of cheaper sockets is easily damaged.) Trim the element to 427 mm (top of element to top of coil) *after* the wire has been soldered to the socket.

The base coil causes the aerial to be rather 'whippy', so a piece of 5 mm plastic knitting needle can be cut to the length of the coil and then forced into it.

The radials are made from four lengths of 3 mm welding rod. These are bent and soldered into the four mounting holes of the socket, and then cut to the lengths shown in Figure 1.

Testing

Using a standing-wave-ratio (SWR) meter connected between your aerial and the transceiver, measure the SWR (or obtain some indication of the reflected power) on transmit. If it is greater than 1.5, switch off the transmitter, trim about 3 mm off the end of the element, and try again. Repeat the process until the measured SWR (or the reflected power) is as low as you can get it.

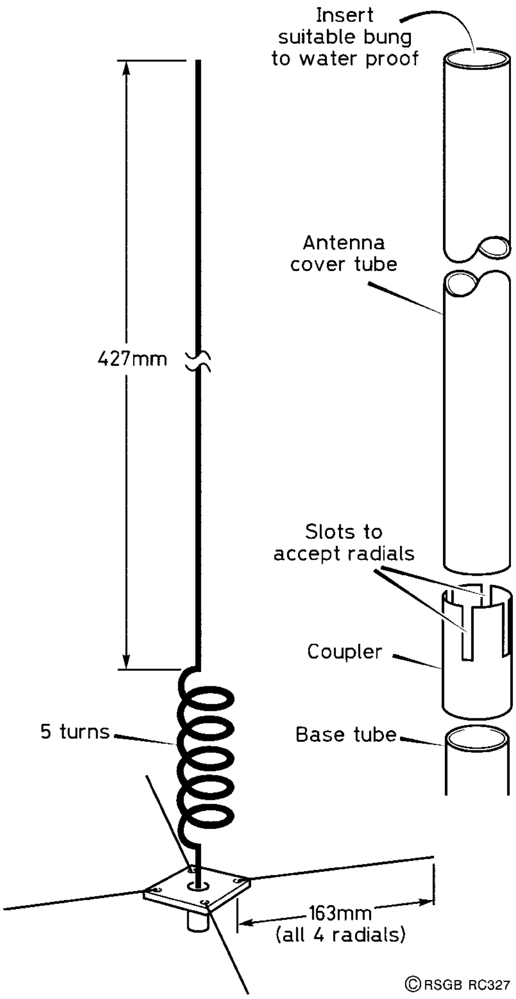
Insert
suitable bung
to water proof
Antenna
cover tube
427mm
Slots to
accept radials
Coupler
5 turns
Base tube
163mm
(all 4 radials)
©RSGB RC327

Figure 1 Construction of the 70 cm antenna

Installation

The aerial is made waterproof by enclosing it in a 22 mm diameter PVC waste water pipe. It is 'weldable', and available at plumbers' merchants, usually by the metre. A coupler is slotted to take the radials (Figure 1). File the BNC socket as required, so that it slides inside the coupler until the radials poke out of the slots. Cut a length of plastic tubing which is 30 mm longer than the aerial, and push it into the coupler. You will then need a plastic bung or screw-top to waterproof the top end.

Plastic welding solution is now applied to the joints, sealing the aerial inside the tube. The coupler has been weakened as a result of making the slots, so it is worthwhile applying PVC tape around this joint until the welding solution sets.

A support for the aerial is made from an off-cut of tube pushed into the lower end of the coupler and held with a self-tapping screw.

Parts list

1	4-hole panel-mounting BNC socket
1 m	3 mm brazing rod (may be available from a small garage)
1 m	1.5 mm brazing rod
1 m	22 mm PVC waste water pipe
1	22 mm straight coupler
	Plastic welding solution

66 A UHF corner reflector aerial

Introduction

The corner reflector is a well-known design and is capable of good performance on the VHF and UHF bands. At UHF, the practical implementation of the corner reflector is an ideal constructional project.

Some details

Expressed quite simply, the aerial consists of a $\frac{\lambda}{2}$ dipole (where λ is the standard symbol for wavelength, making a '$\frac{\lambda}{2}$ dipole' a half-wave dipole).

Nothing new in that, you might say. However, the interesting feature is the reflector, which is not the usual single element, but a 90° metal 'corner', acting rather like a parabolic dish as used for satellite signal reception. The wind resistance of this type of reflector makes it impractical so, to reduce the 'windage' quite significantly, we make the 'corner' from closely spaced rods, as illustrated in **Figure 1**.

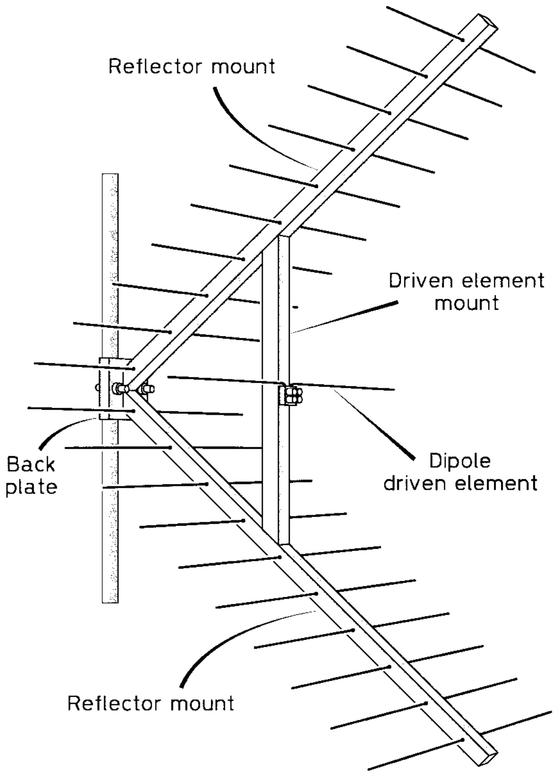

©RSGB DY215

Figure 1 70 cm corner reflector antenna

The reflector consists of a number of 0.6λ rods, spaced from each other by 0.1λ. The aerial frame can be made of metal or wood, but wood is easier to work with, and mounting the elements to the frame is simpler. The prototype was made with wood of 20 mm by 15 mm cross-section, as **Figure 2** shows. The wood was varnished for protection. The elements were made from 1.5 mm diameter copper wire, because a large reel of the wire happened to be available. The wire diameter is not critical; tubing could be used just as successfully. 14 SWG hard-drawn copper aerial wire would be even better than that used in the prototype.

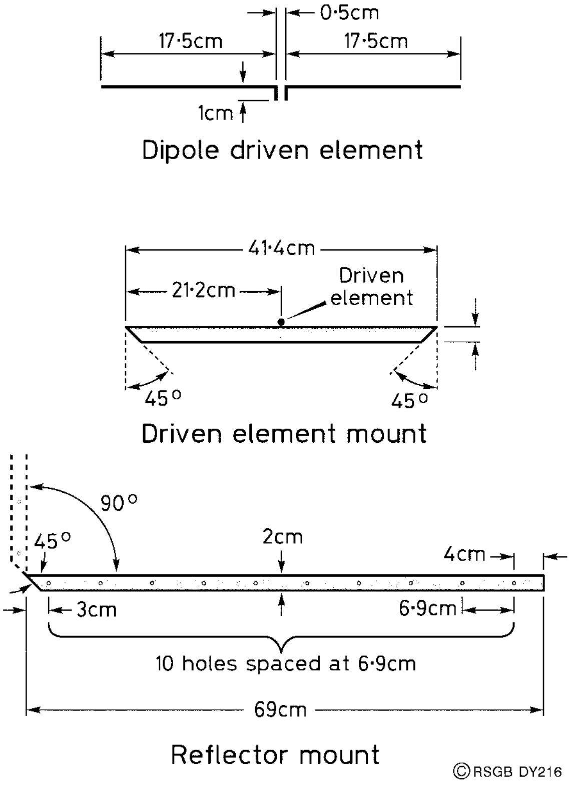
0·5cm
17·5cm
17·5cm
1cm
Dipole driven element
41·4cm
21·2cm
Driven element
45°
45°
Driven element mount
90°
45°
2cm
4cm
3cm
6·9cm
10 holes spaced at 6·9cm
69cm
Reflector mount
©RSGB DY216

Figure 2 Driven element dimensions, together with boom dimensions for driven element and reflectors

Construction

This project is just as much a woodworking project as a radio project! Follow the instructions carefully, and you should have little trouble.

- Cut the booms for the reflectors, as shown in Figure 2. A mitre block is invaluable here in producing the 45° corners.
- Using the dimensions given on the diagram, mark the hole positions for the reflector elements, and then drill holes of a size which holds the elements firmly.
- Cut the driven element boom according to the diagram, and mark the point midway along the longer side, which will assist you later in positioning the driven element.
- Cut the back plate to size (about 120 mm by 80 mm). You may need to alter this size depending on the size of the U-bolt you will be using to clamp the aerial to the mast.

- If you want to be extra cautious in your construction, use the belt-and-braces approach, commonly known as 'screw-and-glue' to fix the booms to each other and to the back plate.
- Fix the reflector booms to the back plate first, then slide in the driven-element boom until it will go no further, then apply the wood glue and screw the two ends tightly to the reflector booms. Leave for the period prescribed by the glue manufacturers for the glue to harden.
- Varnish the whole structure.
- Cut the driven element to the correct size plus a couple of centimetres (the reason for this will be evident in the *Testing* section), and fix it to the centre of its boom (at the position you marked earlier) with a 'chocolate block' connector to which the coaxial feeder cable will eventually be connected.
- Cut and fix the reflector elements in place. If you find that these are a loose fit in the holes then, *for each element*, drill a pilot hole through the boom to intersect the hole for the element. File off the point of a woodscrew, and screw it gently into the pilot hole until it meets the element and grips it in place. You will now see why the point was filed off! Alternatively, you can glue the elements in place.

Testing

Place the aerial on a mast, clear of obstructions. Connect it to a transceiver with a length of coaxial cable, with a standing-wave-ratio (SWR) meter in circuit. Find a clear frequency, identify your transmission and ask if the frequency *really is* clear. If so, key the transmitter again and note the SWR. **Do not stand in front of any aerial when it is radiating!** The length of the driven element must be adjusted to obtain an SWR of less than 2. If you have to shorten the dipole, bend the ends over rather than cut them off. That way, if you go too far, your can lengthen them again! The dipole was initially cut too long intentionally, to allow for adjustment here! Bending the ends over also reduces the risk of physical damage to clothing, skin and eyes. You may like to consider applying the same technique to the reflector elements for that reason alone.

Moving on . . .

Once you have warmed to the idea of the corner reflector as an aerial, you might like to ring the changes regarding the reflector. How about replacing the 20 reflector elements with a wire mesh, such as garden centres sell as 'chicken wire'? Choose the finest mesh if there is a choice. Some extra support may be needed around the edges of the mesh, but you could go on to make a comparison of aerial gain between the two types, using the *UHF Field Strength Meter* described elsewhere in this book.

Materials

Stiff wire or thin-walled tubing for dipole and reflector
Frame – wood, 15 mm by 20 mm cross-section, lengths given in text
Back plate – stout plywood, dimensions given in text
U-bolt to suit mast
Wood screws
Wood glue
50 W coaxial cable for feeder
2-terminal ‘chocolate block’ for dipole connection to feeder
Varnish

67 A switched dummy load

Introduction

A dummy load is a pure resistor of value 50 ohms which can replace your transmitting aerial and enable you to operate the transmitter for test purposes without radiating a signal. It sounds simple enough, but there are two main problems. Firstly, it is impossible to delve into your junk box and emerge with a resistor that will dissipate 100 W PEP and still retain its marked resistance value. Secondly, a ‘pure resistance’ is very difficult to achieve. A pure resistance is a device which has resistance but no reactance. All common resistors have significant reactance at radio frequencies, particularly the wire-wound varieties, which have a helical (i.e. wound like a coil) construction. This is particularly annoying, because wire-wound construction is normally used for large-wattage resistors.

Although all resistors have *some* reactance, not all are quite as bad as the wire-wound type. Carbon film resistors are made by depositing a thin film of carbon on the surface of a small, hollow ceramic cylinder, the thickness of the film of carbon determining the value of the resistor. Provided the lead lengths are kept short, these resistors have a tolerably small reactance, and will be used in this project.

Bearing the load

A 2 W carbon film resistor is hardly going to withstand our 100 W PEP of SSB, so it is obvious that the design of our dummy load must be a little more complex than a single resistor and a switch, despite what **Figure 1** might suggest! In fact, it uses 20 resistors, each of value 1000 Ω (1 kΩ). How does this solve our problem?

Perhaps a little theory is in order here, but no more than is required by the Radio Amateurs' Examination.

When two *equal* resistors of value *r* are combined in parallel (i.e. side by side), the total resistance, R_T, is given by:

$$\frac{1}{R_T} = \frac{1}{r} + \frac{1}{r}.$$

Adding $1/r$ to $1/r$ gives $2/r$, therefore:

$$\frac{1}{R_T} = \frac{2}{r}, \quad \text{i.e. } R_T = r/2.$$

So, by connecting *two* equal resistors in parallel, we get a combined resistance which is *half* the individual resistances. If we combine *three* in parallel, we get a *third* of the resistance, and so on.

Here, we are connecting *20* resistors of 1 kΩ in parallel, so we will produce an overall resistance of *one-twentieth* of the individual resistance, i.e. 1000/20 = 50 Ω, which is what we set out to achieve!

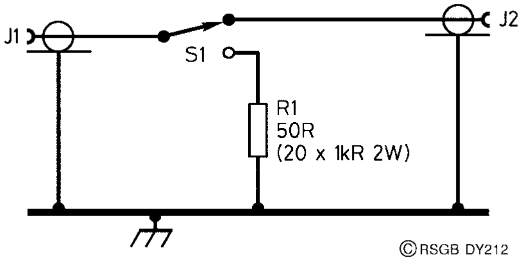

Figure 1 Basic circuit diagram

This is not the only advantage, though. Each resistor is capable of dissipating 2 W, so 20 of them will safely dissipate 40 W, for short periods at least. This power dissipation is approximately the same as 100 W PEP of normal speech, so the design should be capable of use in an 'average' amateur station. The power-handling ability of any dummy load can be improved by providing a 'heat sink' which helps to conduct the heat energy away from the resistors, thus lowering their temperature. One popular heat sink is a can of transformer oil, into which the resistors are immersed. This design uses a rather more mundane heat sink, but which is adequate for the job in hand.

Figure 2 PCB resistor supports

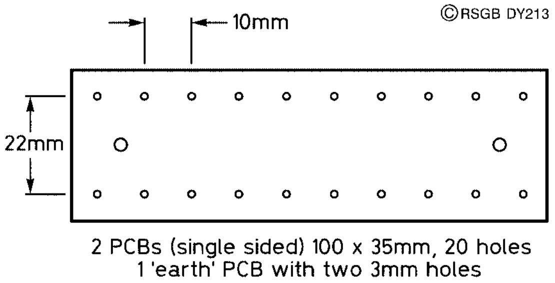
10mm
©RSGB DY213
22mm
2 PCBs (single sided) 100 x 35mm, 20 holes
1 'earth' PCB with two 3mm holes

Figure 3 Switched dummy load, component layout

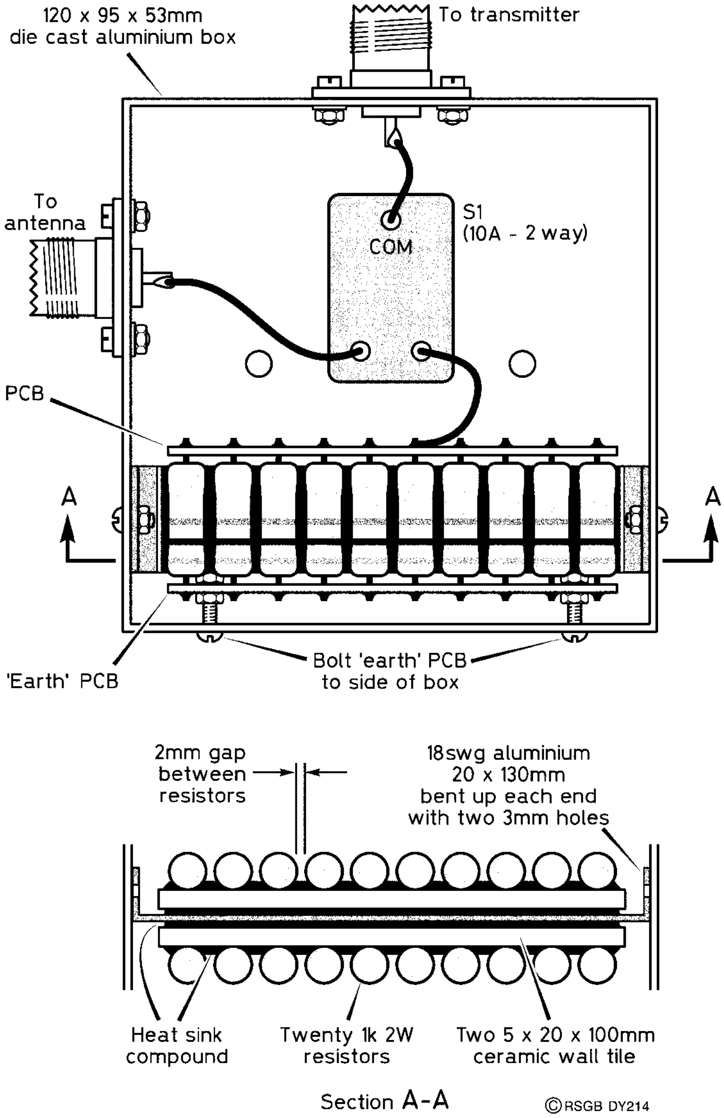
120 x 95 x 53mm
die cast aluminium box
To transmitter
To antenna
S1
(10A - 2 way)
COM
PCB
A
A
'Earth' PCB
Bolt 'earth' PCB to side of box
2mm gap between resistors
18swg aluminium 20 x 130mm bent up each end with two 3mm holes
Heat sink compound
Twenty 1k 2W resistors
Two 5 x 20 x 100mm ceramic wall tile
Section A-A
©RSGB DY214

Pouring the heat away

Two basic PCBs are needed, the details being given in **Figure 2**. Both measure 11 × 35 mm. Each has a set of 20 holes of diameter about 1 mm to take the ends of the resistors. Only one of them has two 3 mm holes which are used for mounting the completed load. **Figure 3** shows how the heat sink is assembled. Solder one end of each resistor into the 'earth' PCB, and then mount this to the box using the two bolts as shown in the diagram. Between each rank of resistors is a 'sandwich' consisting of an aluminium strip and two pieces of ceramic wall tile, to act as a heat sink for the resistors. Heat sink compound is used to provide good thermal contact between the resistors and the tiles, and between the tiles and the aluminium strip. This is shown in the lower part of Figure 3. Thread the loose wires of the resistors through the holes in the second PCB, solder into place, and crop the protruding wires.

Switching

A changeover switch must be used so that the transceiver can be switched between the dummy load and the aerial without unscrewing connectors. For most purposes, an ordinary 10 A 230 V changeover switch will suffice, as found in many electrical shops and DIY stores.

Wire this into the circuit as shown in Figure 1 and Figure 3. Check your wiring. Put S1 in the 'dummy load' position. If you have a multimeter which includes an ohmmeter, measure the resistance across the socket, J1, before connecting it to the transceiver. It should be very near 50 Ω. With S1 in the 'aerial' position, there should be an infinite resistance across J1. Move your ohmmeter to read the resistance across J2. It should be infinite for both positions of S1. If all seems correct, close the box and your dummy load is ready for use!

Parts list

Resistors
  1000 Ω (1 kΩ) 2 W carbon film, 20 required
Additional items
  10 A 230 V changeover (SPDT) switch
  SO–239 sockets, 2 required
  Aluminium box 120 × 95 × 53 mm or similar
  Ceramic wall tile cut as required
  Aluminium strip, 18 SWG
  Nuts and bolts as required

68 A simple Morse oscillator

Introduction

This is an excellent project which uses the 'junk box' as its source of components. If you have trouble in finding the bits for this one, a good source of the components for this and many other similar projects is to be found with the parts list at the end of the project.

The circuit

This is shown in **Figure 1**, and uses an 'unknown' Plessey chip, which makes the overall circuit extremely easy to build. A 0.1 μF capacitor is connected between pins 7 and 8, a speaker (in the popular 8 to 25 Ω impedance range) is connected between pins 8 and 9. If a 9 V battery is connected with its positive terminal to pin 8 and its negative terminal to pin 1, 3 or 5, a tone will be produced in the speaker.

To make this circuit into a good Morse practice oscillator, it is necessary only to insert a Morse key into the supply rail from the battery.

However, there is another refinement which you may care to build into the circuit. The tone from the loudspeaker is different, depending upon which of pins 1, 3 or 5 you use. In the prototype, a single-pole changeover switch was used to select the tones from pin 1 or pin 3, and the Morse key would be connected to the circuit via a small jack socket. There is no need for an on/off switch, as the Morse key performs that function. The switch and the jack socket can be seen in the photograph.

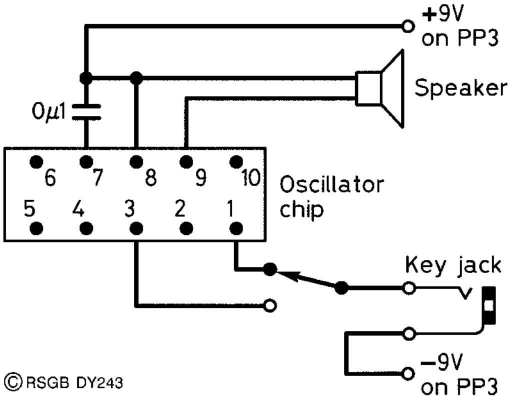

Figure 1 Morse oscillator, circuit diagram

The case

Almost any small loudspeaker will do; the higher the impedance the better. The 'impedance' is the figure usually printed on the rear of the speaker magnet. It is a value given in ohms (Ω). For simple circuits like this, you will usually find that the higher the impedance (within reason), the louder the sound it will produce. Speakers from old transistor radios will work, although their impedances can be rather low sometimes.

Any case big enough to house the components and the battery will be suitable. The prototype used a 'Walkman'-type speaker and case, and is shown in the photograph.

To use the circuit, simply attach a 3.5 mm jack plug to your Morse key, and insert it into the socket. Nothing should happen until you press the key, when a tone should be heard from the loudspeaker.

Another application

Try soldering the two wires of a 'twisted pair' to the 3.5 mm jack plug. Touching the wires together produces a tone from the speaker. This simple circuit can then be used as a 'continuity tester'. Touch the two wires to the ends of a fuse. If there is no sound, then the fuse is blown. There are many other such tests you could perform with this device – to check whether there really *is* a connection between one end of a wire and the other, for example.

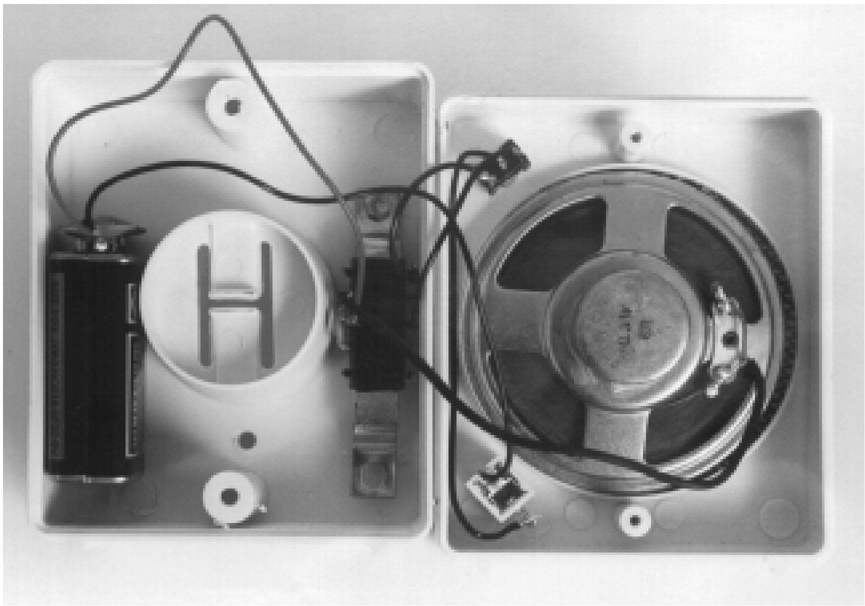

Warning

You must **never** make such tests on any equipment which is connected to the mains supply, even if it is switched off. If you want to make any such tests, make sure you are supervised by someone who understands what you are doing and is competent to advise and supervise you.

Parts list

Plessey oscillator chip
0.1 μF capacitor (disc ceramic)
Loudspeaker 8 to 25 Ω
3.5 mm jack socket (and 3.5 mm jack plug if needed)
Single-pole changeover switch

The Plessey oscillator chip is available from J. Birkett, 25 The Strait, Lincoln LN2 1JF, tel: 01522 520 767.

John Birkett may also provide a kit of parts (the chip, capacitor and loudspeaker).

69 A bipolar transistor tester

Introduction

This is a circuit which will test normal transistors, i.e. npn or pnp. It has the advantage of being able to test devices while they are still connected in their original circuits. However, when such tests are made, **the circuit containing the transistor under test must not be switched on**.

The circuit and how it works

The circuit runs from a 9 V battery such as a PP3 or six AA-type 1.5 V cells. Alkaline cells are to be preferred, as their electrolyte leakage properties are better. The circuit shown in **Figure 1** uses a single CMOS integrated circuit type 4001 or 4011. CMOS circuits require special handling precautions which are described in the project *Christmas Tree LEDs*, elsewhere in this book.

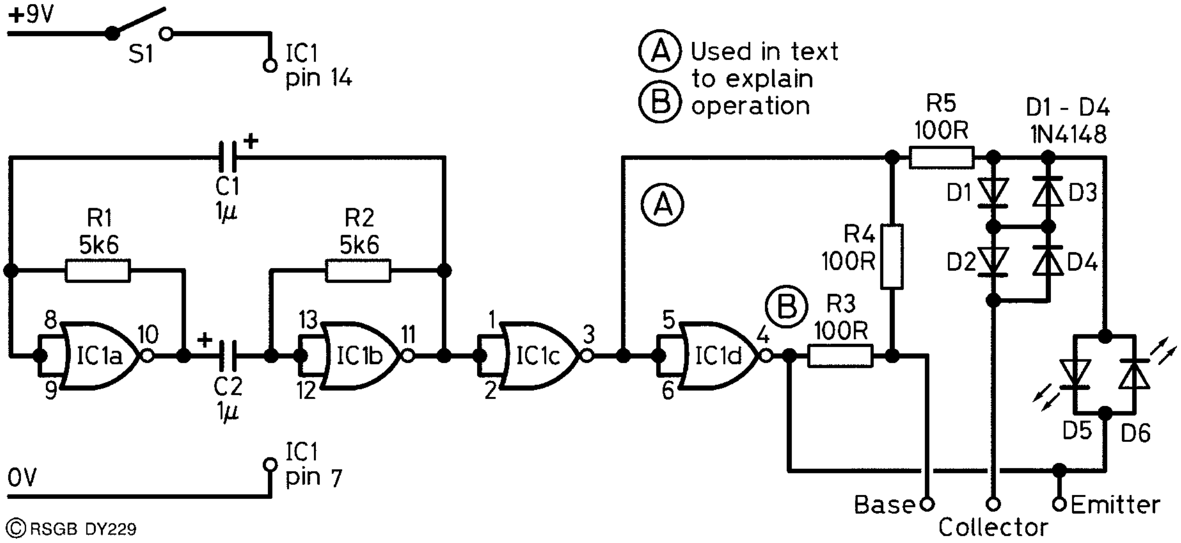
+9V
S1
IC1 pin 14
Ⓐ Used in text to explain Ⓑ operation
C1 1µ
R1 5k6
R2 5k6
8
9
10
IC1a
+
C2 1µ
13
12
11
IC1b
1
2
3
IC1c
Ⓐ
5
6
4
IC1d
Ⓑ
R3 100R
R4 100R
R5 100R
D1 - D4 1N4148
D1
D2
D3
D4
D5
D6
IC1 pin 7
0V
Base
Collector
Emitter
©RSGB DY229

Figure 1 Transistor tester, circuit diagram

Inside the IC are four logic gates (see *Digital Logic Circuits*) which are all connected as inverters, which means that the output signal is always the logical 'opposite' of the input. The first two gates are connected as an oscillator; the circuit being the same as that used in *An Electronic Die*.

The output of the oscillator, at pin 11, is connected to the input of a buffer stage, IC1c, which helps to isolate the oscillator from the circuit that follows it. The buffer output appears on pin 3, which we shall label as *test point A* for future use. Another inverter, IC1d, follows this, its output at pin 4 being labelled *test point B*.

There are two LEDs connected back to back at the circuit output. These are D5 and D6, D5 being red and D6 being green. Whatever the output of the oscillator at any instant, one of the LEDs must be lit and the other unlit. With point A positive and point B zero, the red LED is lit, when A is zero and B is positive, the green LED is lit. Because the oscillator output is repeatedly switching from one polarity to the other, the lit LED is alternately red and green. They switch between the two colours much faster than we can see, so what we think we see are both LEDs lit together.

The two 100 ohm resistors, R3 and R4, provide the bias to the base of the transistor under test. When A is positive and B is zero, the base-emitter junction of the transistor will be forward-biased, and the transistor will switch on (if it is a working npn type). When the transistor is on, it effectively short-circuits D5 (the red LED) and it extinguishes. When point A is zero and B is positive, an npn transistor will be switched off and the green LED (D6) will light. Thus, for a working npn transistor, only the green LED is lit.

If the transistor under test is a pnp type, it will switch on when A is zero and B positive, thus short-circuiting the red LED (D6). When A is positive and B zero, the transistor is off and the red LED (D5) is lit. Thus for a working pnp transistor, only the red LED is lit.

To summarise, the states of the LEDs indicate the following conditions:

- Both LEDs apparently lit: no transistor connected, or transistor permanently open circuit.
- Neither LED lit: a collector-emitter short-circuit is almost certain.
- Red LED alone: pnp transistor in working order.
- Green LED alone: npn transistor in working order.

Construction

The prototype tester was built on a piece of Veroboard measuring 15 strips by 18 holes. Cut the tracks using a track-cutting tool or a 3 mm (⅛ inch) twist drill, as shown in **Figure 2**. Notice that, in this diagram, there is no

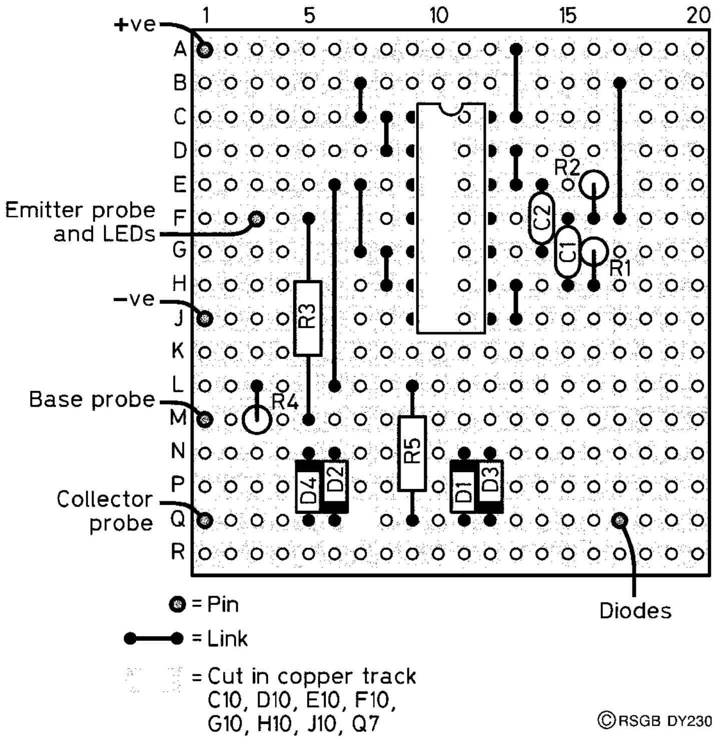
+ve
1
5
10
15
20
A
B
C
D
E
F
G
H
J
K
L
M
N
P
Q
R
Emitter probe and LEDs
−ve
Base probe
Collector probe
R2
C2
C1
R1
R3
R4
R5
D4
D2
D1
D3
Diodes
= Pin
= Link
= Cut in copper track
C10, D10, E10, F10,
G10, H10, J10, Q7
©RSGB DY230

Figure 2 Transistor tester, component layout

track ‘I’, so try to avoid miscounting when you translate diagram positions to real positions on the board. Hold the board up to a strong light to ensure that there is no copper swarf shorting adjacent tracks together, and that you have made the cuts in the correct places.

Having done this, insert and solder Veropins for all the connections to be made to components not on the board itself. Then solder in the IC holder and the wire links. Then insert the other components in the order resistors, capacitors and diodes D1–D4. Some resistors are mounted vertically so that their connections are on adjacent tracks. Double check the diode polarities – it is easy to make a mistake when wiring diodes in anti-parallel! Next, connect up the off–board components, again making sure that the LEDs have the correct polarity. The probe leads for the emitter, base and collector should be made from different colours of wire and terminated in probe clips (small insulated crocodile clips).

Check carefully for dry joints and errant blobs of solder. Plug the IC into its holder, ensuring that it is inserted the right way round, as shown in Figure 2.

Testing

Without a transistor in circuit, and the battery connected, both LEDs should be lit. Connect a known good npn transistor and verify that the green LED lights. Now simulate two transistor faults: disconnect the base lead and both LEDs should light; remove the transistor and connect the emitter and collector leads together. Neither LED should light.

Repeat the tests with a known good pnp transistor. The results should be the same, except that the correct indication should now be a lit red LED. On your computer, make a small label of the bulleted list above, to fit on your tester showing the states of the LEDs and what they mean. It will act as a useful *aide mémoire* when you use the tester in future.

Using

The circuit will test transistors in isolation or in an existing circuit, i.e. prior to use. You can check the lead identifications in component catalogues such as the Maplin catalogue. The tester is ideal for going through the large bags of unmarked transistors that you can buy for a song at rallies. You can sort them into three piles – npn, pnp and dud!

Parts list

Resistors: all 0.25 watt, 10% tolerance or better

R1, R2	5600 ohms (Ω)
R3, R4, R5	100 ohms (Ω)

Capacitors

C1, C2	1 microfarad (μF) electrolytic, 16 V WKG

Semiconductors

D1–D4	1N4148 general-purpose silicon diodes
D5	Red LED
D6	Green LED

Integrated circuit

IC1	CMOS 4001 or 4011

Additional items
  Veroboard, 15 strips by 18 holes
  Veropins
  PP3 battery and connector (or 6 × AA cells in PP3 clip holder)
  Switch, SPST
  Connecting wire
  One each of known working npn and pnp transistors for test purposes.

Source

Components are available from Maplin.

70 The ‘Yearling’ 20 m receiver

Introduction

Published to celebrate the first anniversary of *D-i-Y Radio*, this excellent receiver design forms a suitable ‘second receiver project’ for those who have successfully completed the MW receivers earlier in this series. The receiver is powered from a PP3 battery or from a mains adaptor, and can be built with the help of an experienced constructor, on a prototype board. The circuit diagram and some of the components used are shown in the separate diagram. Headphones or a loudspeaker can be used and, once the radio is completed, a few simple adjustments will make the *Yearling* spring to life!

Building the receiver

Before starting the constructional process, start by identifying all the parts. One by one, tick them off against the parts list. Are their values correct? The varactor diode is a twin type (see circuit diagram overleaf), and must be cut *carefully* down the middle, producing two devices, D1 and D2, with two wires each.

First, solder the IC sockets, followed by the coils (inductors); L1 is pink inside the top, and L2 is yellow inside. Then, solder in the varactors; the lettering on D1 should be next to coil L1, and the lettering on D2 should face resistor R7. After those, the capacitors, wire links and resistors should be soldered to the board. Take care to wire the voltage regulator, IC3, correctly. Solder in the crystal X1 as quickly and deftly as you can – crystals do *not* take kindly to having their leads bent and being fried with a soldering iron! Make sure that the electrolytic capacitors C2, C12, C15 and C16 are fitted the right way round. Most electrolytic capacitors have only the negative lead marked.

Figure 1 shows the rear of the front panel, illustrating the connections from the board and antenna socket to the controls. All normal connecting wires are 22 SWG or thereabouts, with insulation. Their lengths should be about 15 cm, except for the battery lead to the switch, which is about 8 cm. It is recommended that you use different-coloured wire for each connection to a control. **Figure 2** shows this. The variable resistor section of VR5, the AF gain control, uses single screened cable connected to 0 V (ground) at the

Amateur Radio

D-i-Y

Year

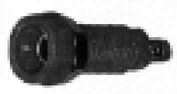
Antenna Socket

Variable Inductor

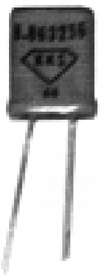
Crystal

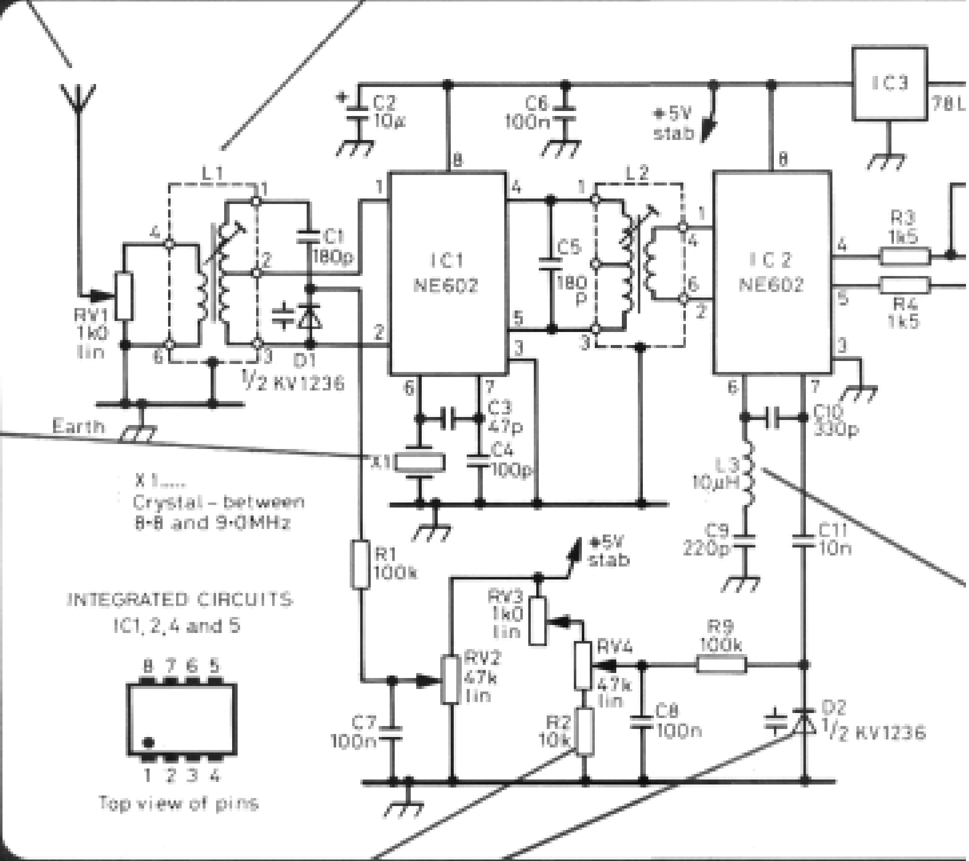
L1
RV1
1k0
lin
C1
180p
D1
1/2 KV1236
Earth
C2
10µ
C6
100n
+5V
stab
IC1
NE602
C5
180
p
L2
IC3
78L
IC 2
NE602
R3
1k5
R4
1k5
C3
47p
C4
100p
X1
C10
330p
L3
10µH
C9
220p
C11
10n
X1.....
Crystal - between
8·8 and 9·0MHz
R1
100k
+5V
stab
RV3
1k0
lin
RV4
47k
lin
R9
100k
RV2
47k
lin
C7
100n
R2
10k
C8
100n
D2
1/2 KV1236
INTEGRATED CIRCUITS
IC1, 2, 4 and 5
8 7 6 5
1 2 3 4
Top view of pins

Resistor

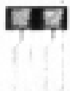
Varicap Diode

Most radio amateurs enjoy building part
is Do-it-Yourself (D-i-Y) Radio, and it is
successfully built yourself. License
allowed (by the Government) to use ho
can see how radio components are dra
circuit diagram for the

D-i-Y Radio, RSGB, Lambda House, Cra

– Do It Yourself

RADIO

rling

Polyester Capacitor

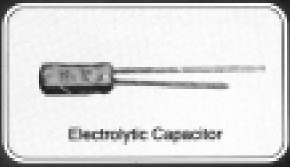

Electrolytic Capacitor

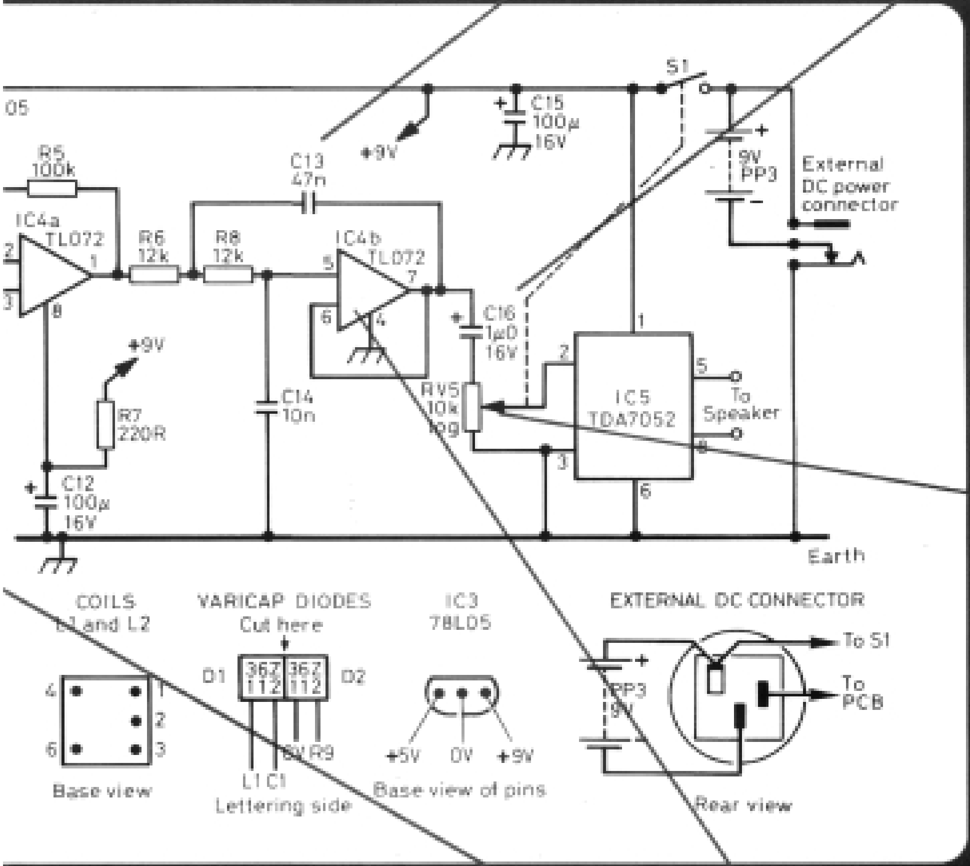

05
R5
100k
IC4a
TL072
2
3
1
8
R6
12k
R8
12k
C13
47n
+9V
C15
100μ
16V
S1
+
9V
PP3
–
External
DC power
connector
IC4b
TL072
5
6
7
4
C16
1μ0
16V
+9V
R7
220R
C14
10n
RV5
10k
log
2
1
5
IC5
TDA7052
To
Speaker
3
6
8
C12
100μ
16V
Earth
COILS
L1 and L2
4
1
2
6
3
Base view
VARICAP DIODES
Cut here
D1
367
112
367
112
D2
0V
R9
L1
C1
Lettering side
IC3
78L05
+5V
0V
+9V
Base view of pins
EXTERNAL DC CONNECTOR
+
PP3
9V
To S1
To
PCB
Rear view

AF Gain Control

of their radio stations themselves, this
very satisfying to use a radio you have
radio amateurs are the only people
me-built transmitters. On this page you
wn on a circuit diagram. This is the full
D-i-Y Radio Yearling.

Fixed Inductor

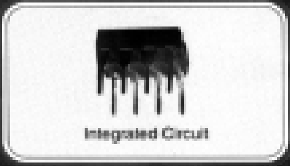

Integrated Circuit

nborne Road, Potters Bar, Herts. EN6 3JE

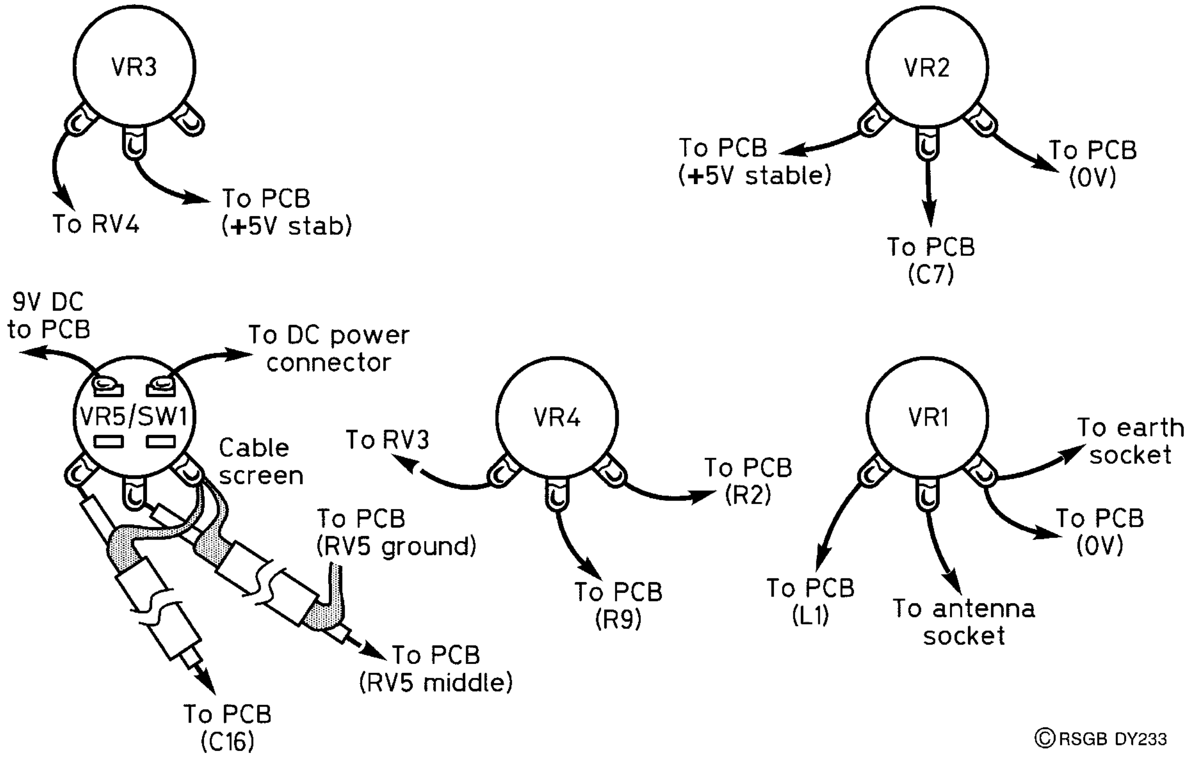
VR3
To RV4
To PCB
(+5V stab)
VR2
To PCB
(+5V stable)
To PCB
(0V)
To PCB
(C7)
9V DC
to PCB
To DC power
connector
VR5/SW1
Cable
screen
To PCB
(RV5 ground)
To PCB
(RV5 middle)
To PCB
(C16)
VR4
To RV3
To PCB
(R2)
To PCB
(R9)
VR1
To earth
socket
To PCB
(0V)
To PCB
(L1)
To antenna
socket
©RSGB DY233

Figure 1 Rear view of the variable resistors. Check the connections carefully to make sure the wires fit the correct holes

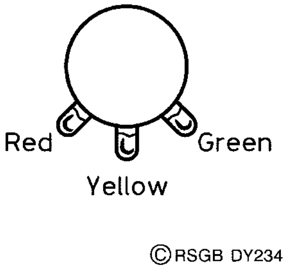
Red
Green
Yellow
©RSGB DY234

Figure 2 It is helpful to use different colour insulated connecting wires. Wires between each variable resistor and the board should be twisted together to give a neat wiring outfit.

board end. Now fit the ICs into their sockets. Make sure that they are the right way round (see circuit diagram) and that each pin lies directly above its corresponding socket before applying *gentle* pressure with the back of the board firmly supported.

Check that all the connections are correct (don't *assume* this – check the ends of each wire against the circuit diagram) and that all your soldered joints are shiny. Lastly, drill the five 10.5 mm diameter holes for the main controls. On the side of the case, a 6.3 mm diameter hole is needed for the speaker socket, and 8 mm holes for the antenna and earth sockets. The external power socket requires an 11 mm diameter hole.

Adjusting it

Before you can make adjustments, and in order to hear *anything* on your *Yearling*, you will need to connect an antenna (aerial) to the antenna socket. About 8 metres of wire, preferably outdoors and as high as possible, is all you need to connect to the socket. Connect the 3.5 mm jack socket for the speaker, and a 3.5 mm jack to your speaker leads. Connect a battery and switch on.

1. Using a very small screwdriver or, better still, a non-metallic 'trimming tool', *gently* screw in the core of L2 as far as it will go, but *don't force it*. Then unscrew it by three turns anticlockwise.
2. Set VR1, VR2 and VR4 to mid-position and rotate the core of L1 anticlockwise until the hissing noise you hear reaches a maximum intensity. Then adjust L2 for maximum noise.
3. If you now tune carefully with the main tuning control, VR4, you should hear some amateur Single-Sideband (SSB) speech signals. You may have to adjust the bandspread (fine tuning) control, VR3, to make the speech sound normal.
4. Having verified that everything is working, switch off and mount the controls on the front panel and the sockets on the side. To do this, it is much safer to disconnect all the controls and sockets, mount them in their final positions, and then wire them up again.
5. Fit the front panel knobs, connect your aerial and switch on again, checking that everything is working. Then, locate the cluster of CW (Morse) signals to be found at the bottom of the 20 metre band, set the bandspread control to mid-position and slacken off the main tuning knob. Turn the knob (but *not* the control!) until the pointer lies a little clockwise of 14.0 MHz. The SSB signals should now lie roughly between the dial centre and 14.35 MHz. Tighten up the knob.
6. Finally, fix the board to the rear panel, and secure the battery (if you are using one). Attach the rear panel to the back of the box, and you are finished!

Listening!

Make a habit of keeping the bandspread control in its centre position when searching for stations; then you can adjust it either way to make the signals readable. Unless you have a very big aerial, it is best to have the 'RF Gain' control, VR1, at maximum. Use the 'Antenna Tune' control, VR2, to give the best signal, and control the volume with the 'OFF/AF Gain' control. You will find some excellent DX stations with your *Yearling* receiver, and it will serve you well.

Parts list

Resistors (all 0.25 W, 5%)

R1, R5, R9	100 kilohms (kΩ)
R2	10 kilohms (kΩ)
R3, R4	1.5 kilohms (kΩ)
R6, R8	12 kilohms (kΩ)
R7	200 ohms (Ω)

Capacitors (all rated at 16 V or more, tolerance *at least* what is quoted)

C1, C5	180 picofarads (pF) polystyrene 5%
C2	10 microfarads (μF) electrolytic
C3	47 picofarads (pF) polystyrene 5%
C4	100 picofarads (pF) polystyrene 5%
C6, C7, C8	100 nanofarads (nF) or 0.1 microfarad (μF) ceramic
C9	220 picofarads (pF) polystyrene 2%
C10	330 picofarads (pF) polystyrene 2%
C11, C14	10 nanofarads (nF) or 0.01 microfarad (μF) ceramic
C12, C15	100 microfarads (μF) electrolytic
C13	47 nanofarads (nF) or 0.047 microfarad (μF) polyester, 5%
C16	1 microfarad (μF) electrolytic

Variable resistors

VR1, VR3	1 kilohm (kΩ) linear
VR2, VR4	47 kilohms (kΩ) linear
VR5	10 kilohms (kΩ) log with switch

Inductors

L1	Toko KANK3335R
L2	Toko KANK3334R
L3	10 microhenries (μH), 5%

Semiconductors	
IC1, IC2	Philips/Signetics NE602 or NE602A
IC3	78L05 5 V 100 mA regulator
IC4	TL072 Dual Op-Amp
IC5	Philips TDA7052 audio amplifier
Additional items	
D1, D2	Varactor diode Toko KV1236 (cut into two sections – see text)
X1	Crystal 8.86 MHz type (from Maplin, etc.)
4 off	8-pin DIL sockets for IC1, IC2, IC4 and IC5
2 off	4 mm sockets aerial (red) and earth (black)
1 off	3.5 mm chassis-mounting speaker jack socket
1 off	DC power socket for external supply (if required)
4 off	Red knobs with pointers
1 off	Tuning knob with pointer (e.g. 37 mm PK3 type)
1 off	Printed-circuit board or prototype board
1 off	Plastic case approx 170 × 110 × 6 mm (e.g. Tandy number 270-224)
1 off	Speaker 8–32 Ω impedance (or headphones)

71 Adding the 80 metre band to the Yearling receiver

Background

You will have noticed that your *Yearling* receiver has a dial which shows coverage of the 80-metre amateur band (3.5–3.8 MHz). This band is used for local contacts during the day, and contacts up to about 1600 miles in darkness. Longer distances are possible, particularly in the middle of winter.

The modifications

Only a few extra parts are required, as you may have noticed from the parts list. A low-pass filter, FL1 (one which passes low frequencies and rejects higher frequencies), is switched into the circuit on 80 m. The circuit of the

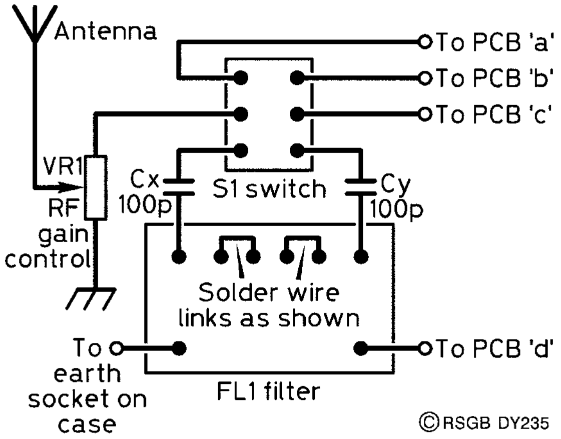
Antenna
To PCB 'a'
To PCB 'b'
To PCB 'c'
VR1
RF
gain
control
Cx
100p
S1 switch
Cy
100p
Solder wire
links as shown
To
earth
socket on
case
To PCB 'd'
FL1 filter
©RSGB DY235

Figure 1 The circuit diagram shows the extra components for 80 m operation. Note the connections to the PCB 'a', 'b', 'c' and 'd'

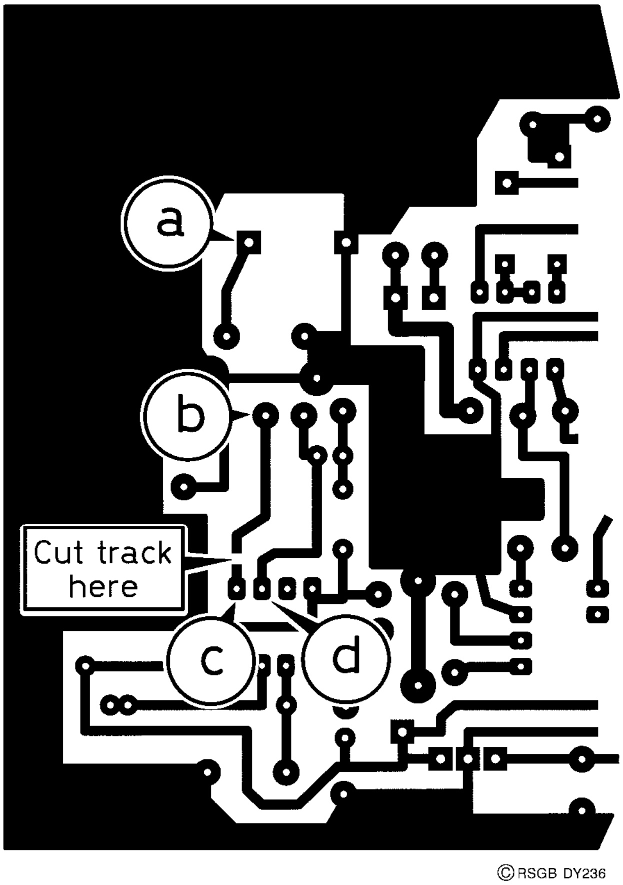
a
b
Cut track
here
c
d
©RSGB DY236

Figure 2 The underside of the PCB. Wires are connected from the switch and filter as shown

switch and its connections is shown in **Figure 1**. Before making the modifications, dismantle the receiver so that you have easy (and safe) access to the case and the track side of the PCB.

1. The first thing to do is to drill a 6.5 mm diameter hole in the side of the case into which the switch fits.
2. Then, using a sharp Stanley knife or scalpel, carefully cut the track on the PCB as shown in **Figure 2**, making a gap about 1 mm wide.
3. Using 10 cm lengths of different-coloured insulated wire, make the four connections, a, b, c and d, to the PCB, as shown in Figure 2.
4. Solder the two links on the filter, and then make the connections to the switch, capacitors and PCB. You will have to disconnect the existing wire between the RF gain control, VR1, and pin 4 of L1 on the PCB.
5. Lastly, check your new connections carefully, then mount the switch in the new hole and fix the filter to the bottom of the case with a little glue, as shown in **Figure 3**. Reassemble the circuit, and replace the back of the case.

More testing!

Firstly, switch your new switch, S1, into the 20 m position, to check that the original circuit still works! If you find that the ‘Antenna Tune’ control peaks at a slightly different position, don’t worry.

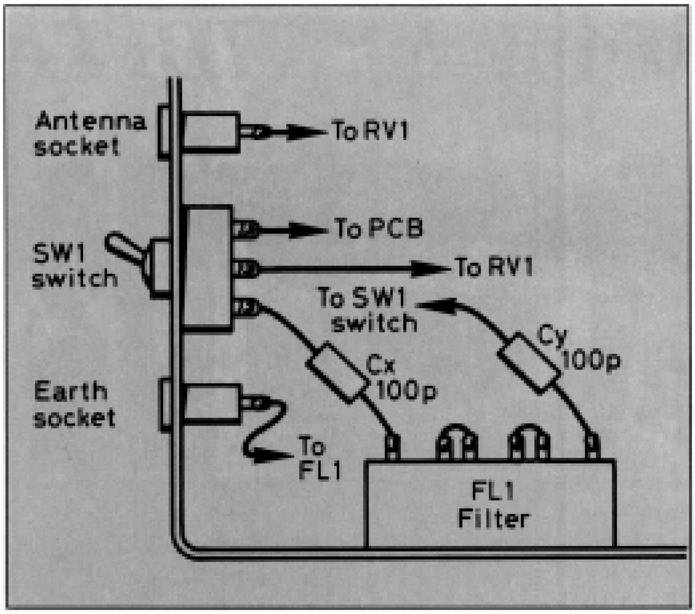

Figure 3 Internal view of the Yearling case. The filter FL1 is attached to the base with glue

Now switch to 80 m, and tune around the anticlockwise end of the dial; you should hear some SSB stations, particularly in the evenings and at weekends, when may people are on the air. At the other end of the travel of the tuning control, you should hear CW (Morse) stations.

The Radio Society of Great Britain broadcasts amateur radio news every Sunday morning on or about 3.65 MHz; the table below has the details. Finally, a good antenna is more important than ever for 80 metre reception – aim for more height and length, and then consider the project concerned with making an Antenna Tuning Unit (ATU)!

Parts list

Capacitors (all rated at 16 V or more, tolerance 10% or better)

Cx, Cy	100 picofarads (pF) polystyrene

Filter

FL1	Toko 237LVS1110 low-pass filter

Additional items

SW1	2-pole 2-way (changeover) toggle switch
7 off	Short lengths of insulated wire of different colours

The RSGB news broadcasts, GB2RS – Sunday mornings

Frequency (MHz)	*Local time*	*Reception area*
3.650	0900	SE England
3.650	0930	Midlands
3.650	1000	SW England
3.650	1100	Yorkshire
3.640	1130	Aberdeen
3.660	1130	Glasgow

The Midlands transmission is repeated at 1800 (6pm) local time on 3.650 MHz. All frequencies are approximate in order to avoid interference, and use lower sideband (LSB). If you also have a 40 metre receiver, there are GB2RS news broadcasts on 7.048 MHz at 0900 local time from Northern Ireland and from 1100 local time from the north midlands.

72 How the Yearling works

Introduction

The *Yearling* was designed to provide an introduction to *Amateur Radio* on the 20 m amateur band. Let's look at how the different sections (or 'stages') of a radio work, and how they fit together to form a complete receiver. **Figure 1** shows a block diagram which you can follow and compare with the circuit diagram of your *Yearling* receiver.

The antenna (or aerial)

Connected to your receiver, it will pick up not only amateur signals, but all other signals as well! This means that the receiver has to select the one signal that interests you, while rejecting all the others. The following stages do just that.

The RF filter

This stage (centred around L1) selects the band of radio frequencies (RF) containing the signal you want, in this case, those having wavelengths around 20 m. Signals from the 40 m band, for example, would not get through.

The crystal oscillator

This is an oscillator circuit designed around a quartz crystal (X1), and has a very stable frequency. It produces a single, very pure frequency to feed into the mixer. A crystal having a frequency between 8.800 MHz and 9.000 MHz is suitable for this circuit. The oscillator and mixer functions are both carried out inside IC1.

The first mixer

Yes, this stage 'mixes' two signals together. In this case, the two signals are (i) from the aerial via the RF Filter, and (ii) from the crystal oscillator. Two

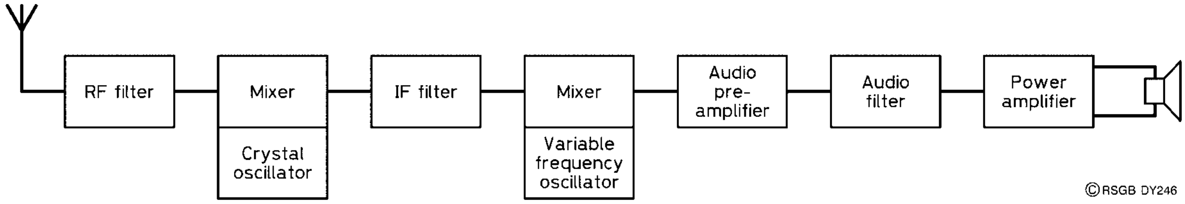

Figure 1 Block diagram of the Yearling, showing how the various stages fit together to make a complete radio receiver

bands of signals emerge from the mixer. The first is centred upon a frequency equal to the incoming signal frequency **added** to the crystal frequency, and the second is centred upon a frequency equal to the incoming signal frequency **subtracted** from the crystal frequency. Look at an example – if the signal is at 14 MHz and the oscillator at 9 MHz, then the mixer outputs will be 14 + 9 = 23 MHz and 14 – 9 = 5 MHz.

Intermediate frequency (IF) filter

It is the purpose of the IF filter (centred around L2) to select only one of these two bands of frequencies emerging from the mixer. In this case, it is the lower band of frequencies (around 5 MHz) which we select. This is because, in general, lower frequencies are easier to handle than higher ones.

Variable-frequency oscillator (VFO)

The VFO (part of IC2) enables us to tune into a particular station, and operates over a band of frequencies between 5 MHz and 5.35 MHz in this receiver. You will notice that IC1 and IC2 are the same type of chip, so that you will be expecting another mixer stage to be associated with the VFO. You are quite right!

The second mixer

This mixer obeys exactly the same rules as those of mixer 1. Sum and difference frequencies are produced, like this. Mixing is between the incoming IF signals (around 5 MHz) and the VFO signals (around 5 MHz), producing output frequency bands centred upon 10 MHz and 0 MHz. The use of the words ‘band of frequencies’ throughout this explanation is intentional. If all the signals were pure, there would be no bands; the bands are produced because of one thing – the modulation imposed on the pure frequencies at the transmitter. So, the ‘bands’ contain the one thing that we want to extract from the signal, and that is the speech or Morse code that the signal contains. The band of frequencies at 0 MHz is just that – the audio frequencies we want in the loudspeaker. Because of this, the audio frequency output of the second mixer is selected and passed on for amplification.

The audio preamplifier

Preliminary amplification of the minute audio signal which emerges from the second mixer is provided by IC4a, which will respond only to audio signals, automatically rejecting the 10 MHz signal.

The audio filter

The bandwidth of normal speech when transmitted by an amateur station is around 3 kHz, so there is no advantage to be gained in amplifying frequencies greater than this. IC4b is known as a low-pass filter, because it passes (lets through) lower frequencies and rejects higher ones.

The power amplifier

IC5 produces the final audio amplification and provides enough power (about 350 milliwatts (mW) to drive a small speaker.

How does it work on 80 m?

If you have fitted the 80 m modification to your receiver, you are probably wondering how the circuit works at this different frequency. Firstly, the filter which you fitted selects the 80 m band instead of the 20 m band. The only other slight difference lies in the way the first mixer stage works. Its job is to produce sum and difference frequencies from the incoming signal and crystal frequencies. On 20 m, it did this by subtracting the crystal frequency (9 MHz) from the incoming frequency (14 MHz) to produce an IF output of 5 MHz. On 80 m, the incoming frequency (3.5 MHz) is subtracted from the crystal frequency (9 MHz) to produce an IF output of 5.5 MHz, which is still within the tuning range of the VFO in the next stage.

73 A field strength meter

Introduction

Having a reliable field strength meter is always useful. There are two basic types – the tuned variety (also known as an absorption wavemeter) and the broad-band variety. A project centred on the tuned variety can be found in this book (*An HF absorption wavemeter*), and the way it works is also presented (*The absorption wavemeter*). This project describes an untuned or broad-band type which is every bit as useful as the tuned type, when it is the level of RF in which you are principally interested.

With the possible exception of the meter, you may have all the parts needed in your junk box. The prototype was distinguished by its meter, found at a rally, and marked 'Safe', 'Dangerous' and 'Explosive' on its scale.

How it works

The basic field strength meter uses the circuit of a crystal set, but with a meter replacing the headphones. A better design, which is used here, is that of a voltage doubler, giving more sensitivity. **Figure 1** illustrates the voltage doubler circuit.

The AC input shown will be our RF input, which will be explained soon. Diode D1 will pass the positive half of the signal and use it to charge up C1 to the peak value of the signal. D2 uses the negative half of the signal to charge up C2 to the same value. Because C1 and C2 are in series, the peak voltage appearing across *both* of them (which is the DC output voltage) is equal to *twice* the peak input voltage, hence the name voltage doubler. If you're wondering why DC is present at the output when AC comes in at the

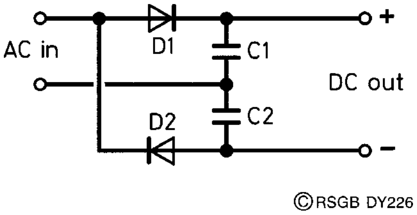

Figure 1 Voltage doubler circuit

input, remember that capacitors pass AC (RF) but block DC; thus the RF is shorted out through C1 and C2, but the steady voltage (DC) remains across the two capacitors. The voltage is thus proportional to the size of the RF signal applied at the input.

The circuit

The voltage doubler is converted into a field strength meter using the circuit of **Figure 2**. A piece of wire serves as the aerial to provide an RF signal across the radio-frequency choke (RFC). A choke is an inductor which is large enough to prevent the RF passing through it – it 'chokes' the RF. This produces the maximum RF signal at the input to the voltage doubler, and the DC output from it is measured on the meter.

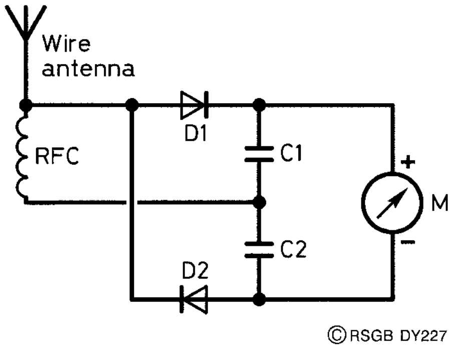

Figure 2 Field strength meter

The parts used are not very critical. The RF choke can have any value between about 1 mH and 2.5 mH. Almost any common diode such as the 1N914 or 1N4148 can be used for D1 and D2. The two capacitors could be any value between 1 nF and 100 nF (0.001 μF and 0.1 μF). The meter should be reasonably sensitive, with a full-scale deflection (FSD) in the range 50 μA to 100 μA. Look for VU meters at rallies – these are ideal.

Construction

The prototype circuit was made on matrix board. If you don't want to use pins with the board, simply push the component leads through the holes and make the connections on the underside of the board, either with the excess component leads themselves, or with ordinary connecting wire.

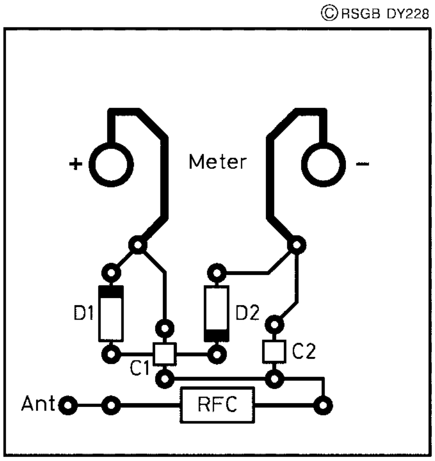

Figure 3 Board layout and interconnections

Component side of board shown.
All connections are under board

Figure 3 shows how the parts are placed and connected. The matrix board can be mounted directly on the meter using its terminal bolts!

In use

Needing no power supply other than an RF signal, just connect it all up and leave it to work! A short length of insulated wire is enough to pick up some RF and display it on the meter. Using a 200 μA meter, about 3 metres of wire gave a good deflection on the meter. To increase deflection, the wire can be wrapped around the aerial lead, provided that the wire is PVC covered and doesn't come into contact with the aerial wire.

This requires a little experimentation. Try a long piece of wire first and adjust its position until the meter needle kicks whenever there is a transmission. It is very reassuring to see the meter moving during a transmission. Although SWR meters also indicate power, they are usually set to read reverse power, and show little or no movement during transmission.

Parts list

D1, D2	1N914, 1N4148 or similar
C1, C2	10 nF disc ceramic
RFC	Miniature axial choke (1 mH)
Meter	Surplus VU meter or similar
Matrix board or similar	

Components are available from Maplin.

74 Preselector for a short-wave receiver

Introduction

A preselector is a simple RF tuned amplifier which is inserted between the aerial and the receiver. It provides some extra gain and may improve the overall performance of the receiver. This project uses a Field-Effect Transistor (FET) amplifier in grounded-gate mode.

The design has a tuned circuit at both the input and output which, with excessive gain and poor construction, would produce only one thing – oscillation! So, to avoid this happening, we will have only a low gain, and use a circuit which provides good isolation between input and output. The grounded-gate FET amplifier fulfils both these criteria. It will also cover a frequency range from about 7 to 30 MHz, which includes most of the HF amateur bands.

The circuit

This is shown in **Figure 1**. The signal from the aerial arrives at an RF transformer, the secondary of which is tuned with capacitor VC1a. The output from the tuned circuit is taken from a tap on the secondary to the source of the FET. The gate is grounded (earthed) and the amplified signal appears at the FET drain, which is then fed to the primary of another RF transformer, which is tuned by VC1b. The output to the receiver comes from the secondary of the RF transformer.

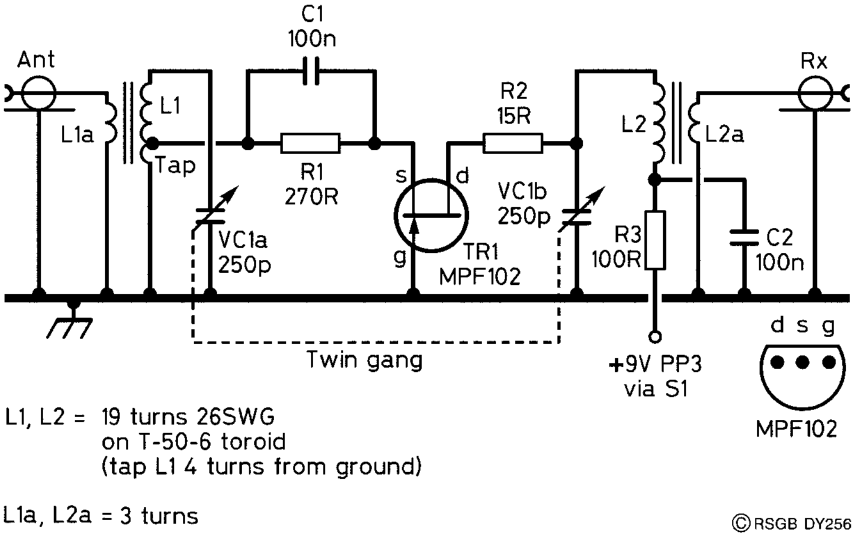
C1
100n
Ant
Rx
L1
L1a
Tap
R2
15R
L2
L2a
R1
270R
s
d
g
VC1b
250p
VC1a
250p
TR1
MPF102
R3
100R
C2
100n
Twin gang
+9V PP3
via S1
d s g
MPF102
L1, L2 = 19 turns 26SWG
on T-50-6 toroid
(tap L1 4 turns from ground)
L1a, L2a = 3 turns
©RSGB DY256

Figure 1 HF preselector, circuit diagram

Notice that the two RF transformers are identical, but they are used ‘back to back’, with the secondary of the first and the primary of the second being tuned. They are tuned with identical capacitors, fitted on the same shaft of a variable capacitor. We say that the two capacitors are ‘ganged’. Because L1 and L2 are the same, and VC1a and VC1b are the same, both RF transformers should be resonant at the same frequency, no matter what that frequency is.

Construction

The final layout should look something like that shown in **Figure 2**. The external connectors and controls being two SO-239 sockets for connection to your receiver and aerial, a tuning control and its associated scale, and an on/off switch.

The circuit can be put together on a plain matrix board, using pins to anchor the components, or simply by pushing the component leads through the board and making connections on the underside. The layout of the prototype is shown in **Figure 3**. Mounting the board to the aluminium box is accomplished with bolts, solder tags and stand-off insulators.

Check your construction against the circuit diagram and against the layout diagram. Wire in the PP3 battery clip, put the switch in the ‘off’ position, and fit the battery. Testing can be carried out without fitting the top of the box.

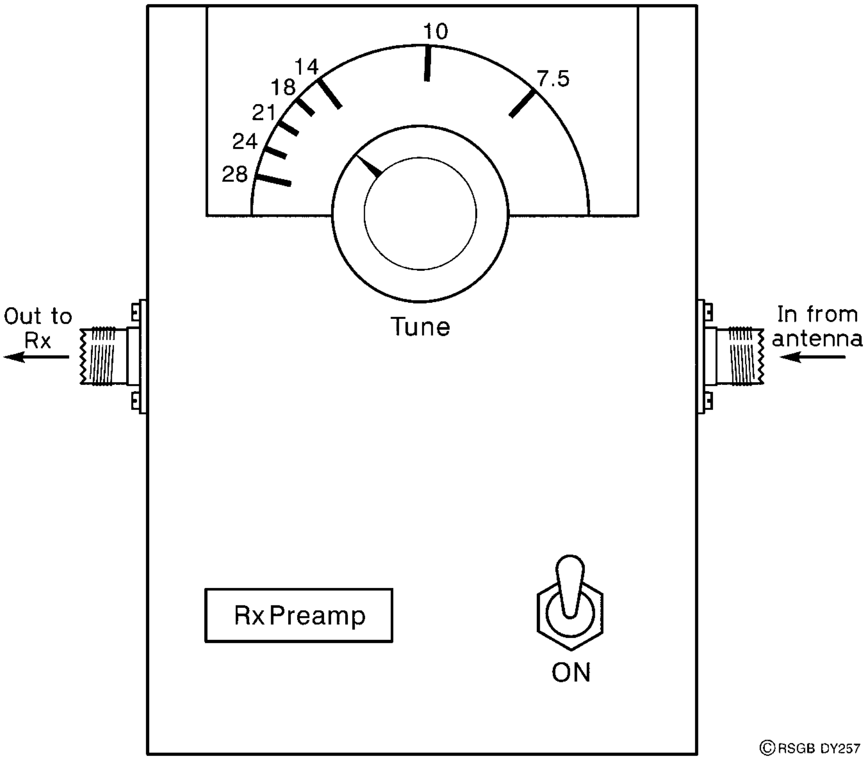

Figure 2 HF preselector, front panel layout

Testing

Don't fit your preselector yet. Tune your radio to a broadcast station, preferably a fairly weak one. Disconnect the aerial and fit your preselector between the aerial and receiver. Switch it on and rotate the tuning knob slowly. You should find a position where your original station is received more clearly than before. If it doesn't work at all, recheck your wiring. Is there a positive voltage on the drain of the FET? If not, work back towards the positive battery terminal. Is there a voltage at the junction of L2 and R2? Is there a voltage at the junction of L2 and R3? Is there a voltage at the junction of R3 and the battery lead? If there isn't a voltage at that point, then you have probably mounted your switch upside down, and it is off, not on! It's a common mistake.

Calibration

This is not obligatory, surprisingly enough. However, if having a frequency scale appeals to you, then using an RF signal generator (or using the services of a friend who has one) is the simplest solution. Feed in a weak modulated

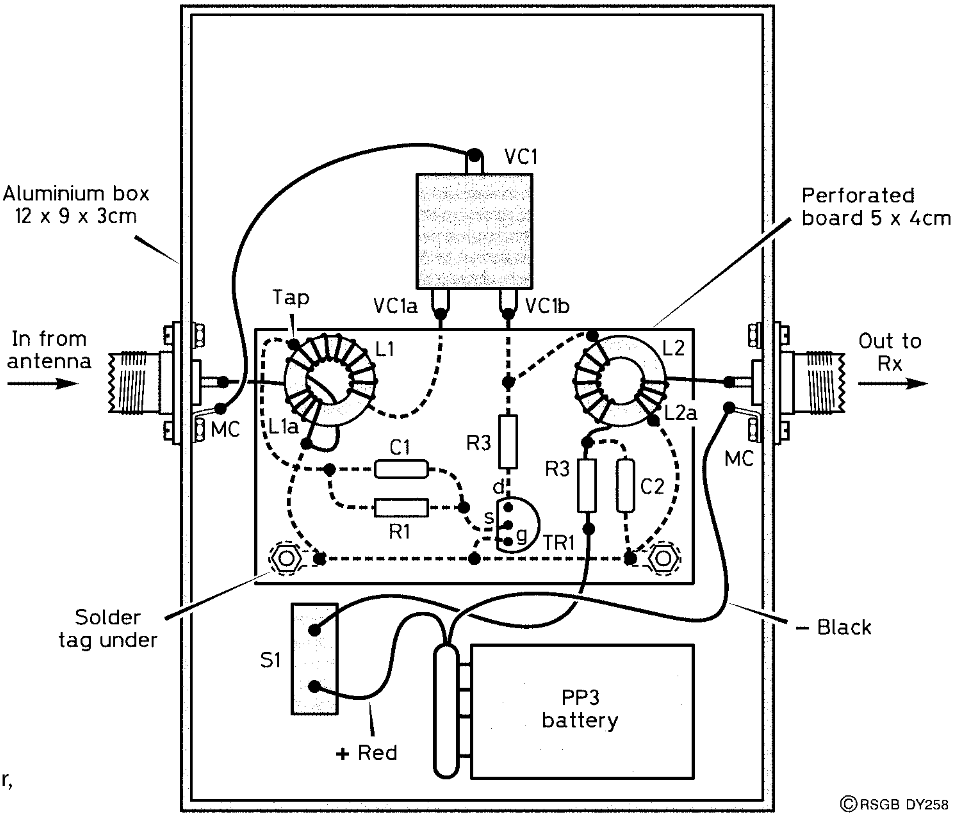
VC1
Aluminium box
12 x 9 x 3cm
Perforated
board 5 x 4cm
Tap
VC1a
VC1b
In from
antenna
L1
L2
Out to
Rx
MC
L1a
L2a
C1
R3
R3
C2
MC
d
s
g
R1
TR1
Solder
tag under
– Black
S1
PP3
battery
+ Red
©RSGB DY258

Figure 3 HF preselector, component layout

signal from the generator to the preselector, and rotate VC1 until the signal is maximum. Mark this frequency on your dial. Repeat the process for the frequencies shown on the dial in Figure 2.

If you have a commercial transceiver, feed its output into a dummy load (such as the type described in *A Switched Dummy Load*, elsewhere in this book) using a distinctive modulating signal such as an idling RTTY signal. Set the receiver to the same frequency with the preselector out of circuit. Insert it into the aerial lead, and search for the signal with VC1 until it gives the maximum deflection on your S-meter, then mark the frequency on your dial.

Parts list

Resistors: all 0.25 W carbon film or better

R1	270 Ω
R2	15 Ω
R3	100 Ω

Capacitors	
C1, C2	100 nF (0.1 μF) ceramic
VC1	250/250 pF polyvaricon
Inductors	
L1, L2	19 turns 26 SWG enamelled copper on T.50.6 toroid L1 has a tap 4 turns from ground end L1a and L2a – 3 turns wound over previous winding
Semiconductors	
TR1	MPF102 FET
Additional items	
Matrix board	To fit aluminium box (see Figure 3)
Aluminium box	12 × 9 × 3 cm
Battery and connector	PP3 9 V
S1	SPST on/off switch
Knob	As required
SO239	Sockets – 2 required

75 An audible continuity tester

Introduction

This is not the only continuity tester in this book. This alone attests to their use, so you may well want to experiment with several designs, then come up with one of your own! The very simplest form of continuity tester is probably a battery and a bulb in series, with the circuit being closed by connecting it to a fuse or other object being tested for continuity. The bulb could be replaced by a buzzer to give an audible indication. The current taken by the buzzer could damage some components, however. An ohmmeter can also be used, and is very popular for the purpose, as it indicates whether the circuit is low, medium or high resistance. This project has the advantages of indicating whether there is no continuity, some resistance or good continuity, and making an audible sound, so that you don't have to move your eyes from the circuit while making the test.

The circuit

Using only one integrated circuit, four components and a battery, this is a particularly simple circuit, as **Figure 1** shows. IC1 is usually used to flash an LED from a 1.5 V source, and to have a low current consumption. By changing the component values, IC1 is made to oscillate at audio frequencies, and we can hear these through the loudspeaker, LS1. Low-impedance speakers can be used, but result in an increased current drain, so the higher the impedance, the better.

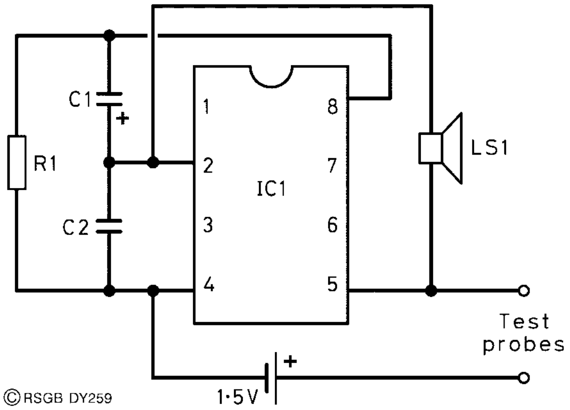

Figure 1 Continuity tester, circuit diagram

Construction

The layout of the circuit on Veroboard measuring 7 strips by 15 holes is shown in **Figure 2**. Start by making the four track cuts which will lie underneath the IC. Use a track cutter or a 3 mm (⅛ inch) twist drill rotated between thumb and forefinger for this. Solder the Veropins in place, followed by the wire link. Then solder in the IC holder, the resistor and the capacitors, making sure that the electrolytic capacitor, C1, is connected the correct way round.

Check, with the board against a bright light, that there are no shorted tracks, either by large blobs of solder or by copper swarf from the track-cutting process. Then insert IC1 into its holder the right way round. The probes can be ordinary connecting wire, the free ends being tinned with solder to prevent wire whiskers from touching components other than the one you are testing.

If you are happy that the circuit and the wiring appear to be correct, fit the battery into its clip. Nothing should happen until the probes are touched together, when you should hear a note from the loudspeaker. If nothing happens, all you can do is to recheck your circuit, as there is nothing else to go wrong!

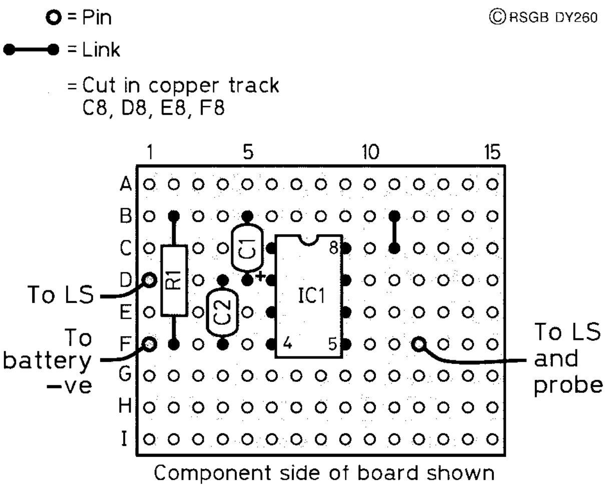
= Pin
= Link
= Cut in copper track C8, D8, E8, F8
©RSGB DY260
1
5
10
15
A
B
C
D
E
F
G
H
I
R1
C1
C2
+
IC1
8
4
5
To LS
To battery −ve
To LS and probe
Component side of board shown

Figure 2 Continuity tester, component layout

Any box can be used; there is no justification for a metal case, unless you want to make use of something, which is to hand, such as a tobacco tin. Alternatively, any suitable plastic box will do.

Use

The tester will give different pitch notes for resistances of different values. The higher the resistance, the higher the pitch from the speaker. Try it with small inductors, and you will learn to recognise the different tones produced by the IC.

Safety notice

Using the continuity tester on components *in situ* is not advisable, as the results could be misleading. It can be dangerous to make measurements *in situ* on equipment, which is operating. If you must make such tests, always disconnect the equipment form its power source first.

Parts list

Resistor

R1	1000 Ω (1 kΩ), 0.25 W carbon, 10% or better

Capacitors

C1	10 μF electrolytic, 16 V
C2	0.1 μF subminiature polyester or ceramic

Integrated circuits

IC1	LM3909N

Additional items

	Veroboard 7 strips by 15 holes
LS1	64 Ω miniature loudspeaker
	IC holder
	Veropins
	Connecting wire
	Test leads
	1.5 V AA alkaline battery

76 An experimental 70 cm rhombic aerial

Introduction

Most commonly used aerials can be classed as *resonant* or *standing-wave* aerials. There is another class known as *non-resonant* or *travelling-wave*. Resonant aerials, such as the dipole, are narrow-band; this occurs because resonance occurs only over a narrow band of frequencies. Travelling-wave aerials, on the other hand, can operate over a wide band of frequencies.

The *rhombic* is an example of a non-resonant or travelling-wave aerial. It is often employed for fixed commercial and military short-wave radio links. Made with wire, it has a diamond shape when looking down on it from above. The four corners are supported on four masts. It is a very effective aerial, and has good gain, a quality which can be judged from the *polar diagram* shown in **Figure 1**. This has been obtained from a computer program, and so is the perfect shape for a rhombic aerial.

Theory (but only a little)

A polar diagram shows graphically the ability of an aerial to radiate (or receive) more effectively in one direction at the expensive of the radiation in other directions. Figure 1 shows the polar diagrams of two aerials, a simple 70 cm dipole and the rhombic described in this project. The dipole has the

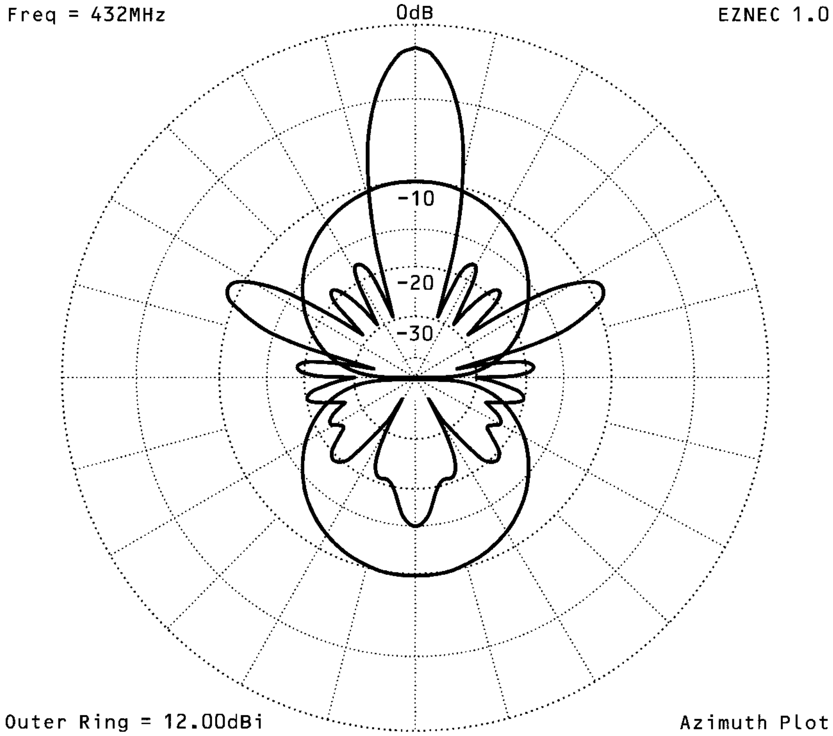

Figure 1 The horizontal polar diagrams of the small rhombic antenna and a dipole compared

©RSGB DY262

well-known figure-of-eight shape, showing a symmetrical angular behaviour about the direction of the aerial and about a direction perpendicular to the aerial. That of the rhombic, on the other hand, is quite irregular by comparison, but is still symmetrical about one axis only, not two. It's this asymmetry that gives the rhombic its gain, by virtue of its *front-to-back ratio*. This is the ratio of the power radiated forwards to that being radiated backwards. Notice the large *lobe* (lump) at the top of the polar diagram; this is the direction in which the aerial transmits best. The lobe in the opposite direction has been reduced significantly, allowing more power to be directed forwards, not backwards

Problems

The rhombic is not found in every amateur's back garden, despite its attractions. To work best each edge of the diamond shape should be about two wavelengths long. Hence, a rhombic for the 20 metre band could be about 80 m from tip to tip! Another disadvantage is that, because of its size, it cannot be rotated.

It will operate well over a range of frequencies. An HF rhombic could work on the 7, 10, 14, 18, 21 and 28 MHz bands. One for the lower VHF frequencies could operate on the 50, 70 and 144 MHz bands.

Its large size is less of a problem at UHF. A portable rhombic can be made for 70 cm which can be used with a hand-held transceiver. The aerial design to be described here will give a gain of up to 9 dB relative to a dipole. This is written as 9 dBd, the second ‘d’ meaning ‘relative to a dipole’. This is equivalent to improving your signal by 1.5 S-points at the receiver or (and you may be surprised by this) by fitting a linear amplifier to your transceiver that would take 5 W input and produce 40 W output to a normal dipole! Consider the relative costs of the two approaches to producing the same received signal. You will also receive everyone else’s signals 1.5 S-points better than before!

Unlike the popular Yagi aerial, this design has no critical dimensions and can be folded up for transport by car or bicycle. It has a fairly high input (or feed) impedance, being fed usually by balanced twin feeder. To feed it with standard 50 Ω coaxial cable, a matching transformer in the form of a *balun* is required. A balun will convert the *bal*anced (symmetrical) aerial impedance to the *un*balanced and lower impedance of the coaxial cable. The details of how to build the balun, which in this case is a half-wave transformer, are included in the constructional details.

Construction

The aerial frame is made up of 1 cm × 2 cm strips of wood fixed to a plywood centre using 30 mm long M4 bolts, as shown in **Figure 2a**. The outer bolts fixing the front and side supports can be removed for folding prior to transportation, as shown in **Figure 2b**.

The wires are fixed to the front and rear supports using screw connectors, sometimes called ‘chocolate block’ connectors. Detail X and Y of Figure 2 show how this is done. The side supports have holes in the ends through which the wire is threaded.

The aerial must be mounted in the horizontal plane using a small shelf bracket attached to the centre plate, using the same bolts that hold the rear support. The other half of the bracket may be mounted to a vertical mast using screws or jubilee clips (hose clips).

The balun

Cut a 23 cm length of coaxial cable; this is to be our half-wave transformer. Cut and remove 2 cm from each end of the sheath. Make the braid into a pigtail at each end. Cut and remove 1 cm of the inner insulator from each

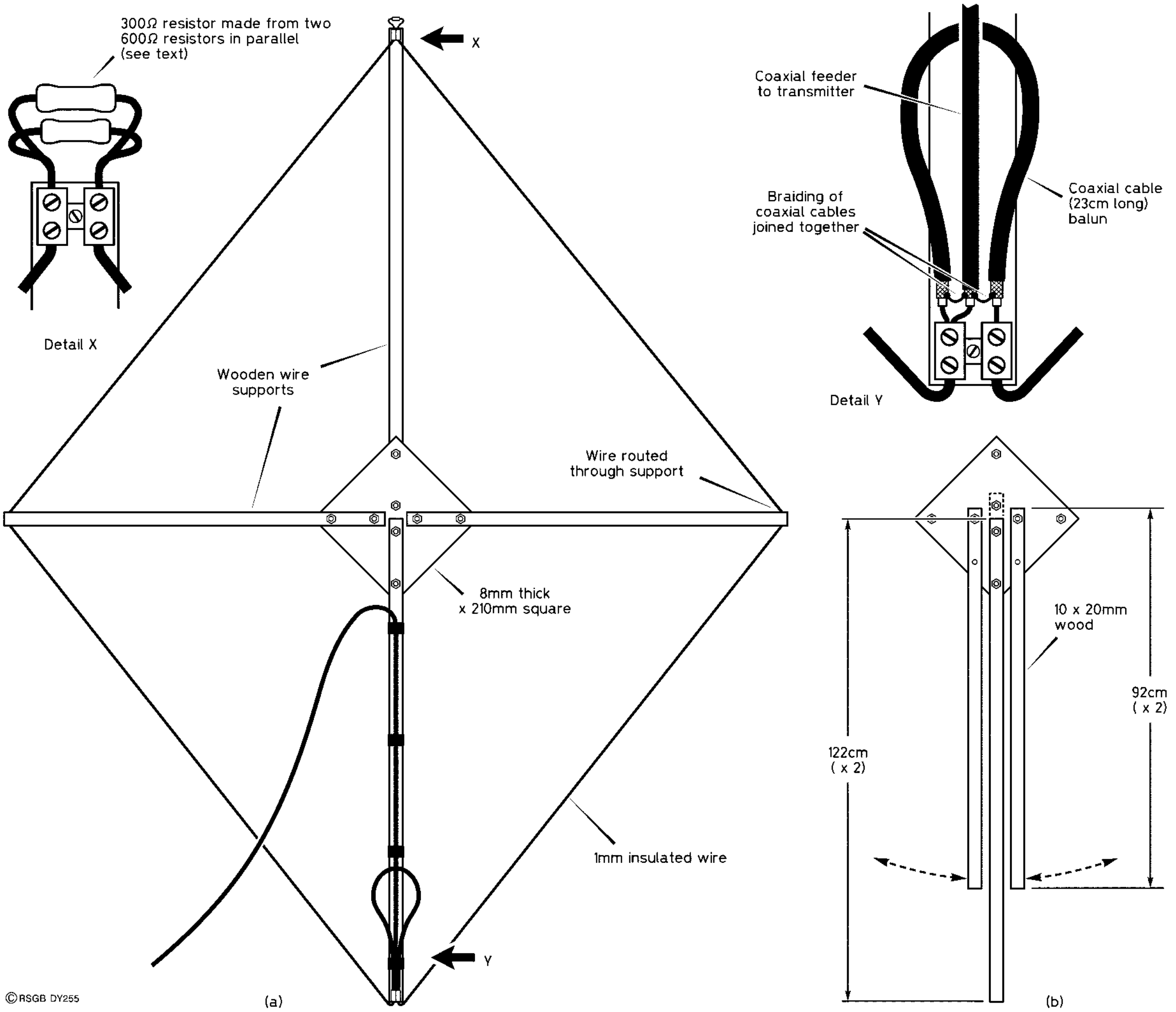
300Ω resistor made from two
600Ω resistors in parallel
(see text)
Detail X
X
Wooden wire
supports
Wire routed
through support
8mm thick
x 210mm square
1mm insulated wire
Y
©RSGB DY255
(a)
Coaxial feeder
to transmitter
Braiding of
coaxial cables
joined together
Coaxial cable
(23cm long)
balun
Detail Y
10 x 20mm
wood
92cm
(x 2)
122cm
(x 2)
(b)

Figure 2 Construction of the UHF rhombic antenna

end, leaving 1 cm of the centre conductor exposed. Fasten this piece of cable to the chocolate block connector together with the coaxial feeder cable as shown in Detail Y of Figure 2. The braids of all three prepared ends are soldered together, but are not connected to anything else. The inner conductor of the feeder from the transceiver is soldered to one end of the centre conductor of the 23 cm piece, and is connected to one end of the rhombic by one side of the chocolate block. The other end of the half-wave transformer also goes to the chocolate block, where it is connected to the other side of the rhombic loop.

The opposite corner of the rhombic, as shown in Detail X of Figure 2, shows that the loop is broken at the chocolate block, and is ‘terminated’ by two 600 Ω resistors in parallel. This makes the aerial a broad-band travelling-wave device, and gives it its directivity and gain. There is a rule of thumb, which includes a safety factor, which says that the terminating resistor must be able to absorb one-half of the maximum transmitter output power. So, if you use two 2 W 600 Ω carbon resistors (*not* wire-wound resistors) in parallel, you can use a transmitter with an output of 8 W, which is more than adequate for 70 cm hand-helds.

Using the rhombic

Fit the aerial to a pole or mast in the horizontal plane, if you intend to use CW or SSB, but in the vertical plane if you want to concentrate on FM work. In the latter case, a wood or fibreglass pole is mandatory. Tune to a local repeater, whose signal strength you know. Rotate the aerial to face the repeater, and you should see that the signal strength is greatly improved! Verify the directional properties of the aerial by rotating it and observing the changes in signal strength on your S-meter.

77 Water level alarm

Introduction

This is a simple device which, when you have built it, you begin to wonder how you ever managed without it! It will sound a buzzer when the level of water reaches a particular point, which you can set and vary at will.

Detecting water

Water is not too difficult to detect electrically, because it is a conductor of electricity. Not a good one, but sufficiently so to carry enough current to control a warning circuit.

The circuit is shown in **Figure 1**. If you imagine a resistor placed between the ends marked ‘water detector’, the circuit is then recognisable as a transistor used as an electronic switch. Whether the switch is off or on depends on the value of this resistor. If the resistance is low, the transistor switches on and the buzzer sounds. If the resistance is high, the transistor switches off and the buzzer stops.

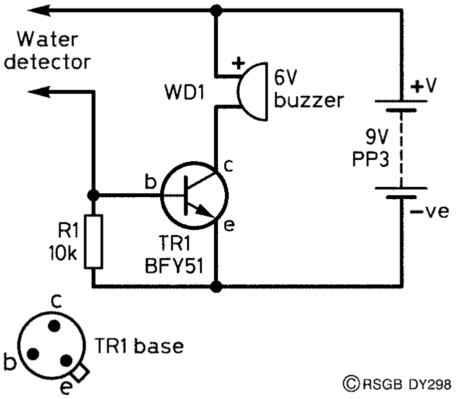
Water
detector
WD1
+
6V
buzzer
+V
9V
PP3
−ve
b
c
e
R1
10k
TR1
BFY51
c
b
e
TR1 base
©RSGB DY298

Figure 1 Circuit diagram of the water level detector

Construction

The unit is made in two separate sections – the detector board (**Figure 2**) and the main board (**Figure 3**). The main board is plain perforated board (without copper strips). The components are pushed through holes in the main board and connected together on the underside. The battery is held on with sticky tape. No on/off switch is used, as the circuit consumes virtually no current when it is not sounding the buzzer.

Unlike some buzzers, this type is polarised, and the red lead must go to the supply as shown in Figs 1 and 3.

The detector board is a piece of standard matrix board – the type with copper strips along one side. It measures 9 strips by 15 holes, and is wired in such a way that alternate strips are connected together, as Figure 2 shows. This gives a greater surface area for the water to touch, thus increasing the sensitivity of detection.

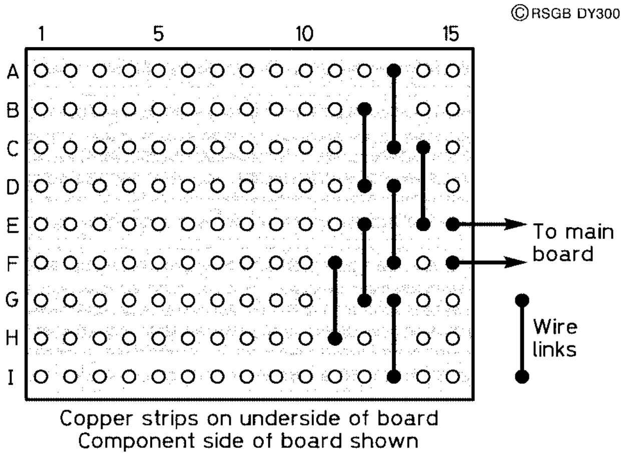
©RSGB DY300
1
5
10
15
A
B
C
D
E
F
G
H
I
To main
board
Wire
links
Copper strips on underside of board
Component side of board shown

Figure 2 Layout of the main board

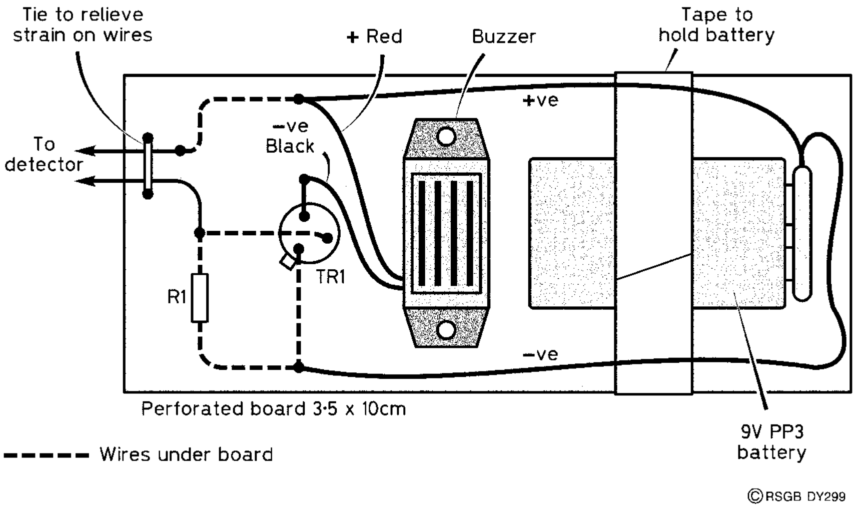
Tie to relieve strain on wires
+ Red
Buzzer
Tape to hold battery
To detector
–ve Black
+ve
TR1
R1
–ve
Perforated board 3·5 x 10cm
9V PP3 battery
Wires under board
©RSGB DY299

Figure 3 Water detector constructed from stripboard

The two boards are connected by ordinary twisted pair, enabling the detector to hang over the edge of the water container, with the buzzer circuit safely out of the way. When positioning the detector board, the left-hand edge (as shown in Figure 2) should face downwards, with the connecting-wire edge facing upwards. When the water level rises, the water causes current to flow between adjacent strips, and the higher the water, the lower the resistance between them. Eventually, the resistance will be low enough to turn on the transistor and sound the alarm.

Parts list

Resistor

R1	10 000 Ω (10 kΩ), 0.25 W

Semiconductor

TR1	BFY51

Additional items

WD1	6 V miniature buzzer (polarised)
	Battery PP3
	Battery connector
	Strip board 9 strips by 15 holes
	Perforated board 3.5 × 10 cm

78 A delta loop for 20 metres

Introduction

HF aerials take up a lot of room when they are straight. Space can be saved by bending them and, at the same time, giving them properties which are quite different from their original linear forms. The delta loops shown in the accompanying diagrams have the shape of the Greek upper-case letter delta (Δ) upside down. It can be upright, except that it is more practical to have the feed point at the bottom than at the top.

Putting the loop together

This is the design for a loop to be used on the 20 metre band. It is a very popular band and carries most of the amateur radio DX traffic. The aerial is light in weight and operates best if the top section is at least 30 ft above the ground, with the feeder point being about 6 ft above ground. All you need is some good wire, some polypropylene rope, some insulators and a method of matching the output of this 75 Ω balanced aerial to the 50 Ω unbalanced coaxial input of your transceiver or receiver.

Each of the three sides must be 7.2 m long. Using tent pegs or six-inch nails, mark out the three corners on the lawn, making sure that each pair of pegs or nails is exactly 7.2 m apart. Put the wire around the nails, together with two insulators on the wire for the two top corners. These insulators can be of the 'dog bone' variety, or of the home-made type, using a piece of flat plastic with holes in each end. The use of these is shown in **Figure 1**. Located at the positions of two of the tent pegs or nails, they must be secured at those points by any convenient means.

Now attach a support rope to the free end of each of the insulators, ready for suspension, as shown in the diagrams, from any convenient supports (house, tower, tree, etc.). The top section of the loop could be taped to a continuous piece of rope between the two supports. You might like to try this if you think it is easier. Then, when you find some dog-bone insulators, you can change the design and see if there is any noticeable change of performance.

Connecting to the radio

If you are the proud possessor of an aerial tuning unit (ATU) with a balanced output, all you will need is a length of twin cable soldered to the

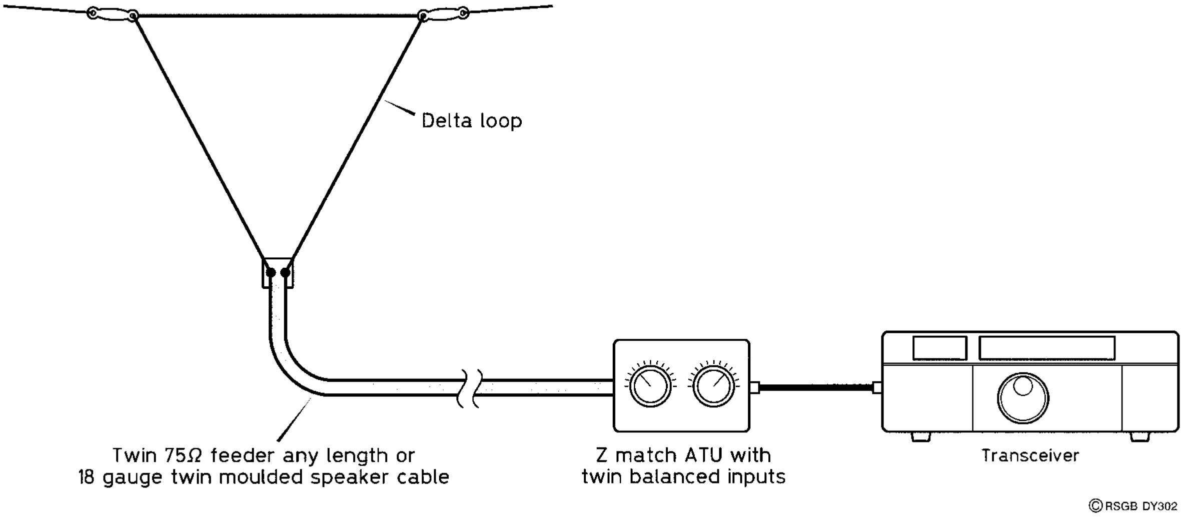
Delta loop
Twin 75Ω feeder any length or
18 gauge twin moulded speaker cable
Z match ATU with
twin balanced inputs
Transceiver
©RSGB DY302

Figure 1 Multiband version of the delta loop connected to the ASTU using 75 Ω twin feeder

ends of your delta loop and connected to the balanced output terminals on the ATU, as Figure 1 shows.

Another way is to buy a ferrite-cored 1:1 balun, and use it as indicated in **Figure 2**. This produces an aerial which will operate over the whole range of amateur and commercial short-wave frequencies, when used with an ATU.

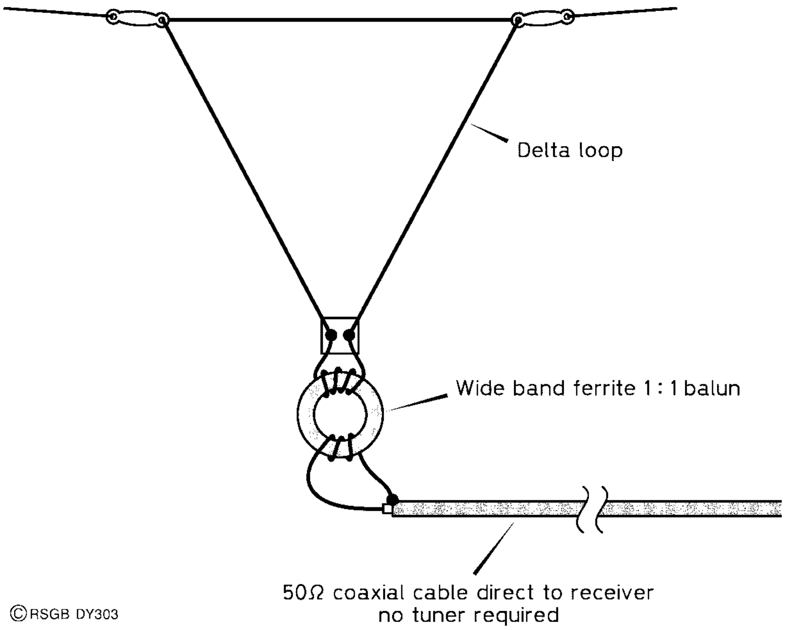
Delta loop
Wide band ferrite 1 : 1 balun
50Ω coaxial cable direct to receiver
no tuner required
©RSGB DY303

Figure 2 Multiband version of the delta loop using 75 Ω coaxial cable. This arrangement requires an ASTU if used with a multiband antenna

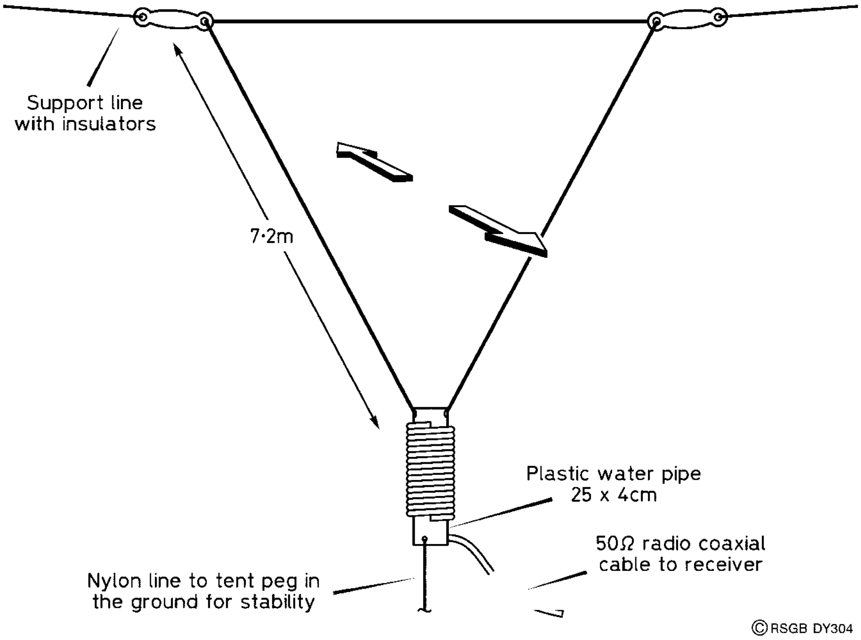

Figure 3 Single band version of the loop antenna using a coaxial balun. The direction of maximum signal strength is indicated by the arrows

However, since this aerial is designed to be a single-band type, there is a simpler answer to the problem of matching the aerial to your 50 Ω coaxial cable – **Figure 3** shows this. Make a tuned balun with a length of plastic water pipe, 25 cm long and 4 cm diameter and a piece of good-quality TV aerial cable. TV aerial cable has a characteristic impedance of 75 Ω, compared with the 50 Ω of the coaxial feeder that we usually use with amateur radio equipment. Use the type with a brown sheath and a closely knit earth braid, not the type having an earth foil inside the sheath.

The length we need for the balun is 3.8 m, but always allow between 3 and 4 cm extra for preparing the ends. Drill two small holes diametrically opposed in the top of the tube; these will be used to anchor the two ends of the delta loop, as shown in Figure 3. Drill another single hole in the bottom end of the tube, which will be used to anchor a nylon line going to the ground to add stability to the loop. Then, after drilling a pilot hole, drill a 5 or 6 mm hole near the top end of the tube, as shown in **Figure 4**. Prepare both ends of the coaxial cable, then feed one end into the tube, far enough for its ends to be soldered to the ends of the loop when the assembly reaches that stage.

Now, close-wrap the cable around the tube until only about 3 cm remain. Holding the cable tightly, drill another 6 mm hole beside the free end of the cable and feed it into this hole. At this point, feed the two ends of the loop into the top two holes and twist it back on itself. Figure 4 shows how this is done. Then, solder the ends of the coaxial balun to the ends of the loop.

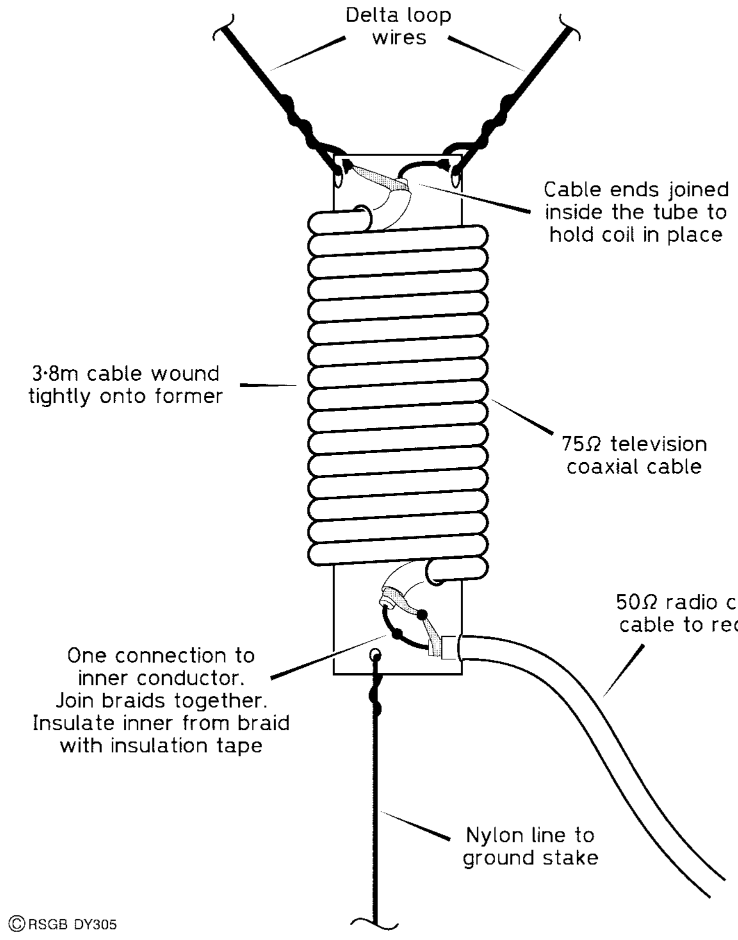
Delta loop wires
Cable ends joined inside the tube to hold coil in place
3·8m cable wound tightly onto former
75Ω television coaxial cable
50Ω radio coaxial cable to receiver
One connection to inner conductor. Join braids together. Insulate inner from braid with insulation tape
Nylon line to ground stake
©RSGB DY305

Figure 4 Construction of the balun

The bottom ends of the balun are then soldered to the 50 Ω coax which goes to your shack and to the transceiver.

Hoist the aerial into position carefully, being careful not to pull too hard on the support lines. Then, take the nylon line from the bottom of the balun to a peg in the ground. This adds stability to the aerial.

Using the delta loop

It is a directional aerial, as Figure 3 shows. It produces maximum power (and has maximum receive sensitivity) along a direction perpendicular to its own plane. Don't be too concerned with which direction it is pointing at first. Give it a try 'on the air' and see how it performs. Then you can contemplate how to point it in your favourite direction, to the USA, or Australia, for example.

Experimenting

You may want to enclose the balun in some sort of weatherproof container. Plastic ice cream containers are favourites for this sort of job. Seal all the holes where wires enter it with silicone sealant or self-amalgamating tape.

If you have used ordinary single-strand or multi-strand wire for your aerial, it will stretch over time under its own weight and that of the balun. Its operating frequency will fall slightly as a result. If you notice a significant difference, then dismantle it, remeasure and fix the sides and erect it again, perhaps facing a different direction. You can buy pre-stretched or hard-drawn wire for such purposes, if you feel that periodic tweaking of your aerial is a chore.

You can make a delta loop for different frequencies simply by scaling the lengths of wire for the loop and for the balun according to the design frequency. If your maths isn't quite up to this, enlist the help of someone well versed either in maths or aerial design!

79 A simple desk microphone

Introduction

A hand-held transceiver is usually the first type used by most novices. It is ideal when used out of doors, in the shack or mobile. However, using one in the shack with an outside aerial is not the best of solutions.

This article describes the design of a desk microphone which can be used either for transceivers using a common microphone and push-to-talk (PTT) line, or for those using a separate PTT line.

It is a simple design, using the barrels of two plastic pens, the lid of an old coffee jar, and the plastic sheath of a DIN plug.

Construction

The end caps and the ink tubes should be removed from two ball-point pens, leaving only the two plastic shells. One of them is used for the vertical microphone support, while the other is used to hold the microphone element.

The vertical support should have the pointed end shaped so that it holds the microphone support at a convenient angle, as **Figure 1** shows.

The base should be prepared next – see **Figure 2**. Four holes are required. Two, A and B, are small holes for the cable to be fed through. A third, C, is slightly larger and is used to take one of the plastic end caps. The fourth hole, D, is determined by the type of switch that you want to use for the PTT function.

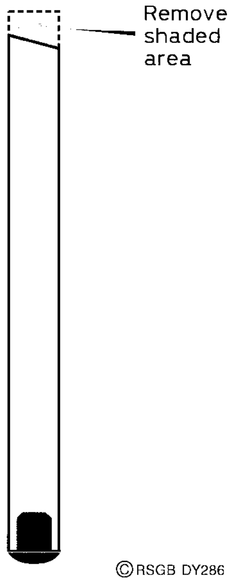
Remove
shaded
area
©RSGB DY286

Figure 1 Vertical support made from ball-point pen shell

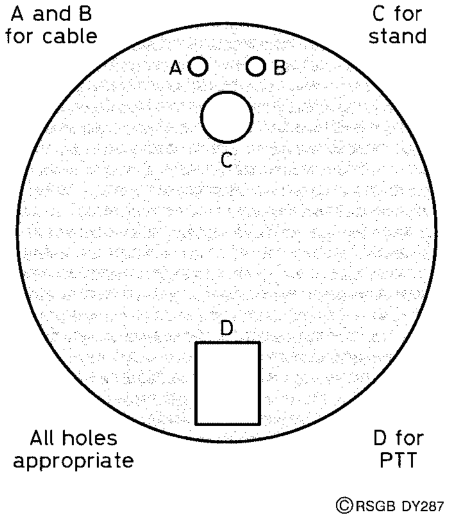
A and B
for cable
C for
stand
A
B
C
D
All holes
appropriate
D for
PTT
©RSGB DY287

Figure 2 The microphone stand base

Some epoxy resin such as Araldite Rapid is used to mount the separate parts. Mix only a small amount at a time, and glue each part of the base assembly and allow to harden before moving on. Be sure to read and follow the instructions for the use of the glue to prevent accidents.

Firstly, glue the end cap to the base, feeding it underneath and pushing it up through hole C. Then, the vertical pen body is placed over the protruding end cap and glued into place, making sure that the top (shaped) end is pointing in the right direction to hold the other plastic tube carrying the microphone. While the glue hardens, place the DIN plug sheath on the second pen body. Place and glue this pen body on the shaped top of the vertical support, making sure that the whole assembly is supported while the glue sets. What you should have now ought to resemble **Figure 3**, but without the microphone and wiring. Leave everything to set for 24 hours.

Any switch of the press-to-make variety can be used for your PTT switch, but choose one which does not apply too much force to operate! If you find the whole assembly is too light and slides over your table, some plasticine can be pushed into the base when all the wiring has been completed.

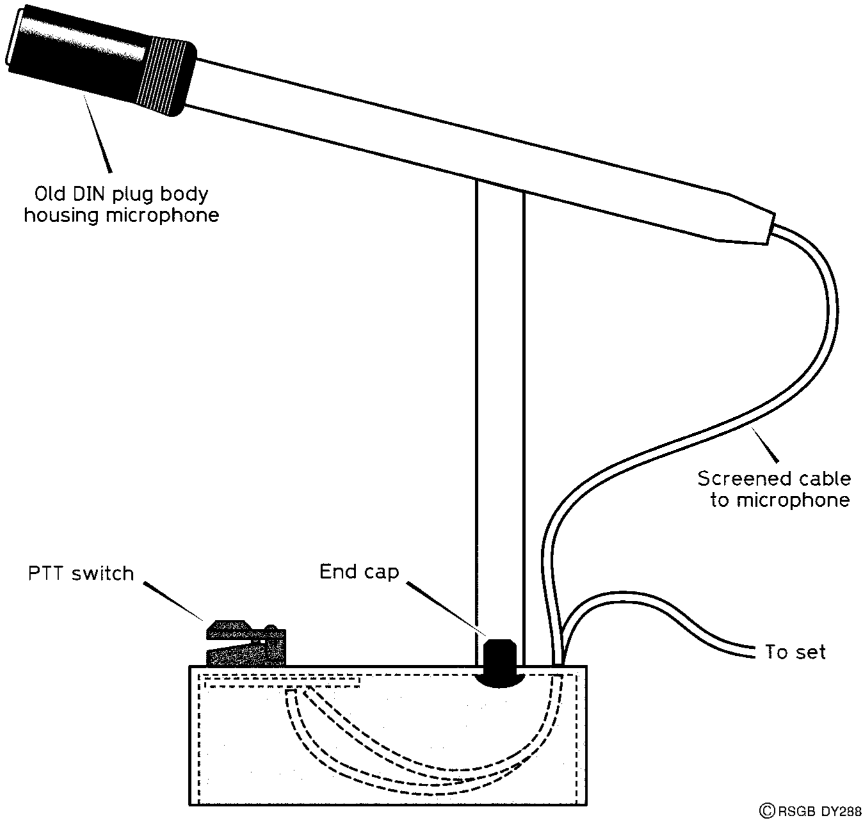

Figure 3 Complete microphone stand assembly

The circuit

A small electret condenser microphone is used, powered by a DC supply. Most hand-held radios use a single screened lead for both the PTT line and the microphone audio lead. In such cases, the circuit of **Figure 4** will operate the PTT function. Note that the PTT switch is in series with the audio lead from the microphone. If you find that the PTT switch does not switch your radio into transmit, reduce R1 from its normal value of 33 kΩ to 27 kΩ. You should find that the original value works with most hand-helds.

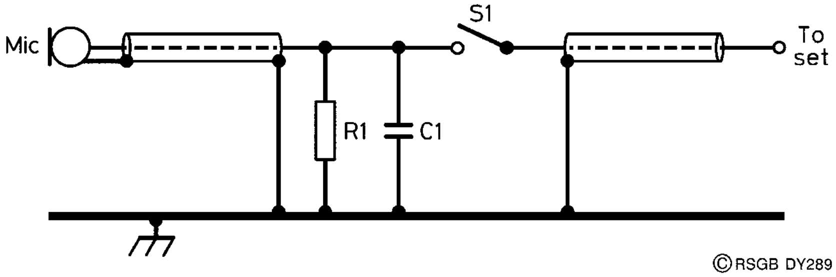

Figure 4 Wiring diagram for hand-held radios

For radios with a separate PTT lead, the circuit of **Figure 5** is used. However, a power supply is needed for the electret microphone. This can be a PP3 battery, or can (in most cases) be derived from the microphone socket on the transceiver. You will need to consult the makers' handbook for this information. Figure 5 shows the PTT switch wired in a 'ground-to-transmit' configuration, which is correct for most base-station transceivers. If you're in any doubt about what your radio needs, consult an experienced friend.

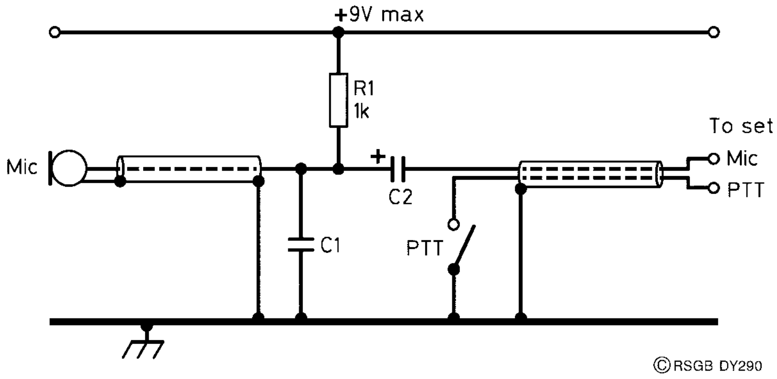

Figure 5 Wiring diagram for radios which use separate PTT lines

The last connections

The microphone element should be connected with screened cable. Tape (masking or insulating) should be wound around the electret insert until it is a snug fit inside the DIN plug sheath. Feed the screened cable through the barrel of the pen until it emerges from the far end. Poke it through hole B and make the connections under the coffee jar lid. The components can all be mounted on the PTT switch or on a small piece of Veroboard mounted under the top surface. The output lead emerges through hole A and is of sufficient length to reach your transceiver. A suitable plug needs to be fitted to it.

Parts list

Resistors: all 0.25 W carbon, 5% tolerance	
R1 (Figure 4)	33 000 Ω (33 kΩ)
R1 (Figure 5)	1000 Ω (1 kΩ)
Capacitors	
C1 (Figure 4 and 5)	0.1 microfarads (0.1 μF) disc
C2	10 microfarads (10 μF) electrolytic
Additional items	
S1	Push-to-make
Mic	Electret microphone (Maplin type FS34W)
Screened cable	As required
Plug	As required by your radio
Veroboard	If required
PP3 battery and clip	If required
Two old plastic ball-point pens	
Lid of coffee jar	
Plastic sheath from DIN plug	
Araldite or similar epoxy resin glue	

80 Morse oscillator

Introduction

This is not the simplest Morse oscillator to build, but it differs from the simple circuits in that it produces a pure note, not a coarse, rasping sound. People who have practised Morse using a non-sinusoidal oscillator sometimes find that they have trouble copying Morse code with a pure tone. As the pure tone is the correct way to receive Morse code, it is important that you should learn to listen to the code from a pure oscillator – so here's one!

The twin-T

There are many oscillator circuits, and there are many variations of the twin-T oscillator that we are going to use. **Figure 1** shows one version of a very useful circuit. All oscillators must have positive feedback in order to work. The feedback determines the frequency of the note produced by the oscillator.

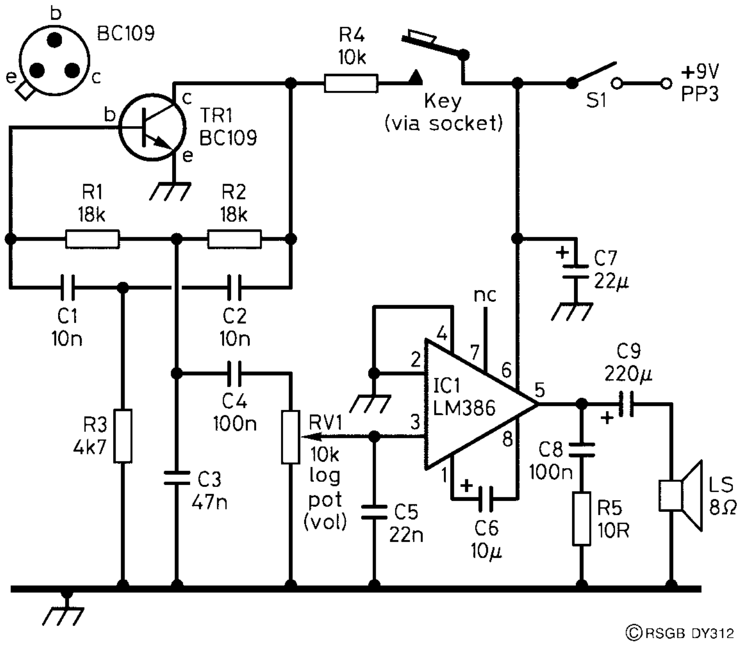
b
e
c
BC109
R4
10k
Key
(via socket)
S1
+9V
PP3
c
b
e
TR1
BC109
R1
18k
R2
18k
C1
10n
C2
10n
C7
22μ
nc
4
7
2
6
5
3
8
1
IC1
LM386
C9
220μ
+
+
+
R3
4k7
C4
100n
C3
47n
RV1
10k
log
pot
(vol)
C5
22n
C6
10μ
C8
100n
R5
10R
LS
8Ω
©RSGB DY312

Figure 1 Circuit diagram of the sinewave oscillator and amplifier

Here, the feedback circuit looks like two letter Ts. If you look at Figure 1, one T is formed by R1, R2 and C3, the other by C1, C2 and R3 – hence the name 'Twin-T'. Notice that the two Ts are connected in parallel between the collector and base of TR1, so any signal appearing at the collector is fed into both Ts. What emerges is then fed back into the base, producing in turn a signal at the collector. And so it goes on, producing a sine wave output.

The oscillator output is fed into an integrated circuit amplifier for output via a small loudspeaker.

Putting it together

The prototype was constructed on plain matrix board (the type *without* the copper strips), as shown in **Figure 2**. The components have their leads pushed through the holes in the board, and connections are made underneath.

Build the amplifier circuit first, using a socket for IC1. Connect the 9 V supply and touch pin 3 with an ordinary piece of wire. If a buzz is heard from the speaker, all should be well. If not, check your circuit and make changes until it does.

Build the oscillator circuit, and connect its output to the volume control VR1 via C4. Set VR1 half-way along its travel and switch on. A note should be heard from the speaker when the Morse key is depressed. The component values making up the twin-T determine the frequency of the note. Try varying them if you think your note is too high or too low. Whatever changes you make, either to the resistors or the capacitors, always ensure that R1 = R2 and that C1 = C2.

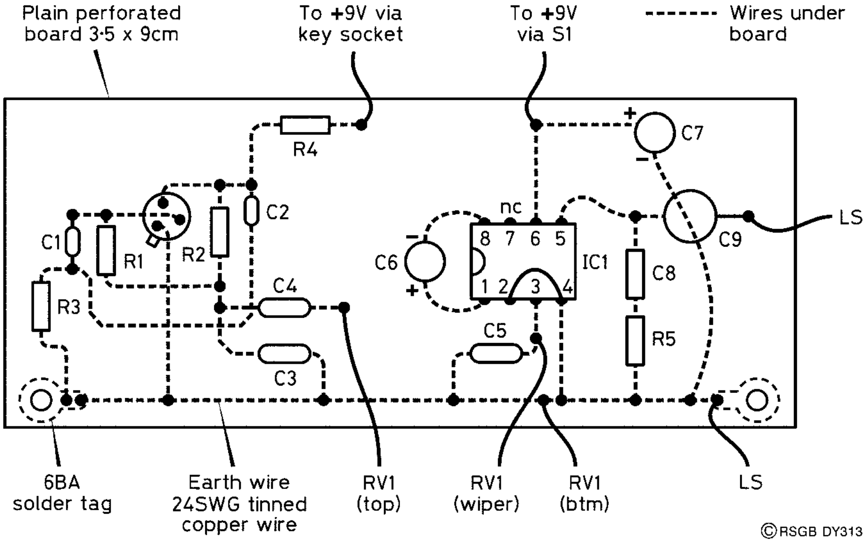
Plain perforated board 3·5 x 9cm
To +9V via key socket
To +9V via S1
------- Wires under board
R4
C7
+
−
C2
nc
LS
8 7 6 5
C9
C1
R1
R2
C6
IC1
C8
1 2 3 4
−
+
C4
R3
C5
R5
C3
6BA solder tag
Earth wire 24SWG tinned copper wire
RV1 (top)
RV1 (wiper)
RV1 (btm)
LS
©RSGB DY313

Figure 2 Component layout and interconnection diagram

Parts list

Resistors: all 0.25 W carbon film

R1, R2	18 000 ohms (18 kΩ)
R3	4700 ohms (4.7 kΩ)
R4	10 000 ohms (10 kΩ)
R5	10 ohms (10 Ω)

Capacitors

C1, C2	10 nF (0.01 μF)
C3, C8	47 nF (0.047 μF)
C4	100 nF (0.1 μF)
C5	22 nF (0.022 μF)
C6	10 μF electrolytic, 16 V WKG
C7	22 μF electrolytic, 16 V WKG
C9	220 μF electrolytic, 16 V WKG

Semiconductor

TR1	BC109

Integrated circuit

IC1	LM386

Additional items

VR1	10 kΩ log potentiometer with switch
LS	8 Ω loudspeaker

Matrix board 3.5 × 9 cm
Small jack socket for key input
Box
PP3 battery and clip

81 A simple 6 m beam

Introduction

The attraction of building your own aerials is an abiding feature of our hobby. You can *buy* almost any shape or size of aerial, but one you have made yourself can often work every bit as well as a commercial device costing ten times as much.

The design

This aerial, designed for use on the 6 m band, is essentially a two-element Yagi, with the elements bent in order to reduce the physical size. It is known

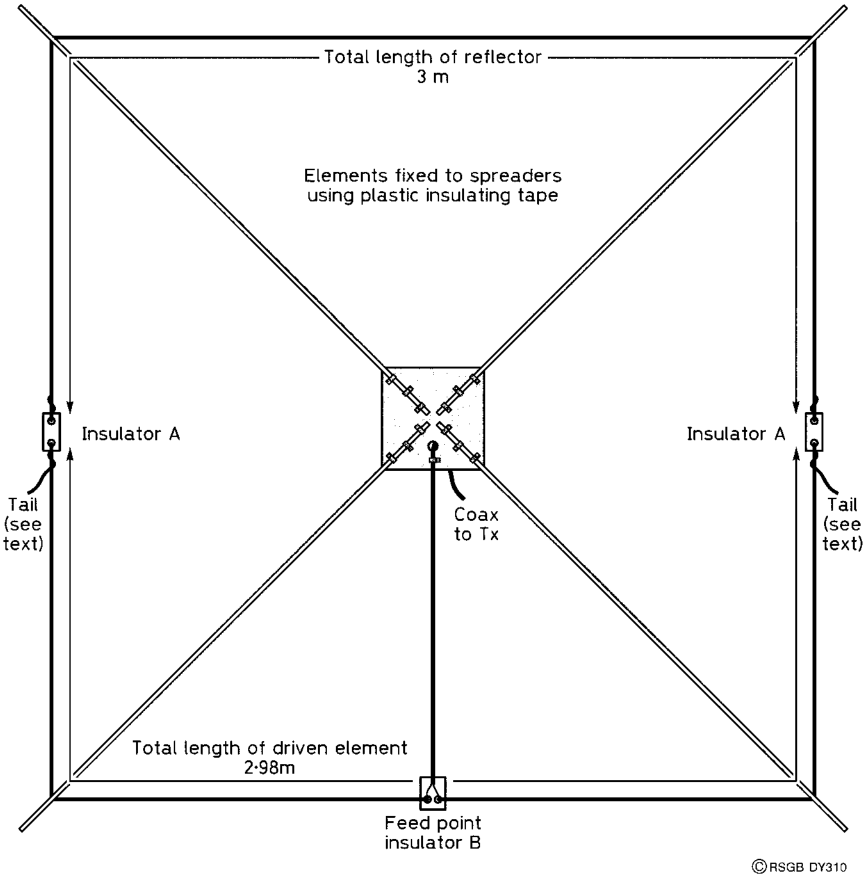
Total length of reflector
3 m
Elements fixed to spreaders
using plastic insulating tape
Insulator A
Insulator A
Tail
(see
text)
Tail
(see
text)
Coax
to Tx
Total length of driven element
2·98m
Feed point
insulator B
©RSGB DY310

Figure 1 Plan view of the complete 50 MHz VK2ABQ antenna

as the VK2ABQ beam, and was designed originally for the 20, 15 and 10 m bands, principally because of its space-saving qualities. It is made using a wooden frame and wire elements, and is ideal for portable operation.

Tools ready?

The beam is shown in **Figure 1**. The driven element is the one whose centre is fed by the coaxial cable, and lies between the two insulators marked A and the feedpoint at B. The reflector is also anchored at the points A, and lies over the upper half of the frame.

The wooden centrepiece is used to support the cross-pieces and to mount the aerial on the mast, using a common shelf bracket. The cross-pieces, known as *spreaders*, can be wooden canes or dowelling, and are mounted to the centrepiece using cable clips and adhesive. **Figure 2** shows how this is done.

If the aerial is to be a permanent installation, the spreaders should be weatherproofed using a good-quality exterior varnish. The wire elements are PVC covered and fed through holes in the spreaders.

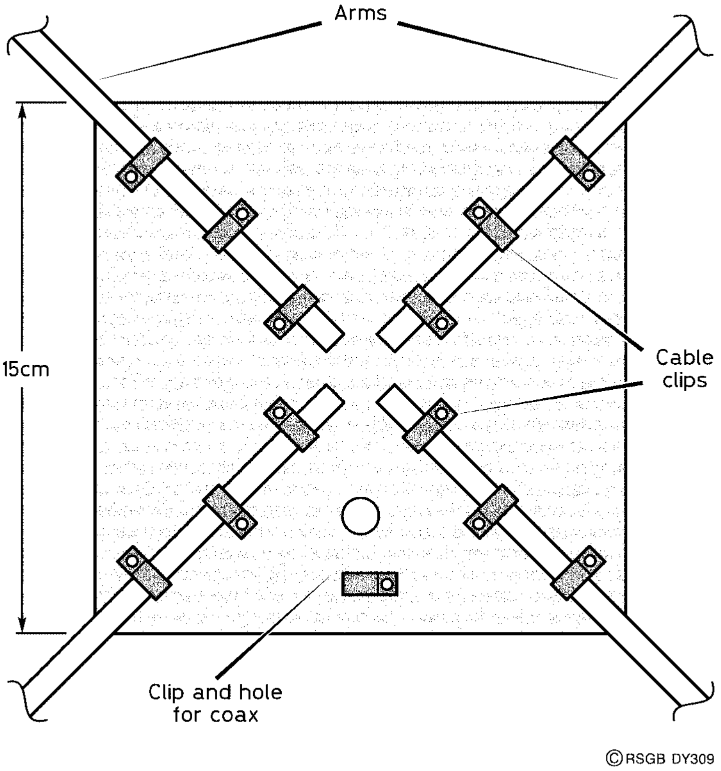

Figure 2 Centre support piece

The end insulators are made of drilled perspex, and the wire passed through the two holes and twisted, as shown in **Figure 3**.

Note: if you have not drilled holes in perspex before, take care! The drill bit must be well-lubricated because it generates a lot of heat (enough to melt the perspex and jam the drill). Never turn the drill for more than a few seconds at a time, and moisten the bit in between. Start with a pilot hole and use bits of increasing size until the hole is the size you want.

Another perspex insulator is used to secure the feeder to the aerial, as shown in Figure 3. The feeder then passes directly to the centrepiece, where it is fastened with a cable clip and then passes down the mast.

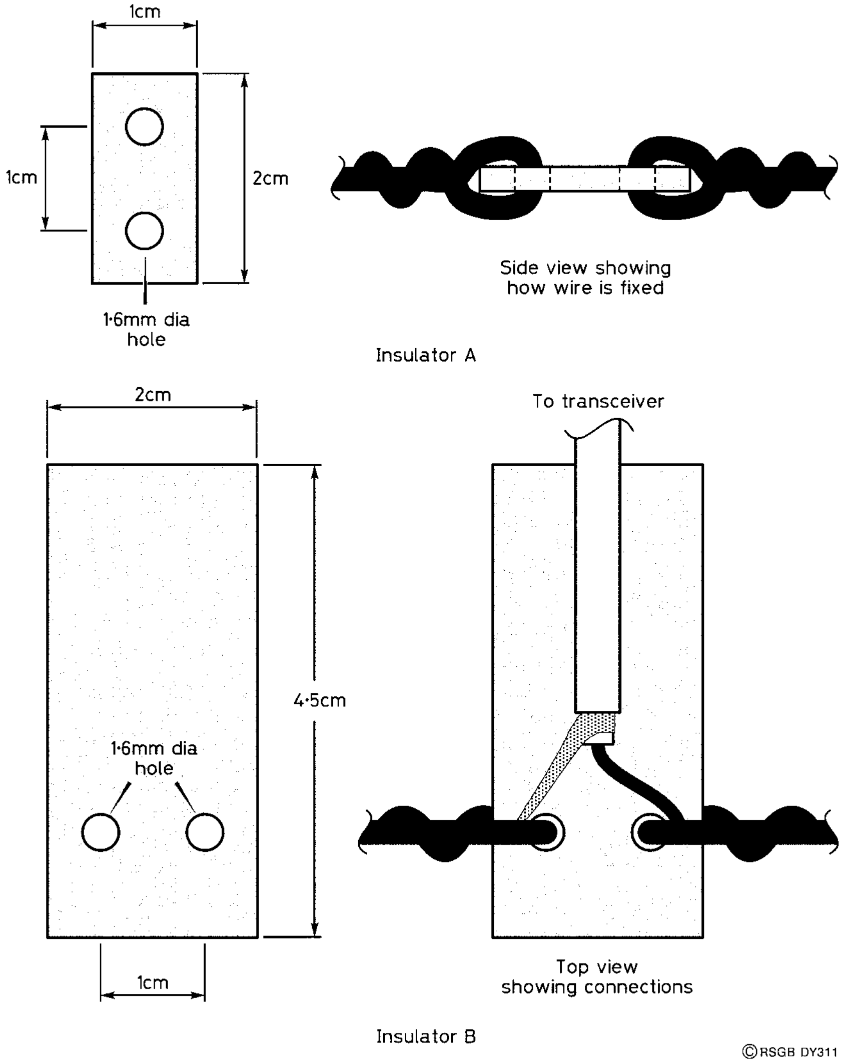
1cm
1cm
2cm
1·6mm dia hole
Side view showing how wire is fixed
Insulator A
2cm
To transceiver
4·5cm
1·6mm dia hole
1cm
Top view showing connections
Insulator B
©RSGB DY311

Figure 3 Details of insulators

Adjustment

The driven element (A–B–A) in Figure 1 is fixed to the end insulators in such a way as to have ‘pigtails’ which are about 10 cm long. Using an SWR meter between the aerial and the transmitter, trim the pigtails equally at each end for minimum SWR in that portion of the 6 m band in which you plan to work. Make sure that the transmitter is off when you trim the ends, as high voltages can be present there. Always listen on the frequency before you transmit and, when you do, ask if the frequency is in use and identify yourself. Use as little power as possible.

The prototype had an SWR of 1.2 at 50.2 MHz and performed well. If you look at Figure 1, which is a view of the aerial from above, you will immediately see that when the aerial is horizontal, it radiates with horizontal polarisation in a direction from the top of the page to the bottom. A metal pole or mast can be used for horizontal polarisation, but if you intend to use the aerial for vertical polarisation, it is much better to use a wooden or fibreglass pole.

Portable use

If you plan to operate portable with this aerial, the only real modifications you need are to the centrepiece and how it supports the spreaders. Instead of using glue and cable clips, nuts and bolts through the spreaders and centrepiece would allow the spreaders to be ‘hinged’ closed for transport.

Materials

Centrepiece	Hardwood, 15 × 15 × 25 mm
4 spreaders	110 cm long (cane or 6 mm dowel)
Wire	1.5 mm PVC-covered copper
50 Ω coaxial cable	RG58 or similar
Cable clips	6 mm (12 off)
Varnish	Polyurethane for waterproofing
Tape	Self-amalgamating, to waterproof all soldered joints
Insulators	See text

82 An integrated-circuit amplifier

Introduction

A simple audio-frequency amplifier is a very useful building-block in many more advanced electronic circuits. It is also a very useful piece of test equipment. To have one spare in the shack can be a life-saver at times.

Planning

Get into the habit of planning your project. How do you want it to look when it is finished. Do you want it in a box? Do you want it 'open-plan', with all the components on view?

The 'minimalist' approach to any project is simply a front panel and a baseboard, on to which all the components are fitted. In this case, the front panel would accommodate the loudspeaker, the volume control and the input and output jack sockets, and the baseboard would support the circuit board. Once you have decided these things, and know the size of the board, speaker and volume control, you can decided how big the panel and baseboard should be.

The amplifier

Instead of building the amplifier from discrete components (i.e. transistors), as was done in the project *An audio-frequency amplifier* (which you will find elsewhere in this book), we are going to use an integrated-circuit (IC) amplifier. External resistors and capacitors are still needed but, compared with the number of components *inside* the chip, these are very few indeed!

The IC we are going to use is the TBA820M. The circuit may look complicated, but with the use of a matrix circuit board it becomes quite simple. **Figure 1** shows the circuit diagram and **Figure 2** the layout on the board. The external connections to the PCB are shown in **Figure 3**.

To avoid having a separate switch to switch off the power supply when the amplifier is not being used, a volume control which incorporates a switch is used.

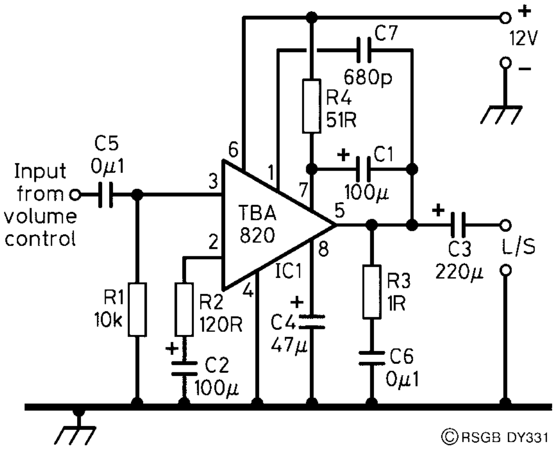
+
12V
–
C7
680p
R4
51R
C5
0µ1
Input
from
volume
control
6
3
1
7
+
C1
100µ
TBA
820
5
2
8
IC1
4
+
C3
220µ
L/S
R1
10k
R2
120R
+
C4
47µ
R3
1R
+
C2
100µ
C6
0µ1

©RSGB DY331

Figure 1 Audio amplifier, circuit diagram

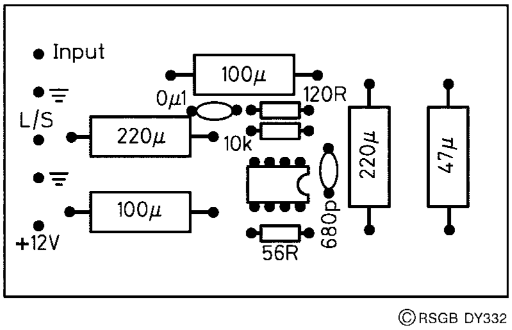
Input
L/S
+12V
100µ
0µ1
120R
220µ
10k
100µ
56R
680p
220µ
47µ

©RSGB DY332

Figure 2 Audio amplifier, component layout

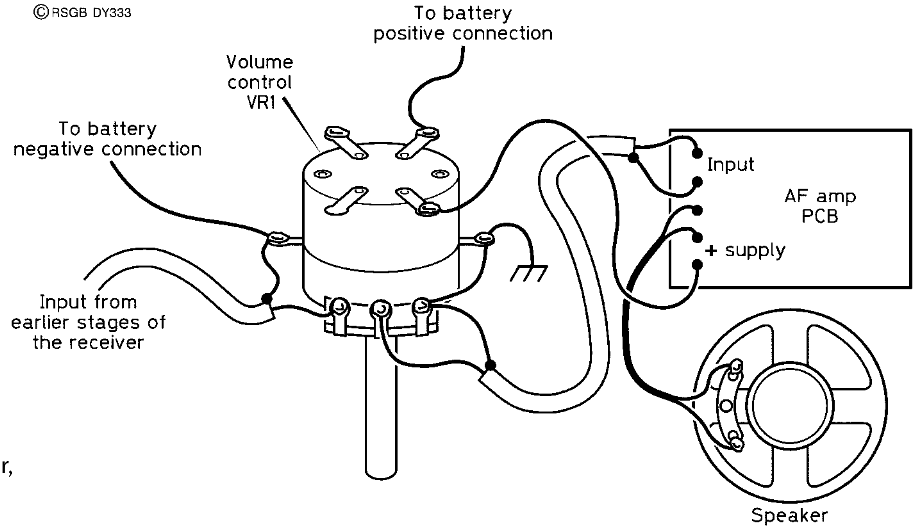
To battery
positive connection
Volume
control
VR1
To battery
negative connection
Input
AF amp
PCB
+ supply
Input from
earlier stages of
the receiver
Speaker

©RSGB DY333

Figure 3 Connecting the circuit board to the speaker, volume control/switch and battery

What power supply?

If the amplifier is to be used only for short periods, as a test instrument, for example, then it can be run from a PP3 battery, which can be mounted on the baseboard. If you intend to use it often, then a connection to an external power supply is preferable. For this option, you may want to consider fixing two terminals to the baseboard for this connection.

83 A novice ATU

Introduction

Having an aerial tuning unit (ATU) is always useful. It is used for adjusting the aerial impedance, as 'seen' by the transceiver, to be the same as that of the transceiver itself, usually 50 Ω. This process is called *matching*, and ensures that the receiver and the power amplifier (PA) stage of the transmitter work efficiently.

This design is due to the late Doug DeMaw, W1FB, and uses readily available components. It will handle up to about 5 W, and operates over the frequency range 1.8 to 30 MHz.

Circuit evolution

The basic circuit of one type of ATU is shown in **Figure 1**. On transmit, the input signal at L1 is coupled into L2, which forms a resonant circuit with C1. The signal across the resonant circuit is fed to the aerial by C2. The combination of L2 and C1 helps to remove signal harmonics, because they are not at the resonant frequency and are shunted to earth. On receive, only those signals which are within the pass-band of L2 and C1 will pass into the receiver. This improves receiver performance by rejecting many out-of-band signals which the simpler receiver doesn't like.

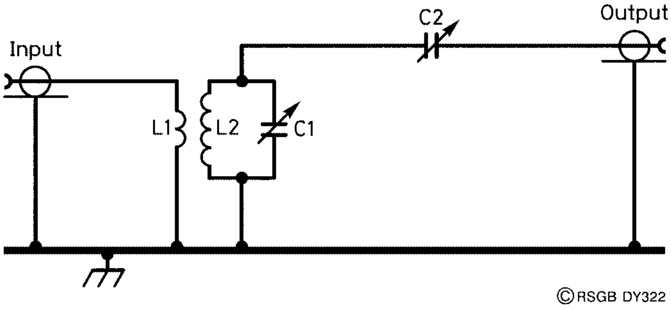

Figure 1 Basic circuit of the ATU

For this design, the all-important resonant frequency is chosen to be in the 80 metre band. If we want to make the circuit operate on other bands, we need to change either L2 or C1. It is easy to change L2 by using a set of switched inductors (L3–L5) as shown attached to S1 in **Figure 2**. Forget about S2 and L6 for the moment.

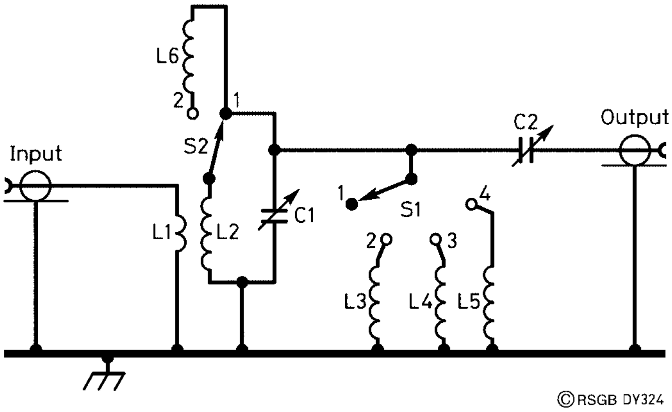

Figure 2 L6 added for 160 m

When S1 is in position 1, the circuit reverts to that of Figure 1. S2 should be brought into circuit only when S1 is in position 1. Looking at the way the circuit is drawn, you can see that when S2 is in position 2, we have two coils in series, which is equivalent to adding more turns to L2. More turns means more inductance, which lowers the resonant frequency still more, and gives coverage of the 160 m band.

Rather than having to remember to flick S2 to the correct position *and* move S1 to position 1 when we want to operate on 160 m, both functions can be combined into one rotary switch with two wafers. This circuit is shown in **Figure 3**.

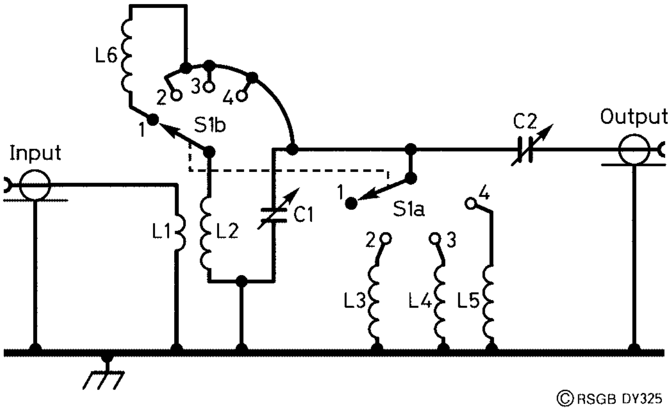
L6
2
3
4
1
S1b
C2
Output
Input
C1
1
S1a
4
L1
L2
2
3
L3
L4
L5
©RSGB DY325

Figure 3 Two-pole switch for all bands

In all positions of S1a *except* position 1, the extra coil, L6, is shorted out. The two halves of the switch, S1a and S1b, move together, as the shaft is turned, so both halves are in position 1 at the same time, position 2 at the same time, and so on. The two switches, S1 and S2, are said to be *ganged*.

Construction

L1 is formed by winding four turns of 22 SWG enamelled copper wire over L2, as in **Figure 4**. L2 already exists on the purchased former. After scraping the enamel off the ends of the wire (with a sharp knife or sandpaper), one end of L1 must be soldered to one end of L2, as shown in Figure 4. The free end of L1 goes to the transceiver aerial socket.

All components except the capacitors are assembled on the switch. Note that the rotor of C1 is earthed, but neither side of C2 is earthed. This means that the metal shaft of C2 is not earthed, and touching it will detune the ATU. Using a plastic knob for C2 will minimise this effect. If you decide to use a metal box, take precautions to ensure that no part of C2 is in electrical contact with the box.

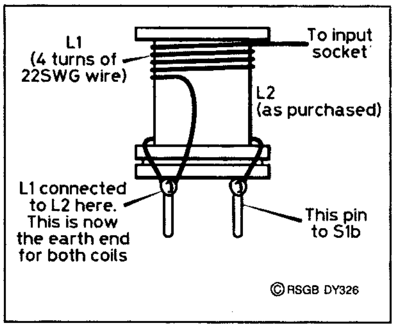
L1
(4 turns of
22SWG wire)
To input
socket
L2
(as purchased)
L1 connected
to L2 here.
This is now
the earth end
for both coils
This pin
to S1b
©RSGB DY326

Figure 4 L1 is wound over L2 as shown

The capacitors are mounted using M2.5 screws. Make sure that the screws do not foul the vanes of the capacitor. If your screws are too long, a few washers between the box and the capacitor will solve the problem!

In use

The best indication of a good match is obtained with a standing-wave meter between the ATU and the transceiver. The controls are adjusted alternately to ‘feel’ your way to a better and better match.

For receive-only use, the same alternate adjustments are used, watching for the maximum signal strength on the S-meter or, for a very weak signal, making the signal from the loudspeaker as large as possible.

Parts list

Capacitors	
C1, C2	350 pF variable
Inductors	
L1	See text
L2	27 µH
L3	10 µH
L4	2.2 µH
L5	1 µH
L6	65 µH
Additional items	
SO239 sockets (2 off)	
S1	2-pole 6-way rotary
Box	as required
Plastic knobs (2 off)	
Stick-on feet (4 off)	
M2.5 screws for capacitors	
Screws, nuts and washers for mounting the input and output sockets.	

84 CW QRP transmitter for 80 metres

Introduction

This is a relatively simple transmitter design having an output of 1 W. The design is not new, having been described before in other amateur radio publications. The components are all available new and the total cost should not exceed £15.

The circuit

Like other simple transmitters (see *An 80 Metre Crystal-Controlled CW Transmitter* and *A Breadboard 80 m CW Transmitter* elsewhere in this book) this one is crystal controlled. This assures frequency stability, but limits the usefulness of the transmitter. The key to increased frequency coverage without a conventional Variable Frequency Oscillator (VFO) is the use of a low-cost 3.58 MHz *ceramic resonator*. The 'pulling' range of a 3.58 MHz ceramic resonator covers the UK novice 80 m sub-band and some of the CW segment below 3.525 MHz.

A ceramic resonator is like a crystal, but not quite as stable in frequency. Its main advantage is its large pulling range.

The block diagram is shown in **Figure 1**. It is very similar to a crystal-controlled transmitter, and includes an oscillator, buffer and final amplifier.

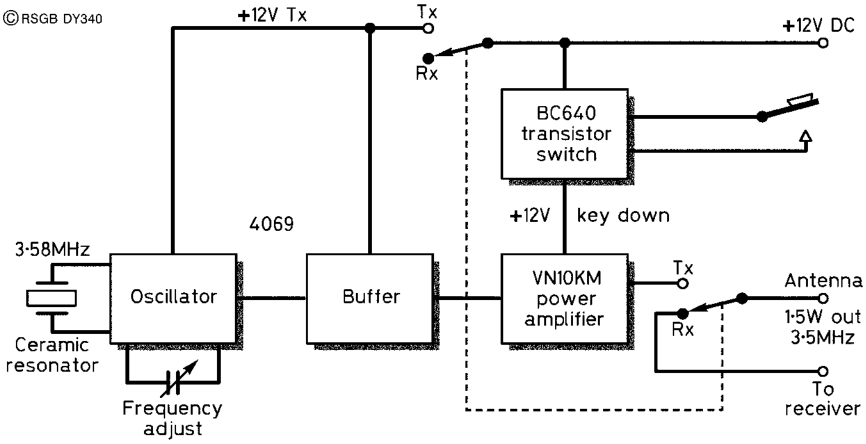
©RSGB DY340
+12V Tx
Tx
Rx
+12V DC
BC640 transistor switch
+12V
key down
4069
3·58MHz
Ceramic resonator
Oscillator
Buffer
VN10KM power amplifier
Tx
Rx
Antenna
1·5W out 3·5MHz
To receiver
Frequency adjust

Figure 1 Block diagram of the ceramic resonator controlled 80 m CW transmitter

This amplifier is keyed, the oscillator remaining running all the time. This improves frequency stability because the oscillator is not being continuously stopped and started by the key. It is switched off while receiving, though, to avoid interference with the received signal. Transmit/receive switching is accomplished by a panel-mounted switch controlling both the aerial, oscillator and buffer switching.

Figure 2 shows the transmitter circuit diagram. An unusual aspect of this transmitter is the use of a digital CMOS integrated circuit (IC) type 4069 for the buffer and oscillator stages. The IC houses six inverters, four of which are used in the circuit. One is used as the oscillator, two are used for the buffer stage, and the fourth provides an output for a direct-conversion receiver, should one be added at a later date.

The frequency of the oscillator is changed by varying the capacitance in the ceramic resonator circuit. This is provided by VC1.

The power amplifier (PA) is a small MOSFET (metal oxide semiconductor field-effect transistor), TR1. This is capable of providing an output power of 2 W but, in this circuit, it is run conservatively to give 1.5 W. The output can be varied by changing the resistance (R5 + R6) in the gate circuit. Attempts

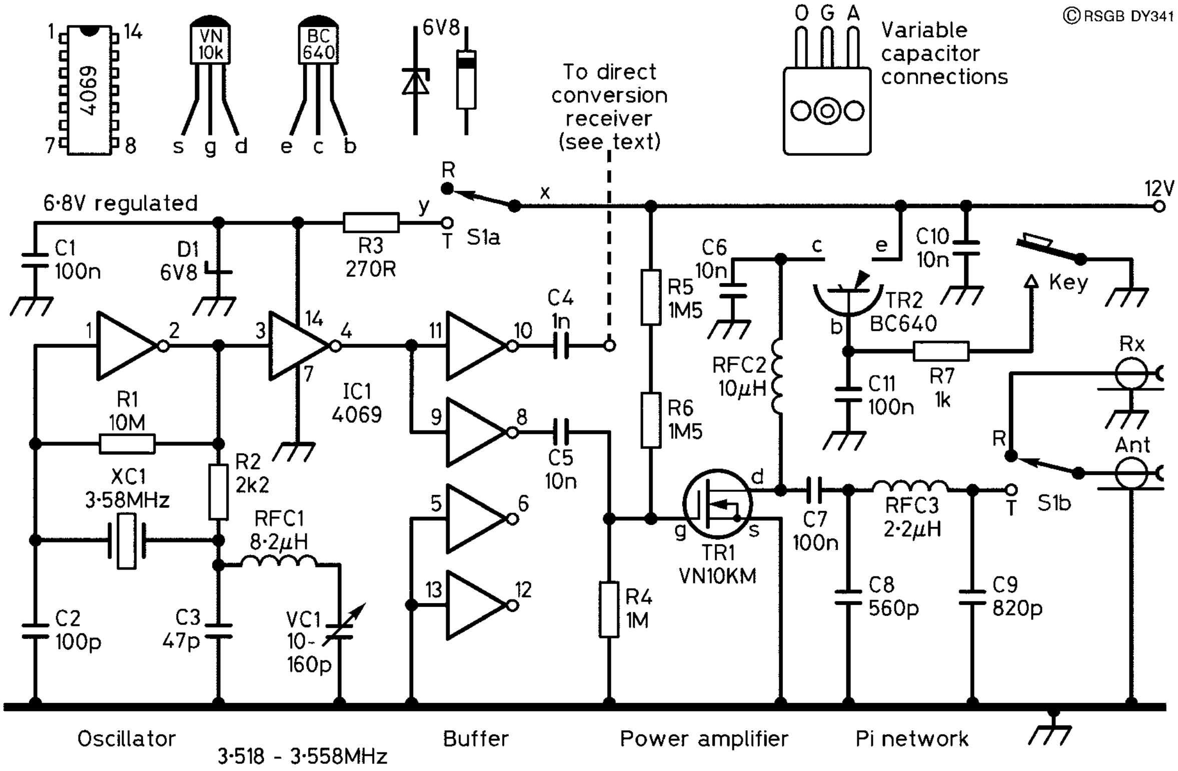
©RSGB DY341
4069
1
14
7
8
VN 10k
s g d
BC 640
e c b
6V8
To direct conversion receiver (see text)
O G A
Variable capacitor connections
6·8V regulated
C1 100n
D1 6V8
R3 270R
y
T
S1a
R
x
12V
C6 10n
c
e
C10 10n
Key
R5 1M5
C4 1n
TR2 BC640
b
R7 1k
Rx
IC1 4069
RFC2 10μH
C11 100n
R1 10M
R6 1M5
R
Ant
XC1 3·58MHz
R2 2k2
C5 10n
d
C7 100n
RFC3 2·2μH
T
S1b
RFC1 8·2μH
g
s
TR1 VN10KM
C2 100p
C3 47p
VC1 10-160p
R4 1M
C8 560p
C9 820p
Oscillator
3·518 - 3·558MHz
Buffer
Power amplifier
Pi network

Figure 2 Circuit diagram of 80 metre transmitter

to raise the output power by decreasing the values of these resistors may result in immediate MOSFET failure.

A pi-network (C8, RFC, C9) provides impedance matching to 50 Ω, together with harmonic suppression. Like all inductors in this transmitter, the pi-network inductor is a pre-wound RF choke. A pi-network is so called because the components are arranged in the shape of the Greek letter pi (π).

Keying is carried out by a pnp transistor switch, TR2. Closing the key earths the base and supplies 12 V to the collector of TR2 and to the drain of the MOSFET, TR1, allowing the PA to operate.

Construction

You must house your transmitter in a metal box, to avoid hand-capacity effects and the radiation of spurious frequencies. Size is not important, provided it is large enough to accommodate the transmitter without cramping the components. You may want to allow space for future additions such as a direct-conversion receiver, break-in keying, sidetone or a small power amplifier. A good size is 5 × 15 × 15 cm. You can make your own box, buy it, or even use a biscuit tin!

Front and rear panel connectors can be fitted first. The choice of these is a personal matter, but a good working choice would be:

(a) Power socket – 2.1 mm panel socket – centre pin positive.
(b) Key socket – ¼ inch jack socket.
(c) Aerial and receiver connectors – panel-mounting SO239 type.

Particular attention must be paid to the mounting of the variable capacitor, VC1. Make sure the hole for the shaft is amply big enough, and if you use screws to mount the capacitor on the front panel, then make sure they are not too long, otherwise they will touch the vanes of the capacitor! Mounting can be by means of glue, sparingly applied and kept well away from the shaft.

A board size of about 6 × 10 cm is adequate. Component layout on the board is suggested in **Figure 3**. The prototype used ordinary matrix board, which is preferable to stripboard for a design like this; stripboard has undue capacitance between adjacent strips. Component leads are fed through holes in the board and are connected underneath. Make sure that leads and connections are rigid because, if they can move, there is always the danger of short-circuiting, and capacitance changes.

To facilitate construction, servicing and testing, it is advisable to use Veropins for connections to the variable capacitor, transmit/receive switch,

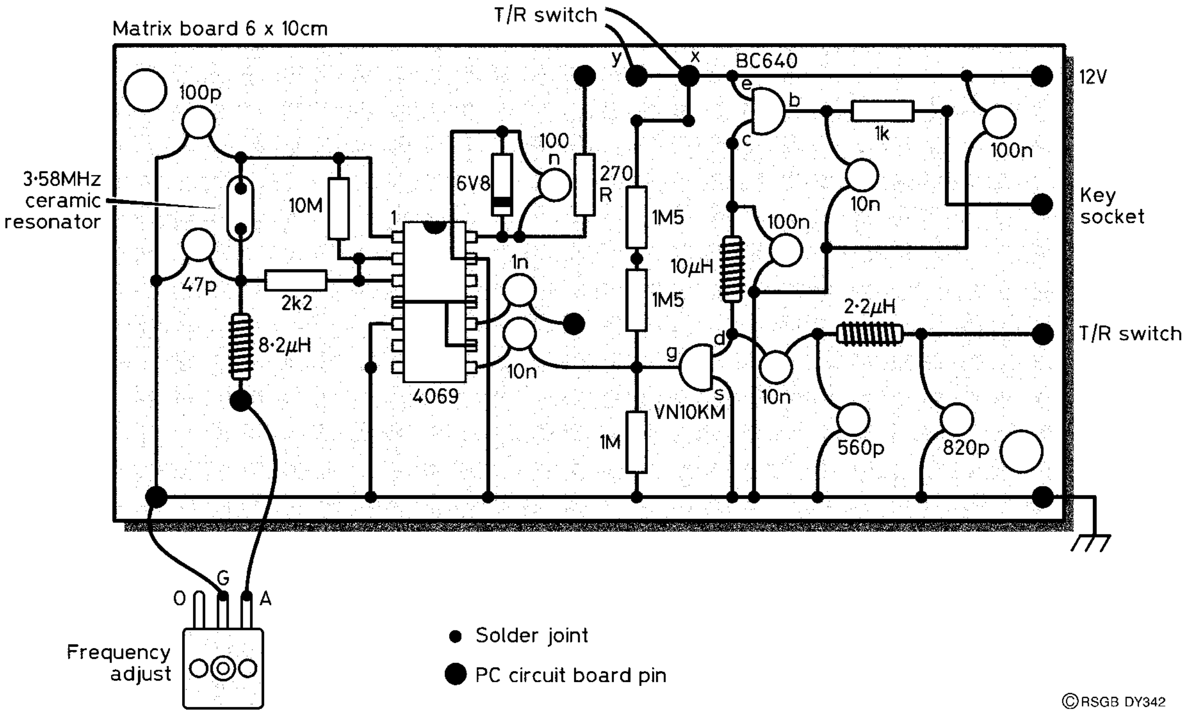
Matrix board 6 x 10cm
T/R switch
y
x
BC640
e
b
c
100p
3·58MHz ceramic resonator
10M
1
6V8
100 n
270 R
1M5
1k
100n
12V
10n
Key socket
100n
10μH
47p
2k2
1n
1M5
2·2μH
T/R switch
8·2μH
d
g
10n
s
4069
VN10KM
10n
1M
560p
820p
G
O
A
Frequency adjust
● Solder joint
● PC circuit board pin
©RSGB DY342

Figure 3 Component layout of the 80 metre transmitter. The transmit/receive switch is not mounted on the board and is not shown in this diagram

aerial and power sockets. Use screws and spacers to mount the circuit board to the box. Mounting the board horizontally assists troubleshooting.

Use a socket for the IC, and observe the CMOS handling precautions given in *An Electronic Die*, elsewhere in this book. When soldering the leads to the ceramic resonator, do it quickly – excessive heat damages the device. The earth lead running acoss the bottom of the board must be connected to the metal case by a short length of stout wire.

Testing

After carefully checking your wiring, both against the circuit diagram and the layout diagram, it is time to test your circuit. You will need a multimeter, an 80 m SSB receiver and a 50 Ω dummy load. A good design of dummy load can be found in the project *A Switched Dummy Load*, also in this book. An RF power meter and frequency counter will also be useful, although if your receiver has a digital frequency readout and S-meter, the latter two items are not really necessary. You will also need a 12 V 1 A power supply unit (PSU) to power the transmitter.

Switch the transmitter to receive and switch on the transmitter. No current should be consumed. Switch to transmit and check that pin 14 of IC1 is 6.8 V positive. With the dummy load connected to the aerial socket, press the key. The voltage on TR2 collector should now be 12 V, dropping to zero when the key is released.

Now check the operation of the oscillator. In transmit mode, you ought to be able to find a strong carrier signal with the receiver, even though the dummy load is connected. Adjusting the variable capacitor should change the frequency. At the lower end of the frequency range, you may find that the oscillator is unreliable in starting, because the circuit is attempting to pull the resonator too low in frequency. If this is the case, set the trimmer at the back of VC1 to minimum capacitance. If your version of VC1 has two trimmers, and you don't know which one to set, set them *both* to minimum capacitance. If there is still a problem, reduce the value of RFC1 to 6.8 or 4.7 μH.

In all probability, the unmodified circuit of Figure 2 will not require any of the changes outlined here. A coverage of 3.518 to 3.558 MHz should be possible, while preserving good frequency stability and reliable oscillation.

A signal probe (see *An RF Signal Probe*, elsewhere in this book) is useful for checking the operation of the oscillator and PA. Alternatively, an RF power meter or the receiver's S-meter can be used. With the PA running, the unit should draw between 200 and 300 mA. If TR1 becomes too hot to touch after a few seconds of transmitting, increase R5 or R6 to limit the transistor's heat dissipation. A small 6.3 V bulb connected across the aerial output is a simple way to check that the PA is working. An orange/white glow when the key is pressed is indicative of correct operation.

The final test is to monitor keying 'quality'. With your dummy load connected, press the key and listen to the note on the receiver's loudspeaker. Then operate the key, sending a string of dits, for example. What you hear should be free of chirps and clicks, as well as being stable in frequency. This test is sometimes better performed with no aerial connected to the receiver, thus preventing receiver overload and its associated plops. No problems should be encountered here.

Frequency tuning

A peculiarity of ceramic resonators is that, every now and again, their frequency changes abruptly by 100 Hz or so, then remains stable for some time. This is certainly noticeable in the received signal, but does not detract from the QSO and no characters are lost as a result. Try to keep the area around the ceramic resonator cool, to avoid temperature variations.

Parts list

Resistors: all 0.25 W, 5% tolerance

R1	10 megohms (MΩ)
R2	2200 ohms (2.2 kΩ)
R3	270 ohms (270 Ω)
R4	1 megohm (1 MΩ)
R5, R6	1.5 megohms (1.5 MΩ)
R7	1000 ohms (1 kΩ)

Capacitors

C1, C7	100 nF
C2	100 pF
C3	47 pF
C4	1 nF
C5, C6, C10	10 nF
C8	560 pF
C9	820 pF
VC1	10–160 pF variable

Inductors

RFC1	8.2 μH
RFC2	10 μH
RFC3	2.2 μH

Semiconductors

IC1	4069
TR1	VN10KM
TR2	BC640
D1	6V8 Zener

Additional items

- Matrix board (see text)
- 14-pin DIL socket
- Pointer knob
- Sockets (see text)

85 An audio booster for your hand-held

Introduction

The audio output from many hand-held transceivers and receivers usually leaves much to be desired, so this little amplifier was designed to increase the output at minimal expense.

All that is needed is a separate amplifier and bigger loudspeaker. This is accomplished using a single integrated circuit (IC), a few components, and a loudspeaker from the junk box. This circuit will enable the output from your hand-held to be heard easily in a car.

The circuit

This is shown in **Figure 1**. It uses only those components necessary to operate the IC amplifier. VR1 is the preset volume control, and varies the signal coming from the 'External speaker' jack on the hand-held before feeding it into the IC for amplification. C1 blocks any constant voltage present on the input.

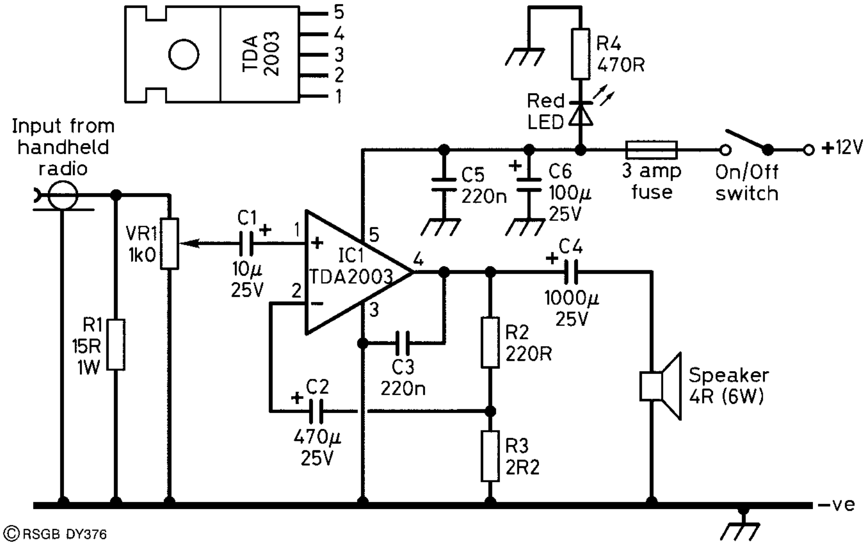
TDA 2003
5
4
3
2
1
R4
470R
Red LED
Input from handheld radio
+12V
C5 220n
C6 100μ 25V
3 amp fuse
On/Off switch
VR1 1k0
C1 10μ 25V
1
5
IC1 TDA2003
4
2
3
C4 1000μ 25V
R1 15R 1W
R2 220R
C3 220n
C2 470μ 25V
Speaker 4R (6W)
R3 2R2
−ve
©RSGB DY376

Figure 1 Circuit diagram of the audio amplifier. The power is derived from the cigarette lighter socket and the fuse is in-line with the lead

The IC output comes from pin 4 and is fed via the electrolytic capacitor, C4, to the loudspeaker. The circuit is provided with an on/off switch, fuse and LED to indicate when the circuit is switched on.

Construction

The box is made of aluminium. This is necessary to help to dissipate some of the heat generated by the IC. Do not build the circuit inside a plastic box unless you take special precautions! The IC has a metal mounting tab with a hole, specifically designed to be mounted to a metal box or other metal heat sink. Apply plenty of heat sink compound between the tab and the box, tighten the nut and bolt, and then wipe off any excess compound. The box will get slightly warm in operation.

The size of the box is not specified. You may want to decide on this when you find a loudspeaker. Choose one which will be able to handle 6 W

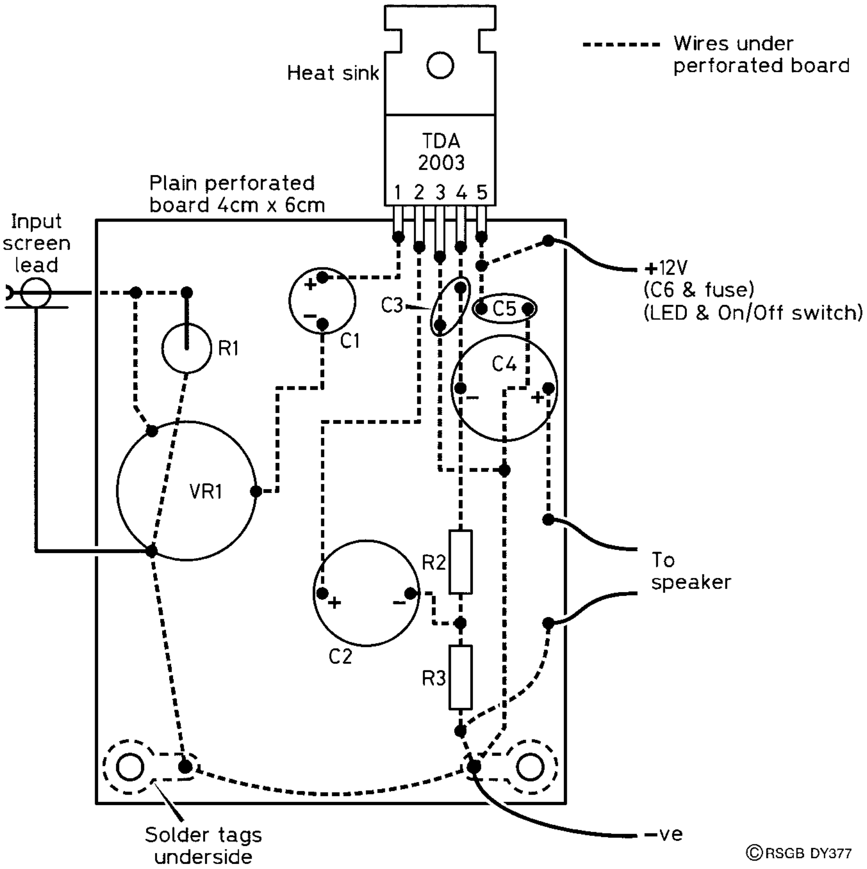
Heat sink
TDA 2003
1 2 3 4 5
------- Wires under perforated board
Plain perforated board 4cm x 6cm
Input screen lead
+12V (C6 & fuse) (LED & On/Off switch)
R1
C1
C3
C5
C4
VR1
R2
C2
R3
To speaker
Solder tags underside
−ve
©RSGB DY377

Figure 2 Layout of the components within the box

output. Drill all the holes in the box first. Holes for the speaker, input phono socket and the LED. The amplifier can be constructed on ordinary matrix board, which can be mounted inside the box with screws and spacers.

The layout is shown, for your guidance, in **Figure 2**. The components are mounted by pushing their leads through the holes in the board and making connections on the underside. The preset volume control, VR1, is set such that the hand-held's volume control is sufficient to control the final output over a good volume range. Use a screened lead from the 'External speaker' jack socket to the phono plug.

An external power supply is needed for this circuit. The normal dry battery which we usually use for small projects in this book will *not* work here, so you will need a proper mains power supply producing a stabilised 12 V. If you are going to use the unit principally in a car, then the cigar lighter socket can supply this voltage easily. Do make sure that the polarity is correct before you switch on!

When you plug the jack plug from your booster into the 'External speaker' socket on your hand-held, its internal speaker will be muted, so don't think that something dire has gone wrong! Adjust VR1 for a good volume range on your booster, when the volume control is turned on the hand-held.

Parts list

Resistors: all 0.25 W, 5% tolerance, unless otherwise stated

R1	15 ohms (15 Ω) 1 W
R2	220 ohms (220 Ω)
R3	2.2 ohms (2.2 Ω)
R4	470 ohms (470 Ω)
VR1	1000 ohms (1k Ω)

Capacitors

C1	10 μF 25 V
C2	470 μF 25 V
C3, C5	220 nF (0.22 μF) Mylar
C4	1000 μF 25 V
C6	100 μF 25 V

Semiconductor

LED	5 mm Red

Integrated circuit

IC1	TDA2003

Additional items
  Heat sink compound
  Nuts and bolts
  Loudspeaker 4 Ω 6 W
  3 A fuse
  On/off switch (SPST)
  Matrix board 4 × 6 cm
  Solder tags
  Plugs and screened cable for connecting lead
  Aluminium box

86 A grid dip oscillator

Introduction

When an inductor and a capacitor are connected, whether in series or parallel, they form a circuit with a natural (or *resonant*) frequency. The circuit stores energy, and this energy is being constantly shifted from the inductor to the capacitor and back again.

The dip oscillator is a simple instrument used to measure the resonant frequency of a tuned circuit without having to make any direct connection to the circuit. The circuit is more commonly known as the *grid dip oscillator* (GDO), from the days when the active device in the circuit was a valve. The FET or Field-Effect Transistor operates in a way which is very similar to that of the valve, so it is not quite a misnomer to call this instrument a grid dip oscillator, too.

The circuit

The GDO uses a calibrated, tunable FET oscillator in the circuit of **Figure 1**. It has a frequency range of 1.6 to 35.2 MHz in four ranges using a set of plug-in coils, shown in **Figure 2**. When the oscillator coil, L1, is placed near an external resonant circuit, some of its RF energy is coupled into the external circuit. A gain in energy of the external circuit must mean a loss of energy in the GDO circuit, resulting in a change of current through TR1, which is measured by the meter, M1.

The current through TR1 is of the order of 5 to 8 mA, but the change of current may be only a few microamps. To measure a very small

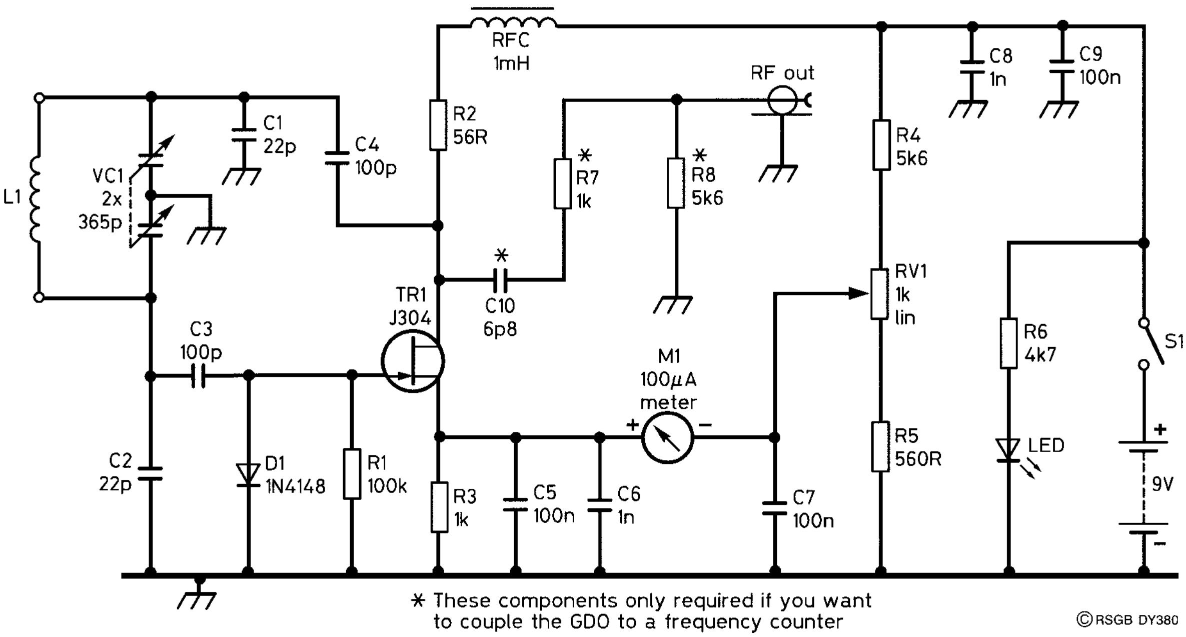
RFC
1mH
RF out
C8
1n
C9
100n
L1
VC1
2x
365p
C1
22p
C4
100p
R2
56R
*
R7
1k
*
R8
5k6
R4
5k6
*
C10
6p8
TR1
J304
RV1
1k
lin
R6
4k7
S1
C3
100p
M1
100μA
meter
+
−
C2
22p
D1
1N4148
R1
100k
R3
1k
C5
100n
C6
1n
C7
100n
R5
560R
LED
+
9V
−
* These components only required if you want
to couple the GDO to a frequency counter
©RSGB DY380

Figure 1 Circuit of an FET GDO. The coils are wound on DIN speater plugs, which provide both a plug-in base and a coil former

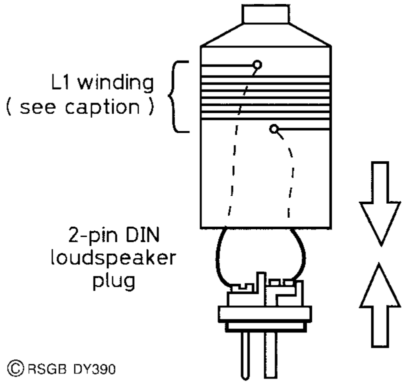
L1 winding
(see caption)
2-pin DIN
loudspeaker
plug
©RSGB DY390

Figure 2 Details of coil construction:
Range 1: 1.6–4.0 MHz 55 turns of 30 SWG
Range 2: 3.3–7.9 MHz 27 turns of 30 SWG
Range 3: 6.3–4.0 MHz 14 turns of 26 SWG
Range 4: 11.9–4.0 MHz 7 turns of 24 SWG.

change superimposed on a much larger standing current, the method of *offset* can be used.

One connection to the meter goes to the source of the FET, while the other goes to a variable offset voltage set by VR1. M1 has a full-scale deflection (FSD) of 100 μA. If the current through TR1 changes, the voltage across R3 changes. When there is no resonance, the voltage at the wiper of VR1 is set to be very slightly greater than that across R3, and there is a 75% FSD meter deflection. When the voltage across R3 decreases very slightly, due to external circuit resonance, a significant 'dip' in the meter deflection is produced, hence the name of the instrument.

The circuit is not difficult to make on standard matrix board. Provided you can follow a circuit and translate it into a good component layout, then this project is probably only an evening's work.

The most important part of the GDO is the tuning capacitor and its associated frequency-calibrated dial. New, air-spaced tuning capacitors can cost you up to £20, so it is worth delving around in junk boxes, or scouring the tables at a local boot sale or rally. The tuning capacitor from an old transistor radio should be perfect. It may even have a slow-motion drive and a dial which can be remarked for the project.

Choose a coil plug and socket arrangement that is practical. Think about crystal holders or phono plugs and sockets. The prototype shown in Figure 1 and Figure 2 used 2-pin DIN plugs, with the coil wound on the outside of the plastic plug cover. Figure 2 shows the coil construction and the winding details.

If you use a variable capacitor, VC1, with a value different from that shown in the parts list, then the frequency ranges will be different. This does not matter, as it will be taken into account during calibration.

Position VC1 so that the dial will be easy to see and to operate, while locating the coil socket as close to it as possible. **Figure 3** shows the traditional layout of the GDO.

Calibration

Because the GDO also radiates a very small amount of energy, a general coverage receiver can be used to calibrate the dial. Don't try to aim for great accuracy and clutter the dial with marks and figures! If you include C10, R7 and R8 in your circuit, you can connect a frequency counter directly to the GDO and leave it in circuit all the time.

The GDO in use

Always try to place the external tuned circuit with its coil coaxial with the plug-in coil, as shown in Figs 1 and 3. If the coils are at right angles, the GDO may not produce any resonance. Set the offset control to give about 75% FSD and slowly tune L1 through its whole range. If no dip occurs, you may have the wrong coil plugged in. When you eventually find a dip, move the external coil further away until only a minute dip is seen. You may have to retune the dip meter as you do this, but it gives a much more accurate reading of frequency.

Remember that you cannot 'dip' a coil by itself – there must always be a capacitor present.

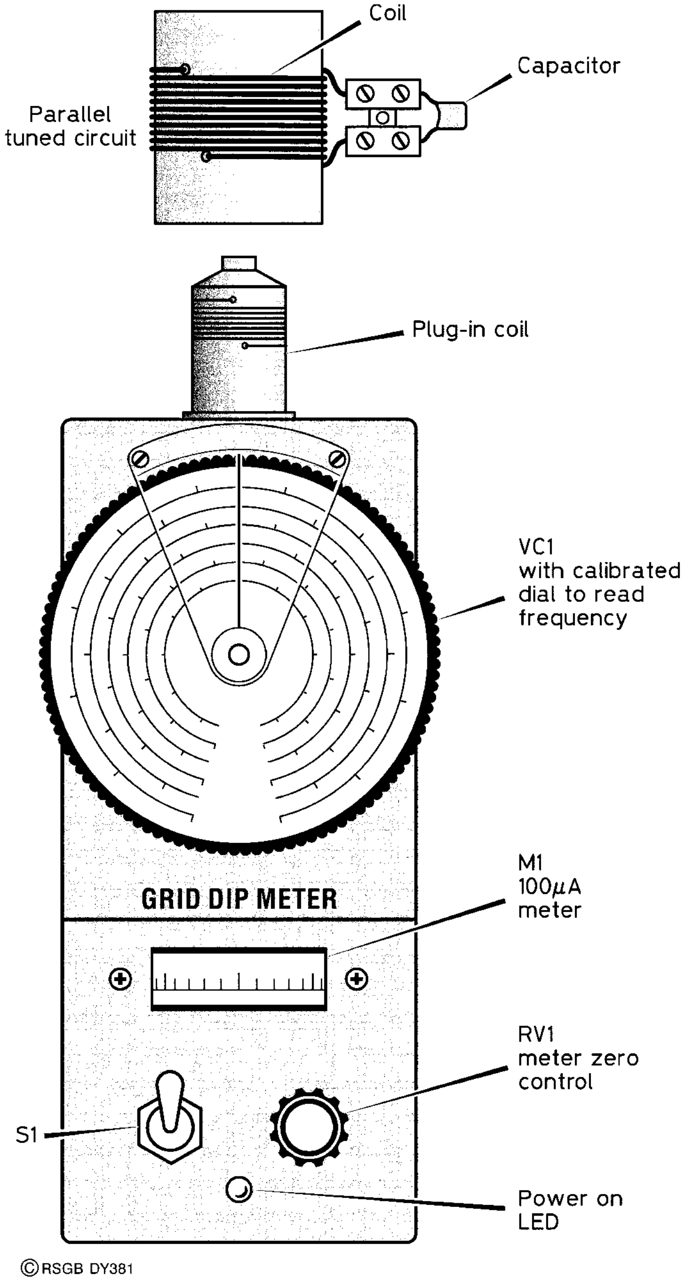
Coil
Capacitor
Parallel tuned circuit
Plug-in coil
VC1 with calibrated dial to read frequency
GRID DIP METER
M1 100μA meter
RV1 meter zero control
S1
Power on LED
©RSGB DY381

Figure 3 Layout of a typical GDO. The dial, meter and the location of the coil to the circuit under test can all be viewed at the same time. It is shown measuring resonance of a tuned circuit

Aerial resonance

Figure 4 shows how to check the resonance of a dipole aerial. Disconnect the coax feed at the aerial and place a short piece of wire, terminated with crocodile clips, across the centre insulator to short together the two ends of the aerial. By placing the GDO close to the shorting link, a dip should be seen on the meter while VC1 is turned. Alternatively, a loop in the element can be made around the coil, as in Figure 4, or the shorting link can be made long enough to loop over the coil. The latter method does not require tampering with the mounting and tensioning of the dipole wires.

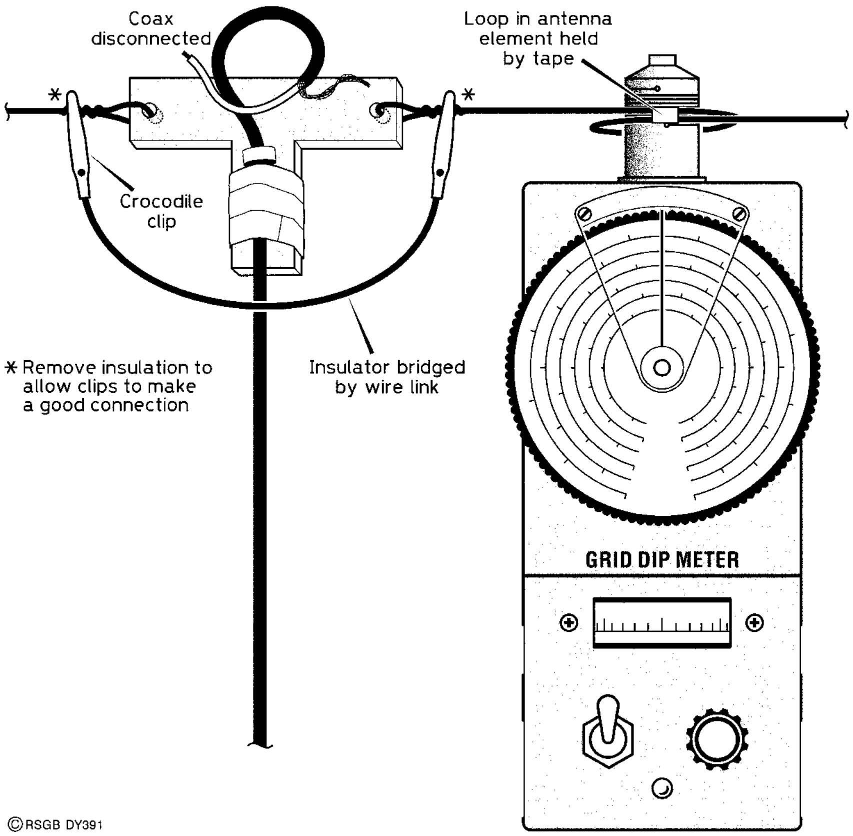
Coax
disconnected
Loop in antenna
element held
by tape
*
*
Crocodile
clip
∗ Remove insulation to
allow clips to make
a good connection
Insulator bridged
by wire link
GRID DIP METER
©RSGB DY391

Figure 4 A wire antenna element can be looped into a single turn coil for increased coupling to the GDO

Parts list

Resistors: all 0.25 W, 5% tolerance

R1	100 000 ohms (100 kΩ)
R2	56 ohms (56 Ω)
R3	1000 ohms (1 kΩ)
R4	5600 ohms (5.6 kΩ)
R5	560 ohms (560 Ω)
R6	4700 ohms (4.7 kΩ)
R7	1000 ohms (1 kΩ)
R8	5600 ohms (5.6 kΩ)
VR1	1000 (1 kΩ) linear

Capacitors

C1, C2	22 pF
C3, C4	100 pF
C5, C7, C9	100 nF (0.1 μF)
C6, C8	1 nF (1000 pF)
C10	6.8 pF
VC1	2 × 365 pF

Semiconductors

TR1	J304 or similar
D1	1N4148
LED	

Additional items

L1	See Figure 2
RFC	1 mH
S1	SPST
M1	100 μA

Source

Components are available from Maplin.

87 A CW transmitter for 160 to 20 metres

Introduction

This very small transmitter is designed to work on any band from top band (160 m) to 20 m, with an RF output of 1 W. It will work on higher frequencies but with a reduced output.

The circuit

The three-transistor circuit is shown in **Figure 1**. It comprises a crystal oscillator using a BC182 transistor which drives a 2N3866 power amplifier (PA) keyed by a ZTX750 PNP transistor. The oscillator and PA are coupled by a capacitor and resistor; this provides a very small amount of positive bias to the PA.

The oscillator can be used as a basic crystal oscillator but, by including a variable series capacitor as shown in Figure 1, the crystal frequency can be 'pulled' slightly, making the oscillator a *variable crystal oscillator* (VXO).

Construction

The PCB layout is shown in **Figure 2** and in the photograph. Although the prototype was built around a PCB, this circuit is equally amenable to

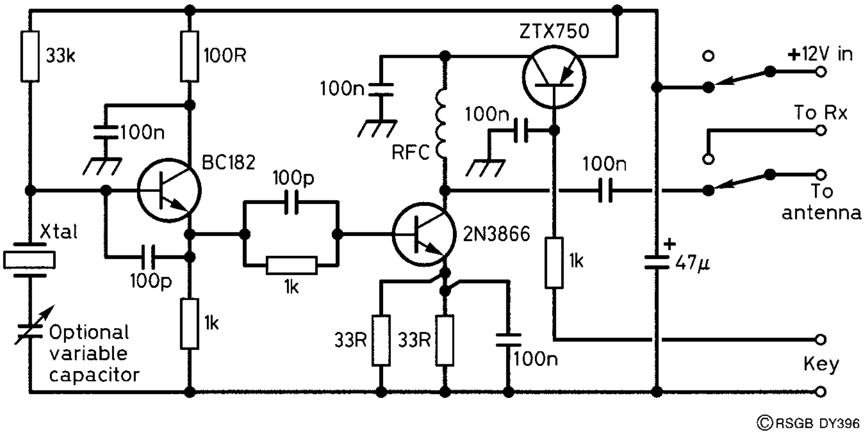
33k
100R
ZTX750
+12V in
100n
100n
To Rx
100n
BC182
RFC
100n
100p
To antenna
Xtal
2N3866
1k
47μ
100p
1k
Optional variable capacitor
1k
33R
33R
100n
Key
©RSGB DY396

Figure 1 Transmitter circuit diagram

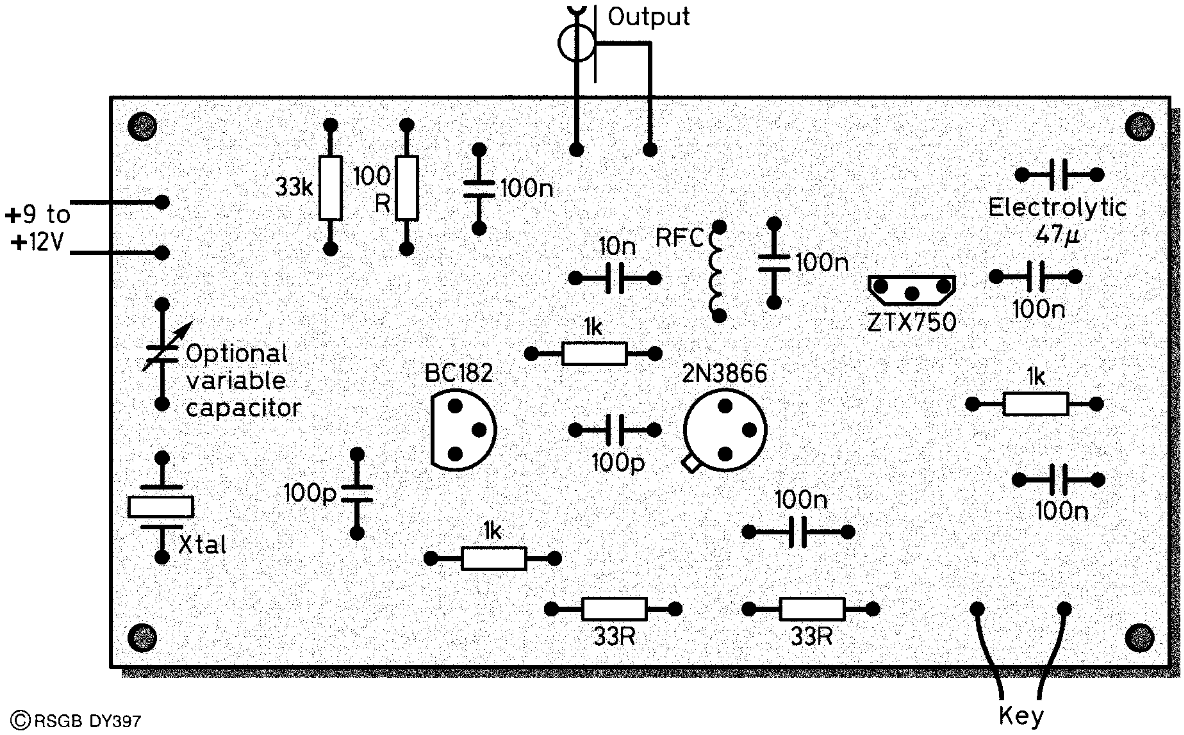
Output
+9 to
+12V
33k
100
R
100n
10n
RFC
100n
Electrolytic
47μ
ZTX750
100n
1k
Optional
variable
capacitor
BC182
2N3866
1k
100p
Xtal
100p
1k
100n
100n
33R
33R
Key
©RSGB DY397

Figure 2 Component placing on the PCB. The external connections are also shown

construction on a matrix board. Populating the PCB is very simple, and you can expect to be able to do this in about one hour. The radio-frequency choke (RFC) is made by winding 10 turns of 33 SWG wire on a ferrite bead. The enamel coating of the wire is intended to vaporise when soldered into the board, thus obviating the need to remove the enamel manually with a knife or sandpaper. However, if you do have problems with the PA either not working or keying intermittently, it is suggested that you investigate the RFC connections immediately!

If you decide to make the VXO version, you will have to cut the track between the crystal and earth, and connect the variable capacitor (250 pF) across the break.

In use

After performing the usual checks on the accuracy of your circuit building and the wiring of the external components, it is time to connect a 12 V battery between the points shown in Figure 2. Do not switch on yet. An aerial needs to be connected to the output via an ATU and a crystal, of frequency matching that of the aerial, fitted. The variable capacitor should give you a tuning range from about 14.058 to 14.064 MHz. 'Netting' (the process of tuning your transmitter to the same frequency as that of a received station) is achieved simply, because the oscillator is always running, and the leakage of the signal (despite the fact that the PA is not powered) is sufficient to bring the two frequencies to zero beat.

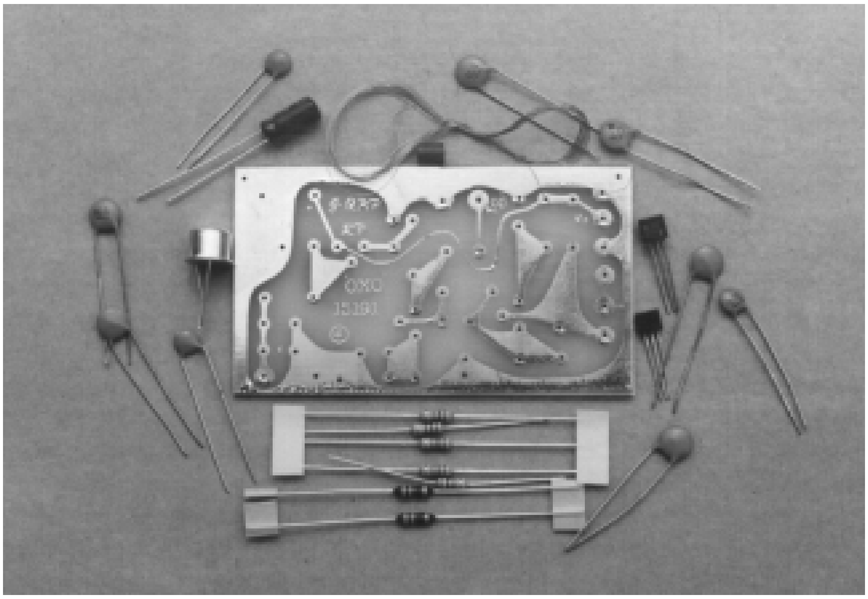

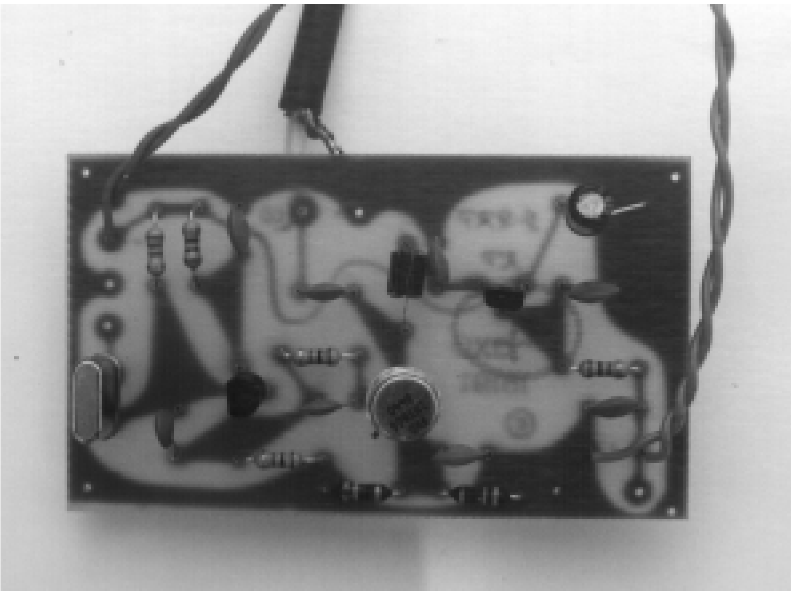

Warning

Note that the transmitter has no filtering; harmonics are not suppressed. It is strongly recommended that you use this transmitter in conjunction with the excellent low-pass filter described in the project *A 7-element low-pass filter for transmitters*, described elsewhere in this book.

88 Matching the end-fed random-wire aerial

Introduction

Many amateurs who do not have the space (or money) for a multi-band beam aerial, make use of the simplest possible alternative – the longest piece of wire that they can erect, with its end connected to the transceiver or receiver by an aerial tuning unit (ATU).

The length of the wire is not of major importance. Any length between 10 m and 80 m, with bends if necessary, will suffice. A good earth connection to the radio is just as important. Bends in the aerial wire can have some interesting effects on the directional properties of the aerial; V- or L-bends, or even a square shape are permitted. The only thing *not* permitted is to fold the wire back on itself in a tight hairpin bend!

Longish wire aerials

The term ‘long wire’ is usually used (incorrectly) to describe an end-fed aerial. How long is a piece of string? It depends what you mean by ‘long’. In aerial parlance, it means ‘long with respect to one wavelength’. Again, this depends on the band you are using. A long wire at 20 m is somewhat different from a long wire at 160 m. However, if you have sufficient real estate for a long wire on the 160 m band it must, by definition, be a long wire on all the other bands, too!

This should put you on your guard when analysing published data about feed-point impedance and the directional properties of a long wire aerial. Such theoretical data relate to a *real*, *ideal* long wire which is straight, horizontal, very high above perfect (conducting) ground, and not obstructed in any way. So your aerial doesn't quite match these criteria? Join the club!

Don't let this dampen your ardour when it comes to evaluating what the longish wire can do for you. The following should explain why.

Feed-point impedance

The impedance at the end of a longish-wire aerial can vary from a few tens of ohms to several thousand ohms, depending on the frequency in use and the wire length. It is also affected by factors such as bends, height above ground, proximity to buildings and wire diameter.

The actual value doesn't matter, provided we can make the aerial *appear* to have a 50 Ω impedance at the aerial socket of the transmitter. This process is what we call *impedance matching*, or simply *matching*. It maximises the power transfer from the transmitter to the aerial, and from the aerial to the receiver.

That is why an aerial tuning unit, or ATU, is almost (but not necessarily) obligatory.

The ATU

Many commercially produced HF receivers and transceivers have single 50 Ω coaxial sockets as their one and only means for connecting an external aerial. This means that an external aerial should have a 50 Ω feed impedance if it is to work efficiently, and it rules out most of the aerials being used by amateurs on the HF bands. Some means is necessary to change the aerial feed impedance to 'match' that of the transceiver. Such impedance-matching, or Z-matching (because Z is the symbol for impedance, just as L is the symbol for inductance) is the rôle of the ATU.

These can be bought and will accommodate either an end-fed or a coax-fed aerial. They can be bought ready for use or in kit form. Whether you want one for receiving only, or for use with a low-power (QRP) or high-power (QRO) transmitter, will determine what you need and how much you pay.

A simple single-band ATU

The simplest form of ATU is shown in **Figure 1**. It is simply a parallel LC (coil and capacitor) circuit, resonant at the chosen frequency, with taps on the coil for the aerial and the coaxial feed to the transceiver. If we assume the circuit is resonant, a high impedance exists at the top of the coil, and a low impedance at the bottom.

We said earlier that the end-fed longish wire presented an impedance which was high (or at least higher than 50 Ω). This explains why the aerial is tapped to the coil near the top, where the impedance is high, and the 50 Ω coax is tapped near the bottom, where the impedance is low.

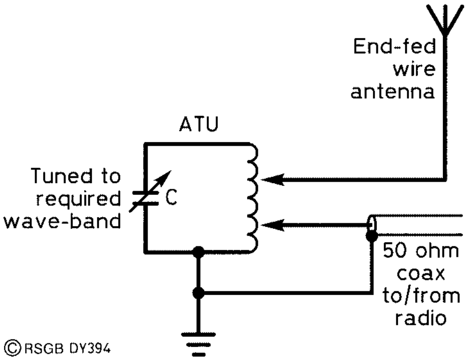

Figure 1 Parallel tuned circuit as single-band ATU for end-fed longish-wire antenna

Because the feed-point impedance of the aerial changes with frequency, so must the point at which the aerial is tapped to the coil to achieve impedance matching. The value of C must be changed also, to ensure that the circuit is resonant, and the 50 Ω tap will require tweaking also.

Setting up an ATU is quite simple. Make up an LC parallel-tuned circuit consisting of 50 turns of enamelled copper wire on an empty 35 mm film plastic container (or similar), tuned with a 500 pF variable capacitor. Make sure the enamel is removed from the ends of the wire before soldering.

Solder the inner wire of the coaxial cable from the radio to the first or second turn of the coil from its grounded end. Then solder the braid to the grounded end. Connect the aerial about one-third of the way down the coil from the top, removing the enamel at the connection point.

Adjust the variable capacitor for maximum noise or signal strength in the receiver. Then, try different tapping points from the aerial, to maximise the signal again. This matches the aerial impedance to that of the tuned circuit. Repeat the process with the coax tap, thus matching the impedance of the radio and the cable to that of the tuned circuit.

You will no doubt find that the tapping process on the coil was not easily accomplished, especially when the enamel must be removed at each tapping point without shorting adjacent turns together. It is therefore logical to produce a design where the taps have been prepared during the winding of the coil, and are selected with a rotary switch. To this end, the following multi-band design is described.

A multi-band ATU

The same type of coil design around a 35 mm film container is used (see **Figure 2**). The tapping points can be prepared in advance by a little judicious planning. Dismantle your original coil and measure the length of wire on it. It will be a little more than a calculation of $50 \times \pi D$ would suggest (where D is the diameter of the film container), due to the lack of tension in the coil and the wire diameter itself. You will need aerial tapping points at turns 1, 2, 3, 5, 10, 15, 20, 25, 30, 35, 40 and 45, corresponding to all the bands from 28 MHz to 1.8 MHz. The coaxial cable tap is fixed at turn 2 (an acceptable compromise). All the turns are counted from the earthy end.

Cut another piece of enamelled copper the same length as you used originally. Then, with the aid of an ordinary calculator, work out the positions of the points where the enamel must be removed for the taps. For example, turn 15 will have to be made $\frac{15}{50}$ of the way along the wire, turn 20 tap made at $\frac{20}{50}$ of the way along. So, if the length of wire is, say 4.7 metres, the two taps in question will be made at $\frac{15}{50} \times 4.7 = 1.41$ m and $\frac{20}{50} \times 4.7 = 1.88$ m from one end. This must be repeated for each of the tap positions, and the enamel removed ready for the wire to be soldered to it. With a

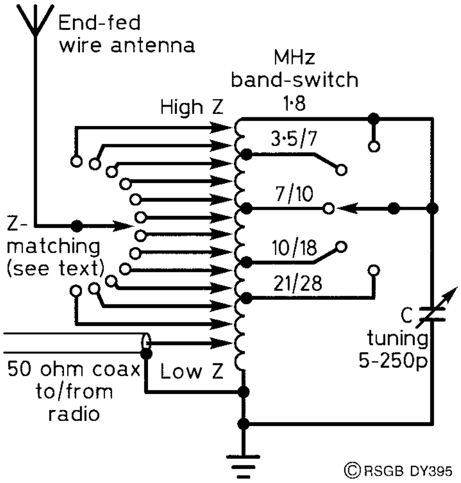

Figure 2 Simple multi-band ATU for end-fed longish-wire antennas

soldering iron, tin each tap point while the copper is shiny, thus ensuring a good, low-resistance connection. Do not solder on the taps yet.

The coil can now be wound as before. The taps can be soldered on, taking the lead from each one to the wafer of a single-pole 12-way rotary switch, the pole being connected to the aerial.

The tuning of the ATU is carried out by the same 250 pF capacitor, with a single-pole 5-way rotary switch used to select the band. Its tapping points will need to be chosen manually, using the method described earlier. Don't attempt to make new tapping points on the coil for this – use the taps available on the wafer of the other rotary switch, and find which is optimum for each band.

Notes

- For the aerial, use PVC-covered stranded tinned copper, of size 16/0.2 mm or 24/0.2 mm.
- Make the wire as long as possible, but anything over 10 m should be OK.
- Keep the wire as high as you can, in the clear and away from obstructions.
- Don't worry about bends, but don't use hairpin bends.
- Use a good insulator to attach your aerial – anything plastic will do.
- Anchor the wire near the point of entry to the building, but use a U-bend to prevent ingress of water.
- The wire can be brought in through the corner of a window, the PVC acting as an insulator. If you must drill a hole in the brickwork, make sure it slopes upwards from outside, so that water is deterred from entering.
- Use a good RF earth (as opposed to an electrical earth) such as half a dozen bare copper wires buried under the lawn in a fan shape. They should be joined together at the point of the fan and strapped to the earth connector of your ATU.

properties that make it particularly suited to our purpose, namely that it works at lower voltage levels than most other common diodes. Since the voltage in our radio comes from weak little radio waves, we need all the help we can get.

We now carry this diode in our catalog, to make it easier to obtain. Radio Shack used to carry them, but they no longer have them in their stores.

We are now ready to build our simplest radio.

A very simple radio with two parts

First let me warn you that this first little radio may not work in your location. It relies on having a very strong local radio station to overcome the limitations of such a simple radio. If it does not work where you are, you can either build its cousins that we will discuss later, or you can drive out closer to a local radio station, and try it there. But because it is so simple, you might try building it just to see what you might be able to pick up.

If your earphone has a jack on the end, cut it off, so you have two long wires coming from the earphone. If the wires are twisted around each other, that is OK, since we only need them to be separate at the very ends.

Remove the covering (called insulation) from the ends of the wires to expose an inch of bare wire. Often you can do this with your fingernail, but a tool called a wire stripper is made for this purpose, and can usually be purchased at the same place you got the earphone or the diode.

Wrap one bare wire around one of the diode's wires. Use some tape to keep it in place. If you know how to solder, you can solder the wires together, but it really isn't necessary for now.

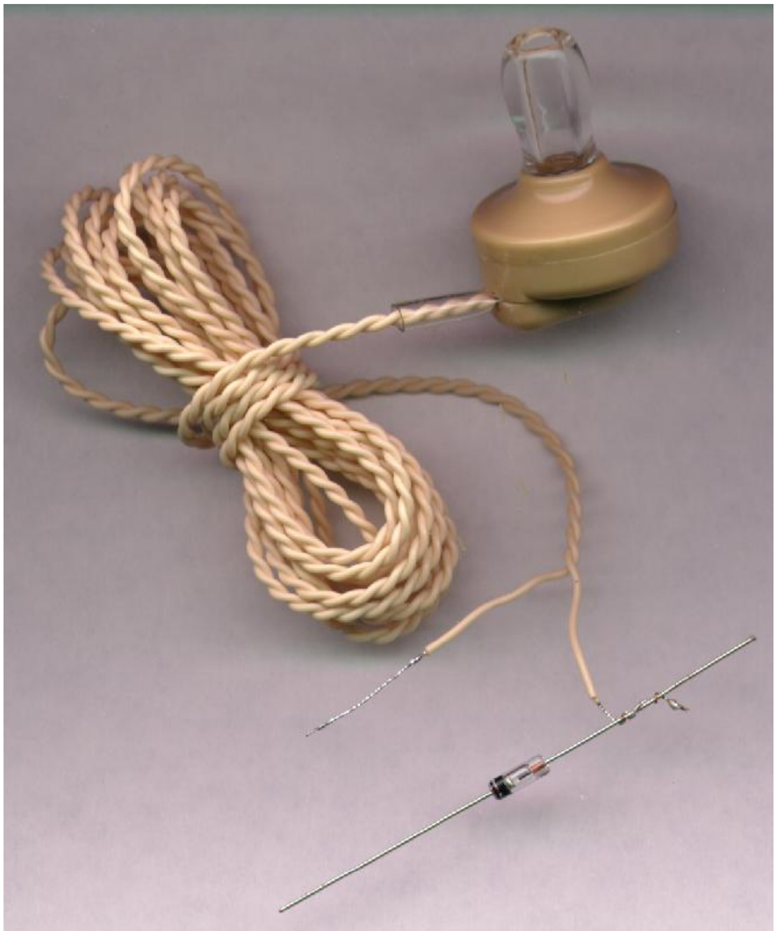

Tape the other diode wire to a cold water faucet. This makes a good connection to the ground, and is thus called a 'ground' connection.

Hold the remaining free bare wire of the earphone in your hand. This makes your body into the antenna for the radio. Put the earphone in your ear. If you are close to a strong AM radio station, you will be able to hear that station faintly in the earphone. You may hear more than one station at once.

If you can't hear anything, you might try a better antenna. You can tape the wire you were holding to a metal window screen, or a long wire. If one end of the long wire is thrown up on a roof or in a tree, you might get better results. Another good antenna is an outdoor TV antenna. Just touch the free earphone wire to one of the antenna terminals where it comes into the TV. If you have a good antenna, you may be able to eliminate the ground connection, using your body as a ground instead, by holding the free diode wire in your hand.

Another simple radio with two parts

Our simple radio has two main drawbacks. One is that the signals are very faint, and can only be heard if you are close to a radio station's transmitting antenna. The other is that you hear all of the strong stations at once, and it is hard to pick out just one song or voice from the mixed up jumble. The first problem is called the 'sensitivity' of the radio. Our radio is not very sensitive. The second problem is called the 'selectivity' of the radio. Our radio is not very selective.

We can solve both problems by using a trick called resonance.

Resonance is a way of taking a little bit of energy, and using it over and over again, at just the right time, to accomplish a big task. We use resonance when we push someone on a swing. It would take a lot of work to lift someone several feet in the air, but we can do this easily on a swing by giving a little push over and over again at just the right time. Timing is important: if we push at the wrong time, the swing can actually lose energy instead of getting higher.

When an opera singer uses her voice to shatter a wine glass, she is using resonance. Her voice gives the glass a little push at just the right time, over and over again, until the glass is moving so far that it shatters. In a similar way, we can slosh all the water out of a bathtub by moving a hand in the water at just the right back and forth speed. Each time the hand moves, the water climbs a little higher, until it is over the top of the tub.

Radio waves can act like the sound waves of the singer's voice, or like the waves in the bathtub. Radio waves can cause electrons to move back and forth in a wire, just like the water in the tub. If the radio waves are moving back and forth at the right frequency, then the electrons in the wire will just be crowding towards one end of the wire when the radio waves start moving them back to the other side. Just like the water in the tub, the electrons will crowd higher and higher at the ends of the wire. These electrons can do work, like moving the brass disk in the earphone to create sound.

We can use resonance to build a radio that can pick up only one station at a time, and make a louder sound in the earphone. This radio will also have some drawbacks (for one thing it will be over 1,000 feet long!) but we will solve these problems in the next radio we build.

Suppose we pick a local radio station we want to hear. For this example we will choose 740 kilohertz on the AM dial. We now need to figure out how long the wire must be to resonate at this frequency. Radio waves travel at the speed of light. This radio wave is going back and forth 740,000 times per second. This means the wave needs to go about a quarter of a mile in one direction, then turn around and go back again, over and over. The actual formula for figuring out how long the wire should be is

$$\frac{\text{936 feet}}{\text{Frequency in Megahertz}}$$

or, for our example:

$$\frac{\text{936 feet}}{.740}$$

or about 1264 feet.

To make our radio, we take half of the wire (632 feet) and attach it to one end of the diode. We attach the other half of the wire to the other end of the diode. We attach one earphone wire to one side of the diode also, and the other earphone wire to the other end. We put the long wire up in the air by attaching each end to a tree (the trees must be about 1264 feet apart). Then we put the earphone into our ear, and listen to the radio.

Now I can think of a couple problems with this radio. It is not the most portable radio. Also, in order to change the station, we need to make the wire longer or shorter.

One solution to the portability problem is to coil the wire up by winding it on a box or a cylinder. Then we can solve the tuning problem by attaching the diode and earphone to the coil at different places (easy to do now that the whole wire is in one small place).

A simple radio with three parts

There are several ways to connect a coil of wire to a diode and earphone to make a radio. In the photos below, we show two possibilities that work.

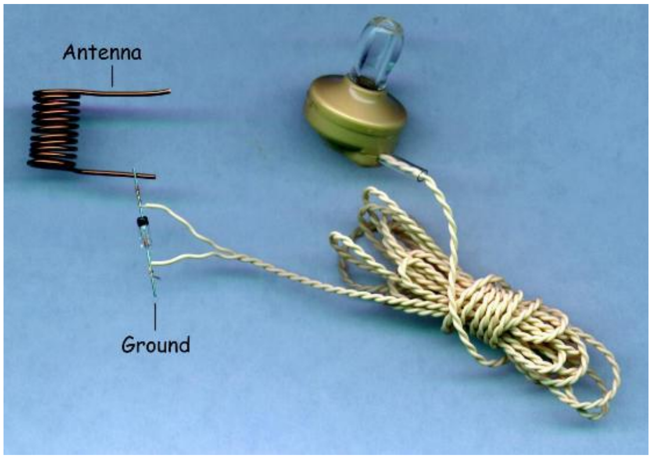

The photos do not show the antenna and ground connections, but instead indicate where they would be attached.

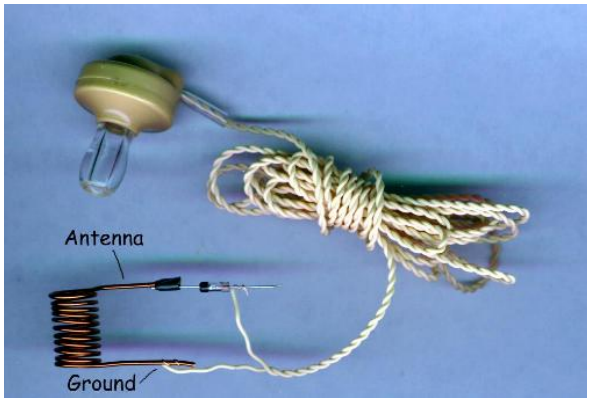

The coil in the photos is also dramatically simplified. A real coil for the AM radio frequencies would be somewhat larger, as we saw when we built our first radio using the plastic bottle.

Often photographs show so much detail that the important parts are easily missed. By using a simplified drawing, we can accentuate the important parts of the circuit and leave out unimportant or distracting details that can interfere with getting the point across.

A simplified drawing of a circuit is called a *schematic*. A schematic for a simple crystal radio might look like this if drawn on a napkin at a party:

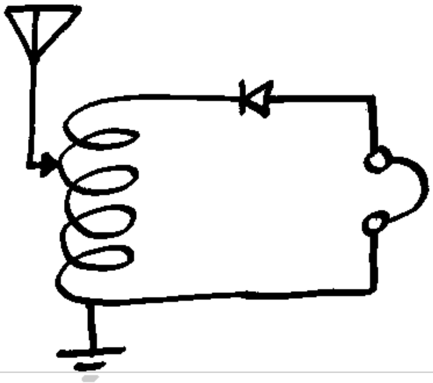

The symbol for a coil looks like a spring. The symbol for an antenna looks like someone used a coat hanger. The symbol for headphones looks like the old fashioned ear-muff style (which are great for crystal radios, since they block out ambient noise in the room). The symbol for the ground looks like what a cartoonist would draw under a cartoon character to represent the earth.

Note that the antenna is attached to the coil in the middle by a small arrow. This indicates that it is attached to a tap in the coil. An arrow is used to indicate a connection that can move, like our clip lead.

The symbol for the diode looks nothing like the little glass tube with wires coming out. Instead of represeting what the diode *looks* like, it represents what the diode *does*.

A diode is a one-way valve for electricity. The electric current flows through the diode in one direction, but is blocked if it tries to flow in the other direction. We will find out why this is important later, when we learn why the radio works. But for now, we will concentrate on building a radio that will let us hear one station at a time, with reasonable loudness.

Power from radio waves -- hooking up a meter to measure the voltage and current

It is useful at this point to be able to measure the effects of changes we make to the radio. We can just use our ears and try to remember how loud it used to be, but it is easier to read a meter, and remember a number. With a meter connected to the radio we can adjust the tuning for the highest meter reading, or make other adjustments as we add new components or replace purchased components with ones we make ourselves.

The meters must be sensitive to very small changes in the amount of electricity flowing in our radio. We will be measuring *current* mostly, but we will add a voltmeter as well, so we can calculate the total amount of energy we are receiving.

Current is the flow of electricity through the circuit, and it is measured in amperes, or amps for short. Voltage is the pressure that pushes the current through the wires. If electicity were water, current would be the amount of water flowing (gallons per minute), and voltage would be the water pressure in pounds per square inch.

Since the amount of current is very small, we will use a meter that measures current in micro-amperes, or at most small fractions of a milliampere. Some examples of microammeters and milliammeters can be seen in the photo below:

0 .2 .4 .6 .8 1.0
DC MILLIAMPERES
GC
0 10 20 30 40 50
D.C. MICROAMPERES
0 10 20 30 40 50
MICROAMPERES
D.C.
WESTON
DIRECT CURRENT
0 20 40 60 80 100
MICROAMPERES

To measure the current in our radio, we will need to have the current flowing through the meter. To do this, we connect the microammeter between the earphone and the ground connection, so that any electricity that is going to flow throught the earphones to make noise is going to have to flow through the meter also. The meter can be connected in two ways, one is forward and one is backward. If the meter is connected backward, the needle will start reading below zero. If this happens, just reverse the connections, so the needle reads above zero.

To measure the voltage, we connect the meter to both of the earphone wires. The schematic diagram now looks like this:

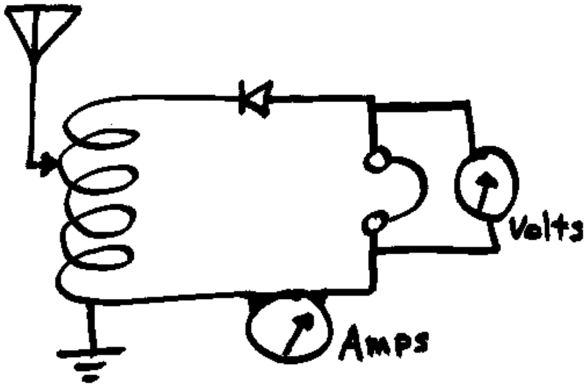

If you have a good antenna, or a strong radio station nearby, the ammeter might read more than 50 microamps. If you have a short antenna, you might get only 5 microamps and still be able to hear the station clearly in the headphones. I put up a 200 foot antenna between two trees over my house, and tuned to a 50,000 watt station about 30 miles away, and now I get 175 microamps of current through my meter. I put the earphone to the mouth of a cone (like a megaphone) and I can clearly hear the radio from across the room when the house is quiet. It doesn't sound as nice and clear as it does with the earphone right up to my ear, but I can follow a conversation easily (it's an all-news station).

The voltmeter in the same radio reads 125 millivolts. Since watts (the measure of how much power we have) is the voltage multiplied by the amperes, we have 0.000175 times 0.125, or 0.0000218 watts, or about 22 microwatts. The station is putting out 50 killowatts, and we are receiving one ten billionth of that power, yet we can hear it across the room.

Try different lengths of antenna, and watch the current go up as the longer antennas catch more of the power from the radio station. Try more that one antenna. Try connecting the ground wire to different things that are connected to the ground, such as pipes, metal fences, etc. As you try each test, make sure you tune the radio again, because your changes may affect the tuning.

Adding a capacitor (or three)

As you tried different antenna lengths, you may have noticed that you had to move the tap on the coil in order to get the station at its loudest. To understand why this happens, and how we can use an understanding of it to improve our radio, we must first understand *capacitance* and how it affects the tuning coil.

A capacitor is simply two pieces of metal with an insulator between them. If a capacitor is connected to a battery, the battery will push electrons onto one piece of metal (called a *plate*) and draw electrons from the other piece of metal. If we remove the battery, the electrons can't go anywhere, so one plate of the capacitor will have more electrons than the other plate.

If we connect the two plates together with a wire, the electrons will rush from the plate that had too many (because electrons have the same charge, and thus repel each other like the north poles of two magnets) to the plate that had fewer electrons. As the electrons rush from one plate to the other, we can make them do work, such as light a light bulb. In this way, the capacitor seems to store the electricity from the battery, for use at another time when the battery isn't there.

Now suppose we connect a coil and a capacitor together like this:

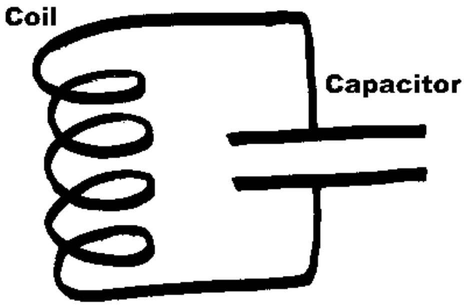

Suppose also that the capacitor has been charged by a battery so the top plate has more electrons than the bottom plate. When we connect the coil, the excess electrons in the top plate immediately start traveling through the coil to get to the plate that has a shortage of electrons.

As the electrons travel through the coil, they create a magnetic field, (remember 'coil' is just another word for 'electromagnet'). The magnetic field grows until the plates on the capacitor have equalized. At this point you would think the current would stop flowing in the coil. But the magnetic field that built up when the current flowed through the coil now starts to collapse.

Just as moving a magnet past a coil will generate a current, a collapsing magnetic field around a coil creates a current too. The current is in the same direction as it was when the magnetic field was created, so the coil ends up pushing electrons onto the bottom plate of the capacitor, and stealing them from the top plate.

By the time the magnetic field around the coil has completely collapsed, the bottom plate of the capacitor has a surplus of electrons, and the top plate has a deficit. You can guess what happens next.

The electrons start flowing back into the coil, this time from the bottom plate to the top. The coil starts building up a magnetic field again, but since the current is now going the other way, what used to be the north pole of the magnetic field is now the south pole, and vice-versa.

The field grows until the capacitor has equalized, then it collapses, and pumps electrons into the top plate of the capacitor. We are now back where we started, and the whole process starts over again!

The coil and the capacitor are resonating, just like the child on a swing, or the water in a bathtub. In fact, this circuit is called a 'tank circuit', like a tank full of water that sloshes back and forth.

We can control the frequency of the oscillations in two ways. We can make the coil larger or smaller, or we can make the capacitor larger or smaller. The coil we built for our radio has taps, which have the effect of making the coil shorter or longer, depending on which tap we connect to the antenna.

Our radio has a coil. But it doesn't have a capacitor. Or does it? Actually, the antenna itself is acting like a capacitor. The capacitance of the antenna is reacting with the *inductance* of the coil to resonate at the frequency of the radio station.

When we change the length of the antenna, it is like changing the size of the capacitor. This is why changing the length of the antenna changed the tuning of the radio, forcing us to move to a different tap on the coil in order to listen to the same station.

There is another way to change the capacitance of a capacitor. We can change the distance between the two plates. If the plates are closer together, the excess electrons on one plate are attracted to the other plate, because when the negatively charged electrons were removed from that plate, it was left with a positive charge.

Because the electrons are attracted to the positive charge, we can pile more of them together, storing more energy. In a similar fashion, when we make a capacitor with the plates farther apart, the positive charge is farther away, and can't help to pull as many electrons onto the negative plate. Thus the amount of energy we can store is less, and we say the capacitor has less *capacity*

We can combine capacitors to raise or lower the capacitance, now that we know how capacitors work. If we put two capacitors together in parallel, we can increase the capacitance, because the top plates are connected together, and the bottom plates are connected together, it is just as if we had one capacitor with large plates.

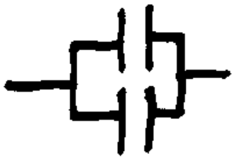

If we connect the capacitors in series, it has the effect of making the plates of the capacitor be farther apart. This can be seen in the illustration below. The bottom plate of one capacitor is connected to the top plate of the other. Electrically, this is the same as making the two plates into one plate in the middle of a capacitor that has twice the distance between the outer plates. The phantom inner plate has no effect, and is drawn as a dotted line in the bottom illustration.

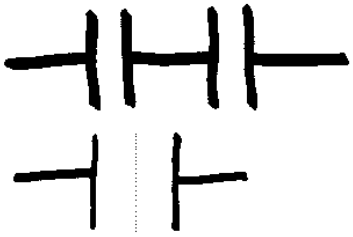

We now know enough about capacitors to use them in our radio. We can use a small capacitor between the antenna and the coil to lower the capacitance of the antenna. This will allow the coil to tune to stations that are higher in frequency. The capacitor is in series with the capacitance of the antenna, so the total capacitance is lower.

The circuit now looks like this:

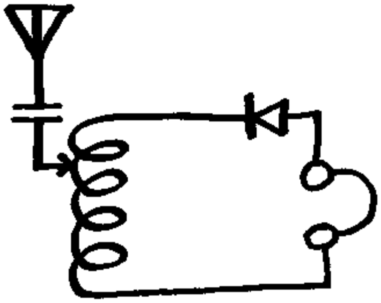

Building your own capacitors

Capacitors are easy to build in the kitchen out of aluminum foil. In fact, our first capacitor will simply be two sheets of foil tucked into a paperback book, with one page separating them, as if they were two bookmarks.

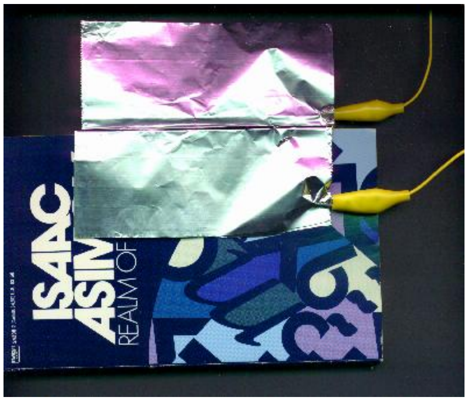

This quick capacitor has advantages and disadvantages. It is quick and easy to build, it can be easily adjusted to vary the capacitance by simply sliding one of the foil strips out of the book a little at a time, thus reducing the capacitance. On the other hand, it is bulky, and comes apart easily, and will change its capacitance when you press down on the book, squeezing the pages closer together. Lastly, it can change capacitance slightly on humid days as the pages of the book absorb moisture.

With only a little more effort, we can make a durable, stable, capacitor using foil and a little waxed paper or plastic wrap.

We start by laying down a sheet of waxed paper. On top of that we lay a sheet of foil. We leave the foil hanging over the top of the waxed paper, so we will have something to which we can attach a wire. We lay another piece of waxed paper over the first piece and the foil. We then lay another piece of foil on the top, overlapping it at the bottom for our other wire. We make sure that the foil sheets are always separated by the waxed paper, so they do not make an electrical connection.

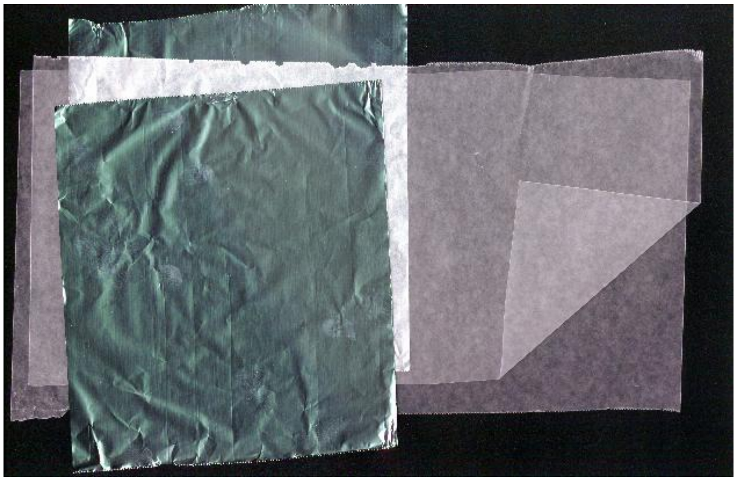

Now we roll the whole thing up like a jelly roll.

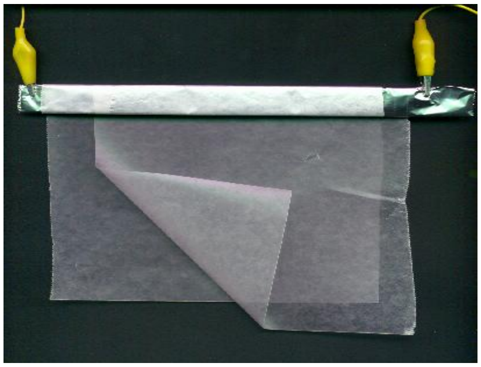

Now we trim up the paper with some scissors, and we can even roll it up the other way to make it smaller.

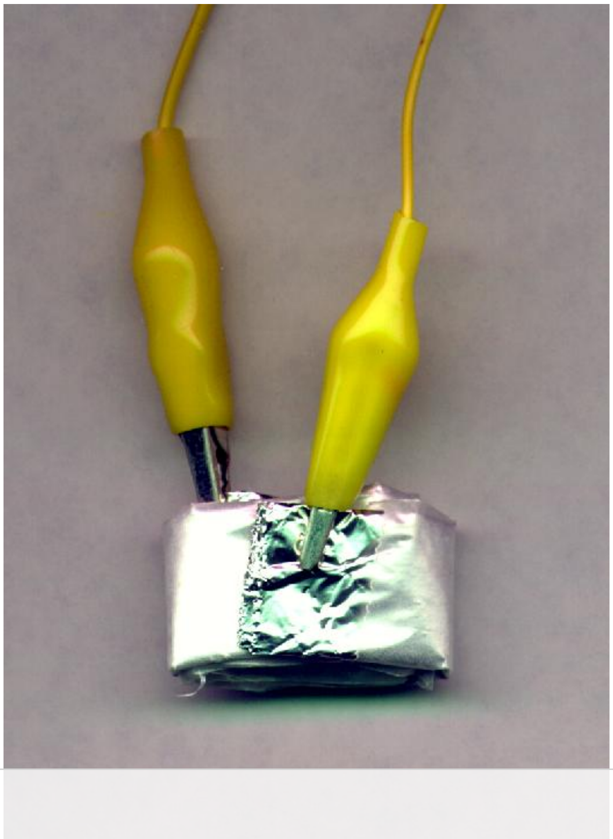

This capacitor is not adjustable like our first one, but we can make several of them, each a different size, and connect the one we want. We can even combine them in parallel or in series to change their capacitance.

We can use the small *fixed* capacitor to tune the antenna, and another *variable* capacitor (like our book capacitor) to tune the coil. We put the variable capacitor in parallel with the coil, to make a tank circuit. The small fixed capacitor lowers the antenna's capacitance, making the circuit tune to a higher frequency. But the variable capacitor adds more capacitance to the circuit, making it tune to a lower frequency. Now we can tune the radio with the taps on the coil, *and* by sliding the foil in and out of the book.

The circuit now looks like this:

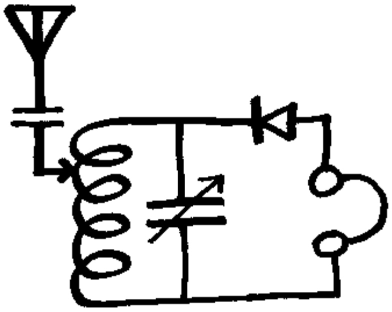

Notice how the variable capacitor has an arrow through it to indicate that it can change its capacitance.

Building your own diodes

During World War I, soldiers in the field made their own radios to listen to programs for entertainment and news. They had access to wire from broken down vehicles, and telephone receivers, but they did not have modern solid state diodes in little glass tubes.

However, it is surprising to find out just how many ordinary objects can act as a diode, letting current flow one way better than another.

The soldiers found that an old rusty razor blade and a pencil lead worked just fine. By lightly touching the pencil lead to spots of blue on the blade, or to spots of rust, they formed what is called a *point contact* diode.

We can replace our store-bought diode with a homemade point contact diode and compare the results. The parts can be attached to the circuit with clip leads, or they can be soldered, as in the photo below. The pencil lead is attached to a safety pin by wrapping it with bare copper wire and soldering it.

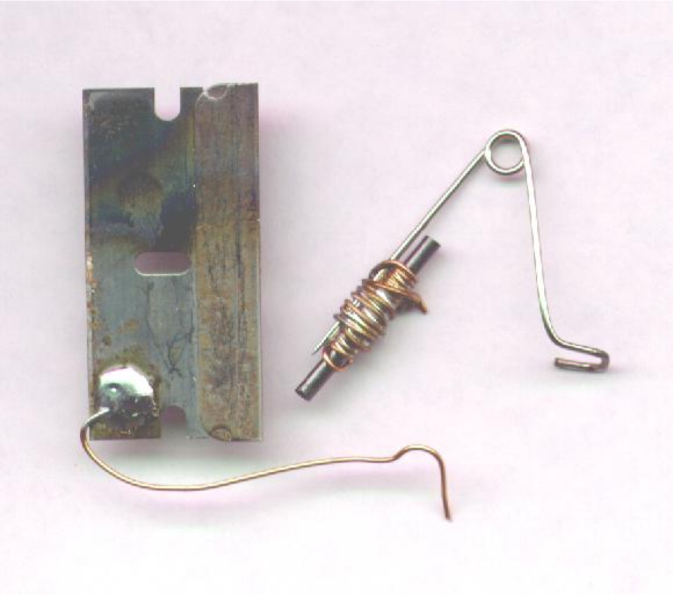

The safety pin acts as a spring to lightly press the pencil lead onto the razor. If the pressure is too hard or not hard enough, the diode will not work, so experiment. The exact spot on the razor is also critical, since some spots will have too much or too little oxide on them to make the diode. Move the pencil lead around on the razor until the sound is loudest, or the meter (if you have attached one) reads highest.

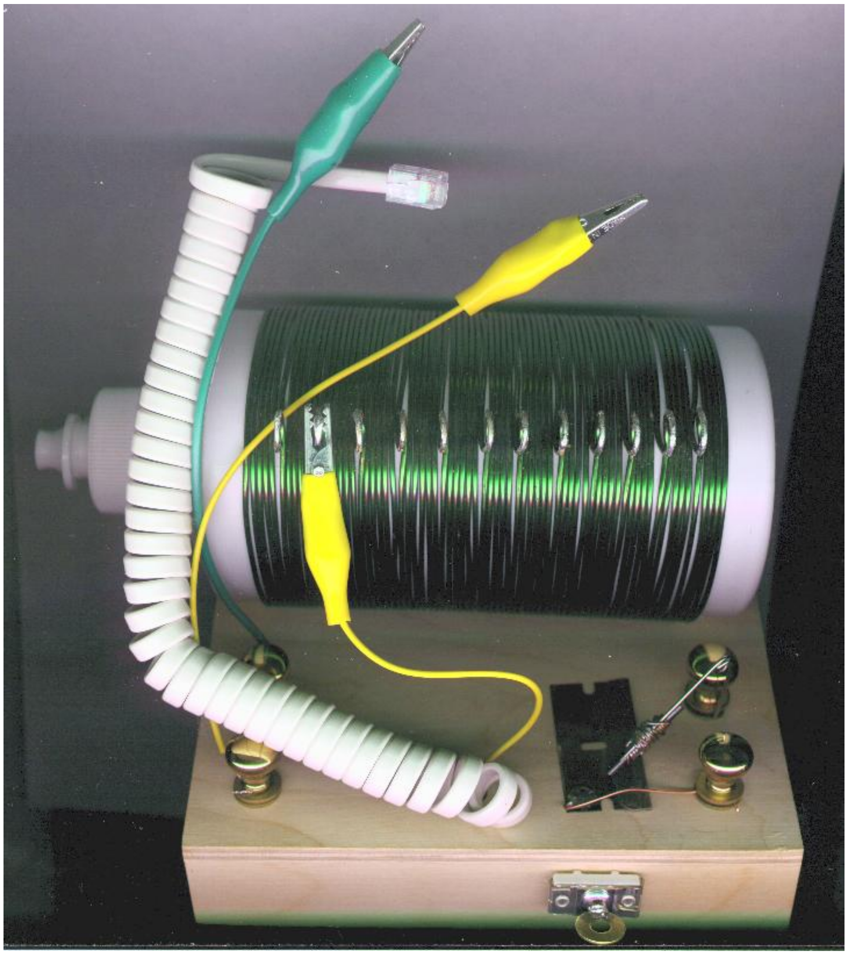

In the photo above, you can see how handy the brass drawer pulls are when we want to attach new types of diodes.

If you don't have a rusty razor blade lying around, you can try other bits of rusty metal. The blade shown above was clean and new, so I put a little salt and water on it, and held it in the flame of a gas stove until parts of it were blue and purple.

You might have other things around the house that can act as diodes. In my rock collection, I found some iron pyrite (fool's gold) and some carborundum (silicon carbide, the blue stone in the photo below). The carborundum works well with a strong pressure, so I simply wrapped some bare copper wire around it, soldered the wire, and then let the jaws of a clip lead supply the pressure. It works quite well. The pyrite needs a gentle touch, so I used the point of a safety pin to gently probe until I found a spot on the pyrite that gave good volume in the radio.

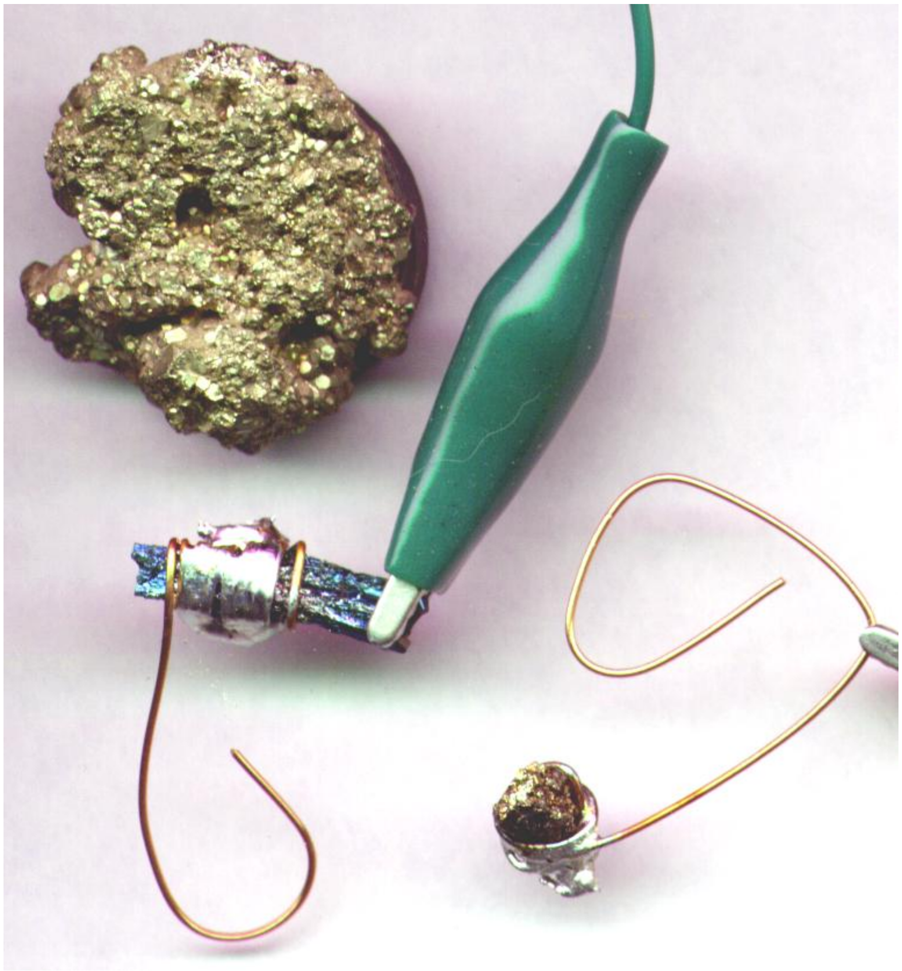

Going further - some quick thoughts

Trading loudness for more stations

In our radio, the diode and earphones are connected directly to the antenna and ground. This connection gets the loudest signal. However, it also *loads* the tuning coil, making it less selective. This means that many lower power or distant stations are drowned out by local strong stations.

We can make the radio more selective by decoupling the tuning coil from the antenna and ground. We do this by adding a small coil. The new coil is attached to the antenna and the ground, and then it is placed inside the main tuning coil.

Wind about five or ten turns of wire around a small coil form such as the plastic container use to package 35 mm film (about 1 inch in diameter). Cut a large hole in the bottom of the plastic bottle on which we wound the large tuning coil. Attach the antenna and ground to the small coil, and

TECHLIB.COM

Circuits | Science | Hobby | Forum
Home | Reference | Misc. | Links

Crystal Radio Circuits

- Simple Crystal Radio
- One-Transistor Amplifier/Detector
- TL431 Crystal Radio Amplifier
- Crystal Radio RF Amplifier
- Very High Gain Crystal Earphone Amplifier
- Simple Two-transistor Radio
- Experiments with Detector Diodes

Make sure to see the simple Reflex Receivers! A little more work yields vastly superior results.

Simple Crystal Radio

The crystal radio gets its name from the galena crystal (lead sulfide) used to rectify the signals. A "cat's whisker" wire contact was moved about the surface of the crystal until a diode junction was formed. The 1N34A germanium diode is the modern substitute for galena and most other germanium small-signal diodes will also work well. Silicon diodes are not a good choice because their much higher barrier potential requires larger signals for efficient rectification. Certain silicon Schottky diodes with low barrier potential will work well but most small-signal Schottky diodes will not perform as well as a garden-variety germanium diode.

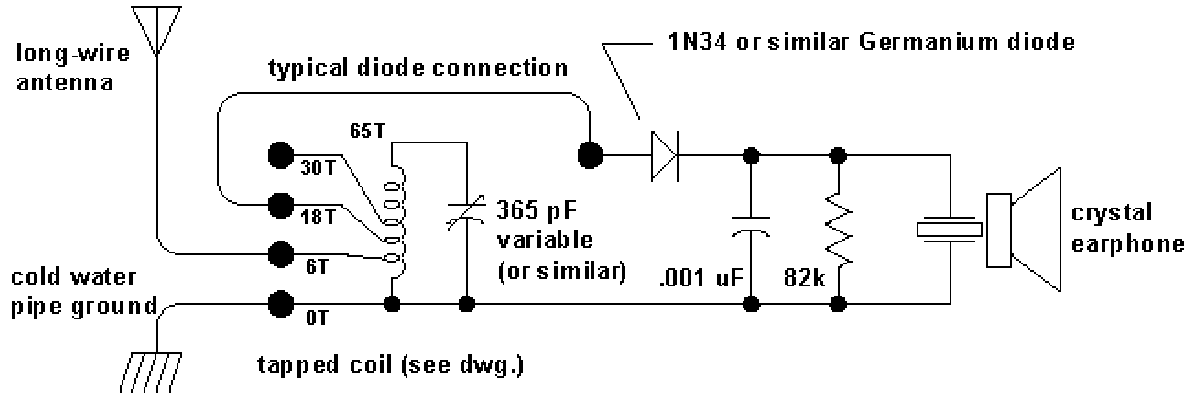
long-wire
antenna
typical diode connection
1N34 or similar Germanium diode
65T
30T
18T
6T
0T
365 pF
variable
(or similar)
.001 uF
82k
crystal
earphone
cold water
pipe ground
tapped coil (see dwg.)

Crystal radio with typical connections for a long wire antenna and good ground connections. The diode is connected for weak signals and moderate selectivity.

The circuit is quite simple but many pitfalls await the novice. The first precaution is most important! The crystal radio works best with a long, high outdoor antenna but the beginner may not fully appreciate the danger of bringing such a wire into the house. Lightning strikes to the antenna will probably destroy the crystal radio but if precautions are not taken, much more damage will result. The best strategy is to incorporate a commercial lightning arrestor with a straight, heavy gauge ground wire leading down to a buried water pipe. It is not sufficient to disconnect the antenna from the receiver during thunderstorms.

Other pitfalls are less dangerous and relate to the receiver's performance. A common mistake when building a crystal radio is to load the tuned circuit excessively. The Q of the tuned circuit must remain high to give selectivity or strong radio stations will all mix together. A good design will usually have low-impedance taps on the inductor for connections to the antenna and diode as shown in the schematic. A long wire antenna with a good ground connection will connect to the lowest impedance tap whereas a shorter antenna with no ground connection may connect to a higher tap. The diode may be experimentally moved to different taps and even across the whole coil for maximum sensitivity. The antenna and diode connection may be made with alligator clips for easy experimentation.

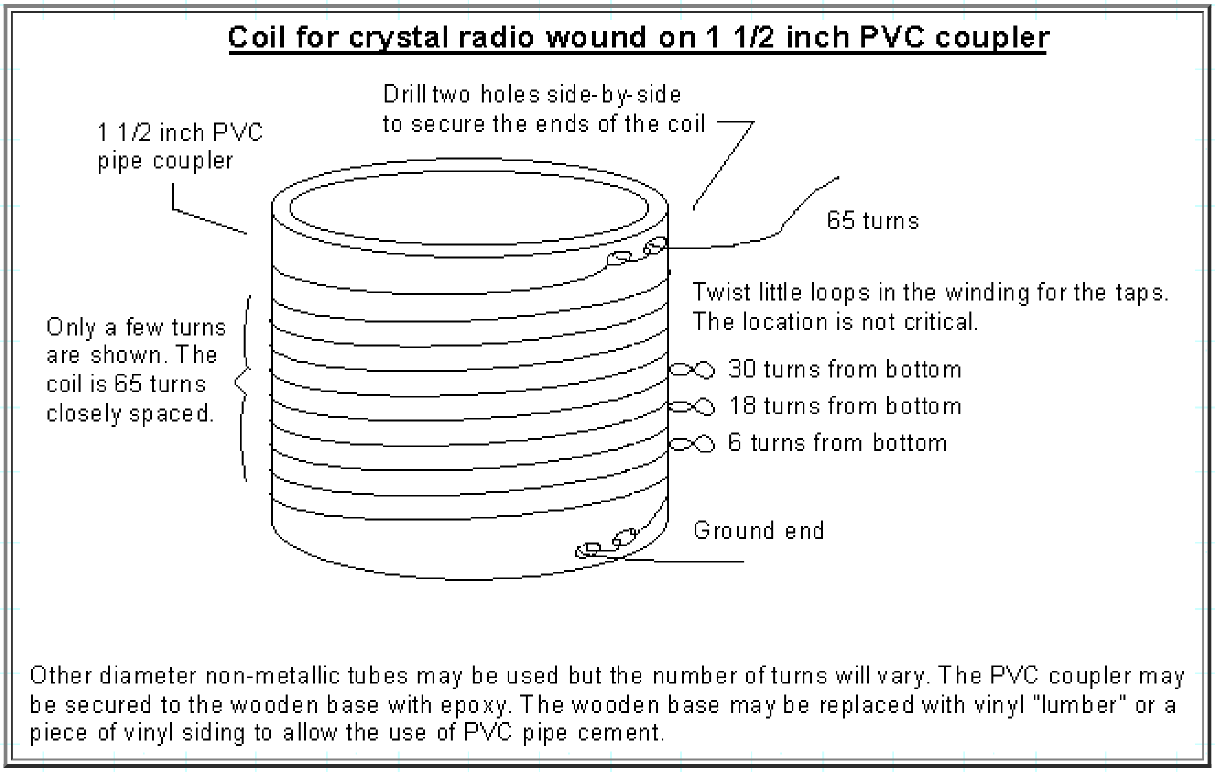
Coil for crystal radio wound on 1 1/2 inch PVC coupler

Drill two holes side-by-side to secure the ends of the coil
1 1/2 inch PVC pipe coupler
65 turns
Twist little loops in the winding for the taps. The location is not critical.
Only a few turns are shown. The coil is 65 turns closely spaced.
30 turns from bottom
18 turns from bottom
6 turns from bottom
Ground end

Other diameter non-metallic tubes may be used but the number of turns will vary. The PVC coupler may be secured to the wooden base with epoxy. The wooden base may be replaced with vinyl "lumber" or a piece of vinyl siding to allow the use of PVC pipe cement.

Another potential problem area is the earphone. Not all crystal earphones are sensitive and the experimenter should test a few to get a "good" one. High impedance dynamic earphones are a bit more reliable and can give excellent results. Try an old telephone receiver or a modern portable tape player headset (some are high-Z and fairly sensitive). Low impedance earphones like those used with many portable radios will not work at all. A simple test is to hold one earphone wire between the fingers while scraping the other lead across a large metal object like a file cabinet. If static is heard in the earphone it will probably work well with the crystal radio.

The variable capacitor is often connected incorrectly. Make sure to connect the rotor to ground and the stator to the "hot" side of the coil. Otherwise, the radio will detune when the capacitor knob is touched. If detuning is noticed then try reversing the connections.

Some experimenters are tempted to omit the 82k resistor which discharges the capacitor on the theory that it wastes precious signal power. With a typical germanium diode, this little "improvement" may work somewhat but only because the diode has significant leakage and the performance will not be predictable. A dynamic earphone may be DC coupled eliminating the need for the resistor.

The coil may be wound on a 1.5 inch PVC pipe coupler as shown in the drawing. These typically have an outer diameter of about 2.2". Drill two small holes at each end to secure the ends of the coil. The wire type is not particularly critical but select a gauge and insulation so that the 65 turns cover about 2/3 of the coupler. An excellent choice is 30 AWG "wrap" wire from Radio Shack. The prototype uses this solid conductor wire with blue insulation. This wirewrap wire is available in 50' lengths on little spools and about 37' will be needed. A "loopstick" coil may be used in place of the coil shown. These coils have an adjustable ferrite core for tuning so a fixed value capacitor may be used in place of the variable capacitor shown. The coil, capacitor and a terminal strip for the other parts may be mounted to a small wooden board. (See photo of receiver with transistor amplifier below.)

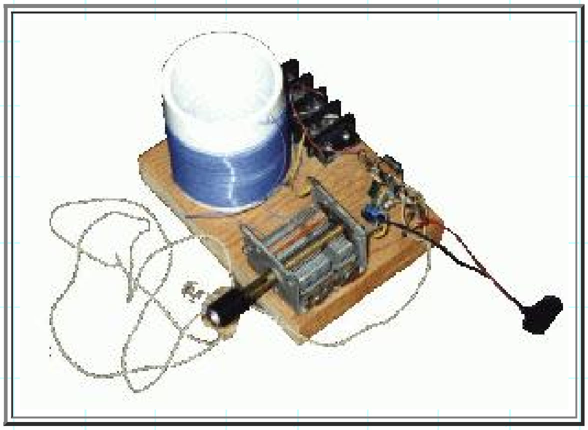

If a metal chassis is used then the coil must be mounted horizontally and above the metal to prevent unacceptable loading.

Here are some alternative construction ideas:

Fahnestock clips make excellent connectors for the antenna and ground wires. The coil may be mounted above the board or chassis with angle brackets by adding another bend, as shown below. The windings may be quickly secured with a single layer of colored "Duck" tape that is now available in more attractive colors than gray or black. The taps to the coil can be located at the rear, near the bottom so that the unavoidable bulges in the tape don't show. An ordinary piece of wood may be quickly finished by applying adhesive-backed PVC film intended for kitchen cabinets. Just stick it on and trim flush with scissors.

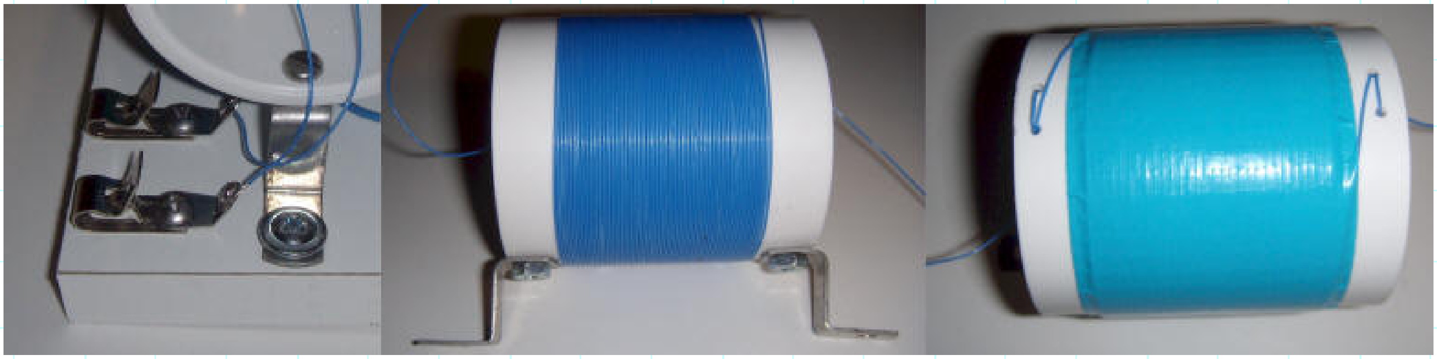

One Transistor Amplifier/Detector

An amplifier may be added to boost the audio level as shown below. The current consumption of this amplifier is quite low and a power switch is not included. Disconnect the battery when the receiver is stored for long periods.

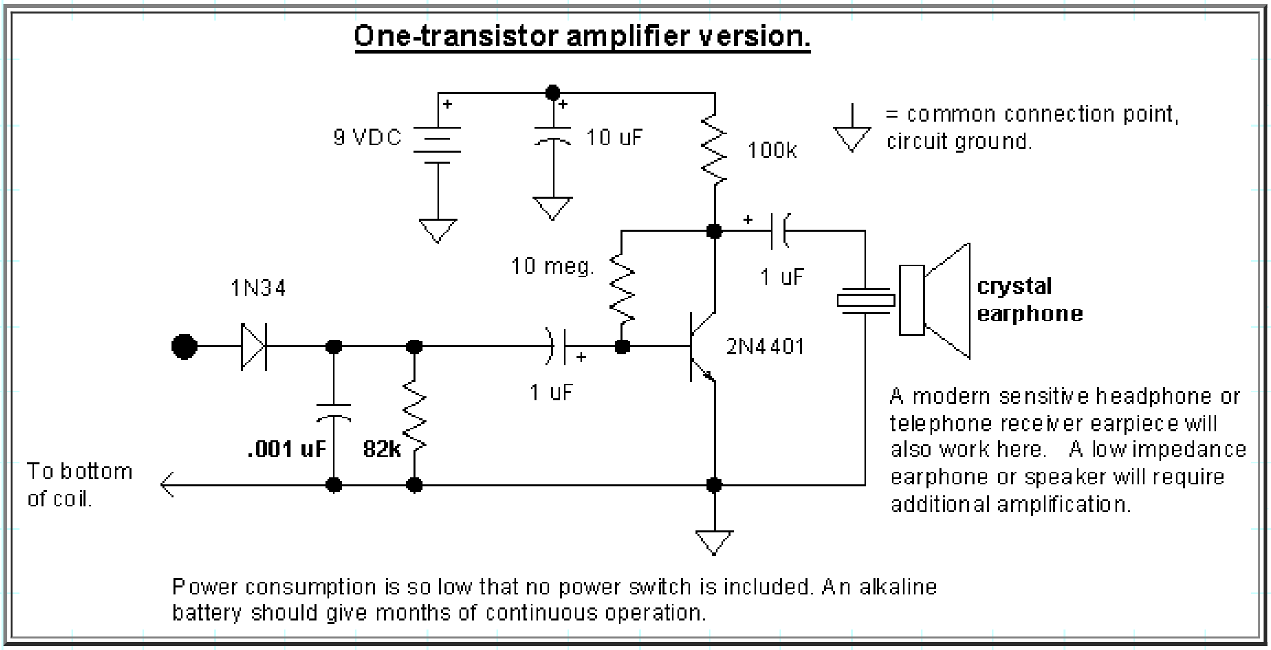
One-transistor amplifier version.
9 VDC
+
+
10 uF
100k
= common connection point, circuit ground.
+
10 meg.
1 uF
1N34
crystal earphone
2N4401
1 uF
+
A modern sensitive headphone or telephone receiver earpiece will also work here. A low impedance earphone or speaker will require additional amplification.
.001 uF
82k
To bottom of coil.
Power consumption is so low that no power switch is included. An alkaline battery should give months of continuous operation.

© 1995, Charles Wenzel

Note: You may use the transistor above as a sensitive detector eliminating the need for the 1N34A diode. Simply leave out the diode, the 0.001 uF, and the 82k resistor. Connect the negative side of the 1 uF directly to the coil. Change the base resistor from 10 meg. to 1 meg. and change the collector resistor from 100k to 10k. Now add a 0.01 uF from the collector to the emitter and the modifications are complete. This detector is quite sensitive and will be overloaded by very long antennas! Use a shorter antenna or a coil tap very near ground if significant distortion is noticed. The circuit draws about 1/2 mA.

Crystal Radio Audio Amplifier

Here is a simple audio amplifier using a TL431 shunt regulator. The amplifier will provide room-filling volume from an ordinary crystal radio outfitted with a long-wire antenna and good ground. The circuitry is similar in complexity to a simple one-transistor radio but the performance is far superior. The TL431 is available in a TO-92 package and it looks like an ordinary transistor so your hobbyist friends will be impressed by the volume you are getting with only one transistor! The amplifier may be used for other projects, too. Higher impedance headphones and speakers may also be used. An earphone from an old telephone will give ear-splitting volume and great sensitivity! The 68 ohm resistor may be increased to several hundred ohms when using high impedance earphones to save battery power.

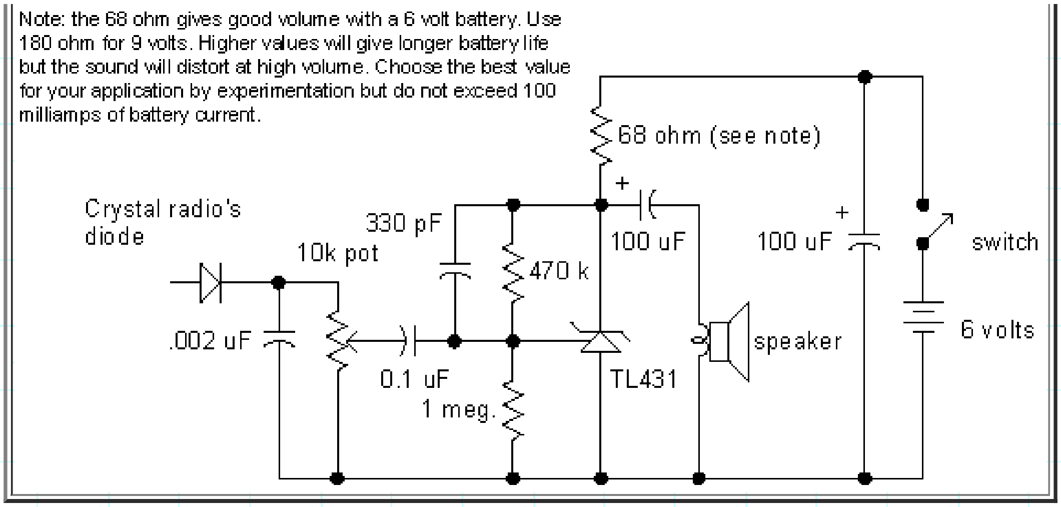

Note: the 68 ohm gives good volume with a 6 volt battery. Use 180 ohm for 9 volts. Higher values will give longer battery life but the sound will distort at high volume. Choose the best value for your application by experimentation but do not exceed 100 milliamps of battery current.

68 ohm (see note)
Crystal radio's diode
330 pF
10k pot
+
100 uF
100 uF
+
switch
470 k
.002 uF
0.1 uF
1 meg.
TL431
speaker
6 volts

To use the circuit as a general-purpose amplifier, apply the input signal to the top of the potentiometer. (Leave out the diode and .002 uF capacitor.) A higher value potentiometer may be used for a higher input impedance.

© 1995, Charles Wenzel

Crystal Radio RF Amplifier

For the more experienced hobbyist...

One of the best places to add a transistor to a simple crystal radio is at the front end in the form of an RF amplifier. The circuit below is a simple but effective amplifier which will give surprising performance improvement. This amplifier can exhibit negative resistance for low settings of the 500 ohm pot which results in extra gain or even oscillation. So, the circuit can actually be considered to be a regenerative receiver with an external detector. The sensitivity is so high that no cold water pipe ground is needed and the antenna is short.

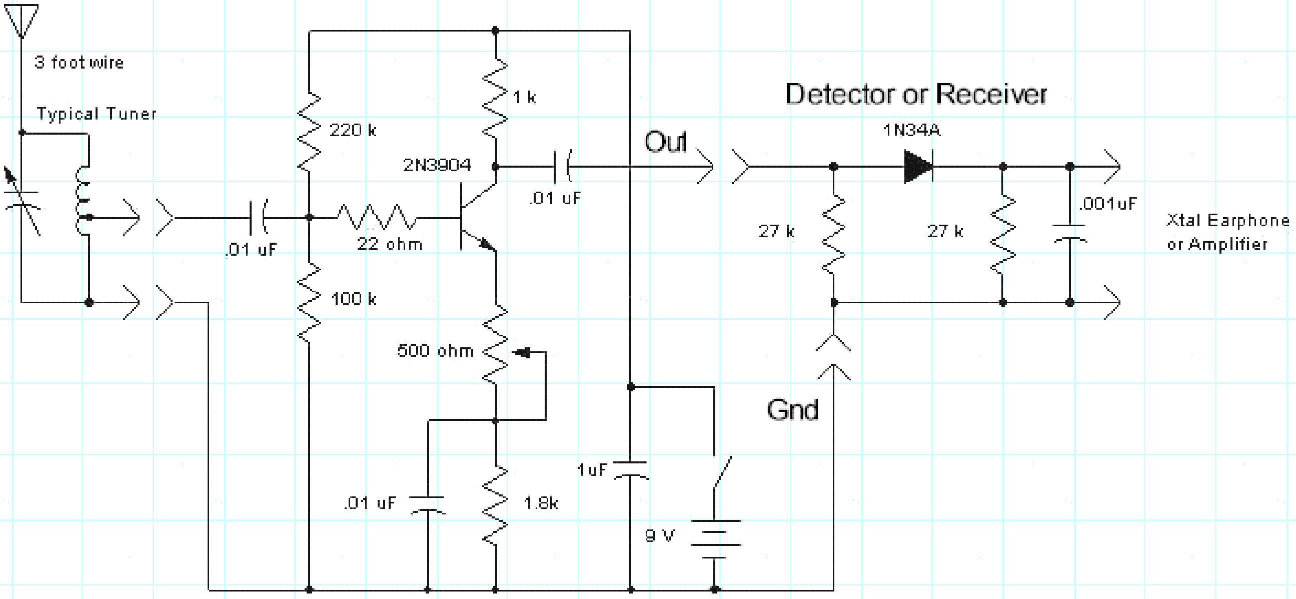

3 foot wire
Typical Tuner
220 k
1 k
2N3904
.01 uF
Out
Detector or Receiver
1N34A
.01 uF
22 ohm
27 k
27 k
.001uF
Xtal Earphone or Amplifier
100 k
500 ohm
Gnd
1uF
.01 uF
1.8k
9 V

The behavior of the amplifier depends on how it is connected to the tuned circuit. When connected to a lower impedance tap as shown in the schematic, the gain will be lower with less tendency to oscillate. Higher taps or even connection directly to the antenna will give higher gain and even oscillation. The 500 ohm pot is adjusted to give adequate gain without squealing as stations are tuned. High regeneration settings will actually narrow the bandwidth of the tank enough to give the sound a "mellow" quality which sounds pretty good in a "tinny" crystal earphone! Lower settings are best when using an audio amplifier and the fidelity is quite good thanks to the linear detector (typical regens use changes in the operating point of the transistor to demodulate the RF). As with any regen, the gain may be increased after the station is tuned in and the circuit will oscillate, locked to the station's frequency.

- Current consumption is about 1 mA which may be reduced by increasing the 1.8k but the RF envelope begins to distort below about 500 uA.
- A 4.5 volt battery may be used if the 220 k resistor is reduced to 68 k.
- The transistor may be just about any NPN small-signal transistor.
- No ground is shown but performance is better with a good ground connected to the bottom of the tuner.
- Longer antennas should be connected to taps instead of across the whole coil. A ferrite loopstick will pick up stronger stations with no antenna at all but use more audio gain after the diode detector and reduce the regeneration to get adequate bandwidth or the sound will be muffled.

Very High Gain Crystal Earphone Amplifier

This simple, one-transistor amplifier provides a voltage gain over 1000 (60 dB) for driving a high impedance ceramic (crystal) earphone. The high gain is achieved by replacing the traditional collector resistor with an unusual constant-current diode that supplies 1/2 mA yet exhibits a very high resistance to the audio. This amplifier will give excellent battery life, drawing only 500 uA.

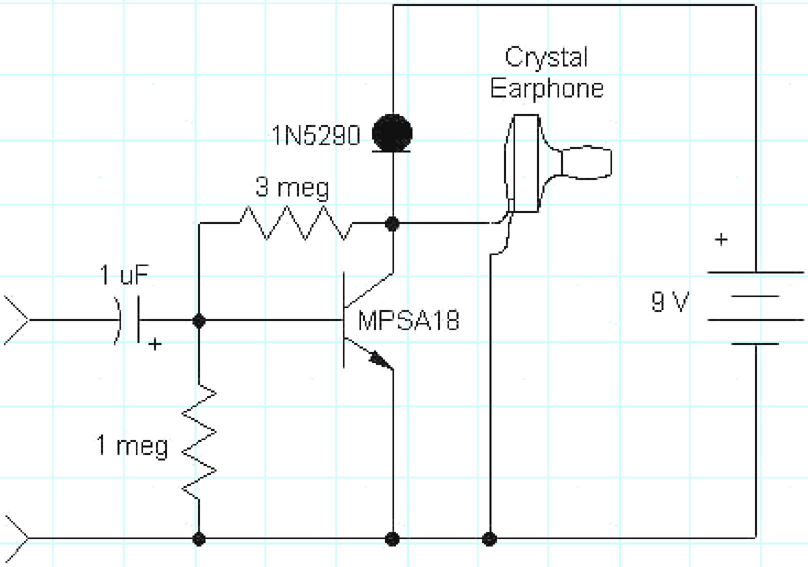

Below is a typical application using it with the first crystal radio circuit on this page. The amplifier provides good volume with a modest antenna. You may want a volume control as with the TL431 project!

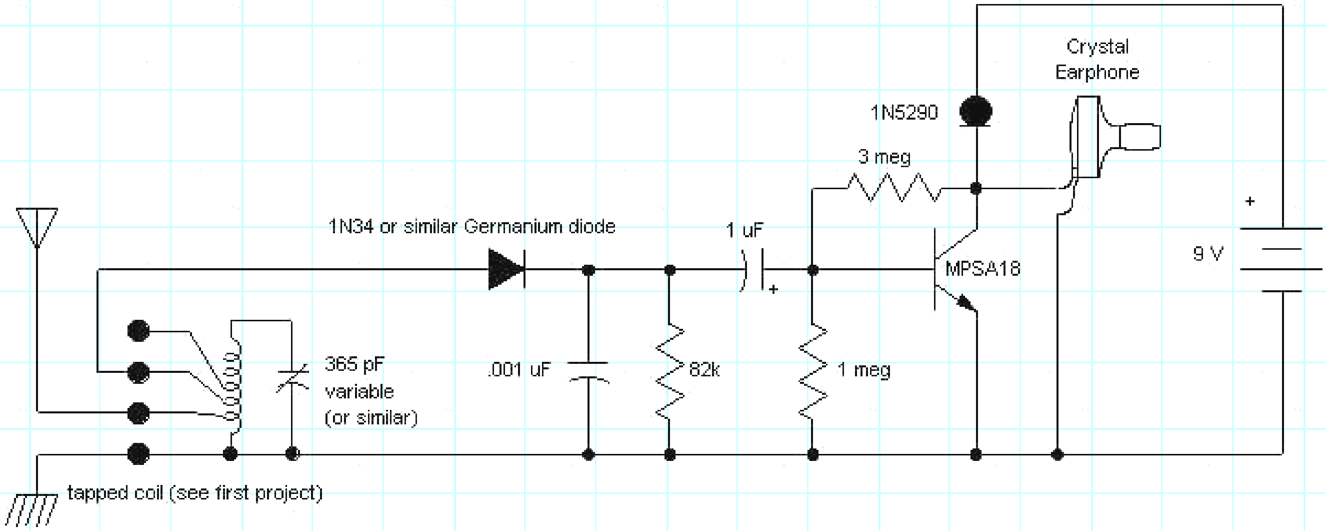

Or use the Crystal Radio RF Amplifier directly above for even more sensitivity with less than 2 mA current drain.

Simple Two-Transistor Radio

Here is a simple radio that was designed to minimize unusual parts; there isn't even a detector diode! The sensitivity isn't as high as the one-transistor reflex but the simplicity is attractive. Strong stations will provide plenty of volume into a crystal earphone or an external amplifier. The AM Loopstick was purchased on eBay but the enterprising experimenter can swipe one from the interior of a cheap radio. If the loopstick has more than one winding, use the one with the most turns. Wind 3 or 4 turns near one end of the winding as seen in the photo. The tuning capacitor in the prototype is from an old radio and the little plastic dial was cut down such that it just fit into the back of a black pointer knob. The fit was tight so no glue was needed. All of the sections of the capacitor were connected in parallel to get the most capacitance for this loopstick.

All the other parts are common. The transistors can be just about any small-signal type. The prototype uses the metal can 2N2222, primarily for looks. Some transistors may have too much high frequency gain; if the circuit squeals, try adding a small resistor in the emitter of the first transistor, maybe 47 ohms, the smaller the better as long as the circuit is stable. The large 47 uF could be smaller in most cases but the circuit can pick up hum if the wires are too long. Don't leave out the large capacitor across the battery, it provides needed low power supply impedance.

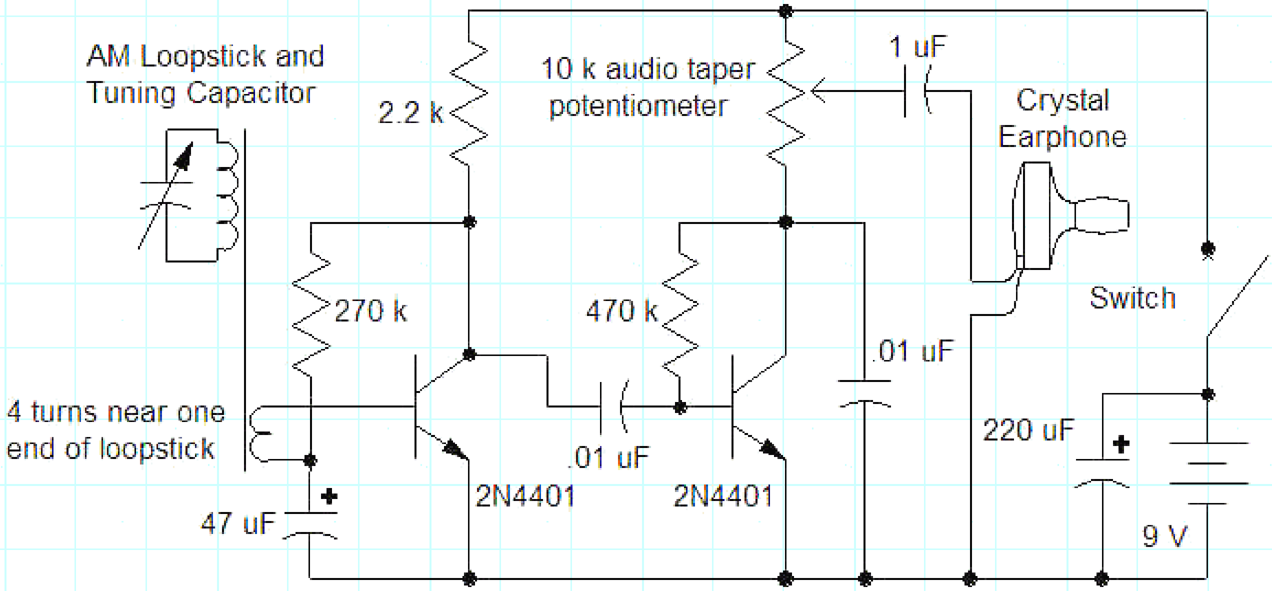
AM Loopstick and
Tuning Capacitor
2.2 k
10 k audio taper
potentiometer
1 uF
Crystal
Earphone
Switch
270 k
470 k
.01 uF
4 turns near one
end of loopstick
.01 uF
220 uF
47 uF
2N4401
2N4401
9 V

The circuit is built on a piece of 3.8" x 2.7" x 0.5" stained and varnished oak. The terminals are copper-plated nails used for weatherstripping. These nails are commonly available in home improvement stores and are also available in brass which is also solderable. Predrill the holes to make nailing easier and use a nail set or larger nail turned upside-down to make it easier to hit only the desired nail. The loopstick is held in position by an ordinary nylon cable clamp and the battery is mounted in a spring holder. The front panel is aluminum that was polished to a nice shine. First sand off all scratches with fine sandpaper. Then remove the sanding marks with ordinary kitchen steel wool. Now polish the surface with the finest steel wool in the paint department, usually "000". Then, for the real shine, polish the surface with a polishing compound like rouge. By the way, those paper-wrapped sticks of polishing compound are easily dissolved by lighter fluid (naphtha). Just put a few drops on a tissue and rub it on the end of the stick to load the tissue with compound. These polishing steps go quickly and you can have a mirror finish in a couple of minutes.

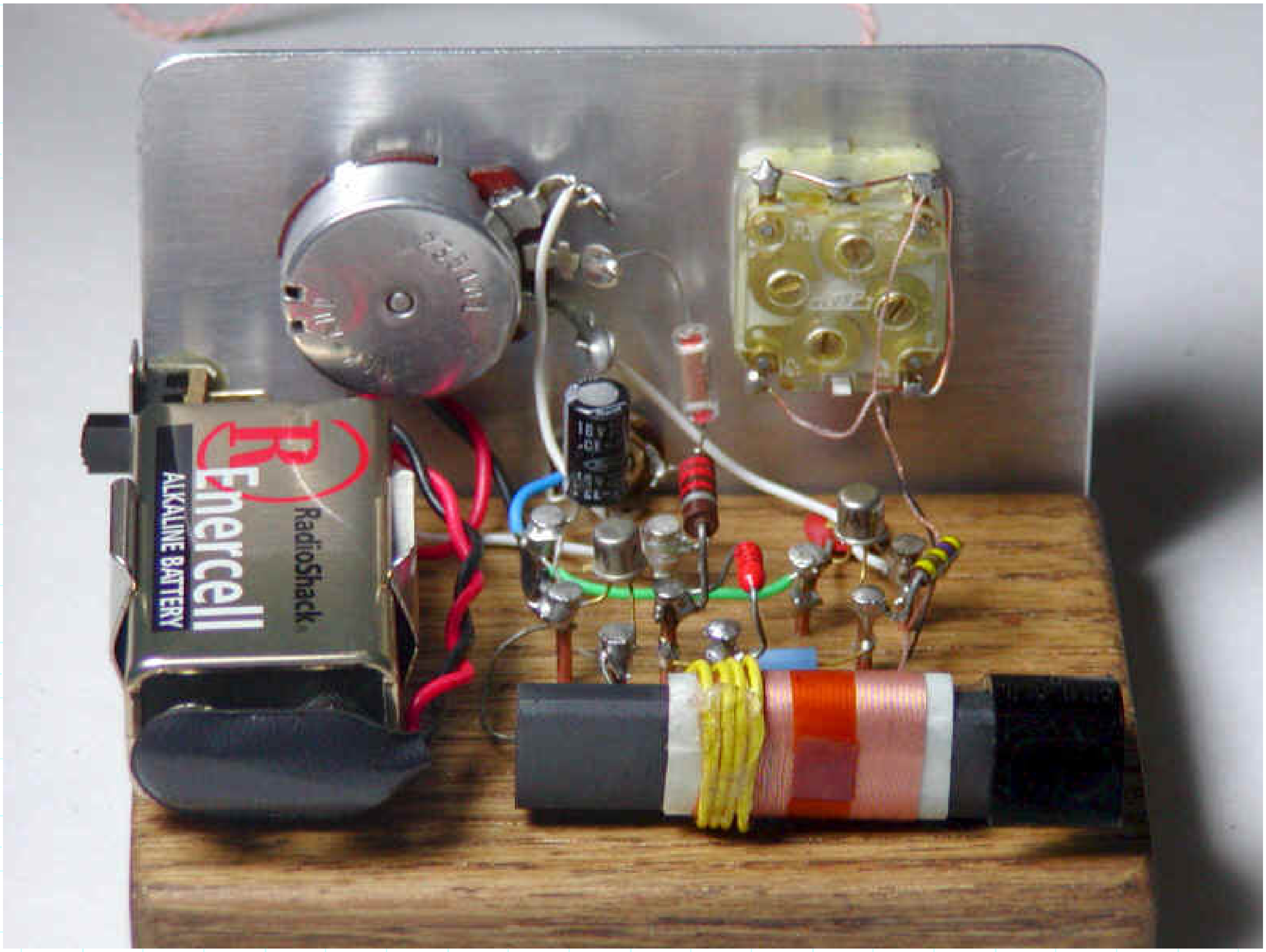

The front panel has a couple of scales for the tuning and volume printed on high gloss report cover stock and sprayed with a protective clear spray. The feet on the bottom are recessed in the oak using a forstner bit. The corners on the aluminum and oak were rounded on a belt sander.

AM
VOL
16
12
9
7
5.3
5
10
0
9
1
8
2
7
6
5

Experiments with Detector Diodes

When building crystal radios or other simple receivers, the experimenter often wonders about the relative performance of the different diodes in the junk box. Here are the results of several experiments using the typical types available to the hobbyist. The source is a low impedance and the load is a fairly high impedance. A particular diode will behave differently with different impedance levels but for low received signal levels these measurements are fairly predictive of the relative performance in most circuits. The diode types include germanium, silicon, Schottky, and even a light-emitting diode! The test setup uses an accurate RF synthesizer, a homemade AM pin diode modulator driven by an audio generator, a simple test fixture, a DC power supply for adding bias current, and a sensitive audio voltmeter. The setup shown below was used to test the diodes at several frequencies with a low modulation index (about 20%) and the near optimum bias current was determined by varying the DC supply.

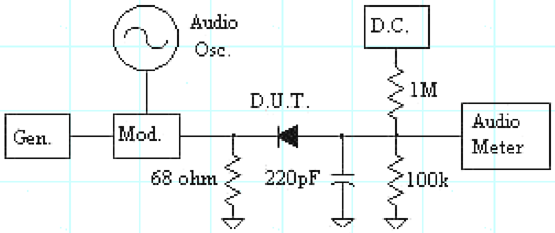

Initially an RF level of about -15 dBm was used but this level was dropped to -25 dBm without much change in the relative results. The best performance was provided by an H-P (Agilent) diode, the 5082-2835 with a tiny 10 uA of DC bias. The -25 dBm (35 mV p-p) results are shown below using the 5082-2835 Schottky diode at 20 MHz as the 0 dB reference point. The dB numbers are the audio level at the audio meter for the different RF carrier frequencies. Some variability is due to the test setup.

Diode Part Number	DC Bias	20 MHz	60 MHz	100 MHz	130 MHz	Notes
5082-2835 Schottky	10 uA	"0" dB	0 dB	-2.5 dB	-4.5 dB	quite good!
1N5711 Schottky	10 uA	-0.5 dB	-0.5 dB	-2.0 dB	-3.5 dB	better at high freq.
1N4454 silicon (similar to 1N914)	20 uA	-8.5 dB	-9.5 dB	-10.5 dB	-11.5 dB	pretty bad!
1N277 (Ge.)	None	-3.0 dB	-4.0 dB	-6.5 dB	-8.5 dB	not great at high freq.
1N34A (Ge.)	None	-3.0 dB	-4.0 dB	-6.5 dB	-8.5 dB	.
1N32	10 uA	-1.0 dB	-1.0 dB	-3.5 dB	-5.0 dB	microwave diode
Red LED	10 uA	-4.0 dB	-4.5 dB	-8.0 dB	-11 dB	not bad for low freq.

Note: -3 dB means that the audio voltage dropped to about 0.7 of the 0 dB level and -6 dB would be a drop to about one half.

The Schottky diodes are the all-around winners if bias is used but they do not perform as well as a germanium diode without bias. Other small-signal Schottky diodes gave nearly the same results as the 1N5711. Other germanium diodes were tried but the results were nearly identical to those shown. (The 1N60 was not available for testing.) The 1N4454 is similar to other ordinary silicon diodes and the results were poor, as expected. Several LEDs were tried and a bright type red led gave fairly good results as shown. Zeners were tried with dismal results in both directions. Large rectifiers like the 1N4001 were similarly poor.

Misc.:

- The optimum bias current is tiny and a very small battery can be permanently wired into a "crystal" radio without a switch. A little photocell battery like those on tiny calculators could do the job.
- Schottky diodes will work better without bias if higher source and load impedances are achieved. Germanium diodes are hard to beat for most low frequency crystal radio designs if no bias is desired.
- Schottky diodes don't need bias when heated by a soldering iron! (Hardly practical information.)

CRYSTAL SETS TO SIDEBAND

By Frank W. Harris, KØIYE
© Frank W. Harris 2002
© Frank W. Harris 2010, Revision 12.

FOREWORD

Dear Radio Amateur,

I began writing this book when I realized that my homebuilt station was almost unique on the air. For me, the education and fun of building radios is one of the best parts of ham radio. It seemed to me that homebrewing was disappearing rapidly, so I wrote articles about it for my local radio club newsletter. My ham friends liked the articles, but they rarely built anything. I realized that most modern hams lack the basic skills and knowledge to build radios usable on the air today. My articles were too brief to help them, but perhaps a detailed guide might help revive homebuilding.

I have tried to write the book that I wish had been available when I was a novice operator back in 1957. I knew that rejuvenating homebuilding was probably unrealistic, but I enjoy writing so I pressed on. I thought of myself as Don Quixote battling the windmills of modern apathy toward the details of technology and science in general. It seemed to be a hopeless quest. But even if it turned out that no one else cared, I found the project satisfying and extremely educational.

Modern book publishers don't print books that will have few readers. Because of its length and numerous color illustrations, this book would be expensive to put on paper. Not surprisingly, several publishers of electronics books had no interest in it. Fortunately, CDs cost no more than a postage stamp, so I began sending out CD copies with my QSLs to those radio contacts that showed an interest in my rig. It was great fun. I often got back thank-you letters and pictures of their homebrew equipment along with their QSL cards. Eventually, the book appeared on several websites and was even translated into Spanish.

This book has now been in circulation for 8 years. It is no longer new, so naturally interest has faded and some of the websites have dropped it. What I should be doing is continuing to build new designs and continually updating it. However, life must be lived in series and I seem to be spending more time hiking in the mountains than I do working on ham radio. My electronics projects these days are nearly all related to mountain rescue. As examples, as I'm writing this, I have several rope force dynamometers to modify waiting on my workbench and some data cables for our climbing equipment test tower that I need to rebuild. My mountain rescue work gets me out in the fresh air and in contact with people face to face. This is important for retired guys like me. I have added to two articles to chapter 16 about PIC programming and class D audio amplifiers that came from my rescue work but should also be interesting to hams. Sigh. Life is too short!

We homebrewers are nearly extinct, but there are still hundreds of us scattered around the world, some are even in the USA. Yes, there *ARE* American homebuilders! We're rare, but thanks to the QRP hobby, the number is growing. Even if we homebrewers don't change the world, I guarantee you will enjoy learning radio technology and building your own equipment. Someday I hope to meet you on the air … or at least on the Internet. Hi!

73s, Frank Harris, KØIYE

Frwharris@live.com

CRYSTAL SETS TO SIDEBAND
A Guide to Building an Amateur Radio Station

By Frank W. Harris, KØIYE
© Frank W. Harris 2010, Rev 12
(Note: This document is in PDF format. If schematics and drawings appear to have insufficient resolution, increase the display magnification.)

Table of Contents
Chapter 1
THE FASCINATION OF RADIO

Exploring the shortwave bands
Growing up in the Morse code era
The joy of building it yourself
A brief history of radio communication
    Henry, Maxwell, Hertz, Tesla, and Marconi.
    Fessenden, Edison, Flemming, DeForest and Armstrong
    The sinking of the RMS Republic and the birth of ham radio
Ham radio in the last 80 years
Becoming a radio amateur

Chapter 2
HOMEBUILDING AMATEUR RADIO EQUIPMENT

What qualifies as homebuilding?
When homebrewing is not appropriate
Barriers to modern homebuilding –
    Time, frequency stability, and lead inductance
Basic electrical knowledge
    Magnets and static electricity
    Voltage, current, resistance, energy, and power
        (Illustrated with drawings of water and mechanical analogies)
    Conductors, Insulators, and semiconductors
    Capacitors, inductors, transformers, and alternators
    Home power distribution, transformers at low and high frequencies

Chapter 3
SETTING UP AN ELECTRONICS WORKSHOP

R&D as recreation
How to build radios (or anything else) in your basement
    Persistence, read books, keep a notebook, and work in small increments

- Minimum tools needed
  - The ARRL Amateur Radio Handbook
  - Soldering irons and small tools
  - Drills & thread taps
  - Wood carving gouges for making PC boards
  - >50 MHz Oscilloscope
  - Frequency counter
  - Quality multimeter
  - Lab power supply
  - Calculator
  - Lab notebook
  - Collection of electronic junk
  - Parts catalogs
  - Capacitance meter
  - Test leads and socket boards
- Nice-to-have tools
  - RF and audio generators, spice software, and spectrum analyzer

Chapter 4
HERTZIAN WAVES IN THE BASEMENT

- The nature of radio waves
- Mechanical and LC electrical oscillators
- Antenna and transmission line theory
- Crystal set components
  - LC tuner
  - PN junction diode detectors
  - P-type and N-type semiconductors
  - Detection of AM signals
- Homebuilding the parts for a crystal set
  - The Jamestown diode
  - The Caribou headphone
- Revisiting Crystal Sets in 2006
  - Learning to troubleshoot
  - Selective tuning
- Recreating Hertz’s radio equipment
  - Transmitting and receiving as simply as possible
  - The 1880 ten-meter communicator
- Proving that radio waves exist and aren’t just capacitive or magnetic coupling
  - Demonstrating standing waves to measure frequency
- Building homebrew transistors
  - Bipolar transistors, PNP and NPN
  - Demonstrating power gain
  - The Boulder Rock Radio

Chapter 5
GETTING ON THE AIR
- DECIDING WHAT TO DO FIRST

- How to earn a license
- The rules of the homebuilding game – Whatever makes you happy!
- Picking an HF band
  - Getting acquainted with the HF ham bands, 160 – 10 meters + 6 meters
- Instant high quality HF communications
- VHF/ UHF handheld transceivers
- Building an antenna
  - Dipoles, regular and folded
  - Multi-band dipoles
  - 80 meters when you don’t have room for a dipole
  - The curtain rod vertical
  - A multi-band vertical antenna
  - Lightning protection

Chapter 6
BUILDING A QRP HOMEBREW

- A single-band, crystal-controlled, QRP module
- The transmitter mainframe
- HF construction methods
  - Making your own PC boards
    - “Dead Bug” and “Gouged Board” construction
    - Superglue “Island Boards”
  - Coax jumpers
  - Shielded boxes
- The complete QRP crystal-controlled transmitter
- Transistor amplifiers and oscillators
- How an amplifier becomes an oscillator
- Class A and Class C amplifiers
- Stabilizing the operating point, bypass caps, and emitter resistors
- Quartz crystals – the key to frequency stability
- The 40 meter QRP circuit
  - Oscillator and buffer
- Inductors, RF transformers, and impedance matching
- Tapped toroid inductors
  - How to wind them (and mistakes you might make)
- The final amplifier stages for the QRP
  - Tuned versus broadband - Use both for best results
  - Bifilar wound, broadband transformers
    - How to wind them (and how you might screw up)
  - Ferrite bead RF chokes, expensive RF power transistors, heat sinks, & output connectors
- Conquering inductors

- Calculating resonance
- Calibrating trimmer capacitors
- Calculating turns on powdered iron and ferrite toroids
- Chebyshev output low pass filters

Keying your QRP
- MOSFET power transistors

A “spot switch” for the QRP

Chapter 7
BUILDING A CODE PRACTICE RECEIVER

- A simple, direct-conversion receiver
  - A great first project for a new ham
  - Excellent sensitivity and good stability
  - Poor selectivity
- Adding 700 Hz audio filtering
  - High pass and low pass filters
  - Cascaded bandpass filters increase selectivity
  - Operational amplifiers
- Building with integrated circuits
- AM broadcast filter
- Getting rid of the image

Chapter 8
POWER SUPPLIES

- Line powered power supplies
- Power supply safety features
  - Isolation, 3-conductor cords, fuses, switches, ratings
- Supply performance and regulation
  - Rectification, ripple, chokes, capacitors, and bleeders
  - Zeners, linear regulators, switching regulators
  - A QRP regulated power supply
- A battery power supply for the radio shack
  - Solar cell charging, low drop-out regulators
  - Battery powered shack lighting

Chapter 9
ACCESSORIES FOR THE TRANSMITTER

- A straight key
- An electronic bug
- Building dummy loads
- “T” type antenna coupler
- A low pass filter
- How to stay legal with a homebrew transmitter

Antenna and power relays
Homebrew QSL cards

Chapter 10
VARIABLE FREQUENCY OSCILLATORS

Drift is a big deal today
Low frequency VFOs drift less than high frequency VFOs
JFET transistors
The oscillator circuit
The buffer, final amplifier, and output filter
The 50 secrets of avoiding drift
    JFETs, single-side PC boards, cast metal box, multiple NPO caps, small variable caps, precision voltage regulation, and more
Vernier tuning
Varactor tuning elements – advantages and disadvantages
A precision power supply
A voltage doubler power supply for battery use
    Square wave generator with a multivibrator
    Squaring up the square wave
    Charge pump, diode/ capacitor voltage doubler
    Schottky diodes for efficiency
Temperature compensation methods
    Positive coefficient capacitive trimmer compensation
        How to adjust the compensator
    Thermistor/ varactor temperature compensation

Chapter 11
Building a VFO for the higher bands (PMOs)

Old approaches that no longer work
    Frequency multiplication
    High frequency oscillators
Pre-Mix Oscillator method of frequency translation
A VFO-controlled QRP module
Crystal oscillators are stable, aren’t they?
    Crystal oscillator circuits
    Butler oscillators and big crystals
Mixers, bipolar transistor and dual-gate MOSFET
    Optimum drive requirements
    Direction of tuning, drift error cancellation
Multistage filters and filter/amplifiers
The QRP final amplifier stages
Spot switches for PMO QRPs

Chapter 12
FINAL AMPLIFIERS

- The basic features of a modern linear power amplifier
- It looked easier in the Handbook
    - Linear “noise mode” operation
- A tuned 50 watt class B amplifier
    - Ferrite balun transformers
- An untuned, sort-of-linear, class B, amplifier
    - Keying the 50 watt transmitter
- A linear Class AB amplifier, this time for sure
    - Single Sideband (SSB) needs a linear
    - Biasing without thermal runaway
    - Clamp diodes prevent runaway
    - Mechanical construction

Chapter 13
BUILDING A HOMEBREW HF RECEIVER

- Building a receiver - an unusual adventure
- What’s a reasonable goal?
- An “adequate performance” HF communication receiver
- Does it have to be so complicated?
- Planning your receiver
    - Direct conversion versus superhetrodyne
    - Why not single conversion?
    - Start with a single-band, single-conversion superhetrodyne
    - How do modern digital receivers do it?
- Receiver construction – build with shielded modules connected by thin coax.
- The 80 meter preselector
    - Reception on 80 meter and 160 meters is aided by a tuned transmatch
- The Variable Frequency Oscillator
- Mixer magic
    - Mixers will give you lots of static – and howls and squeals
    - A practical homebrew mixer made from discrete parts – it’s harder than it looks
    - Dual gate MOSFET mixers
    - Not all MOSFETS work equally well
    - A JFET alternative mixer
- Crystal ladder filters – essential for CW
    - All 9.000 MHz crystals aren’t equal
    - Using the BFO oscillator to match crystals
    - Switch in filters with a rotary switch
- The IF amplifier
    - Lessons learned from a dual-gate IF amplifier
    - The cascode amplifier strip - variable gain with constant Q
- Automatic Gain Control (AGC) - not a luxury

- The product detector
    - Nearly anything works at least a little
- The AF amplifier – a vital part of the signal dynamic range
    - Protecting your ears from strong signals
    - How Hi-Fi should it be?
    - Driving a speaker
- HF converters for the other ham bands
    - Crystal oscillators
    - Band switching
- Receiver power supplies
    - Use a linear regulator, not a switching regulator

Chapter 14
OLD-TECH VACUUM TUBE RADIO

- How old can radio technology be and still be used on the air today?
- Why bother with vacuum tubes?
    - – Glowing filaments, colored plasmas, and Jules Verne glass envelopes
- Power supplies for tubes
    - High voltage power supply safety
- The old-tech QRP transmitter
    - Vacuum tube amplifiers
    - The three roles of the triode filament
    - RF sinewave oscillator
    - Quartz crystals
    - Triode and pentode oscillators
- Old-tech voltage regulation – big, crude, expensive, but beautiful
- The travails of triode tubes
    - The oscillator and buffer
    - The final amplifier – triodes chirp
- The transmitter power supply
    - An inadequate supply from a 1935 radio
    - A good power supply made from cheap, modern, boring parts
    - How to check out junk power transformers
    - A complex but adequate supply made from ancient parts
- It works! No one suspects it's old and it's a success on today's 40 meter band
- An old-tech receiver
    - A super regenerative receiver made from ancient tubes
    - The power supply
    - Super-regen on the modern hambands
        - Lots of fun, but not up to modern QRM & QRPs - back to the drawing board!

Chapter 15
THE NOBEL PRIZE FOR SIDEBAND

- It can't be that hard! Want to bet?

- The sideband generator – how it works
    - The 9 MHz oscillator / amplifier
    - The audio amplifier
    - The balanced modulator
    - Building your own crystal ladder filter
    - Decoupling the power supply leads
    - Getting rid of RF feedback - RF filtering for all inputs
- Tuning and testing
- Using the generator for AM modulation and CW
- Moving the 9 MHz SSB signal to a hamband
    - Move the SSB only once!
- No wonder most ham rigs are transceivers
- Moving the 9 MHz signal to the difficult HF hambands
    - Move the VFO first, then mix it with the SSB 9 MHz.
    - Pick your oscillator and VFO frequencies carefully
    - Hearing your own VFO in the receiver
    - The hardest band – 17 meters
    - 40 meters and 12 meters
    - Covering the widest band – 10 meters
- A linear sideband QRP, VFO-tuned module
    - All stages must be linear and low distortion
    - All gain stages should be broadband to prevent oscillation
    - Sometimes high pass filter output is needed and not the usual low pass
    - Checking out the generator
- Driving a 50 watt linear amplifier

Chapter 16
ANCIENT MODULATION & OTHER TOPICS

- Defining amplitude modulation
- AM modulation for vacuum tube final amplifiers
    - Plate, screen, & cathode modulation
- A "collector modulator"
- Converting a MOSFET keyer into an AM modulator
- Generating AM with an SSB balanced modulator
    - Compensating for non-linearity
- Compression by accident
    - You probably don't need to build a voice compressor
- Ham TV - The old way.
    - Fun with an ancient flying spot scanner TV camera.
- Getting on 60 meter sideband – lots of work, little reward
- PIC programming and class D audio amplifiers

In conclusion,

Homebrew ham radio is never complete - when it works perfectly and does all the latest stuff, the hobby is over. Not likely. Long live homebuilding!

CRYSTAL SETS TO SIDEBAND

A guide to building your own amateur radio station

By Frank W. Harris, KØIYE

3850 Pinon Drive
Boulder, Colorado 80303-3539

© Frank W. Harris 2010, Rev 12

****

Chapter 1

THE FASCINATION OF RADIO

Radio produces action at immense distances with no physical connection that can be perceived by our senses. A modern way to demystify radio is to say that radio is simply a kind of light that our eyes cannot see. To those of us addicted to shortwave radio, it's an adventurous realm that can be explored. When we listen to our radio receivers it is comparable to using the Hubble telescope to explore the heavens. Shortwave is fascinating because you can't predict what you are going to hear. You might hear a radio broadcast from an exotic foreign capitol. You might pick up an SOS from a ship sinking in a storm or maybe weather reports from a radio amateur on Pitcairn Island. The next evening the same frequency band might be completely empty except for two hams on the other side of your own town discussing the Super Bowl. Or you might receive coded messages intended for some undercover spy lurking in our country.

I'm not kidding. I routinely hear such coded messages consisting of groups of letters on the 10.1 and 28.1 MHz ham bands. The codes are usually sent in Morse code, but sometimes you will hear a voice reciting the letter groups. Sometimes the woman announcer finishes by saying, "Thank you for decoding this message!" Since hams are forbidden to use codes or modulation modes that are not easily decoded, these communications are at least illegal.

Yes, it's true that shortwave isn't as vital to world activities as it once was, but if there's any romance in your soul, shortwave is still entrancing and always will be. This book is about using amateur radio to recapture the adventure of early day radio and bring it into the present. It is also about learning electronics and radio technology. If you can get through this book, shortwave radio will still be fascinating, but no longer mysterious.

Admiral Byrd at the South Pole

I first became intrigued by shortwave when I read Admiral Byrd's book on his last expedition to the Antarctic. Admiral Richard Byrd was in the business of launching expeditions to explore the Earth's poles. These expeditions had no inherent commercial value except for book sales and sponsoring grants from companies hoping to gain visibility for their products. In order for Byrd to get those grants, the public had to be sufficiently interested in the expeditions to generate advertising value. With each polar expedition, finding new expedition goals that would be exciting to the public became increasingly difficult. Studying rocks, glaciers, and penguins was scientifically important, but not particularly interesting to the public. By the 1940's all the neat stuff, like walking to the North and South Poles had been done decades before.

On his last expedition to Antarctica, Byrd established a base on the Antarctic coast like all previous large expeditions. However, he succeeded in maintaining public interest by setting up a tiny second outpost on the polar icecap hundreds of miles south of the coast. Then he attempted to spend the Antarctic winter alone in his little under-snow cabin totally isolated from the world in the cold and dark. His one connection with his base camp at "Little America" and the outside world was Morse code radio contact. Other than producing some interesting weather reports, the outpost had little real value. However, it did attract attention. Who could help but be captivated by the ordeal of a man totally isolated, hundreds of miles from the nearest humans? It was like being marooned on the moon, utterly alone.

Byrd's messages were relayed from his big base back out to the rest of the world. As a boy I was fascinated by Byrd's lonely vigil. I imagined what it would be like to be shut off from the world for months on end. I pictured Byrd bundled up in a fur parka huddled over his little table sending and receiving Morse code. His connection with the world was reduced to musical notes barely audible above the soft purring static of the polar night. The Morse tones came into his headphones and he wrote down their meaning, one letter at a time. The decoded messages appeared on his pad, one word at time. He fumbled with his pencil. "Was that a 'C' or a 'K?'" he asked himself. He pushed his indecision aside and kept writing down new letters. Fretting about one letter can destroy the whole sentence. A radiotelegraph operator learns to focus on the stream of characters and not dwell on each one.

After a few weeks in his frozen prison, Byrd began to suffer from headaches, nausea, weakness, and confusion. His Morse code became harder and harder to read and his team back at Little America became greatly concerned. Unknown to Byrd, his cabin heater was leaking carbon monoxide and was slowly killing him. Finally, when Byrd's condition became desperate, his crew drove hundreds of miles over the ice cap through winter darkness, howling wind, and below zero temperatures to rescue him.

Growing up at the end of the Morse code age

As late as 1960, Morse code was still commonly used commercially and by the military. Since Morse code had an exotic sound, news broadcasts were routinely introduced by snippets of code. When the word "NEWS" is spelled out in Morse and repeated rapidly, it makes a pleasant, rhythmic, musical phrase that blends in well with Hollywood-style introduction music. The public often assumed that messages from the other side of the world arrived by Morse code, although in reality its importance had been fading since the 1930s.

The Morse code used for radio communications in the English language

The "dashes" are three times longer in duration than the "dots."

A	**._**	**G**	**_ _ .**	**M**	**_ _**	**S**	**. . .**	**Y**	**_ . _ _**
B	**_ . . .**	**H**	**. . . .**	**N**	**_ .**	**T**	**_**	**Z**	**_ _ . .**
C	**_ . _ .**	**I**	**. .**	**O**	**_ _ _**	**U**	**. . _**		
D	**_ . .**	**J**	**. _ _ _**	**P**	**. _ _ .**	**V**	**. . . _**		

E	**.**	**K**	**_ . _**	**Q**	**_ _ . _**	**W**	**. _ _**
F	**. . _ .**	**L**	**. _ . .**	**R**	**. _ .**	**X**	**_ .. _**

Numbers and Commonly Used Punctuation

1	**. _ _ _ _**	**3**	**. . . _ _**	**5**	**. . . . .**	**7**	**_ _ . . .**	**9**	**_ _ _ _ .**
2	**. . _ _ _**	**4**	**. . . . _**	**6**	**_ . . . .**	**8**	**_ _ _ . .**		
0(zero)	**_ _ _ _ _**	**(, comma)**	**_ _ . . _ _**	**(. Period)**	**. _ . _ . _**	**(/ slash)**	**_ . . _ .**		

When I was a kid in New Jersey, my closest friend was Garth McKenzie. My introduction to ham radio was through his dad, Alexander ("Mac") McKenzie. Mac's call letters were W2SOU and his radio station was crammed into an alcove off the dining room. In the 1940s, quality radio equipment was packaged behind somber black aluminum panels 22 inches wide, eight inches high, and mounted in tall racks. The controls were enigmatic black knobs with strange labels like "grid drive" and "loading." The displays were usually just meters with equally arcane titles such as "S-meter" and "plate current."

The McKenzies had a cabin up in New Hampshire. Mrs. McKenzie and the kids spent most of every summer up at the cabin. Mac went up to New Hampshire on weekends when he could, but most of the time he stayed in touch with his family by radio. A friend of Garth's dad, Mr. Henny, lived near the McKenzies' cabin. He was also a ham, so on Saturday mornings Mac had a regular schedule to talk with Mr. Henny using Morse code, or CW (continuous wave) as it is still known. I was intrigued when I heard about these scheduled contacts and wanted to see Mac operate his station. I arrived at the McKenzie house at the appointed time. Sure enough, right on schedule, Morse code appeared out of the static. Mac wrote down the letters on a pad. I watched over his shoulder and stared at his pencil tip. It was mesmerizing to hear the code and watch the words and sentences appear on the paper. Unfortunately I couldn't understand even one letter of what Mac was sending, so I quickly tired of the one-sided conversation. In spite of that, Morse code had a mysterious, other-world quality and I was hooked.

Among the other equipment in Mac's radio shack was a Loran set. Loran was a long-range direction finder, the 1950 version of today's global positioning system (GPS). Mac demonstrated for me how to find latitude and longitude using a tiny green oscilloscope screen. The little round screen was only 2 or 3 inches wide and peered out from another one of those black, 22-inch wide black rack panels. Mac had it set up just for fun, of course. The Loran was designed for use on a ship and the McKenzie's house certainly wasn't going anywhere. Other toys included a pair of war surplus teletype machines. He could type on one machine and the other would print out his message 5 feet away.

The joy of building it yourself

It was hard for an eight year old like me to imagine getting a ham license and affording all that massive equipment. The Loran was quite alien to me and the teletype machines were interesting, but seemed like clumsy typewriters. What turned me on was Mac's television set. In the late 1940s, television stations were on the air, but no one I knew other than Mac actually owned a TV. That wasn't surprising. TVs cost as much as an automobile. Talk about a luxury! Undaunted, Mac built his own television from old radio parts and an army surplus, five-inch diameter, green oscilloscope tube.

A real, white phosphor, (black and white) TV cathode ray tube cost a fortune back then, so Mac couldn't even afford the picture tube. And because the TV tubes were designed for magnetic deflection and the oscilloscope tube used electric deflection, Mac couldn't just copy the deflection circuits from an RCA TV. Instead, he had to design his own custom picture tube drive and sweep circuits. Perfecting a new circuit meant that it had to be built and tested one small piece at a time. Since Mac had little idea how large the final circuit would be, he couldn't assemble his TV in a cabinet right away. Instead, he built his TV as a giant "breadboard" circuit with all the glowing tubes, wires, resistors, transformers, capacitors, and components all laid out in a huge spider-web matrix.

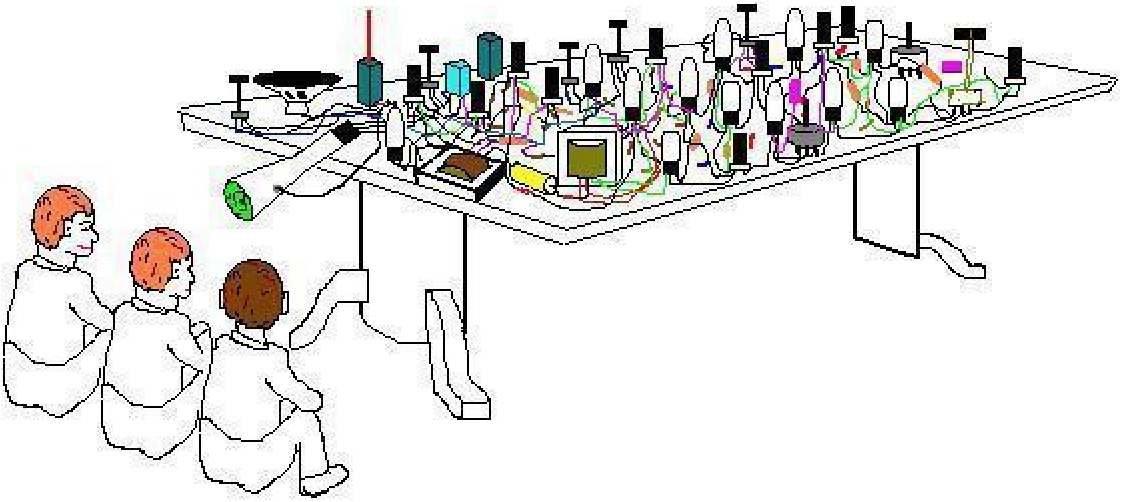

A TV is extremely complicated and a large breadboard was needed. Fortunately Barbara McKenzie was a tolerant woman. For several months the dining room table, including the extension leaves, was completely covered with about four by eight feet of television circuitry. Then Mac moved it upstairs into the master bedroom - probably not much of an improvement from Barbara's perspective. Toward the end of the year the TV began to work. We kids used to come home from school and sit on the floor and watch programs on the tiny five-inch round picture tube dangling off the end of the table. The pictures were in living "black and green." We watched "Zoo parade" with Marlin Perkins and our favorite program, "Flash Gordon," an early Star Trek–like show. TV was different back then. Flash Gordon was 15 minute film clips but most other programs were live. Even the commercials were live. I remember laughing silly over a commercial for a vacuum cleaner in which the fellow plugged the hose into the wrong end of the cleaner. The machine blew dust all over the room while the announcer tried to pretend it was working perfectly.

Eventually Mac installed his TV in an old record player cabinet. To make the picture larger, he put a big magnifying lens in front of the screen. When he watched TV, he propped up the hinged lid of the cabinet at a 45 degree angle and watched the enlarged picture in a mirror mounted on the underside of the lid. Mac showed me that, with patience, you can build almost anything. And, in the long run, it's usually much more rewarding to build a possession rather than to buy it. He also taught me that projects must be built and tested one tiny part at a time. ***If you build it all at once without testing the parts as you go, it might fit in the cabinet, but it almost certainly won't work***. There are very few short cuts.

The complete radio amateur

This book is about building ham radio equipment. To be sure, it's much, much easier to buy the equipment. In fact, commercial ham equipment today is so cheap, that buying it is far less expensive than buying the parts one at a time. The good news is that equipment you build yourself will have a value and meaning for you that can't be purchased. Along the way you'll learn more about electricity than you will ever learn reading the operator's manual of commercial equipment. Most of us will never be an Edison, Marconi, or Armstrong, but we can learn what they knew and we can share some of the thrill they felt when their inventions began to work. When your homemade station is finally on the air, you'll have all the same fun the other hams are having. But unlike the rest of the herd, you will be "***The Complete Radio Amateur.***"

A brief history of radio communication

Radio is based on phenomena that have been known since ancient times, namely static electricity and magnets. These phenomena also produce action at a distance with no visible connection, but only over extremely short distances. In 600 BC the philosopher Thales of Melitus described how, after rubbing amber with cloth, the amber could attract bits of straw. At about the same time the Greeks observed that natural magnetite ore (iron oxide, Fe_3O_4) could attract other bits of magnetite rock. This strange rock was found in a region of Turkey then known as "Magnesia," hence the name. Knowledge of natural magnets eventually led to the discovery of the magnetic compass. Compasses were a Godsend to sailors lost in fog and must have seemed magical to those who first used them. The compass was in wide use in Europe by 1000 AD.

Magnets and electricity appeared to be separate phenomena until 1820 when Hans Christian Oersted noticed that an electric current in a wire generates a magnetic field that can move a compass needle. Faraday and Henry studied and quantified the generation of magnetic fields produced by coils of wire that we now call called "inductors." In one of the all-time greatest triumphs of theoretical physics, James Maxwell published four equations in 1884 that summarized the connection between magnetism and electrical force. Maxwell's equations not only quantified and connected what was already known about these forces, they also predicted that magnetism and electric force could be combined to form a free-flying radiation. From the equations it appeared that these radio waves should be able to propagate great distances through space, much like light and heat. He may or may not have known that light and infra-red heat are also forms of electromagnetic radiation - radio!

What exactly is a radio wave?

An electric field and a magnetic field can both temporarily store energy in free space. For example, a refrigerator magnet generates a magnetic field in the space surrounding it. This magnetic energy hovers in "cloud" or "field" surrounding the metal magnet. Similarly, electric field energy is present in the space between the terminals of an ordinary flashlight battery. Suppose that magnets and charged batteries were sent into outer space and turned loose to float in the void. These devices would still generate their magnetic and electric fields in the vacuum surrounding the devices. However, if the devices could suddenly disappear, the magnetic and electric fields would not be maintained. The fields would quickly collapse and the energy would dissipate in all directions at the speed of light.

A battery or a magnet can be compared to a glass of water on a table. The glass holds the water in place and the water will rest there indefinitely. But if the glass were to suddenly break or vanish, the water would flood out in all directions. If either a magnet or battery floating in free space could be made to suddenly disappear, it would generate a radio wave that would propagate outward in all directions making a spherical shell of expanding waves. It turns out that ***collapsing magnetic field energy in free space is converted into electric field energy.*** Then, ***a moment later, the electric field energy similarly collapses back into a magnetic field.***

One way to look at it is that the collapsing magnetic field forces the storage of that same energy as an electric field in neighboring space. In other words, a collapsing field becomes a “device” that establishes the opposite kind of field in adjacent space. The end result is a wavefront of energy propagating across the void. As it travels, the energy oscillates back and forth in its form between electric and magnet fields. In the vacuum of space there is no dissipation of the original energy except that the energy becomes more dilute as it spreads out in all directions like ripples from a stone thrown into a pond.

The water analogy has other similarities with radio waves. The crests of the ripples on the pond represent the storage of mechanical energy as ***potential energy***. The potential energy is proportional to the height of the ripples or waves. The higher the wave, the more energy it stores. As the water falls back down, the energy from this descent is converted into ***kinetic energy***, that is, the outward velocity. As the wave spreads outward, the water stacks up to form another wave crest, restoring the energy to its potential energy form.

In 1887 Heinrich Hertz, a professor at the University of Bonn, Germany, managed to demonstrate in his laboratory that Maxwell’s radio waves actually existed. From then on other experimenters built “Hertzian apparatus” and tried to use it for communication or remote control. Experiments much like the ones Hertz performed are described in Chapter 4. Using rocks, copper wire, iron, and other materials available in 1880, you can build a short-range communicator to send and receive radio waves from one end of your house to the other. You can even demonstrate “standing waves” on an antenna.

How inventions happen

Big inventions usually begin with a novel observation. Faraday first invented the AC transformer with independent coils. An alternating current (AC) introduced into one coil on the transformer causes a second current to appear in a tightly coupled similar coil a fraction of an inch away. Today we still routinely use transformers to convert the ratio of current to voltage. For example, inside your flashlight battery charger, there is a transformer that converts a tiny current at 120 volts AC into a large current at 1.5 volts AC. If you applied 120 volts directly to your battery, it could be disastrous. Chargers would be quite impractical (or at least horribly inefficient) without transformers. We shall discuss these principles in detail in later chapters.

Getting back to Faraday, he must have marveled when he thought about the implications of electrical energy fed into one coil appearing in a neighboring coil. That is, the energy was “transmitted” across a gap. Yes, the gap may have only been a fraction of an inch, but certainly the thought must have occurred to him, “how far can it transmit?” In a letter in 1832 he proposed to a friend that electric energy could probably travel through space as waves. Unfortunately, he had no evidence, experiments, or equations to support this idea.

Many early radio communication experiments began when the first high frequency transformers were made. Unlike low frequency, like our 60 Hz line current, high frequency transformers of 500 KHz and above readily couple energy several inches through air. High frequency currents couple from one coil to another and begin to resemble radio.

It’s surprisingly easy to build a high frequency transformer and demonstrate crude, short-range radio communications. All that’s needed is a powerful battery, a large coil of wire, and a second coil wrapped around the first coil. The second coil is arranged so that the two ends of the wire are fixed a tiny distance apart, perhaps a sixteenth of an inch. The two ends of the first coil of wire are scratched transiently across the terminals of the battery. Huge currents flow in the first coil and establish a magnetic field around that coil. Since the same space is shared with the second coil, the magnetic field induces voltage across the second coil and a spark appears in the gap on the second coil. In other words, electric current was converted into magnetic energy, jumped across a short distance and then was reconverted back into electrical current. Now if the two coils are moved far apart, there will continue to be energy transmitted from one coil to other. However, with such a crude detection system, a spark probably won't be visible and a much more sensitive detector would be needed to prove that energy was actually transmitted.

Inventions appear when all the conditions are in place

New technologies appear whenever the necessary knowledge and affordable raw materials become available. For example, cell phones could have been built 50 years ago, but they would have been the size of suitcases, served few people and would have only been available to the most wealthy. Even today it’s possible to introduce useful technology too early to be profitable. The Iridium phone system is a world-wide direct satellite telephone system. Unfortunately, the Iridium “phone” is big and clumsy and the phone calls cost a fortune. Sure, you can reliably talk to a guy on dog sled at the North Pole, but there aren’t many people who actually need to do that. As of 2008 the Iridium satellites were still working, but five years ago there was talk of Iridium going bankrupt and crashing the satellites into the Pacific Ocean.

Radio was invented between the years 1884 and 1910 at a time when all the pieces to make it practical were in place. Many inventors had the chance to pursue radio communication, but most turned it down. To be more than a parlor trick, radio had to have a commercial reason for its development. The concept of broadcasting voices, music, and even motion pictures to the masses seems obvious to us now. But in 1900 it wasn’t obvious that radio could be more than an unreliable way to send telegrams. Hardly anyone back then imagined that speech and music might be transmitted.

Nicola Tesla, the archetype “mad scientist”

Nicola Tesla was born in Smiljan near Gospic in Croatia in 1856. In college he studied what was then the exotic field of electrical engineering. He once proposed to his professor that an AC generator could be built that would be simpler than DC power generation and which would have several other advantages. The professor ridiculed his idea mercilessly. Today we call these “alternators.” We use gigantic alternators to generate electricity in all large power plants. And we use little ones in our cars to recharge our batteries. When his father died, Nicola was forced to leave school and go to work. Like most electrical engineers of his time, he worked on DC motors and DC generators. At that time the DC motor was beginning to replace the belt and pulley as a means of powering industrial machinery such as looms and mine hoists.

Tesla migrated to America and arrived almost penniless. He even worked briefly as a ditch digger in order to eat. He applied for work with Edison who tested his skills by assigning him to fix a DC generator on a ship. Tesla rebuilt the generator right on the ship and made it produce more electricity than its original design. Tesla worked briefly for Edison, then he struck out on his own. He built his own small laboratory and worked on gadgets of all sorts. He soon acquired a reputation as a “science wizard.” He enjoyed putting on “magic shows” with giant sparks flying off his fingers and whirling fluorescent light bulbs. His reputation as a science magician encouraged him to put show business into everything he did. After reading his biography, it appears to me that his ability to gain awe and respect through showmanship eventually ruined his career. (1)

As money ran short, Tesla got a job with the Westinghouse Company and developed the alternator into a practical power generator. Tesla’s greatest contribution to the world was the power generation and distribution system which he demonstrated at a brand new power plant at Niagara Falls, New York. He invented the three phase AC alternators, transformers, and high-tension power lines that are still in use world-wide. After Tesla left Westinghouse, he set up his own laboratory in New York City to experiment with uses for radio frequency current.

The missed opportunity

Ship owners have probably always wished they could communicate with ships at sea. Until the late 19th century the fate of a ship might be totally unknown for months or even a year. When the ship finally sailed into homeport, the owner might suddenly learn that he was extremely wealthy. Or the ship might never return and the owner would have lost a huge investment. Being able to communicate a few hundred miles or even a dozen miles out to sea might be life saving in an emergency.

By 1900 scientists knew that “wireless telegraph” could communicate across the English Channel using giant transmitters and antennas, but no one had been able to transmit a message much farther than that. At that time J.P. Morgan was a financier and banker and one of the richest men in the world. Among his empire of enterprises he owned a fleet of ships. If a practical long-range telegraph could be developed, he wanted it on his ships. Marconi already had a good start on a ship-to-shore radio and had already demonstrated short-range ship-to-shore communication, both in England and America. In spite of that lead, Morgan approached Tesla who certainly had the knowledge and experience to develop practical radio communications. J.P. Morgan gave Tesla a big financial grant to do this work. Tesla set up a laboratory in Colorado Springs to invent long distance radio, or so he allowed Morgan to believe.

Unfortunately, merely talking to ships was boring to Tesla. Tesla wanted to develop what he called “The World Telegraphy Center.” Tesla planned to set up a communications center that could not only talk to ships, but also to everyone else on earth. His vision of what he was trying to build sounds to modern ears like a one-way Internet or perhaps CNN. He doesn't seem to have thought about the difficulties of handling all the messages in the world through one single gigantic low frequency transmitter. Back then, there were no Internet servers to organize all that message traffic into digital streams of information. Considering the operating frequency of his transmitters, his data rate would have been limited to a few kilobytes per second rather than the terabytes handled today by a single node on the Internet.

Tesla's radio transmitters were certainly adequate for transoceanic communication. But instead of also developing a sensitive radio receiver, Tesla spent nearly all his effort developing ever larger low frequency radio transmitters. His transmitters were so powerful, he experimented with transmitting electric power as well as information. Tesla proposed using tuned coils to energize fluorescent light bulbs miles away from his transmitter. Yes, his idea worked, but only at an extremely low efficiency. Yes, the lights glowed just as he said, but damp soil, cows, people, barbed wire fences, and every other electric conductor within range would be heated with wasted energy, just like a microwave oven.

Tesla built an enormous "Tesla coil" that produced radio frequency sparks 60 feet long. Always the showman, Tesla liked to be photographed sitting among the sparks and fire, while calmly reading a book. Actually, he used double exposures to create the illusion of sitting among the sparks. Tesla's machine was so huge and had such unique capability that the U.S. Air Force built a copy of it 80 years later for research.

With all this dramatic futuristic activity, Tesla never got around to building the dinky ship-to-shore radio that Morgan was paying him to develop. When he gave Morgan a progress report, Tesla tried to sell Morgan on his futuristic schemes. Morgan was furious at him for not sticking to the assignment and had little interest in any of Tesla's ideas. Morgan did however force Tesla to assign him the ownership of any useful patents that might arise out of the work. Morgan was not known for generosity.

After Morgan gave Tesla a tongue lashing, he also gave him a second chance. But instead of getting serious about ship-to-shore communication, Tesla blew the money on building his "World Telegraphy Center" out at Wardenclyff, Long Island, New York. It was an imposing building with a huge tower housing the Tesla coil transmitter. The communications center came to nothing and Morgan stopped the funds. Thereafter Tesla lived at the Waldorf Astoria Hotel in New York City and became a sort of self-absorbed lounge lizard. He dressed in a Tuxedo and top hat and mooched off his friends.

In the following decades Tesla dabbled in inventing and came up with several interesting devices that were almost good enough to become standard technology. For example, he designed a "bladeless turbine" heat engine, the functional equivalent of a steam engine or the internal combustion engine. There are few successful heat engine designs that are fundamentally different, so inventing a new one was an intellectual triumph. Unfortunately, Tesla's heat engine was not as efficient as other methods and, so far, there have been no good uses for it. He also developed a speedometer gauge that was excellent and was used in several luxury cars. Converting the speed of a rotating shaft into smooth, linear needle movement is much harder than it looks. Unfortunately Tesla's method was more expensive than the meter design that eventually became universally used for that purpose.

Tesla ended up as a lonely old man feeding pigeons in a third rate hotel in New York.

After he died in 1943, it turned out that he had paid his rent for several months by giving the hotel manager a “death ray” to hold as collateral. Tesla told the manager the death ray was worth $10,000. The ray gun was actually a Wheatstone bridge, a sensitive resistance-measuring device commonly found in electrical labs.

Marconi gets the job done

Guglielmo Marconi was born into a prosperous family in Bologna, Italy on April 25, 1874. He was educated in Bologna then later in Florence. He studied physics at Leghorn College. He was fascinated by Hertz’s discovery of radio waves and he became interested in wireless telegraphy in 1890. Starting in 1894, Marconi worked at home building prototypes in his basement.

Today most of us think of a radio receiver as a kind of amplified stethoscope that lets us listen in on the hidden world of the radio spectrum. In Marconi’s time the main precedent for radio was telegraphy. This concept of one telegraph operator banging out telegrams to another operator using Morse code dominated Marconi’s vision of what he was trying to build. In conventional telegraphy the signal over the wire triggered a “sounder” which was a kind of electro-magnetic relay. The sounder made clickity-clack noises that the receiving operator interpreted as dots and dashes.

Similarly, Marconi’s first radio transmission to another room in the house rang a bell when the signal was detected. There were no headphones that a person listened to. Most early experimenters built radios that resembled radio control systems rather than listening devices. As the technology developed, the radio operator gradually became a vital part of the system. The operator’s skill and trained ears became responsible for most of the range and practicality of the system. A trained operator can hear Morse code signals that are no stronger than the atmospheric static. Unlike a simple bell system, an operator can copy one Morse code signal while ignoring another. It took a hundred years for computerized digital signal processing to exceed the ability of a trained radio operator and return to Marconi’s vision of a robotic receiver.

Radio detectors – An early challenge

The most popular early radio detector, the “***coherer***,” was invented by the English physicist Lodge. Coherers were first used with long distance wire telegraph lines. They greatly extended the practical range of a telegraph wire and it was natural that they would be applied to the earliest radio experiments.

A coherer was a small glass vial containing loose powdered carbon or iron filings. This powder contacted two electrodes in the vial. When a small voltage appeared across the powder, it would break down the contact resistance between the powder granules and cause the resistance of the coherer to suddenly drop. The decrease in resistance was used to cause current to flow through the sounder relay. Coherers were often built onto the frame of a sounder so that the vibration of the sounder would keep the powder loose, thereby continually resetting the coherer to its original state.

The set - reset action of a coherer resembles a modern silicon controlled rectifier. A small input current causes a much bigger current to flow. Unfortunately, just like a silicon-controlled rectifier, the current through the coherer doesn’t shut off by itself when the input is turned off. Because coherers turned on and off at rates below 20 cycles per second, the output from a coherer wasn't an audio signal that someone could listen to directly.

At first Marconi’s receiver sat on the table next to the transmitter. Then he was able to transmit across the room and then to other rooms in the house. As his range increased, he moved his operation into an unused granary behind his parents’ house where he could string up antennas. His next triumph was a transmission from the granary to the end of the garden, 100 meters away. During these years the existence of radio was widely known to scientists, but it was believed that radio waves were inherently line-of-sight, much like a signal lantern. Marconi had already observed that he was able to transmit to the receiver when it was behind walls and trees. Since he already knew the experts were wrong, he worked on the big question of whether radio waves could travel over mountains and perhaps over the horizon.

About this time Marconi must have graduated from coherers to some kind of rectifier detector. This detector produced an audio output that an operator could listen to directly with earphones. The earphone was already in use as a component in Alexander Graham Bell's telephone. The earliest rectifiers consisted of “crystals” which consisted of a piece of wire pressed against a crystalline chunk of sulfide ore. Crystal detectors are described in detail in chapter 4.

Marconi had a servant named Mignani. To test his receiver over distances, Mignani manned the receiver while Marconi sent test signals. One of Marconi’s breakthrough improvements was a directional antenna that focused his weak transmitter output directly toward the feeble receiver and thereby extended the range. Radio stopped being a toy the day that Marconi transmitted a test signal two miles over a hill. Mignani signaled back to Marconi that he had received the Morse code letter “S,” by firing a rifle into the air from the hilltop.

Radio reaches across the Atlantic

Following the success of his experiments at home, Marconi became obsessed with the possibility of transmitting a signal across the Atlantic. If he could do that, radio communication could cover the world. There was essentially no interest in radio in Italy. He was even unable to get a patent for his device. An Italian government minister told him that radiotelegraphy “was not suitable for communications.” Marconi moved to England where he patented his method of transmitting signals in 1895. In 1897 he was financed by the British Post Office to continue his experiments. Gradually the range of his transmissions was extended to 8, 15, 30, and 100 kilometers. In 1897 he founded the Marconi Wireless Telegraph Company, Ltd. in London. In 1899 he established a communications service across the English Channel.

Marconi built a huge transmitter, 100 times more powerful than any earlier transmitter, and set it up at Plodu, Cornwall in South-west England. One approach to building very large transmitters was to construct large, high-speed AC alternators. These resembled Tesla’s power generation alternators but ran at such high speeds that they produced a sinewave, not at 60 cycles per second like a modern power plant, but at low radio frequencies, 20,000 cycles per second. Marconi also built a complementary station at St. Johns, Newfoundland and on December 12, 1901, he received the first signals from across the ocean. Actually, modern analysis of the propagation conditions of that first communication suggests that the guys listening to the static may have been imagining Morse code that wasn't really there hiding in the hiss and crackle. Even if the first reports may have been premature, Marconi's team soon established real communication.

The British and Italian Navies promptly adopted his system and ship-to-shore radio became reality. By 1907 his system was available to the public as a transatlantic radio telegram service. Marconi was awarded the Nobel physics prize for 1909. In his later life he continued to experiment with short waves and microwaves. Marconi also briefly served as a statesmen. He was sent as a delegate to the Peace Conference in Paris after World War 1 where he signed the peace treaties with Austria and Bulgaria. He died in 1937.

Radio changes history

Until radios were put on ships, radio communication didn't make much difference to the course of history. Radio telegrams sent across the ocean or between cities were competing with undersea and overland cable telegrams. Ordinary hard-wired telegraph was just as fast as radio but was not vulnerable to atmospheric conditions. However, once radio transmitters were placed on ships, it was only a matter of time before radio was used to rescue the passengers and crew from a sinking ship. This first happened during the sinking of the RMS Republic.

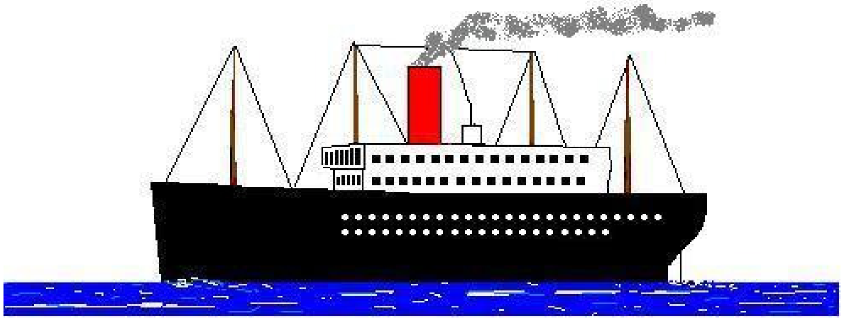

At 5:40 AM, Saturday January 24, 1909, the 15,000 ton passenger steamship RMS Republic was steaming off Nantucket, Massachusetts through thick fog. The Republic was outbound from New York. This ship had a lot in common with the later Titanic. The Republic was owned by the same British White Star Line and was considered "unsinkable." Its hull was divided into multiple compartments by watertight bulkheads so that several compartments would have to flood before the ship could sink. Also like the Titanic, the Republic carried only half as many lifeboats as needed for her 800 passengers and crew.

As the Republic motored through the fog she sounded her foghorn periodically and ship's lookouts listened for other foghorns. The crew heard a foghorn from another ship and responded by sounding a steam whistle. It was the convention at that time that, when ships exchanged whistles, both ships would turn right and thereby avoid the collision. There is some evidence that the Italian passenger ship "Florida" turned left instead. The Florida suddenly appeared out of the fog and crashed into the Republic amidships. Seven people were killed outright by the collision. The Florida struck the bulkhead between the two engine rooms on the Republic, thereby flooding the two largest compartments below the waterline. The engines had to be shut down which also terminated the electricity needed to run the pumps and the radio.

Jack Binns saves the day

The Republic's radio shack was a wooden cabin that had been tacked onto the upper deck. It was located where it was easy to connect to the antennas that were strung up in the ship's rigging. With amazing bad luck, the bow of the Florida sliced part way through the radio cabin pushing equipment aside and disabling the radio. The operator John (Jack) R. Binns was sleeping on a bunk next to his operating station. He said later that if he had been still working at his radio he would have been gravely injured. Binns managed to piece his transmitter back together, but his little cabin was now open to the cold and fog. Because the ship's electricity was off, Binns had to rummage around in the dark below decks to find batteries to run his radio. Without the ship's electricity, Binn's transmitter range was limited to about 75 miles. His best hope was to reach a station on Cape Cod, 65 miles away. Even Binn's telegraph key was broken. He had to hold it together with one hand while he used his other hand to send out history's first emergency SOS.

Actually, the official distress call at that time was not SOS but "CQD." "CQ" meant calling any station, just as it still does today on the ham bands. CQ was supposed to be short for "Seeking You." "D" of course meant danger or distress. Binns managed to contact the station on Cape Cod. Through them, he continued to call for help and talked to rescue ships for 12 hours while the Republic slowly settled into the sea.

Since the Florida was in no danger of sinking, the Republic crew transferred the Republic passengers to the Florida by lifeboat. Meanwhile rescue ships criss-crossed the fog trying to find them. In addition to foghorns, ships of that era were equipped with "submarine bells" which had a greater range underwater than sound from horns was able to travel through air. The submarine bell of the Republic was heard by the liner Baltic and used its sound to guide it the final few miles to the Republic.

In another account of the rescue there was no submarine bell and the contact was made when the Baltic shot off "its very last aerial rocket bomb." The crew of the Republic heard the explosion and gave them the correct bearing by radio. When the Baltic arrived, the remaining crew on the Republic were rescued, then all 2,494 passengers from both the Republic and Florida were transferred by lifeboat to the Baltic. 39 hours after the collision the Republic sank. Meanwhile rescue ships managed to tow the Florida to New York City.

The success of the Republic rescue was a huge news story. Jack Binns was an overnight national celebrity. Radiotelegraph operators were instantly transformed from curiosities into heroes. It's bizarre and even criminal that the White Star Line didn't seem to learn anything from the sinking of the RMS Republic. However, in 1912 the Titanic sinking and the partial success of the Titantic's SOS emergency call further elevated the status of radio communications. The steamship Carpathian sailed 300 miles to rescue the Titanic survivors the morning after the sinking. Another freighter, the Californian, lay at sea anchor just 10 miles away. The captain of the Californian didn't want to risk steaming at night through the icebergs - smart man! The Californian radio operator sent an iceberg warning to Titanic, but the Titanic radio operator told the Californian to get off the frequency because he was handling telegrams for the passengers. The Californian radio operator shut down his radio and went to bed. The night crew on the Californian could plainly see the Titanic but didn't realize the Titanic was in trouble. When the Titanic shot red emergency rockets into the air, the Californian crew thought the rockets were fireworks to amuse the wealthy passengers.

The dawn of amateur radio

In the early days of radio anyone who wanted to get on the air could build his own transmitter and just do it. Frequency assignments were based on whoever was on the air first with the biggest transmitter and the best antenna. Basement inventors began building their own radio equipment and amateur radio was born. In those days there were no licenses so amateurs at first just gave themselves call letters. All radio communications in the beginning were low frequency, 200 meters wavelength and longer. This meant that all communications were happening at what is today AM radio (1,700,000 Hz to 550,000 Hz) and below. Higher frequencies were inefficient to generate and it was believed that higher frequencies were only good for local communication.

When modern amateur radio operators read descriptions of early radio transmitters and antennas, we are impressed by the industrial scale of the commercial equipment. For example, the rotary spark gap transmitter on the Titanic was rated at 5,000 watts. It was so huge and made so much noise and ozone stink, it was installed in a separate room from the receiver. With all that power and size, its reliable range during the day was only 250 miles. At night the range could be as far as a thousand miles. From our modern viewpoint, it's apparent that the biggest weakness of early day equipment was the insensitivity of their receivers. Back then amateur receivers were just as bad and probably worse. Moreover, amateurs were unable to compete with the power used by commercial stations. Consequently, hams were lucky to talk to the next county, let alone any significant distance.

Because early ham radios were limited to short range, radio amateurs banded together to form "relay nets" so that messages could be relayed to destinations many hundreds of miles away. This organization became formalized as the American Radio Relay League in 1914. During World War I the US government banned amateur radio as a national security measure. After the war in 1919 the American Radio Relay League under the leadership of Maxim Percy lobbied the government to allow hams back on the air. The government, especially the military, was unsympathetic. To placate them, the government eventually gave the hams all the "worthless" frequencies above 200 meters. That was roughly everything above present day AM radio.

Flemming develops the vacuum tube diode detector

The vacuum tube diode had actually first been built by Edison. During Edison's work on the electric light, he perfected the technology to put electrodes and filaments in evacuated glass bulbs. Edison was also the first to notice that electrons can flow from a hot filament across empty vacuum toward a positively charged electrode called a "***plate***." However, the electrons could not flow from the plate back to the filament. As a result, the diode behaved as a one-way check valve. Edison experimented with this and wrote about it. This effect became known as the "Edison effect."

Flemming was a British inventor who was intrigued with Edison's discovery and applied it to the detection of radio signals. He worked with Edison who shared his data on the Edison effect. A radio signal received on an antenna consists of a high frequency sinewave voltage that switches from positive to negative and back again hundreds of thousands (or millions) of times per second. These rapid alternating (AC) currents can't power a headphone until they are converted to lower frequency bursts of DC. By passing radio signals through a vacuum diode, one polarity of the signal is "sheared off" leaving just the polarity of the flowing negative electrons. When Morse code signals are received, this results in bursts of DC current that can be used to power a sounder, headphones, or other kind of transducer.

Flemming's detectors became known as the "Flemming Valve." In England thereafter, all vacuum tubes became known as “valves.” In terms of sensitivity, the vacuum tube detector was not more sensitive than a crystal detector, but it was far more rugged, reliable, and versatile. Later on, when it was combined with the vacuum tube amplifier, its sensitivity was vastly improved.

Lee DeForest launches modern electronics

Tesla was not the only radio pioneer with an ego problem. In many ways DeForest had a personality similar to Tesla’s. DeForest was industrious and gifted. While he was a young man he developed several important inventions. One of his cleverest was a telegraph multiplex system that allowed one wire to carry up to six telegraph messages simultaneously. It’s hard to imagine how it could have worked. All he had to work with were switches, relays, transformers, and motors. Synchronizing such a signal separation system is easy with a microcomputer, but doing it with whirring motors and switches is amazing. De Forest’s most important invention was the triode vacuum tube, which he called an “audion.” The audion was the functional equivalent of the modern transistor. DeForest’s triode vacuum tube was essentially a "Fleming valve" with a grid placed between the filament and the plate. DeForest used the audion to make receivers more sensitive by amplifying weak audio and radio signals.

Actually Edison built the first triode while he was studying “the Edison Effect.” Edison added a grid-like electrode between the filament and the plate of his diode to learn more about the flow of electrons across a vacuum. Although Edison made his measurements, he didn’t notice that the grid was like a sensitive “gate” that could control the current flow to the plate.

The grid can be compared to a matador’s one-pound red cape steering a 2,000 pound bull. When little currents are able to control big currents, the device is said to “amplify.” Little signals coming into the grid can modulate the big current into a larger, “amplified” version of the original signal. Notice that the original signal is not “inflated” in some way, but rather, the amplifier just directs the generation of a larger copy of the original signal. Notice that the copy may or may not be a good likeness of the original.

DeForest successfully applied his triode vacuum tube to radio receivers. The triode meant that weak signals could be amplified to a level high enough so that after detection they could drive a loudspeaker. Once amplifiers became available, nearly every kind electronic device known today became at least theoretically possible. Eventually even television was implemented with the descendents of the audion vacuum tube. During World War II the first computers were built using vacuum tubes.

The vacuum tube oscillator cleans up radio transmissions

A few early high power systems used alternators to generate low frequency radio signals. However, the frequency of an alternator is limited by the speed of a mechanical, rotating magnet. In practice, alternators couldn't get much higher than audio frequencies, 20 to 50 KHz. An advantage of an alternator was that it generated a pure sinewave signal. Unfortunately it could only cover the very bottom of the radio spectrum.

In contrast, the vacuum tube amplifier could amplify its own output, causing a self-sustained oscillation at any frequency up to hundreds of MegaHertz. Like an alternator, a vacuum tube oscillator could generate clean sinewave signals that were confined to a single frequency. Up until that time, radio transmissions from all high frequency transmitters were generated by discharging electric sparks, then letting the radio frequency signal components ring in a tuned circuit. "Spark gap" transmitters made a noisy, hissing signal that splattered energy all over the band.

Have you ever heard the clicking static in an AM radio from a passing automobile ignition? Similarly, lightning causes a pop or crash of static that can be heard over the entire radio spectrum. Other than having a filter to limit the noise to one band, early radio transmitters were like little lightning radio noise generators. These early sparkgap transmitters wasted power and frequency space. Because of the availability of vacuum tube oscillators, spark gaps became unusual in the early 1920's and were totally banned in the United States in 1927.

If an inventor is persistent, it's sometimes possible to invent wonderful things without really understanding why they work. This was the case with DeForest's audion. In the short run, ignorance can be OK, but in the long run the inventor had better figure out exactly what he has developed. Specifically, he must be able to explain how it works in his patent application.

Unfortunately, DeForest blew it. His patents failed to give valid explanations of how his vacuum tube devices worked. A dozen years after the invention of the audion, DeForest's patents came into conflict with later patents that were accurately explained. In court DeForest was literally unable to explain how a triode amplified. It is completely understandable how he might have been naive at the time of his inventions. But considering that the vacuum triode was his crowning achievement, it isn't flattering that he never bothered to keep in touch with the field of electronics well enough to find out how his own invention worked. Forty-five years after his invention of the audion, De Forest was still strutting around in formal clothes grandly referring to himself as “The Father of Television.”

Reginald Fessenden transmits speech

Professor Reginald Fessenden is an almost unknown hero of radio development. He is truly one of the most remarkable radio pioneers. Although Marconi first demonstrated transatlantic communication, Fessenden was the first to offer it as a regular service in 1906. Fessenden's most amazing accomplishment was the transmission of speech. He is not remarkable so much for what he did, but rather when he did it. In December 1900 he transmitted a voice signal to his assistant Alfred Thiessen over a distance of one mile from his laboratory on Cobb Island in the Potomac River. “One-two-three-four ... Is it snowing where you are, Mr. Thiessen? Telegraph back if it is.”

“Yes, it is!” tapped out Thiessen over a telegraph. From a modern perspective Fessenden's invention came at least 15 or 20 years before the parts were available to do it gracefully. What Fessenden needed was high power audion tubes. However, he didn't know that. So he transmitted AM modulated voice transmission using spark gap transmitters. That sounds impossible. And perhaps because spark gaps were “dead end technology,” it isn't totally surprising that Fessenden has been largely forgotten. The only articles on Fessenden that I have ever read don't describe his apparatus clearly. However, he seems to have succeeded by means of three major breakthroughs and many minor ones.

Fessenden smooths out the sparks

A smooth, pure carrier wave is needed to produce clear AM modulation, but there is nothing smooth about static from a spark gap. Fessenden reasoned that, if the static crashes occurred often enough, the frequency of the buzzing racket would be too high an audio frequency for human ears to hear. So instead of sparking a few dozen or a hundred times a second, Fessenden's generator sparked 10,000 times per second. Then (apparently) he filtered the heck out of the signal using inductor/ capacitor resonant circuits. The transmitter had its own steam engine power plant and was probably quite powerful. The steam engine had to be running at full speed before the speech became intelligible.

Using high frequency noise to get rid of low frequency noise is reminiscent of the modern "Dolby sound" noise canceling principle. Rather than fight the noise inherent in any analog recording system, the Dolby system deliberately modulates the music with a high frequency sinewave at about 25,000 Hz to obliterate the noise.

I read another account of Fessenden's work in which his voice transmitter was described as a high-speed alternator. Oh, well. Transmitting speech with a high-speed alternator also sounds difficult. Garbled technical descriptions are a serious problem when studying the history of radio. It is hard to figure out exactly what they did.

Fessenden invents the barretter

Fessenden also invented a sensitive new detector that he called a ***barretter***. "Barrette" means "a small bar" in French. Modern barretters consist of a microscopic platinum wire that heats and cools rapidly when bursts of tiny radio frequency currents pass through the wire. As the wire changes temperature, its resistance changes rapidly and this can be used to modulate a DC current passing through a sensitive telephone earphone. Barretters are still used today as microwave detectors.

What I don't understand is how the professor got enough signal strength out of a 1900-era telephone microphone to modulate a powerful transmitter. In my opinion that would have been his third amazing breakthrough. In a commercial amplitude modulated (plate modulated) transmitter from the 1930s and later, the audio sound from the microphone had to be amplified up to half the power output of radio transmitter. For example, for a 1,000 watt transmitter, a 500 watt audio amplifier was needed. There was no way to build such a thing in 1900.

Perhaps he invented the "magnetic modulator." In the 1920s some phone transmitters used a transformer to impress amplitude modulation directly onto the RF signal. The microphone modulated a DC current on the transformer primary while the transmitter ground for the antenna went through the transformer secondary. By saturating the iron in the transformer, the current flowing through the secondary can be radically changed with a small signal, thereby providing the needed amplification. Fessenden must have been an obsessed genius to broadcast speech successfully with steam engines, iron bars, copper wire, and spark gaps.

At the time Fessenden's achievement didn't go unnoticed and he was able to raise money to start a broadcast company. His first public demonstration consisted of speaking and playing "Oh Holy Night" on the violin on Christmas Eve in 1906. However, his commercial progress was slow and by the time AM broadcast was beginning to work well, nearly everyone was stealing his ideas. The Marconi Company eventually licensed his patents in 1914. From a modern perspective, Fessenden's biggest problem was his lack of viable business plans. He invented neat stuff, but had difficulty getting paid for his work.

Edwin Howard Armstrong

Another engineer who understood his own inventions was Edwin Armstrong. He studied electrical engineering during World War One, then, right after graduation, he produced the first of his three great inventions, ***the super-regenerative receiver***. In the early days of radio, it was too expensive to build receivers with large numbers of large, costly, power-hungry vacuum tubes. Armstrong invented a way to use feedback on a vacuum tube amplifier to increase the sensitivity of the receiver by an order of magnitude. Super-regenerative receivers were crude and had to be adjusted just right to avoid an unpleasant squealing noise on top of the stations you were trying to hear. However, when super-regenerative receivers began to be used, the range of radio broadcasts soared to a thousand miles and more.

In spite of their severe limitations, ham radio operators commonly made homemade super-regenerative receivers as late as 1960. “Super-regens” were primitive, but they were a stepping-stone to the next generation of receivers. In the 1920s Armstrong developed ***the super-hetrodyne receiver.*** This is the basic design used today in most modern receivers, from television to cell phones. When I was a young ham, I built a Knight "Ocean Hopper" super-regenerative receiver from a kit. I only bought it because it was cheap. The high-pitched noise of the super-regen was so obnoxious, I couldn’t wait to replace it with a high quality super-hetrodyne communications receiver.

The TRF receiver didn't cut it

To achieve the required amplification without super-regeneration, the obvious solution was to put several radio frequency tube amplifiers in series. These radios were called ***tuned radio frequency (TRF) receivers*** but they were never very popular. Each amplifier in the string had to be tuned separately to the desired station. This meant that the first TRF radios literally had three or four station tuning knobs that all had to be tweaked independently. Alternatively, in the best TRF radios the tuning capacitors were ganged together with gears and shafts so that they always tuned to the same station simultaneously. This method was extremely hard to synchronize and calibrate so that each tuned coil would track precisely over the entire frequency band. Another problem was that they were inherently "low Q" and tended to receive more than one station at once. The only practical TRF receivers were either extremely expensive or were designed to receive just one frequency.

The super-hetrodyne receiver

Armstrong’s solution to the gain problem was to convert the incoming radio signals to a constant ***intermediate frequency***, an ***IF***. The IF frequency remains the same no matter what station the radio was tuned to. In other words, ***a superhetrodyne is a single-frequency TRF receiver with a tunable frequency converter on the front end.*** The IF strip is tuned up just once at the factory. It never needs to be tweaked again. Another important advantage was that the signal on the IF strip could be filtered so that it would only admit a bandpass equal to the actual width of the signal. For example, an AM radio receiver might have an IF bandpass of 20,000 cycles per second (one cps = one Hertz). However, a Morse code signal only needs 100 Hz of bandwidth or less. So when a narrow bandwidth filter is used in the IF, the receiver can select just one signal out of many that may be crowding the band. In contrast, a TRF receiver or even a super-regenerative receiver or may force you to listen to 50,000 Hz of bandwidth at once.

The super-hetrodyne generates the IF by means of a ***local RF oscillator, an "LO."*** The principle can be illustrated with audible sound. When two different frequencies of sound are mixed together, the sound waves cancel and reinforce each other generating frequency components that are the sum and difference of the two signals. For example, if you hit two adjacent keys on a piano simultaneously, the sound is discordant. That's because you're hearing those sum and difference frequencies. In another example, twin-engine, propeller-driven airplanes are equipped with engine speed synchronizers so that the engines don't make an obnoxious "WAH -WAH- WAH" beat frequency sound. This sound is the difference in frequency between the two speeds of the engines. The same principle works with electrical sinewaves at radio frequencies. Unlike the annoying audible examples, in radio the beat frequency is the desired product and that was Armstrong's invention.

A super-hetrodyne is tuned, not by tuning a sharp filter on the antenna, but rather by tuning a sinewave oscillator that is offset from the signal you wish to hear. The amount of offset frequency is equal to the IF frequency. For example, in a household FM radio, if you wish to tune to a station broadcasting on 100 MHz, then the local tuning oscillator is tuned to 110.7 MHz. The difference between the two frequencies is 10.7 MHz which is the IF commonly used in FM radios. Because the tuning oscillator generates only one precise frequency, and because the IF filtering can be quite narrow, the tuning of a superhetrodyne receiver can be extremely selective. The superhetrodyne is made sensitive by putting several IF amplifiers in series. Or, as explained earlier, you can think of the IF amplifiers as a single frequency TRF receiver.

By the late 1920s spark gap transmitters were banned and replaced by vacuum tube sinewave oscillators. The sinewave oscillators generate just one, discrete frequency. After this improvement hundreds of Morse code signals could share a band without interfering with each other. And using Armstrong's superhetrodyne, the receiver could select just one of these Morse code signals.

Beat Frequency Oscillators – where the musical Morse code tone comes from

Although a simple superhetrodyne receives AM broadcasts beautifully, Morse code is essentially inaudible because a pure sinewave signal is unmodulated. During the 1920s Morse code transmitters were often modulated with a mechanical motor driven switch device that made the Morse sound like a buzz and more like an old spark gap. Alternatively, they sometimes deliberately ran the transmitter on unfiltered DC from a rectifier so that the signal was modulated with AC hum from the power line.

To hear Morse code with a superhetrodyne, another oscillator in the receiver called a ***Beat Frequency Oscillator (BFO)*** is needed to produce the musical sound. For example, my first short wave receiver was one of those ancient all-band radios often found in living rooms in the 1940s. The receiver worked fine for receiving foreign AM radio stations like Radio Moscow. However, when I tuned to the ham bands, the Morse code signals were inaudible or just thumping noises. To receive Morse code I had to place a small table radio on top of the big shortwave radio. I used to tune the little radio until I could hear a harmonic of the table radio's tuning oscillator (local oscillator) in the big receiver. This signal beat with Morse code signals and made them audible. This was extremely finicky to adjust and barely practical. A communications receiver of course has a built-in BFO that typically oscillates at an RF frequency 500 to 700 Hz away from the IF. This produces the clear Morse code whistle we are accustomed to hearing, no matter what frequency the local tuning oscillator is tuned to.

Getting rid of the atmospheric static

Radio communication was essentially 100% Morse code until after World War I. Suddenly it became common to hear voices over the radio. In 1921 the first commercially successful Amplitude Modulated voice broadcasts began. With AM broadcasts, the transmitter was not turned on and off like Morse code, but rather the transmitter was left on continuously. The transmitter power was raised and lowered in time with the music or voice audio frequency. This process impressed or "modulated" voice and music onto the signal. In other words, with AM modulation the information content is proportional to the momentary strength of the signal.

So long as the strength of the signal is high and there is no interference from lightning storms, big DC motors, or nearby automobile ignitions, amplitude modulation (AM) radio works quite well. By 1930 AM radio was a standard appliance in American households. But with all that crackling and popping noise on weak signals, AM radio is never really high fidelity except when tuned to strong, local stations.

Atmospheric static is a natural noise signal that has the same amplitude modulated form as man-made AM signals. Therefore it is impossible to get rid of the static without changing the method of modulation. Edwin Armstrong secluded himself in his laboratory in the Empire State Building in New York City and worked to find a way to eliminate the static in voice broadcasts. Armstrong needed to create an audio modulation that could be impressed on a transmitter signal without imitating the natural noise modulation produced by lightning and static discharge.

Armstrong eventually hit upon the idea of modulating the frequency of the signal rather than the intensity of the signal. In other words, as the announcer talks, the frequency of the transmitter moves up and down in time with the audio frequency and amplitude of the sound. The signal amplitude never changes**.** ***Frequency Modulation*** solved most of the noise problem and was the origin of the FM broadcasts we listen to today. Armstrong had been working for the Radio Corporation of American that manufactured superhetrodynes. Because of the superhetrodyne, RCA with its subsidiary, the National Broadcast Company, dominated AM radio for as long as that patent lasted. RCA was run by David Sarnoff who failed to reward Armstrong's contribution to the company. Armstrong left RCA and started his own FM radio broadcast network.

Sarnoff needed FM radio to transmit the audio signal with television so he simply stole the invention from Armstrong. Sarnoff, being able to hire more lawyers, was able to play the patent war two ways in his favor. His lawyers persuaded the court that RCA had invented FM, not Armstrong. Sarnoff also persuaded the FCC to force FM stations to broadcast on VHF frequencies at low power that restricted them to local broadcast. This prevented FM from ever being used by long-range stations like the 50,000 watt "clear channel" AM stations that are still scattered around the U.S. Ultimately Sarnoff won all the battles and ruined Armstrong. Armstrong, who loved to climb on high radio towers, ended his life by jumping out of a 13 story window in 1954.

The transistor miniaturizes electronics

The bipolar transistor was invented in 1947 by Shockley and Bardeen while working at Bell Laboratories. In function, the transistor can be thought of as a "miniature triode vacuum tube." Unlike a tube, a transistor consists of a tiny lump of semiconductor crystal with three wires attached. Like a triode, a control gate called the "***base***" allows a small current to control a much larger current that flows from the "***emitter***" wire to the "***collector***" wire. Unlike tubes, there is no vacuum chamber, no heated filament, no relatively high voltage, and no separate power supply needed to light the filament.

In vacuum tubes, the control grid of a vacuum tube must always be referenced to the negative pole of the circuit. That is, the grid is always operated at just a few volts different from the voltage on the filament (cathode). The vacuum tube plate potential is usually quite high, typically hundreds of volts, and is always positive polarity. In contrast, transistors can run on as little as one or two volts and can be built in two polarities. The control base can be referenced to either the negative pole, (***NPN transistors***) or the positive pole, (***PNP transistors***.) Because they are available in complementary designs, the two types can be used together to form compact, high gain circuits with fewer additional components such as resistors and transformers.

The first transistors were fragile devices called ***point contact transistors*** that never appeared in consumer products. (A homebrew attempt to build point contact transistors is described in chapter 4.) The first widely used transistors were made from germanium and not silicon. Early germanium transistors could only tolerate tiny power levels. For example, the 2N35 transistor would burn out if it dissipated more than 35 thousandths of a watt. In the 1950s they were sold to amateurs for experimentation, but it was difficult to get them to work before they burned up. I bought one for $5 and promptly ruined it. Today that amount of money would be like $50 each. Then in 1960 the Texas Instruments Company perfected silicon transistors and the dominance of the vacuum tube was doomed.

Radio covers the globe

The Sony Company in Japan realized that silicon transistors presented a unique opportunity. They jumped on it and revolutionized AM radio. They made tiny battery powered “transistor radios” that cost a few dollars and fit in a shirt pocket. Because they didn’t need power from the wall and the batteries were small and cheap, suddenly even the poorest people on Earth could afford transistor radios.

In the 1960s another type of silicon transistor, the Field Effect Transistor (FET) appeared. FETs are also three wire devices. The control gate of an FET is actually called the “***gate***” and the gate uses tiny voltages, rather than tiny currents, to control the big current flowing from the “***source***” lead to the “***drain***” lead. Like bipolar transistors, FETs come in two polarities called “***N-channel***” and “***P-channel***.” Today FETs are the basis of most integrated circuits used in computers and, as we’ll see later, they are also valuable in radio circuits.

Becoming a ham radio operator

My serious shortwave listening began during the cold war. My big “all band radio” allowed me to tune the shortwave bands. This radio was three feet high, two feet wide and packaged in a beautiful wooden cabinet. But compared with a real communications receiver, this living room radio was extremely limited. It had poor selectivity and only covered a few ham bands. It had no “band spread” control so the hambands were only about 1/8 inch wide on 4 inches of dial. On the other hand, it wasn’t worthless. I was able to hear occasional AM phone conversations between hams. Later I bought a World War II surplus “morale radio.” These were shortwave receivers that were issued to the American troops so they could listen to broadcasts from home, Tokyo Rose, or other AM modulated stations. Morale radios also had no “beat frequency oscillator” and I could not receive Morse code without resorting to the second-radio-trick described earlier.

I learned Morse code in the Boy Scouts. One requirement for our Boy Scout First Class badge was to send and receive Morse Code using a signal flag. The flag was waved over one's head to the left for "dash" and to the right for "dot." Our Boy Scout manual admonished us to remember what the Dutchman said, "Dots right!" Today's Boy Scouts don't have to learn Morse code. That's a shame. In an emergency, the ability to communicate by tapping through a wall or waving across a canyon might not be a trivial skill. Prisoners throughout the world often communicate by tapping messages through walls using a universal code. The prisoners' code translates the alphabet into tapping in which the letter "A" equals one tap, "B" equals two taps and of course "Z" equals 26 taps. Good Grief! I guess prisoners have lots of time on their hands.

Several of my friends were also interested in shortwave and we started a shortwave listening (SWL) club. For code practice my buddy Eric Raimy (soon to be KØDUA), rigged up a telegraph system of buzzers that communicated between his bedroom, the downstairs hall closet, and the basement. We three middle school students sat at our posts and talked with our slow Morse code until we got our speed up to 5 words a minute and could pass the Novice class ham radio test. A local school principal, Glenn Johnson, WØFQK, taught us about ham radio administered the tests to us.

If you wish to get a ham license, you need to find a local ham radio club in your area. In America ham clubs generally conduct classes for new hams and give the license exams. Check the Internet for local clubs. Failing that, if you live in the USA, check the American Radio Relay League website, www.arrl.org or call (888) 277-5289. The ARRL will help you in many ways. They can provide you with study materials, magazines, books, and hams to contact in your area.

Ham radio in the last 80 years

Like electronics in general, amateur radio has exploded into many different facets since World War II. From a hobby that was originally just Morse code, it has exploded into a long list of capabilities and activities. Each decade has added more and more variations in the ways and methods for amateurs to use radio technology. As a rule, new technologies were first demonstrated ten years before they became commonplace. Of all the ham modalities that have ever been used, only spark gap transmitters are completely extinct. However, hams rarely use AM phone today.

1920s Continuous wave Morse code and early AM radio. (Vacuum tubes and the end of the spark gap transmitter.)

1930s AM phone, Very High Frequency (VHF) 54 MHz and above.

1940s Radio Teletype (RTTY), VHF FM communication, 2-way ham radios in cars.

1950s Single Side Band Phone (SSB), amateur television, slow scan television (a kind of radio Fax), UHF communication, 220 MHz and above, Directional Beam Antennas, radio transmitter hunts.

1960s Amateur microwave communication. Easy-to-use SSB HF transceivers. Moon bounce and meteor bounce communication.

1970s Amateur Satellite communication, VHF/ UHF relay stations.

1980s hand-held transceivers, computerized station log keeping, AMTOR packet message handling (a sort of radio e-mail)

1990s Spread Spectrum communication, radio control, robot beacon stations, PSK-31 messages (computer instant e-mail), IRLP (Internet-connected VHF relay stations), QRP (low power) stations, and amateur radio radio-controlled models, TV on small rockets.

2000s Software controlled radio receivers, complex digital signal processing, remote internet controlled stations, GPS locating for mobile transmitters (APRS) and, in general, the application of all the technologies found in cell phones and computers.

****

From the point of view of variety of new equipment, amateur radio is booming. On the other hand, thanks to the Internet, fax machines, and cellphones, much of this technology can now be used by people who have no license and no interest in how the equipment works. On the other hand, some aspects of ham radio remain unique to this hobby and haven't changed. These are:

1. Morse code.
2. Exchanging QSL (contact confirmation) cards,
3. Homebuilding equipment.
4. Mastering new or exotic modes of communication, such as satellite relay or unusual signal propagation modes like moon bounce.

The future of ham radio frequency bands

The existence of ham radio as a hobby is totally dependent on our governments allowing us to transmit. For this we need allocations of the frequency spectrum. But every day there are more and more commercial uses for wireless communication. These applications fall into two categories. Short-range communications, so called "Part 15 applications," transmit less than 100 feet. They link printers to computers, support cordless phones, open garage doors, and unlock automobiles. Suitable frequencies can be low, say 1 Megahertz and below. Or they can use very high frequencies, 40 MHz and above. The high frequency (HF) range in between 1 and 40 MHz is not desirable for these applications because, when atmospheric conditions are right, signals from great distances, even the other side of the world, can trigger a local device.

For example, when the first Soviet Sputnik satellite was launched in 1957, it broadcast on 20 MHz. Every time the satellite went over the United States, many homeowners discovered their garage doors mysteriously cycling up and down. Through the use of selective receivers and digital coding, most of this kind of radio interference has been eliminated. But when the frequency bands become crowded with signals, even a sophisticated receiver will eventually be paralyzed by the interference and won't respond to the correct code.

Cell phones, global positioning, and Internet links transmit over distances of a few miles or up to satellites to provide reliable communications. The best frequencies for these applications are above 40 MHz because random signals from the rest of the world usually won't interfere. At 500 MHz and above, signals from over the horizon will almost never interfere. These frequencies can be relied upon day and night for dependable communication. In the modern era it's these UHF and microwave frequencies that have the most commercial value.

Hams have a few frequency bands in this high-value part of the spectrum. So far we have been allowed to keep them by sharing them with the military.

Thank goodness for the erratic nature of HF

Since the 1920s the most important ham bands have been the HF frequencies between 1.8 and 29.7 MHz. The good news for amateurs is that it is the unreliability of these bands that makes them attractive to us and unappealing for commercial users. When we hams turn on our receivers, we really don't know what we're going to hear. There may be just static or there may be two local guys discussing their golf game. Or we might hear a ham in Mongolia looking for a chat. It's like going fishing. That's the fun of it.

Bad signs for the future

As ham radio equipment becomes more complex, fewer and fewer hams understand how it works. It's fair to say that the vast majority of hams today are overwhelmed by the complexity of the equipment they use and no longer make a serious attempt to understand it. As a result only a tiny minority attempt to build their own. American FCC licensing for hams reflects this trend. As fewer hams build their own equipment, ham radios have become just another form of consumer electronics. The FCC has become more concerned about manufacturers building fool-proof equipment than it is about hams understanding their radios.

Forty years ago, ham radio license exams in America were administered directly by the Federal Communications Commission in somber and intimidating government offices. The tests were devoted to the technical details needed to be sure hams knew how to keep their homemade transmitters operating within assigned bands. Typical questions included calculations on crystal specifications to determine whether a given crystal could be relied upon to stay within a particular ham band. The privilege of using all the ham frequencies was only granted to the best-trained hams. They had to demonstrate 20 words per minute Morse code speed and take a complex test that included drawing circuit diagrams of various kinds of transmitters and calculating design parameters. Today the Morse code requirement has been eliminated altogether. The tests are now multiple-choice exams given by ham volunteers. In preparation for the exams, modern hams study the actual tests, rather than the material covered by the tests. In terms of engineering knowledge, the modern ham is a lightweight compared with hams 50 years ago.

Another alarming sign is that the number of ham radio operators is relatively fixed and the average age of hams continues to rise. It's a logical conclusion that eventually our frequencies will be taken away for lack of interest. Ham radio will have become another historical pastime in the same category with quilting-bees and buffalo hunting.

A recent threat to ham radio is the broadband power line initiative (BPL). Power companies wish to make money by using their power lines as Internet connection conduits into every house that receives electricity. Unfortunately the power lines leak this radio noise into the sky and produce a strong buzz of noise that would obliterate all weak to moderate radio signals between 2.0 and 80 MHz. This radio noise obliterates all but the most powerful ham radio and shortwave signals. So far, BPL installations are limited to just a few cities and the ARRL has been fighting the FCC to make the government enforce their own regulations on interference to FCC licensed stations. As of 2010, the hams seem to be winning this battle. Similarly, the radio racket from plasma TVs and other modern sources increases the background static and makes ham bands less usable than they used to be. The difference in static between the city and living out in the country is dramatic.

Mastering the technology

As civilization's technology becomes more sophisticated, the knowledge base among our population becomes more and more fractured. Although the amount of knowledge that individual educated people have in their heads may be the same, each citizen knows more and more about less and less. Operating modern ham radio transceivers resembles operating a complex VCR. Yes, the equipment can handle all the modern modes and frequencies. But first you must read the manual and push 48 buttons to select all the right menus and options. It's not easy to operate one of these do-everything wonders. But when you've succeeded, you really haven't learned much about electronics.

Modern radio transceivers remind me of integrated circuits. Transceivers are packed with dozens (or hundreds) of integrated circuits. Some of these contain literally millions of transistors. Like the transceiver as a whole, an integrated circuit cannot be fixed, its contents are mystery, and it is usually a black, rectangular solid with "feet" or leads. Even with the service manual of a modern radio transceiver in front of you, it's hard to get more than a general idea of its block diagram and how it works.

But surely the engineers who designed these wondrous modern radios know how they work! No, not really. Perhaps there are a handful of engineers in the world who have a good grasp of most the technologies in a modern transceiver. But I wouldn't bet on it! Each engineer specializes in assembling or programming modules that are bought from other factories. The modules are sealed and can't be repaired. Exactly what's inside those modules is probably as much a mystery to them as it is to the rest of us.

Over the last century, a radio operator's span of knowledge has continued to shrink. A hundred years ago, the early radio pioneers were not only on the air testing radios, they were also working on the materials to build the components for their radios.

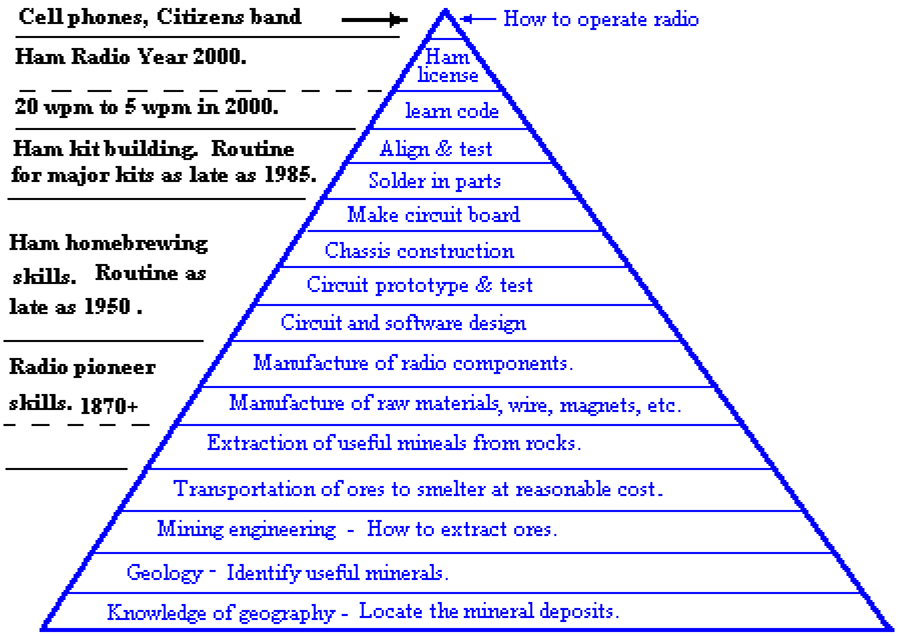
Cell phones, Citizens band
Ham Radio Year 2000.
20 wpm to 5 wpm in 2000.
Ham kit building. Routine for major kits as late as 1985.
Ham homebrewing skills. Routine as late as 1950 .
Radio pioneer skills. 1870+
How to operate radio
Ham license
learn code
Align & test
Solder in parts
Make circuit board
Chassis construction
Circuit prototype & test
Circuit and software design
Manufacture of radio components.
Manufacture of raw materials, wire, magnets, etc.
Extraction of useful mineals from rocks.
Transportation of ores to smelter at reasonable cost.
Mining engineering - How to extract ores.
Geology - Identify useful minerals.
Knowledge of geography - Locate the mineral deposits.

This trend can be illustrated by a “technology pyramid” for ham radio technology. At the very top of the pyramid is the knowledge of how to operate a two-way radio. At the bottom of the pyramid is the prospector who explores the wilderness and first finds the raw materials needed to make a radio and every other modern technology. In the middle are the skills needed to build your own radio equipment.

One way to look at the change is that ham radio has slowly retreated up the pyramid to become just another consumer product like TVs and cellphones. Most hams argue that ham radio can do more things than ever and is therefore more interesting. We can transmit live amateur television, e-mail message nets, and faxes. Fortunately or unfortunately, the same technologies are becoming available to ordinary folks without licenses. Why bother with ham radio?

Homebuilt ham radio and the “QRP”

Building your own radios is commonly known as ***homebrewing***. Up until World War II, ham radio was homebrewing. In the early days decent commercial radio equipment was barely available and during the depression it was unaffordable to average people. If you couldn’t build your own equipment, you probably couldn’t afford to get on the air. In this respect a little poverty isn’t always a bad thing. When life is too easy, it becomes boring.

After 1950 building transmitters at home remained common only among the young and impoverished. Good communications receivers were the hardest to build, so homemade receivers were the first to go. In the 1950s and 1960s pre-fabricated, well-designed kits from companies like Heath and Allied Radio replaced the homebuilt equipment. To have good, usable equipment, all you had to do was solder it together. Finally in the 1980s the kits became so complex, the kit companies couldn’t trust us to do more than solder prefabricated assemblies together. When that happened, the kits became so boring that they disappeared altogether.

Today there is a growing interest in building low power QRP transmitters. These are transmitters with less than 5 watts output. "***QRP***" is an abbreviated way of saying, "please reduce power" in Morse code. The majority of QRP hams are once again building kits. A few modern pioneers are building them from discrete parts. One branch of this hobby builds transmitters in sardine cans to emphasize the small size of the transmitter. Another group uses metal "Altoid" candy boxes that are even smaller. In any case, QRP is the brightest hope our hobby has in remaining technically competent and attractive to young people.

(1) Wizard, The Life and Times of Nikola Tesla, by Marc J. Seifer, Carol Publishing Group, Secaucus, New Jersey, 1996.

CRYSTAL SETS TO SIDEBAND
© Frank W. Harris 2010, Revision 12

CHAPTER 2

HOME-BUILDING AMATEUR RADIO EQUIPMENT

What qualifies as "homebuilding?"

The ultimate homebrewer would be a guy that walks off into the wilderness stark naked. Then, using rocks and sticks, he builds a high tech communications system. Hey, don't laugh! Our ancestors did it. Of course, they had to stay out in the woods for thousands of years to finish the job. Our ancestors were smart, hard-working folks. From our lofty position today it's easy to look down on them as some sort of blue-collar, country bumpkins. Ironically, country bumpkins are often more than they appear. In my opinion, today's family farmers are masters of a wider range of technology than any other group of people in our society. Farmers today can't afford to hire others to repair their old equipment, so they have mastered skills from electrician to biologist. Most farmers today have talents and knowledge that make the rest of us look like specialized robots on an assembly line.

Building your own radio equipment is fun and gives you the satisfaction that comes from mastering knowledge. Homebuilding doesn't need a precise definition. If you're having fun and learning something, do it! You should feel free to invent your own definition. However, the more levels of the technology pyramid you master, the bigger your achievement. Besides, buying ham equipment is like skipping school and buying a diploma. You may get the same privileges, but the result will be empty in many ways. My personal definition of "homebuilding" is that I build my own equipment starting from simple components that (I hope) I understand.

I try never to buy equipment or subassemblies specifically designed for amateur radio. I am proud to be the bane of most of the advertisers in ham radio magazines. I still buy individual electrical components, of course. I just pretend that the electronics industry never got around to inventing radio communications.

An irony of our hobby is that, when the few remaining homebrewers retire from their day jobs, they often build and sell ham radio equipment. These industrious guys manufacture and sell every imaginable ham gizmo. I doubt any of them have noticed that, by making everything readily available, they have discouraged homebuilding.

When is homebrewing NOT appropriate?

Homebuilding should not be done as a way to save money or procure modern equipment. Modern multiband, HF ham transceivers are amazingly cheap. You should be able to buy a quality, used, high frequency SSB transceiver for as little as $500 - $1000. The good news is that many people in our society are so prosperous that parents can afford to give kids toys costing hundreds or even thousands of dollars. The bad news is that the economic incentive for young hams to build something has almost vanished. Also, if you do build your own, even the most skilled homebuilder will not have enough years to build a truly equivalent, modern amateur radio station. The commercial units contain custom integrated circuits, custom cabinets, and are the end result of multiple prototypes and exhaustive testing.

Homebuilding with vacuum tubes

50 years ago hams built reasonably good transmitters and receivers in their basements. A well-built homebrew could cover all the high frequency (HF) bands (1.8 to 30 MHz). Best of all, its signal didn't embarrass you on the air. Homebrew CW and AM phone were routine. On the other hand, homebrew single sideband phone (SSB voice transmission) has always been rare because SSB is complex, or at least finicky, to generate.

Circumstances have changed. If you could examine the complete schematic of a modern ICOM, Kenwood, or Yaesu transceiver, you will need a microscope to read it. Of course if it's an "F" size engineering drawing, you might be able can read it OK, but the schematic wouldn't fit on your living room floor. In contrast, the schematic for a 1970 all-band, CW/ AM phone ham transmitter, complete with parts list, fits comfortably on one page of a ham radio magazine. In short, modern stuff is complicated. If they were just made of "little vacuum tubes," it would still take hundreds of times longer to build them. But there are more differences.

Frequency stability

Nearly all hams today are using transceivers designed around precise, digital frequency synthesizers. The frequency is digitally displayed on a little LCD screen to precision as close as 0.1 Hz. The numbers on the readouts may not be that accurate, but the average guy believes his dial is accurate because he hasn't read the specifications. For example, stability might typically be specified to 1 part in a million. So on the 10 meter band, 28 Megahertz, the *precision* of the display might be "0.1 Hz," but the guaranteed *accuracy* might be +/- 28 Hz. However, if the commercial equipment is reading an error of 28 Hz, it is still probably much more stable than a typical homebuilt oscillator. Consequently, when a homebuilt transmitter goes on the air and drifts more than about 20 Hz, it isn't long before the criticism begins.

Today if a homemade transmitter frequency drifts like ***typical*** rigs of 40 years ago, the complaints will probably drive that ham off the air. In my limited experience, typical vintage ham transmitters from 35 or 40 years ago are likely to transmit poor quality signals that will provoke lots of complaints. On the other hand, if you are using top-of-the-line, best-quality, 40 year old equipment, then the signal quality is likely to be acceptable. Chapter 14 illustrates the effort needed to build old-fashioned equipment suitable for use on the air today. Yes, it can be done. But with such high standards, it's a challenge to make those old parts perform like modern equipment.

Lead Inductance

Another barrier to homebuilding transmitters today is ***lead inductance***. The problem is that simple wires act like coils at high frequency. They resist radio frequency currents in proportion to the frequency - the higher the frequency, the greater the "***impedance***" to current flow. When the currents are tiny, like those used in 1960-era ham vacuum tube transmitters, this problem isn't severe. In an old tube transmitter, the plate voltage ranged from 500 to 2000 volts or even more. In accord with Ohm's law, small currents multiplied times the inductance times frequency = small voltage loss. And as you can see, starting with 500 volts, there was a lot of voltage to lose.

In contrast, most modern transistorized transmitters run on 12 volts. Since 12 volts eliminates much of the power supply complexity, this sounds safe and easy. For operation in a vehicle, you don’t even need a power supply, you just run it directly off the battery using the cigarette lighter outlet. 12 volts DC supplies are safer and sound like good news.

Unfortunately, if your power supply is only 12 volts, the currents in your transmitter will be 50 times higher for the same amount of power. Therefore the voltage drop across lead inductances will be 50 times higher and, starting with 12 volts, you don't have much voltage to lose. Suddenly each component in your circuit acts as if it has an inductor in series with it. These unwanted inductors cause the transistor stages to self-oscillate, or simply to produce no power gain. The bottom line is that it’s much harder to build transistorized transmitters.

The solution to the inductance problem is ***printed circuit boards*** and very short leads between the components. This means that the craftsmanship needed to build a transistorized transmitter is higher than we needed with vacuum tubes. If you don't carefully limit your goals and work within your knowledge and time constraints, the decision to homebrew your station can lead down a long road of frustration.

****

BASIC ELECTRICAL KNOWLEDGE

The first challenge in building radio equipment is acquiring the technical know-how. There are many ways to learn basic electronics. The American Radio Relay League (ARRL) Amateur Radio Handbook is published every year and has (nearly) everything you need to know. This large volume covers all aspects of the hobby. Unfortunately, its size is overwhelming for many readers. Also, so few hams build their own equipment today that serious homebuilding has been de-emphasized in the Handbook. In fact, if all you had were the handbook, you would probably conclude that building your own equipment is impractical. A goal of what you are now reading is to help restore homebuilding to some of the importance it once held in Amateur Radio. This chapter introduces the minimum knowledge you need to get started. If you have trouble grasping the perhaps overly simplified explanations in this book, read about the same topics in the ARRL manual and other texts.

Much of the following is written in a “Dick and Jane” fashion that ignores many of the fine points. For example, if I say, “***all circuits are complete loops***,” a purist might bring up unusual situations in which this doesn’t appear to be true. However, it’s important to have some simple, bedrock concepts to fall back on which will usually serve you well. “Rules of thumb” are essential to keep your reasoning organized. However, keep your mind open to exceptions, because you’ll encounter lots of them. When you do find exceptions, don’t throw out the basic rules of thumb. They usually work well and will keep you sane.

I only use math when it is essential to do the job. A serious problem with engineering schools is that engineering professors are selected from the 2% of the students who naturally think in terms of math. So after graduation, the math geeks are the guys who hang around to educate the next generation. The students who think like “normal people,” go out in the real world to build things that work. Actually, the majority of engineers are so disenchanted by the obsessive/compulsive math-squiggles that they go into management or sales.

Learn a qualitative understanding first

By concentrating on math, professors often fail to teach the basic concepts of how electricity works. A friend of mine, John Anderson, taught third year electronics engineering. He was discouraged to discover that most of his students could cover a blackboard with equations, but couldn't bias a light emitting diode or predict the voltage on the collector of a saturated transistor. Any good technician or experienced ham can do these tasks with hardly a thought.

As John told me, "Engineering is like learning to paint a picture. First you need the skill to paint a horse so it looks like a horse and not a house. It's not important for most artists to be able to calculate the physical chemistry of paint pigments. Those skills are only important to a few specialized members of the art community."

Personally, I worked for decades as a real engineer, but my methods are primarily qualitative. I only use simple math that I completely trust. One day I was working on a project with John Anderson when he turned to me and said, "You spent four years in engineering school and all you learned was that math doesn't work and shouldn't be trusted." I am ashamed to admit he's right. Once I got into the real world, I spent too many months calculating random, useless numbers on paper. Eventually I learned to work primarily from experience and intuition. In other words, I work like a radio ham! I'm making this confession to emphasize that when I use formulas and math, it is important stuff and you won't be able to avoid using it.

In case you believe that building radios in your basement is just idle entertainment, consider this: During my four years in the Air Force my job title was "electronics engineer." I did my technical work entirely with my ham skills. No one ever asked me to write equations. All they wanted from me was a working power supply by Thursday, a working radio control link, an electronic timer, a static electricity protection circuit, or whatever. If I had not been a ham, I suppose the Air Force would have found some paperwork for me to do, but I would not have been able to complete the tasks I was originally assigned.

****

MAGNETIC AND ELECTRIC FIELDS

Magnets - an everyday force field generator

I never get tired of playing with magnets. Other than Earth's gravity, magnets are the most common devices in our everyday world that allow us to experiment with a physical field. An invisible force from the magnet reaches out and grabs paper clips or thumbtacks from a quarter inch away or more. Playing with two magnets is even more interesting. One magnet can attract another, the same way a magnet attracts steel. But when you reverse the direction of one of the magnets, the magnetic force acts as a "repeller" pushing the other magnet away or causing it to spin 180 degrees to the opposite polarity. After it has been reoriented, the second magnet jumps toward the first magnet.

Playing with magnets gets us accustomed to the idea that a magnetic field may be generated by a physical device. We also observe that ***a magnet field is a short-range phenomenon.*** With small magnets the field only seems to be effective for a quarter of an inch or less. With such short range, refrigerator magnets don't seem promising as a starting point for long-range communication.

The magnetic field from a permanent magnet arises from the individual atoms that make up the magnet. Each atom has electrons that orbit around the nucleus. These electrons resemble electrons passing around the coils of an inductor. Unlike the electrons in ordinary atoms, the electrons of iron and other magnetic atoms orbit the atoms aligned to at least a small degree. In contrast, ordinary atoms have electron orbits that race around the entire atom in perfectly spherical shells. Ordinary atoms generate magnetic fields too, but there is no alignment and the fields cancel.

Electric fields are also common in our everyday world, but the effects are usually subtle. During the winter at northern latitudes the humidity in our houses drops and static electricity discharge often startles us when we grab a metal doorknob. Like the magnet, static electricity can move lightweight objects a short distance. It can make non-conductors stick together, or even repel them. Unlike magnets, the objects affected by household electric fields must be featherweight items, like lint or hairs. Static cling in our clothes is an example of oppositely charged materials attracting each other, just like the magnets. In contrast, petting the cat may make the cat's hair stand on end. When hair stands on end, each hair is repelling the hairs around it. Moreover, a bit of humidity makes the phenomenon vanish. Because the affected objects are so lightweight, static electricity seems pretty minor. For both magnetic and electric fields these routine phenomena seem to be strictly short range.

WHAT IS ELECTRICITY?

Electricity consists of electrons flowing through conductive materials. Some physicists might jump in now and start lecturing about holes flowing through P-type semiconductor or chemical ions migrating through electrolytic solutions. But let's forget about exceptions for now. It's easier to visualize current flow as electrons.

Electrons are a form of "charge." "Charge" just means an excess or lack of electrons that gives rise to an electric field. When charge is gathered together in one place, it establishes an electric field with respect to nearby objects that have an opposite charge, or simply have less of the same charge. If a conduction path is established between the two objects, the electric field becomes a ***voltage. Voltage is the force that pushes a current flow*** from one object to the other until the charges are equal.

Electric field force can be startlingly powerful. For example, suppose a copper penny were floating in the vacuum of deep space. Now suppose it were possible to strip all the electrons from the penny and move them in a glass bottle some distance away from the penny. Obviously electric force will try to reunite the penny with its electrons. How far apart would these two objects have to be moved in order for the force driving them together to decrease to one pound of force? Believe it or not, according to my college physics book, the answer is 3.6 million miles. (2) This tells us that, in the real world, gigantic voltages are needed to remove more than just an occasional electron from an object. Electronics merely dabbles with the surface of an immense ocean of power and energy. Removing all the electrons requires converting the penny into pure energy - in other words, $E = mc^2$.

Electrons have a ***negative*** charge. Benjamin Franklin coined the word "negative" based on a static electricity experiment in which electric charge was transferred from a person to a glass tube. That is, the person had loss or "negative" amount of charge. Whereas when the glass transferred charge to a person, the person gained a "positive" or increased amount of charge. (1)

‘Negative” is just a word and could just as well have been called “George” or “Napolean.” The opposite of negative charge occurs when an object has lost electrons and needs to replace them. An object with a dearth of electrons is said to have a “***positive charge***.” When the electrons are given the opportunity to travel, they will flow from a negatively charged region to a positively charged region at (nearly) the speed of light. ***“Positive current flow” is usually taken to mean current flow from positive to negative.*** In other words, positive current flow is the direction opposite to the flow of electrons.

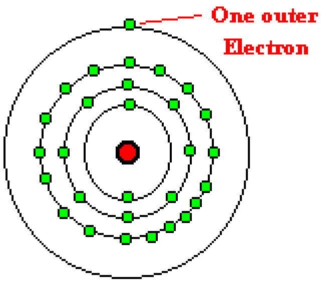

copper

Metal conductor

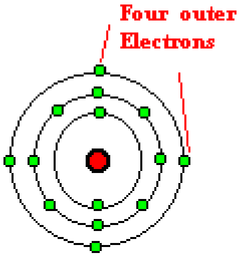

silicon

Semiconductor

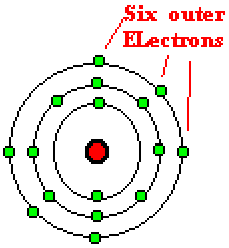

Sulfur

Insulator

Metal Conductors

Materials that allow electrons to travel are called “electrical conductors.” Electrical conductors are usually metals. If you remember your chemistry, metals are on the left side of the periodic table. Metals have an outer electron shell with 1, 2, 3, or 4 electrons. As you may recall, 8 electrons are needed to complete a stable outer electron shell in most elements. The outer electrons orbiting metals are not tightly bound to their atom and these electrons can be easily lured away. In a piece of metal the outer electrons wander away from their home atoms and move over to neighboring metal atoms. These roaming electrons form an electric current. For example, copper has just one outer electron. When one considers price, weight, conductivity, ductility, and corrosion characteristics, copper is the best available metal for making wires.

A copper penny weighs 3.1 grams. If it were possible to remove all of its electrons instead of just a scattering from the outer shells of atoms, how many electrons would you have and what could you do with that many electrons? The answer is 5×10^{29} electrons. This is the number of electrons that pass through a 100 watt incandescent light bulb in 40 hours of operation - all from one penny! (2)

Insulators

The atoms on the right side of the periodic table have their electron shells filled with 8 electrons or are nearly filled with 5, 6, or, 7 electrons. These atoms do not have loose electrons, but they easily accept loose electrons to complete their shells. In this way, insulators take free electrons out of circulation and prevent current flow. These elements are inert gases and materials like sulfur and phosphorous which are extremely poor conductors. Non-conductors are called ***insulators*** and are vital in electronic circuits to confine the electric currents to the intended pathways. For example, sulfur has six electrons in the outer shell and is a crystalline electric insulator. All the lighter elements with 7 and 8 outer electrons are gases. In practice insulators are usually compounds made from two or more atoms with four or more outer electrons. For example, ordinary silicon-dioxide glass is made from silicon with 4 outer electrons and oxygen with 6 outer electrons.

Semiconductors

Just to the right of center in the periodic table are the ***semiconductors***. Semiconductors, such as silicon, germanium, and carbon have a valence of 4. That is, they have half-stable outer electron shells. If enough voltage is applied, electrons can be forced to leave a semiconductor atom and the semiconductor can act like a conductor. Or, if the missing 4 electrons are supplied to the atom, it can become a temporary insulator. This ambivalent nature is exploited in semiconductors to make electronic switches and diodes that can be turned on or off. In addition, combinations of semiconductor elements have other surprising properties. Silicon solar cells turn sunlight into electricity. Light emitting diodes turn electricity into light. Other types of simple semiconductor devices detect magnetic fields, detect gamma radiation, regulate voltage, and serve as reset-able circuit breakers. Needless to say, complex matrix arrays of semiconductors are also manufactured into entire computers on a chip or visual displays.

Ordinary amorphous carbon, such as charcoal, is a poor conductor, not a semiconductor. But when carbon is crystallized into graphite it becomes a semiconductor. Unfortunately, graphite is too heat sensitive be used for transistors. Mixtures of silicon and carbon are used successfully in blue LEDs. In theory, carbon in the form of diamonds might make superb high performance semiconductors. In practice, this has been difficult and hasn't happened yet.

Some semiconductors are made from mixtures of lightweight elements with valences of 3 and 5 that give an average valence of about 4. For example, all light emitting diodes are made of mixtures like this and use a wide variety of elements. As examples, LEDs are made from silicon carbide, indium phosphide, gallium-aluminum-arsenide and aluminum-indium-gallium-phosphide. These produce red, green, yellow, blue, and even infra-red “colors.”

High atomic weight elements make poor semiconductors and insulators

The neat relationship for properties of elements breaks down at the bottom of the periodic table. Heavy elements like lead and tin have 4 electrons in the outer shells. Their location in the table suggests that they should be semiconductors. However, when they react chemically, they usually just lose 2 electrons, not 4. They are good electrical conductors and their physical properties resemble zinc or copper, rather than carbon or silicon.

Arsenic, antimony, and bismuth should be insulators since they have 5 electrons in their outer shell. But as the atomic weight rises, they look and act less and less like crystalline insulators and more and more like heavy metals. Bismuth is a good electrical conductor that closely resembles lead. It wasn't fully recognized as being different from lead until 1735. All the elements heavier than the inert gas radon are conductive metals, regardless of their outer shells.

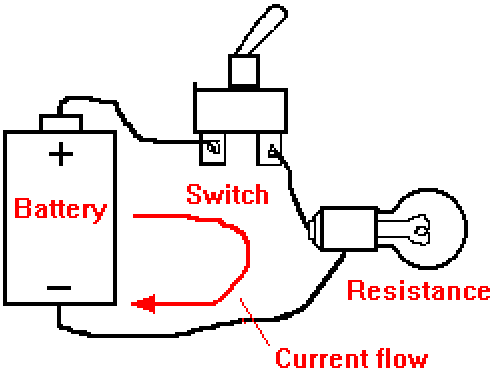

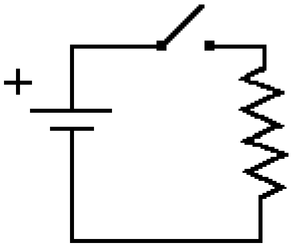

Schematic diagram

Circuit Loops

All circuits are in complete loops. Electrons must flow in complete circles or no work is accomplished. ***Voltage is the force that pushes electrons around a circuit.*** Voltage by itself does nothing. Only when voltage is able to push electrons through a complete circuit, does anything happen. For example, unscrew a light bulb while it's glowing. The light is extinguished because the current no longer passes through the bulb. That was obvious, but notice that current doesn't jump across the terminals inside the bulb socket and it doesn't leak out into the room. There's still plenty of voltage across the light socket terminals. You could put your finger in the socket to test it, but current would flow through your finger and, with household voltage, that is a very bad idea.

Voltage = Electrical Pressure

Voltage, electrical pressure, is measured in **VOLTS**. As you know, an ordinary flashlight battery has about 1.5 volts of electrical pressure. Household power in North America has about 120 volts pressure. You already know that handling 1.5 volts is harmless to people, while handling a hundred times that much is dangerous. Sticking your fingers in light sockets can be fatal. If there is enough voltage present, voltage can push an electric arc right through thin air. This feat requires about 30,000 volts per inch of air. If you're interested, you may calculate the voltage needed to generate a typical thunderbolt during a thunderstorm. Hint: This will be a really big number.

Current = The Flow of Electric Charge

Current could be measured in terms of electrons flowing through a circuit per second but that would be unwieldy. Instead, current is measured in **AMPERES.** An ampere is defined as a coulomb of electrons flowing through a wire in one second. One coulomb equals 6.242 billion-billion electrons. Enough. Let's use amperes! In ham radio, currents larger than an ampere are only common in transmitters. The currents delivered by 12 volt power supplies usually get up into the amperes. However, most ham circuits found in receivers or signal processing deal with currents measured in milliamperes and microamperes.

One milliampere = 1/ 1000 of one ampere. (one thousandth of an ampere.)

One microampere = 1/ 1,000,000 of one ampere. (one millionth of an ampere.)

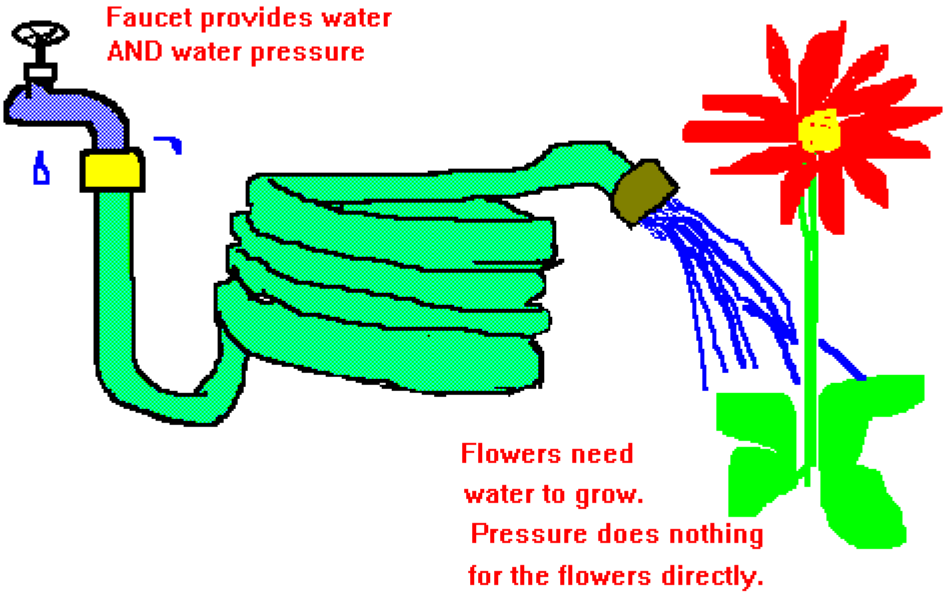
Faucet provides water
AND water pressure
Flowers need
water to grow.
Pressure does nothing
for the flowers directly.

Many people have trouble keeping voltage and current straight in their minds. To the average guy electricity is zappy-shocking stuff. Maybe the following will help. Voltage is analogous to water pressure while electric current is analogous to water. A simple way to remember this is that you can't water flowers with water pressure. Alternatively, there can be lots of water in the reservoir, but pressure is needed to push it through the hose to the flowers.

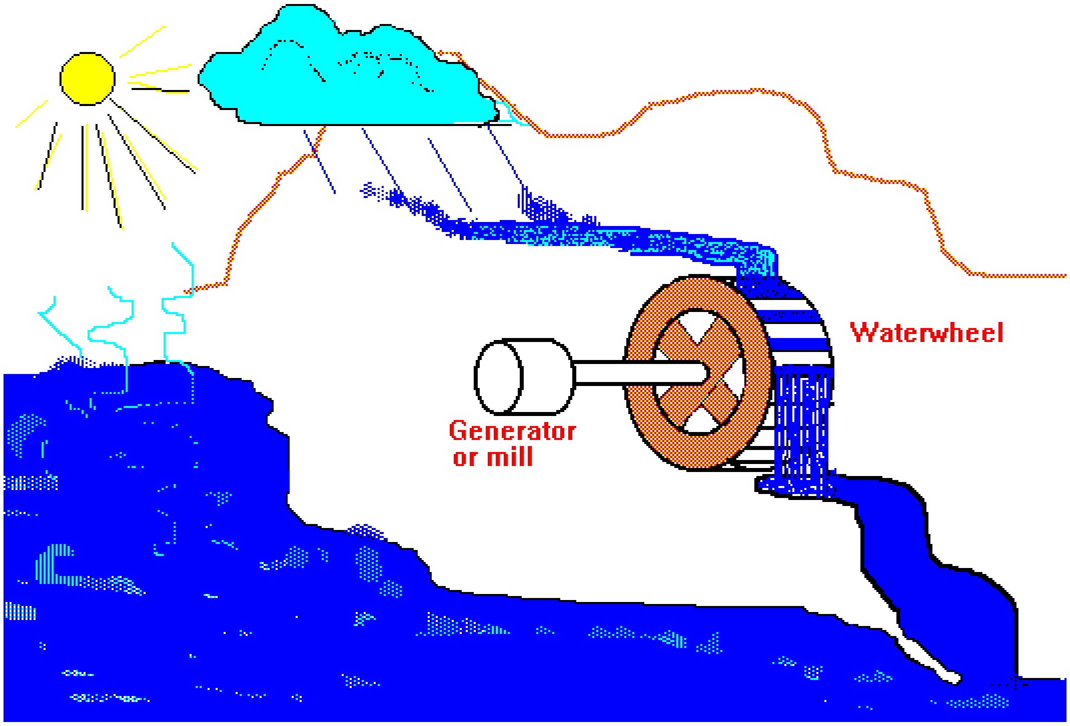
Waterwheel
Generator
or mill

Water falling as rain and flowing from the mountains down the rivers produces a cycle that is reminiscent of a complete electrical circuit loop. Sun shining on the ocean causes evaporation that raises the water up above the mountains where it again falls as rain. In other words, the sun's energy is the "battery" that keeps the cycle going. Lifting the water into the air literally raises the potential energy of the water. Rain gathers as mountain streams and rivers that can flow through dams on the way back to the ocean. The falling water imparts its gravitational energy to the water wheel in the dam, thereby doing useful work. In other words gravity provides the water pressure.

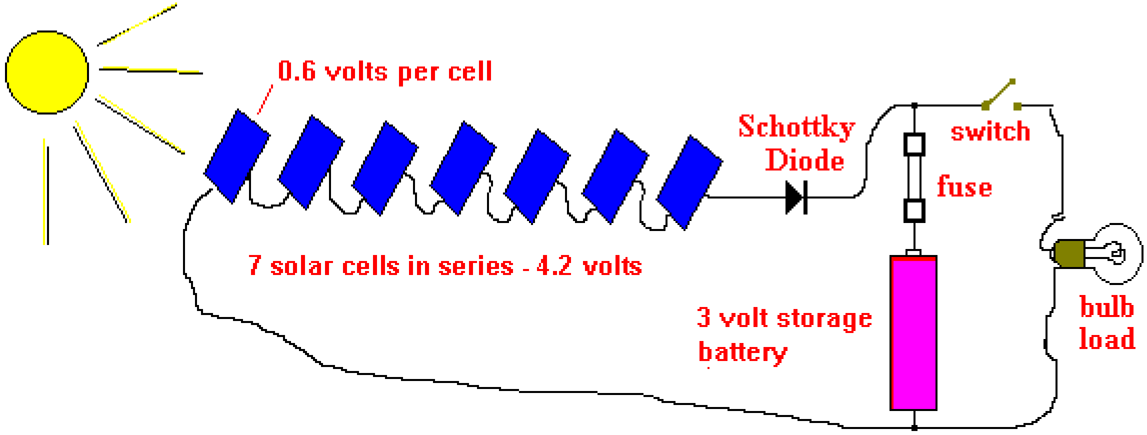

Solar cells are devices that are the electrical equivalent of sunshine raising water vapor to high potential energy through evaporation. In a solar cell, sunshine provides the energy to "boost" electrons up to a higher energy level. In practice, each solar cell generates just 0.6 volt of potential. So, in order to charge a 12 volt car battery, at least 20 solar cells must be placed in series so that the electrons can be boosted up to 12 volts.

Power = the rate of doing work

The rate at which the water wheel does work is its "power." ***Power is defined as energy produced or expended per second.*** In ham applications power is usually measured in watts. Watts are defined as joules of energy per second. ***Joules*** are occasionally used in ham work when designing coils or figuring out how big a capacitor should be. However, you won't need them very often and they aren't needed to get through this book.

The amount of energy that can be extracted from falling water equals the volume of water times the height it falls. Notice in the water wheel above, the water is only pushing the wheel around for the upper half of the diameter of the wheel. The rest of the descent is wasted - not very efficient as drawn.

Power = Height of fall (water pressure) x amount of water falling (current)

In electricity the relationship is the same:

Power = Voltage (electrical pressure) x Current (electrons doing work)

$$P = V \times I$$

Where "I" is electrical current. "I" is the letter used by engineers to represent current. No, I don't know why. Maybe it's because "C" usually stands for electrical capacitance or the speed of light.

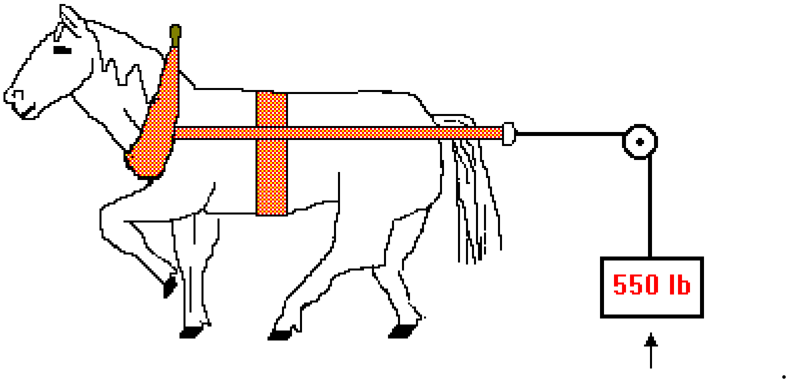

To put watts in familiar terms, power was probably first described in terms of equivalent working horses. For example, horses were formerly used to lift coal or metal ore up a mineshaft. Pulling on a pulley mine hoist, a strong horse might lift 550 pounds up one foot in one second. This was defined as one horsepower. It turns out that:

746 watts = 1 horsepower = 550 pounds lifted one foot each second.

Resistance

In the water cycle analogy above, the rain falls in the mountains then makes a long journey of days back to the sea. This long journey is analogous to a high electrical resistance. In contrast, if the rain fell directly into the sea, the journey would take only a few minutes. This "short circuit" journey would be analogous to a LOW electrical resistance.

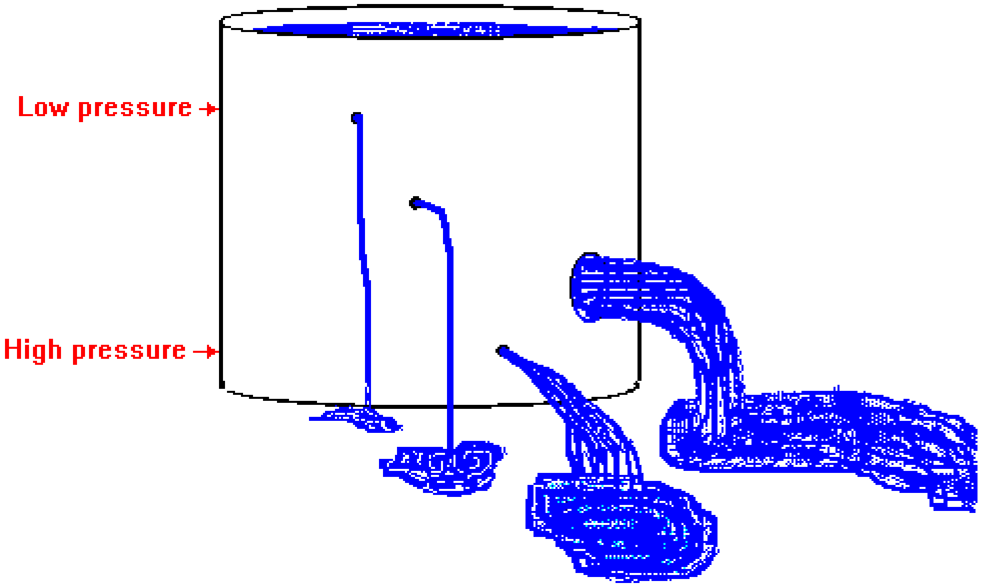

Good water analogies for electrical resistances are narrow water pipes and wide water pipes. For example, suppose a large water tank is suddenly punctured by small bullet holes and a large hole the size of a melon. Obviously the water rapidly flows out of the tank, especially the water coming out of the big hole. The speed at which the water leaves depends on the size of the hole and the water pressure. Because of the weight of the water higher than a hole in the tank, the water pressure is much higher at the bottom of the tank than it is at the top. Therefore, water leaving from a hole at the bottom of the tank squirts out at considerable pressure, while at the top of the tank it just dribbles out slowly.

Ohm's law

Resistance is measured in ohms. The ohm is named after Georg Simon Ohm. The relationship between voltage, current, and resistance is described by Mr. Ohm's law, which says that one volt will drive one ampere through one ohm resistance.

Voltage = (current) x (resistance)

In electrical formulas V = voltage and R = resistance. Ohm's law is written as

$$\mathbf{V = I \times R}$$

If you know any two of the three parameters, you can find the missing value with algebra. Rearranging the letters we can have:

V = IR or **I = V/ R** or **R = V/ I**

Sorry about the math, but this relationship is unavoidable for anyone who wants to succeed with electricity.

For example: You know that household electric current operates at 120 volts. Suppose that you measure the current through a 100 watt light bulb and find it to be 0.833 amperes. What is the resistance of the light bulb?

120 volts = (0.833 amperes) X Resistance

Resistance = Voltage / Current.

Resistance = 120 volts / 0.833 amperes = 144 ohms

The symbol for ohms is usually the capital Greek letter Omega, Ω. For example, 144 ohms might be written as "144 Ω."

Resistors in series

When current must pass through two or more resistors in series, the total resistance is equal to sum of the resistors. For example, two 100 watt light bulbs in series will present twice as much resistance as a single bulb:

Voltage = 120 volts = 0.417 ampere x (144 ohms + 144 ohms)

Because the current must pass through twice as many obstacles, the current flow is half and the bulbs will only burn dimly, if they light at all.

A voltage divider made from resistors

A common use for resistors is to reduce DC or AC voltages. Two resistors in series make a voltage divider. This is analogous to the remaining potential energy of the rainwater after it has returned halfway back to the ocean. In the light bulb example above, the voltage across one of the two light bulbs in series will be 50% of the total. In other words, the voltage will be reduced to 60 volts AC. In the DC electronic circuitry is it common to reduce a voltage to bias a transistor or to control the voltage across a light or other application. The more current required at the lower voltage tap point, the lower the resistances must be to support the extra load at the reduced voltage.

In the example below 10 volts AC is divided to produce 5 volts AC. This same resistive divider could just as well divide 10 volts DC down to 5 volts DC. You'll see shortly that there are other ways to make AC voltage dividers and that is why the odd value of "530 ohms" was used in this drawing.

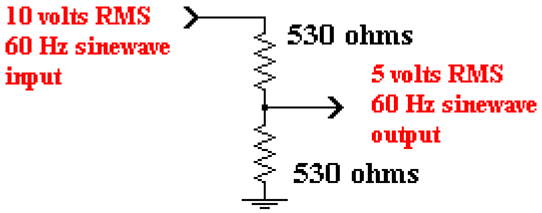

A Voltage Divider Made from Resistors

Resistors in parallel

When two equal resistors are placed in parallel, the resistance seen by the voltage source will be half as much as if there were one resistor. This causes twice as much current to flow. In the case of light bulbs, when you turn on twice as many lights, the current flow doubles, all bulbs burn equally brightly, and your electricity bill doubles.

When the parallel resistors are equal in size, the current they will draw is intuitively obvious. That is, each resistor receives the same current. But when the resistors in parallel are all different, then you will need to calculate the equivalent resistance using the formula:

$$\text{Resistance equivalent} = 1/(1/R_1 + 1/R_2 + 1/R_3 + \ldots\text{etc.})$$

The equivalent resistance is the inverse of the sum of the inverses of the resistances.

If there are just two resistors, then the formula is slightly less ungainly, but probably harder to remember:

$$\text{Resistance equivalent} = R_1 \times R_2 / (R_1 + R_2)$$

For example, suppose a 50 watt light bulb with a resistance of 288 ohms is in parallel with a 100 watt light bulb with 144 ohms resistance. The equivalent resistance would be:

Resistance equivalent = (288 ohms x 144 ohms)/ (288 ohms + 144 ohms)

Resistance = 96 ohms

Real world resistors

The photo below shows some actual resistors. 95% of the time you will be soldering in those tiny little ¼ watt guys at the lower left. The blue and gray square plastic devices in the center are trim pots that allow you to adjust a voltage or resistance to your exact needs. The biggest resistors and the pot at the upper right are wire wound devices that can handle 5 or more watts. The controls at the lower right are the sort of potentiometers used to adjust the volume on a receiver or stereo.

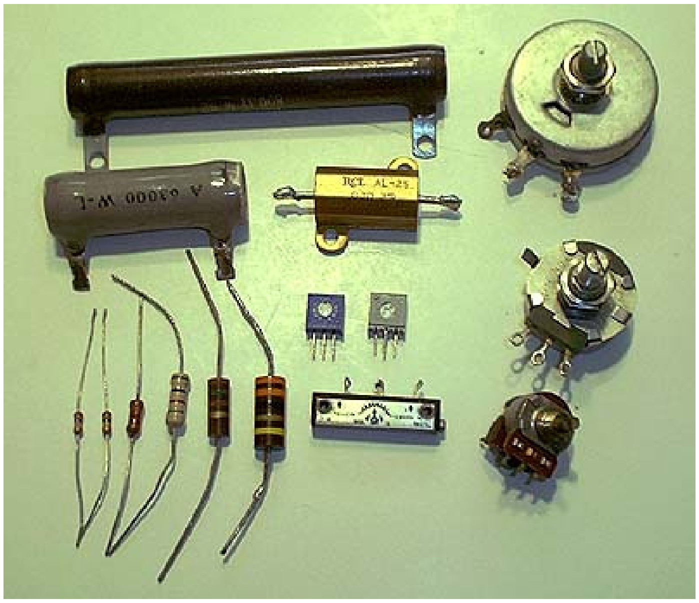

The small radial lead tubular resistors are usually marked with a 3 digit color code starting from the left. For example, at the lower center above there is a big 2 watt, 33 ohm carbon composition resistor whose resistance is labeled orange, orange, black. Orange means "3" and black means "0." The third digit tells you how many zeros to add after the two-digit number, zero zeroes in this case. The fourth colored bar is "gold" which means the resistance value is guaranteed to be within 5% of the above value. If the color were silver, the tolerance would be 10% and no color would mean 20%. These days resistors more imprecise than 5% are rarely encountered.

On the other hand, ***if the 3rd band is silver or gold, the decimal multiplier is less than one, 0.1 or 0.01 respectively***. I suggest you check these with a ohm meter before you use them. 1% precision resistors have another bar for a third significant digit. Occasionally another colored bar, usually yellow like the 33 ohm resistor above, is at the opposite end of the unit. This means that the resistance will not drift significantly after thousands of hours of use. When in doubt, use your ohm-meter!

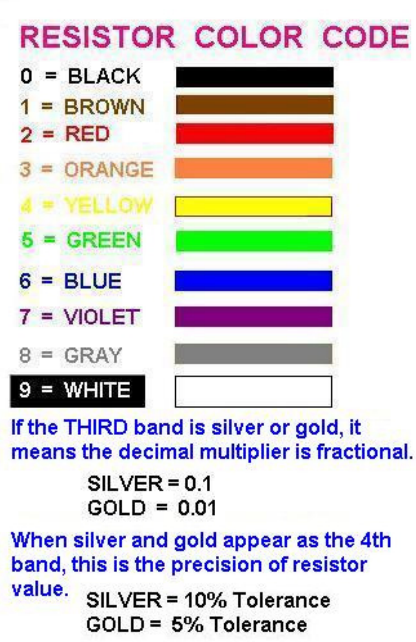

RESISTOR COLOR CODE

0 = BLACK
1 = BROWN
2 = RED
3 = ORANGE
4 = YELLOW
5 = GREEN
6 = BLUE
7 = VIOLET
8 = GRAY
9 = WHITE

If the THIRD band is silver or gold, it means the decimal multiplier is fractional.

SILVER = 0.1
GOLD = 0.01

When silver and gold appear as the 4th band, this is the precision of resistor value.

SILVER = 10% Tolerance
GOLD = 5% Tolerance

Batteries

A battery is a voltage source that delivers electrical current as a result of a chemical reaction. The battery contains two chemicals that react together to produce a third chemical compound. Nearly all of the energy released by the reaction takes the form of electricity. When the supply of the original two chemicals is exhausted, the reaction and the electricity stop. The battery is then “dead.” But as you know, some types of batteries can be recharged. This is done by applying a voltage higher than the nominal battery voltage to the same battery terminals. Current flows back into the cell, reverses the original chemical reaction, and restores the original charged state. For example, lead-acid car batteries are designed to be recharged. Others, such as alkaline batteries, can only be recharged with difficulty and may leak or burst. Lithium batteries are tricky to recharge and this may even cause them to catch fire. For example, the Tesla electric car uses rechargeable lithium batteries, however the batteries are immersed in cold freon to dissipate the heat and avoid a fire. The radiator under the hood cools the freon, not the motor.

The voltage and capacity of batteries can be compared to pressure and volume of water tanks. A really large battery, such as a single battery cell in the bank of batteries in a submarine, can have a huge capacity and may be able to supply hundreds of amperes for hours. But in spite of its great size, it might only be able to supply two volts of electrical pressure. In a submarine there might be a hundred or more of these huge cells in series to supply 200 volts to the electric motors. In terms of analogous water tanks, a large low voltage battery is comparable to a large cattle watering tank, perhaps twenty feet wide but only one foot deep.

A car battery is the biggest battery most people ever see. In order to start a car on cold mornings, the car battery must delivers hundreds of amperes for several seconds. It does this through LARGE copper wire cables. These are the thickest electrical wires under the hood. If you tried to use thin wires for this load, the wires would melt in seconds. How much power does it take to start your car in January?

Power = 12 volts x 200 amperes = 2,400 watts

or, Power = 2,400 watts / 746 watts/ hp = 3.2 Horsepower.

In contrast, an ordinary 9 volt alkaline transistor battery has six tiny 1.5 volt cells in series for a total of 9 volts. Each cell is only rated at delivering about 1/2 ampere (550 milliamperes) for one hour. In terms of power, this is about 4.5 watts.

Power = 9 volts x .55 amperes = 4.5 watts

Notice that a little battery with several cells in series would be analogous to a tall water tank with low volume. For example, suppose a pump is located at the bottom of a well one hundred feet down. The pipe might be only an inch in diameter. Therefore a hundred feet of this pipe will only contain a few gallons. But the pressure at the bottom of the pipe will be huge, 123 pounds per square inch. In municipal water systems the water is stored in big tanks up on towers to provide both water pressure and a significant reserve of potential water flow.

Electrical conductors are comparable to holes punched in the water tank. Fine wires have higher resistance than fat wires made of the same material. Metals like copper and silver have such inherently low resistance that the difference in conductivity may not be obvious. But in a conductor like carbon, a fat stick of carbon will have far less resistance than a thin fiber of carbon such as the carbon fibers Edison used in his first commercial light bulbs.

Electrical Power

What is the power consumed by the light bulb? Or more familiarly, what is the light bulb “wattage?”

Power = Volts X Current

Power = (120 volts) X (0.8333 amperes) = 100 watts.

Another way to calculate power is often useful when the voltage isn't known:

Power = Resistance X $(\text{Current})^2$

$$P = I^2 \text{ x } R$$

Substituting for voltage from above,

Power = (Current x Resistance) X Current

Power = Resistance x $(\text{Current})^2$

Or, if you only have voltage and resistance,

Power = $(\text{Voltage})^2 /$ (Resistance)

$$P = V^2 / R$$

For example, 100 watts = $(120 \text{ volts})^2$ / 144 ohms

Capacitors

Capacitors are devices that store electric charge, much like a rechargeable battery. Capacitors are usually made from two parallel plates of conductive metal separated from each other by a thin layer of insulation. The insulation can be air, paper, plastic, mica, or even vacuum. When the two plates are connected to the poles of a battery, charge rushes into the capacitor by means of wires. Since positive charge is attractive to negative charge, the charge gathers on the plates as if it were “trying” to jump across the gap. Because the voltage from the battery has established the same uniform voltage across the plates, the charge gathered on the plates of the capacitor will maintain that voltage, even after the battery is disconnected from the battery. The charge is held on the plates by the force of attraction from the opposite plate.

The amount of charge a capacitor can store is directly proportional to the area of the plates and how little insulation separates the plates. The less distance between the plates, the more charge it can store. Of course the more thin the insulation, the more likely it is to break down with high voltages. So voltage rating and capacitance are opposing characteristics.

If a charged capacitor has high quality insulation that doesn't leak charge across the gap, the capacitor will store energy in the form of trapped charge indefinitely. For example, large, low voltage capacitors are sometimes used to maintain computer RAM memory when the rest of the computer is turned off. However, with present technology capacitors rarely seriously rival electrochemical batteries as a long-term energy source.

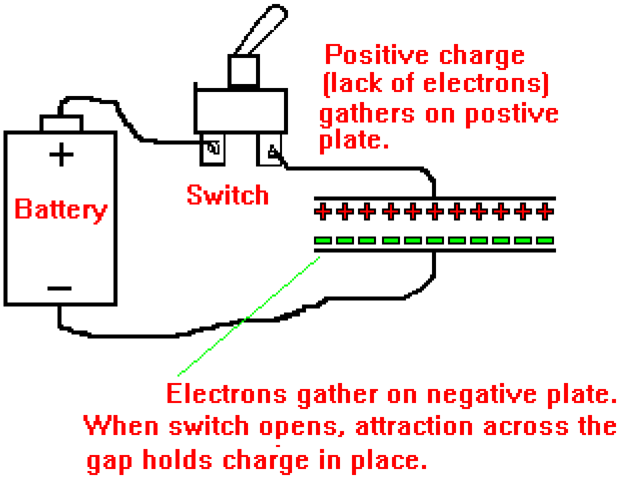
Positive charge
(lack of electrons)
gathers on postive
plate.
+
Battery
–
Switch
Electrons gather on negative plate.
When switch opens, attraction across the
gap holds charge in place.

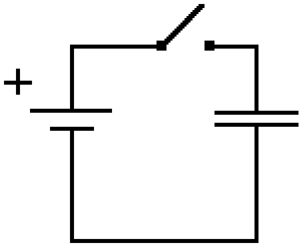

Schematic diagram

In a capacitor the electrons gathered on the plates establish an electric field across the gap. Another way to look at a charged capacitor is that the electric field across the insulation stores energy, just as a magnet stores energy in the magnetic field in and around the body of the magnet.

The Voltage Across a Capacitor Can't Change Instantly.

In electronics, the single most important property of capacitors is that ***the voltage across a capacitor can't change instantly***. The current flowing into or out of a capacitor can change immediately, but the voltage can't change until more charge has been stored or charge has been removed. In other words, current must flow into or out of a capacitor before its voltage can change. In contrast with a capacitor, the voltage across a resistor and the current flowing through it are proportional and change instantly. With resistors, voltage and current are always strictly related by Ohm's law. Ohm's law works with capacitors only when the driving voltage is a continuous sinewave operating at a constant frequency.

Capacitors are commonly used in circuits in four different ways:

1. They store energy for short periods of time.

2. Because capacitor voltage can't change instantly, capacitors are used to smooth out or regulate voltages and hold them constant. For example, direct current (DC) power supplies use large capacitors to hold the direct current at a constant voltage.

3. Capacitors are used to separate AC signals from DC. Capacitors "conduct" alternating current much like resistors, but because of the insulation layer, capacitors block the flow of direct current. When alternating polarity voltage is applied, current rushes in and out of the capacitor. The alternating current *appears* to be flowing through the capacitor.

4. Capacitors are used with inductors to form resonant circuits. Capacitor/ inductor resonant circuits are the core of radio technology and will be discussed in detail later.

Farads

The storage capacity of capacitors is called "***capacitance***" and is measured in **FARADS**.

One farad of capacitance is a huge capacitor. Although you can buy such a novelty, the insulation between the plates is generally very thin and will only tolerate low voltage, usually just a few volts. In low frequency audio electronics and power supplies the capacitors used are usually measure in MICROFARADS, one millionth of a farad.

Microfarads are sometimes written as " µF " where the Greek letter µ is short for "micro." For example, a large 10,000 microfarad capacitor might be labeled "10,000 µF."

Capacitors store electrical charge, which is measured in **COULOMBS.** A coulomb of electrons is 6.242 billion-billion electrons. A common electronics laboratory practical joke involves charging up a large capacitor say, 10,000 microfarads, to greater than 20 volts. The capacitor is then tossed to a victim. “Hey, Joe, catch!” Joe grabs it and receives a big jolt as it discharges through his fingers.

A defibrillator for restoring normal heart rhythm works on this principle. A large capacitor is typically charged to about 400 joules of energy - that is, 400 watts for one second. The capacitor is then discharged through large paddles placed across the victim's chest. Being defibrillated is a horrific jolt and can be fatal if misapplied.

Notice that any conductive object isolated from a second conductor by an insulator comprises a capacitor. For example, your saltwater-filled body is a conductive object. If you’re wearing insulating rubber soled shoes and standing on a metal floor, you are now a capacitor with respect to the floor. Your body capacitance might typically be, say 50 picofarads. A picofarad is one millionth of a millionth of a Farad. This sounds trivial, but it is not.

If the humidity is low and your body becomes charged as you scuff your feet on the wool carpet, your body can be charged up to thousands of volts potential. Fortunately only a tiny amount of charge gathers on your body, typically far less than 1 microcoulomb. Otherwise you might be electrocuted instead of just jolted when you touch a metal object. This phenomenon becomes important when handling integrated circuits and transistors. Discharging your body capacitance into an integrated circuit can instantly ruin it. ***Always ground your body before touching chips and transistors, especially chips made from field effect transistors.***

Capacitors in parallel are calculated like resistors in series

If two capacitors are placed in **parallel**, the storage capacity of the pair is the sum of the two. This is easy to visualize. If the capacitors are equal, then the size of the parallel plates doubles while everything else remains the same.

Total capacitance = C1 + C2 + C3 + …..etc.

When capacitors are in series, the calculation resembles resistors in parallel. Capacitors in series have less ability to store charge, but they gain extra insulation and can withstand higher voltage. The voltage rating of a capacitor is called its ***working voltage*** or “***WV***.”

Equivalent Capacitance $= 1/(1/C_1 + 1/C_2 + 1/C_3 + \ldots\text{etc.})$

If you have just two capacitors in **series** they may be calculated like resistors in parallel

Equivalent Capacitance $= C_1 \times C_2 / (C_1 + C_2)$

The current that flows into a capacitor moment to moment while the capacitor is charging can be calculated using relationships that are based on Ohm’s law. Unfortunately, to express a formula that will work in all cases of changing voltages requires a differential equation. In general,

$$I = (dV/dt) / C$$

Where t stands for time and C = the capacitance in farads. "dV" means "the differential of voltage and dt means the differential of time.

Life is too short for hams to get too far into calculus, unless of course it happens to fascinate you. Fortunately for most of us we rarely need it. There are special cases in which we can do simpler calculations to get what we need. For example, from the point of view of an AC sinewave current, capacitors look like resistors. That is, the capacitor takes a finite amount of time to charge and discharge. Therefore, when a sinewave voltage is applied to a capacitor, the current flowing in and out of the capacitor will be limited by how much voltage is already charged across the capacitor.

At one extreme when the frequency is infinitely high, the capacitor has an equivalent resistance of zero – a short circuit. At the opposite end of the spectrum, to a DC current, the capacitor just looks like an open circuit. That is, it will have infinite resistance. After all, the capacitor is just an insulator separating two conductors. And if the charging voltage hasn't changed in ages, no current will be flowing into or out of the capacitor.

Capacitive reactance

At frequencies between zero and infinity a capacitor resists sinewave current flow as if it were a resistor. You can calculate the equivalent resistance or ***capacitive reactance*** as follows:

Reactance in ohms = 1/ [2 π (Frequency) (Capacitance in Farads)]

$$X_C = 1/(2 \pi f C)$$

Where π is "pi" or 3.1416

For example, the reactance of a 5 μ Farad capacitor at 60 Hz is:

$$X_C = 1/[2 \pi (60 \text{ Hz}) (5 \mu \text{F})] = 530 \text{ ohms}$$

We can use this resistance-like property to attenuate sinewave signals or make voltage dividers. For example, two 5-microfarad capacitors in series can divide the voltage of a sinewave in half.

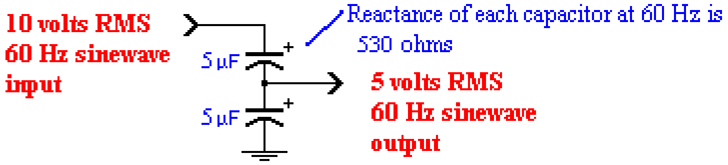

A Voltage Divider Made from Capacitors

Unlike a comparable divider made from 530 ohm resistors, no energy would be dissipated in the capacitors, so they don't get hot and they don't waste energy.

Real world capacitors

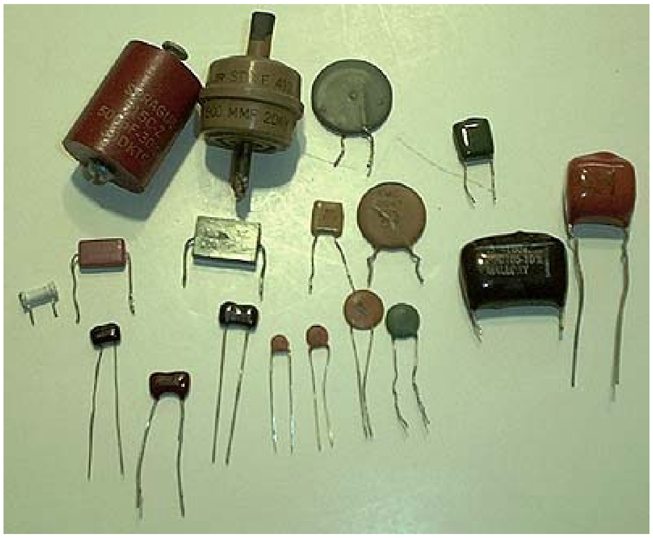

On the left is an assortment of fixed capacitors that have ceramic or mica insulation. Large capacitors either have high capacitance, high voltage ratings, or both. The "door knob" capacitors at the top are rated for 30,000 and 20,000 volts and are used in picture tube power supplies. Mostly you will be using the small mica and ceramic capacitors in the lower row. In ham equipment they typically range from 2 pico-farards (2 micro-micro-farads) up to about 0.5 microfarads.

The material used as insulation also partly determines the capacitance. If a vacuum is used as an insulator, it has a ***dielectric constant*** defined as 1.000. All other materials produce more capacitance, assuming that the area and thickness are equal. Mica, plastics, and various oils all have constants ranging between 2 and 10. Different dielectrics vary in cost and temperature stability. In general, for radio applications, mica and low dielectric constant ceramics have the best temperature stability and the least energy dissipation. There are many kinds of high dielectric constant ceramic capacitor types, but the penalties are always heat and temperature instability. Pure water is a poor to fair insulator, dissipates a great deal of heat, but it has a high dielectric constant of 80.

Electrolytic capacitors

Capacitors larger than roughly 1 microfarad are nearly always ***electrolytic capacitors***. In the capacitive voltage divider shown earlier, the little + signs over the 5 µF capacitor symbols mean they are electrolytic capacitors. It also means that, when you wire them into a circuit, you must wire them so that the + mark is oriented to line up with the polarity of any average DC voltage that might be applied to the capacitor. In the example above, the polarity would not matter because the capacitors are only exposed to AC voltage.

An electrolytic capacitor achieves high capacitance by means of two tricks: First, the insulator is an oxide layer electrochemically plated onto a thin sheet of aluminum or tantalum metal. The oxide layer serves as the insulator and can be microscopically thin. Capacitance is inversely proportional to the distance between the plates, so the thinner the insulator, the higher the capacitance. Because the insulation is chemically bonded to the metal, there are no gaps between the metal and insulation and the contact is as intimate as possible.

The second trick is that, although one pole of the capacitor is the metal sheet, the other pole is an electrolytic solution, a salt bath. The fluid is in extremely intimate contact with the insulation, so that the effective gap between the two conductive bodies is truly tiny, just a few molecules thick. Aluminum oxide and tantalum pentoxide insulation have dielectric constants of 8.4 and 28, which also explains their high capacitance. I was intrigued by a list of dielectric constants in the ARRL handbook that gave the dielectric constant for titanium dioxide as 80. Unfortunately I've never seen any titanium oxide capacitors in catalogs. No doubt they have some severe practical disadvantage - probably the usual heat and low breakdown voltage problems.

Electrolytic capacitors must be wired with the correct polarity

Now the bad news: ***Electrolytic capacitors are polarized***. In other words, if they are used in a circuit with a DC voltage, they must be oriented in the circuit so that one pole is always positive and the other is always negative. If the electrolytic is wired backwards, the DC current will corrode the insulating layer of oxide and the capacitor will eventually short out. When the short occurs, the aluminum capacitor usually makes a loud bang and vents a cloud of foul smelling steam. A shorted tantalum capacitor may even burst into flames. ***It is extremely exciting to have an electrolytic go off in your face.*** There are homebrewers who ***have had*** this experience and the rest who eventually ***will have*** this experience. Always make an extra effort to be sure the polarity is correct. This is especially vital in DC power supplies where the capacitor is continuously subjected to the same polarity.

Fortunately, in low frequency circuits it's OK if the polarity is briefly reversed. For example, if an electrolytic capacitor is used to drive a high power HI-FI speaker, the audio signal voltage is both positive and negative as the music current signal goes through it. But when there is no sound present, the capacitor must be oriented so that the local DC voltage has the correct polarity across the capacitor. There are also special capacitors called ***non-polar electrolytics***. They are made from two electrolytic capacitors back to back. They are generally more than twice as large as a polarized electrolytic capacitor of the same capacitance.

Real Electrolytic capacitors

An assortment of electrolytic capacitors is shown below. The tiny ones on the left are all tantalums. Tantalums are generally less than 100 microfarads and have working voltages less than 35 volts. The modern ones are all encapsulated in plastic or epoxy while the older tantalums often have metal cases. The big capacitors on the right are aluminum electrolytics. The largest is rated at 120,000 μF at 10 volts DC WV.

INDUCTORS

Inductors are the functional opposite of capacitors. Remember that the voltage across a capacitor cannot change instantly. In contrast, ***the* current *through an inductor cannot change instantly.*** Whenever electrons flow through a wire, a magnetic field appears in the space surrounding the wire. This energy floats in the space surrounding the wire. Since the energy is real, it doesn't appear without a cost. The price paid for establishing the field is that energy must be expended in the wire to "charge up" the magnetic field before the electrons can pass through. When the current first begins to flow, the wire appears as a "resistance" for a moment. Then, after the field has been established, this pseudo-resistance drops to zero and the direct current flows through unimpeded.

This property of wires is called "***inductance***." The longer the wire, the more inductance it has. Inductance is measured in **Henries**. It is named after Joseph Henry, an English Physicist who first described this phenomenon. In equations, inductance is usually represented by "**L,**" which doesn't make much sense.

Inductor design

The picture below shows several turns of wire wound into an "***inductor***." A coil of wire like this has more inductance than would be expected from a straight length of the same wire. The magnetic fields from each turn overlap, reinforce each other, and gather in the center of the coil.

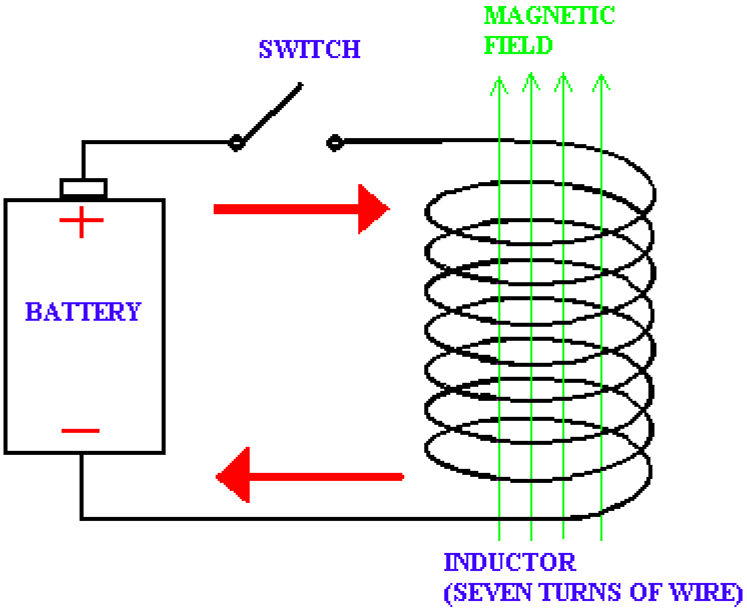

The more turns in a coil, the more inductance it generates. In fact, the inductance is proportional to the square of the number of turns. Therefore a seven-turn coil like the one above has 7^2 or 49 times more inductance than a single turn coil would have. Notice that the magnetic field produced by this coil is a genuine magnet. In fact, it will lift iron filings. Heck, big electromagnets like this are used in junkyards to lift cars! Unlike a refrigerator magnet, this magnet can be turned on and off using the switch.

If you wanted to build a big electromagnet of the junkyard variety, an air-core coil like the one above would be impractical. By the time you had wound enough turns to generate a field strong enough to lift a car, you would need miles of wire. Miles of fine wire would have a huge electrical resistance. This is not the kind of resistance called “impedance” or “reactance”, but just the resistance of the copper metal that makes up ordinary wires. To push big currents through a high resistance takes a high voltage, high current power source. If you actually did that, all those turns of fine wire packed together in a coil would get hotter than blazes. As we saw above, power = current times voltage. A high resistance coil would be in danger of melting because the energy dissipated in it would be so high.

As permanent magnets demonstrate, magnetic fields require no power consumption at all once the field has been established. A current is only needed to generate a magnet field. In a permanent magnet, the electrons orbiting around the iron or other magnetic atoms provide the current. No outside power is needed to keep those electrons orbiting, so a permanent magnet can generate the field forever. In theory, superconducting wire can be wound in a closed loop to produce a permanent magnet by having a current circulating through the loop endlessly without loss to form a permanent magnet. In practice, the unavailability of high current, high temperature superconducting wire makes this dream impractical.

Instead, ordinary iron can be “recruited” to form a magnet even when it is not already crystallized with all the magnetic atoms lined up in the same direction. ***Electro-magnets are constructed by winding a coil around a piece of iron.***

The Current Through An Inductor Cannot Change Instantly

In electronics, the single most important property of inductors is that ***the current through an inductor can’t change instantly***. The voltage across an inductor can change immediately, but the current can’t change until the magnetic field surrounding the inductor has been augmented or decreased. In other words, voltage must change across the inductor to force a change in the current and its corresponding magnetic field. In summary, inductors are the opposite of capacitors. As with capacitors, Ohm’s law works with inductors only when the driving voltage is a continuous sinewave operating at a constant frequency.

Inductors are used commonly in electronic circuits in four ways:

1. Inductors are used to prolong the flow of current into a resistance and hold the current constant. This prevents surges of currents that might otherwise damage other components. They are used in filters of various kinds to regulate voltage and currents.

2. Inductors generate magnetic fields that attract iron and produce motion. Consequently, inductors are a component of relays, solenoids, motors, and loudspeakers.

3. Inductors, as part of transformers or acting alone, are used to change voltage levels up or down.

4. Inductors are used with capacitors to form resonant circuits. Capacitor/ inductor resonant circuits are the core of radio technology and will be discussed in detail later.

Inductance in measured in Henries

The unit of inductance is the ***Henry.*** A Henry is a relatively large inductor. An inductor that large nearly always has an iron core and is often used in low frequency applications like power supplies. At radio frequencies a microhenry is a significant inductance. As you will see, most radio circuits use inductors in the range of 1 to 100 microhenries. The energy handling ability of inductor is rated in ***amperes***. For example, suppose a small inductor is rated at 1 millihenry at 100 milliamperes. If you put more than 100 mA through it, the wire might burn up or, if it has an iron core, the iron might saturate and the inductance will drop dramatically to much less than its 1 mH rating. Iron cores and saturation will be discussed in more detail shortly.

Inductors in series

When current must pass through two or more inductors in series, the total inductance will be equal to the sum of the inductors. When two equal inductors are placed in series, the current must pass through twice as many obstacles, the current flow will charge up the inductors half as rapidly, but twice as much magnetic field energy will be stored.

To calculate the equivalent inductance of several inductors in **series**, just add up the separate inductances, just like series resistors.

Inductance equivalent = $L_1 + L_2 + L_3 + L_4 + \ldots$ ***etc.***

Inductors in parallel

When two equal inductors are placed in **parallel**, the inductance seen by the voltage source will be half as much as if there were one inductor. This causes twice as much AC current to flow and the magnetic field energy will charge twice as fast and be half as large.

When the inductors are equal in size, the AC current they will draw when in parallel is intuitively obvious. But when the inductors in parallel have different inductances, then you will need to calculate the equivalent inductance using the formula:

Inductance equivalent = $1/(1/L_1 + 1/L_2 + 1/L_3 + \ldots$ ***etc.***$)$

That is, the equivalent inductance is the inverse of the sum of the inverses of the inductances. If there are just two inductors, then the formula reduces to the same as it did with resistors:

Inductance equivalent = $L_1 \times L_2 / (L_1 + L_2)$

Calculating sinewave current flow through inductors and the voltage across them can be done using Ohm's law. Unfortunately, just like capacitors, when the current is not a constant sinewave, the calculations are tricky because the current flow is time dependent. To do it correctly in a way that works under all circumstances requires a differential equation, the inverse of the differential equation for capacitance. Specifically,

Voltage $(t) = L\, dI/dt$

Where t is for time and L = inductance in Henries. Why "L"? Inductance could have been "I," but that stands for current. Fortunately, hams hardly ever need to make these calculations. However, when the changing current is a sinewave, the voltage or equivalent resistance ***(inductive reactance)*** of an inductor is easy to calculate.

Inductive reactance

From the point of view of a constant AC sinewave current, inductors look like resistors. At one extreme, if the frequency is infinitely high, the equivalent AC resistance is infinite. At the opposite end of the frequency spectrum, to a DC current the inductor just looks like a long piece of copper wire. In other words, at DC a perfect inductor would have zero resistance.

At frequencies between zero and infinity an inductor resists sinewave current flow as if it were a resistor. You can calculate the equivalent resistance or ***inductive reactance*** as follows:

Reactance in ohms = 2π (frequency) (Inductance in Henries)

$$X_L = 2\pi f L$$

For example, the reactance of a 5 Henry inductor at 60 Hz is:

$$X_L = 2\pi \text{ (60 Hz) (5 H)} = 1885 \text{ ohms}$$

Like capacitors, inductors can be used like resistors to attenuate currents or divide voltages. For example, two 5-Henry inductors in series can divide the voltage of a sinewave in half. Unlike a comparable divider made from 1885 ohm resistors, virtually no energy is dissipated in the inductors.

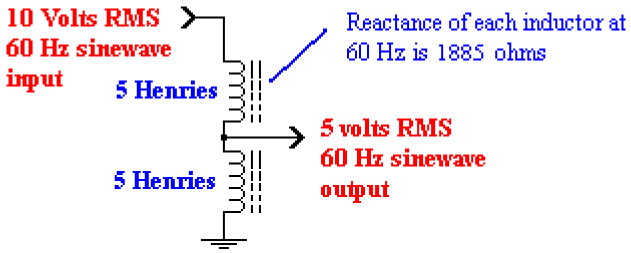
10 Volts RMS
60 Hz sinewave
input
Reactance of each inductor at
60 Hz is 1885 ohms
5 Henries
5 volts RMS
60 Hz sinewave
output
5 Henries

Voltage Divider Made From Inductors

If you wanted 530 ohms reactance like the previous voltage dividers, then you can solve for the inductance starting with 530 ohms. As you can see, smaller reactance means a smaller inductance.

$$530 \text{ ohms} = 2\pi \text{ (60 Hz)(L) }, \quad L = 1.4 \text{ Henries}$$

Iron core inductors

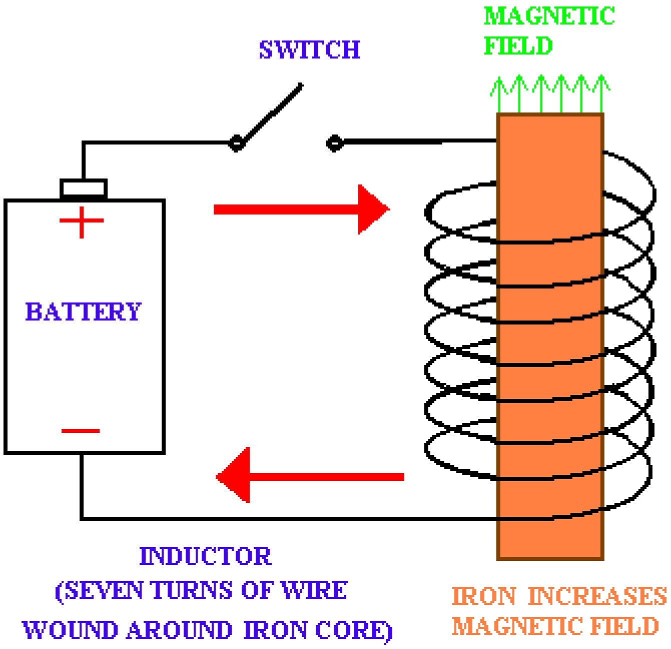

An iron bar placed inside a coil of wire greatly increases the inductance and the magnetic field energy by orders of magnitude. This is how high-force electromagnets are made. In the divider circuit above the dashed lines next to the inductor symbols means that the inductor has an iron core. In practice, any inductor larger than about 0.1 Henry will nearly always have an iron core to achieve such high inductance. If room temperature superconductors ever become practical, this may change. But for now, large inductance means iron core.

Transformers

Transformers are devices consisting of two or more inductors that share the same magnetic field. Suppose the two coils are placed next to each other so that the magnetic field from one coil will pass through the other:

A changing magnetic field couples energy from one coil to another. If we put a voltmeter or oscilloscope across the resistor on the secondary coil, we will see a voltage jump up from zero, then immediately fade back down to zero as the magnetic field in the coils is being established. Once a static magnetic field is established in the primary winding, there will be no voltage across the secondary winding and resistor.

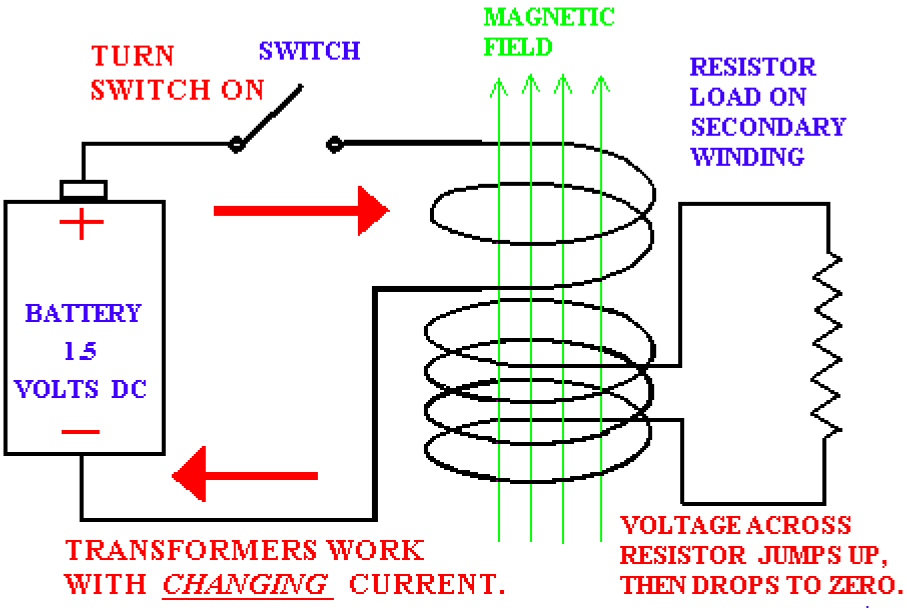
TURN
SWITCH ON
SWITCH
MAGNETIC
FIELD
RESISTOR
LOAD ON
SECONDARY
WINDING
BATTERY
1.5
VOLTS DC
+
—
TRANSFORMERS WORK
WITH *CHANGING* CURRENT.
VOLTAGE ACROSS
RESISTOR JUMPS UP,
THEN DROPS TO ZERO.

Iron core transformers

Putting an iron core inside a transformer vastly increases the inductance of the windings. This greatly enlarges the size and duration of the voltage spike observed across the secondary coil.

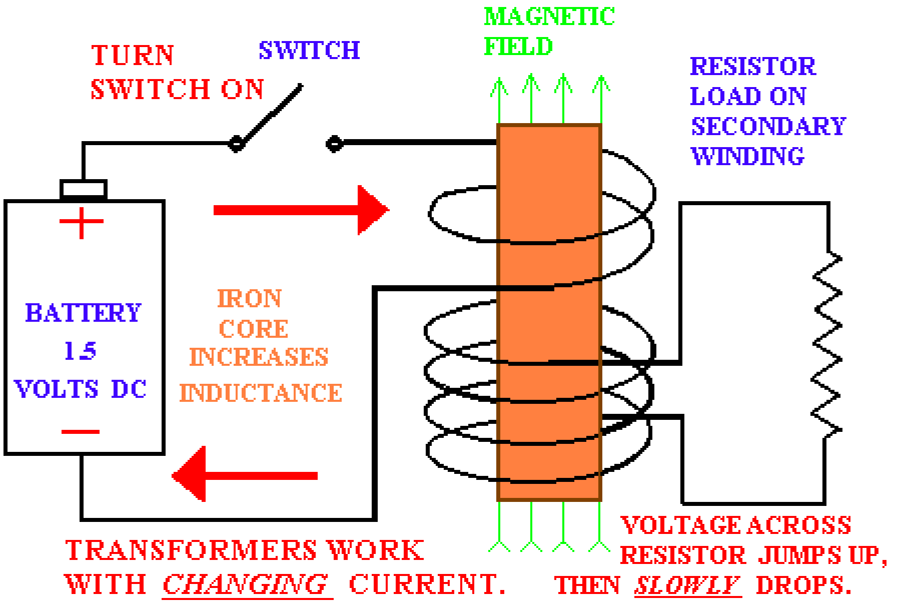
TURN
SWITCH ON
SWITCH
MAGNETIC
FIELD
RESISTOR
LOAD ON
SECONDARY
WINDING
BATTERY
1.5
VOLTS DC
+
—
IRON
CORE
INCREASES
INDUCTANCE
TRANSFORMERS WORK
WITH *CHANGING* CURRENT.
VOLTAGE ACROSS
RESISTOR JUMPS UP,
THEN *SLOWLY* DROPS.

Unfortunately, ordinary iron not only generates magnetic fields, it is also a conductor of electricity. As you know, the copper windings are loops of conductive metal. Similarly the iron core is a solid mass of conductive metal. Consequently, the iron acts like a array of shorted loops competing with the copper windings for current flow. As a result a simple iron transformer like the one above works poorly unless the changes in the current are very slow. That is, so long as the current changes take place slowly, like manually turning an electro-magnet on and off, then it works OK. But as the switch is thrown faster and faster, more and more energy is wasted in circulating currents inside the iron.

Real inductors

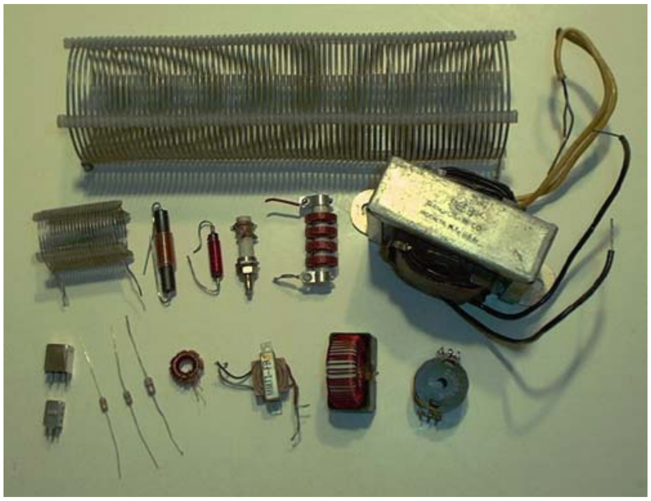

An assortment of inductors is shown on the left. The kinds usually used in modern homebrew ham equipment are in the bottom row. The tiny metal cans are shielded miniature IF transformers used in superhetrodyne receivers. The tiny chokes at the lower left resemble resistors and use the same color codes. I used lots of these in the designs in this book. The big coils, "chokes," and iron transformers were common in vacuum tube transmitters 50 years ago.

An automobile spark coil

Subdividing the iron core into vertical slats of iron (or better yet, silicon steel) ameliorates the problem of circulating currents. This forces the currents to circulate into tiny loops that don’t waste so much energy. The figure directly below shows a simplified ignition system for an automobile. The switch in this case is the breaker points in the distributor. In modern cars, the switch is actually a transistor, but the function is the same as the old-fashioned mechanical breaker points. Since the engine can fire many thousands of times per second, the rates of change of the currents in the primary are fairly high.

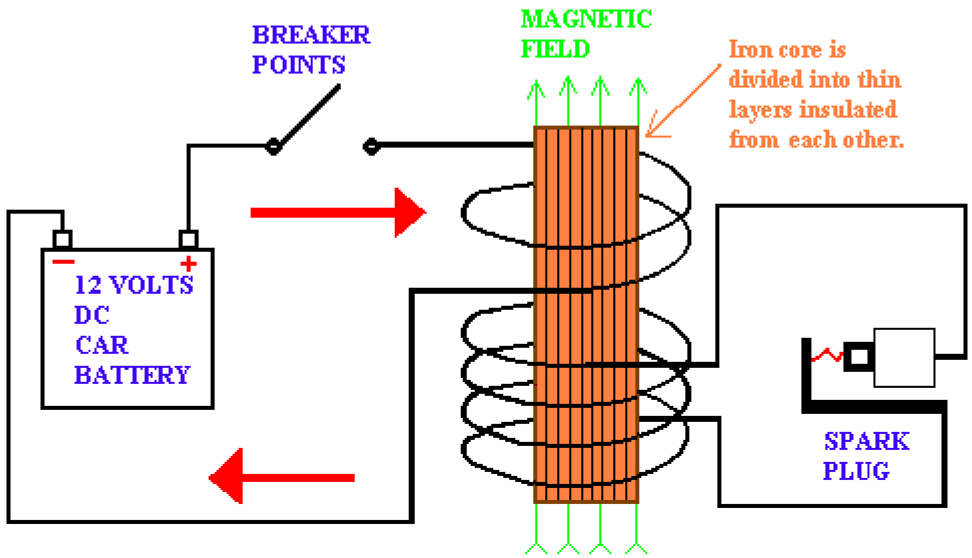
BREAKER POINTS
MAGNETIC FIELD
Iron core is divided into thin layers insulated from each other.
12 VOLTS DC CAR BATTERY
−
+
SPARK PLUG

AUTOMOTIVE IGNITION TRANSFORMER.
12 Volt primary winding, 10,000 Volt secondary winding.

Voltage step up and current step down

A vital characteristic of an automobile ignition transformer is that the voltage across the secondary must be extremely high. Without the high voltage, the spark would not jump across the spark plug gap. ***Voltage is raised by having far more winding turns on the secondary than there are on the primary***. An important use for transformers is to change varying (AC) voltage up or down by increasing or decreasing the number of turns on the secondary.

Ignition systems are an unusual circuit in that the secondary is an open circuit, a spark plug. When the breaker points suddenly ***open***, the inductor current must remain constant. The result is that the voltage soars to an extreme level – theoretically an infinite voltage - causing the spark to jump the gap. In practice, a capacitor across the breaker points is needed to force the energy over to the secondary winding. Unlike the inductor, the capacitor can’t change its voltage suddenly and this prevents sparks from occurring at the breaker points as well as the spark gap.

Power out of the secondary (nearly) equals the power into the primary winding

A transformer “transforms” the ratio of current to voltage for AC currents. In other words, if the voltage is stepped up, the current in the secondary will be proportionately stepped down. The power into and out of the transformer remains the same.

Power in = (Voltage in) X (Current in) = Power out = (Voltage out) X (Current out)

Generation of sinewave currents

An ***alternator*** is a device that generates low frequency AC current. How it works will be discussed in detail shortly. But for now let's just use it to deliver an AC voltage to the primary of a transformer. In contrast to a simple switch and battery, the alternator generates alternating sinewave current. When the sinewave from the alternator is fed into the primary winding of a transformer, the secondary winding will have a continuous sinewave on it, just like the primary. The voltage that appears across the secondary will be proportional to the ratio of the numbers of turns on the primary and secondary. For example, as drawn below, there are twice as many turns on the secondary as on the primary. Therefore the secondary voltage will be twice as high as the primary voltage. Similarly, the secondary current will be half as great as the primary current.

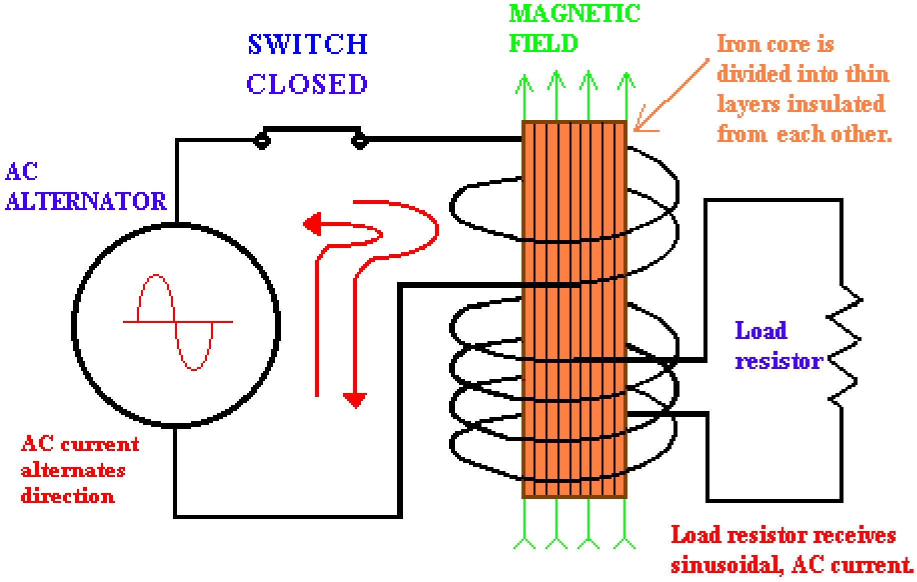
SWITCH CLOSED
MAGNETIC FIELD
Iron core is divided into thin layers insulated from each other.
AC ALTERNATOR
Load resistor
AC current alternates direction
Load resistor receives sinusoidal, AC current.

For example, if we wish to step up 12 volts AC to 120 volts AC with a sine wave, the secondary must have 120 ÷ 12 or ten times more turns.

Earlier I explained the voltage-boosting phenomenon of open circuit inductors. Now suppose we were going to use AC to generate 10,000 volts for an ignition system. If ignitions ran on AC sinewaves, instead of using DC and breaker points, then the secondary winding would have to have 833 time more turns than the secondary, 10,000 volts ÷ 12 volts = 833.

Toroidal or closed loop transformer cores

The best coupling efficiency is achieved when the magnetic field is confined to the iron core and is not allowed to fill the space around the transformer. Also, the amount of inductance per turn of wire that can be achieved is the highest when the iron core is a closed loop. In the drawing below, the thin donut-like, iron plates making up the core are stacked going away from the observer and so they are not shown.

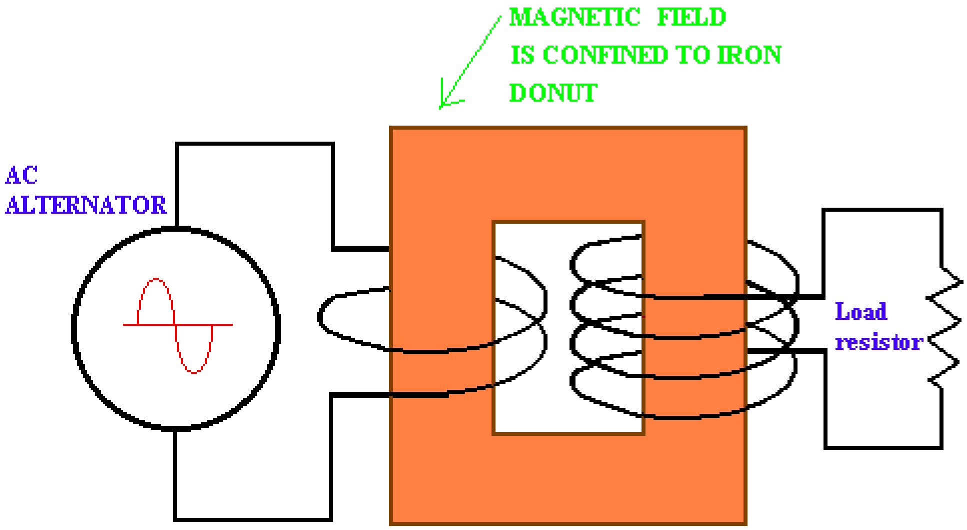
MAGNETIC FIELD
IS CONFINED TO IRON
DONUT
AC
ALTERNATOR
Load
resistor

TYPICAL 60 HZ IRON TRANSFORMER
Iron core made from laminated sheets of silicon steel.

Household power distribution

In North America electric power comes into our homes in the form of 240 volts alternating current on THREE wires. This power source is referenced to ground. The reference point is not one of the two hot wires, but is ground located halfway in voltage between the two live wires. We use the two 240 volt AC wires for our 240 volt stoves and electric clothes driers.

HOUSEHOLD POWER DISTRIBUTION

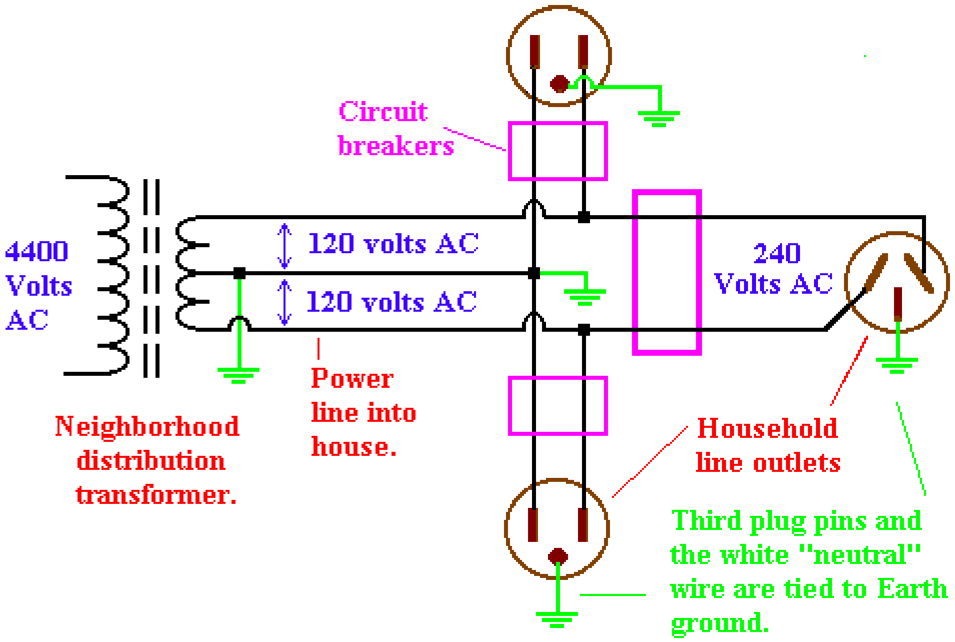
Circuit
breakers
4400
Volts
AC
120 volts AC
120 volts AC
240
Volts AC
Power
line into
house.
Neighborhood
distribution
transformer.
Household
line outlets
Third plug pins and
the white "neutral"
wire are tied to Earth
ground.

Most of our lights and appliances in the U.S.A. are powered with 120 volts AC. This voltage is obtained by tapping between one of the “hot” 240 volt wires and ground. Inside your circuit breaker box there are two ***buss bars.*** These are the 240 volt terminals to which all the circuit breaker units are attached. A third buss located between the first two is the ground terminal. The big 240 volt breakers snap onto the outer terminals. The smaller 120 volt breakers clamp onto the center ground and either one of the two 240 volt side rails. Clever, huh? Now you can obtain two different AC voltages from the two 240 volt AC wires.

Electric power comes into our neighborhoods at very high voltages, 4400 volts AC or more. Large transformers out in the alley convert this power down to 240 volts for several houses in the neighborhood. These step-down transformers are those large, black “pole pigs” up at the top of power poles. Why not just generate electricity at the power plant at 240 volts AC and be done with it? Because copper wire has a significant resistance. If large currents travel through miles of wire, the energy will be dissipated by heating the copper instead of being delivered to the customers. By using very high AC voltages, high power can be delivered with small currents and low loses. For long cross-country power lines the AC voltage is usually hundreds of thousands of volts.

Low frequency transformers at home

120 volt 60 Hz power line transformers are common around the house. Most appliances that contain electronics have a transformer to drop the voltage level down to 12 volts or less. The closed loop iron transformer above is the basic design. Usually, the windings are placed on top of each other rather than on opposite sides of the donut. Also, there are usually two holes in the “donut” with the coils wound on the center post between the two holes.

Safety is an important advantage of transformers. Transformers are often used to isolate AC voltages from ground. For example, if the input current on the primary is referenced to ground, the two secondary wires will only be referenced to each other. Well, OK. Depending on the quality of the transformer design, there may be as much as one milliampere of leakage current if one of the wires were touched to ground. In a well-designed transformer, the leakage can be just a few microamperes. But for all practical purposes the AC voltage on the secondary winding has no relationship to ground.

As was just discussed, our household AC line voltage is referenced to “ground.” Ground is literally the wet dirt in your backyard. This means that if you grasp a grounded water pipe with one hand and stick a finger in a light socket with the other, a large current will pass through your body and you will experience ***electrocution***. This should be avoided unless you are tired of living. We ham radio operators have many opportunities to be shocked, burned, or killed by high voltage. In practice, the most common serious accident for hams is falling from roofs and high towers. We need to be far more concerned about safety than ordinary citizens. Isolation transformers are one way we hams make accidents less likely.

Compared to capacitors, inductors are relatively “imperfect”

In electronics capacitors are used far more often than inductors. One reason is that inductors are more expensive than capacitors. It’s hard for a machine to wind wire around tiny cores, especially a closed loop iron-containing core. Another problem with inductors is that significant lengths of copper wire must be used to build them. This wire inevitably has a significant resistance. This means that every inductor is actually two components; an inductor and an unintentional resistor. In contrast, capacitors, especially small capacitors, can approach perfection. Capacitors have essentially no resistance, no inductance, and their properties are constant over their operating range.

The other limitations of inductors are the limitations of the iron cores. As pointed out earlier, circulating currents in the iron dissipate energy. This loss appears as a warm iron core and wastes energy as if they were yet another resistor in series with the inductor.

Inductor saturation

As current flows through the windings, the magnetic field "recruits" the magnetic fields of the iron atoms to generate a much larger magnetic field than the coil would otherwise generate. This works well until all the iron has been recruited. At this point the iron "saturates" and the coil abruptly returns to being able to generate no more additional magnetic field than its copper windings can produce. As a result, iron core inductors have a limited operating range.

Little bitty inductors made from extremely fine wire wound on iron cores the size of a raisin can have impressive inductances. The catch is that these tiny coils saturate almost immediately and the inductance is only valid at essentially zero current. Seriously, a tiny inductor might be labeled, 100 millihenries @ zero milliamperes. Needless to say, components like this have limited usefulness! Small, low inductance coils with heavy wire and with no iron core are closer to perfect inductors, but without the iron, the magnetic fields surround the coils. These stray fields induce currents in nearby components. Again, if an inductor couples unwanted signals into neighboring circuits, it is not an ideal component.

Inductor core hysteresis

Another problem with iron is ***hysteresis.*** When the current in the coil shuts off, some of the magnetism in the iron remains. Then, when the iron is magnetized in the opposite polarity, energy must be wasted to first restore the magnetism to zero before it is magnetized in the new direction. Again, this is a big departure from an ideal component.

Inductors at radio frequencies

Inductors and transformers are vital to radio circuits. In radio circuits AC currents alternate not at 60 Hz or a thousand Hz, but at millions of Hz. The principles are identical to those explained above, but there are differences that may surprise you. First, in general, the more power an inductor or transformer must handle, the bigger it must be. However, the size of a required inductor varies with the frequency. At low frequencies, like 60 Hz, a transformer is comparable to bailing water with a really big bucket. Since it only delivers "water" 60 times a second, the bucket must be large to deliver a lot of water. Now suppose the same water is delivered with thimbles instead of buckets. If the thimbles are filled and emptied millions of times per second, the same amount of water could be delivered as with the bucket. ***At high frequencies high power can be handled by tiny transformers.***

Powdered iron cores

Unfortunately, many transformer problems are accentuated with high frequencies. For example, it is no longer sufficient to build iron core transformers with sheets of steel. For high radio frequencies the iron must be ground into a fine powder and dispersed in ceramic. There are two kinds of powdered iron cores used at high frequencies. Ordinary ***powdered iron cores*** have a high percentage of ceramic and relatively little iron. This insures low loss in the cores and prevents abrupt saturation at high peak current level. ***Ferrites*** are a sophisticated type of powdered iron core that looks more like black glass but has more iron and much less ceramic content. In spite of having little ceramic in the mixture, ferrites manage to insulate neighboring granules of iron from each other. Ferrites allow much higher inductances to be used at radio frequencies than would otherwise be practical.

Ferrites and powdered iron cores are usually in the form of toroids – donut shaped rings. This is the modern way to build radio frequency inductors and transformers. At lower frequencies ferrites are often formed into "pot cores" which totally enclose and shield an inductor with ferrite. This almost totally confines the magnetic fields so they cannot interfere with neighboring circuits.

Sinewaves – "Pure alternating current"

The origin of household AC sinewave current is a gigantic alternator at your local power plant. Alternators are a kind of AC generator that can be thought of as a modified transformer in which one of the windings is mechanically rotated by an engine or other mechanical power source. An alternator produces a pure sinewave output simply from the natural rotation of its input shaft. The ***stator*** is the stationary part of the alternator. It consists of the iron frame that surrounds the rotating shaft. The stator has a pair of coils mounted on it on opposite sides of the central chamber. These coils are hooked up in series so that the voltages generated at any moment are added together.

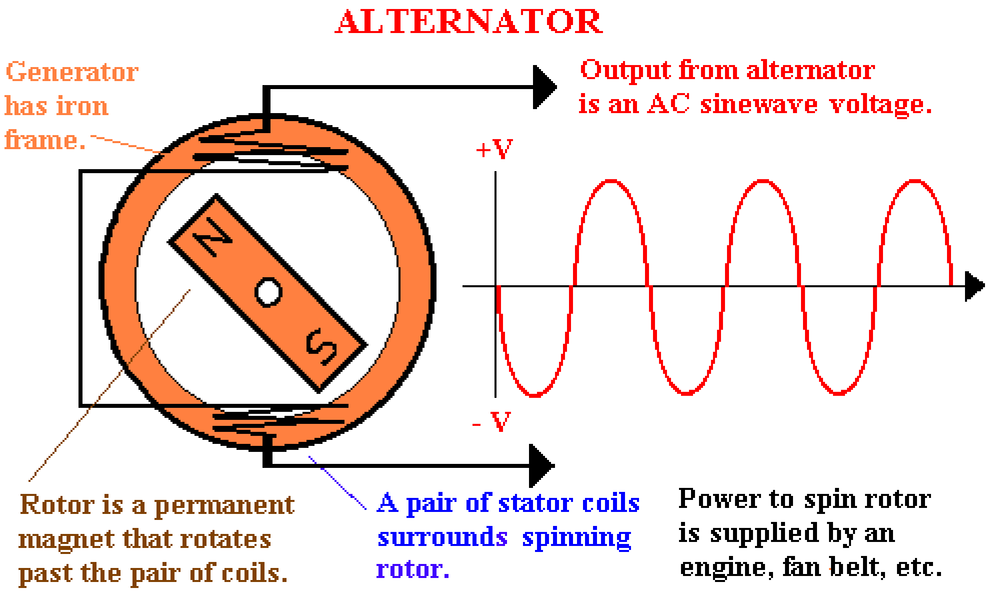
ALTERNATOR
Generator has iron frame.
Output from alternator is an AC sinewave voltage.
+V
N
S
- V
Rotor is a permanent magnet that rotates past the pair of coils.
A pair of stator coils surrounds spinning rotor.
Power to spin rotor is supplied by an engine, fan belt, etc.

The ***rotor*** is the central part that spins. As drawn here, the rotor consists of a simple bar magnet. Actually, the rotor is usually an electro-magnet, but I'm trying to keep it simple. As the north and south poles of the magnet sweep past the coils, they generate voltages of opposite polarity. That is, the negative voltage appears on the bottom terminal while the positive voltage appears on the top terminal. When the rotor spins around to the opposite alignment, the voltage polarity reverses. The sinewave comes about from the changing angle of the magnet with respect to the stator coils. To express it mathematically, the voltage output equals the sine of the angle plus ninety degrees.

Voltage at any given moment = (Peak voltage) Sine (Angle + 90^o)

Because of the 90 degree offset, this waveform is actually a cosine wave, but nobody calls it that.

When viewed abstractly, alternators closely resemble AC transformers as described above. Although there is a small air gap between the rotor and stator iron, the magnetic field travels in essentially a closed loop. The rotor is analogous to the primary winding. Instead of the energy coming in on wires, the energy enters the system mechanically. The moving magnetic field then creates the sinewave as it zooms past the stator coils.

Automobile alternators

Most alternators, such as the one in your car, are more complicated. They have three sets of stator windings crammed into the same frame. These three windings generate three separate, out-of-phase sinewaves. In your car these three outputs are rectified (converted to DC) by an array of six semiconductor diodes to produce 12 volts DC. (We'll describe diodes in chapter 4.) The DC is then used to charge the car battery. Alternator rotors are nearly always made from electro-magnets. They are powered by ***slip rings*** (moving spring contacts) that bring DC current in from outside to the moving rotor. Using an electro-magnet for a rotor allows the output voltage from the alternator to be controlled by varying the strength of the current that maintains the magnetic field. This is how the voltage regulator in your car keeps the battery charged, but not over-charged.

A sinewave voltage from an alternator obviously has a frequency determined by the speed of the input shaft. And because of the direct way the waveform is generated, the sinewave is as natural as the mathematical formula that expresses it. In contrast, modern digital electronics often generate “artificial” sinewaves by reproducing voltages dictated by a series of numbers that represent a sinewave. The end product is a complex “stair step” waveform that resembles a sinewave.

RF sinewaves

In radio technology natural sinewaves are universal in every circuit. For example, the voltages on transmitting or receiving antennas are sinewaves. Radio Frequency AC voltages are called ***RF voltages.*** They are generated by transistor oscillators as pure sinewaves. But unlike the output from an alternator, in radio circuits there are often multiple sinewaves riding on the same wire. For example, a receiving antenna carries sinewaves from ALL of the different radio waves impinging on it. Much of radio technology concerns ways to select and filter one sine wave from among a crowd of sinewaves.

So far we have introduced low frequency electrical theory. Radio frequencies are the same thing, but as we have seen, different aspects of the theory become more important as the frequency goes up. In the next chapter we’ll put together a basement radio lab and get ready to build radios.

---

(1) Chaplin, J. E., *The First Scientific American*, Basic Books (Perseus Books Group), New York, 2006, p109.

(2) Halliday & Resnick, *Physics for Students of Science and Engineering*, John Wiley & Sons; New York, 1960, pp. 558-560.

CRYSTAL SETS TO SIDEBAND

© Frank W. Harris 2010, Rev 12

CHAPTER 3

SETTING UP AN ELECTRONICS WORKSHOP

Research and Development as Recreation

Building an amateur radio transmitter and receiver from the component level up is true research and development. That's why it's fun. Sure, it's much easier than developing products that have never existed before, but conquering the details will be hard and you'll find it plenty satisfying. When you're done, you can brag about your rig to your ham buddies. These days, if you have a 100% homebuilt station, you'll find you are almost unique among other hams.

We can learn how to do R & D by studying the methods of the master inventors. We Americans would probably nominate Thomas Edison as our single most famous inventor. Edison is not credited with the invention of radio, but he did develop many of the components used in 20th century radios. Also many of the techniques to build those components were first developed in Edison's laboratory. Edison fabricated the first vacuum tube diode, the loudspeaker, and the triode vacuum tube, but he never applied them to radio. Edison is best known for a handful of his most important inventions. However, ***Edison's greatest contribution may have been his methods of inventing.***

Persistence

A reporter asked Edison why he had often succeeded in perfecting inventions, while other experimenters who started down the same path were never able to build a practical device. Edison replied, "Most inventors will have a good idea and try out one or two versions of their concept. When it doesn't work, they declare it hopeless and give up. The difference is, I never give up."

Try everything and keep careful notes

The most well-known story about Edison's persistence was his legendary search for the ideal material to make light bulb filaments. Edison was asked if he was discouraged by his failure to find a suitable material after trying hundreds of substances. He replied that it hadn't been a failure. He now knew hundreds of materials that didn't work.

The corollary to Edison's method is that it's essential to write everything down in detail. It's not fun writing the details of experiments that failed. But a year later, any researcher can tell you they have had that "deja vu" sensation halfway through an experiment only to find in their notes that they had tried this before. At the moment when you discover "a pearl of wisdom," it seems so profound that you believe you will never forget it. Wrong! Unless your memory is much better than mine, a year later an old laboratory notebook can be re-read almost as if someone else had written it.

Lots of junk

Edison was asked what a fellow needed to become an inventor. He said, " First you need a large pile of junk. You can't afford the time and money to run down to the store every time you need something. Often the junk pile will supply the parts to try out an idea right away. If you order parts from hundreds of miles away, you may waste weeks just to find out that your idea didn't work."

Subdivide the problem and build for modification

A complex invention like the light bulb consists of many parts. Light bulbs only look simple. What kind of glass can stand the heat of the filament? What kind of wire has the same coefficient of thermal expansion as that of glass? What kind of filament material is optimum? Will the filament leads bond to the glass and keep out the air? What should the resistance of the filament be in order to be compatible with the electric power source? How does the resistance change with temperature? What is the trade off between operating life and brightness? How good a vacuum is needed and how can it be produced? Once the air is pumped out of the glass, how can the bulb be sealed? Answers to all of these questions had to be found and tested one at a time before Edison could make a practical light bulb.

Have you seen the movie Gizmo? Gizmo is a collection of film clips of early 20th century inventors showing off their inventions for the first time. The movie is hilarious because again and again the inventors make the same mistake. They start with ideas that are often fundamentally sound and then they build an entire, polished, beautiful prototype without ever testing the pieces. They make the first test run of their airplane, jet boat, etc. in front of a movie camera. Sometimes they apparently even called in news reporters just to make sure the test would be as humiliating as possible.

My favorite relatively recent invention was McCready's human-powered airplane. In 1965 an English sponsor named Cramer offered a $100,000 prize to the first fellow who could fly over a 20 foot obstacle, fly over a half mile course, fly over a second 20 foot obstacle, then turn around and return to fly over the first obstacle again. The plane had to be 100% powered by muscle power. For 20 years many smart people tried to build the pedal-powered plane. Again and again they made the same error. They worked for months building handsome, handcrafted balsa wood airplanes. Then on the first flight the plane would crash and be smashed beyond repair. After two or three attempted flights, the inventors had spent months or years and were out of money and enthusiasm. In contrast, McCready designed his aircraft out of aluminum poles, piano wire, adhesive tape, and mylar. He designed his plane to be flown, crashed, repaired, and modified. McCready was able to crash and redesign his plane once or twice a day until he got it right. And he made sure there were no cameras around until it was tested and working.

Build for function, not beauty

First and foremost, your equipment should work well. It is natural to feel embarrassed by its crude, homemade appearance. Don't be! Commercial ham equipment has been designed and redesigned several times before you ever see it. Moreover, commercial equipment is packed with custom displays, custom heat sinks, custom cabinets, and even unique integrated circuits. You can't compete with that, so don't try. Instead take pride in your crude prototype.

When you see designs for homebuilt equipment in the ARRL handbook or in ham magazines, the equipment is usually quite attractive. But if you read carefully, you will find that the unit in the picture is prototype number five or even number one hundred. You probably don't have time to rebuild your station five times just to achieve beauty. Also, if you attempt to duplicate one of those units, you'll find you must use exactly the same parts and circuit board the author did. This means buying a custom circuit board and perhaps a complete parts kit from some manufacturer. If you try to substitute parts, I can almost guarantee it won't work.

Get smart guys to help you

This piece of wisdom from Edison may be limited in usefulness to the basement ham, but it's still interesting. Edison was the first fellow to industrialize the process of research and development. Once he acquired financial backing, he hired a whole team to work on his projects. He didn't try to do everything himself. The quantity of Edison's inventions can be partly explained by the number of competent guys he had working for him. Edison and his lab were credited with inventions that represent the work of many lifetimes. Even for a guy who works day and night, there are limits on what one fellow can do. If you're a one-man show, it's important to limit yourself to projects you can complete. Be sure to use the library to find previous work in your area. Although real ham homebuilders are rare, homebuilding is the most fun if you can find someone in your area to share your triumphs and problems.

Assume as little as possible

Edison looked for guys who had the right attitude about R & D. Many job applicants have a knack for sounding good in an interview, but they turn out to be more glib than useful. Edison's interviews began with asking the applicant dozens of practical questions about ordinary objects such as, "What is asphalt made of?" Another of his tricks was to take potential employees out to lunch. If they put salt and pepper on their food without tasting it first, they were in big trouble with Edison. Edison's selection method is probably extreme, but it does illustrate a cardinal rule of research. ***Never assume anything about the project without good data to back up your starting assumption.***

For example, I had known for forty years that resistance in series with a transistor emitter was important for the thermal stability of an RF amplifier. However, I believed that any emitter resistance would inevitably reduce the signal output from that stage. In other words, I thought stability and gain were a tradeoff. Recently I took the time to try different values of emitter resistance over the entire range of possibilities. I was amazed to find that maximum output occurred not at zero ohms, but at a certain significant value, 300 ohms in my specific amplifier. When you do experiments like this, write down the details! You'll want them later.

Many inventions are frustrated by assumptions that turn out to be limiting. For example, weapons inventors were limited for centuries by the concept that firearms had to be ignited by flintlocks. When you consider the practical difficulties of inventing a waterproof, rapid-firing, breech-loading flintlock rifle, it's no wonder that firearms were essentially unchanged for 200 years. Millions of lives were probably saved by this fixation on flintlocks. Narrow-mindedness in inventors isn't always bad for society, but it sure restricts innovation.

The genius of trying

Magic happens when you actually sit down at your workbench and try to do something. You may have thought about the problem in spare moments for weeks, but when you actually have the work in front of your face, ideas pop into your head as if by magic. For this reason, many people, like Edison for example, were well known for working non-stop all night. Once you get the momentum going, it can be wasteful to stop. Otherwise you may not remember all the details when you get back to work hours or days later.

I used to work with a patent attorney, Robert E. Harris, who always put everything off until the last minute. I kidded him about his procrastination. He answered me seriously with an explanation that went something like this: "I do it deliberately," Bob said. "I find that in order to write a patent application, I need complete concentration and nothing gives me that intensity like an approaching deadline. In order to write a good patent, I must have all of the prior inventions in my head at the same time. If I just put in just a few hours, by the next day I will have forgotten important details. For that reason, the day before the deadline I go into seclusion and work all night if necessary."

When you get stuck, do something else for a while

Unfortunately non-stop work slows to a crawl when you run out of significant new ideas to try. As long as you keep sitting in front of the problem, you will keep finding little variations to try, but as the hours go by, you will become more and more tired and your ideas will become less and less creative. To escape from this trap, get up from the workbench and do something else. Take a walk, take a shower, or go to dinner. When you're not in front of the work, you can't do anything with your hands. Since you've been concentrating so long on the problem, your brain will continue to work on the problem long after you leave the workbench. Because you can no longer try out small, uninspired ideas, your mind must wander farther and you'll find you are thinking seriously about radically new concepts.

For example, when Edison was searching for his light bulb filament material, he was stuck on the idea of using an inert metal filament. An inert metal would not react with oxygen or with residual gasses that might remain in the light bulb. If Edison had been able to carry through his "try everything" philosophy, he would have ultimately tested every metal. Therefore, he would have eventually tried tungsten and that's what filaments are made of today. But perhaps tungsten wasn't available 130 years ago.

Edison was particularly stuck on platinum as a filament material. Yes, it was expensive, but it seemed to work beautifully and gave a bright yellow-white light for a few hours. Unfortunately, after a while a segment of the platinum wire would become thin and abruptly melt, thereby ruining the bulb. A related problem was that the resistance of platinum was too low. This meant that a long, very thin platinum wire had to be used to make the filament compatible with his 100 volt power source.

Edison and his team realized that, if they could detect the sudden resistance rise of the filament as it started to fail, they could turn down the current and keep the filament intact. In theory, a temperature/ current regulator could allow the bulb to last indefinitely and would make it almost immune to power surges. The team expended a great deal of effort to invent the regulator, but it never worked well enough. Finally while away from work, Edison thought that the whole idea of a metal filament should be reconsidered. From metals he turned to carbon filaments. Carbon was cheap, had an inherently high resistance, and it didn't melt. Carbon turned out to be a practical answer he could use and in the end carbonized cotton thread became the filaments in his first commercial bulbs.

Developing your own basement electronics laboratory

Before you can begin building ham equipment you will need some basic tools and materials. Notice that, in addition to books, you will need both heavy-duty tools and light-duty ones. If you use tools that are too fragile, you may ruin the tool. If you use tools that are too heavy, you may ruin your project. The following list is incomplete, but it will get you started.

1. Buy an ARRL Amateur Radio Handbook

An R&D hobbiest can't hire a staff of underlings and consultants, but he can get advice from guys who have done it all before. ***The first investment a new American ham should make is The ARRL Handbook for the Radio Amateur.*** This handbook is as big as a phone book. It's published every year and the latest edition has all you need to understand the breadth of our hobby. It will also give you a good background in basic electrical principles and some detailed descriptions of do-it-yourself construction projects. I presume (hope) that similar texts are available in other countries and languages, but I'm not familiar with them.

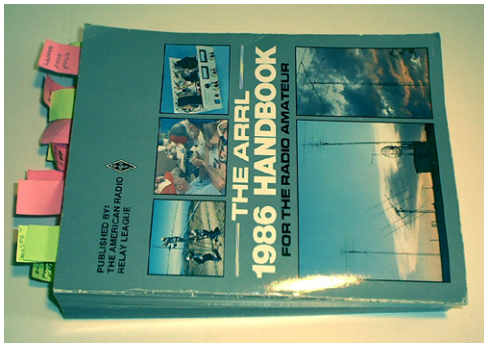

The ARRL Handbook , 1986 edition

Yes, the above handbook is way out of date. I also have a brand new edition, but I have two criticisms of modern ARRL handbooks. First, they are so huge, they discourage most people from sitting down and reading them. The second limitation is that modern hams no longer build complex receivers and transmitters. So projects like this are no longer described in detail. In fact, reading my up-to-date handbook I got the impression that building complex transmitters and receivers is impossible for amateurs. That's not true and that's why I'm writing this book.

Other than that, the latest edition handbook is a great reference to own. It covers all the latest exotic technology and you can use it like an encyclopedia. For actual homebuilding I recommend a handbook from the 1980s. During that decade hams were still building good equipment from discrete transistors. Handbooks from the 1970s and earlier usually describe projects that are unnecessarily primitive. The projects in the Handbooks from the 1990s and later usually contain integrated circuits which don't teach you anything about how your project works. The authors of the present day handbook don't expect anyone to build serious receivers and transmitters. Therefore the projects tend to be "toys" with many limitations. You should be able to find a handbook from the 1980s at a ham radio swap fest. Or maybe you can buy one from a ham in your neighborhood who has no interest in homebuilding.

2. A heavy duty, 100 watt soldering gun and a fine pointed small 25 watt pencil or gun

You need both. The big soldering iron is important for soldering antenna wires and heavy work. A fine pencil is essential to solder the leads on fragile components like transistors and integrated circuits.

3. 60/40 Rosin core solder, fine and coarse sizes

Rosin core solder (60% tin/ 40% lead) is used for joining wires whenever reliable electrical conduction is the primary goal. The rosin flux is built into the wire-like solder, so you never have to apply solder flux. Acid core solder is used for structural purposes. It is not usually recommended for electrical use because the acid continues to corrode the metal for years afterwards. Eventually residual acid may result in a poor electrical connection. However, ***a roll of plumber's acid core solder is nice to have around.*** Sooner or later you will be building an antenna or other project that forces you to solder copper wires to steel. Since nothing else works, a bit of acid core solder or acid solder flux can be a lifesaver.

4. Solder Wick and Alcohol

You'll also need some ***Solder Wick***® for sopping up excess solder. Solder Wick® is just a small coil of fine, multi-strand copper wire specifically made for removing solder from circuit boards. The free end of your roll of wick is placed on top of the excess solder. The wick and solder are then heated with your soldering iron. The wick sucks up the excess solder. Before you can use it again, you must clip off the used end of the wick with clippers. This sounds clumsy, but it works well. There are also pneumatic pump-like devices called ***solder suckers*** that suck up surplus molten solder. They're OK, but I prefer solder wick.

Rawn
T-E 35
ROSIN FLUX
REMOVER
70% ALCOHOL

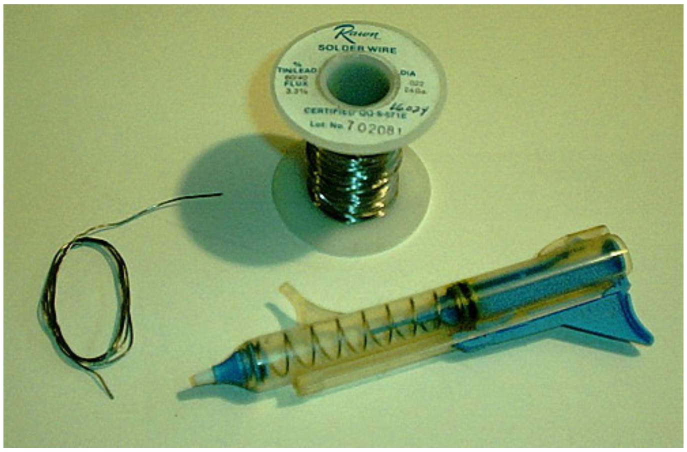

Rosin core solder and solder sucker

Iso-propyl alcohol (rubbing alcohol) is useful for cleaning the excess rosin flux off circuit boards. I put a few drops of alcohol on the board, then use a small plastic brush to scrub off the rosin deposits. Denatured ethyl alcohol also works, but not as well. There are also commercial flux cleaners, but the ones I have used are a bit harsh and tend to dissolve paint.

5. Fine needle nose pliers and diagonal cutters for bending and clipping tiny leads

These should be high quality and your best pairs should be almost small enough to fix watches. In fact, if you use modern "surface mount" electronic components, you will also need fine pointed tweezers to manipulate the parts. ***A cardinal rule about delicate tools is NEVER use them on large wires and parts.*** These fragile tools will be ruined instantly if you try to cut steel wire with the small diagonal cutters or use the needle nose pliers as a wrench. A hemostat can be useful for holding small parts in place while you solder. A non-conductive plastic screwdriver is helpful for adjusting trimmer capacitors in situations where the adjustment screw is floating above ground.

The yellow-handled tool in the collection below is a wire stripper for removing plastic insulation from wires. A pocket knife is also nice to have for stripping insulation off the ends of enameled and Teflon insulated wire. These kinds of insulation cannot be removed gracefully with a wire stripper and must be scraped and carved off with a knife blade. A set of tiny jeweler's screwdrivers is also useful.

6. Heavy needle nose pliers, heavy-duty diagonal cutters and small wrenches

You will often need large versions of the delicate tools. Use them when appropriate.

You will also need a vise, assorted files, a hacksaw, and hole reamers for shaping and mounting PC boards and heat sinks.

7. Microscopes & reading glasses

Even if your eyes are much better than mine, you'll need a strong magnifying glass to inspect your solder connections. Often tiny whiskers of solder or bits of thin wire short out connections. Other times a solder joint looks OK from a distance, but under extreme magnification, the piece is not actually making contact with the desired terminal. These problems usually can't be seen with the naked eye and you might spend hours looking for a problem you could have noticed immediately with a glass. ***I routinely examine each solder connection with my lens before moving on to solder the next component.*** I use pocket microscopes, a strong magnifying lens, and over-the-counter, strong (+3.50) reading glasses for this purpose.

POCKET MICROSCOPE

8. A set of wood carving gouges

"A set of what?" you ask. Wood carving gouges are a kind of delicate wood chisel with a cupped end. I use them for cutting traces on blank printed circuit boards. You will find they work well for making one-of-a-kind prototype printed circuit boards. In my opinion, carving a circuit board is superior to any other method I have seen for homebuilt radio frequency circuit boards. (Other folks prefer the Superglue and disk method described in chapter 6.)

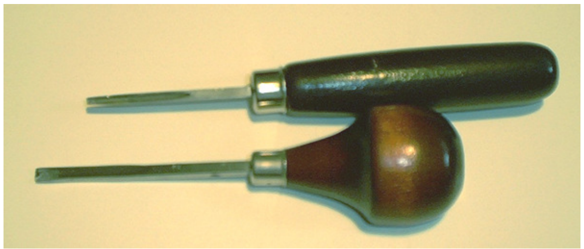

9. An electric drill

An electric drill is used to make holes in PC boards and heat sinks.

9. Chapter 3, Harris

20 Piece
General Purpose
Drill Bit Set
WOOD
METAL
PLASTIC

10. Male Thread taps

Rather than assemble your entire project with machine screws and nuts, you'll find it's often simpler and more professional to tap threads into the aluminum heat sinks and brackets. I find 4-40 and 6-32 threads to be the most useful sizes.

11. Freeze spray

TECHSPRAY
FREEZER

One of the frustrations of electronics is finding intermittent connections. One minute your device works, the next it doesn't. You open up the cover, expose the circuitry and … it starts working again. It's hard to fix something that's working! Spraying refrigerant on suspicious solder joints can often reveal which connection is at fault. Sometimes a heat gun (or hair dryer) can be used the same way, but they are not so focused. Heat guns are useful for mounting heat shrink insulation on terminals.

12. A high quality multimeter

Modern digital "multimeters" measure voltage, current, and resistance. Fancy ones may also measure frequency, conductance, capacitance, decibels, temperature, and other parameters. All multimeters are based on a high input impedance voltmeter. A quality modern meter has an input resistance (impedance) of 10 million ohms. High impedance is needed so that the measurement doesn't load down the circuit and alter what you are trying to measure. Old-fashioned electro-mechanical meters have impedances as low as a few thousand ohms.

A quality meter is also extremely accurate. Measurements of a voltage or resistance will be correct to several decimal points. ***Don't be cheap with this item.*** It is the core of your electronic measurement capability. If you expect to take the meter outdoors to work on the car or up onto the roof to do antenna work, then maybe you should also buy a cheap multimeter that you can afford to break.

13. A high quality oscilloscope

The oscilloscope is another foundation of your laboratory. The traces on the screen tell whether your device is working and how well. Without a scope, you are almost blind. Perhaps the most amazing achievement of people like Edwin Armstrong is that they were able to do their work by inferring the function of circuits from secondary measurements. For example, the DC plate current of an RF amplifier tube dips when resonance is achieved. However, it's so much easier to just look at the signal with a scope and WATCH the actual sinewave while the circuit is tuned.

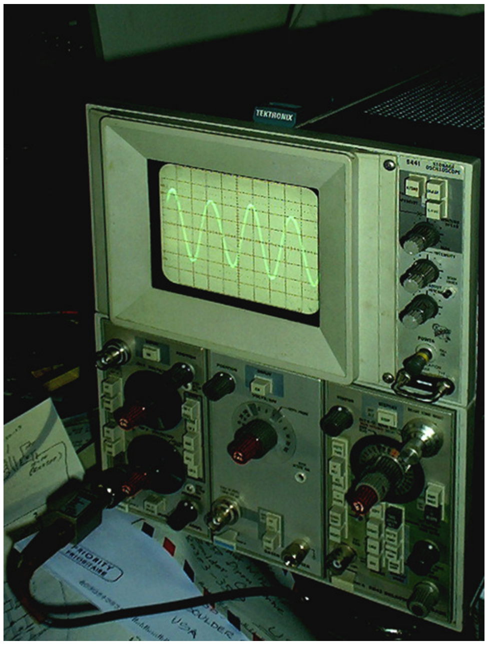

A first rate oscilloscope might cost $10,000 brand new. But there are advantages to living in the 21st century. One of them is that 20-year old $10,000 oscilloscopes are all transistorized, still first-rate quality, and you can buy a used one for $300 or $400 dollars. The prime consideration is that the scope must have a frequency rating higher than the frequencies you will be working with. For HF ham radio, 50 MHz is enough, but you can find used scopes that will go as high as 1,000 MHz.

The oscilloscope is connected to your project by a ***probe***. A probe is a 3 to 6 foot long coaxial cable with a ground wire clip and a little “grabber” at the end that hooks onto the wires carrying the voltage waveforms you want to look at. Probes usually have a 10:1 voltage divider that protects the oscilloscope from high voltages you may be measuring. Not all probes are created equal. ***For high frequency radio work you need a short probe with minimum capacitance***. If you buy a probe, look at the specifications to see what kind of capacitive load you are putting on your circuit.

For example, if your circuit is tuned by a variable capacitor that ranges from 5 to 60 pF capacitance and your probe has 50 pF of capacitance, the probe will totally dominate the circuit tuning. In general, a good RF probe has a short cable and a short ground lead. Of course even 5 pF is a significant load. So when I’m tuning a circuit, I try to tune one stage of an amplifier while I use the scope to monitor the signal in the following stage. That way, the tuning of the first stage will not be affected by the probe.

14. A frequency counter

As soon as you build your first ham transmitter, you will need to prove that your transmitter is operating inside the ham band. Also, we hams are supposed to limit our transmissions to one frequency and not splatter all over the band. By definition, a pure sinewave is a single frequency. My solution is to routinely monitor the frequency and shape of my RF sinewave right at the base of the antenna. If it looks like a good sinewave on the scope, the signal is almost certainly clean.

You can estimate the frequency of the sinewave by using the horizontal marks on the oscilloscope screen to measure how long it takes to complete a complete sinewave. If you count grid squares on the screen, you can measure the frequency of a sinewave to about 10 or 20% accuracy. Unfortunately, that isn’t nearly good enough. For example, if your transmitter frequency is 7.05 MHz, one complete sine wave will take 0.142 microseconds. You won’t be able to read it closely enough be sure it isn’t really 0.143 microseconds and just outside the 40 meter band. You really need a frequency counter.

Some sophisticated oscilloscopes have built-in frequency counters to measure the exact frequency of a signal. Almost certainly, you will have to buy a separate frequency counter to tell you exactly where you are to the nearest Hz. You can buy a new, quality counter for $2,000 or more. Or you can get a new, cheap frequency counter for $200. The best bargains are first-rate used frequency counters. When start-up electronics companies go bankrupt they are often forced to liquidate their equipment at bargain prices. You may be able to buy a good one for under $100.

15. A quality short wave receiver

Aside from using it to hear other hams, a good quality, commercially manufactured shortwave receiver can serve as a laboratory instrument. A modern receiver is so well calibrated that it may be used as a substitute for a frequency counter. Also, it’s important to be able to listen to your own signal in a receiver to be sure that it doesn’t have subtle defects that may be hard to see on an oscilloscope. Ideally, you should have both a counter and a calibrated receiver.

16. A laboratory power supply

A laboratory power supply allows you to apply voltage to a circuit cautiously. This will usually prevent component damage due to wiring errors or other problems. Meters tell you how much current is being drawn and the voltage applied. The voltage is adjusted with a control pot. You can start with one volt than gradually turn it up if the current flow isn't excessive. This particular old power supply is actually three separate supplies. For example, you might use the 5 volt supply to power a microprocessor, while the other two variable supplies could be set up to deliver + 12 volts and – 12 volts for operational amplifier circuits.

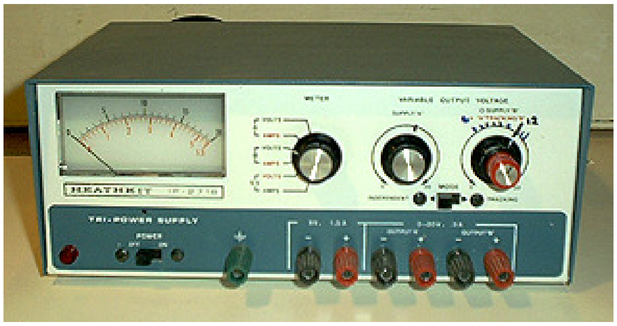

17. An RF frequency generator

This is a tool you will eventually want. But if you have everything above, it isn’t essential. The RF frequency generator allows you to inject a sinewave of known amplitude and frequency into an amplifier so that you can align it. When aligning a homebuilt receiver, it is nice to have a known test signal you can listen to any time you want.

18. A capacitance meter

OFF
000.0
B+K PRECISION
DYNASCAN CORPORATION
820 CAPACITANCE METER

While not essential, you can use a battery-powered capacitance meter to sort out your junk drawer and give a reliable indication of capacitor sizes. I find it extremely useful for determining the maximum and minimum set points on trimmer capacitors.

19. Catalogs of electronics parts suppliers

You will need at least 3 or 4 catalogs. I like RF Parts Company, Jameco, Digi-Key, Newark, Mouser, and Radio Shack. Hopefully you have a Radio Shack store in your area in case you need some routine part in a hurry.

Tech
JAMECO
MOUSER
Digi-Key
RF PARTS

20. Your very own junk collection

OPTICAL
POTS
KNOBS
Relays
METERS
LEVER SWITCH SPDT-MA Toggle MINI
FERRITE CORES
MOLEX
DIAL LIGHTS
XISTORS
XTALS
RF CONNECTORS
BATTERY HOLDER
SPEAKERS
POWER RESISTORS

As Edison said, it is important to collect junk parts so that you aren't continually waiting for parts to arrive in the mail. Considering that a capacitor or potentiometer costs a few dollars new, it is easy to spend hundreds of dollars on a project. If you are **not** getting most of your parts from old TVs and ham swap-fests, your project will cost a fortune and will proceed slowly. Never throw away an old radio or computer without first cannibalizing it for useful parts. Organize your parts in bins, parts drawers, and labeled boxes so you can find them when you need them. If you can't find a part when you need it, junk is just junk.

21. A laboratory notebook

Your memory isn't half as good as you may think. Write down all your experiments, your triumphs, and especially your failures. The experiment that doesn't work is just as important as the experiment that did.

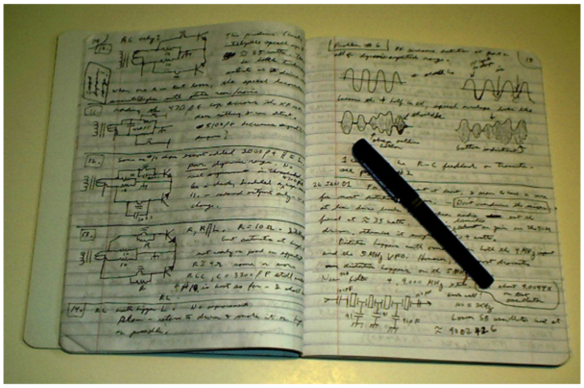

22. A simple calculator

You probably already own a calculator that will be adequate for the simple component value calculations you must do. Square roots may be the most complex calculation you'll perform on your way to your first two-way ham radio contact.

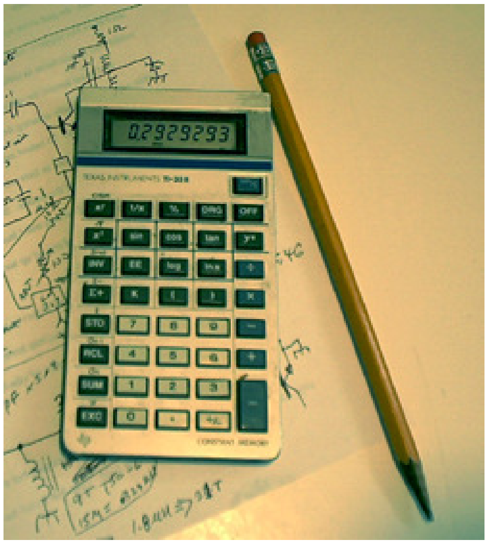

When I was in engineering school, the calculator age had not yet arrived. We young engineers swaggered about carrying big sliderules that hung from our belts like swords. We were very cool. When I sat down in class to take a test, I would ceremoniously draw my sliderule from its sheath and check to see that the upper scale was perfectly aligned with the lower scale. If it wasn't lined up perfectly, I used the blade of my pocketknife as a screwdriver to adjust it. For me this ritual was something like a parachutist checking his harness and D-ring one more time before leaping from the airplane.

For you whippersnappers who haven't used sliderules, these antiques do logarithms, calculate trigonometric functions, take squares and square roots, and do almost everything you can do on a simple "scientific" calculator. The reason for this archaic story is that sliderules don't do decimal points. Therefore, to get the right answer, ***we had to have a FEEL for the math***. We had to be able to estimate the answer so that we would know what order of magnitude it would be and where to put the decimal point. In other words, we had to know what we were doing. The sliderule expanded our skills. It did not replace them.

When calculators came out, engineering students suddenly began to fill their test papers with random numbers. Beginners think that, if they push the buttons, the calculator is doing the thinking for them. Wrong. It turns out that calculators really aren't different from sliderules. The students must estimate the answer in their heads so that they will know if they have pushed the right buttons. Once the students mastered the ability to estimate, calculators became a boon to engineering.

23. Software for experimenters

I hesitate to mention circuit simulation software. I am a rebel against the modern age of smug engineers who have never soldered a wire. I don't like the trend of increasing specialization and generalized ignorance that is spreading through the technical industry. Spice programs are much like calculators. They're wonderful if you can estimate what the circuit should do before you activate the simulation. Once you get very far into building ham gear, you will probably get tired to doing everything the hard old way. There are many different simulation programs available that allow you to emulate a circuit on your personal computer before you build a real circuit. I found it particularly useful for designing frequency filters such as those used in chapter 15. Like most modern conveniences, this one is truly marvelous. However, Spice should complement actual circuit testing, not replace it.

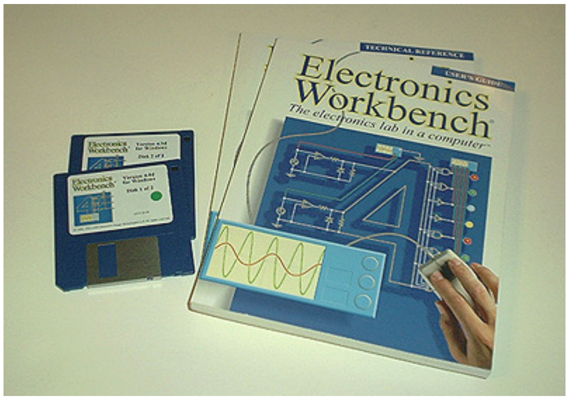

I use an old **Electronics** **Workbench** "Spice" program, but there are probably many newer ones that work as well or better. Mine is a simple program that will not simulate some components, such as crystals or some kinds of circuits like oscillators. I really like it for simulating filters. If I need to build a filter with a certain cut-off frequency, but I don't have the right parts, I can simulate substituting parts and find out how critical the values are.

Actually, it's so much fun trying out circuits with so little work, that it's almost addictive. Also, other problems in my design become obvious that I had never thought of. For example, I built a filter for a ham transmitter that was designed to eliminate interference to the neighbors' TV reception. (See chapter 9.) Any transmitter signal has minor "impurities" in its frequency spectrum. This means that it can easily be radiating weak signals on the TV channels. A filter will reduce these harmonics. Before I built my filter, I took the time to model it on the Spice and found that, as I had planned, it severely reduced interference on channels 2, 3, and 4. Above those channels, the attenuation of possible harmonics was not nearly as great. For the upper UHF channels, there was hardly any attenuation. Once I saw the problem, it was easy to add a couple more stages of filtering to insure that *all* TV channels were protected.

24. Test leads and experimental "socket" circuit boards

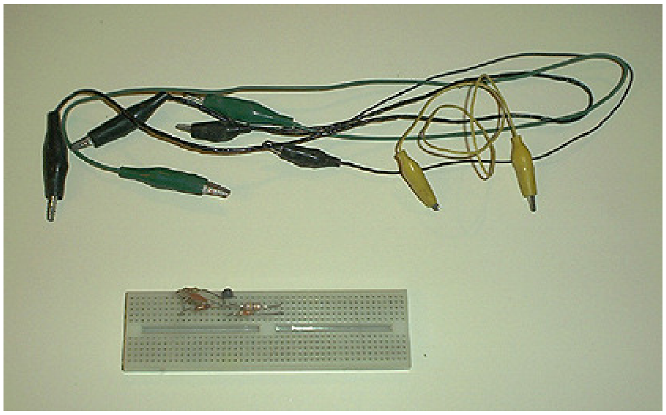

Every electronics lab has handfuls of test leads and a few temporary plug-in circuit boards. Although I routinely use these, I hesitate to recommend them. ***ANY ELECTRICAL CONNECTION THAT ISN'T SOLDERED CAN'T BE TRUSTED!*** Yes, it's true that these gizmos often work, but many times I have been led to believe that parts were bad or that circuits didn't work when they actually worked fine.

Let me illustrate: Once when I was in the Air Force I had to wire some explosive squibs on a 1,500 pound cargo parachute load. The squibs were supposed to explode and deploy the parachute. I passed the bare, scraped copper wires into two tinned metal eyelets, then wound the wire through the eyelets again and again until the eyelet holes were stuffed with clean, bare wire. Then I wrapped the remainder of the bare copper wires tightly around the outside of the metal eyelets and wrapped the whole thing securely in tape. I thought there was no way that the copper wire wasn't in good contact with the eyelets! The huge box fell 3,000 feet and crashed into the dirt at a couple hundred miles an hour. The squibs never fired. I checked the firing circuit with my meter and found the proper 3 volts across the squibs. I unwrapped the tape from the eyelets. The squibs exploded as soon as I tugged on the bare wire. After that day I soldered my squib wires and never had another failure.

If you buy a socket board like the one shown above, be aware that the contacts are easily damaged by stuffing component leads into the board that are too big. For example, ½ watt resistors are barely OK, but 1 watt resistor leads will permanently expand the spring contacts.

25. Pliers for crimping connectors and eyelets

What about the special pliers used to crimp connectors and eyelets? Some metal eyelets and lugs are designed to be installed onto wires using special crimping pliers. Yes, crimped connections can be fairly reliable in the short run. However, in my experience crimped wires pull out easily and after several years they often become open circuit. I have seen dozens of failures on old equipment. I still use eyelets on leads, but after I crimp them, I solder them as well. Surprisingly, eyelets that are screwed down to aluminum metal seem to make a reliable contact with the aluminum. However, I still solder the copper wire onto the tinned eyelets.

In conclusion,

The equipment described above should get you started. Everyone has different ideas about what works best and what is essential. For what it's worth, the above list pretty well describes my workshop.

CRYSTAL SETS TO SIDEBAND
© Frank W. Harris 2010, REV 12

Chapter 4
HERTZIAN WAVES IN THE BASEMENT

Crystal set radios and ancient spark gap transmitters from the 19th century may seem too simple and too far removed from ham radio to be worth building. If you're already familiar with basic electronics and early radio history, you may not learn anything from this exercise. If you can't wait to build a real ham rig, please skip ahead to the next chapter. On the other hand, if you have little electronics experience, there are worlds of lessons to be learned from old technology. And if you've never built simple radios before, you should find it fun.

The nature of radio waves

Before we build transmitters and receivers, let's review radio waves. When we understand what radio waves are, the technology to generate and receive them becomes more obvious. An electromagnetic wave is an oscillation in free space that radiates out away from its source at the speed of light. It is called ***electromagnetic*** because it is both electric field energy and magnetic field energy. The wave oscillates or changes back and forth between these two forms of energy as it travels.

Propagating across vacuum

From our experiences with magnets and static electricity, it's hard to visualize how a magnetic or electric field can travel millions of miles across the vacuum of space. In our experience these fields are tightly localized around the device that generated them. How can a magnetic field exist isolated in a vacuum, perhaps even light years away from the nearest atom? Suppose we could somehow magically generate a magnetic or an electric field in space, miles from the nearest object. Would the field just sit in space forever waiting for an object to pass by and be influenced by the field?

Let's suppose there is a refrigerator magnet floating in the void of space. Its magnet field will be at rest in the space surrounding the magnet, just as it does on your refrigerator. As always, the magnetic field will reach out its usual distance of a quarter inch or so. However, if the magnet were to suddenly vanish, the energy in the field would lose its "container" or "anchor" and be loose in the void.

The same scenario can be proposed for an electric field: If a flashlight battery were floating in space, the electric force would extend perhaps an inch out into the space in a halo around the two battery terminals. Again, if the battery suddenly vanished, the electric field energy would lose its generator and be stranded in the void. Without its anchor, it would spill out in all directions.

Oscillation occurs when two form of energy rhythmically change back and forth

When magnetic and electric fields are turned loose in space, what becomes of them? As James Maxwell first explained over a century ago, electric fields and magnetic fields are intimately related. It turns out that ***a changing or moving electric field generates a changing***

magnetic field and vice-versa. As the energy continues to “spill” out in all directions, the energy oscillates back and forth between these two kinds of fields. This connection is not "obvious” or intuitive. If it were, the Greeks, Chinese, or Egyptians would have described and exploited it long ago.

Mechanical oscillators

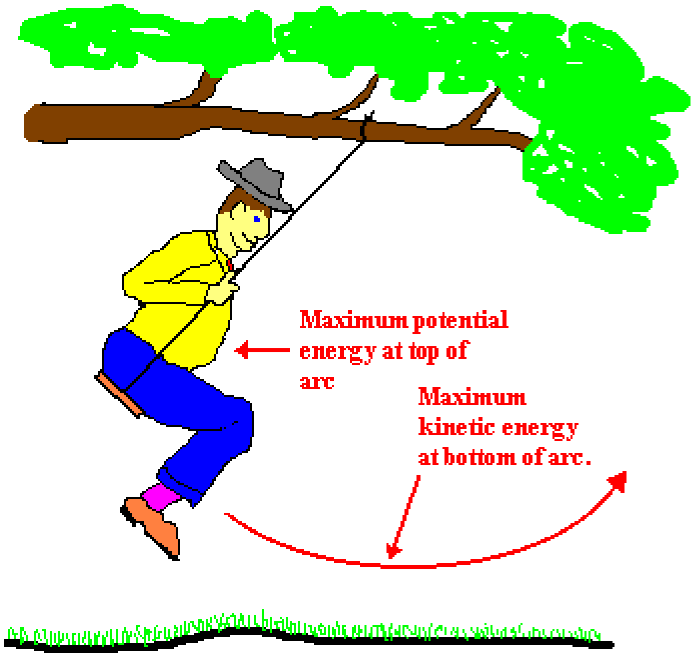
Maximum potential energy at top of arc
Maximum kinetic energy at bottom of arc.

Many physical devices in our world oscillate, so oscillation between electric and magnetic fields should not be a surprise. An oscillation in nature can be described as energy spontaneously transforming from one form of energy into another then back again. For example, as a clock pendulum swings back and forth, the pendulum acquires the kinetic energy of motion as it swings through the bottom of its arc. Then, when the pendulum swings back uphill, the energy contained in the kinetic energy is returned to gravitational potential energy. When the pendulum reaches the top of its swing, it momentarily comes to a complete halt, turns around and races back downhill. When it is at the top, the energy is all “potential.” A boulder sitting on the lip of a cliff doesn’t seem to have any energy until it is nudged off the cliff. The fellow standing at the bottom of cliff can testify that the rock has plenty of energy when it slams into the foot of the cliff. (That assumes, of course, that he survives.)

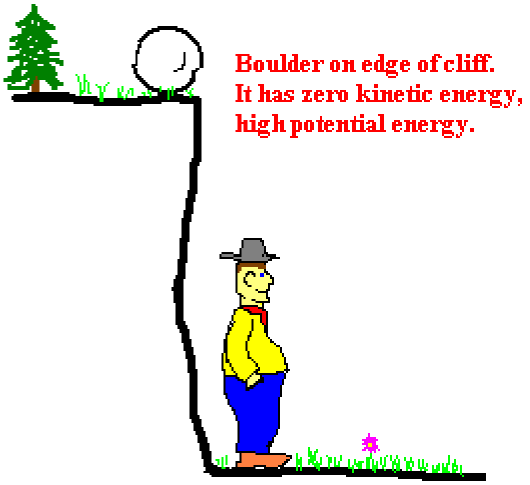
Boulder on edge of cliff.
It has zero kinetic energy,
high potential energy.

To reiterate, an oscillating pendulum switches its energy back and forth between kinetic energy and potential energy. Notice that the length of the pendulum establishes the frequency of the oscillation of a pendulum. This is because gravity is constant and lightweight objects fall just as fast as heavy objects. If you ignore air resistance, the frequency of the pendulum swing is determined solely by the length of the pendulum arm (and of course the acceleration of gravity). This makes a pendulum good for keeping a mechanical clock running uniformly and accurately. Similarly, once the frequency of a radio wave is established, it doesn't change frequency as it races through space and becomes weaker.

In summary, free space (which is literally "nothing"), can support magnetic or electric field energy, but only temporarily. To be maintained, a magnetic field needs to be generated by a device. ***A magnetic field can be generated temporally by a nearby collapsing electric field. As the magnetic field collapses, it produces a temporary electric field in the adjacent space.*** This seesaw produces a radio wave traveling outward across the void at the speed of light.

What's special about the speed of light? Good question! Einstein's laws of relativity teach us that time and space are frozen like a 4 dimensional ice cube. Most people are aware that light speed is the ultimate velocity. What is rarely understood is that light speed is really "the speed of time itself." Relativity is almost never studied deeply, even by engineers. This is perhaps because, until the invention of portable atomic clocks, the evidence proving relativity was less than overwhelming. For example, when the first GPS satellite was launched, the software had "switches" that allowed ground engineers to turn off the adjustments for Special and General Relativity in case relativity turned out to be wrong. They weren't. The crazy theories turned out to be laws. Our universe is weird and profound far beyond what 99.9% of humans have the privilege to notice. However, these subjects belong in another book.

Getting back to 20th century radio: Transmitter antennas are designed to generate either a rapidly changing electric field, or alternatively, a rapidly changing magnetic field. The antenna is placed out in the open with free access to the sky. The electric or magnetic fields around the antenna create the opposite kind of field and the result is a free-flying radio wave. The same antennas work well for receivers. As radio waves flash past the metal antenna elements, electric currents are induced into the structure just as if it were temporarily a capacitor or a secondary loop of a transformer.

The LC circuit, the fundamental electronic oscillator

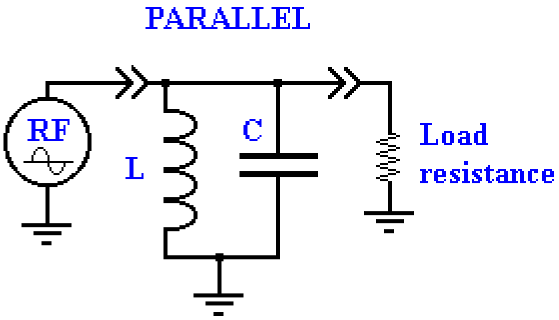

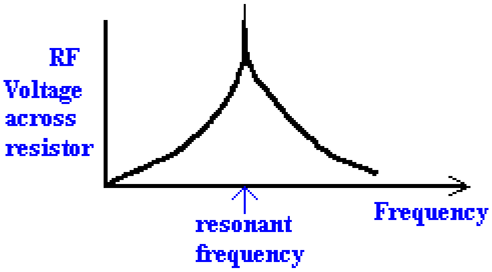

The most fundamental component of all radio transmitters and receivers is the ***capacitor/ inductor parallel resonant circuit***. This basic circuit consists of an inductor wired in parallel with a capacitor. These are called ***LC circuits*** where L is the letter used when calculating inductance and C of course stands for capacitance. If a high frequency sinewave voltage is applied across the parallel LC circuit, there is a specific frequency at which the LC circuit resonates and appears to be an open circuit. At all the other frequencies the LC appears as a load or short circuit. The LC circuit attenuates or eliminates the sinewave at every frequency except one. In this way one radio signal can be “tuned in” preferentially over another.

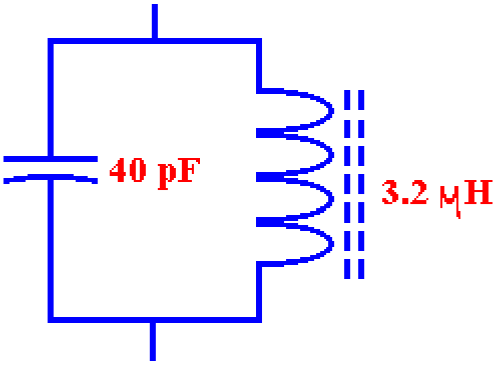

20 Meter RESONANT CIRCUIT

For example, the circuit above resonates at 14 MHz, the 20-meter hamband. The tiny inductor is just 3.2 microHenries. The capacitor is only 40 picoFarads which means 40 millionths-of-a-millionth of a Farad.

The LC circuit is a kind of electric oscillator. It is analogous to a swinging pendulum or a weight bouncing up and down on a mechanical spring. The LC oscillator goes through the same energy cycle as radio waves. First the energy is stored in the magnetic field inside and around the inductor. During the next half cycle, the energy is stored in an electric field between the plates of the capacitor. The energy alternates back and forth between these components until the resistance in the wires dissipates it.

A parallel resonant LC circuit with dual antennas forms a simple transmitter

Significant energy can also be leaked off into the space around the LC circuit as radio waves. Therefore, once we get an LC circuit to oscillate, we are already on our way to generating radio waves. ***If we simply add wires to the ends of the parallel LC circuit, these wires form an antenna to couple the electric field to the free space around it.*** In other words, the simple circuit shown below is a crude radio transmitter.

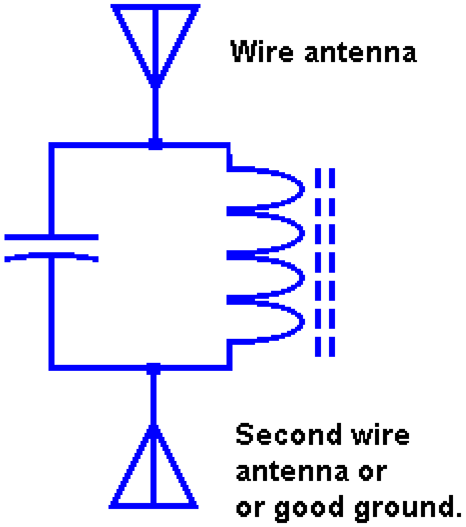

A receiver of radio waves can be built in the same way. Imagine that an identical parallel LC circuit with antennas is tuned to resonate at the same frequency. Now imagine that this second circuit is floating in the void, perhaps miles from the transmitting circuit. When the radio waves radiate past the receiving LC circuit, the electric field component in the radio wave will produce a tiny surge of current in the wires that charge the capacitor. Alternatively, and depending on the orientation of the coil with respect to the radio waves, the magnetic component of the radio wave will induce a tiny voltage to appear across the coil. This is the same as if the inductor were the secondary of a transformer. Once the radio wave has zoomed past, a tiny, miniscule oscillation will remain in the receiving LC circuit, ringing back and forth between the inductor and capacitor.

****

CRYSTAL SET RADIOS

A crystal set radio is an excellent first radio project. They have few parts and are easy to understand and build. Crystal sets were a common toy when I was a kid and my first electronic project was building one. Although we kids played with crystal sets, we didn't really understand them. When they didn't work, we had only the haziest notion of how to fix them. If you bought a commercial toy crystal set, I never saw one that had explanations even remotely as complete as what you are now reading. Starting with the crystal set, later on we can build on what we learn to build transmitters and more elaborate receivers.

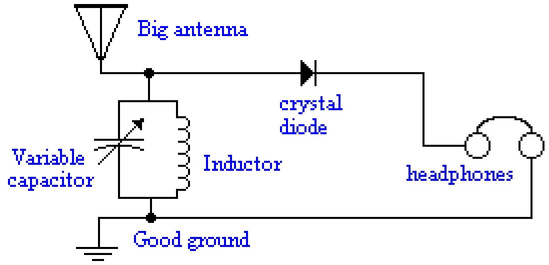

The combination of inductor and capacitor tune in the desired station. That is, the inductor and capacitor oscillate at the frequency of the desired station. The crystal diode rectifies the sinewaves oscillating across the LC circuit. This converts the high frequency sinewaves into low frequency sound frequencies that can be heard in the headphones. Crystal sets have six basic parts. The antenna of course picks up the signal from the air.

The antenna

For crystal sets it is easiest to build an "electric field" antenna. It is usually just a long piece of wire strung out a window or up in a tree. A limitation of a single wire antenna like this is that, when the radio wave generates a voltage on the wire, the current it might produce has no where to go. A simple wire antenna is like one terminal of battery. Yes, the battery has a voltage, but without a connection to the other terminal of the LC circuit, the current has no circuit to flow through. To provide a destination for the current we can add a second antenna. Alternatively we can connect the crystal set to "ground."

The ground

Electrical "ground" is a word that we learn as youngsters, but most people go through their whole lives without ever understanding it. I suspect that the term arose during the early days of telegraph communication in the 1840s. It turns out that wet earth is a fairly good conductor. If you drive two metal stakes into the ground in your backyard and connect a battery to the two stakes, current will flow from one stake to the other. For two stakes about 100 feet apart, the ground has a resistance of about 100 ohms. If you experiment, you learn that most of the electrical resistance to current flow occurs right around the stakes. Once the current gets launched, the electrical resistance only increases slightly as you increase the distance. A metal stake in a backyard in China and the one in your backyard in the US might only have 300 ohms resistance between them. This was a boon for early telegraphers because it meant they only had to string one wire between cities instead of two wires to complete the circuit loop. In practice, using ground as an intercity "wire" is not as reliable as stringing a second wire, but it illustrates the concept of ground.

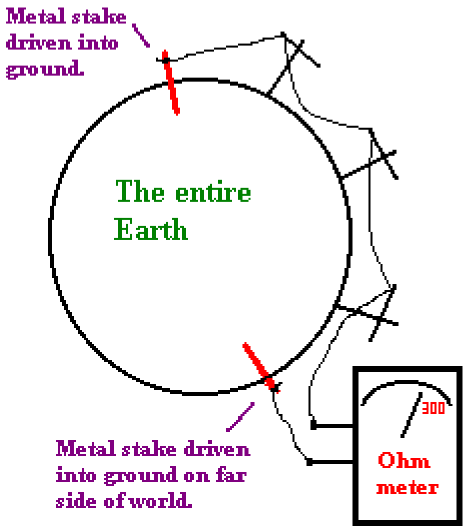
Metal stake driven into ground.
The entire Earth
300
Ohm meter
Metal stake driven into ground on far side of world.

A good electrical connection with ground is an essential component of the crystal set. The most accessible ground for a crystal set or a ham radio station is usually a copper water pipe or a hot water heating pipe. At low frequencies, like standard broadcast AM, ideal electric antennas are very large. Needless to say, being able to use the entire Earth as half of the antenna is often quite convenient.

Dipoles

Unlike low frequencies, at high frequencies, like VHF television or FM radio, the ideal length for an electric antenna is just a meter or so. The higher the frequency, the less distance there is between successive electric and magnetic waves. That is, wavelength decreases with increasing frequency.

Although in theory you could use a stake driven into your lawn for the “ground” path on your TV, it is much simpler to just use a second short antenna oriented 180° away from the “real” antenna. This dual antenna is called a ***dipole*** and is the most common basic antenna design used in ham radio and TV antennas. Although it isn’t obvious, the arrays of thin metal tubes on rooftops are just refinements of the basic dipole antenna. In general, the lower the frequency, the larger the dipole must be in order to work well. A typical ham radio dipole is shown below.

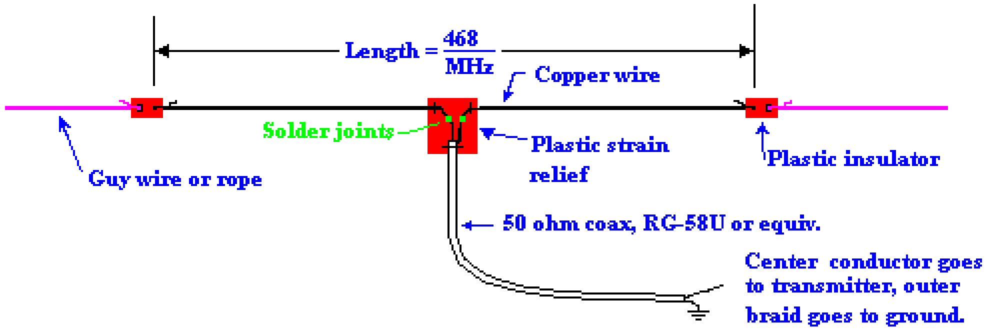
Length = 468 / MHz
Copper wire
Solder joints
Plastic strain relief
Plastic insulator
Guy wire or rope
50 ohm coax, RG-58U or equiv.
Center conductor goes to transmitter, outer braid goes to ground.

For 40 meters the length is 66 feet. The coax feed line can be any length, but shorter is better.

Transmission lines

In the ham radio dipole antenna above, the “arms” stick out in space in opposite directions and snag the passing electrical field. In general, the higher the dipole is above the local terrain, the better the dipole will receive signals. Unfortunately, you and your radio are down on the ground. Climbing up on the roof to listen to the radio or watch TV is inconvenient, to say the least. The problem of how to move radio frequency signals down to the receiver is solved with a ***transmission line.***

A transmission line is a pair of parallel wires separated by insulation. It works very much like a speaking tube in a ship or even the string in a tin can telephone. In all these devices, vibrations are transmitted down a narrow pathway with surprisingly little loss of energy. A terrific example of a mechanical transmission line can be a farmer’s ditch filled with water. Provided the water is flowing slowly, when you throw a big rock into the ditch, the wave from the splash will travel hundreds of yards before it dissipates. A wave in the ditch propagates unchanged for many minutes and travels great distances. In contrast, if you throw the same rock into an open pond, the wave spreads out in all directions and quickly vanishes.

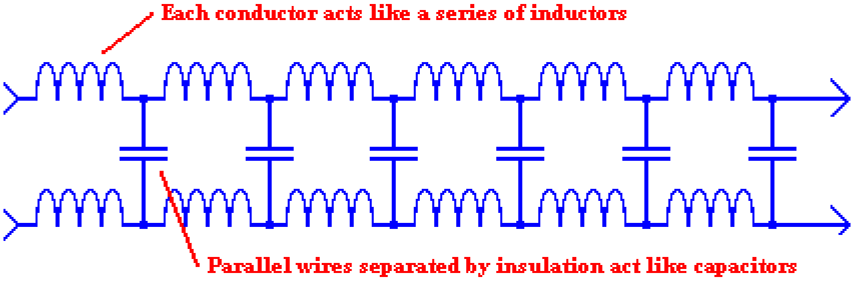
Each conductor acts like a series of inductors
Parallel wires separated by insulation act like capacitors

Two parallel wires separated by insulation make up a transmisison line.

A radio transmission line is a distributed, LC resonant circuit. We saw in chapter 2 that a simple wire has inductance. Similarly any two wires separated by an insulator comprise a capacitor, whether we planned on making a capacitor or not. Consequently, when we run two parallel, insulated wires over any distance, there will be a measurable capacitance between them and the wires themselves will have a significant amount of inductance. To a radio wave, this construction looks like a long, LC circuit without end. As the capacitance and inductance are charged and discharged, the oscillation doesn’t stand still, but rather moves down the pair of wires at nearly the speed of light. As you can see, propagation down a transmission line is analogous to propagation through free space, but it only propagates in one dimension instead of three dimensions. The electric field or voltage generates a current and magnetic field, which in turn generates a new electric field and so on. An example of a simple transmission line consisting of two parallel wires is the flat wire "300 ohm" TV line used to feed older TVs.

Coaxial cable

The round, shielded transmission line used to feed modern TVs is a ***coaxial cable***. Instead of using two separate ordinary wires, the outer conductor of a coaxial cable is a metal cylinder that completely encloses the center conductor. The inductance of the shield conductor is far less than that of a simple wire, but it prevents the radio frequency signals on the inner conductor from leaking out. Even better, it keeps new signals from leaking into the cable and interfering with TV reception.

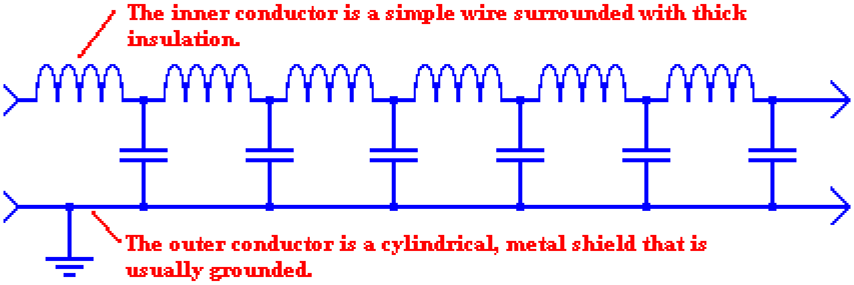
The inner conductor is a simple wire surrounded with thick insulation.
The outer conductor is a cylindrical, metal shield that is usually grounded.

A coaxial cable is a transmission line with a cylindrical, outer conductor shield.

The ham radio dipole antenna described earlier uses type RG-58 coaxial cable to transport the radio frequency signals down into the house. You will notice that the outer shield is connected to ground. This is nearly always the case with coax. You could use cheap TV coax cable for your ham transmitter, but you will find it difficult to work with. The outer shield of cheap coax is just aluminum foil and is difficult to connect mechanically and electrically. In contrast, the outer shield of quality coax is braided copper wire that is easy to cut and solder. It also has considerable mechanical strength.

Transmission line impedance

An abstract characteristic of transmission lines is that, from the point of view of a traveling radio signal, the line “looks like” a specific load resistance. For example, RG-58 coax appears to the radio signal to be a 50 ohm resistor. It isn’t, of course, but the voltage and current levels along the wire suggest that it is. In other words, voltage divided by current at points along the line will give 50 ohms. Another reason not to use TV cable is that TV cable is usually designed for 75 ohms, while most ham equipment is designed for 50 ohms. Similarly, the old flat, brown TV lead-in wire had a characteristic impedance of 300 ohms.

In general, the finer the wire and farther apart the two conductors of a transmission line, the higher the characteristic impedance. Sometimes hams use a wide “ladder line” in which bare copper conductors are separated by an inch or more of air and a few ceramic separators. Ladder line often has an impedance of 600 ohms. Ladder line is useful when transmitter power must be transmitted great distances to get to the antenna. Because a ladder line has little or no insulation in contact with the wires, the small dissipation of energy in the insulation is reduced to the absolute minimum. More importantly, the high impedance means that smaller currents will flow in the wire and less energy will be dissipated in the resistance of the copper. Of course, the trade off is that higher voltages are needed to transmit the same power levels. I know a ham who lives at the bottom of a canyon where radio reception is poor. He uses a long ladder line hung in the trees to connect his transmitter to the antenna, which is located hundreds of yards up on the mountainside.

Diode detectors

The diode is the ***detector*** that converts radio frequency sinewaves into audio frequency electric waves, ready to be converted into sound. The diode is a "one way electricity valve." In plumbing terms, it works like a check valve. The schematic symbol for a diode is an arrowhead pointed at a barrier at right angles to the wire.

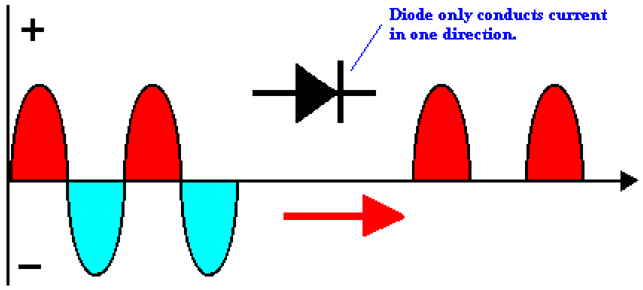

HALF-WAVE RECTIFICATION

(Only one half of wave is exploited.)

In electronics, the convention for ***"positive" current flow is from positive to negative***. Unfortunately, the flow of electrons is from negative to positive. So, what is actually "flowing" from positive to negative is the absence of electrons. Confusing, no? I suspect this convention was established before electrons were understood. Referring to the symbol for a diode, positive current is allowed to pass if it flows in the direction of the arrowhead. Positive current will be blocked by the diode if it attempts to enter the diode from the perpendicular "barrier" side.

Semiconductors usually perform the check-valve function of diodes. A semiconductor is a crystal of an element like silicon or germanium that has a chemical valence of 4. That is, during chemical reactions this element can either take up 4 electrons, or give away 4 electrons. As we shall see shortly, semiconductors can also be fashioned by making crystals out of mixed elements with valences of 3 and 5, or even 2 and 6.

N-type semiconductors

To keep it simple, let's suppose we have a pure crystal made out of silicon, which has a valence of 4. If we put some multimeter (ohm meter) probes across this pure silicon, it will act like an insulator – there will be no significant current flow. However, if we make a new crystal with just a touch of phosphorus impurity in it, suddenly it becomes a conductor. Phosphorus has a valence 5 and is almost the same as silicon in atomic weight. This means that in chemical reactions it normally accepts 3 electrons to complete an outer shell of 8 electrons. But when silicon crystal is contaminated with phosphorus, lone atoms of phosphorus are trapped among a frozen, rigid crystal of silicon. The phosphorus atom fits in the matrix, but it has an extra electron that is "loose" and free to move around the crystal. The electron can't move over to silicon atoms because they are joined with neighboring silicon atoms so that each silicon atom has a stable outer shell of eight shared electrons. However, the extra phosphorous electron can move over to other phosphorous atoms that have already lost their 5th electron. In other words, ***a silicon crystal with just a touch of valence 5 impurity acts like metal.*** It has electrons that are free to migrate through the whole solid. A semiconductor with extra electrons is called an ***N-type semiconductor.***

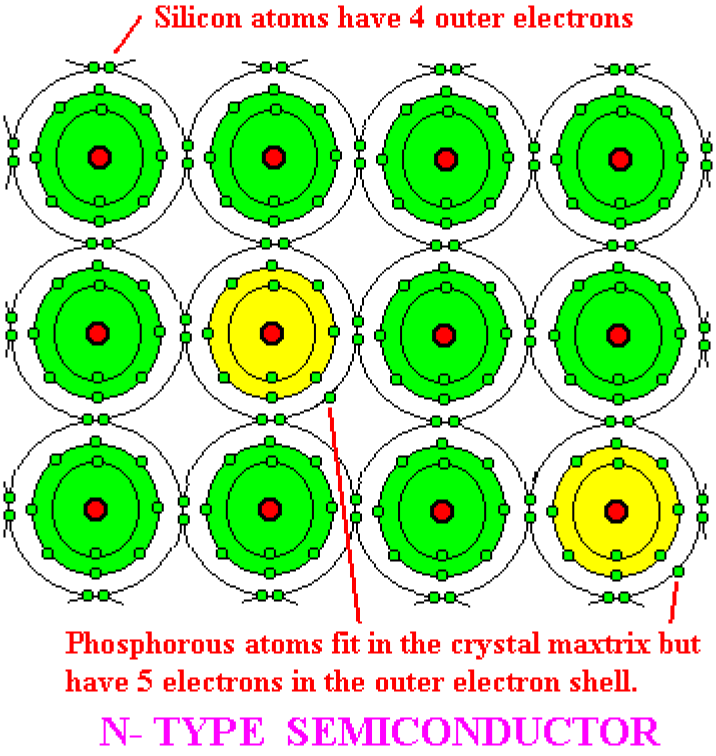
Silicon atoms have 4 outer electrons
Phosphorous atoms fit in the crystal maxtrix but have 5 electrons in the outer electron shell.
N- TYPE SEMICONDUCTOR

P-type semiconductor

P-type semiconductor is a bit abstract. Instead of making a silicon crystal with valence 5 impurity, now suppose we add an impurity such as aluminum, indium, or gallium with a valence of 3. The impurity fits into the crystal matrix, but it needs one more electron to reach an equilibrium of 8 electrons shared with its neighboring silicon atoms. In other words, this semiconductor has "***holes***" in the crystal matrix that can be filled by electrons passing through. Now when you place multimeter probes across a P-semiconductor, it will conduct just like the N-type semiconductor. However, the conduction mechanism is different. With P-type semiconductor, the negative metal probe touching the crystal supplies all the free electrons flowing through the crystal. These electrons are hopping from hole to hole to cross the crystal.

Like phosphorus, aluminum atoms have almost the same atomic weight and size as silicon. Aluminum atoms fit perfectly in the silicon crystal matrix.

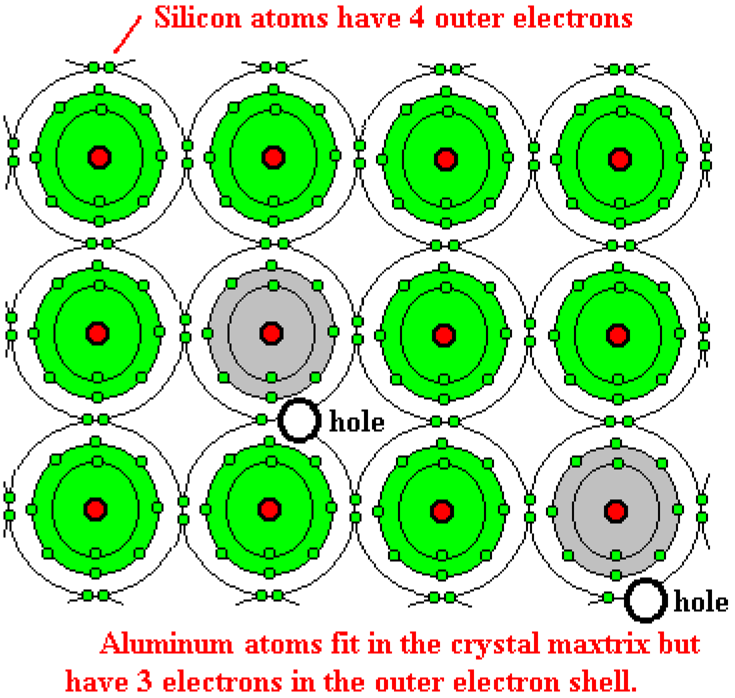
Silicon atoms have 4 outer electrons
hole
hole
Aluminum atoms fit in the crystal maxtrix but have 3 electrons in the outer electron shell.

P - TYPE SEMICONDUCTOR

Diodes are P-N junctions

Semiconductor diodes are constructed by placing P- type semiconductor in contact with N-type semiconductor. In other words, for electrons to flow through the diode, the electrons must enter the N-type crystal and then move across the junction into the P-type where they complete the journey by jumping from hole to hole.

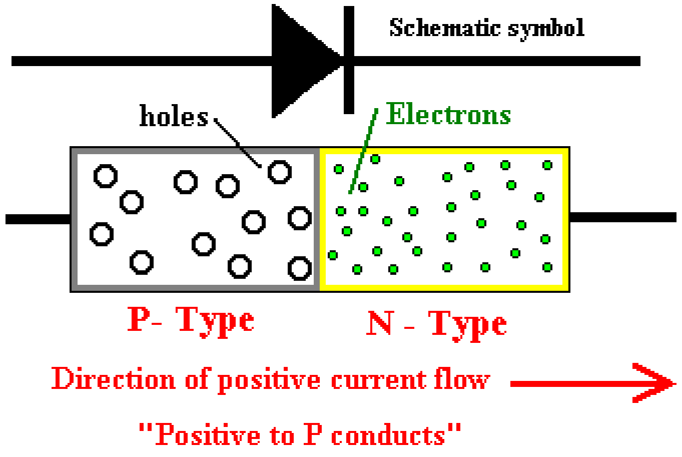
Schematic symbol
holes
Electrons
P- Type
N - Type
Direction of positive current flow
"Positive to P conducts"

P-N SEMICONDUCTOR DIODE

"positive to P conducts"

If we measure the resistance across a diode with an ohmmeter, the resistance is low when we place the positive pole of the meter on the P-type semiconductor and the negative pole on the N-type. OK. Now let's reverse the probes of the ohmmeter. We are placing the positive probe against the N-type semi-conductor and the negative probe against the P-type. Electrons flow off the metal probe and into the P-type semiconductor. No problem so far. On the other side of the diode the extra electrons from the N-type silicon are being attracted or "sucked" into the positive metal probe. Thus the conduction seems to start out all right, but it isn't long before the extra electrons in the N-type silicon along the P-N junction are depleted. All that remains in this region are depleted valence 5 atoms that are now acting like pure silicon. This whole region now acts like pure silicon and the conduction stops.

So why can't the electrons that are migrating through the P-semiconductor holes hop across the P-N barrier and move onto the valence 5 atoms? The reason is the same. The migrating electrons have filled in all the holes in the P-type and the crystal has also become pseudo-pure silicon that is an insulator. When thinking about PN diodes, remember, ***"positive to P conducts."***

Real diodes

ARCHER ELECTRONIC PARTS
SCHOTTKY BARRIER DIODES (2)
ARCHER
FULL WAVE BRIDGE

Commercial diodes come in all sizes. The types suitable for detectors in crystal sets are the little bitty guys on the lower left. The big black diodes on the right are rectifier diodes and diode "bridge" arrays for power supplies. The long black diode is rated at 6,000 volts. The two diodes that are built like machine bolts are high speed, high power units that might be used in a large switching power supply or an industrial RF generator.

Detection of AM radio signals with a diode

In amplitude modulation, (AM) the audio speech signal is impressed onto the radio signal by varying the AMPLITUDE of the radio signal. An AM transmitter literally increases and decreases the output power of the transmitter in time with the speech and music being broadcast. The drawing below shows an unmodulated radio signal of the sort used to send Morse code. The radio frequency sinewave remains the same amplitude throughout the time that the transmitter is keyed. Because the sinewave maintains its amplitude during the "dots" and "dashes," Morse code signals are known as ***continuous wave*** or "***CW***."

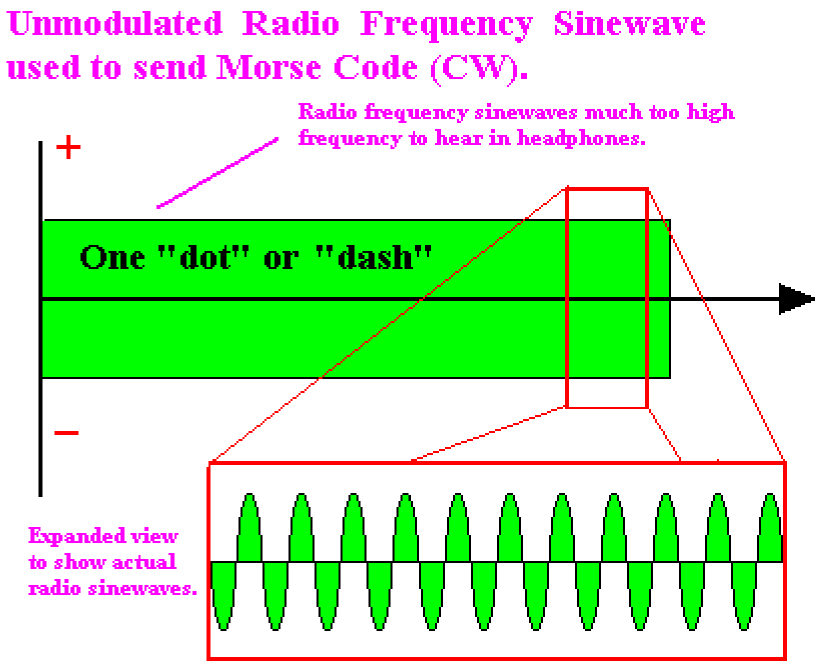
Unmodulated Radio Frequency Sinewave
used to send Morse Code (CW).
Radio frequency sinewaves much too high frequency to hear in headphones.
+
One "dot" or "dash"
–
Expanded view to show actual radio sinewaves.

As the name implies, an ***amplitude modulated (AM) radio broadcast, (550 KHz to 1.750 MHz)*** makes a continuously varying graph of the RF signal that looks like a psychiatrist's Rorschach. But of course the outline of the audio signal is actually made up of hundreds of thousands or millions of RF sinewave cycles.

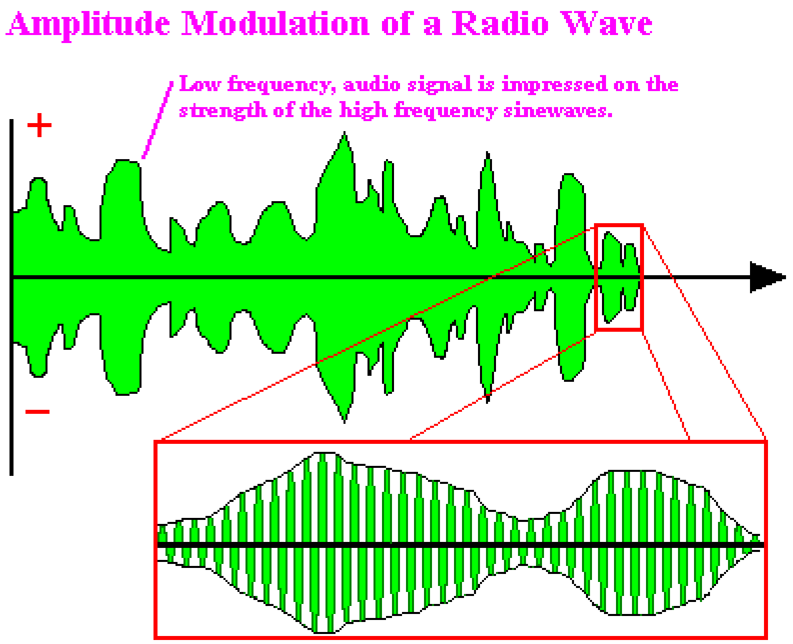
Amplitude Modulation of a Radio Wave
Low frequency, audio signal is impressed on the strength of the high frequency sinewaves.
+
–

The diode detector recovers the audio signal by "shaving off" one of the two polarities of the RF signal. Sinewave currents have both positive and negative polarity. Diodes only allow conduction in one direction. So, when a radio frequency sinewave current is passed through a diode, one of those polarities will not pass and will be eliminated. What remains is a series of narrow, direct current pulses, all with the same polarity.

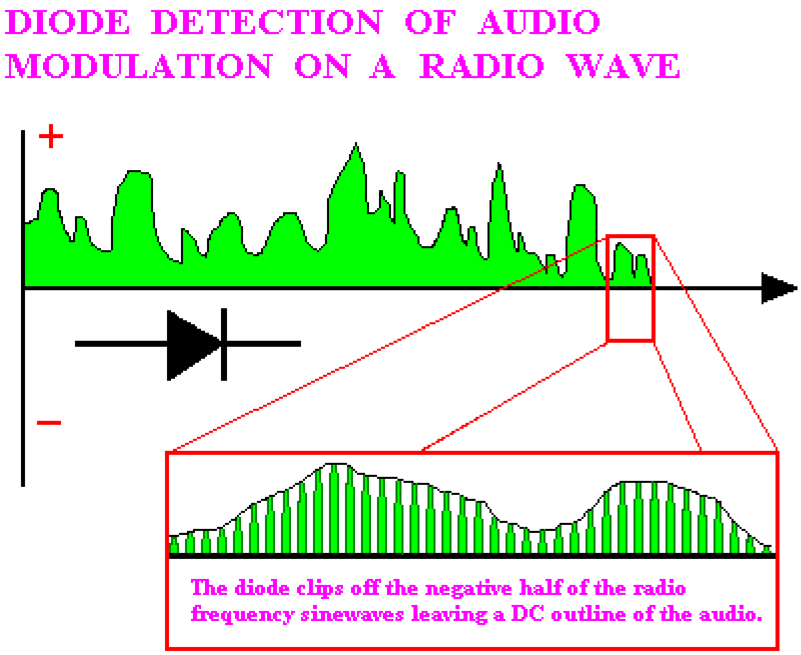
DIODE DETECTION OF AUDIO MODULATION ON A RADIO WAVE
+
−
The diode clips off the negative half of the radio frequency sinewaves leaving a DC outline of the audio.

This detection process, which is also called ***rectification***, produces a varying DC signal that may be passed through a headphone to convert it into sound. Physically, a modern diode is usually a tiny glass cylinder typically ¼ inch long with two wires extending from the ends. It's not much to look at. As will be described below, it's more fun to make a diode out of sulfide ore, or even out of razor blades or safety pins.

Headphones

After the diode has generated the varying DC current representing the audio signal, a device is needed to convert the current into sound. The classic way to do this is to use a magnetic headphone. As we shall describe below, a headphone is an electromagnet that attracts a thin, steel diaphragm and makes it vibrate in time with the speech and music.

A practical crystal set schematic

The crystal set can be extremely simple. A schematic is shown below:

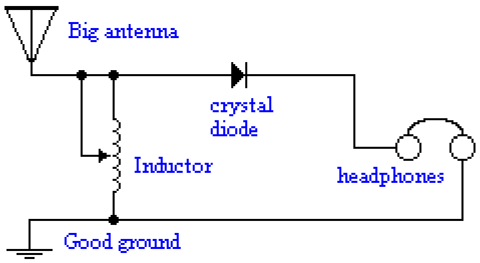
Big antenna
crystal diode
Inductor
headphones
Good ground

Crystal Set Parts List:

Big antenna – 50 feet of wire strung up in a tree will be ideal. Or, use the 40 meter dipole described above. For this application, use the entire dipole assembly as if it were a single piece of wire. Connect the center conductor and the braided outer shield of the coaxial cable together and fasten the resulting "wire" to the "big antenna" location above.

Good ground – A connection firmly clamped to a household copper water pipe would be ideal. Alternately, you may use a second length of wire strung up in another tree. The second wire should be far away from the first wire. I happen to have a 30 meter ham band dipole in my back yard. I used my 40 meter dipole in my front yard as the "antenna" and the 30 meter as a "ground." Or, as it would be known in this application, the 30 meter dipole became a ***"counterpoise."***

Inductor Wind about 20 turns of bare copper wire around a large diameter cardboard tube. Cardboard Quaker Oats boxes are the classic coil form for this purpose, but the cardboard cylinder from a toilet roll will work. In general, the larger the diameter coil, the better it works. I believe large coils work better because the coil is acting as a magnetic antenna, as well as a tuned LC circuit. In other words, a large diameter coil snags more magnetic field component from the radio wave. To tune the crystal set, you need to rig up a slider or shorting clip that allows you to short out some of the coil.

Capacitor Where's the variable capacitor? A capacitor consists of two pieces of metal separated by an insulator. If you wind a big coil of wire around a cardboard tube, then there is capacitance between one loop of wire and all the neighboring loops. "But, hold on! That can't be! They're shorted together!" you say. Yes, you're right. But if you look at an LC circuit as a whole, the inductor is a kind of "short circuit" across the whole capacitor and we know that works OK. The hard part about physics is that you have to learn to think abstractly. A lot of phenomena seem fuzzy and inconsistent. We are forced to "get a feel" for what works and what doesn't. The coil of wire is said to have ***intra-winding capacitance*** that acts the same as if it were a separate capacitor across the whole thing, honest.

Crystal diode It's great fun to build your own diode as described below. However, to get started, you may want to use an ordinary, small silicon diode such as a 1N4148 or a 1N914, which is available at Radio Shack.

Headphones. You may construct a working headphone from ordinary parts as described below. This will be fun and educational, but eventually you will need to buy a good pair. You may buy either old-fashioned high impedance (2000 ohms) headphones or modern low impedance (8 ohms) headphones. The modern ones are extremely efficient, comfortable to wear and have hi-fi sound. The high impedance headphones are historic and little more can be said in their favor.

****

Homebuilt diode detectors

My experimentation with crystal sets as an adult began one day when I was hiking near Jamestown, Colorado. I was scrambling up a yellow-colored abandoned mine dump. Mine tailings up there are mostly yellow, sulfated, powdered rock that consists of broken-down granite or gneiss. Suddenly right in front of my face were chunks of the shiny, black sulfide ore that was the reason for the mine. Without an assay, I don't know exactly what's in this ore, but it's a safe bet that it's a mixture of sulfides of silver, lead, and maybe zinc, a dash of arsenic, tin, and copper. There might even be a trace of gold telluride in those crystals. Galena, which is lead sulfide, is the stuff used in old-time crystal sets to make detector diodes. "Gee! I wonder if I can make a crystal set out of this ore?"

It seems to me I once saw a war movie in which a POW in a Nazi Stalag made a radio out of barbed wire, a razor blade, and silver paper from a chewing gum wrapper. Well, that's Hollywood, but maybe a receiver can be built without using parts specifically manufactured for radios. I happened to have a toy crystal set radio dating from about 1950 in my attic, so I hauled it down and checked it out. The "diode" consists of a tiny chunk of gray galena sticking out of a little puddle of solidified solder. The positive pole of the diode is a metal "cat whisker," a piece of thin copper wire poked against the crystal.

My crystal set project was done in two sessions and is written as two separate articles. Before you build the set pictured below, I suggest you read about the second crystal set as described in another article following this one. The second crystal set encountered some abstract mysteries, so while it worked better, it is harder to understand.

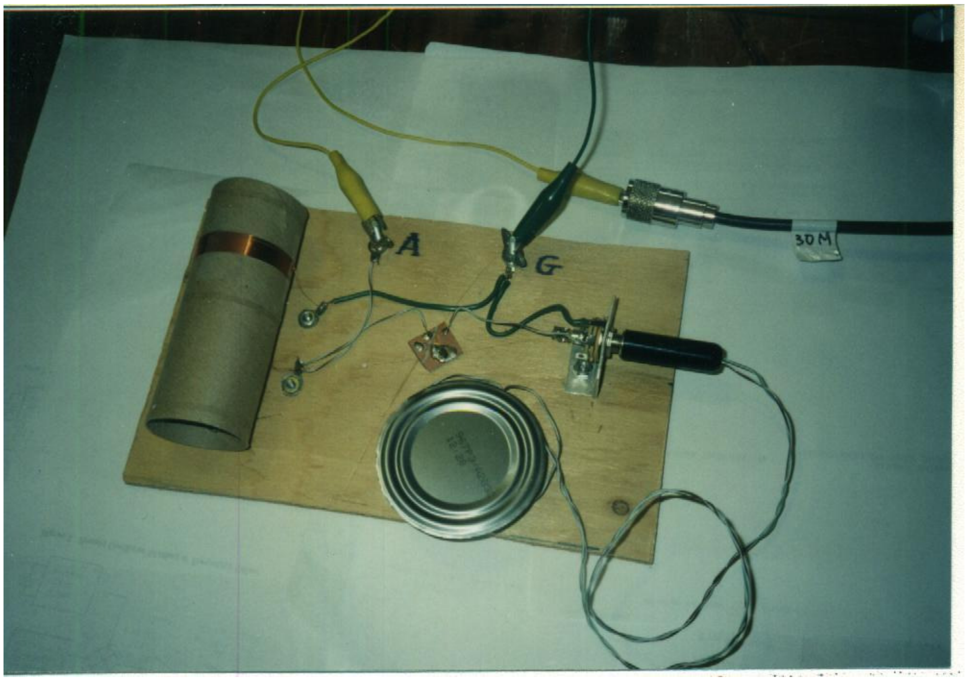

The above picture shows my crystal set with all three homemade components. An LC resonant circuit is vital to select the AM radio band (or other band). An actual tuning capacitor isn't really needed at AM radio frequencies. A big coil, at very least two inches in diameter with 20 to 60 turns wound on a cardboard tube, has enough inter-winding capacitance to resonate in the AM band. From a physics point of view, the circuit in the "practical" schematic is functionally the same as the circuit that contains the variable capacitor across the coil.

Without a variable capacitor, you will have no way to tune in particular stations. A tap on the winding can be added for peaking a station. A "tap" is just a way to short out part of the inductor. Using this method you can crudely, very crudely, select the loudest stations at the top or bottom of the AM broadcast band. Tuning a crystal set is (usually) sloppy. However, the article at the end of this section demonstrates how to accomplish sharp tuning.

The crystal diode rectifies the radio frequency voltage ringing on the LC circuit and the headphones turn it into sound. Some crystal sets also have an audio signal filter or “integrator” capacitor. This capacitor, about 0.01 microfarad, is placed across the headphones. However, in my crystal set, it didn’t do anything useful, so I left it out. Leaving out parts is a good way to find out what they do.

Try leaving out the LC circuit and just connecting the diode and headphone to the antenna and ground. At my house all I could hear was faint static that sounded like power line noise. That implies that power lines generate the biggest AM signals over the entire radio spectrum. In any case, without the LC circuit, I heard no radio stations.

The Jamestown crystal diode

To make my crystal detector out of sulfide ore, I melted a puddle of solder about 3/8 inch wide on a piece of PC board. Then I used tweezers to press a bit of ore into the puddle so that, when it hardened, half of the crystal was exposed. Next I soldered a tiny loop of copper wire onto a pad on the board for a cat whisker. My first diodes used loops of fine copper wire as “cat whiskers.” The copper wasn’t springy enough to poke into the crystal with enough force for reliable performance. It worked well, but with a little vibration, it quickly died.

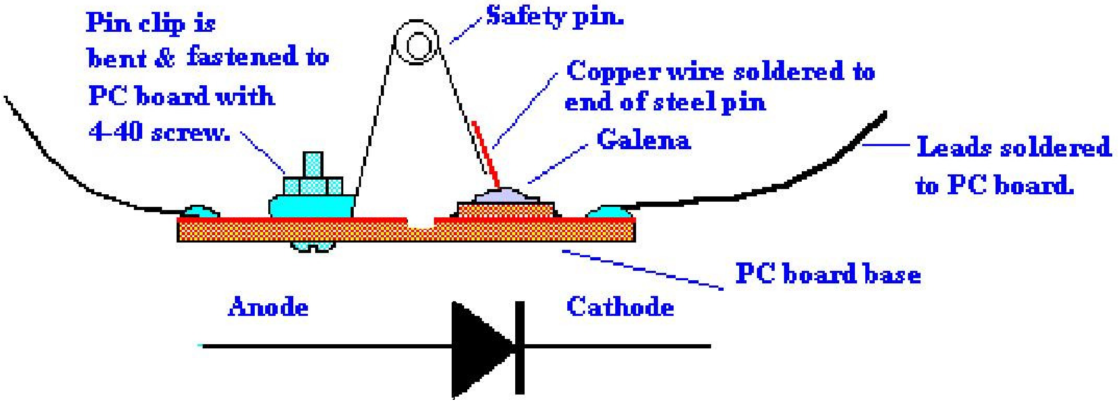
Pin clip is bent & fastened to PC board with 4-40 screw.
Safety pin.
Copper wire soldered to end of steel pin
Galena
Leads soldered to PC board.
PC board base
Anode
Cathode

A safety pin pushes the copper whisker against the galena.

In my next diode I made a copper ring cut from the end of 1/4 inch copper tubing that served as a deep “tub” of molten solder into which I could push the galena. Bob, NØRN, told me that when he was a kid, he used safety pins as cat’s whiskers. Sure enough, the spring-loaded safety pin produced plenty of force and solved the mechanical problem. ***Don’t forget to cut the notch or gap in the circuit board between the galena side and safety pin side*** – they are two separate circuits!

Carbon steel is a semiconductor

When I first put my diode with the safety pin cat whisker into a crystal set, it was stone silent - nothing. No matter how I moved the sharp steel pin around on the galena, the headphones were dead. The pin happened to strike the solder at the edge of the galena and the crystal set came to life with music from KBCU, our loudest local AM station. At first I was mystified. The steel pin rectified well against either solder or copper. The signal was perhaps only 2/3 as loud has it had been with the copper-to-galena diode, but it was much easier to adjust.

It turns out that steel is a carbon-iron semiconductor compound called cementite. The surface of hardened steel is a crystal, perhaps not radically different from the galena (lead-sulfide) crystal. Carbon has a valence of four, just like silicon or germanium. So, if you want to build a crystal set for your kids, ***you don't have to mine galena. Just use a safety pin pressing against copper or solder.***

Another surprise for me was that copper-to-copper, solder-to-solder, or solder-to-copper junctions also rectify and produce weak signals. The contact between the two metal surfaces must be extremely light - just barely touching. This phenomenon is poor for making crystal sets, but it's a warning about bad contacts in electronic equipment. ***Cold solder joints and loose screws can fill your circuit with accidental diodes.***

Copper cat whiskers work best

As shown in the diode construction diagram, I used ***acid core*** solder to attach a piece of copper wire onto the end of the pin. Now the contact point of my diode is between the semiconductor galena and copper rather than semiconductor steel-to-semiconductor sulfide ore. I connected my crystal set to the center conductor of my 40-meter dipole coax and my station ground. I scratched the copper whisker around on the sulfide crystal and suddenly I was again hearing our local station. Using commercial 8-ohm headphones, it was almost painfully loud. Too bad KBCU is mostly rap music.

Where is the P-N junction in these crude diodes?

If you are a thoughtful person, you must be asking, "Where is the P-N junction with the impurities imbedded in the pure semiconductor and all that?" It turns out that you can make crude diodes by throwing together pretty inferior materials. For example, pure galena crystal consists of lead and sulfur that have valences of 2 and 6, that sort of average to 4. But there are also all those other atoms in typical galena ore. These impurities, like silver or copper, have valences like plus 1, while other transition metals like tin have valences of 2 or 4. Let's just assume that because of sulfide (valence 6) and arsenic (valence 5) salts, my ore is a N-type semiconductor. So where is the P-type semiconductor? It turns out that if you press a metal against N-type semiconductor, metal ions will migrate a few microns into the N-type crystal and make a tiny P-type region surrounding the contact point of the "cat whisker."

As you might expect, the disadvantage of such crude diodes is that P-N junction is quite fragile. That is, the check valve function only works with very low voltages and extremely small currents. The P-N junction is easily destroyed if you put large reverse voltage across it or try to pass large currents through it.

I tried measuring the volt/ ampere characteristic of a steel/ copper diode. It was no surprise that it looked like a short circuit on an ohm-meter. I studied it carefully with a high impedance multimeter and 10 megohm series resistors, but it still looked like a short circuit. I believe I just learned that ***the world of RF detection is quite subtle***. At least the legend about POWs in WWII making radios out of barbed wire and razor blades is starting to make sense. The razor blade must have been the semiconductor.

Commercial diodes

A perfect diode would have zero resistance in the forward direction and infinite resistance in the reverse direction. As you can see in the graph below, no real diode is perfect and all diode types have different forward voltage characteristics, depending on the semiconductors used to make them. The reverse characteristics of commercial diodes are all excellent and similar, but each type of semiconductor diode needs a certain level of forward voltage before it will begin to conduct. In the graph below relative current is plotted on the vertical axis. Commercial diodes handle hundreds of milliamperes or even hundreds of amperes. As you'll see below, homemade diodes can barely handle microamperes of current before they fail.

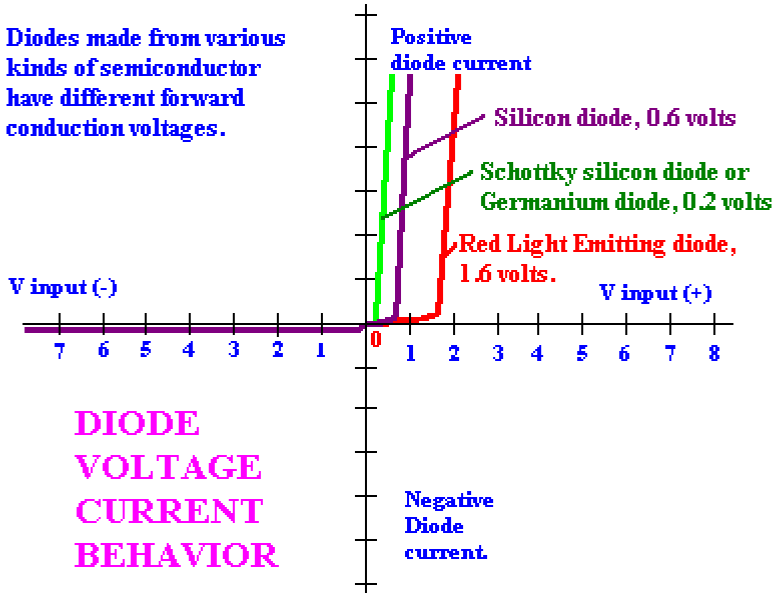

Volt – ampere characteristics of homemade diodes

Once I had my cat whisker adjusted, the Jamestown diode was just as loud as the 1950 crystal set diode. I tried substituting a modern, hot-carrier Schottky diode for the crystal. In theory, a Schottky should be comparable to a point-contact diode. Schottky diodes are commonly used as detectors in relatively modern equipment. I was surprised to discover that the modern Schottky produced as big a signal as the crystal diodes, but no better. If these diodes perform the same, how do their volt/ ampere characteristics compare?

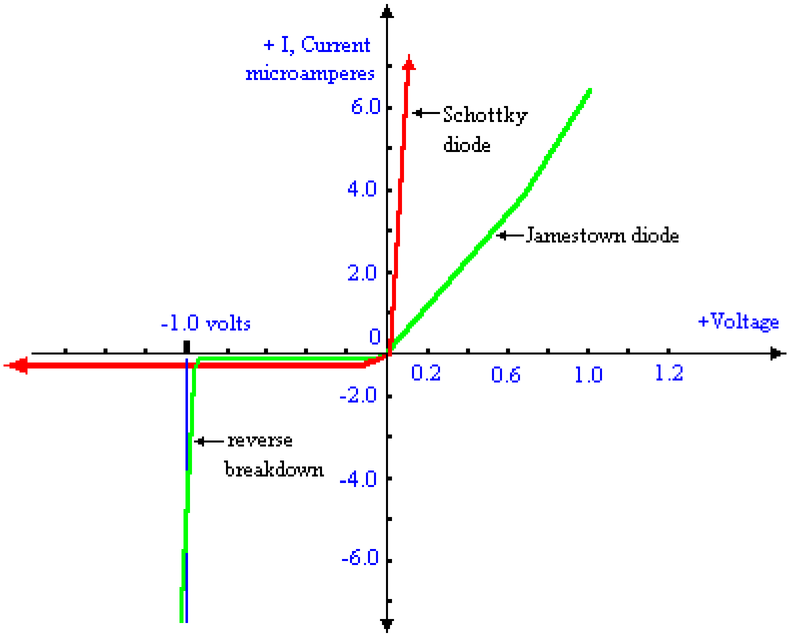

Volt / Ampere curves for the Jamestown diode and a commercial Schottky diode

At big voltages and currents, (milliamperes), the Schottky diode acted as you would expect: It passed big currents (milliamps) above 0.2 volts forward voltage and leaked only 100 nanoamperes with reverse voltage. Notice that if the Schottky were “perfect," the red line would be directly on top of the vertical axis upward, and directly on top of the horizontal axis to the left. However at very low currents, microamperes, the commercial Schottky was pretty nearly perfect with a transition right at zero volts.

In contrast, with big currents the Jamestown diode behaved like a resistor in both directions. At first it wasn’t obvious to me that it could rectify anything. The curves above show the behavior of both diodes at tiny current levels, microamperes, using a 1 megohm (one million ohms) load. For tiny currents, the Schottky and Jamestown diodes were both strongly non-linear at the zero current, zero voltage point. The surprise for me was that, for reverse voltages, the Jamestown diode broke down abruptly at minus one volt. It’s no wonder it conducted so well in both directions with a “low” resistance 10K ohm load. This abrupt, reverse breakdown is called ***avalanche breakdown.*** When it occurs with big currents it usually destroys the diode. As we’ll see in chapter 8, some diodes called ***Zener diodes*** are designed to breakdown at specific voltages without being destroyed.

A homemade headphone

Building my own headphone was the hardest part of my crystal set. A headphone uses a high impedance coil of wire to make a magnetic field proportional to the audio signal. The changing field pushes and pulls against a thin steel diaphragm to produce sound vibrations. Even if you decide to build one of these headphones, I strongly suggest you buy a good pair of headphones so you will have them for your ham rig. Also, with commercial headphones the speech and music will be perfectly clear and loud, even with a safety pin diode.

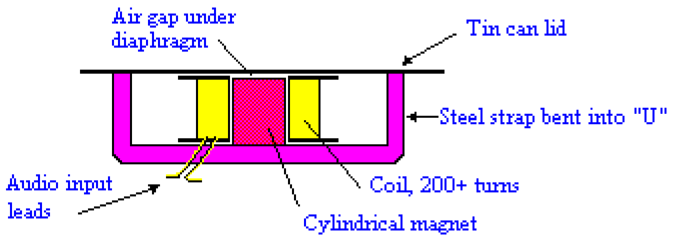

The Caribou headphone

A cross section of my homemade headphone is diagramed above. Its construction is basically the same as old-fashioned high impedance headphones. Yes, the sound is tinny. What did you expect from a headphone diaphragm made from a tin can lid? The coil is hundreds of turns of #36 wire wound on a paper coil form. Inside the coil is a cylindrical magnet I took out of an old loudspeaker. A piece of steel strap conducts the magnetic flux around to the edges of the lid. The magnetic force holds the lid on. By completing the magnetic circuit, the magnetic force is concentrated in the gap between the tin-plated steel and the magnet.

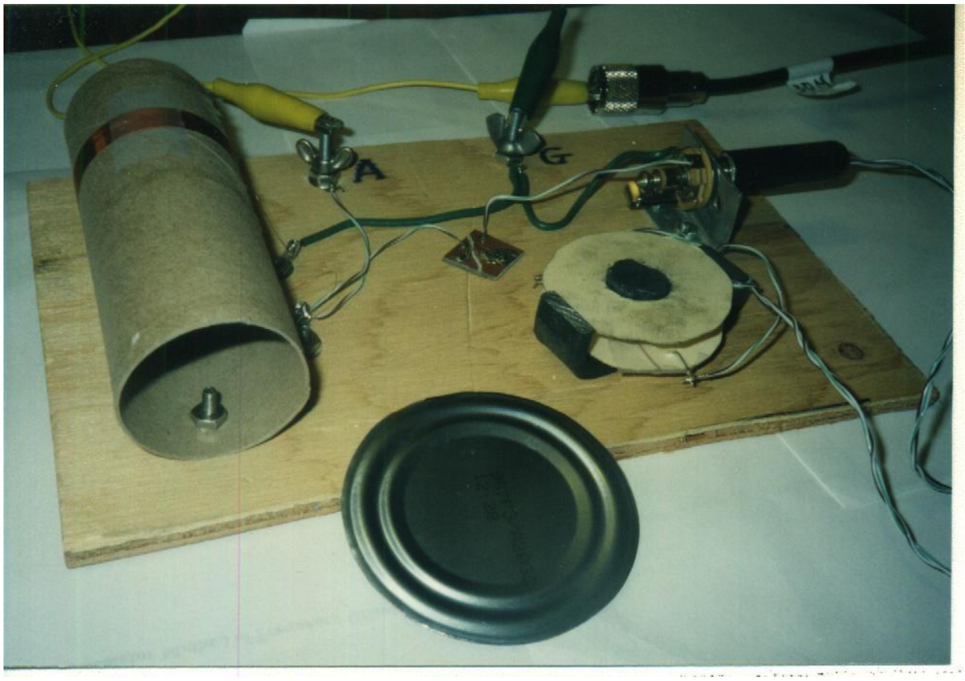

Crystal set showing homemade headphone. The tin can lid diaphragm has been removed.

I started out using a small magnet from an old loudspeaker, but that felt like cheating. Would Heinrich Hertz have been able to use a loudspeaker magnet? Anyway, it seemed to me that the magnet wasn't essential. Why couldn't the coil just magnetize ordinary iron? I tried substituting a big steel nut of the same size. Sure enough, it worked, but the sound was too faint to be audible in the crystal set. However, when I plugged the homemade headphone with the steel nut into my shortwave radio, it was surprisingly loud. Not Hi-fi, mind you, but loud. No, for a sensitive headphone a magnet is needed to overcome the hysteresis.

Hysteresis

What's hysteresis, you ask? Whenever iron is magnetized with a DC coil, the tiny "magnetic domains" in the iron line up to make a big magnetic field. But when the DC current is shut off, some of the magnetic domains remain aligned and leave a residual field. To magnetize the iron in the opposite direction, a current of the opposite polarity must first overcome the residual field. This means that hysteresis interferes with the sensitivity to weak signals. Since crystal sets are powered by the radio waves themselves, sensitivity is vital. A magnet is needed to overcome the hysteresis and "bias" the magnetic field so that it always operates in one direction. I could magnetize iron with a DC powered coil, but then to be a purist, I would need to build a homemade battery. And I would need to smelt and extrude my own copper wires. (Forget I said that.)

I had a sudden inspiration. I dug around in my rock collection and found a piece of magnetite ore from a mine dump at Caribou, Colorado. Magnetite is a specific iron oxide, $Fe_2O_{4,}$, that retains a magnetic field. I machined the magnetite with my bench grinder into a small cylindrical magnet. Unfortunately, the grinding and heat ruined the magnetism. However, fixing it to a big, heavy permanent magnet, I was able to put my magnetite in a strong magnetic field. Then I banged on it firmly against my anvil. Believe it or not, that abuse restored the magnetic field. Behold! - The completed Boulder County rock and toilet roll radio!

How does it perform? Well, frankly the homemade headphone is pathetic and needs lots of R&D. The sound is plenty loud when plugged into a real radio, but installed in the crystal set, I can just barely hear the rap music. Perhaps if I had a thinner steel diaphragm, a headphone for each ear, optimum impedance matching, better craftsmanship, and other refinements, it might approach a commercial headphone. In other words, for serious listening, buy a decent headphone! And, after your done playing with homemade crystals, I suggest you buy some silicon diodes. Type 1N914 or 1N4148 diodes work great in this radio. ***They don't work better than the diode made from sulfide ore***, but they are smaller, more rugged, and don't need to be carefully adjusted.

This crystal set article was first written in 2002. I revisited crystal sets in 2006 and discovered some surprises.

****

CRYSTAL SETS REVISITED

Or, electronics makes us humble

A friend of mine, Jack Ciaccia, WM0G, suggested to his grandson, Rutger Koch, that a broadcast band crystal set like the one described above would be a good science project. Rutger did a terrific job of duplicating the set and the homemade diode. The only obvious difference I could see was that he had used the 5-inch cardboard oatmeal box coil form that I had recommended. But when Rutger put it together, it was totally inert - no sound.

Rutger consulted his granddad who also couldn't find anything wrong. Jack brought the crystal set over to my house and we two old hams scratched our heads and still couldn't find the problem. After five minutes of swapping parts, I dug out my prototype crystal set from the closet and there was the answer: If you go back to the drawing of the homemade diode made out of the piece of PC board and a safety pin, you will see that there is a notch carved in the PC board between the anode and cathode. Rutger had forgotten to cut the copper sheet between the anode and cathode so they were shorted together. We cut the copper sheet and it worked perfectly.

Our eyes only see what they expect to see! Don't take the obvious stuff for granted. By the way, Rutger's science teacher refused to believe rocks could serve as diodes until he put on the headphones. This ancient technology is all new information for whippersnappers.

Does a bigger diameter coil work better?

When I first built my AM broadcast band crystal set described earlier, a toilet paper roll was the largest diameter cardboard form I could find around the house. Although I hadn't noticed any obvious superiority in Rutger's crystal set over mine, I decided to wind a big coil like his and see if my magnetic antenna theory had any validity. That is, the big coil might pick up the magnetic wave components and produce a bigger signal. Perhaps it would be directional like an AM band ferrite core antenna. Since my crystal set had previously been converted to 10 meters (see the next article, below), I just took off the loop antenna and left the 140 pF variable capacitor on the set. Then I installed the oatmeal box coil with about 20 turns on it. Since I already knew that ordinary silicon diodes have the same sensitivity as the rock diodes, I used a 1N4148 diode and my large, low impedance headphone.

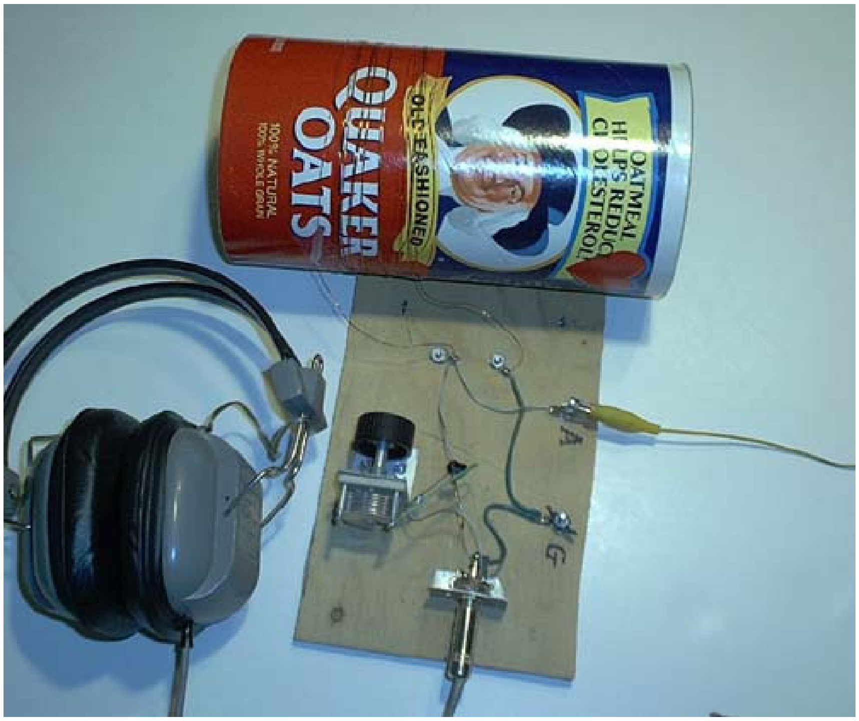

Oatmeal box crystal set - operating without a ground or counterpoise

I hooked up the 40-meter dipole antenna, "A" above, and used the 30-meter dipole as a counterpoise connected to the ground post, "G" above, just as I had before. This was the same way Jack and I had hooked up his grandson's set. The silence was deafening. Good grief! Now what?

After some flailing about, I accidentally disconnected the counterpoise wire as shown above, just leaving the antenna wire connected. Voila! It not only worked well, the tuning was precise enough to tune in my local NPR station at 1490 KHz with excellent selectivity and signal strength. By "excellent" I mean, it was not only loud and clear, I could only hear the one station. Precise tuning with a crystal set! How is this possible? When I padded the relatively small 140 pF variable capacitor with 100 pF fixed capacitors, I was able to tune down the AM broadcast band and uniquely select other stations. The farther down the band I went, the less precise the tuning. I believe the loss of precision at the low end of the band is related to the relative signal strength of the stations. Also, I obviously needed more turns on my coil. If I were starting over, I would use a larger variable capacitor, e.g., 365 pF.

How does this work?

The selectively tuned crystal set seems to be an example of a circuit that doesn't seem to have a complete circuit loop. Without a ground or a counterpoise, where does the current go? I assume that there is some form of capacitive coupling that completes the loop, but I must say this coupling isn't obvious. As you can see, RF sometimes makes current loops that are hard to identify.

Using the backyard antenna as a counterpoise didn't work with the big coil, but I found that clipping the ground terminal, "G" to my station ground did improve signal strength. However, it ruined the precision of the tuning. Hooking the station ground and the counterpoise antenna together still worked, but was weaker and produced poor tuning. Using the backyard 30-meter antenna as the sole antenna, worked just as well as the 40-meter antenna.

I probed with my oscilloscope at various places on the circuit to get some answers. As soon as I attached the oscilloscope probe ground wire, it behaved just like the grounded hook up, just as you would expect. The signal strengths across the coil were on the order of 0.40 volts peak. The power delivered to the 8-ohm headphones was about 20 milliwatts peak. The most surprising observation was that, when I just left the scope probe and its ground wire lying six inches away from the crystal set, I could see the modulation on the screen appear and disappear as I selected specific stations. The voltage signal was only slightly less than when it was directly connected. Apparently there is a large electric field surrounding the crystal set.

As for the magnetic antenna effect, when tuned into a specific station, there was no hint of directionality as I rotated the coil from north to south. In conclusion, the advantage of the big coil seems to be a dramatically higher Q (higher quality resonance). But to observe this advantage, the headphone must be powered exclusively by the signal ringing across the coil. If the current is passing from antenna to ground, the L-C circuit has much less effect on the tuning. And finally, (apparently) when the ends of a high Q coil are connected to two equal antennas, the signals cancel each other.

Perhaps you have your own theories about these phenomena. Feel free to pontificate. Obviously we old hams don't know everything, even about the simple stuff! Keep your brain engaged and question authority.

****

RECREATING HERTZ'S RADIO EQUIPMENT

Most of what I've read about the history of radio was written by non-engineers. They describe the revolutionary apparatus invented by our heroes using old-time radio terms like “earth resonances,” “aether,” and “coherers.” They tell us how far it transmitted, but they give us only the faintest clues about how the gizmo actually worked. Was it a spark gap transmitter? A high-speed alternator? What the heck was a "Tesla oscillator" anyway?

In 1884 James Maxwell published four equations that quantified and connected magnetism with electric phenomena. These equations also predicted the existence of radio waves. The changing magnetic and electric fields related to each other with sine functions. So, once physicists had the equations to stare at, it wasn't too huge a leap to conclude that sinewave-shaped electric and magnetic fields would generate each other in an oscillation and radio energy would propagate through space.

In 1889 Heinrich Hertz, a physics professor at the University of Bonn, Germany, was the first to demonstrate radio waves in the laboratory. Of course he might have done this in 1884, or 1887, depending on which website you visit. Ah, the glorious information age!

That's fascinating, but **HOW** did he demonstrate radio waves? Using 1880 technology, that could not have been easy. How did he know he was detecting waves and not just magnetic coupling from one coil to another? Or if his antenna was capacitive, how did he know he wasn't observing capacitive coupling? ***If I were skeptical about the existence of radio waves, but I understood the full implications of Maxwell's equations, I would be convinced if I could see communication across a distance greater than one wavelength.*** A minimum of one wavelength means that "the alleged electromagnetic wave" would change from magnetic to electric field energy then back again at least once. ***Of course, I would also want to see evidence of standing waves and a way to measure frequency.***

Demonstrating Hertzian waves

Suppose you were living in the year 1884 and Maxwell had just predicted the existence of radio waves. Using components available in your time, how would you generate Hertzian waves and get those waves named after you instead of Hertz? If you are able to generate radio waves, how could you prove to a skeptic that you had actually done it? Hertz managed this feat and apparently his demonstration was convincing. Otherwise the unit of measurement for frequency would not be the ***Hertz***. One Hertz (Hz) equals one cycle (one complete oscillation) per second.

When I first had the idea of pretending to be Dr. Hertz, I was not able to find a description of his apparatus. That was a good thing because it forced me to invent my own method to demonstrate Hertzian waves. If you already know enough about electricity to be able to handle the challenge, then get out your 1880 hardware and build a transmitter and receiver that will transmit at least one wavelength. If you don't know how to begin, keep reading.

Transmitting and receiving as simply as possible

The only detail of Hertz's apparatus I found described was that he detected his waves by means of a loop of wire. The wire had such a large current and voltage induced into it by the radio waves that a visible spark jumped across a gap in the circle of wire. Wow! It must have been a big radio signal that would induce that much energy into a loop of wire. And if the signal was that big, how far away from the transmitter had the loop been? I suspect the signal strength had to be high and the loop had to be very close, like a foot or two away.

If were a skeptic who already knew about Faraday's transformers, this demonstration would not convince me. How would I know that radio waves had propagated across the one foot distance to the loop? Maybe all I was seeing was a big magnetic field that reached from one coil to another.

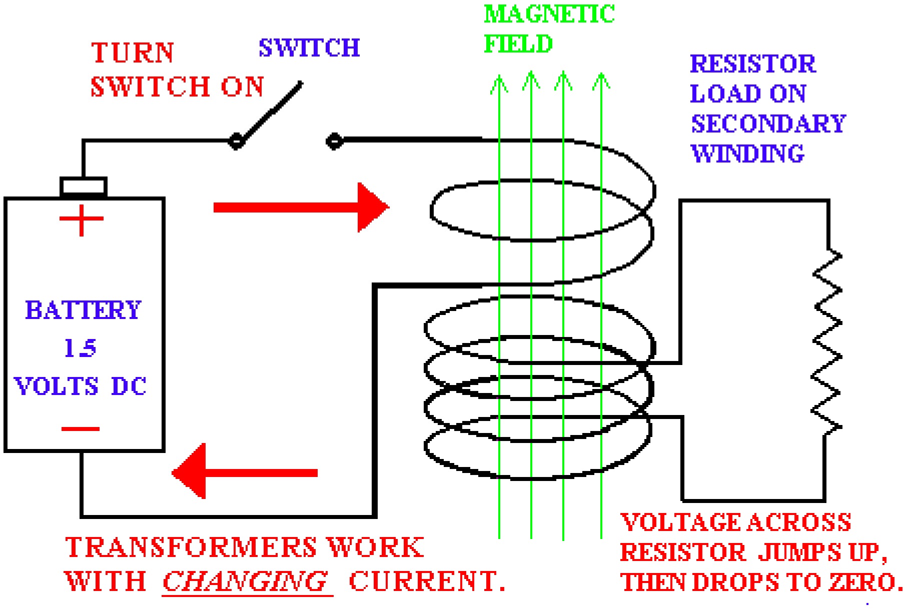
TURN SWITCH ON
SWITCH
MAGNETIC FIELD
RESISTOR LOAD ON SECONDARY WINDING
+
BATTERY 1.5 VOLTS DC
—
TRANSFORMERS WORK WITH *CHANGING* CURRENT.
VOLTAGE ACROSS RESISTOR JUMPS UP, THEN DROPS TO ZERO.

Maybe Hertz's demonstrator was just a transformer?

A transformer is a *magnetic* device that works by transmitting a changing magnetic field from one coil to another. Coils (inductors), convert the energy of an electric current moving through a wire into magnetic field energy that hovers in a cloud like region around the coil. If a second coil is close enough to the first coil to be inside the magnetic "cloud," then if the magnetic field is changed, then an electric force, a voltage, will be generated in the second coil. Transformers and the above diagram were discussed in chapter 2.

Inductors store magnetic energy in the space around them so long as current is passing through the coil. The energy will remain in space so long as the current keeps flowing in the same direction through the coil. But when the current stops flowing, the magnetic energy becomes "stranded" in space. The magnetic energy then returns to the coil and induces a voltage in that coil in a direction that would force the current to continue flowing. That is, the induced voltage will try to keep the status of the magnetic field and coil the same. If the current is gone for good, then the magnetic field will collapse completely and the energy will dissipate into the coil and any circuit connected to it. But if the first coil is open circuit and even high voltages cannot restore current flow, then the field will collapse into the second coil. If the induced voltage is able, it will cause current to flow in the second coil to maintain the field. Or, as in the case of Hertz's loop detector, the induced voltage caused a big spark to jump across a gap where the resistor is located in the illustration above.

But I thought transformers were always made from iron

You may be thinking that transformers are not air-filled flimsy coils but rather large iron things like the big steel cylinder on the power pole in the alley behind your house. Yes, those big iron things are transformers but they are designed for low frequency power lines. An AM radio frequency is on the order of a million Hz frequency, while the power company supplies current at

60 Hz. Therefore, instead of having air between the two coils, power transformers have iron. The magnetic field from the coil magnetizes the iron temporarily. Recruiting iron and turning it into a magnet increases the magnetic field a thousand times or more. With a huge magnetic field stored in the iron, the power transformer can transmit big amounts of energy with only 60 direction changes per second. A similar transformer without an iron core could transmit the same amount of energy, but would have to repeat the magnetic field cycle perhaps a thousand times more often to transfer the same amount of total energy.

Why do you suppose the power company doesn't use 1million Hertz and do away with all that iron? After all, at one million Hertz the RF voltage would still be a burn hazard, but it could not electrocute anyone and would be considerably safer. Unfortunately, at one million Hz the power lines would act like antennas and radiate the energy into the sky instead of delivering it to your house. This is the same reason hams object to using the power lines to distribute high-speed Internet connections (BPL). BPL broadcasts noise all over the shortwave bands and makes them nearly unusable.

How far should it transmit to demonstrate the existence of Hertzian waves?

To be sure that the waves are Hertzian and not just magnetic fields, I would be impressed by the demonstration if the detector (the receiver) were more than one wavelength away. A wavelength is the distance that a radio wave travels during the time it cycles from magnetic field, to electric field, and back to magnetic field.

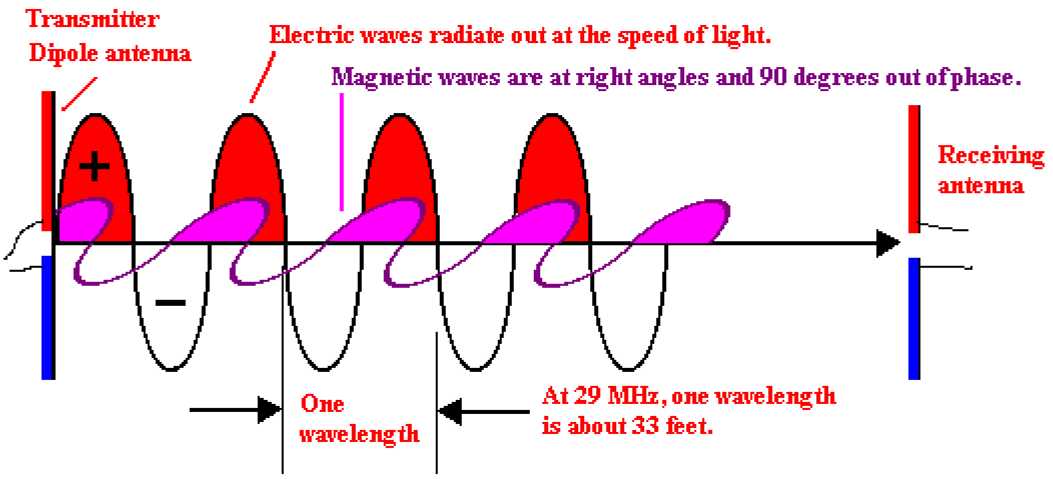
Transmitter
Dipole antenna
Electric waves radiate out at the speed of light.
Magnetic waves are at right angles and 90 degrees out of phase.
+
−
Receiving antenna
One wavelength
At 29 MHz, one wavelength is about 33 feet.

The speed of light is 186,000 miles per hour, or 300,000,000 meters per second. (A meter is about 39 inches) A wavelength is the distance a wave travels while going through one cycle of magnetic to electric energy conversion. The wavelength of the forty-meter amateur radio band (7 MHz) is obviously 40 meters. It turns out that typical, (total) antenna lengths for radio transmitters are either one half wavelenth or one quarter wavelength. On 40 meters, a typical vertical pole antenna is one quarter wavelength or 10 meters (33 feet) tall. The 10 meter ham band extends from 28.0 MHz to 29.7 MHz.

What exactly is the frequency in Hertz of the 10 meter ham band? To convert wavelength to frequency, divide meters per second of light speed by the wavelength:

$$f = c / \lambda$$

Where f represents frequency, c represents the speed of light and λ represents wavelength

Speed of light/ wavelength = Frequency in Hz.

300,000,000 meters/ second / Ten meters = 30 Million Hz (30 MHz) frequency

Remember that the AM radio band extends from 550,000 Hz to 1.7 MHz. The old analog channel 2 television started at 54 MHz . So the 10 meter ham band is roughly halfway between AM Radio and TV.

Getting back to the Hertzian demonstration, if I wish to transmit one wavelength, on 40 meters, my loop would have to be 132 feet away from my transmitter. Frankly, I don't think Hertz's loop detector will work at that range. And if it did, I would be arrested for using a transmitter that powerful. As a rule of thumb, the American FCC will not object to experiments like this if the radio waves don't go past 50 feet at easily detected signal strengths. On the other hand, maybe I could use a higher ham band like 10 meters. Now I only have to go 33 feet. If I go up to UHF frequencies, the wavelength could a foot or less, but those frequencies would be hard to generate and harder to measure with 1884 technology. I have since been told that Hertz actually used 4 meters wavelength for his demonstrations.

Designing the 10 meter transmitter

For the sake of symmetry I used two identical LC circuits for my transmitter and receiver. To imitate what little I knew about Hertz's apparatus, I used one loop of wire about a foot in diameter. I knew from experience that seeing sparks on the receiver side was hopeless, so I also made the receiver an LC circuit that I knew would trap an oscillation from the transmitter.

At this point, you may want to depart from the year 1880. It all depends on how pure you wish to be playing the historical game. For a capacitor I used a modern 140 pF variable capacitor. That way, I could adjust the capacitance and tune the oscillation to a particular frequency. If I wanted to be a purist, it would not be hard to make a homemade capacitor made from sheets of metal with paper for insulation between the plates. Personally, I was confident that such a capacitor would work. I just didn't want to spend hours to make one.

My first problem was how to start the oscillation in the transmitter LC. In theory, by shorting a battery across the loop, it will charge the loop with a big current limited only by the internal resistance of the battery. Then when the battery is removed, the coil's magnetic field will discharge forcing a voltage to appear across the capacitor. The loop will then be shorting out the capacitor and the oscillation will begin.

Like the variable capacitor, this project will go faster if you use modern tools to make sure your components are working. For example, to see if my transmitter was really transmitting, I used a ham band receiver tuned to ten meters. Sure enough, when I clicked the battery on the capacitor terminals, I could hear a click in the receiver loudspeaker. And when I tuned the capacitor, I could get the sound to reach a sharp maximum volume at a specific setting of the capacitor. Of course, if this were 1880, I would have to do everything by guess, trial, and error. Those old guys were darn sharp.

Designing the receiver

The next problem was how to detect when the receiver loop was oscillating due to waves from the transmitter. What to use for a detector? The 1880 solution would be to use a detector called a ***coherer.*** Coherer detectors were developed for the wire telegraph. By the time a signal had propagated miles down a telegraph wire, the signal was often too weak to close a mechanical relay. Coherers were used to "amplify" a weak Morse code signal. They were described in chapter 1, page 10. Unfortunately, coherers are a low frequency device, under 20 Hz. They are suitable for detecting a weak DC Morse code signal on a cross-country telegraph wire, but will not respond to the modulation in the audible range. I doubted that they would be useful for tiny radio frequency signals. Being lazy, I didn't build one to find out. Besides, my crystal detector made out of local rocks certainly fit the 1880 criterion. I decided to build a 10 meter crystal set.

Are headphones 1880 technology? Yes, just barely. Alexander Bell built his first telephone in 1879. He used a headphone designed like the homemade device described earlier. Actually, for my 10-meter receiver I used old commercial high impedance headphones instead of the homemade earphone. I couldn't afford to waste any sensitivity.

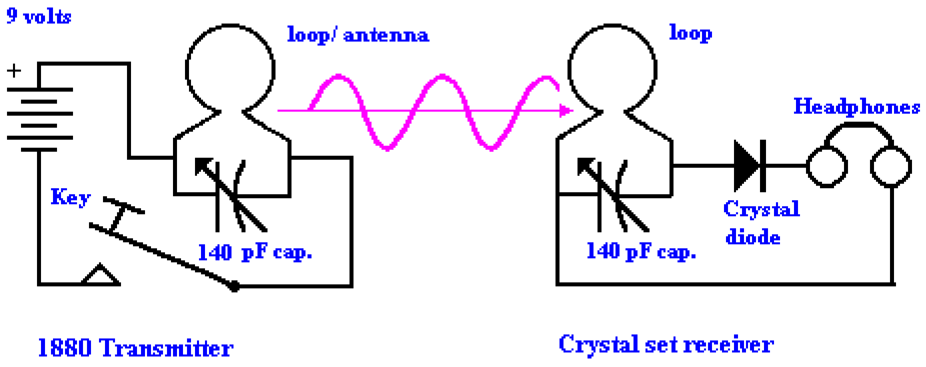

The simplest 10-meter communicator

I began with identical loop antenna/ inductors about 1 foot diameter for both receiver and transmitter. Instead of Hertz's spark gap, I put 140 pF variable capacitors across both loops to give me tunable LC circuits.

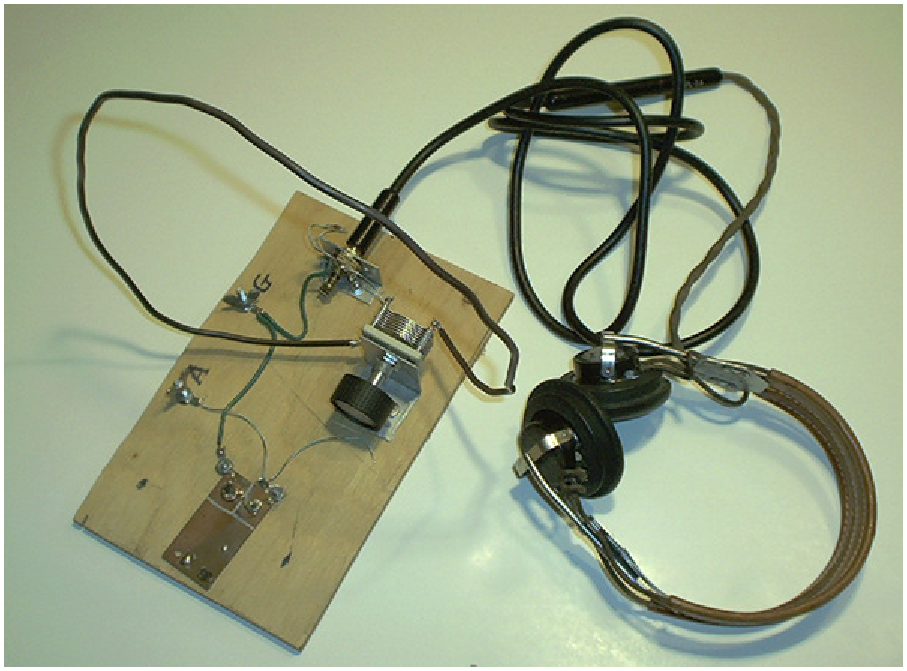

10-meter crystal set receiver

To charge the LC circuit, I used a crude ***spark gap***. I just touched battery terminals across the LC circuit while listening to the crystal set. I positioned the crystal set a foot away from the transmitter loop. It wouldn't be "radio" communication, but it would at least tell me if I was on the right track. I made sparks on the transmitter loop while tuning the capacitor. When tuned to just the right spot, I could suddenly hear obvious clicking in the headphones. I was surprised how sharp the tuning had to be. The big loop had relatively low inductance, so the capacitor had a tuning range of over 30 MHz. Tuning was probably not sharp by modern standards, but the adjustment was critical. In any case I achieved a range of 12 inches from the transmitting loop. Progress! Well, it's much farther than the obvious range of a refrigerator magnet.

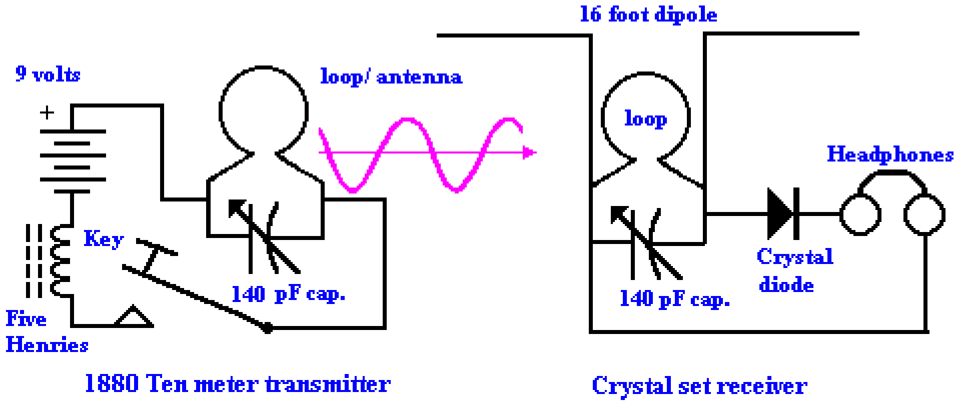
16 foot dipole
9 volts
+
loop/ antenna
loop
Headphones
Key
Five Henries
140 pF cap.
140 pF cap.
Crystal diode
1880 Ten meter transmitter
Crystal set receiver

Improved spark gap transmitter & receiver

By adding a big iron core inductor in series with the battery, I got a much bigger, more sustained spark and a much louder signal in the crystal set. The inductor was the primary of an iron core filament transformer that I had in my junk box. The secondary of the transformer was left open circuit. Actually, I tried several transformer windings and inductors until I found one that gave me the biggest visible spark. The battery was six D-cell alkaline batteries in a plastic battery holder from Radio Shack. My telegraph key was made out of two pieces of printed circuit board separated from each other by a piece of wood. (See Chapter 9.)

Using an extra inductor in this way is analogous to the automobile ignition system described in chapter 2. Without the big inductor, the only inductance to store energy was the single loop of wire. So when I tapped the battery wire on the LC circuit, very little energy was stored in the capacitor and inductance. The spark in a spark gap transmitter happens when the wire charging the inductor is broken, just like opening the breaker points in a car ignition. The more energy stored in the system, the bigger the spark when the connection is opened.

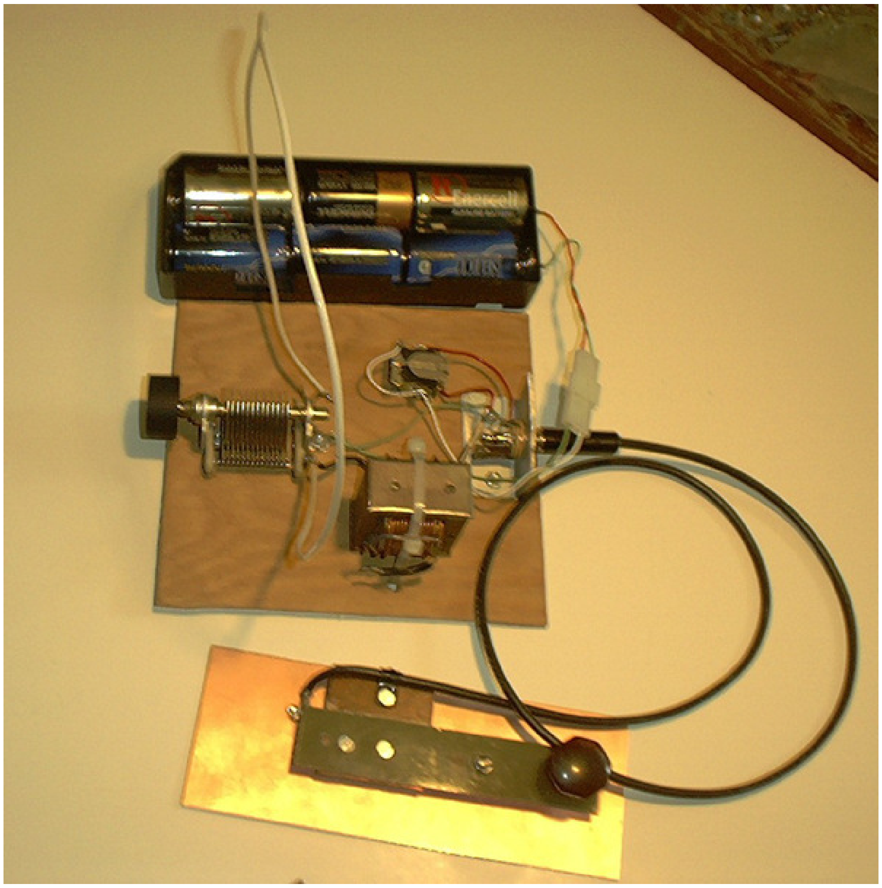

10 meter sparkgap transmitter

I enlisted my XYL to listen to the headphones while I moved the transmitter across the room. (Wives are known as “XYL” in Morse code. XYL stands for “former young lady.") Now that I had the inductor and a larger spark, I got clear out to a range of 10 feet! I explained to Katie that she was doing the same job as Marconi’s assistant, Mignani. When Mignani heard the repeated “S” in Morse, he fired a rifle into the air. “Don't I get a rifle?” Katie asked.

A relay for automatic keying

Unfortunately, a wavelength at 29 MHz is 33 feet and I was still short. Hmmmm ... How to get the last 23 feet of range? First I added a relay to key the inductor. Yes, they had relays in

1880. The relay had nothing to do with extending the range, but it did enable me to continue doing experiments without having a Mignani to do the listening for me. You know, fire the rifle and all that.

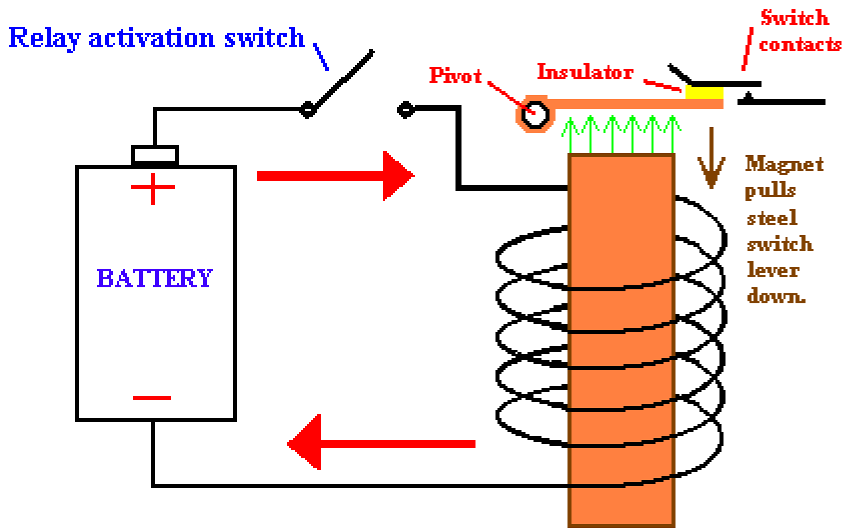

RELAY CONSTRUCTION

A relay is an inductor/ electromagnet wound around an iron core. A hinged piece of steel is suspended by a spring near the iron core. When current passes through the electromagnet, the steel hinge is attracted to the iron with an audible “clunk.” The hinge, in turn, mechanically closes a switch that can be entirely independent of the electromagnet circuit. In this way, one current can control an independent circuit. In my transmitter the switch contacts on the relay became my “spark gap.” Every time the switch opened, a big spark jumped across the switch contacts, not unlike the spark in an automobile ignition spark plug.

Using a relay, I could use my electronic telegraph key, a “bug” set on “dots,” to key the transmitter automatically. This homemade bug is described in Chapter 9. If you don’t have one of those, you’ll have to have your “Mignani” key the transmitter for you. With the transmitter making a continuous buzzing signal, I could move the receiver around my house. The signal sounded just like the automobile ignition noise that you sometimes hear in your AM radio.

More range = bigger, higher antennas plus bigger batteries

I could easily increase the range of the transmitter by using more and bigger batteries and a larger series inductor. If I really wanted to extend the range, I could add a dipole designed for 10 meters and put it up in the air about 50 feet. In fact, this is exactly what the early guys did – they made bigger and bigger transmitters and antennas. However this was 2002, not 1880. The trouble with using a wavelength of 10 meters is that, if I were to increase the effectiveness of my transmitter, I might easily hear it with my crystal set 33 feet away. Unfortunately, someone else might also hear it in Australia. That would be bad since sparkgap transmitters have been banned since 1927.

The simplest improvement I could make to the receiver was to add a 16-foot dipole antenna. The dipole consisted simply of two eight-foot wires soldered onto the sides of the receiver tuning capacitor. The dipole was oriented at right angles to the direct path to the transmitter. That did it. Now I could plainly hear the signal from the basement to the other end of the 2nd floor of my house, over 50 feet. That was well beyond one wavelength range. Success!

Looking for standing waves

To measure wavelength, I set up a long wire transmitting "antenna" about 50 feet long across the floor and upstairs. I reduced the transmitter batteries from 9 volts down to 3 volts. Then I turned on the transmitter. I took the dipole off the receiver and then used the receiver loop as a "probe." Walking along the wire, I was able to hear peaks and dips in reception every 6 feet or so along the wire. What I was hearing was "standing waves." When the RF current reaches the end of an open wire, it bounces back along the wire. The returning waves cancel and reinforce the outgoing waves making the peaks and nulls I was hearing. A large number of peaks means the wire length is different than one wavelength and the standing waves are complicated. If the wire was exactly one wavelength, I would hear just two peaks – they would be the two humps of a single sinewave.

Next I cheated. Since I already knew the frequency was 29 MHz, I calculated what the wire length should be for one wavelength. I trimmed the wire to exactly that distance and tried again. As expected, there was a single pronounced dip in the center of the wire. The sinewave signal was reflecting back and forth from one end of the wire to the other, with a dip, the zero crossing, in the middle. When the reflections don't come out even, you get many multiple dips and peaks.

Of course, knowing the answer before you start is not what Hertz experienced. He had to figure out all the details the hard way. Also, knowing the answer ahead of time biases the result. The exact alignment and distance of the receiver loop with respect to the wire were critical, so there's a chance I was just hearing what I wanted to hear. Craftsmanship and scrupulous honesty are essential when doing science. My frequency measurement obviously more needs work.

****

HOMEBREW TRANSISTORS

Here is another project you may enjoy. It will also introduce you to the basic principles of bipolar transistors. Schockley and Bardeen first invented bipolar transistors while working for the Bell Laboratories in 1947. Actually, I've read that the basic principles for field effect transistors were described in German patents from the 1930s. However, field effect transistors (FETs) were not developed into useful components until the 1970s. We shall first use an FET in chapter 6, so FETs are discussed there.

The purpose of transistors is to "amplify" small signals. Saying it another way, transistors control big currents or voltages using tiny control signals. Transistors can amplify a tiny signal that is audible only with sensitive headphones. They can make it strong enough to run a loudspeaker or even deafen a stadium during a rock concert. Alternatively, a transistor can be used to allow a little control signal to turn on a huge current and voltage. For example, an engineer at a power plant might push a keypad on a computer with a fraction of a milliampere of current flowing through the switch. This action is amplified and results in megawatts of power at hundreds of thousands of volts flowing toward a city.

The homebrew transistor

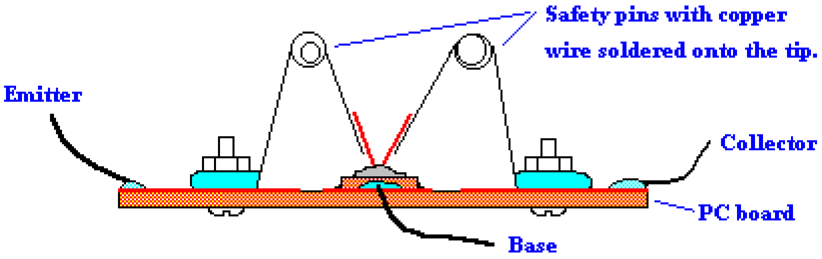

A point contact transistor

The first bipolar transistors were the “point contact” type. They were much like the galena diode described earlier. After my diodes worked so well, I wondered if I could make a transistor.

Idealized construction of a PNP bipolar transistor

A ***bipolar*** junction transistor consists of two diodes made on the same semiconductor crystal. In the experiment above, I am attempting to use galena as the N-type semiconductor. The diodes are connected back-to-back so that it will look like an open circuit from the terminals called “emitter” and “collector.” The two copper points are supposed to touch the galena so close together that the tiny semiconductor region between the two points can be biased by the base current. The bias current is supposed to electrically convert the semiconductor region into a “conductor” and thus turn the two back biased diodes “ON.” Don't forget to cut notches in the copper sheet to separate the three terminals.

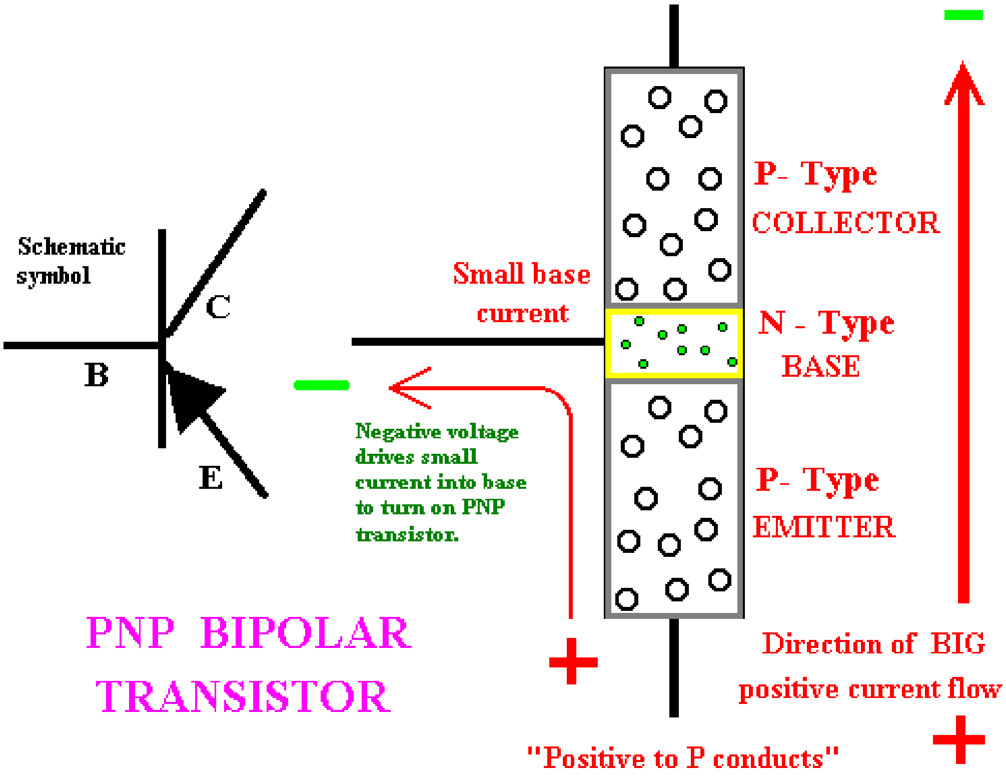
Schematic symbol
C
B
E
Small base current
Negative voltage drives small current into base to turn on PNP transistor.
P- Type
COLLECTOR
N - Type
BASE
P- Type
EMITTER
PNP BIPOLAR TRANSISTOR
Direction of BIG positive current flow
"Positive to P conducts"
−
+

In theory, the metal from the copper points diffuses into the surface of the crystal and makes a tiny region of P-type semiconductor where the copper touches the semiconductor. Unfortunately the "emitter" and "collector" are identical and aren't optimized for their different roles as they are in commercial bipolar transistors. I also tried making a copper/steel/copper transistor, but without any measurable breakdown voltage, it just acted like a short circuit.

NPN transistors

A nifty advantage of bipolar transistors is that they can be built two ways. By reversing the P-type and N-type semiconductors, an NPN transistor can be built that operates exactly like a PNP transistor, except all the polarities and current directions are reversed. The advantage of having two polarities is that the circuits can often be simplified by using both kinds in the same circuit. In practice, NPN transistors are usually slightly more robust and less likely to fail at high power loads. For this reason, the power amplifier stages in modern transmitters are almost always N-channel devices. On the other hand making an NPN transistor out of crude crystals and safety pins is inherently difficult!

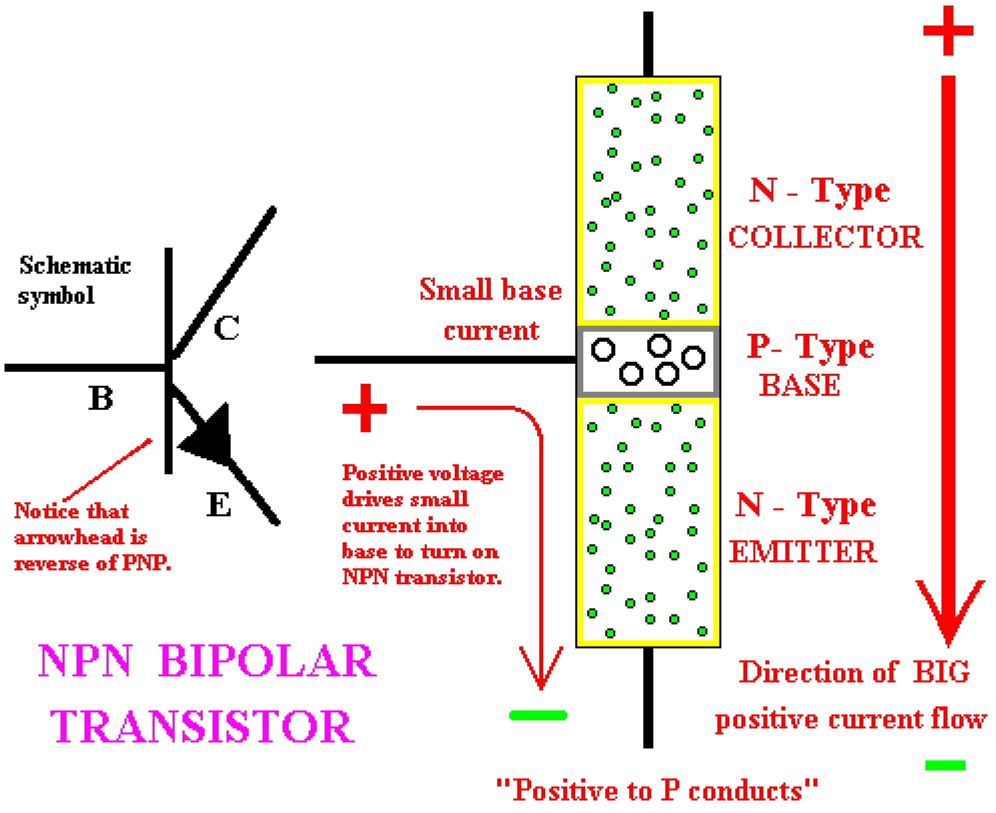
Schematic symbol
C
B
E
Notice that arrowhead is reverse of PNP.
NPN BIPOLAR TRANSISTOR
Small base current
+
Positive voltage drives small current into base to turn on NPN transistor.
N - Type
COLLECTOR
P- Type
BASE
N - Type
EMITTER
+
Direction of BIG positive current flow
"Positive to P conducts"

Proving the homebuilt PNP transistor has gain

I measured the static volt/ ampere characteristics of my galena, point-contact transistor but couldn't show any gain with static DC currents even in the microampere range. Before I gave up, I thought I would try it as an amplifier in the crystal set. Maybe I could demonstrate gain in the subtle world of RF detection.

I decided my "transistor" was most likely to work as an ***emitter follower.*** In an emitter follower, there is no voltage gain, only current amplification. The load, the headphones, would be located between the emitter and the positive side of the battery. That would match the impedance between the high impedance detector and my low impedance (8 ohm) commercial headphones. Because galena diodes break down with typically 1 volt of reverse bias, I used a 9 volt battery with a 10K pot in series so I could limit the voltage on the collector to less than 1 volt.

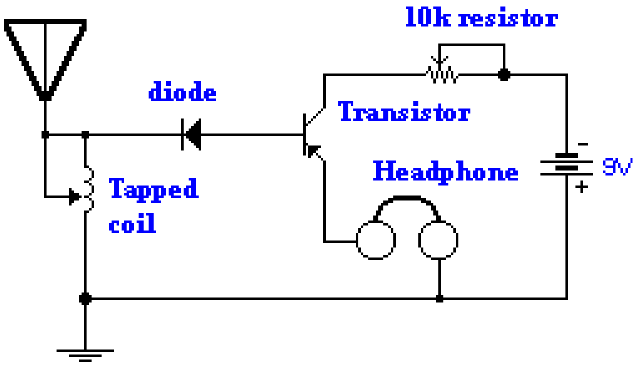

Rock radio schematic

As I moved the emitter pin around on the crystal, a loud radio station abruptly appeared in the headphones. I disconnected the battery. Sure enough, the music was much louder with the battery connected.

An accidental crystal microphone

With the battery in place, but the signal diode disconnected, I heard a faint “sea shell” roaring sound - you know, like a live microphone. Tapping on the transistor assembly, I heard the scratching sound greatly amplified in the headphones. I seemed to have built a “crystal microphone"! I replaced the homebrew transistor with a real PNP transistor, a 2N3906, which amplified just as well but had no microphone-like characteristic.

Commercial crystal microphones consist of a lump of Rochelle salt held between two electrodes. When exposed to sound, mechanical vibrations cause the salt to generate a tiny audio frequency electric voltage. Crystals like this are called piezoelectric crystals and perhaps I had just built one. Piezoelectric quartz crystals are used to regulate frequency and are discussed in chapter 6.

Repairable transistors

While listening to a radio station, I slowly increased the battery voltage by lowering the resistance of the 10K pot. As the collector-to-emitter DC voltage rose higher, the volume increased higher and higher. I monitored the average collector to emitter DC voltage with a high impedance voltmeter. Then suddenly the voltage and sound crashed. I lowered the voltage again, but the sound didn’t return. Good grief! I blew my transistor! No sweat. I just scratched the collector pin around on the crystal until I found a new “sweet spot” and I was back in business. Repairable transistors! Now there’s a concept. After several trials I found sweet spots as high as 5 volts before the transistor died.

Now that I had an amplifier on my crystal set, I replaced the commercial headphone with the homebuilt “Caribou headphone” which I described earlier. You may remember that this headphone was made from a piece of magnetite ore and a tin can lid for a diaphragm. Sure enough, the sound was loud enough to understand actual words, rather than just distant music. More progress!

Is this gain or what?

In fairness, most of the "gain" or amplification I observed was just DC bias to the headphone which helps overcome the hysteresis of the steel components in the headphone. I replaced the transistor with a variable resistor so that the battery's sole function was to bias the headphone. That produced extra sound, especially for the homemade headphone that has such a weak magnet. After switching back and forth between resistor and transistor, the transistor was clearly louder, about 5 dB (decibels) louder measured with my multi-meter.

As I fiddled with the contact points, I eventually ruined the crystal. Under the microscope I could see copper smeared on the surface of the galena. When I built replacements, they never worked as well. I spent another morning trying to conclusively demonstrate gain by building an RF oscillator. I never got so much as a peep out of it, although as you would expect, a real 2N3906 PNP transistor worked just fine.

Real transistors

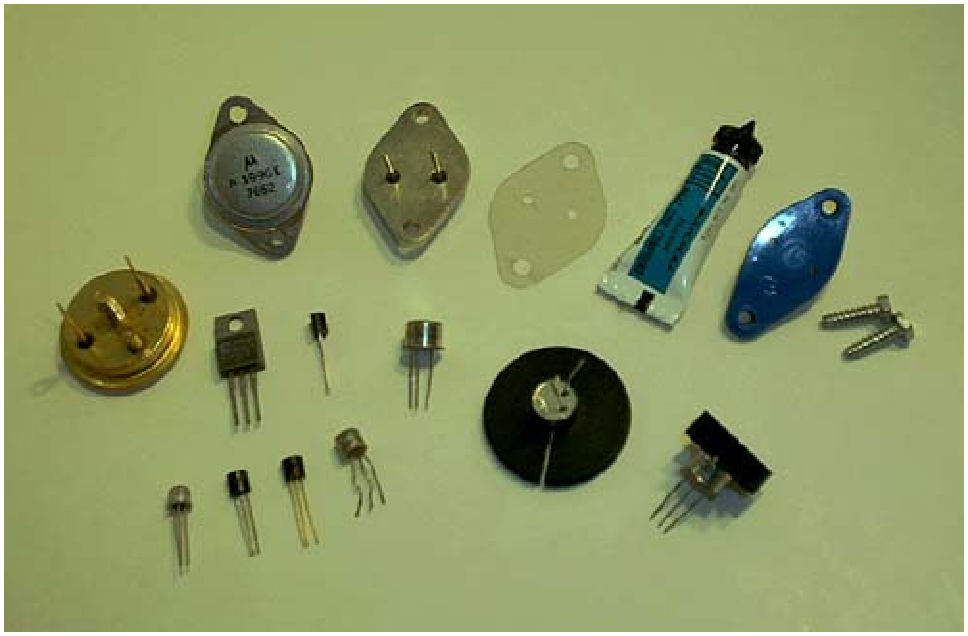

Above is an assortment of bipolar transistors. The little ones on the lower left include 2N3906 and 2N3904 devices used for low power oscillators and amplifiers. The medium sized transistors are used for amplifiers in the range of ¼ to 5 watts. Notice that two of them are wearing black aluminum heat sinks. The large power transistors in the back row are designed for power supplies and high power amplifiers. A mounting kit for case style TO-204 (formerly TO-3) transistors at the upper right consists of a mica insulator, silicon grease, and a socket with screws. The beautiful gold device is an obsolete high power germanium transistor.

In conclusion

Yes, Virginia, there are homemade transistors. But science that isn't reproducible isn't science. Without better basement technology, my homemade transistors have no future except perhaps as microphones. Oh, well. The reward is the journey, rather than the destination. Keep thinking and dreaming!

****

CRYSTAL SETS TO SIDEBAND
© Frank W. Harris 2010, REV 12

CHAPTER 5

GETTING ON THE AIR

Deciding what to do first

Earn an amateur license!

If you're going to build your own ham station, you will need a national government radio station license to transmit. In the US the Federal Communications Commission issues ham licenses. In this country committees of senior hams in our local areas give the tests rather than government officials. The simplest first step is to contact a ham in your area. He or she will tell you what to do. Failing that, if you are an American, contact the American Radio Relay League, www.arrl.org. Most other countries have similar organizations of radio operators. They will fix you up with study guides and point you to the nearest ham radio club where you can meet other hams and take your test. Working on a license can be an interesting side project while you build your receiver and transmitter.

Learning Morse code

If you’re going to build and operate a homebuilt transmitter, it will have to be a CW (Morse Code) transmitter. Building a single sideband (voice) transmitter is too difficult for beginners. Even for experienced guys that takes hundreds of hours of effort. Until recently, all nations required Morse code at least 5 words per minute (WPM) speed to earn a license for operating between 1.8 MHz and 30 MHz. That has changed and many countries including the US now issue licenses without Morse code skill. Fortunately Morse code is not hard to learn for most people - besides, it's fun! The average conversation on the air takes place at roughly 10 to 15 WPM, but there are usually a few beginners poking along at 5 WPM. The senior hams at the bottom of the CW bands zoom along at 25 or 30 WPM. Don’t feel bad if you can’t do that. I can’t either.

To practice code, you can listen to it with a shortwave receiver and you will also need to practice sending with a key. During the evenings in the US the ARRL station W1AW broadcasts code practice sessions. These sessions start at 5 words a minute then gradually increase the speed. Scheduled times for the almost daily broadcasts can be found in the QST magazine. If I haven't been on the air for a while, I tune in the low ends of 40 meters or 20 meters and copy a few CW (code) stations to restore my speed and confidence before I start sending. I write down each letter as I receive it. The really good operators hear entire words and don't have to write down anything! I've never been that good. I spend too much time soldering.

Planning your station

Most hams are content to buy one of those marvelous commercial transceivers. They plug it into the wall, put up an antenna, and start talking. In contrast, a growing number of hams are building small, low power transmitters called ***QRP transmitters***. QRP is an old Morse code abbreviation meaning “turn down the power.” In modern usage QRP means a transmitter delivering less than 5 watts of RF power. QRP guys are usually fellows who have been using commercial rigs for years and want to try something more adventurous and primitive.

Most QRP enthusiasts buy tiny transmitter kits with pre-printed and pre-drilled circuit boards. The assembly procedure consists of following a picture, plugging components into the correct holes, and soldering the leads. These kits teach soldering technique and you will learn what the components look like. Unfortunately, if the kit is well designed, it offers little chance to learn how the circuits work.

A few hams start with a circuit from the ARRL handbook or QST magazine. Following the circuit diagram, they build their own version of the circuit out of individual components. Their components are bought or scrounged one at a time and rarely exactly match the components used by the author of the article. A QRP like this usually doesn't work properly without sweat and rebuilding. This sounds like frustration, but the compensation is that now you're learning something! This kind of project is educational whether we planned to learn anything or not. Frustration, persistence, and triumph is what ham radio used to be.

Rules of the game

Scratch-building your own station is a kind of game. ***Our chief opponents are the amount of free time we have and the modern frequency stability standards.*** In my fantasy it would be neat to build all the components from raw materials and invent and design totally new circuits from scratch. Unfortunately, these tasks are too much for me. Rather, I use discrete parts I scrounge from junk and components I buy through catalogs or local stores. For circuit design I borrow circuit blocks from examples in the ARRL handbook and other sources. Then I assemble the circuit blocks into my project.

After I retired and had plenty of time, I got out my ARRL handbooks and studied plans for homebuilt receivers and transmitters. I quickly encountered my other nemesis, ***integrated circuits:*** Tracing the circuit diagram of a typical modern receiver, one finds that the antenna is connected to a chip marked "RF amplifier." Then the signal goes to another chip labeled "mixer." The mixer receives a local oscillator signal from a large chip called "frequency synthesizer." The output from the mixer goes to a "crystal filter" sealed in a tiny metal can. Well, you get the idea. ***Modern radios are just block diagrams soldered onto printed circuit boards.***

I find that 1980 technology is usually a good compromise. Few integrated circuits appear in the parts lists from that era. And if you're a purist, most of the integrated circuits can be replaced with discrete components. In summary, I try to use individual transistors and passive components exclusively. This teaches me how the circuit works and I can still build equipment that's reasonably modern. One compromise I have had to make was integrated circuit voltage regulators. In order to hold the frequency constant, the voltage supplied to the oscillators must be more stable than I know how to build using discrete parts.

Another compromise I've made is the use of integrated circuit operational amplifiers. An operational amplifier is an array of transistors that functions as though it were a "perfect transistor." It is usually easy to substitute real transistors, but you need to use many more of them to get the same performance. After a while I became bored making the same complex circuit again and again using discrete transistors. So, my rule is that, if I have proved I can build the equivalent of an integrated circuit with discrete parts, I feel I have "earned the right" to use certain kinds of integrated circuit.

My rules have taught me how the circuits work. I've learned plenty. And by golly, when I'm done, I really feel that I built it myself! Because I have struggled to understand it and make it work, I rarely have the feeling I am using "magic." You don't have to use my rules. Make up your own. If you're having fun, those are the right rules.

Planning your station

A simple receiver and a QRP transmitter are a great start on a homebuilt station. (See Chapters 6 and 7) However, before you can go on the air with your transmitter, you will also need:

- A ham radio license. Contact local hams in your area or, in the US, the American Radio Relay League in Newington, Connecticut. (888) 277-5289.
- A 12 volt DC power supply for the QRP. (Chapter 8.)
- An antenna (See end of this chapter, Chapter 5)
- An antenna tuner (Chapter 9)
- A telegraph key for Morse code. (Chapter 9)
- Basic test equipment to ensure that your transmitter is operating correctly. (Chapter 9).

Equipment for later on:

- A selective, sensitive, all-band hamband receiver (Chapter 13)
- A Class B or linear amplifier. (Chapter 12) Although QRP is fun, you will probably eventually wish your signal were more powerful. With a loud signal more stations will hear you without a struggle and your contacts won't say "73" (good bye) so quickly. There are two ways to achieve a bigger signal. Build a better, high gain (directional) antenna, higher above the ground. The other way is to build a linear (or Class B) amplifier to raise the QRP power from 5 watts up to 50 or more watts.
- **A lowpass filter**. (Chapter 9) If you have a high power Class B amplifier, it is a good idea to build a low pass filter to make sure that harmonics of your signal are not interfering with the neighbors' cellphones and television.
- **VFO**. As you'll soon learn, being restricted in frequency mobility by crystal control is awkward and you will eventually want to build a variable frequency oscillator to replace those clumsy and expensive crystals. (Chapters 10 and 11)
- **Single Side Band Generator.** (Chapter 15) After you have been a ham for a few years, you will sooner or later want to get on phone (voice) so you can talk to the majority of hams. Otherwise you'll be stuck just talking to funny old Morse code geezers like me. Almost no one ever builds their own SSB. So if you get that far, you will have joined an extremely exclusive club of homebuilding hams.

Divide and conquer

Building an entire ham radio station from discrete components is a huge project. The secret to having fun is to build it in carefully planned pieces that can be used and enjoyed right away. How big the challenge becomes depends on where you start. If you're a retired electrical engineer like me, the project begins by wandering down to the basement, dusting off the workbench and getting started. If you are still working for a living, your free time for this effort may be just a few hours a week. If you have never worked as an electronics technician or as an engineer and you have no formal electronics training, your road to homebuilt ham radio will be lengthy. If you have no electronic experience at all and little free time, then maybe buying a receiver and a QRP kit may be the best way for you to begin.

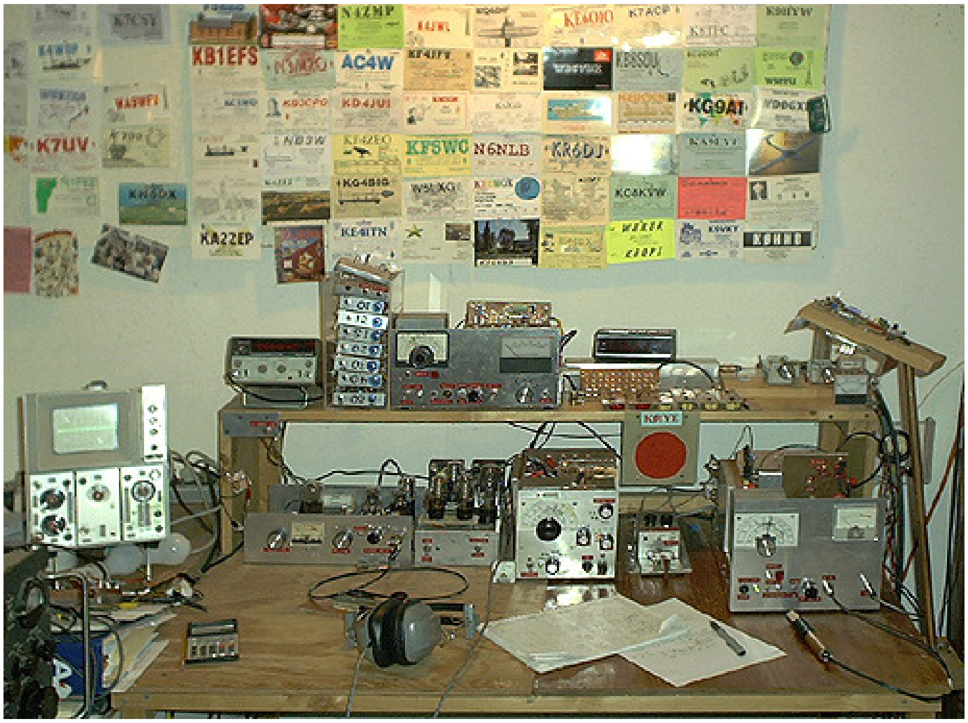

The author's 100% homebuilt station. It isn't flawless and it isn't the latest technology, but it talks around the world and sure has been fun. Oh, by the way, the oscilloscope and frequency counter are *NOT* homebuilt.

Many guys try to do too much too soon. This can easily result in many hours invested in useless junk that does nothing except get hot and generate acrid smoke. A better plan is to focus on small projects that can make you proud as soon as possible. Try to decide what it is about ham radio that turns you on and then zero in on those aspects. Your ambitions will always be greater than the time you have to achieve your dreams. Plan carefully.

Pick an HF band

50 years ago it was practical to build simple ham transmitters that tuned several bands. Unfortunately today a ham transmitter must meet strict frequency stability standards. Actually, your government is only interested in having you stay inside the hamband and radiate a reasonably narrowband signal. However, other hams will almost always complain about your signal if you drift more than about 50 Hertz per minute. This degree of precision can be hard to achieve, but it's easiest if you start with crystal frequency control and just build for one hamband at a time. Later you may expand your capability as your time, knowledge, and enthusiasm permit.

There are ten different HF bands and seventeen VHF, UHF, and microwave ham bands. In addition, a ham is free to operate anywhere he likes above 300,000 MHz, provided output power is limited. As the frequency goes up, the basic principles are the same, but the construction methods change radically and the precision and craftsmanship needed soar with the frequency. This book only considers the HF bands because scratch building equipment for VHF and higher frequencies needs expensive test equipment and a good understanding of HF building principles. In other words, walk before you run.

40, 30, and 20 meters

In my experience, the easiest and most versatile frequencies to build for are the 40, 30 and 20 meter hambands. (7 MHz, 10 MHz, and 14 MHz.) The signals on these bands are strong and these bands are usually "open" throughout the year. For Americans it's easy to talk to hams all across North America on all three bands. Talking around the world is relatively easy on 30 and 20 meters. Optimal antennas are relatively small. On 40 meters (7 MHz) a dipole is 66 feet long while on 20 meters (14 MHz) a dipole is only 33 feet long. A vertical antenna on 20 meters is only 16.5 feet high. A typical suburban house is about 50 feet long so it's usually possible to put up a 40 meter dipole somewhere on the property.

40 meters (7.000 to 7.300 MHz)

This is a great band for CW contacts within about 2,000 miles. However, for us Americans, often only the bottom 50 KHz is usable. Yes, we're allowed to operate from 7.00 MHz to 7.30 MHz. Unfortunately, foreign broadcast stations often obliterate the upper half of the band. On the other side of the world most of this band is available for commercial shortwave broadcasts. Also, from 7.05 to 7.10 the band is filled with foreign hams talking on single side band phone (SSB, voice transmission). We Americans can't use phone below 7.150 MHz, so we can't join in. Consequently, most of the time American hams just use the bottom 50 KHz for CW. 40 meters can be useful for local sideband phone contacts during the daytime. For example, in my state there is a noon SSB net where guys around the state chat with each other. However at night most of 40 is unusable.

30 meters (10.10 to 10.15 MHz)

Thirty meters resembles the CW band on 40. It's only 50 KHz wide and no phone is allowed. Typically there are only 4 or 5 CW stations on the air. It's a terrific band for homebuilders. The power output on this band is limited to 200 watts worldwide, so it is a good band for talking to DX (foreign stations). Nearly everyone on 30 is using comparable equipment, so you won't have to compete with kilowatt stations using giant, multi-element beam antennas. Another advantage is that contesting isn't allowed on 30. Ham organizations often hold weekend contests to see how many stations one can talk to in a 24 hour period. All the bands except 30, 17, and 12 meters become packed with guys just exchanging call letters. It's pretty boring if you aren't competing in the contest.

20 meters (14.000 to 14.350 MHz)

Twenty meters is the single most popular band. It is large and, depending on the 11 year sunspot cycle, usually usable 24 hours a day and "open" year round. It normally has hundreds of stations. In fact, 20 meters is often so crowded, you will need a receiver with great selectivity to carry on a conversation. If you can copy fast Morse code, foreign CW stations are easily heard at the bottom of the band. Because of the crowds and high speed, talking to them isn't so easy. Most of the band is filled with powerful stations on single side band phone. It is hard to build a homebuilt station that will compete on the 20 meter phone band.

17 meters (18.068 to 18.168 MHz)

Seventeen meters is a fine little CW and phone band. It's mostly inhabited by single side band phone stations. However, there are usually a couple CW stations at the bottom of the band. When conditions are good, many of these stations are DX on the other side of the world. It isn't my first choice for a QRP CW transmitter, but other than scarce CW stations, there's nothing wrong with 17 meters.

15 meters (21.000 to 21.450 MHz)

Fifteen meters is slightly harder to build for than 20 meters, but it's an excellent band for beginners. 15 meters is a large band, 450 KHz wide. It's not crowded but there are plenty of stations. 21.150 MHz is a great place to meet beginner CW operators. 21.040 to 21.060 MHz is where most of the QRP CW guys hang out. 15 meter QRP transmitters are crystal controlled and are usually locked into that frequency range. After your code speed gets up over 20 WPM, you can move down to the bottom of the band where the DX (foreign) stations and high-speed, old-time CW operators hang out. On the other hand, every now and then a station on the other side of the world will answer your slow speed CQ. Actually, these are usually the best DX QSOs because the other fellow obviously is no hurry to talk to anyone else.

Extra care in construction and antenna tuning are needed as the frequency goes up. You may find that your homemade receiver is slightly less sensitive than it was on 20 Meters and your transmitter probably won't produce as clean an output waveform as it did on 20 meters. On the other hand, a 15 meter vertical antenna is only 11 feet tall and a dipole is only 22 feet long. Another reason for selecting 15 meters is that ***a 40 meter dipole (or 40 meter vertical) will work well on both 40 meters and 15 meters***. This simple antenna is ideally matched to both bands.

Some hams successfully string 15 meter antennas in their attics or along the ceiling of a large second story room. When I was in high school, a friend of mine, Al Beezer, KØKZL, had his 15 meter antenna strung across his second floor bedroom. He attached florescent light bulbs at each end of his dipole. The bulbs flashed every time he pushed the telegraph key. Great show biz! Yes, Al's signal would have been stronger if his antenna had been 10 feet higher outside in a tree, but his indoor antenna was adequate to make contacts all over the U.S.

Open bands and sunspots

The disadvantage of 17 meters and higher bands is that these bands are often not "open." When the sunspots are few, the Earth's upper atmosphere is minimally ionized. When the band is "dead" the radio signals are passing through the ionosphere right out into space. The sunspots occur in an 11-year cycle and 2008 happened to be one of those years when the upper bands are quiet. As of 2010, the high bands are just now beginning to be active again. 15 meters is often closed for months at a time, especially in the summer. During these times all you'll hear is the hiss of static. On the other hand, 15 meters can still be used for local communications during these periods. The higher the frequency, the less often the bands are open. Ten and 12 meters are even more erratic.

10 and 12 meters

I found building receiver converters for 10 and 12 meters was as easy as it was for 15 meters. However, building the transmitters was more difficult. My final amplifiers and antennas were hard to tune. It took me several attempts and many modifications before I got more than a few watts of power output on 10 meters. For example, my transmit-receive antenna relay had too much reflection (high SWR) to work properly on those bands. I had to bypass the relay and use a separate antenna for the receiver. However, once I was able to transmit just two watts on 10 meters, I could easily talk to other continents. In that sense, I believe 10 meters is the best QRP band. But as I said, sadly, months often go by when there is nothing to hear on 10 meters but static.

When I was in high school, several of my friends and I built handheld 10 meter walkie-talkies. They put out 1/4 watt into a 2-foot long antenna with a loading coil. One of my biggest thrills in ham radio was standing on my chimney in Colorado and talking to a guy in New Jersey - 3,000 kilometers away. That was nothing. My friend, Bob (KØIYF, now NØRN), was walking down the street and talked to a fellow in Morocco. Fifty years later he still remembers the call letters, CN8NN.

6 meters

Six meters is a huge band, 50 to 54 MHz . It closely resembles 10 meters in that a few watts will talk around the world using high ionospheric propagation. Some years, on ham field day at least, (the last weekend in June) it is crowded with signals from all over the world. The catch is that 6 meter propagation is rarely “open.” You might have to listen for many months before you hear stations. With good craftsmanship, it’s possible to build homebrew equipment for 6 meters using the technology discussed in this book. However, I don’t recommend it until you have become bored with 10 meters.

80 and 160 meters

80 meters (3.5 to 4.0 MHz) and 160 meters (1.8 to 2.0 MHz) are large, lightly used bands. Of all the hambands, transmitters are most easily built for these bands. There are plenty of SSB phone stations on 80 meters at night, but relatively few CW stations on these bands. Good receivers are hard to build for these bands because the atmospheric noise is so heavy that it takes a sophisticated receiver to pull signals out of the static. Noise from lightning is especially bad in the summer. Also, if your homebuilt receiver isn’t well filtered, it may be overwhelmed by your local standard broadcast AM stations, which are located just below 160 meters. A modern commercial high-end transceiver can often hear dozens of stations while a simple homebuilt receiver might only hear a few. During the day, these bands buzz with noise and are usually only good for across town. Sometimes at night, if the thunderstorm noise isn’t too bad, there are lots of stations from all over the continent. It also helps to live out in the country far from power lines, plasma TVs, auto ignitions and heaven only knows what else is in city background static.

Another barrier to 80 and 160 meters is that large antennas are needed for good performance. On 80 meters a vertical antenna must be about 66 feet high and a dipole antenna is 135 feet long. For 160 meters, ideal antennas are twice those sizes. In my opinion 160 is the noisiest and least popular band. Yes, you can get on 80 and 160 with small antennas equipped with “loading coils” and antenna tuners. Unfortunately, in my experience, if you don't have a big antenna, you will rarely get out of town. On the other hand, once you have a full size antenna, I was surprised to discover that QRP transmitters can talk all across the continent on 80 meters.

60 meters

Since July 4, 2003 American hams have been allowed to transmit single sideband phone (USB SSB) on five discrete channels, 5.332, 5.348, 5.368, 5.372, and 5.405 MHz. This band is noisy and is shared by several services. Each channel is like an old time party-line phone on which everybody tries to talk at once. Moreover, at any one time, you will find that two or three of the channels are occupied by commercial teletype (RTTY) stations. It's interesting, but frankly it isn't worth the trouble for a beginning ham.

In summary, I recommend starting off with CW on 40 and/ or 15 meters. A 40 meter dipole works well on both bands and the equipment is relatively easy to build. Both bands will let you talk all over your local continent with ease. On 15 meters you will have the chance to work foreign stations even with simple equipment and minimal code skill.

Phone (voice) transmissions

For a homebrewer in the modern world, building an SSB phone transmitter is a difficult project. Actually, amplitude modulation (AM) voice transmitters, like those used on standard broadcast radio, are relatively easy to build and are still legal for ham radio. To convert a CW transmitter into an AM transmitter, all you have to do is impress an audio voice signal onto the power supply line of the transmitter’s last stage of amplification. Occasionally you can still hear AM stations on 10 meters, 80 meters, and 160 meters. But in general, AM is rarely used today by hams.

The modern replacement for AM is ***single sideband (SSB) phone***. SSB is similar in principle to AM phone, but it takes up about a third of the bandwidth of an AM signal. Sideband can be thought of as AM radio with half of the signal plus the carrier wave removed with precise filters. This allows more stations to share the band and makes your transmitted power three times more effective. The SSB signal must be generated at a fixed frequency then translated up or down to the desired frequency using mixer technology. To maintain the purity of the signal, all of the amplifiers along the chain must operate perfectly in linear mode so that the speech is not distorted.

A sideband generator is described in chapter 15 of this book. Scratch-built sideband phone is definitely an advanced project. To be honest, I found mine hard to get working and spent many months on the project. I rebuilt some stages of the transmitter several times. Perhaps homebuilt single sideband should be something to look forward to when you have loads of free time. After you have worked on an SSB transmitter for a while, you may decide that good old Morse code isn’t so bad after all!

On the other hand, receiving single sideband is easy. A filter made from a single, off-the-shelf, microprocessor crystal is all you need to receive clear, interference-free speech. The homemade receiver described in chapter 13 works great for SSB.

Instant high quality communications

Perhaps you have read all of the above and decided that your primary goal is to talk to people by radio as soon as possible. For you, building equipment is a secondary interest. If that describes you, you should probably buy a modern high frequency (HF) transceiver then begin work on your ham license as soon as possible. Modern commercial ham stations are usually "transceivers" that combine the transmitter and receiver in one unit. Many of the circuit blocks in transmitters and receivers are almost identical. Logically enough, the transceivers use the same circuits for both send and receive.

A modern, top-of-the-line HF transceiver – the Yaesu Mark V Field, FT1000MP

I recommend buying a transceiver that can tune all ten high frequency bands. If you buy a modern, quality **used** transceiver, the price can be well under $1,000. This will give you a (nearly) complete station that is more sophisticated than you or I could build with years of effort. Even if you don't have your license yet and can't transmit legally, you should buy or borrow some kind of receiver as soon as you can. That way, you can at least listen to shortwave radio. Really, listening is more than half the fun. Many commercial transceivers or receivers can receive the entire shortwave spectrum from 1.8 to 30 MHz. This includes all the commercial broadcast, foreign stations, and time and weather stations. Listening will make studying for your ham license much less theoretical. If you want to experiment with the built-in transmitter, you may not transmit on the air without a license. However, you can practice loading a "dummy load" instead of an antenna. If you have an antenna tuner, it's fun to load up an ordinary filament lightbulb as a dummy antenna and watch the light pulsate as you talk or send Morse code. See Chapter 9.

Occasionally I consider buying a modern transceiver like the one above. These transceivers are "easy" in the sense that they can instantly perform virtually every feature and communication method you have ever heard of. For example, when the new 60 meter band opened, it was possible to reprogram the Yaesu FT1000MP to cover 60 meters just by pushing the right combination of buttons. And, as you might expect, practically everyone on 60 meters was using an FT1000MP. The bad news is that the instruction manuals make operating your VCR look like child's play. All those buttons, menus, and operating modes are enough to drive you to drink.

Yes, I occasionally think about buying one, but for me, it wouldn't be satisfying. As soon as I had read the manual and tried out the features I was interested in, the transceiver would become boring. I can name several hams who have superb transceivers gathering dust in their home stations. They haven't turned them on in months or even years. On the other hand, if your ham radio bliss is working all the 335 odd call signs in the world, then you will need the best possible equipment. Building your rig yourself is going to slow you down for many years or decades. Maybe that's a good thing. You decide!

Pursuing a dream

When I was in high school 50 years ago, television was “exotic.” Although black and white TV receivers were common in our living rooms, almost no one could afford a TV camera. Because I couldn’t afford a camera, their unavailability made them fascinating. I spent hundreds of hours building and experimenting with TV cameras. Only one of my five camera projects actually worked properly. (See chapter 16.) Later when I was grown, my wife bought me a fantastic TV camcorder which today is a big yawn. Because camcorders are routine today, to me the camcorder was just another way to take baby pictures. I have hardly done a thing with it. ***Familiarity and ease of acquisition produce boredom.***

VHF/ UHF Hand-helds

The simplest and cheapest kinds of commercial ham transceivers are VHF hand-held walkie-talkies for 2 meters (144 MHz) and/or 70 centimeters (420 MHz .) These walkie-talkies are slightly more adventurous than a cellphone, but in my opinion they aren’t as useful or interesting unless you are using IRLP, described below. Hand-helds also require a ham license, but in the US at least, a VHF "technician" ham license has the easiest test. Another limitation of hand-helds is that the communication is basically line-of-sight. In other words, if you are behind a mountain and want to talk to the other side of the mountain, you need a relay station (a repeater) that just happens to be on top of the mountain or in some location that is visible to both you and the guy you’re talking to.

The best feature of hand-helds is ***IRLP***. This is an Internet-connected network of VHF and UHF repeater relay stations or "nodes." The idea is that you use your handheld transceiver to key in codes into your local repeater node. The repeater uses the codes to connect you via the Internet with other nodes anywhere in the world. The repeater node in Boston, Australia, or Berlin then rebroadcasts your VHF signal as if it were local. Guys in those distant cities hear you and reply as if you were located in their own town. This system is mostly used for rag-chewing and isn't burdened with contests and guys who are just interested in collecting call letters and QSL cards.

A 2 meter hand held transceiver (walkie-talkie) with a spare battery pack.

Personally, I use my 2 meter walkie-talkie as an emergency radio for hiking. Honestly, I think a cellphone would probably serve me better in an emergency, but I've been too cheap to buy one. Most owners of VHF walkie-talkies use them like citizens band radios to talk around town. In general, the quality and range of the commercial amateur radio hand-helds is far superior to citizens band radios. On the other hand, citizens band and the "Family Radio Service" (FRS) are other "quick and dirty" ways to get on the air without a license.

If you buy a hand-held, be sure to buy an extra battery pack or two. To make mine reliable in an emergency, I replaced the rechargeable batteries inside the battery packs with alkaline batteries. Although they aren't rechargeable, alkalines store several times more energy and tolerate cold temperature much better than rechargeables. I found that the original NiCads went dead almost immediately when I was out skiing.

HF versus VHF and UHF ham radio

Among serious hams, the spectrum is generally divided into two halves, "High Frequency" (HF = 1.8 to 30 MHz) and "Very High Frequency" (VHF = 50 MHz) extending up to Ultra High Frequency (UHF = 220 MHz and above). Above UHF are the microwave bands with frequencies measured in gigahertz and wavelengths in centimeters. Most HF hams like long distance communication with either CW or voice. There are also HF packet-radio modes and slow speed picture transmission which resembles e-mail with a very slow modem. RTTY and PSK-31 are modes that resemble instant e-mail messaging on the Internet.

There are two kinds of VHF/UHF guys. The vast majority of them just own hand-helds, but a few UHF hams are extremely technical and do high tech experimentation. This can be ham-television, exotic voice modulation like spread-spectrum and transmissions using satellites. One way the microwave guys have fun is trying to see how far they can communicate between mountain tops.

Much like my effort to build a TV camera, what you do with great effort is more rewarding than buying a gadget and copying the crowd. So it shouldn't be a surprise that, for the technical UHF guys, a lot of the fun is organized around trying to talk to distant stations using ionosphere substitutes like satellites, or bouncing radio waves off the moon or even the aurora and falling meteors. Instead of using repeater relay stations, these guys get their kicks by using these exotic substitutes.

****

BUILDING AN ANTENNA

Every shortwave radio station needs an antenna and building one should be your first project. Shortwave reception with a modern receiver can be adequate with as little as a few feet of wire dangling off the antenna connector. Many commercial short wave receivers built for non-hams have coil antennas (magnetic antennas) built inside the receiver case. Tiny antennas like this are adequate for hearing powerful stations such as Voice of America, Deutsche Welle, (Radio Germany), Radio Moscow, or the BBC (British Broadcasting Corporation).

Amateur radio stations, by law, are limited to 1,500 watts Peak Envelope Power (PEP) and most of them run 100 watts or less. This means that hams are hundreds of times weaker than the commercial stations mentioned above and a big antenna is important to hear them. The lower the frequency, the bigger the antenna you will need. A good transmitting antenna needs to be at least ¼ wavelength long. A good transmitting antenna for shortwave radio signals must be high enough to send signals toward the horizon with as few obstructions as possible. Although an adequate receiving antenna can often be small and close to the ground, a large, high wire antenna works well for both jobs.

The long wire

The simplest, most versatile antenna is a “long wire.” It is literally a long piece of wire, ideally 1/4 wavelength long or larger. For example, for 40 meters, it must be at least 33 feet long and should be strung up in a tree or other place as high as possible. In general, if you connect a random length of long wire to a transmitter, it will not be resonant with your hamband and you must “match it” with an impedance transformer called a ***transmatch*** or ***antenna tuner***. Fortunately, an antenna tuner can be quite simple and can be put together in a couple hours. An example is described in Chapter 9.

The 40 meter dipole

If I could only have one antenna, this is the one I would build. This simple antenna works well for both 40 meters and 15 meters (7 MHz and 21 MHz). As will be explained later, extra antenna arms can be added so it will work on most of the other hambands. The advantage of a dipole antenna is that they have an impedance of roughly 50 ohms that is easily loaded by a ham transmitter. So, although an antenna tuner is still helpful, this antenna will transmit quite efficiently without much tuning.

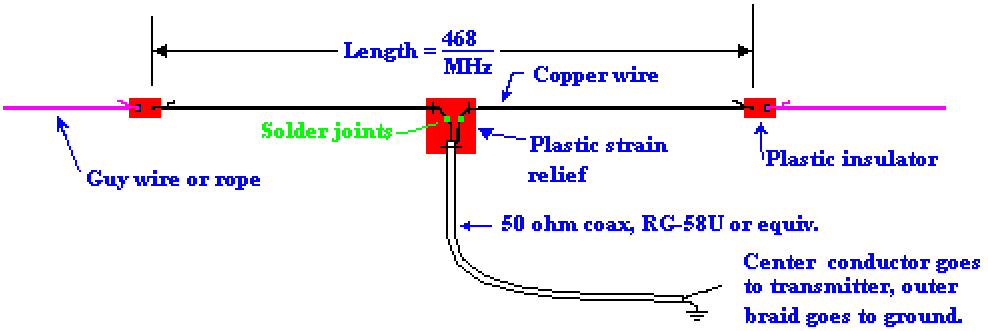
Length = 468/MHz
Copper wire
Solder joints
Plastic strain relief
Plastic insulator
Guy wire or rope
50 ohm coax, RG-58U or equiv.
Center conductor goes to transmitter, outer braid goes to ground.

For 40 meters the length is 66 feet. The coax feed line can be any length, but shorter is better.

Where do I get antenna materials?

Radio Shack stocks RG-58U coax, insulators, and heavy, multistrand copper wire that are perfect for dipoles. I usually make my insulators out of plastic scrap from a local plastic fabrication shop or I use ABS plastic pipe sections from the hardware store. The total length of a dipole in feet is easily calculated. Divide the constant number 468 by the frequency in megahertz. For 40 meters, 468 / by 7.05 MHz is 66 feet long. If you have trees, a roof, or other tall objects on which you can mount an antenna, you’re in business. Hang it as high as you can. I use relatively light nylon rope to secure the ends to trees. The nylon stretches as the wind blows and prevents the antenna from breaking. Steel or aluminum guy wires don't stretch and are more likely to break although I suppose stiff springs could also be used to take up the slack.

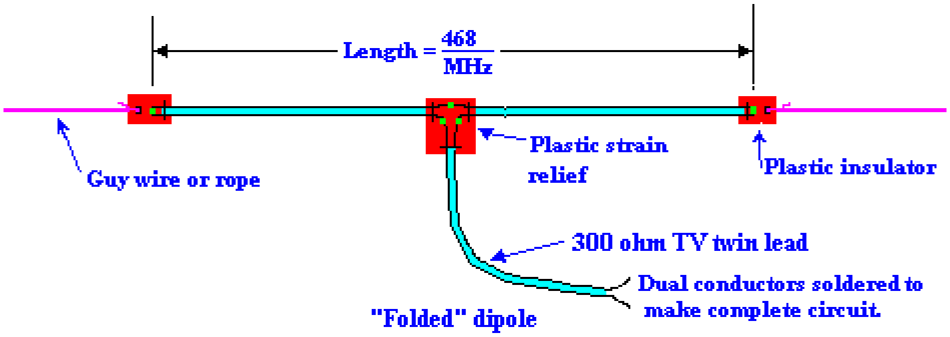

The folded dipole

Homebrew dipoles come in two basic flavors, “regular” and “folded.” Back when I was young and impoverished, my favorite dipole was “folded” and made entirely from 300 ohm TV twin lead. The dual conductors of all three arms of the “T” are soldered together to complete the circuit in a “loop.” I had one antenna cut for 40 meters that I used primarily for 40 and 15. It even worked OK on 75 meters for the local AM phone round tables our club used to hold. Aside from being made from cheap wire, another advantage of the folded dipole is that flat 300 ohm twin lead can slip through a slightly open window without mashing a fat, round cable or without drilling holes in the window sash.

Yeah, I know. 300 ohm balanced folded antennas need a ***balun coil*** and a transmatch to match the transmitter. When I was a happy, dumb novice, I just loaded it directly with my unbalanced π (pi) output class C vacuum tube transmitter and it worked very well, thank you. Then, when I learned how balanced antennas are supposed to be driven, I bought some fancy balun coils. I was disgusted when I observed no difference in performance.

Multi-band Dipoles

Today we have four more HF bands than 40 years ago and I can’t resist trying them all. Any band I can’t work MUST be wonderful, right? I have a “regular” 40 meter dipole in the front yard and a 30 meter one in the back yard. I guess I have room for an 80 meter Zepp, but I’ve been too lazy to build one. Jack Ciaccia, WM0G, uses the gutters and drainpipes on his house as an 80 meter “longwire.” This sounded like a terrific solution. Unfortunately, my laziness soared when he began talking about “the buried radial ground field.” In other words, I was supposed to dig long ditches across the lawn in several directions and bury multiple ground wires. My wife would be thrilled with that.

A year later Jack told me that deep heroic ditches are unnecessary. He waters his lawn to soften the dirt and roots. Then he digs his "ditchs" with a large pizza roller cutter. He cuts a slit in the turf a couple inches deep. He slips the radial wires into the slits and, when the turf grows back, the wires are sealed under the grass.

Several years ago the 80 meter band was exceptionally “hot” with dozens of CW stations on every night. I was eager to get on the air. I tried loading my 40 meter dipole on 80. I managed to do it, but there were hundreds of volts at the station end of the coax. (With a 40 meter antenna, the transmitting impedance is a tad high on 80 meters.) After a couple hours of calling and tweaking, no one answered me. It seemed clear that I wasn't getting out and I needed a better plan.

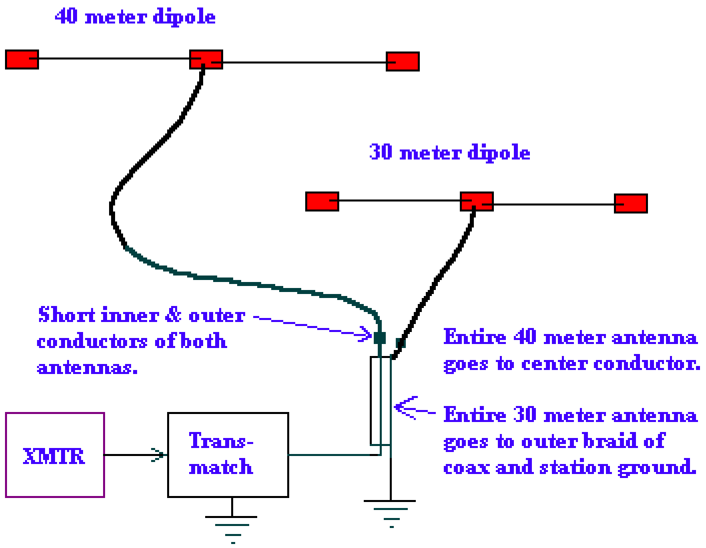
40 meter dipole
30 meter dipole
Short inner & outer conductors of both antennas.
Entire 40 meter antenna goes to center conductor.
Entire 30 meter antenna goes to outer braid of coax and station ground.
XMTR
Trans-match

Then the cerebral lightbulb dawned! Why not use the entire 40 meter dipole, coax and all, as a longwire? For a counterpoise I first used my station ground, which is heavy copper wire connected to a water pipe. This loaded better than the 40 meter dipole but still didn't work very well. Next I connected the entire 30 meter dipole, coax and all, to the station ground. My 30-meter dipole is in the back yard and well separated from the 40 meter dipole. Presto! I suddenly had a low impedance 80 meter antenna that loaded beautifully. There was one little glitch: My wife stormed downstairs and told me that I was obliterating the TV pretty much on all channels. Oh well, after that experience I got on 80 after she went to bed. I think (pray) the TVI is limited to inside my house. My theory is that I'm loading up the house wiring and the TV is immersed in the antenna near field. More recently we bought a much more modern TV. It seems to be immune to not only my 80 meter signal, but my signals on all the HF bands.

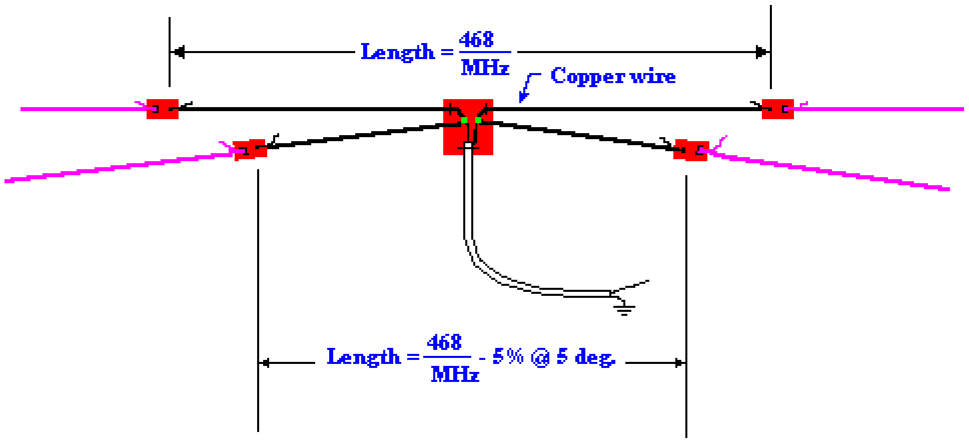
Length = 468/MHz
Copper wire
Length = 468/MHz - 5% @ 5 deg.

The multi-band dipole

When I tried to get on 20 meters with my 40 meter dipole, I had a high impedance problem similar to using the 40 meter dipole on 80. However, using the antenna tuner described in chapter 9, I still managed to work folks. Consulting the ARRL handbook, I added 20 meter elements to my 40 meter dipole to turn it into a two-band dipole. A simple formula for the length in feet of both of the secondary elements taken together is 468/ frequency in MHz minus 5%. Each arm of the dipole is offset from its neighbor by about 5 degrees. This lowered the impedance on 20 meters to something near 50 ohms and has been quite successful.

Now suppose you want to add elements to the 40/20 meter dual band dipole to cover other bands, such as 30, 17, 12, and 10. Yes, it's possible to do this. Steve, WØSGC, built a five-band dipole just like this. The ARRL handbook doesn't recommend this because such an all-band array tends to radiate every stray harmonic your transmitter may generate. In other words, it works too well! On the other hand, three-banders, such as 30/17/12 meters or 40/20/10 meters are recommended. Notice that you don't need to add an element for 15 meters because the 40 meter elements already work well on 15 meters. Therefore a tri-bander built for 40, 20, and 10 is actually a four-band antenna.

The length of the third dipole is calculated just like the second dipole. Suppose you already have the 40 and 20 meter dual band dipole and want to add another dipole for 10 meters (28.1 MHz). The length would be 468/ 28.1 MHz or 16.65 feet minus 5% or 15.82 feet. (15 feet 9.8 inches). The third antenna would also be offset from the center (20 meter) dipole by 5 degrees. Do *NOT* decrease the length by 5% a second time.

A limitation, or possibly an advantage, of a dipole is that it transmits the most energy perpendicular to the wire. It also radiates significant power in line with the dipole. That is, dipoles are somewhat directional. Therefore, if you wish to talk to South America or over the North Pole to Russia, ideally in the US your dipole should be oriented east to west. Conversely, if you wish to talk across the US to the east and west coasts, the dipole should be oriented north and south. This isn't a big deal in performance, but if you have your choice in places to string the antenna, it is something to consider.

The curtain rod vertical

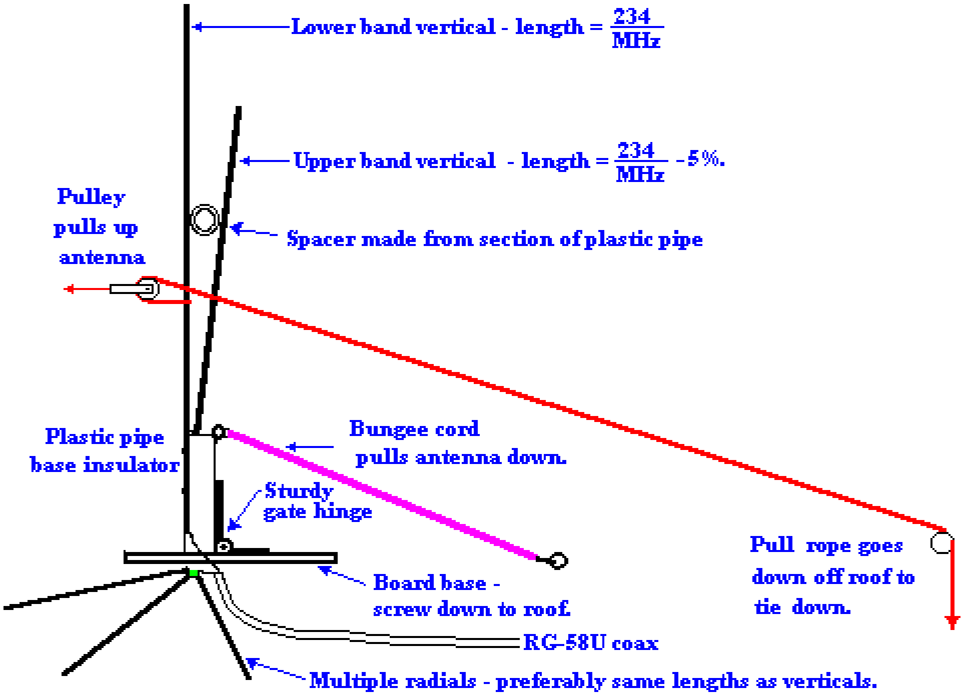
Lower band vertical - length = $\frac{234}{MHz}$
Upper band vertical - length = $\frac{234}{MHz}$ - 5%.
Pulley pulls up antenna
Spacer made from section of plastic pipe
Plastic pipe base insulator
Bungee cord pulls antenna down.
Sturdy gate hinge
Pull rope goes down off roof to tie down.
Board base - screw down to roof.
RG-58U coax
Multiple radials - preferably same lengths as verticals.

A dual band, dual vertical, hinged for lightning protection

I wanted antennas for 10 and 15 meters that would be higher off the ground and less directional than my dipoles. Vertical antennas are simply metal poles ¼ wavelength long sticking straight up. They are really just dipoles standing on end. Therefore, although we talk about the vertical part, we also need the counterpoise portion connected to the coax cable outer shield. The counterpoise consists of "radials" preferably the same length as the vertical extending out along the roof or ground in all directions. In theory, like a dipole, a single radial will work, but in practice multiple radials equal or longer than the vertical work best.

I started with a 15 meter vertical consisting of an 11 foot curtain rod lashed to my chimney. (234 divided by 21.1 MHz) Four thin wire radials fanned out down the roof in four directions. It worked well the first time I loaded it. The fun of using a real curtain rod is that, when you tell someone your antenna is “a vertical made from a curtain rod,” they usually comment on it as if you had built a nuclear reactor out of a washing machine. Interesting psychology. Obviously these guys have never built anything themselves.

Trap antennas

Next I tried to add 10 meter capability by adding a parallel LC resonant trap tuned to 10 meters about 8 feet up the vertical. The idea is that, if a tuned LC resonant circuit is added at the right distance up the antenna, the trap will prevent RF current from traveling up the antenna past the trap. In other words, for higher frequency bands the trap makes the antenna appear shorter than it really is. In this way a long antenna can also work well for a frequency much higher than its length would suggest. After I built a trap resonant for 10 meters, I adjusted the distances above and below the trap dozens of times. But even with an antenna tuner, I could never get it to load well on either band. It turns out that homemade trap antennas are hard to build.

You may have noticed that in an LC resonant circuit, the sizes of the L and the sizes of the C are not fixed. In other words, the LC circuit will be resonant at the desired frequency so long as:

$$\omega = 1 / (LC)^{1/2}$$

or, $$\omega^2 = 1 / (LC)$$

where ω = (2 x π x Frequency)

π = Pi, 3.1416.

L = inductance

and, C = capacitance

(an exponent of ½ means the square root)

$2 \times \pi \times \text{Frequency} = 1 / (LC)^{1/2}$

Unfortunately, the sizes of the L or C do matter in a trap. This is because a capacitor or an inductor shift the phases of the current and voltage and affect the reflections of the waves along the wires. To design a trap properly, the distances above and below the LC as well as the sizes of the L and the C all have to work together to make the whole apparatus look like a 50 ohm resistor as seen by a radio wave. Good luck!

It turns out that this is a problem for Smith charts. These are a way of mapping out inductive reactances, that is, current phasing, so that they cancel and the antenna will look resistive. I haven't figured out a practical way to do this and the handbooks don't help, unless you copy their trap antenna designs precisely. In the end, I just applied the same dual-dipole principle as the multi-band dipoles discussed earlier. I added a second vertical to the first, separated by about 5^o degrees and about 5% shorter than it would have been if it were standing alone. I added two more radial wires cut for 10 meters, but I'm not convinced they were essential. Once again, it worked perfectly on both bands the first time I loaded it.

Caution: ***Remember that "the vertical antenna" begins at the exact inch where the coax inner conductor first separates from the outer braid.*** A common mistake is to ignore an extra 6 inches of inner conductor that may be adding to the effective length of the driven element. The dual band vertical has worked beautifully. I've also used it on 12 meters successfully. Much to my surprise, it also worked on 17 meters, although the SWR (Standing Wave Ratio) isn't great. In other words, when there are standing waves on the antenna because the length is wrong, it takes extra voltage to drive it. I was probably losing some power on 12 and 17 meters.

Two years later, 10 and 15 meters were no long open but 17 meters was still active. To make the 15 meter vertical work well on 17, I tacked on an additional 1.9 feet in length and added some 16 foot long radial wires which are suitable for 20 meters. After this modification the antenna loaded easily on 17 meters and I got good signal reports. I can summarize verticals by saying that all you need to do is calculate the length in feet using the formula 234 / frequency and it will work well the first time you try it.

Kaboom! What about lightning?

Looking up at the metal poles on my roof, they looked more like lightning rods than curtain rods. I'm probably more paranoid than necessary, but I've had bad experiences with lightning. When I was a novice, fuses blew in my rig during a thunderstorm. A buddy of mine had his oscilloscope fried and a hole blown in his roof. On mountaintops I've felt my hair stand on end and heard my ice axe hiss as we galloped down the mountain in terror. Once I had the grim duty of carrying three deceased lightning victims down from Arapaho Glacier. Lightning scares the heck out of me and that vertical looked like a disaster waiting to happen.

My solution was the base hinge and bungee cord shown above. During the summer I raise the antenna by tugging on the rope whenever I need it, then lower the vertical when I'm done. The hinge also has stealth advantages. One important point is that the rope pulls the antenna ***UP***, while the bungee makes sure it folds back ***DOWN***. I tried to have the bungee pull it up, but that worked poorly. There was too much torque on the vertical and the bungee wouldn't let the antenna fold flat against the roof.

How lightning forms and how likely it is to occur

According the National Lightning Safety Institute of Louisville, Colorado, lightning begins with "leaders" of corona coming down from the clouds. These leaders of charged air produce high voltage with respect to the ground. In response "streamers" of charge rise up from the ground and are most easily launched skyward by sharp conductive objects. When the streamers meet the leaders the ion path from cloud to ground is complete and the current flow enters a brief "spark phase" when small currents begin to flow. This can easily transition into "arc phase" in which the air becomes fully ionized and large currents are free to flow. Once the path is complete, a lightning bolt begins. Lightning can be postponed or even prevented if the streamer paths can be suppressed and streamers don't merge with the leaders. ***Fortunately, except for the highest peaks and tall metal towers, lightning strikes are rare and capricious.*** Your best bet for safety, (but not for DX), is to keep your antennas no higher than the local trees and buildings.

Fuzzy metal brushes on an antenna are controversial

A relatively new approach to lightning prevention is to bleed the excess local charge off a vulnerable metal tower with metal brushes before it can attract a full-fledged lightning bolt. In contrast, ***the National Lightning Safety Institute (www.lightningsafety.com) strongly condemns using lightning rods of any sort.*** They say that lightning rods are allowing the streamer to begin earlier and are therefore attracting lightning. The fact that this is controversial demonstrates that lightning really is capricious. After 200 years of research it still isn't obvious whether lightning rods are good or bad! This tells us that, statistically speaking, lightning is not a big risk.

The Lightning Safety Institute had a research project up on a bare, rocky ridge high above timberline at 11,500 feet above sea level near Nederland, Colorado. The location looks like a good place to avoid during a lightning storm. I have been up there three times and examined their array of various lightning rods. They have pointed ones, square ones, and round ones. The rods are connected to ground through fuses. The first time I was up there, they were using 20 ampere fuses but none of them had blown. The next year they were using 2 ampere fuses and still none of them had blown. That must have been frustrating!

On the other hand I have encountered two 60 foot high spruce trees that were struck by lightning. The trees were literally exploded into giant chunks of log weighing hundreds of pounds. These trees were located among thousands of other 60 foot high trees in relatively flat high altitude forests. What was special about those trees? Apparently nothing. Those trees were just unlucky. In summary, ***lightning is extremely random and difficult to predict.***

The 2005 ARRL Handbook doesn't discuss antennas as lightning rods. Instead ARRL stresses grounding antenna towers and placing lightning arrestors on lead-in coax and wires. They don't even mention using spiky points on antennas. ARRL emphasizes that ***it is difficult or impossible for a ham to protect from DIRECT strikes.*** This is because the spike of current produced by a lightning bolt is so huge and so sudden that, lightning not only delivers thousands of amperes, the abrupt rise in current means that any wire or metal conductor we could install will have too much inductance to ground to shunt the whole pulse away. The current pulse and big voltages will divide and jump to other objects. The best we can do is protect our house and equipment from surges due to *nearby* strikes. For example, instead of having tall, grounded spiky metal poles projecting from the roof ridge, it would be better to simply put a grounded wire along the ridge of the roof.

ARRL recommends surge protectors on power cords to prevent damage to line powered equipment. Battery powered ham stations obviously don't need surge protectors and battery-power has other advantages as described in chapter 8.

In summary, unless you live on a mountaintop or have a huge metal tower, lightning is not an imminent threat and you don't have to be paranoid. If you're like me, your antennas are no higher than the local power lines and trees. I haven't had any incidents since 1960. However, I still crank my vertical antenna down in the summer whenever I'm not on the air and I unplug my antennas from my rig.

CRYSTAL SETS TO SIDEBAND
© Frank W. Harris 2010, REV 12

Chapter 6

BUILDING A HOMEBREW QRP

Among the guys I work, QRPs seem to be the most common homebrew project, second only to building antennas. Therefore this chapter describes a simple QRP design I have settled on. I use my QRPs as stand-alone transmitters or I use them to drive a final amplifier to produce higher power, 25 to 100 watts.

It's true that before you build a transmitter you'll need a receiver. Unfortunately, a good selective, all-band ham receiver is complicated to build and most guys don't have the time and enthusiasm to do it. (See chapter 13.) On the other hand, chapter 7 describes building a simple, 5-transistor 40 meter receiver that I have used with the QRP below to talk to other hams. This simple receiver will work best during off hours when 40 meters isn't crowded. It can also be used to receive Morse code for code practice.

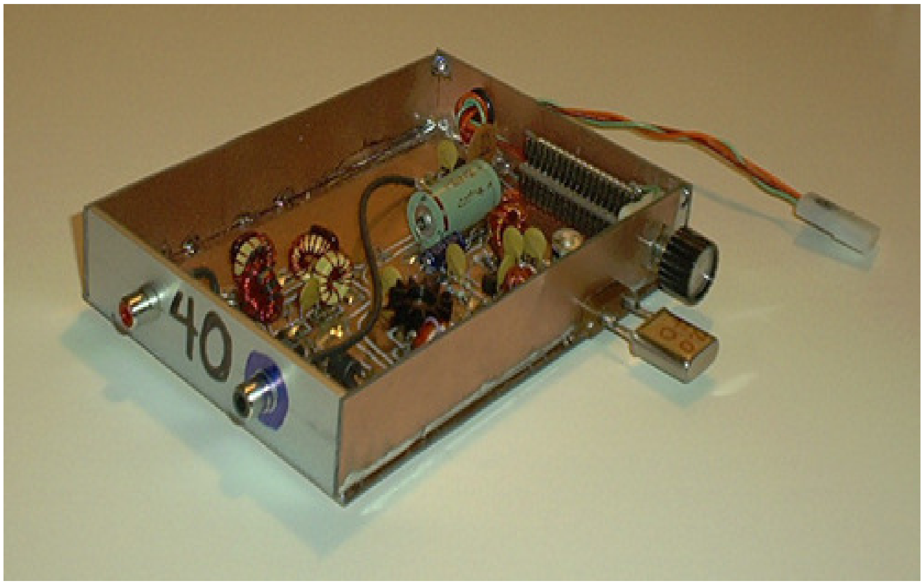

A 40 meter QRP module.

The QRP transmitter shown on the left is designed exclusively for 40 meters, (7.000 to 7.300 MHz.) The twelve volt power supply comes in through the pig-tail wire on the right. The telegraph key plugs into the blue-marked phono-plug socket on the right of the aluminum heat sink. The antenna output is the red-colored socket on the left end of the heat sink.

The transmitting frequency of the QRP module is controlled by a quartz crystal. That's the silver rectangular can plugged into the box on the right front. The knob on the far right is a variable capacitor for adjusting or "pulling" the crystal frequency over a range of about 3 KHz. As we shall explain in chapter 10, a variable frequency VFO that can cover the entire band is hard to build. I don't recommend starting out with a VFO. You need success, not frustration.

Start simple

The easiest QRP is just a crystal-controlled oscillator. It usually has one to three amplifier stages to raise the power to between 1/2 to 5 watts. This will get you on the air quickly.

Unfortunately crystal control means you must order $18 custom crystals for your favorite frequencies in each band. Because the oscillator only has a very narrow tuning range, you can usually only call CQ (calling any station). As luck would have it, it always seems as though the other fellow calling CQ is just out of your crystal range.

The transmitter "mainframe"

You could use this bare-board little transmitter just as shown above. However, it will be more convenient to mount it in a larger chassis equipped with an ON/OFF switch, pilot lights, an antenna relay, and other niceties. My whole transmitter assembly is shown below.

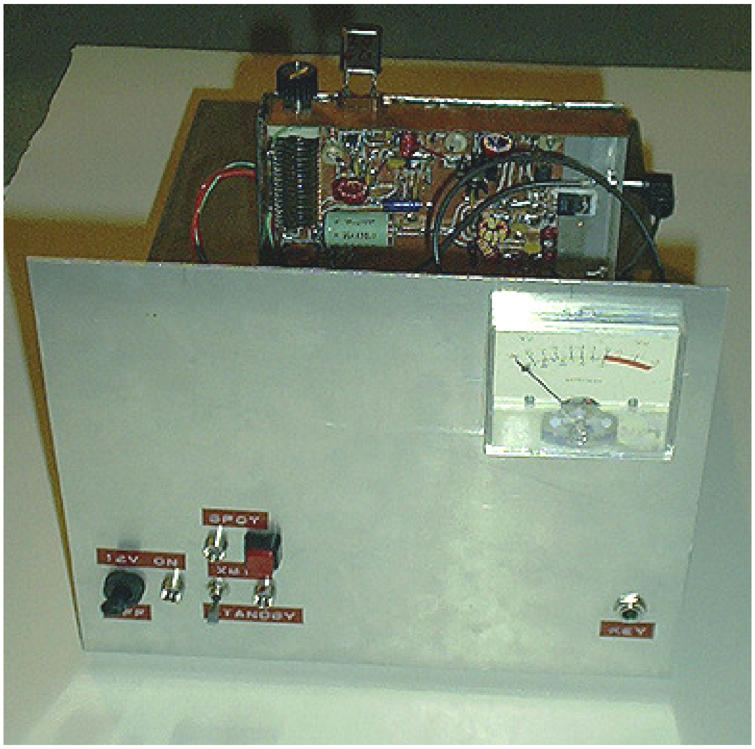

Transmitter mainframe

The transmitter "mainframe" is a box to contain your transmitter modules. The large meter at the upper right indicates the current drawn by the final amplifier. It's helpful to know how much power the transmitter is consuming. The switches and lights on the lower left are the main power switch and the transmit/ receive switches. There are also LED lights to indicate which switches are active. The large red button is a "spot" switch. With the spot switch engaged, the QRP oscillator may be tuned to another station without actually transmitting. How big you make your mainframe depends on your plans and ambitions. As you can see, I had *really* big plans. By the time you add multi-band capability, a VFO, power supplies, and single-sideband capability, even a large box may be too small. My previous mainframe was too small so I made this one extra large to eliminate crowding.

HF construction methods – building your own circuit boards

Back in the vacuum tube days we built hand-wired HF transmitters. Inside they had long skinny wires running every which way. However, transistors generate high power with low voltage and big currents. As a result, the inductance of the wiring must be kept as low as possible. This means you must build with printed circuit boards or the transmitter simply won't work.

For example, once I connected an output lead from a 21 MHz (15 meter band) transistor power amplifier to the base of the next stage with a bare wire 4 inches long. If this had been a vacuum tube circuit, this wire coupling would have worked well. But with high current transistors, the wire acted like an RF choke. That is, the wire blocked current flow as if it were an inductor. Sliding an oscilloscope probe along the wire, I could see an 80% drop in drive voltage from the output of the first amplifier to the input of the next stage. In contrast, a wide circuit trace printed on a circuit board acts like a coaxial cable and has very low inductance. With a PC board the same drive to the next stage can be nearly 100%, not 20%.

If you have already developed your own method of making printed circuit boards, you may skip ahead. But if you have never done this before, read on.

MAKING PROTOTYPE RF CIRCUIT BOARDS

PC board tools and materials

I usually use double-sided printed circuit (PC) board material such as Radio Shack part # 276-1499A. This board has layers of thin copper bonded to both sides of a 1/16th inch fiberglass sheet. I only solder components onto the surface of one side. The solid surface of grounded copper on the opposite side provides distributed capacitance over the whole circuit. I believe this gives extra stability and works as a grounded shield to prevent radiating signals to whatever circuits might be adjacent to the bottom of the board.

Traces on a PC board act like transmission lines

A transmission line, such as the coaxial cable that delivers your TV signals, is remarkably efficient. As explained in chapter 4, the wire in the center of the coax acts like a distributed inductance that rings with the distributed capacitance between the center wire and the outside insulation. In other words, a coaxial cable is a distributed LC circuit and does not dissipate voltage and energy as you would otherwise expect. In fact, the energy losses that do occur are only the heating losses in the insulation and the simple resistance losses that happen in copper wire.

A two-sided circuit board can be designed to act like a circuit wired with pieces of coax. The circuitry is mounted on one side of the board while the other side is ground. Each trace has a small amount of inductance and each trace has a proportional amount of capacitance with the grounded copper sheet on the far side of the board. The result is that signals are almost unchanged as they travel along the traces of a circuit board. It would be interesting to build the exact same RF power amplifier on single-sided and double-sided boards and compare the difference in performance at HF frequencies. I'm convinced double-sided is far better, but I've never done a controlled experiment.

In microwave printed circuit boards the traces are deliberately designed to act like transmission lines for the particular frequency that is being used. The precise, optimum dimensions for these "microstrip" traces can be calculated for maximum performance. Fortunately for us, high frequency circuit boards like ours don't have to be designed with so much care.

Etched circuit boards

There are at least four ways to make homebrew RF boards. Originally I tried etching "real" printed circuit boards with ferric-chloride solution. It's slow, messy, and lots of work. Etching boards is an art that isn't easy to master. You may under-etch or you may over-etch, often on the same board! If you succeed, you'll have a result close to a commercial product.

Unfortunately, when you use a circuit layout from an ARRL handbook or QEX magazine, you are making the assumption that you can buy all the exact parts the author used. Good luck! The parts you can actually buy may not fit on the board. I found that the worst limitation of etched PC boards was that I couldn't build and test my circuits one transistor stage at a time. If I needed to add another component or add another amplifier stage, I had no room for it. ***The key to success in homebuilding is building AND TESTING one small circuit at a time.***

Gouged PC boards

My favorite method for making RF boards is carving them into a bare, two-sided PC board with a small wood-carving gouge. The gouge is a chisel with a cupped end, about 0.5 to 3 mm across. I have several, but my favorite is about 1 mm across. I got it from the Traditional Woodworker Co., (www.traditionalwoodworker.com)

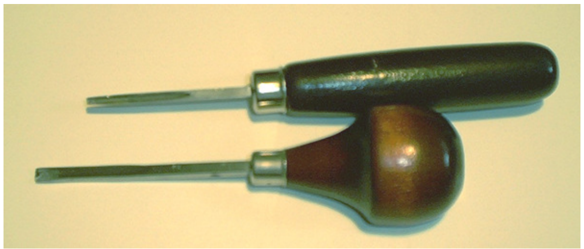

Two small wood-carving gouges

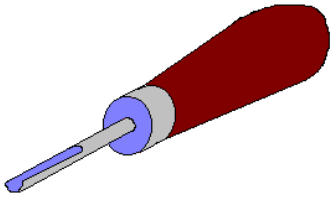

Notice the cupped end of the wood-carving gouge

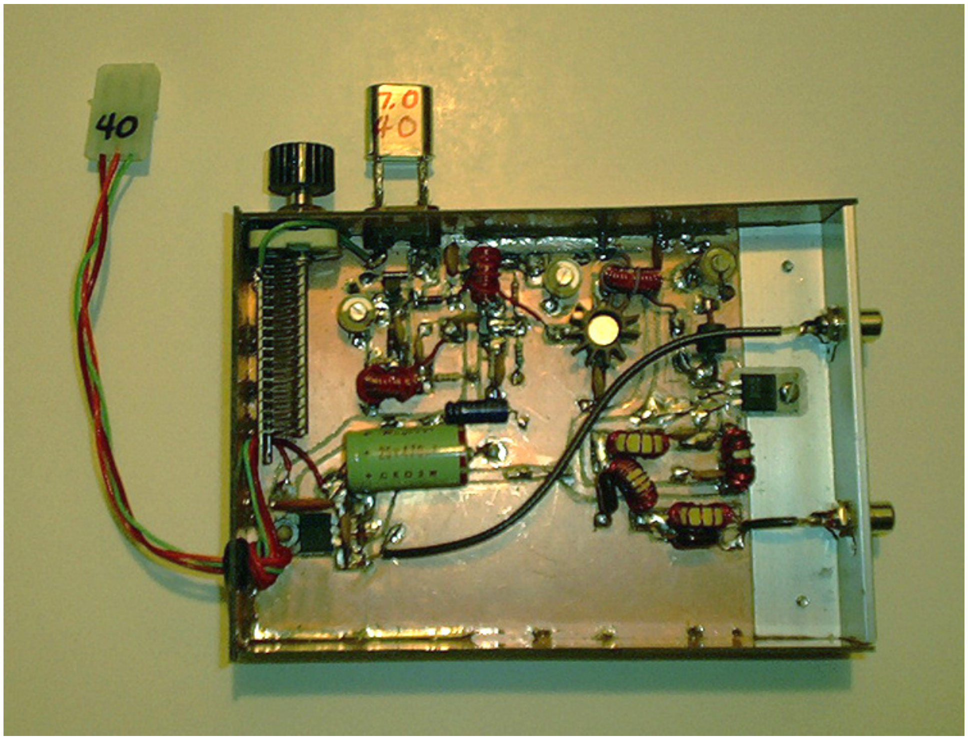

This QRP board was made with "gouged board" construction. Nearly all of the visible copper is "ground." The aluminum angle on the right serves as a heat sink for the output transistor.

When you push the gouge along the board at a high angle and twist your wrist back and forth, the gouge carves out a little trench through the copper. Hint: Keep your elbow high and the gouge won't skate across the surface and into your hand. Ouch! Obviously two "trenches" can isolate a strip of copper that serves as wire or "trace." Often I just isolate little islands of PC trace onto which I solder short component leads. I usually solder components to the surface, rather than drilling mounting holes for each lead.

PC boards for integrated circuits

A wood-carving gouge works fine for RF circuits made from discrete transistors, but is impractical for integrated circuits. For through-hole ICs I use 0.1 inch center punch-boards and wire-wrap methods. There is an example of this construction method in chapter 7. Another method of working with integrated circuits is to etch (or buy) tiny boards for mounting each type of IC. That is, you make 8 pin boards, 14 or 16 pin boards, etc. These little boards have big pads connecting each pin so that you can solder jumpers down onto your relatively crude handmade board. Isolated little IC boards are especially good for working with "surface mount" ICs which are otherwise too tiny to handle.

Coax jumpers

If you need to, it is still possible to run long wires from point to point. But instead of wire, you must use shielded coax. Sometimes I "pipe" an RF signal from one end of a PC board to the other without significant loss of voltage or power. If there's no room for a wide trace, I use a piece of skinny RG-174 coax transmission line for this purpose.

Another application for a coax jumper is to shield an audio or DC signal from the RF. If you run a bare wire across a PC board full of RF currents, the low frequency signal will be contaminated with RF signals by the time it arrives at the other end. For example, in the QRP board above, my DC (Morse code) keying signal is transported across the board with a length of coax. Because of the way my mainframe transmitter chassis was wired, my QRP board plugs into the mainframe with the telegraph key input at the wrong end of the board. The telegraph key must switch the 12 volt power MOSFET transistor on an off, but unfortunately it was located on the opposite end of the board from the key input. To shield this low power DC line, I used a coax jumper as shown in the photo.

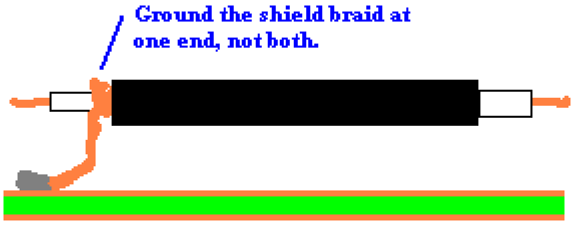

COAX JUMPER CABLE

Just ground the coaxial shield at one end. The most common error when using coax jumpers on a board is to ground both ends of the shield. This introduces a "current loop" that can act something like a turn of a winding on a transformer. This current loop may pick up stray currents and the coax might do more harm than good.

Disk and Super-glue boards

Another practical RF PC construction method is the "disk and Super-glue" method. Small disks or strips are punched or cut out of PC board material. The disks and strips are glued onto the PC board using superglue to form contact nodes and traces. An advantage of this method over gouged board is that the ground sheet is unbroken. That is, when you solder a component lead to ground, you will know that the entire PC board is "ground," and not just areas or strips of "ground." A continuous sheet of copper means that there will be virtually no voltage differences between one region of the board and another. The higher the frequency or the higher the power levels, the more important this advantage becomes.

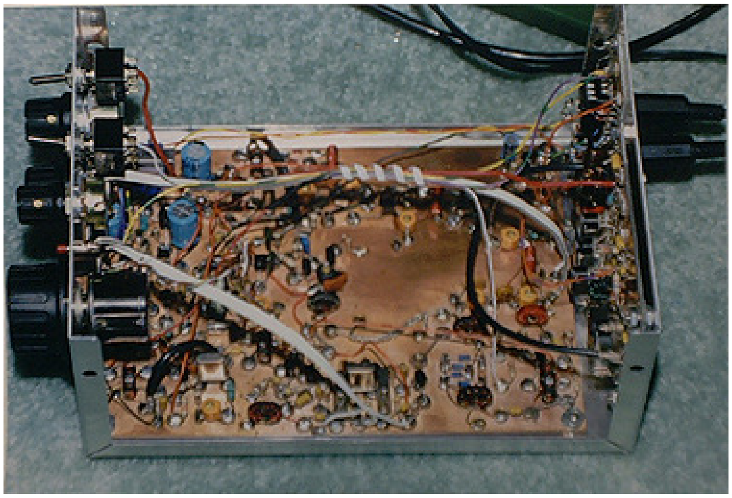

Mike Fitzgibbon, NØMF, built the homebrew QRP transceiver shown on the left using the super-glue technique. I met Mike on the air while he was using this transceiver. The transceiver puts out about a watt and uses a super-regenerative receiver. We communicated easily across 1,400 miles so apparently it works very well.

Dead Bug construction

"Dead bug" construction resembles super-glue boards in that the PC board is used as an unbroken sheet of grounded copper metal. Traces are not cut or etched into metal. Instead of using disks glued to the board, the ungrounded component leads are soldered together up above the board as needed. Since the ICs and transistors are upside down, they resemble "deceased insects" with their feet in the air. Of course you must keep the leads as short as possible. When currents are low, lead inductance isn't a problem. For the low power stages (less than 100 milliwatts) I have found "dead bug" construction OK, but I don't recommend it for RF power amplifiers.

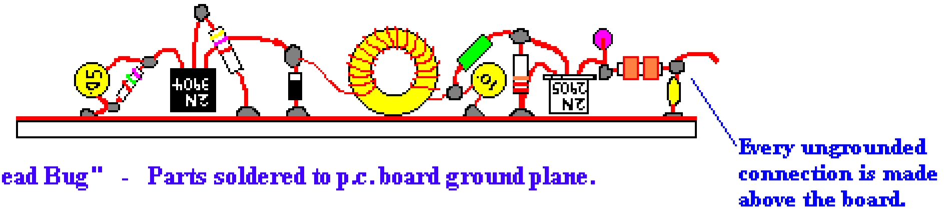

"Dead Bug" - Parts soldered to p.c. board ground plane.

Every ungrounded connection is made above the board.

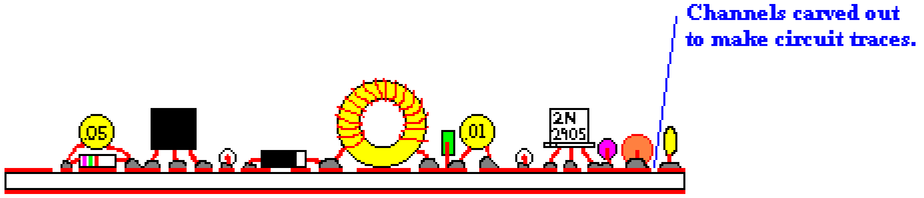

Channels carved out to make circuit traces.

"Gouged Board" - Parts soldered to islands and traces on 2-sided p.c. board.

"Dead Bug" versus "Gouged Board" construction.

When you get into “high power,” greater than 100 milliwatts, you should minimize the lead inductance by using wide traces on a PC board. Once I built a 10 meter power gain amplifier stage using extremely compact dead-bug construction. It worked well and gave me about 0.5 watts output. Since the stage was a success, I rebuilt it using gouged board construction. Using carved traces, the same circuit and components gave me 1.0 watt, a 100% improvement. When I have to connect an RF power lead across the board for a half an inch or more, I cut a wide trace on the board or I use a short piece of coax. For big, 10 ampere RF currents in a power amplifier my traces might be ½ inch wide or more. If I’m short of space, sometimes I use a wide rectangle of PC board material standing on edge as a wide, low inductance “wire.” The antenna relay shown in chapter 9 uses this technique for the transmit circuit.

Boxes out of PC boards

It’s often necessary to shield circuit boards from other circuits. That is, the circuit needs to be completely enclosed in a metal box so that it does not radiate or receive radio signals to or from nearby circuit boards. A gouged board circuit can be shielded by building up the edges into a box. Solder inch-high strips of PC board material around the periphery of the board. Since the copper surface solders nicely, it’s easy to build up a board into a sturdy open-top box. And because a gouge–board doesn’t need to have through-hole components, the circuit inside is completely isolated from the grounded outside of the box and is enclosed on five sides.

To make a lid for the box, fold a flat piece of thin aluminum or copper sheet into a mating shallow box that fits over the top. The top just needs to be snug enough to hold it in place. This technique is essential for receivers and SSB transmitters as illustrated in chapters 13 and 15.

****

The complete crystal controlled QRP transmitter

The block diagram below shows the basic modules of the completed QRP transmitter from battery to antenna.

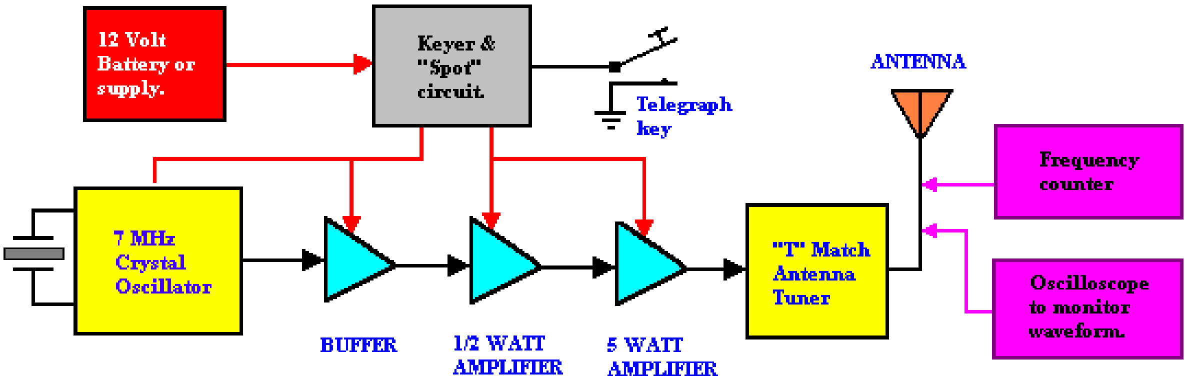

40 Meter 5 Watt QRP Transmitter Block Diagram

The bad news is that there are lots of circuit blocks. Two of the modules on the right, the oscilloscope and frequency counter, are test instruments you need to be sure you are operating on the correct band. These are items you will almost certainly buy. The power supply that I prefer is a simple car battery or a 12 volt deep-discharge marine battery. It's desirable, but not essential, to regulate the battery power to about 11 volts. In this way, the transmitter always has the same supply voltage, whether the battery is being charged or nearly dead. You can also build a power supply that plugs into the wall or buy a supply. Building a ten-watt, 12 volt AC plug in power supply for your QRP is fairly easy. But later, when you get into high power amplifiers needing 100 or 200 watts, building high power AC line-powered supplies is much harder and you may want to return to a battery. Power supplies are discussed in chapter 8.

The oscillator, buffer, and amplifiers will be discussed next. The keyer circuit is discussed at the end of this chapter. A receiver is not shown above and is assumed to use a separate antenna. If you hook your antenna to the transmitter and receiver simultaneously, the receiver is likely to be damaged. Suitable antennas are discussed in chapter 5. Accessories like the telegraph key and the T-match antenna tuner are discussed in chapter 9.

Amplifiers and oscillators

In order to generate a radio signal, we first need to make an oscillator. A spark gap generates radiowaves by the electronic equivalent of banging on a bell. On the modern ham bands we need continuous, pure sinewaves. Think of it as the radio equivalent of an electronic organ. You hold down a key and a pure tone never stops. In fact, that's the origin of the term "continuous wave" or CW for radiotelegraphy. You may have already asked yourself, "If I'm pounding on a telegraph key, what's continuous about that?"

Continuous waves at Megahertz frequencies are most easily generated using electronic oscillators. A continuous wave oscillator is an amplifier that amplifies feedback from its own output. Before we can understand an oscillator, we need to cover bipolar transistor amplifiers.

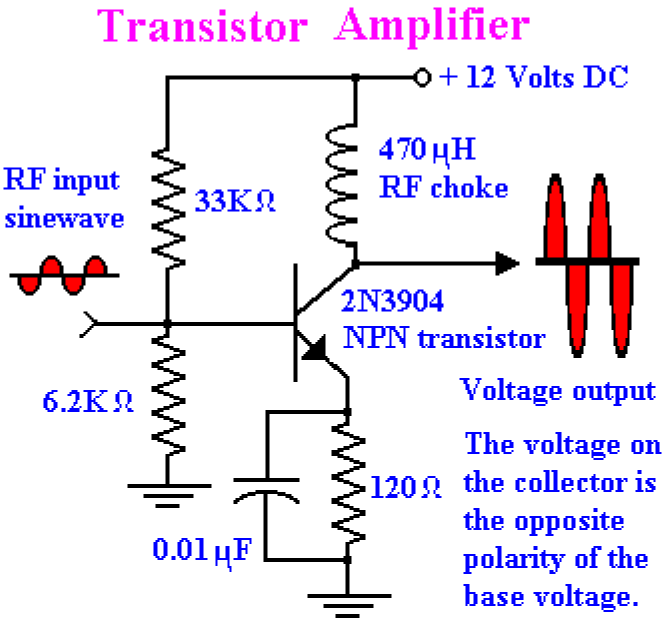

The voltage on the collector is the opposite polarity of the base voltage.

What is a linear amplifier?

The circuit at the left is a simple transistor RF (radio frequency) amplifier made from an NPN transistor. If you prefer, you could reverse all the polarities and use a PNP transistor and it will work just as well. The above amplifier is designed to be ***linear***. Linear means that it can amplify big signals or tiny signals equally well over a wide range of frequencies. In other words, although designed for radio frequencies, this amplifier is "hi-fidelity" and is analogous to the audio frequency amplifiers found in your stereo. Another name for a linear amplifier like this is ***Class A amplifier.***

As we saw in chapter 4, a bipolar transistor can be thought of as a structure made by merging two PN junction diodes. When a silicon diode is forward biased, (positive to P conducts) relatively large currents, (milliamperes), can pass through the diode. However, current will flow only when the forward offset voltage, (about 0.6 volts) is exceeded. Just like a diode, when a silicon NPN transistor is used, nothing significant will happen until the base voltage rises above positive 0.6 volt. When this occurs, a relatively small base current will “convert” the tiny region of P type semiconductor into a conductor. Big currents are then free to flow from collector to emitter. In the 2N3904 transistor used above, the transistor “gain” causes a big current to flow that is 100 times larger than the base current.

Suppose that we want to amplify a sinewave that has both positive and negative polarities. A transistor amplifier with a minimum of parts would be just a base connected to the input. That amplifier would only amplify the upper part of the positive half of the sinewave, above 0.6 volts. All the rest of the sinewave would be below the turn-on threshold. The purpose of the 33K ohm resistor in the above circuit is to turn the transistor “halfway on,” like a faucet. Now when a sinewave current comes in on the base, the negative portion of the sinewave will turn the transistor *LESS ON*. And when the positive half of the sinewave arrives, it will turn the transistor *MORE ON*.

What does the 470 microhenry inductor “choke” do?

The choke produces a voltage output without wasting energy. We could use a resistor instead, but a resistor would waste energy and become hot. This amplifier is designed to generate a big RF voltage. The output is the voltage that rides on the transistor collector. The current flowing through the inductor isn’t really doing anything useful besides dropping the voltage between the 12 volts supply and the collector. Therefore, we use a large inductor relative to ham band frequencies. RF cannot pass through the choke, but DC current can. Remember that ***the current through an inductor cannot change instantly.*** The inductance is high enough so that, during each half-sinewave cycle, the inductor is too large to charge significantly with current. Consequently, a big RF voltage appears on the collector and little energy is wasted.

Stabilizing the transistor operating point

The **33K ohm resistor** is the component that turns the transistor “half–on.” It ***biases*** the transistor on by injecting a small current into the base. The **120 ohm resistor** serves to add a small amount of negative feedback so that the transistor will not turn on too hard when the transistor becomes warm. As big DC currents flow through the 120 ohm resistor, a DC voltage will appear across it. This voltage “raises” the emitter voltage off ground, thereby decreasing the base to emitter voltage. When this voltage difference drops, the current flowing into the base decreases, thereby turning the transistor a bit less on. So long as this resistor isn’t too large, it just serves to bias the transistor and make it more stable with temperature changes.

If you replace the 120 ohm resistor with a short circuit, you’ll find that the amplifier still works, but the transistor will run hot to the touch. The **6.2K ohm resistor** makes sure that the charge in the base of the transistor always has a way to leave and turn off the transistor. The 6.2K resistor also stabilizes the operating point and insures that the circuit will work every time you build it.

Bypass capacitors

So what's the 0.01 microfarad capacitor across the emitter resistor for? This is a ***bypass*** capacitor. An RF amplifier can be thought of as two circuits superimposed on each other. One circuit establishes the static DC voltages and currents needed for stable, linear operation. The other circuit handles the transient RF sinewaves that modulate the static voltages and currents. As explained above, DC current passing through the 120 ohm resistor causes a DC voltage drop from emitter to ground. Radio frequency current passing through the resistor would also cause an RF sinewave to appear on the emitter. This voltage would be subtracted from the voltage on the collector and therefore would decrease the output. You may remember that ***the voltage across a capacitor can't change instantly.*** The bypass capacitance is chosen so that, although it will charge up to some DC voltage, at ham band frequencies the capacitor voltage will not change significantly. For a capacitor this large, the fractions of a microsecond that the RF voltage oscillates up and down are not important. The result is that emitter will have zero RF voltage on it and all the RF voltage will appear on the collector. Another way to look at bypass capacitors is that they "shunt" RF to ground, but don't affect the DC.

How an amplifier becomes an oscillator

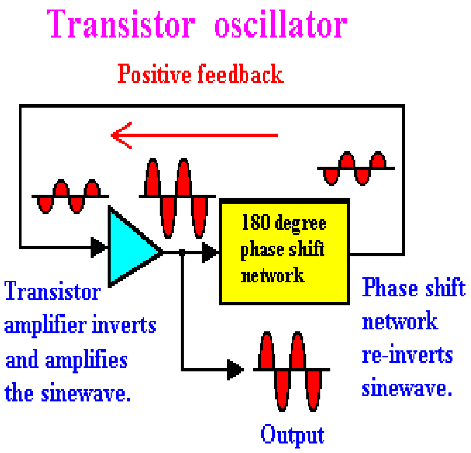
Transistor oscillator
Positive feedback
180 degree phase shift network
Transistor amplifier inverts and amplifies the sinewave.
Phase shift network re-inverts sinewave.
Output

You already know that a public address system turns into an audio frequency oscillator when you place the microphone close to the loudspeaker. This results in an ear- splitting, screaming oscillation. Electronic oscillators work by feeding the output back into the input.

Notice that a one-stage transistor amplifier inverts the polarity of the input waveform. Rising voltage on the base causes the voltage on the collector to fall. In theory at least, if you were to feed the output of this amplifier back to the base circuit, the opposite polarity of the output would cancel out any signal that begins to form on the input. For this reason, if you wish to make an oscillator, the waveform needs to be turned up side down or ***phase inverted*** before the feedback is introduced to the transistor base.

The need for a phase inverter circuit gives us the opportunity to use this inverter circuit as a filter that also restricts the oscillation to just one frequency. You have already met parallel L-C resonant circuits. A series L-C circuit will attenuate every frequency except its resonant frequency. A series L-C circuit located between the collector and base of a transistor amplifier will oscillate on a specific frequency determined by the series inductance and capacitance.

Series LC resonant circuit

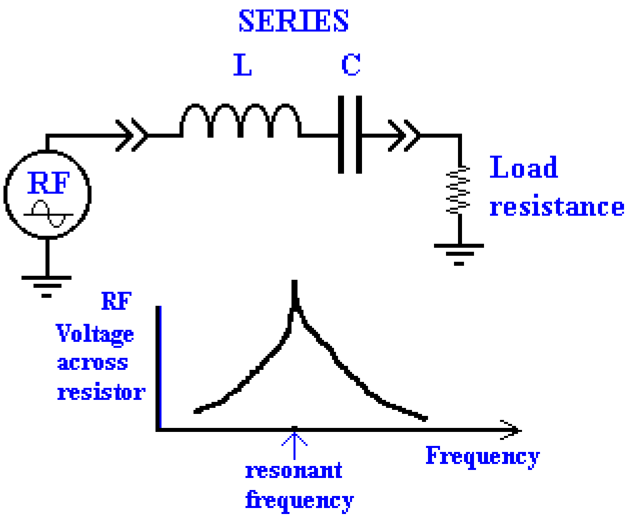
SERIES
L
C
RF
Load resistance
RF Voltage across resistor
resonant frequency
Frequency

In chapter 4 you were introduced to the parallel resonant circuit. The series resonant circuit also resonates at a specific frequency and can be used in much the same way. The parallel resonant circuit appears as an ***infinite resistance*** or ***infinite "impedance***" at a specific frequency while shorting out RF voltage at all non-resonant frequencies applied across it. In contrast, the series resonant circuit looks like ***zero resistance*** at the resonate frequency. As shown above, it only passes one RF frequency current efficiently to the load resistor. To all other frequencies, it appears as a large inductance or as a tiny capacitance.

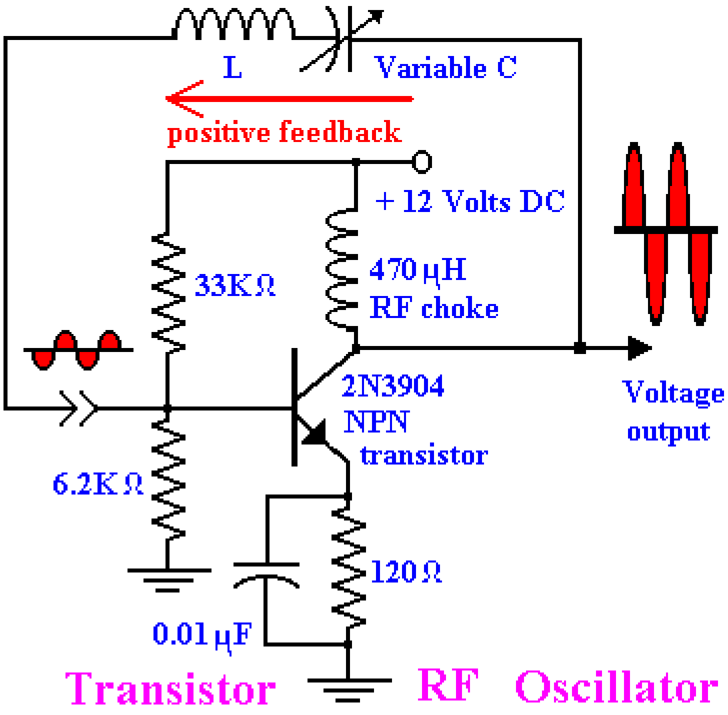
L
Variable C
positive feedback
+ 12 Volts DC
33KΩ
470 μH
RF choke
2N3904
NPN
transistor
Voltage output
6.2K Ω
120 Ω
0.01 μF
Transistor
RF Oscillator

A working transistor oscillator

The circuit on the left oscillates in the range of 1 to 30 MHz, depending on the sizes of the inductor L and capacitor C. Unfortunately, if you build it, you'll find this oscillator too unstable for use in ham radio. However, it does produce a strong oscillation. Once you get it running, you can illustrate the importance of the 33K forward bias resistor. When you disconnect this resistor, the oscillator will continue to oscillate as though nothing happened. Well, almost nothing. If you look closely on your oscilloscope, you'll see that the bottom loops of the output sinewave are severely distorted. Without the 33K, the transistor turns off whenever the input drops below 0.6 volts.

Now turn off the power supply for a moment. Turn the supply back on and the oscillator will be "dead." The output will be just a straight line on your oscilloscope. Without this 33K forward bias, the amplifier cannot "see" its own tiny random noise output and the oscillation never begins. That is, the random noise is way below the 0.6 volt threshold. Restore the 33K and the oscillation will restart immediately.

A non-linear amplifier with no forward bias is called a ***Class C amplifier***. They are useful for amplifying CW signals that are greater than the forward base drop. That is, the drive for a Class C amplifier must be greater than 0.6 volt, otherwise there will be no output. Class C is not useful for amplifying voice or music signals because the 0.6 volt threshold cuts off much of the waveform and distorts it severely. The good news about class C amplifiers is that they run at higher energy efficiency than a class A. 65% versus 30% is typical.

Quartz crystal oscillators - the key to modern frequency stability

Electronic wrist watches today use quartz crystals as the frequency standard that gives cheap everyday watches precision that rivals the old time mechanical chronometers that once were vital for navigation. Analog color TVs use crystals as a time standard so that the color separation can accurately separate red, blue, and green. Computers use crystals as clock standards so that modems will run accurately and remain in step with the sending station.

Quartz is a type of natural, crystalline, silicon dioxide glass. It is an extremely good insulator. Household glass is also mostly silicon dioxide, but the atoms are arranged in an amorphous structure, something like a random pile of bricks dumped from a truck. In contrast, quartz has a regular crystalline structure as though the bricks were neatly stacked in layers. The quartz is not perfectly pure, but contains ionized, contaminant atoms that are trapped in the crystal lattice. Notice that contaminant ions ***do NOT*** convert the quartz into a semi-conductor. Natural quartz from Brazil and a few other places occurs as large, clear crystals with sharp flat faces and uniform structure. This material can be cut and polished into thin, flat sheets that are cut up into tiny squares. Each square is then mounted between two metal plates as if they were building a simple capacitor with quartz glass as the insulator. Since WWII quartz for crystals is manufactured artificially in ovens.

When voltage is applied across the quartz, the charged metal ions contaminating the quartz are physically attracted to the charged plates as shown below. The thin quartz sheet literally bends back and forth as the voltage changes polarity. If the voltage is an RF voltage, the polarity changes back and forth millions of times per second and the crystal vibrates at that frequency.

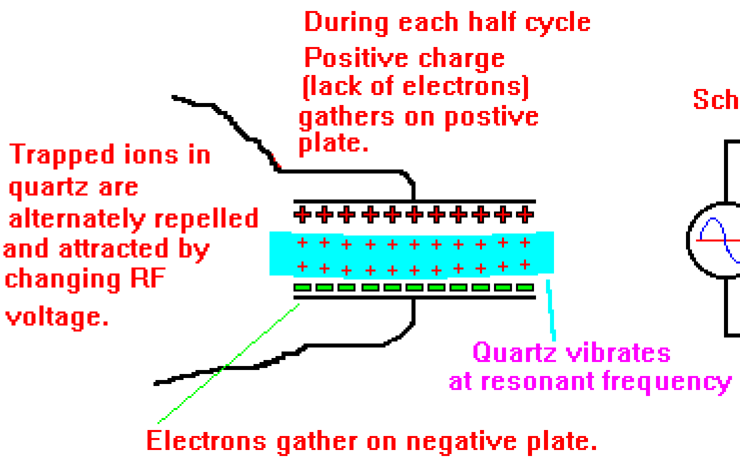
During each half cycle
Positive charge
(lack of electrons)
gathers on postive
plate.
Trapped ions in
quartz are
alternately repelled
and attracted by
changing RF
voltage.
Quartz vibrates
at resonant frequency
Electrons gather on negative plate.

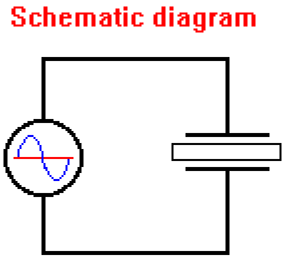
Schematic diagram

When AC voltage is applied to a crystal, it vibrates most easily at a frequency proportional to the mass and physical dimensions of the quartz. In other words, each crystal can be fashioned to vibrate at a specific frequency. As you would expect, big thick crystals vibrate at low frequencies. Thin tiny quartz chips vibrate at high frequencies. Even though crystals oscillate at millions of vibrations per second, the vibration is mechanical and they resemble musical instruments. Just like musical instruments, crystals also have overtones or higher harmonic frequencies. Some crystals are designed to operate at the 3rd or 5th overtones, rather than the fundamental frequency. Overtone crystals can oscillate at frequencies as high as 100 MHz or more.

Although the oscillation is mechanical, it also acts as though it were oscillating electrically. Quartz crystals act like a series L-C circuit. They are equivalent to the L-C components we used to tune the crude oscillator above. The equivalent capacitance Cx and Lx are completely dependent on the physical dimensions of the crystal wafer. Think of a quartz crystal as an extremely stable, L-C circuit. Since the crystal is constructed from two metal plates applied across the quartz, one can also say that the crystal has an element of parallel resonance as well as series resonance. The resistor Rs represents mechanical heating.

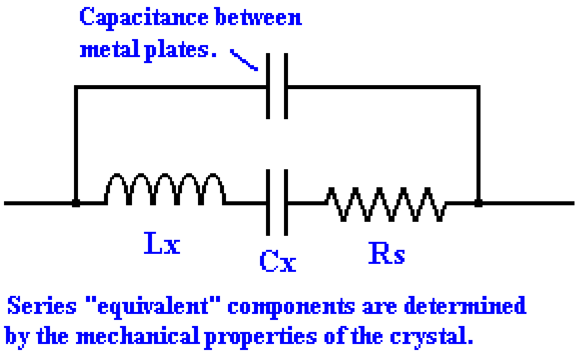
Capacitance between
metal plates.
Lx
Cx
Rs
Series "equivalent" components are determined
by the mechanical properties of the crystal.

Equivalent circuit of a crystal.

Typical crystals are shown below. Two large ancient, ham crystals are shown at the upper right. These can be good parts, provided they are still working after 50 or more years. Sometimes dead, ancient crystals in FT-243 holders can be taken

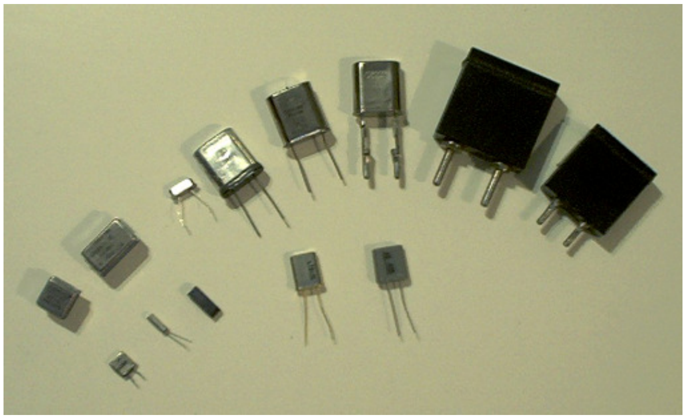

apart, cleaned with alcohol, and made to work again. Careful! The bare quartz wafer is quite fragile. But if the crystal was already inert, breaking it isn't a loss.

An assortment of quartz frequency control crystals

The three big silver-colored crystals in center rear are size HC-33 and are modern-manufactured, first-rate parts. The two medium size HC-49 crystals in the front center are microprocessor crystals. They are good parts, but often they aren't available for the exact frequency you happen to need. Sometimes it's necessary to order custom HC-49s from companies like International Crystal Manufacturing, Inc. (ICM) for about $18 each. The little bitty crystals at the left front should be used with caution. Little crystals get warm easily and their frequencies drift. The square and rectangular "crystal blocks" at the extreme left are complete crystal controlled oscillators packaged in little cans. These are intended for computer work, not RF frequency control. So far, I have never encountered a crystal block oscillator that didn't run hot as a pistol and drift like crazy. Unless you've checked it out, ***don't use an oscillator block***!

Common Crystal Oscillator Circuits

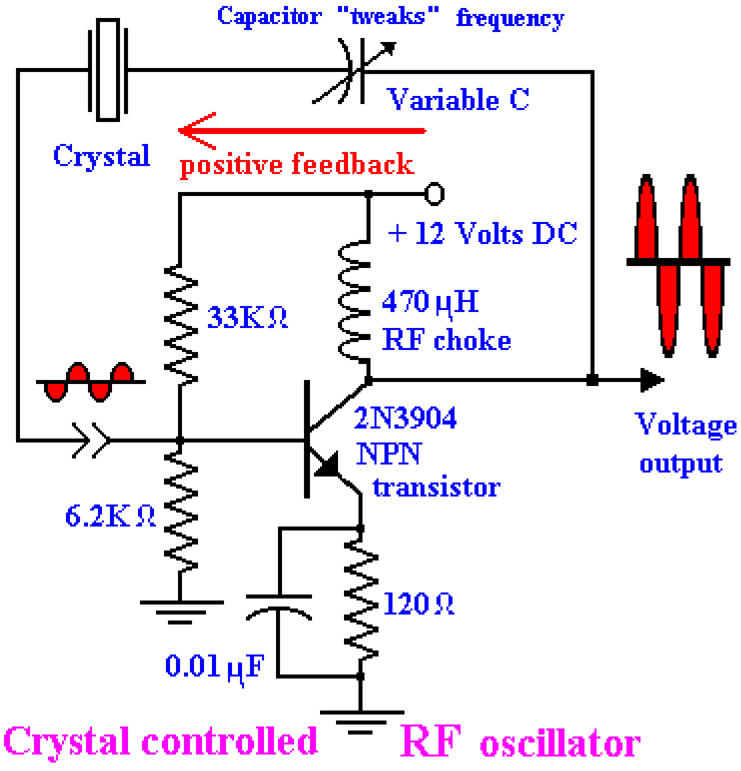

There must be a dozen or more transistor crystal oscillator circuits. The circuit on the left is practically the same as the crude LC oscillator illustrated earlier. The variable capacitor allows you to "tweak" the frequency a kilohertz or more above, on, or below the nominal frequency printed on the crystal case. The circuit above is useful and you will occasionally see it used in homebuilding projects. Yes, if you like, you can leave out the variable capacitor. However, it is not a good idea to put big sinewave voltages directly on the crystal. For example, if you leave out the variable capacitor, the collector voltage will be directly on the crystal. Too much RF voltage on a crystal can heat it and cause frequency drift. Crystal heating will cause the oscillation frequency to begin dropping the moment you turn on the oscillator. In extreme cases, such as exposing the crystal to big voltages in a vacuum tube oscillator, the voltage will literally crack it and ruin it. For these reasons, although this circuit is easy to explain, I rarely use it.

Series-cut and parallel-cut crystals

So called ***series cut*** crystals are designed for use with a capacitor in series with the crystal as shown above. This means that when the series-cut crystal has a specified size of capacitor in series, the crystal will oscillate at the exact frequency on the label. Otherwise, it might be a kilohertz off or more. Similarly, ***parallel-cut*** crystals are labeled to operate with a specific parallel capacitance as shown below.

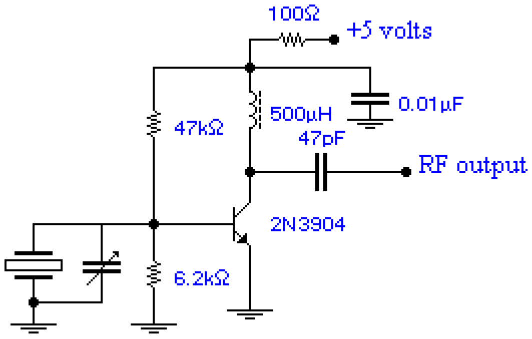

The oscillator shown on the left has the crystal connected to the base of the transistor. A variable capacitor is connected in parallel with the crystal. Again, when you first turn on the oscillator, the exact oscillator frequency may be hundreds or even thousands of Hertz off the nominal frequency. By adjusting the capacitor, the oscillation can be the exact frequency you need.

Butler is better

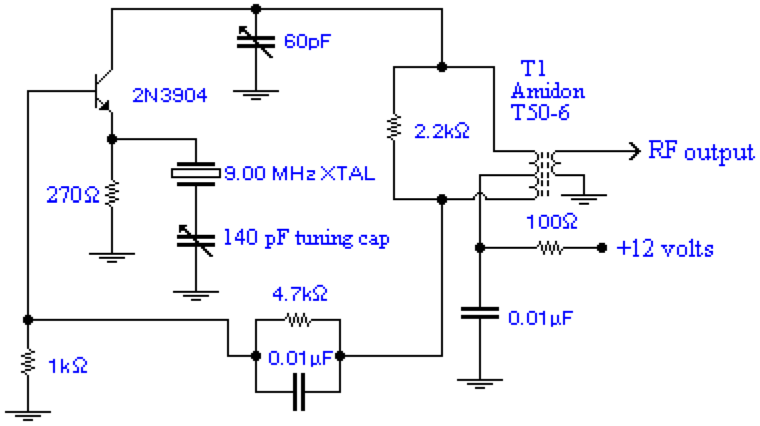

A Butler crystal oscillator

I checked out every oscillator in my receiver and transmitter and discovered that some of them didn't have the warm-up drift when turned on. The stable ones were Butler oscillators as shown on the left. Notice that ***the crystal and its capacitor are in parallel with the emitter resistor.*** I don't really know why, but this circuit is stable the moment it is turned on.

Maybe it's because the crystal just shunts the emitter resistor so the crystal voltage is low and it receives little heat energy. Anyway, Butlers typically drift no more than a hertz or two per minute. Two of my oscillators showed zero Hertz drift during the first minute. One of them, my receiver BFO shown above, was on the same Hertz a half hour later!! Now that's what crystal control is supposed to be like. Also, it wasn't necessary to put these oscillators in heavy-walled, sealed metal boxes. Even built on an exposed board out in the breezes, they are impressively stable.

On high frequencies, like 30 MHz, the circuit as shown above may not oscillate right away. If you're sure it's wired right and has 12 volts applied to it, then try increasing the 270 ohm resistor to 470 ohms. If the output is too weak, sometimes it can be increased by lowering the same 270 ohm resistor to say, 150 ohms. Also, big crystals tend to put out much larger signals than small ones.

Depending on your application, the Butler also has the advantage that the auxiliary tuning capacitor can pull the frequency lower than the base-connected oscillator shown earlier. I studied the oscillator circuits in my ARRL annual handbooks. In one of the examples of a modern commercial transceiver in the 1998 handbook there was a version of the Butler oscillator. These oscillators were touted as "low phase noise" but the low initial drift advantage wasn't mentioned.

Most modern equipment use anonymous integrated circuit (IC) oscillators, like "NE602." Among other circuit blocks, these ICs contain oscillators. To wire it, the crystal is just connected to pins on the little square package. Heaven only knows what's in there! I guess only we homebrewers care.

****

The 40 meter QRP circuit

The circuit shown below is the core of the QRP transmitter. The crystal oscillator on the left is a butler oscillator. The transistor RF amplifier on the right is similar to the example shown earlier, but is tuned with an L-C circuit for a specific band. Also, its output is a transformer winding that "matches" the high impedance output (roughly 600 ohms) to a 50 ohm dipole antenna.

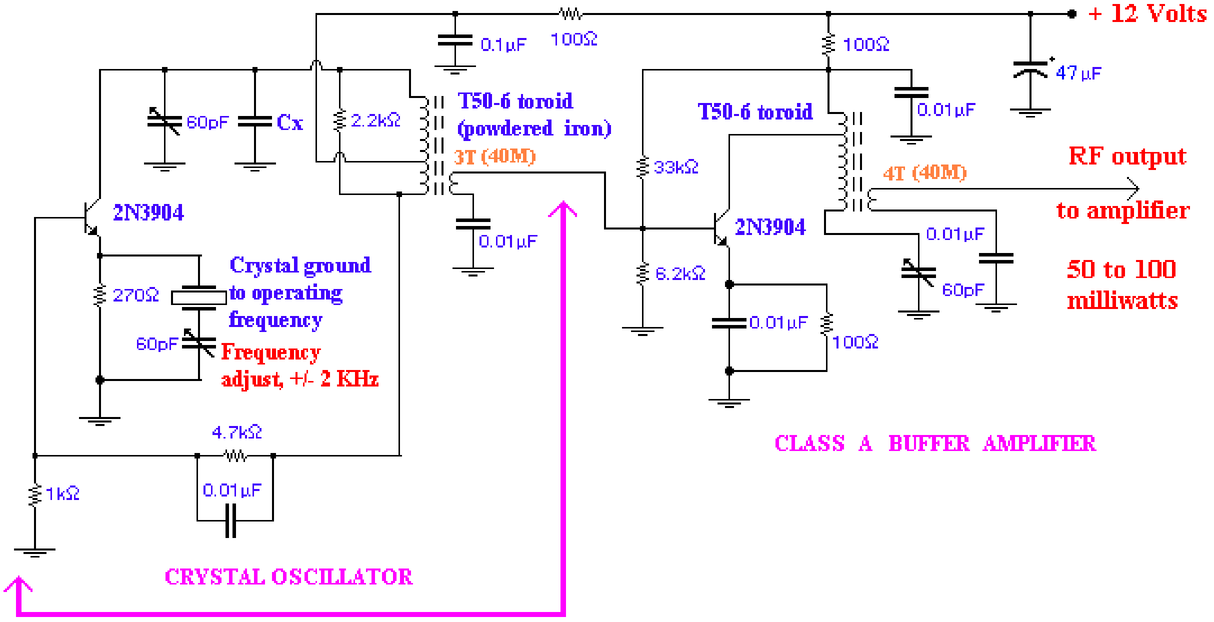
+ 12 Volts
0.1µF
100Ω
100Ω
47µF
60pF
Cx
2.2kΩ
T50-6 toroid
(powdered iron)
3T (40M)
T50-6 toroid
0.01µF
RF output
to amplifier
33kΩ
4T (40M)
2N3904
0.01µF
2N3904
0.01µF
6.2kΩ
60pF
50 to 100
milliwatts
270Ω
Crystal ground
to operating
frequency
60pF
Frequency
adjust, +/- 2 KHz
0.01µF
100Ω
CLASS A BUFFER AMPLIFIER
4.7kΩ
1kΩ
0.01µF
CRYSTAL OSCILLATOR

The Crystal Oscillator and Buffer

The diagram above shows oscillator and buffer stages for the QRP. This circuit puts out about 1/8 watt and could be used directly. Of course, you'd have to have a terrific antenna for anyone to hear you. How much power you get from this circuit depends on the crystal you use. I used a big, new, size HC-33 that worked so well, that instead of 5 watts, my QRP (including the final amplifiers described below) put out 15 watts with 12 volts on the power supply. Oops! When I run it on 6 fresh flashlight batteries (9 volts) it puts out about 7 watts.

The Butler oscillator circuit comes from a receiver project in the 1986 ARRL handbook and I've used it successfully many times. The tuned amplifier-buffer is part of a QRP design in the 1979 handbook. This basic amplifier circuit gives high voltage gain and can also serve as a tuned, active filter. Or, when it's coupled to a low impedance secondary, as it is here, it works as a power gain stage.

Building this QRP is really several small projects. Build the oscillator and get it working before you build the buffer amplifier. The important thing is that it works, not beauty or compact size. When building a project, use circuits you understand and parts you can get.

The diagram above doesn't include a detailed parts list. Parts lists are nice, but only if you can actually buy those exact parts. Manufacturers and distributors change every year and listing specific manufacturers will only frustrate you. The inductor cores are the only critical part. They should be powdered iron and designed for approximately the right frequency range. To work with a brand like Amidon (CWS Bytemark), Micrometals, or Fairrite, you need the A_L inductance value so you may calculate the number of turns. This process is explained later in this chapter.

Inductor specifications for the QRP oscillator

Band	Toroid type	Primary turns	Secondary turns
80 Meters	T50- 15	41 turns, tapped 1/3	4 turns & 5 turns
40 Meters	T50- 15	30 turns, tapped 1/3	3 turns & 4 turns as shown.
Or,	T50- 2	51 turns, tapped 1/3	5 turns & 7 turns
30 Meters	T50- 6	36 turns, tapped 1/3	7 turns & 9 turns
20 Meters	T50- 6	28 turns, tapped 1/3	3 turns & 4 turns
17 Meters	T50- 6	22 turns, tapped 1/3	3 turns & 5 turns
15 Meters	T50- 6	22 turns, tapped 1/3	3 turns & 4 turns
12 Meters	T50- 6	16 turns, tapped 1/3	3 turns & 3 turns
10 Meters	T50- 6	16 turns, tapped 1/3	3 turns & 3 turns

Turns are NOT critical – if it tunes up well in the center of the variable capacitor range, that's the right number. (The toroids are CWS Bytemark, formerly Amidon)

Tapped toroid inductors

The crystal oscillator and following tuned amplifier stages of this QRP transmitter use tapped coils for the primary windings of the output transformers. The tap is 1/3 of the way from the power supply end of the coil. The DC current from the 12 volt source flows into the primary winding, then after 1/3 of coils, it leaves through the tap and enters the transistor which turns on every half cycle and shunts the current to ground. The opposite end of the primary winding is connected to the variable capacitor making a resonant LC circuit. At this end of the winding there is no exit for DC current. So 2/3 of the coil is only used for resonant "ringing" of RF currents. That is, most of the inductor is devoted to oscillating at a tuned frequency.

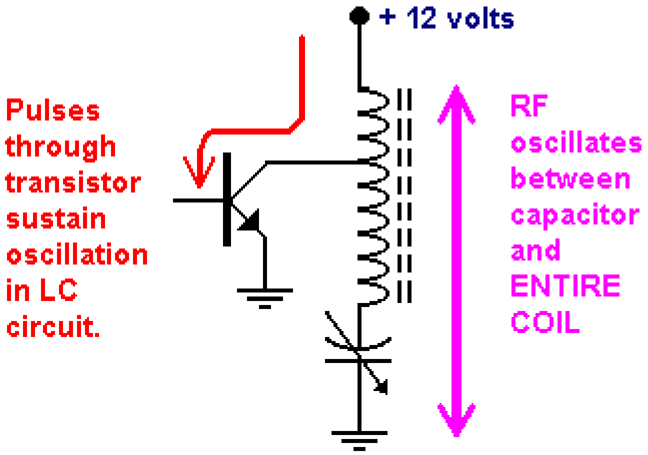
+ 12 volts
Pulses through transistor sustain oscillation in LC circuit.
RF oscillates between capacitor and ENTIRE COIL

Tapped Inductor forms high Q tuned circuit

Short pulses of DC current through the 1/3 end of the coil serve to prime the oscillation and keep it going. It is comparable to an adult pushing a child on a swing. The adult only pushes briefly at one end of the arc of the swing. The pendulum system of the swing does most of the oscillating, not the adult. This tapped coil makes the oscillations much larger than they would be if the winding weren't tapped. Also, the oscillation is much more confined to a specific frequency. In other words, the "Q" or quality of the resonant circuit is higher.

A mistake you might make

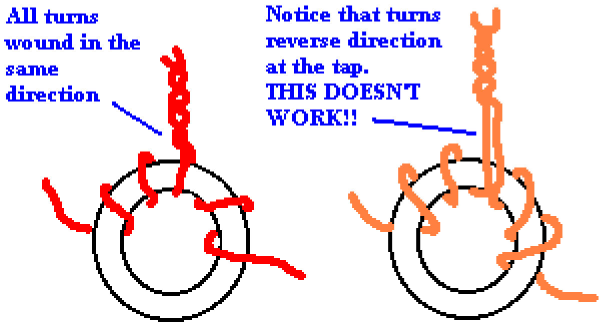
All turns wound in the same direction
Notice that turns reverse direction at the tap.
THIS DOESN'T WORK!!

WINDING A TAPPED COIL

The drawing on the left shows the right way and wrong way to wind a tapped inductor. A tapped inductor is supposed to be a single coil that has an external wire connected to some spot along the coil. In other words, the coil must be wound from beginning to end in the same direction. In the drawing above, the correct tapped coil has three turns wound in one direction, then three more turns wound in the same direction. If the coil winding direction reverses at the tap, the inductance of the first half of the coil will cancel out by the inductance of the second half. In other words, the device on the right has no inductance at all! The device on the right isn't even an inductor, it is just a "wire."

Q = equals quality

Inductors and capacitors have a quality factor called "Q." It is defined as the reactance of the device, X, divided by the resistance of the device, R. ***Reactance is the property of capacitors and inductors that resists the flow of AC current and acts like a resistor.*** Inductors are made from lengths of copper wire wound into a coil. Because of the resistance of the copper metal, the resistance of inductors wound with fine wire and many turns can be significant, even without the reactance component. Therefore, inductors are sometimes “low Q.” In contrast, the resistance of the short wires in capacitors is rarely significant, so capacitors are almost always “high Q.” In large capacitors that have dielectric plastic (or electrolytic) insulation between the plates, this insulation dissipates energy and appears as a “resistance” and lowers the Q.

Frequency multipliers

By the way, the tapped coil amplifier can also be used as a frequency multiplier. Frequency multipliers can be explained by the adult-pushing-the-kid-on-the-swing analogy. Instead of the adult giving the swing a push with every cycle, suppose the adult only pushes every other cycle or every 3rd or 4th cycle. In this way the frequency of the pushes can be some fraction of the natural frequency of the swing. Notice that the frequency of the kid's motion on the swing depends on the length of the swinging pendulum and has little to do with the frequency of the adult pushing. In other words, the length of the pendulum must be "tuned" to the frequency of the pushes. Or, the push frequency can be tuned to the swing frequency or an even fraction of the swing frequency.

For example, let's apply a 7 MHz sinewave to the input. Now suppose the LC circuit is tuned to 14 or 21 MHz. While the amplifier is running, your scope will reveal a 7 MHz sinewave on the transistor collector. Meanwhile, the whole coil and capacitor will be oscillating at the multiple of the frequency of the input determined by the LC circuit. To make use of the higher frequency, tap into it with a tiny capacitor, like 2 pF, between the coil and the trimmer capacitor. A bigger capacitor will tend to load the ringing LC circuit too much and kill the oscillation. Alternatively, the higher frequency oscillation can be sampled with a one or two turn secondary winding wrapped around the primary coil.

If inductive and capacitive resistance is called “reactance,” what’s impedance?

Impedance is the sum or total result of all the reactances and resistances in a circuit. When we say a circuit has an impedance of “100 ohms,” we are saying that every component in the circuit works together to behave like a 100 ohm resistor *at that particular frequency*. For example, as mentioned earlier, most ham antennas, receivers, and transmitters are designed to work with 50 ohm impedance loads.

Impedance matching

If you wish to transfer power from one circuit to another and you don’t have a perfect voltage source, then you can transmit the most power if you “match” the impedance of the load with the internal impedance of the voltage source. In other words, real voltage sources always have internal resistance that limits how much energy you can suck out of the source. For example, a fresh battery has low internal impedance. In contrast, a dead battery (usually) has the same chemically generated voltage, but when the battery is exhausted, there is now a huge resistance in series with the voltage. This limits you to very little power that can be drawn out of the battery. However, even with a nearly dead battery, you can always maximize the power transfer by matching the load resistance to the internal resistance, no matter how high it may be.

The buffer stage and power supply decoupling

A ***buffer amplifier*** follows the oscillator. The purpose of a buffer is to isolate the oscillator from the final amplifier(s) as well as amplify the oscillator signal. Although it makes no intuitive sense, changes on the output load of an amplifier can couple back to the input of the amplifier and affect the oscillator. This feedback makes subtle changes in the frequency of the oscillator and contributes to ***chirp***. Chirp is a change in tone of the Morse code that makes it sound like a bird. One way to decrease chirp is to add a buffer amplifier between the oscillator and final amplifiers.

Ideally, all "communication" between one amplifier and the next is through the intended path between the output transformer and the base of the next stage. Unfortunately, other ways amplifiers "talk" to each other can be through the power supply or even through skinny ground traces. You can minimize the ground communication by using large ground traces or by soldering feed-through ground points through the PC board to the unbroken grounded copper sheet on the far side.

Power supply communication is more difficult. Suppose a big sinewave signal is being generated in a final amplifier. As the current rises through the large output transistor, this big current loads down the power supply voltage and can cause it to drop. The amplifier driving the final then loses some of its supply voltage, just when it needs to supply the rising current. The result of this interaction is that the sinewave can jitter and become unstable. On an oscilloscope the sinewave loses its focus and becomes blurry. A frequency counter will usually read below the crystal frequency and will be unstable. This instability can be greatly improved by decoupling each stage as shown below:

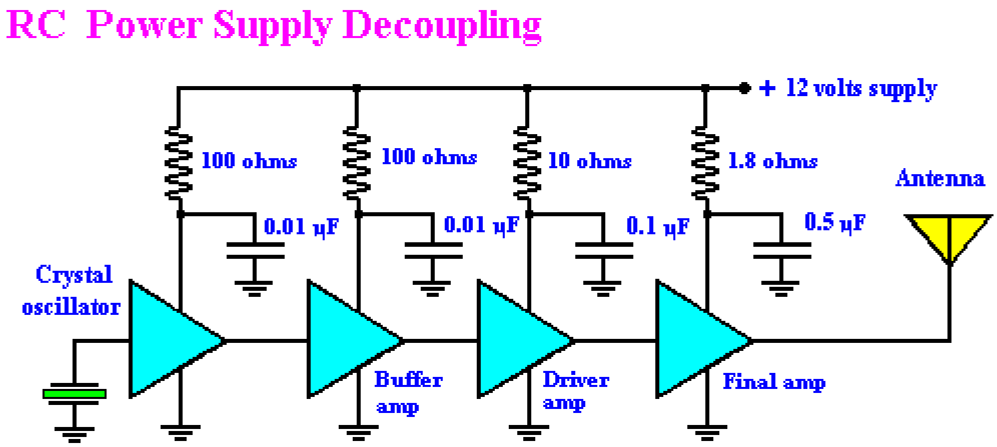

The resistor / capacitor feeding each stage slows the changes in the power supply voltage, RF cycle by cycle. That is, the R-C time constant of each network should be much longer than the variations in the power supply voltage.

R-C time constant is just ***resistance times capacitance = time constant in seconds***

e.g., 1.8 ohms X 0.5 x 10^{-6} farads = 0.9 microseconds

This limits the voltage change to the output stage much slower than an RF cycle. That is, 1/ 0.9 µ sec = 1.1 MHz

These RC decouplers prevent the voltage changes from feeding back to earlier stages. For the stages that draw the most current you need smaller resistance so that the DC voltage loss across the resistors will be as low as you can tolerate. Also, ***The more current each stage draws, the more capacitance it needs.*** The capacitors should be low loss types designed for RF bypass applications such as disk ceramic or mica. If you leave out these RC circuits, the QRP will still work fine ***IF*** you have a stable 12 volt supply. However, if you use a weak supply, such as old flashlight batteries, it will be very difficult to maintain a clean sinewave output.

LC decouplers

The first QRP I built was from a design in the 1979 handbook. This design used inductors instead of resistors in the decoupler circuits. Since there is no appreciable DC voltage drop across an inductor, inductors look like a good idea. Unfortunately my circuit was extremely unstable until I took out the LC decouplers completely. It was only later when I tried to run the QRP on flashlight batteries that I discovered that RC decouplers were useful.

The final amplifier stages for the QRP

I was surprised by the big signal I got out of the oscillator stage described above using the large HC-33 size crystals. If you use small HC-49 crystals, the power output will probably not be so great and the first stage of the two-stage amplifier described below will be needed. QRP is defined as less than 5 watts. As explained earlier, you shouldn't have any trouble getting enough power from this little transmitter.

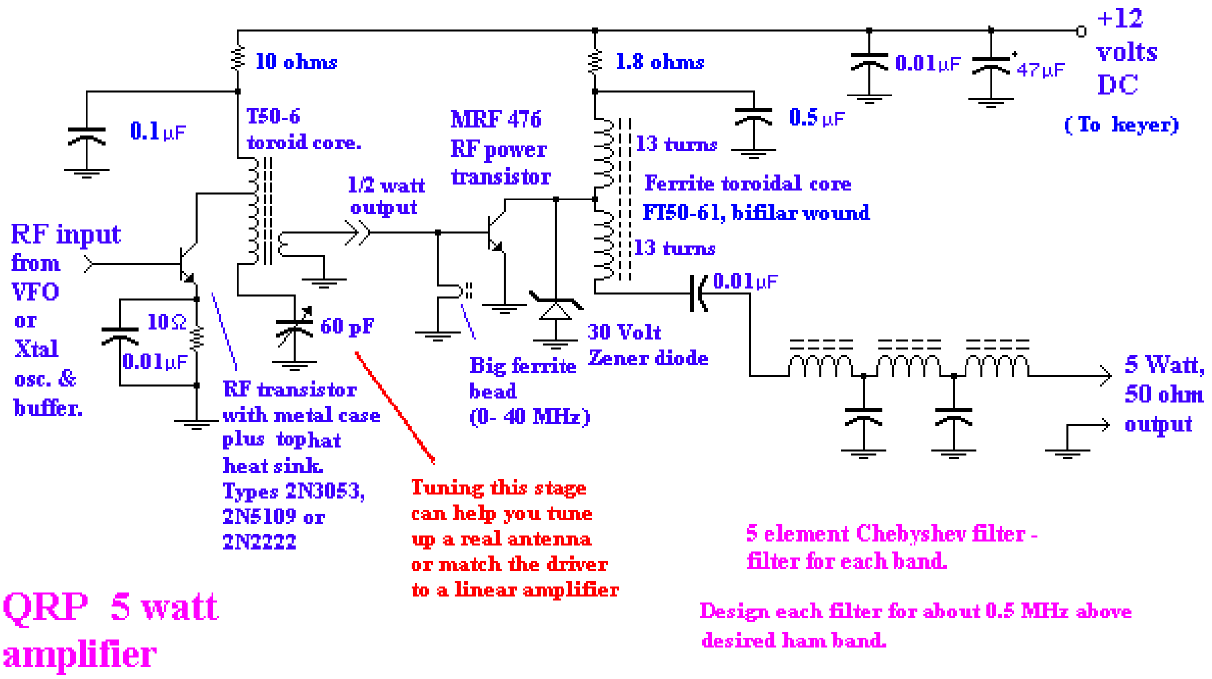
+12 volts DC
(To keyer)
10 ohms
1.8 ohms
0.01µF
47µF
0.1µF
0.5 µF
T50-6 toroid core.
MRF 476 RF power transistor
13 turns
Ferrite toroidal core FT50-61, bifilar wound
13 turns
1/2 watt output
RF input from VFO or Xtal osc. & buffer.
10Ω
0.01µF
60 pF
0.01µF
30 Volt Zener diode
Big ferrite bead (0- 40 MHz)
RF transistor with metal case plus tophat heat sink. Types 2N3053, 2N5109 or 2N2222
Tuning this stage can help you tune up a real antenna or match the driver to a linear amplifier
5 Watt, 50 ohm output
5 element Chebyshev filter - filter for each band.
Design each filter for about 0.5 MHz above desired ham band.
QRP 5 watt amplifier

A QRP Two Stage Power Amplifier

The last two of the four transistor stages in the QRP are shown above. The first stage is tuned and resembles the buffer stage following the oscillator. Depending on the number of turns on the coils in the tuned amplifier stages, these stages can tune two or more bands. For example, with 27 turns on the primary, you can (barely) cover 20 to 10 meters. Make the tap about 1/3 of the total number of turns, or about 9 turns in this case. The secondary would be about 6 turns. The input of a transistor amplifier is usually a low impedance. Therefore the transformer lowers the voltage and raises the current to "match" the impedance.

The buffer that follows the oscillator circuit is ***Class A*** because it's biased "on" with the 33K resistor. The emitter R-C circuit keeps the class A amplifier stable so it won't run away when the temperature rises. Class A is best when the signal levels are small. In the two amplifiers above, the first (tuned) amplifier runs ***Class C.*** This means that the input voltage swing must be far larger than the base offset voltage, 0.6 volts. This Class C runs at relatively high power levels, so you'll need a bigger transistor, 2N3053, 2N2222 or equivalent. This moderate power transistor has a metal case like a tiny tin can to help conduct away the heat. Moreover, a "top hat" heatsink should be clipped onto this transistor to help radiate the heat. Look in your catalogs under heatsinks and you should find a large assortment.

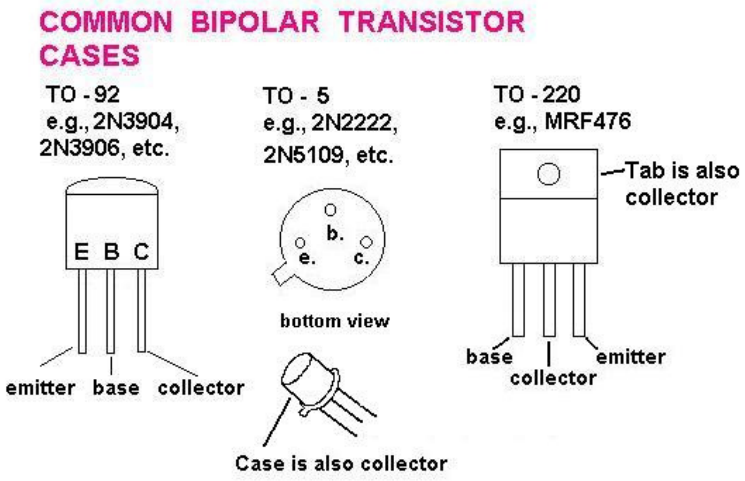
COMMON BIPOLAR TRANSISTOR CASES
TO - 92
e.g., 2N3904, 2N3906, etc.
E B C
emitter
base
collector
TO - 5
e.g., 2N2222, 2N5109, etc.
b.
e.
c.
bottom view
Case is also collector
TO - 220
e.g., MRF476
Tab is also collector
base
collector
emitter

Bifilar wound transformers

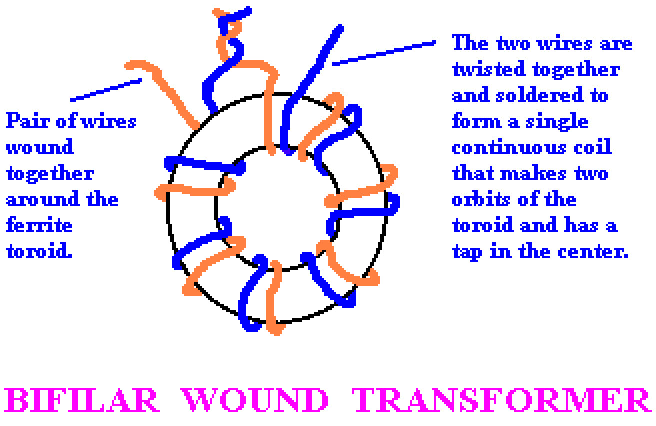
Pair of wires wound together around the ferrite toroid.
The two wires are twisted together and soldered to form a single continuous coil that makes two orbits of the toroid and has a tap in the center.
BIFILAR WOUND TRANSFORMER

The transformer for the broadband final amplifier stage is a "bifilar wound" transformer. This broadband transformer is untuned and will work on *ANY* HF hamband. The toroid core is an Amidon T50-61 ferrite instead of powdered iron. The ferrite provides much higher inductance than you would obtain if the same coil were wound on powdered iron cores, such as the T50-6. That is, the A_L factor is much larger for a ferrite core. The high inductance means that the input signal to the transformer will be passed along to the output before the inductor has a chance to charge. There is nothing resonant about this transformer.

Although the transformer is wound with two parallel wires, the two wires are soldered together to make one winding that orbits the toroid twice. Bifilar wound transformers are a kind of tapped coil. As a result, the impedance (voltage) can be stepped up or down by connecting the output to either all of the coil for high voltage, or just half of the coil to step down.

Before you wind one of these coils, examine the drawing extremely carefully. It works only when wound exactly as drawn. This is really easy to screw up! ***If you connect one of the wires to itself, that wire becomes a shorted turn that will dissipate most of your RF energy.*** Before you solder your bifilar into a circuit, use an ohmmeter to confirm that all three terminals have zero ohms betweens them. If you find that the "centertap" is an open circuit with respect to the other two wires, you have connected it wrong. I have made this error twice and wasted considerable time finding it.

Tuned amplifiers versus broadband amplifiers

The QRP described above uses two amplifier stages that are tuned by trimmer capacitors and a broadband final amplifier that needs no tuning. Tuned amplifiers almost always work – that’s a big advantage! A disadvantage of tuned amplifiers is that the tuning can be critical and may not cover the entire band. Tuning all the stages in a transmitter turns out to be a bad idea. As my car battery power supply discharged, sometimes my signal would “crash” halfway through a QSO. On the scope I could see a sudden loss of amplitude on the output and distortion would appear in the sinewave. My contact would say, “You’re signal is breaking up.”

Another advantage of tuned stages is that, when you tune up one stage, by subtle shift in the sinewave phase, you’re also adjusting impedance matches in the following stages. It works much like an antenna coupler matching an antenna. I confess I don’t fully understand this, but the phenomenon is real and useful.

Noise mode

In contrast, broadband amplifiers just pass along whatever signal they receive. So, if you build all your stages as broadband amplifiers wired in series, they won’t work unless you’ve done a wonderful job of matching impedances at every stage. When a broadband amplifier is mis-matched, it goes into “noise mode.” That is, you put a sinewave into the input and the amplifier puts out a blast of noise that contains only a ghost of the sinewave you hoped to amplify. If you don’t have at least one tuned stage in your QRP, you’ll have nothing to adjust when the final goes into noise mode.

The output stage in the QRP is a broadband amplifier that uses a five-element Chebyshev filter to suppress high frequency harmonics in the output. Chebyshevs are described in detail below. The filter component values can be found in tables and formulas in an ARRL Handbook, e.g., 1986 or later. This broadband amplifier is a universal circuit you may add to your bag of tricks. For example, in my ten meter QRP board I used two broadband stages in series to increase power from 2 to 9 watts. The first broadband stage used a 2N3053 and it didn’t need its own Chebyshev filter.

Expensive RF transistors

The final amplifier uses a Motorola MRF-476 transistor. They are pricey, $13 or more.

According to the RF Parts Co. 2006 catalog, the type 2CS2166 is direct replacement for the MRF-476 and costs only $1.95. I haven' t tried any, so I can't vouch for this. There must be hundreds of cheap transistors that can put out 5 or 10 watts at HF frequencies. Unfortunately, the cheap transistors I have actually tried needed 24 to 80 volts collector supply voltage to deliver the same power. The virtue of Motorola “MRFs” for HF seems to be that they deliver big power with a 12 volt supply. To me it was worth the cost.

Ferrite bead RF chokes

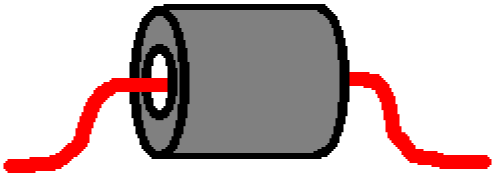

FERRITE BEAD RF CHOKE

Notice the bead RF choke on the MRF-476 base. It’s just a short piece of wire through a big ferrite iron bead. This eliminates low frequencies from the output and turns a “roller coaster,” chaotic output waveform into a clean sinewave. This simple component looks as though it would act as a short circuit to ground. But even if you run the amplifier on a low frequency like 80 meters, the bead is vital and doesn’t reduce output power. If this sounds illogical to you, unsolder the choke and look at your waveform.

****
**

CAPACITORS

Capacitors are rarely a problem

In contrast to inductors, capacitors are rarely a problem. If the label on the capacitor says 330 picofards, that’s probably what it is. Small mica capacitors are often labeled with 3 numbers. The last number is the number of zeros following the first two digits. For example, a 330 picofarad capacitor would be labeled "331." 33 picofarads has no zeros so it would be "330" which is confusing. If the label is puzzling, a capacitance meter usually tells you everything you need to know about the unlabeled capacitor in your junk box.

Electrolytic capacitors can be a little tricky. Those are the caps with the plus and minus labels on the leads. You have to orient them with the correct polarity. The positive end must be aligned with the positive voltage in your circuit. Also, there are a few rules about what kind to use where. Aluminum electrolytics have large capacitances and are often used to stabilize power supplies. They are packaged in little plastic-coated aluminum cans and have + marked on the positive lead. They do not conduct RF particularly well so it may sometimes be advisable to shunt an aluminum electrolytic cap with a ceramic or mica RF bypass.

Tantalum electrolytics have the largest capacitances relative to their small physical size. They are usually packaged in tiny plastic rectangular blocks or blobs of colored epoxy. Again, they have plus and minus lines on them to indicate polarity. Sometimes the positive is just marked with a dot of colored paint. They conduct RF well. That is, they make good RF bypass capacitors as well as low frequency ripple stabilizers. On the other hand, tantalums have low voltage ratings and you should not place them directly across a battery. Sudden rises in voltage across a tantalum can cause it to break down and burst into flames. For example, I once designed a battery-powered circuit with a 10 volt rated tantalum wired across a 9 volt battery. It wasn't long before I had a charred capacitor and circuit board.

CONQUERING INDUCTORS

Inductors are often the problem

As you gather parts for the above QRP transmitter or any typical ham project, you may discover that the hardest parts to locate are the inductors. “Where do I buy a Miller # 233 anyway?” Or, maybe the parts list says, “6 turns on a Stackpole 4-12 toroid core.” Who sells Stackpole cores? Where do I get a catalog? Then you think, “I know what to do! I’ve got an iron core-thing that LOOKS just like the picture. I’ll use that!” With this optimism you are well on your way to building a useless piece of junk destined for your attic. All of us who have wound inductors have wished our multimeters had “inductance scales.” Sorry. Handheld inductance meters barely exist. And if you have one, it probably won’t tell you much about core losses, leakage inductance, saturation, and winding resistance.

A little math is as good as an inductance meter

Normal humans hate mathematics. However, simple calculations and toroid cores with known characteristics are a tool that let you wind the exact inductor you need. Your LC circuits will resonate in the right band, your filters will attenuate and pass the right frequencies, and your equipment will work! The ability to wind the right inductance is a vital skill.

LC circuits. What size inductor do I need?

Radio technology is based on LC circuits. Sure, some circuits use piezo devices like crystals and “SAWs” that only ***act*** like LC circuits. But they are LC circuits none-the-less. The C part is easy enough. Just pick a capacitor value. But how big does the inductor need to be?

$\omega^2 = 1/LC$ where $\omega = 2\pi$ (frequency in Hertz)

L = inductance in Henries.

C = capacitance in Farads.

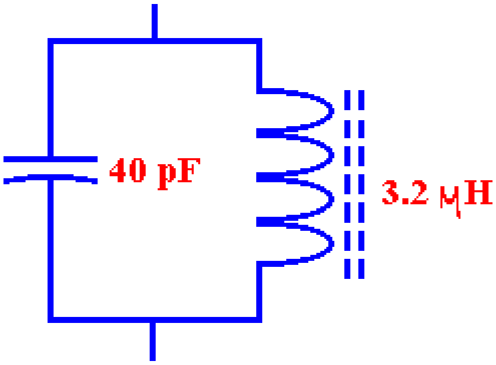

20 Meter RESONANT CIRCUIT

Suppose you’re building a transistor amplifier stage that has an LC resonant tank circuit. Your trimmer capacitor has a range of 5 to 60 picofarads. You need an inductor that resonates with that capacitor. Let’s say the band is 20 meters, 14.1 MHz. We’ll design the inductor to resonate with say, 40 pF. That way, you can tweak the frequency if necessary.

Using the above formula:

$[2 \times 3.1416 \times 14.1 \times 10^6 \text{ MHz}]^2 = 1/ (40 \times 10^{-12} \text{ Farads})(L)$

Solving for L, $L = 3.18 \times 10^{-6}$ Henry or 3.18 microhenries.

Now that we know how big the inductor must be, we need to wind a coil on a toroid core.

Using CWS (Amidon) powdered iron cores

There are many quality brands of ferrite and powdered iron core toroids. I like CWS Bytemark cores (Formerly Amidon) simply because I know where to buy them. Go to www.bytemark.com or www.coilws.com. The important issue is that ***the cores must have a known inductance factor, A_L.*** A_L allows you can calculate how many turns you need for a given inductance.

Ferrite vs. powdered iron

Ferrite cores have high iron content and produce high inductance for a given number of turns. In ham projects, ferrite is usually used for inductors that are too large to charge significantly for several cycles of the design frequency. In other words, ferrites are used for RF chokes and waveform transformers. For example, the output transformers of *WIDEBAND* linear amplifiers are usually ferrites. The final amplifier of the QRP circuit above uses a T50-61 ferrite toroid. Ferrites are rarely if ever used for resonant, tuned LC circuits.

In contrast, powdered iron cores contain far less iron and more ceramic. They are used for high-Q resonant circuits and filters. Notice the powdered iron cores are color coded for their permeability. In contrast, all the ferrites are just plain black, semi-shiny toroid cores.

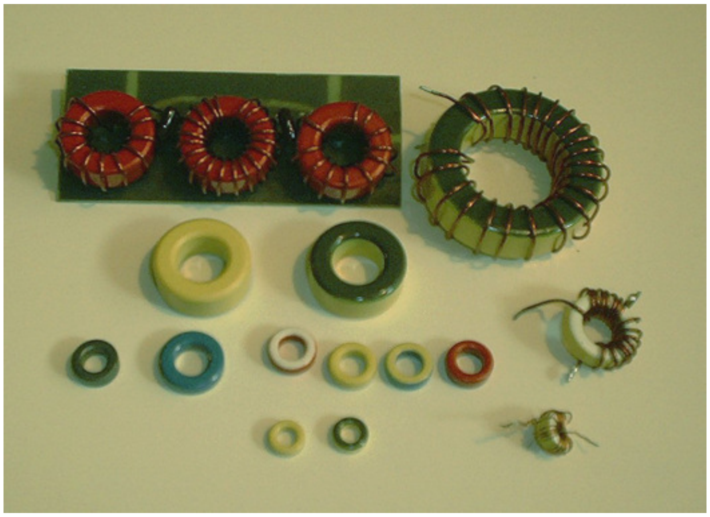

First: Select the type and size of toroid core appropriate for the frequency and power level. There are several grades of powdered iron and ferrite cores that are designed for different frequency ranges. For resonant circuits or filters on the upper HF hambands, I usually ***use type 6 powdered iron (color code yellow and black)***. For the lowest HF hambands 80 and 160, I sometimes use type 2 (red/black) or type 15 (red/white). These give me more inductance for the same number of turns and allow me to use heavier gauge (lower resistance) wire. Type 17 powdered iron (yellow-blue) is useful for 30 MHz and above. Type 26 (yellow-white) cores are most useful for frequencies well below the ham bands. They are often found in switching power supplies.

The power level is proportional to the size of the toroid core. I've had bad luck with little bitty size T-37 cores. They produce low gain and are hard to wind. Don't use them unless you are pressed for space. On the other hand, type T-50 works fine from receiver circuits up to a few watts in a QRP. Type T-68 is good for 10 watts or more. Type T106 and T200 handle 200 watts or more and are used for output filters in high power linear amplifiers.

Second: Calculate the number of turns needed for a given inductance. The wire size is simply the largest gauge that will fit conveniently in the toroid. Don't go nuts with thick wires. On the other hand, for low frequency, high inductance usually means many turns of wire. The coil will work best when you use the highest diameter wire that will fit in the core without overlapping the turns. For each core type, there is a constant, "A_L," that is proportional to the square of the number of turns.

The number of turns = 100 (Inductance in microhenries)/ A_L) $^{1/2}$

Suppose we are using a T-50 type 6 powdered iron core to design a 3.18 microhenry inductor:

The T-50-6 core has an inductance constant, A_L, equal to 40. Note that when calculating, the "micro" part isn't in the calculation. 3.18 microhenries is just "3.18."

Number of turns = 100 $(3.18 / 40)^{1/2}$ = 28 turns.

Note: a bracketed number to the power ½ is another way of writing "the square root of."

You should be able to get 28 turns of # 30 enameled magnet wire onto the core. If not, use a finer wire, say # 34. If the wire doesn't cover up most of the core, use a fatter wire, say # 26.

****

Calibrating the trimmer capacitors

I used small, cylindrical variable "trimmer" capacitors to tune the first three stages of my QRP module. They are about the size of a grape and are adjusted with a small screwdriver. A problem with these devices is that you can't tell what capacitance they are set to by just looking at them. Since the adjustment screw goes around and around without stops, there is no way to know when they are at maximum or minimum.

Suppose you tune up an amplifier stage and you find that it produces the maximum signal at a certain setting. If you knew that this set point were the maximum capacitance point, you could guess that probably you need more capacitance to get the best performance. You could fix this by soldering a small capacitor, say 30 pF in parallel with the trimmer. Alternatively, you could add a few more turns to the inductor. On the other hand, if you knew that the best performance set point happened coincide with the minimum capacitance point, then you could conclude that the primary winding of the transformer probably has too many turns and needs to be reduced.

My solution is to measure the capacitor with a capacitance meter, then mark the minimum and maximum set points. With my ceramic trimmers, the maximum is 60 pF and minimum is about 7 pF. I use a fine felt-tip marker pen to indicate the maximum capacitance point. Minimum capacitance is 180 degrees from maximum. Ideally, when I have the amplifier tuned up for maximum signal, I will find that the screw is adjusted to roughly half of the capacitor's range. This means I have a good working amplifier stage and can't improve the LC circuit further.

Chebyshev filters

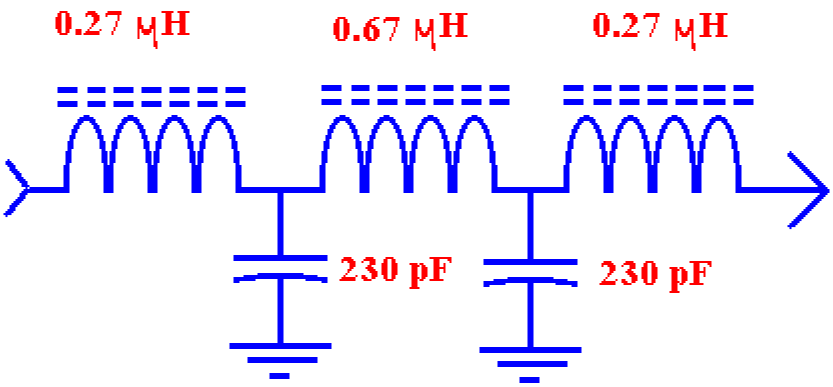

20 Meter CHEBYSHEV FILTER

LC circuits are vital to ***tuned*** amplifier stages. But many modern amplifiers are ***untuned*** or ***linear*** and just need a filter to prevent harmonics above the transmitting frequency. The output stage of nearly every transistorized linear transmitter final has a 5-element filter to prevent harmonics above the operating ham band. Your ARRL handbook has long, long, detailed explanations about designing different kinds of Chebyshevs. Most guys look at all those pages and think, "Like heck I'm going to figure out all that!" On the other hand, you usually just need a 5 element low pass and here is how you design it:

First: What operating impedance do you need? Let's assume 50 ohms. That's the most common.

Second: What frequency is it for? Let's assume we want to attenuate everything above the 20-meter band. That is, 14.35 MHz.

Third: Calculate the "normalized" capacitance and inductance for the filter:

$$\mathbf{Cs = 1 / 2 \pi (Z) f}$$

where π = pi, 3.1416 , Z = impedance in ohms and f = frequency

Capacitance, Cs = 1 / 2 π (50 ohms) (14.35 MHz) = 222 pF

The normalized inductance is defined as:

$$\mathbf{Ls = Z / 2 \pi f}$$

Inductance, Ls = 50 ohms / 2 π (14.35 MHz) = 0.55 microhenries (μH)

Fourth: Multiply the normalized C and L by the ***design factors*** for each of the five elements.

The values of the five elements are:

L1 = **0.4869** Ls = (.4869) (0.55 μH) = 0.27 μH

C2 = **1.05** Cs = 1.05 (222 pF) = 230 pF

L3 = **1.226** Ls = 1.226 (0.55 μH) = 0.67 μH

C4 = **1.05** Cs = 1.05 (222 pF) = 230 pF

L5 = **0.4869** Ls = (.4869) (0.55 μH) = 0.27 μH

For the capacitors, 220 pF (a standard value) will work just fine.

****

Heat sink for the output transistor

Use a big metal heat sink on your MRF-476 and keep it cool. I haven't damaged any, even with as much as 15 watts RF CW output. Notice that a bare MRF-476 is only rated at 3 watts dissipation. I use large aluminum heat-sinks made from 1" aluminum angle from the hardware store. The metal tab on the MRF-476 is the transistor collector and must be kept insulated from the grounded heat sink. I insulate the transistors from the aluminum with mica insulators. I use a film of silicon grease on the mica to improve heat conduction. Look for "transistor mounting kits" in your parts catalog index. The mounting hole for the MRF-476 is tapped into the aluminum with 4-40 thread so that no nut is needed.

Zener over-voltage protection for the output transistor

The 30 volt Zener helps protect your MRF-476 from open circuit operation. The Zener doesn't seem to be vital, but it doesn't hurt performance either. With these precautions I have found these transistors extremely hard to damage. However, the book does say 3 watts only, so try not to run them so hard or so continuously that they become too hot to touch.

Connectors for your QRP module

I use "RCA" audio phono plugs and coax for RF connections between HF modules. Yes, I should be using BNC or other connectors designed specifically for RF. However, phono plugs are cheap, easier to wire, and there seems to be no difference in performance, so long as the power level (current level) is under roughly 10 watts and frequency is below 30 MHz. After all, phono plugs only have twice as much capacitance (4 pF more) than an equal length of coax. 30 MHz is probably as high as one should go with phono jacks, especially at high power levels. Small card edge connectors are another way to snap PC board modules together. For example, I use PC board card edge connectors for plug-in Chebyshev filters so I can change bands easily. You don't need gold-plated card edge fingers. However, you should use wide traces and connect several pins in parallel wherever you're worried about keeping inductance or contact resistance low.

****

Keying CW Transmitters

Some QRP designs put the telegraph key in series with the emitter of a transistor amplifier stage just before the final amplifier. To transmit, the key is pressed connecting the emitter to ground. When the operator wants to "spot" his signal to find out where the transmitter is with respect to a station, he or she turns on the power supply, but doesn't key the transmitter. Without drive to the final, the oscillator and first buffer stage deliver a loud (but not overwhelming) signal to the receiver.

Supposedly, an amplifier with no emitter current or no base drive won't produce a signal. That sounds right and in the old days it worked OK with tubes. Unfortunately, in my basement the transistors don't care and deliver RF anyway. The capacitor across the key and even the capacitance in the coax going to the key provide enough AC current to produce drive through the emitter to deliver drive to the output stage. All I ever achieved with emitter keying was to lower the signal amplitude between the dots and dashes. If you just key the oscillator and leave the other stages active, that works fine as long as the oscillator is completely controlling them. Unfortunately, when you let up the key, the following stages will often self-oscillate at whatever frequency they please.

A P-Channel MOSFET Keys the DC Power Supply

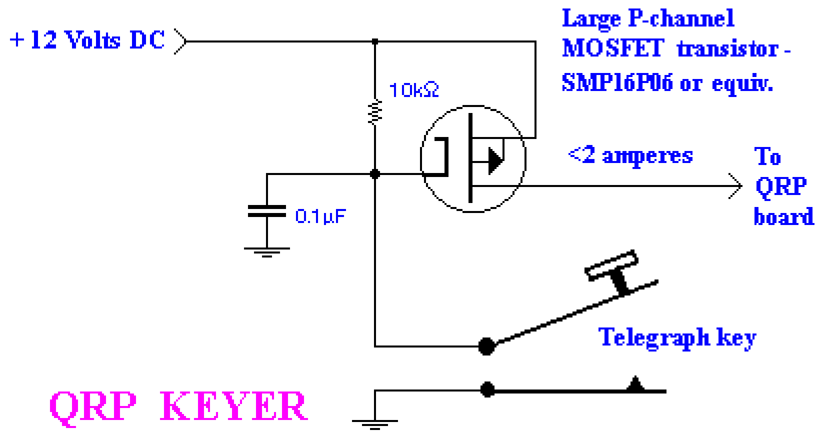

I like to key the transmitter by keying the entire power supply on and off. This circuit is not elegant, but it works. For micropower QRPs, like homebrew transmitters in Altoids boxes and sardine cans, I notice they just put the key in series with the supply. Keying the battery directly works, but for keying 5 or more watts of DC power, it's better for the key contacts to switch the current with a solid-state (transistor) switch. I use a P-Channel power MOSFET as shown above. The resistor and capacitor in the gate circuit soften (extend) the turn-off time to prevent key click.

David, VK6DI, pointed out to me that my circuit shorts out the charge in the capacitor thereby turning the 12 volt signal on abruptly. I agree with him when he says this *should* cause key clicks. As he points out, I could slow the turn-on by putting a small resistor in series with the key. It must be small, like less than 1K so the gate turn on voltage will charge to be nearly 12 volts. Or better yet, we could put another cap across the 10 K resistor. These steps may be "gilding the lily." The above circuit has worked well for me. Perhaps it is the relatively slowly charging capacitors in the rest of the transmitter circuitry that soften the turn-on time and prevent key clicks. He also suggested keying with an operational amplifier to turn the transistor on and off gradually. He also suggest saving money by using bipolar transistors instead of MOSFET. Those ideas all sound good to me, but I haven't tried them.

One of the mysteries of CW transmitters is that producing good-sounding CW with a vacuum tube circuit is harder than with transistors. I had lots complaints about my vacuum tube key clicks and chirp. (See chapter 14.) But in my experience, avoiding clicks and chirps with transistors has been much easier.

If you have any doubts about the sound of your signal, listen to it in a quality receiver with the receiver antenna input shorted and the RF gain turned way down. This will simulate how your signal will sound from hundreds or thousands of miles away. Chirps, clicks or power supply noise on your signal will be obvious.

****

MOSFET field effect transistors

What is that weird transistor called "MOSFET," you ask? Metal Oxide Semiconductor Field Effect Transistors work differently than the bipolar transistor you met in chapter 4. Fortunately, they are easy to understand. A MOSFET transistor consists of a piece of either P or N-type semiconductor mated to a capacitor. The semiconductor strip or layer is necked down in the middle so that the center is very narrow. The narrow region also serves as one of the two conductive plates that make up the capacitor. The other plate of the capacitor is the metal control gate. The gate is insulated from the transistor by a thin film of glass. The control gate is analogous to the base of a bipolar transistor. When voltage is applied to the capacitor, (that is, applied between the control gate and the piece of semiconductor), charge gathers on the conductive surfaces around the gate insulator. When charge gathers in the semiconductor, it changes the ion density in the semiconductor and changes its conductivity.

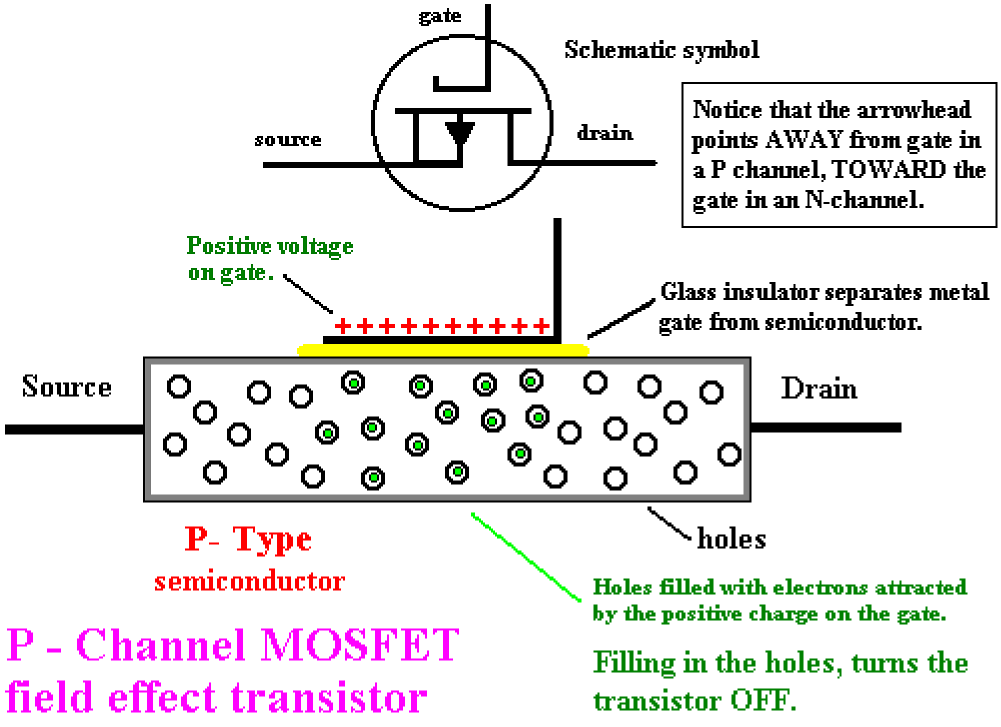
gate
Schematic symbol
source
drain
Notice that the arrowhead points AWAY from gate in a P channel, TOWARD the gate in an N-channel.
Positive voltage on gate.
++++++++++
Glass insulator separates metal gate from semiconductor.
Source
Drain
holes
P- Type semiconductor
Holes filled with electrons attracted by the positive charge on the gate.
P - Channel MOSFET field effect transistor
Filling in the holes, turns the transistor OFF.

N-channel MOSFET

Just like bipolar transistors, a complementary version that works with the opposite polarities can be made by replacing the P-type semiconductor with N-type.

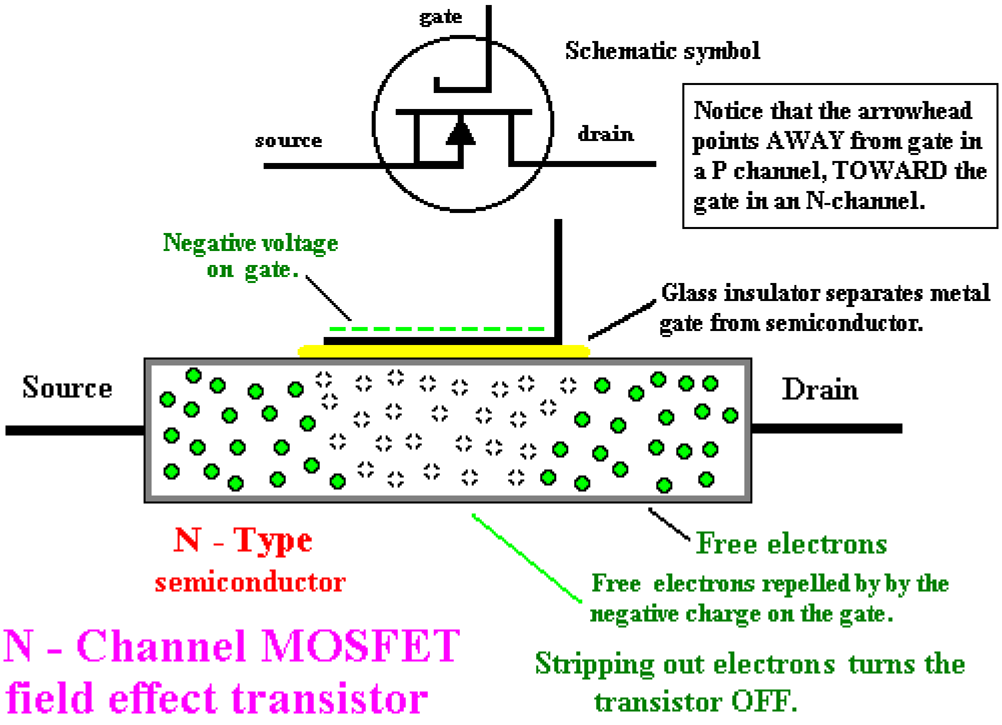
gate
Schematic symbol
source
drain
Notice that the arrowhead points AWAY from gate in a P channel, TOWARD the gate in an N-channel.
Negative voltage on gate.
Glass insulator separates metal gate from semiconductor.
Source
Drain
N - Type semiconductor
Free electrons
Free electrons repelled by by the negative charge on the gate.
N - Channel MOSFET field effect transistor
Stripping out electrons turns the transistor OFF.

Depletion-type and Enhancement-type MOSFETs

A simple MOSFET, like the one diagrammed above, is normally turned about half on. That is, when the gate voltage is zero, the transistor has a significant resistance, perhaps 300 ohms. This type of MOSFET is called a ***depletion-type.*** When one polarity of gate voltage is applied, the transistor turns full on. When the other polarity is applied, all the holes are filled in or all the free electrons are drawn out and the transistor becomes an insulator. That is, if the semiconductor is N-type, the extra electrons are pulled out of the crystal and the crystal becomes an insulator. Depletion MOSFETs are usually used in low power receiver applications.

Enhancement type MOSFETs

Enhancement type MOSFETs have been cleverly designed to be more convenient for typical power applications. When the gate voltage is zero with respect to the lead attached to the transistor semiconductor called "the source," the enhancement MOSFET is turned off. The SMP16P06 or the IRF9541 P-channel MOSFETs used in this book are enhancement types. This means that when the telegraph key is open, the QRP is fully off. Then when the key is depressed, the MOSFET turns full on and turns on the QRP. Power MOSFETs like these are also equipped with an internal diode to protect the transistor from being used with the wrong polarity. When the MOSFET is correctly biased, the protection diode looks like an open circuit and doesn't interfere. But if the polarity is reversed, such as when it is exposed to a ringing inductor, the diode turns on and shorts out the piece of semiconductor and protects it from being fried.

Compared to bipolar transistors, MOSFETs have two major advantages:

1. There are no PN junctions in a MOSFET. This means that there is no PN junction to breakdown and no PN junction to cause temperature sensitivity. Because big MOSFET transistors are hard to break, they are often the best choice in high power applications.

2. MOSFETs are a voltage-controlled device. Once a voltage is established on the gate, no current is required to maintain the ON state of the transistor. MOSFETs are often ideal for power supply applications and switching big DC currents. And, they can be used for high powered audio amplifiers. Power MOSFETs, like the P-channel used to key the QRP transmitter, aren't so well suited for use as a high frequency (RF) power amplifier. The capacitor input means that every time the input voltage changes, the capacitor input must be charged or discharged. For very high frequencies, the driving circuit needs to stuff current into the gate and pull it out tens of millions times per second. This makes driving a power MOSFET like the one above impractical for high power radio frequency transmitters.

For example, from time to time QST ham magazine has had transmitter plans using high power MOSFETs for 20 to 50 watt transmitters. The catch was, they didn't recommend them for hambands higher than about 40 meters. Above that frequency, the power needed to develop the drive current had to be so high, that the driver power began to be similar to the output power. In recent years however, special RF MOSFETs have been developed that can be used for transmitters running at hundreds of megahertz. However they are quite pricey and I haven't seen hamband transmitter designs using them.

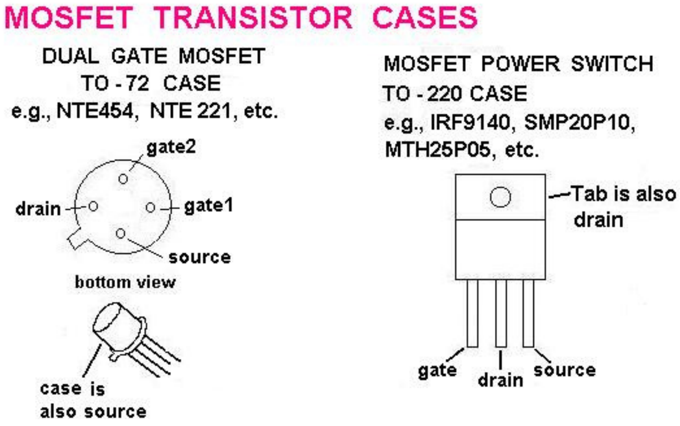

Power enhancement MOSFETs made by different companies have various designs that make them more rugged, handle larger currents, work at high frequencies, and dissipate more waste heat. Names like HEXFETs and V-MOS are examples of sophisticated power MOSFET designs. Dual gate MOSFETs are used to make mixers, product-detectors and RF amplifiers in receivers. As the name implies, they have two control gates instead of one. They will be discussed in chapters 7 and 13.

****

SPOT SWITCHES

Suppose you hear another ham calling CQ and you'd like to answer him. How can you tell whether your crystal and variable capacitor can reach his frequency? The answer is a "spot switch." The idea is to turn on the oscillator so you can hear your own transmitter in the receiver, but your signal will not be broadcast until you are ready to call the station. Using the spot switch, you turn on the oscillator and maybe the buffer, but not the amplifiers. This is done by separating the +12 volt power into the QRP into two wires, one for the oscillator and a second for the amplifiers.

In the circuit below, diodes are used to power the oscillator from either wire under the appropriate circumstances. Strictly speaking, the diode in series with the spot switch isn't really needed, because the open spot switch will turn the oscillator off while the transmitter is in stand by. However, the voltage supplied to the oscillator will be more equal if diodes are in both supply lines.

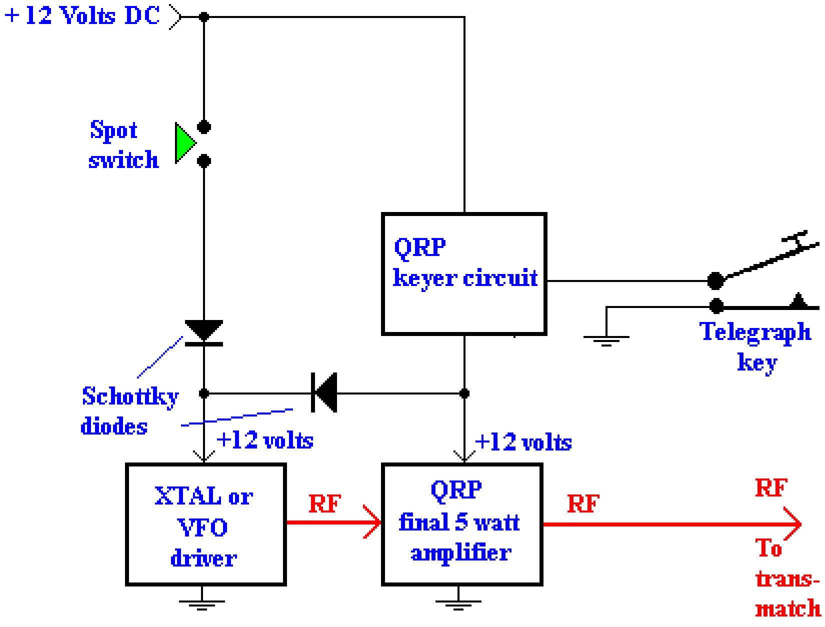
+12 Volts DC
Spot switch
QRP keyer circuit
Telegraph key
Schottky diodes
+12 volts
+12 volts
XTAL or VFO driver
RF
QRP final 5 watt amplifier
RF
RF
To trans-match

QRP KEYER WITH "SPOT" BUTTON

The "spot" momentary contact button turns on the oscillator and buffer stages with power through a Schottky power diode capable of handling 100 milliamperes or more. Then, when the telegraph key activates the final amplifier, the oscillator gets its power from a second Schottky diode connected to the P-channel MOSFET keyer.

I have spent hours trying to accomplish these functions without big MOSFETs and diodes, but I have yet to succeed. Have a go at it! Maybe you can be more clever than I was.

Building accessories and checking out your QRP

Building power supplies for your QRP module is explained in chapter 8. Other important accessories, such as a telegraph code key and an antenna coupler are discussed in Chapter 9. Checking out your QRP is also covered in chapter 9.

CRYSTAL SETS TO SIDEBAND
© Frank W. Harris 2010, REV 12

Chapter 7

A CODE PRACTICE RECEIVER

This chapter describes how to build a simple ***direct conversion*** ham receiver for the 40-meter CW band. This receiver can be used to listen to CW ham stations for code practice. It will also receive ham single sideband (SSB, voice) stations and foreign shortwave broadcasts. When combined with the CW QRP transmitter described in chapter 6, you could even use it to talk to other hams.

The main virtue of this receiver is its simplicity. The basic design has only 5 transistors and is an excellent first project for a new ham. It can be powered with a 9 volt battery. However a 12 volt battery made from small AA flashlight batteries will last much longer and provide more volume in your headphones.

ON/
VOL
700HZ

This receiver has good sensitivity and good stability. Unfortunately, by modern standards the selectivity of the basic receiver is poor for receiving CW. On many evenings the 40 meter CW band will be too crowded to listen to a single station easily. Adding the 700 Hz audio filter described later in the chapter will improve it greatly. Also, if you have a powerful AM broadcast station in your neighborhood, you may need to build a high pass filter to get rid of that interference. The broadcast filter is the little box shown to the right of the receiver.

When you're ready for a first rate receiver with selectivity as well as sensitivity, you may want to buy a commercial receiver or begin work on a superhetrodyne receiver as described in chapter 13. It is possible to get the same performance out of a direct conversion receiver, but that will require super-selective audio filters, image-canceling, and other R&D that I've never attempted. If you're interested, go for it!

Direct conversion receivers

A direct conversion receiver (DCR) has 4 basic circuit blocks. They are a band-pass filter, a variable frequency oscillator (VFO), a mixer or “product detector,” and an audio amplifier.

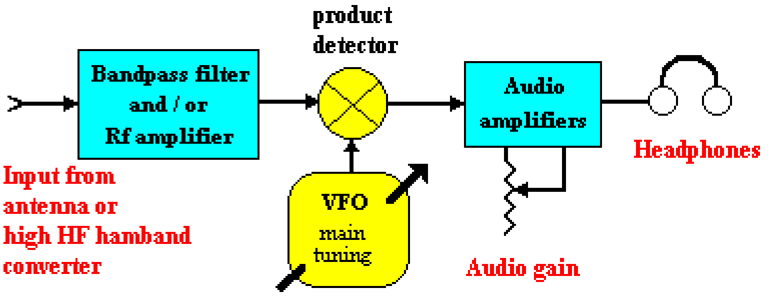
product detector
Bandpass filter and / or Rf amplifier
Audio amplifiers
Headphones
Input from antenna or high HF hamband converter
VFO main tuning
Audio gain

Direct Conversion Receiver

The frequency tuning of a DCR is accomplished with an adjustable frequency oscillator called a VFO (variable frequency oscillator). This sinewave signal is mixed with the incoming radio signals in a special amplifier called a ***product detector***. The signals heard in the headphones are those that differ from the VFO frequency by a difference within the audio range, say 20 Hz to 2.5 KHz. For example, suppose you are listening to a Morse code station on 7.100,000 MHz. You might tune the VFO to 7.100,700 MHz. That is, you tune 700 Hz above the actual transmitting frequency. The frequency difference will be heard as an audible, musical, 700 Hz Morse code tone.

The difference frequency is filtered, amplified, and passed on to the earphones - pretty simple radio! Considering how few parts it has, it's amazing how sensitive it is. I measured the performance of mine and found it could easily detect a 0.5 microvolt signal on 40 meters.

The input filter

The product detector has a weak diode-like characteristic. That means it is prone to act like a crystal set and detect strong signals present on the antenna that are not near the VFO frequency. Consequently, the signals coming in from the antenna should first be filtered to restrict signals to approximately 7 MHz. As you may recall from your experience with crystal sets, the tuning selectivity of a simple input LC filter is usually limited. However, a simple filter can get rid of background low frequency AM broadcast signals or strong foreign AM shortwave broadcast stations operating near 7 MHz.

40 Meter Hamband Passband Filter

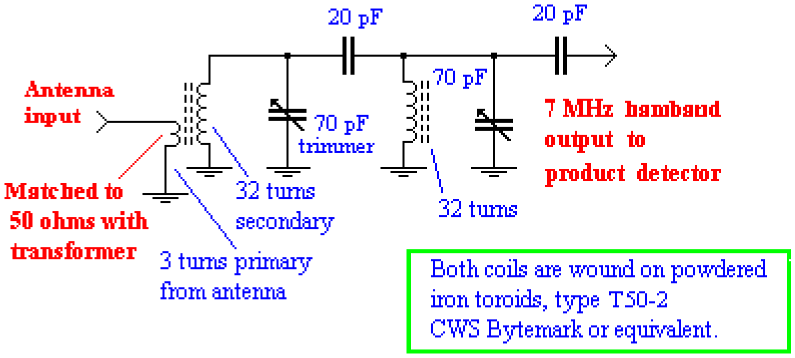
20 pF
20 pF
70 pF
Antenna input
70 pF trimmer
7 MHz hamband output to product detector
Matched to 50 ohms with transformer
32 turns secondary
32 turns
3 turns primary from antenna
Both coils are wound on powdered iron toroids, type T50-2 CWS Bytemark or equivalent.

The input filter consists of two tuned LC circuits. The input from the antenna is wired as an RF transformer that matches the high impedance filter and high impedance product detector with a low impedance antenna. In other words, the antenna is assumed to be approximately 50 ohms, like a typical ham dipole with a coax lead wire. The transformer steps up the RF voltage about ten times using the 1:10 turns ratio. If the receiver will be used with a high impedance long wire or whip antenna, the input lead should simply connect to the top of the trimmer capacitor through a small blocking capacitor.

While listening to CW signals on 40 meters, tune the trimmer capacitors until the signals are as loud as possible and any background voice stations disappear. If this filter is insufficient to get rid of local AM broadcast signals, a 1.8 MHz high pass filter is described later in this chapter. Also, you might try reducing the 20 pF coupling capacitors between the LC stages to 5 pF or less.

This input filter section could also include an RF amplifier stage. However, this would probably not be useful on 40 meters because atmospheric noise is relatively high below about 20 meters. On the other hand, if you would rather build this receiver for 20 meters or above, an RF amplifier will be a great help to sensitivity. RF input amplifiers are discussed in chapter 13.

The RF tuning oscillator (VFO)

The circuit below shows the VFO tuning oscillator. As shown, the inductor and capacitor values will tune the 40 meter ham band and the shortwave broadcast bands just above and below 7 MHz. The circuit is a tunable oscillator comparable in function to the crystal-controlled oscillator used in the QRP transmitter in chapter 6. Instead of a quartz crystal, the frequency control is a tunable LC circuit that has a range of hundreds of KHz. This VFO is too crude to control the frequency of a transmitter. It would drift too much and the guys you are talking to would complain about your unstable signal. However, because you will be listening to a KHz of audio at once, a drift of 50 Hz per minute will be scarcely noticeable. Building a first rate VFO for a transmitter is described in chapter 10.

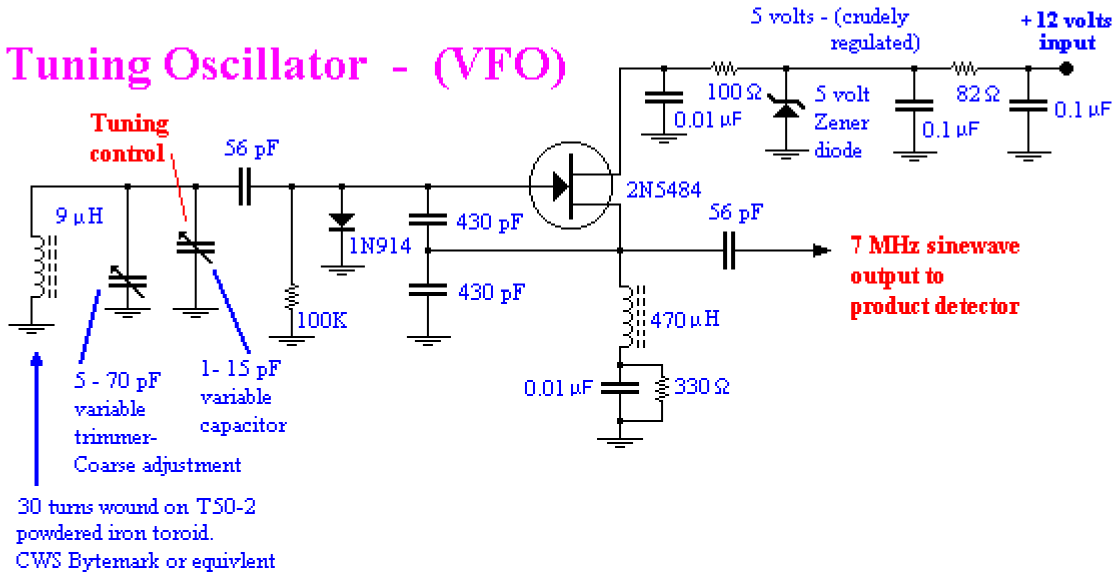
Tuning Oscillator - (VFO)
5 volts - (crudely regulated)
+ 12 volts input
Tuning control
56 pF
100 Ω
0.01 µF
5 volt Zener diode
82 Ω
0.1 µF
0.1 µF
2N5484
9 µH
1N914
430 pF
56 pF
7 MHz sinewave output to product detector
430 pF
100K
470 µH
5 - 70 pF variable trimmer- Coarse adjustment
1- 15 pF variable capacitor
0.01 µF
330 Ω
30 turns wound on T50-2 powdered iron toroid. CWS Bytemark or equivlent

JFET TRANSISTORS

This tuning oscillator uses a Junction Field Effect Transistor, (JFET). They are ideal for building VFOs. Unlike bipolar transistors, the main current from the JFET drain to the source does not pass through any PN junctions. PN junctions change their characteristics with temperature. Therefore VFOs made from bipolar transistors tend to drift more than JFETs. JFETs work on the same principle as a MOSFET transistor, but the control gate is a P-N junction diode rather than a tiny capacitor. Power MOSFETs were introduced in chapter 6 as a way to key the QRP power supply on and off.

P-N junctions

As you learned in chapter 4, ***a semiconductor P-N junction conducts when positive voltage is applied to the P-type semiconductor***. Conversely, it does not conduct when positive is applied to the N-type semiconductor side. In the crystal set, we were dealing with such tiny currents (microamperes) that the transition at zero volts was plainly visible. Commercial diodes handle much larger currents, milliamperes or even amperes. With commercial diodes the forward conduction doesn't fully start until a forward offset is reached. For an ordinary P-N junction silicon diode, this is around 0.6 volts. For a commercial Schottky diode, this offset is around 0.2 volts. This threshold means that if we apply a positive voltage to the P-type semiconductor side below the threshold, say, 0.1 volts, the current will be far less than a milliampere. Since very little current flows, the diode is still essentially "OFF."

P-N junctions can work like capacitors

In a MOSFET, gate current is prevented from flowing into the main channel by means of a layer of glass insulation. In a JFET, the gate appears to be a forward biased diode but doesn't conduct much current unless the drive exceeds the forward breakdown voltage, about 0.6 volts DC. ***With zero volts input on the gate, the JFET is already partly turned on.*** In other words, with zero volts on the gate the transistor is already operating as a Class A amplifier and, unlike a bipolar transistor, no bias resistor from the drain supply is needed.

JFETs operate at input voltages less than 0.6 volts

The JFET turns full on, (that is, it saturates) when the gate voltage approaches or exceeds 0.6 volts. Notice the diode clamp that prevents the gate voltage from ever exceeding 0.6 volts. To turn the JFET completely off, reverse voltage must be applied to the gate so that the charge carriers, holes and electrons, are completely depleted. For example, in the N-channel JFET shown below, the transistor turns completely OFF when the gate goes negative with respect to the source lead of the transistor.

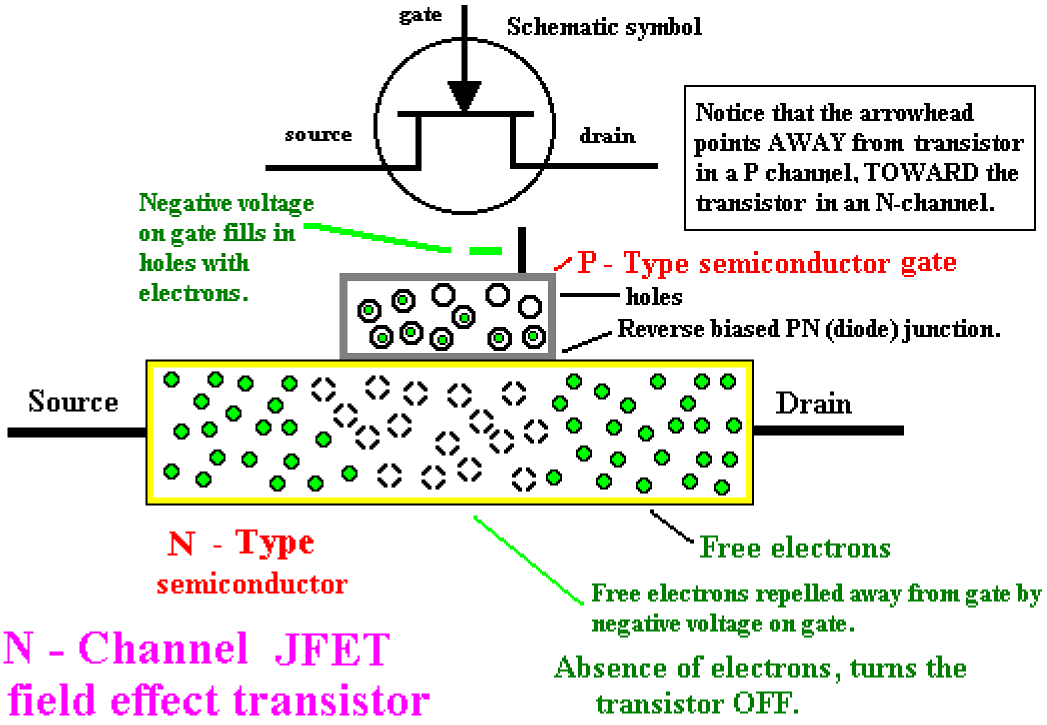
gate
Schematic symbol
source
drain
Notice that the arrowhead points AWAY from transistor in a P channel, TOWARD the transistor in an N-channel.
Negative voltage on gate fills in holes with electrons.
−
P - Type semiconductor gate
holes
Reverse biased PN (diode) junction.
Source
Drain
N - Type semiconductor
Free electrons
Free electrons repelled away from gate by negative voltage on gate.
N - Channel JFET field effect transistor
Absence of electrons, turns the transistor OFF.

As you would expect, P-channel JFETs are mirror images of N-channel JFETs with respect to construction and polarity. To turn a P-channel fully off, its diode gate is back biased with positive voltage as shown below.

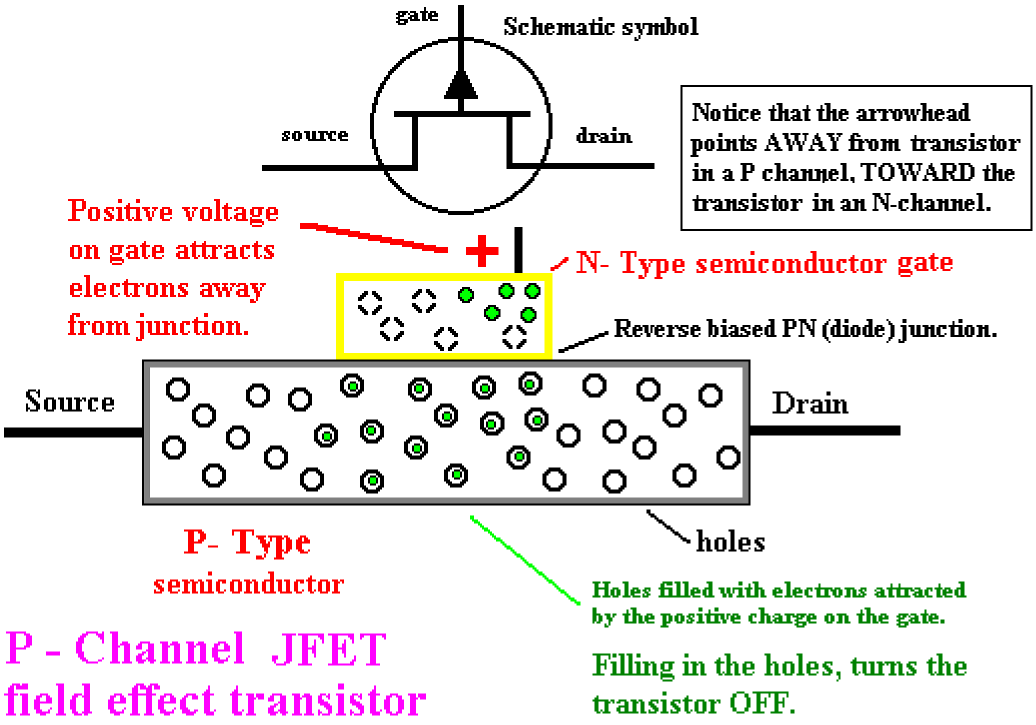
gate
Schematic symbol
source
drain
Notice that the arrowhead points AWAY from transistor in a P channel, TOWARD the transistor in an N-channel.
Positive voltage on gate attracts electrons away from junction.
+
N- Type semiconductor gate
Reverse biased PN (diode) junction.
Source
Drain
P- Type semiconductor
holes
Holes filled with electrons attracted by the positive charge on the gate.
P - Channel JFET field effect transistor
Filling in the holes, turns the transistor OFF.

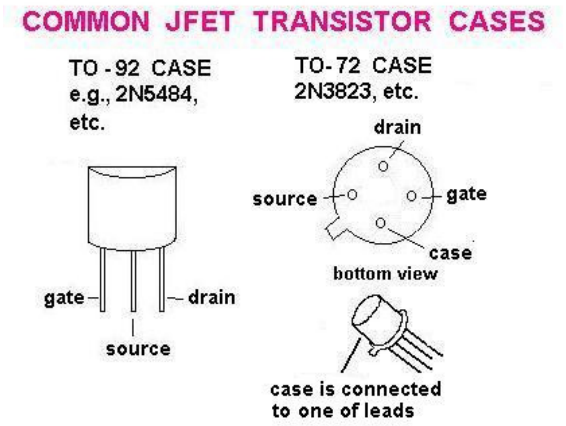
COMMON JFET TRANSISTOR CASES
TO-92 CASE
e.g., 2N5484,
etc.
gate
drain
source
TO-72 CASE
2N3823, etc.
drain
source
gate
case
bottom view
case is connected
to one of leads

The product detector

This receiver project uses a ***product detector***. The product detector amplifies and mixes the VFO signal with the signal coming in from the antenna. The resultant mixture of radio signals is amplified and sent to an audio amplifier. Since the only audio component present is the *DIFFERENCE* between the two radio signals, that's what the audio amplifier passes on to the earphones.

The product detector circuit described here uses a dual gate N-channel MOSFET transistor. A dual gate MOSFET works like a single gate MOSFET (or JFET). But as the name implies, it has two control gates modulating the drain-to-source current instead of just one gate. With two gates this MOSFET is convenient for mixing two separate signals because the output of the amplifier is an amplified mixture of the two inputs. Mixers and product detectors can be built with single gate MOSFETs and JFETs, but with a dual gate MOSFET, both inputs are amplified in one stage.

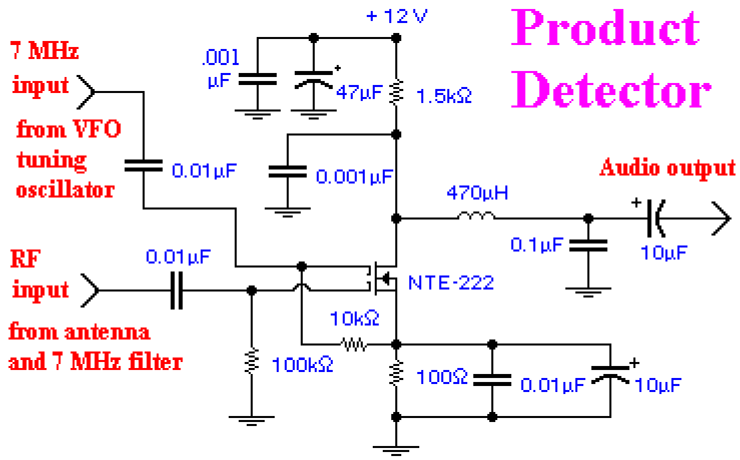
Product
Detector
+12V
7 MHz
input
from VFO
tuning
oscillator
.001
µF
47µF
1.5kΩ
0.01µF
0.001µF
470µH
Audio output
0.1µF
10µF
RF
input
from antenna
and 7 MHz filter
0.01µF
NTE-222
10kΩ
100kΩ
100Ω
0.01µF
10µF

The gates of the MOSFET are tiny capacitors and therefore have extremely high impedance. The amplifier design is a ***common source*** design, which implies that the source is essentially at ground. This also means the load resistance, 1.5 K ohms, is on the drain side (+12 volt side) of the transistor. Because the desired output is an audio frequency signal, the load impedance is a resistor, not an inductor. In order to work at audio frequencies, an inductor would have to be huge and for this application would have little advantage over a resistor. The 470 microhenry RF choke (inductor) and the 0.001 and 0.1 microfarad capacitors serve as a filter to remove the RF from the audio output. Without the filter, the audio amplifier would be more likely to rectify, (i.e. detect), strong RF signals that are able to get into the product detector and are not within the desired audio frequency difference from the VFO.

Modern designs often use integrated circuit mixer modules for product detectors. Personally, I prefer the dual gate MOSFET design simply because I know what's in there. It isn't just another mysterious integrated circuit. In this application any of the following dual gate MOSFET transistor types will work fine. I have successfully used the NTE221, NTE222, NTE454, or NTE 455. These are all "replacement" transistor types. So far as I know, any of the original 3N-series dual gates such as the 3N140 will also work. The 40763 by RCA is available from RF Parts Co. or from Mouser Electronics and should also work. I haven't personally tried it but I've seen it used in similar circuits.

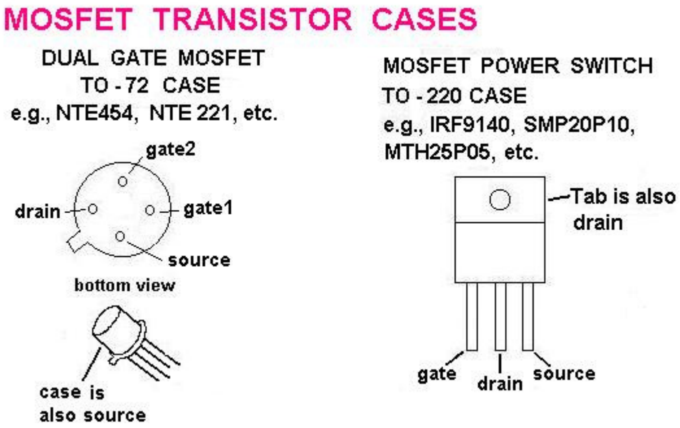
MOSFET TRANSISTOR CASES
DUAL GATE MOSFET
TO-72 CASE
e.g., NTE454, NTE 221, etc.
gate2
drain
gate1
source
bottom view
case is
also source
MOSFET POWER SWITCH
TO-220 CASE
e.g., IRF9140, SMP20P10,
MTH25P05, etc.
Tab is also
drain
gate
drain
source

A JFET product detector

Since this book was first written, dual gate MOSFETs have become increasingly expensive and hard to find, at least in my neighborhood. In the event that you need a substitute, here is a JFET circuit that works but, in my opinion, isn't quite as sensitive. It is practically the same circuit but instead of introducing the local oscillator signal into a separate gate, it is introduced across the source resistor.

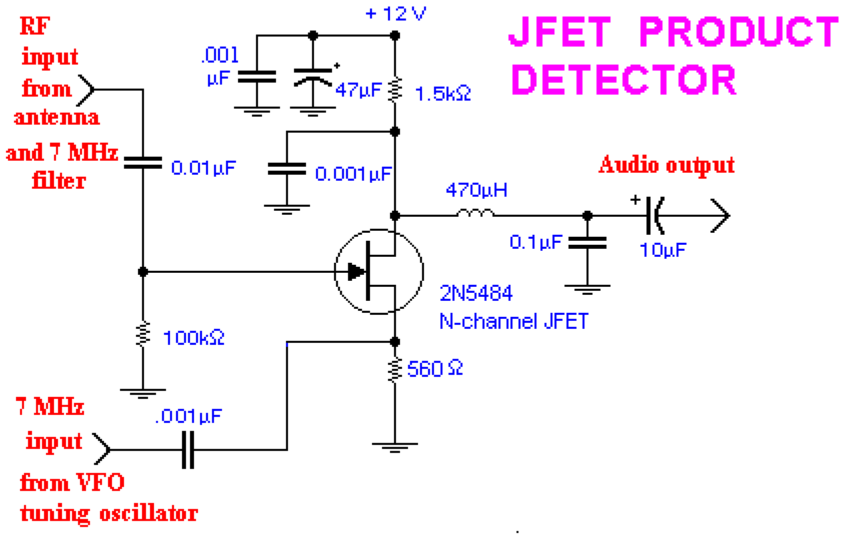
JFET PRODUCT DETECTOR
RF input from antenna and 7 MHz filter
+12V
.001 µF
47µF
1.5kΩ
0.01µF
0.001µF
470µH
Audio output
0.1µF
10µF
2N5484
N-channel JFET
100kΩ
560 Ω
7 MHz input from VFO tuning oscillator
.001µF

An article in the May 2008 issue of World Radio News describes a DCR similar to mine designed by Wayne McFee, NB6M. The interesting difference is that the JFET product detector was implemented with two JFETs in series. In that way both inputs, the VFO and the antenna as well, both go to gates and the output signals are added together, literally one on top of the other. The antenna amplifier JFET takes the place of the 560 ohm source resistor above. Since the McFee design has often been copied, I gather it works well, but I haven't tried it as a product detector. I did try to use it as a mixer in a superhetrodyne, but its performance was not significantly different than a single JFET mixer.

A generic R-C coupled audio amplifier

The output from the product detector is an audio signal that needs to be amplified before it goes to the earphones or speaker. Once again, most designs seen in ham magazines use integrated circuits marked "audio amplifier." For example, the LM386 is a typical one-chip audio amplifier. I've used these and they usually work well. Of course I didn't learn anything from the experience.

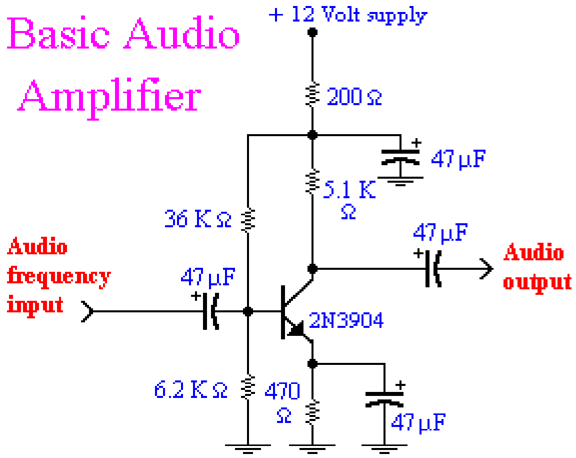

The diagram above shows a basic audio frequency ***RC coupled*** linear amplifier. If you look closely at the component values, you'll see it is similar to the class A RF amplifier we used as a buffer amplifier in the QRP transmitter in chapter 6. The differences are the large sizes of the capacitors and the absence of inductors. A 5.1 K Ω resistor is used as a load where you would expect to find an inductor in an RF amplifier. You could use an inductor here, but at audio frequencies it would have to be huge and expensive. Notice that the 36 K Ω resistor turns the amplifier partly on so that an audio sinewave will "fit" between zero volts and 12 volts. Otherwise the amplifier would only amplify positive signals larger than about 0.6 volts. Without the forward bias, the negative half of the sinewaves would be clipped off and the sound would be distorted. Because of the 0.6 volt base voltage offset, small signals wouldn't be heard at all.

The 470 ohm resistor biases the transistor, giving it a small amount of DC negative feedback. When the transistor conducts DC voltage, the voltage across this resistor in effect decreases the voltage on the base, thereby turning the transistor more off. This reduces the amplifier gain slightly, but makes the amplifier less likely to oscillate or run hot. The adjacent 47 µF capacitor allows the AC audio signal to go to ground without passing through the resistor. This allows us to have it both ways: The amplifier is more stable and the audio amplification is essentially unchanged.

The DCR audio amplifier

My complete audio amplifier is an extensively modified version of an example in the ARRL 1986 handbook. It is the same circuit used in the receiver in chapter 13. It looks like three straight-forward R-C coupled amplifiers in series. But the original design had extra filter components I didn’t quite understand. Every part that I didn’t understand, I left out. That's how I learn how circuits work. The audio amplifier was dead as a doornail when I first turned it on.

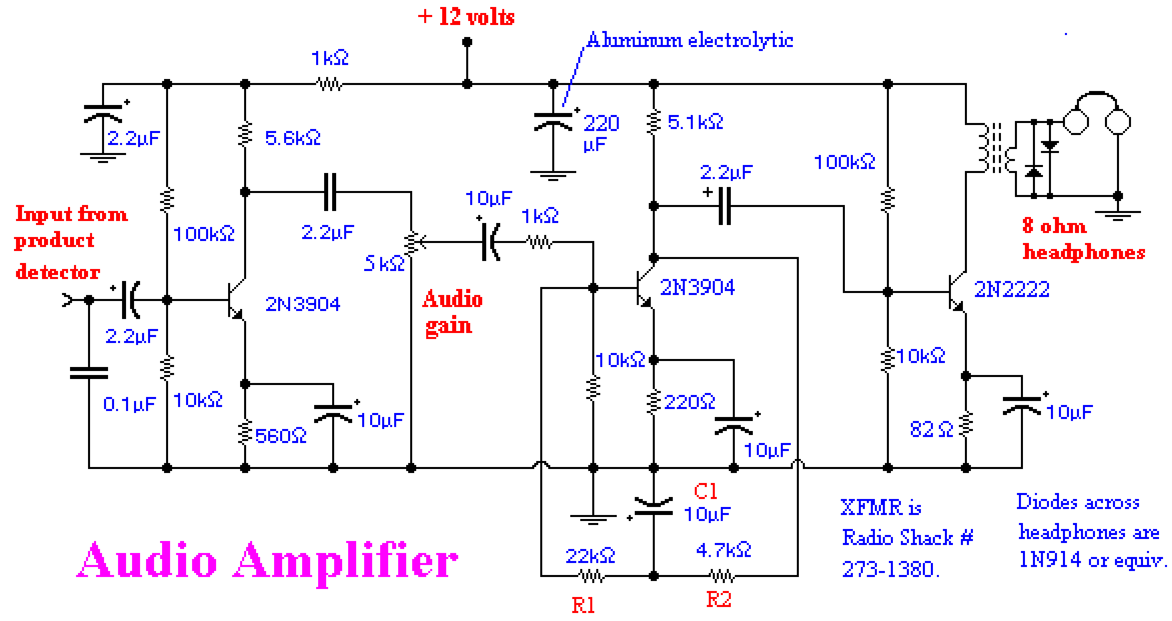
+ 12 volts
Aluminum electrolytic
1kΩ
2.2µF
5.6kΩ
220 µF
5.1kΩ
2.2µF
100kΩ
Input from product detector
100kΩ
2.2µF
10µF
1kΩ
5 kΩ
8 ohm headphones
2N3904
Audio gain
2N3904
2N2222
2.2µF
10kΩ
10kΩ
0.1µF
10kΩ
220Ω
10µF
560Ω
10µF
82Ω
10µF
10µF
C1
10µF
XFMR is Radio Shack # 273-1380.
Diodes across headphones are 1N914 or equiv.
Audio Amplifier
22kΩ
4.7kΩ
R1
R2

An audio Automatic Gain Control (AGC)

I was particularly puzzled by the low frequency feedback link, R1, R2, and C1. I couldn't understand what sort of "low frequency filtering" the designer was trying to accomplish. But, when I turned it on, the amplifier seemed completely dead. I put those mysterious components back in the circuit and ... voila! The earphones came to life. It turns out that this loop biases the amplifier "on" for weak signals and biases it "off" for loud signals. It's a kind of audio Automatic Gain Control (AGC) circuit.

Remember that for a bipolar transistor to turn on, the input signal must be greater than 0.6 volts or no current will flow into the base. In a "Class A" amplifier a small DC current is injected into the base. This increases the base voltage above 0.6 volts so that it's always turned on. The low frequency feedback adjusts the bias for weak and strong signals. When the signals are weak, the middle transistor is turned off, so it's collector voltage is high (+12 volts) and unchanging. This big collector voltage is leaked into C1 through R2. The voltage on C1 then biases the base through R1 to provide a forward bias for the transistor, turning it on and raising its sensitivity. Conversely, when the signals are strong, the collector has a big current flowing but a low average voltage from the collector to ground. This lower voltage biases the transistor more "off."

Protecting your ears from strong signals

This triple amplifier has enough gain to almost deafen you when you encounter a strong signal. Therefore it's essential to add a clamp circuit to limit the voltage to the headphones to less than about one volt. In practice, with sensitive, modern 8 ohm headphones, I found that less than one volt peak is plenty of volume for me. Eventually I put in two ordinary silicon 1N914 diodes "shorted" with opposite polarities across the headphones. This limits the positive and negative sound peaks to just 0.6 volts and my ears have been adequately protected.

How Hi-Fi should it be?

The original circuit was also sprinkled with 0.1 microfarad bypass capacitors as if the designer were trying to kill all higher frequency sounds and shunt most of the audio to ground. Since I was worried about having enough gain, I left out the bypasses. The amplifier worked well without them, but the sound of the static had an obnoxious, piercing high pitch that irritated my ears. I put the bypass on the input back in and, as I expected, the audio sounded more “bass” and became somewhat weaker. However, getting rid of that piercing, hissing static was well worth the loss of gain. Experiment!

The original design also had no emitter bypass capacitor. That's the 10 µF capacitor across the 220 ohm resistor. Not having this bypass capacitor reduces the gain because some of the audio voltage signal is wasted across the 220 ohm emitter resistor. Since I didn't want to waste any gain, I put in the capacitor and my gain jumped up noticeably. This bypass has no disadvantage that I could detect.

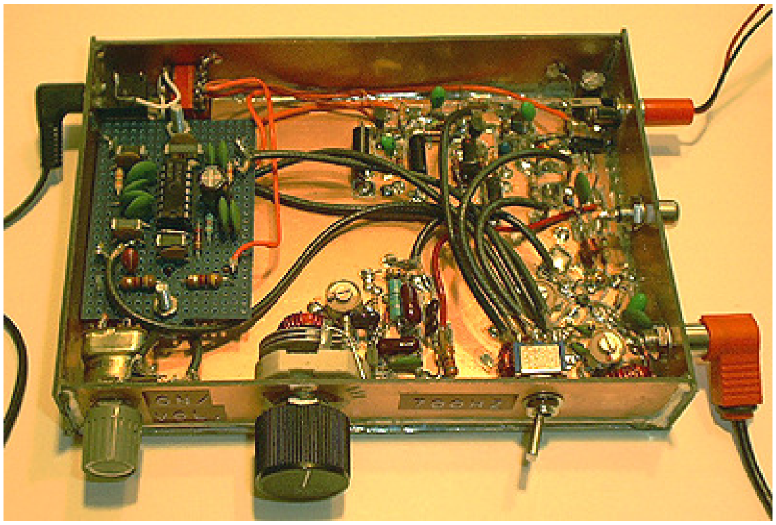

Mechanical construction

I built my receiver in a box made from soldering pieces of two-sided PC board together. This is the same construction described for the QRP transmitter described in chapter 6. The black coax RG-174 cables running around the inside of the box connect the optional 700 Hz filter module described later in this chapter. The coax shields the wires from stray RF that might be present. Only one end of the coax should be grounded.

The tuning knob ideally should be a mechanical, vernier, planetary-gear device to spread out the 40 meter CW band and make the signals easier to tune. Alternatively, one can use two capacitors in parallel with a smaller variable capacitor, say 3 to 15 picofarads (pF), for fine tuning. In this way, out of the 180 degrees of useful knob rotation, more degrees can be devoted to the CW band and less to the nearby broadcast stations. A 70 pF trimmer capacitor tunes the whole frequency spectrum around 40 meters to align the small capacitor to the ham band. The low capacitance variable capacitor is called an ***electrical bandspread.*** In my receiver the 15 pF fine-tuning capacitor has the big black knob, while the physically tiny 70 pF ceramic screw adjust trimmer is directly behind it.

The receiver power supply

The receiver will work well on 9 volts at about 10 milliamperes. However, when the voltage drops below about 8 volts, the volume and sensitivity will fall off dramatically. So, if you use a small alkaline 9 volt "transistor radio" battery, it will only work well for about 1/3 of the energy stored in the battery. ***To get the most from alkaline batteries, you need to be able to run them down to 2/3 of their original voltage.*** In this case the receiver would have to work well down to 6 volts. I recommend making up a 12 volt battery out of AA cells. Radio Shack and other companies sell battery holders that carry 6 or 8 cells to give you 9 or 12 volts. Or if you are using the receiver with the QRP transmitter described in the last chapter, the receiver can share the QRP power supply.

Hearing the transmitter VFO

When I was trying to use the DCR with my QRP transmitter, I found I couldn't hear my own transmitter VFO (crystal oscillator) over the stations I wanted to call. To fix that, I ran a coax cable over to the QRP and imported a weakly-coupled transmitter crystal oscillator signal into the receiver box. The center connector on the right side of the receiver is for that purpose. The receiver has no direct connection to the transmitter oscillator. The connector merely projects a small piece of wire into the receiver box like an antenna to broadcast the weak oscillator signal.

****

FIXING THE LIMITATIONS OF DIRECT CONVERSION

One of the DCR limitations I had read about did not need to be fixed. Because a DCR has nearly all its gain in the audio amplifiers, these amplifiers are supposed to be prone to ***microphonics***. In other words, DC receivers are supposed to be sensitive to vibration. When you touch a knob or bump the table, you are likely to hear that sound amplified in the earphones. I observed no microphonics at all. I suspect microphonics may have been a common problem when using vacuum tube amplifiers.

A DCR detects both sidebands at once

Unfortunately, the product detector mixer detects both the upper and lower sidebands simultaneously. This is OK for listening to AM radio stations that broadcast both sidebands, but it has poor selectivity on a crowded CW band. A good CW receiver can select a bandwidth as small as 500 Hz or less. A simple direct conversion receiver will have a bandwidth of 10,000 Hz or more, depending on the audio filtering.

Although audio filtering is helpful for a DCR, the big problem is that you'll still be listening to audio from both "sidebands" at once. CW by definition is extremely narrowband. It covers a few Hz at most. Therefore, CW doesn't really have sidebands. However, to hear the signal as a musical tone, a receiver must be tuned either above or below the actual frequency by an amount equal to the frequency of the desired musical tone. Most modern hams are using transceivers that automatically listen about 700 Hz above or below their transmitting frequency. This offset frequency is adjustable, but is nearly always set between 500 to 1000 Hz. By convention, ***On 40 meters and below, they are listening BELOW the transmitting frequency.*** On HF bands 30 meters and above hams listen ABOVE the transmitting frequency. This convention has its origin in the method for generating single sideband phone that was formerly common 40 years ago. The old SSB design isn't used anymore, but the convention lives on. The new 60 meter band is specified for upper sideband only and is an exception. The result is that if you answer a station on the upper sideband side while the station is listening on the lower sideband, he will never hear you because modern equipment only hears one sideband at a time.

Sharp audio filters aren't enough

A sharp audio filter on the DCR output does not correct this problem. For example, suppose you build a sharp audio filter to select 700 Hz audio tones only. Yes, you'll hear your 700 Hz offset CW station just fine. But you're likely to be confused by an ***image*** signal that you can also hear 1.4 KHz away on the opposite sideband. That is, 2 X 700 Hz = 1.4 KHz. When you answer a CQ, this confusion will make it hard for you to know where your contact will be listening. Remember, when he switched his commercial transceiver to that band, he selected upper or lower sideband and didn't have to worry about this effect.

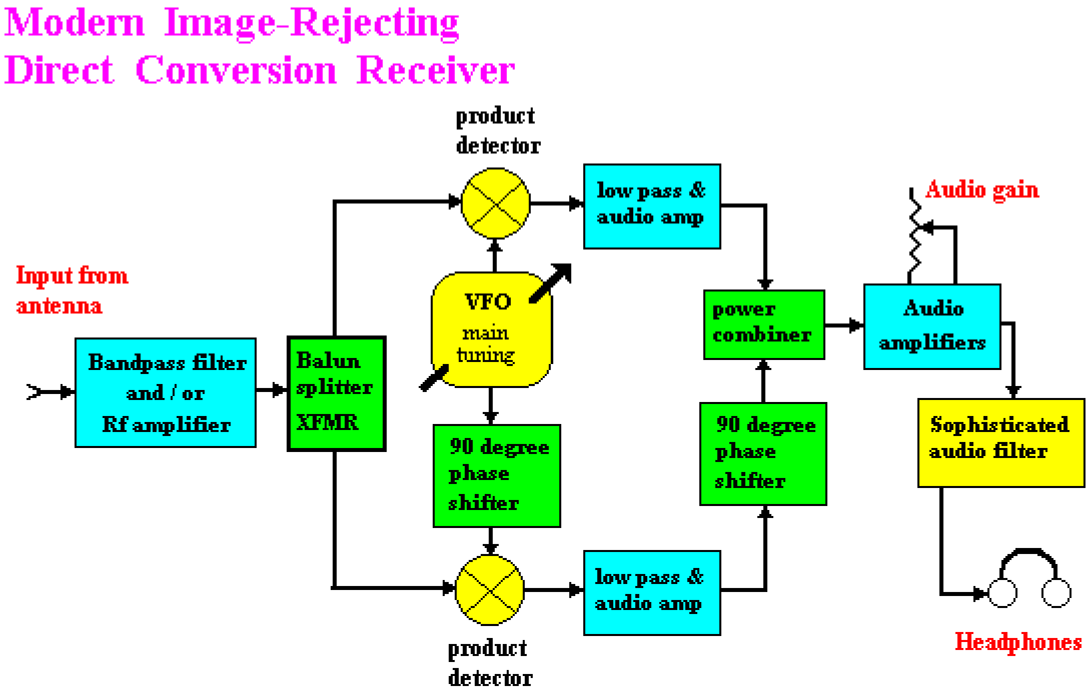
Modern Image-Rejecting
Direct Conversion Receiver
product detector
low pass & audio amp
Audio gain
Input from antenna
VFO
main tuning
power combiner
Audio amplifiers
Bandpass filter and / or Rf amplifier
Balun splitter XFMR
90 degree phase shifter
90 degree phase shifter
Sophisticated audio filter
low pass & audio amp
product detector
Headphones

A sophisticated direct conversion (DC) receiver eliminates one sideband by essentially building two DC receivers. One sideband is cancelled out by phase shifting the input VFO RF signal and also phase shifting the resulting audio signal to remove one sideband. So, although a DCR is "simple," you might need to build two of them, plus a sophisticated audio filter. Personally, I decided that a dual conversion superhetrodyne was the best way to go.

AM broadcast interference

In my city there is a powerful AM radio station that broadcasts at high power during the day. When I put an oscilloscope on my 40 meter dipole lead wire, I see a 1.75 peak volts RF signal on the wire even with a 50 ohm resistor connected to the antenna. This represents 31 milliwatts of power! No wonder my crystal set was so loud.

Power = Voltage (RMS) squared ÷ load resistance = V_{RMS}^2 /R

RMS voltage = Voltage peak x (0.707)

Power = $[(1.75 \text{ volts peak}) \times (0.707)]^2 \div 50 \text{ ohms} = 31 \text{ milliwatts}$

The ham signals I was trying to listen to were buried in the voltage storms from the local AM rap-music station. The input filter for the DC receiver described earlier in this chapter was apparently inadequate to get rid of such a large signal. So, although I could hear ham stations, in the background I could always hear the rap music leaking through like a crystal set. I needed a better filter. The good old '86 ARRL handbook had such a filter design. It worked the first time without tweaking and reduced the RF voltage coming in on my antenna down to about 0.15 volts peak without affecting the strength of the 40 meter signals.

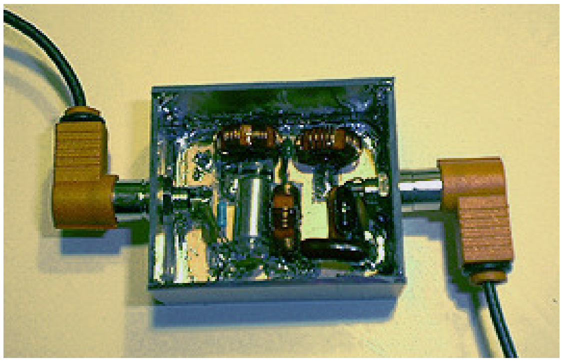

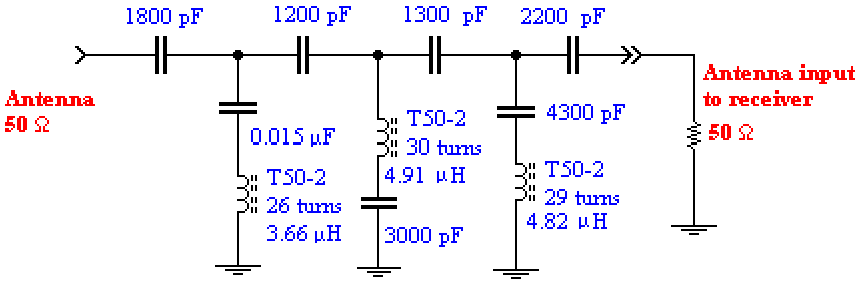

An AM broadcast filter

A 700 Hz audio filter – why you need one

As described earlier, if you intend to answer a fellow calling CQ, you must be sure you are listening on the same side of his transmitter frequency that he is. Because your direct conversion receiver has no image canceling, you will hear his Morse code both above and below his actual transmitting frequency. If you zero in your transmitter on the wrong side of his signal, your transmitter will be 1.4 KHz (2 X 700 Hz) away from where he is listening and he'll never hear you. Unless of course, his receiver is as primitive as yours.

For example, on 40 meters, as you tune ***UP*** the band you'll first hear his Morse code on the low side of his actual frequency. That is the lower sideband. If you continue tuning *VERY* carefully, the tone you hear will fall to a low pitch, then disappear, then rise back up to the same musical tone again. That is the upper sideband.

If you wish to answer the fellow's CQ, listen on the fellow’s lower sideband on 40 meters. Now set your transmitter to "spot" and tune your transmitter's crystal-pulling capacitor ***DOWN***-frequency until you first hear an equal tone. If you have a little musical talent, you should be able to match the tones pretty well. ***In summary, when listening for CQs, tune UP the 40 meter band. When spotting your transmitter oscillator on top of the station you wish to call, tune the transmitter DOWN the band.*** If you were on 30 meters or above you would reverse these instructions, because upper sideband is the convention up there.

Those darn modern hams are listening with exquisitely narrow audio filters. They will usually only hear you if your signal is making a tone within a couple hundred Hz above or below their listening frequency. When they switch in the really sharp digital filters, they only hear stations within a few Hz of their listening frequency.

An analog audio filter

If we had ideal, small cheap components, the easy way to build analog audio filters would be to use inductors and capacitors the same way we build radio frequency filters. We might even imagine that there would be audio frequency quartz crystals for building super precise audio filters.

Unfortunately, in the real world those parts don't exist. In order to work at such low frequencies, the inductors would have to be gigantic and expensive. Since that approach isn't practical, we use resistor/ capacitor networks. ***These RC networks don't resonate, but rather they just attenuate some frequencies more than others.*** A large capacitor charges more slowly than a small one. When combined with a resistor, the frequency of the sinewave across either the capacitor or the resistor will be attenuated at either the high frequencies or the low frequencies, respectively.

700 Hz Low Pass & High Pass Filters

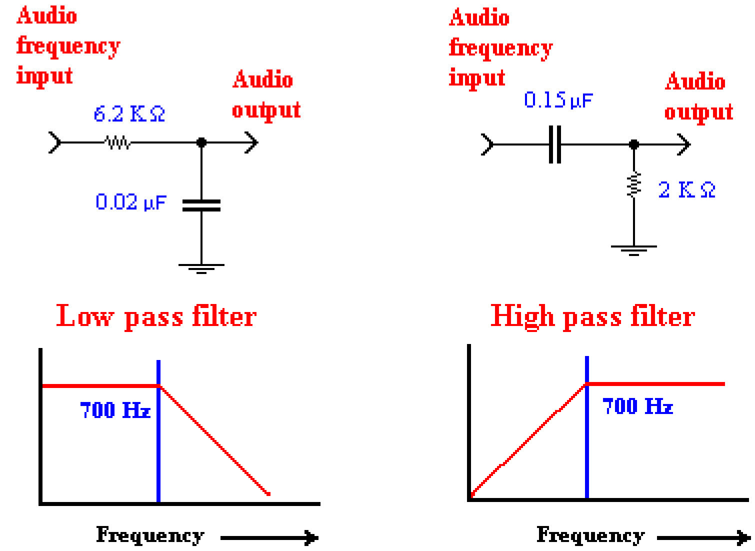

Low pass filter

As shown above, a simple resistor and capacitor makes a low pass filter when the output is taken off the capacitor. If the capacitor is large, relatively high frequency sinewaves will be "shorted" to ground. Remember, ***the voltage across a capacitor can't change instantly***. In contrast, the effect of smaller capacitors will be negligible to low frequency sinewaves that have time to charge the capacitor during each half cycle. The resistor loads down the circuit and draws current into the capacitor. Similarly, during the next half cycle, the resistor has sufficiently low resistance to fully discharge the charged capacitor and get ready for the next half cycle.

High pass filter

When the R and C are reversed and the output is taken off the resistor, then the RC filter becomes a high pass filter. High frequencies pass right through a capacitor to the output as if it weren't even there. Remember, ***the voltage across a capacitor cannot change instantly.*** Therefore, to high frequencies the capacitor looks like a wire with zero AC voltage drop across it. At high frequencies there isn't time to charge. At low frequencies, the capacitor charges up completely and the AC voltage drop across the capacitor approaches the entire voltage of the input sinewave. These ***RC coupled filters*** all have a particular frequency called the ***break point*** where the capacitor becomes relatively insignificant in the circuit. A simple RC filter attenuates the sinewave voltage ten times (20 decibels) for a tenfold change in frequency away from its break point.

Compensating for attenuation

A problem with RC filters is that the resistors attenuate ALL the frequencies to some degree, not just the ones we don't want. Fortunately if an RC network is combined with an audio frequency linear amplifier, we can compensate for the attenuation. By amplifying the output, we can restore the signal strength of the desired frequencies back to its original strength.

Bandpass filter = RC networks plus amplifier

To make a bandpass filter we need to combine a low pass filter with a high pass filter and then amplify the result. The circuit below is a one-stage R-C (resistor/ capacitor) audio filter/ amplifier that accentuates 700 Hz tones and attenuates tones above and below roughly 600 and 800 Hz.

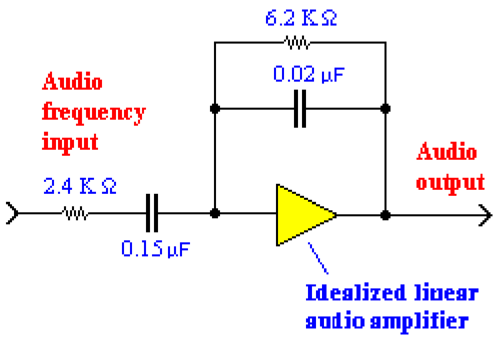

Single Stage 700 Hz Bandpass Filter

In this filter the Rs and Cs are not arranged as you might expect. However, if you carefully analyze how they affect signals passing through the circuit, they work the same as the separate circuits described earlier. That is, low frequencies and high frequencies are attenuated and the desired 700 Hz frequency range is relatively unaffected. The input RC, the 2K Ω and 0.15 µF capacitor, attenuate low frequencies because low frequencies, (for example, 60 Hz and direct current), can't get through the capacitor. Therefore the input network is a high pass filter.

Note that the 6.2K Ω ohm and 0.02 µF capacitors are wired between the input and output of the linear amplifier. This negative feedback network is the low pass filter. This combination of components tends to "short out" the amplifier and attenuate everything above about 700 Hz. Note that the small 0.02 µF capacitor has little effect on low frequencies because it charges essentially immediately. Because the capacitor charges almost instantly, its voltage follows the low frequency sinewave curve without significantly attenuating it. However, at high frequencies a 0.02 µF capacitor becomes charged more slowly than the high frequency sinewave is changing. Between half cycles, the 6.2K Ω ohm resistor has low enough resistance to be able to discharge the 0.02 µF capacitor significantly so that the network is restored to block the next half cycle.

The two resistors, the 2.4 K Ω and 6.2 K Ω, also control the gain in this circuit. The triangle amplifier symbol is assumed to have infinite gain. It doesn't of course, but this symbol generally means an operational amplifier, which is supposed to act like a perfect transistor with unlimited gain. The gain is set by the external resistors. There's no need to go into the details, but, the ratio of the feedback resistor to the input resistor is 6.2 K / 2.4 K = 2.5 times. That means, the gain is approximately 2.5. To say it another way, the voltage lost across the 2.4 K resistor while entering the amplifier is compensated by the much larger voltage gain across the 6.2 K. This more than makes up for the signal voltage lost in the filtering process.

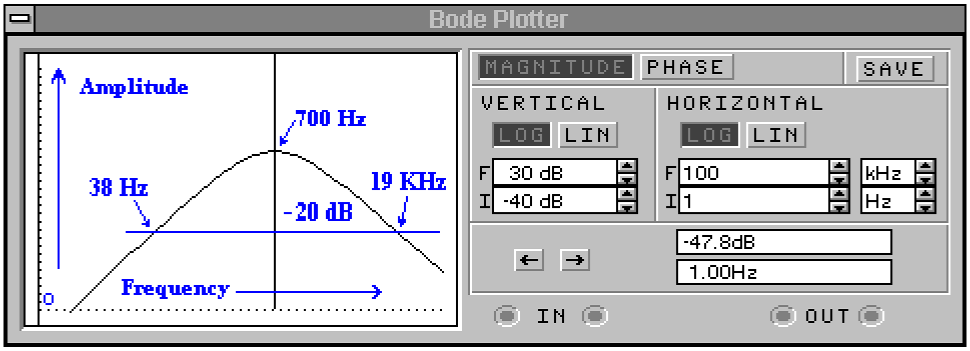

VERTICAL: LOG
F 30 dB
I -40 dB

HORIZONTAL: LOG
F 100 kHz
I 1 Hz

-47.8dB
1.00Hz

Single Stage 700 Hz Audio Filter

The frequency plot above shows the response of the single stage filter. As you can see, this filter is pretty crude. In order for the voltage of a sinewave to be attenuated more than 10 times (20 decibels), the frequency must be below 38 Hz or above 19 KHz. To be blunt, this won't help much. It will decrease some of the unpleasant high frequency noise from static, but won't be particularly helpful in relieving the QRM (interference from other stations).

More precise filtering can be achieved by putting multiple audio filters in series. The plot below shows four of the same filters in series. This is a big improvement and makes it easier to listen to just one station at a time. It also eliminates the high frequency noise and makes listening less tiring.

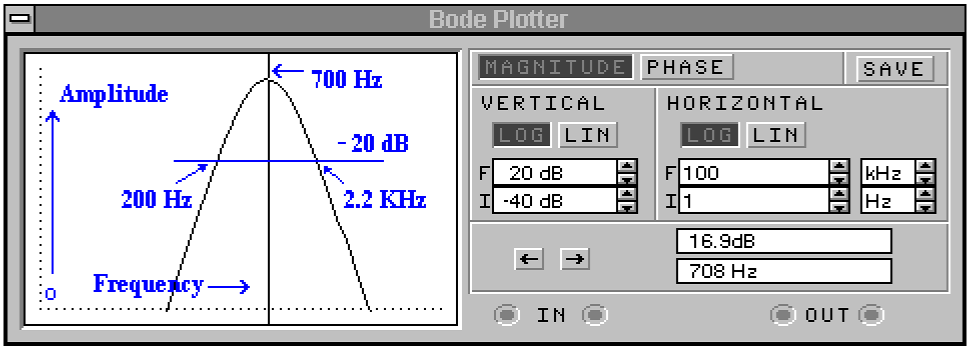
Bode Plotter
Amplitude
700 Hz
- 20 dB
200 Hz
2.2 KHz
Frequency
0
MAGNITUDE
PHASE
SAVE
IN
OUT

VERTICAL LOG LIN
F 20 dB
I -40 dB
HORIZONTAL LOG LIN
F 100 kHz
I 1 Hz
16.9dB
708 Hz

Four Stage 700 Hz Audio Filter

Implementing the filter

We can build filters that work just like the graphs above using transistor linear AF amplifiers made from discrete components. Here is what a single stage 700 Hz filter would look like:

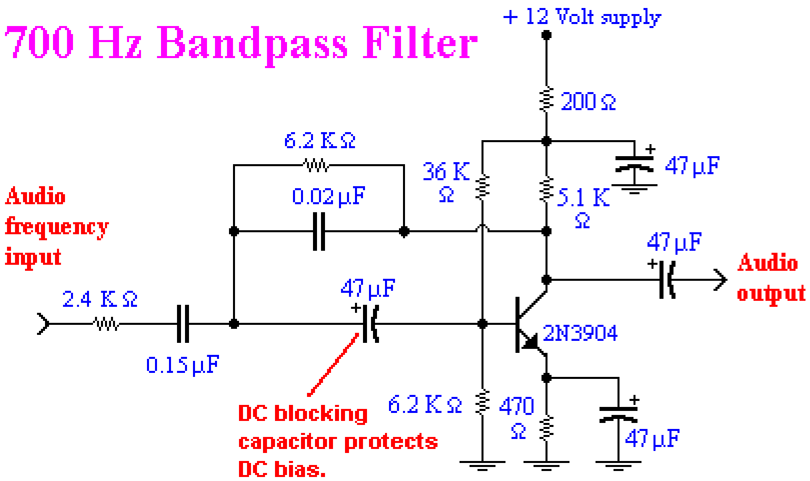
+ 12 Volt supply
200 Ω
6.2 K Ω
36 K Ω
47 µF
0.02 µF
5.1 K Ω
Audio frequency input
47 µF
Audio output
2.4 K Ω
47 µF
2N3904
0.15 µF
6.2 K Ω
470 Ω
47 µF

700 Hz Bandpass Filter

DC blocking capacitor protects DC bias.

Operational amplifiers – "perfect" amplifiers

Notice the big blocking capacitor needed to prevent the DC bias for the class A amplifier from being changed by the 6.2 K ohm resistor. In any case, normal non-eccentric hams do not make audio filters using linear amplifiers made from discrete parts. Instead, they use integrated circuit operational amplifiers. Being obsessive-compulsive, I thought about wiring all those bias resistors and 47 µF capacitors. Then I decided, what the heck! I'm checked out on audio amplifiers. I'm going to use a quadruple op-amp integrated circuit.

Op-amps are integrated circuits containing 12 to 30 transistors and resistors in a network that produces a "perfect" or idealized amplifier. How much gain should a "perfect" amplifier have? The best answer is that the gain should be infinite and then feedback can be used to produce any gain you like. Therefore real operational amplifiers have ***near infinite gain***, and ***near infinite input impedance,*** and a ***voltage source output***.

Typically there are two or four operational amplifiers on a single chip. The symbol for an operational amplifier is a triangle. The typical op-amp has an output pin and two input pins. For example, look at pins #1, # 2, and #3 below. A ***positive*** voltage on the ***positive input*** pin makes the output pin voltage rise, that is, go positive. In contrast a ***positive*** voltage on the ***negative*** input pin makes the output pin go down. ***Because of the feedback, the op-amp will "attempt" to drive its output pin up or down until feedback from the external circuit components cause the two inputs to be equal in voltage.*** If the feedback is insufficient to produce equilibrium, the output will bump into either ground or the voltage supply, whichever happens first.

To bias an operational amplifier so that a large sinewave will "fit" between zero volts and 12 volts output, the amplifier must have a baseline about halfway between the supply voltage and zero. In this case the resting level will be about 6 volts. In the filter below, the positive input is connected to a 6 volt reference level made from a voltage divider connected between 12 volts and ground.

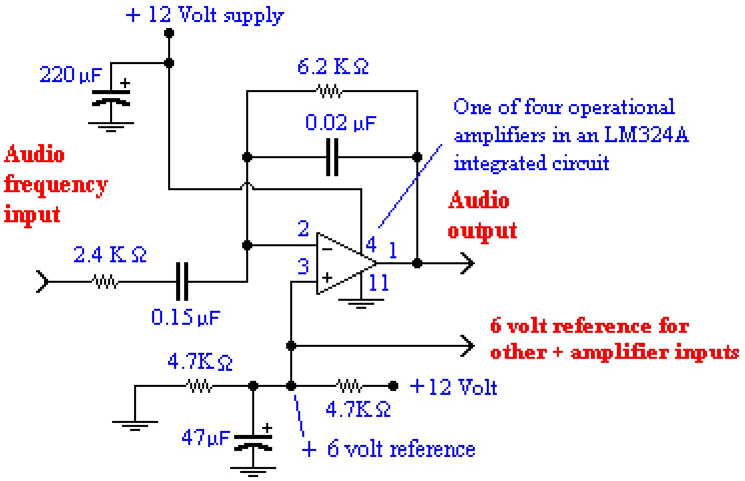
+ 12 Volt supply
220 µF
+
6.2 KΩ
0.02 µF
One of four operational amplifiers in an LM324A integrated circuit
Audio frequency input
Audio output
2.4 KΩ
2
3
4
1
11
0.15 µF
6 volt reference for other + amplifier inputs
4.7KΩ
+12 Volt
4.7KΩ
47µF
+
+ 6 volt reference

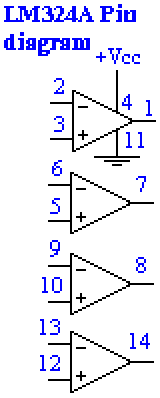
LM324A Pin diagram
+Vcc
2
3
4
1
11
6
5
7
9
10
8
13
12
14

Single Stage 700 Hz Bandpass Filter

Homebrewing with integrated circuits

Homemade "gouge boards" described in chapter 6 are usually much too crude to use with integrated circuits. Therefore when I build with ICs, I use "perf-boards." These are plain, 1/16 inch thick fiberglass boards. They have no copper layers but have a grid of tiny holes drilled on 1/10 inch centers. The older ICs have their pins spaced 1/10 inch apart and the two rows of pins are spaced apart in even multiples of one-tenth inch. Modern ICs are often available either as old-style, through-hole or modern surface-mount. To use through-hole, you could just push them through the perf-board and then solder your components and wiring on the pins on the other side.

I prefer to ***use IC sockets*** rather than hard solder them on. If you accidentally damage an IC and it is installed with an IC socket, it can be instantly replaced. Without a socket, you'll seriously damage your wiring when you tear it apart to solder in a new one. The IC socket also has pins lined up just like an IC, so you push it through the board and solder your wires to the socket pins on the bottom.

For homebrew work you need the big, old time ICs with 1/10 inch spacing. Modern surface mount ICs have pin spacing as tiny as 20 thousands of an inch. They are designed to be installed with mass production robots and are extremely tricky to solder in your basement. One practical (but extremely time-consuming) home-brew approach is to make little etched circuit boards that accept the surface mount ICs. The little adapter boards have traces that serve to expand the pin lead spacing radially out to large pads that human fingers can work with directly. Using hemostat clamps, tweezers, and a microscope, you solder the IC onto the miniature board. After you have inspected your pin connections under a microscope, you can then solder these adapter boards onto a normal size perf-board.

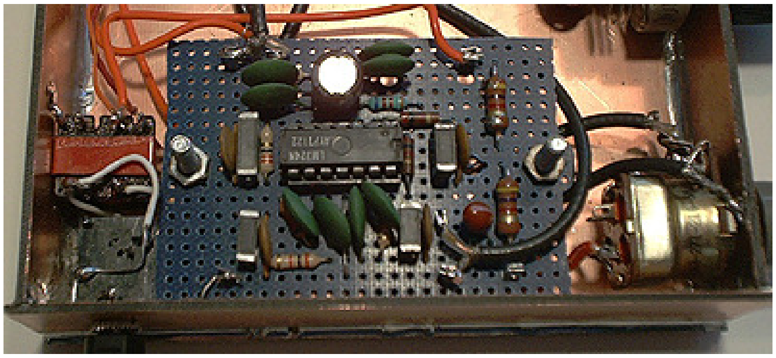

A four-stage 700 Hz bandpass filter built on a perf-board.

Using the diagram for the single stage filter shown earlier, simply feed one output into the next input until you have four stages in series. All four op-amps can share the same 6 volt resting level reference.

Label your pins!

Keeping track of the pin numbers isn't easy unless you label them. I like to cut out a narrow strip of sticky white label paper and paste it between the pins. I write the pin numbers on the paper with a fine felt-tip pen adjacent to the respective pins. ***Pin # 1 is at the end of the integrated circuit that has the notch.*** Sometimes pin # 1 is indicated with a little number or a round indentation in the black plastic. The notch and indentation may be seen in the above photo. In any case, the pin numbers seen from the top are numbered counterclockwise. However, as seen from the bottom of the IC socket where you do your soldering, ***the pin numbers go around clockwise starting from the end with the notch in the plastic case.***

There are two kinds of IC sockets suitable for ham work - ***solder tail*** and ***wire wrap***. Wire wrap sockets have big tall pins that allow thin wires to be wrapped around them with a special tool. I prefer the solder-tail type. They are harder to work with, but aren't so tall, are less messy and don't take up so much room. Just to cover the subject of IC sockets, large complex computer ICs use "grid sockets" and "solderball sockets" with hundreds of contact points. Unfortunately, working with these is nearly impossible at home because that would require extreme precision and building multi-layer PC boards.

Wiring the 700 Hz filter into the circuit

I designed the op-amp filter to have a slight amount of amplification. Unfortunately, by the time I put four in series, I had as much gain as one of the audio amplifier stages. Therefore, the output of my filter goes to the input of the second AF amplifier stage and bypasses the first stage. When I returned the signal to the input of the 1st stage, it was much too loud and unstable.

You could bypass the 2nd stage, but you would lose your AGC. A block diagram of my receiver is shown below. A double-pole, double-throw DPDT switch substitutes the filter for the 1st stage AF amplifier.

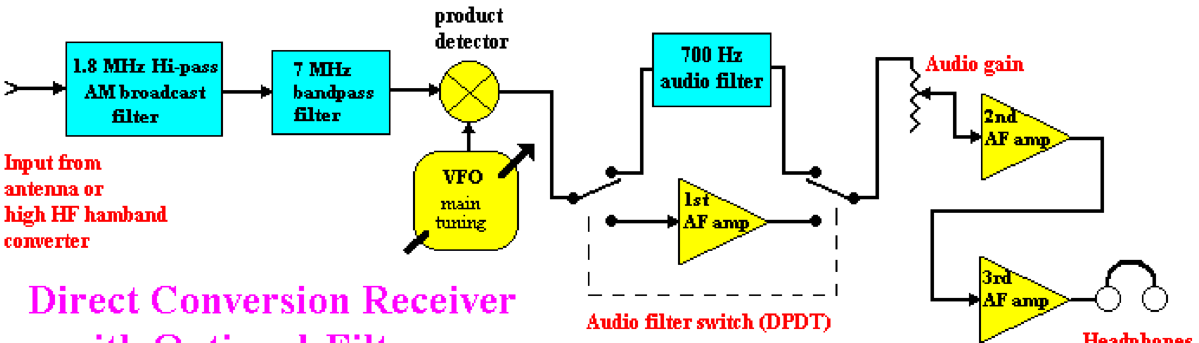

Direct Conversion Receiver with Optional Filters

The digital solution to audio filtering

Actually, op-amp filters are also old-fashioned. The modern way to build audio filters is to use digital filtering. Basically, a microcomputer program measures the width of individual sinewaves with a counter-like algorithm. For example, a 700 Hz sinewave cycle is 1.43 milliseconds long. The digital filter measures the height and widths of the sinewave humps. If the widths are the desired frequency, then it just recreates waveforms of that length and amplitude. Sinewaves that are longer or shorter are ignored. Averaging and integrating math algorithms are used to massage the data and arrive at the most accurate and useful representation of the original spectral content of the signal. In practice, a digitally filtered passband of just one Hz is easy to build but is very hard to tune in with your VFO tuning knob. Therefore the passband of a digital filter is usually adjustable. When the passband is fairly wide, say 200 Hz, tuning isn't so difficult.

In conclusion,

The block diagram above shows the receiver I built. I packaged the AM broadcast band filter outside the receiver so I could use it with other receivers. A switch allows me to put in the 700 Hz filter when I need it. If you use it with single sideband phone, it will get rid of much of the noise, but voice quality will be poor. To tune in single sideband, you have to turn off the filter and tune your VFO to exactly the right spot to make the modulation intelligible. AM voice stations must also be tuned in perfectly, otherwise they make an obnoxious whistle. The whistle occurs because the AM broadcast includes a fixed carrier wave that is comparable to a Morse code signal with the key held down. To get rid of the whistle, you must "zero beat" the VFO so that there is no audio frequency difference between the carrier wave and your local VFO. When you do that, the whistle vanishes and you will just hear the audio sidebands.

So? How well does the direct conversion receiver work? If this were the year 1935, all your ham friends would probably be insanely jealous of its terrific performance. Unfortunately, by modern standards it is a bit of a toy. It's adequate for shortwave listening and an experienced CW operator can use it to talk to people. I managed to use it in conjunction with the QRP transmitter in chapter 6 to talk to hams in other states. I guess that makes me an experienced operator. Novice hams could easily use it for communications when the band isn't crowded. It has plenty of sensitivity, just not much selectivity.

CRYSTAL SETS TO SIDEBAND
© Frank W. Harris 2010, REV 12

Chapter 8

POWER SUPPLIES

Once you progress past crystal sets, electronic projects almost always need a power supply. Low power projects, like a small receiver, can be powered with flashlight batteries or even a little 9 volt “transistor” battery. Other convenient sources of low power DC are ***adapter plugs***. These are the black cubes a few inches square that plug into the wall and have a long, skinny cord that plugs into your recorder or small appliance. They deliver 6, 12, or other DC voltage at a few watts. Adapter plugs have the safety of a battery with the convenience of wall power. On the other hand, their voltage may have high AC ripple noise and need to be filtered.

If you plan to power your QRP with household power, you’ll need more than a few watts. A 5-watt transmitter needs at least 10 watts of power with good voltage regulation. Otherwise AC ripple on the DC will go right out over the air as a hum or buzz. When you decide to plug a homebuilt circuit into the wall, you must confront some significant safety issues. Line-powered power supplies convert 120 volts AC into DC voltage at the required levels of voltage and current. Actually, a transmitter can also be thought of as an energy conversion device. It converts direct current into radio frequency current. In this chapter I’ll describe some power supplies you could use for your QRP transmitter.

Line-powered power supplies

The ideal power supply is called a ***voltage source.*** A voltage source is a power supply that can supply infinite amounts of current without the slightest waver in the voltage. For example: As every northern resident knows, starting a car can be difficult when it’s below zero. A cold battery does not supply as much current as a warm battery. So, when you turn the ignition key on a frigid morning, the battery voltage crashes.

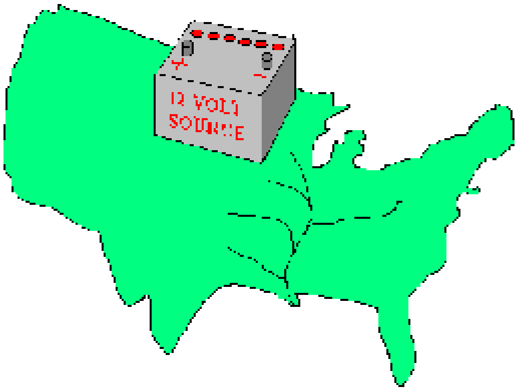

On the other hand, if you had a battery the size of North Dakota, the voltage would not drop a microvolt when you started the engine. Moreover, you could start all the other cars in Minnesota simultaneously without voltage drop. Of course, there are also other practical issues here. For example, your battery would need zero resistance battery cables, zero resistance connectors, etc. Well, you get the idea: The ideal voltage source should not lose any voltage, no matter how much power it supplies. Using ham radio vernacular, a good power supply is a "stiff" supply.

Lab power supplies

A reliable, line-powered laboratory power supply is useful for checking circuit boards. No lab should be without one. A big advantage of commercial lab supplies is that the voltage is adjustable from zero to some high level like 20 volts. Meters show you the current and voltage at every moment, so you know what is happening. Variable voltage allows you to power up a new circuit *CAREFULLY*. You can start with a few tenths of a volt and see what happens. If the circuit is shorted, you can find out with one volt applied to the circuit board, rather than blasting it with 12 volts right away. This helps you avoid burning up expensive transistors. Many lab supplies put out two or even three separate supply voltages at once. Another feature of some lab power supplies is that they automatically limit the available current to some maximum that you select.

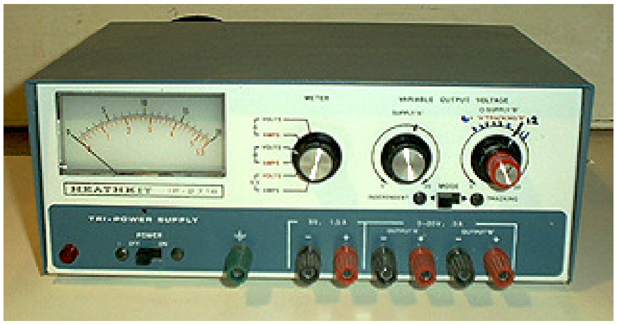

Bench power supplies are quite generic and there are many modern ones that will serve you well. Remember, to power your QRP you need about 1.0 ampere at 12 volts DC. A typical modern, transistor QRP transmitter runs on a 12 volt power supply but its efficiency is only about 50%. Therefore,

10 watts = 12 volts x 800 milliamperes

Homebuilt power supplies for use with rechargeable batteries or line power are described below. If you aren't familiar with power supply design, a discussion of the basic principles follows.

Simple wall-powered supplies for 120 Volts AC

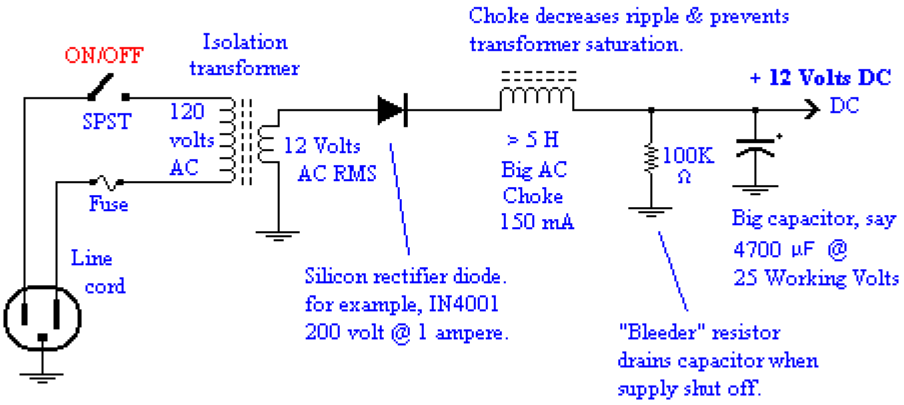
ON/OFF
SPST
Isolation transformer
120 volts AC
Fuse
Line cord
12 Volts AC RMS
Silicon rectifier diode. for example, IN4001 200 volt @ 1 ampere.
Choke decreases ripple & prevents transformer saturation.
> 5 H
Big AC Choke
150 mA
100K Ω
"Bleeder" resistor drains capacitor when supply shut off.
+ 12 Volts DC
DC
Big capacitor, say 4700 µF @ 25 Working Volts

SIMPLE LINE-POWERED
12 VOLT POWER SUPPLY.

(NOT SUFFICIENT FOR QRP)

The diagram above illustrates the simplest, safe, generic, line-powered power supply you can build. Unfortunately, this supply is too poorly regulated to power a transmitter. However, it illustrates the minimum safety features and it's easy to explain.

The following discussion assumes that the reader lives in North America where the standard household line voltage is 120 volts AC RMS @ 60 Hz. The safety issues explained here are applicable to other regions of the world. However, voltages, connector types, wire color codes, and ground configurations are often different. For example, in Europe the standard is 220 volts AC RMS @ 50 Hz.

Power supply safety

Metal enclosures. The supply should be enclosed in a box to insure that children (and you) won't get fingers across the 120 volts AC. Ideally the box should be made of metal so that, in case of a short circuit, a fire is highly unlikely.

Another safety design philosophy is called ***double insulation.*** In this scheme the electronics are housed in a plastic box and extra effort is made to insure that the internal wires are properly protected so that shorts and loose wires are highly unlikely. A double insulated plastic box does not necessarily need a ground wire in the line cord. However, in ham work, metal boxes shield circuitry from stray radio waves and are usually the best choice.

Line cord. The line cord should be the modern, three-wire type with the (green) ground wire securely connected to the metal box. In case a loose wire in the box causes the hot side of the AC line to touch the metal box, the ground wire will safely shunt the AC current to ground.

The line cord should pass into the metal box through a ***rubber grommet*** so that the metal edge can't cut through the insulation on the wire and cause a short circuit. Once inside the box, the cord should be held captive by a clamp, properly known as a ***strain relief.*** The strain relief insures that if the power supply is ever yanked by its cord, the live wires will not be ripped loose and short out.

The wires in a line cord are usually color-coded. The "hot" wire usually has ***black*** insulation while the "neutral" or return wire is ***white***. The third ***green*** safety wire is connected to power line ground. It should be connected directly to the metal chassis. The neutral wire is also supposed to be connected to the house ground buss out in the circuit breaker box. Looking at a North American household three-prong socket, the round pin is the "ground" and is connected to the green wire. The wider, flat pin is the hot side and the narrower flat pin is neutral. Unfortunately, sometimes wall sockets are wired wrong, so it's better not to bet your life on the orientation of the flat pins.

Fuse. The first destination of one of the two power wires, preferably the "hot" black lead, should be a ***fuse***. As you probably know, fuses are little pieces of solder-like lead mounted in a glass case. When the current exceeds some calibrated level, like one ampere, the lead melts and the circuit opens. Fuses, of course, can only be blown once and can't be reused. A fuse is represented on the diagram by the squiggle in the diagram labeled "1A," meaning "one ampere." The electrical standards allow a fuse to be considerably larger than necessary, like 5 amperes and still give adequate protection against shorts. Small ***circuit breakers*** serve as reset-able fuses and are available in low current levels such as 3 or 5 amperes.

Switch. The power switch can switch just one side of the line, such as the single-pole, single-throw (SPST) switch shown above. Or, it is even safer to switch both sides of the power line at once using a double-pole, single-throw switch (DPST). The switch should be rated for at least 125 volts AC and 3 amperes.

The transformer

After the power switch, the line current usually goes to the primary winding of a transformer. The transformer has two functions: first, it ***isolates your power supply from the household supply*** and from ground. This makes electrocuting yourself much less likely. As explained earlier, the AC lines supply 120 volts AC referenced to ground. The transformer secondary delivers AC power that has no relation to ground at all.

For example, I don't recommend actually trying this, but suppose you were to plug a well-designed transformer into a wall socket. And suppose that this transformer has high voltage secondary wires left dangling open circuit: Because of the isolation, you could touch either secondary wire without being shocked, even if your other hand were hanging onto a grounded water pipe. Of course if you touch both high voltage wires simultaneously, they will blast you.

A secondary winding is isolated from ground - like a battery floating in mid-air

Think of isolation as a battery hanging from a balloon. Electric circuits require a closed loop in order for current to flow. For the dangling battery, the current can only flow from one end of the battery to the other. There is no relationship to ground. If a person standing on the ground reaches up and touches the battery, no circuit loops are completed, so no current flows.

5. Chapter 8, Harris

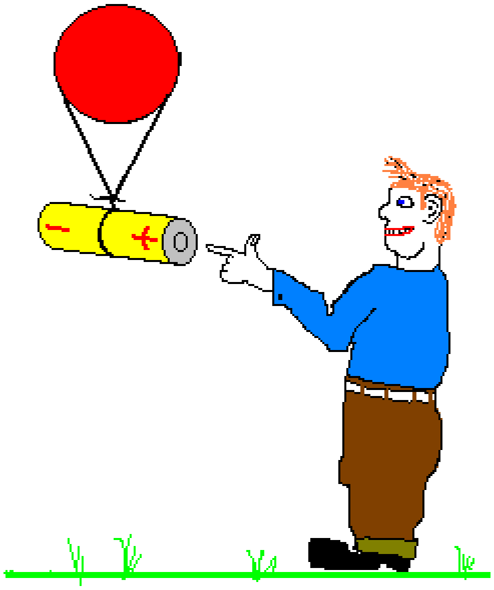

As explained in Chapter 2, transformers can change the ratio of current to voltage in proportion to the number of turns of wire around the core. To convert 120 volts AC to 12 volts AC, the turns ratio between primary and secondary coils would be 10:1. Or, if you needed to generate 1200 volts AC for an oscilloscope cathode ray tube (**CRT**), the turns ratio would be 1:10.

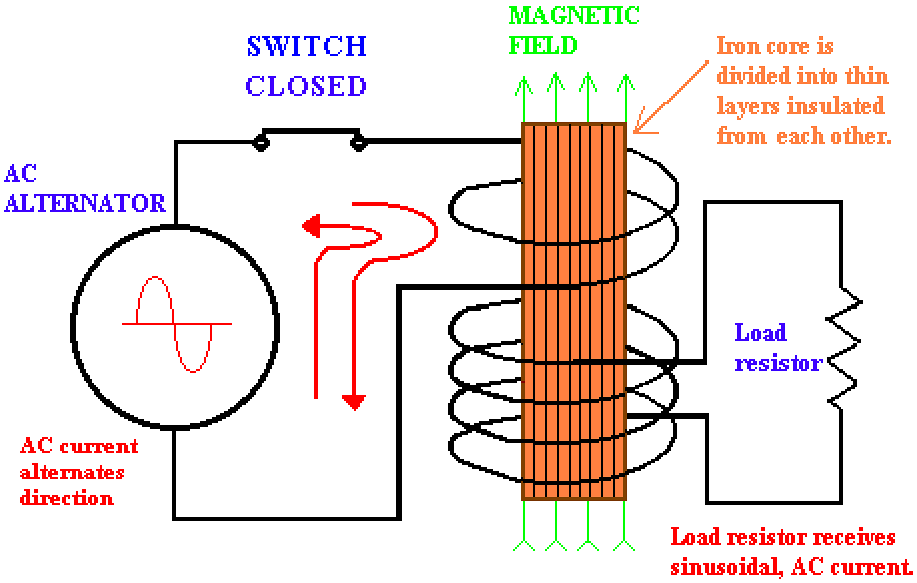

SWITCH CLOSED
MAGNETIC FIELD
Iron core is divided into thin layers insulated from each other.
AC ALTERNATOR
Load resistor
AC current alternates direction
Load resistor receives sinusoidal, AC current.

Transformers are not 100 % efficient. They are made from copper wire that has a significant resistance and iron that dissipates a small amount of energy as heat every time a magnetic field is generated or its polarity is reversed. In general, the larger the transformer, the larger the diameter wire used on the windings, the higher the efficiency will be.

Iron has a definite and abrupt limit on how much magnetic field it can support. Once all the iron has been magnetized, the iron will contribute no more magnetic field, no matter how much more current flows through the primary. Obviously, the larger the iron core, the more energy the iron can pass on to the secondary before the iron saturates. As a general rule, the larger the transformer, the more power it can pass through its windings.

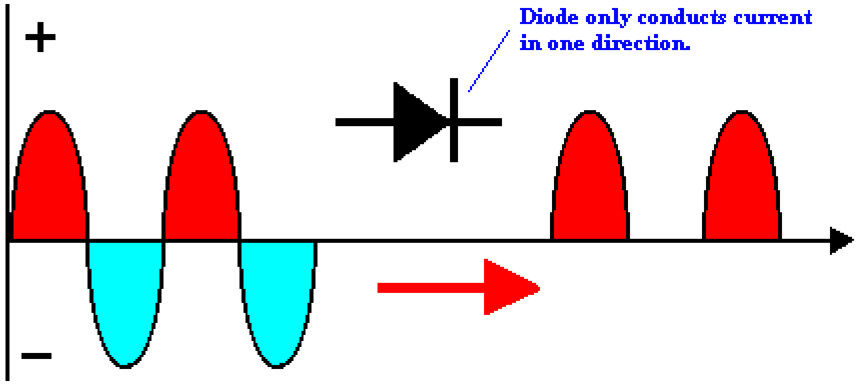

HALF-WAVE RECTIFICATION

(Only one half of wave is exploited.)

Rectification

Most electronic devices require DC voltage to work properly. AC is converted to DC by means of rectifier diodes. Rectifiers are high-power, high-current, high-voltage versions of the diodes used in crystal sets. Referring to the simple power supply shown earlier, the transformer is followed by a rectifier diode. ***The diode only passes positive current flow in the direction of the arrow.*** This means that current leaving the diode is in the form of "humps," or half sine waves. Only half of the sinewave is passed through, so a single diode rectifier is called a ***half-wave rectifier.*** By definition these humps are "DC" since they have only one polarity. Unfortunately, for most electronics applications bumpy, intermittent sine-wave halves are grossly inadequate. For example, if you use them to power a CW transmitter, your signal will sound like an unpleasant buzz and it will take up 120 Hertz of bandwidth. In the 1920s this was often done deliberately to modulate Morse code and make it audible in a receiver without a BFO. If you tried to run a computer microprocessor on these humps without filtering, the processor would reset 60 times per second.

Peak Volts, RMS volts, and DC volts

Obviously sinewaves don't stand still at any one voltage, so how are they measured? If the sinewave reaches peaks of plus 12 volts and minus 12 volts, then most of the time the output from the rectifier will be much less than 12 volts. Also, the positive voltage seems to cancel the negative voltage so that the arithmetic average of any sinewave voltage is zero. Obviously that measurement isn't useful either.

Rules are needed to name sinewave voltage and current. As you might expect, ***peak voltage*** is the voltage difference between zero and the most positive extreme of the waveform.

Peak-to-peak or ***PP*** voltage means the voltage difference between the most negative peak and the highest positive peak.

As you may know, the voltage that comes out of North American wall sockets is officially named ***120 volts AC RMS***. During my lifetime American line voltage has also been nominally called “110 volts AC,” “115 volts AC,” “117 volts AC,” and now it is called "120 volts AC." Confusing, no? Anyway, the two power wires coming into your house from the power pole out in the alley nominally have 240 volts AC across them. Of course it was formerly called two times 110 volts AC or 220 volts AC. Some people still call it 220 volts. Line voltage is designed so that each of the two wires from the power pole is at 120 volts AC with respect to ground.

As explained in chapter 2, inside your circuit breaker box these two wires are connected to two big metal “buss” bars. There is a third, grounded metal bar that runs down between the two active buss bars. The individual circuit breakers snap onto these bars like cars on a railroad track. Heavy power circuits for your electric stove and clothes drier clip clear across the ground buss to engage both hot lines for a total of 240 volts. Small circuit breakers just clip from one side to the ground bar to obtain 120 volts for ordinary low-power circuits.

The ***RMS*** or ***Root Mean Square*** of a sinewave voltage is the peak voltage divided by the square root of two. RMS voltage can be thought of as "the effective average voltage." It can be used to calculate AC RMS current, power, and AC resistance and impedance. For example, 120 volts RMS is a sinewave with a peak voltage of 1.414 times the RMS voltage. In other words,

120 V RMS x $\sqrt{2}$ = 120 x 1.414 = 169.7 Volts Peak. Therefore, ordinary household line voltage could be expressed as ***120 Volts RMS, 170 Volts Peak*** or ***340 Volts Peak-to-Peak***.

Filter capacitors - filtering out the ripple

The half sinewave bumps, which are properly known as “ripple,” must be smoothed out into a continuous DC voltage. This is accomplished by means of a ***low pass filter.*** In this case low pass means that the filter only passes frequencies well below 60 Hz. DC is of course zero Hertz, which is the lowest frequency possible. The half-wave rectified supply illustrated earlier is equipped with an L-C “L” filter. L- filters are simply two-element filters that represent the two legs of the letter “L.”

Capacitors conduct AC and prevent the flow of DC. Inductors conduct DC but resist the flow of AC current. In an effective DC filter, the component values will be huge, like 10 henries and 5,000 microfarads. This is because the frequency we want to filter out, 60 Hz, is extremely low and big components are needed to have an effect on such a slowly changing sinewave.

Bleeder resistor across the capacitor

The purpose of the bleeder resistor across the filter capacitor is to discharge it when the supply is not in use. Remember that high quality capacitors will hold their charge for many hours, sometimes days. Bleeders usually aren’t important with a low voltage supply like 12 volts. But if this were a 500 volt supply, a person could get a nasty jolt or burn if they were to touch the capacitor. This could happen even though the supply is no longer turned on or plugged in.

If you were to build the half-wave supply shown earlier and put a 10 watt load on it, with an excellent transformer and a really huge capacitor, perhaps the DC voltage will at least be continuous. However, there will still be a 60 Hz sinewave ripple or waves impressed on top of it. If you used it to power a transmitter, the tone of the Morse code would have a distinct rough sound as the DC level varies at 60 times per second. When hams gave you a signal report they would say that your signal tone was a number much less than 9. For example, they might send “UR RST 593.” (RST means Readability, Strength and Tone. Hams use a scale of 1 to 5 for readabilitity, and 1 to 9 for both strength and tone.) Because of the harsh note of your tone, you might only rate a “3.” For this reason, transmitter power supplies always use dual rectifiers to produce ***full-wave rectification***.

Full-wave rectifiers

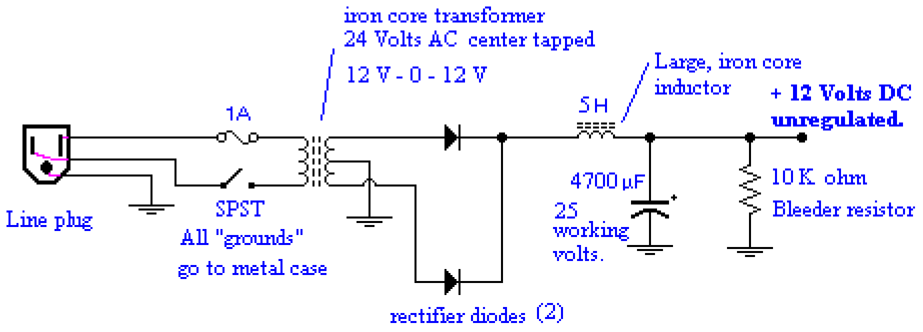
iron core transformer
24 Volts AC center tapped
12 V - 0 - 12 V
1A
Large, iron core inductor
5 H
+ 12 Volts DC unregulated.
4700 µF
+
25 working volts.
10 K ohm
Bleeder resistor
Line plug
SPST
All "grounds" go to metal case
rectifier diodes (2)

FULL - WAVE RETIFICATION USING A TAPPED POWER TRANSFORMER

Full-wave rectifiers convert both halves of the sinewave into useful DC current. The DC voltage is now a succession of “humps” with no “off” intervals. With twice as many “humps” per second, the voltage is much easier to filter. Full-wave rectification is a big step toward producing a DC source that resembles the smooth continuous voltage available from a battery.

There are two ways to achieve full wave rectification. The circuit above uses two diodes. What you probably didn’t notice at first is that the secondary of the transformer has *TWO* 12 volt RMS AC windings. By having two separate windings, and wiring them in series, one of the windings can be positive at all times. This allows positive current to flow through one of the two diodes at all times and greatly decreases the ripple. Notice that, if we wanted, we could reverse the polarity of the diodes and produce the same waveform with the opposite polarity. That is, if we wanted a ***NEGATIVE voltage source*** referenced to ground, reversing the diodes would do that.

Bridge rectifiers

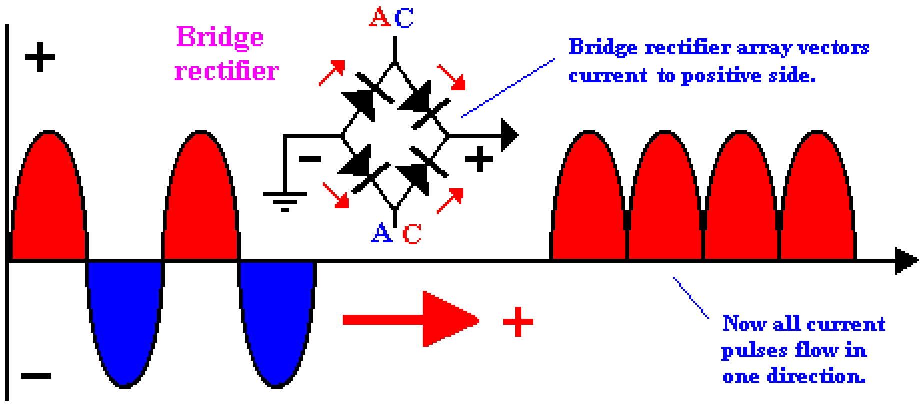
+
Bridge rectifier
AC
Bridge rectifier array vectors current to positive side.
−
+
AC
+
−
Now all current pulses flow in one direction.

FULL WAVE RECTIFICATION

(Both polarities of sinewave are exploited.)

The second way to achieve full wave rectification is to use a bridge rectifier made from four individual diodes. This configuration allows us to get full wave rectification from a single secondary winding. The four diodes are soldered in a diamond pattern as shown above. The AC voltage source is applied across the top and bottom of the diamond. The two diodes on the right are pointed so that the positive current will always flow to the positive side. The left side is wired to ground and the diodes point in such a way the negative current is always vectored toward ground. To say it another way, the positive current always flows "UP" from ground.

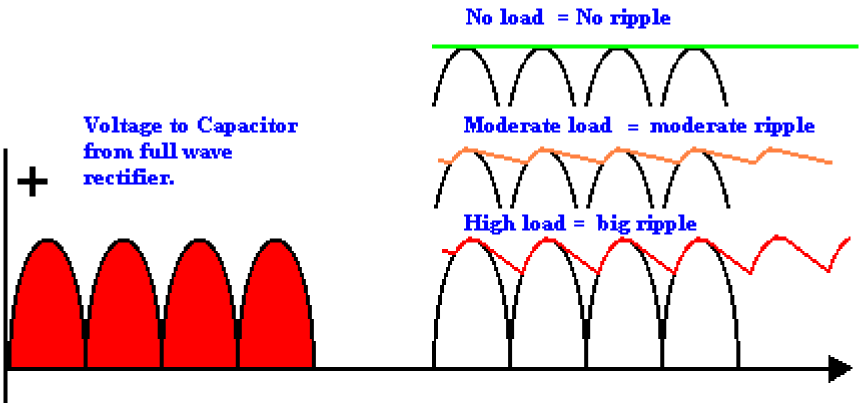
No load = No ripple
Voltage to Capacitor from full wave rectifier.
+
Moderate load = moderate ripple
High load = big ripple

Effect of Load Current on Power Supply Ripple.

Ripple

Because the rectifier is supplying current in the form of "humps," the voltage output across the capacitor will also vary up and down. If the capacitance is large the voltage doesn't drop to zero during the "valleys," but it can drop fairly low if the current drain from the supply is large. The more current that is sucked out of the capacitor, the lower the voltage "valleys." This is illustrated in the drawing above. If the choke is large enough, it will work with the capacitor to smooth out the height of the voltage peaks and raise the level of the valleys dramatically.

With no load current, output voltage charges to the peak voltage

Notice that, if the power supply is not connected to an external load, the output voltage will rise up to the maximum voltage that comes through the rectifiers. For a 12 volt RMS transformer output winding, this is approximately the square root of two (1.414) multiplied times 12 volt RMS volts, or about 18 volts. Referring to the diagram, with no external load the only load on the capacitor is the tiny one milliampere current passing through the bleeder resistor. This means that the variation in voltage between the humps will be extremely small. In summary, if there is no load on the power supply, to a voltmeter it looks like a "regulated" 18 volt power supply with essentially no ripple. If your 12 volt circuit might be damaged by 18 volts, you must not connect a supply like this to your circuit.

When you first turn it on, the voltage will be 18 volts for a moment before the load current flow is established and the choke and capacitor pull the peak voltage down. RMS (Root Mean Square) refers to the *AVERAGE* voltage of the rectified sine wave "humps." If the output from the big capacitor were fed directly into your QRP without a regulator, the capacitor would charge toward the peak voltage of the "humps." Depending on how much current your QRP was drawing, 18 volts might destroy some of the components in your QRP. As the load on the power supply is increased, the output voltage would drop down toward the rated RMS voltage, 12 volts. As the rated transformer load, say 3 amps, is exceeded, the DC voltage will probably drop below 12 volts.

Power transformer saturation

A second purpose of the choke (series inductor) is to prevent the peak current from the transformer from exceeding the current rating of the transformer. If the choke were not located between the rectifier and the capacitor, the current from the "humps" would only flow into the capacitor when the voltage from the rectifiers is higher than the voltage already stored in the capacitor. It is like a tidal pool alongside the ocean. The pool can only fill with water when the tide rises higher than the level of the pool. The result of these pulses of charging current is that the power transformer must provide much more peak current than it may be rated for. If a transformer is rated for, say 10 amperes RMS, and these surges of current are drawn in bursts just a few microseconds long, then the peak current might be 100 amperes. Since the transformer doesn't have enough iron for that, the iron saturates and the transformer inductance momentarily collapses. Suddenly, the transformer will act like a few turns of copper wire shorting out the AC source. Saturation causes the windings and the transformer to heat rapidly and perform poorly.

Substituting big chokes with small, cheap resistors

In the real world, most low voltage power supplies like this solve the problem with a cheap resistor instead of a large, expensive iron core inductor. Resistors waste energy, but what the heck! You can use the resistor with an extra-large filter capacitor that costs less and weighs less than a choke with equivalent filter value. Or, as we'll see shortly, a linear voltage regulator can put a load on the capacitor all the time so that the current flow is not just in short bursts.

Regulators

To provide pure DC at a constant voltage over a wide range of load current, you need a regulated power supply. The regulator's first task is to "trim off" unwanted peak voltage and provide a DC voltage equal to (or similar to) the transformer rated RMS voltage. Regulators solve the 18 volt over-voltage problem described above. Its second purpose is to maintain constant voltage even when the load resistance is changing continuously or during a line voltage brown out. A regulator circuit is usually added to a power supply like the one above. There are two basic designs for regulators, ***linear regulators*** and ***switching regulators***.

Linear regulators are a sort of automatic variable resistor placed in series with the output of a simple supply like the ones we have been discussing. The regulator uses feedback from the load voltage to change the size of the "automatic resistor" and hold the load voltage constant. For example, in the above power supply the regulator input voltage might vary from say 15 to 18 volts, but the regulator would change its resistance to hold the output constant at 12 volts DC. ***A linear regulator not only insures that the load voltage is always the same, it also "trims" off the ripple.***

Switching regulators are more complex circuits that usually involve inductors (or transformers) and switching transistors. They start with unregulated DC and turn it back into AC power. This AC power is then passed through a transformer to generate whatever voltage is needed above or below the original DC voltage. In an equivalent method, the unregulated DC is pulsed through an inductor to generate higher or lower voltages. Some switching regulators work directly off the household line. In other words, the 120 volts AC is rectified without a transformer and results in roughly 120 volts DC that is then converted into AC to drive a small, high frequency transformer or a step-down inductor. For amateur radio work, switchers usually make radio noise that you will hear in your receiver. Yes, commercial radio equipment often use switchers in their designs, but in my experience, getting rid of the switching noise is extremely difficult.

In contrast to switching regulators, some switchers just boost the voltage, but don't regulate the output voltage. These are often called ***charge pumps***. The noise from these unregulated charge pumps can be insignificant because they only switch at one frequency and they don't "dither" back and forth trying to hold the output voltage constant. To summarize, ***regulated switchers are noisy*** and I have pretty well given up trying to use them inside ham equipment.

Zener diode regulators

The simplest regulator is made from a ***Zener diode***. It is a "***linear regulator***" because there are no abrupt pulsed signals involved and it works entirely with DC. It dissipates the unwanted voltage as heat. A Zener diode is a modified silicon diode subtly different from the rectifiers used in the power supplies above. Zener diodes are made from silicon that has been doped (contaminated) with extra ions. The extra ions cause the diode to break down at a specific lower reverse voltage when a high reverse voltage is applied. Zener diodes are deliberately manufactured to be inferior rectifiers, but when used as regulators, they are quite useful.

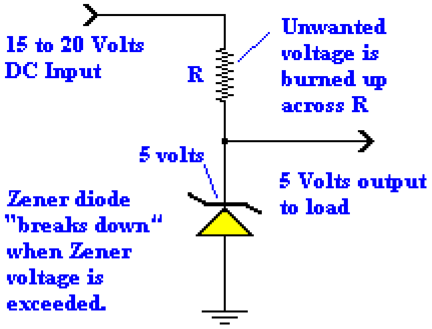
15 to 20 Volts DC Input
R
Unwanted voltage is burned up across R
5 volts
5 Volts output to load
Zener diode "breaks down" when Zener voltage is exceeded.

Why ordinary rectifiers don't behave like Zeners

Diodes rectify because the reverse voltage is insufficient to force electrons through the P-side of the P-N junction that has no free electrons. However, when enough voltage or "electrical force" is applied, the P-N barrier breaks down in an ***avalanche breakdown***. When this breakdown happens at high voltage and high current, the sudden heat generated often ruins the diode (or bipolar transistor). You may remember the homemade crystal diode discussed in chapter 4. When this crude diode was reversed biased, it broke down abruptly at about one volt. At any voltage above 1 volt, it acted like a short circuit and regulated the load voltage to one volt. At low voltage levels, such a breakdown doesn't necessarily destroy the diode if the diode doesn't overheat. This ***Zener voltage*** level can also be used as a voltage reference.

Doping silicon diodes makes Zeners

When big, tough, modern 400 volt silicon diodes are "doped" with extra ions mixed into the semiconductor, the avalanche can happen at lower voltages anywhere below 400 down to as little as 3 volts. Because the load voltage is low, and the load resistance presumably limits the current to a safe level, the heating in the silicon is mild enough that the diode survives the breakdown. In practice, Zener diodes are available from about 3 to 50 volts. 100 volt Zeners would have to be capable of dissipating a great deal of heat or they would be easily destroyed.

In the circuit above the Zener diode breaks down at 5 volts. If one of these diodes is placed across a load, the Zener diode will clamp the voltage to 5 volts so the voltage across the load never rises above 5 volts. Of course the input voltage must always be higher than 5 volts and the unwanted voltage will be dissipated across the resistor in series with the power supply. Ohm's law tells us that at least half of the resistor current must pass through the Zener on its way to ground. So the Zener diode itself is dissipating as much energy as the intended load. As you can see, Zener diode regulators have poor energy efficiency.

Real versus ideal Zeners

If a Zener diode behaved "perfectly," the voltage across it would be the Zener voltage, no matter how many amperes flowed through the diode. This is illustrated by the **green** curve in the graph below. Unfortunately, the Zener voltage rises with large currents as shown below on the **red** curve. Notice that when the Zener diode is wired backwards, it acts like a normal, forward-biased silicon diode. Conduction starts at about 0.6 volts. What is called "forward" conduction for a normal diode is called reverse conduction for a Zener diode.

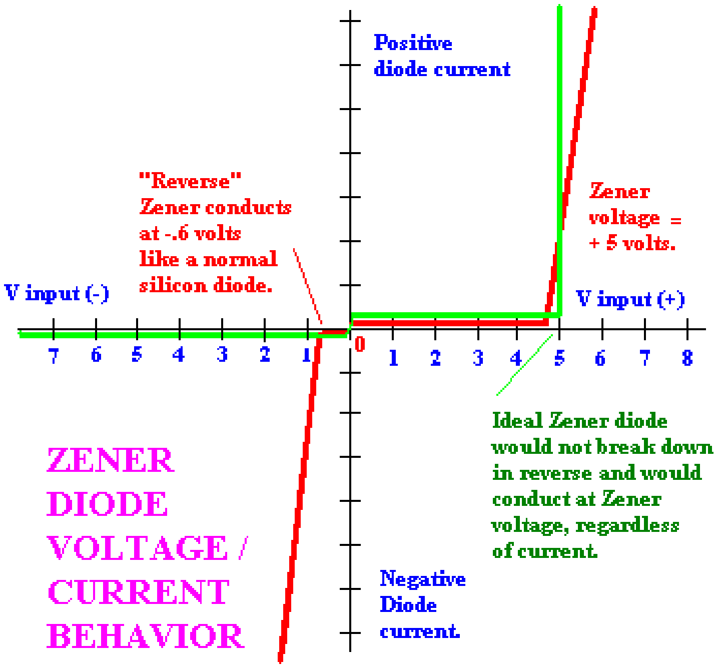

Zener diodes are impractical for heavy current regulation. They not only dissipate energy in the series resistor R, they also burn up energy in the diode. Moreover, as the diode current rises, the Zener voltage can be significantly different from its nominal value. In practice Zener regulation is used for light loads over a narrow range of input voltage. As you will see later, Zeners can be used as ***VOLTAGE REFERENCES*** to run a transistor regulator. In this way, the regulation can be fairly good and energy isn't wasted.

This will be illustrated in an application later in this chapter.

A homebuilt, line-powered power supply for a QRP

Now let's be practical: The 12 volt supply below worked well for me and can be assembled entirely from parts at your local Radio Shack.

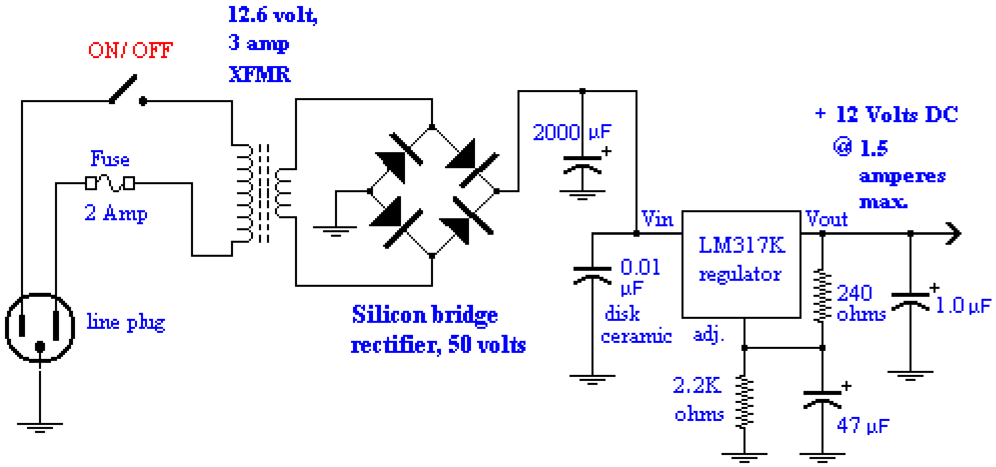
12.6 volt, 3 amp XFMR
ON/ OFF
Fuse
2 Amp
line plug
Silicon bridge rectifier, 50 volts
2000 µF
+
Vin
0.01 µF disk ceramic
LM317K regulator
adj.
Vout
+ 12 Volts DC @ 1.5 amperes max.
240 ohms
1.0 µF
2.2K ohms
47 µF

This 18 watt regulated line-powered supply should be built in a metal chassis. The round ground pin from the line cord should be connected to the chassis. In that way, if the "hot" black wire should break and touch the chassis, the metal would not become dangerous to touch. A 2 to 5 ampere fuse in series with the line input protects the supply. The ON/OFF switch should be rated for 250 volts DC and at least 1 ampere. The line cord should pass through a rubber grommet on the way into the chassis. Once inside, the line cord should be clamped to the chassis so that, if someone picks up the power supply by the cord, the force will be on the clamp and not on the solder joint where the line cord is fastened to the fuse and switch.

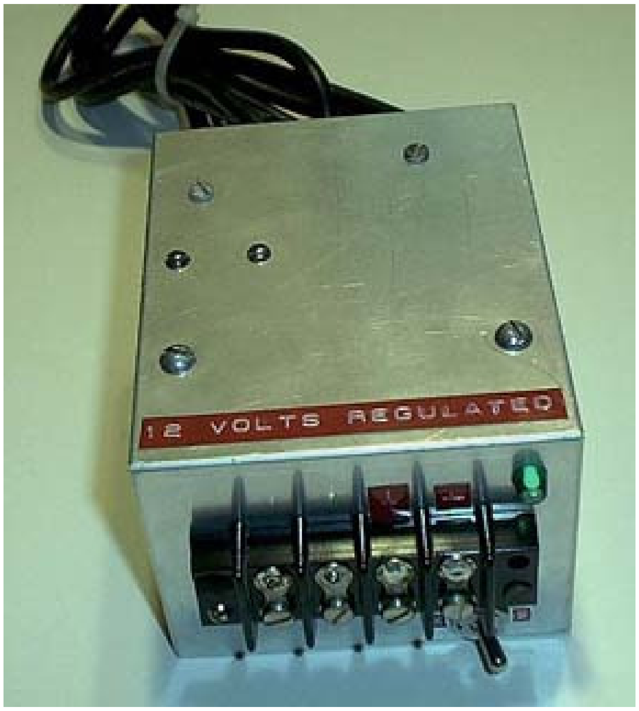

The transformer is rated at 12.6 volts RMS at 3 amperes. I used Radio Shack part # 273-1511. The rectifier is a generic silicon bridge rectifier rated for at least 50 volts @ 3 amperes. Alternatively you could use 4 individual rectifiers arranged as shown above. The filter capacitor following the rectifier just needs to be large and greater than about 25 working volts. The other three capacitors serve to stabilize the output and prevent oscillations and ripple. Adding a 12 volt dial light so you will know when it is on is a nice touch.

The LM317K programmable voltage regulator

The LM317K regulator chip is packaged in a TO-204 (formerly known as TO-3) metal case and can deliver 1.5 amperes. To dissipate the heat, the TO-204 case should be bolted to the metal chassis and insulated by means of a mica washer and silicone grease. Mounting kits for this purpose are also available at Radio Shack. This regulator chip works by regulating the voltage between the V_{out} terminal and the ”adjustment” lead on the regulator. The regulator regulates this voltage difference to 1.2 volts. This low voltage allows the designer to regulate voltages equal to or higher than 1.2 volts. In this 12 volt application, the regulator passes current through a 240 ohm resistor and regulates the voltage across the 240 ohm resistor to 1.2 volts. This results in a “current source” that drives a constant current through the 2.2K resistor to ground. This increases the total regulated output voltage from 1.2 up to 12 volts. In other words, 240 ohms is about 10% of the sum of 240 ohms plus 2200 ohms. Because you can “program” the total regulated voltage, this same regulator can be used to regulate voltages from 1.2 volts up to about 20 volts.

Another reason you need good voltage regulation is that the tuned stages in your QRP don’t take kindly to changing the supply voltage. If you tune up the antenna using 12 volts, and the voltage later goes up or down, the tuning of some stages may change slightly and your signal might crash in the middle of a QSO (conversation).

The LM317K “clips off” any voltage above 12 volts so your QRP will never see a higher voltage. Notice that the LM317K is a ***linear regulator***. This means that it turns any input voltage that is above the regulation voltage into heat. This seems wasteful, but linear regulation makes no radio noise and produces a flat output voltage with practically no ripple. Notice that the choke (or a resistor) is not necessary here because the regulator is drawing current from the capacitor nearly all the time. This means that the current is flowing from the transformer nearly all of the time and will not be saturated by high current pulses.

This power supply violates my goal to never use integrated circuits. If there are purists who share my aversion to integrated circuits, the "low dropout" regulator shown below can be substituted for the LM317K regulator. A larger power supply than this would be more versatile and could power a bigger transmitter. Obviously, the higher the power level, the more cumbersome the regulation becomes. Running a 100 watt transmitter requires a 300 watt regulated power supply. I sidestepped this project by running my entire ham station on a solar-charged 12 volt battery which is described later in this chapter.

A low dropout voltage regulator for use with a battery

As you will read shortly, two 6-volt lead-acid golf-cart batteries make a fine, high-current 12 volt power supply. Unfortunately, running your transmitter on batteries means that, when the battery is ready to be recharged, its voltage will drop down to 11 or even 10.5 volts. This means that (ideally) your QRP must be tuned up to run on 10.5 or 11 volts and then regulated to that level. Remember that ***a linear regulator can only deliver LESS than its input voltage***. So if you're running on a 12 volt battery, you need a regulator that wastes as little voltage as possible. This waste voltage is called ***dropout.*** The regulator shown below receives nominal 12 volts from a battery. The active regulation element is a big P-channel MOSFET transistor. This transistor can be arbitrarily large. The bigger it is, the less voltage it will waste across its internal resistance and the lower the "drop out."

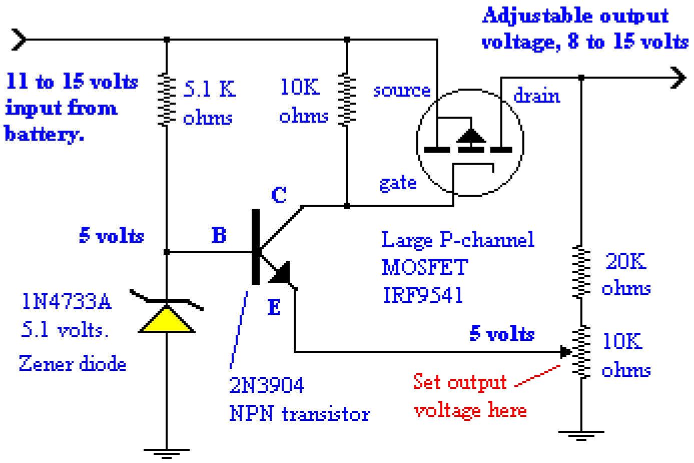
Adjustable output voltage, 8 to 15 volts
11 to 15 volts input from battery.
5.1 K ohms
10K ohms
source
drain
gate
C
5 volts
B
Large P-channel MOSFET IRF9541
20K ohms
1N4733A 5.1 volts. Zener diode
E
5 volts
10K ohms
2N3904 NPN transistor
Set output voltage here

Low Dropout 12 Volt Power Supply

The P-channel MOSFET transistor turns on (conducts current) when its gate voltage is pulled downward toward ground. So when the battery voltage drops, the gate voltage must be pulled down (toward zero volts) to turn the MOSFET more on. When the battery input voltage rises, the gate voltage must be raised to turn the transistor more off and restore the output to the set voltage.

The gate voltage control is done with an NPN bipolar transistor. The transistor compares the reference voltage across a 5 volt Zener with a fraction of the output voltage across the sense resistor pot on the lower right. The slider on the sense resistor contacts the resistance at a level that produces about 0.6 volts less than 5 volts. As the output voltage rises, the voltage on the sense resistor rises. This in turn increases the 2N3904 emitter voltage. The difference in voltage between the base and emitter drops, causing the drive to the NPN transistor to drop. As the NPN transistor turns more off, the gate voltage on the MOSFET rises toward the battery voltage and turns the MOSFET more off and lowers the regulated output voltage.

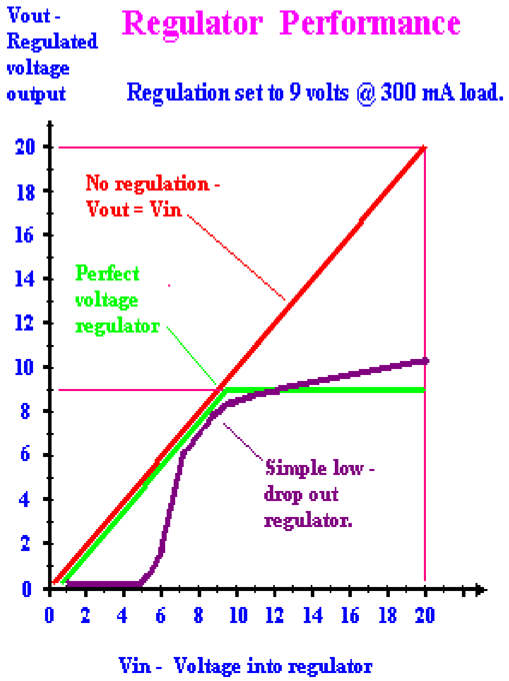

The **purple** plot on the left shows the performance of this simple power supply when it has been set to 9 volts with a 300 milliampere load. The **red** line shows the voltage the load would receive if there were no regulation. That is, Vin = Vout. The **green** line illustrates the performance of an ideal or perfect linear regulator. If the transistors had infinite gain and the Zener diode always produced precisely the same reference output voltage, then the green curve is what you would get. As soon as the regulator is given 9 volts input, it would deliver exactly nine volts, no matter what the load was. As you can see, the simple regulator isn't radically different from the ideal, but it is far from perfect.

A precision, temperature-compensated, low-dropout regulator

More nearly perfect regulation can be achieved by replacing the NPN transistor with an operational amplifier. Operation amplifiers, "op-amps," are integrated circuits composed of many or even dozens of transistors. Op-amps perform as though they were nearly perfect transistors. Similarly the LM336 voltage reference is an integrated circuit made from bunches of transistors that perform as though they were a nearly "perfect" Zener diode. As the voltage across the regulator is varied, the voltage across the Zener remains constant within a couple thousandths of a volt.

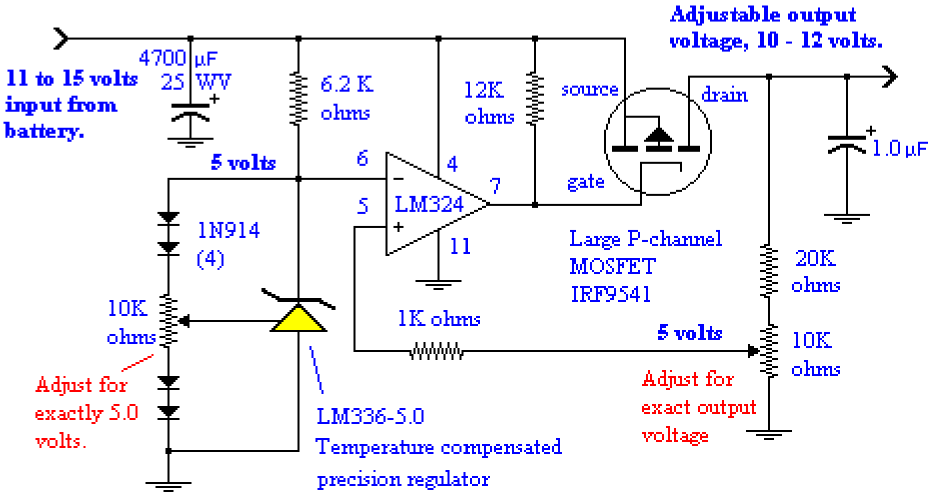
Adjustable output voltage, 10 - 12 volts.
11 to 15 volts input from battery.
4700 µF
25 WV
6.2 K ohms
12K ohms
source
drain
gate
5 volts
6
4
7
5
LM324
11
1.0 µF
1N914 (4)
Large P-channel MOSFET IRF9541
20K ohms
10K ohms
1K ohms
5 volts
10K ohms
Adjust for exactly 5.0 volts.
Adjust for exact output voltage
LM336-5.0 Temperature compensated precision regulator

Low Dropout 12 Volt Power Supply with Precision Regulator

The op-amp is the triangle in the center. This op-amp has 14 pins and the numbers shown are the pin numbers. This particular chip contains four op-amps. Only one is used in this circuit. Pins 6 and 5 are the inputs. The input pins function in such a way that the op-amp changes the output voltage (pin 7) to “try” to keep the two input pins at the same voltage. That is, so long as the voltage on pin 6 is identical to the voltage on pin 5, the output remains constant. When the **positive** pin 5 has a higher voltage than the **negative** pin 6, the output pin will shoot positive as high as it can go. When the negative pin 6 is higher than pin 5, the output pin will zoom negative as low as it can go.

Unlike simple transistors, the op-amp has nearly infinite voltage and current gain. Gains like 100,000 or a million are common. Also, the inputs draw essentially no current. They have nearly infinitely high input resistance. This means that op-amps are ideal to “monitor” some condition, such as the output voltage from the power supply, and then change the OP-AMP output pin voltage in order to restore the supply to equilibrium. Looking at the diagram, pin 6, the negative input, always rests at 5.00 volts, so long as the battery voltage input is above 5 volts. The output voltage pot at the lower right is adjusted to produce the desired output voltage. The big MOSFET transistor will remain turned on to the exact degree to deliver the voltage you set. This set point is the place where 5 volts appears on the pot. As soon as the op-amp “sees” that pins 5 and 6 have the same voltage, the output voltage on pin 7 holds still and stops changing.

A precision Zener diode

The LM336 precision Zener diode is not only extremely accurate, it will maintain its accuracy over a wide temperature range. Speaking of temperature change, don’t forget to bolt your MOSFET to a big heat sink. At high input voltages and big load currents, you may expect the MOSFET to get boiling hot without a heat sink. Keeping it cool keeps its internal resistance low and improves the regulation. The metal flange on the transistor should be insulated from the heat sink and metal chassis with a mica insulator and silicone grease.

As we shall see in chapter 10, building a variable frequency oscillator that reaches modern standards of frequency drift requires the use of a super-regulated power supply. This in turn will require you to use regulators made from integrated circuits. Sigh. Of course you could prove me wrong by building one with discrete parts.

Packaging power supplies

The photo below shows the underside of my all transistor CW transmitter. It is based on the modules described in chapters 6, 10, 11 and 12. The low dropout 12V QRP power supply is located just to the left of the center. The circuitry is mounted on a commercial perforated circuit board. This circuit board is standing on edge but the blue trim pots that adjust the LM336 and output voltage can be seen. The P-channel power MOSFET that controls the output is just above the board. The transistor is insulated from the chassis with a plastic screw and a gray silicon pad.

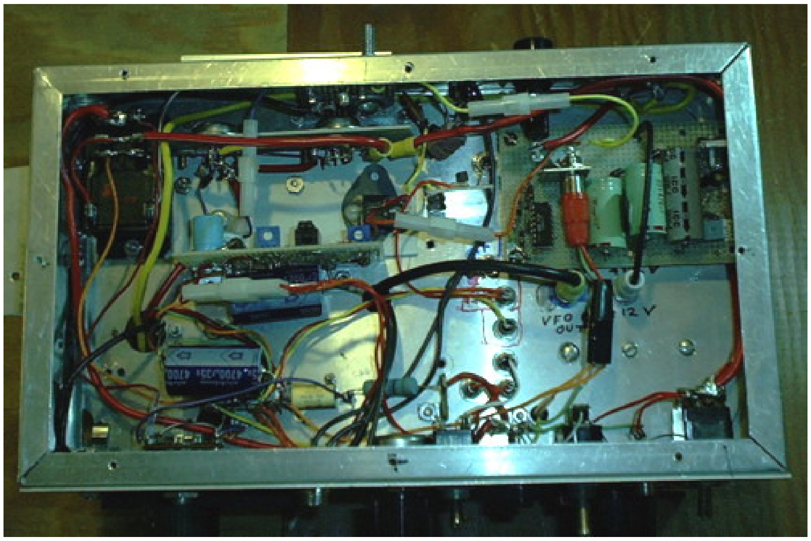

The perf-board circuit at the upper right with the three large capacitors is the precision supply for the VFO. The two large TO-204 transistors mounted vertically are the CW keying transistors for the 100 watt linear amplifier. The black relay at upper left is the main DC relay that engages the 12 volt battery during transmit.

Each module is mounted to the chassis with Molex or other connectors so that it may be easily repaired or replaced. Phono plug connectors are used as feed-through connectors wherever needed.

****

BATTERIES AS A POWER SUPPLY

The beauty of batteries

As I mentioned earlier, I run my whole station on solar power stored in a 12 volt battery. Now I know this sounds like tree-hugging, liberal silliness, but my solar powered station came about quite logically and has many advantages for a homebuilt station. There are lots of hams that do this - and no wonder! A lead acid storage battery is a wonderful power supply. It puts out huge currents whenever you need them and the voltage regulation is excellent. The best advantages are that batteries make no switching power supply RF noise and there's no waste heat from a big linear regulator.

My return to ham radio began in 1997 when I hauled my 1967 homebuilt mobile ham rig down from the attic and blew off the dust. Since my old mobile rig runs on 12 volts, I couldn't even try out the receiver without a heavy-duty 12 volt power supply. I considered building or buying a supply but after 10 seconds I decided that was ridiculous. I happened to have some 6 volt golf cart batteries in the garage. I hauled two of them down to the basement and put on an automotive 12 volt trickle charger. After a day, the batteries recovered and I could fire up the rig. AM phone was extinct, but there was nothing to stop me from getting on CW, which I did with good success. Well, that's not quite true. My adventures included a drifting VFO and a blown 200 watt charge-pump switcher, but those problems had nothing to do with the power source.

My old vacuum tube rig was a power hog. The receiver drew 3.5 amperes at 12 volts. That's 42 watts just to listen! The transmitter was much worse, of course, but it wasn't dramatically worse than a transistorized rig. After all, a transmitter can't radiate big power if it doesn't draw big power. Moreover, a modern, linear, class A transistor rig can be even more inefficient than an old Class C vacuum tube transmitter. Either way, my transmitters draw 6 to 20 amperes.

Compared to a golf cart or an automobile starter motor, a ham rig is a low power device. 20 amperes should be plenty of current. A deep-discharge storage battery is best, but there's no reason you can't use an old car battery. Those of us who live in snow country routinely discard car batteries when they can no longer supply 400 amperes on a frigid January morning. But even an old car battery will usually supply 20 amperes for five minutes without appreciable voltage drop. All you need to do is keep a small, one-ampere charger on it continuously. For anyone on a budget, a free used battery is a darn cheap high current regulated supply.

Requirements of a line-powered 20 ampere, 12 volt supply

I don't plan to build an equivalent line-powered power supply. But if I did, I wouldn't build a switching power supply. I've had zero luck running ham equipment on switchers. If the switcher is more than just an asynchronous charge pump and actually regulates the output voltage, then it will make RF noise which you'll hear in your receiver.

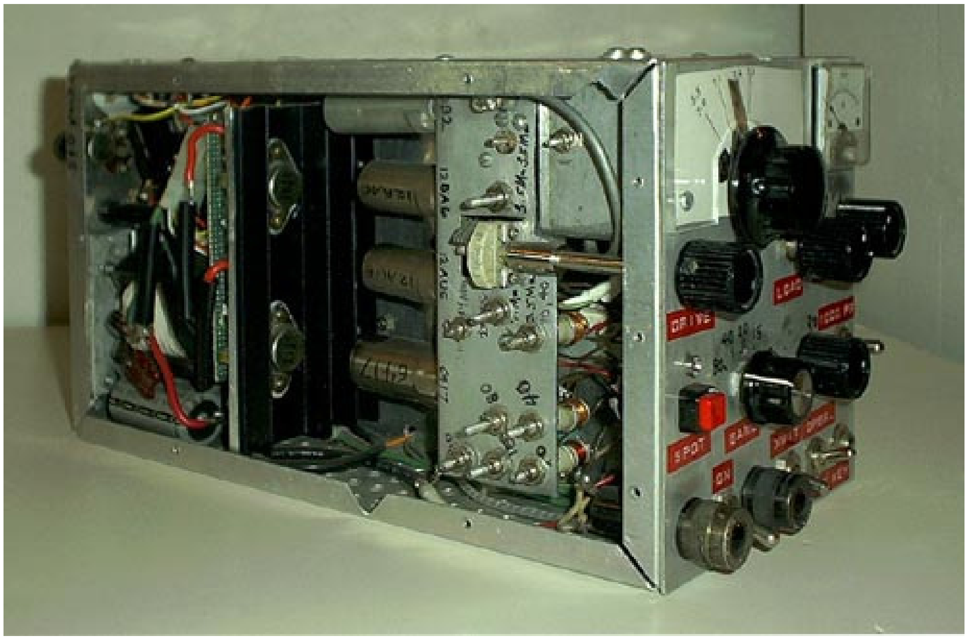

A 1967 homebrew all-band transmitter with a switching power supply running on 12 volts DC

The 150 watt switching supply is at the left rear. Originally the supply was a simple charge pump that used germanium power transistors. It consisted of a free-running multivibrator (square wave oscillator) circuit that drove a step-up transformer with no feedback. The supply worked well and didn't produce audible hash in my receiver. Then one day a germanium transistor blew and I was unable to find a suitable replacement. Germanium transistors seem to be extinct in the modern world – they can’t really compete with MOSFETs. I rebuilt the supply using a modern pulse width modulator regulation system and MOSFET power transistors. An elegant feature of the new switcher was that the output voltage was adjustable. Just by turning a knob I could vary the transmitter output power from 20 to 80 watts.

Unfortunately, the switcher and the RF amplifier stages “talked to each other.” That is, when the load increased, the supply increased the pulse width modulation to compensate, but not without a slight delay. This subtle ripple or “jitter” was hard to see on a scope, but no matter how much I filtered the DC high voltage, the jitter appeared as a slightly rough note on the CW signal. In practical terms, I kept getting 598 RST reports.

There was an elaborate 12 volt switcher in QEX magazine a few years ago that solved the jitter noise problem. I was delighted to see that I wasn’t the only one who noticed “jitter noise.” The problem is real, but after seeing the complicated Rube Goldberg solutions, my desire to build one vanished.

If I had to build such a supply again, one method would be to build a charge pump or other unregulated switched system that boosted (or lowered) the input voltage to just above the desired voltage. Then I could use a linear voltage regulator to accomplish the regulation function with a minimum of waste heat generated. This is the low noise approach I used to power my super-regulated miniature VFO power supplies described in chapter 10.

One idea I’ve had for a high power charge pump would be to use a manually adjusted pulse width for the AC generation. This way the output voltage could be manually set so that the linear regulator would waste as little energy as possible. This regulation margin would depend on how consistent my line voltage was and how stiff the charge pump conversion system was. For example, starting with 120 volts AC, the charge pump might lower the voltage to say, 18 volts DC. Then the linear could reduce the 18 volts DC down to 12 volts.

Going solar

Because my automotive trickle charger was feeble, I was charging non-stop. Even then, I was having trouble keeping the battery charged for a couple hours of daily operating. However, I already owned a 12 watt, 12 volt solar panel which I installed on the roof.

21. Chapter 8, Harris

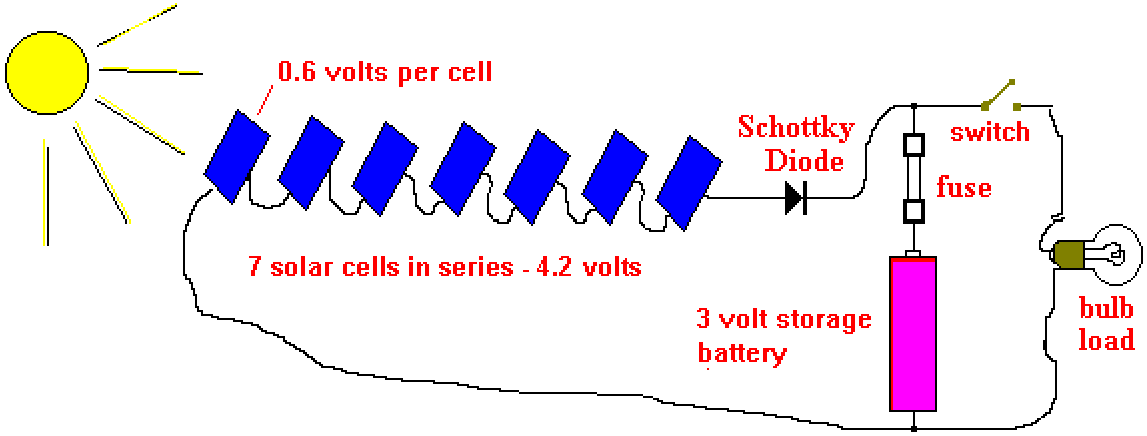

A simple solar charger circuit

Solar cells are a kind of silicon diode. They are arranged in series so that the forward voltage drop of each diode adds up to some voltage greater than the voltage of the storage battery. For example, at 0.6 volts per solar cell, we need at least 20 cells in series to raise the panel voltage higher than the 12 volt battery. Typically, an open circuited panel puts out 20 volts in bright sunshine. This extra capability insures that it will continue to charge a 12 volt battery all day and implies that there are roughly thirty 0.6 volt cells in series. Solar cells are interesting to play with. I was surprised to discover that if you put your hand over just one of the series cells, it turns off the whole string, something like Christmas tree lights wired in series. This means that just one wet leaf stuck on your panel can turn off the whole array.

The panel output current is proportional to the sine of the angle the sunlight makes with the panel. If your object is to produce the maximum kilowatt-hours during the entire year, then the angle should equal the latitude. I mounted my panel at 45°, which at 40° latitude, gives me better performance during the winter. Optimizing for winter is smart because the days are short and the panel is often covered with snow. Actually 50° may be better here in Colorado. Snow slides off a steep incline and higher angles are more resistant to hail damage. On the other hand, mounting them on a steep frame may make them subject to wind damage. There is no way to win. Many people just mount the panels flat on whatever roof they happen to have. If you wish to sacrifice some percentage of the output for beauty or mounting convenience, that's your decision.

The "lunar discharger"

It's bad form to connect a solar panel directly to a storage battery. The solar cells are forward biased diodes with respect to the battery. Therefore, whenever the sunlight quits, the panel voltage may drop below the battery voltage and a small battery current will reverse and flow through the solar panel. In other words, at night, the solar charger becomes a "lunar discharger." (Never mind. That was a silly joke.) If there are enough extra solar cells, then the night-time voltage drop may still exceed the battery voltage. In any case, the discharge problem is typically prevented by a silicon diode in series with the panel to insure that current is never allowed to flow from the battery back into the panel. Since this check valve diode has a forward voltage drop that wastes energy, you may as well use a big Schottky diode that will only penalize you with 0.2 volts instead of 0.6 volts loss.

Another small issue is lightning protection. My panel is on the roof where, in theory, it might attract a lightning bolt. I'm not really worried, but I have a connector on the lead from the panel so I can disconnect it from my shack when I go away on trips or whenever a storm is particularly frightening. If I lived on a mountaintop or if my house were higher than the other houses in my neighborhood, I would be more concerned.

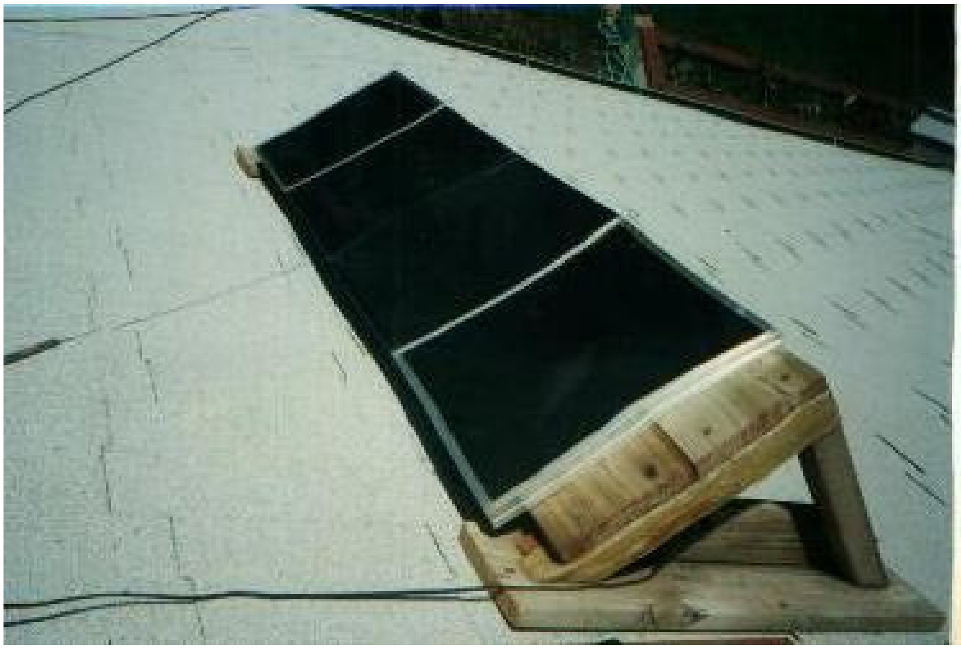

Solar panel on the roof

I built my panel frame out of varnished wood. The wooden boards screwed down to the roof retained water against the shingles and rapidly rotted. I replaced the boards with 2 inch aluminum angle stock which seems to be a more permanent fix.

Conservation helps

My solar panel charges about as much as the line-powered trickle charger, about 1 ampere. Consequently I still had to use line power to charge occasionally whenever I stayed on the air too long. A breakthrough came when I built my new transistorized receiver, the one described in Chapter 13. The new receiver draws 120 mA. That sure beats 3,500 mA. Since then, I haven't needed my line-powered charger.

12 Volt power distribution

A storage battery can deliver hundreds of amperes, so it's wise to isolate the battery from your rig and solar panel with an appropriate fuse. A 30 ampere fuse should be about right. I have a master switch to isolate the battery from the ham rig when I'm not using it. A little LED pilot light tells me when it's connected and a voltmeter warns me if the battery is not fit for use. I also have charging and discharging ammeters (ampere meters) so I know the status of my supply at all times. I find the discharge meter helpful for monitoring my transmitters. You can even tune the antenna while watching the battery current. The wire in front of the charge meter goes to my 12 volt desk lamp.

Battery charge monitor

Maintaining clean battery contacts is critical to keeping the supply voltage constant under load. Every few months, I clean the contacts using baking powder and water. I know the battery contacts need cleaning when the pilot light on my transmitter begins to flicker noticeably while I'm transmitting. I top off the batteries as needed with distilled water to keep plates in the cells from being exposed to the air. A gel-cell battery shouldn't have these corrosion and fluid loss problems. However, if you overcharge a sealed gel-cell, the water cannot be replaced and the battery will be permanently damaged.

The 12 volt storage battery power system

The storage batteries are shown above. The transistorized inverter to the left of the batteries converts 12 volts DC to 120 volts AC for those rare times when a storm knocks out our commercial electricity. On those occasions I have simply strung an extension cord upstairs to power the TV and few lights. The glass and rubber device is a ***hydrometer*** for measuring the specific density of the battery acid. When the glass float rides high in the green zone, the battery is fully charged.

Storage battery safety

My power system is small and I rarely have to add distilled water to my batteries. This tells me that my use of stored energy and the output of my small solar panel are well-matched. If I were having to add water constantly it would mean that I was wasting energy and converting the water into gaseous hydrogen and oxygen. This situation could be potentially dangerous. If I had a large system powering more of our household needs, the likelihood an explosion would be serious. Therefore, a large storage battery array should be located out in a well-ventilated place like a garage. Moreover, the state of the batteries should be continuously monitored with an electronic regulator to be sure that the batteries are not overcharged and generating significant amounts of hydrogen. Charge control regulators are mandatory in any large system.

I once read an article written by a fellow who was in charge of maintaining remote microwave relay stations. Wind turbines powered the stations, but the power systems were not equipped with any means to turn off the charging current to the batteries during windstorms. The results were catastrophic explosions inside the stations. Eventually, the phone company realized the problem and put in charge control regulators.

In summary, I get a kick out of operating like a fully solar-powered Field Day station. Because the whole station runs on 12 volts, in theory I'm always ready for Field Day and there's nothing to stop me from going mobile. One of my back-to-nature dreams is that, during a power failure, I can go right on operating and thumb my nose at the evil corporate power monopoly. Twice during power failures I have galloped downstairs in the dark to make my first 100% truly independent radio contact. But before I could even tune the transmitter, the *#@%!$ lights came back on. Oh, curse those efficient fellows at Xcel Energy Company!

****

LED THERE BE LIGHT

Emergency Station Lighting

As explained above, I run my whole station on solar energy stored in a 12 volt battery. An advantage of battery power is that I operate free from the power grid. This saves very little money, but it's quite romantic. Of course at night, even on batteries I need to keep a log and throw the correct switches. Using candles or a flashlight clenched in my teeth didn't seem very sophisticated.

My first idea was to dangle a 12 volt automotive light bulb over my shack. My buddy Bob, N0RN, uses such a light in his tent on Field Day. This solution is completely practical, but it seemed old-tech and boring. It also draws about 10 watts of power. If I am going to all the trouble to build a custom 12 volt light, I figured it should at least be ***an interesting light***. Modern lighting methods produce more light per watt than tungsten bulbs. Also, energy conservation is important when your station is powered by a battery charged with a small solar panel.

A homebrew LED and flourescent station lamp

White LEDs – a modern marvel.

I happened to notice some "white LEDs" for sale down at Radio Shack. (Part # 276-320) I thought cynically that they were probably really dim and not really white. I bought one and hooked it to a variable power supply. I slowly turned up the voltage until I obtained the rated 20 mA DC current. Hey! Not bad! The beam of white light has a slightly bluish cast, but otherwise, it throws a spot of light out to a couple feet. It's rated at 1100 mcd. That's impressive brightness when you consider its beam angle is 100^o. One LED makes a perfect key chain light and many LED flashlights are now on the market.

I wrote this section some years ago. Since then, screw-in 12 volt light bulbs consisting of large clusters of white LEDs have become available. I now have one in the shack that draws 3 watts and delivers considerably more light than the two systems below. Even so, there are lessons to be learned from my experimenting below. The switching power supply described for the fluorescent light is a design I have used often for several applications.

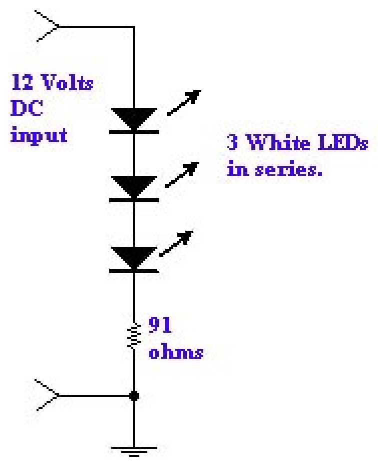

I'm told these white LEDs are made from sophisticated layers consisting of yellow and blue LEDs and a phosphorous that glows white. As LEDs go, it has a rather high voltage drop, 3.6 volts. I figured if I put three in series plus a single dropping resistor, it could run on 12 volts. The current is limited with just one 91 ohm resistor. The resulting triple light consumes just 0.25 watts and is adequate for keeping a log and operating the station. That's amazing when you think about it. We routinely squander hundreds of watts to illuminate entire rooms when all we really need is ¼ watt to read our newspaper or book.

Flourescent lighting

Unfortunately, "adequate" isn't the same as "comfortably bright." So my next project was to build a 12 volt powered compact fluorescent. I fiddled for a couple hours trying to build a power oscillator capable of producing the required 800 Volts AC needed to ignite the miniature fluorescent bulb. I had several transformer and switching transistor problems that I was having trouble solving. I also fried my voltmeter because I forgot to change the range when I put it on the high voltage. There's nothing like dead test gear to make you think differently.

I remembered that I still had a 6 volt fluorescent lamp that I used to use for lighting my pup tent while camping. One day I rolled up the tent while the light was still hanging from the tent ceiling. Oops! The plastic housing was splintered and the remains of the light assembly were still in my junk box. I found the parts of the light and resoldered the broken wires. I connected it to a 6 volt power supply and it still worked fine. Next I built a shiny reflector/ lampshade out of sheet aluminum and it produced loads of light. Unfortunately, I need a 12 volt light, not a 6 volt light. I didn't have a second light to put in series with the first, so I used a big dropping resistor to run the light on 12 volts. That worked OK, but it seemed pretty crude. Besides, I was already "cheating" because I hadn't built the fluorescent high voltage supply.

Nifty "buck-type" switching power supply

I replaced the dropping resistor with a switching power supply to reduce 12 volts DC down to 6 volts. I got the design from the National Semiconductor data book. Using the switcher the fluorescent light draws about 2.5 watts at 12 volts. The switcher is about 80% efficient. That is, it dissipates 10% of the total energy. In contrast a resistor would have dissipated 50% of my energy. I have used this little supply design for several home projects so far and found it thoroughly reliable. If you're a homebuilder, this simple, buck-switcher belongs in your bag of tricks. If you're bored with building a 12 volt light, maybe this little switcher will interest you.

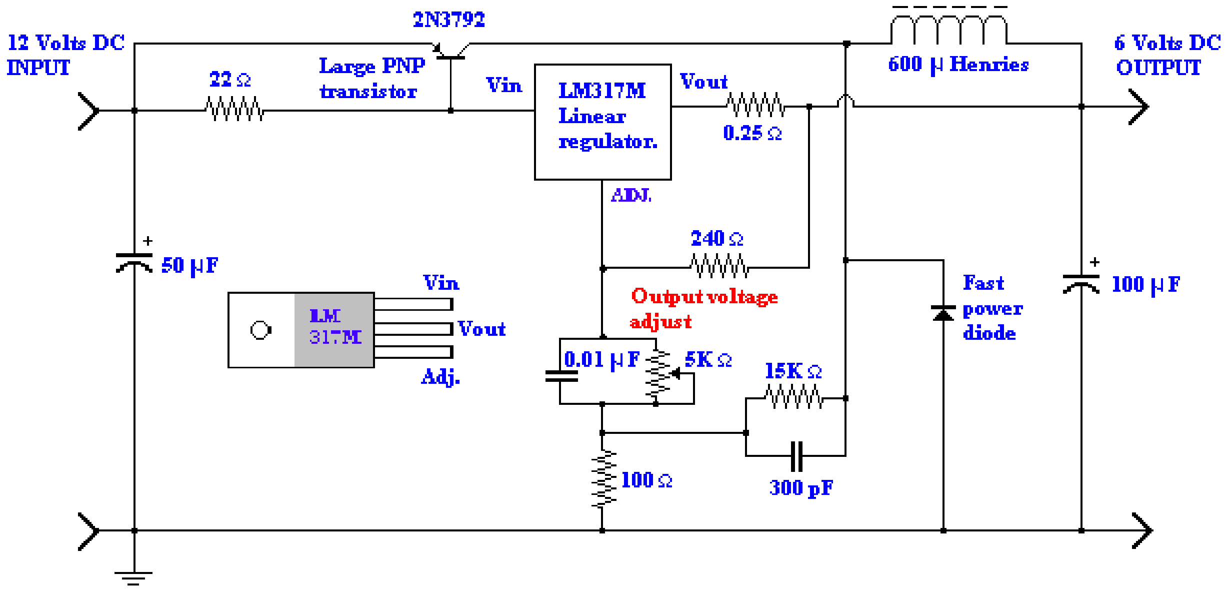
12 Volts DC
INPUT
22 Ω
2N3792
Large PNP
transistor
Vin
LM317M
Linear
regulator.
Vout
0.25 Ω
ADJ.
600 μ Henries
6 Volts DC
OUTPUT
+
50 μF
Vin
LM
317M
Vout
Adj.
240 Ω
Output voltage
adjust
0.01 μF
5K Ω
15K Ω
300 pF
100 Ω
Fast
power
diode
+
100 μF

STEPDOWN SWITCHING POWER SUPPLY

(12 Volt DC to 6 Volt DC Converter)

Switching power supply made from a linear regulator

How can this work?

This switcher uses a LINEAR regulator to generate the pulsewidth modulation needed to implement a switching power supply. This is essentially the same linear regulator used in the suggested line-powered QRP power supply. But, when you put a scope probe on the inductor, sure enough! The linear regulator is switching full on and off in rectangular pulses. Apparently the 15K Ω resistor and 300 pF capacitor are coupling a feedback pulse from the inductor into the regulator to cause it to switch full on and full off. The pulse width varies with load and responds just like a real switcher. Aside from my amazement that linear regulators can work as switchers, I was also surprised that I hear very little switcher noise in my receiver. The RF noise it produces is apparently sufficiently isolated from the receiver that I rarely hear hash competing with the ham signals. On the rare occasions when I do hear it, I simply turn off the fluorescent.

Subjectively, the fluorescent produces roughly twice as much light as the LEDs. In other words, for ten times as much power the light seems to be twice as bright. On the other hand, it illuminates a much wider area than the LEDs, so I just run both of them simultaneously. The LEDs illuminate my log and scratch pad while the fluorescent lights up the station as a whole. I’ve used this light for hours at a time. It’s bright enough that I’m content with it and I don’t think about running across the room to turn on “the real lights.” In summary, battery power and energy conservation are entertaining games. And, if we’re really unlucky, someday our hobbies might even be useful in a community crisis.

CRYSTAL SETS TO SIDEBAND
© Frank W. Harris 2010, REV 12

Chapter 9

ACCESSORIES & SKILLS

A MORSE CODE KEY

"Straight keys"

A telegraph key for sending Morse code is an easy mechanical project. All that's needed is a reliable spring-loaded switch. The fingers contact a simple knob on a lever. The operator rests his wrist and forearm on the table and grips the knob with two fingers and a thumb. The operator pushes down on the knob, closing the switch contacts. A "straight" key like this is used to send both dots and dashes. The timing of the dots and dashes is totally dependent on the operator.

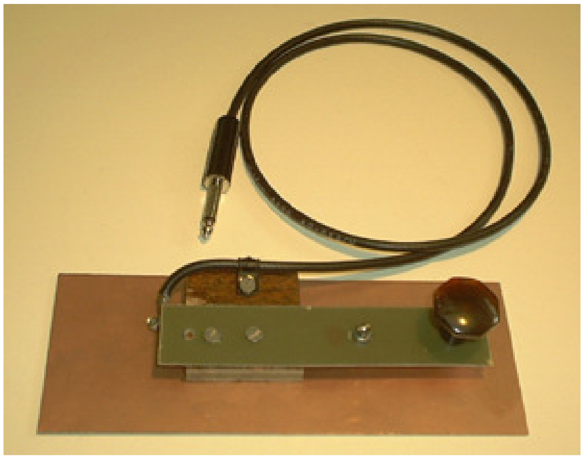

A simple telegraph key

Commercial straight keys, especially the old ones, are made from elegant machined brass parts and springs on engraved Bakelite plates. The exact tension and gap width can be adjusted to each operator's preference. This sophistication is nice, but not essential for a beginner sending at low speeds. A straight key is good for code speed up to about 15 words per minute. Faster speeds are tiring and hard to send well.

The straight key shown above was made from two pieces of single-sided printed circuit board. The upper sender lever board serves as the spring and its underside is the conductive surface. The switch contact consists of a machine screw that touches the grounded lower board when the lever is pushed down. Two nuts on the screw adjust the contact distance. Most operators like to adjust the switch contacts so that the travel is about 1/32 inch. The spring action of the lever should be strong enough so that it breaks the contact sharply when the lever is released, but not so much force that it is tiring to use. The two pieces of circuit board are insulated from each other by a small block of plywood. The knob is a plastic pull handle from a drawer.

Several articles in QST have appeared in recent years describing how to make keys out of household junk. Most of these keys are really "paddles," rather than straight keys. Paddles are keys that are pressed side-to-side instead of just downward. Paddles have two switch contacts and are used to control automatic "keyers" of the kind found in modern transceivers. When the paddle is pushed to the left, the keyer automatically generates perfect dashes. When the paddle is pushed to the right, the keyer pauses the exact length of time and then automatically makes perfect dots. It would not be hard to adapt the mechanisms described in those magazine articles to make a “straight,” up-and-down key.

Mechanical bugs

No, we’re not talking about a mechanical glitch, we’re talking about a type of telegraph key. The next step up in sophistication from a straight key is a mechanical ***“bug.”*** This telegraph key is activated with a paddle. When pushed to the right, it automatically makes dots, so long as the operator deflects the paddle. Mechanical bugs make the dots with a weighted beam that swings back and forth propelled by a weak spring. A dot occurs whenever the swinging lever closes the “dot switch.” When the bug paddle is pushed to the left, it closes the “dash switch.” The operator must make each dash manually. So unlike a modern keyer, the operator provides the timing for dashes. Mechanical bugs like this were standard among commercial radio-telegraph operators and hams for many years. Railroad telegraph operators also used them.

You can still buy commercial mechanical bugs. At one time there were even complex versions that generated both dots and dashes automatically. A mechanical bug is a difficult basement project without a machine shop. It would require a great deal of patience to make a reliable mechanical bug, but it can be done.

Homebrew electronic keyers with automatic dots and dashes are a common homebrew project. Most hams just buy a keyer kit that has a tiny, pre-programmed “PIC” microcomputer chip that does all the difficult timing chores. Since a pre-programmed chip didn’t fit my rules for “homebuilt,” I built a homebuilt electronic bug that makes automatic dots but requires manual dashes.

****

A HOMEBREW ELECTRONIC BUG

No matter how I adjusted the screws and cleaned the burned contacts on my 45 year old mechanical bug, the dots sounded more like static. Of course, if I had bought a quality bug in the first place, it would still be working. While I was trying to get the old bug working properly, it occurred to me that my key was the only part of my rig that wasn’t homebrew. Ah ha! - a challenge!

I didn’t see how I could build a decent mechanical bug with my limited tools. However, I figured that an all-electronic key that generated both dots and dashes automatically couldn’t be too hard. I began prototyping a logic circuit based key on a large plug-in board. I quickly discovered that implementing automatic dashes wasn’t so simple. The dashes had to be timed with respect to the dots and there could be no overlap. Moreover, the spaces between dots and dashes should be enforced regardless of how inept the operator might be. I soon had about 20 CMOS ICs wired in a complex mess of logic circuits that nearly worked. But no matter how many more gates I added, I always seemed to have glitches. This was becoming frustrating. Also, my new key was going to end up as a foot-long circuit board. I lowered my sights to building a simple electronic bug built with op-amp oscillators. The new key would have automatic dots, but manual dashes. In other words, it would be the electronic equivalent of a mechanical bug.

A homebrew electronic bug

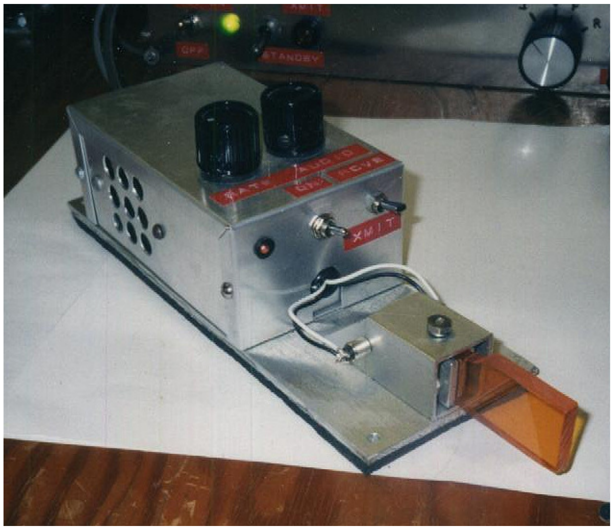

The mechanical parts

The mechanical requirement was to make two momentary-contact, spring-loaded switches controlled by a single paddle. I suppose I could have used printed circuit board switches like the straight key discussed earlier. However, my solution was to use two miniature, push-button momentary switches. The switches can provide both the switching and spring action. I mounted them on opposite sides of a piece of square aluminum channel so that the buttons face each other. The back of the plastic paddle arm rests between the two buttons and pushes one or the other as needed. I cut out the paddle from a plastic sheet that I bought at a local plastics shop. By the way, scrap plastic is a great resource for material to make antenna insulators, stand-offs, boxes, etc. The bug is packaged in a commercial aluminum box and screwed onto a thick aluminum plate. I glued a piece of sticky rubber from an old mouse pad onto the bottom to prevent sliding from side to side.

After I got my bug working, I had trouble sending accurately. That is, I kept sending extra dots or half-formed dots. The biggest problem turned out to be insufficient return spring force. I supplemented the spring force of my button switches with coil springs that I stuffed into the aluminum channel on both sides of the plastic paddle. Afterwards I was surprised how much easier it was to send good code.

Another annoyance was that the key kept sliding on the table, so I screwed metal "railings" onto my tabletop to confine the key. The key continued to creep away from me so I finally just screwed it down to the table. Now I can bang away and the key stays put. Good sending is just plain hard. Any advantage you can give yourself is worth doing.

The electronic parts

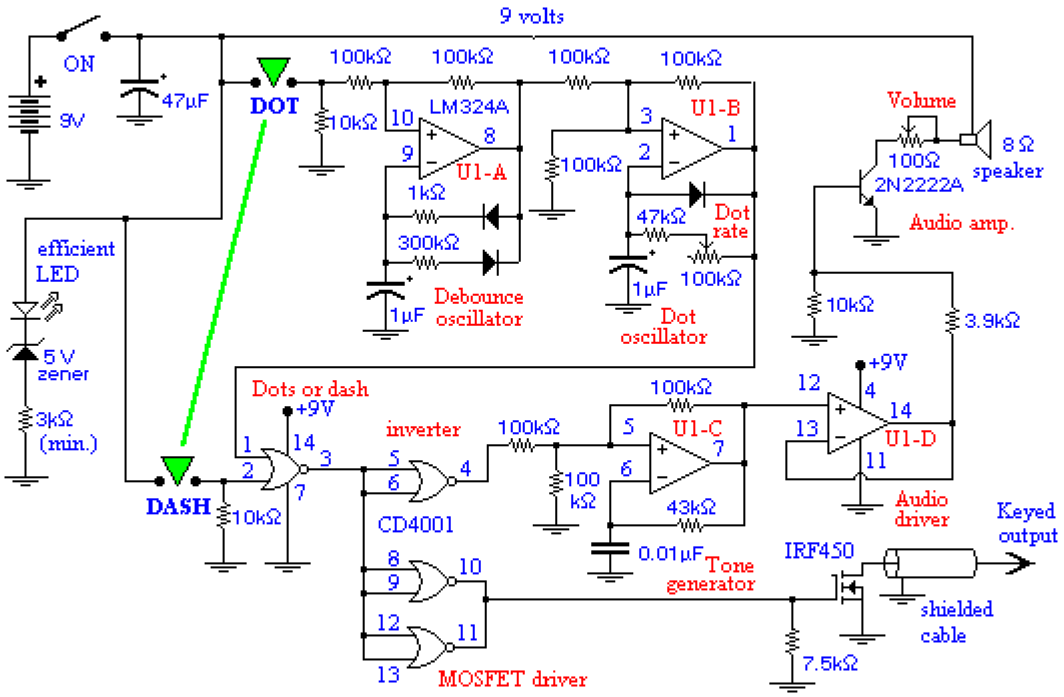
9 volts
ON
+
9V
47µF
DOT
100kΩ
100kΩ
100kΩ
100kΩ
10kΩ
10
LM324A
9
8
U1-A
1kΩ
300kΩ
1µF
Debounce oscillator
100kΩ
3
U1-B
2
1
47kΩ
Dot rate
100kΩ
1µF
Dot oscillator
Volume
100Ω
2N2222A
8 Ω speaker
Audio amp.
10kΩ
3.9kΩ
efficient LED
5 V zener
3kΩ (min.)
DASH
10kΩ
Dots or dash
+9V
1
2
14
3
7
inverter
5
6
4
CD4001
100kΩ
100 kΩ
100kΩ
5
U1-C
6
7
43kΩ
0.01µF
Tone generator
+9V
12
4
13
14
U1-D
11
Audio driver
Keyed output
IRF450
8
9
10
12
11
13
MOSFET driver
7.5kΩ
shielded cable

Circuit Diagram for the Bug

My bug has two unusual features. It has a **built-in sounder** and it can **key any positive voltage to ground up to 400 volts.** The transistor that does the actual keying is a 400 volt N-channel MOSFET power transistor. I built the bug while I was still using a vacuum tube transmitter with a 6146 final. Positive 300 volts cathode voltage was on the key, so I needed high voltage capability. A keying relay would have worked, but after my experience with the old mechanical bug, I didn't want big currents flowing through mechanical contacts. The IRF450 N-channel MOSFET transistor handles any positive signal from 5 volt logic up to several amperes at high voltage. Of course you may use whatever size MOSFET is appropriate for your transmitter.

The built-in sounder consists of a small speaker and an audio oscillator. A potentiometer adjusts the volume. When I was on the air, I used to listen to my own signal in the receiver. Unfortunately the sound from my receiver was distorted from the transmitter being so close. The distortion led to poor sending. Too much coffee may have been another reason. Anyway, by listening to clean tones from the sounder, my “fist” is likely to be as good as it can be. The sounder can also be used as a stand-alone code practice sender.

My transmitter and receiver are more primitive than commercial transceivers. I still have to turn on the transmitter and mute the receiver manually. To do this more quickly, I installed a small toggle switch on the bug just to the right of the key paddle. The switch is just a contact to ground. Shielded leads connect the switch to both the transmitter and receiver. The ground lead triggers the Transmit-Receive relay in the transmitter and sets the receiver power supply to standby. It isn’t exactly “break-in keying,” but it’s OK for the time being.

Circuit Description

The bug runs on a 9 volt alkaline battery turned on by a small toggle switch. An efficient red LED that lights on just one milliampere reminds me to turn it off. When it's time to change the battery, a 5 volt Zener diode in series with the LED makes the LED dim when the voltage gets down to about 6 volts. When not sending, the bug draws 1.2 milliamperes. This means the 550 milliampere-hour alkaline 9 volt battery will last over two weeks if I forget to turn it off.

All oscillators in the circuit are implemented with an LM324 quad op-amp. Op-amps were discussed in more detail in chapter 7. The LM324 is a great old component. It doesn't work at high frequencies like modern op-amps, but that can be a good thing because it doesn't oscillate unexpectedly. And unlike nearly all the old op-amps, such as LM458, LM741 or LM301, the LM324 almost always acts like an "ideal op-amp." It only needs one supply voltage and it rarely surprises you with "practical limitations."

The other component I haven't discussed before are the CMOS logic circuits, specifically the CD4001 NOR logic chip. Logic circuits consist of small arrays of transistors used for making simple binary decisions. For example, if you have two binary (high or low) inputs and you wish to know when either or both of the two inputs is high (a "one"), then these two input lines can be connected to the inputs of a two input "***OR gate***." When either or both lines go high, the output gate of an OR gate will switch from low to high. In this circuit we are using a "***NOR gate"*** which is also known as a "NOT OR gate." These work the same as an OR gate except that when either or both the inputs go high, the output will switch in the opposite way. That is, the output line will drop from high to low. "***AND gates"*** and "***NAND gates"*** are similar except that the output changes only when both inputs go high simultaneously.

The 4000 series CMOS logic gates are also old-tech and quite slow relative to modern CMOS logic chips. The "CD" part just identifies the particular brand of gate. The 4000 number identifies the particular construction technology of the logic gate. These 1970s era parts are immune to interference from RF and tolerate up to 18 volts peak. Modern CMOS gates are limited to lower voltages and are designed for performing logic at high speed. High speed isn't needed here and just makes the chips vulnerable to interference and self-oscillation.

It's necessary to "debounce" the dot key. First I just tried turning on the dot-generating oscillator with the dot switch. Because my timing wasn't always matched to the oscillator speed, and because the contacts didn't always close solidly, the dots often sputtered like my old bug. Referring to the circuit diagram, the dot switch now keys a multivibrator (U1-A). This generates a narrow, repeating pulse that sets the maximum dot rate. The actual dots are formed by a second op-amp multivibrator, (U1-B). The pulses from (U1-A) start a new dot, provided that the second oscillator, (U1-B), is ready to start one. The dot rate is adjustable over a wide range using the 100K pot.

The dots and dashes are combined with a CMOS 4001 NOR gate to make a single keying signal. That is, the NOR gate puts out an output if either a "dot" or a "dash" is present. Two more NOR gates serve as a signal polarity inverter and driver for the keying transistor and for the audio oscillator. The NOR gate output turns on the audio oscillator (U1-C) by pulling up the hysteresis 100K feedback resistor. The audio oscillator in turn drives an op-amp voltage follower (U1-D) and the 2N2222 speaker driver. The audio volume control is simply a 100 ohm pot in series with the 8 ohm speaker.

I've used my bug for three years. I've had hundreds of QSOs with it and replaced the battery about twice a year. Once the paddle began to stick "on," so I had to oil the machine screw pivot on the paddle. Otherwise, the bug has given me good service and little trouble. This is an easy home project because it doesn't involve RF and is almost guaranteed to work.

DUMMY LOADS

A checkout of a homebrew transmitter begins with a low inductance dummy load. Once you are able to deliver a pure, stable sinewave into your dummy load, then you may graduate to an antenna.

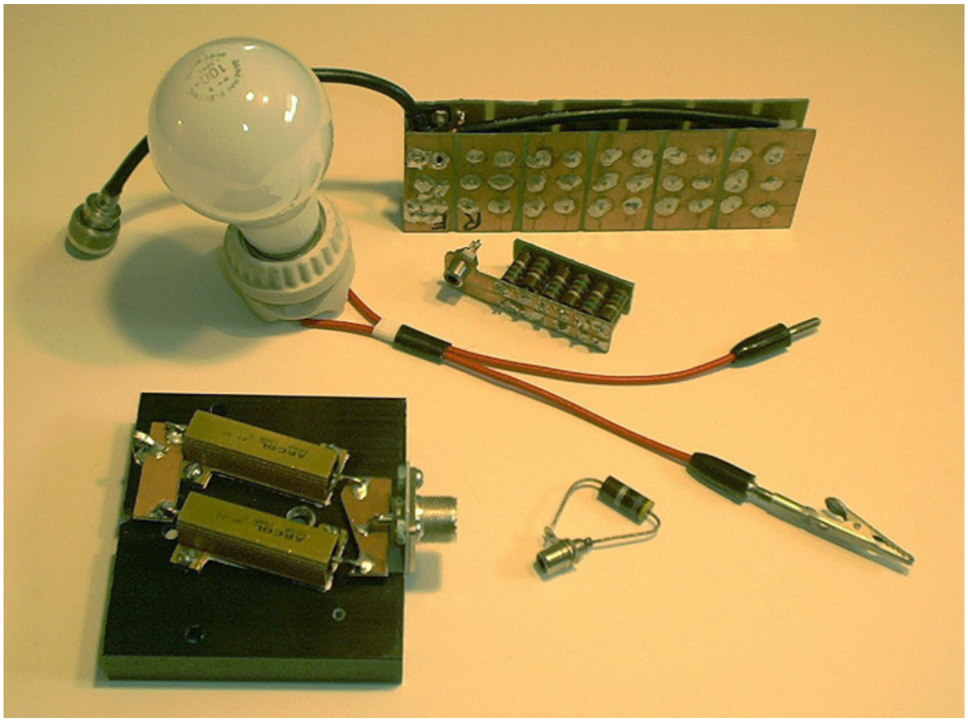

Homebrew dummy loads ranging from 2 to 100 watts.

A dummy load is just a big resistor that can stand the power from your transmitter without smoking or catching fire. In order that it resembles a well-designed antenna, the dummy load should be a pure resistance with little accidental inductance or capacitance. Big, non-inductive resistors are pricey and must be specially ordered. That's why most hams buy commercial dummy loads rated for power equal or greater than their transmitter's maximum power. Dummy loads are almost always designed for 50 ohms because that is the standard design impedance for most transmitter amplifiers and many antennas.

For a QRP, a dummy load can be as simple as a 47 or 51 ohm 2 watt carbon composition resistor. Single, 50 ohm resistors are rare items because the standard resistor values are set up logarithmically. If you don't drive it continuously, five watts won't hurt it. Be careful, though. The heat may not damage the resistor, but it can melt oscilloscope probe tips clipped onto the hot metal resistor leads. For higher powers, you can build a good dummy load out of arrays of carbon composition or other low inductance resistors.

A dummy load is one of the few simple quality homebrew test instruments that a guy can make for himself in a few hours. A big dummy load resistor can be made from an array of low power, low inductance resistors. Metal film resistors and carbon composition resistors usually have low inductance and make good loads. For example, I happened to have a whole package of old 1 watt, 150 ohm carbon composition resistors. By placing them in the appropriate series/ parallel arrangement I made a 10 watt, 50 ohm dummy load. If you need a 50 watt dummy load, it can be made from an array of 25 two watt resistors. You will have to be clever selecting the resistance and arranging them so that the final resistance is 50 ohms.

For high frequencies, like the 10 meter band and above, it is important to use a low inductance connection with the resistance. Therefore, the leads into the dummy load should be coaxial cable. I used circuit boards with wide traces to connect the individual resistors in an array. In the large load it was also necessary to connect one end of the array to the other internally. I first used a simple wire for this and it had significant inductance. I replaced the wire with a piece of coax grounded at one end and the inductance disappeared.

Beware of high power, "low inductance" resistors

I bought a pair of supposedly "low inductance" 100 ohm, 50 watt resistors. I put them in parallel and mounted them on an aluminum heat sink. They should have made a terrific high power dummy load. Alas, there was considerable inductance in these resistors and, even in parallel, at 14 MHz they had twice the impedance I expected. In summary, this big dummy load is useful, but it isn't 50 ohms.

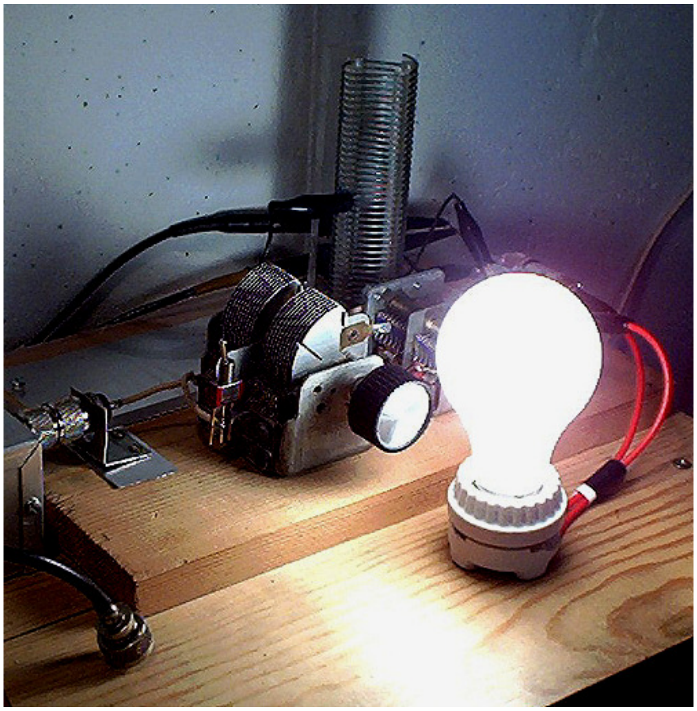

Light bulbs as dummy loads

An ordinary filament light bulb can work as dummy load. They are far from ideal and should not be your only dummy load. However, they do have advantages: They are fun to use and great show-and-tell devices. However, at less than 100 watts, their impedances are much higher than 50 ohms and the impedance rises as they get hot. On the other hand, if you need a dummy load that simulates a high impedance antenna, then a light bulb is useful. Also, if you're checking out a power meter, a light bulb gives an obvious indication of when the power output is maximum. Also, when a 40 watt bulb shines at full brilliance, obviously it is dissipating about 40 watts. Similarly, small Christmas tree light bulbs can be useful for checking out QRP transmitters.

****

BUILDING A "T- MATCH" ANTENNA COUPLER

Modern transmitters are usually designed for 50 ohm antennas. The Chebyshev output filters found in modern transmitters only work at this design impedance. When you load them with a mismatch, you won't believe the scrambled waveform that comes out. In other words, they only filter properly at 50 ohms. Unfortunately, real antennas are usually higher or lower impedance and need some form of impedance transformer to make them appear as a 50 ohm load to the transmitter.

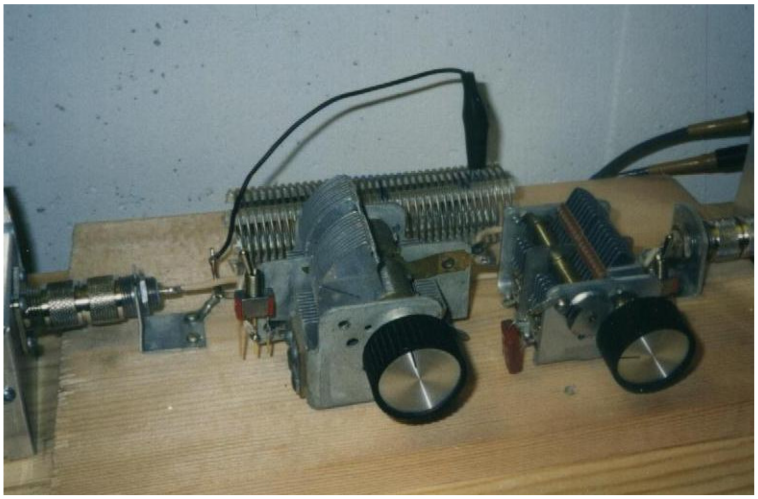

Old vacuum tube "Pi output" transmitters had built-in antenna couplers, but none of the transistorized designs in the ARRL annual handbooks have couplers. So after I built a transistorized transmitter, I needed an antenna coupler. The coupler designs in the handbooks were complicated with all-band capability and built-in power / SWR meters. They're nice, but I wanted to get on the air quickly. I consulted Bob, NØRN, for advice. Bob had built several couplers and he thought a simple T-Match was best. It just consists of two variable capacitors and a variable inductor.

A Simple T-Match Antenna Coupler

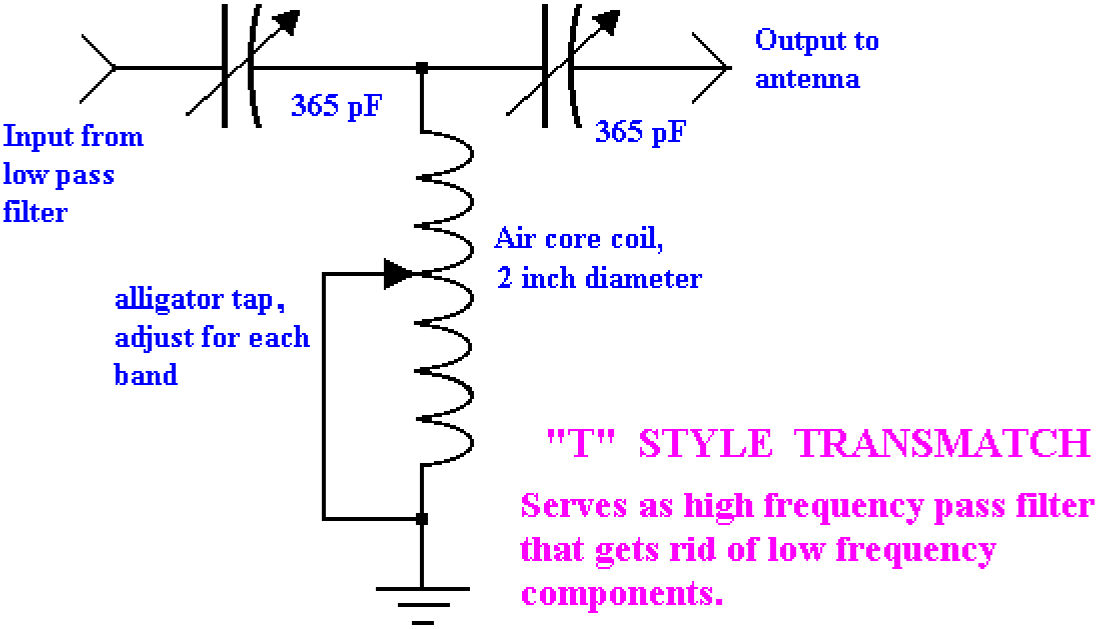
Input from low pass filter
365 pF
365 pF
Output to antenna
Air core coil, 2 inch diameter
alligator tap, adjust for each band
"T" STYLE TRANSMATCH
Serves as high frequency pass filter that gets rid of low frequency components.

How it works

The idea behind the T-Match is to resonate the variable capacitor on the left with the inductor to ground. When tuned to resonance the oscillation produces sinewave voltages across the L and C that can be much larger (or smaller) than the sinewave voltage that arrives on the input. Because the voltage across the inductor can be huge, the coupler can “match” the impedance of a high impedance antenna. The oscillating L and C work like a transformer, stepping the voltage up or down. For example, to drive a 300 ohm antenna, the impedance of a 50 ohm transmitter output will have to be made to "look like" it's six times higher in order deliver the same power. For this to be true, the voltage across the antenna will have to be 2.5 times higher.

The variable capacitor on the right is not nearly so critical. You will find that for most situations keeping the capacitor at full capacitance, 365 pF, produces the best signal. Sometimes I can get a little better sinewave output or slightly more amplitude by adjusting this capacitor. But usually, I just leave it alone. For the lower bands, 80 and 160 meters, you may need to add more capacitance in parallel with both variable capacitors using switches. My variable capacitors are dual-section types from old broadcast radios. I use small toggle switches to add the capacitance from the second sections. In addition, the right hand capacitor has a fixed 1000 volt, 200 pF mica capacitor across the second section so I have plenty of coupling capacitance for 160 meters.

Packaging

My friend Bob showed me a T-Match he had built in a wood and plastic box.

“Why didn’t you use a metal box?” I asked.

“Oh, wood was easier and it doesn’t matter,” explained Bob. “Maybe I get more RF radiation here in the shack, but otherwise, there’s no need for a metal cabinet.”

I was anxious to get on the air, so I rummaged through my junk and came up with two dual section 365 pF broadcast radio variable capacitors. One of them was out of a 1935 radio. I thought using an ancient component had nostalgic charm. I also found a big piece of open “Air Dux” coil I could use for the inductor. I made the inductor “variable” by using an alligator clip to short out the unwanted length of coil. For RF connectors, I used SO-239 UHF connectors from Radio Shack. After an hour’s work I had screwed all the parts down onto a pine board and I had a T- match.

Improving performance on the higher bands

I originally used a simple 12 gauge wire connecting the grounds on the input and output connectors. This simple wire looks like a significant inductor on 17 meters and above. This small inductance can occasionally make loading difficult or impossible with some antennas. I replaced the wire with a low inductance, 3 inch wide sheet of metal and the difficulty largely disappeared. Try to maintain at least a half an inch of spacing between this ground sheet and your coil. Another modification that was helpful was mounting the coil vertically. This minimized the capacitive coupling between the coil and the sheet metal ground plane. I used clear plastic and epoxy to build brackets and insulate the bottom of the coil.

I tested the coupler by using a few ordinary light bulbs as dummy loads. Bulbs have a wide range of resistance, depending on the wattage and how hot the filament happens to be. The T-Match worked great and the bulbs burned brightly. It worked just as well on my real antennas, so I went on the air and began working folks right and left. So, if you just want a coupler that works, the story is now over. If you like, you may stop reading.

The T-Match as an aid for receiving on 80 and 160 meters

One surprise from this project was that my T-Match has become vital for receiving weak signals on 80 and 160 meter ham bands. These lower bands are quite close to the standard broadcast AM band, (550 – 1700 KHz). As you know, these stations are extremely powerful – as much as 50,000 watts. If you have such a station within 50 miles, or a less powerful station close to your house, there will be big RF voltages riding on your antenna at all times.

With these AM elephants bellowing in my neighborhood, my home-built receiver had trouble filtering them out and hearing the low power ham stations hundreds of miles away. The front end of a receiver can be easily overwhelmed and this problem may not be obvious. I couldn't hear the AM broadcasts in my headphones but the ham bands were mostly static and I didn't know there were any ham signals present. But when my receiver shared the antenna with the transmitter, the signal strength of the weak signals on the low bands peaked dramatically when the T-Match was properly tuned for the transmitter. The T-Match serves as a “high pass” filter that reduces the signal strength of the broadcast stations dramatically. Not only was I suddenly hearing dozens of CW stations on 80 meters, I found myself working QRP stations halfway across the country. I didn’t know that was even practical.

Using the T-Match on receive has another benefit. Since my receiver is designed for a 50 ohm input, when I peak the received signal strength using the T-Match, the 50 ohm transmitter is also (nearly) perfectly matched. So on any band, before I try to load the antenna with my transmitter, I peak the received signals with the T-Match.

Adding the whistles and bells

The interesting part of the T-Match story is what happens when you add the refinements - you know, the metal case, the power meter, and all that. After I had used my primitive coupler for a while, I wanted something more impressive. I already owned a metal enclosure that was about the right size. It took me considerably longer than an hour to build another coupler, but eventually I had a professional-looking T-Match in a metal box. I tried it out and ... it didn’t work worth a darn. It couldn’t match much of anything. It seemed to be a wonderful standing wave generator, but a poor antenna coupler.

I consulted Bob about my problem. He nodded knowingly. “Oh, yeah. It doesn’t work in metal cabinets. That’s why I used wood and plastic.” Now he tells me. “However,” he went on, “you could convert the coupler to a ‘π match.' That should work in a metal box since the capacitors go directly to ground and the accidental capacitance to the metal case will become part of the circuit. Also, you could use a powdered iron core inductor. The magnetic flux is confined to the core, so powdered iron cores work well in small metal enclosures.”

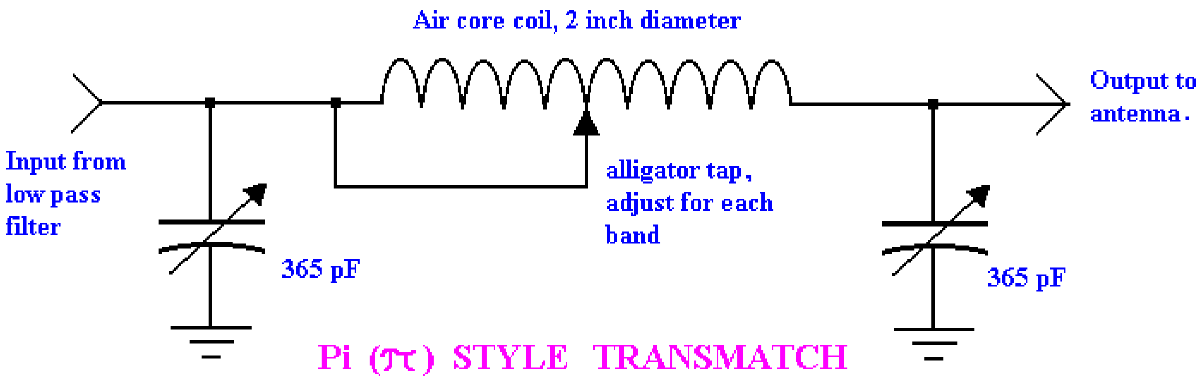
Air core coil, 2 inch diameter
Output to antenna.
Input from low pass filter
alligator tap, adjust for each band
365 pF
365 pF
Pi (π) STYLE TRANSMATCH

The π - Match

I tried out these ideas. Yes, I got the π to work, but I learned that the T-Match was better. If you examine the π-Match circuit, you'll see it is a low pass filter. That is, the high frequency noise gets shunted to ground through the capacitors, while low frequency components pass through the inductor. I was already using a low pass filter to suppress TVI, so the π-Match was redundant. When I used the π-Match I noticed the output waveform usually had low frequency, "roller coaster" distortion and sub-harmonic modulation on it that resembled AM modulation. In contrast, the T-Match serves as a high pass filter that removes these distortions. It turned out that the low pass TVI filter and the T-Match work together as a band pass filter to maintain a pure sinewave.

I found a powdered iron core in my junk box large enough to handle 100 watts and I wound a multiple tap winding on it. Yes, the powdered iron core inductor worked, but it became hot and was clearly inferior to the air core. Since I didn't have any data on the core I was using, I ordered a big new core from CWS Bytemark with known properties. I bought a CWS (Amidon) T200-6. Sure enough, it heated up too. Long live air-core inductors!

My other experiment was built-in power meters. I installed dual meters for forward and reverse power. Cool, huh? Too bad they worked so poorly. The meters were voltage sensitive rather than power sensitive. For example, they read twice as much "power" into a 100 ohm load as into a 50 ohm load. I knew the meters weren't correct because all the other data I had from my oscilloscope, DC power input to the final, etc., all told me that the power into 50 ohms and 100 ohms should have been equal. I discovered I could compensate for this error by placing the power meters on the transmitter side, the 50 ohm side, of the antenna coupler. That way the voltage was constant for the same power level.

I soon retired the fancy coupler to the junk pile and I'm still using the piece of wood with the 1935 capacitor. (Now you know why I have so much junk.) In order to monitor my output power and frequency purity, I monitor the input to the antenna with scope probes going to a frequency counter and to an oscilloscope. The scope and the frequency counter readout are far more sensitive to the problems of a homebrew rig than an SWR meter will ever be.

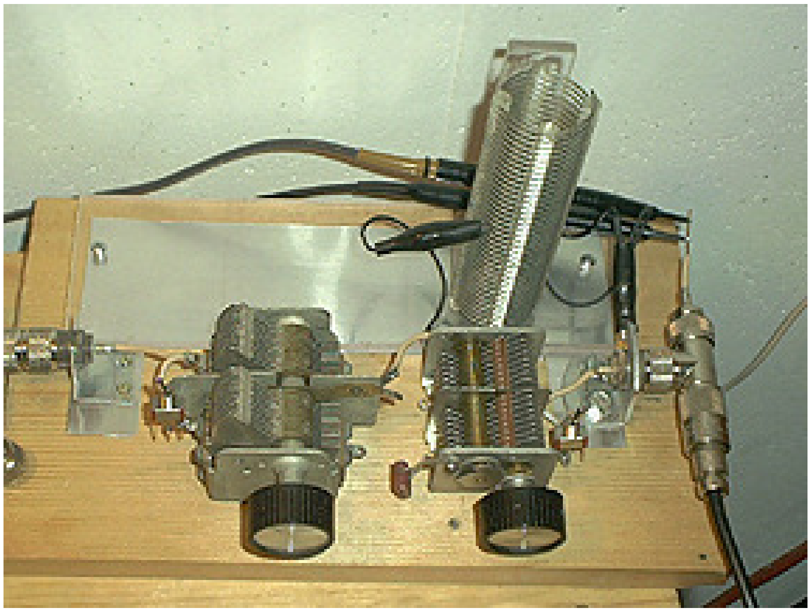

The T-Match monitored by oscilloscope and counter probes. The low pass filter is on the left. The antenna coaxial cable exits on the right. Note the flat sheet of aluminum that serves as a low inductance ground plane. It's covered with plastic to help isolate it from the coil.

****

BUILDING A LOWPASS FILTER

Back in 1997 I carried my old vacuum tube HF transmitters down from the attic and looked at them critically. Hmmm ... Class C output stages. Class C produces distorted sinewaves and distortion causes harmonics. I hadn't been on the air in 30 years, but I knew from studying for my new license that spectral purity standards were higher than they used to be. At the very least, if I used my ancient transmitters, I would be risking TV interference (***TVI***) complaints from the neighbors. Consequently a transmitter lowpass filter became my first project.

I'm told that today hardly anyone with a modern transceiver needs or uses lowpass filters. However, if you do any homebrewing, especially with Class C or Class B output stages, a lowpass is a good idea. When placed directly on the transmitter output, I've found that the filter doesn't reduce power output and it doesn't make the antenna hard to tune, even when operating QRP. I think of my lowpass filter as "insurance" against angry neighbors and pink slips. So why not use one?

Your first question probably is," Do I need one for my little 5 watt QRP?" The answer is, "Probably not." However, as soon as you get complaints, your neighbors will eternally label you as the cause of every bit of snow on their TV screens or virtually any problem with their sets. Once the complaint syndrome begins, being "legal" and "meeting FCC requirements" aren't enough. It is much better to do everything you can to be sure the neighbors never realize that you might interfere with their TVs.

Electrical Design of the Filter

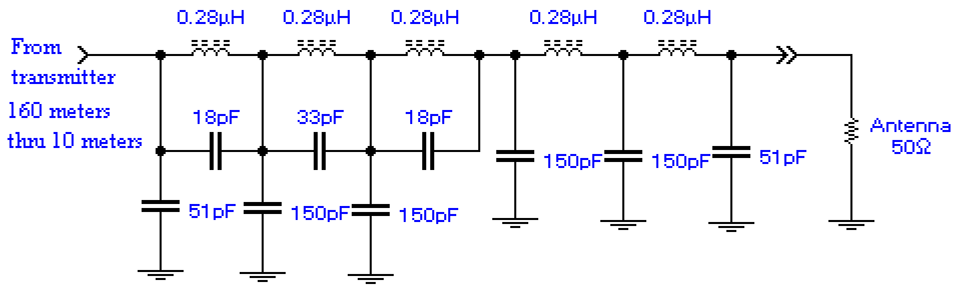

A five-stage lowpass filter

I started with a three stage elliptic lowpass filter from the good-ol' ARRL handbook. Later I acquired a "Spice" program for analyzing circuits, Electronics Workbench. I turned it loose on my lowpass and settled on the circuit shown above. The capacitors to ground tell us this is a lowpass filter. In other words, the capacitors shunt high frequencies to ground that might otherwise radiate in the TV bands, while lower HF frequencies will pass through the filter via the inductors. This filter is designed to pass all the ham HF bands from 10 meters (30 MHz) and below. As you will see shortly, it attenuates all frequencies above 10 meters.

Notice that there are 33 pF and 18 pF capacitors across three of the inductors. Each inductor is part of an LC resonant trap that appears as high series impedances to TV channels 2 and 4. However, the higher channels are also attenuated at least 105 dB. The main lesson I learned from Spice was that component values are surprisingly tolerant. I had used Chebyshev filter tables that have values to 4 decimal points and sophisticated "bell curves" for the different component values. That elegance made me think filters had to be precise. Wrong! Well, that's what Spice and my experience tells me. Notice that I made all the inductors identical and use only two values of capacitors to ground.

Like all complex reactive filters, this one is designed for specific impedances, in this case 50 ohms. That means that the filter only works correctly when the input and output impedances are 50 ohms. For that reason, it's inserted into the antenna line right after the transmitter and just before the antenna coupler. If you are using a power meter designed for 50 ohms, it would go between the low pass and the antenna coupler. The antenna can have any impedance from near zero to hundreds of ohms, so the coupler serves as a transformer to match 50 ohms to whatever is needed.

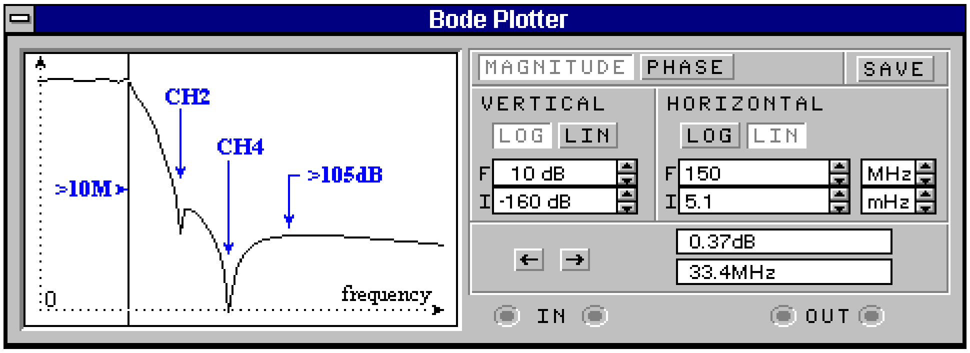

VERTICAL
F 10 dB
I -160 dB

HORIZONTAL
F 150 MHz
I 5.1 mHz

0.37dB
33.4MHz

Lowpass filter Bode plot

Construction Details

The filter is housed in a long, skinny aluminum box. Capacitance to ground is part of the design. Therefore, provided the leakage capacitance between the coils and the metal box isn't extreme, this stray capacitance won't degrade the performance. Just keep the coils at least 1/2 inch away from the grounded metal walls and the filter will work well. I used air core inductors, which meant that I had to rivet four pairs of overlapping metal partitions into my box to prevent coupling between the five coils. Without the metal shields, air core inductors would couple to each other like transformer windings. I prefer air core because simple coils don't dissipate significant energy. I'm also cheap.

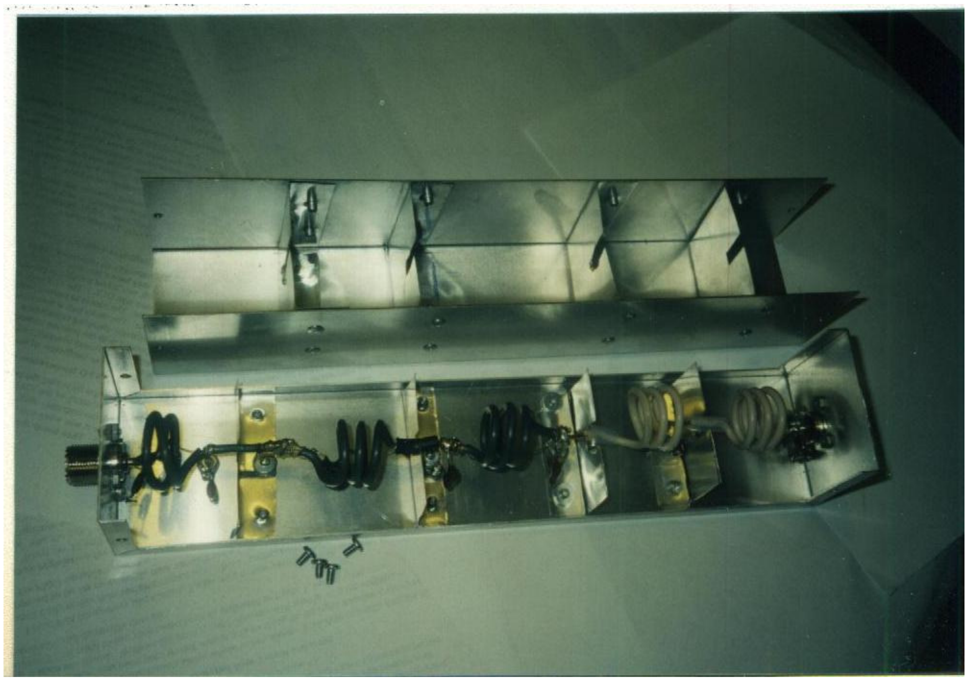

Inside view of the filter. Each coil is isolated from its neighbors by notched partitions.

I wound my coils out of 12 gauge insulated copper wire stripped from Romex household three-conductor wiring cable. If you follow the dimensions in Table 1 exactly, you should get about the same resonant frequencies I did. Depending on the shape of your box, you may use either large coils with 2 turns or smaller diameter coils with 3 turns. The inductance will be about the same. The important issues are that the three LC traps should resonate well above the 10 meter band and there should be no significant attenuation below 30 MHz.

Table 1. **Air core inductor specifications**.

	turns	**diameter**	**length**	**inductance**
	2 turns	1.25 inches	0.5 inch	0.28 microhenries
or,	3 turns	1 inch	0.75 inch	0.28 microhenries.

How to measure air core inductance

How did I know when I had 0.28 microhenries? I calculated what the resonant frequency should be for 0.28 microhenries and a convenient fixed capacitor. I wound a coil, put the capacitor in parallel, and put the LC circuit across a frequency generator output. Using an oscilloscope to watch the voltage, I adjusted the frequency to find the resonant peak and then rewound the coil until it peaked at the calculated frequency.

The inductors could also be wound on powdered iron cores. An advantage of powdered iron is that the magnetic flux is confined in the cores and you can cram the filter into a smaller metal box without metal shields between filter sections. However, powdered iron core inductors will still have stray capacitance to ground.

All capacitors should be mica and capable of handling voltages consistent with your power level. If you're running more than 100 watts, it would be prudent to use capacitors with working voltages of 1,000 volts or more. I used 500 volt caps with my 100 watts and so far haven't had any capacitor failures. Two identical 500 volt caps in series will give a working voltage approaching 1,000 volts, but of course the effective capacitance will be half the value of each. Use real RF connectors on your filter, UHF SO-239 or equivalent.

Testing the filter

I tested my filter by terminating it with a 50 ohm load and then running an RF signal generator through it. There was no significant attenuation or distortion of the sinewave from 160 to 10 meters. Above 10 meters the output signal dropped to almost nothing. Similarly, when I loaded lightbulbs and 50 ohm dummy loads on all HF bands, 80 through 10 meters, I could see no differences with or without the filter.

So far, my only TVI complaints have been from my own family. On 15 meters, they see minor flickering on the picture. Not surprisingly, 17 Meters bothers channel 4 (4 x 18 MHz = 72 MHz) and 30 meters tears up channel 6 (8 x 10.1 MHz = 80.8 MHz.) I believe my biggest remaining TVI problem is my open-chassis transmitter with the PC boards exposed to the breezes. Obviously I need a cabinet.

When we bought a modern TV, all my interference problems vanished. Fortunately, digital TV and the increasing use of cable TV, Direct TV satellites, and 2.4 gigahertz digital cell phones, makes annoying the neighbors less and less likely. In February 2009 in the United States UHF digital television replaced all analog channels 2 through 13. In other words, the need for a low pass is rapidly passing. Considering that we are living in "The Wireless Age," the future of our TVI problems is surprisingly bright.

****

STAYING LEGAL WITH HOMEBREW TRANSMITTERS

On Field Day I watched while up-to-date hams tuned up their rigs. I was amazed by how casually they assumed their signals were matched to the antennas and not suffering from rough CW tone, harmonics, or drifting frequency. They plugged beam antennas and dipoles directly into sophisticated transceivers and were instantly in business. I didn't see any antenna couplers, power meters, low pass filters, dummy loads, oscilloscopes, frequency counters, or any of the tools we homebuilders use to stay legal.

Investigate before you radiate

Many hams these days are building QRP transmitters. Before they go on the air with a homebrew transmitter, they should learn how to check it out. Out there on the bands are Official Observers (OOs) who send out little white SWL (Short Wave Listener) cards. Ten years ago when I got back on the air with my homebrew, vacuum tube antique, I quickly got an OO card from New Jersey. OO cards are *NOT* suitable for framing. The card informed me that my CW signal was drifting and had an obnoxious tone. Very embarrassing! Don't just go on the air and hope everything is OK. ***Investigate before you radiate!***

Listen to your own signal

A good way to find out how your signal would sound to another station is to load a dummy load then listen to it with your receiver. I routinely use an old Collins receiver for this purpose. The tricky part is desensitizing the receiver so that the signal strength simulates a signal on the air. I did this by shorting a coax connector and plugging it into the antenna jack on the back of the receiver. I also clamped ferrite blocks on the line cord and speaker wires to help keep RF out of the receiver cabinet. After I made these changes, the signal was sufficiently weak that I was able to hear key-clicks, rough note, or "warble" instabilities. Now I could understand what my contacts were complaining about.

My 1967 homebrew transmitter was a mobile unit with a switching power supply. No matter how I filtered the supply output, I couldn't get rid of the switching hash. The hash was extremely hard to see on the oscilloscope, but in the receiver, I could hear a distinctly rough note. No wonder I kept getting those 598 reports!

Oscilloscopes

In my opinion, an HF transmitter would be hard to build without a quality RF oscilloscope. Looking at the waveforms on a scope makes it easy to tune up a stage, or at least to get the tuning approximately right. When I operate, I keep a 10:1 scope probe right on the antenna feedline. That way, there's no question about what's being transmitted. A typical probe is rated for several hundred volts and has one megohm load resistance and five picofarads capacitance. This shouldn't bother your tuning and your transmatch can compensate for any slight mismatch.

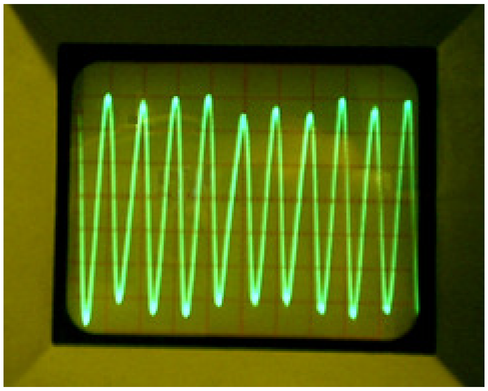

A well-tuned 30 meter CW signal on the antenna lead

Usually on the lowest HF bands, like 80 and 40, you'll see a truly perfect sinewave on your antenna. But the higher the frequency, the less perfect the waveform will probably be. The 30 meter signal on the left is sharply focused, but it has a bit of low frequency modulation. This is quite all right.

Beware of blurs

On the other hand, the 30 meter signal shown below is mistuned. Notice how only the first half sinewave is synchronized. After that, the waves are a blur of different, overlapping frequencies. Also, the reading on your frequency counter will be low and out of the desired band. Usually tuning the transmatch instantly corrects this.

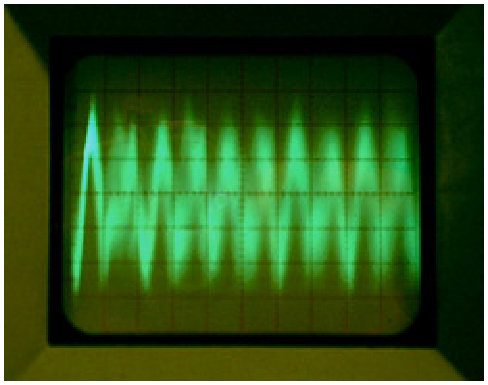

A poorly tuned 30 meter CW signal

Look for the zero crossings

At even higher frequencies, like 15 and 10 meters, your sinewave may have more low frequency modulation and be slightly blurry as illustrated below. The higher the frequency, the more difficult purity becomes. Also, above 20 meters it becomes increasingly hard to know what's artifact from the oscilloscope and probe and what's actually going out the antenna. ***A reliable indication that all is well is when the picture focuses sharply enough to see the zero crossings clearly.***

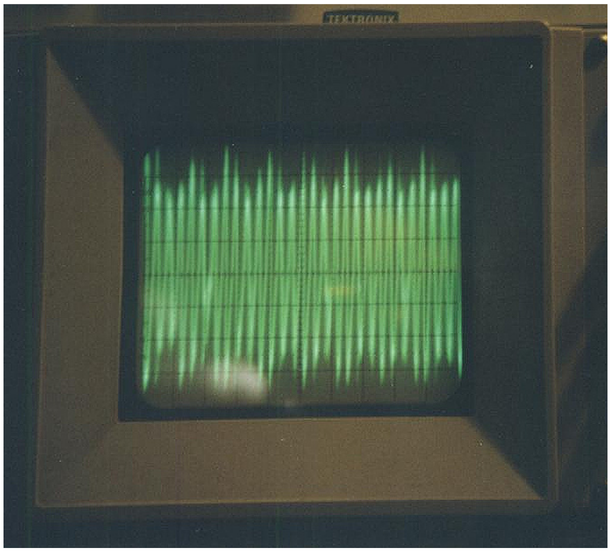

An adequate signal on 10 meters

In contrast, the 10 meter signal shown below is too blurry. Tune your transmatch!

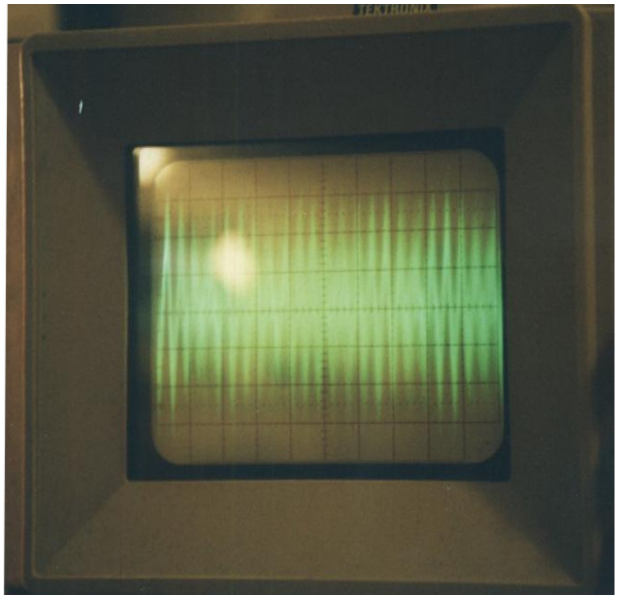

Tuning for maximum power out

I tune my transmatch for maximum amplitude of a sharply focused sinewave at my antenna lead. This gives me an approximate measurement of antenna impedance. For example, suppose I first I load my 50 ohm dummy load on 80 meters and get about 50 watts. In other words, the scope shows a 70 volt peak sinewave. Next I try to load my 40 meter dipole on 80 meters. Using the transmatch, it loads up and produces a sharp 300 volt peak sinewave. Unfortunately, the 300 volts peak on the coax with a 50-watt signal suggests really high antenna impedance, about 900 ohms. If the antenna looked like 50 ohms, I would see about 70 volts peak, just like the dummy load. In my experience, with a match that poor, it may be safe for the transmitter, but nobody will hear me. Let's look at the calculations:

Measuring transmitter power with an oscilloscope

To measure the power output of your transmitter into an antenna or a dummy load, you can look at the RF voltage across the load with an oscilloscope. RMS power is the heating value of the electric current. That is, one watt RMS equals one joule of energy per second. Hams usually use RMS power instead of peak power or other possibilities. To find the power we first need to measure the RMS voltage of a sinewave. You can measure the peak voltage by counting the grid squares on the screen. Then, to get the RMS voltage, you divide the peak voltage by the square root of two. For example, suppose the RF sinewave on the screen is 100 volts peak:

100 volts Peak $/ \sqrt{2}$ = 100 volts / 1.414 = 100 volts x 0.707 = 70.7 volts RMS .

Rather than calculating the square root of 2 or its inverse every time, you can proceed more quickly by remembering the factors 1.414 and 0.707. For example, suppose the RF voltage seen on the oscilloscope screen is 100 volts peak, then the RF power delivered to a 50 ohm antenna or dummy load resistor would be:

$$\text{Power} = I_{RMS} \times V_{RMS} = (V_{RMS} / \text{Resistance}) \times V_{RMS} = (V_{RMS})^2 / \text{Resistance}.$$

In this case, Power output = $(70.7 \text{ volts RMS})^2 / 50$ ohms = 100 watts.

I remember my electrical engineering professor turning purple in the face when a student used the V^2/R formula for power during class. “You must be a ham! You guys are always using that formula, but we never teach it because it assumes that the current and voltage in phase.” He was absolutely right on both accounts. That is why the power is first measured on a resistive dummy load. Dummy loads have no capacitive or inductive components to make the current and voltage components out of phase. Then, logically, if the transmitter is drawing the same current from the 12 volt battery, the power input must be the same. Therefore we can assume that the antenna is receiving the same power as the dummy load did we are justified in calculating the power to the antenna.

Peak Envelope Power (PEP)

The U.S. government regulations for ham radio operators limit the power we can use to a specified amount of "***Peak Envelope Power***." This varies with different hambands and class of license. As examples, on 30 meters we are only allowed to use 200 watts PEP, but on 20 meters, 1,500 watts PEP is permitted.

PEP is defined as "the average power supplied to the antenna transmission line by a transmitter during one RF cycle at the crest of the modulation envelope taken under normal operating conditions." In other words, it is the peak RMS power for the highest part of a modulated waveform. So, if you were operating with single sideband phone, the measurement with an oscilloscope would measure the highest peak voltage seen on the scope multiplied by 0.707. Then, just as we did above, we would square this RMS voltage value and divide by the impedance of the antenna, e.g. 50 ohms. A CW transmitter output is measured the same way.

An oscilloscope for HF ham work should be rated for at least 50 MHz. A brand new oscilloscope like this will cost at least $2,000, but is not a good value. My Tektronics 5441 scope originally sold for $11,000 in 1976. Today that scope or equivalent ones can be found used in electronic surplus stores or mail-order catalogs for $400 or less.

An FM radio will detect serious frequency impurity

A serious splatter problem (noise mode) will make a roar of static in your FM radio. On the other hand, sometimes the FM radio just cuts out and goes silent. That may not be a transmitter problem. It could be that the radio’s IF is just overwhelmed by your signal. Alternatively a harmonic of your transmitter frequency might happen to line up with your FM station and obliterate it. Again, the FM radio just goes silent.

DC power supply ampere meter

It is always nice to know how much current the transmitter is drawing. Linear amplifier stages run at roughly 50% efficiency, while the transmitter circuitry as a whole will be somewhat less than that, say 40%. DC current is another indication of the SWR (Standing Wave Ratio). That is, SWR describes how well your antenna is tuned. When operating correctly, 50 watts output should draw about 8 to 10 amperes at 12 volts. If it reads 15 or 20 amperes, it means you are way out of tune and the final amplifier is heating rapidly.

Frequency counters

The most persistent problem a homebrewer faces is frequency drift. Commercial transceivers use frequency synthesizer chips slaved to unusually stable crystal oscillators. They also have built-in displays that read frequency to a fraction of a Hertz. This amount of precision is almost never justified by the specifications. But hey! The guy you’re talking to hasn’t read the specs on his transceiver and he believes his readout is gospel. Since we

A typical HF frequency counter

When I operate, I have two scope probes on the antenna line. One goes to the scope and the other goes to a frequency counter. A good counter is vital because most of the frequencies you must measure are too close to the edge of the ham band to simply count squares on the oscilloscope screen. When the transmitter is operating properly, the counter display is solid to within 10 Hz and doesn't dance around. The same conditions that cause blurry sinewaves on the scope cause the counter to read low and jump about. For example, if you are loading up on 15 meters and the counter reads something like "20.6XXXX" with the last few digits changing every second, you are mistuned. Don't be satisfied until the counter reads what it should and stays there. That is, you should see a relatively solid reading of a legal frequency, like "21.120XX." Counters can be a bargain if you buy them used. My used Hewlett-Packard counter originally sold for about $2000, but I paid $60 for it.

Lowpass filters

When operating homebrew equipment it's a safe bet you will occasionally generate out-of-band harmonics, especially while loading your antenna. A lowpass filter is simple insurance against generating interference above some design frequency. Another way I minimize the out-of-band noise problem is to keep a chart of the transmatch settings for each band tacked up on the wall. So when I switch bands, the transmatch is already fairly well adjusted before I key the transmitter.

Not so useful instruments

Power meters I have not found power meters particularly useful. I have a commercial one that gives readings that correspond well with the light produced by a light bulb dummy load. My homemade power meters aren't that good. The difficulty is that power meters only tell you the total power delivered to the antenna. I need to know much more than that to avoid bad signal reports and OO cards.

Grid dip meters In the old days we used "grid dip meters" for frequency measurement. A grid dip meter measures the frequency of a resonating coil by placing a secondary coil next to the target coil. RF currents induced into the coil produce a drop or "dip" in the meter reading when the tuning knob on the meter is tuned to the resonating frequency. The approximate frequency is then read off the dial. Grid dip meters aren't at all accurate, but they got us onto the right ham band. Today's inductors are usually wound on powdered iron toroids. Grid dip meters don't work well on toroids because the magnetic field is trapped inside the closed iron loop. There is practically no leakage outside the toroid for the grid dip meter to tap into.

Spectrum analyzers A spectrum analyzer would be great for homebuilders. They measure the purity of transmitter signals by graphing all the frequency components. Pictures of analyzer displays are often shown in homebrew transmitter articles. Unfortunately, even a used spectrum analyzer with sufficient precision is pricey - many thousands. Since I can't afford one, they don't do me a bit of good. Without this tool, I have to "interpret" what I see on the oscilloscope and the frequency counter. Fortunately, that isn't hard.

****

ANTENNA RELAYS

Switching from transmit to receive

If you simultaneously connect both your receiver and your transmitter to the antenna, you may burn up the front-end stages of your receiver when you transmit. When you first go on the air with your QRP transmitter, you will no doubt be irritated by the need to use a separate antenna for the receiver. A clumsy solution would be to rig up a manual switch to move the antenna back and forth from transmitter to receiver every time you stop sending. That's inefficient, to say the least. If you have to throw more than one switch to go between send and receive, you will be at a huge disadvantage when trying to work DX or in contests. Actually, even one switch is not up to modern standards.

Modern transceivers have "break-in keying." When they stop sending, the receiver instantly comes back on automatically. Non-homebuilders aren't even aware that switching antennas is a problem. I haven't yet mastered building break-in keying and still use one switch to go from send to receive. Even so, by the time my receiver comes back up after transmitting, I usually only hear the last three or four characters of my call sign from my contact, "…IYE". The other guy has already switched over and sent "KØ" before my receiver was back on the air. As you can see, one switch operation is the bare minimum.

Requirements for an antenna relay

An antenna relay is usually a ***single-pole, double-throw*** switch, sometimes called ***SPDT***. The external antenna line is connected to the moving contact. This contact is normally resting against the receiver contact. When the relay is activated, the moving contact switches over to the transmitter contact.

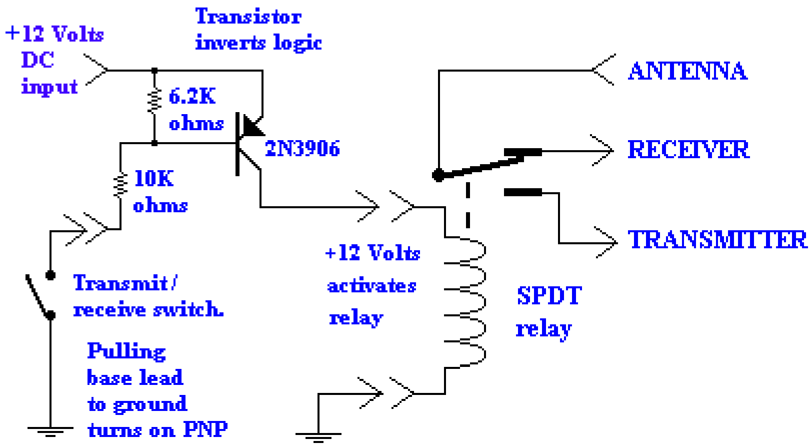
+12 Volts
DC
input
Transistor
inverts logic
6.2K
ohms
2N3906
10K
ohms
Transmit /
receive switch.
Pulling
base lead
to ground
turns on PNP
+12 Volts
activates
relay
SPDT
relay
ANTENNA
RECEIVER
TRANSMITTER

Schematic of antenna relay The optional transistor inverter turns on the relay with a high impedance line switched to ground.

Unfortunately, ordinary relays have too much inductance on 10 meters. Inside the relay RF current must travel through a length of wire perhaps one or two inches long. This wire is not a coaxial transmission line or a wide, low inductance strip. The result is that ordinary relays often work poorly on 10, 12, and 15 meters. By “poor” I mean that no matter how you adjust the antenna coupler, you can’t quite deliver a sharply focused sinewave to the antenna. (See the above article on checking out homebrew transmitters.) In fact, the final amplifier may remain in “noise mode” and not produce a sinewave at all. You can demonstrate that the problem is the relay by bypassing it and connecting the final amplifier directly to the lowpass or to the antenna coupler. Once the relay is out of the circuit, you will often find that the problem is cured and the antenna coupler will load the antenna just fine.

There are, of course, commercial coaxial cable relays to solve this problem. I built a usable antenna relay out of a *TINY* 10 amp relay rated at 120 volts AC (Radio Shack Part # 275-248A). The important specification is the “tiny” part. Since the relay is miniature, the wires inside the relay are very short. I kept the external transmitter RF leads as short as possible by implementing the "wires" with little, low-impedance rectangles of PC board. The connections only have to travel about 5/8 inch from the center conductors of the SO-239 RF connectors to the relay pins.

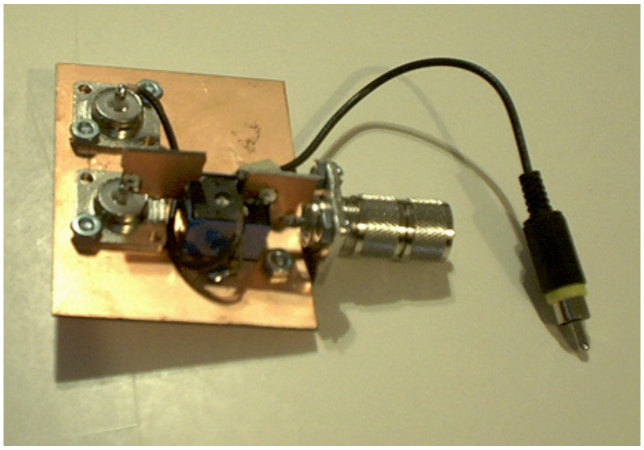

Homebuilt antenna relay

The receiver antenna connection is not nearly so critical. Therefore it was implemented with a three-inch length of RG-174 coaxial cable. The cable shield is only grounded at one end. These precautions reduced the inductance of the relay leads sufficiently. Now when I load my 10 meter vertical antenna, it works as well with or without the relay in the pathway. The relay coil was energized with another short length of RG-174 going to the phono plug on the right. I shielded the 12 volt signal wire because I was trying to keep RF out of my power supply. This goal was easier for me using the logic inverter to drive the relay coil. Don't bother building the inverter if you don't need it.

Turning on the transmitter power supply

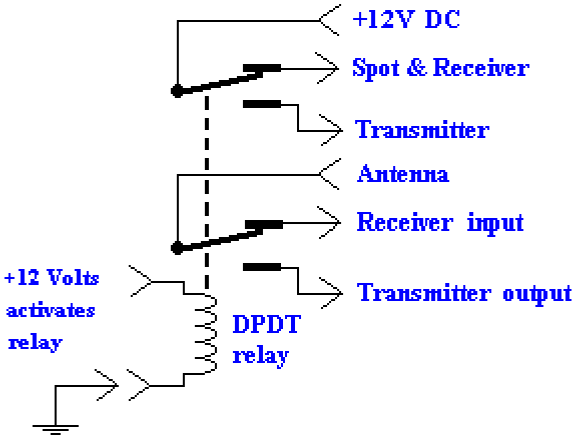

In many homebuilt transmitter designs the antenna relay has another set of contacts to turn on the DC power supply for the transmitter. A VFO-controlled transmitter is much more complex than a crystal controlled QRP. There are usually one or more oscillators running during transmit that must be turned off during receive. Otherwise you will hear them as whistles in the receiver. The "double" relay above in older ham transmitters is usually a ***Double Pole, Double Throw***, a ***DPDT***. The DC supply line goes to a moving contact.

At rest this contact energizes the "spot" circuit and perhaps also the receiver. The "spot switch" allows you to turn on the transmitter crystal oscillator or VFO to figure out where your signal will be in relation to the fellow you're listening to.

Grounding one side of the relay coil activates both the antenna relay and the DC supply relay. In other words, this control line always has 12 volts on it. When the line is grounded by means of the send-receive switch, this line activates the transmitter. This same signal can also be used to "mute" or inactivate the receiver. Alternatively, the DC supply relay in the transmitter could turn the power to the receiver off and transfer it to the transmitter.

I have built three transmitters that used a DPDT relay to perform both the antenna switching and the DC power switching. Except for the 10 meter problem explained earlier, these big relays all worked well ... at first. But eventually, the DC power side always became intermittent. For this reason, if you are going to run a 50 or 100 watt transmitter, plan on switching at least 20 amperes. I recommend using a 30 ampere relay and maybe yours won't become intermittent. Of course, the bigger the relay, the more difficult it will be to use one section of the relay as an antenna relay. I finally gave up and used separate relays for the power and antenna. When I push the little send / receive toggle switch on my electronic key "bug" box, it activates the coils of both relays.

Avoid power relays

Yet another problem with relays is that, the larger the relay, the more DC current it takes to activate it. A big 20 or 30 ampere relay coil can draw 100 to 200 milliamperes of current just to turn it on. Better yet, don't use a DC power relay. Using the QRP keyer described in Chapter 6 as a model to follow, you may use P-channel MOSFET power transistors to turn on the transmitter and/ or receiver. A heavy duty P-channel MOSFET power switch is illustrated below.

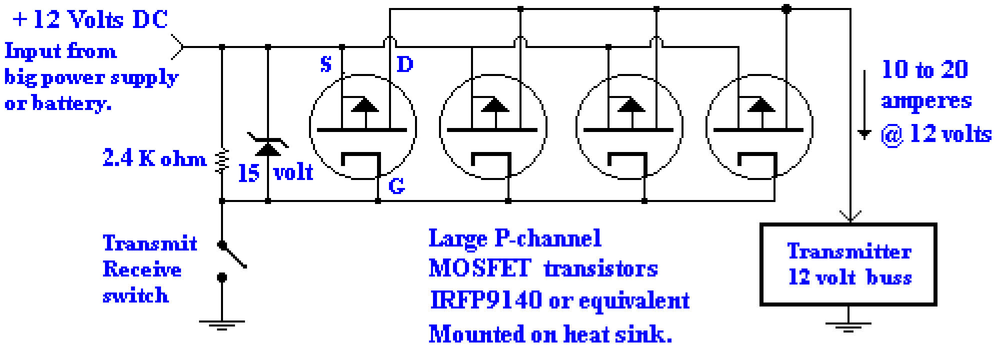

TRANSMITTER POWER SWITCH

The MOSFET power switch above uses four P-channel MOSFETs in parallel. When they are turned on, MOSFETs look like low resistances, 0.15 ohms or less. This means that little voltage is wasted across the switch. The larger the MOSFET, the lower the resistance. By putting several in parallel, the resistance can be lower still. Large TO-204 package MOSFETs can be bolted to a heat sink to keep the transistor temperature as low as possible. Otherwise, as the temperature rises, the internal resistance of the transistors will also rise.

P-channel MOSFETs are the easiest to use because you can turn them on by pulling their gates down to ground. In my first transmitter I used two big TO-204 case MOSFETs. Their heating was negligible when bolted to a large heat sink. In my second transmitter I used four smaller TO-220 case MOSFETS, type RFP30P05 p-channel MOSFETS. (Notice that you can decode the type number: 30 = 30 amperes, P = p-channel and 05 stands for 50 volts.)

Actually, for a given size and voltage rating of a transistor, N-channels are superior to P-channels. N-channels usually have about 1/3 of the on-resistance and are more difficult to damage. Unfortunately, to turn on N-channel MOSFETs, you would need a 24 volt power supply to pull the gates ***up*** to about 12 volts above 12 volts. However, if you have built a VFO like the one described in chapter 10, you already have a 22 volt power source you can use.

The 2.4 K ohm gate-to-source resistor insures that the transistors turn off when the switch opens. The 15 volt Zener diode on the gates is advisable because, in my transmitter, the gates of the transistor are in parallel with the antenna relay coil. When the transmit/ receive switch is opened, a big voltage can appear across the coil and may damage the transistor gates.

* * *

HOMEMADE QSL CARDS

I do not miss the QSL cards from the old days. Fifty years ago the most common American QSL cards were bought from World Radio Labs. They had a map of the US with one's call letters printed in red. It seemed as though everyone who didn't send out World Radio cards was sending Allied Radio cards. Allied cards were just glossy white cards with big orange call letters and a couple of orange stripes. These designs were quite attractive, but since they were the vast majority of what you received, QSL cards were pretty monotonous. Back then, if you had bought QSL cards that would have been competitive with today's commercial cards, they would have cost a fortune. Hand-drawing my own cards was fun, but only for the first 2 or 3 cards. After that, it was far too time consuming to be practical. I managed to make a crude template and print some rough black and white cards that were barely acceptable, but they sure looked primitive.

Since the age of Xerox, computers, digital cameras, and the color printers, making your own color QSLs has become a breeze. Other than knowing how to run a computer, it is totally non-technical. I dare say that typical 12 year olds can make great cards, even if their parents can't. Considering how easy it is, I'm disappointed that so few guys make their own cards. These cards have much more "soul" than mass-produced commercial cards.

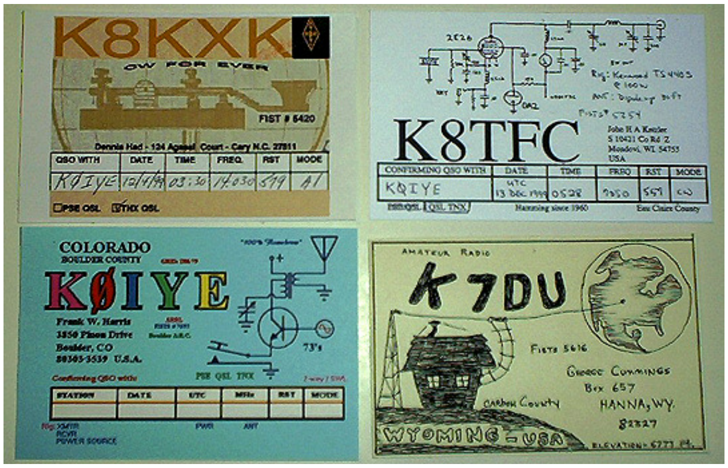

Here are a few examples of homebrew cards. The two on the left were made using Microsoft Paint, Mac Draw or a similar, standard drawing program. The ones on the right were made using hand drawings that were later Xeroxed. They were printed onto plain 5 by 8 inch index cards and then trimmed to postcard size. Not much to it, really!

My buddy Jack, KØHEH, used his digital camera to take a picture of a nearby mountain. Then he used a free QSL software program to overlay his call letters on the scenery and ended up with a beautiful card. The cards can be printed on a color printer using photo quality paper. The result is very classy.

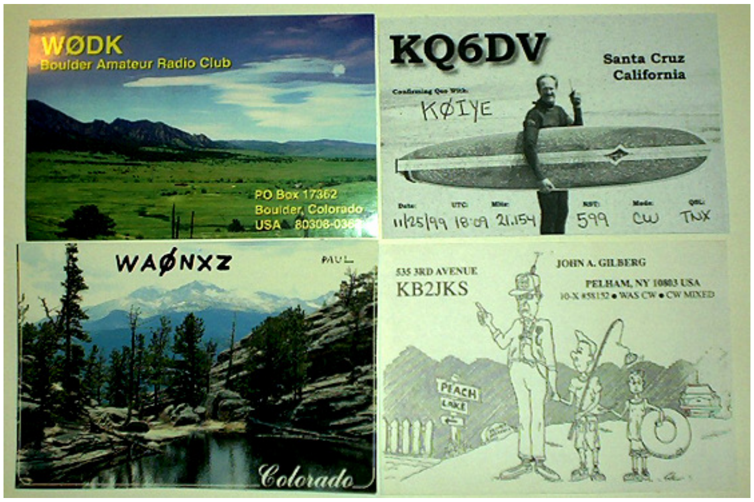
WØDK
Boulder Amateur Radio Club
PO Box 17362
Boulder, Colorado
USA 80308-036
KQ6DV
Santa Cruz
California
KØIYE
11/25/99 18:09 21.154 599 CW TNX
WAØNXZ
PAUL
Colorado
535 3RD AVENUE
KB2JKS
JOHN A. GILBERG
PELHAM, NY 10803 USA
PEACH LAKE

Four more examples are shown above. The Boulder Amateur Radio Club card at the upper left, WØDK, could be made with a digital camera as was just described. Alternatively, one could begin with a regular color photo and a scanner. The photo card on the lower left, is from Paul, WAØNXZ. This card could have been done with a scanner or digital camera, but this particular card happens to be a commercial postcard with hand-lettered call sign at the top. It's a beautiful card but buying postcards is pretty expensive.

The guy with the surfboard holding the walkie-talkie at the upper right, is Tom, KQ6DV. He used his own photo but kept the cost down by not using color and photo quality paper. The card at the lower right, is from John, KB2JKS. He made a complex hand-drawing and then Xeroxed it onto cards.

There are loads of ways to make QSL cards today. And all of these cards are more interesting than commercial ones. Yes, printing cards a few at a time is expensive. But most of us don't send out zillions of cards anyway. Making them yourself lets you modify them as often as you like. You aren't stuck with some error on the card for the next two hundred copies. Even if it's just QSL cards, homebrew is more fun!

CRYSTAL SETS TO SIDEBAND
© Frank W. Harris 2010, REV 12

Chapter 10

Variable Frequency Oscillators (VFOs)

Sooner or later you'll become frustrated with being stuck on a few crystal-controlled frequencies. You would like to have a frequency-tuning knob that covers the entire band and not just a few kilohertz. This sounds simple, but isn't. It's difficult because, without the stability of a crystal, an ordinary RF oscillator will drift hundreds of hertz while you're sending. The fellow you're talking to probably has a modern transceiver with a narrow, stable passband. From his perspective your signal quickly drifts out of his passband. His digital readout may be calibrated to tenths of a Hertz and he will take great delight in telling you about *YOUR PROBLEM.*

Drift is a big deal today

In the old days, like 1950, receiver passbands were usually huge, like 10 or 20 kilohertz. So you could drift quite a distance before your contact even noticed, let alone could no longer hear you. Besides, everyone drifted a little back then, so it wasn't worth mentioning. In the really old days, like 1930, the signals drifted so much, that hams often tuned their receivers with one hand while they copied down the Morse code with the other.

A stable variable frequency oscillator can replace a crystal oscillator. This chapter summarizes what I learned in my odyssey through six VFO prototypes. My early VFOs drifted hundreds of Hz and I got loads of complaints. After I added temperature compensation, I got the instability down to 20 Hz drift per minute. Some stations notice 20 Hz drift and a few even told me about it. Eventually I built a super-regulated power supply for the VFO and got the drift down to about 5 Hz. I suggest you avoid embarrassment and work directly toward 5 Hz. According to the ARRL handbook, +/- 5 Hz is about as stable as you can achieve without phase lock loop technology.

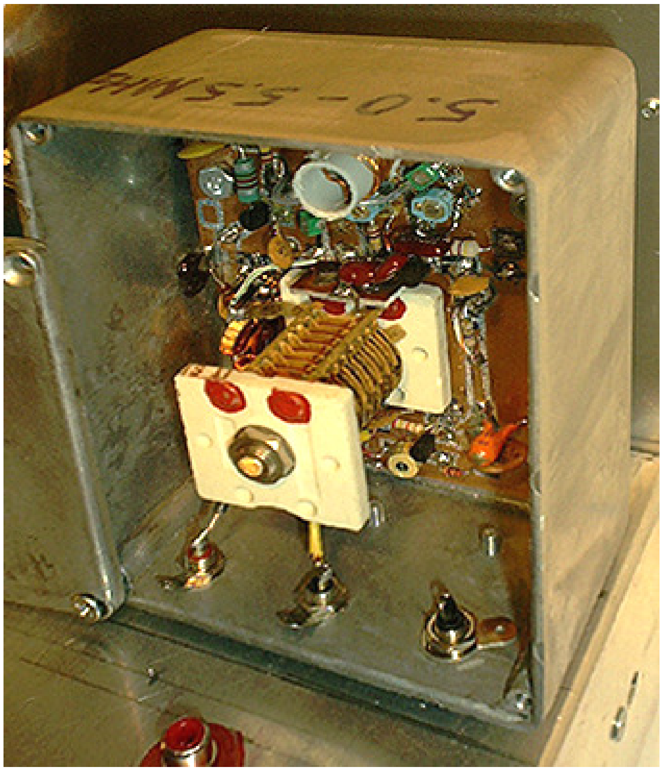

A 5 MHz VFO tuned by a mechanical variable capacitor

I was recently talking to a brand new local ham on 40 meters CW. We were chatting at a leisurely 5 words per minute. He, of course, was using a wondrous modern transceiver. It took us an hour to communicate very little information. Finally he tired of the strain and called me on the phone. "By the way," he said, "were you aware that you drifted over 50 Hz during our QSO?" He thought he was informing me of a grievous defect in my rig. Less than 1 Hz per minute? I was delighted that the drift was so low.

The test equipment needed to build a VFO are a precision multimeter for measuring voltage down to millivolts and a frequency standard. A super-accurate, modern receiver is OK, but a frequency counter is better for this application. The hardest part about building a stable VFO is following all the detailed instructions on how to do it. If you're like me, you'll have trouble believing that all that trivia is really necessary. Yes, you can cut a few corners, but the more compromises you make, the more your VFO will drift.

Low frequency VFOs drift less than high frequency VFOs

It's practical to build a VFO that operates directly on 160 meters (1.8 to 2.0 MHz) or on 80 meters (3.5 to 4.0 MHz). This makes transmitter design simple. For these bands you just amplify the VFO signal directly. For example, if you had a crystal-controlled QRP designed for 80 meters, an 80 meter VFO could be plugged directly into the crystal socket. You might want to attenuate the signal somewhat before directly substituting a crystal, but the VFO can replace the crystal and allow you to tune all over the band. If you can build a really first rate VFO for 40 meters, that band would also be practical. Unfortunately, for a given level of sophistication and precision, frequency drift is directly proportional to frequency. You will probably find that somewhere above 5 MHz, there is too much instability to keep the drift under 5 Hz per minute. In order to build a low drift VFO for 40 meters or above, a low frequency VFO can be "converted" up to the desired high frequency. Conversion means adding the VFO to a high frequency crystal oscillator signal and then extracting the sum frequency with a filter. Frequency conversion is covered in chapter 11.

Considering the size of most HF ham bands, a VFO needs a tuning range of at least 0.5 MHz. The higher the basic frequency of the VFO, the wider the tuning range you can achieve. ARRL handbook VFO projects have various frequencies ranging from 1.75 to 9 MHz. My CW transmitter VFO ranges from 3.5 to 4.0 MHz. My receiver and SSB-transmitter VFOs tune from 5.0 to 5.5 MHz. In retrospect, if I were starting over, I would have built the 5 MHz VFO first, since that frequency turns out to be more versatile. The disadvantage of a 5 MHz VFO is that it can't be used on any hamband directly and every hamband needs a frequency converter circuit. (Yes, a 5 MHz VFO can reach 60 meters. But CW isn't permitted on 60 and generating SSB on the same frequency on which you're transmitting requires a phase shift technique that is rarely used today.)

JFET transistors

Junction Field Effect Transistors (JFETs) are ideal for building VFOs. Unlike bipolar transistors, the main current from the drain to the source does not pass through any PN junctions. PN junctions change their characteristics with temperature. Therefore, VFOs made from bipolar transistors tend to drift more than JFETs. JFETs work on the same principle as a MOSFET, but the control gate is a P-N junction diode rather than a capacitor. JFETs were explained and used in the VFO oscillator in the direct conversion receiver in chapter 7.

THE VFO CIRCUIT

The basic oscillator

In principle the VFO is almost the same as a quartz crystal oscillator. The crystal is electrically equivalent to an LC resonant circuit. Therefore, to tune a VFO we use either a variable capacitor or a variable inductor to change the resonant frequency. So far as I know, suitable variable inductors can't be bought or built in a home workshop. That means that the tuning element will have to be a variable capacitor, much like the one you used to adjust the crystal frequency in your QRP.

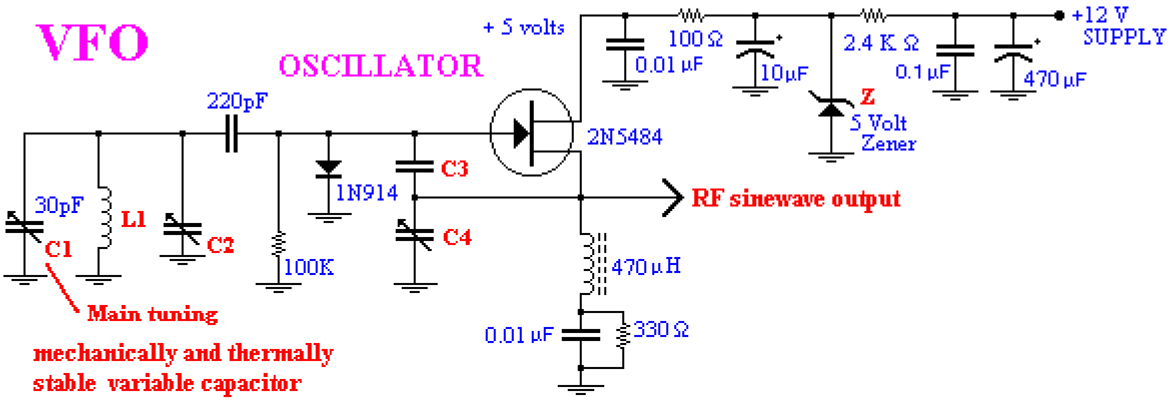

The circuit above is essentially what you will find in your ARRL handbook. It uses a JFET transistor. The oscillator is a ***Colpitts*** and can be recognized by the capacitive divider feedback, C3 and C4. Whenever the source voltage goes up, some of this change is coupled to the gate through C3. This turns the JFET more ON. That is, the feedback is positive which sustains the oscillation. The basic LC resonant circuit that determines the frequency is C1 and L1. C2 is a trimmer capacitor to help adjust the desired tuning range.

So what are the values of C1, L1, C2, etc.? The answer isn't simple. We start with a quality variable capacitor for C1 as described below. For various reasons, **C1 will probably be about 30 pF.** Starting with this capacitor and the need for a 0.5 MHz tuning range, the other values must be determined by trial and error. The values are extremely hard to calculate because C3, C4 and even the 1N914 diode are part of the capacitance. Don't try too hard to get the values right until you have studied this entire chapter and formulated a plan for your VFO. As you will see later, C2, C3, and C4 are part of the temperature compensation strategy and will need to be determined by those considerations first. When you have decided what to do about C3 and C4, then you can work on L1 and C2.

The 470 µHenry choke in the source circuit serves as the load "resistor." That is, the oscillator sinewave output is locked in between the operating voltage, +5 volts, and ground. Since the sinewave voltage appears across the transistor, the remainder, 5 volts minus the sinewave, must appear across a load of some sort. This choke could be functionally replaced with a resistor, except that a resistor dissipates energy and gets hot. The last thing you need in your VFO is another source of heat. The choke inductance is high enough so that the current through it changes very little during one RF cycle and its effect on the frequency of oscillation is minimal and unchanging.

While we're on the subject, you might experiment with substituting an LM336-2.5 volt precision Zener diode for the 5 volt diode. This would of course cut the amplitude of the oscillation in half, but it will also decrease the circuit heating to one fourth. It may not work out for you, but anything that reduces heating, increases stability.

The 1N914 diode on the gate seems counter-productive, but it is used as a clamp to keep the base of the JFET transistor out of the forward conduction region. That is, when the oscillation gets too vigorous and the gate P-N junction begins to turn on at 0.6 volts, the diode will help prevent operation in this range. The VFO as a whole is powered by 12 volts. However, 5 volts derived from a Zener diode regulator (Z) power the oscillator stage and its buffer. This is done to dissipate as little energy as possible in the oscillator and prevent the components from heating.

The entire VFO circuit

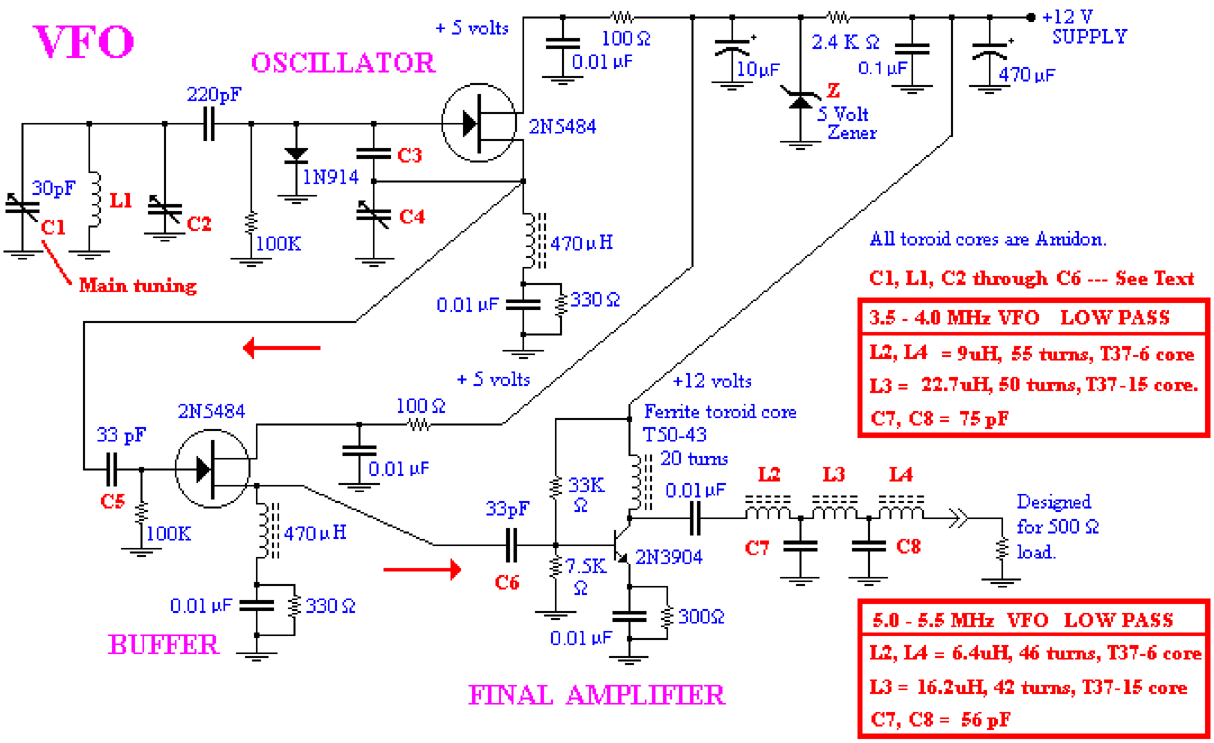
VFO
OSCILLATOR
+ 5 volts
100 Ω
0.01 µF
10µF
2.4 K Ω
0.1 µF
+12 V SUPPLY
470 µF
Z
5 Volt Zener
220pF
2N5484
C3
1N914
30pF
L1
C1
C2
100K
C4
470 µH
Main tuning
0.01 µF
330 Ω
All toroid cores are Amidon.
C1, L1, C2 through C6 --- See Text
3.5 - 4.0 MHz VFO LOW PASS
L2, L4 = 9uH, 55 turns, T37-6 core
L3 = 22.7uH, 50 turns, T37-15 core.
C7, C8 = 75 pF
+ 5 volts
100 Ω
2N5484
33 pF
C5
0.01 µF
100K
470 µH
0.01 µF
330 Ω
BUFFER
+12 volts
Ferrite toroid core T50-43
20 turns
33K Ω
0.01 µF
33pF
C6
7.5K Ω
2N3904
0.01 µF
300Ω
L2
L3
L4
C7
C8
Designed for 500 Ω load.
FINAL AMPLIFIER
5.0 - 5.5 MHz VFO LOW PASS
L2, L4 = 6.4uH, 46 turns, T37-6 core
L3 = 16.2uH, 42 turns, T37-15 core
C7, C8 = 56 pF

Here is the entire VFO circuit, excluding the temperature compensation. As you can see, the VFO also contains a buffer stage and a final amplifier. Temperature compensation consists of special circuits that replace either C2 or C4.

Buffer

The buffer stage separates the oscillator from the final amplifier. Otherwise there would be a subtle connection between the load on the VFO output and the oscillator. Believe it or not, without the buffer, if you change the load on the VFO output slightly, the frequency will change too. The buffer is connected to the other stages by C5 and C6. These capacitors should be as small as possible to reduce the connection between the oscillator and final amplifier.

Final amplifier

The final amplifier raises the VFO output up to the level needed to drive the transmitter. The stage being driven in the transmitter is usually a "mixer" which we shall discuss in the next chapter. But, if the VFO is designed for the 80 meter ham band, then the next stage after the VFO could be a string of class C amplifiers to increase the output power to the final level, say 5 or 50 watts. The VFO final stage will need to deliver at least two volts peak sine wave. Since we have deliberately kept the oscillator and buffer signals tiny, the final amplifier must be a linear or Class A amplifier to get the signal level up to 2 to 5 volts peak. The drive to this transistor is biased with a 33K resistor so that this stage is always turned on.

Low pass filter

The drive to the transmitter should be as pure a sinewave as possible to avoid radiating harmonics outside the hamband. A low pass filter on the output attenuates most of the harmonics above the desired frequency range. A Chebyshev low pass filter is shown designed for a 500 ohm load. In Chapter 6 you were introduced to Chebyshev filters designed for 50 ohms. However, here we have no need for power, just voltage. Therefore, the filter is designed for 500 ohms, which is plenty of drive current for this application. Values are given for both a 80 meter VFO (3.5 to 4.0 MHz) or a 5 MHz VFO (5.0 to 5.5 MHz).

****
**

The 50 secrets of avoiding drift

Suppose you were to build the above VFO without reading the details in the following paragraphs: When you first turn it on, you will be disappointed to find that it drifts a hundred Hz per minute or more. ***The drift is caused by temperature change***. The components expand and contract with temperature and this causes small changes in the capacitance and inductance of the components. Air wafting across the board doesn't allow the temperature to stabilize. ***Drift is prevented by preventing temperature change and by choosing components that change as little as possible with temperature.***

VFO building is an art form as arcane as Grandma's secret pie crust recipe or the fine points of building Cub Scout Pinewood Derby racers. As you'll see, there must be 50 ways to improve the drift problem. I have never built a VFO that was completely "stable" and probably never will. But perhaps that's because I only know the 14 secrets listed below. If you apply as many of these as possible, you should get within the 20 Hz target - and maybe even under 5 Hz.

Secret # 1. Use Junction Field Effect Transistors (JFETs). The first secret of a stable VFO is using a JFET instead of a bipolar transistor. As described earlier, a field effect transistor is better because it is less sensitive to temperature. I have used 2N3823, 2N5484, and 2N4416 N-channel JFET's for VFO oscillators. My impression is that any small N-channel FET works well.

Secret # 2. Seal the VFO in a cast metal box. Simply protecting the VFO from air currents makes a huge improvement. Use a heavy cast metal box so that the temperature will at least change slowly. In contrast, a flimsy sheet-metal aluminum box will heat and cool relatively rapidly. On the other hand, *ANY* box is a huge improvement over not having the circuit sealed from air currents.

Secret # 3. Use single-sided PC board. A double-sided PC board is constructed like a capacitor. That is, thin metal sheets are bonded to a layer of insulator. Unfortunately, the resulting capacitor has a significant temperature coefficient. As temperature increases, the board material expands (thickens) and the capacitance across the board drops. If the VFO is built on traces and islands that have changing capacitance to ground, the frequency of the oscillator will drift slightly.

Secret #4. Mount the oscillator PC board away from the metal case on standoffs. Following the same principle as above, do not mount the single-sided PC board flush against the metal case. By standing the board up and away from the case, the capacitance between the traces and the metal case can be minimized.

Secret #5. Choose and mount all components affecting the oscillator LC circuit carefully. All the L and C components in the oscillator should be designed for minimum temperature drift. Referring to the diagram, it is not just capacitors C1 and C2 that affect the frequency. Capacitors in series with the 220 pF capacitor, C3, C4, and even C5 affect the frequency. To at least a tiny degree, ALL components in contact with these capacitances can affect frequency drift. These include the diode, the RF choke, the transistor, and the 100 K resistor.

Secret # 6. Mechanical variable capacitors should be chosen carefully. Although appropriate mechanical variable capacitors are hard to find, they may be the best solution for you. Pick a capacitor of about 30 to 60 PF, not larger. High capacitance variable capacitors are too sensitive to temperature change. Smaller ones don't tune far enough. ***Don't use a capacitor with aluminum plates*** – they warp too much with temperature. ***Brass is the best metal. Try to find a capacitor with thick, widely-spaced plates.*** Paper-thin plates are compact, but warp too easily.

An assortment of variable caps is shown below. The big brass variable cap just to the right of center is rather big physically. However, out of this group it is the best for a VFO. The capacitor to its right is quite stable, but its capacitance is too small. The tiny green plastic trimmer caps in the clear plastic strip are positive temperature coefficient variables useful for VFO temperature compensation. The round ceramic white and silver trimmer just above them is my favorite 60 pF trimmer. I used dozens of them in the projects of this book. The small variable cap at the bottom center has surprisingly large capacitance, 250 pF. Unfortunately it's too unstable for a VFO.

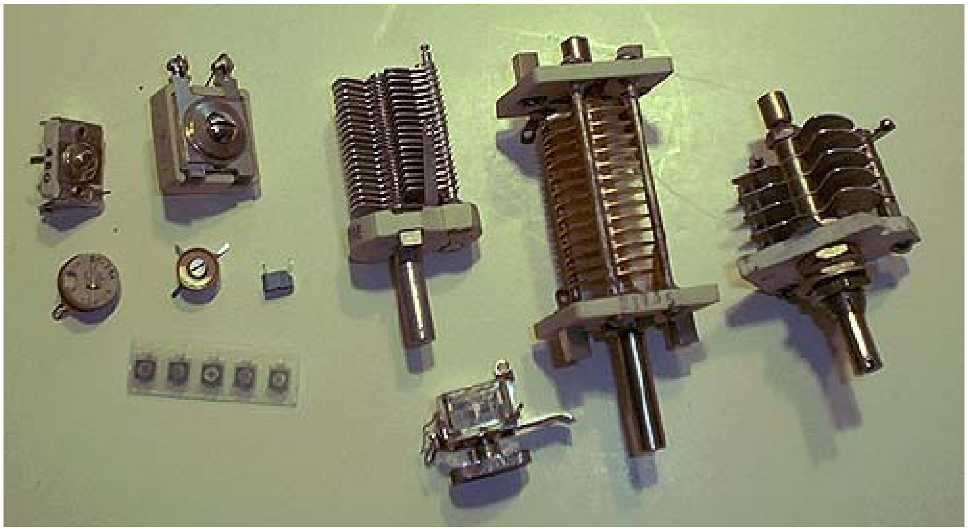

If the capacitor tuning is linear with degrees of rotation, the frequency it produces will be somewhat non-linear. Ideally, the capacitor plates should have a non-linear shape that allows it to tune an LC resonant circuit so that the frequency will be linear. Rotate the capacitor through its range and you'll see that a compensated capacitor has rotor plates that are not simple half circles. As they rotate, they do not mesh with the stator plates at the same point. The non-linear correction isn't a big deal, but it is something to be aware of. RF Parts Co. sometimes has adequate capacitors in stock. See www.rfparts.com.

Secret # 7. Varactors are the most stable tuning element. It's hard to buy mechanical variable capacitors that are mechanically and thermally stable. Collins Radio formerly tuned their VFOs with special powdered-iron slug-tuned coils, but I've never seen any for sale. A varactor capacitor controlled by a quality pot can be a good solution to these problems. Varactors are a kind of silicon diode biased with DC voltage. In my experience varactors are an order of magnitude more thermally stable than mechanical capacitors. They are at least two orders of magnitude more mechanically stable. You can slap the VFO with your hand and, although other components may vibrate, the varactor will not change its capacitance.

Unfortunately, varactors produce a non-linear scale on the frequency tuning knob. This means that the high frequency end of the VFO range will be extremely detailed while the low end may be compressed into a few degrees of rotation. For this to be usable, ideally the potentiometer should be non-linear to compensate. Varactors are discussed in detail below.

Secret # 8. Use NP0 fixed capacitors. When selecting fixed capacitors, look for type C0G (formerly known as NP0). These are supposed to have minimum temperature change. Use these for *ALL* fixed capacitors affecting the LC circuit.

Secret # 9. Use multiple C0G(NP0) capacitors in parallel to achieve a given value. For example, you might use fixed capacitors for C1 or C2 in the above circuit. For each value needed it is better to use several small ones in parallel rather than one large capacitor. The temperature of a small capacitor stabilizes quickly, whereas heat builds up more slowly in a larger capacitor.

Secret # 10. Temperature compensation for the LC circuit is essential. It took me four prototypes to accept this, but temperature compensation is as important as putting the VFO in a box. Lots of guys claim to have succeeded without it, but I never have. Not using temperature compensation implies that every capacitor and inductance in the VFO must have a zero temperature coefficient. Alternatively, all negative coefficients must be balanced precisely with components that have equivalent positive temperature coefficients. Good luck doing that! Compensation circuits are described below.

Secret # 11. Use an air core inductor. As usual, it is most convenient to use a powdered iron toroid core. Unfortunately, powdered iron changes its permeability (magnetism factor) with temperature. Therefore, by not using iron, another variable is eliminated. I have successfully used old plastic ball point pen caps as little coil forms for air-core inductors. I bore little holes in the plastic to accept tiny pieces of stiff copper wires to serve as wiring terminals.

If you do use powdered iron, among the CWS (Amidon) cores, type #7 is supposed to have the best thermal stability. CWS Bytemark (Amidon) #6 cores have worked reasonably well for me but maybe #7 would be a few percent better. If you make a coil out of turns of copper wire on a plastic form, the copper will change its dimensions slightly with temperature too. And because an air core inductor requires more turns of wire, there is more opportunity for the copper to change its dimensions, its interwinding capacitance, and also its resistance. Finally, an air-core coil will couple like a transformer to nearby parts whereas a powdered iron toroid couples far less. Sigh. Nothing is perfect.

After you have your coil wound and working over the right frequency range, epoxy or clamp it to the board. Without the epoxy, the frequency will warble with the slightest vibration. I once tried to use slug-tuned coils. They were convenient to adjust, but were mechanically and thermally unstable.

Secret # 12. Precision voltage regulation for the VFO supply is vital for precision frequency stability. The 12 volt supply for the VFO as whole must be regulated. Ordinary voltage regulators like the LM317 or LM7812 gave me regulation within 0.1 volt. This was OK for frequency stability down to about 20 Hz drift. But to get down to less than 5 Hz, I needed to regulate my VFO power supply to a few millivolts. To achieve this, I built a precision supply that just powers the VFO. The less current the supply has to deliver, the more constant its output voltage will be. The supply is discussed in detail below.

Secret # 13. The VFO should draw as little power as possible. The less power drawn, the less heating that occurs inside the VFO box. Also, the less power drawn, the easier it is to build a precision voltage supply to drive the VFO. That is why the VFO was designed for a 500 ohm load rather than 50 ohms like most ham RF circuits. The VFO as a whole should draw less 20 mA DC. 10 mA would be even better.

My next attempt at a VFO will house the oscillator and buffer inside the cast metal box and I’ll put the "high power" final stage outside the box so its heat can't affect the oscillator. Another idea I'm going to try is to run the oscillator on even less voltage, perhaps 1.5 to 2.5 volts.

Secret # 14. Forget tube oscillators. You old timers may be tempted to use a tube oscillator. I first tried to update an old tube VFO, but tubes get hot and make temperature compensation too difficult. You’ll have plenty of trouble without this extra burden. You may use bipolar transistors for the final amplifier in your VFO, but not for the oscillator. For good measure you may as well use a JFET for the buffer as well.

Vernier mechanical tuning and frequency indicator

Because a VFO must be tuned precisely to the other fellow's frequency, it is vital to use a vernier tuning gear between the tuning knob and the variable capacitor. In my opinion, the tuning knob should rotate at least three times around for each revolution of the capacitor. Without vernier tuning, it will be exceedingly difficult to tune your receiver or transmitter accurately to the other fellow's frequency. A planetary reduction gear is generally mounted on the front panel. Machine screws clamp it to the ¼ inch capacitor tuning shaft that protrudes from the VFO box. Tuning gear systems are usually combined with a dial and pointer that you can calibrate. A paper dial can be marked with ink for the calibrations. A plastic cover protects the paper from moisture.

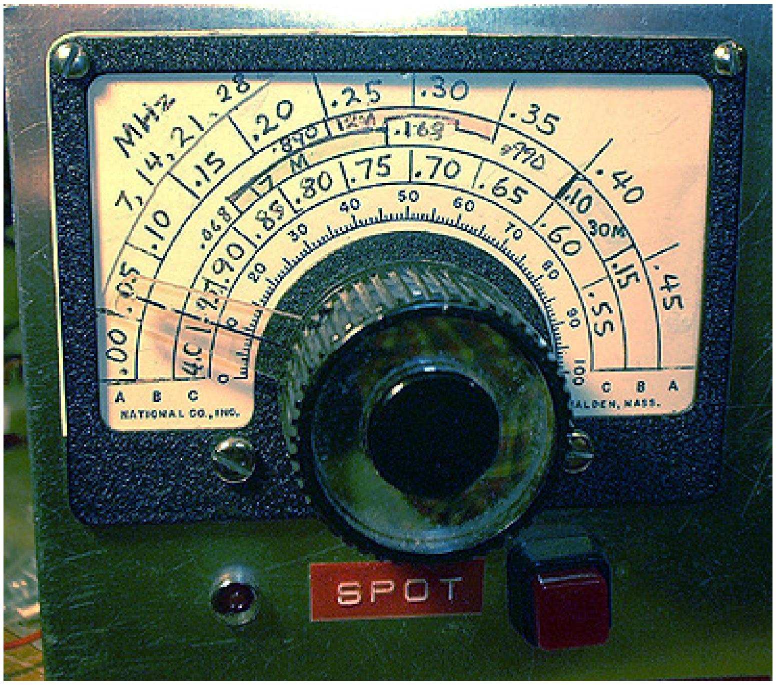

I used a National Company brand vernier dial on one of my VFOs shown at left. Recent interest in QRPs has made these available again after being absent for some years. Look for the ads in ham magazines. Unfortunately, these dials are pretty pricey. For some of my VFOs I used a military surplus reduction gear that had no pointer and dial. I made the pointer out of super-thin PC board painted black. The dial calibrations were on thin white cardboard covered with 1/8 inch Lucite plastic and screwed onto the front panel.

If you use a varactor variable capacitor as described below, you may be able to find a multi-turn potentiometer for your tuning vernier. This solves the vernier problem, but doesn't offer a way to calibrate the dial. Some hams have built elaborate frequency counters or digital voltmeters as solutions for VFO calibration. All I can say is beware of digital circuitry in your ham equipment. Homebuilt digital circuits (in my experience) always generate radio noise that will interfere with hearing weak signals. Commercial manufacturers make digital technology look easy. But so far, all my digital toys have generated a static "hiss" that I have regretted.

****

Varactor Tuning

While building a new receiver, I had to construct another VFO. In this prototype I explored replacing the mechanical tuning capacitor with a varactor. Back-biased P-N junctions block the flow of charge as if they were capacitors. They not only act like capacitors, when back-biased, ***they are capacitors***. The interesting aspect of this behavior is that, by biasing them with a DC voltage, say 0 to 10 volts, the capacitance can be tuned like a variable capacitor. As more and more voltage is placed across the diode, the ion contaminants in the semiconductor are used up and the charge it can store is diminished. In other words, PN junction diodes change their capacitance inversely proportional to the voltage on them. ***Varactors are voltage-variable capacitors.***

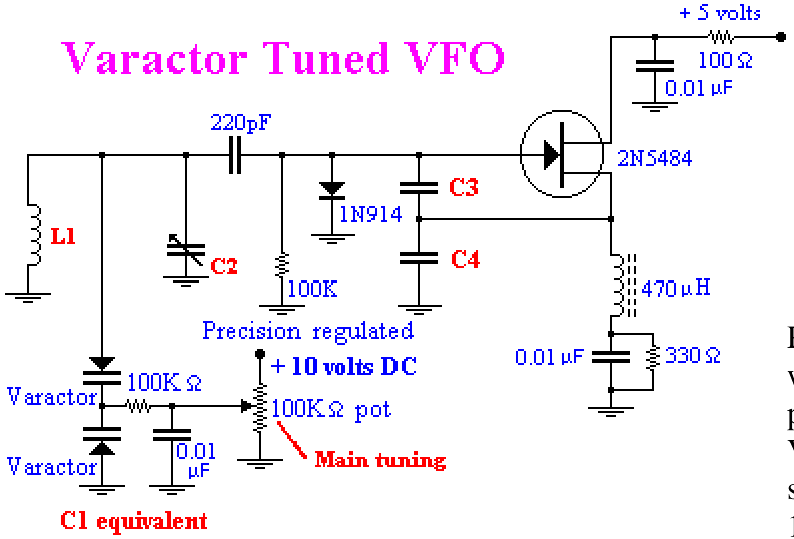
Varactor Tuned VFO
+ 5 volts
100 Ω
0.01 µF
220pF
2N5484
C3
1N914
L1
C2
100K
C4
470 µH
Precision regulated
+ 10 volts DC
0.01 µF
330 Ω
100K Ω
Varactor
100K Ω pot
Varactor
0.01 µF
Main tuning
C1 equivalent

Varactors are specialized P-N junction silicon diodes that were designed just for this purpose. However, I have seen VFO circuits that use ordinary silicon diodes like the 1N914 or 1N4148 as varactors. One of the disadvantages of most varactors is that they usually don't have much capacitance. 5 to 20 pF is typical. Because of this, I didn't believe I could get enough tuning range from a varactor. It turned out that I could compensate by using two or more varactors in parallel and by decreasing the size of the feedback capacitors C3 and C4. Tuning range wasn't a problem.

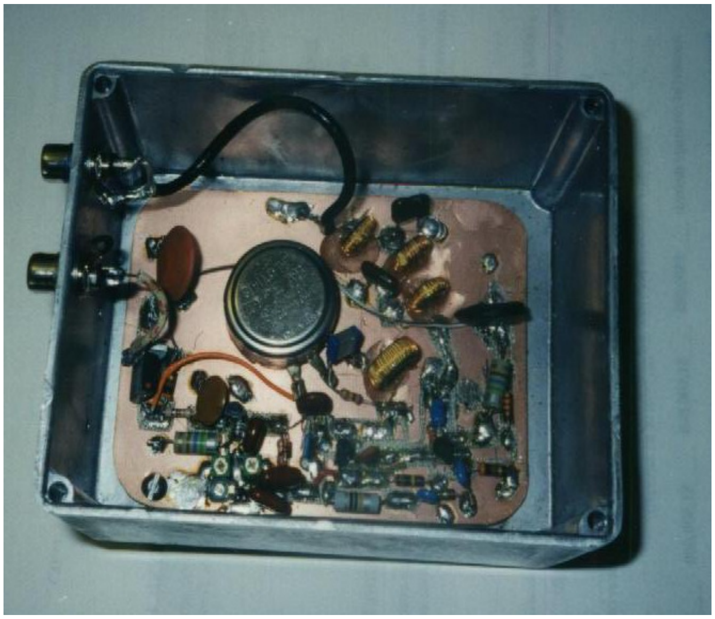

A VFO tuned by varactor. The round potentiometer adjusts the voltage on the varactor.

Since then I bought a varactor, Motorola type MV104, that has a capacitance of 110 pF. (!!!) This device has the potential to provide the large tuning range needed for covering the 10 meter band. Alternatively, it can be operated over a narrow range of bias voltage and thereby largely solves the non-linearity problem.

The advantages of varactor tuning are:

1. Varactors are mechanically stable. Assuming the potentiometer driving your varactor is mechanically stable, then the resulting VFO will be mechanically stable. You can bang the table with your fist and the frequency will barely warble in the receiver.

2. Varactors are more thermally stable than mechanical capacitors. When testing a mechanical capacitor VFO circuit with a hair dryer, I found that the blast of hot air on the circuit board caused the frequency to soar or plunge hundreds of Hz, sometimes even KHz. When my varactor VFO is given the same treatment, the frequency change is an order of magnitude less.

3. Varactors are available. Good mechanical tuning capacitors are hard to find. In contrast, varactors can always be purchased from Digi-Key, Mouser, and other companies.

4. Varactors are tiny. Some of the ones I used are the size of a grain of sand. Soldering them on my circuit board required patience, sharp tweezers, and a jeweler's loupe. A VFO module made with a varactor can be much smaller than a VFO made with a mechanical variable capacitor. Also, the varactor can be tuned by a potentiometer located at some distance away on a front panel.

5. Varactor VFOs may be tuned by phase-lock loops. Since the varactor VFO is tuned by a variable DC voltage, it can be part of a modern phase-lock loop design. A homebrew VFO doesn't have to be confined to old technology.

The problem with varactors

Alas, varactor-tuned VFOs are non-linear. However, you can turn this to your advantage. As the DC voltage is changed across the varactor, the frequency change it produces is not linear. When the voltage is first applied, the holes and electrons in the PN junction are filled in readily and decrease the capacitance rapidly. After the first big change, more and more voltage must be applied to fill in more holes and deplete the electrons in the N-type semiconductor. In other words, the wider the tuning range, the more non-linear the relationship between applied voltage and frequency. This exaggerates the tuning of the high-end frequency end of the band. For example, if the varactor is pushed to its maximum capacitance range, then 75% of the voltage tuning range might be needed to cover the upper 25% of your total frequency range.

Suppose you are primarily interested in CW operation. The CW portion of the band is always at the bottom of the hamband with phone at the high frequency end. The linearity problem can become an advantage by designing the VFO frequency converter in your transmitter or receiver so that, for each band, the high frequency end of the VFO range covers the low CW end of the ham band. That is, the VFO frequency should be ***subtracted*** from a higher crystal controlled converter frequency.

CW signals have little bandwidth, a few Hz, and the CW bands are often crowded. In contrast, although phone bands are sometimes crowded, the phone bands are several times wider and the phone signals themselves each cover 3 KHz. In other words, good bandspread, (small frequency change for big knob rotation), is important for the CW band and not so important for the phone band. Yes, tuning in SSB phone requires fine-tuning. But you may find that tweaking the speech quality is more easily done with the receiver BFO knob than with the VFO tuning.

For example, in a transmitter, the VFO might range from 5.0 to 5.5 MHz. To transmit on

40 meters (7.00 to 7.30 MHz), the transmitter could use a 12.5 MHz crystal controlled local oscillator to cover from 7.0 to 7.5 MHz. That is, 12.5 MHz minus 5.5 MHz = 7.0 MHz. In this way, the high end of the VFO tuning covers the *LOW* end of the ham band. In contrast, if you use a low frequency crystal, 2.0 MHz (e.g., 2.0 MHz + 5.3 MHz = 7.300 MHz), the high frequency end of the VFO will cover the upper end of the phone band where bandspread isn't important. If you're confused, frequency converters are explained in detail in chapter 11.

Frequency offset on transmit

A transmitter VFO has a problem you may not have thought about. With an old-fashioned, wideband CW receiver, when you tune past a CW signal, you'll hear a whistle that changes from a high pitch to a low pitch or "zero beat," then rises back up to a high pitch as you continue to change frequency. If your receiver were tuned precisely to the other guy's frequency, he would be at the "zero beat." His CW signal would have such a low pitch you wouldn't be able to hear it.

To fix this, modern transceivers automatically add a frequency offset between receive and transmit, usually 700 or 800 Hz. Also, fancy transceivers don't receive the lower half of the signal unless the operator selects LSB on the front panel. In general, the lower sideband (LSB) is used on 160, 80, and 40 meters while the upper sideband, (USB) is used on 60 meters, 30 meters and above. With a modern, narrow bandwidth receiver, modern hams may not even be aware that there are two sidebands. The upshot is that, when you answer a CQ with your homebrew VFO, you must tune in the correct direction about 700 Hz above or below his zero beat point. Otherwise he won't even hear you. Old timers used to tune around, but modern guys don't. I believe I had this problem when I first got on the air with my homebrew rig. Few stations seemed to be able to hear me. Yet, when I did talk to someone, I got good signal reports.

The simplest solution is to use the narrow bandwidth filters on your receiver so that you can only hear the upper or lower sideband at one time. Then when you zero beat your signal on top of the fellow calling CQ, you only hear your VFO when you are on the correct sideband. Narrow passband filters for a homebrew receiver are discussed in chapter 13.

VFOs for transceivers

If you build a VFO for a transceiver, the VFO will be used during both receive and transmit. As explained above, it may be useful to add a varactor tuned offset adjustment to the VFO so that you can send and receive on slightly different frequencies. To do this, add an auxiliary low capacitance varactor adjustor in parallel with the main tuning. The technology is the same as for varactor tuning, but the tuning range will be less than one KHz.

****

A precision power supply for the VFO

One of my VFOs had a relentless upward drift of 200 Hz per hour. I was puzzled until I noticed that my 12 volt VFO power supply had a subtle downward drift. Ordinary voltage regulators are crude compared to temperature compensated regulator ICs. Regulators like the LM317 or LM7812 drift hundredths of a volt per minute, especially if the load is more than 100 milliamperes. A simple Zener diode regulator can easily allow one or two tenths of a volt change on the oscillator. The solution is to build a precision voltage regulator. This 12 volt regulator should be located outside the VFO box. So far as possible, ***anything that generates heat should be kept outside the VFO box.***

Precision Zener diodes

The trick to building a precision, temperature-compensated power supply is a precision Zener diode reference. Ordinary Zeners vary their regulation voltage with temperature. Precision Zeners are integrated circuits that behave like Zener diodes, but have temperature compensation circuitry and can be adjusted to exactly the rated voltage. The LM336–5.0 Zener diode can be adjusted with a trim pot to precisely 5.000 volts DC. It is designed to have the best temperature compensation at that exact voltage. Other than the pot and the diodes, this part is used like an ordinary Zener.

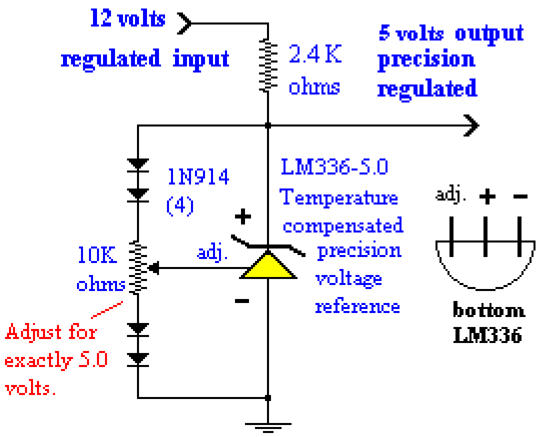

Precision 5 volt Zener

Yeah, I know. I promised I wouldn't use ICs that I couldn't build myself. I'm afraid precision regulators are beyond me. I've tried, but my attempts never exceeded the precision of the most ordinary IC voltage regulators. I did manage to build a pretty good voltage reference consisting of Zeners regulating other Zeners in a cascade in which some of the resistors were thermistors. Unfortunately, the final output was only stable at tiny voltages and tiny load currents and I couldn't successfully leverage it up to useful voltages and currents.

A precision 12 volt power supply

The precision voltage regulator shown below comes from the National Semiconductor Data Book for linear integrated circuits. This regulator will hold the VFO voltage constant to within about 2 millivolts. It uses a programmable LM317T regulator. (The "T" at the end identifies the TO-220 case style.) The output of the big regulator is “modified” with a precision Zener reference regulator to hold the output voltage constant. The LM317 regulates the voltage across the 1.2 K ohm resistor to about 1.2 volts. The precision reference then regulates the voltage across the 620 ohm resistor to exactly 6.2 volts - that is, 1.2 volts plus 5.000 volts. Since the voltage across the 620 ohm resistor is held constant, the current passing through the 1K potentiometer is also held constant. Therefore, adjusting the 1K pot can adjust the total voltage.

In order to maintain this high degree of precision, ***the above regulator just powers the VFO.*** When I tried to run other transmitter stages with the precision power supply, the load rose to several hundred milliamperes. Although the regulator chips were operating within specifications, the extra load ruined the millivolt precision regulation. I retrofitted my old capacitor-tuned VFO with the same precision regulator and found the frequency stability became as good as my newer varactor VFO.

I try to keep my VFO supply current down to less than 10 milliamperes. If you need more than a 3 volts peak output voltage, you’ll probably find this difficult. In the full circuit shown above, by the time you get a 5 volts peak, pure sinewave, the current drawn from the supply is likely to approach 20 milliamperes. One way to compensate is to adjust the voltage from the precision supply to the bare minimum needed to provide the output waveform you need. That is, instead of using regulated 12 volts, use regulated 8 or 10 volts.

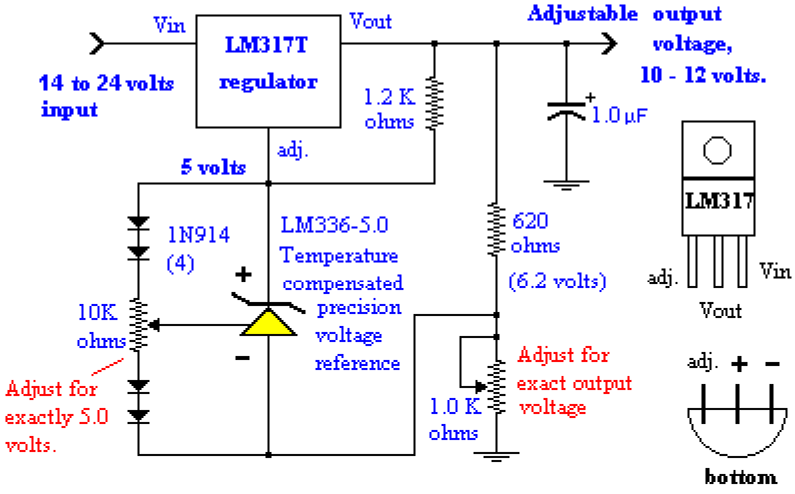
Vin
LM317T
regulator
Vout
Adjustable output voltage, 10 - 12 volts.
14 to 24 volts input
1.2 K ohms
+
1.0 µF
adj.
5 volts
LM317
1N914 (4)
LM336-5.0 Temperature compensated precision voltage reference
620 ohms
(6.2 volts)
adj.
Vin
Vout
+
−
10K ohms
Adjust for exactly 5.0 volts.
Adjust for exact output voltage
1.0 K ohms
adj.
+
−
bottom LM336

Precision Supply for the VFO

Separate regulation for the VFO oscillator

If your VFO is consuming more than about 10 mA, you may want to compensate for the temperature change inside the box by powering the oscillator inside the VFO box with 5 volts from a separate precision 5 volt or 2.5 volt Zener instead of using an ordinary 5 volt Zener.

A voltage doubler for battery use

If you're line powered, the regulator described above will probably work just fine, unless there is a brown out. On the other hand, if your transmitter is battery powered, its output voltage will fall well below 12 volts as the battery discharges. One solution is to run the VFO on regulated 9 volts. That way, even with 10 volts left on the battery, the regulator will still be providing 9.000 volts. However, if your VFO uses a varactor as the tuning element, you may need at least 10 volts to get the maximum tuning range from the varactor. My solution to this frustration was to first double the unregulated battery voltage. By starting with nominally two times the battery voltage, my VFO will always have at least 12 volts. Because the VFO only draws 10 to 20 mA DC, this DC voltage doubler does not have to be powerful.

****

A DC voltage doubler

Square wave generation

Any time you wish to raise a DC voltage, it is necessary to use the existing DC supply to generate a source of AC voltage. For example, the AC voltage could be applied to a transformer to produce as high an AC voltage as you need. The high AC voltage would then be rectified back into DC at the required high voltage. Instead of using a transformer, the voltage doubler described below uses a capacitive ***charge pump*** to raise the voltage. As you'll see, this technique is a specialized kind of rectification.

The first task is to convert the DC supply to square wave pulses. It turns out there are many ways to generate square waves using integrated circuits. For example, I used an operational amplifier IC to generate the dots in the electronic bug in chapter 9. You may use an IC if you like, but you might enjoy doing it the hard way. As usual, if you're new to electronics, you'll learn some interesting concepts.

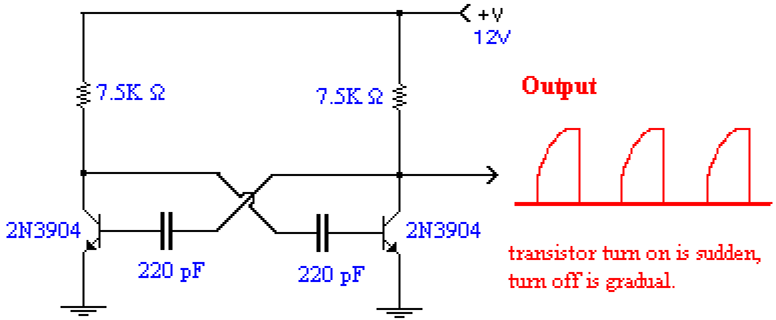

transistor turn on is sudden, turn off is gradual.

SIMPLE ASTABLE MULTIVIBRATOR

A simplified ***astable multivibrator*** oscillator is shown to the left. As you can see, this circuit consists of two grounded emitter amplifiers wired so that any change on the collector of one transistor is immediately coupled to the base of the other. Let's assume that the capacitor on the left is charged to a low voltage, say 1 volt. The capacitor on the right is charged to nearly 12 volts. The capacitor on the left is charging toward 12 volts through the 7.5 K resistor on the right. As the positive charge flows out of the capacitor and into the base, the left transistor turns on. This pulls its collector and its respective capacitor toward ground.

Since the voltage across the capacitor on the right can't change instantly, the voltage on the right base is pushed downward to roughly –12 volts. This extreme negative voltage turns off the right hand transistor. The right hand capacitor with its –12 volts will discharge toward zero volts since there is now no source of voltage to maintain negative 12 volts. This discharge takes a significant length of time because the current must bleed into it through the 7.5 K ohm resistor. Eventually, the voltage on the right transistor's base rises above +0.6 volts, which will turn that transistor back on. When the right transistor turns on, it pushes the left hand capacitor back down to –12 volts turning off the left transistor.

A practical multivibrator

The simplified multivibrator described above oscillates OK, but it doesn't always start spontaneously. In fact, you will find that it only runs at a specific range of supply voltage and must be turned on abruptly. If you turn up the supply voltage gradually, the multivibrator remains stable with one or both sides turned off and the oscillation never begins.

The unreliability of the simple multivibrator can be fixed by biasing the transistors partly *ON* with the 100K ohm resistors. These resistors insure that the capacitors will always be charging or discharging. Now the circuit will generate square waves even with very low voltages. When the power supply voltage is turned up slowly, the oscillation will always start. Also, the *ON* state of each transistor is maintained longer and a better square wave results.

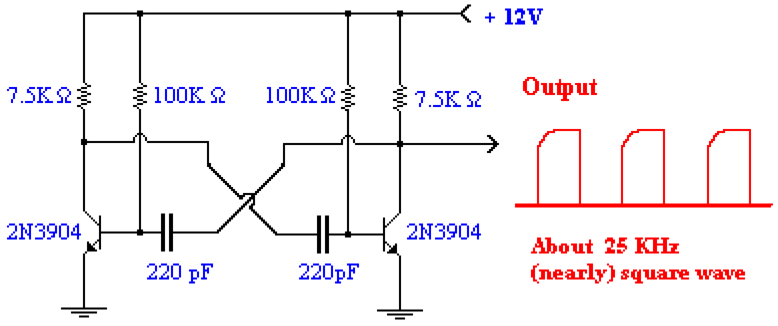

ASTABLE MULTIVIBRATOR

Bistable multivibrators are RAM flip-flops

This is off the subject, but suppose that the two capacitors in the above circuit were replaced with high resistance resistors. Since there would be no reactance to charge or discharge, the circuit would "lock up" with one transistor *ON* and the other *OFF*. This is called a ***stable-*** or ***bistable-multivibrator.*** If a pulse is introduced to one transistor or the other, the circuit can be made to "flip-flop" to the opposite stable state in which the *OFF* transistor turns *ON* and the *ON* transistor turns *OFF*. The ***flip-flop*** circuit is the basis of static RAM memory (SRAM). One flip-flop can store one bit of information. So long as the supply voltage is applied, the circuit will "remember" one bit of information indefinitely. Or, it will remember it until another pulse comes along and resets it to the opposite state. In an integrated circuit SRAM literally millions of flip-flops are printed on a chip storing megabits of information.

Squaring up the low power square wave to drive a doubler charge pump

Now that we have a low-power, square-wave voltage, we need to clean up the waveform and amplify it so that we can produce about 30 mA squarewave AC for our VFO power supply. This is done with a simple buffer amplifier to make it square. A square waveform is important because the less time that the circuit spends "half turned on," the more efficient the power supply will be.

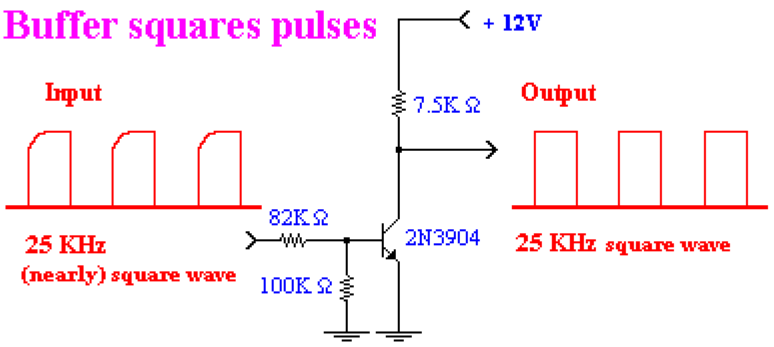
Buffer squares pulses

This square wave buffer is just a high gain voltage amplifier. During the sloping, rising input voltage, the upward slope of this voltage is exaggerated by 10 or 20 times. This diminishes the rise time until the rise time becomes negligible.

Diode voltage doubler

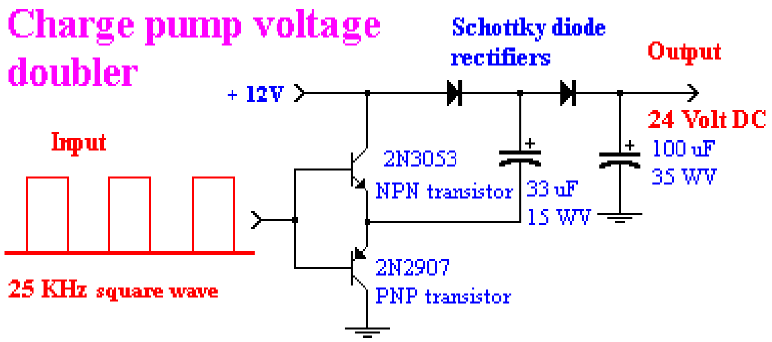

How can we "double" DC voltage using diodes? The idea is to repeatedly charge the 33 µF capacitor to +12 volts as though that capacitor were a rechargeable battery. Then the capacitor is switched out of its "charging mode" and added like a battery on top of the existing 12 volt supply. In other words, 12 volts plus 12 more volts equals 24 volts. This pulsing, 24 volt signal then charges up the large 100 µF storage capacitor on the right to 24 volts. Provided the current drawn from the 24 volts supply is small, the capacitor on the right can maintain a relatively constant voltage approaching 24 volts.

As shown above, the PNP and NPN transistors work together to pull the left hand capacitor up and push it down. When the PNP transistor is turned on, the bottom end of the capacitor is connected to ground. In this condition the diode on the left charges up the capacitor to 12 volts. When the PNP turns off and the NPN turns on, the bottom end of the capacitor is suddenly shoved "up" and tied to the 12 volt supply line. Since the top of the capacitor is suddenly 12 volts higher than the supply line, the diode on the left is back-biased and can no longer charge it. Instead, the diode on the right is now forward biased and will discharge the 12 volts into the storage capacitor on the right. The capacitor on the right charges toward 24 volts thereby creating twice the original voltage.

Complementary transistors in action

In chapter four I mentioned that it was often convenient to have transistors that work with opposite polarities. As the drive to one transistor turns it on, the same polarity turns its complement transistor off. The output is taken off the emitters. In effect, the complementary transistors connect the output back and forth between ground and the 12 volt supply without any resistors becoming hot and wasting energy.

By the way, the logic circuits in your personal computer are almost entirely implemented with integrated circuits made from Complementary N-channel and P-channel MOSFET transistors. These ICs are called ***CMOS***. By avoiding resistor loads on the FET drains, heating is minimized and switching speed is maximized.

Use Schottky rectifiers for best efficiency

Now we have loads of extra voltage, even when the battery supplying the transmitter is nearly exhausted, so there is still plenty of voltage for the precision regulator to generate 12.000 volts. The supply works best if the two diodes in the voltage doubler are power Schottky diodes. These rectifier diodes only waste about 0.2 volts per diode when conducting current.

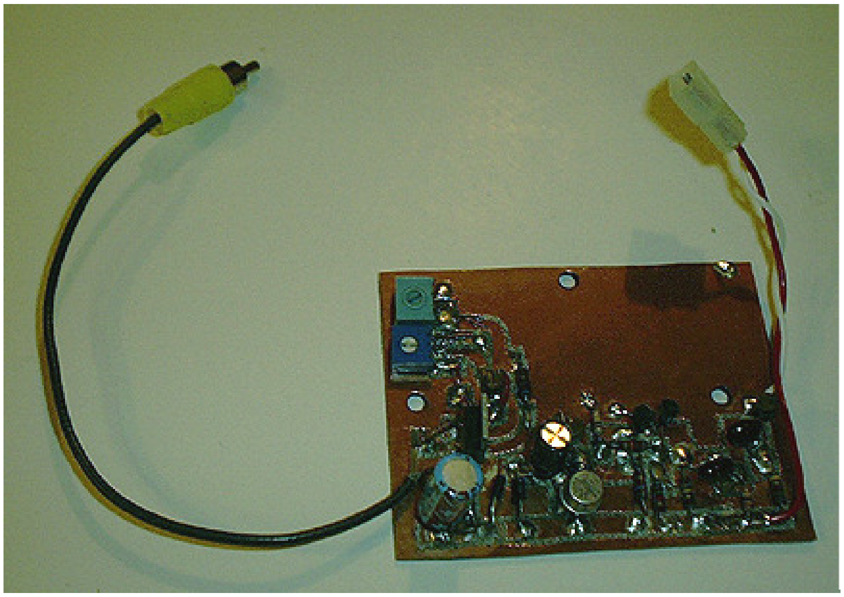

The down side of Schottky rectifiers is that they can usually only tolerate about 30 volts maximum, but that's plenty for this application. In the circuit shown on the left I used ordinary 1N4001 silicon rectifier diodes. Each of these wastes about 0.7 volts when turned on. Consequently, my 24 volt supply only delivers about 22 volts when the input is 12 volts.

VFO power supply board. It has a variable battery voltage input, precision regulated 12 volts output.

****

TEMPERATURE COMPENSATON

Supposedly, good VFOs can be built without temperature compensation. Personally, I've never succeeded, but don't let me be a wet blanket! Go ahead and try. Just leave room on the PC board to add compensation later.

The challenge is that most VFO parts such as variable capacitors, PC boards, etc. all have negative temperature coefficients. That is, their capacitance drops as temperature rises. The strategy of compensation is to substitute C2 or C4 with a capacitor that has an adjustable ***positive temperature coefficient***. Therefore by paralleling the capacitance with a capacitor that has an equal but opposite positive temperature coefficient, the capacitance change will be cancelled. I suggest you start with this method of temperature compensation since it is the easiest.

Positive coefficient trimmer capacitive compensation

If you look in Digi-Key, Mouser, or other parts catalogs you'll occasionally find trimmer capacitors with positive temperature coefficients. These are capacitors that increase their capacitance with rising temperature. I used 20 to 40 pF variable trimmer capacitors made by Sprague-Goodman, type GCL. I replaced some of the fixed capacitance in the capacitive divider, C2 and C4, with two positive coefficient trimmers in parallel. ***The clever part is obtaining both the correct positive coefficient and the correct amount of positively changing capacitance. This is done by using a similar negative coefficient trimmer in parallel with the positive trimmer.*** By balancing the positive coefficient trimmer directly against an ordinary negative coefficient trimmer, you can produce a total positive coefficient value that compensates for the actual negative coefficient of the rest of your LC circuit.

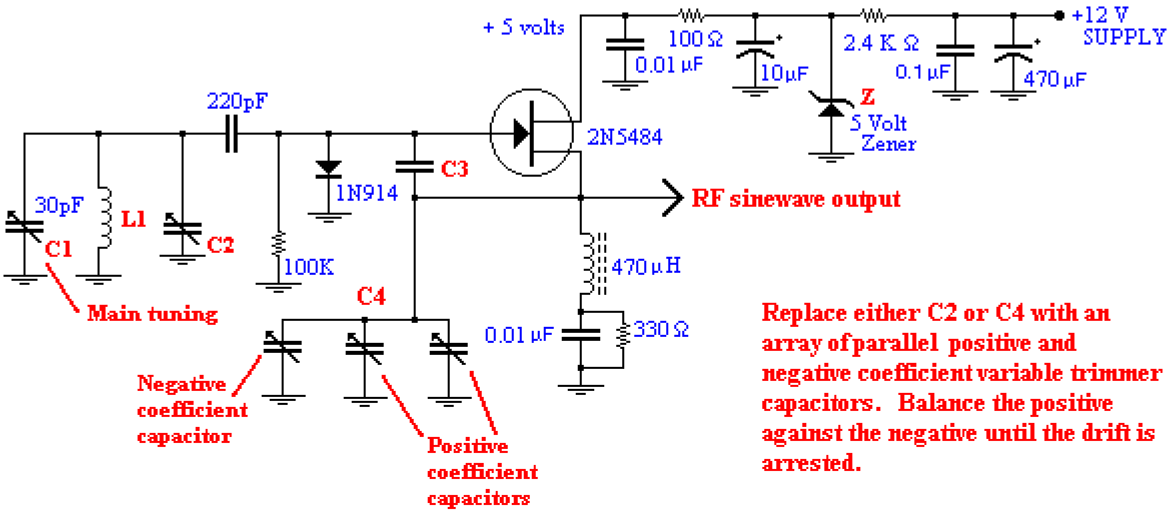

Replace either C2 or C4 with an array of parallel positive and negative coefficient variable trimmer capacitors. Balance the positive against the negative until the drift is arrested.

POSITIVE TEMPERATURE COEFFICIENT COMPENSATION

Adjusting VFO temperature compensation

Put a 500 ohm load on the VFO output and put probes across the load for your scope and frequency counter. Check that your voltage supply is set where you want it, say 12.000 volts. It should be rock stable within a millivolt or two. If you're using a mechanical variable capacitor, set the capacitor to maximum capacitance. Now adjust the positive and negative coefficient trimmers to midscale at the desired low-end frequency of the VFO. For example, with a 5 to 5.5 MHz VFO, you would set it to 5.000 MHz (or just below). Check that you can still tune up to 5.5 MHz using the main tuning capacitance. If you can't, you may have to change C3, or fiddle with C4 or C2. Remember that C2 can also be a positive or negative coefficient trimmer, depending on what you need.

Now put the lid on the box but don't screw it down. Watch the frequency drift on your frequency counter. You will almost certainly see it ***drift relentlessly up or down.*** If it is going down, assume this is caused by increasing temperature in the box. Turn a positive trimmer to *LESS* capacitance, then adjust a negative trimmer to restore the frequency back to where you started. Repeat this again and again until the drift direction reverses. Now it should be going relentlessly *UP*. Now take off just a bit of the negative trimmer until the drift comes to a halt. When you get them balanced well, the frequency will still change, but now it will ***wander up and down but soon return to the same frequency***. That is, the frequency will no longer march continuously in the same direction. When you get to this point, you have done as well as you can with your present set of components. With luck, during any given minute it won't wander up or down more than 5 Hz.

****

Thermistor temperature compensation

You may find that your positive trimmer capacitance compensator just doesn't give you enough positive compensation to do a good job. A ***thermistor*** compensation circuit may be just what you need. I had good results with the ARRL handbook circuit below.

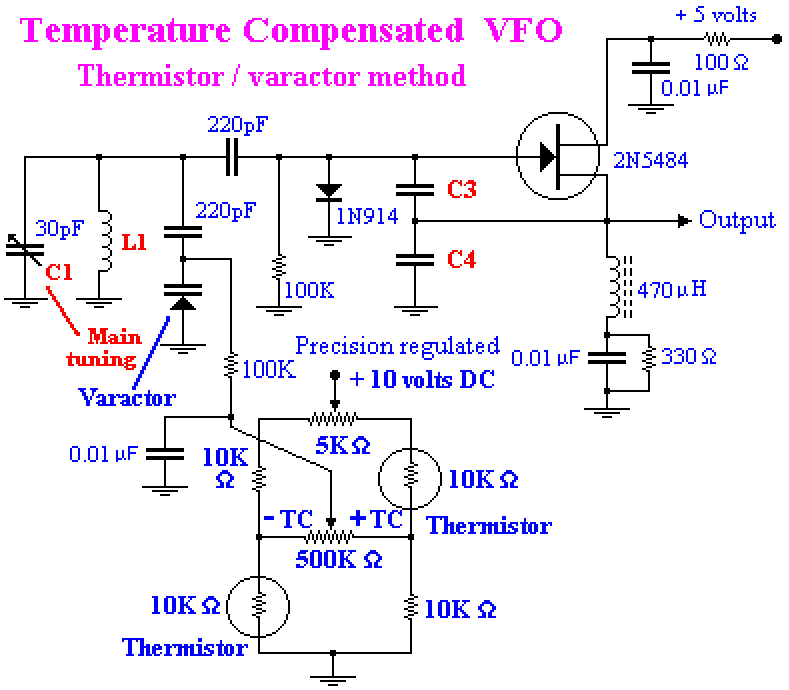
Temperature Compensated VFO
Thermistor / varactor method
+ 5 volts
100 Ω
0.01 µF
220pF
2N5484
C3
1N914
Output
30pF
L1
220pF
C1
100K
C4
470 µH
Main tuning
Varactor
100K
Precision regulated
+ 10 volts DC
0.01 µF
330 Ω
0.01 µF
10K Ω
5K Ω
10K Ω
Thermistor
- TC
+ TC
500K Ω
10K Ω
Thermistor
10K Ω

Resistors are usually designed to change as little as possible with temperature. However thermistors are resistors made from semiconductor that has a large temperature coefficient. The coefficient can either be positive or negative, and both kinds can be used in the above circuit. The thermistors are placed in a bridge circuit with a thermistor on each side of the bridge. The entire circuit is powered by a precision regulated supply, 5 to 10 volts. By adjusting the 5K and 500K trimmer pots at the top and middle, the degree and direction of the compensation can be selected. Tuning the 500K pot to the right selects more positive compensation. Tuning to the left selects less positive compensation or possibly even negative compensation. If you need more capacitance, you can always parallel two (or more) varactors. The tune up is done much like the capacitive compensation version. The good news is that I found the thermistor-varactor compensation much easier to adjust.

Temperature regulation

Another approach to temperature compensation is to hold the temperature constant by heating the VFO and regulating the temperature with a thermostat. I built such a device inside the lid of my cast metal box. The heater had a resistive heater element and a thermistor-controlled temperature regulation feedback system. The good news was that it regulated the box temperature to within 0.1 degree Fahrenheit. The bad news was that it took the temperature at least 30 minutes to stabilize and, when installed in the transmitter, the regulator was trying to heat the entire transmitter and perhaps the whole room. In other words, a heated VFO box will need extensive thermal insulation in order to be practical. I gave up on this idea.

In conclusion,

Precision voltage regulation, temperature compensation and careful adjustment can produce a homebuilt VFO that doesn't drift like a homebuilt. When you get on the air and describe your rig as "all homebrew," the stations you work will (sometimes) shower you with praise. Enjoy every compliment. If you needed as many prototypes as I did to develop a reasonably good VFO, you deserve kudos. Actually, the best complements I've had were when I yacked with guys for an hour without any complaints about my drift. This was especially true if I was able to avoid discussing my funky homebrew transmitter. For all they knew, I was using a first rate modern transceiver.

CRYSTAL SETS TO SIDEBAND
© Frank W. Harris 2010, REV 12

Chapter 11

Building a VFO for the higher bands (PMOs)

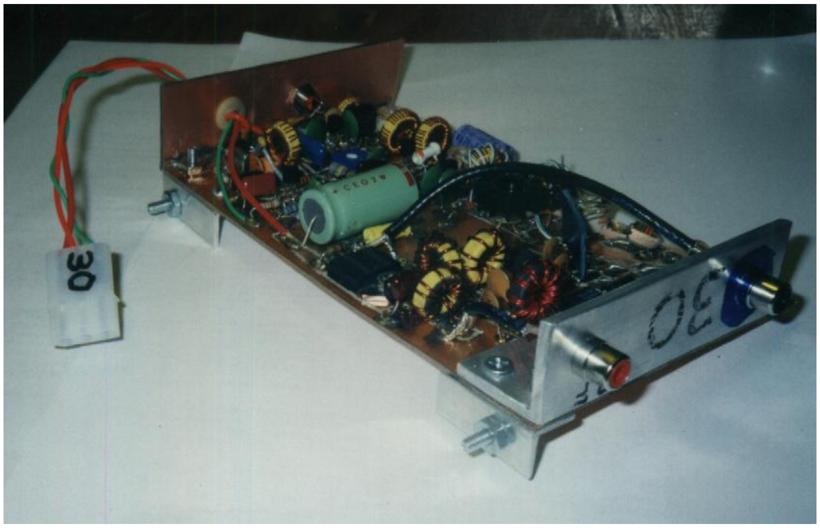

A QRP CW module for 30 meters is shown above. In many ways this module resembles the crystal controlled QRP module described in chapter 6. However, it uses a crystal-controlled premix oscillator (PMO) to convert the low frequency VFO sinewave up to the desired ham band. This particular module receives an 80 meter VFO signal and converts it to cover the 30 meter hamband, 10.100 to 10.150 MHz. The VFO signal and DC power come in at the back. The 5 watts of RF output exits from the red phono jack at the left front on the heat sink. The telegraph key port marked with blue is at the right. The Chebyshev output filters are visible in the foreground. The PMO filter amplifiers are at the rear. It would be more professional if it were enclosed in a metal shield, but I like to see all the components. With a CW QRP, you can get away with not using shielding. Later you will find that shielding is essential for SSB.

You can't multiply frequency anymore

In the old days, it was customary to build a VFO for 1.8 to 2.0 MHz or 3.5 to 4.0 MHz. Then for higher frequencies, we ran the signal through successive frequency multiplier amplifiers to get 7, 14, 21, and 28 MHz. A frequency multiplier was simply an amplifier tuned to the second or third harmonic of the input frequency. By using an amplifier tuned to multiples of the base frequency, the desired harmonic could be selected. For example, the tapped coil-tuned amplifiers described in chapter 6 work well for this purpose.

If your driving oscillator is crystal-controlled, then frequency multiplication is still practical. However, if your VFO drifts more than about 2 Hz /minute, you may get complaints on the upper bands. For example, if you have an 80 meter VFO you will have to multiply the frequency eight times to raise the frequency up to 28 MHz giving a drift of 16 Hz. But, if your VFO drifts 5 Hz, then the multiplied signal will drift 40 Hz at 28 MHz.

Fortunately, carefully built high frequency crystal oscillators can be quite stable even up to 30 MHz. The solution to the drift problem is to "add" a low frequency VFO to a stable high frequency crystal oscillator. These crystal oscillators are called ***Pre-Mix Oscillators*** or ***PMOs***. A ***mixer*** performs the frequency addition by literally combining two sinewave signals. The composite signal contains, not only the original signals, but also signals with frequencies that are the sum and difference of the original frequencies. Filters follow the mixer to extract and amplify the desired frequency component. The process is illustrated by the block diagram of a 20 meter QRP transmitter shown below:

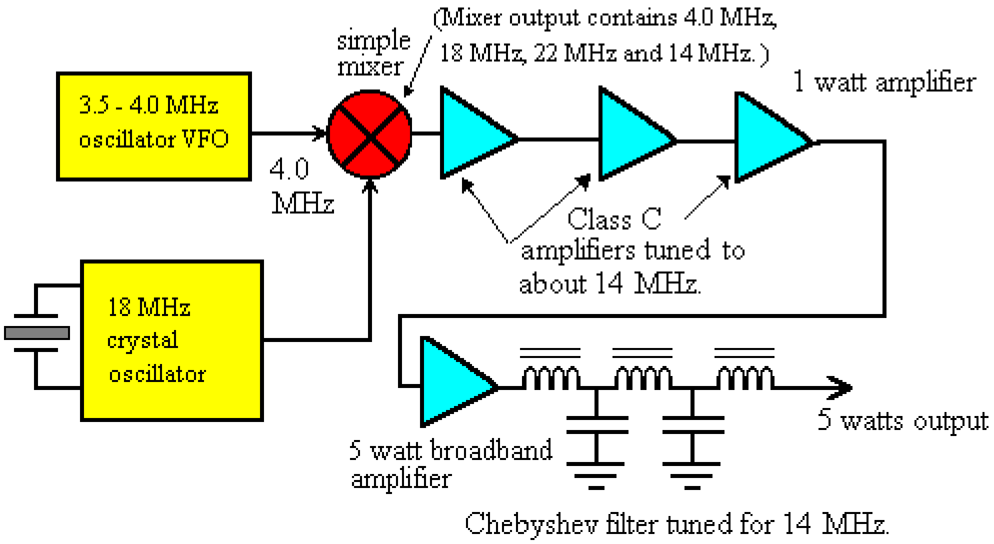
(Mixer output contains 4.0 MHz, 18 MHz, 22 MHz and 14 MHz.)
simple mixer
3.5 - 4.0 MHz oscillator VFO
4.0 MHz
1 watt amplifier
Class C amplifiers tuned to about 14 MHz.
18 MHz crystal oscillator
5 watts output
5 watt broadband amplifier
Chebyshev filter tuned for 14 MHz.

20 Meter 5 Watt QRP Transmitter Block Diagram

PreMix Oscillator method of frequency translation

In the diagram above, an 80 meter VFO is "converted" to 20 meters. The 80 meter sinewave is mixed with the output from an 18.000 MHz crystal oscillator. When the VFO is set to 4.0 MHz, the output from the mixer is a messy looking waveform that contains several frequencies, namely - 4.0 MHz, 18 MHz, 22 MHz, and 14 MHz. By tuning the next three amplifier stages to 14 MHz, the "contamination" goes away and we get a pure 14.0 MHz sinewave tunable up to 14.5 MHz. The crystal oscillator may contribute a Hz or two of drift, but basically, the drift on 20 Meters is the same as it was on 80 Meters. The mixer is comparable in function to those used in superhetrodyne receivers, but PMO mixers are much less critical. Low noise and extreme image canceling aren't necessary because both input signals can be as large as you like.

****

CRYSTAL OSCILLATORS ARE STABLE, AREN'T THEY?

A few years ago I thought I had the VFO problem conquered. I had just enjoyed a nine-month "VFO vacation." During this time my signal was so stable, that no one ever commented on it. I was quite proud of myself. Then I built QRP modules to get on 17 and 30 meters. Suddenly the complaints started up again and I was mystified. After all, I was using the same VFO. What had changed?

I checked out my VFO. I discovered that, when it was cold, it drifted downward 20 Hz the first minute. Then, after a few more minutes, it stabilized and the drift was plus or minus a 2 or 3 Hertz /minute. Of course, by definition, whenever I start sending the VFO is cold. Therefore, unless I send for minutes on end, it must always be drifting. But even so, that didn't explain the 100 Hertz complaints.

Surely, it couldn't be the frequency converter crystal oscillators, could it? Crystals drifting!!?? Blasphemy! I checked out my 17 and 30 meter crystal oscillators. The 30 meter oscillator drifted downward 50 Hz the first minute, 25 Hz the 2nd minute and eventually stabilized 150 Hz below the starting frequency.

Use HC-49 or larger crystals

The problem with my 30 meter converter turned out to be the crystal. The crystal was a teeny-weeny can, about 1/4" square and a 1/32" thick. It had come from my junk box and I don't know what size number it was. However, I've since concluded that all the tiny ones in my collection aren't as stable as the HC-49 or larger crystals. Little bitty overtone crystals are particularly bad. Yes, eventually they settle down and become reasonably stable. But by then you have switched the QSO over to the other guy. Now your oscillator is cooling off again so it will be ready to drift during your next transmission.

Not all tiny crystals are bad. I have some little half-size 9.00 MHz HC-49 crystals that work extremely well in my receiver BFO and IF filters. I guess the lesson is that you need to check out your oscillator stability during that first critical minute. The drift after 5 minutes is interesting, but not very important for a ham transmitter.

TTL oscillator blocks

The 17 meter board had one of those sealed-in-a-can, TTL oscillator blocks. They're like an integrated circuit with the crystal and oscillator sealed in the same package. I used one because it was the right frequency and I happened to have it in my famous junk collection. Yes, I remembered to run it at a correct 5 volts TTL voltage, not 12 volts. My oscillator started out at the right frequency, but it ran surprisingly hot. Then it dropped at 25 Hz per minute. Although the drift slowed, the frequency never stopped sinking. I happen to have a big bag of various frequency oscillator blocks and they all did that. Every one I tried was terrible! All except the really high frequency ones - like 50 or 100 MHz. ... Those were really terrible. Some moved as much as 500 Hertz a minute. The only good news is that they're consistent. They all drift downward. It did not occur to me to try running the oscillator blocks at lower voltage, say 3 volts or lower.

Solutions to crystal drift

I could leave the crystal oscillators running continually. That might work with the ones that stabilize, but then I might be forced to listen to a harmonic of the oscillator signal in the receiver. No, thanks. I already have a couple whistling artifacts in my receiver. Notice that vacuum tube oscillators might actually have an advantage in the warm-up problem. Since the tube filaments run continuously, a vacuum tube oscillator is always hot and a few milliamperes of plate current aren't going to change its temperature much. In the old days, temperature controlled ovens were used to hold crystal oscillators at constant temperature. I don't know how you feel about it, but that's too drastic for me.

All oscillator circuits are not equal

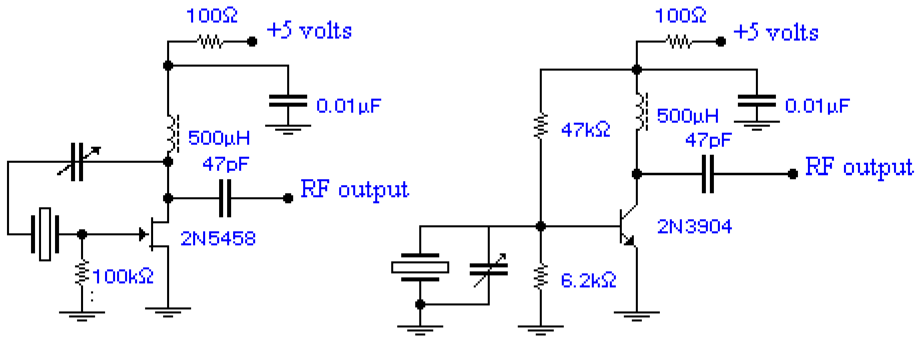

Two common crystal oscillator circuits

The drawing above shows two common crystal oscillator circuits that I used in some of my first QRP PMO converters. The variable capacitors are used to trim the frequencies to the exact Hertz. ***Both oscillators have the crystal connected to the base or gate.*** (Remember that feature and you'll know which oscillators to be wary of.) In my experience, these base-connected crystal oscillator circuits drift downward when first turned on. Finally, after a minute or two, they stabilize.

Plan your frequency converters so the drifts cancel.

Suddenly I understood why I received no complaints while using my 40-, 20-, and 15-meter crystal-controlled frequency converters. They used oscillators with the crystals connected to the bases like those above. However, the crystal frequencies were 4 MHz **above** the target band. As the oscillators drifted downward for the first minute, typically at 20 Hertz per minute, my 4 MHz VFO was also drifting downward at the same rate. For example, (25 MHz - 20 Hz crystal drift) minus (4 MHz - 20 Hz VFO drift) = 21.000,000 MHz. The result was a relatively constant frequency and no complaints. After a couple minutes the drifting stopped and the crystal oscillators were slightly more stable than the VFO.

Notice that if these converter crystal oscillators had been **below** the target hamband, then the drifts would have been added instead of subtracted. It turns out that my receiver was designed this way. Oops! Oh well, no ham has ever complained about my receiver during a QSO. And of course receiver oscillators can run non-stop, so initial drift isn't a big deal.

Butler is better

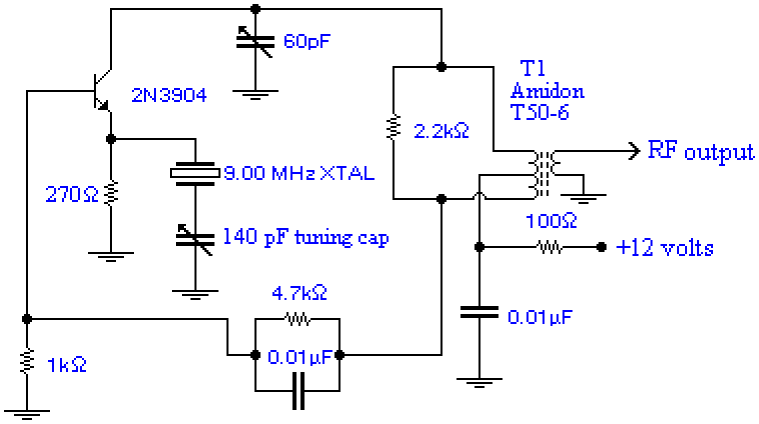
60pF
2N3904
T1
Amidon
T50-6
2.2kΩ
RF output
270Ω
9.00 MHz XTAL
140 pF tuning cap
100Ω
+12 volts
4.7kΩ
0.01µF
1kΩ
0.01µF

A Butler crystal oscillator

I checked out every oscillator in my rig and discovered that some of them didn't drift when turned on. The stable ones used the Butler crystal oscillator above. Notice that ***the crystal and its capacitor are in parallel with the emitter resistor.*** The equivalent FET oscillator circuit works as well and maybe better. I don't really know why, but this circuit is stable the moment you turn it on. I suppose it's because the crystal has little voltage across it and receives little power to heat it. Anyway, Butlers typically drift no more than a hertz or two per minute. Two of my oscillators showed zero Hertz drift during the first minute. This is the same oscillator I recommended in chapter 6. Depending on your application, Butler also has the advantage that the series capacitor can pull the frequency lower than the base-connected oscillators above.

A VFO controlled QRP

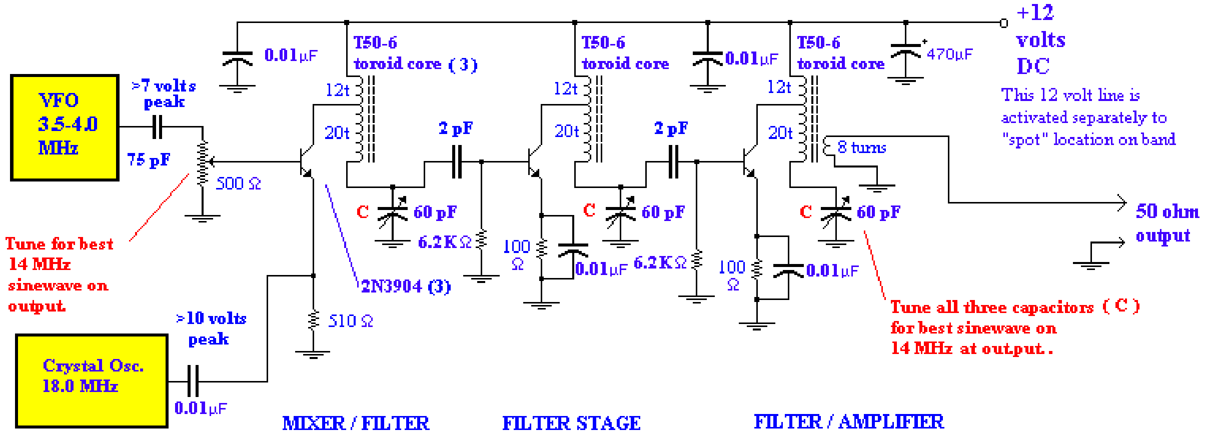
VFO
3.5-4.0
MHz
>7 volts
peak
75 pF
500 Ω
Tune for best
14 MHz
sinewave on
output.
0.01µF
T50-6
toroid core (3)
12t
20t
2N3904 (3)
510 Ω
>10 volts
peak
Crystal Osc.
18.0 MHz
0.01µF
C
60 pF
6.2KΩ
2 pF
100
Ω
0.01µF
T50-6
toroid core
12t
20t
C
60 pF
6.2KΩ
2 pF
100
Ω
0.01µF
T50-6
toroid core
12t
20t
8 turns
C
60 pF
470µF
+12
volts
DC
This 12 volt line is
activated separately to
"spot" location on band
50 ohm
output
Tune all three capacitors (C)
for best sinewave on
14 MHz at output.
MIXER / FILTER
FILTER STAGE
FILTER / AMPLIFIER

QRP DRIVER FOR 20 METER CW (14 MHz)

My "standard design" for a CW QRP driver is shown above. Unfortunately each QRP driver covers just one band. However, once it's tuned and working, it covers the entire band without further tuning or fussing. I should mention that the same filter train can be designed so that it could be tuned up for several different bands, for example 20 meters through 10 meters. But of course, changing bands would mean changing the crystal oscillator and retuning the entire chain for the new band – not exactly convenient band switching.

So far, I've built seven versions of this design covering 40 through 10 meters. Taken as a whole, the entire circuit shown above can be considered as a 14 MHz VFO. In other words, it took me all those parts to generate a stable sinewave on 14 MHz. A simple crystal oscillator using 14 MHz crystals generates the same result, but of course it will only tune a few KHz at best. Life is hard for us homebuilders in the 21st century.

So why didn't I build one driver that works on every band? Back in the vacuum tube days, that was easy to do. However, the spectral purity and stability of our signals back then was awful. Also tubes are much easier to work with. If you're a lightly-equipped tinkerer like me, you'll find that getting even one band working to modern standards is much harder than it looks. I suspect that's why hardly anyone does this kind of scratch-built homebrewing anymore. I strongly recommend you ***start simple***.

Changing the direction of tuning

In the 20 meter QRP driver above, the 80 meter VFO is mixed with an 18 MHz local oscillator. Note that the oscillator could also run at 10.5 MHz. As an experiment I ran my 20-meter QRP with both 18 MHz or 10.5 MHz crystals. All I had to do was change the crystal and it worked well. The 14 MHz filter remained tuned up without any adjustment. The operational difference is that ***the direction of the VFO tuning reverses.*** As explained in the last chapter, this can be useful if you are using a varactor VFO and need to have the high frequency end of the VFO range tuning the lower end of a ham band.

The mixer needs a big local oscillator drive signal

The mixer stage in the above converter is just an RF amplifier made out of a bipolar 2N3904 transistor, much like the amplifiers in the filter train. This mixer is a class C amplifier stage with a 500 ohm emitter resistor. We can use class C because the input signals are much larger than 0.6 volts. One input, usually the VFO, feeds into the transistor base in the usual fashion. However, this drive level is adjustable with the input pot.

The local oscillator input is applied across the emitter resistor. I usually inject the higher frequency input across the resistor, but I've done it both ways. Unlike the input to the base, the input to the resistor has no amplification gain whatsoever. In order for the emitter signal to produce a big signal on the collector, all of the signal amplitude must be impressed onto the 500 ohm resistor.

What I learned the hard way is that, ***the local oscillator drive must be strong enough to turn the mixer stage full on and off like a switch, cycle by cycle.*** I use an input sinewave of at least 20 volts peak-to-peak. A wimpy little 2 volt crystal oscillator signal will produce little difference-frequency component in the output and it will take many stages of filtering to extract the desired frequency. To get 20 volts pk-to-pk signal drive, I had to amplify the output of the crystal through an amplifier stage before it went into the mixer. I threw out two boards before I figured this out. (I'm a slow learner.)

On the other hand, the second input signal, the VFO, can be small because it is amplified by the transistor. Later, when you're tuning up the whole filter/ amplifier string for best output, you'll find that the maximum output and purity occurs at a specific setting of the input pot. The optimum VFO input level is not simply maximum input.

The LC filter/ tank circuit on the mixer collector is tuned to the desired sum or difference frequency. Using the formulas in your CWS (Amidon) T50-6 core literature, calculate the inductance needed to go with your trimmer cap to resonate at the desired band, just like we did back in chapter 6. I found that T37 cores were too small and didn't produce the gain per stage that I got out of the T50s. In contrast, the T68 cores were unnecessarily large.

Dual gate MOSFET mixers

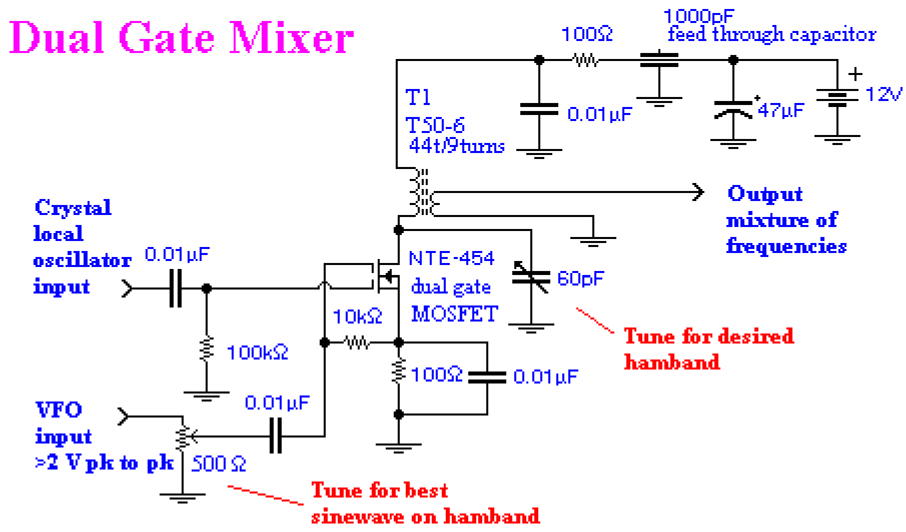
Dual Gate Mixer
100Ω
1000pF
feed through capacitor
+
12V
T1
T50-6
44t/9turns
0.01µF
47µF
Output
mixture of
frequencies
Crystal
local
oscillator
input
0.01µF
NTE-454
dual gate
MOSFET
60pF
10kΩ
Tune for desired
hamband
100kΩ
100Ω
0.01µF
VFO
input
>2 V pk to pk
0.01µF
500 Ω
Tune for best
sinewave on hamband

The principal advantage of the bipolar transistor mixer shown earlier is that ***it's cheap***. I began using dual-gate MOSFET transistor mixer stages in a receiver project and found them superior in several ways. A dual gate MOSFET is a small RF transistor with *TWO* input gates. Otherwise, in principle a dual gate works just like the power MOSFETs described in chapter 6. Since both gates have lots of voltage gain, small signals may be used on both inputs. I found that each gate only needs 2 volts peak-to-peak and the output is much easier to tune and filter. Unfortunately dual-gate MOSFETs cost at least $7 each compared with 20 cents. This saves me some complexity and my later converters usually used the pricey mixer. I have used NTE-221, NTE-222 and NTE454. This application is quite uncritical and I believe any dual gate will work well. You will find that mixers in superhetrodyne receivers are not so uncritical.

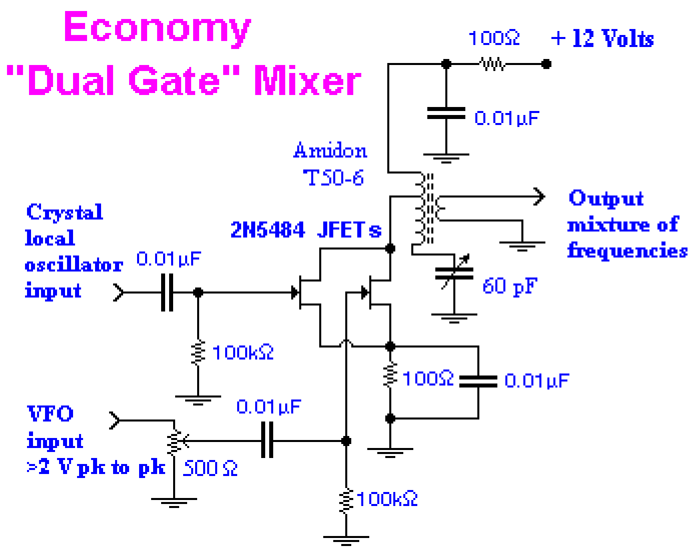
Economy
"Dual Gate" Mixer
100Ω
+ 12 Volts
0.01µF
Amidon
T50-6
Output
mixture of
frequencies
Crystal
local
oscillator
input
0.01µF
2N5484 JFETs
60 pF
100kΩ
100Ω
0.01µF
VFO
input
>2 V pk to pk
0.01µF
500 Ω
100kΩ

The economy dual gate mixer

The "dual gate mixer" shown above is made from two paralleled JFETs. Two JFETs are about a tenth the price of the MOSFET and solve the cost and availability problem. Like the dual gate MOSFET, this circuit has the advantage that both inputs have gain. The two circuits are interchangeable for this PMO application. Notice that the VFO input goes to a 500 ohm pot and may be adjusted.

If you are really miserly, you can use the same trick with parallel bipolar transistors. If both input signals are small, both transistors will need forward bias, the 33K resistors. If one of the inputs is large enough, say 5 volts peak or larger, you won't need the forward bias for that input. This circuit should be plenty sensitive for a receiver mixer. However, because it has PN junctions, it will be noisier than the dual–gate MOSFET mixer and I don't recommend it for receivers.

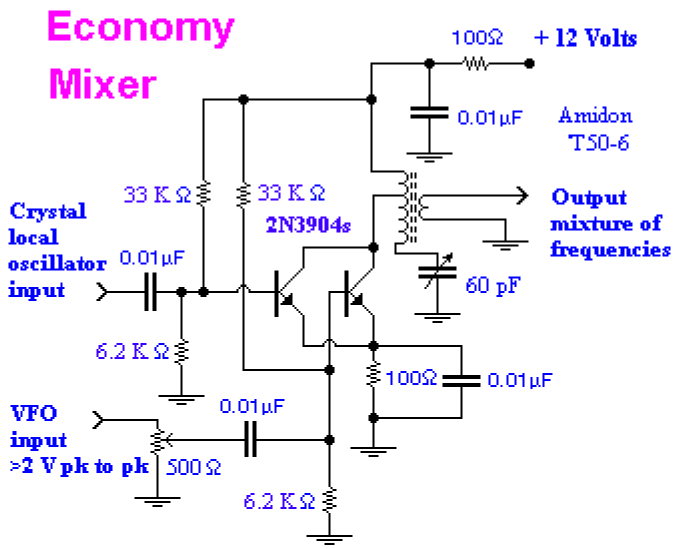

Tuning the mixer

When you first apply the two input frequencies to the mixer stage, the scope will show a messy, complicated waveform on the collector (or drain). It will be impossible see what setting of the trimmer capacitor on the drain is best. The first step is to disconnect the input frequency that is farthest from the desired output frequency. Now it will be easy to tune the trimmer for maximum gain. For example, in the 20 meter filter train above, tune the first mixer/ amplifier to the crystal frequency of 18 MHz. When it's peaked at 18 MHz, look to see if the trimmer is at or near maximum or minimum capacitance. If it is near minimum capacitance, but not quite there, the mixer stage should tune up well at the lower frequency, 14 MHz. If the gain at 18 MHz is maximum while the capacitance is tuned to either minimum or maximum, then you know that your toroid coil has too many or too few turns.

Later in the receiver chapter (13) and the sideband chapter (15) there are examples of un-tuned, broadband mixers that could also be used. Broadband mixers don't need tuning and they are less likely to self-oscillate. However, they have less gain and you may need more stages to reach the same power level.

Filtering the desired frequency from the mixture

At the mixer drain (or collector) there are four frequency components and you must filter the one you want into a pure sinewave. I first considered using complex Chebyshev filters as described in the Handbook. I was discouraged by the many inductors and capacitors needed to achieve the desired "pure" output. Eventually I realized that Chebyshev filters have the virtue that they can be designed for arbitrarily low impedances, that is, high power signals. In the PMO application here the signal is extremely high impedance and some relatively simple parallel LC filters are adequate. Chebyshevs are not needed. This approach is illustrated in the SSB transmitter in chapter 15.

My CW transmitter uses two sharply tuned amplifiers as filters, just like an IF amplifier in a receiver. The ease of filtering depends on how far the desired frequency is from the local oscillator and other mixer products. For example, on 14 MHz, the 4.0 MHz VFO is 28% of the desired frequency. 14 MHz compared with the 18 MHz crystal oscillator is 77% of the desired frequency. This is pretty close but not a problem. Now suppose we use a 32 MHz crystal on 10 meters (that is, 28 MHz). With a 4.0 MHz VFO, the desired frequency is 88% of the crystal frequency. You will find that tuning this is much more “tweaky,” but still practical. In general, having the crystal oscillator *BELOW* the desired frequency makes tuning the filters easier.

In summary, the tuned amplifier approach is reliable for CW and is extremely hi-Q. It can separate frequencies that are quite close. However, it is not suitable for use in an SSB transmitter because the hi-Q amplifiers tend to self-oscillate during pauses in the speech signal.

Bandpass "filter amplifiers"

Each amplifier stage is essentially like the bipolar transistor mixer shown earlier. However, the emitter resistor is bypassed with a capacitor so that, from the point of view of the RF, the emitter is connected to ground. The purpose of the RC in series with the emitter is to stabilize the gain and reduce the DC current drawn by the stage. You may use either class A or class C amplifiers. I often use class A amplifiers, meaning the stage is biased “on” at all times with a 33K resistor, just as we did back in chapter 6. Class A amps draw more current than class C amplifiers that are basically the same circuit. However, they handle wave components of any amplitude. Or to say it another way, class As work over a wider range of input amplitudes and don’t introduce harmonics that must be filtered out.

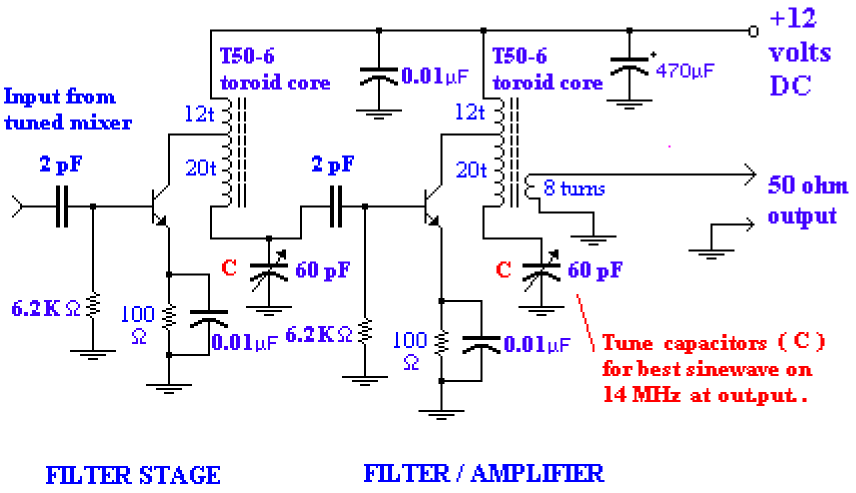
+12 volts DC
T50-6 toroid core
0.01µF
T50-6 toroid core
470µF
Input from tuned mixer
12t
12t
2 pF
20t
2 pF
20t
8 turns
50 ohm output
C
60 pF
C
60 pF
6.2KΩ
100 Ω
0.01µF
6.2KΩ
100 Ω
0.01µF
Tune capacitors (C) for best sinewave on 14 MHz at output.
FILTER STAGE
FILTER / AMPLIFIER

FILTER FOR 20 METER CW (14 MHz)

An RF filter/ amplifier stage

Two stages of filter amplifiers are shown above. Including the tuned mixer stage, three stages of tuned amplifier were enough for any HF band using an 80 meter VFO. However, as explained earlier, when you are trying to separate two frequencies that are only 10% different, using only 3 filter stages is just barely practical. If you are having trouble getting a pure sinewave, just add another tuned stage. As shown above, the above stages are class C. If desired, you could forward bias these amplifiers on with 33K ohm resistors and convert them to class A amplifiers. In this way, they could handle smaller signal levels.

Use tiny interstage coupling capacitors

The BIG SECRET in making "filter amplifiers" work is to ***use tiny coupling capacitors between amplifier stages.*** Notice the 2 PF capacitors between stages in the diagram above. The purpose of these amplifier stages is ***filtering, not power gain.*** The LC circuit "rings like a bell" when the input contains a frequency that resonates with the LC. This ringing exaggerates the desired frequency component. If you load the LC circuit by trying to couple significant power to the next stage, it's like putting your hand on a ringing bell - the ringing will be damped and the filter effect dies. ***To avoid damping the ringing, use tiny 2 pF capacitors.*** OK, on 80 Meters maybe 5 pF is acceptable. And on 10 meters 1 pF would be better. However, 2 pF works over the whole HF spectrum. Notice that if you use large coupling capacitors, say 50 pF, that 50 pF becomes part of the LC resonance and will dominate the tuning and lower the Q. Also remember that your scope probe contributes another 5 pF or so. To make a final adjustment of a filter stage, you must put the probe on the output of the stage ***following*** the one you are adjusting.

With one stage of filtering after the mixer, the waveform will still look "messy" on the scope. But after two stages of filtering it should be possible for your frequency counter to lock onto the correct frequency. As you tune the VFO, the reading on the counter should track solidly with no drifting and dancing digits. When properly tuned, the sinewave will look nearly perfect on the oscilloscope after two stages of filtering. When you first try to tune up all three stages at once, you may be frustrated, but keep trying. When your counter "locks" onto the desired band, look at the ringing on the trimmer of the last stage with the scope while tweaking all the previous stages for the best sinewave. Notice that perfection occurs when you trim the input level of the VFO on R1. Now you see why the input is applied through a trim pot.

Where to get crystals for your local oscillators

Yes, you need a separate crystal for each band. Fortunately, when using an 80 meter VFO, standard microprocessor frequencies can cover the major ham bands. (e.g., 11 MHz, 18 MHz, 25 MHz, and 32 MHz cover 40, 20, 15, and 10 meters, respectively.) Mouser Electronics and Digi-Key sell these for about $1 each. For WARC bands and 160 Meters you may have to spend some money or be creative. As explained earlier, don't use those TTL oscillator blocks. The cheap ceramic resonator "crystals" are also a poor idea. Significant drift is not worth saving a few dollars.

The QRP power amplifier stages

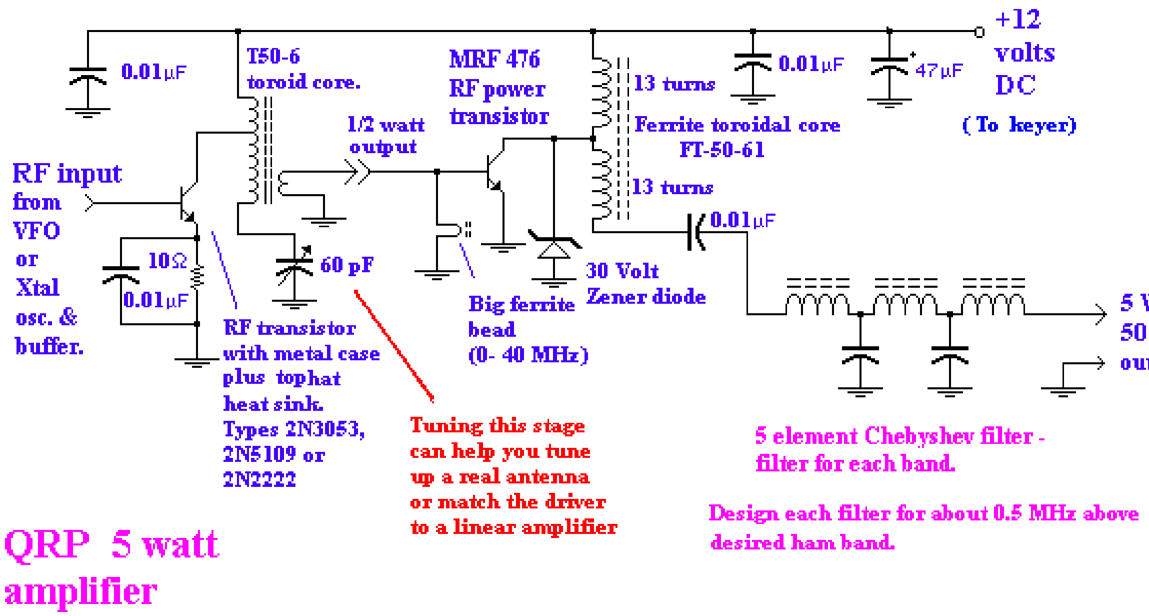
0.01µF
T50-6
toroid core.
MRF 476
RF power
transistor
13 turns
0.01µF
47µF
+12
volts
DC
(To keyer)
1/2 watt
output
Ferrite toroidal core
FT-50-61
13 turns
0.01µF
RF input
from
VFO
or
Xtal
osc. &
buffer.
10Ω
0.01µF
60 pF
30 Volt
Zener diode
Big ferrite
bead
(0- 40 MHz)
5 Watt,
50 ohm
output
RF transistor
with metal case
plus tophat
heat sink.
Types 2N3053,
2N5109 or
2N2222
Tuning this stage
can help you tune
up a real antenna
or match the driver
to a linear amplifier
5 element Chebyshev filter -
filter for each band.
Design each filter for about 0.5 MHz above
desired ham band.
QRP 5 watt
amplifier

Your VFO now tunes the HF band of your choice. To increase the gain of this signal up to 3 to 5 watts, you will need two or three stages of power gain as described earlier in chapter 6. My QRP boards use two power amplifier stages. The first is a tuned stage. The second is a broadband amplifier followed by a Chebyshev lowpass filter designed for 50 ohms. This scheme seems to combine the advantages of both systems. For example, suppose I connect the QRP output to a non-inductive 50 ohm dummy load. All seven QRP boards I have built have no trouble delivering a clean sinewave into a dummy load. That is, tuning is easy until you have to connect it to a real antenna or to a final amplifier.

Tuning the QRP output to an antenna or amplifier

Suppose after tuning it up with a dummy load, I connect the QRP to a high power final amplifier or to an antenna tuner. Suddenly I discover that the QRP output is badly distorted. The output stage broadband amplifier may even go into “noise mode.” If I had designed both stages as broadband, I wouldn’t have anything to adjust. Strange as it seems, tweaking a tuned stage will usually match up the broadband output stage to my final amplifier. In general, the lower the frequency, the easier it is to match stages and the antenna. Matching 80 and 40 meters is as easy as falling off a log. 10 meters is tough and I still haven’t gotten my linear final (described in chapter 12) to put out more than about 20 watts on that band. Needless to say, I am in awe of guys who homebrew transistorized UHF rigs.

In summary, when you build an amplifier input that is nominally designed for “50 ohms resistive,” you may find that it has lots of reactance (inadvertent capacitance and/ or inductance) and is quite different than planned. Notice that Chebyshev filters are designed for specific input and output impedances. In other words, filters don’t filter when they are mismatched.

Spotting with a VFO

Now suppose you are on an upper band using your new VFO driven QRP. You hear a station calling CQ. But if you key your transmitter and adjust the VFO to the same frequency as the station, you will be sweeping the band with your transmitter and making yourself unpopular. Besides, you have probably wired the antenna and power supplies so that the receiver shuts off while transmitting. In order to hear your own QRP in the receiver, you need to turn on those parts of the QRP that generate the desired frequency, but not the power amplifiers. You also want the antenna to remain connected to the receiver and you don't want the receiver to shut down. To hear the VFO we need to direct power from the "Spot" switch to the VFO power supply. But we also need to turn on the QRP PMO sections that generate the desired frequency.

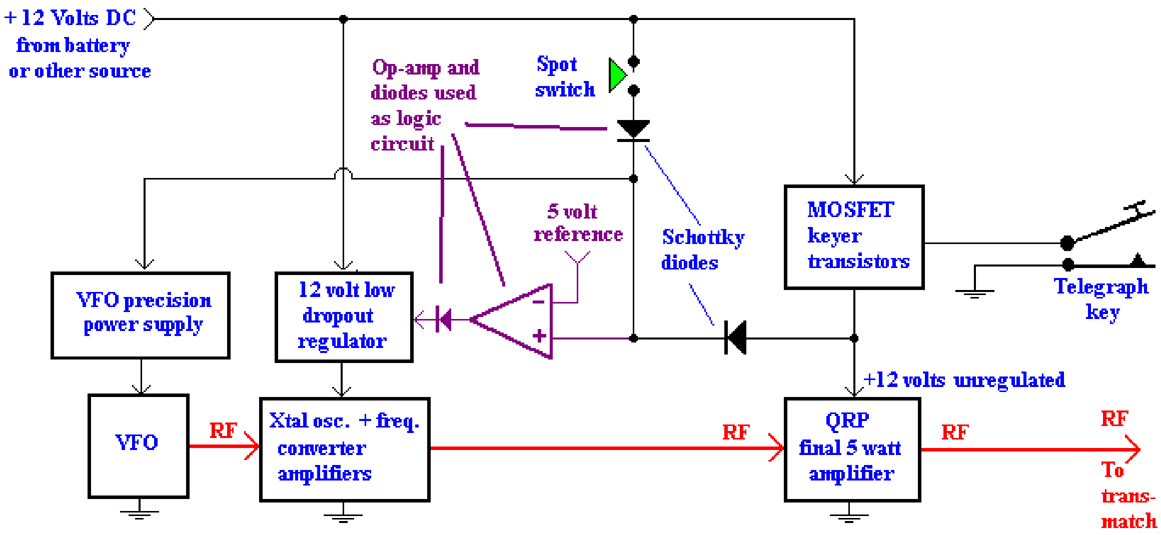
+12 Volts DC from battery or other source
Op-amp and diodes used as logic circuit
Spot switch
5 volt reference
Schottky diodes
MOSFET keyer transistors
Telegraph key
VFO precision power supply
12 volt low dropout regulator
VFO
RF
Xtal osc. + freq. converter amplifiers
RF
+12 volts unregulated
QRP final 5 watt amplifier
RF
RF
To trans-match

QRP KEYER WITH "SPOT" BUTTON

If you are running your transmitter on battery power as I do, then the low dropout supply only has to run the crystal oscillator and the filter/amplifiers. The power amplifier (or amplifiers) can run directly off the battery and that unregulated 12 volt line can be activated by the MOSFET keyer circuit. Now when you push the spot switch, the VFO and the low power sections of QRP will turn on, without shutting down the receiver. Below is the low dropout regulator with the turn-ON control line from the Spot diodes.

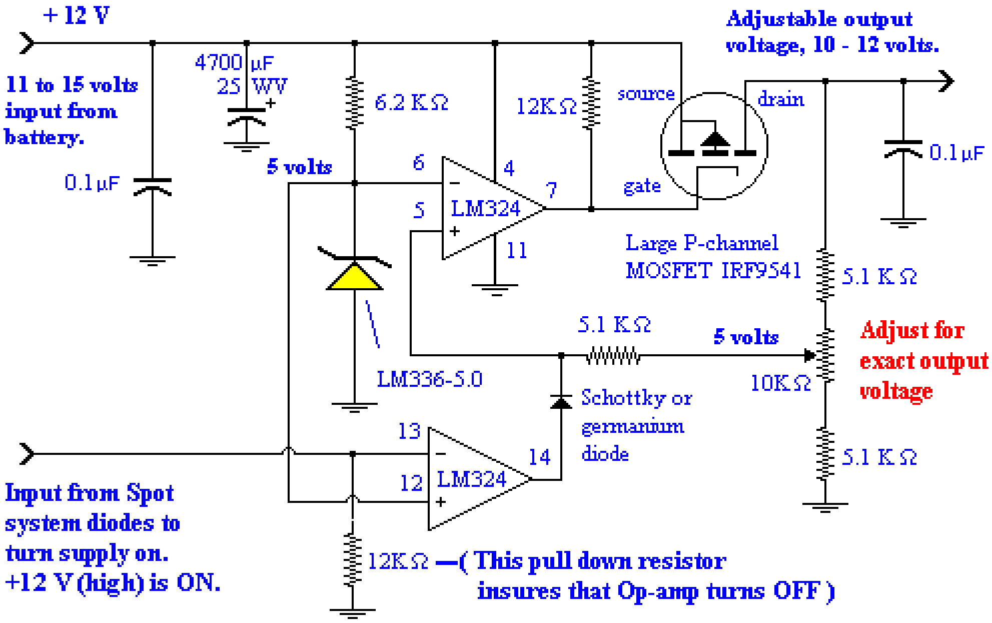
+ 12 V
11 to 15 volts input from battery.
0.1 µF
4700 µF
25 WV
+
6.2 KΩ
5 volts
6
4
7
5
LM324
11
12KΩ
source
drain
gate
Adjustable output voltage, 10 - 12 volts.
0.1 µF
Large P-channel MOSFET IRF9541
5.1 KΩ
LM336-5.0
5.1 KΩ
5 volts
Adjust for exact output voltage
10KΩ
Schottky or germanium diode
13
14
12
LM324
5.1 KΩ
Input from Spot system diodes to turn supply on. +12 V (high) is ON.
12KΩ —(This pull down resistor insures that Op-amp turns OFF)

Low Dropout 12 Volt Power Supply with Remote

When I began trying to get on the air with my first CW QRP, my goal was simply to reach 15 meters. This chapter summarizes what I learned on that first journey. Since one band was plenty of challenge, I didn't worry about expanding the QRP to cover multiple bands. After a while I began to wish I could operate on 40 meters, so I just built another version of the same QRP design. Because soldering is such fun, I eventually cloned the unit for every HF CW band. This was grossly inefficient and, in retrospect, is a bit embarrassing, especially since I included a separate keyer MOSFET on every band module. Surely you can improve on that!

If you are an experienced builder and want make a single multi-band QRP, then go look at the Single Side Band (SSB) QRP modules in chapter 15. In that system the amplifiers are wide-band/ linear and made separate from the band filters. If everything works as planned, arrays of filters for each band can be switched in with only one set of linear amplifiers. In the CW PMOs described in this chapter, the filters are tuned amplifiers that are easier to make work, but much less versatile. Beginners beware. Multi-band units are like flying submarines – A vehicle that does both is almost certainly a poor airplane and a lousy submarine. It is very difficult to build one machine that performs multiple functions well.

CRYSTAL SETS TO SIDEBAND
© Frank W. Harris 2010, REV 12

Chapter 12

SIMPLIFIED QRO AMPLIFIER DESIGNS

When I first got back on the air as a retiree, I built a QRP that put out 4 watts on 15 meters. I spent two days answering CQs and calling CQ. Unfortunately no one heard me. I came to the (incorrect) conclusion that QRP is a hobby for guys with expensive, huge beam antennas, not guys with verticals and dipoles. Without that extra 10 dB of gain, I figured my QRP signal must be down in the noise. The other way to get 10 dB of gain is an amplifier. Transmitting more than 5 watts is called ***QRO***. In this chapter I shall describe my efforts to build a good linear amplifier. Having QRO power when I need it has made my transmitter a reliable communications system. With 50 or more watts, your contacts won't have to work so hard to hear you and rag-chewing becomes practical.

Now that I'm older and wiser, I realize that low power and simple antennas weren't my biggest problem. In the old days our receivers had passbands several KHz wide and most beginners were stuck with a couple crystals for each band. After we called CQ, we tuned up and down the entire band looking for replies. In contrast, today guys with modern receivers are usually just listening to a few hundred Hz. When I put my 4 watts on 15 meters, I didn't realize that using the upper sideband is the standard convention for 20 meters and higher. On 15 meters the other stations were tuned to the upper sideband while I was often zeroed in on the lower sideband. My old (1967) homebrew receiver was so wideband, it wasn't obvious to me which sideband I was on. When I answered those CQs, I was usually off their frequencies by about 1.4 KHz.

In this chapter I describe three different final amplifier designs that I used successfully on the air. I don't recommend building the first one. I describe it here because it was educational. It was a ***tuned class B amplifier***. It worked, I learned from it, and it covers 20 through 10 meters. Unfortunately, it was too hard to tune. Moreover, you would probably have a hard time finding a dual-section, butterfly variable capacitor like the one I used to tune the output to resonance.

The second amplifier is an ***untuned class B*** design. It works on all bands and is suitable for CW only. Someday when you graduate to single sideband phone you can upgrade this amplifier to the third design, which is an all band ***Class AB linear amplifier***. Rather than just read the descriptions of the finished products, you may gain some insights by reading my odyssey of how I worked my way up to a real linear amplifier.

The quest to build a 50 watt amplifier

I began my QRO project by searching my 1998 ARRL handbook for linear amplifier construction projects. I found three examples of linear amplifiers. One of them, "An HF 50 Watt Linear Amplifier," was a complex schematic that covered two pages. The other two examples were buried in diagrams of elaborate transceivers that seemed to be "illustrative" rather than something I was encouraged to build. I could almost hear a deep baritone voice saying, “For your own safety, do not build this at home.”

I studied the HF 50 watt linear amplifier project closely. It used a pair of MRF-477 transistors. I looked them up in my RF Parts Company catalog. It said, “call for pricing.” That was ominous. It turned out that a pair of them was $45. I also worried about all those feedback loops and clamps that protected the transistors from **overdrive**, from **excess SWR**, from **excess collector voltage,** and from **thermal runaway**. In addition, there were at least three kinds of frequency compensation feedback. *In short, the schematic seemed to be telling me that high power RF transistors are extremely fragile.*

The article gave me the impression that, if all those protection circuits didn’t work perfectly the first time I turned it on, my pricey transistors would turn into toast before I could say, “expletive deleted.” I had never built a high power transistorized RF power amplifier before. My only comparable previous projects had been 200 and 300 watt switching power supplies. Until I got those projects working, they had devoured numerous $20 transistors like popcorn. I was extremely wary of this project.

I retreated to my 1979 handbook and found a more primitive linear transistor amplifier project. This one also had thermal protection, but at least they implied that it didn’t have to work perfectly in concert with a flawlessly adjusted bi-directional power meter. The most reassuring feature was that MRF-454 transistors seemed to be the cheapest power transistors available, about $13 each. These transistors are big, rugged, and able to dissipate a great deal of heat. Just in case, I bought two extra sets of transistors.

At this point you may be thinking that, even at $13, those are expensive transistors! Why doesn’t he use a cheaper power transistor that has adequate power and frequency ratings? The answer is that the MRF-454 will produce high power with a 12 volt power supply. Sure, if you’re willing to build a 48 volt, 200 watt DC power supply, you can find dozens of really cheap transistors that will work well.

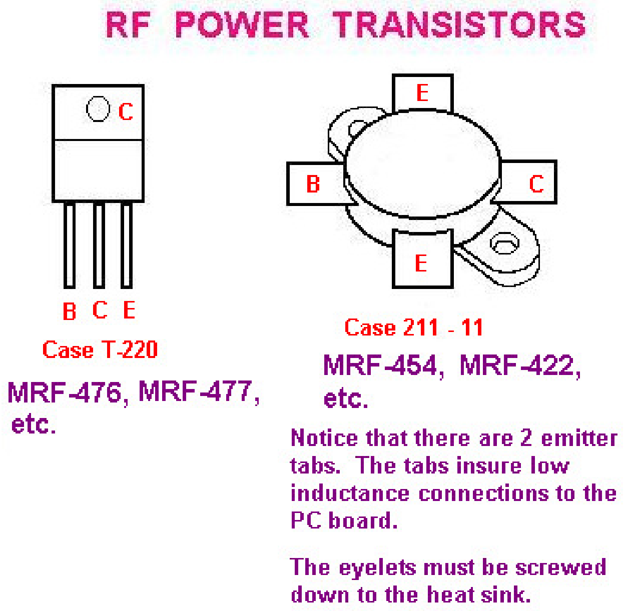

Notice that there are 2 emitter tabs. The tabs insure low inductance connections to the PC board.

The eyelets must be screwed down to the heat sink.

Basic features of a modern linear amplifier

A typical modern amateur radio linear final amplifier has six basic features:

First, ***Two separate power transistors are driven with a center-tapped transformer.*** The driver transformer is wired so that it turns on one transistor for half of the sine wave while the other transistor is turned off. During the next half cycle, the first transistor turns off while the second transistor turns on.

Second***, it is a class B design operating in "push pull."*** An advantage of class B is that, when there is no RF on the input, both transistors are nearly completely turned off. This means that they don't get hot and don't waste energy. Even when running with forward bias to make the amplifier linear, very little forward bias is needed and the efficiency approaches 50%. Also, class B tends to cancel out even harmonics.

Third***, the transistors must be cooled with a large heat sink.*** Large power transistors are designed to be bolted onto a heat sink. They have metal flanges with mounting holes for this purpose. For a 100 watt amplifier, the heat sink is typically a large, aluminum casting perhaps five or six inches on a side. Heat sinks usually have multiple cooling fins ¾ of an inch high**.**

Fourth, ***the output from the two transistors drives a second, center-tapped, untuned high inductance transformer.*** Since this output transformer is untuned, it can amplify nearly *ANY* RF signal over a wide range of frequency. So long as its circuit board is properly designed and the input signal is pure, the output will be a pure sinewave. The tricky part of this design is that, if any noise or "complex waveform" is introduced into the circuit, the amplifier may run away and produce wideband noise - more about this problem later.

Fifth, ***a push-pull linear amplifier is not really running "Class B" but rather it runs "Class AB."*** This means that a small amount of forward DC bias is injected into the bases of both transistors to turn them slightly on at all times. By having the transistors already turned on, they respond instantly when a tiny input signal appears on the bases. Without the bias, an input signal would have to exceed some threshold limit before the transistors could turn on. The advantage of matched pairs of transistors is that the forward DC bias for each transistor can be equal and low.

Sixth, ***a low pass output filter limits the frequency components in the output waveform.*** In other words, the filter suppresses harmonics so that, if you're transmitting on 40 meters, nobody will be able to hear you on 20, 15, or 10 meters. Each band you operate on needs a separate filter that clips off harmonics that would radiate at higher frequencies. You can get by with using the same filters for 12 and 10 meters and for 15 and 17 meters. I built my filters on little circuit boards that I plug into a card edge connector on the main board. I use several connector pins in parallel to keep the inductance of the connection as low as possible.

****

It looked easier in the handbook

When I began my work, I built the linear as close as I could to the drawings in the 1979 handbook, although as usual I had to substitute some parts. After I had carefully tested the forward base bias regulator circuit, I gingerly put 12 volts on the output transistors. Without any

RF input drive, the transistors immediately ran away and drew huge currents. Gee, something must be wrong with the bias circuit! I soon discovered that *ANY* forward bias caused the transistors to run away. Obviously, the guy who designed this amplifier used MRF-454s that behaved differently than mine.

Next I disconnected the bias circuit and powered it up again. In other words, I was hoping it would run as an untuned, push-pull Class B “sort-of-linear” amplifier. This time at least the transistors didn’t run away. I put an RF signal on the input and found that the amplifier was operating in what I call “noise mode.” As you know, linears are supposed to act like hi-fi amplifiers. They uncritically amplify whatever frequency signals you put in. If you put 80 meters or 10 meters in, you are supposed to get amplified 80 meters or 10 meters out.

Noise mode

An unfortunate mode of operation for a linear is “HF broadband noise generator.” I put in a clean, filtered 5 watt sinewave and I got a blast of wideband noise from the output that made my FM radio roar like a waterfall. Using an oscilloscope, the waveform across the dummy load looked like dancing grass in a tornado. My new amplifier had terrific power output into a dummy load. Unfortunately, little if any power was at the desired frequency.

Class Bs are easier

I concluded that my linear had at least two fundamental problems. I had already encountered “noise mode” while I was building my first 15 meter QRP. At that time I hadn’t found many cures for that disease, even at the milliwatt level. So I wasn’t optimistic about fixing it at the 50 watt level. I was tired of not having a working transmitter, so I decided to start over and build a simple class B *TUNED* amplifier. I was almost certain I could get that to work. Of course a tuned class B would only work on two or three bands without changing the output transformer and tuning capacitor. However, that was better than being off the air, possibly for months.

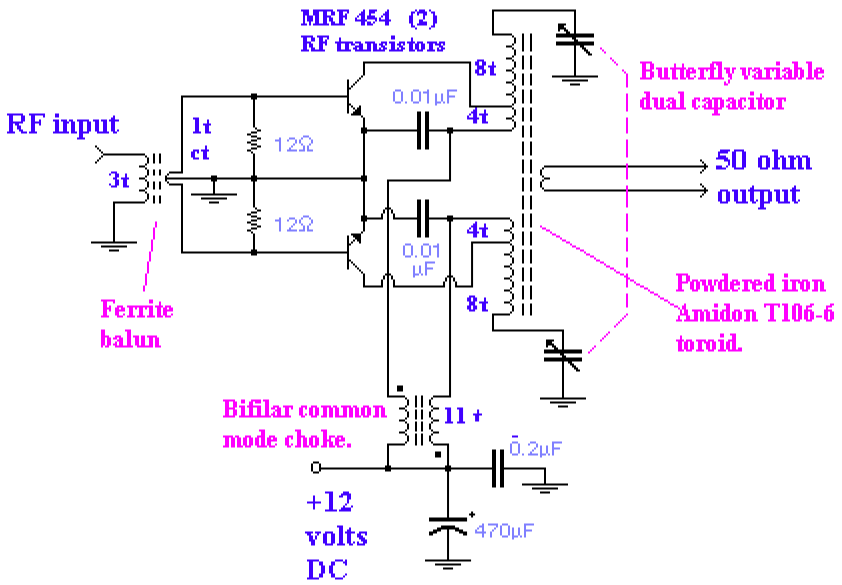
MRF 454 (2)
RF transistors
8t
0.01µF
4t
Butterfly variable
dual capacitor
RF input
1t
ct
12Ω
3t
12Ω
50 ohm
output
0.01
µF
4t
8t
Powdered iron
Amidon T106-6
toroid.
Ferrite
balun
Bifilar common
mode choke.
11 t
0.2µF
+12
volts
DC
470µF

Amplifier # 1. **A simple class B tuned amplifier**. This works, but I don't recommend it.

The tuned class B worked well. The only trouble I encountered was getting the input transformer to match properly and deliver the required big drive currents. After two unsuccessful attempts at winding powdered iron toroid input transformers, I tried the ferrite balun transformer from the linear amplifier. Success! Ferrite balun transformers really are different from powdered iron toroids. They match those low impedance power transistors when nothing else will. At least *SOMETHING* from the linear design worked.

Ferrite balun transformers

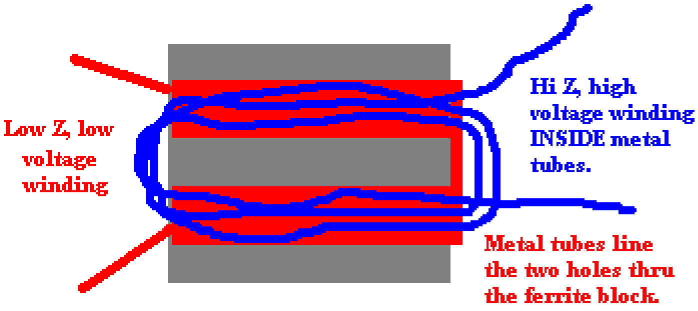
Low Z, low voltage winding
Hi Z, high voltage winding INSIDE metal tubes.
Metal tubes line the two holes thru the ferrite block.

BALUN TRANSFORMER

(Cut-away to show construction)

Think of the ferrite balun transformer as two large ferrite beads placed side by side. The beads are simply hollow cylinders made from high $\mathbf{A}_L$ ferrite. When a coil is wound around them, the ferrite produces a large inductance with very few turns. Sometimes the two beads are cast as a single block of ferrite with two cylindrical holes side by side through the block. In principle, the transformer is just like the transformers you have met before. That is, it consists of two coils wound on the same iron core. The high impedance, higher voltage winding has the most turns and is just 3 or 4 turns of wire wound through the hollow centers of both beads. So far, this is pretty ordinary.

The tricky, unobvious part is the low voltage, low impedance winding. What I haven't mentioned yet is that the hollow centers of the two beads are lined with non-ferrous metal tubes. The high impedance, high voltage winding is passed through these tubes. At one end of the assembly, the two tubes are connected together electrically so that they form a "U" passing through both beads. This "U" is the entire low impedance winding. Like any transformer winding, it has two output leads and these are the two legs of the "U." The center tap of the low impedance winding is the connection between the two tubes, on the right in the above drawing. That is, the bottom of the "U" is soldered to the PC board ground. The ungrounded ends of the tubes go to the balanced, low impedance transistor inputs.

For the input balun transformer, I made my "U" out of tubular mesh from the outer conductor of a length of RG-174 coax. I forced holes in the sides of the outer braid mesh to bring the insulated secondary winding wire in and out of the braided tubing. This is tricky to make and you may have to try a couple times. Using thin-walled brass or copper tubing would be easier and more elegant. I used Teflon insulated multi-strand wire for my secondary windings to be sure there would be no short circuits between primary and secondary. I bought the small input balun ferrites from CWS Bytemark. These small ferrites consist of a single, flat block of ferrite with two parallel holes molded through the longest dimension.

Bifilar wound RF choke

The power to the final is delivered by a small bifilar wound choke. Bifilar wound transformers were discussed in chapter 6. Wind about a dozen turns of a pair of #26 wires onto a small ferrite core. The two wires are wound on the core simultaneously as if they were one wire. An FT50-61 CWS toroid ferrite core will work well. The exact type of core or number of turns is not critical. Just be sure that the RF that appears from one winding will generate an opposite voltage in the other winding. If you don't, the two transistor collectors will be effectively shorted together. That's the meaning of the dots next to the coils on the diagram. Yeah, I get as confused as you do about dot marks. Leave the leads on one coil long enough to swap the ends when it doesn't work!

The tuned class B worked, but I don’t recommend it

The diagram of the tuned class B was shown earlier. Depending on the range of the ganged, dual tuning capacitor, it can tune between 10 and 20 meters. I got on the air using CW and talked to lots of people with my 50 watt Class B. I was pleased, but whenever I changed frequency more than about 50 KHz, I had to retune the amplifier. Using a scope, my procedure was to tune the amplifier and a “T” type transmatch for maximum amplitude with minimal low frequency artifacts. When tuned, it produced a clean sinewave output and I could see no evidence that the lack of forward bias was distorting the output. Just to be on the safe side, I ran the output through the multistage, TVI low pass filter described in chapter 9. This TVI filter is designed to work with any HF band since it cuts off above 10 meters. The amplifier ran quite cool and I didn’t burn up any transistors, even though I had omitted all those exotic feedback safety circuits.

The disadvantages of my tuned class B was that it was a bit tweaky and tended to go out of tune whenever the battery voltage declined. The best reasons for not building one are that the class B untuned amplifier described below works better and doesn’t use any hard-to-find parts.

A CLASS B, UNTUNED, SORT-OF-LINEAR AMPLIFIER

Episode two of the power amplifier saga

15 meters was dead in the evenings so I wanted to get on 40 meters. Rather than build a new tuned Class B amplifier just for 40, I went back to work on the linear. First I ordered a data manual for Motorola RF transistors. When I got my manual, I discovered that the MRF-454 was the only transistor in its class that ***WASN’T*** recommended for linear operation. The manual didn’t say why it wasn’t, but I thought to myself, “No wonder MRF-454s are so cheap and no wonder they run away with forward bias!” I got out my RF Parts Company catalog and priced all the similar RF transistors that were recommended for linear operation. They were all much more expensive than the MRF-454, but I picked out the cheapest and ordered a matched pair of MRF-422s. The output from the transistors goes to a large balun ferrite transformer. Large balun transformer assemblies are also available from RF Parts Company. I have used the 1 inch and 1.5 inch versions (PN # T1 & T1.5). Both seemed to work fine with no signs of saturation.

When my new transistors arrived, I put them in the linear and ... they ran away, just like the MRF-454s. Each transistor had swamping resistors that connected each base to ground. I lowered the values of these again and again until the transistors stopped running away. Of course by that time I had zero RF output. Apparently forward bias works for everyone else but the laws of physics are different at my house. On the other hand, unless you are planning to go on SSB (voice), you don’t really need a linear amplifier. Old fashioned class B or class C amplifiers work fine for CW.

Trouble with Chebyshev output filters

I tried again to run my “linear” amplifier as an untuned class B. As before, it just produced high power noise. I looked more closely at my Chebyshev output filters. Were they defective in some way? I had built them from the 1979 handbook linear amplifier plans. I had followed the winding instructions exactly using identical CWS (Amidon) 106-6 toroid cores. To test the 15 meter Chebyshev filter, I substituted it for the TVI low pass output filter on my working tuned Class B amplifier. It should have worked fine, but instead, the old tuned Class B amplifier went into noise mode, just like the new amplifier. Very little power arrived at the dummy load and the filter cores got quite hot. Something was wrong with the filter.

It turned out that the parts list or coil turns listed in the table were just plain wrong. I started over and redesigned my plug-in filters using the Chebyshev design tables from the 1998 ARRL handbook. I used the procedure described for making 5-element low pass Chebyshev filters that I previously explained in chapter 6. The main difference between the QRP filter and the big amplifier filter is that, for 100 watts, you need large powdered iron cores. Instead of T50-6, I used T106-6. For 40 meters and below you might want to use T106-2 cores. T68 cores might be big enough, but I haven't tried them at high power levels.

It turned out that the instructions in the 1979 manual described filters that were more appropriate for several bands below each band listed. For example, the 15 meter filter was designed about right for 160 meters, and so on. The filter for 160 meters would have been correct for low frequency transmissions from submerged nuclear submarines. Something in the T106 core specifications must have changed over the years.

I checked out the redesigned 15 meter filter on my old amplifier and … it worked perfectly! Progress! Next I put the new filter on my new linear amplifier and held my breath. Behold ... it was still operating in noise mode. Now I was mad. I got out my wood-carving gouge and cut the PC board traces to the transistor bases. Now they were free from all that R-C-L frequency compensation network gobble-de-gook. I wired the bases up just like the tuned Class B linear input above. As far as I can tell, it works perfectly. It puts out 100 watts of lovely sinewave on 40 meters and doesn’t blow transistors. I only get 50 watts on 15 meters because my driver isn’t as powerful. I soon worked dozens of stations on 15 and 40 and got excellent signal reports.

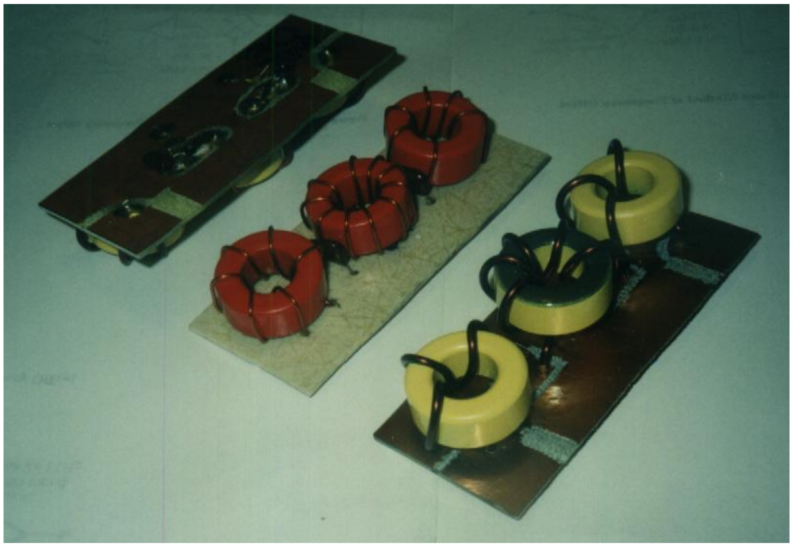

Chebyshev output low pass filters for an untuned class B. They plug into a card edge connector.

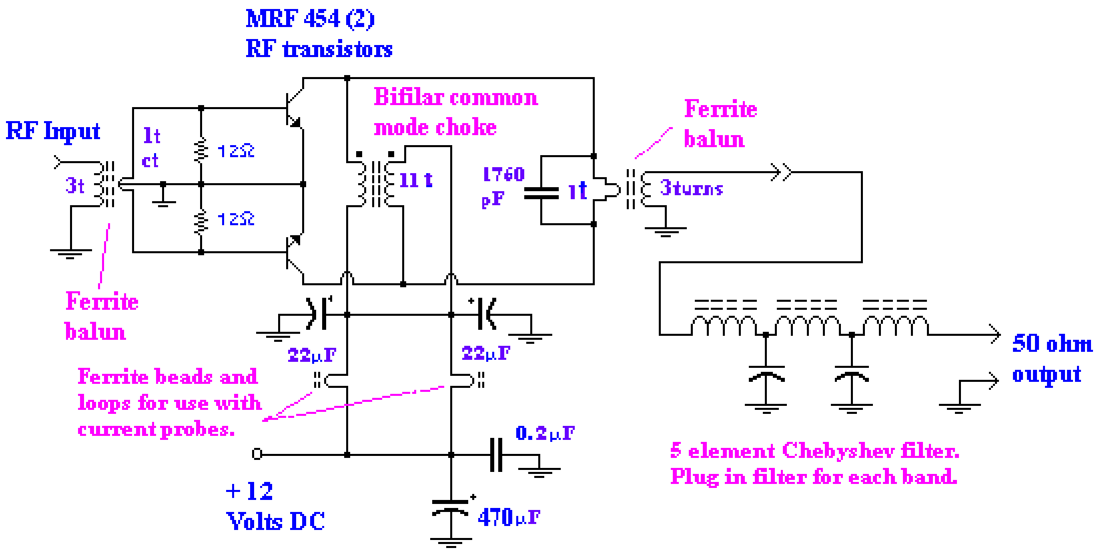
MRF 454 (2)
RF transistors
Bifilar common
mode choke
Ferrite
balun
RF Input
1t
ct
12Ω
12Ω
3t
11 t
1760
pF
1t
3turns
Ferrite
balun
22µF
22µF
Ferrite beads and
loops for use with
current probes.
0.2µF
+12
Volts DC
470µF
50 ohm
output
5 element Chebyshev filter.
Plug in filter for each band.

Amplifier # 2. **A simplified class B untuned amplifier**

Why didn’t I need fancy frequency compensation feedback loops? The answer may be that the purpose of the feedback loops was to equalize the outputs on all bands. Also, I am driving the linear with complete QRP transmitter drivers that have their own Chebyshev output filters. In other words, the input signals are quite pure. I have observed that any defect in the input sinewave is faithfully reproduced in the output. Even without forward bias to make it class AB, it is “hi-fi” to a large degree.

After all my worry, I never did ruin any transistors. On one occasion I was testing the linear at 80 watts output. I was happily looking at the scope when I smelled something burning. Oops. I had forgotten to screw the big heat sink back on. I shut off the linear and no harm was done. RF power transistors aren't so fragile after all.

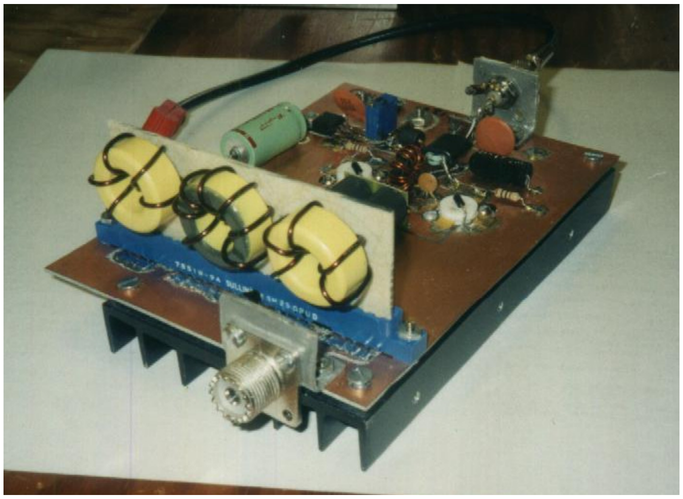

The completed linear amplifier. A 15 meter low pass filter is plugged into the output.

You may already be using an adequate low pass filter

I have occasionally had trouble loading antennas on 17 meters and above. Even with the T-match coupler described in chapter 9, sometimes I couldn't get a good sinewave signal on the antenna lead wire. The waveform was contaminated with low frequency modulation(s) and the frequency counter was usually reading low and wouldn't lock. Sometimes I have been able to correct the problem by using an output filter for the next band higher than the one I was on. For example, if it wouldn't load on the correct 15 meter filter, it would sometimes load well using the 10 meter filter.

Finally it occurred to me that I was already using the multistage 30 MHz cut-off low-pass TVI filter described in chapter 9. Therefore, for those high bands my 10 meter output filter on the final was redundant. I built a "blank" filter that was just a piece of RG-58 coax that shunts from one end to the other of a blank PC board plug. To summarize, using the TVI filter by itself is another alternative for your bag of tricks.

Keying the 50 watt linear amplifier

When I first began using my finals, I kept them turned on all the time. That is, whenever a QRP signal appeared at the input, the final was supposed to amplify it. When there was no input to the amplifier, there should have been no output. This way, I could leave the 12 volt power supply connected all the time. Unfortunately, every so often the final would begin oscillating all by itself at some random frequency outside the ham bands. I often had two or three successful QSOs without trouble. Then suddenly, for no reason that I could determine, the oscillation began.

I have been told that professionals solve this problem by loading down the input of the final amplifier with a resistor, perhaps as low as 100 or even 50 ohms. I haven't experimented much with this approach, but I'm certain that it cuts drive to the final and lowers my output power. If I were you, I'd experiment some more with this cheap approach. However, if you don't like your results, you can always fall back to my keyer circuit shown below.

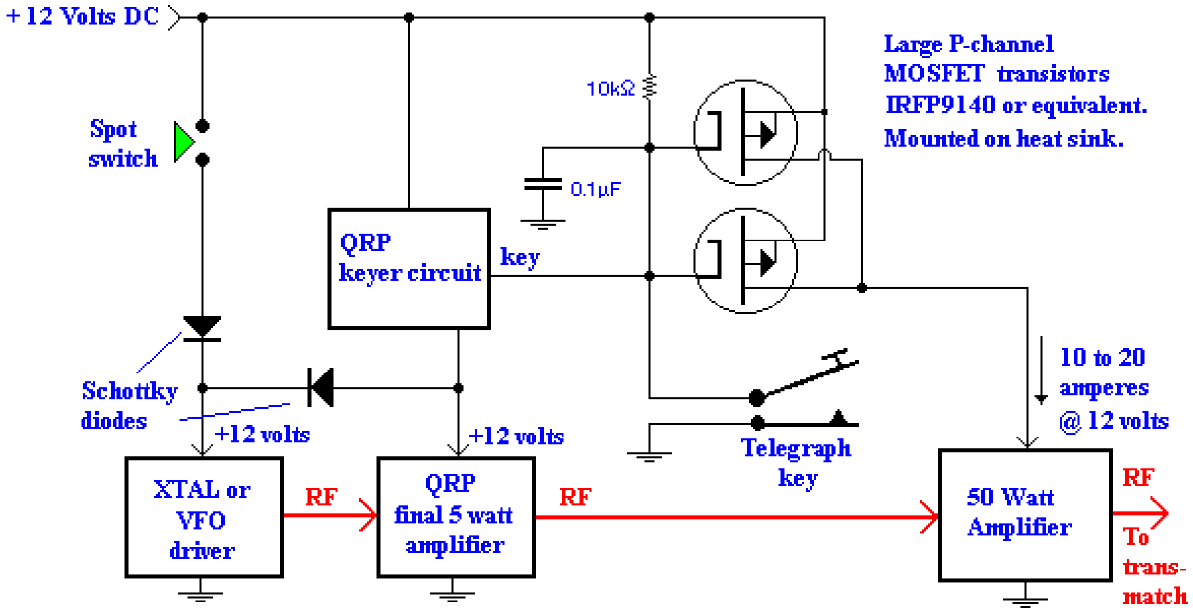
+ 12 Volts DC
Spot switch
10kΩ
0.1µF
Large P-channel MOSFET transistors IRFP9140 or equivalent. Mounted on heat sink.
QRP keyer circuit
key
Schottky diodes
+12 volts
+12 volts
10 to 20 amperes @ 12 volts
Telegraph key
XTAL or VFO driver
RF
QRP final 5 watt amplifier
RF
50 Watt Amplifier
RF
To trans-match

50 WATT TRANSMITTER KEYER

My solution was to build a giant version of the same MOSFET power switch I used to key my QRP modules. Naturally I had to use humongous P-channel MOSFETs with low on-resistance. Referring to the figure above, the telegraph key pulls the MOSFET gates down turning on the MOSFETs and bringing power "down" to the amplifier. The MOSFETs are mounted on a small sheet of 1/8 inch thick aluminum plate which in turn is bolted to the thin aluminum chassis. Even at heavy current loads I haven't noticed the MOSFET transistors becoming warm to the touch. The cases are the drain connection and must be insulated with mica or equivalent insulators. A light layer of silicon heat sink grease on the mica insulator fills the air gaps and improves heat conduction.

Power bipolar and MOSFET transistors and the necessary size TO-204 mounting kit

As with the QRP module, turning on the power abruptly by shorting out the 0.1 microfarad capacitor with the key looks like a recipe for producing key clicks. I haven't had any trouble yet with this, but I wouldn't be shocked if someone received complaints of key clicks. David, VK6KI, suggests turning the keyer on gradually with an op-amp keyer will prevent any abrupt turn-on and turn-off. He also suggests using cheaper bipolar PNP power transistors. Use whatever works for you. It's your homebrew and it's fun to try different ideas.

N-Channel MOSFETS have much lower ON resistance and are harder to damage. They can also be used as keyers and fewer transistors can do the same job. Of course, the problem is that, instead of pulling the gates down to ground as we did above, now the gates have to be pulled UP to 10 or 12 volts above the 12 volt supply. If you have built the VFO described in chapter 10, then already you have a convenient 22 volt source. This is the keyer in my CW transmitter:

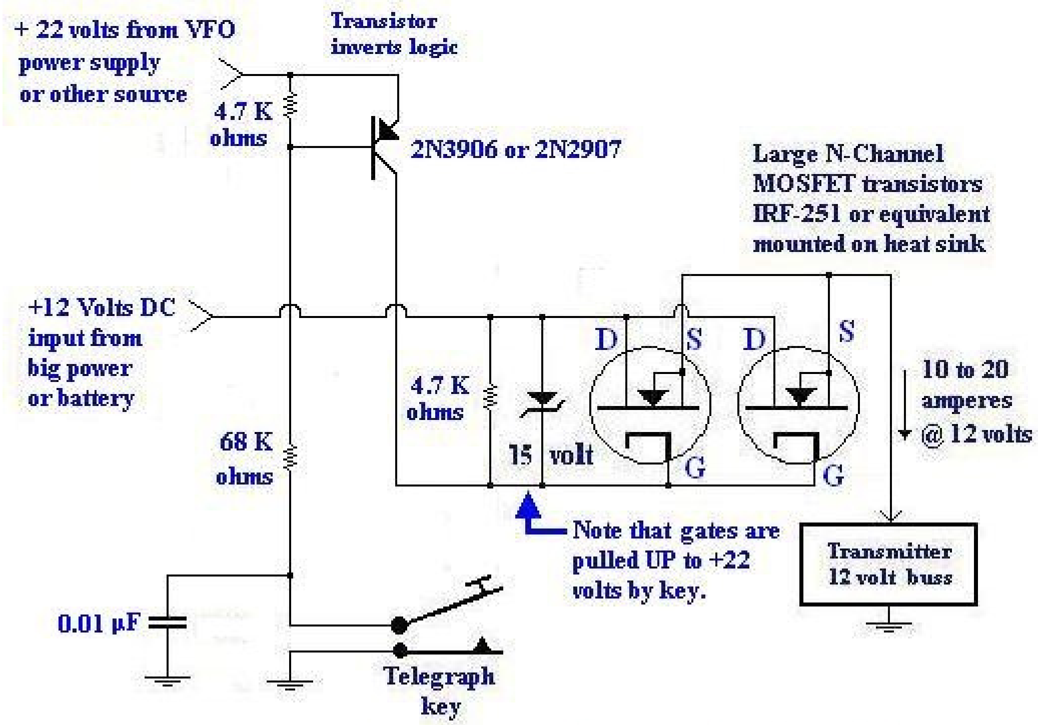
+ 22 volts from VFO
power supply
or other source
Transistor
inverts logic
4.7 K
ohms
2N3906 or 2N2907
Large N-Channel
MOSFET transistors
IRF-251 or equivalent
mounted on heat sink
+12 Volts DC
input from
big power
or battery
D
S
D
S
4.7 K
ohms
15 volt
G
G
10 to 20
amperes
@ 12 volts
68 K
ohms
Note that gates are
pulled UP to +22
volts by key.
Transmitter
12 volt buss
0.01 µF
Telegraph
key

HIGH POWER TRANSMITTER KEYER
Using N-Channel MOSFET transistors

What had I learned up to this point?

1. The ARRL Handbooks are excellent but occasionally they print errors. Don't believe everything you read in a parts list.

2. I understand the need for class A forward bias. In practice, for CW use it isn't necessary. However, I was still mystified how the experts do it without a runaway. In my experience, the transistors ran away instantly with the first milliamp of forward bias. This happened even when

the transistors were stone cold. It couldn't have anything to do with temperature feedback being defective because there wasn't time for heating to occur. Moreover, runaway wasn't related to the RF drive, because it happened with or without RF input.

3. Ferrite balun transformers are impressive components. They produce tight coupling at really low impedances and they don't need tuning.

4. If your Chebyshev output filter doesn't work the first time, check it out carefully for solder splash shorts on the PC board. If you're sure it should work, but it doesn't, go to the design table in a recent ARRL handbook and redesign the filter yourself starting from scratch. If it still doesn't work, try a different core size.

5. A Chebyshev filter in a QRP driver worked poorly when I designed it with T68-6 cores. But the same filter worked great when I rebuilt it using smaller T50-6 cores. Sorry, but I have no idea why. Sometimes it helps to be open-minded and try things that may seem silly. Persistence is your ultimate weapon!

6. Finally, it seems to me that much of the complexity in designs in QST and QEX is great in theory, but sometimes unnecessary in practice. The guys who wrote those articles are over-educated. Their sophistication often discourages us. Don't let them rain on your parade! Build it simple and work up from there.

****

A LINEAR AMPLIFIER, THIS TIME FOR SURE

Adding linear bias to the Class B amplifier

As explained above I was able to run my final amplifier as an untuned Class B, but when I applied DC bias to make it "linear," the amplifier "ran away." That is, it drew huge currents and blew fuses. In the end, I left it as a class B broadband amplifier. Class B amplifies both halves of the driving sinewave, so there's only a small cross-over non-linearity. So who cares if it isn't class AB? At the end of my project I was happily working guys on CW and I always used a lowpass filter to suppress harmonics. So what's the fuss over "linear" amplifiers? *I* certainly didn't need one!

Sideband needs a linear

Eventually I got bored with HF CW and built an SSB exciter. I fed SSB English speech into my class B "nearly-linear." All that came through were the voice peaks. It sounded something like African click language. I couldn't understand the speech, but I finally got the message: ***The key virtue of linears for sideband is that they amplify all AMPLITUDES.*** Yes, linears also amplify all *frequencies* equally. They are a sort of "RF Hi-fi," but the broadband part isn't so important. A sideband amplifier only has to amplify a signal 3 KHz wide. In theory at least, a tuned class A amplifier would work fine on sideband, even if you had to retune it every time you QSY a hundred kilohertz. In practice, a tuned amplifier on sideband would tend to self-oscillate every time you stopped talking.

When class C or class B amplifiers are operated CW, a big drive signal comes in on the base(s) of the final amplifier and you get a big, constant output signal with about 10 dB gain over the input. The big drive signal exceeds the base forward voltage drop of the output transistors by a wide margin. ***When operating CW the amplifier only operates at one amplitude***, so the base voltage threshold problem never arises.

Why does sideband need a linear when AM modulation was so easy?

Sideband is different. The drive from an SSB exciter has a range of amplitudes and all must be amplified equally. Actually, it's useful to amplify the low amplitude signals *more* than the high amplitude voice peaks. This raises the average power and makes the SSB signal better able to compete with noise and QRM.

At first glance SSB doesn't seem so different from AM. So why was AM phone so easy to build back in the old days? Amplitude modulators usually modulated the final, not the driver. For AM we tuned up the tube final on CW, then ***we modulated the power supply, or the screen grid, or the cathode of the final tube amplifier with audio frequency***. The final amplifier acted like a mixer that produced sum and difference frequencies. With no conscious engineering effort we produced two audio frequency sidebands in addition to the original RF carrier. Another advantage of an AM transmitter final is that, whenever there is no speech, it is still generating a carrier, just like CW. Therefore it can't self-oscillate.

Guys who could afford big AM modulation transformers modulated the power supply for the final. We cheapskates modulated the screen grid or cathode of our final amplifier tube. Either way, we were modulating the gain of the final amplifier, not the drive signal. Now that I think about it, I guess I knew that some guys used *LINEAR AMPLIFIERS* to boost the power of their low power AM exciters. The fog clears. Yes, you can broadcast AM with class B or even Class C final amplifiers, but you can't amplify a low level AM or SSB drive signal without a linear.

Biasing a linear amplifier without thermal runaway

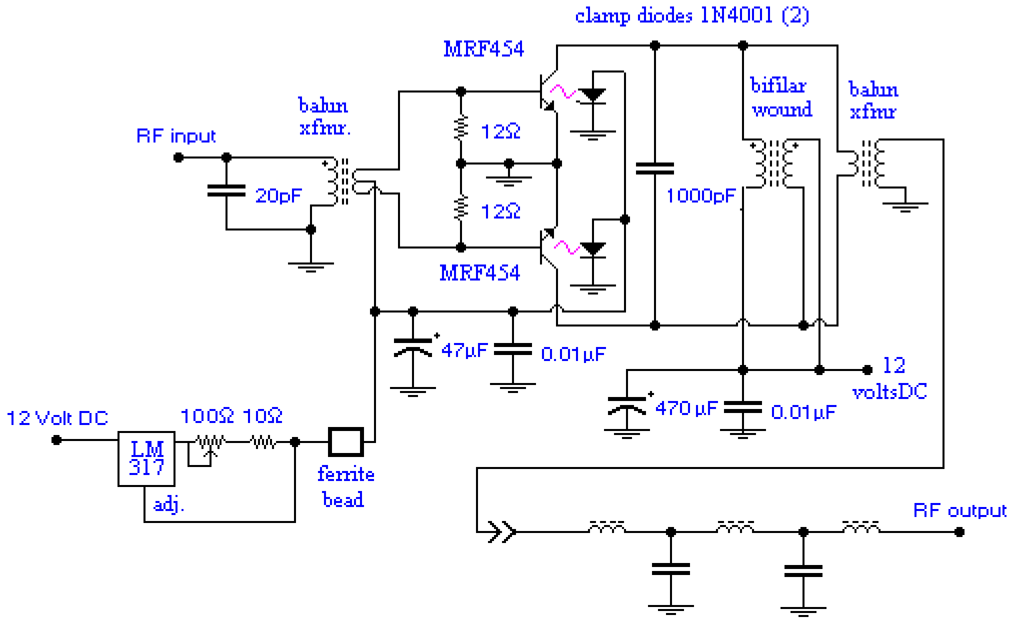

Chebyshev filter for particular ham band
See ARRL Handbook

Amplifier # 3. **A push-pull linear amplifier**

The entire linear final is shown above. The bias circuit at the lower left solved my run-away problem and gave me the linear amplification I needed. I found the bias circuit lurking in a big schematic describing a commercial transceiver in the 1998 ARRL handbook. This deceptively simple circuit performs three functions:

* It provides an adjustable, constant DC bias current into the transistor bases.

* Diodes (1N4002) shunt the bases to ground, limiting how high the base voltages are allowed to rise.

* The diodes heat up with the transistors and provide temperature compensation.

An LM317 voltage regulator is used as an adjustable current source to feed roughly 100 milliamperes of DC into the bases of the high power output transistors. The LM317 is a three-terminal, 1.2 volt voltage regulator. The regulator output passes through a variable current set-resistor. The voltage drop across the resistor is monitored by the "adjustment" lead of the regulator so that the voltage across the resistor is held constant at 1.2 volts. The regulator is RF-isolated from the amplifier by a big ferrite bead on a short chunk of wire. The rest of the circuit is almost identical to the final in an SSB transceiver in the ARRL 1986 handbook. The 1000 pF capacitor across the collectors is found in all these push-pull circuits. I took mine off and I got a few percent more gain. I don't know what it's supposed to do.

Clamp diodes prevent run-away

The critical parts of the bias circuit are the two rectifier diodes which clamp the base voltage (V_{be}) to ground. Their most important function is that they instantaneously clamp the base voltage to less than roughly 0.8 volts, like a Zener diode. They limit how high the drive voltage to the transistors can rise when the whole assembly gets hot. Hot transistors draw more current, causing more heating. This positive feedback is called thermal runaway. As the input drive to the transistors increases, the DC base-to-emitter voltage of the output transistors rises which causes the collector to draw bigger and bigger currents. For example, while operating sideband, the DC base voltage rises to well over 0.8 volts on voice peaks. If it were allowed to rise to 0.9 or 1.0 volt, runaway might result.

The second function of the diodes is that they also change with temperature and automatically compensate for temperature rise. ***They are strapped across their respective transistors so they are in thermal contact with the transistor cases.*** As the transistors and piggyback diodes heat up, the forward voltage drop of the diodes ***decreases*** with increasing temperature. When cold, the V_{be} peaks might start at 0.8 volts, but as the transistors heat up, the V_{be} voice peaks will try to rise higher still. So, under the same bias current level, but at a higher temperature, the diode clamp voltage might typically drop from 0.8 to 0.7 volts or even lower. The decreasing voltage drags the base voltages down, preventing runaway. While holding the telegraph key down, I could watch the DC base voltage slowly sink, while the total DC current drawn by the entire amplifier remained constant.

The difference between this bias circuit and the ones I tried earlier is that there were no clamp diodes. Yes, there was a reference diode mounted on the heat sink for temperature compensation, but my output transistors ran away immediately before the temperature had a chance to rise. A current source with just a temperature compensation circuit has no instantaneous clamping function.

Mechanical construction

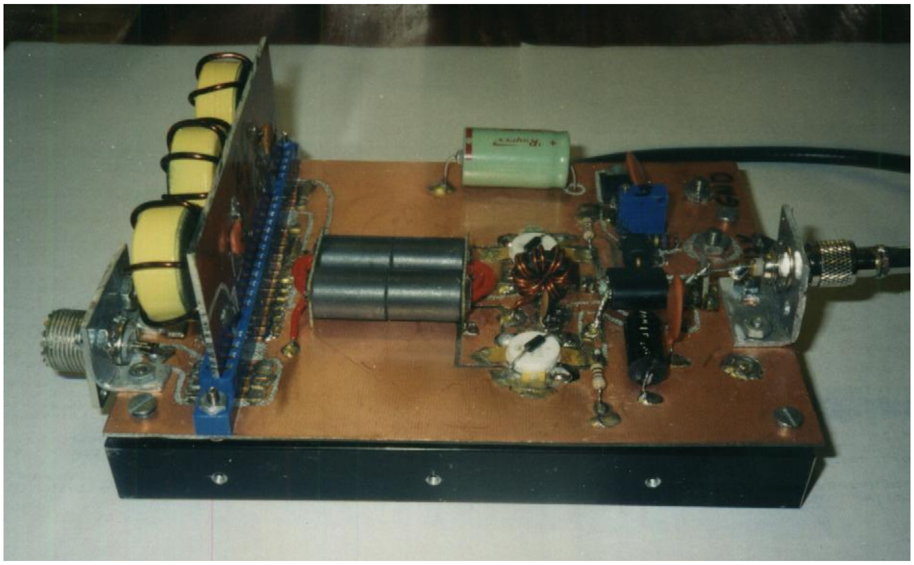

The linear 50 watt amplifier. Notice the large ferrite balun output transformer in the center of the assembly. The bias circuit is at the right rear. Also notice how the base clamping diodes are strapped across the output transistors so that any transistor heating will immediately be passed on to the diodes. The holes in the PC board adjacent to the white output transistors are where the machine screws bolt the transistor bodies to the heat sink. A thin layer of silicon grease between the transistors and the heat sink is advisable, but mica insulation is not needed. These mounting tabs are not electrically connected to the transistor leads.

The completed amplifier module is made from a two-sided circuit board screwed down into holes tapped in a large, finned heat sink. The major RF traces are wide, about 3/8 inch or more, to keep the inductance down. This is a simple circuit and all the traces were cut into the board with a small wood-carving gouge. ***It is vital to arrange the collector and emitter traces so that they are symmetrical and equal in every way to the other transistor.*** Otherwise, one transistor will have more trace capacitance and inductance than the other. The input and output transformers are the balun type which were described earlier.

The input comes in through a mini-UHF connector. Mini-UHF connectors are the only affordable, small RF connectors I have found. At $5 a mated pair they are expensive compared to phono plugs. But if you must pay $12 a pair for SMA, SMB, or SMC connectors, $5 looks cheap. The output connector is a big UHF SO-239 that, strangely enough, has always been affordable. For high power, >5 watts, I suggest avoiding connectors like phono plugs which weren't designed for RF power. You can get away with phono plugs on 20 meters and below, but the standing waves make it difficult or impossible to tune the amplifier on 10 meters. Also, if you plan to amplify sideband with this, it's a good idea to build a shield cover to fit over the top of the board. As you can see, I didn't. However, all commercial transmitters have their linears sealed in extremely tight metal enclosures. In my SSB transmitter all the low level modules are sealed individually in separate metal boxes.

I built my Chebyshev low-pass filters for the linear on separate PC boards that plug into the linear board using a card edge connector. To keep the inductance low, I soldered eight connector pins in parallel at each end of the card connector and the rest of the pins were used for ground. To change bands, I plug in another low-pass filter. If you like, you can use little RF reed relays to switch in different filters. Personally, I like the old-time flavor of plug-in coils. It is probably practical to use smaller ferrite toroids, such as T68-6, but I haven't tried them at this power level.

Why stop at 100 watts?

I opened up an issue of QST (June 2006) and discovered plans for a homebrew 600 watt linear amplifier. It uses two pairs of MRF-150 transistors to reach this level. It is quite similar to my linear in many respects. I was surprised to see that it's practical to get so much power from so few parts.

I was tempted to build one until I remembered Bob Hamilton's experience with his kilowatt amplifier. Bob, NØRN, used to live in the city like I do. He quickly found that his new kilowatt interfered with the neighbor's TVs and stereos. In contrast, his 100 watt transmitter had never bothered them. Bob fought the problem for several months and even had his station checked out by a government accredited expert. It turned out that his station was operating correctly and in accordance with the law. The difficulties were in the sensitivity of the neighbor's appliances. Bob was "legal" but that didn't really change anything. His neighbors were still mad at him and weren't willing to buy new TVs, cordless phones, etc. In the end, Bob mothballed the linear in a closet and went back to 100 watts. Since then Bob has moved out into the country where his nearest neighbor is hundreds of meters away. He resurrected the kilowatt amplifier and has had no complaints from his neighbors. Signal strength falls off with the square of the distance and 300 meters is vastly different from 30 meters distance.

Surprisingly, in the US the maximum power on most HF bands has been increased. Formerly transmitter power was defined as the DC power input to the final stage. For example, in the old days a "50 watt novice transmitter" might have 500 volts DC on the tube plate and draw 100 milliamperes. That is, 500 volts times 0.1 amps = 50 watts. Since the efficiency of a class C amplifier is no greater than 65%, "50 watts" was really only 32.5 watts PEP. Similarly an old "kilowatt" Class B amplifier, the kind you find at hamfests, is 50% efficient and delivers at most 500 watts to the antenna. But now on many bands we are allowed 1,500 watts PEP measured at the antenna. That can be three times as much power as an old “kilowatt.” And, when you consider that SSB concentrates all the power into a single sideband, the modern SSB kilowatt has at least 6 times the effective power as an old AM modulated kilowatt.

In conclusion,

In the end I was able to build a linear using the cheap MRF-454 transistors, just like the handbook said - even if their darn bias circuit didn’t work. I talk to guys on sideband and they seem to understand what I’m saying. If my final weren’t close to linear, they wouldn’t have understood a word.

How much power the linear amplifier delivers depends on how much power I put in and what band I’m on. Not all my QRP drivers are equal. On 10 meters I only get 20 watts, while on 40 or 80 meters I get as much as 120 watts. As usual, lower frequencies are easier. The efficiency is about 50 percent. For example, to get 120 watts output I need 18 amperes input from the battery.

17. Chapter 12, Harris

Hmmmm ... Now that it works, I wonder what would happen if I unsoldered the base clamping diodes? It should run away, of course. After all, these are the same individual transistors I used before. It better run away! So I unsoldered the diodes. It didn’t run away. The temperature compensation no longer worked, but it didn’t run away. I have no idea why it didn’t. Sometimes electronics drives you crazy! Persistence is your only weapon against the innate perversity of inanimate objects.

CRYSTAL SETS TO SIDEBAND
© Frank W. Harris 2010, REV 12

Chapter 13

BUILDING A HOMEBREW HF RECEIVER

The Vanishing Art

The 1986 ARRL Amateur Radio Handbook reported that hardly anyone was building homebrew ham receivers. Since then, handbooks have had fewer and fewer practical details on building radios. Out of hundreds of contacts, so far I've only worked three guys, George, K7DU, Mike, NØMF, and Biz, WDØHCO, who were using homebrew receivers for the QSO. Two of these receivers were made from vacuum tubes. The third homebrew receiver that I worked was an old-fashioned super-regenerative design. However, the super-regenerative was implemented with modern FET transistors and integrated circuits. All three of these receivers seemed to have no trouble hearing me on 40 meter CW. I talked to one other fellow, Gil, N1FED who told me he had just finished a vacuum tube receiver. Unfortunately, it was performing so poorly he was still using his modern transceiver on the air. Gil told me he didn't like transistors. I guess he found printed circuit boards and those pesky oscillations too much trouble.

In spite of this pessimism, ***you CAN build transistorized receivers that work reasonably well.*** I built mine because I was intrigued by mysterious circuits like "balanced mixers," "product detectors," "cascode amplifiers," and "crystal ladder filters." Before this project, I could recite the purposes of these circuits, but I had no "feel" for how they worked and why receivers are designed the way they are. What better way to learn than to build one?

What's a reasonable goal?

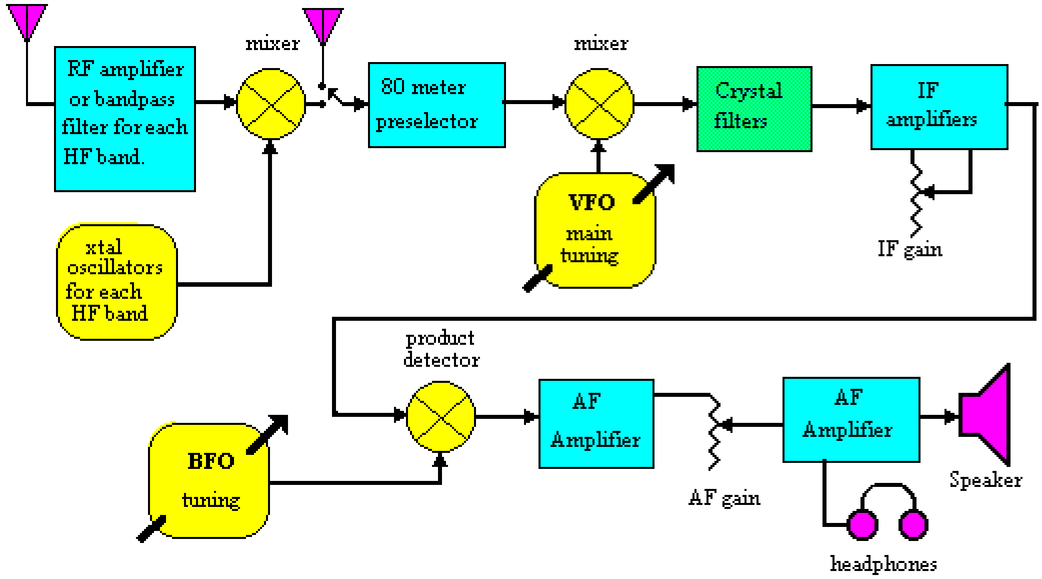
RF amplifier or bandpass filter for each HF band.
mixer
80 meter preselector
mixer
Crystal filters
IF amplifiers
VFO main tuning
IF gain
xtal oscillators for each HF band
product detector
AF Amplifier
AF Amplifier
BFO tuning
AF gain
Speaker
headphones

An "adequate performance" communications receiver

My receiver is based on the “High Performance Communications Receiver” designed by W7ZOI and K5IRK described in most of the annual ARRL Handbooks in the 1980s. In my opinion “High Performance” is optimistic, but certainly “adequate performance” is realistic. I define ***adequate sensitivity and noise figure*** to mean that I can hear the DX and QRPs that other stations are working. Before I built the receiver described here, I often had the impression I was hearing only the loudest signals. For me, ***adequate selectivity*** means that it's good enough for CW QSOs in the evening on 20 and 40 Meters. On these bands there are often dozens of narrow CW signals operating within a few hundred Hertz of each other. With a 10 KHz bandpass you may hear many stations simultaneously and not be able to copy any of them.

Adequate sensitivity will allow you to hear most QRP signals. I believe that 45 years ago hardly anyone had receivers that were “adequate” for QRP contacts. When I was a novice, my first transmitter was a 7-watt homebrew for 40 and 80 meters. It was a close copy of a design in the 1957 Handbook. I know it worked OK, because I talked to my novice buddies around town. Unfortunately I hardly worked anyone outside of town. It wasn’t until I bought a 50-watt commercial kit, just like all the other novices, that I was able to talk to all the same stations my friends were working. I was still using the same dipole antenna, so I can only assume the improvement was the extra power.

The sensitivity of the receiver described in this chapter is well under 0.5 microvolt on 80 meters and lower bands that have no RF preamplifiers. On the upper bands where the receiver has preamplifiers I could hear a calibrated signal source at 0.02 microvolts. Wow! No wonder I can hear those QRPs. In the old days sensitivity less than 1 microvolt was considered hot stuff.

Another issue is ***adequate stability.*** When your receiver is equipped with sharp crystal filters, it is vital that the VFO and crystal oscillators are stable enough that the signals you're listening to do not drift in and out of your passband. If you build a VFO like the one described in chapter 10, you will have no drift troubles.

Does it have to be so complicated?

Looking at the block diagram above, each one of those blocks represents one to three transistor stages. The front-end converter has three transistor stages for each separate HF band. That means you need to build about 20 transistor amplifier or oscillator stages for the converters to cover all the bands. You’re probably wondering if there isn’t some simple receiver you can build that will get you on the air rapidly. The best news is that you can build the above receiver in stages. The core of the all-band HF receiver is a quality 80 meter receiver. You can build that first, then at least you’ll be in business on 80 meters. In the beginning you can also do without the loudspeaker and multiple crystal filters. After you have a functioning receiver, you can add features and the converters to hear the other hambands.

Yes, you can build a less complicated receiver, but I doubt it will be “adequate.” My direct conversion receiver in chapter 7 worked well, but was not selective enough. There are modern super-regenerative receiver designs with two or three integrated circuits that *MAY* be pretty good, but I haven’t built one, so I’m not convinced. Chapter 14 describes a vacuum tube regenerative receiver that was great fun to build and quite good for listening to foreign short wave broadcast stations. Unfortunately, it was *NOT* selective and sensitive enough for ham communications. In summary, ***YES, a decent ham receiver does have to be complicated.***

A homemade ham receiver built in 1967

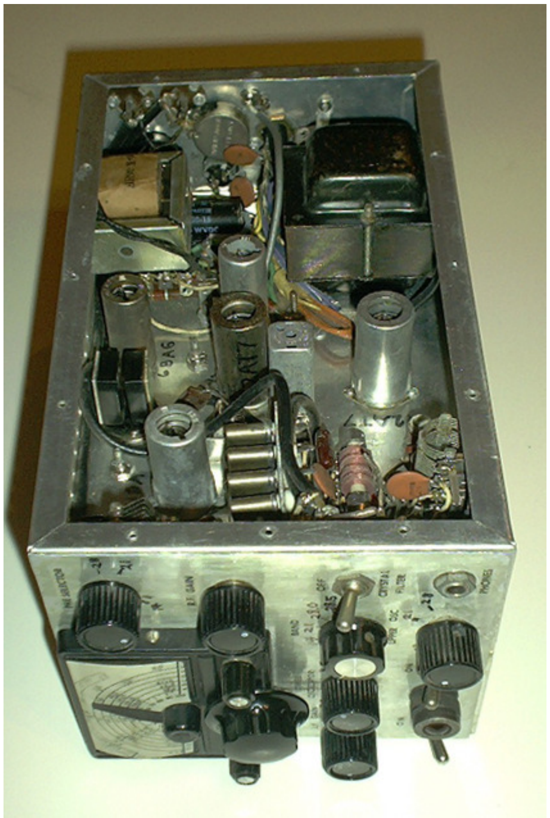

A homebuilt ham receiver from 40 years ago is shown above. It has 11 tubes, a simple crystal filter, and covers 80 through 10 meters. It runs on either 12 volts DC or 120 volts AC. It doesn't cover the WARC bands. Yes, it works OK. But compared to the all-transistor receiver described in this chapter, it is insensitive, noisy, and has poor selectivity.

Being realistic, any receiver you build is unlikely to match the performance of high-end commercial rigs. But every time your receiver brings in DX on a new band or whenever you conquer one of the dozens of glitches you will encounter, you'll have a thrill and pride you'll never get from a commercial rig. If you decide to build your own version of the W7ZOI / K5IRK receiver, I recommend you find a copy of an old ARRL Handbook from the 1980s and Xerox their original descriptions. You'll find they built most circuit blocks differently than I did. Going back to the original description may give you some useful ideas. Perhaps their version will work out better for you.

PLANNING YOUR RECEIVER

Superhetrodynes offer crystal filters for CW

A superhetrodyne uses a mixer to produce a constant intermediate radio frequency (IF). This intermediate frequency signal is always at the same frequency so it can be filtered with fixed crystals or mechanical filters to establish bandpass widths for CW and upper and lower SSB. Before you commit to any design, make sure you can buy the critical parts you need, especially the crystal or mechanical filters for your IF. For example, many receiver designs use a 455 KHz IF. Unfortunately, I have yet to find an easy source for 455 KHz crystals for building filters and

BFOs. Consequently, I have avoided this frequency. Among homebuilt ham receivers the most common IF frequency seems to be 9 MHz.

Why not single conversion?

I had always wondered why homebuilt all-band HF receivers are almost always dual conversion. It turns out that the fundamental challenge of homebuilt receivers and transmitters is building a stable VFO. Yes, you can build a reasonably stable VFO, but homebrew VFOs usually don't have much tuning range. 0.5 MHz is typical. And, in order to drift as few Hertz as possible, the VFO needs to be relatively low frequency. Homebrew VFOs are usually in the range of 2 to 7 MHz. The disadvantage of a low frequency VFO is that its harmonics will appear as one or two loud whistles on some upper HF ham bands.

Compared to the practical VFO range of 2 to 7 MHz, the HF spectrum is huge, 1.8 to 30 MHz. Right away one can see that a homebrew direct conversion 10 meter receiver is difficult because it needs a stable VFO that will tune 28 to 29.7 MHz. This problem can be solved by "converting" the VFO oscillator up to 28 MHz using a crystal controlled oscillator and a mixer plus 28 MHz filter/amplifiers. This complexity ruins the simplicity of direct conversion. If you're going to operate above 40 meters, you may as well build dual conversion like the rest of us.

How do modern receivers do it?

Modern receivers use integrated circuit frequency synthesizers to generate stable VFO signals anywhere they like. Sometimes modern HF receivers escape from artifact images and harmonics by using an IF frequency way up in the VHF range. In addition, after the initial mixer stage, some commercial receivers use multiple conversions to get the signal back down to an audio output. At each conversion stage, different kinds of filtering are applied. For example, the Yaesu FT1000MP has four down-conversions from an 89 MHz IF. (!) This includes the digital signal processor with its 32 KHz input.

In a superhetrodyne the VFO interacts with the incoming RF signals to produce an intermediate (IF) frequency. A 5 MHz VFO implies that the IF is going to be within 5 MHz of the band or bands it covers. Such a receiver might cover 28 MHz, but that would imply an IF of 23 MHz or possibly 33 MHz. The lower bands would be out of range unless the VFO could be tuned over many MHz. Consequently, a single conversion homebrew superhetrodyne can only cover one band well and can't possibly cover the whole spectrum.

In some old ham designs the VFO tuned 5.2 to 5.7 MHz. They used a 1.7 MHz IF and either subtracted or added the IF to the VFO frequency to cover either 80 or 40 Meters. Specifically, 5.7 MHz ***minus*** 1.7 MHz = 4.0 MHz and 5.3 MHz ***plus*** 1.7 MHz = 7.00 MHz.)

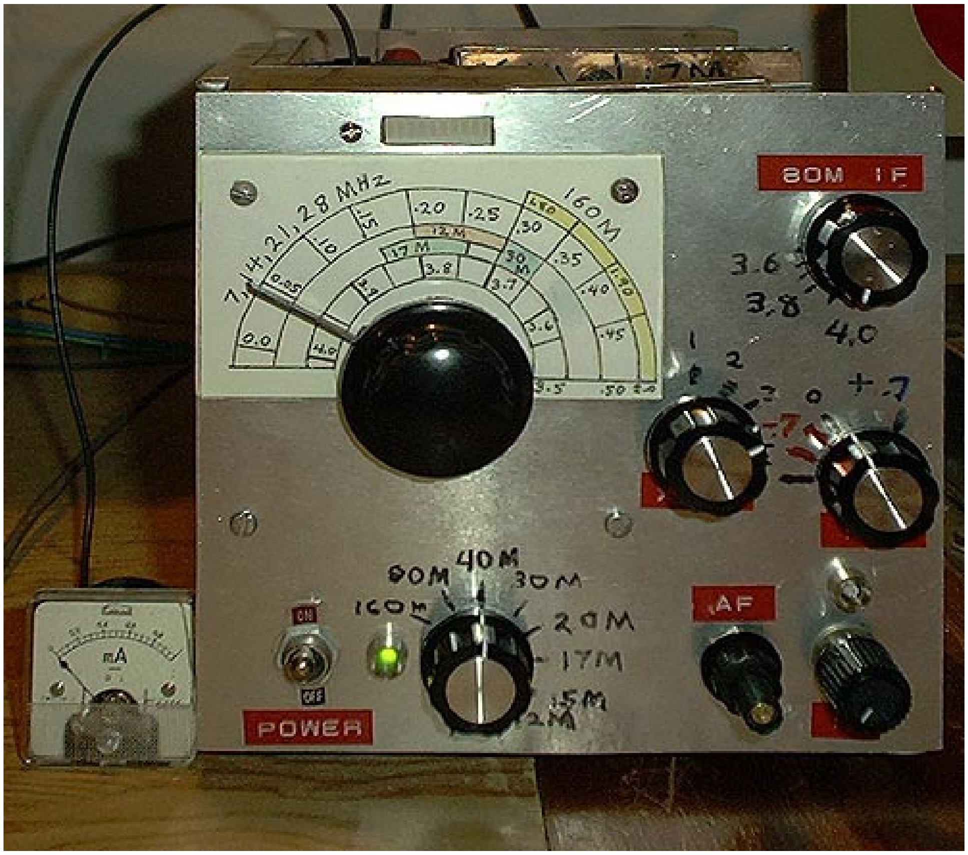

Homebuilt all-band dual conversion HF ham receiver.

Start with a single band, single conversion superhetrodyne

My "adequate receiver" is shown above. At the upper right is the preselector tuning for peaking or decreasing the sensitivity of the 80 meter IF. This control is useful as an attenuator when tuning in SSB stations. Above the frequency dial is a bargraph S-meter. The S meter is useful for zeroing in my transmitter VFO onto a station I wish to call. Unfortunately, the bargraph produces digital switching noises that often interfere with weak signals. When I substitute the analog meter on the left, the noise vanishes. Difficulties like this are typical of the real world of building your own equipment. Try to learn from these challenges and find pleasure in meeting them. Building perfect equipment is not a practical goal for us homebrewers.

Begin by building the best single-band receiver you can. The core of my receiver tunes 80 Meters. To cover the other eight HF bands, separate crystal controlled converters convert the signals down (or up) to 80 meters. This makes the receiver “dual conversion” on every band except 80 Meters. In the old vacuum tube days each homebuilt amplifier stage was designed to work on several bands. This meant that the front panel of a 1970 homebrew receiver usually had a bunch of tuning knobs to adjust each stage. Today, a transistor amplifier or mixer stage is physically tiny, so it’s practical to build a separate converter and amplifier pre-tuned for each band. My experience was that building the converters to translate the bands to 80 meters was relatively easy. Unfortunately, building a decent 80 meter core receiver was much harder. Maybe with the help of this chapter, you’ll find the core receiver easier to build than I did.

My 80 meter receiver has a 9 MHz IF. 9.00 MHz crystals are available for less than $1 from Digi-Key and Mouser. The low price is important because, depending on your filter plans, you may need as many as 11 or more 9 MHz crystals. The downside of 9.00 MHz was that I had to wait months for the crystals to arrive.

An 8 MHz IF wasn't such a good idea

At first I used the more common 8.0 MHz crystals. Unfortunately, to receive 4.0 MHz, the VFO had to tune 4.0 MHz. I expected the 4.0 MHz VFO signal would be "a little birdie" that would mark the high end of the band. I thought this "edge-of-band marker" would be kind of convenient. Instead, the "birdie" was more like a screaming siren that overwhelmed the IF and made the upper end of the 80 meter ham band unusable. So when the 9.0 MHz crystals eventually arrived, I rebuilt everything for 9 MHz. Now the VFO (the big tuning knob on the above receiver) tunes 5.0 to 5.5 MHz to cover 4.0 to 3.5 MHz. That is, 5.0 MHz + 4.0 MHz = 9.0 MHz. Of course the BFO frequency also had to change from 8 MHz to 9 MHz.

An unusual adventure

Once your receiver begins to work, you'll have interesting glitches. Until I got my 80-meter preselector filter working, I usually heard rap music from my local 1190 KHz AM radio. Also, the 31 meter shortwave band is just above the 9 MHz IF. Before it was aligned, I was hearing sermons from HCJB in Quito, Ecuador. Later, my 20 meter converter was overwhelmed by Dr. Scott, a Los Angeles evangelical minister, who preaches on 13.8 MHz. Once I had my modules tuned and sealed, Dr. Scott and his friends were silenced. Actually, I got a kick out of these problems.

Building a receiver revived my interest in shortwave listening. I've had shortwave radios since I was a kid. Some of them, like my Army surplus Collins R-388/URR, were excellent. In spite of this, I rarely listened when I wasn't actively hamming. But once my homebrew receiver(s) began to work, I found myself exploring the bands as never before. For instance, on 80 meters I was amazed to hear hams from all over the continent. I had heard about guys who work DX and earn WAS certificates (Worked All States) on 80 meters, but I never really believed it. I have even worked QRP stations on 80 meters. 80 meters is usually so noisy, I didn't know that was possible. Until I built this receiver, I had never heard "spy code stations" before. Some of them have an announcer reading what sound like random letter groups, but most of them are CW "spy stations" sending Morse code 5 letter groups, just like the WWII Enigma signals. I often hear them on 10 and 30 meters. As I built converters for each of the HF bands, it was like hearing those bands for the first time. The "WARC bands" were interesting because they were near to shortwave broadcast bands I hadn't listened to in years.

Building with modules

Aside from the need to shield circuit blocks from one another, a homebrew receiver with a single big board full of discrete components has another problem. If you build the whole thing at once without buying a kit and pre-cut board, I guarantee it won't work. ***To make homebrew stuff that works, you have to develop your own technology based on parts you can get and circuits you understand.*** Thinking this way has always been difficult for me. Rather than "building a receiver," I had to lower my sights and build one circuit at a time, e.g., "an oscillator," "a mixer," "an audio amplifier," etc. Then I put the blocks together to complete my project. Some of these circuit blocks didn't work the first time so I had to build a new block.

There were various reasons the modules didn't work. Usually, I wasn't able to buy the exact parts used in the circuits I was copying. Or my craftsmanship or shielding wasn't adequate. Sometimes I never did figure out why one version of a circuit block was superior to another. By building my receiver using separate little shielded modules for each circuit block, I could replace a circuit block whenever I managed to build an improved version. Otherwise, I would have ruined the whole big board.

On rare occasions my circuits didn't work because there were errors in circuit diagrams in QST magazine or in the handbooks. I found some serious errors in my 1979 ARRL Handbook and a minor one in my 1998 edition. Perfect editing is not possible, so we shouldn't expect it. If painstaking R & D is new for you, prepare for a long battle. On the other hand, you'll learn a lot and victory will be especially sweet.

Mechanical construction

Homebrew receivers should be built in big cabinets. That way there's lots of room to add features and swap modules. The table in my shack is small, so I shoe-horned my receiver into a fairly compact package. In general, roomy boxes with lots of panel space are a much better idea. For example, when I discovered that my bargraph wasn't practical, there was no room on the panel for an analog meter. My long-range dream is to recycle my receiver modules into a larger cabinet. Eventually I want to add DSP, frequency counters, phase lock loops, and other goodies that intrigue me. With a big enough chassis, a receiver can grow and improve continually.

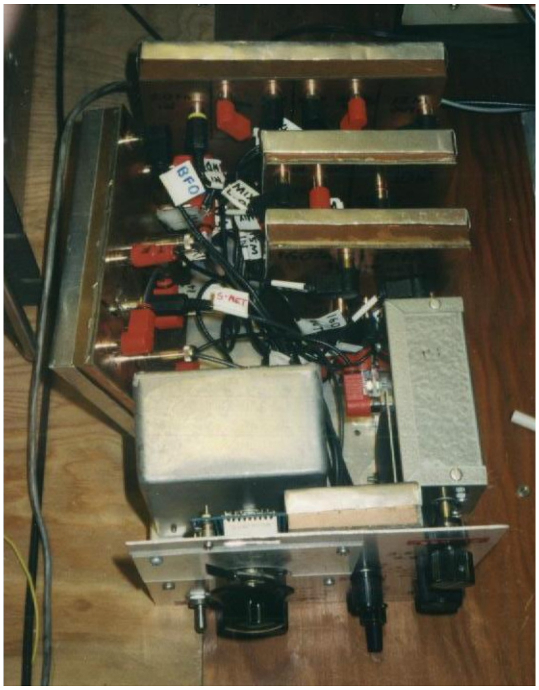

The receiver modules are interconnected with right-angle phono plugs and skinny coax.

Metal boxes shield each circuit block

A basic reason for building a receiver in metal-shielded modules is that capacitive coupling from one circuit block to the next can degrade performance. For example, my first 80-meter receiver module was built on one board. A crystal filter determined the selectivity. I made two "plug-in" crystal filters so I could use different filters for CW and for SSB. Using my 20 meter converter, I was tuning around 20 meters one day. I could hear lots of hams, but I was bothered by poor sensitivity and poor selectivity. I thought, "What's wrong with this receiver today?" I soon discovered there was no filter at all plugged into the 80 meter receiver board. What I was hearing was just stray coupling between the mixer and the IF amplifier. Amazing! So if you want band-pass filters with 50 dB skirt attenuation, you're going to need at least that much isolation between the stages. That means you need metal shields between all stages, coax interconnects and lots of bypass capacitors.

The metal-shielded modules could be small circuit boards mounted in commercial boxes. What I usually do is make shallow rectangular boxes out of pieces of two-sided circuit board material soldered together. The circuit is then carved into the floor of the box using small wood-carving chisels. The press-fit lid of the box is made from thin sheet aluminum folded over the corners of the PC board box.

Egg carton construction of the "mother board" of an 80 meter receiver

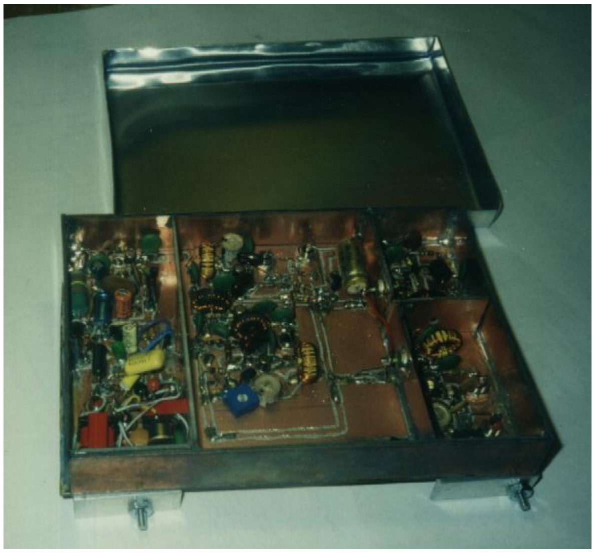

The compartment on the left is the product detector and audio amplifier. The large area in the center is the IF amplifier and AGC. The two modules on the right are the mixer and an optional RF pre-amplifier for the crystals.

If you plan on having more than one circuit block on the same board, you can isolate circuit blocks from each other using circuit board barriers soldered in place. The result is "egg-carton construction." Power can be routed between compartments using feed-through capacitors. If you are concerned that a circuit block might not work, you can wire your circuit on a separate square of PC board and then mount it on the floor of the desired compartment. All of these techniques are illustrated above. I constructed the IF amplifier in the center compartment on a separate PC board.

Shielded modules and shielded cable interconnects

For connections between stages I use thin **RG-174 coax and phono plugs**. Right angle phono plugs are not designed for RF and they have too much capacitance, about 4 picofarads. However, they're cheap, available, easy to wire, and don't take up too much room. I don't pretend that phono plugs are OK for 6 meters and VHF. I got some feel for their frequency limit when I discovered that switching from a phono plug to a UHF PL-259 on the input to my 50 watt transmitter amplifier vastly improved its performance on 10 meters. So far, I haven't seen any problems when working with lower frequencies and far lower power levels. Fortunately, receivers have tiny currents and low voltages. Most phono plugs have plastic bodies. That means that about 3/8 inch of the center conductor is not shielded. For my crystal filter module I used metal-bodied phono plugs that are at least a slight improvement over the plastic.

TV cable connectors, the ones that use the cable inner wire as a male connector pin, are electrically superior to phono plugs, but they are dreadfully intermittent. Personally, I've found them almost unusable and I wish the TV industry would junk them. There is a TV-compatible first-rate connector called an "F" type that is excellent, but pricey. In general, proper RF connectors like BNC, SMA, or TMA cost $2 - $6 each. Some of them are hard to assemble and your receiver could easily contain $300 worth of "proper" connectors. Also, most of these connectors are too long to fit gracefully in a small receiver. Right-angle phono plugs are quite short and quite cheap. Hurrah for phono-plugs!

Use plastic knobs

One odd little problem I encountered was that touching the metal control knobs or the front panel sometimes causes scratchy noise in the headphones when I listened on the higher bands. Yes, the metal panel was grounded and the chassis was wired to the station ground. The station ground is a heavy 12-gauge wire that grounds all the various metal boxes to a copper water pipe next to the station. I don't really have an explanation for the noises, but I switched from metal to plastic knobs and the annoying scratchy noises greatly improved. That was a shame - the metal knobs looked much more sophisticated.

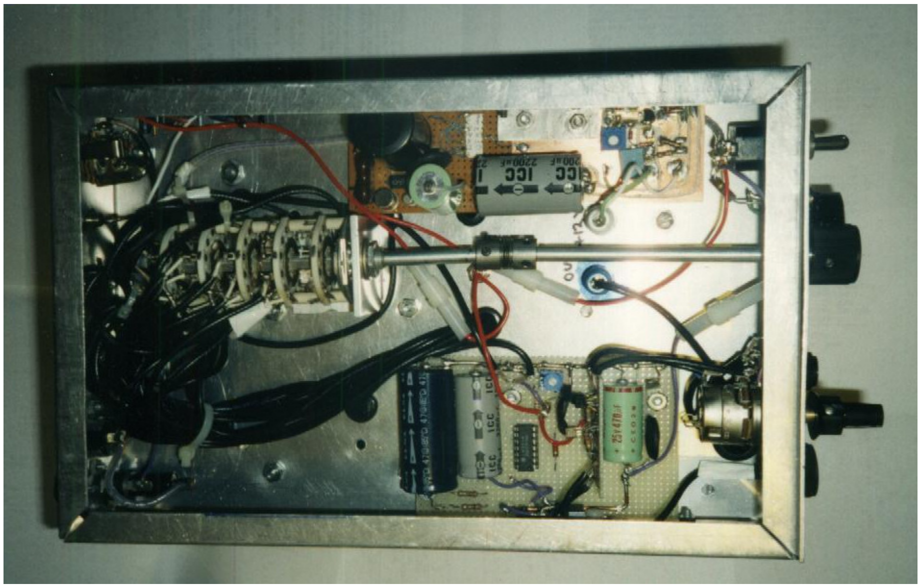

Bottom view of the receiver

Band switching and power supplies

The precision power supply for the VFO is at the top right. The low drop-out regulated supply for the rest of the receiver is at the bottom right. These are the same circuits used earlier with the transmitter VFO and QRP transmitter modules. The band-switch is the multi-wafer ceramic switch on the left. The black wires on the left are skinny coax that interconnect the inputs and outputs of the converters for every ham band except 80 meters. It is desirable to cover the bottom with a metal plate to help keep stray signals out of the power leads.

80 Meter input preselector

The "front end" of the 80 meter receiver is a mixer. No RF pre-amplifier is needed on 80 meters because, if the receiver works well, then the atmospheric noise coming in from the antenna will be louder than the receiver internal noise. In this situation an RF amplifier won't help. However, the mixer does need a sharp bandpass "preselector" filter to keep out the low frequency AM radio and limit the input signals to 3.5 to 4.0 MHz. It's especially important to filter out AM broadcast stations. These stations put big voltages on the antenna and tend to overwhelm the mixer unless they are well attenuated. The high pass AM broadcast filter module described in Chapter 7 for the direct coupled receiver can be used and works well.

The mixer subtracts the VFO frequency, (5.0 to 5.5 MHz) from the IF frequency, (9 MHz), to tune 80 meters, (3.5 to 4.0 MHz). The two inputs to the mixer are the VFO signal and the antenna signal. When I first examined the 1986 ARRL design, I was disappointed to see that the pre-selector had a primitive variable capacitor that the operator was supposed to tune for maximum gain for a particular part of the band. After all, the bandpass filters for the other HF bands were fixed and not accessible from the front panel. I attempted to build my own fixed bandpass filter, but my filters had too much attenuation (poor sensitivity) and sometimes let in AM broadcast stations. That is, it was like listening to a crystal set.

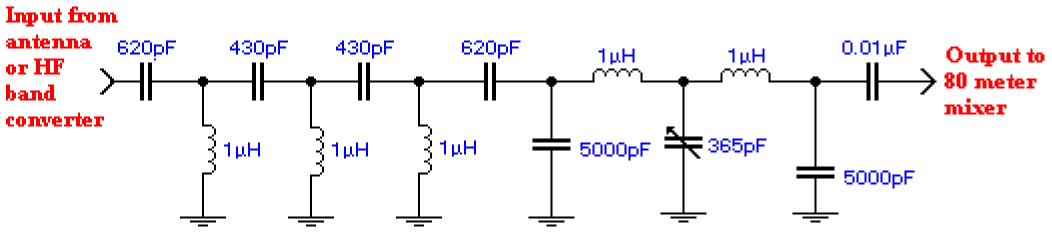
Input from antenna or HF band converter
620pF
430pF
430pF
620pF
1µH
1µH
0.01µF
Output to 80 meter mixer
1µH
1µH
1µH
5000pF
365pF
5000pF

The recommended 80 Meter preselector filter for the core receiver mixer input.

So I returned to the ARRL design with the 365 pF variable capacitor. It had so much attenuation on 80 meters, I couldn’t hear a thing. I ran the ARRL circuit on a Spice program, and, according to Spice, it should have worked well. However, mine didn’t. I must have had a bad part somewhere, but I couldn't find the problem. Using trial and error, I removed some parts and ended up with the circuit shown below. My filter works pretty well, although, according to my Spice program, it is not as good as it should be. Sigh ...

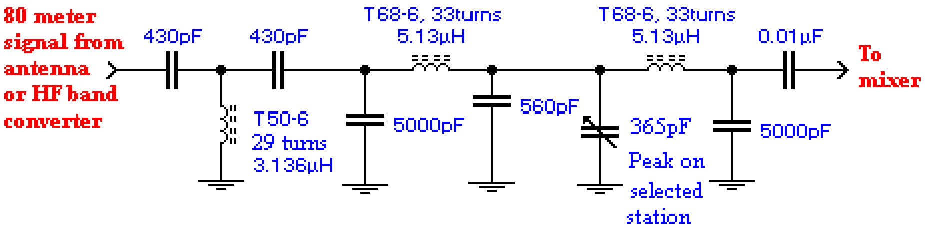
80 meter signal from antenna or HF band converter
430pF
430pF
T68-6, 33turns
5.13µH
T68-6, 33turns
5.13µH
0.01µF
To mixer
T50-6
29 turns
3.136µH
5000pF
560pF
365pF
Peak on selected station
5000pF

My version of the 80 Meter preselector filter for the mixer input

I mounted the preselector in a shielded box up on the front panel where the peaking capacitor was accessible. Someday, when I figure out how to build a better preselector, I shall replace the whole module.

The 365 pF variable capacitor acts like an attenuator

On the positive side I found that the variable peaking capacitor is quite useful as an attenuator for receiving strong single sideband signals. In other words, strong SSB phone signals are often much more intelligible when the preselector is mistuned and signal strength is decreased. If I didn't have this capacitor as a variable sensitivity control, I would have to build an attenuator.

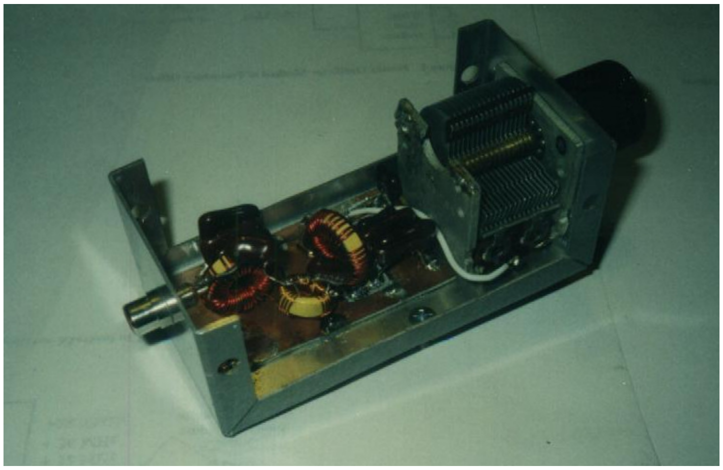

The preselector is built in a little box up front behind the front panel.

The variable frequency oscillator (VFO)

The receiver VFO is the same design as the 5 MHz transmitter VFO discussed in chapter 10. The big tuning knob controls the VFO. Actually, in superhetrodynes the VFO is usually called a ***local oscillator*** or ***LO***. The range and stability of the VFO determine what VFO and IF frequencies are practical. Like a transmitter VFO, a receiver VFO should be stable to less than 5 Hz/ minute drift, although less than 20 Hz will probably be OK.

Unfortunately, if the VFO frequency is too low, it probably won't span enough Hz to cover the bands you're interested in. Notice that 10 meters is so huge, 1.7 MHz wide, that you will have to cover it with multiple converters. So far, my receiver just tunes the first 500 KHz of 10 meters, which includes all the CW activity. I guess I'll worry about the rest when I get on 10 meter SSB. And that will probably happen when the sunspots return. The VFO frequency and its harmonics should avoid the ham bands as much as possible. Frequencies with round numbers are best because they make calibration easy. Put all these factors together and you'll find your choices are fairly limited.

In fact, once you've built separate transmitters and receivers and used them on the air, the disadvantage of having to tune two VFOs makes it obvious why most modern rigs are transceivers. When planning a homebrew receiver and transmitter, you might consider using a common VFO module. A common VFO would be a great help on the air. But, if you use just one VFO, you will have to master the 500 to 800 Hz send/ receive frequency-offset problem. Also, an isolation amplifier is needed to keep the cable connecting one to the other from loading down the receiver VFO. And finally, when used on the upper bands, each crystal oscillator in the converter and each transmitter PMO must be on the corresponding frequencies in order for the receiver to listen on exactly the same frequency used by the transmitter. In other words the converter oscillators need to be shared too. All this represents huge amounts of work. It makes the most sense to build a transceiver from the beginning, starting with a gigantic mainframe cabinet.

How much VFO signal do you need?

As explained in chapter 10, the stability of a VFO is partly dependent on generating the minimum heat possible inside the VFO box. Therefore, unlike me, you should decide out how much VFO signal voltage you will need before you build the VFO. The required VFO voltage is the level needed to drive your mixer. MOSFET and JFET mixers only need 1 or 2 volts peak, so if you're going to use one of these, there is no good reason to generate 5 volts, then throw away most of it in a potentiometer at the input to the 80 meter mixer. Instead of running the VFO on 12 volts or 5 volts, you can reduce the supply voltage externally to 3 volts or whatever minimum voltage you actually need. Remember that low voltage means low heat which means low drift.

A varactor tuned 5 MHz VFO. Tuning is accomplished with the big round potentiometer.

Mixer magic

The purpose of a mixer is to translate the frequency of an incoming radio signal to a constant intermediate frequency (IF) that can be amplified and filtered more easily. Mixers combine a local oscillator sinewave with the incoming radio signal to make a composite signal. The new signal contains the original frequencies, plus the new sum and difference frequencies. Mixers intended for moving a VFO up to a high band were described in chapter 11. Mixers for that purpose can be quite crude and will work well. Unfortunately, receiver mixers are much more difficult because the incoming signal can be so tiny.

One way to look at mixers is that a big local oscillator sinewave keys the incoming RF signal totally on and off, much like a sequence of fast Morse code dots. The lesson is that the local oscillator (the VFO tuning knob) must be a big signal while the RF input signals may be arbitrarily small.

The ARRL Handbook presents several different mixer designs made with discrete diodes, inductors, and transistors. However, in most ARRL receiver designs since the 1980s, including the W7ZOI/K5IRK receiver, the mixer is an integrated circuit or little canned assembly labeled “mixer.” I guess everyone else was having mixer trouble too, so they resorted to integrated circuits. Two years ago there was a receiver project in QST that used an IC that contained both the mixer and the VFO. I’ll bet these marvels work fine, but the contents of these ICs are a mystery. Use one if you want.

Mixers will give you lots of static... and squeals, howls, and squawks

So far I’ve built five different mixer designs from discrete parts. First I built a classic balanced mixer with ferrite cores and a hot carrier diode ring. When I turned it on, I heard loud, roaring static in the headphones. “Oh goodie!” I thought, “Listen to all that atmospheric static! It must be working!” I soon figured out that the static was coming from the mixer and the IF amplifiers and had nothing to do with the outside world. I had just learned a basic truth about mixers: Mixers aren’t just prone to generate “a little background noise.” They often produce gigantic Niagara Falls noise that obscures everything coming in the antenna. However, once I had proper mixer input levels and resonant circuits tuned up as best I could, the noise disappeared and I began to hear stations. Unfortunately, as I tuned across the band, there were loud whistles like marker beacons every few KHz. In between the whistles, I could sometimes barely hear strong stations.

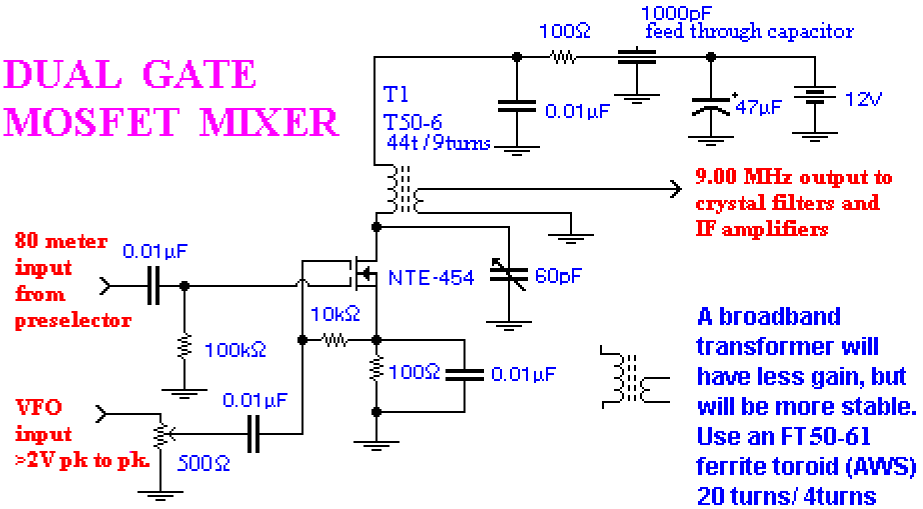
DUAL GATE MOSFET MIXER
100Ω
1000pF feed through capacitor
T1
T50-6
44t / 9turns
0.01µF
47µF
12V
9.00 MHz output to crystal filters and IF amplifiers
80 meter input from preselector
0.01µF
NTE-454
60pF
10kΩ
100kΩ
100Ω
0.01µF
A broadband transformer will have less gain, but will be more stable. Use an FT50-61 ferrite toroid (AWS) 20 turns/ 4turns
VFO input >2V pk to pk
0.01µF
500Ω

A practical mixer

My favorite mixer is shown above. Most of the others suffered from noise and “birdies” and had poor sensitivity as well. Unlike diode mixers, the operation of the MOSFET mixer is obvious. It is essentially an ordinary tuned transistor RF amplifier. The radio signals come in on one control gate. This modulates the large current passing from drain to source of the transistor. The small voltage on the control gate controls the large drain current thereby amplifying the original signal. A second input gate amplifies the local oscillator signal. This means that 2 volts of VFO signal is plenty. The local oscillator signal is so strong it turns the drain to source current totally on and off, “chopping” the input RF signal into tiny segments. The big output current from the transistor becomes an amplified “mixture” of the two input signals.

I originally used a tuned version of this mixer in which the transformer primary was tuned with a capacitor to resonate at 9 MHz as shown above. That works OK, but is rather tweaky and subject to noise and birdies. I now prefer an untuned version (also shown above) which has slightly less gain, but isn’t so easily misaligned. Instead of the powdered iron core and 60 pF trimmer, I used a CWS (Amidon) FT50-61 ferrite core with 20 turns primary and 4 turns secondary.

Because of the high gain of the MOSFET, the VFO sinewave signal can be low amplitude, typically 2 volts peak, and still fully chop the radio signals on and off. In contrast, a diode ring mixer needs a big local oscillator signal, 12 or more volts peak, to chop the signal.

Other transistor mixer designs use junction FETs or bipolar transistors. These designs use the emitter resistor as the VFO input port. The emitter or source has no gain, so these designs also need big local oscillator signals.

A small local oscillator signal is especially helpful for building the converter or converters you will need for each of the other nine HF bands. Each of these converters must work through its own mixer. If the dual gate mixer were not used in the converters as well, each of the eight local oscillators you will build would need separate amplifiers to get the local oscillator signal up to 12 volts peak. I used a common, broadband pre-amplifier for the crystal oscillator inputs to a common untuned mixer. That is, the mixer has a ferrite, broadband transformer as described above.

All dual gate MOSFETs are not equal

Alas, a dual gate MOSFET mixer isn't a guaranteed success either. When I first built a MOSFET mixer, I couldn't buy any of the transistors recommended in the handbook. I first tried a generic part, the NTE221 transistor. This produced the usual oscillations and insensitivity. I was getting discouraged, but I tried the similar NTE454 and *IT WORKED!* The only obvious difference in the specifications was that the gate shut-off voltage was smaller. In other words, the NTE454 was more sensitive. Since then, I've discovered the NTE222 seems to work as well as the NTE454. The NTE455 seems too sensitive. In my circuit it produced whistles, birdies, and noise. On the other hand, the NTE455 worked great as a product detector (Chapter 7).

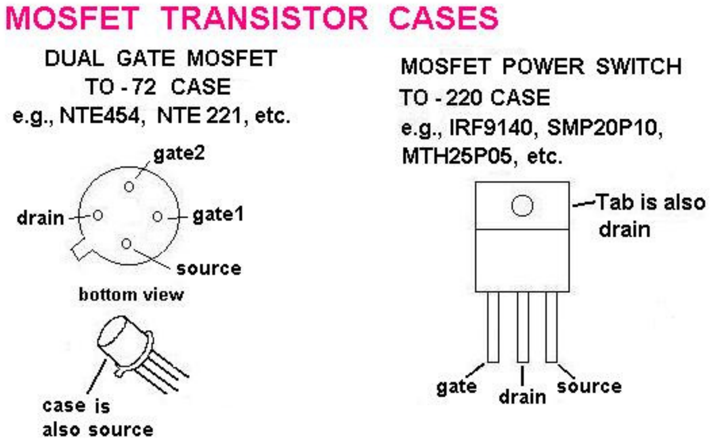

I first tried to use two dual-gate MOSFETs in a "semi-balanced" configuration. The handbook said this is superior to the simple mixer shown above. Balanced designs are supposed to help cancel images, by that I mean, birdies. My semi-balanced mixer produced the usual oscillations, but at least the noise and sensitivity were adequate and I began to hear weak 80 meter stations with my receiver.

Reading up on mixers, I learned that mixers are only happy when they receive the exact input levels. That's why I put a pot on my VFO drive to inject the optimum level. As I turn up the VFO drive to the mixer the output signal strength rises abruptly then levels off. Higher levels of VFO contribute only slightly more gain, but much more noise. I adjust the VFO input to where the gain first begins to level off. (Note: My VFOs are designed to work into a 500 ohm load, hence the 500 ohm pot.)

The sensitivity of mixers to ideal levels explains why most modern transceivers have input attenuators so that they can be adjusted to tolerate strong signals. I got a QSL card from a guy who wrote, “Sorry about the 529 signal report. After we signed off, I discovered I had the attenuator on.” As mentioned earlier, the 80 meter preselector filter may be deliberately mistuned so that it acts as an attenuator to limit signal strength.

Note: Reception on 80 meters and 160 meters is best with a tuned transmatch

By accident I discovered that reception on the two lowest HF bands is much better when the receiver is sharing the antenna with the transmitter and the antenna is tuned with the “T” antenna coupler described in chapter 9. In my neighborhood at least, the signals from the local AM radio stations are so high that they tend to overwhelm the 80 meter mixer. This results in a lack of audible signals on 80 and 160. I didn't realize I had a problem because I wasn't hearing the AM stations in the headphones. However, when I tuned up the transmatch, suddenly numerous ham signals appeared. The obvious conclusion is that my receiver preselector filter is not selective enough. Even my old Collins receiver is greatly improved by a tuned antenna coupler.

In chapter 7 I described a high pass filter for the direct conversion receiver to attenuate the AM broadcast signals. I have found that this little filter also improves the 80 meter receiver board performance. Because I had built it in a little module with phono plug connectors, I could simply plug it in without building anything new or without altering the rest of the receiver.

A JFET mixer

Since this book was first written, dual gate MOSFETs have become increasingly expensive and hard to find. One part you might try is the RCA 40673 available from Mouser for $7. I haven't tried them, but my friend Jayram, VU2JN in India, has used them in receiver mixers with great success. Dual gate mixers are his favorite design too.

In the event that you need a substitute design, here is a JFET circuit that works but in my opinion, isn't as sensitive. It is practically the same circuit, but instead of introducing the local oscillator signal into a separate gate, it is introduced across the source resistor. Surprisingly the optimum local oscillator signal input level for this circuit is only about 1 volt peak-to-peak. For this reason the signal is first passed through a 500 ohm pot. I would have expected that, for a circuit like this, the optimum peak voltage would have approached the supply voltage so that the transistor would be entirely turned on and off. I guess the answer is that this is a ***depletion type*** MOSFET that is already half turned off with no bias. Consequently, it doesn't take as much drive to turn it either full on or full off. I also attempted to use the dual parallel JFET PMO mixer design described in chapter 11, but it worked poorly in my receiver.

There is another dual JFET mixer circuit that puts the two transistors in series. In this way, the two voltage signals both have gain and are stacked on top of each other, so their output voltages add. In contrast, the parallel JFET version produces signals that tend to cancel each other. I tried this variation and it seemed no better than the single JFET version above. In other words, it was insensitive. It looks as though it would produce more signal strength and I always wonder if I have missed something. Perhaps with more R&D it can be improved.

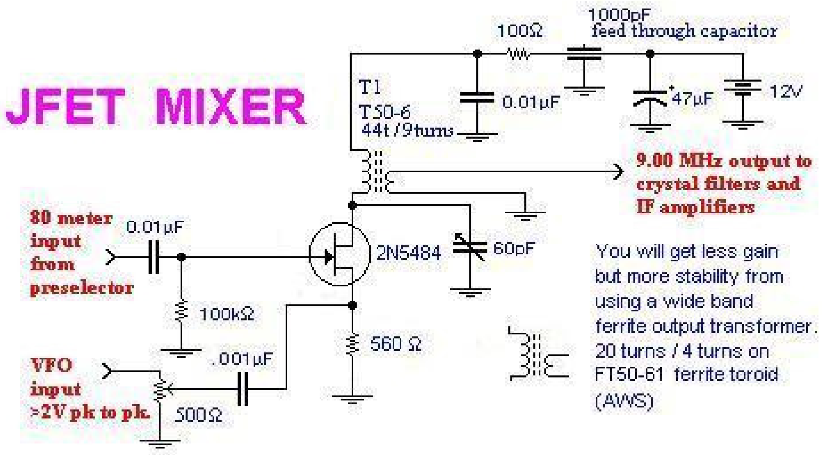
JFET MIXER
100Ω
1000pF
feed through capacitor
T1
T50-6
44t / 9turns
0.01µF
47µF
12V
9.00 MHz output to crystal filters and IF amplifiers
80 meter input from preselector
0.01µF
2N5484
60pF
100kΩ
560 Ω
VFO input >2V pk to pk
.001µF
500Ω

You will get less gain but more stability from using a wide band ferrite output transformer. 20 turns / 4 turns on FT50-61 ferrite toroid (AWS)

Crystal filters and BFOs

Crystal IF filters give you the selectivity you will need for working CW. They eliminate interference from nearby stations and also eliminate a great deal of the atmospheric noise. The output from the superhetrodyne mixer is a weak, broadband, IF frequency signal that needs amplification and filtering before it is ready to be detected. The bandpass filtering is usually done right after the mixer.

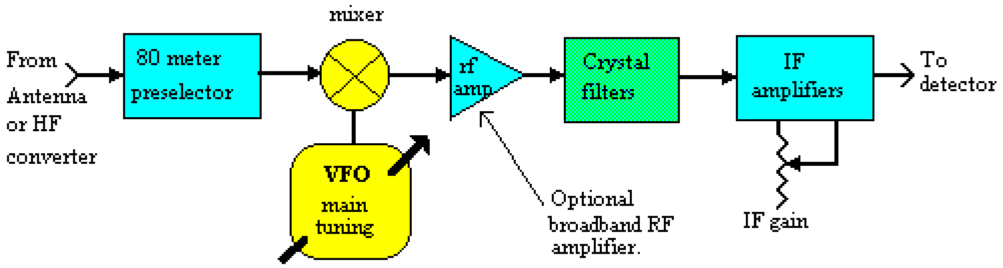
mixer
From Antenna or HF converter
80 meter preselector
rf amp
Crystal filters
IF amplifiers
To detector
VFO main tuning
Optional broadband RF amplifier.
IF gain

Location of crystal filters in a superhetrodyne

The filter could be a "mechanical filter" if you are using a low frequency IF like 455 KHz. But if your IF is 9.0 MHz like mine, then you'll need one or more crystal filters. Before I describe building crystal filters, I'll discuss the Beat Frequency Oscillator (BFO). You will probably need the BFO as a tool to select the crystals for your filters.

The Beat Frequency Oscillator (BFO)

A beat frequency oscillator is an RF oscillator that operates on the intermediate frequency of a superhetrodyne. The BFO mixes with the IF signal to make CW and single sideband transmissions audible and/ or understandable. Without the BFO CW signals would be inaudible or just thumping noises at best. Single sideband phone would be unintelligible “Donald Duck” sounds. In single sideband the transmitter filters out the basic carrier frequency leaving just one of the modulation sidebands. The BFO serves to restore the carrier sinewave, in effect returning the sideband signal to its original amplitude modulation.

During detection, the audio signal passed on to the loudspeaker is the difference between the IF frequency and the BFO frequency. For example, when listening to CW signals, the IF frequency might be 9.000,000 MHz. The BFO frequency might be 9.000,**700** MHz. What you would hear in your headphones is a musical tone of the difference frequency, 700 Hz. If that pitch sounds unpleasant to you, adjust the BFO frequency to say 9.000,500 MHz to produce a musical tone of 500 Hz. For the musical tone to stay constant, the BFO oscillator must be quite stable. Therefore, we use a crystal oscillator and pull the frequency up or down using a variable capacitor, just like we did in the crystal controlled QRP in chapter 6.

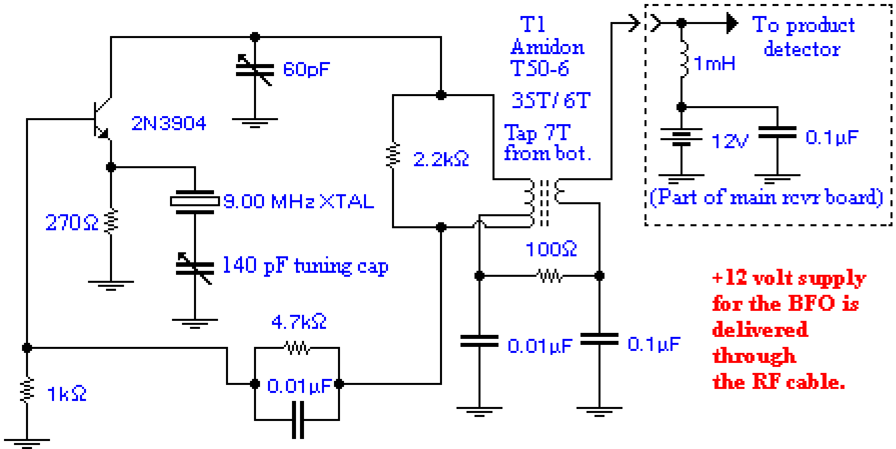

+12 volt supply for the BFO is delivered through the RF cable.

The Beat Frequency Oscillator

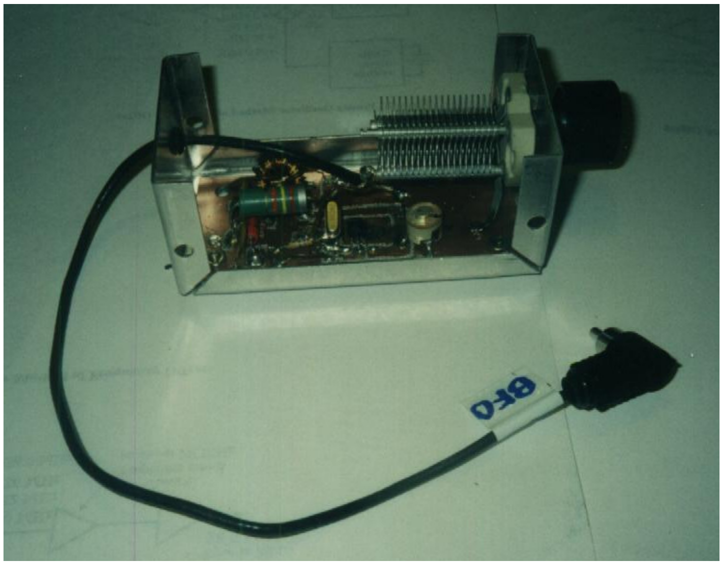

The Beat Frequency Oscillator (BFO) module

This BFO is taken directly from the W7ZOI and K5IRK receiver in the 1986 ARRL Handbook. Its unusual feature is that its DC power supply rides on the same line as the RF output. This makes it easy to install the BFO in a little metal box up on the front panel remote from the main receiver board. A variable capacitor on the front panel "pulls" the BFO crystal frequency above and below the nominal frequency. The BFO, together with the crystal filter, allow you select upper or lower sidebands. The BFO is connected to the main board with a length of thin coax jumper cable. If you wish to use this oscillator for matching 9 MHz crystals for filters, I suggest you install the BFO crystal in a small IC or transistor socket.

The BFO frequency tuning range should extend above and below the bandpass of your crystal filters. ***When the BFO frequency is below the center of the filter bandpass, you are listening to the upper sideband. When the BFO is tuned above the filter bandpass, you're hearing the lower sideband.*** Adjusted for the upper sideband, tuning down the CW band will cause the whistle pitch of a signal to start high, drop down to a low pitch, and then disappear. Adjusted for the lower sideband, tuning down the band will cause the pitch to start low then climb up high and disappear. When the BFO is tuned to the center of the filter bandpass, the tone will start at a medium pitch, drop to a low pitch, briefly become inaudible, then climb back up to medium pitch and disappear.

When tuned to sideband phone signals, the signal is harder to make understandable if you select the wrong sideband. Consequently, it will be important to calibrate the BFO tuning knob so that you will know when you are listening to upper or lower sideband.

Ladder filters

Building crystal filters was easy, once I figured out how. Most commercial transceivers use modular crystal filters that have specific bandwidths and are sealed in little cans, something like an integrated circuit. I made my filters from discrete crystals.

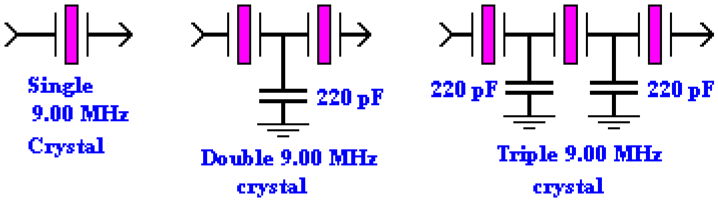

One, two, and three crystal “ladder filters”

A ladder filter is just two or more crystals in series with capacitors bypassed to ground at the nodes. The bandpass width is inversely proportional to the number of crystals and the capacitance to ground. ***In general, the lower the capacitance on the nodes, the wider the bandwidth.*** Large capacitors, like 50 ohms reactance, will give narrow bandwidths and higher attenuation. Long “ladders” of course produce even narrower bandwidths and more attenuation. If all the crystals are all identical, then the “skirt attenuation” on either side of the peak passband becomes steeper and steeper as more crystals are added.

Crystal filters with two or three crystals are selective enough for “CW after dinner.” By that I mean you can work the busy CW bands early in the evening with adequate signal separation. A single crystal filter made from just one 9 MHz crystal is good for single sideband phone (SSB). With three or four crystals the bandwidth becomes so narrow that SSB is totally unintelligible. With strong CW signals and lots of QRM, triple or quadruple crystal ladders are extremely useful. If you tune down to the bottom of 20 meters in the evening, you’ll usually hear a cacophony of CW stations all trying to work the same 5 or 6 DX stations in exotic countries. With only one crystal switched in, you’ll hear practically everyone at once. Switch in two and three crystals and suddenly, you’re not only listening to one clear station, but most of the background noise has gone too.

How many crystals can you use in a ladder?

One limit on how many crystals you can use in series depends on the precision with which you match the crystals. I didn’t match my first set of filters and they produced more attenuation than filtering. I then proceeded to waste time getting more gain out of my IF amplifier. Then, after I achieved the gain, the selectivity wasn’t much better than single crystals.

Finally, I tested my crystals one by one by putting them in the BFO and measuring the frequency with a frequency counter.

I didn't expect that the frequency filtering characteristics would be exactly the same frequency as when the crystal is used in an oscillator. However, I figured I could at least select sets of crystals that were similar. When I ran them in the oscillator, I was shocked to discover that the crystals were as much as 2.5 KHz different! No wonder it worked so poorly. I had put 9.001 MHz crystals in series with 9.003 MHz crystals. I had built a "crystal barrier," rather than a crystal filter.

Luckily I had bought twenty 9.000 MHz microprocessor crystals. That sounds extravagant, but they're under $1 each at Digi-Key or Mouser. Because I had a wide selection of 9.000 MHz crystals, I was able to match up two crystals that were within a few Hz of each other. And I was also able to match up a group of 3 crystals that were within 50 Hz. This time when I put matched crystals in my ladders, the improvement was dramatic. When I switched from a single crystal to a double crystal, the signal strength barely dropped. With three crystals, the signal strength just dropped slightly more.

In theory, you can make perfectly matched crystals by placing tiny trimmer capacitors in parallel (or series) with each crystal. Then you would trim the crystal and capacitor assemblies in the oscillator one at a time so that each crystal/capacitor assembly oscillates on precisely the same frequency. I've only tried this once, but it worked poorly and I went back to the simpler design.

Another limit on the number of crystals in a filter is the shielding and RF isolation between the mixer and the IF amplifier. If the shielding is poor your IF amplifier will "hear" the signal from the mixer without the signals ever passing through the crystal filters. In my receiver, 5- or 6-crystal ladders aren't worth building.

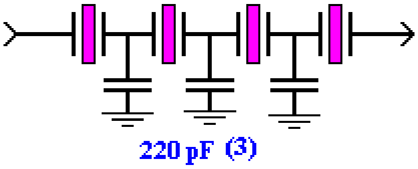

Four matched 9.000 MHz crystals

Switch in your filters with a rotary switch

I first built my filters as "plug-ins," but I soon discovered it was too hard to change them in the middle of a QSO. Eventually I wired them to a rotary switch in a shielded box. As I became more experienced with using the filters, I began to use the triple filter more and more. Finally I built a quadruple filter and now I use it routinely. I find it works well with the S meter (strength meter) as a way to tune my transmitter to zero beat with another ham's signal. I just sweep my transmitter VFO across the band until the S meter jumps up to maximum. This occurs when the VFO frequency converted into the IF matches the frequency of the four matched crystals. The offset of the BFO and the tone of the Morse code signal take care of themselves. That is, if the fellow is on the upper sideband, then the S meter only responds to my VFO when I am on the same BFO offset that he is using. This happens because only one sideband is audible at a time with 3- or 4-crystal ladder filters. By the way, one of my 20 crystals behaved erratically. While I watched, the frequency wandered about +/- 200 Hz. Consequently, I didn't use it.

Series and parallel cut "XTALS"

There are two kinds of simple crystals, series and parallel cut. As I understand it, the difference is the oscillator circuit for which they are designed to be used. For example, a series-type crystal is intended to be used in series with a specific capacitance in an oscillator. When this exact capacitance is used, it will oscillate at the rated frequency, for example, 9.000 MHz. In contrast, if you use a parallel cut crystal in the same circuit, it might oscillate at 9.004 MHz. You may use either kind of crystal, but your filter frequency may not be exactly 9.000 MHz.

All 9.000 MHz crystals aren't equal

In my experience larger crystals, size HC-49 or bigger, work well. The tiny surface mount or half-size HC-49 crystals need more signal strength from the mixer output to push the signals through the filter. The broadband "optional" amplifier shown below will give you extra gain if you need it. It's the same design that will be used later as a RF amplifier for the higher HF bands. If you are short of MOSFETs, a JFET alternate circuit is also shown below.

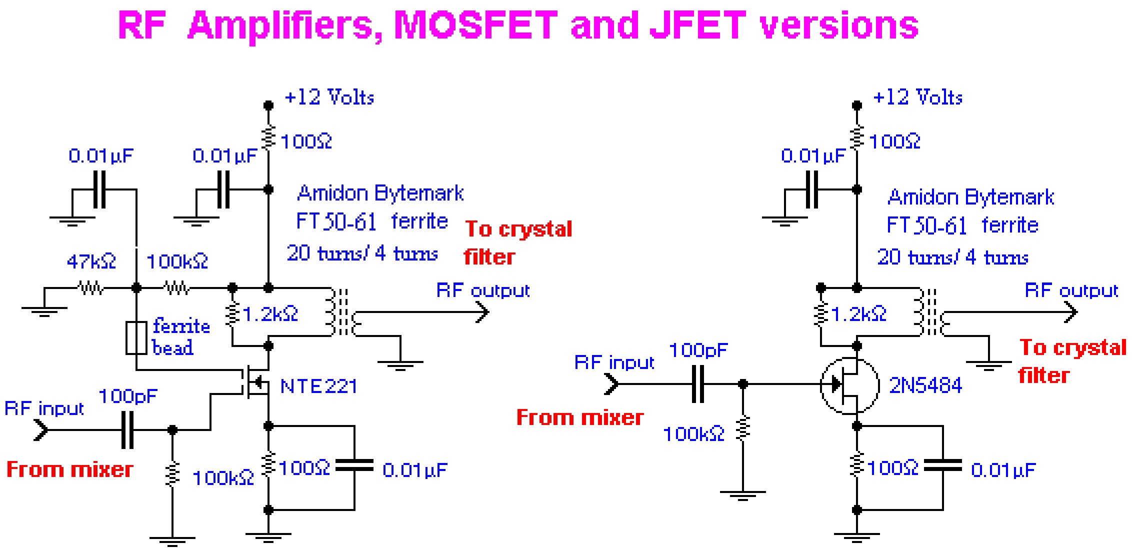
RF Amplifiers, MOSFET and JFET versions
+12 Volts
100Ω
0.01μF
0.01μF
Amidon Bytemark
FT50-61 ferrite
20 turns/ 4 turns
To crystal filter
47kΩ
100kΩ
1.2kΩ
RF output
ferrite bead
NTE221
RF input
100pF
From mixer
100kΩ
100Ω
0.01μF
+12 Volts
100Ω
0.01μF
Amidon Bytemark
FT50-61 ferrite
20 turns/ 4 turns
1.2kΩ
RF output
To crystal filter
RF input
100pF
2N5484
From mixer
100kΩ
100Ω
0.01μF

Broadband RF amplifiers Place one between the RF mixer and the crystal filter.

I also observed a difference between different brands of crystals. My crystals from the ICM company were extremely similar to each other. For ladder filters, this would be the brand to buy. My crystals from the ECS company had much wider frequency variation from one crystal to the next. This isn't helpful for building ladder filters, but if you need crystals that are slightly different from each other to build wider pass-band filters, or perhaps SSB oscillators that must operate 2.5 KHz above and below the nominal frequency, then buy the ECS crystals. I found them both useful.

By trial and error I found that the untuned impedance step-down transformer shown in above pre-amp circuit performs better than leaving the transformer out and simply using an inductor on the drain then coupling to the crystals with a capacitor. In other words, this implies that the crystals with their bypass capacitors look like roughly 50 ohms, but must be matched to the high impedance output of the MOSFET transistor. There's a design in my ARRL handbook that uses impedance step-up transformers to match the filter. I find it hard to believe that design is optimum unless the crystals they used behave differently than mine.

The second input gate in the above dual gate amplifier is used to set the DC bias and make the amplifier class A. A voltage divider delivers about 4 volts DC to the gate. The ***ferrite bead*** is a tiny inductor (RF choke) that helps insure that the MOSFET doesn't oscillate. The ferrite bead is literally a 1/8 inch cylinder with a tiny hole through the center. For example, you could use a CWS type (Amidon) FB43-101 bead. The type isn't critical. I used several different kinds of beads and have had no trouble with oscillation. If it does oscillate, remove the 0.01 µF bypass cap from the 100 ohm source resistor. The resulting negative feedback should kill the oscillation at the expense of a small amount of gain.

The IF amplifier

The IF amplifier is another tricky part of a superhetrodyne. It's a high-Q (sharply resonant) amplifier that must handle signals with a range of 100 dB or more without oscillation or noise. This is a huge dynamic range. The gain on the IF amplifier stages should be adjustable using an IF gain control. Too much gain and you will have noise and squeals. Too little gain and you can't hear weak DX stations.

Moreover, if you used miniature half-size HC-49 crystals to build your bandpass filters, you will need even more gain to stuff signals through the significant attenuation of the filters. In the last section I described a simple RF amplifier that can be placed between the mixer and crystal filter to overcome this difficulty.

Oscillations in an IF amplifier come in several flavors. As you tune the LC circuit of an IF amplifier stage you will hear squeals, harsh roaring, silent dead spots, and gentle static. The setting that brings in the loudest signals is surprisingly noise free. The first time I turned on my receiver, I quickly learned that ***most of the receiver noise is coming from the mixer and IF amplifiers***, not from the outside world. The noise comes from a maladjusted mixer or too much IF amplifier gain.

Although I was able to tune up my 80 meter receiver using a signal generator, real 80 meter ham signals worked the best for me. There's no point in simulations when you have the real thing. A problem with tuning up the IF on 80 meters is that 80 may be dead during the day. During the summer, 80 may not be so hot in the evening either. As a result, you might consider building a converter for 20 or 40 meters early in the project. For most of the 11 year sunspot cycle these bands usually have signals anytime, day or night, all year long. Since your 80 meter receiver may not yet be working, you can adjust your converter by feeding the output into a commercial receiver tuned to 80 meters. Then, after you have the converter working, you'll be confident that there are lots of real signals for your 80 meter receiver to hear.

Impedance matching the crystal filter to the IF amplifier

Looking at examples on crystal filter circuits in handbooks from various years, I found circuits that seem to assume the filters are low, medium, or even high impedance. In most of my attempts, I seemed to get the best gain when my filters were assumed to be relatively low impedance, say 50 to 100 ohms. That's why the optional amplifier described above used a step down transformer output. I tried step-up, step-down and no transformer to feed the signal into the IF amplifier shown directly below. Step up worked best on the two versions as shown below.

A dual-gate MOSFET IF amplifier

My first versions of the IF amplifier used dual gate MOSFET amplifiers, similar to the crystal filter preamplifier described earlier. Just what you need - another circuit made from 7 dollar transistors! The gain of each MOSFET transistor can be controlled by varying the DC bias on one of the two control gates. This control voltage can be generated by either the IF gain knob or by the automatic gain control circuit. In short, the dual gate MOSFET looks ideal for IF stages. Unfortunately, I had lots of trouble with squeals and noise.

With struggle, I eventually managed to get the dual-gate version working. Originally it consisted of two dual-gate tuned amplifiers in series. Each amp was similar to the dual gate amplifier on the previous page. Instead of the fixed 4 volt bias on the second gate, those gates were biased with the variable IF gain/AGC voltage. When I started, I used tuned LC circuits as the impedances between the + supply and the two MOSFET drains. This approach was quite unstable and prone to oscillations. I decreased the instability but retained enough of its sensitivity by replacing the tuned drain LC circuits with ferrite cores - that is, RF chokes. I moved both tuned LC circuits to the input and output respectively. These LCs must be tuned carefully for the maximum signal strength without squeals. The circuit below is the result.

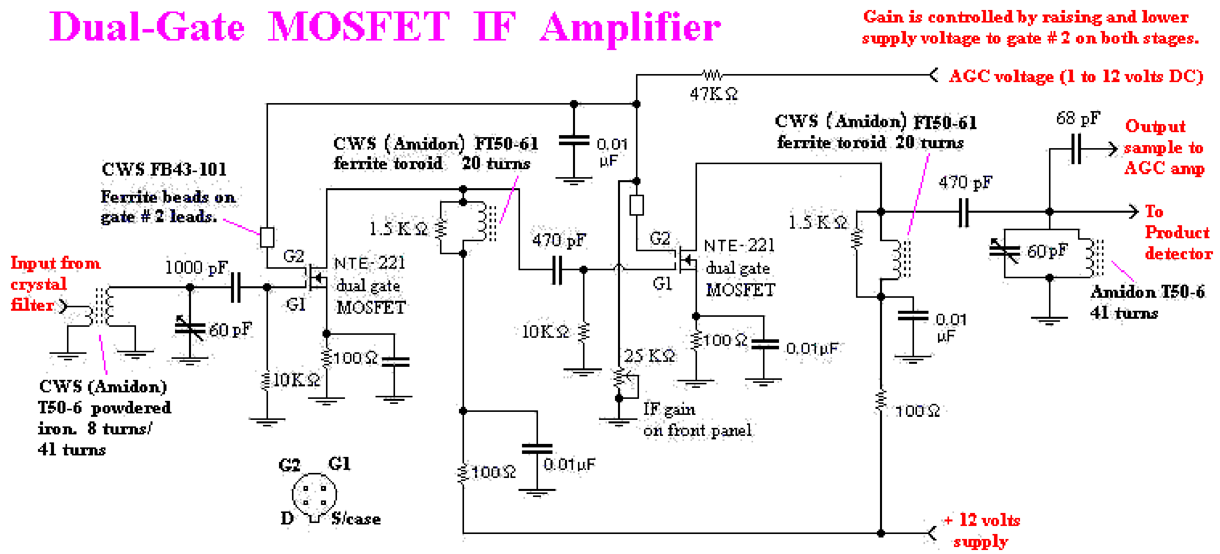
Dual-Gate MOSFET IF Amplifier
Gain is controlled by raising and lower supply voltage to gate # 2 on both stages.
AGC voltage (1 to 12 volts DC)
47KΩ
CWS (Amidon) FT50-61 ferrite toroid 20 turns
0.01 µF
CWS FB43-101 Ferrite beads on gate # 2 leads.
CWS (Amidon) FT50-61 ferrite toroid 20 turns
68 pF
Output sample to AGC amp
470 pF
To Product detector
1.5 KΩ
G2
NTE-221 dual gate MOSFET
G1
470 pF
1.5 KΩ
G2
NTE-221 dual gate MOSFET
G1
60pF
Amidon T50-6 41 turns
Input from crystal filter
1000 pF
60 pF
10KΩ
100Ω
10KΩ
25 KΩ
100Ω
0.01µF
0.01 µF
CWS (Amidon) T50-6 powdered iron. 8 turns/ 41 turns
IF gain on front panel
100Ω
G2 G1
D S/case
100Ω
0.01µF
+ 12 volts supply

The main lesson I learned from this circuit was that high gain amplifiers shouldn't be tuned unless the input signal is continuous. Whenever there is little or no signal to amplify, a high gain tuned amplifier will amplify any noise that may be present and may begin to oscillate at the tuned frequency. To prevent this, you are better off designing your tuning function to be separate from the amplification. Notice the tuned circuits at the input and output of the IF strip as a whole. The two amplifiers above are wideband but relatively low gain.

I suspect that a similar circuit would work OK with ordinary bipolar transistors, but I haven't tried it. The IF gain would control the forward bias level of the two transistors, that is, the bias that makes them class A. A gain/ bias system like this is used in the cascode amplifier shown below.

In the circuit above, the input LC attempts to limit input noise to just 9 MHz noise. Similarly, the output noise is similarly clipped with the second LC. Chapter 15 discusses PMOs for single sideband generation, which have a similar problem. Every time there is a pause in the speech being transmitted, a tuned amplifier will usually oscillate. Consequently, all the frequency filtering is done in separate passive filter blocks, rather than being incorporated into the amplifiers.

In summary, the above amplifier stages are broadband and use ferrite transformers that pass on any radio frequency signal to the output with little or no filtering effect. On the other hand, ferrite transformers have lots of internal inductance and capacitance. They will resonate at lower frequencies, perhaps even audio frequencies. Therefore 1.5K Ω resistors were placed across the transformer primaries to lower the Q still more. If you are still plagued with squeals and motorboating oscillation, you might try removing the 0.01 µF emitter resistor bypass capacitors.

Cascode amplifiers - variable gain with constant Q

I had heard of ***cascode amplifiers*** but didn’t have a clue why they were wonderful. I had already built two other IF strips before I settled on the circuit shown above. Reading in an old handbook, I spotted the IF amplifier shown below. The handbook said that simple transistor amplifiers were poor for IF amplifiers because, when you try to change the gain of a single transistor, the Q of the output tank circuit changes and you get squeals and noise. “Yes!! Yes!!” I cheered. “That’s my problem!” The next circuit uses two (cheap) bipolar transistors in each stage in a “cascode” configuration that (hopefully) cures this difficulty.

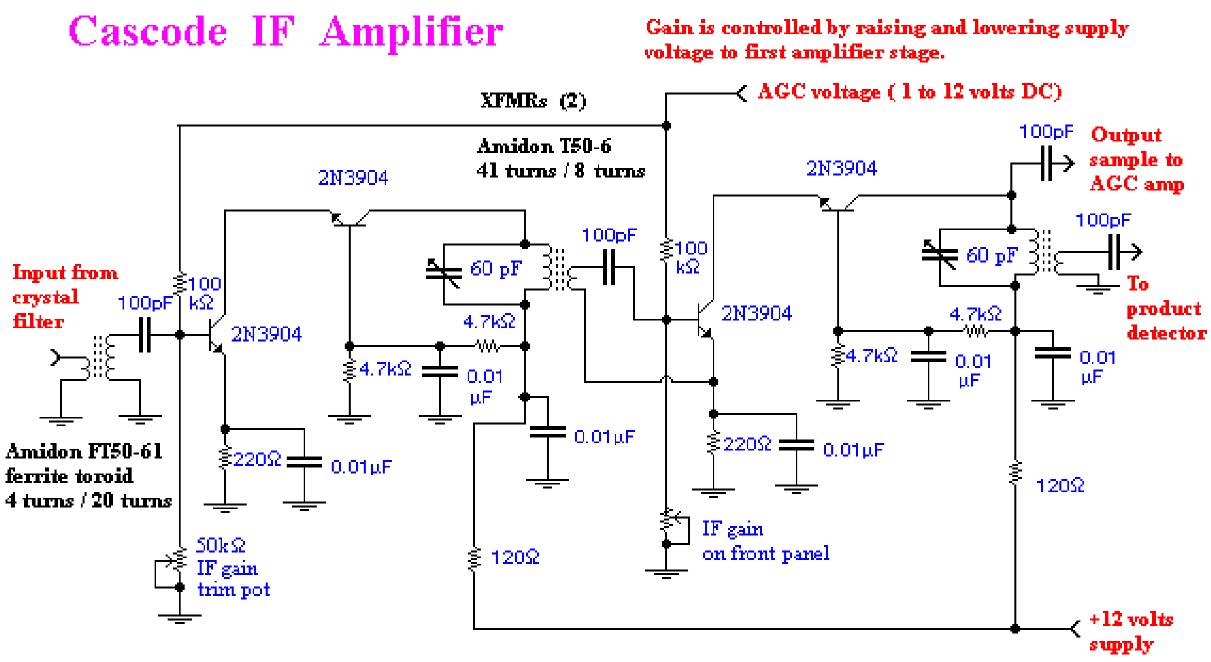
Cascode IF Amplifier
Gain is controlled by raising and lowering supply voltage to first amplifier stage.
AGC voltage (1 to 12 volts DC)
XFMRs (2)
Amidon T50-6
41 turns / 8 turns
2N3904
2N3904
2N3904
2N3904
100pF
Output sample to AGC amp
100pF
To product detector
Input from crystal filter
100pF
100 kΩ
60 pF
100pF
100 kΩ
60 pF
4.7kΩ
4.7kΩ
0.01 µF
4.7kΩ
4.7kΩ
0.01 µF
0.01 µF
0.01µF
220Ω
0.01µF
220Ω
0.01µF
Amidon FT50-61
ferrite toroid
4 turns / 20 turns
120Ω
120Ω
50kΩ
IF gain
trim pot
IF gain
on front panel
+12 volts supply

The input transistor is wired as an ordinary grounded emitter amplifier with its high input impedance. The clever part is that the second transistor is wired to the first in a grounded base configuration. This gives the amplifier a super-high output impedance which supposedly makes it immune to changing the DC bias on the first stage. Besides, the phrase "cascode amplifiers" sounded cool and I wanted to use some. This cascode amplifier has worked well for me.

It's interesting to see what happens when one tunes up an IF amplifier with a scope probe on the IF amplifier output. As expected, the audio signals ride on the IF frequency signal, just like amplitude modulation should. When the amplifier is tuned for optimum signal reception, the scope shows that the amplifier is producing the most modulation on the IF signal. But when the output is tuned slightly differently to produce the largest 9 MHz signal, the reception is OK, but not the best. I had not realized that these two attributes aren't the same thing.

How to tune an IF amplifier

When you first turn on your 80 meter receiver module, there will be several modules with variable capacitors and pots that all need to be adjusted. Let's assume that you have checked out the VFO and it is providing enough voltage sinewave to drive your particular mixer. You should also have a look at the waveform of your BFO after it has arrived at the receiver module. It should be a smooth, clean sinewave. A clean sinewave is more important than high amplitude. You may want to readjust the trimmer capacitor on the BFO oscillator for the best waveform.

To align the IF you could just twiddle pots and trimmers and hope you hit a favorable combination of settings. I've tried that - it doesn't work very well. First, clip an oscilloscope probe onto the output capacitor going from the IF amplifier to the product detector. At this point your BFO comes to your rescue. Clip a test lead onto the connector where the BFO plugs into the 80 meter receiver board. This signal line also has 12 volts riding on it, so connect the other end of the test lead to a small capacitor, say 20 pF. This will filter out the DC. Use the other end of the capacitor as a signal probe to inject a 9 MHz test signal into your IF amplifiers. Since the oscilloscope probe is at the output of the IF, start by aligning the trimmer and input pot of the last amplifier.

Next move the BFO test signal down to the input of the first amplifier and adjust the trimmer capacitor and input trim pot. Peak the 9 MHz signal at the output as before. Now remove the BFO signal test probe but leave the oscilloscope probe in place.

At this point, if there are strong 80 meter signals at the input to the mixer, you should begin to be able to hear them. Now tune each stage carefully for maximum sound level. Notice that maximum sound level is not the same setting as maximum 9 MHz carrier in the output.

Automatic Gain Control (AGC) is not a luxury

The automatic gain control is a receiver feature that holds the signal level relatively constant while tuning in signals of varying strength. Before I built one, I thought an AGC was in the same category with digital readouts and beautiful cabinets. Why do I need one? Am I too lazy to turn the IF gain up and down? It turns out that an AGC has many advantages. The main one is that it helps you achieve the gigantic signal strength dynamic range (100 decibels) that you need in a practical ham receiver. After I built an AGC, I realized it was it also greatly improved my problems with noise and oscillations.

Although I had been happy with the performance of my IF without an AGC, I could never get rid of the "noise zone" in my IF gain control. That is, I had to keep the IF gain below a certain level or it would produce a roar of receiver-generated static. Apparently, IF amplifier stages are only happy when they are processing signals of a limited range of amplitude. Noise and oscillations happen when the signals in the final IF amplifier are too large. With an automatic gain control, it was easier to tune the IF so that the IF gain control acts like a "volume control" without a noise zone. I have to admit that each of my crystal filter ladders behaves differently and I still have noise when I turn the IF gain too high. I usually just leave the IF gain set very low and let the AGC do the adjusting.

The S meter and other uses for the AGC

Another benefit of an AGC is that, when I switch in higher selectivity crystal filters, the AGC compensates for the filter attenuation to a large degree. Also, when you put a meter on the AGC signal level, you have made an ***S meter*** - in other words, a "strength meter." The S meter taught me that what you hear in the headphones doesn't always correlate with the signal strength in the IF strip. In other words, the S meter is reacting to big IF frequency signals, not the level of modulation on those carrier signals.

The best use for an S meter is to tune the transmitter VFO to match the receiver. In other words, if you are answering a CQ, you can tune your transmitter right in on top of the fellow you want to call. First, you need to switch in a 3- or 4-crystal filter. Then, as you slew your transmitter VFO across the frequency, the S meter will soar when you are lined up right on top of him. Without using this technique, "zero beating" the VFO is time consuming. Modern transceivers don't have this synchronization problem because the receiver and transmitter are using the same VFO.

I originally used a digital bargraph S meter that looks very racey. Unfortunately, like virtually everything digital, it makes a hissing radio noise when it changes level and I don't need that. I tried hard to filter the bargraph circuit, but as usual, I couldn't get rid of the hiss. I finally replaced it with an old-fashioned analog meter and the noise vanished.

Automatic Gain Control (AGC)

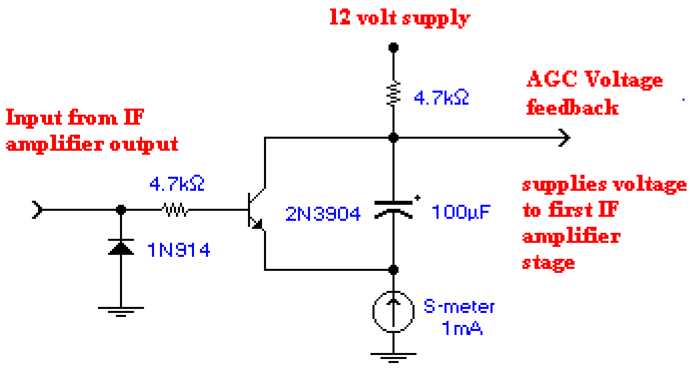
12 volt supply
4.7kΩ
AGC Voltage feedback
Input from IF amplifier output
4.7kΩ
2N3904
100µF
supplies voltage to first IF amplifier stage
1N914
S-meter 1mA

An AGC works by sampling the output level of the last IF amplifier stage. Signals are detected like a crystal set using a diode and averaged with a capacitor to produce a DC level proportional to the IF signal strength. This DC level is then amplified and used to bias the IF amplifiers. For example, the above circuit can deliver the positive voltage bias on IF amplifier stages made from dual gate MOSFETs. Or if the IF amplifiers are made from bipolar transistors, the same circuit can put Class A bias current into the bases of the transistors. For big signals, the AGC automatically turns off the bias and runs the transistors "Class C." When signals become weak, the bases are biased "ON" so that the signals don't have to exceed the 0.6 volt input barrier.

The product detector

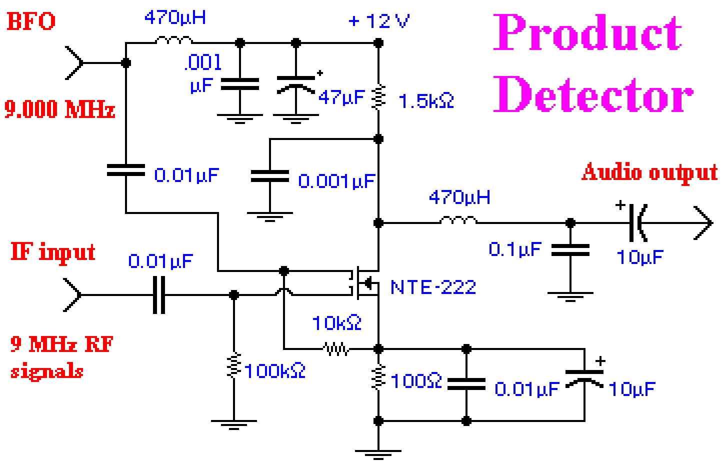
BFO
9.000 MHz
470µH
+12V
.001 µF
47µF
1.5kΩ
Product Detector
0.01µF
0.001µF
Audio output
470µH
0.1µF
10µF
IF input
9 MHz RF signals
0.01µF
NTE-222
10kΩ
100kΩ
100Ω
0.01µF
10µF

My product detector is basically the same circuit I used as my mixer. Product detectors are "direct conversion mixers" that mix an RF "beat frequency" (BFO) signal with the IF frequency to produce a difference frequency which is the audio signal. This is the same detector type used in the simple receiver in chapter 7. A 470 microhenry RF choke keeps the RF out of the audio output. To say it another way, the choke keeps the .1 µF cap from shorting out the RF while letting audio frequencies pass on to the AF amplifier.

Notice that the 12 volt DC power supply for the BFO oscillator passes through another choke and goes out to the BFO oscillator box on the front panel. That is, the DC power input for the BFO and 9 MHz RF output from the BFO share the same wire. The 470 microhenry choke prevents the 9 MHz signal from shorting to the power supply line.

Product detectors are exactly what're needed for CW or SSB. However, when you tune in an AM broadcast station, it will have a whistling overtone on it until you tweak the BFO perfectly to get rid of the whistle. If you plan to listen routinely to short wave AM broadcast stations, you'll probably want to replace the IF crystal filter with a short circuit. I have a setting on my crystal filter rotary switch that is a simple short. Otherwise, the 3 KHz width of a single crystal will be too narrow and the sound will be "low-fidelity." Another change you might consider is to put in a switch to bypass the product detector and use an ordinary diode detector for AM signals. In other words, detect it like the crystal set in chapter 4. Any of the four dual gate MOSFET transistor types mentioned earlier will work well for a product detector. Here is a JFET version that also works OK.

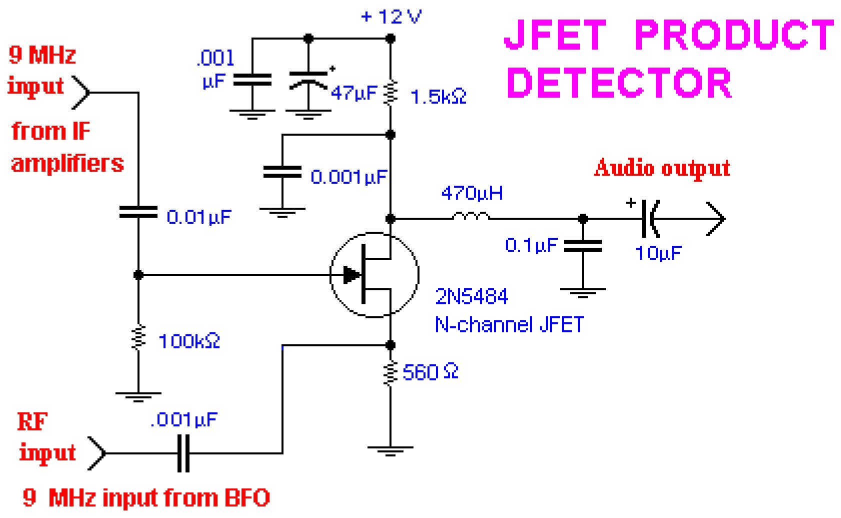

With product detectors, anything works at least a little

In my experience, receiver RF mixers that produce a quality IF output are extremely finicky and are often plagued with low sensitivity and oscillations up and down the band. In contrast, a product detector is amazingly uncritical. I haven't tried to make one out of wood shavings or pebbles from the street, but I wouldn't be surprised if I could still hear signals.

For example, I built a new IF strip and product detector in hopes of improving the noise problem. It worked, but I was slightly disappointed with the sensitivity. I was inspecting my dual gate MOSFET product detector when I realized that I had soldered the MOSFET in 90^{o} out of alignment. In other words, the drain was connected to the RF input gate, the source was connected to the drain circuit, and the BFO input was connected to the source. Delighted that I had discovered my problem, I correctly soldered in a new transistor. When built properly, it worked better - but not dramatically better.

In another experiment I disconnected the RF input so that the input to the product detector was just stray coupling from the IF strip. Signals were weak, but it still worked amazingly well! Finally, I disconnected the BFO input. I was relieved to confirm that it no longer tuned and received ham-band signals. Instead it worked like a crystal set and received the loudest signals on or near the ham band input. For example, on 17 meters, it brought in the Deutsche Welle (Radio Germany) loud and clear.

The AF amplifier

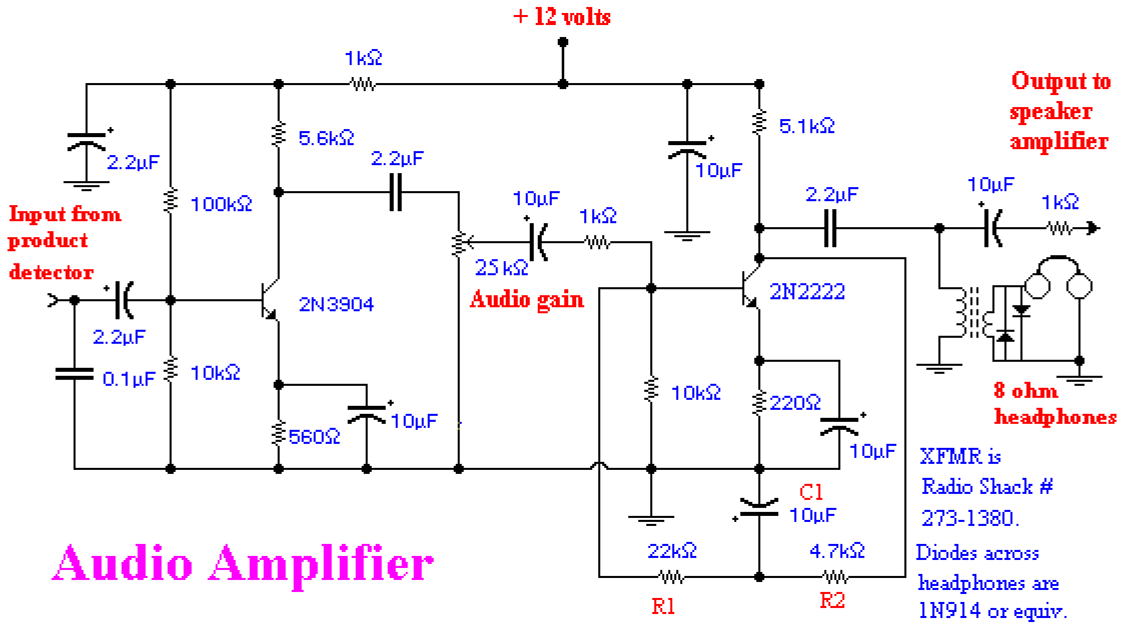
+ 12 volts
1kΩ
5.6kΩ
2.2µF
2.2µF
100kΩ
Input from
product
detector
2.2µF
0.1µF
10kΩ
2N3904
560Ω
10µF
10µF
1kΩ
25 kΩ
Audio gain
10µF
5.1kΩ
2.2µF
10µF
1kΩ
Output to
speaker
amplifier
2N2222
10kΩ
220Ω
10µF
C1
10µF
22kΩ
R1
4.7kΩ
R2
8 ohm
headphones
XFMR is
Radio Shack #
273-1380.
Diodes across
headphones are
1N914 or equiv.
Audio Amplifier

The output from the product detector is an audio signal that needs to be amplified before it goes to the earphones or speaker. Most ARRL designs use integrated circuits marked "audio amplifier." The LM386 is a typical one-chip audio amplifier. I've used these and they usually work great. But of course I didn't learn anything from the experience. So this time I built my audio amplifier out of discrete parts from an example in my 1986 handbook. It looked like two straight-forward "R-C coupled amplifiers" in series. But the design had extra filter components I didn't understand. Every part that I didn't understand, I left out. That was my education. The audio amplifier was dead as a doornail when I first turned it on.

An audio Automatic Gain Control (AGC)

I was particularly puzzled by the low frequency feedback link, R1, R2, and C1. I couldn't understand what sort of "low frequency filtering" the designer was trying to accomplish. But, when the amplifier seemed completely dead, I put these mysterious components back in the circuit. Voila! The earphones came to life. It turns out that this loop biases the amplifier "on" for weak signals and biases it "off" for loud signals. It's a sort of audio AGC circuit.

Remember that for a bipolar transistor to turn on, the input signal must be greater than 0.6

volts or no current will flow into the base. In a “Class A” amplifier a DC signal is added to the base. This increases the base voltage above 0.6 volts so that it’s always turned on. In this way a class A amplifier can amplify signals much smaller than 0.6 volts. The low frequency feedback adjusts the bias for weak and strong signals. When the signals are weak, the second transistor is turned off, so it’s collector voltage is high and unchanging. This big collector voltage is leaked into C1 to provide a forward bias for its own base, biasing it on and raising the sensitivity. Conversely, when the signals are strong, the collector has a big current flowing but a low average DC voltage from the collector to ground. This lower voltage biases the transistor more “off.”

Protecting your ears from strong signals

The audio amplifier is able to blow your ears off when you encounter a strong signal. Therefore it’s essential to add a clamp circuit to limit the voltage to the headphones to less than about a volt. I first did this with back-to-back 5 volt Zener diodes across the headphone jack. In practice, with sensitive modern 8 ohm headphones, I found that less than one volt peak is plenty of volume for me. Eventually I put in two ordinary silicon 1N914 diodes “shorted” in opposite directions across the headphones. This limits the positive and negative sound peaks to just +/- 0.6 volts and my ears haven’t been blasted since.

How Hi-Fi should it be?

The original circuit was also sprinkled with 0.1 microfarad bypass capacitors as if the designer were trying to kill all higher frequency sounds and shunt most of the audio to ground. Since I was trying to get more gain wherever I could, I left out the bypasses. The amplifier worked well without them, but the sound of the static had an obnoxious, piercing high pitch that irritated my ears. I put the bypasses in and, as I expected, the audio sounded more “bass” and became somewhat weaker. However, getting rid of that piercing, hissing static was well worth the loss of gain. Experiment!

The original design also had no emitter bypass capacitor, the 10 microfarad capacitor across the 220 ohm resistor. Not having this bypass capacitor reduces the gain because some of the audio voltage signal is wasted across the 220 ohm emitter resistor. Since I needed gain, I put in the capacitor and my gain jumped up noticeably. This bypass has no disadvantage that I could detect.

Sharp audio filters

Many receivers have audio filters that limit the audio frequency of signals passed onto the headphones. This can be useful for separating CW signals that are nearly on the same frequency. If I didn’t have my multiple crystal filter selections, I would definitely want audio filters. But in practice, when QRM (interference) happens to me, the guy that is interfering usually has the same audio tone pitch as the fellow I’m trying to listen to. Obviously in this case an audio filter would not help. But if you want to add one later, it is never too late. Unlike IF crystal filters, audio filters can be added external to the receiver. There is an example of a 700 Hz audio filter in chapter 7.

Driving a speaker

If you don’t need a speaker, you don’t need a third amplifier stage. By the same token, an 8 ohm speaker plugged into the 8 ohm headphone output is much too faint. Also, 0.6 volts peak is not nearly enough to drive a loudspeaker.

In the original handbook W7ZOI/ K5IRK AF amplifier design, the third stage was an emitter-follower for driving either a speaker or low impedance headphones. The advantage of this design was that the emitter-follower drove the speaker directly and there was no need for a high-to-low impedance transformer. The speaker was located between the emitter and ground while the collector was connected to the positive supply. It looked straightforward to me, so I tried it. Unfortunately the emitter-follower distorted the sound and "motorboated" on strong signals. That is, the sound ran in "putt-putt" bursts. I tried several modifications to solve these problems but was never able to fix it. I gave up on the emitter-follower and used another impedance step-down audio transformer to drive the low impedance speaker. I happened to have a handful of tiny speaker transformers in my transformer junk box, so for me this was an easy solution.

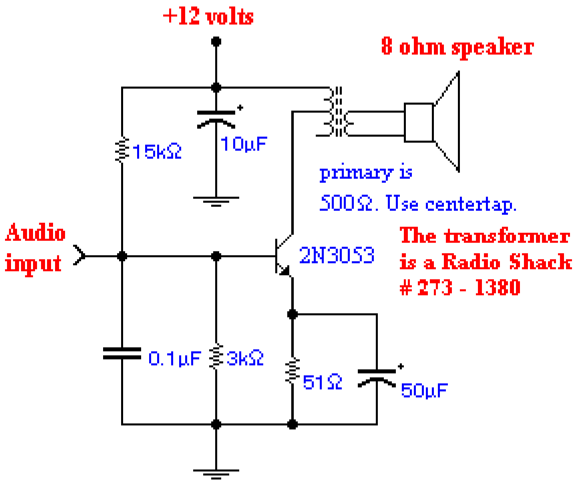

An optional extra amplifier for driving an external speaker

You'll find that a big speaker sounds much better than a little one. A speaker small enough to fit in the receiver itself will sound "tinny." I eventually wired my speaker output to a remote, 12 inch wide speaker. This little amplifier doesn't really provide enough drive for such a large speaker, but it is OK.

****
**

Converters for the other HF bands

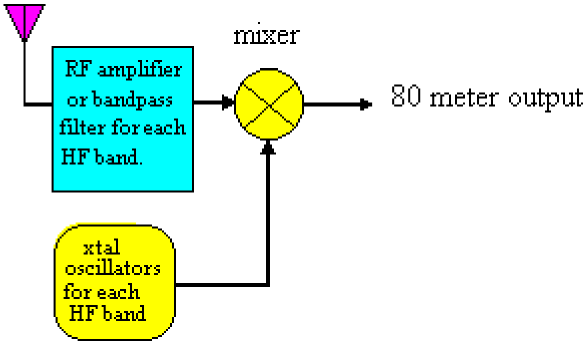
mixer
RF amplifier or bandpass filter for each HF band.
80 meter output
xtal oscillators for each HF band

Block diagram of a converter for an HF hamband

I used the RF amplifiers and crystal oscillators out of the W7ZOI / K5IRK receiver. I built these modules close to what was described in the handbook and they worked right away. For my mixer module I used the same dual gate MOSFET circuit I developed for the 80 meter receiver. I had some difficulty with the low frequency preselector filters, so I used other designs as will be described.

In my receiver all converters for bands other than 80 meters share the same dual gate MOSFET mixer. Band switching would be easier if each converter had its own mixer. On the other hand, those dual-gate MOSFETs are pricey transistors, so do what you like. Each band needs its own crystal-controlled oscillator and a pre-tuned bandpass filter or “preselector” to limit the input to the desired band. Bands above 30 or 40 meters need an RF amplifier. Below 20 or 30 meters, the signals and noise are stronger and an RF amplifier on the antenna input shouldn’t be needed. I have had some trouble with weak signals on 30 meters, so I now regret that I didn’t add an RF amplifier on 30 meters as well. On the other hand, maybe the signals are just plain weak.

Each ham band needs its own HF converter

In the old vacuum tube days, tubes were physically large and it would have been extravagant for a homebuilder to use a separate converter for each band. My old homebrew vacuum tube receiver had a single multi-band converter that had to be tuned manually for each upper band.

Transistors and powdered iron cores are tiny, so today we can easily house a complete HF converter into a few cubic inches. Moreover, each converter runs on just one low DC voltage. This simplicity means that the power can be routed to the converter using the same coaxial line that receives the output from the converter. Since each converter is optimized for just one band, it can be tuned up once and forgotten.

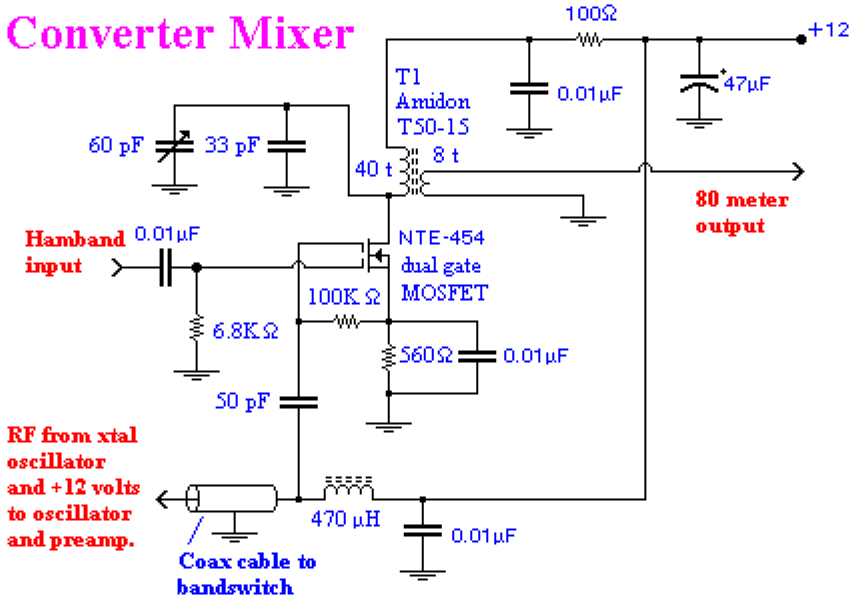
Converter Mixer
100Ω
+12V
T1
Amidon
T50-15
0.01µF
47µF
60 pF
33 pF
40 t
8 t
80 meter
output
Hamband
input
0.01µF
NTE-454
dual gate
MOSFET
100K Ω
6.8K Ω
560Ω
0.01µF
50 pF
RF from xtal
oscillator
and +12 volts
to oscillator
and preamp.
470 µH
0.01µF
Coax cable to
bandswitch

All my converters share the same converter mixer shown above. A rotary bandswitch brings in the filtered and amplified hamband RF signal on the upper input. The lower input takes in the RF from the local crystal oscillator for each band. In addition, the lower "input" is also an output that delivers 12 volt power to the oscillator and the preamplifier for that band.

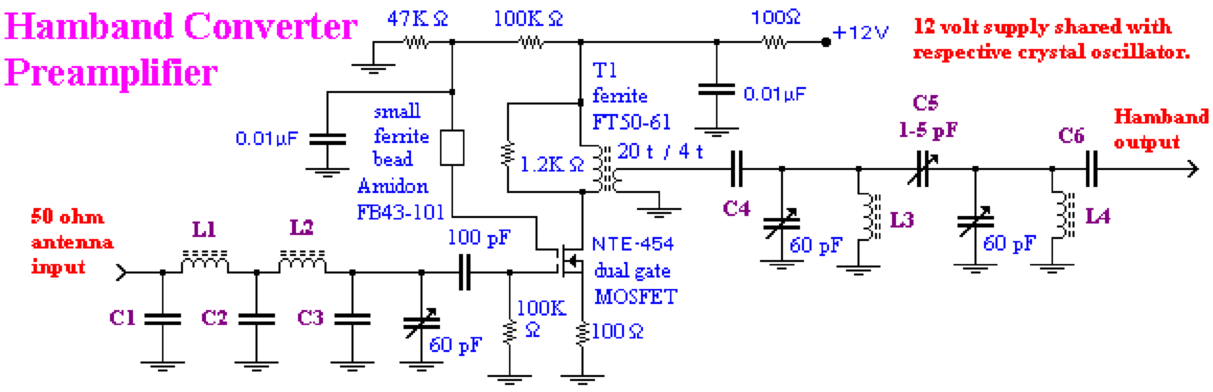
Hamband Converter
Preamplifier
47K Ω
100K Ω
100Ω
+12V
12 volt supply shared with
respective crystal oscillator.
T1
ferrite
FT50-61
0.01µF
0.01µF
small
ferrite
bead
Amidon
FB43-101
1.2K Ω
20 t / 4 t
C5
1-5 pF
C6
Hamband
output
C4
L3
L4
60 pF
60 pF
50 ohm
antenna
input
L1
L2
100 pF
NTE-454
dual gate
MOSFET
C1
C2
C3
60 pF
100K
Ω
100 Ω

Table of values for converter preamplifier and filters

Hamband	**C1**	**C2**	**C3**	**C4** **C6**	**C5**	**L1**	**L2**	**L3** **L4**
	(pF)	**(pF)**	**(pF)**	**(pF)**	**(pF)**	**(µH)**	**(µH)**	**(µH)**
30 meters	300	680	33	33	4.1	0.68	3.36	1.16
		T50-6 CWS (Amidon) toroids				**(13t)**	**(29t)**	**(17t)**
20 meters	220	500	22	27	4	0.58	2.50	1.16
		T50-6				**(12t)**	**(25t)**	**(17t)**

17 meters	180	390	none	22	3.9	0.40	1.94	1.44
		T50-6				**(10t)**	**(22t)**	**(19t)**
15 meters	150	330	none	20	3	0.40	1.60	1.44
		T50-6				**(10t)**	**(20t)**	**(19t)**
12 meters	120	200	none	12	2.1	0.26	1.44	0.58
		T50-6				**(8t)**	**(19t)**	**(12t)**
10 meters	110	250	none	12	1.6	0.26	1.16	0.40
		T50-6				**(8t)**	**(17t)**	**(10t)**

Preselectors for the lower hambands

As explained earlier, hambands lower than about 30 meters shouldn't require an RF preamplifier. The only purpose of the preselector is to limit the input signals to the particular hamband. This design shown below is the one recommended by W7ZOI and K5IRK .

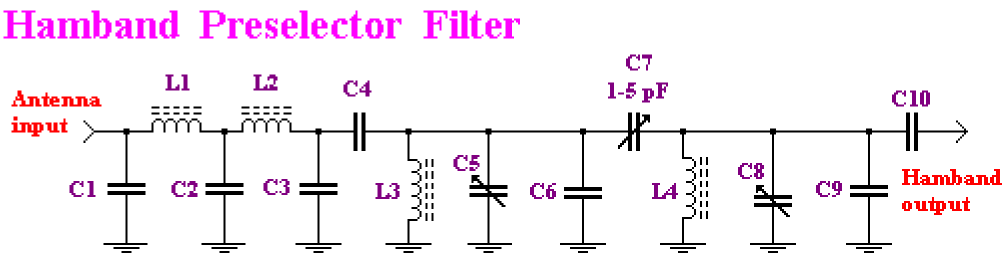

Table of values for preselector filter (without amplifier)

Hamband	**C1 C3 (pF)**	**C2 (pF)**	**C4 C10 (pF)**	**C5 C8 (pF)**	**C6 C9 (pF)**	**C7 (pF)**	**L1 L2 (µH)**	**L3 L4 (µH)**
30 meters	300	600	32	180	50	4.1	0.68	1.16
		CWS (Amidon) T50-6					**(13 turns)**	**(17 turns)**
40 meters	430	860	42	180	50	4.6	1.16	2.50
		CWS (Amidon) T50-6					**(17 turns)**	**(25 turns)**
160 meters	1720	3440	none	250	120	20	4.64	30
		CWS (Amidon) T68-2					**(29 turns)**	**(79 turns)**

My 160 meter filter is similar to the original handbook design and was extrapolated from their 40 meter values. For 30 and 40 meters I used the following filter design that I adapted from an article in QEX:

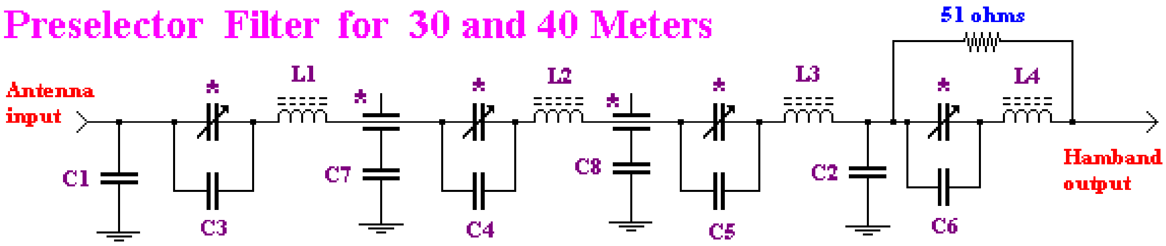

Preselector Filter for 30 and 40 Meters

* All variable caps are 60 pF trimmers

* C7 and C8 are .001 µF (1000pF) feedthrough caps augmented with 0.001 µF disk ceramic caps.

Table of component values for alternate preselector filter

Hamband	**C1, C2**	**C3, C4, C5, & C6**	**C7,C8**	**L1,2,3**	**L4**
	(pF)	**(pF)**	**(pF)**	**(µH)**	**(µH)**
30 meters	710	33	2000	3.5	4.6
			CWS (Amidon) T50-6	**(30 turns)**	**(34 turns)**
40 meters	1000	47	3000	5.0	6.6
			CWS (Amidon) T50-6	**(35 turns)**	**(41 turns)**

The above design is a series of 4 series LC circuits, each in its own egg carton compartment soldered onto the PC board. The signals are fed from one little chamber to the next using .001µF feed-through capacitors. Extra capacitors were soldered on to bring the values up to the values listed above. These filters tuned up well and I have no complaints about the receiver performance on 40 meters and above. 80 and 160 meters are more difficult, as discussed earlier and below.

160 meters is hard

160 meters presents two problems. First, standard broadcast signals are huge and just below 1.800 MHz. If you don't attenuate the broadcast band signals, you may be plagued with the loudest local AM stations appearing in your IF. You may find the AM broadcast filter described in chapter 7 useful. It worked best when placed in between the 160 meter preselector and the input to the 80 meter receiver board. It was not so helpful in series with the antenna jack outside the receiver. My 160 meter preselector was primarily a high pass filter.

The second problem with 160 meters is its proximity to 80 meters. When I first got my 160 meter converter working, I immediately heard hams and thought I was successful. It was several evenings before I discovered that some of those hams were actually on 80 meters. Whoops! When I tuned in a ham, I then switched to 80 meters to see if he was still there. If he disappeared, then I knew he was on 160 meters. Not really up to snuff. Since my core receiver is designed for 80 meters, the bandpass filtering on the 160 meter converter must be quite selective to keep out both the broadcast signals and the 80 meter signals. I'm still not satisfied with my present filter. As discussed elsewhere, 80 and 160 meters work much better when you listen to them using your transmitter "T-match" antenna tuner.

It can be hard to find an appropriate, cheap crystal for 160 meters. 5.5 MHz is available as a standard frequency and looked ideal. Unfortunately, this produces a large whistling artifact on 2.00 MHz. 5.6 MHz worked OK because the artifact is now on 2.1 MHz, completely out of the 1.8 to 2.0 MHz ham band.

Still another difficulty with 160 meters is that, if you build a multistage filter like those shown for 40 and 30 meters, the variable capacitors will be physically quite large. I didn't have room for such a large preselector, but if you started out with a big enough chassis that shouldn't be a problem. To scale the 40 meter filter to 160 meters, multiply the value of each capacitor and inductor times four. Good luck!

Crystal oscillators for the converters

These oscillators are nearly identical to the BFO oscillator described earlier. The 12 volt power for each oscillator is delivered via the cable from the bandswitch.

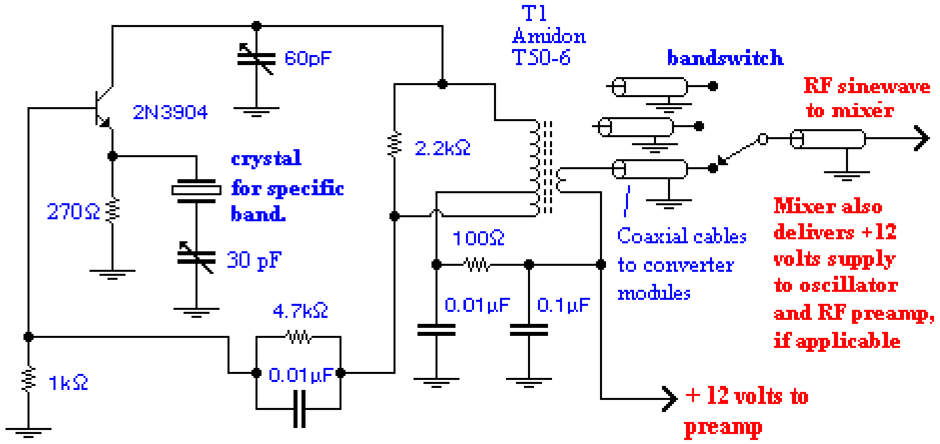
T1
Amidon
T50-6
60pF
bandswitch
RF sinewave
to mixer
2N3904
2.2kΩ
crystal
for specific
band.
270Ω
30 pF
100Ω
Coaxial cables
to converter
modules
Mixer also
delivers +12
volts supply
to oscillator
and RF preamp,
if applicable
0.01µF
0.1µF
4.7kΩ
0.01µF
1kΩ
+ 12 volts to
preamp

Converter Crystal Oscillator(s)

Data table for crystal oscillators

Hamband	Crystal freq. (MHz)	T1 primary total turns	Primary tap turns	Secondary turns
160 meters	5.600	39	8	7
40 meters	11.000	30	7	6
30 meters	13.800	26	6	5
20 meters	18.000	28	5	4
17 meters	21.900	20	4	4
15 meters	25.000	20	4	4
12 meters	28.200	17	4	4
10 meters	32.000	14	3	3

All toroids are CWS (Amidon) T50-6 powdered iron

Since I am converting to 80 meters, my receiver uses the same local oscillator frequencies I used in my CW transmitter, which has an 80 meter VFO. For most bands I use microprocessor crystals cut for frequencies 4.0 MHz above the bottom of the desired ham band. For example, I used 11 MHz for 40 meters, 18 MHz for 20 meters, 25 MHz for 15 meters, and 32 MHz for 10 meters. The WARC bands are only 100 KHz wide. This allows you to use crystals that are any convenient frequency covered by the 3.5 to 4.0 MHz frequency range of the core receiver. To make your choice, you might consider what other interesting bands are close by. For example, if your 30 meter choice allows you to hear 10.00 MHz, you can check your UTC clock setting from WWV. Ham communications are recorded in logbooks using UTC time so that everyone in the world is on the same time standard. For wide bands like 15 meters and especially 10 meters, you have less choice about crystal frequencies to use.

You could also use local oscillator frequencies below the desired band. For example, to cover 15 meters, you might use a 17.5 MHz crystal oscillator. This will convert the 15 meter band to range from 3.5 MHz to 3.95 MHz. Notice that the ***direction*** of VFO tuning is reversed from using local oscillator frequencies above the desired band. Whatever you do, it’s good to be consistent so that direction of tuning and the calibration can be consistent. Also, if the local oscillator frequencies are round numbers, like 11.0, 18.0, 25.0, and 32.0 MHz, then decimal calibrations on the VFO will be identical. Unfortunately, to get quality crystals for the WARC bands that didn’t drift, I eventually had to have custom crystals ground for me. Maybe cheap isn't so important after all.

Mechanical construction of converters

I built my converters on two-sided PC board material. Strips of PC board were soldered onto the main board to provide the walls and partitions in the egg carton construction. A folded aluminum, press-on cover keeps out most stray RF. The assembly shown below contains four converters for 20, 40, 30, and 12 meters. It looks daunting, but I built them one band at a time over several months. I only built another converter when I was ready to listen to a new band.

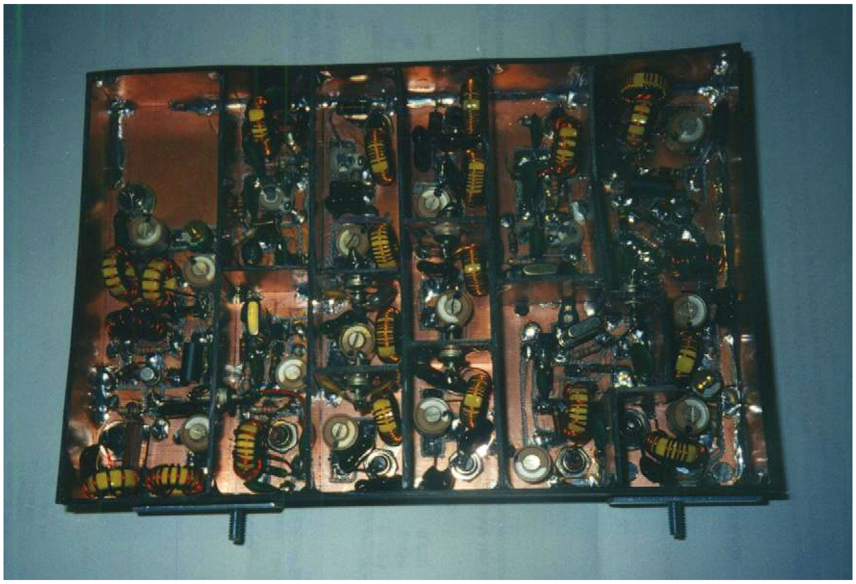

Converter modules for 20 meters, 30 meters, 40 meters, and 12 meters

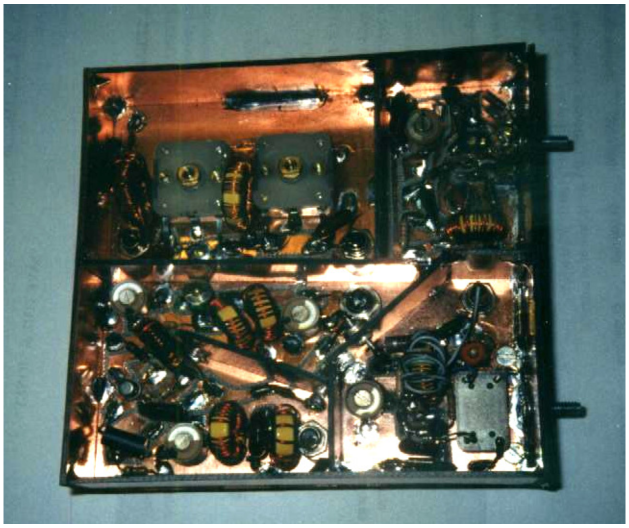

This module contains converters for 160 meters and 17 meters.

Band switching

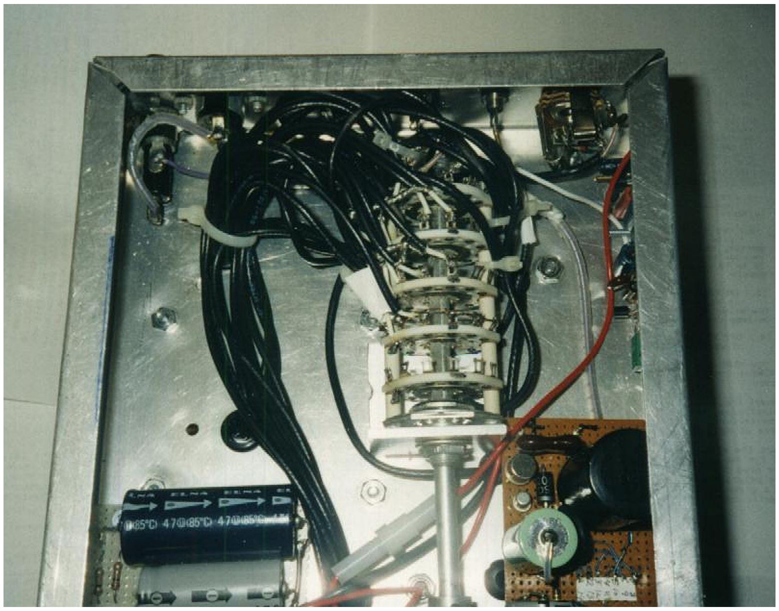

The very first module in my receiver that worked properly was the converter for 15 meters. I checked it out by using it with my old receiver tuned to 80 meters. I immediately noticed that it worked better than the old receiver tuned to 15 meters. Later, when the new 80-meter receiver began to work, I was able to listen to 15 meters by plugging the 80 meter receiver to the 15 meter converter directly with phono plug cables. Next I built converters for 10 meters and 40 meters. Whenever I switched bands, I moved my cables around like plugs on an old-fashioned telephone switchboard.

Eventually, as I built more and more converters, this process became increasingly cumbersome. Finally I gritted my teeth and spent a Saturday afternoon wiring up the bandswitch shown above. To switch from one band to another, a rotary switch selects the desired input filter/ RF amplifier and local oscillator for each band. Like the BFO design described earlier, the 12-volt power for each local oscillator and RF preamplifier arrives riding on the local oscillator RF signal to the mixer. In other words, a single switching circuit carries both the local oscillator signal and the 12 volt power. Even with this reduction in complexity, the 9 band rotary switch becomes a confusing pile of RG-174 coax spaghetti. Be sure to label each plug and socket. I used a fold of white plastic tape on each cable and labeled it with a fine-point permanent marker pen.

I've had trouble buying quality RG-174. One 100 foot roll I bought was mechanically and electrically just fine. Unfortunately, the insulation had a chemical stench that was quite obnoxious and, for all I know, carcinogenic besides. The smell did not dissipate with age. I even tried hanging it on a fence post in the backyard and letting the sun and rain work on it for a few months. It still stank, so I threw it out. Another roll of RG-174 had no smell, but the inner multi-strand conductor was so fragile, that the slightest bending on a solder joint breaks the wire. To use it, I had to strain relief each connection with a drop of epoxy. In summary, if possible, sniff your coax and look at it closely before buying it.

Receiver muting

While transmitting it is extremely distracting to hear your own signal blasting your ears in the receiver. Even with the antenna switched over to the transmitter, the receiver will still be overwhelmed by your own loud signal. A CW signal will usually sound distorted and will interfere with your sending. I have found it best to turn the receiver off altogether while transmitting. That is why I built a Morse code sounder into my electronic bug so I would have a clear tone to listen to while I am sending.

Turning the receiver off by hand is too slow, so you will need to build a muting circuit. I trigger mine with the same transmit/ receive signal line used to turn on the transmitter. I have found that the simplest solution is turning off some or all of the power supply lines. I first tried just turning down the audio and IF gain pots by pulling the pot wipers down to ground using the collectors of transistors. This was inadequate, so I have since gone to keying the receiver power supply on and off, just as I did in the transmitter. I leave the receiver VFO supply on constantly to prevent any drifting due to having the VFO circuitry cool off and warm up repeatedly.

Receiver power supplies

The VFO power supply used in the receiver is the same one described in chapter 10. You could also use the line powered regulated supply described in chapter 8. I run my station on a 12 volt battery, so I just needed the low drop-out regulator shown below. The power supply requirements for the receiver are similar to the low power stages of the transmitter. I used the same supply that I developed for use with my QRP modules in chapter 8. This supply is over-designed for a receiver and can supply at least a few amperes.

The mute capability was implemented using an unused op-amp as a comparator. The mute input lead comes from the transmitter. Or in my case, it comes from the transmit/ receive switch on my homebuilt bug. During transmit the mute line is pulled low to ground. During receive, the 12 K and 47 K resistors pull the mute line up to 12 volts.

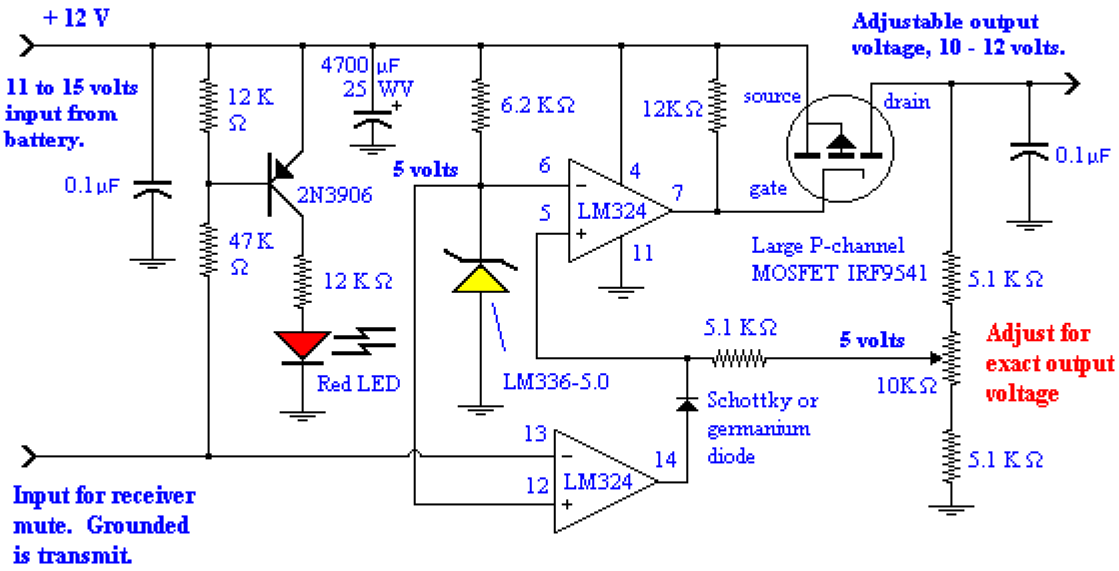
+ 12 V
11 to 15 volts input from battery.
0.1µF
12 K Ω
47 K Ω
2N3906
4700 µF 25 WV
12 K Ω
Red LED
5 volts
6.2 K Ω
LM336-5.0
6
5
4
7
11
LM324
12K Ω
source
drain
gate
Large P-channel MOSFET IRF9541
Adjustable output voltage, 10 - 12 volts.
0.1µF
5.1 K Ω
5.1 K Ω
5 volts
10K Ω
Adjust for exact output voltage
5.1 K Ω
Schottky or germanium diode
13
12
14
LM324
Input for receiver mute. Grounded is transmit.

Low Dropout 12 Volt Power Supply with Mute Capability

Referring to the lower op-amp, when the positive input wired to the 5 volt reference is higher than the mute line, the op-amp output, pin 14, rises up to 12 volts. This causes current to flow down through the diode to the positive input of the upper op-amp, pin 5. When the positive input (pin 5) is higher than the 5 volt reference on pin 6, the upper op-amp output (pin 7) rises to 12 volts, thereby turning the power MOSFET off and shutting off the current supply to the receiver.

Red mute indicator light

While the receiver is being muted, the base of the 2N3906 transistor is pulled down thereby turning on the transistor and lighting the red LED. My LED was an extra bright type that only needed 1 milliampere to be adequately visible. You may need a resistor smaller than 12 K for good brightness.

Use linear regulators, not switching regulators

I tried to use switching supplies for this receiver, but reception was always impaired by switching supply noise. Switchers cause buzzing and hissing on every ham band. RF filtering was a great improvement, but it was never good enough. Building a low-noise receiver is hard enough without fighting a built-in noise generator. For me, one of the great mysteries of modern receivers is how they incorporate all those digital displays and processors without producing a hash of noise that would make the receivers insensitive to weak signals on the upper bands. I can even hear my Hewlett Packard digital counter, which is located 2 feet from the receiver and operates on a different power source.

Tuning signals with your new receiver

I found that it was easy to tune in CW stations. First, you should have calibrate your beat frequency oscillator so that you can select 500 to 700 Hz above and below the 9.000 MHz IF. For 40 meters and below, set the BFO to 700 Hz ***above*** 9.000 MHz. For 30 meters and above set the BFO for 700 Hz ***below*** 9.000 MHz. This will insure that you are listening using the lower and upper sidebands, respectively. If you have built a multi-stage crystal filter, this function will be obvious as you tune up and down the band. With just one crystal filter or no filter you will find you can adjust the IF gain as high as you like. But as more crystals are switched in, the automatic gain control will raise the gain automatically to compensate for the attenuation of the crystals. The S-meter reading will be meaningful with only one or zero crystals in line. If you try to obtain more gain by increasing the IF gain pot setting, the IF strip may oscillate. So for multiple crystals, just turn the IF gain way down.

Tuning in single sideband is also easy, but only if you know the tricks. When tuning in any sideband signal, set your BFO to 9.000 MHz. Think of it as replacing the carrier that was removed from the original amplitude modulated signal. For upper sideband modulation, the modulation will enter your IF in the range of 9.000 to about 9.003 MHz. By beating with your BFO signal, the difference will generate the original audio signal, roughly 200 Hz to 3 KHz.

The main trick to easy SSB tuning is to use the absolute minimum RF and IF gain that will allow you to hear the voice. To make the voice loud enough, you will probably have to crank up the AF gain quite high to compensate. Since my receiver doesn't have an external RF gain, I use the 80 meter preselector capacitor as an attenuator to reduce the RF input signal to a minimum. A single crystal filter helps minimize noise and interference from other stations. Using two or more crystals will chop off the upper audio frequencies and make the voice unintelligible. If you set up the receiver as just described, it will be easy to make the voices intelligible as you tune up the band. However, to make the tuning perfect so that the voice is undistorted and sounds like your friend's normal voice, you will have to tune precisely. You may find it helpful to tune the last few Hz using the BFO control knob rather than the local oscillator tuning knob.

In summary,

If you've already built homebrew transmitters, you'll find that building a receiver is easier than building your multiband, VFO-controlled CW transmitter. The hardest parts are the 80-meter pre-selector, the 80 meter mixer, and the IF amplifier. After that, it's pretty straightforward. Receivers don't need RF power gain stages, so the currents are low and the circuit board layout is forgiving.

In closing, remember that when you build a transmitter, you will hear no end of complaints about your signal until you get ***ALL*** the bugs out of it. But when you're on the air with your homebrew receiver, your contacts will never complain about your receiver. Or at least you will never hear them complain.

CRYSTAL SETS TO SIDEBAND
© Frank W. Harris 2010, REV 12

Chapter 14

VACUUM TUBE RECEIVERS AND TRANSMITTERS

My old-tech QRP project began the day I acquired an ancient type 216A vacuum tube. It's a wondrous Jules Verne-looking device with a round glass ball mounted on an elaborate nickel-plated brass base. It's the simplest three-element "triode" with all its guts in plain view. It has the filament in the center with a wire "grid" and a plate on each side of the filament. The base is imprinted with several patent numbers dating from 1907 to 1918. I decided I would use it to build a one-tube, crystal-controlled transmitter, just like the old days. As you'll see, there were problems with this dream.

The transmitter described in this chapter uses the most ancient vacuum tube technology I could, while still sounding pretty good on the air. Some of the parts I used were rare and unobtainable as new parts. However, if you can't find similar components at local swapfests, you can always buy them on-line. It's also possible to substitute more modern vacuum tubes that are still commercially available.

How old can you go?

I got a huge kick out of building the spark gap transmitter and crystal set receiver. My only regret was that I couldn't use them on the air. I wondered how old fashioned a rig can still be used today. About 1920 hams began to replace their spark gap transmitters with single tube LC tuned oscillators. In principle, these VFOs weren't that different than the VFOs described in chapter 10. Unfortunately the old VFOs were crude and drifted several kilohertz per minute. Obviously one of those state-of-the-art 1925 transmitters was not going to be acceptable today.

Spark gap transmitters were totally banned in 1927 because they splattered all over the band and wasted spectrum. After the international reorganization of the ham bands in 1929, the hams were confined to what were then considered "narrow" portions of the spectrum. Crystal control for hams was first introduced in QST magazine in 1927. The May 2001 QST had an article on ham transmitters used in the 1920s. The article warned that these early machines shouldn’t be used on the air. That’s no fun! In my local ham newsletter Yardley Beers, WØJF, described a transmitter he built in 1930 that sounded promising. It was a 10-watt, one tube oscillator with a type 210 triode. I figured that a triode oscillator with crystal control would be about as old-tech as I can go.

Why bother with tubes?

A homebuilder who has never built with vacuum tubes has missed an adventure. I told my electrical engineer nephew that I was building a vacuum tube transmitter. There was silence on the phone. “Why?!” he finally blurted. I was startled. To us geezers, “why” is obvious.

Let’s use trains as an analogy: I once had the privilege of riding on the French TGV (Le Tran Gran Vitesse) that translates to something like, “The Really Fast Train.” At 240 kilometers per hour the TGV is indeed speedy. When it passes another train on the adjacent track, the whooshing roar lasts less than two seconds. When TGV crosses a trestle, the ground drops away under you so quickly you feel as though you were taking off in a rocket. Modern ham equipment is like the TGV. Its attractions are sophistication and high performance.

Every summer narrow gauge steam trains puff and chug up and down Colorado canyons. These old-time anachronisms are packed with tourists. Steam trains aren’t popular for their performance, but rather for their sound and appearance. Vacuum tube transmitters, especially ones from the 1930s or 40s, don’t have impressive performance, but their visuals are great! Transmitters back then had great big coils, “bread slicer” variable capacitors, and wondrous glowing vacuum tubes. Their plate current and grid drive meters look as though they belong in a power plant.

Vacuum tubes have the same function as transistors. They are constructed like filament light bulbs. Inside an evacuated chamber, usually glass, they have incandescent orange filaments and intricate metal screens and plates. The lightbulb-like filament is heated to incandescence to drive electrons off its surface. The metal mesh, called the control grid, regulates the relatively large current passing through the vacuum between the filament (the cathode) and the metal plate. The grid is analogous to the base of a bipolar transistor or the gate of a field effect transistor.

Vacuum tubes were the first practical small-signal amplifying device. They dominated electronics for over 50 years. Most sophisticated analog circuits and even computer circuits were first implemented with tubes. When transistors became available, it was relatively easy for vacuum tube engineers to redesign the old circuits using transistors.

If tubes are so easy, why did my first two transmitters fail?

I've always been curious why my first two home-built transmitters did not succeed 53 years ago. My first transmitter was a 7 watt 80 & 40 meter rig built with plans in the 1957 ARRL handbook. It worked well around town, but hardly anyone heard me outside of town. Now that I've built modern QRPs and receivers, I realize that the average ham receiver back then was so poor that hardly anyone could hear a QRP. My second homebuilt transmitter was supposed to put out 30 watts but never worked properly. I lacked the test equipment and knowledge to find out why. Eventually I bought a commercial transmitter kit, just like the kits all the other novices were using.

In the transmitter described below, the power supply turned out to be my biggest obstacle. I'm convinced that was also my biggest problem back in 1957. Yes, my power supply was able to supply the required power, but now as then, my power supply was too weak or "soft" to supply the needed power without a significant voltage drop. Whenever the transmitter drew current from the supply, the voltage crashed causing the transmitter to run in bursts called "motor-boating." Since I didn't have a high frequency oscilloscope back then, I couldn't see what was happening.

The power supply

Vacuum tubes work by conducting electrons across a vacuum. Personally, I'm amazed that's even possible. But once you get past that, it shouldn't be a surprise that a vacuum tube has a relatively high resistance. So if you want to pass big currents, you need big voltages to push the current across the void.

Power = Voltage x Current

To get big power you need either big voltage or big current, or both. The transmitter described here needs at least 250 volts to deliver 5 watts output. And because of its low energy conversion inefficiency, plan on supplying 15 watts DC power instead of just 5 watts. For example,

15 watts = 250 volts x 60 milliamperes

High voltage power supply safety

The chief disadvantage of high voltage is the danger of burns and shock. Vacuum tubes almost always operate at high voltages, typically over 100 volts. For high power ham amplifiers the plate voltage can be 1000 volts or even 3000 volts DC. The bad news is that this can cause a severe jolt and burn when your finger touches the **DIRECT CURRENT** supply. I watched a friend adjusting the resonant cavity of UHF diathermy machine with 3000 volts DC on the plate. He finished, smiled and looked up at me as if to say, “There! All done!” As he looked up, he stopped paying attention to the high voltage terminal. It blew a centimeter wide crater in his knuckle exposing the tendon. The good news about high voltage DC is that, although it may knock you across the room, badly burn you, and scare the heck out of you, high DC voltage rarely kills anyone. It can kill, but usually doesn’t.

Shock

Technically, the word ***shock*** means electrocution, so DC voltage usually doesn’t “shock” hams. On the other hand, a 1000 volt DC power supply has 1000 volts of **ALTERNATING CURRENT** on its rectifier. If you get your hands across that, death is quite possible, even likely. Any power supply that plugs into the AC mains will have at least 120 volts AC on it and that’s plenty to kill you. All power supplies are potential killers if you touch high AC voltage. In contrast, when 60 Hz AC voltage is below about 20 volts, it approaches being “safe.” However, if you wet both hands, grip the two leads firmly so that 10 milliamperes of AC current passes through your chest, even 12 volts AC can kill you. In summary, it pays to be extremely cautious with low frequency (50 – 60 Hz) AC voltage. Whenever possible, unplug the power supply before you work on your line powered equipment.

RF burns

Another safety issue is RF burns. High frequency, high voltage, radio frequency AC voltage doesn’t shock you because your nerves and muscles can’t respond fast enough to be polarized by the changing voltage. However, RF voltage can burn large holes in your hand. If you touch the plate circuit of your vacuum tube transmitter, you can expect at least a small hole in your skin. And, the higher the RF voltage you are using, the bigger the hole it will make. In summary, any voltage, AC, DC or RF, higher than roughly 50 volts deserves great respect. Your first lesson on working with vacuum tubes should be ***don’t touch anything until the power is turned off.*** And, after the power is turned off, ***be wary of large capacitors that may be charged to a high DC voltage.***

****
**

The Vacuum Tube QRP Transmitter

Once I began planning my tube QRP, reality began to sink in.

* First, I only have one antique 1918 tube. I'd hate to break it.

* Second, this tube has no cathode. It just has a simple filament. This means that the electrons come directly off the filament. A simple tungsten or carbon filament gives off electrons reluctantly, so these early tubes had low operating currents. That meant that my QRP transmitter would be extremely weak, perhaps a few milliwatts. There was a good reason that many hams used spark-gap transmitters until 1927. Power tubes weren't available until the 'twenties. It was hard to generate any significant power output until tubes were developed that could handle hundreds of milliamperes of current, instead of just a few milliamperes. Also, I'll bet power tubes cost a fortune back then.

* According to my 1979 ARRL handbook, "It is almost impossible to build a one tube CW transmitter that doesn't chirp to some degree." I guess chirp was normal in the old days. But today your signal will be roundly criticized if you dare to chirp at the modern hams who rule the airwaves. So much for a one-tube transmitter.

* A related problem is that, the better a keyed vacuum tube oscillator sounds, the less output power you will get from it. In other words, a keyed oscillator that sounds good puts out little or no useful RF power. I began to understand that my 100 milliwatt transmitter might easily end up as a 20 milliwatt transmitter.

In my opinion, 1930s technology is about as early as you can go and not embarrass yourself. I eventually built my transmitter using tubes that I had salvaged from old radios built during the 1930s. I like tubes from that era because they are big and spectacular. More modern tubes from the 40s and 50s are only large if they handle high power. Moreover, their glowing filaments are largely hidden by their cathode sleeves. Some of them have metal envelopes and don't glow at all.

In the design described below I used a simple triode as a final amplifier. I believe the filament has been treated with thorium so that it emits electrons readily. Triodes have relatively low gain and were not very linear in their amplification characteristics. The tubes I used were type 68 and type 71. If you can't scrape up some old tubes, the more modern 6L6 is still available commercially. To eliminate the chirp, I used two more tubes to drive the final. So my single tube QRP ended up a 3-tube transmitter. And all I get out of it is 5 watts! The oscillator runs at very low power and can be keyed with little chirp. The oscillator is followed by a "buffer" amplifier, which amplifies the tiny oscillator signal up to the level where it can drive the final. My oscillator and buffer are type 41 tubes, which are small power pentodes. They look ancient, but were made in the 1930s and are equivalent to the more modern 6V6 or 6K6 tubes. The 6V6 and 6K6 were manufactured during the 1940s and 1950s and can still be bought from RF Parts Company and other companies.

Vacuum tube amplifiers

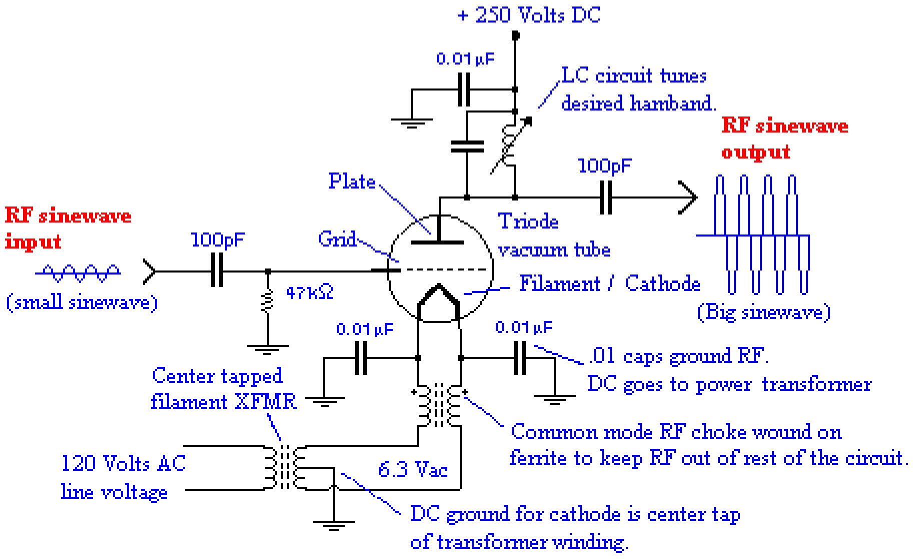
+ 250 Volts DC
0.01 µF
LC circuit tunes desired hamband.
RF sinewave output
100pF
Plate
Triode vacuum tube
RF sinewave input
100pF
Grid
Filament / Cathode
(small sinewave)
47kΩ
0.01 µF
0.01 µF
.01 caps ground RF. DC goes to power transformer
(Big sinewave)
Center tapped filament XFMR
Common mode RF choke wound on ferrite to keep RF out of rest of the circuit.
120 Volts AC line voltage
6.3 Vac
DC ground for cathode is center tap of transformer winding.

A vacuum tube RF amplifier

The simplest tube is a three-element device. The ***control grid*** is analogous to the gate or base of a transistor. The ***plate*** is analogous to the drain or collector and the glowing filament, which acts as the ***cathode***, is comparable to the source or emitter. Tubes are principally voltage-controlled devices, although there is a small current that flows into the control grid. The control grid voltage modulates the much larger current that flows from the filament to the plate. So as you would expect, the input signal is placed on the grid and the output signal is the current passing through a load resistor or load inductor, as shown above.

A filament - three circuits in one

Referring to the diagram above, what is all that stuff wired to the filament? This is the kind of complexity that drives beginners in electronics into other lines of work. Actually, once you unravel the parts, it isn't so bad. The difficulty with using an ancient filament tube is that the filament has three jobs to do simultaneously.

Signal cathode function First, the filament serves as a ***cathode*** to launch our radio frequency current from the chassis ground toward the plate. That is, it serves the same role as an emitter in a bipolar transistor or a source in an FET. RF conducts easily through 0.01 microfarad capacitors, so they conduct RF up from ground into the filament. The common mode RF choke keeps the RF out of the filament transformer. (These chokes were described in chapter 12.)

Heater function Second, the filament must be heated in order to boil the electrons off the filament and into the vacuum. The filament heat function is provided by a 6.3 volt, high current AC signal. So it's necessary to pass a big, low voltage AC current through the filament to warm it up, just like a toaster element. On the other hand, we don't want 60 Hz AC impressed on the signal we are trying to amplify. If we didn't take special steps to prevent it, the guys we are talking to would hear a big 60 Hz hum on our signal.

To keep the AC out of the radio frequency signal, the single loop of filament is powered by AC voltage that is referenced to ground in the center of the filament winding. That is, one side of the filament is connected to 3.15 volts AC, while the other half of the filament is connected to 3.15 volts AC with the opposite polarity. In the center of the filament, the voltage gradient passes through zero volts. The 60 Hz AC filament current biases the current flowing across the vacuum. However, because of the grounded tap on the transformer winding, half of the filament is shedding extra electrons, while the other half is shedding fewer electrons. As a result, the 60 Hz signal is cancelled and the RF signal the amplifier produces is not modulated with 60 Hz. Whew!

All this filament complexity is a nuisance, so newer tubes like the 6V6 and 6L6 have a metal tube surrounding the filament that is heated by the filament. When hot, it is the cathode that gives off the electrons. In that way, the heater function is electrically separated from the signals on the cathode.

DC bias function Third, every tube or transistor circuit has an optimum direct current operating point for the kind of amplifier you are trying to build. For example, if you want a high fidelity audio signal amplified with as little distortion as possible, the bias of a tube is turned half way *ON* at rest. That is, in this way it resembles an N-Channel JFET and is biased to operate class A. In class A, as the audio signal goes positive and negative from the resting point, the amplification will be equal for rising and falling voltage levels. If the operating point were wrong, the upper or lower half of the signal might be amplified more than the other, or perhaps cut off altogether. Any distortion would ruin the sound of the music. Unlike bipolar transistors, the tube is already half-turned-on with no forward bias resistor going to the grid from the plate voltage.

In the case of a CW transmitter RF amplifier, the bias for a tube is simple: We just ground the DC on the cathode to provide the correct bias. In the case of a vacuum tube CW transmitter, we use this DC ground connection to turn the amplifier on and off. To send Morse code the telegraph key is used as a switch between ground and the center tap of the filament winding on the transformer.

The RF sinewave oscillator

Modern radio began with the vacuum tube sinewave oscillator. An RF oscillator consists of an RF amplifier that amplifies its own output. That is, a bit of feedback from the output is returned to the input causing the tube to produce a large sinewave output signal. Tube RF oscillators operate on the same principles as the transistor RF oscillators discussed in chapter 6 and chapter 10. If building an adequate transistorized VFO is difficult, you really don't want to use a tube VFO on the air today. I tried it and the criticism wasn't pleasant. Consequently, my old tech QRP uses crystal control.

Quartz crystals

While building my transmitter, I discovered that the old crystals in my junk box no longer oscillated. When new, crystals are so reliable that it never occurred to me that old crystals might not work. My crystals ranged in age from 20 to 50 years and hadn't been used in decades. The ones that still oscillated did so only under extreme stimulation. Fortunately for me, they were large, old-style FT-243 crystal holders that can be unscrewed and disassembled. I took them apart and cleaned the parts with alcohol. Two of them began to oscillate again, but the others remained inert. I suspect the rubber gasket gives off a vapor that condenses on the crystal and literally gums up the vibrations. In summary, if you try to use old crystals and the oscillator doesn't work, or if it only delivers a tiny signal, you need to clean the crystal or buy a new one.

Triode and pentode oscillators

To be "old-tech," I started with triodes without cathodes. The oldest power tubes I had in that category were the type 68 and 71 tubes. I quickly re-learned that the output signal from a triode varies with the power supply voltage. Those other grids in tetrode and pentode tubes make the amplified signal relatively independent of the power supply voltage. To say it another way, a pentode tube has voltage/current characteristics that resemble NPN or N-channel transistors. This was particularly important for the crystal oscillator. The supply voltage sags slightly when the key is depressed and causes chirp. When using a triode, I could never quite eliminate the chirp, so I ended up using a type 41 pentode. Pentodes work fine.

Circuit "boards"

I used a genuine old-tech circuit board made from a pine tree. A wooden board lets me move parts around inside the metal chassis and swap parts with little effort.

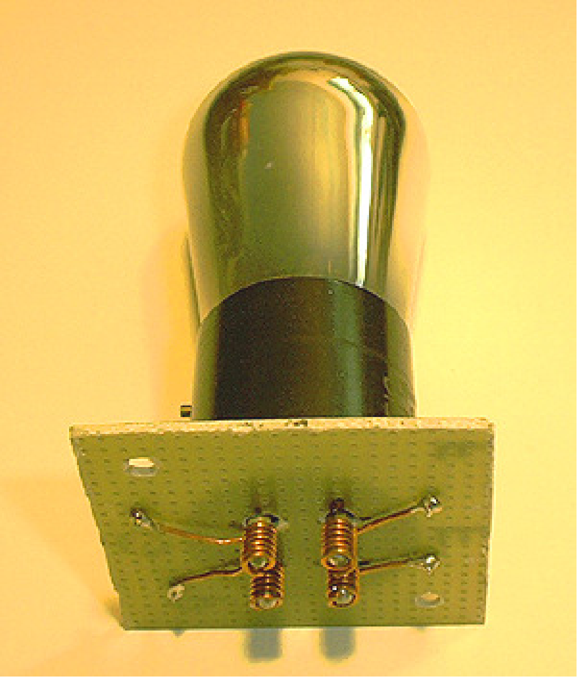

Since 6-pin and short 4-pin tube sockets aren't available anymore, I had to make my own tube sockets by drilling holes into a fiberglass perf-board. To grip the tube pins, I wrapped solid copper wire around them. This worked much better than I expected. I can even unplug and replace tubes.

The best oscillator circuit

There are several different vacuum tube oscillator circuits. The one below is the easiest to understand and produces the biggest output signal. The crystal, which is usually wired in series with a capacitor, is connected between the plate (the output) and the grid input. This puts big voltage on the crystal and produces an extremely strong oscillation. I was able to get 10 watts out of this oscillator with a 300 volt supply. That would be wonderful, but this oscillator is difficult to key on and off for Morse code. Mine started up so slowly it was only able to generate dashes. The oscillator simply would not start in time to send "dots." It sounded more like a sick bird than Morse code. Its chirp was beyond "bad."

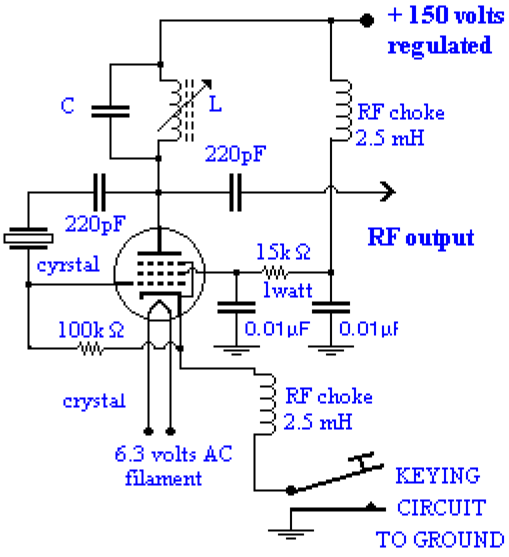
+ 150 volts
regulated
C
L
RF choke
2.5 mH
220pF
220pF
cyrstal
RF output
15k Ω
1watt
0.01µF
0.01µf
100k Ω
crystal
RF choke
2.5 mH
6.3 volts AC
filament
KEYING
CIRCUIT
TO GROUND

A powerful but unstable tube oscillator

A better tube oscillator for CW

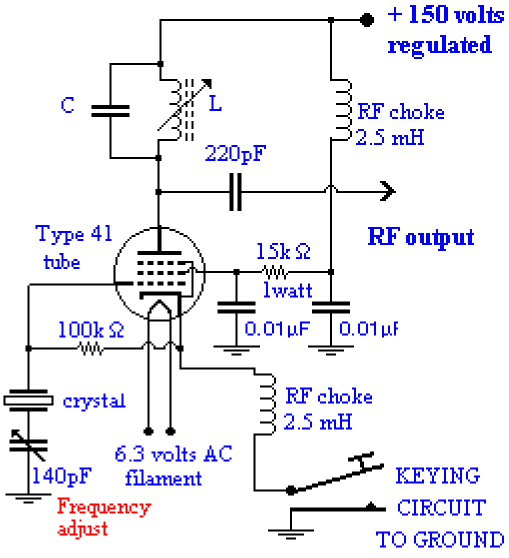
+ 150 volts
regulated
C
L
RF choke
2.5 mH
220pF
RF output
Type 41
tube
15k Ω
1watt
100k Ω
0.01µF
0.01µf
crystal
RF choke
2.5 mH
6.3 volts AC
filament
140pF
Frequency
adjust
KEYING
CIRCUIT
TO GROUND

The circuit on the left is the oscillator used most often in old vacuum tube transmitters. The first question you probably have is, “Where’s the feedback?” There is no visible connection between the input grid and the output plate. However, there is feedback between the cathode and the grid by means of the 100K resistor. There is also feedback by stray (accidental) capacitance between the grid and plate. Remember that a vacuum tube is already turned halfway “on,” even when the grid voltage is zero. If there is any random change in the plate voltage (and/or cathode voltage), a tiny voltage will be coupled to the grid. When this change in voltage appears across the crystal, the crystal will respond, thereby reinforcing an oscillation at the crystal resonant frequency. The advantage of this oscillator is that it keys well and makes good Morse code. The bad news is that the output signal is small. Mine only put out 6 volt peak sine waves. This compared with 30 volt sinewaves when I used plate-to-grid feedback. As seen in the diagram, I eventually settled on using regulated 150 volts plate voltage.

Notice that there is a 1-watt power rating for the 15K Ω resistor. Because of the high voltages, resistors in tube circuits often dissipate at least ½ watt. Wherever the power rating isn't mentioned, assume it is ½ watt. In modern ham transistor circuits it is rare that a resistor needs to be more than ¼ watt and it's easy to forget that power dissipation is even a consideration.

Old time voltage regulation

My tube oscillator was powered by a simple, line-powered transformer high voltage supply. When the oscillator was in “standby,” the plate voltage on the filter capacitor rose to a level proportional to the height of the power line sinewaves. Then, when I pushed down on the telegraph key, the voltage crashed to a lower steady state. This caused a noticeable “chirp” in the tone.

To fix this I used a 150 volt regulator tube. These tubes are the old-time equivalent of a Zener diode. Regulator tubes are simply glass vials containing low pressure inert gas. They have two electrodes. A central wire “cold cathode” runs down the center and a cylindrical plate collects the electrons from the cathode. When voltage is applied across the two electrodes no conduction occurs until the voltage reaches a threshold voltage where the inert gas ionizes into a glowing plasma. As the current through the gas increases, more and more gas is ionized and the voltage across the tube remains approximately constant. By designing tubes of different dimensions and using different mixtures of gases, the tubes regulate at different voltages. To regulate my oscillator supply I used a 0D3 tube that regulates at 150 volts. This trims the ripple off my supply voltage and greatly improves the sound of the CW. The 0D3 regulator glows with a lovely violet color.

The travails of tubes

After building my own rig, I now suspect that “normal” signal quality was really poor back in the bad old days. Yes, I was able to build a one-tube, crystal-controlled oscillator that keyed well and wouldn’t be criticized for key clicks, chirps, drift, sag, or any of the other vacuum tube diseases. Unfortunately, it only put out 300 milliwatts. As the ARRL handbook said, it’s almost impossible to build a one-tube transmitter than doesn’t chirp. I think they’re right. I needed at least a two-stage transmitter.

The final amplifier

I still wanted to use the type 68 triode as my final. 68s have a simple round shape like tubes from the 1920s and they have a big, open filament that looks like an orange, glowing, bread toaster element. You can almost imagine dropping little Pop Tarts down in there for breakfast. The type 71 worked as well as a 68, but the glass has the newer, double-curved shape. Unlike more modern tubes, all three elements of these triodes are plainly visible. In order to operate a triode filament as a “cathode,” the filament transformer winding must be insulated from ground and the DC taken off through a centertap. Since both sides of the filament are DC inputs for the cathode function, the cathode choke should be a common mode design.

I made the common mode choke from an FT50-61 modern ferrite toroid core. Oh, well, close enough. They had pretty good powdered iron cores back then. I wound two 12 turn windings around the core as if I were making a bifilar transformer as described in chapter 6. Instead of connecting the windings together as described earlier, this time I kept them electrically separate and wired them as shown below. The windings at one end of the pair were wired to the filament, the windings at the other end were connected to the filament transformer as shown below. The dots on the transformer mean that if positive voltage appears at the end of the winding with a dot, then positive polarity will appear on other winding at the location of the second dot.

This is the reverse of the way the common mode was wired in chapter 12. I just looked at my linear amplifier and compared the choke wiring with the choke winding in the tube final. The dot marks below are correct as drawn, or at least, both of these units work well when built that way. Feel free to reverse the windings and see what happens!

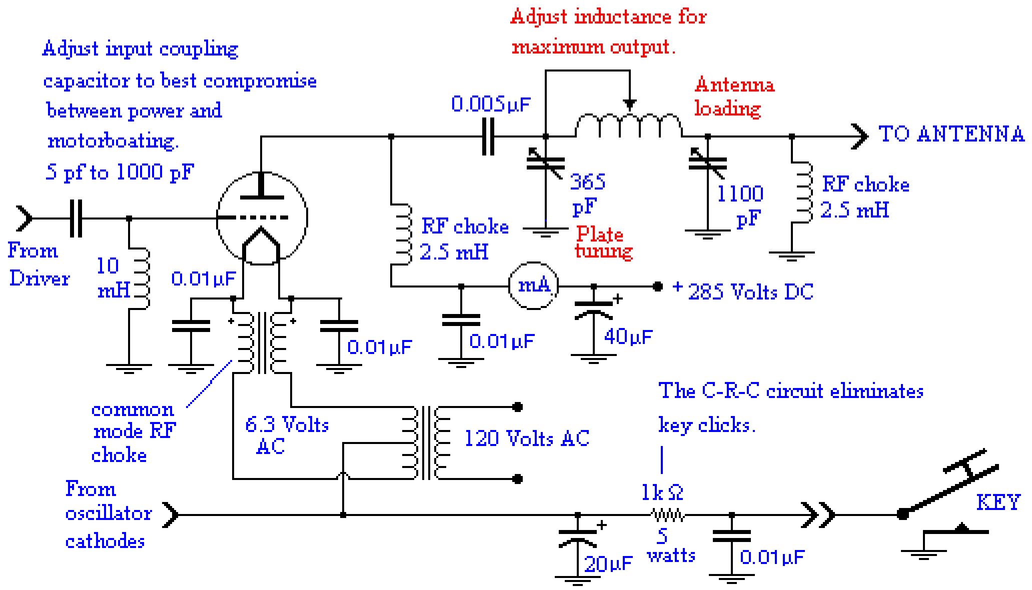
Adjust inductance for maximum output.
Adjust input coupling capacitor to best compromise between power and motorboating.
5 pf to 1000 pF
0.005µF
Antenna loading
TO ANTENNA
365 pF
1100 pF
RF choke 2.5 mH
From Driver
10 mH
0.01µF
RF choke 2.5 mH
Plate tuning
mA
+ 285 Volts DC
0.01µF
0.01µF
40µF
common mode RF choke
6.3 Volts AC
120 Volts AC
The C-R-C circuit eliminates key clicks.
From oscillator cathodes
1k Ω
5 watts
20µF
0.01µF
KEY

My triode final is shown above. Rather than amplify the input signal, the triode preferred to oscillate all by itself. When the final wasn't self-oscillating, the drive from my puny driver oscillator was insufficient to produce any useful power. I wondered how typical novice transmitter kits 40 years ago managed with only two tubes. I concluded, "It must be those high gain pentode finals!" I threw together a 6L6 final and tried it out. No change. The 6L6 had the same gain as my ancient triodes. If this was going to work, I needed more grid drive. I finally concluded that those old Heathkits, Knight kits, and Viking Adventurers were just darn well engineered. They achieved terrific performance with few parts.

Key clicks

Notice the C-R-C filter in the keying circuit. That's what I had to add to get rid of the key clicks. Key clicks are sharp, unpleasant clicking sounds on the Morse code caused by the dots and dashes turning on and off too rapidly. Although my transistorized CW transmitters seem to have rise and fall times just as abrupt, my transistorized rigs have had practically no problem with clicks. I don't understand this, but the C-R-C filter in front of the key causes the tubes to key on and off gently and the tube transmitter no longer clicks.

Buffer amplifiers

After struggling for a few hours, I eventually added a "buffer amplifier" made from another type 41 pentode between the oscillator and the final. Now I had enough drive for the triode final.

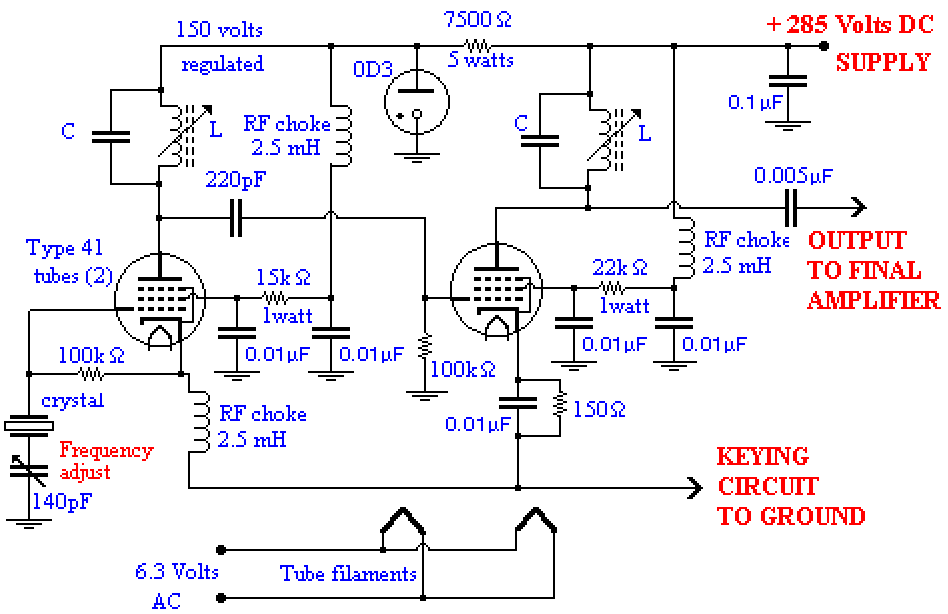
150 volts
regulated
7500 Ω
5 watts
+ 285 Volts DC
SUPPLY
0D3
0.1 µF
C
L
RF choke
2.5 mH
C
L
220pF
0.005µF
Type 41
tubes (2)
15k Ω
1watt
22k Ω
1watt
RF choke
2.5 mH
OUTPUT
TO FINAL
AMPLIFIER
100k Ω
0.01µF
0.01µF
0.01µF
0.01µF
100k Ω
crystal
RF choke
2.5 mH
0.01µF
150 Ω
Frequency
adjust
140pF
KEYING
CIRCUIT
TO GROUND
6.3 Volts
AC
Tube filaments

Oscillator and buffer stages

The oscillator tube drives a buffer amplifier to increase the drive to the final. Notice the 0D3 regulator tube for the oscillator plate supply. A 7.5K ohm resistor drops the 285 volt supply down to 150 volts while the regulator tube holds it at that level.

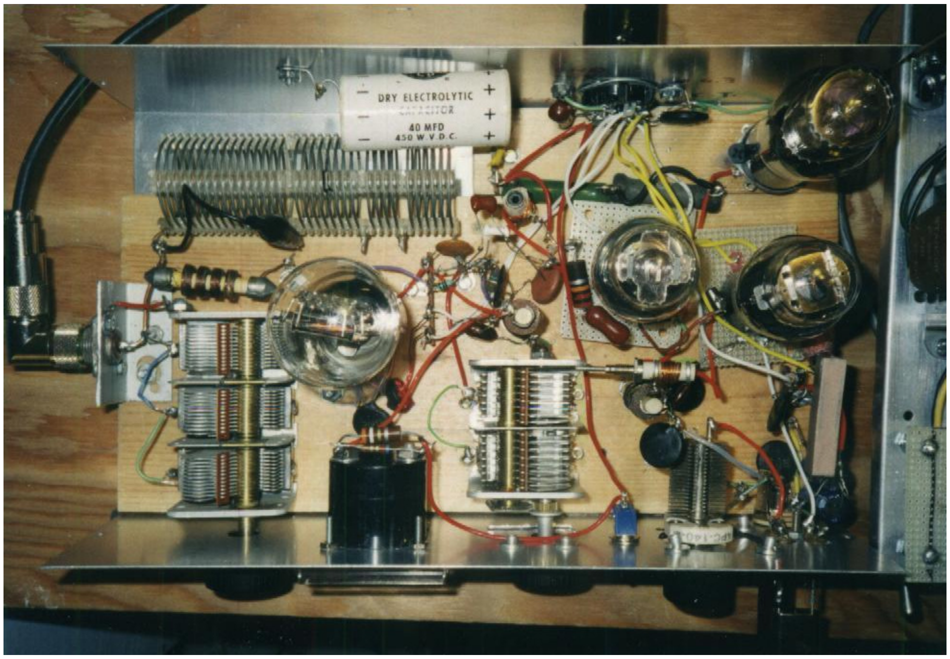
DRY ELECTROLYTIC
40 MFD
450 W.V.D.C.

The completed 40 meter transmitter

The final amplifier is on the left. The oscillator, buffer, and regulator tube are at the right rear. The plate coil is the long, tapped coil at the rear. In theory it was supposed to work on several bands, which is why the coil has the multiple sections. However, so far I have only gotten it to work on 40 meters. The plate coils for the oscillator and buffer were wound on slug-tuned coils, which were the popular way to do it back in the vacuum tube days.

Triodes chirp

While up to my ears in vacuum tubes, it dawned on me that those homebrew “kilowatt linear amplifiers” in the old ARRL handbooks made from triodes can not really be linear. The output voltage of a triode varies with the supply voltage, not just the grid voltage. Well, they could be linear, but the operating range along the load line would have to be really narrow and the supply would have to be perfectly regulated. I doubt those designs meet either of those criteria.

THE POWER SUPPLY

An inadequate power supply made from old-fashioned parts

Before I could try out my QRP transmitter, I needed a power supply. I first tried to use a power supply salvaged from a junked 1935 radio. Unfortunately this power supply didn’t deliver enough power for a QRP transmitter but, as always, it was educational.

OFF
PLATE VOLTAGE

A power supply made from parts salvaged from an ancient radio.

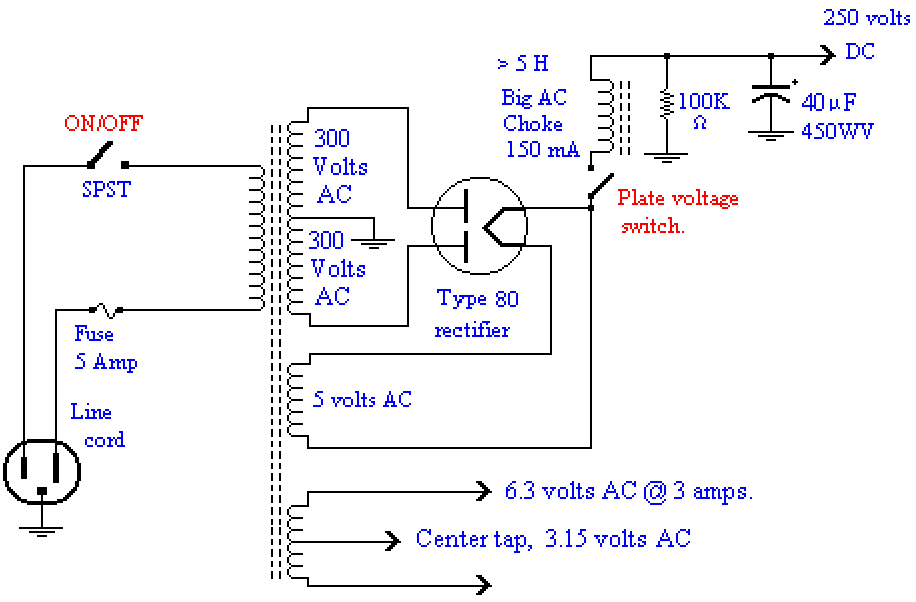

Schematic of the 1935 power supply

The AC power from the wall comes into the supply and first encounters a 5 ampere fuse. In case the supply is shorted, the fuse can blow and prevent further damage. Five amperes is a larger fuse than needed, but it's within electrical standards for small appliances. A switch connects the power transformer primary winding with the AC line.

The transformer is equipped with three secondary windings. A high voltage winding puts out about 300 volts at 30 milliamperes. Actually, it has two 300 volt windings that allow us to achieve full wave rectification with just two diodes. The type 80 tube contains two anodes that share a common cathode/filament. A filament winding delivers 5 volts at about 2 amperes. This winding is used exclusively to heat the rectifier filament. The winding also has the 300 volts DC on it, so it must remain separate from other circuits. An independent filament winding delivers 6.3 volts at about 3 amperes.

The 6.3 volt winding is center-tapped. In other words, it is a 3.15 volt winding in series with a second 3.15 volt winding. The center tap is essential if you wish to use an ancient tube that has no separate cathode and must use the filament as a cathode. The center tap offers balanced access to the filament circuit to complete the main circuit between the plate and filament. Without a centertap, 60 Hertz power line noise will be impressed on the current and will appear in the transmitted signal.

Tube rectifiers

Before the high voltage can be used, it must be rectified and converted to DC. Vacuum tube rectifiers are a kind of diode and only conduct electricity in one direction. Specifically, they conduct electrons from the filament (or heated cathode) to the plate. The round circle symbol with the single filament and two plates is a dual, "full wave" rectifier. It converts both halves of the AC signal into a single DC signal in which both rectified halves of the sinewave have positive polarity. Both sides of the high voltage winding are connected to the two plates so that, at any moment, one side is always conducting while the other is not. The two positive voltages are added together at the common filament for a single, continuous, positive output.

Ripple filtering

In order for Morse code to sound clear and pure, the voltage driving the transmitter must have no ripple left over from the sinewave. This is done with an "L" filter. An inductor and a capacitor make the two arms of the "L." The choke, being a large 5 Henry inductor, doesn't allow the current flowing through it to change rapidly. This prolongs the charging period of the capacitor and helps smooth out the ripple. Many designs add a second big filter capacitor on the input side of the choke. This practice causes big surge currents to flow into the first capacitor and may saturate the transformer. The saturation in turn causes the transformer to heat and probably isn't as effective as putting both capacitors on the downstream side of the choke.

So much for theory. In practice, a receiver needs far less DC current than even a small transmitter. One difficulty was that the old-fashioned type 80 tube rectifier cost me 50 volts drop all by itself. Another problem was that the L and C filter were insufficient to prevent the ripple from making the Morse code sound rough and harsh.

But worst of all, the transmitter drew so much current that the voltage dropped precipitously and it often transmitted in "bursts" called ***motorboating.*** One cure is to use the smallest input coupling capacitance to the grid of the final amplifier. With luck, this can stop the motorboating and still give good output. With a soft (sagging) power supply the coupling cap may need to be as low as 5 pF on 40 meters.

I also tried to use an ancient type 83 mercury vapor rectifier tube. This is a vacuum tube rectifier that contains a touch of mercury vapor gas inside. This lowers the resistance of the tube and raises the current capacity of the power supply. The type 83 was a big improvement, but still not adequate for on-the-air use. I've had quite enough bad signal reports, thank you.

Checking out an old transformer

Big power transformers have always been costly. That's why my 1957 novice transmitter power supply was so wimpy - I thought I couldn't afford a bigger one. The old ham radio handbooks suggested using power transformers from vacuum tube TV sets. However, junk vacuum tube TVs are rare today. If you have such an old TV, remove the transformer from the chassis carefully being sure to label which pair of wires (usually black) go to the AC line.

Suppose some friend from your ham club gives you a power transformer and assures you it is "perfect" for building a transmitter. He hands you a 5 pound lump of iron with a nest of unmarked black wires protruding from the bottom of the transformer. You think to yourself, "Now what?"

The first step in checking out an old transformer is to strip the ends of the wires so you can measure the resistance of each winding. If you're lucky, you will find that, once you clean off the dirt and tar, the wires are color-coded. The most common color codes are as follows:

two black wires - the 120 volt primary winding.

two red wires - the high voltage secondary for the plate supply, typically 1000 volts centertap for a 50 watt transmitter.

one yellow and red stripped wire - the center tap of the high voltage winding. This wire will be grounded in a full-wave rectifier circuit.

two yellow wires - a 5 volt filament winding is usually good for 1 or 2 amperes. This is used exclusively for the rectifier filament.

two green wires - a 6.3 volts filament winding. This supplies the filament current for the transmitter tubes.

one yellow and green stripped wire - This is the center tap for the 6.3 volt winding. You'll need this if you use an old-time triode vacuum tube final amplifier. If you use a more modern tube with a separate cathode, then just tape off the end of the centertap wire and fold it up neatly. Don't cut it off. Someday you may want it.

Often there are no visual clues telling you which winding is which. Or sometimes the windings are just solder eyelets mounted on Bakelite strip at the edge of the transformer. Using your ohm-meter, measure the resistances of the windings picking them out one by one. Label them with tape and write down the resistances. The highest resistance winding will be the high voltage winding. For example, it might read, say 40 ohms. The high voltage center tap will have half as much resistance as the whole winding measured end to end. The 120 volt primary winding will have the next highest resistance. The lowest resistance windings will be the two filament windings. These will have an ohm or two at most.

Once you think you have the windings figured out, you need to test them before bolting and soldering the transformer into your transmitter. If this isn't going to work, you want to know as soon as possible. Now that you are considering plugging it in, you have a great opportunity to electrocute yourself, trip circuit breakers, and start fires. But if you're careful and proceed thoughtfully, testing a transformer can be done safely. The idea is to limit the current that flows into the transformer. That way nothing terrible will happen even if you screw up or if the transformer is internally shorted.

Testing a power transformer

First you must expose each of the many windings and keep them un-shorted. You will need to measure the open circuit AC voltage of every lead pair. It would be best to set up your voltmeter with test clips so that you don't need to touch them with bare fingers. Another way is to tape the wires to a board so they hold still. To start you must connect the winding you believe to be the primary to a source of 60 Hz voltage. How safe you want to be is your decision. But AT VERY LEAST don't plug the test transformer into the power mains without soldering a 10K ohm resistor in series with the primary winding. That way, even if the primary is a dead short, little current will flow and nothing bad will happen. Go through the windings one by one and you should be able to confirm the guesses you made earlier from the resistance measurements. Whatever you do, don't just connect it to a line cord and plug it into the wall.

An adequate tube QRP power supply made from modern parts

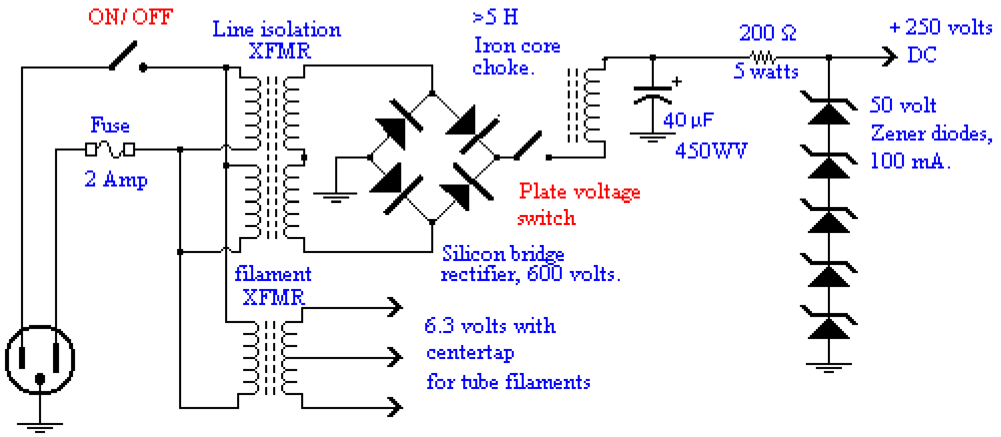

When my first supply worked poorly, I built another made from modern transformers and semiconductors. This supply was adequate to run the vacuum tube transmitter. Ideally it produces 250 volts at 200 milliamperes. This time there was no significant voltage drop across the modern bridge rectifier and the modern Zener diode string clipped off the ripple and gave me a pure CW tone. That was easy.

I couldn't find an adequate, triple-winding transformer, but there are alternatives. Magnetek makes an isolation transformer that can isolate an appliance from ground for shock protection. (Magnetek / Triad part # VPS230-350 sells for about $30.). It has four 120 volt windings. Each winding is rated at 350 milliamperes - plenty "stiff." That is, this supply behaves like a good voltage source over a wide range of current. These two pairs of windings may be connected in series or parallel. For this application you could put two windings in parallel for a 120 volt primary. The secondary windings can then be connected in series to give you a 240 volt secondary rated at 350 milliamperes. Notice the dot markings on the windings. These tell you the phase direction of the sinewave voltage. On the primary side, you must connect the leads so that the dots should be connected together. That way, the two windings will work together.

On the secondary side, the dotted end of one winding goes to the un-dotted end of the other winding. Notice that if you connect the secondary windings so that the two dotted ends or the two non-dotted ends are tied together, the voltages will cancel each other instead of add. Instead of 240 volts you would get zero volts.

The other power requirement for vacuum tubes is a 6.3 volt supply to light the filaments. Each tube filament needs about 0.5 to 1.2 amperes of current. Notice that roughly 5 watts is expended just to heat up each tube! Tube transmitters are inherently inefficient.

I used four modern silicon rectifier diodes that were rated for 600 volts peak inverse voltage and I wired them to make a bridge rectifier. Since you aren't using a tube rectifier, you won't need a 5 volt filament winding like the one shown earlier in the 1935 power supply. However, you will need a separate 6.3 volt filament transformer for your transmitter tube filaments. To regulate the output voltage I used five 50 volt Zener diodes in series.

My supply made from modern parts can put out 42 watts and that's sufficiently "stiff" to deliver my measly 10 watts without sag. Using the modern silicon rectifiers I got 4 watts RF out of my transmitter.

A power supply for the tube transmitter constructed from old style parts

Although I now had the rig working, I still wanted to use my old time parts exclusively. Unfortunately, to use an old tube rectifier with full-wave rectification, I needed 480 volts centertap. So I added a second transformer. This worked, but the extra voltage drop across the type 83 tube rectifier cost me half my 4 watts! Finally I added yet another transformer and got the average, well-regulated DC voltage up to about 285 volts.

I tried using gigantic filter capacitors, but they weren't available in 1935 and they caused long-term sag. I was surprised to discover that a total of two 40 microfarad caps were plenty when used with voltage regulators. To regulate my 285 volts DC I used one 0A3 (75 volts) and two 0C3 (105 volts) gas tube voltage regulators in series. 0A3s glow with a spectacular orange color while 0C3s glow violet.

In order to regulate while the transmitter is keyed, the voltage regulator tubes should be at least slightly glowing at all times. For best regulation, you may have to reduce the 200 ohm current limiting resistor to 150, 100, or even zero ohms. You will probably find that the regulators only stay lit while the key is down when the transmitter is perfectly tuned to the antenna. When it is out of tune, the transmitter draws too much current and the regulator tubes will be extinguished and essentially take themselves out of the circuit.

I have seen these old gas regulator tubes offered for sale at as much as $30 each! Hopefully you can find some at a swap fest. Notice that six teeny 100 mA 50 volt Zeners cost less than $2 each. Anyway, when all was said and done I had 5 watts of pretty good sounding CW. Behold! I have created the world's most inefficient transmitter!

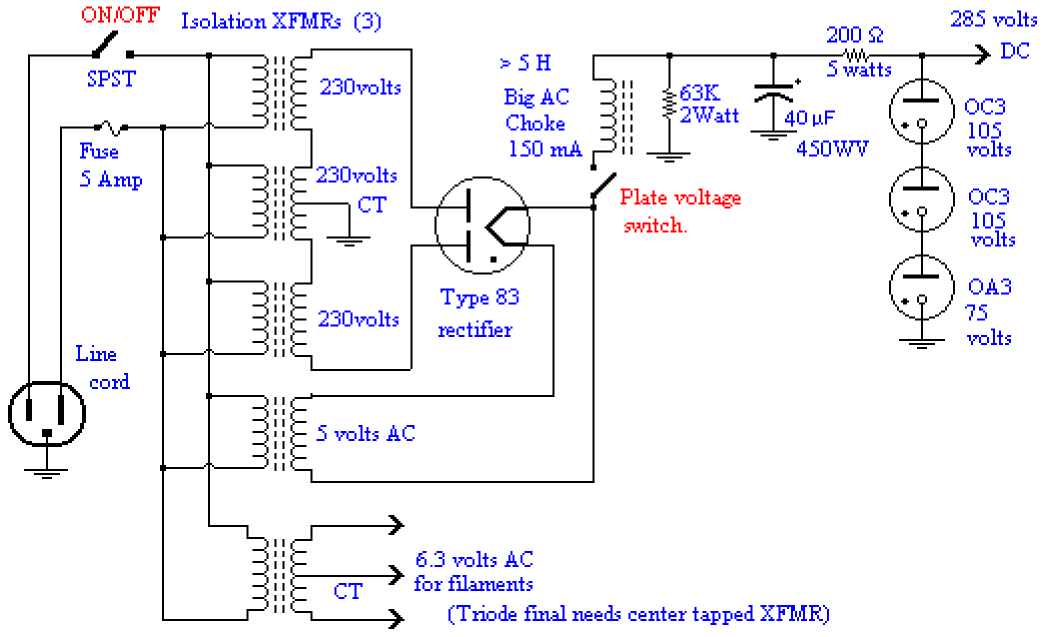
ON/OFF
Isolation XFMRs (3)
SPST
Fuse
5 Amp
230volts
230volts
CT
230volts
Line
cord
5 volts AC
CT
6.3 volts AC
for filaments
(Triode final needs center tapped XFMR)
Type 83
rectifier
> 5 H
Big AC
Choke
150 mA
63K
2Watt
40 µF
450WV
200 Ω
5 watts
285 volts
DC
Plate voltage
switch.
OC3
105
volts
OC3
105
volts
OA3
75
volts

Schematic of the regulated QRP supply made with old-style parts

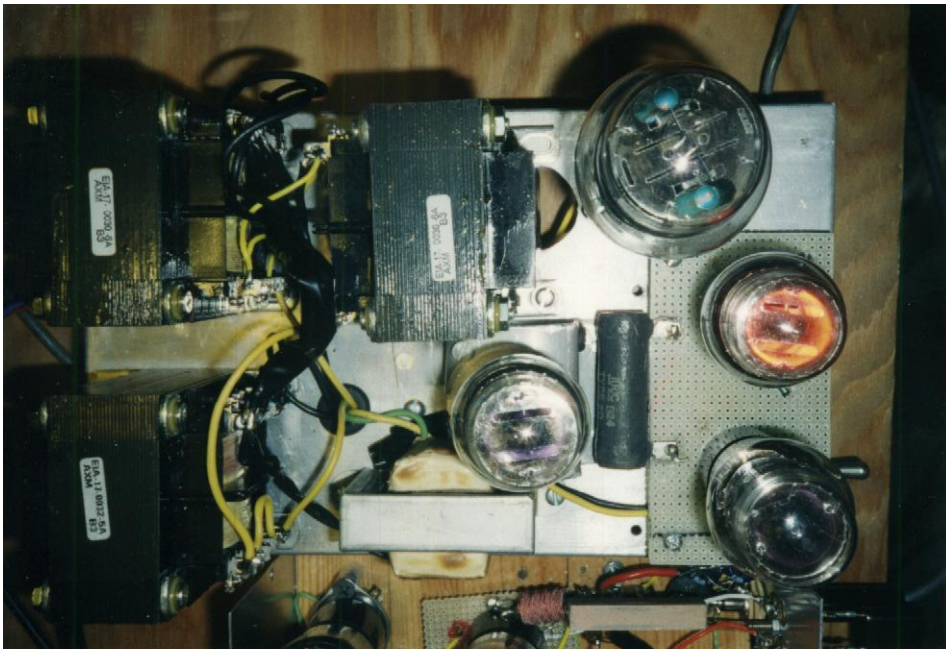

It isn't that you can't build modern high tech stuff with 1935 parts, it's just that the assembly becomes huge and expensive. Meeting modern specs with old parts reminds me of a scene in the movie Back to the Future III. In this time travel movie Doc Brown replaced a burned up integrated circuit from the future with a tube circuit the size of a laundry basket. It must have been a low scale integration chip!

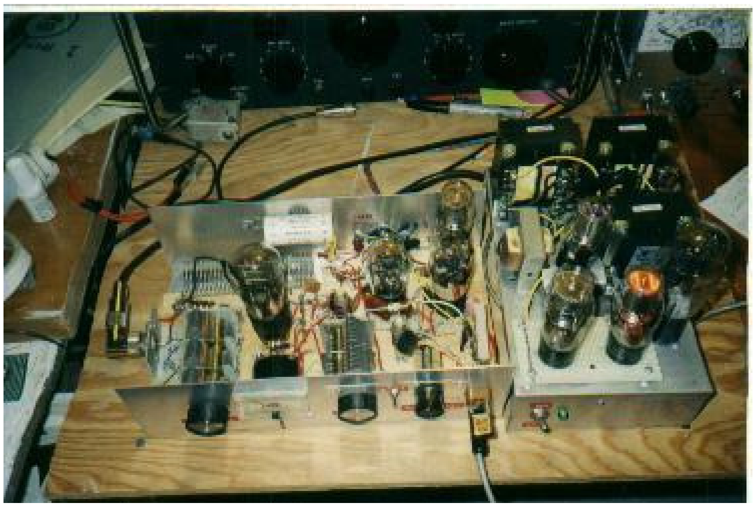

Old-tech QRP

A happy ending

In conclusion, old parts aren't efficient in any way. On the other hand, you get to use a whole lot of them to accomplish very little and they look terrific. The voltage regulators glow orange and violet, but when you key the transmitter the glow flashes on and off inversely with the CW. When I first turn the mercury vapor rectifier on, mercury condenses on the glass in a silver-white fog. Then as it heats, the mercury evaporates and the tube runs with a glorious blue glow with bright orange filaments. Great show biz! Oh, yeah. I forgot to mention the “mercury vapor rectifier hash” that you get to listen to on receive. Oh, and another thing - don’t place your transmitter supply transformers close to your receiver. The receiver may pick up a 60 Hz hum. You’ll probably have to turn off the transmitter high voltage supply during receive. There are lots of reasons these old parts are extinct.

Every year my local ham club has a "homebrew night." Hams who have built something during the previous year bring in their devices and give a brief show and tell. Over the years I have showed off most of the projects described in this book. They include all the most difficult projects, the SSB transmitter, the dual conversion receiver, etc. It's fascinating to me that the project my fellow hams remember best is the vacuum tube QRP. I demonstrated it with a Christmas tree light bulb for a dummy load and they all oohed and aaahed over the pretty glowing regulator tubes and lights. In the years since, they always act disappointed that I haven't brought in another vacuum tube light show. Food for thought!

So anyway, I can successfully work people with my old-tech QRP. QRP is quite popular. I call CQ, mention “/QRP” and I get lots of responses from other fanatics. Who else would answer an American “Ø area” RST 449 CQ? Usually we can barely hear each other, but we’re having fun. However, you never know! I got a 599 from Albuquerque. Maybe he was just kidding. Anyway, how old-tech can **YOU** go?

****

OLD-TECH SUPER-REGEN

How old fashioned a receiver can you use on the air today?

Once I had my old-tech, 1935-era CW transmitter working, I thought it would be fun to complete my ancient rig with an old-tech receiver. To be successful on the CW bands, I figured I needed a superhetrodyne with a crystal filter. In 1967 I built a pretty good vacuum tube CW receiver so I had little interest in doing the same project again. (There is a photo of this receiver in chapter 13.) The 1930 tubes would be much larger than 1960 tubes, but otherwise it would be the same project.

Yardley Beers, WØJF, told me about the regenerative receiver he built in 1930. I doubted it would work well enough for today’s bands because I had a previous experience with regenerative receivers. So Yardley didn’t inspire me to build another one. Then a few months later I worked Mike, NØMF, who lives about 1,000 miles from me. He was using a homebrew super-regenerative receiver made from modern FETs as described in QEX magazine. Obviously it was sensitive and selective enough to hear my QRP on 40 meters. Maybe a super-regenerative receiver wasn’t so silly.

Ancient tubes

I have never been able to bring myself to throw out vacuum tubes. I have boxes of them in my attic. I’ve always been especially fascinated by the earliest tubes. They have lightbulb-like glass envelopes, bayonet sockets and numbers like 201, 216, 224, 227, and 301A. This was my chance to build a really old-tech receiver.

I started by building an RF amplifier out of a type 201 tube. I had no specs for a 201, but I gently increased the filament voltage until I got a nice warm, orange glow out of the filament. That occurred at 2.5 volts and 0.3 milliamperes. Unfortunately, it didn't amplify the RF noise from the antenna at all. The signal on the plate was smaller than the input on the grid. I tried reversing the grid and plate wires … no luck. No matter how I biased the grid, the tube didn't turn on. I increased the filament current to produce a yellow glow and promptly burned out the filament. Whoops! Starting over, I tried to build a regenerative oscillator, but it refused to oscillate or amplify. I suspect that these tubes from the 1920s are extremely high impedance.

Old-tech receiver

Since my oldest tubes seemed inert, I gave up on the triodes before I ruined more of them. From there I retreated to a high-tech type 224. This tube is a tetrode. It has two grids and even has a cathode. Pretty modern, huh? 224s are the early version of the types 24 and 24A tubes which were commonly used in the 1930s. The 224s worked immediately in the same circuits I had tried before and soon my super-regen was picking up 40 meter hams and 31 meter commercial broadcasts. The 31 meter stations are mostly religious sermons, but the sound quality was pretty good. One of my 224 tubes has a test label on it from the Marshing Radio and Electric Company at 246 Main in Longmont, Colorado dated December 19, 1932. The note on the label written in pencil says, "doubtful." Shows what they knew!

Like the older tubes, the type 224 and 24 tubes have low voltage (2.5 volt) filaments. A major difference is that the filaments run at much higher power levels to heat the heck out of the cathode so it will shed electrons readily. The high temperature gives the tube much lower internal resistance. The brawny filaments resemble toaster elements and are hard to burn out. The bad news is that each tube consumes 4.4 watts just to run the filament.

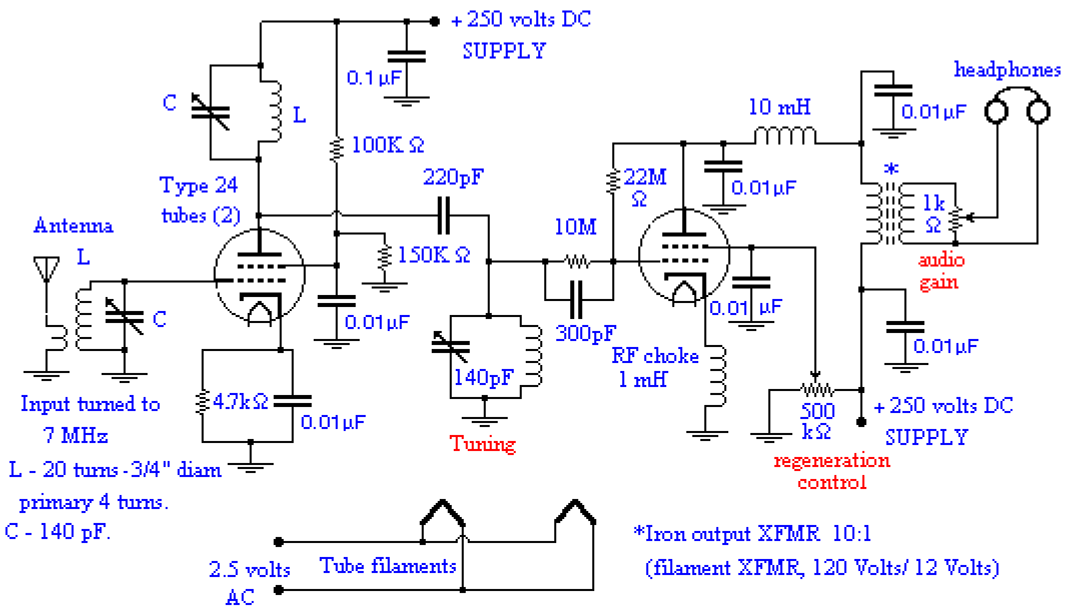
+ 250 volts DC
SUPPLY
0.1 µF
C
L
100K Ω
headphones
10 mH
0.01µF
220pF
22M Ω
0.01µF
Type 24
tubes (2)
*
1k Ω
Antenna
L
10M
150K Ω
audio
gain
C
0.01µF
0.01 µF
300pF
RF choke
1 mH
0.01µF
140pF
Input tuned to
7 MHz
4.7kΩ
0.01µF
Tuning
500 kΩ
+ 250 volts DC
SUPPLY
regeneration
control
L - 20 turns -3/4" diam
primary 4 turns.
C - 140 pF.
2.5 volts
AC
Tube filaments
*Iron output XFMR 10:1
(filament XFMR, 120 Volts/ 12 Volts)

Circuit diagram of the old-tech super-regenerative receiver

Motorboating can be a good thing

Super-regenerative means that the amplifier is deliberately designed to "motorboat." That is, the power supply is deliberately given a high source impedance with a series inductance (10 mH). This causes the tube to oscillate in short bursts. Normally, when RF circuits motorboat it's a darn nuisance and tough to get rid of. It's ironic that motorboating was my big problem with the transmitter and here it is vital to achieve the receiver sensitivity. The 10 mH choke can be seen in series with the plate of the tube on the right. The motorboating typically occurs at a rate of 1 KHz.

In a super-regen motorboating does three useful jobs: Maximum sensitivity and selectivity are obtained during the feedback near the peak of each burst. The sidetone caused by this oscillation makes an audio squeal that serves as a BFO. If you are listening to AM modulation, the regeneration control can be turned down until just below the point where the squeal begins.

Super-regeneration re-broadcasts what you're hearing

Unfortunately, if you wish to use a super-regenerative receiver today, it must be a two stage device. An RF amplifier (or other device) must isolate the oscillator from the antenna. Otherwise, you'll be rebroadcasting everything you're listening to on your receiving frequency. Even with an RF isolation amplifier, you'll be radiating a few milliwatts of RF signal. To demonstrate this, disconnect the antenna from your modern communications receiver so there are no signals audible on the band. Now tune the modern receiver to the same frequency as your super-regenerative receiver. Suddenly, coming out of the modern loudspeaker you'll hear exactly what you're hearing in the headphones of your super-regen. The super-regen really does rebroadcast the signals it picks up!

Super-regenerative design

When you think about it, a one-tube super-regenerative receiver is actually quite amazing. One tube is providing RF amplification, detection, audio amplification, and a BFO all in one tube. 80 years ago this economic efficiency made it possible for hams to afford to build receivers at a time when a single tube was a major investment.

Regeneration is accomplished by feeding back the output from the plate to the grid. In my circuit this is done with the 22 megohm resistor from plate to grid. The gain of the tube is adjusted by the pot that controls the voltage on the second grid. The second grid is called the ***screen grid***. The oscillation provides far more gain than a simple RF amplifier can achieve. It takes good sensitivity to hear hams on the HF bands. Moreover, the self-amplification greatly increases the circuit Q, making the amplifier much more selective than it would otherwise be.

For an audio output transformer I used an iron 12 volt filament transformer, i.e., 120 VAC to 12.6 VAC. It works fine and was easier than trying to find a real old time audio transformer. The secondary side goes to the low impedance headphones. The 120 volt primary side goes to the tube plate circuit.

The power supply

Bob, K6VGA, told me about the homebrew super-regen receiver he built back in 1935. He said its worst fault was AC hum. It was so bad, he replaced the AC power supply with batteries. After that story I figured I would have to filter the heck out of the power supply with a big choke, big capacitors, and a voltage regulator from the 1940 era. However, a voltage regulator turned out to be overkill. Unlike my old-tech QRP transmitter, the power supply for the super-regen turned out to be remarkably easy. In other words, simple power supplies like the ones I tried to use for my transmitter worked fine for the receiver. Yes, the receiver hums slightly, but too little to bother me.

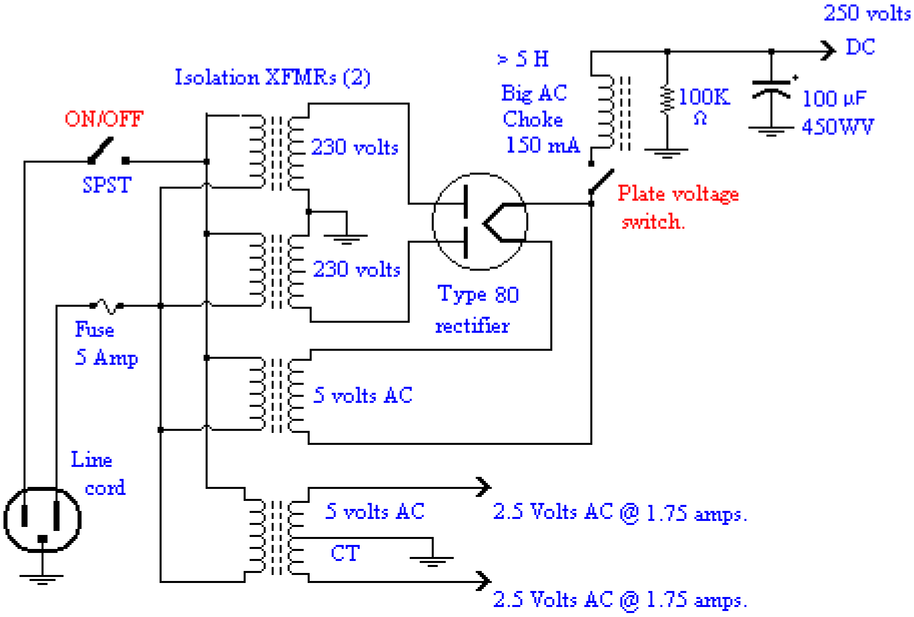
250 volts
DC
> 5 H
Big AC
Choke
150 mA
100K
Ω
100 µF
450WV
Isolation XFMRs (2)
ON/OFF
SPST
230 volts
Plate voltage
switch.
230 volts
Type 80
rectifier
Fuse
5 Amp
5 volts AC
Line
cord
5 volts AC
2.5 Volts AC @ 1.75 amps.
CT
2.5 Volts AC @ 1.75 amps.

An old-tech power supply for the super-regen.

As explained earlier, you can't buy a new, old-time tube transformer with both high voltage and rectifier filament windings anymore. Brent, KD0GLS, pointed out to me that tube transformers are still manufactured by Hammond Mfg. for tube hi-fi amplifiers. Some modern hi-fi addicts say tube amplifiers sound better, so there is still a demand for them. Perhaps one of these transformers would work. In any case, 5 volt filament transformers are still available as separate units. A center-tapped 5 volt, 2 ampere transformer is needed to provide the 2.5 volt filament voltage for the type 24 tubes. You could also use a 5 volt winding and wire the two tube filaments in series - that is what I did. The tube filament transformer must be separate from the transformer winding used for the rectifier. Otherwise all the filaments would have +250 volts DC on them.

In my receiver I used an old plate transformer I happened to have. However, you may use two of the same transformers I used for my old-tech QRP transmitter. For a choke I used the primary of yet another filament transformer. Because the current drain is so low, a voltage regulator tube isn't needed.

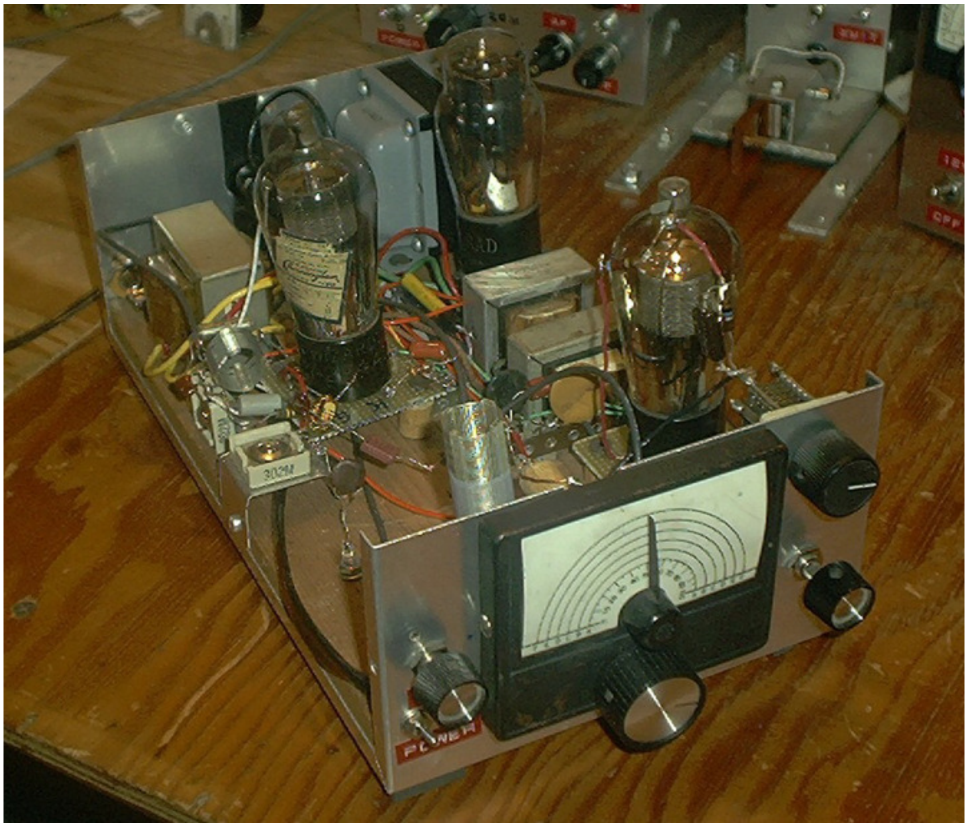

Super-regen on the modern hambands

In summary my super-regen worked quite well for listening to shortwave broadcast stations on the 31 meter band, around 10 MHz. These are extremely powerful stations like Die Deutsche Welle (Radio Germany) and various religious stations. To hear AM stations, the regenerative control is turned down just below where it begins to squeal.

To hear CW stations, the regeneration is turned up just above the squeal. The squeal modulates the Morse code and makes it audible. Unfortunately, as I feared, my super-regen isn't sensitive and selective enough to be used on the air today. Yes, I heard lots of ham CW stations on 40 meters, including guys I had worked before. The stability seemed OK once a signal was tuned in. But that's because I was hearing at least 20 KHz bandwidth at once. In other words, I was hearing half of the CW stations on 40 meters simultaneously. The worst problem is that the reception is "flaky." If I just touch a control, the pitch of the CW signal jumps to inaudible. Or it leaps from a nice tone to a harsh, static-like sound. When I try to tune it in better, the signal is likely to vanish altogether. Yes, I can work guys with this receiver, but it would be a stunt, rather than practical. How does it compare with the little direct conversion receiver described in chapter 7? There's no comparison. The transistorized direct conversion receiver is far better in every way.

In conclusion, the modern FET super-regenerative designs apparently achieve much higher Q and stability than I got using gigantic old-time parts with their stray capacitance and leakage inductance. But the project was fun and I learned a great deal. On the other hand, there's no reason I couldn't build a 1930 superhetrodyne using 6 or 7 ancient tubes and 1927 crystals. That would fill a huge chassis and require loads of work. Until I can muster the enthusiasm, I guess a practical old-tech receiver will remain one of those unfinished projects in my attic. But wait! I just got a QSL from Biz, WD0HCO. He claims that when he worked me he was using a super-regen made from two type-30 tubes. He found the design in a 1932 ARRL handbook. Hmmm… Maybe I’m being too hasty.

CRYSTAL SETS TO SIDEBAND
© Frank W. Harris 2010, REV 12

Chapter 15

THE NOBEL PRIZE FOR SIDEBAND

How sideband works

In the beginning of the book I mentioned that Glenn Johnson, WØFQK, was an elementary school principal who made a project out of recruiting kids into ham radio. We were in the 8th grade walking down the street minding our own business when Glenn ran out of his house and grabbed us. "Come on in boys and I'll show you how sideband works!" Glenn's wife served milk and cookies while Glenn worked bunches of guys on 20 meter sideband phone. I sat quietly and watched while Glenn effortlessly operated massive equipment that cost enough to buy a car. I was fascinated by ham radio, but I didn't learn much about how sideband worked. I had the impression that sideband was *MODULATION FOR MILLIONAIRES* and too complicated to homebrew. The 1957 ARRL handbook's opaque descriptions of "phase shifters" and "balanced modulators" only confirmed my opinion.

Today SSB is affordable, but the technology is still exotic to the average ham. I overheard a conversation at my local ham club meeting that went something like this: "I once knew a guy who built his own sideband rig." "*REALLY*! That's amazing. Are you sure it wasn't a kit?" The implication was that homebrewing sideband was about the same level as a Nobel Prize in physics. So, is anyone interested in the Nobel Prize for Sideband? If you've already built homebrew QRPs, VFOs, and a receiver, then sideband is the next logical project. SSB uses all the same basic circuits. Besides, you won't really understand sideband until you've built one.

You begin with the sideband generator

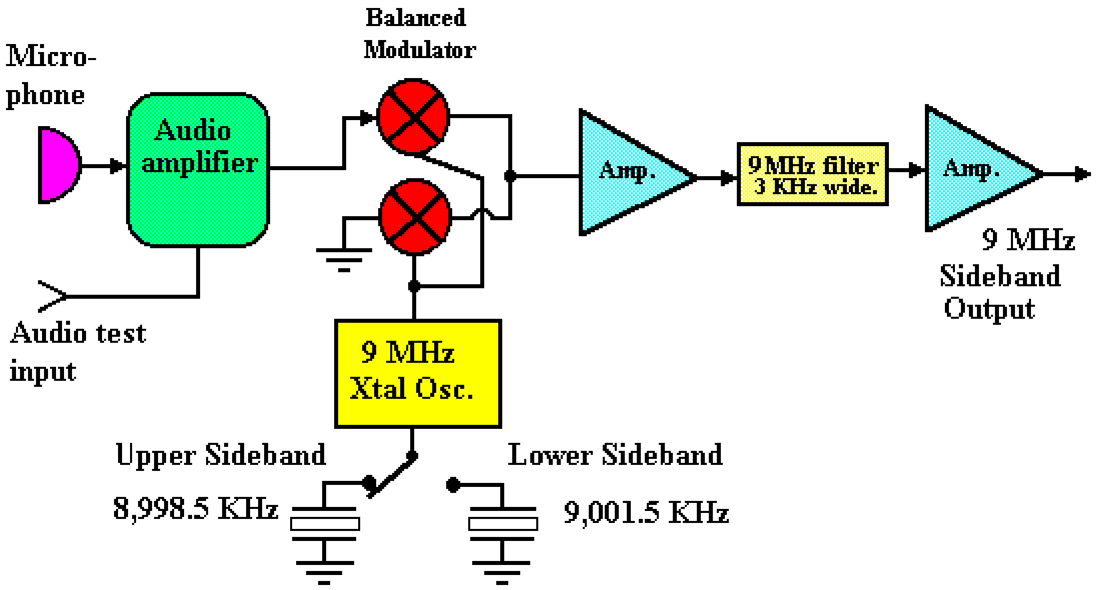
Micro-
phone
Audio amplifier
Balanced Modulator
Amp.
9MHz filter 3 KHz wide.
Amp.
9 MHz Sideband Output
Audio test input
9 MHz Xtal Osc.
Upper Sideband
8,998.5 KHz
Lower Sideband
9,001.5 KHz

There are different ways of generating an RF sideband phone signal, but the most straight-forward one I've seen is outlined above. The block diagram shows the five circuit blocks needed to generate a sideband signal on 9.000 MHz. This generator is similar to one found in the 1986 ARRL handbook. After the 9 MHz SSB signal has been generated, it must be moved to the desired ham band using a mixer and a high frequency VFO of the correct frequency range.

The circuits you've used in previous chapters are the audio amplifier, the crystal filter, the RF oscillator/ amplifier, and the conversion modules to move the VFO signal to ham bands. The audio amplifier design is similar to the one in the homebrew receiver in chapter 13. The 9 MHz RF oscillator/ amplifier uses the same technology used in the QRP described in chapter 6. In theory, the VFO could be the VFO signal from your receiver. When I started this project, I figured if the sideband generator didn't work, I would at least have a CW signal that was slaved to my receiver so that it would be easier to zero-beat my signal with the guy I was trying to talk to. Unfortunately, that goal turned out to be harder than it looked.

Don't burn your bridges

If you're thinking about modifying a working CW transmitter to sideband, I don't recommend it. If you already have a working QRP driver based on chapters 6 or 11 from this book, those designs are full of tuned amplifiers and mixers. Tuned amplifiers tend to self-oscillate when used for sideband. To have a good chance of working, every gain stage should be converted to broadband. If you convert your old transmitter, you are likely to have months of struggle in which you aren't on the air at all. Start from scratch! Don't ruin a rig that works!

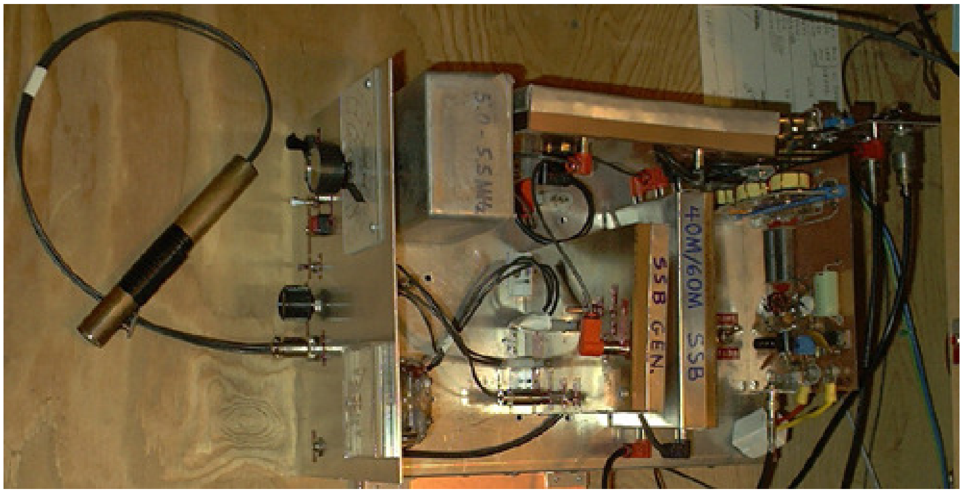

Homebuilt SSB transmitter

How sideband really works

Ordinary broadcast band AM modulation transmits three separate signals. These are the carrier signal and two sidebands of speech modulation. Single sideband begins with AM, but a cancellation process removes the carrier leaving a double sideband signal (DSB). Next one of the two sidebands is filtered out with a crystal filter. Let's begin with the crystal oscillator:

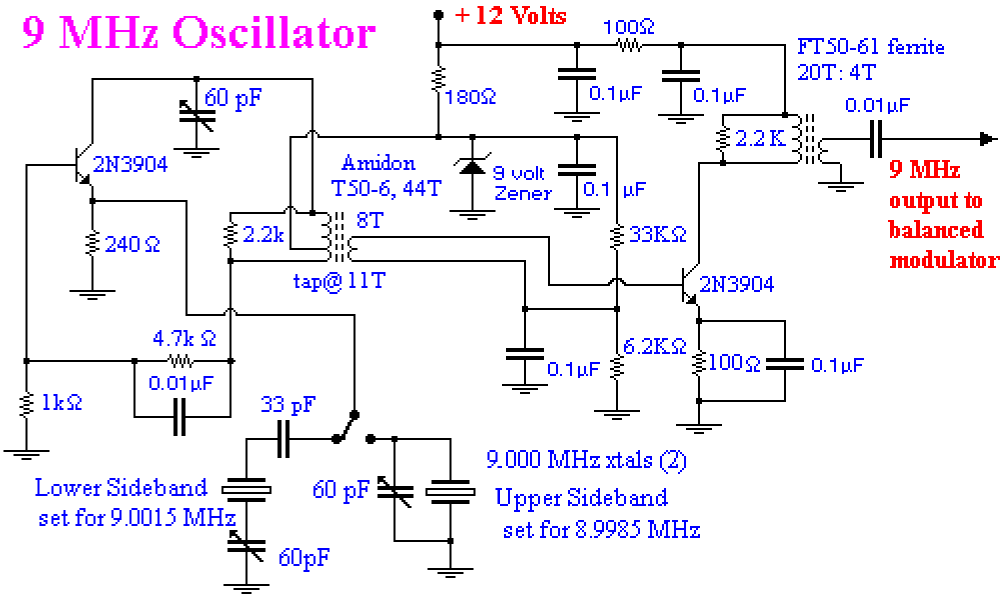
9 MHz Oscillator
+12 Volts
100Ω
FT50-61 ferrite
20T: 4T
60 pF
180Ω
0.1µF
0.1µF
0.01µF
2N3904
Amidon
T50-6, 44T
9 volt
Zener
0.1 µF
2.2 K
9 MHz
output to
balanced
modulator
240 Ω
2.2k
8T
33KΩ
tap@ 11T
2N3904
4.7k Ω
0.01µF
0.1µF
6.2KΩ
100Ω
0.1µF
1kΩ
33 pF
9.000 MHz xtals (2)
Lower Sideband
set for 9.0015 MHz
60 pF
Upper Sideband
set for 8.9985 MHz
60pF

9 MHz sinewave oscillator / amplifier

A stable, fixed frequency RF sinewave signal is generated by a crystal-controlled 9 MHz oscillator and amplifier that resemble the 7 MHz QRP transmitter described in chapter 6. The crystal oscillator has two crystals. Each crystal has tuning capacitors so that the frequency can be pulled about 1.5 KHz up and down. This allows the two AM sidebands to be lined up properly with the ladder-style crystal sideband filter that follows the balanced modulator. The filter shears off the unwanted upper or lower sideband.

A switch enables the oscillator to select two crystal/ capacitor pairs so that the operator can switch between upper and lower sideband. The crystal filter that removes the unwanted sideband is at 9.000 MHz, almost exactly. Notice that the upper sideband is generated by a sinewave 1.5 KHz below 9.000 MHz. The lower sideband is generated by a sinewave 1.5 KHz above 9.000 MHz. To pull the crystal above 9 MHz, the tuning capacitor is in series with the crystal. To push the crystal below 9 MHz, the tuning capacitor is in parallel with the crystal.

Just like the receiver project in chapter 13, you will need a bunch of inexpensive 9 MHz microprocessor crystals. How much capacitance is needed in parallel or series with a crystal depends on the individual crystal. The low side oscillator is the difficult one. Start by selecting the lowest natural oscillation frequency among your collection of crystals. For some crystals 8.9985 MHz might be reached with the capacitor in series with the crystal. For other crystals the parallel capacitance method is necessary and you may even have to pad the trimmer with an additional fixed capacitor. You may find that none of your crystals will go that low without becoming unstable and losing control of the frequency. You might even need to buy a custom crystal. As always, the oscillator collector circuit LC must be tuned to the region where the frequency "locks in" to the crystal.

The audio amplifier

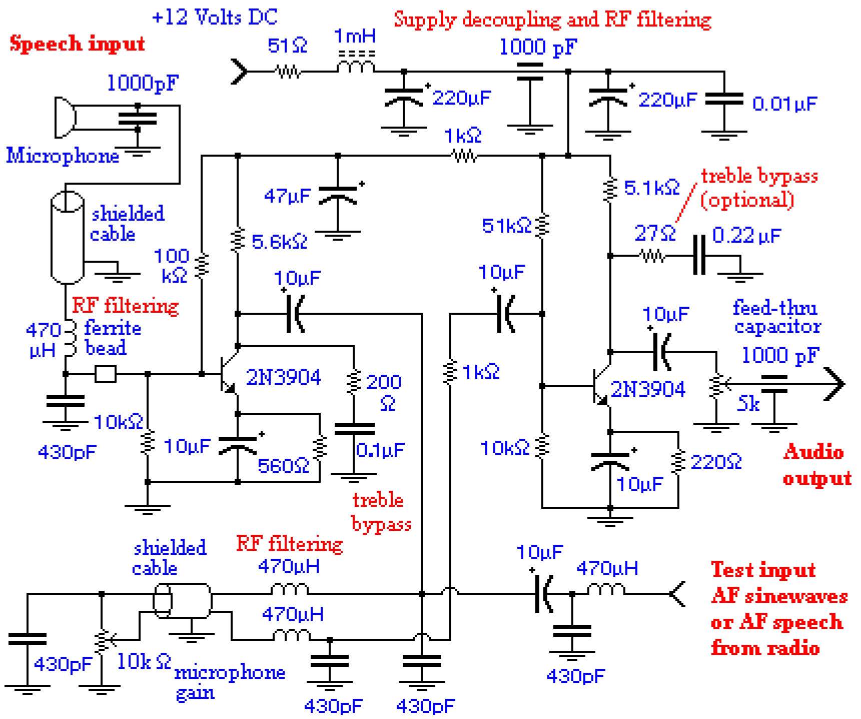
Speech input
+12 Volts DC
51Ω
1mH
Supply decoupling and RF filtering
1000 pF
1000pF
220µF
220µF
0.01µF
Microphone
1kΩ
47µF
5.1kΩ
treble bypass (optional)
shielded cable
5.6kΩ
51kΩ
27Ω
0.22µF
100 kΩ
10µF
10µF
RF filtering
470 µH
ferrite bead
10µF
feed-thru capacitor
1000 pF
2N3904
200 Ω
1kΩ
2N3904
5k
10kΩ
430pF
10µF
560Ω
0.1µF
10kΩ
Audio output
220Ω
10µF
treble bypass
shielded cable
RF filtering
470µH
10µF
470µH
Test input
AF sinewaves
or AF speech
from radio
470µH
430pF
10k Ω
microphone gain
430pF
430pF
430pF

An audio amplifier with test input

The microphone needs a high gain audio frequency (AF) amplifier before it can drive the balanced modulator. The audio amplifier is pretty routine except that heroic effort is needed to shield it from RF. Notice the RF chokes and bypass capacitors on the two audio inputs, the audio gain pot and the 12 volt power input. Because crystal mikes have a puny output, it took me two stages to get the signal up to roughly 5 volts peak. My crystal microphone exaggerated the high frequencies, so I attenuated the high frequencies with series RC ***treble filters*** on the collectors of both amplifier stages. You may be tempted to add another stage of audio gain. Don't! It's much better to run the audio gain wide open than to have extra gain and keep the gain turned low. Surplus gain just invites noise and sensitivity to RF feedback. If you like, you can replace most of this circuit with an IC, but as always, building your own amplifier with discrete parts will be more educational.

After you get the amplifier working, look at the audio waveform critically. You may find that the negative and positive voice peaks are not symmetrical. If this is the case, you may have too much or too little forward bias on the second stage. That is, you may want to increase or decrease the resistance of the 51 K ohm resistor biasing the second 2N3904. Ideally, the amplifier should be able to deliver a 10 volt peak-to-peak symmetrical signal.

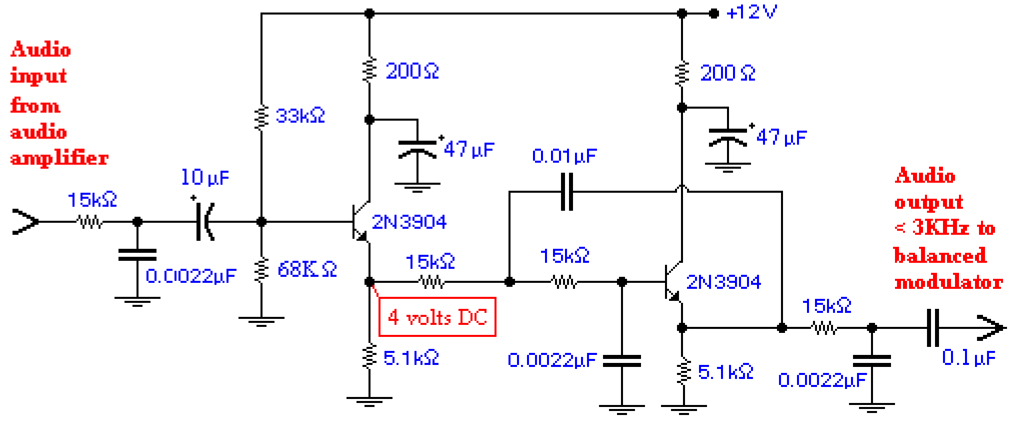

AUDIO LOWPASS FILTER (< 3 KHz)

I added the ***Butterworth*** filter shown above to be sure the bandwidth of my final signal would be less than 3 KHz. Like the treble filters discussed earlier, your generator may not need this. The Butterworth sharply cuts off practically all audio signals higher than 3 KHz. In contrast, the treble filters will just emphasize the lower frequencies. The filter uses two transistor amplifiers wired in the ***emitter follower*** configuration. Notice that the load resistor (5.1K Ω) for each transistor is wired between the emitter and ground, rather than between the collector and the positive supply.

This audio filter is probably unnecessary because the crystal filter trims the bandwidth down to 3 KHz anyway. So, you might try the SSB generator first without the filter, then add it later if necessary.

The advantages of emitter followers

Emitter followers have the advantages that the input impedance is extremely high and the output impedance is very low. High input impedance means they will not load down or affect the input signal strength. Low output impedance means they deliver big currents into low resistance loads. Another feature of the emitter follower is that ***the voltage gain is less than unity.*** That is, they don't amplify voltage. This is an advantage here because it insures that the amplifier will not oscillate. Butterworth filters are usually implemented with operational amplifiers. Until this filter, I had never built a Butterworth with transistors. Yes, simple transistors work too.

No matter what audio amplifier circuit you use, it will be sensitive to RF interference from any strong RF signal in your radio shack. For example, if you are using a simple antenna coupler with no shielding like mine, those RF signals will tend to feedback into your microphone cable. To prevent this I added RF chokes, bypass capacitors, and a ferrite bead in series with the microphone input. Since my microphone gain pot is remote from the audio module, the shielded wires to the pot are also filtered with RF chokes and bypass capacitors. Even the output from the amplifier passes through a capacitive feed-through (bypass) capacitor on its way to the balanced modulator.

The 0.1 µF audio output capacitor

Notice that the output from the filter circuit is not the usual big 10 µF capacitor but is only a 0.1 µF cap. This cap goes to the audio input of the balanced modulator. If you use a larger capacitor here, the capacitor will take time to charge when the transmitter is first keyed and power is applied to the audio amplifier. A big cap would take two seconds or so to charge and would cause the balanced modulator to turn on causing a brief, shrieking whistle to go out over the air. The input resistance to the balanced modulator is very high, about 100,000 ohms. Therefore, the time constant of 100K ohms times 0.1 µF is about 0.01 second. This allows for 100 Hz audio and is plenty Hi-Fi for SSB ham work.

Decoupling for the power supply lead

The 12 volt power supply lead for the audio amplifier also has a large RF choke (one millihenry) in series with the lead and passes through another feed-through capacitor bypass. In addition, the power supply lead is isolated or ***decoupled*** by means of the 51 ohm resistor and the large 220 microfarad bypass capacitors. The purpose of these capacitors is to insure that the voltage supplied to the amplifier cannot change as fast as the audio signals. All of the modules in an SSB transmitter, except the final amplifier, need to be decoupled from changes in the 12 volt supply level. Otherwise, as you talk into the microphone, the current drawn by the high power final varies rapidly and the voltage delivered to each module will rise and fall in time with the speech. Because the voltage is rising and falling, the RF output from each module will rise and fall too. This feedback produces surges in the radio signal that sound like noise superimposed on the speech. In fact, it makes nearly the same buzzing roar as RF interference.

The final amplifier draws too much current for it to be practical to decouple the 50 watt linear amplifier supply lead. In fact, it is the big 10-ampere surges of current drawn by the final that cause the noise in the rest of the transmitter. In general, the less current drawn by a circuit block, the more extreme the decoupling must be. For example, the audio amplifier has a series 51 ohm resistor and 440 microfarads bypass. In contrast the 5 watt RF driver stage has only a one ohm resistor and a 0.1 microfarad capacitor.

Microphones are important

Not all microphones are equal. I tried three different crystal microphones. Two small Radio Shack microphone cartridges gave a “tinny” sound. I was able to compensate with RC bypass networks on the collectors of all three transistors to limit the high frequency components (treble) of the speech. For example, notice the 200 ohm resistor and 0.1 µfd capacitor combinations going to ground from the first two transistor collectors. I also tried a 40 year old Hallicrafters crystal microphone designed for mobile radio. It worked fine without the RC bypasses. Next I tried two tiny condensor-type electret microphones. One was too “bassy” and made a low frequency hum. The other, a Radio Shack PN # 270-092A, worked perfectly. To bias the electret with about 4 volts, I used a 3.9K resistor in series with a 7.5K resistor to step down the 12 volt supply.

When I finally had a working SSB generator, I was able to use either my ancient Hallicrafters mike or a homebrew mike built around the Radio Shack electret. To shield the electret I mounted it in a length of 3/4 inch copper pipe. I soldered copper disks (PC board) to the top and bottom of the pipe section so that the microphone would be well shielded. The electret is force-fit into a snug hole at the top of the pipe. I put a 1000 pF capacitor across the microphone to further reduce RF interference. The electret is connected to the transmitter by two short pieces of RG-174 coax: one for the audio signal and the other coax for the 4 volt DC supply line for the electret. Keep the microphone cord(s) as short as practical. A long cord invites RF interference.

The microphone case/ copper pipe also contains a push-to-talk SPST button switch. This switch turns on the transmitter just like the switch mounted on my electronic bug. The switch lead-in has its own separate piece of RG-174 coax. So as you can see, my microphone cord is actually three parallel RG-174 coax cables. Obviously my mike cord should be a single, shielded three-conductor cable. But since I didn't have such a thing, I used the three separate single-conductor shielded coax cables. To connect the mike to the transmitter, I used a (fairly) standard microphone connector that has 4 inner conductors plus the outer shield ring. I found this connector pair at Radio Shack and by some miracle it was the same connector used on my old Hallicrafters crystal mike.

The balanced modulator

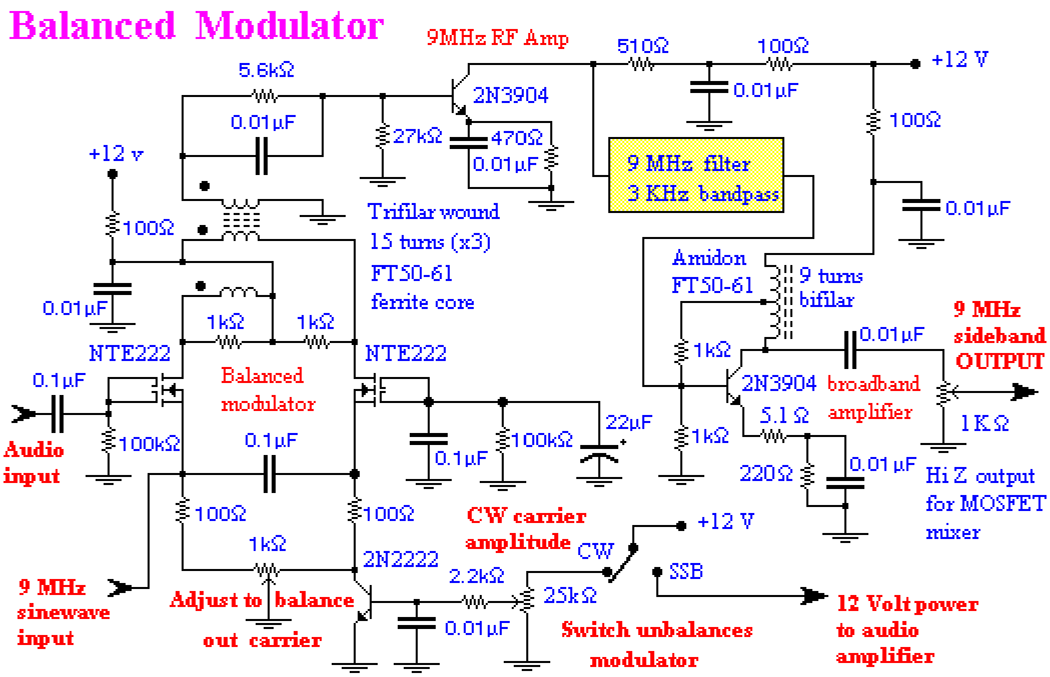
Balanced Modulator
9MHz RF Amp
510Ω
100Ω
+12 V
5.6kΩ
2N3904
0.01µF
0.01µF
100Ω
27kΩ
470Ω
0.01µF
+12 v
9 MHz filter
3 KHz bandpass
0.01µF
100Ω
Trifilar wound
15 turns (x3)
FT50-61
ferrite core
Amidon
FT50-61
9 turns
bifilar
0.01µF
1kΩ
1kΩ
NTE222
NTE222
9 MHz
sideband
OUTPUT
1kΩ
0.01µF
0.1µF
Balanced
modulator
2N3904
broadband
amplifier
5.1 Ω
1 K Ω
Audio
input
100kΩ
0.1µF
0.1µF
100kΩ
22µF
1kΩ
220Ω
0.01 µF
Hi Z output
for MOSFET
mixer
100Ω
100Ω
CW carrier
amplitude
+12 V
CW
1kΩ
2N2222
SSB
9 MHz
sinewave
input
Adjust to balance
out carrier
2.2kΩ
25kΩ
0.01µF
Switch unbalances
modulator
12 Volt power
to audio
amplifier

The balanced modulator is the "carrier cancellation circuit." It is a kind of dual mixer in which an audio signal is mixed with the 9 MHz sinewave to produce an AM modulated RF signal, exactly like AM radio. An AM signal has a carrier signal just like the CW input plus the two RF sidebands caused by the audio modulation. What's different about a balanced modulator is that it consists of two mixers in parallel. The second mixer has no audio input so its output is simply another CW signal, just like its RF input. The two mixers share a common output transformer that has three windings - two primaries and one output winding. There is a primary for each mixer.

The clever part happens when the primaries generate magnetic signals in the transformer iron. The windings are oriented so that the two primaries work against each other. The CW signals in both windings are "balanced" with an adjustment pot so that they exactly cancel. This means that the only signals that appear in the secondary winding are the two sideband signals. ***In***

summary, a balanced modulator produces a double sideband signal with no carrier.

The transistors are dual-gate MOSFETS with the gates shorted together. The idea is to use transistors without any P-N diode junctions. According to the handbook, P-N junctions act like varactors and distort the speech slightly. I used dual gate MOSFETs simply because they were the only small RF MOSFET available. Single gate small RF MOSFETs are fine, but you may not be able to find any. I haven't tried JFETs, but they might work too. The above circuit was adapted from the 1986 Handbook.

To use this modulator for CW, there is a CW/ SSB switch that unbalances the modulator and allows some carrier to pass through to the filter. Notice that when the switch is in the SSB position 12 volt power is fed to the microphone audio amplifier, thereby turning it on. When the switch is in the CW mode, pure CW carrier is sent to the filter. Unfortunately, this CW signal will be hard to use for actual CW contacts because the crystal filter will tend to remove most of the carrier again. However, this small 9 MHz sinewave is useful for tuning an antenna with a "test signal" or for "spotting" the VFO on top of the station you wish to call. For real CW operation, you could bypass the SSB filter with another switch or you could build a separate 9 MHz oscillator.

This sideband generator can also be modified to generate amplitude modulation. This is discussed in chapter 16. You can also go on the air with double sideband, DSB. Many homebrew sideband builders take this shortcut. It will sound like sideband, but the signal will be twice as wide as single sideband.

The sideband filter

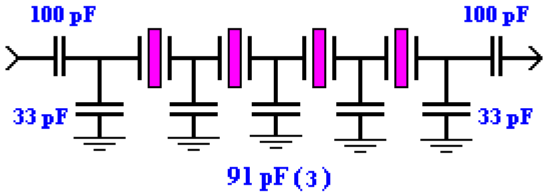

Four matched 9.000 MHz crystals

You can buy sideband filters that select a 3 KHz passband, typically 9,000 KHz to 9,003 KHz. Sometimes matched oscillator crystals are also available that will position the RF signal optimally to line up with a particular filter.

In chapter 13 we made a 4-crystal CW receiver crystal ladder filter that was quite similar. The difference is the sizes of the accompanying shunt capacitors. In the receiver the shunt capacitors were 220 pF. However, ***The smaller the shunt capacitors, the wider the pass band of the filter.*** The 91 pF capacitor value in the above filter was scaled from a sideband filter used in a sideband transmitter designed around an 8.000 MHz SSB generator that used 100 pF caps. This filter seems to work, so I haven’t had to experiment.

The homebrew way to build the filter is to buy a bunch of 9.000 MHz microprocessor crystals from Mouser or Digi-Key for less than a dollar each. Using the RF oscillator shown earlier and a frequency counter, measure the frequencies of each of your crystals. When used as filters, their natural frequencies may not be the same as in your oscillator, but their *RELATIVE* frequencies will be comparable. Pick four crystals that are as closely matched as you can. Matching within 100 Hz should be adequate. I tried tuning each crystal to the same frequency using trimmer capacitors in parallel with the crystals. This proved to be unnecessary and I later removed the trimmers. This application is not nearly as critical as crystal filters for receiving CW in a receiver.

Using a signal generator as a test signal, these filters seem to peak very close to their nominal frequencies. For example, a crystal might oscillate at 9.0015 MHz in your test oscillator, but the filter will peak quite close to 9.000 MHz. So far I have built three of these filters and each one worked well and was centered more or less on 9.000 MHz.

The crystal oscillators are misaligned for CW

The crystal filter is set up to cut off either the upper or lower sideband. When you unbalance the balanced modulator to produce a CW signal, you will indeed get a small sinewave signal through the filter that may be strong enough to tune the antenna coupler or adjust the transmitter. However, because the center of the SSB crystal filter (9.0000 MHz) is lined up 1.5 KHz away from both crystal oscillators (8.9985 MHz and 9.0015 MHz), this CW signal will be far weaker than the maximum power you would get on voice peaks in SSB.

When I built a comparable filter for 8.000 MHz, the filter was centered on 7.995 MHz, 5 KHz low. I don't understand why this filter was different, because on an oscillator, the crystal frequencies were all well above 8.000 MHz. Because the filter frequency was so low, the upper sideband 8 MHz oscillator had to run on 7.9935 MHz. To drive the oscillator that low, I had to put a huge 150 pF cap across the 8.000 MHz crystal. I was amazed that the frequency locked in. Better engineering practice would dictate a custom crystal. A cheaper solution would be to start off with a different collection of cheap microprocessor crystals that center properly on 8.000 MHz.

MECHANICAL CONSTRUCTION

Sorry! You must shield sideband.

When building CW transmitters some of us think it's cool to have the pretty little parts out in the open where we can admire them. Unfortunately a huge problem with SSB is RF feedback. To help prevent that, you must enclose all the modules of your SSB transmitter in metal. All the connections between modules should also be shielded cables. I started out using shielded power cables to my RF modules, but eventually I worked out a way to mount the Molex connectors directly to the shielded boxes. In this way I could plug the module onto the chassis with no long exposed wires.

Actually, I still leave the linear amplifier out in the breeze, but all the low power modules and the power supplies are well shielded. If I were building a new linear, I would design in a shield from the start. All my modules consist of a two-sided PC board with walls of PC board soldered onto the edges to make a box. Then I fold an aluminum cover over the top to provide the lid.

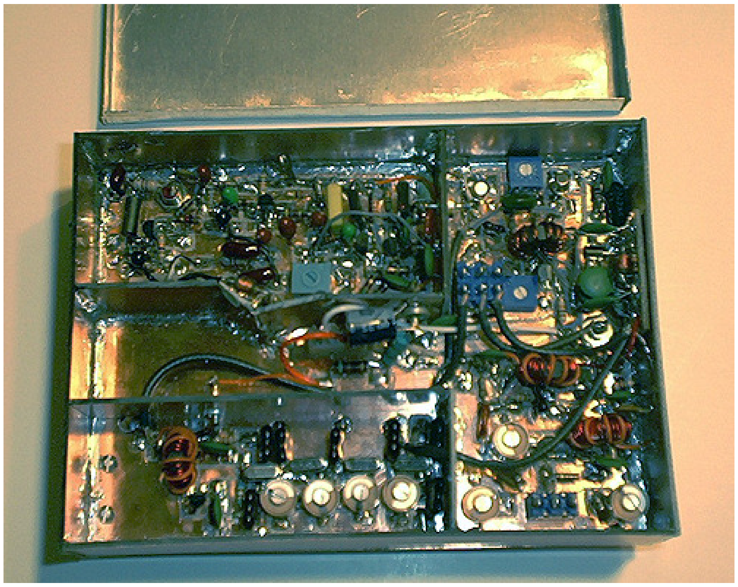

The completed 9MHz SSB generator. An aluminum lid fits over the top of the box.

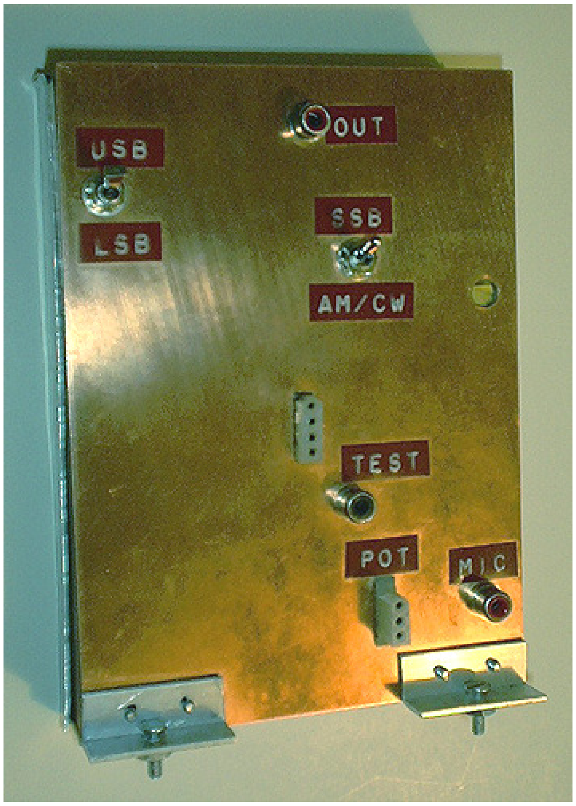
OUT
USB
LSB
SSB
AM/CW
TEST
POT
MIC

The 9 MHz SSB generator as seen from the control side

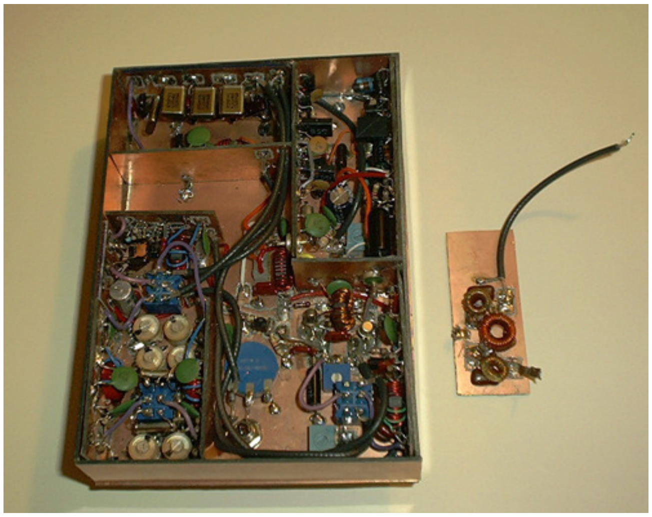

A dual frequency 9 MHz and 8 MHz SSB generator module is shown above

After learning that 9 MHz was impractical on 17 meters, I built a new SSB dual frequency generator. The 8 MHz crystals are in a row at the top left. The 9 MHz crystals are in a row just beneath them. The two frequency oscillators are at the bottom left. The balanced modulator is at the lower right. Notice that many of the components are mounted on inch high PC board strips and soldered in vertically. This allowed me to cram in about 50 percent more components than would otherwise be possible.

A major advantage of this technique is that circuits that don't work well can easily be removed and replaced. The loose little board on the right is a low pass filter that I took out when it appeared to be unnecessary. I removed it with little or no damage to the module. As you can see, some components, especially those directly wired to controls and connectors, are mounted on the "floor" of the box. If you wish to replace the circuitry on the "floor," you can unsolder all the old components and start over with a new board that makes a "patch" over the old circuit board laying out all the circuitry on the floor of the module box in the usual way.

Tuning and testing the sideband generator

The essential tools for tuning up your generator are ***a frequency counter, an oscilloscope, and a good ham receiver.*** Ideally, you need one of those same modern receivers that will be listening to your signal. The SSB generator above can generate a sideband signal on 9 MHz. Keep in mind that when you listen to the 9 MHz signal on your ham receiver, unless your generator is well shielded, you'll still hear the carrier signal and the suppressed sideband leaking from your 9 MHz oscillator. That's because those signal components are present on your circuit board and your receiver will have little difficulty in hearing them.

To check out the generator introduce audio from a Walkman radio into the test input allowing you to align the generator. Tune the Walkman to a talk radio station and inject the audio into the test input. Speech should not only be understandable in your ham receiver, the fidelity should be good enough that you can easily recognize the person's voice. When you turn off the BFO and set up the receiver for AM modulation, speech should become unintelligible. Music should always sound awful. If music sounds pleasant, your bandwidth is too high.

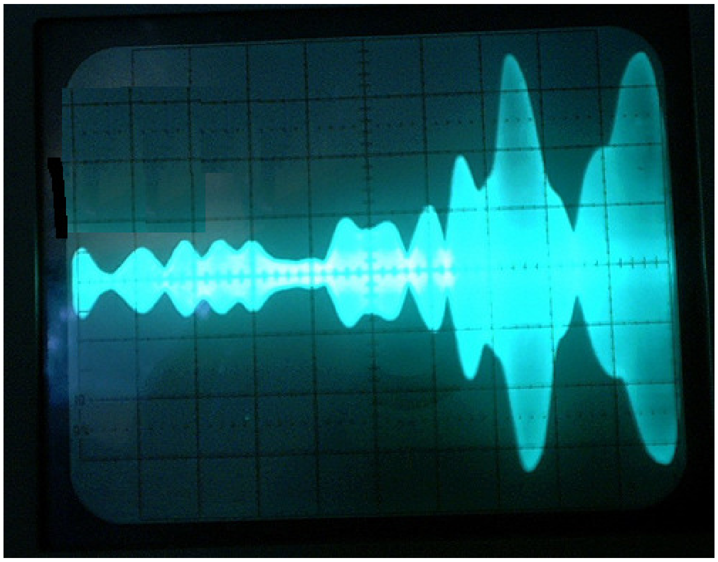

The 9 MHz SSB voice signal seen on an oscilloscope.

An SSB voice RF signal should appear on your scope as shown above. The audio modulation is symmetrical about the zero axis. In between syllables or words, the signal strength drops to near zero. The edges of the sinewave bursts should be reasonably sharp, meaning that the frequency should remain pure with varying levels of speech. When there is no audio input, there should be essentially no RF output.

Checking out the microphone and microphone pre-amplifier can be complicated. It's hard to listen to your own voice critically. Also, the audio from the ham receiver loudspeaker will feedback into your microphone. My solution was to put the microphone up against a high fidelity Walkman headset. Then I wrapped the headset in cloth to muffle the sound. To hear how it sounded on the air, I listened to the sideband generator signal in the ham receiver using headphones. Unfortunately, when used with the 50 watt amplifier, the RF from the dummy load and transmatch interfered with the Walkman so this technique only worked well for 5 watts. I was able to partly test the 50 watt linear by listening to my own voice while wearing headphones with no receiver antenna plugged in. I could at least confirm that there was no RF feedback.

Audio signal generator testing

It is instructive to feed an audio tone from an audio oscillator into the audio input test jack. This input is shown on the audio amplifier diagram. As you sweep the audio spectrum from 20 Hz to 3 KHz, watch the sideband generator RF output on the oscilloscope. Unlike AM modulation, there should be no audio frequency modulation visible on the radio signal. That is, for each audio sinewave frequency you should see a pure, CW-like signal. Another way to think about single sideband is that it is a kind of extremely narrow-band frequency modulation. As the audio frequency changes, the signal frequency shifts up and down in direct proportion. Unlike AM modulation, the amplitude of the transmitted signal shouldn't change when you introduce a constant amplitude audio frequency. That is, with SSB, you shouldn't see sinewaves impressed on the signal amplitude proportional to the frequency. The amplitude should only change with speech amplitude, not with speech frequency. In contrast, pure FM modulation does not change its amplitude with speech amplitude or with audio frequency.

Suppressing power supply RF feedback and low frequency coupling

My first sideband contact said, “Sorry, old man! I hear some hissing noises, but I can’t understand a word you’re saying.” It turned out that the power supply leads in the generator and other modules in the transmitter needed low frequency decoupling. In the sideband generator this consisted of the 51 ohm resistor and the two 220 microfarad caps on the 12 volt line. Without decoupling, the audio turns into noise as the generator competes with the final amplifier(s) for operating voltage. That is, the 12 volt supply voltage surges up and down with the speech and the amplifiers exaggerate this.

After these improvements my next contact could understand me, but he said my voice was “raspy with popping sounds.” I didn’t have laryngitis, so I asked Jack Quinn, KØHEH, about the criticism. He instantly diagnosed the problem: “That’s RF feedback. Improve the shielding of your microphone and audio amplifier.” I placed the 1000 pF capacitor directly across the microphone, the 430 pF bypass capacitors and the 470 microhenry inductors in series with the inputs and power line. Also, the power and audio output pass through feedthrough capacitors to further attenuate the RF. When RF feedback is really bad, the signal turns into a roar of noise that can sound similar to low frequency power supply decoupling problems.

****

The Hard Part --- Moving the SSB signal to a hamband

Is 9 MHz a hamband?

To get on the air you need to amplify the 9 MHz sideband signal up to 50 or more watts. Unfortunately, the last I heard, 9 MHz isn’t a hamband. Unfortunately, the hardest part of this project turns out to be moving the 9 MHz signal to the band(s) of your choice. Alternatively, we could all write to the WARC to ask them to establish a little 3 KHz hamband centered on 9.000 MHz. Maybe not.

Although moving the SSB to a hamband is the most difficult part of sideband, maybe it won’t be so bad if you don’t make the mistakes I did. Six principles I learned the hard way were:

*** Move your sideband signal only once.*** Double conversion might appear convenient, but it’s extremely hard to do without distortion. In other words, don’t do the hardest task twice.

*** In the conversion between 9 MHz and your HF band, make sure that the mixer input frequencies are far away from the final frequency.*** For example, to get on 20 meters, it is practical to add a 5 MHz VFO to 9 MHz to get 14 MHz. On the other hand I found that it was impractical to move a 4 MHz sideband signal to 21 MHz using a 25 MHz crystal oscillator. Every time I stopped talking, a significant 25 MHz signal went right out through the transmitter output filters.

*** Plan your VFO and sideband frequencies so they do not divide evenly into the desired ham band frequency***. For example, 2 times 9 MHz is 18 MHz. This fact makes using a 9 MHz SSB generator and/or a 9 MHz VFO to generate 17 meters extremely difficult. Notice that 6 MHz is also difficult because 3 times 6 MHz = 18 MHz.

Don’t use tuned amplifiers and mixers. When you stop talking, tuned amplifiers tend to oscillate by themselves at the frequencies to which they were tuned. Actually, getting rid of the noise and oscillations when you're *NOT* talking is harder than making the speech intelligible. Unlike CW, it is best to use broadband mixers and amplifiers and to put all your ham band filtering into two passive filter networks. Sideband is different from CW!

** **Beware of having too much gain in your SSB generator and frequency converter.*** I originally had unnecessary broadband amplifiers in the generator output and also right after the converter mixer. These extra amplifiers amplified noise. Every time I wasn't talking, they often began to self-oscillate.

**** It sometimes helps to connect all ground connections to the outside layer of your two-sided PC board.*** The ground connections for all high current RF stages must be extremely low inductance. Otherwise, if your board layout isn't well designed, RF voltages on all the ground traces inside the PC board box will "bounce up and down" with the currents in the power amplifier stages. This feedback introduces noise into the mixer stage and makes the QRP module difficult or impossible to adjust. If you are using 2-sided PC boards, solid grounds can be added by drilling the PC board at each ground connection and soldering wire shunts through the board to the unbroken sheet of grounded copper outside the PC board.

Getting on 20 and 80 meters

Hetrodyne converter for the SSB generator

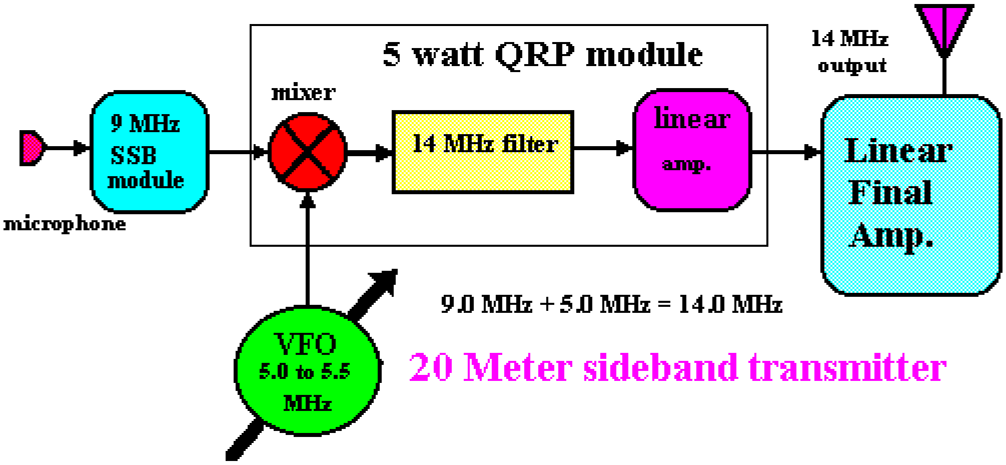

20 Meter sideband transmitter

When starting with a 9 MHz sideband signal, 20 meters is the easiest hamband to reach. For 20 meters, a 5.00 to 5.35 MHz VFO signal is mixed with 9 MHz to give 14.0 to 14.35 MHz. 9 MHz is 36% different from 14 MHz. Consequently, building a filter to extract the 20 meter component and suppress the 9 MHz signal is relatively easy.

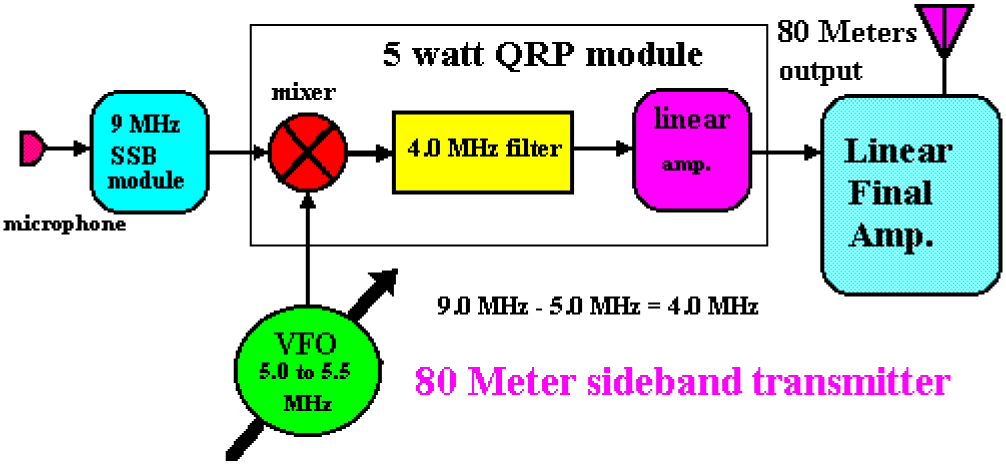

Now suppose that you wish to move the 9 MHz sideband signal down to 80 meters. 9 MHz minus 5 MHz is 4.0 MHz. The phone band extends right up to 4.0 MHz. So the 80 meter (75-meter) output signal can be as little as 20% different from the VFO signal. Filtering the 80-meter signal is almost twice as hard as 20 meters. What happens if your filtering is inadequate? Every time you stop talking, your linear final amplifier will be transmitting a sinewave carrier on your VFO frequency, 5 MHz. As we'll see, when you start with a 9 MHz sideband signal, all the other HF bands are harder than 20 meters.

Spice will make filter design bearable

I found my Spice program essential for simulating filters before I built them. I use a rather elderly program called Electronics Workbench, version 4. If I had had to calculate the component values or build all the filters by trial and error using real components, this project would have completely stalled. Using the program I could enter a possible filter design with trial values. When I ran the simulation, it plotted the frequency response instantly. It would inevitably be way off from what I had intended. But then I could quickly plug in dozens of values until I got the response I wanted. I get discouraged just thinking about building and testing those filters the old way and plotting the results at each frequency with real components.

Below is a copy of a page from my notebook showing one of the most difficult filters I built. It was a low impedance band pass for 17 meters. After it looked right on the computer, I built a real one and it tuned up immediately using the variable cap in the center. The simulations are amazingly trustworthy. I found that when the simulation predicted small capacitor values, it was wise to use a variable capacitor in that location. If the capacitance or inductance came out extremely small, for example, a few picofarads or hundredths of a microhenry, I changed the design until all the components had significant values. Once I had a design on paper, I built the real filter and tested it using a frequency generator and scope to confirm that it really behaved as simulated.

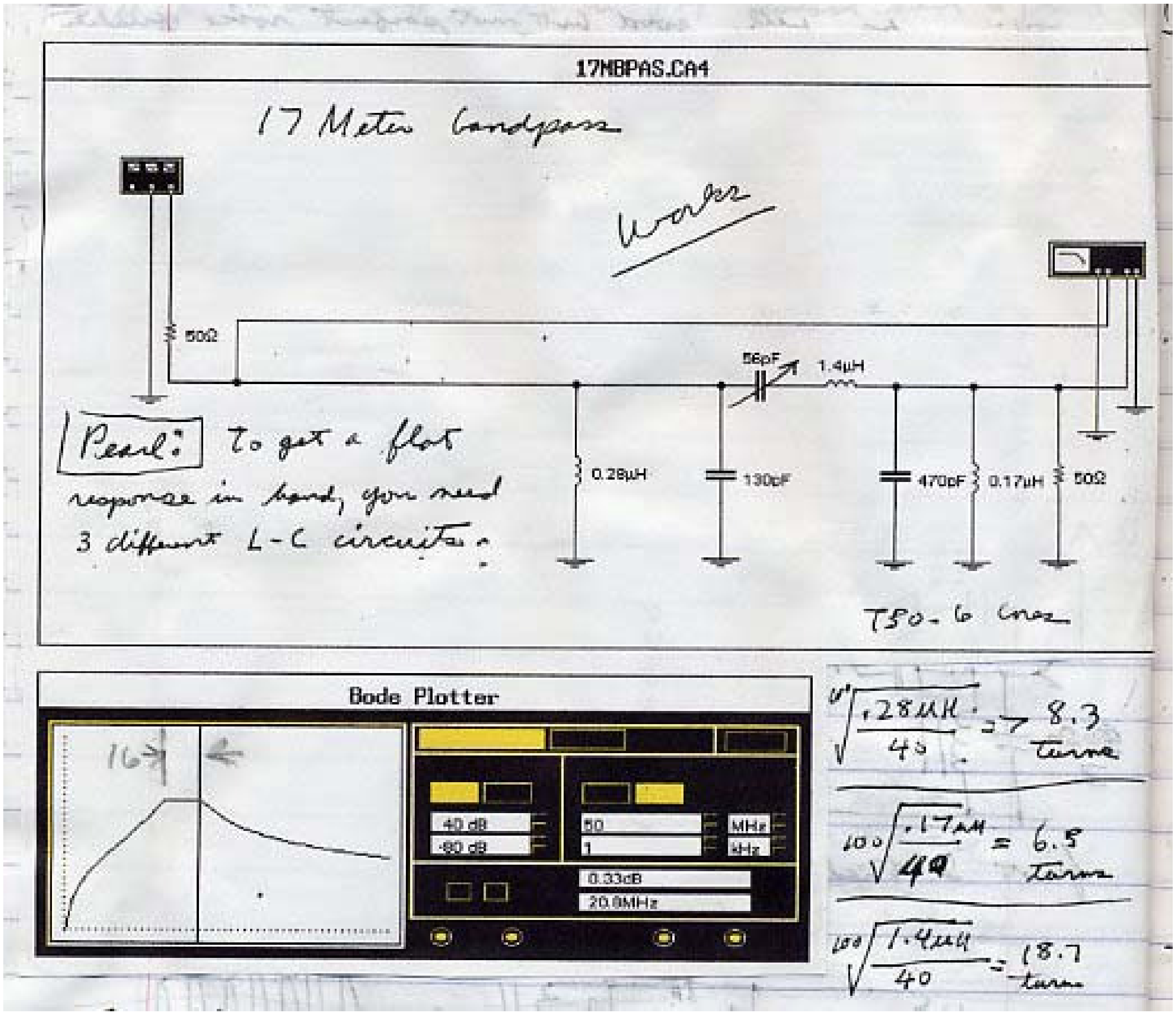
17MBPAS.CA4
17 Meter bandpass
works
50Ω
56pF
1.4µH
0.28µH
130pF
470pF
0.17µH
50Ω
Pearl: To get a flat response in band, you need 3 different L-C circuits.
T50-6 Cores
Bode Plotter
16
40 dB
-80 dB
50
1
MHz
kHz
0.33dB
20.8MHz
100 √(.28µH / 40) => 8.3 turns
100 √(.17µH / 49) = 6.5 turns
100 √(1.4µH / 40) = 18.7 turns

Self-oscillation

When you build a high-Q amplifier-filter stage, it tends to oscillate on its own whenever there is no signal coming into the input. This means that, in between words, your QRP module may be oscillating on some random frequency on or near the ham band you are using. Sometimes this oscillation can be suppressed by placing a 50 or 100 ohm resistor across the input of the offending stage. Another method is to place a 1K to 2K resistor across the RF transformer primary on the collector or you may add a small resistance (such as 10 or 20 ohms) of un-bypassed resistor in series with the emitter of the transistor. Unfortunately, these tricks are usually not enough. ***The best solution is to use untuned broadband amplifiers!*** Even with broadband amplifiers, you will still have to use some or most of these tricks to keep them from oscillating.

No wonder most rigs are transceivers

There is a great deal of similarity between a sideband receiver and a sideband transmitter. Once you've built a receiver, it dawns on you that the transmitter has most of the same modules and that you are building the same circuits twice. On the other hand, using the same circuit modules for both tasks takes finesse. Unfortunately we homebrewers have enough problems without that extra complexity.

Ideally, it would be best to use the 5.0 MHz VFO from your receiver. That way the transmitter frequency and receiver frequency can zero beat exactly. When you answer a CQ, you don’t want to take the time to tune the transmitter VFO. When I got on the air with my separate 5 MHz VFO, I found that, by the time I had it precisely zero beat with the guy calling CQ, he was often already talking to somebody else.

Unfortunately, using the receiver VFO isn’t simple. If you simply connect it to the transmitter by a long cable, the receiver will suddenly acquire intermodulation, noise, and whistles. To get past this, the VFO signal must be isolated from the receiver by an isolation amplifier. Also, the 9.000 MHz BFO and the sideband generator oscillator must be on the exact same frequencies. It really should be the same 9 MHz oscillator. Furthermore, the VFO band conversion oscillators in the receiver and transmitter must be aligned to within a Hz or two. Hmmm … this isn't so easy after all. The transceivers solve the problem by using the same oscillators for all those tasks so alignment isn't a problem. I believe a homebrewer can enjoy all the convenience of a transceiver, but only if we design it as a transceiver from scratch. It seems to be impractical to merge an SSB transmitter with an existing receiver.

An 80 meter sideband QRP driver

As explained above, an untuned mixer should be followed by a ***totally passive filter***. That is, the filter should be just a network of LC circuits. It should have no transistors. Instead, the gain is provided by two or ***three untuned stages in series***. Three high-gain broadband amplifiers in series can work without oscillating, provided that their input has very little noise. The basic design shown below has the advantage that, it can be used on any HF band. To change bands, you plug in different filter sections shown in the green boxes below.

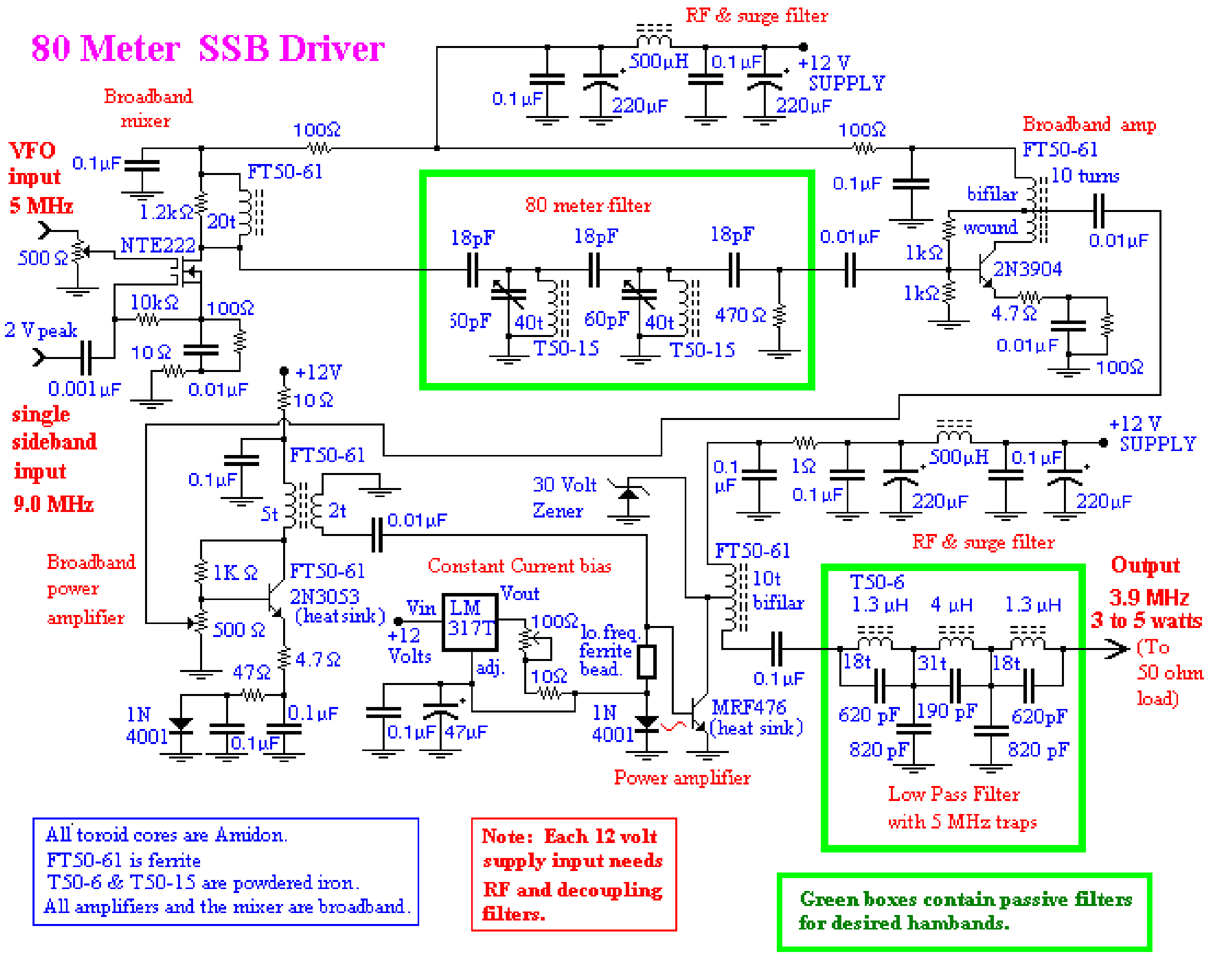
80 Meter SSB Driver
RF & surge filter
500µH
0.1 µF
+12 V SUPPLY
0.1 µF
220µF
220µF
Broadband mixer
100Ω
VFO input 5 MHz
0.1µF
1.2kΩ
FT50-61
20t
500 Ω
NTE222
2 V peak
10kΩ
100Ω
10 Ω
0.001µF
0.01µF
single sideband input 9.0 MHz
+12V
10 Ω
80 meter filter
18pF
18pF
18pF
50pF
40t
60pF
40t
470 Ω
T50-15
T50-15
100Ω
0.1µF
0.01µF
1kΩ
1kΩ
Broadband amp
FT50-61
10 turns
bifilar
wound
0.01µF
2N3904
4.7 Ω
0.01µF
100Ω
FT50-61
0.1µF
5t
2t
0.01µF
30 Volt Zener
+12 V SUPPLY
500µH
0.1 µF
1Ω
0.1 µF
0.1 µF
220µF
220µF
RF & surge filter
Broadband power amplifier
1K Ω
FT50-61
2N3053 (heat sink)
500 Ω
4.7Ω
47Ω
1N 4001
0.1µF
0.1µF
Constant Current bias
Vin
LM 317T
Vout
+12 Volts
adj.
100Ω
lo. freq. ferrite bead.
10Ω
0.1µF
47µF
1N 4001
MRF476 (heat sink)
Power amplifier
FT50-61
10t bifilar
0.1 µF
T50-6
1.3 µH
4 µH
1.3 µH
18t
31t
18t
620 pF
190 pF
620pF
820 pF
820 pF
Low Pass Filter with 5 MHz traps
Output 3.9 MHz 3 to 5 watts
(To 50 ohm load)

All toroid cores are Amidon.
FT50-61 is ferrite
T50-6 & T50-15 are powdered iron.
All amplifiers and the mixer are broadband.

Note: Each 12 volt supply input needs RF and decoupling filters.

Green boxes contain passive filters for desired hambands.

All transistor stages are broadband, including the mixer. Note the 1.2K resistor across the primary of the mixer ferrite core transformer. Without this or other feedback device, the broadband amplifiers tend to generate an uncontrolled signal whenever the SSB input drops to zero between spoken words. The un-bypassed 4.7 and 10 ohm emitter resistors in three of the amplifiers also help prevent oscillations. To reduce surging and RF feedback to the driver module I filtered my 12 volt input leads with RF chokes, big capacitors, and small ceramic capacitors.

Unlike the CW QRP modules, the SSB module output stage needs forward bias to operate in linear mode. The LM317 circuit current source supplies the needed current. This temperature compensated circuit was used in the 50 watt linear in chapter 12 and is overkill for this 5 watt final, but what the heck! With experimentation I found that just 20 milliamperes of forward bias is plenty to operate linear and give good speech quality. In theory, a 560 ohm, one watt resistor can provide this bias much more cheaply. However, I haven't tried this.

A passive high impedance (500 ohm) 80 meter filter is placed between the mixer and the first broadband amplifier. The output stage filter is the usual 50 ohm Chebyshev low pass except that capacitors have been placed in parallel with the inductors. These make the inductors resonate at 5 MHz and provide extra attenuation to get rid of the 5 MHz VFO signal. Notice that the equivalent of five parallel and series LC circuits were used to clean up the 80 meter signal. In contrast, as you'll see on the next page, a 20 meter QRP (5 MHz + 9 MHz) can be done with just two LC circuits and an output high pass filter.

Beware of residual VFO signal

Once I had the 80 meter driver (shown above) working well, I fed it into my linear amplifier described in Chapter 12. With the amplifier it produced 60 to 100 watts on voice peaks. However, whenever I stopped talking, the frequency counter probe on my antenna shifted to the 5 MHz VFO frequency, even though the amplitude of the signal on the scope screen looked negligible. When I turned up the scope amplitude, sure enough, there was a 5 volts peak sinewave on the output. That represented about 0.25 watts of 5 MHz sinewave. To get rid of it, I rebuilt the final (high power) amplifier 80 meter low pass output filter using another “elliptical” filter design. Once again each inductor has a parallel capacitance that resonates at 5 MHz and keeps the 5 MHz out of the antenna. I used the values shown below. After that change the residual, no-speech signal was only 1 volt peak and the counter measured it as the correct output frequency, 3.9 MHz.

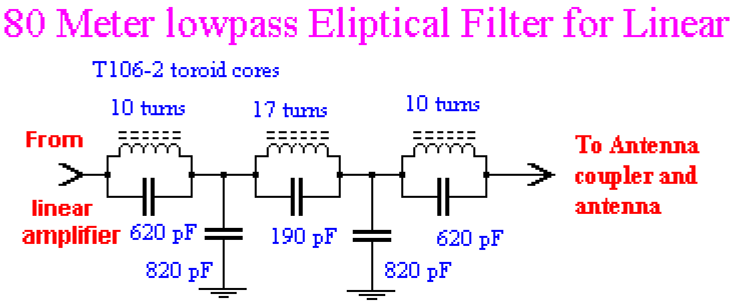
80 Meter lowpass Eliptical Filter for Linear

If you are like me, you will have a devil of a time getting your SSB drivers to produce intelligible speech without hissing and noise problems. All I can tell you is to keep your brain mulling over your difficulties. Shield and filter your prototype until the darn thing works. Keep careful notes so you don't make the same mistakes twice. Persistence will win in the end.

Hearing the transmitter VFO in your receiver

One consequence of shielding and filtering every low power module of my SSB transmitter was that I could not hear my own VFO signal. This made it impossible to tune the VFO to a station I might wish to talk to. Eventually I solved the problem by connecting a tiny capacitor, 10 pF, to the receiver antenna terminal on the antenna relay. This capacitor is connected to a shielded cable that runs over to my QRP module and plugs into a shielded phono connector on the side. Inside the QRP module an open-ended wire runs over to the first of the three broadband amplifiers (not the mixer) and passes one turn through the ferrite toroid core. This sampler wire is not a complete winding turn. It is *NOT* connected to ground or anywhere else. The wire just serves as a tiny capacitor or antenna to sample a bit of VFO signal for the receiver. During transmit the antenna relay disconnects the sampler wire from the transmitter to prevent it from causing feedback from the antenna into the transmitter.

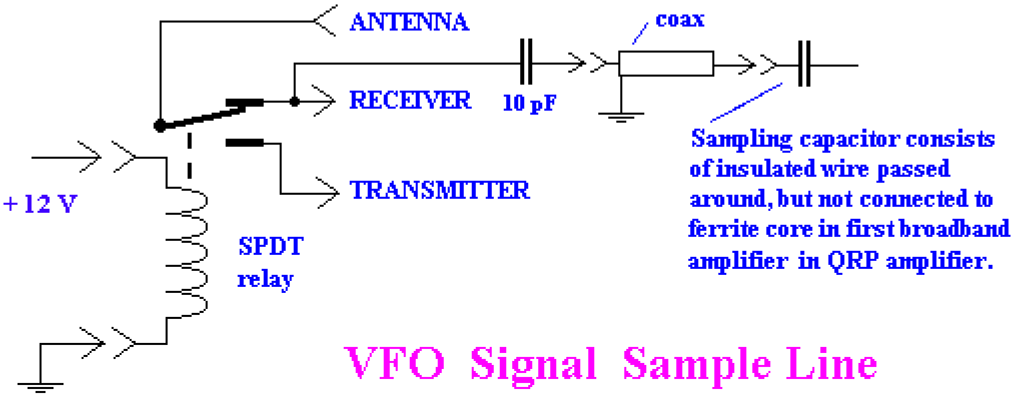
ANTENNA
coax
RECEIVER
10 pF
+12 V
TRANSMITTER
SPDT
relay
Sampling capacitor consists of insulated wire passed around, but not connected to ferrite core in first broadband amplifier in QRP amplifier.

VFO Signal Sample Line

When I push the "Spot" button on my transmitter, 12 volts power is connected to all the low power modules of the transmitter. During "Spot" mode, the 12 volt power is *NOT* connected to the last two power amplifier stages of the QRP. The 12 volts goes to the 5 MHz VFO, the SSB generator, the VFO frequency converter (if one is used on that particular band), the mixer, and first broadband amplifier of the SSB QRP module. The first amplifier is the first place in the transmitter where the actual broadcast frequency is present for sampling. Because the final frequency depends on the 9 MHz signal from the SSB generator, there will be no signal from the SSB generator unless you are actually talking or the generator is set to "CW." In summary, to hear the VFO without transmitting, the SSB generator must be set to "CW," the "Spot" button must be pushed, and the receiver antenna needs a tiny capacitive coupling to the first broadband amplifier. In SSB, even simple things are complicated. See the fun most hams are missing?

Adding 20 meter capability to the 80 meter QRP module

In the 80 meter module above there are two filter networks. You can put this same QRP module on 20 meters by switching in a 14 MHz filter after the mixer and by replacing the low pass on the output with a high pass filter. On 80 meters the 9 MHz and 5 MHz unwanted signals are above 4 MHz. Therefore, the 80 meter QRP module has a low pass filter. In contrast, on 20 meters the unwanted frequency components are below 14 MHz. Therefore, a high pass is desirable for 14 MHz. The filters to put the module on 20 meters are shown below. My module uses two DPDT switches to switch back and forth between the 80- and 20 meter bands.

20 meter Bandpass Filter

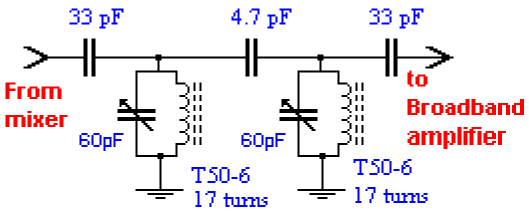
33 pF
4.7 pF
33 pF
From
mixer
60pF
T50-6
17 turns
60pF
T50-6
17 turns
to
Broadband
amplifier

20 meter
Output Hi-Pass Filter

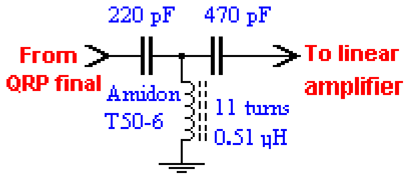
220 pF
470 pF
From
QRP final
Amidon
T50-6
11 turns
0.51 µH
To linear
amplifier

Feedback and distortion – don’t overdrive!

A frequent problem I ran into was using too much drive on a stage. For example, I built my prototype for 20 and 80 meters. And after many changes and fussing I got it to work. Then I reproduced the circuit for use on 17 and 12 meters. However, the second time I knew what I was doing and the new circuit was much "cleaner." The result of my compact, pretty wiring was higher efficiency and more power out of each amplifier stage. Instead of 3 watts output, now I had 6 or 8 watts or more and I was overdriving the linear final. Excess drive gave my signal a rough, rasping sound and made the speech hard to understand. To fix this, I had to go back and decrease the output by various methods. I decreased transformer turns driving bases, used more negative feedback (emitter resistance), etc. Finally I put the 500 ohm pot in front of the 2nd broadband amplifier. This enabled me to deliver just what I needed and no more. Now the voice quality was acceptable. The pot worked so well, I went back and installed one in my 80/ 20 meter driver module.

Moving an SSB signal to the "difficult" hambands

As explained above, the easiest hambands to reach with your SSB generator are 80 and 20 meters. Unfortunately, on weekends 20 meters is the most crowded ham band. It's full of guys running 1,500 watts peak into Yagi beam antennas 50+ feet in the air. If that weren't bad enough, their sideband transmitters are exquisitely designed to get the most modulation out of every watt. If you do get on 20 meters with your little homebuilt, it will probably average 20 to 40 watts on voice peaks. Combine that with your dipole antenna and it's going to be hard for those big guys to hear you. On the other hand, a band like 15 meters (or possibly 17 meters) is less crowded and you are more likely to make solid, enjoyable contacts there.

Getting on 15 meters

How do we move the 9 MHz to 21 MHz with a 5 MHz VFO? My solution was to move the 5 MHz VFO to 12 MHz. Then I added my 12 MHz VFO to 9 MHz to get 21 MHz. (12 MHz + 9 MHz = 21 MHz) After mixing, the 21 MHz signal was 43% different from the nearest frequency component and filtering was relatively easy. Unfortunately, moving the VFO to 12 MHz is quite cumbersome. However, it uses technology you have already mastered. So, in the long run, I believe moving the VFO is the easiest way to go.

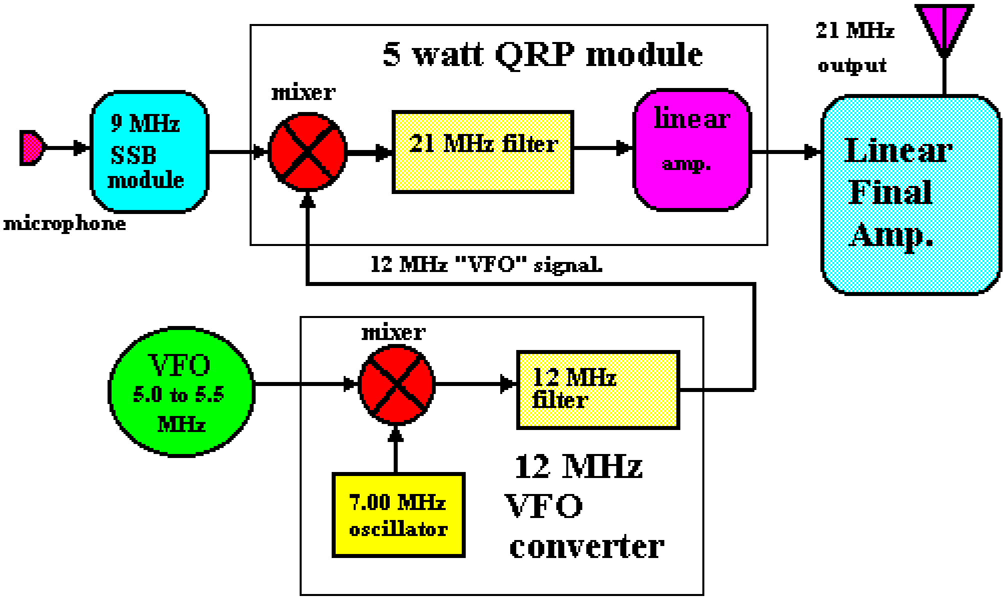
5 watt QRP module
mixer
9 MHz SSB module
microphone
21 MHz filter
linear amp.
21 MHz output
Linear Final Amp.
12 MHz "VFO" signal
mixer
VFO 5.0 to 5.5 MHz
12 MHz filter
7.00 MHz oscillator
12 MHz VFO converter

Generating a 12 MHz VFO signal

The VFO signal is just a sinewave. Therefore, compared to moving sideband, moving a sinewave to 12 MHz is relatively easy. There are no frequency or amplitude modulation components present, so the 12 MHz signal is easily filtered and purified. To move the 5 MHz VFO to 12 MHz, mix it with a 7.00 MHz signal from a crystal oscillator. Think of the VFO mover as a primitive frequency synthesizer.

A 5 MHz to 12 MHz VFO converter

The frequency converter contains the same circuits I used in my CW QRP boards in which I used an 80 meter VFO to drive a CW signal on each HF band. I used the dual gate mixer because it was simpler than the bipolar transistor mixers I used in my first QRP boards. If you start with an 80 meter VFO, it can be combined with a 8.5 MHz sinewave to give 12 MHz. You get the idea.

12 MHz VFO Converter for 15 Meters

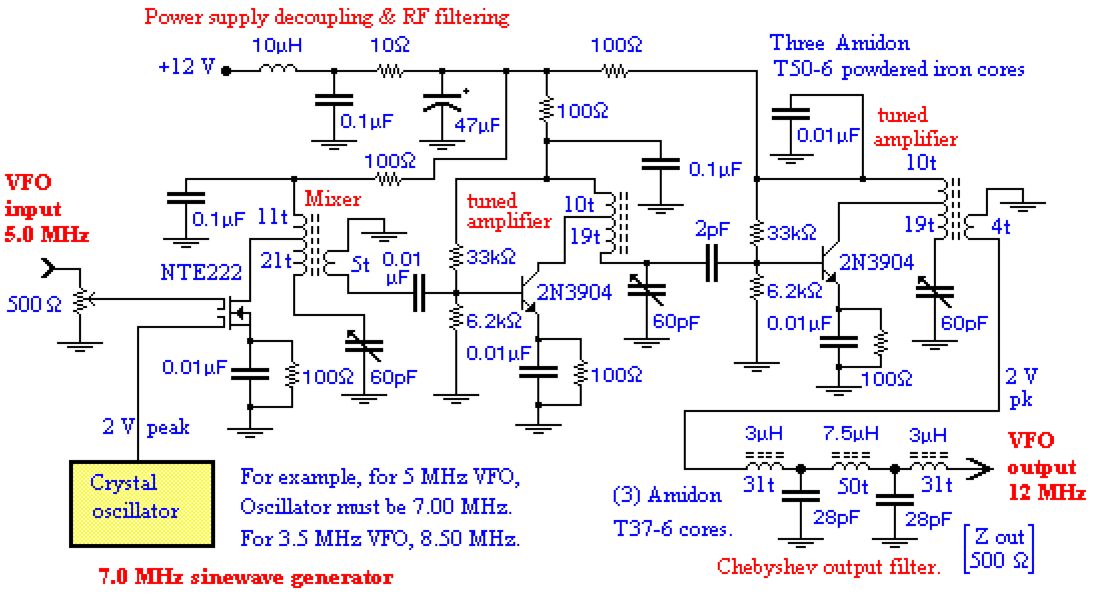
Power supply decoupling & RF filtering
+12 V
10µH
10Ω
0.1µF
47µF
100Ω
100Ω
100Ω
Three Amidon
T50-6 powdered iron cores
0.01µF
tuned
amplifier
10t
0.1µF
VFO
input
5.0 MHz
0.1µF
11t
Mixer
tuned
amplifier
10t
19t
2pF
33kΩ
19t
4t
NTE222
21t
5t
0.01
µF
33kΩ
2N3904
2N3904
500 Ω
6.2kΩ
6.2kΩ
60pF
0.01µF
60pF
0.01µF
100Ω
60pF
0.01µF
100Ω
100Ω
2 V
pk
2 V peak
3µH
7.5µH
3µH
VFO
output
12 MHz
Crystal
oscillator
For example, for 5 MHz VFO,
Oscillator must be 7.00 MHz.
For 3.5 MHz VFO, 8.50 MHz.
(3) Amidon
T37-6 cores.
31t
50t
31t
28pF
28pF
Z out
500 Ω
Chebyshev output filter.
7.0 MHz sinewave generator

The VFO converter moves the 5 MHz VFO up to 12 MHz

Before you build this, I suggest you look at the VFO movers for 17 meters and 10 meters shown later. I believe these newer VFO converter designs are more stable, more versatile, and easier to adjust.

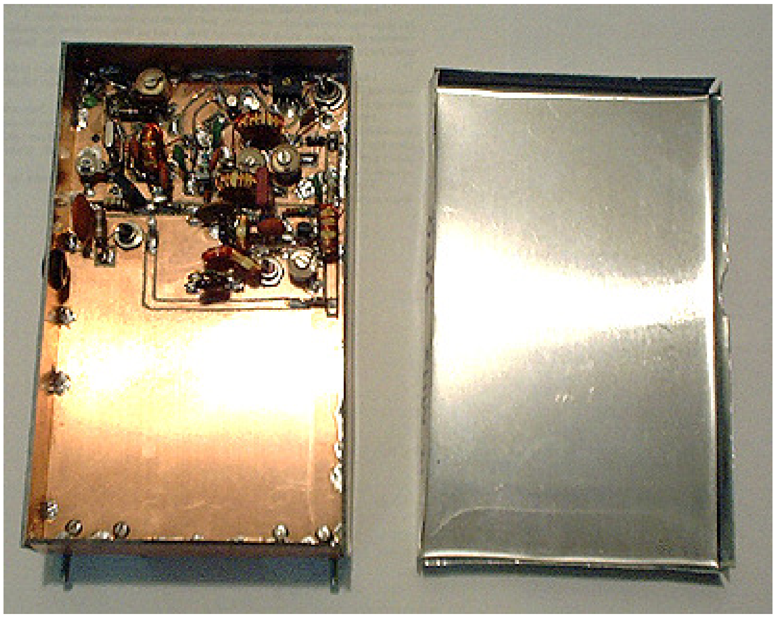

The 12 MHz VFO converter. I left room for a second VFO converter for some other band.

A linear sideband QRP module for 15 meters

Once you have generated a stable 12 MHz VFO, it needs to be mixed with the 9 MHz sideband signal to get on 21 MHz. Using the same design as the 80 meter QRP shown earlier, now all you need are the two passive filters to go into the QRP module. The bandpass filter is the same design as before, but I used a 5-element high pass filter on the output to get rid of the signals below 15 meters that tend to appear when I'm not talking. The two filters are shown below.

15 meter Bandpass Filter

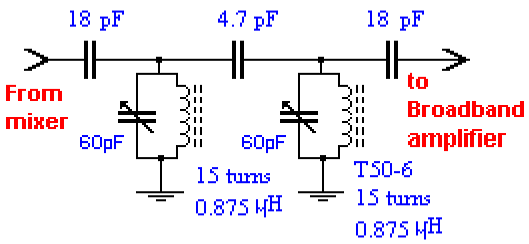

15 meter

Output Hi-Pass Filter

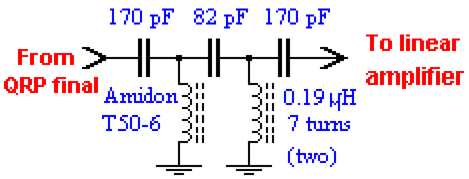

Notice that the output of the QRP driver described above has a *HIGH PASS* filter designed for 50 ohms that works best for driving a final amplifier. If you wish to run the driver "barefoot" and go on the air with just 5 watts peak, you'll also need a low pass filter, just like the ones you built for the CW QRP drivers. The most troublesome unwanted frequencies are the second harmonic of the 12 MHz VFO (24 MHz) and the second harmonic of the SSB generator, 18 MHz.

A 21 MHz Driver module. The box lid is shown above the board.

Notice the "pigtail" power supply cable in the above driver. This was an invitation to RF interference. This cable acts like an antenna. I later modified the Molex connector so that it is imbedded in the bottom wall of the module. Now the module is plugged directly into the metal transmitter chassis with no exposed cable. RF feedback has not been a problem.

17 meters can be really tough

My first problem with 17 meters was that the sweep oscillator in my oscilloscope makes an 18 MHz signal that my receiver picks up right in the middle of the 17 meter band. Another odd problem with 17 meters was that it interferes with my cordless phone and my stereo. The other bands don't cause this interference so it's a mystery to me. More importantly, I'm nervous about what might be happening over at the neighbors.

Anyway, even without those troubles, 17 meters is particularly difficult when starting with a 9 MHz SSB generator. 18 MHz is the second harmonic of the 9 MHz SSB signal. Therefore the 17 meter frequency converter will also amplify the second harmonic of the sideband signal. This means that, although there may be a good signal on 17 meters where it is supposed to be, (for example, 18.130 MHz), there will also be a small sideband-like signal on 18.000 MHz. Of course the frequency deviation of the unwanted signal will have twice the audio frequency modulation. In the old days it was routine to move low frequency VFOs to high frequencies using frequency multiplier amplifiers. It is difficult to avoid building a multiplier/ amplifier here and it will be hard to get rid of unwanted 18 MHz signals with simple filters.

In addition, if you generate a 9.130 MHz VFO signal to add to the 9.000 MHz SSB to get 18.130 MHz, you will also be transmitting the second harmonic of the 9 MHz VFO sinewave signal. That is, if the desired frequency is 18.130, there will also be another small second harmonic sinewave transmitted on 18.260.

Practical approaches to getting on 17 meters

In spite of these harmonic troubles I first pressed on with 9 MHz. To avoid the 2nd harmonic of a 9 MHz VFO, I added my 5 MHz VFO to a 22 MHz oscillator producing a 27 MHz VFO signal. Then I subtracted the 9 MHz from 27 MHz to get 18 MHz. ***Notice that when you subtract an SSB signal from a higher frequency, the upper sideband becomes the lower sideband and vice versa.*** The VFO worked fine, but occasionally I discovered that I was loading up on 18.000 MHz, the second harmonic of the SSB generator signal, not the correct frequency component. In short, the behavior with a 9 MHz SSB generator was too flaky to trust.

The best way to get on 17 meters is to start over with a different SSB generator frequency, say 8.0 MHz. Then you can combine that with a 26 MHz VFO and it will work OK. In other words, 26.13 MHz - 8 MHz = 18.13 MHz. With this approach you aren't using any 9 MHz frequency components and second-harmonic, out-of-band emissions will be much less likely. I strongly recommend this approach. I really don't think a 9 MHz SSB generator can be reliable on 17 meters without first moving the 9 MHz SSB signal up above 18 MHz. That is, you would have to move the sideband signal twice. My solution was to start over and build a new SSB generator with two filters switch-able between 8 MHz and 9 MHz. My 8 MHz crystal filter was just like the 9 MHz filter, but the 91 pF capacitors were proportionately increased to 100 pF.

There are lots of ways to screw up on 17 meters. For example, ***start with a 6 MHz SSB signal and the 3rd harmonic of the signal will be on 18.000 MHz and will be just as bad as 9 MHz.*** You can also screw up with 8 MHz. For example, the 8 MHz SSB can be added to a 10.15 MHz VFO to get 18.15 MHz. Unfortunately, the 2nd harmonic of 10.15 MHz is 20.30 MHz. This continuous sinewave is close enough to 18 MHz to go right out over the air whenever you stop talking.

17 Meters with an 8 MHz SSB generator:

As shown below, I eliminated the VFO harmonic problem by generating a 26 MHz VFO. The second harmonic of an 8 MHz SSB generator is 16 MHz. Fortunately it wasn't hard to avoid accidentally tuning it to 16 MHz. Notice again that the SSB is subtracted from a higher frequency so ***the SSB generator must be set to the lower sideband to get upper sideband on 17 meters.***

26 MHz VFO converter for 17 meters

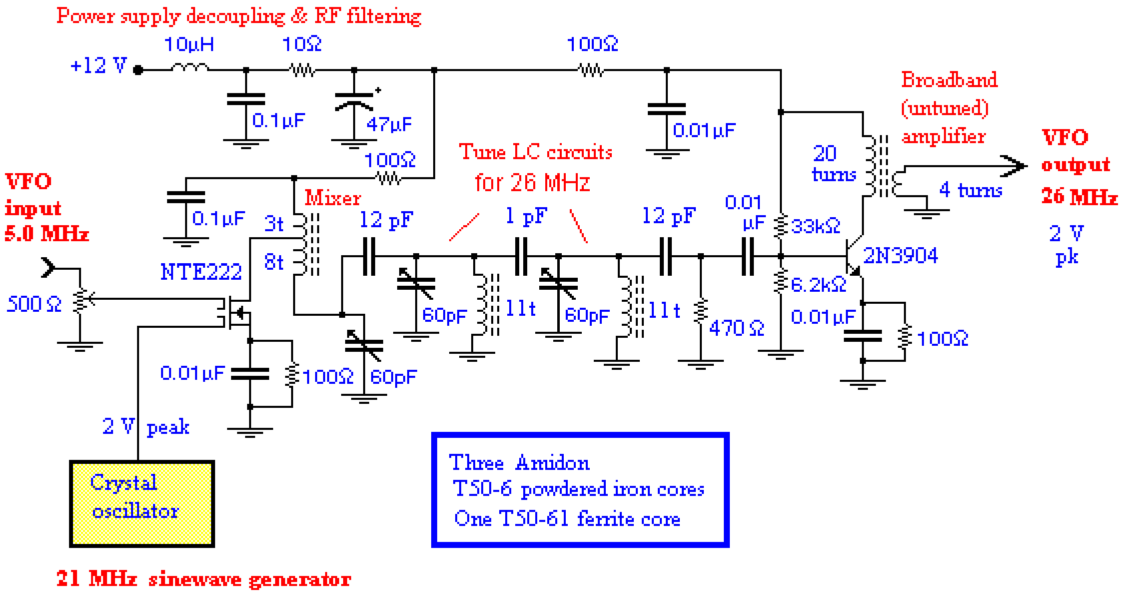
Power supply decoupling & RF filtering
+12 V
10µH
10Ω
100Ω
0.1µF
47µF
0.01µF
100Ω
Broadband (untuned) amplifier
VFO output 26 MHz
2 V pk
20 turns
4 turns
Tune LC circuits for 26 MHz
VFO input 5.0 MHz
500 Ω
0.1µF
Mixer
3t
8t
NTE222
12 pF
1 pF
12 pF
0.01 µF
33kΩ
2N3904
6.2kΩ
60pF
11t
60pF
11t
470 Ω
0.01µF
100Ω
0.01µF
100Ω
60pF
2 V peak
Crystal oscillator
21 MHz sinewave generator
Three Amidon T50-6 powdered iron cores One T50-61 ferrite core

The output of the 5 watt driver needs both a low-pass to get rid of the 26 MHz artifact and a high pass filter to avoid the low frequency "bursting" problem. Rather than choosing between high pass or low pass, I used a second 18 MHz bandpass filter that severely attenuates both 8 MHz *and* 26 MHz. This flat response bandpass filter is designed for 50 ohms. In contrast, the sharply tuned bandpass filter following the mixer is designed for high impedance, 500 ohms input and output. The 500 ohm bandpass filter is easier to design, but it can't handle any power and the impedance is incorrect for the QRP output.

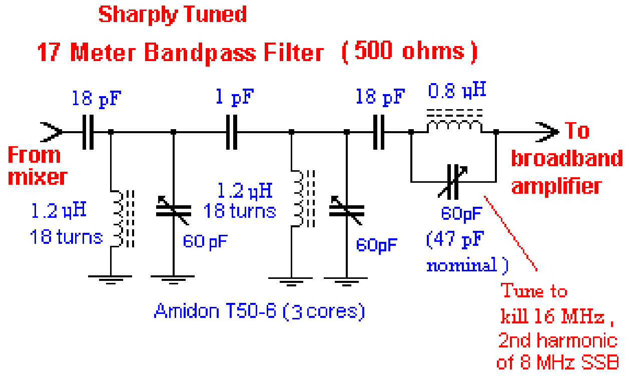
Sharply Tuned
17 Meter Bandpass Filter (500 ohms)
18 pF
1 pF
18 pF
0.8 µH
To broadband amplifier
From mixer
1.2 µH 18 turns
60 pF
1.2 µH 18 turns
60pF
60pF (47 pF nominal)
Tune to kill 16 MHz, 2nd harmonic of 8 MHz SSB
Amidon T50-6 (3 cores)

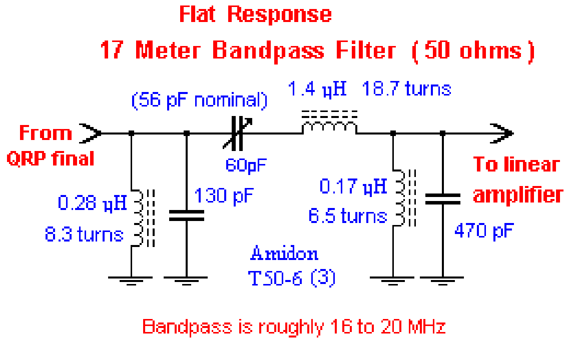
Flat Response
17 Meter Bandpass Filter (50 ohms)
(56 pF nominal)
1.4 µH 18.7 turns
From QRP final
60pF
To linear amplifier
0.28 µH 8.3 turns
130 pF
0.17 µH 6.5 turns
470 pF
Amidon T50-6 (3)
Bandpass is roughly 16 to 20 MHz

The filter on the right has a flat response from about 16 MHz to 20 MHz. In order to get the flat response, the three LC resonant circuits all have different values. And, of course, the three L-Cs interact, so a simple calculation wouldn't work anyway. I derived this circuit by trial and error using my Spice program. The component values must be rather precise otherwise the response has sharp peaks. Notice the fractional turns on the cores. These should help you be aware that wrapping the wire a tad more or less might make a difference. I found that making the smallest capacitor(s) variable was useful for final tune up. A few picofarads can make a big difference. Anyway, it was a struggle but eventually it worked. Tuning the series variable cap peaks it up nicely on 18 MHz.

Filters for 12 meters

In contrast to 17 meters, 12 meters was much easier. I generated a 15.9 MHz VFO signal by adding my 5 MHz VFO to a 10.700 MHz crystal oscillator. The QRP combines the 9 MHz SSB signal with the 15.9 MHz VFO to produce 24.9 MHz. The high impedance bandpass filter following the mixer is tuned for 24.9 MHz. At the output of the QRP module all the unwanted frequencies are well below 12 meters. Therefore the low impedance output filter is a simple high pass.

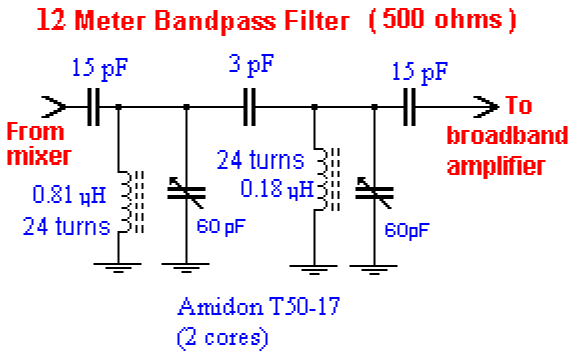
12 Meter Bandpass Filter (500 ohms)
15 pF
3 pF
15 pF
From mixer
To broadband amplifier
0.81 µH
24 turns
60 pF
24 turns
0.18 µH
60pF
Amidon T50-17
(2 cores)

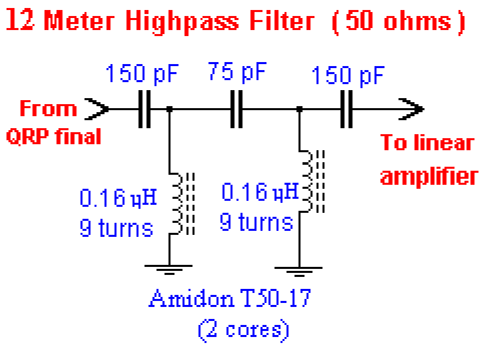
12 Meter Highpass Filter (50 ohms)
150 pF
75 pF
150 pF
From QRP final
To linear amplifier
0.16 µH
9 turns
0.16 µH
9 turns
Amidon T50-17
(2 cores)

The driver for 12 and 17 meters is shown below. Notice how this module has its Molex power plug mounted on the bottom. It plugs directly into the transmitter chassis and avoids exposed wires. This QRP driver also has an aluminum cover to shield the circuitry from RF.

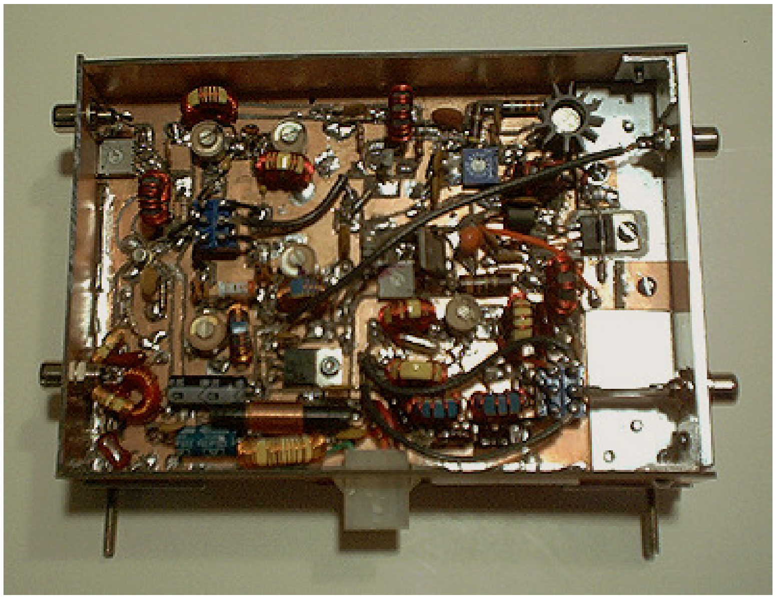

Getting on 40 meters SSB

Forty meters was also straight-forward. I generated a 16 MHz to 16.5 MHz VFO signal by adding the 5 MHz VFO to an 11.000 MHz crystal oscillator. Then I subtracted the 9 MHz

SSB to get 7.00 to 7.50 MHz. Because of the subtraction, the upper sideband 9 MHz SSB signal generated the lower sideband signal on 40 meters. The 40 meter QRP mixer filter uses almost the same tunable bandpass filter used in the DC receiver in chapter 7. Since all the frequency components are well above 7 MHz, a low impedance lowpass filters the output. As usual, 40 meters was kind to me and it worked right away. I was immediately able to check into my local state-wide 40 meter noon net.

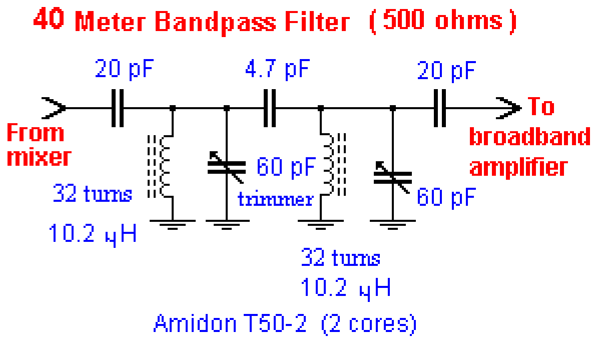
40 Meter Bandpass Filter (500 ohms)
20 pF
4.7 pF
20 pF
From mixer
To broadband amplifier
32 turns
10.2 μH
60 pF
trimmer
60 pF
32 turns
10.2 μH
Amidon T50-2 (2 cores)

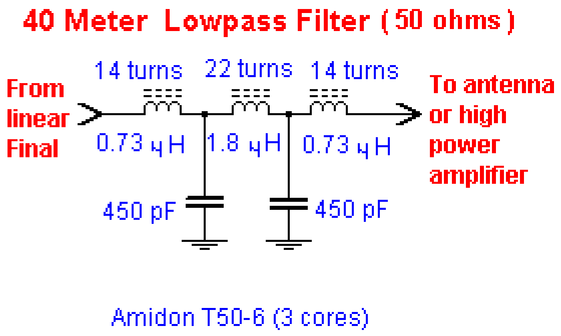
40 Meter Lowpass Filter (50 ohms)
14 turns
22 turns
14 turns
From linear Final
To antenna or high power amplifier
0.73 μH
1.8 μH
0.73 μH
450 pF
450 pF
Amidon T50-6 (3 cores)

SSB on 10 meters

The tricky part about 10 meters is that it's 1.7 MHz wide, 28.0 MHz to 29.7 MHz. My VFO only tunes a range of 0.5 MHz. Therefore it takes 4 crystal PMOs to cover the whole band. I solved the problem with a quadruple frequency crystal oscillator. A 6 position rotary switch switches in the appropriate crystals for 28, 28.5, 29.0, and 29.5 MHz. My oscillator uses the other two switch positions to cover 12 meters and 40 meters. The crystals aren't switched directly, but rather, they are grounded one at a time by means of diodes that are biased *ON* by means of a 12 volt DC signal passed through the 4.7 K resistors. The advantage of this DC switching method is that, if you like, the rotary switch can be far away from the oscillator up on the front panel.

Be sure to minimize the capacitances between the oscillator emitter and ground. All those crystals connected to the emitter will be affected by the extra capacitance. This will tend to pull each crystal frequency away from its nominal value. Too much capacitance and it may not oscillate at all.

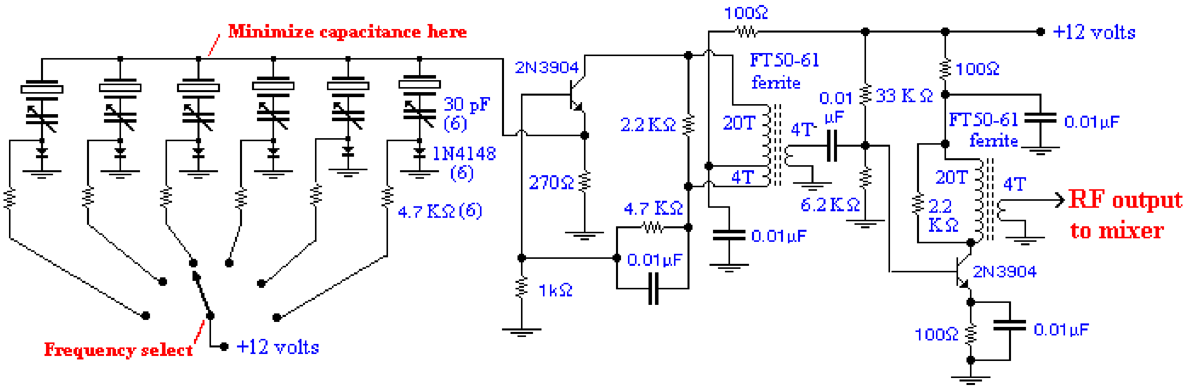
Minimize capacitance here
100Ω
+12 volts
2N3904
FT50-61
ferrite
0.01 µF
33 K Ω
100Ω
FT50-61
ferrite
0.01µF
30 pF
(6)
1N4148
(6)
4.7 KΩ (6)
2.2 KΩ
20T
4T
270Ω
4T
4.7 KΩ
6.2 KΩ
20T
4T
2.2
KΩ
RF output
to mixer
0.01µF
0.01µF
0.01µF
1kΩ
2N3904
100Ω
0.01µF
Frequency select
+12 volts

MULTI - CRYSTAL OSCILLATOR

The multi-crystal oscillator covers a wide range so the oscillator could not be tuned and had to be broadband. Therefore the oscillator frequency is entirely controlled by the crystal frequency. Obviously the particular crystal must spontaneously oscillate on that overtone frequency and not a lower, primary frequency. And, because the oscillator stage isn't resonant, its signal output is tiny, tenths of a volt. Consequently, I had to pass the oscillator signal through a broadband amplifier to make it large enough, about 2 volts p-p, to be sent to the mixer to be mixed with the 5 MHz VFO.

After the mixer each VFO signal must be filtered to select the desired frequency component. My 6 position rotary switch has a second section that allowed me to switch in a filter for each frequency. I found that just two tuned filters could cover the whole 10 meter band. I didn't need 4 separate filters after all.

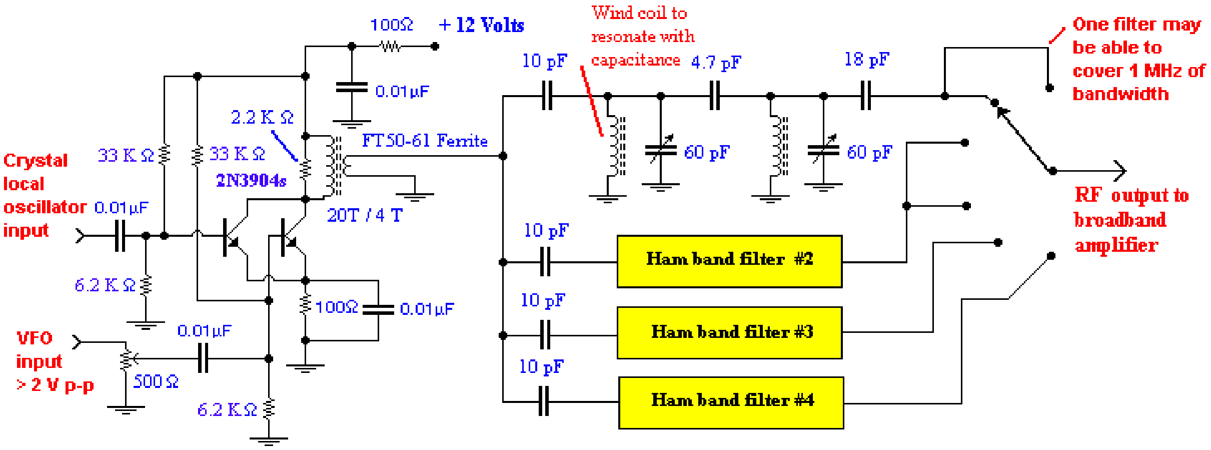
100Ω
+ 12 Volts
Wind coil to
resonate with
capacitance
4.7 pF
18 pF
10 pF
One filter may
be able to
cover 1 MHz of
bandwidth
0.01µF
2.2 K Ω
FT50-61 Ferrite
60 pF
60 pF
Crystal
local
oscillator
input
33 K Ω
33 K Ω
2N3904s
0.01µF
20T / 4 T
RF output to
broadband
amplifier
10 pF
Ham band filter #2
6.2 K Ω
100Ω
0.01µF
10 pF
Ham band filter #3
VFO
input
> 2 V p-p
0.01µF
500 Ω
10 pF
6.2 K Ω
Ham band filter #4

MULTIPLE FREQUENCY MIXER

The tuned filters are high impedance and connected to the mixer with 10 pF capacitors. Because all the filters are connected to a low impedance transformer winding, the mixer can drive all of them at once. As shown the total load is only 40 pF. The rotary switch on the right then picks out the desired filter output. Another broadband amplifier amplifies the filter output before it goes to the SSB QRP driver where it is combined with the 9 MHz SSB signal. By the way, I tried to use the switching-by-grounded-diode trick to switch filters but that idea worked poorly. For 10 meters this VFO converter generates VFO signals from 19 to 21 MHz.

Frankly, my 10 meter QRP only seems to works well up to 29 MHz. It doesn't work properly above that frequency because my QRP mixer filter isn't broadband enough. I guess I need two high impedance tuned filters. Also, since 10 meters has been dead for several years now, I have yet to make a 10 meter SSB contact. Needless to say, my 10 meter SSB is still a work in progress.

Getting on 60 meters SSB - don't bother!

Before the 60 meter (5 MHz) SSB frequency became available to American hams on July 4th 2003, I thought it would be fun to get on the air before commercial gear became available for this new frontier. I naively thought that homebrewers would own the band for at least a little while. My sad story about my attempt to be a pioneer on 60 meters is told in Chapter 16.

Checking out the QRP module

You will have to experiment with how to listen to your sideband signal at a signal level that simulates what it would sound like if you were receiving it off the air. I run my QRP into a 50 ohm dummy load. Then I disconnect my receiver from the antenna and leave the antenna coax from the receiver lying on the bench a few feet from the sideband generator. If that’s not strong enough, I clip a test lead to the transmitter ground, then clip the other end to the center conductor of the receiver antenna coax connector.

I start by feeding a 9 MHz sinewave (CW) signal into my QRP board and then tune the filters to produce the largest, stable sinewave output into the dummy load. I monitor the frequency with a counter to be sure the VFO is controlling the frequency properly over the entire ham band. I adjust the VFO input level to produce the maximum output signal. However I just use the minimum 5 MHz VFO level that achieves this. When you are *NOT* talking, excess VFO will tend to induce signals on unwanted frequencies. I increase the DC bias to the output transistor until that transistor draws about 20 milliamperes DC more than it does with the bias set to minimum.

After the CW mode is working properly, I switch to SSB and use an audio signal injected into the "test input" of the SSB generator. I use speech from a talk radio station as supplied by a little Walkman radio. If you're lucky, the speech should sound pretty good in your ham receiver. If it doesn't, decrease the drive to the last two stages of the QRP using the 500 ohm pot. You may also need to reduce the number of secondary turns on the transformer feeding the output transistor. For example, instead of 3 turns, 2 turns or even one turn may be optimum for your particular QRP board. I found that overdriving stages was a common cause of poor audio.

Driving a QRO linear amplifier

A QRP sideband transmitter is just fine for communication around town. One or two watts are plenty for talking a few miles. But unless you have a great antenna and good conditions, you won’t talk to many stations with just a few watts. For distant stations a linear amplifier will be a big improvement. Building a 50 watt linear is explained in chapter 12. If you are going to have trouble with RF feedback and insufficient power supply decoupling, a big linear amplifier will bring out these troubles. RF from my antenna coupler feeds back to my

Walkman radio and (usually) makes that speech source impractical for testing. I usually listen to the receiver with headphones with the receiver volume turned way, way down. Don't deafen yourself! When you speak into the microphone, your voice should sound clear, as though you were talking on a public address system. It should not sound rough and gravelly.

Watch the output waveform across the dummy load with your scope. The waveform should look just like it did coming out of the 9 MHz generator. You will probably find that speech sounds terrible before you have everything adjusted. It may just be bursting, sputtering sounds.

Adjusting DC bias to the final

You'll have to adjust the DC bias to the final (50 watt) amplifier for optimum speech quality. As you increase the bias current, watch the DC current drawn by the entire transmitter. It should not be more than about 2 amperes when you are not talking. As you talk, the current should jump up to 6 to 12 amperes, depending on the drive levels, the band you're on, etc. As always, the higher the frequency, the more difficult it will be to obtain clear speech. ***DC bias for the high power amplifier that is adequate for one band, may not be enough for another.*** This little pearl of wisdom cost me days of frustration.

If the speech still sounds bad, RF chokes and RC decoupling filters for power leads to each module can help the problem. Also, filter the DC power line entering the transmitter chassis and the remote "mute" line going to the receiver. If troubles persist, filter all the wires entering your transmitter. Sometimes a clamp-on ferrite filter block around power leads or cables can be helpful.

Finally, SSB works best with a good antenna

A high gain, beam antenna is highly desirable. As you listen to the other SSB stations, you'll find that most strong signals come from a beam antenna. Directional antennas improve the signal by focusing most of the RF energy toward the guy you are talking to. Think of beam antennas as being comparable to the mirror reflector in a flashlight. The mirror concentrates the energy in one direction only.

In conclusion

My first real sideband contact was with W9WFE, a fellow about a thousand miles away. When I explained to him that my rig was homebrew, he said, “Well, it certainly sounds like sideband to me. It seems to work!” Sweet success.

My sideband transmitters are still in the experimental category. You will find that it takes a great deal of tweaking and fussing to get SSB tuned so it sounds good and doesn’t radiate on unplanned frequencies. You won’t believe how many diseases your SSB transmitter will create for you to conquer! Sideband is not a project for impatient people.

Shortly after I got my sideband working, I tried to arrange a schedule with Doug, KD6DCO, in California. We failed to make contact. In that weak moment I thought I should stop messing with homemade junk and buy a modern transmitter. No, wait. If I want to communicate with Doug, all I have to do is write him an e-mail or call him on the phone. I’m already on-line and long distance calls are cheap. If I wanted to use radio, I could even talk with him by cellphone. No, it was back to the drawing board for me.

And after some major redesign, my next schedule with Doug was successful, but my signal was pretty weak out in California. That's OK. I have to keep reminding myself that, so long as my station falls short of what is technically possible, my hobby continues. Woe to me if I ever finish. Long live homebrew!!

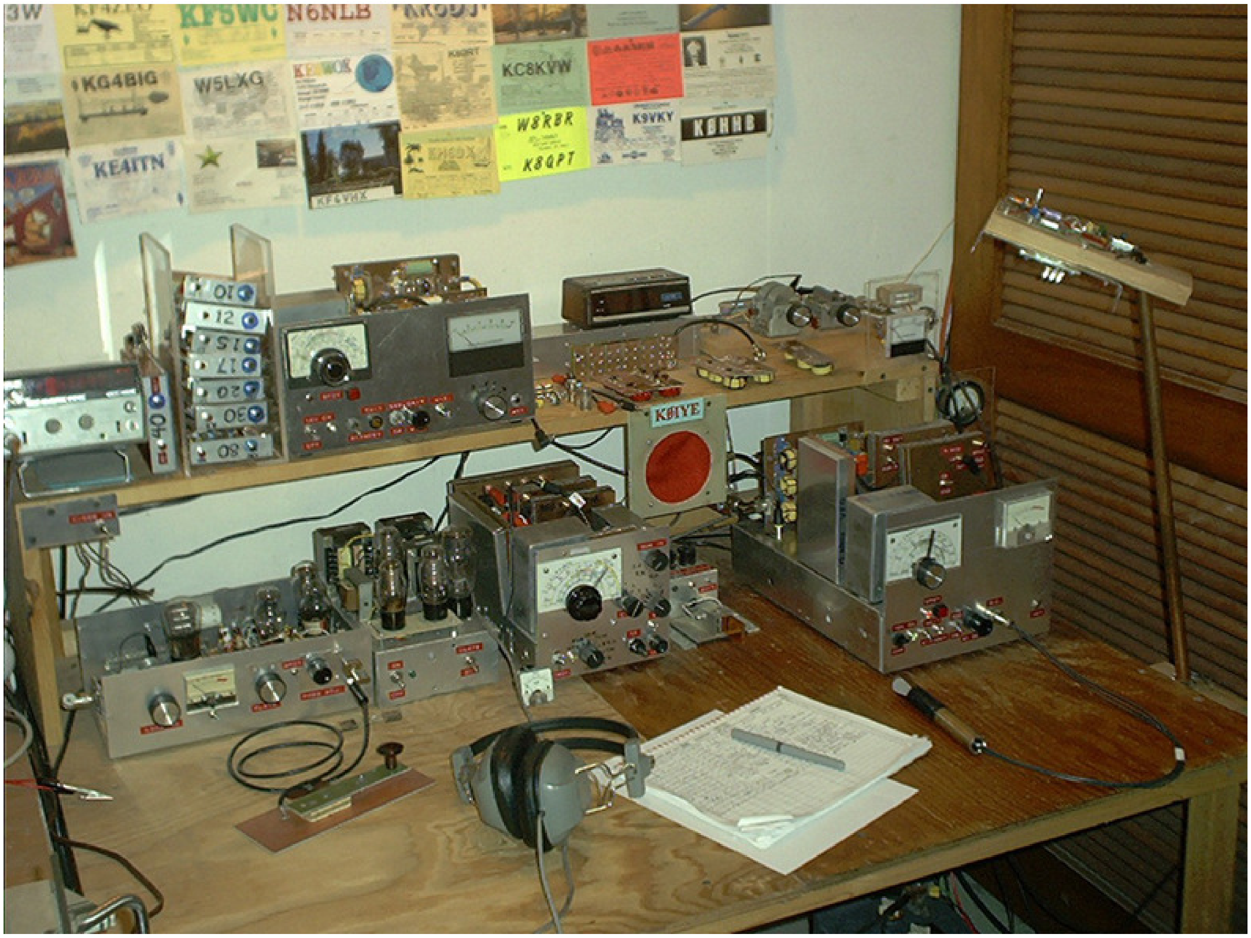
KF5WC
N6NLB
KG4BIG
W5LXG
KC8KVW
W8RBR
K9VKY
K8HHB
KE4ITN
K8QPT
KF4VHX
K8IYE
10
12
15
17
20
30
40
80

CRYSTAL SETS TO SIDEBAND
© Frank W. Harris 2010, REV 12

Chapter 16

ANCIENT MODULATION

And other topics

When I got back into ham radio 11 years ago, my ham friends told me that amplitude modulation (AM) was extinct. I was under the impression that SSB was the only mode of HF phone permitted. Later I learned that AM isn't actually illegal and there are a few diehards using AM on the 75 and 10 meter phone bands. I've also heard AM stations on 15 and 160 meters. In short, you might find a use for it. Besides, it's an interesting challenge to AM-modulate a transistorized transmitter.

Homebuilt AM

Back in the vacuum tube days many of us built our own transmitters and AM modulators. My first AM transmitter was a Heathkit DX-20. That was a 50 watt, CW-only, kit-built, vacuum tube transmitter to which I added a homebuilt AM modulator. Unlike SSB, AM could be added onto an existing CW transmitter. Rather than generate a low power AM signal and then amplify it with a linear amplifier, in the old days the usual method was to AM-modulate the final amplifier of the CW transmitter.

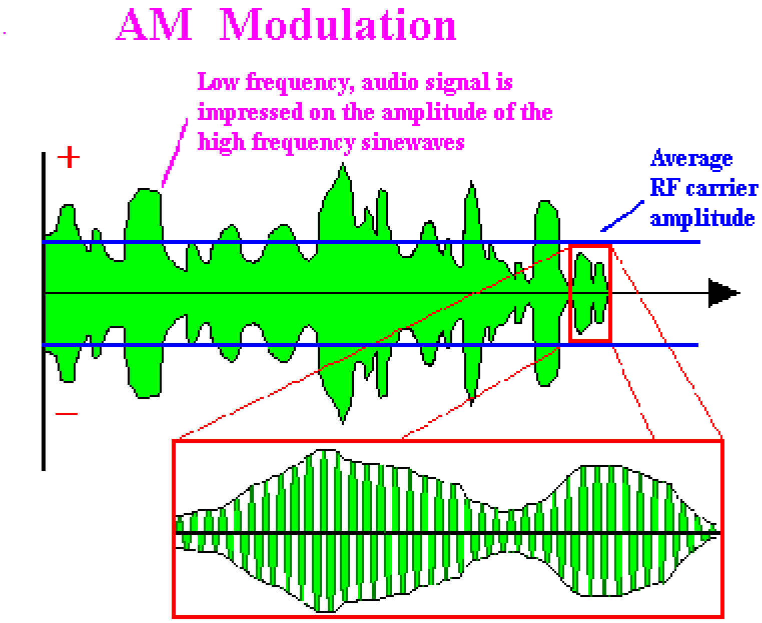
AM Modulation
Low frequency, audio signal is impressed on the amplitude of the high frequency sinewaves
+
–
Average RF carrier amplitude

On an oscilloscope, the hallmark of AM is that, when you are **not** speaking, the RF carrier wave runs continuously at an average power. That is, in AM the highest peak power and zero power only occur at the very highest voice peaks. Although I could see these transient peaks on the scope, when I tried to catch one with a storage scope, they are statistically rare and I couldn't catch a zero power level. The waveform below was typical of what I saw.

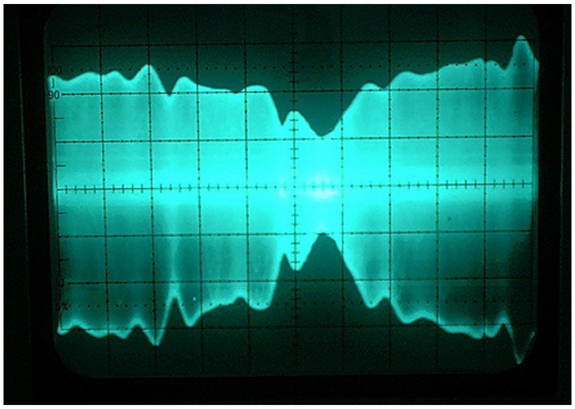

In contrast to AM, the RF output amplitude in SSB is always **zero** whenever you aren't talking. Notice in the SSB oscilloscope picture below that each RF blip representing the audio starts from zero. It doesn't start from a halfway, continuous carrier level.

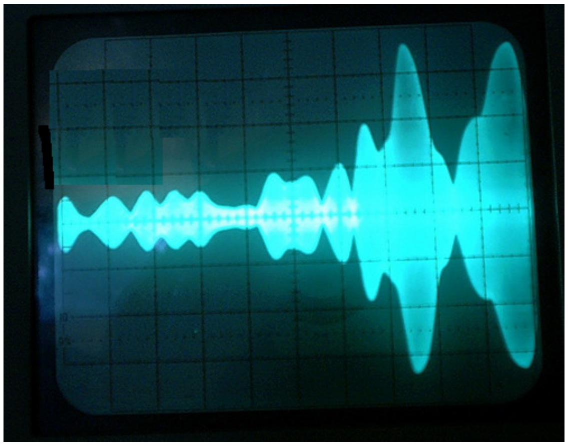

Plate, screen, and cathode modulators

Formerly, there were three common methods of AM modulation. The "Mercedes" method was to use a ***plate modulator transformer***. These large iron transformers impressed the audio signal onto the DC supply current. That is, as you talked, the DC input current rose and fell above and below the level of what it would be for a CW sinewave. For a 100 watt transmitter, this transformer was about the size of a softball, weighed a ton, and cost like crazy. The transformer was driven with a big audio amplifier that put out at least 50% of the CW carrier power. In other words, the plate modulator circuitry was nearly as large and expensive as the rest of the transmitter.

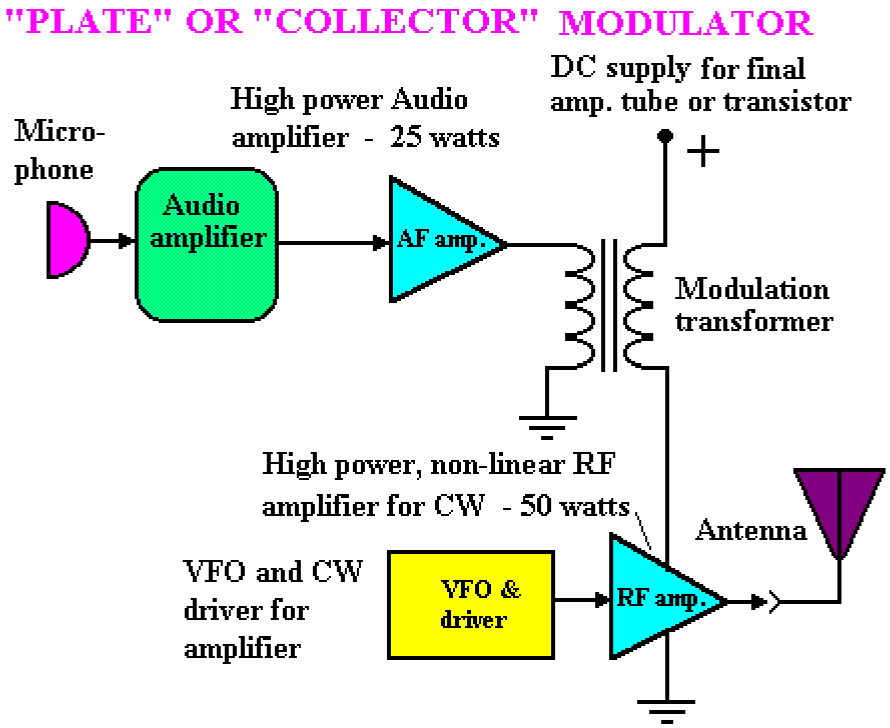
"PLATE" OR "COLLECTOR" MODULATOR
Micro-phone
Audio amplifier
High power Audio amplifier - 25 watts
AF amp.
DC supply for final amp. tube or transistor
+
Modulation transformer
High power, non-linear RF amplifier for CW - 50 watts
Antenna
VFO and CW driver for amplifier
VFO & driver
RF amp.

The "Toyota" and "Bicycle" approaches to AM modulation were to modulate the gain of the final amplifier tube by impressing the audio on the screen or cathode, respectively. Screen modulators usually sounded pretty good. Cathode modulation, sometimes called Heizing modulation, tended to produce "down modulation" which meant that power decreased whenever you talked. It sounded just fine, but was inefficient use of RF power output. These methods required less audio power than plate modulation and were easy for a high school kid like me to afford and build.

Modern AM construction

Now forward to 2003. Most modern SSB transceivers have the capability to generate AM modulation. To get into this mode, you read your manual for 20 minutes, bring up menu #26, push button numbers 14, 7, and 12 and you're done. That wasn't hard, I guess. But did you learn anything?

Let's suppose that you're a homebrew fanatic and wish to scratchbuild your own AM rig using transistors. Is that hard? Hmmmmm. Well, for one thing, transistors don't have cathodes and screen grids. Emitters are analogous to cathodes but, as explained above, cathode modulation wasn't all that great. Another difference between tubes and transistors is that, for the same power levels, ***the final amplifier transistor has DC currents about 50 times larger.*** So for DC supply modulation, you must impress 5 or 10 ampere audio signals onto the 12 volt DC power supply line. The modulation transformer will have to be just as large, but it will need a super low impedance output winding.

Modulating a transistorized 50-watt CW transmitter

I have a 25-watt, plate modulator transformer from the 1960s designed for use with a transistorized audio amplifier modulating a vacuum tube transmitter. In other words, it was designed to modulate a vacuum tube final amplifier, but the modulator itself was transistorized. In those days high power, high frequency transistors didn't exist, so transmitters were built with tubes, but audio circuits could be built with transistors. Since my transformer had low impedance primary windings, I thought I could "run it backward" and supply enough audio current drive to build an AM "collector modulator." I happen to have an old 10-watt vacuum tube hi-fi amplifier, so I used that to drive the high impedance winding on my modulation transformer. Sure enough, even with music my AM modulation sounded great when I broadcast into a dummy load. However, it only modulated about 30% of the carrier amplitude. That is, I was wasting most of my RF power. I could have built a 25-watt vacuum tube output audio amplifier, but I had a more modern idea. Why not use my MOSFET CW keyer as an audio modulator?

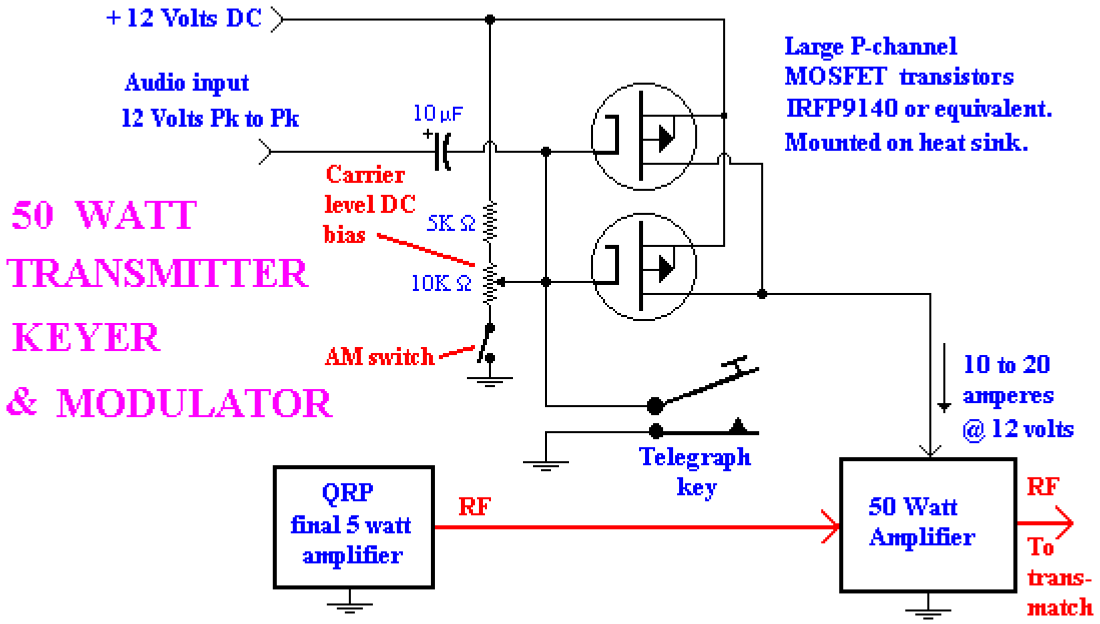
+ 12 Volts DC
Audio input
12 Volts Pk to Pk
10 µF
+
Carrier
level DC
bias
5K Ω
10K Ω
AM switch
50 WATT
TRANSMITTER
KEYER
& MODULATOR
Large P-channel
MOSFET transistors
IRFP9140 or equivalent.
Mounted on heat sink.
10 to 20
amperes
@ 12 volts
Telegraph
key
QRP
final 5 watt
amplifier
RF
50 Watt
Amplifier
RF
To
trans-
match

The above keyer was originally designed to turn the DC power to my final on and off with a telegraph key. My AM modulation scheme was to turn the MOSFETs half-on with a simple DC potentiometer, then modulate the gates with a 12 volt P-P audio signal. Because I was driving MOSFETs with a low (audio) frequency signal, hardly any power was needed. This simple scheme worked pretty well, but it was extremely finicky to adjust. It was easy to have too much bias or too little and too much modulation or too little. The difficulty was that the gate voltage versus drain current transfer characteristic is rather non-linear. With feedback and a more sophisticated drive circuit, I believe this method can be made to work well.

The SSB approach to AM

At this point in my R&D, I had not yet succeeded in building a practical SSB transmitter. So rather than invest more time on "obsolete modulation," I went back to work on SSB. I figured that, if I ever got the SSB working, it would be easy to downgrade my SSB generator to AM. This turned out to be true. I tried out several variations. ***The method that was simplest and***

worked the best was bypassing the SSB crystal filter with a switch and unbalancing the balanced modulator circuit using a circuit that resembles the CW switch.

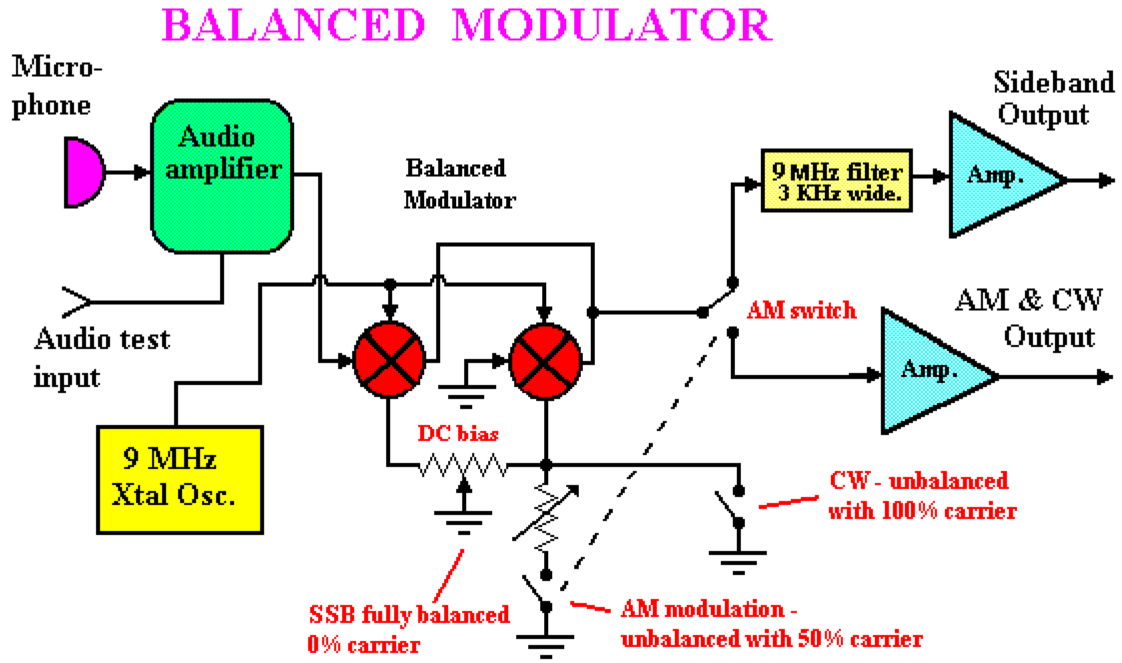
BALANCED MODULATOR
Micro-phone
Audio amplifier
Audio test input
9 MHz Xtal Osc.
Balanced Modulator
DC bias
9 MHz filter 3 KHz wide.
Amp.
Sideband Output
AM switch
AM & CW Output
Amp.
CW - unbalanced with 100% carrier
SSB fully balanced 0% carrier
AM modulation - unbalanced with 50% carrier

AM resembles CW in that a sinewave carrier is generated continuously. However, the same "unbalance" switch used as a SSB/ CW mode switch can't be used for AM. When modulation is applied, the instantaneous power must rise above and below the no-speech carrier level. Ideal AM modulation drives the carrier alternately between zero and 200% of the carrier level. Because there is a limit on the signal amplitude available, ***the carrier must be set to 50% of the level used for CW.*** This provides a modulation amplitude range of +/- 100%. A separate AM mode, double-pole switch bypasses the SSB filter and unbalances the modulator 50%. The AM switch is in series with an adjustable 5K ohm resistor that unbalances the modulator just enough to produce the 50% carrier.

The audio gain pot and your voice level should be adjusted to produce voice peaks twice the carrier level. Compared to SSB, you'll find that AM modulation is quite HI-FI. While testing the generator and transmitter on an 80 meter dummy load, music retransmitted from a Walkman was quite acceptable. In contrast, when using SSB, speech sounds OK, but music is really terrible. The principle difference is that the sideband filter greatly attenuates frequencies below 300 Hz whereas AM preserves the low frequencies. Speech transmitted on SSB can sound like the person's normal voice, but music on SSB is really bad. It's just as well. The last I heard, ham music is still illegal.

COMPRESSION BY ACCIDENT

Or, sometimes we get lucky

A modern single sideband generator processes the amplified audio from the microphone before the audio is fed into the balanced modulator. This "***compression***" process attempts to equalize the voice peaks so that as many voice elements as possible are transmitted with full Peak-Envelope-Power. Without this process, most of what you have to say will be transmitted with far less than the nominal peak power. When most of your sentences are reduced to QRP muttering, your intelligibility suffers.

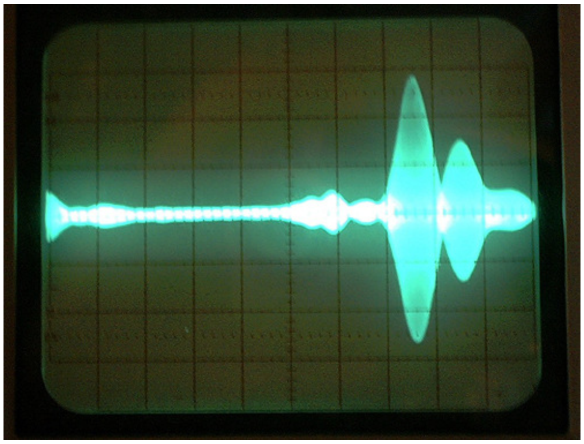

In other words, without compression, the single sideband RF envelope of a spoken word is close to zero most of the time. It would look something like the waveform shown above. A compressor circuit attempts to leave the peaks alone while proportionally amplifying the subtle, low voltage waveform wiggles near the horizontal axis. I guess the latest transceivers use digital processing to accomplish this feat. However, 15 years ago a compressor circuit usually performed the following tasks:

1. It amplified the whole audio waveform.
2. It clipped off the highest audio peaks.
3. And finally, it filtered the clipped audio with a 300 Hz to 3KHz bandpass filter.

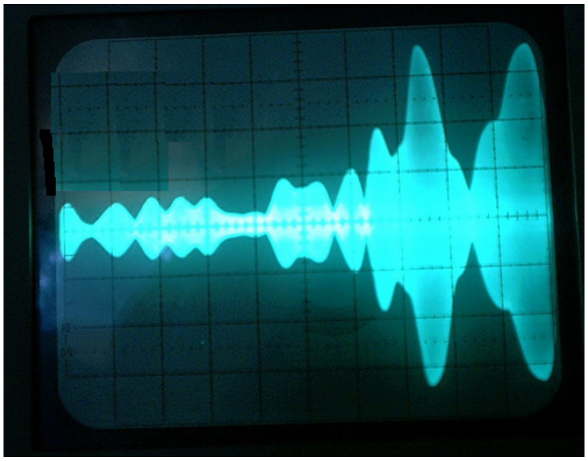

After compression, the same RF sideband waveform might look something like the above picture. The idea is that all the tiny stuff near zero has been expanded. (These waveforms aren't actual before-and-after pix, but they illustrate the principle.) After transmission some modern receivers "re-expand" the waveform to try to restore the original waveform. This entire process is called ***companding.*** However, for me, building a homebrew SSB that worked at all seemed plenty difficult. Consequently I didn't worry about secondary issues like "companding."

A crystal filter does more than clip the unwanted sideband

In the beginning I was afraid my RF signal might be too wide. So, because it was relatively easy, I built a 3 KHz audio low pass filter. It turned out that I didn't need it. Once I had passed the 9.000 MHz RF double sideband signal through the crystal filter to cleave off the unwanted sideband, I found that the filter had also removed virtually everything above 3 KHz anyway. Also, when I adjusted the original sinewave frequency to get rid of every trace of the carrier, I found the filter had also clipped off the lower 300 Hz of the audio. It's remarkable how normal a voice can sound without the lower 300 Hz. Voices are quite lifelike. In any case the crystal filter accomplished the same frequency filtering that the ARRL Handbook specified for the audio compressor. Interesting!

An SSB transmitter has several linear amps in series

After the SSB RF signal has been generated at a milliwatt level, the signal must be amplified and converted to the desired hamband. Including the mixer, this meant that my SSB signal had to pass through 5 stages of amplification to get to 100 watts peak. Each linear stage is forward biased so that even tiny signals will be amplified. Without this bias, all you hear are the voice peaks. In other words, an unbiased amplifier cuts off all the little audio signals a compressor tries to accentuate. I knew that the linearity of all these stages in series couldn't possibly be "perfectly linear." But since voices sounded good, I stopped worrying about linearity.

Where has all the AM modulation gone?

I didn't realize that my RF amplifiers were significantly non-linear until I added an Amplitude Modulator mode to my SSB generator. I listened to my little 9 MHz AM generator in my all band shortwave receiver. It sounded great and looked like 100% modulation on the scope. Next I fed the signal from the 9 MHz AM generator into my 80 meter "linear" QRP module which put out about 3 watts on 80 meters. Yes, it worked, but the signal was nearly all carrier. Instead of 100% modulation, on 80 meters I only had about 5% modulation. Where did that huge carrier signal come from? What happened to my modulation?

Transistors aren't linear

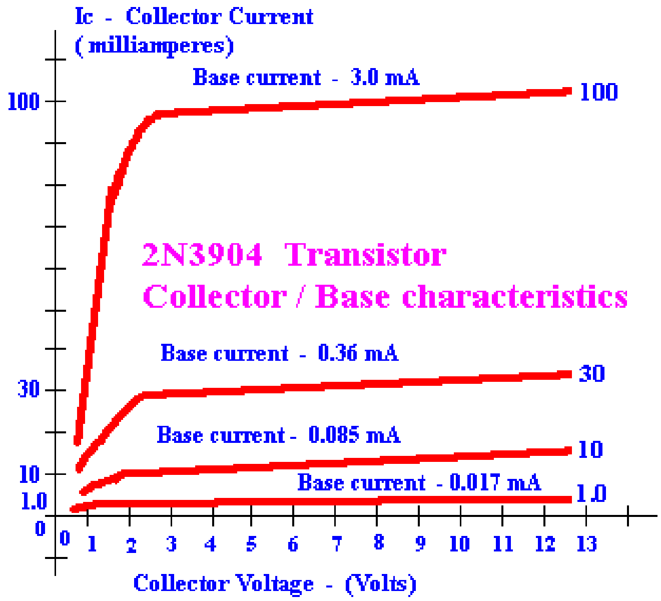

"Linear" implies that big signals will be amplified just as much as the small ones. However, if the raw output of the transistor covers most of the collector operating range, then it happens that small signals are amplified more than big ones. I have two 2N3904 transistors in my chain of amplifiers, so the Base/ Collector current characteristics for this transistor are shown above. Notice that one milliampere of collector current requires 0.017 milliamperes base current. But to get 10 milliamperes of collector current takes 0.085 milliamperes. That's 5 times more base current to get 10 times more collector current. But if you want 100 milliamperes of collector current, you need 3.0 milliamperes of base current. That's an additional 35 times more base current. Sure looks non-linear to me. ***BEHOLD, A NON-LINEAR COMPRESSOR!***

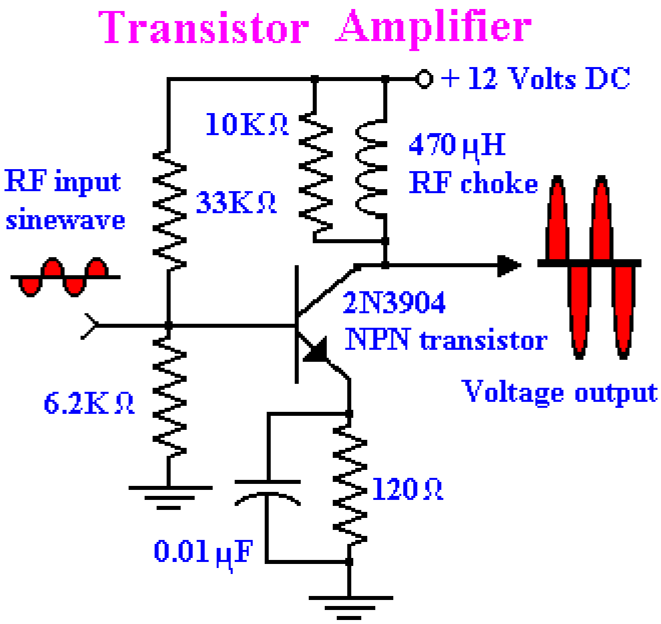

Transistor Amplifier
+ 12 Volts DC
10KΩ
470 μH
RF choke
RF input
sinewave
33KΩ
2N3904
NPN transistor
6.2KΩ
Voltage output
120 Ω
0.01 μF

The "linear" amplifier above illustrates an accidental compressor circuit. The 33K resistor biases the transistor ON so that even tiny RF signals will be amplified. (By the way, the 10K resistor across the inductor keeps the amplifier from oscillating when there is no input signal.) The main reason for the 120 ohm resistor is to provide negative DC feedback to make the amplifier thermally stable. Without the emitter resistor, the amplifier works, but the transistor runs extremely hot. The emitter resistor also makes the amplifier more linear than the transistor characteristic would suggest because the feedback restricts the transistor to a narrower range of operation. However, 120 ohms feedback makes it a long way from linear. 470 ohms is much better, but still far from perfect.

Oh, well, why fight it? To fix my AM mode, I reduced the imbalance of the balanced modulator to just a few percent of voice peaks. This gives me roughly 50% carrier by the time it arrives at the final amplifier. And as for the SSB, it already works well. Apparently I had a pretty darn good compression system all along and I didn't even know it. Imagine! A happy accident! They sure don't happen often.

****

HAM TELEVISION - The Old Way

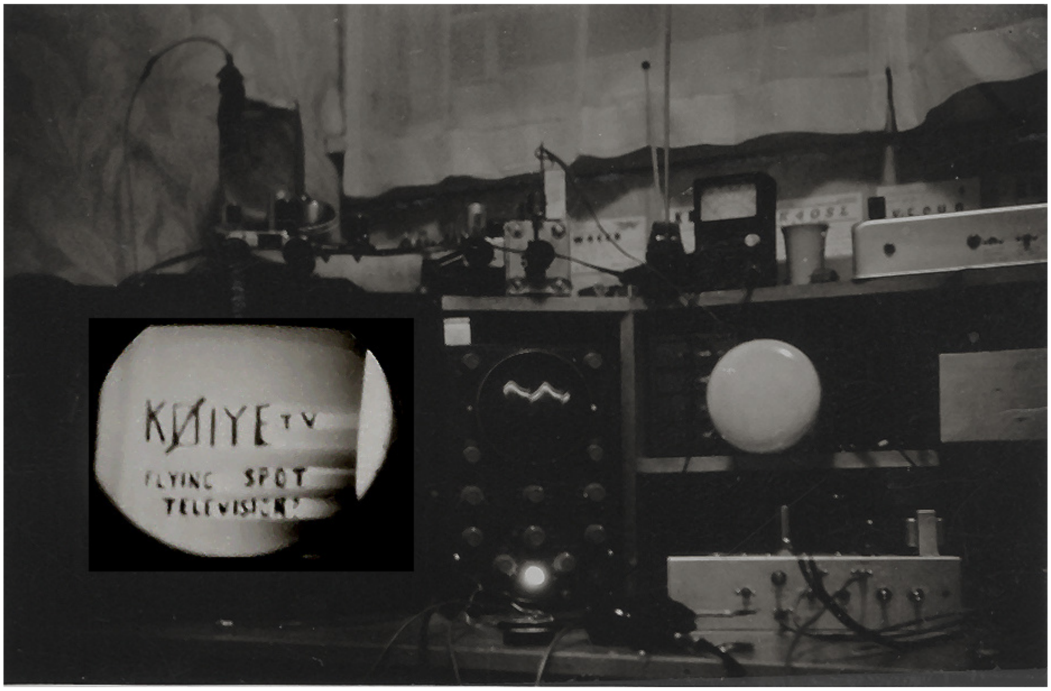

First, a glimpse of modern amateur TV

I was inspired to write this article for my local newsletter by Jim Andrews, WAØNHD. At our August 2003 ham club meeting he gave a wonderful hour-long presentation on modern Amateur TV. He showed us oscilloscope and spectrum analyzer images of audio, RF, and video waveforms and block diagrams of all the required black boxes. At the end of his presentation he included a run-down on the available commercial ham TV equipment. It turns out that you can put together a first-rate ATV station for under $1000. Less than that if you already own a suitable camera, antenna, etc. If you plan to support your local emergency service and televise forest fires, floods and riots, then a station like Jim's is what you need. On the other hand, if you just like to play around with circuits like I do, then building something from scratch may be more satisfying.

Homebuilt ham TV

Paradoxically, one of the best attributes of the old days was our relative poverty and the lower level of technology. Many services and gizmos that are routine today existed 50 years ago but were rare or unaffordable. Long distance telephony, walkie-talkies, RTTY (radio teletype), and TV cameras are obvious examples. Ham radio allowed us high school kids to play with these toys decades before they were cheap or even available to ordinary adults. Since our toys were the latest technology, we were extremely excited about them. If you show a TV camera to modern kids, they fall asleep.

Television fascinated me as much as short wave radio. So after I had a working HF station, I wanted to get on TV. In the 1950s, the hard part of ham TV was the camera. The cheapest way to get one was to buy a WWII Navy surplus flying bomb camera. The Navy built radio-controlled flying bombs that could crash into enemy ships, Kamikaze-style. After much searching I was finally able to buy a camera with its huge iconoscope camera tube. Unfortunately, by the late 1950s finding an iconoscope that still worked was difficult and mine didn't. In contrast, TV monitors were easy to get. I toured the TV repair shops and bought old sets for a few dollars that their owners didn't want to pay to have repaired.

A flying spot camera

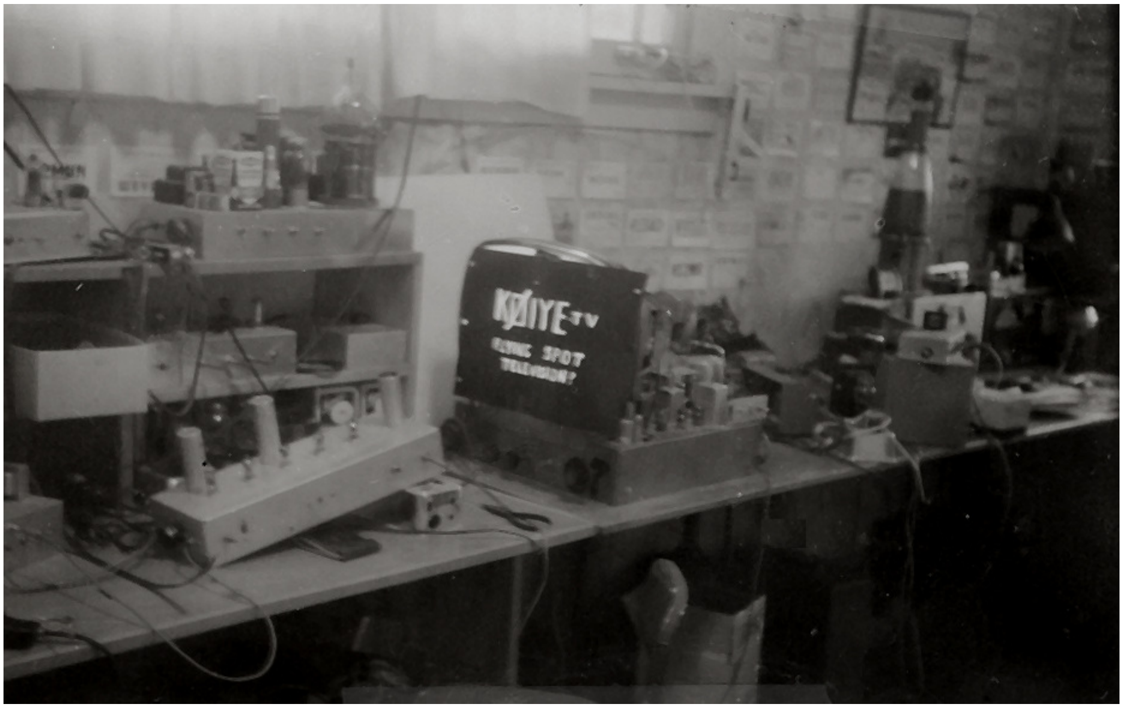

Since TV camera tubes were out of reach, I resorted to using a TV set as a scanner. Large paper or grease-pencil transparencies were taped to the TV tube. A 914 photo-multiplier tube was mounted on the tilted aluminum box on the left. The photo tube "looked" at the light from a blank TV picture raster passing behind the slide. The flying spot of light scanned the slide, one line at time. It took five stages of amplification to boost the signal to the required few volts. I usually used the retrace blanking pulses from the TV to make crude synchronization pulses. The pulses were combined with the varying light signal to make a complete TV signal. I wired a polarity switch on my combiner circuit so my slides could be either black on white or vice versa.

The TV signal was relatively high frequency and was easily separated from the constant background lighting signal. Consequently it wasn't necessary to operate in the dark or enclose the scanner in a box. The crude blanking pulse sync worked, but it pulled the image down and to the left. The easy way to get real sync pulses was to receive channel 4 from Denver, 30 miles away, then remove the picture. My light signal was recombined with the channel 4 pulses then sent on to the monitor.

Fun with flying spots

Obviously we couldn't televise forest fires with this scanner, but we played with it in other ways. When you're in high school, silliness can be great fun. Aside from televising test patterns, I liked to do silhouette hand puppets and rude finger gestures. My friends and I used to draw up transparencies of signs like, "Help! I'm trapped in your TV set" or we would draw single frame cartoons and slides showing supposedly humorous TV ads.

Synchronizing the picture to channel 4 had a more nefarious advantage. I could broadcast images and superimpose them on top of real channel 4 broadcasts. For example, I had a tiny cutout of a vulture that I could place on David Brinkley's shoulder during the NBC Huntley-Brinkley evening news. Alternatively, a giant black silhouette of a hand might slide into the picture and tickle him under the chin.

Normally I just broadcast this entertainment around the house. However, the kids across the street were interested so I thought, "It's only a few milliwatts. I'll just broadcast it over there on channel 3, which happened to be vacant in our area. It's 200 feet instead of the legal 50 feet maximum, but beyond that, how far can it go?" My little slide and puppet show for the kids worked great. Nothing more came of "The Frank Show" until several years later when Jim Synder, WØUR, was visiting my shack. I told him about my former flying spot project and he said, "So you were the culprit!" He described how his brother was watching TV in their living room across town from my house, about 5 miles away. Suddenly he began hollering for Jim to come look at the TV. Jim sprinted into the room just in time to see the word "HELP!" written on top of channel 4. A few moments later the mysterious signal disappeared. I dug out my old slides for Jim and he thought he recognized the perpetrator. This was interesting because I never deliberately broadcast on top of channel 4 using an outside antenna. Even in high school, I wasn't that reckless. On that other hand, if my lower sideband was on channel 3, the upper sideband would have been on ... channel 4.

Broadcasting properly on 420 MHz UHF

Unfortunately, being a poorly equipped kid, I was unable to generate and receive a signal on 70 CM over any distance. I scratch-built a little 420 MHz transmitter that appeared to work OK. That is, a 50 ohm ½ watt resistor on the output got hot and all the stages "dipped" when tuned. For all I could tell, it was working. At that time I had never even heard of filtering the output with a resonant cavity filter to get rid of the lower sideband, so the vestigial sideband issue was blissfully ignored. That is probably why Jim saw it across town. I also built an alleged 420 MHz converter that received my own signal, although I had no real knowledge of what frequency I was actually sending and receiving. Another barrier was that none of my ham friends were interested in putting up UHF antennas, building converters, and all that. They all were too busy with DX, building kilowatt finals, walkie-talkies, RTTY, and so on. We all had different interests and high school was a busy time.

As you can see, my ham TV project wasn't a complete success. It illustrates the difficulties with homebrew VHF and UHF. To be sure you're producing a quality signal on the right frequency, you need expensive UHF test equipment. Moreover, you need a high level of craftsmanship to control the unwanted oscillations. If 10 meters is tricky, imagine getting 0.70 meters to work right! The only advantages are that you can use low QRP power levels and compensate by building small, high gain antennas. Most of the difficulty with the high HF bands happens when you try to generate high powers over one watt. In contrast, with a small rooftop antenna a few milliwatts of VHF or UHF can get you around town.

In the final analysis, my TV project was loads of fun and I learned plenty. When you scratch-build, the rewards are usually quite different from store-bought ham radio and can be quite unexpected. For example, who would have thought QRP television could work so well?

****

GETTING ON SIXTY METERS

A ham band frontier?

Every generation or so, hams have been granted a new ham band. When 30 meters first opened up, Bob Hamilton, NØRN (former KØIYF), was ready with his homebrew CW transmitter. The bad news was that he was practically the only American on the band. The good news was that 30 meters had been open for years in Europe and suddenly ***Bob was hot DX.*** Everyone was calling him. This delightful situation lasted about a year, but as soon as the commercial equipment covered 30 meters, Bob's private DX band became history. After hearing this story I wanted to try out 60 meters when it opened on July 4, 2003.

What 60 meters is

60 Meters consists of five, 2.8 KHz wide channels with center frequencies on 5532, 5348, 5368, 5373 and 5405 KHz. They may be used by General Class and above. The only mode permitted is USB SSB and 50W effective radiated power. ARRL recommends that we set the carrier frequency to 1.5 KHz below the centers of each channel and restrict the audio to between 200 and 2.6 KHz. Unfortunately this band is shared with several other services.

Modifying your present SSB rig

You may already own a 60 meter rig. Bob has a collection of transceivers and set out to put one of them onto 60M. His first try was an old Drake R4 that, with the proper crystal, supposedly "covers everything from 1.8 to 30 MHz." It turned out that the R4 could be used on any HF frequency ***except 5.0 to 5.5 MHz.*** The IF and VFO are in that range and spurious signals would result. Next he tried a Kenwood TS-180. His particular unit was missing a surface-mount, broadband filter that needed to have been assembled on a PC board at the factory. It's possible to kludge a filter, but it wouldn't be easy because there is little room for it.

In modern rigs that use VHF IF frequencies, the only barrier preventing transmitting on 5 MHz is the firmware that vetoes illegal frequencies. Bob was able to put his Kenwood TS-50 on 60M by just clipping a diode to reprogram the microprocessor. The miniature Yaesu FT-817 can be modified by tearing the unit apart and soldering and unsoldering some tiny programming jumpers on a board. The Yaesu FT-1000MP was the easiest. He turned off the unit and then turned it back on while holding down two buttons. That brought up a menu that allowed him to select frequency ranges for different regions of the world. At the bottom of the list was "general coverage." This permits operation over the entire spectrum, even on standard broadcast.

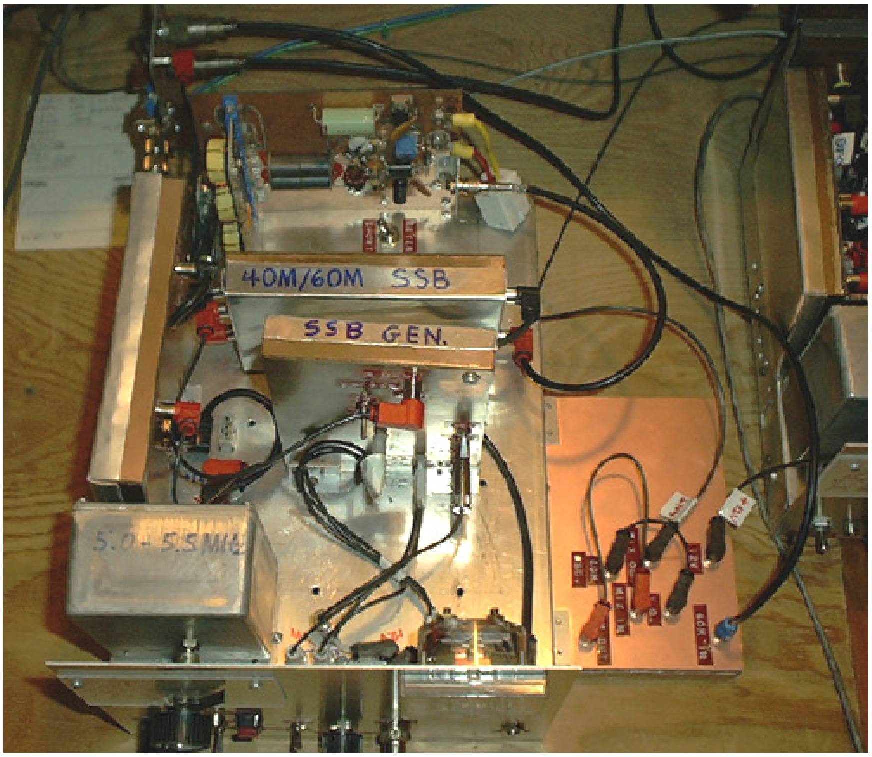

A homebrew 60 meter rig

My 60 meter transmitter and receiver converter are shown above. The copper box on the right is the receiver converter. The transmitter consists of 5 main modules: These are a 5 MHz VF0, a frequency converter for the VFO to make a 14 MHz VFO signal, a 9 MHz SSB generator, the mixer/ driver to combine 14MHz with 9 MHz SSB to make 5 MHz SSB and, at the rear, a 50 watt linear amplifier. To get on 60, I added a crystal and 60 meter filters to two of my transmitter modules and I built the receiver converter box. The converter has jumpers all over it because I eventually plan to use it on 6 and 10 meters.

July 4th 2003

I knew 60 meter SSB wouldn't be as much fun as 30 meters, but I naively hoped I might meet some eccentric homebrewers there. As midnight July 3rd approached I tweaked my homebrew 60 meter rig on a dummy load. I could hear 4 strong RTTY stations between 5.3 and 5.4 MHz that ran continuously. One of them was parked pretty much on top of channel 2. At 11 p.m., just before the band opened, I found faint SSB stations on all five new channels way down in the noise. I couldn't copy any of them, but they were definitely there under the RTTY and static. They may have been those special-licensed experimental hams or perhaps they were the Coast Guard, far, far away. At 10 minutes after midnight, some relatively loud signals appeared on channel 4. One of them was Milo, NL7SA. These Alaskan hams were gushing about being the first to try out the new band. They said they were using Alaskan CB radios, but for me they were only about Q2-S2. For many years there have been 5 MHz SSB CB channels assigned just for Alaska. I had no graceful opportunity to call them and after a few minutes they faded.

The next night I heard two Alaskan fishermen trading shrimp stew recipes. One recipe featured shrimp, chipped beef and cabbage. Yuk! The other guy said he had just motored 35 miles back from his favorite fishing spot. They never gave call signs, so I believe I was hearing an Alaskan 5 MHz SSB CB channel. After a few days, more and more stateside hams appeared during the evening hours. Channels 1 and 5 had the least RTTY interference and were the most popular. Most guys were using FT1000MPs but I also heard a Kenwood TS-930 and an ICOM 746. Each channel turned into an impromptu "net" with the loudest guy serving as net control. For example, on July 10, Bob, K9CGD, in Longmont, Colorado served as moderator on channel 5 for several hours. As for me, apparently my signal was too puny to get more than a "Sorry old man, you're too weak." For one thing, my antenna is probably inadequate. However, my rig has been successful on four other SSB bands, so I'm mystified about what's wrong.

In conclusion,

60 meters is a band inhabited by RTTY, buzzing data signals, crashing static and two or three channels of weak SSB signals. Basically, it's a whole continent of Yaesu owners trying to get their 60 meter WASs (Worked All States) sharing two or three "party line" channels. After just 7 days, some guys were already claiming as many as 28 states.

When you tune in, you will hear the two loudest guys talking about how neat 60 meters is and that they're going to sign off now so others will have a chance. And ten minutes later the same two guys will still be talking about shutting up, but not actually doing it. The other annoying category is the East Coast guys who are apparently talking to England. But of course, we can't hear England from Colorado and yet it seems impolite to interrupt, even though static is all that's on the channel.

So, what did I learn from getting on 60 meters? I learned that I better be prepared to be disappointed when I start a project and that perhaps I should have done more homework about what units might be the first on the air. I was naive to think I might be among the first. Oh, well, the Alaskan gourmet fishermen were interesting. Hmmmm ... I'm going back to 20 and 30 meter CW. As for something to build, modern ATV still looks interesting. I ought to be able to do a better job this time.

****

HOMEBREWING WITH PICs

By Frank Harris, KØIYE

Difficult, but still practical

The little ham kits advertised in QST often contain a PIC microcontroller. For example, when a modern ham says he "built a homebrew CW keyer," what he probably did was buy a kit with a pre-programmed PIC microcomputer and a PC board. There are perhaps 6 or 7 parts that he must solder. Such a kit may produce a great keyer, but it won't produce much sense of accomplishment. The guy who designed the kit has already enjoyed the homebrew fun, not the guy who buys the kit.

The PIC controllers you find in the kits look simple. They usually have 8, 14 or 20 pins in an old-fashioned DIP package. They look so innocent, how can those little things be hard to program? 20 years ago, microprocessor projects involved building an 8"x 11" PC board full of parts to complete a system. You needed a microprocessor IC, a crystal, EPROM, the RAM, A-to-D converters, D-to-A converters, a modem chip, a timer or two, perhaps some EEPROM and the peripheral interface adapter chips. These ICs were interconnected by busses of 24 different data and address lines. Whew! No wonder microprocessor homebrewing was rare!

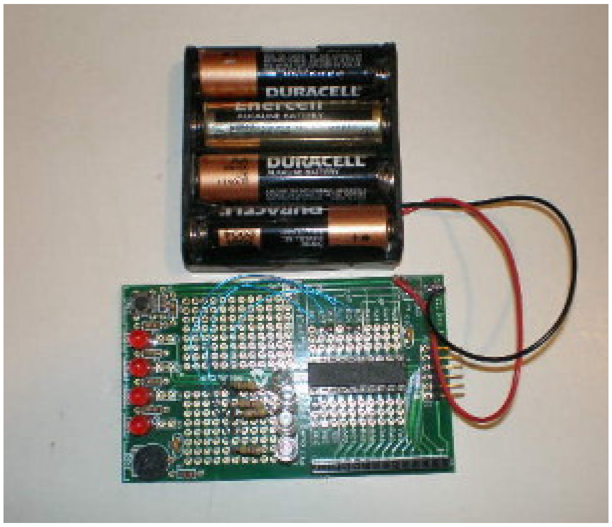

PIC16F690 development board.
A 20 pin PIC version.

The good news about PIC micros is that they do ***EVERYTHING internally – no giant PC board is needed***. Another neat feature is flash programming – 14K or more. That's a lot of ROM, really! You can build your project with its own 6 pin interface that allows you to download a new program into it in seconds. Supposedly the programming will last for 40 years, but I doubt I'll be around long enough to find out. As I recall, the old UV-erasable EPROMs were only good for 10 years.

Selecting the desired pin and feature configuration

The bad news about PICs is that programming them has big challenges: The first obstacle is turning off the wondrous capabilities that you DON'T want. This includes programming nearly every pin on the package to be assigned as a comparator input, voltage reference input, LED driver, analog input, etc, etc. Typically each pin has 3 to 5 possible roles it can serve. These capabilities often interfere with each other. If two capabilities are mutually exclusive, obviously it won't work.

Because of all the capabilities you don't need, each version of the PIC typically has a **300 page set of datasheets.** Also, whatever your application, there's a PIC processor for every imaginable job. There are literally over a hundred PIC chip variations. Each one of those is available in 5 or 6 different physical packages that are mostly tiny kinds of surface mount. A few years ago electronics parts catalogs were the size of magazines. Today catalogs like Mouser and Digi-Key are the size of the L.A. phone book with thousands of fragile pages. That's why most people use the on-line catalogs that have built-in search engines to help you find your needle in the haystack. The on-line catalogs are so huge they make the phonebook size catalogs look like abstracts.

Primitive RISC programming

Another odd problem with the PIC is the "reduced instruction set code" (RISC). RISC is often extolled as good, but I'll be darned if I see why. RISC runs very fast. It allows you to do almost nothing really quickly – up to 20 MHz! It's like programming the 8008 Intel processor from 1970. There's just one accumulator, no index register, no divide command, no multiply and no byte test. There are only two simple conditional branch commands that just skip the next instruction when the carry or zero flag is set. Therefore, every branch needs a "**goto**" following the "**skip if**" command. However, the PIC assembler is quite good. But I'm easy to please. In the bad old days I used to hand-calculate addresses and write out programs in pencil on ruled sheets. Then I Scotch- taped the sheets together to make long scrolls.

In fairness, the PIC professionals use C language programming which bundles huge chunks of RISC programming into clumps usable with a single command. ProBasic for PICs is also popular. Unfortunately, C or Basic programming can't do all the unique stuff that you need for your project. Therefore, you still have to write some of your own subroutines using RISC machine language. It was obvious to me that I had to master RISC anyway, so my first project was done 100% with RISC.

PIC development systems

PIC development systems consist of a little capsule the size of a playing card that plugs into a USB port. The capsule then plugs into the similar sized sample circuit board on which the PIC is mounted. Once you get launched, you can build the flash programming right into your own project board and you may no longer need the sample development board. Of course if you do that, you may lose the ability to step through programs, read internal registers, and so on. I bought two development systems for about $25 each, one for the PIC16F690 and the other for the more capable PIC16F887.

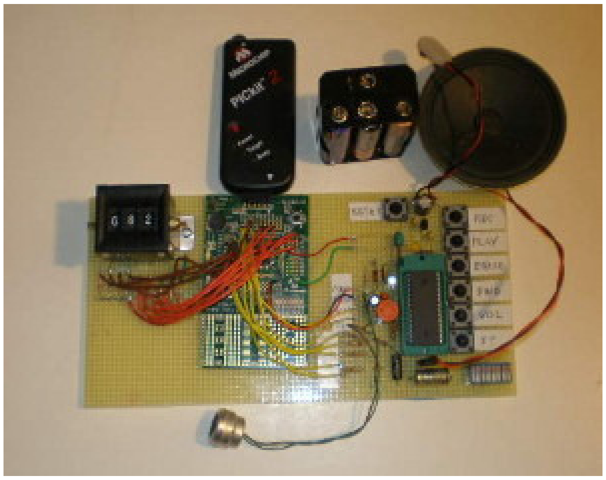

A PIC16F887 development board wired into a speech chip programmer. The black plastic programmer unit is plugged into the development board.

The little development boards come with a disk containing zillions of data files, help instructions and example programs. It also includes an editor, a debugger, the assembler and the PIC ROM programmer. It does NOT contain a C or Basic compiler. Compilers range from $150 to $500 or more, depending on which PIC families they are designed for.

The Debugger Program

In theory, the debugger program combines all your functional programs, such as the editor, assembler and even 3rd party compilers, so that they may be operated together "seamlessly." If you can make it work on your project, I imagine it would be wonderful. You could step through programs, watch constants change, set breakpoints, read internal registers, etc., etc. Sadly, I could never get mine to work. Fortunately, all the parts like the compiler, editor, programmer and so on, can all be called independently, so I pressed on without the debugger. Basically I use lots of trial and error and intuition.

A talking dynamometer

My only PIC project so far has been a talking rope tension meter, (a.k.a, dynamometer), for the rescue group. My original plan was to use the PIC to run a voice synthesizer chip, something like the old "Speak and Spell" Texas Instruments technology. Ironically, the only programmable speech chip I could buy on the Internet was a PIC micro that had been programmed to be a speech synthesizer. Since PICs include multiple oscillators, a modem, D-to-A converters, etc, I shouldn't have been surprised to find that my programmable speech synthesizer chip was just a programmed PIC microcontroller. Yes, they really are versatile.

It took me weeks to get the serial interface to work. The instructions didn't mention that the speech synthesizer needed both receive AND transmit lines. There is apparently a "clear to send" confirmation that had to be sent back that wasn't mentioned in the data sheets. Anyway, I soon had it working as well as it was going to. Sadly, I was the only one who could understand what it was saying. Not good for rescue situations. "Did it say let go of the rope?" ... AHHHHHHHhhhhhhhh … Splash! Apparently not.

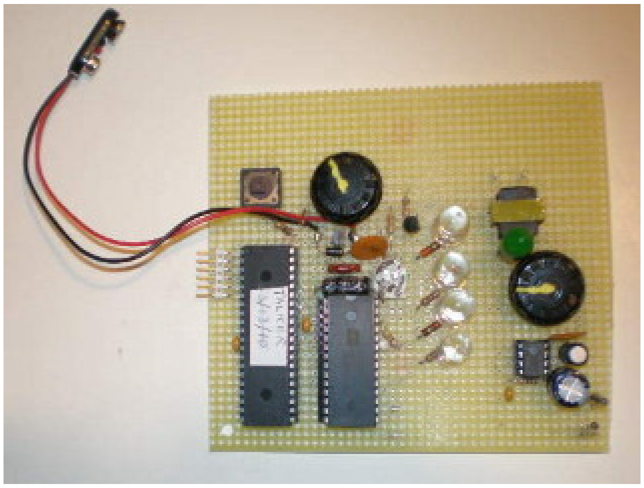

A PIC16F887 breadboard for developing software to drive the speech chip. Note the 6 pin connector on the left used to program the large, 40 pin PIC microcontroller.

My next attempt used an obsolete American speech recording chip that works well, but has a limited audio frequency response (ISD25120, 1.7 KHz). The speech chip can be hard to understand, but by putting it through a severe hi-pass audio filter, I managed to accentuate its high frequencies and compensate pretty well. But when I tried to buy more speech chips, they were extinct. I finally found a newer chip, the ISD17240, that is superior in every way – except for its modern Serial Peripheral Interface (SPI). Programming that sucker is way beyond what mortals can do with just data sheets and trial and error. Not only must I figure out how to get the PIC to send the instructions and receive the responses, the speech chip has its own complementary SPI digital nightmare I have to decipher. No luck so far. Oh, and by the way, the PIC "weak pull-up" feature for digital input lines is so weak, it's prone to the slightest noise – use pull up resistors!

In summary,

PIC homebrewing is NOT for people with limited patience and time. It turns out that there is a good reason for buying those little pre-programmed PIC kits – you'll be done in less than 6 months! On the other hand, if you keep the project simple, such as just directly driving LEDs and reading switches, it's not so bad. But when you have to talk to other PIC processors or peripheral chips, the fun begins. …. My next project is going to be a CW keyer because it's basically lights and switches. That's where I should have started.

****

CLASS D AUDIO AMPLIFIERS

By Frank Harris, KØIYE

Big hi-fi sound, high efficiency, low voltage

Class D amplifiers chop an analog sinewave signal into a pulse-width-modulated stream of rectangular pulses. In other words, the density of high frequency (400 KHz) rectangular pulses varies with time and represents the amplitude of the wavy simple analog signal. When applied to audio amplification, a chopped series of high frequency square pulses can drive a 4 ohm loudspeaker with surprisingly good fidelity. Because the amplifier output only has two states, the highest voltage and zero voltage, there is no time spent half on and half off. Consequently there is little power dissipated in the chip. A tiny chip can deliver big power, 1 to 3 watts or more. And it does all this with just 5 volts or less.

Why you might want to use one, or not

Suppose you are building a small, portable battery-powered receiver to go with your QRP transmitter and want to use 6 or 9 volt batteries. If you look at the data sheets for the good old LM388 or LM386 analog amplifiers, you'll see they are rated to work from 5 to 15 volts. If you are listening with sensitive earphones, a low voltage supply will work fine. On the other hand, if you want audio that's both Hi-Fi and LOUD, then you'll discover you need big voltage to go with your 4 inch or larger speaker. Yes, low voltage can be loud with analog chips using just 5 volts and, but the distortion is intolerable.

It is theoretically possible to use op-amps and comparators to build your own Class D audio amplifier. But when you diagram what has to be done, you quickly discover that you would need a huge array of components, a big circuit board and lots of trial and error. Class D amplifiers are the reason other people's tiny cellphones can be so annoying - they're both loud AND hi-fi and they accomplish this with tiny low voltage batteries.

EMI – digital hash in your receiver

If you've ever home-built a sensitive ham receiver, you probably discovered that a pulse-width-modulated regulated power supply or even a simple digital display, will make noise in your receiver that destroyed much of the sensitivity you worked so hard to achieve. The class D audio chips I used, **Maxim MAX9719** and **MAX9759**, have "spread spectrum" PWM systems to reduce EMI. Moreover, you have your choice of 3 different sample frequencies that supposedly make it less likely that the hash will appear outside your desired ham band. Unfortunately, the wires from this board up to my speaker radiated VHF RF noise like crazy. Since my project needed to work around VHF hand-helds, I was relieved to discover that I could silence the racket by stringing 5 large ferrite beads on the speaker wires. Unfortunately, the noise was still unacceptable for use around my HF receiver. Sure, I could still hear the hams over the hiss, but that's not good enough. Oh, well. At least Class D worked for my immediate application.

And now more bad news

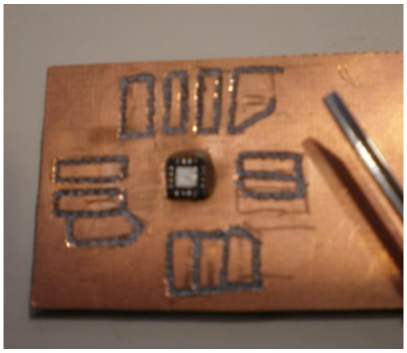

Modern chips are often surface mount and really tiny. Even surface mount with projecting leads are becoming scarce. The tool is the tip of a 1.5 mm wood carving gouge used to cut the pads.

You have probably already noticed that the modern world does not want us to build cool stuff in our basements. The complexity is bad enough. But modern surface mount is providing the coup-de-grace to homebrewers like us. In the old days, chips came in DIP packages with big pins. They have 0.1 inch spacing that we can plug into sockets. Even the early surface mounts weren't impossible to solder. Yes, 0.50" spacing was a little tight, but the leads stuck out from the body of the chip and most of us figured out ways of soldering them. For example, you can buy tiny adapter circuit boards to accept the “SO” surface mount chips. The periphery of the little board has pads big enough to accept reasonable sized wires. I’ve even managed to successfully solder SOs with 0.020” spacing onto the little boards.

Today many of the new chips aren’t available in DIPs or even SOs. When I ordered my Class D amplifiers, I studied the 5 or 6 versions of surface mount and ordered the biggest one. It had no leads that stuck out, but the contacts were at least on the periphery of the tiny squares and I hoped I could solder wires to those.

I was shocked when I opened the package and found tiny square black flakes 4 mm on a side. Using a microscope, I could read the labels on the chips. Yes, those were the “big” versions I had ordered. Sure glad I didn’t get the little ones! The 16 contacts were positioned around the circumference of the chip, but they were just metallic dots. My first reaction was that the project was over. But then I thought: These chips are too cheap to bother sending back. I might as well have a go at it.

How to solder chips you can barely see

First you must give up caffeine and booze. If you have benign familial tremor like some members of my family, I’m sorry, you’re already screwed. I talked about this at the local ham club meeting and I was told that the professional way to solder surface mount is using low temperature solder available in a paste. The paste is applied to the circuit board using a syringe with an extremely fine needle. The chip is then lowered onto its pads and heated with a heat gun. I haven’t tried this, but because of the small size, it isn’t going to be easy, fine gauge needle or not.

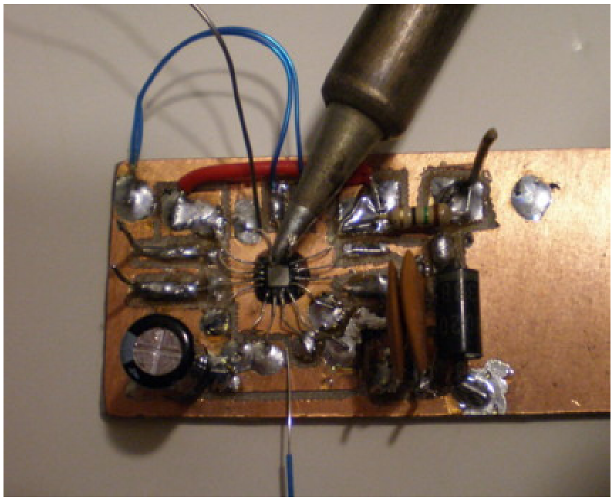

Sharpen your soldering tip and set the temperature to 600 degrees. I use 24 gauge solder and stripped wire-wrap wire to make the connections. Bring your iron in radially as shown and solder quickly before you fry the chip.

Even if I had the solder paste and needle applicator, I would still need the PC board with the tiny traces and pads. I use two methods of making PC boards. For big DIP chips I use perf-boards with pre-drilled holes spaced on 0.1 centers. RF circuits need real printed-style circuit boards to reduce lead inductance. I make crude but effective PC boards using a fine wood gouge to carve off the copper between the traces. I start by gluing the chip down to the PC board with 5 minute Epoxy. Tiny hand-held gouges intended for wood carving are hard to find. I bought mine from www.TraditionalWoodworker.com. “Two Cherries” brand, micro-carving tools are the smallest I’ve found. They have cupped ends as small as 1mm. However, a 1/8” (3mm) V-parting gouge can be made to work.

Next, you'll need a really fine-tipped, temperature controlled soldering iron. I use a Weller model WES51. I filed the tip to make it as sharp as possible. Also, the solder needs to be the thinnest you can find. I use 0.022" (24 gauge) diameter solder. I use stripped wire-wrap wire, about 30 gauge, to contact the chip pads. I solder the easy end of the jumper first onto the big pads and then bend the tiny wire down into position over the tiny dot. You have to solder each contact as swiftly and deftly as possible. You might want to practice before you ruin too many chips. The higher the temperature, the easier it is to make the contact, but the more likely you'll kill the chip – a bit of a dilemma there. 600 degrees seems to be a good compromise.

My class D amplifier board

Here's the final product installed in my talking meter project. One interesting observation is that over half the sound I get out of my box depends on the resonance chamber behind the speaker. That empty space in the speaker cabinet isn't poor planning - it's vital. One of my major challenges in this project was making it loud enough. But the first time it was used on a practice mission, my teammates demanded that I install a volume control. Success!

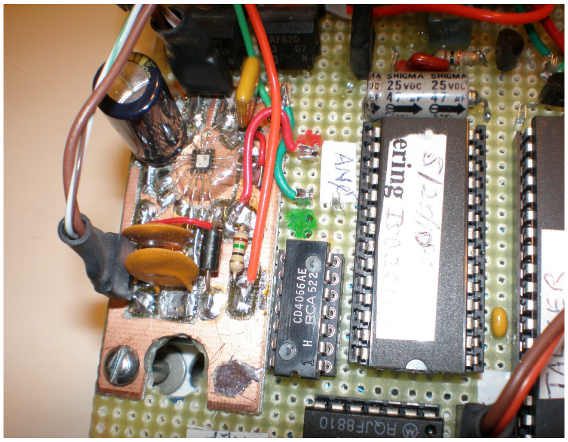

I like to pot the working chip with a cover of clear epoxy to protect it.

In summary, I was totally amazed that I was able to make this work. Who says we old guys have lost the touch? It rarely hurts to try.

* * *

BRIAN R. PAGE, N4TRB

CRYSTAL RADIO TO THE RESCUE

There was a time when crystal radios played a role in national defense. That time was the mid-1950s, when the United States was newly finding itself under the shadow of a Soviet H-bomb threat.

On 1 November 1952, the U. S. had detonated the first full-scale hydrogen bomb. The Soviet Union quickly followed with its first H-bomb test in 1953. These fusion bombs drastically raised the stakes in the nuclear arms race. The 1952 U. S. test, for instance, released over 450 times the energy of the puny, by comparison, fission bomb that was dropped on Nagasaki, Japan.

That bomb hastened the end of World War II, but the advent of the H-bomb raised the specter of a truly civilization-ending conflict. The end of civilization means the end of commercial electrical power. So what technology remains once the commercial mains are dead and all batteries exhausted?

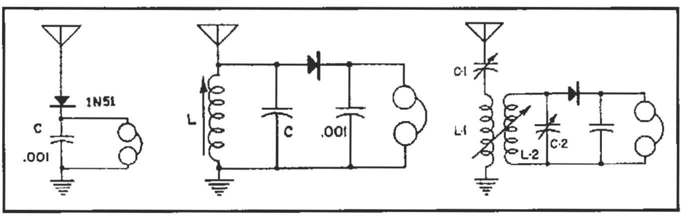

Fig. 1. G.E.'s sample crystal set circuits.

I've always been fascinated by the crystal radio, that most elemental radio circuit, requiring no source of power other than that which it teases from the ether. With a crystal set on the receiving end, the broadcast stations themselves supply the power. There is no need for commercial power or even batteries. Thus, the good folks at General Electric, in their bi-monthly *G-E Ham News,* publicized "Operation Crystal" to encourage the development of crystal radio receivers for use in times of national crisis.

Operation Crystal was the cover story for the January-February 1955 issue of *G-E Ham News*. The project challenged "old-timer and novice alike" to "make a valuable contribution to the public welfare." The story inside fleshed out the scenario: "Disaster. . .no radio, no newspaper, no telephone. . .hurricane? Explosion? A-bomb? H-bomb? How many bombs? What should you do to save yourself and your family?"

The writer sketches a situation that day-by-day becomes more desperate. "Is help on the way? Or are worse disaster areas taking up all the relief? Should you start hiking? Which way? North? South? East? West? Will there be more bombs?"

Finally, of course, our hero the crystal set rides to the rescue: "Superhuman effort has kept one broadcasting station on the air, perhaps on the Conelrad frequency. But you can't hear the official reports and advice without power or batteries." CONELRAD (Control of Electromagnetic Radiation) was the predecessor to the Emergency Broadcast System. Developed in 1951, CONELRAD was an elaborate system to alert and inform the public during disaster situations (originally anticipated to be atomic bomb attacks from fleets of Soviet bombers) using a network of radio stations.

Upon notification, all commercial broadcast stations would leave the air except for select AM stations on 640 kHz and 1240 kHz. These stations were to broadcast the Civil Defense announcements and cycle on and off the air. Having all active stations on just two frequencies, it was hoped, would prevent Soviet bombers from using radio direction finding to deliver nuclear death and destruction. Remember, this was *way* before GPS.

Since, in this picture, the commercial mains are down, and batteries are dead, ordinary radios are useless, the *Ham News* article continued with "What is the answer? You, the radio amateur have the answer—right in your junk box. You guessed it. A crystal diode detector."

Thence followed a selection of really simple crystal radio circuits (Figure 1) along with a challenge to the readers: "*G-E Ham News* will award $10 tube certificates for the three most

1717 TIDEWELL TRACE, LAWRENCEVILLE, GA 30043
BRIAN@GLADLYLEARN.COM

outstanding emergency crystal diode designs submitted for publication for each issue during 1955." And readers responded to such an extent that "Larry Lighthouse," the principal editor of *Ham News* had to enlist an editorial assistant, "Danny Diode" to handle the operation.

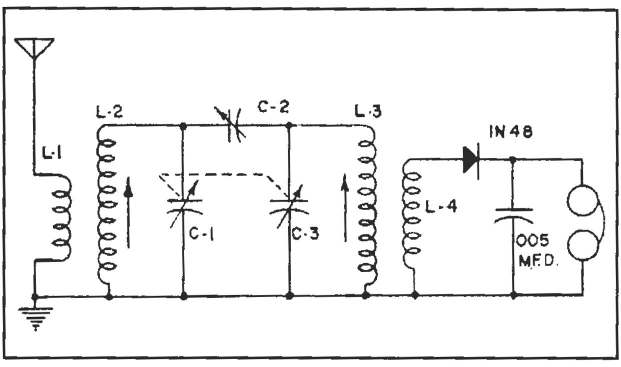

Fig. 2.The knob twister's special.

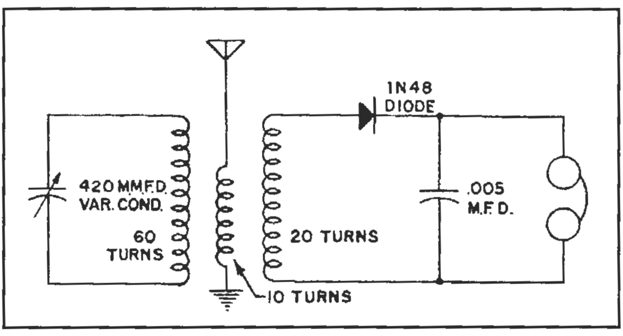

Fig. 3. Drinking glass receiver.

The March-April 1955 issue featured the first three winners. These little rigs carried imaginative names, "Knob Twister's Special," "Furschlugginer Model," and "Drinking Glass Receiver." The "Knob Twister's Special" (Figure 2) came from William Patzer, W8RWX, and featured five variable adjustments (although two were ganged). A 1N48 or 1N52 diode was specified for rectification and two vari-loop-stick coils (L2 and L3) were coupled with L1 and L4 via 30 turns of No. 36 wire wound over the primaries. C1 and C3 were ganged 365 pF variables. C2 was a 3-30 pF compression trimmer.

The Drinking Glass Receiver demonstrated the ingenuity of eighth-grade student, Jack Lambuth. As described by Danny Diode, "the 10-turn antenna and the 20-turn detector coils are wound over the 60-turn coil with a layer of plastic electrician's or scotch tape" on a 2-inch water glass form using No. 26 wire (Figure 3).

The third installment of Operation Crystal shifted back to simple circuits, one of which required no ground connection (Figure 4) and another that was the essence of simplicity (Figure 5), obviously from someone who had only a meager junk box. As Danny Diode commented, the simple set "makes up in volume and simplicity what it lacks in selectivity." It took advantage of the fact that CONELRAD stations broadcast on only two frequencies. Thus, In an actual emergency, interference probably wouldn't occur.

The simple circuits of the third round must have inspired the *Ham News* readers to new heights of creativity, because the September-October issue carried a dual-antenna monster with a triple-ganged variable capacitor (Figure 6). Note how this circuit uses a tuned ground

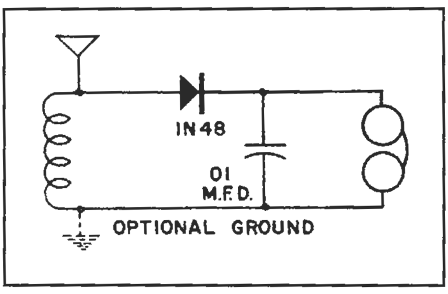

Fig. 4. No ground required!

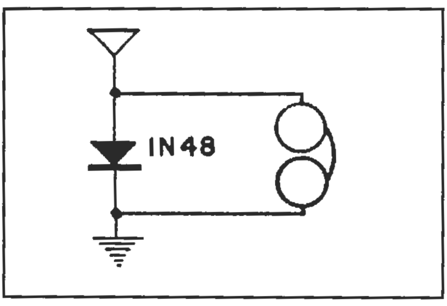

Fig. 5. For those with a meager junk box.

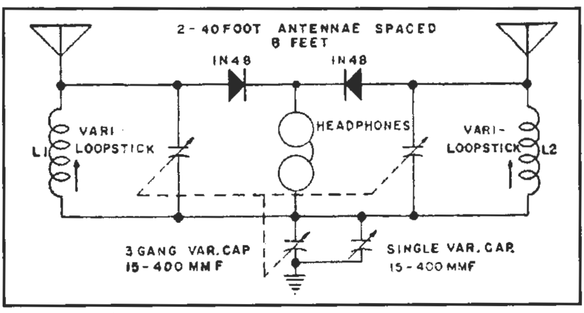
2 - 40 FOOT ANTENNAE SPACED 8 FEET
IN48
IN48
VARI LOOPSTICK
L1
HEADPHONES
VARI-LOOPSTICK
L2
3 GANG VAR. CAP 15 - 400 MMF
SINGLE VAR. CAP. 15 - 400 MMF

Fig. 6. Dual antenna monster.

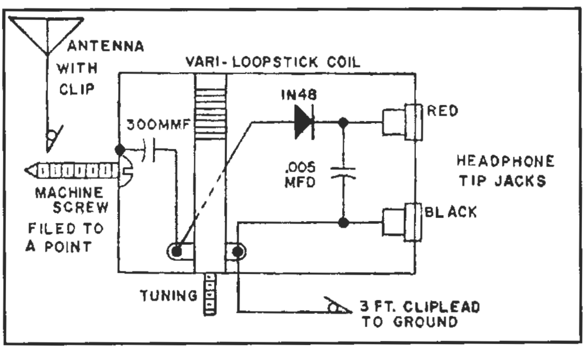
ANTENNA WITH CLIP
VARI - LOOPSTICK COIL
IN48
300MMF
RED
.005 MFD
HEADPHONE TIP JACKS
MACHINE SCREW FILED TO A POINT
BLACK
TUNING
3 FT. CLIPLEAD TO GROUND

Fig. 7. Receiver doubles as signal tracer.

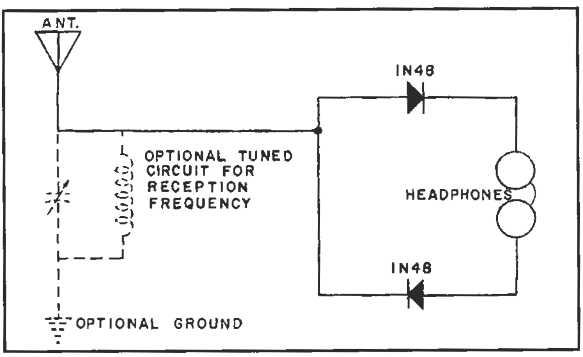
ANT.
IN48
OPTIONAL TUNED CIRCUIT FOR RECEPTION FREQUENCY
HEADPHONES
IN48
OPTIONAL GROUND

Fig. 8. Ground unnecessary—tuner optional!

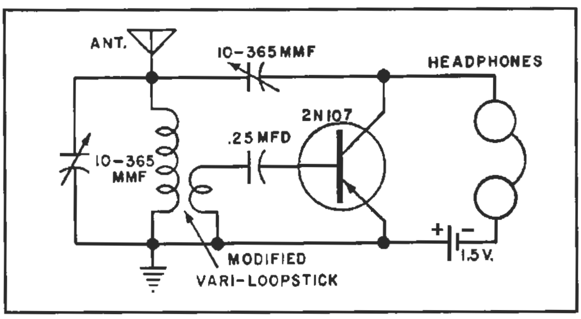
ANT.
10-365MMF
HEADPHONES
2N107
.25 MFD
10-365 MMF
MODIFIED VARI-LOOPSTICK
1.5V.

Fig. 9. Transistorized regenerative circuit.

connection. And for his entry, L. Walty of Walla Walla, Washington created a crystal set that could do double duty. First, of course, it was a crystal set; but while waiting for a national disaster, Walty's rig could be used as a signal tracer (Figure 7). You gotta give these guys credit for ingenuity.

The dual antennas model, in turn, must have inspired other dual approaches because the following issue contained sets with dual detectors. Two were fairly standard full-wave rectification circuits. The third was an inventive circuit by R. J. Baker, W8JIA, which contained no ground connection and only an optional tuner (Figure 8). The caption includes the comment, "Several of the local hams glanced at the circuit and told me, 'Impossible!!, There's no ground return path'." Had Danny Diode not vouched for the efficacy of this circuit, I'm afraid I'd side with the local hams who expressed their skepticism.

The final sets of Operation Crystal appeared in the March-April 1956 issue of *G-E Ham News* and these three really broke new ground. Technically, the rigs violated a basic rule of the contest because each used battery power. Quite significantly, the *Ham News* publishers in GE's Electronic Tube Department had then introduced the brand new GE 2N107 hobbyist transistor. One circuit used a conventional detector circuit and added a one-transistor amplifier driven with 3 volts of battery power. A variation replaced the diode detector with a second 2N107 to fashion a two-transistor "crystal" set.

The third transistor set incorporated a simple and elegant regenerative circuit (Figure 9). Powered by a 1.5 volt battery and using two 365 pF variable capacitors for tuning along with high-impedance headphones, the little rig added eight turns of wire around a standard loop stick inductor to feed the

(continued on page 47)

CRYSTAL, *cont. from page 44*

base of the 2N107.

Fortunately, the national disasters envisioned by Larry Lighthouse never came to pass. But the contest for no-power receivers stimulated the creativity of many hams who developed interesting twists on a fundamental circuit that had been in use since the dawn of radio communications.

This story of Operation Crystal contains an interesting postscript. Several times I alluded to the principal writer of *G-E Ham News*, an individual using the pen name of Larry Lighthouse (and probably Danny Diode, too). Old Timers *might* remember Larry Lighthouse, but they *must* certainly remember "Hashafisti Scratchi" the *CQ* magazine columnist famous for his humor and tall tales.

Hashafisti and Larry Lighthouse were one and the same, both creations of George Henry Floyd, Jr., WA4DGA (ex-W2RYT). Floyd was an engineer and executive at General Electric. He was born in 1917 and just recently passed away on 22 November 2008. Floyd was 91.

The complete *G-E Ham News* issues featured in this article may be found on the writer's web site at: http://n4trb.com/AmateurRadio/GE_HamNews/ge_ham_news.htm.

Radio a cristallo e dintorni - Le Radio di Sophie - Crystal sets

Radio a cristallo doppio circuito per O.M.

Un progetto di Luciano Loria

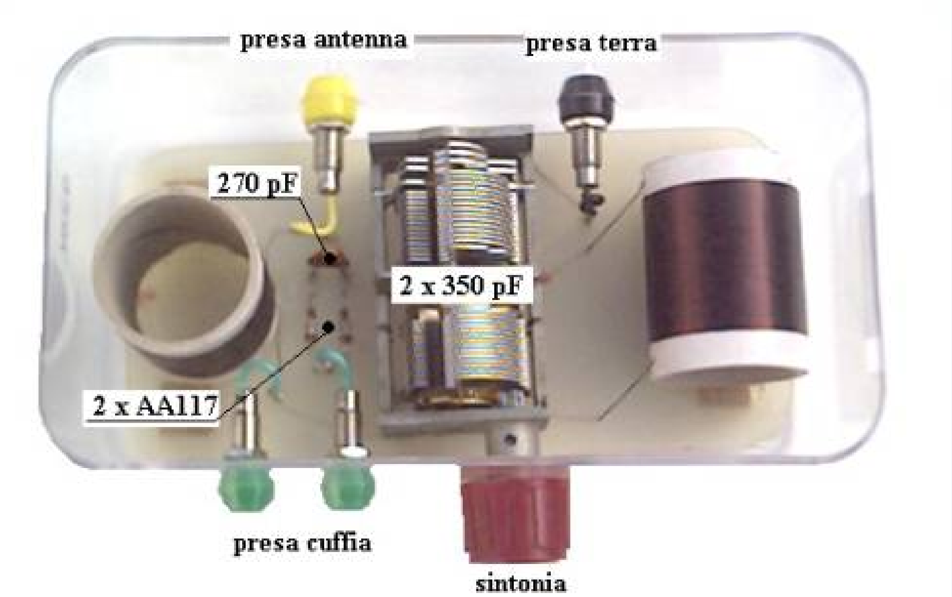

Il circuito più semplice, ed anche molto efficace, per la realizzazione di una radio a cristallo, consiste di un circuito LC accordato sulla frequenza da ricevere e da un rivelatore (cristallo di galena, diodo al germanio).
Vediamo cosa accade se questo circuito si raddoppia: due bobine, due condensatori variabili, due diodi.

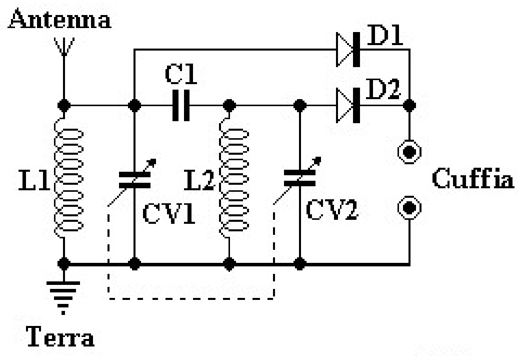

Materiale occorrente

L1; L2 = 90 spire filo rame smaltato Ø 0,3 mm
su tubo isolante Ø 3 cm
CV1; CV2 = variabile in aria 350 + 350 pF
C1 = condensatore ceramico 220÷330 pF
D1; D2 = diodo al germanio (AA117)

Schema elettrico

Nello schema elettrico notiamo che L1, CV1, D1 costituiscono un primo circuito di sintonia e di rivelazione, così come L2, CV2, D2 che costituiscono il secondo circuito; questi sono accoppiati tramite il condensatore ceramico C1 da 220/330 pF. Per ottenere risultati apprezzabili, ed una migliore selettività*, occorre che le due bobine siano montate con i rispettivi assi ortogonali fra loro (vedere foto).

* La selettività dipende anche da C1, per aumentarla si può diminuire la capacità fino a circa 33pF.

NOTE COSTRUTTIVE

Evitare, per la costruzione, l’uso di telai metallici, anche le bobine non vanno sistemate troppo vicine al variabile.

Il condensatore variabile ha due identiche sezioni comandate da un alberino unico; L1 e L2 sono avvolte su tubo di cartone di derivazione casalinga, infatti, si tratta del tubo dove sono avvolte le pellicole trasparenti o d’alluminio per cucina.

Il contenitore trasparente, sempre nell’ambito di un riciclaggio intelligente dei materiali che si trovano in ogni casa, è l’astuccio di certe famose praline al cioccolato (nella foto è visto senza il coperchio).

Naturalmente occorrono, anche: una buona antenna filare esterna; una presa di terra efficace; una cuffia ad alta impedenza; in tal caso l’ascolto, della stazione locale in OM, è garantito e risulta migliorato rispetto a quello ottenuto con il circuito base.

luciano.loria@gmail.com

Torna alla Pagina delle radio a cristallo

Un ottimo ricevitore FM a cristallo

Carlo Bramanti

Vi comunico gli ultimi risultati che ho raggiunto dopo molte intense ricerche per la miglior soluzione per la galena FM. La J antenna andava egregiamente, ma sopportava solo un collegamento brevissimo. Quando mi sono deciso di appendere verticalmente un dipolo di piattina, invece del classico posizionamento di detto, ho scoperto un mondo nuovo. Funziona egregiamente anche con una discesa, in piattina, di 6,5 metri. Solo in quel modo sono riuscito ad ascoltare costantemente e confortevolmente anche senza amplificazione. Ma non mi sono fermato lì: vista la forte corrente che si ricava da un diodo collegato al dipolo (oltre 50 microampere) ne ho usato uno per alimentare un EPAD. Ciò vale per l’FM ed anche per il Kit che ho fatto per la SANDIT.

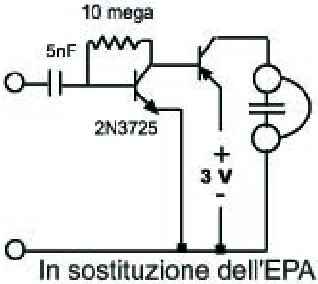
10 mega
5nF
2N3725
+
3 V
-

In sostituzione dell'EPAD: fa fortissimo ma occorre una pila

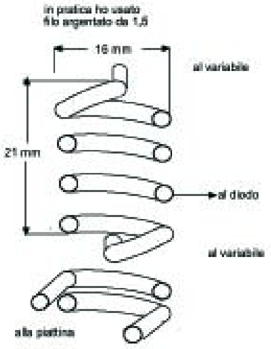
in pratica ho usato filo argentato da 1,5
16 mm
al variabile
21 mm
al diodo
al variabile
alla piattina

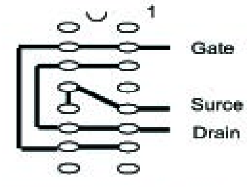
1
Gate
Surce
Drain

ALD110800vista dal sotto

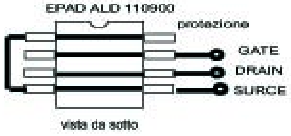
EPAD ALD 110900
protezione
GATE
DRAIN
SURCE
vista da sotto

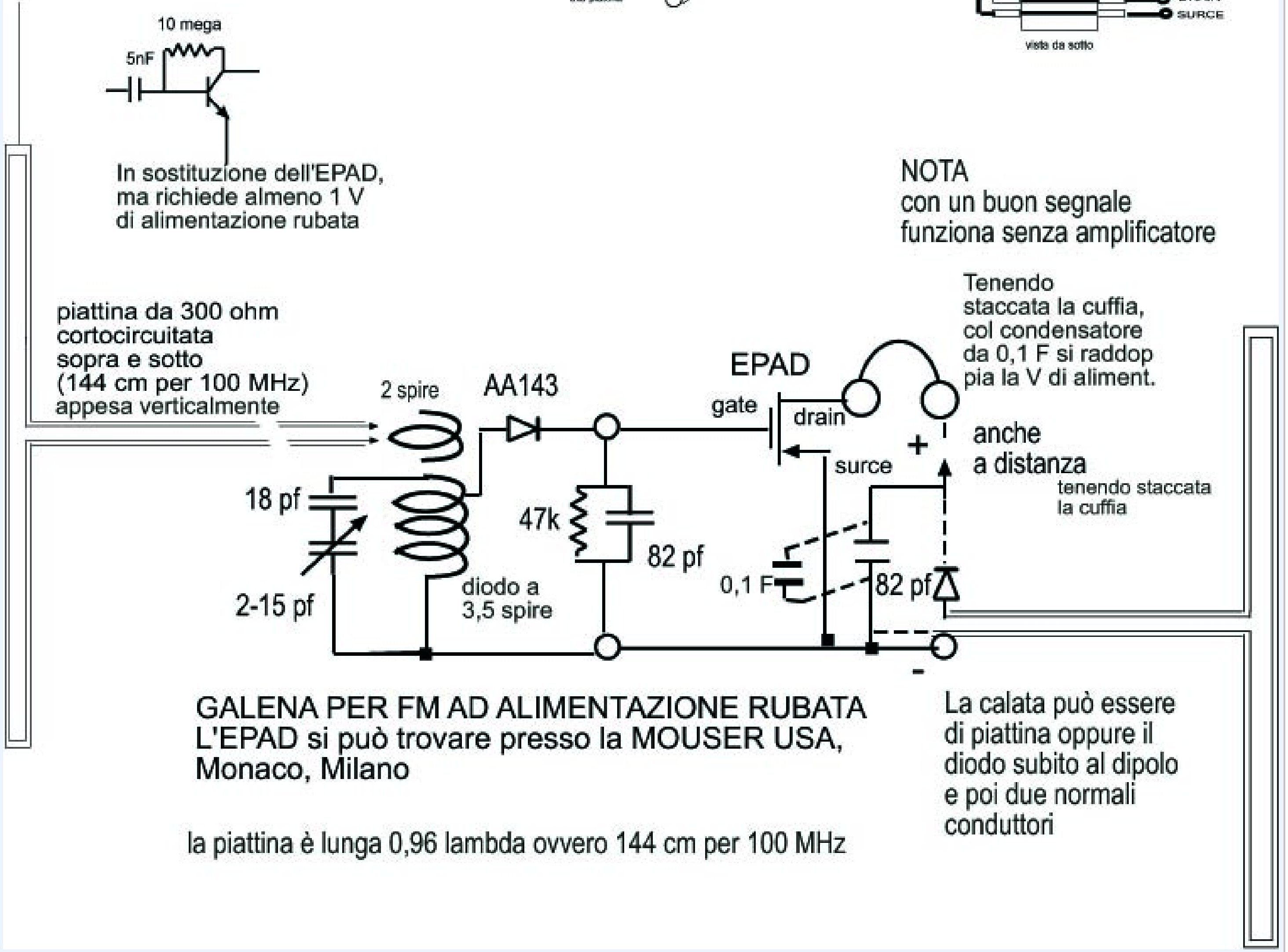
10 mega
5nF
2 spire
AA143
EPAD
gate
drain
surce
+
18 pf
2-15 pf
47k
82 pf
diodo a
3,5 spire
0,1 F
82 pf
-

In sostituzione dell'EPAD, ma richiede almeno 1 V di alimentazione rubata

NOTA
con un buon segnale
funziona senza amplificatore

piattina da 300 ohm cortocircuitata sopra e sotto (144 cm per 100 MHz) appesa verticalmente

Tenendo staccata la cuffia, col condensatore da 0,1 F si raddoppia la V di aliment.

anche a distanza tenendo staccata la cuffia

GALENA PER FM AD ALIMENTAZIONE RUBATA

L'EPAD si può trovare presso la MOUSER USA, Monaco, Milano

La calata può essere di piattina oppure il diodo subito al dipolo e poi due normali conduttori

la piattina è lunga 0,96 lambda ovvero 144 cm per 100 MHz

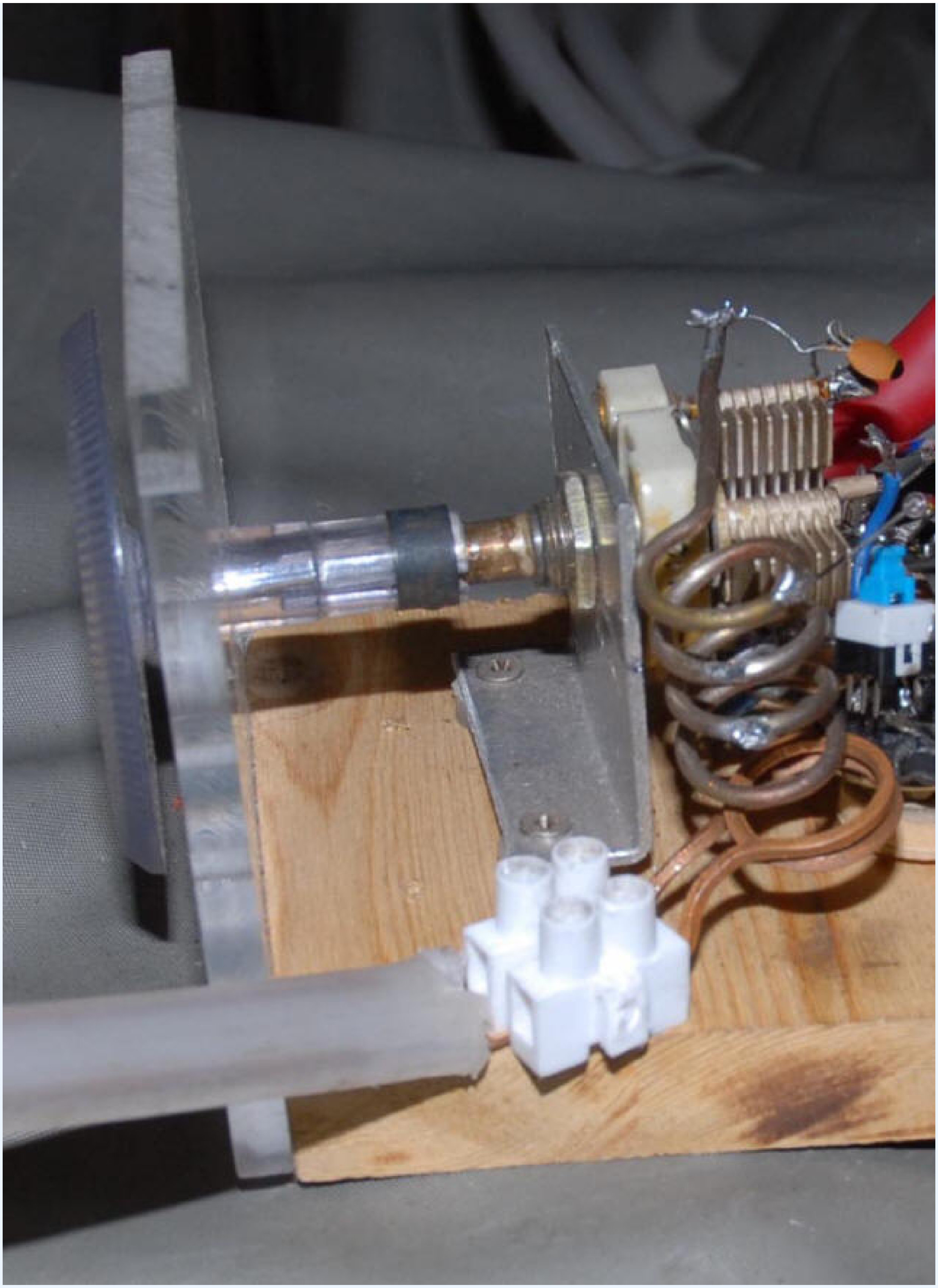

La realizzazione con la semplice bobina illustrata mi ha dato intensità maggiore che con la cavità, anche se un poco più di rumore. Una realizzazione fatta usando per bobina un anello di tubo di ottone da 6 mm e 11 cm di diametro, che teoricamente ha un Q 4 volte maggiore di quella classica, mi ha deluso. Dunque avanti con la semplice bobina raccomandandomi di osservare attentamente il disegno allegato, seguirlo a lettera e non arrendersi quando alle prime prove non funziona:capita sempre così.

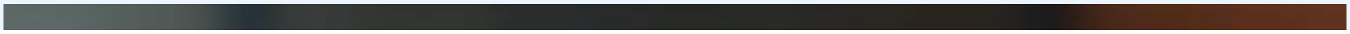

Radio a galena FM

per la “banda commerciale” 88 – 108 MHz

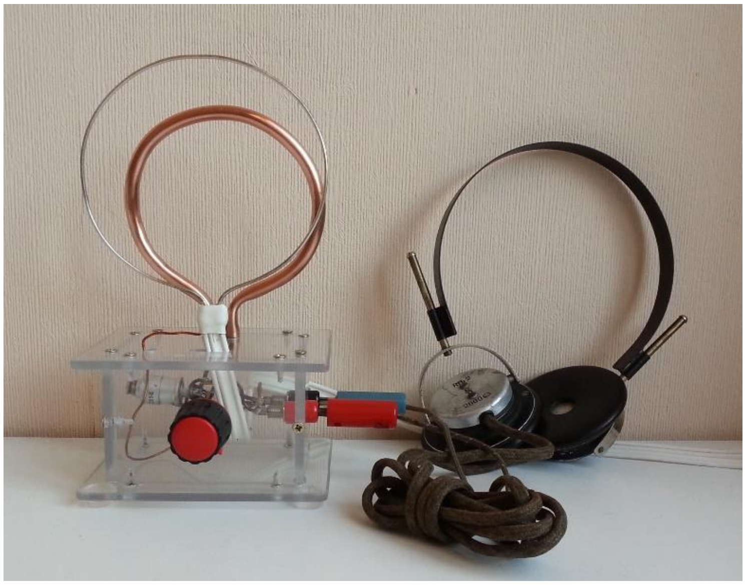

Figura 1: Radio a galena a montaggio ultimato.

Sono passati già otto anni da quando decisi di sperimentare con il primo ricevitore a galena. Lessi come costruirne uno proprio da questo sito (Realizza il tuo Primo Ricevitore a cristallo) e decisi di montarlo pensando di continuo che non avrebbe mai potuto funzionare veramente. La sorpresa nel “sentire qualcosa” da un vecchio auricolare telefonico è stata grande, e da allora la Vecchia Signora ha rapito la mia passione per l’elettronica, dirottandola verso la radiotecnica. Nonostante questa galena fosse pensata per ricevere segnali ad onde medie (come tutte le prime galene che si rispettino) è interessante notare che, già allora, quell’unica stazione che riuscivo a ricevere ho scoperto poi appartenere alla gamma delle onde corte… perché non è da pochi anni che le onde medie sono in crisi!
Stazioni abbastanza potenti per permettere un ascolto soddisfacente in cuffia da qualsiasi punto del Pese sono ormai quasi esclusivamente disponibili in onde corte, gamma che di sera offre segnali fortissimi provenienti da posti anche al di là del globo. Ma la radio a cristallo era pensata per permettere l’ascolto “dell’emittente locale”, e non di segnali provenienti da chissà dove, ed in chissà quale lingua! Sono cambiati i tempi, dunque… cambiamo anche la radio a glena, no? Eppure oggi le moderne tecnologie ed il più basso costo dei materiali permettono di realizzare ricevitori a cristallo di grandi prestazioni, proprio quando, paradosso, il tempo della radio a galena è da considerarsi ormai concluso.
L’unico territorio abbastanza inesplorato è quello della gamma broadcasting F.M., e ciò ha spinto gli appassionati amministratori e lettori di questo sito a sperimentare con le VHF, come è stato possibile notare dai progetti di radio a cristallo per FM che hanno fatto capolino da qualche anno ad oggi. Se si vuole sperimentare, d’altro canto, bisogna farlo subito giacché si sente parlare sempre di più di D.A.B. e di digitalizzazione…
È possibile ascoltare una radio a galena in FM solo se si abita nelle immediate vicinanze di un ripetitore; circa 10km di aperta campagna e di costa marina (ho provato) sono troppi per un ricevitore a cristallo.

Il caso ha voluto che andassi ad abitare in una città dove ci sono dei ripetitori, e questo mi è bastato per decidermi a sperimentare, rispondendo all'appello lanciato su questo sito qualche anno fa.
Questo articolo, già comparso nei numeri di giugno e di luglio-agosto della rivista Radiokit Elettronica, nasce come proposta di realizzazione del circuito presentato dal lettore Carlo Bramanti su questo sito (Galena per FM). Ad ogni modo il progetto da me proposto differisce alquanto dall'originale per la cruciale scelta del diodo rivelatore, del quale presento un piccolo "lavoro di ricerca", e per la presenza di un nuovo componente.

1. Schema elettrico

Il circuito elettrico della galena FM, riportato in Figura 2, risponde all'esigenza di essenzialità tipica di ogni ricevitore a cristallo. Lo schema elettrico deve essere semplice, soprattutto se si vogliono ricevere deboli segnali, per di più nella banda delle onde ultracorte.

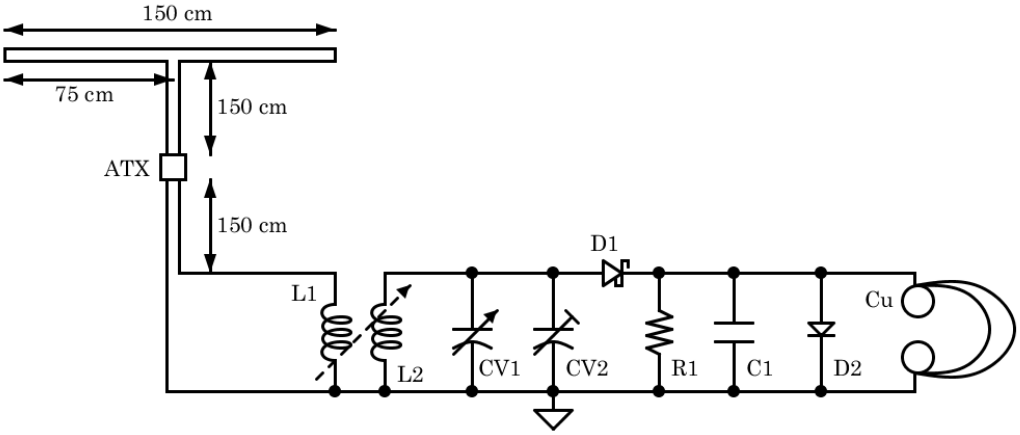

Figura 2: Schema elettrico della radio a galena FM.

I componenti adoperati sono:

- L1 = vedi testo (sezione Bobine L1 e L2);
- L2 = $0.137\mu H$, vedi Figura 7;
- CV1 = Johnson 160-211-1 $(2.7 - 10.8)pF$ per sezione;
- CV2 = trimmer tubolare da $(5 \div 15)pF$;
- D1 = diodo Schottky Skyworks modello SMS7630-001;
- D2 = diodo di segnale 1N4148;
- R1 = $47k\Omega$, $1/4$W;
- C1 = $100pF$ ceramico a disco;
- Cu = cuffie ad alta impedenza ($2k\Omega$ o superiore);
- ATX = connettore ATX femmina e header pin;
- Due connettori banana femmina.

Il segnale ricevuto dall'antenna è inviato alla bobina L1 che, insieme all'induttanza L2, forma un trasformatore RF in aria. L'induttore L1 è quindi il classico "primario aperiodico" il cui accoppiamento alla bobina di sintonia L2 è variabile a piacere. Il circuito accordato per la sintonizzazione delle stazioni è completato dall'ancor più classico condensatore variabile CV1 e dal correttore CV2, indispensabile

per la taratura del ricevitore. Seguono poi il diodo rivelatore D1 ed il filtro di deenfasi, costituito dal resistore R1 e dal condensatore C1.
Un componente piuttosto insolito, presente nella mia radio, è il diodo D2, in parallelo all'uscita; del tutto assente nelle classiche galene, questo diodo serve per proteggere il rivelatore, ed il motivo della sua presenza verrà spiegato più avanti.
Un paio di cuffie ad alta impedenza Cu completano il tutto, e possono essere sostituite da un qualsiasi amplificatore (va bene anche l'ingresso del microfono di un PC).

2. Realizzazione

Come in tutti i ricevitori a cristallo la realizzazione è sempre un passo importante; ottime saldature e collegamenti cortissimi sono la chiave per ottenere una radio a galena funzionante. Di seguito vi riporto nel dettaglio le varie fasi del montaggio arricchite di consigli empirici e note pratiche in parte sperimentate per la prima volta con questo ricevitore.

2.1. Antenna

Per questa radio le antenne possibili sono tante, ma ho condiviso la scelta del dipolo ripiegato già proposta da Carlo Bramanti per diversi motivi: Perché ha una resistenza di irradiazione di circa 300Ω, maggiore rispetto a quella del classico dipolo (75Ω), quindi tensioni di segnale più alte ai capi dell'antenna; perché è piuttosto compatta (basta con le antenne chilometriche ed ingombranti!) ma, soprattutto, perché ho trovato un modo facile e veloce per realizzarla. Sono riuscito a reperire una piattina per i vecchi impianti elettrici (una rimanenza di una ferramenta) che comprendeva pure il conduttore di messa a terra. Dopo averlo eliminato ho ritagliato la piattina in tre parti, di lunghezza indicata in Figura 2, per realizzare così un dipolo ripiegato ed un cavo di discesa. L'antenna a montaggio ultimato si presenta come in Figura 3. In pratica ho spelato i due conduttori alle estremità di uno dei pezzi da $150cm$ per poi saldarli tra loro.

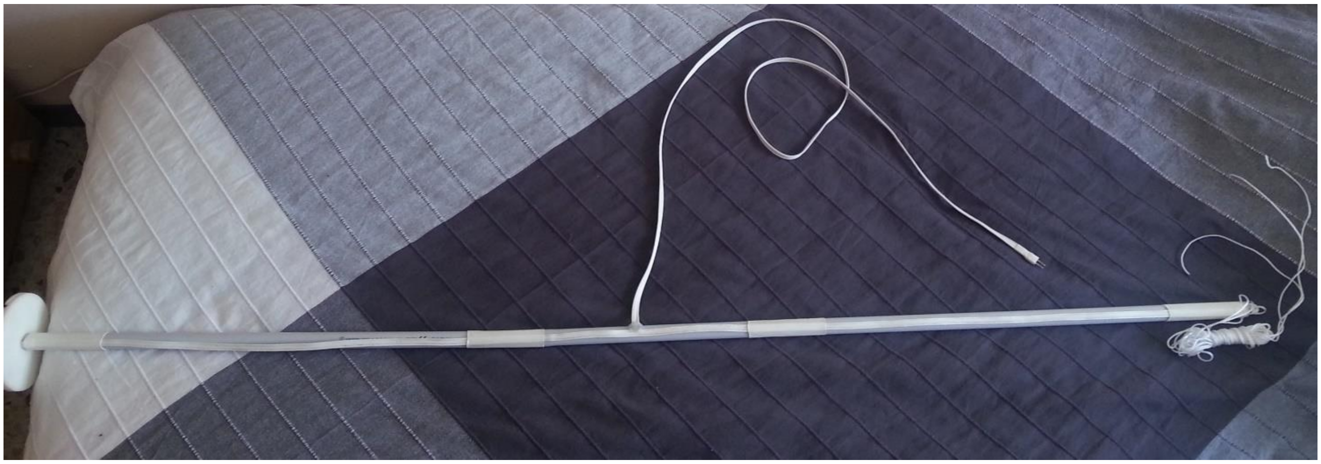

Figura 3: Dipolo ripiegato a montaggio ultimato.

Ho poi tagliato al centro uno dei due fili per realizzare la presa centrale per "il cavo di discesa", quello che termina con il connettore header pin (Figura 3). Il particolare della presa è visibile in Figura 4.
Chi non riuscisse a reperire una piattina (ormai non più in commercio) può sempre acquistare un dipolo

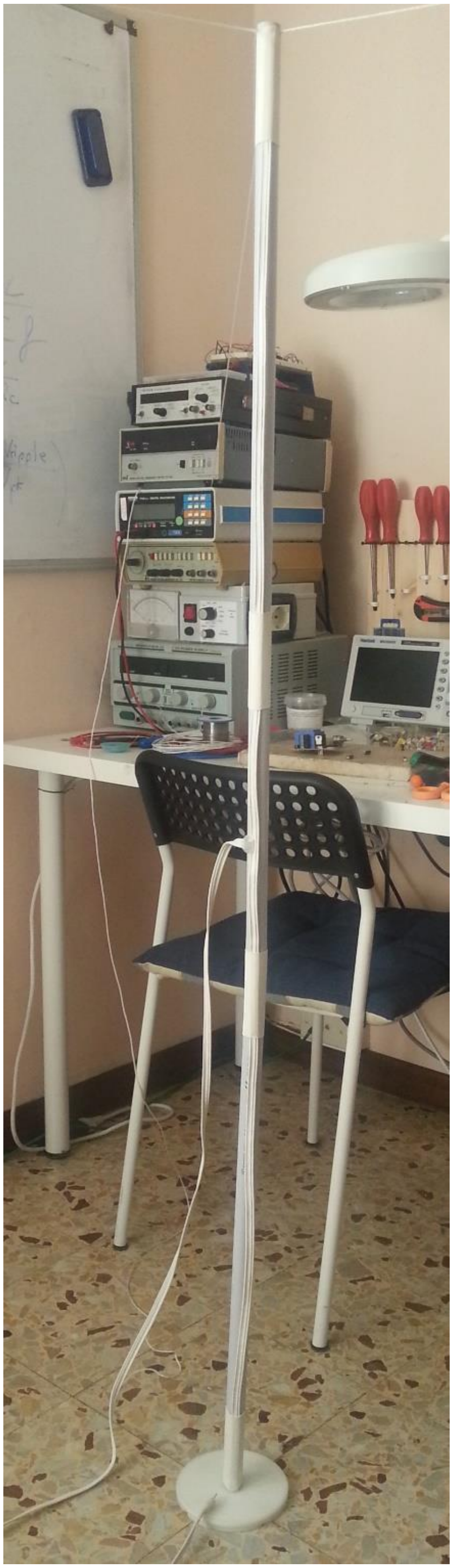

Figura 6: Esempio di impiego del dipolo ripiegato.

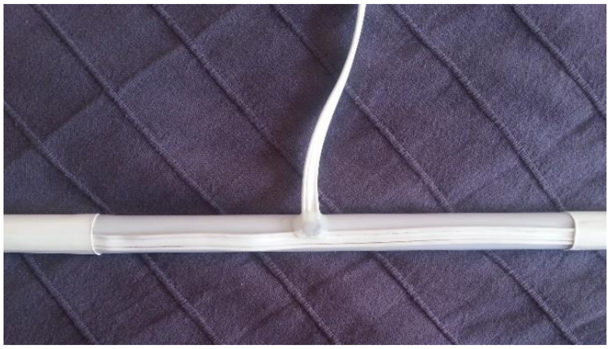

Figura 4: Particolare della presa per il cavo di discesa.

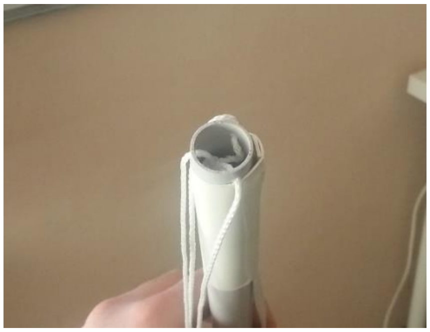

Figura 5: Particolare dei cordini fissati alla canalina.

ripiegato già pronto all'uso, di quelli venduti per i sintonizzatori degli impianti stereo domestici. Di solito anche queste antenne sono realizzate con una piattina di colore rossiccio/rosa trasparente.

Per poter mantenere l'antenna sempre "in piedi" ho pensato di fissarla su una canalina da $2cm$ di diametro, e lunga $152cm$ (poco più del dipolo, insomma). Per il fissaggio (peccato che non si possano usare viti metalliche…) ho adoperato degli spezzoni di tubo di gomma termorestringibile bianchi, lunghi circa $10cm$, e disposti a distanza quasi regolare, come visibile in Figura 3.

La presa centrale (Figura 4) è stata invece fissata con della colla termica trasparente. Sempre dalla Figura 3 si può notare anche una matassina di cordino bianco; è una corda da $1.5mm$ di diametro, di polipropilene (abbastanza anti ingarbugliamento) ritagliata in tre pezzi lunghi ciascuno due volte la lunghezza del dipolo. Economica, questa corda mi ha permesso di fissare l'antenna in qualsiasi punto di una stanza, realizzando dei rapidissimi tiranti. Basta fare un paio di nodi ed il dipolo ripiegato si mantiene in piedi, come si può vedere dalla foto di Figura 6. Questa soluzione, tanto semplice quanto funzionale, permette anche di appendere l'antenna, per fissarla in punti più in alto, lontani dalle pareti e dal pavimento. Ai non addetti ai lavori preciso che è indispensabile che l'antenna sia posizionata verticalmente, come di Figura 6, perché verticale è la polarizzazione dei segnali irradiati dai ripetitori FM che si vogliono ricevere; posizionando l'antenna orizzontalmente (ho anche provato!) non si riceve nessuna stazione. Per assicurare i cordini alla canalina ho praticato tre fori, come in Figura 5, ed ho unito le tre corde in un unico grosso nodo, riscaldato poi alla fiamma. Ho stampato in 3D anche una piccola base (il disco bianco di Figura 6), che contribuisce alla stabilità del dipolo, anche se non è strettamente necessaria. Chi non ha una stampante 3D può sostituire la base magari con un piatto di plastica (di quelli rigidi) ed un tappo di sughero…

Il dipolo così realizzato è accordato circa al "centro banda". Ricordando infatti che in un dipolo ripiegato i nostri $150cm$ rappresentano metà della lunghezza d'onda del segnale che si vuole sintonizzare, la frequenza corrispondente è data dalla relazione:

$$f = \frac{c}{\lambda} = \frac{3 \cdot 10^8 m/s}{2 \cdot 1.5m} = \ 100MHz$$

$75cm$ rappresentano quindi un quarto della lunghezza d'onda relativa al segnale a $100MHz$.

Per trasportare facilmente l'antenna e la radiolina ho pensato di interrompere il cavo di discesa con un connettore header pin maschio, che andrà poi innestato ad un connettore ATX, di quelli usati negli alimentatori dei PC per portare i $12V$ alla scheda madre. Ho adoperato questo tipo di connettore perché, oltre che essere robusto, presenta una distanza tra i due conduttori che è simile a quella della piattina, e magari il disadattamento introdotto lungo la linea del cavo di discesa è ridotto al minimo (trascurando la discontinuità dell'isolante).

2.2. Bobine L1 e L2

La piattina del cavo di discesa termina direttamente con la bobina L1. Questa è stata realizzata con un filo di rame argentato da $1.6mm$, avvolta a formare una sola spira circolare di circa $9.5cm$ di diametro. A fine realizzazione la ho subito protetta con dello spray di vernice trasparente, che ha la funzione di preservare il prezioso metallo dall'ossidazione pur non coprendone il colore originale.

Questa bobina così fatta è molto bella a vedersi; l'attenzione per l'estetica è una delle linee guida che ho cercato di seguire per realizzare questa radio, ed è una diretta conseguenza del tipo di montaggio scelto: scatola in plexiglass per avere tutto il circuito in vista. In ogni caso la sua induttanza non è molto importante; le prove che ho fatto hanno confermato che l'importante è che il diametro della spira sia pari o superiore al diametro medio della bobina L2. Non è neanche necessario che la bobina L1 sia accordata con un condensatore in serie, perché ci ho provato e non ho ottenuto migliorie.

Ho realizzato poi la bobina L2 piegando in una sola spira circolare un tubo di rame da $7mm$ di diametro; tutte le sue dimensioni sono indicate nell'immagine di Figura 7.

Questa è una delle soluzioni possibili per avere un'induttanza con un alto fattore di merito (purtroppo ad oggi non ho un Qmetro funzionante, e quindi non ho potuto misurarlo). L'induttore ottenuto ha quindi delle ridotte capacità parassite (una sola spira) ed un ridotto effetto pelle, avendo a disposizione un'ampia superficie di conduttore. Il tubo che ho adoperato infatti è cavo, di quelli impiegati per gli impianti dei climatizzatori, e reperito ad un brico. Vi auguro di trovarne uno con un diametro inferiore al mio, perché già $7mm$ è il massimo possibile per realizzare un'induttanza da $0.137\mu H$.

Per diametri maggiori, infatti, il tubo diventa veramente difficile da lavorare. Lo è stato già il mio, che sono riuscito a sagomare aiutandomi con una piccola morsa, dopo averlo piegato gradualmente su supporti di diametro decrescente. Al fine di ridurre l'effetto pelle è indispensabile che il rame non sia minimamente ossidato. Teoricamente, se non si tocca il tubo con le mani (il sebo della nostra pelle poi a contatto con l'ossigeno ossida il metallo) quindi adoperando i guanti, e se si riesce a lavorarlo in poco tempo, per poi subito proteggerlo con della vernice, sarebbe possibile evitare che si ossidi. In realtà ciò è difficile a farsi; il tubo è già molto difficile da piegare afferrandolo con le mani nude, figuriamoci con i guanti! … ma con un po' di pasta abrasiva si può eliminare tutto l'ossido ripristinando la lucentezza tipica del rame. Ed è questa la soluzione che ho usato per rendere bella la mia realizzazione, spruzzando quattro mani di vernice trasparente acrilica sulla bobina, dopo averla lucidata.

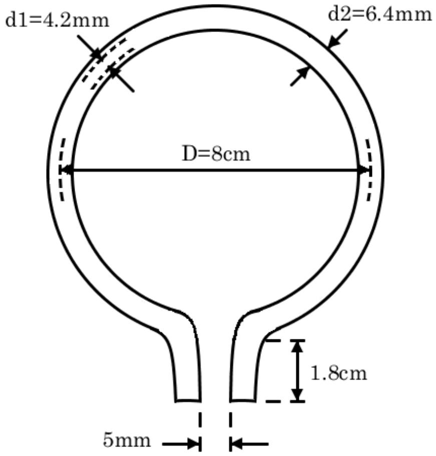

Figura 7: Dimensioni della bobina realizzata.

Per calcolare il valore dell'induttanza L2 ho faticato davvero tanto (con mia grande sorpresa!) a trovare la formula. Sono riuscito a trovarne una approssimata sul Radio Engineering Handbook (F. E. Terman) e, anche se non è di facile impiego, ve la riporto qualora doveste ottenere un'induttanza da $0.137\mu H$ con un tubo di diametro minore del mio (cosa che continuo ad augurarvi). Con riferimento alla Figura 7, l'induttanza di un tubo ripiegato a forma di Ω è:

$$L = 0.01595 \cdot D \cdot \left[2.303 \cdot \log_{10}\left(\frac{8 \cdot D}{d_2}\right) - 1.75 - \frac{d_1^2}{2 \cdot (d_2^2 - d_1^2)} + 2.303 \cdot \frac{d_1^4}{(d_2^2 - d_1^2)^2} \cdot \log_{10}\left(\frac{d_2}{d_1}\right)\right]$$

dove:

- L = induttanza in micro Henry;
- D = diametro medio della spira in pollici (come al solito!);
- d_1 = diametro interno in pollici;
- d_2 = diametro esterno, sempre in pollici.

Scusate per i pollici, ma se avessi incluso anche il fattore di conversione, la formula sarebbe diventata illeggibile. In ogni caso vi ricordo che, per passare dai pollici ai centimetri, basta dividere i centimetri per 2,54. Con un foglio Excel od una calcolatrice scientifica si può provare a variare il diametro D, fino ad ottenere il valore di induttanza voluto; il Terman consiglia comunque la possibilità di semplificare la formula per frequenze molto alte, come segue:

$$L = 0.01595 \cdot D \cdot \left[2.303 \cdot \log_{10}\left(\frac{8 \cdot D}{d_2}\right) - 2\right]$$

2.3. Condensatore variabile, rivelatore e filtro di deenfasi

Il condensatore variabile CV1 collegato in parallelo alla bobina L2 è del tipo ad aria con rotore a farfalla, ad alto fattore di merito, pensato per alte frequenze. È composto da due sezioni identiche, ma io ne ho usata una soltanto (avevo solo questo nel cassetto!). Considerando che il montaggio introduce una capacità parassita di almeno $10pF$ (è un valore molto ottimistico, lo so, ma con collegamenti cortissimi ci si può riuscire), e che il trimmer CV2 ha una capacità minima di $5pF$, è possibile sintonizzarsi sull'intera gamma FM con il variabile scelto. Il segnale radio così sintonizzato viene poi demodulato dal diodo D1 con la tecnica del classico "rivelatore a pendenza".

Il circuito RC-parallelo formato da C1 e R1 costituiscono, come già accennato, il filtro di deenfasi. Tale filtro deve essere del primo ordine, e con frequenza di taglio di $3.2kHz$ circa. Ho calcolato quindi i valori di R1 e C1 ipotizzando di collegare all'uscita del circuito un amplificatore ad alta impedenza (maggiore di $47k\Omega$, nel mio caso). A rigore avrei dovuto calcolare il condensatore C1 sulla base dell'impedenza di $4k\Omega$ offerta dalle mie cuffie, ma la qualità del segnale ascoltato non permette di apprezzare queste differenze. Ho provato infatti a collegare in parallelo alle cuffie un condensatore di valore idoneo, più alto, ma sono rimasto deluso nel non sentire alcuna differenza! Ho preferito allora lasciare valori bassi di C1, anche perché questo shunta a massa il segnale RF residuo, quindi è bene che sia ceramico, di piccola capacità. Il resistore R1, d'altronde, essendo collegato in parallelo all'uscita, tenderebbe a cortocircuitare il già debole segnale che si vuole ascoltare, quindi è bene che il suo valore sia più alto possibile.

2.4. Diodo di protezione

I'1N4148 serve da protezione per il delicato diodo rivelatore. Data la bassa tensione di soglia dell SMS 7630 – 001, dal datasheet si legge che la massima tensione inversa applicabile *Vr* è di circa $1V$. Peccato che le tensioni a vuoto indotte dalle cuffie, quando le si urtano un po' tra loro (o semplicemente le si appoggiano sul tavolo) sono di circa $6V_{pp}$, durata $1ms$ circa, e possono arrivare anche a $20V_{pp}$! Le cuffie infatti, come ogni trasduttore, possono funzionare "al contrario", fungendo da microfono, e generando segnali di valore elevato, data la loro alta impedenza. Per questo motivo, in parallelo all'uscita (quindi alle cuffie) ho inserito un "veloce" 1N4148, che tosa tutti i picchi maggiori di circa $0.6V$, impedendo quindi la rottura del rivelatore. Il lettore attento avrà certamente notato che il diodo "di protezione" potrebbe tosare anche il segnale radio che si vuole ascoltare; in realtà ciò non può mai accadere poiché, con le cuffie inserite, il segnale di uscita ha un valore massimo di picco di circa $250mV_{pk}$. Questa condizione sarebbe comunque da verificare, giacché la tensione del segnale di uscita dipende molto dalle cuffie adoperate. Il massimo valore di $250mV_{pk}$ è assicurato dato che, per ottenere questa misura, ho collegato alla radio direttamente l'uscita dell'oscillatore modulato, che quindi mi ha fornito i segnali più ampi possibili sintonizzabili con la galena. La presenza del diodo D2, in ogni caso, introduce una capacità trascurabile in parallelo al filtro (circa $5\ pF$), dunque la sua frequenza di taglio non viene alterata.

Un paio di connettori a banana completano il tutto, permettendomi di calzare le mie Siemens da $4k\Omega$ e godermi l'ascolto di qualche stazione (cuffie ad alta impedenza sono strettamente necessarie dati i debolissimi segnali in gioco). Ai meno esperti raccomando di collegare in parallelo all'uscita un'induttanza da $100mH$ o superiore, qualora decideste di optare per un paio di auricolari piezoelettrici (es. buzzer) al posto delle classiche cuffie ad alta impedenza, che magari non avete a disposizione. Tale induttanza servirebbe per cortocircuitare la tensione continua che altrimenti si accumulerebbe ai capi del condensatore C1, polarizzando il diodo e bloccandone l'azione rivelatrice, impedendovi quindi di sentire alcunché. Il resistore R1, comunque, dovrebbe già assolvere a questo compito, ma, a mio avviso, sarebbe opportuno verificare caso per caso.

2.5. Cablaggio

Per il cablaggio del circuito ho usato del filo di rame smaltato da $2mm$ di diametro per collegare la bobina L2 e lo statore del CV1, mentre per il diodo rivelatore è bastato un filo da $1{,}8mm$. Come è possibile notare dalla foto del montaggio (Figura 8) i collegamenti sono ordinati e corti, con fili che non sono mai piegati ad angolo retto, cosa da evitare assolutamente in alta frequenza.

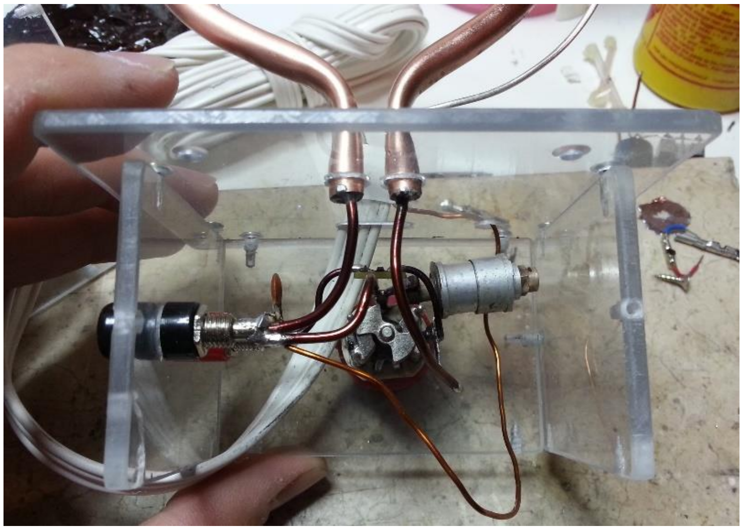

Figura 8: Vista posteriore del cablaggio

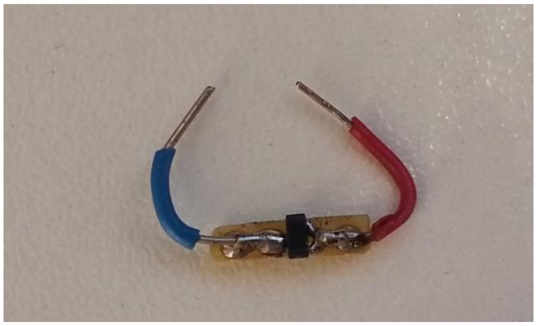

Figura 9: Particolare del diodo sulla basetta millefori.

Il filo di rame smaltato più chiaro e più sottile che si allunga verso il basso, visibile in Figura 8, serve per collegare un capo della bobina d'antenna con quello che è stato scelto come il riferimento del circuito. In questa maniera si può fare a meno della presa di terra per assicurare un percorso di ritorno per il segnale, ottenendo gli stessi risultati. Questo filo serve anche per mantenere la bobina d'antenna nella posizione scelta per realizzare l'accoppiamento voluto con la bobina di sintonia.

Il diodo rivelatore che ho impiegato nel mio circuito ha un contenitore tipo SOT-23, quindi pensato per il montaggio in SMD. In realtà il diodo non è troppo piccolo per essere saldato con una punta standard di un saldatore qualsiasi, quindi mi è bastato ritagliare una striscetta di una fila di basetta millefori, con quattro piazzole. Dalla Figura 9 si può vedere che il diodo occupa lo spazio tra una piazzola e la successiva; gli altri due reofori esterni sono stati usati per saldare i due fili che mi sono serviti per provarlo.
I collegamenti del diodo al resto del circuito sono visibili in Figura 10.

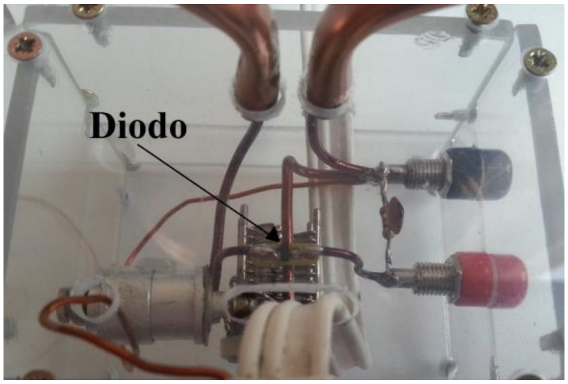

Figura 10: Cablaggio del diodo rivelatore.

A montaggio ultimato, nella sua scatola di plexiglass da $3.5mm$ di spessore, la radiolina appare come nelle immagini di Figura 11, Figura 12, Figura 13 e Figura 14.

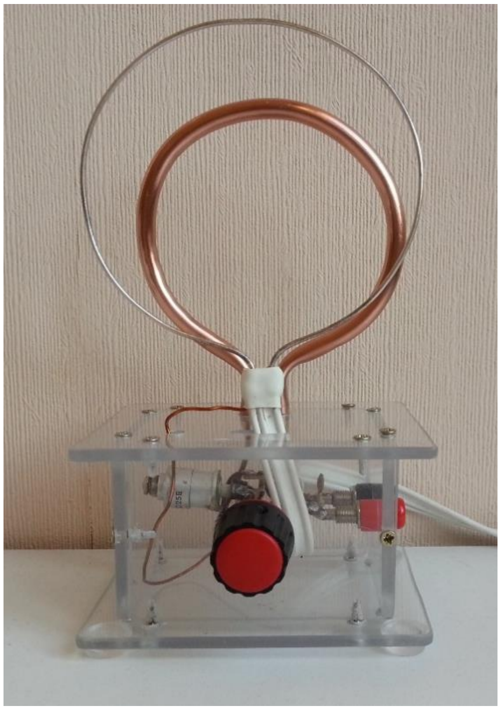

Figura 11: Vista frontale.

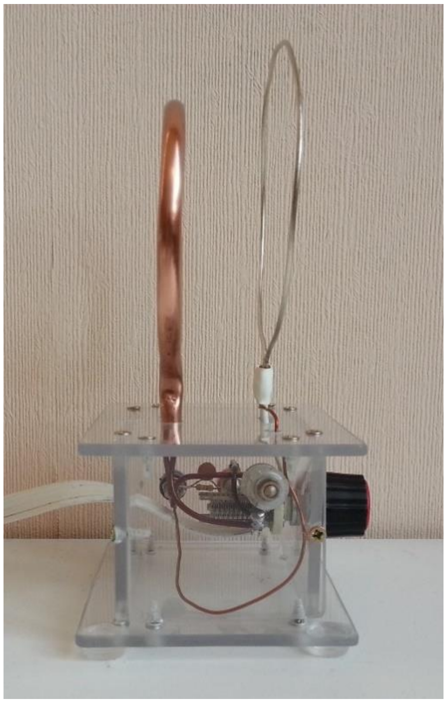

Figura 12: Vista laterale sinistra.

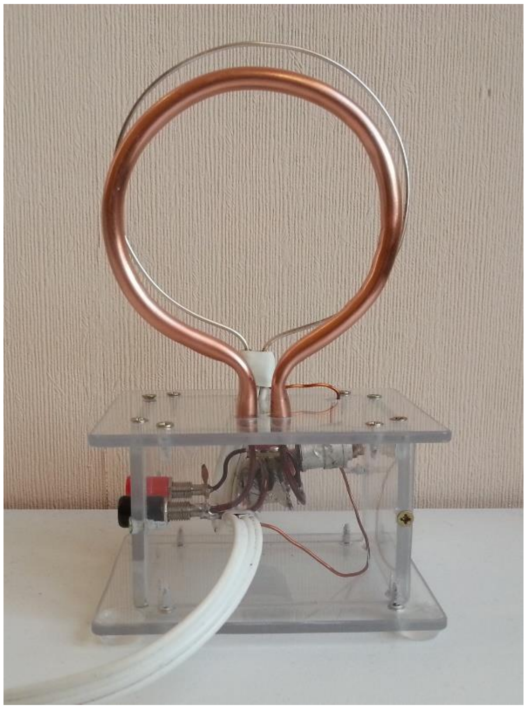

Figura 13: Vista posteriore.

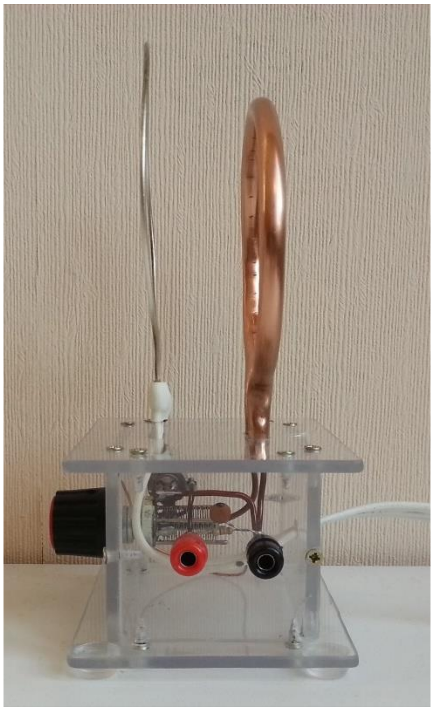

Figura 14: Vista laterale destra.

La realizzazione di un mobile con questo materiale è da sconsigliare per chi, come me, non ha un laboratorio attrezzato con apparecchi di precisione per il taglio e la foratura. Il plexiglass mette molto in risalto i minimi graffi, ed è piuttosto difficile da lavorare e "tagliare dritto". Tuttavia non ho rifatto la scatola avendo impiegato molto tempo per realizzarla. Per fissare i vari pannelli tra loro ho adoperato un metodo che mi ha permesso di ottenere un montaggio solido: ho forato il plexiglass con una punta da $2mm$, per poi inserirvi delle viti autofilettanti dal diametro poco superiore. Il risultato estetico ottenuto non è un granché (a parer mio), perché nei pannelli si vedono delle protuberanze biancastre al posto delle "viti affogare nel vetro", effetto che speravo di ottenere. In ogni caso il contenitore completamente trasparente ha il suo fascino, anche se costringe a posizionare i vari fili di collegamento in un certo modo, che possa essere insieme bello da vedere e fatto per ridurre al minimo le capacità parassite e la lunghezza dei collegamenti.
Con questo credo di avervi raccontato tutto… della complicata teoria dice che i segnali FM ricevuti sono molto deboli quindi, per realizzare una soddisfacente radio a galena su questa gamma bisogna avere l'antenna migliore (spero di averla!) le bobine con il più alto fattore di merito (le abbiamo) un buon condensatore variabile e delle cuffie ad alta impedenza (non credo ce ne siano da più di $4k\Omega$). Ma se il segnale è debole, ci vuole un ottimo diodo rivelatore, con bassissima tensione di soglia, e magari a basso rumore (e piccole capacità parassite). Ad una fiera dell'elettronica trovai degli strani diodi, venduti come "diodi tunnel". Avevano i terminali cortissimi, per essere innestati in degli appositi connettori, con un contenitore cilindrico plastico, marrone. Erano gli 1N82A; ne presi un po', convinto che un giorno li avrei usati per chissà quale strana applicazione. Dopo aver visionato il datasheet scoprii che erano in realtà dei diodi mixer adatti per le bande VHF/UHF. Avevo a disposizione anche i classici OA90, OA95, AA117, e 1N542. Dopo aver montato un primo prototipo di galena ho provato tutti questi diodi, scoprendo che la situazione era talmente tanto critica che solamente due degli esemplari di 1N82A permettevano un ascolto soddisfacente in cuffia. Nel tentativo di fare tutte le prove possibili per migliorare al massimo la mia radio, ho deciso di acquistare un po' di diodi RF, del tipo "mixer" o "zero-bias", a bassa tensione di soglia, e di cercare il migliore tra questi.

3. Alla ricerca del "diodo migliore"

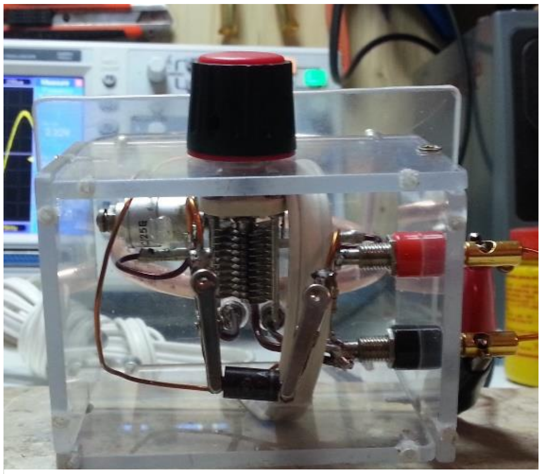

Figura 15:Particolare dei coccodrilli montati per provare i candidati diodi

Per riuscire in questa impresa ho pensato di non fare troppe cerimonie, testando magari il diodo in "circuiti equivalenti" a quelli della mia radio, ma di fare alcune misure collegando il diodo nello stesso circuito nel quale poi l'avrei usato: la mia radio, appunto. Per provare agevolmente tutti i rivelatori candidati ho saldato dei coccodrilli come in Figura 15. Dopo qualche prova preliminare con l'oscillatore modulato mi sono subito reso conto che la misura del segnale rivelato dopo il filtro di deenfasi era tutt'altro che banale, perché il segnale BF ai capi delle cuffie era molto distorto e rumoroso, oltre che "piccolo".
Dopo tante prove sono pervenuto all'allestimento del banco di misura illustrato in Figura 16 e Figura 17, e schematicamente rappresentato in Figura 18.

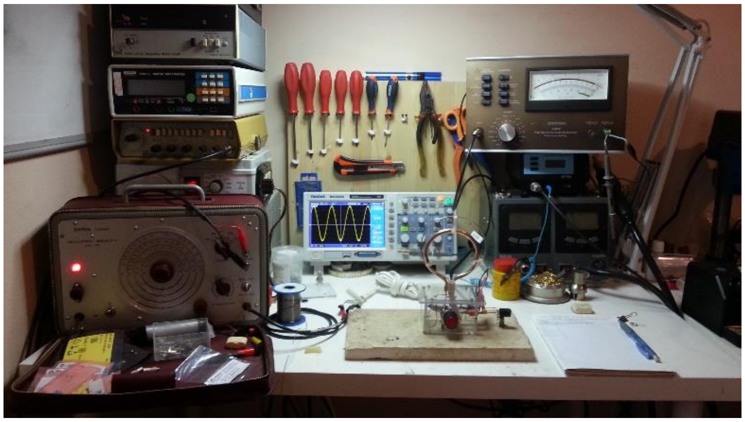

Figura 16: Banco di misura.

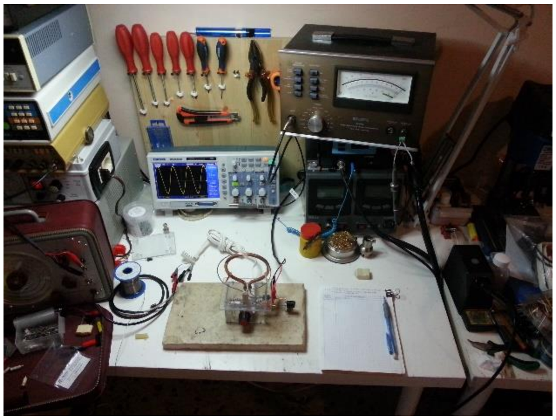

Figura 17: Vista superiore del banco di misura.

Le impostazioni degli strumenti, in parte già presenti in Figura 18, sono:

- **Generatore di funzioni**: segnale di uscita sinusoidale, ampiezza $400mVpp$ (con uso dell'attenuatore di 30 dB), frequenza $1kHz$, senza offset;
- **Oscillatore modulato**: ingresso per modulazione esterna selezionato, frequenza (misurata con l'oscilloscopio) di $88MHz$, attenuatore regolato per la massima ampiezza del segnale d'uscita;
- **Noisemeter**: amplificazione $1mV$ fondo scala, filtro CCIR inserito.

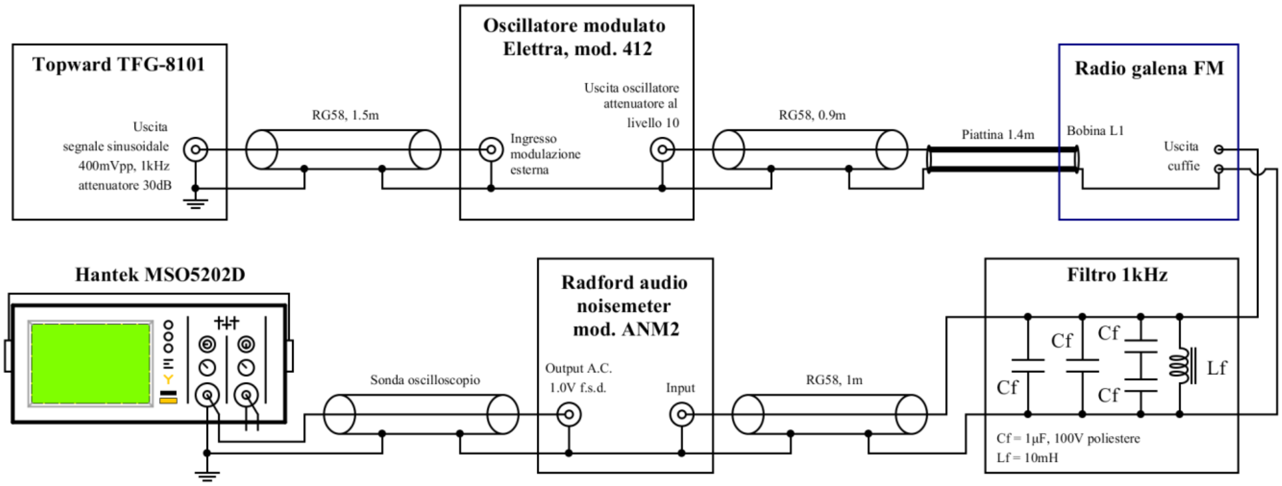
Topward TFG-8101
Uscita
segnale sinusoidale
400mVpp, 1kHz
attenuatore 30dB
RG58, 1.5m
Oscillatore modulato
Elettra, mod. 412
Ingresso
modulazione
esterna
Uscita oscillatore
attenuatore al
livello 10
RG58, 0.9m
Piattina 1.4m
Radio galena FM
Bobina L1
Uscita
cuffie
Hantek MSO5202D
Sonda oscilloscopio
Radford audio
noisemeter
mod. ANM2
Output A.C.
1.0V f.s.d.
Input
RG58, 1m
Filtro 1kHz
Cf
Cf
Cf
Cf
Lf
Cf = 1μF, 100V poliestere
Lf = 10mH

Figura 18: Schema del banco di misura messo a punto.

Gli accorgimenti che mi hanno permesso di ottenere un segnale rivelato pulito sullo schermo dell'oscilloscopio sono stati:

- La sistemazione dei cavi di collegamento lontano dalla galena.
- La scelta di un segnale di modulazione esterno perché quello generato internamente è distorto. L'ampiezza del segnale modulante è stata scelta la massima possibile per non avere sovramodulazione;

- La scelta di un segnale di $88MHz$ per l'oscillatore modulato perché il segnale dello strumento a quella frequenza è più ampio. In ogni caso non sarei mai riuscito a sintonizzarmi sull'estremo alto della gamma con le capacità parassite aggiunte dai coccodrilli di Figura 15;
- Il doppio filtraggio del segnale audio ottenuto sia con un circuito accordato alla frequenza di $1kHz$, collegato all'uscita della radio, sia con il filtro CCIR del noisemeter. La frequenza di $1kHz$ è stata scelta perché la curva CCIR non amplifica né attenua a questa frequenza. Ho prediletto poi questo filtro del noisemeter perché è quello che presenta il maggior filtraggio della banda audio.

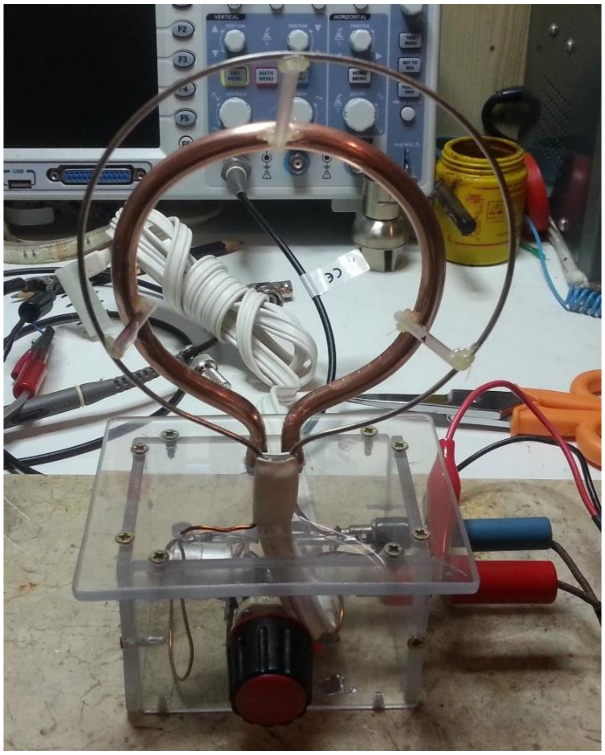

Figura 19: Vista frontale dell'accoppiamento fisso delle bobine.

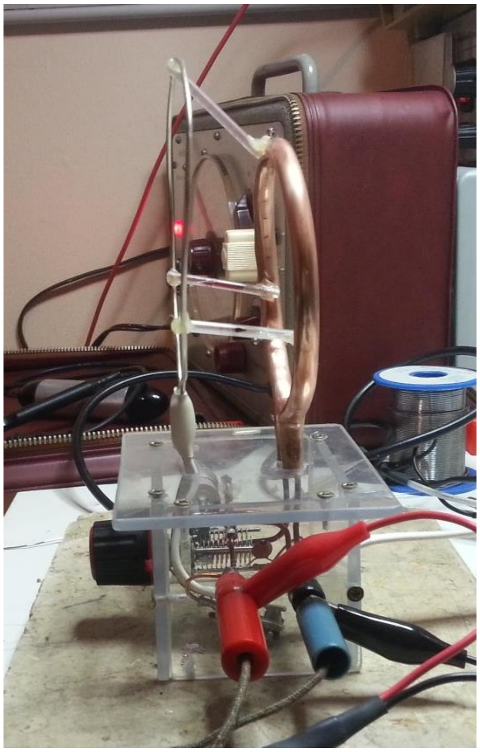

Figura 20: Vista laterale dell'accoppiamento fisso delle bobine.

Prima di eseguire le misurazioni ho atteso più di 20 minuti per permettere a tutti gli strumenti di "riscaldarsi".
Per ottenere delle misure coerenti tra un diodo e l'altro ho agito in questo modo:

- Per prima cosa ho fissato la bobina d'antenna alla bobina di sintonia per mezzo di tre spezzoni di tubetto di plastica (ex contenitore di inchiostro di penna Biro) incollati con colla termica, come indicato nelle foto di Figura 19 e Figura 20. In questa maniera l'accoppiamento tra L1 e L2 è stato sempre lo stesso per tutti i diodi;
- Ho inserito nel circuito dapprima il diodo 1N82A, che ho scelto come riferimento; mi sono accertato che il massimo segnale udito in cuffia durante l'operazione di sintonizzazione corrispondesse alla massima ampiezza del segnale sinusoidale misurato con l'oscilloscopio e con il noisemeter;
- Ho eliminato le cuffie perché tendevano a sporcare il segnale; dopo di che ho misurato tutti gli altri diodi;

- Ogni volta dopo aver inserito un nuovo diodo ho risintonizzato il ricevitore fino ad ottenere la massima deviazione dell'indice del noisemeter;
- Prima di misurare un nuovo diodo ho rimisurato l'1N82A per accertarmi di riottenere lo stesso valore letto dopo la prima misurazione;
- Ho eseguito le misure una di seguito all'altra, senza toccare i cavi di collegamento e restando seduto sempre nello stesso punto, onde evitare di variare gli accoppiamenti capacitivi tra il mio corpo e la radio ed i cavi;

Vi faccio presente che l'ingresso del noisemeter è ad alta impedenza ($1M\Omega$), quindi tutte le misure della Tabella 1 rappresentano i valori del segnale ad $1kHz$ misurato all'uscita della radio come se fosse "a vuoto". L'inserzione di un resistore da $4.7k\Omega$ per simulare il carico introdotto dalle cuffie ha avuto come effetto quello di ridurre troppo l'ampiezza del segnale in uscita, rendendolo ancora più rumoroso; quindi non l'ho usato. In ogni caso il filtro passivo ad $1kHz$, che idealmente non avrebbe dovuto "caricare" il circuito, ha ridotto di molto l'ampiezza del segnale rivelato, segno che, in realtà, un piccolo carico era comunque introdotto. Per questa ragione, onde visualizzare un bel segnale sullo schermo dell'oscilloscopio, ho adoperato l'uscita del noisemeter, che può essere usato come amplificatore calibrato. Un esempio del segnale demodulato si può intravedere sullo schermo dell'oscilloscopio nelle foto di Figura 16 e Figura 17. Ho adoperato anche l'oscilloscopio perché il Radford è un voltmetro a vero valore efficace, e avrebbe potuto fornirmi un massimo di tensione laddove in realtà l'ampiezza picco-picco del segnale rivelato non era massima. Con l'uso del noisemeter e dell'oscilloscopio ho potuto controllare volta per volta che il massimo segnale rivelato (Vpp) coincidesse con la massima deviazione dell'indice dello strumento.

Ad essere rigoroso avrei dovuto adattare l'impedenza di uscita dell'oscillatore modulato (50Ω) all'impedenza della piattina, che si aggira intorno a 300Ω. Ho deciso di non farlo perché l'adattatore d'impedenza fornito insieme all'oscillatore modulato, essendo passivo, riduceva di molto l'ampiezza del segnale generato. Il ricevitore ha comunque funzionato bene anche senza adattatori, segno che il disadattamento del circuito era trascurabile.

Nella Tabella 1 ci sono tre colonne, ed ogni colonna si riferisce alla misura di un esemplare del diodo indicato nella corrispondente riga. Questo perché, laddove il portafogli me lo ha consentito (alcuni di questi diodi sono piuttosto costosi) ho acquistato tre esemplari dello stesso diodo, sulla scorta del fatto che dei sei diodi 1N82A che avevo, soltanto tre hanno funzionato a dovere, mentre soltanto due erano utilizzabili come rivelatori. Ho rieseguito le misure per più di una volta, ed i valori riportati sono da intendersi con una "precisione" (accuratezza) di $0.05mV_{eff}$ in più o in meno.

Le misure riportate mi sono servite quindi per vedere quali tra i diodi provati mi fornisse, nelle stesse condizioni degli altri, un segnale rivelato più ampio; non si deve quindi andare a guardare il decimo di milliVolt in più o in meno. Inoltre può essere accaduto che, per i diodi dei quali ne ho acquistato un solo esemplare, me ne sia capitato uno un po' sordo, ma non è detto che questa sia una prerogativa di quel particolare modello di diodo.

Dopo aver eseguito le misure ho guardato la tabella, ed il vincitore come miglior diodo rivelatore si è *rivelato* essere il... SMS7630-001! Meno male, dovrei dire, perché l'1N23A, l'MA423A e l'1N405A costano 15 euro l'uno! L' SMS7630-001 invece mi è costato solo 1 euro. In realtà questo era il risultato che speravo di ottenere, perché dimostra che, per realizzare un'ottima galena, possono andare bene dei diodi di moderna concezione, e che quindi non si deve andare alla ricerca di chissà quale esoterico surplus da chissà quale negozio e/o fiera.

Diodo	Segnale rivelato $[mV_{eff}]$		
	primo	secondo	terzo
1N82A	0.80	0.85	-
AA137	0.35	0.35	0.30
DDC2353	1.00	0.97	-
ND4991	1.00	0.97	-
SMS7630-001	1.50	1.50	-
AAY30	0.28	0.30	0.32
BAT17-04	0	0	0
AA114	0.10	0.15	0.10
OA95	0.10	0.10	-
1N914	0	0	-
1N618	0.15	-	-
1N76A	0.80	-	-
1N23A	0.80	-	-
MA423A	0.70	-	-
1N405A	0.87	-	-

Tabella 1: Misure del segnale rivelato con i diodi.

Guardando poi i datasheet dei diodi della Tabella 1 risulta chiaro perché l'SMS7630-001 sia il migliore: è quello che presenta la più bassa tensione di soglia, ma anche la più bassa tensione inversa! Cosa che mi ha costretto ad aggiungere un 1N4148 in parallelo all'uscita. Ho aggiunto questo diodo dopo aver fatto "il servizio fotografico" alla radio, ed è per questo motivo che non è presente nelle foto.

4. Taratura e messa a punto

Per poter sintonizzare tutta la gamma FM bisogna eseguire una piccola taratura agendo sul trimmer tubolare. Per fare ciò ho aperto la scatola da uno dei lati, come in Figura 21, ho collegato l'uscita dell'oscillatore modulato ad un filo di un metro circa, che poi ho avvicinato all'antenna. Dopo aver impostato lo strumento per generare un segnale modulato di $87MHz$, con il condensatore variabile tutto chiuso, ho avvitato il trimmer con un tubetto di colla termica, per non toccarlo con le mani, fino ad ottenere il massimo segnale in cuffia. Il ricevitore è così tarato e "pronto all'uso".

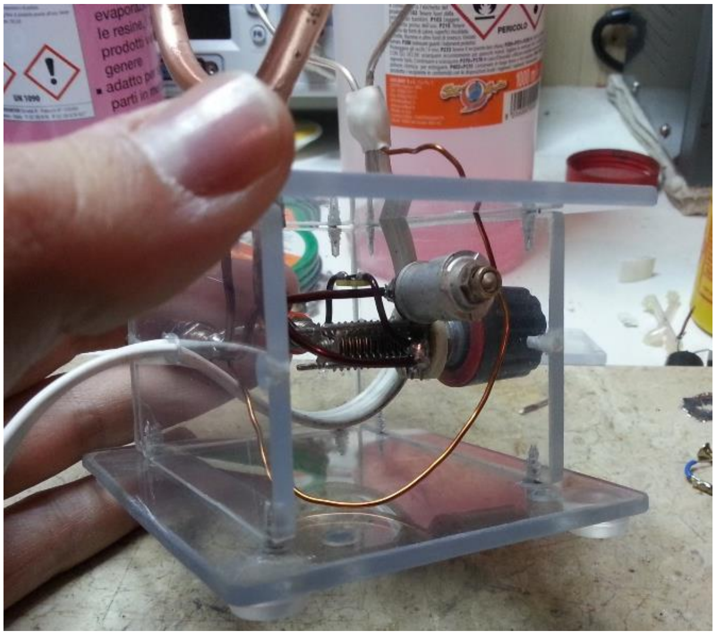

Figura 21: Taratura del ricevitore.

5. Prova d'ascolto

Usare la radio è molto semplice; basta montare l'antenna come per esempio in Figura 6, collegarla alla radio avendo cura di posizionare in orizzontale il cavo di discesa, cosicché non possa fungere da seconda antenna, calzare le cuffie e ruotare il condensatore variabile. Separare le stazioni interferenti è sempre possibile grazie al sistema di accoppiamento tra la bobina d'antenna e quella di sintonia, che permette di posizionarle anche ad angolo retto e ad una distanza di ben $5cm$, come è possibile notare nelle foto di Figura 22 e Figura 23.

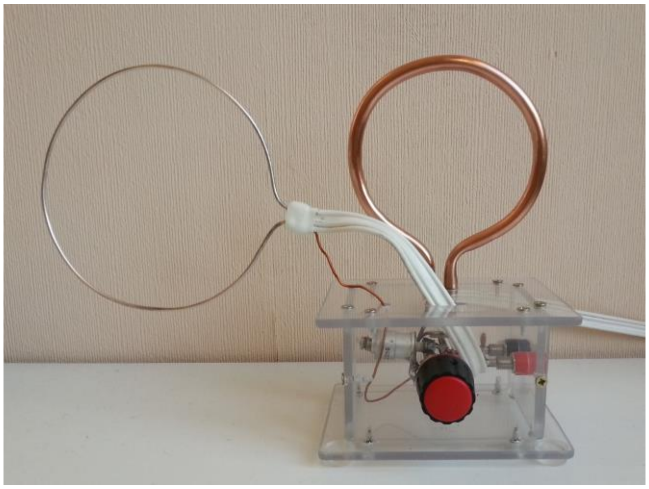

Figura 22: Esempio di accoppiamento lasco ottenibile tra le bobine L1 e L2.

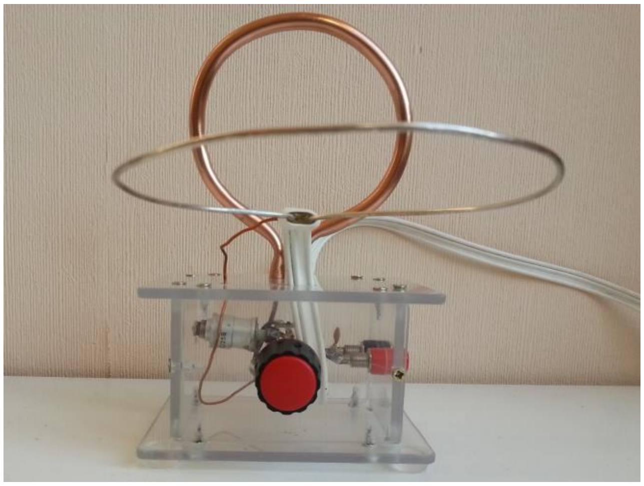

Figura 23: Esempio di accoppiamento minimo ottenibile tra le bobine L1 e L2.

I risultati ottenuti variano molto in base alla posizione del radiotecnico rispetto all'antenna ed all'installazione del dipolo ripiegato. In ogni caso con l'antenna montata nella mia stanza sono riuscito a ricevere un paio di stazioni, la più forte delle quali aveva un'ampiezza di circa $60mV_{pp}$. Con la stessa antenna sono riuscito a ricevere quattro stazioni collegando al posto delle cuffie un amplificatore. Con un'installazione migliore, sul balcone, sono riuscito a ricevere cinque stazioni, tre delle quali ben udibili.
Chi non ha già sperimentato con una radio a galena non può immaginare la gioia di poter ascoltare delle stazioni in lingua italiana, a qualsiasi ora del giorno e senza il bisogno del classico impianto antenna-terra. Un'antenna lunga almeno 10 metri è spesso difficile da installare, mentre una buona presa di terra è spesso impossibile da trovare, e puntualmente costringe il radiotecnico a posizionarsi in un punto scomodo della propria abitazione sia durante le sperimentazioni che durante le sessioni d'ascolto. Con un'antenna compatta come il dipolo ripiegato, che include anche la presa di terra, la radio a galena diventa un'oggetto portatile, come in effetti lo è sempre stato il circuito del ricevitore in sé. Sperimentare con le VHF può essere sicuramente difficile ma, a conti fatti, credo proprio che i vantaggi ottenuti compensino ampiamente le avversità con le quali ogni sperimentatore è chiamato a scontrarsi.
Con questo non mi resta che augurarvi un buon ascolto con tante sperimentazioni attorno ad una galena simile alla mia e alle altre già presentate su questo sito, sperando di esservi stato d'aiuto nel darvi molte idee per la realizzazione, e che la mia galena, insieme alle altre, vi abbia convinto ad avventurarvi in questa difficoltosa quanto gratificante impresa.
Per eventuali ulteriori chiarimenti resto sempre a disposizione; potete contattarmi all'indirizzo email damiano.cirielli@gmail.com.

Damiano Cirielli, 4/09/2017

Just for Fun: Crystal Radios

Many years ago, one of my sons asked how a radio worked, and I suggested that the best way to answer his question was for us to build a crystal radio, as I had when I was about his age. As it turned out, it wasn't quite as easy as I had remembered, largely because some of the key components are much harder to find now. But, we found the parts, had a lot of fun and learned some things.

A couple of years after we built our first radio, we learned about a great organization, the X-tal Set Society (http://www.midnightscience.com). The Society has a web site, publishes books and a newsletter with lots of plans, and sells many of the hardest to find parts, as well as complete kits. If you are at all interested in crystal radios, be sure to visit the web site and buy some books and parts!

Here is some of what we learned and plans for the radio we came up with:

What does this have to do with proteins?

The truth is that many distinguished protein scientists have had very successful careers without knowing much about crystal radios! But, there is a bit of a connection. NMR spectroscopy, which is one of the most powerful modern tools for studying protein molecules, uses some of the same technology as radio communication. In fact, the circuitry of an NMR probe is almost identical to that of the tuning circuit for a radio, as described below.

Now, this may not be a sufficient motivation for you to learn about crystal radios, especially if you found this web site in the hopes of learning something about proteins. In this case you should probably take a look at some of our scientific publications (publications.html).

Introduction

The distinguishing feature of a crystal radio is that it requires no electric power: The sound is generated from the energy of the radio waves themselves, without any amplification. The audio signal is weak and is only powerful enough to drive earphones, but the sound can be wonderfully clear.

In the early years of the 20th century, crystal sets were the first radios in many homes. Even after radios with vacuum tubes (and then transistors) became widely available and relatively inexpensive, building a crystal radio was a very common introduction to electronics for many kids.

The Crystal

As suggested by the name, the key component is a crystal. Originally, the crystal was a mineral called galena. When a metal wire is brought into contact with galena, a rectifying junction is formed. That is, electricity flows through the connection with lower resistance in one direction than in the other direction. This provides a way of converting the high frequency AM radio signal into the much lower frequency audio signal that is then converted into sound in the earphones. Extracting the audio signal from the modulated radio signal is called "detection", and the crystal with the wire that touches it (called a "cat's whisker") is a "detector". The picture below shows a crystal detector I built, along with its modern equivalent, the diode.

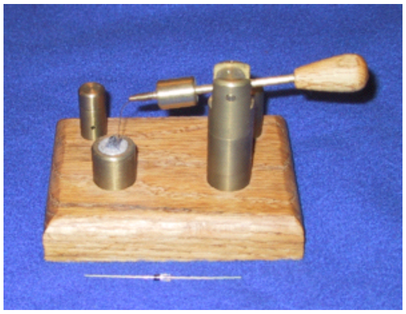

This is not an exact reproduction of an antique detector, but it is similar in style and serves the same functions of holding the crystal steady and allowing the cat's whisker to be carefully positioned. Not every position on the crystal will work, and it takes lots of patience to find a good spot and carefully adjust the cat's whisker for the best signal. Each crystal is different, and some work better than others. When radios were first installed on ships, each operator would often have his own favorite crystals that he would carry with him from one ship to another.

In the 1920's crystal detectors were largely replaced with vacuum tubes, which amplified the signal as well as detected it. During World War II, however, the crystal detector was brought back to prominence as a key component in the very high frequency circuits necessary for radar. Through a great deal of fundamental research in solid-state physics, methods were developed to reliably produce efficient detectors that did not require any of the fiddly adjustment of the original devices. Solid-state diodes became essential components of nearly all electronic devices, and most crystal sets are now built using readily available germanium diodes, like the one shown above. (A package of 10 can be bought at Radio Shack for about a dollar.) If, though, you would like to build a radio with a real crystal, galena crystals and wire for the cat's whisker can be purchased from the X-tal Set Society.

After World War II, the diode and the knowledge of solid state physics that was gained in developing it became the basis for the first transistors. Transistors, in turn, led to integrated circuits, and integrated circuits, of course, led to inexpensive computers and all of the other electronic marvels we now take for granted. The simple crystal detector is the direct ancestor of modern solid-state electronics.

The Tuning Circuit

The second major requirement for a radio receiver is a means of separating out signals with different frequencies, i.e. tuning in different stations. The simplest efficient circuit for this purpose consists of just a coil of wire and a capacitor connected in parallel, as shown below using the standard symbols for electronic components.

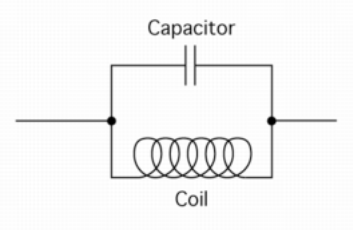

Typically, the coil for a crystal radio is formed by wrapping a single layer of enamel-coated copper wire over a cardboard or plastic tube a few inches long and an inch or two in diameter. The coil has a very important electronic property: Although direct current passes through it with low resistance, its resistance to a current that rapidly changes direction (an alternating current) increases as the frequency of the reversals increases. (The resistance to an alternating current is more properly called "impedance".) The impedance depends on the dimensions of the coil and the frequency of the current.

As indicated by its schematic symbol, a capacitor in its simplest form is composed of two metal plates separated by air or some other non-conducting material. This may seem like an odd arrangement for an electronic component, since it doesn't look as though electricity could pass through it. Indeed, direct current can't. But, an alternating current can. The impedance of a capacitor decreases as the frequency increases; exactly the opposite of the effect of a coil.

Now, think about what would happen if the circuit was connected to currents of different frequencies. At a very low frequency, the current will pass readily through the coil, but not the capacitor. At very high frequencies, the current passes through the capacitor but not the coil. For a given combination of coil and capacitor, there will be a magic frequency at which the impedance of the circuit is a maximum. At this frequency, electrical charge alternately builds up on the two plates of the capacitor and flows back and forth through the coil. The circuit is said to "resonate", just as a weight on a spring will resonate at a characteristic frequency.

In a radio, one side of the tuning circuit is connected to an antenna. For maximum efficiency, the other side of the circuit is connected to a metal rod or pipe buried in the ground. Radio waves are converted into alternating electrical currents in the antenna. Signals with frequencies that are higher or lower than the resonant frequency pass through the tuning circuit to the ground. (If the radio has a means of amplifying the weak signal from the antenna, as all modern radios do, the ground connection isn't necessary.) If, however, there is a signal that matches the resonant frequency, the energy of this signal can be drawn off to create the audio signal. In a simple crystal radio, the crystal detector and headphones are connected to the tuning circuit as shown below:

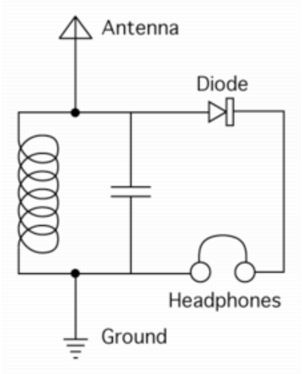

Detailed instructions for building a radio based on this circuit can be found below.

If the radio is to receive signals of more than one frequency, there has to be a means of adjusting either the coil or the capacitor. An adjustable coil can be made by scraping the enamel insulation off of the wire along a section of the coil running along its length and then arranging a slider that can be moved to contact the coil wire at different points. This is the method used in our radio, as shown in the picture below:

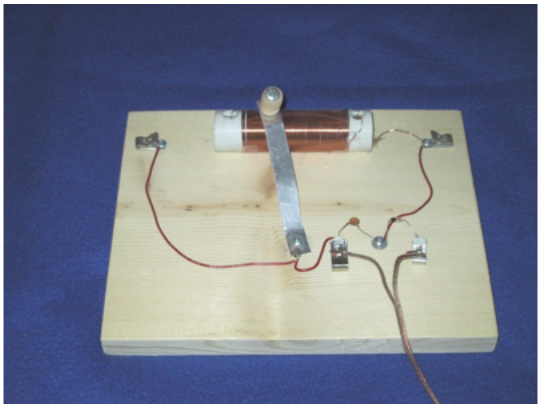

There are also adjustable capacitors. These usually have two sets of metal plates, one of which is mounted on a shaft that is turned to adjust the overlap between the two sets, as shown below:

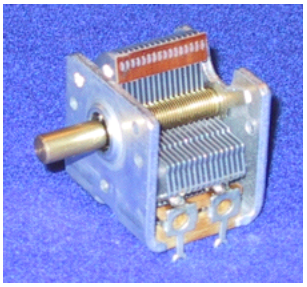

Capacitors suitable for AM radios aren't so easy to find anymore, but the X-tal Set Society (http://www.midnightscience.com/catalog5.html) has found someone to manufacture them, and they are available for a quite reasonable price. Using one of these, you can build a crystal radio without having to make a slider, or you can build a set with both components adjustable. The latter option makes it possible to cover a wider range of frequencies and can increase the sensitivity of the set.

Headphones

The final critical component of a crystal set is a suitable earphone or pair of headphones. This isn't a easy as it might appear: The readily available headphones for portable radios, tape players, CD players, etc. won't work. At all. The reason is that these modern headphones are designed to be used with amplifiers and have a low impedance. A crystal set requires an earphone with a high impedance to match the other components of the circuit.

There are two major alternatives. The least expensive is a small earphone that uses a piezo-electric crystal to convert the electrical signal into sound. Most earphones in stores are not of this type, but crystal earphones can be found at some shops that sell electronic parts for hobbyists or ordered from the X-tal Set Society, as shown on the left in the picture below:

This type of earphone is found in most inexpensive crystal radio kits. Some seem to work better than others.

Better results can be obtained using a pair of high-impedance electromagnetic headphones, as shown on the right side of the figure above. These used to be quite common, but, like variable capacitors, are much less so now. But, you may be able to find a pair in someone's attic or at an Army-Navy Surplus store. I bought the pair shown above at a surplus store in San Diego. Suitable headphones will have a DC resistance, which you can measure with a simple Ohm meter, of about 2000 Ohms. With a good set of headphones, the sound from a crystal set can be plenty load and surprisingly clear.

Plans

Here are instructions for building the crystal radio shown above, in a pdf (Adobe Acrobat) file: Crystal Radio Instructions (./downloads/radioInstructions.pdf)

These instructions were included in a book entitled "Crystal Radio Projects" (http://www.midnightscience.com/projbook.html) published by the X-tal Set Society. The book includes lots of other radio plans, most of which are more elaborate than this one.

THE UNIVERSITY OF UTAH (http://www.utah.edu)

Department of Biology (http://www.biology.utah.edu)
257 South 1400 East
Salt Lake City, UT
84112-0840 (http://bit.ly/1X3LTFL)
Webmaster (mailto:goldenberg@biology.utah.edu)
Disclaimer (http://www.utah.edu/disclaimer/)

KEVIN'S WEBSURFER HANDBOOK V
FOR CRYSTAL RADIO

INDUCTANCE, INDUCTORS and MEASURING Q

Kevin Smith
2013

NOTES:

NOTES:

INTRODUCTION

Finally, I have cobbled together a handbood dedicated to that one crystal radio component virtually all builders fabricate for themselves, the coil. This has been a long time in comming together. As I have increased my knowledge of set design and theory, I have found a deeper need for good technical background reading. Inductors, as will quickly be found, are highly mathematical beasties. Their properties, those we most wish to know such as Quality (Q), Resistance, Losses, etc are devilishely difficult to measure. As such, I have not spent as much time digging into this as I have for diodes, antennas, grounds etc.

I gather together in this volume articles with the central theme of inductance. Some of these are introductory and have appeared in earlier handbooks, others are specialized and will be found here alone. I wish to include both theory as well as practical articles on their design and construction. As an addition, you will find two non-crystal radio projects which otherwise rely on induction and induction coils for their performamce. These describe a cool VLF-ULF coil. Hopefully this volume gives the crystal set designer / builder a good base of understanding for this chief set component.

Of general note, the web is a marvelous source of data and information. Many long-time crystal set builders, and many others have created dedicated sites to disseminate information and resources, to share their creations and knowledge. I am eternally in your debt. All of the material in this handbook is copyright for which I have not sought permission. Therefore this is not presented for publication or copy. It is only my personal resource. I encourage anyone finding this copy to pursue ON THE WEB the web pages identified within. I include the name of the author and web address of each section. I wish to sincerely thank every author presented for their excellent pages and ask forgiveness for my editing into this handbook.

Kevin Smith
2013

www.lessmiths.com/~kjsmith/crystal/cr0intro.shtml

By the way, the speakers sound terrific! I'm not sure how much of the improvement was from the litz wire, but it certainly didn't hurt the sound!

CONTENTS:

1 **Designing an Air-Core Inductor**
Kenneth Kuhn

15 **Professor Coil**
Dan Petersen

21 **Variometer Design**
Claudio Girardi

25 **Get/Make a "Coilmaster"**
QSLnet

39 **Understanding LC Tank Q, Impedance**
David Wagner

47 **Measuring the Q of LC Circuits**
Dick Kleijer

57 **Q Factor Measurements on L_C Circuits**
Jacques Audet

79 **Experiments with Coils and Q-Measurement**
Wes Hayward

111 **Radio-Frequency Resistance and Inductance**
Hund A, and DeGroot HB 1925

137 **Coil Q**
Kevin Smith

148 **An Analytical Approach to coil Q**
Kevin Smith

166 **RLC Circuits**
York County ARS K4YTS

180 **Damped oscillations in RLC circuits**
Barbara Dziurdzia

190 **RLC Circuits**
Rice University

200 **Oscillations and Resonance in LRC Circuits**
Durham University

214 **Logarithmic Decrement…**
Roy Beardmore

218 **Transients and Oscillations in RLC Circuits**
Will Chemelewski

236 **Mutual Inductance**
Dr. David Cattell

244 **Mutual Inductance of Two Coils**
Wayne Storr

252 **Litz Wire – Here's the best way to prep**
Maxamillion

256 **Notes**

I tried several of these methods, and none were satisfactory in producing a nice clean tinned wire end, without breaking or damaging the strands. However, I did eventually come up with a simple, fast and very effective method, so I thought I would share it with the other DIY'ers here, in case they are crazy enough to build something with litz wire.

A Dremel tool with a stainless steel wire wheel attachment works very well. I found that the stainless wheel works faster and lasts longer than the carbon steel wheel, with the added advantage that the small pieces of the wheel brush that wear off during the process are non-magnetic, so they won't attach to nearby driver magnets (don't need little pieces of wire inside my EMIMs and EMITs!). Simply remove ~ 1/2" of the insulation, then untwist the various strands from each other (the Cardas Litz has 3 counterwound layers of different diameter wire to untwist, do them one at a time) and flatten them into a flat fan shape. Set the Dremel to about 1/2 of maximum speed and apply to the coated wire, ensuring that the wheel is ALWAYS turning toward the ends of the wire (otherwise they will entangle with the spinning wire wheel and snap off!). 10 seconds on each side of the fanned wire is very effective at removing the varnish coating without destroying the wires. Then re-fan the wire in a different direction and repeat the process 2 or 3 times to ensure that all the wires get stripped. You will be left with nice clean varnish-free wires, which can then be tinned as usual.

As a test of this method I measured the AC impedance of each of the wires of the same length at 10kHz using my LCR meter, and found that they all were very consistent, indicating that I was effectively and consistently removing the varnish and using all the wires in the bundle.

Litz Wire - Here's the Best Way I've Found to Prep

Maxamillion
http://www.audiokarma.org/forums/archive/index.php/t-282190.html

I just completed building outboard crossovers for my Infinity RSIIb speakers, and I was determined to wire the crossovers together and to the speakers using Cardas Copper Litz wire, due to its flatter AC impedance as frequencies increase. The Cardas Litz wire consists of many fine strands of wire, each individually insulated, which all need to be stripped and tinned together at each end in order to form a proper termination. Quite a job!

I searched the literature for methods and there were many, including:

1) The use of a solder pot at >800 degrees (too expensive to purchase one just for this project),

2) Using a hot soldering iron and lots of flux (very, very messy and very, very ineffective in my case!),

3) Scraping the coating off with an exacto knife, (not attempted due to the tedium factor),

4) Burning the coating off with a torch, followed by sanding (I found it works, but you end up oxidizing the wires and even incinerating some of the smaller strands),

5) Individually sanding each strand (I'd be doing this job for years as each wire has dozens of strands and I have 6 tweeters, 6 midranges and 4 woofers to wire up to 18 capacitors and 20 inductors!).

DESIGNING AN AIR-CORE INDUCTOR

by Kenneth A. Kuhn
http://www.kennethkuhn.com/students/crystal_radios/designing_inductors.pdf

Introduction

This chapter describes the mathematical process for designing an air-core inductor comprised of a single layer solenoid winding over a rigid coil form. Although the development of the mathematics is a bit complicated the final result is simple to apply. For practical reasons, this chapter will make use of English rather than metric units. A well designed and constructed air-core coil has better performance than those with ferrite cores. Ferrite acts as a flux multiplier and has the advantage that the physical size of the inductor can be reduced. That is very important for small radios and the chief reason ferrite is used. The price paid for small size is loss of performance but that loss is generally negligible in active radios. The loss is not bad for crystal radio performance and many good crystal radios have been built using ferrite core inductors. But ferrite is not required. Purists correctly argue that the coil should be air core as that is how early radios were built. A mediocre ferrite core inductor will work considerably better than a poorly designed air-core one and that has probably led to the popularity of ferrite as the process for designing good air-core inductors is not widely known. This chapter reveals those secrets.

Analytic equation

The classic equation (which you can find in any book or article about winding inductors) for calculating the inductance of a given single layer coil is (Reference 2):

$$L = \frac{r2*n2}{9*r + 10*l} \qquad \text{Eq. 1}$$

where:
L = inductance in microhenries
r = coil radius in inches (center of coil to center of conductor)
n = number of turns
l = coil length in inches (center of starting turn to center of ending turn)

This equation is generally accurate to around one percent for inductors of common dimensions. It is more convenient to work with coil diameter and Equation 1 can be written as:

$$L = \frac{d2*n2}{18*d + 40*l} \qquad \text{Eq. 2}$$

where d is the coil diameter in inches (center of conductor to center of conductor)

Example: What is the inductance of a coil has a diameter of 2.5 inches, a length of 2.33 inches, and has 72 turns?

$$L = \frac{2.5*2.5*72*72}{18*2.5 + 40*2.33} = 234 \text{ uH}$$

Development of design equations

Equations 1 and 2 are fine for determining the inductance of an existing coil but are very awkward to apply to the design of a desired coil as there are many variables. Any time there are a multitude of variables then the possibility of optimum combinations or relations should be explored. In the following development the number of variables is reduced by finding

When the coefficient of coupling, k is equal to 1, (unity) such that all the lines of flux of one coil cuts all of the turns of the other, the mutual inductance is equal to the geometric mean of the two individual inductances of the coils. So when the two inductances are equal and L_1 is equal to L_2, the mutual inductance that exists between the two coils can be defined as:

$$M = \sqrt{L_1 L_2} = L$$

Example No1

Two inductors whose self-inductances are given as 75mH and 55mH respectively, are positioned next to each other on a common magnetic core so that 75% of the lines of flux from the first coil are cutting the second coil. Calculate the total mutual inductance that exists between them.

$$M = k\sqrt{L_1 L_2}$$

$$M = 0.75\sqrt{75\text{mH} \times 55\text{mH}} = 48.2\text{mH}$$

In the next tutorial about Inductors, we look at connecting together *Inductors in Series* and the affect this combination has on the circuits mutual inductance, total inductance and their induced voltages.

---

relations to known constants. We first replace the coil length by a factor that relates it to the diameter.

l = k*d Eq. 3

where
l = coil length in inches
k = a dimensionless constant
d = coil diameter in inches as before
Substituting Equation 3 into Equation 2 gives:

```
          d2*n2
L = ---------------------
     18*d + 40*k*d
```

which reduces to

```
        d*n2
L = --------------          Eq. 4
     18 + 40*k
```

It can be shown that the value of k that minimizes the length of wire to wind the coil is 0.450. However, other research indicates (see Reference 1) that the value of k that minimizes coil losses is approximately 0.96 even though that value uses about twenty percent more wire. Factors contributing to coil losses include:

- * Ohmic losses in the wire including skin-effect
- * Dielectric losses in the coil form and nearby materials
- * Dielectric losses in the insulation around the wire
- * Induction losses in nearby materials

There are also losses caused by adjacent turns being too close together. It has been found (see Reference 1) that the optimum spacing (wire center to wire center) of adjacent turns is between about 1.3 to 2.0 times the diameter of the conductor. Coils for crystal radios are commonly wound using what is known as magnet wire (thin enamel insulation) and the turns are tightly wound next to each other corresponding to a spacing factor slightly greater than 1.0 (the thin insulation is of finite thickness). Although it is less than the optimum discussed it works well.

Without some special technique (such as a lathe) it can be very difficult to manually wind a coil with controlled spacing between the turns. One easy method for achieving a spacing factor of 2.0 is to wind two wires tightly side by side at the same time and then remove one of the windings when finished. Smaller spacing factors can be achieved using a smaller diameter wire for the spacer but the difficulty of controlling two wires will increase. It might occur to someone to use a wire with a thicker insulation so that a spacing is naturally formed with a tight winding. The problem with this method is that the insulation may increase dielectric losses and become self defeating –although this may be a small issue –be sure to try it before tossing the concept. This method can work great if Teflon wire is used as that is a very low-loss material and the internal wire strands are silver plated.

Equation 4 can be used to determine the optimum coil diameter for a given inductance and wire size. We note that the coil length is the number of turns divided by t (turns per inch of the wire). We also note that the coil length has previously been related to the coil diameter by the constant, k. Thus:

n = k*d*t Eq. 5

Substituting Equation 5 into Equation 4 gives:

However, the above equation assumes zero flux leakage and 100% magnetic coupling between the two coils, L_1 and L_2. In reality there will always be some loss due to leakage and position, so the magnetic coupling between the two coils can never reach or exceed 100%, but can become very close to this value in some special inductive coils. If some of the total magnetic flux links with the two coils, this amount of flux linkage can be defined as a fraction of the total possible flux linkage between the coils. This fractional value is called the **coefficient of coupling** and is given the letter k.

Coupling Coefficient

Generally, the amount of inductive coupling that exists between the two coils is expressed as a fractional number between 0 and 1 instead of a percentage (%) value, where 0 indicates zero or no inductive coupling, and 1 indicating full or maximum inductive coupling. In other words, if $k = 1$ the two coils are perfectly coupled, if $k > 0.5$ the two coils are said to be tightly coupled and if $k < 0.5$ the two coils are said to be loosely coupled. Then the equation above which assumes a perfect coupling can be modified to take into account this coefficient of coupling, k and is given as:

Coupling Factor Between Coils

$$k = \frac{M}{\sqrt{L_1 L_2}} \text{ or } M = k \text{ x } \sqrt{L_1 L_2} \quad H$$

the size, number of turns, relative position or orientation of the two coils. Because of this, we can write the mutual inductance between the two coils as: $M_{12} = M_{21} = M$.

Hopefully we remember from our tutorials on *Electromagnets* that the self inductance of each individual coil is given as:

$$L_1 = \frac{\mu_o \mu_r N_1^2 A}{\ell} \text{ and } L_2 = \frac{\mu_o \mu_r N_2^2 A}{\ell}$$

Then by cross-multiplying the two equations above, the mutual inductance that exists between the two coils can be expressed in terms of the self inductance of each coil.

$$M^2 = L_1 L_2$$

giving us a final and more common expression for the mutual inductance between two coils as:

Mutual Inductance Between Coils

$$M = \sqrt{L_1 L_2} \quad H$$

$$L = \frac{k2*t2*d3}{18 + 40*k} \qquad \text{Eq. 6}$$

We will use 0.96 for k and t will be that of the particular wire we have available. Solving Equation 6 for the optimum diameter gives:

$$d\_optimum = 4*(L/t2)1/3 \qquad \text{Eq. 7}$$

Figure 1 shows a plot of this Equation 7 for common wire sizes. In all cases the turns are close-spaced. The lower curves are for common enamel insulated magnet wire. The two upper curves are for vinyl insulated house wire which can be considered if a large diameter coil form is available. To use the curves, select the desired inductance and the wire size that will be used. Look up the optimum coil form diameter and then use the closest practical form you have to that size. The optimum is broad so do not worry about being exactly on it. Note that the true diameter is the sum of the diameter of the coil form and the diameter of the wire since by definition the coil diameter is measured between opposite centers of the wire.

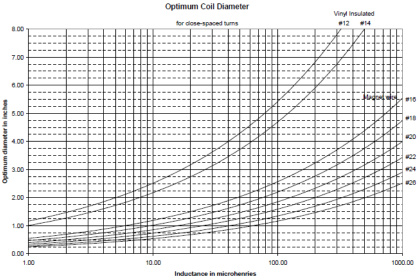

Figure 1: Optimum coil diameter

The following table provides typical values for t (turns per inch) for some common wire sizes:

Table 1: Wire data

Gauge	t	Comments
12	6.2	Vinyl insulated house wire
14	7.7	Vinyl insulated house wire
16	19	Enamel insulated magnet wire
18	24	ditto
20	31	ditto
22	39	ditto
24	50	ditto
26	62	ditto

The length of the winding will be the number of turns divided by the turns per inch of the wire. That is:

Mutual Induction

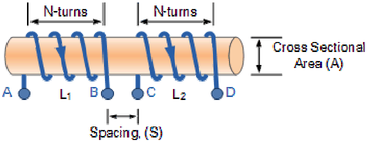

Here the current flowing in coil one, L_1 sets up a magnetic field around itself with some of these magnetic field lines passing through coil two, L_2 giving us mutual inductance. Coil one has a current of I_1 and N_1 turns while, coil two has N_2 turns. Therefore, the mutual inductance, M_{12} of coil two that exists with respect to coil one depends on their position with respect to each other and is given as:

$$M_{12} = \frac{N_2 \Phi_{12}}{I_1}$$

Likewise, the flux linking coil one, L_1 when a current flows around coil two, L_2 is exactly the same as the flux linking coil two when the same current flows around coil one above, then the mutual inductance of coil one with respect of coil two is defined as M_{21}. This mutual inductance is true irrespective of

The mutual inductance that exists between the two coils can be greatly increased by positioning them on a common soft iron core or by increasing the number of turns of either coil as would be found in a transformer. If the two coils are tightly wound one on top of the other over a common soft iron core unity coupling is said to exist between them as any losses due to the leakage of flux will be extremely small. Then assuming a perfect flux linkage between the two coils the mutual inductance that exists between them can be given as.

$$M = \frac{\mu_0 \mu_r N_1 N_2 A}{\ell}$$

- Where:
- μ_o is the permeability of free space ($4.\pi.10^{-7}$)
- μ_r is the relative permeability of the soft iron core
- N is in the number of coil turns
- A is in the cross-sectional area in m^2
- l is the coils length in meters

---

l = n/t Eq. 8

We now substitute Equation 8 into Equation 2 and solve for n

d2*n2 –18*d*L –40*(n/t)*L = 0 Eq. 9

t*d2*n2 –40*L *n –18*d*t*L = 0 Eq. 10

Solving for n gives:

$$n = \frac{20*L + sqrt(400*L2 + 18*t2*d3*L)}{t*d2} \qquad \text{Eq. 11}$$

Although a precise value (in inches) for the length of wire required can be calculated using trigonometry for a spiral, a very close value can be calculated as

w = pi*d*n Eq. 12

Remember that d is the sum of the coil form diameter and the diameter of the wire. This approximation assumes that the diameter of the wire is very small in comparison to that of the coil form. Also remember to allow an extra couple of inches for connecting leads at each end of the coil.

Example: A 300 uH coil is needed. The expected Q should be over 350. What coil diameters and wire sizes could possibly meet this?
Solution: Using Figure 2 it can be seen that wire sizes #12, #14, #16, #18, and #20 could achieve the required Q. Using Figure 1 the required coil form diameters are:

Optimum

Gauge	diameter
#12	7.75”
#14	6.75”
#16	3.75”
#18	3.15”
#20	2.65”

A piece of 4.5” OD PVC pipe is available and #14 electrical wire is available. From Table 1, #14 insulated wire will make about 7.7 turns per inch. Thus, the effective diameter is 4.5 plus 1/7.7 = 4.63 inches. Using Equation 11 the number of turns required is 86. Using Equation 8 the length of the winding is 11.2 inches. The length/diameter ratio is 2.4 which is a bit longer than the optimum of 0.96. The length of wire required is given by Equation 12 and is 1,251 inches. The length would have been 1,058 inches if the optimal diameter could have been used. This extra length will cause somewhat higher losses –it might still meet the desired spec though. This is about as far as I would go in rounding to an available coil form diameter.

Estimation of Inductor Q

All inductors have an equivalent series resistance loss as discussed earlier and is comprised of a number of components. We measure the quality factor or Q of the inductor by computing the ratio of inductive reactance at the frequency of interest to the series loss resistance as follows:

$$Q = \frac{XL}{Rs} \qquad \text{Eq. 13}$$

where

Q is the dimensionless “quality” factor of the inductor

---

magnetic flux generated by the first coil will interact with the coil turns of the second coil inducing a relatively large emf and therefore producing a large mutual inductance value.

Likewise, if the two coils are farther apart from each other or at different angles, the amount of induced magnetic flux from the first coil into the second will be weaker producing a much smaller induced emf and therefore a much smaller mutual inductance value. So the effect of mutual inductance is very much dependant upon the relative positions or spacing, (S) of the two coils and this is demonstrated below.

Mutual Inductance between Coils

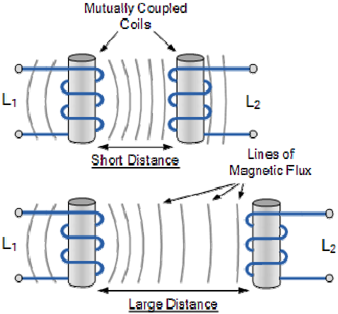

Mutual Inductance of Two Coils

Wayne Storr

http://www.electronics-tutorials.ws/inductor/mutual-inductance.html

In the previous tutorial we saw that an inductor generates an induced emf within itself as a result of the changing magnetic field around its own turns, and when this emf is induced in the same circuit in which the current is changing this effect is called **Self-induction**, (L). However, when the emf is induced into an adjacent coil situated within the same magnetic field, the emf is said to be induced magnetically, inductively or by **Mutual induction**, symbol (M). Then when two or more coils are magnetically linked together by a common magnetic flux they are said to have the property of **Mutual Inductance**.

Mutual Inductance is the basic operating principal of transformers, motors, generators and any other electrical component that interacts with another magnetic field. Then we can define mutual induction as the current flowing in one coil induces an emf in an adjacent coil. But mutual inductance can also be a bad thing as "stray" or "leakage" inductance from a coil can interfere with the operation of another adjacent component by means of electromagnetic induction, so some form of electrical screening to a ground potential may be required.

The amount of mutual inductance that links one coil to another depends very much on the relative positioning of the two coils. If one coil is positioned next to the other coil so that their physical distance apart is small, then nearly nearly all of the

XL is the inductive reactance in ohms at the frequency of interest

Rs is the equivalent series resistance in ohms at the frequency of interest

Note that inductive reactance, XL, is calculated as

$XL = 2*\pi*F*L$

where

F is the frequency in Hz

L is the inductance in henries

The equivalent series resistance is the net of ohmic losses including skin effect, dielectric losses in distributed capacitance and coil structure, absorption losses by nearby conducting media, magnetic losses in nearby magnetic media, etc. With care these losses can be kept small but it takes very little loss to reduce the Q of an inductor from 400 to 200. The magnitude of Rs can be measured on sophisticated impedance equipment but it is hard to calculate the effect of all factors. Figure 2 shows an estimated value of Q at 1 MHz considering typical losses assuming the coil is wound optimally and is not disturbed by nearby lossy materials. Use the figure only as a guideline as your specific results may be better or worse. The expected Q at 540 kHz will be between about 50 to 70 percent of what is shown and the expected Q at 1.6 MHz will be around 1.2 to 1.5 times that shown.

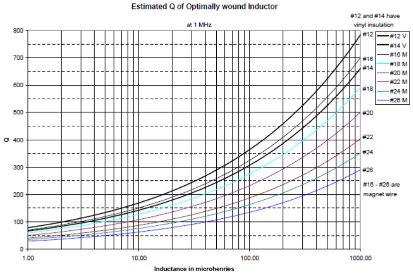

Figure 2: Estimated Q of Inductor

The Q we obtain from Equation 11 is for the unloaded coil (i.e. antenna and crystal detector not connected). The net loaded Q will typically be significantly smaller but ideally (as discussed in another chapter) would be in the general range of one hundred. Thus, we would like to start with an unloaded Q of several hundred. As can be seen in Figure 2 the Q of the inductor can be made higher by using larger diameter wire. From Figure 1 this also means using a large diameter coil form. This is a very important conclusion –high Q coils need to be physically large.

Type of wire

The only material to consider for the wire is copper. A variety of styles of copper wire is readily available. The most basic choice is between solid or stranded. Although a variety of arguments can be made for and against each, in practical terms you will not notice any difference in performance although one

---

The power delivered to an ideal transformer[2] on its primary circuit will match the power output on its secondary circuit and we can write $I_p V_p = I_s V_s$ or

$$\frac{I_s}{I_p} = \frac{V_p}{V_s} = \frac{N_p}{N_s}$$

---

[2] Here we neglect the resistances of the windings and the energy expended in magnetizing and demagnetizing the iron core.

$$E_s = -N_s \frac{\Delta\Phi}{\Delta t}$$

In the primary coil the induced emf is due to self-induction and is given by Faraday’s law as

$$E_p = -N_p \frac{\Delta\Phi}{\Delta t}$$

The term $\Delta\Phi/\Delta t$ is the same in both equations since the same amount of magnetic flux Φ passes through both coils. Dividing the equations gives

$$\frac{E_s}{E_p} = \frac{N_s}{N_p}$$

If the resistances of the coils are negligible the terminal voltages V_s and V_p of the coils are nearly equal to the magnitudes of the emfs E_s and E_p. Hence we can write

$$\boxed{\frac{V_s}{V_p} = \frac{N_s}{N_p}}$$

This is called the ***transformer equation***.

or the other may have physical advantages for your particular construction method. Avoid wires that are plated as those will have higher losses since skin-effect will cause most of the conduction to be in the plating which has higher resistance than copper. Avoid wires with rubber or cheap plastic insulations as dielectric losses will be higher. An exception is silver plated Teflon wire as that has the best conductivity and the lowest dielectric losses –but it is expensive.

For use in low to medium frequency inductors there is a special wire called Litzengrad or just Litz for short. It is designed to minimize skin-effect losses and is made by assembling many strands of enamel insulated magnet wire together to form a wire that has a large surface area. Litz wire is not easy to find and tends to be expensive. If you are going to use Litz wire then make sure that other losses as previously discussed are minimized. Otherwise Litz wire will make little if any difference and will be wasted effort and expense. Avoid belief in a variety of myths about skin-effect. Although it is true that skin-effect is more severe on large diameter conductors, a larger diameter still conducts better than a smaller diameter at any frequency. This can be seen in Figure 3 which shows the frequency dependence of the resistance per meter factor of some common wire sizes.

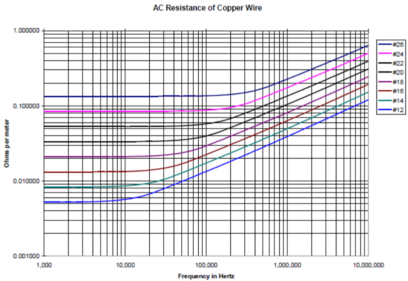

Figure 3: AC Resistance of Copper Wire

Coil Forms

From a loss standpoint air is the best coil form material there is. The obvious problem is that air has no structural strength. However, there are methods used by commercial inductor companies that employ a minimal structure so that the coil form is around 99 percent air. Manually, you can achieve the effect by first winding the coil using large diameter solid copper wire (i.e. #18, #16, #14, etc.) on a rigid coil form and then carefully sliding the winding off of the form. The stiff wire will retain the shape and you can easily space the turns to the optimal discussed previously. You will need a few supports to keep the whole thing from being too loose.

A popular coil form is some kind of cardboard tube that you have salvaged from a variety of sources such as used for paper towels or shipping tubes. These are great if you are using small diameter wire as small wire will not self support. Plastic pipe

Transformers

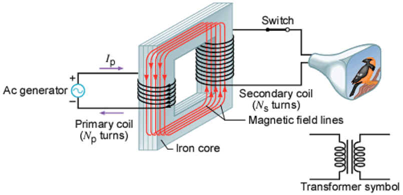

A ***transformer*** is a device for increasing or decreasing an AC voltage. It consists of an iron core with two coil wrappings: a primary coil with N_p turns and a secondary coil with N_s turns. See the figure at the right.

Assume an AC source, such as an AC generator, produces an alternating current I_p in the primary coil. The primary coil creates a changing magnetic flux Φ in the core which links the turns of both the primary and secondary coils.[1] In the secondary coil the induced emf arises from mutual induction and is given by Faraday's law as

---

[1] The details of the operation of a transformer, which also involve the magnetization and demagnetization of the iron core, is complex. Here we assume, for simplicity, a transformer operating in steady-state with a purely resistive load.

Assume that the solenoid carries a current I. Then the magnetic flux in the solenoid is

$$\Phi = \mu_0 \frac{NI}{l} A. \; L = \frac{N\Phi}{I} = \frac{N}{\cancel{I}} \mu_0 \frac{N\cancel{I}}{l} A$$

$$L = \mu_0 \frac{N^2}{l} A \text{ or } \boxed{L = \mu_0 n^2 A l} \text{ where } n = \frac{N}{l}.$$

(Note how L is independent of the current I.)

---

is another material you might consider. None of these materials are made with any consideration about high frequency dielectric losses but because only a small amount of material is used the losses are probably minimal.

Wood is a convenient coil form and has low losses if very dry. Common sizes that have been used are 2x2, 4x4, or a pair of 2x4 combined to make 4x4. Round dowel rods may also be used but their diameters are often much less than optimum. When using a square coil form there is a logical question about how that affects the inductance calculations. A simplistic (but good) answer is to use an effective circle diameter that has the same area as the square form since area is a strong factor in inductance. Losses with a square form will be somewhat higher than for a circular form. Rectangular forms (such as a single 2x4) have even higher losses in comparison –it takes more wire to encompass a given area.

Winding the coil

Counting turns is a tedious and error prone task. It is much simpler to cut the length of wire needed and then wind that until finished. The resulting turns count will be very close if not exact. Wire is springy and will jump off the form in a tangled mess if not restrained. Start by securing the wire at one end of the form and have a means for easily (preferably with one hand) securing the opposite end when you finish. It is tempting to use some kind of adhesive tape and that will work if you are careful and understand what you are doing. The forces will build and the tape may give way which will result in a frustrating mess of tangled wire. Make sure the tape can not slip. A good way to secure the ends is to first drill a hole in the tube at the starting and end points. Then feed the starting end through the starting hole and bend the wire such that it naturally resists tension and secure the wire with tape. Stuff the loose end length inside the tube so it is out of the way while winding.

It is best to wind the coil by hand as the set up for using a lathe is not worth the trouble for a single coil. There are a number of "poor man's" lathes such as a power drill that have been used but I do not recommend that as you are more likely to make a mess or cause injury than you are to wind a coil. It only takes a couple of minutes to wind a coil by hand so take the time to think what you are doing. It is important to keep the winding tight at all times. The wire will spring off if it ever gets loose. You will very likely have some fractional turn as a result of your calculations. I recommend that you round that to the nearest integer as it is not worth the trouble of making measurements to stop at a specific fractional turn.

References:

1. Electronic and Radio Engineering, fourth edition, Frederick Emmons Terman,
McGraw-Hill Book Company, 1955, pages 30 – 33.
2. The Radio Amateur's Handbook, 44th edition, 1967, The American Radio
Relay League, Newington, Conn., page 26

Let the coil have N turns. Assume that the same amount of magnetic flux Φ links each turn of the coil. The net flux linking the coil is then $N\Phi$. This net flux is proportional to the magnetic field, which, in turn, is proportional to the current I in the coil. Thus we can write $N\Phi \propto I$. This proportionality can be turned into an equation by introducing a constant. Call this constant L, the ***self-inductance*** (or simply ***inductance***) of the coil:

$$N\Phi = LI \quad \text{or} \quad L = \frac{N\Phi}{I}$$

As with mutual inductance, the unit of self-inductance is the henry.

The self-induced emf can now be calculated using Faraday's law:

$$\mathsf{E} = -N\frac{\Delta\Phi}{\Delta t} = -\frac{\Delta\left(N\Phi\right)}{\Delta t} = -\frac{\Delta\left(LI\right)}{\Delta t} = -L\frac{\Delta I}{\Delta t}$$

$$\boxed{\mathsf{E} = -L\frac{\Delta I}{\Delta t}}$$

The above formula is the ***emf due to self-induction***.

Example

Find the formula for the self-inductance of a solenoid of N turns, length l, and cross-sectional area A.

The above formula is the ***emf due to mutual induction***.

Example

The apparatus used in Experiment EM-11B consists of two coaxial solenoids. A solenoid is essentially just a coil of wire. For a long, tightly-wound solenoid of *n* turns per unit length carrying current *I* the magnetic field over its cross-section is nearly constant and given by $B = \mu_0 nI$. Assume that the two solenoids have the same cross-sectional area *A*. Find a formula for the mutual inductance of the solenoids.

The magnetic flux in the primary coil is

$\Phi_1 = \mu_0 \frac{N_1 I_1}{l_1} A$ where l_1 is the length of the primary coil.

$$M = \frac{N_2 \Phi_2}{I_1} \quad \text{But } \Phi_2 = \Phi_1 : \quad M = \frac{N_2 \Phi_2}{I_1} = \frac{N_2}{\cancel{I_1}} \mu_0 \frac{N_1 \cancel{I_1}}{l_1} A; \quad \boxed{M = \mu_0 \frac{N_1 N_2}{l_1} A}$$

(Note how M is independent of the current I_1.)

Self-Inductance

A current-carrying coil produces a magnetic field that links its own turns. If the current in the coil changes the amount of magnetic flux linking the coil changes and, by Faraday's law, an emf is produced in the coil. This emf is called a *self-induced* emf.

"Professor Coyle" Coil Calculators

By Dan Petersen

http://www.crystalradio.net/professorcoyle/index.shtml

Instructions for Professor Coyle:

Professor Coyle is intended to be a mathematical aid in determining the inductance of a coil and the resonant frequency of a coil-capacitor tuned circuit.

To Use The Coil Calculator:

The coil calculator will calculate the inductance of both closewound or non-closewound or "spaced" coils. To do a closewound coil, enter the coil diameter and the desired number of turns into the appropriate places. Left-click on any of the red numbers to change their value, then press <enter>. Look at the list below the entered numbers ans you will note a series of inductances and lengths paired with the gauge of wire to be used. These are the inductances and coil lengths of a closewound coil as a function of the wire guage. For example, enter a form diameter of 1.75" and 78 turns into the calculator. Now if you wanted to use #28 enamelled wire, you would need a space of 1.11" on the coil and the inductance would be 244.9 uH.

A "spaced" coil is one that has a gap between turns. An example is a coil that is 1 inch long and has 8 turns of #24 wire "spaced" along the 1 inch length.There is a significant gap between turns. To calculate the inductance, enter the coil diameter, the number of turns and the coil length into the calculator. The calculated inductance will appear in blue just under the entered parameters.

To Use the Resonance Calculator:

Enter the known inductance and capacitance into the calculator section at the top of the page. The resonant frequency and the reactance of the L-C circuit will be given in blue immediately below. Example: You have a 40-400pF variable capacitor and a 240uH inductor and you want to find the tuning range. Enter the inductor value, press <enter>, and the highest variable capacitor value into the calculator and press <enter>. You should see a resonance of .514 MHz. Left-click on the capacitance value and now enter 40. You should now see a resonance of 1.624 MHz. In this example would be perfect for tuning the AM broadcast band.The result is a theoretical value. The actual frequency will differ slightly due to distributed capacitance and measurement tolerances

Professor Coyle can be copied for non-commercial purposes only and is meant to be an aid in calculation. The author is not responsible for the results calculated nor any damage to equipment affected by these calculations. This program is not to be bought or sold. Please credit the author, Dan Petersen, when copying this program. Close cover before striking. Place all trays in the upright position before landing. Put the toilet seat down. Exact change only. Don't kiss the dog on the nose.

needle). Hence the secondary coil encloses a *changing* magnetic field. By Faraday's law of induction this changing magnetic flux induces an emf in the secondary coil. This effect in which changing current in one circuit induces an emf in another circuit is called ***mutual induction***.

Let the primary coil have N_1 turns and the secondary coil have N_2 turns. Assume that the same amount of magnetic flux Φ_2 from the primary coil links each turn of the secondary coil. The net flux linking the secondary coil is then $N_2\Phi_2$. This net flux is proportional to the magnetic field, which, in turn, is proportional to the current I_1 in the primary coil. Thus we can write $N_2\Phi_2 \propto I_1$. This proportionality can be turned into an equation by introducing a constant. Call this constant M, the ***mutual inductance*** of the two coils:

$$N_2\Phi_2 = MI_1 \text{ or } M = \frac{N_2\Phi_2}{I_1}$$

the unit of inductance is $\frac{\text{wb}}{\text{A}} = henry\ (\text{H})$ named after Joseph Henry.

The emf induced in the secondary coil may now be calculated using Faraday's law:

$$\mathcal{E}_2 = -N_2\frac{\Delta\Phi_2}{\Delta t} = -\frac{\Delta\left(N_2\Phi_2\right)}{\Delta t} = -\frac{\Delta\left(MI_1\right)}{\Delta t} = -M\frac{\Delta I_1}{\Delta t}$$

$$\boxed{\mathcal{E}_2 = -M\frac{\Delta I_1}{\Delta t}}$$

Mutual Inductance
Dr. David F. Cattell
faculty.ccp.edu/faculty/dcattell/Sp12/.../Mutual%20Inductance.doc

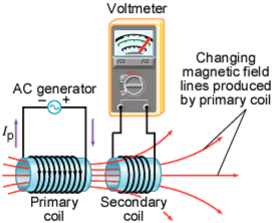

Suppose we hook up an AC generator to a solenoid so that the wire in the solenoid carries AC. Call this solenoid the *primary coil*. Next place a second solenoid connected to an AC voltmeter near the primary coil so that it is coaxial with the primary coil. Call this second solenoid the *secondary coil*. See the figure at the right.

The alternating current in the primary coil produces an alternating magnetic field whose lines of flux *link* the secondary coil (like thread passing through the eye of a

---

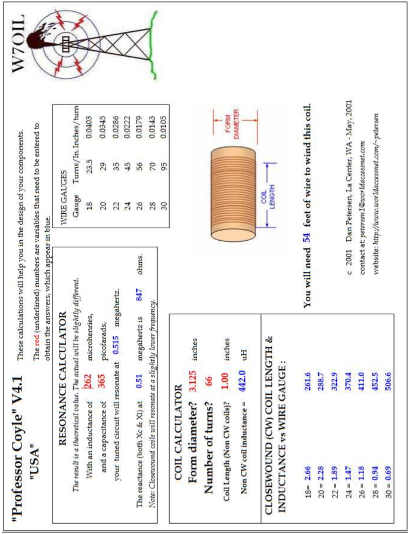

"Professor Coyle" V4.1
"USA"

These calculations will help you in the design of your components.

The red (underlined) numbers are variables that need to be entered to obtain the answers, which appear in blue.

W7OIL

RESONANCE CALCULATOR

The result is a theoretical value. The actual will be slightly different.

With an inductance of 262 microhenries,
and a capacitance of 365 picofarads,
your tuned circuit will resonate at 0.515 megahertz.

The reactance (both Xc & Xl) at 0.51 megahertz is 847 ohms.

Note: Closewound coils will resonate at a slightly lower frequency.

WIRE GAUGES

Gauge	Turns/In.	Inches/turn
18	23.5	0.0403
20	29	0.0345
22	35	0.0286
24	45	0.0222
26	56	0.0179
28	70	0.0143
30	95	0.0105

COIL CALCULATOR
Form diameter? 3.125 inches
Number of turns? 66
Coil Length (Non CW coils)? 1.00 inches
Non CW coil inductance = 442.0 uH

CLOSEWOUND (CW) COIL LENGTH & INDUCTANCE vs WIRE GAUGE :

18= 2.66	261.6
20 = 2.28	288.7
22 = 1.89	322.9
24 = 1.47	370.4
26 = 1.18	411.0
28 = 0.94	452.5
30 = 0.69	506.6

You will need 54 feet of wire to wind this coil.

c 2001 Dan Petersen, La Center, WA - May, 2001
contact at: *petersen1@worldaccessnet.com*
website: *http://www.worldaccessnet.com/~petersen*

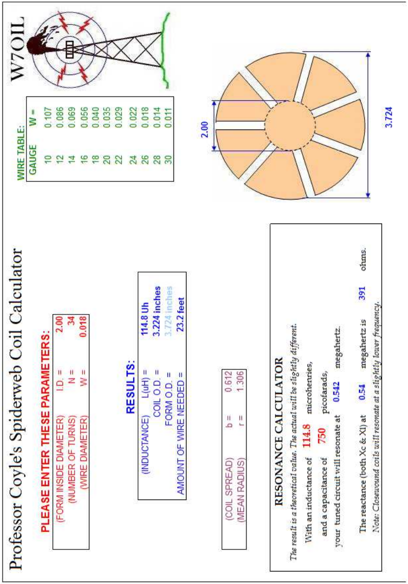

Professor Coyle's Spiderweb Coil Calculator

PLEASE ENTER THESE PARAMETERS:

(FORM INSIDE DIAMETER) I.D. = 2.00
(NUMBER OF TURNS) N = 34
(WIRE DIAMETER) W = 0.018

RESULTS:

(INDUCTANCE) L(uH) = 114.8 Uh
COIL O.D. = 3.224 inches
FORM O.D. = 3.724 inches
AMOUNT OF WIRE NEEDED = 23.2 feet

(COIL SPREAD) b = 0.612
(MEAN RADIUS) r = 1.306

RESONANCE CALCULATOR

The result is a theoretical value. The actual will be slightly different.

With an inductance of 114.8 microhenries,
and a capacitance of 750 picofarads,
your tuned circuit will resonate at 0.542 megahertz.

The reactance (both Xc & Xl) at 0.54 megahertz is 391 ohms.

Note: Closewound coils will resonate at a slightly lower frequency.

W7OIL

WIRE TABLE:

GAUGE	W =
10	0.107
12	0.086
14	0.069
16	0.056
18	0.040
20	0.035
22	0.029
24	0.022
26	0.018
28	0.014
30	0.011

Figure 10: Peak to peak voltage across capacitor as a function of frequency with theoretical voltage fit to data with R = 202 Ω.

the critical resistance as a function of capacitance there is a problem. But the issue is most likely due to the nature of the measurement as it is difficult to tell exactly when the circuit transitions from under-damped to over-damped. Finally, the model accurately predicts the resonant frequency of a RLC circuit, however, other details of the behavior such as the Q factor are not described correctly in this experiment.

This fit is shown in Figure 10. It can be seen that this does shift the resonance from where measured (to 485 Hz) but the Q factor is the same as in the experiment. So while not entirely responsible for the difference between measurement and theory, as evidenced by the resonant frequency difference, uncounted resistances play a part in the disparity.

4 Conclusions

RLC circuit behavior is described well by oscillating system theory which describes the charge on the capacitor as a 2nd order differential equation. It is demonstrated in this experiment that the theory can be used to reliably find the inductance of a RLC loop by measuring the period of oscillation as a function of capacitance. The predicted exponential damping dependence on resistance is also verified and it is shown that that total resistance of the circuit can be found accurately using a resistance versus log decrement plot. However, when applying the theory to measure the inductance using

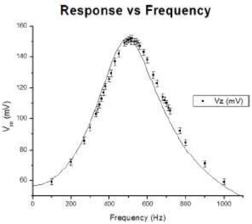

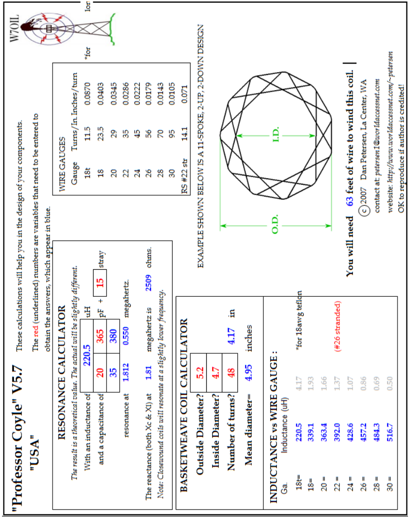

"Professor Coyle" V5.7
"USA"

These calculations will help you in the design of your components.
The red (underlined) numbers are variables that need to be entered to obtain the answers, which appear in blue.

WIRE GAUGES		
Gauge	Turns/In.	Inches/turn
18t	11.5	0.0870
18	23.5	0.0403
20	29	0.0345
22	35	0.0286
24	45	0.0222
26	56	0.0179
28	70	0.0143
30	95	0.0105
RS #22 str	14.1	0.071

*for

W7OIL

RESONANCE CALCULATOR

The result is a theoretical value. The actual will be slightly different.

With an inductance of 220.5 uH
and a capacitance of 20 365 pF + 15 stray
35 380
resonance at 1.812 0.550 megahertz.

The reactance (both Xc & Xl) at 1.81 megahertz is 2509 ohms.

Note: Closewound coils will resonate at a slightly lower frequency.

EXAMPLE SHOWN BELOW IS A 11-SPOKE, 2-UP, 2-DOWN DESIGN

BASKETWEAVE COIL CALCULATOR

Outside Diameter? 5.2
Inside Diameter? 4.7
Number of turns? 48 4.17 in
Mean diameter= 4.95 inches

INDUCTANCE vs WIRE GAUGE :

Ga.	Inductance (uH)		
18t=	220.5	4.17	*for 18awg teflon
18=	339.1	1.93	
20 =	363.4	1.66	
22 =	392.0	1.37	(#26 stranded)
24 =	428.6	1.07	
26 =	457.2	0.86	
28 =	484.3	0.69	
30 =	516.7	0.50	

You will need 63 feet of wire to wind this coil.

ⓒ 2007 Dan Petersen, La Center, WA
contact at: *petersen1@worldaccessnet.com*
website: *http://www.worldaccessnet.com/~petersen*
OK to reproduce if author is credited!

Ed Note: I include these screen shots because I believe this is the premiere coil calculator available. I highly encourage any/all crystal radio builders to get a copy, modify it at will, and use it, use it a lot!

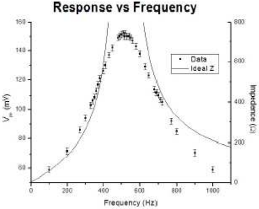

Figure 9: Peak to peak voltage across capacitor as a function of frequency. The data points go to the left axis and the Ideal Z goes with the right axis.

bandwidth of 360 − 15 Hz5. Then, using $Q = f_0/\Delta f$ it was found that the quality of the circuit was 1.4 − 0.1.

The theoretical Q is 52.8 and it is therefore quite obvious why the theoretical impedance curve is much sharper than the measured curve. The plot shows a poor agreement between the theoretical response and the actual results. The peak is much sharper for the ideal case and the bandwidth is much lower as seen in a comparison of Q values. These discrepancies are due to components not behaving ideally and other loses in the circuit such as resistance in the wires or imperfect connections on the bread board. Additionally, the actual voltage falls off faster than expected for the same reasons. Building on this idea, if we assume that the inductance and capacitance of the circuit are the same as measured and fit the function to the data by changing the resistance in equation 14, we find the best fit around R = 202 Ω.

function of frequency. We observed frequencies between 100 and 1000 Hz and set the oscilloscope to average 16 samples and measure the amplitude automatically. The results are shown in Table 6. These values are then

Table 6: Voltage versus frequency. Errors for all values are − 2 mV.

Frequency (Hz)	Vpp (mV)	Frequency (Hz)	Vpp (mV)	Frequency (Hz)	Vpp (mV)
100	59	450	142	600	138
200	72	480	149	630	129
270	86	490	150	650	123
300	94	500	151	680	114
330	103	510	152	690	112
340	106	515	150	700	110
350	109	520	152	710	107
360	113	530	150	720	105
370	117	540	150	770	92
380	121	550	149	800	85
400	126	560	147	900	71
410	130	580	143	1000	59
430	137				

plotted along with equation 14 in Figure 9. Equation 14 is important because RM is so large that the peak to peak voltage (Vpp) should approximately equal V0Z=RM and thus is directly related Vpp. Using this plot we found the resonant frequency to be 515 − 10 Hz. Based on the measurements of the inductor, resistor, and capacitor the theoretical resonance occurs at 514 Hz which is in good agreement with our data. We also measured the bandwidth by finding the two frequencies at which Vpp fell to 70.7% of its peak value. These two frequencies were 350 − 10 Hz and 710 − 10 Hz giving a

---

Variometer Design

Claudio Girardi
http://www.qsl.net/in3otd/variodes.html

Variometer Design

A variometer consists usually of two coaxial coils, connected in series, where the coils relative position can be varied in some way. If L1 and L2 are the self-inductances of the first and second coil and Lm is the mutual inductance between the two, the total inductance can be written as Ltot=L1+L2±2Lm. The mutual inductance is defined as the flux linked by the turns of an inductance when the other carries a unit current; this of course depends not only on the coil length and diameter but also on their relative position.
For two coaxial coils, like the ones in the picture, the mutual inductance is at maximum when they are also concentrical, i.e. D=0.

If the two concentrical coils are rotated, so that their axes are not parallel any more, the mutual inductance decreases, reaching zero (almost) when the angle is 90°. Continuing the rotation beyond 90° the mutual inductance increases again, but this time with the opposite sign. So, according to the above formula,

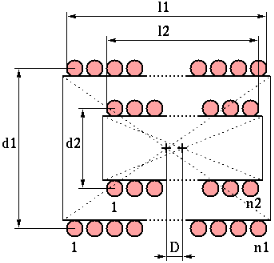

rotating a coil 180° gives a variation of the total inductance of 4Lm.

Two coaxial coils If the two coaxial coils are moved along their axis instead (D<>0), the mutual inductance decreaes, eventually reaching zero at infinity, but does not change sign. This means that there is less variation in the total inductance just moving the coils with respect to rotating them.

The form below computes the main parameters for the two types of variometer described above. The self-inductances are calculated using the formulas in [1], while the mutual inductances use the formulas from [2], with some minor corrections.

These formulas agree very well with the results of some electromagnetic simulations I have done, see the page for details.

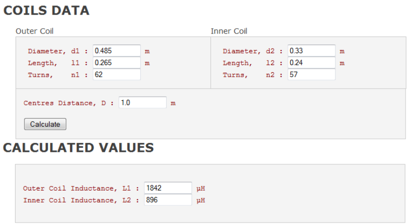

COILS DATA

Outer Coil

Diameter, d1 : 0.485 m
Length, l1 : 0.265 m
Turns, n1 : 62

Inner Coil

Diameter, d2 : 0.33 m
Length, l2 : 0.24 m
Turns, n2 : 57

Centres Distance, D : 1.0 m

CALCULATED VALUES

Outer Coil Inductance, L1 : 1842 µH
Inner Coil Inductance, L2 : 896 µH

In a typical variometer the two coils are concentric (i.e. D=0) and the overall inductance is varied rotating the inner coil; in this case we have:

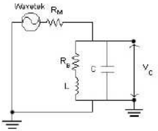

Figure 8: Circuit layout for part D

When analyzing AC circuits it is convenient to use impedances to determine various aspects of the response. For example, with this circuit, the impedance can be used to determine the magnitude of the voltage across the RLC element as

$$V_C = V_0 \frac{R}{R_M + Z} \tag{13}$$

Impedances are complex numbers in general and for a capacitor Z = 1=j!C, for an inductor Z = j!L, and for a resistor Z = R where j = p □1. As can be seen in Figure 8 the RLC loop has the capacitor in parallel with a series addition of the resistor and inductor. Using the addition rules for circuits the result is that

$$|Z| = \frac{\sqrt{R^2 + (\omega L)^2}}{\sqrt{\left(1 - \frac{\omega^2}{\omega_0^2}\right)^2 + (\omega RC)^2}} \tag{14}$$

where !0 is the resonant frequency.

With this in mind we set out to observe this response by measuring the peak to peak voltage across the capacitor as a difference and based on this it would seem all of our measurements of Rcritical were too low by at least 15%. However, when the resistance is adjusted upwards that much it is readily apparent that the circuit has moved into being over-damped based on the waveform. Therefore, it must be concluded that this is not a reliable way to measure an inductance at least partially due to the subjectivity of the measurement.

3.5 Part D: RLC response to sinusoidal signal

For this section a different circuit was used than in parts A, B, and C. The circuit is shown in Figure 8. RM was 101 k , L was 93 mH, RB was 5.69 , and C was 1.03 ┬ F. Also, the Wavetek was set to output a sine wave with an amplitude of 16 V.

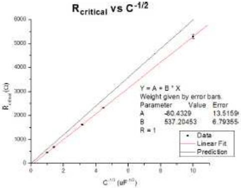

Figure 7: Plot of Rcritical versus 1/p C showing a strong linear relationship. Additionally, a plot of the theoretical behavior given the measured inductance is shown.

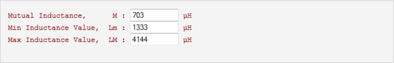

Mutual Inductance, M : 703 µH
Min Inductance Value, Lm : 1333 µH
Max Inductance Value, LM : 4144 µH

When the coils can not rotate but only move along their axis (coaxial coils), assuming a distance D between the two coils centres, we have:

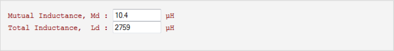

Mutual Inductance, Md : 10.4 µH
Total Inductance, Ld : 2759 µH

The total inductance value for this case is computed assuming a positive mutual inductance (coils not rotated 180°).

References:

[1] R. Lundin, "A Handbook Formula for the Inductance of a Single-Layer Circular Coil," Proc. IEEE, vol. 73, no. 9, pp. 1428-1429, Sep. 1985.

[2] F.E. Terman, "Radio Engineers' Handbook," London, McGraw-Hill, 1st ed., Sep. 1950.

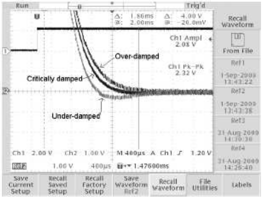
Run
Trig'd
Δ: 1.86ms Δ: 4.00 V
@: 2.00ms @: −20.0mV
Recall Waveform
Ch1 Ampl 2.08 V
From File
Over-damped
Ch1 Pk−Pk 2.32 V
Ref1
1-Sep-2000 13:43:22
Critically damped
Ref2
1-Sep-2000 13:43:38
Under-damped
Ref3
31-Aug-2000 14:30:30
Ref4
31-Aug-2000 14:26:40
Ch1 2.00 V Ch2 1.00 V M 400µs A Ch1 ⨍ 1.20 V
Ref2 1.00 V 400µs 1.47600ms
Save Current Setup
Recall Saved Setup
Recall Factory Setup
Save Waveform Ref2
Recall Waveform
File Utilities
Labels

Figure 6: Sample circuit response for this part of the experiment

Table 5: Rcritical versus capacitance.

C (μF)	$R_{critical}$ (Ω)
1	460 ± 10
0.5	678 ± 10
0.1	1620 ± 30
0.05	2320 ± 30
0.01	5300 ± 100

the error values a change in the response was observed that we believe was significant enough to warrant a high degree of confidence (assumed to be 95%) in that range.

A plot of these critical resistances versus 1/p C is shown in Figure 7. Additionally, the ideal behavior is plotted using L = 90 mH. The first result to notice is that the linear fit is very strong and that the error in the slope is only 2.3% with 95% confidence. However, it is also clear that the plot implies a different value for the inductance of the circuit than previously found. Based on the slope of the fit, the inductance of the circuit is 72 − 3 mH with 95% confidence4. This gives a 20%

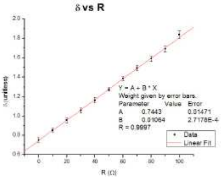

Figure 5: Graph of versus R showing a good fit and match to theory.

3.4 Part C: Dependence of resistance for critical damping on capacitance

For this part of the experiment the same circuit and waveform is used as in part B. The critical resistance is the resistance at which the circuit is critically damped. Therefore b2 = 0 and by solving equation 3 we find

$$R_{critical} = 2\sqrt{\frac{L}{C}} \qquad (12)$$

So a plot of R vs 1/p C should be linear. We determined the critical resistance by adjusting the decade resistance box and observing the resulting waveform. Three example waveforms are shown in Figure 6. To find where the circuit was critically damped we allowed it to be slightly under-damped and then increased the resistance until no overshoot was observed. We did this for 5 different capacitances and our results are summarized in Table 5. Unfortunately we did not have one definitive way to determine the error in our measurement. However, when the resistances were adjusted by 7

Get/Make a "Coilmaster"
http://www.qsl.net/k5bcq/COIL/COIL.html

Get/Make a "Coilmaster"

This is "The New Modern Coilmaster" made by MoReCo (MOrris REgister COmpany), Inc. of Council Bluffs, Iowa.

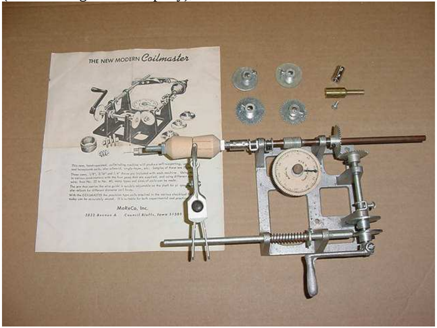

If you want to wind your own coils, and many people did, you need to build one of these. They are really pretty simple as you can see and with a little ingenuity you can make one. The turns counter is driven off a worm gear. The spring in front is to maintain contact between the wire guide assy and the different shaped cams, moving it back and forth. The cam is mounted right behind the crank handle. Each Coilmaster is equipped with four 32 pitch gears, with 39, 40, 42, and 44 teeth. This would give it a spindle to cam ratio of approximately 0.9:1 to

1.1:1. It says enameled wire must first be treated to give it "grip" by passing it through a quick-drying solution, such as resin dissolved in alcohol or other similiar material. Had you guessing too, huh ? Extra cams were $0.75 and gears were $1.00.

A Homebrew "Coilmaster"

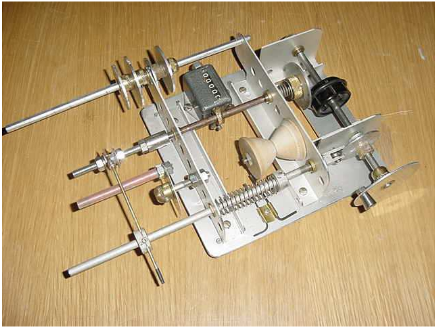

This is homebrew "Coilmaster" using "whatever is available" parts. It has a variable spindle to cam ratio from 0.9:1 to 2.1:1, various cams, various wire feed heads, wire spool holder, turns counter, etc......even comes with two allen wrenches. It is shown with a small plastic bobbin coil spool mounted. Additional cams are located on the wire spool holder shaft and also prevent the wire feed spool from coming too close. The wooden spools are mounted on the spindle for large inner diameter coils.

$$\delta = \left(\frac{R}{2L}\right)\frac{2\pi}{\sqrt{\frac{1}{LC}-\left(\frac{R}{2L}\right)^2}} = \pi\frac{R}{L}\sqrt{LC}\frac{1}{\sqrt{1-\frac{R^2C}{4L}}} \approx \pi\sqrt{\frac{C}{L}}R$$

(11)

when R2C=4L 1. In this experiment, the maximum value of R2C=4L is 0.08 and can fairly safely be ignored to test for linearity. Figure 5 shows versus R and a strong linear relationship is observed. Based on the fitting parameters the x intercept of the line is not R = 0

. This is because R is referring only to the resistance of the decade box and ignores the resistance of the coil and the function generator output. This residual resistance is calculated to be 70 − 4

with 95% confidence. The combined resistance of the inductor coil with the stated output resistance of the Wavetek is 69.4

giving a good agreement with the above measurement.

Table 4: versus resistance.

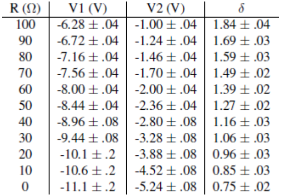

R (Ω)	V1 (V)	V2 (V)	δ
100	-6.28 ± .04	-1.00 ± .04	1.84 ± .04
90	-6.72 ± .04	-1.24 ± .04	1.69 ± .03
80	-7.16 ± .04	-1.46 ± .04	1.59 ± .03
70	-7.56 ± .04	-1.70 ± .04	1.49 ± .02
60	-8.00 ± .04	-2.00 ± .04	1.39 ± .02
50	-8.44 ± .04	-2.36 ± .04	1.27 ± .02
40	-8.96 ± .08	-2.80 ± .08	1.16 ± .03
30	-9.44 ± .08	-3.28 ± .08	1.06 ± .03
20	-10.1 ± .2	-3.88 ± .08	0.96 ± .03
10	-10.6 ± .2	-4.52 ± .08	0.85 ± .03
0	-11.1 ± .2	-5.24 ± .08	0.75 ± .02

Table 3 with the errors in voltage determined by multiplying the minimum change by 2.3 The error becomes worse as the peak number increases

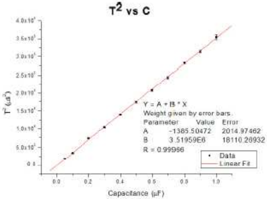

Figure 4: Plot of T2 versus C showing a straight line as predicted by theory.

Table 3: Verification of constant

Peak #	Voltage (V)	δ
1	$7.48 \pm .04$	N/A
2	$3.44 \pm .04$	0.78 ± 0.01
3	$1.56 \pm .04$	0.79 ± 0.03
4	$0.76 \pm .04$	0.72 ± 0.06

because the data was only acquired once with a set vertical resolution on the scope, and thus the uncertainty as a percentage of the value increased as well.

With the consistency of established we proceeded to measure as a function of R. The first and second peaks were measured and used to determine . This data appears in Table 4. Again the error was determined by multiplying the smallest change in V on the scope by 2. Plugging equations 2 and 4 to equation 5 gives

---

The whole thing is made from aluminum scrap. Since aluminum is soft (and easy to work), the wear points are strengthened with brass bushings from old potentiometers. Be sure to grease all the bearing surfaces or they will jam. All shafts are 1/4" (steel or aluminum) cut to size. The collars are made from old knob inserts after the plastic/bakelite is removed via an application of cold chisel. The cams are made from knobs with thick aluminum skirts (plastic/bakelite removed). Some required a little hammer tweaking to tighten the skirt. Since the cam is soft it rides on a small ball bearing (from an old PC disk drive) cam follower.

The drive "gear" is a knob with a groove cut into it. A small "O" ring is installed into the groove and contacts/drives the aluminum skirt of another knob. A short spring is located behind the skirted knob to insure pressure on the "O" ring. Ran into one problem on my coil winder. The "O" ring kept "walking" off the knob on the right angle drive. After a little study, I decided that the "O" ring footprint is not "zero", so the inside of the track is actually running at a different ratio than the outside of the track. This tends to cause an outward force on the "O" ring in the plane of the "O" ring drive shaft. If the centerline of the "O" ring drive shaft is below that of the driven plate shaft centerline, you also have the additional outward force due to the direction of rotation of the driven plate. The way I solved the problem is to equalize the forces by locating the "O" ring drive shaft centerline ABOVE the center line of the driven plate (for this size plate, "O" ring, etc it turns out that 3/32"-1/8" works best). No more problem, everything runs true at any ratio. This was the first Engineering Change to my coil winder .....the driving shaft bushings are now mounted in vertical slots.

The crank is yet another aluminum skirted knob. The mechanical counter was something from a swapmeet and lends itself to being worked with a beveled collar. I could have used a cam to work the counter, but the counter reset knob would have been difficult to get to. The wire feed is a small hobby shop brass tube run through a drilled knob insert with one set screw used to hold the tube in place and the other used to tighten it to the shaft. This tube easily feeds #24 AWG and smaller wire but a larger tube could be used and you can make several wire feed heads. Most coils I need will be #28 AWG or smaller. It has a bail made from a safety pin on the end. The wire is fed into the back of the tube and wire tension is controlled with your fingers. A small piece of shrink tubing was put on the end to further protect the wire. The wire feed head rests on an adjustable bar to control height (yet another knob with aluminum skirt). Fun project.

The additional holes in the braces are for better cooling and to decrease weight ...Hi, Hi.

......and we have the "Cheapo-Winder"

95% confidence1. Based on this value for the inductance it is determined that for C = 1 uF and R = 70 Ω.

$$\left(\frac{1}{LC}\right) = 1.11x10^7 \gg 1.51x10^5 = \left(\frac{R}{2L}\right)^2 \tag{9}$$

and therefore, according to equation 4

$$T^2 \approx (2\pi)^2 LC \tag{10}$$

This relationship is plotted in Figure 42. Again, the R value shows a strong linear relationship between the values as predicted and the error in the slope is only 0.9% with a 95% confidence interval.

Finally, the inductance of the coil was measured using a Z-meter and found to be 92− 1 mH. This is in good agreement with the value determined from the 1/T2 versus 1/C relationship.

3.3 Part B: Log decrement dependence on resistance

For this section of the experiment the same circuit and waveform was used as in part A. The capacitance was set to 1 ┬ F. First we observed the voltage transient at R = 0, 50, 100, 150, 200, 250, and 300 . Above 100 it became difficult to see more than 2 oscillations. Next we verified that the log decrement ( ) did not change going from one set of peaks to the next. We did this by setting the resistance to 0
and measuring the height of the first 4 peaks allowing 3 successive s to be calculated. The results are summarized in

value. Then the error of the single period was just the error of the total time divided by the numer of periods for that measurement. According to equation 4 a plot of 1/T2 versus 1/C should give a straight line with a

$$m = \frac{1}{4\pi^2 L} \tag{7}$$

Figure 3 shows this relationship along with the best fit line. The error bars were calculated using the propagation of errors formula:

$$Error = \sqrt{\frac{4}{T^6} T^2_{error}} \tag{8}$$

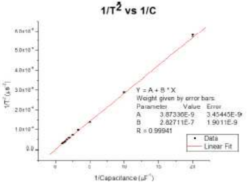

Figure 3: Plot of 1/T2 versus 1/C showing a straight line as predicted by theory.

In this plot the 'Error' column refers to the standard error and it can be seen that the R value indicates a strong linear relationship. Using the relationship in equation 7 it was determined that the inductance of the circuit was 90− 1 mH to

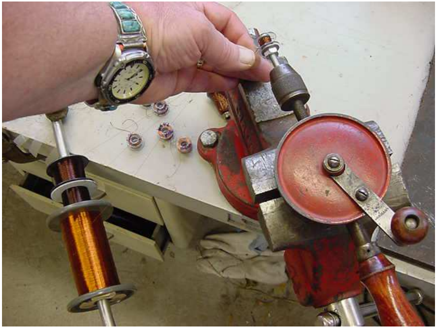

This is "The Old Cheap Econo-Winder" used by many, including myself. It offers the advantages of: no cams to change, no gears to change, 4x chuck to crank ratio, infinitely adjustable, full control, able to wind multiple strands (above photo shows three #38 AWG strands being fed onto a spool). You are the turns counter....multiply the drill crank turns by 4.3 (this drill gear ratio). Turn the crank with one hand, tension and feed the wire in a pattern with the other hand. If you can "pat your head with one hand and rub your belly with the other", you can wind coils by hand. I you can't ....see above.

A few winding hints/ideas/observations

Your results will probably vary some from mine, but the general ideas will hold.

* It was pointed out to me that the main advantage of a winding machine is to allow a "basket weave" pattern which reduces adjacent turn/layer capacitance and increases coil "Q". Basket weave also tends to stay in place better than level wound coils because it's a diagonal criss-cross ....and it looks cool. You can manually wind a very acceptable "basket weave" pattern by hand with a hand drill or with a coil winder. It doesn't have to be "exactly correct" to see the advantage of higher coil "Q". Hand winding will generally require a spool or something to keep the end turns from collapsing. Coil winder made coils can be wound "free standing" if they are not too tall and you use a 2:1 spindle to cam ratio (my opinion).

* Using Litz wire or multi strand (individually insulated) wire increases the surface area which reduces resistance at RF frequencies due to "skin effect" (concentrates the RF current on the outer surface of the wire) and increases coil "Q". Litz wire is useable at frequencies of 2-3Mhz or LOWER. Above 2-3Mhz, solid wire performs as well as the Litz (from Therman's book). The manufacturers of Litz wire twist and alternately place the conductors relative to each other to reduce eddy currents, capacitance, keep impedance constant, etc. The inductance of one strand is about that of two strands, is about that of the whole bundle for Litz wire ....even though the "individual" coil inductances are, in effect, "paralleled". This is because of the close coupling of the wires.

* When winding with served Litz wire, you will find that it lays down easily and makes nice coils. Solid wire (single or multistrand) will also lay down easily if you feed it across a block of beeswax to give it "stick". You will find that the Litz wire does have higher "Q" but it's "bulkier" than the equivalent solid wire (will take up more space). Litz wire will have a tendency to kink as it comes off the spool due to the wire twist.

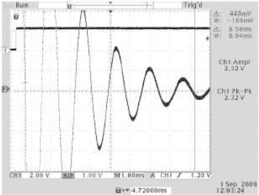

Figure 2: Representative transient signal of voltage across capacitor in this part of the experiment.

used. This was done for 11 different capacitances over the range of 0.05-1 μF. The results are shown in Table 2. The error for the total time was taken to be the resolution of horizontal axis of the Tektronix multiplied by 2. We assumed that this large range gives a 95% confidence in the

Table 2: Measurement of period of oscillation as a function of capacitance.

C (μF)	# of periods	Total time (ms)	Single period (ms)
1	5	9.40±.08	1.88±.02
0.9	5	8.86±.04	1.77±.01
0.8	4	6.72±.04	1.68±.01
0.7	5	7.82±.04	1.56±.01
0.6	6	8.64±.04	1.44±.01
0.5	5	6.62±.04	1.32±.01
0.4	6	7.08±.04	1.18±.01
0.3	7	7.14±.04	1.02±.01
0.2	9	7.70±.04	0.856±.004
0.1	13	7.64±.04	0.588±.004
0.05	9	3.75±.02	0.417±.002

with serial number Phys-943034 was used to measure the resistance of the inductor coil #7. A Wavetek function generator was used in all parts and had the serial number Phys-943026. Additionally a Tektronix TDS3012B oscilloscope with serial number F28819 was used to measure the transient response in all parts.

3.2 Part A: Frequency dependence on capacitance

In this section, along with parts B and C, the Wavetek and circuit were connected as shown in Figure 1. VC was measured using the Tektronix scope and Rout was assumed to be 50 . R in the figure was a decade resistance box and C was a decade capacitance box. L was a large coil inductor (#7) which we measured the resistance of using the DMM before turning on the Wavetek. The resistance was found to be 19.36 . The Wavetek was set to output a 8 V unipolar square waveform with a frequency of 10 Hz and a duty cycle of 50%. The oscilloscope was set to trigger on the leading edge of the unipolar output of the Wavetek.

After all the elements had been setup, the resistance box was set to 0 and it was verified that the period was constant for each oscillation as shown in Table 1. Additionally a representative scope output is shown in Figure 2.

Table 1: Verification of constant period.

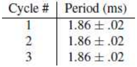

Cycle #	Period (ms)
1	1.86 ± .02
2	1.86 ± .02
3	1.86 ± .02

Then we measured the period of oscillation for different capacitances by determining the time to complete multiple cycles and dividing the result by the number of cycles

---

* To measure the capacitance between conductors of bifilar or Litz wire, measure only a foot or so of wire and extrapolate to the total length (via resistance measurements). If you measure the capacitance on the spool, you will see a very small (in error) reading. The reason is that you are trying to measure capacitance across two really good RF chokes with a tester which uses an RF frequency to determine capacitance.

* Litz wire is sometimes a little confusing. For example: 6/44 unserved Litz wire means 6 strands of #44 AWG wire without an overall wrap (unserved). The AWG "equivalent circular mills" for 6/44 Litz is #36 AWG. This means that 6/44 Litz has the same circular mills and DC current handling capability as one #36 AWG wire. However, the RF current resistance of 6/44 Litz is less than half that of the #36 AWG wire and approaches that of #28 AWG wire. What all this means for the coil is that 6/44 Litz will have a higher "Q" than the equivalent solid wire of equal DC current capability (#36 AWG), but it will take up more space. Unserved 6/44 takes up the space of #34 AWG and served 6/44 takes up the space of #32 AWG.

* From what I've seen, manufacturers of small IF transformers in the 100Khz to 455Khz range use single #38 AWG wire (good), bifilar #40-42 AWG wire (better) or trifilar #44 AWG wire (best) for the windings (allows 18-28ma DC). The smaller JW Miller 100Khz IFs also have a ruberized "powdered iron?" layer inside the aluminum can. I believe the reasons for this layer is to further increase permeability and to isolate the coil from the can. The fact that 100Khz IFs require 5-6mH of inductance dictates a physically large coil, the large coil results in increased capacitive coupling to the can, which reduces "Q" and overall inductance. The higher permeability and resultant higher "AL" value (uH/100turns) is needed to keep the overall coil size manageable in the smaller 3/4" square IF shields.

* Spindle to cam shaft ratios for small IF type coils seem the be 2:1. This means the wire feed head traverses from one side to the other in one spindle revolution and traverses back during the second spindle revolution. I don't think a Morris "Coilmaster" can do that. A 1:1 spindle to cam shaft ratio means the wire feed head traverses to one side and back during one spindle revolution.

* Tuning slug/cup material permeability can vary substantially (just like toroids) and Amidon lists permeabilities up to 35 for their slug tuned coil forms. The 0.01 to 0.50Mhz "3" material has a permeability of 35. As an example: one 455Khz IF (which you would expect to have a high permeability slug) measures 0.7mH without a slug, 1.5mH (max) with it's powdered iron slug and 2.5mH (max) with another powdered iron slug from a JW Miller 100Khz IF (obviously with a higher permeability). Some of the physically small coils, like those in 3/8" cans use very high permeability ferrite cups to reduce coil size requirements.

* When swapping tuning slugs for one you just had to drill out because it broke, make sure the permeabilities are about the same or it won't tune correctly. Could never get that Heathkit HF osc to peak....huh ?.

* One source of low frequency, high permeability slugs is old TV horizontal osc coils. If you need to shorten a powdered iron slug, scribe it with a hacksaw and snap it by hand. A simple method for determining relative permeability is to swap slugs in a given coil form and measure the inductance.

* Coil spools solve the problem of the end turns "collapsing" on small coils, especially if you are not using a coil winder. The coil spools used are plastic "Singer" Class 15 (11/32" tall winding) or "PFAFF" (7/32" tall winding) sewing machine bobbins available at any sewing store (or the wife's sewing room ....but be careful, they don't share your enthusiasm). Coil winding using the wife's sewing machine "bobbin winder"

$$\delta = \ln \frac{q(t_{max})}{q(t_{max} + T_1)} = aT_1 \qquad (5)$$

The last equality in equation 5 can be made based on the form of q(t) where e^{-at} is the attenuating factor. Lastly, the quality factor Q of the circuit is defined as

$$Q = 2\pi \frac{total\ stored\ energy}{decrease\ in\ energy\ per\ period} = \frac{\pi}{\delta} = \frac{\omega_1 L}{R}$$

(6)

with $\omega_1 = 2\pi f_1$ being the damped angular frequency. As can be seen, a lower resistance leads to a higher quality while a higher inductance increases Q.

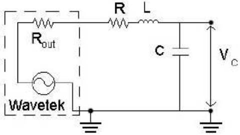

Figure 1: Circuit layout for parts A, B, and C.

3 Experiment

3.1 Equipment

A decade capacitor and decade resistor were used in parts A-C. The serial number for the capacitor was A-1678 and for the resistor was 48288. A HP 34401A Digital Multimeter (DMM)

$$b = \sqrt{\left(\frac{R}{2L}\right)^2 - \left(\frac{1}{LC}\right)} \tag{3}$$

There are three distinct types of solutions depending on whether b2 is positive, negative or zero.

When b2 = 0 the circuit is said to be critically damped. In this case the two roots of the characteristic equation are real and the same value. Therefore, the charge in the capacitor falls to zero exponentially and quicker than for any other value of b.

When b2 > 0 the two roots of the characteristic equation are real and again the charge drops to zero in an exponential fashion. However, the falloff of charge is slower than for b2 = 0. This can be seen by observing that at later times the decay constant is (a □ b) < a. With this type of response the circuit is said to be over-damped.

Finally, when b2 < 0 the two roots are imaginary and thus the charge oscillates about 0 before finally decaying to 0 (assuming a 6= 0). Under these conditions the circuit is said to be under-damped. The frequency of the oscillation is

$$f_1 = \frac{1}{T_1} = \left(\frac{1}{2\pi}\right)\sqrt{\left(\frac{1}{LC}\right) - \left(\frac{R}{2L}\right)^2} \tag{4}$$

where T1 is the period of the oscillation. Because the charge is still decaying the logarithmic decrement δ can be defined as the natural log of the ratio between two successive peaks of charge

does not work and can shorten your life. A "Singer" Class 15 or "PFAFF" bobbin has a thicker core and lends itself to being drilled out to 19/64", 9/32", or 17/64" which allows them to be put onto standard 1/4" slug tuned cores. Trim the spool with wire nippers and since it's an outside curved edge, a finger nail clipper works well. Trim the spools after you have wound them to keep the newly created rough edge from cutting small gauge wire or smooth the spool edge before winding. If the spools are not the correct height, cut them in half or make some "washers". It's important to secure these solidly to the coil form or the winding will tend to "walk" the washer ends outward. When done, carefully remove the washers and add some beeswax/Q dope to seal the outer turns.

* To keep the initial turns from slipping add a layer of double sided tape to the core or drip some beeswax onto the core (preferred).

* Many types of magnet wire allow direct tinning of the wire which burns off the insulation in the process. This is MUCH easier than scraping the insulation off with a razor blade, especially for the smaller gauges. Burning off the insulation over a flame is not advised.

* Aspect ratio is important. A fixed length of wire was wound on a given form and slug. The coil winding height on one was 3/8" and 3/16" on the other. The 3/16" winding (larger outer diameter) measured 40% (max) more inductance. Therman says optimum inductance is achieved if the winding cross section is square. Slug tuned coils may be different.

* To give the wire more "stick" as you are winding, feed it across and into a beeswax block (available at art/craft and sewing stores).

Factors Affecting Coil Q ...some experiments

EXPERIMENT #1 ....Skin Effect

Four coils were wound by hand. Coil Q will increase if the turn-to-turn capacitance is reduced. This can be accomplished by using a "basket weave" pattern. The "basket weave" coils have roughly 1-2 spool length traverses of the wire per revolution of the spool (coil Q will be even higher on fixed pattern wound coils made using a coil winder). Q can also be increased by using multi strands of smaller gauge wire due to increased surface area (skin effect) at frequencies below about 3Mhz. It's a substantial improvement in Q to use two strands of #39 AWG vs one strand of #36 AWG, or better yet .....six strands of #42 AWG vs one strand of #36 AWG. All three of those examples have the same current handling capability. Many small, higher quality, IF transformers are wound with two strands of smaller gauge wire. Larger, high quality IF transformers are wound with Litz wire, dipped, etc.

All coils were trimmed to 2.0 mH with no slug, and 3.0 mH with a slug. Q Data is "unloaded Q" as measured on a Boonton 260-A.

At 200Khz, all coils trimmed to 2mH, no tuning slug

Coil #1 #34 AWG, level wound, 325pf, Q of 52
Coil #2 #34 AWG, basket weave wound, 320pf, Q of 55
Coil #3 two #38 AWG, basket weave, 310pf, Q of 78
Coil #4 three #38 AWG, basket weave, 320pf, Q of 83

At 170Khz, all coils trimmed to 2mH, no tuning slug

Coil #1 #34 AWG, level wound, 440pf, Q of 55
Coil #2 #34 AWG, basket weave wound, 447pf, Q of 57
Coil #3 two #38 AWG, basket weave, 433pf, Q of 78

RLC loop when an external voltage was applied. The capacitance was varied and the periods of the oscillations were used to determine the inductance in the circuit. Next we measured the log decrement as a function of resistance to verify the response is approximately linear and to estimate the total resistance of the circuit including the inductor and the function generator. Following that we determined the resistance required for critical damping as a function of capacitance. Using this we verified the theoretical result that the critical resistance is proportional to 1/ p C. Finally, we measured the voltage across the capacitor in a different RLC circuit driven by a sinusoidally varying voltage. The peak-to-peak voltage was measured as a function 1 of frequency to determine the resonant frequency, the bandwidth, and the quality factor Q. We also compared the resonant frequency with the theoretical value.

2 Theory

The governing equation for a resistor, inductor and capacitor in series with a voltage source is

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{q}{C} = V(t) \qquad (1)$$

This is the equation for an oscillator with damping and a driving function. Solving the characteristic equation gives two roots s1; s2 = a − b with

$$a = \frac{R}{2L} \qquad (2)$$

Transients and Oscillations in RLC Circuits

Will Chemelewski with Brian Enders

http://courses.physics.illinois.edu/phys401/Files/Misc/Sample%20report%20RLC.pdf

Abstract

Transient responses of RLC circuits are examined when subjected to both long time scale (relative to the decay time) square wave voltages and sinusoidally varying voltages over a range of frequencies about the resonant frequency. In general, a good correspondence is found between theory - describing the charge in the system in terms of a 2nd order differential equation with a harmonic oscillator form - and experiment. It is demonstrated that the inductance can be accurately measured using period of oscillation versus capacitance measurements. Furthermore, the exponential decay of the response is described well by the model and the resonant frequency of a sinusoidal external voltage is accurately predicted. However, some discrepancies were found though not necessarily a result of theoretical failures. One problem is the failure to predict the inductance of a circuit based on the critical resistance variation with capacitance, although the problem could lie in how the measurement is conducted. Additionally, the quality and bandwidth of a RLC element is poorly predicted but this could also be a result of experimental problems.

1 Purpose

The purpose of this experiment was to observe and measure the transient response of RLC circuits to external voltages. We measured the time varying voltage across the capacitor in a

Coil #4 three #38 AWG, basket weave, 446pf, Q of 85

At 150Khz, all coils adjusted to 3mH with the tuning slug

Coil #1 #34 AWG, level wound, 400pf, Q of 82
Coil #2 #34 AWG, basket weave wound, 380pf, Q of 85
Coil #3 two #38 AWG, basket weave, 380pf, Q of 109
Coil #4 three #38 AWG, basket weave, 385pf, Q of 114

EXPERIMENT #2 ....Coil Form Dielectrics

Other factors which effect coil Q are the coil form material and the adhesive used to hold the windings. Any coil form which can absorb moisture, like the old cardboard oatmeal boxes and toilet paper tubes used for crystal sets, is a problem unless the cardboard is treated with varnish, shellac, etc ....but it's cheap.

Some have suggested using old plastic pill bottles which is OK if they are sturdy so the coil turns don't move. Recently some testing has shown that coils wound on those amber pill bottles have just has high a "Q" as those wound on the Amphenol 5-pin forms. Some PVC is the same way. What does affect "Q" is the material you use to coat the windings (materials to be tested later ...beeswax, refined paraffin, varnish, laquer, other "glues"). Here is an example of coils wound (adjacent turns) with #28 AWG magnet wire:

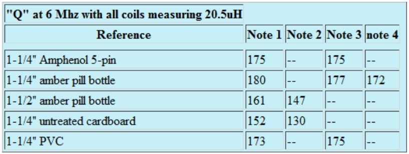

"Q" at 6 Mhz with all coils measuring 20.5uH

Reference	Note 1	Note 2	Note 3	note 4
1-1/4" Amphenol 5-pin	175	--	175	--
1-1/4" amber pill bottle	180	--	177	172
1-1/2" amber pill bottle	161	147	--	--
1-1/4" untreated cardboard	152	130	--	--
1-1/4" PVC	173	--	175	--

Notes:

Plain wiring, no adhesives, no tape to hold windings

Use a light coat of clear fingernail polish to hold windings (strips across windings didn't work), Q reading 15-20 points lower when first applied, above readings are after 48 hours.

Use Elmer's Stix-All silicone to hold windings (4 strips across coil windings)

Use Scotch black electrical tape to hold coil windings (1-1/2 turn)

EXPERIMENT #3 ...Turn Spacing

The highest Q coil for a given inductance is "air wound" with space between the turns (less turn-to-turn capacitance), the next is space winding on a good coil form (ceramic, phenolic, etc). Space winding on a coil form by winding string or another wire along with the coil wire and removing the "spacer" later is good. Ceramic and some phenolic coil forms came with wire groves to space the turns. You do have the problem of having to stabilize the separated turns with Q-Dope, paraffin, Duco cement, etc.

Here is some data on two coils:

---

stripped. You will be left with nice clean varnish-free wires, which can then be tinned as usual.

As a test of this method I measured the AC impedance of each of the wires of the same length at 10kHz using my LCR meter, and found that they all were very consistent, indicating that I was effectively and consistently removing the varnish and using all the wires in the bundle.

By the way, the speakers sound terrific! I'm not sure how much of the improvement was from the litz wire, but it certainly didn't hurt the sound!

found it works, but you end up oxidizing the wires and even incinerating some of the smaller strands),

5) Individually sanding each strand (I'd be doing this job for years as each wire has dozens of strands and I have 6 tweeters, 6 midranges and 4 woofers to wire up to 18 capacitors and 20 inductors!).

I tried several of these methods, and none were satisfactory in producing a nice clean tinned wire end, without breaking or damaging the strands. However, I did eventually come up with a simple, fast and very effective method, so I thought I would share it with the other DIY'ers here, in case they are crazy enough to build something with litz wire.

A Dremel tool with a stainless steel wire wheel attachment works very well. I found that the stainless wheel works faster and lasts longer than the carbon steel wheel, with the added advantage that the small pieces of the wheel brush that wear off during the process are non-magnetic, so they won't attach to nearby driver magnets (don't need little pieces of wire inside my EMIMs and EMITs!). Simply remove ~ 1/2" of the insulation, then untwist the various strands from each other (the Cardas Litz has 3 counterwound layers of different diameter wire to untwist, do them one at a time) and flatten them into a flat fan shape. Set the Dremel to about 1/2 of maximum speed and apply to the coated wire, ensuring that the wheel is ALWAYS turning toward the ends of the wire (otherwise they will entangle with the spinning wire wheel and snap off!). 10 seconds on each side of the fanned wire is very effective at removing the varnish coating without destroying the wires. Then re-fan the wire in a different direction and repeat the process 2 or 3 times to ensure that all the wires get

1-1/4" coil form (amber pill bottle)
both coils 13.8uH +/- 0.1uH
both coils #28 AWG wire
coil "A" 16 turns are adjacent
coil "B" 20 turns are on .025" centers +/- (spacer was removed)
both coils, no adhesives/coatings to hold the turns in place

At 3.4MHz ..... coil "A" Q=107, coil "B" Q=175

Granted, these are unloaded Q measurements on a Boonton 260-A and loaded Q will be less, however the improvement will be carried over and the "trend" is there.

What does this mean for the signal (loaded Q will be less dramatic):

Sensitivity ...the Boonton uses a 20mVAC "e" signal and measures "E" with a very sensitive voltmeter, so a Q of "107" means "E" is 2.14VAC ....a Q of "175" Q means "E" is 3.5VAC ....higher sensitivity.

Selectivity ....Q is also the -3db voltage BW divided into the frequency (Q = Fo/-3dbBW ...or -3dbBW = Fo/Q). For a Q of "107" the -3dbBW is 31.8Khz. For a Q of "175" the -3dbBW is 19.4Khz .....better selectivity.

EXPERIMENT #4 ....Coil Dope

"E6000" Craft Glue was used on a coil with a Q of "175" ...absolutely no change in Q as you put the material on, as it dries, or 24 hrs later, even for a very heavy application of glue. I believe it's some kind of silicone based product .....waterproof, clear, and remains flexible. My wife didn't even tell me she had some "really good coil dope" material in her crafts pile.

"CG Clear Ice" fingernail polish was used on a coil with a Q of "107" and the Q dropped to "97".

Considering the amplitudes of successive cycles when cos(ω dt- ϕ) = 1

$$x_n = X_0\, e^{-\zeta\omega_n t_n} \quad \text{and} \quad x_{n+1} = X_0\, e^{-\zeta\omega_n(t_n + T_d)}$$

$$\text{Therefore} \quad \frac{x_n}{x_{n+1}} = e^{\zeta\omega_n T_d}$$

The logarithmic decrement δ is ln(xn / xn+1).. Normally n =1 and n+1 is therefore 2.

$$\delta = \ln\left(\frac{x_n}{x_{n+1}}\right) = \zeta\omega_n T_d = \zeta\omega_n \frac{2\pi}{\omega_d} = \zeta\omega_n \frac{2\pi}{\omega_n\sqrt{1-\zeta^2}}$$

The damping ratio ζ can be expressed in terms of the logarithmic decrement δ as follows

$$\zeta = \frac{\delta}{\sqrt{(2\pi)^2 + \delta^2}}$$

It is clearly possible to determine the damping ratio experimentally for a mechanical system be initiating vibrations and measuring the amplitude of the vibrations.

Logarithmic Decrement....

Roy Beardmore 17/01/2013

http://www.roymech.co.uk/Useful_Tables/Vibrations/Free_Vibrations.html

The rate at which the amplitude of vibrations decays over time provides a very useful method of identifying the degree of damping. See the figure below for the plot of a typical underdamped vibration.

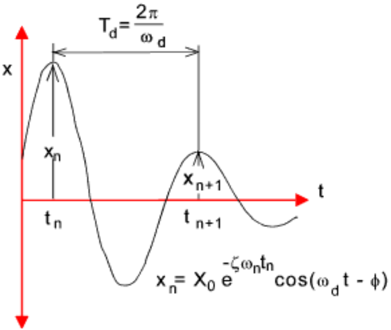

The equation for of motion for underdamped oscillations arrived at above can be used to establish the amplitude of any of the cycles. i.e.

$$x_n = X_0 e^{-\zeta\omega_n t_n} \cos(\omega_d t - \phi)$$

Understanding LC Tank Q, Impedance, and Losses

David Wagner

http://wiki.waggy.org/dokuwiki/crystal_radio/tank

A resonant inductor-capacitor (LC) tank is the beating heart of a crystal radio, filtering the signal of interest from the sea of radiofrequency (RF) energy received by an antenna. Properly loaded, a high-quality (high-Q) tank will give a crystal radio greater sensitivity and selectivity.

For a parallel tank, a smaller inductor and larger capacitor can result in higher Q.

$$Q=\omega R_p C = \frac{R_p}{X} = \frac{R_p}{\omega L}$$ 1)

Tank Impedance

When tuned to the station of interest, the crystal radio's LC tuning tank between the antenna and detector presents a very high resistive impedance *to ground*. Estimating the magnitude of this effective shunt resistance is not difficult, but it is important to keep in mind how the tank impedance is in parallel with both the source and the detector/load impedance.

If the source and load impedances are low in comparison to the tank's shunt resistance, this resistance may be neglected, the tank loss will be fairly low, and the antenna can be matched directly to the serial combination of the detector and the audio load. However, passing a low impedance signal significantly loads the tank and results in broader tuning. To minimize tank loading, the signal impedance will approach the tank's shunt resistance, and this resistance must be considered as part of the driven load. Estimating the parallel LC tank's effective shunt resistance can be done by measuring or assuming the tank's unloaded Q.

$$R_{tank}=2\pi f\ L_{tank}Q_{unloaded}$$

Tank Loading

For efficient energy transfer, the antenna impedance should be matched to the parallel combination of the load and shunt impedances. In effect, the tank must be kept fully 'powered up' to maintain its narrow frequency response while reflecting the rest of the signal into the load.

$$R_{src}=\frac{1}{\frac{1}{R_{load}}+\frac{1}{R_{tank}}}\quad\Rightarrow\quad\frac{1}{R_{load}}+\frac{1}{R_{tank}}=\frac{1}{R_{src}}$$

The LC tank shunt resistance can then be used to estimate upper limits on the tank's loaded Q and the matched source and load impedances.

$$R_P=\frac{1}{\frac{1}{R_{src}}+\frac{1}{R_{load}}+\frac{1}{R_{tank}}}=\frac{1}{\frac{1}{R_{src}}+\frac{1}{R_{src}}}=\frac{1}{2}R_{src}$$

$$Q_{loaded}=\frac{R_P}{2\pi f\ L_{tank}}=\frac{R_{src}}{4\pi f\ L_{tank}}$$

$$R_{src}=2R_P$$

This is for very small signals.

oscillation of IL relative to the oscillation of I changes by about 180 deg. Increasing the value of the Q-factor makes this variation of phase more abrupt. However, the most important feature of the steady state solution described by Equations (19-21) is that one can reduces the frequency width of the peak in the amplitude of IL, increases its height, and chooses its frequency position, by an appropriate choice of L, r and C. That the enhancement can be made to occur in a selected, narrow frequency band by tuning the parameters of a resonant circuit is particularly useful in applications, as this makes it possible to isolate a signal occurring at a particular frequency (e.g., the emission from a radio station) from an irrelevant background.

6. To conclude

More details on LR and series LRC circuits can be found in Young and Freedman [1], chapters 30 and 31. The Physics and Maths of LRC circuits will also be studied in the Level 2 Electronics course, amongst many other things.

References

[1] H. D. Young and R. A. Freedman, University Physics, 13th Ed., Pearson Addison-Wesley, San Francisco (2012).

$$R_{load} = R_{det} + R_{audio} = 2R_{det}$$

This may not be correct since R_load(RF) may be only about half of R_load(audio) and R_load(DC).

Assume all loss is through the LC tank shunt resistance (R_tank); losses due to circulating current resistance are small in comparison to primarily capacitive losses to ground.

$$R_{src} = \frac{1}{\frac{1}{2R_{det}} + \frac{1}{R_{tank}}}$$

$$\Rightarrow R_p = \frac{1}{2} R_{src} = \frac{1}{\frac{1}{R_{det}} + \frac{2}{R_{tank}}}$$

$$\Rightarrow Q_{loaded} = \frac{1}{\left(\frac{2\pi f\, L_{tank}}{R_{det}} + \frac{2}{Q_{unloaded}}\right)} =$$

$$\frac{Q_{unloaded} R_{det}}{2\pi f\, L_{tank} Q_{unloaded} + 2R_{det}}$$ 2)

$$\Rightarrow Q_{loaded} = \frac{1}{\left(\frac{4\pi f\, L_{tank}}{R_{load}} + \frac{2}{Q_{unloaded}}\right)} =$$

$$\frac{Q_{unloaded} R_{load}}{4\pi f\ L_{tank} Q_{unloaded} + 2R_{load}}$$

The tank loss can now be estimated.

Efficiency (%): $\frac{R_{tank}}{R_{tank} + 2R_{det}} = \frac{\pi f\ L_{tank} Q_{unloaded}}{\pi f\ L_{tank} Q_{unloaded} + R_{det}} =$

$$\frac{1}{\frac{R_{det}}{\pi f\ L_{tank} Q_{unloaded}} + 1}$$

Tank Loss (dB): $10\log\frac{R_{tank}}{R_{tank} + 2R_{det}}$ dB

Tank Loss (%): $\frac{R_{det}}{R_{tank} + 2R_{det}} = \frac{R_{det}}{2\pi f\ L_{tank} Q_{unloaded} + 2R_{det}}$

Circulating Current Losses

The preceding should be valid so long as the tank's Q is limited primarily by its shunt resistance. To check this assumption, verify the tank's circulating current series resistance (R_circ) is

---

at the natural frequency of the LRC circuit, $f_0 = \omega_0/2\pi$. (This last feature might not be completely obvious from the form of the equation if you are not much used to analysing the behaviour of functions.) At its maximum, the amplitude of oscillation of IL is about a factor Q larger than the amplitude of oscillation of I. (Note that $I_{L0} = QI_0$ when $\omega = \omega_0$. The amplitude of IL can exceed that of the total current, I0, because the current passing through the inductor includes a current circulating round the LCR circuit formed by the inductor and the capacitor; this circulating current may be large when $f \approx f_0$.)

This enhancement is an example of a resonance, a common phenomenon in many different fields of Physics and Engineering: When an oscillating system is excited by a periodic perturbation whose frequency (almost) coincides with one of its natural frequencies, the response of the system to the perturbation is greatly magnified when damping is light. (See Young and Freedman [1], Section 14.8, for a short discussion of this matter.)

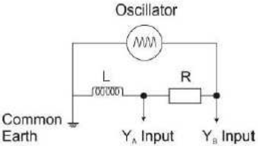

Figure 3: The circuit considered in tasks 1 and 2.

One also sees from Equation (20) that when the frequency of the generator sweeps through the resonance, the phase of the common feature of the evolution of dissipative systems kicked away from their equilibrium position by a sudden perturbation.

The steady state solution is particularly interesting when the total current generated by the signal generator varies sinusoidally. Setting $I = I_0 \cos(\omega t)$, as given by Equation (2), Equation (15) now reads

$$LC\frac{d^2 I_L}{dt^2} + rC\frac{dI_L}{dt} + I_L = I_0 \cos(\omega t) \qquad (18)$$

This last equation can also be solved exactly using general methods (which you will study in your Maths course). Doing so results in the following steady state solution:

$$I_L = I_{L0}\sin(\omega t - \phi) \qquad (19)$$

where the phase-angle f is defined by the equation

$$\tan\phi = Q\frac{\omega^2 - \omega_0^2}{\omega\omega_0} \qquad (20)$$

and

$$I_{L0} = \frac{I_0}{\sqrt{\frac{1}{Q^2}\frac{\omega^2}{\omega_0^2} + \left(\frac{\omega_0^2 - \omega^2}{\omega_0^2}\right)^2}} \qquad (21)$$

Equation (19) shows that in the steady state the current through the inductor oscillates at the frequency of the signal generator, $f = \omega/2\pi$. Equation (21) tells us that when the Q-factor is much larger than 1, then, as a function of f, IL strongly peaks

much lower than what it would need to be to be solely responsible for the LC tank's unloaded Q.

$$R_{\circ} << \frac{2\pi f\, L_{tank}}{Q_{unloaded}} = \frac{1}{2\pi f\, C Q_{unloaded}}$$

For the 'worst case' broadcast band values, this series resistance is small, but not impossible to beat.

$R_{\circ} << \frac{2\times\pi\times f\ \times 90\times 10^{-6}}{1000} \approx$ 0.3 Ω at 520 kHz, and 1 Ω at 1720 kHz.

For comparison, the 30' (10 m) of 12 AWG (2 mm) solid wire sufficient to make a stout cylindrical air-core inductor of this value is about 0.05 Ω.

One thing to note is how much circulating current resistance by itself is equivalent to the tank loading resistance (again by itself) resulting in the same tank Q.

$$Q = \frac{R_{tank}}{2\pi f\, L_{tank}} = \frac{2\pi f\, L_{tank}}{R_{\circ}}$$

$$\Rightarrow \quad R_{tank} R_{\circ} = \left(2\pi f\, L_{tank}\right)^2$$

$$R_{tank} = \frac{\left(2\pi f\, L_{tank}\right)^2}{R_{\circ}} \quad ;$$

$$R_o = \frac{\left(2\pi f\ L_{tank}\right)^2}{R_{tank}}$$ [3]

Thus, at 1 MHz, one ohm of circulating current resistance is equivalent to loading a 100 µH tank with about 400 kΩ, while it would load a 200 µH tank only one-quarter as heavily, about 1.6 MΩ. Since parallel loading dominates most crystal radio tank designs, it seems worthwhile to use tanks with smaller inductors. However, parallel loading decreases Q with increasing frequency, and BCB reception requires an increasing Q to maintain constant bandwidth, so some trade-offs may be worth considering.

Now, if it were true that the unloaded tank Q can be modeled accurately from only the inductance (L_tnk) and parallel tank resistance (R_tnk), then the Q should always decrease with frequency, but this is not the case. Perhaps a combination of parallel and serial (circulating) resistance can be used.

$$Q_{unloaded} = \frac{1}{\dfrac{2\pi f\ L_{tank}}{R_{tank}} + \dfrac{R_o}{2\pi f\ L_{tank}}}$$

Tank Design

Typical ferrite toroid inductor Q is 300.[4]
Extremely good tanks have R_tank increasing from 1 MOhm to 2 MOhm across the broadcast band.[5]

[1] Q-Factor

$$LC\frac{d^2 I_L}{dt^2} + rC\frac{dI_L}{dt} + I_L = I \qquad (15)$$

The solutions of this equation will depend on how the total current I varies in time. For a square wave signal, I can be taken to jump rather abruptly, but periodically, from one value to another, staying constant between the jumps. Between any two consecutive jumps, the time evolution of IL is thus governed by the equation

$$LC\frac{d^2 I_L}{dt^2} + rC\frac{dI_L}{dt} + I_L = I_0 \qquad (16)$$

where I0 is a constant. The relevant solution of this equation is

$$I_L = A\exp(-\alpha t)\cos(\omega_0' t + \chi) + I_0 \qquad (17)$$

with the values of the constants A and depending on the precise variation of I during the jumps. (What these values are is not important for us.)

It is clear from Equation (17) that immediately after the jump in I, the current across the inductor oscillates at the natural frequency of the circuit, about a value equal to I0, with an amplitude of oscillation decreasing in time due to the exponential factor. Thus IL tends towards I0, and becomes indistinguishable from I0 once the initial oscillation has sufficiently decrease in amplitude. The time-evolution of IL can thus be divided into two parts: At first, immediately after the jump in I, IL displays transients (the free oscillations), which disappear gradually and leave place to a steady state solution (IL = I0). The occurrence of transients is a very is large (which will be the case for the circuit built in this practical).

As can be easily checked from Equation (10), the amplitude of the oscillation decreases by a factor 2 over any interval of time of duration $\Delta t_{1/2} = \ln 2 / \alpha$, that is, over a number of periods of oscillation equal to $n_{1/2} = \Delta t_{1/2} / T_0$. (We neglect the difference between ω_0 and ω_0'.) Given that $\alpha = r/(2L)$ as noted above, and that $T_0 = 1/f_0$, we have, in good approximation when Q is large,

$$Q = \frac{\pi n_{1/2}}{\ln 2} \qquad (14)$$

This last result will be used in Task 5. An important conclusion one can derive from it is that a large Q-factor means that the oscillations last for a long time, i.e., that the damping is light.

However, the case of interest here is not that of an LRC circuit in isolation but that of Figure 4 below, in which an LRC circuit is connected to other components. In the experiment, the function generator, or "oscillator", may generate a signal which is either a square wave or a sinusoidal wave, and, due to the large resistance of the resistor R, the current flowing through the whole circuit is practically the same as if the capacitor and inductor were not present. We denote by I the total current flowing through the whole circuit. Therefore we now have $I_C + I_L = I$ instead of $I_C + I_L = 0$.

The result is that Equation (9) acquires a non-zero right-hand side and becomes

---

2)

$$Q_{loaded} = \frac{1}{\left(\frac{1}{R_{det}} + \frac{2}{R_{tank}}\right) 2\pi f\, L_{tank}} =$$

$$\frac{1}{\left(\frac{1}{R_{det}} + \frac{2}{2\pi f\, L_{tank} Q_{unloaded}}\right) 2\pi f\, L_{tank}} =$$

$$\frac{1}{\left(\frac{2\pi f\, L_{tank}}{R_{det}} + \frac{2}{Q_{unloaded}}\right)} = \frac{Q_{unloaded} R_{det}}{2\pi f\, L_{tank} Q_{unloaded} + 2 R_{det}}$$

3) This gives the same results as Series - Parallel Impedance Conversion Calculator//.

4) Experiments with Coils and Q-Measurement, Wes Hayward, w7zoi, October, 2007 (Update 01Dec07.)

5) Experiments with LC circuits part 10

$$\omega_0' = \sqrt{\frac{1}{LC} - \frac{r^2}{4L^2}} \qquad (11)$$

[Equations (10) and (11) are not correct when the argument of the square root function is zero or negative; however, we do not need to worry about this as in the practical r, L and C are such that the argument is positive.]

The time evolution described by Equation (10) is a damped oscillation: the current oscillates, now at a natural angular frequency ω_0' rather than ω_0, and the amplitude of the oscillation decreases exponentially in time. The oscillations disappear because the energy they contain is dissipated – here it is dissipated as heat due to the resistance of the coil. The "quality" of an oscillator, i.e., the length of time the oscillations last once started, is usually quantified by a dimensionless parameter denoted Q, the Q-factor of theoscillator. Here Q is given by the equation

$$Q = \frac{2\pi f_0 L}{r} \qquad (12)$$

where f0 is the natural frequency of the circuit as defined above, $f_0 = 1/2\pi\sqrt{LC}$.Thus $Q = \omega_0 L/r$, and, from Equation (11),

$$\omega_0' = \omega_0\sqrt{1 - \frac{1}{4Q^2}} \qquad (13)$$

The displacement of the natural frequency from its zeroresistance value is therefore small when the quality factor

T_0. $f_0 = 1/T_0$, is called the natural frequency of the circuit. (Recall the relations between frequency, angular frequency and period: $T_0 = 2\pi/\omega_0$ and $\omega_0 = 2\pi f_0$.)

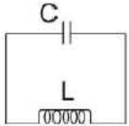

Figure 2: An LC circuit

It is a simple matter to correct Equations (6) and (7) for the internal resistance of the inductor. (With this resistance taken into account, the circuit shown in Figure 3, is a “parallel LRC circuit.) If the resistance of the inductor is r, then $V_L = L\, dI_L/dt + rI_L$, so that Equation (6) now gives

$$LC\frac{d^2 I_L}{dt^2} + rC\frac{dI_L}{dt} + I_L = 0 \qquad (9)$$

This last equation has exactly the same form as the equation of motion of a classical harmonic oscillator for the case where the oscillating mass is submitted to a frictional force proportional to its velocity – see, e.g., Section 14.7 of Young and Freedman [1]. The equation can be solved analytically using general methods. The result is that

$$I_L = A\exp(-\alpha t)\cos(\omega_0' t + \chi) \qquad (10)$$

where A and are two arbitrary constants (i.e., two constants which can have any value), a = r / (2L) , and

---

MEASURING THE Q OF LC CIRCUITS

Dick Kleijer crystal-radio.eu
http://www.crystal-radio.eu/enqmeting.htm

In theory we can determine the Q of a circuit as follows:

Step 1

Couple a RF signal generator to the LC circuit.The coupling between generator and LC circuit must be loose, otherwise the output resistance of the generator will load the circuit and reduces the Q.

Step 2:

Set the generator to the frequency at which you want to measure the Q.Adjust the LC circuit (turn the tuner capacitor) so you have maximum voltage over the circuit, the circuit is now in resonance, this frequency is the resonance frequency of the circuit (f.res).

Step 3:

Measure the voltage over the LC circuit at resonance frequency (f.res).

Step 4:

Vary the generator frequency a little above and below f.res. and determine the two frequenties were the voltage over the circuit is 0.707 times the value at f.res.The voltage reduction to 0.707 times, is the -3 dB point.One -3 dB point, is lower in frequency then f.res, this frequency we call: fl.The other -3 dB point is higher in frequency then f.res, this frequency we call: fh.

Step 5:

Calculate the bandwidth BW: **BW= fh - fl.**Calculate the Q: **Q= f.res / BW**

For performing these 5 steps, we can use the following test setup

Test setup 1 Measuring the Q with a signal generator and a probe.

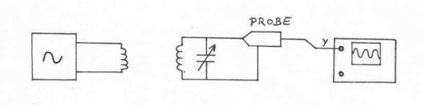

In the schematic above you see from left to right the following components.
A signal generator
A coupling coil
The LC circuit
A 1:100 oscilloscope probe
A oscilloscope

Connect the output of the signal generator to the coupling coil having e.g. 50 turns.
Place the coupling coil at about 20 cm from the coil of the LC circuit.
The coupling coil don't have to be high Q.

These currents and voltages and the charge of the capacitor may vary in time. The instantaneous voltage across the capacitor is given by $V_C = q/C$. Since $I_C = dq/dt$ and $I_C = -I_L$, we can write

$$C\frac{d}{dt}(V_C) = -I_L \tag{6}$$

From Equation (1), we also have VL= L dIL/dt. (As above, we assume that there is no significant potential drop across the coil originating from its internal resistance.) Since C L V =V , we arrive at the equation

$$LC\frac{d^2 I_L}{dt^2} = -I_L \tag{7}$$

You may recognize from the Foundations course that this equation has exactly the same form as the equation of motion for a mass attached to a spring and sliding on a frictionless table (see Chapter 14 of Young and Freedman [1]). Its solutions have thus the same mathematical properties as the solutions found in the mechanical problem, although the Physics context is completely different. Therefore IL (and thus q, IC, VC and VL) oscillate harmonically in time, i.e., like a sin or a cos function. Repeating the calculation done for the mechanical harmonic oscillator, one finds that the angular frequency of these oscillations is

$$\omega_0 = \sqrt{\frac{1}{LC}} \tag{8}$$

and that their period, T_0. is $2\pi\sqrt{LC}$. The inverse of

$$V_L = -\omega L I_0 \sin(\omega t) = \omega L I_0 \cos(\omega t + \pi / 2) \tag{4}$$

Comparing equations (2) and (4), we see that the maxima of VL occur a time $\pi/(2\omega)$ before the maxima I, i.e. a quarter of a cycle earlier: one says that the voltage leads the current by 90 degrees.

From the last two equations we also see that the maximum of the voltage across the inductor, V_L^{MAX}, is $\omega L I_0$, while the maximum of the voltage across the resistor, V_R^{MAX}, is RI_0. We thus have

$$V_L^{MAX} = \left(\frac{\omega L}{R}\right) V_R^{MAX} = \left(\frac{2\pi f L}{R}\right) V_R^{MAX} \tag{5}$$

This relation will be used in Task 2 to obtain the inductance of the inductor provided. Let us now consider the circuit shown in Figure 2. To begin, we'll assume that the resistance of the wires connecting the capacitor and the inductor and the resistance of the wire forming the inductor (its internal resistance) are both negligible. The voltage across the capacitor and the voltage across the inductor will be denoted by VC and VL, respectively, the capacitance and inductance of these two components by C and L, the current flowing through the capacitor by IC, that flowing through the inductor by IL, and the charge of the capacitor by q. For convenience, we measure VC , VL , IC and IL from left to right. Hence VC =VL and C L I = -I . (To see why there is a minus sign in the last equation, note that the current must flow in opposite directions in the upper and lower parts of the circuit.)

Because of the 20 cm distance, there is a loose coupling between the coils.

Connect the probe to the LC circuit.
The earth connection of the probe must be connected to the housing of the tuner capacitor.
The probe is connected to the oscilloscope.
The probe provides a small loading of the circuit, so the Q will not reduce so much.
There are also 1:1 and 1:10 probe's, but these will load the LC circuit too much.
The 1:100 probe I use has a input resistance of 100 M.Ohm, and a input capacity of 4 pF.

The output voltage of the generator must be set so high, that the oscilloscoop gives a clear picture of the RF signal.
Because the 100 times attenuation in the probe, the signal generator output must be set fairly high.
When measuring low Q circuits, I must set the generator output to it's maximum of 20 Volt peak-peak.

For measuring the Q: perform the 5 steps described on the top of this page.

The frequency adjustment is done by hand, by turning the frequency knob of the generator.

Test setup 2 Measuring the Q with a sweep generator and a probe.

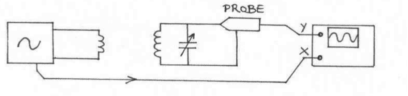

In this schematic you see from left to right the following components:

A sweep signal generator
A coupling coil
The LC circuit
A 1:100 oscilloscope probe
A oscilloscope

This method uses a sweep generator, this is a signal generator where the frequency is constant varying between two set values.
I use a sweep function generator of brand "Hung Chang" with model number G305, it can produce signals up to 10 MHz.
It has a "sweep output" which gives a voltage going up and down with the frequency sweep.

The "sweep output" is connected with the X input of the oscilloscope, the oscilloscope is placed in the X-Y mode.
Now the lightspot on the scope runs from left to right and back over the screen, this makes a frequency scale with on the leftside the startfrequency and on the rightside the stopfrequency of the sweep generator.
The sweepfrequency must be set at about 10 Hertz, this means the frequency is running 10 times per second from startfrequency to stopfrequency and back.

The Y input of the oscilloscope is connected via the 1:100 probe with the LC circuit.

The RF output of the sweep generator is connected to the coupling coil, which is placed about 20 cm from the coil of the LC circuit.

L is always positive. Hence, when the current increases, so does the potential drop across the coil, which means that the coil effectively opposes the increase in current. Likewise, when the current decreases, the potential difference across the coil is negative, which tends to increase the current and opposes its change. These effects are a manifestation of Lenz's law, which states, in the words of Young and Freedman [1], that 'the direction of any magnetic induction effect is such as to oppose the cause of the effect'.

The coil does not act like a resistor following Ohm's law, though, since VL is proportional to the rate of change in current, not to the current itself. For instance, consider the LR circuit of Figure 1. We'll denote the inductance of the inductor by L, the resistance of the resistor by R, and the frequency of oscillation of the emf produced by the function generator (or oscillator) by f. The values of L, R and f relevant in this practical are such that the total potential drop across the circuit is dominated by the drop across the resistor, which means that the current flowing through the circuit, I, is almost the same as if the inductor was not present. Let us thus assume that

$$I = I_0 \cos(\omega t) \tag{2}$$

with I0 a constant and Then, if VR is the voltage across the resistor, we have, by Ohm's law

$$V_R = RI = RI_0 \cos(\omega t) \tag{3}$$

If the resistance of the wire forming the inductor is negligible, then the potential difference between its two ends, VL, is identical to the voltage induced by the change of current, Vind. In this case, from Equation (1), happens when the circuit enters resonance. By the end of the practical you will have seen how resonant circuits can be used to tune radios.

While the work will focus on measurements on a relatively simple circuit, it actually addresses the general issue of how “under-damped” oscillating systems behave when subjected to a periodic perturbation, which is an issue of wide significance in Physics and Engineering. The main features of the currents and voltages measured during the session – their free damped harmonic oscillation, the excitation of transients by a sudden perturbation, the occurrence of resonances under a periodic forcing, and the dependence of the height and width of the resonance peaks and of the accompanying phase changes on the Q-factor of the system – are in fact quite general. They are found in any physical system governed by the same type of differential equations as the LRC circuit considered here.

2. Preparation

A current passing through a coil produces a magnetic field – this is the principle of electromagnets and transformers. If this current changes as a function of time, then so does the magnetic field. As you will see in the Foundations course, the change in the magnetic field present inside the coil induces a potential difference Vind between the two ends of the wire forming the coil, and this potential difference is proportional to the rate of change in the current [1]. By definition, the inductance of the coil (or more generally, of an inductor) is the quantity L such that

$$V_{\text{ind}} = L\frac{dI}{dt} \qquad (1)$$

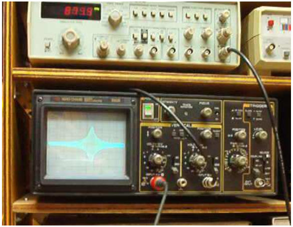

At the top: the sweep signal generator.

Under: oscilloscope with the curve of the LC circuit on the screen.

We can turn the tuner capacitor and get the curve of the LC circuit on the oscilloscope screen.
Adjust with the amplitudeknob of the sweep generator the hight of the peak of the curve to 2.83 cm.
(The peak-peak distance is then: 2x2.83=5.66 cm).

Determine the width of the curve at 2 cm high, this is the -3 dB point (because 2.83x0.707= 2).

Calculate the bandwidth:
BW= (stopfrequency-startfrequency) x curve width at -3 dB / total screenwidth.

And the Q:
Q= f.res / BW

The great advantage of this method is that changes in resonance frequency of the LC circuit, can direct be seen on the screen.
Also changes in Q can direct be seen, because the hight of the peak will change then.

At high Q circuits, we can see the hight of the peak halve for instance, when we touch the (insulated) litzwire with our fingers.

Test setup 3 measuring the Q with a sweep generator and a amplifier.

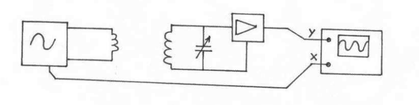

In this schematic we see from left to right the following components:

A sweep signal generator
A coupling coil
The LC circuit
A amplifier
A oscilloscope

When using a 1:100 probe between LC circuit and oscilloscope there are two problems:

a- Because the 100 times attenuation of the signal, the amplitude on the oscilloscope will often have a very low level.
b- The probe can give dielectric losses, which reduces the Q.

To solve these two problems, I replaced the probe by a selfmade amplifier with a gain of 1x.
The amplitude on the oscilloscope will now be 100 times higher then with the 1:100 probe.

Question: Compare this last equation to Equation (18) of Section 2, and show, by identifying the constants between these two equations, (i) that, within a change of notation and of physical dimensions, the amplitude IL0 defined by Equation (21) is nothing else than the amplitude A defined by Equation (14.46) of Young and Freedman; and (ii) that the ratio b / km is the inverse of the Q-factor of the harmonic oscillator. (Hence, in Figure 14.28 of Young and Freedman, the amplitude A is more and more peaked at the resonance for higher and higher values of the Q-factor.)

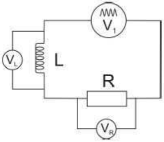

Figure 1: The LR circuit considered in Section 2.

1.What it’s about

This session explores the behaviour of “LRC circuits”, also known as “LCR circuits” and “RLC circuits”, namely circuits containing resistors (R), inductors (L) and capacitors (C). It builds on the second Circuits Skills experiments, in which you looked at circuits containing resistors and capacitors and measured voltages using an oscilloscope. Here the objective of your measurements will be to find the inductance and internal resistance of an inductor and the Q-factor and frequency of oscillation of an LRC circuit. In the process, you will investigate the response of a driven LRC circuit to different frequencies, look at the effect of damping, and see what

Oscillations and Resonances in LRC Circuits

Durham University, UK
http://level1.physics.dur.ac.uk/projects/script/lcr.pdf

Preparatory Task:

Read the entire script and familiarise yourself with the objectives of the session. In particular, read through Section 2 in advance of the session, as the theory developed in this section is essential for understanding the Physics behind the results you'll obtain. You also need to write in your lab book a brief plan of what you will be doing, task by task, and answer the question below. This practical will involve measuring timedependent voltages with an oscilloscope, error analysis, plotting diagrams (with error bars), and least square fitting using the LINEST function of Excel. If you are not confident in having already acquired the necessary skills in these matters, make use of the online resources available at the url http://level1.physics.dur.ac.uk/general/index.php.

In Sections 14.7 and 14.8 of Young and Freedman [1], you'll find a discussion of the dynamics of a damped harmonic oscillator consisting of a mass m sliding on a table under the effect of a spring of force constant k, a frictional force proportional to the velocity of the mass, and an additional force varying with time sinusoidally. In the notation of Young and Freedman, the equation of motion for the position of the mass is

$$m\frac{d^2x}{dt^2} = -b\frac{dx}{dt} - kx + F_{\max}\cos\omega_d t$$

This equation can also be written in the equivalent form

$$\frac{m}{k}\frac{d^2x}{dt^2} + \frac{b}{k}\frac{dx}{dt} + x = \frac{1}{k}F_{\max}\cos\omega_d t$$

The input of the amplifier uses a FET (Field Effect Transistor) and a capacitive voltage divider, which will load the circuit only very little.

A complete schematic of the amplifier you will find here

For the rest, this test setup is the same as test setup 2.

test setup 4 Measuring the Q with a DDS signal generator and a amplifier.

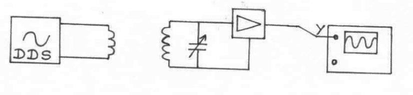

In this schematic you see from left to right:

A DDS signal generator
A coupling coil
The LC circuit
A amplifier
A oscilloscope

DDS means "Direct Digital Synthesis".

The output signal is in a DDS generator made in a digital way.
The great advantage of this kind of generator is the accuracy of the frequency setting.
The output has also a very low distortion.
The DDS generator I use, is a build yourself electronic kit from the company ELV .
You can also buy a complete build and tested module.

The specification are:

Output frequency: 0.1Hz -- 20 MHz.
Output voltage: 0 -- 4 Volt peak peak (not loaded).
Output impedance: 50 Ohm.
Minimum stepsize of frequency setting is 0.1 Hz. (up to 10 MHz output frequency).
Minimum stepsize of frequency setting is 1 Hz. (10 to 20 MHz output frequency).

As stepsize you can also select, e.g. 10 Hz, 100 Hz, 1 KHz, etc.

I now use the DDS generator because the used sweep generator could not be set accurate enough on frequency.
And also the frequency changes slightly during the measurement.

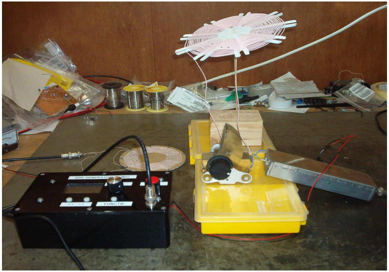

Photo of the test setup.
The coil laying on the table is the coupling coil.

You can get an accurate measure of the width by finding the frequencies just above and just below the resonance at which vC is reduced by $1/\sqrt{2}$ from the peak value. The width $\Delta\omega$ is then $2\pi\Delta f$, where the 2π converts from frequency in Hertz to angular frequency. Do your measured values of $\Delta\omega$ and τ satisfy Eq. 12?

The FET amplifier is connected via short wires to the tuner capacitor.
The coil of the LC circuit is placed at the top of a wooden stick, so it has not much influence from surrounding obstacles.

The windings are laying in a horizontal plane, so the coil picks up less signal from radiostations which can influence the measurement.

During measurements on high Q circuits, I tune the DDS generator in 10 Hz steps.
This test setup is in my opinion very reliable for determining the circuit Q.

Tips for measuring the Q:

During measurements don't come with your hands too close to the LC circuit, because this has influence on frequency and Q.
Keep a minimum distance of 20 cm.

Don't lay the coil of the LC circuit during measurement on the table, but keep a minimum distance of 20 cm from wooden or metal objects.

Driven oscillations

To study the driven solution we need to provide a sine-wave signal source, as shown in Fig. 5. Our source is equivalent to an ideal sine-wave generator in series with a 50 resistor, labeled Rg in the diagram. Since Rg is large enough to affect the damping of the circuit, we add the 3.3 ! resistor in parallel to reduce the effective resistance of the generator. The other change from Fig. 4 is to replace the LabPro with a DMM, set to read AC voltage. (The DMM can read AC voltages between about 30 Hz and 1000 Hz. It is unreliable outside of that range.)

Connect the circuit as shown in Fig 5, and then vary the driving frequency to find the frequency at which vC is maximum. This identifies the resonance frequency, in Hertz. Next, you should plot vC as a function of frequency, taking care to get enough data around the resonance frequency to clearly define the curve. This goes very quickly if you enter f and vC directly into Graph.cmbl and then pick data points to trace the regions of interest. Does your plot look like Fig.3?

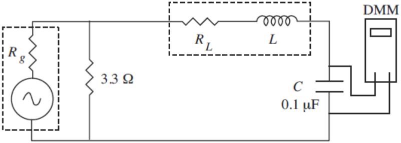

Fig. 5 Circuit for measuring driven oscillations in an RLC circuit.

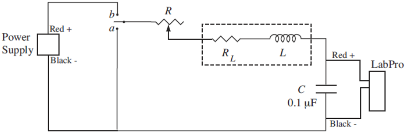

Fig. 4 Circuit for measuring free oscillations in an RLC circuit.

a will start the discharge, and should result in a plot resembling Fig. 2. There may be some irregularity at the beginning, due to bouncing when the switch first makes contact, but you can ignore that in your analysis.

When you have a reasonable looking plot, select the portion of the data after any effects of switch bounce, and fit it to a damped sine wave. You should be able to obtain estimates of the time constant and the angular frequency for your circuit. (Delete the fit or restart RLC.cmbl before taking more data. Otherwise you will have to wait while the program tries to fit the new data.)

Now increase the variable resistor R slightly and obtain another voltage vs time plot. Describe what happens to the time constant and to the shape of the curve as R becomes progressively larger.

Critical damping occurs when R is just big enough that the voltage no longer crosses zero during the decay. Try to find this setting, and measure the value of the variable resistor using a DMM. The exact value is not very clear, but make a reasonable attempt. Reminder: You must disconnect R from the circuit to get an accurate measurement with the DMM.

Q Factor Measurements on L-C Circuits

Jacques Audet, VE2AZX
http://www.arrl.org/files/file/QEX_Next_Issue/Jan-Feb_2012/QEX_1_12_Audet.pdf

The author reviews existing measurement techniques and offers insight into loaded and unloaded Q factors as applied to LC circuits and antennas. A simpler method is proposed that uses an SWR analyzer along with a spreadsheet that easily computes the unloaded Q.

Introduction

The Q factor gives a figure of merit for inductors and capacitors. It is the ratio of reactance to resistance. For filters, it relates directly to the circuit selectivity: The higher the Q, the better the selectivity and the lower the insertion loss of the filter. For oscillators, higher Q also means that lower phase noise is produced. In the case of antennas, a lower Q is generally preferred, giving a larger SWR bandwidth.

Transmission methods are traditionally used for making quality factor (Q) measurements on L-C circuits. This implies that a signal source and an RF voltmeter or spectrum analyzer must be available for such measurements. These are not always available however. Since SWR analyzers are becoming commonplace in many amateur radio rooms, it then becomes tempting to use this instrument for Q measurements on L-C circuits.

Let’s review the existing methods that are currently used for Q measurements that only require scalar measurements — that is, no phase measurements are required. The last method details how the SWR analyzer can be used to measure the unloaded Q of L-C circuits.

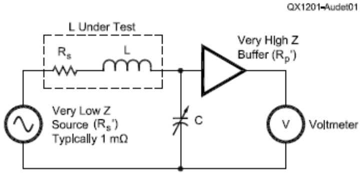

Figure 1 — Block diagram of the classical Q meter as used in the HP 4342A.

1 – The Classical Q Meter Method

This is the technique you would use if you had access to a Q Meter, such as an HP / Agilent Q Meter model HP 4342A. See Figure 1. A very low impedance source is required, typically 1 milliohm, and a very high impedance detector is connected across the L-C circuit. The unloaded Q (QU) of a single reactance component is given by Equation 1.

$$Q_U = \frac{X_S}{R_S} = \frac{R_P}{X_P} \quad \text{[Eq 1]}$$

where XS and RS are the series reactance and loss resistance, and RP and XP are the corresponding parallel loss resistance and reactance components.

In the test set-up we need to make the source resistance RS′ as small as possible, since it adds to the coil or capacitor resistance. Similarly, the detector resistance, RP′ shunting

---

demonstrating the intimate relation between time and frequency response parameters. In obtaining this result, we assumed , so that at the peak.

EXPERIMENTAL PROCEDURE

The RLC circuit is assembled from a large solenoid, a capacitor on the circuit board, and an additional variable resistance to change the damping. The circuit can be charged up with a DC power supply to study the free oscillations, or driven with a sine wave source for forced oscillations.

Free oscillations

To study the homogeneous solution we will use LoggerPro to record the voltage across the capacitor as a function of time. The required circuit is shown in Fig. 4. Since the coil is not made with superconducting wire, we account for the wire resistance with RL. The variable resistor, shown by a resistor symbol with an arrow in the middle, should be set to minimum resistance initially. (You can check with the DMM in ohmmeter mode, before connecting R into the circuit.) The power supply can be set to maximum output.

Start LoggerPro with the file RLC.cmbl. This will configure the program to collect data at 10,000 samples per second, triggered when the voltage decreases across 9.5 V. Set the switch to position b to charge the capacitor, and then start data acquisition. Flipping the switch to position

At the maximum, the oscillation amplitude is considerably greater than the driving amplitude. In fact, if τ were infinite (no damping), the response would be infinite at ω_0.

Since the shape of the peak in vC characterizes the resonance, it is convenient to have some parameter to specify the sharpness of the peak. Traditionally, this is taken to be the full width $\Delta\omega$, shown in Fig. 3, at which the voltage or current have fallen to $1/\sqrt{2}$ of their peak value.

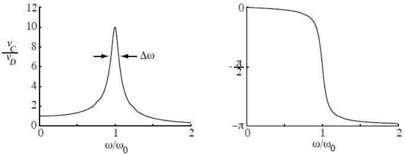

Fig. 3 Amplitude and phase of capacitor voltage as a function of frequency.

The reason for this choice is that the power dissipation is proportional to i^2, so these frequencies correspond to the points at which the power dissipation is half of the maximum. Using Eq. 9 we find that the width is related to τ by

$$\Delta\omega = \frac{2}{\tau} \quad (12)$$

the L-C should be much higher than the RP of the component under test. Note that the L-C circuit is commonly represented as having a resistor in series (RS) with L and C or a shunt resistor (RP) in the parallel model to represent the losses.

Measurement consists of setting the source frequency and adjusting the tuning capacitor, C, for resonance to maximize the voltage across the resonating capacitor C. The Q is calculated using Equation 2.

$$Q = \frac{\text{Voltage Across } C}{\text{Source Voltage}} \quad \text{[Eq 2]}$$

Note that the source voltage is in the millivolt range, since it will be multiplied by the Q factor. A 10 mV source and a Q of 500 will give 5 V across the L-C circuit. In order to preserve the high impedance of the detector even at the higher frequencies, a capacitive voltage divider is used.

This circuit measures the unloaded Q called QU provided that the series resistance RS of the inductor under test is much higher than the source resistance RS′ and the detector parallel resistance RP′ is much larger than the L-C circuit RP.1

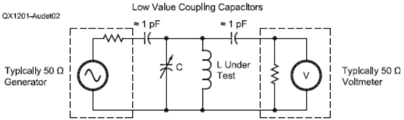

Figure 2 — The signal generator and the source are weakly coupled to the L-C circuit under test, allowing measurement of the 3 dB bandwidth.

2 – Transmission Method Using Coupling Capacitors

A signal is coupled into an L-C parallel circuit using a low value capacitor and extracts the output signal using the same value of capacitor.2 See Figure 2. Note that inductive coupling is also possible, as used in transmission line (cavity) resonators. The – 3 dB bandwidth (BW) is measured at the resonant frequency f0, and the loaded Q (QL) is calculated with Equation 3.

$$Q_L = \frac{f_0}{BW} \quad \text{[Eq 3]}$$

The obtained bandwidth, BW, is a function of the coupling, and another calculation is required to get QU, the unloaded Q. When the input to output coupling is identical, we use Equation 4.

$$Loss = 20\log\left(\frac{Q_U}{Q_U - Q_L}\right) \quad \text{[Eq 4]}$$

where Loss is in positive dB. The ratio QU / QL may now be calculated using Equation 5.

$$\frac{Q_U}{Q_L} = \left(\frac{10^{\frac{Loss}{20}}}{10^{\frac{Loss}{20}} - 1}\right) \quad \text{[Eq 5]}$$

1/# is referred to as "critical damping" because the charge reaches zero in the shortest time without changing polarity.

The solution for sine-wave driving describes a steady oscillation at the frequency of the driving voltage:

$$q_C = A\sin(\omega t + \phi) \quad (8)$$

We can find A and ϕ by substituting into the differential equation and solving:

$$A = \frac{v_D / L}{\left[\left(\omega_0^2 - \omega^2\right)^2 + \left(2\omega / \tau\right)^2\right]^{1/2}} \quad (9)$$

$$\tan\phi = \frac{2\omega / \tau}{\left(\omega^2 - \omega_0^2\right)} \quad (10)$$

These two equations are plotted in Fig. 3, where we use the fact that $v_C = q_C/C$ to plot the voltage across the capacitor relative to the driving voltage. (Our apparatus does not allow us to observe the phase of the response, so we won't consider that further.) The angular frequency for maximum amplitude is given by

$$\omega_{peak} = \omega_0\left[1 - 2/\left(\omega_0\tau\right)^2\right]^{1/2} \quad (11)$$

Where qC0 and ! are determined by initial conditions, and

$$\omega_1 = \omega_0\left[1 - (\omega_0 \tau)^{-2}\right]^{1/2} \quad (5)$$

This solution is plotted in Fig. 2 for a case where the capacitor is initially charged and no current is flowing. (For a mass on a spring the equivalent situation would be to pull the mass aside and release it from rest.) Evidently there are oscillations at ω_1, approximately equal to ω_0, within an exponential envelope. Note that the amplitude falls to 1/e of the initial value when $t = \tau$.

As $\omega_0 < 1/\tau$ gets smaller (larger resistance R), ω_1 becomes smaller and finally imaginary. The corresponding solutions do not oscillate at all. For $\omega_0 < 1/\tau$, there are two exponentials

$$q_C = A_1 e^{-t/\tau_1} + A_2 e^{-t/\tau_2} \quad (6)$$

where τ_1 and τ_2 differ somewhat from τ. When $\omega_0 = 1/\tau$, the solution is slightly simpler:

$$q_C = \left(A_1 + A_2 t\right) e^{-t/\tau} \quad (7)$$

If the capacitor is initially charged, these results tell us we will get a monotonic decay for sufficiently large R. The case "0 =

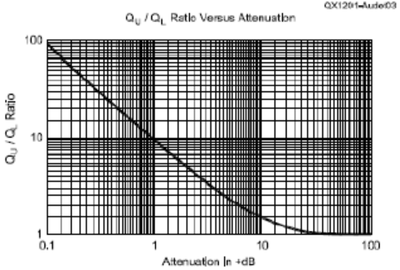

Figure 3 — Graph showing the ratio of unloaded Q (QU) to loaded Q (QL) as a function of the L-C circuit attenuation at resonance. Applies to a single tuned circuit with identical in / out coupling. When the attenuation is large (> 20 dB) the coupling is minimal and the loaded Q closely approaches the unloaded Q.

This equation is plotted in Figure 3. Note that this method applies well to transmission line resonators with equal impedance inductive coupling loops at the input / output. Note also that these two methods do not require knowledge of the L or C values to compute the Q factors.

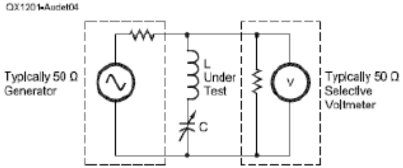

Figure 4 — Here the L-C circuit is connected in series and across the generator - detector. At resonance the L-C circuit has minimum impedance and by measuring the attenuation, the effective series resistance ESR of the L-C series may be computed at the resonant frequency f0.

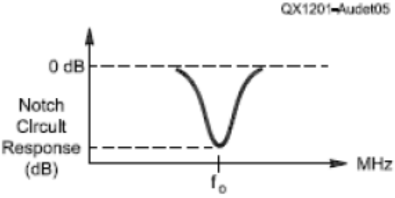

Figure 5 — Frequency response of the notch circuit of Figure 4, where f0 is the frequency of maximum attenuation.

3 – Shunt Mode Transmission Method With L-C in Series

with $\tau = 2L/R$ and $\omega_0^2 = 1/LC$. Two situations will be of interest. We will examine the free oscillations, when vD is exactly zero, meaning that the sine-wave generator has been replaced with a short circuit. We will then connect the sine wave generator and calculate the response as a function of frequency.

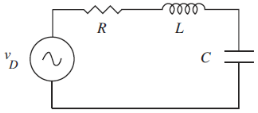

Fig. 1 Idealized series RLC circuit driven by a sine-wave voltage source.

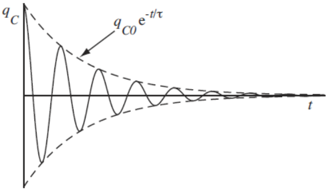

Fig. 2 Damped oscillation, showing decay envelope

If the resistance in the circuit is small, the free oscillations are of the form

$$q_C = q_{C0}e^{-t/\tau}\cos(\omega_1 t + \phi) \quad (4)$$

RLC Circuits
Rice University
http://www.owlnet.rice.edu/~phys102/Lab/RLC_circuits.pdf

It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiment, it's wrong.
Richard Feynman (1918-1988)

OBJECTIVES

To observe free and driven oscillations of an RLC circuit.

THEORY

The circuit of interest is shown in Fig. 1, including sine-wave sources. We start with the series connection, writing Kirchoff's law for the loop in terms of the charge qC on the capacitor and the current $i = dq_C/dt$ in the loop. The sum of the voltages around the loop must be zero, so we obtain

$$v_L + v_R + v_C = v_D \sin(\omega t) \quad (1)$$

$$L\frac{d^2q_C}{dt^2} + R\frac{dq_C}{dt} + \frac{q_C}{C} = v_D \sin(\omega t) \quad (2)$$

For reasons that will become clear shortly, we rewrite this as

$$\frac{d^2q_C}{dt^2} + \frac{2}{\tau}\frac{dq_C}{dt} + \omega_0^2 q_C = \frac{v_D}{L}\sin(\omega t) \quad (3)$$

Here the L and C are series connected and placed in shunt across the transmission circuit. See Figure 4. (This method is also detailed in the reference of Note 2.) It basically computes the effective series resistance (ESR) of the combined L and C, based on the attenuation in dB at resonance, using a selective voltmeter to prevent source harmonics from affecting the measurement. The ESR may be calculated using Equation 6.

$$ESR = \frac{Z}{2\left(10^{\frac{dB}{20}} - 1\right)} \quad \text{[Eq 6]}$$

where dB is a positive value of attenuation and Z is the source and detector impedance. Once the ESR is known, it is only necessary to compute:

$$Q_U = \frac{X_L}{ESR} \quad \text{[Eq 7]}$$

where XL = 2π f0L, and L has already been measured separately.

Note 3 gives a reference to a spreadsheet that I have developed to eliminate the need to measure L and C. Two more attenuation points are required to compute these values. There is also a version for crystals, which computes the equivalent R L C values and other related parameters.

4 – Reflection Measurement Using an SWR Analyzer

This method has been recently developed to make use of an SWR analyzer, thus eliminating the need for the source detector combination. Adjust the SWR analyzer to the resonant frequency of the circuit. The L-C circuit may be coupled to a link coil on the SWR analyzer, which provides variable coupling. As shown in Figure 7, the amount of coupling is adjusted until the SWR drops to 1:1. The frequency is recorded as f0. Then the frequency is offset above or below f0 to obtain an SWR reading between 2 and 5. Now plug the new frequency and SWR values in the spreadsheet that I provide to calculate the unloaded Q factor. Note that the L or C values are not required to compute the Q factor.

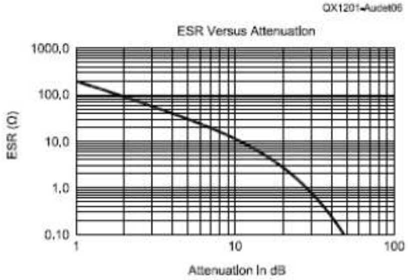

Figure 6 — Effective series resistance as computed from the attenuation in dB, in a 50 Ω system.

When $R = 0 \Rightarrow R_{total} = R_p \Rightarrow R_p = \beta_{exp} \cdot 2L$

3. Calculate experimental angular frequency of oscillations

$$\omega_R = \frac{2\pi}{T}$$

4. For the resonance conditions in RLC circuit, calculate capacitance C for each inductance L using the equations:

5. Calculate the uncertainty of the capacitance _C

6. For the case of critical damping, compare the experimental value of critical resistance RC with its theoretical value calculated on the basis of the equation:

$$R_{Ctheory} = 2\sqrt{\frac{L}{C}}$$

7. Write the conclusions.

Literature:

1. Halliday, Resnick “Fundamentals of Physics - 8th edition”, John Wiley 2007

2. Zięba “Pracownia Fizyczna Wydziału Fizyki I Techniki Jądrowej AGH”, Uczelniane Wydawnictwo Naukowo-Dydaktyczne 1999.

Updated: 11.02.2011 by Barbara Dziurdzia resistor and repeat the observations. Investigate how the increase in resistance R affects the oscillation parameters.

6. Increase R until the critical aperiodic oscillations occur and write down the Rcritical in the measurement table.

7. Change the inductance to L2 at the decade coil and for resistances R0=0, R1, R2 repeat all measurements according to 2-6. Investigate how the increase in inductance L affects the oscillation parameters.

8. Write down the results in Table 1.

6. Data Handling

1. Calculate the log decrements $\delta = \ln\left(\frac{U_i}{U_{i+2}}\right)$ and the experimental damping coefficient β exp of oscillations: $\beta_{\exp} = \frac{\ln \frac{U_i}{U_{i+2}}}{T}$ for all combinations (L1, R0), (L1, R1), (L1, R2), (L2, R0), (L2, R1), (L2, R2)

2. Calculate the parasitic resistance of a coil Rp taking advantage from the fact that at the beginning there is always set R=0 at the resistance decade. the total resistance Rtotal in the RLC circuit consists of the resistance of a decade R and the parasitic resistance of a coil R_p: $R_{total} = R + R_p$.

$$\beta_{\exp} = \frac{R_{total}}{2L}, \quad R_{total} = \beta_{\exp} \cdot 2L$$

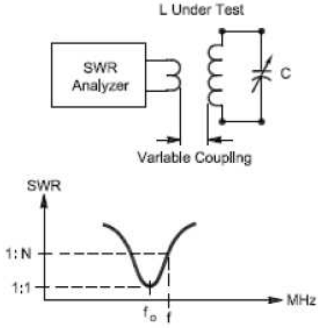
L Under Test
SWR Analyzer
C
Variable Coupling
SWR
1:N
1:1
MHz
f0
f

Figure 7 — Set-up for measuring the unloaded Q of L-C resonant circuit. The coupling is adjusted for 1:1 SWR at resonance. The link should have approximately 1 turn per 5 or 10 turns of the inductor under test.

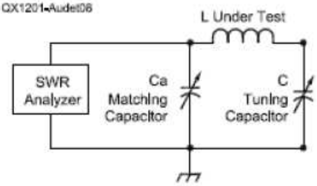
QX1201-Audet06
L Under Test
SWR Analyzer
Ca Matching Capacitor
C Tuning Capacitor

Figure 8 — Here the variable coupling to the SWR analyzer is provided by the capacitive divider formed by Ca and C. Ca is typically 10 to 50 times larger than C.

As shown in Figure 8 the coupling may also be realized with a second variable capacitor, Ca, which serves as an impedance divider. Both capacitors have their frame grounded, along with the SWR analyzer. This makes the construction and operation of the SWR - Q meter much easier.

Both circuits transform the L-C circuit effective parallel resistance RP to 50 + j 0 ohm as required by the analyzer to obtain a 1:1 SWR at resonance. The capacitance value of Ca will be in the order of 10 to 50 times the value of C, the main tuning capacitor. A slight interaction will be present between these two adjustments when adjusting for 1:1 SWR.

Note that a variable inductive divider can be used, in series with the inductor under test, instead of the capacitive divider of Figure 8. In this case the SWR analyzer is connected across this variable inductor, which allows matching to the SWR analyzer 50 ohm input. I do not recommend this as the Q of the variable inductor is probably not going to be as high as the matching capacitor, Ca. The Q of these matching components have a second order effect on the measured QU.

At a Q factor of 300, a 1% error in SWR gives a 1% error in Q. Assume that the measured SWR is 4 and the SWR meter has a resolution of 0.1. If the error in SWR is 0.1, then the error in Q is 2.5%. Other errors include the Q of the tuning capacitor, and the Q of the matching component, Ca, or the link coil to a lesser degree.

2. Set the inductance at the decade coil to the certain value L1 . Set the resistance at the decade resistor to R0=0. Observe the damping oscillations on the oscilloscope selecting the proper time base setting and vertical gain.

3. Measure on the oscilloscope display the period of oscillations T (Fig. 8). To increase accuracy, measure periods of a few successive oscillations and calculate the average value which will be further used to get the experimental angular frequency ω exp of oscillations:

$$\omega_{exp} = \frac{2\pi}{T}$$

4. Measure on the oscilloscope display the amplitude of maxima U2, U4, (Fig. 8)

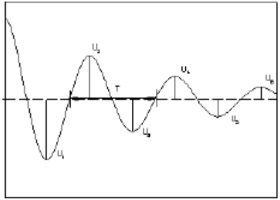

Fig. 8.. Damping oscillations

5. Keep the inductance L1 at the decade coil unchanged and set the another resistance R1 at the decade resistor. Observe the damping oscillations for the settings (L1, R1) and repeat the measurements 3-4. Set the resistance R2 at the decade

connected in series . The oscilloscope monitors the potential difference uC(t) across the capacitor as a function of time.

5. Measurements

1. Set the RLC circuit as shown in Fig. 6 . Fig. 7 shows the decade coil, the decade resistor and the capacitor with its charging system.

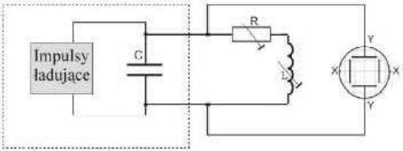

Fig. 6. The measurement system

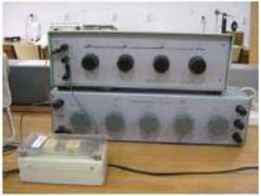

Fig. 7. The decade coil, the decade resistor and the capacitor with its charging system.

Here are some simulation results:

At Q = 200 the average error on the calculated Q is – 0.5% when the test frequency is below resonance. Above resonance, the average error drops to – 0.05%.

At Q = 50 the average error on the calculated Q is –1.0% when the test frequency is below resonance. Above resonance, the average error drops to –0.7%.

These small errors come from the reactance of the coupling element (link coil or Ca in the capacitive divider).

The following errors were obtained by simulation with an inductor Q of 200.

The matching capacitor Ca should have a Q of 500 or larger. A Q value of 500 for Ca gives – 6% error and a Q value of 1000 for Ca gives – 0.6% error. Interestingly the link coil Q may be as low as 50 and the error is only – 0.4%. The losses of the transmission line between the SWR analyzer and the Q measuring circuit should be kept to a minimum. An attenuation of 0.05 dB gives –1.6% error on the calculated Q.

So far the Q factor measurement included both the combined Qs of the L and C, since it is difficult to separate their individual Q values. When the capacitor Q (Qcap) is known the inductor Q (Qind) may be calculated like parallel resistors:

$$\frac{1}{Q_{ind}} = \frac{1}{Q} - \frac{1}{Q_{cap}} \quad \text{[Eq 8]}$$

Note that the combined Q of the L-C circuit is 17% lower with Qind = 200 and Qcap = 1000, compared to having a capacitor with an infinite Q.

Measuring the Q of Low Value Capacitors (< 100 pF) Using the Delta Q Method

Two Q measurements are required:

First measure the Q of a test inductor (preferably having a stable high Q) and record it as Q1 and record the amount of capacitance C1 used to resonate it.

Second, connect the low value capacitor to be tested across the tuning capacitor, decrease its capacitance to obtain resonance again. Note the measured Q as Q2 and the amount of capacitance C2 used to resonate. (Q2 should be less than Q1). Compute the Q of this capacitor as follows:

$$Q = \frac{(C_1 - C_2) \times Q_1 \times Q_2}{C_1 \times (Q_1 - Q_2)} \quad \text{[Eq 9]}$$

Note that (C1 – C2) is the capacitance of interest. It may be measured separately using a C meter. If Q2 = Q1 the Q of the capacitor becomes infinite.

Unloaded and Loaded Q

Looking back at method 1, the measured Q is the loaded Q and it approaches the unloaded Q as the source impedance goes toward zero and the detector impedance is infinite.

With method 2, the measured Q is the loaded Q and a correction is required based on insertion loss to obtain the

---

For damped oscillations, the log decrement is equal to:

$$\delta = \ln\left(\frac{V_1}{V_2}\right) = \frac{R}{2L}T$$

Critical damping in an RLC circuit is achieved when:

$$\frac{1}{LC} - \left(\frac{R}{2L}\right)^2 = 0,$$

from which we obtain:

$$R_{critical} = 2\sqrt{\frac{L}{C}}$$

We vary the resistance and search for a signal that has no undershoot and has a maximum decay rate. Typical oscilloscope traces for the overdamped, underdamped and critical damping are shown in Fig. 5.

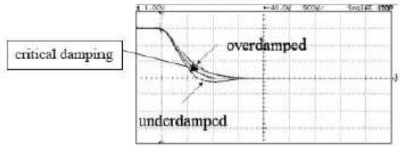

Fig. 5. Typical oscilloscope traces for critical damping critical damping

4. Equipment

RLC circuit consists of a capacitor C combined with its charging system, a decade resistor and a decade coil – all

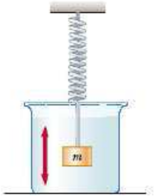

Fig. 3. A mechanical analogous to the RLC circuit – a block–spring system moving in a viscous medium.

The log decrement is determined from the ratio of the voltages of successive maxima of the damped oscillatory signal. This measurement is taken from the oscilloscope display, as shown in the Fig. 4. The figure shows the damped signal for two different values of resistance for fixed value of capacitance. The arrows indicate the first and second maxima.

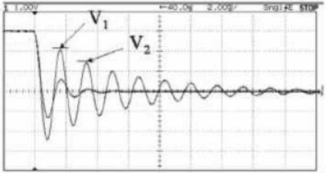

Fig. 4. Typical oscilloscope trace showing log decrement measurement

---

unloaded Q. In this case the L-C circuit under test “sees” the same Q as we have measured, (if it could look at its environment by looking towards the source and detector).

Such is not the case in methods 3 and 4. In method 3 the L-C circuit will typically “see” a 50 ohm source and a 50 ohm detector that are effectively in parallel. This adds 25 ohm in series with the ESR of the L-C circuit, and reduces the Q that it sees. This loaded Q sets the selectivity of the notch filter created by the circuit.

In method 4, the L-C “sees” the SWR analyzer internal impedance, say 50 ohm. Remember that the Rp value is transformed to 50 ohm at the analyzer, and effectively is in parallel with the analyzer’s internal impedance. This means that the actual Q factor seen by the L-C circuit is reduced by half. This value is the loaded Q of the L-C circuit. In this case the coupling factor equals 1.5

Now consider a 50 ohm dipole antenna. The series R-L-C model may be applied to a resonant dipole antenna, which presents approximately 50 ohm at the feed point. When the dipole is fed by a low loss transmission line the SWR meter connected at the transmitter will measure the antenna unloaded Q at an SWR = 2.62 as calculated from equation 3: Q = f0 / BW. Note that the spreadsheet given in Note 4 will calculate the dipole Q factor under these conditions.

When the antenna is fed by a 50 ohm source, its effective bandwidth will double, since the total resistance seen by the dipole is now the radiation resistance plus the transmitter output resistance.

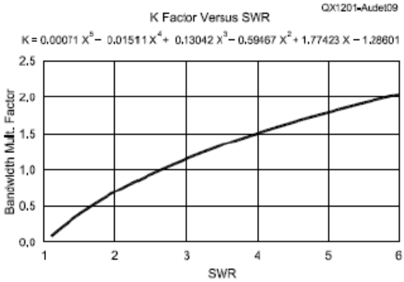

Figure 9 — This graph shows the K factor versus SWR. At SWR = 2.62 the correction factor K equals 1 while at SWR = 5.83 the K factor is 2.

Here we still have the loaded Q = unloaded Q / 2. This assumes that the transmitter output impedance and the feed line are 50 ohm. The Q derived from the bandwidth at SWR points of 2.62 gives the unloaded Q value, independently of the transmitter output impedance.

In the general case, the complex impedance of the transmitter reflected at the antenna will modify its effective bandwidth and its actual resonant frequency. Note that the SWR meter connected at the transmitter will not show this effect since it can only measure the unloaded Q of the dipole. Also, the bandwidth of the matching circuits at the transmitter will affect the effective impedance seen by the antenna and modify its loaded Q and bandwidth.

is a circuit damping constant

The equation **** describes a sinusoidal oscillation with an exponentially decaying amplitude $Qe^{-\frac{Rt}{2L}}$.

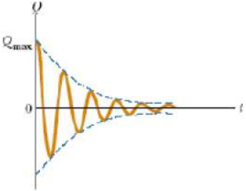

Fig. 2 Charge versus time for the damped RLC circuit.

The RLC circuit is analogous to the damped harmonic oscillator illustrated in Fig. 3 and described by the equation:

$$m\frac{d^2x}{dt^2}+b\frac{dx}{dt}+kx=0$$

Q corresponds to the position x of the block at any instant, L to the mass m of the block, R to the damping coefficient b, and C to 1/k, where k is the force constant of the spring.

$$\frac{dU}{dt} = Li\frac{di}{dt} + \frac{q}{C}\frac{dq}{dt} = -i^2 R$$

$$i = \frac{dq}{dt} \text{ and } \frac{di}{dt} = \frac{d^2 q}{dt^2}$$

$$L\frac{d^2 q}{dt^2} + R\frac{dq}{dt} + \frac{1}{C}q = 0$$

$$\frac{d^2 q}{dt^2} + 2\frac{R}{2L}\frac{dq}{dt} + \frac{1}{LC}q = 0$$

$$\frac{R}{2L} = \beta, \qquad \frac{1}{LC} = \omega_0^2$$

$$\frac{d^2 q}{dt^2} + 2\beta\frac{dq}{dt} + \omega_0^2 q = 0$$ ***

The solution to this differential equation for damped oscillations in an RLC circuit is:

$$q = Qe^{-\frac{Rt}{2L}}\cos(\omega t + \varphi)$$ ****

where:

$$\omega = \sqrt{\omega_0^{\,2} - \left(\frac{R}{2L}\right)^2}$$

is the angular frequency of the damped oscillations,

$$\omega_0 = \frac{1}{\sqrt{LC}}$$

is the angular frequency of the undamped oscillations.

$$\omega = \sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2}$$

$$\beta = \frac{R}{2L}$$

One way to measure the effective bandwidth of the dipole might be to insert an RF ammeter in series with one dipole leg. Find the frequencies where the current is reduced by ~ 30% and compute the effective bandwidth this way. In the receive mode the dipole antenna will also exhibit the same loaded Q = unloaded Q / 2 if the receiver impedance is 50 ohm. Your antenna effective bandwidth may be twice as large as you really thought!

Summary of the Four Methods Presented in this Article

1 – Classical Q Meter. See Figure 1. This shows a set-up as used by the HP / Agilent model HP 4342A.

This technique uses a transmission method that requires a very low impedance source and very high impedance detector, which are not easy to realize.

Set the frequency and adjust the reference capacitor for resonance.

The measurement approximates the unloaded Q. Corrections are difficult to apply.

You don’t have to know the L-C values.

2 – Transmission Method Using Coupling. See Figure 2. This technique requires a source and a voltmeter.

Find the –3 dB transmission bandwidth points. The measurement requires two low value, high Q coupling capacitors.

It measures the loaded Q as used in the test circuit.

You need to correct the loaded Q value for attenuation to find the unloaded Q, if the attenuation is less than 30 dB.

The accuracy of the Q measurement remains somewhat dependant upon the Q factor of the coupling capacitors. Refer to Figure 2. With 1 pF coupling capacitors having a Q of 1000, the error on Q was – 1.4%.

The L or C values are not required.

3 – Shunt mode transmission method with LC connected in series. See Figure 4.

This method requires a source and a selective voltmeter to prevent source harmonics from affecting the measurement. Compute the ESR from the minimum attenuation measured at resonance.

The L or C values are required to compute the Q value.

The author’s spreadsheet computes the series/parallel R-L-C values from 3 attenuation measurements. This technique is also useful for crystal measurements.

This is potentially the most accurate method, if the attenuation is measured with high accuracy using a vector network analyzer (VNA).

4 – Reflection measurement using an SWR analyzer. See Figures 7 and 8.

oscillations of charge, current and potential difference continuously decrease in amplitude, and the oscillations are said to be damped.

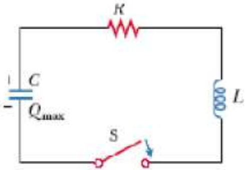

Let's write the equation for the total electromagnetic energy U in the circuit at any instant. As we know the resistance doesn't store electromagnetic energy, so we have:

$$U = U_B + U_E = \frac{Li^2}{2} + \frac{q^2}{2C} \quad *$$

This total energy decreases because energy is transferred to thermal energy. The rate of this
transfer is:

$$\frac{dU}{dt} = -i^2 R \quad **$$

where the minus sign indicates that U decreases.
By differentiating * with respect to time and then substituting the results in ** we obtain:

Damped oscillations in RLC circuits

by Barbara Dziurdzia
AGH University of Science and Technology in Cracow
http://home.agh.edu.pl/~lyson/downloads/Manual_8.pdf

1. Goal

To observe damped oscillations in the RLC circuit and measure the amplitude, period, angular frequency, damping constant and log decrement of damped oscillatory signals. To find the critical resistance for which the critical damping occurs.

2. What to learn?

Transfer of energy in LC circuit. The electrical-mechanical analogy. Differential equation describing damped simple harmonic motion in the RLC circuit.. Solution of this equation. Angular frequency of the damped oscillator. Damping constant. Angular frequency of the undamped oscillator. Forced oscillations and resonance. Kirchhoff's rules. Log decrement of damped oscillatory signals. Critical damping. How does the oscilloscope work?

3. Background

A circuit consisting of an inductor, a capacitor, and a resistor connected in series (Fig. 1) is called an RLC circuit. The resistance of the resistor R represents all of the resistance in the circuit. .With the resistance R present, the total electromagnetic energy U of the circuit (the sum of the electrical energy and magnetic energy) decreases with time because some portion of this energy is transferred to thermal energy in the resistance. Because of this loss of energy, the

---

This technique requires a matching capacitor or variable link coupling. There will be a slight interaction between the matching components and the main tuning capacitor.

Adjust two variable capacitors or vary link coupling to obtain an SWR of approximately 1:1

Toroidal inductors are difficult to test with variable link coupling. Use the capacitive divider method instead. With link coupling, keep the link close to the grounded end of the coil, to minimize capacitive coupling. In general air wound coils are more delicate to test since they are sensitive to their environment.

The L and C values are not required.

Offset the frequency to have an SWR increase. Note the frequency and SWR.

The author’s spreadsheet computes the unloaded Q factor directly.

This technique requires the least equipment of any of the methods.

Best accuracy is obtained with a link coil or a variable capacitive divider.

While all four methods do not require phase or complex impedance measurements, method 4 only requires an SWR analyzer. The accuracy of the SWR analyzer method has been validated from RF Simulations and by actual measurements.

Since this method relies on SWR measurements, the accuracy of the measuring instrument should be verified. The SWR = 2 point may easily verified by placing two accurate (≤ 1%) 50 ohm resistors in parallel at the analyzer output using a tee connector. Note the reading obtained at the frequency of interest and compute the SWR offset from the ideal value of 2.00. Then perform the Q measurements around this same SWR value, possibly above and below the resonant frequency and average the Q results.

To compensate for inaccuracies in determining the frequency (f0) of 1:1 SWR, it is recommended that you perform the Q measurement at two frequencies above and below the f0, as done in the spreadsheet provided in Note 4.

Note 5 covers other methods of measuring Q using a vector network analyzer, mostly suitable for use at microwave frequencies.

Jacques Audet, VE2AZX, became interested in radio at the age of 14, after playing with crystal radio sets and repairing old receivers. At 17, he obtained his first ham license, and in 1967 he obtained his B Sc degree in electrical engineering from Laval University. He then worked in engineering functions at Nortel Networks, where he retired in 2000. He worked mostly in test engineering on a number of products and components operating from dc to light-wave frequencies.

His areas of interest are in RF simulations, filters, duplexers, antennas and using computers to develop new test techniques in measurement and data processing.

Notes

Let us summarize what we have learned about parallel and series RLC circuits:

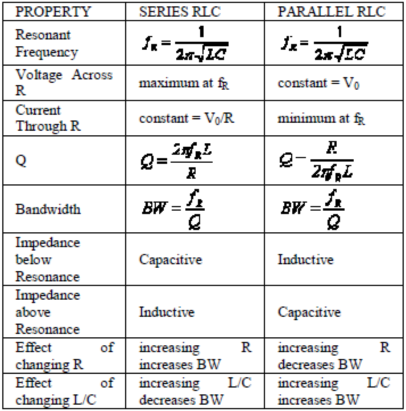

PROPERTY	SERIES RLC	PARALLEL RLC
Resonant Frequency	$f_R = \frac{1}{2\pi\sqrt{LC}}$	$f_R = \frac{1}{2\pi\sqrt{LC}}$
Voltage Across R	maximum at f_R	constant = V_0
Current Through R	constant = V_0/R	minimum at f_R
Q	$Q = \frac{2\pi f_R L}{R}$	$Q = \frac{R}{2\pi f_R L}$
Bandwidth	$BW = \frac{f_R}{Q}$	$BW = \frac{f_R}{Q}$
Impedance below Resonance	Capacitive	Inductive
Impedance above Resonance	Inductive	Capacitive
Effect of changing R	increasing R increases BW	increasing R decreases BW
Effect of changing L/C	increasing L/C decreases BW	increasing L/C increases BW

frequency at which the current increases to 141.4% of the minimum. The 141.4% current points were chosen because they correspond to a doubling of the power.

$BW = f_U - f_L$

Q is defined as:

$$Q = \frac{R}{2\pi f_R L}$$

Bandwidth and Q are also related as follows:

$$BW = \frac{f_R}{Q}$$

The graph below illustrates the definition of BW.

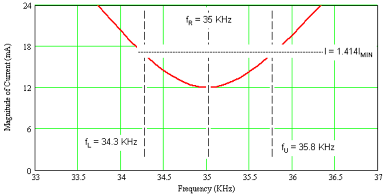

The bandwidth is 35.8 - 34.3 = 1.5 KHz. Now we can determine Q:

$$Q = \frac{f_R}{BW} = \frac{35 KHz}{1.5 KHz} = 23.3$$

As a check, we can use our other definition to determine Q:

$$Q = \frac{R}{2\pi f_R L} = \frac{1000}{42.89} = 23.3$$

As one would expect, both results agree.

1Hewlett Packard Journal, September 1970,
www.hpl.hp.com/hpjournal/pdfs/IssuePDFs/1970-09.pdf

2Wes Hayward W7ZOI "Two Faces of Q,"
w7zoi.net/2faces/twofaces.html

3There is an Excel spreadsheet to perform the various Q calculations on my website: ve2azx.net/technical/Calc_Series-Par_RLC.xls for RLC circuits.

4The MathCad™ and the corresponding Adobe PDF file as well as the Excel file may be downloaded from the ARRL website at www.arrl.org/qexfiles/. Look for the file 1x12_Audet.zip.

5 Darko Kajfez "Q Factor Measurements, Analog and Digital,"
www.ee.olemiss.edu/darko/rfqmeas2b.pdf

Appendix 1

In the SWR analyzer method, the equivalent parallel resistance RP of the L-C circuit is transformed to 50 ohm to give 1:1 SWR at resonance by using an adjustable link coupling or a variable capacitive divider. As the frequency is varied around resonance, it may be set at two frequencies where the absolute value of the reactance, X, is equal to RP.

Abs(X) = RP

Normalizing these impedances gives impedances of 1 for RP and j for X. Since these are in parallel, the resulting impedance, Z, will be:

$$Z = \frac{1 \times j}{1 + j} = 0.5 + 0.5j$$ [Eq A-1]

Taking the absolute value of Z we get Z = 0.707. This is the –3 dB point, since this value is 3 dB below the initial value of 1. The complex reflection coefficient is:

$$\rho = \frac{(0.5+0.5j)-1}{(0.5+0.5j)+1} = -0.2+0.4j$$ [Eq A-2]

The absolute value of the reflection coefficient $|\rho|$ is 0.447. The return loss RL in dB is given by:

RL = –20 log $|\rho|$ = 6.99 [Eq A-3]

The corresponding SWR is given as:

$$SWR = \frac{1+\rho}{1-\rho} = 2.62$$ [Eq A-4]

At the – 3 dB points, the bandwidth (BW) is related to the unloaded Q factor as follows:

$$Q_U = \frac{f_0}{BW}$$

Therefore, measuring the bandwidth, BW, at SWR = 2.62 allows us to compute the unloaded Q (QU) of the circuit. So far, we need three measurement points. One point at frequency f0 that gives 1:1 SWR and two points at fL and fH below and

The shape of the band-stop response of a parallel RLC circuit depends on the value of R and L/C as shown in the next graphs.

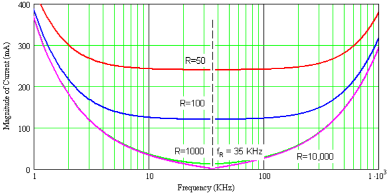

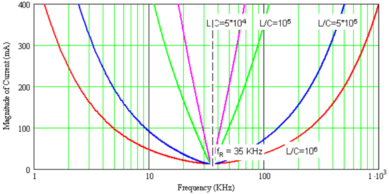

Note that the band-stop characteristic becomes narrower as the value of R increases. This the opposite of what we saw in the series RLC circuit. The response also becomes narrower as the L/C ratio decreases, which again is the opposite of what we saw in the series RLC circuit. We can summarize this behavior by using two additional parameters, Q and the bandwidth, BW. We define bandwidth for parallel RLC circuits, as the difference in frequency between the upper frequency at which the current increases to 141.4% of the maximum and the lower frequencies near the resonant frequency, while allowing others to pass.

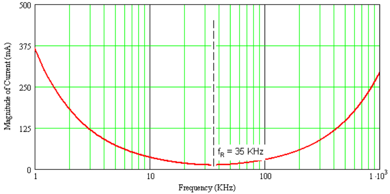

The graph below shows the variation in phase shift of the current through the parallel RLC circuit.

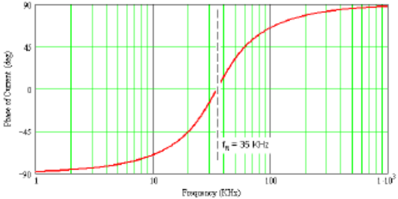

Notice that the current has no phase shift at the resonant frequency. At frequencies below the resonant frequency the current lags the applied voltage and the circuit acts like an inductor. At frequencies above resonance, the current leads the applied voltage and the circuit acts like a capacitor.

---

above resonance that give an SWR of 2.62. The last two points are symmetrical around f0, so only one is measured as f = fH.

Since fL × fH = f0 [Eq A-5]

The other frequency fX is assumed to be at:

$$f_X = \frac{f_0^2}{f_H} = \frac{f_0^2}{f}$$ [Eq A-6]

The bandwidth BW1 may now be calculated:

$$BW_1 = f - f_X = f - \frac{f_0^2}{f}$$ [Eq A-7]

Since we want to be able to measure at any SWR value, we need to add a correction factor, K, resulting from the use of bandwidth BW1.

$$Q_U = \frac{K\, f_0}{BW_1}$$ [Eq A-8]

The correction factor, K, is the bandwidth multiplication factor. It is a function of the SWR. I derived K using Mathcad™ numerical calculations and then normalized the value. I used a fifth order polynomial to fit the computed data. See Figure 9. Note 4 gives the Mathcad™ and Adobe PDF files that I used, as well as more details on the derivations. The complete equation as used in my spreadsheet for the unloaded Q factor becomes:

$$Q_0 = \frac{f_0 \cdot \left(1.77423 \cdot SWR - 0.59467 \cdot SWR^2 + 0.13042 \cdot SWR^3 - 0.0151 \cdot SWR^4 + 0.00071 \cdot SWR^5 - 1.28601\right)}{\left| f - \frac{f_0^2}{f} \right|}$$

[Eq A-9]

Let us look at a specific parallel RLC circuit and explore its behavior.

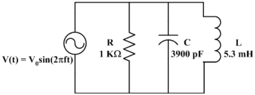

The graph below shows the variation of the admittance of this parallel RLC circuit with frequency.

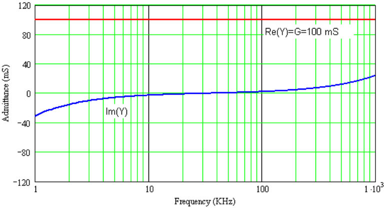

We can compute the resonant frequency from the component values:

$$f_0 = \frac{1}{2\pi\sqrt{LC}} = \frac{0.159}{\sqrt{(5.3*10^{-3})(3900*10^{-12})}} = \frac{0.159}{\sqrt{2.067*10^{-11}}} = \frac{0.159}{4.546*10^{-6}} = 35.0\ \text{KHz}$$

Note that this parallel RLC circuit has the same resonant frequency as the series RLC circuit examined earlier. In general, the resonant frequency depends only on L and C and not whether they are connected in parallel or series.

The next graph shows the variation in the current through the parallel RLC circuit. Notice that the current is a minimum at the resonant frequency. This is an example of a "band-stop" circuit response. A parallel RLC circuit can be used to block

Inductive susceptance:

$$B_L = \frac{-j}{2\pi fL}$$

Capacitive susceptance:

$$B_C = j2\pi fC$$

And of course, conductance is defined as follows:

$$G = \frac{1}{R}$$

By analyzing the parallel RLC circuit using admittance, we will be able to use nearly all of what we have already learned about series RLC circuits. Admittances in parallel add, just like impedances in series. We can immediately write the equation for the admittance of the parallel RLC circuit by adding the admittances of the three components:

$$Y = G + B_C + B_L = \frac{1}{R} + j2\pi fC - \frac{j}{2\pi fL}$$

This can be simplified to the following expression:

$$Y = G - j2\pi fC\left(1 - \frac{1}{4\pi^2 f^2 LC}\right)$$

The imaginary part of the admittance vanishes when:

$$4\pi^2 f^2 LC = 1$$

We define the frequency at which this occurs to be the resonant frequency, given by the following formula:

$$f_R = \frac{1}{2\pi\sqrt{LC}}$$

The equation for the admittance of the parallel RLC circuit can be simplified further:

$$Y = G + j2\pi fC\left(1 - \left(\frac{f_R}{f}\right)^2\right)$$

Experiments with Coils and Q-Measurement

Wes Hayward, w7zoi,
October, 2007 (Updates 01Dec07, 08Dec08.)
http://web.archive.org/web/20101226143919/http://w7zoi.net/coilq.pdf

I recently became interested in building a “Zero Power Receiver (ZPR)”, a circuit that would receive the AM broadcast band while using no external power. There would be no batteries or other sources of energy. This is, of course, the classic crystal set that most of us built in our youth. But building a contemporary ZPR is a different exercise than it was for those youthful interludes. The main difference is the presence of considerably more science in the process than we used in the construction of that first crystal set. There are some really wonderful web sites out there that present much of this work. Another difference is that the modern ZPR may not even use a “crystal,” or diode detector. Instead, it may use a field effect transistor. The most recent published design that I know about was presented by Bob Culter, N7FKI, in QST for January, 2007. Bob used a new zero threshold MOSFET in his ZPR.

The w7zoi amateur radio set up and listening post, September 30, 2007. A ZPR resides on top of the Icom R75 communications receiver.

Those interested in finding out more about the modern crystal set should go to the world wide web with a good search engine such as Google. The first place to go is the Birmingham, Alabama Crystal Radio Group. They have a collection of links listed on their web page that will get you to many of the other really good sites. Birmingham really is the center for much of this activity, probably the result of some local activity that grew into something greater. These guys are to be commended for work that gets a lot of us thinking and trying something different. Be sure to look at their photo collection and at the results from their contests.

I'll not spend too much more space with further discussion of crystal sets here. This note is related to some of the measurement methods that I have used for Q measurements. But there is one piece of data that I do want to emphasize, for this was the thing that got my attention: The gang from Birmingham hold an annual listening contest that goes on for

The voltage response of the series RLC circuit can be completely characterized by its resonant frequency, f_R, and Q or the bandwidth, BW.

Parallel RLC Circuits

A parallel RLC circuit, also called a tank circuit, is one in which R, L, and C are in parallel, as shown below.

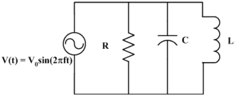

Much of what we just learned about series RLC circuits will carry over into parallel RLC circuits. Before we begin we need to define a new concept known as admittance. Admittance is the reciprocal of impedance and its symbol is Y:

$$Z = \frac{1}{Y}$$

The unit of admittance is the Siemens (S). 1 Siemens = 1 ohm^{-1}. Like impedance, admittance is a complex number, consisting of a real part and an imaginary part. The real part is called conductance, G, and the imaginary part is called susceptance, B:

$Y = G+jB$

Admittance is very handy when analyzing parallel circuits because the total admittance of a group of components in parallel is simply the sum of their individual admittances. Our formulas for inductive and capacitive reactance can be converted to formulas for susceptance:

Another way to look at this is to say that the Q is the ratio of the reactance of the inductor to the resistance, at resonance. High Q RLC circuits have a very sharp response. Low Q RLC circuits have a broad response.

We have talked about the bandwidth of a circuit without defining exactly what it is. Now we will define it.

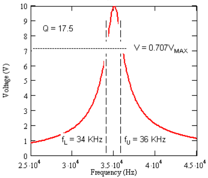

Referring to the figure above, bandwidth (BW) is defined as the difference in frequency between the upper frequency at which the voltage drops to 70.7% of the maximum and the lower frequency at which the voltage drops to 70.7% of the maximum. The 70.7% voltage points were chosen because they correspond to a power decrease of 50%. Thus:

$BW = f_U - f_L$

Bandwidth and Q are also related as follows:

$$BW = \frac{f_R}{Q}$$

In the graph above, Q = 17.5, and f_R = 35 KHz. The bandwidth is:

$$BW = \frac{35000}{17.5} = 2000\,Hz = 2\,KHz$$

over a week in the dead of winter. They then collect the logs to see who is hearing the good stuff. The folks who win are not just listening to the local stations in their area. Rather, they log stations from hundreds, and even thousands of miles away! A winning log may well contain over a hundred stations. One recent winner was from Hawaii.

Q-Measurement using a Q-Meter

The basic ZPR has at least one tuned circuit that serves to tune the receiver to a desired station. Most of the receivers that are built by those who listen to far distant broadcasts will have several. While there is considerable lore on the web that present an algorithm for coil construction, this does not provide the numbers we need for design. If we are to really be able to analyze the circuits that we build, even in a circuit as primitive as a crystal receiver, we must do measurements on key components to characterize, and hence, model them. We must have resonator Q.

Shown below is the basic topology of a Q-Meter, the HP-4342A.

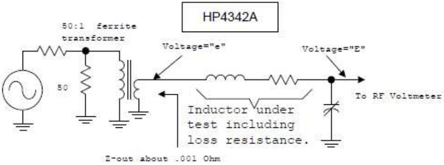

A relatively modern Q Meter. This particular model is perhaps the last of its kind.

This instrument has a built in signal generator (22 kHz up to 70 MHz) that supplies energy to a ferrite transformer with an extremely low output impedance, around one milliohm. The output has a magnitude "e". This level is maintained with a detector and feedback loop. The signal is then applied to the inductor under test, which is attached to terminals on top of the instrument. The inductor is tuned to resonance with a high quality, calibrated, built-in variable capacitor. The voltage across the capacitor is measured with a very high impedance RF voltmeter. The value of this voltage, E, is then directly related to Q. The lower the net series resistance, the higher E will be.

Some amateur experimenters have built homebrew Q meters using the HP scheme. Calibration may be a challenge, but the basic concepts are fundamental and would form a sound basis for experimental efforts. I might be tempted to try some of the very wideband op-amps that are available today as a way to generate the ultra low output impedance drive.

A modern measurement performed in industry or at an institution of higher learning will probably be done with a network analyzer. We will not go into any detail here, for that will take us far from our immediate goals. However, we should state that it is now quite possible for the amateur experimenter to build his or her own network analyzer. Some of the "antenna analyzers" now on the market edge in this direction. A network analyzer was described a few years ago in QEX. In my opinion though, the finest example of a homebrew analyzer is that presented by Paul Kiciak, N2PK. Paul’s vector (meaning that it gives impedance magnitude and angle, or real + imaginary) is found on the web at http://n2pk.com/ and offers measurements to 60 MHz with a 100 dB dynamic range.

Although there are many combinations of L and C that will give the same resonant frequency, the shape of the curve representing VR depends on the L/C ratio as shown below.

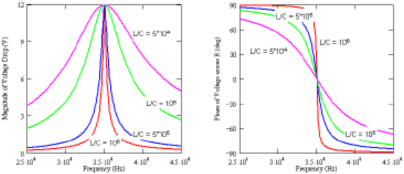

As the L/C ratio increases, the response becomes sharper.
The value of series resistance, R, also affects the shape of the V_R curve as shown below:

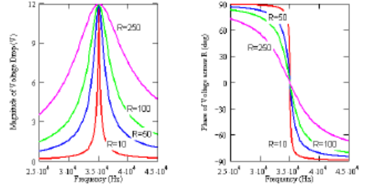

As the series resistance gets larger, the voltage curve gets broader.

In order to make sense of all these possibilities, it is necessary to introduce some new parameters. This first is Q, the quality factor of the series RLC circuit. Q is defined as:

$$Q = \frac{2\pi f_R L}{R}$$

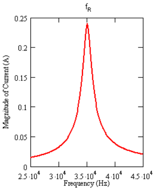

The magnitude of the current is maximum at the resonant frequency and $|I_{max}| = V_0/R = 12/50 = 240$ mA. This type of circuit behavior is called a band-pass response. If a complex signal containing many different frequencies was applied to the series RLC circuit and the output taken across R, the output would contain mostly frequencies near the resonant frequency. The next graph shows the phase of the current through the series resonant circuit.

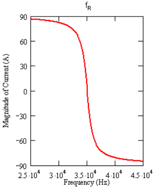

At frequencies below the resonant frequency, the current leads the voltage, which is characteristic of an RC circuit. At frequencies above resonance, the current lags the voltage, and the series RLC circuit looks like a series RL circuit. At resonance, the current and voltage are in phase.

A Still Viable Classic Q Measurement

In earlier times when a network analyzer was not as available to the experimenter as it is today, resonator (another term for tuned circuit) Q was measured by examining the bandwidth of the circuit. The Q is then just the ratio of the center frequency to the bandwidth. A system for this is shown below.

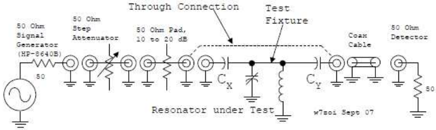

A classic scheme for measurement of Q. CX and CY are adjusted for high loss. See text.

The use of this method is very general and is not restricted in frequency. I've used it to evaluate low frequency inductors used in audio filters as well as components for microwave filters. Like most RF measurements, a 50 Ohm system is used for both a driving RF source and for the load. It is vital to have digital frequency readout. A counter can be attached to an analog calibrated generator. The generator is followed by a step attenuator. This can be homebrew and is trivial for measurements up through the HF spectrum. All we will need for this measurement is a 3 dB step. However, it should be fairly accurate. (See EMRFD chapter 7 for info on attenuators.) The step attenuator is followed by a fixed pad to establish a Z0 environment. This pad should ideally be right next to the following test fixture that contains the tuned circuit we are measuring. The test fixture output is routed to a 50

Ohm detector. This can be a power meter, a spectrum analyzer (my usual choice), or a 50 Ohm terminated oscilloscope. If the ‘scope option is picked, the termination should be at the input to the oscilloscope and not out at the test fixture end of the connecting cable. With a 50 Ohm terminator at the ‘scope, the test fixture will see 50 Ohms at all frequencies, no matter what the cable length might be.

The first step in use is to put a jumper (usually a BNC barrel) from the fixed pad to the cable leading to the detector, shown as a dotted line in the figure above. The level in the detector is then noted. If you are using a ‘scope, calculate the power in dBm. Power in dBm is usually read directly if you are using a power meter or spectrum analyzer. Then connect the test fixture with its tuned circuit and tune the signal generator and/or variable capacitor to obtain a detector peak. The peak response will always be less than was seen with the "through" connection. That direct connection represented a matched impedance case where all of the available power from the signal generator is transferred to the detector.

It’s now time to extract some numbers. Set the step attenuator to 3 dB and then tune the generator for a peak in output response. Peak the resonator capacitor if necessary. Note the frequency for this peak response. Now comes an important part of the procedure that is easy to bypass. Note the power seen in the detector and record it in your notebook. It should be at least 30 dB below the maximum available power that we measured with the through connection. Let’s assume, for the moment, that it is.

We now carefully note the detector level. If we are using a spectrum analyzer or an oscilloscope, we adjust the signal generator amplitude just a bit so the trace is right on a cursor

---

$$Z = R + j2\pi fL\left(1-\left(\frac{f_r}{f}\right)^2\right) = 50 + j6.2832f*5.3*10^{-3}\left(1-\left(\frac{3.5*10^4}{f}\right)^2\right)$$

$$= 50 + j\left(0.0330f - \frac{40.79*10^6}{f}\right)$$

The graph below shows how Z varies with frequency:

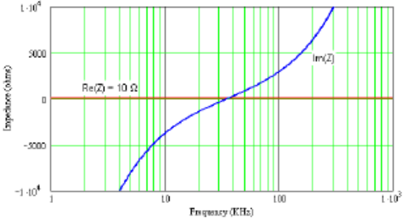

We can find the current through the series RLC circuit by using Ohm's Law:

$$I = \frac{V}{Z} = \frac{12}{50 + j\left(0.0330f - \frac{40.79*10^6}{f}\right)}$$

rather than simplify this ugly expression further, we will graph the current as a function of frequency:

called the resonant frequency. It is possible to derive an equation for the resonant frequency in terms of L and C from the condition $4p_2f_2LC = 1$:

$$f_R = \frac{1}{2\pi\sqrt{LC}}$$

where:
f_R = resonant frequency (Hz)
L = inductance (H)
C = capacitance (F)

The resonant frequency depends on the product of L and C, not the individual values, and is independent of R. We can use the formula for the resonant frequency to simplify our equation for the impedance of the series RLC circuit as follows:

$$Z = R + j2\pi fL\left(1 - \left(\frac{f_R}{f}\right)^2\right)$$

Let us examine the behavior of a series RLC circuit in more detail by studying the circuit shown below:

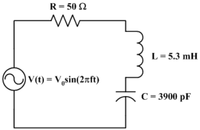

.

We can immediately calculate the resonant frequency:

$$f_R = \frac{1}{2\pi\sqrt{LC}} = \frac{0.159}{\sqrt{(5.3*10^{-3})(3900*10^{-12})}} = \frac{0.159}{\sqrt{2.067*10^{-11}}} = \frac{0.159}{4.546*10^{-6}} = 35.0\,\mathrm{KHz}$$

We can substitute in values for fR, R, and L to get the total impedance:

line on the display. An adjustment of oscilloscope vertical position may also be used for this.

Next, remove the previously added 3 dB attenuation. This will cause the response to increase, doubling the detector power. (An oscilloscope voltage response will go up by a factor of 1.41.) Now carefully tune the signal generator to a higher frequency until the response has dropped to produce exactly the same level that we had at the peak. Record this upper frequency in your notebook. Next, the generator is tuned back toward and through the peak until the previously noted amplitude is again obtained. This occurs at a lower bandwidth edge. The difference between the upper and the lower frequencies is the resonator bandwidth. The Q is then the ratio.

$$\mathrm{BW} = \mathbf{F}_{\mathbf{upper}} - \mathbf{F}_{\mathbf{lower}}$$

$$\mathrm{Q} = \frac{\mathbf{F}_{\mathbf{center}}}{\mathbf{BW}}$$

Consider now the earlier assumption that there was at least 30 dB loss through the test fixture containing the resonator and connecting capacitors. This loss relates to the choice of CX and CY shown in the schematic diagram for the classic measurement. First, it is important that the values be approximately the same value. This guarantees that any loading by the Z0 source will equal the loading by the Z0 detector. If CY was much larger that CX, for example, we could have a situation where the detector would severely load the resonator, but we could still have 30 dB or more attenuation. So make the two loads about the same. A loss of 30 dB or more means that the dominant loss mechanism is the loss within the tuned circuit and not the loss related to loading by the source or detector.

The way we adjust the loading at the ends of this single resonator filter is by picking CX and CY. The values we use will depend upon the capacitor in the resonator. If, for example, we have a resonator capacitance of 300 pF, we can probably get our 30 dB loss with values of perhaps 3 pF for CX and CY. I would probably grab 1 pF for each part from the junk box. Exact details will change as we go to much higher frequency. The concepts are the same though. For example, when evaluating a VHF helical resonator, or similar LC-like structure, we might have nothing more than a pair of coaxial connectors mounted in the resonator wall. The normal center pins from the connectors may well be enough coupling. Indeed, I have encountered situations where it was necessary to recess the connectors in the wall so that the center wires are partially hidden. Good measurement results are guaranteed if the 30 dB rule is maintained and symmetrical loading is achieved. If the Q exceeds 500 or 1000, it may be useful to go to loss greater than 30 dB. Alternatively, a correction can be applied. See page 58 of Introduction to RF Design (ARRL, 1994.)

Some of these details are not intuitive. A good way to build some intuition is to do simulations in LT Spice, which is free from Linear Technology on the web. (Bravo for LT—many thanks!) You can then "build" inductors in software, with appropriate series resistance. Then sweep the filter and see if you get the right Q by observing the 3 dB points.

The procedure outlined is general and applies to any frequency. The use of low impedances allows extreme flexibility. Some folks have measured Q with a similar approach where a small probe coil is driven by the generator and is placed near the resonator being studied. An oscilloscope

This expression in more complex than those derived in earlier sections for the series RL and RC circuits and as a result, the behavior of a series RLC circuit is much more interesting. Let us consider several interesting cases:

Case 1: f <<1.

If the frequency is very close to 0 Hz, the j2pfL term in the original expression for Z is approximately 0. In this case the impedance of the parallel RLC circuit can be approximated by the following equation

$$Z \approx R - \frac{j}{2\pi fC}$$

At very low frequencies, the series RLC circuit behaves like a series RC circuit.

Case 2: f >> 1

When the frequency is very large, the -j/(2pfC) term in the original expression for Z is approximately 0. In this case the impedance of the series RLC circuit can be approximated by the following equation

$$Z \approx R + j2\pi fL$$

At very high frequencies, the series RLC circuit behaves like a series RL circuit.

Case 3: $4p_{2}f_{2}LC = 1$

When this condition occurs, the imaginary part of Z is zero. The impedance of the series RLC circuit is real and is equal to the resistance:

Z = R

The circuit behaves as if the inductor and capacitor were not present. The negative reactance of the capacitor and positive reactance of the inductor add up to 0, creating a condition is known as resonance, which occurs at a specific frequency

RLC Circuits in: Electronics for Radio Amateurs

York County ARS K4YTZ
http://www.ycars.org/EFRA/Module%20A/AC%20Circuit%20Analysis%205.htm

RLC Circuits

An RLC circuit contains resistors, inductors and capacitors. In this section we shall look at the two simplest RLC circuits, the series RLC circuit and the parallel RLC circuit.

Series RLC Circuits

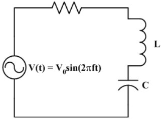

Since the resistor inductor and capacitor are in series, the total impedance of the series RLC circuit is the sum of the individual impedances:

$Z = R + X_L + X_C$

After substitution of the formulas for X_L and X_C one gets the following expression for Z in terms of f, R, L and C:

$$Z = R + j2\pi fL - \left(\frac{-j}{2\pi fC}\right)$$

This can be simplified by combining the imaginary terms:

$$Z = R + j2\pi fL\left(1 - \frac{1}{4\pi^2 f^2 LC}\right)$$

---

is then used with a 10X probe and a small value series capacitor as a detector. The series capacitor is important, for even with a 10X probe, the impedance at RF might still excessively load the resonator. It is important to keep the source probe inductor well away from the coil. The concepts are the same and can all be modeled with appropriate calculations.

The EMRFD Q-Measurement Scheme

An alternative scheme is presented in Experimental Methods in RF Design (ARRL, 2003) in Chapter 7, page 7.36. In this method, a tuned circuit containing the inductor to be measured is configured as a series resonator and is then connected as a shunt element. The basic scheme is shown below.

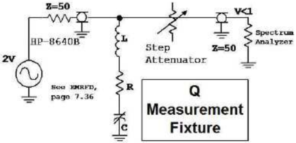

Alternative Q measurement scheme.

A steady strong output is seen in the detector as the signal generator is tuned. As resonance is reached, the output dips down. The reactance of the inductor is cancelled by that of the capacitor exactly at resonance, leaving nothing but the loss resistance, R, to attenuate the signal reaching the detector. The higher the Q, the greater the dip will become. A more detailed diagram is shown below.

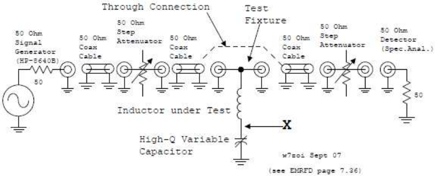
Through Connection
Test Fixture
50 Ohm Signal Generator (HP-8640B)
50
50 Ohm Coax Cable
50 Ohm Step Attenuator
50 Ohm Coax Cable
50 Ohm Coax Cable
50 Ohm Step Attenuator
50 Ohm Detector (Spec.Anal.)
50
Inductor under Test
X
High-Q Variable Capacitor
w7zoi Sept 07
(see EMRFD page 7.36)

Alternative Q measurement system. See text for explanation.

The procedure for using this method begins with setting the signal generator to the desired frequency. The level is noted in the spectrum analyzer. A power meter or 50 Ohm terminated oscilloscope can also be used if the signal generator is known to be low in harmonic content. (Not all signal generators fulfill this criterion though, so be careful. The HP-8640B that I use has very low harmonic output. If you are concerned, use a low pass filter with the generator.) The capacitor in the test fixture is then tuned to produce a dip. Use the highest possible amplitude resolution you have in your detector. I go to a 2 dB/division mode with my spectrum analyzer. After an initial dip has been set with the capacitor, use the signal generator to select the lowest dip response. Then use an amplitude control on either the detector or generator to move the observed signal to a cursor line.

Record the signal generator frequency. Then take the test fixture out of the system, realized in my setup by substituting a BNC barrel for the test fixture. The signal will increase at the detector. Increase the attenuation in the step attenuator until the same response at the cursor line is observed. This will provide an attenuation or dip value. Interpolate as needed to

---

EE. Bucher, 1919, "Wireless Experimenter’s Manual", Wireless Press, New York.

Wes Hayward, W7ZOI, Experiments with Coils and Q-Measurement, October, 2007 (Updates 01Dec07, 08Dec08. http://web.archive.org/web/20101226143919/http://w7zoi.net/coilq.pdf

Steve Ratzlaff, AA7U kindly made detail measurements on the Q of the capacitors and coils used in this report.

Dick Kleijer , 2013, Measuring the Q of LC Circuits,
http://www.crystal-radio.eu/enqmeting.htm

Barbara Dziurdzia, 2013, Damped oscillations in RLC circuits, AGH University of Science and Technology in Cracow,
http://home.agh.edu.pl/~lyson/downloads/Manual_8.pdf

Rice University EE., RLC Circuits,
http://www.owlnet.rice.edu/~phys102/Lab/RLC_circuits.pdf

Durham University EE., Oscillations and Resonances in LRC Circuits,
http://level1.physics.dur.ac.uk/projects/script/lcr.pdf

This is slightly low to Steve's measured 240Q. I am of the opinion that the original tank measurement may have been a bit on the low side. Still, again the results are within the ballpark and the technique works if you have a good measure on the Q of the capacitor used in your tank. Alternatively, starting with a known coil Q, the technique will work equally well to determine the Q of an unknown capacitor.

Conclusion:

If you have access to, or own a digital oscilloscope, or have been hankering to purchase one, then the time may be now. The analytical method presented above is simple in practice and robust in theory. I would also say that, by working with your wave forms, you learn to recognize your data and soon get a good "feel" for what is good and what is not. The technique also produces, should you wish, an auditable report of the wave and analysis. It is hard to argue against the results with the waveform sitting there staring right at you!

Measuring your coil Q should not be difficult.

A copy of the evaluation spreadsheet in xls format can be found here CoilQbyDecrement_kjs.xls

.

Bibliography:

Jacques Audet, VE2AZX., Q Factor Measurements on L-C Circuits., QEX - January/February 2012, pp 7 - 11.
http://www.arrl.org/files/file/QEX_Next_Issue/Jan-Feb_2012/QEX_1_12_Audet.pdf

---

obtain data within 0.1 or 0.2 dB. Record this value in the notebook with the previous center frequency.

Finally, unsolder the inductor from the circuit at the point marked with the large X in the figure. Then attach a capacitance measuring meter to measure the C value. I used one of the AADE L/C meters. (Just Google AADE on the web.) The C value is recorded. The F and C data allows calculation of the effective inductance. The inductor equivalent series resistance can then be calculated from the attenuation, which then leads directly to Q. The equations are included in the EMRFD presentation.

An advantage of this scheme is that the Q value is directly related to the attenuation value, which is relatively easy to determine with an accurate step attenuator. (I use a surplus HP-355C and HP-355D combination.) The detector operates directly on a null point rather than along the slope of an attenuation function where frequency drift can complicate the results. But the method is sensitive to the driving and load impedance at the test fixture. Values other than 50 Ohms will compromise results It may even be important to actually use 50 Ohm cables; the 52 Ohms of RG-58 might not be close enough. A conservative approach would place a 50 Ohm pad (6 to 10 dB) at the edge of the test fixture.

This photo shows the test fixture that we used with the EMRFD scheme. The inductor in this case is a ferrite toroid wound with Litz wire. This variable capacitor is believed to be made by Jackson Brothers, Ltd, from England and is the highest Q variable capacitor we had in our junk box. More on this below.

Comparing the methods.

Most of the measurements we have done in recent times used the EMRFD method. In years past, we
would use a Q meter when available, but most often, the 3 dB BW method. Short of a calibrated Q-Meter, the EMRFD scheme is quicker than the 3 dB resonator bandwidth measurement. In one recent case, I measured the Q of an inductor with both methods. The inductor was a basket weave coil, described below. Using the classic 3 dB bandwidth method, the Q was measured as 579 at 1 MHz. Insertion loss was 48 dB for this measurement. The same coil with the same variable capacitor measured Q=650 at 1 MHz with the EMRFD scheme. That’s about a 10% spread between the two schemes.

The coil used in this experiment was a 5 inch diameter solenoid wound in a basket weave pattern for L=265 uH. The wire was 175/46 Litz, which is 175 strands of #46. The coil sat on a piece of plastic suspended on a stack of wood blocks during measurement in an effort to avoid external loads.

Experiments with various coils.

APC cap with red/black leads (149 pF meter): 937Q at 904kc (Q meter alone and coil 1060Q)
APC cap with my Litz leads (149 pF meter): 995Q at 904kc (Q meter alone and coil 1060Q)

Steve's measurements above provide important datapoints with which to calibrate my technique. Note first of all that the Litz coil is not a modest 506Q but a robust 1140Q at 1MHz. The reason for my low measurement was my assumption of a high-Q capacitor. Steve's measurements of two of my APC's are 995 and 937, hardly very high. There is an obvious need for a correction. Wes Hayward in his excellent notes on Q measurements provides the key in the form of the following formula:

$$Ql = (Qc * Qres) / (Qc - Qres) \qquad (14)$$

where Ql is the desired coil Q, Qres is the measured Q of the tank/resonator, and Qc is the tank capacitor Q.

Substituting the values Qres = 532 and Qc = 995 the equation simplifies thus:
$Ql = (937 * 506) / (937 - 506)$
$Ql = 1100$.

The result is distressingly close to Steve's measured value of 1140Q. This shows the possible level of accuracy in the analytical technique used, both reassuringly high and very "Litz-like".

Taking the first solenoid and substituting the values Qres = 167 and Qc = 937 the equation simplifies thus:
$Ql = (937 * 167) / (937 - 167)$
$Ql = 203$.

my opinion that the capacitor Q ought to be high enough as to not influence the coil Q measurement in any significant manner. I said that for three reasons: 1) Just about every published article starts with such an assumption and 2) I had not done much research on the subject of capacitor Q and so had no particular expectations other than assumption 1 and finally 3) I had no measurements or specific knowledge of the Q for the capacitor used in my setup.

Reason 1 is a punt and should be rejected, reason 2 has been remedied and I give some analysis on the interesting subject of capacitor ESR (Equivelant Series Resistance) on a separate page so that readers may set their own expectations. For the third reason I have received the kind assistance of Steve Ratzlaff, AA7U who offered to make actual Q measurements on the coils and APC capacitors with his HP4342A Q-meter, results follow:

Litz basket coil,
550kc 1127Q;
1000kc 1140Q,
1700kc 674Q (194.2 uH on aade.com LC meter)

Solenoid coil,
550kc 205Q,
1000kc 240Q,
1700kc 222Q (218.2 uH on aade.com LC meter)

Using Steve's Litz wound ferrite rod coil (1000kc 1047Q) as the reference for external caps:

100 pF silver mica (99.2 pF meter):859Q at 1068kc (Q meter alone and coil 1025Q)

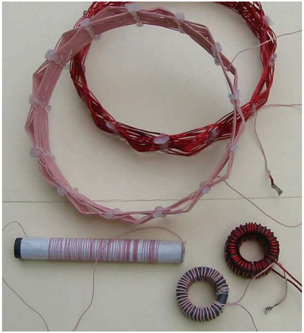

Some of the coils investigated in this study. The basket weave coils were wound with the fixture shown below. The two toroids shown are Amidon FT-114A-61 and used #18 enamel wire or 175/46 Litz wire. The rod is similar to an Amidon R61-050-400 and is wound with 50/46 Litz wire.

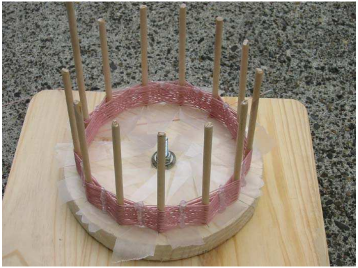

Fixture used to wind basket weave coils. Strips of waxed paper are placed under the coil cross over lines before application of hot melt glue.

Toroid Experiments

Most of the filters I build at HF use toroid cores. Powdered iron cores seem to offer higher Q at HF (3 to 30 MHz) than I can obtain with other practical, reasonable sized forms. The -61 and -63 Ferrite materials from Amidon offer high Q in the MF region from 300 kHz to 3 MHz. Powdered iron can provide high Q, but have such low permeability that the number of turns becomes excessive. So the first experiments I did used ferrite toroids.

Not all of the data collected will be presented here. However, I'll give a few of the data points of interest. There are some surprises.

qc = 18.55 exp(-t / 2*Rp*123)*cos(6.283*1.044*t + 1)

Where t = 0 to 50 uS
R = 0.627 Mohm in this case for best match.

Q = Rp / 2pifL = 627000 / 6.283 * 1.044 * 189 = 506

The results above for a big litz coil clearly demonstrate its superiority to a more common tapped solenoid. Published data though has let me to expect a big litz coil Q to be in the 1000+ range. I have run this test backwards and forwards, checking all possible causes for this unexpected low Q with no difference in the final value. In fact I am quite impressed with the repeatability of the technique and the correspondence between the results from the decrement and those from modeling the wave form.

To be certain I even tested the coil using the old -3dB method. As stated previously, I am uncomfortable using a signal generator in addition to the scope, and I have already found trouble connecting the 1:100 scope probe directly to the tank, as is necessary in this method. With that said, the results yield a bandwidth of 2.3kHz about a nominal frequency of 1018.8kHz for a tank Q = 450. I suspect the method (with the problems noted) lowballs somewhat the final result. In any case, this confirms the test to have a moderate Q and nothing like the expected values one reads online.

Notes on Capacitor Q and its impact on the measurement value.

This is where we begin to think more seriously about the capacitor used to resonate the tank. The assumption of very high Q needs to be questioned. I stated in the setup discussion

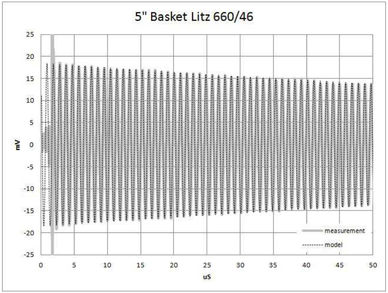

Figure 6. Second oscillogram with modeled damped wave response.

The coil tested values as follows:

$n = 45$
$T = (t_n - t_o) / n = (47.82 - 4.70)/45 = 0.958$ uS
$f = 1/T = 1/0.958 = 1.044$ MHz

$A_o =$ $= 18.05$ mV
$A_n =$ $= 13.65$

$d = 1/n \ln(A_o/A_n) = 1/45 \exp(18.05/13.65) = 0.00621$
$Q = \pi / d = 3.142 / 0.00621 = 506$

And:

$q_c = q_o * e^{(-t/2R_pC)} * \cos(2\pi f * t + \phi)$
$C = 123$ pF

First, the best toroid I found for BC band use was 38 turns of #18 enamel wire nearly filling a FT-114A-61 core. Inductance was 195 uH on an AADE L/C meter. Q=383 at 1 MHz and peaked up to almost 400 at 811 kHz. Q remained well over 300 at both 500 and 1610 kHz.

A slightly more practical and easier to wind coil used a FT-114A-61 form with 41 turns of #22 enamel wire. At 1 MHz, the inductance was 240 uH with a Q of 384. The Q dropped nearly to 300 at the band extremes.

The receiver I was building was to use a variable capacitor with only 188 pF maximum C. So a higher L was needed. I managed to get 56 turns of #22 enamel wire on a FT-114A-61 form for L(AADE) of 449 uH and Q=344 at 1 MHz. Q was 316 at 549 kHz and 274 at 1.55 MHz. Although compromised, this seemed a practical coil for a first zero power receiver.

A question that immediately came up was what could be done with Litz wire on toroids. The results were disappointing. 39 turns of 175/46 Litz wire on a FT-114A-61 form yielded L=244 uH and Q=318 at 1 MHz. The Q improved to 352 down at 550 kHz. A second FT-114A-61 core was wound with 40 turns of smaller 50/46 Litz wire, resulting in L=248 uH and Q=298 at 1 MHz. Again, low frequency Q improved slightly. My conclusion at this point is that the 175/46 Litz wire is not justified for ferrite toroids, although that conclusion is certainly more general than my data can support. I have not investigated any of the expensive Litz "rope" using over 600 strands of #46 wire. (See later note.)

Variable Capacitor Q (data updated 08Dec08 with more discussion at the end.)

Many of the web sites I read that deal with crystal sets and similar receivers state that the absolute best possible variable capacitors should be used. They then define these as being those capacitors with ceramic insulators, with silver plating being preferred on the metal. Double bearing designs are also recommended. Some sort of special means to minimize resistance related to moving contacts is deemed important. While all of these things are intuitively reasonable, they are not numbers.

The Q measurement schemes I have presented all relate to a determination of resonator Q. Actual inductor Q can only be obtained if one knows the Q of the capacitors in a tuned circuit, or if can be established that capacitor Q is very high. The folks building Q meters go to great extremes to optimize the variable capacitor Q within the instrument. (See Kito and Hasegawa, "Measuring Q" Easier and Faster,(HP Journal, September, 1970.)

Let me again emphasize that all of the numbers presented here, with all measurement methods, relate to resonator Q. We can calculate capacitor Q only if we independently know the inductor Q.

This still leaves us with the question of what we should be using for variable capacitors, both in our receivers and for Q measurement. To begin an evaluation, I took the highest Q toroid that I had built at the time (38 turns of #18 enameled wire on a FT-114A-61 core) and attached it to various variable and fixed capacitors from my junk box. Measurements were done at or near 1 MHz using the EMRFD measurement scheme. The data, although less than conclusive, is interesting. Here are some results:

decrement method. An exact match will not always be the case, but they ought to be fairly close. The match between theory and the decrement method gives confidence that this analytical technique is as robust as it is simple.

qc = qo * e^(-t/2RpC) * cos(2pif * t + phi)
qc = 12.2 exp(-t/2*Rp*106)*cos(6.283*1.055*t + 1)

Where t = 0 to 50 uS
R = 0.237 Mohm in this case for best match.

Q = Rp / 2pifL = 237000 / 6.283 * 1.055 * 215 = 167

Example Two:

To be sure the technique works I have also assessed the quality of a 5inch diameter basket weave coil wound from 660/46 litz wire. Such coils are rightly considered "performance coils" for their low skin-effect and low resistivity.

This leaves the resistance of the coil (R) as the primary variable. Model the above equation in the spreadsheet by adjusting R until the model results match the measured oscillation from the coil. From the above example we see Figure 5 with the modeled damped wave overlaid on the test oscillation.

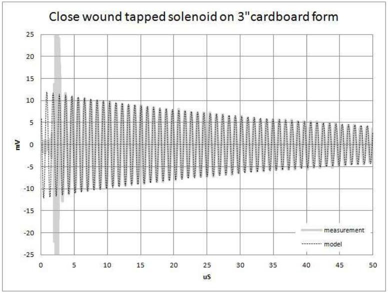

Figure 5. Oscillogram with modeled damped wave response.

Here the background data from the test is in grey and overlaying it is the dashed model in black. The fit is nearly perfect using a resistance of 0.22 Mohms. Relating resistivity to Q we have the following equation (5):

$Q = Rp / 2pifL$

Substituting f = 1.055 MHz, C = 106 pF, and R = 0.237 Mohms the result is Q = 167, exactly the same as with the

Capacitor comparisons, all operated near 100 pF.

1. 107 pF Silver Mica. This consisted of two series connected 214 pF 1% 500 volt units.
1. Resonator Q=341.
2. Mica compression trimmer, 30-300 pF, marked GMA40400, Resonator Q=314.
3. "Jackson Brothers" (that’s a guess) dual section, 497 pF/section max, built-in gear drive, Resonator Q=383. This was the best variable capacitor in my junk box.
4. TRW four section. This was the most elegant looking capacitor I had in my stash. It came from a Tektronix 191 sine wave generator. Resonator Q=354. An OK cap, but not the best.
5. Collins TCS Transmitter VFO capacitor, built-in gear drive. Resonator Q=357.
6. ARC-5 Receiver variable. (BC-454?, all three sections in parallel yield 47-514 pF) Resonator Q=378.
7. BC-221 capacitor, 13 to 188 pF, Resonator Q=364.
8. Hammarlund single bearing, 150 pF max. Resonator Q=279. (ugh).

Much of the lore seems correct. However, the best measured resonator Q came from the Jackson Brothers dual section, not from antiquity, but from just slightly more recent times. The beautiful TRW capacitor, while good, was not at the top of the list. The aging single bearing capacitor was not expected to be great, and there was no surprise. I was surprised that the Silver Mica fixed capacitors were not at the top. But perhaps these were rejects. Although they were brand new unused capacitors, they were purchased at a surplus emporium and loss may have been the reason they were declared surplus. Additional measurements are called for in this area.

Additional information regarding capacitor measurements is appended at the end of this note.

Litz Wire Connection.

Classic lore suggests that it is vital to solder all strands in a Litz wire bundle at both ends. If one strand is left unsoldered, the argument is that current can no longer flow in that wire, so it will not contribute to the expanded surface area that leads to improved Q with Litz wire.

I had measured less than stellar Q with some 175/46 Litz wire on a toroid form and wondered if perhaps I did not manage to get all of the strands soldered. So I did an experiment while using the mica compression trimmer as the capacitor (poor capacitor choice.) After doing my best to solder all strands by carefully applying rosin flux from a SMT rosin pen, I heated and tinned the ends with an 800 degree tip in my Weller soldering station. The ends looked very good in a visual inspection.

I then started doing Q measurements near 1 MHz. After each one, a few strands of the Litz wire were cut and a 1 inch section was removed near one end. The strands were removed from the capacitor end of the test resonator. Here’s a plot of Q from 0 to 152 clipped strands. N was estimated by counting the number of pieces of wire after a clipping operation.

To test the validity of this technique we need a bit more theory, just to be certain this is working. With the data in spreadsheet form, it becomes possible to model the coil from a theoretical standpoint. The model can then be plotted with and compared to the acquired data to test for correlation. Various web pages, especially university-sponsored lab exercises provide a good source for practical techniques and their theoretical underpinnings. Some good sources I utilized include:
"Damped oscillations in RLC circuits" by Barbara Dziurdzia at AGH University of Science and Technology in Cracow, "RLC Circuits" a lab note from Rice University, and "Oscillations and Resonances in LRC Circuits" from Durham University in the UK. And very many others.

From the above we find the general equation for damped electrical oscillations as:

qc = qo * e^(-a*t) * cos(2pif * t + phi) (13)

Where
qc is the charge in volts
qo is the initial charge
a is the damping factor (=Rs/2L (series) and = 1/2RpC (parallel)
R is the coil resistance in ohms
L is the coil inductance in uH
t is the time in uS
f is the frequency in MHz
and phi is the phase angle

The phase angle and qo charge are initial conditions and adjustments should be minor to fit the calculated oscillation to actual. L, the inductance comes from the calculation we made with equation (12), and frequency comes from equation (9).

An example to show off a bit. This test was made with a close-wound tapped solenoid on a 3" cardboard form. The graphic in Figure 4 shows some initial contact hash followed by a lovely damped oscillation. Measurements on this coil showed a test frequency of 1.055 MHz, L of 215 uH and C = 106 pF. Q was calculated at 167. A decent coil but not great.

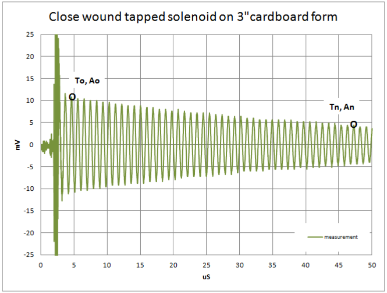

Figure 4. Oscillogram of a coil under test.

$n = 45$
$T = (t_n - t_o) / n = (47.24 - 4.59)/45 = 0.948$ uS
$f = 1/T = 1/0.948 = 1.055$ MHz

$A_o =$ $= 11.2$ mV
$A_n =$ $= 4.80$

$d = 1/n \ln(A_o/A_n) = 1/45\exp(11.2/4.80) = 0.01883$
$Q = \pi / d = 3.142 / 0.01883 = 167$

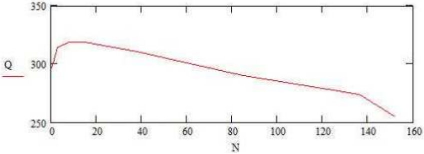

Q versus number of removed strands in a Litz wire toroid. There are 175 total strands in this wire.

This is certainly not a definitive experiment. It should be repeated with a coil where the Litz wire seemed to actually make a difference, and that was not the case with the toroid. However, the classic lore is not supported. But this makes sense after some thought (and some comments from colleague KK7B.) The weak dependence suggests that even though a cut at the end of a wire means that conduction is eliminated at the end of the wire, this does not mean that it is removed from the interior of the coil. Strand to strand capacitance guarantees that similar current will flow in each. The current in each forces the flow to the surface. If a strand is cut, there will still be some current flow in that strand back from the cut. The capacitance between strands connects the strands to those that are still connected at the end.

I hope to repeat this experiment with a different coil and a better resonator capacitor.

The Basket Case

Solenoid coils tend to have the highest Q if the wires are spaced by about 1 wire diameter. This lore is found in Doug DeMaw's book on Ferromagnetic Inductors (Prentice Hall,

1981) and is quoted on the Amidon web page. This rule of thumb is reasonable. The highest inductance will occur with tight spacing. If the spacing increases beyond this, there is a greater chance that some of the magnetic flux from one wire in a coil will escape from the side of a coil and not link to all turns.

If we plot the inductance of a solenoid of constant pitch versus the number of turns, L is proportional to N2 when N is small. However, the parabolic shape relaxes toward a linear dependence as N becomes large, especially as the length exceeds the diameter. Linear L versus N is akin to adding inductors in series where each one is isolated from the others.

When one turn is directly next to the next turn with no gap, the current flowing in one turn will force some of the current in the next wire to move away. The overall effect produces non-uniform current in the wire surfaces. This is alleviated with a slight spacing between turns. But too much spacing decreases inductance too much.

A classic coil type is the so called basket weave. A couple of coils of this type are shown, along with a coil winding fixture on page 7 of this report. There are an odd number of rods in the winding jig. This means that as we weave the wires among the rods, one turn will not be directly adjacent to the next turn. This makes the coil behave as if it was a solenoid spaced by 1 turn. But there is no real gap, so the coil length, and hence, inductance is the same as a tightly wound one would be.

I had never built a basket weave coil, yet this is one of many designs that the folks describe on the web sites mentioned. So a couple of coils were built. The first used #22 enamel wire. The coil diameter is 5 inches wound on 13 posts with a total of

5000 data points to the computer. At 1 MHz that is about 100 data points per cycle, more than enough sampling to avoid any question of aliasing.

The decrement measurement is based on measuring two peaks. The data is in digital format so one should eliminate "estimation error" by actually reading the value off the spreadsheet, noting both time (in uS) and amplitude. Before taking readings, care must be taken to center the oscillations around zero. Amplitude measurements assume that the waveform is symmetrical around zero yet the scope may or, more likely, may not be set perfectly for this. In the spreadsheet the solution is simply to take the average of all the readings and subtract that from each.

Type the values for “to” (initial peak time), “tn” (final peak time), Ao (initial peak amplitude) and An (final peak amplifude), and the number of periods analyzed “n”.

T = tn - to in uS (8)
f = 1/T in MHz (9)

d = 1/n ln(Ao/An) (10)
Q = pi / d (7)

Next, disconnect the battery and coil and measure carefully the capacitance, C, that was used in the test. With this you can calculate the actual coil inductance from the frequency and capacitance. Why buy an expensive L meter?

L (uH) = 1E6 * ((1/ 2pif)^2) / C (pF) (12)

Example One:

10uS / division for a longer recording. Set the scope trigger appropriately to capture the data.

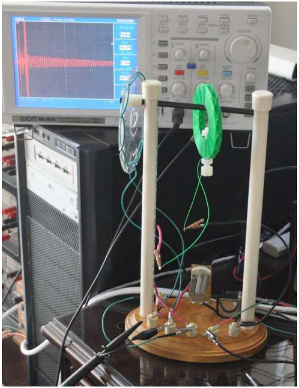

Figure 3. Photo of the test setup.

Tap, or "key" the battery to pulse the tank and review the waveform on the screen. Setting the scope on its measurement mode allows reading the wave frequency with each pulse. Adjust the capacitance after each pulse to bring the tank to the desired frequency, near 1 MHz in my case. Once at the desired frequency you are collecting data. When you get a good clean oscillation free of contact hash and noise, send the data to the computer for analysis. For my scope set at 5uS/div it returns

---

36 turns. The 1 MHz Q was a poor 244 with Q slowly increasing with frequency. Inductance was 265 uH A similar coil was then wound with 175/46 Litz wire. Again, inductance was 265 uH. The 1 MHz Q was a spectacular 650. The Q dropped off at both 530 and 1630 kHz, but was OK over most of the range. This is the highest 1 MHz Q I had ever seen at that time. I remember frequent Q values well over 1000 with some helical resonators at 500 MHz, but that is common at UHF.

Ferrite Rod Experiments

The next resonator type to be investigated was the ferrite rod. The only new part that I had around was an Amidon R33-050-750, meaning that it was built from mix 33 type ferrite with a diameter of 0.5 inch and length 7.5 inch. This material has an initial permeability of 800, which is high enough to make me wonder about it as a high Q antenna material. Owing to the high permeability, the number of turns needed should be much less than the 70 turns suggested in the N7FKI paper. I put a layer of paper on the core and then would 36 turns of #22 enamel wire over the central 3 inches. The inductance was 158 uH. Additional turns would be needed to reach the desired 240 uH.

Then Q was measured. It was a miserable 58.6 at 1 MHz. On the other hand, I did see several local radio stations in the spectrum analyzer display during testing. It may be a viable antenna, but it is not much as a resonator. I didn’t even bother changing to Litz wire.

The next experiment was with a junk box ferrite rod. This was a piece that was 3.85 inches long with a diameter of 0.475 inch. After putting a layer of paper on the core, I wound 37 turns of #22 enamel over the inner 2.8 inches. The inductance was 80 uH. A 1 MHz Q measurement was a more respectable 293.

The dimensions of this rod were very close to an Amidon R61-050-400. The inductance constant in the Amidon literature, after conversion, was 43 nH/turn2, so the 37 turn winding should have an inductance of 59 uH if it was wound evenly over the total core. My winding was bunched near the center which would increase the inductance a bit, so 80 uH would be reasonable. I concluded that the material is probably -61 mix, or a similar mix with initial permeability of 125, perhaps from a different vendor. I calculated that 65 turns over the same part of the core would yield 250 uH.

Based upon the measurements above, I wound 62 turns of 50/46 Litz wire on a single layer of paper over the junk box rod. The result was a spectacular Q of 610 at 1 MHz. The Q dropped off to 577 at 550 kHz and 500 at 1551 kHz, both still quite respectable. This high Q was a major surprise.

N7FKI has a larger R61-050-750 ferrite rod in his parts box and has offered it for some measurements. That is on the list of "To Do" experiments.

Classic Solenoid Coil

Among the goodies in the N7FKI junk box were some large cylinders that looked like they would work well as coil forms and yield reasonable Q. The outside diameter was 4.2 inches. After earlier coil construction experiments, about 55 feet of Litz wire remained on a spool (purchased on EBay from Paul Cianciolo, W1VLF), so I put solder lugs on each end and wound a coil. The first one had the wires spaced approximately

pF. These are silver or nickel plated (depending on the source of information) and ceramic (steatite) insulated and should have a high Q with little impact on the test. At least this is the general assumption used in most Q-measurement articles I have run across. In actual fact the capacitor Q is not infinite and not even that high so resistive losses associated with the capacitor can be expected to lower the measured Q. One needs to keep this in mind and know that this measurement is on the resonator, not on the coil itself. More on this topic in a later section. The APC capacitance range is limited but sufficient for my intended test frequencies of 1.0 to 1.1 MHz, this is a test, not a radio.

Keying, or pulsing the circuit is somewhat tricky. In order to obtain a clean wave trace the pulse contact needs to be fast and sharp. Contact "bounce" and arcing, especially at radio frequency are difficult to avoid and pulsing the circuit several times may be needed prior to finding a clean trace. I initially used a transmitter key but found the contactor at RF to be messy with visible arcing. The best method I have found for pulsing the circuit is to connect a pointed probe to one side of the circuit and tap it sharply but lightly against the hot side of the battery. After a few tries one gets a "feel" for the contact. Pulse the circuit and get a wave, things are this simple.

Procedure:

Hang the coil under test on a stand where it is clear of other components or metal objects. Connect the coil as per the schematic. Connect the probe and scope to the coupling coil leads and set the scope for a good scale, (Figure 3). I generally use 5uS / division horizontal scale which gives me 40 to 50 oscillations at about 1 MHz. For higher Q coils I may go to

Setup:

The following schematic shows the setup, (Figure 2). In this configuration, I have used a simple 9V battery to stimulate the coil and have the scope connected to the tank via a coupling coil. An earlier configuration with a power supply resulted in too much noise and ripple to obtain a good reading. The tank is coupled lightly to the scope with a coupling coil. This coupling coil is simply two turns of hookup wire about 4 inches in diameter placed an inch or two from the coil to be tested and I have found no reason to space it further. Between the coupling coil and oscilloscope I utilize the 1:100 probe although testing with my 1:10 probe showed little impact. In this way, the coil under test is clean and well isolated from the measurement circuit.

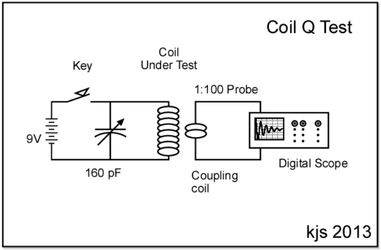

Figure 2. Schematic of test circuit.

For the tank I have used a small CT1C150 military spec cap ("APC"-type) with a capacitance range from about 20 to 160

with a 1 wire diameter gap. Measured inductance with the AADE L/C meter was 275 uH. The Q at 1 MHz was 353; OK but disappointing. The winding length was 2.9 inches.

The coil was then rewound with close spacing, resulting in 53.25 turns over a length of 1.52 inches. Inductance with the AADE meter was 378 uH. A 1 MHz Q measurement showed a higher effective L of 419, but with Q=219.

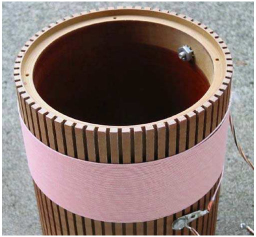

Close spaced solenoid coil.

I have not discussed the self resonant frequency of any of the inductors described in this note. However, it is easily measured by merely inserting the inductor in a 50 Ohm line. The fixture on page 6 of this note can be used. The "through" wire between the BNC connectors is removed and the coil is inserted. The signal generator is then tuned to find the first null frequency. This null is the result of stray parallel C resonating with the inductance to form a trap. The parallel resonant frequency was 4.0 MHz for the close spaced inductor shown in the above photo. This inductor had an inductance of 378 uH, which was measured at about 500 kHz with the AADE L/C meter. This yields an equivalent parallel capacitance of 4.2 pF. Using this and the 378 uH inductance measurement we calculate an effective 1 MHz L of 403 uH. This is still below the value we measured. The AADE number may well be in error because of the coil's parallel capacitance. Additional number crunching should yield a better model.

Update: 01Dec07

A Spider Coil.

Many of the crystal radio related web sites present information on Spider coils. I picked up a piece of Lexan Polycarbonate at Home Depot and cut it to form a nine segment coil form. The diameter ranged from 4.4 to 6.1 inches. This form was then wound with 39 turns of the 175/46 Litz wire. A photo follows.

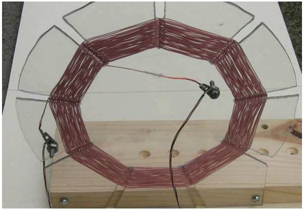

The inductance was 373 uH, but the Q results were disappointing at best: 550 kHz, Q=499; 1.002 MHz; Q=474; 1.504 MHz, Q=382. The EMRFD Q measuring scheme was used.

be related to the circuit components as follows, (Bucher, 1919):

d = pi (R / 2pifL) (3)

The angular frequency is often expressed as w= 2pif so this can be substituted into equation (3) above to simplify as

d = pi R / w L (4)

We turn now to looking at Q and its relation to the circuit components. The general expression for Q is:

Q = 2pifL / R (5)

Simplifying to

Q = w L / R (6)

With the term w L in both expressions, we can solve for w L in each and set them equal to each other:

QR = piR / d so:

Q = pi / d (7)

Dreadful, wasn't it? The inspiration here is that with a simple determination of the logarithmic decrement, deriving the circuit Q is a trivial exercise even a geologist can manage. No signal generators, no attenuators, no multiple measurements. All this should serve to reduce sources of error and give a faster and easier way to determine the quality of that most central component of your crystal radio, the coil.

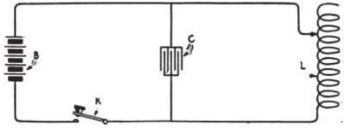

FIG. 8. Simple oscillation circuit illustrating the production of radio frequency currents.

Figure 1. Simple oscillation circuit, Bucher 1919.

As a professional Geologist, I must admit that most math involving more than ten fingers and ten toes (for advanced computations) is not my specialty. As such I will keep this aspect to an absolute minimum. Still, a bit of algebra may be useful to explain damping and logarithmic decrement. RF oscillations in a tank circuit are damped due to losses (primarily resistive) associated with that circuit. The amount of damping thus is related to the quality, or Q of the tank. In old texts the damping is generally determined by measuring the "Logarithmic Decrement" or the amplitude of successive oscillation peaks and taking the log of the ratio.

$$d = \ln(A1/A2) \tag{1}$$

A more generalized version of this formula taking into account measurements over many periods can be expressed as:

$$d = 1/n \ln(Ao/An) \tag{2}$$

Where n is the number of periods analyzed, Ao is the amplitude of the first peak and An is the amplitude of the peak n periods away. This equation allows a very simple determination of the log decrement with high accuracy. Because damping is due to resistive losses in the circuit, it can

Using Litz Rope.

The initial experiments were encouraging, although I was frustrated in not having achieved higher Q values. A review of many sites on the web quoted "Q over 1000" as if it was to be expected. One of the most informative web sites that I found was from Europe, http://www.crystal-radio.eu/enlctest.htm . He built numerous coils and then did careful Q measurements. Many of his measurements were consistent with mine. There was one central theme that emerged: If Q in excess of 1000 is desired, it is probably achievable only with Litz wire using many more strands than the 175 that I had used. Specifically, the wire that really made the difference in the measurements from Europe was 660/46 Litz. This is 660 strands of #46 wire. I’ve seen this stuff referred to as "Litz Rope," although that term is often reserved for the bundles that are up to two inches in diameter. The 660/46 wire is manufactured by Kerrigan-Lewis. They will sell to individuals, but have a 2 pound minimum. That’s a lot of rope. You can find smaller pieces of this wire for sale on the web, but with a high price tag.

Then my experiments expanded to the next level. This was the result of an unexpected package that showed up in my mail box. I opened this offering to discover a 54 foot piece of the 660/46 Litz rope, plus a most interesting letter. My benefactor was Steve, AA7U. Steve’s letter also mentioned the coil form that many of the crystal set guys use, a 4.5 inch OD coupler for large PVC pipe. I quickly purchased some. The part was found in the bins at Home Depot where it is their part number 405* and is marked "4 IN COUPLING HUBXHUB" and "(STYRENE)". The form cost was just over $1. Steve had done careful measurements of coils using this form and 660/46 Litz Rope.

I quickly wound the coil shown below.

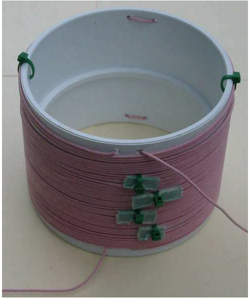

The photo does not really tell us about coil length. This coil has the 44 turns spaced to occupy 3 inches of the 3.6 inch form length. The wire is held in place at the ends with holes in the form. Small cable ties are then used to keep the windings from moving. I wanted to preserve the integrity of the Rope for additional experiments, so did not want to apply any glue. I cut some small strips of Plexiglas with filed notches. These strips are then placed over a pair of holes drilled in the form so that three wire turns are covered. The cable tie is then inserted.

The results were outstanding; finally, Q>1000. The details with the EMRFD Q measuring scheme:
F=550 kHz Q=1199
1011 kHz Q=1375
1600 kHz Q= 973 Steve's results with a similar coil were 1276, 1426, and 1129 at approximately the same frequencies.

frequency. The tank is generally energized with a signal generator and measured with an oscilloscope or sensitive digital voltmeter. Other components often may include attenuators, SWR analyzers, Spectrum analyzers etc. Sources of error enter with coupling, loading, uncertainty as to the actual internal resistance of the source generator, and the need for several independent measurements which are then multiplied or divided together, adding additional error. All this works very well for the engineer with good bench practice, good equipment ($$), and the patience to perform the measurements several times to check repeatability. All these things plus the required anality coefficient the current author lacks.

In this paper I propose an alternative approach that takes an analytical look at the oscillating tank waveform and determines the coil Q from that. This technique dispenses with a majority of the equipment involved with traditional methods but does require a digital oscilloscope. In recent years the affordability of these scopes has greatly increased. If you have a digital scope, or access to one, or were looking for an excuse to purchase one, then read on, this technique may be for you.

Some Theory:

The inspiration for this technique is certainly not new. I first found the following simple circuit, (Figure 1) on page 22 of Bucher's 1919 "Wireless Experimenter's Manual". Having a good deal of experience capturing and evaluating damped radio oscillations for a Spark Gap transmitter, I immediately saw the utility of the circuit for a simple Q determination based on the damping decrement of the tank.

An Analytical Approach to the measurement of coil Q

http://www.lessmiths.com/~kjsmith/crystal/coilqm.shtml

Kevin Smith

Background:

Measuring the values of your radio components is an interesting and fun part of the Crystal Radio Hobby. Not only does it help you design better radios, but helps to understand the physics behind the wires and plates. The radio coil, one of the principal components is generally that part wound by hand. All other components, diode, variable capacitor are usually bought ready-made. The coil sits at the heart of the set and yet it remains devilishly difficult to measure and characterize. Even cheapo inductance meters, (I mean, my cheapo meter) do not properly measure the coil inductance at radio frequency. Coil quality, well that is a whole other matter. Quality, or the Q-factor as it is often called is a bench-intensive and measurement-intensive proposition that even the most dedicated radio fan may shy away from.

If you wish to know your coil Q-factor you have the choice of tracking down and paying a lot of money for an old HP or Boonton Q-Meter, or setting up a test bench and making the required measurements yourself. A number of techniques for measuring the Q of a coil have been published and excellent summaries can be found in "Q Factor Measurements on L-C Circuits" by Jacques Audet, and "Experiments with Coils and Q-Measurement" a web page by Wes Hayward.

All the techniques discussed involve variously setting up a parallel or series tank containing the coil to be tested and an accompanying capacitor to tune it to the needed measurement

---

Steve used a traditional 3 dB bandwidth measurement to establish Q, but with the coil tuned by the variable capacitor in a HP4342A Q meter. That Q meter stops at values of 1000.

Although I was pleased to have finally reached a Q goal, the measurement correlation was of equal significance to me. This data is further validation of the EMRFD Q measurement scheme. The Q values that are slightly lower than those that Steve obtained can be attributed to losses in the variable capacitor I used and to experimental error. These values are consistent with those reported in the crystalradio. eu web site listed above.

Recent examination of capacitor Q. (December 8, 2008.)

The earlier information regarding variable capacitor Q was confusing. That info, starting on page 8, has been modified to emphasize that my measurements are of total resonator Q. Although I have yet to do additional measurements, I did a few calculations that are illuminating.

We begin with analysis to derive a suitable formula. This is shown below in a MathCad work sheet.

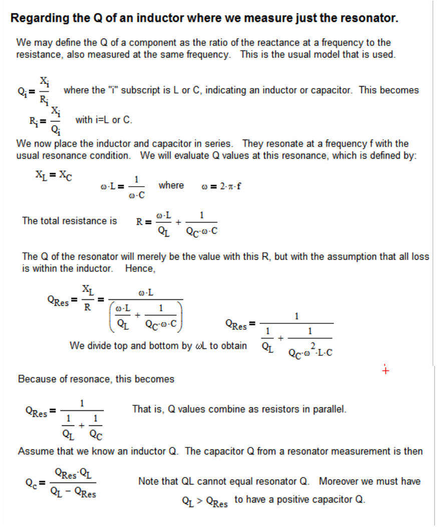

Regarding the Q of an inductor where we measure just the resonator.

We may define the Q of a component as the ratio of the reactance at a frequency to the resistance, also measured at the same frequency. This is the usual model that is used.

$Q_i = \frac{X_i}{R_i}$ where the "i" subscript is L or C, indicating an inductor or capacitor. This becomes

$R_i = \frac{X_i}{Q_i}$ with i=L or C.

We now place the inductor and capacitor in series. They resonate at a frequency f with the usual resonance condition. We will evaluate Q values at this resonance, which is defined by:

$X_L = X_C$

$\omega \cdot L = \frac{1}{\omega \cdot C}$ where $\omega = 2 \cdot \pi \cdot f$

The total resistance is $R = \frac{\omega \cdot L}{Q_L} + \frac{1}{Q_C \cdot \omega \cdot C}$

The Q of the resonator will merely be the value with this R, but with the assumption that all loss is within the inductor. Hence,

$$Q_{Res} = \frac{X_L}{R} = \frac{\omega \cdot L}{\left(\frac{\omega \cdot L}{Q_L} + \frac{1}{Q_C \cdot \omega \cdot C}\right)}$$

We divide top and bottom by ωL to obtain $Q_{Res} = \frac{1}{\frac{1}{Q_L} + \frac{1}{Q_C \cdot \omega^2 \cdot L \cdot C}}$

Because of resonace, this becomes

$Q_{Res} = \frac{1}{\frac{1}{Q_L} + \frac{1}{Q_C}}$ That is, Q values combine as resistors in parallel.

Assume that we know an inductor Q. The capacitor Q from a resonator measurement is then

$Q_c = \frac{Q_{Res} \cdot Q_L}{Q_L - Q_{Res}}$ Note that QL cannot equal resonator Q. Moreover we must have $Q_L > Q_{Res}$ to have a positive capacitor Q.

MathCad work sheet showing analysis.

Armed with the final formula in the above sheet, we can examine the capacitor Q data in more detail. We did this with Excel, beginning with a summary of the data and a "plot"

2Layer Bank	32/38	9.40	327	219
3Layer Bank	32/38	13.00	336	162
2Layer Bank	24	14.20	323	143
2Layer Bank	28	16.80	323	121
4Layer Bank	32/38	19.50	360	116
3Layer Bank	24	21.00	333	100
3Layer Bank	28	24.00	333	87
4Layer Bank	24	29.00	336	73
4Layer Bank	28	32.50	336	65
Double Layer	28	inf	355	0

showing the results. The plot is just a bar chart and serves only to illustrate that most of the results are close to each other. Only one resonator Q was below 300, the result of a single bearing variable capacitor. None of the resonator Q values exceeded 400. This is shown below.

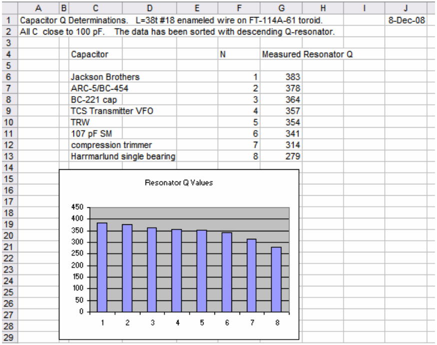

Capacitor Q Determinations. L=38t #18 enameled wire on FT-114A-61 toroid. 8-Dec-08

All C close to 100 pF. The data has been sorted with descending Q-resonator.

Capacitor	N	Measured Resonator Q
Jackson Brothers	1	383
ARC-5/BC-454	2	378
BC-221 cap	3	364
TCS Transmitter VFO	4	357
TRW	5	354
107 pF SM	6	341
compression trimmer	7	314
Harrmarlund single bearing	8	279

Sorted and plotted Capacitor impact on Resonator Q.

Having generated the bar chart, we proceeded to use the spread sheet to do some “what if” investigations. The first analysis assumed that out best capacitor was very good, so the highest resonator Q that we measured was almost the inductor Q value. If that really was the case, we can then calculate capacitor Q values. This case is at the top of the following page with QL=385.. After that, we assumed even higher inductor Q values and calculated the resulting capacitor Q results. See the next page.

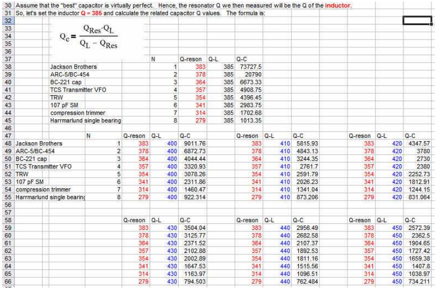

Assume that the "best" capacitor is virtually perfect. Hence, the resonator Q we then measured will be the Q of the inductor.
So, let's set the inductor Q = 385 and calculate the related capacitor Q values. The formula is:

$$Q_c = \frac{Q_{Res} \cdot Q_L}{Q_L - Q_{Res}}$$

	N	Q-reson	Q-L	Q-C
Jackson Brothers	1	383	385	73727.5
ARC-5/BC-454	2	378	385	20790
BC-221 cap	3	364	385	6673.33
TCS Transmitter VFO	4	357	385	4908.75
TRW	5	354	385	4396.45
107 pF SM	6	341	385	2983.75
compression trimmer	7	314	385	1702.68
Harrmarlund single bearing	8	279	385	1013.35

	N	Q-reson	Q-L	Q-C	Q-reson	Q-L	Q-C	Q-reson	Q-L	Q-C
Jackson Brothers	1	383	400	9011.76	383	410	5815.93	383	420	4347.57
ARC-5/BC-454	2	378	400	6872.73	378	410	4843.13	378	420	3780
BC-221 cap	3	364	400	4044.44	364	410	3244.35	364	420	2730
TCS Transmitter VFO	4	357	400	3320.93	357	410	2761.7	357	420	2380
TRW	5	354	400	3078.26	354	410	2591.79	354	420	2252.73
107 pF SM	6	341	400	2311.86	341	410	2026.23	341	420	1812.91
compression trimmer	7	314	400	1460.47	314	410	1341.04	314	420	1244.15
Harrmarlund single bearing	8	279	400	922.314	279	410	873.206	279	420	831.064

Q-reson	Q-L	Q-C	Q-reson	Q-L	Q-C	Q-reson	Q-L	Q-C
383	430	3504.04	383	440	2956.49	383	450	2572.39
378	430	3125.77	378	440	2682.58	378	450	2362.5
364	430	2371.52	364	440	2107.37	364	450	1904.65
357	430	2102.88	357	440	1892.53	357	450	1727.42
354	430	2002.89	354	440	1811.16	354	450	1659.38
341	430	1647.53	341	440	1515.56	341	450	1407.8
314	430	1163.97	314	440	1096.51	314	450	1038.97
279	430	794.503	279	440	762.484	279	450	734.211

Spreadsheet showing a variety of capacitor Q values resulting from assumed inductor Q (in blue) and the measured resonator Q values (in red.)

The results are interesting. If the best measured Q is indeed mostly that of the inductor, we see that most of the capacitor Q values are high with even the worst still over 1000. If we slowly allow the inductor Q to increase in our "thought experiment," the capacitor Q values begin to drop. However, the best of the lot are still several thousand. While none of this conjecture is really hard data, it certainly makes us feel more comfortable with inductor Q values close to 400 and that it is valid to assume high capacitor Q.

Examination of the central equation is revealing. It is clear that if the inductor Q becomes much higher, we will be able to

Spider	165/46	1.94	232	750
Basket weave	175/46	2.56	265	650
spider	660/46			641
Spider	100/45	2.41	238	620
Spider	100/45	2.44	241	620
Spider	100/45	1.73	149	540
spider	175/46	4.94	373	474
Loose Basket	32/38	5.30	317	376
Spider	40/44	4.16	248	375
Spider	40/44	4.16	225	340
Spider	40/44	2.93	154	330
Spider	32/38	7.50	331	277
Loose Basket	24.00	7.60	317	262
Basket weave	22.00	6.82	265	244
Loose Basket	28.00	8.50	317	234
Spider	24.00	9.50	327	216
Basket weave	32/38	10.40	332	201
Spider	28.00	10.60	327	194
Honeycomb	32/38	12.00	355	186
Spider	28.00	12.00	330	173
Basket weave	24.00	13.80	323	147
Basket weave	28.00	16.40	323	124
Honeycomb	24.00	18.50	347	118
Honeycomb	28.00	27.50	347	79
ferrite rod	50/46	2.58	250	610
toroid	22	3.93	240	384
toroid	18	3.20	195	383
toroid	22	8.20	449	344
toroid	175/46	4.82	244	318
toroid	50/46	5.23	248	298
ferrite rod	22		80	
ferrite rod	22	16.94	158	58.6

Radio-Frequency Resistance and Inductance

Qu = 2pifL / Rs

		1Mhz		
Coil Type	Wire	Rs	L	Q
		awg	ohm	uH
solenoid	660/46	0.90	200	1400
solenoid	660/46	0.91	200	1375
solenoid	175/46	2.88	265	579
solenoid	12	2.98	230	485
solenoid	14	3.28	230	440
solenoid	16	3.52	230	410
solenoid	18	4.01	230	360
solenoid	50/46	4.89	275	353
solenoid	32/38	6.20	327	331
solenoid	20	4.66	230	310
solenoid	16	7.60	319	264
solenoid	28	8.80	360	257
solenoid	22	5.78	230	250
solenoid	24	8.10	319	247
solenoid	28	10.10	360	224
solenoid	28	10.30	360	220
solenoid	50/46	10.84	378	219
solenoid	24	6.72	230	215
solenoid	28	9.60	327	214
solenoid	28	9.70	319	207
solenoid	26	8.03	230	180
solenoid	28		327	
Basket weave	660/46	1.03	186	1134
Spider	660/46	1.38	241	1000++
Basket weave	660/46	1.08	186	1082
Spider	660/46	0.90	150	1000+
spider	660/46			816

obtain much better data about the capacitors. The most recent inductor built has a Q that is three times higher than the one used for the capacitor difference. The next experiment is obvious and will be in the next update to this note.

Acknowledgements

Thanks go to my son Roger, KA7EXM, to colleagues Rick, KK7B, and Bob, N7FKI, for numerous discussions. Bob also provided some critical parts for my experiments. Special thanks go to Steve Ratzlaff, AA7U, for the crucial Litz wire that so greatly enhanced the experiments.

Do = Diameter of the finished cable over the strands in inches = 0.056
K = Constant depending on the number of strands = 2

Rac/Rdc = 1.0003 +2(660*0.0016 / 0.056)^2 * (0.0016*1000 / 10.44)^4 = 1.393

Therefore the AC Resistance of 660/46 litz wire is approximately:

The A.C. resistance is: 1.39 * 7.27 = 10.13 ohms/1000ft.

A 5inch diameter basket weave coil made from 660/46 Litz wire having an inductance of 230 uH will typically require some 40 turns or about 53ft of wire. At 10.13 Ohm/kft, that comes to an AC wire resistance of 0.53 ohms for the coil alone. Were all the losses represented by series resistance of the wire, the coil would have, at 1 MHz an unloaded Q = 2700! Naturally, wire resistance is not the only source of loss in the tank. There is a capacitor, metallic objects intruding into the magnetic field of the coil, diaelectric losses, eddy current and other losses. Its no wonder that the best coils just top out above Q = 1000 or so.

This is where I leave things. Time to get measuring. Below I provide my input data, have at it!

Kevin Smith

*Note that in 1925 the importance of coil Rs was understood, but the factor Q was apparently not used. I have taken (quite painfully) the L and Rs data from their plots and calculated the resulting Q. I recommend to those interested to download the pdf of their paper.

The following calculation is made to determine the AC resistance of the Litz wire used in a coil:
From different Litz wire manufacturing sites one can find tables and data allowing the calculation of the resistance of your favorite litz wire.

The formula for the D.C. resistance of any Litz construction is:

Rdc = Rs (1.0515)^Nb * (1.025)^Nc / Ns

Where:
Rdc = Resistance in Ohms/1000 ft.
Rs = Maximum D.C. resistance of the individual strands (4544 for 46awg wire)
Nb = Number of Bunching operations (assume = 2)
Nc = Number of Cabling operations (assume = 1)
Ns = Number of individual strands (assume = 660)

Rdc = 4544 (1.015)^2 (1.025)^1 / 660 = 7.27 ohms / 1000 ft.

The ratio of AC resistance to DC resistance of any Litz construction is:

Rac/Rdc = S + K (N Di / Do)^2 * G

Where:
S = Resistance ratio of individual strands when isolated (1.0003 for 46awg wire)
G = Eddy Current basis factor = (Di * sqrt(f) / 10.44)^4
F = Operating Frequency in HZ (assume 1MHz)
N = Number of strands in the cable = 660
Di = Diameter of the individual strands over the copper in inches = 0.0016

http://digital.library.unt.edu/ark:/67531/metadc66512/m2/1/high_res_d/metadc66512.pdf

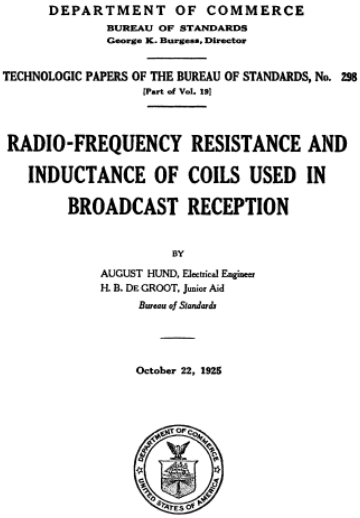

DEPARTMENT OF COMMERCE
BUREAU OF STANDARDS
George K. Burgess, Director

TECHNOLOGIC PAPERS OF THE BUREAU OF STANDARDS, No. 298
[Part of Vol. 19]

RADIO-FREQUENCY RESISTANCE AND INDUCTANCE OF COILS USED IN BROADCAST RECEPTION

BY

AUGUST HUND, Electrical Engineer
H. B. DE GROOT, Junior Aid
Bureau of Standards

October 22, 1925

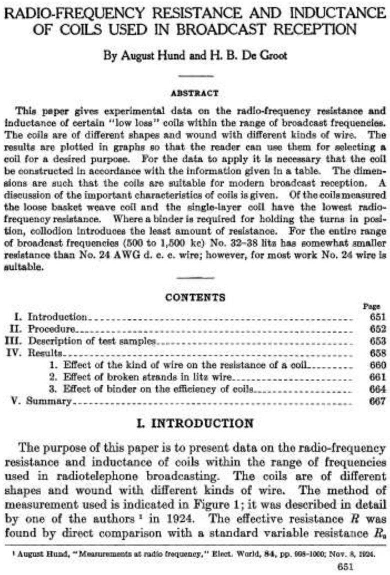

RADIO-FREQUENCY RESISTANCE AND INDUCTANCE OF COILS USED IN BROADCAST RECEPTION

By August Hund and H. B. De Groot

ABSTRACT

This paper gives experimental data on the radio-frequency resistance and inductance of certain "low loss" coils within the range of broadcast frequencies. The coils are of different shapes and wound with different kinds of wire. The results are plotted in graphs so that the reader can use them for selecting a coil for a desired purpose. For the data to apply it is necessary that the coil be constructed in accordance with the information given in a table. The dimensions are such that the coils are suitable for modern broadcast reception. A discussion of the important characteristics of coils is given. Of the coils measured the loose basket weave coil and the single-layer coil have the lowest radio-frequency resistance. Where a binder is required for holding the turns in position, collodion introduces the least amount of resistance. For the entire range of broadcast frequencies (500 to 1,500 kc) No. 32-38 litz has somewhat smaller resistance than No. 24 AWG d. c. c. wire; however, for most work No. 24 wire is suitable.

CONTENTS

Page
I. Introduction 651
II. Procedure 652
III. Description of test samples 653
IV. Results 658
  1. Effect of the kind of wire on the resistance of a coil 660
  2. Effect of broken strands in litz wire 661
  3. Effect of binder on the efficiency of coils 664
V. Summary 667

I. INTRODUCTION

The purpose of this paper is to present data on the radio-frequency resistance and inductance of coils within the range of frequencies used in radiotelephone broadcasting. The coils are of different shapes and wound with different kinds of wire. The method of measurement used is indicated in Figure 1; it was described in detail by one of the authors[1] in 1924. The effective resistance R was found by direct comparison with a standard variable resistance R_s

---

[1] August Hund, "Measurements at radio frequency," Elect. World, 84, pp. 998-1000; Nov. 8, 1924.

651

also includes Q vs f curves at constant Rp values. It is clear that, while tuning across the BCB spectrum, the Rp of the coil changes continuously.

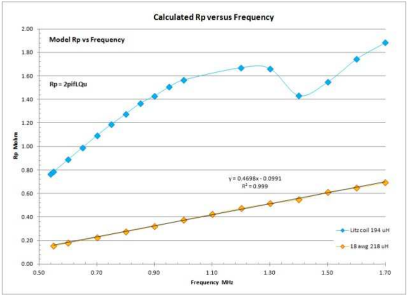

Calculating the coil Rp from the data (Q:C:f), I have reproduced the above graph of the coil Rp vs frequency. It is evident that the trend follows a perfectly straight line for the low-Q coil. I would expect a similar straight-line relation for the litz coil but again things look strange. My impression here is that of a much steeper Rp vs f relation where the calibration on the Q-meter breaks down near 1MHz (Rp ~1.5 MOhm) and the data wanders until a second linear trend (with about the same slope) is re-established between 1.4 to 1.7MHz.

Now time for a few calculations, just how good can a coil Q get?

tuned a big litz coil with cheapo capacitors. Having an idea of your set's Ql will allow a better selection of the proper diode for matching.

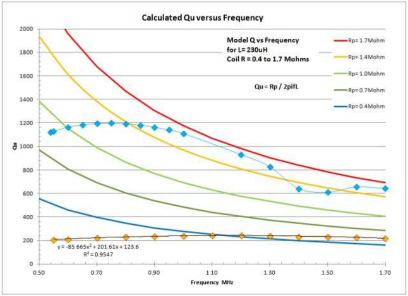

The above plot is from a series of measurements on two of my coils kindly made for me by Steve Ratzlaff, AA7U. The blue diamonds are the measurements on a 660/46 Litz basket beast and the yellow diamonds are measurements on a modest 18awg tapped solenoid wound of a cardboard form. The data from the low Q coil form a lovely continuum across the BCB band while the litz coil data above 1MHz or so seem to be declining and erratic. Steve measured the coil a great many times and found good consistant results. The measurements are excellent but I must say, I am a bit suspicious that the high-Q coil may be at the limit of the HP Q-meter calibration, or perhaps 660/46 litz Q tops out at 0.7-0.8 MHz and crashes above that. (I have not yet found time to persue this). The plot

after the test coil and the condenser C had been tuned to resonance. The apparent inductance was calculated from the formula

$$L = \frac{25350}{f^2 C}$$

where the inductance L is in microhenries, frequency f in kilocycles per second, and the measured resonance capacity C of the series condenser is in microfarads.

II. PROCEDURE

The method employed was to compare at radio-frequency coils of several types commonly used in receiving sets. The coils had been adjusted to the same self-inductance at a frequency of 1 kc per second. This represents the minimum value of the apparent inductance, larger values being obtained at the broadcast frequencies. The direct-current resistance was different for the various coils, since some shapes required more wire than others for the same inductance at 1 kc.

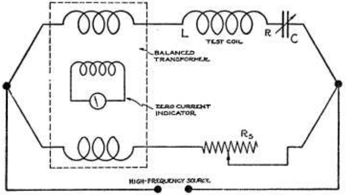

FIG. 1.—*Arrangement for the determination of radio-frequency resistance and inductance*

The quality of coils can be considered in terms of a number of different properties,[2] each of which is of importance in the use of

---

[2] It may help the reader to recall that the function of the coil is essentially to introduce inductance in the circuit, a given amount of inductance being introduced with the minimum possible length of wire and of resistance. On account of the capacity action of a coil the apparent inductance is usually much larger at broadcast frequencies, since the decrease due to skin effect action is small in comparison. The capacity action of a coil tends to transfer more or less energy across the turns than along them, thus turning the coil with increasing frequency gradually into a condenser with a resistance which is due to the insulation between the turns.

Technologic Papers of the Bureau of Standards, Vol. 19

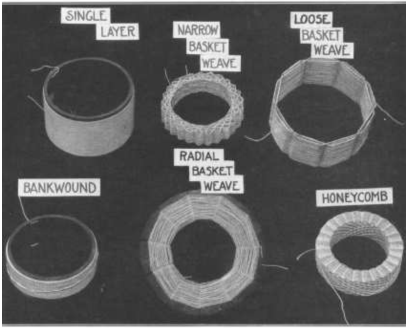
SINGLE LAYER
NARROW BASKET WEAVE
LOOSE BASKET WEAVE
RADIAL BASKET WEAVE
BANKWOUND
HONEYCOMB

FIG. 2.—*Low-loss coils used for measurement*

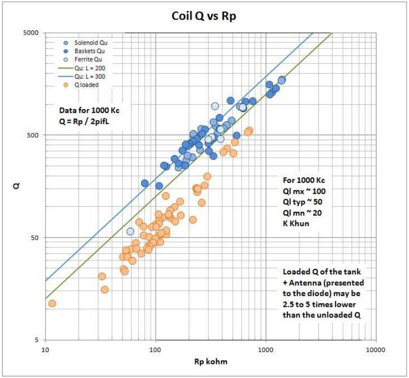

On the plot I also post in orange circles my estimate of what the loaded Q presented to the diode (Antenna + Tank) might look like. Loaded Q will always be lower than the unloaded Q by several times. I have estimated that high-end sets (big litz, silver-plated ceramic insulated caps, best wiring practices) may lower the Q by about 2 1/2 times (B Tongue's performance set has Q = 700). At the low end (vintage components, small solid wire coils, taps) the load may lower Q by up to 5 times. Ken Khun in his excellent web book states that typical sets at 1Mhz have a loaded Q between 20 and 100, 50 typical and this is where most of the data falls. I scaled the divisor by Qu to produce the above plot but note that this is merely an estimate. Somewhere some bloke has no doubt cardboard form, sealed, of course. With various "open" coils, spiderweb, basket weave, diamond weave, etc. you may expect to double that.. possibly. The advantage in open coils derives from first the separation between adjacent wire turns which reduces self-capacitance in the coil, and secondly from the obvious lack of a form. All materials used in the coil will have some amount of dielectric losses associated with them and the less material used the better. Air core coils are best in this respect. I am not here to speak about ferrite coils as I have no experience with them. Frankly, they seem (to me) a bit like cheating.

The following graph presents data that I have scoured off the web. My primary sources include Hund and Groot, 1925*, Wes Hayward, Dave Schmarder, Ken Khun, Mike Tuggle, Steve Ratzlaff, and Dick Kleijer. 10,000 thanks for those who post their data on the web! The following plot gives the Q value as a function of coil Rs (series resistance. In striving for a high Q coil, in effect one is eliminating losses and lowering the series resistance as much as possible. There is more to it of course. The Q formula Qu = 2pifL / Rs tells us that Q is also a function of the coil inductance L and the frequency f of the measurement. All measurements chosen for plotting are made around 1 Mhz and most of the coils are in the L = 200 - 350 uH range. So, despite the formula, in this plot the coil Rp is the main driver. From the plot it should be apparent that winding a coil with Q = 200 or better should be a no-brainer. If your coil Q is less, you just aren't trying. At the high end, Q's > 1000 seem to be pretty extreme and these coils are expensive. You better be using big Litz 660/46 and use a basket design (although the two best coils on the plot were solenoids).

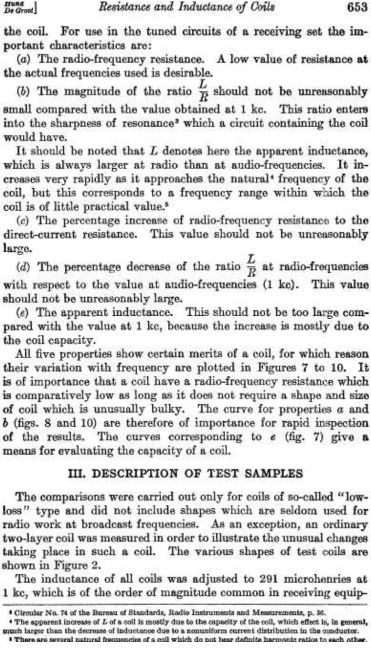

the coil. For use in the tuned circuits of a receiving set the important characteristics are:

(*a*) The radio-frequency resistance. A low value of resistance at the actual frequencies used is desirable.

(*b*) The magnitude of the ratio $\frac{L}{R}$ should not be unreasonably small compared with the value obtained at 1 kc. This ratio enters into the sharpness of resonance[3] which a circuit containing the coil would have.

It should be noted that L denotes here the apparent inductance, which is always larger at radio than at audio-frequencies. It increases very rapidly as it approaches the natural[4] frequency of the coil, but this corresponds to a frequency range within which the coil is of little practical value.[5]

(*c*) The percentage increase of radio-frequency resistance to the direct-current resistance. This value should not be unreasonably large.

(*d*) The percentage decrease of the ratio $\frac{L}{R}$ at radio-frequencies with respect to the value at audio-frequencies (1 kc). This value should not be unreasonably large.

(*e*) The apparent inductance. This should not be too large compared with the value at 1 kc, because the increase is mostly due to the coil capacity.

All five properties show certain merits of a coil, for which reason their variation with frequency are plotted in Figures 7 to 10. It is of importance that a coil have a radio-frequency resistance which is comparatively low as long as it does not require a shape and size of coil which is unusually bulky. The curve for properties *a* and *b* (figs. 8 and 10) are therefore of importance for rapid inspection of the results. The curves corresponding to *e* (fig. 7) give a means for evaluating the capacity of a coil.

III. DESCRIPTION OF TEST SAMPLES

The comparisons were carried out only for coils of so-called "low-loss" type and did not include shapes which are seldom used for radio work at broadcast frequencies. As an exception, an ordinary two-layer coil was measured in order to illustrate the unusual changes taking place in such a coil. The various shapes of test coils are shown in Figure 2.

The inductance of all coils was adjusted to 291 microhenries at 1 kc, which is of the order of magnitude common in receiving equip-

---

[3] Circular No. 74 of the Bureau of Standards, Radio Instruments and Measurements, p. 36.

[4] The apparent increase of L of a coil is mostly due to the capacity of the coil, which effect is, in general, much larger than the decrease of inductance due to a nonuniform current distribution in the conductor.

[5] There are several natural frequencies of a coil which do not bear definite harmonic ratios to each other.

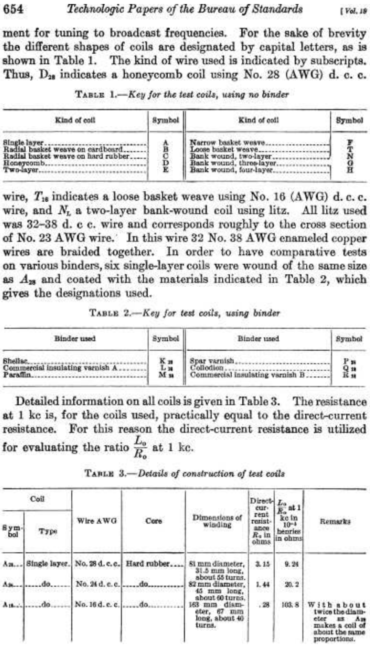

654 *Technologic Papers of the Bureau of Standards* [Vol. 19

ment for tuning to broadcast frequencies. For the sake of brevity the different shapes of coils are designated by capital letters, as is shown in Table 1. The kind of wire used is indicated by subscripts. Thus, D_{28} indicates a honeycomb coil using No. 28 (AWG) d. c. c. wire, T_{16} indicates a loose basket weave using No. 16 (AWG) d. c. c. wire, and N_L a two-layer bank-wound coil using litz. All litz used was 32–38 d. c c. wire and corresponds roughly to the cross section of No. 23 AWG wire. In this wire 32 No. 38 AWG enameled copper wires are braided together. In order to have comparative tests on various binders, six single-layer coils were wound of the same size as A_{28} and coated with the materials indicated in Table 2, which gives the designations used.

TABLE 1.—*Key for the test coils, using no binder*

Kind of coil	Symbol	Kind of coil	Symbol
Single layer	A	Narrow basket weave	F
Radial basket weave on cardboard	B	Loose basket weave	T
Radial basket weave on hard rubber	C	Bank wound, two-layer	N
Honeycomb	D	Bank wound, three-layer	G
Two-layer	E	Bank wound, four-layer	H

TABLE 2.—*Key for test coils, using binder*

Binder used	Symbol	Binder used	Symbol
Shellac	K_{28}	Spar varnish	P_{28}
Commercial insulating varnish A	L_{28}	Collodion	Q_{28}
Paraffin	M_{28}	Commercial insulating varnish B	R_{28}

Detailed information on all coils is given in Table 3. The resistance at 1 kc is, for the coils used, practically equal to the direct-current resistance. For this reason the direct-current resistance is utilized for evaluating the ratio $\frac{L_o}{R_o}$ at 1 kc.

TABLE 3.—*Details of construction of test coils*

Coil		Wire AWG	Core	Dimensions of winding	Direct-current resistance R_o in ohms	$\frac{L_o}{R_o}$ at 1 kc in 10^{-4} henries in ohms	Remarks
Symbol	Type						
A_{28}	Single layer	No. 28 d. c. c.	Hard rubber	81 mm diameter, 31.5 mm long, about 55 turns.	3.15	9.24	
A_{24}	do	No. 24 d. c. c.	do	82 mm diameter, 45 mm long, about 60 turns.	1.44	20.2	
A_{16}	do	No. 16 d. c. c.	do	163 mm diameter, 67 mm long, about 40 turns.	.28	103.8	With about twice the diameter as A_{28} makes a coil of about the same proportions.

Coil Q:
http://www.lessmiths.com/~kjsmith/crystal/coilq.shtml
Kevin Smith

Introduction:

I have been building and studying crystal radios for some time now and slowly begin to learn a few things about these marvelous sets. In this section I begin my exploration of coil, coil quality and that mysterious dimensionless factor, Q... This page thus is a bit preliminary as I have not yet begun any measurements to determine the Q of my coils. Please bear with me.

My purpose here is to present some facts and data that has resulted from my explorations of the web. I have often wondered at what the quality factor of my coils should be and as often realized that I really do not have any expectations as to what is possible. Now, having dome some research, I can with some confidance say that this is solved. Coils as used in crystal radio broadcast band reception typically employ coils with Q factors ranging from 100+ (pretty lousy) through the several 100's (decent) and on up to 1000 or more for those remarkable Big-Litz wonders ($$$). Knowing the expected (or actual) Q of your coil is important in so far as it impacts the choice of diode to be used in the set. On my page of Diode Calibration I present a summary graphic which indicates how the diode Rd value relates to the tank parallel resistance Rp. This Rp in turn is a function of the coil Q. Schezzam! So here we are.

In your set construction, with a good effort and good engineering practice one can easily expect to wind a solenoid coil in the Q = 200 range without much trouble, even with a

TABLE 3.—*Details of construction of test coils*—Continued

Coil		Wire AWG	Core	Dimensions of winding	Direct-current resistance R_e in ohms	$\frac{L_e}{R_e}$ at 1 kc in 10^{-5} henries in ohms	Remarks
Symbol	Type						
A_L	Single layer.	No. 32-38 d. s. c. litz.	Hard rubber	81 mm diameter, 59 mm long, about 65 turns.	1.25	23.27	
B_{28}	Radial basket weave.	No 28 d. c. c.	Cardboard 1.5 mm thick, 11 spokes, with slots 2 mm wide.	Inside diameter 65 mm.	3.08	9.45	
C_{28}	do	do	Hard rubber 3 mm thick, 11 spokes, with slots 1 mm wide.	Inside diameter 64 mm.	3.59	8.11	
C_{24}	do	No. 24 d. c. c.	do	do	1.65	17.62	
C_L	do	No. 32-38 d. s. c. litz.	Hard rubber 3 mm thick, 13 spokes, with slots 2.5 mm wide.	do	1.24	23.46	
D_{28}	Honeycomb	No. 28 d. c. c.	Air, and just enough collodion to hold coil together.	Inside diameter 57 mm; four diagonal retarded winding; 25 pins 2 mm in diameter.	3.31	8.79	Pins are set on inside diameter and coil builds up along radius.
D_{24}	do	No. 24 d. c. c.	do	Inside diameter 65 mm; otherwise as above.	1.35	21.55	Do.
D_L	do	No. 32-38 d. s. c. litz.	do	do	1.11	26.21	Do.
E_{28}	Double layer.	No. 28 d. c. c.	Hard rubber	81 mm inside diameter.	2.75	10.58	Very poor co...
F_{28}	Narrow basket weave.	do	Air, and just enough collodion to hold coil together.	Outside diameter 76 mm; 25 pins of 1.5 mm diameter; four diagonal retarded winding.	2.97	9.8	Pins are set on outside diameter and coil builds up along axis.
F_{24}	do	No. 24 d. c. c.	do	do	1.30	22.38	Do.
F_L	do	No. 32-38 d. s. c. litz.	do	do	1.07	27.2	Do.
G_{28}	3-layer bank wound.	No. 28 d. c. c.	Hard rubber, coil held together by collodion.	Inside diameter 81 mm; length of coil 10 mm.	2.70	10.78	
G_{24}	do	No. 24 d. c. c.	do	Inside diameter 81 mm; length of coil 14 mm.	1.17	24.87	
G_L	do	No. 32-38 d. s. c.	do	Inside diameter 81 mm; length of coil 17 mm.	.88	33.06	
H_{28}	4-layer bank wound.	No. 28 d. c. c.	do	Inside diameter 81 mm; length of coil 7 mm.	2.46	11.83	
H_{24}	do	No. 24 d. c. c.	do	Inside diameter 81 mm; length of coil 11 mm.	1.14	25.51	
H_L	do	No. 32-38 d. s. c. litz.	do	Inside diameter 81 mm; length of coil 12 mm.	.86	33.85	
K_{28}	Single layer.	No. 28 d. c. c.	Hard rubber, using shellac as a binder.	Same as A_{28}	3.11	9.36	Used for binder test.
L_{28}	do	do	Hard rubber, using commercial insulating varnish A as binder.	do	3.20	9.09	Do.
M_{28}	do	do	Hard rubber, using paraffin as a binder.	do	3.14	9.27	Do.

TABLE 3.—*Details of construction of test coils*—Continued

Coil		Wire AWG	Core	Dimensions of winding	Direct-current resistance R_o in ohms	$\frac{L_o}{R_o}$ at 1 kc in 10^{-6} henries in ohms	Remarks
Symbol	Type						
N_{28}	2-layer bank wound.	No. 28 d. c. c.	Hard rubber, coil held together by collodion.	Inside diameter 81 mm.	2.65	10.975	
N_{24}	do	No. 24 d. c. c.	do	Inside diameter 81 mm; length of coil 20 mm.	1.25	23.27	
N_L	do	No. 32–38 d. s. c. litz.	do	Inside diameter 81 mm; length of coil 25 mm.	.95	30.63	
P_{28}	Single layer.	No. 28 d. c. c.	Hard rubber, using spar varnish as a binder.	Same as A_{28}	3.14	9.27	Used for binder test.
Q_{28}	do	do	Hard rubber, using collodion as a binder.	do	3.11	9.36	Do.
R_{28}	do	do	Hard rubber, using commercial insulating varnish B as a binder.	do	3.21	9.07	Do.
T_{28}	Loose basket weave.	do	Air, and just enough collodion to hold coil together.	9 diagonal alternating winding; 9 pairs of pins set on a circle of 92.5 mm diameter; diameter of pins 2.5 mm; and small spacing between a pair of pins 10 mm.	3.19	9.13	
T_{24}	do	No. 24 d. c. c.	do	do	1.42	20.5	
T_L	do	No. 32–38 d. s. c.	do	do	1.21	24.05	

Figures 3 to 6 show how coils of these types are wound. Figure 3 illustrates the method used for winding a "three-layer bank wound"

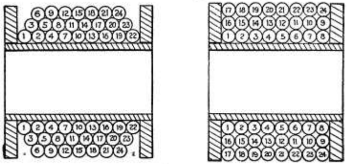

FIG. 3

Three-layer bank winding

Ordinary three-layer winding (useless for currents of the broadcasting range of frequency)

---

7. All the insulating materials which were used as binders caused very slight increases in the resistance of the coils. Collodion seems best, and also has the inherent advantage of drying rapidly after application to the coil. This is of especial advantage in the construction of a bank-wound coil.

WASHINGTON, May 25, 1925.

V. SUMMARY

1. The various experimentally obtained curves given in this paper can be used as design bases for comparing coils of six types for any frequency in the broadcast range. For these data to apply it is necessary that the coils be constructed in accordance with the information given in Table 3. The coil dimensions are such that the coils are applicable to modern broadcast reception. A statement of the important characteristics of coils is given in the introduction.

2. The curves shown in Figures 9 and 11 give the changes of resistance and of $\frac{L}{R}$ with frequency. High values of these ratios do not in all cases correspond to high values of radio-frequency resistance. In some cases, for instance, a particular coil has a relatively high value for $\frac{\Delta R}{R_o}$, although the actual radio-frequency resistance is not large, because the direct current resistance is comparatively low.

3. The curves of Figures 8 and 10 give the actual radio-frequency resistance and ratio of inductance to resistance at various frequencies.

4. Of the coils measured the loose basket weave coil and the single-layer coil, and next to them the radial basket weave coil wound on hard rubber, have the lowest radio-frequency resistance. The four-layer bank-wound coil and the honeycomb winding have the highest resistance. This can not, however, be generalized to other frequency ranges. For instance, for low-frequency sets (20 to 100 kc) the multilayer bank-wound coil and the honeycomb coil have relatively low resistance, and besides are good coils mechanically while the loose basket weave coil has no special advantage and the single-layer coil can not be used on account of excessive size.

5. There appears to be little reduction of resistance at the lower frequencies in spacing the turns, so that the advantage of getting a smaller resistance is small compared with the disadvantage of requiring a coil twice as long.

6. The use of Nos. 32–38 litz gives coils of somewhat lower resistance than coils wound with solid wire of the same cross section. No. 24 AWG solid wire has less resistance than No. 28 wire, and No. 16 wire for a certain range has less resistance than Nos. 24 and 28 wire. If solid wire is used it does not appear necessary to use wire larger than No. 24 AWG. This conclusion can not, of course, be extended outside the broadcast frequency range; for instance, No. 16 solid wire would be better for frequencies above 5,000 kc.

---

coil in comparison with an ordinary three-layer coil. Figure 4 illustrates the method used in making the "narrow basket weave" coil, in which the coil is built up along the pins, which are removed after the coil is finished. Figure 5 illustrates the method used in winding the " loose basket weave" coil, which requires a pair of pins for each corner. A spacing of 10 mm between the two pins of a pair was

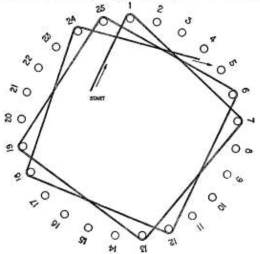

FIG. 4.—*Narrow basket weave. Four diagonal retarded winding (using 25 pins set on a circle. Coil builds up along axis)*

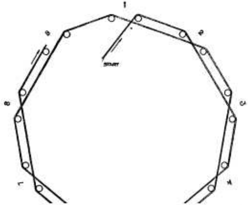

FIG. 5.—*Loose basket weave. Nine diagonal alternating winding (using nine pairs of pins set on a circle. Coil builds up along the axis)*

used. Figure 6 shows the method used in making the "honeycomb" coil. The zigzag winding is illustrated by the view of the entire cylinder surface. The winding is built up along the radii of the coil.

IV. RESULTS

The results are exhibited in graphical form. Figure 7 gives curves for the apparent self-inductance of the various test coils as a function of frequency. As on all curve sheets, the dotted curves indicate the different coils using No. 32–38 litz. It is seen that the "loose basket weave" type of coil and the "single-layer" coils give the lowest apparent inductance over the entire range of broadcast frequencies (500 to 1,500 kc). This indicates that the coil capacity is comparatively low, while the ordinary two-layer coil (E_{28}) acts more or less like a condenser, since the coil capacity is exceedingly large.

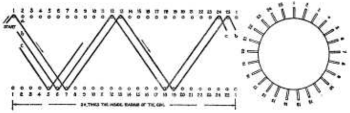

FIG. 6

Honeycomb coil. Four diagonal retarded winding (using 25 pins. Coil builds up along radius) *Cross-section of winding frame with the 25 pins*

Figure 8 gives the curves for the radio-frequency resistance. It is seen again that the values of the resistance vary greatly. Naturally, the ordinary two-layer coil (E_{28}) has very large radio-frequency resistance within the broadcasting range. Though its direct-current resistance is only 2.75 ohms, the effective resistance at 500 kc is 162 ohms, at 580 kc it is 465 ohms, and at 748 kc it is 1,800 ohms. The resistance increases very rapidly until it is mostly due to the dielectric resistance across the insulation between the layers. A three-layer coil wound in the ordinary way (ordinary multi-layer coil) and adjusted, like all coils discussed in this paper, to 291 microhenries at 1 kc has a radio-frequency resistance of several thousand ohms at a frequency as low as 400 kc.

Figure 9 gives the curves for the ratio of the increase of the radio-frequency resistance over its direct-current value to the direct-current value. It is seen that the loose basket weave and the single-layer coils have low values; next comes the radial basket weave coil,

---

TABLE 5.—*Variation with frequency of radio-frequency resistance (in ohms) for different binders*

Frequency in kilocycles	A_{28}, no binder	Q_{28}, collodion	R_{28}, commercial insulating varnish B	L_{28}, commercial insulating varnish A	K_{28} shellac	M_{28} paraffin
300	3.9	3.8	3.9	3.8	3.7	3.9
400	4.4	4.3	4.4	4.3	4.1	4.6
500	5.1	4.9	5.0	4.8	4.8	5.2
600	5.9	5.5	5.7	5.3	5.6	6.0
700	6.7	6.3	6.6	6.0	6.6	6.8
800	7.6	7.1	7.5	6.8	7.6	7.8
900	8.5	8.0	8.5	7.7	8.8	9.0
1,000	9.4	8.9	9.6	8.8	10.1	10.3
1,100	10.3	10.0	10.8	10.1	11.4	11.8
1,200	11.3	11.0	12.0	11.5	12.8	13.3
1,300	12.2	12.1	13.3	13.1	14.2	15.0
1,400	13.2	13.2	14.6	14.7	15.7	16.7
1,500	14.2	14.4	15.9	16.4	17.2	18.5

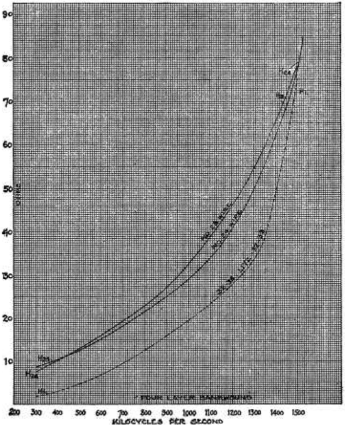

some cases the effective resistance of a coil using a binder came out smaller by a fraction of an ohm than for the coil using no binder. The binder seemed to make an accurate resistance adjustment more difficult.

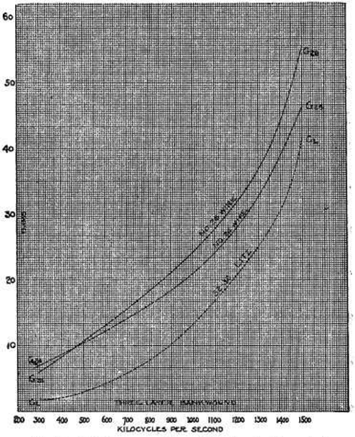

FIG. 18.—*Radio-frequency resistance of three-layer bank-wound coils*

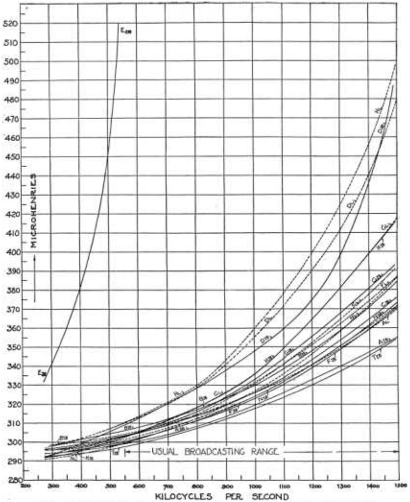

FIG. 7.—*Apparent inductance at different frequencies (minimum inductance is 291 microhenries for all coils)*

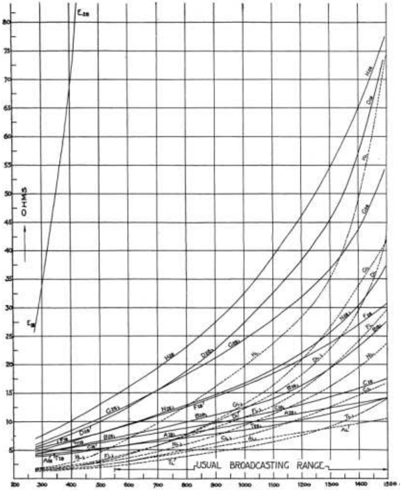

FIG. 8.—*Radio-frequency resistance at different frequencies*

2981—21. (Face p. 658.) N

3. EFFECT OF BINDER ON THE EFFICIENCY OF COILS

The binders used in this test were each applied to a single-layer coil, and just enough of it to cover the entire surface. The varying nature of the different binders probably resulted in different thicknesses. The coils in every case were dried thoroughly. The measurements were a little difficult when a binder was used, even though great care was taken to have the binder dried. For instance, in

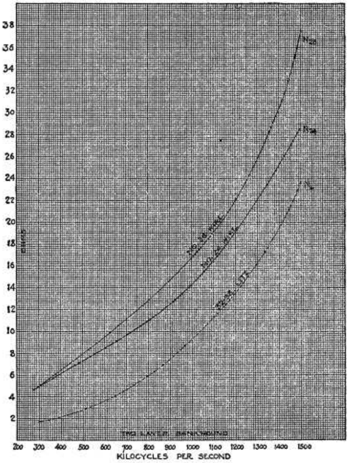

FIG. 17.—*Radio-frequency resistance of two-layer bank-wound coils*

Table 4 shows that it is not so serious to have a few strands broken, since the radio-frequency current apparently finds its way back across the strands. Even if, say, six strands are broken, the radio-frequency resistance is 3.4 ohms as compared with 3.1 ohms for perfect strands.

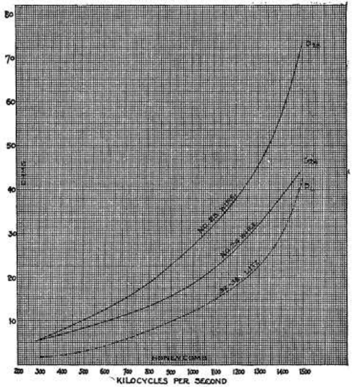

FIG. 16.—*Radio-frequency resistance of honeycomb coils*

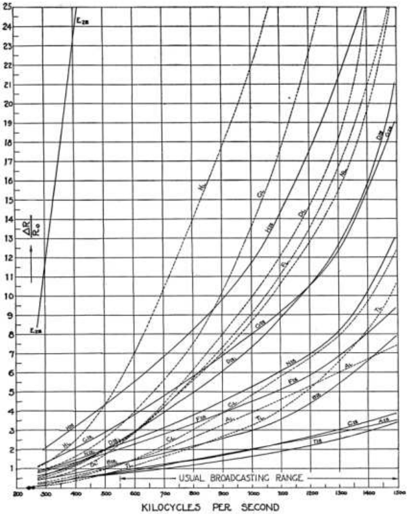

FIG. 9.—*Ratio of the increase of resistance over its direct-current value to the direct-current value*

[illegible]—25. (Face p. 668.) N

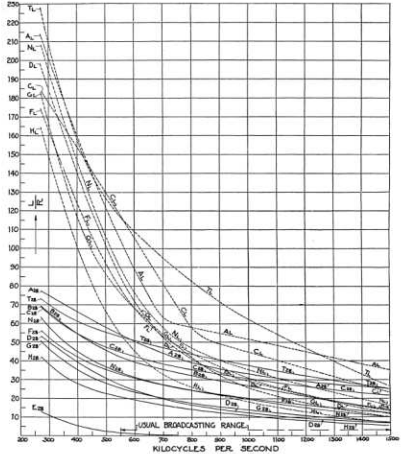

FIG. 10.—*Ratio of inductance to resistance*

10960—25. (Face p. 658.) No.

TABLE 4.—*Resistance of litz wire Nos. 32–38, at 750 kc*

Broken strands	Resistance	Broken strands	Resistance	Broken strands	Resistance	Broken strands	Resistance
	Ohms		*Ohms*		*Ohms*		*Ohms*
0	3.1	8	3.6	16	5.4	24	9.5
1	3.2	9	3.8	17	5.6	25	10.8
2	3.2	10	3.8	18	6.1	26	13.5
3	3.3	11	4.2	19	6.4	27	14.4
4	3.3	12	4.4	20	7.4	28	16.5
5	3.3	13	4.4	21	7.6	29	21.7
6	3.4	14	4.4	22	7.8	30	42.4
7	3.5	15	4.7	23	8.4	31	51.6

Litz used is Nos. 32-38 and has 32 strands of No. 38 AWG enameled wire braided together.

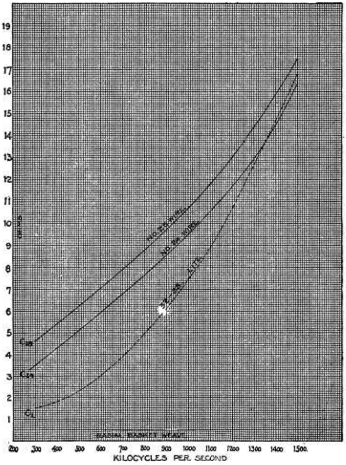

convenient use in receiving sets. Figure 12 shows a single-layer coil the turns of which are spaced by a distance equal to the diameter of the wire. There seems to be no material reduction in resistance except at the higher frequencies.

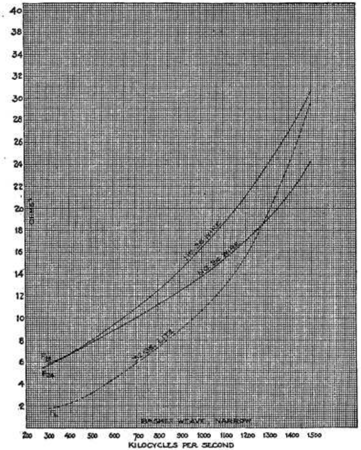

FIG. 14.—*Radio-frequency resistance of narrow basket-weave coils*

2. EFFECT OF BROKEN STRANDS IN LITZ WIRE

The question is often asked concerning litz wire as to the effect on the resistance of some of the strands being broken or not connected at the respective ends.

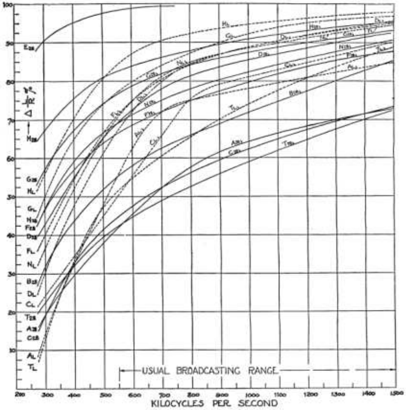

FIG. 11.—*Percentage decrease in the ratio of inductance to resistance below the value at 1 kc*

50081—25. (Face p. 658.) N

using hard rubber for the core, while the four-layer bank wound, three-layer bank wound, and the honeycomb coils show considerable percentage increases.

Figure 10 gives the curves for the ratio $\frac{L}{R}$. It is always smaller than the ratio $\frac{L_o}{R_o}$ at 1 kc. Since the apparent inductance and the radio-frequency resistance increase with the frequency, this ratio in certain cases has only small changes although both L and R are changing rapidly. Nevertheless, these curves show about the same comparisons as the resistance curves.

Figure 11 gives the curves for the percentage decrease in the ratio $\frac{L}{R}$ at different frequencies.

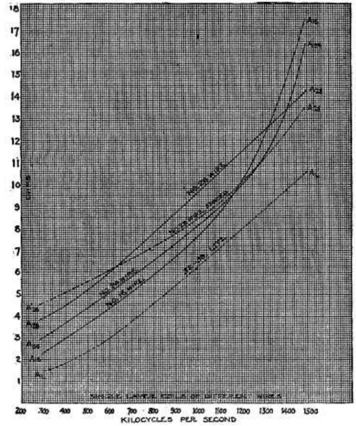

1. EFFECT OF THE KIND OF WIRE ON THE RESISTANCE OF A COIL

In order to show the effect of the size of the wire, curves for the resistance for No. 28, No. 24, and No. 16 d. c. c. are plotted and compared among themselves and against No. 32–38 litz. These comparisons are shown in Figures 12 to 19.

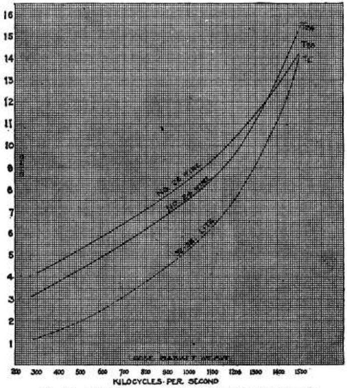

Fig. 13.—*Radio-frequency resistance of loose basket-weave coils*

It will be noted that in all cases litz, which corresponds roughly to No. 23 solid wire as regards its cross section, had the lowest effective resistance. If solid wire is used, it appears unnecessary to use wire larger than No. 24, although No. 16 gives, for the lower frequencies, resistances which are slightly lower. Such a large size of wire would, however, make the winding of certain types of coils more difficult and the size of the finished coil too large for

KEVIN'S WEBSURFER HANDBOOK IX
FOR CRYSTAL RADIO

CAPACITOR LOSSES and WIRE RESISTANCE

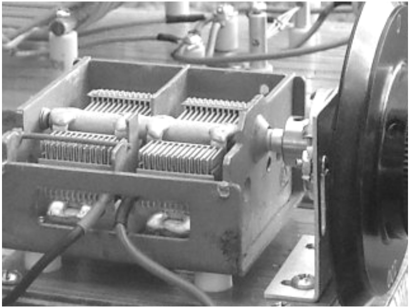

Kevin Smith
2013

Printing / Binding Instructions

1. Choose “fit to page” in print menu
2. Print document double sided on letter size paper
3. Cut the entire printed document in half
4. Fold over making sure the page numbering is continuous
5. For the cover: Print just the first page on card stock paper
   Cut the cover in half as well
6. Assemble the covers on the document
7. Punch the left side for a binding, spiral or comb as desired

http://www.lessmiths.com/~kjsmith/crystal/catalog.shtml

KJ Smith

NOTES:

NOTES:

INTRODUCTION

This handbook stems from my studies of set resonance, LRC circuits and the quality of the tank components. My measurements on coil Q led me to believe I need to take the other components of the tank more seriously, noteably the capacitor and even the wire from which the coil is wound. In many radio articles I have read the Q-factor of the tank variable capacitor is generally assumed to be very high. High enough as to be disregarded as a significant contributor to the tank Q. Recent discussions on the Radio Board suggest that this may not be the case but in fact contribute significantly to tank resistive losses.

The first several chapters of this handbook then address capacitors, both from a general standpoint and then from the question of capacitor losses. Several articles on measurement techniques are included although it is noted that the capacitors they address appear to botton out at the 1uF range and do not include the smaller MW band picofarad world. Also they seem to be aimed at electrolyte caps. Still, this should stimulate the gray matter and suggest possible attack angles on the problem for crystal radio.

Next I address the question of wire resistance, primarily with a couple handy tables of wire data. High-End coils are wound from litz wire and so this was a partictular focus as I do not know much about this type of wire. A good page on litz wire resistance and data required to make your own calculations should help greatly, followed by papers on general litz considerations. With this edition I hope I will have covered most technical/theoritical aspects of crystal radio and can get back to the happy job of design and construction, this time with a more thorough understanding of what the heck I am up to.

Of general note, the web is a marvelous source of data and information. Many long-time crystal set builders, and many others have created dedicated sites to disseminate information and resources, to share their creations and knowledge. I am eternally in your debt. All of the material in this handbook is copyright for which I have not sought permission. Therefore this is not presented for publication or copy. It is only my personal resource. I encourage anyone finding this copy to pursue ON THE WEB the web pages identified within. I include the name of the author and web address of each section. I wish to sincerely thank every author presented for their excellent pages and ask forgiveness for my editing into this handbook.

Kevin Smith
2012

www.lessmiths.com/~kjsmith/crystal/cr0intro.shtml

to adopt wires of Litz construction to reduce the change in internal inductance and resistance as a function as frequency. However it would still probably be good advice to ensure that each wire bundle was of reasonably large diameter, and the bundles were closely spaced.

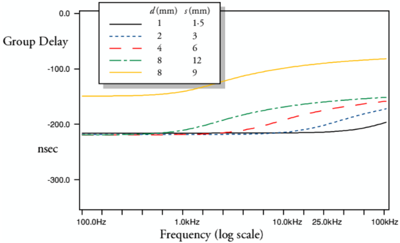

Figure 14 Delays as affected by wire diameter

Figure 14 plots the group delays for the same cables as were used as examples for Figure 13. Looking at these it can be seen that for the examples which share the same value the smallest diameter tends to push the variation in delay as a function of frequency up into the ultrasonic range. Hence a small diameter might be desired if this result is wished. However, by using the thickest wires with a closer spacing we obtain the result shown with the orange line. This reduces the overall level of delay.

In practice it is questionable whether delays of the magnitudes shown would ever be audible. If so, the general advice would seem to be to choose reasonable large diameter wires with a close spacing in order to minimise the effects of resistance and inductance. Since the characteristic impedance of the cables tend to be low compared to those of a typical load the cable series resistance and inductance seem to have far more effect than cable capacitance. This being so, it is the cable resistance and inductance which we should regard as the primary threat to reliable signal transfer. If the change in group delay over the audio band is felt to be significant, a suitable remedy would be

---

CONTENTS:

2 **Capacitor Theory**
University of Waterloo

6 **Basic Capacitor Formulas**
AVX

8 **The best variable capacitor for crystal radio**
Dave Schmarder

18 **Types of capacitor**
Wikipedia

52 **Sensitivity and selectivity issues**
Ben Tongue

74 **Capacitor Losses**
Conrad Hoffman

80 **Lossy Capacitors**
Dr. Gary L. Johnson

100 **ESR of Capacitors**
QuadTech

108 **High Frequency Measurements Web Page**
Douglas C. Smith

116 **Measuring Capacitor ESR**
Geoff

122 **Capacitor Testing Challenges and Solutions**
Dale Cigoy

132 **Measuring Loss in Variable Air Capacitors**
Alan Payne

158 **Capacitor ESR - Methods of Measurement**
Doug Jones

166 **Crystal Radio Capacitors and ESR**
Kevin Smith

174 **Litz Designing Calculations**
HM Wire International, Inc.

178 **American Wire Gages (AWG)**
Conrad Hoffman

180 **Stranded Wire Data Chart**
Dr. Gary L. Johnson

184 **Litz Coil wire Resistance Calculation**
Kevin Smith

186 **Skin Effect, Proximity Effect, and Litz Wir**
Charles R. Sullivan

192 **Skin effect and Litz wire**
Kouichi Hirabayashi

196 **Skin Effect, Internal Impedence**
Jim Lesurf

222 **Notes**

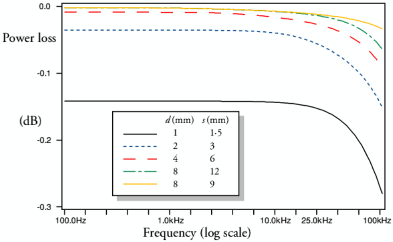

Figure 13 Signal loss as a function of wire diameter

Having examined the effect of varying the wire spacing we can now look at the effect of the choice of wire diameters. Figure 13 shows the power loss as a function of frequency for our standard twin feeder, but using various wire diameters and spacings. Four of the plotted lines have the same wire diameter (d) to spacing (s) ratio. This is to ensure that they all actually have the same characteristic impedance and hence the same nominal inductance and capacitance per length. The variations between these four examples are therefore only due to changes in the internal resistance and inductance of the wires.

Comparing the four plots which share the same value we can see that the lowest losses occur when the wire diameter (and hence spacing) is largest. This seems to clash with the desire to choose a low wire spacing. However the final plot (shown as a solid orange line) is for the largest wire diameter used for the previous four examples, but with a much smaller spacing. This reduces the inductance and hence reduces the level of high frequency power loss.

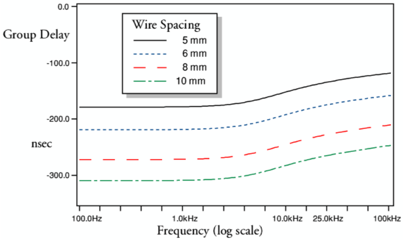

Figure 12 Delays as affected by wire spacing

Figure 12 shows the time delay plots for the same set of spacings, etc, as used for Figure 11. As we would expect, as the spacing is increased and the cable inductance rises, the delay also rises. From the results shown in Figures 11 and 12 we can reasonably expect that we can generally hope to reduce the high frequency losses and overall group delay by choosing a small wire spacing and minimising the cable inductance.

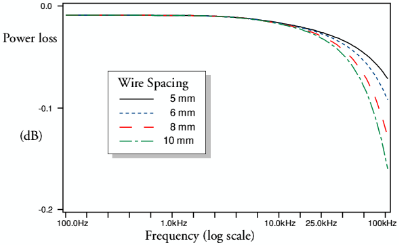

Figure 11 Signal loss as a function of wire spacing

Figure 11 shows the effect of various choices of wire spacing, with a common wire diameter of 4mm (i.e. as previously, a wire radius of 2mm). Note that this has a different vertical scale to earlier power/frequency plots to accommodate the greater losses produced by the wider-spaced cables. Looking at the results we can see that the effect of moving the wires apart is to increase the relative loss at high frequencies. This is simply a consequence of the increase in wire spacing producing a corresponding rise in the cable's inductance per unit length. The combination of the cable inductance and the load resistance act as a conventional low-pass filter. Since the wire diameter is the same for all the lines shown in figure 11 they all have the same value of wire internal impedance.

Secondly, the above assumes we have maintained the diameter of the wire when changing to a lower conductivity material. As we can see from figure 9 this would produce quite high signal losses, and would make the response more sensitive to any frequency dependence of the load impedance.

Thirdly, we can interpret the reduction in the dependence of the delay upon frequency as being due to the relatively high cable resistance ‘swamping’ the inductance. i.e. the wire resistance is now so high that the cable/load system begins to look more like a simple resistive potential divider with a relatively negligable inductance.

In practice we would probably choose to increase the wire diameter to keep down the overall level of wire resistance. This increase in wire diameter would also increase the value at any frequency, thus lowering the frequency at which phase changes may occur.

7. Effects of wire diameter and spacing

For all the examples considered so far we have used the same values of wire diameter and spacing. This has allowed us to compare various choices of conductor, etc, on a ‘like for like’ basis. In practice, however, one of the most common types of variation from cable to cable is in the choice of the diameters of the wires, and the spacing between them. We can now use a convenient standard material (Copper) and explore the effects of varying the spacing and diameter of the wires. As before we will assume a 3 metre length of twin-feeder, and an 8 Ohm resistive load as the standard for the purposes of comparisons.

Capacitor Theory

https://ece.uwaterloo.ca/~lab100/ls1notes.pdf

Any arrangement of two conductors separated by an electric insulator (i.e., dielectric) is a capacitor. An electric charge deposited on one of the conductors induces an equal charge of opposite polarity on the other conductor. As a result, an electric field exists between the two conductor surfaces and there is a potential difference between them. The electric field anywhere between the conductor surfaces is directly proportional to the magnitude of the charge Q on the conductors. And the potential difference V is also directly proportional to the charge Q. The ratio Q/V is thus a constant for any electric field distribution as determined by the shape of the conductors, the distance of separation, and the dielectric in which the field exists.

The ratio Q/V is called the capacitance, C, of a particular arrangement of conductors and dielectric. Thus, C = Q/V, where Q and V are in units of coulomb and volt. C has the units farad (F).

$$C = \frac{\varepsilon_r \quad \varepsilon_o \quad A}{d} \qquad (F) \qquad (farads)$$

The simple theoretical expression for the capacitance value of a parallel plate capacitor is where

A = plate area [m2] = cross section of electric field,

d = distance between plates [m],

Eo = permittivity of free space = 8.854 x 10-12 F/m and

Er = relative permittivity of the dielectric between the plates [dimension less].

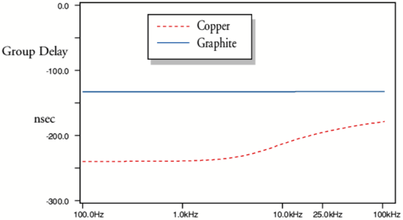

Figure 10 Delays of Graphite and Copper

Figure 10 shows the phase delay behaviour of graphite and copper. In this case the behaviours of the two materials are clearly very different. At low frequencies the inductance of the wires is not dependent upon the resistance level. Hence we do not see the ‘scaling’ effect we encountered when looking at signal losses which there tends to offset the reduction in at any frequency produced by the lower conductivity. From figure 10 therefore we can say that the variation in phase delay with frequency in the audio band is much smaller with graphite than with copper. The graphite also shows a lower overall delay as well as better uniformity. Hence low conductivity materials may be useful if we wish to reduce differential phase/time effects.

However we must draw such a deduction with care for three reasons.

Firstly, the phase changes at audio frequency are very small, even using copper. It is therefore questionable that they would be audible.

$$\frac{100R_0}{100f_0} = \frac{R_0}{f_0}$$

i.e the same in both cases!

This example is not exact and the behaviour is more complex than described here. However this example shows why it turns out that both materials show a noticeable change in signal loss at much the same frequency. Hence for a given wire diameter changing the choice of material has much less effect upon the frequency dependence of the signal loss than we might expect. In fact, we have already seen this to a lesser extent. If we examine the earlier results for fill factor we can see that the main effect of a fill factor of less than unity is to displace the loss curve to a higher or lower overall loss, not to changes its shape.

Using the same wire diameter, etc, as before, do see a marked increase in the low frequency loss due to the lower conductivity. This is simply due to the wires having a much higher d.c. resistance when we change to a material like graphite which has a relatively low conductivity. In practice the level of loss in the example chosen here (around -5 dB) would be far too high for a loudspeaker cable. When using such a lower conductivity material we would therefore probably choose to employ wires of much greater diameter (and hence cross-sectional area) to ensure a low level of loss. However it should be noted that this will mean a change in the value at any given frequency and alter the cable’s behaviour in other ways.

This calculated value is based on the assumption that the charge density on the plates is uniformly distributed. In practice there is always a concentration of charge along the edges. This charge concentration is at the sharp corners of the plates. Thus for a given voltage, the actual total charge is always greater than the theoretical total charge.

Practical Capacitor Model

The low frequency lossy model is introduced to allow you to measure capacitance with a
capacitance measuring bridge instrument.

Discrete component measurements and models of the physical capacitor need to be considered here. Consideration of a finite subdivision of the physical device may assist in arriving at a lumped discrete equivalent circuit model of the practical physical device (ie: a capacitor).

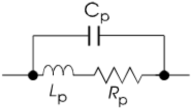

**Figure 1.
Capacitor Model
With Parasitic L & R**

First consider the physical device, the capacitor. We have two conducting surfaces separated by some distance. Hence a defined capacitor . Assume there is a voltage applied to the capacitor. The dielectric has some loss1 and will conduct a small urrent. One can consider this as a distributed loss within the length of the dielectric material between the two conductors. Hence a resistance in parallel with the capacitor.The lossy dielectric has length. Hence one could conceive a distributed inductance2 in series with the distributed resistance. Hence the inductor as shown in figure 1.

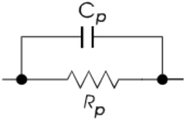

Figure 2. Capacitor Model With R Only

At very low frequencies (say 1 kHz) the parasitic inductance can be ignored3 for the parallel plate capacitor you will be using. Hence the model will only have the capacitor and resistor in parallel as seen in figure 2.

In this experiment the parallel model resistance is very large. Hence the dissipation of energy is small. The resistor will dissipate very little energy. Thus we will have a capacitor with a low dissipation factor, D. One can calculate the resistance value knowing the value of D measured after balancing the bridge.

The capacitance measuring bridge has to be adjusted so that both the capacitance and the resistance of the practical model are balanced. This is the reason for adjusting both the CRL and the DQ dials of the Capacitance Measuring Bridge.

1) The dielectric has a finite resistivity, so when an electric field is across it, a small but finite electric current flows through it. This current, or charge flow, is referred to as a loss (from what an ideal device would do).
2) Because the plates of the capacitor are conductors with current flowing through them, they have a small amount of self inductance.

---

around 0·05dB greater than the loss at d.c. This is lower than the the drop of around 0·1dB we would obtain using materials like copper, but the improvement is surprisingly small given the factor of 1000 in conductivity.

To understand this behaviour we can consider an example where we have two rods of material which have the same diameters and lengths. One material has a conductivity which is 100 times lower than the other. As a result the end-to-end d.c. resistances of the rods also differ by a factor of 100. The frequencies at which the a.c. resistances of the rods double due to the effects of finite skin depth (internal impedance) also differ by a factor of 100. We can now define a specific value of the ratio of skin depth to diameter (and hence) which is the value for which the a.c. resistance is double the d.c. resistance. For the rod of lower conductivity material this value of occurs at a frequency 100 times higher than for the rod of higher conductivity material.

As a result of the above we would expect the effect of skin effect to be shifted up in frequency by a factor of 100 when we change from the higher to the lower conductivity rod. However, we must then also take into account another difference which turns out to be significant. From the above we can say that the higher conductivity rod can be said to have a d.c. resistance, , which increases to at a frequency we can label . The lower conductivity rod will have a d.c. resistance of which rises to at a frequency . Another way of looking at what happens is therefore that the higher conductivity rod changes its resistance by an amount when we change the frequency by an amount . Whereas the lower conductivity rod changes resistance by when we change the frequency by . This means that in each case the overall rate of change of resistance as we change the frequency will be

6. Effect of low conductivity material

Having noticed the effect of the relatively low conductivity of aluminium some manufacturers and writers have argued that it would be better to deliberately use a material with a particularly low conductivity. Carbon (graphite) has frequently been mentioned in this context. To estimate the effect of using such a material we can now use figures 9 and 10. These show the relative loss and phase delay for the two materials and our standard example of cable.

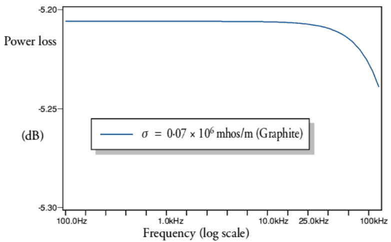

Figure 9 Signal loss for Graphite

Figure 9 shows a result which at first examination seems strange. The graphite has a conductivity around three orders of magnitude lower than that of copper. The skin depth in graphite at any frequency is therefore around 30 times greater. For a given wire diameter this might lead us to expect any increase in loss as the frequency rises to be shifted up to much higher frequencies in graphite than in copper. Yet figure 9 does not seem to show this. The increase in signal loss at 100kHz is

3) The change in current and direction is so slow, the self inductance effect is almost too small to measure.

http://www.avx.com/docs/Catalogs/cbasic.pdf

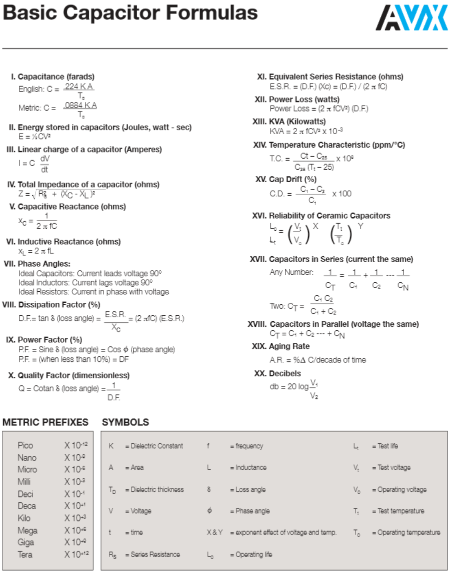

Basic Capacitor Formulas

I. Capacitance (farads)

English: $C = \frac{.224\,K\,A}{T_D}$

Metric: $C = \frac{.0884\,K\,A}{T_D}$

II. Energy stored in capacitors (Joules, watt - sec)

$E = \frac{1}{2}CV^2$

III. Linear charge of a capacitor (Amperes)

$I = C\,\frac{dV}{dt}$

IV. Total Impedance of a capacitor (ohms)

$Z = \sqrt{R_S^2 + (X_C - X_L)^2}$

V. Capacitive Reactance (ohms)

$x_C = \frac{1}{2\,\pi\,fC}$

VI. Inductive Reactance (ohms)

$x_L = 2\,\pi\,fL$

VII. Phase Angles:

Ideal Capacitors: Current leads voltage 90°
Ideal Inductors: Current lags voltage 90°
Ideal Resistors: Current in phase with voltage

VIII. Dissipation Factor (%)

D.F.= tan δ (loss angle) = $\frac{E.S.R.}{X_C}$ = (2 πfC) (E.S.R.)

IX. Power Factor (%)

P.F. = Sine δ (loss angle) = Cos ϕ (phase angle)
P.F. = (when less than 10%) = DF

X. Quality Factor (dimensionless)

Q = Cotan δ (loss angle) = $\frac{1}{D.F.}$

XI. Equivalent Series Resistance (ohms)

E.S.R. = (D.F.) (Xc) = (D.F.) / (2 π fC)

XII. Power Loss (watts)

Power Loss = $(2\,\pi\,fCV^2)$ (D.F.)

XIII. KVA (Kilowatts)

$KVA = 2\,\pi\,fCV^2 \times 10^{-3}$

XIV. Temperature Characteristic (ppm/°C)

$T.C. = \frac{Ct - C_{25}}{C_{25}\,(T_t - 25)} \times 10^6$

XV. Cap Drift (%)

$C.D. = \frac{C_1 - C_2}{C_1} \times 100$

XVI. Reliability of Ceramic Capacitors

$\frac{L_o}{L_t} = \left(\frac{V_t}{V_o}\right)^X \left(\frac{T_t}{T_o}\right)^Y$

XVII. Capacitors in Series (current the same)

Any Number: $\frac{1}{C_T} = \frac{1}{C_1} + \frac{1}{C_2} \cdots \frac{1}{C_N}$

Two: $C_T = \frac{C_1\,C_2}{C_1 + C_2}$

XVIII. Capacitors in Parallel (voltage the same)

$C_T = C_1 + C_2 \cdots + C_N$

XIX. Aging Rate

A.R. = %Δ C/decade of time

XX. Decibels

$db = 20 \log \frac{V_1}{V_2}$

METRIC PREFIXES

Pico	$X\,10^{-12}$
Nano	$X\,10^{-9}$
Micro	$X\,10^{-6}$
Milli	$X\,10^{-3}$
Deci	$X\,10^{-1}$
Deca	$X\,10^{+1}$
Kilo	$X\,10^{+3}$
Mega	$X\,10^{+6}$
Giga	$X\,10^{+9}$
Tera	$X\,10^{+12}$

SYMBOLS

K = Dielectric Constant
A = Area
T_D = Dielectric thickness
V = Voltage
t = time
R_S = Series Resistance

f = frequency
L = Inductance
δ = Loss angle
ϕ = Phase angle
X & Y = exponent effect of voltage and temp.
L_o = Operating life

L_t = Test life
V_t = Test voltage
V_o = Operating voltage
T_t = Test temperature
T_o = Operating temperature

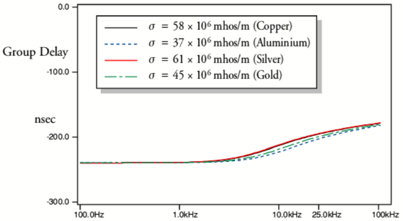

Figure 8 Effect of material upon delay

As with variations in fill factor, the variations between the materials used as examples here seem quite small. Hence it is not obvious that changing, say, from copper wires to silver ones of the same diameter and spacing is likely to produce an audible change as a result of differences in the internal impedance of the wires. Curiously, although aluminium is generally regarded as a 'poor' choice for wiring in audio systems, its relatively low conductivity does mean that the effects of internal impedance tend to occur at slightly higher frequencies than copper.

It is interesting to note that the nominal behaviours of copper and silver are quite similar compared to their difference from aluminium or gold. However it is worth bearing in mind that most of the metals used for wires are far from being of elemental purity. This means that the actual conductivity of a wire may differ by perhaps 5% or more from the standard values assumed for the above calculations. This in itself is unlikely to be significant compared to the much larger range in conductivities for the materials used here as examples.

5. Effect of material conductivity

We can now examine the effect of various choices of conductive material. Figure 7 illustrates the effect of using four different choices of metal for the wires. The same wire diameter and spacing is assumed as previously, and the fill factor is assumed to be unity.

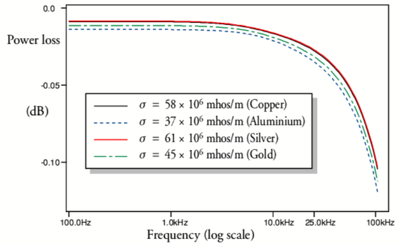

Figure 7 Effect of material upon loss

Figure 8 shows the related results for signal delay values.

What is the best variable capacitor for my crystal radio?

By Dave Schmarder

http://makearadio.com/misc-stuff/capacitors.php

This is a question that has been asked of me many times. It would be easy for me to just say, "Just buy some holy grail capacitors.". Unfortunately, they are not so common anymore, and when you see how much they cost at the large auction venue, you'll forget about the high cost of the litz wire that goes with them.

The purpose of this paper is to discuss various types of variable capacitors, showing their good and bad points. I have come up with some answers that I have not seen discussed elsewhere. Perhaps this work will help you find the right variable capacitor for your radio.

This paper deals nearly exclusively with a crystal radio detector unit in a two section radio. Discussion of an antenna tuner section can be found on my antenna tuner page.

Before we get down to the nitty-gritty brass tacks of the actual capacitors that you should get, let's set the basics for this page. First, anyone that wants to build a dx set should visit the crystal radio DX'ers best friend, Ben Tongue's website. Specifically his paper dealing with variable capacitors as used in crystal sets. In his article 24, section B1, his charts plainly show the difference between a holy grail and an everyday capacitor.

One even wonders how a crystal set works with such a lousy capacitor. Well, it really isn't that bad. But with losses highest at the high end of the band, plus the very squeezed tuning (10% of the capacitor rotation covers 30 to 40% of the

---

Looking at figure 5 we can see that changing the fill factor from being unity down to 0·7 has a surprisingly modest effect upon the signal loss behaviour.

Figure 6 shows the delay behaviour for the same set of fill factor values.

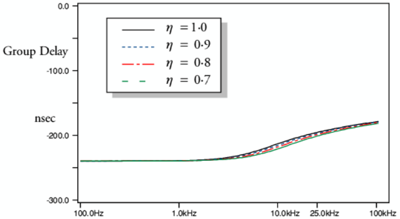

Figure 6 Effect of fill factor upon delay

Again, the changes are relatively modest. On the basis of these results it does not seem as if the fill factor value of stranded wires is likely to have a noticeable effect for twin feeders of this type. That said, it may be more significant for wire diameters and spacing other than those used for this example. More generally, however, it seems that a Litz construction (with the strands individually stranded) is probably usually required for multistranded wires to have audio properties that differ noticeably from solid wires. That said, multistrand is often physically more flexible, and may be easier to connect, so may be more convenient than solid-core for other practical reasons.

Again it is not clear that this would be noticeable in a practical situation.

4. Effects of fill-factor

For the sake of example, the cables considered until now were assumed to be made of copper with an effective fill factor value of unity. To assess the typical effects of a change in fill factor, or of the chosen conductive material, we can recalculate the behaviour of a cable similar to that used before, but with various fill values or materials. As before we therefore assume a 3 metre length of twin-feeder with 2mm diameter wires spaced with their centers 6mm apart.

Figure 5 shows the power loss for four differing values of the fill factor, using copper wires.

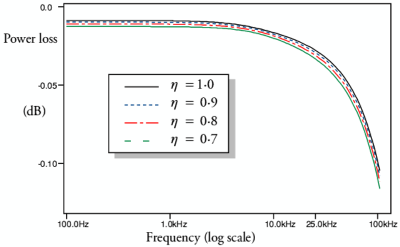

Figure 5 Effect of fill factor upon power loss

---

frequency range), there is a lot of bad karma at the high end of the band.

Below are some examples of capacitor traits. They are not in any particular order of importance or the goodness of one capacitor over another.

A variable capacitor is made up of two major sections, a stator or stationary part, and a rotor or moveable section. These sections are made with plates, also called vanes or blades. The frame is usually connected to the rotor and is one connection of the capacitor. There are some instances where the rotor is isolated from the frame, but not too often.

Dielectric

This is the material that separates the stator plates from the rotor plates. An air dielectric is most common, and is very low loss. Air has a standard dielectric constant of 1.0. Sometimes other materials are placed between the rotor and stator plates, mainly so the plates can be closer together. This allows for a smaller capacitor size, but it is at the expense of higher losses. I like using air capacitors the best.

Insulators

Insulators are the non-conducting material that holds the stator plates to the capacitor frame. The insulator material is considered by most builders the prime criteria in selecting a good capacitor. Two major types of insulating materials used are ceramic and phenolic. Ceramic is consider the very best and the only thing to use in a dx crystal set.

However, new ceramic insulated variable capacitors are expensive. For less aggressive radios, and very aggressive budgets, phenolic is the choice.

Plate Plating

Some of the variable capacitors that you see will have a silver plating. Ben warned us in his article that what appears to be silver plating might not be. The new capacitors sold have bare aluminum plates. They are higher resistance than the silvered plates. Of all the capacitor spec's, this is the least important. But most of the older capacitors that have ceramic insulators will also have a better plating.

Wiper Arms

The wiper arms are those little springy thin pieces of metal that attach to the frame and contact the rotor. Variable capacitors are rated by the number of wipers and how much surface area is in contact. The tightness of the contact should also be considered. Tight is good.

Inexpensive capacitors sometimes have no wiper arms. While they do work, their use in a good crystal set is discouraged. Other cheaper capacitors have just one small wiper at the back frame. The contact area is small, and the contact pressure is fairly light. These are ok for TPT sets. (Radios that are low performance because the builder uses coil forms made with cardboard Toilet Paper Tubes.)

The maximum number of wipers is determined by how many frame surfaces are reachable. In a single gang capacitor, two, one at the front and one at the back is the maximum. In a two gang, it's four (front, back and two in the middle), and so on.

Most hobbyists don't consider the wiper arm quality when they are looking for the capacitor for their dream radio. I consider wipers to be almost as important as the insulator material. I have tested phenolic insulator capacitors that had very nice

---

$$v = \frac{c}{\sqrt{\varepsilon_r}}$$

where is the velocity of light in vacuo, and is the relative permittivity. Here so we would expect a velocity of around 2· 44 × 108 m/s. Hence we might expect a propagation delay when using 3 metre cable of only around 12 - 13 nanoseconds. Yet we see values over ten times greater than this.

The reason for this discrepancy is that microwave/RF engineers usually arrange to match the source and load impedances to the cable’s characteristic impedance. This is far from being the case here. The load resistance is 8 Ohms, but the cable has a characteristic impedance of around 75 Ohms. Since the cable is also very ‘short’ in terms of the signal wavelengths it primarily acts as an inductor and resistor placed in series with the ‘low’ load impedance. At low frequencies the signal delay is therefore approximately equal to which – ignoring internal effects – in the example equals around 164 nanoseconds. Hence the group delays we see are in broad agreement for what we would expect as a result of cable inductance in this example. The dashed line (equivalent to Litz wire) shows a value of 164 nanoseconds as internal inductance has been suppressed. This indicates that although the solid conductor (or a similar stranded wire) have a value of around 240 nanoseconds at low frequencies only around 80 nanoseconds of this is due to the wire’s internal inductance.

Despite the above, to an acoustic engineer the delays seem very small. Since sound propagates at a velocity of around 330 m/sec it will only travel around 16 microns in 50 nanoseconds. In effect, therefore, the above differential group delay is equivalent to a sound source that seems 16 microns nearer at high audible frequencies than at low audible frequencies.

lines we can see that the internal impedance causes the group delay to depart from being uniform. If no ‘skin effect’ were present the group delay would be almost steady at 164 nanoseconds across the whole band plotted. However when the internal impedance of the wires are taken into account the group delay varies from around 240 nanoseconds at low frequency, to around 212 nanoseconds at 10kHz and around 195 nanoseconds at 25 kHz. The effect of the internal impedance is therefore to produce a differential group delay of around 50 nanoseconds across the audio band. Interestingly, high frequencies will tend to arrive slightly earlier than low frequencies.

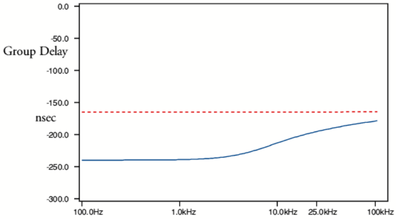

Figure 4 Group delay versus frequency

Curiously, the above time delay values can be seen as ‘large’ or ‘small’ depending upon your starting point. To a microwave/RF engineer with experience of transmission lines a delay of around 200 nanoseconds seems oddly large for a twin feeder only three metres long. Since we are assuming that no magnetic materials are present and a twin feeder propagates a TEM wave we might expect a propagation velocity of

---

wipers and while they didn't test as well as ceramic types, they were still very decent.

Capacitance Minimum to Maximum Ratio

Here is something that almost nobody considers when buying their capacitors, until the builder tries to tune the entire band with a low maximum to minimum capacitance ratio and finds out that it doesn't work. The US-Canadian broadcast band is from 530 to 1710 kHz (while the US is 540 - 1710 and Europe only goes to 1606 kHz). This means that if you are going to tune the entire band, without switching in additional capacitance or with coil taps (yeech), close attention has to be paid to the maximum to minimum capacitance ratio.

An example of a capacitor that will not tune the entire MW band in one swoop is a capacitor that is available on the surplus market. This capacitor has everything that one would want in a capacitor, except a good maximum to minimum capacitance ratio! A picture of it is shown below. Each section measures from about 300 to 35 pF. This is ratio of about 8.6:1. The medium wave band at my location needs a total ratio of at least 20:1 if it is expected to tune without band switching. This includes the distributed capacitance of the coil, and the diode input. At 20:1 there is very little leeway and the coil has to be constructed with extra care. The cheap capacitors fall well within this range, as they have a ratio of about 24:1.

The way you can tell about a capacitors ratio at a glance is to look at the plate structure. If you see a many plates and they are small, this means that the ratio is low. This is because when completely open, the capacitor still has a fair amount of surface between the plates.

Capacitor Frequency Linearity

Those of you that have tuned a radio using a capacitance linear capacitor such as the inexpensive ones mentioned earlier, noticed the top of the band dial squeeze. This is because the capacitance is linear to the rotation. If the capacitor shaft is rotated half way, the capacitance is approximately at half the total value. This means that the center of the dial is at around 800 kHz, about 370 kHz out of the total of nearly 1200 kHz that comprises the medium wave broadcast band. The other rotation half has to cover 3/4 of the band. It gets a lot worse the closer you get to the top of the band.

But what if you could move the capacitor 50% and have half of the band covered? This is what is called a straight line frequency capacitor. In practice having a completely straight line frequency capacitor is nearly impossible, but we can come close.

There are three basic types of capacitor construction. One is where the shaft is centered on the capacitor frame, right to left. The capacitor plates are semicircular. This operates as a straight line frequency capacitor (because if you plot the capacitance vs. the rotation, there is a straight line result).

Another type which was very popular in the old superhet tube radio days was with the shaft and plates offset from the center. These types can also have a modified plate shape, which produces a slight "S" curve when the capacitance vs. rotation is plotted. These, mostly dual gang capacitors are widely sold, but not as popular as the inexpensive single gang linear variable capacitors.

You can expect to pay twice as much for these as the inexpensive capacitors, but most have the advantage of a dual wiper, which lowers the losses. The tuning of this capacitor

Figure 3 shows the power loss in the cable as a function of frequency. The solid (blue) line shows the loss when the wire's internal impedance is taken into account. The broken (red) line shows the loss that would occur if we ignored the internal impedance. Hence the difference between the two shows the results in this case of 'skin effect'. In principle, the broken line shows what we might expect when using wires of 'Litz' construction with a fill factor approaching unity. The solid line shows what we might expect from either a solid wire or a multistranded wire of closely packed thin strands.

For the example chosen, at low frequencies the bulk resistance of the copper wire causes a power loss of around 0·008dB. At 10kHz the loss rises to 0·009dB if the internal impedance were absent, and 0·016dB with the internal impedance taken into account. At 25kHz these values rise to 0·012dB and 0·028dB respectively. Hence the change in relative signal level from near-d.c. to 25kHz, with internal effects taken into account is around 0·020dB. This is quite a small change so it is not obvious that it would be noticed in a practical audio system.

Figure 4 shows the group delay as a function of frequency. This measure is more convenient that plotting any change in phase since a propagation delay will manifest itself as a phase change that rises linearly with frequency. When signals have to propagate through a system the group delay is therefore a better indicator of any time domain effects that might alter the signal pattern. A uniform group delay value implies that all components will arrive with the same relationship as when they were transmitted.

As with figure 3, figure 4 shows two two lines, the solid (blue) one taking internal impedance into account, and the broken (red) line neglecting internal impedance. By examining these

For illustration we can use a case where we have copper cables (assume = 5·8 × 107 mhos/metre) which have wire radii of 2mm with their centers 6·5mm apart. We can also take as an example a twin feeder length of 3 metres. (i.e. 6 metres of wire, in total, in the circuit) and assume a nominal load of 8 Ohms. We can also assume a perfect voltage source and set the source resistance to be 0 Ohms. The cables are assumed to be insulated with a dielectric whose relative permittivity value is 1·5. Most practical dielectrics tend to have a higher value, but the effect is diluted since part of the surrounding E-field is in the air, so this value seems a reasonable one to assume for the purposes of an example.

The results obtainable for this situation are indicated in the following graphs. The dimensions chosen for the wires here are not meant to represent any specific cable, but seem typical for a relatively inexpensive form of wiring.

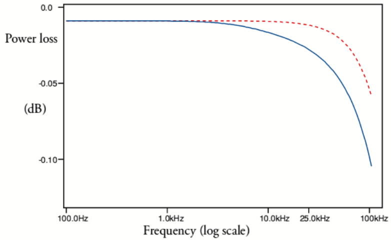

Figure 3 Power loss in transmission versus frequency

---

does help the high end of the band frequency crunch. It is at least tolerable.

The third type is a closer to a straight line frequency capacitor. This has highly shaped plates plus the offset shaft position. There is a much better frequency spread across the dial, especially at the top end of the band.

Insulator Loss vs. Capacitor Plate Loss Ratio

As Ben so clearly shows in his charts, as the capacitance is increased on a particular variable capacitor, the losses go down. In his example of the cheap capacitor, the losses go down by over a factor of 10 when the shaft is rotated to about 1/3 of the way up. The more expensive capacitor also trends up, but at less steep rate. This is because the losses of the dielectric and wiper arms are summed with the losses between the capacitor plates.
High quality capacitor with a lot capacitance minimum to maximum ratio.

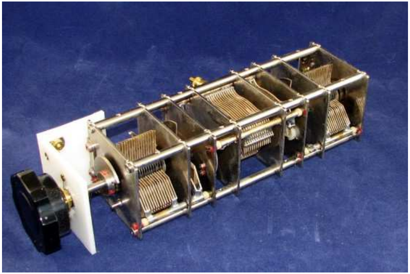

This is the so called "waysalot" capacitor. I used a pair of these in my #75 dx crystal set. It is the best quality capacitor I have. This one has a vernier attached and is used as a test capacitor for determining unloaded Q of the coil under test.

The so called Holy Grail capacitor. The dream of every dx crystal radio builder.

conductor in each triangle is equal to . However the area of each triangle will be . Hence the fill factor will be approximately equal to

$$\frac{\pi}{2\sqrt{3}} \approx 0{\cdot}9069...$$

Real strands will tend to deform slightly when compressed and may not have perfectly smooth, circular cross sections. They may also not be perfectly packed. Hence we can expect the actual value of the fill factor to vary accordingly in practice. In Litz bundles the insulation layers on the strands will also move the conductors apart by a small amount, thus reducing the fill factor value. Usually, we can expect this effect to be small as the layer of insulation is likely to be very thin. In most cases we can therefore tend to assume that the fill factor is reasonably close to unity so this is a reasonable assumption for general analysis. That said, in some specific cases we can take stranding into account by modifying the effective conductivity by an appropriate fill factor value.

3. Twin Feeder Loudspeaker Connections – an example

These internal values can now be used when we wish to evaluate the effect of a cable upon the transmitted signals. We can treat the connecting cables as a form of transmission line, and modify the inductance and resistance per unit length to take the internal values into account. At audio frequencies the cables are normally many orders of magnitude shorter than the highest signal wavelength so it is reasonable to then use lumped values which just multiply the per-unit-length value by the chosen length. For the sake of example here we will consider loudspeaker cables nominally in the form of ‘twin feeder’. For a given cable length this means we must remember to double the values for internal resistance and inductance as a pair of wires are used.

illustrated in figure 2. This shows a close-packed array of strands, each of radius .

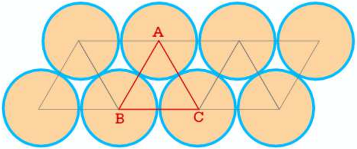

Figure 2 Cross section of a packed bundle of strands.

Since the strands all have a circular cross-section we find that even if they are tightly packed into a hexagonal array there will be spaces in between the places where they touch. Hence a bundle of small packed strands that are in electrical contact can be regarded as similar to a solid but which has some air inclusions which mean that overall wire cross section is only partly filled with conductor. Given a value for the 'fill factor', , (the fraction of the cross section which is filled with conductor) we can treat the wire as being equivalent to a solid conductor of the same wire diameter, , but having an effective conductivity of . Thus the main electronic effect of using a bundle of strands is to dilute the effective conductivity and lower its apparent value.

To estimate the value of the fill factor we can note that the array has a symmetry similar to a packed array of equilateral triangles whose sides all have a length of . Taking one of these elemental triangles we can see that it contains three 60 degree sectors of conductor. Hence the cross sectional area of

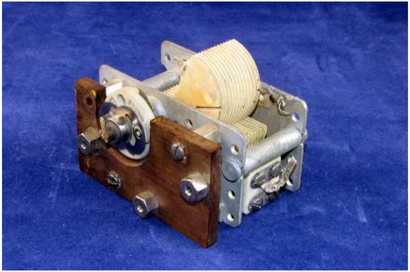

This is known as Holy Grail (everyone should bow and say "I'm not worthy, I'm not worthy"). This capacitor is about 475 pF can was removed from a high quality test instrument. If you can find one for less than seventy-five bucks, grab it.
Examples of capacitor styles with special attention to the plate shape.

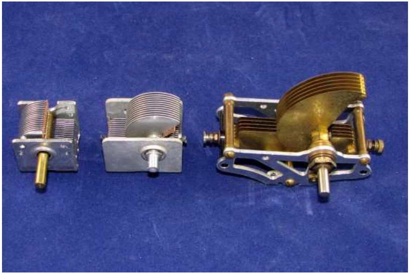

Here are examples of three different variable capacitors frequency vs. capacitance plots. The one on the left is the inexpensive ones available everywhere. The shaft is centered left to right and the plates have a semicircular shape.

The middle one also has the semicircular shape, but the shaft is offset from the center of the frame and the plates. This offers better frequency linearity than the capacitor on the left.

Finally the capacitor on the right is a great example of a straight line frequency plot. The shaft is offset, and the plates are very long. It is only near the bottom of the tuning range is where a lot of capacitance is added with little turning of the shaft. This allows for a more even frequency spread across the tuning range.

Litz wire (as stranded but with insulation between the individual strands)

For solid-core wires the above analysis can be used immediately to compute the internal impedance and deduce the effects it may have in a given situation. For the stranded and Litz wires we need to take the stranding and the effect of inter-strand contacts or insulation into account.

Litz wires consist of a bundle of very thin, individually insulated conductors. The insulation ensures that the current flows in all of the wires in the bundle as the charge cannot migrate towards the surface of the bundle. The entire cross section of conductor bundle is therefore used by the charge transport. Provided that the individual strands are thin enough, the strands all have individual radii that are small compared to the skin depth at audio frequencies. Hence the overall properties of the Litz bundle tends to be similar to that of a single wire of the same diameter of the bundle but where ‘skin effect’ is apparently absent.

In practice, most of the multi-strand wires used for audio purposes have no insulation on the individual strands. This means they do not behave like a Litz wire. In stranded wires without insulation between the individual strands charge may cross from strand to strand. Hence current will tend to preferentially flow near the skin of the bundle of wires, just as it does with a single solid conductor of similar overall diameter. Hence when the strands are thin but in electrical contact with their neighbours we can expect the effect of internal impedance to be similar to that of a solid wire of a diameter similar to the bundle of strands. There is, however, one factor we should take into account. This arises when the strands do not ‘fill’ the bundle and their are air gaps. This is

The standard ‘h.f.’ approximation is that both and will be essentially equal to . However by inspecting the results shown in figure 1 we can see that it is possible to do better than this and a more reliable approximation would be

$$\frac{R}{R_0} \approx \frac{q}{2\sqrt{2}} + 0{\cdot}26 = \frac{r_0}{2\delta} + 0{\cdot}26 \qquad (9)$$

$$\frac{\omega L}{R_0} \approx \frac{q}{2\sqrt{2}} - 0{\cdot}02 = \frac{r_0}{2\delta} - 0{\cdot}02 \qquad (10)$$

This approximations are show in figure 1 by the broken lines. The approximation for is shown by a short-dashed red line, and that for by a longer-dashed green line. It can be seen that these approximations are likely to be reasonably accurate in the region .

2. Types of Wire

In general, electronic signals are conveyed using a ‘pair of wires’. These are used to form a closed loop (path) between the signal source and the load around which charge may flow. Broadly speaking we can then define a ‘cable’ to consist of a pair of wires. The most common forms of cable used in audio are ‘Twin Feeder’ and ‘Co-ax’. The basic properties of these are discussed elsewhere. Each of the wires may be a single, solid, length of conductor. More usually, however, each wire will consist of a bundle of conducting ‘strands’. Multistrand wires have properties which may differ from that of a single, solid wire of similar cross-section. We can therefore treat wires as falling into three general categories as outlined below.

Solid core (i.e. just one strand of conductor per wire)
Stranded wire (a bundle of thin strands of conductors)

Types of capacitor

From Wikipedia

http://en.wikipedia.org/wiki/Types_of_capacitor#Ohmic_losses.2C_ESR.2C_dissipation_factor.2C_and_quality_factor

Theory of conventional construction

Dielectric is placed between two conducting plates (electrodes), each of area A and with a separation of d

In a conventional capacitor, the electric energy is stored statically by charge separation, typically electrons, in an electric field between two electrode plates. The amount of charge stored per unit voltage is essentially a function of the size, the distance and the plates' material properties and the material, the dielectric, placed between the electrodes, while the potential between the plates is limited by the Dielectric breakdown field strength .

Nearly all conventional industrial capacitors except some special styles such as "feed-through capacitors", are constructed as "plate capacitors" even if their electrodes and the dielectric between are wound or rolled to a winding. The capacitance formula for plate capacitors is:

$$C = \frac{\varepsilon A}{d}.$$

The capacitance C for conventional capacitors increases with area A of the electrodes and with permittivity ε of the dielectric material and decreases with the distance d. The capacitance is therefore greatest in devices made from

---

results of doing this are illustrated in figure 1. These are plotted versus so that the relevant nominal skin thickness is also normalised in terms of the wire radius. The solid lines plotted show the relevant values calculated from the above expressions.

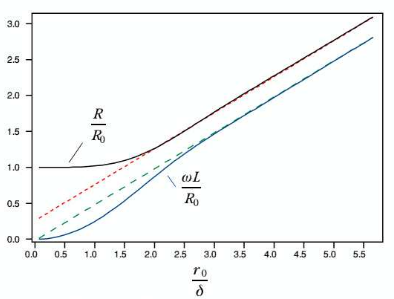

Figure 1 Wire impedance versus skin depth

Unfortunately, the expressions provided by A & S only cover the range for which roughly corresponds to . Above this value the Bessel functions become hard to evaluate and their combination tends to lead to a situation where a set of large values cancel to give a moderate result. Hence for computational simplicity we can use a simpler approximation for the situations where is 'large'. Here this may be defined as where this ratio has a value greater than 5.

The detailed analysis in Ramo leads to the following expressions which can be used to determine the relevant wire resistance and inductance per unit length for a conductive wire of circular cross section at frequencies above d.c.

$$R = \frac{R_S}{\sqrt{2}\,\pi r_0}\left[\frac{\mathrm{Ber}\{q\}\,\mathrm{Bei}'\{q\} - \mathrm{Bei}\{q\}\,\mathrm{Ber}'\{q\}}{(\mathrm{Ber}'\{q\})^2 + (\mathrm{Bei}'\{q\})^2}\right] \qquad (3)$$

$$\omega L = \frac{R_S}{\sqrt{2}\,\pi r_0}\left[\frac{\mathrm{Ber}\{q\}\,\mathrm{Ber}'\{q\} + \mathrm{Bei}\{q\}\,\mathrm{Bei}'\{q\}}{(\mathrm{Ber}'\{q\})^2 + (\mathrm{Bei}'\{q\})^2}\right] \qquad (4)$$

where

$$R_s \equiv \frac{1}{\sigma\delta} \quad ; \quad q \equiv \frac{\sqrt{2}\,r_0}{\delta} \qquad (5,6)$$

and is the ‘skin depth’ value which may be calculated via

$$\delta \equiv \frac{1}{\sqrt{\pi f \mu \sigma}} \qquad (7)$$

where is the signal frequency (as opposed to the signal’s angular frequency).

In fact, using the above expressions we can show that

$$\frac{R_s}{\sqrt{2}\,\pi r_0} \equiv \frac{R_0 q}{2} \qquad (8)$$

and specifying this factor in terms of and may be more convenient when performing calculations.

Now, , etc are Bessel functions. We can find numerical expressions for evaluating these in a text like Abramowitz and Stegun. Using these we can compute values. For the sake of clarity is is useful to plot values normalised in terms of . Some

materials with a high permittivity, large plate area and small distance between plates.

Theory of electrochemical construction

Besides the conventional static storage of electric energy in an electric field, two other storage principles to store electric energy in a capacitor exist. They are so-called electrochemical capacitors.in In contrast to ceramic, film and electrolytic capacitors, supercapacitors, also known as electrical double-layer capacitors (EDLC) or ultracapacitors do not have a conventional dielectric. The capacitance value of an electrochemical capacitor is determined by two high-capacity storage principles. These principles are:

* electrostatic storage within Helmholtz double layers achieved on the phase interface between the surface of the electrodes and the electrolyte (double-layer capacitance) and the

* electrochemical storage achieved by a faradaic electron charge-transfer by specifically adsorpted ions with redox reactions (pseudocapacitance)

Double-layer capacitance and pseudocapacitance combine to a form a supercapacitor.

Common capacitors and their names

Capacitors are divided into two mechanical groups: Fixed capacitors with fixed capacitance values and variable capacitors with variable (trimmer) or adjustable (tunable) capacitance values.

The most important group is the fixed capacitors. Many got their names from the dielectric. For a systematic classification these characteristics can't be used, because one of the oldest, the electrolytic capacitor, is named instead by its cathode construction. So the most-used names are simply historical.

The most common kinds of capacitors are:

* Ceramic capacitors have a ceramic dialectric.

* Film and paper capacitors are named for their dielectrics.

* Aluminum, tantalum and niobium electrolytic capacitors are named after the material used as the anode and the construction of the cathode

* Supercapacitor is the family name for:
  * Double-layer capacitors were named for the physical phenomenon of the Helmholtz double-layer
  * Pseudocapacitors were named for their ability to store electric energy electro-chemically with reversible faradaic charge-transfer
  * Hybrid capacitors combine double-layer and pseudocapacitors to increase power density

* Seldom-used Silver mica, glass, silicon, air-gap and vacuum capacitors were named for their dielectric.

Capacitors in each family have similar physical design features, but vary, for example, in the form of the terminals.

misunderstood, and hence people occasionally tend to invoke this frequency dependent behaviour as the implied basis for all kinds of claims regarding the ‘sounds’ of different types of cables. The purpose of this analysis is to throw some light into this area and help provide some understanding of the effects of using conductors of finite conductivity.

In engineering textbooks, the consequences of finite conductivity and wire size are treated in terms of an ‘Internal Impedance’. This term is probably more useful that ‘skin effect’ as it acts as a reminder that the effects arise due to the fields internal to the conductor. The internal impedance per unit length of a wire is considered in Ramo et al. From this was may draw the following results as a starting point.

The d.c. (i.e. very low frequency) impedance of a wire which has a circular cross-section and is uniform may be said to consist of a resistance per unit length of

$$R_0 = \frac{1}{\pi r_0^2 \sigma} \quad (1)$$

where is the bulk conductivity value appropriate for the material used to manufacture the wire, and is the radius of the wire. The resistance is in Ohms/metre if we are using S.I. units (which will be assumed from now on).

The wire will also exhibit an effective inductance per unit length at very low frequency due to its internal fields. At very low frequencies this has the value

$$L_0 = \frac{\mu}{8\pi} \quad (2)$$

where is the permeability of the material. In general we can assume that this equals the value for free space

$$\mu_0 = 4\pi \times 10^{-7} \quad \text{Henries/Metre}$$

Skin Effect, Internal Impedence, and types of wire

Jim Lesurf
University of St. Andrews, St Andrews, Fife KY16 9SS, Scotland
http://www.st-andrews.ac.uk/~www_pa/Scots_Guide/audio/skineffect/page1.html

1. Skin Effect and cable impedance

There is a considerable amount of discussion (not to say argument!) amongst those interested in audio and Hi-Fi about the possible effects of cables upon ‘sound quality’. This tends to lead to some people adopting almost ‘theological’ viewpoints that differ fundamentally from the views held by others. One of the technical factors which is sometimes claimed to affect sound quality is what is usually called ‘Skin effect’.

In most electronics textbooks, the properties of cables and wires are considered as a form of transmission line. The text may mention briefly the skin effect without exploring this in detail. More often, however, the only parameters that tend to be considered are the capacitance per length, inductance per length, and their relationship with the signal’s nominal propagation velocity and the characteristic impedance of the system. The fact that normal materials have a finite conductance (or resistance) is not usually considered beyond its effect upon the d.c. (and low frequency) resistance of the cables and the resulting implied signal power losses.

In reality, when we transmit alternating signals along conductive lines we may experience effects due to what is generally called ‘skin effects’. This subject is widely

---

In addition to the above shown capacitor types, which derived their name from historical development, there are many individual capacitors that have been named based on their application. They include:

* Power capacitors, motor capacitors, DC-link capacitors, suppression capacitors, audio crossover capacitors, lighting ballast capacitors, snubber capacitors, coupling, decoupling or bypassing capacitors.

Often, more than one capacitor family is employed for these applications, e.g. interference suppression can use ceramic capacitors or film capacitors.

Specialized devices such as built-in capacitors with metal conductive areas in different layers of a multi-layer printed circuit board and kludges such as twisting together two pieces of insulated wire also exist.

Dielectrics

The most common dielectrics are:

* Ceramics
* Plastic films
* Oxide layer on metal (Aluminum, Tantalum, Niobium)
* Natural materials like mica, glass, paper, air, vacuum

All of them store their electrical charge statically within an electric field between two (parallel) electrodes.

Beneath this conventional capacitors a family of electrochemical capacitors called Supercapacitors was developed. Supercapacitors don't have a conventional dielectric. They store their electrical charge statically in

* Helmholtz double-layers (Double-layer capacitors)

and additional electrochemical with faradaic charge transfer

* with a pseudocapacitance (Pseudocapacitors)
* or with both storage principles together (Hybrid capacitors).

The most important material parameters of the different dielectrics used and the appr. Helmholtz-layer thickness are given in the table below.

Table of Key parameters

I0(z) = Modified Bessel Function of First kind of order 0
I1(z) = Modified Bessel Function of First kind of order 1

The first mathematical discussions on the skin effect of wires were done by Maxwell, who started modern electromagnetics in 1873. This was followed by significant contributions by the creative genius Heaviside, who followed the theoretical work of Maxwell in 1884-1887, and by Poynting in 1884-1885. Experimental confirmation was first done by Hughes in 1886. Numerical engineering calculations on the applicable level was first done by Lord Rayleigh for the skin effect on infinite planes in 1886, and on cylindrical conductors by Lord Kelvin using ber-bei functions in 1899. Afterwards, there was work by Hertz and J.J. Thomson, who took over experimental work by Maxwell. The first person to use the word "skin effect" is considered to be J.Swinburne in 1891. Until recently, there were continuous challenges by many brilliant theorists and engineers, including H.B.Dwight and A.E.Kennelly.

For your information, the word "Litz wire" comes from the German word "Litzendraht" (twisted strands).

Kouichi Hirabayashi, (C) 2003

"if we increase the number of wires, we can make a cable with less skin effect. If company A has a product with 19 wires, we make one with 30 wires. If they beat us, we will increase the number of wires by one...".

As a consequence, many audio cables with Litz wire were born. Is this strategy effective?

This is the question.

Footnote 1

Obtaining the current distribution of a single cylindrical conductor under high frequency is relatively simple, and the solution is shown below. The current distribution shown in Fig. 1 has been calculated using this equation. However, only the magnitude is shown. Note that the phase of the current also depends on the position. As we get deeper into the conductor from the surface, the phase of the current delays and may even become backward!

Ir / Ia = I0(sqrt(j)*k*r) / I0(sqrt(j)*k*a)
Here,
Ir = current density (A/m^2) at radius r (m)
Ia = current density at outer surface of conductor (A/m^2)
a = conductor radius (m)
j = sqrt(-1)
k = sqrt(w*u*g)
w = angular velocity (rad/s)
= 2*PI*f
PI = 3.14159265..
f = frequency (Hz)
u = permeability (H/m)
g = dielectric constant (S/m)

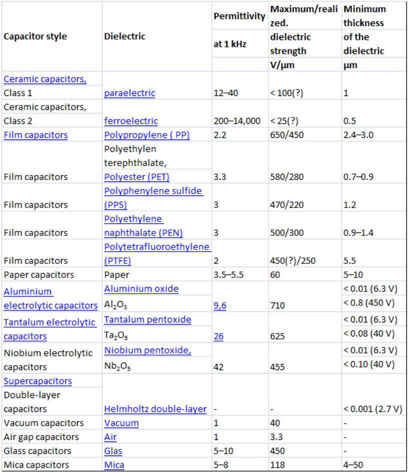

Capacitor style	Dielectric	Permittivity at 1 kHz	Maximum/realized. dielectric strength V/µm	Minimum thickness of the dielectric µm
Ceramic capacitors, Class 1	paraelectric	12–40	< 100(?)	1
Ceramic capacitors, Class 2	ferroelectric	200–14,000	< 25(?)	0.5
Film capacitors	Polypropylene (PP)	2.2	650/450	2.4–3.0
Film capacitors	Polyethylen terephthalate, Polyester (PET)	3.3	580/280	0.7–0.9
Film capacitors	Polyphenylene sulfide (PPS)	3	470/220	1.2
Film capacitors	Polyethylene naphthalate (PEN)	3	500/300	0.9–1.4
Film capacitors	Polytetrafluoroethylene (PTFE)	2	450(?)/250	5.5
Paper capacitors	Paper	3.5–5.5	60	5–10
Aluminium electrolytic capacitors	Aluminium oxide Al_2O_3	9,6	710	< 0.01 (6.3 V) < 0.8 (450 V)
Tantalum electrolytic capacitors	Tantalum pentoxide Ta_2O_5	26	625	< 0.01 (6.3 V) < 0.08 (40 V)
Niobium electrolytic capacitors	Niobium pentoxide, Nb_2O_5	42	455	< 0.01 (6.3 V) < 0.10 (40 V)
Supercapacitors Double-layer capacitors	Helmholtz double-layer	-	-	< 0.001 (2.7 V)
Vacuum capacitors	Vacuum	1	40	-
Air gap capacitors	Air	1	3.3	-
Glass capacitors	Glas	5–10	450	-
Mica capacitors	Mica	5–8	118	4–50

The capacitor's plate area can be adapted to the wanted capacitance value. The permittivity and the dielectric thickness are the determining parameter for capacitors. Ease of processing is also crucial. Thin, mechanically flexible sheets can be wrapped or stacked easily, yielding large designs with high capacitance values. Razor-thin metallized sintered ceramic layers covered with metallized electrodes however, offer the best conditions for the miniaturization of circuits with SMD styles.

A short view to the figures in the table above gives the explanation for some simple facts:

* Supercapacitors have the highest capacitance density because of its special charge storage principles
* Electrolytic capacitors have lesser capacitance density than supercapacitors but the the highest capacitance density of conventional capacitors because its thin dielectric.
* Ceramic capacitors class 2 have much higher capacitance values in a given case than class 1 capacitors because of their much higher permittivity.
* Film capacitors with their different plastic film material do have a small spread in the dimensions for a given capacitance/voltage value of a film capacitor because the minimum dielectric film thickness differs between the different film materials.

Ceramic capacitors

A ceramic capacitor is a non-polarized fixed capacitor made out of two or more alternating layers of ceramic and metal in which the ceramic material acts as the dielectric and the metal acts as the electrodes. The ceramic material is a mixture of finely ground granules of paraelectric or ferroelectric materials, modified by mixed oxides that are necessary to achieve the capacitor's desired characteristics. The electrical behavior of the ceramic material is divided into two stability classes:

The red, purple, green and light green lines correspond to the current density in the conductor cross-section for 10kHz, 100kHz, 1MHz and 10MHz respectively. The current density at the center of the conductor decreases from 98% of the surface for 10kHz, to 41% for 100kHz, to 0.4% for 1MHz, and to just 0.00000006 % for 10 MHz.

As a result of the skin effect, the effective cross-section of the conductor decreases and alternating current resistance increases. Therefore, in coil windings, conductor loss increases and the "Q" (quality factor), or quality, decreases with the increase in frequency.

As a countermeasure, from early times "Litz wire" was used for bar antenna coils in shortwave radios and mid-frequency transformer windings. Litz wire is many thin insulated conductor strands twisted together.

Many people considered applying the concept behind Litz wire to "audio cables", and came up with the idea as shown in Fig. 2.

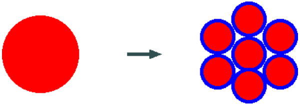

Fig. 2 Splitting a single wire conductor into many insulated conductors (conductors in red, insulators in blue)

Here, we split a single wire conductor into many insulated conductors with the same conductor cross-section area. For example, if we split into seven wires, we get the configuration shown in Fig. 2. People who think of these products think that

Skin effect and Litz wire

Kouichi Hirabayashi
http://www.mogami.com/e/puzzle/pzl-21.html

Current flowing through a cylindrical conductor distributes evenly along the conductor cross-section when the current is a direct current. However, with the increase in frequency, the effect where the current concentrates at the conductor surface and the current density decreases at the center, is known well as the skin effect (footnote 1).

For example, Fig. 1 shows the current density in the conductor cross-section of an annealed copper wire with a diameter of 1mm for frequencies 10kHz, 100kHz, 1MHz and 10MHz.

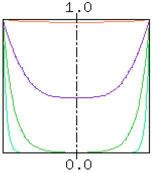

Fig. 1 Skin effect for annealed copper wire with a diameter of 1mm

The x-axis in Fig. 1 is the distance from the conductor center (center of the figure represents the center of the conductor and both ends show the conductor surface), and the y-axis is (internal current density/current density at conductor surface).

---

* Class 1 ceramic capacitors with high stability and low losses compensating the influence of temperature in resonant circuit application. Common EIA/IEC code abbreviations are C0G/NP0, P2G/N150, R2G/N220, U2J/N750 etc.

* Class 2 ceramic capacitors with high volumetric efficiency for buffer, by-pass and coupling applications Common EIA/IEC code abbreviations are: X7R/2XI, Z5U/E26, Y5V/2F4, X7S/2C1, etc.

The great plasticity of ceramic raw material works well for many special applications and enables an enormous diversity of styles, shapes and great dimensional spread of ceramic capacitors. The smallest discrete capacitor, for instance, is a "01005" chip capacitor with the dimension of only 0.4 mm × 0.2 mm.

The construction of ceramic multilayer capacitors with mostly alternating layers results in single capacitors connected in parallel. This configuration increases capacitance and decreases all losses and parasitic inductances. Ceramic capacitors are well-suited for high frequencies and high current pulse loads.

Because the thickness of the ceramic the dielectric layer can be easily controlled and produced by the desired application voltage, ceramic capacitors are available with rated voltages up to the 10 kV range.

Some ceramic capacitors of special shapes and styles are used as capacitors for special applications, including RFI/EMI suppression capacitors for connection to supply mains, also known as safety capacitors,[9][10] X2Y® capacitors for bypassing and decoupling applications,[11] feed-through capacitors for noise suppression by low-pass filters[12] and ceramic power capacitors for transmitters and HF applications.[13][14]

Film capacitors

Film capacitors or plastic film capacitors are non-polarized capacitors with an insulating plastic film as the dielectric. The dielectric films are drawn to a thin layer, provided with metallic electrodes and wound into a cylindrical winding. The electrodes of film capacitors may be metallized aluminum or zinc, applied on one or both sides of the plastic film, resulting in metallized film capacitors or a separate metallic foil overlying the film, called film/foil capacitors.

Metallized film capacitors offer self-healing properties. Dielectric breakdowns or shorts between the electrodes do not destroy the component. The metallized construction makes it possible to produce wound capacitors with larger capacitance values (up to 100 µF and larger) in smaller cases than within film/foil construction.

Film/foil capacitors or metal foil capacitors are use two plastic films as the dielectric. Each film is covered with a thin metal foil, mostly aluminium, to form the electrodes. The advantage of this construction is the ease of connecting the metal foil electrodes, along with an excellent current pulse strength.

A key advantage of every film capacitor's internal construction is direct contact to the electrodes on both ends of the winding. This contact keeps all current paths very short. The design behaves like a large number of individual capacitors connected in parallel, thus reducing the internal ohmic losses (ESR) and

simplicity facilitates finding those optimal diameters. Other models (such as [14] and the similar analysis in [17]) also model large strand diameters and circular bundle configurations accurately, but they fully calculate only internal, and not external, proximity effect, and so are not useful for the present purposes.

CONCLUSION

The number of strands for a minimum-loss litz-wire winding may be found by evaluating the tradeoff between proximity-effect losses and dc resistance. Of the factors leading to increased dc resistance in a litz-wire winding, only the space allocated to strand insulation varies significantly with the number of strands in a well designed construction. A power law can be used to model insulation thickness in the region of interest. Combining this with standard models for eddy-current loss results in an analytic solution for the optimal number of strands. The simplest model for loss, using only the first terms of a series expansion can be used since good designs use strands that are small compared to a skin depth. Experimental results correlate well with the simple model. Stranding for minimum loss may lead to many strands of fine wire and thus excessive expense. Minimum loss designs constrained by minimum strand size or maximum number of strands have also been derived.

where Fr is a factor relating dc resistance to an ac resistance which accounts for all winding losses, given a sinusoidal current with rms amplitude Iac. As shown in Appendix A, we can approximate Fr by

$$F_r = 1 + \frac{\pi^2 \omega^2 \mu_0^2 N^2 n^2 d_c^6 k}{768 \rho_c^2 b_c^2}$$

(2)

where ; is the radian frequency of a sinusoidal current, n is the number of strands, N is the number of turns, dc is the diameter of the copper in each strand, is the resistivity of the copper conductor, bc is the breadth of the window area of the core, and k is a factor accounting for field distribution in multi-winding transformers, normally equal to one (see Appendix A). For waveforms with a dc component, and for some non-sinusoidal waveforms, it is possible to derive a single equivalent frequency that may be used in this analysis (Appendix C). In an inductor, the field in the winding area depends on the gapping configuration, and this analysis is not directly applicable.

The analysis described here considers the strands of all litz bundles to be uniformly distributed in the window, as they would be in a single winding using N turns of wire the diameter of the litz strands. In fact, the strands are arranged in more or less circular bundles. In this sense, the analysis of [15] may be more accurate, but this difference has very little effect on the results. The most important difference between the model used here and the model in [15] is the greater accuracy of [15] for strands that are large compared to a skin depth. The simpler model is used because it is accurate for the small strand diameters that are found to be optimal, and because its

---

parasitic inductance (ESL). The inherent geometry of film capacitor structure results in low ohmic losses and a low parasitic inductance, which makes them suitable for applications with high surge currents (snubbers) and for AC power applications, or for applications at higher frequencies.

The plastic films used as the dielectric for film capacitors are Polypropylene (PP), Polyester (PET), Polyphenylene sulfide (PPS), Polyethylene naphthalate (PEN), and Polytetrafluoroethylene or Teflon (PTFE). Polypropylene film material with a market share of something about 50% and Polyester film with something about 40% are the most used film materials. The rest of something about 10% will be used by all other materials including PPS and paper with roughly 3%, each.[15][16]

Table Characteristics of plastic film materials for film capacitors

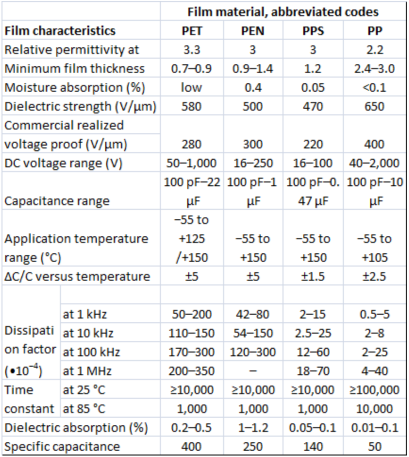

Film characteristics		PET	PEN	PPS	PP
		Film material, abbreviated codes			
Relative permittivity at		3.3	3	3	2.2
Minimum film thickness		0.7–0.9	0.9–1.4	1.2	2.4–3.0
Moisture absorption (%)		low	0.4	0.05	<0.1
Dielectric strength (V/µm)		580	500	470	650
Commercial realized voltage proof (V/µm)		280	300	220	400
DC voltage range (V)		50–1,000	16–250	16–100	40–2,000
Capacitance range		100 pF–22 µF	100 pF–1 µF	100 pF–0.47 µF	100 pF–10 µF
Application temperature range (°C)		−55 to +125 /+150	−55 to +150	−55 to +150	−55 to +105
ΔC/C versus temperature		±5	±5	±1.5	±2.5
Dissipation factor $(\bullet 10^{-4})$	at 1 kHz	50–200	42–80	2–15	0.5–5
	at 10 kHz	110–150	54–150	2.5–25	2–8
	at 100 kHz	170–300	120–300	12–60	2–25
	at 1 MHz	200–350	–	18–70	4–40
Time constant	at 25 °C	≥10,000	≥10,000	≥10,000	≥100,000
	at 85 °C	1,000	1,000	1,000	10,000
Dielectric absorption (%)		0.2–0.5	1–1.2	0.05–0.1	0.01–0.1
Specific capacitance		400	250	140	50

Some film capacitors of special shapes and styles are used as capacitors for special applications, including RFI/EMI suppression capacitors for connection to the supply mains, also known as safety capacitors,[17] Snubber capacitors for very high surge currents,[18] Motor run capacitors, AC capacitors for motor-run applications[19]

(the effect of current in other bundles) [14], [15]. However, the distinction is useful only as a form of bookkeeping. The actual losses in one strand of a litz bundle are simply a result of the total external field, due to the currents in all the other strands present. Another approach to calculating the loss in a litz winding is to look at it as a single winding, made up of nN turns of the strand wire, each with current i/n flowing in it, where n is the number of strands, N is the number of turns of litz wire, and i is current flowing in the overall litz bundle. The loss in the litz winding will be the same as in the equivalent single-strand winding as long as the currents flowing in all the strands are equal [16]. Other methods of calculating loss in litz wire also assume equal current in all strands [14], [12], [17]].

The objective of twisting or weaving litz wire, as opposed to just grouping fine conductors together, is to ensure that the strand currents are equal. Simple twisted bunched-conductor wire can accomplish this adequately in situations where proximity effect would be the only significant problem with solid wire. Where skin effect would also be a problem, more complex litz constructions can be used to ensure equal strand currents. Thus, in a well-designed construction, strand currents are very close to equal. However, our results remain valid even when simple twisting results in significant skin effect at the litz-bundle level. This is because the the bundle-level skin-effect loss is independent of the number of strands, and is orthogonal [15] to the strand-level eddy-current losses.

We represent winding losses by

$$P_{loss} = F_r I_{ac}^2 R_{dc}$$

(1)

All of these methods give similar results for strands that are small compared to the skin depth [12]. (See Appendix B for a discussion of one minor discrepancy.) The solutions for optimal stranding result in strand diameters much smaller than a skin depth. In this region the distinctions between the various methods evaporate, and the simplest analysis is adequate. More rigorous analysis (e.g. [12]) is important when strands are not small compared to a skin depth. In this case, losses are reduced relative to what is predicted by the analysis used here, due to the self-shielding effect of the conductor.

Previous work, such as [2] , [3], [4] has addressed optimal wire diameter for single-strand windings. The approach in [2], [3], [4] is also applicable for litz-wire windings in the case that the number of strands is fixed, and the strand diameter for lowest loss is desired. As discussed in Section V, this can be useful for cost-sensitive applications, if the number of strands is the determining factor in cost, and the maximum cost is constrained. However, this will, in general, lead to higher-loss designs than are possible using the optimal number of strands.

SKIN EFFECT, PROXIMITY EFFECT, AND LITZ WIRE

Skin effect is the tendency for high-frequency currents to flow on the surface of a conductor. Proximity effect is the tendency for current to flow in other undesirable patterns---loops or concentrated distributions---due to the presence of magnetic fields generated by nearby conductors. In transformers and inductors, proximity-effect losses typically dominate over skin-effect losses. In litz-wire windings, proximity effect may be further divided into internal proximity effect (the effect of other currents within the bundle) and external proximity effect

High pulse current load is the most important feature of film capacitors so many of the available styles do have special terminations for high currents

Electrolytic capacitors

Electrolytic capacitors have a metallic anode covered with an oxidized layer used as dielectric. The second electrode is a non-solid (wet) or solid electrolyte. Electrolytic capacitors are polarized. Three families are available, categorized according to their dielectric.

* Aluminum electrolytic capacitors with aluminum oxide as dielectric
* Tantalum electrolytic capacitors with tantalum pentoxide as dielectric
* Niobium electrolytic capacitors with niobium pentoxide as dielectric.

The anode is highly roughened to increase the surface area. This and the relatively high permittivity of the oxide layer gives these capacitors very high capacitance per unit volume compared with film- or ceramic capacitors.

The permittivity of tantalum pentoxide is approximately three times higher than aluminium dioxide, producing significantly smaller components. However, permittivity determines only the dimensions. Electrical parameters, especially conductivity, are established by the electrolyte's material and composition. Three general types of electrolytes are used:

* non solid (wet, liquid)—conductivity approximately 10 mS/cm and are the lowest cost

* solid manganese oxide—conductivity approximately 100 mS/cm offer high quality and stability
* solid conductive polymer (Polypyrrole)—conductivity approximately 10,000 mS/cm,[21] offer ESR values as low as <10 mΩ

Internal losses of electrolytic capacitors, prevailing used for decoupling and buffering applications, are determined by the kind of electrolyte.

Table Some important values of the different electrolytic capacitors

loss. This paper presents a method of finding that optimum, using standard methods of estimating the eddy-current losses.
Optimizations on magnetics design may be done to minimize volume, loss, cost, weight, temperature rise, or some combination of these factors. For example, in the design of magnetic components for a solar-powered race vehicle [1] (the original impetus for this work) an optimal compromise between loss and weight is important. Although we will explicitly minimize only winding loss, the results are compatible with and useful for any minimization of total loss (including core loss), temperature rise, volume or weight. This is because the only design change considered is a change in the degree of stranding, preserving the overall diameter per turn and overall window area usage. This does not affect core loss or volume, and has only a negligible effect on weight. However, the degree of stranding does significantly affect cost. Although we have not attempted to quantify or optimize this, additional results presented in Section V are useful for cost-constrained designs. The analysis of eddy-current losses used here does not differ substantially from previous work [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13] ([10] gives a useful review). Although different descriptions can be used, most calculations are fundamentally equivalent to one of three analyses. The most rigorous approach uses an exact calculation of losses in a cylindrical conductor with a known current, subjected to a uniform external field, combined with an expression for the field as a function of one-dimensional position in the winding area [12]. Perhaps the most commonly cited analysis [11] uses "equivalent" rectangular conductors to approximate round wires, and then proceeds with an exact one-dimensional solution. Finally, one may use only the first terms of a series expansion of these solutions, e.g. [9].

Skin Effect, Proximity Effect, and Litz Wire

Charles R. Sullivan
Thayer School of Engineering, Dartmouth College
http://www.dartmouth.edu/~sullivan/litzwire/litz.html

INTRODUCTION

A salient difficulty in designing high-frequency inductors and transformers is eddy-current effects in windings. These effects include skin-effect losses and proximity-effect losses. Both effects can be controlled by the use of conductors made up of multiple, individually insulated strands, twisted or woven together. Sometimes the term litz wire is reserved for conductors constructed according to a carefully prescribed pattern, and strands simply twisted together are called bunched wire. We will use the term litz wire for any insulated grouped strands, but will discuss the effect of different constructions.

This paper addresses the choice of the degree of stranding in litz wire for a transformer winding. The number of turns and the winding cross-sectional area are assumed to be fixed. The maximum cross-sectional area of each turn is thus fixed, and as the number of strands is increased, the cross-sectional area of each strand must be decreased. This typically leads to a reduction in eddy-current losses. However, as the number of strands increases, the fraction of the window area that is filled with copper decreases and the fraction filled with insulation increases.This results in an increase in dc resistance. Eventually, the eddy-current losses are made small enough that the increasing dc resistance offsets any further improvements in eddy-current loss, and the losses start to increase. Thus, there is an optimal number of strands that results in minimum

---

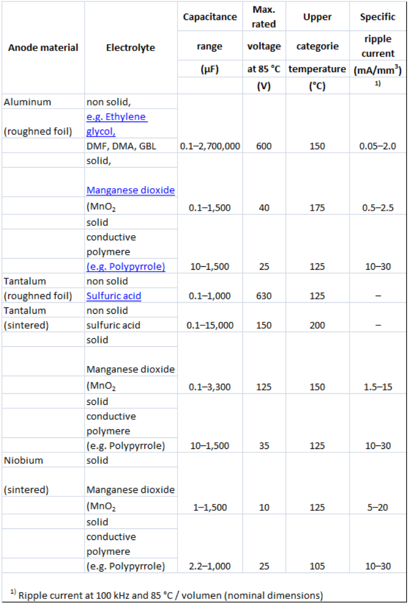

Anode material	Electrolyte	Capacitance range (µF)	Max. rated voltage at 85 °C (V)	Upper categorie temperature (°C)	Specific ripple current (mA/mm^3) 1)
Aluminum (roughned foil)	non solid, e.g. Ethylene glycol, DMF, DMA, GBL	0.1–2,700,000	600	150	0.05–2.0
	solid, Manganese dioxide (MnO$_2$	0.1–1,500	40	175	0.5–2.5
	solid conductive polymere (e.g. Polypyrrole)	10–1,500	25	125	10–30
Tantalum (roughned foil)	non solid Sulfuric acid	0.1–1,000	630	125	–
Tantalum (sintered)	non solid sulfuric acid	0.1–15,000	150	200	–
	solid Manganese dioxide (MnO$_2$	0.1–3,300	125	150	1.5–15
	solid conductive polymere (e.g. Polypyrrole)	10–1,500	35	125	10–30
Niobium (sintered)	solid Manganese dioxide (MnO$_2$	1–1,500	10	125	5–20
	solid conductive polymere (e.g. Polypyrrole)	2.2–1,000	25	105	10–30

1) Ripple current at 100 kHz and 85 °C / volumen (nominal dimensions)

The large capacitance per unit volume of electrolytic capacitors make them valuable in relatively high-current and low-frequency electrical circuits, e.g. in power supply filters for decoupling unwanted AC components from DC power connections or as coupling capacitors in audio amplifiers, for passing or bypassing low-frequency signals and storing large amounts of energy. The relatively high capacitance value of an electrolytic capacitor combined with the very low ESR of the polymer electrolyte of polymer capacitors, especially in SMD styles, makes them a competitor to MLC chip capacitors in personal computer power supplies.

Bipolar electrolytics (also called Non-Polarized capacitors) contain two anodized aluminium foils, behaving like two capacitors connected in series opposition.

Electolytic capacitors for special applications include motor start capacitors,[22] flashlight capacitors[23] and audio frequency capacitors.[24]

Electrical characteristics

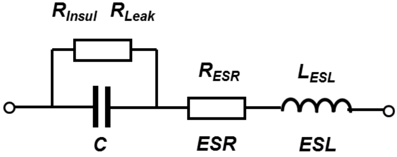

Series-equivalent circuit

Do = Diameter of the finished cable over the strands in inches = 0.056
K = Constant depending on the number of strands = 2

Rac/Rdc = 1.0003 +2(660*0.0016 / 0.056)^2 * (0.0016*1000 / 10.44)^4 = 1.393

Therefore the AC Resistance of 660/46 litz wire is: 1.39 * 7.27 = 10.13 ohms/1000ft.

A 5inch diameter basket weave coil made from 660/46 Litz wire having an inductance of 230 uH will typically require some 40 turns or about 53ft of wire. At 10.13 Ohm/kft, that comes to an AC wire resistance of 0.53 ohms for the coil alone. Were all the losses represented by just the wire, the coil would have, at 1 MHz an unloaded Q = 2700!

Naturally, wire resistance is not the only source of loss in the tank. There is a capacitor, metallic objects intruding into the magnetic field of the coil, diaelectric losses, eddy current and other losses. Its no wonder that the best coils just top out above Q = 1000 or so.

kjs

Litz Coil wire Resistance Calculation

Kevin Smith

From different Litz wire manufacturing sites one can find tables and data allowing the calculation of the resistance of your favorite litz wire.

The formula for the D.C. resistance of any Litz construction is:

Rdc = Rs (1.0515)^Nb * (1.025)^Nc / Ns

Where:
Rdc = Resistance in Ohms/1000 ft.
Rs = Maximum D.C. resistance of the individual strands (4544 for 46awg wire)
Nb = Number of Bunching operations (assume = 2)
Nc = Number of Cabling operations (assume = 1)
Ns = Number of individual strands (assume = 660)

Rdc = 4544 (1.015)^2 (1.025)^1 / 660 = 7.27 ohms / 1000 ft.

The ratio of AC resistance to DC resistance of any Litz construction is:

Rac/Rdc = S + K (N Di / Do)^2 * G

Where:
S = Resistance ratio of individual strands when isolated (1.0003 for 46awg wire)
G = Eddy Current basis factor = (Di * sqrt(f) / 10.44)^4
F = Operating Frequency in HZ (assume 1MHz)
N = Number of strands in the cable = 660
Di = Diameter of the individual strands over the copper in inches = 0.0016

Discrete capacitors deviate from the ideal capacitor. An ideal capacitor only stores and releases electrical energy, with no dissipation. Capacitor components have losses and parasitic inductive parts. These imperfections in material and construction can have positive implications such as linear frequency and temperature behavior in class 1 ceramic capacitors. Conversely, negative implications include the non-linear, voltage-dependent capacitance in class 2 ceramic capacitors or the insufficient dielectric insulation of capacitors leading to leakage currents.

All properties can be defined and specified by a series equivalent circuit composed out of an idealized capacitance and additional electrical components which model all losses and inductive parameters of a capacitor. In this series-equivalent circuit the electrical characteristics are defined by:

* C, the capacitance of the capacitor
* Rinsul, the insulation resistance of the dielectric, not to be confused with the insulation of the housing
* Rleak, the resistance representing the leakage current of the capacitor
* RESR, the equivalent series resistance which summarizes all ohmic losses of the capacitor, usually abbreviated as "ESR"
* LESL, the equivalent series inductance which is the effective self-inductance of the capacitor, usually abbreviated as "ESL".

Using a series equivalent circuit instead of a parallel equivalent circuit is specified by IEC/EN 60384-1.

Frequency dependence

Most discrete capacitor types have more or less capacitance changes with increasing frequencies. The dielectric strength of class 2 ceramic and plastic film diminishes with rising frequency. Therefore their capacitance value decreases with increasing frequency. This phenomenon for ceramic class 2 and plastic film dielectrics is related to dielectric relaxation in which the time constant of the electrical dipoles is the reason for the frequency dependence of permittivity. The graphs below show typical frequency behavior of the capacitance for ceramic and film capacitors.

For electrolytic capacitors with non-solid electrolyte, mechanical motion of the ions occurs. Their movability is limited so that at higher frequencies not all areas of the roughened anode structure are covered with charge-carrying ions. As higher the anode structure is roughned as more the capacitance value decreases with increasing frequency. Low voltage types with highly-roughened anodes display capacitance at 100 kHz approximately 10 to 20% of the value measured at 100 Hz.

Frequency dependence of capacitance for ceramic and film capacitors

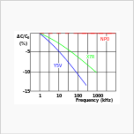

Frequency dependence of capacitance for ceramic class 2 capacitors (NP0 class 1 for comparisation)

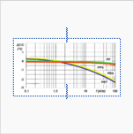

Frequency dependence of capacitance for film capacitors with different film materials

Wire-Gauge Ampacity

http://wiki.xtronics.com/index.php/Wire-Gauge_Ampacity

Material	Melt pt °C	Resistivity (Ω-m) at 20 °C
Silver	961.78	1.59×10-8
Copper	1084.62	1.72×10-8
Aluminum	660.32	2.82×10-8
Iron	1538	1.0×10-7
Tin	231.93	1.09×10-7
Platinum	1768.3	1.1×10-7
Lead	327.46	2.2×10-7
Tin/lead solder	183	1.44x10-7

Resistivities at room temp:

http://www.epanorama.net/documents/wiring/wire_resistance.html

Element	Electrical resistivity (microohm-cm)
Aluminum	2.655
Copper	1.678
Gold	2.24
Silver	1.586
Platinum	10.5

This clearly puts silver as the number one conductor and gold has higher resistance than silver or copper. It's desireable in connectors because it does not combine well with other materials so remains relatively pure at the surface. It also has the capability to adhere to itself (touch pure gold to pure gold and it sticks together) which makes for very reliable connections.

Voltage dependence

Capacitance may also change with applied voltage. This effect is more prevalent in class 2 ceramic capacitors. The permittivity of ferroelectric class 2 material depends on the applied voltage. Higher applied voltage lowers permittivity. The change of capacitance can drop to 80% of the value measured with the standardized measuring voltage of 0.5 or 1.0 V. This behavior is a small source of non-linearity in low-distortion filters and other analog applications. In audio applications this can be the reason for harmonic distortion.

Film capacitors and electrolytic capacitors have no significant voltage dependence.

Voltage dependence of capacitance for some different class 2 ceramic capacitors

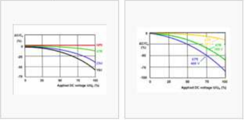

Simplified diagram of the change in capacitance as a function of the applied voltage for 25-V capacitors in different kind of ceramic grades

Simplified diagram of the change in capacitance as a function of applied voltage for X7R ceramics with different rated voltages

Impedance

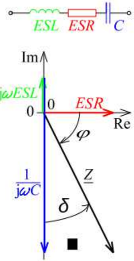

Simplified series-equivalent circuit of a capacitor for higher frequencies (above); vector diagram with electrical reactances XESL and XC and resistance ESR and for illustration the impedance Z and dissipation factor tan δ

In general, a capacitor is seen as a storage component for electric energy. But this is only one capacitor function. A capacitor can also act as an AC resistor. In many cases the capacitor is used as a decoupling capacitor to filter or bypass undesired biased AC frequencies to the ground. Other applications use capacitors for capacitive coupling of AC signals; the dielectric is used only for blocking DC. For such applications the AC resistance is as important as the capacitance value.

The frequency dependent AC resistance is called impedance \scriptstyle Z and is the complex ratio of the voltage to the current in an AC circuit. Impedance extends the concept of resistance to AC circuits and possesses both magnitude and phase at a particular frequency. This is unlike resistance, which has only magnitude.

$$Z = |Z|e^{j\theta}$$

The magnitude \scriptstyle |Z| represents the ratio of the voltage difference amplitude to the current amplitude,

16	2,426.30	4.27	0.058	19/29
16	2,600.00	4.00	0.059	26/30
16	2579.85	4.02	0.059	65/34
16	2625	3.99	0.059	105/36
14	4480	2.31	0.073	7/22
14	3830.4	2.7	0.073	19/27
14	4100	2.53	0.073	41/30
14	4167.5	2.49	0.073	105/34
12	7168	1.45	0.096	7/20
12	6087.6	1.7	0.093	19/25
12	6500	1.75	0.095	65/30
12	6548.9	1.58	0.095	165/30

Stranded Wire Data Chart

http://www.wetechnologies.com/technical-info/Stranded%20Wire%20Chart.pdf

AWG	Area mil	Ohms kft	~OD	strands
36	28.00	371.00	0.006	7/44
34	43.75	237.00	0.007	7/42
32	67.27	164.00	0.008	7/40
32	76.00	136.40	0.009	19/44
30	112.00	103.20	0.012	7/38
30	118.75	87.30	0.012	19/42
28	141.75	64.90	0.015	7/36
28	182.59	56.70	0.016	19/40
27	219.52	51.47	0.018	7/35
26	277.83	37.30	0.019	7/34
26	250.00	41.48	0.021	10/36
26	304.00	34.43	0.020	19/38
24	448.00	23.30	0.024	7/32
24	396.90	26.09	0.023	10/34
24	475.00	21.08	0.024	19/36
24	384.40	25.59	0.023	41/40
22	700.00	14.74	0.030	7/30
22	754.11	13.73	0.031	19/34
22	650.00	15.94	0.030	26/36
20	1,000.00	10.32	0.035	10/30
20	1,216.00	8.63	0.037	19/32
20	1,031.94	10.05	0.036	26/34
20	1,025.00	10.02	0.036	41/36
18	1,769.60	5.86	0.048	7/26
18	1,600.00	6.48	0.047	16/30
18	1,900.00	5.46	0.049	19/30
18	1,627.29	6.37	0.047	41/34
18	1,625.00	6.39	0.047	65/36
16	2,828.00	3.67	0.060	7/24

\scriptstyle j is the imaginary unit, while the argument \scriptstyle \theta gives the phase difference between voltage and current.

In capacitor data sheets, only the impedance magnitude |Z| is specified, and simply written as "Z" so that the formula for the impedance can be written in Cartesian form

$$Z = R + jX$$

where the real part of impedance is the resistance \scriptstyle R (for capacitors \scriptstyle ESR) and the imaginary part is the reactance \scriptstyle X.

As shown in a capacitor's series-equivalent circuit, the real component includes an ideal capacitor C, an inductance L (ESL) and a resistor R (ESR). The total reactance at the angular frequency ω therefore is given by the geometric (complex) addition of a capacitive reactance (Capacitance) $X_C = -\frac{1}{\omega C}$ and an inductive reactance $X_L = \omega L_{\mathrm{ESL}}$.

To calculate the impedance \scriptstyle Z the resistance has to be added geometrically and then Z is given by

$$Z = \sqrt{ESR^2 + (X_\mathrm{C} + (-X_\mathrm{L}))^2}.$$

The impedance is a measure of the capacitor's ability to pass alternating currents. In this sense the impedance can be used like Ohms law

$$Z = \frac{\hat{u}}{\hat{\imath}} = \frac{U_\mathrm{eff}}{I_\mathrm{eff}}.$$

to calculate either the peak or the effective value of the current or the voltage.

In the special case of resonance, in which the both reactive resistances

$$X_C = -\frac{1}{\omega C} \text{ and } X_L = \omega L_{\mathrm{ESL}}$$

have the same value $(X_C = X_L)$ then the impedance will only be determined by {ESR}.

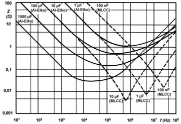

Typical impedance curves for different capacitance values over frequency showing the typical form with a decreasing impedance values below resonance and increasing values above resonance. As higher the capacitance as lower the resonance.

The impedance specified in the datasheets often show typical curves for the different capacitance values. With increasing frequency as the impedance decreases down to a minimum. The lower the impedance, the more easily alternating currents can be passed through the capacitor. At the apex, the point of resonance, where XC has the same value than XL, the capacitor has the lowest impedance value. Here only the ESR determines the impedance. With frequencies above the resonance the impedance increases again due to the ESL of the capacitor. The capacitor becomes to an inductance.

As shown in the graph the higher capacitance values can fit the lower frequencies better while the lower capacitance values can fit better the higher frequencies.

28	159.79	64.9	0.01264	1.4
29	126.72	81.83	0.01126	1.2
30	100.5	103.2	0.01002	0.86
31	79.7	130.1	0.00893	0.7
32	63.21	164.1	0.00795	0.53
33	50.13	206.9	0.00708	0.43
34	39.75	260.9	0.0063	0.33
35	31.52	329	0.00561	0.27
36	25	414.8	0.005	0.21
37	19.83	523.1	0.00445	0.17
38	15.72	659.6	0.00396	0.13
39	12.47	831.8	0.00353	0.11
40	9.89	1049	0.00314	0.09
46		4544	0.0016	

American Wire Gages (AWG) Sizes and Resistances
http://hyperphysics.phy-astr.gsu.edu/hbase/tables/wirega.html

AWG	Area CM	Ohms kft	Dia in	Max I
0	167810	0.0618	0.40965	328
0	133080	0.078	0.3648	283
0	105530	0.0983	0.32485	245
1	83694	0.124	0.2893	211
2	66373	0.1563	0.25763	181
3	52634	0.197	0.22942	158
4	41742	0.2485	0.20431	135
5	33102	0.3133	0.18194	118
6	26250	0.3951	0.16202	101
7	20816	0.4982	0.14428	89
8	16509	0.6282	0.12849	73
9	13094	0.7921	0.11443	64
10	10381	0.9989	0.10189	55
11	8234	1.26	0.09074	47
12	6529	1.588	0.0808	41
13	5178.4	2.003	0.07196	35
14	4106.8	2.525	0.06408	32
15	3256.7	3.184	0.05707	28
16	2582.9	4.016	0.05082	22
17	2048.2	5.064	0.04526	19
18	1624.3	6.385	0.0403	16
19	1288.1	8.051	0.03589	14
20	1021.5	10.15	0.03196	11
21	810.1	12.8	0.02846	9
22	642.4	16.14	0.02535	7
23	509.45	20.36	0.02257	4.7
24	404.01	25.67	0.0201	3.5
25	320.4	32.37	0.0179	2.7
26	254.1	40.81	0.01594	2.2
27	201.5	51.47	0.0142	1.7

Aluminum electrolytic capacitors have relatively good decoupling properties in the lower frequency range up to about 1 MHz due to their large capacitance values. This is the reason for using electrolytic capacitors in standard or switched-mode power supplies behind the rectifier for smoothing application.

Ceramic and film capacitors are already out of their smaller capacitance values suitable for higher frequencies up to several 100 MHz. They also have significantly lower parasitic inductance, making them suitable for higher frequency applications, due to their construction with end-surface contacting of the electrodes. To increase the range of frequencies, often an electrolytic capacitor is connected in parallel with a ceramic or film capacitor.[36]

Many new developments are targeted at reducing parasitic inductance (ESL). This increases the resonance frequency of the capacitor and, for example, can follow the constantly increasing switching speed of digital circuits. Miniaturization, especially in the SMD multilayer ceramic chip capacitors (MLCC), increases the resonance frequency. Parasitic inductance is further lowered by placing the electrodes on the longitudinal side of the chip instead of the lateral side. The "face-down" construction associated with multi-anode technology in tantalum electrolytic capacitors further reduced ESL. Capacitor families such as the so-called MOS capacitor or silicon capacitors offer solutions when capacitors at frequencies up to the GHz range are needed.

Inductance (ESL) and self-resonant frequency

ESL in industrial capacitors is mainly caused by the leads and internal connections used to connect the capacitor plates to the outside world. Large capacitors tend to have higher ESL than small ones because the distances to the plate are longer and every mm counts as an inductance.

For any discrete capacitor, there is a frequency above DC at which it ceases to behave as a pure capacitor. This frequency, where XC is as high as XL, is called the self-resonant frequency. The self-resonant frequency is the lowest frequency at which the impedance passes through a minimum. For any AC application the self-resonant frequency is the highest frequency at which capacitors can be used as a capacitive component.

This is critically important for decoupling high-speed logic circuits from the power supply. The decoupling capacitor supplies transient current to the chip. Without decouplers, the IC demands current faster than the connection to the power supply can supply it, as parts of the circuit rapidly switch on and off. To counter this potential problem, circuits frequently use multiple bypass capacitors—small (100 nF or less) capacitors rated for high frequencies, a large electrolytic capacitor rated for lower frequencies and occasionally, an intermediate value capacitor.

Q factor

The quality factor (or Q) of a capacitor is the ratio of its reactance to its resistance at a given frequency, and is a measure of its efficiency. The higher the Q factor of the capacitor, the closer it approaches the behavior of an ideal, lossless, capacitor.

---

1. Calculate the D.C. resistance of the Litz construction using Formula 3.
Rdc = 681.90 (1.015)^2 (1.025)^1 / 400 = 1.80 ohms / 1000 ft.

2. Calculate the A.C. to D.C. Resistance Ratio using Formula 2.

$$\frac{R_{AC}}{R_{DC}} = 1.000 + 2\left(\frac{400 \times 0.0045}{0.1040}\right)^2 \times \left(\frac{0.0045 \quad 50000}{10.44}\right)^4 = 1.2068$$

3. The A.C. resistance is, therefore, 1.2068 x or 1.80 ohms/1000ft.

*****

Following is an example of the calculations required to evaluate a Type 2 Litz construction consisting of 650 strands of 46 AWG single-film polyurethane-coated wire operating at 1.0 MHZ.

1. Calculate the D.C. resistance of the Litz construction using Formula 3.
Rdc = 4544 (1.015)^2 (1.025)^1 / 660 = 7.27 ohms / 1000 ft.

2. Calculate the A.C. to D.C. Resistance Ratio using Formula 2.

Rac/Rdc = 1.0003 +2(660*0.0016 / 0.056)^2 *
(0.0016*1000 / 10.44)^4 = 1.393

3. The A.C. resistance is, therefore, 1.39*7.27 = **10.13** ohms/1000ft.

The D.C. Resistance of a Litz conductor is related to the following parameters:

1. Individual strands AWG.
2. Number of strands in the cable.
3. Factors relating to the increased length of the individual strands per unit length of cable (take - up). For standard Litz constructions a 1.5% increase in D.C. resistance for every bunching operation and a 2.5% increase in D.C. resistance for every cabling operation are roughly correct.

The formula derived from these parameters for the D.C. resistance of any Litz construction is:

$$R_{DC} = \frac{R_S\,(1.015)^{N_B}\,(1.025)^{N_C}}{N_S}$$

Where: Rdc = Resistance in Ohms/1000 ft.
Rs = Maximum D.C. resistance of the individual strands (taken from Table 2).
Nb = Number of Bunching operations
Nc = Number of Cabling operations
Ns = Number of individual strands

Following is an example of the calculations required to evaluate a Type 2 Litz construction consisting of 400 strands of 38 AWG single-film polyurethane-coated wire operating at 500 KHZ. This construction, designed with two bunching operations and one cabling operation, would be written 5 x 5 / 16 / 40 AWG (HM Wire uses x to indicate a cabling operation and / to indicate a bunching operation.)

The Q factor of a capacitor can be found through the following formula:

$$Q = \frac{X_C}{R_C} = \frac{1}{\omega C R_C}$$

Where:

ω is frequency in radians per second,
C is the capacitance,
X_C is the capacitive reactance, and
R_C is the series resistance of the capacitor.

Ohmic losses, ESR, dissipation factor, and quality factor

The summarized losses in discrete capacitors are ohmic AC losses. DC losses are specified as "leakage current" or "insulating resistance" and are negligible for an AC specification. AC losses are non-linear, possibly depending on frequency, temperature, age or humidity. The losses result from two physical conditions:

* line losses including internal supply line resistances, the contact resistance of the electrode contact, line resistance of the electrodes, and in "wet" aluminum electrolytic capacitors and especially supercapacitors, the limited conductivity of liquid electrolytes and
* dielectric losses from dielectric polarization.

The largest share of these losses in larger capacitors is usually the frequency dependent ohmic dielectric losses. For smaller components, especially for wet electrolytic capacitors, conductivity of liquid electrolytes may exceed dielectric losses. To measure these losses, the measurement frequency must be set. Since commercially available components offer capacitance values cover 15 orders of magnitude, ranging from pF (10–12 F) to some 1000 F in supercapacitors, it is not possible to capture the entire range with only one frequency. IEC 60384-1 states that ohmic losses should be measured at the same frequency used to measure capacitance. These are:

* 100 kHz, 1 MHz (preferred) or 10 MHz for non-electrolytic capacitors with CR ≤ 1 nF:
* 1 kHz or 10 kHz for non-electrolytic capacitors with 1 nF < CR ≤ 10 μF
* 100/120 Hz for electrolytic capacitors
* 50/60 Hz or 100/120 Hz for non-electrolytic capacitors with CR > 10 μF

A capacitor's summarized resistive losses may be specified either as ESR, as a dissipation factor(DF, tan δ), or as quality factor (Q), depending on application requirements.

Capacitors with higher ripple current I_R loads, such as electrolytic capacitors, are specified with equivalent series resistance ESR. ESR can be shown as an ohmic part in the above vector diagram. ESR values are specified in datasheets per individual type.

The losses of film capacitors and some class 2 ceramic capacitors are mostly specified with the dissipation factor tan δ. These capacitors have smaller losses than electrolytic capacitors and mostly are used at higher frequencies up to some hundred MHz. However the numeric value of the dissipation factor, measured at the same frequency, is independent on the capacitance value and can be specified for a capacitor series with a range of capacitance. The dissipation

---

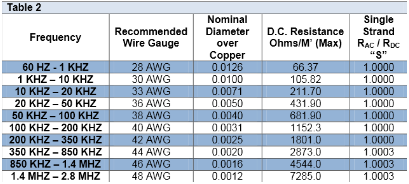

Table 2

Frequency	Recommended Wire Gauge	Nominal Diameter over Copper	D.C. Resistance Ohms/M' (Max)	Single Strand R_{AC} / R_{DC} "S"
60 HZ - 1 KHZ	28 AWG	0.0126	66.37	1.0000
1 KHZ – 10 KHZ	30 AWG	0.0100	105.82	1.0000
10 KHZ – 20 KHZ	33 AWG	0.0071	211.70	1.0000
20 KHZ – 50 KHZ	36 AWG	0.0050	431.90	1.0000
50 KHZ – 100 KHZ	38 AWG	0.0040	681.90	1.0000
100 KHZ – 200 KHZ	40 AWG	0.0031	1152.3	1.0000
200 KHZ – 350 KHZ	42 AWG	0.0025	1801.0	1.0000
350 KHZ – 850 KHZ	44 AWG	0.0020	2873.0	1.0003
850 KHZ – 1.4 MHZ	46 AWG	0.0016	4544.0	1.0003
1.4 MHZ – 2.8 MHZ	48 AWG	0.0012	7285.0	1.0003

After the individual wire gauge has been decided and assuming that the Litz construction has been designed such that each strand tends to inhabit all possible positions in the cable to just about the same extent, the ratio of A.C to D/C resistance of an remote Litz conductor can be determined from the following formula.

$$\frac{\textbf{Resistance to Alternating Current}}{\textbf{Resistance to Direct Current}} = S + K\left(\frac{N D_I}{D_O}\right)^2 G$$

Where:
S =Resistance ratio of individual strands when isolated (taken from Table 1 or 2)

G =Eddy Current basis factor = $\left(\frac{D_I\sqrt{F}}{10.44}\right)^4$
F =Operating Frequency in HZ
N =Number of strands in the cable
Dl =Diameter of the individual strands over the copper in inches
Do =Diameter of the finished cable over the strands in inches
K =Constant depending on N, given in the following table:

N	3	9	27	Infinity
K	1.55	1.84	1.92	2

Litz Designing Calculations
HM Wire International, Inc.
http://www.litz-wire.com/New%20PDFs/Litz_Designing_Calculations.pdf

Typically, the designing engineer that uses Litz knows the operating frequency and RMS current required for his application. Since the main advantage of a Litz conductor is the reduction of AC losses, the first thought in any Litz design is the operating frequency. The operating frequency not only influences the authentic Litz construction, it also determines the individual wire gauge.

Ratios of alternating-current resistance to direct-current resistance for an isolated solid round wire (S) in terms of a value (X) are shown in Table 1.

Table 1

X	0	0.5	0.6	0.7	0.8	0.9	1.0
S	1.0000	1.0003	1.0007	1.0012	1.0021	1.0034	1.005

The Value of X for copper wire is determined by Formula 1.

$$X=0.271\ D_M\sqrt{F_{MHZ}}$$

Where: Dm = Wire Diameter in mils
Fmhz = Frequency in Megahertz

From Table 1 and other practical data, the following table of recommended wire gauges vs. frequency for most Litz constructions has been prepared.

Table 2

---

factor is determined as the tangent of the reactance (X_C - X_L) and the ESR, and can be shown as the angle δ between imaginary and the impedance axis.

If the inductance ESL is small, the dissipation factor can be approximated as:

$$\tan\delta = ESR\cdot\omega C$$

Capacitors with very low losses, such as ceramic Class 1 and Class 2 capacitors, specify resistive losses with a quality factor (Q). Ceramic Class 1 capacitors are especially suitable for LC resonant circuits with frequencies up to the GHz range, and precise high and low pass filters. For an electrically resonant system, Q represents the effect of electrical resistance and characterizes a resonator's bandwidth B relative to its center or resonant frequency f_0. Q is defined as the reciprocal value of the dissipation factor.

$$Q = \frac{1}{tan\delta} = \frac{f_0}{B}$$

A high Q value is for resonant circuits a mark of the quality of the resonance.

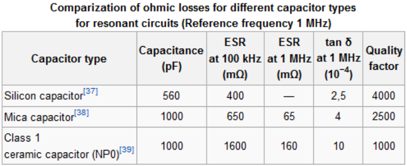

Comparization of ohmic losses for different capacitor types for resonant circuits (Reference frequency 1 MHz)

Capacitor type	Capacitance (pF)	ESR at 100 kHz (mΩ)	ESR at 1 MHz (mΩ)	tan δ at 1 MHz (10^{-4})	Quality factor
Silicon capacitor[37]	560	400	—	2,5	4000
Mica capacitor[38]	1000	650	65	4	2500
Class 1 ceramic capacitor (NP0)[39]	1000	1600	160	10	1000

Equivalent series resistance

In a non-electrolytic capacitor and electrolytic capacitors with solid electrolyte the metallic resistance of the leads and electrodes and losses in the dielectric cause the ESR. Typically quoted values of ESR for ceramic capacitors are between 0.01 and 0.1 ohms. ESR of non-electrolytic capacitors tends to be fairly stable over time; for most purposes real non-electrolytic capacitors can be treated as ideal components.

Aluminium and tantalum electrolytic capacitors with non solid electrolyte have much higher ESR values, up to several ohms, and ESR tends to increase with frequency due to effects of the electrolyte. A very serious problem, particularly with aluminium electrolytics, is that ESR increases over time with use; ESR can increase enough to cause circuit malfunction and even component damage,[1] although measured capacitance may remain within tolerance. While this happens with normal aging, high temperatures and large ripple current exacerbate the problem. In a circuit with significant ripple current, an increase in ESR will increase heat dissipation, thus accelerating aging.

Electrolytic capacitors rated for high-temperature operation and of higher quality than basic consumer-grade parts are less susceptible to become prematurely unusable due to ESR increase. A cheap electrolytic capacitor may be rated for a life of less than 1000 hours at 85°C (a year is about 9000 hours). Higher-grade parts are typically rated at a few thousand hours at maximum rated temperature, as can be seen from manufacturers' datasheets. Electrolytics of higher capacitance have lower ESR; if ESR is critical, specification of a part of larger capacitance than is otherwise required may be advantageous.

References:

AVX Basin Capacitor Formulas.
http://www.avx.com/docs/Catalogs/cbasic.pdf

Ben Tongue 1999, Sensitivity and selectivity issues in crystal radio sets including diode problems; measurements of the Q of variable and fixed capacitors, RF loss in slide switches and loss tangent of various dielectrics.
http://www.bentongue.com/xtalset/24Cmnts/24Cmnts.html

Wikipedia, Types of Capacitors, Ohmic Losses, Dissipation Factor and Q.
http://en.wikipedia.org/wiki/Types_of_capacitor#Ohmic_losses.2C_ESR.2C_dissipation_factor.2C_and_quality_factor

Dr. Gary L. Johnson, 2001. Lossy Capacitors
http://www.eece.ksu.edu/~gjohnson/tcchap3.pdf

Polymer capacitors usually have lower ESR than wet-electrolytic of same value, and stable under varying temperature. Therefore polymer capacitors can handle higher ripple current. From about 2007 it became common for better-quality computer motherboards to use only polymer capacitors where wet electrolytics had been used previously.[2]

The ESR of capacitors of relatively high capacity (from about 1 µF), which are the ones likely to cause trouble, is easily measured in-circuit with an ESR meter.

Typical values of ESR for capacitors

Type	22 µF	100 µF
Standard aluminum	0.1 - 3.0 Ω	0.05 - 0.5 Ω
Ceramic	<0.015 Ω	

AC current

An AC load only can be applied to a non-polarized capacitor. Capacitors for AC applications are primarily film capacitors, metallized paper capacitors, ceramic capacitors and bipolar electrolytic capacitors.

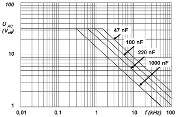

The rated AC load for an AC capacitor is the maximum sinusoidal effective AC current (rms) which may be applied continuously to a capacitor within the specified temperature range. In the datasheets the AC load may be expressed as

* rated AC voltage at low frequencies,
* rated reactive power at intermediate frequencies,
* reduced AC voltage or rated AC current at high frequencies.

Typical rms AC voltage curves as a function of frequency, for 4 different capacitance values of a 63 V DC film capacitor series

The rated AC voltage for film capacitors is generally calculated so that an internal temperature rise of 8 to 10 °K is the allowed limit for safe operation. Because dielectric losses increase with increasing frequency, the specified AC voltage has to be derated at higher frequencies. Datasheets for film capacitors specify special curves for derating AC voltages at higher frequencies.

If film capacitors or ceramic capacitors only have a DC specification, the peak value of the AC voltage applied has to be lower than the specified DC voltage.

AC loads can occur in AC Motor run capacitors, for voltage doubling, in snubbers, lighting ballast and for power factor correction PFC for phase shifting to improve transmission network stability and efficiency, which is one of the most important applications for large power capacitors. These mostly large PP film or metallized paper capacitors are limited by the rated reactive power VAr.

Bipolar electrolytic capacitors, to which an AC voltage may be applicable, are specified with a rated ripple current.

Insulation resistance and self-discharge constant

The resistance of the dielectric is finite, leading to some level of DC "leakage current" that causes a charged capacitor to lose charge over time. For ceramic and film capacitors, this

---

134	1025	0.0006250	1600	0.852
48	1712	0.0019341	517	6.425

On the main plot then I have included for Ben's Cap B a second Q vs. ESR curve to illustrate the estimated benefit of using a higher tuning capacitance range. The new curve slides mostly to lower ESR (especially at the high end of the tuning range) and also to slightly lower Q. The improvements made appear modest from the modeled data and, for the case of a high-Q capacitor the possible improvements may be slight indeed. It is comforting to know that at least there should be no negative impact. Lower ESR certainly improves tank overall efficiency. This is a win-win situation where the set designer wins twice!

In seeking the "Holy Grail" tank capacitor I would add to the list of valuable features. We have ceramic stator supports, silver plating on the vanes, and good wiper contacts as features desired. I would add to this an extended tuning range if the capacitor is truly to be considered "Holy". With a standard 230 uH coil the tuning for the BCB ranges from 40 to 380 pF or so. With a 180 uH coil the capacitance needed ranges from 50 to 490 pF, (stray capacitance not accounted for).

Two questions immediately come to mind with this analysis:

1) To how low a coil inductance can you go?
2) If you find a few of these super-caps can you send me one?

kjs 6/2013 (in France)

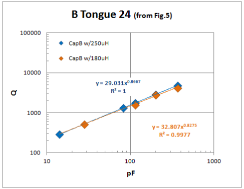

provides the estimated Q, and calculated D, and ESR for the capacitances from Fig 5 (with 20pF stray added). The values of Q do not improve by moving to higher capacitances but the ESR of the cap is slightly (at low tuning freq) and well improved (at the high tuning freq). This is the goal. It is worthy to note that the tuning frequencies do not line up with the original data from Ben's Fig 3 as seen on my main plot. While Ben specifically noted that his figure 3 included 20pF strays, in his figure 5 he speaks in absolute capacitances. I did my best to correct for this but there remains some ambiguity in the presentation. The point remains though; moving to higher capacitances in the tank will lower the tank losses due to ESR.

250 inductor (w/20pF stray)

tank pF	KHz	D	Q	ESR ohm
385	513	0.0002083	4800	0.177
220	679	0.0003448	2900	0.404
134	870	0.0005714	1750	0.917
101	1000	0.0007638	1309	1.502
34	1726	0.0034973	286	23.046

180 inductor (w/20pF stray)

tank pF	KHz	D	Q	ESR ohm
385	605	0.0002381	4200	0.172
220	800	0.0003571	2800	0.355

---

resistance is called "insulation resistance Rins". This resistance is represented by the resistor Rins in parallel with the capacitor in the series-equivalent circuit of capacitors. Insulation resistance must not be confused with the outer isolation of the component with respect to the environment.

The time curve of self-discharge over insulation resistance with decreasing capacitor voltage follows the formula

$$u(t) = U_0 \cdot \mathrm{e}^{-t/\tau_\mathrm{s}},$$

With stored DC voltage U_0 and self-discharge constant

$$\tau_\mathrm{s} = R_\mathrm{ins} \cdot C$$

Thus, after τ_s voltage U_0 drops to 37 % of the initial value.

The self-discharge constant is an important parameter for the insulation of the dielectric between the electrodes of ceramic and film capacitors. For example, a capacitor can be used as the time-determining component for time relays or for storing a voltage value as in a sample and hold circuits or operational amplifiers.

Class 1 ceramic capacitors have an insulation resistance of at least 10 GΩ, while class 2 capacitors have at least 4 GΩ or a self-discharge constant of at least 100 s. Plastic film capacitors typically have an insulation resistance of 6 to 12 GΩ. This corresponds to capacitors in the uF range of a self-discarge constant of about 2000–4000 s.[42]

Insulation resistance respectively the self-discharge constant can be reduced if humidity penetrates into the winding. It is partially strongly temperature dependent and decreases with increasing temperature. Both decrease with increasing temperature.

In electrolytic capacitors, the insulation resistance is defined as leakage current.

Leakage current

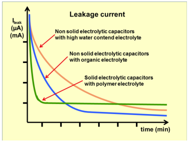

The general leakage current behavior of electrolytic capacitors depend on the kind of electrolyte

For electrolytic capacitors the insulation resistance of the dielectric is termed "leakage current". This DC current is represented by the resistor Rleak in parallel with the capacitor in the series-equivalent circuit of electrolytic capacitors. This resistance between the terminals of a capacitor is also finite. Rleak is lower for electrolytics than for ceramic or film capacitors.

The leakage current includes all weak imperfections of the dielectric caused by unwanted chemical processes and mechanical damage. It is also the DC current that can pass through the dielectric after applying a voltage. It depends on the interval without voltage applied (storage time), the thermic stress from soldering, on voltage applied, on temperature of the capacitor, and on measuring time.

The leakage current drops in the first minutes after applying DC voltage. In this period the dielectric oxide layer can self-repair weaknesses by building up new layers. The time required depends generally on the electrolyte. Solid

---

pF		f	Q	
tank	cap	MHz	Cap A	Cap B
375	355	520	19000	4750
190	170	730	14000	3230
114	94	943	14400	1610
101	81	1000	*13148*	*1828*
60	40	1300	11600	870
35	15	1710	9800	460

for I = 250 uH and 20pF stray in tank
(Q data in italics is computed)

The quest for high Q is indeed a quest for low ESR as long as you work in the capacitance range needed. One interesting piece of advice I have occasionally seen is that, to increase your tank Q, it is better to use lower-value coils and higher-value tuning caps. As much of the losses in the tank derive from losses in the coil this makes excellent sense. I wish to ask the question what about those higher-value caps? The plot immediately suggests that moving to higher capacitance drives down the ESR, this is a good thing. It may or may not improve the cap Q but I am not sure it matters as long as resistive losses are minimized. To better set my expectations I have again turned to Ben's Cap B data.

The small plot below takes data presented in his Figure 5. Data presented include Q at various capacitances and frequencies. I cross-plotted his data for pF and Q for the appropriate frequencies using both 250 uH and 180 uH inductances. Reading the actual Q off the tiny plot was an estimate at best, but it is interesting. This allows one to estimate his cap Q at any capacitance setting. The table below the likely capacitor data will be specific to caps in the uF range, or 10^6 higher that what we need in a crystal radio tank. Be careful not to get enamored with the high Q values occasionally quoted, check the capacitance. From the plot it should be apparent that not all Q's are created equal. Larger-value caps have lower ESR and so the Q's are often agreeably high. I have plotted a few specific cap measurements from various sources.

Two measurements should stand out; these are indicated by the larger blue diamonds. These are measurements by Bill Hebbert and presented by Ben Tongue in section B of his Article 24. I reproduce his Figure three below, with grateful apologies for the liberties I have taken,

By replotting Ben's figure I can calculate the relation between Q and frequency (table below) and so compute the estimated Q at any frequency. On my main plot I give Ben's Capacitor A and B for 1.0 MHz given the 250 uH inductor. Ben states that the plot assumes 20pF stray added. Since the capacitance needed to resonate at 1 MHz and 250 uH is 101 pF, I make the assumption his caps were set to 81 pF. This accounts for why the diamonds are set slightly off the 100 pF line of the plot. Given his Q measurements it is easy to compute the ESR for each cap. This is the world of the crystal radio builder. I include the ESR at the BCB endpoints for each cap as well to give the full picture.

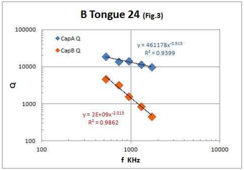

electrolytes drop faster than non-solid electrolytes but remain at a slightly higher level.

The leakage current in non-solid electrolytic capacitors as well as in manganese oxide solid tantalum capacitors decreases with voltage-connected time due to self-healing effects. Although electrolytics leakage current is higher than current flow over insulation resistance in ceramic or film capacitors, the self-discharge of modern non solid electrolytic capacitors takes several weeks.

A particular problem with electrolytic capacitors is storage time. Higher leakage current can be the result of longer storage times. These behaviors are limited to electrolytes with a high percentage of water. Organic solvents such as GBL do not have high leakage with longer storage times.

Leakage current is normally measured 2 or 5 minutes after applying rated voltage.

Dielectric absorption (soakage)

Dielectric absorption occurs when a capacitor that has remained charged for a long time discharges only incompletely when briefly discharged. Although an ideal capacitor would reach zero volts after discharge, real capacitors develop a small voltage from time-delayed dipole discharging, a phenomenon that is also called dielectric relaxation, "soakage" or "battery action".

Values of dielectric absorption for some often used capacitors

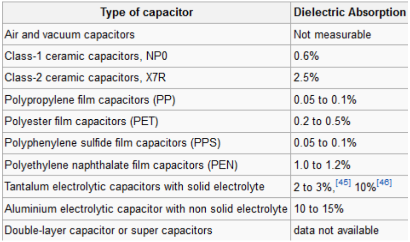

Type of capacitor	Dielectric Absorption
Air and vacuum capacitors	Not measurable
Class-1 ceramic capacitors, NP0	0.6%
Class-2 ceramic capacitors, X7R	2.5%
Polypropylene film capacitors (PP)	0.05 to 0.1%
Polyester film capacitors (PET)	0.2 to 0.5%
Polyphenylene sulfide film capacitors (PPS)	0.05 to 0.1%
Polyethylene naphthalate film capacitors (PEN)	1.0 to 1.2%
Tantalum electrolytic capacitors with solid electrolyte	2 to 3%,[45] 10%[46]
Aluminium electrolytic capacitor with non solid electrolyte	10 to 15%
Double-layer capacitor or super capacitors	data not available

In many applications of capacitors dielectric absorption is not a problem but in some applications, such as long-time-constant integrators, sample-and-hold circuits, switched-capacitor analog-to-digital converters, and very low-distortion filters, it is important that the capacitor does not recover a residual charge after full discharge, and capacitors with low absorption are specified.[47] The voltage at the terminals generated by the dielectric absorption may in some cases possibly cause problems in the function of an electronic circuit or can be a safety risk to personnel. In order to prevent shocks most very large capacitors are shipped with shorting wires that need to be removed before they are used.[48]

dependant. It is important to remember that not all Q's are created equal. What we want is to keep the ESR to a minimum.

On the plot there are a number of data trends presented. Note first that both plot axes are logarithmic, it makes a difference on how one thinks of the data. Additionally, note that the plot is specific (mostly) for data at 1.0 MHz. I make the assumption that a cap that is superior at that frequency is likely to be superior across the BCB.

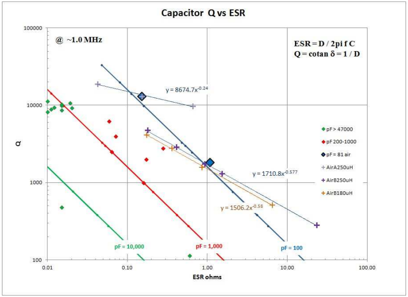

The first things to discuss are the three slanting lines. These are computed from a table of Dissipation Factors at 1.0MHz by Dr. Johnson. I have supplied the capacitances and calculated the resulting Q and ESR values. When one goes looking at capacitor data sheets or summary tables one is likely to find data computed at 100 KHz and 1.0 MHz (mercifully) but that

Crystal Radio Capacitors and ESR

Kevin Smith

In my ongoing effort to understand crystal radios and the theory that makes them work I have recently turned to setting my expectations concerning capacitance. The tuning capacitor is, next the coil itself, the principal component of the set. Few builders fabricate their own tuning caps so I suppose it is possible to take them a bit for granted. I have been reading on the radio forums some threads concerning radio Q and the subject of the variable cap Q seems to be getting more attention. So, what is the deal?

I set about my study by gathering what I could find on the web and at the close of this note I include a short reference list of useful pages from which I extracted data the ideas. I wish to extend special thanks to Ben Tongue who, as usual, has provided the essential measurements and background. The plot below of capacitor Q versus Equivalent Series Resistance (ESR) summarizes my findings. It is a bit busy as there are a number of interesting and important ideas represented in the relationships presented.

Any discussion of component quality needs first to acknowledge that the fight to increase the component Q is in reality a fight to eliminate (not possible) or reduce as much as feasible the resistive losses associated with that component. It helps then to look not just at Q, but at how Q relates to the component ESR. There are two fundamental equations that describe this relation: 1) Q = cotan theta = 1/D. D is the dissipation factor and is an attribute of the dielectric material used in the construction of the capacitor. Equation 2) ESR = D / 2pi f C. This equation tells us that capacitor losses expected due to resistance are both angular frequency and capacitance

ARTICLE 24

Sensitivity and selectivity issues in crystal radio sets including diode problems; measurements of the Q of variable and fixed capacitors, RF loss in slide switches and loss tangent of various dielectrics

By Ben H. Tongue
http://www.bentongue.com/xtalset/20MeaAGs/20MeaAGs.html

Part A: The Issues.

The sensitivity and selectivity in a crystal radio set can be impacted by many factors, including:

- A1: Ineffective coupling of the antenna to the tank circuit.
- A2: Resistive RF losses in the tank circuit.
- A3: Inappropriate diode.
- A4: Unintentional diode damage from exposure to electrostatic discharge (ESD) or voltages higher than the minimum reverse voltage rating. This is one cause of the loss of weak-signal sensitivity over time.
- A5: Relations between antenna-ground system, diode RF input and tank loss resistance, as effecting selectivity and loaded Q.
- A6: Audio transformer loss.

A1. There is a practical minimum limit to the possible impedance transformation ratio of the series resistance of the antenna-ground circuit to the shunt value desired across the tank circuit, when the transformation means is just a series capacitor between the antenna and the top of the tank. This problem occurs when the capacitor required for the desired

Do you really want a linear ohm meter scale? How about a LOG scale? Analog or digital?
What are the reasonable good and bad limits of ESR? What range in ohms? How sensitive does it need to be?
Input - To protect or not protect? What about charged caps?
Power - Battery or A/C adaptor? Battery life?
Physical Size
Cost

All of the above choices were carefully considered in the design of the Capacitor Wizard®. The end product was to be easy to use, durable, small & lightweight, accurate, measure ESR in circuit and inexpensive. I think I have met these design goals with the Capacitor Wizard®!

impedance transformation becomes so large that it causes the tank inductance to resonate below the desired frequency (See Article #22, Part 2). The solution here is to use a lower value inductance for the tank or to tap the antenna down on the tank (towards ground). This is why some experimenters find that low inductance tank circuits seem to work better than those of higher inductance. If one does not use a series capacitor for impedance transformation, the antenna may be just tapped down on the tank. Another alternative is to connect the antenna-ground system to a low value unturned inductance that is coupled to the tank.

The main advantage of using a series capacitor connected to the top of the tank (for impedance matching) is that it moves the undesired short wave resonance of the antenna-ground circuit (present in every single tuned crystal radio set) to the highest possible frequency and reduces its strength. A disadvantage is that unless a high enough Q variable capacitor is used, insertion power loss is increased, especially at the high frequency end of the band. See Part B, Section 1 of this Article and Article #22, Part 7.

A2. Resistive RF losses in the tank circuit are affected by: 1) Losses in any capacitor used for tuning or RF coupling. 2) Physical size of the coil and such items as length/diameter ratio, cross-section size and shape, and turns spacing (to reduce coil proximity losses). 3) Loss tangent in the coil form material, wire insulation and all dielectric material penetrated by the electric field of the coil. 4) Wire size and plating, if any. Silver plating is good but tin plating is bad, especially at the high end of the band. 5) Wire construction such as litz, solid or un-insulated stranded. The latter should be avoided. 6) Switches (if used). 7) Magnetic coupling from the coil to nearby lossy metallic objects. 9) Capacity coupling from "hot" high impedance RF points through a lossy RF return path to ground. See Article #22, point 7 of Part 10. Some comments: The loss from the loss tangent of the dielectric material used for the mounting base of detector stands can be nontrivial. Loss tangent of the material used for a front panel can cause dissipative loss if terminals provided for the connection of an external diode are too close together. See Part B, section 3 of this Article.

A3. A diode with too low an axis-crossing resistance (too high a saturation current) will resistively load the tank too heavily, causing a low loaded Q that results in loss of selectivity and sensitivity. A diode with too high an axis-crossing resistance (too low a saturation current) will increase selectivity because it only lightly resistively loads the tank circuit. The disadvantage is that sensitivity is reduced (many people liken this to the diode having too high a "turn-on" voltage.) and considerable audio distortion is generated. A little reverse or forward DC voltage bias will usually fix up these performance problems. See Article #9.

A way to check whether weak signal performance would be improved if one used a diode with a different saturation current (Is), but without experimenting with DC bias, is as follows: 1) Give the diode a one second or so spray with an aerosol "component cooler" . The reduction in temperature will temporarily substantially reduce the Is of the diode. If performance improves during the subsequent warm-back-up period, but before reaching room temperature, the diode has too high a room temperature Is. 2) Heat the diode by holding a hot soldering iron next to it for 5 seconds or so or give it a quick blow from a hot hair dryer. If performance improves during the subsequent cool-back-down period, but before reaching room temperature, the diode has too low a room

voltage across the test terminals to maintain its constant current. This voltage would turn on solid state devices when measuring ESR in circuit. For this reason the constant voltage method was chosen for the Capacitor Wizard.

For convenience we don't want a linear scale because it would require a Range Switch. I modified the basic constant voltage diagram above to include a resistor in parallel with the constant voltage generator that has the desirable effect of logging the ohm scale. It also serves to lower the output voltage to a level far below that required to turn on solid state devices.

Conclusion:

It should now be obvious that a DC ohm meter cannot measure ESR! A Capacitance meter is equally unsuitable for ESR measurements. ESR doesn't exist under the STATIC measurement condition of a DC ohm meter. Capacitance tells one nothing about a capacitors ESR. To measure ESR a controlled DYNAMIC test condition must be created. To create this test condition the engineer needs to consider a number of "Design Judgement Calls" based upon his experience and the desired end result. Electronic design is an exercise in compromise. There are many ways to design an ESR meter and each has its advantages and disadvantages. The quality of the end product is the result of wise choices made by the design engineer. Some of the important design considerations for an ESR meter are:

Which to use - constant voltage or constant current? At what magnitudes?
Test Signal - Sine wave or square wave? At what frequency?

with a linear meter scale. We could use either digital or analog meter displays.

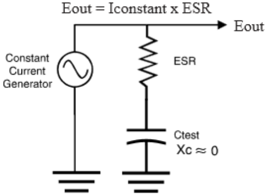

Constant Current Method

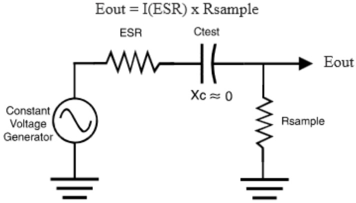

Constant Voltage Method

Constant current has the advantage of conveniently measuring low values of ESR down in the milli ohm range. It however has the disadvantage of applying a theoretical infinitely large

temperature Is. Is, for the usual Schottky diode, changes by about two times for each 10° C. temperature change. Germanium diodes probably act the same. Aerosol component coolers are available from most Electronics Distributors, including Radio Shack.

A4. Semiconductor diodes are subject to damage from exposure to electrostatic discharge (ESD) and voltages higher than their reverse breakdown values. Easy-to-check-for damage shows up as an increased reverse current when tested with a reverse DC voltage close to or less than the specified minimum breakdown voltage, as compared to an undamaged diode. Other types of damage can also occur. The effect on detector performance can be anywhere from a mild to a very great reduction of weak-signal sensitivity. This type of damage has the effect of placing a resistor across the diode, which reduces tank Q and adds a parasitic, unneeded audio load resistance. If one has an old VOM having a d'Arsonval moving-coil meter movement (not a digital type) such as the Triplett 360, Weston 980 or Simpson 260, one can do a quick check of the back leakage of a Schottky diode by measuring its back resistance on the X1,000 range. If the needle does not move, the diode has probably not been damaged from ESD. This test does not apply to so-called zero-bias Schottky detector diodes because of their usual low reverse breakdown voltages. Undamaged germanium diodes normally have greater reverse leakage (lower back-resistance) than the usual Schottky detector diodes. This does not effect performance in crystal set applications. The germaniums seem less subject to the electrostatic damage problem precisely because their lower back-resistance (and high reverse breakdown voltages) tends to bleed off any electrostatic charge that might accumulate, maybe from handling. Diodes with the lowest leakage can be selected with a simple test as follows: Connect a 3-4.5 volt

DC source, a 4.7k to 10k resistor, the diode and a DVM set to read DC current in series. Polarize the battery so that the diode is back biased. If the current is 2 uA or less, the parasitic back leakage resistance is greater than 2 Megohms or so and all should be well, as far as weak signal loss is concerned. The resistor is used to prevent damage to the test diode if it is accidentally connected in the forward direction. This test may be used to sort out the best weak-signal sensitivity diodes and exclude damaged ones. It is probable that for a selection of diodes having the same part number, ones with the lowest reverse leakage current will deliver the best weak-signal performance.

I believe poor weak-signal results some people have reported with the Avago (formerly Agilent) 5082-2835 and HSMS-2820 diodes can be laid at the doorstep of ESD damage. I experienced my first problem with these diodes by storing a few new ones in a propylene vitamin pill bottle for a while. When I wanted a new one I picked up the bottle, shook it (I think,) to see if something was inside and withdrew a diode. It performed badly, so I checked its back resistance with my Triplett 360 VOM as suggested above and found a low back resistance reading. Checking all the others, I found they were all bad. Static electricity did them in... beware! It was pretty stupid of me to not use anti-static bags. BTW, I've never had problems with those diodes if stored in paper bags. The bags must have had enough moisture content to provide sufficient conductivity to bleed off any static charge that might accumulate.

It is difficult to properly impedance match these diodes (see Part 4 of Article #5) because of their low saturation current (high axis-crossing resistance). They work best if 3 or or so are connected in parallel to raise the effective saturation current. If

---

Disadvantages: Not suitable for in circuit use, will not find open caps, keying must be highly accurate and short duration or ESR readings will be wrong, moderately complex, moderate cost, requires periodic calibration

We now have a method to correct for capacitive reactance by using a high test frequency and a valid formula to compute ESR. Let's investigate our formula E=IR.

The formula E = IR has three unknown quantities and it is not usable until we get the formula into the form of two known and one unknown quantities . The quantity we are trying to find is "R" so let's make it the unknown. We are left with "E" (voltage) and "I" (current) as the two variable quantities to be measured. Using conventional test equipment we can drive a capacitor with an ac 100khz signal, measure both "E" and "I" and, using the formula E = I x R, compute "R" (ESR). That method will work but it is clumsy, slow, and requires two pieces of test equipment - an AC volt meter, and an AC ammeter. If we were to hold one of the two variables in the test signal at a known CONSTANT VALUE we could measure the other variable with only one piece of test equipment! We could then calibrate its scale to represent the COMPUTED VALUE of "R"!! We could hold the test signals current(I) constant and measure the voltage(E) and then use the result to compute the value of "R". Or we could hold the test signals voltage(E) constant (the Capacitor Wizards method) and measure the test signals current(I). The result could be displayed on an analog or digital meter calibrated in ohms. These two methods of measuring resistance are well known by electronic measurement engineers as the "CONSTANT CURRENT" and "CONSTANT VOLTAGE" methods. Since the formula E = IR is a linear expression, both methods will yield an ohm meter

(1) Use a high test frequency: This is the technique the Capacitor Wizard® uses to "Zero Out" capacitive reactance.

Advantages: Will find open capacitors within the specified range of capacitance, Simple circuit, Reasonable Cost, no calibration required

Disadvantages : Only valid above a certain capacitance value, dependent upon test frequency

Xc = capacitive reactance
Zc = Total AC resistance (Vector Sum of Xc and ESR)
ESR = Equivalent Series Resistance
C = capacitance in Farads
F = Frequency in Hertz

Proof:

$$\text{If} \quad Xc = 1 / (6.28 * F * C) = 0 \quad \text{AND} \quad Zc = \sqrt{ESR^2 + Xc^2}$$

$$\text{THEN} \quad Zc = \sqrt{ESR^2 + 0^2} = \sqrt{ESR^2} = ESR$$

(2) Electronically subtract capacitive reactance from measured AC ohms:

Uses: Suitable for capacitor manufacturer

Advantages: Measure ESR of any capacitor, can also be adapted to measure capacitance

Disadvantages: Requires operator skills, not suitable for in circuit use, circuit complex, cost more, easily damaged, will not find open capacitors, requires periodic calibration

(3) Sample/Hold Keying:

Uses: Inexpensive method to measure ESR of any capacitor

Advantages: Measure ESR of any capacitor

---

one wishes to use only one of these diodes, it could have a low forward bias applied to get the same result (see Article #9). Keep in mind that these diodes do not do very well on strong signals when optimally impedance matched for weak signal reception because of their low reverse voltage breakdown rating. Avago specifies that rating as 15 volts for these diodes. If a rectified voltage approaching 7.5 volts appears across the DC load (now operating in 'peak-detector' mode), a maximum reverse voltage of about 15 volts will appear across the diode for every cycle of the RF carrier. When the diode is operated at this signal level or greater, a 'bucking' rectified current flows through the diode reducing audio output and causing audio distortion. This problem is eliminated by using diodes having a greater reverse breakdown voltage such as most ITT FO 215 and most 1N34A units.

A5. A question asked by many designers is this: What is the best approach when deciding how to 'impedance match' the transformed antenna-ground-system resistance to the RF input resistance of the diode detector? This would be a no-brainer if the resonant resistance of the tank was infinite, but it is not.

Table 1 - Definitions of Terms

IPL	Insertion power loss
Lo	Inductance of the tank
Ql	Loaded Q of the tank
Qo	Unloaded Q of the tank
Rag	Actual antenna-ground-system resistance
Ragt	A antenna-ground-system resistance transformed its value at the top of the tank
Rd	Actual input RF resistance of the diode.
Rdt	Input RF resistance of the diode as transformed to its value at the top of the tank.
Ro	Resistor representing all losses in the tank
Xa	Antenna-ground-system impedance transformation.
Xd	Diode input RF resistance transformation.

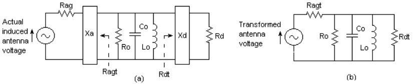

Fig. 1 - Simplified schematics of a single-tuned crystal set.

Simplified Schematic

The presence of finite tank loss increases IPL from the theoretical zero of the matched impedance case. It also reduces selectivity. Fig. 1a shows a simplified schematic of a single-tuned crystal radio set. The input impedance transformation Xa might take the form of capacitor in series with the 'Rag' and the 'induced voltage source' series combination or a tap on the tank. The diode input RF resistance transformation might take the form of a tap on the tank or a capacitor in series with the diode, along with other components. See Son of Hobbydyne and Hobbydyne II at http://www.hobbytech.com/crystalradio/crystalradio.htm. Table 2 shows calculated data for the simplified schematic shown in Fig.1b.

Table 2 - Calculated Insertion power loss (IPL), ratio of loaded to unloaded Q and impedance match (return loss*) in a single-tuned circuit.

---

does not really exist as a physical entity so direct measurements across the ESR resistor are no t possible! However, if a method of correcting for the effects of capacitive reactance is provided, and considering all ESR resistances are INPHASE, ESR can be calculated and measured by using the basic electronics formula E = I x R! This is the Basic Electronics foundation used to design the Capacitor Wizard®!

Before the formula E = I x R can be used to compute ESR, we must find a way to "ZERO OUT" or otherwise negate the effects of capacitive reactance. Here's why: Consider a series AC circuit that has an equal capacitive reactance (Xc) and resistance (ESR) of 10 ohms. The total value of these resistances when measured with an AC ohm meter will be the "VECTOR SUM" of the CAPACITIVE REACTANCE and the INPHASE resistance measured in ohms. In this example the total resistance measured by the AC ohm meter will be 14.14 ohms! This vector sum is always larger than the INPHASE resistance(ESR). For this reason an AC ohm meter is not a valid device to measure ESR unless a means of negating the effects of capacitive reactance is applied. Three equally valid basic techniques come to mind: (1) Raise the AC ohm meters test frequency so high that the capacitive reactance is so low compared to the expected ESR that the effects of capacitive reactance can be ignored. (2) Electronically subtract the capacitive reactance portion of the vector sum from the total measured AC resistance. (3) Use a sample/hold keying circuit and only "key on" the AC ohm meter when the PHASE ANGLE between the measured AC resistance and the test signal is at 0 or 180 degrees (INPHASE). Each technique has its advantages and disadvantages which will be explained below.

Capacitor ESR - Methods of Measurement

by Doug Jones (Capacitor Wizard)

http://www.midwestdevices.com/pdfs/Tnote3.pdf

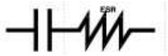

E = I x ESR

ESR would be easy to measure if it really existed as the single physical resistor represented by the drawing above! Unfortunately you can't just hang two test leads across the ESR resistor in the drawing above and measure the ESR! The drawing is just a symbolic representation of the total effect of the many physical quantities that make up ESR! This article is intended to give the reader an understanding of ESR and an introduction into the complexities of measuring ESR.

ESR is a DYNAMIC quantity of a capacitor. ESR does not exist as a STATIC quantity therefore it cannot be measured by a conventional DC ohm meter. ESR exists only when alternating current is applied to a capacitor or when a capacitors dielectric charge is changing states. ESR can be considered to be the TOTAL INPHASE AC resistance of a capacitor. INPHASE means the current(I) and the voltage(E) if measured across the resistive component ESR(R) of a capacitor are INPHASE with each other, just as if ESR(R) were pure resistance (no capacitive or inductive reactance). ESR includes the DC resistance of the leads, DC resistance of the connections to the dielectric, capacitor plate resistance, and the INPHASE AC resistance of the dielectric material at a particular frequency and temperature. The combination of components that make up ESR are symbolized by a resistor in series with a capacitor as shown above. This symbolic resistor

---

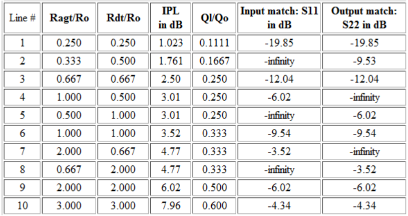

Line #	Ragt/Ro	Rdt/Ro	IPL in dB	Ql/Qo	Input match: S11 in dB	Output match: S22 in dB
1	0.250	0.250	1.023	0.1111	-19.85	-19.85
2	0.333	0.500	1.761	0.1667	-infinity	-9.53
3	0.667	0.667	2.50	0.250	-12.04	-12.04
4	1.000	0.500	3.01	0.250	-6.02	-infinity
5	0.500	1.000	3.01	0.250	-infinity	-6.02
6	1.000	1.000	3.52	0.333	-9.54	-9.54
7	2.000	0.667	4.77	0.333	-3.52	-infinity
8	0.667	2.000	4.77	0.333	-infinity	-3.52
9	2.000	2.000	6.02	0.500	-6.02	-6.02
10	3.000	3.000	7.96	0.600	-4.34	-4.34

*Return loss is a measure of the "goodness" of an impedance match. A value of minus infinity indicates a 'perfect impedance match' (all the available input power is delivered to the load). A value of zero indicates a 'perfect impedance mismatch' (all of the available input power is reflected back to the source, and none is delivered to the load). An intermediate value indicates the amount of available power that is reflected back to the source by the load. The usual way of referring to mismatch of a two-port network is by using S parameters. S11 is the input reflection coefficient and S22, the output reflection coefficient. The magnitude of a voltage reflection coefficient is 20*log|{(Rload-Rsource)/(Rload+Rsource)}|. One reference is Radiotron Designer's Handbook, Fourth Edition, pp 891-892. In our case, consider out little circuit (at resonance) to be a zero length transmission line having attenuation. Note that the Handbook was written before 'S" parameters were widely used.

Table 1 shows the tradeoff between IPL and selectivity. The lower Rag and Rd become, the lower IPL becomes, but Ql/Qo drops (poorer selectivity). Higher values of Rag and Rd result in greater IPL and greater Ql/Qo (greater selectivity). Ramon

Vargas has suggested that many people consider the parameters on line 4 to be close to a practical optimum, and they are. Line 3 shows alternate parameters for achieving the same selectivity at an IPL 0.51 dB less. A general rule may be stated that for any given value of Ql/Qo, equal values for Rag and Rd will result in the least possible IPL. In this case, input and output return losses will be equal. It appears to me that the parameters in lines 3 or 6 are probably the ones to shoot for in most design calculations.

An effect I have observed is that one cannot simultaneously attain a perfect impedance match at both input and output ports in systems of the type shown in Fig. 1a (simultaneous conjugate match). As shown on lines 4 and 5 in Table 2, Ragt and Rdt can be arranged to provide a perfect match at one port, but then the other port will be mismatched. Ro could be replaced by a series resistor (not a real-world crystal radio set anymore) and one would still not be able to arrange a simultaneous impedance match. If circuit losses were represented by proper values of both series and shunt resistances, it would then be possible to attain a simultaneous perfect match at both input and output.

An audio transformer is one passive device that has series and shunt loss components. If it is so designed, it can provide a simultaneous perfect impedance match at both input and output when loaded by its designed-for load resistances (ignoring reactance effects). If operated at any other impedance level, such as doubling the source and load resistances, simultaneous perfect input and output impedance matching cannot be attained. This info is of mainly theoretical interest for most crystal radio set applications except when one tries to operate an audio transformer at considerably higher or lower source and load resistance values than it was designed

5. WELSBY V G : ‘The Theory and Design of Inductance Coils’, Macdonald & Co, Second edition 1960.

6. The Plumbers Handbook :
http://www.kembla.com/assets/Uploads/general-PDFs/The-Plumbers-Handbook.pdf

Issue 1 : August 2013
Issue 2 : September 2013 : the addition of Annex 1 Broadcast Capacitor

© Alan Payne 2013
Alan Payne asserts the right to be recognized as the author of this work.
Enquiries to paynealpayne@aol.com

---

for. If this is done, IPL is considerable increased compared to using it the impedances for which it was designed.

A6. Audio transformer loss. See Articles #1 and #5.

Part B: The Measurements.

Loss measurements at various frequencies on some components, with the loss expressed as a parallel resistance (Rp) in parallel with the capacitance of the component or the loss tangent, or equivalent Q of the component:

*

B1: Q and equivalent parallel loss resistance (Rp) of two variable capacitors vs frequency when resonated with a 250 uH inductor, in parallel with a total circuit stray circuit capacitance of 20 pF. Rp and Q vs. frequency at four different capacitance settings is also shown.

*

B2: Equivalent parallel loss resistance of some DPDT slide switches.

*

B3: Loss tangent of some coil form and sheet plastic (front panel) materials.

*

B4: Q of inductors and L/C resonators.

*

B5: High Q fixed value ceramic capacitors.

B1: Measurements at 520, 730, 943, 1300 and 1710 kHz, made on two different variable capacitors used in crystal radio sets, are shown in the two graphs below. Capacitor A is a 485 pF variable capacitor that was purchased from Fair Radio Sales. It has a ceramic insulated stator and silver plated brass plates with silver plated wiper contacts. Capacitor B is a small 365 pF air-variable capacitor purchased from the Xtal Set Society. Its plates are made of aluminum and the stator support insulators are made of a phenolic plastic. This capacitor is similar to those sold by Antique Electronic Supply and others. See note at the end of this section B1.

Fig. 2 shows Rp plotted against frequency, with the capacitor adjusted at each frequency to a value 20 pf lower than that required to resonate a 250 uH inductor. This allows for a stray capacitance of 20 pF, in an actual circuit. Any losses that may be in the stray capacitance are assumed assigned to the inductor. The plot shows how, in actual practice, the Rp of variable capacitors A and B vary when tuned across the broadcast band. Fig. 3 shows how the Q of the total capacitance (including the 20 pF stray capacitance) varies across the BC band for each capacitor. Do not make the mistake, when looking at the two graphs below, of thinking that they represent Rp and Q of capacitors A or B vs. frequency, with the capacitor set to a fixed capacitance. The capacitance is varied as a function of frequency, along the horizontal axis of the graphs, to a value that would resonate with a 250 uH inductor. Figs. 4 and 5 show Rp and Q of capacitor B as a function of frequency, at four different fixed capacity settings (the frequency at each capacitance setting is always a value that brings about resonance with the 250 uH inductor).

The higher loss of this capacitor compared to that of the variable capacitor can be partially attributed to the difference in conductivity – copper versus silver, but this would give a difference of only about 6-7%. The additional loss is probably due to the longer conducting path in the fixed capacitor, because the connection to the outer tube was half way down its *outside* surface, so current to the inside had about twice the distance to travel. It became clear that even if the conductors were silver plated and the construction changed to minimise the conduction path it was unlikely that a loss significantly lower than the variable capacitor would be achieved, which was the aim.

11. References

1. MOULLIN E B : ‘A Method of Measuring the Effective Resistance of a Condenser at Radio Frequencies, and of Measuring the Resistance of Long Straight Wires’. Proc. R. Soc. Lond. A. 1932 137 116-133 doi:10.1098/rspa.1932.0125 (published 1 July 1932).

2. JACKSON W : ‘The Analysis of Air Condenser Loss Resistance’ Proc IRE, Vol22 Number 8, Aug 1934.

3. FIELD R F & SINCLAIR D B : ‘A Method for Determining the Residual Inductance and Resistance of a Variable Air Condenser at Radio Frequencies’. Proc IRE, Vol 24, Number 2, Feb 1936.

4. BOCK E M : ‘Low-Level Contact Resistance Characterisation’, AMP Journal of Technology, Vol 3, November 1993.

If the plate is silver (as here) , then p=1.629 10-8, and u = 4pi10-7, so Zwall = 8 10-4 at 10 Mhz. Evaluating ***l*** and w is somewhat difficult, but if we take semicircular plates of radius r, then when the capacitance is maximum the rotor and stator plates will fully overlap. The average length from the centre of the rotor (where current enters) will be r/2 and the average width (taken as the average circumference) will be 2π(r/2)/2, so the ratio ***l*** / w will be 1/π. As the spindle is rotated the overlap between the rotor and stator will reduce in direct proportion to the capacitance, so ***l*** / w will vary as:

$$\boldsymbol{l} / w \approx 1/(\pi\ C/Cmax) \qquad \text{A1.2}$$

So at maximum capacitance we can expect the loss resistance in the plates to be, Rv = 2 Zwall / (πN) = 0.034 mΩ at 10Mhz, for N=15. This is a factor of around 1000 less than that measured, and so any variation with spindle rotation would be insignificant, and as experiment confirms.

Appendix 2 : Standard Capacitor

An attempt was made to produce a standard capacitor, having a very low loss. This was made using two coaxial copper pipes, 15mm and 22mm diameter, about 100mm long. The insulation between the two pipes was two bands of PTFE tape, located at the ends of the tubes. Capacitance was around 25pf.

However this capacitor had a lower Q than the variable capacitor by about 12%. It was thought the PTFE tape may be lossy (it was plumber's tape), and it was replaced with thin expanded polystyrene foam, but with the same results.

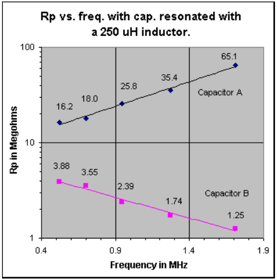

Fig. 2

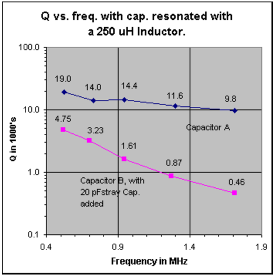

Fig. 3

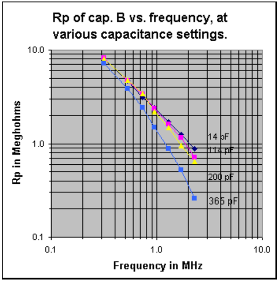

Fig.4

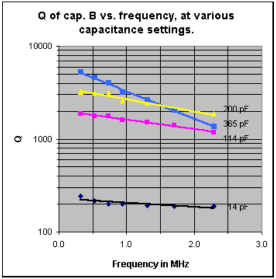

Fig. 5

capacitor uses a relatively inexpensive insulator, such as porcelain with a loss tangent of 0.0075, then the capacitance of its insulators is 4.8 pf (see Section 8.5), around half the minimum capacitance.

The contact resistance is also seen to be much larger at 32 mΩ compared with 10 mΩ. However, the metallic loss is about 1/3rd that of the Photo 1 capacitor (0.0039 cf 0.01), despite the metalwork being aluminium rather than silver. This lower metallic loss is because of the very much shorter path which the current takes. Its Q at 1 MHz and 4 MHz is given below :

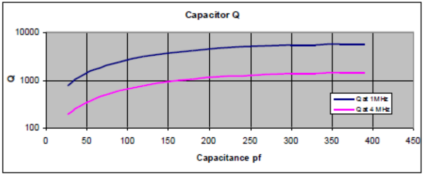

Appendix 1 : Estimated Loss in the Plates

The resistance of two plates will be twice that of one plate, and if we have N pairs of plates the resistance Rv will be :

Rv = 2 Zwall ***l*** / (N w) A1.1

where Zwall = p /d = (p d u f)0.5
p is the resistivity of the metal
u is the permeability of the metal
l is the length of the current path through the plate
w is the width of the current path through the plate

Subsequent to the above work, the loss of the broadcast-band tuning capacitor shown below was measured, and gave the following equation for its series resistance:

$$Rcap = 0.032 + 5800 /(f\ C2) + 0.0039\ (f\ 0.5)\ \text{ohms} \qquad 10.1$$

where f is in MHz and C in pf

The accuracy of the above equation is estimated at ± 13%.

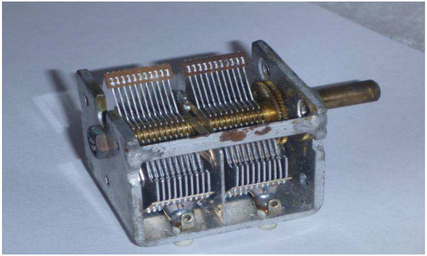

Photograph 3 : Broadcast-band Capacitor

This capacitor has 2 sections each with a range of 10 pf to 380 pf, and the above equation is for one section.

It is interesting to compare the losses of the capacitor shown in Photo 1 with that of this broadcast-band capacitor. In common with most capacitors of this type it has a very short insulation path and this accounts for the high dielectric loss factor of 5800 in the above equation, over 7 times the loss of the capacitor shown in Photo 1. Assuming this broadcast-band

---

Discussion:

There are two main sources of loss in an air-dielectric variable cap: 1) Loss in the dielectric of the stator insulators, and 2) Resistive losses in the metal parts. Of course, there is also the important very low-loss capacitor made from the air dielectric between the plates. The losses in 2) include resistive loss in the plates and in the wiper that connects the rotor to the frame. Resistive loss in the plates is very small at low frequencies, but increases with increasing frequency because of skin effect.

Notice, that for capacitor B, Rp (vertical scale) in Fig. 4 is about the same for all capacitance settings at the lowest frequency plotted (310 kHz). Rp is approximately constant over the full 14-365 pF capacitance range. This is because we are varying a high Q air dielectric capacitor of very low loss, in parallel with a low Q fixed capacitance made up of the lossy phenolic stator supports. Most of Rp comes from the loss in the phenolic. In Fig. 5, again at the low frequency end of the horizontal axis, observe that Q is a direct linear function of the overall capacitance, as it should be. The Q of the air-variable part, taken alone, is much higher than that of the capacitor made up of the phenolic insulators. The main loss, here, comes from the approximately fixed shunt Rp provided by the phenolic stator supports. At 14 pF the C from the air cap is relatively low compared to that from the phenolic insulators. At 365 pF the C from the air cap is much higher than that from the phenolic supports.

Things change at higher frequencies. The reactance of the capacitor drops. This, combined with the series resistance of the plates and wiper now come into play as an additional factor reducing the Q. If this series resistance (Rs) were the only resistance affecting Q, the equation for Q would be: Q=(reactance of the capacitor)/Rs. One can see from this equation that the introduction of Rs makes the Q drop when frequency increases. Up to now, the main loss came from the parallel loss resistance of the phenolic supports. The Q from a setting of 365 pF drops quite rapidly with increase of frequency because of this series resistance. Skin effect makes the effect worse by increasing Rs, the higher one goes in frequency. Notice that at the low frequency end of Fig. 5, Q is approximately proportional to the capacitance, as it should be if the main loss is the fixed shunt resistance Rp, coming from the phenolic stator insulators.

Fig. 5 shows an approximately constant Q (vs. frequency) when the capacitor is set to 14 pF. The loss causing this low, constant Q, comes from the loss tangent of the capacitor formed from the phenolic stator insulators. At low frequencies, Q increases when the capacitor is successively set to 114, 200 or 365 pF because engaging the plates adds a high Q air dielectric capacitor component in parallel with the low Q capacitor formed from the phenolic dielectric supports. As frequency increases, with the capacitor set to 365 pF, one can see that the Q drops at a faster rate than it does when set to 200 or 114 pF. This is because Rs (being in series with the air capacitor, that dominates at the 365 pF setting), acting with its lower capacitive reactance results in a capacitor of lower Q (Q=reactance of air capacitor at 365 pF/Rs).

In the Figs. 2 and 3, the capacitor is always set for a circuit capacity value that resonates with 250 uH. This means that at 520 kHz, the varicap is set so the circuit capacity is 375 pF. At 1710 kHz, the circuit capacity is set to 34.7 pF. Even though the capacitor Q goes from 19,000 to 9800 as frequency goes from 520 to 1710 kHz, Rp increases as frequency increases

The series resistance of a variable air capacitor at radio frequencies is shown to be given by :

Rcap = Rs + α /(f C2) + β (f 0.5) ohms 9.1

where Rs , α and β are constants for any particular capacitor.

f is the frequency and C the capacitance

The first factor Rs is due to the contact resistance, and this is largely due to the wiping contacts. The second factor is due to the dielectric loss in the supporting insulator, and the third factor is due to the resistivity of the metalwork, particularly that supporting the plates.

Measurement of this resistance requires the capacitive reactance to be tuned-out with a low loss inductor and it is shown that this can be achieved with straight copper pipes arranged as a shorted two-wire transmission line.
For the capacitor shown in Photo 1, the series resistance is given by:

Rcap = 0.01+ 800 /(f C2) + 0.01 (f 0.5) ohms 9.2

where f is in MHz and C in pf

The accuracy of the above equation is estimated at ± 13%.

10. ANNEX 1 : BROADCAST CAPACITOR

In respect of dielectric loss, Jackson says ‘The overall capacitance C of the capacitor may be regarded as composed of a fixed and imperfect portion C1 due to the supporting insulation, in parallel with a variable air portion devoid of energy loss. The power factor tan δ of the portion C1 seems usually to be substantially constant over a wide frequency range’.

It is interesting, although not particularly useful, to calculate capacitance of the supporting insulation, C1. This can be determined from Equation 6.1.2, by setting C = C1, since then the loss is not diluted by the air capacitor and the series resistance due to the dielectric loss will then be :

$Rc = 800/(f\ C12)$ 8.5.1

Also from the definition of tan δ :

$Rc = \tan\delta/(2\pi f\ C1)$
for f in Hz and C in farads

$= 106 \tan\delta/(2\pi f\ C1)$ 8.5.2
for f in MHz and C in pf (as equation 8.5.1)

Equating 8.5.1 and 8.5.2 :

$C1 = 2\pi\ 800 / (106 \tan\delta)$ pf 8.5.3

The material of the insulator is unknown but is possibly Steatite, for which tan δ = 0.003, and then the capacitance C1 due to the supporting insulation is 1.7 pf.

9. SUMMARY

---

because the circuit capacity must be reduced from 375 to 34.7 pF to tune the tank from 520 to 1710 kHz (Rp=Q/(2*pi*f*C).

If the question is posed: 'Is it more important to have ceramic insulated stators or silver plated plates on a variable capacitor used in a BC band crystal radio set?', the answer is that ceramic insulated stators are the way to go.

Capacitor A: The main practical conclusion that can be taken from Fig .2 above is that Rp of capacitor A is very high over the whole band and varies roughly proportionally to frequency, over the frequency range of interest: 520-1710 kHz In fact, Rp is so high that it will not contribute any appreciable loss even when used in high performance crystal radio sets using a high Q tank inductor. Another plus is that its Rp increases with increasing frequency, further reducing any effect on loaded Q, sensitivity or selectivity at the high end of the band. The silver plating on the brass plates is beneficial because the resistivity of silver is about 25% of that of brass. Practically speaking, the silver should have little effect on the operation of a crystal radio set in the BC band, but short wave is another matter. The Q of the capacitor drops at higher frequencies, especially when set to a high capacitance value. Silver plating can materially improve performance at higher frequencies by providing a needed higher Q. Be aware that some capacitors are made with cadmium plated brass plates. Silvery-whitish colored cadmium has a resistivity 4.6 times that of silver. This higher resistivity will somewhat reduce Q at high frequencies and at high capacity settings. Some people mistakenly assume, because of the silvery-whitish color, that cadmium plated plates are really silver plated.

Capacitor B: One can see from Fig. 2 that Rp of capacitor B varies approximately inversely with frequency. The DLF (dielectric loss factor) of the phenolic stator support insulators is the main cause of this loss, over the whole frequency band. Towards the high end of the band, some loss is contributed by the series resistance in the capacitor plates, the rotor shaft wiper contact and skin effect. This loss effect is greater than that in capacitor A because the resistivity of the aluminum plates is 1.7 times that of silver. Practically speaking, this effect is minimized at the high end of the band because the plates are mostly disengaged. The Rp of the capacitor will have its greatest effect in reducing sensitivity and selectivity at the high frequency end of the band because that is where its value is the the lowest. See note at the end of this Article.

The usual crystal radio set uses shunt capacitor tuning with a fixed tank inductance. This configuration causes the tank reactance to be highest at the high end of the band, thus further reducing loaded Q, sensitivity and selectivity, for a given value of Rp. At the low and medium frequency parts of the band, Rp of the capacitor is so high that its effect is small in many crystal radio sets. Highest performance crystal radio sets made with high Q inductors, with careful attention to impedance matching may experience a noticeable reduction in sensitivity and selectivity at the high end of the band when using this or other capacitors using phenolic stator insulation. This is because the Rp of the capacitor becomes comparable to the higher equivalent Rp of the high Q tank inductor at the high end of the band as compared to its value at the low end. To clarify this, an ideal condition would exist if Rp of the capacitor and inductor were infinite. If this were to be the case, and good impedance matching of antenna-ground system to tank, diode to tank and headphones to diode existed, all of the power intercepted by the antenna-ground system would be delivered to the headphones and maximum sensitivity would occur. Any loss present in capacitors, inductors or

remaining 30% is grooved. So we might anticipate the loss increasing considerably, and if it is assumed by a factor of 4, to 12 mΩ, this agrees well with the measurements as the following shows : a similar situation exists for the moveable vanes, and they are supported by the central spindle and this has very deep grooves. So by the same argument its resistance is estimated to be 15 mΩ, giving a total due to metallic loss of 27 mΩ, close to the measurements of total metallic loss of 28 mΩ at 10 MHz.

So it seems that the current flow in the support pillars is seriously interrupted by the presence of the plates, and this is a major source of loss.

8.4. Conductor Loss Budget

Bringing together the estimated conductor losses discussed above gives the following for 10 MHz (all figures are very approximate) :

Loss in vanes	0.03 mΩ	i.e. negligible
Pillar supporting static vanes	12 mΩ	
Shaft supporting moving vanes	15 mΩ.	
Slip ring contact resistance	10mΩ.	
Two solder tags	1 mΩ	
Total	38 mΩ	

The measurements indicate a total of 42 mΩ, so the above seems a reasonable estimate.

8.5. Dielectric Loss and Capacitance of Plate Supports

leaving about 4.5 mΩ attributable to each of the sliding contacts. This is somewhat larger than Bock's measurements (ref 4) of contact resistance of silver contacts of 1 mΩ, but the contact force here may be lower than his 100 gm, and also the contacts could be worn.

8.2. Loss in Plates

There is no indication in the measurements that the loss varies with capacitance setting, and so the loss in the plates is presumably very small, and this is supported by the analysis given in Appendix 1. Field & Sinclair (ref 3) found the same and stated 'the change in series resistance with rotor displacement is therefore small despite the change in current distribution in the plates'.

8.3. Loss in Plate Support Pillars

The static plates are supported by two silver-plated cylindrical bars, which are bolted to the end ceramic. If only one pillar is connected to the external circuit (the normal configuration), then the whole of the current flows through this pillar. Its diameter is 6.5 mm, and it has an effective length of 70 mm, and at 10 MHz such a bar would have a resistance of around 3 mΩ, assuming the whole circumference carries current. However part of the circumference is interrupted by the plates, and these will seriously affect the current flow. To see this, imagine a cylindrical conductor with circular washers along its length, forming a structure similar to a thermal radiator. Current flowing from one end to the other will now have to flow outwards across one face of each washer, across its thickness, and down the other face, and this will increase the resistance considerably. In the capacitor here, around 70% of the circumference is interrupted by plates and in addition the

transformers reduces sensitivity. In this discussion we are dealing primarily with losses in the resonating capacitor and tank inductor, both referred to as their respective values of Rp. The higher the value of these Rps, compared to the transformed antenna-ground source resistance across which they appear, the lesser the loss they cause. One approach to counter the effect the drop in Rp as frequency increases is to change to two-step inductive tuning by dividing the band into two sections as described in Article #22.

Technical Note: At any frequency, a real world capacitor of value C1 having a Q of Q1 (Q1>10), can be quite accurately modeled as a series combination of an ideal no-loss capacitor of value C1 and a resistor (Rs) equal to: (reactance of C1)/Q1. Alternatively, at any frequency, a capacitor of value C1, having a Q of Q1 (Q1>10) can be modeled as a parallel combination of an ideal capacitor of value C1 and a resistor (Rp). The resistor, Rp in this case, has a value of:: (reactance of C1)*Q1. Note that since capacitor reactance is a function of frequency, the value of the resistor will, in general, vary with frequency. In crystal radio set design it is sometimes convenient to model the tuning capacitor loss as a parallel resistor, other times as a series resistor.

Credit must go to Bill Hebbert for making the time consuming, difficult, precision measurements required for Figs. 2-5.

B2: Slide switches used in the Crystal Radio Set described in Article #22 have dielectric losses, as do all switches. To get a handle on this loss, samples of several different types of DPDT switch were measured at 1500 kHz. Each switch except the last three, below, were wired as a SPDT unit by paralleling the two sections. The Q of the capacitance appearing across the open contacts was then measured and Rp was calculated. Rp usually varies approximately inversely with frequency and therefore causes more loss at the high end of the band than at the low end. Contact resistance of all switches was found to be very low. It is unknown how well that characteristic will hold up over time. Note the extremely low loss of the Switchcraft 56206L1. The loss is so low that this switch is overkill in most crystal radio set applications. For applications in which the crystal radio set builder wants to use the lowest loss DPDT slide-switch available, this switch is the best I have found.

Table 3 - Equivalent parallel Loss Resistance (Rp) caused by the loss tangent of the dielectric of some DPDT Slide Switches at 1.5 MHz.

Brand name, model number and Size of switch	**Rp in Megohms**
ARK-LES (std. size)	2.3
Radio Shack 275-403A (std. size)	3.5
Radio Shack 275-407A (sub mini)	4.1
Stackpole S7022X (std. size)	4.9
Switchcraft 46206EE-6 (std. size)	5.1
Radio Shack 275-327B (mini)	5.7
CW (mini)	6.0
Switchcraft 46206LR (std. size)	6.5
C&K L202-1 (mini)	6.8
Switchcraft 56206L1 (mini)	13.3
Radio Shack 275-327B connected as SW3 in Article #22 (HI band)	4.1
Switchcraft 56206L1 connected as SW6 in Article #22	9.4
Radio Shack 275-327B connected as SW6 in Article #22, **See Graph--->>**	3.5

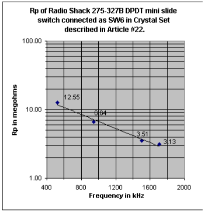

Rotary selector switches using ceramic insulation should have very low loss, even lower than the Switchcraft 56206L1. Quality switches using brown phenolic insulation probably have losses similar to slide switches using similar material. I

random uncertainty in bandwidth of around ± 1%. These random errors and quantisation errors were reduced by using the analyser's averaging of 16 measurements, to gave an overall precision (narrowness of scatter) in the measurement of Q of less than ± 1%.

The *accuracy* of the bandwidth measurements is less easy to estimate, but the errors in the AIM 4170 are probably less significant than other practical matters, such as small changes to the experimental set-up, such as the size of the coupling loop with respect to the tuned circuit. These practical aspects probably account for ±5%. In addition there is the ±1.5% error in matching the measurements (para 6.1) to the model.

So the overall measurement accuracy is assessed at ±9%, for the measurement of the overall Q and its comparison with theory. Since the capacitor resistance is about half the total this translates to an uncertainty in the capacitor resistance of ±13% (assuming the errors add in quadrature).

8. ANALYSIS OF LOSS FACTORS

8.1. Contact Resistance

There are a number of contacts contributing to the contact resistance. There is a sliding contact to the rotor, with two contacts in series, and for connection to the external circuit there is a solder tags bolted to each of the rods supporting the stator plates, and another to the static part of the sliding contact. The measurements show a combined contact resistance of 10 mΩ. The pressures on the solder tags are likely to be much greater than the slider pressure, and so their contact resistance will probably be of the order of 1 mΩ each or less, The proximity multiplier and the radiation resistance contribute a very small resistance and so uncertainties here will have a negligible effect on the results.

The major contributor is the resistance of the copper pipes, and the main uncertainties here are the resistivity and permeability of the copper, and each have an uncertainty of around ±6 %, so the uncertainty on Rwall is ±6 %. Also there was an uncertainty in the room temperature of ± 5o during the experiments, giving a further uncertainty in resistance of ±1%.

The SRF multiplier increases the resistance by only 5% and so the uncertainty here is say ±0.5% of total resistance.

The leads to the capacitor contribute a surprisingly large 15% to the total resistance, but it is very difficult to calculate this accurately and the uncertainty to the total resistance is assumed to be ±1.5 %

These uncertainties are uncorrelated and so they can be added in quadrature, giving an overall uncertainty in the *calculations* of ±7%.

The measurements also have uncertainty and error. Measurements were of Q and carried-out on an AIM 4170 analyser. This measures Q by taking the two frequencies either side of resonance where the phase is 45o, and so the error here is related to the uncertainties in measuring frequency and phase. The analyser gives frequency to 6 places, for example 24.3652 MHz, but with a Q of say 1000, the bandwidth will be only 0.0244 MHz. Assuming the last figure is uncertain (i.e. ±0.0001) this will give an random uncertainty on the measurement of bandwidth of around ± 0.5%. The uncertainty on phase is not known, but if we assume ±1o, this will give a

would expect that the slope of the Rp vs. frequency graph of the other slide switches to be the similar that shown above.

B3: The loss tangent of an insulating material is the reciprocal of the Q of a capacitor made of that material. Some of the insulating materials listed below are used as front panels, detector stand bases, wire insulation and coil forms in crystal radio sets. A capacitor formed by the use of one of these materials, connected across a high impedance point and ground, will contribute a loss proportional to the loss tangent and capacitance.
Table 4 - Loss Tangent of some Insulating Materials
used in Crystal Radio Sets, Measured at 1.5 MHz.

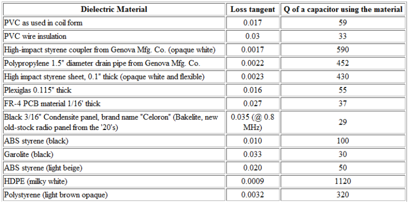

Dielectric Material	Loss tangent	Q of a capacitor using the material
PVC as used in coil form	0.017	59
PVC wire insulation	0.03	33
High-impact styrene coupler from Genova Mfg. Co. (opaque white)	0.0017	590
Polypropylene 1.5" diameter drain pipe from Genova Mfg. Co.	0.0022	452
High impact styrene sheet, 0.1" thick (opaque white and flexible)	0.0023	430
Plexiglas 0.115" thick	0.016	55
FR-4 PCB material 1/16' thick	0.027	37
Black 3/16" Condensite panel, brand name "Celoron" (Bakelite, new old-stock radio panel from the '20's)	0.035 (@ 0.8 MHz)	29
ABS styrene (black)	0.010	100
Garolite (black)	0.033	30
ABS styrene (light beige)	0.020	50
HDPE (milky white)	0.0009	1120
Polystyrene (light brown opaque)	0.0032	320

Note: GE's version of polycarbonate is called Lexan. A review GE's spec. sheets of various grades of Lexan show loss tangent values at 1.0 MHz ranging between 0.006 and 0.026 . Many grades are specified 0.01.

B4: Please see Parts 10 and 11 of Article #26 as well as Table 4. Also see Table 2 of Article #22 and Article #29.

B5: Sometimes, when working with high Q tank circuits, a need pops up for a fixed capacitor with a value between say, 100 and 1000 pF that will not degrade overall circuit Q. Generic NPO disc capacitors in that range usually have a Q of around 2000-3000 at 1 MHz. Table 4 shows some caps having higher Q values. The only downside to the high Q caps is that they are SMD types and require some skill when soldering pigtail leads to them to easy connecting to one's circuit. The capacitors were measured singly, in parallel or in series, aiming for values approximating 500 pF. This was for convenience in measurement. I used solid tinned copper wire having a diameter of about 0.010" for my pig-tails. The source for the strands was a piece of stranded hook-up wire.

Table 4 - Q of easily available capacitors in the 100 - 1000 pF range

	Type	**Value in pF**	**Voltage**	**Q at about 920 kHz**	**Mfg**	**Mfg. part number**
1	Polypropylene	Two 1k in series=500	630	2,000	Xicon (Mouser)	1431-6102K
2	Polystyrene	One 470	50	6,200	Xicon (Mouser)	23PS147
3	Generic NPO disc	One 220	500	2,800	3/8" dia.	NA
4	Multilayer, hi Q SMD	Two 1k in series=500	50	8,000	Murata	ERB32Q5C1H102JDX1L
5	Multilayer, hi Q SMD	One 470	100	18,000	Murata	ERB32Q5C2A471JDX1L
6	Multilayer, hi Q SMD	Two 220 in parallel=440	200	20,000	Murata	ERB32Q5C2D221JDX1L
7	Multilayer, hi Q SMD	Two100 in parallel=200	500	40,000	Murata	ERB32Q5C2H101JDX1L

Note: Solder flux contamination on a dielectric is the enemy of high Q because it usually provides a resistive leakage path. If one gets solder flux on the insulation of a variable cap or switch, remove it with a commercial flux remover. This is important when a DX crystal radio set is involved.

#24 Published: 03/25/2002; Revised: 06/10/2008

6.4. Typical Resonance

A typical resonance curve is shown below, from which the vector analyser calculates the Q.

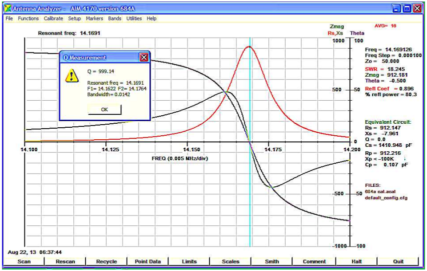
Antenna Analyzer - AIM 4170 version 684A
File
Functions
Calibrate
Setup
Markers
Bands
Utilities
Help
Zmag
Rs,Xs
Theta
1000
100
500
50
0
-500
-50
-1000
-100
14.100
14.125
14.150
14.175
14.200
FREQ (0.005 MHz/div)
Q Measurement
OK
Scan
Rescan
Recycle
Point Data
Limits
Scales
Smith
Comment
Halt
Quit

AVG= 16

Resonant freq: 14.1691

Q = 999.14

Resonant freq = 14.1691
F1= 14.1622 F2= 14.1764
Bandwidth= 0.0142

Freq = 14.169126
Freq Step = 0.000100
Zo = 50.000
SWR = 18.245
Zmag = 912.181
Theta = -0.500
Refl Coef = 0.896
% refl power = 80.3

Equivalent Circuit:
Rs = 912.147
Xs = -7.961
Q = 0.0
Cs = 1410.948 pF
Rp = 912.216
Xp < -100K
Cp = 0.107 pF

FILES:
684a cal.acal
default_config.cfg

Aug 22, 13 06:37:44

7. ACCURACY

To evaluate the likely uncertainty, it is useful to consider the individual contributions to the loss of the measurement circuit. At 10 MHz these are :

Copper Pipes (ignoring proximity and SRF) :	0.04 Ω
Proximity multiplier on above :	1.009
SRF multiplier on above :	1.048
Radiation resistance :	0.0005 Ω
Lead resistance :	0.0061 Ω

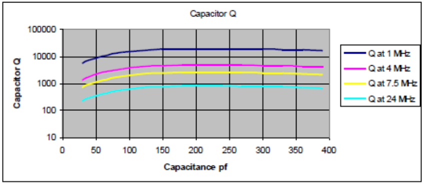

6.2. Test of Ground Effects

The measured Q was as large as 1000, and so it is possible that the floor of the room and other surfaces could be introducing loss. To evaluate this an aluminium sheet, approx 0.8mx 0.5m, was placed on the floor below the tuned circuit, a distance of about 0.8m. A measured Q of 953 was unaffected, to within the repeatability of the measurement.

6.3. Standard Capacitor

The measurement uncertainty could be reduced if a fixed capacitor could be made with a very low loss since this could be used as a standard to calibrate the system. However attempts to make such a capacitor were not successful in that it had a higher loss than the variable capacitor to be measured (see Appendix 2).

Capacitor Losses

Conrad Hoffman, 2012
http://conradhoffman.com/cap_losses.htm

Dielectrics

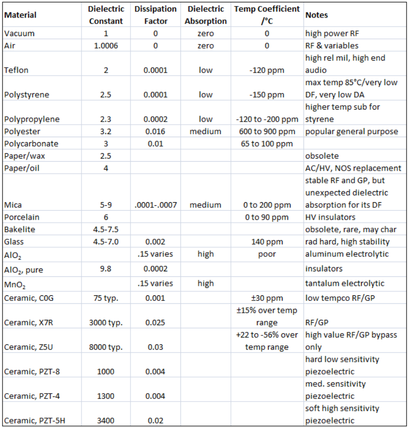

Material	Dielectric Constant	Dissipation Factor	Dielectric Absorption	Temp Coefficient /°C	Notes
Vacuum	1	0	zero	0	high power RF
Air	1.0006	0	zero	0	RF & variables
Teflon	2	0.0001	low	-120 ppm	high rel mil, high end audio
Polystyrene	2.5	0.0001	low	-150 ppm	max temp 85°C/very low DF, very low DA
Polypropylene	2.3	0.0002	low	-120 to -200 ppm	higher temp sub for styrene
Polyester	3.2	0.016	medium	600 to 900 ppm	popular general purpose
Polycarbonate	3	0.01		65 to 100 ppm	
Paper/wax	2.5				obsolete
Paper/oil	4				AC/HV, NOS replacement
Mica	5-9	.0001-.0007	medium	0 to 200 ppm	stable RF and GP, but unexpected dielectric absorption for its DF
Porcelain	6			0 to 90 ppm	HV insulators
Bakelite	4.5-7.5				obsolete, rare, may char
Glass	4.5-7.0	0.002		140 ppm	rad hard, high stability
AlO_2		.15 varies	high	poor	aluminum electrolytic
AlO_2, pure	9.8	0.0002			insulators
MnO_2		.15 varies	high		tantalum electrolytic
Ceramic, C0G	75 typ.	0.001		±30 ppm	low tempco RF/GP
Ceramic, X7R	3000 typ.	0.025		±15% over temp range	RF/GP
Ceramic, Z5U	8000 typ.	0.03		+22 to -56% over temp range	high value RF/GP bypass only
Ceramic, PZT-8	1000	0.004			hard low sensitivity piezoelectric
Ceramic, PZT-4	1300	0.004			med. sensitivity piezoelectric
Ceramic, PZT-5H	3400	0.02			soft high sensitivity piezoelectric

Capacitors are constructed of two or more electrodes, separated by a dielectric. The dielectric is commonly ceramic, plastic film, oiled paper, mica, or air. Each one has advantages and disadvantages in regards to dielectric constant, losses,

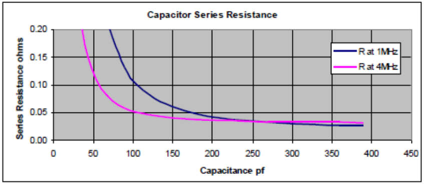

It is useful to know the contributions which each factor in Equation 6.2.1 makes to the overall resistance, and for f = 4 MHz this is shown below:

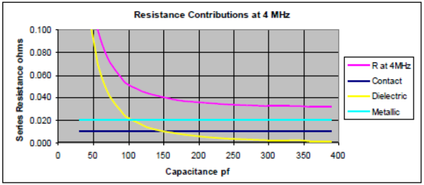

This shows that at low capacitance the dielectric loss dominates and at high capacitance the metallic loss dominates.

The capacitor Q is given by 1/ (ω C Rcap), and is given below:

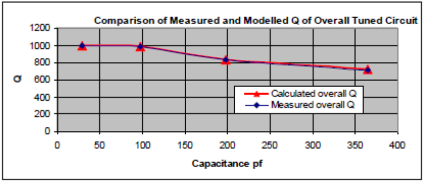

The agreement is within ±1.5%.

It should be noted that the frequency is not constant in the above graph, and is 24.5 MHz at 27pf, 10.2 MHz at 198 pf, 14 MHz at 98 pf and 7.56 MHz at 365 pf, all frequencies corresponding to resonance with the inductor. These combinations of capacitance and frequency give relatively high losses, so that the overall Q to be measured is not excessive (below 1000). Also, when the capacitance is set to its minimum of 27 pf the dielectric loss is dominant, even at 24.5 MHz, and so allows the constant α to be determined fairly accurately. Similarly at 365 pf and 7.56 MHz the metallic loss and contact loss dominate permitting these to be determined.
So the series resistance of this capacitor is given by:

Rcap = 0.01+ 800 / (f C2) + 0.01(f 0.5)   6.1.2

This equation is plotted below, for f = 1 MHz and 4 MHz:

---

temperature coefficient, and, of course, cost. High dielectric constants result in smaller capacitors, but usually with poorer properties than the lower constant materials. Some properties of various dielectric materials are shown above:

The data in this table comes from many sources, not all of which are in agreement. If you have better information, want to fill in any blanks or have additional entries, please email me with the data and its source.

Low Frequency Losses

How do we specify loss? If you ask most engineers about capacitor loss, they will mumble something about "loss tangent", then disappear for an emergency coffee refill. There are several different ways of expressing capacitor losses, and this often leads to confusion. They are all very simply related, as shown below.

If you drive a perfect capacitor with a sine wave, the current will lead the voltage by exactly 90°. The capacitor gives back all the energy put into it on each cycle. In a real capacitor, the current will lead the voltage by a bit less than 90°. The capacitor will dissipate a small fraction of the energy put into it as heat. Real capacitors can be modeled, at least to a first order, as a perfect capacitor in series with a resistance. This resistance is referred to as the effective series resistance or ESR, and is only valid for a single frequency. For the examples below, assume a 0.47µF capacitor, driven at 5000 Hz, 35 volts RMS, and showing a phase angle of -89.5°. Note that the relationships shown only apply to sine waves. Capacitors can also be modeled using parallel elements, but we'll limit our discussion to the series model. You'll find a collection of "handy formulas" on this site that includes conversions between series and parallel models, plus other useful data.

Dissipation factor, or "D" as it is usually marked on test bridges, is the tangent of the difference between the phase angle of a perfect capacitor, and the capacitor in question. In our example, -90°- -89.5°=-0.5° The tangent of -0.5° is -0.00873. We take the absolute value so D=0.00873. Since this number is directly read from most test bridges, other parameters are often calculated from it. It is also known as the loss tangent and is sometimes expressed as a percentage. D=.873%

This is probably a good time to mention that by general agreement capacitive reactances are negative and inductive reactances are positive. A vector impedance meter would display the phase angle of the capacitor in question as "-89.5". Xc, the capacitive reactance would be -1/(2*pi*f*c) or -67.7255 ohms. The sign is often omitted.

Power factor, or "PF", is less common than it used to be, at least as applied to capacitors. It is the cosine of the phase angle itself. In our example, COS(-89.5)=0.00873. Note that for small angles, PF is essentially equal to D, and both are approximately equal to the phase angle expressed in radians. For large angles the situation is quite different. A power factor of 1 is 100% resistive loss whereas D can exceed 1 and approaches infinity for 100% loss.

Q is the quality factor, a dimensionless figure of merit. It is the reciprocal of D. In our example, 1/.0087=114.58, so Q=114.58.

6.1. Procedure and Initial Measurements

The Q of the total circuit was measured with an AIM 4170 network analyser. This analyser could have been connected across the variable capacitor, but the resistance of the tuned circuit at resonance is very high at this point (around 60kΩ) and the analyser is not very accurate at this level. In addition the analyser will introduce stray capacitance which will upset the measurements. The analyser was therefore connected to a 'tap', on the coil. For this the ground terminal of the analyser was connected at a point which was assumed to be at zero potential wrt ground i.e. half-way across the end pipe forming the transmission-line short. The active port of the analyser was connected to one of the copper pipes at a distance of 150 mm from the end of the pipe, and the connection lead taken down the centre of the transmission line, forming a small loop (see Photograph 2). This gave a resonant resistance at the analyser of around 700 Ω (at 14 MHz). The measurement of Q was not particularly sensitive to the exact dimensions of this coupling loop (≈ ± 5%), but the dimensions above gave the highest Q.

The measured Q was compared with the theoretical Q given by:

$$Q_{theory} = \omega L / R_s \qquad 6.1.1$$

Where L is as measured, or as given by Equation 4.4.1
Rs is given by Equation 5.1

The capacitor resistance Rc was assumed to conform to Equation 2.1, and the constants Rs, α and β were adjusted in the model to give the best correlation between the measured Q and the theoretical Q, and this was achieved with $R_s = 0.01$, $\alpha = 800$, and $\beta = 0.01$, to give the following :

Rr is the radiation resistance = 31200 (A/λ2)2
Rleads is the resistance of the leads connecting the capacitor
Rcap is series resistance of the capacitor being measured, see below.

It was assumed that the solder joints connecting the pipes together had negligible resistance, as did those connecting the metal tape to the capacitor.

Plumbing pipes were used here and the copper in these has a higher resistivity than normal copper wire, and lies between 1.92 10-8 and 2.30 10-8 Ωm (ref 6). The average of these is 2.11 10-8 and this was used here. The relative permeability μr of pure copper is unity, and it is assumed to be the same for plumbing copper.

The frequency fr was derived from the measurement of inductance as 60 MHz (para 4.3).

The capacitor was connected to the copper pipes via copper straps, having a combined length of 80 mm and a width of 11 mm and thickness 0.25 mm. The resistance of these was calculated from :

$$Rstrip = Rwall\ l / [2(w+ t)]\ Kfringing \tag{5.2}$$

where Kfringing = 1.06 +0.22 Ln w/t +0.28 (t/w)2

The fringing factor allows for the increased resistance due to the current crowding towards the edges of the strip.

6. MEASUREMENTS

Effective series resistance, or "ESR" is the value of resistance in series with a perfect capacitor that produces the phase angle error. It can be calculated by dividing D by ωC (2 pi F C). In our example, .0087/(6.28*5000*.00000047)=0.589, so ESR=0.589 ohms.

Capacitive reactance is the negative reciprocal of ωC. -1/(6.28*5000*.00000047)=-67.725, so X=-67.725 ohms.

Total impedance of a capacitor is obtained by taking the absolute value of the root sum of the squares of capacitive reactance and ESR. (67.72522+0.58922)1/2=67.727, so Z=67.727 ohms.

Capacitor current is the RMS voltage divided by the total impedance. 35/67.7=0.52 amps.

Power dissipation in the ESR component is calculated from the RMS voltage times current times the ratio of ESR to total impedance. 35*.52*(.589/67.727)=0.16 watts. Or, use I^2 times ESR. The resulting temperature rise depends on the size and heat sinking of the capacitor.

Verifying capacitor ESR on the bench requires both care and good instrumentation. Because ESR is usually small, test lead resistance and poor connections can easily contribute more resistance than the capacitor. Noise pickup from long leads or a hostile environment can make readings unstable. Note that neither D nor ESR is constant with frequency. ESR will decrease rapidly as frequency increases, then drop at a much slower rate above about 1 kHz, until it starts rising again at self resonance. Be sure to make measurements in the same frequency and voltage range that the parts will be used in- test it the way you use it. Be aware that in simulations ESR is a constant unless you take special measures to make it frequency dependent.

Dielectric Absorption is another imperfection. Briefly, the dielectric refuses to give up its full charge, and a previously discharged capacitor will self charge. This can be modeled with additional C-R pairs in parallel with the main capacitor. Dielectric absorption is a particular problem in capacitors used in integrators. There is some debate as to its importance in audio applications. Much that has been written about dielectric absorption remains obscure; at least two standard tests exist, but there is very little published data for specific parts.

Other Effects Capacitors are not 100% linear and may contribute a small amount of distortion to signals, particularly the electrolytic types. Capacitance may vary with DC voltage. Each dielectric type will have some characteristic temperature coefficient. Some capacitors are microphonic, and may also "sing" when driven at high voltages.

This relationship is useful later in determining the change of *resistance* with frequency, which Welsby gives as $1/[1-(f/fr)2]2$.

4.4. Calculated Inductance

In the initial design of the experiments it is useful to be able to calculate the approximate inductance, and the following has an accuracy of around 5% :

$$L = \mu o\ p/(2\pi)\ [\ \mathrm{Ln}\ (2p/r) + 0.25 - \mathrm{Ln}\ (p2/A)]\ SL \qquad 4.4.1$$

where p is the perimeter and A the area
SL is the inductance SRF factor $1/[1-(f/fr)2]$
$fr = 0.85(150/\ length)$

5. CALCULATED TOTAL LOSS OF RESONANT CIRCUIT

The series resistance of the resonant circuit, including the capacitor being evaluated, is given by :

$$Rs = Rwall\ [l pipes\ /\ (\pi\ dw)]\ SR\ P + Rr + Rleads + Rcap \qquad 5.1$$

where dw is the diameter of the copper pipes
$Rwall = (\rho\ \pi\ \mu o\ \mu r\ f)0.5$
ρ is the resistivity of the pipes (see below)
$\mu o = 0.4\pi\ \mu H/m$
µr is the relative permeability of the copper pipes (see below)
SR is the resistance SRF factor $=1/[1-(f/fr)2]2$
P is the proximity factor $1/\ [1- (dw/D)2]0.5$ (see Moullin ref1)
D is the spacing of the pipes (centre to centre)

4.3. Inductance changes with Frequency

The inductance of the coil changes with frequency and it was therefore measured over the frequency range of interest, and also at some lower frequencies to establish the low frequency inductance, with the following results (f is in MHz and L in μH) :

f	L	Equation 4.3.1	Error
0.5	1.1947	1.1948	0.0%
5	1.1988	1.2031	0.4%
7.459	1.2096	1.2135	0.3%
10.21	1.2464	1.2303	-1.3%
14.098	1.3147	1.2645	-3.8%
24.426	1.4322	1.4320	0.0%

It is known that the change of inductance with frequency follows approximately an equation of the form (Welsby ref 5 p37):

$$L = Lo / [1-(f/fr)2] \qquad 4.3.1$$

where Lo is the low frequency inductance
f is the frequency
fr is the Self Resonant Frequency

This equation becomes less accurate as the Self Resonant Frequency is approached, but its accuracy can be improved if a value for fr is chosen which is slightly lower than the actual SRF. A good fit with the measured inductance above is given by fr = 60 MHz (about 85% of the actual SRF), and the results of the equation with this value are shown in the table above.

Lossy Capacitors

Dr. Gary L. Johnson December 10, 2001
Manhattan,Kansas

http://www.eece.ksu.edu/~gjohnson/tcchap3.pdf

1 Dielectric Loss

Capacitors are used for a wide variety of purposes and are made of many different materials in many different styles. For purposes of discussion we will consider three broad types, that is, capacitors made for ac, dc, and pulse applications. The ac case is the most general since ac capacitors will work (or at least survive) in dc and pulse applications, where the reverse may not be true.

It is important to consider the losses in ac capacitors. All dielectrics (except vacuum) have two types of losses. One is a conduction loss, representing the flow of actual charge through the dielectric. The other is a dielectric loss due to movement or rotation of the atoms or molecules in an alternating electric field. Dielectric losses in water are the reason for food and drink getting hot in a microwave oven.

One way of describing dielectric losses is to consider the permittivity as a complex number,
defined as

$$\epsilon = \epsilon' - j\epsilon'' = |\epsilon| e^{-j\delta} \quad (1)$$

where
e’ = ac capacitivity
e” = dielectric loss factor

(Photograph 2). The parallel pipes and the connection strips will subsequently be referred-to as the coil.

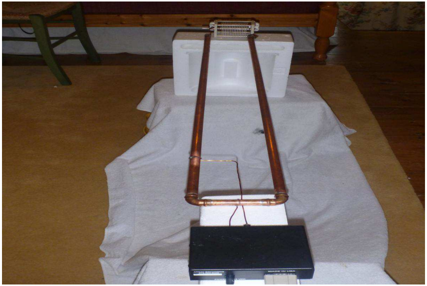

Photograph 2 : The Measurement Apparatus

4.2. Inductance of Capacitor

The total inductance of the tuned circuit will include the inductance of the capacitor. The value for this is not known and is difficult to measure, however Field & Sinclair (ref 3) report that their measurements of capacitor inductance ‘compare well with that calculated from the dimensions of the leads, stator supports, and rotor shaft’. So for the capacitor measured here, its inductance is assumed to be represented by a rod of 7.75 mm diameter and 90 mm long, and a rod of these dimensions was inserted during the inductance measurements. Assuming the inductance of this rod is given by L = u ***l*** /2pi [ln (2 ***l*** /a) - 1] henrys, then this will have a free space inductance of 0.05 μH, about 3.5% of the total inductance.

require a large number of measurements. For instance, if there are 5 capacitance settings and each is measured at 5 frequencies this gives a total of 25 separate measurements, each having to be done with high precision. In addition more than one coil would be necessary to cover the frequency range and, for instance, Moullin used 5 coils. However, with the benefit of Equation 2.1 the task is considerably reduced, since it is now only necessary to carry-out sufficient measurements to determine the factors Rs, α and β. In this respect notice in Equation 2.1 that if C is small the dielectric loss will tend to dominate, and if C is large the metallic loss and contact resistance will tend to dominate. So measurements should be made at the minimum and maximum capacitance at the least, and also with other values for added confidence.

Only one coil was used for the measurements, and this gave resonance between 7.5 and 24 MHz with the variable capacitor. On reflection some measurements at a lower frequency may have improved the accuracy, particularly of the dielectric loss, but this would have required a second larger coil.

4. INDUCTANCE

4.1. The Coil

The inductance coil was made from two copper pipes, each 15.03 mm outside diameter and 0.985 meter long. These were spaced at 115 mm (centre to centre), and shorted at one end with a section of copper pipe.

The capacitor was connected across the open ends with two short flexible copper strips, the whole forming a tuned circuit

---

δ = dielectric loss angle

Capacitance is a complex number C∗ in this definition, becoming the expected real number C as the losses go to zero. That is, we define

$$C^* = C - jC'' \qquad (2)$$

One reason for defining a complex capacitance is that we can use the complex value in any equation derived for a real capacitance in a sinusoidal application, and get the correct phase shifts and power losses by applying the usual rules of circuit theory. This means that most of our analyses are already done, and we do not need to start over just because we now have a lossy capacitor.

Equation 1 expresses the complex permittivity in two ways, as real and imaginary or as magnitude and phase. The magnitude and phase notation is rarely used. Instead, people usually express the complex permittivity by e and tan δ,where

$$\tan\delta = \frac{\epsilon''}{\epsilon'} \qquad (3)$$

where tan δ is called either the loss tangent or the dissipation factor DF.

The real part of the permittivity is defined as

$$\epsilon' = \epsilon_r \epsilon_o \qquad (4)$$

where er is the dielectric constant and eo is the permittivity of free space.

Dielectric properties of several different materials are given in Table 1 [4, 5]. Some of these materials are used for capacitors, while others may be present in oscillators or other devices where dielectric losses may affect circuit performance. The dielectric constant and the dissipation factor are given at two frequencies, 60 Hz and 1 MHz. The righthand column of Table 1 gives the approximate breakdown voltage of the material in V/mil, where 1 mil = 0.001 inch. This would be for thin layers where voids and impurities in the dielectrics are not a factor. Breakdown usually destroys a capacitor, so capacitors must be designed with a substantial safety factor.

It can be seen that most materials have dielectric constants between one and ten. One exception is barium titanate with a dielectric constant greater than 1000. It also has relatively high losses which keep it from being more widely used than it is.

We see that polyethylene, polypropylene, and polystyrene all have small dissipation factors. They also have other desirable properties and are widely used for capacitors. For high power, high voltage, and high frequency applications, such as an antenna capacitor in an AM broadcast station, the ruby mica seems to be the best.

Each of the materials in Table 1 has its own advantages and disadvantages when used in a capacitor. The ideal dielectric would have a high dielectric constant, like barium titanate, a low dissipation factor, like polystyrene, a high breakdown voltage, like mylar, a low cost, like aluminum oxide, and be easily fabricated into capacitors. It would also be perfectly stable, so the capacitance would not vary with temperature or

---

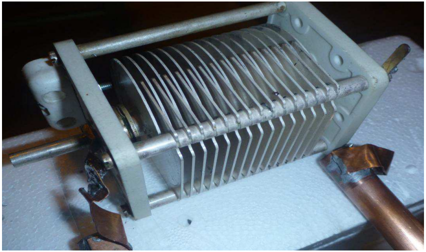

Photograph 1 : Variable Capacitor

3. MEASUREMENT PRINCIPLES

The measurement principle is that the reactance of the capacitor is tuned-out with a low loss coil, measurements are then made of the series resistance of this tuned circuit, and the calculated resistance of the coil subtracted to give that of the capacitor.

In practice the resistance values are too low to be measured accurately (≈ 0.04Ω), and so the Q (i.e. bandwidth) of the tuned circuit is measured at resonance and compared with the calculations of the Q, assuming a given loss in the capacitor. This assumed capacitor loss is then adjusted manually in the theoretical model to make the overall calculated Q agree with the measured Q.

To fully characterise the capacitor over its full capacitance range and over its useful frequency range would normally

So the measured value of Rs lay between 4.5 mΩ and 50 mΩ. The low value in this range is consistent with the value of contact resistance measured by Bock (ref 4) who gives 1 mΩ for silver plated contacts, but the high values are too large to be due to contact resistance unless the contacts were particularly dirty or worn and this seems unlikely. It is significant then that Field & Sinclair (ref 3) say that this resistance is 'joulean loss in the metal structure'. Such metallic loss will be frequency dependent due to skin effect (i.e. proportional to f-0.5), whereas contact resistance seems to be independent of frequency. So for capacitors with a long metallic path Equation 1.1 an additional term needs to be added:

$$Rcap = Rs + \alpha /(f\ C2) + \beta\ (f\ 0.5) \qquad 2.1$$

The factors Rs, α and β are assumed to be constant for any capacitor, and it is the objective here to derive them for the capacitor shown in Photograph 1 (also the capacitor shown in Annex 1). This capacitor has square ceramic ends with sides of 65mm, and an overall length of 90mm, not including the protruding shaft. Contact to the rotating shaft is at one end only, via a spring loaded washer which is stationary and makes sliding contact with a lip on the shaft. All metalwork is silver plated. The capacitor has a range of 25 – 390 pf.

voltage. No such dielectric has been discovered so we must apply engineering judgment in each situation, and select the capacitor type that will meet all the requirements and at least cost.

Capacitors used for ac must be unpolarized so they can handle full voltage reversals. They also need to have a lower dissipation factor than capacitors used as dc filter capacitors, for example. One important application of ac capacitors is in tuning electronic equipment. These capacitors must have high stability with time and temperature, so the tuned frequency does not drift beyond some specified amount.

Another category of ac capacitor is the motor run or power factor correcting capacitor. These are used on motors and other devices operating at 60 Hz and at voltages up to 480 V or more. They are usually much larger than capacitors used for tuning electronic circuits, and are not sold by electronics supply houses. One has to ask for motor run capacitors at an electrical supply house like Graingers. These also work nicely as dc filter capacitors if voltages higher than allowed by conventional dc filter capacitors are required.

The term power factor PF may also be defined for ac capacitors. It is given by the expression

$$\mathrm{PF} = \cos\theta \qquad (5)$$

where θ is the angle between the current flowing through the capacitor and the voltage across it.
The capacitive reactance for the sinusoidal case can be defined as

$$X_C = \frac{1}{\omega C} \quad (6)$$

where ω =2 πf rad/sec, and f is in Hz.

In a lossless capacitor, e"= 0, and the current leads the voltage by exactly 90o. If e" is greater than zero, then the current has a component in phase with the voltage.

$$\cos\theta = \frac{\epsilon''}{\sqrt{(\epsilon'')^2 + (\epsilon')^2}} \quad (7)$$

For a good dielectric, e'>>e", so

$$\cos\theta \approx \frac{\epsilon''}{\epsilon'} = \tan\delta \quad (8)$$

Therefore, the term power factor is often used interchangeably with the terms loss tangent or dissipation factor, even though they are only approximately equal to each other.

We can define the apparent power flow into a parallel plate capacitor as

$$S = VI = \frac{V^2}{-jX_C} = jV^2\omega C^* = jV^2\frac{\omega A}{d}(\epsilon' - j\epsilon'') = V^2\frac{\omega A}{d}\epsilon_r\epsilon_o(j + \mathrm{DF})$$

(9)

By analogy, the apparent power flow into any arbitrary capacitor is

that proximity effect was minimal, and shorted at the far end. It therefore resembled a shorted two-wire transmission line, and the capacitor to be measured was connected to the open ends. This type of inductor is used here and is shown in Photograph 2.

A number of authors have measured the loss of variable capacitors (refs 1, 2, 3), and found that the series resistance conformed to the following equation :

Rcap = Rs + α /(f C2) ohms 1.1

where f is the frequency
C is the capacitance

The first factor Rs is a constant resistance, often accredited to the contact resistance of the slip rings, or equivalent. The second factor is due to the loss in the insulators, where again factor α is a constant. Jackson (ref 2) summarises his own measurements and those of previous workers as follows :

Rcap = 0.0045 + 0.028 [106/(f C2)] ohms (Jackson) 1.2
Rcap= 0.05 + 0.091 [105/(f C2)] ohms (Moullin) 1.3
Rcap= 0.007 + 0.012[105/(f C2)] +10/(f 2C2) ohms (Dye) 1.4
Rcap= 0.02 + 0.022 [103/(f C2)] ohms (Wilmotte) 1.5

where f is in MHZ and C in pf

Notice that Dye has an additional term related to a constant resistance across the dielectric, but this term is only significant at very low frequencies, and so is not considered here.

2. EVALUATION OF LOSS COMPONENTS

Measuring the Loss in Variable Air Capacitors

Alan Payne 2013

http://g3rbj.co.uk/wp-content/uploads/2013/10/Measurements_of_Loss_in_Variable_Capacitors_issue_2.pdf

The resistive loss of variable air capacitors is difficult to measure because they have a very high Q. The method described here uses a twin-wire transmission line made from copper pipe as the inductor to tune-out the reactance of the capacitor.

Previous authors have shown that the capacitor loss comes partly from resistive loss and partly from dielectric loss in the insulating supports. They have attributed the resistive loss to that in the rotating contacts, but it is shown that there is also a major contribution from the metallic part of the structure, although the plates themselves have negligible resistance.

1. INTRODUCTION

The loss resistance of high quality variable capacitors is difficult to measure because of the presence of the very high reactance. This must be tuned-out with an inductor, which must be of very high Q if its resistance is not to dominate the measurements, and its own resistance must be calculable to a high degree of accuracy, because this cannot be independently measured. Conventional helical coils cannot be used therefore because they do not have a sufficiently high Q, and, more importantly their loss resistance cannot be calculated with sufficient accuracy. Moullin (ref 1) solved this problem by making his inductor from two parallel conductors, spaced so

---

$$S = P + jQ = V^2\omega C(j + \mathrm{DF}) \quad (10)$$

Table 1: Dielectric Constant er, Dissipation Factor DF and Breakdown Strength Vb of selected materials.

Material	ϵ_r 60 Hz	ϵ_r 10^6 Hz	DF 60 Hz	DF 10^6 Hz	V_b V/mil
Air	1.000585	1.000585	-	-	75
Aluminum oxide	-	8.80	-	0.00033	300
Barium titanate	1250	1143	0.056	0.0105	50
Carbon tetrachloride	2.17	2.17	0.007	<0.00004	-
Castor oil	3.7	3.7	-	-	300
Glass, soda-borosilicate	-	4.84	-	0.0036	-
Heavy Soderon	3.39	3.39	0.0168	0.0283	-
Lucite	3.3	3.3	-	-	500
Mica, glass bonded	-	7.39	-	0.0013	1600
Mica, glass, titanium dioxide	-	9.0	-	0.0026	-
Mica, ruby	5.4	5.4	0.005	0.0003	-
Mylar	2.5	2.5	-	-	5000
Nylon	3.88	3.33	0.014	0.026	-
Paraffin	2.25	2.25	-	-	250
Plexiglas	3.4	2.76	0.06	0.014	-
Polycarbonate	2.7	2.7	-	-	7000
Polyethylene	2.26	2.26	<0.0002	<0.0002	4500
Polypropylene	2.25	2.25	<0.0005	<0.0005	9600
Polystyrene	2.56	2.56	<0.00005	0.00007	500
Polysulfone	3.1	3.1	-	-	8000
Polytetrafluoroethylene(teflon)	2.1	2.1	<0.0005	<0.0002	1500
Polyvinyl chloride (PVC)	3.2	2.88	0.0115	0.016	
Quartz	3.78	3.78	0.0009	0.0001	500
Tantalum oxide	2.0	-	-	-	100
Transformer oil	2.2	-	-	-	250
Vaseline	2.16	2.16	0.0004	<0.0001	-

The power dissipated in the capacitor is

$$P = V^2\omega C'' = V^2\omega C(\mathrm{DF}) \quad (11)$$

Example

Find the real and reactive power into a ruby mica capacitor with area A = 0.03m2, and a dielectric thickness d = 0.001 m, if the voltage is 2000 V (rms) at a frequency f = 1 MHz.

$$S = V^2 \frac{\omega A}{d} \epsilon_r \epsilon_o (j + \mathrm{DF})$$

$$= (2000)^2 \frac{2\pi(10^6)(0.03)}{0.001}(5.4)(8.854 \times 10^{-12})(j + 0.0003)$$

$$= j36040 + 10.8$$

The capacitor is absorbing 36040 capacitive VARs (Volt Amperes Reactive) and 10.8 Watts. The real power of 10.8 W appears as heat and must be removed by appropriate heat sinks.

The real power dissipation in a capacitor varies directly with frequency if the dissipation factor remains constant, and also with the square of the voltage. At low frequencies, the voltage limit is determined by the dielectric strength. At high frequencies, however, the voltage limit may be determined by the ability of the capacitor to dissipate heat. If the ruby mica capacitor in the previous example could safely dissipate only 10 W, but was to be used at 5 MHz, the operating voltage must be reduced from 2000 V to keep the losses in an acceptable range.

The dissipation factor varies significantly with frequency for some materials in Table 1. The actual variation must be determined experimentally. If interpolation or extrapolation seems necessary to find the loss at some frequency not given in

---

Test System Safety

Many electrical test systems or instruments are capable of measuring or sourcing hazardous voltage and power levels. It is also possible, under single fault conditions (e.g., a programming error or an instrument failure), to output hazardous levels, even when the system indicates no hazard is present. These high-voltage and power levels make it essential to protect operators from any of these hazards at all times. It is the responsibility of the test system designers, integrators and installers to make sure operator and maintenance personnel protection is in place and effective. Protection methods include:

* Design test fixtures to prevent operator contact with any hazardous circuit.
* Make sure the device under test is fully enclosed to protect the operator from any flying debris.
* Double insulate all electrical connections that an operator could touch. Double insulation ensures the operator is still protected, even if one insulation layer fails.
* Use high-reliability, fail-safe interlock switches to disconnect power sources when a test fixture cover is opened.
* Where possible, use automated handlers so operators do not require access to the inside of the test fixture.
* Provide proper training to all users of the system so they understand all potential hazards and know how to protect themselves from injury.

Dale Cigoy is lead applications engineer at Keithley Instruments (keithley.com).

suitable soak period. After the capacitors have been tested, the voltage source should be set to zero, and then some time allowed so the capacitors can discharge before they are removed from the test fixture.

Note that in Figure 4 the capacitors have a discharge path through the normally closed contacts of the relays. To prevent electric shock, test connections must be configured in such a way that the user cannot come in contact with the conductors, connections or DUT. Safe installation requires proper shielding, barriers and grounding to prevent contact with conductors.

More complex test systems that combine leakage measurement with capacitance measurements, dielectric absorption and other tests, if desired, are possible. A simplified schematic of such a test system using an LCZ bridge and a picoammeter with a voltage source is shown in Figure 5.

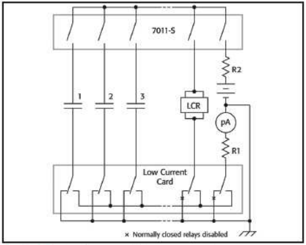

FIGURE 5. Capacitance and IR measurement system.

the Table, the best assumption would be that DF varies linearly with log f. That is, if DF = 0.02 at f = 102 and 0.01 at f = 106, a reasonable assumption at f = 104 would be that DF = 0.015.

The basic circuit model for a capacitor is shown in Fig. 1. Any conductor, whether straight or wound in a coil, has inductance, and the capacitor inductance is represented by a series inductance Ls. The effects of conductor resistance and dielectric losses are represented by a series resistance Rs. Leakage current through the capacitor at dc flows through a parallel resistance Rp. In some manufacturer's databooks, these are called ESL, ESR, and EPR, where

Ls = ESL = equivalent series inductance
Rs = ESR = equivalent series resistance
Rp = EPR = equivalent parallel resistance

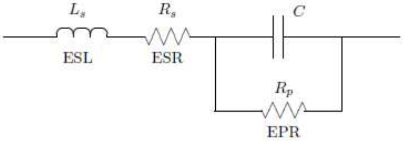

Figure 1: Equivalent Circuit for a Capacitor

This circuit indicates that every capacitor has a self-resonant frequency, above which it becomes an inductor. This is certain to puzzle a student making measurements on a capacitor above this frequency if the student is not aware of this fact. Rs is readily measured by applying this frequency to a capacitor, measuring the voltage and current, and calculating the ratio. The capacitive and inductive reactances cancel at the resonant frequency, leaving only Rs to limit the current. The resistance

Rp will always be much larger than the capacitive reactance at the resonant frequency, so this resistance can be neglected for this computation.

The self-resonant frequency of a high capacitance unit is lower than that for a low capacitance unit. Hence, in some circuits we will see two capacitors in parallel, say a 10μF in parallel with a 0.001μF capacitor as shown in Fig. 2. At first glance, this seems totally unnecessary. However, the larger capacitor is used to filter low frequencies, say in the audio range, while the small capacitor filters the high frequencies which are above the self-resonant frequency of the large capacitor.

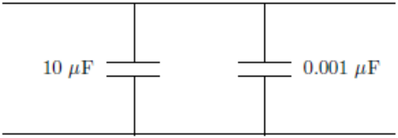

Figure 2: Capacitors in Parallel to Filter Two Different Frequencies

An example of the variation of Rs and self-resonant frequency fres with capacitance value and rated voltage is given in Table 2. These are metallized polypropylene capacitors designed for switch-mode power supplies by the company Electronic Concepts, Inc. This application requires a low Rs and a high fres so capacitors designed for other applications will tend to have higher values of Rs and lower values of fres.

---

that can integrate easily with a switching system capable of higher channel counts.

Configuration examples.

Producing enough useful data for statistical analysis requires testing a large quantity of capacitors quickly. Obviously, performing these tests manually is impractical, so an automated test system is required. Figure 4 illustrates such a system, which employs an electrometer with built-in voltage source, as well as a switching mainframe that houses a low current scanner card and a Form C switching card. In this test setup, a single instrument provides both the voltage sourcing and low current measurement functions. A computer controls the instruments to perform the tests automatically.

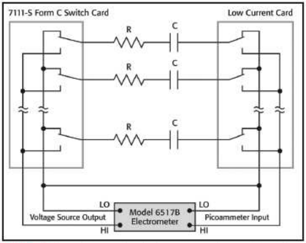

FIGURE 4. Example capacitor leakage test system configuration.

One set of switches is used to apply the test voltage to each capacitor in turn; a second set of switches connects each capacitor to the electrometer's picoammeter input after a

Test Hardware Considerations

A variety of considerations go into the selection of the instrumentation used when measuring capacitor leakage:

- Although it is certainly possible to set up a system with a separate voltage source, an integrated one simplifies the configuration and programming process significantly, so look for an electrometer or picoammeter with a built-in variable voltage source. A continuously variable voltage source calculates voltage coefficients easily. For making high-resistance measurements on capacitors with high voltage ratings, consider a 1000 V source with built-in current limiting. For a given capacitor, a larger applied voltage within the voltage rating of the capacitor will produce a larger leakage current. Measuring a larger current with the same intrinsic noise floor produces a greater signal-to-noise ratio and, therefore, a more accurate reading.
- Temperature and humidity can have a significant effect on high-resistance measurements, so monitoring, regulating and recording these conditions can be critical to ensuring measurement accuracy. Some electrometers monitor temperature and humidity simultaneously. This provides a record of conditions, and permits easier determination of temperature coefficients. Automatic time-stamping of readings provides a further record for time-resolved measurements.
- Incorporating switching hardware into the test setup allows automating the testing process. For small batch testing in a lab with a benchtop test setup, consider an electrometer that offers the convenience of a plug-in switching card. For testing larger batches of capacitors, look for an instrument

The operating frequency range will obviously be less than the self-resonant frequency. This could easily be as low as a few kHz for large motor run capacitors.

Table 2: Capacitor Resistance and Self-resonant Frequency for Electronic Concepts Type 5MP Metallized Polypropylene capacitors.

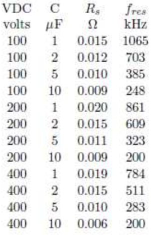

VDC volts	C μF	R_s Ω	f_{res} kHz
100	1	0.015	1065
100	2	0.012	703
100	5	0.010	385
100	10	0.009	248
200	1	0.020	861
200	2	0.015	609
200	5	0.011	323
200	10	0.009	200
400	1	0.019	784
400	2	0.015	511
400	5	0.010	283
400	10	0.006	200

Capacitors will always be rated for a working voltage as well as a specific value of capacitance. This voltage will always be well under the breakdown voltage of the dielectric. It may be specified either as a dc voltage or an ac voltage, depending on the application. Motor run capacitors are always operated on ac, so the voltage is specified as, say, 370 or 480 VAC. They can also be operated on dc, with a dc rating of at least √2 times the ac rating. Electrolytic capacitors that can only be operated on dc will have their working voltage expressed as WVDC. This is the maximum dc voltage, plus the peak of the ac ripple voltage, that should be continuously applied to a capacitor to prevent excessive deterioration and aging. Capacitors that are sometimes used on ac, and sometimes on dc, usually have working voltages expressed as WVDC, as was the case for the polypropylene capacitors in Table 2.

We will now present some more detailed background information on several of the dielectrics shown in Table 1.

Mica is a natural material that can be easily split into thin layers. It is very stable and does not deteriorate with age. The maximum capacitance is on the order of 0.03 μF. Mica capacitors tend to be quite expensive and the larger sizes are used only in critical applications like radio transmitters.

Glass capacitors were first developed during World War II as a replacement for mica capacitors when supplies of mica were threatened. Glass capacitors exhibit excellent long-term parametric stability, low losses, and can be used at high frequencies. They are used in aerospace applications where capacitance must not vary. The maximum size is limited to about 0.01μF

Paper capacitors use kraft paper impregnated with non-ionized liquid electrolyte as the dielectric between aluminum foils. These capacitors can provide large capacitances and voltage ratings but tend to be physically large. The thickness of the paper layers and the particular electrolyte used can be varied to produce a wide range of electrical characteristics, which is the reason for not listing paper in the dielectrics of Table 1.

Ceramic capacitors are made with one of a large number of ceramic materials, which include aluminum oxide, barium titanate, and porcelain. These are very widely used as bypass capacitors in electronic circuits. The older style is a single

The series resistor also adds Johnson noise – the thermal noise created by any resistor – to the measurement. At room temperature, this is roughly 6.5×1010 amps, p-p. The current noise in a 1 TΩ feedback resistor at a typical 3 Hz bandwidth would be ~8×10-16 A. When measuring an insulation resistance of 1016Ω at 10 V, the noise current will be 80% of the measured current.

Alternate test circuit.

Greater measurement accuracy can be achieved by including a forward-biased diode (D) in the circuit (Figure 3). The diode acts like a variable resistance, low when the charging current to the capacitor is high, then increases in value as the current decreases with time. This allows the series resistor used to be much smaller because it is only needed to prevent overload of the voltage source and damage to the diode if the capacitor is short-circuited. The diode used should be a small signal diode, such as a 1N914 or a 1N3595, but it must be housed in a light-tight enclosure to eliminate photoelectric and electrostatic interference. For dual-polarity tests, two diodes should be used back to back in parallel.

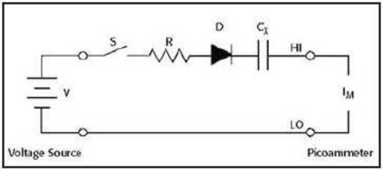

FIGURE 3. Alternative test circuit that incorporates a small-signal diode.

1MΩ·µF to 100 MΩ·µF. For example, a 4.7 µF aluminum cap specified as 50 MΩ·µF is guaranteed to have at least 10.6 MΩ insulation resistance.

Capacitor leakage test method.

Figure 2 illustrates a general circuit for testing capacitor leakage. Here, the voltage is placed across the capacitor (CX) for the soak period; then the ammeter measures the current after this period has elapsed. The resistor (R), which is in series with the capacitor, serves two important functions. First, it limits the current, in case the capacitor becomes shorted. Second, the decreasing reactance of the capacitor with increasing frequency will increase the gain of the feedback ammeter. The resistor limits this increase in gain to a finite value. A reasonable value is one that results in an RC product from 0.5 to 2 sec. The switch (S), while not strictly necessary, is included in the circuit to allow control over the voltage to be applied to the capacitor.

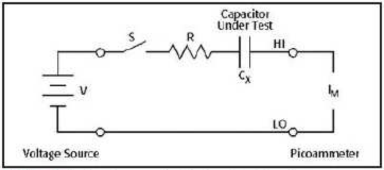

FIGURE 2. A simple capacitor leakage test circuit.

layer of dielectric separating two conducting plates and packaged in a small disc. The newer style is the monolithic, which appears in a rectangular package. It consists of alternating layers of ceramic material and printed electrodes which are sintered together to form the final package. The self resonant frequency is on the order of 15 MHz for the 0.01µF size, for a total lead length of 0.5 inch, and on the order of 165 MHz for the 0.0001µF = 100 pF size. Ceramic capacitors are classified into Class 1 (r<600) and Class 2 (r>600) by the Electronics Industries Association (EIA). Class 1 ceramic capacitors tend to be larger than their Class 2 counterparts, and have better stability of values with changes in temperature, voltage, or frequency. The best Class 1 capacitor, with nearly constant characteristic, will be labeled 'C0G' using EIA designators, but is often referred to as 'NP0' which stands for 'negative-positive-zero' (temperature coefficient).

Plastic-film capacitors use extremely thin sheets of plastic film as the dielectric between capacitor plates, usually in a coil construction. They have low losses and good resistance to humidity. Four common materials are:
1. Polycarbonate
2. Polypropylene
3. Polystyrene
4. Teflon

Generally speaking, as one moves down in this list from polycarbonate to Teflon, the capacitors get larger, better, and more expensive.

Historically, low-budget Tesla coilers have used either saltwater capacitors (as Tesla himself did) or homemade rolled polyethylene and foil capacitors. The saltwater capacitors are lossy and heavy. They are one of the first components to be replaced as a new coiler grows with his hobby. The homemade foil capacitors are limited by corona onset to about 7000 V. They are oil filled, so if the container leaks or is tipped over, one has a real mess. Serious coilers would look for commercial high voltage, high current, pulse rated capacitors made by specialty companies like Maxwell. These worked fine, but were expensive if purchased new, and difficult to find on the used market.

This all changed in 1999 when the Tesla coil community moved en masse to multi mini capacitors (MMC). These are small commercial capacitors that one buys by the sackful from Digi-Key and connects in series and parallel strings to get the required ratings.

There are several different capacitor series, some of which are more suitable for Tesla coil work than others. One should look for capacitor types that are rated for “high voltage, high frequency, and high pulses”. A dissipation factor of 0.1% at 1 kHz is good, as is polypropylene for the dielectric. The ECWH(V) and ECQP(U) series of capacitors are good. The Panasonic ECQ-E capacitors utilize metallized polyester (Mylar) which has a higher dissipation factor (1% at 1 kHz) than the polypropylene capacitors.

Capacitor ratings have received a great deal of attention by Tesla coilers. A given capacitor might be rated at 1600 VDC and 500 VAC. The AC rating is always for rms values, so the peak voltage would be about 700 V for a rating of 500 VAC. A capacitor in a Tesla coil primary experiences full voltage reversal, so it would seem wise to find the AC rating, multiply by √2, and divide that into the peak voltage available from the iron core transformer to find the number of series capacitors to use.

source and the capacitance. This is especially true for capacitance values >1 nF. A typical series resistor would be 10 kΩ to 1 MΩ.

Capacitor Leakage

Leakage is one of the less-than-ideal properties of a capacitor, expressed in terms of its insulation resistance (IR). For a given dielectric material, the effective parallel resistance is inversely proportional to the capacitance. This is because the resistance is proportional to the thickness of the dielectric, and inverse to the capacitive area. The capacitance is proportional to the area and inverse to the separation. Therefore, a common unit for quantifying capacitor leakage is the product of its capacitance and its leakage resistance, usually expressed in megohms-microfarads (MΩ·µF). Capacitor leakage is measured by applying a fixed voltage to the capacitor under test and measuring the resulting current. The leakage current will decay exponentially with time, so it’s usually necessary to apply the voltage for a known period (the soak time) before measuring the current.

In theory, a capacitor’s dielectric could be made of any non-conductive substance. However, for practical applications, specific materials are used that best suit the capacitor’s function. The insulation resistance of polymer dielectrics, such as polystyrene, polycarbonate or Teflon, can range from 104 MΩ·µF to 108 MΩ·µF, depending on the specific materials used and their purity. For example, a 1000 pF Teflon cap with an insulation resistance higher than 1017 Ω is specified as >108 MΩ·µF. The insulation resistance of ceramics such as X7R or NPO can be anywhere from 103 MΩ·µF to 106 MΩ·µF. Electrolytic capacitors such as tantalum or aluminum have much lower leakage resistances, typically ranging from

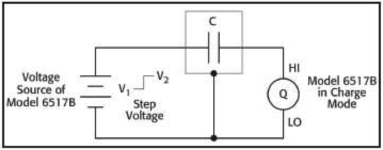

FIGURE 1. Capacitance measurement using an electrometer with an integrated voltage source.

Then, the voltage source should be turned on and the charge reading noted immediately. The capacitance can then be calculated from:

$$C = \frac{Q_2 - Q_1}{V_2 - V_1}$$

where

Q2 = final charge
Q1 = initial charge (assumed to be zero)
V2 = step voltage
V1 = initial voltage (assumed to be zero)

After the reading has been recorded, reset the voltage source to 0 V to dissipate the charge from the device. Before handling the device, verify the capacitance has been discharged to a safe level. The unknown capacitance should be in a shielded test fixture. The shield is connected to the LO input terminal of the electrometer. The HI input terminal should be connected to the highest impedance terminal of the unknown capacitance. If the rate of charge is too great, the resulting measurement will be in error because the input stage becomes temporarily saturated. To limit the rate of charge transfer at the input of the electrometer, add a resistor in series between the voltage

Experience of coilers, however, has been that there are sufficient safety factors built into the capacitors that using the DC rating is quite acceptable. For example, a 15 kV transformer has a peak voltage of 15√2 = 21. 2 kV. Divide 21.2 by 1.6 for 1600 VDC capacitors to get 13.25. Use either 13 or 14 capacitors in series.

The voltage rating is not determined by the dielectric breakdown voltage as much as it is by the Ionization or Corona Inception Level . This refers to the level at which ionization or partial discharges can begin to occur inside a bubble of entrapped air or within an air-filled void within the solid dielectric system. If one had perfect dielectrics and could always exclude any entrapped air, derating for this phenomenon would not be necessary. Corona inside a capacitor will chemically degrade the dielectric material over a period of time until the capacitor ultimately fails. Manufacturers of capacitors might select a voltage rating whereby their capacitors will last for one million hours before this occurs. In Tesla coil use, 100 hours of actual operation is a long time. This helps explain why coilers can exceed the manufacturer ratings without immediate problems.

There is evidence that operating a capacitor above its ac voltage rating will shorten its life by a factor of the overvoltage ratio raised to the 15th power. Suppose we have a capacitor rated at one million hours at 500 VAC. For economic reasons we are thinking about operating it at 700 VAC. The life reduction factor is (700/500)15 = 155. 57. Dividing one million by 155.57 gives an expected life of 6428 hours or almost one year of continuous operation. This is more than adequate for most Tesla coil applications.

One brand of capacitor is the WIMA. Terry Fritz comments about them: "The switching power supplies we build have WIMAs in pulsed duty similar to that seen in Tesla coil use.

The real indicator of how long they live is how warm they get. If they heat to about 5 degrees C above the ambient, then they last forever. At about 8 degrees above ambient, we see occasional failures. At 10 degrees, they get to be a problem. In many situations, we go beyond the WIMA chart derating in pulsed applications with no problem at all, just being sure they don't get hot."

WIMA experts say that partial discharges occur only above a certain voltage level and only on ac. Frequency does not seem to be a factor. They agree that the ac rating can be exceeded in Tesla coil applications, but without a serious investigation, were mentioning factors like 1.25 to 1.5.

Some capacitors that are candidates for a MMC use actual metal foil for electrodes while others use metallized plastic. The metal foil devices are physically larger for a given rating and are much more robust in Tesla coil service.

Dielectric material, electrode thickness, and lead size all limit the rate at which charge can be added to or removed from a capacitor. This limit is expressed as the dV/dt limit. The peak current I peak can be determined from the dV/dt value by the equation

$$I_{peak} = \frac{dV}{dt}C \qquad (12)$$

The coulomb's function of an electrometer can be used with a step voltage source to measure capacitance levels ranging from <10pF to hundreds of nanofarads.

The unknown capacitance is connected in series with the electrometer input and the step voltage source.

The calculation of the capacitance is based on this equation:

$$C = \frac{Q}{V}$$

Figure 1 illustrates a basic configuration for measuring capacitance with an electrometer with an internal voltage source. The instrument is used in the charge (or coulombs) mode, and its voltage source provides the step voltage. Just before the voltage source is turned on, the meter's zero check function should be disabled and the charge reading suppressed by using the REL function to zero the display. (The purpose of zero check is to protect the input FET from overload and to zero the instrument. When zero check is enabled, the input of the electrometer is a resistance from roughly 10 MΩ to 100 MΩ, depending on the electrometer used. Zero check should be enabled when changing conditions on the input circuit, such as changing functions and connections. The REL function subtracts a reference value from actual readings. When REL is enabled, the instrument uses the present reading as a relative value. Subsequent readings will be the difference between the actual input value and the relative value.)

Capacitor Testing Challenges and Solutions

by Dale Cigoy

http://www.circuitsassembly.com/cms/component/content/article/159/10164-capacitor-testing-challenges-and-solutions

Chip caps are prone to leakage, so consider these test methods for minimizing electrical failures.

Capacitors are widely used for bypassing, coupling, filtering and tunneling electronic circuits. However, to be useful, their capacitance value, voltage rating, temperature coefficient and leakage resistance must be characterized. Although capacitor manufacturers perform these tests, many electronics assemblers also perform some of these tests as quality checks. This article looks at some of the challenges associated with capacitor testing, as well as some of the test techniques used.

A capacitor is somewhat like a battery in that they both store electrical energy. Inside a battery, chemical reactions produce electrons on one terminal and store electrons on the other. However, a capacitor is much simpler than a battery because it can't produce new electrons – it only stores them. Inside the capacitor, the terminals connect to two metal plates separated by a non-conducting substance known as a dielectric.

A capacitor's storage potential, or capacitance, is measured in farads. A one-farad (1 F) capacitor can store one coulomb (1 C) of charge at one volt (1 V). A coulomb is 6.25´1018 electronics. One amp represents a rate of electron flow of 1C of electrons per second, so a 1 F capacitor can hold one amp-second (1 A/s) of electrons at 1 V.

---

where appropriate multipliers should be used to get everything in volts, seconds, and farads. By way of reference, a 1000 pF capacitor rated for 10,000 V/µs will withstand 10 A surges. A capacitor with a dV/dt rating less than 1000 V per 5µs probably does not have metal foil endplates and should not be used.

The peak current that a Tesla coil primary capacitor must provide is determined by the capacitor voltage and the surge impedance. If the secondary were removed, the surge impedance would be

$$Z_s = \sqrt{\frac{L_1}{C_1}} \qquad (13)$$

where C1 is the capacitance in the Tesla coil primary and L1 is the inductance of the primary. When the secondary is in place, the surge impedance will increase from the above value, so the peak current required from the capacitor will be less. Using the above expression for Zs should be a worst case calculation. If it tells you that there might be 100 A flowing after the gap shorts, and your individual capacitors are rated at 10 A each, then you would need ten parallel strings. Since this figure includes some factor of safety, you might be able to get by with somewhat fewer strings. One WIMA expert said that the maximum amperage rating is not as critical as the Corona Inception Level mentioned above.

The polypropylene capacitors useful for Tesla coil work have a self-healing mode. A partial discharge will eventually cause a local failure. A burst of energy through this short will vaporize everything around the short, effectively removing the short.

The capacitor continues to operate, but with a slightly lower capacitance due to some of the electrodes being removed. It is a good idea to carefully measure the capacitance of the MMC periodically. A decrease of even 1% could indicate that the capacitors are being stressed, and that total failure is possible.

Terry Fritz reports testing some WIMA MKS 4 (K4) capacitors rated at 1μF and 400VDC on a large DC supply. These are metallized polypropylene dielectric capacitors that are encapsulated in epoxy filled 3×1 cm rectangles. Starting at about 900 V there are a few little snaps, indicating that the dielectric had punched through and the arc blasted the thin metal layer on the other side. Going to 1200 V gave a few more snaps. Around 1500 V there were many snaps and the capacitor shorted out.

When a capacitor cleanly blows apart, that is a sign that they were destroyed by over voltage. When a capacitor puffs up and looks melted and burned, that is from too much current. It is always a good idea to think about possible failure modes, and put the capacitors in appropriate enclosures to protect bystanders.

I had a student build a small design project once, that illustrates this point. He used an electrolytic capacitor rated at about 25 VDC in some circuit for 120 VAC usage, and I did not notice it before testing. He was proudly demonstrating performance, using a variac to bring up the voltage. At about 90 V, the electrolytic capacitor exploded with a noise somewhere between a pistol and a small shotgun. The contents of the capacitor, about a cubic centimeter of plastic fibers, hit the student in the middle of the chest. He was not injured, but for a brief moment he thought he had been killed by this exploding capacitor. I suspect he still pays

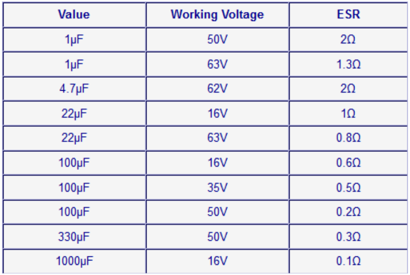

Value	Working Voltage	ESR
1µF	50V	2Ω
1µF	63V	1.3Ω
4.7µF	62V	2Ω
22µF	16V	1Ω
22µF	63V	0.8Ω
100µF	16V	0.6Ω
100µF	35V	0.5Ω
100µF	50V	0.2Ω
330µF	50V	0.3Ω
1000µF	16V	0.1Ω

Conclusion

All the capacitors had an ESR of 2 ohms or less and most of the values that you would use in a power supply (22µF and up) had an ESR of less than 1 ohm. As a rule of thumb, higher valued capacitors with higher working voltages had a lower ESR, although there were some exceptions so this cannot be guaranteed.

Keep in mind that the ESR for standard electrolytic capacitors increases with lower temperatures. For example, it will typically increase by 20 to 30 times when the temperature drops from 25°C to -40°C.

---

more attention to capacitor voltage ratings than most people.

Many coilers recommend placing resistors across the MMC to bleed off the charge after power is removed. This can eliminate some very unpleasant surprises when making adjustments between runs. One can get true high voltage resistors but they are expensive. Most use a long chain of ordinary resistors therefore. Digi-Key sells 0.5 W carbon film resistors rated at 350 volts each for about two cents each in quantity. At this price, there is little point in not using adequate safety factors. Suppose we feel comfortable in operating a 0.5 W resistor at 0.1 W and at the rated peak voltage of 350 V each. We solve for the resistance as

$$R = \frac{V^2}{P} = \frac{(350/\sqrt{2})^2}{0.1} = 612\ \ \text{k}\Omega$$

where we round off to the nearest standard value of 620 kΩ.

The number of resistors needed for a string across say a 15 kV transformer secondary would be

$$N = \frac{15000\sqrt{2}}{350} = 60\ \ \text{resistors}$$

We then check the time constant to get a measure of how long it takes to discharge a

Table 3: Dielectric Constant r and Dissipation Factor DF of Water.

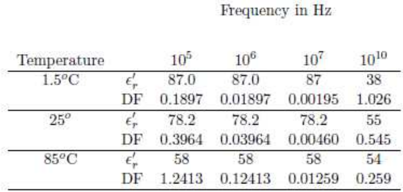

		Frequency in Hz			
Temperature		10^5	10^6	10^7	10^{10}
1.5°C	ϵ'_r	87.0	87.0	87	38
	DF	0.1897	0.01897	0.00195	1.026
25^o	ϵ'_r	78.2	78.2	78.2	55
	DF	0.3964	0.03964	0.00460	0.545
85°C	ϵ'_r	58	58	58	54
	DF	1.2413	0.12413	0.01259	0.259

capacitor chain. If the capacitance happened to be 27 nF, the time constant would be
$\tau = RC = (60)(620000)(27\times10{-}9) = 1$ second

A capacitor will be mostly discharged after five time constants or 5 seconds in this example. There are obviously many other design possibilities. If we assumed the resistors would accept a higher voltage before arcing over, then we would want to use higher ohm values to keep heating within bounds. Terry Fritz tested some Yageo 10 MΩ 0.5 W carbon film resistors to see where they would actually fail. At 4000 V the resistor started to turn brown and smoke a little. This should be expected since it is now drawing 2 W, four times its power rating. At 5300 V, it gave a very satisfying crack and arced along its outside surface. After that, the 2 MΩ resistor measured 11.1 MΩ. So a resistor specified at 350 V actually failed at 5300 V, a safety factor of 15. Terry sees no problem with using these 0.5 W resistors up to 1000 V if the resulting wattage is not excessive.

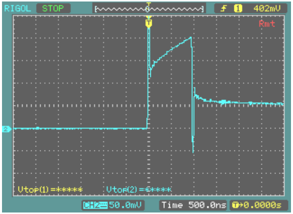

Results

The following table lists the results for a representative collection of aluminium electrolytic capacitors in my parts collection. The manufacturers were Elete and Lelon and they were bought from Futurlec and Jaycar (in Australia). They represent the typical components that a hobbyist would have access to.

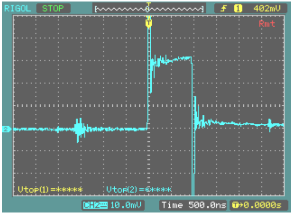

This method will work for values of ESR from 0.1Ω to greater than 10Ω. Alas, it is not very accurate - but then the ESR of a capacitor is not a precise value anyway and I was only looking for an approximation.

The main problem with this technique is that it is not very effective with capacitors below 10µF. With these small values the charge on the capacitor will rapidly rise making it difficult to estimate the height of the pulse before the effect of charging takes over – although, with some guesswork it can be still used for values down to 1 µF.

This is illustrated in the screenshot on the left. The capacitor on test was a 2.2µF 63VW electrolytic and it is possible to estimate that the start of the pulse is about 130mV (vertical scale is 50mV/division). This gives an ESR of 1.3Ω.

Equivalent Series Resistance (ESR) of Capacitors

QuadTech, Incorporated
http://www.lowesr.com/QT_LowESR.pdf

Questions continually arise concerning the correct definition of the ESR (Equivalent Series Resistance) of a capacitor and, more particularly, the difference between ESR and the actual physical series resistance (which we'll call Ras), the ohmic resistance of the leads and plates or foils. The definition and application of the term ESR has often been misconstrued. We hope this note will answer any questions and clarify any confusion that might exist. Very briefly, ESR is a measure of the total lossiness of a capacitor. It is larger than Ras because the actual series resistance is only one source of the total loss (usually a small part).

The Series Equivalent Circuit

At one frequency, a measurement of complex impedance gives two numbers, the real part and the imaginary part: Z = Rs + jXs. At that frequency, the impedance behaves like a series combination of an ideal resistance Rs and an ideal reactance Xs (Figure 1).

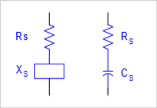

Figure 1: Equivalent Series Circuit Representation

If Xs is negative, the impedance is capacitive and the general reactance can be replaced with a capacitance of:

Test Setup

The method that I used is quite simple and is shown in the diagram below.

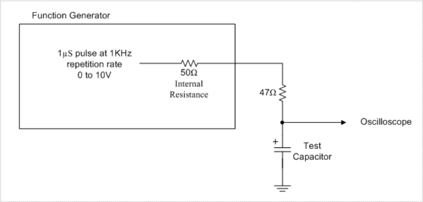

I used a function generator to generate narrow pulses (1µS wide) with a slow repetition rate (1KHz). This drove a voltage divider with the test capacitor in the bottom leg and 100Ω in the top leg. Because of the narrow pulse, the capacitor did not have enough time to build up a charge, so the voltage across it represented the voltage drop caused by its ESR.

With the voltage of each pulse set at 10V this arrangement meant that the height of the pulse (in mV) across the test capacitor divided by 100 was equivalent to the capacitor's ESR in ohms. For example, a pulse height of 80mV represented an ESR of 0.8Ω.

This is illustrated on the right. The vertical sensitivity of the oscilloscope was 10mV per division and the height of the pulse was about 30mV. This means that the capacitor's ESR was approximately 0.3Ω (the capacitor under test was a 330µF 50VW aluminium electrolytic).

Measuring Capacitor ESR

by Geoff

http://geoffg.net/Measuring_ESR.html

Recently, while selecting components for a power supply, I came across the following statement in the datasheet for the voltage regulator... "The output capacitor must have an ESR of less than 3 ohms."

When I looked up the retailers who supplied the capacitors in my parts bin I found that none of them listed the ESR for the capacitors that they sold. So, somehow, I needed to measure this value.

After some research I came up with the simple approach described below. It relies on some good test equipment and, while you may not want to repeat the test, the results should be informative.

ESR

Equivalent Series Resistance (ESR) is the internal resistance that appears in series with the device's capacitance. No capacitor is perfect and the ESR comes from the resistance of the leads, the aluminium foil and the electrolyte. It is often an important parameter in power supply design where the ESR of an output capacitor can affect the stability of the regulator (ie, causing it to oscillate or over react to transients in the load).

You can buy ESR meters but I just wanted to know if any of the capacitors in my parts bin had an ESR of less than 3 ohms. So, how to test them?

$$\mathrm{Cs} = \frac{-1}{\omega \mathrm{Xs}}$$

We now have an equivalent circuit that is correct only at the measurement frequency. The resistance of this equivalent circuit is the equivalent series resistance ESR = Rs = Real part of Z

Add the Dissipation Factor

If we define the dissipation factor D as

$$\mathrm{D} = \frac{\text{energy lost}}{\text{energy stored}} = \frac{\text{Real part of Z}}{(-\text{Imaginary part of Z})}$$

$$\text{Then}\ \ \mathrm{D} = \frac{\mathrm{Rs}}{(-)\mathrm{Xs}} = \mathrm{Rs}\omega\mathrm{C} = (\mathrm{ESR})\omega\mathrm{C}$$

If one took a pure resistance and a pure capacitance and connected them in series, then one could say that the ESR of the combination was indeed equal to the actual series resistance. However, if one put a pure resistance in parallel with a pure capacitance (Figure 2a), the ESR of the combination is

Real part of Z = Real part of

$$\frac{1}{\frac{1}{\mathrm{Rp}} + j\omega\mathrm{Cp}} = \frac{\mathrm{Rp}}{1 + \omega^2\mathrm{Cp}^2\mathrm{Rp}^2}$$

as illustrated in Figure 2b. From Figure 2a, however, it is obvious that there is no actual series resistance in series with the capacitor.

Ras = 0

But ESR > 0
Therefore ESR > Ras

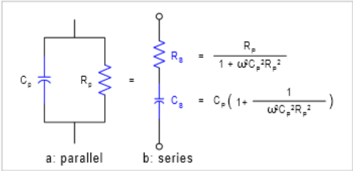
C_p
R_p
=
$R_s = \frac{R_p}{1 + \omega^2 C_p^2 R_p^2}$
$C_s = C_p\left(1 + \frac{1}{\omega^2 C_p^2 R_p^2}\right)$
a: parallel
b: series

Figure 2: Series and Parallel Equivalent Circuits

QuadTech manufactures a wide variety of LCR meters designed for accurate measurements of C, Df and ESR. The design of the QuadTech 1930 LCR Meter, has been specifically optimized for high speed, low ESR measurements on tantalum and other types of capacitors. Fast settling time and quick open/short recovery significantly reduce test times especially on capacitors greater than 100µ F at 120Hz. The 1930 Low ESR Capacitance Meter performs most impedance measurements (C, Df, ESR, Z, R, X, L Q, Y, G, B, θ and DCR) over a frequency range of 100Hz – 100kHz. Actual voltage and current is measurable across the DUT.

The ESR of a Real Capacitor

Actual capacitors have three main sources of loss:
1. Actual series resistance: There is some resistance in the leads and plates or foils. This is the resistance of conductors

higher inductance in this test setup. I believe this was due to the fact that the smaller capacitor did not fill the package and internal lead inductance caused the effect. In this case, a 1 uF capacitor was a better choice than a 0.1 uF!

One of the advantages of this test is that the output waveform is the transient response of the capacitor. The voltages developed across the capacitor in this test are directly related to what will happen in a real circuit if the current risetime from the generator is similar to what the capacitor will see in its intended application.

was repetitive and the slight slope on the left half of the waveform was the end of the exponential fall from 5 volts. If a single pulse on a digital scope was used, the slope to the left of the Ldi/dt spike would be zero.

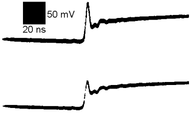

Figure 5. 1uF Capacitor

Figure 6 shows the result for a 1 uF radial ceramic capacitor (square case). Note the low inductance and undetectable ESR. Note also that the slope of the 1/C exponential rise is flatter indicating more capacitance than the 1 uF capacitor of Figure 5. This may be due to the fact that the electrolytic capacitor used for Figure 5 may have lower capacitance near zero voltage than at its operating voltage whereas the ceramic capacitor has a more constant capacitance with voltage. The inductance corresponding to the lower trace is estimated to be 4.4 nH.

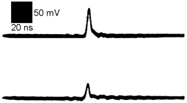

Figure 6. 1uF Ceramic Capacitor

It is interesting to note that a 0.1 uF ceramic capacitor in the same size package as the 1 uF of Figure 6 showed a slightly

and is always low. It causes a power loss I2Ras where I is the current flowing in the capacitor. This causes $D_1 = \omega \mathrm{RasC}$

2. Leakage resistance: There is some actual parallel resistance due to leakage current in the capacitor. We'll call this RL. It is the resistance of the capacitor at DC and it is a high resistance. For plastic capacitors it can be 10^12ohms (GΩ) or higher. It causes a loss of E^2/R_L where E is the applied (rms) voltage and $D_2 = \frac{1}{\omega R_L C}$. This loss is usually negligible except at very low frequencies.

3. Dielectric loss: This is due to two phenomena molecular polarization and interfacial polarization (dielectric absorption). It is not the purpose of this note to discuss its cause, but only its effect. It causes a D that can change with frequency in most any manner that is not abrupt. It acts neither as a series resistance, nor as a parallel resistance that would pass DC. One model is a parallel resistance, RD, which is variable with frequency, with DC blocked by a large series capacitance (CB) as illustrated in Figure 3.

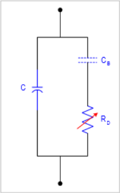

Figure 3: Parallel Resistance

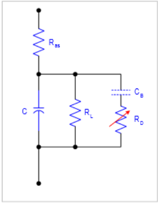

Figure 4: All 3 Capacitor Losses

If the series C were infinite, the D would be $D_3 = \frac{1}{\omega R_D C}$ but it must be remembered that it is not inversely proportional to ω because RD changes with frequency.

All three sources of loss are illustrated in Figure 4. This circuit would have a total dissipation factor equal to D1+ D2+ D3.

Plot of Total Dissipation

For the circuit in Figure 4, the total dissipation factor is

$$D = D_1 + D_2 + D_3 = \omega RasC + \frac{1}{\omega R_L C} + \frac{1}{\omega R_D C}$$

Figure 5 illustrates a general plot of this combination series and parallel resistance in a capacitance measurement.

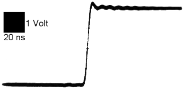

Figure 3. Input From Pulse Generator

Figure 4 shows data from a 4 uF electrolytic capacitor. The ESR offset is about 50 mV yielding an estimate of the ESR of just over one Ohm. Notice that there appears to be some oscillations on the 1/C part of the slope. This could be scope probe resonance or a resonance in the capacitor. The data was taken with a standard 10X Hi-Z probe, so the probe is suspect. I have seen capacitors with pronounced oscillation from internal resonance. If you are planning to put a large capacitor in parallel with a smaller one, especially if they are constructed from different technologies, it would a good idea to check out the impulse response of the combination using this method. It is possible for the smaller capacitor to resonate with the inductance of the larger one, causing an unexpected result.

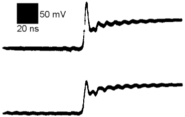

Figure 4. 4uF Capacitor

Figure 5 shows the result for a 1 uF capacitor of the same construction as the 4 uF capacitor tested in Figure 4. Note that the inductance is similar to the 4 uF capacitor, but the ESR is slightly lower. Since an analog scope was used, the waveform circuit to not significantly affect the initial current very much. For this frequency range a generator with a risetime of one to two nanoseconds will do.

If you need to check the capacitor using faster risetimes, it would be best to build the test setup on a small circuit board with a ground plane and controlled impedances. At this point, the parasitic capacitance of the 50 Ohm resistor would also be an issue to be taken into account. Fortunately, such accuracy is often not needed. Especially, if one is just comparing the relative performance of several capacitors.

Data

Figure 3 below shows the initial rise from the generator. The black square indicates the vertical voltage and horizontal time scales. The open circuit voltage was a little over 4 Volts with about a 5 nanosecond risetime. The data in Figures 3 through 6 were taken with an analog scope some years ago. Figures 4 through 6 show data obtained from several leaded capacitors (as opposed to surface mount). Two traces were taken for each capacitor. The lower trace was measured at the capacitor body where the leads entered and the upper trace included the minimum amount of lead to practically connect the capacitor to a printed wiring board. The upper trace would not be needed for modern surface mount capacitors unless one wanted to model the connection inductance from the capacitor to the point of interest on a printed wiring board.

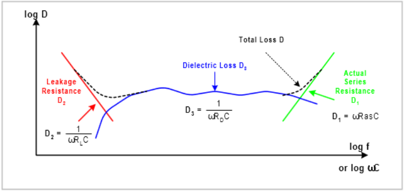

Figure 5: Total Dissipation Factor Plot

At any given frequency we could measure the real part of the impedance, Rs, or measure C and D and calculate it

$$Rs = ESR = \frac{D}{\omega C} = \frac{D_1 + D_2 + D_3}{\omega C}$$

Therefore ESR = $Ras + \frac{1}{\omega^2 C^2 R_L} + \frac{1}{\omega^2 C^2 R_D}$ and thus ESR > Ras, always.

Usually, ESR is very much larger than Ras. However, when ω C is large at high frequencies, high capacitances or some combination, the actual series resistance can cause the largest part of the total D. (See plot.) For very large capacitors (like 0.1F), ESR can be very nearly equal to the actual series resistance even at low frequencies (such as 120 Hz). For most capacitors at low frequencies, the actual series resistance is only a small part of ESR.

For complete product specifications on the 1900 Series Impedance Meters, or any of QuadTech's products, please visit us at http://www.quadtech.com/products. Call us at 1-800-253-1230 or email your questions to info@quadtech.com.

1) $dv/dt = i/C = 50$ mA/C
where C is the value of the capacitor at this low voltage and the risetime of the current << RC.

The offset between the baseline and the beginning of the exponential rise is just the voltage that the current, 50 mA for this case, develops across the ESR of the capacitor. The ESR can be easily estimated in this case by dividing the voltage offset (labeled ESR in Figure 2) by 50 mA.

Parasitic inductance in the capacitor will cause the spike in the waveform shown in Figure 2 exceeding the value of the dotted line along its length. If the current rise were in fact a ramp with constant slope and very sharp corners (high di2/dt) then the spike would be a square pulse of value:

2) $E = L*di/dt$
where L is the parasitic inductance of the capacitor.

The current rise from the generator used for the data in this article was not a ramp with very sharp corners and constant slope (the case for most generators I have used). That characteristic of the generator combined with probe effects led to a peaked shape to the Ldi/dt spike as shown in Figure 2. Using equation 2 the inductance of the capacitor can be calculated. Often, one does not need to calculate the inductance or ESR but just choose a capacitor from several available ones that has the lowest inductance and/or ESR.

Soldering the components onto a BNC connector as in Figure 1, works up to 300 MHz. I estimate the inductive reactance of the loop formed by the capacitor and resistor to be about 20 Ohms at 300 MHz (estimating the inductance at 10 nH). This is small enough relative to the 100 Ohms of resistance in the

For a pulse length that is long with respect to the RC time constant, one will see an exponential rise to the open circuit voltage of the pulse source. For the purposes of this discussion, we will be looking at the first couple of hundred millivolts of a 5 volt exponential rise. An example of this appears in Figure 2.

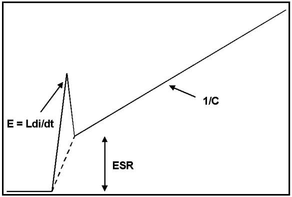

Figure 2. Initial Rise

Figure 2 shows the beginning of the exponential voltage rise across the capacitor when the generator pulse starts. The vertical scale is about 200 mV and the horizontal time is a small fraction of the RC time constant of 100 Ohms and the capacitor being measured. Since the capacitor voltage is still very small compared to the 5 Volt open circuit output of the generator, the current through the capacitor may be presumed to be constant and equal to the generator open circuit voltage divided by 100 Ohms, 50 mA in this case.

The risetime of the current will be the same as the generator voltage. If the rise is a ramp with a constant slope and the capacitor had no inductance, the initial rise shown in Figure 2 would follow the dotted line and then the slope would change to the initial slope of the exponential rise determined by:

High Frequency Measurements Web Page
Douglas C. Smith
http://www.emcesd.com/tt020100.htm

Measuring Capacitor Self-inductance and ESR

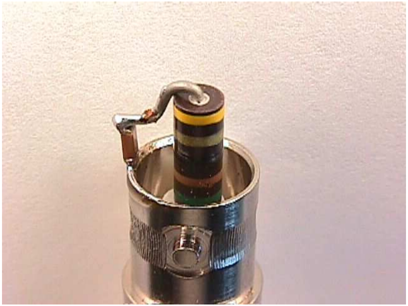

Figure 1. Test Circuit for Measurement of Capacitor Self-inductance and ESR

Technical Background

The parasitic parameters of a capacitor, that is its equivalent series resistance (ESR) and its inductance, affect the way the capacitor performs in circuits. Some applications are very sensitive to these parameters. For instance, a bypass capacitor used between power and ground in a digital circuit must be able to supply current quickly to nearby active devices. If it has too much inductance it will not be able to do this.

Similarly, the transient response of a capacitor used to divert a current pulse due to electrostatic discharge is very important to the ability of the capacitor to do its job.

So how can the parasitic parameters of a capacitor be measured? One could certainly connect the capacitor to a network analyzer and get a very good characterization. Such an instrument can be quite expensive though. Even the less expensive capacitance measuring instruments may not be available when needed. Both instruments may not provide the information in an easily usable form. If you have a pulse generator (preferably with a 50 Ohm output impedance) and an oscilloscope, you can easily measure the transient response of a capacitor. From this data the capacitor's ESR and inductance can be determined.

First, construct the simple network shown in Figure 1 at the end of a 50 Ohm coaxial cable fed from a 50 Ohm pulse generator. A 50 Ohm resistor is used in Figure 1 to terminate the coax during the rising edge and provide a total of 100 Ohms of source impedance. The resistor shown is a 51 Ohm 1/2 watt carbon composition resistor with one lead trimmed so that the resistor just seats with the trimmed lead fully inserted into the BNC connector. It may be necessary to put a little solder bump on the resistor lead so that it stays firmly in the BNC connector. The capacitor to be tested is connected between the end of the resistor and the shell of the BNC connector. An oscilloscope is connected directly across the capacitor using leads as short as possible to connect the probe. Probes with a resistive input impedance of 500 to 1000 Ohms are recommended. Standard 10X "Hi-Z" probes often have rising edge effects that will distort the part of the waveform used for the calculations.

NOTES:

KEVIN’S WEBSURFER HANDBOOK X
FOR CRYSTAL RADIO

MY WEB PAGE VISITED

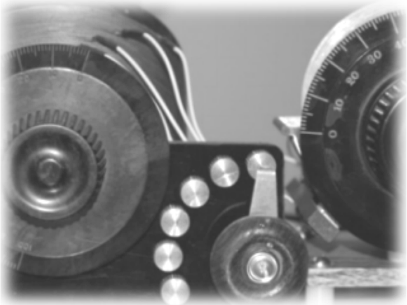

Kevin Smith
2012

Printing / Binding Instructions
1. Choose “fit to page” in print menu
2. Print document double sided on letter size paper
3. Cut the entire printed document in half
4. Fold over making sure the page numbering is continuous
5. For the cover: Print just the first page on card stock paper
   Cut the cover in half as well
6. Assemble the covers on the document
7. Punch the left side for a binding, spiral or comb as desired

www.lessmiths.com/~kjsmith/crystal/cr0intro.shtml

KJ Smith

NOTES:

NOTES:

INTRODUCTION

In March of 2009 while shopping for some sensors on ebay, quite on a lark I typed "Crystal Radio" as a search term not really knowing what to expect. My search was not totally random, I have always been interested in radio communications and in radio propagation as a geophysical probe of the ionosphere, (more about that here). Additionally as a child I had once a small "Spy Pen" crystal radio what always facinated me, and as an adult I had given a crystal radio toy kit to my second son who showed much interest in computers and electrical things. He has since gone on to his Master's degree in EE and now workd designing computer chips at Intel in Silicone Valley. So, maybe it was the radio?

Vaguely I thought to find a crystal radio or radio kit of some sort to play with, something more serious than a cheap plastic toy. What I found was suprising, so much there. I explored the listings, odds and ends mostly of course, its ebay after all, and a few completed radios new and vintage, and one in partictular which caught my eye. There were several kits by someone calling himself Xtalman including one he calls Dunwoody which seemed to exactly fit my idea, a kit, easy enough for a beginner but serious enough to satisfy a more educated mind. I decided to pull the trigger, the radio was offered "buy it now", so I did! So began my journey into a black hole of interest, somebody really ought to post warnings on all crystal radios.. "WARNING, may cause serious refocusing of lifestyle".

In the time since, I have plunged deeply into the web literature, joined the Xtal Set Society, and built homebrews of my own in addition to purchasing a few sets, (described separately). In the following pages I present each radio with the ideas and motavations that move me to own, admire, and use them. It has become a facinating journey for me which I wish to share so that others may understand and catch the bug for themselves, so that other radio builders may view and comment, with suggestions perhaps for improvement, and so us newbies may heed the WARNING, and STILL take the plunge!

The Xtal Set Society Dedicated to once again building and experimenting with radio electronics Great source for plans and books, kits and parts. I joined the society as soon as I finished my Mystery Set.

Dave Schmarder Dave's Homemade Radios, literally 10's of different sets all with interesting features and each a pleasure to look at. Much technical background and good design ideas can be found as well. I like this page.

Scott's Crystal Radios Mount Your Own Crystals Here I learned, and successfully applied, a good technique to mount my own crystals.

Part 2, Inspiration:, Sites with radio plans and projects, so many ideas.

Mike Peebles Crystal Radios - Peebles Originals!

Tom's Handmade Radios Nice selection of radios, superb craftmanship, for sale!.

Birmingham Alabama Crystal Radio Group Nice contest page, many lovely contest radios

Jim's Crystal Radio Page Crystal radios with ferrite cores, and a Mystery Set.

Larry J Solomon Solomon Radios, several variations on the Mystery Set and others..

Rainer Steinfuehr Gollum´s Crystal Receiver World.

So Many more, if I missed your site my sincere apologies.

The Stay Tuned Website is probably the site I visit most often of all. This excellent effort contains an enormous amount of information, vintage plans and projects, Contest Radios, Museums, and more. Download Professor Coyle's coil calculators, you will use them constantly. This is the only source for MB Sleeper's 101 radio receiving circuits I have found.

The Smith-Kettlewell Technical File ENERGY-FREE RADIO: THE CRYSTAL RADIO SET REVISITED This is also a great introductory page with a great deal of design tips and background. There is no author and its embedded in an eye research center website, oh well.. I highly recommend this page.

Owen Pool WB4LFH Crystal Radio Resources Owen Pool's great site with much fundamental background, projects, teaching, contest results and so much more.

Steve Mcdonald VE7SL Radio Notebook Crystal Radio DXing Great site for discussions of DXing with crystal radios and double-tuned design fundamentals.

Mike Tuggle DESIGNING A DX CRYSTAL SET Mike's discussion of his Lyonodyne-17 and the concepts behind its design.

Dick Kleijer crystal-radio.eu
Dick Kleijer's site contains great discussions of theory and the results of numerous experiments with his crystal radios.

CONTENTS:

1) Resources for Theory and Design.

RADIO LABORATORY.. p 003
- Antenna Tuning p 007
- Coil Q p 019
- Q Analytical p 029
- Capacitors and ESR p 047
- Diode Protocol p 055
- Diode Test p 065
- Radio Test p 099

A Library of Crystal Hookups and Theory: p 113

2) Radios I have built with my own little mitts:

SPARK GAP TRANSMITTER.. p 119
- Coil p 122
- Interruptor p 128
- Spark Gap p 134
- Oscillation p 139
- Performance p 143

SLEEPER #1.. p 155

DUNWOODY.. p 159

MYSTERY.. p 165

HOMEBREW p 171

DE FOREST.. p 177

BREMER-TULLY.. p 185

GLOBAL DX'TER.. p 189

FLEMING.. p 195
MINSTREL BOY.. p 203
HAMMARLUND.. p 209
TEFLON.. p 219

3) Accessories on Breadboards

LOADING COIL.. p 229
WAVE TRAP p 233
AUDIO AMPLIFIER p 235

4) Radios I have acquired (read: paid waaay too much for, with no regrets)

STEINITE.. p 243
AMERICAN RADIO STORES.. p 249
BBC COMMERATIVE.. p 253

5) Other thoughts, speculations, ideas, ravings...

JACQUEMARD.. p 261
Steel Galena, my thoughts:.. p 265
KGOW1650, a local PEST!!.. p 275

6) Acknowledgements:.. (links) p 279

8) Notes p 282

Acknowledgements:
http://www.lessmiths.com/~kjsmith/crystal/ack.shtml

Radio hobby in general and certainly crystal radio is a mentored hobby with most practitioners able and willing to help one get started and do well in this hobby. In my researches for information I have been greatly aided by the many helpful web resources built by this community. I wish to offer here my sincere thanks for your efforts. I certainly would not have come as far as I have without this help! The following list is only partial to the number of sites I visited and if your page is not listed I offer my apologies. This list represents those sites I found most significant and helpful in terms of 1) Crystal Radio Fundamentals and 2) Set Design Inspiration.

Part 1, Fundamentals:, Sites where I spend most of my time.

Lance Borden BORDEN RADIO COMPANY The start of my adventure, serious crystal radio kits, a seriously fine gentleman.

Kenneth A. Kuhn Crystal Radio Engineering Kenneth Kuhn's excellent pages on the fundamentals and design of crystal radios and all their associated parts, excellently written, accessable to the non-specialist yet precise and clear. I spent hours studying these documents.

Alan R. Klase CRYSTAL SET DESIGN 102 Alan Klase's site is chock-a-block full of great work and ideas for interesting sets. His Design 102 gave me good insights in the various incarnations a crystal radio may take.

Darryl Boyd Stay Tuned Website: see #58 Crystal Radio Circuits 101

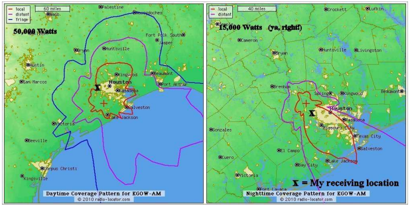
local
distant
fringe
60 miles
50,000 Watts
Palestine
Nacogdoches
Waco
Fort Polk South
Jasper
Bryan
Huntsville
Austin
Kingwood
Beaumont
San Marcos
Houston
Port Arthur
X
Pasadena
Galveston
Lake Jackson
Victoria
Beeville
Corpus Christi
Kingsville
Daytime Coverage Pattern for KGOW-AM
© 2010 radio-locator.com
local
distant
40 miles
Crockett
Lufkin
15,000 Watts (ya, right!)
Cameron
Huntsville
Livingston
Bryan
Brenham
Spring
Kingwood
Beaumont
Houston
X
Pasadena
Missouri City
Texas City
Gonzales
Galveston
El Campo
Cuero
Lake Jackson
Bay City
Victoria
Port Lavaca
X = My receiving location
Nighttime Coverage Pattern for KGOW-AM
© 2010 radio-locator.com

More disturbingly, it takes the joy out of the hobby. Mercifully, my handy dandy wave trap goes a long way to solve much of the trouble, but not far enough. I need to direct my efforts in trapping this station now.

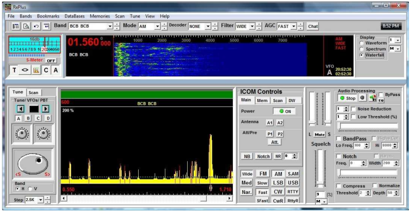

Band BCB BCB
Mode AM
Decoder NONE
Filter WIDE
AGC FAST
01.560 000

RxPlus
File Bands Bookmarks DataBases Memories Scan Tune View Help
Chat
8:52 PM
S-Meter OFF
T C A
BCB BCB
1000 2000 3000 4000 5000 6000 7000
AM WIDE FAST
VFO A 20:52:30 02:52:30
Display
Waveform
Spectrum
Waterfall
Tune Scan
Tune/ VFOs/ PBT
A B C D
Band
Step 2.5K
600
200 %
0.550
1.710
ICOM Controls
Main Mem Scan DW
Power ON
Antenna A1 A2
Att/Pre P1 P2 Att.
NB Notch NR
Wide FM AM S.AM
Med Slow LSB USB
Nar. Fast CW RTTY
SFast CwR RttyR
L Mute S
Squelch
Audio Processing
Stop
ByPass
Noise Reduction
Low Threshold (%)
BandPass
Lo Freq. 100 Hi 8000
Notch
Freq. Width 200
Compress
Normalize
Threshold 2 Depth 50

Question: Is this spectrum normal? Is it common for one station to so dominate the scene? I sometimes wonder if they arent cheating on their power output there. I know, probably not. My next project may be finding a way to use this energy to charge my batteries or light my house, or heck, to power my radios! The following plot shows that while by day I am merely in trouble, at night things are near hopeless. The band scan is a night spectrum. Sad thing is, I don't partictularily care to listen to sports on radio, (or television for that matter).

CHAPTER I

Resources for Theory and Design

Some thoughts on KGOW 1560 "The Game".
http://www.lessmiths.com/~kjsmith/crystal/kgow.shtml

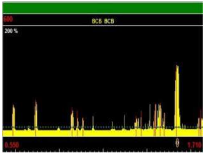

Over the months of playing with my hobby, my radios, learning, reading, i have come across a small fly in the ointment of an otherwise very enjoyable hobby. From time to time while listening to my radios I would find the airwaves filled with a single station, KGOW blasting across my entire tuning range. It was curious and interesting at first, radios would receive strong clear signals with or without a crystal or rectifier which seemed to me thoroughly at odds with what I supposed I was learning about how radios work. Discussions on Rap-n-Tap provided some, but not deep insights into my problem. Mostly I was advised to check twice and thrice all my connections. The most frustratingly bothersome aspect was that, it affected all my radios. Maybe my antenna and/or ground was responsible. Even when hooking up my "Sleeper #1" (shunting a diode across the phones) gave me the problem, in fact gave me the problem in spades. Turns out I am able to receive KGOW clearly merely by attaching the antenna and ground across the phones, no diode, no nada.

Looking somewhat deeper into the mystery, I chose to look closer, not at my antenna and ground, but at the BC band itself. Using my Icon R75 and RxPlus software, I made a band scan of the BC activity. What I found was most troubling for my location. The plot below shows a screen capture of the scan. Clearly KGOW stands out 2 to 3 times stronger than any other local station. It dominates the scene and saturates my post.

My radio "laboratory"

http://www.lessmiths.com/~kjsmith/crystal/lab.shtml

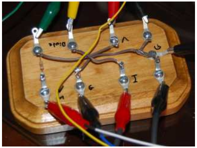

Having now acquired several crystal radios by purchase (two) or building my own (four), I begin to wonder how they stack up against each other and against other radios I read about and dream of on the net. It can be fairly common to hear radio performance described in qualitative terms, "high Q", "sharp tuning", "highly sensitive", etc. Without measured numbers to back them up, it is extremely difficult to judge the subjective notes of different radio users. It is inevitable I believe that in this hobby that one ends up getting together as much test equipment as possible, (read "affordable") in order to know more objectively about the radios one builds and/or uses. This page is the tale of my own effort to equip a small "laboratory" of sorts for just such purpose.

My first needs were related to my building projects and included meters for measuring capacitance and inductance. I have excellent-quality bench meters for other hobby projects including a Keithley 192 and 181 voltmeter with IEEE 488 computer interfaces. These are voltmeters only though, nothing else. My first purchase then was a capacitance/inductance meter. It worked great for capacitance, but the lowest range for inductance, 20mH unfortunately is useless. So, this meter was quickly followed by a stand-alone inductance meter ranging down to 200uH. I use these frequently when making or measuring coils, and measuring capacitors purchased on ebay, better to design a radio around a specific componant.

My next needs follow logically, what about the radios I build?, how do they do? For the non-electrical engineer, testing radios is a bit of a daunting task and this page is by no means intended to be a tutorial. I do give my experiences so that perhaps someone with the same idea may benefit and possibly avoid the mistakes I have certainly made. I describe my equipment and the protocols used in my testing, but I STRONGLY refer the reader to the following two pages for technical guidance and explanations, most of what I do came directly from here:
Dick Kleijer crystal-radio.eu http://www.crystal-radio.eu/enqmeting.htm
Gollum´s Crystal Receiver World
http://home.snafu.de/wumpus/gollum/testing.htm article by C.A. Lauter
I heartily express my appreciation to our European friends for making available on the web these excellent technical tutorials for crystal radio testing.

*****

Testing and understanding your set from the Antenna at the front, the tank in the middle and on to the Diode at the back and then the Radio complete, the whole shebang!

ANTENNA TUNING UNIT.. Design notes and Models, Effect of earth resistivity..

COIL Q.. Summary of web research on expected values of Q for various coil designs.

---

with a surface area of 1,600 square feet. The native lead occurred along a part of the margin of the ore body in a zone about 9 inches thick separated from the wall rock by clay gouge. The ore generally was argentiferous steel galena partly or completely altered to cerussite and anglesite with minor minerals including wulfenite, litharge (?), quartz and limonite. The native lead occurred as threads, sheets to one-half inch thick and pods up to 4xIX2.S inches.

sheared, massive, and 'steel' (a dense, extremely fine-grained, almost cryptocrystalline type). Sheared and steel galena textures were caused by post-ore shearing of coarse-grained galena and it was not uncommon to find the three varieties side-by-side. These textural variations had important implications for exploration because the more massive and free-grained types, which are less friable, were not affected as much by oxidation. As galena was released by erosion from the oxidized vein subcrops, it tended to develop a protective white rind of the Pb sulfate anglesite, and to collect downhill from the vein as a dispersion (float) train.

THE GEOLOGY AND MINERAL DEPOSITS OF THE TOBY-HORSETHIEF CREEK MAP AREA, NORTHERN PURCELL MOUNTAINS, SOUTHEAST BRITISH COLUMBIA., OPEN FILE 1990-26. By Alasdaire Pope
mine consist of massive panidiomorphic galena, sphalerite, pyrite, sucrosic cerussite and banded dolomite, galena, sphalerite and pyrite. In hand specimens bleischweif texture (steel galena), indicating deformation by simple shear (McClay, 1980) is frequently seen.

PROCEEDINGS OF THE THIRTY-THIRD ANNUAL MEETING OF THE N4INERALOGICAL SOCIETY OF AMERICA AT BOSTON, MASSACHUSETTS.
NATIVE LEAD, PRESIDIO COUNTY, TEXAS JORRN T. I_ONSDALAEN D KATHRYNO . DICKSON University of Texas, Austin, Texas
Approximately 3,000 pounds of ore containing native lead was recovered from three small tabular ore bodies in the western part of the shafter District, presidio county, Texas. Two of the ore bodies were vein-like in limestone with clay gouge walls. The metallic lead was localized at the centers of the veins. The third ore body was flat lying disk-shaped about 4 feet thick

DIODE TEST.. Describes protocols and results of testing many different diodes, diode classes, crystals, and even two different vacuum tube diodes.

DIODE PROTOCOL.. Describes a useful protocol for determination of Diode Is, n, and Ro.

DIODE CALIBRATION.. A simple set of modelled characteristics with variations in Is and n.

RADIO TEST.. Describes protocols and results of radio performance testing on my sets.

Conclusion

So, for an investment in a few key pieces of good equipment, yes at some non-trivial cost, you can appreciate the hobby so much more. I have an oscilloscope as well, but so far have not explored its value. I prefer the recording and spreadsheet methods as they give a better record of the results and allow graphing and easy comparison. Knowing how your sets perform, being able to test your diodes, knowing your coil inductance are all important aspects of knowing your radios. I certainly recommend this!
Ciao!

Steel galena owes its fine-grained texture either to mechanical deformation of larger crystals or to a beginning transformation to anglesite. It is supposed to be rich in silver but this is by no means always true. $PbS+4O=PbSO_4$ (Anglesite)

Geology of Colorado and Western are Deposits By Arthur Lakes, 1893.
A coarse galena is generally poor in silver, while fine grained "steel galena" is generally rich in silver, but the reverse may also be the case.

Waldschmidt, 1921. Hecla East Vein, Coeur d'Alene
The coarsest galena examined showed cleavage faces from 2 to 4 mm, square, but generally it is much finer grained than this. The fine-grained galena is commonly known as steel galena because freshly fractured surfaces have a steel-like luster. In all the specimens examed, this steel galena appeared to be in an intimate mixture of galena, tetrahedrite, sphalerite, pyrite, gnague, and other minerals present in the ore.
It appears that the original ore body consisted chiefly of galena, sphalerite, pyrite, pyrrhotite, tetrahedrite, freibergite, quartz and siderite. When movement took place along the O'Neil Fault and possibly along with minor slips, the ore body was grreatly crushed, during which time the galena acted as a putty-like mass for the suspension of broken fragments of the harder minerals. This breaking resulted in the formation of the typical steel galena.

Great mining camps of Canada 1. The history and geology of the Keno Hill Silver Camp, Yukon Territory.
Keno Hill mineralization "silver-lead-zinc veins in lastic metasedimentary terranes". Galena is present mainly as the common coarse-grained, well-crystallized and friable type, but also displays several other textures, such as fine-grained,

Steel Galena
Tintic Mining District: Mammoth Mine
Utah
TENTHS
TO INCH
1
2

Steel Galena from the Tintic Mining District, a classic silver locality in Utah. (The excessively sparkally surface in the photo is from the flash.) Note also the presence of Anglesite alteration, (the clearish anglesite crystals separated from the galena by a dark finer grained almost greenish alteration zone). This is good radio material, but you may need hunt a bit more for hot spots than the finer-grained stuff above.

Steel Galena Notes off the web:

USGS Bull 625, 1917
Galena of the Coeur d'Alene District contains appreciable silver, is generally not so well crystallized, and has a much less perfect cleavage.

Mineral Deposits, Waldemar Lindgren, 1919

ANTENNA TUNING

Notes on simulations for a Tuggle Front End

http://www.lessmiths.com/~kjsmith/crystal/atu.shtml

By Kevin Smith

Introduction:

When designing my radios I always make extensive use of Mike Peebles and Dan Petersen's "Professor Coil" spreadsheet. This tool along with the many tutorials and explanations online have vastly eased the design work around homebrew coil winding. We can now model a coil with good accuracy prior to the actual job and thus better match the coil to the other components we plan to include. One quickly becomes aware that most coils for the broadcast band vary around a nominal 220 or so uH. More than this and most variable capacitors will have too much bottom-end capacitance to tune the top of the band. Less than this and you need caps with fairly high values when fully meshed, 500+pF or more.

All the above is pretty straight forward. On the other hand, when I have been reviewing double-tuned set designs, I often note that the coil used for the ATU (antenna tuning unit) will have an inductance quite a bit lower than the main tank coil. Professor Coil doesn’t address this aspect of coil design and I have found little discussion online concerning ATU design. Mostly this seems to be dealt with in passing or as a digression when dealing with other subjects. The most useful web sites for addressing this aspect of crystal set design include Dick Kleijer's excellent work and Ramon Vargas's detailed analysis of the "Tuggle Front End", that’s about it. Neither site discusses the ramifications of varying different parameters such as the Earth Resistance, Coil Inductance, etc although Kleijier's site includes a great calculator page to allow one to ask these questions. In my explorations of ATU design (primarily Tuggle) I have made extensive use of Kleijier'e page and for this I am deeply in his debt. My hat's off to Dick, thanks so much.

The following discussion presents the results of many models cranked through the design calculator. I wished to understand what factors play main roles and which have minor parts.

For a conventional detector-circuit coil one can pretty much control all the main factors, really only the inductance and capacitance of the circuit. The main and only surprise comes in the form of "stray capacitance" resulting from the spacing of the coil wires. This can readily be 1) guessed at and 2) minimized by good coil winding technique. For the antenna circuit by contrast many of the needed parameters, earth resistance, antenna capacitance and inductance, and antenna resistance in the forms of actual wire resistance and radiation resistance are, for most of us, unknown and only guessed at. The following figure illustrates the antenna and ATU with a list of the main "components" that need to be understood and/or modeled.

The Antenna Equivalent Circuit consists of an AC voltage source Va (the signal of interest), in series with some radiation resistance Rr, antenna capacitance Ca, antenna inductance La, antenna resistance Rr, and finally a ground

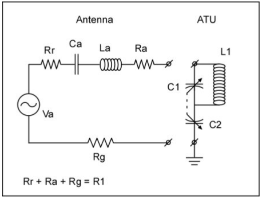

Steel Galena
Leadville Mining District
Colorado

From Leadville Colorado, the specimens below are the best examples of "steel galena" I have yet come across. The samples have a dull weathered surface skin deep hiding a massive, fine-grained lead-silver mining ore of a classic Rocky Mountain type. On a fresh surface the sample has a fine texture and flat grey color. This is not museum material, and its difficult (to impossible) to locate on ebay, but its the real mccoy when it comes to crystal radio. Don't even bother with the pretty pretty Missoury galena, it rightly belongs behind a glass display, not in a radio. Chips from this sample, potted in woods metal, rectifies strongly almost regardless of where the cat's whisker touches, instant sound.

Steel galena is difficult to find for sale due primarily to its lack of excellent crystal cleavage. Galena, as is well known, has beautiful LARGE crystal structure. Who would buy that ugly-looking grey junk?

Kevin

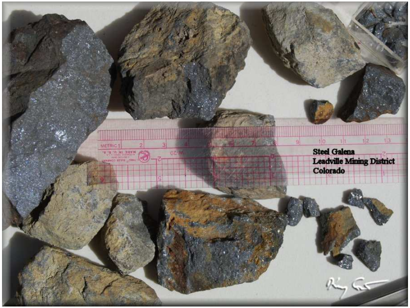

resistance Rg for the return path to complete the circuit. In addition the ATU consists of a coupling capacitance C2, and a tank with inductance L1 and capacitance C1. The capacitance and inductance of the ATU is needed to tune out the reactance of the various components of the antenna for which we do not have data. What to do?

Without an antenna analyzer (expensive) one can only estimate the many component values based on published antenna models. Ken Khun's engineering page has a number of nice models and a good discussion of the antenna parameters we will be trying to understand and use. For my personal setup my antenna consists of about 75' of 14awg wire averaging about 10-12' high and with another 25' of lead-in. For such an antenna Khun models a 30m antenna 3m high which approximates my situation closely enough. Wire resistance is negligible as is the radiation resistance which he shows to vary between some 0.1 to 1.5 ohms. Such an antenna is capacitive by nature and will have about 220 - 375pF along with some small 20uH inductance, not too far off a standard "dummy" antenna. This pretty much leaves earth resistance Rg as the main unknown parameter.

For Rg one has the option of punting and taking the "Standard" value of 25 ohms. In my modeling I have found this parameter to be critical and highly sensitive. I do not recommend guessing here, it is recommended one go to the internet page of their state, county, or local government and search for reports on ground or soil resistivity (or conductivity). As this is an important agricultural and engineering parameter, it has been surveyed for most places and should be available with some effort. Effort well rewarded. Earth resistivity for the Texas Gulf Coast is a mercifully low 10 - 15 ohm meters, but for many regions this will not be the case.

Earth resistivity depends on a number of variables including the material, moisture, mineral salt, and temperature. A table of typical ranges in ohm meters for some different soils follows:

Loam	5	-	50	ohm meter
Clay	4	-	100	
Sand/Gravel	50	-	1,000	
Limestone	5	-	10,000	
Sandstone	20	-	2,000	
Granite	1,000	-	2,000	
Slates	600	-	5,000	

Moisture up to about 17% dramatically lowers resistivity, mineral salts are needed and pure water is an insulator, and as the temperature approaches freezing the resistivity also rises dramatically. All this factors into your estimation of Rg. The actual resistivity seen by the circuit depends on the earth resistivity and the type of grounding system you have installed. The more metal in the ground, and deeper, the lower the resistance. Know thy earth! A map of USA soil resistivity follows:

"Substitution of other atoms for lead in galena is not very extensive. Among the elements which do accur in small amounts are: Sb, As, Bi, Ag, Ti, Zn, Cd, Fe, Mn and Cu. In many cases these may be present in impurity minerals (acanthite Ag2S, sphalerite, chalcopyrite, etc.) rather than in the galena structure."

Why is this interesting for the radio engineer? The following is my own theory, (speculation mostly). Most of you who have worked with galena detectors have noted that hot spots generally occur on or adjacent to fractures, lines or discontinuities on the crystal face, seldomly on the smooth lustrous face itself. Rectification occurs at the junction of two minerals, or of a mineral with a cat's whisker metal. Many explanations have been offered, most having to do with the position and nature of the crystal lattice at the point of contact, and the perceived presence of "impurities" (atoms other than lead) in the galena crystal matrix. As noted above, "Steel Galena" is a often an ore of silver containing not only galena, but abundant admixtures of associated minerals generally too small to see and thus seldomly described except in detailed mining reports as above. In addition, the very small grain size of common steel galena allows numerous physical and mineral discontinuities on the small detector surface area, and thus numerous hot spots. I speculate that many/most of these hot spots may in fact be at the junctions of galena and other minerals along with the light contact of the ever-so-sharp cat's whisker itself. Other galena (lead) deposits such as the Missouri / Tri-State region are not associated with silver mineralization (Tri-State ores typically have only about 30ppm Ag). These show large museum quality crystal faces but have poor detector properties.

"The coarsest galena examined showed cleavage faces from 2 to 4 mm, square, but generally it is much finer grained than this. The fine-grained galena is commonly known as steel galena because freshly fractured surfaces have a steel-like luster. In all the specimens examed, this steel galena appeared to be in an intimate mixture of galena, tetrahedrite, sphalerite, pyrite, gnague, and other minerals present in the ore."....
"It appears that the original ore body consisted chiefly of galena, sphalerite, pyrite, pyrrhotite, tetrahedrite, freibergite, quartz and siderite. When movement took place along the O'Neil Fault and possibly along with minor slips, the ore body was greatly crushed, during which time the galena acted as a putty-like mass for the suspension of broken fragments of the harder minerals. This breaking resulted in the formation of the typical steel galena."....

In this ore, galena is NOT the silver-bearing mineral. Rather, the silver is in the freibergite and to a lesser extent in the tetrahedrite. Very interesting. Quoting from classic mineralogy texts on galena I find the following:

Dana's Manual of Minerology: (Hurlbut)
"Pb 86.6, S 13.4 per cent. Silver is usually present, probably as admixtures of silver minerals such as argentite or tetrahedrite."

Elements of Mineralogy: (Mason and Berry)
"Commonly, galena is very nearly pure PbS. The silver, arsenic, and antimony reported in chemical analyses are largely due to the inclusions of argentite of tetrahedrite, small amounts of which are difficult to detect in a black opaque mineral."

An Introduction to the Rock Forming Minerals: (Deer, Howie and Zussman)

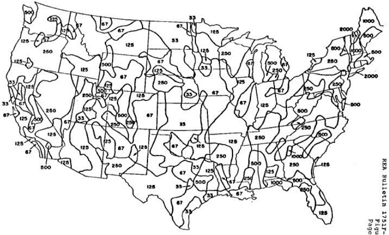

FIGURE 2

Estimated Average Earth Resistivity in U.S. (ohm·meters)

Modeling:

All the above discussion is great but... What is important really? The following section presents the results of applying different parameters to get a feel for the sensitivity of it all. To begin with I need explain my base assumptions. All the models are based on Kleijer's calculation spreadsheet which requires inputs for the following parameters:

Frequency
Coil inductance
LC circuit Q unloaded
* Complex impedance of antenna or
Series Resistance (Rg+Ra+Rr)
Series Capacitance Ca
Series Inductance La

* If the series values are given, the complex impedance is calculated.

1) Frequency is chosen per your interest, I have taken models at F = 550-1100-1700 kHz.

2) Coil inductance one has control over when winding. I have chosen base values that resonate with about 400pF tank capacitance, with some sensitivities.

3) Unloaded Q I do not have. For modeling I have taken Vargas's measures for a 4.5" coil wound with 660/46 Litz wire. I assume this will be about as good a coil as one can wind. I also made a sensitivity for lower Q. Q is a function of frequency and I have used the appropriate value of Q to match the modeled frequency.

4) Series Resistance is basically what the earth and ground system will deliver plus a small 1-2 ohm contribution from Ra and Rr. I have modeled four cases 10, 15, 25 and 50 ohms. The first two cases reflect my needs on the Gulf Coast with low resistivity soil. The 25 ohm case is a "typical" or "standard" grround. Finally, many will have much higher resistivity soils and you need to know just how difficult things can get!

5) Series capacitance I based for my specific antenna on the model of Ken Kuhn (Mathematical Model of Wire Antenna). The capacitance of a 30m antenna 3m high he calculated to range from 220pF and low frequencies to 375pF and high frequencies. I input the correct capacitance to match the frequency modeled.

6) Series inductance I just input 20 uH every time.

Some thoughts on "Steel Galena" and its use as a crystal detector/rectifier for radio waves.

http://www.lessmiths.com/~kjsmith/crystal/sgalena.shtml

I have recently become interested in crystal radios and am especially interested in the "cystal" in the radio. This is due to my background as a geologist / mineralogist / petrologist / oil exploration geologist. I have read in several radio websites concerning steel galena describing it as argentiferous as well as fine grained and useful as an excellent detector with numerous hot spots. While commonly associated with silver ores, I have never read of galena itself as being much silver bearing as a mineral. This puzzle sent me to the web to see what I could find on this "steel" variety.

Steel Galena does indeed associate most closely with silver ore deposits, especially in the classic Rocky Mountain silver districts like Leadville in Colorado, Coeur d'Alene in Idaho, and Tintic / Park City of Utah. In all cases the ore bodies themselves are in sheared and altered sedimentary rocks. Mineralization is due to hydrothermal movements of water with movement aided by the sheared, and crushed zones having fine to extremely fine grain size, hence its name "steel".

To give a sense of the deposits themselves I wish to quote a couple accurate petrological descriptions:

Waldschmidt, 1921: Hecla East Vein, Coeur d'Alene

The data and plots for four different scenarios follow:

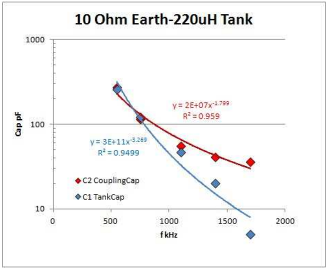

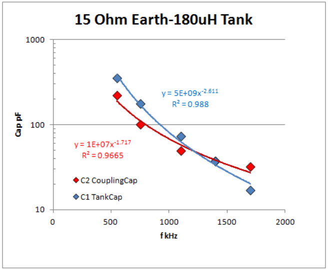

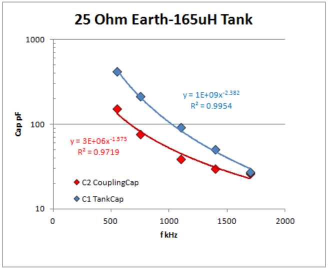

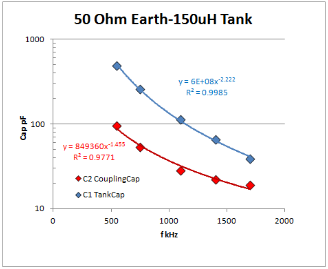

The above plots show the calculated capacitance versus frequency for four different models of Earth Resistivity / Tank Inductance. In each case I maintain the same log capacitance

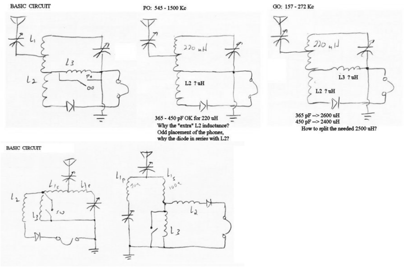
BASIC CIRCUIT
L1
L2
L3
Po
Go
PO: 545 - 1500 Kc
220 uH
L2 ? uH
GO: 157 - 272 Kc
220 uH
L3 ? uH
L2 ? uH
BASIC CIRCUIT
L2
L1s
L1e
L3
sw
L1p
L1s
L2
L3

365 - 450 pF OK for 220 uH
Why the "extra" L2 inductance?
Odd placement of the phones,
why the diode in series with L2?

365 pF --> 2600 uH
450 pF --> 2400 uH
How to split the needed 2500 uH?

Innards Photograph from an actual set..

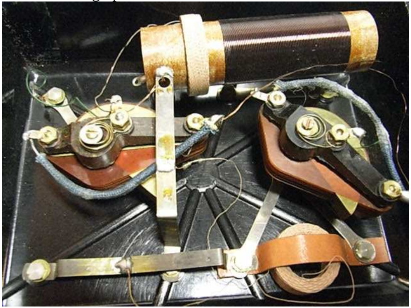

calculations from above:
LW tuning with assumed 450pF cap is 2300 - 2400 uH

The following photos show the set and the references mentionned above.

(some useful translations for my anglophone friends:
PO = Petits Ondes = Short waves [so named prior to the utility of the HF bands was known]
GO = Grandes Ondes = Long Waves
Bobine Addition = Additional Coil)

Jacquemard Schematic and layout..

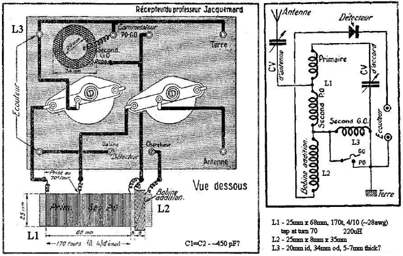
Récepteur du professeur Jacquemard
L3
Commutateur PO-GO
Terre
20 mm
Second. GO
34 mm
Ecouteur
Galène
Détecteur
Chercheur
Antenne
Prise au 70e tour
Vue dessous
Bobine addition.
25 mm
Prim.
Sec. PO
L2
L1
68 mm
3
8
170 tours fil 4/10 émail
C1=C2 - ~450 pF?
Antenne
Détecteur
CV d'antenne
Primaire
L1
Second P.O
CV d'accord
Second G.O.
L3
GO
PO
Ecouteur
Bobine addition.
L2
Terre

L1 - 25mm x 68mm, 170t, 4/10 (~28awg)
tap at turn 70 220uH
L2 - 25mm x 8mm x 35mm
L3 - 20mm id, 34mm od, 5-7mm thick?

Circuit Analysis/Redraws of above..

---

vs frequency scale for easy comparison. The cases modeled represent increasing earth resistance presented to the ATU from a very low 10 ohms to a fairly high 50 ohms. Actual earths can go up to two orders of magnitude higher. I chose L1 inductances such that the maximum needed capacitance would approach 500pFs, easily found on many variable caps. With respect to the coupling cap (C2, red curve), as the earth resistivity increases the needed capacitance declines. The tank cap value needs to resonate with the inductor and increases with smaller-value inductance. In searching for a 500pF max cap value, the choice of inductance needs to decrease with increasing earth resistance. Parameters other than Rg, L1 and C1-C2 have minor impact on the models.

Conclusions:

From the plots one can readily see that above 10 ohm earth the two variable capacitors on the ATU do not track well. As the earth resistivity increases, the worse the tracking. Most locations are not blessed with a low resistivity earth and this needs to be factored into the ATU design. There is little to be done, changing the design coil inductance will change the needed capacitance on BOTH C1 and C2, (larger L1 leads to smaller C1 and C2 for resonance and vice versa). A possible solution well worth trying is to use a dual-gang capacitor where one section has a different value than the other, the above plots give an easy way to decide the max values needed. Use the smaller gang on the coupling circuit and the larget on the tank. Experience tells us that ganging the capacitors on a "Tuggle" front end works well, but from the models I have to imagine one might squeeze a bit more performance by giving up the convenience of one-dial tuning on the ATU, especially where your earth has a fairly high resistance.

The best design concept will be to know first and critically your earth resistance. If this is just a guessed-at parameter then I strongly urge you to put down the Litz, set aside the silver-plated caps, store the holy-grail diodes and go do some research on your local ground. All the time, expense, and effort on the greatest "state-of-the-art" crystal receiver will be wasted if the ground is not attended to. If the resistivity of your earth is high in ohm-meters, get more metal in the ground or consider a counter-poise. The following table from LEM Instruments shows the relation between earth resistivity in ohm meters and actual earth resistance in ohms as a function of the earthing method used.

Models and Data:

10 ohm / 220 uH Lowest resistivity Earth Case sensitivities

Ohm Rg	pF Ca	uH La	uH L1	Qu	kHz f	cpl C2	tank C1
10	220	20	220	475	550	442	228
10	300	20	220	400	1100	68	36
10	375	20	220	200	1700	44	-3
10	220	20	220	700	550	273	255
10	260	20	220	640	750	114	122
10	300	20	220	585	1100	55	47
10	340	20	220	440	1400	41	20
10	375	20	220	300	1700	36	5
10	220	20	200	700	550	304	287
10	300	20	200	585	1100	58	54
10	375	20	200	300	1700	38	7

Professeur Jacquemard Poste a Galene
http://www.lessmiths.com/~kjsmith/crystal/jmrd.shtml

A bit of a mystery this one, I am hankering to build a breadboard repro of the set if only to play with this bizarre (to me) circuit.

data and information on the set is not small, but not enough:

from the schematic:
L1: form 25mm dia, 68mm long selenoid, 170 turns 4/10 wire
calculation- 0.4mm-26awg wire, 220uH coil
L2: ID 25mm, W 8mm. OD, #turns, guage not given
L3: ID 20mm, OD 34mm. W, #turns, guage not given

estimation from photo:
L2: OD about 32-35mm, wire guage is small in the 32+ range
L3: W maybe 6-8mm, wire guage is small in the 32+ range

from book:
MW = 545 - 1500 Kc
LW = 157 - 272 Kc

from calculations:
MW tuning with 220uH typically 365 - 450 pF

from clues in the two publications
1. most Vcaps are 450pF
2. most GO coils 120turns x two coils

15 ohm / 180 uH Low resistivity Earth Case sensitivities

Ohm	pF	uH	uH	Qu	kHz	cpl	tank
Rg	Ca	La	L1		f	C2	C1
15	220	20	180	475	550	337	328
15	300	20	180	400	1100	61	63
15	375	20	180	200	1700	40	10
15	220	20	180	700	550	220	352
15	260	20	180	640	750	100	176
15	300	20	180	585	1100	49	73
15	340	20	180	440	1400	37	37
15	375	20	180	300	1700	32	17
15	220	20	130	700	550	313	511
15	300	20	130	585	1100	59	110
15	375	20	130	300	1700	38	30
15	220	20	205	700	550	195	302
15	300	20	205	585	1100	45	61
15	375	20	205	300	1700	30	13

25 ohm / 165 uH "Standard" resistivity Earth Case sensitivities

Ohm	pF	uH	uH	Qu	kHz	cpl	tank
Rg	Ca	La	L1		f	C2	C1
25	220	20	165	475	550	213	396
25	300	20	165	400	1100	48	84
25	375	20	165	200	1700	32	22
25	220	20	150	700	550	164	462
25	300	20	150	585	1100	41	102
25	375	20	150	300	1700	27	31

25	220	20	165	700	550	151	416
25	260	20	165	640	750	76	213
25	300	20	165	585	1100	39	91
25	340	20	165	440	1400	30	50
25	375	20	165	300	1700	26	27
25	220	20	200	700	550	130	335
25	300	20	200	585	1100	35	72
25	375	20	200	300	1700	24	20

50 ohm / 150 uH High resistivity Earth Case sensitivities

Ohm	pF	uH	uH	Qu	kHz	cpl	tank
Rg	Ca	La	L1		f	C2	C1
50	220	20	150	475	550	128	476
50	300	20	150	400	1100	34	108
50	375	20	150	200	1700	24	35
50	220	20	150	700	550	96	490
50	260	20	150	640	750	53	256
50	300	20	150	585	1100	28	113
50	340	20	150	440	1400	22	65
50	375	20	150	300	1700	19	39
50	220	20	125	700	550	110	595
50	300	20	125	585	1100	31	139
50	375	20	125	300	1700	21	49

Kevin Smith
09/2011

CHAPTER V

Other Thoughts and Speculations

Coil Q:

http://www.lessmiths.com/~kjsmith/crystal/coilq.shtml

Kevin Smith

Introduction:

I have been building and studying crystal radios for some time now and slowly begin to learn a few things about these marvelous sets. In this section I begin my exploration of coil, coil quality and that mysterious dimensionless factor, Q... This page thus is a bit preliminary as I have not yet begun any measurements to determine the Q of my coils. Please bear with me.

My purpose here is to present some facts and data that has resulted from my explorations of the web. I have often wondered at what the quality factor of my coils should be and as often realized that I really do not have any expectations as to what is possible. Now, having dome some research, I can with some confidance say that this is solved. Coils as used in crystal radio broadcast band reception typically employ coils with Q factors ranging from 100+ (pretty lousy) through the several 100's (decent) and on up to 1000 or more for those remarkable Big-Litz wonders ($$$). Knowing the expected (or actual) Q of your coil is important in so far as it impacts the choice of diode to be used in the set. On my page of Diode Calibration I present a summary graphic which indicates how the diode Rd value relates to the tank parallel resistance Rp. This Rp in turn is a function of the coil Q. Schezzam! So here we are.

In your set construction, with a good effort and good engineering practice one can easily expect to wind a solenoid coil in the Q = 200 range without much trouble, even with a cardboard form, sealed, of course. With various "open" coils, spiderweb, basket weave, diamond weave, etc. you may expect to double that.. possibly. The advantage in open coils derives from first the separation between adjacent wire turns which reduces self-capacitance in the coil, and secondly from the obvious lack of a form. All materials used in the coil will have some amount of dielectric losses associated with them and the less material used the better. Air core coils are best in this respect. I am not here to speak about ferrite coils as I have no experience with them. Frankly, they seem (to me) a bit like cheating.

The following graph presents data that I have scoured off the web. My primary sources include Hund and Groot, 1925*, Wes Hayward, Dave Schmarder, Ken Khun, Mike Tuggle, Steve Ratzlaff, and Dick Kleijer. 10,000 thanks for those who post their data on the web! The following plot gives the Q value as a function of coil Rs (series resistance. In striving for a high Q coil, in effect one is eliminating losses and lowering the series resistance as much as possible. There is more to it of course. The Q formula Qu = 2pifL / Rs tells us that Q is also a function of the coil inductance L and the frequency f of the measurement. All measurements chosen for plotting are made around 1 Mhz and most of the coils are in the L = 200 - 350 uH range. So, despite the formula, in this plot the coil Rp is the main driver. From the plot it should be apparent that winding a coil with Q = 200 or better should be a no-brainer. If your coil Q is less, you just aren't trying. At the high end, Q's > 1000 seem to be pretty extreme and these coils are expensive. You better be using big Litz 660/46 and use a basket design (although the two best coils on the plot were solenoids).

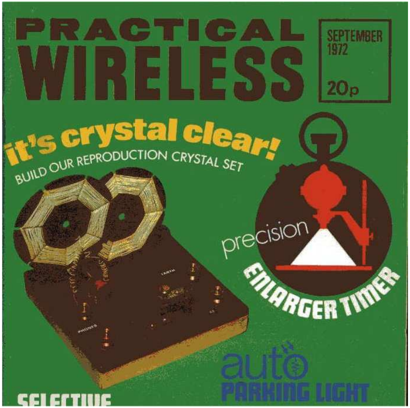

Practical Wireless, September 1972 cover.

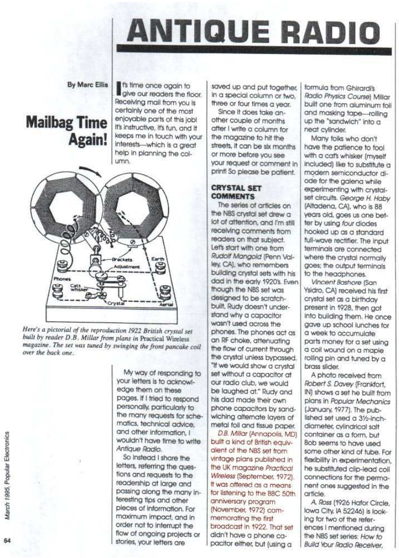
Brackets
Adjustment
Earth
Phones
Cats whisker
Crystal
Aerial

ANTIQUE RADIO

By Marc Ellis

Mailbag Time Again!

It's time once again to give our readers the floor. Receiving mail from you is certainly one of the most enjoyable parts of this job! It's instructive, it's fun, and it keeps me in touch with your interests—which is a great help in planning the column.

My way of responding to your letters is to acknowledge them on these pages. If I tried to respond personally, particularly to the many requests for schematics, technical advice, and other information, I wouldn't have time to write *Antique Radio*.

So instead I share the letters, referring the questions and requests to the readership at large and passing along the many interesting tips and other pieces of information. For maximum impact, and in order not to interrupt the flow of ongoing projects or stories, your letters are saved up and put together, in a special column or two, three or four times a year.

Since it does take another couple of months after I write a column for the magazine to hit the streets, it can be six months or more before you see your request or comment in print! So please be patient.

Here's a pictorial of the reproduction 1922 British crystal set built by reader D.B. Millar from plans in Practical Wireless *magazine. The set was tuned by swinging the front pancake coil over the back one.*

CRYSTAL SET COMMENTS

The series of articles on the NBS crystal set drew a lot of attention, and I'm still receiving comments from readers on that subject. Let's start with one from *Rudolf Mangold* (Penn Valley, CA), who remembers building crystal sets with his dad in the early 1920's. Even though the NBS set was designed to be scratch-built, Rudy doesn't understand why a capacitor wasn't used across the phones. The phones act as an RF choke, attenuating the flow of current through the crystal unless bypassed. "If we would show a crystal set without a capacitor at our radio club, we would be laughed at." Rudy and his dad made their own phone capacitors by sandwiching alternate layers of metal foil and tissue paper.

D.B. Millar (Annapolis, MD) built a kind of British equivalent of the NBS set from vintage plans published in the UK magazine *Practical Wireless* (September, 1972). It was offered as a means for listening to the BBC 50th anniversary program (November, 1972) commemorating the first broadcast in 1922. That set didn't have a phone capacitor either, but (using a formula from Ghirardi's *Radio Physics Course*) Millar built one from aluminum foil and masking tape—rolling up the "sandwich" into a neat cylinder.

Many folks who don't have the patience to fool with a cat's whisker (myself included) like to substitute a modern semiconductor diode for the galena while experimenting with crystal-set circuits. *George H. Haby* (Altadena, CA), who is 88 years old, goes us one better by using *four* diodes hooked up as a standard full-wave rectifier. The input terminals are connected where the crystal normally goes; the output terminals to the headphones.

Vincent Bashore (San Ysidro, CA) received his first crystal set as a birthday present in 1928, then got into building them. He once gave up school lunches for a week to accumulate parts money for a set using a coil wound on a maple rolling pin and tuned by a brass slider.

A photo received from *Robert S. Davey* (Frankfort, IN) shows a set he built from plans in *Popular Mechanics* (January, 1977). The published set used a 3½-inch-diameter, cylindrical salt container as a form, but Bob seems to have used some other kind of tube. For flexibility in experimentation, he substituted clip-lead coil connections for the permanent ones suggested in the article.

A. Ross (1926 Hafor Circle, Iowa City, IA 52246) is looking for two of the references I mentioned during the NBS set series: *How to Build Your Radio Receiver,*

Popular Electronics, March 1995

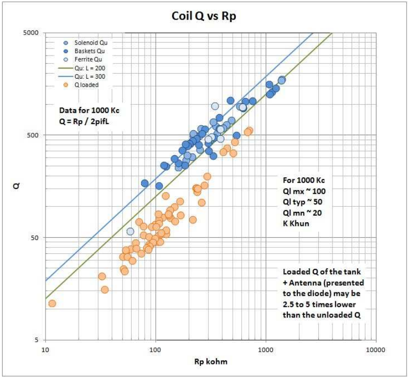

On the plot I also post in orange circles my estimate of what the loaded Q presented to the diode (Antenna + Tank) might look like. Loaded Q will always be lower than the unloaded Q by several times. I have estimated that high-end sets (big litz, silver-plated ceramic insulated caps, best wiring practices) may lower the Q by about 2 1/2 times (B Tongue's performance set has Q = 700). At the low end (vintage components, small solid wire coils, taps) the load may lower Q by up to 5 times. Ken Khun in his excellent web book states that typical sets at 1Mhz have a loaded Q between 20 and 100, 50 typical and this is where most of the data falls. I scaled the divisor by Qu to produce the above plot but note that this is merely an estimate. Somewhere some bloke has no doubt tuned a big litz coil with cheapo capacitors. Having an idea of your set's Ql will allow a better selection of the proper diode for matching.

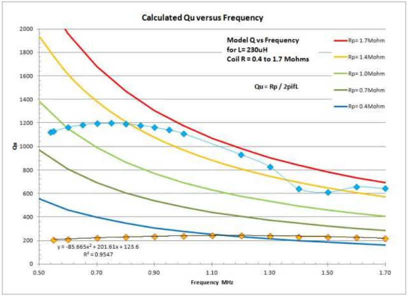

The above plot is from a series of measurements on two of my coils kindly made for me by Steve Ratzlaff, AA7U. The blue diamonds are the measurements on a 660/46 Litz basket beast and the yellow diamonds are measurements on a modest 18awg tapped solenoid wound of a cardboard form. The data from the low Q coil form a lovely continuum across the BCB band while the litz coil data above 1MHz or so seem to be declining and erratic. Steve measured the coil a great many times and found good consistant results. The measurements are excellent but I must say, I am a bit suspicious that the high-Q coil may be at the limit of the HP Q-meter calibration, or perhaps 660/46 litz Q tops out at 0.7-0.8 MHz and crashes above that. (I have not yet found time to persue this). The plot

---

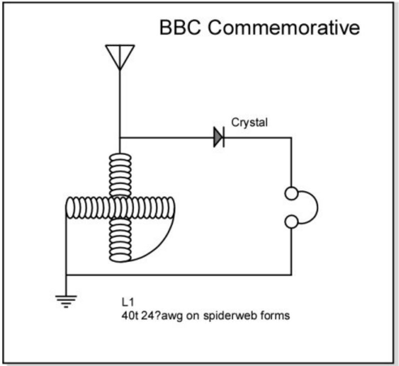

Circuit Diagram of same.

plans? Or, was this a kit? As I cannot locate an actual copy of the Sept 72 journal, I still can only speculate. I do hope it is a true homebuilt.

The following photos show the set and the references mentionned above.

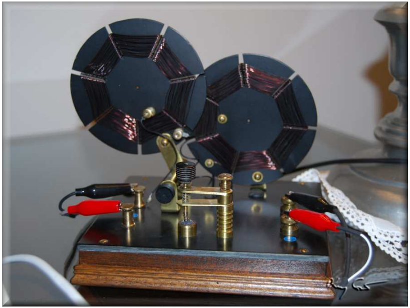

Portrait of the beauty..

also includes Q vs f curves at constant Rp values. It is clear that, while tuning across the BCB spectrum, the Rp of the coil changes continuously.

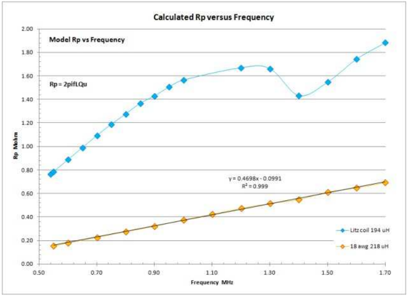

Calculating the coil Rp from the data (Q:C:f), I have reproduced the above graph of the coil Rp vs frequency. It is evident that the trend follows a perfectly straight line for the low-Q coil. I would expect a similar straight-line relation for the litz coil but again things look strange. My impression here is that of a much steeper Rp vs f relation where the calibration on the Q-meter breaks down near 1MHz (Rp ~1.5 MOhm) and the data wanders until a second linear trend (with about the same slope) is re-established between 1.4 to 1.7MHz.

Now time for a few calculations, just how good can a coil Q get?

The following calculation is made to determine the AC resistance of the Litz wire used in a coil:
From different Litz wire manufacturing sites one can find tables and data allowing the calculation of the resistance of your favorite litz wire.

The formula for the D.C. resistance of any Litz construction is:

Rdc = Rs (1.0515)^Nb * (1.025)^Nc / Ns

Where:
Rdc = Resistance in Ohms/1000 ft.
Rs = Maximum D.C. resistance of the individual strands (4544 for 46awg wire)
Nb = Number of Bunching operations (assume = 2)
Nc = Number of Cabling operations (assume = 1)
Ns = Number of individual strands (assume = 660)

Rdc = 4544 (1.015)^2 (1.025)^1 / 660 = 7.27 ohms / 1000 ft.

The ratio of AC resistance to DC resistance of any Litz construction is:

Rac/Rdc = S + K (N Di / Do)^2 * G

Where:
S = Resistance ratio of individual strands when isolated (1.0003 for 46awg wire)
G = Eddy Current basis factor = (Di * sqrt(f) / 10.44)^4
F = Operating Frequency in HZ (assume 1MHz)
N = Number of strands in the cable = 660
Di = Diameter of the individual strands over the copper in inches = 0.0016

BBC 1922 - 1972 Commerative
http://www.lessmiths.com/~kjsmith/crystal/bbc.shtml

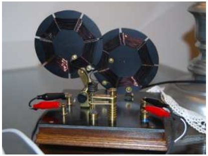

A bit of a mystery this one, I found this radio on ebay and it was just SO pretty that I had to have it. The circuit is a simple variometer of pancake-coil design. Being a variometer has its drawbacks such as limited sensitivity and gawd-awful selectivity. If this piece wasn't so darned cute I would let it adorn the shelf indefinately. The set does have one mysterious element for me. Who made it? When? Some research on the web resulted in two interesting discoveries and an added mystery. On Darryl boyd's excellent web site I found in the plans section a mention of this very set, or one very near to it taken from the March 1995 issue fo Popular Electronics. The reference was to the September 1972 British journal Popular Wireless, the text I have shamelessly copied and placed below, the critical text highlighted in red. I subsequently found an image of the cover of the Prac Wireless issue with the set nicely featured. So, I have to assume this set is homemade at or soon after the publication of the plans in that issue. I assume the set was made in USA, but cannot be sure at all.

The new mystery? well, if you look closely at the radio I have, the drawing in the Pop Electronics article and the photo of the radio on the Prac Wireless cover, you will see virtually the EXACT SAME set right down to the hardware and placement. Only a few minor differences. My mystery then is this, was this an extremely faithful homebuilt version from detailed

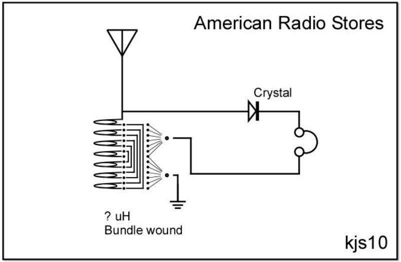

Circuit schematic for the set.

AMERICAN RADIO STORES, INC.

Portrait, set and box.

Do = Diameter of the finished cable over the strands in inches = 0.056
K = Constant depending on the number of strands = 2

Rac/Rdc = 1.0003 +2(660*0.0016 / 0.056)^2 * (0.0016*1000 / 10.44)^4 = 1.393

Therefore the AC Resistance of 660/46 litz wire is approximately:

The A.C. resistance is: 1.39 * 7.27 = 10.13 ohms/1000ft.

A 5inch diameter basket weave coil made from 660/46 Litz wire having an inductance of 230 uH will typically require some 40 turns or about 53ft of wire. At 10.13 Ohm/kft, that comes to an AC wire resistance of 0.53 ohms for the coil alone. Were all the losses represented by series resistance of the wire, the coil would have, at 1 MHz an unloaded Q = 2700! Naturally, wire resistance is not the only source of loss in the tank. There is a capacitor, metallic objects intruding into the magnetic field of the coil, diaelectric losses, eddy current and other losses. Its no wonder that the best coils just top out above Q = 1000 or so.

This is where I leave things. Time to get measuring. Below I provide my input data, have at it!

Kevin Smith

*Note that in 1925 the importance of coil Rs was understood, but the factor Q was apparently not used. I have taken (quite painfully) the L and Rs data from their plots and calculated the resulting Q. I recommend to those interested to download the pdf of their paper.

Radio-Frequency Resistance and Inductance

Qu = 2pifL / Rs

		1Mhz		
Coil Type	Wire	Rs	L	Q
		awg	ohm	uH
solenoid	660/46	0.90	200	1400
solenoid	660/46	0.91	200	1375
solenoid	175/46	2.88	265	579
solenoid	12	2.98	230	485
solenoid	14	3.28	230	440
solenoid	16	3.52	230	410
solenoid	18	4.01	230	360
solenoid	50/46	4.89	275	353
solenoid	32/38	6.20	327	331
solenoid	20	4.66	230	310
solenoid	16	7.60	319	264
solenoid	28	8.80	360	257
solenoid	22	5.78	230	250
solenoid	24	8.10	319	247
solenoid	28	10.10	360	224
solenoid	28	10.30	360	220
solenoid	50/46	10.84	378	219
solenoid	24	6.72	230	215
solenoid	28	9.60	327	214
solenoid	28	9.70	319	207
solenoid	26	8.03	230	180
solenoid	28		327	
Basket weave	660/46	1.03	186	1134
Spider	660/46	1.38	241	1000++
Basket weave	660/46	1.08	186	1082
Spider	660/46	0.90	150	1000+
spider	660/46			816

Inside I expected to find a typical tapped selenoid, the box is certainly large enough for one. What I found suprised and rather shocked me. It still does! The coil is bundle-wound, a mass of 26-30 dcc wire about 2.25 inches diameter. I was unaware that such a design would work. It is impossible to check how the tap wires are attached, they seem to just get wound into the coil with no apparent connection. It appears to be continuous wires from one switch set wound around and around and then exiting to the second switch set. That sounds like eight independent coils wound together. Another odd circuit characteristic is the fact that the tap points between the two tap switches are wired to each other (see photo and schematic). Perhaps connecting in such a manner combines them into a single longer coil, hard to speculate. How this coil works is different from anything I have experienced in crystal radio. Completing the schematic, rectification takes place via an ordinary cat's whisker and small potted galena crystal, whew! This is certainly an interesting design if not otherwise efficient or selective. It is a welcome addition to my small menagerie of vintage sets.

A
G
NES
AMERICAN RADIO STORES, INC.

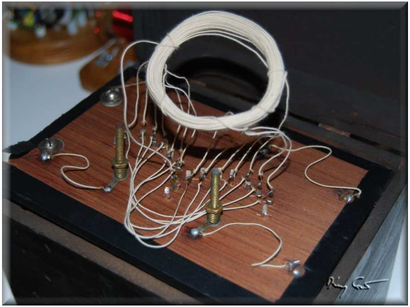

Spider	165/46	1.94	232	750
Basket weave	175/46	2.56	265	650
spider	660/46			641
Spider	100/45	2.41	238	620
Spider	100/45	2.44	241	620
Spider	100/45	1.73	149	540
spider	175/46	4.94	373	474
Loose Basket	32/38	5.30	317	376
Spider	40/44	4.16	248	375
Spider	40/44	4.16	225	340
Spider	40/44	2.93	154	330
Spider	32/38	7.50	331	277
Loose Basket	24.00	7.60	317	262
Basket weave	22.00	6.82	265	244
Loose Basket	28.00	8.50	317	234
Spider	24.00	9.50	327	216
Basket weave	32/38	10.40	332	201
Spider	28.00	10.60	327	194
Honeycomb	32/38	12.00	355	186
Spider	28.00	12.00	330	173
Basket weave	24.00	13.80	323	147
Basket weave	28.00	16.40	323	124
Honeycomb	24.00	18.50	347	118
Honeycomb	28.00	27.50	347	79

ferrite rod	50/46	2.58	250	610
toroid	22	3.93	240	384
toroid	18	3.20	195	383
toroid	22	8.20	449	344
toroid	175/46	4.82	244	318
toroid	50/46	5.23	248	298
ferrite rod	22		80	
ferrite rod	22	16.94	158	58.6

2Layer Bank	32/38	9.40	327	219
3Layer Bank	32/38	13.00	336	162
2Layer Bank	24	14.20	323	143
2Layer Bank	28	16.80	323	121
4Layer Bank	32/38	19.50	360	116
3Layer Bank	24	21.00	333	100
3Layer Bank	28	24.00	333	87
4Layer Bank	24	29.00	336	73
4Layer Bank	28	32.50	336	65
Double Layer	28	inf	355	0

American Radio Stores, Inc.

http://www.lessmiths.com/~kjsmith/crystal/ars.shtml

An interesting vintage crystal set recently came up for bidding on Ebay, an American Radio Stores, Inc radio in good working condition. This set interested me because, from the photos provided, it looked to have a fairly classic 1920's circuit with tuning accomplished via two switches to taps on the coil. I presumed that they are units and 10's taps as outlined on the classic Bureau of standards Circuar #121 set. A modest bid brought the radio to my doorstep a couple weeks later.

Looking at the radio when it arrived showed me that this was, even for the time, a fairly inexpensive set lightly built with cheap but workable parts. As expected, the radio has fair sensitivity and abysmal selectivity. (Recall, I bought it for the classic design not its DX potential). Somewhat to my consternation, the front panel was attached to the box by four small brads that could not be removed. No chance to view the inductor in the interior without compromising the vintage nature of the radio. I agonized over the question of whether to tear into the box or preserve the radio. After deliberating some 30 - 40 nanoseconds I was scheming for the most efficient and least destructive procedure to access the interior. My handy Dremmel neatly sliced the small brads one by one leaving no marks on the box or panel.

An Analytical Approach to the measurement of coil Q

http://www.lessmiths.com/~kjsmith/crystal/coilqm.shtml

Kevin Smith

Background:

Measuring the values of your radio components is an interesting and fun part of the Crystal Radio Hobby. Not only does it help you design better radios, but helps to understand the physics behind the wires and plates. The radio coil, one of the principal components is generally that part wound by hand. All other components, diode, variable capacitor are usually bought ready-made. The coil sits at the heart of the set and yet it remains devilishly difficult to measure and characterize. Even cheapo inductance meters, (I mean, my cheapo meter) do not properly measure the coil inductance at radio frequency. Coil quality, well that is a whole other matter. Quality, or the Q-factor as it is often called is a bench-intensive and measurement-intensive proposition that even the most dedicated radio fan may shy away from.

If you wish to know your coil Q-factor you have the choice of tracking down and paying a lot of money for an old HP or Boonton Q-Meter, or setting up a test bench and making the required measurements yourself. A number of techniques for measuring the Q of a coil have been published and excellent summaries can be found in "Q Factor Measurements on L-C Circuits" by Jacques Audet, and "Experiments with Coils and Q-Measurement" a web page by Wes Hayward.

All the techniques discussed involve variously setting up a parallel or series tank containing the coil to be tested and an accompanying capacitor to tune it to the needed measurement frequency. The tank is generally energized with a signal generator and measured with an oscilloscope or sensitive digital voltmeter. Other components often may include attenuators, SWR analyzers, Spectrum analyzers etc. Sources of error enter with coupling, loading, uncertainty as to the actual internal resistance of the source generator, and the need for several independent measurements which are then multiplied or divided together, adding additional error. All this works very well for the engineer with good bench practice, good equipment ($$), and the patience to perform the measurements several times to check repeatability. All these things plus the required anality coefficient the current author lacks.

In this paper I propose an alternative approach that takes an analytical look at the oscillating tank waveform and determines the coil Q from that. This technique dispenses with a majority of the equipment involved with traditional methods but does require a digital oscilloscope. In recent years the affordability of these scopes has greatly increased. If you have a digital scope, or access to one, or were looking for an excuse to purchase one, then read on, this technique may be for you.

Some Theory:

The inspiration for this technique is certainly not new. I first found the following simple circuit, (Figure 1) on page 22 of Bucher’s 1919 "Wireless Experimenter’s Manual". Having a good deal of experience capturing and evaluating damped radio oscillations for a Spark Gap transmitter, I immediately saw the utility of the circuit for a simple Q determination based on the damping decrement of the tank.

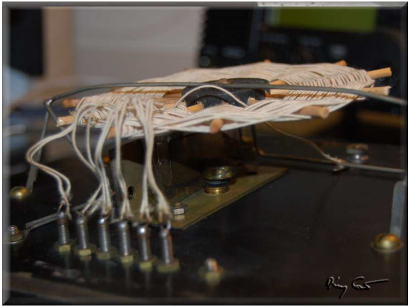

A peek, such as it is, at the set condensers and cam tuning assembly.

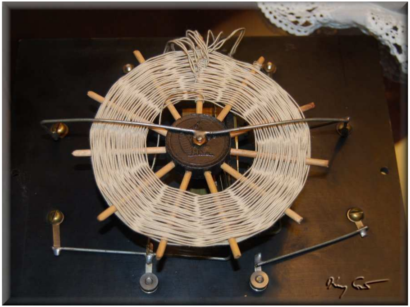

The spiderweb coil with vintage cloth-covered wires, taps and wireing.

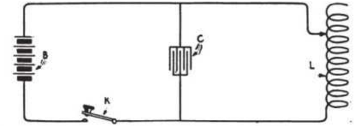

FIG. 8. Simple oscillation circuit illustrating the production of radio frequency currents.

Figure 1. Simple oscillation circuit, Bucher 1919.

As a professional Geologist, I must admit that most math involving more than ten fingers and ten toes (for advanced computations) is not my specialty. As such I will keep this aspect to an absolute minimum. Still, a bit of algebra may be useful to explain damping and logarithmic decrement. RF oscillations in a tank circuit are damped due to losses (primarily resistive) associated with that circuit. The amount of damping thus is related to the quality, or Q of the tank. In old texts the damping is generally determined by measuring the "Logarithmic Decrement" or the amplitude of successive oscillation peaks and taking the log of the ratio.

$$d = \ln(A1/A2) \qquad (1)$$

A more generalized version of this formula taking into account measurements over many periods can be expressed as:

$$d = 1/n \ln(Ao/An) \qquad (2)$$

Where n is the number of periods analyzed, Ao is the amplitude of the first peak and An is the amplitude of the peak n periods away. This equation allows a very simple determination of the log decrement with high accuracy. Because damping is due to resistive losses in the circuit, it can be related to the circuit components as follows, (Bucher, 1919):

d = pi (R / 2pifL) (3)

The angular frequency is often expressed as w= 2pif so this can be substituted into equation (3) above to simplify as

d = pi R / w L (4)

We turn now to looking at Q and its relation to the circuit components. The general expression for Q is:

Q = 2pifL / R (5)

Simplifying to

Q = w L / R (6)

With the term w L in both expressions, we can solve for w L in each and set them equal to each other:

QR = piR / d so:

Q = pi / d (7)

Dreadful, wasn't it? The inspiration here is that with a simple determination of the logarithmic decrement, deriving the circuit Q is a trivial exercise even a geologist can manage. No signal generators, no attenuators, no multiple measurements. All this should serve to reduce sources of error and give a faster and easier way to determine the quality of that most central component of your crystal radio, the coil.

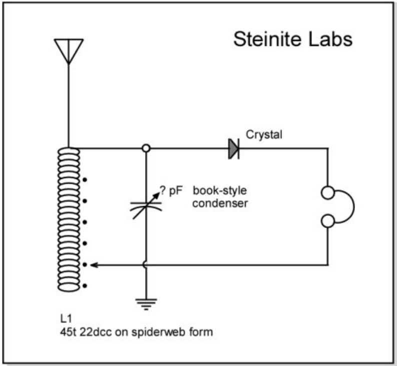

Circuit as best I can make out

Portrait of the Steinite set

Setup:

The following schematic shows the setup, (Figure 2). In this configuration, I have used a simple 9V battery to stimulate the coil and have the scope connected to the tank via a coupling coil. An earlier configuration with a power supply resulted in too much noise and ripple to obtain a good reading. The tank is coupled lightly to the scope with a coupling coil. This coupling coil is simply two turns of hookup wire about 4 inches in diameter placed an inch or two from the coil to be tested and I have found no reason to space it further. Between the coupling coil and oscilloscope I utilize the 1:100 probe although testing with my 1:10 probe showed little impact. In this way, the coil under test is clean and well isolated from the measurement circuit.

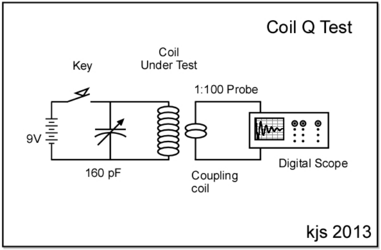

Figure 2. Schematic of test circuit.

For the tank I have used a small CT1C150 military spec cap ("APC"-type) with a capacitance range from about 20 to 160 pF. These are silver or nickel plated (depending on the source of information) and ceramic (steatite) insulated and should have a high Q with little impact on the test. At least this is the general assumption used in most Q-measurement articles I have run across. In actual fact the capacitor Q is not infinite and not even that high so resistive losses associated with the capacitor can be expected to lower the measured Q. One needs to keep this in mind and know that this measurement is on the resonator, not on the coil itself. More on this topic in a later section. The APC capacitance range is limited but sufficient for my intended test frequencies of 1.0 to 1.1 MHz, this is a test, not a radio.

Keying, or pulsing the circuit is somewhat tricky. In order to obtain a clean wave trace the pulse contact needs to be fast and sharp. Contact "bounce" and arcing, especially at radio frequency are difficult to avoid and pulsing the circuit several times may be needed prior to finding a clean trace. I initially used a transmitter key but found the contactor at RF to be messy with visible arcing. The best method I have found for pulsing the circuit is to connect a pointed probe to one side of the circuit and tap it sharply but lightly against the hot side of the battery. After a few tries one gets a "feel" for the contact. Pulse the circuit and get a wave, things are this simple.

Procedure:

Hang the coil under test on a stand where it is clear of other components or metal objects. Connect the coil as per the schematic. Connect the probe and scope to the coupling coil leads and set the scope for a good scale, (Figure 3). I generally use 5uS / division horizontal scale which gives me 40 to 50 oscillations at about 1 MHz. For higher Q coils I may go to

---

A Vintage Steinite

http://www.lessmiths.com/~kjsmith/crystal/steinite.shtml

Steinite standard from the mid-1920's. This set has lost its decal but not its classic look. It has a good condition enclosed crystal detector which works very well, binding and phone posts and labels in the typical Steinite lettering. Six taps on the interior spider-web coil and a mica and brass plate "book-style" condenser complete the tuning of this unit. Inside all the components are original with original soldering. It is unfortunate that the condenser is not working as expected. Peering in between the coil tap wires I can see what appears to be 3 or 5 brass plates separated by mica dielectrics. The tuning dial on top turns a shaft with a small cam that alternately compresses the plates together and lets them spring apart. I can see only the outside plate returning to its open position when the cam is open, the others remain compressed so the variation in capacitance is minimal. Well, I can enjoy one station. In fact the tuning is very broad so when "running" the radio has two or more stations competing for my attention. Well, its lovely just the same.

The following photos show the set and its interior.

10uS / division for a longer recording. Set the scope trigger appropriately to capture the data.

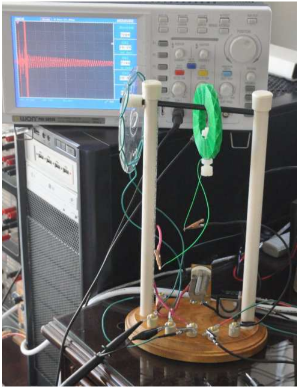

Figure 3. Photo of the test setup.

Tap, or "key" the battery to pulse the tank and review the waveform on the screen. Setting the scope on its measurement mode allows reading the wave frequency with each pulse. Adjust the capacitance after each pulse to bring the tank to the desired frequency, near 1 MHz in my case. Once at the desired frequency you are collecting data. When you get a good clean oscillation free of contact hash and noise, send the data to the computer for analysis. For my scope set at 5uS/div it returns

5000 data points to the computer. At 1 MHz that is about 100 data points per cycle, more than enough sampling to avoid any question of aliasing.

The decrement measurement is based on measuring two peaks. The data is in digital format so one should eliminate "estimation error" by actually reading the value off the spreadsheet, noting both time (in uS) and amplitude. Before taking readings, care must be taken to center the oscillations around zero. Amplitude measurements assume that the waveform is symmetrical around zero yet the scope may or, more likely, may not be set perfectly for this. In the spreadsheet the solution is simply to take the average of all the readings and subtract that from each.

Type the values for “to” (initial peak time), “tn” (final peak time), Ao (initial peak amplitude) and An (final peak amplifude), and the number of periods analyzed “n”.

T = tn - to in uS (8)
f = 1/T in MHz (9)

d = 1/n ln(Ao/An) (10)
Q = pi / d (7)

Next, disconnect the battery and coil and measure carefully the capacitance, C, that was used in the test. With this you can calculate the actual coil inductance from the frequency and capacitance. Why buy an expensive L meter?

L (uH) = 1E6 * ((1/ 2pif)^2) / C (pF) (12)

Example One:

CHAPTER IV

Radios I have acquired (read: paid waay too much for, with no regrets)

An example to show off a bit. This test was made with a close-wound tapped solenoid on a 3" cardboard form. The graphic in Figure 4 shows some initial contact hash followed by a lovely damped oscillation. Measurements on this coil showed a test frequency of 1.055 MHz, L of 215 uH and C = 106 pF. Q was calculated at 167. A decent coil but not great.

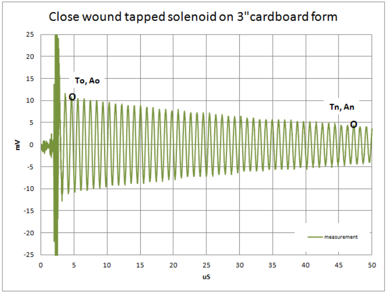

Figure 4. Oscillogram of a coil under test.

n = 45
T = (tn - to) / n = (47.24 - 4.59)/45 = 0.948 uS
f = 1/T = 1/0.948 = 1.055 MHz

Ao = = 11.2 mV
An = = 4.80

d = 1/n ln(Ao/An) = 1/45exp(11.2/4.80) = 0.01883
Q = pi / d = 3.142 / 0.01883 = 167

To test the validity of this technique we need a bit more theory, just to be certain this is working. With the data in spreadsheet form, it becomes possible to model the coil from a theoretical standpoint. The model can then be plotted with and compared to the acquired data to test for correlation. Various web pages, especially university-sponsored lab exercises provide a good source for practical techniques and their theoretical underpinnings. Some good sources I utilized include:
"Damped oscillations in RLC circuits" by Barbara Dziurdzia at AGH University of Science and Technology in Cracow, "RLC Circuits" a lab note from Rice University, and "Oscillations and Resonances in LRC Circuits" from Durham University in the UK. And very many others.

From the above we find the general equation for damped electrical oscillations as:

qc = qo * e^(-a*t) * cos(2pif * t + phi) (13)

Where
qc is the charge in volts
qo is the initial charge
a is the damping factor (=Rs/2L (series) and = 1/2RpC (parallel)
R is the coil resistance in ohms
L is the coil inductance in uH
t is the time in uS
f is the frequency in MHz
and phi is the phase angle

The phase angle and qo charge are initial conditions and adjustments should be minor to fit the calculated oscillation to actual. L, the inductance comes from the calculation we made with equation (12), and frequency comes from equation (9).

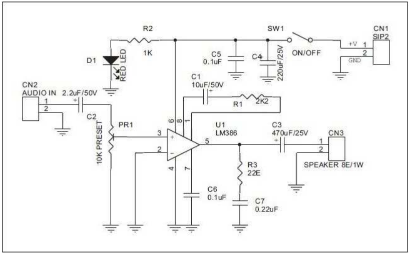
R2
1K
SW1
ON/OFF
CN1
SIP2
+V 1
2
GND
D1
RED LED
C5
0.1uF
C4
220uF/25V
C1
10uF/50V
CN2
AUDIO IN
1
2
2.2uF/50V
C2
PR1
10K PRESET
R1
2K2
6
8
1
U1
LM386
3
2
5
4
7
C3
470uF/25V
CN3
1
2
SPEAKER 8E/1W
R3
22E
C6
0.1uF
C7
0.22uF

It even looks nice from behind, vintage binding posts, naturally..

Schematic of the kit, (AnyKits.com) A005. The kit is configured with a maximum gain between 45 and 50, too high for headphones but too low really for listening with the speaker easily. You will note the location of R1 (in series with a 10uF capacitor) between pins 1 and 8 on the LM386. If you short the resistor leaving only the capacitor, then the gain will increase to the maximum of 200. If you do such a project, I recommend you download the LM386 datasheet from National Instruments.

This leaves the resistance of the coil (R) as the primary variable. Model the above equation in the spreadsheet by adjusting R until the model results match the measured oscillation from the coil. From the above example we see Figure 5 with the modeled damped wave overlaid on the test oscillation.

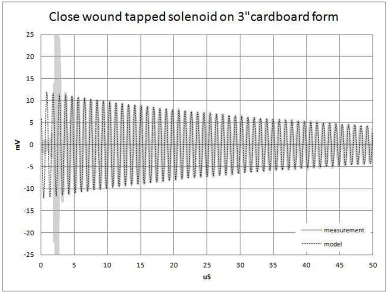

Figure 5. Oscillogram with modeled damped wave response.

Here the background data from the test is in grey and overlaying it is the dashed model in black. The fit is nearly perfect using a resistance of 0.22 Mohms. Relating resistivity to Q we have the following equation (5):

Q = Rp / 2pifL

Substituting f = 1.055 MHz, C = 106 pF, and R = 0.237 Mohms the result is Q = 167, exactly the same as with the decrement method. An exact match will not always be the case, but they ought to be fairly close. The match between theory and the decrement method gives confidence that this analytical technique is as robust as it is simple.

qc = qo * e^(-t/2RpC) * cos(2pif * t + phi)
qc = 12.2 exp(-t/2*Rp*106)*cos(6.283*1.055*t + 1)

Where t = 0 to 50 uS
R = 0.237 Mohm in this case for best match.

Q = Rp / 2pifL = 237000 / 6.283 * 1.055 * 215 = 167

Example Two:

To be sure the technique works I have also assessed the quality of a 5inch diameter basket weave coil wound from 660/46 litz wire. Such coils are rightly considered "performance coils" for their low skin-effect and low resistivity.

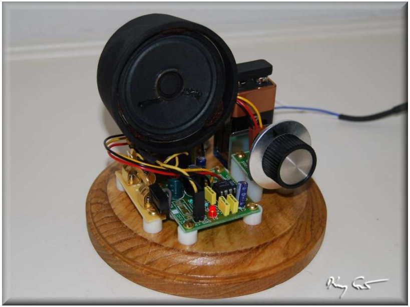

I was initially worried that the audio input impedance may cause me trouble. I read so much about this, impedance matching, transformers, etc etc, almost a fetish really. I have subsequently learned that the LM386 input impedance is about 800 Megohms, the 10k pot being the limiting factor. So, with considerable relief, I concluded this ought not present any problems with the crystal radio circuits. I find the amplifier works very nicely with my radios. The sound is as good as the Radio Shack unit.

The amplifier certainly adds to the vintage "look" of my setup. Well, you be the judge, I hope I come close, please check out the photos...

From the front, wooden base, full display of components, big adequate volume control..

crystal radios one builds with passion, creating something more than a mere radio, but a work of art, a piece of furniture. Still, the technology itself is primitive dating back to the dawn of radio. When I look at the mini amplifier, it looks like a toy yet inside there is the most wonderful technology, solid state integrated circuitry. We take this advanced technology and slap it inside a cheap plastic box, an appliance. Naturally, I don't complain about the price, but for my project I wish for something above.

The solution struck me as obvious, give the same care and respect for the amplifier as I do the radio. I also sought to maintain a "vintage" visual aspect for the unit, breadboard layout, antique binding posts and a large adequate volume control knob, (the pitiful volume control wheel on my RS mini has already come apart once and threatens to break at any time). I wanted a nice lacquered base and sturdy speaker enclosure. The enclosure was the main challenge. I finally ended up with a heavy plactic goblet from a local thrift shop cut down to a tapered cylinder into which the speaker can be placed tightly. The rest of the set is modular with plug-in connections for the audio in, power, speaker, and volume control potentiometer. A question, do I add a brass knife switch for the on-off? I worry I am out of room on the base. Still pondering this.

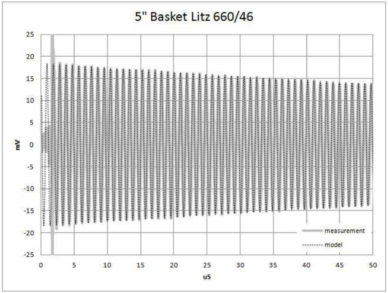

Figure 6. Second oscillogram with modeled damped wave response.

The coil tested values as follows:

n = 45
T = (tn - to) / n = (47.82 - 4.70)/45 = 0.958 uS
f = 1/T = 1/0.958 = 1.044 MHz

Ao = = 18.05 mV
An = = 13.65

d = 1/n ln(Ao/An) = 1/45exp(18.05/13.65) = 0.00621
Q = pi / d = 3.142 / 0.00621 = 506

And:

qc = qo * e^(-t/2RpC) * cos(2pif * t + phi)
C = 123 pF

qc = 18.55 exp(-t / 2*Rp*123)*cos(6.283*1.044*t + 1)

Where t = 0 to 50 uS
R = 0.627 Mohm in this case for best match.

Q = Rp / 2pifL = 627000 / 6.283 * 1.044 * 189 = 506

The results above for a big litz coil clearly demonstrate its superiority to a more common tapped solenoid. Published data though has let me to expect a big litz coil Q to be in the 1000+ range. I have run this test backwards and forwards, checking all possible causes for this unexpected low Q with no difference in the final value. In fact I am quite impressed with the repeatability of the technique and the correspondence between the results from the decrement and those from modeling the wave form.

To be certain I even tested the coil using the old -3dB method. As stated previously, I am uncomfortable using a signal generator in addition to the scope, and I have already found trouble connecting the 1:100 scope probe directly to the tank, as is necessary in this method. With that said, the results yield a bandwidth of 2.3kHz about a nominal frequency of 1018.8kHz for a tank Q = 450. I suspect the method (with the problems noted) lowballs somewhat the final result. In any case, this confirms the test to have a moderate Q and nothing like the expected values one reads online.

Notes on Capacitor Q and its impact on the measurement value.

This is where we begin to think more seriously about the capacitor used to resonate the tank. The assumption of very high Q needs to be questioned. I stated in the setup discussion

A breadboard Audio Amplifier

http://www.lessmiths.com/~kjsmith/crystal/amp.shtml

A little project stemming from my persistant difficulty in hearing. (Losing one's hearing is never fun to admit, but combined with crystal radio, well, its trouble!) I hope I do not offend any "purists" with this project. Still, I understand that many crystal radio owners keep a small amplifier too, for "those times" when, well, you know.. So, this amplifier project, its for, well, those times.. (I suspect Radio Shack would have discontinued their nifty Mini Amplifier years ago had it not been for the crystal radio community.)

My idea was to build my own unit rather than use a store-bought amplifier. Additionally, I wished to put my own stamp on it, make it unique. I started with a kit for the basic layout and electronics. I wanted to base things on the NI LM386 chip which is a good part and hard to break. I found several distributers of what appears to me as the exact same kit and so chose the cheapest $6 dollar kit. NI sells the LM386 for 34 cents a pop and would even provide a few "samples" but then I would be stuck laying things out and wireing.. Been there, done that. A printed circuit board is just SO much easier to use.

One of my inspirations for this project also stems from the nature of modern electronics in general, and the Radio Shack mini amplifier in partictular. Using my RS mini, I am always struck by the contrast between electronics new and old. With

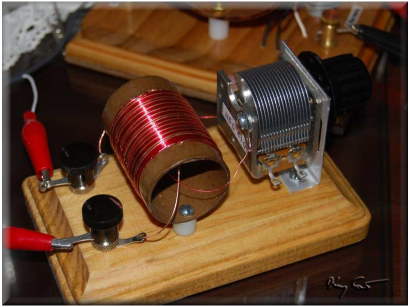
Wave Trap

my opinion that the capacitor Q ought to be high enough as to not influence the coil Q measurement in any significant manner. I said that for three reasons: 1) Just about every published article starts with such an assumption and 2) I had not done much research on the subject of capacitor Q and so had no particular expectations other than assumption 1 and finally 3) I had no measurements or specific knowledge of the Q for the capacitor used in my setup.

Reason 1 is a punt and should be rejected, reason 2 has been remedied and I give some analysis on the interesting subject of capacitor ESR (Equivelant Series Resistance) on a separate page so that readers may set their own expectations. For the third reason I have received the kind assistance of Steve Ratzlaff, AA7U who offered to make actual Q measurements on the coils and APC capacitors with his HP4342A Q-meter, results follow:

Litz basket coil,
550kc 1127Q;
1000kc 1140Q,
1700kc 674Q (194.2 uH on aade.com LC meter)

Solenoid coil,
550kc 205Q,
1000kc 240Q,
1700kc 222Q (218.2 uH on aade.com LC meter)

Using Steve's Litz wound ferrite rod coil (1000kc 1047Q) as the reference for external caps:

100 pF silver mica (99.2 pF meter):859Q at 1068kc (Q meter alone and coil 1025Q)

APC cap with red/black leads (149 pF meter): 937Q at 904kc (Q meter alone and coil 1060Q)
APC cap with my Litz leads (149 pF meter): 995Q at 904kc (Q meter alone and coil 1060Q)

Steve's measurements above provide important datapoints with which to calibrate my technique. Note first of all that the Litz coil is not a modest 506Q but a robust 1140Q at 1MHz. The reason for my low measurement was my assumption of a high-Q capacitor. Steve's measurements of two of my APC's are 995 and 937, hardly very high. There is an obvious need for a correction. Wes Hayward in his excellent notes on Q measurements provides the key in the form of the following formula:

Ql = (Qc * Qres) / (Qc - Qres) (14)

where Ql is the desired coil Q, Qres is the measured Q of the tank/resonator, and Qc is the tank capacitor Q.

Substituting the values Qres = 532 and Qc = 995 the equation simplifies thus:
Ql = (937 * 506) / (937 - 506)
Ql = 1100.

The result is distressingly close to Steve's measured value of 1140Q. This shows the possible level of accuracy in the analytical technique used, both reassuringly high and very "Litz-like".

Taking the first solenoid and substituting the values Qres = 167 and Qc = 937 the equation simplifies thus:
Ql = (937 * 167) / (937 - 167)
Ql = 203.

Wavetrap, a useful gadget

http://www.lessmiths.com/~kjsmith/crystal/wt.shtml

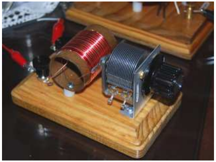

This little gadget is most handy for crunching those local behemoths that always seem to be directing their strongest signals right in your direction. My wave trap is a bit of a combination of 1) Osterhoudt's QRM coil, 2) a design I found online on Owen Pool's excellent web site, and 3) a bit of my own seeing as how I had some, but not all of the ingredients needed from both traps. That is, I used what I found in the junk box.

The trap is formed on a stiff cardboard paper core with 1 1/4 inch outside diameter. On this I drilled holes for winding the tank coil, 110 turns of #30 awg enameled wire. Both Poole and Osterhoudt recommend #32, but 30 is what I had. Over this I would a coupling coil, 15 turns space wound of #22 awg enameled wire. Tuning of the tank is completed with a 410 pF var cap.

Tuning is sharp and deep, it takes a bit of getting used to, but then, schezzam!!!.. problem station gone! I luv it.

This is slightly low to Steve's measured 240Q. I am of the opinion that the original tank measurement may have been a bit on the low side. Still, again the results are within the ballpark and the technique works if you have a good measure on the Q of the capacitor used in your tank. Alternatively, starting with a known coil Q, the technique will work equally well to determine the Q of an unknown capacitor.

Conclusion:

If you have access to, or own a digital oscilloscope, or have been hankering to purchase one, then the time may be now. The analytical method presented above is simple in practice and robust in theory. I would also say that, by working with your wave forms, you learn to recognize your data and soon get a good "feel" for what is good and what is not. The technique also produces, should you wish, an auditable report of the wave and analysis. It is hard to argue against the results with the waveform sitting there staring right at you!

Measuring your coil Q should not be difficult.

A copy of the evaluation spreadsheet in xls format can be found here CoilQbyDecrement_kjs.xls

.

Bibliography:

Jacques Audet, VE2AZX., Q Factor Measurements on L-C Circuits., QEX - January/February 2012, pp 7 - 11. http://www.arrl.org/files/file/QEX_Next_Issue/Jan-Feb_2012/QEX_1_12_Audet.pdf

EE. Bucher, 1919, "Wireless Experimenter’s Manual", Wireless Press, New York.

Wes Hayward, W7ZOI, Experiments with Coils and Q-Measurement, October, 2007 (Updates 01Dec07, 08Dec08. http://web.archive.org/web/20101226143919/http://w7zoi.net/coilq.pdf

Steve Ratzlaff, AA7U kindly made detail measurements on the Q of the capacitors and coils used in this report.

Dick Kleijer , 2013, Measuring the Q of LC Circuits, http://www.crystal-radio.eu/enqmeting.htm

Barbara Dziurdzia, 2013, Damped oscillations in RLC circuits, AGH University of Science and Technology in Cracow, http://home.agh.edu.pl/~lyson/downloads/Manual_8.pdf

Rice University EE., RLC Circuits, http://www.owlnet.rice.edu/~phys102/Lab/RLC_circuits.pdf

Durham University EE., Oscillations and Resonances in LRC Circuits,
http://level1.physics.dur.ac.uk/projects/script/lcr.pdf

of the inductance versus rotation angle and a photo of the Variometer mounted in its stand, ready for use.

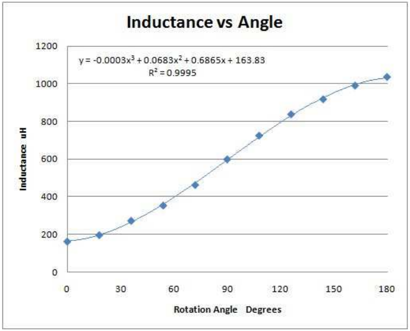

Perhaps the less-than-optimal condition was a factor, but I was in luck to win the coil with an only bid. I had no illusions that this was a long-shot deal, the chances of the variometer being functional were poor. My expectations were not deceived when it arrived sometime later. Nope, it didnt work. The question was, could it be restored, and just how did the thing work, how were the connections made anyway? Inspection quickly led to three obvious broken connections and an understanding as to how the connections were intended. A few dabs of solder later and under the meter I had a fully-functionning variometer! All that was left was to give the vintage coil a lovely stand to show it off.

Measurements of the Inductance show a large variation, from 163 uH to 1036 uH. This should give good duty to electrically "lengthen" my woefully short antenna. Shown below is a plot

---

Crystal Radio Capacitors and ESR

Kevin Smith

In my ongoing effort to understand crystal radios and the theory that makes them work I have recently turned to setting my expectations concerning capacitance. The tuning capacitor is, next the coil itself, the principal component of the set. Few builders fabricate their own tuning caps so I suppose it is possible to take them a bit for granted. I have been reading on the radio forums some threads concerning radio Q and the subject of the variable cap Q seems to be getting more attention. So, what is the deal?

I set about my study by gathering what I could find on the web and at the close of this note I include a short reference list of useful pages from which I extracted data the ideas. I wish to extend special thanks to Ben Tongue who, as usual, has provided the essential measurements and background. The plot below of capacitor Q versus Equivalent Series Resistance (ESR) summarizes my findings. It is a bit busy as there are a number of interesting and important ideas represented in the relationships presented.

Any discussion of component quality needs first to acknowledge that the fight to increase the component Q is in reality a fight to eliminate (not possible) or reduce as much as feasible the resistive losses associated with that component. It helps then to look not just at Q, but at how Q relates to the component ESR. There are two fundamental equations that describe this relation: 1) Q = cotan theta = 1/D. D is the dissipation factor and is an attribute of the dielectric material used in the construction of the capacitor. Equation 2) ESR = D / 2pi f C. This equation tells us that capacitor losses expected due to resistance are both angular frequency and capacitance dependant. It is important to remember that not all Q's are created equal. What we want is to keep the ESR to a minimum.

On the plot there are a number of data trends presented. Note first that both plot axes are logarithmic, it makes a difference on how one thinks of the data. Additionally, note that the plot is specific (mostly) for data at 1.0 MHz. I make the assumption that a cap that is superior at that frequency is likely to be superior across the BCB.

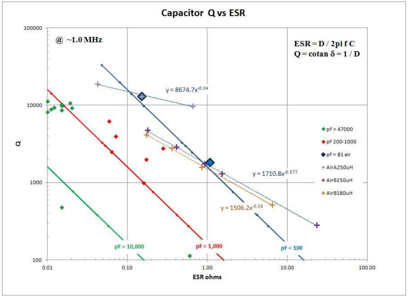

The first things to discuss are the three slanting lines. These are computed from a table of Dissipation Factors at 1.0MHz by Dr. Johnson. I have supplied the capacitances and calculated the resulting Q and ESR values. When one goes looking at capacitor data sheets or summary tables one is likely to find data computed at 100 KHz and 1.0 MHz (mercifully) but that

Vintage Variometer / Loading Coil

http://www.lessmiths.com/~kjsmith/crystal/lcoil.shtml

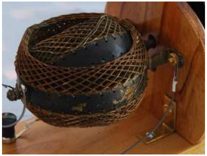

My receiving station is pretty typical I expect, decent ground, so so antenna and I have been hankering for some time to add a loading coil to my station to better match that antenna with my radios. Lack of urgency or just plain laziness perhaps have kept me from working on this project. Part of my hesitancy has been the idea of utilizing a variometer as a loading coil, and these are not easy to find.

Variometers are variable induction coils that function by being connected in series and changing the mutual inductance between them. Most variometers though have a relatively small range of inductances making them of limited suitability for, well, just about anything in crystal radio. They find a home mostly these days as a control for regeneration in such circuits. Alternatively they make a nice variable inductance load on your antenna, the purpose I was seeking. Only, I really wished to find/make something with quite a lot of wire to get a good range of inductances going. It was by some degree of luck I believe that I noticed some time back on ebay a rather beat-up looking variometer on offer. The variometer, shown below, was a basketball weave design and importantly, was wound in at least three layers suggesting a lot of inductance.

the likely capacitor data will be specific to caps in the uF range, or 10^6 higher that what we need in a crystal radio tank. Be careful not to get enamored with the high Q values occasionally quoted, check the capacitance. From the plot it should be apparent that not all Q's are created equal. Larger-value caps have lower ESR and so the Q's are often agreeably high. I have plotted a few specific cap measurements from various sources.

Two measurements should stand out; these are indicated by the larger blue diamonds. These are measurements by Bill Hebbert and presented by Ben Tongue in section B of his Article 24. I reproduce his Figure three below, with grateful apologies for the liberties I have taken,

By replotting Ben's figure I can calculate the relation between Q and frequency (table below) and so compute the estimated Q at any frequency. On my main plot I give Ben's Capacitor A and B for 1.0 MHz given the 250 uH inductor. Ben states that the plot assumes 20pF stray added. Since the capacitance needed to resonate at 1 MHz and 250 uH is 101 pF, I make the assumption his caps were set to 81 pF. This accounts for why the diamonds are set slightly off the 100 pF line of the plot. Given his Q measurements it is easy to compute the ESR for each cap. This is the world of the crystal radio builder. I include the ESR at the BCB endpoints for each cap as well to give the full picture.

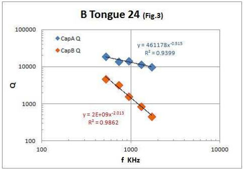

pF		f	Q	
tank	cap	MHz	Cap A	Cap B
375	355	520	19000	4750
190	170	730	14000	3230
114	94	943	14400	1610
101	81	1000	*13148*	*1828*
60	40	1300	11600	870
35	15	1710	9800	460

for I = 250 uH and 20pF stray in tank
(Q data in italics is computed)

The quest for high Q is indeed a quest for low ESR as long as you work in the capacitance range needed. One interesting piece of advice I have occasionally seen is that, to increase your tank Q, it is better to use lower-value coils and higher-value tuning caps. As much of the losses in the tank derive from losses in the coil this makes excellent sense. I wish to ask the question what about those higher-value caps? The plot immediately suggests that moving to higher capacitance drives down the ESR, this is a good thing. It may or may not improve the cap Q but I am not sure it matters as long as resistive losses are minimized. To better set my expectations I have again turned to Ben's Cap B data.

The small plot below takes data presented in his Figure 5. Data presented include Q at various capacitances and frequencies. I cross-plotted his data for pF and Q for the appropriate frequencies using both 250 uH and 180 uH inductances. Reading the actual Q off the tiny plot was an estimate at best, but it is interesting. This allows one to estimate his cap Q at any capacitance setting. The table below

---

CHAPTER III

Accessories on Breadboards

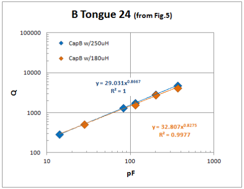

provides the estimated Q, and calculated D, and ESR for the capacitances from Fig 5 (with 20pF stray added). The values of Q do not improve by moving to higher capacitances but the ESR of the cap is slightly (at low tuning freq) and well improved (at the high tuning freq). This is the goal. It is worthy to note that the tuning frequencies do not line up with the original data from Ben's Fig 3 as seen on my main plot. While Ben specifically noted that his figure 3 included 20pF strays, in his figure 5 he speaks in absolute capacitances. I did my best to correct for this but there remains some ambiguity in the presentation. The point remains though; moving to higher capacitances in the tank will lower the tank losses due to ESR.

250 inductor (w/20pF stray)

tank pF	KHz	D	Q	ESR ohm
385	513	0.0002083	4800	0.177
220	679	0.0003448	2900	0.404
134	870	0.0005714	1750	0.917
101	1000	0.0007638	1309	1.502
34	1726	0.0034973	286	23.046

180 inductor (w/20pF stray)

tank pF	KHz	D	Q	ESR ohm
385	605	0.0002381	4200	0.172
220	800	0.0003571	2800	0.355

134	1025	0.0006250	1600	0.852
48	1712	0.0019341	517	6.425

On the main plot then I have included for Ben's Cap B a second Q vs. ESR curve to illustrate the estimated benefit of using a higher tuning capacitance range. The new curve slides mostly to lower ESR (especially at the high end of the tuning range) and also to slightly lower Q. The improvements made appear modest from the modeled data and, for the case of a high-Q capacitor the possible improvements may be slight indeed. It is comforting to know that at least there should be no negative impact. Lower ESR certainly improves tank overall efficiency. This is a win-win situation where the set designer wins twice!

In seeking the "Holy Grail" tank capacitor I would add to the list of valuable features. We have ceramic stator supports, silver plating on the vanes, and good wiper contacts as features desired. I would add to this an extended tuning range if the capacitor is truly to be considered "Holy". With a standard 230 uH coil the tuning for the BCB ranges from 40 to 380 pF or so. With a 180 uH coil the capacitance needed ranges from 50 to 490 pF, (stray capacitance not accounted for).

Two questions immediately come to mind with this analysis:

1) To how low a coil inductance can you go?
2) If you find a few of these super-caps can you send me one?

kjs 6/2013 (in France)

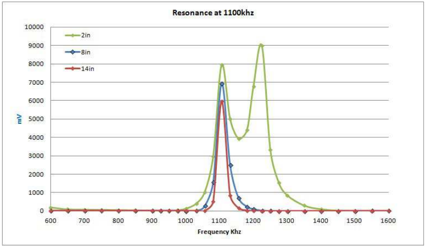

A look at the resonance curves above (including an over-coupled 2-inch spacing case) is very interesting. I include this as a nice demonstration of resonance at various conditions of coupling. The 2-inch coupling case clearly demonstrates the problem of over-coupling in a radio. Two resonance peaks show up and tuning brings in different stations at the same time. Total power is still very high, just not useful. The 8-inch case is pretty close to critical with good power and the two-peak problem eliminated. Bandwith is still pretty high at 22khz (the curves are made with the 4 parallel HP's). Finally when separated to 14 inches the power loss is not too bad and bandwidth has narrowed to 10 giving a loaded Q of 107 as mentionned above.

It ain't Litz, but PD Close!
kjs 11/2012

tuggle tuner on this set is in the 50-80 ohm range giving an excellent match and maximum power transfer to the tank.

At the back end things get more complicated, a diode is present! For my testing I have included a step to determine, under test conditions of course, the optimum load resistance for maximum power transfer to the phones. Maximum power output means that the load resistance matches the output resistance of the tank/diode combination. The principal factors which influence this include the frequency and tank coil Q (which don't change from test to test), the diode Rd and characteristic, and the signal voltage being rectified.

Signal level and diode characteristic are inter-related and together determine the resistance of rectification. At very low signal voltages rectification takes place on the "square-law" portion of the diode with large swings in resistance with signal sine wave. Larger signal levels rectify higher on the characteristic curve (peak-detection region) with consequant lower resistances and less variation. The Rl shown on my table thus is specific for my test. My radio signal is in the 6-8 mV region placing it in the peak-detection portion of HP5082-2835 diode and possibly borderline square-law / peak detection on the FO-215. This explains why the optimum Rl is lower that of the HP but pretty close to the FO-215. It may also help explain why the FO bottomed out on the 14-inch separation, there was just no longer much signal for it to work with.

References:

AVX Basin Capacitor Formulas.
http://www.avx.com/docs/Catalogs/cbasic.pdf

Ben Tongue 1999, Sensitivity and selectivity issues in crystal radio sets including diode problems; measurements of the Q of variable and fixed capacitors, RF loss in slide switches and loss tangent of various dielectrics.
http://www.bentongue.com/xtalset/24Cmnts/24Cmnts.html

Wikipedia, Types of Capacitors, Ohmic Losses, Dissipation Factor and Q.
http://en.wikipedia.org/wiki/Types_of_capacitor#Ohmic_losses.2C_ESR.2C_dissipation_factor.2C_and_quality_factor

Dr. Gary L. Johnson, 2001. Lossy Capacitors
http://www.eece.ksu.edu/~gjohnson/tcchap3.pdf

Pin = Power into the set in microwatts
Pout = Power out the back end (into the phones) also in microwatts
Eff = Set efficiency or Sensitivity and is the ratio on Pout to Pin * 100
BW = -3dB Bandwidth of the set at 1100khz (f res)
QL = Loaded Q of the set = f res / BW
Rx = Input resistance of the set in ohms
Rl = Load resistance on the tank in kohms
Rd = Junction resistance of the diode in the set

The power numbers are fairly straight forward and represent the power calculated across the dummy antenna resistor and load resitor. Input to the set is from a signal generator with a 200mVpp signal output. Losses due to the dummy were accounted for and the input power is that presented to the set itself. Looking at the Efficiency and QL numbers one sees nicely the tradoff between sensitivity and selectivity. The larger the spacing between the sets, the higher the QL, but at a sacrifice of sensitivity. While the numbers look low, they are still good for crystal sets and the 107 QL is a superlative number. By comparison, I have looked at a High-Performance built by Dave Schmarder (see my Radio Test page) with big Litz and super capicitors and this set max's out around QL = 120-125. I would opine that any QL above 100 is "Performance Quality".

Note for a bit the Rx numbers. Impedance matching for maximum power transfer starts at the antenna and antenna tuner. The dummy antenna used for the measurements mimics a fairly typical longwire with an output impedance between 25 and 50 ohms. Of the many radios I have measured, most have set input resistances between 100 and 250 or so ohms. The

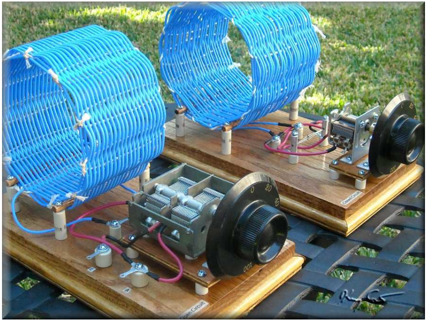

Performance summary:

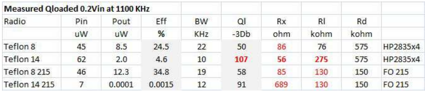

Measured Qloaded 0.2Vin at 1100 KHz									
Radio	Pin	Pout	Eff	BW	Ql	Rx	Rl	Rd	
	uW	uW	%	KHz	-3Db	ohm	kohm	kohm	
Teflon 8	45	8.5	24.5	22	50	86	76	575	HP2835x4
Teflon 14	62	2.0	4.6	10	**107**	**56**	**275**	575	HP2835x4
Teflon 8 215	46	12.3	34.8	19	58	85	130	150	FO 215
Teflon 14 215	7	0.0001	0.0015	12	91	689	130	150	FO 215

As mentionned at the start of the article, this set was designed as a high-performance rig. Lets take a look at the numbers and assess how well I did. On the spreadsheet above I have indicated the essential measures of the set in four configurations, two spacing tests (8 inches and 14 inches) with 4 parallel HP 5082-2835 diodes and two spacing tests with the holy grail FO-215 diode. First an explanation of the units:

Diode Modeling Update

Kevin Smith

http://www.lessmiths.com/~kjsmith/crystal/dmodel.shtml

The following explains my first attempt to compare a measured diode characteristics with calculated plots based on the Shockley Equation.

I have been taking serious look into diode modeling and measurement recently in an attempt to better understand these solid-state successors to the humble crystal and cat's whisker detector. There is much good information on diodes to be found in many excellent web sites of course, but there is nothing quite like actually making the measurements and working with them to get a good feel. This page reports some of my protocols and test setups which I have found to be useful.

Most probably the single best web article for the measurement of diode Is and n parameters is given by Ben Tongue in his Article 16. In this article he describes an interesting circuit for the measurement as well as protocols for making them. For my purposes I did not wish to make this a construction project and felt I might get along with good quality meters and my already-built "Diode Test Jig". Essentially Ben Tongue's method consists of making two precision measurements of Voltage and Current through a diode at small signal levels, essentially about 3 and 6 times Is (sufficiently low that the voltage drop across the series resistance Ro can be ignored). The measurements are then substituted into the Shockley equation Is*{exp[(qe/(n*k*T))*(V-I*Rs)]-1} and solved simultaneously for the two data pairs. Mike Tuggle has provided a nice excel spreadsheet to do the math. This spreadsheet, Cal_n_Is.xls forms the basis of my technique and

I am thus indebted to both Ben Tongue and Mike Tuggle for making this project feasible.

In addition to simply calculating the main parameters Is and n, I also wished to measure enough points to plot a characteristic curve, and to compare the measured curve with one calculated directly from the Shockley equation. My first spreadsheet then combined Tongue/Tuggle calculation protocol and a graphical view of the match between measured and theoretical. I did make a few methodology modifications which I thought/hoped would allow added accuracy:

1) Using the full Shockly equation without simplifying for an assumed 25dC room temperature. While the 25dC assumption is generally good and probably within the measurement error (or perhaps not), including the actual measured temperature eliminates doubts of inaccuracies due to this parameter. With the power of PC's and spreadsheets there is no reason simplify the equation.

2) Used Vd = 0.04 and 0.05 V (adequately close to Ben Tongue's recommended 0.039 and 0.055 V) whenever possible. Note for some diodes with significantly different forward voltage drops, Si diodes in partictular I had to use higher values for Vd.

3) Reporting: I decided that, as there is no unique solution to n and Is, each is dependant on the valuse Vd and Id, I feel that reporting both Vd and Id is necessary for repeatability. Naturally I also include the ambient temperature in the report as well, and

4) I included a calculation for Ro (or Rg if you are using Tongue's reference) as that is the goal of the exercise.

from 4 to 7 or more khz off peak tuning when I remove my hand from the dial. The set is fairly modular and readily modified so I am keeping my eyes open for better quality caps. I am afraid I must learn to live with hand capacitance. Oh well, we see about the next set!

Portrait of set, ummmm..

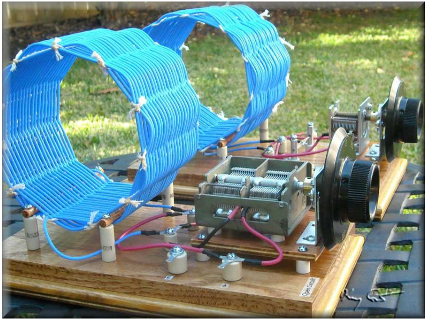

With high-efficiency coils, very loose coupling is permitted with acceptable power levels, thus allowing very narrow bandwidths to be achieved. This can only be accomplished with separate base boards. To further minimize losses I have used short direct lengths of large 16 AWG stranded copper hookup wire and importantly, I have insulated all components and electrical connections from the board with ceramic standoffs. Finally, for the coils I have utilized a very lovely sky-blue teflon-insulated, silver plated wire of 18 AWG, strictly military spec, ($$$). I know, I KNOW.. This isn's a Litz Blitz set and yes, the performance suffers a little because of this. Still, I have good performance measures to tell me I am certainly on the right path. Note also that wire of such size and insulation is quite thick and to get the desired inductance required very large coils of some 4 3/4 inch diameter. The coils dominate the set visually and provide the name.

The tuning capacitors I would allow are most likely the weakeast link in the chain. I have chosen fairly generic though not bad-quality caps. The antenna tuner has a dual-gang cap with sections 465pF and 375pF respectively, well matched for my needs. Note, from my studies of antenna tuning (elsewhere on my pages) I have noted that a tuggle tuner does not necessarily require equal-size sections on the tuning cap. In fact, I would state that tracking is better if the ground section max's out at a lower capacitance that the tank section. The cap has aluminum blades and ceramic insulation and is a decent capacitor. For the tuning tank I chose a standard 410pF single gang unit from the Crystal Set Society. These are aluminum blade caps with phenolic insulation, OK but not superb. Both tank and atu include 3:1 reduction gears to help spread the band and allow more careful tuning. I did not allow for hand capacitance which I have noted is indeed a problem. When I tune the set to resonance with a signal generator, I find I am

A screen shot shown below:

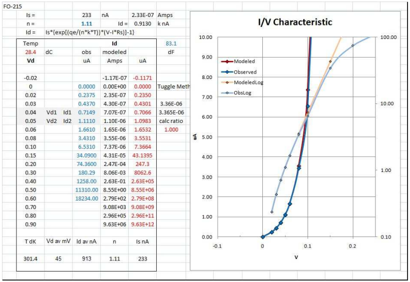

FO-215

Is =		233	nA	2.33E-07	Amps
n =		1.11	Id =	0.9130	k nA
Id =	Is*{exp[(qe/(n*k*T))*(V-I*Rs)]-1}				
Temp			Id		83.1
28.4	dC	obs	modeled		dF
Vd		uA	Amps	uA	
-0.02			-1.17E-07	-0.1171	
0		0.0000	0.00E+00	0.0000	Tuggle Meth
0.02		0.2375	2.35E-07	0.2350	
0.03		0.4370	4.30E-07	0.4301	3.36E-06
0.04	Vd1 Id1	0.7149	7.07E-07	0.7066	3.365E-06
0.05	Vd2 Id2	1.1110	1.10E-06	1.0983	calc ratio
0.06		1.6610	1.65E-06	1.6532	1.000
0.08		3.4310	3.55E-06	3.5531	
0.10		6.5310	7.37E-06	7.3664	
0.15		34.0900	4.31E-05	43.1395	
0.20		74.3600	2.47E-04	247.3	
0.30		180.29	8.06E-03	8062.6	
0.40		1258.00	2.63E-01	2.63E+05	
0.50		11310.00	8.55E+00	8.55E+06	
0.60		18234.00	2.79E+02	2.79E+08	
0.70			9.08E+03	9.08E+09	
0.80			2.96E+05	2.96E+11	
0.90			9.63E+06	9.63E+12	
T dK	Vd av mV	Id av nA	n	Is nA	
301.4	45	913	1.11	233	

To use the sheet one need only adjust the voltage Vd to that shown in the first column and record the current Id. If only Vd1/Vd2 and Id1/Id2 are measured then the next step is to adjust the value for "n" until the match between the two equations is exact. I even provided a simple ratio calculation to easily test for a match.

While the math is good and results excellent, I quickly found a couple deficiencies in my method,

1) measurements require high precision and its nearly impossible to land EXACTLY at the voltage required. I needed only to get quite close, let the meters stabilize for 2 - 5 minutes, and then record BOTH Vd and Id.

2) the work is rather tedious and, for a large number of diodes it pays dividends to measure only the needed Vd and Id and let the plot aside. My second spreadsheet thus dispenses with the plot. It is with my second spreadsheet that all my data tables and results are posted.

3) better comparisons between diodes can be made if one targets specific Currents rather than Voltages, discussion on this below.

A view of my current data reporting spreadsheet showing ALL input parameters as well as determined values of n, Is, and Rd.
Data input fields in blue: (target Id1 close to 1.0 uA and Id2 close to 0.5 uA).
Calculated fields in red and black.
Adjustment field "n" in green: (adjust value of n until the ratio of the two calculations = exactly 1).

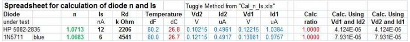

Spreadsheet for calculation of diode n and Is							Tuggle Method from "Cal_n_Is.xls"						
Diode		n	Is	Rd	Temperature		Vd2	Id2	Vd1	Id1	Calc	Calc. Using	Calc. Using
under test			nA	k Ohm	dF	dC	V	uA	V	uA	ratio	Vd1 and Id2	Vd2 and Id1
HP 5082-2835		1.0713	12	2206	80.2	26.8	0.10215	0.4961	0.12215	1.0384	1.0000	4.124E-05	4.124E-05
1N5711	blue	1.0683	6	4541	80.0	26.7	0.12115	0.4917	0.13981	0.9757	1.0000	7.931E-05	7.931E-05

(Note: the engineers in the crowd will have noticed in the above report than I am carrying a degree of precision not justified by the level of accuracy in my measurements. The final results of n, Is, and Rd should be rounded off to not more than three significant figures.)
A copy of my spreadsheet can be downloaded here..

While measuring a good number of diodes, both Germanium, Schottky, Silicon, and even a few LED's, I noted that Tongue's recommended measurements at Vd = 0.039 and 0.055 V resulted in current reports spread out over more than two orders of magnitude. As the determination of Is and n is not unique but a function of Id, I felt uneasy by this method. While most Germanium diodes I have measured have a fairly narrow range of Forward Voltage drops, (Vf), that of Schottky's can range up to a tenth of a volt. This is the factor responsible for the huge range of measured Id values, see the following plot of I/V characteristics for some selected diodes:

My Teflon Warrior

http://www.lessmiths.com/~kjsmith/crystal/teflon.shtml

With this current radio I am making my entry into the "Performance Radio" constructors club. I have tried to follow all the best practices of high-end crystal set design while giving the set my own personal breadboard mark. Two central themes guided my design on this set, simplicity and simplicity. From the circuit diagram below you will note that I have chosen, not surprisingly, a simple double-tuned circuit layout with a Tuggle front end and a Tank/Detector circuit with no circuit decorations whatsoever. No taps. No selectivity enhancement. No Benny. No Decoration. No nada. This is a basic radio wherein I have striven to keep losses, including insertion losses, to an absolute minimum. Should I wish to include selectivity enhancement and/or benny at a later time, they will go on an outboard platform along with transformer and phone plug. The radio itself is simplicity realized.

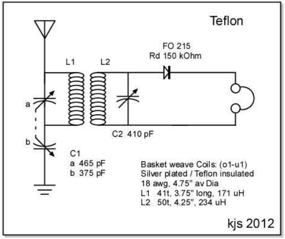

As can be seen in the set photos below, this design follows standard practice of placing the open and closed circuits of the set on separate baseboards. This is important as it allows easy and wide adjustment of the coupling between the coils.

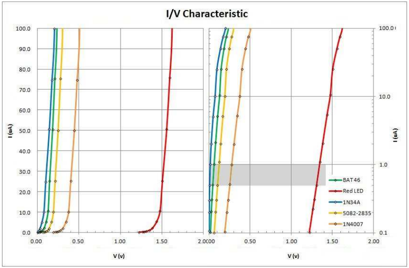

In order to get what I feel is a better comparison between diodes, I have decided to target not some pre-determined voltage, but rather target currents of Id = 0.5 and 1.0 uA. On the above plot on the right (with Log Id vs Vd scale) it will be clear that I am aiming at the same part of the characteristic curve regardless of Vf. Hopefully this will allow good comparison between diodes with rather different Vf, even as far as including Silicon and LED's in my mix. On va voir....

Shockley equation Id = Is*{exp[(qe/(n*k*T))*(V-I*Rs)]-1}
where:
n = ideality factor
Is = Saturation current in Amps
Id = Diode Current in Amps
Vd = Diode Voltage
k = boltzmann = 1.38E-23 J/K
T = temp K = 300
K = dC + 273.15 Kelvin
qe = electron charge = 1.609E-19 cmb

Often simplified to:
Id = Is*(exp(Vd/(0.0256789*n))-1) [at room temperature]

Diode Resistance Rd = VT * n / Is
Where:
T = 300K
VT = k*T/qe = 0.0256789
SO: Rd = k * T / qe * n / Is

I cannot express enough my indebtedness and thanks to Ben Tongue and Ken Khun for their great work and web documentation in matters Diode!

A photo of the current "bench" setup. Keithey 195A pico-ampmeter + Keithley 192 DMM bench meter.

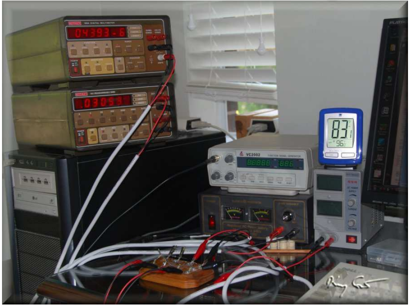

Kevin 09/2011

---

effects I can easily hear when tuning the set. Oh well.. Better be forewarned next time! This does not necessarily mean that double-tuned rigs MUST be built in modular fashion with separate tuning and detection units, but it certainly suggests it.
With the lack of surface real-estate to lengthen the coil separation, I made a small modification on the set as seen in the third photo above. There you see the possibility to rotate the outer coil away from vertical thus further lowering the coupling. I can get a max 40 degree rotation which just gets the coils to critical coupling. No more ghosts on the band! Still, I think my next double-tuned rig will indeed be modular. Experience is such a good teacher!

kjs 07/2011

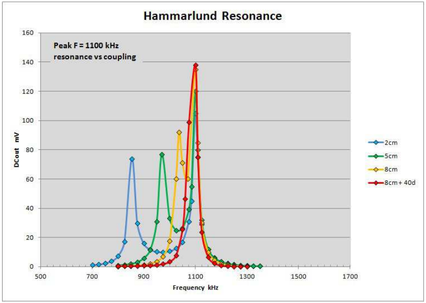

A few words on resonance: When making my measurements I noted a curious effect in the resonance curves. They became increasingly asymmetrical with wider spacing of the coils and actually had a double hump at the maximum spacing of 10 cm. This is the exact opposite of how I expect resonance to work. I set to measuring the Vout across a significant portion of the MW band for each of the coil separations and present a graph below:

It should be readily apparent from the above graph that my set is badly over-coupled even at its widest separation of 10cm. I suspect critical coupling for the set is with a coil separation in the 15 - 20 cm range. Yet one more design consideration to keep in mind when making a new set! The double hump causes two problems for the set, one: it tunes each station twice and two: has the possibility to bring in two stations at the same time if they have the right separation on the band. Both these

---

Spreadsheet for calculation of diode n and Is (@1.0uA)

Diode		n	Is nA	Rd k Ohm
Germanium Diodes				
FO-215	ITT	1.11	175	163
FO-215	ITT	1.10	196	144
1N277	black	1.50	2293	17
1N277	black	1.60	2296	18
1N270	bonafide	1.28	915	36
1N270	blue	1.26	857	38
1N270	blue	1.66	2310	19
D18	russia	1.20	194	160
D18	russia	1.27	193	170
GAZ 51	Tesla	1.14	140	211
GAZ 51	Tesla	1.52	617	64
OA 5	Tesla	1.85	3384	14
OA 5	Tesla	1.48	1996	19
D9E	russia	1.42	2161	17
D9E	russia	1.54	2414	16
1N34A	bonafide	1.82	2153	22
1N34A	bonafide	1.25	1392	23
1N34A	green	1.57	1389	29
1N34A	green	1.30	1185	28
1N34A	green	1.32	1554	22
1N34A 37	orange	1.10	1458	19
1N34A 37	orange	1.36	1370	26

1N34A	red	1.34	833	42
1N34A	red	1.32	993	34
D310	russia	1.01	1215	21
Diode		n	Is nA	Rd k Ohm
GD 402A	russia	1.55	970	41
GD 402A	russia	1.71	1360	33
UK A	russia	1.56	1565	26
UK A	russia	2.00	5766	9
UK B red	russia	1.64	2049	21
UK B red	russia	1.85	3777	13
UK C ora	russia	1.92	3462	14
UK C ora	russia	1.95	3354	15
UK D blk	russia	2.00	4901	10
UK E blk	russia	2.00	5862	9
UK F blk	russia	1.82	3523	13
UK G blue	russia	1.86	3123	15
Average		1.39	1334	54
Schottky Diodes				
HP 5082-2835		1.07	12	2206
1SS98		1.08	12	2302
1SS98		1.06	13	2188
1N5711	blue	1.07	6	4541
1N5711	blue	1.07	6	4816
1N60		1.09	178	158
1N60		1.10	161	176
BAT 46		1.12	141	204

Sensitivity on the other hand, well, it sucks. 2 to 11 percent Vdc/Vpp are very low numbers, many of my single tuned sets range from S = 10 to 50 and my Minstrel Boy getting over 60 percent. What happened? The data show the SEC to be involved here. Note that with capacitance in series the sensitivity is about 10% but when shunted across the tank the selectivity improves, but the sensitivity crashes down to 2-3%. I really expected better for this set. This is my second double-tuned rig, my first was a simple set using solid magnet wire close-wound on 3" cardboard cores and having 6 twisted taps, all Q lowering. This sets' coils have no taps and no cores and ought to have significantly better Q. The primary difference between the two sets is the presence of the benny and SEC on my new set. I am worrying about insertion losses more than offsetting the improvements in the coils.

The severe loss of sensitivity when using the SEC on this set also reminds me of my experience with my Fleming set, a modified MRL#2 using SEC. There I also found loss of sensitivity when the SEC was in the circuit. At the time I concluded that the circuit was inappropriate for testing the SEC concept, now I am not so sure. This set employs a high-performance circuit (if not high-performance parts) and is fully correct for the SEC inclusion. One additional test I intend to make when time allows is to completely remove the SEC and Benny ornamentation from the set and re-test the sensitivity. I suspect that insertion losses are killing me. Dave Schmarder has the most experience with this design feature, I wonder what his thoughts are? Have you tested a set with and without the SEC? What have you found? I am suspicious of this feature now.

capacitor which places the capacitance either 1) in series with the top of the tank and detector, or 2) shunted parallel across the tank. I make a full set of measurements with the SEC cap set fully to each the series and shunt positions. The above paired measurements are made for each of three coil coupling positions. The two coils can be shifted along two rails to bring them from essentially touching to a maximum separation of about 10 cm, or about one coil diameter. This is the max separation obtainable on my breadboard as the Hammarlund caps are quite large and take up much space. To evaluate the impact of coupling, I measured with the coils separated 1cm, 5cm and the max 10cm.

OK, enough with the preliminaries let's look at the data!

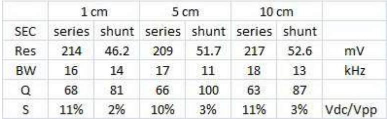

	1 cm		5 cm		10 cm		
SEC	series	shunt	series	shunt	series	shunt	
Res	214	46.2	209	51.7	217	52.6	mV
BW	16	14	17	11	18	13	kHz
Q	68	81	66	100	63	87	
S	11%	2%	10%	3%	11%	3%	Vdc/Vpp

The table at left presents the measurements (Res = resonant peak Vout mV : BW = bandwidth and -3dB : Q = under load : S = sensitivity = Vdc/Vpp) vertically and the different cases horizontally. Bandwidths for the cases all range between 16 - 18 kHz for SEC in series and 11 - 14 kHz for SEC in shunt. These are pretty good numbers for a standard (non-DX Contest) rig and correspond to Ql in the 65 - 85 range. Context for this comes from Ken Khun's excellent engineering pages where he notes for crystal sets that loaded Q normally ranges between 10 (poor) to 50 (typical) to 100 (very good). My set here then is pretty good in this respect and confirms my eardrum assessment above.

BAT 46		1.18	172	177
1N34A ?	schottky	1.11	317	90
1N34A ?	schottky	1.24	482	66
1N5819		1.19	750	41
1N5819		1.14	807	36
Diode		n	Is	Rd
			nA	k Ohm
1SS16	NEC	6.24	14288	11
1SS16		2.27	4101	14
Average		1.13	276	1137
Silicon Diodes				
1N914		1.95	5	9293
1N914		2.02	7	7618
1N4148		2.00	6	8934
1N4148		2.00	6	9323
1N4007		1.51	1	64311
1N4007		1.51	1	77954
6A10		1.57	2	19275
6A10		1.54	2	20379
1N4736A	Zener	1.15	0	1.91E+09
1N4736A	Zener	1.14	0	2.50E+09
KB 130	russian	1.14	0	5.83E+08
KB 130	russian	1.16	0	4.36E+08
UK H	russian	1.94	8	6263
UK H	russian	1.77	4	12548
UK I	russian	1.78	0	260253

KD 401A	russian	1.51	0	232930
KD 401A	russian	1.38	0	1539619
D 220	russian	1.26	0	322405
D 220	russian	1.24	0	537707
D 223A	russian	1.20	0	10262661
Diode		n	Is nA	Rd k Ohm
D 223A	russian	1.20	0	8565213
UK J	russian	1.46	4	9849
UK J	russian	1.40	2	20013
Average		1.52	2	2.37E+08
Light Emitting Diodes				
Red		2.15	0	1.70E+12
Red		2.07	0	5.52E+12
Amber		1.56	0	2.09E+17
Amber		1.62	0	6.46E+16
Yellow		1.59	0	8.03E+17
Yellow		1.38	0	2.87E+20
Green		2.05	0	4.23E+13
Green		2.07	0	4.34E+13
Water Green		3.27	0	4.85E+13
Water Green		1.93	0	5.79E+19
White-Gn		1.44	0	1.05E+20
White-Gn		3.04	0	6.88E+09
Average		2.01	0	3.76E+19

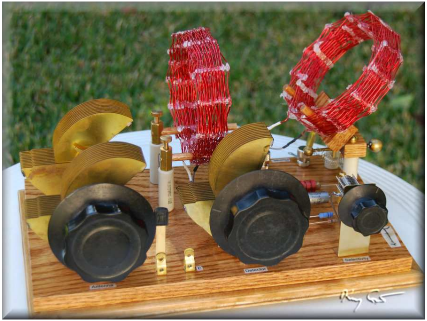

A few words of the performance of my Hammerlund set are in order. When I first started using the set I noted with satisfaction that, by my eardrum meter, the set seemed to indeed tune quite sharply. This is certainly my principal goal in making this double-tuned set. The only fly in the ointment was that the radio seemed quite insensitive, the stations coming in very weakly and needing amplification. The SEC control did have an interesting and dramatic effect when used, but it also hugely lowered an overall low sensitive set. With these thoughts in mind, I have spent some time making some preliminary measurements on the sets' resonance, sensitivity and selectivity.

To characterize the set completely, I measure under a number of different cases. I wish to test the effect of the selectivity enhancement circuit (SEC). The SEC uses a 20pF differential

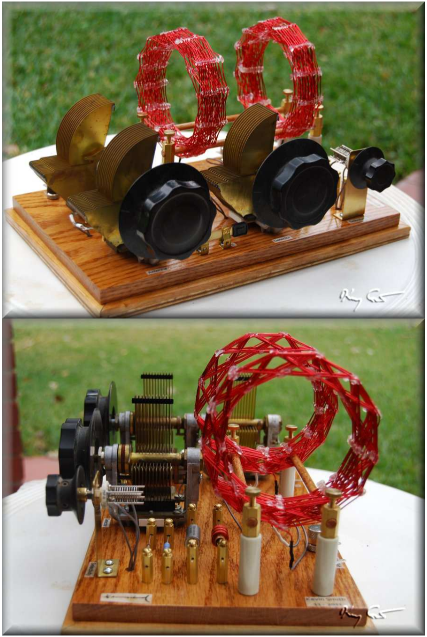

Diode Test

Kevin Smith

http://www.lessmiths.com/~kjsmith/crystal/dtest.shtml

So, what is the best diode out there to use? Do crystals always show poor sensitivity compared to germanium? What about Fleming's marvelous device, the vacuum diode? I start by using a test jig, the design shamelessly modified from Dan Petersen in the Nov 2009 Crystal Set Society Newsletter, (if you aren't a member, go sign up to be.. NOW):

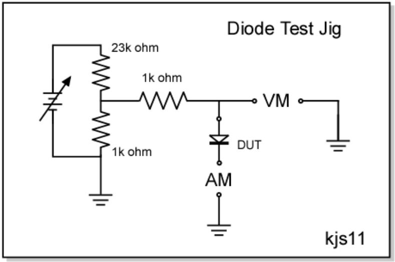

This is a very simple circuit allowing power input, attaching a diode for testing, and sampling both the curent and voltage across the diode under test, (DUT). I made a simple modification by adding a voltage divider, about 20:1 at the power input. The power supply can be connected either at the front across the divider, or after the divider in front of the 100k current limiting resistor for higher voltage readings. I get the input voltage from a modest 18V, 2A regulated DC power supply. This unit has both volt and amp controls with digital readouts, nice to use and dosnt take up much room on my desk. Here is a case where I recommend you look at getting a new unit. I looked long and hard on ebay, but found few bargins and few units with the same features. Still, the supply regulator was never designed to adjust the output to precise small values making it somewhat difficult/tedious to adjust. Adding a small rheostat to the test jig may aid for making fine adjustments on the voltage. Another piece of equipment you will need for this test is an amp meter. I use a nice small digital meter I found new on ebay for $15 shipping included. It reads amps to a lowest range of 200uA, more than enough precision for the task. For that price get two. The DVM is, of course, my trusty Keithley workhorse. Easy protocol as follows:

1) Set up the test with the diode and meters clipped to the jig.

2) Adjust the power supply until the reading on the DVM meter is where you plan to measure. Adjust the voltage (DVM readout) in steps from 0 to 1.0 Volt in 0.1 V increments and record the current (Amps) in a spreadsheet.

3) Repeat at each tenth volt recording while the spreadsheet makes a sweet graph.

Test Results

First, I offer a quick look at my diode modeling setup. I am currently upping the ante on resolution in order to determine diode parameters Is and n, hoping for success soon! Setup and protocol..

Over a period of a few weeks I ran I-V characterization tests on a good number and variety of diodes, crystals and my two tubes to see how very thing stacks up. The following presents the resulting curves, diode photos, and some puzzling questions/conclusions I churned up in the process. I start off with the realization that when one orders diodes, it is by no

Something was needed here! At left is a photo of a very sweet brass shaft converter machined by my clever and capable EE elder brother, many beaucoup thanks for your addition to the project! Not being an engineer myself I was able to follow very poor practice and just wired the whole lot together, crossed my fingers and hooked the sucker up.. Ya Hey! There be sounds here!

The set is playing as I type these words and I am most hopeful for a successful performance goal. Although much testing and measuring yet remains to be done, the set by ear calibration tunes quite sharply. The coils have 4.2 inch OD and 3.2 inch ID and can slide upwards of 4-5 inches separation for looser coupling. The selectivity control has a dramatic effect on sensitivity and, I am sure, on selectivity as well.. to be measured soon so stay tuned.

The photos below present the set in its final functionning form.

tried this before on a version of an MRL-2 but it was a disappointment, not the right set for the test. As this set was to employ a high-performance circuit, an SEC would be appropriate. The set itself is only moderate performance though, the caps having fairly high-loss brass plates and my coils wound with solid wire. This set also ploughs new ground for me in my first basketweave coils. Basketweave coils recuce stray capacitance and increase coil Q by eliminating the need for lossy coil formers. As these were my first, I merely used solid wire I had in the house. As they are wound of 20awg solid wire, they are quite springy and I had to hot-glue the bejeezes out of them to get them to stay put. I do one day expect to build a truly hi-performance set, but this is just a stepping stone to test components and design. As such, I tossed a benny into the circuit for good measure.

The design calls for non-tapped coils to increase coil Q. While the coil Q is not great, it is certainly considerably better than any of my earlier cardboard tube or even PVC selonoids and it ought to have a decent tank parallel resistance. Matching (hopefully) to this I used a germanium FO-215 diode from Mike Peebles. The combination in these diodes of low n (~1.1) and Is (190) gives a diode with moderately high junction resistance on the order of 150+ kOhm which should work well with the solid-wire coils in the tank.

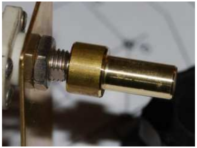

The selectivity enhancement circuit called for a cool 20pF differential cap which I found at Surplus Sales of Nebraska. The cap itself is a cute little item but the shaft was a short 1/4in long and only 0.19in in diameter.

means certain what one will get. This seems especially true for the supposedly ubiquitous 1N34A. Unless you see the part number actually on the diode, you probably need to test it to know what it really is. So, lets take a look!

Germanium Diodes

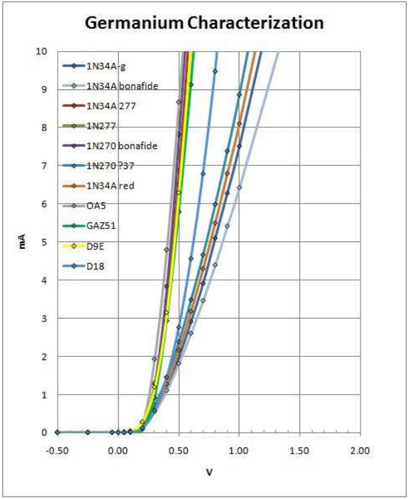

Hammarlund, Double-Tuned

http://www.lessmiths.com/~kjsmith/crystal/ham.shtml

This fine radio features three beautiful and high-quality Hammarlund square-law 510pF variable capacitors. I have been hankering for some time to build a radio with more than just OK to mediocre performance, which translates to a double-tuned set with a properly tuned front end. The capacitors mentioned before were, at first, a poor choice for this project as I needed for the antenna tuner a two-ganged variable cap and the Hammarlund units were only single. Additionally, I only had one! With what I can only call considerable luck, two additional identical capacitors were listed over the last several months on ebay giving me three total. The caps are high quality and very heavy construction and I figured I ought to be able to dismantle them enough to give a common shaft to two thus ganging them as needed. This turned out to be the case and the project was a go.

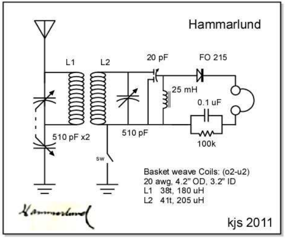

Additionally, with each radio I build, I like to test and learn new circuit concepts and ideas. This radio was to be made with a farily simple double-tuned circuit, a tuggle-tuned front end and variable coupling. I also wished to test and play with the "Hobbydyne" or selectivity enhancement ideas presented by Dave Schmarder on his excellent site. I had

1N34A series

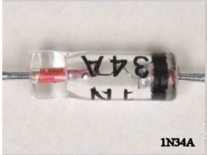
1N34A

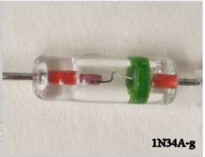
1N34A-g

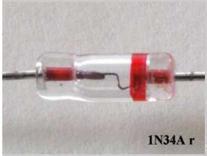
1N34A r

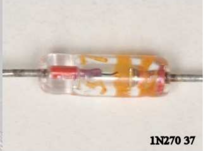
1N270 37

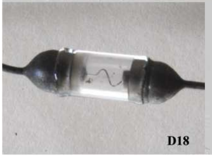
D18

1N270-277 series

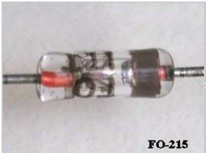
FO-215

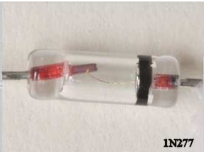
1N277

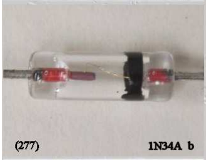
(277) 1N34A b

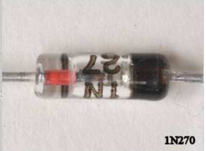
1N270

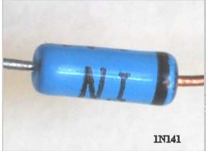
1N141

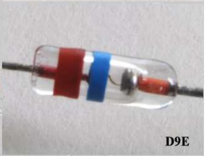
D9E

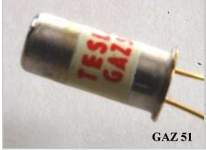
GAZ 51

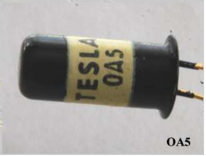

OA5

When I did get around to hooking the set up to the signal generator I found to my relief that indeed the set has good (not great) selectivity and really excellent sensitivity. Bandwidth around a nominal 1100KHz signal to -3dB is 22KHz giving this set a loaded Q of 51. This is right in the middle range of set loaded Q to be expected for a "typical" crystal set given by Kenneth Khun in his notes on Resonant Circuits. Importantly as well, this set does a much better job at rejecting KGOW than my other sets. I attribute this to the main coils being toroids. KGOW blasts down my direction with a very powerful ground wave and I suspect that the other set coils, mostly open selenoids and / or spiderweb, are picking up the signal and flooding my set across the tuning range. Toroid coils confine the magnetic fields within the torus and thus couple poorly if at all to surrounding fields, including KGOW. Thank you Magnavox for making such lovely coils!!

View from above.

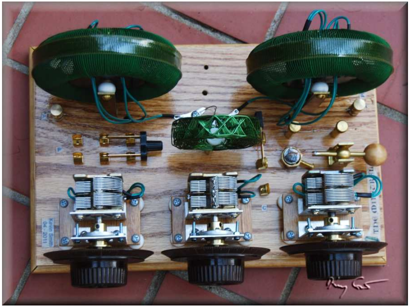

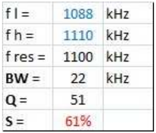

f l =	1088	kHz
f h =	1110	kHz
f res =	1100	kHz
BW =	22	kHz
Q =	51	
S =	61%	

Performance summary for the set is quite pleasing. When I first started listening to the radio I immediately got the sinking feeling that the selectivity would be no better then many of my earlier sets, a disappointing prospect especially as this set was meant to be a higher performer. Part of the problem what just the sets' general quirky and difficult three-dial tuning method. I soon got used to fiddling with the dials and found I could usually "tune out" most any nearby or interfering station from the one of interest, it just took some effort.

Here we see a set of Germanium diodes and photos of the diodes under test. Immediately you should notice two sets of curves, what I call the "1N34A" set and the "1N270-277" set. In this analysis, the second set appears clearly superior in sensitivity and deservedly so. In the photos it is clear that one bunch of diodes I purchased as 1N34A were in fact better 1N277, the black band and thin gold contact wire as well as common curve unite these. "True" 1N34A's have a more robust contact wire. I also purchased a 1N270 and received pretty orange diodes with the number "37" stamped on it, a fairly robust contact wire and a curve that most closely resembles a 1N34A, go figure. Mercifully, I DID manage to get some bona-fide 1N34A's with the part number on the diode, as well as a single bona-fide ITT 1N270 diode with the part number. These are my base for comarison.

In addition to the standard Germanium Diodes tested above, I also have latched my hands on a few "other" germanium diodes for the fun of testing. These include two vintage Russian types, a D9E (in the 270 class) and a D18 (transitional between 270 and 1N34A). Additionally, I got a few FUZZ diodes popular, I suspect, with the guitar gadget crowd. These are larger packages in metal cases so I cannot see how the internal contact is made other that they are stated to be gold bonded point contact type. Clearly these are highly sensitive diodes in my 270 class, the OA5 looking best of the lot. Interestingly, I recently purchased from good Mr. Peebles a few of his "Holy Grail" ITT FO-215 diodes. My measured characteristic for one of these is a dead laydown on the gold-bonded OA5. From the photos, these are in different packaging with the FO-215 in a traditional glass casing. What makes the FO-215 so great is the fact that its resistance Rd is an interesting 150k ohm or so which matches very well with typical moderate Q tank curcuits. Interestingly, I find that an old Russian D18 diode has very similar characteristics to the FO-215 and so much also be placed into the "Holy Grail" class of germanium diode. Take your pick!

Spreadsheet for calculation of Germanium diode n and Is (modified from Mike Tuggle's spreadsheet)

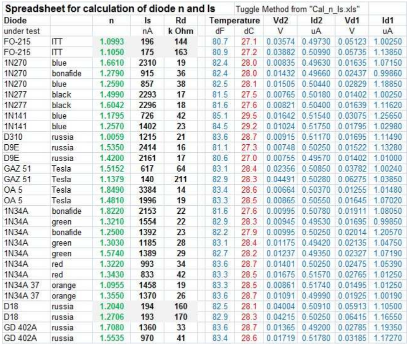

Spreadsheet for calculation of diode n and Is Tuggle Method from "Cal_n_Is.xls"

Diode		n	Is	Rd	Temperature		Vd2	Id2	Vd1	Id1
under test			nA	k Ohm	dF	dC	V	uA	V	uA
FO-215	ITT	1.0993	196	144	80.7	27.1	0.03574	0.49730	0.05123	1.00250
FO-215	ITT	1.1050	175	163	80.9	27.2	0.03882	0.50990	0.05735	1.13850
1N270	blue	1.6610	2310	19	82.4	28.0	0.00835	0.49630	0.01635	1.07150
1N270	bonafide	1.2790	915	36	82.4	28.0	0.01432	0.49660	0.02437	0.99860
1N270	blue	1.2590	857	38	82.5	28.1	0.01505	0.50440	0.02829	1.18850
1N277	black	1.4990	2293	17	81.5	27.5	0.00765	0.50180	0.01402	1.00250
1N277	black	1.6042	2296	18	81.6	27.6	0.00821	0.50400	0.01639	1.11620
1N141	blue	1.1795	726	42	85.1	29.5	0.01642	0.51540	0.03075	1.25650
1N141	blue	1.2570	1402	23	84.5	29.2	0.01024	0.51750	0.01795	1.02980
D310	russia	1.0059	1215	21	83.6	28.7	0.00915	0.51170	0.01695	1.11490
D9E	russia	1.5350	2414	16	81.1	27.3	0.00748	0.50250	0.01522	1.13280
D9E	russia	1.4200	2161	17	80.6	27.0	0.00755	0.49570	0.01402	1.01000
GAZ 51	Tesla	1.5152	617	64	83.1	28.4	0.02356	0.50850	0.03782	1.00240
GAZ 51	Tesla	1.1379	140	211	82.9	28.3	0.04491	0.50280	0.06275	1.03850
OA 5	Tesla	1.8490	3384	14	83.4	28.6	0.00664	0.50370	0.01255	1.01480
OA 5	Tesla	1.4810	1996	19	83.3	28.5	0.00865	0.50550	0.01645	1.07020
1N34A	bonafide	1.8220	2153	22	81.6	27.6	0.00995	0.50780	0.01911	1.08050
1N34A	green	1.3210	1554	22	82.9	28.3	0.00945	0.49530	0.01695	0.99850
1N34A	bonafide	1.2500	1392	23	82.2	27.9	0.00995	0.50250	0.02014	1.20570
1N34A	green	1.3030	1185	28	83.1	28.4	0.01175	0.49420	0.02135	1.04750
1N34A	green	1.5740	1389	29	82.7	28.2	0.01237	0.49350	0.02327	1.07190
1N34A	red	1.3220	993	34	83.6	28.7	0.01401	0.50250	0.02475	1.05390
1N34A	red	1.3430	833	42	83.3	28.5	0.01675	0.51570	0.02765	1.01250
1N34A 37	orange	1.0955	1458	19	83.3	28.5	0.00861	0.51740	0.01495	1.01250
1N34A 37	orange	1.3550	1370	26	83.6	28.7	0.01091	0.49990	0.01925	1.00190
D18	russia	1.2040	194	160	82.5	28.1	0.04004	0.50910	0.05913	1.10500
D18	russia	1.2706	193	170	82.9	28.3	0.04215	0.50250	0.06415	1.16550
GD 402A	russia	1.7080	1360	33	83.6	28.7	0.01365	0.49200	0.02785	1.19350
GD 402A	russia	1.5535	970	41	83.4	28.6	0.01719	0.51780	0.03185	1.17270

Averages: n = 1.38 Is = 1315 nA Rd = 52 kOhm

The above spreadsheet is based on measurements of the diodes shown above. For the determination of Is and n, I chose at random two examples from my collection of various diodes, (all germanium in this case) and measured via a modified version of the methodology outlined by Ben Tongue and Mike

Unfortunately the primaries do not have taps. 180 uH is a bit low for MW work and would require a variable cap with a fairly high top end. I chose three identical 2-gang 270 pF caps tied in parallel to give three 20 - 540 pF tuning. The link circuit consists of two serial 35 pF secondaries for a total of 70 pF, not enough to tune with so I would need to add further inductance. Here is another new feature for me. I made for this set my first honeycomb-style coil, green, of course. I gave the link coil 110 uH for a total link inductance of 180 uH, just as in the primaries. All tune with the three variable caps. Additionally I added a series-parallel knife switch for the antenna circuit capacitance. For my setup the series position will work best, but this will add utility later on. (One green feature I tested and subsequently abandoned was a green LED diode, but even with bias I found it to be hopelessly insensitive).

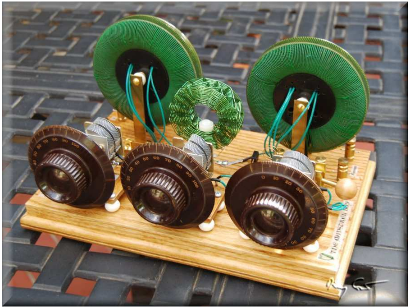

Portrait of set, Wearin O'the Green..

sharpen the tuning overall. This method forms a triple-tuned set with its consequent quirkiness and occasionaly annoying tuning. Still, I find this part of the fun, I am not looking for tuning efficiency, just sharpness.

The lovely green color and large size of the coils would be a prominant design feature and this immediately suggested an Irish theme for the set. I settled on naming it "The Minstrel Boy" from a patriotic Irish song written by Thomas Moore, reproduced at the bottom of this page. Of course the name also reflects the idea that I expect to listen to music on the set, it will be a minstrel for me. This decision then set in motion the need for other green features, green hookup wire and making a link coil from green magnet wire. Finally a 1N34A diode I bought a couple years back from Dave Schmarder with the green band to cap things off.

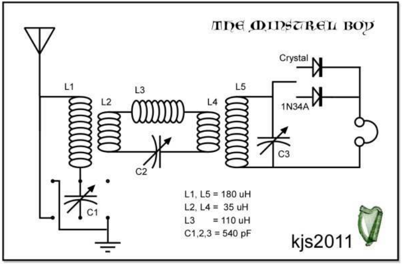

The toroid coils have identical features including a primary with 180 uH and a secondary coupling coil with 35 uH.

Tuggle. For this work I have chosen to set Id2 about 0.5uA and Id1 about 1.0uA and then read the needed voltage. This is the inverse of the normal method but the justification is that for any diode regardless of forward voltage drop the measurement is made at the same part of the LOG I vs V characteristic. This allows comparison between all diodes. I have included the actual room temperature in the calculations in order to eliminate this variable as a source of doubt or error. All the measured parameters as well as the determined results are presented in order to facilitate repeatability.

Schottky Diodes

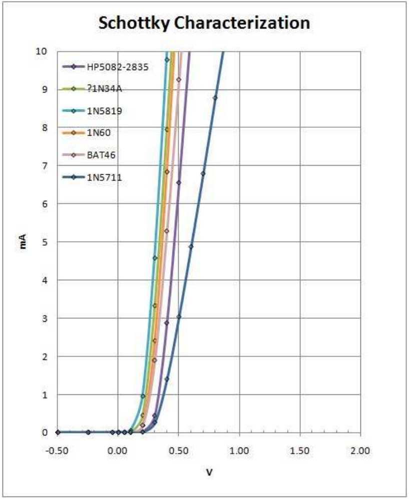

The Minstrel Boy, my triple-tuned triple-threat!

http://www.lessmiths.com/~kjsmith/crystal/minstrel.shtml

I am always on the lookout to find interesting and different circuits with which to construct my radio projects. About a year or more ago I found on ebay an interesting deal for a set of three vintage toroid Magnavox (Model D) air-core coils. They are wound from a beautiful lime-green cloth-covered wire and I just had to make an effort. I assumed, correctly as it turns out but not obvious from the description, that they would come with internal coupling coils. I also assumed they would have a tap somewhere on the main coil but in this I was to be disappointed. I was also disappointed to discover that in one of the three coils the secondary connection was broken. Oh well, my project needed only two coils and I knew these would make a fine radio. I already had a link-coupled design in mind.

Toroid coils are interesting in that they confine the magnetic field within the center axis of the toroid. Because of this, stray coupling seldomely causes a problem, but then again, coupling in general is not easy. So, I had the idea these would be my ticket to exploring designs with capacitive or link couplings between circuits. Eventually I began scouring the net and my books/references for circuits employing link coupling. This coupling method seems to come in two main varieties, tuned (such as found in the Crystal Set Societys' "My Marconi" set), and untuned (as found say in MRL's #42 set). Tuning the link is not needed, but it was something I wanted in an effort to virtually no difference between the Narrow and Broad performance. Oh well... At least I understand now why Osterhoudt removed the choke from the MRL#2 set.

Q for this single-tuned radio is comparable to other radios I have built. This is a circuit demonstration radio presenting Fleming's valve. Nor have I finished with the SEC concept. The circuit utilized here is not ideal for the SEC, ground is taken off the top of the coil, its an odd, if interesting, configuration. I will give the SEC another shot with some other radio. Overall, I am very pleased with the set and enjoy listening to the local Spanish music format station with the warm glow of the tube keeping me company. Oh that glow....

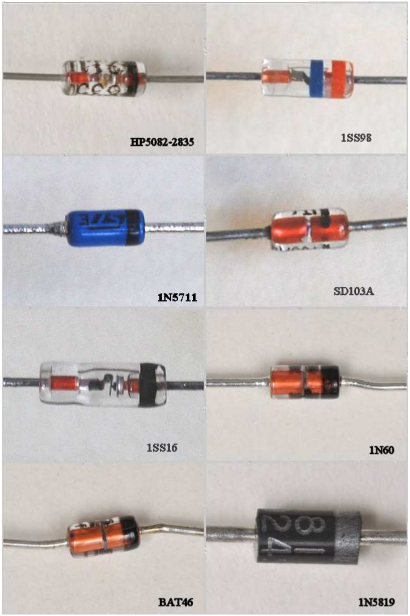

Schottky diodes are very sensitive diodes that work excellently in crystal radios. Their construction and theory are different and I confess to not fully understanding these components. Still, from the characteristic curves, they are excellent! Again, note that you dont always get what you bargain for. Here I found what was supposed to be 1N34A's to be some sort of Schottky of unknown pedigree. The 1N5819, while having the lowest forward voltage drop, has a rather very high Junction Capacitance and will not perform well. Posts on the RaidoBoard Crystal Radio Forum however highly recommend the 1N5711 for crystal sets although the characteristic curve dosnt look that fabulous. Many web pages out there on Shottky diodes, I recommend you do your homework. Of all the schottky's, I note that Ben Tongue recommends most highly the HP5083-2835. The high resistance Rd makes them useful for DX sets with very very high Q tanks. Even so the diodes need to be paralleled with up to 4 or 5 diodes to match correctly the Rd with the tank Rp. I have found these somewhat hard to find and expensive, especially when one requires using several in parallel. I recently measured a few 1SS98 diodes and I discover them to have characteristics extermely similar to the HP and I feel they deseve more attention. Sadly, on ebay they seem to be just as difficult to find and expensive. No free ride!

Spreadsheet below for calculation of Schottky diode n and Is (modified from Mike Tuggle's spreadsheet)

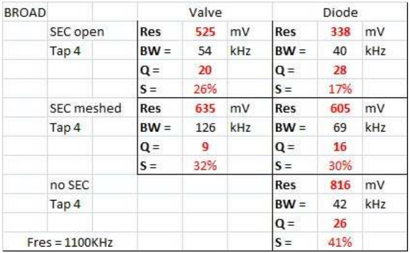

BROAD		Valve			Diode		
	SEC open	Res	525	mV	Res	338	mV
	Tap 4	BW =	54	kHz	BW =	40	kHz
		Q =	20		Q =	28	
		S =	26%		S =	17%	
	SEC meshed	Res	635	mV	Res	605	mV
	Tap 4	BW =	126	kHz	BW =	69	kHz
		Q =	9		Q =	16	
		S =	32%		S =	30%	
	no SEC				Res	816	mV
	Tap 4				BW =	42	kHz
					Q =	26	
Fres = 1100KHz					S =	41%	

Looking at the performance in the Broad setting we find the SEC has a definite impact. As before, the valve has higher sensitivity than the diode, but here we see a classic tradeoff between sensitivity and selectivity. Turning the SEC to its maximum selective position (lowest capacitance) helps, (Q increases from 9 to 20 with the valve) but at the price of lower sensitivity, (Decrease from 32% to 26%). More disturbingly, the overall sensitivity in the broad setting, which ought to be higher than in the narrow setting, (that is, greater than 40%) is quite the opposite, (less than 30%). Additionally, setting the SEC to its most selective position merely returns the radio to a selectivity already found in the narrow setting (narrow Q = 18, broad Q = 20) while lowering the sensitivity. So, I consider the SEC in this circuit a bit of a failure. My suspicion is that the insertion losses associated with the SEC are damaging and no real selectivity boost is given that was not already found in the narrow setting. To address this, I removed the choke and minimized the capacitance of the SEC, effectively removing it from the radio. Re-testing the radio now showed a return to good sensitivity and moderate Q. Actually, with this test I see

(Russian D9E, a 1N270 equivalent) and the tube/valve (Russian 6C19P) in place.

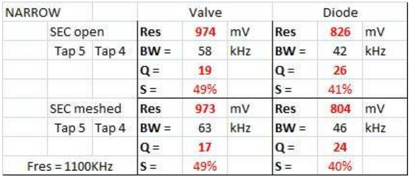

NARROW			Valve			Diode		
	SEC open		Res	974	mV	Res	826	mV
	Tap 5	Tap 4	BW =	58	kHz	BW =	42	kHz
			Q =	19		Q =	26	
			S =	49%		S =	41%	
	SEC meshed		Res	973	mV	Res	804	mV
	Tap 5	Tap 4	BW =	63	kHz	BW =	46	kHz
			Q =	17		Q =	24	
Fres = 1100KHz			S =	49%		S =	40%	

Looking first at the radio in the Narrow switch setting. The table at left indicates the peak output DC signal in mV at resonance, the measured bandwidth at -3dB above and below resonance, calculated radio Q and Sensitivity (Sensitivity here is simply 100 * DCout / ACin). The table shows that the SEC circuit itself provides no change in performance. This is expected as noted above, the SEC doesn't impact the circuit in the narrow setting. I ran the valve at 6.3 V on the heater filament and am surprised to see that the radio sensitivity is higher with the valve than with a diode. Dissappointingly, the valve had a poor effect on the radio Q, going from Q = 25 with the diode to Q = 18 with the valve. I have to imagine the valve has a rather different impedance / capacitance characteristic from the diode. Still, the good sensitivity is pleasing to see.

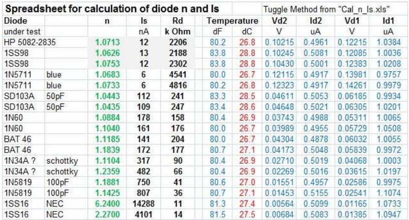

Spreadsheet for calculation of diode n and Is							Tuggle Method from "Cal_n_Is.xls"			
Diode		n	Is	Rd	Temperature		Vd2	Id2	Vd1	Id1
under test			nA	k Ohm	dF	dC	V	uA	V	uA
HP 5082-2835		1.0713	12	2206	80.2	26.8	0.10215	0.4961	0.12215	1.0384
1SS98		1.0626	13	2188	83.8	28.8	0.10245	0.5081	0.12085	1.0036
1SS98		1.0753	12	2302	83.8	28.8	0.10430	0.5001	0.12383	1.0208
1N5711	blue	1.0683	6	4541	80.0	26.7	0.12115	0.4917	0.13981	0.9757
1N5711	blue	1.0733	6	4816	80.2	26.8	0.12323	0.4917	0.14261	0.9979
SD103A	50pF	1.0443	112	241	83.3	28.5	0.04611	0.5053	0.06185	0.9934
SD103A	50pF	1.0435	109	247	83.4	28.6	0.04648	0.5021	0.06305	1.0201
1N60		1.0884	178	158	80.4	26.9	0.03743	0.4988	0.05311	1.0065
1N60		1.1040	161	176	80.0	26.7	0.03989	0.4955	0.05729	1.0508
BAT 46		1.1185	141	204	80.0	26.7	0.04304	0.4878	0.06032	1.0055
BAT 46		1.1839	172	177	80.7	27.1	0.04173	0.5048	0.05839	0.9972
1N34A ?	schottky	1.1104	317	90	80.4	26.9	0.02710	0.5019	0.04068	1.0003
1N34A ?	schottky	1.2359	482	66	80.4	26.9	0.02269	0.5016	0.03615	1.0197
1N5819	100pF	1.1881	750	41	80.6	27.0	0.01551	0.4957	0.02586	0.9975
1N5819	100pF	1.1425	807	36	80.7	27.1	0.01453	0.5155	0.02541	1.1074
1SS16	NEC	6.2400	14288	11	81.3	27.4	0.00564	0.5099	0.01165	1.0733
1SS16	NEC	2.2700	4101	14	81.5	27.5	0.00684	0.5083	0.01385	1.0947

Averages: n = 1.48 Is = 1275 nA Rd = 1030 kOhm

The above spreadsheet is based on measurements of the diodes shown above. For the determination of Is and n, I chose at random two examples from my collection of various diodes, (all schottky in this case) and measured via a modified version of the methodology outlined by Ben Tongue and Mike Tuggle. For this work I have chosen to set Id2 about 0.5uA and Id1 about 1.0uA and then read the needed voltage. This is the inverse of the normal method but the justification is that for any diode regardless of forward voltage drop the measurement is made at the same part of the LOG I vs V characteristic. This allows comparison between all diodes. I have included the actual room temperature in the calculations in order to eliminate this variable as a source of doubt or error. All the measured parameters as well as the determined results are presented in order to facilitate repeatability.

Silicon Diodes

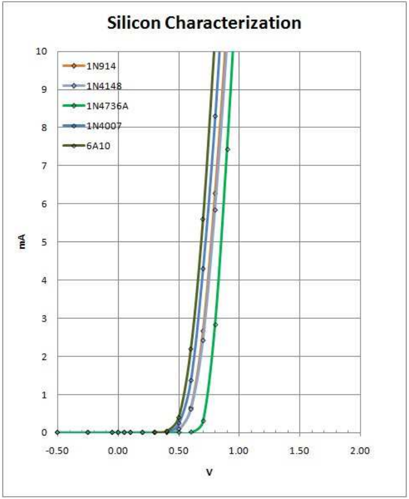

The glow.... Oh the glow!

PERFORMANCE:

So, is is a keeper? Does it perform well? Is Fleming's valve a revolution? The following tests answer all these questions and them some. We start by noting from the circuit diagram above that this is essentially two radios depending on whether it is switched in the Narrow or the Broad setting. Importantly, in the narrow position the SEC option is essentially out of the circuit. This will be seen in the measurements as well. A note here, all measurements were made at a nominal 1100KHz with a signal generator and the output of the generator was a sine wave with 2.0V p-p. This is a strong signal and may impact the results, especially concerning sensitivity, but it makes for easier measurements and ready comparison between sets. For this set I made identical measurements with both the diode

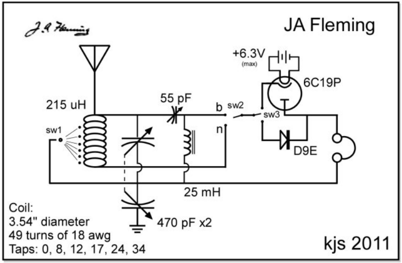
JA Fleming
+6.3V
(max)
6C19P
215 uH
55 pF
b
sw2
sw3
n
sw1
D9E
25 mH
Coil:
3.54" diameter
49 turns of 18 awg
Taps: 0, 8, 12, 17, 24, 34
470 pF x2
kjs 2011

The actual set with its controls, taps, tuning and SEC shown below:

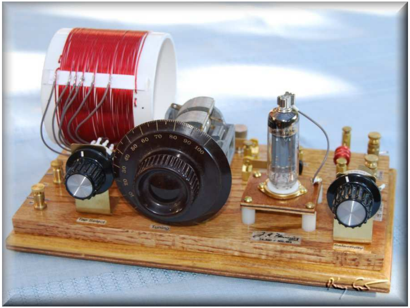
Tap Select
Tuning
Selectivity

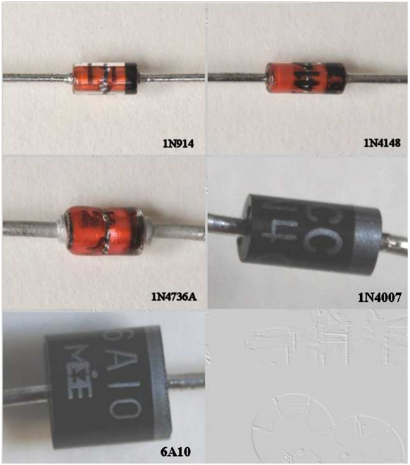
1N914
1N4148
1N4736A
1N4007
6A10

Silicon diodes have good characteristics, but an unacceptably high forward voltage drop making them a very poor choice for crystal radio unless used with bias. The 1N4736A is a Zenner diode.

Spreadsheet for calculation of Silicon diode n and Is (modified from Mike Tuggle's spreadsheet)

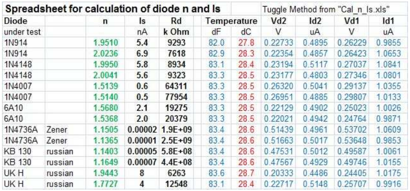

Spreadsheet for calculation of diode n and Is

Tuggle Method from "Cal_n_Is.xls"

Diode		n	Is	Rd	Temperature		Vd2	Id2	Vd1	Id1
under test			nA	k Ohm	dF	dC	V	uA	V	uA
1N914		1.9510	5.4	9293	82.0	27.8	0.22733	0.4895	0.26229	0.9855
1N914		2.0236	6.9	7618	82.9	28.3	0.22354	0.4857	0.26423	1.0653
1N4148		1.9950	5.8	8934	83.1	28.4	0.23194	0.5117	0.27037	1.0841
1N4148		2.0041	5.6	9323	83.3	28.5	0.23177	0.4803	0.27346	1.0801
1N4007		1.5139	0.6	64311	83.3	28.5	0.26320	0.5041	0.29137	1.0355
1N4007		1.5140	0.5	77954	83.3	28.5	0.26951	0.4885	0.29807	1.0133
6A10		1.5680	2.1	19275	83.3	28.5	0.22129	0.4902	0.25023	1.0026
6A10		1.5368	2.0	20379	83.3	28.5	0.22021	0.4942	0.24764	0.9871
1N4736A	Zener	1.1505	0.00002	1.9E+09	83.4	28.6	0.51439	0.4961	0.53702	1.0609
1N4736A	Zener	1.1365	0.00001	2.5E+09	83.4	28.6	0.51663	0.5017	0.53648	0.9853
KB 130	russian	1.1403	0.00005	5.8E+08	83.4	28.6	0.47531	0.5012	0.49587	1.0061
KB 130	russian	1.1649	0.00007	4.4E+08	83.4	28.6	0.47567	0.4929	0.49746	1.0155
UK H	russian	1.9443	8	6263	83.6	28.7	0.20333	0.4486	0.24405	1.0175
UK H	russian	1.7727	4	12548	83.1	28.4	0.22717	0.5148	0.25707	0.9916

Averages: n = 1.52 Is = 2 nA Rd = 2.4E8 kOhm

The above spreadsheet is based on measurements of the diodes shown above. For the determination of Is and n, I chose at random two examples from my collection of various diodes, (all silicon in this case) and measured with a modified methodology to that outlined by Ben Tongue and Mike Tuggle. In this case, with the radically different forward voltage drop of these diodes from Germanium or Schottky diodes, I have kept the Id values constant (about Id2=0.5uA and Id1=1.0uA) and varied Vd. I have included the actual room temperature in the calculations in order to eliminate this variable as a source of doubt or error. All the measured parameters as well as the determined results are presented in order to facilitate repeatability.

Additionally I purchased a small variety of tubes for testing. The results of this study are discussed in full under the Radio Lab : Diode Test section of my page. Of all the tubes tested, I finally chose the Russian 6C19P as the tube having the most sensitive characteristic and a handsome appearnce. The tube is a power-hungry beast though, (1.1 Amp current draw at its rated 6.3 V), and this radio requires a power supply, but its worth it! The following graph plots the characteristics of seven tubes I tested for the project.

While the project radio was to feature the Fleming valve, I naturally wish to have an interesting circuit as well. As always, it would be a breadboard set and single tuned for ease of use. I also wished to utilize and test Dave Schmarder's selectivity enhancement circuit (SEC) in the set. This last item to compensate for the typical poor selectivity in single-tuned circuits. One circuit that caught my eye from an early time was Modern Radio Labs' MRL#2 set. To this set I would add the SEC modification. In such configuration I notice that it bears a remarkably close resemblance to Set #6 in W.J. Mays' Boys Book of Crystal Sets (BBCS). The MRL#8 set also shows design affinities and includes the choke missing in MRL#2, a fact noted by Mr. Osterhoudt himself. In the end my circuit design, shown below, must be considered a close, (but not exact), member of a family of radio circuits consisting of MRL#2, MRL#8, and BBCS#6.

Most avenues of research proved futile until he recalled his prior work on the Edison Effect. He literally retrieved the original bulbs he had experimented on previously and set up a test to apply the effect for radio wave rectification. His experiments were certainly successful. This idea proved a novel and unanticipated use of a prior-known effect amounting to true discovery and thus was awarded British (24,850) and American (803,684) patents in 1905, these successfully defended against De Forest's subsequent infringements. Pondering this interesting history, I began wondering about the Fleming diode and thinking, in crystal radio, this is not amplification, so why not? Also, one must consider the glow, oh the glow… I just had to build this.

Finding a correct tube for the project is a bit of a story in itself. I have never worked with tubes and, while it appeared natural that diode tubes ought to exist as well as triode/pentode/etc (they do), I had no idea what tube was what nor how to know what would work for radio rectification. Many diode tubes I found are for power supply work and may not be appropriate for radio detection at all, (actually they work fine). Jumping into the search both feet, I waded through the datasheets of many/most diode tubes to get to know them well enough to choose.

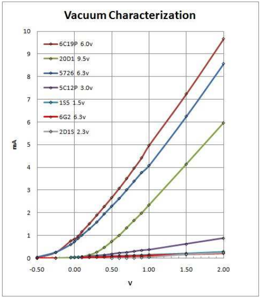

Light Emitting Diodes

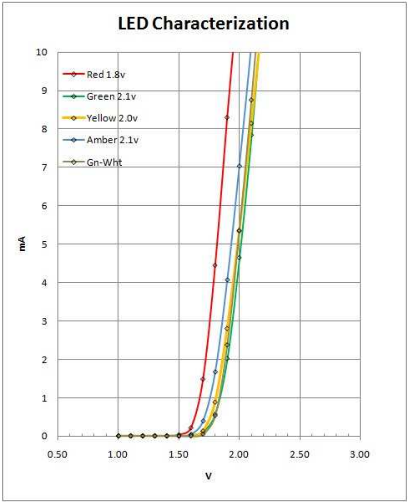

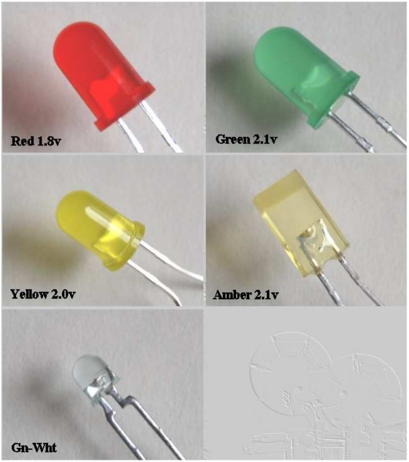

As long as I am measuring various and sundry diodes, I figure I ought to include that most ubiquitous of modern diode, the LED. Found everywhere, these diodes are rapidly becoming the low-energy light source of choice for many lighting applications. I have read occasionally of someone asking whether these ought to be useful for radio applications as well. To this question the answer is generally a resounding "No!". The turn-on voltage is waaay above any reasonable value

Powered by Fleming's amazing valve

http://www.lessmiths.com/~kjsmith/crystal/flm.shtml

Crystal radio circuits come in a seemingly endless variety of forms and possibilities. That and the historical context of their importance at the dawn of the radio-age makes them a fascinating hobby. I enjoy making these radios, learning and exploring new and interesting circuit ideas, incorporating vintage components into my sets, and finding historical connections between the radios and the fathers of the technology.

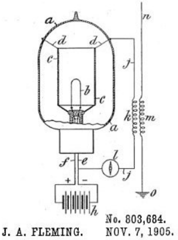

My latest project involves a passive receiver rectifying the signals with a vacuum diode. Professor John Ambrose Fleming had consulted for Edison Electric in Britain and had experimented extensively with Edison two-electrode bulbs in the 1880's, presenting his results on the "Edison Effect" to the Royal Society of London in 1889, and the Royal Institution in early 1890, and finally his definitive paper to the Physical Society of London in 1896. Later after the turn of the century he began consulting for Marconi seeking better mechanisms for radiotelegraphic detection of radio signals.

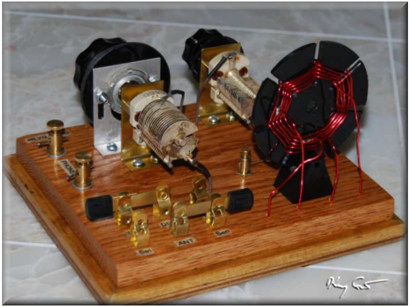

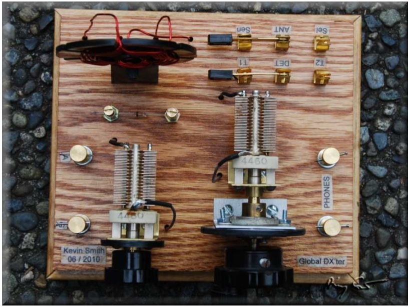

expected to be delivered by an antenna to a crystal set. Still, this simple answer avoids the actual question, what in fact does the characteristic curve of a LED really look like? Where is the turn-on voltage with respect to the published junction voltage, (assuming you can find that).

To provide just such a look I visited my local electronics store and bought a small handful of LED's, most with the junction voltage listed and took them home to measure. Typical LED junction voltages seem to range from about 1.8v to 2.1v or more. The turn-on voltages look closer to 1.6v-1.7v. Anyone used to working with Carborundum crystals or silicon diodes will be used to biasing their rectifier to get good sensitivity. These LED's once on have a very sharp rise and with a proper bias should work quite well as detector diodes. As a bonus you will get a sweet glow as well. OK, not as cool as the glow of a vacuum diode, but certainly more sensitive!

Spreadsheet for calculation of Light Emitting Diode n and Is (modified from Mike Tuggle's spreadsheet)

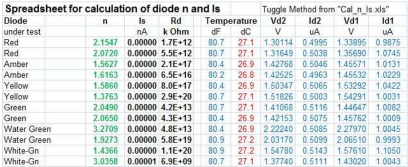

Spreadsheet for calculation of diode n and Is						Tuggle Method from "Cal_n_Is.xls"			
Diode	n	Is	Rd	Temperature		Vd2	Id2	Vd1	Id1
under test		nA	k Ohm	dF	dC	V	uA	V	uA
Red	2.1547	0.00000	1.7E+12	80.7	27.1	1.30114	0.4995	1.33895	0.9875
Red	2.0720	0.00000	5.5E+12	80.7	27.1	1.31649	0.5038	1.35690	1.0745
Amber	1.5627	0.00000	2.1E+17	80.4	26.9	1.42768	0.5046	1.45571	1.0131
Amber	1.6163	0.00000	6.5E+16	80.2	26.8	1.42525	0.4963	1.45532	1.0229
Yellow	1.5860	0.00000	8.0E+17	80.4	26.9	1.50347	0.5065	1.53292	1.0422
Yellow	1.3763	0.00000	2.9E+20	80.7	27.1	1.51826	0.5003	1.54291	1.0031
Green	2.0490	0.00000	4.2E+13	80.7	27.1	1.41068	0.5116	1.44647	1.0082
Green	2.0650	0.00000	4.3E+13	80.4	26.9	1.42153	0.5075	1.45762	1.0009
Water Green	3.2709	0.00000	4.8E+13	80.4	26.9	2.22240	0.5085	2.27970	1.0045
Water Green	1.9273	0.00000	5.8E+19	80.9	27.2	2.03170	0.5099	2.06510	0.9993
White-Gn	1.4366	0.00000	1.1E+20	80.9	27.2	1.54780	0.5143	1.57610	1.1050
White-Gn	3.0358	0.00001	6.9E+09	80.7	27.1	1.37740	0.5111	1.43020	1.0043

Averages: n = 2.01 Is = Tr nA Rd = 3.8E19 kOhm

The above spreadsheet is based on measurements of the diodes shown above. For the determination of Is and n, I chose at random two examples from my collection of various diodes, (all led's in this case) and measured with a modified methodology to that outlined by Ben Tongue and Mike Tuggle. In this case, with the radically different forward voltage drop of these diodes from Germanium or Schottky diodes, I have kept the Id values constant (about Id2=0.5uA and Id1=1.0uA) and varied Vd. I have included the actual room temperature in the calculations in order to eliminate this variable as a source of doubt or error. All the measured parameters as well as the determined results are presented in order to facilitate repeatability.

Why know the values of Is, n, and Rd of your diodes?

I have been attempting to get a handle on the main diode parameters Is, n, and Rd and how they impact the operation of a crystal radio. My method of choice is to use a graphical approach, plotting charts where variations between plotted parameters become apparent and easy to visualize. I have developed an interesting chart from the measurements above that compares many diodes with respect to Is, n, and Rd, and to understand how this fits into a matching situation with a fuzzy indication of typical tank impedance limits.

Ultimately one seeks a diode with an Rd that matches with and conjugate to the impedance it sees from the tank (+antenna). In the following discussions I will start with the cross-plot of Is to Rd and then proceed to a look at how received RF power plays a part in the selection and why low Is and n are desired.

What I found then, was a farily sensitive but, to be truthful, woefully selective set tuning from 6 to about 17-18 MHz. Not bad for a single-tuned rig.

Final thoughts. Upon returning to my home base, I quickly hooked the set to my longwire to make first impressions of actual reception. In midday I was able to pick up a spanish language format broadcast and two english language religous format stations and a brief encounter with WWV. At night the bands are flooded with competing stations, success! Note that all this with my handy dandy Radio Shack mini-amplifier. When I connected a crystal earphone and used my poor hearing, there was nothing as expected. One pity is the ongoing presence, even on this set, of KGOW 1.65MHz forever audible in the background. I consider the set a success as I was not expecting selectivity in a single-tuned rig. Still, with the bands dominated by religous and spanish broadcasts, I cant say it differs much from the local Houston AM fare. Oh well...

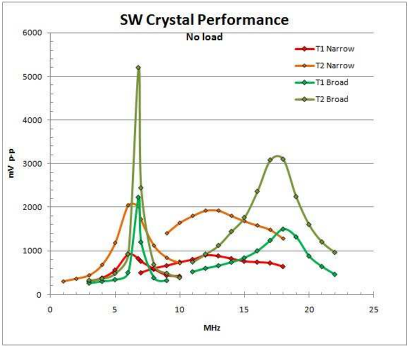

The plot above charts the received signal strength (p-p mV) versus input frequency for the four possible switch configurations (Detector on tap 1 or 2, Antenna on selective [top of coil] or broad [on tap t1]. These four conditions were tested with the tuning cap fully meshed (low frequency) and fully open (high frequency) to give an idea of the actual tuning range as compared to the calculated range. All measurements were made with no load on the output, and the antenna trim cap in the full meshed position. Other conditions not shown were also tested.

The results indicate that placing the antenna low on the coil, on the T1 position gives the best sensitivity and selectivity in tuning. The detector, a 1N277 germanium diode placed high on the coil at T2 also gives dramatic increase in sensitivity.

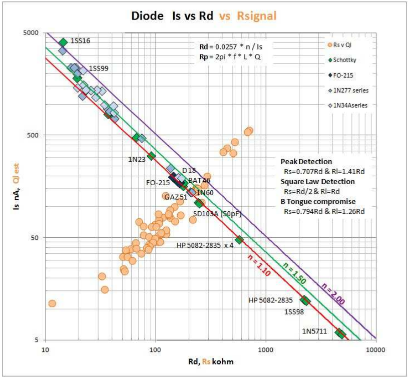

The plot above shows the relationship between diode Is and Rd. Rd is calculated via the equation Rd = VT * n / Is where VT is the thermal voltage = k * T / q = 0.0257V at room temperature and Is and n are from the measured diodes. k is Boltzmann's constant = 1.38E-23 J/K and q is the electron charge = 1.609E-19 coulombs. On such a plot I can show lines of constant n, and plot the values for individual diodes for which I have spent considerable effort to determine the parameters Is, n and Rd (where Is is the diode saturation current, n is the ideality factor, and Rd is the diode resistance). On the plot it should be evident how changes in n or Is impact a diodes' Rd.

Additionally I have plotted RF signal Rs as orange circles from published coil unloaded Q (see my section on Coil Q). To convert from Qu to Ql I have made the assumption that loaded Q is about 1/5 unloaded Q for low-end sets while in performance sets it may approach 1/2.5 the unloaded Q. This is only an estimate then, but starting from actual data. To make the conversion to Rs I simply plugged the data into the standard equation R = 2pi*f*L*Q using a scaling factor to divide down Qu to Ql from high Q (1400/2.5) to low Q (100/5). This provides a nice visual display of the expected range for Rs in many crystal sets. I plot the Rs against the value of Ql used. Note that these data are for the case at about 1MHz and inductances between 200 and 300uH, as published.

Matching the diode to the tank is a matter of finding a diode with both sensitive qualities (low n) and an Rd close to the Rs presented to the diode (as per Ben Tongue's suggestions for Peak and Square Law detection). Those diodes with Is's in the 100 - 250 nA (100-280 kohm for n=1.1) or so range have the possibility to match while connected to the top of the tank coil without the use of Q-lowering taps, while untapped matching to high-end high-performance big-Litz baskets will require diodes with Is's in the 35 - 60 nA (500-800 kohm for n=1.1) range. Most "typical" germanium diodes have high Is values (>500, >60kohm) and will require a tap. Diodes such as FO-215, BAT 46, 1N60 and GAZ51 can be matched to many tanks without taps. For tanks with high quality Litz coils one will be using HP 5082-2835 or 1SS98 diodes, generally 3 to 4 in parallel to lower the Rd to the desired range. (Curiously, I have not found any diodes with Rd values in the 300 to 4500 kohm range.) Note that both resistance and impedance are frequency-specific so a match at one frequency will not remain matched across the broadcast band.

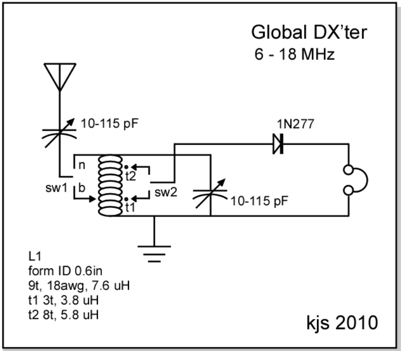

Initial testing on the set was dismal. NO RECEPTION. There followed many hours of double checking all wireing, solder joints, fidelity to the schematic, etc. Everything checked out, no good. Time to seek professional help, which happily corresponded to my annual visit home. There my radio engineer brother took the radio and gave it his close attention, puzzled over it as had I, and concluded a defective part somewhere. This he soon found in the tuning capacitor which had had a connection slip off its set point. This easily corrected, the radio seemed ready for testing. Unfortunately he did not have an antenna handy (he works in uhf+ research stuff) so all I could do was hook the set to a signal generator and test its performance, summarized below.

at 3 turns and 8 turns from the ground (center) of the coil. I ended with 9 turns on the final set with testing giving me a 7.6 mH coil. The final schematic presented below.

A
G
PHONES
Global DX'ter
Kevin Smith
06 / 2010

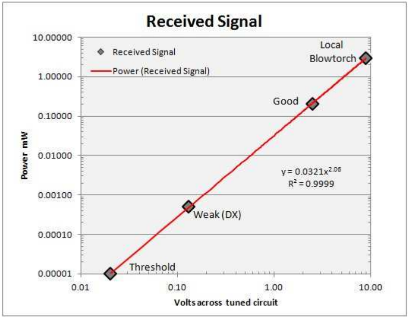

The above is an interesting plot taken from the text discussion in "Crystal Set Analysis" by Berthold Bosch. It presents received signal strength across the tuned circuit for various scenarios from threshold audibility to local blowtorch. This plot should be considered a single example specific to his location and antenna/ground system. In the text he describes his antenna an an inverted L 43m long (140') and 10m high (32'), an excellent antenna most of us do not have the real estate to erect, but offsetting this is a poor ground with Rg = 210ohms. Given the offsetting conditions I would imagine this is a good generic example of what received signal voltage and power levels presented across the LC circuit to the diode are likely to be. Bosch cited 40nV as the threshold of audible detectability (impedance matched conditions with 16kohm RF impedance / 4kohm DC phones) and for this plot I pushed it back to 20nV as it gave a superior regression fit.

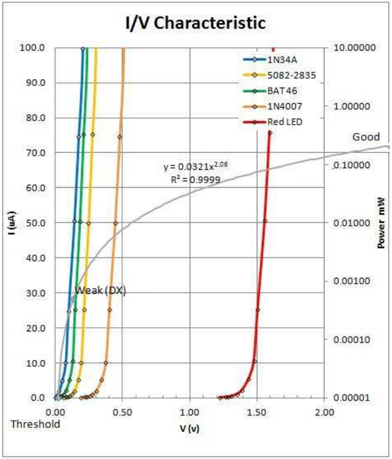

This plot at left takes some actual diode I/V measurements rather than the models presented below and plots them along with the above RF voltage across the LC circuit versus RF power in mW. Here one sees that pushing for very very low Vf (via high Is / low n combinations) pushes the limit of detectable signal power from the antenna. At such minute power levels I imagine the reverse leakage current becomes significant and probably offsets the low operating point gains sought.

Global DX'ter
http://www.lessmiths.com/~kjsmith/crystal/sw2.shtml

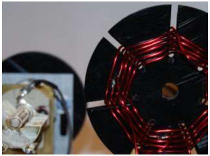

Finally, a shortwave set comes off the design bureau and out into the world. This is a type of set I have been planning and pondering on for some time. With many crystal radio design possibilites this will certainly not be the only SW set I plan, but it is my first effort, my chance to "do it" on shortwave. For a first effort I settled on a fairly simple design, a single-tuned rig with some degree of control, but not overly much. I took as my base a design by Rainer Steinfuehr on his Crystal Reciever World site. This design features a high Q spider coil with two taps, and adjustable capacitance on both the antenna and tuning circuits. It made a rather lovely small set indeed.

The specifics of the design were entirely my own and were based on the frequency range I wished to target and the specifications of the materials I had available for the project. In partictular, I had available small good-quality 10 - 115 pF variable caps for which my clever elder brother had lathed lovely 1/4 inch shaft extenders for use with standard tuning knobs. My target spectrum was about 5 to 18 MHZ as that is where the dominant international BC action takes place. With luck that would also allow receiving WWV at 10 and 15MHz as well. Digging into the design spreadsheets, it seemed an inductor with 7 to 9 mH would best deliver the goods. I made a spider form from a circlar cut of 1/8 inch Delrin plastic and wound 11 turns of 18awg magnet wire, carefully soldering taps

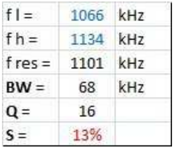

f l =	1066	kHz
f h =	1134	kHz
f res =	1101	kHz
BW =	68	kHz
Q =	16	
S =	13%	

Performance summary of my Bremer-Tully Set is so so with a set Q (loaded) of 16, -3db bandwidth of 68 kHz and a sensitivity calculation only 13%, not so great. I rather expected as much as this is a single-tuned set of very simple circuit design. Compensating slightly is the fact that it tunes with a single dial which is certainly more user-friendly. It does require a wave trap to cut out powerhouse KGOW blazing away and flooding the entire BC band on this set if left untrapped. Most of my radios have higher sensitivity so this set definitely benefits from my audio amplifier.

Bank winding in six banks: 4 wires / side, 10 windings per bank, times 6 = 60 turns. Coil form 3 inches in diameter, space wound 0.8 inches wide calculates to 376 uH, pretty close to my measured 377 uH. Measurements: Primary B-P = 20 uH. Secondary F-G = 377 uH, G-T = 210 uH, T-F = 65 uH. I know, 65 + 210 not equal 377, I measured it several times to be sure!

Crystal Galena and Pyrite Diodes

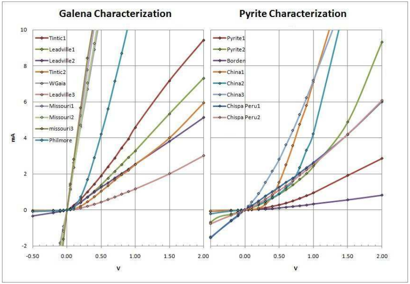

The above curves demonstrate the wide variation in properties and qualities that can be found in natural crystals of galena or pyrite, the two most common and best quality natural stones. In each crystal test I first poked around the crystal for some time to 1) determine a typical sensitivity for the crystal in question and 2) to locate the best hot spot with which to test. This turns out to be a non-trivial exercise on the diode test setup. In a crystal radio one need only listen for the loudest spot. With the test setup one needs test both the forward and reverse current in order to determine if the whisker is on a hot spot or not. Very tedious work! (In retrospect, if I were starting over making a test jig, I would definately add a DPDT switch to readily change between forward and reverse current measurements. I'd probably toss in a rheostat for fine adjustments as well).

For many crystals there are limited number of possible hot spots, but these may be hot indeed. For most of my "Steel Galena" samples (Tintic Utah, or Leadville Colorado) there are numerous hot spots under virtually every place I touch the probe, but in general the sensitivity is good to so so. These crystals are very kind to work with in terms of finding spots and avoiding frustration. Mirror galena on the other hand may have quality hot spots, but any hot spots at all are rare and frustratingly difficult to locate. Here my Philmore detector crystal shines with an almost ideal "Galena" response. To chase down this rabbit I purchased some lovely mirror galena from Sweetwater Missouri. I broke off a few appropriate-size chunks to pot in woods metal and test. At first I was very excited with the high currents I was seeing at moderate voltages. Figured I had struck gold. When these crystals failed miserably to rectify anything in my radios, I re-measured things in both forward and reverse directions. These crystals obey Ohm's law and act like typical resistors, not suitable for radio work at all.

For my pyrite crystals the work has been especially tedious and frustrating. With one of the crystals one "hot spot" alternated, entirely on its own, between hot and bad while I was making the measurements. I would start over and over, sometimes getting interesting readings then suddenly it would drop to low values and I'd start over, back and forth. I present this data as best as I have measured, and I don't intend to go back! You see at least one of the crystals, my "China 1", (from a lead/silver mine in Hunan) gave a sweet classic-looking curve. More to the point, "ideal" curves for natural minerals, are difficult to come by. Most crystals you use will be less than ideal. The good news is that, while listening to your set, poking about for a good spot is far easier than what I have

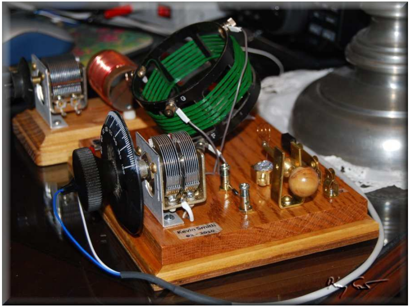

Circuit schematic for the set.

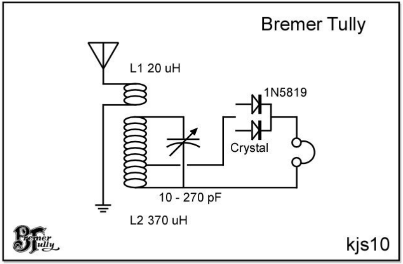

tune down to 500 kHz. The closest variable cap I have found is a cute little 2-gang 10-270pF unit. This is as low as I have found at the low end for a standard-configuration variable cap. The goal then is to try and keep the circuit capacitance as low as I can, which I translate into keeping the circuit as simple as possible.

My circuit is further constrained by the coil itself. It is essentially an "oatmeal box" type setup with an aperiodic small inductance primary and large inductance secondary. Tuning then must be single on the detector side only. My only option is whether I connect the crystal/diode to the tap post or to the top of the coil. I went with the tap, so shoot me! One complication I maintained, as I do on all my sets, is to keep a switch between using a crystal and cat's whisker or using a diode. In this case I chose a Schottky 1N5819, the most sensitive diode I have yet tested, (diode test). I have yet to test just how high on the BC band I can tune, but KGOW at 1560 kHz sure gives me trouble.

---

gone through to produce these curves. Your ear will take care of you!

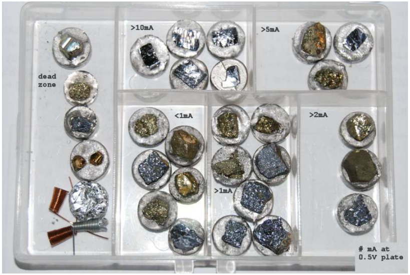

In the photo I indicate groupings based on an easy measure of performance. I note the current in milliamps for each set where the plate voltage is set at 0.5V. The greater the current the better are your chances to get a sensitive crystal, assuming Ohm's Law is not followed! Crystals in the "dead zone" on the left will have their woods metal re-melted for new detector crystals and the bad ones tossed, its tough love for crystals. I find that easily half the potted crystals I make are tossed this way, and only a few can be considered superlative.

Vacuum Diodes

When I decided early on to construct a radio based on the vacuum diode (see my Fleming Radio section), I had to find a suitable tube for the project. That began with an extensive search through the datasheets checking vacuum diode properties and reviewing the characteristic curves for a large number of diode tubes. In the above figure I plot the characteristics as best I can on a commpn plot for comparison. I was hoping to find a good candidate with sensitive characteristics and low energy consumption. As you will find, such a beast did not exist. Following this research began a period ot purchase and testing.

I have tested seven different tube types including 6.3V rectifier diodes that take a lot of power to run and are not really suitable for battery use, a 9v dual diode tube (20D1), a pentode/diode tube (1S5) that is designed to run on a battery at 1.5V and 50mA, and a couple miscellaneous but cute tubes. In testing the 1S5, I found the filament never glowed incandescant at 1.5V and, on measuring, was barely sensitive to anything. I cannot imagine this tube would make much of a diode for crystal radio use. I tested different manufacture 1S5 tubes from two different suppliers. No dice. Finally, I tried to push the tube to operate at higher-than-specified voltages. At 4.3V the I-V curve was still very flat, but shifted slightly higher, crossing the Amp axis at about 0.4 mA, (see graph). At 5.5V one of my tubes gave up the ghost and I didnt do more. I pretty much rule the 1S5 out for my crystal radio work. Looks like i'll be needing the power supply when I get around to building/operating that Fleming Radio.

A question remains as to what in fact is different about these tubes to give such different characteristics. While the I/V

A Bremer- Tully Transformer

http://www.lessmiths.com/~kjsmith/crystal/brt.shtml

M. D.
Bremer Tully
532 S. Canal St.
CHICAGO, ILL.
AC-1
AIR CORE TRANSFORMER
SERIES BANK WINDINGS
TYPE

I recently won on eBay and lovely NOS Bremer Tully air core transformer its inspection tag still faithfully wired to the base, (and so it will remain on the final radio). It is absolutly gorgeous and I was the only bidder when it sold, for such a great price! Honestly, I was not totally sure what I was getting when I bid other than it looked like a coil and, from the ebay photo, maybe would work for shortwaves. Calling it a transformer was confusing although I understand all coupled coils are transformers. Perhaps that is why I was the only bidder, well, so much the better! When it arrived I got a good look, very interesting winding technique. The tuning coil is wound with somthing like #30 dcc or dsc wire. The winding is in 6 bank-wound groups of about 10 windings each. Inside is a small primary antenna coil. I promptly got out my handy-dandy inductance meter and measured 377 uH on the tuning and 20 uH on the primary. In addition, there is a tap on the tuning coil one "stack" in. 377 uH is a bit on the high side for BC work and I'm not sure I'll ever get the top end of the band but still, the transformer works beautifully.

My calculations for a setup using a 377uH coil suggest that a minimum total capacitance of 23 pF will allow tuning up to 1700kHz. This is a tall order, most variable capacitors I have found have a minimum value in the 20 - 25pF range. Add another 15 - 25pF of stray circuit capacitance and I am way over the line. The good news is that only 270pF is needed to

characteristic of most vacuum diodes pretty much follow the Langmuir-Child Law, the steepness of the I/V characteristic is largely due to the geometry of the tube elements and the volume of electron space-charge between them. JB Calvert's Theory of Vacuum Tubes informs one that this is the property called Perveance in a vacuum diode. Measuring this requires plotting I raised to the 2/3 power against V and taking the slope of the best-fit line through the data. This slope raised to the 3/2 power is the perveance. The following chart illustrates graphically the relation bewtween the tube characteristic and perveance. For the 5726 tube I plot the characteristic in dark blue for the tube heated with the design 6.3 volts. In light blue I have plotted the measurements and raised them to the 2/3 power. One sees immediately that the new curve is linear. The slope of the curve raised to the 3/2 power yields the perveance factor for the tube.

(Note in addition that when running the tube at a more sensitive point with 4 volts only on the heater the perveance is slightly less than spec.).

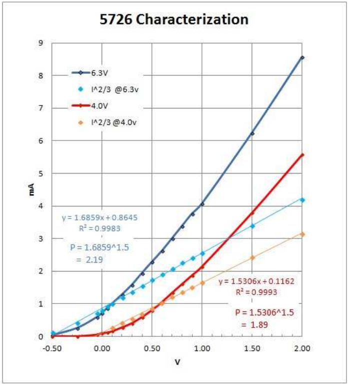

With my measurements calculating the perveance is practical and quickly done, results as follows:

Tube	Perveance	R2
6C19P	2.54	0.997
5726	2.19	0.998
6DN3	5.40	0.997
20D1	2.02	0.985
5C12P	0.32	0.999
1S5	0.063	0.999
2D1S	0.155	0.975
6G2	0.050	0.999

however, is NOT the inventor of the oscillating and regenerative feedback circuit, which forms the basis of all radio amplification and truly opened radio to a mass mode of communication. This honor goes to E.H. Armstrong, the man who dug into the tube and worked out correctly for the first time its basic theory and operation. The two men fought a bitter 19-year litigation battle over this honor that in the end was awarded to De Forest by the Supreme Court in what amounts to a technicality.

From what I have read of De Forest’s writings, patents, and exchanges with Armstrong in the Proceedings of the Society of Radio Engineers, it is pretty clear that De Forest never really understood how his tubes worked. So, he was never in a position to develop the tubes to their full potential. Armstrong for his part seems to have treated De Forest with academic contempt. Armstrong went on to invent the super-heterodyne (1918) and super-regenerative (1922) amplifier circuits and created almost single-handedly FM radio (1933) as we know it today. His life ended in tragedy, representing the closing chapter in the age of great individual inventors, finally crushed by the greed of corporate America. Virtually all advances following his were the product of team efforts in company or government research laboratories.

If therefore you have misgivings as to my honoring De Forest with a radio, please understand I do so not for his Supreme Court victory but to honor the inventor of the triode tube, a debt which even Armstrong would always acknowledge. Should you still feel uncertain, then just enjoy the sweet irony in honoring the father of the triode tube with a crystal radio.

De Forest circuit diagram (aka Sleeper #18).

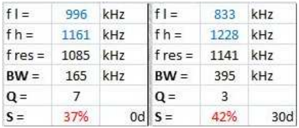

f l =	996	kHz	f l =	833	kHz
f h =	1161	kHz	f h =	1228	kHz
f res =	1085	kHz	f res =	1141	kHz
BW =	165	kHz	BW =	395	kHz
Q =	7		Q =	3	
S =	37%	0d	S =	42%	30d

Performance summary of my DeForest is really just abysmal, it has a low set QL of 7 reflecting a huge bandwidth of 165 KHz but my sensitivity calculation is 37% for the condition of tight coupling. This set clearly trades Q and bandwidth for sensitivity. The set is technically double tuned with variable capacitance in the detector circuit and variable inductance via taps on the antenna circuit. The inductors are loosely coupled via rotating the inner inductor with respect to the outer. What I discover in fact is that rotating the inductor changes the tuning and so it "feels" more like a variometer rather than a variocoupler. Rotating the rotor 30 degrees to loosen the coupling does not narrow the bandwidth, in fact it increases to 395 KHz. I have much to test still on this set, it is a disappointment as I expected a much better QL. Measurements were made with the detector circuit farily optimized for a test frequency of about 1100 kHz, but I wonder if I can better optimize the tap settings for the antenna circuit. More to come.

What’s in a name?

The name De Forest in the history of radio is one few engineers can hear without a fairly strong opinion. Dr. Lee De Forest is the acknowledged inventor of the triode radio tube. De Forest

The perveance numbers above range from 0.02 to over 2 mA/V^3/2 (with >2 being good for small-signal detection) and clearly shows my choice of the 6C19P to be an excellent one. I am surprised to see a huge 5.4 for the 6DN3 diode, a color television damping diode. It takes a Novar 9-pin socket and is a large tube so doing more with this tube may have to wait a bit.

Calvert's 2001 published data

Tube	Perveance
6AL5	2.42
6H6	0.50
7Y4	0.58
2X2A	0.017
6V3-A	2.3
6AX4GT	1.42
6AV6	0.085
1A3	0.075

The high-perveance diode tube types I tested show an interesting property in that their characteristic curves do not pass through the origin of the graph. These curves are wholly acceptable as crystal radio detectors although they will require power to operate. What I have noticed with such tubes is that when I turn them off after using them with a radio, on cooling they go through a period of much increased sensitivity (very loud) before fading to nothing. It is as though running them at the full 6.3V lowers their full potential as crystal radio rectifiers. The reason is in the high perveance of the tubes, the anode is already proximate to the space charge before any plate voltage is applied. Effectively, while the plate (anode) "Zero" voltage is measured with respect to ground, the plate itself is still positive with respect to the cathodic space charge.

Small currents will continue to flow even with a negative bias to the plate, (Contact potential).

In order to explore this idea, I ran a series of tests with the tubes running at lower operating voltages (effectively diminishing the space charge thus lowering the perveance) as shown:

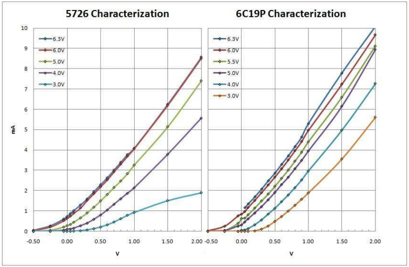

Here you see the impact of lower operating voltages. With the tube operating between 3 and 4 volts the characteristic curves begin to pass through the origin as in regular solid-state diodes and crystals. It is precisely in this voltage range that I found the radio sensitivity, as measured by loudness to my ear, to be most pronounced. It appears that the diode characteristic needs to pass through the origin of the I-V graph for highest sensitivity. This takes me to the 20D1 tube which operates at 9v but only 200mA. This will still take a power supply to use, but the good thing is that the characteristic curve passes

---

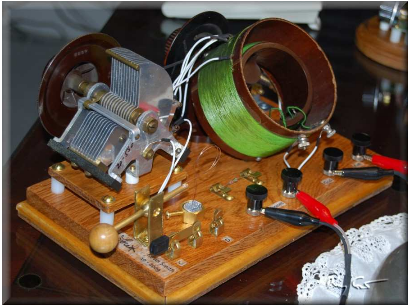

View from behind with my cat's whisker crystal detector, SPDT knife and 1N34A diode.

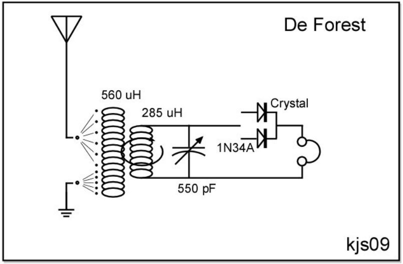

View from above highlighting the tap wiring, SO much work!

through the origin. Running at its design parameters the tube looks a lot like the 6C19P at 4v.

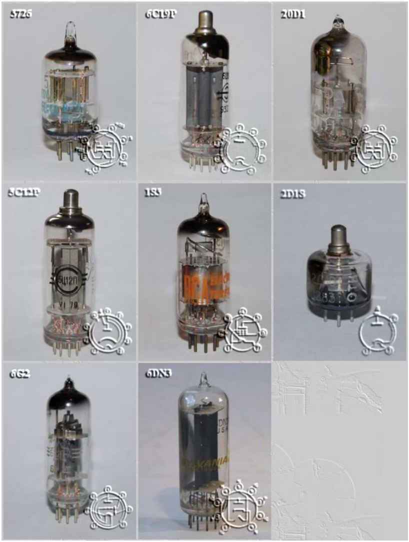
5726
6C19P
20D1
5C12P
1S3
2D1S
6G2
6DN3

Photo of the lab with my diode test jig, amp meter, and diode vacuum tubes under test.

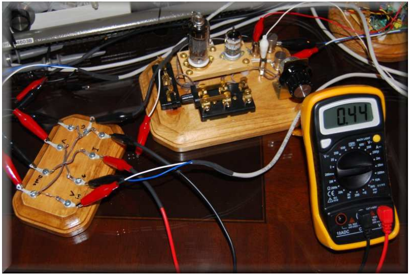

on each switch. The wood knobs have a nice easy action, the switch passing readily between contact buttons.

The radio works quite well, in its tightly-coupled setting it is quite sensitive with a strong signal. Its selectivity on the other hand leaves much to be desired. This can be expected from the circuit layout with only the secondary circuit capacitively tuned and the detector tied at the top. I was hoping that by varying the coupling I could achieve some increase in selectivity, but it de-tunes the station with an additional unacceptable loss of sensitivity. Take this outside the urban environment with such crowded airwaves and the set ought to perform admirably. Signals come in strong with the steel galena nearly as well as with the diode. I like this set.

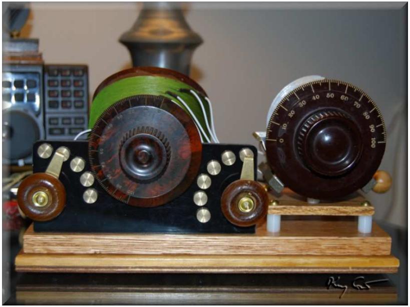

Portrait of set, ummmm..

As most of the components for this set were ready made, the construction planning felt a bit much like assembling a kit. I had originally intended to use modern tap-selector switches as well. In order to make the construction challenging, and more importantly, to maintain the vintage look and feel of the set, I quickly abandoned the modern switches and chose to fabricate instead classical wiper switches in the early style. Of course, I also include my own style cat’s whisker and SPDT knife switching to a more modern diode. The diode is my only nod to post war technology. In the crystal cup I have some lovely Utah steel galena potted in wood’s metal, finely crystalline, hot as hell all over its surface.

As one can clearly see, my obvious preference is for the breadboard style of early radio. Much of the fascination and beauty of these sets lies in the component hardware. To enclose them behind panels or put them in boxes seems such a waste. The base is a box with about ¼ inch of space inside to hide the bolts and wires. I did need a small panel to place the wiper switches on. This is made of 1/8 inch black Delrin plastic resin, a very hard and nice to work with material. In fact though, it is a bit too shiny, too “plasticy” for my taste. Well, that is what I had.

Fabricating the wiper switches, buttons, and leads to the inductor was an odyssey in its own right, amply satisfying my desire for more challenge. Attaching the leads to the primary, soldering without damaging the windings, and the added touch of shrink-wrap “mittens” was a labor of love. I do hope it is justified by the final presentation. The rest of the construction and wiring was a piece of cake by comparison. The wiper switch knobs are wood with a felt buffer against the panel. Behind the panel compression springs keep a steady pressure

RADIO TEST

Kevin Smith
http://www.lessmiths.com/~kjsmith/crystal/rtest.shtml

OK, now that you have built a radio, or several, how do they perform? Without a bit of objective testing you will never really know. I have often read descriptions of radio performance such as "tunes very sharply" or "separates two closely spaced stations" and/or etc. These are qualitative descriptions that do not really give you much information. To know your radio, it takes some measuring. I am making the assumption here that, as a homebrew radio builder, you will find tinkering and measuring your sets just as interesting as their design and construction.

For testing, I have pretty much lifted the excellent procedure outlined by Charles Lauter with a few modifications to suit my own philosophy. I will give my own protocol but first let’s look at the needed equipment and test setup. Testing a crystal set requires two pieces of homemade equipment, a dummy antenna in front of the set which allows measurement points, and a measurement load at the back end. Other pieces of equipment include a signal generator, an rms voltmeter, a resistance meter, and a good digital voltmeter. Circuit schematics for the homebrew components are shown below.

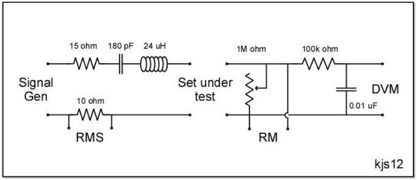

For the tests I originally found an old analog signal generator on ebay which I sadly found to be wholly inadequate. The unit was by no means cheap but the dial was difficult to read with any precision and the "play" in the dial made readings hopelessly inaccurate. Finally, it gave no indication of the attenuation or voltage output levels, rms or pp, of the signal. I have since found a nice, reasonably-priced new digital unit, (max's out at 2Mhz so strictly for AM band work) that gives good readings, gives me the output voltage level in pp, and can even produce different wave forms. I recommend if you are seeking to equip your lab, go this way from the start!

To calculate the signal input power you will need a good rms voltmeter. This is a non-trivial item and, after my experience with the vintage signal generator I was leery of getting a used meter. In any case, an expensive dedicated single-use piece of equipment seemed extravagant. For this measurement I have found entirely acceptable results using an oscilloscope which, of course, is useful for so much more and belongs in your lab regardless!

It is on the back end of the measurement where I diverge from Lauter's protocol significantly. He recommends placing in all

---

De Forest, Vintage beauty

http://www.lessmiths.com/~kjsmith/crystal/deforest.shtml

My latest effort, I call this my “DeForest” after the principal component of the radio. Over the last few months of visiting ebay and finding essential radio parts and browsing, I have stumbled across an occasional vintage piece that I just loved and needed to bid aggressively on. My intent is not to just stash these in a drawer somewhere for the pleasure of owning, but to use them in projects. This project is built around a lovely De Forest Radio Telephone and Telegraph Company Vario-Coupler. The vario-coupler, wound in beautiful apple-green silk covered wire 24 gauge or so, has a tapped primary stator measured at 560 uH and an untapped internal rotor of 285 uH. Other vintage components in this set include a Windham square-law tuning variable capacitor, 20 to 550 pF, and an Atwater-Kent 3 ¼ inch tuning knob. The second 3 inch knob on the vario-coupler is a vintage Radion Dial made by the American Hard Rubber Co.

I built the radio according to Sleeper hook-up #18. It is a fairly straightforward hookup but with sophisticated features including a double-tuned design, a tapped primary tuned via single and multiple taps, and variable coupling. The secondary coil is tuned with a vintage variable capacitor. As the secondary coil is the rotor inside the primary, there is no provision for tapping and the detector circuit is connected to the top of the coil. This set could use some selectivity enhancement circuitry I suppose, but oh well...

radios to be tested a 1N34A diode and then use a standard 2k ohm load resistance to make the output measurements. This may give a standard and comparable result, but it does not insure a good impedance match and does not reflect the intended usage and setup of many crystal sets, especially sets intended for high-performance. Here I am interested to learn just what is the optimal load resistance for my set under the test conditions. Note the caveat "under the test conditions". This may not reflect actual usage unless you take precautions. You will want to set the generator to an input voltage similar to what you expect the radio to receive in actual operation. You will probably want to signal level work the diode in its peak-detection region but not so high as to saturate the diode, (In making a loaded Q measurement you want to make sure that the diode impedance stays constant; otherwise, the change in diode impedance (as the tank attenuates the signal before the diode) will change the load on the tank and the bandwidth measurement won't really reflect the Q of the tank). Finally, for the output voltage you need a good voltmeter with very high input impedance that will not load down the set under test. This really means a good DVM. I use a used Keithley 192 bench meter with a 2M ohm input impedance, and as a backup a used Keithley 180 bench electrometer. Overkill perhaps, but for just in case...

Getting back to the load resistance then, my protocol is to find the optimum resistance to deliver the maximum power (not voltage!) to the load. This is a bit of a tedious process but will amply rewarded. Having tuned and peaked the set to the measurement frequency, I then take a series of resistance/voltage readings to find the maximum power output. In my circuit this is accomplished with the variable 1M pot. The resistance meter and voltmeter cannot be connected to the load at the same time making this a tiresome road to follow.

Once I find the optimum load resistance, I set the pot to that and proceed to the main testing for sensitivity and bandwidth/Q measurements.

Test Protocol as follows:

1) Connect the set to be tested between the dummy antenna and measurement load as per the above diagram.
2) Connect the signal generator to the dummy antenna and set the frequency and voltage output. I typically test in the center of the BCB at 1100khz and use a 200mVpp output sine wave.
3) Set the load resistance to somewhere around 50k ohms. The exact value is not critical as it will be adjusted later.
4) Connect the DVM and tune the set to peak the output voltage. Care here is amply rewarded, especially with double-tuned sets.
5) Once peaked it is necessary to adjust the generator frequency to re-peak the output voltage as hand-capacitance or other factors may have prevented perfect tuning. You now have your set properly tuned for maximum output. Presumably this also means the best possible impedance match between the set and dummy antenna. This is the resonant frequency (f res).
6) Record the frequency and output voltage.
7) Now find the optimum load resistance. This is done by making a series of paired measurements of load resistance and output voltage. Power in uW = mVout * Rl khoms.
8) Set the load resistor pot to the value where power output is highest, this is the optimum load for the test. Record this.
9) With the signal generator re-peak the output voltage and re-record the resonant frequency (f res) and output voltage (Vout mV).
10) Multiply mVout by 0.707 to find the -3dB level.

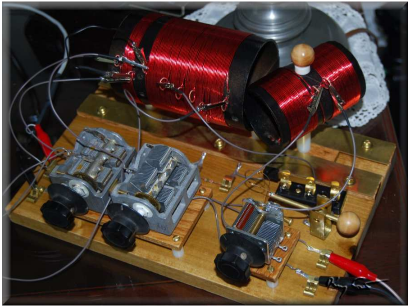

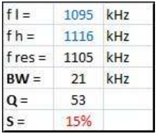

f l =	1095	kHz
f h =	1116	kHz
f res =	1105	kHz
BW =	21	kHz
Q =	53	
S =	15%	

My homebrew in its Galenatron configuration shows fair Q at 53 and poor-fair sensitivity of 15%. There is a 1cm separation between the antenna coil and detector coil which may help. In many radios of this design I see the coils separated by 2-3mm only. As I wound the coils on a single form, there is no experimenting here, just wondering. The variocoupler here is only a trap and was not used in this measurement. With the return to a single-tuned set you need expect performance downgrade, but greater ease of use for general listening. I feel this configuration is a good compromise for easy listening without the hassel fo manipulating two (or more) tuning dials. It is clearly superior to the Dunwoody and De Forest single-tuned sets.

general listener radio and far far from the performance achievable in high-quality DX designs. But it is teaching me as intended.

Inspired by Dan Petersen's Galenatron, I wished to play with a vario-coupler design. This hookup uses a single form with an antenna coupling coil of 10 turns and a tapped secondary coil with 62 turns. The first 2-gang vcap is unused. The single-gang variable cap tunes the rotor. So far the best position seems to be keeping the rotor in-line with the main coil. More learning to come.

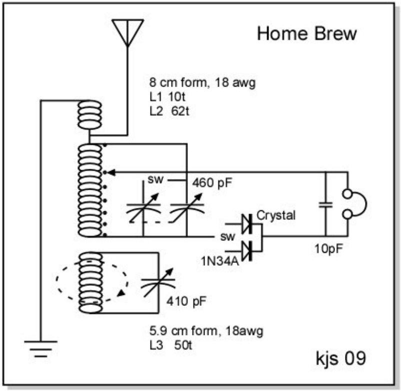

11) Adjust the frequency above and below f res until the output level falls to the level calculated in step10 above. These frequencies are f high and f low. Record these.
12) The -3dB bandwidth = f high - f low. Loaded Q (QL) = f res / BW
13) Take the generator output peak-peak voltage recorded in step 2 above and convert it to rms voltage as follows: mVrms = mVpp / 2.829. This is Lauter’s RF voltage E1 mV.
14) Measure the RF voltage across the 10 ohm resistor on the dummy antenna. This is Lauter’s RF voltage E2 mV.
15) Input current to the dummy antenna = Lin = E2/10 (mA)
16) Input power to the dummy antenna = Pin = E1 * Lin (uW)
17) Power loss in the dummy antenna = Pda = Lin^2 * 25 (the dummy antenna resistance)
18) Power delivered to set = Px = Pin - Pda uW
19) Set input resistance = Rx ohms = Ex / Lin where Ex = E1 - (Lin * 25)
20) Set output power (uW) = Pout = Vout^2 / Rl in ohms
21) Set % efficiency (sensitivity) = 100 * Pout / Pin

Having built a fair number of sets, and with varying quality, all the above protocol was developed to test my sets. What follows now is a discussion of how my sets stack up against each other and some notes on what one may expect in a crystal radio performance-wise. I give a summary table below of the essential data gathered on each of my sets.

Table 1

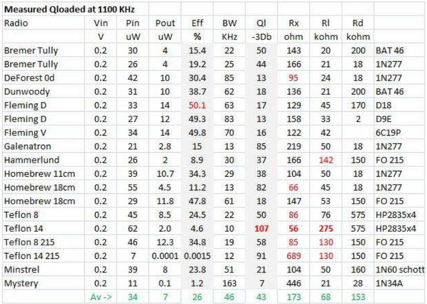

Measured Qloaded at 1100 KHz

Radio	Vin	Pin	Pout	Eff	BW	Ql	Rx	Rl	Rd	
	V	uW	uW	%	KHz	-3Db	ohm	kohm	kohm	
Bremer Tully	0.2	30	4	15.4	22	50	143	20	200	BAT 46
Bremer Tully	0.2	26	4	19.2	25	44	166	21	18	1N277
DeForest 0d	0.2	42	10	30.4	85	13	95	24	18	1N277
Dunwoody	0.2	31	10	38.7	62	18	136	21	200	BAT 46
Fleming D	0.2	33	14	50.1	63	17	129	45	170	D18
Fleming D	0.2	27	12	49.3	83	13	158	33	2	D9E
Fleming V	0.2	34	14	49.8	70	16	122	42		6C19P
Galenatron	0.2	21	2.8	15	13	85	219	50	18	1N277
Hammerlund	0.2	26	2	8.9	30	37	166	142	150	FO 215
Homebrew 11cm	0.2	39	10.7	34.3	29	38	104	50	18	1N277
Homebrew 18cm	0.2	55	4.5	11.2	13	82	66	45	18	1N277
Homebrew 18cm	0.2	29	11.8	47.8	61	18	147	53	150	FO 215
Teflon 8	0.2	45	8.5	24.5	22	50	86	76	575	HP2835x4
Teflon 14	0.2	62	2.0	4.6	10	**107**	**56**	**275**	575	HP2835x4
Teflon 8 215	0.2	46	12.3	34.8	19	58	85	130	150	FO 215
Teflon 14 215	0.2	7	0.0001	0.0015	12	91	689	130	150	FO 215
Minstrel	0.2	39	8	23.8	51	21	104	50	160	1N60 schott
Mystery	0.2	11	0.1	1.2	163	7	446	21	28	1N34A
	Av ->	34	7	26	46	43	173	68	153	

The above table of measurements were made at a nominal frequency of 1100khz and input voltage of 0.2Vpp as indicated on a digital signal generator following the protocol outlined. I give the radio name under test should you wish to refer to my pages for more on each set. The columns as follows:

Vin = 0.2V pp (0.71mV rms)
Pin = input power in uW into the set (after the dummy antenna)
Pout= output power across the load resistor in uW
% Eff = 100 * Pin/Pout
BW = bandwidth in khz at -3db
Ql = calculated loaded Q of the set
Rx = the set input resistance presented to the antenna
Rl = measured optimum load resistance for maximum power transfer
Rd = Diode junction resistance Rd used in the set

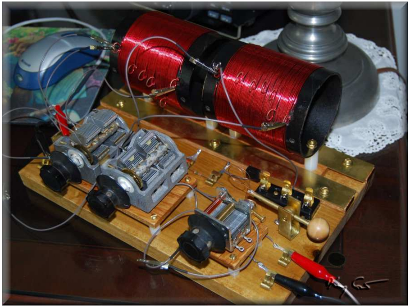

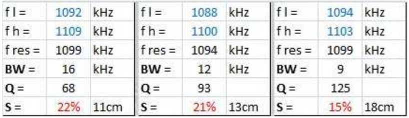

f l =	1092	kHz	f l =	1088	kHz	f l =	1094	kHz
f h =	1109	kHz	f h =	1100	kHz	f h =	1103	kHz
f res =	1099	kHz	f res =	1094	kHz	f res =	1099	kHz
BW =	16	kHz	**BW =**	12	kHz	**BW =**	9	kHz
Q =	68		**Q =**	93		**Q =**	125	
S =	22%	11cm	**S =**	21%	13cm	**S =**	15%	18cm

Here I show the performance summary for my set with three coil separations, 11cm, 13cm and 18cm. With these charts you see the value and trade-offs in loose coupling between the antenna and detector circuits. With the coils at 11cm separation they are slightly over-coupled but still with relatively good Q, 68, and fair sensitivity at 22%. Separating the coils to 13cm and then to 18cm loosens the coupling allowing narrower bandwidths and higher Q (getting to 125, woohoo!), but at the sacrifice of much sensivitity, down to 15%, still reasonable. This is an excellent performance, I must be doing something right somewhere. Of course this is still a

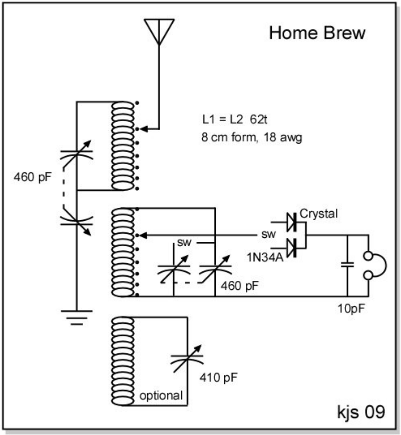
Home Brew
L1 = L2 62t
8 cm form, 18 awg
460 pF
Crystal
sw
sw
1N34A
460 pF
10pF
410 pF
optional
kjs 09

From the data above some interesting observations can be made.

Figure 1.

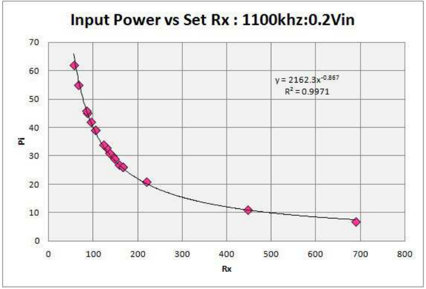

In figure 1 above I show the relation of the input power versus the set resistance. The quality of the fit reflects that fact that this is largely a calculated value, but it does illustrate the importance of having a low input impedance to match that of the antenna. This is not often considered in discussions on impedance matching but is critically important. The antenna has an impedance in the 25 - 50 ohm range, not easy to get this in a set without quality parts and construction. My sets typically have an input impedance in the low 100's ohms and only two of my sets really get down to the 50-60 ohm range giving a good match to the antenna. Both of these are double-tuned sets with a tuggle front end.

If you are serious about getting DX and doing a lot of heavy lifting, you will need a good antenna tuner as part of your set design.

Figure 2.

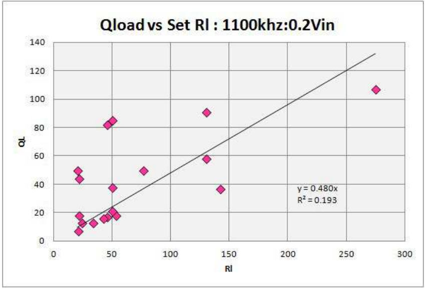

At the back end of the set things get a bit messier and are not east to show with a simple two-component plot. Here I have chosen to highlight the importance of the proper load resistance on the set loaded Q. In my set testing I have tried to pay close attention to the importance of the load on the tank and diode. Early on I considered construction of an output transformer unit that would work with different radio’s as well as with different phones. Such a design can be found on Ben Tongue’s excellent and highly technical web site. The plans were well beyond what I was able/willing to take on so I turned my attention to discovering just what load resistance exactly do my sets require for maximum power transfer. After all, if I found only a small range typical values, I might design

My second homemade, A double-tuned open test bed
http://www.lessmiths.com/~kjsmith/crystal/homebrew.shtml

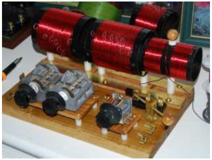

Having learned my lesson with the Mystery set, I have set myself to make a double-tuned rig with open wireing for testing multiple configurations. The concept is to seek the maximum amount of flexibility in hookup and coil design. This will mean interchangeable coils with or without taps, and wireing with allegator clips. I have chosen to hard place in the set three capacitors, 2 double gang 20 - 460 pF and one 20 - 410 single unit. The detector, as always, will switch between a diode and my own crystal stand. The modular design will allow testing coils mainly although vcaps will also easily be tested with jumpers. The number of leads and allegator clips, not all used with any one configuration will give this a bit messy look, but this is not a beauty contest but for learning. As always I include the option of using either a germanium diode or actual crystal.

The photos and circuits below show two hookups currently completed.

Double tuned circuit with tapped coils and an optional QRM trap (unused in photo). This configuration alone will keep me busy quite a while! I am learning what "sharp" tuning is. The antenna coil is tunes series-parallel but the first, (left) 2-gang vcap while I only use one gang of the second vcap. For lower inductance coils I can use the second gang in parallel.

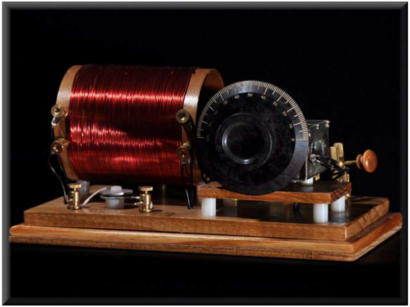

a generic transformer that would work well with many or all minus the complicated switching.

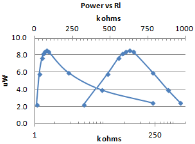

In my testing protocol therefore, I included the important step of determining the optimum load resistance for the set under test conditions. I certainly part company with the school of thought that says to test all sets exactly the same way, including same diode (1N34A typically) and same test load. Sets are designed to work best when impedance matched and at the back end that generally means a load matched to the tank/diode combination that delivers the maxumum power. At right above I show a small plot where I varied the load resistance over a wide range and calculated the output power of the set. (I give the plot with both a linear and a more appropriate log scale). This should demonstrate the importance of optimizing the load resistance.

Optimum load resistance depends on both the tank Q and that portion of the diode detecting the signal. That in turn depends on the signal level. Very weak signals will be rectified in the square-law region of the diode characteristic and require relatively high (and variable) load resistances. Normal to strong signals will be rectified primarily along the peak-detection portion of the diode characteristic and require lower load resistance with less variation. Not shown in the above spreadsheet, I also ran measurements with a 2.0Vpp input. With this stronger signal, rectification was taking place farther out on the diode characteristic. As expected, the measured optimal load resistances for this case was similar to or lower than that for the 0.2Vpp input case.

With all that said, the plot in figure 2 is messy but the impression is that of increasing Q with increasing Rl. I view the plot as having a number of small groupings of measurements with similar conditions, each showing higher Q's associated with higher load resistance.

Figure 3.

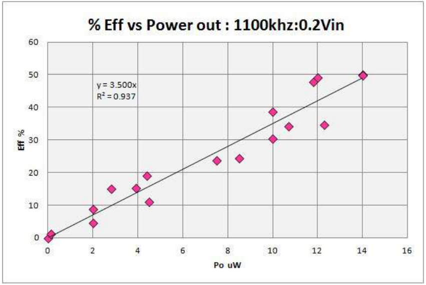

Figure 3 is most obvious with the clear benefit of good set efficiency on output power. An efficient set will better approach the maximum power transfer desired. High Q plays a part here. Still, there is always the tradeoff between selectivity (high Q) and sensitivity. An efficient set delivers more power to the load, but to go after Q you will ultimately be sacrificing sensitivity.

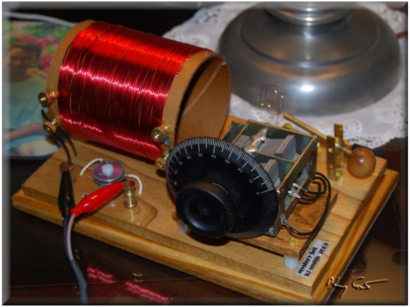

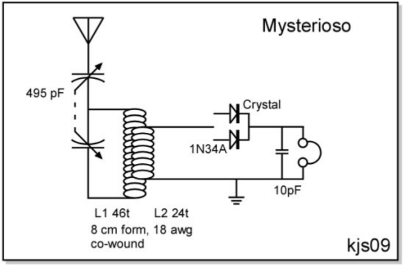

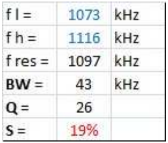

f l =	1073	kHz
f h =	1116	kHz
f res =	1097	kHz
BW =	43	kHz
Q =	26	
S =	19%	

Performance summary for the radio. My Mystery set, despite initial mis-calculations on the coil winding specs, turns out to be a fair performer. the bandwidth around a nominal 1100Khz is 43Khz giving a set Q of 26. This result is superior to both my DeForest and my kit Dunwoody, not bad for a first try! Sensitivity comes in at 19%, second after the DeForest. So, I consider this a successful entry into the world of scratch-built crystal radios.

Figure 4.

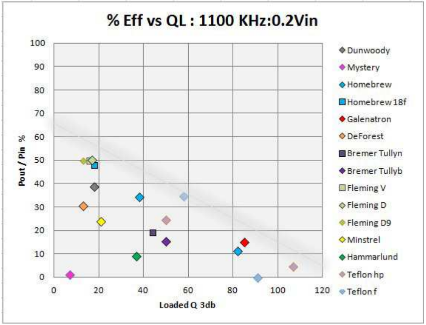

Here is the bottom line, Q verses Sensitivity, efficiency in this case. On this plot it is readily evident that I have built my fair share of loser radios, but also a good number of sets which push boundary of what can be achieved in terms of compromise between Q and sensitivity. Looking at the plot it is fairly easy to imagine a sloping line or zone running from 0 Q at 60% efficiency down to 120-130 Q at 0% efficiency. Beyond this one will probably never go with tank and antenna tuner alone. It will be time to start adding traps to your bag of tricks. Recall that the above is for the 1100khs frequency and that Q is a function of frequency.

In order to explore this final conclusion, I have taken a look at Dave Schmarder's excellent bandwidth analysis of his #50 contest set. This set is a superb work of craftsmanship and skill, experience and usage of the best ($$$) materials available. I take this to be just about the most of what can wrung out of the sky with a crystal radio. Dave’s protocol is rather different from my own and he did not make a specific -3dB bandwidth measurement so what I do from here on takes many liberties with his work, my sincere apologies in advance!

Figure 5.

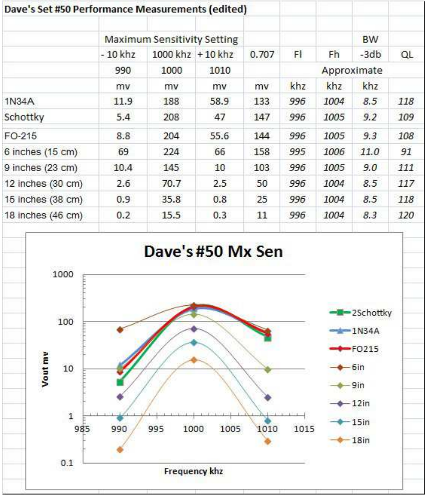

Dave's Set #50 Performance Measurements (edited)

	Maximum Sensitivity Setting						BW	
	- 10 khz	1000 khz	+ 10 khz	0.707	Fl	Fh	-3db	QL
	990	1000	1010		Approximate			
	mv	mv	mv	mv	khz	khz	khz	
1N34A	11.9	188	58.9	133	*996*	*1004*	*8.5*	*118*
Schottky	5.4	208	47	147	*996*	*1005*	*9.2*	*109*
FO-215	8.8	204	55.6	144	*996*	*1005*	*9.3*	*108*
6 inches (15 cm)	69	224	66	158	*995*	*1006*	*11.0*	*91*
9 inches (23 cm)	10.4	145	10	103	*996*	*1005*	*9.0*	*111*
12 inches (30 cm)	2.6	70.7	2.5	50	*996*	*1004*	*8.5*	*117*
15 inches (38 cm)	0.9	35.8	0.8	25	*996*	*1004*	*8.5*	*118*
18 inches (46 cm)	0.2	15.5	0.3	11	*996*	*1004*	*8.3*	*120*

More study, more puzzling over the problem, more calculations, more frustration, more more study.. Did I say that? Short of complete disassembly, which I was loath to do, I was unsure where to go with it. Reading all I could, I came across a set with a modification with a second cap tuning the antenna, this I could test! My antenna, by the way, is a 100ft 14g stranded insulated copper wire with an additional 30ft or so of lead-in. Tuning the antenna seemed promising and when placing a cap the set suddenly found its voice! With a working radio, I had the dilemma of what to do with the case. No second cap would fit. The large vintage capacitor could easily be replaced by two small modern caps though and testing showed that in reality they tuned very effectively with a dual-gang capacitor which is what I settled for in the end. The set still dosnt tun below about 700 Khz or so. Posing the question on the Xtal Set Societie's excellent discussion forum Rap'n Tap brought elightenment from Golfguru who wisely asked the length of my winding, pointing out that the inductance of the coil should be calculated from the total length of the windings, INCLUDING the bifilar primary. Running back to Professor Coyle with a 3.27 inch length gave me 112 uH and a lower tuning range of 660 kH or so. So, "Mystery" solved...

The following is a final schematic of the set and photos of the set as it exists. I listen to it most evenings as I work.

cap, all displayed on a breadboard box with wireing hidden inside.

The risk here was inexperience and it certainly came home! My 18 AWG wire turns per inch were rather different from values assumed in the "Professor Coyle" spreadsheet, both because the sheet assumes bare wire and because my ability to close wind is not that great! My first attempt was for experience, but not for getting a useful coil. Rethinking and adjusting allowed success, almost, on the second attempt and I got 46 out of a planned 48 turns on the main coil. Assembly was difficult but in the end I had a lovely set that wouldn't bring in anything! Well, almost nothing, some weak interfering stations at the cap's full-open position. My first photo is of this incarnation.

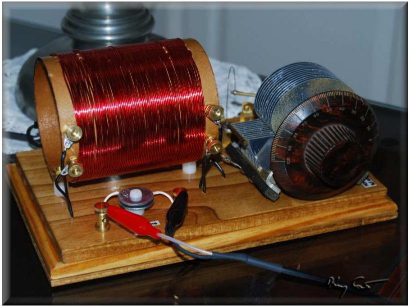

In figure 5 above I have attempted to cast Dave’s measurements into something similar to what I am doing, which is looking at the -3db BW and Q at 1100khz. I finally produced the above spreadsheet where I simply smoothed the data for the 1000khz case and estimated the bandwidth at -3dB. This makes a LOT of assumptions of course. Mostly I assume that the +-10khz measurements are well above the resonance skirt so my smooth curve is not unreasonable.

The #50 set then appears to have, at 1000khz, a -3db bandwidth of around 8-11khz depending on diode and/or spacing between the tank and antenna unit giving loaded Q's of 91-125. These numbers fall in the range of what my own data lead me to expect. Ken Kuhn's page on Resonant Circuit suggests the "typical" QL at 1Mhz ranges between 20 and 100 and Dave’s #50 set is certainly beating the best of that.

In the final figure 6 I have taken my estimated results above and looked at how Q and peak voltage (assuming constant conditions other than the spacing between the antenna and detector units so this ought to reflect power as well). It is quite apparent that the set is rapidly losing output power while Q improvements are flattening out as the spacing between tank and antenna tuner grows to 18 inches. It would appear that a -3db loaded Q of 125 is near the limit of what can be expected in a crystal set.

Figure 6.

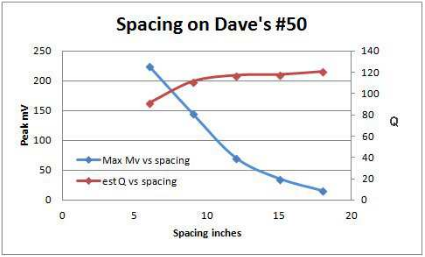

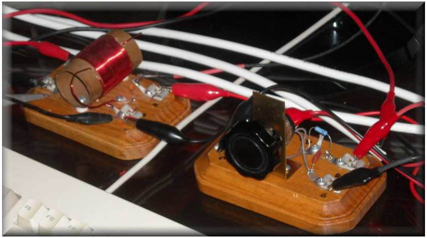

My first homemade, A Mystery Set

http://www.lessmiths.com/~kjsmith/crystal/mystery.shtml

Having made the Dunwoody and experienced the thrill of listening to the local broadcasts on a radio of my own, unpowered save for what the stations themselves provide, I was thoroughly hooked on crystal radio. I hit the web for all I could find, which is considerable. My many thanks to all of you who have provided such useful and informative pages. I categorize my effort along two axes, 1) finding and reading sites with background theory and advice on crystal radio, and 2) sites of crystal radio builders to know what is possible and find inspiration. I was determined to produce a set of my own from scratch. Of all the possibilities, and there are SO MANY, I soon settled on the Mystery Set as an interesting and challenging goal. I have the impression that many, perhaps most of the experienced builders who make their presence felt on the web have built this, or modifications of this set. A worthy place to start, worthy company.

My intent, after studying so many variations of this set was to settle on the original set as published in the 1932 Brisbane Sunday Mail. If I can do well, then only can I consider modifications. My plan was to take the basic circuit, but otherwise use my own forms and wire, calculating appropriate inductor dimentions and winding specific for my needs. I sought a square coil with large dimentions, 8cm mailing tube wound with 18 AWG magnet wire, switchable detector between a diode and homebrew, and a lovely vintage 500pF

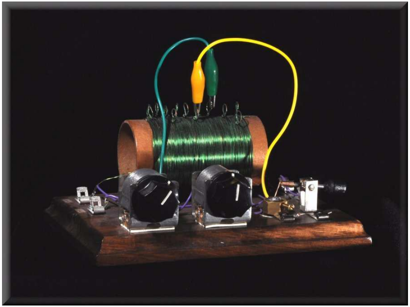

My library of Crystal Radio Hookups and theory...
http://www.lessmiths.com/~kjsmith/crystal/catalog.shtml

This section of my page started out as a place to put my compilation of crystal radio circuits. That compilation became a catalog featuring circuit diagrams hand-drawn by myself from various sources (as indicated on each page, see catalog reference). When I first entered the crystal radio hobby, I was amazed and more than a bit intimidated by the sheer variety of circuit possibilities. I began to make my own circuit drawings in order to get a handle on this in addition to finding interesting circuits to construct myself.

To organize the circuits from fairly easy to rather complicated, I put together a simple classification based on the number of Inductors and Tuning Capacitors in each circuit and a few of their main distinguishing features. Better classifications may exist, but this works well enough for my purpose. A legend follows so anyone may break the code. The catalog provides only minimal information about each circuit. Although an experienced builder may make the calculations to build a radio from the data provided, I mostly intend this for information only. I highly recommend if you wish build a set to consult the original reference.

Later as my experience and familiarity with crystal radio has grown, and after having built and tested a certain number of sets, more detailed and technical information has become necessary. Scouring the web has turfed up an amazing number of extremely well written and helpful web resources. While the number of excellent web sites is large, there have been a certain few that I found myself returning to time and again. From these I would print out large sections and read carefully, scribbling notes on the side, generating more questions and learning, learning, always learning.

Eventually I began to assemble what were, to me, the most useful and informative pages into a handy book format. I decided to print the darned thing out to keep as a sort of Crystal Radio Handbook, something to take with me on the go, or just to read on the couch. I am NOT a "kindle" type.. The project was no trivial thing. I knew what I wanted, but casting everything into a common (more or less) format, organizing things, formatting figures, etc, it was a lot of work. For my handbooks, a glance at the table of contents shows a book organized to lead the reader along the path I followed. The authors are many and I wish to thank 10,000 times all the people who have contributed so much to this hobby on the web. My apologies if I did not choose your pages for the book, the choices of what to include and what to leave out was purely my own. The final Handbooks, there are three now, I feel represent university-level material from advanced introductory to university senior / graduate presentations on Crystal Radio. Published handbooks on crystal radio are traditionally directed primarily to the "boy scientist" adolescent, introductory in nature. I hope this these handbooks fill a gap rather than pander.

Finally, to help as reference, I eventually formatted two particularly interesting web pages as complete volumes in themselves. Ben Tongue's pages are essential reading for the serious hobbiest, even if over my head, (read.. waaaaaaay over my head). At nearly 400 pages, I have broken it into two volumes. Kenneth Khun's web book is just great, technical, but not overly, well explained as one would expect from a professor. Thanks for your efforts and for all those others whose pages I missed.

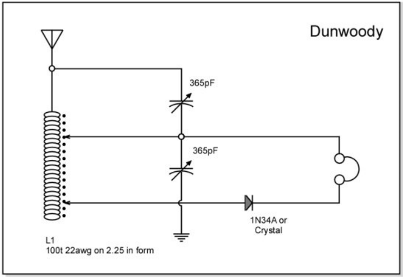

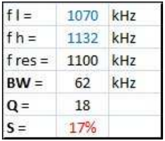

f l =	1070	kHz
f h =	1132	kHz
f res =	1100	kHz
BW =	62	kHz
Q =	18	
S =	17%	

Performance summary of my Dunwoody is so so with a set Q (loaded) of 18 but my sensitivity calculation is 17%, not so great. This may be all that can be wrung out of a single-coil radio, who knows? Most of my radio's have higher sensitivity so this set definitely benefits from my audio amplifier.

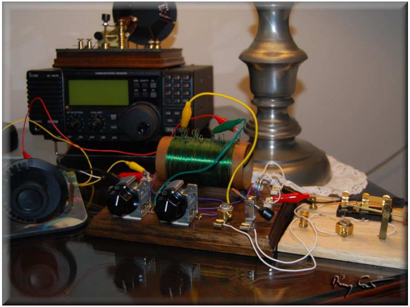

Circuit schematic for the set.

I intend these documents to be printed on double-side letter paper, folded over or cut in half and then bound. The foldbook paging system in the document only makes sense when this is done.

Crystal Radio Handbook: V1 Catalog of Crystal Radio Hook-Ups (ready for printing and binding..129pp)
pdf format, 15.7 meg size

Crystal Radio Handbook: V2 Crystal Radio Theory A Handbook of useful technical articles, (Bookfold format for printing, 262pp).
pdf format, 2.21 meg size

Crystal Radio Handbook: V3 Solid-State and Vacuum Diodes A Handbook of useful technical articles, (Bookfold format for printing, 268pp).
pdf format, 4.47 meg size

Crystal Radio Handbook: V4 Antenna / Ground and Antenna Tuning A Handbook of useful technical articles, (Bookfold format for printing, 164pp).
pdf format, 2.13 meg size

Crystal Radio Handbook: V5 Inductance, Inductors, and Measuring Q Technical look at Coils, (Bookfold format for printing, 272pp).
pdf format, 5.58 meg size

Crystal Radio Handbook: V6 Antennas and Propagation A walk down memory lane, historical and personal, (Bookfold format for printing, 234pp).
pdf format, 9.49 meg size

Crystal Radio Handbook: V7 Vacuum Diodes Dedicated entirely to vacuum diodes, (Bookfold format for printing, 210pp).
pdf format, 11.00 meg size

Crystal Radio Handbook: V8 Resonance Resonance and Coupled Circuits, (Bookfold format for printing, 252pp).
pdf format, 17.70 meg size

Kenneth Khun's Web Book on Crystal Radio Engineering (ready for printing and binding.., 113pp)
pdf format, 1.11 meg size

Ben Tongue's Web Book on Crystal Radio Systems Volume I (ready for printing and binding.., 185pp)
pdf format, 1.37 meg size

Ben Tongue's Web Book on Crystal Radio Systems Volume II (ready for printing and binding.., 208pp)
pdf format, 1.59 meg size

Ramon Vargas Patron's Web Book Contributions to Crystal Radio (ready for printing and binding.., 111pp)
pdf format, 1.65 meg size

This Web Page.. (because I'm so vain.., 238pp)
pdf format, 6.72 meg size

The radio now has a NOS Philmore crystal detector stand. In this photo I am testing a Pyrite crystal sitting on a crumpled aluminum foil bed, successfully detecting 1590 Khz locally here in Houston. In the background is my trusty Icon R75 giving me the straight poop on what station I am listening to. I was never really happy with the Philmore crystal holder though, it had to be improved.

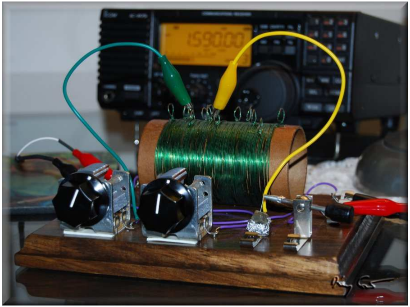

Current configuration on the Dunwoody now sporting a good brass crystal cup with set screw, also I have lifted the coil off the wood base with 1/4" standoffs. I am not sure how much of a difference that made, but seems good practice. This photo of the set taken while prototyping a homebrew detector stand of my own. I figured if I was to become a radio builder, I would need master the detector. I am a geologost so no radio is a crystal radio without a crystal, I mean, an actual crystal!

me back to my roots as a minerologist as well. In my fun I have subsequently learned to pot my own crystal detectors in woods metal, and tested a number of minerals as detectors, though in no systematic way. I did immediately notice that Galena and Pyrite, the most frequently cited minerals as detectors, are metallic sulfides so I have acquired additional samples of Stibnite, Molybdnite, Argentite, and Acanthite, only successful with the Molybinite. Seems a high degree of symmetry is needed as well.

The following photos are of the radio as originally constructed and a couple morphs as I learned to love this great hobby.

My Dunwoody in its original configuration as per the kit instructions. Clearly seen is my first-ever attempt at winding a coil, first attempt at taps in said coil, and a fine radio just the same.

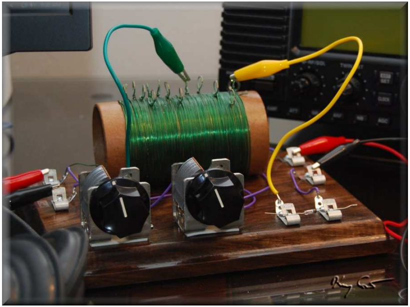

CHAPTER II

Radios I have built with my own little mitts…

Dunwoody, a kit from Borden Radio right here in Houston

http://www.lessmiths.com/~kjsmith/crystal/dunwoody.shtml

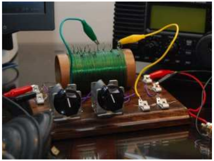

My first real intro to this fascinating world of crystal radio came with a set I purchased on Ebay from Borden Radio Company, Lance Borden. Lance is an engineer at NASA and radio amateur with a business selling crystal radio's and parts. When I first ordered my kit he was very helpful and encouraged me quite a lot. I appreciate the well-designed radio as much as his help in my introduction. The radio itself is a single-tuned set with dual 365 caps tuning on the antenna and detector sides of the single coil. The coil features numerous taps, a concept I was quite unfamiliar with and curious about when I first considered the set for purchase. I figured I had much to learn and the set seems a good and serious way to begin.

The Dunwoody comes with a 1N34A diode detector, two 365 variable capacitors, fahnestock clips, a sturdy cardboard coil form and magnet wire to make the inductor. The included instructions are clear and easy to follow. One item not included or part of the radio is a crystal detector, but the instructions give a hint by stating that the spacing for the germanium detector are the same as for a Philmore detector. That sent me to the web and introduced me to the crystal in crystal radio. I thoroughly had the bug by this time as I searched for, and eventually acquired a Philmore detector and crystal. In fact the galena crystal is remarkably hot and often performs every bit, in my opinion, as well as my germanium diode. This has led

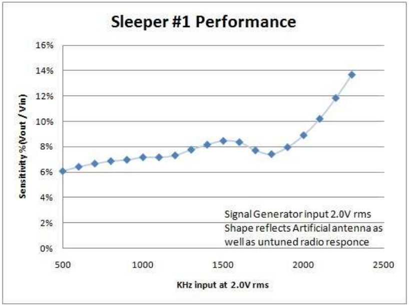

Hooking parallel configuration to my signal generator can give an idea of the performance of the radio across the spectrum. I measured the mV output in 100 KHz increments from 500 to 2300 KHz. Across the broadcast band the set has faitly flat responce with a slight tendancy to resonance about 1500 KHz and then increasing sensitivity from 1900 KHz on up. Keep in mind that the set is hooked to a dummy antenna rather than my actual long wire. Still this gives an idea of the overall untuned nature of the circuit.

When I hook in a series configuraiton I am unable to get a measurement, only 0V DC output. I am working on why this may be the case. Drives me nuts.

A Spark-Gap Transmitter

http://www.lessmiths.com/~kjsmith/crystal/spark.shtml

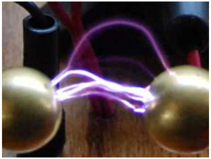

Why a spark-gap transmitter? And why here on a crystal radio page? Well, the second question I'll answer first as it is easiest. The transmitter is a vintage technology device that is no longer current in the modern world and so fits thematically with the purpose of my crystal radio page. The first Question is more interesting, why build it? In late 2009 my elder brother, founder and owner of Northwest Research Engineering, LLC., in Seattle, challenged me to construct a spark-gap transmitter. He enjoyed seeing my vintage-style crystal sets and we enjoyed discussing the early history of radio together. So, it seemed a logical step to progress in this direction. Additionally, as his work is in cutting edge ionospheric research, he has little opportunity to study such retro technologies up close and personal.

> DISCLAIMER:
>
> I am not a licensed radio-amateur and in any case a spark-gap transmitter is not legal to use in the United States. As such, this project was taken on as a technology-demonstration exercise only and is not intended for actual use. Such use is not permitted. Nor is this page intended to promote the construction and use of such transmitters by others. I DO give detail on the construction and theory of the transmitter in order to better understand and appreciate the technologies used at the dawn of the modern wireless era.

The concept behind the project is to build an operating spark-gap transmitter using technology as close as possible to that used at the turn of the 20th century. Technology that would be recognizable to Marconi or Fleming. This meant no chips, no transistors, no diodes, (OK, a galena and cat's whisker is technically a diode, but you know what I mean...) no solid-state. It does not mean using vintage components as these are for the most part unobtainable and/or prohibitively expensive. With each component new or homebrew the question is: would Fleming recognize this? Sometimes I needed to stretch things somewhat, especially concerning the relays, but still I believe I have remained true to this original goal.

The transmitter was conceived as two modules, a power module and an oscillation module. Size constraints forced me to break the oscillation module further into two breadboards, one for the variable HV capacitance and the other for the oscillation transformer. The power module is the most complex piece having on it the induction coil, interrupter, capacitor, and the spark gap itself. To keep this page length manageable and allow enough space to give a good description of each component, its theory and function, I have broken the descriptions into separate links below.

INDUCTION COIL..

COIL INTERRUPTOR..

SPARK GAP..

OSCILLATION CIRCUITS..

PERFORMANCE TESTING..

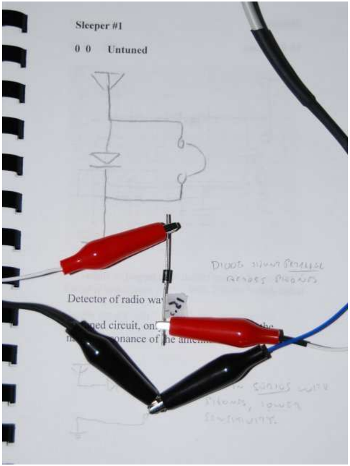
Sleeper #1
0 0 Untuned
Diode shunt parallel across phones
Detector of radio wa
ned circuit, on
nance of the ant
in series with phones, lower sensitivity.

Sleeper #1

The hookup in its parallel and series configuragions. In either case I really cannot hear any signal with just a headphone or crystal earplug. With amplification I can get good signal strength with the parallel hookup giving slightly more sensitivity than the series hookup.

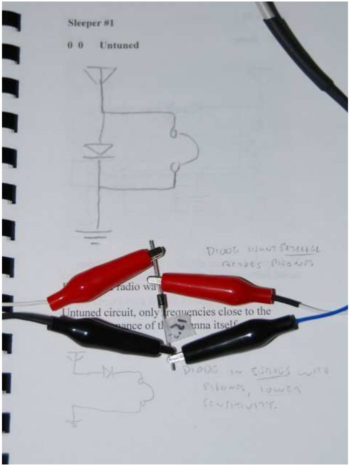
Sleeper #1
0 0 Untuned
DIODE SHUNT PARALLEL ACROSS PHONES
radio wa
Untuned circuit, only frequencies close to the
nance of the antenna itself
DIODE IN SERIES WITH PHONES, LOWER SENSITIVITY.

As shown, the radio passes most loudly a local powerhouse spanish format station. It is quite sensitive as one can imagine. Listening closely and you can certainly hear the other stations on the broadcast band competing for attention. My question now, do I make a breadboard extravaganza of this set? On va voir...

The following circuit diagram and photo shows the complete transmitter and how the modules integrate together:

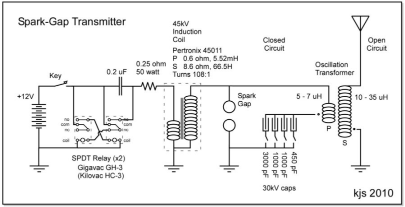
Spark-Gap Transmitter
45kV
Induction
Coil
Pertronix 45011
P 0.6 ohm, 5.52mH
S 8.6 ohm, 66.5H
Turns 108:1
Closed
Circuit
Open
Circuit
Key
0.2 uF
0.25 ohm
50 watt
Oscillation
Transformer
+12V
Spark
Gap
5 - 7 uH
10 - 35 uH
no
com
nc
coil
P
S
SPDT Relay (x2)
Gigavac GH-3
(Kilovac HC-3)
3000 pF
1000 pF
1000 pF
450 pF
30kV caps
kjs 2010

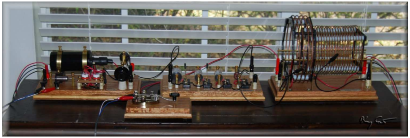

Induction Coil

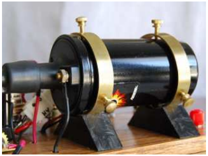

Marconi's first successful radio employed an induction coil as its main energy source. Other methods were quickly adopted that were more efficient and more powerful, but the reliable induction coil remained common throughout most of the early period of radio. Fleming describes a typical "10-inch" coil used for radio work in his day. By 10 inches, he refers to the length of spark obtainable with the coil. These coils are large and today mostly unobtainable except as specialty reproductions at outlandish prices. Fortunately the venerable induction coil remains in manufacture as the common (increasingly uncommon unfortunately) canister ignition coil found in most cars built through the 1970's. In Fleming's terms such a coil might be called a "1-inch" coil, woefully inadequate for transmitting messages to China perhaps, but ideal for this project. Standard coils generate up to about 30,000 volts and "sport" coils can put up to 45,000 volts.

When I first began researching this project I was surprised by the difficulty in finding useful technical specifications for ignition coils of various manufacture, or even of knowing what constituted a "good" versus a "bad" spec when such specs were found. What little information is generally published includes the maximum voltage naturally, often the secondary resistance, sometimes the primary resistance and the turns ratio, (ratio of primary to secondary windings). At first look there doesn’t seem to be much correlation between voltage and turns ratio or

Sleeper #1

http://www.lessmiths.com/~kjsmith/crystal/sleeper.shtml

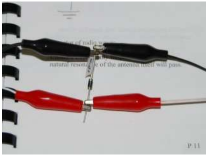

OK, this is a bit of a cheap shot for a radio, but.. it IS a radio!

It is always interesting to get down to bare essentials, to do the absolute minimum and still get a working instrument. It is not for nothing that MB Sleeper included a circuit diagram for this hookup, it is a functionning radio. This circuit is essentially a detector of radio broadcast energy. It has no tuneing functionality. Resonance is extremely broad although weakly peaking as determined by the impedence and capacitance of the antenna itself. In part my idea is to see if it is possible to determine this peak region.
I present the set photo with a 1N5819 Schottky diode but it works well with virtually any diode, only the sensitivity changes. I even tried a Zenner diode but had to attach the Radio Shack mini-amp and crank it up. Still, it worked ok. So, more to come..

even resistances. It took a lot of research to learn the needed theory behind Kettering ignition coils and much of what is found on the web is inadequate, oversimplified, poorly founded theoretically, or downright wrong. I will try here to explain in moderate detail how the Kettering Ignition System works, the theory, and therefore how the various components (windings, primary, secondary, capacitor, resistances, etc) work together. With the right theoretical underpinnings, the specs do make a lot of sense.

Induction coils operate by storing large amounts of energy in the form of strong magnetic fields. As currents pass through the windings their electric fields generate magnetic fields, and fields in one coil induce fields in the other, primary and secondary. As long as a current is maintained in one winding its magnetic field is sustained and a corresponding field is sustained in the second winding. Conversely, moving the coil through a magnetic field or removing the field entirely will generate an electric field, (measured as voltage) in the coil. The ratio of turns between the windings determines the ratio of voltages between the coils. Although the input voltage to the primary coil is small, only 12V for a car battery, it has a large current and a great deal of magnetic energy will be stored in the primary coil. By shutting off the primary voltage source the magnetic field of the primary collapses. As this magnetic field collapses it cuts the secondary coil generating a very high voltage output. The rate of field collapse (dB/dt) determines the strength of the voltage output of the secondary. This high voltage output spikes only when the primary current is shut off causing the field collapse. Coils therefore require an interrupter to repeatedly switch the current to the primary on and off. The trick is to select materials with needed properties to achieve the results wanted. The Kettering Ignition system use in this project is described next.

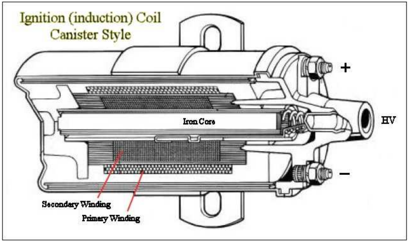
Ignition (induction) Coil
Canister Style
+
HV
Iron Core
−
Secondary Winding
Primary Winding

The Kettering ignition system consists of a 12v battery (the source of emf), an induction coil, a capacitor, and a switch of some sort, (points, relay, vibrator etc..) in a car its points. For the coil the primary and secondary are wound in series and connected together inside the canister. The coil has inductance and impedance/resistance. As the system has LRC, (L = Inductance, R = Resistance, C = Capacitance) the circuit has a resonant frequency and this frequency determines the unit time period "dt" in the equation V = L di/dt. Even without the capacitor there remains some distributed capacitance in the coil and so opening the points will never instantaneously collapse the magnetic field, but it will be extremely fast just the same. The field is maintained by the current from the battery to ground, opening the points removes the ground and stops the current.

As the circuit has a more or less fixed L and R from whatever coil you have available, you can only vary C by choosing the capacitor carefully. For cars, this typically means a capacitance

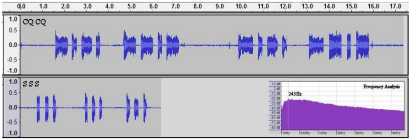
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
13.0
14.0
15.0
16.0
17.0
CQ CQ
SSS
1.0
0.5
0.0
-0.5
-1.0
Frequency Analysis
243Hz

Does it work? What does it sound like on an AM radio? I recorded the above plot off a small AM radio tuned to the top of the broadcast band. Plenty of signal to go around, I understand why these transmitters are no longer allowed to operate! I give you a couple CQ's and an SSS in commemoration of Marconi's first transatlantic radio reception on December 12, 1901. Fleming designed the transmitter power station at Poldhu for the transmission.

CQCQ Yes... it really IS a transmitter..
SSS

73

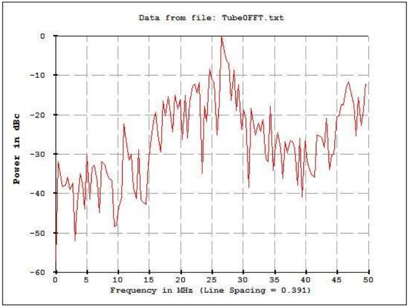

If I can get my arms around the frequency and learn to control this, I see this as an excellent gap for my set. But, would Professor Fleming recognize this? I must fall back to my vacuum relay justification and restate that Fleming, as the inventor of the vacuum tube, was no stranger to placing elements inside a vacuum. He would certainly recognize this.

****

Transmits:

in the 0.1 to 0.2uF range. This capacitance gives a resonant frequency such that the dt will deliver a 30kv jolt to the plugs from the secondary of the coil within the period allowed by the engine timing.

A smaller dt (shorter unit time) delivers a higher secondary V (do the math) but standard coils are not designed for this and the internal insulation will probably not handle the tension. Additionally, the primary also delivers a voltage back to the switch/points/relay. A smaller dt will also result in a higher V off the primary. Of more importance, the smaller dt means that the back emf hits the points early before they have had much time to open. Either the higher V and/or the early arrival will cause arcing at the points. Arcing at the points creates a low resistivity path to ground and leaks the power away from the coil and so the secondary will also not deliver much V to the gap.

Conversely, a dt too long will slow the rate of charging and prevent the coil from being charged sufficiently in the time available (dwell time). The field collapse will be slower lowering the voltage out of the secondary, V = L di/dt. A smaller capacitance means higher frequency/smaller dt and vice versa.

As dt is determined by the resonance of the LRC circuit, the variable you have to work with is generally the choice of capacitor value. Modern high performance coils also factor in. they have lower R (resistance) and lower L (inductance). lower L hurts because the energy stored in the coil is as follows: Energy = 1/2 LI^2. Reducing L lowers the stored energy. lower current (I) also lowers stored energy. Lower R increases the amount of current flow per unit time as well as increasing the frequency of the resonant circuit. Larger di and smaller dt means a much larger V. High performance coils are designed to deliver 45kv instead of 30kv and must have the insulation appropriate to this. The "performance" of these coils comes from two favorable factors, the high voltage assures sparking in high compression and the small dt allows sparking at high RPM's in 8 cylinder engines. That is, it's the small dt that allows sparking at very fast cycle times. They still have roughly 100:1 turns ratio.

Photo of the induction coil / ignition coil mounted horizontally on the power breadboard. Other components seen include the interrupter relays, the capacitor used to determine the coil resonant frequency, and the active spark gap. The thing really works!

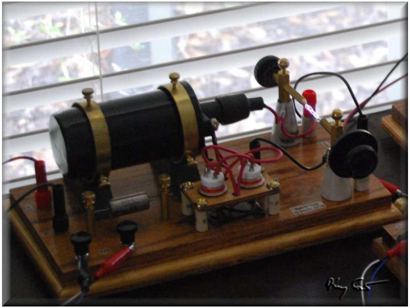

The following are a few easy specs on some popular ignition coils, only a small sampling:

oxidized. This results in an uneven and "messy" spark. More importantly, I suspect that there is a LOT of arcing taking place at the gap as well. Arcing is a continuing current that effectively shorts the closed circuit to ground. This would explain why the audio-frequency coil oscillations are present even when the sparking is present. The vacuum gap prevents oxidation and produces a clean spark with every pulse. The electrodes are also made of either zinc or tungston or some other refractory metal. The lesson to be learned, make your electrodes of something more durable than brass and get some steel wool to FREQUENTLY clean the tips!

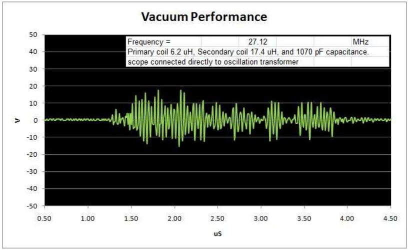

circuits have different wavelengths and so will set in motion a variety of unwanted harmonics. Presumably, with an actual antenna close to the design specs the transmitter will perform flawlessly!

****

The above analyses are presented with the transmitter and gap working in their "normal" operation mode. You will recall in my section on the spark gap that I also have tested a UHF vacuum tube spark gap. The following plots give some early interesting results needing presentation here. The RF oscillogram is quite typical of what this spark produces. You will immediately notice that instead of an RF pulse consisting of several distinct phases as before, this gap produces very clean almost pure HF tones with varying amplitude in the 27MHz range. These two differences, unphased tone and high frequency need explaining. For the difference in frequency I am really at a theoretical loss, all the formula I have tell me that the frequency results from inductance and capacitance, two parameters that did not change. I can only speculate that, as the spark resistance ought to be significantly lower in the vacuum gap, this may be the cause. More testing and more reading needed!

As for the tone itself, again I need to speculate and hypothize. If you look again at the initial audio-frequency oscillogram, you note the obvious induction coil audio oscillations as well as the initial spark. Normally these oscillations should not be seen unless the electrodes are separated far enough to prevent sparking. Initially in my testing this was the case, but with time the coil oscillations became a normal part of my responce. Additionally, with time I have had to adjust the electrodes ever closer to get consistant sparking, thus the frustration. I hypothize that the electrodes themselves are getting quite

PerTronix 45011: V max = 45000 V, P res = 0.6 ohm, S res = 8.6 ohm, P ind = 5.5 mH
Mallory 29217: V max = 58000 V, P res = 1.4 ohm, S res = 10.0 ohm, P ind = 6.6 mH
MSD 8202: V max = 45000 V, P res = 0.7 ohm, S res = 5.2 ohm, P ind = 8.0 mH
Taylor 718203: V max = 45000 V, P res = 0.7 ohm, S res = 4.7 ohm, P ind = 8.0 mH

For my project I have used the PerTronix coil above although I imagine most will do nicely. The Kettering discussion, appropriately placed here in the Induction Coil write-up, is needed understanding in the next section on the Coil Interrupter selection.

Coil Interrupter

In order to function properly an induction coil requires an interrupter circuit. The high-voltage output of the secondary winding results from the magnetic field collapse of the primary when its current source is interrupted. A single interruption will cause a single HV spike and thus a single spark. Repeated sparks require repeated charging and closings of the primary winding. For my project I was seeking a pulse-rate some 60Hz or better to give an audible tone at the receiver. Too rapid though and the induction coil will not have enough time to charge between cycles. Putting things into automotive terms, a 4-cylendar engine idling at 750 rpm must cycle the coil every 40mS, or 25Hz, no problem. For an 8-cylendar engine running at 2500 rpm you are down to 6 mS cycling (170Hz) and many coils are unable to work well at such rates. High performance coils are intended work well even at such high rates. For the transmitter I feel a cycle frequency between 60 and 100Hz would be ideal.

It would be lovely if cycle timing were the only constraint, general purpose relays typically run around 60Hz. Indeed this was my initial design, woe is me. Another concern the interrupter needs to deal with is the presence of a back emf from the induction coil primary winding. With a 100:1 turns ratio, a coil that sends a 45kV pulse out the secondary will also send back to the interrupter a 450V pulse. Small general purpose relay contactors remain rather close together even at their widest separation and so are prone to arcing with modest

Open circuit:
Secondary coil = 17.4 uH (L1)
Distributed inductance = 0 uH (Lo) (no distributed inductance)
Distributed capacitance = 25 pF (Co) (Scope probe in X10 position)
Wavelength = 59.6 sqrt((L1 + La)*Co)
Wavelength = 39m = 7.63 MHz

The two following plots give spectral analyses of the oscillations in the closed and open circuits.

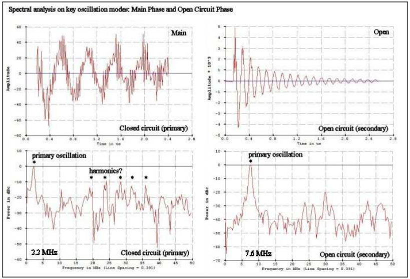

This analysis gives a pretty close prediction to the values actually found on the testing and gives confidance that things are behaving as per theory. Other frequencies noted are harmonics and/or interference between the open and closed circuits. The original design was set so that both open and closed circuits resonated together at 120m. Under test the two

Right about now the observant reader must have noted that the frequencies measured, 2.11 MHz on the closed circuit, and 7.66 MHz on the open circuit, (and variations in-between) do not compare well with the design frequency desired. This certainly needs to be looked at. Recall my original design was for 2.5 MHz on both circuits with the following parameters:

Closed circuit:
Primary coil = 2 uH
Capacitance = 2000 pF
Wavelength = 1884 * sqrt(L*C) = 119m = 2.52 MHz

Open circuit:
Secondary coil = 20 uH (L1)
Distributed inductance = 20 uH (Lo) La=Lo/3 = 6.65 uH
Distributed capacitance = 150 pF (Co)
Wavelength = 59.6 sqrt((L1 + La)*Co) = 121m = 2.48 MHz
Wavelength = 1884 * sqrt(L*C) = 121m = 2.49 MHz

Under testing conditions I have made some significant changes to the above. Specifically, the lumped inductance and capacitance are changed to P = 6.2uH (closed circuit), S = 17.4uH (open circuit), and the capacitance = 1070pF (closed circuit). Also, the design assumed a "generic" antenna with distributed values La = 6.65uH and Co = 154pF. Under test there is no antenna and the scope probe gives about 25pF capacitance and no inductance. Substituting the test values provides the following analysis:

Closed circuit:
Primary coil = 6.2 uH
Capacitance = 1070 pF
Wavelength = 1884 * sqrt(L*C) = 153m = 1.96 MHz

voltages. When I first hooked up my relay I got modest sparking from the gap and contact arcing that was quickly burning up the relay. It was my belief that the capacitor was there to slow the process and allow time for the relay contacts to separate before the back emf hit. I tried capacitors of increasingly high capacitance, they did not cure the problem. It was time to hit books and work the test bench.

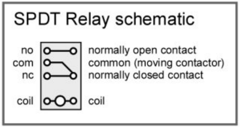

What follows was a better understanding of relays in general, and SPDT relays wired in a buzzer configuration in particular. An SPDT relay has a common contactor set between two posts. The contactor has a spring which keeps it in contact with one of the posts, the "Normally Closed" (NC) position. The other post thus represents the "Normally Open" (NO) position. The relay also contains a small coil that, when energized will push the contactor towards the NO position and so breaking the NC contact. In a buzzer configuration, breaking the NC contact also de-energizes the relay coil so there will be no more push towards the NO post. The contactor will use what little momentum it has to oppose the spring and move to the NO position. In reality this momentum is not sufficient and so no contact with the NO post is ever made. This is why one always uses the NC post with the SPDT buzzer relay. A further negative consequence of this is that the contactor never separates far from the NC post. The close proximity allows breakdown of the dielectric potential of the air between the posts and arcing to occur.

Solutions to this arcing problem come in two flavors, the dielectric and the contactor separation. One solution used in high power relays is to enclose the contactors in a high-dielectric oil to suppress arcing. This is very effective but the fluid is difficult to work with, expensive, and it slows the contact frequency greatly. Recall that I am seeking at least a 60Hz contact frequency. Another way is to place the contactors in a vacuum. While not as good a dielectric as oil, it is significantly better than air and has the advantages of being easy to work with and maintains high frequency operation. Several companies manufacture vacuum relays and, while expensive, this is the route I chose. Note: there is of course a third solution to abandon relays and go solid state. This is fine if all you desire is cool sparks, but it is not in my directive.

Even vacuum relays when wired as a buzzer have the contact separation problem. If the contactors remain very close then arcing can still cause trouble. The method to force the contactor to move to the fully open as well as fully closed position, allowing maximum separation, is to wire two relays in series. The resulting contact frequency (duty cycle) will be half that of a single relay. Happily the vacuum relays operate at a frequency of about 200-240Hz. That provides a 100-120Hz operation in a two-relay configuration, very sweet. Additionally, the two-relay output consists of a nice square-wave signal bringing joy to the professional radio engineer. I know this because the circuit configuration was suggested to me by my clever elder radio engineer brother. Fei chang gan xie! Of course, this solution doubles the expense of an already expensive relay. I have tried the configuration with general purpose relays and find the resulting 30Hz operation unacceptably slow.

---

frequency. High decrements, (in the 1920's a spark-gap transmitter was required by law to keep a decrement of 0.2 or below) indicate a poorly engineered / poorly operated station. Calculations for the wave decrement as follows:

Decrement = 1/n ln (Xo/Xn) where Xo is the amplitude of one wave and Xn is the amplitude of another wave n periods following.

Damping Ratio = 1/sqrt(1+(2pi/D)^2) where pi = 3.14 and D is the decrement from above.

A particularly nice example of an isolated fully-quenched oscillation follows. I assume in this event that enough energy remained present in the closed circuit to allow an occasional isolated single spark to jump the gap. Such a spark produces a nice ring in the open circuit without the complexity associated with the main sparking from the induction coil. This plot shows the open circuit ringing nicely at 7.66 MHz frequency with an acceptable 0.12 decrement.

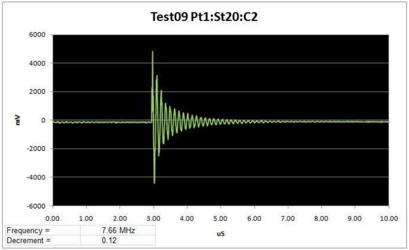

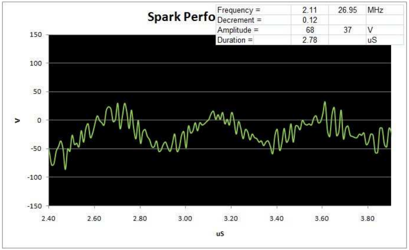

Frequency =	2.11	26.95	MHz
Decrement =	0.12		
Amplitude =	68	37	V
Duration =	2.78		uS

Finally, when the energy loss finally quenches the sparking at the gap, the open, secondary circuit then can freely oscillate, here with a frequency of 10.76 MHz and lasting some 3.3 uS. This last phase is the desired transmission frequency and the open circuit normally oscillates into the antenna. Decrement on this last phase is an excellent 0.04. Early spark gap transmitters struggled to reduce or eliminate the initial two phases of the event with quenched gaps.

Why is decrement important? Most RF damping in a spark-gap transmitter results from interference between oscillations in the open and closed circuits of the oscillation transformer. When these waves interfere two separate wave energy peaks are radiated from the antenna causing loss of signal strength at the desired frequency and adjacent channel interference. A good spark-gap transmitter should quench the closed circuit as quickly as possible after the energy is transferred to the open antenna circuit. Once quenched the open antenna can resonate at its natural frequency. Lacking quenching as with my transmitter, the oscillation transformer needs to be carefully tuned to resonance and loosely coupled to radiate at a single

One final and somewhat late note concerning the vacuum relays. These relays, (Gigavac GH-3 or Kilovac HC-3) are load-switching devices designed to operate to 18 amps. The Pertronix coil on the other hand runs nicely at a cool 20 amps (12 V and 6 ohm, Ohm's Law here). The duty cycle is plenty slow enough to allow a longer charging time so all those amps shouldn't really be needed. I have added a 0.25 ohm ballast to the circuit to lower the current to about 14 ohms. Those cute little relays are very expensive and I have already damanged (temporarily) one of them by pushing 20 amps through. With the ballast I DO find the spark power, measured in terms of brightness, noise level, and max length is diminished somewhat, such a pity.

So, would Fleming recognize this? It is where I push the boundaries as I doubt that vacuum relays existed in his time. Still, relay action was well understood in Fleming's day. As for enclosing them in a vacuum? Well, I defend myself by stating that Fleming himself invented the vacuum diode. Enclosing electrical devices into vacuum bulbs was not a new idea even then. So, while the vacuum relay may not have existed, relays themselves did and Fleming would have been comfortable with the vacuum technology. I think...

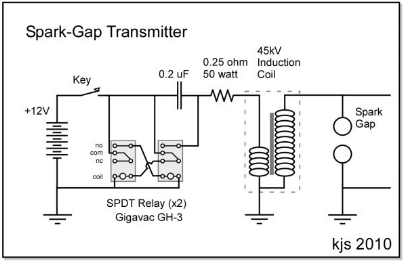

Circuit diagram of the power module indicating the wiring concept for the two-relay solution. And finally, a photo of the set with the two relays in position and ready to go.

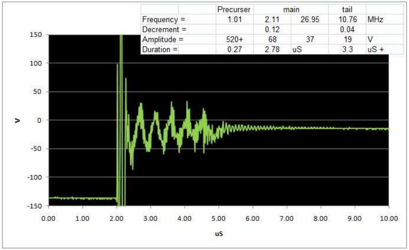

	Precurser	main		tail	
Frequency =	1.01	2.11	26.95	10.76	MHz
Decrement =		0.12		0.04	
Amplitude =	520+	68	37	19	V
Duration =	0.27	2.78	uS	3.3	uS +

Zooming in to the radio frequency world the above plot shows the radio frequency wave in some considerable detail. This plot indicates that the oscillations given off from a spark gap are anything but pretty. The complexity results from the ringing going through several phases with the closed and open circuits, and induction coil each having their imprint. The initial (precursor) phase of the event, lasting here about 0.27 uS is the high-amplitude energy (upwards of 520+ V) of the spark as the induction coil pours its energy directly into the circuit at the gap. This is followed by the main phase, a complex 68 V amplitude ringing lasting some 2.8 uS. The ringing consists of two combined waveforms, one, a high amplitude (68 V) 2.11 MHz wave with a second lower amplitude (37 V) high frequency 27 MHz wave superimposed. This phase results while the closed (primary) circuit still retains sufficient energy to cause sparks at the gap with each main oscillation. The higher frequency wave comes from the open circuit (secondary) inductively coupling to the closed. A detail look at the main phase seen below:

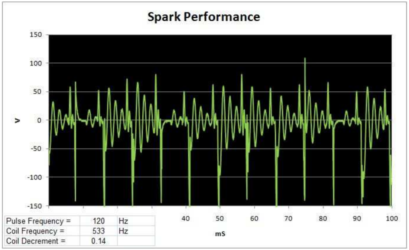

This first plot above shows the response at audio frequency. In the 100 mS plotted we see 11 pulses at 120 Hz frequency. This represents the pulse frequency of the dual relay system, two relays at a nominal 240 Hz each. Each pulse starts with a high amplitude spike associated with the spark discharge and followed by two to three damped oscillations with a frequency of 533 Hz. This oscillation is associated with the natural frequency of the induction coil itself. Damping of waves is a natural oscillation feature and is refered to as the wave Decrement. Decrement is the natural log (Log to the base e) of the amplitudes of any two successive peaks in the wave train. It results from the loss of energy in the circuit during the oscillation period. Energy is lost due to resistive heating, radiation of waved into space, and interference between oscillations in the primary and secondary oscillation circuits, especially where they are closely coupled. Where resonance is high the interference losses are minimized.

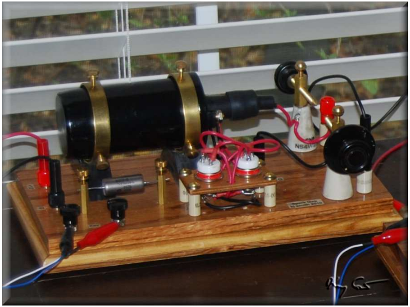

Spark Gap

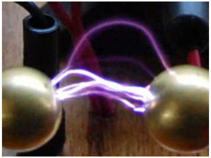

The heart of the system is the probably the simplest component, the spark gap itself. After all the required background theory for the induction coil and relay setup, this component was a walk in the park. Still I need note that my gap is not intended for serious duty and so does not require extra features such a cooling fins and/or etc. Below is an early mock-up of the gap, a simple set of 1/2" brass balls tapped for a 1/4" threaded brass support rod. The entire assembly set off the base with lovely tapered ceramic standoffs.

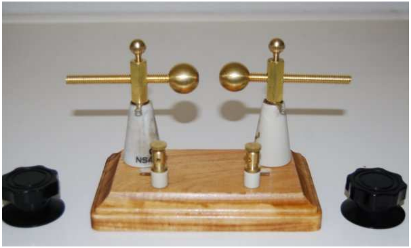

Early use of the spark gap, especially with a weak induction coil, showed up room for improvement. When sparking the

---

Performance Testing the Transmitter

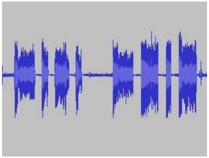

If you build it, will it work? And how well? This page attempts in a preliminary manner to answer these questions and provide a first look at the performance through the eyes of an oscilloscope. It will take much testing and observation/head scratching before I ferret out all the phenomena present in the set. This is one of the objectives of the project, to study and learn about spark-gap technology. The plots below are a first look, some will raise more questions than answers, but that’s the fun!

All of the scope plots presented here are made with a single hookup to the scope. The scope probe was attached to the set antenna output of the oscillation secondary coil. Many other observation points remain to be made and observed and I will update this page accordingly when needed. I present observations of the three primary oscillation frequencies expected from the set: A) the pulse frequency from the relay, B) the resonant frequency of the induction coil + power capacitor, and C) radio frequency oscillations of the transmitter itself. So little work so far, so many questions remain!

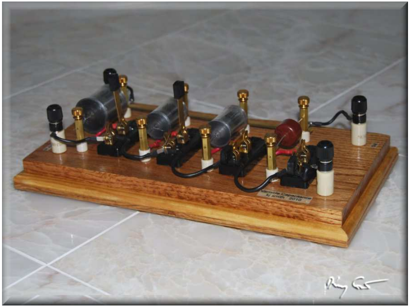

The capacitance module simply consists of high-tension modern capacitors switchable into or out of the circuit in parallel. The original intention, still in play, is to use Leyden Jars, but that is another matter. The four capacitors have values of 500, 1000, 1000, and 3000 pF giving plenty of combinations to choose from. Changing inductance is a matter of clipping to the coil at the right place. Finally, coupling the coils is a matter of sliding the secondary along rails into or out of the primary winding. The following photos give portraits of the two oscillation circuit modules.

When one walks into a room, nothing quite says "Transmitter present!" like a big oscillation transformer!

---

brass balls build up an oxidation layer fairly quickly. This layer increases the dielectric between them and, especially when the voltage is marginal, (my first coil was old) the sparks had trouble to jump the gap. I needed to clean the balls frequently with steel wool. The 1/2" balls spread out the field flux over a small area. One improvement was to switch from a ball-gap to a gap between sharply pointed tips. The photos below show both configurations. The points concentrate the field and creates a higher potential. This makes it easier for the sparks to form. Switching between configurations is as easy as screwing on or off the balls which cover the pointed tips.

Adjustment of the gap length is easily accomplished by turning the end knobs of the threaded supports. Small top thumb-screws lock the posts at the desired position. CAUTION: Do NOT touch any metal components while the spark gap is in operation. Handle by the insulated knobs only.

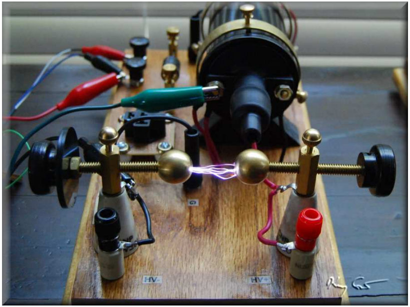

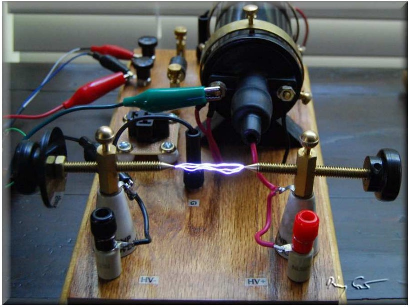

In the photos above the sparks are being produced without any ballast resistance and are just over 2cm long. With the ballast I subsequently added to protect the vacuum relays, the spark power is lowered from 20 amps down to about 14 amps. The sparks obtained are smaller, about 1 to 1.5 cm. According to Fleming an induction coil capable of producing sparks up to 10 inches (25cm) in air will give a spark of 6-7 mm when connected to a load. I find with the tank circuit connected to the gap I need adjust the spark points very carefully and close, about half a mm or less to obtain consistant sparks. This can be a frustrating exercise. Air diaelectrics also have the problem of arcing when the air between the electrodes begins to ionize. A solution to this is to enclose the spark gap in a vacuum. It just so happens that occasional inexpensive vacuum spark gaps are listed on ebay. These tubes are intended for radar (UHF) applications and are rated in the few hundred volts range. As

The oscillation transformer design thus becomes a matter to choose coil lengths and diameters to give a range of values inclusive of Lsecondary = 20uH and Lprimary = 2uH. To make the transformer loosely coupled the secondary needs to slide within the primary. I chose the following parameters:
Primary: Diameter = 7", Length = 2.0", Turns = 4, uH = 3.81 max, 1/8" copper tubing
Secondary: Diameter = 5", Length = 6.0", Turns = 22, uH = 27.4 max, 1/16" copper tubing

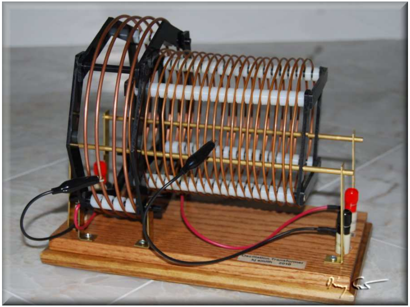

to imagine an amateur inverted L antenna with about 20 uH distributed inductance and 150 pF distributed capacitance. As antennas are magical mystical things that nobody (meaning myself) really understands this appears entirely reasonable. In any case, the transmitter is tunable to bring it to proper resonance (hopefully) when attached to an actual antenna. These then are my fundamental design parameters:

Design frequency = 120m
Antenna Co = 150 pF
Antenna Lo = 20 uH (20000cms)
La = Lo/3 = 6.67 uH (6667cms)

L1 = Loading, or secondary coil = ?

From Bucher:
Wavelength = 59.6 sqrt((L1+La)*Co)

Assuming a design frequency of 120m and solving the above equation for L1, I find:
L1 = 20 uH (20000cms)
20uH then is the value we wish for the oscillation transformer secondary winding.

For the primary winding we find in Bucher the following useful equations:
L = lambda^2 / 3550000 * C, and/or
C = lambda^2 / 3550000 * L

Therefore one starts with lambda = 120m and then either assume a capacitance or an inductance. For this project I found a good match with 2000 pF and 2 uH. To be safe both inductance and capacitance can be varied around these two design numbers.

my gap is tiny now, I felt this may be a good solution, or, alternatively I might have some fun blowing up a cool vacuum gap. The following photo shows that in fact the vacuum gap performs handsomely as a gap for my transmitter. The gap length is mercifully fixed at about a half mm and the low diaelectric allows good sparking with flat (rather than pointed) spark surfaces about 2mm in diameter. I certainly will experiment more with this kind of gap.

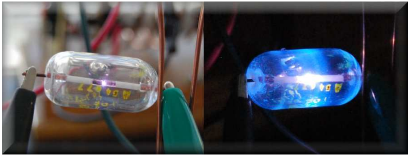

Thats'a some glow...!

****

Other Sparks:

Of note, the gap itself need not remain strictly a set of pointed, flat, and/or round surfaces. Once you have a working High Voltage power supply, there are other gaps/toys waiting to be tried.
The photo below is a quick cobbled together Jacob's Ladder just to see: Will it work? The copper wire is not perfectly smooth nor can one adjust the spacing between the two verticals. Still, I can tell this will be a fun side trip with my HV power supply. Stay tuned for more.

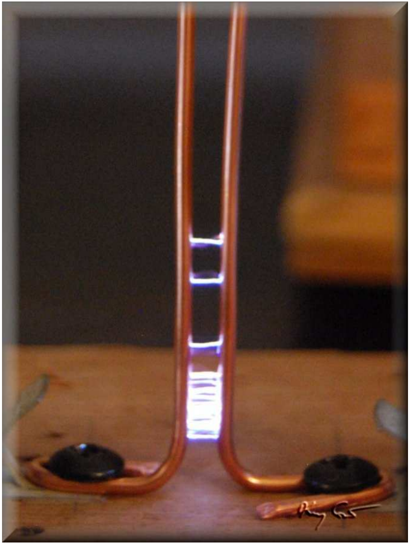

Schazzam!!

Oscillation Circuits: HV Capacitance and Transformer

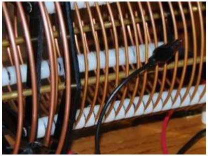

The oscillation circuit in the transmitter performs the twin functions of converting the spark's energy to trains of radio frequency oscillations and transferring them from the closed tank to the open antenna for transmission. Like any tank circuit the main components consist of capacitance and inductance, both variable in order to tune the circuit to the needed frequency. The closed circuit loosely couples to the open circuit to achieve the best resonance and lowest decrement. The design aspects of this section is where the radio theory begins and induction/relay theory ends, we are talking radio here now.

The design for the transmitter starts with a notional target transmission frequency and the design equations from two primary sources: Fleming, 1910, The Principals of Electric Wave Telegraphy and Telephony., and Bucher, 1920, The Wireless Experimenters Manual. Additionally it requires some estimate, guesstimate really, of the antenna distributed capacitance and inductance.

As the project is demonstration only, I have selected a frequency at the low end of the HF range just above the MW broadcast bands. The designs in the 1920's typically aimed at about 200m or 1.5 MHz. To keep my design close I chose 120m, or about 2.5MHz. It is a fairly quiet part of the spectrum. Keep in mind this is a notional project not intended for actual broadcast. I used charts and data presented in Bucher

KEVIN’S WEBSURFER HANDBOOK I
FOR CRYSTAL RADIO

A CATALOG OF CRYSTAL HOOK-UPS

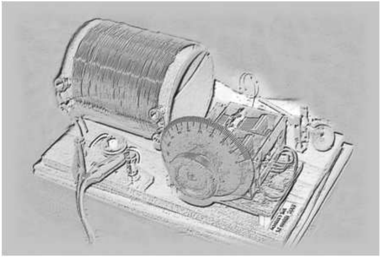

Kevin Smith
2009
rev 2015

Printing / Binding Instructions

1. Choose “fit to page” in print menu
2. Print document double sided on letter size paper
3. Cut the entire printed document in half
4. Fold over making sure the page numbering is continuous
5. For the cover: Print just the first page on card stock paper
   Cut the cover in half as well
6. Assemble the covers on the document
7. Punch the left side for a binding, spiral or comb as desired

http://www.lessmiths.com/~kjsmith/crystal/catalog.shtml

KJ Smith

Notes:

Notes:

Hookup	**Page**	**coil**	**cap**	**features**
Contents	iii			
Intro Note	vii			
Legend	viii			
Chapter 1 Broadcast				
Sleeper #1	2	0	0	Untuned
Sleeper #2	3	1 S	0	Basic
American Radio Stores	4	1 B	0	TC
Sleeper #9	5	1 S	0	TC
MRL #12	6	1 S	0	2SL
Wagner: XSS v10, n4	7	1 S	0	Veractor
Sleeper #5	8	1 S	1	Series
Sleeper #6	9	1 S	1	Parallel
Sleeper #7	10	1 S	1	SP switched
MRL #1	11	1 S	1	LWT
MRL #17	12	1 S	1	TC
Steinite	13	1 W	1	TC
Dunwoody	14	1 S	2	TC
Babani LD	15	1 S	2	TC
MRL #13	16	1 S	2	SL LWT
Schmarder #30	17	1 L	3	Loop RFC
Schmarder #43	18	1 W	2	LD LD
MRL #2	19	1 S	2 G	LWT
KJS Fleming	20	1 S	2	Valve
BBCS #6	21	2 S	2	TC LWT
Sleeper #11	22	2 S	0	LD SL
MRL #11	23	2 S	1	TC
BBC Comm	24	2 W	0	VM
Sleeper #12	25	2 S	0	VM
Sleeper #14	26	2 S	1	VM
MRL #22	27	2 S	1	VC TC
Sleeper #15	28	3 S	0	VM TC TC
MRL #29	29	3 S	1	VC TC

Hookup	**Page**	**coil**	**cap**	**features**
Schmarder #11	30	3 S	2	TC VM
MRL #15	31	3 S	2	VC
Sleeper #18 DeForest	32	2 S	1	VC TC
MRL #35	33	3 S	3	TC VC
Secor, 1920	34	4 S	0	VM
StayTuned #76	35	3 W	0	LC TC
KJS Bremer-Tully	36	2 D	1	LC
MRL #10	37	2 L	1	TC
DP Galenatron	38	3 S	2	VC TC
XSS #59	39	3 D	2	LD TC
MRL #24	40	3 D	2	LD TC
MRL #43	41	2 D	2	Counterwound
McCall	42	2 S	1	VC
MRL #3	43	2 S	2	VC TC
MRL #30	44	3 D	2 G	LC LWT
Sleeper #16	45	2 S	0	LC SL
Sleeper #17	46	2 S	1	LC SL
KJS Teflon	47	2 S	3G	LC
Tuggle L17	48	2 W	3 G	LC TG IWT
Tuggle ML17	49	2 W	3 G	LC Mossfet
Poole #5	50	2 S	2	TC LC
XSS #58	51	2 S	2	TC LC
Polk Symmetry	52	2 S	3 G	TG 4SL
Poole off the shelf	53	2 S	3 G	TC LC
KJS Hammarlund	54	2 S	3 G	TG SEC
XSS #54	55	3 S	3	LC WT
KJS ID #3	56	2 S	3 G	TG TC
Dejan	57	2 S	4 G	LC TG IWT
M Hampton, BRG 2009	58	2 S	5 G	TG LC
W Thelen, BRG 2006	59	2 W	5	TG LC
Sleeper #19	60	4 S	1	LC SL LD
Kinzie 47b	61	3 S	3	LD LC
RTVE p129	62	2 SW	2	FW

XSS	www.midnightscience.com/index.html Crystal Set Society, (M Peebles) Newsletter #

References:

BBCS	Boys Book of Crystal Sets W J May 1954
Babani	Crystal Set Construction BB Babani,
BCRG	Birmingham Crystal Radio Group http://www.crystalradio.us/
GE Ham	GE Ham News, V10, n6
Gollum	home.snafu.de/wumpus/gollum/
Kinzie	in XSS newsletter #47
Klase	www.skywaves.ar88.net/xtal/xtal.htm
MRL	Modern Radio Labs HB 17, 25 Elmer G Osterhoudt
Poole	bellsouthpwp2.net/w/u/wuggy/
Schmarder	www.schmarder.com/radios/crystal
Sleeper	Radio Hook-Ups Milton B Sleeper 1920
Solomon	Solomonmusic.net/Radios.html
Stay Tuned	www.crystalradio.net/crystalplans

Hookup	Page	coil	cap	features
Schmarder #1	63	2 W	4	
Boursin 1939	64	2 S	1	MW LW
Jacquemard	65	3 S	2	MW LW
Proton Mystery	66	2 C	1	TF
MRL #8	67	2 L	1	TF CK
Peebles XSS #9	68	2 L	1	TF
KJS Mystery	69	2 C	2 G	TF TG
Tuggle Mystery	70	2 C	2 G	TF TG
Solomon Double G	71	2 C	2 G	TF
MRL #39	72	2 L	2 G	TF TG
Gollum Mystery	73	2 C	4 G	TF
Proton Mystery Plus	74	3 C	1	TF
MRL #4	75	3 L	1	TF
Solomon Mystery Plus	76	3 C	2	TF
Solomon Crazy Plus	77	3 C	2 G	TF
MRL #23	78	3 L	2 G	TF LWT
Solomon Almighty	79	3 C	3 G	TF
GE Ham News	80	2 L	4 G	FW DA
Pop Mechanics	81	3 M	3 G	FW
XSS #62	82	5 M	3 G	FW TD
Sleeper #21	83	2 S	3 G	SL CT
BBCS #11	84	2 S	3 G	TC
Coupled Trap	85	4 S	2	IWT
Sleeper #20	86	4 S	2	LK
Hogan 1922	87	4 S	2	LK TC
XSS #63	88	4 S	2 G	LK VC
Ford, Pop M 1945	89	4 S	3 G	LK
Radio Craft ST#31	90	4 S	3 G	Band select
KJS Minstrel Boy	91	5 T	3	LK
Petersen XSS #107-8	92	5 S	4 G	LK CK
Barket, RW 1924	93	8 W	2	LD LC

Hookup	**Page**	**coil**	**cap**	**features**	
Chapter 2 Shortwave					
KJS Global DX	96	1 W	2	TC	
Poole SW	97	2 S	1	LC	
XSS #56	98	2 S	2	TC	
H Lash SW	99	2 S	4 G	LC	
AKlase SW	100	3 S	3 G	LC	LD
Harthun SW Mystery	101	3 D	1	TF	SW
Chapter 3 Misc					
Wave traps	104		Circuit Additions		
Benny & SEC	105		Circuit Additions		
Detector Bias Circuits	106		Circuit Additions		
Matching Transformer	107		Circuit Additions		
Transistor detectors	108		Circuit Additions		
Phone Amplification	109		Circuit Additions		
Dummy Antenna	110		Radio Measurement		
Performance Test	111		Radio Measurement		
Diode Test	112		Radio Measurement		
KJS Spark Gap	113		Transmitter		
Marconi 1896	114		Coherer detector		
Lodge 1898	115		Coherer detector		
Tesla 1899	116		Coherer detector		
Marconi 1900	117		Coherer detector		
DeForest Smythe 1902	118		Coherer detector		
Anderson XSS #53	119		Heterodyne		
Fleming Valve	120		Vacuum Diode		
Pratt 1915 IRE v2 n2	121		Crystal+Audion		
Audion Valve	122		Armstrong Circuit		
Armstrong Regen	123		Dec 1914		
AF amplifier	124		LM 386		
References	125				
Notes	126				

AF Audio Amplifier

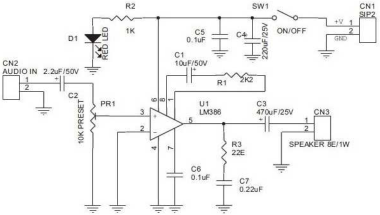
R2
1K
SW1
ON/OFF
CN1
SIP2
+V
1
2
GND
D1
RED LED
C5
0.1uF
C4
220uF/25V
C1
10uF/50V
R1
2K2
CN2
AUDIO IN
1
2
2.2uF/50V
C2
PR1
10K PRESET
3
2
6
8
1
4
7
5
U1
LM386
C3
470uF/25V
CN3
1
2
SPEAKER 8E/1W
R3
22E
C6
0.1uF
C7
0.22uF

C1	10 uF
C2	2.2 uF
C3	470 uF
C4	220 uF
C5-6	0.1 uF
C7	0.22 uF
D1	LED 3mm
R1	2.2 k Ohm (short R1 to get max gain = 200)
R2	1k Ohm
R3	22 E Ohm
R4	10 k Ohm variable
SW	switch
IC	LM 386

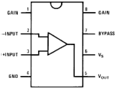
GAIN
1
–INPUT
2
+INPUT
3
GND
4
8
GAIN
7
BYPASS
6
V_S
5
V_{OUT}

Introduction:

The following catalog features circuit diagrams hand-drafted by myself from various sources (as indicated on each page, see catalog reference). When I first entered the crystal radio hobby, I was amazed and more than a bit intimidated by the sheer variety of circuit possibilities. I began to make my own circuit drawings in order to get a handle on this in addition to finding interesting circuits to construct myself.

To organize the circuits from fairly easy to rather complicated, I put together a simple classification based on the number of Inductors and Tuning Capacitors in each circuit and a few of their main distinguishing features, but this does not always feel right. Sets are arranged broadly by groups, single-tuned, variometers, double-tuned, etc. Better classifications may exist, but this works well enough for my purpose. A legend follows so anyone may break the code.

I provide only minimal information about each circuit. Although an experienced builder may make the calculations to build a radio from the data provided, I mostly intend this for information only. I highly recommend if you wish build a set to consult the original reference.

Please note that this compilation represents only a small fraction of the circuit possibilities in the world of crystal radio. Competition sets in particular can get very complex. I have tried to find a wide range of styles and ideas with examples from many circuit families. The reader is encouraged to pursue more on the web or via whatever sources are available.

Kevin Smith

Legend:

Coil

1, 2, 3+ number of coils
S = single winding selenoid
W = spiderweb, other non-selenoid
D = dual windings on one form
M = multiple windings on one form
C = co-wound, bifiliar
L = layered windings

Capacitor

1, 2, 3+ number of capacitors
G = ganged

Features

VM = variometer
VC = variocoupler
LC = loose coupler
FW = full wave receiver
LD = loading coils
CK = rf choke
CT = capacitively tuned
TG = tuggle tuning
TF = telefunken variation
SL = slider tuned
TC = tapped coils
LK = link coupling
LWT = linked wave trap
IWT = wave trap inductively coupled
MW = Medium Wave
LW = Long Wave
SEC = Selectivity Enhancement Circuit
B = Benny circuit

Armstrong Regenerative 12-1914

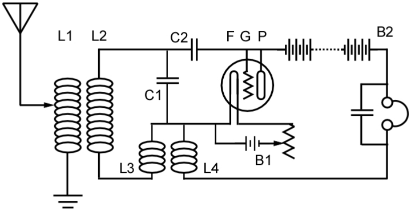

L1-2	coupled tuning circuit	
L3-4	feedback circuit (transformer)	
C1	tuning capacitor	
C2	circuit choke	
B1	A-battery	filament current
B2	B-battery	plate potential
Single triode vacuum (Audion) tube		
F	Filament	supplies electron current
G	Grid	control space charge around filament
P	Plate (Wing)	rectifier / oscillator

Armstrong Circuit
Audion Valve

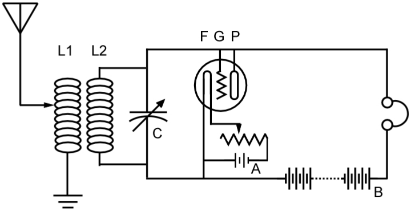

Single Audion tube rectifies signal and provides amplification.

L1-2	Coupling coils
C	Tuning capacitor
A	Filament current
B	Supply of plate potential

Vacuum Tube elements

F	Filament	supplies electron current
G	Grid	control space charge around filament
P	Plate (Wing)	rectifier / oscillator

Unit prefix’s

m	milli	10^-3
u	micro	10^-6
n	nano	10^-9
p	pico	10^-12

Old radio text notation
1 cm = 1 nH
1 mmf = 1 pF

Poole’s Quick Fix’s
Add a loading coil for short antennas
Add series capacitance for long antennas
Add coupling coil in front of tank
Tap antenna down on coil
Tap diode down on tank coil

To consider:
Add a Benny, even without matching transformer
Add SEC to unload diode
Add Bias to the detector

Pratt 1915   IRE v3, n2

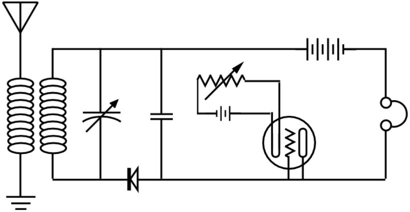

Combined crystal detector and Audion amplifier circuit to test amplification on distant signals.

Fleming Valve

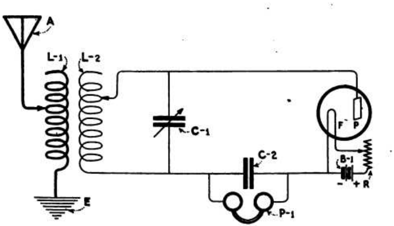

Figure 20a
Simple circuit for the two-electrode vacuum tube as an oscillation detector in radio telegraphy.

Rectification of signal with diode tube

L1-2	Coupled coils
C1	Tuning capacitor
C2	Shunt capacitor
R	Variable resistor to adjust temperature on filament
B1	Battery for heater (filament current)
V	Vacuum diode (Fleming Valve) rectifier

P35: 1918_Bucher Vacuum tubes in wireless communication

CHAPTER 1

BROADCAST BAND HOOK-UPS

Sleeper #1

0 0 Untuned

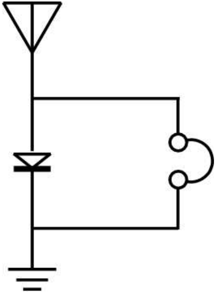

Detector of radio waves

Untuned circuit, all frequencies, especially those close to the natural resonance of the antenna itself will pass.

Diode shown in parallel with the phoned, may also be placed in series, but will have lower sensitivity.

Anderson XSS #53

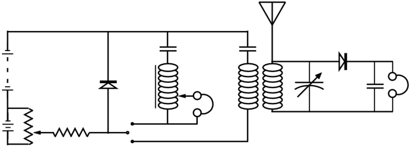

Heterodyne Circuit

DeForest – Smythe 1902

“Receptor”

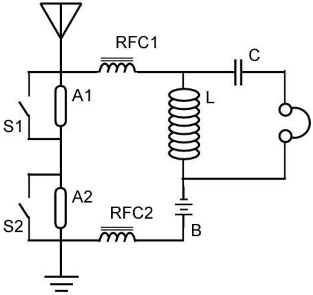

A1-2	Anti-Coherers
RFC1-2	Choke coils
C	Tuning condenser
L	Inductor
B	Battery
S1-2	switches

Sleeper #2

1S 0 Basic

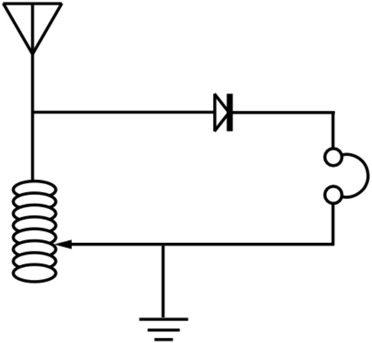

Variable inductance circuit with tapped or slider-tuned coil.

American Radio Stores Inc.

1B 0 TC

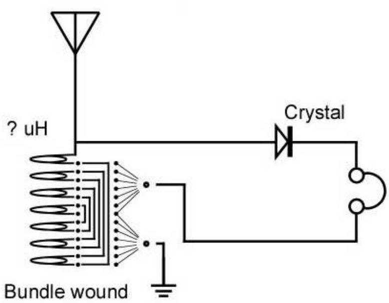

Variable inductance circuit with 2 sets of taps, Units and Multiples

The fascination of this radio is its tapped, bundle-wound coil. Taps are formed to wind into the main bundle.

Marconi 1900

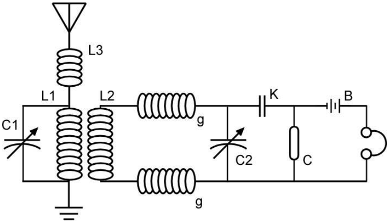

L1-2	Coupling inductors
L3	Ant loading coil
g	Loading inductor
C1-2	Tuning condensers
C	Coherer
K	“Stopping“ condenser
B	Battery

Coherer and telephone circuits as per Lodge Patent

Tesla 1899

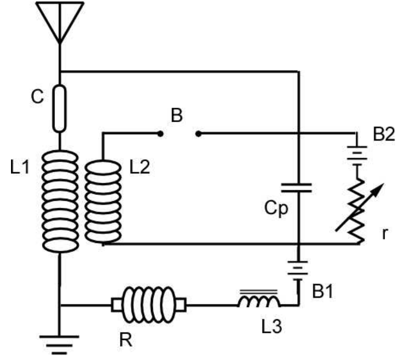

L1	Secondary inductor
L2	Primary inductor
L3	RF Choke
Cp	Condenser 500 pF
C	Coherer
B	Interrupter “Break”
B1-2	Batteries
r	resistor
R	Relay (sounder or headphones)

Sleeper #9

1S 0 TC

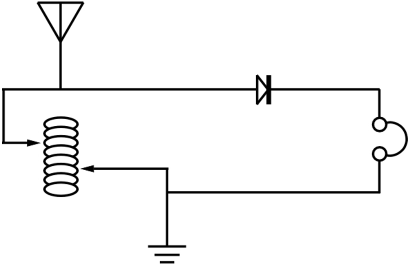

Variable inductance circuit with double switch to taps on coil, coarse and fine spacing switches separate.

MRL #12

1S 0 **2SL**

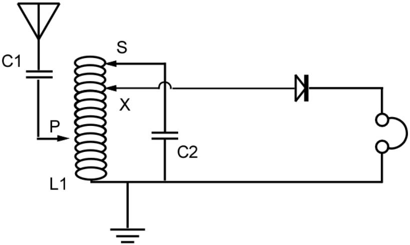

Form	2” x 4 ½”	
L1	120t,	22 awg

C1-2	Fixed mica, 150 pF
P	primary slider
S	secondary slider
X	crystal slider

Lodge 1898

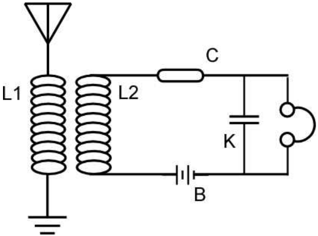

L1	Variable antenna inductor
L2	Coupling coil to detector circuit
C	Coherer
B	Battery, current triggered through coherer
K	Condenser of large capacity

This detects radio energy impulses, damped waves ala Morse.
Secondary coil aperiodic, tuning un-necessary.
First receiving circuit to place detector circuit away from antenna circuit thus reducing load resistance and allowing antenna to be tuned.

Marconi 1896

Untuned receiver

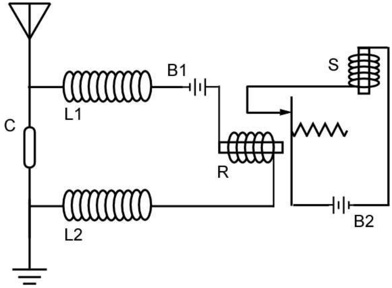

L1-2	Choke coils
C	Coherer
B1-2	Batteries
R	Relay
S	Morse sounder

Coherer tapper not shown

Wagner: XSS v10, n4

1S 0 Veractor

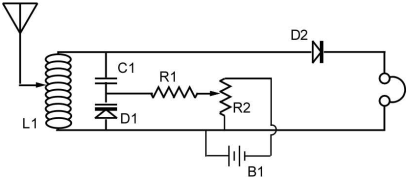

L1	260 uH, tapped
C1	1000 – 5000 pF
D1	MV1662 veractor
D2	1N34A
R1	100k ohm
R2	5-10Meg Pot
B1	9 – 15V DC

Tuning of set via variable capacitance diode (veractor) rather than traditional Vcap.
Crystal Set Sociery, 2000, v10, n4.

Sleeper #5

1S 1 Series

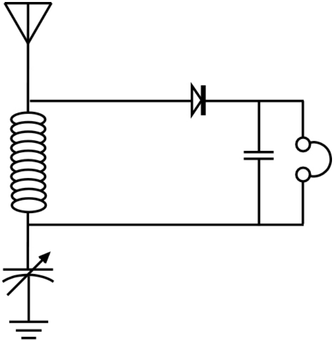

Series tuning capacitor

The lower the capacity used, the shorter the wavelength, (higher frequency) tuned.
Adding series capacitance lowers overall capacitance and thus “shortens” the antenna.

KJS Spark Gap Transmitter

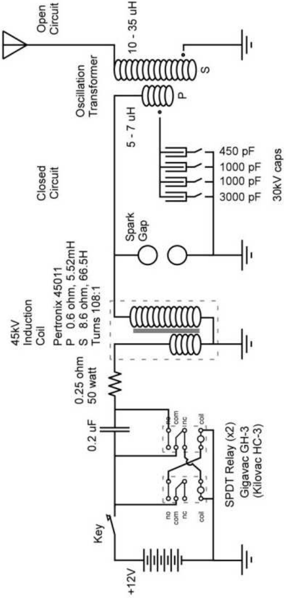
Open Circuit
10 - 35 uH
Oscillation Transformer
S
P
5 - 7 uH
450 pF
1000 pF
1000 pF
3000 pF
30kV caps
Closed Circuit
Spark Gap
45kV Induction Coil
Pertronix 45011
P 0.6 ohm, 5.52mH
S 8.6 ohm, 66.5H
Turns 108:1
0.25 ohm 50 watt
0.2 uF
no
com
nc
coil
no
com
nc
coil
SPDT Relay (x2)
Gigavac GH-3
(Kilovac HC-3)
Key
+12V

Diode Test

Petersen (modified)

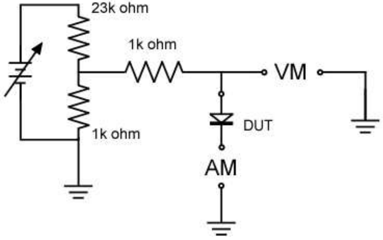

R1-2	1k ohm
R3	23k ohm
B	Variable power supply (18 – 24V max)
VM	Volt meter (digital)
AM	Amp meter (digital)
DUT	Diode Under Test

Sleeper #6

1S 1 Parallel

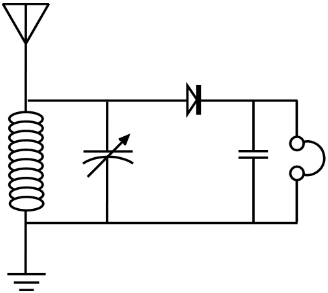

Parallel tuning capacitor

The lower the capacity used, the greater the wavelength, (lower frequency) tuned.
Adding capacitance in parallel increases overall capacitance thus “lengthening” the antenna.

Sleeper #7

1S 1 SP switched

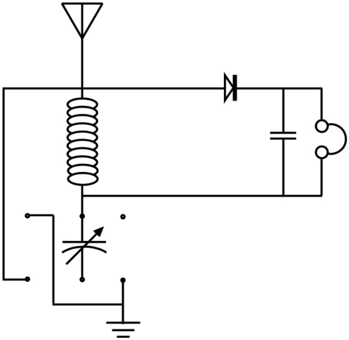

Series / Parallel switched tuning capacitor

DPDT switch to choose between the two prior circuits in a single set.

Performance Test
C Lauter

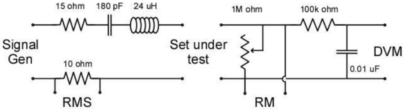
15 ohm
180 pF
24 uH
Signal Gen
10 ohm
RMS
Set under test
1M ohm
100k ohm
DVM
0.01 uF
RM

Artificial antenna Measurement load

Artificial Antenna

L1 =	20 uH
R1 =	15 ohm
R2 =	10 ohm
C1 =	200 pF

Measurement Load

R1 =	1M ohm pot
R2 =	100k ohm
C2 =	0.01 uF

Hookup to measure crystal set performance. Connect dummy antenna between Signal generator and set, use load between set and DVM. Follow protocol of Lauter: http://www.oldradioworld.de/gollum/testing.htm

Note**:** Modification of the Load (pot) made to determine impedance match at output. kjs

Dummy Antenna

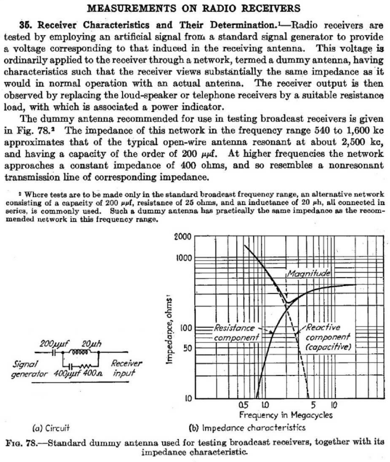

MEASUREMENTS ON RADIO RECEIVERS

35. Receiver Characteristics and Their Determination.[1]—Radio receivers are tested by employing an artificial signal from a standard signal generator to provide a voltage corresponding to that induced in the receiving antenna. This voltage is ordinarily applied to the receiver through a network, termed a dummy antenna, having characteristics such that the receiver views substantially the same impedance as it would in normal operation with an actual antenna. The receiver output is then observed by replacing the loud-speaker or telephone receivers by a suitable resistance load, with which is associated a power indicator.

The dummy antenna recommended for use in testing broadcast receivers is given in Fig. 78.[2] The impedance of this network in the frequency range 540 to 1,600 kc approximates that of the typical open-wire antenna resonant at about 2,500 kc, and having a capacity of the order of 200 μμf. At higher frequencies the network approaches a constant impedance of 400 ohms, and so resembles a nonresonant transmission line of corresponding impedance.

[2] Where tests are to be made only in the standard broadcast frequency range, an alternative network consisting of a capacity of 200 μμf, resistance of 25 ohms, and an inductance of 20 μh, all connected in series, is commonly used. Such a dummy antenna has practically the same impedance as the recommended network in this frequency range.

200μμf
20μh
Signal generator
400μμf 400Ω
Receiver input
2000
1000
Magnitude
Impedance, ohms
100
50
10
Resistance component
Reactive component (capacitive)
0.5
1.0
5
10
Frequency in Megacycles

(a) Circuit (b) Impedance characteristics

FIG. 78.—Standard dummy antenna used for testing broadcast receivers, together with its impedance characteristic.

From Terman's Radio Engineer's Handbook, first edition, 1943, pp 973 and 974.

MRL #1

1S 1 **LWT**

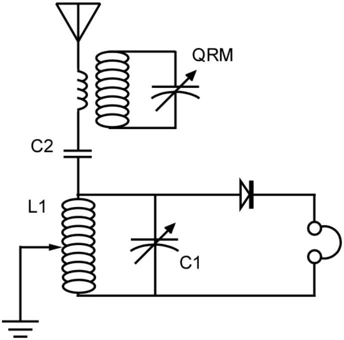

Form	2" x 4 ½"
L1	75t, 22 dcc ~175 uH Taps: 3, 6, 9, 12, 15, 20, 25, 30, 35, 40, 45 50
C1	365 pF
C2	fixed, 15 – 100 pF depending on antenna (or just make it variable)

MRL #17

1S 1 TC

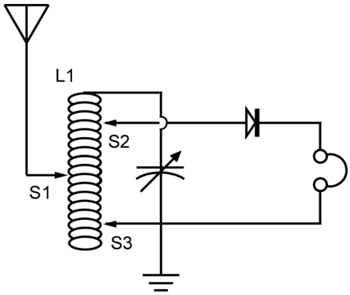

Form	2” x 4 ½”		
L1	111t,	24 dcc	~340 uH
	Taps: 20, 36, 46, 51, 56, 66, 76/83, 89, 96, 111		

S1 and S2 to taps 20, 36, 46, 51, 56, 66, 76
S3 to taps 83, 89, 96, 111

Simple Amplification for your Phones

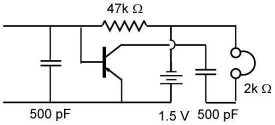
47k Ω
2k Ω
500 pF
1.5 V
500 pF

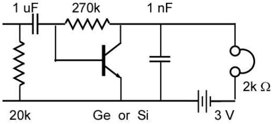
1 uF
270k
1 nF
2k Ω
20k
Ge or Si
3 V

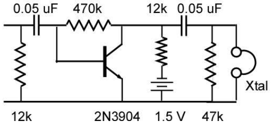
0.05 uF
470k
12k
0.05 uF
Xtal
12k
2N3904
1.5 V
47k

A few extra components can add up to 10 dB of volume to your phones. This may offset some of the lost sensitivity when used with highly selective sets..

Simple circuits with Transistor Detectors

http://www.noniandjim.com/Jim/RadioProjects/UsingElectronics/UsingElectronics.html

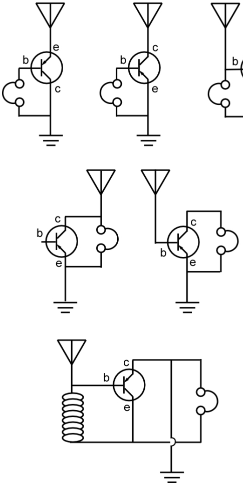
b
e
c
b
c
e
b
c
e
c
b
e
c
b
e
c
b
e

Transistors may be substituted for diode detectors to make interesting configurations while remaining un-powered. These form a starting point, see what works.

Steinite

1W 1 **TC**

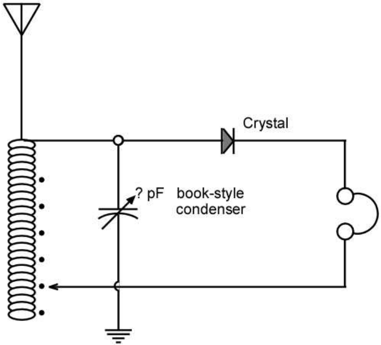
Crystal
? pF
book-style
condenser

Basic radio from 1920’s

Spiderweb coil
ID OD
45t, 22 dcc, 6 taps

C1 mica book-style ~250 uH

Borden Radio Dunwoody

1S 2 TC

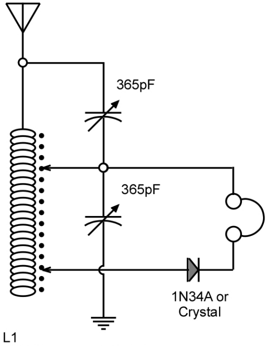
365pF
365pF
1N34A or
Crystal
L1

Form	2” dia	
L1	100t,	22 awg
C1-2	365 pF	

Impedance Matching Transformer

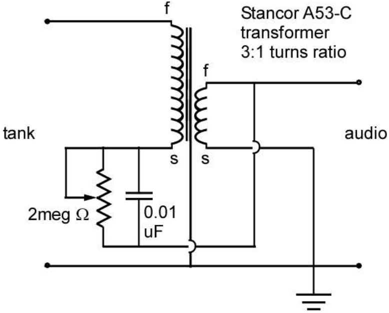
f
Stancor A53-C
transformer
3:1 turns ratio
f
tank
audio
s
s
2meg Ω
0.01
uF

I have generally avoided including schematics and info concerning impedance matching transformers as I do not have much experience with them. The above general schematic I take from Ben Tongue’s article 01. Note that he includes a benny as a standard part of the transformer circuit.

The transformer provides a 16:1 transformation ratio between the typical 12k ohm phones and a tank Rd of around 192k ohms.

Detector Bias Circuits

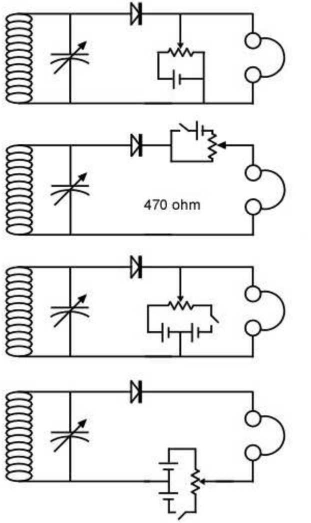

Babani Long Distance

1S 2 TC

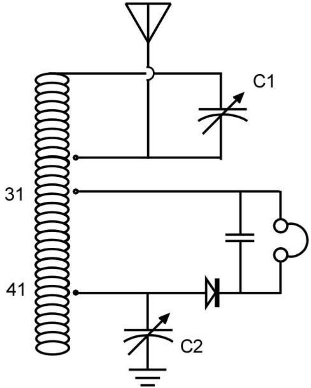

Form 2 ½” x 3”

L1	51t, 24 swg ~170 uH Taps 2, 4, 6, 8, 10, 12, 14, 16, 31, 41
C1-2	365 pF

MRL #13

1S 2 SL, LWT

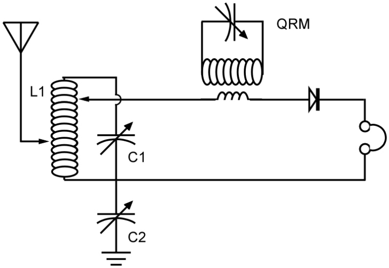

2-Slider with QRM

Form 2 ½ x 4 ¼”
L1 120t, 22 awg ~460 uH
C1-2 365 pF not ganged

Benny

A Benny serves to equalize the DC load resistance on the diode to the AC audio load.

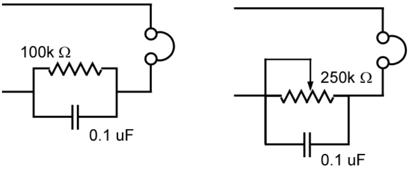

Examples of Benny circuits, with or without variable resistance.

Selectivity Enhancement Circuit

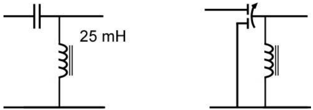

Example of a circuit to include for selectivity enhancement. Simple cap, small trimmer, or differential cap (shown on left). Cap ~ 20pF, choke 25+- mH.

Wave Traps

Inductively coupled:

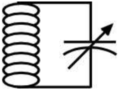

Coil should have same diameter as set tuning coils, placed in-line to couple.

Directly coupled:

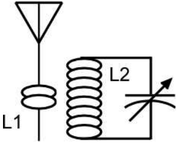

Form 1” x 1 ½”

L1	110t, 30 awg on base layer
L2	15t, 22 awg space wound on top
C1	410 pF

Schmarder #30

1L 3 RFC B

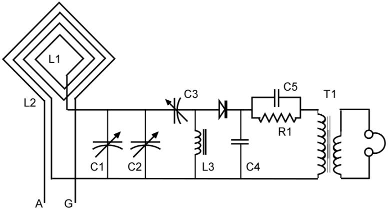

L1	Loop 30x36“, holes spaced 3/8“ 13t, 660/46 litz
L2	2t, 40/38 litz coupling winding
L3	2.5 mH RFC
C1-2	365 pF (ganged?)
C3	3-30 pF trimmer
C4	500 pF
C5	0.1 uF
R1	100 k

C3-L2 is the Selectivity Enhancement Circuit.
C5-R1 is the Benny

Schmarder #43

1W 3 CK

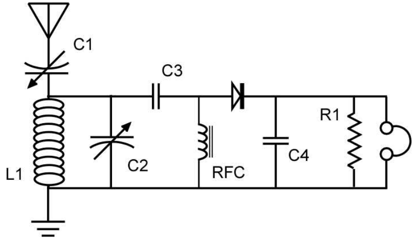

L1	spider, 1“ ID, 54t, 165/46 L
C1-2	365 pF (ganged?)
C3	10 pF trimmer (or 5-50 var)
C4	180 pF
R1	47 k (w/crystal earphones only)
RFC	27 uH

C3-RFC is the Selectivity Enhancement Circuit.

CHAPTER 3

MISCELLANEOUS

MRL #2

1S 2G LWT

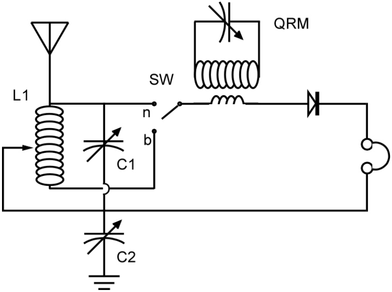

Form 2“
L1 90t, 22 dcc ~220 uH
taps, 5, 10, 16, 23, 31, 40, 50, 61, 73, 90

C1, C2, 365 pF, may be ganged
Switch broad / narrow tuning

QRM may be placed in tank as well to cut out strong locals.

This set “looks” a lot like May’s BBCS#6 set without the choke, (p 20).

KJS Fleming

1S 2G Valve

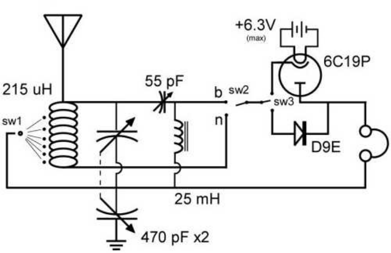
+6.3V
(max)
6C19P
215 uH
55 pF
b
sw2
sw3
sw1
n
D9E
25 mH
470 pF x2

Valve Detector of Professor Fleming
Patented 1904

Form 3.54”	
L1	49t, 18awg, taps 0,8,12,17,24,34
C1a-b	470 pF ganged

This is MRL #2 with Selectivity Enhancement Add, superficially resembling BBCS #6, p21.

Harthun Shortwave Mystery XSS #62

3D 1 TF SW

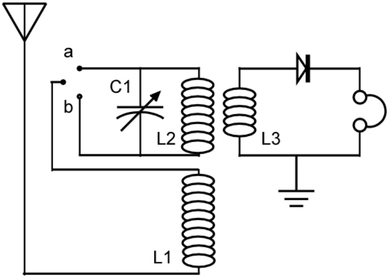
a
b
C1
L2
L3
L1

Form 2 3/8” x 2 ½”	
L1	20t, 22 awg, close wound
Space 1/2” dual winding	
L2	13t, 20 awg
L3	6t, 22 awg, bifilar with L2
C1	15 – 250 pF

The antenna may be connected directly to post A or B bypassing the coil L1.

Al Klase SW

3S 3G LC, LD

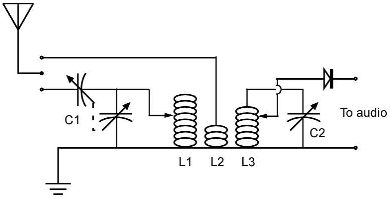
C1
C2
To audio
L1
L2
L3

L1	2”, 18t, 14 bare awg, clip to tap
L2	2”, 2t, 14 bare awg
L3	2”, 8t, 14 bare awg, clip to tap, ~4.3 uH

C1	100 pF 2 gang
C2	365 pF
C3	fixed 0.5 uF

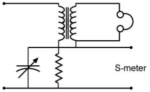

R1	100 k
T1	100 – 600 k

Back end to audio transformer and S-meter

WJ May   BBCS #6

2S 2 TC, LWT

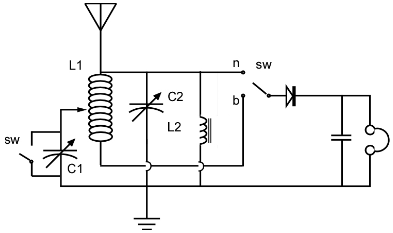
L1
n
sw
b
C2
L2
sw
C1

Forms 2” x 4”

L1	90t, 22 dcc ~230 uH
	Taps 5, 10, 15, 25, 30, 40, 50, 60, 70, 80
L2	choke 110t, 38 dcc right angle to L1

C1-2	500 pF

This set looks similar to MRL #2   (p19),
See also MRL #8 on p67 and MRL #23 on p78.

Sleeper #11

2S 0 LD, SL

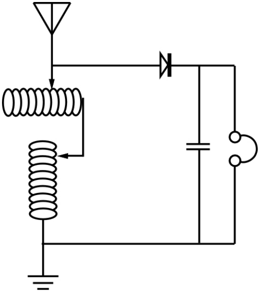

Slider tuned radio with loading coil in front end
Coils in series

Loading coil increases the wavelength tuned.

H Lash SW

2S 4G LC

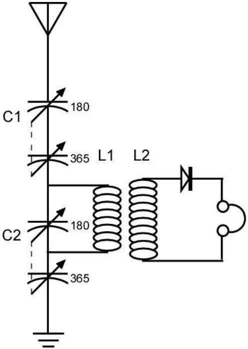

Shortwave

L1	2”, 9 1/2t, 22 awg
L2	2 1/8”, 30t, 22 awg on top layer
C1-2	365 pF – 180 pF 2-gang

XSS #56 SW

2S 2 TC

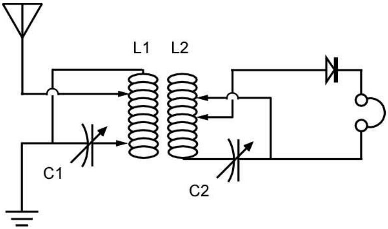

Shortwave Set

Forms 3 ½“ x 3 ½“
L1-2 space wound, 6-8 tpi, 14 awg, ~38 uH

C1-2 100 pF

MRL #11

2S 1 TC

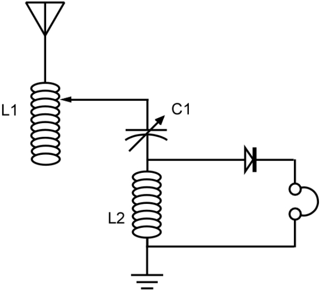

Form 2 ½ x 4 ¼”
L1 132t, 24 dcc Loading, ~625 uH
12 taps, spaced each 11 turns
L2 40t, 20 dcc 2“ space wound ~80 uH
Coils at right angles, not coupled

C1 350 pF

BBC Commemorative (Prac Wireless Sep 1972)
AKA Sleeper #12

2W 0 VM

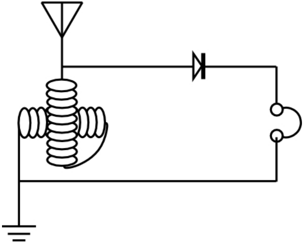

Variometer
Pancake-style with 2 spiderweb coils

Tuning accomplished by varying the mutual inductance between two coils connected in series.

Owen Poole SW

2S 1 LC

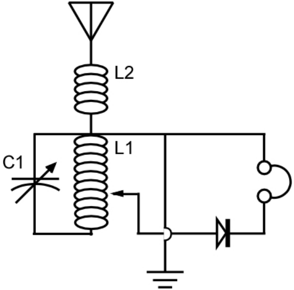

Form 1 5/8”

L1	14t, 24 awg hookup wire ~12.4 uH
L2	5t, 24 awg hookup 1/8” from first coil
C1	10 – 60 pF

KJS Global DX

1W 2 TC

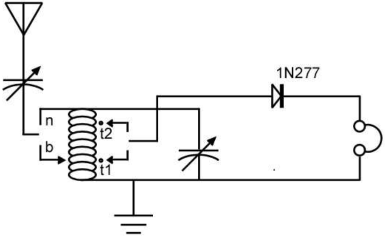

Spiderweb Form ID 0.6“, OD 2.0“

L1	9t, 18 awg	7.6 uH	
	Tap 3		3.8 uH
	Tap 8		5.8 uH
C1-2	10 - 115 pF (tunes ~ 6 – 18 MHz)		

Set based on Gollum SW Set

Sleeper #12, 13

2S 0 VM

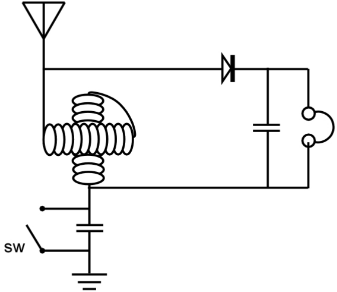

Variometer

Optional capacitor in series to ground
to tune upper BC band

Sleeper #14

2S 1 VM

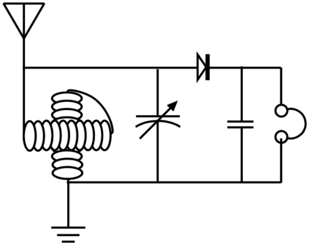

Variometer
Parallel tuned circuit

Variable condenser to tune wavelengths greater than the circuit with variometer alone.

CHAPTER 2

SHORTWAVE HOOK-UPS

120 M	2.3	2.5 MHz	
90	3.2	3.4	
60	4.75	5.06	
49	5.9	6.2	active
41	7.3	7.35	
31	9.4	9.9	active
25	11.6	12.1	active
22	13.5	13.9	acrive
19	15.1	15.8	active
16	17.5	17.9	
15	18.9	19.0	

MRL #22

2S 1 VC, TC

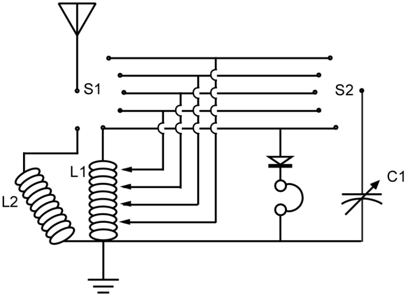

Variometer, coils in series

L1 stator	2“ 54t, 22 dcc Taps : 14, 27, 39, 48, 54
L2 rotor	1 1/2", 11t, 22dcc
C1	365 + Pf
S1-2	Roraty switches

Sleeper #15

3S 0 VM, TC

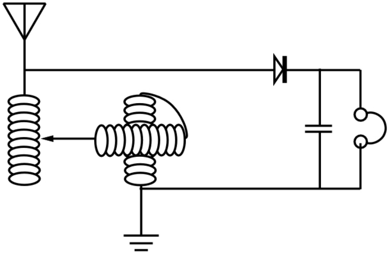

Variometer
With tapped loading coil in series with the variometer. Tap separation equal to the inductance range of the variometer.

Barkett, Radio World 1924, (st #20)

8W 2 LD LC VM

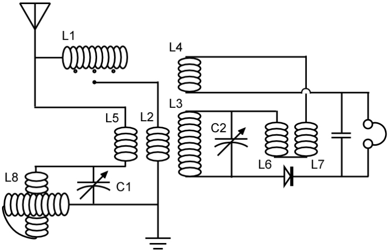

The Ultimate Coil Fantasyland:

Coils L1-L5 basket weave
Coils 6-7 honeycomb

L1	3" loading, 50t, 20 awg, taps	~175 uH
L2	4", 25t, 20 awg	~ 90 uH
L3	3", 60t	~230 uH
L4	4", 30t	~125 uH
L5	3", 30t	~ 80 uH
L6-7	25t	
L8	variometer	
C1-2	365 pF	

Petersen XSS 107-8

5S 4G LK CK

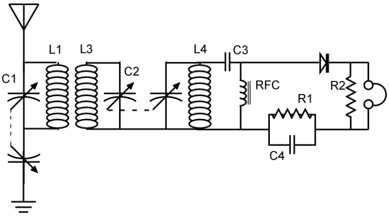

L1	4 x 3“, 44t, #26 hookup, stator	~250 uH
L2	1 ¾ x 3“, 15t gap 15t, rotor	
C1	365 pF 2-gang	
L3-4	3 ¼”, 68t, 20 awg, right angles	~300 uH
C2	365 pF 2-gang	
C3	10 pF	C3/RFC SEC circuit
RFC	2.5 uH	to lightly load detector
C4	0.1 uF	C4/R1 is a Benny
R1	22 k	
R2	120 k	for high-impedence phones

MRL #29

3S 1 VC, TC

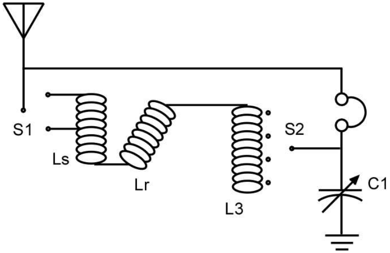

Variometer

Form 2“ x 4 ½“

Ls	3 ½“ x 2 ½”, 60t, 24 dcc, gap ~360 uH Tap at 30
Lr	2 ½“ x 2”, 64t, 28 dcc, gap ~300 uH
L3	loading, 2“ x 4 ½“, 100t, 24 awg Tap each 25t ~300 uH
C1	365 pF

Schmarder #11

3D 2 TC VM

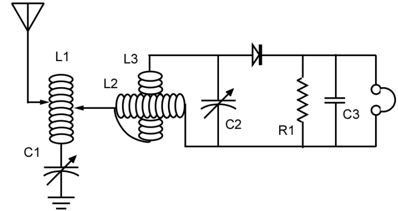
L1
L3
L2
C2
R1
C3
C1

L1	3 1/8”, 50t, tap each 10t
L2	stator, 3 1/8“, 40t, 40/44 L
	L1-2 dual wound on same form, 1/8” apart
L3	rotor, 2 1/8“, 48t, 40/44 L

C1-2	500 pF
C3	220 pF
R1	33k (w/crystal earphones only)

Coupling between antenna coil and stator.

KJS Minstrel Boy

5T 3 LK

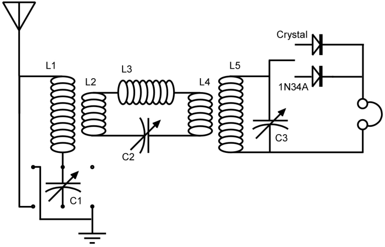
Crystal
L1
L3
L5
L2
L4
1N34A
C3
C2
C1

Link coupled Triple tuned set

L1-L5	180 uH Air-core torroids (Magnavox)
L2-L4	35 uH coupling coils within above
L3	110 uH Honeycomb,

C1,2,3	540 pF each

Series/Parallel switch on antenna circuit cap

Radio Craft ST #31

4S 3G **Band Selection**

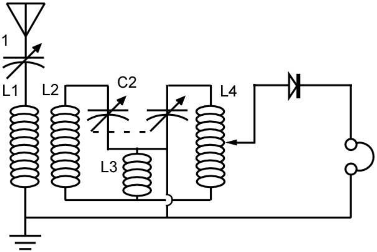

L1	4“, 80t, 28awg, stator	~840 uH
L2	2“, 80t, 28awg, rotor	~300 uH
L3	1”, 48t, 28 awg	~ 50 uH
L4	3”, 90t, 24 awg	~515 uH

C1	365 pF
C2	365 pF 2-gang

turns on L3 controls degree of band selection

How does this work? How to “select” band?
Make L3 variable?

MRL #15

3S 2 VC

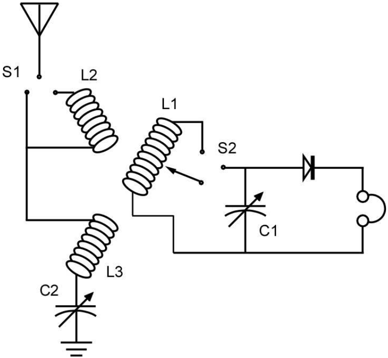

L1	3“ x 4”, 24t, 20 dcc, ½“ gap, 45t, 22 dcc Tap at 30
L2-3	2” x 2”, 60t, 24 awg
C1-2	? 365 pF

KJS DeForest (aka Sleeper #18)

2S 1 LC, TC

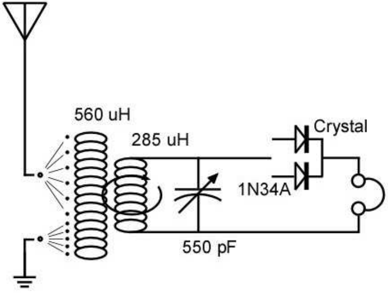

Basic Variocoupler

L1	Primary: stator tapped fine / coarse switches	560 uH
L2	Secondary: rotor Variocoupler NOS De Forest Co	285 uH
C1	550 pF to tune secondary circuit Used, Windham Co	

Note: The radio doesn’t couple well and tunes broad as a barn. Pity…

Ford, Pop Mech 1945, (st #111)

4L 3G LK

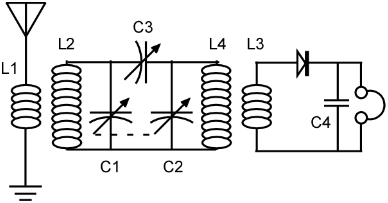

Form 1” x 2 ½”

L1/L3	30t, 30 awg, top layer	~ 28 uH
L2/L4	140t, 30 awg, base layer	~235 uH
C1-2	365 pF 2-gang	
C3	3-30 pF trim	
C4	0.001 uF	

Coils at right angles

Option to connect all three circuits to ground to increase sensitivity.

XSS #63

4S 4 LK

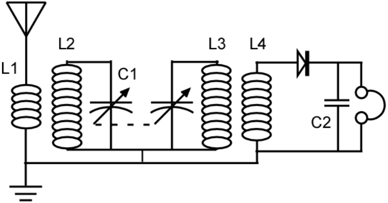

Link coupled variocoupler

L1 rotor	1 1/16"x 7/8", 16t, 28 awg, gap in coil	
L2 stator	1 7/8"x 2 3/8", 72t, 28 awg	~230 uH
L3	1 7/8", 70t, 28 awg,	~220 uH
L4	1 7/8", 35t, 28 awg,	~75 uH
C1	365 pF, 2-gang with trimmers	
C2	0.001 uF	

MRL #35

3S 3 TC, VC

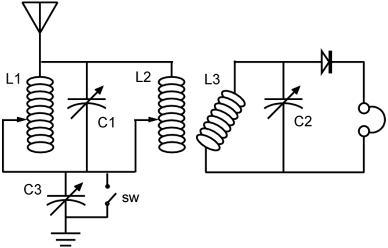

Forms 2" x 4 ½"

L1	100t, 24 dcc	~300 uH
	Tap every 10t	
L2	stator	100t, 24 dcc, tap every 10t
L3	rotor,	1 ½" x 1 ½", 35t, 32 awg
C1-3	365 pF	

Secor, 1920

4S 0 VM

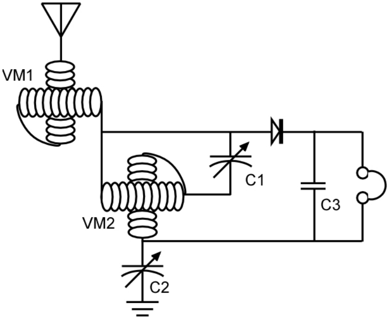
VM1
C1
C3
VM2
C2

Double variometer

VM1	loading
VM2	tuning
C1-2	365pF
C3	0.001 uF

Hogan 1922

4S 2 LK, TC

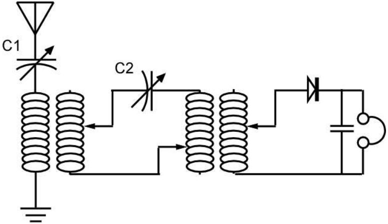

Link coupled variation with sliders or taps

C1	365 pF
C2	500 pF or higher

Sleeper #20

4S 2 LK

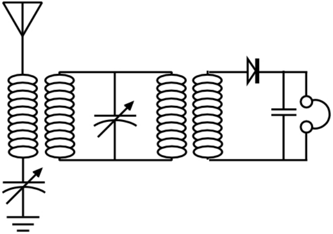

Link coupled set with loose coupling

Primary and intermediate circuits tuned with variable condensers while secondary circuit is untuned, (aperiodic).

Stay Tuned #76 (ref not known)

3W 0 LC, TC

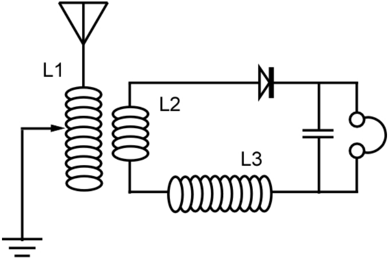

Tapped Spiderweb Rig

Coils ID 1 ¾“

L1	50t, 22 dcc, over 1 under 2, 20 spokes Tap every second turn
L2	12t, 22 dcc, over 2 under 2, 19 spokes
L3	50t, 22 dcc, over 2 under 2, 19 spokes

Visit Darryl Boyd’s excellent website to see the plans for this set. It is a visually interesting set that will please you.

KJS Bremer Tully Set

2D 1 LC

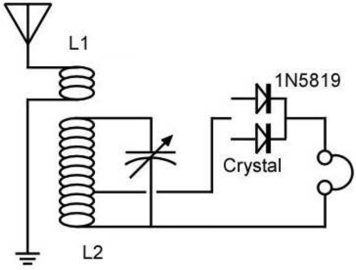

Basic tuner with co-wound antenna coupling coil

Dual wound primary and secondary on open form

L1	20 uH
L2	370 uH
C1	10 - 270pF

This set features a lovely bank-would coil with separate primary and secondary. The high primary inductance necessitated a low-value variable cap with as low a minimum as possible.

Coupled trap

4S 2 IWT

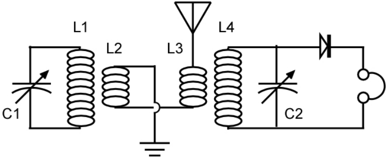

Coupling coil circuit with trap

L1-L4	1.5” form, 90t 28 awg, 220 uH
L2-L3	1.5” form, 45t 28 awg, 85 uH
C1-C2	365+ pF

(Reference unknown, clipped from an old magazine)

BBCS #11

2S 3G TC

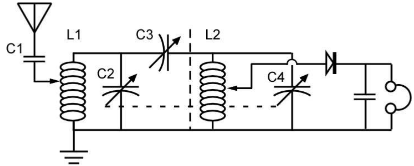

Forms 2 ½” x 2”

L1-2	50t, 28 swg (~25 awg)	~180 uH
C1	100 pF	
C2 C4	500 pF 2-gang	
C3	50 pF trimmer	

Band Pass Circuit
Grounded screen between coils to prevent coupling

Capacitively coupled

MRL #10

2L 1 TC

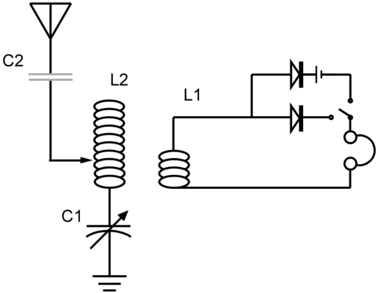

Form	2 x 4 ¼”

L1 40t, 20 dcc start 1/8“ from end ~95 uH
L2 160t, 28 dcc on top layer loading ~1100 uH
Taps 5, 10, 20, 60, 100, 160

C1	365 pF
C2	150 pF optional

Coils not coupled.

Dan Petersen Galenatron

3S 2 VC, TC

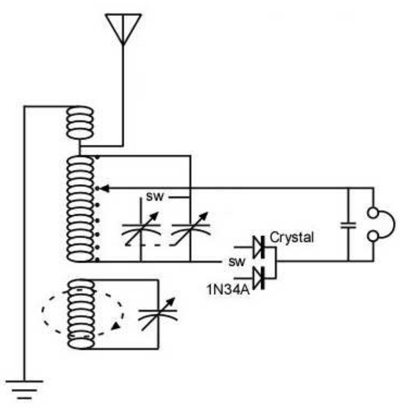

L1	3.15", 50t, 22 awg	~230 uH
L2	3.15", 10t, 22 awg	
L3	2.15"x2", 50t, 22 awg, gap in coil	
C1-2	400 pF	

Sleeper #21

2S 3G SL, CT

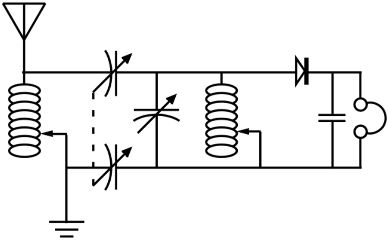

Primary and secondary circuits tuned by varying the coil inductance with sliders while coupling between primary and secondary is varied capacitively, (ganged capacitors)

XSS #62

5M 3G FW, TD

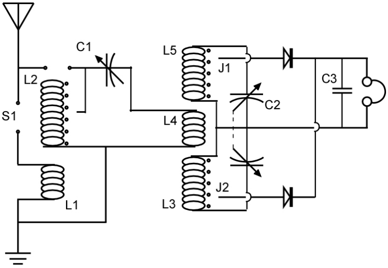
C1
L5
J1
L2
C3
C2
S1
L4
L3
J2
L1

Fullwave Receiver (variation on Pop Mech, P72)

Form 1 7/8” x 2”

L1	72t, 28 awg, Taps 0, 16, 30, 44, 58, 72
L2	17t, 28 awg

Form 1 7/8” x 3 ½“

L3,5	72t, 28 awg, Taps 0, 18, 36, 54, 72
L4	17t, 28 awg

S1	6-position switch

XSS #59

3D 2 LD, TC

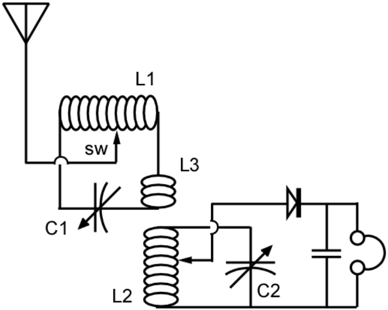
L1
sw
L3
C1
L2
C2

L1	2 ¾”x 3”, 42t, 22 awg	~130 uH
L2	2 ¾”x 4”, 67t, 22 awg	~255 uH
L3	2 ¾”x 4”, 15t, 22 awg 1/8” from L2, same form	~ 25 uH
C1-2	365 pF	
Sw	4-6 position switch	

MRL #24

3D 2 LD, TC

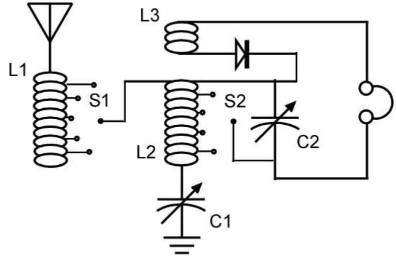

Forms 2” x 4½”

L1	100t, 24 dcc, loading coil	~300 uH
	Taps 0, 25, 50, 75, 100	
L2	90t, 24 dcc,	~260 uH
	Taps 5, 25, 50, 90	
L3	25t, 24 dcc 1/8” from L2, same form	~ 40 uH
C1-2	365 pF	

Popular Mechanics (Dec-1950): **BBCS #4**

3M 3G FW

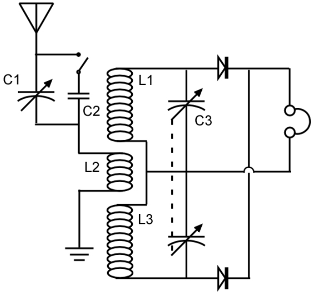

Full wave receiver

Form 1” x 4”

L1	43t, 32 awg	~ 60 uH
L2-3	120t, 32 awg	~270 uH

1/8“ between windings

C1	365 pF ?
C2	fixed, ? pF
C3	365 pF 2-gang with trimmers

This set is a fun candidate to make with a 5726 dual Diode vacuum tube.

EL Anderson, (GE Ham News, 1955, v10, n6)

2L 4G FW DA

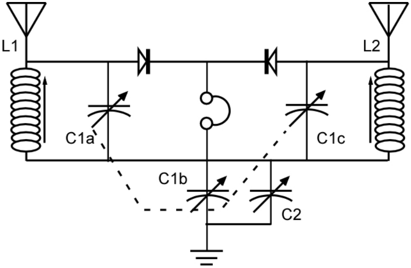

Twin-Crystal Set

L1, L2	Variable Loopstick Coil
C1	15-400 pF 3-gang
C2	15-400 pF To trim balance between antennae

MRL #43

2D 2 Counterwound

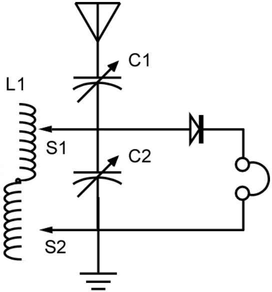

Form 2” x 4 ½”

L1	Use 27 awg ~375 uH each section Wind 100t, anchor then reverse wind 100t Taps every 10t on each section
C1-2	365 pF (not ganged ?)

Gordon McCall

2S 1 VC

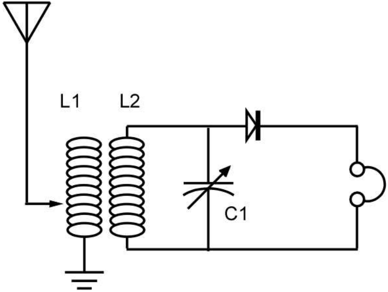

Variocoupler

L1	4", 57t with gap, #22 awg
L2	3", 57t with gap, #22 awg
C1	? 365 pF

Solomon Almighty

3C 3G TF

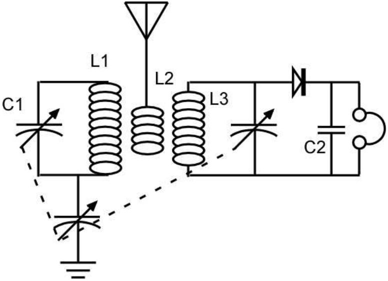

Telefunken Style

Form 3"

L1	50t, 24 awg, base layer
L2	15t, 24 awg, top layer
L3	25t, 24 awg, bifilar with L1
C1	500 pF, 3-gang
C2	0.001 uF

MRL #23

3L 2G TF, CK, QRM

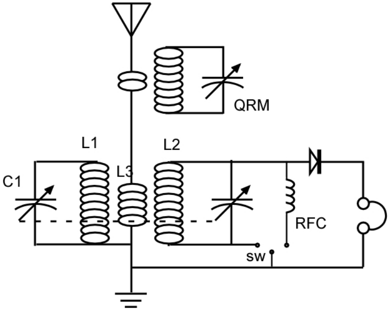

Telefunken Style with RFC and QRM
See also MRL #8 on page 67.
With family resemblance to sets on pages 19-21.

Form 2“ x 4 ½“

L1	75t, 22 dcc, base layer	~175 uH
L2	75t, 22 dcc, middle layer	~190 uH
L3	15t, 22 dcc top layer	~ 19 uH

C1	365 pF, 2-gang
RF choke	85 mH

MRL #3

2S 2 VC, TC

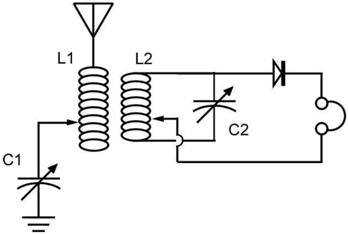

Variocoupler

L1	3“, 100t, 24 dcc, 4 taps	stator	~590 uH
L2	2“, 75t, 24 dcc, 4 taps	rotor	~205 uH

C1-2	365 pF

MRL #30

3D 2G LC, LWT

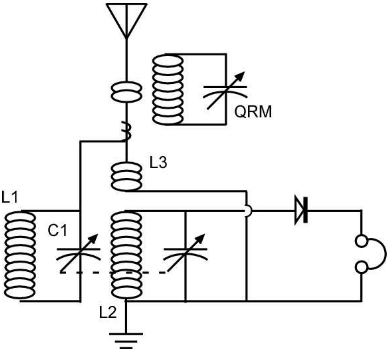

QRM	1”x 1½”, 110t, 30 awg, 15t, 24 dcc space	
Forms 2” x 4½”,		
L1-2	75t, 22 dcc, couple to ant	~175 uH
L3	8t, 22 awg 1/8” from L2, same form	~5-6 uH
C1	365 pF, 2-gang	

Solomon Crazy +

3C 2G TF

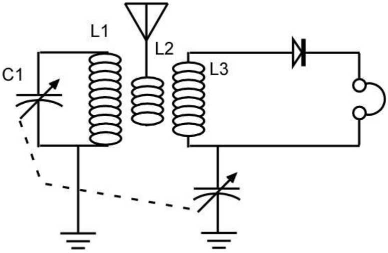

Telefunken Style

Form 2.5”

L1	45t, 24 awg
L2	15t, 26 awg on top layer
L3	25t, 30 awg, bifilar with L1
C1	500 pF

Solomon Mystery +

3C 2 TF

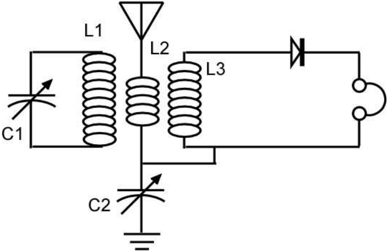

Telefunken Style

Forms 2.76”

L1	45t, 22 awg
L2	15t, 22 awg on top layer
L3	25t, 30 awg, bifilar with L1
C1-2	350 pF not ganged

Sleeper #16

2S 0 LC, SL

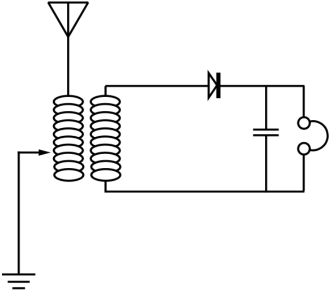

Loose coupler

Aperiodic secondary, responds to any wavelength with which the Primary coil may be tuned.

Sleeper #17

2S 1 LC, SL

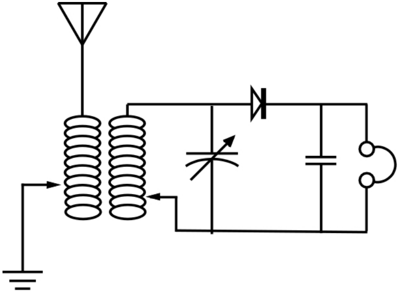

Loose coupler with tuned secondary

Both Primary and secondary circuits may be tuned to resonance giving high efficiency of energy transfer. Changing coupling between the two coils can vary the degree of selectivity attainable, at the expense of sensitivity.

MRL #4

3L 1 TF

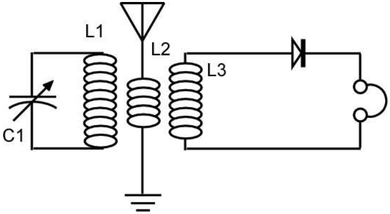

Basic Telefunken Set

Form ?

L1	75t, 22 dcc, base layer	~175 uH
L2	20t, 22 dcc, middle layer	
L3	10t, 22 dcc top layer	
C1	365 pF	

Proton Mystery +

3C 1 TF

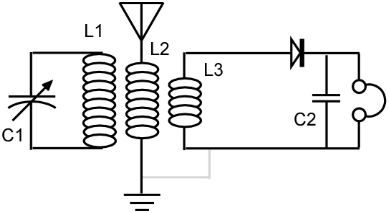
L1
L2
L3
C1
C2

Telefunken Style

Form 3“

L1	45t, 23 awg, base layer
L2	15t, 23 awg, top layer
L3	25t, 29 awg, bifilar with L1

C1	365 pF
C2	0.001 uF

KJS Teflon

2S 3G LC

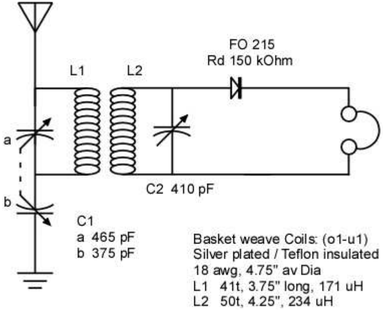
FO 215
Rd 150 kOhm
L1
L2
a
b
C2 410 pF
C1
a 465 pF
b 375 pF
Basket weave Coils: (o1-u1)
Silver plated / Teflon insulated
18 awg, 4.75" av Dia
L1 41t, 3.75" long, 171 uH
L2 50t, 4.25", 234 uH

Simplicity in a double-tuned rig.

Silver-plated, Teflon-insulated Coils, over 1 / under 1
18 awg, 4.75” av diameter

L1	41t, 3.75” long, 171 uH
L2	50t, 4.25” long, 234 uH
C1a	465 pF
C1b	375 pF
C2	433 pF

Mike Tuggle Lyonodyne 17

2W 3G LC, TG, 2WT

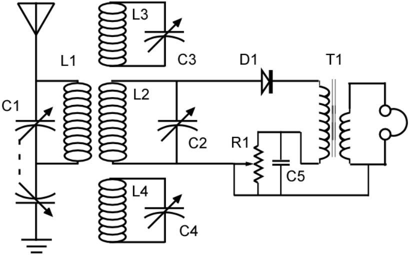
L3
C3
D1
T1
L1
C1
L2
C2
R1
C5
L4
C4

L1	51t ferrite	147 uH
L2	36t basket	185 uH
L3	36t basket	185 uH
L4	42t basket	187 uH

C1	15 – 470 pF, 2-gang
C2, C3, C4	15 – 497 pF
C5	0.02 uF
R1	500 kΩ
T1	UTC A-27
D1	RS 12101 3RT

Gollum Mystery

2C 4G TF

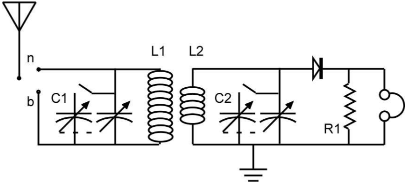
n
b
L1
L2
C1
C2
R1

Telefunken Style

Form 3.15”

L1	48t, 26 awg ~250 uH
L2	28t, 22 awg, bifilar with L1
C1-2	?365 pF, 2-gang
R1	68 kohm

Capacitor gangs switchable in parallel to add capacitance.

MRL #39

2L 2G TF, TG

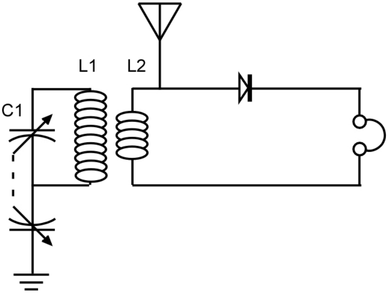

Telefunken Style

Form 2” x 4 ½”

L1	80t, 22 dcc primary on bottom layer 190uH
L2	40t, 20 dcc secondary on top layer ~ 70uH

C1	pF 2-gang

Mike Tuggle Lyonodyne Mod

2W 3G LC, TG, MOSFET

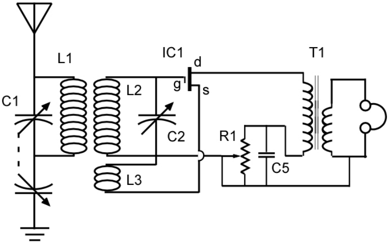

L1	51t ferrite	147 uH
L2	36t basket	185 uH
L3	8t basket	

Outboard traps not shown

C1	15 – 470 pF, 2-gang
C2	15 – 497 pF
C5	0.02 uF
R1	500 kΩ
IC 1	ALD 110900 APA MOSFET
T1	UTC LS-12

Owen Poole Project Crystal Radio #5

2S 2 TC, LC

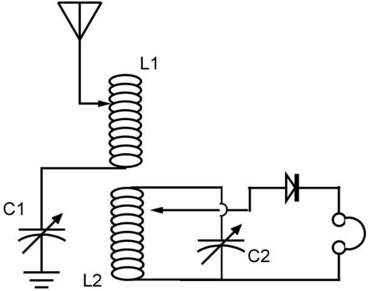

Antenna tuner w/secondary

Form 1 ¾“		
L1-2	152t, 24 awg Tap each 10 turns	~390 uH
C1-2	? 365 pF	

(essentially the same as MRL #3)

Solomon Double gang

2C 2G TF

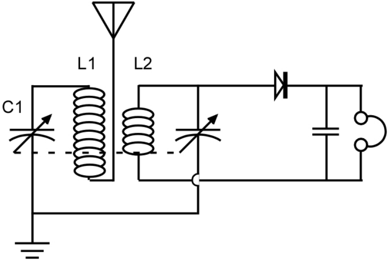

Telefunken Style

Form 6 cm (3.275”)

L1	50t, 24 awg primary
L2	25t, 24 awg secondary co-wound
C1	500 pF 2-gang

Mike Tuggle Mystery

2C 2G TF, TG

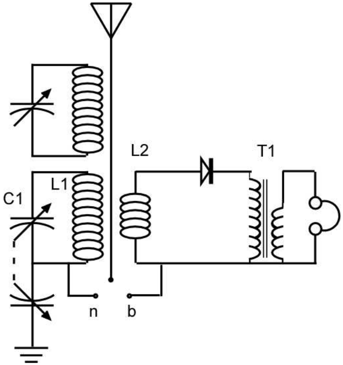

Telefunken Style w/QRM

L1	36t, 660/46 litz 5 inch basket weave
L2	18t, 420/46 litz
C1	500 pF 2-gang
T1	UTC LS-12

XSS #58

2S 2 TC, LC

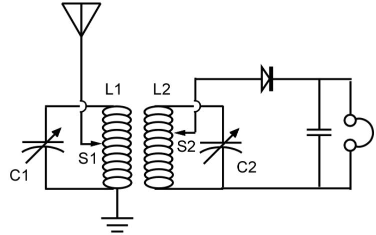

Double loose coupler Navy Style

L1	3 ½”x3 ½”, 48t, 22 awg Taps 0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 44 48
L2	2 3/8” x 3 ¾”, 72t, 22 awg Taps 0, 6, 12, 18, … 72
C1-2	365 pF

Tom Polk Simply Symmetry

2S 3G TG, 4SL

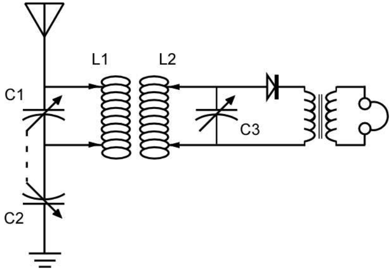
L1
L2
C1
C3
C2

Form 2” x 5 ½”

L1-2	125t, 18 awg ~240 uH
C1-2	365 pF 2-gang
C3	365 pF

Radio Shack 70v Line Transformer
Sliders

KJS Mystery

2C 2G TF, TG

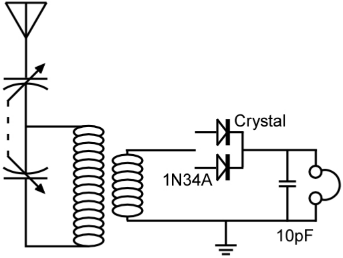
Crystal
1N34A
10pF

Telefunken Style

Form 8 cm (3.275”)

L1	46t, 18 awg primary
L2	24t, 18 awg secondary co-wound
C1	495 pF 2-gang

Peebles XSS #9

2L 1 TF

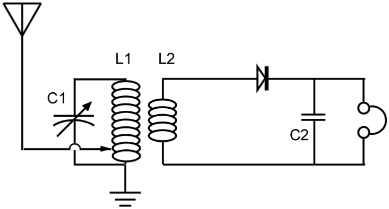

Telefunken Style

Form 1 ¼“ x 1 5/8“

L1	110t, 28 awg, base layer	~205 uH
L2	30t, 28 awg, top layer	~ 34 uH

C1	365 pF
C2	0.001 uF

Long antenna tap at 10t from ground

Owen Poole Off the Shelf

2S 3G TC, LC

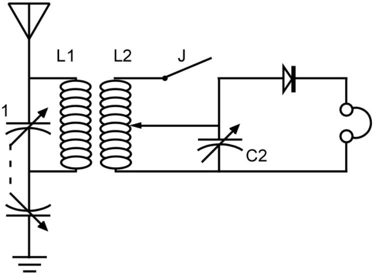

Form 4“ x 4”

L1-2	60t, 20 awg hookup wire Tap at 20
C1	365 pF 2-gang
C2	365 pF
J	optional jumper

KJS Hammarlund

2B 3G TG, SEC, B

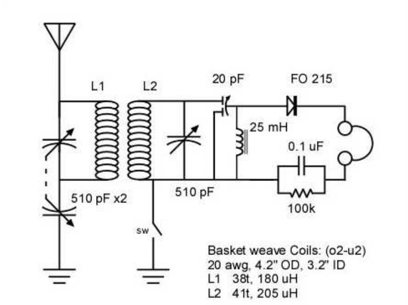
L1
L2
20 pF
FO 215
25 mH
0.1 uF
510 pF x2
510 pF
100k
sw
Basket weave Coils: (o2-u2)
20 awg, 4.2" OD, 3.2" ID
L1 38t, 180 uH
L2 41t, 205 uH

Basketweave coil 4.2" OD – 3.2" ID

	Over 2 – under 2		
L1	38t	20 awg	180 uH
L2	41t	20 awg	205 uH
C1-3	510 pF		C1-C2 are ganged
C4	20 pF differential		
C5	0.1 uF		
RFC	Choke: 25 mH		
R1	100k ohm		
D1	ITT FO-215 germanium		

MRL #8

2L 1 TF, CK

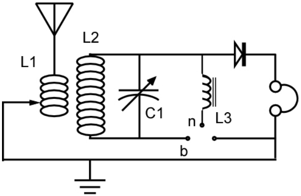
L2
L1
C1
n
L3
b

Telefunken Style
See also MRL #23 on page 78.
With family resemblance to sets on pages 19-21.

Form	2" ?
L2 75t, 22 dcc bottom layer	
L1 20t, 20 dcc space wound 2 ½" on top layer	
	Taps 4, 6, 7, 8, 9, 10, 11, 12, 14, 16, 18, 20
L3	BC band RF Choke, ~ 50ohm DC
	Single or 3-section ok
C1	? 365 pF

Proton Mystery

2C 1 TF

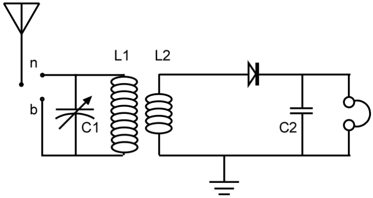

Telefunken Style

Form 3”

L1	50t, 23 awg primary
L2	25t, 29 awg secondary co-wound
C1	500 pF
C2	0.001 uF

Helps to add Loading coil in series with antenna.

XSS #54

3S 3 LC, WT

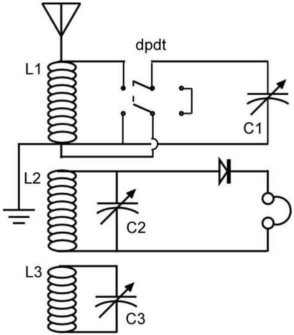

Forms 3 ½”x 3 ¾”

L1,2,3	55t, 20 awg, no taps	~250 uH
C1,2,3	500 pF	

Series / parallel DPDT switch over L1

Diode on top of tank, why? Candidate for SEC?

KJS ID #3

2S 3G TG, TC

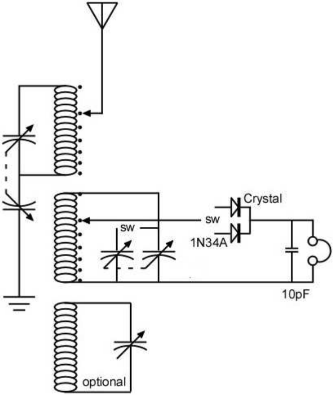
Crystal
sw
sw
1N34A
10pF
optional

Forms 3.15”	
L1	62t, 18 awg Tap 11, 16, 22, 31, 43, 62
L2	62t, 18 awg
C1-2	450 pF, 2-gang

Professeur Jacquemard

3S 2 MWLW

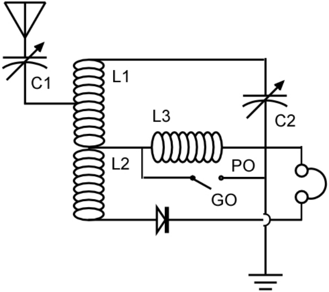
C1
L1
L3
C2
L2
PO
GO

Medium Wave / Long Wave Receiver

L1	25mm x 68mm, 170t 4/10 wire (~28awg) ~ 220 uH
L2	Honeycomb ID 25mm, OD 35mm, W 8mm #turns ?, In-line with L1
L3	Honeycomb ID 20mm, OD 34mm, W 6mm #turns ?. Coil for LW
C1-2	450 pF

Boursin Selective 1939

2S 1 MW LW

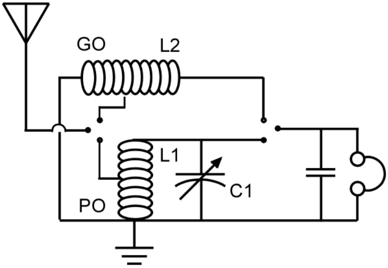

Medium Wave / Long Wave Set

L1 MW 30mm, 90t, tap at 45. ~220 uH
L2 LW ID 10-15mm, OD 20-25mm, W ?
280t, tap at 140
C1 500 pF
DPDT switch between LW-MW

GO = Grandes Ondes = LongWave
PO = Petits Ondes = Medium Wave

Dejan’s contest rig

2S 4G LC, WT, TG

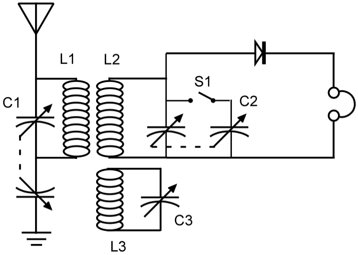

(Shown without audio transformer)

Contest Rig

L1	ferrite core
L2	basket weave
L3	spiderweb
C1-3	?

M Hampton, BRG 2009

2S 5G TG LC

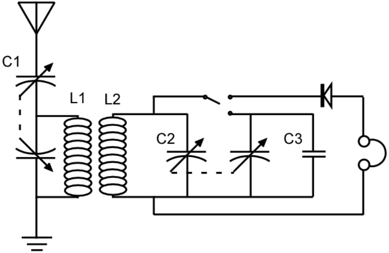
C1
L1
L2
C2
C3

Audio transformer not shown

Form 3“

Schmarder #1

2W 4

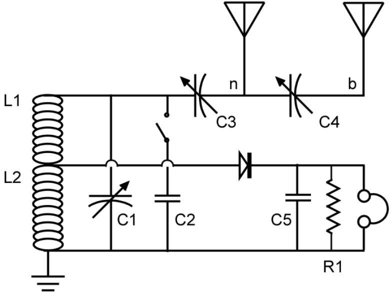
n
b
L1
C3
C4
L2
C1
C2
C5
R1

L1-2	spider, 1.5“ ID, 3.5“ OD	
	L1 15t, 40/44 L	~16 uH
	L2 33t	~80 uH
C1	356 pF	
C2-4	270 pF	
C5	.001 pF	
R1	33k (w/crystal earphones only)	
n	narrow tuning	
b	broad tuning	

RTV Electronics p129 George P Pearce

2SW 2 FW

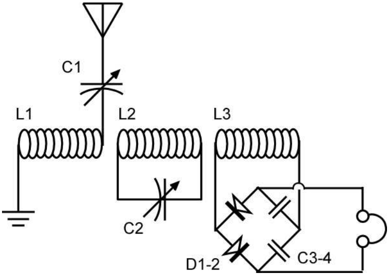
C1
L1
L2
L3
C2
D1-2
C3-4

L1, L3	Solenoid
	2” od, 1.2” l, 63t, 27 awg, ~150 uH
L2	Basket o1,u1
	2.25” od, 1.25” l, 80t, 21 awg, ~350 uH
	Coils spaced about 1” apart
C1, C2	380 pF
C3, C4	1000 pF
D1, D2	1N35

W Thelen BRG 2006

2W 5 TG LC

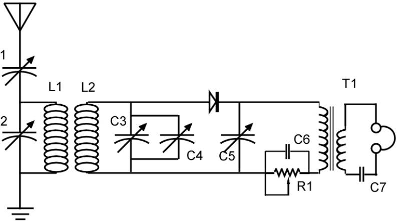
C1
L1
L2
T1
C2
C3
C4
C5
C6
R1
C7

L1	spider, OD 5.7”, 9-point, 36 ½t, 660/46
	L ~160 uH
L2	spider, OD 5.85 », 9-p, 37 ½ t, L~171 uH
C1	450 pF
C2	435 pF
C3	500 pF
C4-5	50 pF trimmers
C6	0.1 uF
C7	0.47 uF
R1	500 kΩ
T1	Vintage RCA

Sleeper #19

4S 1 LC, SL, LD

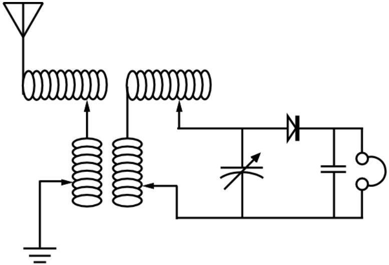

Loose coupler with loading coils

Primary and secondary circuits each have separate loading coils in order for the secondary circuit to be tuned as well. Loading coils at right angles to tuning coils.
In secondary circuit, condenser placed across both tuning AND loading coil.

Kinzie 47b

3S 3 LD, LC

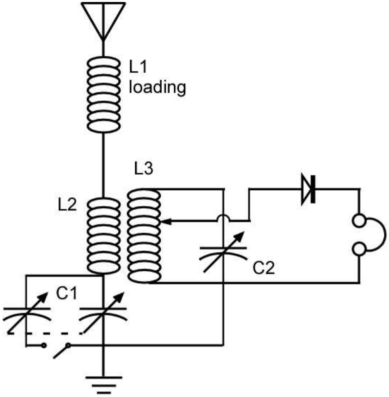

Double tuned set with loading coil

NOTES:

KEVIN'S WEBSURFER HANDBOOK II
FOR CRYSTAL RADIO

DESIGN – THEORY – TESTING

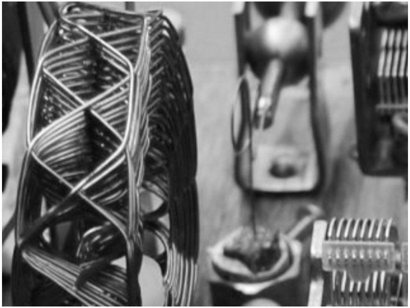

Kevin Smith
2011

Printing / Binding Instructions

1. Choose “fit to page” in print menu
2. Print document double sided on letter size paper
3. Cut the entire printed document in half
4. Fold over making sure the page numbering is continuous
5. For the cover: Print just the first page on card stock paper
   Cut the cover in half as well
6. Assemble the covers on the document
7. Punch the left side for a binding, spiral or comb as desired

http://www.lessmiths.com/~kjsmith/crystal/catalog.shtml

KJ Smith

NOTES:

NOTES:

INTRODUCTION

The contents of this handbook consist of a number of web-based documents that provide significant and detailed insights into the design and theory of crystal radio. For the crystal set hobbyist, it is not sufficient to build from kits or merely follow published instructions or designs. It is important to gain deeper understanding of the theory behind the sets and learn from experienced members, to learn about the tricks and tips they employ to construct interesting and advanced sets. The web is a marvelous source of data and information. Many long-time crystal set builders have created dedicated sites to disseminate information and resources, to share their creations.

To further my understanding I have spent many long hours surfing the many excellent crystal radio websites. The wealth of information is outstanding, even a bit overwhelming at times. Pages range from simple to advanced, up to pages where one likely needs their EE degree to comprehend the material. When I first started I spent time on the basic web pages, but as I grew comfortable and began to consider more advanced features for my sets, I spent ever more time on more advanced web pages. Whenever I found material especially well presented and useful, I would print it out into notebooks and study for hours at a time. With time I came to find that within the large number of web resources available, there exists a canon of sites particularly well suited for study. While my personal needs and tastes ultimately dictate those pages I find useful, I do believe that these pages are the best of the lot.

In this booklet I have taken the sections that I spend most time studying and have arranged them into sequence from basic theory and design considerations to more advanced material on particular aspects of the hobby. Finally I include several sites with advanced methodology for testing and experimenting with the sets I have built. Each section is useful and understandable to the hobbyist with moderate to good experience. The handbook will help the beginner to quickly get up to speed and allow the experienced builder to find endless new ideas. This is not a book of Hookups or circuit designs, that is covered in my Catalog of Crystal Hookups.

All of the material in this handbook is copyright for which I have not sought permission. Therefore this is not presented for publication or copy. It is only my personal resource. I encourage anyone finding this copy to pursue ON THE WEB the web pages identified within. I include the name of the author and web address of each section. I wish to sincerely thank every author presented for their excellent pages and ask forgiveness for my editing into this handbook.

Kevin Smith
2011

www.lessmiths.com/~kjsmith/crystal/cr0intro.shtml

Dielectric losses in coils.

When a coil is wounded on a coilformer, there will be dielectric losses in the former. Also in the insulation of the wires there will be dielectric losses, especially when windings are touching each other.

For the coilformer it is important to use a material with low DF value, and to use as few as possible of this material between the windings so the capacitance and losses are kept to a minimum.

Other dielectric losses.

Also in the wires in the receiver dielectric losses can occur. For instance if a wire carrying radio frequencies is placed very close to a other conductor. The two conductors form a capacitor, with the wire insulation as insulator, in this insulation losses can occur.

We can reduce these losses by placing the wire at some distance of the other conductor (some cm. or more).

If we connect this capacitor to a coil, then we have a LC circuit. If the coil has no loss, the maximum Q of the LC circuit will be:

Q= Zc / Rs = Zc / (Zc . DF) = **1 / DF**

So, if we use a capacitor with nylon insulation, the Q of the LC circuit will have a maximum value of: Q= 1 / 0.026 = 38. This is a very low value.

So, a LC circuit with airspaced tunercapacitor has a infinite Q??? No, because there are more losses, like the resistance of the coilwire, the resistance of the capacitor plates, the dielectric losses in the coil, etc.

It is also not possible to use a tunercapacitor with only air insulation. The rotor (turnable part) and stator (non turnable part) must also be connected to each other by insulation material. If for instance 4/5th of the capacitance is caused by air insulation, and 1/5th by nylon insulation blocks, then the Q will increase by a factor 5 compared to a full nylon insulation. So the Q can then be 5x38=190.

The higher the frequency of the circuit, the higher the percentage of capacitance caused by the insulationblocks, and the lower the Q. Instead of nylon it is better to use a insulator with a lower DF value, and preferable also a low εr value.

For a high Q the following is also important:

Use the insulation material only on places where the distance between rotor and stator is high, this helps to reduce the capacitance and losses caused by the insulator.

Don't use more insulation material then necessary.

CONTENTS:

1 **Crystal Set Design 102**
Al Klase
11 **Crystal Radio Antennas**
Owen Poole
31 **Crystal Radio DXing**
Steve McDonald
45 **The Crystal Radio Revisited**
Smith-Kettlewell
75 **Designing an air-core Inductor**
Kenneth Kuhn
89 **Antenna Matching**
Kenneth Kuhn
93 **Variations: Slider-Tuned Set**
Owen Poole
97 **A Crystal Radio Wave Trap**
Owen Poole
103 **Bandspreading Techniques**
Owen Poole
109 **Selectivity Enhancement Circuits**
Dave Schmarder
113 **Contra Wound Coils**
Dave Schmarder
125 **Basket/spiderweb coil design**
PA2MRX
139 **Dedicated Impedance Matching**
Darryl Boyd
159 **Designing a DX Crystal Set**
Mike Tuggle
169 **Lyonodyne Version 17 Crystal Set**
Mike Tuggle

177 **Testing your Crystal Set**
Owen Poole
187 **Crystal Set Testing**
Charles Lauter
199 **Radio Test**
Kevin Smith
213 **Measuring the Q of LC circuits**
Dick Kleijer
217 **Diodes**
Dick Kleijer
227 **Diode I/V Characteristic**
Kevin Smith
233 **Experiments, antenna tuner**
Dick Kleijer
237 **Experiments with a detector unit**
Dick Kleijer
239 **Resonant Circuit**
Kenneth Kuhn
249 **Tuned LC circuits**
Dick Kleijer
257 **Coupled circuits**
Dick Kleijer
263 **Insulation materials**
Dick Kleijer
269 **Notes**

Nylon	3.33	0.026
Plexiglas:	2.76	0.014
Polyethylene (PE)	2.26	<0.0002
Polypropylene (PP)	2 .25	<0.0005
Polystyrene (PS)	2.56	<0.00007
Polytetrafluoroethene (PTFE, Teflon)	2.1	<0.0002
Polyvinyl chloride (PVC)	2.88	0.016

A capacitor wil have at certain frequency a impedance of:

Zc=1/(2.pi.f.C)

Zc= impedance of the capacitor (Ohm)

pi=3,14

f= frequency (Hertz)

C= capacitance of the capacitor (farad).

Because of the losses caused by the insulator, it looks like there is a resistor connected in series with the capacitor. This is the series resistance (Rs) of the capacitor.

If the insulation of the capacitor has a certain DF value, the series resistance wil have a value of:

Rs=Zc . DF

In the ideal case a capacitor has no loss, if we send a AC current through the capacitor there is no loss of energy. In practice there will always be losses, if we send a current through the capacitor a part of the energy will get lost by heating up the capacitor.

These losses are caused by:

a- The resistance of the conductors (plates), these losses are left out of consideration here.

b- The dielectric losses in the insulation. One insulator gives more dielectric losses then the other, this is indicated by the dissipation factor DF. The dissipation factor "DF" is sometimes also called: "tangens delta".

The lower the value of DF, the better the quality of the insulator.

In the following table some values of εr and DF for different insulation materials.

Material	Dielectric constant (εr) at 1 MHz.	Dissipation factor (DF) at 1 Mhz.
Vacuum:	1.000000	0.00000 ?
Air:	1.000585	0.00000 ?
Acrylonitrile butadiene styrene (ABS)	2.8 -- 3.8	0.006 -- 0.011
Glass	4.84	0.0036

CRYSTAL SET DESIGN 102

Copywrite 1997, 1998 - Alan R. Klase - All rights reserved.
http://www.skywaves.ar88.net/xtalset102/xtalset102.htm

Introduction:

Welcome to Crystal Set Design 102. It is assumed that you've already completed the 101 level course elsewhere, and know, at least, something about electronics in general and crystal radios in particular. If not, hit the books. The stuff in your public library will get you started. The early chapters of the "Radio Amateurs Handbook" are especially concise and approachable.

I've been a little disappointed at the lack of meaningful crystal set technical information on the web and in current literature. This is my attempt to at least partially remedy this situation. This work is the result of about ten years part-time investigation of crystal radios from an engineering perspective. Serious development of passive receivers pretty much came to an end with the introduction of reliable vacuum tubes around 1920. A lot of crystal sets, both commercial and home brew, have been designed in the interim, but most are mediocre performers. So the mission turned out to be one of rediscovering the the secrets of the age when spark was king.

Getting Started:

My approach is to build crystal radios out of quality vintage radio parts. They are available in great profusion at amateur radio "hamfests" and antique-radio meets (and in my garage). If you don't have access to these sources, I'd suggest you start at Radio Shack: Buy their Crystal AM Radio Kit, 28-177, for $6.99. This is actually not a bad crystal set, and it contains a coil, a variable capacitor, a germanium diode, and, most importantly, a reasonably sensitive high-impedance ear phone. You're also going to need an antenna. This generally means wire up in the air. The attic may be the next logical alternative. Apartment dwellers may be in trouble unless you're near an AM radio station, or can arrange a "stealth" antenna of some sort.. Again, if you don't (yet) have a junk box, get Radio Shack's Outdoor Antenna Kit, 278-758, for ten bucks. Wire has always been expensive, and hard to find retail, so keep your eyes open for bargains. Another RS item, that's almost indispensable, is a set of mini-alligator jumper cables, 278-1156, 10 for $3.99. These are how you make temporary "breadboard" hookups while experimenting with new circuits. You'll also want basic electronic hand tools and a soldering iron.

Circuits:

The simplest radio you can build is just a diode detector and a headset. With a reasonable antenna and ground you will hear the strongest stations, albeit all a once. This is not much of a

---

INSULATION MATERIALS

Dick Kleijer crystal-radio.eu
http://www.crystal-radio.eu/enisolators.htm

For insulation materials used in crystal receivers the following electrical properties are important: insulation resistance, dielectric constant and dissipation factor.

Insulation resistance.

This is the resistance between two conductors having a insulation material in between. In most cases the insulation resistance is high enough for use in crystal receivers.

Dielectric constant.

Two conductors with insulation in between form a capacitor. If the conductors have the shape of parallel plates then we can calculate the capacitance as follows:

$$C=0.0885\ \varepsilon r\ A / d$$

C= capacitance in pF (picofarad)

εr=dielectric constant of the insulator (ε = the Greek letter Epsilon)

A= area of the plates in square cm. d= thickness of the insulation in cm

The value of εr indicates the increase of capacitance compared to air insulation.

Dissipation factor.

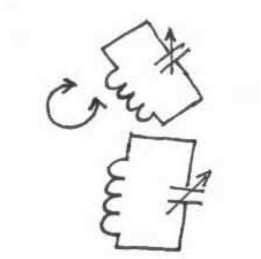

Capacitive coupling
Coupling is adjusted by a variable capacitor with a very low value (e.g. 1 pF). There must be no coupling via the magnetic field, so the coils must be placed at a right angle. Capacitive coupling is also useable when magnetic coupling is difficult, for instance when using ringcore or potcore coils.

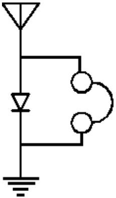

radio. but it will give you some indication that you have enough signal strength to continue experimenting.

The primary problem with the above set is that it offers no selectivity. We'll solve this problem by building a tunable filter. This will generally consist of a coil and a capacitor forming a tuned circuit. Either the capacitor or inductor, or both need to be variable so circuit can be tuned to different stations. The classic values are 250 uH and 365 pF to tune the broadcast band. This is the basic crystal set schematic you'll find almost everywhere. It works better than just a diode, but has some serious short comings that are easily remedied.

The Importance of Impedance Matching:

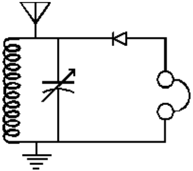

In a crystal set, all the audio power that arrives at your ear drum came from the distant transmitter. If the transmitter is hundreds of miles away, the amount of power captured by even a good antenna is reckoned in nanowatts . At each point in the set we must strive to transfer at least a reasonable percentage of the available power to the next stage. Perfect impedance matching is not necessary, don't fret over a 2 to 1 mismatch, but let's eliminate as many of the 10 to 1 and 100 to 1 mismatches as we can.

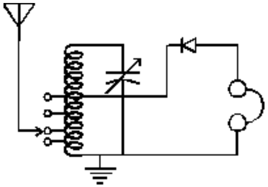

Single-Tuned Sets

In the case of the simple set in the last example, lets do two things: Tap the antenna input "down" on the coil. The impedances of the antenna and tuned circuit will vary with frequency, so it's a good idea to provide multiple taps and a selector switch or movable jumper. As a starting point tap the coil at about 5%, 10%, 20%, and 50% of the total number of turns from the ground end of the coil. Secondly, connect the detector to the 50% tap. This does two things, both of them beneficial: It provides a better match to the detector when it's connected to the usual sort of crystal set headphone that has an impedance of about 10K ohms, resulting in a louder signal. It's also reduces the loading on the tuned circuit, increasing it's Q and consequently it's selectivity.

These improvements result in a better than average crystal radio. With 50-75 feet of wire up in the air, you should hear daytime 50KW stations out to 40-50 miles, and night-time skywave stations will come in form hundreds of miles away. This is essentially the circuit I arrived at for my Cub Scouts a few years back. (See *The Den Two Crystal Radio* in *Crystal*

There are several methods for adjusting coupling. The methods shown below are mostly used in crystal receivers.

The coils are side by side
Coupling is adjusted by varying the distance between the coils.

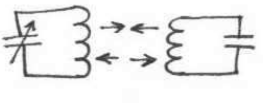

Coils behind each other.
Coupling is adjusted by varying the distance between the coils.

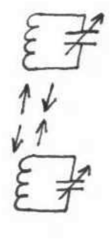

One coil is turnable.
Coupling can be adjusted by varying the angle between the coils. With both coils in the same direction, the coupling is maximal. Turning the angle towards 90 degrees, the coupling will decrease. With a angle of 90 degree, there is no coupling at all.

Overcoupling
The coupling is too high because the distance between the coils is too small.

The bandwidth is too high.

The curve shows two peaks, the higher the coupling , the higher the distance between the peaks, and the deeper the dip between the peaks.

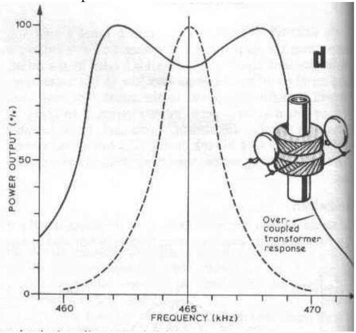

Set Projects published by The Xtal Set Society) You'll also notice it's almost exactly the same circuit used in the Radio Shack set

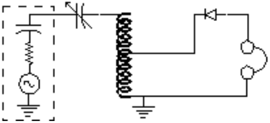

Another effective way to match the antenna to the tuned circuit is to use a variable capacitor as shown in the drawing. A cap in the 300-500 pF range is appropriate. Tune in a station then change the coupling and retune and see if there's an improvement. Some sets have the coupling capacitor in the ground lead instead of the antenna lead. Electrically it's the same thing, but sometimes it's more convenient for mounting and grounding and eliminating hand capacitance effects.

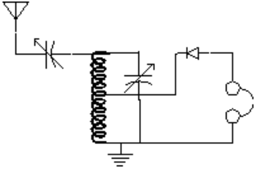

With the previous circuit, connected to a good sized antenna, you'll find you can set the tuning capacitor to it's minimum value, or even remove it completely and still get good performance. To understand how this "series tuned" receiver works, I've inserted the equivalent circuit of the antenna in the drawing. Any Macroni antenna less than a quarter wavelength long appears to be a capacitor in series with a small resistor, known as the radiation resistance of the antenna, and an RF voltage source. This is almost always the case as a quarter wavelength, even at the top of the broadcast band, is 187 feet. Our aim is to make as much RF current as possible flow from our antenna. When the series value of the antenna and tuning capacitors and the inductor are tuned to the frequency of interest, the inductive and capacitive reactance's cancel out, leaving only the DC resistance of the inductor. This results in maximum current flow in the tuned circuit, and maximum voltage applied to the detector. A receiver of this sort will usually require a variable inductor to cover the entire broadcast band. In such cases, keeping the detector connected to the optimal point on the coil presents challenges.

One classic variation on the series tuned receiver, known as the "two-slider tuner", dispenses with the tuning capacitor entirely. Instead a sliding contact on the coil varies the inductance in the antenna circuit. A second slider connects the detector at the best point on the coil. Total inductance should probably be the better part of a millihenry

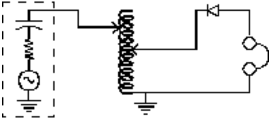

Critical coupling
There is maximal powertransfer.

The bandwidth is higher than in a single circuit, and also also higher than a undercoupled circuit.

The response curve is flat in the top over a smal part.

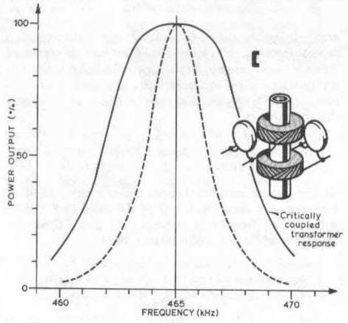

Undercoupling

The coupling between the circuits is too low, because the distance between the coils is too high.

There is no maximum powertransfer from one coil to the other.

The bandwidth is higher compared to a single circuit.

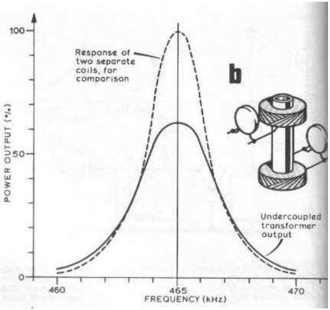

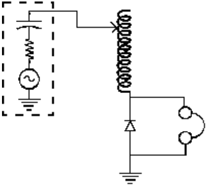

Yet another, historically common circuit, connects the detector in series with the variable inductor. This is not such a great idea from the standpoint of impedance matching, but it is simple. Series tuned sets all have difficulties with short antennas, because their capacity is low.

Another way to implement the series tuned set, that improves the match to the detector, is to connect the detector and headset across a second coil in the series circuit. For the broadcast band, this coil has a fixed value of approximately 80 uH. The variable capacitor is optional, especially if the larger inductor is continuously variable.

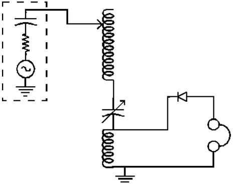

Single-tuned crystal sets, whether series or parallel tuned, leave a lot to be desired in terms of selectivity. Yes, the nose selectivity can be improved by increasing the Q of the tuned circuit, but the skirt selectivity remains hopelessly broad. The obvious solution is to use more than one tuned circuit. The point of diminishing returns is between 2 and 3 circuits for a crystal set due to cumulative losses and tuning difficulties.

The Two Circuit Receiver

The classic solution is the "two circuit" tuner. The antenna circuit is series tuned by a variable capacitor and an inductor, while the detector circuit is connected to a parallel LC circuit. The amount of mutual inductance, or coupling, between the primary and secondary circuit is generally made variable. This allows light coupling to be used to obtain the sharpest tuning, while increasing the coupling increases sensitivity at the expense of selectivity. This is essentially the same architecture used to great effect in the communication receivers of the wireless era. If you're seeking better performance for your crystal set this increased complexity is well worthwhile.

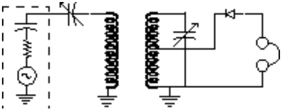

There are several way to implement variable coupling in the two-circuit set. The navy style "loose couplers" used a secondary coil that telescoped inside the primary. Other possible schemes include variable taps on the low end of the secondary coil, link coupling between the two coils, or the use of a small variable inductance common to both circuits.

COUPLED CIRCUITS

Dick Kleijer crystal-radio.eu
http://www.crystal-radio.eu/enkoppelfactor.htm

In a receiver having two tuned circuits, both circuits must be tuned to the same frequency.
De 2 coils of the tuned circuits must have a certain distance in between, via the magnetic field around the coils, the circuits are coupled to each other. The distance between the coils determines the degree of coupling, the smaller the distance, the higher the coupling.

The -3dB bandwidth of a coupled circuit is higher than with a single tuned circuit.

The suppression of frequencies at some distance from the tuning frequentie is in a coupled circuit better than with a single circuit, this is good for suppressing strong local station.

In the pictures below you see the response of the coupled circuit at different degrees of coupling. For comparison the response of a single tuned circuit is shown as a dotted line.

The pictures are taken from a radiohandbook, and are about a coupled filter for 465 kHz, but the theory is usefull for all coupled circuits, so also for crystal receivers with two tuned circuits.

Here we see, that at a constant value of Rp, the Q will decrease with increasing frequency (f). A parallel resistance across the circuit will especially give a reduction of Q at higher frequencies.

The value of Q will both depend on series and parallel resistance, it is possible that a LC circuit gives a increasing Q at increasing frequency (because of series resistance) and than Q will reduce (because of the parallel resistance). In this case we have a peak in Q somewhere in the frequency band.

Often the losses caused by the parallel resistance are the highest, and we only see a reduction of Q at higher frequencies in the medium wave band. Than there is also a peak in Q, but this occurs at a frequency lower than we use.

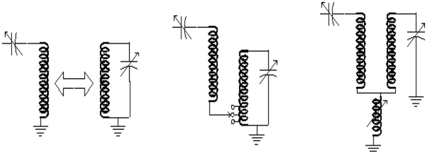

Variable coupling schemes for two-circuit tuners.

Additional Details

The primary inductance will want to have a maximum value of about 500 uH to reach down to 530 KHz. However, a lesser value in needed to reach the 1600 KHz end of the band. I've been building coils with five or six evenly spaced taps. This lets you tune around for an optimum match to the antenna.

Install a DPDT switch to allow the primary circuit to be operated in a parallel-tuned mode: The input end of the primary capicator is grounded, and the antenna attached to the top of the coil. This allows effective operation with short antennas.

If a non-fixed detector is used in a double-tuned circuit, it's a good idea to include a buzzer to generate a local signal to adjust the detector. The circuit is a low-voltage mechanical buzzer, a battery, a push-button switch, and a one-or-two-turn link to the secondary coil. A low voltage relay with it's normally-closed contact wired in series with the coil is a good substitute for a buzzer.

insulators of the tuning capacitor Dielectric losses in materials placed near the coil or tuning capacitor, look here for more information about this subject. If the capacitor plates are not clean: dielectric losses in the dirt and oxide on the capacitor plates.

Also there can be a leaking resistance in insulators, e.g. by moisture.

All these losses together makes a parallel resistance (Rp) across the LC circuit.

Magnetic losses

This occurs when a magnetic material (iron) is placed near the coil. Non magnetic materials (plastic, wood, aluminium etc.) don't give magnetic losses.

Why is the Q factor not constant for all frequencies?

The series resistance Rs gives a reduction of Q. If we leave other losses out of consideration, the Q will have a value of:

Q= 2.pi.f.L / Rs

If the value of Rs is constant, the value of Q will increase with increasing frequency (f). A series resistance in the circuit will especially give a reduction of Q at lower frequencies.

On the other hand: if we only look at the losses caused by the parallel resistance Rp, the Q will have a value of:

Q=Rp / (2.pi.f.L)

Rs=(2.pi.f.L) / Q

The Q factor of a unloaded LC circuit is determined by the following factors: series resistance in the circuit, parallel resistance across the circuit and magnetic losses.

The series resistance in the LC circuit,

the lower the series resistance the higher the Q. The total series resistance in the circuit (Rs) is the sum of:

The wire resistance of the coil, thicker wire or litzwire with much strands helps to reduce wire resistance.

The resistance of the capacitor plates, silvered plates gives the lowest resistance.
Plates with oxide gives more resistance than clean plates.

Contact resistance between rotor and frame of the variable capacitor, preferable the variable capacitor has a spring connected to rotor and frame, this provides a low resistance. When the contact is made with a slidercontact at the rotor, this must be clean and free of oxide.

The parallel resistance across the LC circuit,

the higher the parallel resistance, the higher the Q.

Parallel resistance across the circuit (Rp) is caused by dielectric losses.

There are: Dielectric losses in the coilformer Dielectric losses in the insulation of the coilwire Dielectric losses in the

---

CRYSTAL RADIO ANTENNAS

Owen Poole

http://bellsouthpwp2.net/w/w/wuggy/antenna.htm

If there is such a thing as the perfect antenna, then why does every issue of every ham magazine have an antenna article? Fortunately, for the crystal radio nut, most ham antennas are less than optimum for general listening use, and require special care and feeding due to the need to have the transmitter and antenna somehow matched to near perfection and for the antenna and its associated parts robust enough to handle power without arcing and sparking while also keeping the transmitted signal out of the shack and also out of the neighbor's tv set.

Crystal radio antennas have some special requirements:

a. large frequency range
b. large signal gathering capability
c. efficient in getting and then transferring the received signal to the set.

Your basic problem is normally just getting enough wire in the air to have a resonant circuit. For 99 percent of us, the answer is: you can't. Even the shortest resonant antenna for the top of the AM broadcast band, a quarter wavelength *Marconi* antenna, would have to be about 137 feet long; the bottom of band calls for about 425 feet. Unless you are one of the fortunate few to be able to erect a bona fide antenna farm everything you do from this point on will be a compromise of one sort of another. This means you have to start thinking about your antenna system rather than just what you hang up outside the house.

First, the antenna itself: I recommend you use stranded copper wire, at least 20 gage. Try to get at least 50 feet of it "out the window" and up in the air as high and as far from the house as possible. My experience is that anything shorter gets real hard to tune, and gives away a whole bunch of signal to boot. Longer is better, higher is better. Use plastic strips or ceramic separators or whatever you can get to ensure the antenna wire is not touching anything that could be grounded, including the tree you may be using for a support. I usually just use stranded wire with a tough plastic coating, wrapping it over tree limbs as I go, and connect it at the far end with heavy nylon fishing line. **Do not go over or under power lines.** If you are coming in through the bottom of a window, as many of us do, be sure to use insulated wire for the final lead in. If you have to put it inside, such as up in the attic, make sure you are not hanging it under a metal roof or ridge vent, otherwise it could very well shield your wire from the signals you are trying to get. High straight antennas are probably best, but do the best you can, and don't be afraid to bend it in order to get more wire out there. Contrary to popular wisdom, you can tune a bent antenna. If you like to try out different stuff, put an alligator clip on the end in the shack to make quick connections to your latest xtal wonder rig. If you replace your antenna, or otherwise have room, a second antenna can come in handy, if for nothing more than comparison purposes. My second antenna consists of 50 feet of wire out the window, up to the roof, and over to a tree. This is about what I expect my students to get up, and I always use it for testing kid radios. If your main antenna is mostly long and horizontal, a shorter, and more vertical antenna might give some pleasant results,

The table below gives the value of impedance and voltage on the tap's of the coil. Also the current is given which can be given at the tap, the current will increase at a lower tap.

Tap	Voltage on tap	current	impedantie
100 %	100%	100%	100%.Rp
90%	90%	111%	81%.Rp
80%	80%	125%	64%.Rp
70%	70%	142%	49%.Rp
60%	60%	166%	36%.Rp
50%	50%	200%	25%.Rp

The used diode must also have about the same resistance as the impedance at the tap.

Because of the lower voltage at the tap's the diode efficiency will decrease however, and this will reduce the sensitivity of the receiver. The best option is the diode at the top of the coil, and the use of a load impedance with a high enough value.

Series resonance

In a series connected LC circuit, the impedance will be low at resonance. If there are no losses in coil and capacitor, the impedance will be zero Ohm at resonance. But here are also always losses, in serie resonance we keep a certain resistance Rs. The higher the Q, the lower the series resistance Rs.

The Q of a LC circuit will decrease when we connect a antenna or detector to it, this is because the antenna or detector gives extra parallel resistance to the circuit. By doing this the selectivity of the circuit will reduce.

Parallelresonance

In a parallel LC circuit, the impedance will be high in resonance.

If the coil and capacitor has no losses, the impedance would even be infinite at resonance. In practice this is not possible,there will always be losses, for instance in the resistance of the coilwire.

So, the impedance is not infinite, but has a certain value, it looks like there is a resistor connected parallel to the LC circuit, this we call the parallelresistance of the circuit "Rp".

Rp=2.pi.f.L.Q

Parallelcircuit with taps on the coil

As discribed above the parallel LC circuit has a certain parallelresistance.

For maximum sensitivity in a receiver, the load impedance (speaker or audiotransformer) must have about the same value. If we only have a speaker or transformer with a lower impedance, then we can connect the detector diode at a tap on the coil, instead of the top of the coil. The circuit will then not be loaded too much, and the Q and selectivity will not drop.

particularly with distant stations and higher frequencies. Since you will probably be using what is considered a random wire antenna, connected at one end to the set, it may be of some value to know that the horizontal part of the antenna "points" in the direction of best reception off the free end.

Next, get a good ground. This is the other half of your antenna, and a good earth ground is essential. By earth ground I mean a good electrical connection to the earth. (I once had a student who tried just laying the ground wire on the ground - didn't work) I know of several hams who invest more time, effort and money in their ground than any other part of their antenna system. Mine is currently a tight connection to a close by cold water pipe. If you have plastic pipes, this doesn't work. I have also used pipes hammered into the ground, the ground rod the telephone company provides where their service comes into the house, and even the screw holding on the cover of electrical plug cover plates in a pinch (this is usually electrical ground, and may even lead electrically to the earth somewhere - works better for some locations than others). I haven't resorted to burying assorted lengths of wire in the ground yet, but many dedicated types do. If you are located in the desert or on a granite mountain, and good earth grounds are hard to come by, you need to read up on dipole antennas, ground radial systems and counterpoises; good luck. If you can not make a short connection to a good ground, and your ground lead is long enough to act as an antenna itself, you might want to try this trick: use a piece of coaxial cable between the set and the ground connection, using the inner core for the connection; connect the shield to the core at both ends with a capacitor, say, about 0.01 uF or so, to shield the ground lead from rf. I haven't tried this myself, but it makes sense, I think.

Whatever you have done up to this point, your antenna is (1) too long or too short (usually the latter), (2) is not resonant at the frequencies of interest, and (3) will not efficiently transfer whatever signals you get to your set. Guaranteed. Despite all these deficiencies, it will probably work pretty well, at least on the stronger stations. If you are happy at this point, declare victory and quit. If, however, you have a free hand left with which to flog yourself while you tune with the other, or really want to go after them, read on.

The first quick fix you can make is to effectively add additional length to your antenna by putting a large coil of wire between the antenna and the set. Wind 50 to 100 feet of 20 to 24 gage wire around a core between 2 and 5 inches in diameter, and tap it about every 10 turns or so. Attach one end of this coil to the antenna. Use an alligator clip to connect the xtal set to various taps until you get the loudest signal. While this doesn't add much to the antenna in the way of length, it can make the antenna closer to resonance with the frequency you are trying to dig out.

Quick fix number two, which can be done after or in lieu of number one, is to add some capacitance between the antenna and the set. This seems to work best if your antenna is too long for the frequency of interest. Radio Shack uses this technique in some of their Bunch-in-One electronic project kits, using a fixed 10 pF capacitor for a short antenna and a 100 pF capacitor for long antennas. Short refers to antennas 50 feet or less in length. I prefer to use a variable capacitor if I go this route. Putting the capacitor in series with the antenna this way lowers the inherent capacitance of the antenna. One of the more popular designs a number of years back put the

---

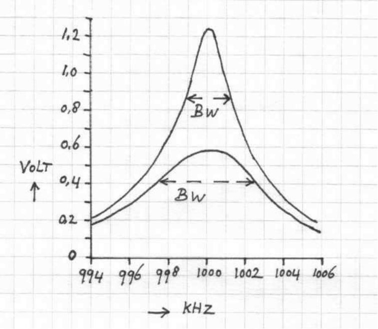

Curve of frequency respons for two tuned circuits.

At the upper curve, the bandwidth is 2.4 kHz.

The Q is 1000 / 2.4 = 416

In the lower curve the bandwidth is 5 kHz.

And here the Q is 1000 / 5 = 200.

The Q of a circuit can vary from less then 100 for coils made with massive wire, up to 400 or more for coils made with litzwire.

Bandwidth:

At resonance frequency the impedance of a parallel LC circuit will reach it's highest value, the voltage over the circuit will then also reach it's highest value.

Above and below the resonance frequency the voltage will decrease.

There are two frequency's where the voltage is 0.707 times the voltage at f.res.

One frequency is just below f.res, this is frequency fl.

And one frequency ia above f.res, this is frequency fh.

The voltage reduction to a factor 0.707 is a reduction of 3 Db

The current is there also reduced to 0.707, and so the power in the circuit is halve the power of f.res. (0.707 x 0.707 = 0.5).

The bandwidth of the tuned circuit is: **BW =fh - fl**

Circuit Q:

The Q of the circuit is:: **Q=f.res / BW**

The higher the Q, the smaller the bandwidth, which is better for separating adjacent stations. And also, the higher the Q, the higher the voltage which the received station gives over the circuit, so a more sensitive receiver.

variable capacitor in between the xtal set and the ground connection. If it is the same size as the main tuning capacitor in the rig, you can even gang them and make tuning a bit less tiresome, with little or no adverse effect on results. This actually works very well. If you are using a set that has the detector or the antenna attached directly to the top of the tank circuit, and don't want to make taps in the tank coil for better impedance matching, this fix might be(I should say is) essential, and is certainly worth trying. The problem with connecting the antenna to the top of the tank circuit coil is complex, to say the least. I observed that the first effect is to lower the resonant frequency of the tank circuit drastically, even with a very short (a few feet) antenna. If you're going to try to get away with this, and don't have a variable capacitor to spare, you can make a gimmick capacitor by twisting a piece of wire connected to the set around the bitter end of the antenna. I did it another way, using some 18 gage twin lead wire; turned out that the capacitance of the twin lead was about 1.2 pF per inch, and the circuit resonance from the antenna was not affected if you used enough twin lead (one side connected to the set, the other to the antenna) to give you about 1 pF per foot of antenna.

Quick fix number three, which is an excellent idea even if you don't do the first two, is to connect the antenna to a tap on the tuning coil. This gives a better match between the impedance of the antenna and the tuned circuit of the rig. Techtronics does this with their nice little crystal set, providing two places to connect the antenna (really three, but they don't mention the third tap, for some reason. It is between the other two.) Probably the absolute worst place, in my opinion, to have the antenna connected is at the top of the set tuning coil along with the connection to the detector, yet that is what several web page offerings would have you do. The tap you use will have to be determined experimentally, but I find that low on the coil close to ground helps.

Quick fix number four, which I recommend with fix number one: Use an antenna coupling coil . Wind the coupling coil on approximately the same diameter core as the main tuning coil, using about one turn for every six in the main tuning coil, and then put it in line (on the same axis) with the main tuning coil, with the ability to move them apart. You will get better selectivity, and can sharpen it up a bit more by moving the two coils apart - what the trade calls loose coupling. Oh yeah, the antenna and ground are connected only to the coupling coil, although you might try connecting the ground to the set as well. The "one for six" rule is just a rule of thumb. If you don't mind experimenting a bit, pick the number of turns on the coupling coil that gives you the best coverage across the band, particularly if you are going to use your set without a tuner. It will still be a compromise, and the best number of turns will depend on the length of your antenna and your location, but will give better results. Of course, you can dispense with the coupling coil if you are inductively coupling the antenna coil directly to the set's tank coil as in the basic tuner below.

If you have tried one or more of the tricks above, and are happy, fine. If not, I don't blame you, so it is time to get serious about tuning the antenna and coupling it to the set for optimum sensitivity and selectivity. In other words, it is time to talk antenna tuning and coupling methods, subtitled "how I went from 5 to 50 stations , and then to 98, heard on my xtal set".

TUNED LC CIRCUITS.

Dick Kleijer crystal-radio.eu
http://www.crystal-radio.eu/enlckring.htm

Frequency:

A tuned circuit made of a coil L (unit: Henry) and a capacitor C (unit: Farad) has the following resonance frequency:

f.res = 1/(2.pi.root (L.C))

Example: a coil of 0.2 mH (0.0002 Henry) is connected to a 500 pF tuner capacitor (0.000,000,000,500 Farad).

The lowest frequency we can tune to is 503 kHz.

If we turn the value of the tuner capacitor to a value of 48.8 pF, the resonance frequency will be 1611 kHz.

Here in Europe 1611 kHz is the highest frequency on medium wave.

So, with this LC circuit we can tune over the entire MW band.

In practice both coil and detectordiode will have a certain capacitance, so we must set the tuner capacitor to a lower value for tuning at 1611 kHz.

If the coil has too much capacity it is possible that we can't reach the highest frequency.

In this case we can add some space between the turns of the coil, this will reduce capacity.

$$R_{shunt} = 2 * \pi * 1\text{ MHz} * \text{loaded } Q * L \qquad \text{Eq. 8}$$

As an example, if we have a 240 uH coil and we want the loaded Q to be 20 at 1 MHz then the total shunt load should be 30,000 ohms. Assuming coil losses are small this means that the detector impedance should be 60,000 ohms and that the antenna impedance should be transformed (via a tap near the ground end of the coil) up to 60,000 ohms. This particular scenario is a practical one that can be built.

Concluding remarks

Although it is discussed in other chapters it should be mentioned here that we generally want to make the detector impedance as high as possible in order that we can use the highest inductance practical. This results in the maximum signal voltage applied to the detector thus lowering the losses in the detector circuit.

A significant factor to be aware of is stray capacitance across the inductor. This capacitance results from the natural physics of the winding (conductors separated by insulation). This stray capacitance is typically in the 10 to 50 pF range depending on how the inductor is wound and is in shunt with the tuning capacitance. The effect is to limit the upper frequency that the inductor can be tuned too. The effect is worse in large value inductors and that typically limits the maximum practical inductance to something less than 1 millihenry. There are some advanced winding methods that minimize stray capacitance but those are beyond the scope of this chapter.

A properly employed antenna tuner will do a lot of "heavy lifting" for you. First, it can make your antenna resonant at the frequency of interest, and bring up the signal level at the same time. Second, it will screen out a lot of unwanted stations, including those out of band ghosts. Third, it will transfer the signal efficiently to the main tuning and tank circuit, and do so such that you can vary the selectivity of your set as a bonus. This seems like a set of tall orders, but it isn't nearly as hard as it may seem. I will discuss mostly sets using air core solenoid tuning coils (that is, toilet paper rolls wrapped with a single layer of wire or the like), since that's what I usually use and have the most experience with. Before we start, let me say that an antenna tuner is not just another stage of passband filtering, such as the Miller 595 crystal set used. A key to a passband tuner design is when you see a two or more section variable capacitor, with one section in the gang working with one coil and the other section working with the second coil. Also let me add that an antenna tuner is essential for serious work.

Basic tuner : Wind a coil on a core the same diameter as your xtal set's main tuning coil. You may want to add about 30 to 50% more turns, but it should be at least as many turns as are on the main tuning coil. Put a tap at least every 10 turns (every 5 is better). Put the two coil axes in line, and connect the end of the antenna coil that is closest to the xtal set to ground. Attach the antenna to taps on the antenna coil using an alligator clip until you get the best signal.

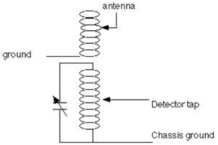

No direct connection to the xtal set is needed. You should find that you can separate the two coils by a distance of one coil diameter or more and get a good signal; in fact, moving them closer together may even make things worse. The closer the antenna is tapped to ground on the coil, the higher the frequency. You can leave the unused antenna coil turns "dangling", as it were, or attach the antenna to the other end and short the unused turns with the tap - both ways are acceptable. You should be able to notice a difference in signal strength of a station by moving just one or two taps along the coil. If you find that you never need to use more than half the coil for the best signal, even at the bottom of the AM band, don't worry; you may need them later. In this setup, you are essentially making the antenna appear the correct length, so you are in effect adding length to compensate for the short antenna. If you can't inductively couple the tuner to the set's tuning coil, you can attach your earth grond to the chassis ground of the set, and attach the ground end of the antenna coil to a suitable tap on the set's tank coil.

The Teenie Tuner: This is a variant of the basic tuner, using a coil wrapped around a ferrite core. Here's how to make one: Get a ferrite core of some sort, either a rod or a bar. Wrap some typing paper over it and glue the paper into a sleeve - put a spacer between the paper sleeve and the core while winding the coil so that the core will slide easily in and out of the coil after it is wound. Over the paper sleeve, wind a coil of some 50 to 90 turns of magnet wire in a single layer; no taps needed. Glue or tape the coil onto the sleeve. Now, mount the sleeve only on a support, such as a scrap of wood, or in line with the set's coil. Connect one end of the tuner coil to the antenna, the other to ground, and vary the tuning by sliding the ferrite rod in and out of the coil. You should get a range of

be used. The details would be a chapter in of itself but I can tell you that the practical absolute minimum inductance that is useful for a resonator in the AM broadcast band is around 40 microhenries as it is a challenge to obtain a high Qu for inductances less than this in that frequency range. Thus, only the last column of Table 2 is useful.

Determination of maximum practical inductance: This calculation is done the same way as before except that the lowest value of loaded Q is used. That results in a higher inductance. Table 3 is a summary calculation. Again, the top two rows are for reference only.

1,000,000 Hz, resonant frequency
50,000 Hz, BW **20.0 Qloaded**
Rdetector transformed across entire inductance

Qunloaded	**2K**	**5K**	**10K**	**20K**	**50K**	**100K**
1,000,000	8.0E-6	19.9E-6	39.8E-6	79.6E-6	198.9E-6	397.9E-6
1,000	7.8E-6	19.5E-6	39.0E-6	78.0E-6	195.0E-6	389.9E-6
500	7.6E-6	19.1E-6	38.2E-6	76.4E-6	191.0E-6	382.0E-6
250	7.3E-6	18.3E-6	36.6E-6	73.2E-6	183.0E-6	366.1E-6
125	6.7E-6	16.7E-6	33.4E-6	66.8E-6	167.1E-6	334.2E-6

Table 2: Maximum inductance in Henries

Conclusions:

The conclusions from studying Tables 2 and 3 are that low values of inductance are needed for low impedance circuits and that high values of inductance are needed for high impedance circuits. Typical values of inductance used for the resonant circuit in crystal radios ranges from around 100 microhenries up to around 700 microhenries with more common values in the 200 to 400 microhenry range.

We can work Equation 7 backwards to determine a good value for the net shunt resistance across the coil at 1 MHz as follows.

two across the coil as Rsignal. It was discussed previously that losses in the inductor can be represented by an effective resistance across the entire coil. We will refer to this loss resistance across the coil as Rloss. The parallel combination of Rsignal and Rloss is the net resistance across a lossless inductor. We will refer to this net resistance as Rshunt. From Equation 5 we can write

XL = Rshunt / loaded Q Eq. 6

Thus, using a frequency of 1 MHz

Lmin = XL / (2 * π * 1 MHz)= Rshunt / (2 * π * 1 MHz * loaded Q) Eq. 7

The detector impedance is typically in the 2,000 to 50,000 ohm range. With antenna matching the net resistance, Rsignal, will be half this range as discussed above. Table 2 shows a summary calculation for the minimum value of inductance to use to achieve a loaded Q of 100 for various loads either directly or transformed across the coil. The Q of 1,000,000 row represents essentially infinite Q and the Q of 1,000 is not normally attainable and are shown for reference only.

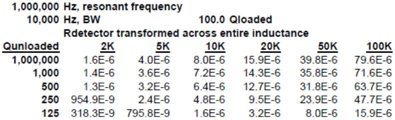

1,000,000 Hz, resonant frequency
10,000 Hz, BW **100.0 Qloaded**
Rdetector transformed across entire inductance

Qunloaded	**2K**	**5K**	**10K**	**20K**	**50K**	**100K**
1,000,000	1.6E-6	4.0E-6	8.0E-6	15.9E-6	39.8E-6	79.6E-6
1,000	1.4E-6	3.6E-6	7.2E-6	14.3E-6	35.8E-6	71.6E-6
500	1.3E-6	3.2E-6	6.4E-6	12.7E-6	31.8E-6	63.7E-6
250	954.9E-9	2.4E-6	4.8E-6	9.5E-6	23.9E-6	47.7E-6
125	318.3E-9	795.8E-9	1.6E-6	3.2E-6	8.0E-6	15.9E-6

Table 2: Minimum Inductance in Henries

All of the inductances in Table 2 are small –some very small. But our target was the absolute minimum inductance that could

inductance of about 10:1 this way, effectively lengthening your antenna as he rod is pushed further into the coil. As with the basic tuner, you can set the tuning coil alongside the main coil, in line with it, or connect it's ground end to the antenna tap of the set. Just for grins, affix the teenie tuner to the back of your am superhet and see the improvement in your reception.

Series tuner : Here we add a variable capacitor, the same size as the one in your xtal set, in series with the antenna coil used in the basic tuner. Traditionally, you put it between the end of the antenna coil and the ground connection. Now you have two tuner adjustments to make (not counting distance of the antenna coil from the main tuning coil). For the first time, I recommend you get the set tuned to a moderately strong station, then put the tuner capacitor in the mid position, and then find a tap which allows you to hear the station. Then you can use the capacitor to bring in the station more sharply. You will probably find you need to use more turns of the coil than you did with the basic tuner, since the capacitor "shortens" the antenna. Ultimate signal strength may not show much improvement, but it should help selectivity. This setup is a favorite of the "pros". As before, the antenna and main tuning coils are in line. You should have to make fewer tap changes with a series tuner compared to the

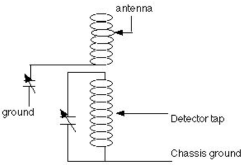

basic tuner, since the capacitor does some of the tuning for you.

Parallel tuner : Use the same coil and capacitor as with the series tuner, except this time the capacitor is connected to both ends of the coil. One end of the coil, the one closest to the set, is connected to ground, and the antenna is connected to the coil tap that gives the best signal; incidentally, I have seen a few circuits where the antenna just connects to the other end of the coil from the ground - I don't recommend this, even though it works fairly well, sometimes, maybe, I think. Adjust the capacitor as needed to peak the signal. This setup is often used when the antenna is "very" short.

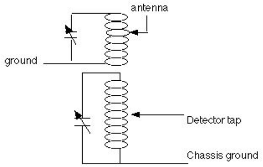

My recommendation to add turns to the antenna coil will sometimes let you use series tuning across the whole band. If your two coils are the same length, I think you will find that the series arrangement works best at the high end of the band, and the parallel arrangement is needed at the bottom; the longer your antenna, the lower in frequency you can use a series tuner. As with the series tuner, a parallel tuner gives you more selectivity than does the basic tuner.

Although there are an infinite number of combinations of inductance and capacitance that will resonate at a desired frequency there is only a limited range of practical values that can be used. An excellent question to ask is if there is an optimum inductance. If there is then that is what we will use. There is not a specific answer to that question other than there is an identifiable range of inductance that provides the best overall results. We will determine the extremes starting with the minimum practical value that will result in the highest allowable loaded Q followed by calculating the maximum practical value that will result in the lowest acceptable loaded Q. Any practical value of inductance between these two extremes can be used.

Determination of minimum practical inductance: As discussed previously, the highest loaded Q of resonance that is practical to use is such that the 3 dB bandwidth is about 10 kHz. This permits the full double sideband width to be detected. In extreme cases where we are willing to forfeit some audio fidelity to achieve better selectivity the bandwidth can be reduced to about 6 kHz –but we are not going to consider that case here. At 1 MHz (the rough center of the AM broadcast band), a 10 kHz bandwidth occurs with a loaded Q of 100. The antenna and detector circuits may be operating from taps on the inductor for the purpose of impedance matching (the method to design the taps is discussed in another chapter). The impedance at each tap can be transformed to an equivalent impedance across the entire coil. We will assume that the antenna circuit and detector circuit are impedance matched via this tapping process (this provides the much needed maximum power transfer from the antenna to the detector). Thus, for the impedance matched condition the antenna and detector impedance will transform to the identical impedance across the inductor. We will refer to the net parallel impedance of these than about 50 so that the bandwidth is not too narrow. An excessively narrow bandwidth makes tuning very sharp and also distorts the audio. The bandwidth needs to be at least 10 kHz with 20 to 50 kHz being common. The following table illustrates the maximum, typical, and minimum loaded Q values to use across the AM broadcast band. Keep in mind that the unloaded Q of resonance should be significantly higher – preferably in the hundreds. The maximum Q value is for the minimum bandwidth of 10 kHz. The typical Q is a good value to try to achieve although realistically it is likely to be lower. The minimum Q has a fairly wide bandwidth and will not be able to separate stations well.

Frequency	QLmax	QLtyp	QLmin
0.5 MHz	50	25	10
1.0 MHz	100	50	20
1.7 MHz	170	85	34

Table 1: Practical ranges for Q of resonance in the AM broadcast band

The resonator has an effective resistance across it that represents coil losses. This resistance will be referred to as RQ. Ideally, RQ is infinity (i.e. no losses) but realistic values are typically in the many tens to hundreds of thousands of ohms. There are two loads on the resonator –one is effective source resistance, RS, of the matching network to the antenna and the other is the effective load resistance, RL, of the diode detector and audio transducer. For maximum power transfer from the antenna to the audio transducer the source and load resistances should be equal.

How to determine the required inductance

A broadband tuner : This is not a classic design, to my knowledge, and I stumbled on it by accident, but it does a lot of heavy lifting, and helped me well over the single digits in stations heard. I thought I was copying faithfully a design out of an ARRL handbook for dealing with random length antennas, but ended up with something a bit different, and won't try to analyze how it works, but here it is: Don't sweat the coil/capacitor specs; what you have in the series or parallel tuner will work fine. However, the connections shown going to receiver antenna and ground are actually connected to that coupling coil I mentioned up in quick fix number four. The tuner's orientation with respect to the set is not a factor, since the signal is transferred via the coupling coil. It will generally be closer to the main tuning coil than described in the three tuners above. I like this tuner arrangement for casual listening, as I don't have to play with it much over the band. Find the settings that work best for you. It is a particularly good setup to use when you are finding out about a new xtal set, and don't want to have to mess with the tuner every hundred kHz or so, as you do with a more selective tuner, or don't want to "tailor" a tuner just for the new rig. This setup works well when you can't inductively couple the antenna tuner to the set. It also works without the coupling coil by connecting the end of the antenna coil to a low tap on the main tuning coil of the set.

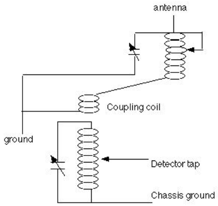

All of the above tuning methods can be accomplished using a simple coil and capacitor arrangement, which I made up according to the schematic below:

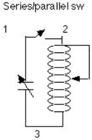

Points 1, 2 and 3 are connecting points for antenna/ground as needed. Antenna can be connected directly to coil taps if desired.

A selective tuner: Richard O'Neill used this tuner effectively in the 1999 DX contest (and everybody used it in the 2000 contest), and it is perhaps best known in connection with the Tuggle Circuit, which is described in detail in a couple of fine (and inexpensive) publications of the The National Radio Club / DX Audio Service. Essentially, it is the same as a parallel tuner, except, you use a ganged capacitor, such as a dual 365 pF capacitor (some are still around), or even a "superhet" variable capacitor with an antenna and an oscillator section. One section (the larger one for a superhet capacitor) is in parallel with the antenna coil, and the other is between the parallel circuit and ground. I have tried this and it seems to work fine, and is very sharp. As before, try attaching the antenna to taps on the coil for best results; I find attaching it to the top of the coil works fine. Several popular crystal sets have used same double capacitor arrangement. If you don't use ganged capacitors, tuning this baby can be really tricky, since the two (make that three) capacitors interact a lot. When hooking this up to a ganged dual capacitor, the antenna and the top of the antenna coil are hooked to one stator. The bottom of

Q = reactance / Rseries Eq. 6

Inductors and capacitors have internal losses which appear as a resistance load either in series or shunt with the component. Losses in picofarrad capacitors used at broadcast frequencies tend to be very small and thus these capacitors have a high Q (typically many hundreds) or as it is more commonly referred to in capacitors, low dissipation (the reciprocal of Q). Inductors have the dominant loss. This loss is commonly the series resistance of the wire (including skin-effect) but can also include losses in any magnetic medium used in the inductor. For good crystal radios we need inductors to have a Q in the hundreds.

Resonant circuits are not used in isolation. The antenna is coupled to the resonator as well as the diode detector. Both of these represent a resistive load which lowers the Q of resonance. This is referred to as the loaded Q. That may sound bad but it is actually fine. The loads represent power that we are interested in. We prefer that the resonator have low losses or high-Q. That permits the maximum transfer of power from the antenna to the detector.

Since any resistive load such as the antenna radiation resistance or the detector resistance lowers the net Q (i.e. Qloaded) of resonance, it is important to start with a high unloaded Q known as Qu. The Qu of an LC resonant circuit is primarily determined by the Q of the inductor as its losses tend to dominate. Typical values for the inductor Q range from around 50 to over 500. If the tuning capacitor is of very good quality (i.e. air or silver mica insulation, etc.) then its Q will typically be in the 500 to over 2000 range. Although we like for the unloaded Q of the resonator to be as high as practical for low losses, it is desirable for the loaded Q to be no more

The required tuning range, Cmax / Cmin, of a variable capacitor can be determined using Equation 2 as follows

$$\frac{\mathrm{Cmax}}{\mathrm{Cmin}} = \frac{\dfrac{1}{(2 * \pi * \mathrm{Fmin})^2 * L}}{\dfrac{1}{(2 * \pi * \mathrm{Fmax})^2 * L}} = \frac{\mathrm{Fmax}^2}{\mathrm{Fmin}^2} \qquad \text{Eq. 3}$$

For the AM broadcast band Fmax is generally 1.7 MHz and Fmin is 0.54 MHz. Thus, the ratio of Cmax to Cmin should be around 10.

Tuning inductor

The classic ancient method of making a variable inductor was to provide a movable wiper to connect to individual turns of the winding. Modern methods use a movable ferrite rod to vary the inductance. Similarly to the development of Equation 3, the ratio of Lmax to Lmin is the square of the desired frequency ratio and works out to be around 10 for the AM broadcast band.

Discussion of Q

Q can be defined in several ways and each way is compatible with the others as shown below.

Q = resonant frequency / bandwidth3 db Eq. 4

Q = Rshunt / reactance Eq. 5

the coil goes to the rotor (frame), and the ground to the second stator. Note that with decent Q coils, you will experience some hand capacitance effects. I simply used some 9/32 " aluminum tubing on the 1/4 " capacitor shaft, then used about 2 inches of 1/4 " dowel rod as an extension, and put the tuning knob on the dowel; worked fine. Sometimes, when tuning the top of the band (higher frequencies), I can spread the tuner out a bit by making it a series tuner. All I have to do is disconnect the first stator from the top of the coil. Yeah, this means the ground is not attached to the rotor as in a proper series tuner, but it works fine, particularly if you have already done the hand capacitance elimination step.

As long as you have gone this far, why not try connecting the detector to the top of the detector coil instead of to a tap. Works pretty well with a crystal earplug, and cuts out the shortwave intruders. With magnetic phones, using an impedance matching transformer per Ben Tongue's page is rally the thing to do. However folks, with the tuner you already have an additional filtering stage in operation, and can sacrifice a bit of selectivity with the high detector tap; anything to get rig of the intruders in a pinch, particularly in the early evening when they can take over your set.

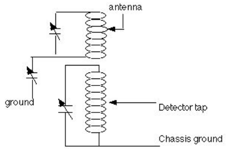

While I mentioned putting the two coils on the same axis, you can also place them side by side, as Mike Tuggle does with his Lyonodyne #17 (getting 2500 mile contacts), using an antenna tuning coil wound with Litz wire on a ferrite rod; just make sure they are parallel for maximum coupling. You can vary the coupling by either separating them from each other or by rotating one of them from the parallel position, as is done with variocouplers. Minimum coupling occurs for a given distance when the axes of the two coils are perpendicular.

For you people using toroids in your xtal set, I have only to offer what you see in the "quick fixes". My tuners essentially rely on loose coupling to the main coil, and I really haven't played with the other type tuning coils enough. Al Klase uses toroids in his very fine crystal set, and you should look at his page for coupling to that. Also look at his work on XTAL Sets 102 for ideas.

In your quest for sensitivity and selectivity, you may now be ready to try a wave trap, sometimes known as a QRM coil, to get rid of offending, overbearing bandmaster stations. The premise is simple; trap the strong signal and it won't bury the weaker ones. While I can separate two local stations that are 40 kHz apart, with one of them at least 3 dB louder than the other one, I often can't eliminate the louder station entirely, and can often hear it in other parts of the dial except when I am between stations. Then there is the signal from the secret Radio Habana transmitting site, which pokes through at odd times. Crystal radios tend to go with the signal that is the loudest, hanging on until something better comes along. A decent wave trap just might be the answer. . One basic design is around a high resistance coil, 32 gage wire or so, in parallel with a variable capacitor, and a 15 turn link around the coil which connects antenna to set (or wherever you want to

Note that the Q=100 curve has a band-pass of 10 kHz which is about the minimum that can be used to recover the entire double-sideband signal.

Tuning capacitor

Generally, when the capacitor is variable, the inductor is fixed and visa-versa. Mechanical variable capacitors (often 10 –365 pF or similar) were very popular in older times but are more difficult to find in modern times. A modern substitute can be made using one or more rotary switches as will be illustrated later.

The well know equation for the resonant frequency of an inductor and capacitor is

$$F = \frac{1}{2 * \pi * sqrt(L * C)} \qquad \text{Eq. 1}$$

where
F = resonant frequency in Hz
L = inductance in henries
C = capacitance in farads

For a given inductance the required value of capacitance can be determined from

$$C = \frac{1}{(2 * \pi * F)2 * L} \qquad \text{Eq. 2}$$

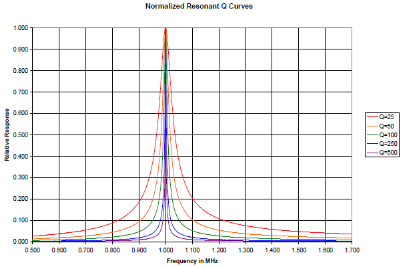

Figure 1: Normalized resonance curves at 1 MHz

Figure 2 shows a zoomed in version of Figure 1.

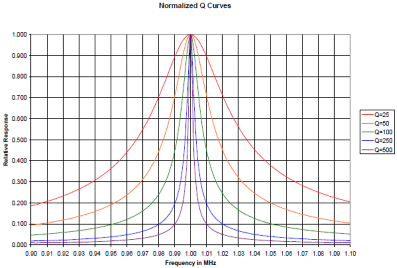

Figure 2: Normalized resonance curves at 1 MHz

try it. The idea is that the resonant LC circuit not only is set into motion by the offending signal but also dissipates it in the high resistance windings of the coil. The link lets the other signals bypass the trap with little loss. Here is a link to a nice short page on wave traps. http://pages.ca.inter.net/~funk/trap.GIF These traps, with no coupling coil, work best for out of band intruders when used as in-line traps in the antenna line, or when inductively coupled to your set's tank circuit. I think the one I use is better for in-line use, and you can find a description, picture, and discussion of it and trapping in general here . I realize that a wave trap might not be considered technically a part of the antenna system, but as long as it is in front of the main tuning circuit, I will consider it so. Wave traps can be used as either rejectors, to block an offending station, or as acceptors, which can be sometimes used to boost a desired station or, more commonly, to add another stage of filtering to increase overall selectivity.

Antennas for the short waves, some notes and preliminary tests (a/o Nov 01):

When you go for the short wave bands, and I think you'll be pleasantly surprised when you do, you will find that even your 50 foot long antenna is more than a quarter wavelength long on even the 49 meter broadcast band. This is a bonus, but also calls for some different ways of holding your head on tuning the antenna to resonance and then coupling it to the detector circuit. If you're like me, you use some sort of coupling coil to the detector tank circuit, and a simple coupling coil will only make the antenna appear longer. My first attempts to couple the antenna to the set and tune it at the same time were somewhat frustrating, especially since I was trying to tune from about 5.8 - 15 MHz. A combination of coils and capacitors that worked on one band did not on another. Then Steve Holden made this timely posting on the Crystal Set Radio Club forum, and I got a better appreciation of the problem:

In the October 1934 edition of Short Wave Craft, Editor Hugo Gernsback writes about Short Wave antennas on page 344. On page 345, fig. 6 shows a schematic of an antenna tuner that is a Coil in series with the antenna, with a parallel Variable Cap (C1); under the Coil is another Varicap (C) that goes to ground.

Gernsback writes, "In figure 6 we have endeavored to strike a happy medium, that is, an antenna that can be tuned and will respond to the SW broadcast bands, 19, 25, 31, 49 meters. The length of the antenna from A to B, that is the flat top and including what leadin may exist, should be 75'. The ground lead should be as short as possible, not over 4-5' long. With C, C1, and L it is possible to tune this antenna to any of the 4 bands mentioned. On some bands it will be a Hertzian and on others it will function as a Marconi. On the 49 meter band a Hz ant. will need to be 80' long. By setting C to a minimum the system becomes, in effect, not grounded. Therefore L and C1 can tune it up to an effective length of 80'. In the 31 meter band this antenna functions as a 3/4 wave Marconi. C should be adjusted to approximately half, and the tuning done with C1. In the 25 meter band,it is also a 3/4 wave marconi, and it is necessary that the effective length be reduced to 60'. This is accomplished by the adjustment of C with C1 set to minimum capacity. In the 19 meter band it is possible to make

RESONANT CIRCUIT

by Kenneth A. Kuhn

http://www.kennethkuhn.com/students/crystal_radios/resonant_circuit.pdf

The resonant circuit plays a very important role in a radio receiver. Its primary purpose is to be a tunable narrow band-pass filter for selecting a desired station while rejecting undesired stations. An important secondary purpose is to be a means for transforming the
low impedance antenna up to the high impedance detector.

A resonant circuit is comprised of an inductor and a capacitor, one or both of which may be variable. Energy flows back and forth between the inductor and capacitor at the resonant frequency. Energy in the inductor is stored in a magnetic field and energy in the capacitor is stored in an electric field. If there were no losses, this cyclic process could continue forever. In reality there is some loss each time the energy moves. This loss is expressed as a resistance. A common term used to account for this loss is Q which is the inverse of the loss. Thus, a high-Q resonance has very low loss and the cyclic process can continue for many cycles. A common example is a bell which continues to ring long after being hit.

Figure 1 shows the frequency response of a resonant circuit at 1 MHz for a range of Q.

voltage across the LC circuit

The voltmeter measures the detected DC voltage

The load resistor is formed by resistor R parallel with the voltmeter with 10 MΩ input resistance

The value of capacitor C is: 100 nF

The RF signal is supplied by a DDS signal generator, the signal is via a coupling coil coupled to the LC circuit.

All measurements are done at a frequency of about 1500 kHz.

At 1500 kHz, the impedance of the unloaded detector unit is: 1.97 MΩ.

There are 4 (combinations of) diodes tested.

HSMS-282K Schottky diode (2 diodes in one case parallel)

HSMS-282K Schottky diode (1 diode)

5082-2835 Schottky diode (2 diodes parallel)

5082-2835 Schottky diode (1 diode)

the antenna function as a 5/quarter wave Marconi. The necessary length here is 75 feet so we can use C for tuning and C1 should be set at minimum." In the next paragraph he talks about coil making, 20 turns DCC (double cotton covered} on a 1" diameter form. He ecommends inductive coupling to the radio. From context I think the caps are standard MW units.

In other words, what Gernsback was saying was that I should use a Tuggle type tuner, but independently tune the two capacitors. Since I was already tuning through some 10 MHz of bandwidth with a 180 degree rotation of the dial, the thought of trying to get three capacitors to tune to the same frequency was somewhat daunting, particularly since my simple sets are typically calibrated by "feel". So, I started playing around a bit more, and finally (I think) came up with a coil/cap combination that gets me back to one knob antenna tuning. I used a coil of about 5 uH in parallel with the 160 pF section of one of the little polyfilm tuning capacitors, attached the antenna to one end of the tank circuit, and the ground to the 60 pF section of the capacitor. I got a dip using a grid dip oscillator across the tuning range of the detector tank circuit, and it seems to work pretty well with both my 50 foot and my 160 foot antennas. Now I am at least back to two handed tuning - hi.

Here is another approach I tried, and it works about as well (if not better), and I don't really know why. I make a tank circuit duplicating the tank in the detector circuit, and then attached just the antenna to the top of it. Darned if it didn't work well, and I was able to separate the two coils by as much as a couple or more coil diameters. Haven't tried it yet, but it might be that you can get the two circuits to track, and use a ganged capacitor. Mike Tuggle gets his best results without a ground, so maybe I'm on to something here. Yeah, someone out there is chuckling as they read this, having known -this all along, but I'm the one with the web site, so I'll take the credit - hi.

Final warning! When you do anything to the antenna system, you can expect it to affect some other aspect of the crystal radio's tuning. When you are ready to start making adjustments, have a comfortable chair, a tall drink, maybe something to make notes with, and the time to be patient. And oh yeah, I didn't even mention the possibility of using different antennas, different grounds, loop antennas, no ground at all.... and did I mention lightning arrestors?

If you noodle around on the web a bit, you will find some excellent antenna articles. I am just trying to get you started. As I said before, every antenna is a compromise - just try to get a setup that works for you, and then you can get back to building and testing your own crystal set creations.

A word or two about small loop antennas, portable antennas and the like: You will find at times references, plans and testimonials in praise of the small loop antennae. If you live near strong stations, they may work for you. Where I live, they are almost useless, and do require a ground. I personally believe that if you are going after the distant stations with a loop you will be very disappointed. These are specialty antennas and have limited value - but where they do work, they get high marks. As an aside, I suspect that if you can't get at least 30 feet of honest antenna out the window, you might

---

EXPERIMENTS WITH A DETECTOR UNIT

Dick Kleijer crystal-radio.eu
http://www.crystal-radio.eu/enantunittest1.htm
On this page measurements are described on my detector unit 1 .

In this detector unit is as diode used the HSMS-282K (2 diodes parallel), and the load resistance by transformer unit 1 is 1.6 MΩ.

For the choice of diode and load resistance I assumed that maximum sensitivity would occur if load resistance and diode resistance were about equal to the impedance of the unloaded LC circuit (this assumption is however not correct).

Via some experiments I will now determine if the sensitivity and selectivity can be improved by the use of another detector diode and/or another load resistor.

Also frequency shift of the detector circuit is measured as a function of the voltage across the circuit.

Test setup used for the measurements on this page.

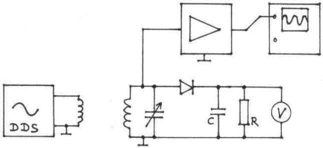

The oscilloscope measures via the measuring amplifier the try a large diameter loop antenna instead. If you are new to this business, start at the top with an honest antenna (please try for at least 50 feet out the window), and then, when you start getting experience, and have a set you know works, try what you like. Trying to hook up a first crystal set with a small loop is an invitation to frustration.

Made with a resistor of 50 Ohm, a coil of 20 uH, and a capacitor of 200 pF. The 50 Ohm output resistance of the signal generator is in parallel with the 50 Ohm resistor.

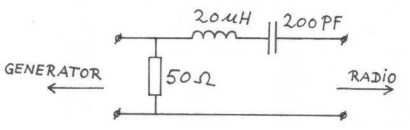

The dummy antenna makes the receiver input is driven with a impedance of 25 Ohm in series with 20 uH in series with 200 pF.. This is about the impedance of a average longwire antenna. With the dummy antenna connected to the generator (but without the receiver connected), the output voltage of the generator is adjusted to 50 mV peak-peak.

Now with various settings of C1, the frequency range of the circuit, the voltage across coil L1, and the Q (across coil L1) are measured.

The measured voltages are Volt peak-peak.

The frequency range can be slightly influenced by the measuring amplifier which was connected across coil L1.

The circuit L1,C2 has a high impedance (e.g. 1 M.Ohm). But the antenna has a low impedance (e.g. 10 Ohm), tuning capacitor C1 forms a impedance match between this high and low impedance.

With a certain value of C1, there wil be maximum power transfer from antenna to the circuit L1,C2. Then there is maximum voltage across the coil L1, and maximum sensitivity of the receiver. At low frequencies we must for instance adjust C1 to 100 pF, and at high frequencies to 20 pF, but these values are depending on the (lenght of) antenna we connect to it. For the circuit Q however, the lower the value of C1, the higher the Q. More information about this, you wil find here.

Coil L1 is wound with litzwire 660x 0.04mm (660/46 AWG), on a polypropylene former. This coil is described here as coil L12, only the outermost winding is removed to reduce inductance a little bit. This reduced the total wirelenght from 15 to 14.5 meters.

The frame of this antenna unit is made of 8mm polyethene sheet.

Tuning capacitor C2 is driven via a 1:5 vernier drive, so we can tune it very accurate. Tuning capacitor C1 has no vernier drive.

Experiments:

During the measurements the antenna unit is connected to the signalgenerator via a dummy antenna.

The dummy antenna.

CRYSTAL RADIO DXing

Steve McDonald

http://members.shaw.ca/ve7sl/crystal.html

If you are anything like most radio 'nuts', you probably had a crystal set when you were a kid. Thinking back to those good old days, I recall that I could only hear two local stations on my crystal radio. I had always believed that you just couldn't hear very far on a crystal set.....until I ran across this picture!

It was a picture of Al Klase's wonderful "Crystal Radio DX Contest" station! I was amazed to learn that it was even possible to hear 'DX' on a crystal radio and, apparently, there was a large group of very active crystal radio DXers. I was hooked on learning more about this fascinating possibility!

PUSHING THE BOUNDARIES... With some further investigation I turned up the 2001 contest winner, Mike Tuggle in Hawaii. Mike had logged 63 stations and Cuba, all on the AM broadcast band! It turned out that Mike, after years of experimenting, had developed an extremely efficient DX set, which he called the "LYONODYNE 17".

Mike's elegant set is not only a work of art but also 'state of the art', using large diameter basketwound Litz-wire coils and high quality, low loss ceramic tuning capacitors. To be 'legal' in the crystal set community, the radio must be completely "passive", meaning no amplification of any form is allowed. The only power used must come from the antenna, making these DX Contest results even more amazing!

LOCATION! LOCATION!... Crystal-set DXers are up against two problems:

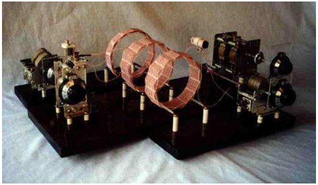

SELECTIVITY versus SENSITIVITY. If you live in a city or near a large city that has several 'blowtorch' AM signals, your main focus will be on separating signals so that the weaker DX can be heard between the blowtorches. On the other hand, if you live in the country, you don't have to worry so much about

EXPERIMENTS WITH AN ANTENNA TUNER UNIT

Dick Kleijer crystal-radio.eu
http://www.crystal-radio.eu/enantunittest1.htm
Here some measurements done with my "antenna unit1".

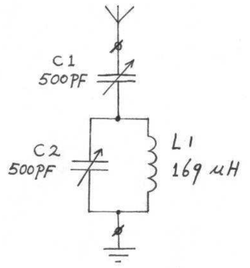

Schematic of the antenna unit 1.

The antenna unit has a tuned circuit L1, C2.

If we only look at L1 and C2, then the tuning range is 550 - 2184 kHz. But if we also connect the antenna and earth, the frequency of the circuit will decrease, so we can also receive the lowest mediumwave frequency of 530 kHz.

Variable capacitor C1 and the antenna and earth are also a part of the tuned circuit, but the antenna and earth also give reduction of circuit Q.

Kevin Smith
09/2011

---

separating strong locals; you can run things 'wide-open' enabling you to log some of the weaker DX signals. Living away from the city certainly makes life simpler for crystal radio DXing but don't despair if you are a city-dweller. One of the DX Contest winners lived in San Jose and still managed to log plenty of DX!

QUICKY BASICS

The diagram shows a simple basic crystal radio. In the city, with a good antenna, such a set will hear plenty of locals but it is unlikely to hear any DX. Take it out to the campground for the weekend however and even a simple set like this is capable of logging DX. Since most of us live in or near the city, we need to look at ways of transforming the basic set into a DX machine!

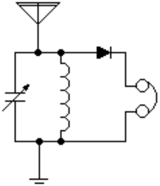

Basic Crystal Radio

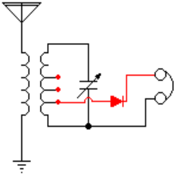

Loose-coupled Antenna Circuit

In order to separate the strong locals, the tuned circuit (L-C) must have as high a 'Q' as possible. Placing the diode and headphone load at the top of the circuit will result in strong signals but poor selectivity. The impedance of the headphones will load down the tuned circuit, lowering its 'Q'. The result is very poor selectivity. Tapping the diode/headphone load at a lower point in the tuned circuit results in less 'Q-lowering' of the circuit. The result is a higher degree of selectivity. You don't, however, get a free ride! The higher selectivity diode tap point results in less overall voltage being fed to the headphones. Signals are weaker but easier to separate! The challenge of building an efficient tuner is to find the right balance between selectivity and sensitivity!

There are a few more improvements that will help the 'basic set'. More sensitivity as well as more selectivity can be realized in the antenna circuit. A separate antenna coupling coil that will couple the antenna into the main tuning coil will result in a further overall improvement. If the position of the antenna coil can be adjusted in relation to the main tuning (detector) coil, a degree of coupling can be reached that offers a good balance point between adequate sensitivity and improved selectivity. Often the antenna coil and the detector coils are mounted on a long rod or dowel that allows easy adjustment of the 'loose-coupling' circuits. Also note that improved detector tuning can be achieved by optimizing the diode 'tap' location on the main tuning coil.

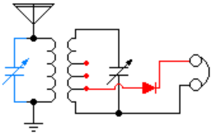

Parallel-tuned Antenna Coupler

Further improvements in both sensitivity and selectivity can be achieved by actually tuning

---

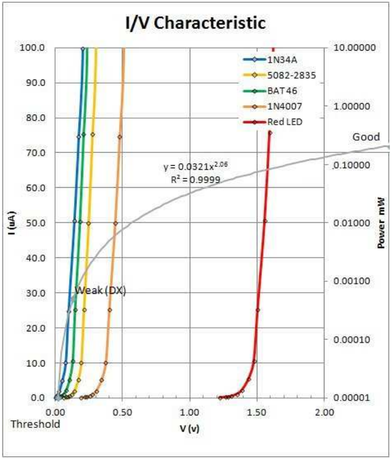

This plot at left takes some actual diode I/V measurements rather than the models presented below and plots them along with the above RF voltage across the LC circuit versus RF power in mW. Here one sees that pushing for very very low Vf (via high Is / low n combinations) pushes the limit of detectable signal power from the antenna. At such minute power levels I imagine the reverse leakage current becomes significant and probably offsets the low operating point gains sought.

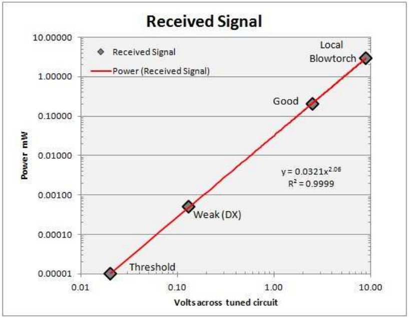

The above is an interesting plot taken from the text discussion in "Crystal Set Analysis" by Berthold Bosch. It presents received signal strength across the tuned circuit for various scenarios from threshold audibility to local blowtorch. This plot should be considered a single example specific to his location and antenna/ground system. In the text he describes his antenna an an inverted L 43m long (140') and 10m high (32'), an excellent antenna most of us do not have the real estate to erect, but offsetting this is a poor ground with Rg = 210ohms. Given the offsetting conditions I would imagine this is a good generic example of what received signal voltage and power levels presented across the LC circuit to the diode are likely to be. Bosch cited 40nV as the threshold of audible detectability (impedance matched conditions with 16kohm RF impedance / 4kohm DC phones) and for this plot I pushed it back to 20nV as it gave a superior regression fit.

the antenna system to resonance. Your type of antenna system as well as the part of the AM band that you are tuning will determine which system works best. The series-tuned antenna coupler is best suited for tuning large antennas, especially those with large horizontal runs ("top hats"). The parallel-tuned antenna coupler is better suited for tuning shorter antennas, such as vertical wires, masts or whips - those without top hats. A series of test clips or a simple switching system can be used experimentally to determine what works best for you. *Making this effort is important as it goes a long way towards avoiding initial disappointment with the crystal set's performance.* Such a system as this is called a 'double-tuned' set. With care and attention to construction, such systems are capable of hearing DX signals.

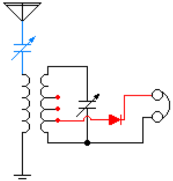

Series-tuned Antenna Coupler

Adding a tuning capacitor in series with the antenna permits a long antenna to be tuned very effectively by a parallel-tuned antenna coupler. In fact, this is the method-of-choice for many crystal set listeners. To avoid the complication of an extra tuning knob, a single-knob, two-gang variable capacitor can be used as shown here. The added variable capacitor can be placed instead in the ground leg. This 'enhanced parallel-tuned' coupler can be very effective in peaking up desired signals while knocking down nearby interference. I have found the use of a tapped coil in this circuit to be very beneficial in achieving optimal Q for the portion of band being tuned. One might even decide on using separate 'high-band' and 'low-band' coils for this circuit.

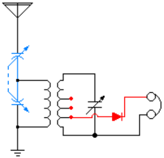

Enhanced Parallel-tuned Coupler

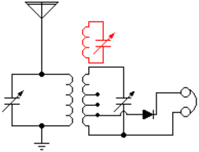

Loose-coupled Wavetrap

Although you may now be able to hear some DX, chances are you are still bothered by at least one (if not several) local 'pest' signals. Further gains in selectivity can be had by the addition of another tuned circuit, called a 'wavetrap'. By coupling this circuit to the detector coil, this

---

Additionally I have plotted RF signal Rs as orange circles from published coil unloaded Q (see my section on Coil Q). To convert from Qu to Ql I have made the assumption that loaded Q is about 1/5 unloaded Q for low-end sets while in performance sets it may approach 1/2.5 the unloaded Q. This is only an estimate then, but starting from actual data. To make the conversion to Rs I simply plugged the data into the standard equation R = 2pi*f*L*Q using a scaling factor to divide down Qu to Ql from high Q (1400/2.5) to low Q (100/5). This provides a nice visual display of the expected range for Rs in many crystal sets. I plot the Rs against the value of Ql used. Note that these data are for the case at about 1MHz and inductances between 200 and 300uH, as published.

Matching the diode to the tank is a matter of finding a diode with both sensitive qualities (low n) and an Rd close to the Rs presented to the diode (as per Ben Tongue's suggestions for Peak and Square Law detection). Those diodes with Is's in the 100 - 250 nA (100-280 kohm for n=1.1) or so range have the possibility to match while connected to the top of the tank coil without the use of Q-lowering taps, while untapped matching to high-end high-performance big-Litz baskets will require diodes with Is's in the 35 - 60 nA (500-800 kohm for n=1.1) range. Most "typical" germanium diodes have high Is values (>500, >60kohm) and will require a tap. Diodes such as FO-215, BAT 46, 1N60 and GAZ51 can be matched to many tanks without taps. For tanks with high quality Litz coils one will be using HP 5082-2835 or 1SS98 diodes, generally 3 to 4 in parallel to lower the Rd to the desired range. (Curiously, I have not found any diodes with Rd values in the 300 to 4500 kohm range.) Note that both resistance and impedance are frequency-specific so a match at one frequency will not remain matched across the broadcast band.

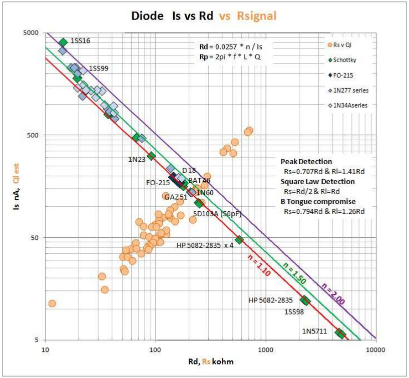

The plot above shows the relationship between diode Is and Rd. Rd is calculated via the equation Rd = VT * n / Is where VT is the thermal voltage = k * T / q = 0.0257V at room temperature and Is and n are from the measured diodes. k is Boltzmann's constant = 1.38E-23 J/K and q is the electron charge = 1.609E-19 coulombs. On such a plot I can show lines of constant n, and plot the values for individual diodes for which I have spent considerable effort to determine the parameters Is, n and Rd (where Is is the diode saturation current, n is the ideality factor, and Rd is the diode resistance). On the plot it should be evident how changes in n or Is impact a diodes' Rd.

circuit can be tuned to the same frequency of the pest signal, absorbing much of its signal-killing strength at the detector, while not adversely affecting the strength of wanted signals. More than one wavetrap may be employed to tackle more than one pest signal. With a high Q wavetrap, very loose coupling is usually best.

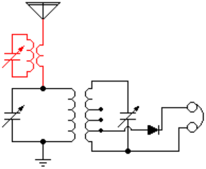

In-line Wavetrap

Another method that some find successful is the use of an 'in-line' wavetrap in the antenna circuit. The smaller secondary winding can be wound either on top of the primary or on one end. One or two in-line wavetraps can often be used to effectively knockdown the strong pest signals.

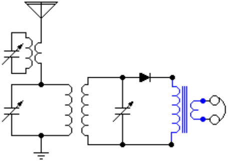

Impedance-matched Output

Another important improvement can be made in the headphone circuit. If you are using the old standard style of headphones (Brush, Baldwin etc) they must be high impedance phones. The modern low-impedance 'walkman' style phones will load-down the detector circuit too much, seriously degrading selectivity and signal strength. Even high impedance phones will cause unwanted loading of the detector circuit. Using a low-loss audio transformer with a very high impedance primary between the phones and the detector will reduce the problem and often great gains in selectivity can be achieved. Look for a high quality audio transformer with a primary impedance of at least 100K ohms and a secondary to match your phones. Try to avoid the small printed circuit transformers used in transistor radios - this is a case of the bigger, the better.

One of the greatest improvements made on my own DX set was changing from a pair of vintage high-quality Baldwin phones to a pair of headphones made from an old Navy 'sound-powered' handset. Most crystal radio DXers are utilizing this style of headphone. They are extremely sensitive! Old handsets are always available on e-Bay and can often be had at a bargain price. Often the earpiece transducer as well as the microphone element are identical and a single handset will yield a nice set of phones. The elements can be removed and placed inside an old set of headphone covers or hearing protectors. I think that going 'sound-powered' was the best single improvement that I was able to make when my DX set was evolving

---

Why know the values of Is, n, and Rd of your diodes?:

by Kevin Smith

http://www.lessmiths.com/~kjsmith/crystal/deyeball.shtml

I have been attempting to get a handle on the main diode parameters Is, n, and Rd and how they impact the operation of a crystal radio. My method of choice is to use a graphical approach, plotting charts where variations between plotted parameters become apparent and easy to visualize. I have developed an interesting chart from the measurements above that compares many diodes with respect to Is, n, and Rd, and to understand how this fits into a matching situation with a fuzzy indication of typical tank impedance limits.

Ultimately one seeks a diode with an Rd that matches with and conjugate to the impedance it sees from the tank (+antenna). In the following discussions I will start with the cross-plot of Is to Rd and then proceed to a look at how received RF power plays a part in the selection and why low Is and n are desired.

OA81 (germanium)	0.0225	800	34
OA95 #1 (germanium)	0.0272	600	45
OA95 #2	0.0221	821	33
OA95 #3	0.0271	604	45
OA95 #4	0.0304	502	54
AA116 (germanium)	0.0441	256	106
AA119 #1 (germanium)	0.0320	461	59
AA119 #2	0.0363	370	73
AA119 #3	0.0428	272	100

The HSMS282K is the same as the HSMS2820, only the HSMS282K has 2 equal diodes in one package.
The Is value of the HSMS282K, the HSMS286K and the 5082-2835 is lower than the value in the datasheet, this has also been noticed by other people.

Ben Tongue wrote me, that the Is value of the 5082-2835 has been reduced over the years by the manufacturer.
Also temperature has big influence, I measured at 18 °C, in datasheets the Is value is given at 25 °C.
A increase from 18 °C to 25 °C will increase the Is by 60 %.

BUILDING YOUR DX SET... Start by thoroughly checking out the many links shown at the bottom of the page. There is a huge amount of good information available on the web. The 'Yahoo Crystal Radio Group' as well as the 'Rap'n Tap' group both provide an interactive forum for asking questions. I would suggest that you consider an experimental loose-coupled test bench. This will allow you to easily make coil changes and adjust coupling between circuits. A fine example of a loose-coupled set is shown below. It was built by Larry Pizzella in San Jose and was used by him to win the 1999 DX Contest!

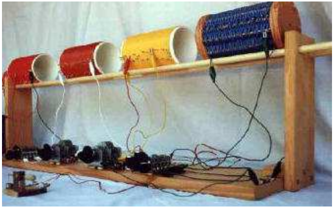

Although the loose-coupled coil rack is not visible in the photo, my own DX set was built along these same principles.

Many different coil styles, wire diameters and antenna tuning circuits were tried over the space of several months.

This is the schematic diagram of the final arrangement used in my 2002 DX Contest set.

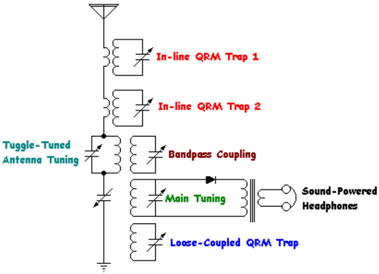
In-line QRM Trap 1
In-line QRM Trap 2
Tuggle-Tuned Antenna Tuning
Bandpass Coupling
Main Tuning
Sound-Powered Headphones
Loose-Coupled QRM Trap

VE7SL - High-Performance DX Crystal Tuner

This is a 'triple-tuned' arrangement. There are vernier reduction drives on all tuning capacitors as tuning is very sharp on all circuits. Because of the great abundance of 'blowtorch' signals

More information about measuring the Is you can find on the website of Ben Tongue , in his articles number 4 and 16.

I measured the Is value of several diodes, and calculated the diode resistance RD.
Also some European germanium types are measured.
Several diodes are measured of the type OA95 and AA119

Diode	VD (Volt) at 1 µA	Is (nA)	RD (kΩ).
HSMS282K	0.1341	7.9	3428
HSMS282K 2 parallel	0.118	14.5	1867
HSMS286K (1 diode)	0.1116	18.3	1479
5082-2800	0.1871	1.14	23756
5082-2835	0.1464	5.04	5373
5082-2835 2 parallel	0.1289	9.5	2850
BAT 82	0.136	7.3	3710
BAT 85	0.0686	90.8	298

Measure the voltage across the diode (VD).

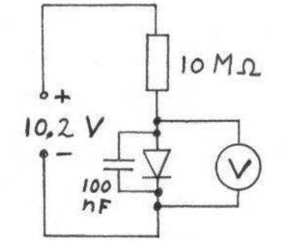

Circuit diagram for measuring the Is value of a diode.
The voltage across the diode is about 0.2 Volt.
The voltage across the resistor is about 10 Volts, so the current is about 1 μA.
The voltmeter must have a resistance of at least 10 MΩ.
The 100 nF capacitor reduces the influence of radio signals and hum on the measurement.

Calculate Is with the formula:

formula 3: Is = ID / (e^ (VD /(0.0257xn))-1)
Is = saturation current of the diode in nA
ID = current through the diode in nA, (1 μA = 1000 nA)
e = base of the natural logarithms, this is about 2.718
^ = raise to the power of
VD = voltage across the diode in Volt
n = ideality factor of the diode, if you don't know this value, take for instance: n= 1.08

within 15 miles of my location (seven 50 KW locals) heavy trapping of pests is required. Both sensitivity and selectivity are good; on evenings of good propagation, signals from Alberta and California can become 'pests' in themselves! Both in-line traps and loose-coupled traps have proven to be very effective.

Not visible in the photo are the in-line wavetraps shown here. Calibrated dials help to quickly zero in on pest frequencies. During the DX Contest 80 different stations were logged on the AM band, the furthest DX catch being WOAI in San Antonio, Texas. U.S. stations in WA, OR, ID, CA, UT, NV, MT and CO as well as Canadians in BC, AB, and SK were heard. Nine 1 KW stations along the west coast were logged as was the little 500 watter CKBX in 100 Mile House, B.C. As well as the DX set performed however, there was still much room for improvement!

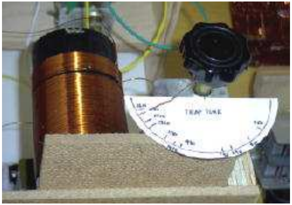

BUILDING HINTS

1. Avoid winding coils on wood or cardboard forms. These materials, even when sealed, will absorb 'Q -killing' moisture and impair both selectivity and sensitivity. Use either styrene sewer pipe couplers, thin-walled black ABS plumbing pipe or white PVC tubing. For really high-Q coils, experiment with 'basket-weave' or 'rook' style windings.

2. Avoid small guage wire for coils. Large diameter coils made of solid wire (#12-#18) are easy to wind. Try to space the coil winding so that each turn is at least one wire-diameter away from its neighbor. Stay away from taps if possible. Taps can quickly rob 'Q' from a coil.

3. Make your coils big! No smaller than four inches in diameter is a good starting point.

4. Use the best quality tuning capacitors that you can find. The ones with ceramic insulation are best. Try to avoid mounting them directly against moisture absorbing surfaces, such as wood. Insulate them with ceramic or plexi spacers. Use insulated shaft extensions to avoid hand-capacitance effects.

5. Devise a method of comparing different diodes. A switching arrangement that will accomodate more than one diode will allow some real-time 'A-B' diode comparisons. Test your diodes on a DX signal as a diode that produces the best signal on a strong local may not be the best one for hearing weak DX. Most older 1N34 diodes appear to be hotter than those manufactured today. Consider the use of Schottky diodes that

T = diode temperature
^ = raise to the power of

Ideality factor n

The ideality factor n of a diode indicates how good the diode performs with regard to an ideal diode.
A (not existing) ideal diode has a value of n=1.
At low input signals, the maximum available detected output power is proportional to 1/n.
So doubling the n will halve the output power (this only applies at weak signals).

Diode capacitance

Between the two connections of the diode there will be a certain capacitance (capacitor value), when this capacitance is fairly high (e.g. 10 pF) the tuning range at high frequencies is limited.
At increasing reverse voltage across the diode the capacitance will reduce, also the detected voltage in a crystal receiver is such a reverse voltage.

Through this, the frequency of the circuit can shift upwards when receiving strong signals.
On the next page: experiments with a detector unit you find in table 3 a measurement about the frequency shift.

Measuring the Is value of a diode.

We can measure the Is value of a diode as follows:
Send a small current through the diode, the value of this current (ID) must be about 1 μA.

With 3 diodes parallel, the value of RD shall be divided by 3 etc..

Diode resistance when using bias current.

We can decrease the value of RD by sending a small DC bias current (e.g. 0.1 uA) in forward direction through the diode.
The higher the bias current, the lower RD will be.

With the following formula we can calculate the diode resistance RD, when we make use of a DC bias current.

Formula 2: RD = 0.000086171 x n x TK /(Ib + Is)

RD= diode resistance at certain DC bias current Ib (unit: Ohm)
n= ideality factor of the diode
TK = Temperature in Kelvin (= temperature in ºC + 273)
Ib= DC bias current through the diode in A
Is =saturation current of the diode in A

A diode with a certain RD value at a certain bias current, gives the same receiving performance as a diode without bias current with the same RD.

Influence of temperature on "saturation current: Is"

The saturation current (Is-value) is strongly depending on temperature.
A temperature increase of 1 ºC will increase the Is value by about 7 %.
In datasheets, the Is value is most times given at 25 ºC.
If the diode temperature is not 25 ºC, but another value "T", then we must multiply Is with a factor 1.07^(T-25).

have proven to be extremely sensitive when used in high-Q systems.

6. Don't wait too long before considering sound-powered phones. They are a must if you want to hear the weak ones. Watch e-Bay for some good bargains.

7. Use vernier dials/drives on all controls as the high-Q circuits often tune very sharply. Detector tuning and antenna tuners should have a logging scale so that tuning is repeatable.

8. Get your antenna as high as possible. If it can't be very long, then give it lots of top-loading as shown below.

9. Pay attention to your ground system. Buried wires acting as a ground counterpoise work well. Tie into the copper waterpipe system if possible.

10. Keep notes on your experiments.

11. Don't give up! Eventually you will break through and hear DX. On nights of good propagation the band will come alive with signals.

The values of n and Is can (sometimes) be found in the diode datasheet.

In the following table some types of schottky diodes, with the values for n, Is and Rd, the maximum reverse voltage and the diode capacitance at zero voltage.

type diode	n	Is at 25 ºC	RD at 25 ºC	maximum reverse voltage	capacitance	
5082-2835	1.08	22 nA	1260 kΩ	8 Volt	1 pF	datasheet
BAT85	?	?	± 200 kΩ ??	30 Volt	10 pF	datasheet
HSMS 2820	1.08	22 nA	1260 kΩ	15 Volt	1 pF	datasheet
HSMS 2850	1.06	3000 nA	9.07 kΩ	2 Volt	0.3 pF	datasheet
HSMS 2860	1.10	38 nA	743 kΩ	4 Volt	0.3 pF	datasheet

To decrease the value RD, we can connect more diodes in parallel, with two the same diodes parallel the value of RD shall halve.

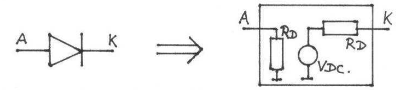

Equivalent circuit of a diode at low input voltages

Via this link you find a measurement on several schottky diodes, which shows detection in the square law region takes place at input voltages below 200 mVpp.

Diode resistance RD.

At zero voltage, diodes have a certain resistance.

This resistance at zero Volt we call RD.

The lower the reverse leaking current of the diode, the higher resistance RD.
When detecting small signals (in the square law region) the input of the diode also behaves like a resistor with value RD.

But how do we know the RD of a diode?

We can calculate it with the formula:
formula 1: RD = 0.000086171 x n x TK / Is
RD = diode resistance at zero Volt (unit: Ohm)
n = ideality factor, the lower this factor the better, between 1.0 and 1.1 is a very good value.
TK = temperature in Kelvin (= temperature in °C + 273)
Is = saturation current (unit: A)
x = multiply

THE CRYSTAL RADIO SET REVISITED

No Author, Smith-Kettlewell

http://www.ski.org/Rehab/sktf/vol11no4Fall1990.html

Abstract

Although this is an article presenting variations on designs of crystal radio sets, I tried to put enough theory in here for universal appeal. For example, impedance matching with transformers is discussed, a formula for calculating single-layer air-core inductors is given, and the "Daniell cell" (a voltaic cell made from a lemon) is also described.

Introduction

[The editor would like to acknowledge the writings and accomplishments of Mr. Joseph Amoros of Richmond, Virginia, who published crystal radio circuits between 1941 and 1953. He sold my father boxes of crystal-set parts when he closed his mail-order business in 1957, and he brought us much joy in experimentation.]

A Bill Gerrish passion is the "crystal set"--the simplest possible radio receiver. It got its name for its crystal detector. Once upon a time, this detector was made by scratching around on a small chip of galena with a rusty wire, hoping to find the happy accident of oxides that would rectify the radio signal and present the amplitude envelope to your headphones.

Actually, various metallic sulfides were used as "crystals." "Galena," lead sulfide or PbS, is the most common ore of lead, and this was often used. "Iron pyrites, iron disulfide or FeS2 (sometimes called "fool's gold") was another ore used. Even Carborundum (a trade name for silicon carbide, SiC) was sometimes employed as the crystal. A blue-steel wire with a sharpened point was the "cat's whisker."

Modern crystal sets use germanium PN-junction diodes for their detectors. These are better diodes than those created from oxides, and they are a lot less fuss.

The purist's crystal radio set is a passive receiver; the power that "runs" it is what you steal from the radio station you are listening to. I leave them connected for years at a time.

With my father's help, I built my first one at age six, and I have hobbied around with them ever since--finding new and better coils now and then, or just putting them in differently shaped containers for the lark of it. The constant faint chatter from earphones hanging on a hook (in my lab upstairs and in the garage) is as friendly as the sound of a familiar clock to me.

Why do I entertain this frivolous pursuit? It's culturally enriching; it causes me to listen to talk shows, music, and yammering advertisers that I would normally leave the room to avoid. I suppose it brings back memories of a kind father revealing his best "shop tricks" to an impatient child who was barely listening.

How Do They Work?

Well, a big long antenna gets in the way of all those waves, and a tuned circuit (a coil and capacitor) resonates to absorb the maximum amount of energy from the station you want--barely attenuating the others. Than comes the heart of the set, the detector.

The power losses in the diode are in this region very low compared to the rectified power.

Situation 2: Recifying in the square law region

If the input voltage is low, lower then the voltage drop of the diode (at 1 mA) then the situation is completely different.
The input of the diode behaves for the RF signal like a resistor with value RD.
The output of the diode behaves like a DC voltage source in series with a resistor, the value of this resistor is also equal to RD.
The value of the DC voltage source is square law related to the amplitude of the RF input signal.
So double input voltage, gives 4 times as much detected DC voltage at the output
In the square law region the output voltage of the diode will be much lower then the input voltage, the diode gives much power loss between input and output.

The lower the input voltage, the higher the losses.
The higher the input voltage, the lower the diode losses.
When further increasing the diode input voltage, we gradually come into the linear detection region.
When receiving weak stations, detection takes place in the square law region.

Between the linear and square law region, there is a region not linear, and not square law but somewhere in between.
This region is not discussed here.

best test several germanium diodes in our receiver and then choose the best. The diode resistance RD of germanium diodes is most times rather low, and only useable in crystal receivers with a low Q (low sensitivity and low selectivity). For high performance receivers, we can better use a suitable schottky diode.

Schottky diodes have a voltage drop of about 0.25 Volt. The differences in properties between two diodes of the same type are often small. Schottky diodes with the correct resistance RD are very useable in high quality crystal receivers.

The given voltage drop is normally measured at a forward current of about 1 mA. Also if we measure the voltage drop of a diode with a multimeter, the test current shall be about 1 mA. But also below this voltage drop the diode can conduct current, and can rectify a RF (radio frequency) signal. Only the current through the diode is then much smaller. When receiving very weak stations, the current through the diode can be e.g. only 10 nA. At such a low current, the voltage drop of the diode is also much lower then at 1 mA.

Detected voltage as function of the input voltage. If we rectify a RF signal with a diode we can distinguish two situations.

Situation 1: Rectifying in the linear region

If the input voltage is high enough (well above the voltage drop of the diode at 1 mA), the output voltage of the diode will be about proportional to the input voltage.

So double input voltage, gives about double output voltage.

The output voltage is almost equal to the peak value of the input voltage.

By itself, a radio signal has no audible component. Stations suitable for this kind of detection have the amplitude of their radio-frequency (RF) signal "modulated" (the power made greater or lesser) by the "intelligence?". (Intelligence on AM radio? Nah!) In order to hear these fluctuations in amplitude, you have to strip away the radio-frequency "carrier" and detect the "envelope" of the amplitude variations.

The diode detector is a rectifier-filter system that feeds the earphones the instantaneous peak value of the RF signal. This peak value is the "amplitude" of the signal, and when you retrieve that information, you get amplitude variations that are the modulating audio.

The rectifier, conducting only in one direction, charges a filter capacitor which then follows the peaks of the RF signal. The filter capacitor is the 0.005uF unit found throughout these circuits, and the diode is the "crystal detector."

Antenna Requirements

The crystal set needs a long antenna and some sort of ground return. A water main is usually the best ground, but a mounting screw of a wall socket, a steam radiator, or a wire fence may be good enough. Avoid using gas pipes as grounds.

There is no substitute for the antenna, however; it must be long. A couple of hundred feet of wire is a luxury. Fifty feet will sometimes do, if it is clear of tall buildings and trees.

My home lab antenna is a continuous overhead central-heating duct which is 50 feet long. (This one is rather funny, too, since it is grounded at the furnace.) My garage antenna is an unused wire in the telephone line. (Be careful of this one. Some telephone installations still use this third yellow wire for the ring signal. first, make sure that there is no DC from the yellow to the red and green, then get someone to ring you and make sure that no high-voltage AC appears between the yellow and the other wires. Finally, always couple to such a thing through a 0.01uF 500-volt capacitor.)

If you can string up a wire antenna, try to keep it from lying on surfaces and bumping into aluminum window frames. You can run a wire antenna inside an attic; hang it from short pieces of string under the rafters (the string acting as insulators). You can use any kind of insulating scheme to keep the antenna from resting against large surfaces. Holding it with plastic strapping of some sort, you can stretch the wire taut between elevated fixtures on a roof so that it doesn't rest on the shingles. On the other hand, unlike transmitting antennas, touching non-metallic objects here and there along the way will not be fatal to an antenna's performance.

That it be long is the important antenna consideration, but beyond this, not having it double back on itself is to be preferred. If you can make it long, but it must go out and comeback, keep the two runs as far apart as possible, and if it can be arranged, avoid running the wire in exact parallel directions.

Luckily, I live where there hasn't been a lightning storm in years. If your situation is otherwise, cultivate the habit of grounding your outdoor antenna when not in use, and if you hear a rumble of thunder, ground it quickly with a clip lead and run like a turkey.

DIODES

Dick Kleijer crystal-radio.eu
http://www.crystal-radio.eu/endiodes.htm

Which diode can I use best in my crystal receiver?

Maybe you think a diode with a voltage drop as low as possible, then also small signals at the detector circuit are detected. But diodes with a low voltage drop, also have a high reverse current (leaking current), this will load the detector circuit heavier, the Q of the detector circuit reduces, and with that also the voltage across the LC circuit. At a lower input voltage the diode will give much more losses, and it can happen that despite the lower voltage drop of the diode, you have less voltage at the load resistor. Besides that, reduction of circuit Q will also gives a less selective receiver.

For every 20 mV less voltage drop, the reverse current will approximately double.

Germanium, silicon, en schottky diodes. Depending on the material they are made from, we can distinguish germanium diodes, silicon diodes and schottky diodes. There are some more types, which are not discussed here.

Silicon diodes have the highest voltage drop (about 0.5 Volt) and are for this reason not very useable for crystal receivers. Unless we use a small DC bias current, which brings the diode already a little bit in conduction.

Germanium diodes have a low voltage drop (about 0.1 - 0.2 Volt) and are often used in crystal receivers. The properties like voltage drop and reverse current can vary a lot between two germanium diodes of the same type. In practice we can

Notes on the Various Components

The crystal set lends itself to parts which have been "scrounged" from far and wide. Don't ever let a dead radio, a 1960-vintage circuit board, or a dead telephone slip through your fingers. From telephones, you can get passably good medium-impedance earphones. From old radios, you can get coils and variable capacitors, output transformers for matching to low-impedance earphones, and cabinets to build your crystal sets into. From old circuit boards with solid-state stuff on them, you can get germanium diodes, or transistors that can be used as diodes. If the family finds the catsup in an orange-juice bottle some day, they will just have to accept the fact that you've used the nice cylindrical plastic catsup squeeze bottle as a coil form.

The Diode Detector

Radio Shack still sells 1N34A germanium diodes--ten for a dollar as cat. No. 276-1123. Mouser also sells germanium diodes--cat. No 333-1N34 and 333-1N60 ($.17 and $.25).

Though not always the case, these diodes are in glass envelopes which break rather easily if you are zealous in bending their leads. Also, their leads are often plated iron, and if you fool around too much in soldering them, the plating will come off, the iron will oxidize, and you will find the process of soldering tiresome, if not impossible. Solder 'em quickly, but deliberately, and treat them gently.

You can use a germanium transistor for its base-emitter diode. Typically, people tie the collector and the base together for one diode connection, and use the emitter for the other connection. If the transistor is burned out, you can often tie the collector and emitter together for one end of the "diode, and use the base lead as the diode's other end.

Silicon diodes and transistors won't work (unless the radio station is breathing down your throat). This is because silicon junctions have a comparatively high forward voltage--typically 0.6V as opposed to 0.3V for germanium junctions. On the other hand, you can apply an external bias voltage to reduce this unwelcome threshold to zero, and a circuit for that is given later.

Another possibility is use of the "Schotky Barrier rectifier"; these begin conducting at 1/3 volt, and they make a loud crystal set. Their disadvantage is that they have a very high capacitance; this is true because they use a large-area metal barrier in the junction. Near 0 volts, their typical capacitance is over 100pF, so use of them means that you have trouble tuning the top of the band. Nevertheless, they are fun to try, and a typical number is the 1N5819.

[You have all heard of the razor-blade detector. You rest a pencil lead (one intended for a mechanical pencil) on the edge of the blade; the oxide inevitably present on the thin edge conducts best in one direction. Like the galena and cat's whisker business, this is not a very good diode, and it's a lot of fuss. Interestingly, it was invented by a prisoner; he probably couldn't get down to Radio Shack.]

Commercially Made Coils

Rather than winding your own coils, you can buy them. The store-bought ones, because there is less market than there used to be, can be quite expensive. Once common place, a ferrite-

Because of the 20 cm distance, there is a loose coupling between the coils.

Connect the probe to the LC circuit.

The earth connection of the probe must be connected to the housing of the tuner capacitor.

The probe is connected to the oscilloscope.

The probe provides a small loading of the circuit, so the Q will not reduce so much.

There are also 1:1 and 1:10 probe's, but these will load the LC circuit too much.

The 1:100 probe I use has a input resistance of 100 M.Ohm, and a input capacity of 4 pF.

The output voltage of the generator must be set so high, that the oscilloscoop gives a clear picture of the RF signal.

Because the 100 times attenuation in the probe, the signal generator output must be set fairly high.

When measuring low Q circuits, I must set the generator output to it's maximum of 20 Volt peak-peak.

For measuring the Q: perform the 5 steps described on the top of this page.

The frequency adjustment is done by hand, by turning the frequency knob of the generator.

point is higher in frequency then f.res, this frequency we call: fh.

Step 5:

Calculate the bandwidth BW: **BW= fh - fl.**Calculate the Q: **Q= f.res / BW**

For performing these 5 steps, we can use the following test setup

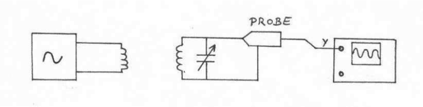

In the schematic above you see from left to right the following components.

A signal generator A coupling coil

The LC circuit

A 1:100 oscilloscope probe

An oscilloscope

Connect the output of the signal generator to the coupling coil having e.g. 50 turns.

Place the coupling coil at about 20 cm from the coil of the LC circuit.

The coupling coil don't have to be high Q.

---

slug-tuned broadcast coil, the good-ol' Miller 9001, now costs you $11. (The tapped one, Miller 9011, is nearly $13 now.)

Nevertheless, trying circuits with those Miller slug-tuned coils is worth doing, because they have such a high Q. Besides having the advantage of a ferrite core, they are wound with "Litz wire." [To conquer the "skin effect" (the tendency for RF current to use only the outside surface of a wire), Litz wire is made of many thin strands of wire, all insulated from one another. The result is low resistance at radio frequencies.] Moreover, those Miller coils are "scramble wound"; their turns do not lie parallel to one another, thereby greatly reducing the distributed capacitance.

Store-bought coils have a funny mounting arrangement. They have a springy metal cap that fits into a 0.3-inch diameter hole. A small tab 1/4 of an inch away from center keeps the coil from turning in its mounting as you adjust the slug. The mounting plate must be thin metal--an aluminum chassis or a fabricated bracket.

Homemade Air-Core Coils

It is easy to make your own air-core coils by winding magnet wire, or even wire-wrap wire, around a cylindrical form.

"Magnet wire" is copper wire which has been coated with enamel insulation; enamel is used because it is so much thinner than other insulating materials. You strip its insulation with medium-grade sandpaper; you test to see if you have gotten down to bare wire with a continuity tester.

There are various types of enamel--some very tough and some fragile. For radio coils, it doesn't matter which kind you use.

(These differences are important to electromagnetics people--designers of motors and transformers.)

Radio Shack has an assortment of "magnet wire"; they give you 40 feet of 22-gauge, 75 feet of 26-gauge, and 200 feet of 30-gauge--Radio Shack No. 2781345. Any electronics supplier or electric motor shop can provide you with magnet wire. Belden Wire Company and Vector Electronics Company sell magnet wire by weight--1/2- and 1-pound spools.

For example, Belden Wire Co. sells "solderable magnet wire" whose enamel is destroyed in the soldering process (as long as the soldering is done at 800 degrees or more). This enamel chafes easily, but it is good enough for these coils; it can be gotten from Newark Electronics. A half pound spool of 22-gauge has the Belden No. 8051, Newark cat. No. 36F1314 (about $20, ouch). A half-pound spool of 24-gauge is Belden No. 8052, Newark cat. No. 36F1315. (about $20).

Though it frightens me to tell you this, Mouser has magnet wire, but only in large quantities. For example, they list an 8-pound 22-gauge spool of magnet wire--cat. No. 501MW22H (4000 feet). I bought one; it only cost me $80, and it is very nice wire made by Phelps Dodge. You can use it as antenna wire as well, but that's a lifetime supply for you and your neighbors.
Typical coil forms for the broadcast band are cores from paper rolls--toilet-tissue cores or kitchen-towel cores, cardboard mailing tubes, or plastic tubes (perhaps bottles with the tops and bottoms cut off). Plastic ones are the best; paper and cardboard take on moisture that causes losses and lowers the Q.

MEASURING THE Q OF LC CIRCUITS

Dick Kleijer crystal-radio.eu
http://www.crystal-radio.eu/enqmeting.htm

In theory we can determine the Q of a circuit as follows:

Step 1

Couple a RF signal generator to the LC circuit.The coupling between generator and LC circuit must be loose, otherwise the output resistance of the generator will load the circuit and reduces the Q.

Step 2:

Set the generator to the frequency at which you want to measure the Q.Adjust the LC circuit (turn the tuner capacitor) so you have maximum voltage over the circuit, the circuit is now in resonance, this frequency is the resonance frequency of the circuit (f.res).

Step 3:

Measure the voltage over the LC circuit at resonance frequency (f.res).

Step 4:

Vary the generator frequency a little above and below f.res. and determine the two frequenties were the voltage over the circuit is 0.707 times the value at f.res.The voltage reduction to 0.707 times, is the -3 dB point.One -3 dB point, is lower in frequency then f.res, this frequency we call: fl.The other -3 dB

Figure 6.

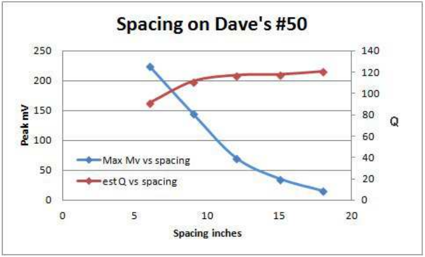

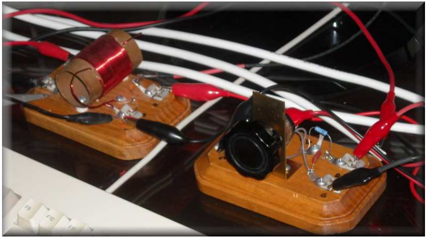

Securing a coil against a panel, especially a wood panel, will cause losses and will add distributed capacitance. Therefore, mount your coils on spacers.

My current favorite plastic forms are the "pill bottles" (vials, 30-dram and 60-dram sizes). These are available from sympathetic pharmacists (you have to convince one that these are for your crystal set).

The 60-dram vial measures 2 inches in diameter--more or less--and 5-1/4 inches long. The 30-dram vial is also 2 inches in diameter, but about 3 inches long (they leave a little room for cotton at the opening). The minor disadvantage of these bottles is the slight taper--smaller at the bottom. Because of the taper, all coils should be wound from the bottom up, so that succeeding turns drift toward the coil, not away from it.

A pill bottle is rather nice for mounting; you just screw the cap to the board and snap the bottle into place so that it stands vertically. This way, spacing the coil away from the board is automatically taken care of.

The ends of windings are anchored by weaving them in and out of a couple of holes. Consider the process of installing a 17-turn antenna winding at one end of a coil form. About 1/2 inch from the end of the form, you poke two holes which are perhaps 1/2 inch apart. You thread the magnet wire through one of the holes--from outside--and pull it through about 8 inches. Then, you poke the wire through the neighboring hole from the inside and pull it taut (giving you 7-1/2 inches of free end now. You now repeat the process; poke the wire down through the first hole and pull it taut, then poke it back out the second hole and pull it taut. This will fix the end firmly in place.

Crumple up the free end and stuff it into the form where it will be out of your way. Next, get a piece of tape ready and stick it to the edge of your workbench where you can grab it handily. Turning the form slowly, wind your 17 turns, every once in a while pressing them together with three or four fingernails so that there are no amateurish spaces between turns. When you are finished (or when you get tired and need to take a break), grab your piece of tape off the edge of the bench and slap it across the coil to hold it in place.

Once the winding has been completed and the tape is holding it, cut the wire to an 8-inch length. Poke two more holes just beyond the winding and weave the free end in and out as before.

Normally, margins of 5/8 inch are left at the ends of the form to permit screw holes to be drilled for mounting. Sometimes, as with the double-tuned circuit using pill bottles, you need to place a winding at the very end of the form; in this case, passing the first half-turn through another pair of holes opposite the ones fixing the end of the wire will keep wraps from spilling off the end of the form.

As an example of anchoring the end-most turn, consider the problem of winding 20 turns of wire around a plastic soda-bottle cap (making a coupling coil for some of these circuits). Very near the closed end of the cap, two pairs of holes are drilled, diametrically opposed. The wire goes in and out of the first pair and is pulled through to a length of 10 inches; then, the end is woven in and out of the opposite pair of holes to firmly "fix the end." A ridge at the open end of the cap means that the last turn can be anchored in one pair of holes; there is no danger of turns falling over the ridge.

In figure 5 above I have attempted to cast Dave’s measurements into something similar to what I am doing, which is looking at the -3db BW and Q at 1100khz. I finally produced the above spreadsheet where I simply smoothed the data for the 1000khz case and estimated the bandwidth at -3dB. This makes a LOT of assumptions of course. Mostly I assume that the +-10khz measurements are well above the resonance skirt so my smooth curve is not unreasonable.

The #50 set then appears to have, at 1000khz, a -3db bandwidth of around 8-11khz depending on diode and/or spacing between the tank and antenna unit giving loaded Q's of 91-125. These numbers fall in the range of what my own data lead me to expect. Ken Kuhn's page on Resonant Circuit suggests the "typical" QL at 1Mhz ranges between 20 and 100 and Dave’s #50 set is certainly beating the best of that.

In the final figure 6 I have taken my estimated results above and looked at how Q and peak voltage (assuming constant conditions other than the spacing between the antenna and detector units so this ought to reflect power as well). It is quite apparent that the set is rapidly losing output power while Q improvements are flattening out as the spacing between tank and antenna tuner grows to 18 inches. It would appear that a -3db loaded Q of 125 is near the limit of what can be expected in a crystal set.

skill, experience and usage of the best ($$$) materials available. I take this to be just about the most of what can wrung out of the sky with a crystal radio. Dave’s protocol is rather different from my own and he did not make a specific -3dB bandwidth measurement so what I do from here on takes many liberties with his work, my sincere apologies in advance!

Figure 5.

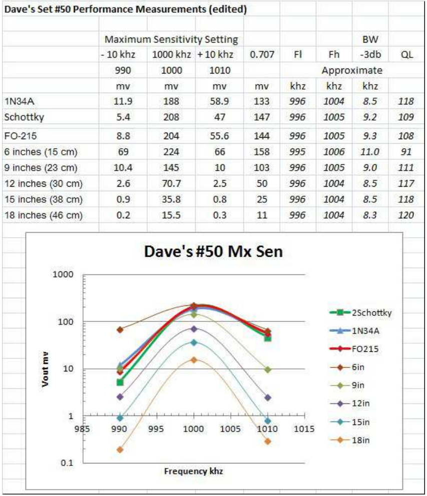

Dave's Set #50 Performance Measurements (edited)

	Maximum Sensitivity Setting						BW	
	- 10 khz	1000 khz	+ 10 khz	0.707	Fl	Fh	-3db	QL
	990	1000	1010		Approximate			
	mv	mv	mv	mv	khz	khz	khz	
1N34A	11.9	188	58.9	133	*996*	*1004*	*8.5*	*118*
Schottky	5.4	208	47	147	*996*	*1005*	*9.2*	*109*
FO-215	8.8	204	55.6	144	*996*	*1005*	*9.3*	*108*
6 inches (15 cm)	69	224	66	158	*995*	*1006*	*11.0*	*91*
9 inches (23 cm)	10.4	145	10	103	*996*	*1005*	*9.0*	*111*
12 inches (30 cm)	2.6	70.7	2.5	50	*996*	*1004*	*8.5*	*117*
15 inches (38 cm)	0.9	35.8	0.8	25	*996*	*1004*	*8.5*	*118*
18 inches (46 cm)	0.2	15.5	0.3	11	*996*	*1004*	*8.3*	*120*

You will see that most of the coils in these circuits require a tap one-quarter or one-third of the way up from ground. A tap can be arranged in two ways. The most elegant way is to sand a little bare spot in the wire when you get to that point in the winding, and solder a pigtail of insulated stranded wire to this bare spot. The other way, while it is a little more crude, is to bend the enameled wire double for an inch or so, then twist it several times so that a loop will emerge from the windings at this point in the coil.

Of course, you can have as many taps as you want. This is one way to match the detector most properly to the earphones--getting sufficient voltage to drive the earphones, while preserving the Q by minimizing the load from the earphones.

When it comes to connecting the coil into the circuit, cut the ends to slightly longer than the desired length and sand about 3/4 of an inch bare and make the connections. If the tap is a twisted loop, don't cut it, but just sand the end of the loop bare and make that connection. Test all stripped segments with a continuity tester; you cannot feel whether you have gotten down to copper, or just roughened the enamel.

Traditional dimensions for radio coils is to design them in such a way that the diameter and length are nearly the same. (Having the length equal the diameter makes the so-called "square coil" which is popularly used.)

There is an approximate formula given in the Reference Data for Engineers, copyright 1964 by IT&T, by which you can calculate the inductance of single-layer coils within a few percent. This expression is:

L (inductance in microHenries) equals n squared times r squared, divided by the quantity 9 r plus 10 l.

Lower-case "l" is the length of the coil in inches, "r" is the radius in inches, and "n" is the number of turns.

For the broadcast band, we want the inductance to fall somewhere around 220 microHenries. To save you the algebra, a 2-inch-diameter form will require secondaries of 80 turns of 22-gauge wire, 77 turns of 24-gauge wire, or 70 turns of 26-gauge wire.

[Because the 60-dram and 30-dram vials are tapered, it is good to make the windings 5% too long to compensate for this.]

Notes on Wire Gauging

The copper wire gauging system used in the U.S. is "the American Wire Gauge" (AWG). The wire diameter changes by a factor of 2 every 6 gauges. Thus, since 26-gauge wire is 0.016 inches in diameter, 20-gauge wire is 0.032 inches in diameter, and 32-gauge wire is 0.008 inches. (These are approximate dimensions, and the enamel adds perhaps 0.001 inches.) Since cross-sectional area depends on the square of the radius, this area changes by a factor of 2 every 3 gauge sizes.

Now, if you can remember three wire sizes, you no longer need a wire table. Good-ol' 22-gauge wire (0.0253 inches in diameter) reminds me of my micrometer--25 thousandths per turn, or 40 turns per inch. Wire of 24-gauge (0.0201 inches) gives you 50 turns per inch. Wire of 20-gauge is about 32 thousandths.

Figure 4.

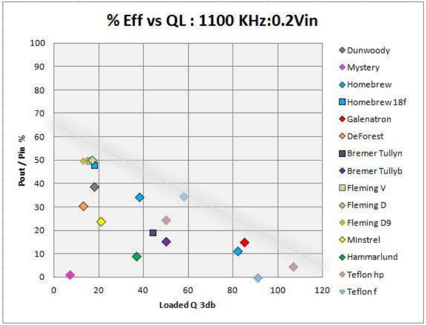

Here is the bottom line, Q verses Sensitivity, efficiency in this case. On this plot it is readily evident that I have built my fair share of loser radios, but also a good number of sets which push boundary of what can be achieved in terms of compromise between Q and sensitivity. Looking at the plot it is fairly easy to imagine a sloping line or zone running from 0 Q at 60% efficiency down to 120-130 Q at 0% efficiency. Beyond this one will probably never go with tank and antenna tuner alone. It will be time to start adding traps to your bag of tricks. Recall that the above is for the 1100khs frequency and that Q is a function of frequency.

In order to explore this final conclusion, I have taken a look at Dave Schmarder’s excellent bandwidth analysis of his #50 contest set. This set is a superb work of craftsmanship and optimal load resistances for this case was similar to or lower than that for the 0.2Vpp input case.

With all that said, the plot in figure 2 is messy but the impression is that of increasing Q with increasing Rl. I view the plot as having a number of small groupings of measurements with similar conditions, each showing higher Q's associated with higher load resistance.

Figure 3.

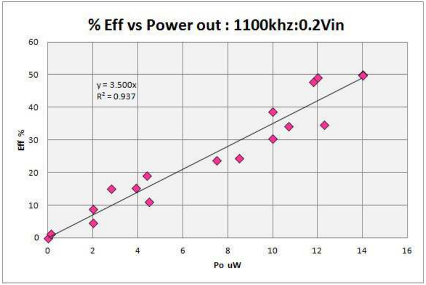

Figure 3 is most obvious with the clear benefit of good set efficiency on output power. An efficient set will better approach the maximum power transfer desired. High Q plays a part here. Still, there is always the tradeoff between selectivity (high Q) and sensitivity. An efficient set delivers more power to the load, but to go after Q you will ultimately be sacrificing sensitivity.

Tuning Capacitors

These are the most hard-to-come-by items in our modern age. Big radios don't have mechanical tuning any more, and little radios will soon give it up. The best tuning capacitors are being harbored in antique radios. Built into metal frames, they have a stationary set of plates ("stator") and a movable set ("rotor"). Many of these are multi-section--usually dual.

Unfortunately, most of the dual ones do not give us two usable sections, because the sections are different sizes and values. The smaller one, used to tune the local oscillator in superheterodyne sets, can be ignored and only the large one used. However, some low-budget radio makers of the 1950s used dual variable capacitors whose sections are identical; for double-tuned crystal sets, these are worth "breaking and entering" for.

Because many salvaged capacitors are fitted with dial-cord pulleys instead of knobs, you may end up making a knob for the odd-sized shaft after the pulley has been pried off.

The rotor is usually common to all sections. Since it is usually part of the frame, and since you must operate it with your hand, it is grounded. Connection to the rotor is almost always done via a mounting screw; solder lugs on the frame are rare. Each stator almost always has two solder lugs, one on each side of the stator, and these are electrically identical.

Radio Shack doesn't even have a suitable capacitor. The only one they show of sufficient range is a "compression-type trimmer" that you have to tune with a screwdriver.

The most readily available modern tuning capacitor is a Mouser unit suited for pocket radios. It has two main sections that go from 5 to 266 picofarads, and four 7pF screwdriver-adjustable trimmers which we can ignore for these sets. The Mouser cat. No. is 24TR218 ($2.71). The only trouble is that they don't supply a knob for the accursed thing; you have to make your own and secure it with a 2.6mm metric screw.

This Mouser unit is festooned with solder lugs. Two sets of three long ones are for the trimmers, which we don't need. There are three short lugs, two on one edge and one on the far corner of an adjacent edge. The lug which is off by itself goes to the rotors of both the large sections; the two coming out the same edge are the stators of these sections.

You can mount the Mouser capacitor by sticking its one free edge--or its back--to the chassis or breadboard with double-sided foam tape.

The required range is perhaps 30pF to 365pF. Actually, most go slightly higher than this--400pF or so, and their lower limit can be nice and low--5pF for each section of the Mouser ones. The Mouser units require that you connect the two main sections together, giving you 10pF to 532pF. The important thing is to have enough range to cover the whole broadcast band (given that the coil will have a lot of distributed capacitance).

(For double-tuned sets with matching coils, the Mouser units can be hooked up another way. Two of their 2-gang units are used; the rotors are connected together, the stators of Section A of the two units are connected together, and the stators of Section B are connected together. In this way, you have ganged capacitors; they both have to be at maximum to tune the bottom of the band, and they both have to be minimum to get to the top of the band.)

---

a generic transformer that would work well with many or all minus the complicated switching.

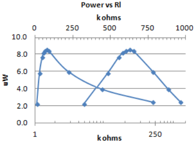

In my testing protocol therefore, I included the important step of determining the optimum load resistance for the set under test conditions. I certainly part company with the school of thought that says to test all sets exactly the same way, including same diode (1N34A typically) and same test load. Sets are designed to work best when impedance matched and at the back end that generally means a load matched to the tank/diode combination that delivers the maxumum power. At right above I show a small plot where I varied the load resistance over a wide range and calculated the output power of the set. (I give the plot with both a linear and a more appropriate log scale). This should demonstrate the importance of optimizing the load resistance.

Optimum load resistance depends on both the tank Q and that portion of the diode detecting the signal. That in turn depends on the signal level. Very weak signals will be rectified in the square-law region of the diode characteristic and require relatively high (and variable) load resistances. Normal to strong signals will be rectified primarily along the peak-detection portion of the diode characteristic and require lower load resistance with less variation. Not shown in the above spreadsheet, I also ran measurements with a 2.0Vpp input. With this stronger signal, rectification was taking place farther out on the diode characteristic. As expected, the measured

If you are serious about getting DX and doing a lot of heavy lifting, you will need a good antenna tuner as part of your set design.

Figure 2.

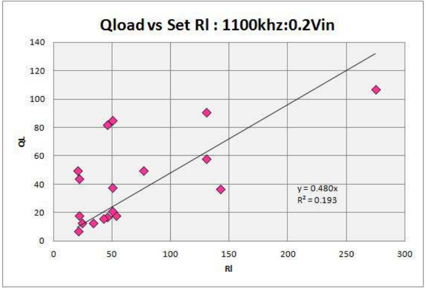

At the back end of the set things get a bit messier and are not east to show with a simple two-component plot. Here I have chosen to highlight the importance of the proper load resistance on the set loaded Q. In my set testing I have tried to pay close attention to the importance of the load on the tank and diode. Early on I considered construction of an output transformer unit that would work with different radio's as well as with different phones. Such a design can be found on Ben Tongue's excellent and highly technical web site. The plans were well beyond what I was able/willing to take on so I turned my attention to discovering just what load resistance exactly do my sets require for maximum power transfer. After all, if I found only a small range typical values, I might design

Caution: When you transport or store open-frame variable capacitors, be sure that the plates are fully meshed. If a capacitor is partly open, the plates of the rotor can be easily bent out of shape.

LC Resonance

As any radio ham knows, the expression for resonance, whether it's a series- or parallel-tuned circuit, is:

f equals 1 over 2 pi times the square root of L times C.

L is the inductance in Henries, C is the capacitance in Farads, and lower-case f is the frequency in Hertz.

Consider the example of 220 microHenries and 400 picoFarads. Expressed in scientific notation, the capacitor is 4 times 10 to the minus 10th power, and the coil is 2.2 times 10 to the minus 4th. Their product is 8.8 times 10 to the minus 14th. The square root of this is 2.97 times 10 to the minus 7th (you have to divide the exponent by 2 when you take the square root). Multiplying this by 2 times pi (2 times 22/7ths), we get 18.7 times 10 to the minus 7th, or 1.87 times 10 to the minus 6th. The reciprocal of this is 0.535 times 10 to the 6th Hz (the exponent changes sign when you bring it up into the numerator). Ten to the 6th Hz is a megacycle, so the result is 0.535 megaHertz.

Earphones

These are becoming a troublesome item to get, but I've solved it. Crystal sets require high-impedance earphones. By high-impedance, I mean 2,000 ohms or more. Most of the traditional hobby types were electromagnetic; a permanent magnet put a bias on a diaphragm by pulling it in a bit, and then coils around the poles of the magnet would cause the field to fluctuate, causing the diaphragm to move back and forth.

Sometimes seen in junk collections, brands to look for are Trim, Brandes, Telephonics, Cannonball, Scientifics, Baldwin, and Utah. Short of these, you can use the earpieces from broken-down telephone handsets--fashioning a headband as best you can. Trim "Dependables" used to be issued by the Library of Congress for talking book listening. If you know someone who has those, get them somehow, even if you have to rough 'm up a bit.

Military surplus phones, such as those made by Utah, were all medium-impedance--600 ohms. These and telephone earpieces must be coupled to the crystal set via an autotransformer (described herein).

Another type, which affords high impedance, is the so-called "crystal earphone." Older ones used a crystal of "Rochelle Salt" as a Piezo-electric element to drive a foil diaphragm. Old ones of these are most likely bad now; the Rochelle Salt takes on moisture and loses its properties over time. New ones have ceramic elements.

They don't look resistive; they electrically appear as capacitors. For this reason, the diode detector may have to be provided with a load resistor in order to perform well. Finally, if you have some around and you are playing with active circuits, you should know that crystal earphones are destroyed by DC, so other than on your crystal sets, put a resistor across these earphones and capacitively couple them.

From the data above some interesting observations can be made.

Figure 1.

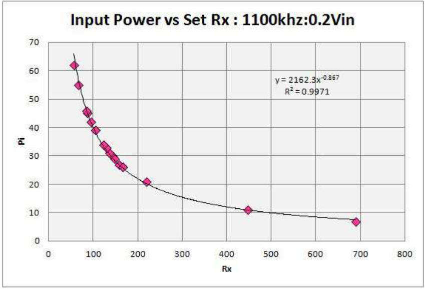

In figure 1 above I show the relation of the input power versus the set resistance. The quality of the fit reflects that fact that this is largely a calculated value, but it does illustrate the importance of having a low input impedance to match that of the antenna. This is not often considered in discussions on impedance matching but is critically important. The antenna has an impedance in the 25 - 50 ohm range, not easy to get this in a set without quality parts and construction. My sets typically have an input impedance in the low 100’s ohms and only two of my sets really get down to the 50-60 ohm range giving a good match to the antenna. Both of these are double-tuned sets with a tuggle front end.

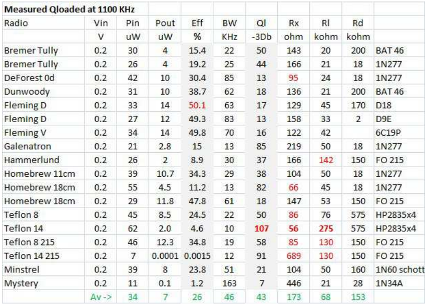

Measured Qloaded at 1100 KHz										
Radio	Vin	Pin	Pout	Eff	BW	Ql	Rx	Rl	Rd	
	V	uW	uW	%	KHz	-3Db	ohm	kohm	kohm	
Bremer Tully	0.2	30	4	15.4	22	50	143	20	200	BAT 46
Bremer Tully	0.2	26	4	19.2	25	44	166	21	18	1N277
DeForest 0d	0.2	42	10	30.4	85	13	95	24	18	1N277
Dunwoody	0.2	31	10	38.7	62	18	136	21	200	BAT 46
Fleming D	0.2	33	14	50.1	63	17	129	45	170	D18
Fleming D	0.2	27	12	49.3	83	13	158	33	2	D9E
Fleming V	0.2	34	14	49.8	70	16	122	42		6C19P
Galenatron	0.2	21	2.8	15	13	85	219	50	18	1N277
Hammerlund	0.2	26	2	8.9	30	37	166	142	150	FO 215
Homebrew 11cm	0.2	39	10.7	34.3	29	38	104	50	18	1N277
Homebrew 18cm	0.2	55	4.5	11.2	13	82	66	45	18	1N277
Homebrew 18cm	0.2	29	11.8	47.8	61	18	147	53	150	FO 215
Teflon 8	0.2	45	8.5	24.5	22	50	86	76	575	HP2835x4
Teflon 14	0.2	62	2.0	4.6	10	**107**	**56**	**275**	575	HP2835x4
Teflon 8 215	0.2	46	12.3	34.8	19	58	85	130	150	FO 215
Teflon 14 215	0.2	7	0.0001	0.0015	12	91	689	130	150	FO 215
Minstrel	0.2	39	8	23.8	51	21	104	50	160	1N60 schott
Mystery	0.2	11	0.1	1.2	163	7	446	21	28	1N34A
	Av ->	34	7	26	46	43	173	68	153	

The above table of measurements were made at a nominal frequency of 1100khz and input voltage of 0.2Vpp as indicated on a digital signal generator following the protocol outlined. I give the radio name under test should you wish to refer to my pages for more on each set. The columns as follows:

Vin = 0.2V pp (0.71mV rms)
Pin = input power in uW into the set (after the dummy antenna)
Pout= output power across the load resistor in uW
% Eff = 100 * Pin/Pout
BW = bandwidth in khz at -3db
Ql = calculated loaded Q of the set
Rx = the set input resistance presented to the antenna
Rl = measured optimum load resistance for maximum power transfer
Rd = Diode junction resistance Rd used in the set

Mouser has high-impedance "crystal earphones," Mouser No. 25CR060 ($1.56). In general, pairs of magnetic earphones are connected in series to get their impedance as high as possible. Pairs of crystal earphones are connected in parallel. Earphones of the wrong impedance can be used if you match the impedances with transformers. For example, small loudspeakers can be used if you drive them with an audio-output transformer; the speakers, connected in series, go to the low-impedance secondary of the transformer (a 4- or 8-ohm secondary), and the primary winding goes to the earphone connections at the crystal set. Medium-impedance earphones, such as telephone receivers, can be matched with an autotransformer; if you have an output transformer with a center-tapped primary winding, the medium-impedance earphones, connected in series, go between the center tap and one side of the winding, while the whole winding goes to the earphone connections on the set (the secondary of the transformer is left open).

Impedances are transformed by the square of the turns ratio. For example, when you drive medium-impedance earphones with half the winding as described above, you see four times the earphone impedance across the whole winding.

Anyhow, I made a pair of earphones using two 8-ohm 2-inch loudspeakers connected in series (an earphone impedance of 16 ohms). Then, I used a miniature output transformer--8 ohms to 1.2K--to drive them. (The series-connected speakers are connected across the 8-ohm secondary winding, while the 1.2K primary goes to the earphone connections on the set. If it has one, the centertap on the primary is unused.) Since the secondary is looking at 16 ohms, the impedance appearing across the primary is 2.4K. I glued the speakers to stiff cardboard disks, then covered the backs with disposable drinking cups. The headband was fashioned out of coathanger. They could further be improved by putting foam shrouds to seal around the ears, thus improving acoustical coupling.

I tried my best to make "Walkman" earphones work, but to no avail. Since they are open earphones, and have small diaphragms, they are somewhat inefficient.

Mouser has a fine output transformer for matching to low-impedance headphones (like my PM speakers and drinking cups); it has a center-tapped primary of 20K, and a center-tapped secondary of 4 ohms--Cat. No. 42KM019 ($1.63). A less drastic impedance-matching transformer is the 8-ohms to center-tapped 1.2K--Mouser No. 42KM003.

Vacuum-tube radios have salvageable transformers of two types, center-tapped primaries for 2-tube "push-pull" amplifiers and high-impedance nontapped primaries for "single-ended" circuits. I kind of remember that 10K to 3.2 ohms was common.

Try everything--half the primary, the full primary, and try various output taps if the transformer has them.

Crystal sets don't give you much surplus power, so acoustic coupling to the ears is a critical factor.

Matching the Detector and the Earphones

Since the impedances of earphones you are likely to find are "by guess and by gosh," and the Q they are effecting is largely unknown, matching by experimentation is the best approach. The following effects can be noted to aid in troubleshooting:

11) Adjust the frequency above and below f res until the output level falls to the level calculated in step10 above. These frequencies are f high and f low. Record these.
12) The -3dB bandwidth = f high - f low. Loaded Q (QL) = f res / BW
13) Take the generator output peak-peak voltage recorded in step 2 above and convert it to rms voltage as follows: mVrms = mVpp / 2.829. This is Lauter's RF voltage E1 mV.
14) Measure the RF voltage across the 10 ohm resistor on the dummy antenna. This is Lauter's RF voltage E2 mV.
15) Input current to the dummy antenna = Lin = E2/10 (mA)
16) Input power to the dummy antenna = Pin = E1 * Lin (uW)
17) Power loss in the dummy antenna = Pda = Lin^2 * 25 (the dummy antenna resistance)
18) Power delivered to set = Px = Pin - Pda uW
19) Set input resistance = Rx ohms = Ex / Lin where Ex = E1 - (Lin * 25)
20) Set output power (uW) = Pout = Vout^2 / Rl in ohms
21) Set % efficiency (sensitivity) = 100 * Pout / Pin

Having built a fair number of sets, and with varying quality, all the above protocol was developed to test my sets. What follows now is a discussion of how my sets stack up against each other and some notes on what one may expect in a crystal radio performance-wise. I give a summary table below of the essential data gathered on each of my sets.

Table 1

Once I find the optimum load resistance, I set the pot to that and proceed to the main testing for sensitivity and bandwidth/Q measurements.

Test Protocol as follows:

1) Connect the set to be tested between the dummy antenna and measurement load as per the above diagram.
2) Connect the signal generator to the dummy antenna and set the frequency and voltage output. I typically test in the center of the BCB at 1100khz and use a 200mVpp output sine wave.
3) Set the load resistance to somewhere around 50k ohms. The exact value is not critical as it will be adjusted later.
4) Connect the DVM and tune the set to peak the output voltage. Care here is amply rewarded, especially with double-tuned sets.
5) Once peaked it is necessary to adjust the generator frequency to re-peak the output voltage as hand-capacitance or other factors may have prevented perfect tuning. You now have your set properly tuned for maximum output. Presumably this also means the best possible impedance match between the set and dummy antenna. This is the resonant frequency (f res).
6) Record the frequency and output voltage.
7) Now find the optimum load resistance. This is done by making a series of paired measurements of load resistance and output voltage. Power in uW = mVout * Rl khoms.
8) Set the load resistor pot to the value where power output is highest, this is the optimum load for the test. Record this.
9) With the signal generator re-peak the output voltage and re-record the resonant frequency (f res) and output voltage (Vout mV).
10) Multiply mVout by 0.707 to find the -3dB level.

If the set distorts on weak stations, either the diode is not seeing enough voltage (the tap is too low), or the impedance of the earphones is too high and the detector is not sufficiently loaded. If the set is nonselective ("broad as a barn"), the impedance of the earphones is loading the coil and greatly reducing its Q, or the tap is too high on the coil.

3DB tapping:

The serious experimenter will want to build his secondary winding with a selection of taps for the detector. Since the impedance--and the power--go up as the square of the turns ratio, adopt the systematic scheme of doubling the impedance every tap (going in steps of 3dB). This would require taps at the following percentages of the winding (starting at the grounded end): 18%, 25%, 35%, 50%, 70%, and the full winding. Notice that these percentages are related to each other by the square root of 2; the voltage and current either go up by 1.414 or down by 0.707 as you select neighboring taps, while the impedance and power are either doubled or halved as you select neighboring taps.

Solder these taps to successive positions on a 6-position rotary switch. The arm of the switch will then go through the diode, then through the earphones to ground, with the earphones being shunted by 0.005uF (see the circuits to follow).

Crystal Radio Circuits

I give my impression of the performance of any set I have in my collection. I tested them on my little furnace duct antenna, doing so in the daytime.

Some of these have variable coupling between the antenna and the main tuning coil. This creates a sort of combination volume control and selectivity adjustment. In general, the looser the coupling, the more selective the set becomes, although this is at the cost of sound volume. If a station is blustery enough to over-ride all interference, tighten the coupling, turn it up and dance to it.

* * *

I won a science prize once for making the world's smallest receiver. The circuit follows:

Circuit for the World's Smallest Receiver

The antenna goes through the diode to ground. The earphones are in parallel with the diode.

It gets everything all at once. This can be rather fun. You can hear zealous spiritual leaders in shouting matches with "rap musicians." When a famous politician's speech is carried by all the networks, you can hear three or four of him, all accompanied by drum machines.

[It is worth noting that at one time, this was good enough; in the very early days of radio, all the stations were trying to transmit at a wavelength of 300 meters--all more or less at 1 mHz.]

* * *

Perhaps the most common of circuits uses the distributed capacitance of the coil to make a half-baked LC circuit. A long coil is required, perhaps 200 turns of No. 26 wire on a cardboard or plastic form (such as a mailing tube, the core from a roll of kitchen towels, or the 60-dram vial). Then, for mechanical stability, "dope" (coat) the coil with model cement. By sanding to and fro with medium sandpaper, create a strip of exposed copper across the windings so that a metal "tuning slide" can make contact with the coil. Next, fashion some sort

radios to be tested a 1N34A diode and then use a standard 2k ohm load resistance to make the output measurements. This may give a standard and comparable result, but it does not insure a good impedance match and does not reflect the intended usage and setup of many crystal sets, especially sets intended for high-performance. Here I am interested to learn just what is the optimal load resistance for my set under the test conditions. Note the caveat "under the test conditions". This may not reflect actual usage unless you take precautions. You will want to set the generator to an input voltage similar to what you expect the radio to receive in actual operation. You will probably want to signal level work the diode in its peak-detection region but not so high as to saturate the diode, (In making a loaded Q measurement you want to make sure that the diode impedance stays constant; otherwise, the change in diode impedance (as the tank attenuates the signal before the diode) will change the load on the tank and the bandwidth measurement won't really reflect the Q of the tank). Finally, for the output voltage you need a good voltmeter with very high input impedance that will not load down the set under test. This really means a good DVM. I use a used Keithley 192 bench meter with a 2M ohm input impedance, and as a backup a used Keithley 180 bench electrometer. Overkill perhaps, but for just in case...

Getting back to the load resistance then, my protocol is to find the optimum resistance to deliver the maximum power (not voltage!) to the load. This is a bit of a tedious process but will amply rewarded. Having tuned and peaked the set to the measurement frequency, I then take a series of resistance/voltage readings to find the maximum power output. In my circuit this is accomplished with the variable 1M pot. The resistance meter and voltmeter cannot be connected to the load at the same time making this a tiresome road to follow.

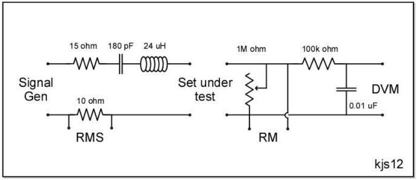

For the tests I originally found an old analog signal generator on ebay which I sadly found to be wholly inadequate. The unit was by no means cheap but the dial was difficult to read with any precision and the "play" in the dial made readings hopelessly inaccurate. Finally, it gave no indication of the attenuation or voltage output levels, rms or pp, of the signal. I have since found a nice, reasonably-priced new digital unit, (max's out at 2Mhz so strictly for AM band work) that gives good readings, gives me the output voltage level in pp, and can even produce different wave forms. I recommend if you are seeking to equip your lab, go this way from the start!

To calculate the signal input power you will need a good rms voltmeter. This is a non-trivial item and, after my experience with the vintage signal generator I was leery of getting a used meter. In any case, an expensive dedicated single-use piece of equipment seemed extravagant. For this measurement I have found entirely acceptable results using an oscilloscope which, of course, is useful for so much more and belongs in your lab regardless!

It is on the back end of the measurement where I diverge from Lauter's protocol significantly. He recommends placing in all

---

of slider (perhaps a ground-off hacksaw blade anchored at one end) which can be used to make contact with the bared coil windings anywhere along the length of the coil.

What you have created is a coil with adjustable length. Only one end of the coil is connected to anything; the other coil connection is the slider. A circuit for the simplest tuned set is as follows:

Simplest Tunable Crystal Radio

The coil slider is grounded. One end of the coil goes to the antenna; also connected to this end of the coil is one end of the diode. The other end of the diode goes through the earphones to ground. The earphones are shunted by a 0.005uF capacitor.

The disadvantage of this circuit is that it is broad as a barn--for two reasons. First, the tuned circuit is low Q because of the resistance of all that wire used in the coil; tuning the set with a variable capacitor means that a shorter coil could be used. Second, no attempt is made to efficiently match the earphone circuit to the "tuned circuit"; unless very high-impedance earphones are used, the load that they present lowers the Q even more.

* * *

Standard LC-Tuned Crystal Radio Circuit

This was copied from a Knight Kit (Allied Radio) set purchased in the 1950s. The variable capacitor goes from 22pF to 400pF. It uses a Bakelite coil form measuring 1-1/2 inches in diameter and 5 inches long. The coil, wound with 22-gauge magnet wire, has a primary (antenna coupling) winding of 17 turns placed at the cold end of a secondary (main tuning) winding of 120 turns. The detector is taken from a tap which is one-fourth the way up from the cold end-- of the secondary30 turns from the grounded end. The separation between the two windings is 1/4 inch.

The antenna coil, 17 turns of wire at one end of the form, has its outside end going to the antenna, while the end nearest the secondary is grounded, This end of the secondary is also grounded. The other end of the secondary goes to the stator of the variable capacitor; the rotor of the capacitor is grounded. A tap on the secondary which is 30 turns from the ground end goes through the diode, then through a 0.005uF disc ceramic capacitor to ground. The earphones are connected across this ceramic capacitor.

This can be built on a 2-inch coil form; just change the proportions. For example, 80 turns of 24-gauge wire--or 85 turns of 22-gauge wire-- on a 2-inch-diameter form will do; I like the detector tap to be at 25 turns. The antenna primary can stay at 15 or 17 turns; an embellishment is to wind a 25-turn antenna winding with a tap 9 turns away from the antenna end, using two antenna binding posts for selectable coupling.

This set is very selective, but not very loud. I am able to get ten stations, seven of which can be isolated enough to listen to. It can be jazzed up a bit by providing it with variable antenna coupling:

Variable Coupling on the Standard Set

Wind 20-turns of 26-gauge wire on the plastic cap of a plastic soda bottle. Affix this cap to a nonmetallic stick of some sort, and fashion a sliding arrangement that permits the cap to be moved in and out of the grounded end of the secondary

---

RADIO TEST

Kevin Smith
http://www.lessmiths.com/~kjsmith/crystal/rtest.shtml

OK, now that you have built a radio, or several, how do they perform? Without a bit of objective testing you will never really know. I have often read descriptions of radio performance such as "tunes very sharply" or "separates two closely spaced stations" and/or etc. These are qualitative descriptions that do not really give you much information. To know your radio, it takes some measuring. I am making the assumption here that, as a homebrew radio builder, you will find tinkering and measuring your sets just as interesting as their design and construction.

For testing, I have pretty much lifted the excellent procedure outlined by Charles Lauter with a few modifications to suit my own philosophy. I will give my own protocol but first let’s look at the needed equipment and test setup. Testing a crystal set requires two pieces of homemade equipment, a dummy antenna in front of the set which allows measurement points, and a measurement load at the back end. Other pieces of equipment include a signal generator, an rms voltmeter, a resistance meter, and a good digital voltmeter. Circuit schematics for the homebrew components are shown below.

automatically. These measurements will definitely determine which are the best Xtal sets, circuits, or components.

Happy Testing

---

winding--from an inch separation to where the coupling coil is encircled by the cold end of the secondary.

Scrap the 17-turn antenna winding on the main coil form. One end of the bottle-cap coil is grounded, while the other end goes to the antenna.

Popular Mechanics Variation

A variation on this basic circuit appeared in the January 1977 issue of Popular Mechanics. The coil was wound on a "salt box" (3-1/4 inches in diameter). The junction between the antenna coil and the tuned secondary is a tap on an overall winding--14 turns below the tap for the antenna coil and 32 turns above the tap for the secondary.

This "tap" is grounded. The antenna goes to the end of the 14-turn segment. The rotor of the capacitor is grounded. The stator goes to the end of the 32-turn segment. The end of the 32-turn coil also goes through the diode, then through the earphones to ground, with the earphones being shunted by 0.005uF.

The editor has not built this Popular Mechanics set, but the closeness of its primary and secondary windings--delineated only by a tap on the continuous coil--will probably make for a loud set with poor selectivity.

* * *

A Double-Tuned Crystal Radio Circuit with Variable Coupling

Two 2-inch-diameter 30-dram pill bottles (vials) are used. The bottoms and caps are drilled to accommodate a wooden dowel or plastic rod. One of the vials is fixed by winding tape around the rod inside the bottle and under its cap; the other is permitted to slide a distance of perhaps 2 inches. When moved toward one another, the vials must be able to touch bottoms.

Matching coils are wound on the vials--85 turns of 22-gauge or 80 turns of 24-gauge wire. These are started as close to the bottoms as you can get, so that they can be brought quite close together. The secondary is tapped 25 turns up from the bottom, or even better, taps are placed at 15 turns, 20 turns, 28 turns, 40 turns, and 56 turns. (If multiple taps are created, connect them to successive positions of a 6-position rotary switch, with position six going to the hot end of the full winding.)

The rotors of two 400pF variable capacitors are grounded. The antenna goes to the far end of the untapped coil; the bottom of this coil goes to the stator of one of the capacitors. The bottom end of the tapped coil is grounded, with its top end going to the stator of the other capacitor. The tap (or the arm of the rotary switch) goes through the diode, then through the earphones to ground. The earphones are shunted by 0.005uF.

Because the coils are matched (except for the load the detector puts on the secondary), you can get away with a ganged 2-section variable, although having separate variable capacitors gives you some freedom to "detune" the set as an aid to interference elimination. Remember, the two sections of a ganged tuning capacitor must be identical.

With this set, I was able to receive ten stations, eight of which could be isolated well enough to listen to.

* * *

2. Increase the Signal Generator output until the DVM reads 1.414 volts DC. This will now be the new value for the input voltage "E1". Maintain this value for the rest of these tests.
3. Decrease the effective Xtal set coupling until the DVM reads 1.000 volts DC. This may effect the Xtal set tuning. If so retune the Xtal set to the Signal generator frequency and repeat steps two and three. Remember to keep E1 constant.
4. Again measure the bandwidth using the previous method above. Now go back and measure the sensitivity again.
5. This process can be repeated over and over for better and better selectivity.

This bandwidth will be less (narrower) than when the Xtal set was adjusted for maximum output.

Xtal Sets with more tuned circuits and variable coupling will have in better selectivity.

Now lets see what this reduced coupling did for the Type SCR-54-A. The bandwidth at -3db is now only 9 kHz and the -20 db bandwidth is now 73.6 kHz. Lets see what this change did to sensitivity. Crystal Set efficiency decreased from 21.5% to 12 35% and total efficiency decreased from 6.83% to 3.67%. These efficiency numbers were expected because they are about a -3db decrease which is the same as the decrease in step 3 above.

The set of three measurements above is then repeated at each of the test frequencies listed above.

All of the above measuring may sound like a lot of work but once the equipment is in place the tests do not take long to perform. I use a computer spreadsheet so that after I enter the measured values, all of the calculations are performed

Now lets again look at the little "Beaver Baby Grand " Xtal Set for Selectivity.
F test = 600.4 kHz, F low = 543.4 kHz, F test - F low = 57 kHz, 57 kHz x 2 = 114 kHz. This is the Basic or "-3db" Bandwidth. Now "F min" = 434 kHz, F test - F min = 166.4 kHz, 166.4 kHz x 2 = 332.8 kHz = the "-20db" Bandwidth.
As you can see this is not a very selective Xtal set. For the intended use, this was probably not a problem in 1922. In today's world, the Set would be receiving several stations at the same time in almost any city in the US.

Now lets look at the selectivity of the Type SCR-54-A. F test=600 kHz, F low= 589.3 kHz, F min = 556 kHz. This means that the half power or -3db bandwidth is 21.4 kHz and the -20db bandwidth is 88 kHz. With this Set we have the ability to possibility select a station we want to listen to.

BANDWIDTH AT REDUCED SENSITIVITY:

More complex Xtal sets have some means of improving Selectivity at the expense of Sensitivity. There are a number of ways of accomplishing this. In most cases reducing the coupling between the antenna circuit and a second tuned circuit is the method used. This is the purpose of the "Loose or Slide Coupler" that is used in many of the better Xtal Sets. In some Sets changing a coil tap and re-tuning a capacitor will accomplish the change in coupling. It may require some experimenting to find the best adjustment method and control settings for any given Xtal Set. The procedure is as follows;
1. Set up the Signal Generator and the Xtal set the same as in the sensitivity measurement above.

A Slug-Tuned Crystal Radio Circuit Using the Tapped Miller Coil

This is quick to build. The Miller 9011 tapped variable inductor is used. (This tap is 30% up from the cold end.) The end of the smaller portion of the winding is grounded. The other end goes to the stator of a 400pF variable capacitor, the rotor of which is grounded. The antenna goes to the tap. The tap also goes through the diode, then through the earphones to ground, with the earphones being shunted by 0.005uF.

The slug is adjusted to "center" the band in the range of the tuning capacitor. Simpler yet, since the coil is variable, the capacitor can be a fixed one--330pF across the ends of the coil. The only disadvantage is that the range of inductance is only about 4 to 1; since the frequency varies as the square root of this, the frequency range can only be 2 to 1 (from 0.6 to 1.2 mHz, for example).

This set is very loud; however, because the antenna coupling is much too high, it's broad as a hippopotamus. Four stations were received, and only three could be isolated enough to listen to.

A profound improvement can be had by fashioning a coil for the antenna. Wind 20 turns around a plastic bottle cap from a soda bottle (which is 1-1/8 inches in diameter). With bee's wax or cement, fix the cap over the end of the Miller coil. On this new primary winding, the end nearest the open end of the bottle cap is grounded; the other end goes to the antenna.

Wow! With this arrangement, I received eleven stations; eight could be isolated enough to be listened to.

Double-Tuned Crystal Radio Circuit with Slug-Tuned Coils

The coils, one Miller 9001 and one Miller 9011 are mounted near each other on a strip of aluminum. They are pointing the same direction and mounted at 1 inch between centers. (One could make this coupling adjustable by using two strips of aluminum in a scissor-like arrangement, but moving these coils detunes them, so you would have to adjust their slugs each time.)

One end of the 9001 (untapped) coil goes to the antenna; its other end goes to the stator of one section of a dual-ganged variable capacitor. The rotors are grounded. The end of the smaller portion of windings on a Miller 9011 coil (tapped at one-third) is grounded; the other end of this coil goes to the stator of the second variable section. The tap on the 9011 goes through the diode, then through the earphones to ground, with the earphones being shunted by 0.005uF.

Back the slugs out 15 turns from maximum (maximum being all the way in). Check to see if the capacitor tunes the whole band; if not, move the slugs in to shift the stations downward, or move them out to shift every thing upward as necessary. Next, find a station near the top of the band and adjust the secondary (the tapped coil) for maximum volume; retune the capacitor and adjust the slug a second time.

This is the best set I've ever had. It's quite loud, and I can receive thirteen stations, eleven of which can be isolated enough to listen to.

DVM and again adjust the Generator above value of "E1". Keeping the value of "E1" constant removes the output resistance of the Signal Generator and keeps it from effecting the bandwidth to be measured. The higher the resistance or impedance of the Signal Generator the more interaction there will be to obtain the final value frequency at which the Xtal Set output is 0.707 Volts DC. This is because as the Set is tuned to other than the selectivity test frequency the impedance looking into the dummy antenna (the Signal Generator load) changes. >>> Keep the value "E1" constant for this and all of the following tests. <<< Note the frequency. Call this frequency "F low".

2. Subtract "F low" from the Sensitivity test frequency used above. Multiply this difference frequency by two. This value is the basic (half power) "-3db" Bandwidth " of the Xtal Set for comparison purposes. Normally one would measure the frequency above the peak or resonant frequency at which the output also drops to 0.707 Volts and then subtract F low from this frequency to obtain the bandwidth. This measurement was done in this manor because in some cases if you try to increase the frequency to the point where the output voltage drops to 0.707 Volts a number of resonant frequencies of the Xtal Set will prevent obtaining a good measurement. (usually due to self resonance of the Xtal Set coil or coils) The error due to this method is very small.

3. For the next test, reduce the Signal generator frequency until the Xtal Set output voltage decreases to 0.1 VDC. (-20db). Call this frequency "F min"

4. Subtract "F min" from the Sensitivity test frequency above. Multiply this difference frequency by two. This is the "-20db" Bandwidth.

editions of Crystal Clear by Maurice L. Sievers and was described by Alan Douglas in the September 1978 edition of Radio Age.
Test frequency used = 600 kHz. Measured values; E1 = .605 VRMS, E2 = .048 VRMS.
Now the computed values; Iin = .0048 A, Pin = .0029 W, Pda = .000576W, Px = .00233W, Pout = .0005W, The Xtal Set efficiency = Pout / Px = 21.48% <<<<< Now lets add the dummy antenna. Rx = 101 Ohms, Pbest = .0073W then the total efficiency is 6.8%. It is Interesting that there is more energy or power loss in the set but the total efficiency is better by about 4.3 times. This Set has more coils and capacitors which is why the losses are greater but the sets ability to better match the antenna is the reason that the overall efficiency or sensitivity is better. The set is much more selective as will be shown below.

SELECTIVITY or BANDWIDTH:

These tests are a little more complex than the sensitivity tests because selectivity is more complex to define.

Make the sensitivity measurement above and before making any adjustments to
the Xtal Set or to the Signal Generator output level make the following measurements. I recommend using a frequency counter to measure the following frequencies.

1. Decrease the Signal Generator frequency to below the first test frequency and until the Xtal set output decreases to 0.707 volts DC (-3db). Readjust the output level from the Signal Generator until the value of Voltage "E1" is the same as that used above. Again adjust the frequency for 0.707 VDC on the

Circuit for the "Rocket Radio"

If you happen to find one of these in a junk shop, they're a lot of fun. They use an untapped slug-tuned coil. Pulling up on the nose of the rocket withdraws the slug and raises the frequency. You will notice the conspicuous lack of the filter capacitor; the capacitance of the crystal earphone doubles as the filter.

One end of the coil goes to the antenna. This junction also goes through the diode, then through the earphone (high-impedance crystal earplug) to the other end of the coil. The earphone is shunted by 220K. Depending on the make, the coil might have enough windings so that the distributed capacitance is the "C" of the LC circuit. Some had a capacitor shunting the coil.

The only ground return is you. The main secret behind these radios was the intimate acoustic coupling from the earphone to your tympanum.

* * *

A Hybrid Crystal Radio Circuit with Lots of Adjustments

The secondary winding is an air-core coil--85 turns of 22-gauge on a 2-inch form; it is multiply tapped and the detector is taken from a 6-position switch. The primary is a Miller 9001 untapped inductor. Since the characteristics of the coils are markedly different, you need separate variable capacitors (of identical type).

The Miller coil is mounted on a sliding bracket that allows its penetration into the cold end of the secondary to be adjusted. A knob is made--having a 4-40 threaded hole--and is fitted to the screw of the slug; a "jam nut" ahead of the knob secures it firmly.

The antenna goes through the Miller coil, then through C1 to ground. The bottom end of the secondary is grounded and C2 is across this coil. The taps go to positions of a rotary switch. The arm of this switch feeds the detector.

Adjusting the slug changes the Q of the primary. Sliding the primary in and out changes the selectivity and the volume. Selecting taps optimizes the match to the detector (changing the selectivity and the volume). Juggling the capacitors finds optimal settings to cope with various interference.

The tally isn't in yet on its performance; the instrument is too hard to run. I'm still chasing the last of thirteen stations, and I think I imagine a fourteenth. Watching you operate it can really impress your sister, although, in my high-school days, it seemed to invoke little enthusiasm from girl friends.

* * *

The 15-Minute Set Using RF Chokes

Mouser sells a style of miniature RF choke which has "radial leads"--both leads come out one end. This set uses two of them, Mouser cat. No. ME434-1120-221L (220 microHenries). If you butt them together--head to head--and hold them together with tape, you have the makings of a fine double-tuned crystal set. I "haywired" the whole arrangement on one side of an old open-frame ganged variable capacitor (no board was used). On the other hand, these chokes, together with two of the Mouser 24TR218 variable capacitors, could be put into a fancy little package.

The antenna goes through RFC1 to the stator of C1. The rotors of C1 and C2 are grounded. RFC2 is in parallel with C2. The

The resistance of the dummy antenna is 25 Ohms. The ideal resistance of the Xtal Set is also 25 Ohms. This means that the best possible efficiency of a Xtal Set and its antenna and ground is 50%. This would mean a Xtal Set that is 100% efficient. Compute the power into the Dummy antenna if Rx was 25 Ohms; P best = (E1 x E1) / (25 + 25).
h. Now compute the total efficiency of the Xtal Set with the dummy antenna. Total efficiency = Pout / P best <<<<<<<<<<<<

Now lets look at some numbers from an actual Xtal Set. The set is "The Beaver Baby Grand vest pocket Radio Receiving Set". This is a little set that was made in 1922 and is shown on the cover of the book "Crystal Clear - Volume 2" by Maurice L. Sievers.
Test frequency used = 600.4 kHz. Measured values; E1 = 1.26 Vrms, E2 = .015 Vrms Computed values; Iin = .0015 Amps, Pin = .00189 W, Pda = .0000563 W, Px = .00183 W, Pout = .0005 W, Xtal Set efficiency = Pout / Px = 27.27% <<<<<<
Now for the rest of the story; Rx = 815 Ohms, P best = .0317 W, and the total efficiency Pout / P best is now only 1.57%. As you can see, the efficiency of the set alone is not too bad, but when it is used with a standard antenna it is very poor.
This set was apparently designed as a low cost Xtal Set to receive a one and only nearby Radio Station. This Set is an example of no matter what type of coil or coil wire is used, the performance would not change enough to matter.

Now lets look at a better Xtal Set. The set is from world war one and its U.S. Army designation is "Type SCR-54-A". This set is also known as a BC-14A. The Set is shown in both

2. Set the Signal Generator to produce an output level of about one volt rms.
3. Adjust the Xtal Set for maximum output as indicated by the DVM.
4. Adjust the Signal Generator frequency for maximum output that may be slightly different from the nominal test frequency due to the inability of the Xtal Set to precisely tune if the Set uses coil taps for tuning.
For extremely accurate tuning make the phase of the voltage across the 10 Ohm resistor in phase with the Signal Generator output using a two channel Oscilloscope.
5. Adjust the Signal Generator output level to give a 1.000 volt DC reading on the DVM.
6. Measure the RF Voltage at the input of the Dummy antenna. Call this value "E1".
7. Measure the RF Voltage across the 10 Ohm resistor. (Between the Signal generator common and the Xtal Set ground terminal) Call this value "E2".

8. Now some calculations;
a. Compute the input current: Iin = E2/10 (Amps)
b. Compute the power into the Dummy antenna; Pin = E1 x Iin (Watts)
c. Compute the power loss in the Dummy antenna: Pda = (Iin x Iin) x 25
d. Compute the power going into the Xtal Set: Px = Pin - Pda
e. The power output from the Xtal Set = Pout = Eo x Eo / 2000 = 0.0005 Watts
f. Now for the Xtal Set Efficiency = Pout / Pin = .0005 / Pin <<<<<<< Note that this value does not include any mismatch with the Dummy antenna.
g. Compute the Xtal Set input Resistance. (Assuming that Xtal Set is tuned)
Rx = Ex / Iin , where Ex = E1 - (Iin x 25)

stator of C2 goes through the diode, then through the earphones to ground. The earphones are shunted by 0.005uF.
I declare, when I counted ten stations on this set (eight of which I could isolate), I jumped right up to "Do the Locomotion" with my soldering iron.

[These chokes are nifty. Good for 50mA, values we are likely to use elsewhere are also available: 2.2mH are ME434-1120-223K, and 10mH are ME434-1120-104K. The "L" appending the low-value 220uH unit stands for "low."]

* * *

Biasing the Detector

Crystal detectors have a serious flaw; a station must be strong enough to overcome a forward voltage drop before it begins to drive the earphones. This can be compensated for with a bias battery. Moreover, someday, germanium diodes may be unavailable; active circuits are taking their place (creating "perfect rectifiers"). A silicon diode, such as the 1N914, can be used if you get rid of its 0.6V forward voltage drop.

"Battery?!" you cry in protest. It's not so bad, in that the current drain is only 15 microamps; a D cell will last its shelf life. This detector will work in any of the sets.

Biased Detector Circuit

The tap on the coil goes to the cathode of a 1N914 diode. To provide the detector with a DC load, the anode goes through 22K, in parallel with 0.005uF, to ground. The anode also goes to the positive end of a 1uF electrolytic (tantalum is best), with the negative end of this cap going through the earphones to ground. The junction of the earphones and the electrolytic goes to the negative terminal of a cell, with the positive of the cell going through 100K to the diode's anode.

Selectivity is quite different, a number of definitions are required. The most basic and popular definition of bandwidth is the so-called half power or "-3db" bandwidth. To obtain this number subtract the frequency below the frequency of maximum output where the output decreases to 0.707 times the peak value (F below) from the frequency above the maximum output frequency again where the output decreases to 0.707 times the peak value (F above). Or bandwidth = (F above) minus (F below) In the actual tests below I modify this definition slightly to simplify the measurements.

Selectivity has a second and important characteristic. It is related to Shape Factor. Filters are what make any Radio selective. Unfortunately simple filters are far from perfect. If a filter were perfect the bandwidth would be the same for all values of attenuation. The term Shape Factor refers to a bandwidth at a greater attenuation than -3db. For our tests we will use the bandwidth at -20db for comparing different Xtal Sets. (-20db = E max. x 0.1) & (-3db = E max. x 0.707)

I could also define the out of band selectivity in terms of Filter Skirt Slope (db / octave) but I decided not to go there. I am also not going to talk about circuit quality or "Q" because there seems to be a lot of confusion as to its meaning in overall Xtal Set performance.

Lets Test!!

SENSITIVITY OR EFFICIENCY TESTS:

Finally we are ready to make the Crystal Set Sensitivity or Efficiency measurement as follows;

1. Set the Signal Generator to the first test frequency. (unmodulated)

All Crystal Sets produce a DC output voltage that is proportional to the RF input level and this same proportionality applies to modulated signals producing audio output signals. Proof of this is too long to present here.

Test setup:
Set up the test equipment, the test circuits and the Crystal Set as follows; Connect the output of the unmodulated Signal generator to the 15 Ohm resistor in series with the 200 pF capacitor in series the 20 uH coil. Connect the other end of the coil to the Antenna terminal of the Crystal Set. Connect the low or ground output of the Signal Generator to the 10 Ohm resistor. Connect the other end of the 10 Ohm resistor to the Low or Ground input terminal of the Crystal Set. Note; This Crystal Set terminal is often not in common with (connected to) either of the output terminals. Place the Crystal set on an insulated surface such as a book for RF isolation purposes.

Connect the precision 2000 Ohm resistor across, or in parallel with, the headphone terminals. Connect the 100 kOhm resistor to one of the headphone terminals. Connect the other end of this resistor to one input of the DVM. Connect the other headphone terminal to the other input of the DVM. Connect a .01 uF capacitor in parallel with the DVM. Place the DVM and all DVM leads on a second RF insulated surface like another book.
Replace galena or other diode types with the 1N34 type. I use clip leads.

Test definitions or characteristics definitions:
The only item that I had to define so far were the values for the Dummy antenna. Efficiency by general definition is the ratio of the power output of any device to the power input to it or Pout / Pin.

DESIGNING AN AIR-CORE INDUCTOR

by Kenneth A. Kuhn

http://www.kennethkuhn.com/students/crystal_radios/designing_inductors.pdf

Introduction

This chapter describes the mathematical process for designing an air-core inductor comprised of a single layer solenoid winding over a rigid coil form. Although the development of the mathematics is a bit complicated the final result is simple to apply. For practical reasons, this chapter will make use of English rather than metric units. A well designed and constructed air-core coil has better performance than those with ferrite cores. Ferrite acts as a flux multiplier and has the advantage that the physical size of the inductor can be reduced. That is very important for small radios and the chief reason ferrite is used. The price paid for small size is loss of performance but that loss is generally negligible in active radios. The loss is not bad for crystal radio performance and many good crystal radios have been built using ferrite core inductors. But ferrite is not required. Purists correctly argue that the coil should be air core as that is how early radios were built. A mediocre ferrite core inductor will work considerably better than a poorly designed air-core one and that has probably led to the popularity of ferrite as the process for designing good air-core inductors is not widely known. This chapter reveals those secrets.

Analytic equation

The classic equation (which you can find in any book or article about winding inductors) for calculating the inductance of a given single layer coil is (Reference 2):

$$L = \frac{r2*n2}{9*r + 10*l} \qquad \text{Eq. 1}$$

where:
L = inductance in microhenries
r = coil radius in inches (center of coil to center of conductor)
n = number of turns
l = coil length in inches (center of starting turn to center of ending turn)

This equation is generally accurate to around one percent for inductors of common dimensions. It is more convenient to work with coil diameter and Equation 1 can be written as:

$$L = \frac{d2*n2}{18*d + 40*l} \qquad \text{Eq. 2}$$

where d is the coil diameter in inches (center of conductor to center of conductor)

Example: What is the inductance of a coil has a diameter of 2.5 inches, a length of 2.33 inches, and has 72 turns?

$$L = \frac{2.5*2.5*72*72}{18*2.5 + 40*2.33} = 234 \text{ uH}$$

Development of design equations

Equations 1 and 2 are fine for determining the inductance of an existing coil but are very awkward to apply to the design of a desired coil as there are many variables. Any time there are a multitude of variables then the possibility of optimum combinations or relations should be explored. In the following development the number of variables is reduced by finding

recommended by the IRE in 1930 for general radio testing. The circuit consists of a 25 Ohm resistor in series with a 200 pF capacitor in series with a 20 uH coil. I split the 25 Ohm resistor into a 15 Ohm resistor and a precision 10 Ohm resistor. The 10 Ohm resistor is connected between the low or ground side of the signal generator output and the input ground terminal of the crystal Set. This was done so that input current to the Crystal Set can be easily measured. The input current is then equal to voltage measured across the 10 Ohm resistor divided by 10.

The second test circuit is the Crystal Set output load that will be used in place of headphones. Headphones all have three values of impedance; the first is simply the DC resistance, the second is the AC impedance at audio frequencies, and the third is the AC impedance at RF frequencies. In order to remove the characteristics of different headphones from the tests I simply replace the headphones with a precision 2,000 Ohm resistor. In these tests I am only interested in measuring the DC output voltage from the Crystal Set so a simple low-pass filter that does not effect the output voltage is added. It is simply a 100 kOhm resistor in series with the DVM and a .01 uF capacitor in parallel with the DVM.

Test frequencies:

The following are the test frequencies to be used. They are 400, 600, 800, 1000, 1200, 1400, and 1600 kHz. Many of the older Crystal sets will not tune to frequencies above 1000 kHz but do as many as the set can tune. This next aspect of the tests may confuse some people. All tests to be conducted are made using an UNMODULATED input signal. This increases the accuracy of the tests and eliminates the need for a Signal Generator that can produce an accurate percentage of modulation. (Most Generators cannot)

Test equipment:
The first item to consider for testing is the test equipment. These tests require three pieces of test equipment; a RF Signal Generator, a RF Voltmeter, and a Digital Multimeter. The Signal Generator should have low output impedance. Most Generators by HP, GR, and others have output impedance of 50 Ohms. When terminated this impedance drops to 25 Ohms. Many so-called Function Generators that have a Sine Wave Output will work just as well for these tests. The impedance does not matter for the sensitivity tests but can be factor in the selectivity tests. The RF Signal generator does not need to have any modulation capability at all. A less common piece of test equipment is the RF Voltmeter. This meter must be able to measure RF voltages from a few milliVolts to over one Volt. A high input impedance of over 10,000 Ohms is also required. A well-calibrated Oscilloscope will meet these requirements but will be less accurate. I use a HP model 3400A RMS Voltmeter. The third piece of test equipment is a battery powered 3-1/2 digit Digital Voltmeter (DVM) or Digital Multimeter (DMM). Everyone that does anything electrical or electronic should have one of these. The DVM must be battery powered to make it RF isolated from ground. Standard diode or diodes need to be used. Some sets use two. I decided to use the popular and common 1N34 type germanium diode.

Test circuits:
The next items are two simple test circuits. The first is the circuit to connect the Signal Generator to the Antenna and Ground terminals of the Crystal Set. This circuit is known as an "Artificial Antenna", "Dummy Antenna", or "Standard Input Circuit". Its purpose is to simulate a typical "Long Wire" or "Marconi" antenna so that the Crystal Set can be tuned in the normal manor. The values chosen to be used are those

relations to known constants. We first replace the coil length by a factor that relates it to the diameter.

$$l = k*d \qquad \text{Eq. 3}$$

where
l = coil length in inches
k = a dimensionless constant
d = coil diameter in inches as before
Substituting Equation 3 into Equation 2 gives:

$$L = \frac{d2*n2}{18*d + 40*k*d}$$

which reduces to

$$L = \frac{d*n2}{18 + 40*k} \qquad \text{Eq. 4}$$

It can be shown that the value of k that minimizes the length of wire to wind the coil is 0.450. However, other research indicates (see Reference 1) that the value of k that minimizes coil losses is approximately 0.96 even though that value uses about twenty percent more wire. Factors contributing to coil losses include:

- * Ohmic losses in the wire including skin-effect
- * Dielectric losses in the coil form and nearby materials
- * Dielectric losses in the insulation around the wire
- * Induction losses in nearby materials

There are also losses caused by adjacent turns being too close together. It has been found (see Reference 1) that the optimum spacing (wire center to wire center) of adjacent turns is between about 1.3 to 2.0 times the diameter of the conductor. Coils for crystal radios are commonly wound using what is known as magnet wire (thin enamel insulation) and the turns are tightly wound next to each other corresponding to a spacing factor slightly greater than 1.0 (the thin insulation is of finite thickness). Although it is less than the optimum discussed it works well.

Without some special technique (such as a lathe) it can be very difficult to manually wind a coil with controlled spacing between the turns. One easy method for achieving a spacing factor of 2.0 is to wind two wires tightly side by side at the same time and then remove one of the windings when finished. Smaller spacing factors can be achieved using a smaller diameter wire for the spacer but the difficulty of controlling two wires will increase. It might occur to someone to use a wire with a thicker insulation so that a spacing is naturally formed with a tight winding. The problem with this method is that the insulation may increase dielectric losses and become self defeating –although this may be a small issue –be sure to try it before tossing the concept. This method can work great if Teflon wire is used as that is a very low-loss material and the internal wire strands are silver plated.

Equation 4 can be used to determine the optimum coil diameter for a given inductance and wire size. We note that the coil length is the number of turns divided by t (turns per inch of the wire). We also note that the coil length has previously been related to the coil diameter by the constant, k. Thus:

$n = k*d*t$ Eq. 5

Substituting Equation 5 into Equation 4 gives:

CRYSTAL SET TESTING

Charles A Lauter. (email: Lautron@aol.com)

Background:

The following are standards and methods for testing Crystal Sets for the purpose of comparing the performance of different Crystal Sets, Crystal Set circuits, and Crystal Set components in an objective manor. Standards for testing all other types of radios have been in effect since about 1930. Crystal Sets have been in use for almost 100 years but no performance measurement standards exist that I can find. I have read many-many claims of performance for various circuits, coil types, wire types and so on but all without performance numbers that have been measured. There are only two parameters or characteristics of all Crystal Sets that are important. They are the sensitivity or efficiency and the selectivity or bandwidth characteristics. The ability of a Crystal Set to match its antenna is very important and I am including it with the efficiency or sensitivity characteristics. These tests cover Sets designed to tune all or most of the AM broadcast band of frequencies using a standard antenna. More about the antenna below. I am not measuring characteristics such as extended tuning range or the ability to work with very long antennas or any other features not common to most Crystal Sets. In all fairness, these tests were very difficult or impractical to perform back in the 1920s or earlier. Today these tests can be performed with what I consider a modest investment in test equipment. These tests will allow me to compare my collection of Crystal Sets with each other and with other designs of circuits and components. They will also permit comparison with other people's Sets and designs. The reasons for the following test equipment and test circuits will become apparent as the test procedures are followed.

loudness. Similarly, two 3 dB steps would be sensed as "equal" changes in loudness.

$$L = \frac{k2*t2*d3}{18 + 40*k} \qquad \text{Eq. 6}$$

We will use 0.96 for k and t will be that of the particular wire we have available. Solving Equation 6 for the optimum diameter gives:

$$d\_optimum = 4*(L/t2)1/3 \qquad \text{Eq. 7}$$

Figure 1 shows a plot of this Equation 7 for common wire sizes. In all cases the turns are close-spaced. The lower curves are for common enamel insulated magnet wire. The two upper curves are for vinyl insulated house wire which can be considered if a large diameter coil form is available. To use the curves, select the desired inductance and the wire size that will be used. Look up the optimum coil form diameter and then use the closest practical form you have to that size. The optimum is broad so do not worry about being exactly on it. Note that the true diameter is the sum of the diameter of the coil form and the diameter of the wire since by definition the coil diameter is measured between opposite centers of the wire.

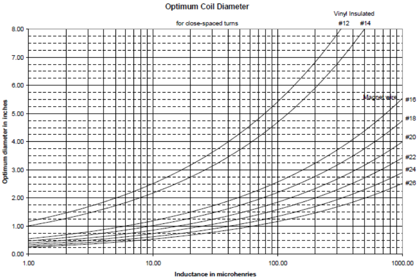

Figure 1: Optimum coil diameter

The following table provides typical values for t (turns per inch) for some common wire sizes:

Table 1: Wire data

Gauge	t	Comments
12	6.2	Vinyl insulated house wire
14	7.7	Vinyl insulated house wire
16	19	Enamel insulated magnet wire
18	24	ditto
20	31	ditto
22	39	ditto
24	50	ditto
26	62	ditto

The length of the winding will be the number of turns divided by the turns per inch of the wire. That is:

---

know how much better (or worse) a change to it is, and can express it in deciBels (dB). Say that you are taking voltage measurements across a resistor as I do. If the voltage goes from 20 to 40 milliVolts, the ratio of voltages is 2; but wait. Power varies as the square of voltage, so the power ratio is 2 squared, or 4. The logarithm of 4 is 0.6, and 10 times that is 6, or 6 dB. Incidentally, if the measured voltage had gone down from 20 to 10, you would have gone down by 6 dB - it works in both directions. So, all you have to do is divide the larger voltage (or current) by the smaller, square the result (this is your power ratio), and multiply the logarithm of that number by 10 to get the change in dB. Round your answer to the nearest whole number.

If this seems like too much work, here is a short table of common power ratios, the corresponding dB change, and what it means to your ear:

Power ratio	dB	Effect
2	3	barely detectable change
4	6	noticeable
5	7	a little better than 6
10	10	oh yeah
20	13	a smidgen better than 10
100	20	oh yeah again -
200	23	(just to show you the effect of multiplying two numbers by adding their logs)

Note: to the ear, an increase of 10 dB, followed by an increase to 100 dB above the original power level (another 10 dB increase) would be sensed as two equal step increases in

Sony is tuned to just as the cap is closing. That is the minimum f for the LC. More tweeking will determine what f the LC will tune as it is coming fully open. Those 2 f indicate the frequency range of that coil and cap combination.

This will work for MW BCB (midwave broadcast band) using a MW transistor radio, but be sure that the coil on the internal ferrite antenna is parallel with the coil under test.

DECIBELS FOR DUMMIES (how to sound like a tekkie)

The ear is a marvelous device which can process a wide range of noise levels. At a frequency of 1000 Hz, the weakest sound detectable is at a power of about 1 picowatt per square meter, normal speech is about at 1 microwatt, hearing damage starts at about 1 milliwatt (smoky bar with a loud band), and physical pain is felt at 1 watt. Since the ear doesn't respond to absolute power changes in a linear manner, and to cut this trillion to one ratio down to size, we use a logarithmic scale instead. The common logarithm of a number is the exponent or the power to which 10 must be raised in order to obtain the given number. E.g. the logarithm of 100 (10 to the second power) is 2, of 1000 is 3, and so on. Power ratios can be expressed as a logarithm, known as a Bel (after Alexander Graham), but more commonly as a deciBel, obtained by multiplying the logarithm of the power ratio by 10. This gives us a range of hearing loudness from 0 to 120 decibels, relative to the threshhold of hearing (1 picowatt per square meter), a much easier, and more useful set of numbers to bandy about. When working with your crystal set, you normally want to

$$l = n/t \qquad \text{Eq. 8}$$

We now substitute Equation 8 into Equation 2 and solve for n

$$d2*n2 - 18*d*L - 40*(n/t)*L = 0 \qquad \text{Eq. 9}$$

$$t*d2*n2 - 40*L*n - 18*d*t*L = 0 \qquad \text{Eq. 10}$$

Solving for n gives:

$$n = \frac{20*L + \mathrm{sqrt}(400*L2 + 18*t2*d3*L)}{t*d2} \qquad \text{Eq. 11}$$

Although a precise value (in inches) for the length of wire required can be calculated using trigonometry for a spiral, a very close value can be calculated as

$$w = pi*d*n \qquad \text{Eq. 12}$$

Remember that d is the sum of the coil form diameter and the diameter of the wire. This approximation assumes that the diameter of the wire is very small in comparison to that of the coil form. Also remember to allow an extra couple of inches for connecting leads at each end of the coil.

Example: A 300 uH coil is needed. The expected Q should be over 350. What coil diameters and wire sizes could possibly meet this?
Solution: Using Figure 2 it can be seen that wire sizes #12, #14, #16, #18, and #20 could achieve the required Q. Using Figure 1 the required coil form diameters are:

Optimum

Gauge	diameter
#12	7.75”
#14	6.75”
#16	3.75”
#18	3.15”
#20	2.65”

A piece of 4.5” OD PVC pipe is available and #14 electrical wire is available. From Table 1, #14 insulated wire will make about 7.7 turns per inch. Thus, the effective diameter is 4.5 plus 1/7.7 = 4.63 inches. Using Equation 11 the number of turns required is 86. Using Equation 8 the length of the winding is 11.2 inches. The length/diameter ratio is 2.4 which is a bit longer than the optimum of 0.96. The length of wire required is given by Equation 12 and is 1,251 inches. The length would have been 1,058 inches if the optimal diameter could have been used. This extra length will cause somewhat higher losses –it might still meet the desired spec though. This is about as far as I would go in rounding to an available coil form diameter.

Estimation of Inductor Q

All inductors have an equivalent series resistance loss as discussed earlier and is comprised of a number of components. We measure the quality factor or Q of the inductor by computing the ratio of inductive reactance at the frequency of interest to the series loss resistance as follows:

$$Q = \frac{XL}{Rs} \qquad \text{Eq. 13}$$

where

Q is the dimensionless “quality” factor of the inductor

---

resonant frequency of your coils. If your coil has a lot of self capacitance, it will show a dip at a lower frequency somewhere in the HF region than one that has little self capacitance. I found the oscillator particularly useful when working with short wave xtal sets. It allows you to check the tuning range of your tank circuit with enough accuracy to put you into a favorite short wave band. It really came in handy when trying to figure out the details of a shortwave antenna tuner. By coupling the dipper loosely to the tank coil of the tuner, you can check the resonant frequency of the antenna/tuner combination. It seems that with the short wave bands, your favorite xtal set antenna is actually longer than a quarter wavelength, and tuning it to a band can be somewhat tricky. The GDO helped me get the antenna to resonance.

Here is another testing gimmick courtesy of Steve Holden:

> Unless you are following somebody else's plan to the letter and using all parts that are exactly the same as he had, there will always be some "cut and try" to be done during the "breadboard" stage. Since all of my electronic test equipment combined would fit in a shoebox I use a few tricks that Pros would not use. For example if I want to find out what f range a coil and cap will tune I use a Sony shirt pocket SW (short wave) radio. I turn the tuning of the Sony to an open spot in the f range I think the LC will tune, then, with whip not extended and the radio volume set to hear noise softly, I place the LC under test near the whip and turn the varicap. If I guessed right, at some point the volume from the Sony will come up. This is because the LC circuit is functioning as a tuned loop antenna. Some more experimenting will allow me to determine what f the

Another useful piece of test equipment, also excellent for chasing dx stations, is a receiver with a digital frequency readout. My Radio Shack set, a pocket portable, fits the bill and costs about 50 bucks or so, depending on how the sales are running. It tunes in 10 kHz increments on the AM band, and I am confident the actual frequency is accurate to within a hundred Hz or less. You can use this to check the output of your sig gen, and to check its frequency; your less expensive and older signal generators have analog tuning and readout, and their calibration can be suspect. It also can be used to calibrate the dial of your crystal set in conjunction with the signal generator to at least the accuracy of a standard analog pocket superhet. You might consider this last to be gilding the lily, but having a known "dial marker " or two can keep you oriented when you're searching the dial.

A meter capable of inductance and capacitance measurements is nice to have, and I use a friend's whenever I make a new coil or want to check an unknown capacitor. Haven't seen fit to spring for one myself however. If I need to, I can always check the inductance of a new coil using a known capacitance with it in a tank circuit and the signal generator. Q meters? They're out of my league, so I depend on the kindness of strangers to tell me how to optimize coil Q.

I recently acquired a Heathkit grid dip oscillator, and found it to be a useful tool, even though its frequency range starts at 1.5 MHz. Actually, it produced harmonics, or probably parasitic oscillations into the BC band, but isn't calibrated for them; I should probably wind a plug in coil just for the BC band, but that's for another day. So what can you do with one of these? Well, you can start by using it to check for the

XL is the inductive reactance in ohms at the frequency of interest
Rs is the equivalent series resistance in ohms at the frequency of interest

Note that inductive reactance, XL, is calculated as

XL = 2*π*F*L

where
F is the frequency in Hz
L is the inductance in henries

The equivalent series resistance is the net of ohmic losses including skin effect, dielectric losses in distributed capacitance and coil structure, absorption losses by nearby conducting media, magnetic losses in nearby magnetic media, etc. With care these losses can be kept small but it takes very little loss to reduce the Q of an inductor from 400 to 200. The magnitude of Rs can be measured on sophisticated impedance equipment but it is hard to calculate the effect of all factors. Figure 2 shows an estimated value of Q at 1 MHz considering typical losses assuming the coil is wound optimally and is not disturbed by nearby lossy materials. Use the figure only as a guideline as your specific results may be better or worse. The expected Q at 540 kHz will be between about 50 to 70 percent of what is shown and the expected Q at 1.6 MHz will be around 1.2 to 1.5 times that shown.

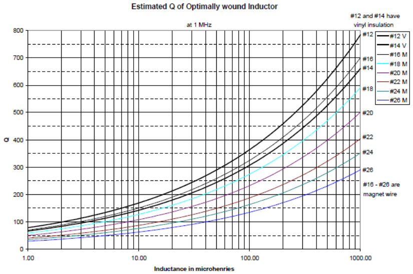

Figure 2: Estimated Q of Inductor

The Q we obtain from Equation 11 is for the unloaded coil (i.e. antenna and crystal detector not connected). The net loaded Q will typically be significantly smaller but ideally (as discussed in another chapter) would be in the general range of one hundred. Thus, we would like to start with an unloaded Q of several hundred. As can be seen in Figure 2 the Q of the inductor can be made higher by using larger diameter wire. From Figure 1 this also means using a large diameter coil form. This is a very important conclusion –high Q coils need to be physically large.

Type of wire

The only material to consider for the wire is copper. A variety of styles of copper wire is readily available. The most basic choice is between solid or stranded. Although a variety of arguments can be made for and against each, in practical terms you will not notice any difference in performance although one

---

hobby, and if I can have this much fun with a handful of cheap parts and some 20 year old, bottom of the line test equipment, working on a table set up in the laundry room, so be it. I have tried to use my "super convertible" as an rf signal generator in the regeneration mode, and it works pretty well. Just put a 2k resistor in place of the headphones, and you can used it as a nice, unmodulated signal. Yeah, it drifts some, but as long as you don't have an antenna hooked up to it, it's pretty stable. Shouldn't be to hard to couple a one transistor audio oscillator to it for some modulation. Transtronics sells their RF Signal Generator SK-303 kit for about 28 bucks. Maybe I should give it a try.

The multimeter connected across your earplug resistor also is most useful in comparing different detectors. What you are usually seeking is the best sensitivity, so first find a weak but stable station (or detune your set to make one), which gives you about 10 mV or so across the resistor (I use this value with the 47kohm resistor). Then you can readily compare different detectors by swapping them in and out of the circuit. With a digital voltmeter with readouts to the nearest 0.1 millivolts, go for a reading of about 2 mV with the station very weak but still clear. This goes faster if you use a couple of alligator clip leads for the connection. In a 10-pack of 1N34A diodes, I have found the voltage developed across the diode to vary as much as 30%; at this level, that equates to about a 2 to 3 dB difference in signal strength, even though the audible difference is small. For stronger signals, the difference in detector sensitivity is less, but doesn't matter.

Use the same setup to check out the effectiveness of your wave traps when you're cutting how the high powered stations.

the BC band, where the antenna tuner coil alone improved the signal from a 135 foot antenna by as much as 12 dB. When I put it on my Radio Shack crystal set, after first adding the obligatory 47 kohm resistor across the earphone, I found the sensitivity for it and my selective crystal set to be within a few millivolts of each other. The list of what you can do from this point is long enough to keep you busy for hours on end. One use that I also found very useful is to compare the sensitivity of different detectors; sometimes within a batch you can see differences of as much as 10 dB. On to phase 3:

For this part, you need an rf signal with some audio modulation. I happened to find an ancient but usable rf signal generator in my school lab which fit the bill. It has to be variable in frequency, and you need some way to check its output frequency, because we are going to try to measure the selectivity of your set. I put the output of the generator across the antenna coil of the selective crystal set, and adjusted the signal to a comfortable level on the voltmeter described above, about 50 mV. Then I adjusted the signal frequency up and down by 25 kHz and 50 kHz, and measured the signal output across the earphone. From a center frequency of 1100 kHz, the signal strength dropped by 12 dB at 25 kHz off, and by 20 dB at 50 kHz off. Running the same test on the Radio Shack set, the change was 2 dB and 5 dB, respectively. I don't know if my set is really any good since I haven't done much comparison testing, but compared to it, the RS kit selectivity sucks. For both rigs, selectivity improved some at lower frequencies, and was a bit worse at the top of the band. At this point, some electronic engineer type is probably reading this and shouting obscenities at his monitor for my sloppy, unsophisticated, and jerry built methods. Well, this isn't Nature magazine and I'm not trying to sell cold fusion. It's a

or the other may have physical advantages for your particular construction method. Avoid wires that are plated as those will have higher losses since skin-effect will cause most of the conduction to be in the plating which has higher resistance than copper. Avoid wires with rubber or cheap plastic insulations as dielectric losses will be higher. An exception is silver plated Teflon wire as that has the best conductivity and the lowest dielectric losses –but it is expensive.

For use in low to medium frequency inductors there is a special wire called Litzengrad or just Litz for short. It is designed to minimize skin-effect losses and is made by assembling many strands of enamel insulated magnet wire together to form a wire that has a large surface area. Litz wire is not easy to find and tends to be expensive. If you are going to use Litz wire then make sure that other losses as previously discussed are minimized. Otherwise Litz wire will make little if any difference and will be wasted effort and expense. Avoid belief in a variety of myths about skin-effect. Although it is true that skin-effect is more severe on large diameter conductors, a larger diameter still conducts better than a smaller diameter at any frequency. This can be seen in Figure 3 which shows the frequency dependence of the resistance per meter factor of some common wire sizes.

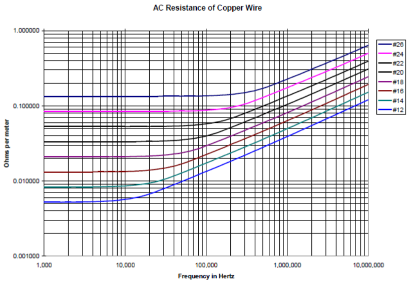

Figure 3: AC Resistance of Copper Wire

Coil Forms

From a loss standpoint air is the best coil form material there is. The obvious problem is that air has no structural strength. However, there are methods used by commercial inductor companies that employ a minimal structure so that the coil form is around 99 percent air. Manually, you can achieve the effect by first winding the coil using large diameter solid copper wire (i.e. #18, #16, #14, etc.) on a rigid coil form and then carefully sliding the winding off of the form. The stiff wire will retain the shape and you can easily space the turns to the optimal discussed previously. You will need a few supports to keep the whole thing from being too loose.

A popular coil form is some kind of cardboard tube that you have salvaged from a variety of sources such as used for paper towels or shipping tubes. These are great if you are using small diameter wire as small wire will not self support. Plastic pipe

with it? First, and obviously, you can use it to tune a station for the best signal. On a strong station, you will be amazed at how much the signal strength can vary and it not make a heck of a lot of difference in earphone volume. This is because the ear is a marvelous signal processor itself, and can accommodate a range of loudness, or signal power to the ear of 90 decibels, that's ten to the ninth power, otherwise known as a billion to one between the softest whisper it can detect and the point at which the ear starts to suffer damage with chronic exposure. Multiply that loudest sound by another thousand, you are now at a trillion to one, and the sound is so loud that it can cause physical pain. In order to handle this range of signal power, the response of the ear is logarithmic, and the power of a sound has to change by a factor of two for you to readily detect it. Since power varies as the square of voltage, this means you usually need to see a forty percent change in voltage, or about a 3 dB change in power to notice the difference in loudness. The voltmeter, being more responsive than the ear to small changes, can help you fine tune to a station, and then compare the difference in signal strengths from different stations using a number rather than an impression. Once you have played around with this a bit, it's on to phase two:

Now that you have a signal strength meter, you can start using it to optimize the settings on your rig. Using it on a receiver similar to Radio number four, which you can find on my project page, I found that the optimum tap for sensitivity and selectivity was at about 35 percent of the coil from the ground end. The 20 turn tap for the detector (for sharp tuning) showed a 9 dB loss of signal strength compared to the best tap, while giving less than a 2 dB gain in selectivity. As expected, using an antenna tuner helped, particularly in the low end of seriously load down the circuit. The digital multimeter has a virtually imperceptible effect on the signal. I tried once putting a galvanometer with a 50 uA movement in line between the detector and the earphone but the reading, which was only about 4 uA, as expected, wasn't enough to make this useful, being at the very bottom of the scale. Anyway, for my money, the high impedance voltmeter makes a much more useful tuning indicator.

If you are using magnetic headphones, which don't have the 47k resistor across them, you will not get much of an indication on the meter. Just put a 47k or some other reasonably high value resistor in series with the headphones and measure the voltage across it - yeah, the resistor cuts your signal a tad, but you're only putting it in long enough to sharpen up the rig's tuning anyway. Incidentally, there is nothing magic about the size of the resistor; I just happen to have several on hand. If you use a larger resistor, you will get a higher reading on the voltmeter, but will also have a corresponding smaller signal in your magnetic headphones. Use the smallest resistor that gives you a usable reading on the meter - this is not an absolute signal level you are measuring, but rather a tuning aid. Just to see what I could do, I tried out a cheap multimeter with a 1000 ohms/volt meter sensitivity. If I tuned in a reasonably strong station, and just put the meter in place of the 47k resistor, I could indicate enough level on the 2.5V scale, in the neighborhood of 200 mV or so, to make some use of the meter. This is pretty low on the scale, but still enough to see the effects of tuning the rig.

Okay, so now you have a tuning indicator/signal level indicator or whatever you want to call it. What can you do

is another material you might consider. None of these materials are made with any consideration about high frequency dielectric losses but because only a small amount of material is used the losses are probably minimal.

Wood is a convenient coil form and has low losses if very dry. Common sizes that have been used are 2x2, 4x4, or a pair of 2x4 combined to make 4x4. Round dowel rods may also be used but their diameters are often much less than optimum. When using a square coil form there is a logical question about how that affects the inductance calculations. A simplistic (but good) answer is to use an effective circle diameter that has the same area as the square form since area is a strong factor in inductance. Losses with a square form will be somewhat higher than for a circular form. Rectangular forms (such as a single 2x4) have even higher losses in comparison –it takes more wire to encompass a given area.

Winding the coil

Counting turns is a tedious and error prone task. It is much simpler to cut the length of wire needed and then wind that until finished. The resulting turns count will be very close if not exact. Wire is springy and will jump off the form in a tangled mess if not restrained. Start by securing the wire at one end of the form and have a means for easily (preferably with one hand) securing the opposite end when you finish. It is tempting to use some kind of adhesive tape and that will work if you are careful and understand what you are doing. The forces will build and the tape may give way which will result in a frustrating mess of tangled wire. Make sure the tape can not slip. A good way to secure the ends is to first drill a hole in the tube at the starting and end points. Then feed the starting end through the starting hole and bend the wire such that it naturally resists tension and secure the wire with tape. Stuff the loose end length inside the tube so it is out of the way while winding.

It is best to wind the coil by hand as the set up for using a lathe is not worth the trouble for a single coil. There are a number of “poor man’s” lathes such as a power drill that have been used but I do not recommend that as you are more likely to make a mess or cause injury than you are to wind a coil. It only takes a couple of minutes to wind a coil by hand so take the time to think what you are doing. It is important to keep the winding tight at all times. The wire will spring off if it ever gets loose. You will very likely have some fractional turn as a result of your calculations. I recommend that you round that to the nearest integer as it is not worth the trouble of making measurements to stop at a specific fractional turn.

References:

1. Electronic and Radio Engineering, fourth edition, Frederick Emmons Terman,
McGraw-Hill Book Company, 1955, pages 30 – 33.
2. The Radio Amateur’s Handbook, 44th edition, 1967, The American Radio
Relay League, Newington, Conn., page 26

TESTING YOUR CRYSTAL SET

(22 Nov 01)
Owen Poole
http://bellsouthpwp2.net/w/u/wuggy/testing.htm

If you like to fiddle around with your crystal sets, then you need some way to do comparisons of before and after when you want to try something new. If course, the obvious is to listen. If it sounds louder, it is. If you don't hear the other station that is close in frequency any more, then selectivity is better. If you hear everything you want, then your set is working fine, and you need go no further. If you are trying to push it a bit, however, some reasonably objective yet quantitative measurements are helpful for you to optimize your set and to compare the performance of different sets, antennas, modifications, etc. Here are a couple of things I do with my rigs:

Get a DC voltmeter which will measure small voltages in the millivolt range. I use either a digital multimeter or a RS analog multimeter with about 20 thousand ohms per volt sensitivity (less sensitive meters will load down the circuit unacceptably, and will probably not measure down in the millivolt range, but go ahead and try if you have to). The RS analog model I have has a 0.6 VDC (600 millivolts) scale which works pretty well. I then connect the meter across the 47 kohm resistor which is also across the crystal earphone. Using this setup, my closest local BC station, less than a kw about 6 miles away, gives me a reading of about 190 millivolts. A station with a 50 kw daytime signal about 75 miles distant comes in at about 30 mV. With the analog meter, there is a little insertion loss, about 10 mV, but not enough to

ANTENNA MATCHING

by Kenneth A. Kuhn
http://www.kennethkuhn.com/students/crystal_radios/antenna_matching.pdf

As discussed in an earlier chapter the ten to thirty meter antenna used for crystal radios has a very low resistance and a high capacitive reactance. The ground resistance discussed previously is typically in the several tens of ohms and is effectively in series with the antenna. As an example, an antenna/ground system may have an impedance of 20 –j1000 ohms at 1 MHz. For maximum power transfer from the antenna to the resonant circuit the input impedance of the set should be a conjugate match –that is have similar resistance but the reactance will be equal in magnitude but opposite in sign. For the example this means that the crystal set should have an input resistance of about 20 ohms and a reactance of about +j1000 ohms (160 uH) at 1 MHz. The positive reactance is obtained by an inductance in series with the antenna circuit. This inductance should be variable to tune out the capacitive reactance of the antenna across the AM broadcast band. Tuning is not sharp as the Q of resonance is low. Table 1 shows some typical values. Note that the inductance tuning range becomes wider for longer antennas since the capacitive reactance drops rapidly as the length approaches one-quarter wavelength at the upper end of the AM band. It is important that this series inductance have very low losses or the advantage of using it will vanish –a lossy inductor could be worse than nothing.

Antenna Length	**Typical Antenna XC @ 550 kHz**	**Typical Antenna XC @ 1.7 MHz**	**Antenna series Inductor tuning range**
10 m -	j4040 ohms -	j1250 ohms	1200 –120 uH
15 -	j2680 -	j780	780 – 70
20 -	j1990 -	j530	580 – 50
25 -	j1580 -	j360	460 – 30
30 -	j1290 -	j240	370 – 20

Table 1: Antenna series inductance tuning range

One effect of not tuning out the reactance of the antenna is that the resonant frequency of the tuned circuit will shift because the antenna becomes a reactive load. One way to know if the series inductance has been tuned to the right value is that the station is received at the calibration point –assuming the radio tuning was calibrated.

The next issue is creating a low input impedance of several tens of ohms. There are two ways to do this and they are essentially the same. One method is wind a turn or so of wire near the ground end of the coil of resonant circuit –one end of the wire goes to the series inductor to the antenna and the other end connects to ground. This small winding transforms the low antenna/ground impedance to a high impedance across the coil. The other method is to make a tap a turn or so above the ground end of the coil of the resonant circuit to accomplish the same effect. If the winding or tap is too few turns then there is an impedance mismatch and a weak signal will result although selectivity will be relatively sharp. If the winding or tap has too many turns then the coil is overloaded by the antenna/ground impedance which also results in weak signals and the selectivity will be broad. The optimum is the point of

Lyonodyne 17

Special Thanks to:
Mike Tuggle

L1 51 turns 100/44 litz on 5/8" dia. x 1-3/16" long ferrite rod, u(eff) = 10.33;
length = 1-35/64"; L = 146.6 uH; Co = 4.54 pF; Q @ 0.5 MHz = 649

L2 36 turns 660/46 litz, basket-wound over-1, under-1 on 5" dia. 13-point form;
length = 1.9375"; L = 185.6 uH; Co = 7.40 pF; Q @ 0.8 MHz = 1134

L3 36 turns 660/46 litz, basket-wound over-1, under-1 on 5" dia. 13-point form;
length = 1.9375"; L = 185.7 uH; Co = 7.40 pF; Q @ 0.8 MHz = 1060

L4 42 turns 660/46 litz, basket-wound over-1, under-1 on 4-1/4" dia. 13-point form; length = 2.25". L= 186.7 uH; Co = 6.45 pF; Q @ 0.8 MHz = 1082

R1 500 k-ohm pot.

S1 DPDT miniature toggle

T1 UTC A-27: input transformer --100 k-ohm primary; secondary taps, 50 to 600 ohms -- around 200 ohms seems best with the RCA phones below.

PHONES: RCA MI-2045 sound powered phones, known as "Big Cans" in the trade.

All variable capacitors have vernier drive dials.

proper impedance match although it is not very critical. The issue is how to determine the proper point. A second issue is that the required impedance level varies across the AM broadcast band. Thus, the tap should be variable.

If the coil is wound on a ferrite toroid core which enables a high degree of flux coupling from turn to turn then it is fairly easy to calculate at what turn a tap should be for the desired impedance transformation. However, our coil is typically an air-core solenoid which has a complicated flux relationship. Calculation is difficult and very error prone. The best way is to make a variety of taps and measure the impedance using laboratory methods and note the results. I will present the data of just such an experiment on a typical coil for crystal radios when this article continues …

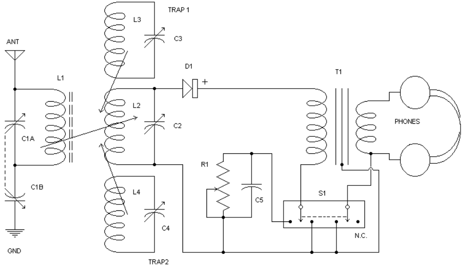
TRAP 1
L3
C3
ANT
D1
+
T1
L1
L2
C2
PHONES
C1A
R1
C1B
L4
C5
S1
C4
N.C.
GND
TRAP2

Lyonodyne-17 Parts (as of 2/5/01)

C1 two gang, 15-470 pF per gang, variable capacitor, high parallel leakage
resistance, Rp.

C2 R/C (or TRW) 15-497 pF straight line frequency variable capacitor, Rp on the order of 20 megohms. No. 107-C-522; FSN 5910-546-9239. Fair Radio Sales cat. no. C123/URM125

C3 same as C2.

C4 same as C2.

C5 0.02 uF polystyrene.

D1 selected (by listening) Radio Shack 12101 3RT germanium diode.

took out the tree at the other end. The shorter flat top worked so well I brought it here to Hawaii, where real estate for long antennas is at a premium. It's about all I can fit gracefully onto this lot. Ground is tapped directly into the city water system.

Circuit Details

The antenna-ground system is tuned by the primary, L1-C1, and is coupled to the secondary, L2-C2, via loose L1 - L2 coupling (physical distance 3 to 4"). L1 and L2 are positioned radially (side-to-side), while the coils of two QRM traps (high-Q, parallel L-C ckts.) are coupled axially to L2. L1 has virtually no effect on L2-C2 tuning and Q, but the QRM traps do affect tuning, especially near the QRM frequencies. So, their coupling is kept to the minimum which will still allow reception of the stations of interest. The set is capable of receiving stations on frequencies adjacent to murderous locals. The low-Z sound-powered phones are impedance-matched to the secondary tank , L2-C2, and detector with a UTC A-27 100 k-ohm-to-200 ohm input transformer. The single12101 diode detector is estimated to present about 200 k-ohm resistance in the forward direction. These impedances, in series, allow you to place the detector-phones directly across the tuned circuit without undue loading. Taps, used in the past, are a mess -- especially with litz wire. Further, they kill a coil's Q, and are to be avoided if at all possible.

VARIATIONS ON A SLIDER-TUNED SET

Owen Poole

http://bellsouthpwp2.net/w/u/wuggy/slider.htm

My first crystal set used a prewound coil, a slider for tuning, a catwhisker and galena crystal, and had a large square mica capacitor (remember those?) across the headphone terminals, feeding a single 1000 ohm headphone. For a variety of reasons, which I don't remember 40 plus years later, it only got one or two stations in Atlanta, GA, but, even at that, it was still magic. Back in an era when responsible parents believed children should remain at least one economic level below their parents, having a radio of any type in my room was doing pretty well, and life was good. This type of set is tricky to use, since maintaining contact between the coil and the slider is touchy, and then you have to find a good spot on the galena crystal with the catwhisker, also easily knocked out of adjustment. Still, it has its advantages; it's inexpensive and it works. The headphone was a cost driver, and throwing in a variable capacitor could easily double the cost. A step up on the ladder came when AM tuning coils with slug-tuned ferrite cores came on the market, which were used in the tank circuit with a fixed capacitor, and feeding a germanium diode. I had a "pocket" receiver which had these as well as a pretty sensitive earphone which was built into the case.

About a year and a half ago I had a nostalgia fit and built another slider, but used a crystal earplug, a 1N34, and even added a 365 pF variable capacitor, remembering the tricky slide contact problem, but wanting to be able to vary the L/C ratio over a range of frequencies. The set worked pretty well, but lacked selectivity. This is a drawing of the set:

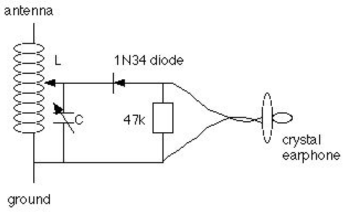

I found that the best place to connect the antenna and ground depended on the antenna used and probably on local conditions. Sometimes connecting the antenna to the slider worked well. I tried the detector tap at the top of the coil, as some designs do, but selectivity really suffered. Some selectivity improvement was seen when I wound a smaller coil for an antenna loading coil and put it up inside the tuning coil. Again, you can still play with where the antenna and ground are connected; you can put the antenna or ground to the loading coil, and try the other on either end of the tuning coil or on the detector connection. You can also make your earth ground to both the loading coil and the tuning coil. The best arrangement is what works best for you. An operating benefit of connecting the detector to the slider is that you don't hear anything unless the slider is making contact with the coil.

(3) if necessary, pad the variable capacitors with high-quality air trimmers to get to lowest frequency and spread out top-end tuning

(4) choose coil design which peaks in Q in the central BCB range -- operating near the self-resonant frequency is self-defeating

Superior quality tuning capacitors (silver plated with ceramic standoffs) were used in the present version. The two prime tuning capacitor candidates I had were a 500 pF and a dual, 470 pF per gang, both silver-plate, ceramic insulated. (Typical parallel leakage resistance measurements on these were 20 megohms, by the Boonton 260-A.) Accordingly, the set was designed about them.

I credit Bill Bowers of Oklahoma with exposing me to the virtues of finer-strand litz for medium frequencies, the use of larger coil diameter-to-length ratios (ca. 5-to-1) than theory suggests, and for going back to 'over-1, under-1' winding pattern. They are not as pretty and have lower L's than the same-sized 'over-2, under-1's', which I used for many years, but they do have substantially higher Q's.

And credit to Al Klase for preaching the virtues of sound-powered phones long enough and loud enough that I finally had to listen -- despite my being firmly convinced nothing could ever surpass Brush crystal phones. The sound-powered's are a real ear-opener, literally, and an order-of-magnitude improvement.

Antenna has been, for some time, a 50-foot, 4-wire flat top (wires spaced 1 ft. apart), maybe 25 feet high. At my previous location, I used to have a 105-ft. long-wire, but high winds

Summary

The "Lyonodyne-17" is a much-advanced version in a series of DX crystal sets evolving from circa 1974. I've been particularly interested in crystal sets since 1959, when I first discovered you could actually DX with them. It's a completely passive set -- no amps, no bias on the detector. The circuit is double-tuned and uses super-high quality components: silver-plated variable capacitors and high-Q litz wire basket-wound coils. Isolated coil Qs are in the 700s across the BCB and over 1000 for much of the BCB. In-set Qs are less, of course. Phones are high quality, surplus sound powered ("balanced armature") units, matched to the high-Z secondary tank/detector by a high quality line transformer. Detector is a single 12101 3RT, sold by Radio Shack as a "1N34 Type." It's the best I've found so far. With traps and all, it takes up the better part of a desk top.

Description

After several abject experimental failures, I concluded that the magical "high L-to-C ratio" for tuned circuits is a crock -- in spite of what theory might suggest. That, and the high Qs obtained in 200/44 litz wire (now, 660/46) coils got me to re-thinking my set's design and construction. This led to a completely rebuilt crystal set -- same circuit, but redesigned components per the following principles:

(1) design for the lowest L/C ratio as practical

(2) use highest quality variable capacitors available

My latest design attempt, which shows promise in the breadboard stage, goes back to a fixed capacitor in the tank circuit , but incorporates an antenna loading coil and separate detector taps, two features which I consider essential to selectivity in a single tuned set. The drawing for this rig is shown below:

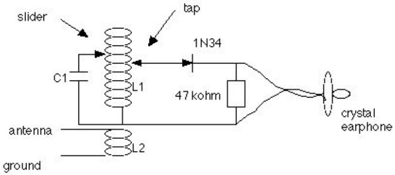

My coil was 100 turns of #26 magnet wire close-wound on a toilet paper core, with detector taps at 10, 20, and 30 turns from the bottom (connected) end. The antenna coil, L2, is 15 turns of #26, also on a tp core. I chose to separate the coils so I could move the antenna coil about to vary coupling, but could probably have done about as well winding both coils on the same core with about a 1/8 to 1/4 inch separation between them. Note: I found that the loading coil had to be placed at the bottom end of L1 as shown; placing it at the open end didn't work very well at all (for me). For a slider, I just straightened out a large paper clip, soldered it with a stranded hookup wire into a soldering lug, and then used a wood screw to hold the slider to the base I used. Just for grins I rolled my own fixed capacitor, C1, making it about 300pF.

On the air results were pretty good. The detector tap is selected by an alligator clip. I got the separation I wanted between the two strong locals that are 50 kHz apart, sensitivity is ok, BUT, the rig design still needs some work before I offer it to my students. Current problems and possible solutions: a). Only using about half the coil, and it tunes pretty fast - use larger wire, maybe a smaller diameter coil form, and maybe a smaller value capacitor. (probably do at least two of the three)

b). Not happy with the paper clip slider; too flexible so you get backlash, and doesn't ride smoothly - use a metal strip and maybe add another washer where it connects to the base.

c). With detector tapped directly to the coil, you always hear something, whether or not the slider is making contact. I guess I can live with this.

Lyonodyne Version 17 Crystal Set

from Mike Tuggle

http://www.crystalradio.net/crystalsets/lyonodyne/index.shtml

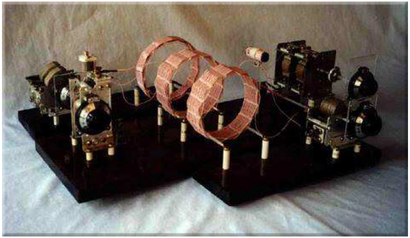

Builder and Operator Mike Tuggle

The "Lyonodyne-17" is an advanced DX crystal set evolving over the years. It was a first place winner of the DX "Open Class"contest two years in a row in 2000 and 2001. Mike placed second in 2003 after handicapping himself by using a rock (lead telluride) as the detectors. The radio incorporates the "Tuggle" front end (which will be shown later), and very high quality parts.

Note:

This is not intended as a "How To" article for a beginner. The following is supplied as information only. Some of the techniques used in the construction of this radio are very advanced and not recommended for a beginner or even an intermediate builder.

My own experiences over the years have led me to shed my skepticism of the incredible DX reports made by old-timers back in the early days. Component technology was not what it is today, but circuit technology was, and the sparsely populated bands back then had to be a lot more DX-friendly. You may want to pull an antique crystal set off the shelf and give it a spin. I'd encourage it.

Some of these old sets are extremely well built. The construction of the military BC-14A/SCR-65 never ceases to amaze me--and others were built just as well. But to start out, I'd recommend conceding to modern technology by using a good, modern diode and sound-powered phones with matching transformer. With just a little luck, prepare to be amazed all over again!

A CRYSTAL RADIO WAVE TRAP

Owen Poole

http://bellsouthpwp2.net/w/u/wuggy/traps.htm

This is a little circuit that is a real giant killer when you are trying to nose up to a loud local to look for the little guys. Basically, it is a tuned circuit placed in the antenna path to your rig, which traps and then dissipates the signal, or at least a bunch of it, from a station, allowing you to reduce interference between stations. There is nothing fancy to it, but a little care in construction and operation can make it really work well. Here is one of mine:

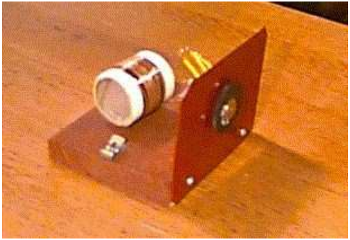

I used a 1 1/4 inch diameter cardboard core, and wrapped 100 turns of #32 enameled wire around it, closely spaced, and attached the ends of the coil to a 200 pF variable capacitor. This forms the tank circuit. I also wound over the first coil a second one, using 15 turns of #22 enameled wire, and ran each end to fahnstock clips - the antenna is connected to one clip, and the second goes to the antenna tuner, the set, or what have you. To use, just tune your receiver to the offending station, and then adjust the trap's capacitor until it reduces the signal. About 15 seconds is all the time you need to become a pro. The #32 wire, being of relatively high resistance, dissipates the signal that makes the tank circuit resonate (real engineers have assured me this is a bunch of baloney, and they are probably right, but my little trap works fine for all that - smart money says to make it of larger wire). . The #22 wire transfers the signal inductively from the antenna circuit while offering a low resistance path for other frequencies. It tunes fairly sharply, so tune slowly. You could install a bypass switch for times when the trap isn't needed, but I just tune to trap out of the way, with no signal loss to that I can notice. If you have more than one station you want to knock out at the same time, put more traps in series with this one. I have two which I use to cut out the two strong signals from two stations dominating the top of the band only 50 kHz apart. They are still too close together to listen between them, but I can certainly get much closer above and below them now. I might try a few more turns on the smaller coil to see if it sharpens up the signal a little. This was a first attempt, and I haven't tried to optimize it. Had I wound the coil for about 400 uH, I could have gotten full band coverage - actually, in my QTH, there are no loud ones below about 690 kHz, so this setup works fine for me with the 200 pF capacitor.

While you are winding the 100 turn coil, go ahead and put a tap in it, say at 30 turns or so. Then you can use the trap for a crystal set as well by adding a detector and earphones across the 30 turns. This lets you do at least two things: 1. It is a nice little crystal set all by itself. 2. You can operate it as a

closest some 2400 miles away? As it turns out, Hawaii is an ideal DX location--for distance if not for sheer numbers of stations. With no regional stations, sunrise-sunset activity is non-existent.

West coast 500-watt stations, 2500 miles distant, have been heard here. High-power stations in Cuba (4800 mi.) and the Caicos Islands (5500 mi.) have also been heard, thanks to a mostly water path. Stations from the "interior" also make it over. Last year, it was neat to hear KRVN in snowbound Nebraska (3600 mi.) using a homemade cat's whisker mineral detector. This must be what it was like in the old days.

With low station powers and crowding, medium-wave broadcast band DXing represents the greatest challenge. On short wave there's no limit to the distance--reception is truly worldwide. I occasionally hear the South African broadcasts nearly 12,000 miles away; Johannesburg and Hawaii are nearly at antipodes. From the mainland, Australia (11,000 miles) was a routine catch. Figures 5 and 6 show my "12,000-mile" crystal set.

Crystal Set DX Activities

Crystal set DX activities include discussion forums sharing ideas and results, an annual DX listening contest now into its sixth year, and the occasional set-building contest where some remarkably high-caliber craftsmanship comes to the fore. All of these activities are served up on the Internet.

The following site is a highly recommended grand portal to the wonderful world of crystal set DXing: Owen Pool's Crystal Radio Resources at http://www.thebest.net/wuggy. The comprehensive set of links here covers all aspects of the hobby. This site is headquarters for the annual crystal set DX (XSDX) contests open to everyone. The contest usually takes place in late January. Watch for announcement of specific dates and rules.

antenna. The rule of thumb is, if you can hear them on a radio, you can hear them on a crystal set.

The best times to listen are at sunrise and sunset, when stations are signing on or off, raising or dropping their powers, and changing their antenna patterns. These circumstances make for a jumble of regional stations ripe for the picking. Deep night is usually the best time for flat-out DX.

In 15-year stints at two locations in Maryland, I accumulated logs of over 600 and 400 stations, respectively. The most distant receptions, at 1800 to 2200 miles, were a few powerful stations in the Caribbean area and adjacent South America. I was aided, no doubt, by the largely water path between them and me. The farthest overland station was in Denver at about 1500 miles.

When I moved to Hawaii, I wasn't sure what to expect. The local Honolulu stations were givens. But what about the outer islands? And all-importantly the next stations, mainlanders, the

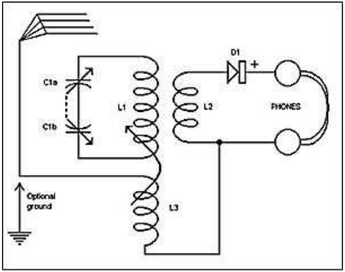

Fig. 6. Schematic of the '12,000 mile-er.'

---

crystal set alone, listen to it while you adjust the trimmers on the capacitor for best frequency coverage - particularly handy if you use a different size wire, different number of turns, different capacitor or different diameter core. Don't leave out the smaller coil of #22, or whatever you use; I tried this trap without the small coil, and it literally wiped out the whole BC band, no matter what frequency it was tuned to - on the other hand, if this is built to be resonant at the shortwave broadcast freqs, it cuts out a larger portion of the offending band.

More fun; put it in the ground path instead, and it works about the same. Put it before or after the antenna tuner (if you use one) - works about the same. Use it as a crystal set along with the other set, and you can listen to two stations at the same time - no, you can't listen to the same station with both sets if they are in series - the one rig cuts out the other. You can even use a trap to provide power to a small transistor amplifier stage if the trapped signal is large enough.

On the air testing with a dc millivoltmeter across the 47k resistor in my detector circuit (see my testing page) shows a signal reduction on "trapped" stationsof about 6 to 14 dB with the trap. I suppose if you have a real monster in your neighborhood, you can put two traps on it for a double stomp;
I couldn't see much advantage to the second trap when I tried this. I have not found the need to use a trap bypass switch when the trap isn't in use; I just shove the tuning cap to the other end of the band, and don't notice a reduction in set sensitivity elsewhere. I also notice that on really strong locals, the difference in signal level of 6 dB or so isn't very noticeable, and tuning the trap "by ear" is a bit difficult. When I tune off the offending station a bit, however, I can see the

effect of the trap more readily. I made mine as an outboard unit, but you can just as easily incorporate into your main set, just don't make the coils of the two circuits in line (on the same long axis) with each other or you will mutually couple them, which makes for confusing tuning.

One of the unexpected benefits of a trap, as I found out, is that it can enhance the operation of an otherwise sensitive but not so selective crystal set. Using a trap with slightly modified Radio Shack crystal set, I heard a brand new station coming in from several hundred miles away that was ordinarily masked by a stronger station.

Trapping with the pros: Probably the most effective trap is an inductively coupled one. Here's how to make and use it: Make a coil of the same dimensions as your set's main tuning coil, and parallel it with a variable capacitor, also like the one your set uses. Now, place the new trap coil in line (on the same axis) with the main tuning coil, and, with the set tuned to an offending station, tune the trap capacitor to null the station. You can vary the depth of the null by moving the two coils apart. If your set has a separate antenna tuning circuit, the trap coil should be on the side of the set tuning coil away from the antenna coil. I have one of these and have gotten over 30 dB of fairly sharp signal rejection with it. Be advised; when using an inductively coupled trap, expect some interaction with the tuning coil. Using an inductive trap can give some interesting results you don't get from one placed in the antenna line.

More trap fun: If you have a set using in-line antenna and main tuning coils, inductively coupled to each other, here is

These phones are low impedance and must be matched to the high-impedance tank (L2-C2) and detector diode by an audio transformer having, typically, 50- to 600-ohm and 50- or 100-kilohm windings (Figure 4). The quality of the transformer is very important, so that its insertion loss is small. UTC input transformers have a good reputation for low loss. The final touch is a comfortable set of headphone cushions for good acoustical coupling to the ears and exclusion of outside noise.

DX Experiences

Under favorable conditions, medium wave or broadcast band

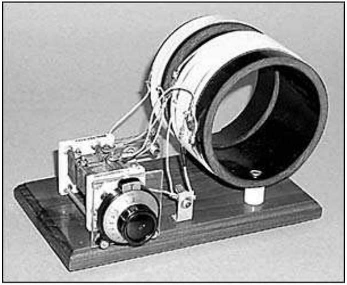

Fig. 5. A '12,000-mile' short-wave receiver.

DX crystal sets can receive hundreds of stations--some of them thousands of miles distant. In fact, DX crystal set performance is comparable to any other radio (powered or not) short of a full-blown communications receiver with its own outside capability to put up one of these 'mega' antennas. So it behooves us to heed principles 1 and 2.

The front-end (antenna tuner) design shown here is one of several that can be used. This particular design tunes a wide variety of antenna-ground systems. Other DX crystal sets use a simple series tuning circuit effectively--especially on long antennas.

The diode detector provokes more mystery, controversy and debate than any other component. Some folks favor low-resistance germanium diodes like the 1N34A or rock stands with galenas. Others swear by high-tech Schottky diodes. The emerging truth is, there is no universally perfect diode.

The 'best' diode depends specifically upon the set it is used in. Per the second principle, the diode needs to match the tank circuit feeding it and the transformer and phones that it feeds. With some care, a high-Q tank with litz wire coil and a modern, military-grade variable capacitor can attain a resistance of from several hundred kilohms up to a megohm. Only Schottky diodes and a few modern germanium diodes have resistances this high. The old catswhisker-rock stand detectors have far lower resistances.

The most practical approach to selecting the right diode is to apply an A-B listening test to a number of them. Two diodes are mounted in a test stand arranged to quickly switch between them. Using a fairly weak station, one can test a pile of diodes, pair-wise, keeping the winner after each test, until the ultimate one is found. But it's most important to realize that this diode is 'best' for the particular set it was tested in. It may be a quite poor performer in a different set.

Surplus, sound-powered (more properly, balanced-armature) phones have become the industry standard for DX crystal sets. Baldwin Type C's were an early example. Now, post-WWII surplus units made by RCA and US Instruments Corp. (USI) are preferred.

something else to do with your inductively coupled trap. When trying to dig out a weak station, get as close to it as you can, then tune the trap for a null. Now, take the trap coil and place it between the antenna and main tuning coils. With some careful tuning, you can often isolate the weak station. The trap is giving you an additional stage of filtering and selectivity, and you are now using "triple" tuning. In this configuration, the trap is now functioning as a transfer circuit. You will notice that best performance is with the coils separated from each other by a coil diameter or so.

Larry Pizzella also reports good results with in-line traps using largish coils, with the tuned circuit portion of the trap inductively coupled to a separate coil connected to the antenna line which can be moved away from the main trap coil to provide variable coupling. If you look at his big radio, you can see a couple of his outboard traps as well as his inductively coupled trap (the large blue coil on the far right).

I would like to report that trapping my two strong locals opened up the top of the band for me; it didn't completely, but I can now get stations as close as about 20 kHz.

I have been able to successfully trap the HF ghosts somewhat that bedevil me early in the evening, but feel the need to work on this some more. Multiple hf intruders are really tough, but the hf trap I made doesn't use a coupling coil, and is rather broad, so gets a largish portion of the lowest shortwave broadcast band, and supresses a couple of stations sometimes.

We have, from left to right: antenna-ground system, front-end (antenna) tuner, secondary (detector) tuner, diode detector, audio matching and finally, phones. Maximum transfer of available signal power from one stage to the next happens when the impedances of these stages are matched to each other--from antenna to phones--and indeed, from phones to one's ears.

Components

A good outside antenna-ground system is essential for DXing. Loop antennas do not have enough pickup to be effective. An inverted L longwire 20 to 30 feet high and 50 feet long is a good start. The one truth to antenna design is, "higher is better." Any effort expended to raise the antenna, even a few feet, will be amply rewarded. In the most recent crystal set DX contest a third design principle emerged: A huge antenna can overcome shortcomings in the first two principles.

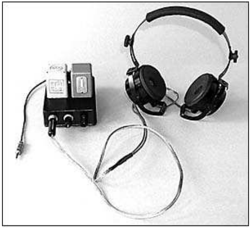

Fig. 4. RCA balanced-armature phones, with matching transformer.

The winner suspended 140 feet of litz wire near vertically using helium-filled balloons. Another high-scorer just happened to have a 140-foot tall base-insulated tower and four 1000-foot Beverage antennas to complement his junk box set. Most of us have neither the real estate nor the rigging

Design Principles

Two fundamental principles underlie DX crystal set design and construction: use of low-loss components and proper impedance matching between stages.

Avoid the temptation to construct the set with vintage components. The result may be a handsome set that is only a fair performer. I have such sets, and they sit up on the shelf and look nice. However, for ultimate performance, one should rely on top quality (usually modern) components and materials.

This is especially important for tuning capacitors, coil wire and forms, and detector diodes. Layout and construction, particularly in the RF-carrying sections (antenna through detector), should follow good HF practices: short direct leads, careful insulation of components, and avoidance of switches, taps and other trappings in the 'hot', RF-carrying sections.

The crystal set, like any other radio, consists of a series of stages--each with a function, each coupled to the next. The stages of a simplified crystal receiver are diagrammed in Figure 3. The resistances shown are used for detailed circuit analysis that will not discussed here.

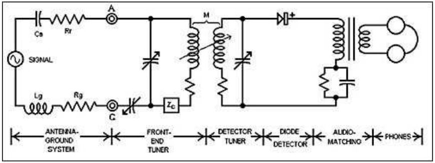

Fig. 3. The stages of a typical DX crystal set (see text).

BANDSPREADING TECHNIQUES

(20 June 99)

Owen Poole

http://bellsouthpwp2.net/w/u/wuggy/spread.htm

Bandspread in a crystal set? You must be nuts! Well, maybe so, but let's give it a shot anyway. Bandspreading is nothing more than making it possible for you to tune your set more slowly. Why bother? Well, if you are in a crowded band, such as in a metropolitan area with lots of strong stations, you have probably already gone for more sophisticated crystal radio designs to allow you to separate the signals. Even in my location, with two peanut whistles only 50 kHz apart, sometimes I still have to tune very carefully when my rig is set up to be selective so that I can get the station I want, and not hear the other. In the HF (shortwave) region, with stations fading out and then roaring back with regularity, being able to tune slowly allows me to catch stations that are just out of the mud, but are more often buried by the stronger stations that usually dominate, and which catch my attention easily. With HF ham rigs, the ability to tune slowly is an absolute necessity; otherwise you sweep right past stations that are only in your selectivity skirt for a few hundred Hz. Most crystal sets use a tuning capacitor that has only a180 degree rotation in going from around 10 to 360 pF or so - on the BC band, this gets you through the 1.2 MHz-wide band rather quickly - on HF, you cover several MHz, and a quick sweep through the frequencies can cause you to miss stations if you are only expecting what you normally hear.

There are two basic types of bandspreading: Mechanical and Electrical

Mechanical: The simplest way to slow your rate of tuning down, or at least let you tune more sharply, is to use a larger tuning knob; your ability to control your tuning rate varies directly with the diameter of the knob, assuming you grab it on the outside edge. Another method that used to be popular was to have a small diameter shaft turning a larger diameter one connected to it by a dial cord; the cord was usually kept tight by a small spring. A third way, still much in use, is to have a planetary dial drive which makes several rotations to drive the capacitor through one half a rotation. Dial cord drives are tricky, and planetary drives can cost more than the rest of the set, so I usually go for the largest knob I can find. Another thing to watch out for with mechanical bandspread drives, should you try one, is backlash or slop - you reverse your direction, and lose a couple of degrees of turn before the capacitor starts to move. For inductive tuning, usually consisting of moving a ferrite core in and out of the coil, the once readily available AM tuning coils used to come with a metal cap through which a screw protruded that allowed you to vary inductance very slowly using a small knob to twist the core in an out. Compression capacitors also use a screw to let you make several turns from stop to stop, but aren't ordinarily used for main tuning, and are usually for screwdriver use only. If you do stick a knob on a compression cap with some epoxy, a soda bottle cap should work, and will allow you about 3 or more revolutions of rotation stop to stop.

Electrical: Nothing magic about this either. Here you use two or more capacitors to change the amount of capacitance a turn of the knob gives you. Shown below are some of the more common ways to electrically bandspread your tank circuit:

So, what makes a crystal radio a "DX" set? Well, this one is double-tuned (L1-C1 and L2-C2) for selectivity to tune weak DX stations in the RF jungle we now live in. Some DX sets resort to triple tuning for even more selectivity -- depends on how many hands you have to do the tuning. The wavetraps (L3-C3 and L4-C4) can be tuned to reject strong unwanted stations in the manner of the notch filter on a communications receiver.

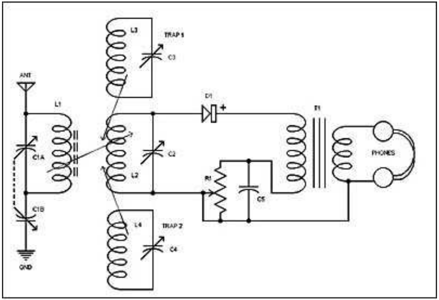

Fig. 2. Schematic of the 'Lyonodyne-17.'

Coils L2 through L4 are basket-wound litz wire for low loss and high Q. Variable capacitors C1 through C4 are highest-quality, silver-plated, ceramic-insulated units--as perhaps only the military could specify. All these components are isolated from the mounting board by ceramic standoff insulators. Overkill maybe, but why take chances?

particularly interested in crystal sets since 1959, when I first discovered you could actually DX with them.

The Internet has aided crystal set DX activity by making exchange of ideas between like-minded enthusiasts much, much easier. This article only touches the surface from a personal perspective. The reader is encouraged to pursue the online resources, described below, covering all aspects and providing further examples of this fascinating hobby.

In this article I'll cover general design considerations for building DX crystal sets, but will leave the actual construction specifications up to you.

Equipment

Figures 1 and 2 show an example of a DX crystal set. I call it the "Lyonodyne-17." The design evolved from an earlier version described in the November, 1978 OTB. As you can see, this is not your grandfather's crystal set.

Antenna (right) and detector (left) tuners are mounted on the middle board. Wave traps are located fore and aft on separate boards, making it possible to adjust the coupling by moving the boards. The antenna coil (L1) is wound on a short ferrite rod. The matching transformer unit for my RCA "Big Cans" sound powered phones is at front, left.

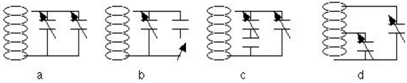

Figure a. is one of the favorites. If one capacitor is significantly larger than the other, it is called the band set capacitor, and the smaller one is the bandspread capacitor. Capacitor size ratios of from 2:1 up to 10:1 should be used. Keep in mind with all of these that minimum circuit capacitance is the sum of the minimum of both capacitors. Since the polyfilm caps I usually use, because they are cheap, only run to about 200 pF, I sometimes just put 2 in parallel to get full AM band coverage (and 360 degrees of rotation at the same time). Incidentally, since these capacitors are really two sections, a 150 pF and 50 pF, you can get full band coverage with a 200 pF capacitor, with both sections tied together, and only the 150 pF section on the other; use the 150 pF section for the top part of the band and it will slow your tuning rate down a bit.

Figure b. is sort of a poor man's bandspread, switching in one or more fixed capacitors to increase capacitance in increments, then adjusting the variable as needed. I used this with my peanut special to get to the bottom of band. The fixed capacitor should be no larger than the range of the variable to prevent skipping over frequencies.

Figure c. has a fixed capacitor in series with one of the variables. This reduces the tuning range of the series variable from about the minimum of the variable to the value of the larger capacitor in the series branch. Good to use if you don't have any small value variables.

Figure d. taps one variable down on the coil, reducing its effect on tuning range. In this case, assuming it is the same value as the band set capacitor for example, its effect capacitance-wise over its full range is approximately equal to the percent of the coil it parallels. If you use a switch to select different coil taps, you can thus vary your bandspread.

A note on using fixed capacitors: I have found that some are better than others, and a lossy one can reduce set sensitivity when the fixed cap is switched in. I have seen this with the readily available ceramic disc capacitors. This may not happen to you, but just be aware of the possibility. You might want to try different types here. Try to use good quality switches as well, and make good solder connections....

Finally, to use the K.I.S.S. principle, use a capacitor with a range no bigger than you need. That is, use only enough range to cover that part of the spectrum in which you are interested. Using a single 365 pF capacitor on the HF bands, one turn of the dial will sweep through several MHz, and no matter how selective your crystal set, you will be sweeping through a lot of dead frequencies and then sweat tuning in and then separating the stations you want to hear.

You can pretty much do with inductors (coils), what you can do with capacitors, and some of the old designs used a lot of variable inductances. Putting inductors in parallel has the

---

DESIGNING A DX CRYSTAL SET

by Mike Tuggle
46-469 Kuneki St.
Kaneohe, HI 96744-3536
E-Mail: tuggle@apacs.com
http://www.antiquewireless.org/otb/dxxtal.htm

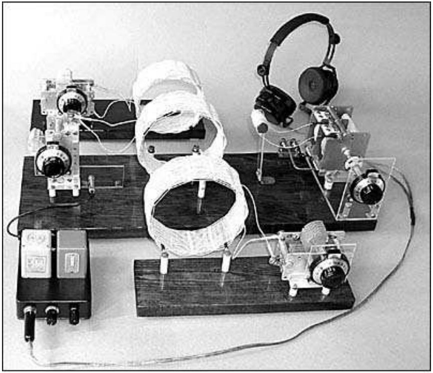

Fig. 1. The 'Lyonodyne-17,' an example of a well-designed DX crystal set.

Long distance reception with crystal radios is once again becoming a serious pursuit among hams and other radio hobbyists. They are discovering that the DX capabilities of these receivers have been greatly underrated. I've been

T1 Stanley audio transformer Part Number 227-E-1-3-98
Pri 100 ohm Sec. 100M (100k) ohm
(Fair Radio Sales number T3/AM-20)
T2 Bogen T725 4 watt P/A transformer (Check the bottom of this page)
R1 250k ohm variable resistor #
R2 2 to 3 Meg ohm variable resistor
SW1 12 position single pole rotary switch (Radio Shack)
SW2 12 position single pole rotary switch (Radio Shack)
SW3 6 position single pole rotary switch (Radio Shack)
SW4 DPST Toggle switch
SW5 SPST Toggle switch
C1 .1 uf capacitor
C2 4.7 uf capacitor non polar (Radio Shack)
C3 2.2 uf capacitor non polar (Radio Shack)
C4 1.0 uf capacitor non polar (Radio Shack)
C5 .5 uf capacitor non polar (Radio Shack)
C6 .05 uf capacitor non polar - (Radio Shack)
C7 .1 uF Capacitor #

Notes: # Not used in UltiMatch, only in UltiMatch 2

same effect as putting capacitors in series, and vice versa for series inductors. My uncle Elmer (I must have had one sometime) told me, however, that circuit Q was highest when you used as much inductance as your frequency range allowed, and to let the capacitor do all the "heavy lifting". Still, some of those old variable inductors look pretty elegant, and I am sure they work pretty well, even if they are a bit more complex to build.

switching the absolute phasing, don't worry about it. Many cannot and it's one less thing for you to concern yourself.
Position 1 on SW1 permits use of an efficient 4, 8, or 16^ speaker on very loud signals, or use of headsets with very low impedance ratings similar to those speakers.

Output Using Input2/Output Combo
The 'High' side terminal of input 2 may be used as the 'return' for the headset for the xtal detector function if it is desired to select a winding ratio 'inside' the normal selections. Connect the headset to the 'high' side output terminal (from SW3) and it's other lead to the 'high' side of input 2. SW1 is then used to select a tap for the 'low' side of the headset connection, SW2 selects the 'high' side as before. This 'inside' option excludes selection/use of the position '1' 8^ speaker tap.Some headsets will match up better using this option.

Absolute phasing may then be reversed/corrected by reversing the switching order of SW1 and SW2 , thus reversing the 'high' and 'low' sides of the headphone conections without need to disconnect and swap them, once the best ratio selection is found by ear. Correct "Absolute phase" refers to the the condition where a pressure 'push' into the studio mic gives a pressure 'push' from the speaker or headphone. Absolute phase reversal results in an unnatural result of a 'push' giving a 'pull'. Some listeners find clearer & more lifelike voice sounds when absolute phasing is correct.

Steve Bringhurst

Parts List
(UltiMatch and UltiMatch 2)

to the headphone being used. The position giving the loudest & clearest sound is usually the best setting. SW3 is used to select bass response balance for the headset and input device in use, the 1st position is 'direct'. This adjustment will have a more subtle effect than a regular tone control.

Input 2
Input 2 is used for powered devices like tube & transistor detectors & amplifier ckts with SW4 in the 'low' position. SW1 is used to select the proper input match for the powered device. Always start from the highest setting first when using this function, too low a setting may damage the powered device. Using SW2 selects the best match for the headset as before, or position 1 can drive a loudspeaker directly. The loudest & clearest sound is usually obtained when SW1 & SW2 are optimum.

Output
Any set of headphones or single earpiece, such as a telephone earpiece, may be connected to the 'output' terminals. It is usually preferable to have 2 earpieces connected in series, rather than parallel, & properly phased so that the sounds seem to occur 'inside' ones head between the ears and not 'at' either ear. Switching the connections to one earpiece will correct phasing. Some listeners can hear a definite difference in intelligibility when the headphone connections to the transformer device are switched around (which reverses 'absolute phase'), the clearer sound has correct 'absolute phase' and is of course preferable. It is generally a one-time adjustment, but not always for every station, program, ect . This can change if the station or studio is careless about engineering practices and allows reversed absolute-phase signals to occur. If you cannot hear a difference when

SELECTIVITY ENHANCEMENT CIRCUITS

http://makearadio.com/misc-stuff/dxnotes.php
Dave Schmarder’s Hints and tips…

I would like to talk about selectivity enhancement circuits (SEC). Some refer to these as the "Hobbydyne". Jim Frederick in Florida came up with the Hobbydyne circuit. The earliest publication of a SEC I have seen was in Australia in the early thirties. Jim Frederick's Site Here is the Australian page. So you see that what appears to be new, sometimes isn't. But Jim's Hobbydyne was new. Also his first (to my knowledge) recent use of the SEC is credited to Jim.

The SEC and Hobbydyne do the same thing. They allow unloading of the tank circuit, thus increasing the loaded Q, which means better selectivity. It also provides a better match between the tank and diode (and audio circuits further down the line). This is better than the old fashioned way of matching and unloading - tapping the coil. The basic SEC consists of a small value variable capacitor (usually around 20-30 pf), and a small RF choke. I found that a 27 millihenry value is good. Some have used a reverse biased diode in place of the choke. The choke or diode is to complete a DC path so the detector can work properly.

Jim discovered this feature while experimenting with his crystal sets in 2003. He sent me an ie-mail and invited me to test this circuit. I tested it on my, at that time current dx receiver, my #35 radio. It didn't take me very long to see the value in this discovery. I knew that this was going to be a regular part of my dx crystal radio designs, so I looked around for some inexpensive chokes. I found some 27 mh chokes and found they worked great. The next development from Jim was his Hobbydyne circuit. Using the SEC described above had the issue of the resonant frequency shifting when the selectivity was changed. It is easy to understand as the capacitance was being changed and this affected the tuning. Jim's solution was to add a differential capacitor. A differential capacitor is a variable capacitor with one rotor and two stators. As the capacitance on one side increases, the other side decreases. This part was used mostly in the high end variable oscillators for better stability.

Jim later made an enhancement to his basic hobbydyne by adding a small trimmer on the ground leg of the differential capacitor. This allowed for better tracking across the differential capacitor range. I found there are two Hobbydyne hookups. One is what Jim came up with and the other is what I came up with. I don't claim any discovery here, it was most likely a wiring error. But, isn't that how some everyday items were discovered? Below are some circuits showing various SEC configurations. Figure (a. is a directly connected diode to the top of the tank. This is the worst possible way to directly connect a diode for most builders. Only if a high impedance load is used at the output, will this method work well. Otherwise, you can tap the coil and connect the diode to that tap. That way you won't end up with a bad mismatch and severe tank loading. This causes loss of sensitivity and selectivity. Figure (b. is a basic SEC. Just a small trimmer capacitor and a dc path rf choke. This is a good place to start, especially if you don't have a differential capacitor. The rest of the figures show various SEC and the Hobbydyne connections. It is kind of a take-your-pick situation. Figure (e. has the

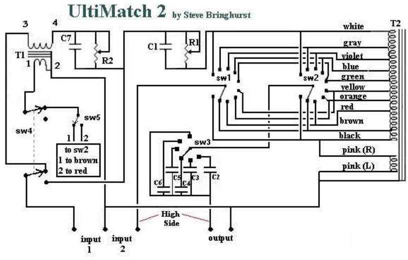

Directions for use of Ulti Match 2

Input 1
General directions for using the UltiMatch with crystal detector receivers; Crystal detector input is made at input 1 with the diode to SW4 & the detector return to the other terminal. 1st position of SW-4 is for higher impedance detectors and the 2nd position is for lower impedance detector ckts & mineral detectors. SW5 selects an over-all ratio shift to better match some low impedance headsets at pos#2, pos#1 is 'normal'. SW5 doesn't function with SW4 in 'low'. R1 is used to adjust for lowest strong-signal distortion with lower impedance detectors, R2 is adjusted for lowest strong-signal distortion with very high impedance detectors like Schottky barrier diodes, otherwise, it's 'normal' setting is 'minimum' with the adjustments for 'low' SW4 setting and germanium and similar detectors on SW4 'high' being made with R1. SW2 allows selection of the best match, by ear, of the input device

UltiMatch and UltiMatch 2(new)
by Steve Bringhurst

This is a "All in one box" unit
This incorporates the Select to Match ("S-T-M") and the Universal Match-Box ("Uni-Match")
all in one unit. Get the most out of your components!!

Original UltiMatchCircuit
Original circuit has been left here so those who have already built the circuit can see easily the differences in the two circuits.
This will help if you would like to "convert" to the new circuit.

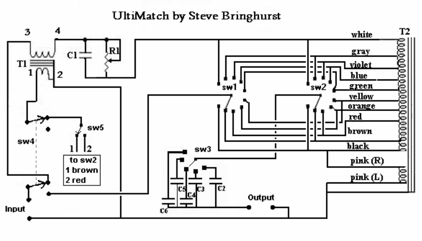
UltiMatch by Steve Bringhurst
3
4
T1
1
2
C1
R1
sw4
sw5
1
2
to sw2
1 brown
2 red
Input
sw1
sw2
sw3
C6
C5
C4
C3
C2
Output
T2
white
gray
violet
blue
green
yellow
orange
red
brown
black
pink (R)
pink (L)

UltiMatch 2(new)

---

additional tracking capacitor. That is recommended if you can spare the capacitor. Another hookup would be for you to use two small capacitors in a quasi differential capacitor. You can manually increase one capacitor while decreasing the value of the other one. This will give you that capacitive divider effect.

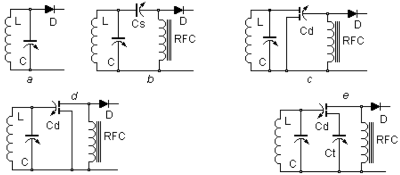
L
D
C
a
Cs
RFC
b
Cd
c
d
e
Ct

L= Main Coil, C= Main Tuning Cap, CS Selectivity Variable (20 pf), D= 1N34A etc.
Cd= Differential Cap (20 pf) Ct= Tracking Cap (10 pf), RFC= 27 mh Choke

Crystal Radio Selectivity Enhancement Circuits (c) 2006, D. Schmarder

T2 Bogen T725 4 watt P/A transformer (Check the bottom of this page)
sw1 12 position single pole rotary switch (Radio Shack)
sw2 12 position single pole rotary switch (Radio Shack)

Universal Match-Box
"Uni-Match"
by Steve Bringhurst

Another idea from Steve Bringhurst is the "Uni-Match" or Universal Match-Box. This simple box will match just about any
type headphones to just about any type radio. Not designed just for crystal radios, but for just about any type radio.
"Say you want to use your S/Ps or Brush-Clevite piezo's, or your Telephonics TDH-39s or?? on a tube or xistor regenny, MK-484 TRF, the output of your regular radio, ect, ect. You could do any and all with this." Steve Bringhurst

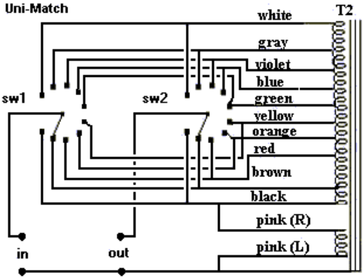

Starting at the first position on the switches and going clockwise the impedance should be approximately
40k, 20K, 10k, 5k, 2.5k. 1.2k, 600, 300, 150,75, 8 (all in ohms)

CONTRA WOUND COILS

http://makearadio.com/misc-stuff/dxnotes.php
Dave Schmarder’s Hints and tips…

Hi Friends. Welcome to my contra coil page. The coils that I am showing on this page are used as the main tank coils for crystal radios (and tube radios too). I believe that this improved type of tank coil will be one that you will want to incorporate in your next crystal set. I owe the ideas on this page to Ben Tongue. A while ago he published this very interesting article on his web site. The article describes a radio he built that has a constant receiving bandwidth across the MW broadcast band. Reading this article is very worthwhile.

My goal isn't to have constant bandwidth, but improved performance where it is really needed, at the top end of the band. There are several difficulties in tuning the high end of the band and I believe the contra coil will improve your dx reception.

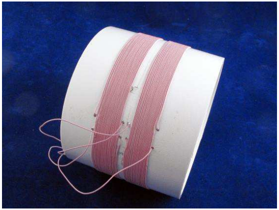

The contra coil in its simplest form is two equal coils wound on the same coil form, but in opposite directions. The coils are connected in series to tune the low end of the band and in parallel for the high end. How the coil is wound and connected is where the secret lies. With the windings wound as they are, the losses are low and the Q is high. The variable capacitor also operates in the sweet spot, further cutting losses. On the high end of the band, the litz strands are doubled with the parallel connection. That's got to be good!

Ben's coils have extra taps for the diode connection. The tap used is dependant on the part of the band that is tuned, along with connecting the coils in series or parallel. He splits the band into 4 segments. For now, I'm leaving off the taps and going with a hobbydyne type circuit that will allow for variable tank loading and matching.

The first coil shown is my test prototype cylinder coil. I felt I should build one cylinder type to get the feel of how this coil will operate. The coil form is a styrene sewer pipe coupler. I bought it at Home Depot. The outside diameter is 4-1/2 inches (11,5 cm). There are 22 turns of 165/46 litz wire on each winding. Each coil calculates out to 27 feet (8,25 m), including 6 inch (15 cm) wire leads. The start (**s**) windings are in the center, and the finish (**f**) windings are towards the form edges. When winding the coil, start at the center. Wind the first coil. Then when winding the second coil, beginning again at the center. The two start windings will wind in the same direction around the form. That makes the coil wound reverse, or contra wound around the form.

To connect the coils in series connect the start of one coil to the finish of the other coil. The other leads are connected to the

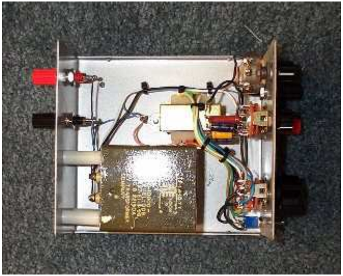

put in for 10k Hi "Z" headphones (Radio Shack)

Notes:
1) The range of C1 can be from .05uf to .22uf. It is not that critical.
2) Only ten of the twelve positions are used on SW1
4) #1 lead on T1 goes to SW3. It is split and one lead goes to brown on SW1 and the
other goes to red on SW1. This allows you to shifts the ratio range a full
step for better use of the Stromberg Carlson elements.
5) Experimentation is the key to getting the right values on this unit.
What is right for you and your set up is what counts.
6) See photo below for proper phase ID of the two "pink" wires on the Bogen 725
transformer.

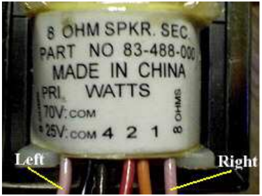

My version of Steve's S-T-M Stanley/Bogen

rest of the circuit. For parallel operation, the two start wires are connected together as the two finish wire are connected together. The connections to the rest of the radio are from these points.

When the coils are in series, the inductances add (plus the mutual coupling). This is because the coils are in phase. It is just that the wiring connections jump to different positions on the forms. When the coils are in parallel, they are also still in phase, but at 1/4 the original inductance value.

The inductance is approximately 240 µh in series and 60 µh in parallel.

It is likely that you will want to use some kind of switch to go from series to parallel. Try to find a low loss switch, such as a ceramic rotary switch, or use thumb nuts, and brass links mounted on low loss materials. The circuit is shown below. It is important that the coil and switch be wired exactly as shown. If your radio doesn't work, check the wiring first. I included the physical wiring pictoral below. This is from my #64 contra radio. The picture of the switch is the top view, while the pictoral is how it is wired from the bottom.

Design Your Own Contra Coil

Now how do you design your own contra coil? Here is the place to start. You may want to skip this part for now and look below at the pre-designed coils. If your tuning capacitor matches one of the situations below, you don't need this section.

You will need a few things before you start. First, bookmark this page on crystalradio.net. This link sends you to Dan Petersen's *Professor Coyle* calculator. This takes nearly all the math out of designing your own coil. You want to select the cylinder coil calculator as this one has the resonance calculator.

If you build a coil, you should have an L/C meter. I use the one produced by AADE. It is important to be able to balance the inductances of the two coils. If you don't have one, you can just shoot for the best by using the designs that I have made.

If you want to test the actual tuning ranges, you will need an accurate signal generator, a capacitance meter and a scope or other rf level indicator. If you are building one coil, you can just put it in your crystal set and check out the tuning range by listening for the stations.

It is best now to go over some of the basic design criteria, map out what we have and discover a few truths and speculate on some assumptions. So here we go:

The contra coil has a 4:1 inductance ratio between series and parallel. 240:60 µh for example.

S-T-M Stanley/Bogen
"Select to Match" Impedance Matching Circuit
Circuit Design by Steve Bringhurst "baldy3823"

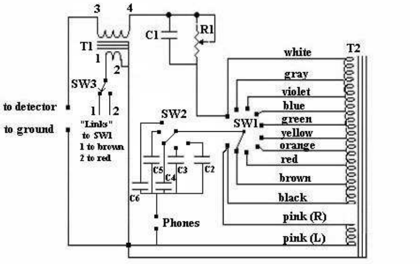

T1 Stanley audio transformer Part Number 227-E-1-3-98
Pri 100 ohm Sec. 100M (100k) ohm
(Fair Radio Sales number T3/AM-20)
T2 Bogen T725 4 watt P/A transformer (Check the bottom of this page)
R1 250k ohm variable resistor
SW1 12 position single pole rotary switch (Radio Shack)
SW2 6 position single pole rotary switch (Radio Shack)
SW3 SPDT toggle switch (Radio Shack)
C1 .1 uf capacitor (Radio Shack)
C2 4.7 uf capacitor non polar (Radio Shack)
C3 2.2 uf capacitor non polar (Radio Shack)
C4 1.0 uf capacitor non polar (Radio Shack)
C5 .5 uf capacitor non polar (Radio Shack)
C6 .05 uf capacitor non polar - not shown in circuit can be

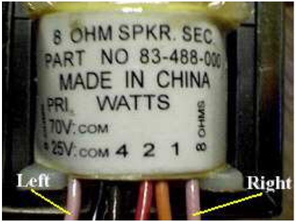

My version of Steve's S-T-M Calrad/Bogen

The S-T-M is excellent for comparison of headphones. It works very well for one or two sets of phones as well. It will run a range of headphones, from magnetic to sound powered.

The low end of the tuning range should be 530 khz. Better to design to 520 khz. The larger the value of your variable capacitor, the lower this is likely to be.

The high end of the series coil connection should reach above 1000 khz. This may not be possible. Just so the next condition is met, all is ok.

The low end of the parallel tuning should overlap the series high end by 30-50 khz.

The tuning ranges should span the dial over a total of 240-300 degrees.

A variable capacitor as low as 15-280 pF can be used but a higher value is recommended.

The variable capacitor shouldn't be over 500 pF.

Figure on about 25 pF capacitance added by the radio detector circuit and coil distributed capacitance.

An air trimmer capacitor of 75 pF is recommended. This helps with dial spread.

If you build the coil too large, you will lose dial spread but you will tune the whole band.

If you make the coil too small, you may not be able to get that 30-50 khz mid band overlap. (This assumes that you would have to increase the trimmer a lot more.)

The higher that the total maximum to the total minimum capacitance ratio is, the wider the tuning range will be. This means that the dial spread would decrease. Look at the pictures below in the *Other Adjustments* section

Being that the main tuning capacitor range and the fixed added values (diode circuit and coil distributed capacitance) is pre-determined, it comes down to juggling the coil value and the trimmer capacitor value to get the best spread with full tuning range.

R1 250k ohm variable resistor
SW1 12 position single pole rotary switch (Radio Shack)
C1 .1 uf capacitor
C2 Value of C2 depends on the set up of the headsets.
(all non polar)
  1) one 600 ohm impedance element = 1 uf
  2) two 600 ohm impedance. elements wired in parallel = 2.2uf
  3) two 600 ohm impedance elements wired in series = .5 uf
  4) 10k Hi "Z" headphones = .068uf to .05uf
  5) two low "Z" elements in series like Stromberg Carlsons can be run on 2.2 uf, but 4.7uf might be a better match up.

Notes:

1) 1 uF is a good fixed compromise value for C2 if several different impedance headsets will be frequently used.
2) The range of C1 can be from .05uf to .22uf. It is not that critical.
3) Only ten of the twelve positions are used on SW1
4) White lead on T1 goes to the rotary switch SW1 with yellow
5) Experimentation is the key to getting the right values on this unit. What is right for you and your set up is what counts.
6) See photo below for proper phase ID of the two "pink" wires on the Bogen 725 transformer.

S-T-M
"Select to Match"
by Steve Bringhurst

The "Select to Match" was designed to allow a person to experiment with different headphones on the same crystal radio.
If you are using only one set of headphones only, you might want to build the
dedicated circuit below which will save you space, time and money.
STM introduction

S-T-M Calrad/Bogen

"Select to Match" Impedance Matching Circuit
Circuit Design by Steve Bringhurst "baldy3823"

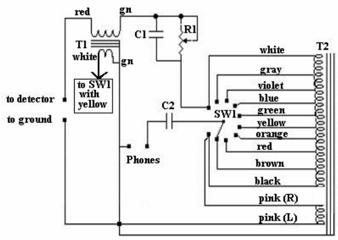

T1   Calrad 45-700 audio transformer (Ocean Electronic)
T2   Bogen T725 4 watt P/A transformer (Check the bottom of this page)

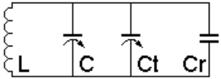

L= Coil Under Design - Series & Parallel
C= Main Tuning Capacitor Min & Max Value
Ct= Trimmer Capacitor 75 pf max
Cr= Radio Capacitances, 25 pf fixed

Crystal Radio Contra Coil Design

Here is an example: To start, you have to measure and add the capacitances for when the tuning cap is at minimum and maximum. Add 25 pF for the extra radio capacitances. Also include a starting value of 10 pF for the trimmer setting. The values are reached by rocking the values back and forth until a suitable value for the coil is found.

Let us assume a 15-350 variable capacitor plus 35 pF for the extra capacitances described above. You can see them in the schematic shown above too. This means that the circuit capacitance ranges from 50 to 385 pF. These are starting values, and the minimum and maximum values are likely to be more like 80 to 415 pF as the trials go on. The starting value for the inductor is 240 and 60µh.

Starting with the low end, plug in a value of 240 µH and 385 pF into Professor Coyle. That is a pretty close 524 khz. Lucky, hµH? Ok, now use the 240µh coil and plug in the minimum total capacitance. That comes to 1453 khz. That is too wide of a frequency spread. Instead of going further, we will change the maximum and minimum capacitor value and start over.

Now plug in a higher value of maximum capacitance and adjust the coil value to a 520-530 khz range. Let's try 415 pF. That drops it down to 504 khz with a 240µh coil. Let's reduce the value of the coil. 225µh brings the frequency in at 521 khz. Now let's try the minimum value of 80 pF with a 225µh coil. That sets it at 1186 khz.

Ok, it is time to try the high range. Since the low range is 225µh, the high coil will be about 56 µH. The low end with 415 pF now tunes 1044 khz. We have more than enough overlap, but the parallel coil low end tuning range is a tad high.

So let's go back and turn up the trimmer 20 more pF. This will give us a capacitance range of 100 pF to 435 pF. 225µh with that capacitance tunes 1061 to 509 khz. The 56µh coil and 435 pF tunes to 1020 khz. That is a 41 khz overlap. This looks like a good value to go, but let's make one more tweak.

How about raising the inductance to 232µh. The tuning range is pretty good but the low to high band split is a little higher than I like. The capacitance at 100 to 435 pF is good. Remember that the trimmer will take care of the inaccuracies.

So 230 µH and 435 pF tune to 501 khz and at 100 pF tunes to 1045 khz. The parallel coil is now 58µh. That tunes down to

STANLEY TRANSFORMER COMPANY

Stanley Match built by Mike Tuggle

The StanleyMatch is nothing radically different from other phone matching units, just a way to get more flexibility from the Stanley/UTC TF-1A-10-YY input transformers still available at a bargain from Fair Radio. By some nifty switching, one can combine two single 100 k-to-100 ohm units to get five different input-output impedance pairs. The key component is the 4-pole, 5-position switch. These come in the form of 4-deck rotary switches, as 2-deck switches with two independent 5-position sections on each deck, as shown here (actually, the sections are 6-position with one position blocked out) -- or possibly as single deck switches with two independent sections on each side! On a 4-deck switch several positions will likely need to be blocked out. That's what those tabs at the collar are for.

The 2 x 3 x 5 in. mini-box is about as small as I would suggest going -- unless you are a micro-circuitry freak and/or don't mind the smell of insulation burning on your soldering iron. Even so, I had to pre-wire several of the components before installing them in the box. Details like exact pot and capacitor value and connector types are user's choice. I just happened to have 'standardized' on 1/4-in. phone plugs and 1/8-in. miniature output jack on my sets.

The low impedance (50 - 200 ohms) outputs make this unit especially suitable for low impedance sound powered phones. I suppose you could Bogenize the low impedance end to match higher impedance magnetic phones. I've been quite happy with SP's and haven't seriously tried other types. The test points (TP) mounted on the rear apron are for audio voltage output measurements if so desired.

1002 khz. That is a 43 khz overlap. The trimmer will let you adjust to 1041 to 998 or 5 channel overlap.

Now a couple of things: First, you won't get the coil that close, but if it is plus or minus 5µh, that is fine. The trimmer will iron out the glitches.

Notice that we didn't talk about the high end of the band in the parallel coil configuration. This isn't important as it will always fall above 1700 khz.

You will have good dial spread with these values. The actual dial spread will depend on the shape of the capacitor plates (straight line capacitance, or straight line frequency). You will have better dial spread than with regular wound coil.

If you are unsure, wind the coil on the large side. It can be taken apart and adjusted if you are real far off. It is better to have less dial spread than not being able to tune the band.

This works with a cylinder wound coil or spider coil.

So to recap, we found a **232/58µh coil** would be good with a **15-350 pF capacitor** with a **60 pF trimmer** (with the fixed 25 pF for the radio and coil capacitance).

Recently Jeff Welty whipped up this page for calculating spider contra coil. It is very complete, from entering the data to printing a coil form template.

So here is the calculator. I'm sure this will take the misery and doubt out of your coil building.

I wish you the best success with your coil. (and luck too) :)

Dial Spread

There is a special feature that comes with the contra coils. That is, the tuning range on each band is somewhat wide. This means that a capacitor with a big capacitance ratio isn't really needed. But what happens is that the dial spread is not as good as it could be. Take a look at the two pictures. They are the same radio, kind of a before and after shot.

The top picture shows the tuning covers a much narrower part of the dial scale as the bottom picture. You have to look close at the numbers as the ranges are a little different. The left side is the low band (530 - 1000) and the right side is the high band (900 - 1700).

The difference is that I placed a small trimmer capacitor (about 75 pF maximum) across the main variable capacitor. Now the capacitance ratio of my variable is about 6:1 (90-550pF). Before the ratio was about 23:1 on this capacitor (20-475pF).

Stanley Match
by Mike Tuggle

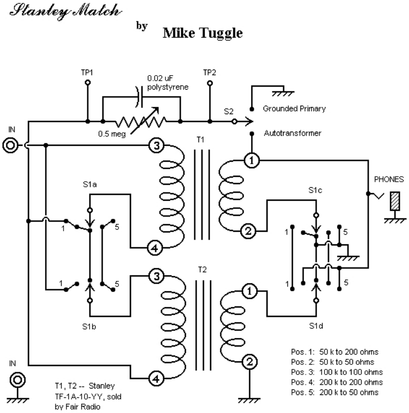
Stanley Match
by
Mike Tuggle
TP1
0.02 uF
polystyrene
TP2
S2
Grounded Primary
Autotransformer
IN
0.5 meg
T1
3
4
1
2
S1a
1
5
S1c
PHONES
T2
S1b
S1d
IN
Pos. 1: 50 k to 200 ohms
Pos. 2: 50 k to 50 ohms
Pos. 3: 100 k to 100 ohms
Pos. 4: 200 k to 200 ohms
Pos. 5: 200 k to 50 ohms
T1, T2 -- Stanley
TF-1A-10-YY, sold
by Fair Radio

The StanleyMatch combines two of the bargain Fair Radio transformers with a 4-pole, 5-position switch to provide several high-to-low impedance matches especially suitable for low impedance sound powered phones.

StanleyMatch by Mike Tuggle

Strombergs are lower AC Impedance than most other sound powered headsets and can
benefit greatly with this circuit. This will help them perform closer to the other "Big Can" decktalkers.
Also works great if using the "mic" elements out of decktalkers.
"Mic" elements are also lower impedance.

I did try a more aggressive type of band spreading using both a trimmer capacitor along with a padder type. But since I am stuck with a fixed 4:1 coil ratio, I had some trouble getting the dial spread to work correctly on both ranges. However with the way I did the dial spread, I get about a total of 2 turns of the knob in each range. A regular coil set, such as my #63 is just under 3 turns of the knob to cover the entire band. Another contravantage.

Your actual situation will be different, depending on the actual components you use. But after you build your set with a contra coil, investigate using a trimmer to widen your dial spread.

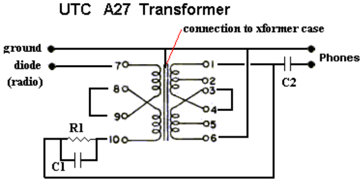
UTC A27 Transformer
connection to xformer case
ground
diode
(radio)
Phones
7
8
9
10
1
2
3
4
5
6
R1
C1
C2

Circuit for a UTC A27 Transformer
An excellent very low loss transformer
(This circuit has been revised on 29 January 2004)

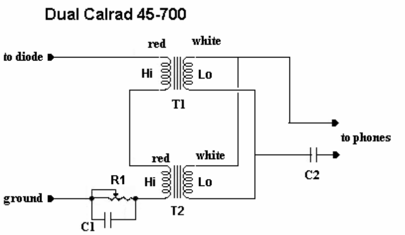
Dual Calrad 45-700
to diode
red
white
Hi
Lo
T1
to phones
red
white
R1
Hi
Lo
ground
T2
C1
C2

Dual Calrad 45-700 xformers
Excellent when using Strombergs Carlson and Dynalec Decktalkers.

C2 .47uf capacitor

Typical Dedicated Circuit With Phones

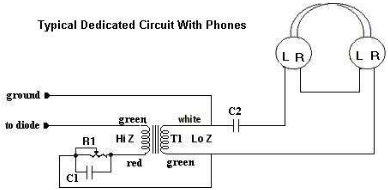

Dedicated Circuit With Phasing Switch

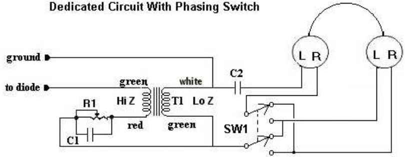

Shown out of phase or opposed

BASKET / SPIDERWEB COIL DESIGN

PA2MRX

http://www.pa2mrx.nl/UK/coil%20homebrewing_uk.htm

The tuning circuits of my xtal sets are made out of old fashioned type of coils as seen in old magazines and photo`s, these can be Honeycomb, Basket weave, Spiderweb and other forms of coils. In German these are: Honigwaben, Korbspule, Spinnwebenspule. Basket weave coils are also mentioned in books as "Lorenz" coils, and can be made in different shapes and sizes, as the Honeycomb and Spiderweb coils.
These coils were made because of their low own capacitance, this gives your coil a higher Q because the windings are further apart from eachother, there is more air between the layers.

You can make your coils using enameled copper wire, or silk or cotton insulated wire. Better if you make coils for the middlewave band using Litze wire, this wire is made out of many insulated smaller wires, from 3 up to 1500 or more small wires. But the more wires, the more expensive the wire.

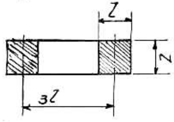

Fig. 19

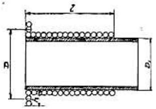

Fig. 20

A honeycomb coil is not more then a one layer cilinder coil. Many variations are made of this type of coil. After many tests and research they came to the conclusion that the selfinduction is going through a maximum when the cross section has a rectangular form (Thickness is same as width) like in fig.19 is the same, and when the average diameter is the same is 3x the thicknes of the winding

Calculation of Spiderweb/basketweave/cilinder coils can be done with the "prof. coyle" programme, but I didn`t find a formule to calculate honeycomb coils, these calculations were so complex they only published the data to create these honeycomb coils. I have the data below as found in old books:

Used capacitor in table 1 is 1000cm (+/-1100pF) wire size 0,5mm. The coil form has a diameter of 5cm, thickness is 4cm.

# Windings	Wire Length	Wavelength
25	4m	180 - 430 meter
35	6m	200 - 550 meter
50	9m	250 - 700 meter
75	14m	400 - 1000 meter
100	20m	500 - 1300 meter
150	30m	700 - 2000 meter
200	42m	1000 - 2700 meter
250	50m	1300 - 3600 meter
300	63m	1600 - 4200 meter

Source : Bastelbuch fur radioamateure, 1925

Coilform in table 2 is 5cm in diameter and a width of 2,5cm. Own wavelength is the wavelength of the coil with its own capacitance.

Winding #	I mH	Self WL m	Wire mm	WL m	Range with C 50-1000pF m
25	0,05	50	0,56	4 meter	90 - 400

DEDICATED IMPEDENCE MATCHING CIRCUITS

Darryl Boyd

http://www.crystalradio.net/soundpowered/matching/index.shtml#Dedicated

This "fixed" setup is designed for standard 600 ohm impedance elements wired in series.
It is a standard matching transformer circuit.

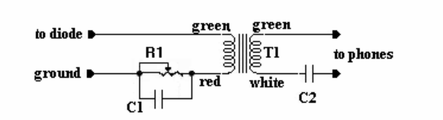

The following is another dedicated circuit for two 600 ohm series wired elements.
The transformer is set up more as a autoformer.
I think it performs a little better than the above circuit.

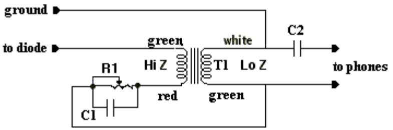

Try both circuits. They use the same parts!

T1	Calrad 45-700 audio transformer (Ocean Electronic)
R1	250k ohm pot (a mini pot would work great!)
C1	.1uf capacitor

35	0,09	90	0,56	6 meter	120 - 550
50	0,16	110	0,56	9 meter	175 - 750
75	0,3	150	0,56	14 meter	250 - 1000
100	0,6	200	0,56	20 meter	300 - 1500
150	1,5	280	0,56	30 meter	500 - 2000
200	2,5	350	0,4	40 meter	650 - 3000
250	4	400	0,4	50 meter	800 - 3500
300	6	500	0,4	60 meter	1000 - 4500
400	10	600	0,4	80 meter	1250 - 6000
500	15	700	0,3	120 meter	1500 - 7000
1000	70	1500	0,2	225 meter	3000 - 15000

Source: Der radioempfangsapparat, 1924

Coils as in the upper tables are made on a coilform as seen on description and photo`s below, but you can alter these by varying number of pegs used.

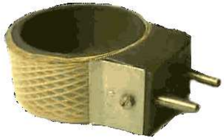

The Honeycomb coil

Coil form with its measurements are in the following drawings. You can use your own measurements and the number of pegs. But the above values are for these coils here.

You can make your form out of wood, but I made it out of alluminium on a lathe, the pegs are made out of knitting pins, bought these at the weekly market here, if you buy these with 2 pointed sides you can make two pegs of `m, so you need to buy half of the knitting pins you need for the coil form.

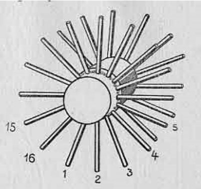

Abb. 59. Der Spulenkörper mit den eingesetzten Drahtstiften.

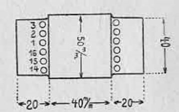
3
2
1
16
15
14
50m/m
40
20
40m/m
20

Abb. 58. Spulenkörper zum Wickeln von Honigwabenspulen.

A Honeycomb coil is getting larger in diameter the more layers are wound on it, but the width stays the same.

You can use your own style the make your coil, over2/under2, over2/under3, over3/under2. Do it as you like and look at the results.

Winding method seen at the left: Over1/under 4.

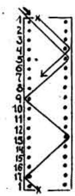
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
1
x
x

Abb. 58. Wicklungsſchema für Honigwabenſpulen.

Fertig montierte Honigwabenſpule.

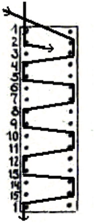
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
1

Over2/under0

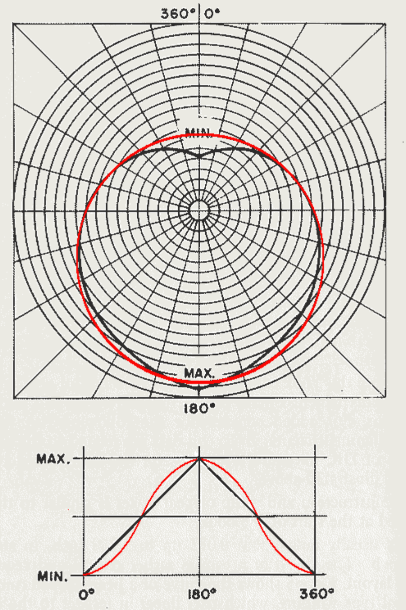
360°
0°
MIN.
MAX.
180°
MAX.
MIN.
0°
180°
360°

Figure 55—Layouts of a heart-shaped cam.

I also made a aircoil out of 6mm copper brake piping, 11cm in diameter, 8 windings, space between winding is 6mm.

These type of coils were used very often in old radio equipment.

Self resonance at 17.690Mhz, -3db point at 17.980 and 17.350, makes a bandwidth of 630Khz.

Q of this coil at 17.690Mhz is 17.690/0.63=28

How do I get I veeery high Q at shortwave,..I think never, even a aircoil I made with a diameter of 25cm had a low Q.

21-7-2008

I also made myself a coil winder according to the Gingery Coil Winder manual. With this type of winder you can make yourself some nice small radio coils for your project.

The coil winder in the above links are in principle the same as the Gingery winder, only the cam has been changed into a heart shaped one. This type of cam has the advantage that the wire does go around the circumference of the coilform in a triangle shaped fashion, instead of a sinus with the round excentric cam. The heart shaped cam gives the coil a much higher stability when winding the coil. Below a scan from WM. Querfurth`s book of the heart shaped cam with the round excentric cam in red to compare the two cams.

If you use only one side of the coil form, you can wind your spiderweb coil on this coilform, but you have to fix it with glue, the needlework will not work here.

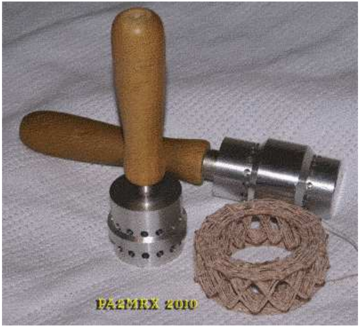

November 2010
On a lathe I made two new coilforms to wind coils. Both have two rows of 15 holes in it to put in some knitting pins to wind the wire around.

The sizes are as mentioned in old books for the coils mentioned above, one 4.5cm thick and one 2.5cm thick with a inner diam of 5cm.

Winding method: Over two, directly to other side and again over two at the other side. (over 2/under 0)

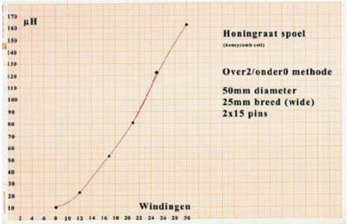

I wound the smallest coil mentioned above (2.5cm wide). Precisely according the description in the books with 25 windings. The value should be 50µH, but after I measured it with my AADE meter and another LCR bridge the value was around 122µH. Same for wire length, I took 4 meters and wound the coil till I got no wire left,..8 windings with a value of 12µH. So just wind your coil, measure and put some more wire on it or take some windings off till you get what you want.

For the above honeycomb coil I made a table for future use when I need to make some coils. I made 3 and put the value and windings on mm-paper, I connected the points to get a rough guess of inbetween points.
Then I made the coils I needed and measured them and also put them inside the table. It only goes to 162µH because I didn`t wind a bigger one.

Table is for honeycomb coils using the "Over2/Under0" method, wound onto a 5cm in diameter form, 2.5cm wide with 2x15 pins.

was resonant at 11.110Mhz. -3dB point was at 11.220 and 10.930, this makes a bandwidth of 290khz.

.

So this is the frequency where the coil is resonant with it`s own capacitance. The Q of this coil is 11.110mhz/0.29mhz=38
Not really a good choise to use for a chort wave crystal receiver.

Its own capacitance is approx. 13pF.
If I use this coil on shortwave, the selectivity would go any better then this 290khz. Q is only getting better at a lower frequency.

I also measured the self resonance of my 440x46 spiderweb coil of my 2006 BTTF receiver.

Resonant at 2.523Mhz, -3dB point at 2.539 and 2.505Mhz, this makes a bandwidth of 34Khz.

The Q of the coil at this frequency is 2.523/0.034=74.

Here I see that the Q is much better at lower frequency as i measured in another chapter on my site.

Self capacitance is around 16pF

If I wanted to make myself a 10-15Mhz shortwave crystal receiver with a bandwidth of around 9Khz, then I
would have to use a coil with a Q of around 1300, this means a self resonance point that is at a much higher level, and
a even lower capacity,..even lower as 5pF,..I still want to seperate the station,..won`t I.

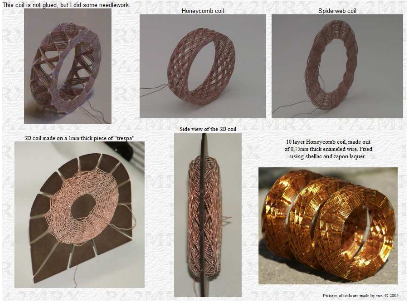
This coil is not glued, but I did some needlework.
Honeycomb coil
Spiderweb coil
Side view of the 3D coil
3D coil made on a 1mm thick piece of "trespa"
10 layer Honeycomb coil, made out of 0,75mm thick enameled wire. Fixed using shellac and zapon laquer.
Pictures of coils are made by me. © 2005
PA2MRX

Pictures of coils are made by me. © 2005

Fixation of the coils so they do not fall apart can be done using some needlework, another used methode from the beginning of the 20th century is to fix the coil using "zapon-laquer". This a laquer on a nitrocellulose basis. Long time ago you could make this yourself by dissolving celluloid film negatives into ethanol. Beware: film negatives these days are not made out of celluloid. Now you submurge the coil into the solution you get and after that you get it out of the solution and let it dry. You also can use "shellac" to fixate the coils, this is available plentyfull on Ebay,..a kilo for 8 Euro`s or so.

Self resonance of coils

I made a cilinder aircoil with a value of 15µH. Without a capacitor connected I measured the resonance of this coil. It

The basket weave coil. (lorenz coil)

You can create your basket weave coils using a piece of wood or something. I used a piece of plexiglass with a thickness of around 1cm and drilled holes in it, I used steel nails to put in the holes. The nails are placed in a cicle of around 12cm. I wound the wire around the nails using the over1/under2 methode. But this can be re-arranged by you using anothe method,..over1/under1, over1/under3,..you name it.

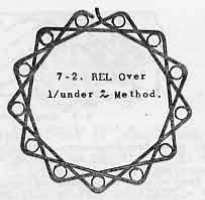
7-2. REL Over
1/under 2 Method.

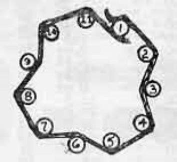
1
2
3
4
5
6
7
8
9
10
11

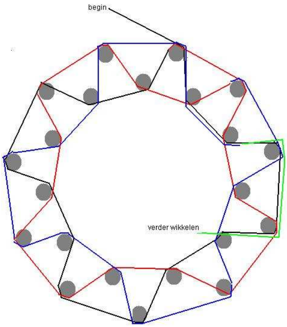

over2/onder methode

Another type of coil is the basket weave coil made using two rings of nails to wind your wire through. At the right you see 2x11 steel nails. The methode of winding here is over2/under2, but you also can use over1/under2 or something else.

Note: Of you follow the wire in the over2/under2 methode, you get to the starting point after 3 circulations of the wire. Now you layed down just one layer. This also includes the coils mentioned above.

This is what they mean by the windings do not see eachother, there is lots of air between every layer, this gives you a coil with a better Q.

This way I made several basket weave coils and some of these had one layer after 8 windings.

In table 1 they used 25 windings and 4meters of wire on a coilform of 5cm in diameter. These 25 windings are the windings,..not the layers. You have to watch out not lay down 25 layers.

A basket weave coil stays the same diameter while the width is getting more.

Below some pictures of coils I made and on their form. (this is my first and old coilform)

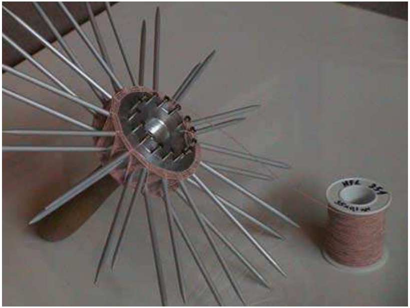

Coil being made using the form with knitting pins

KEVIN'S WEBSURFER HANDBOOK III
FOR CRYSTAL RADIO

DIODES:
SOLID-STATE & VACUUM

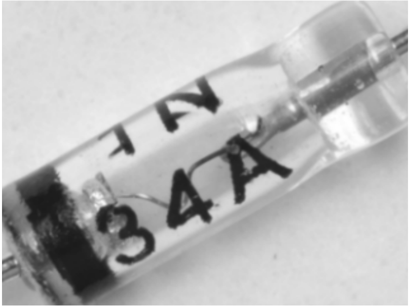

Kevin Smith
2011

Printing / Binding Instructions

1. Choose “fit to page” in print menu
2. Print document double sided on letter size paper
3. Cut the entire printed document in half
4. Fold over making sure the page numbering is continuous
5. For the cover: Print just the first page on card stock paper
   Cut the cover in half as well
6. Assemble the covers on the document
7. Punch the left side for a binding, spiral or comb as desired

http://www.lessmiths.com/~kjsmith/crystal/catalog.shtml

KJ Smith

NOTES:

INTRODUCTION

The contents of this second handbook concern themselves primarily with diodes, diode theory, and aspects of older vacuum diodes. The rectifier diode forms an essential part of the crystal radio circuit, it has a fascinating history, and few parts of the set deliver as much confusion (to me at least) and speculation, comment and even controversy. This document results from my web explorations of diode theory in general and diodes as crystal radio rectifiers in particular. Additionally, I have held a certain fascination with the vacuum diode ever since when reading on the history of radio, I found that this rectifier was actually discovered and patented prior to the lowly galena crystal from which the sets take their name.

In this booklet I have taken the sections and pages that I spend much/most time studying and have arranged them into sequence from basic theory and design considerations to more advanced material and finally on specifically vacuum diodes in particular. Some vacuum tube chapters were especially painful to include as they required pasting in each page of a scanned document. Still, I felt, and feel that this document is especially worthy and useful as a source of vacuum diode theory. Each section is useful and understandable to the hobbyist with moderate to good experience. The handbook will help the beginner to quickly get up to speed and allow the experienced builder to find endless new ideas. This is not a book of Hookups or circuit designs, that is covered in my Catalog of Crystal Hookups, nor is it a tutorial on Crystal Radio, that can be found in my Handbook Volume 1.

Much of the material in this handbook is copyright for which I have not sought permission. Therefore this is not presented for publication or copy, and certainly not for profit. It is only my personal resource. I encourage anyone finding this copy to pursue ON THE WEB the web pages identified within. I include the name of the author and web address of each section. I wish to sincerely thank every author presented for their excellent pages and ask forgiveness for my editing into this handbook.

Kevin Smith
2011

www.lessmiths.com/~kjsmith/crystal/cr0intro.shtml

NOTES:

NOTES:

CONTENTS:

1 **How to build a sensitive crystal receiver**
Dick Kleijer

15 **Crystal Set Analysis**
Berthold Bosch

39 **Diode Characteristics**
Kenneth A. Kuhn

55 **I-V characteristics of real p-n diodes**
B. Van Zeghbroeck

59 **Germanium diode I-V Characteristics**
Jose Maria Roman

77 **Diodes**
Dick Kleijer

87 **Diode forward voltage drop**
David Knight

93 **Measuring Is and n of a diode**
Ben H Tongue

103 **Diode Modeling**
Kevin Smith

113 **Diode Test**
Kevin Smith

147 **Diode Test**
Mike Tuggle

155 **Diode Model Notes**
Mike Tuggle

163 **Effect of Diode Rd on Audio Power**
Dan McGillis

171 **Practical Considerations**
Ben H Tongue

197 **Radio Detectors and How They Work**
John W Campbell Jr

201 **Theory of Vacuum Tubes**
J. B. Calvert

247 **Vacuum Diode or Diode Valve**
Competition Science Vision

259 **The Diode Valve**
Graham Dixey

266 **Notes**

and assuming a constant value for the load Rl then, by drawing a separate load line for each value of supply voltage, the dynamic characteristic can be obtained, as shown below.

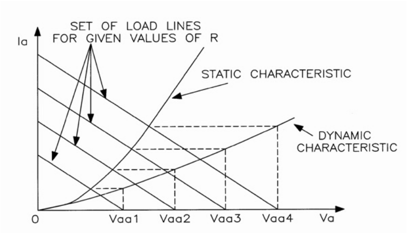
SET OF LOAD LINES
FOR GIVEN VALUES OF R
Ia
STATIC CHARACTERISTIC
DYNAMIC
CHARACTERISTIC
0
Vaa1
Vaa2
Vaa3
Vaa4
Va

Obtaining the dynamic characteristic for a diode valve and its load.

The points on the dynamic characteristic are obtained by projecting, horizontally, the intersection of a load line and the static characteristic until it in turn intersects a vertical line drawn from a supply voltage value. Since the dynamic characteristic is drawn for a range of values of the supply voltage, this implies that the latter is varying, in other words it is an alternating supply rather than DC, as is the case in rectification.

current in the circuit will vary by using a graphical construction.

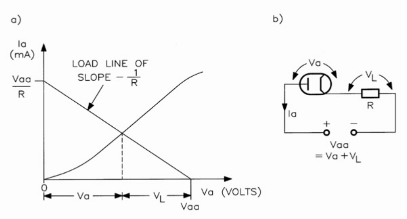

(a) The load line for a diode valve. (b) the diode in series with a linear resistive load.

The second image shows the diode in series with its load and defines the voltages and currents in question. A 'load line' of slope -l/R is drawn between the two points: Ia = Vaa/R Va = 0 and Ia = 0, Va = Vaa. As in solid state practice, the end points of a load line define zero conduction and maximum conduction, the latter being dependent upon the values of supply voltage and load resistance. Any other points on the load line imply intermediate levels of conduction. By dropping a construction line from the intersection of the load line and the static characteristic, we can see how the total supply voltage Vaa is divided into the two separate voltages Va (the voltage across the diode) and Vl (the voltage across the load).

The Dynamic Characteristic

By taking a number of different values of supply voltage Vaa (as would happen if the supply was alternating, for example)

How to build a sensitive crystal receiver

Dick Kleijer crystal-radio.eu
http://www.crystal-radio.eu/engev.htm

[Calculator included at the bottom of the web page to calculate the components of your receiver for maximum sensitivity at weak signals].

In this article some information about the design of a crystal receiver with maximum sensitivity.

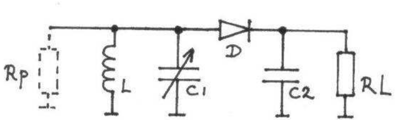

Circuit diagram 1
Circuit diagram of the crystal receiver, which we are going to design for maximum sensitivity at weak signals.
This can be a detector circuit of a 2 circuit receiver.
But also a receiver with loop antenna.

RP represents the losses in coil L and tuner capacitor C1

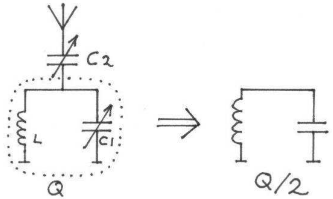

Circuit diagram 2
If an antenna is connected via a matching network (in this case C2) to the LC circuit, this will reduce the circuit Q.
When optimal matched to the antenna, the Q will be halve the value of the unloaded Q (of L,C1).
For a calculation of the values of C1 and C2, click here.

For the calculation of maximum sensitivity: replace the LC circuit + antenna by a LC circuit with an unloaded Q equal to halve the original value.
Coil value L stays the same.

You get maximum sensitivity if there is maximum power transfer from LC detector circuit to load (loudspeaker).
The LC circuit has a certain parallel resistance RP, this is not a real resistor, but a virtual resistor caused by the losses in coil and tuner capacitor.

If we know the Q-factor of the detector circuit, and the induction of the detection coil, we can calculate the value of RP as follows:

RP = 2.pi.f.L.Q (Ohm)

The Anode Slope Resistance ra

It is worth introducing this parameter at this time since it is one that we shall make use of later in discussing the performance of more complex valves. It is defined as shown below,
Defining anode slope resistance for a diode valve.

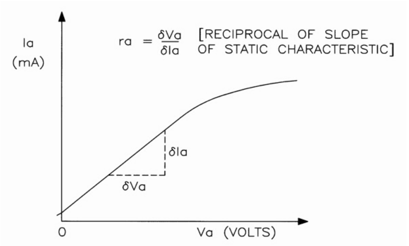

and is the value of resistance obtained by dividing a small change in anode voltage by the corresponding change in anode current. It is therefore the reciprocal of the slope of the static characteristic, and varies with the operating point, although fairly constant over much of the space charge limited region. This is a real value of resistance, since it represents the opposition of the valve to alternating quantities.

Series Circuit Operation

It is usual to operate a diode valve, which is clearly a non-linear device, in series with a resistive load, the latter being a linear device. It is possible to predict how the voltages and

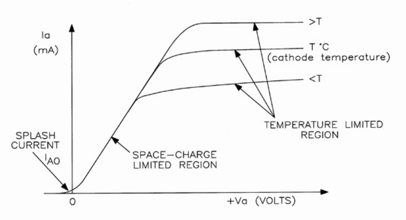

The static characteristic of a diode valve.

Above three curves have been drawn for different values of cathode temperature, although in practice, as explained earlier, the cathode is held at a constant temperature.

It is interesting to note that:

A) The current is not exactly zero when the anode voltage is zero, but has a value (Iao) of a few micro-amperes. This is known as the 'splash current' and is the result of a few high energy electrons that manage to cross the inter-electrode gap even without an attracting potential.
B) In the space-charge limited region, the characteristic is nearly linear (actually following the 'three-halves' power law: Ia is proportional to Va-3/2).
C) In the temperature limited region there is little change in la even though there are large changes in Va. This is because the anode is collecting electrons at the same rate as they are being emitted by the cathode.
D) No significant current flows when the anode is negative with respect to the cathode.

pi = 3.14
f = the frequency (Hertz).
L = induction of detector coil (Henry = H).
Q = the quality factor of the unloaded LC circuit.

The value of RP is depending on the frequency, use for instance the value of RP in the middle of the frequency range. For medium wave for instance, we can use the value of RP at 1 MHz.

Maximum power transfer from LC circuit to load.
If the load resistor RL was directly connected across the LC circuit, it would be simple:
maximum power transfer would occur if RL=RP.
The loaded Q of the LC circuit is then halve the value of the unloaded Q.

But of course we also need a diode between LC circuit and load RL, to demodulate the RF signal.
In the rest of this article we assume the RF signal on the LC circuit is not modulated, so the output of de diode is a DC voltage.

Maximum sensitivity at strong signals.
With a very strong signal across the LC circuit, the diode shall work in the linear detection region.
If the input voltage is high enough, the diode will give only very few power losses, compared to the rectified power.
In the calculation (with strong signals) I assume the diode has no losses at all.
This can in practice not be reached, but it rather simplifies the calculations.

If the diode has no losses, the DC voltage across the load resistor will be equal to the peak voltage of the RF signal.
This peak voltage is 1.41 times the RMS value of the input signal.

In this situation, maximum power transfer to the load RL will occur if:
RL = 2 x RP
The loaded Q of the LC circuit is then halve the value of the unloaded Q.

However with strong signals, maximum sensitivity is not a very important subject, output power is already high enough to be heard.
We can better design the receiver for maximum sensitivity at weak signals.

Maximum sensitivity at weak signals.
When receiving (very) weak signals, the diode shall work in the square law detection region.

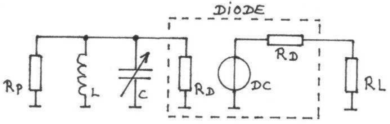
DIODE
RP
L
C
RD
DC
RD
RL

Circuit diagram of the diode.

With the equivalent circuit of the diode at low signal levels

which may be detected by an ammeter placed in, say, the anode lead. The picture below shows an illustration of this.

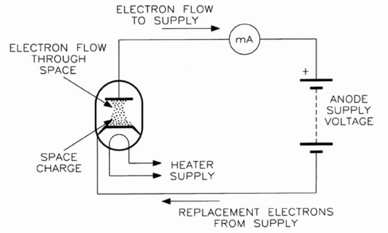
ELECTRON FLOW TO SUPPLY
mA
ELECTRON FLOW THROUGH SPACE
+
ANODE SUPPLY VOLTAGE
SPACE CHARGE
HEATER SUPPLY
REPLACEMENT ELECTRONS FROM SUPPLY

The flow of current in a diode valve.

Diode Static Characteristics

We now start getting into the ways in which specifications for thermionic devices are presented. For the diode, these illustrate clearly the dependence of anode current upon anode voltage.

passed forward into semiconductor phraseology - cathode, anode, emitter, collector!).

Since electrons are negative charged particles, they will only be attracted to the anode if this is given a positive potential with respect to the cathode. This explains why the valve only conducts in one direction, from cathode to anode, and not vice versa. It also explains the choice of the word 'valve' to describe the device, since a valve is, by definition, a one-way device. The Americans, however, never cottoned on to this terminology and always refer to them as 'vacuum tubes'.

The magnitude of the current flowing in a diode depends upon the number of electrons emitted and the magnitude of the voltage applied to the anode (known as the anode voltage Va). The amount of electron emission depends upon the temperature of the cathode, which is fixed by the voltage supply to the heater, this being a constant value. The only true variable is, therefore, the anode voltage. The action of the latter in controlling the anode current can be explained as follows.

As we now know from the foregoing, the cathode is normally surrounded by a cloud of electrons known as the space charge. With zero anode voltage there is no current flow, and there is a state of equilibrium between the electrons being emitted and those falling back onto the cathode's surface, The application of a small positive voltage to the anode causes some of the space charge electrons to be attracted to the anode, resulting in a small anode current flow. The gaps created by these electrons leaving the space charge are filled by further emission from the cathode. Electrons arriving at the anode flow to the positive supply terminal, while at the same time an equal number of electrons leave the negative supply terminal for the cathode. This gives rise to a continuous current flow around the circuit,

The input of the diode behaves like a resistor with value RD parallel to the LC circuit.
The diode output is like a DC voltage source in series with a resistor RD.
Maximum power transfer from DC voltage source to load RL occurs when: RL = RD.

In the square law detection region, the detected DC voltage is proportional to square of the RF input voltage.
The power in the load resistor is proportional to the square of the detected DC voltage.
In other words, the power in load RL is proportional to the 4th power of the voltage across the LC circuit.
So it is important to make the voltage across the LC circuit as high as possible, this is done by making the impedance of the loaded LC circuit as high as possible.

If we make all impedances equal, so RL = RD = RP the loaded Q of the LC circuit would be equal to 0.5 times the unloaded Q.
The voltage across the LC circuit is then also 0.5 times the unloaded voltage.
The output voltage is then proportional to: 0.5^4 = 0.0625
If we however make RD and RL 3 times as high, so RL = RD = 3xRP the loaded Q of the LC circuit will be 0.75 times the unloaded Q.
The voltage across the LC circuit is then also 0.75 times the unloaded voltage.
The output power is then proportional to (0.75^4)/3 = 0.3164/3 = 0.1055.
We divide by 3 because the load resistor is now 3 times higher.
So we now have 0.1055 / 0.0625 = 1.6875 times more output power (compared to the situation: RL = RD = RP).
And also, the loaded Q is now 1.5 times higher, so a better selectivity for the receiver.

In the next Excel file: diode.xls the relative output power is calculated for many values of RD and RL, where RD=RL=3xRP is the most sensitive combination.

Conclusion: At weak signals there is maximum power transfer from LC circuit to load, so maximum sensitivity for the receiver if:
RL=RD=3xRP The loaded Q is then 0.75 times the unloaded Q.

Increasing the circuit Q

Increasing the unloaded Q of the LC circuit, will make the impedance of the LC circuit higher.
Because of this the voltage across the LC circuit will increase, the diode then will work more efficient, and output power increases.
More information about increasing the Q, you can find here.
Very good crystal receivers make use of LC circuits with an unloaded Q above 1000.
A high Q also gives a selective receiver.

Increasing coil induction

The impedance of the LC circuit can also be increased by increasing the coil induction (μH value).
However making the induction higher, can also result in a lower circuit Q, compensating a part of the impedance increase.
And a lower Q of course gives worse selectivity.
Practical values for induction of medium wave receivers are 200-300 μH.

http://www.r-type.org/articles/art-010b.htm

Valve Technology - A Practical Guide
A series of articles from 1993 by Graham Dixey C.Eng., MIEE republished by kind permission of Maplin Magazine.

The Diode Valve

The diode valve is so called because it has just two electrodes – the cathode and the anode.

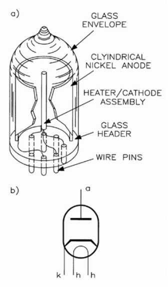

(a) Construction of a modern diode valve (indirectly heated type), (b) circuit symbol for a diode valve.

These correspond to the two electrodes of the original diode valve mentioned above, the cathode being the electrode that is heated and emits electrons, and the anode being the electrode that collects the electrons (notice also that these terms have

Choice of the diode

If we know the diode resistance RD, we can calculate the diode "saturation current" (Is value) with the formula:

Is = 0.000086171 x n x TK / RD

Is = saturation current of the diode in A
n = ideality factor of the diode, if you don't know the value, then take n=1.08
TK = temperature in Kelvin (= temperature in ºC + 273)
RD = diode resistance in Ω

Now search for a diode which has a Is value close to the calculated Is value.
Or connect several diodes with a low Is parallel, to reach together the desired value.
More information about diodes you can find here.

Diode connected to a tap on the coil.

The best sensitivity is always reached if the diode is connected to top of the LC circuit, and RL=RD= 3xRP.
But not always it will be possible to make the load resistance RL high enough.
If we don't have a load (audio transformer) with a high enough impedance, the LC circuit will be loaded to heavy, and the Q will reduce.
A method to prevent this, is to connect the diode on a tap somewhere on the coil, instead of the top of the coil.
The Q of the loaded circuit then will increase.
With the calculator at the bottom of this page (click here) the effect on output power can be calculated when using a tap on the coil.

Audio transformer.

When making a receiver with a high Q factor, the load resistance will have to be very high, for instance some Mega-Ohms.

In that case we need a transformer which transforms this high impedance down to the impedance of the used loudspeaker.

Transformers with such a high input impedance are hardly to find.

They can however be self made, look for instance at my transformer unit 2

At high input impedances, in practice both efficiency and bandwidth of a transformer will reduce.

Loudspeaker

Of course we use a loudspeaker with a sensitivity as high as possible.

More information about this, you can find here.

Calculate optimal components for detector circuit

With calculator 1, diode and load resistor are calculated for which the sensitivity of the detector circuit at weak signals is maximum.

Fill in the yellow coloured fields, and click on "calculate calculator 1".

These values can then be taken over in calculator 2.

Then you can self change in calculator 2 all parameters of the detector circuit (all yellow fields), and calculate the effect of this on sensitivity and loaded Q.

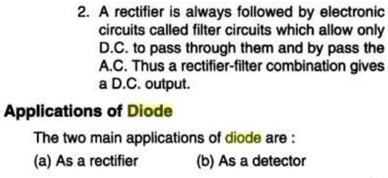

2. A rectifier is always followed by electronic circuits called filter circuits which allow only D.C. to pass through them and by pass the A.C. Thus a rectifier-filter combination gives a D.C. output.

Applications of Diode

The two main applications of diode are :

(a) As a rectifier (b) As a detector

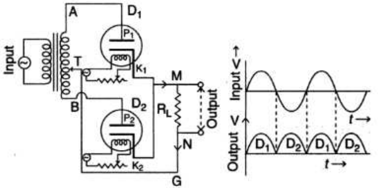

(i) Circuit (ii) Input-Output waveforms

positive. Therefore, the second diode conducts but the first remains passive. Thus, the two diodes conduct alternately. But in both half cycles of A.C., the current in load resistance R_L flows in the same direction. So, we get continuous D.C. at output. For full-wave rectifier,

Note :

$$I_{D.C.} = \frac{2I_0}{\pi}$$

$$I_{rms} = \frac{I_0}{\sqrt{2}}$$

$$\eta_{max} = 81.2\%$$

Ripple factor, $r = 0.48$

The output of rectifier is fluctuating or pulsating. To make it smooth, filters are used.

Notes : 1. The output obtained from a half wave or full wave rectifier is not a pure D.C. but it is a pulsating D.C. (mixture of A.C. and D.C.)

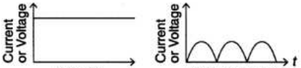

(i) (D.C.) (ii) Pulsating D.C.

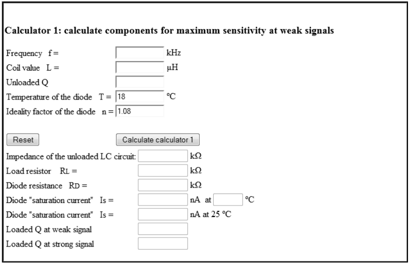

Calculator 1: calculate components for maximum sensitivity at weak signals

Frequency f = kHz
Coil value L = µH
Unloaded Q
Temperature of the diode T = 18 °C
Ideality factor of the diode n = 1.08

Impedance of the unloaded LC circuit: kΩ
Load resistor RL = kΩ
Diode resistance RD = kΩ
Diode "saturation current" Is = nA at °C
Diode "saturation current" Is = nA at 25 °C
Loaded Q at weak signal
Loaded Q at strong signal

Calculator 1: calculate components for maximum sensitivity at weak signals
Frequency f = kHz
Coil value L = µH
Unloaded Q
Temperature of the diode T = ºC
Ideality factor of the diode n =

Impedance of the unloaded LC circuit: kΩ
Load resistor RL = kΩ
Diode resistance RD = kΩ
Diode "saturation current" Is = nA at ºC
Diode "saturation current" Is = nA at 25 ºC
Loaded Q at weak signal
Loaded Q at strong signal

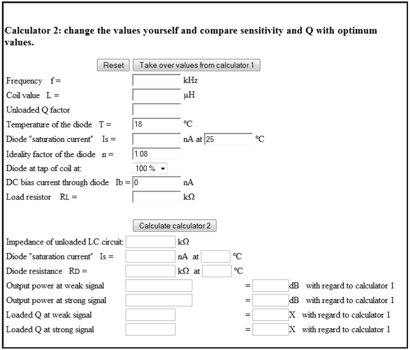

Calculator 2: change the values yourself and compare sensitivity and Q with optimum values.

Frequency f = kHz
Coil value L = µH
Unloaded Q factor
Temperature of the diode T = 18 °C
Diode "saturation current" Is = nA at 25 °C
Ideality factor of the diode n = 1.08
Diode at tap of coil at: 100 %
DC bias current through diode Ib = 0 nA
Load resistor RL = kΩ

Impedance of unloaded LC circuit: kΩ
Diode "saturation current" Is = nA at °C
Diode resistance RD = kΩ at °C
Output power at weak signal = dB with regard to calculator 1
Output power at strong signal = dB with regard to calculator 1
Loaded Q at weak signal = X with regard to calculator 1
Loaded Q at strong signal = X with regard to calculator 1

Calculator 2: change the values yourself and compare sensitivity and Q with optimum values.

Frequency f = kHz
Coil value L = µH
Unloaded Q factor
Temperature of the diode T = ºC
Diode "saturation current" Is = nA at ºC
Ideality factor of the diode n =
Diode at tap of coil at:
DC bias current through diode Ib = nA
Load resistor RL = kΩ

Impedance of unloaded LC circuit: kΩ

Operation

- A.C. input is applied the plate 'P' becomes positive and negative alternately.
- For the positive half cycle of input, the plate is positive with respect of cathode. So that diode conducts and plate current flows through the tube load R_L and secondary. So output occurs across load. R_L as shown in fig. (ii).

Important Points

- During – ve half cycle of A.C. input, →Plate is negative with respect to cathode. So the diode does not conduct, and no voltage appears across output.
- For half wave rectifier; (I_0 is peak value of current)
  - → Average value of current $I_{D.C.} = I_0/\pi$
  - → Root mean square value of current $I_{rms} = I_0/2$
  - → Maximum efficiency $\eta_{max} = 40{\cdot}6\%$
  - → Ripple factor $r = 1{\cdot}21$

(B) **Full wave rectifier**—It converts full A.C. into D.C. In this rectifier two diodes are used which conduct alternately. The output is obtained across the load resistance R_L. Here in first half cycle of A.C., plate of first diode remains positive and plate of second diode remains negative. Hence, the first diode conducts and the second does not. In second half cycle of A.C., plate of the first diode becomes negative and that of second diode becomes

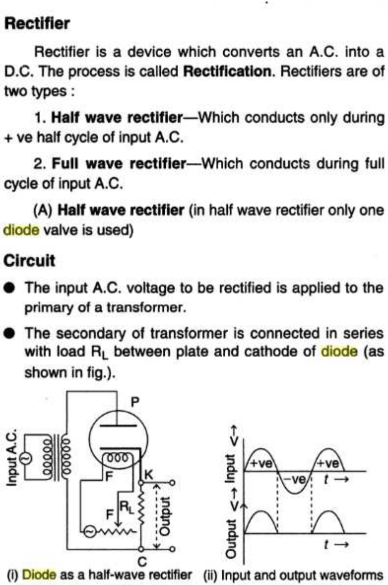

Rectifier

Rectifier is a device which converts an A.C. into a D.C. The process is called **Rectification**. Rectifiers are of two types :

1. **Half wave rectifier**—Which conducts only during + ve half cycle of input A.C.

2. **Full wave rectifier**—Which conducts during full cycle of input A.C.

(A) **Half wave rectifier** (in half wave rectifier only one diode valve is used)

Circuit

- The input A.C. voltage to be rectified is applied to the primary of a transformer.
- The secondary of transformer is connected in series with load R_L between plate and cathode of diode (as shown in fig.).

(i) Diode as a half-wave rectifier (ii) Input and output waveforms

Diode "saturation current" Is = nA at ºC
Diode resistance RD = kΩ at ºC
Output power at weak signal = dB with regard to calculator 1
Output power at strong signal = dB with regard to calculator 1
Loaded Q at weak signal = X with regard to calculator 1
Loaded Q at strong signal = X with regard to calculator 1

Example 1.
We have a LC circuit with a 200 µH coil, the unloaded Q is 1000 and the frequency is 1000 kHz.
As load we use a 100 kΩ audio transformer, because this is the only one we have.
What is the output power at weak signals compared to the optimal load resistor.

Fill in the following values, in calculator 1:
f = 1000 kHz
L = 200 µH
Q = 1000
Click on "calculate calculator 1" (The optimal load resistor and diode are now calculated).
Click on "take over values from calculator 1"

Change load resistor RL in calculator 2 in 100 kΩ.
Click on "calculate calculator 2"
We now see the sensitivity at weak signals is 9.97 dB lower compared to the optimal load.
It's also interesting to see, the Q at strong signals is only 38

Example 2.

We have 2 receivers, both with a 200 µH coil.

One receiver has an unloaded Q of 100, the other receiver has an unloaded Q of 1000, at 1000 kHz.

Both receivers have optimal diode and load resistor.

What is the difference in output power between the 2 receivers?

Fill in the following values, in calculator 1:

f = 1000 kHz

L = 200 µH

Q = 1000

Click on "calculate calculator 1" (The optimal load resistor and diode are now calculated).

Click on "take over values from calculator 1"

Reduce the Q in calculator 1 to: 100.

Click on "calculate calculator 1" (The optimal load resistor and diode are now calculated).

Click on "calculate calculator 2".

The result is, the high Q receiver gives 1000 times more output power (+30 dB) at weak signals.

At strong signals, the difference in output power is 10 times (+10 dB).

Example 3.

We have a simple crystal receiver, with a 200 µH coil, and an unloaded Q of 200.

As diode we use a germanium diode with Is = 400 nA (at 25 ºC).

The load resistor is 47 kΩ.

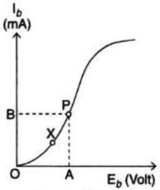

as D.C. plate resistance. The D.C. plate resistance R_p is given by

$$R_p = \frac{OA}{OB}$$

Note : Since plate characteristic curve is not a straight line—Therefore, D.C. plate resistance is variable. Hence, D.C. plate resistance may be calculated at the actual operating point.

(ii) **A. C. plate resistance**—The ratio of a small change in plate voltage across a diode to the resulting change in plate current is known as A.C. plate resistance.

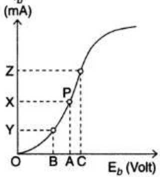

$$r_p = \frac{\Delta V}{\Delta I}$$

Note : The A.C. plate resistance at operating point P can be found by considering small equal changes of plate voltage on either side of the operating point (*i.e.*, AB = A.C.)

Change in plate voltage = BC

Change in plate current = YZ

$$\therefore \text{ A.C. plate resistance at } p,\ r_p = \frac{BC}{YZ}$$

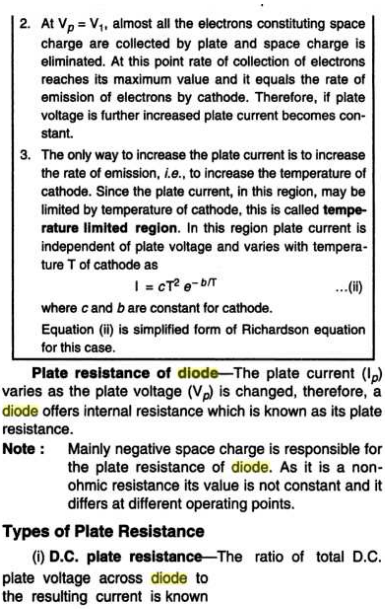

2. At $V_p = V_1$, almost all the electrons constituting space charge are collected by plate and space charge is eliminated. At this point rate of collection of electrons reaches its maximum value and it equals the rate of emission of electrons by cathode. Therefore, if plate voltage is further increased plate current becomes constant.
3. The only way to increase the plate current is to increase the rate of emission, *i.e.*, to increase the temperature of cathode. Since the plate current, in this region, may be limited by temperature of cathode, this is called **temperature limited region**. In this region plate current is independent of plate voltage and varies with temperature T of cathode as

$$I = cT^2 e^{-b/T} \qquad \ldots\text{(ii)}$$

where *c* and *b* are constant for cathode.

Equation (ii) is simplified form of Richardson equation for this case.

Plate resistance of diode—The plate current (I_p) varies as the plate voltage (V_p) is changed, therefore, a diode offers internal resistance which is known as its plate resistance.

Note : Mainly negative space charge is responsible for the plate resistance of diode. As it is a non-ohmic resistance its value is not constant and it differs at different operating points.

Types of Plate Resistance

(i) **D.C. plate resistance**—The ratio of total D.C. plate voltage across diode to the resulting current is known

We are going to compare the sensitivity at 1000 kHz, with a high quality receiver, with L = 200 μH and unloaded Q = 1000.
This receiver has optimal matched diode and load resistor.

The values of the simple receiver can directly be entered in calculator 2.
We don't have to calculate first the optimal load and diode, because this receiver isn't using optimal load and diode.
So, in calculator 2 we fill in:
f = 1000 kHz
L = 200 μH
Q = 200
Is = 400 nA (at 25 ºC).
RL = 47 kΩ.

Fill in the values of the other receiver, in calculator 1:
f = 1000 kHz
L = 200 μH
Q = 1000
Click on "calculate calculator 1" (The optimal load resistor and diode are now calculated).
Click on "calculate calculator 2" the differences between both receiver are calculated.

Result: the simple receiver has 29 dB less output power at weak signals.
The simple receiver has a loaded Q of only 60 (at weak signals).

At strong signals, the loaded Q even reduces to 17, but with such a low sensitivity it's unlikely to get a strong signal across the LC circuit.

We note from fig. (i) that curves coincide at low voltage (in space charge limited region) but saturation current increases as shown by portion CD of characteristics.

(2) **Low voltage characteristics**—The plate current is practically zero at zero plate voltage (however, plate current is not exactly zero). If we measure plate current using a micro-ampere plate current of the order of few micro-ampere occurs. This is due to "the fact that even at zero plate voltage a few electrons may have sufficient K.E. to reach the plate and constitute a plate current of few micro-ampere." From the fig. (ii) the plate current becomes exactly zero at a particular negative value V_C of plate voltage, called **cut-off voltage.** The max. kinetic energy of emitted electron is related to V_C as :

$$K_{max} = e V_C$$

Important Points

1. If we increase plate voltage from zero volt to a value V_1 in steps, more and more electrons are attracted from space charge to plate (though at the same time same no. of electrons are emitted from cathode so that space charge is maintained) so plate current increases. As the plate current is limited by space charge, this region of curve is called **space charge limited region**. In this region, plate current is related to plate voltage by equation.

$$I_p = K . V_p^{3/2} \qquad ...(i)$$

where K is a constant.

Equation (i) is known as **Child-Langmuir's law.**

V-I characteristics of Diode

(1) **High voltage characteristics**—Fig. (i) shows high voltage plate characteristics when plate current is measured in (mA). Keeping temperature of cathode constant, say (T_1) (*i.e.* filament current constant) when plate voltage increased from zero volt in steps and corresponding value of plate current i_p measured in m.A. the variation occurs as shown in curve OAB. If the cathode temperature is increased from T_1 to T_2.

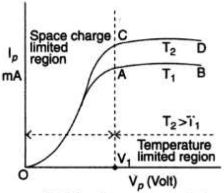

(i) High voltage characteristics

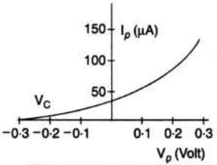

(ii) Low voltage characteristics

Fig. : V-I characteristics of diode.

Crystal Set Analysis

by Berthold Bosch, DK6YY

http://www.oldradioworld.de/gollum/analysis.htm

(First published in German in 1993/94; see references at end. Updated 12/03/2002)

(See some of Berthold Bosch´s realized xtal sets here)

Contents: 1.Voltages and Powers in Set, 2. Antenna/Earth as Signal Source, 3. Set with Parallel- Tuned Circuit, 4. Diode Properties, 5. RF Matching, 6. AF Matching, 7. Computer Simu- lations, 8. Series-Tuned Circuits .

Since my schoolboy days I have been fascinated by crystal radio reception: radio in its most basic form. However, in many cases I was not really satisfied with what I read on the subject in the literature. In the treatises I came across, the descriptions often remained rather vague, presenting little convincing foundations. Partly they were rather speculative and even presented contradicting conclusions. For this reason I found it advisable to carry out my own investigations. My intention was to obtain more quantitative results, for example as regards the best diode and the understanding of the obvious interdependence between the radio-frequency (RF), audio-frequency (AF), and DC subcircuits, what it meant for an optimum design. In the following I present results obtained over the years. Only medium-wave reception is considered.

1.Typical Values of Voltages and Powers in Set

Let us first see of what order of magnitude the RF and AF voltages and powers are which we have to deal with.

According to amplitude-modulation theory, the AF power contained in the total AM signal of power PRF is given by m2

PRF / (2+m2) where m is the modulation factor. If we assume m=0.5 we thus have 11 percent of AF power in the AM signal. Sometimes broadcasting stations use modulation factors of up to m=1 (100 percent) which then causes a correspondingly higher AF power component. At my urban location, in the West of Germany (Ruhr District), the strongest station (15 km away) produces an electric field strength of 0.18 V/m and, with my antenna and earth arrangement, an RF power of about 3 mW is available in the crystal set. Hence 330 µW of AF are contained in the RF if we assume m=0.5. A practical (linear) diode detector coupled to a tuned circuit delivers 70 to 80 percent of this to the AF load. This means that ideally I can expect about 240 µW of AF being available from my local station, sufficient for moderate operation of a loudspeaker.

In the crystal set that I am going to investigate (Fig. 2 below) I measured the following RF voltages across the tuned circuit when RF and AF matching existed (Secs. 2, 5 & 6):
a) Tuned to the local station (WDR 2, 720 kHz, 200 kW, 15 km away):8.9 V

b) From of my "district station" (DLF, 549 kHz, 100 kW, 35 km, field strength 40 mV/m, 0.2 mW of RF power):2.3 V.

c) At night - with a wave trap for the local station - more than a dozen stations appear from all over Europe with 1 to 5 mV/m, producing 130 mV across the circuit as a mean value (0.5 µW of RF). Such low voltages will move the working point over only a rather limited part of the diode characteristic where the relative curvature is low. Consequently, the detector efficiency now drops to below one percent.

d) Good headphones produce an audible signal down to 10 pW of applied AF power.

plate current. These electrons flow through the external circuit and finally return to the cathode, thus making up the supply of electrons lost by emission. On increasing plate potential more electron will gain sufficient kinetic energy so as to reach anode, hence the plate current increases.

Important Points to Note

- The current flows in the diode only when plate is made positive relative to cathode.
- No current can flow when plate is negative relative to cathode.
- Within a diode electrons can flow only from cathode to plate.
- Due to unidirectional conduction property the diode acts like a valve.
- The diode automatically starts conduction when the plate is positive or stops conduction when the plate is negative (due this property) the diode may act as a rectifier, converting A.C. into D.C.)

Characteristics of Diode

Characteristics of vacuum diode can be best studied by finding the relation between plate voltage and plate current for a given cathode temperature.

The circuit for determining the plate characteristic of an indirectly heated vacuum diode is shown in fig.

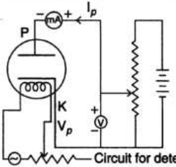

Circuit for determining plate characteristics of diode

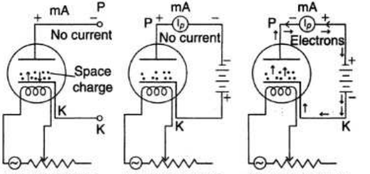

(i) Plate is open (ii) Plate is given –ve potential (iii) Plate is given +ve potential

(1) **Anode (plate) at zero potential relative to cathode**—This situation is shown in fig. (i)—The emitted electrons do not have sufficient kinetic energy so as to reach the anode. However, a few electrons may reach the anode on account of their kinetic energy, constituting negligible current. The emitted electrons accumulate near the cathode and form a cloud of electrons. This is known as space charge. At a certain stage the number of electrons forming the space charge becomes constant for a given operating temperature. This space charge becomes a source of electrons that can be attracted to the plate, if it is at a positive potential.

(2) **Anode (plate) at negative potential relative to cathode**—This situation is shown in fig. (ii). The emitted electrons are repelled back due to retarding potential of anode. For a particular negative potential of anode plate current may be zero.

(3) **Anode (plate) at positive potential relative to cathode**—This situation is shown in fig. (iii). The electrons constituting the space charge are attracted to the plate. This flow of electrons from cathode to plate is known as

Employing a signal generator and using a sensitive diode (see below) I found that an RF power of about 10 nW is required to generate this 10 pW of lower-limit AF. The detector efficiency has at this very low RF level thus fallen to a mere one per mille. Obtaining an RF power of 10 nW in my set requires a field strength of about 0.3 mV/m. According to estimates based on groundwave propagation theory, a 1000 kW transmitter operating near 1500 kHz should generated 0.3 mV/m at a distance of 190 to 200 km; the electric field strength is roughly proportional to Radicle (PTX)/(f2d2), where PTX = transmitter power, f = frequency, and d = distance. The particular example is chosen because at 1440 kHz I can during the day just hear the signal of RTL Luxembourg, being 195 km away and reported to radiate 1200 kW. The voltage measured across the tuned circuit was 40 mV in this case. To be able to receive RTL I carefully have to suppress the local as well as the district station.

e) When I connect an AF amplifier to the crystal set, a number of stations located about 150 to 250 km away can additionally be heard via groundwave propagation in the daytime. The diode thus provides (some) detector action at RF levels even lower than 10 nW. But the AF power generated is then too small to produce an audible signal in the phones directly.

The above numbers show the considerable variations in the RF voltage generated across the tuned circuit. Hence it is not surprising that a crystal diode found to be best suited for DX reception is not necessarily the optimum choice for achieving best loudspeaker operation from the local station. But before I present results on diode behaviour I am going to describe the circuit I employed. Let us start with antenna and earth as an integral part of the total circuit.

Ed Note: Evaluation of Received power from above:

		threshold	weak	medium	strong
RF power	W	1.00E-08	5.00E-07	2.00E-04	3.00E-03
RF Field	mV/M	0.3	2.5	40	180
RF voltage	V	0.02	0.13	2.5	8.9
	mV	*20.0**	130	2500	8900
		(* 40 in text: 20 provides superior fit)			
	Antenna:				
	Inverted L design, 43m (140')				
	Height = 10m (32')				
	Earth Rg = 210 Ω	Counterpoise Rg = 25 Ω			

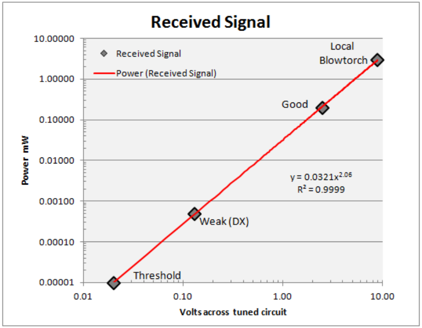

Competition Science Vision:
C.S.V./January/2007/1432

Vacuum Diode or Diode Valve

A vacuum diode consists of two electrodes : a cathode; and an anode enclosed in an evacuated glass tube. Its operation is based upon thermionic emission, it is also called thermionic diode.

Anode (or plate)—The anode is a hollow cylinder made of Ni or molybdenum and surrounds the cathode.

Cathode—A cathode may be directly heated or indirectly heated type. Usually an indirectly heated cathode is provided (except in high power application).

As shown in fig. (i) an indirectly heated cathode is in the form of nickel cylinder coated with oxides of barium or strontium. Inside the cathode, a heater filament of tungsten is inserted.

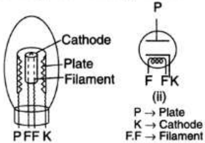

(i) (i) Construction of diode
(ii) Symbol of diode

Fig. (ii) shows it symbol :

Operation of diode—When electric current is passed through the diode it emits a large number of electrons. These emitted electrons may be accelerated or retarded by applying positive or negative potential to anode relative to cathode.

when the tube is in operation. It can just be seen glowing orange in the diode portion of the tube. On one side is the diode with the triode on top; on the other is the pentode, all of which are clearly displayed. The tube cost $3.10, an excellent price considering all the experiments that can be done with it.

The center tap of the input transformer is not used in this circuit. The triode has grid leak bias. It is not possible to measure this bias with a DMM because of the large grid resistor, but the plate current is only about 280 µA. The diode and triode are referred to the negative end of the filament, but cathode bias is provided for the power pentode. The bias in operation is about -9.3 V, giving a plate current of about 5 mA and a screen current of about 1 mA. The output transformer can be any transformer on hand. I used the P-T31 that has been used in other places, which matches 8Ω to 5 kΩ. The load resistance should preferably be about 12 kΩ. However, it drove a good speaker at considerable volume with a rather small input. A three-tube battery radio could be made with this tube, an IF amplifier stage, and a pentagrid converter, a tube type which will be treated next.

---

2. Antenna and Earth as Signal Source

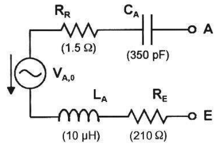

(Fig. 1: Equivalent circuit of antenna/earth combination.)

I use an inverted L-type antenna of 43 m length, about 10 m above ground. The earth connection is provided by three metal rods of 2 m length each, driven into marly, i.e. a not particularly well conducting, soil. The antenna/earth combination can be represented by the equivalent circuit shown in Fig. 1, giving measured values for the various elements. The antenna capacitance is denoted by CA, the inductance by LA, RE is the earth loss resistance, RR the radiation resistance, and VA,O the antenna source voltage. The two last elements increase in value with the antenna's height and length. For the source voltage I measured a value of 1.6 V, using a selective RF voltmeter. The knowledge of this quantity, which is produced by the strong local station, permits to easily determine the earth resistance. For it 210 ohms were obtained, a relatively high value. Reducing it by installing a better ground system would pay high dividend. Note added in 2002: Meanwhile I installed an extensive counterpoise net in the garden as earth terminal. This reduced the earth resistance to about 25 ohms, with an associated marked increase in available RF power.

Maximum power is transferred to the load, i.e. from the antenna to the crystal set connected to A-E, when we arrange for impedance matching and for resonance in the resultant antenna/earth series circuit. The set will presents, in general, an inductance which is too small for achieving resonance in the antenna circuit. Therefore, an additional coil must be inserted (Fig. 2). The resistance of about 400 ohms in the series-tuned circuit (200 ohms source resistance plus 200 ohms resistance of set when matched) yields a Q-factor of only 4. But this is still helpful regarding sensitivity, but also selectivity (sharpness of tuning), since the delivered current (voltage) is increased by this factor of 4, meaning 16 times in power.

3. Crystal Set with Parallel-Tuned Circuit

As a sort of "standard set" I used and investigated the popular arrangement shown in Fig. 2 which employs two tuned circuits. The inductance LC couples the antenna to the coil of the tuned circuit, the degree of the variable coupling chosen so that matching is achieved. The fixed L1 has a somewhat larger value than required for tuning the antenna/earth circuit of Fig. 1. The variable capacitor C1 is then used for tuning to resonance. The numbers for L1 and C1 apply to my particular case. C3 serves as AF storage capacitor for obtaining a maximum of AF amplitude at the phones, and it additionally provides a short for the RF. In practice, however, it often can be omitted without audible drop in AF. - To be able to properly match the diode detector to the parallel-tuned circuit, the detector branch is hooked up either to the top of coil L2 (wound with Litz wire onto a suitable ferrite rod) or to one of 11 taps provided on it. In this way the diode can be connected, via a switch, to 12 resistance values along the tuned circuit. Such a fine adjustment was required for the investigations reported in Sec. 5. The unloaded tuned circuit has a resonance resistance of 105 k ohms (at 1000 kHz), which drops to 52 k ohms when matched to the antenna/earth. (These are not particularly high resistance values because of the many leads from the taps to the switch.) According to the switch position chosen the diode can so be connected to 12 resistance values that vary between 52 k ohms and 100 ohms. When the diode is set to the tap that provides matching the total resonance

---

circuit at the left. AVC is essential when diode bias is used, to keep the bias voltage in the proper range. The audio signal, of course, rides on the bias connection. Diode bias is not suitable for high-mu triodes such as the 6AV6 or 6AT6, since they are very sensitive to the bias level. The 6R7 is a medium-mu triode ($\mu = 16$). The 6SR7 is a single-ended equivalent, and will also work in this circuit. The signal generator should be adjusted to peak the output, at 455 kHz. I obtained an audio output of 13 V peak-to-peak with a 30% modulated AM signal of 0.2 V peak-to-peak, for a gain of about 289. The tuned circuit provides a large part of this gain.

The 6AQ6, 6AT6 and 6BF6 are all miniature dual-diode triodes with the same basing as the 6AV6, the 6BF6 being medium-mu and more appropriate for diode bias and transformer coupling. The 6SQ7, 6SR7 and 6ST7 are similar octal tubes, again all with the same basing. This was a very popular and useful tube type, with many options available.

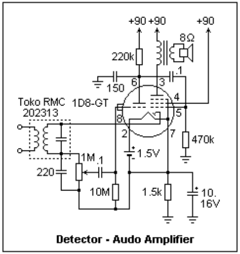

Detector - Audo Amplifier

The 1D8-GT diode-triode-pentode was designed to provide a complete audio system for a battery-powered radio. The circuit is shown at the right. The filament takes 0.1 A at 1.4 V, easily supplied by a D cell. The B+ supply is 90 V, for which batteries were available, consisting of 60 small cells. This circuit has a drain of less than 7 mA, which is quite acceptable, and hardly more than a transistor radio with the same audio output. The filament is hard to see the plate of the 6AT6. The RF ripple was negligible, but there was some 60Hz pickup, not surprising with a high-gain breadboarded circuit out in the open. It was interesting to show the AM signal and the audio output simultaneously on the oscilloscope.

The circuit will function without the tuned circuit, as you can easily demonstrate. The signal generator output must then be increased, because the resonance gain is considerable. The circuit will also function with both diode plates connected to one side of the secondary, and the 1M potentiometer and filter capacitor to the other side. In this case, the ripple frequency will be halved, but the output voltage will be doubled. There is not a great advantage in using a full-wave detector. Look at the DC voltage across the potentiometer as the input amplitude is changed. This voltage, negative with respect to ground, can be used for AVC, automatic volume control, as is discussed elsewhere. With the full-wave detector, it varied from -0.5V for 10V p-p audio output to -1.0V for 30V audio output. In practice, the AVC voltage is further filtered to remove the audio.

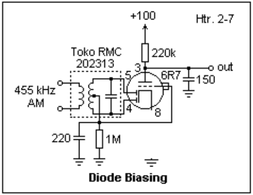

The rectified signal voltage is of the correct polarity to provide grid bias for the amplifier tube in this circuit. Therefore, the grid of the triode can be connected directly to some point on the diode load resistor that provides the proper bias, in place of the volume control potentiometer and coupling capacitor shown above. This arrangement is called *diode bias*, illustrated by the

---

resistance then drops to 26 k ohms. - On the right in Fig. 2 the equivalent circuit of the headphones is given which we require later on.

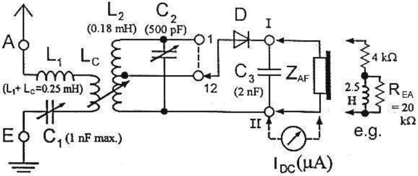

(Fig 2: Wiring diagram of crystal set with parallel-tuned circuit and tuned antenna/earth.)

We have now the task to match (1) on the RF side the diode branch, via the tuned circuit and the antenna coupling, to the antenna source, and (2), regarding the AF, the phones or the loudspeaker to the diode. Both procedures are strongly interrelated. But first we must find out more about the diode and its dynamic resistance.

4. Crystal-Diode Properties

All semiconductor diodes principally show a dependence of current I on applied voltage V (static characteristic) like I=IR,0*exp (V/nVT) - IR,0, where IR,O denotes the reverse saturation current at high negative voltages, VT the temperature voltage of 26 mV at room temperature, and n an ideality factor between 1 and 2.

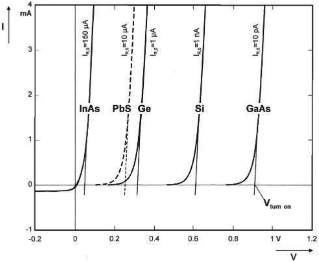

(Fig. 3: Theoretical current-voltage characteristics of various p-n diodes; equal diode areas assumed.)

Reverse breakdown is neglected. The relationship applies to both to p-n diodes and to Schottky diodes. However, the composition and physical meaning of IR,O differs completely for these two diode types (diffusion current resp. field current). Important is the fact that the exponential rise of the diode current occurs the faster, i.e. the turn-on voltage (as a kind of threshold) becomes the lower, the higher the saturation current IR,O is. Calculated characteristics for p-n diodes made from germanium, silicon, two modern compound semiconductors and from galena (PbS) are given in Fig. 3, assuming equal diode areas. The associated reverse currents are stated. The difference in diode behaviour is caused by differing electronic material properties (band gap). Real curves, in particular of the natural crystals, are flatter because of parasitic elements, mainly of diode series resistance. Regarding low turn-on voltage a galena p-n crystal diode is theoretically even slightly better than one made from germanium.

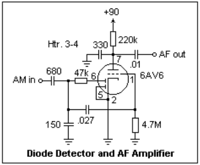

Diode Detector and AF Amplifier

A detector-amplifier circuit is shown at the left. This circuit is fed by a signal generator at AM in, and the audio output is taken from AF out. The amplifier uses grid-leak bias, which effectively clamps the peak of the audio wave at 0 V. The stronger the input signal, the more negative the grid bias, which makes the most of what is there. With an AM input of peak value 0.95 V and modulation amplitude 0.37 V, the peak to peak output was 3 V, for a gain of -8, which is quite good. I used a radio frequency of 2 MHz, and modulation of 1 kHz. At lower radio frequencies, the filtering of the output becomes progressively worse; this could be optimized if desired. This detector gives good fidelity with high sensitivity.

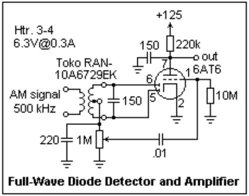

Full-Wave Diode Detector and Amplifier

A full-wave diode detector is shown at the right, using transformer input. The 6AT6 is similar to the 6AV6, but its mu is 70 instead of 100. The transformer is a Toko RAN-10A6729EK, available from Digi-Key. It is specified as 0.63 mH, and for 200 kHz. Adjust the signal carrier frequency until the tuned circuit resonates; in my case, this was 500 kHz. For an input signal from the signal generator of 0.18V peak-to-peak (modulation amplitude 0.04V), I got 26V peak-to-peak at

The typical superheterodyne receiver used a diode detector, followed by an audio voltage amplifier. These functions were usually combined in a single envelope, the diodes and the amplifying triode or pentode sharing the same cathode. A very good tube of this type was the 6AV6, whose structure is shown at the right. The diode plates are at the top of the envelope, with the cathode easily seen between them. The IF output was usually from an air-core, slug-tuned transformer, making it easy to use full-wave rectification. Only one of the diodes could be used for half-wave rectification, which was usually quite adequate. The triode is a high-mu triode with a maximum plate voltage of 300 V, and maximum plate dissipation of 0.5 W. The maximum cathode current is not specified, but should not exceed a few milliamperes. The tube has an unusually high plate resistance, and the amplification factor is over 100. It is a good triode for a voltage amplifier. My example gave g_m = 1.3 mS, r_p = 67k, μ = 90 at about 500 μA plate current, and V_g = -1.0 V. The transconductance varies strongly with plate current, reaching 2.2 mS at 1 mA, where μ = 116.

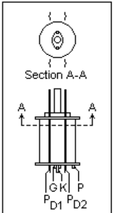

6AV6 Structure

For Schottky diodes principal curves like those in Fig. 3, relating to a particular crystal material, cannot be given. In their case the characteristics strongly depend on the kind of the metal electrode and on the processing parameters. But measured curves for Schottky as well as for p-n diodes are given in Fig. 4, including fairly good examples of galena (PbS) and iron pyrite (FeS) detectors; see also the Table below for identifying the diode types. Again one notices the relatively low turn-on voltage of the two natural crystals which equal or even are below that of the germanium diodes 1N34 and OA5, thus being well suited for low-level detection. (Fig. 4: Measured current-voltage characteristics of various semiconductor diodes (reverse currents indicated))

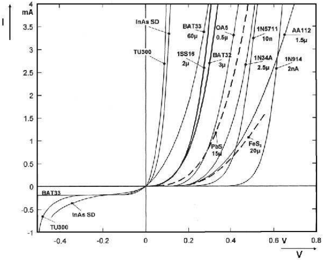

Very interesting is the performance of modern low-barrier Schottky diodes made from silicon, like the NEC 1SS16 (almost identical: 1SS99, BAT32, BAT63), which show turn-ons at 0.15 to 0.18 V. And, indeed, they show superb performance at low levels. One should expect that the InAs Schottky diode (which was specially made for my experiments) and the TU 300, a backward diode made by

Siemens, would according to the curves shown be even more sensitive detectors. But this is not the case.

As mentioned, a low turn-on voltage is inevitably associated with a high reverse current. This current reaches values of a few hundred µA for the InAs diode, as also for the TU300 and the Schottky diode BAT33. If the reverse current, i.e. an unwanted back current, reaches such high values we have strong counteracting effects, and ultimately the detector action disappears completely. Anticipating the results of computer simulations described in Sec. 7 one can state that diodes like the 1SS16 show the optimum relation between low turn-on and still acceptable reverse current, thus making them the best choice of presently available diodes as regards detector sensitivity.

To show and compare the capability of various diodes the Table summarizes values of measured AF voltages and of rectified currents, for 1 and 100 µW of available RF power. A power of 1 µW is in my set typical for DX stations at night, and 100 µW for stations 30 to 50 km away. As is seen the 1SS16 leads the field. - For 3 mW of RF (my local station) I obtained with a 1SS16 a DC current of 715 µA, which increased to 1.85 mA in the short-circuit case (AF/DC load = 0), and to 2.95 mA when under these conditions the set was retuned.

---

positive in a resistance-coupled amplifier, disturbing the bias. This is not a maximum in the same sense as the maximum plate dissipation or maximum cathode current are, since damage to the tube is not in question. There is no trouble in the grid leak detector, and values of several megohms are permissible.

There is no analogy to the grid-leak detector with FET's, which otherwise behave pretty much like vacuum tubes, because the leakage current is far too small for the desired behavior. As a gate-bias detector, however, the FET is suitable, and a BJT biased near cutoff makes an excellent detector, as we saw in the page on Amplitude Modulation.

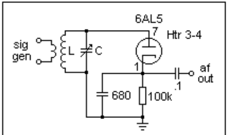

Diode Detector

A diode detector, using a standalone diode, is shown at the left. The output is the same amplitude as the modulation on the carrier that is input, very closely, so the detector could not do any better. Often, the diode was included in the first AF amplifier tube, sharing the same cathode, and dual diodes were also available for full-wave detection, which doubles the output. If you have a tuned circuit at this point, it would be easy to test a diode detector, and high voltages are not required, just the heater power. A solid-state diode would work the same way, but the 0.7 V drop would make it less sensitive, and it would not even respond to weaker signals. Using a Ge diode (1N34A) would help, but not eliminate, the problem. The thermionic diode is ideal in this application.

coupled by about 9 turns slipped over the loopstick, providing an AM signal about 30% modulated with 1 kHz, in the broadcast band. The output was roughly constant, no matter what the plate voltage (45 to 105 V), input amplitude, value of grid resistor (from 1M to 3M) or grid capacitor (100 pF to 220 pF). The grid bias with no signal was about -0.5 V with a 1M grid resistor.

The circuit works by using the small grid current when the grid is slightly negative. This grid current increases rapidly and very nonlinearly, so square-law detection is possible. The grid resistor establishes a quiescent point around which the signal oscillates. The rf part is bypassed through the grid capacitor, while the af part remains and causes a voltage drop across the grid resistor, which is then amplified in the usual way. With a 100k plate resistor, the output af voltage could approach 1 V peak-to-peak. Another stage of audio amplification would give a quite satisfactory result, but I was hoping for one tube. (The 12AX7 has two triodes in the same envelope--if such a tube were used, the receiver would be, strictly speaking, single tube.) With most other kinds of detectors, such a hope would be quite vain, it must be admitted.

If the input signal amplitude is increased, the grid bias starts to decrease because the grid becomes positive on the peaks, and the detection becomes a little more linear. If the grid is biased to cutoff with a C supply, then there is clear linear detection, since the signal is effectively rectified. Such a circuit is called a grid-bias detector, and the signal amplitude must be restricted so that the grid does not go positive, when distortion would result.

You may notice in the tube manuals that a "maximum grid resistance" is given for a tube; for the 6J5, it is 1M. The reason for this is positive ion current, which may drive the grid more

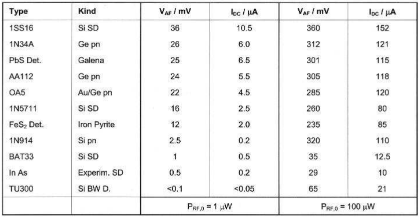

Type	Kind	V_{AF} / mV	I_{DC} / µA	V_{AF} / mV	I_{DC} / µA
1SS16	Si SD	36	10.5	360	152
1N34A	Ge pn	26	6.0	312	121
PbS Det.	Galena	25	6.5	301	115
AA112	Ge pn	24	5.5	305	118
OA5	Au/Ge pn	22	4.5	285	120
1N5711	Si SD	16	2.5	260	80
FeS_2 Det.	Iron Pyrite	12	2.0	235	85
1N914	Si pn	2.5	0.2	320	110
BAT33	Si SD	1	0.5	35	12.5
In As	Experim. SD	0.5	0.2	29	10
TU300	Si BW D.	<0.1	<0.05	65	21
		$P_{RF,0} = 1\ \mu W$		$P_{RF,0} = 100\ \mu W$	

(Table: Measured values of AF voltage (across phones of 4 k at DC) and of rectified DC current for various diodes and two levels of RF power)

Sometimes a DC bias from a battery is applied for shifting the operating point of the diode closer to the turn-on voltage and so improving the detection efficiency. By this method the AF voltage obtained can be increased, for example when using the 1N5711 and the 1N914 at low RF levels. The 1SS16 group of diodes, however, hardly gains from a DC bias. Only at RF powers below about 200 nW I was able to measure a certain rise in AF voltage. At the lowest detectable RF level of 50 nW (Sec. 1), the AF voltage increased by 20 percent (i.e. power by 45 percent) when the optimum bias was applied. But this effect was measurable only, being still too small to be noticed by the ear.

Note added in Jan. 2002: Backward diodes (BWD), like the TU300, are good detectors at extremely low RF signal levels, below about 1 nW with associated voltages of only a few mV. This is due to the relatively sharp bend in the BWD characteristic at zero volts. The generated AF signal is, however, too small for operating phones directly and calls for an AF amplifier. Then stations can be copied which are not heard when in such a set-up with AF amplifier a "normal" sensitive diode, like the 1SS16, is used instead of the BWD.

5. RF Diode Resistances and RF Matching

For achieving best performance it is required to RF match the diode to the tuned circuit. The dynamic resistance of the diode depends on the amplitude of the RF voltage applied to it, and on the kind of AF load impedance.

In AM tube radios the detector diode operates at a high level (linear detection) and has a load consisting of a large (ohmic) resistor shunted by a small capacitor. Calculations show that in this case the RF diode resistance, as presented to the tuned circuit, is roughly half of the ohmic load resistance. In a crystal set the calculation is somewhat more complicated since there the RF voltage on the diode is generally lower and the diode load is more complex (see equivalent circuit of phones in Fig. 2). Hence I preferred to measure the RF diode resistance RD. The measurements were carried out under actual working conditions using a signal generator. Figs. 5(a) and (b) show the results obtained for high-impedance phones with 4 k ohms DC resistance and for low-impedance ones with 120 ohms, respectively. The RF frequency used in these measurements was 1000 kHz, the modulation frequency 1 kHz with a modulation factor of 0.4 (given by the signal generator). Figs. 5 give the measured diode resistances, as a function of the RF power applied, for an 1SS16 (also some in parallel), a silicon p-n diode 1N914, and for natural galena as well as carborundum (silicon carbide; SiC) crystals. The diode circuit was in turn connected to the various taps on the coil L2 .When the RF voltage measured across the tuned circuit dropped by a factor of radicle (2)=1.41 compared to its value without diode, matching was achieved. Then the RF diode resistance equalled the RF resistance of the tuned circuit at the tap point. To avoid

---

Unfortunately, I do not have any actual circuits that used a 1V6, but the amplifier shown at the left may be studied. It gave a gain of -14 with a plate current of 357 µA, so the effective transconductance was 298 µS. I substituted a 100k plate resistor, but the gain only increased marginally, to about -17. By experimenting, you may be able to get more gain from the tube. At low plate currents, the amplifier oscillated in an odd manner, and there were other peculiarities seen while fiddling around, but I did not take the time to track them down.

The Grid-Leak and Diode Detectors

Detectors are discussed in Amplitude Modulation, where the two types of square-law and linear are explained, and the grid-leak detector is mentioned. A circuit using a 6J5 triode is shown at the right. The audio output transformer T can be replaced with a 10k or 100k resistor, and the output observed with an oscilloscope. I could find only a 1k to 8Ω transformer, really designed for use with transistors. The circuit works, but I was not impressed by its output. Asking for loudspeaker output from a detector is being very optimistic, but I had some hopes. Perhaps a more suitable transformer would give better results.

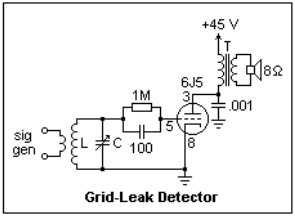

Grid-Leak Detector

The tuned circuit can be one from a crystal set, with a loopstick and variable capacitor. The signal generator is

These tubes make vacuum-tube experiments very easy without requiring high-voltage supplies, only what is found in a normal transistor lab.

Sub-Miniature Tubes

The smallest regular electron tubes were the *subminiature* tubes, designed for battery-powered portable apparatus, from hearing aids to radiosonde transmitters. Typically, the lead wires were brought out through the press seal without a base, but there was also a subminiature base used for a few tubes. As an example, you can work with the 1V6, a pentode-triode with a filament taking 40 mA at 1.25V (the filament supply can be a D cell). The tube is 40mm x 10mm x 7mm, approximately. The maximum plate voltage is 45V, and typical plate currents are less than 1 mA.

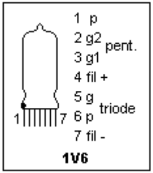

A sketch of the 1V6 and its connections is shown at the right. On the press seal, there is a red dot that identifies what I have called pin 1. The filament is located at the center of the tube, with the pentode on one side and the triode on the other. The triode shows a μ of about 10, r_p = 18.6 kΩ, and g_m = 440 μS. The pentode has g_m as high as 694 μS above 500 μA plate current, but it drops considerably at lower plate currents. The two devices can be tested as usual, taking care not to exceed 45V or 1 mA. Note that the pentode and the triode share the same filamentary cathode.

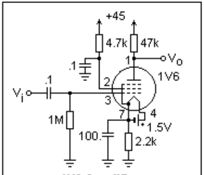

1V6 Amplifier

---

an error one must readjust the coupling to the antenna when the diode is connected to the first found (V/1.4) tap point and then repeat the search for the now somewhat altered (V/1.4) tap. A second iteration further improves the result, but not much.

As in principle to be expected from the characteristics, the diode resistances vary rather widely, from some 100 ohms to some 10 k ohms, with lower values obtained when the DC resistance of the phones is low. The galena detector shows values only moderately higher than those of a single 1SS16. The silicon diode 1N914 presents high values due to its high turn-on, which even more applies to carborundum.

(Fig. 5:Measured RF diode resistances versus available RF power: (a) for high-impedance phones (4 k ohms at DC), (b) for low-impedance phones (120 ohms at DC)

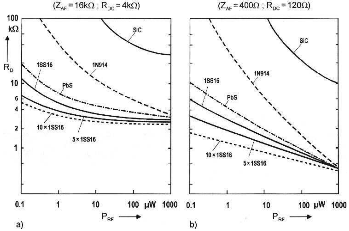

The data obtained then indicate that the optimum tap position on coil L2 (for matching) depends on the diode type, the strength of the received station, and on the DC resistance of the phones. The larger the value of the diode resistance is, the higher must the tap position be up the coil. Sometimes it was suggested in the literature to have a fixed tap at a point of about 1/4 to 1/3 of the windings counting from the earth point. In the present case the tuned circuit has a resistance of approximately 6 k at the 1/3 tap point. As Fig. 5a shows, this indeed is a rather good choice for a galena detector when high-impedance phones are used and weak stations received. Impedance matching requires that the reactances of source and load cancel out. But in our case the resistance of the tuned circuit has no reactive part at resonance, and the reactance of the diode, caused mainly by the diode junction capacitance of at most a few pF, can be neglected.

Connected to a particular tap, the diode resistance is (auto-)transformed up and appears in parallel to the resonance resistance of the tuned circuit. This means that not all of the available RF power reaches the diode since a reasonable fraction of it is dissipated in the resistance of the tuned circuit. In order to really transfer the maximum of power from the antenna to the diode branch, the diode (of generally low resistance compared to that of the tuned circuit) should be connected untapped to the top of the coil L2. This, however, strongly reduces the selectivity of the set and requires a readjustment of the coupling of the tuned circuit to the antenna. With high incident high RF power (and/or low impedance of the phones) the tuned circuit can, under these conditions, become loaded to such an extend that variations of the capacitor C2 have no tuning effect any longer, which means that C2 is obsolete and can be omitted. The diode circuit is then aperiodically coupled to the (tuned) antenna circuit, while the coil L2 merely acts as the secondary winding of the transformer which matches the diode to the antenna.

6. AF Matching

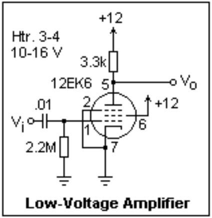

Low-Voltage Amplifier

The 12EK6 (also known as 12DZ6 and 12EA6) is a screen-grid pentode for low-voltage use. A voltage amplifier using this tube is shown at the left. This tube was generally used as an IF or RF amplifier, so the load would have been a resonant circuit, not a resistor, giving much higher gain. However, the resistive load is easy to experiment with. There is no voltage to waste for cathode bias, so grid-leak bias is used. When a signal is applied, the bias will adjust itself so that the top of the input waveform is clamped to 0 V. Vary the input amplitude to verify this. This circuit works with an input of up to about 2 V peak-to-peak, perhaps best for an input of around 1.0 V p-p. The output voltage is then about 5.8 V p-p, the plate potential is 6.0 V, plate current 1.8 mA, and gain about -5.5, corresponding to a transconductance of 1.8 mS, about right from measurements of the characteristics and the position of the operating point in this case. The transfer characteristic is rather curved, leading to considerable distortion for larger amplitudes. Nevertheless, it is remarkable that the circuit works so well at such a low voltage.

As a triode (plate, suppressor and screen connected together), the 12EK6 has $\mu = 8.5$, $g_m = 6.3$ mS and $r_p = 1.35$ kΩ. The transconductance varies considerably with plate current, which gives rise to distortion for large signal amplitudes.

current, which was stepped up by a transformer and then rectified to DC again (sometimes by contacts on the vibrator) for the B supply. Vibrators were electrically very noisy, and had short lives, but were widely used until solid-state equivalents became available. They came in cylindrical metal cans, and looked like electrolytic capacitors. The 0Z4, mentioned elsewhere, was a rectifer specifically for vibrator supplies. Another possibility was the *dynamotor*, a motor-generator with a single rotating armature and brushes, which could supply more power than a vibrator and was more reliable. However, they were too expensive for general commercial use, though widely used in the military.

When 12 V became the automotive standard, vibrator supplies or their solid-state equivalents, could still be used. Tubes with 12.6 V heaters were already common. Many had the heater center-tapped, so they could be used equally well on 6.3 V. However, 12 V is high enough to be used directly as the plate supply if the tubes are specially designed. A range of tubes was produced that could be used on 12 V only, without any high voltage at all. We have already discussed the 12DL8 and 12K5 space-charge-grid tetrodes, and how the space-charge grid allows larger plate currents for small plate voltages. There were approximately 15 tube types designed for 12V use, of which only three, the 12DL8, 12DS7 and 12K5 are space-charge grid tetrodes, behaving like triodes. The other tubes are of conventional construction, in general having low plate currents, less than 1 mA in many cases. These include twin-diode triodes (12AE6, 12FK6, 12AJ6), sharp cutoff pentodes (12AF6, 12BL6, 12CX6, 12EK6), remote cutoff pentodes (12CN5, 12DZ6), twin-diode remote cutoff pentode (12F8), pentagrid converter (12AD6), pentagrid (variable gain) amplifier 12EG6, and a twin-diode power screen-grid tetrode, 12J8. The last tube can deliver 20 mW output power, half of what the 12DL8 can do.

If a crystal ear phone is used or the diode detector is followed by an amplifier (generally of high input impedance) one has to design for maximum voltage at the detector output. Here, we rather have to deliver a maximum of power to the phones. Hence the impedance of the phones (or the speaker) as the AF load should have such a value that a maximum of AF power is transferred to it. The AF source resistance RG is at low RF levels (square-law detection) approximately given by the reciprocal of the slope of the diode characteristic at the operating point. At higher RF levels (linear peak detection) it is determined by the current spikes flowing through the diode. In so far, RG nearly equals the diode resistances as shown in Figs. 5. The tuned circuit presents an AF short.

I determined the equivalent circuit of a pair of high-impedance Telefunken phones (4 k ohms at DC) at 1 kHz by using a measuring bridge and obtained the quantities given in Fig. 2. REA is caused by the electro-acoustical transducing process. The AF source has to provide the real power for REA as well as, necessarily, for the DC coil resistance, and foremost the reactive power for the phone coils (2.5 H) that are to move the membranes. In order to obtain the maximum of power transfer the magnitude (amount) of the overall phones´ impedance ZAF (16 k ohms for my phones) must match the AF source. Again I preferred to experimentally find the optimum AF load: I connected in turn 14 phones and speakers of different impedance to the set, partly connecting two of them in series or parallel, which in total provided 20 load impedance values between 80 ohms and 75 k ohms in magnitude. From the AF voltage measured across these load impedances I determined the AF power. The coupling to the RF signal generator was readjusted to retain RF matching each time the AF load was changed. Fig. 6 shows the obtained results when using a) a diode 1SS16 at low RF power (1 µW) and b) with a 1N914 at higher power (1 mW). The optimum AF load impedance turned out to be, resp., 1.2 and 3 k ohms. In order to simplify matters the diode was in this experiment fixed to the 1/3 tap at the tuned circuit. This meant a compromise as regards matching and generally did not produce quite the maximum of achievable AF power. The dashed curves of higher AF power in Fig. 6 were obtained when the diode was connected to the top of the tuning coil (as discussed above).

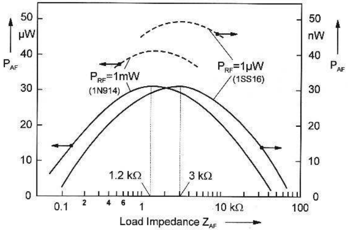

(Fig. 6: measured AF power versus AF load impedance.)

In order to present my measurement results in a more general form, Fig. 7 shows the AF power obtained as a function of the AF impedance now divided by the respective occurring source (= diode) resistance. The curves indicate that the maximum is reached when ZAF has a value of 50 to 70 percent of the diode resistance. The simplifications introduced above, like choosing the fixed 1/3 tap, are probably the reason for not reaching a higher percentage. But we can say to be roughly correct with our predictions. - Sometimes it is suggested to match just available phones (speaker) to the diode by using a suitable transformer. I found this only helpful if the mismatch was extremely high. In the other cases the winding and iron losses of the transformer, as well as the inductive shunt, tend to

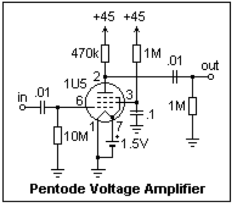

Pentode Voltage Amplifier

A pentode voltage amplifier with a gain of about -33 is shown at the right, using a 1U5, a sharp-cutoff pentode which comes in a 7-pin miniature package. Note the polarity of the filament supply, which is important, since it supplies a little grid bias as well. Since the voltages and currents are low, you can use the same 1/4W resistors used with transistors. The capacitors must have an appropriate voltage rating, of course. The 10M resistor is a "grid leak" that will prevent the grid from going positive. It charges on the positive excursions of the grid to provide whatever bias is necessary. The plate current was 61 μA, and the screen grid current was 20 μA. The screen grid potential is above the plate's in this circuit. Pentodes offer the capacity for controlling the plate current through the screen voltage, so that the plate voltage can be what is required. For higher voltages, the screen grid is normally at the same or lower potential as the plate, on the average. The grid voltage cannot be measured accurately with a DMM or scope, since the bias conditions will be disturbed. However, you will find that the amplifier works very well. Measure the gain the usual way, with a function generator and oscilloscope.

Radios for motor cars had 6 V DC available, which was right for 6.3 V heaters, but could not be used for the B supply--even now, 6 V is inconveniently low for transistors. The usual solution was a *vibrator* supply. Contacts on the magnetically-driven vibrator armature converted 6 V DC to alternating

The 1LH4 is a loktal battery triode, which is shown in a voltage amplifier at the left. A similar octal tube is the 1H5-GT, and either will do in this experiment (but the basing is different!). A diode plate is included, connected to pin 4, and using the negative side of the filament as its cathode. The diode is not used in this circuit. The maximum plate voltage is 110V, but this can be exceeded a little in testing. I measured the characteristics for plate voltages of 125V and less, at grid voltages of 0.0, -0.5 and -1.0V (relative to the negative end of the filament). The plate current was always less than 1 mA. My results were $\mu = 62$, $g_m = 340$ μS, and $r_p = 182$kΩ. These values are close to the published values. The characteristics are noticeably curved in this region, so the parameters will vary with plate current. The apparently low value of transconductance is quite reasonable for the small plate currents at which the tube is used.

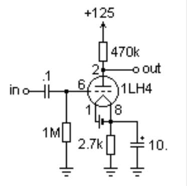

Voltage Amplifier

The amplifier shown gave a voltage gain of G - -40, with an input of 0.4V peak-to-peak. The plate current was about 100 µA, which made the plate voltage 77V and the grid bias -0.26V. The bandwidth of the amplifier was quite good, roughly from 10Hz to 50kHz. It should be noted that the total power drain of this circuit is no more than 83 mW, of which most is the filament power. This is very economical for a normal-sized tube. The tube remains quite cool in service, incidentally.

---

dissipate more AF power than is gained by providing the right transformation ratio. One also has to consider that the human ear cannot register small changes in acoustical power. Alterations like those shown above the dashed line in Fig. 7 will hardly be noticed by the ear. Thus, it seems that the exact value of the AF load impedance is not of paramount importance as regards noticeable output power. But the general principle holds that a number of small improvements in matching, each of which will not produce any audible effect for itself, might in sum indeed be noticed by the ear.

So it turns out as an interesting and important feature that a high-impedance AF load, which is associated with a high DC resistance, will produce a high diode resistance (= AF source resistance), and vice versa. This means that the circuit has a self-optimizing tendency towards the matched condition. Regarding RF selectivity of the set, as another important quantity, a high AF impedance - leading to a high diode resistance - is of advantage. But the influence of the AF impedance in this respect is not particularly pronounced. I measured an increase in -3db RF bandwidth by a factor of 2.5 when the AF load was decreased from 100 k ohms. At 720 kHz (local station) this bandwidth was 20 kHz in my set, which yields a total loaded Q factor of 36 - leaving room for improvement (see Secs. 2 & 3).

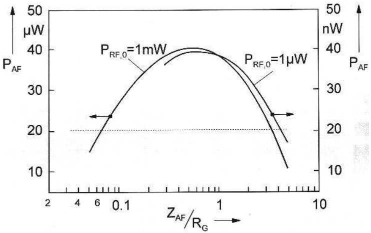

(Fig. 7: Measured AF power versus AF load impedance normalized to AF source resistance.)

7. Computer Simulation of Crystal-Set Behaviour

As a summarising investigation I simulated the overall performance of the circuit shown in Fig. 2, using the analysis program SPICE. An RF frequency of 1 MHz, an AF of 1 kHz, and a modulation factor of m=0.4 were assumed. The tap on the coil L2 was held fixed at 1/3 of the windings from earth. The I(V) equation given in Sec. 4 served to describe the diode behaviour.

---

because of space-charge effects. When you measure the V-I characteristic of a diode, there appears to be a small positive plate voltage when the measured voltage is zero, and a little current flows. The reason is that the electrons are emitted with some kinetic energy, and a negative space charge is created around the cathode. The actual emission to the anode comes from the minimum of this potential well, which is several volts negative, and so the plate may be slightly more positive than this minimum, in spite of contact potentials.

Battery Tubes

In the early days, nearly all radios (and other electronic apparatus) were powered by batteries, usually primary cells. With the rise of central power distribution, apparatus could be powered (and storage batteries charged) from the AC line, which was a practically unlimited, cheap source so far as radios were concerned. Transformers made high-voltage power supplies easy to build. Low-voltage filaments were largely replaced by 6.3V heaters and unipotential cathodes since power was now cheap. However, there still remained a demand for portable radios beyond the reach of power lines, and these were necessarily powered from batteries. The emphasis here had to be on small power drain, since battery power is expensive. Typical batteries were 1.5V and 3.0 V for filaments ("A" batteries), and 45V or 90V for plate supplies ("B" batteries), both the usual Leclanche "dry" cells. Filaments were generally used, because of their much greater efficiency in mA per watt. Typical battery tubes had filaments taking only 50 mA at 1.4V, only 70 mW. Plate currents were not high, since a few mA would be quite sufficient. For experiments with battery tubes, a D cell in a holder is an adequate filament supply, good for about 100 hours of service. Even an AA cell will do for experiments.

What happens in a vacuum tube is illustrated in the figure below. Suppose metal 1 is the cathode and metal 2 is the anode or grid, and the connections are made by a third metal 0 (there could be several such metals, but they all will act like metal 0). At the left, metals 1 and 2 are connected by a wire of metal 0. If metal 0 includes a DMM, it would show 0 V and, of course, no current would flow. However, the Fermi levels of all three metals must line up for equilibrium, making a level surface on which the electrons will not tend to roll one way or the other. It is clear that in this state of equilibrium, the surfaces of metals 1 and 2 will not be at the same electrostatic potential. An oxide-coated cathode has a work function of about 1.0 V, while a nickel plate or grid has a work function of about 5.0 V. When the Fermi levels align, the cathode is 4.0 V positive with respect to the other electrode.

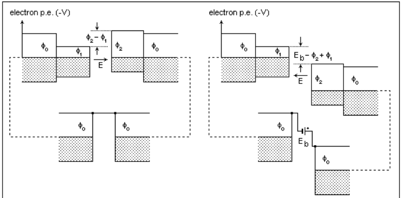

Effects of Contact Potential

At the right, we have connected a source of emf E_b into conductor 0. A source of emf maintains a difference in the Fermi levels equal to its voltage. With the polarity shown, the cathode 1 has now become negative with respect to the anode 2, by an amount equal to the applied voltage less the difference in work functions. If the device is a thermionic diode, current can now flow from cathode to anode. The actual plate voltage is about 4 V less than the meter says, because of the contact potential. In practice, things are a bit more complicated,

---

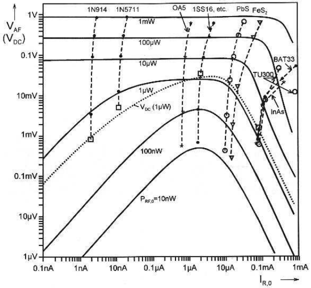

(Fig. 8: Simulated AF voltage across high impedance phones reverse saturation current of diode with available RF power as parameter (m=0.4). Measured values for various diodes included.)

The simulation results are presented in Fig. 8, which shows the AF voltage VAF AF obtained across phones of 16 k ohms AF impedance (4 k ohms DC) as a function of the diode reverse saturation current IR,O. Parameter is the available RF power at the tuned circuit, ranging from 10 nW (lowest sensible level; see Sec. 1) to 1 mW (about local-station level). Measured values pertaining to various diodes are entered for comparison. These measurement values lie in part slightly below, partly somewhat above the simulated curves, but agree in general. In practical diodes, particularly Schottky diodes, the effective reverse current increases noticeably with reverse voltage (i.e. with increased RF power). This had to be considered when inserting the measured values in Fig. 8, and accounts for the slight bending of the vertical lines to the right. The dotted curve in Fig. 8 shows the simulated DC voltage across the phones for 1 µW of RF power.

The simulated curves for VAF drop at the right side for diodes with high IR,O because of the adverse effect of reverse current as mentioned in Sec. 4. The decrease at the left of Fig. 8 for low RF powers results from the high turn-on voltage shown by diodes having a low IR,O. In these cases the associated smaller RF voltages increasingly fail to reach the turn-on of the diodes. The curves make clear how important it is to choose a diode which has the right value of reverse current. A definite maximum in sensitivity, especially pronounced at low RF powers, is found for diodes having a reverse saturation current of a few µA. Particularly the 1SS16 diode class is in this range, but also the OA5 and 1N34 perform not too badly, and good specimen of galena (PbS) crystals behave still satisfactorily. Hence this result is quite in agreement with what we already have found in Sec. 4. - When the AF voltages in Fig. 8 are used for calculating the AF power, the AF/RF detection efficiency can be worked out. It is found that the efficiency drops drastically for low RF levels, with one per mille being reached at 10 nW of RF. This agrees with the observations described in Sec. 1.

In the simulation the reverse breakdown voltage, at which in practical diodes the current starts to rise rapidly, was not included. For 1 mW, the highest RF power considered, the diode resistances have dropped to around 3 k ohms (Fig. 5a), calling for a low tap position on the coil. There the RF voltage is relatively low. My local station, with 3 mW of RF, produces 2.4 V (i.e. a peak-to-peak value of 6.7 V) at the required tap point, so that the negative peak only just reaches the breakdown voltage of -6 V for a 1SS16. In consequence, reverse breakdown seems not to be a particular limiting factor, even if the voltage across the tuned circuit might be somewhat

---

you measure with a DMM are not the actual voltages between electrodes, as seen by the free electrons, but differ by up to a few volts from the true potential differences.

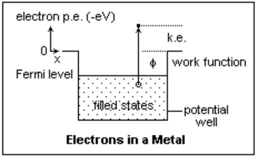

Electrons in a Metal

To understand this, let's review electrons in metals. One or more of the outer electrons of each metal atom is free to wander in the field of the ion cores, effectively binding the metal together. The positive ions create a potential well, as shown in the figure, and the electrons occupy the states of lowest energy. Note that electron energy increases upwards; since electrons are negatively charged, this means that voltage gets more negative upwards. There are not many more states than electrons, so they fill up the states up to some energy called the Fermi energy at 0 K. At ordinary temperatures, there is little difference, since the thermal energy is small compared to the kinetic energy of the electrons near the Fermi level.

If an electron is given enough energy, by thermal agitation or absorption of light, it may be separated completely from the metal and wander freely. The amount of energy necessary to just get outside is called the *work function*. Anything more goes into kinetic energy of the electron. The free electron can now be accelerated or decelerated by electric fields outside the metal in space. Work functions measured by thermionic, photoelectric and contact potential measurements agree roughly, but they depend sensitively on the surface preparation.

increases very rapidly beyond this point. The filament glows brilliantly, like an incandescent lamp, since its operating temperature is about 2400K, not the 900K of an oxide-coated filament. The filament current should not be allowed to exceed 1.6 A. If the power supply has current limiting, it can be useful here. By setting the plate voltage at near 200 V, you can see the saturation current as a function of filament current.

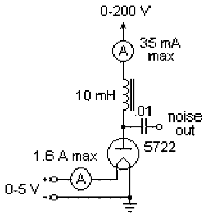

Testing the 5722

For two or more reasonable values of the saturation current, say 5 mA, 12 mA and 20 mA, record the current as a function of plate voltage and plot your results. For I_f = 1.5 A, the plate current saturated for about 50 V on the plate, approaching a value of about 12 mA. It is easy to find out what plate voltage to use to ensure saturation when making shot noise in this way. It is very difficult to make noise measurements in the usual breadboarding environment. I thought it just possible to have seen some on my 100 MHz scope with a plate current of 20 mA, without amplification. See the page on Noise for more discussion of noise measurements.

Contact Potential Effects

This is a good place to mention contact potentials, since you will probably run into their effects when experimenting with low-voltage tubes. Because of contact potentials, the voltages

---

higher in case there is a better Q factor. Assuming a constant available RF power, the RF voltage is proportional to the square root of the Q (resonance resistance). The low-barrier silicon Schottky diodes, which show a reverse breakdown in the range of -5 to -8 V, are thus well suited for use from the lowest to the highest levels of RF power generally occurring in crystal sets.

8. Some Remarks on Series-Tuned Circuits

Historically the first crystal sets, in the pre-broadcasting days, were of the kind shown in Fig. 9a. There the values of L1 and C2 pertain to my particular antenna/earth situation. For maximum power transfer in the circuit of Fig. 9a the combination of crystal plus load in parallel should match the impedance of the antenna source. In the latter the earth resistance represents the main resistive part which in the then primarily commercial stations had values of only 10 to 50 ohms. On the other hand, the crystal-diode resistances were around a few k ohms so that a considerable mismatch existed. For this reason one soon changed to the arrangement of Fig. 9b with diode and load now in series, and with the possibility to match the diode branch to the antenna by choosing the right tap on the coil L1. To increase the selectivity of the set a second tuned circuit was eventually introduced, as e.g. shown by Fig. 2.

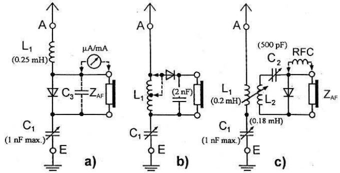

(Fig 9: Series-tuned circuits: a) Diode directly in tuned antenna circuit, b) Diode across tuning inductance (preferably tapped), c) Diode in separate series-tuned circuit, coupled to the series-tuned antenna/earth circuit.)

Using modern low turn-on diodes (1SS16 etc., Sec. 4) and having in general a higher earth resistance than in commercial stations, the circuit of Fig. 9a is however quite effective. Possibly paralleling of diodes is of advantage, depending on the actual source and load resistances. With two 1SS16 and employing a moving-coil speaker via a suitable transformer as the load I obtained an AF power of 180 µW from my local station. With ten 1SS16 and two moving coils of 16- ohms speakers in series as load, the obtained AF power of 210 µW approached the maximum possible after Sec. 1. Ideally the diode resistance should, in my case, about equal the 210 ohms of the antenna source (Fig. 1). Reverse diode current is not harmful, nor a possibly low reverse breakdown voltage. We have here a “current-controlled” case where voltages across the diode remain low with associated high currents, a few tens of a mV and some mA when I used the ten diodes. In contrast, voltage control is - more or less - experienced when parallel-tuned circuits are employed where high(er) voltages and low(er) currents exist. A set according to Fig. 9a, then, is a most simple hook-up for effectively receiving the nearest station. Substantially higher selectivity, approaching that of the set of Fig. 2, is offered by the arrangement with two series-tuned circuits shown in Fig. 9c. Since there any resistance (loss) in the circuit made up by L2 and C2 should be kept low for achieving a high Q factor, the diode(s) - preferably paralleled again - and the phones/speaker should be of low impedance. The RF choke might help to improve performance. The diodes found in the left of Fig. 4, particularly the backward diode TU300 (which is of little use on parallel-tuned circuits), operate excellently in the arrangement of Fig. 9c.

---

some surprise, that this correlated successive electrons so that they were emitted regularly to maintain a constant current, and therefore the shot effect was nearly completely eliminated. That is, a normal diode has no shot effect noise in its plate current.

The noise diode is designed so that at reasonable plate voltages, all electrons emitted by the filament are immediately drawn to the plate without forming much of a space charge. Since the electrons are emitted randomly, the anode current will show the full shot effect noise. This is done by purposely making the filament to have low emission. To do this, a tungsten filament is used. Noise diodes give us the opportunity to observe a tungsten filament, as well as temperature saturation.

An available noise diode is the 5722, whose basing is shown at the right. The 7-pin miniature tube was made as late as 1977, and now costs about $14, which is probably not much more than when it was new. The maximum plate voltage is given as 200 V, and the maximum plate current as 35 mA, so apparently the plate can dissipate 7 W. The plate has wings that make a good dissipation probable.

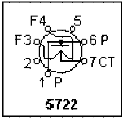

A circuit for testing the 5722 is shown at the left. Note that an RF choke is put in the plate lead to act as a load for the current fluctuations. This choke should be rated for the plate current employed. I connected a variable DC supply to the filament as shown, to pins 3 and 4, leaving the center tap alone. This supply should be rated at 2 A or more. Increase the filament voltage gradually, looking for the glow. There will be no plate current until the filament current reaches about 1.3 A, but it

As an example of the small signal diodes that are often combined with a triode or pentode in the same envelope, and share the same cathode, the 6AV6 or 6AT6 furnish good examples. The 6AV6 has its heater at pins 3-4, cathode at pin 2, and the signal diode plates at pins 5 and 6. The maximum current for each diode is 1 mA. I connected the two plates together for measurement, and took the current up to 3 mA, for which a plate voltage of 6.4 V was required. The curve of I against $V^{1.5}$ sagged a little at low currents, but the upper part was quite linear, showing a perveance of 0.085 mA/$V^{1.5}$ for one plate. The incremental resistance was 4.55kΩ, and V/I was 5.05kΩ at 1 mA. The current for one plate obeyed the formula $I = 0.15 + 0.085V^{1.5}$ mA. In this tube (and similar ones) the plates are flat, one on each side of the cathode.

The 1A3 seems to be the smallest signal diode of all. It was designed for portable measuring apparatus. The heater takes 0.15A at 1.4V (a D cell), connected to pins 1 and 7 of the 7-pin miniature envelope. The cathode is at pin 3, the anode at pins 2 and 6. The peak inverse voltage is 330V max., the maximum plate current 5 mA, and the average plate current 0.5 mA DC. Maximum heater-cathode potential is 140V. The anode is only a few millimeters high; most of the envelope contains only vacuum. The measured perveance was 0.075 mA/$V^{1.5}$.

The Noise Diode

A special kind of diode should be mentioned here, because experiments with it are quite interesting. It is the *noise diode*, intended for the specific purpose of producing wide-band RF noise through the *shot effect*. Shot effect noise is fluctuations in the anode current due to the random collection of electrons. We have already mentioned that the anode current is controlled by the space charge around the filament. It was discovered, to

Modern 8/16- ohms headphones or, for stronger stations, directly the moving coil of a speaker are effective AF loads. - Backward diodes, being scarce these days, are tunnel diodes which have the typical current hump reduced to a flat region at about 200 µA height. The I(V) curve of the TU300 shown in Fig. 4 has in reality reversed polarity. For reasons of comparison with the other diodes the polarity was changed in the graph. Other BWD types are: AEY17 /29, 1N3539 /3543, TU1B.

In conclusion, the investigations sketched here have certainly enlarged my knowledge on crystal-set design, with the identification of the “best diode” and noticing the tendency of self-optimization which makes the set a sort of good-natured device. Other rewarding topics could not be covered, as there are, for example, more complex circuits for increased selectivity and for DX. Also short-wave crystal sets are fascinating since they provide DX from all over the world with simple designs.

Based on:

B. Bosch and M. Bussmann: Zur Empfindlichkeit von Kristallgleichrichtern und Halbleiter- dioden beim Detektorempfang.Funkgeschichte Nr. 93 (1993), pp. 275 - 285.
B. Bosch: Anpassungs- und Schaltungsfragen beim Detektorempfang.Funkgeschichte Nr.98 (1994), pp. 211 - 225.

Loktal equivalent to the types 80 or 5Y3 that are now much more expensive.

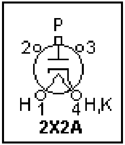

An excellent diode for observing the Langmuir-Child law is the 2X2A. This tube has a 4-pin base like the 82 phanotron discussed below, and the large, bell-like anode is brought out to a cap at the top of the ST envelope. The oxide-coated cathode thimble is easily seen. The heater takes 2.5V at 1.75A, so it can use the same transformer as the type 82. The rated DC current is 7.5 mA, and the maximum voltage is 4500V. A plate voltage of about 60V is needed to reach 7.5 mA plate current, so measurements can be made over a wide range of voltages. Plot your results as $I^{2/3}$ vs. V. A straight line will be found, that intercepts the V axis at -1.2V. The perveance of the 2X2 is found to be 0.0165 $mA/V^{3/2}$. The unusually low value is due to the large cathode-anode spacing.

The 6V3-A is a strange miniature tube with a cap on top that is the cathode connection. Its heater, connected to pins 4 and 5 of the 9-pin miniature base, takes 1.75 A at 6.3 V. The plate is connected to pins 2, 7 and 9. This tube is designed for the rugged service of a television *damper diode*. During horizontal retrace, the damper diode conducts, charging the boost capacitor while absorbing the large inductive kick. The peak inverse voltage is 6000 V, the peak current 800 mA, and the average current 135 mA. The large-diameter cathode tube and long plate imply a large perveance, which, in fact, is about 2.3 $mA/V^{1.5}$. This tube happens to be very cheap, but would serve as an excellent half-wave rectifier for practically any purpose. There are other damper diodes, such as the 6W4 and the 6AX4GT (perveance 1.42), that would have similar characteristics.

The 6H6 is an octal dual signal diode like the 6AL5, in a unique small metal envelope. The heater is connected to pins 2-7, the cathodes to 4 and 8, the plates to 3 and 5. 3 and 4 are one diode, 8 and 5 the other, and completely independent. It can be used for any reasonable service, such as AM detection, as a full-wave rectifier, or as a voltage doubler, so long as the current per plate is 8 mA or lower, and inverse voltages do not exceed 420 V. The voltage between heater and cathodes should not exceed 330 V. Measure the plate current as a function of the plate voltage up to 10 mA (the plate voltage will be about 7 V), and plot the current against the 3/2 power of the voltage. I obtained a rather straight line, showing agreement with Langmuir-Child, with a permeance of 0.5 $mA/V^{1.5}$. At 8 mA, the incremental resistance was 590Ω, and V/I = 785Ω. The 12H6 and 7H6 are similar tubes with different heater ratings and basing.

The 7Y4 is a typical small full-wave rectifier with an indirectly-heated cathode, like the more common 6X4 (miniature) and 6X5 (octal). This "Loktal" tube is inexpensive. Many of the common rectifier diodes are rather costly, for the curious reasons associated with the current tube market. The heater, taking 6.3V at 0.5A, is connected to pins 1-8 (as with all Loktal tubes). The cathode is pin 7, and the plates are pins 3 and 6. The peak inverse voltage is 1250 V, the peak current 180 mA, and the average dc current 70 mA. The heater-cathode voltage should not exceed 450 V. Measure the plate voltage for currents up to, say, 50 mA, and plot the results as for the 6H6. Again, we find a straight line and a perveance of 0.58 $mA/V^{1.5}$. Note that the plate voltage varies considerably as the current changes, from 4 V at 7 mA, to 16 V at 40 mA. Compare these voltages with those for a mercury-vapor *phanotron* as discussed in the next section. The 7Z4 is a somewhat larger full-wave rectifier (with perveance 0.40), the

---

Diode Characteristics

by Kenneth A. Kuhn

Oct. 3, 2007, rev. Sept. 3, 2009, draft –more to come

Introduction

This paper examines various electrical characteristics of a typical silicon junction diode. Useful mathematical relations are shown and illustrated with plots.

All of the plots are based on a typical sample of a very common small signal diode, the 1N4148 using a spreadsheet, diode_plots.xls, written by the author and posted on his web site: http://www.kennethkuhn.com . Because parameters vary from diode to diode, these plots should be interpreted as representative rather than absolute. Variations from diode to diode would typically be in the plus or minus several percent. To illustrate temperature effects, several of the plots are created using three temperatures: the high temperature is 100 C, the medium temperature is 25 C, and the low temperature is 0 C.

Forward current versus voltage

The standard equation for current through a diode is:

$I = IS * (\exp(V/(n*k*T/q)) - 1)$ Eq. 1

Where:
I is the current through the diode
IS is the reverse saturation current
V is the voltage across the diode (can be positive or negative)

n is a junction constant (typically around 2 for diodes, 1 for transistors)
k is Boltzmann's constant, 1.38E-23 Joules/Kelvin
T is temperature in Kelvins
q is the magnitude of an electron charge, 1.609E-19 coulombs

Looking at Equation 1 it would appear that the current should decrease as the temperature increases. The exact opposite is what really occurs. The reverse saturation current, IS, is a strong positive function of temperature as discussed below. The increase in IS with temperature more than offsets the effect of T in the exponential above.

The junction constant, n, is typically a constant at low currents and varies as the current becomes significant and may also vary somewhat with temperature. For this discussion, n will be taken as constant.

The sub-expression, kT/q, has units of voltage and is referred to as the thermal voltage and is typically around 26 millivolts at room temperature.

VT = k*T/q Eq. 2

Figure 1 shows a plot of Equation 1 for three temperatures. The strong temperature dependence can be clearly seen. There is a very important thing to note about the plot in Figure 1 –it is not possible to directly obtain the plot by applying a voltage and measuring the resulting current. Doing that will quickly destroy the diode because of a condition known as thermal run-away. Any power applied will heat the diode which in turn will result in the current increasing thus further heating the diode because of the negative temperature coefficient. This can be easily seen in Figure 1. The only way to collect the data for the plot is to apply a current and measure the resulting voltage

---

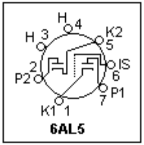

The 6AL5 dual diode, whose basing is shown at the right (7-pin miniature socket), is a typical signal diode. IS is an internal shield between the diodes. The two diodes and the shield are easily seen through the glass envelope, and you should notice how close the plates are to the cathodes. The close spacing means a large perveance, so only small plate voltages are required. Don't connect this tube directly across high voltages! A peak inverse voltage of 330 V can be resisted, and the DC plate current should not exceed 9 mA. Peak currents can go up to 45 mA if necessary, however. I measured the perveance as 2.42 $\text{mA/V}^{1.5}$, for one plate, a large value. The 6AL5 gives 9 mA with a plate voltage of only about 2.5 V! The heater, connected to H-H, pins 3 and 4, takes 0.3 A at 6.3 V.

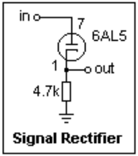

Try the 6AL5 in the circuit shown at the left, which is a basic signal rectifier with a 4.7k load resistor. Feed it with the signal generator, and compare the output and input with the oscilloscope. Try input peak-to-peak voltages of only 2 V or so. You will notice that there is no "diode drop" with the 6AL5--it acts like a perfect diode, rectifying down to small voltages. We know how to do this with a semiconductor diode and an op-amp, but here it's done quite simply. The 6AL5 has an incremental resistance of only about 237 Ω, and is nearly linear. It is easy to run a plate voltage versus plate current curve with a low-voltage power supply. Keep the load resistor, and subtract the voltages at plate and cathode to find the plate voltage.

Thermionic diodes have now been completely superseded by semiconductor diodes, largely for economic reasons, physical size and the need for a filament supply. A silicon diode capable of carrying 1 A is available for $0.04 or so, and takes up very little room. However, diodes can teach us a lot about thermionic emission and other interesting things. They do work rather well, and it is good to make their acquaintance.

The forward voltage (in the direction of current flow) of a diode is always relatively low, less than 15 V or so. The plate current is roughly proportional to the 3/2 power of the anode-cathode voltage (Langmuir-Child law), and the proportionality factor is called the *perveance*. The perveance depends on the geometry of the tube, increasing with larger area and closer spacing. It's remarkable that most diodes agree with Langmuir-Childs so well, in spite of different geometries. Since the voltages are low, contact potentials may affect your measurements. Contact potentials are discussed below in the section on low-voltage tubes. The easiest way to find the perveance is to plot $I^{2/3}$ against V, and to draw the best straight line. The intercept gives the value of the "true" zero plate voltage, and the slope, raised to the 3/2 power, is the perveance. Perveances range from 0.02 to 2.4 mA/$V^{3/2}$ for a representative assortment of 12 diodes of all types. There is no turn-on voltage drop for a thermionic diode, as there is for a silicon diode. Conduction begins immediately when the plate is positive with respect to the cathode, and stops immediately when the plate goes negative. It is easy to measure the V-I characteristic of a diode with a low-voltage DC supply, a voltmeter and an ammeter. I use a 100Ω resistor in series to make adjustment easier and safer. Thermionic diodes are not as easy to destroy as semiconductor diodes, and will take a good deal of abuse.

---

across the diode. Diodes are current operated devices and should never have a direct voltage applied from a low-impedance source.

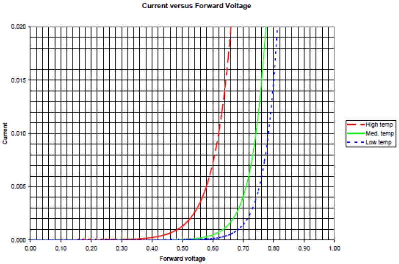

Figure 1: Forward Current versus Forward Voltage

The -1 term in Equation 1 can be omitted when the forward voltage is greater than about 0.1 volts at room temperature. A chart of the approximation and the exact is shown below.

For n = 1, VT = 0.026 volts

	Approximation	Exact	
V	exp(V/(n*VT))	[exp(V/(n*VT)) -1]	error%
0.00	1.00	0.00	infinite
0.02	2.16	1.16	86
0.05	6.84	5.84	17
0.10	46.8	45.8	2.2
0.15	320	319	0.3
0.20	2191	2190	0.05

Logarithmic forward voltage versus current

It is useful to plot the data in Figure 1 with the axis swapped as is done in Figure 2. An interesting observation in Figure 2 is that the forward voltage is a very linear function of the logarithm of the current through the diode. This characteristic is useful in building electronic logarithmic converters. It should also be observed that the slope is a function of temperature and that creates a complication as temperature compensation is required.

Equation 1 can be solved for the forward voltage:

$V = n*VT*\ln[(I / IS) + 1]$ Eq. 3

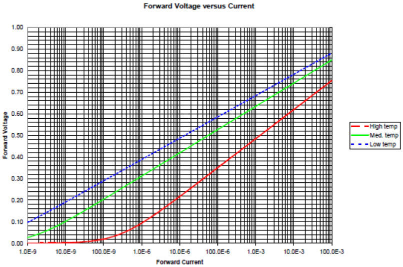

Figure 2: Forward Voltage versus Forward Current

Reverse saturation current

When negative voltages are applied to the diode the current becomes constant at –IS as the exponential term in Equation 1

Diodes

Thermionic diodes, like semiconductor diodes, are divided into *signal* diodes that handle small currents at low voltages, and *rectifier* diodes that handle large currents, often with large inverse voltages. A diode has an electron-emitting cathode and an electron-receiving anode or plate. The arrangments of cathodes and plates in commercial tubes, and what they are called, are shown in the figure. Signal diodes are also often added to a triode or pentode, sharing the same cathode and with one or two plates. Current flows only from plate to cathode, and this unidirectional conduction is the purpose of a diode. Diodes cannot amplify.

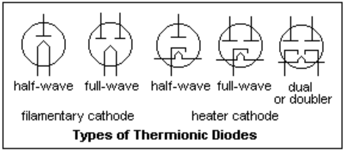

Signal diodes always have indirectly-heated cathodes, so they are easy to use. It is only necessary to make sure that the heater-cathode voltage does not exceed specified limits, usually a few hundred volts. Rectifier diodes often have filamentary oxide-coated cathodes, since these cathodes are more efficient when large currents are needed, requiring less power. We are considering only vacuum diodes, *kenotrons*, in this section. Thermionic gas diodes, or *phanotrons*, will be treated below, since they have rather different properties.

Bakelite) to support the contact pins mechanically, taking the strain off the pressed-glass seal, which was made of lead glass. Around 1935, the metal envelope was developed, but there was still a base. The all-glass "miniature" tube was made possible by the "button seal" that supported the contact pins mechanically as well as bringing them through the glass, allowing the base to be eliminated and tube size to be reduced. The insides, or "cage," was the same size as in previous tubes, however. It is supported on its leads, which are welded to the contact pins before the envelope is fused in place and evacuated. The button seal is also used, in a larger form, on tubes designated by GB at the end of the type designation, and by Loktals. The final step in manufacture was "flashing" the getter, usually barium or magnesium, to perfect the vacuum by adsorption of any remaining gases, leaving a shiny coating. This was generally done by heating a loop inductively by RF from outside.

quickly approaches zero. That is why it is referred to as the reverse saturation current. The current is independent of applied voltage once a small voltage magnitude is exceeded. This current is very small and is typically in the low nanoampere region. The reverse saturation current is a strong function of temperature as illustrated in Figure 3. The following equation is a simplified model for the reverse saturation current and provides excellent results in normal operating regions.

IS = IK * exp-Eg/nVT Eq. 4

Where:
IS is the reverse saturation current
IK is a constant derived from n and IS at a known temperature
Eg is the bandgap voltage for silicon (ranges from about 1.20 to 1.28 volts)
n is the junction constant (typically around 2 for diodes, 1 for transistors)
VT is the thermal voltage as previously discussed

The reverse saturation current should not be confused with an imperfection in diodes known as leakage current from a high value shunt resistance across the diode jucntion. Leakage current is often many times larger than IS. Thus, IS can not be directly measured and must be computed using data from the forward bias region (see the section on measuring diode characteristics).

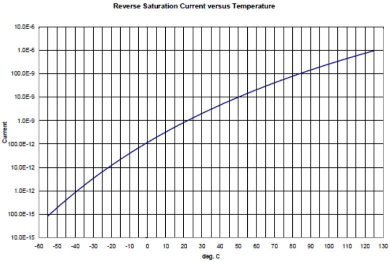

Figure 3: Reverse saturation current, IS, versus Temperature

Near Zero Bias

A theoretical plot of the diode current for the near zero bias condition is shown in Figure 4. Note that once the reverse voltage exceeds about -0.2 volts in magnitude that the reverse current is the value for IS and is independent of reverse voltage –this is why it is referred to as the reverse saturation current. Only the plot for room temperature (25 C) is visible. The low temperature plot is too small to observe on this scale except for the small forward bias region. The high temperature plot is very steep passing through zero volts. This illustrates why it is important to keep diodes cool when reverse current is an issue.

The plot in Figure 4 does not include the effect of leakage shunt resistance across the junction resulting from imperfections in the manufacturing process. This resistance varies from hundreds of millions to tens of billions of ohms. This shunt resistance results in an increasing current with applied reverse voltage and the current can become many

so the locking action guarantees that the tube will stay in the socket in spite of the small pins. The tubes are roughly the same size as an octal GT tube. Most are one size, but a few power tubes have a slightly longer envelope. There are no grid caps on any Loktal tube, and the heater connections are always to pins 1 and 8. Among the thoughtful features of loktal design, the type number appears in a hexagon on the top of the tube where it is visible from above, not on the side as on octal tubes. There is a dimple on the base corresponding to the key of the central pin, making the tube easy to orient for insertion. It seems that a lot of getter was used, so the tops of the envelope appear heavily silvered. The available types are only those used in AM and FM receivers. There are, nevertheless, enough types for a broad variety of experiments, and the prices are not excessive, so you may want to standardize on Loktals. Type numbers beginning with 7 have 6.3 V heaters, while type numbers beginning with 14 have 12.6 V heaters. There are some 7xx and 14xx tubes that are not Loktal, and some tubes that actually take a 7 V heater supply. One loktal rectifier, the 5AZ4 (a 5Y3 equivalent), has a 5 V filament. Loktal tubes designed specifically for battery-powered equipment had 1.4V filaments. The type numbers began with "1L." There were also rectifier and beam power loktals with 35, 50 and 70-volt heaters for AC/DC sets with series heater connections.

A tube designated simply 6N7 will be a metal-envelope octal tube with a 6.3V heater. A 6N7GT will have a cylindrical glass envelope. A 6N7G would have a shouldered glass envelope of the graceful shape designated ST. The electrical characteristics of such tubes were the same, whatever the envelope shape.

A very important part of vacuum-tube technology was bringing the metal leads through the glass envelope. Coefficients of expansion must be exactly matched, and the seal must be strong. Originally, tubes had bases (usually "duodecal" tubes were used in TV sets. Miniature tubes were not miniature, simply tubes with a button seal and all-glass envelope closely fitting a normal-sized cage. Subminiature tubes were actually miniature. Sometimes connections to grid, plate or (rarely) cathode were made to caps at the top of the tubes. In small tubes, these caps have a diameter of 1/4". The pins are numbered consectively clockwise, starting from the left of the index key for the octal, or to the left of the wider space, for the miniature, always looking at the bottom of the tube. This is shown for the octal base at the left. Pin numbers are given in the circuit schematics here. Most sockets have pin numbers marked. You will need to get sockets for the tubes you study, one for each type of socket. Solder wires to the tab at each pin that can be inserted in the solderless breadboard. I use the resistor color code for the pin numbers. A convenient octal socket fixture is available that comes with screw terminals for making connections. It was intended for relays, but is very useful for tube experiments. Heater connections for octal tubes are typically (not always!) to pins 2 and 7, and often to pins 3-4 on 7-pin, or 4-5 on 9-pin, miniature tubes. Sometimes halves of the heater can be connected in series or parallel, for two different voltages. Sockets were originally mounted in holes punched in aluminum chassis, secured by locking rings or by screws and nuts with a mounting plate. The chassis was, not surprisingly, the ground or common.

The "Loktal" tube was an excellent idea that was never universally adopted, mainly because miniature tubes took over in the 1950's. Since loktal was a trade name, RCA used "lock-in" instead, and you sometimes see "loctal." The loktal tube has an 8-pin button-seal (like the seal on miniature and octal GTB tubes). A natural metal base (of some aluminum alloy, apparently) shields the base of the tube and has a central pin with a circumferential locking groove. The pins project only 6 mm, and are 1.4 mm in diameter, much smaller than octal pins,

times IS. The effect of this resistance is most important at low temperatures where IS is very small. At high temperatures the effect of large IS swamps the resistance.

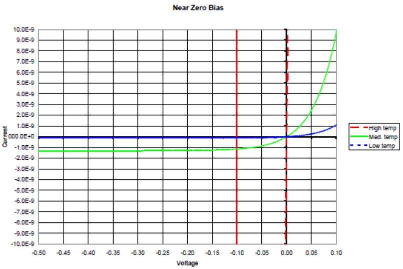

Figure 4: Near Zero Bias

(the vertical red line at -0.1 volts is a Microsoft bug and not part of the actual plot)

The plot in Figure 5 includes the effect of leakage shunt resistance. For this plot the resistance is 100 Mohms which was chosen for illustration and is very low compared to typical values of 1000 Mohms or more.

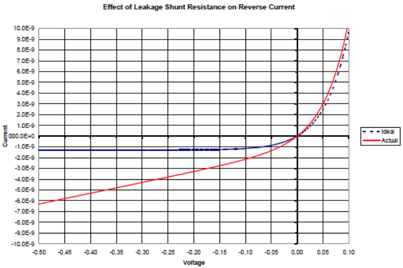

Figure 5: Effect of leakage shunt resistance

Temperature effect on forward voltage
With a constant current, the forward voltage drop of a diode has a very linear negative slope with temperature. The slope is also a function of current which is why constant
current must be used. This characteristic is useful for building electronic thermometers. This characteristic is shown in Figure 6. Note that the slope becomes steeper as the current through the diode is reduced. For practical temperature measurements it is important that the diode not heat with the applied current. The power dissipation should be less than 1 milliwatt. This means that the constant current should be no more than
about 1 milliampere. The plot in Figure 6 is based on Equation 3 but the x-axis is temperature instead.

main B+ supply. This cannot be done, of course, if the B+ voltage is adjusted using a variable transformer.

An all-in-one economical supply for vacuum-tube measurements is shown below. It uses an inexpensive isolation transformer from All Electronics, and can be made for about $30.00. The most expensive single part is the aluminum chassis. The grid potentiometer could be a precision 10-turn pot, but this would be expensive, and an ordinary carbon or plastic potentiometer (1/2 W or better) will be satisfactory. The maximum plate voltage of 120V and maximum plate current of 35 mA is adequate for many measurements. If you use a three-wire line cord, ground the chassis to the green wire. If you use only a two-wire line cord, it is probably better not to ground the chassis.

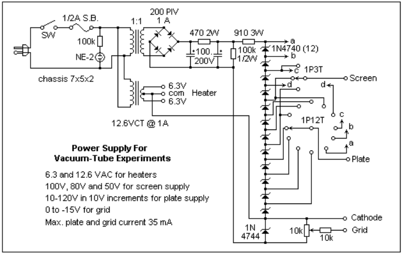
1/2A S.B.
SW
100k
NE-2
chassis 7x5x2
1:1
200 PIV
1 A
470 2W
910 3W
100 200V
100k 1/2W
1N4740 (12)
a
b
c
1P3T
d
Screen
6.3V
com
6.3V
Heater
12.6VCT @ 1A
1P12T
c
b
a
Plate
Cathode
Grid
1N 4744
10k
10k
Power Supply For Vacuum-Tube Experiments
6.3 and 12.6 VAC for heaters
100V, 80V and 50V for screen supply
10-120V in 10V increments for plate supply
0 to -15V for grid
Max. plate and grid current 35 mA

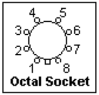
1
2
3
4
5
6
7
8
Octal Socket

Vacuum tubes come with metal or glass envelopes, and in latter days with either the familiar octal arrangment of 8 pins, or as miniature glass tubes with 7 or 9 pins. There were earlier bases with four, five or six pins. Later, 12-pin miniature "compactron" or

The RC ripple filter is worth the expense. Waveforms are shown at the left. The waveform at node "a" is the familiar one for a "tank" capacitor, and the ripple is fairly large. Since the impedance of a 100 μF capacitor is only 26Ω at 60 Hz, the ripple is reduced by a factor of almost 25. At 300V output and a load of 12 mA, the ripple is less than 0.1V, a very satisfactory result. Note that all that is left in the ripple is the 60 Hz component. The filter would work even better on a full-wave rectifier, but here it is very satisfactory, better and more economical than larger capacitors. Of course, a filter choke could be used for an even better result and less voltage drop, but this would double the cost of the supply.

You will also need a heater transformer, which can be quite small if supplying only one tube that requires 0.3A. The transformer can be put in a box with an on-off switch and convenient terminals. Ground the heater supply (at a center tap if one is provided) to the B+ ground, to avoid excess voltages between heater and cathode. If you have a 12.6V CT (center-tapped) secondary, you can supply both 6.3 and 12.6 V heaters. Many 12.6V miniature tubes can also be connected for 6.3V. Tubes whose designations begin with "1" have filaments that can be supplied from a single D cell. Obtain a holder for the cell so connections are easy. 6.3 V was chosen to be compatible with 6 V car batteries, but the supply is usually AC. Many rectifier diodes use a 5 V filament or heater supply, apparently for historical reasons.

A "C" supply, for the grid bias in measuring characteristics, can be any isolated low-voltage supply of say, 15V, and a potentiometer can be used to pick off a variable voltage, since little current is involved. A separate high-voltage supply for screen grids may also be convenient, though it is easy to pick off the necessary voltages with a Zener or a VR tube from the

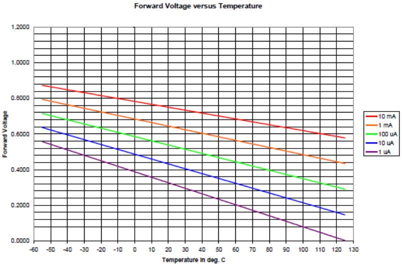

Figure 6: Forward Voltage versus Temperature

Temperature slope of forward voltage

The slope of the forward voltage versus temperature curve varies with the magnitude of the constant current as shown in Figure 7. This plot is made by taking the derivative of Equation 3 with respect to current.

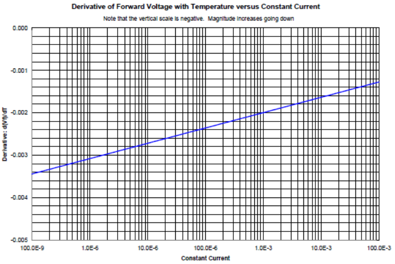

Figure 7: Derivative of Forward Voltage with Temperature versus Constant Current

Forward dynamic resistance

The operating point on Figure 1 represents a conductance (or its reciprocal function, resistance) for small signals (less than about 20 mVpp). The resistance is the reciprocal of the slope of the line through the operating point. This makes it possible to have an electrically variable resistor –the resistance is inversely proportional to the current. Electronically variable resistors are useful for building multipliers and gain control circuits. By taking the reciprocal of the derivative of Equation 1 (leaving off the -1 term) we have the forward dynamic resistance.

rd = n * VT / I Eq. 5

Observe that this resistance is a strong function of temperature. A plot of Equation 5 is shown in Figure 8.

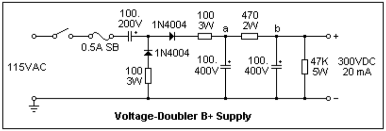

An idea for an inexpensive B+ supply is shown at the right. The greatest expense is for the capacitors, which will cost about $15. It is based on a half-wave voltage doubler, and gives 300V for a 115V rms input. It cannot supply large currents, but is perfectly satisfactory for anything but power amplifiers. If supplied from a variable transformer, it becomes a variable supply for all voltages from 0 to 300. Note very carefully that one side of the supply is connected to the AC line, and this must be the grounded side, for your safety, and to avoid ground loops. You cannot ground the positive terminal of this supply to get a negative voltage supply (for use as a C supply, for instance). An isolation transformer, if you have one, would eliminate this hazard. If you don't have an isolation transformer, use a polarized plug to guarantee that the white wire is connected to the circuit ground. If you have a good ground, consider the old trick of connecting only one wire in the power cord, and using the ground to complete the circuit. It is best to observe the power ratings of the resistors and the voltage ratings of the capacitors. This circuit has been tested, except for the fuse. If the 0.5A slow-blow fuse fails, try a 1.0A. This fuse is to turn things off if a capacitor fails; nothing valuable is protected here, but it saves mess.

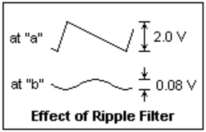

directly, because of the ground hazard. A variable transformer (Variac) is an *autotransformer* that does not isolate the output from the power line ground. I earnestly recommend that you do not work on AC circuits without isolating them from the service ground. The 110/220 adapters commonly available in 50W and 300W sizes, used for shavers and other small loads, should not be used, since they are autotransformers and do not provide isolation. They are, in fact, quite dangerous things, and should be used with great care. The supply was built in a 5" x 7" x 3" aluminum box, with an octal socket for the VR tube on top. The socket can be left vacant when the fixed voltage output is not required.

The voltage regulator requires a certain minimum current (about 5 mA) to function properly. If you are only drawing a few milliamperes from the supply, connect a 12k bleeder resistor across the output. Otherwise, the regulator will not adjust down to the lower voltages. Or, 220Ω and 10k fixed resistors, and a 15k pot, could be used at the voltage regulator, which would draw the necessary minimum current. The VR tube can be replaced by a high-voltage Zener diode.

A 25W isolation transformer is available at the date of writing from All Electronics (See the Your Laboratory page for a link) for $4.50. This transformer is surplus from the Power One firm, and is an excellent value. Solder a jumper between tabs 1 and 3, and another between tabs 2 and 4. The 120V input is connected between 1-3 and 2-4. The output tabs are marked B. This transformer would work well in the circuit above, or it could be put in a box and wired with line cord and output receptacle as a general isolation transformer. It should supply 200 mA without trouble, ample for our purposes.

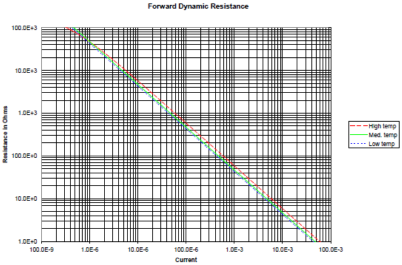

Figure 8: Forward Dynamic Resistance versus Current

Bulk Resistance

All diodes have a bulk resistance, Rb, that is in series with the diode. This resistance is typically very small (typically around 1 ohm or less) and its effects are insignificant at lower currents. At higher currents the effect is to reduce the voltage across the junction by the current multiplied by the resistance. This results in the steep theoretical exponential current versus voltage curve transforming into a linear slope. Equation 6 is Equation 1 with the effect of Rb included. Equation 6 (which omits the -1 term as that is not applicable for the higher diode currents) does not have a closed form solution as there is no known way to manipulate the equation such that either I or V is not involved with either a logarithmic or exponential term. Numerical solutions are the only way to solve Equation 6. Newton’s method was used to generate Figure 9.

$I = IS * \exp((V-I*Rb)/(n*k*T/q))$ Eq. 6

Figure 9 is a plot of Equation 6 and shows the effect of a 1 ohm bulk resistance for the room temperature case of the diode. Notice how the current is significantly less than theoretical for higher diode voltages. Note that the slope is nearly linear.

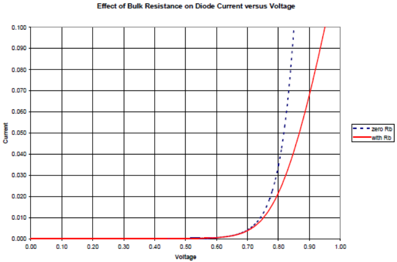

Figure 9: Effect of bulk series resistance

Sensitivity to light

The PN junction will respond to light –particularly longer wavelengths such as red and infrared. This feature can be a problem for diodes in transparent cases such as the 1N4148. The effect of light is to significantly increase the reverse conduction. A voltage can appear across the diode in response to light in high impedance circuits. This can be observed by connecting a diode (in a transparent case such as a 1N4148) to a DVM on the 2 volt scale and varying the applied light. In situations where light sensitivity could be a problem it is common to enclose a transparent diode in an opaque sleeve.

precautions and normal care, you will be fine. The solderless breadboard, DMM and oscilloscope can handle these voltages quite well. All the usual resistors and potentiometers are not afraid of 150V, so long as power ratings are observed. 1W and 1/2W resistors may be required in some places. Capacitors must be able to stand the voltages across them; many of those used with transistors will not be adequate. Keep a separate kit of capacitors rated at 100 V and above for this work. High-voltage capacitors are not needed everywhere, only in the plate circuits and for coupling from plate circuits.

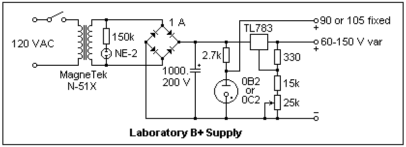

The circuit of the laboratory B+ supply that I use for vacuum tubes is shown at the right, and the supply itself is shown in the photograph below. It provides a regulated variable 60-150 V output, and a regulated fixed voltage output (for screen supply) created by a VR tube. VR tubes can be exchanged for different voltages. The MagneTek N-51X 115 V-115 V isolation transformer is available from Antique Electronics (see references), and a cheaper one from All Electronics. The transformer secondary has a DC resistance of 22Ω, which limits the surge current satisfactorily without having to add a series resistor. By no means eliminate the isolation transformer and use the 120 V household supply

were manufactured for voltages of 75, 90, 105 and 150 that were used like Zener diodes, handling from 5 to 40 mA. There is more information on glow discharges in Relaxation Oscillators, and on Zeners in Voltage Regulators. VR tubes are treated here because of their association with vacuum tubes, and the higher voltages involved.

Experiments

Vacuum tubes generally operate at higher voltages than transistor circuits. Like transistors, vacuum tubes are happier at higher voltages, which for receiving-type tubes, typically would be, say, 200 to 250V. It was once quite common to make DC power supplies for such voltages, using a transformer with a center-tapped secondary (say 200-0-200V), and a rectifier with double anodes and a common cathode, feeding a filter consisting of capacitors of 8 or 16 μF, and a choke of 10 H or so. It was not convenient to make a bridge rectifier with vacuum diodes (three separate filament transformers are necessary), so full-wave rectification with a center-tapped secondary was usual. These days it is rather difficult (and expensive) to acquire all these things, with the possible exception of the capacitors, which are now available up to millifarads at voltages up to 450 V.

The voltages normally used with receiving tubes are not high enough to be really dangerous, though a shock will not improve your day. If you are eager for new experiences, I can save you the trouble of finding out by saying that a DC shock is kind of like a hammer blow, not the zap of an AC shock, and does not paralyze, as an AC shock does. Shocks are given by current, not simple contact, so a good and old rule is to work with one hand in your pocket around high voltage. Always turn things off before making any adjustments or changes, of course, and be neat. Avoid touching bare metal. With these

---

There is a special type of diode known as a photodiode which is specifically made for sensing light. The case is transparent and often includes a lens to focus the light on the junction for increased sensitivity. Technically, it is usually infrared emission between around 800 to 1100 nanometers rather than visible light that these diodes are sensitive to. The sensitivity in the visible spectrum is usually poor except for what is known as blueenhanced diodes. These diodes can be operated either in an unbiased voltage mode or a current mode using a fixed bias around 5 volts. The current mode is very linear over a wide range. The current is sensed by a virtual ground electrometer.

Measuring Diode Characteristics

The mathematical model for diode operation can be completed by knowing IS and n. Neither of these are directly measurable but both can be calculated from data obtained by measuring the diode current at two voltages with temperature constant. We start with the following equation that omits the minus one term since our measurements will be at voltage levels high enough so that term is insignificant. Thus we have:

$I = IS * \exp V/(n*VT)$ Eq. 6

Equation 6 can solved for V by

$V = n*VT*\ln(I/IS)$ Eq. 7

$VH = n*VT*[\ln(IH) - \ln(IS)]$ Eq. 8A
$VL = n*VT*[\ln(IL) - \ln(IS)]$ Eq. 8B

Where H refers to the high voltage and high current measurement point and L refers to the low voltage and low current measurement point. These two points should be at currents less than 1 mA so that junction heating is insignificant. Good points to use are nominally 500 uA for the high value and 5 uA for the low value.

Subtracting Eq. 8B from Eq. 8A gives

$(VH - VL) = n*VT*(\ln(IH) - \ln(IL)$ Eq. 9

Eq. 9 can now be solved for n

$$n = \frac{VH - VL}{VT * [\ln(IH) - \ln(IL)]} \quad \text{Eq. 10}$$

As a reality check, n is typically between around 1.6 and 2.0 for diodes and should be very close to 1.0 for transistors. Suspect some error if n is less than 1.0 or greater than about 2.2. The most common type of errors are using an inaccurate temperature or that the high and low measurements were made at different temperatures.

We can now calculate IS for the particular temperature the data was measured at. That temperature, T, must be known and the corresponding value of VT used in the following equation.

$IS = IL * \exp{-VL/(n*VT)}$ Eq. 11

As a reality check, IS should be around 1 nanoampere give or take perhaps a factor of ten for diodes. For transistors, IS will be extremely small –picoamperes or less.

The proof that these two values are correct is that they produce results that agree with measured data over a wide range when the model is plotted. The use of only two points for the calculations makes it extremely important to keep

also acted as an electrostatic shield between control grid and plate, reducing the Miller capacitance to extremely small values, 0.003 pF in the 6SK7. If the screen and suppressor grid are connected to the plate, the pentode operates as a triode.

A very curious and ingenious kind of tube was the *electron-ray* tube, used on receivers to give a visual indication of the accuracy of tuning to a station. Don't confuse it with the *cathode-ray* tube that uses a guided electron beam for oscilloscopes and TV receivers. It showed a luminous disk, with a dark sector. The dark sector was made as small as possible to achieve accurate tuning. It worked from the AGC (automatic gain control) voltage of the receiver. This is a feedback signal that tries to keep the signal amplitude constant at the output of the intermediate frequency amplifiers, increasing the gain for weak signals and decreasing it for strong. It is usually a negative voltage produced by rectifying the IF output. The tube has a thermionic cathode and a conical anode or *target* covered with cathodoluminescent phosphor (like a CRT), which glows from the 3 or 4 mA of plate current that flows when it is across 125V or more (up to 250). Control electrodes, of which there are two on opposite sides of the 6AF6, make two dark sectors that are widest when at 0V, and narrow as the control voltage approaches the target voltage. The control voltage is typically provided by the plate of a triode controlled by the AVC, such that full negative AVC cuts off the triode and makes the sector as small as possible. All this was cheaper and more graphic than a pointer meter.

Another kind of tube that we'll look at here is the glow-tube voltage regulator. The voltage across a glow discharge depends on the gas and the cathode material, and is almost idependent of the current through the discharge in the "normal glow" region, in which the glow does not completely cover the cathode, and expands to accommodate more current. Tubes toward the plate, to suppress the escape of secondary electrons. This is provided by a third grid, the *suppressor grid*, which is usually connected to the cathode. The tube with three grids: control, screen and suppressor, or grids 1, 2 and 3, is called a *pentode*, which turns out to be a superior voltage amplifier, fully equivalent to a transistor. A typical small pentode, the 6SJ7, has a plate resistance of over a megohm, and a transconductance of 1.6 mS.

An ingenious modification of the pentode has electrodes that shape and concentrate the electron beam instead of a suppressor grid, the negative space charge of the electrons doing the same work. These are called *beam power* tubes, and were good for power work, as the name indicates. A typical example, the 6L6, had a transconductance as high as 6.0 mS, and the smaller 6V6 about 4.0 mS. Both types were widely used for high-fidelity audio amplifiers, and tube amplifiers still have proponents. The same tubes were used in small amateur radio transmitters, which shows the versatility of vacuum tubes.

Receiving pentodes were also classified as *sharp cutoff* or *remote cutoff*, an example of designing tubes to fit their applications. A remote cutoff pentode had a grid with variable spacing, so that areas of wider spacing let electrons through when the grid was made more negative, when areas of smaller spacing were cut off. This effectively reduced the transconductance of the tube, decreasing its gain in response to increased negative grid bias, which was used for AGC (automatic gain control) in IF amplifiers. The 6SK7 was a very popular remote cutoff pentode used as an RF and IF amplifier. The 6SJ7, on the other hand, was a *sharp-cutoff* pentode, used as an audio voltage amplifier. The amplification factor has little significance with pentodes, as with transistors, and transconductance is the important parameter. The screen grid

measurement errors small. Much better accuracy is obtained by using a spread of data over a wide operating range (say currents from 10 microamperes to 1 milliampere) and then using linear regression to identify n and IS. That method is straightforward and is not shown here. For single situations one would enter data in a spreadsheet and then manually tweak n and IS in the spreadsheet model to minimize the sum of error^2 between measured diode voltage and the voltage computed from the model.

The model for the complete diode characteristic including the effect of temperature on IS can be determined by substituting the computed values for n and IS from above and the thermal voltage based on the measured temperature into the following equation.

IK = IS * exp(Eg/(n*VT)) Eq. 12

As an example, at 25 C the calculated values for a sample diode was 1.83 for n and 1.3 nanoamperes for IS. Eg was taken to be 1.23 –roughly a middle value. IK was then computed to be 338 amperes. These are the values used to create all the plots in this paper.

More to come …

shall examine these circuits in detail in the experiments, but this provides the background.

In order to provide higher voltage gain, the plate resistance must be reduced somehow. We recall that with a transistor, the analogous collector resistance was very large, and there was no problem with voltage gain. The plate resistance is the result of the effect of plate voltage on the space charge. This effect is not necessary for control, which is provided by the control grid, so what we need is to eliminate the effect of the plate voltage on the space charge. This is done by introducing another grid, the *screen grid* between the control grid and the plate. If this grid is held at a constant potential, the space charge is "screened" from the effects of changes in plate voltage. The screen grid is usually bypassed to ground by a capacitor, whose reactance at the lower corner frequency should be smaller than the resistance connecting the screen grid to B+. Some, but few, electrons are removed by the screen grid, since it is again a coil of fine wire. With this change, the plate characteristics become (nearly) horizontal lines, as for the transistor, and the plate resistance becomes large, approaching a megohm. The resulting tube is called a *screen-grid tetrode*.

Although tetrodes worked as expected, they had a serious defect. It happens that speedy electrons colliding with the plate knock out *secondary* electrons. In a triode, these are rapidly sucked back to the positive plate, and the same happened in a tetrode when the plate potential was higher than the screen grid potential. In normal operation however, especially with large voltage gain, the plate voltage has a large swing, and can become less positive than the screen grid. Now all these secondary electrons (and some of the primary ones, too) are attracted to the screen grid, and there is a definite sag in the characteristic in this region. To prevent this, it is necessary to establish an electric field at the plate that is always directed amplifiers, the allowable plate currents can be quite a bit larger, and hundreds of watts output is possible with relatively small tubes. The interelectrode capacitances of the 6J5 are on the order of 3.4 pF, and are significant at high frequencies.

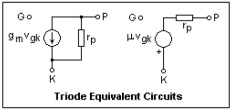

Triode Equivalent Circuits

The small-signal equivalent circuits for the triode are shown at the right. The Norton source circuit is exactly the one for the FET. In the case of the triode, the plate resistance is always important, and cannot be neglected. The other circuit is just the Thévenin source corresponding to the Norton source. It shows the significance of the amplification factor, and is useful for triodes because of the rather small plate resistance. We did not find the Th%eacute;venin source for transistors very useful, and did not introduce a parameter analagous to the amplification factor for this reason.

It's easy to see that the maximum voltage gain achievable with a triode is μ, if the load resistance is much higher than the plate resistance. The rule we derived for the gain of a transistor amplifier as the ratio of the collector and emitter resistances holds here as well, expressed by $\mu = g_m r_p$. The quantity analogous to r_e is $1/g_m$, which is 333Ω for the 6J5. Mu is a rather modest number, so triodes are not good for high voltage gain. They make very good power amplifiers, however, since large currents can be controlled. Using the Thévenin source, the gain of a usual common-cathode amplifier (analogous to a common-emitter amplifier) is simply a voltage divider problem. There must be a plate resistor in series with the plate, or else the voltage would never change, but it should be as large as possible, and has only a small effect on the gain. We

---

Principles of Semiconductor Devices
http://ecee.colorado.edu/~bart/book/book/chapter4/ch4_4.htm
by B. Van Zeghbroeck, 2011

Chapter 4: p-n Junctions
4.4.4. **I-V characteristics of real p-n diodes**

The forward biased I-V characteristics of real p-n diodes are further affected by high injection and the series resistance of the diode. To illustrate these effects while summarizing the current mechanisms discussed previously we consider the I-V characteristics of a silicon p+-n diode with Nd = 4 x 1014 cm-3, tp = 10 us, and up = 450 cm2/V-s. The I-V characteristics are plotted on a semi-logarithmic scale and four different regions can be distinguished as indicated on Figure 4.4.5. First, there is the ideal diode region where the current increases by one order of magnitude as the voltage is increased by 60 mV. This region is referred to as having an ideality factor, n, of one. This ideality factor is obtained by fitting a section of the curve to the following expression for the current:

$$J = J_s e^{V_a/nV_t} \qquad (4.4.39)$$

The ideality factor can also be obtained from the slope of the curve on a semi-logarithmic scale using:

$$n = \frac{\log(e)}{V_t\,\text{slope}} = \frac{\frac{1}{\text{slope}}}{59.6\,\text{mV/decade}} \qquad (4.4.40)$$

where the slope is in units of V/decade. To the left of the ideal diode region there is the region where the current is dominated by the trap-assisted recombination in the depletion region described in section 4.4.3.2. This part of the curve has an ideality factor of two. To the right of the ideal diode region, the current becomes limited by high injection effects and by the series resistance.

High injection occurs in a forward biased p-n diode when the injected minority carrier density exceeds the doping density. High injection will therefore occur first in the lowest doped region of the diode since that region has the highest minority carrier density.

Using equations (4.4.1) and (4.4.2), one finds that high injection occurs in a p+-n diode for the following applied voltage:

$$V_a = 2V_t \ln \frac{N_d}{n_i} \qquad (4.4.41)$$

or at Va = 0.55 V for the diode of Figure 4.4.5 as can be verified on the figure as the voltage where the ideality factor changes from one to two. For higher forward bias voltages, the current no longer increases exponentially with voltage. Instead, it increases linearly due to the series resistance of the diode. This series resistance can be due to the contact resistance between the metal and the semiconductor, due to the resistivity of the semiconductor or due to the series resistance of the connecting wires. This series resistance increases the external voltage, Va*, relative to the internal voltage, Va, considered so far.

$$V_a^* = V_a + IR_s \qquad (4.4.42)$$

Where I is the diode current and Rs is the series resistance.

These four regions can be observed in most p-n diodes although the high-injection region rarely occurs, as the series resistance tends to limit the current first.

---

varied, the plate current and voltage vary along the load line. The quiescent or operating point can be selected at some point along the DC load line, and so the DC grid bias can be found. This grid bias can be obtained from a C battery or equivalent, or from a cathode resistor, just as an emitter resistor is used with a transistor.

A cathode bias resistor is often bypassed by a capacitor if its negative feedback effect is not desired in dynamic operation. The reactance of the capacitor at the lower corner frequency should be equal to the resistance looking into the cathode (normally $1/g_m$ in parallel with the cathode resistor). The size of the cathode resistor has only a small effect on the size of the bypass capacitor. The capacitor never has a large voltage across it, and can be a low-voltage electrolytic. Only part of the cathode resistor can be bypassed if some feedback is desired.

A typical triode, of which the 6J5 is chosen here as the example, has a $\mu = 20$, $r_p = 6.7$k and $g_m = 3.0$ mS. Of course, the relation $\mu = g_p m_p$ always holds. The value of transconductance may seem small, compared to a transistor, but it should be remembered that it refers to a high-voltage plate circuit, and that the input impedance is infinite, so the power amplification is extremely large. The 6J5 is called a "medium-mu" triode. The similar 6SL7, a "high mu" triode, has $\mu = 70$, $r_p = 44$k and $g_m = 1.6$ mS. The 6SL7 is a *dual* triode, two independent valves sharing the same heater. The dual version of the 6J5 is the 6SN7.

The maximum plate voltage of the 6J5 is 300V, and the maximum plate current is 20 mA. Its maximum *plate dissipation* is 2.5W (product of average plate current and average plate voltage). This gives an idea of the ratings of receiving tubes used as voltage amplifiers. As power characteristics are curved, these quantities vary for different currents and voltages.

What we desire to represent is a function of two independent variables $I_p(V_p,V_g)$, which can be represented as a surface in three dimensions. Our various characteristic curves are orthogonal views along one or another of the axes. There are three such views possible, each directly related to one of the three parameters, of which we generally use only two, the ones mentioned here.

A vacuum tube carrying a current I with a plate voltage V dissipates power VI, just as if it were a resistor. However, the process is different. In a resistor, an electron gives up small amounts of energy to the lattice as it is accelerated and then is scattered. In a vacuum tube, the electron acquires a kinetic energy as it is accelerated, which it gives up all at once when it collides with the plate. It is not really correct to ascribe this to the "plate resistance," as some texts do, which is an incremental ratio. Since the plate is in a vacuum, the resulting heat can only be radiated or conducted down the supports. Much of the radiated heat is infrared, which is absorbed by the glass tube envelope. Note how plates are blackened to raise their emissivity and often provided with fins. Really large tubes had plates externally exposed capable of air or water cooling with elaborate seals to the glass parts. *Plate dissipation* is always a limiting factor in power applications.

The line marked *load line* shows the difference between the supply voltage V_{bb} and the voltage drop in a resistance R_L in series with the tube at any plate current, giving the plate voltage directly. The series resistance is mainly a plate resistor, but if there is a cathode resistor (for purposes of biasing) it should be included. There are generally different load lines for static (DC) and dynamic (AC) operation. As the grid voltage is

---

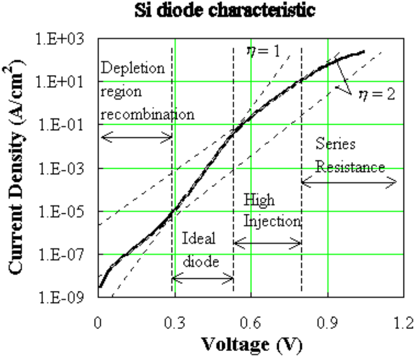

Figure 4.4.5: Current-Voltage characteristics of a silicon diode under forward bias.

characteristic, and a plot of I_b against V_g for a fixed value of V_p is called a *transfer characteristic*. From a family of either characteristics, the complete circuit behavior of the tube can be predicted. Unlike transistors, tubes of different types can have quite different (though similar) characteristics, so characteristic curves are much more important.

Let's begin with idealized plate characteristics for a triode, shown at the right. These are curved lines, but we represent them by straight lines for ease of understanding the various slopes and distances involved, which will be constant. Actual characteristics are not really far from straight lines, anyway. There is one curve for each grid voltage represented, which differ by a constant amount, here 2V. The horizontal distance between them, represented by the line labelled μ, is the amount the plate voltage must increase to hold the plate current constant; it represents the relative influence of plate and grid on plate current, and is the *amplification factor*. The vertical distance between them shows how much the plate current changes for a change in grid voltage. The ratio is a conductance, called the *transconductance*, denoted by g_m, measured in siemens (mho). The slope of the curves is the ratio g_m/μ, called the *plate conductance* g_p. With vacuum tubes, the reciprocal r_p was always used, called the *plate resistance*. Since actual

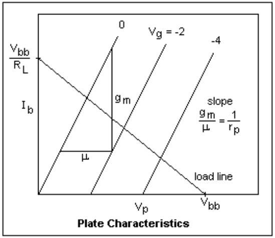

Plate Characteristics

rectifier tubes, because they gave more current per watt of heating power. The indirectly heated, equipotential cathode that could be supplied by AC rather than by battery power was widely used after 1930.

A tube with just cathode and anode is called a *diode*, a term that has survived into the semiconductor age. Diodes were used for power or signal rectification, just like their semiconductor relatives. A "full-wave" diode has two anodes. When a control grid is provided, the tube is called a *triode*, and is used for amplification. Let's first study the peculiar circuit behavior of triodes, which will lead us to the reason for the addition of more grids, and the creation of the *pentode*, which turns out to act very much like a transistor.

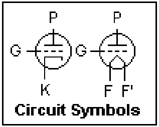

Circuit symbols for a triode are shown at the right, and other tube symbols are derived from it. The connections are plate P, grid G, and cathode K or filament F, F'. These are analogous to the collector, base and emitter of a transistor, with the same polarities and direction of current flow as an NPN transistor. The circle is a part of the symbol. Grid connections can be to the right or left of the symbol, as convenient. A gas tube is indicated by a dot in the lower right-hand part of the circle. The heater of an indirectly heated cathode is usually not shown. A cold cathode (operating by positive-ion bombardment) is shown as a small circle. There are examples of these symbols below.

The important variables are the independent variables V_p and V_g, the plate and grid voltages (with respect to the cathode), and the dependent variable I_p, the plate current. A plot of I_p against V_p for a fixed value of V_g is called a *plate*

Germanium diode I-V Characteristics: Ideal behavior analysis

Jose Maria Roman
https://www.uam.es/personal_pdi/ciencias/jmroman/presentations/jmroman_iv.pdf

Photovoltaic Cell and Module Technology
EN547
March 28, 2004

Abstract

This paper is devoted to the analysis of experimental results for a Germanium diode, and its comparison with the ideal diode theory. In particular, the I-V characteristics is measured for two different temperatures, showing a qualitative agreement with the predictions of the theory. Deviations from the ideal theory are presented and taken into account in a more detailed analysis of the forward bias characteristic. The temperature dependence of the reverse saturation current yields a value of the contact potential in good agreement with the known value of the gap for germanium.

1 Introduction

We analyze in this paper the experimental results for the I-V characteristics for a Germanium transistor ACY20, and compare them to the ideal diode theory. The ideal theory is based in simple assumptions of how the junction of two opposite doped semiconductors behaves when put together [1, 2, 3].

Measurements of the voltage and current across the diode at two different temperatures, t = 20dC and t = 80dC, are taken, showing a qualitative agreement with the predictions of the theory. Possible deviations from the ideal theory [1, 2, 3] are considered in a more detailed study of the forward bias characteristics, leading to several ranges of voltage where the ideal or ideal together with other mechanisms seem to dominate the current through the diode. The temperature dependence of the reverse saturation current allows us for obtaining an estimate of the contact potential, the potential barrier established between the two semiconductors of the junction. This value is in good agreement with the known value of the gap for germanium.

In section 2 we present the main features of semiconductors and junctions. Section 3 develops the simplest theory for the ideal diode, with a detailed calculation of the reverse saturation current from the underlying theory. Non-ideal mechanisms are outlined in section 4. The experimental set-up and measurement procedure are given in section 5, and section 6 contains the experimental results and their analysis. Finally our conclusion are presented in section 7.

2 The ideal diode

A diode is formed by the junction of two semiconductors doped with opposite sign, i.e., a p-n junction. The different electric character of the dopants at both sides of the junction establishes an electric field across the junction which determines the properties of the current which crosses the diode in the two different directions.

A n-type semiconductor is doped with elements called donors, because have electrons in excess of the intrinsic semiconductor elements (a semiconductor free of impurities or lattice defects is called intrinsic), e.g. phosphorous in silicon. The heavy doping density of donors, ND, determines the level of majority carriers, electrons in this case, so that nn0 = ND, while the

---

number indicates the heater voltage. 6 means 6.3V, 12 means 12.6V. Heaters are very forgiving of variations in voltage, but it is best to try to use the recommended voltages. AC is generally used, supplied from a small transformer. It is necessary to make sure that the difference in voltage between heater and cathode does not exceed 90V, so the heater AC supply should be grounded. It is usual to ground the center tap on the transformer for this purpose. Tubes for battery radios have plain filaments that are both heater and emitter, and must be supplied with DC. Their type numbers begin with "1" and are intended to be used with a standard dry cell of 1.5V. Larger tubes for high voltages have to use thoriated-tungsten filaments at 2000K,a bright yellow, to avoid damage from positive-ion bombardment, small as it may be. Heaters are sometimes called filaments, and the heater supply the filament supply, out of linguistic inertia.

Before the 1930's, each manufacturer used arbitrary designations for his tubes, and there was no uniformity or system. Tubes were first systematically identified in the U.S. by three-digit numbers, where the first digit denoted the manufacturer. For example, a type x10 was a power triode, a type x36 a screen-grid tetrode, where x was the manufacturer's number (usually omitted). Later, a new system was introduced where the first digit gave the filament or heater voltage, and the last digit gave the number of functional electrodes. This scheme was introduced by the RMA (Radio Manufacturer's Association) in 1934. A letter between these digits was assigned in order of introduction. For example, a 2A3 was a power triode with a 2.5 V filament (this popular tube is, remarkably, still in use for hi-fi amplifiers because of its low distortion!). This system was not comprehensive enough, and in the final system, the first number designated the heater voltage, but the remainder of the designation was arbitrary. Filaments were customarily used, especially in power and so electrons can pass through without hindrance. When it is made negative, it opposes the effect of the anode in creating an electric field, but does not attract any electrons, and so draws no current (except for the positive-ion current mentioned above). If it is made sufficiently negative, it can *cut off* the plate current entirely. If it is made positive, it can enhance the plate current, but then draws some *grid current* itself. The grid provides a sensitive control, using negligible power, of the large plate current, so the vacuum tube is a powerful amplifying device.

Early radio sets were battery-powered (domestic electrification was in its infancy, and absent in rural areas, when radio began), and a convention was established for identifying the batteries required. The filaments required low voltages at high currents (2W or more each), and their supply was called the A battery. The plates required high voltages at small currents, perhaps 90V at a few mA, and their supply was called the B battery. Grid bias was required, to hold the grids negative, demanding low voltages at small currents, and the corresponding battery was the C battery. The notations A and C were later little-used (except for actual battery radios), but the plate supppy came to be generally known as B+, and the letter B appears in subscripts of quantities referring to the plate circuit.

The cathodes of receiving tubes consist of a sleeve of nickel alloy coated with a compound of alkaline-earth oxides (Ba and Sr, usually). Inside is a tungsten heater wire insulated from the cathode with BeO or alundum (aluminum oxide) ceramic insulation. These cathodes must be heated only to about 850K (a dull red) to emit electrons in the amount necessary. Most receiving tubes require 6.3V or 12.6V for their heaters, at about 0.30A or 0.15A, respectively. Every tube type is identified by a type number, such as 6J5, where the first

---

density of minority carriers is determined from the mass action formula

$$np = n_i^2 = N_V N_C e^{-E_g/kT}$$

(2.1)

which leads to

$$n_{n0} = N_D, \qquad p_{n0} = \frac{n_i^2}{N_D} = \frac{N_V N_C}{N_D}\, e^{-E_g/kT}$$

(2.2)

where NV and NC are the density of states at the top of the valence and at the bottom of the conduction band respectively, and Eg is the band gap of the semiconductor.

Equally a p-type semiconductor is doped with acceptors, which have less electrons than the main element of the material (boron in silicon), and then trap one electron leaving a vacancy in the valance band, which is represented as a hole. Again, the doping density of acceptors, NA, determines the hole concentration in a heavily doped material, and the electron concentration is obtained from the mass action formula:

$$p_{p0} = N_A, \qquad n_{p0} = \frac{n_i^2}{N_A} = \frac{N_V N_C}{N_A}\, e^{-E_g/kT}$$

(2.3)

Now that we know the main features of n-type and p-type semiconductors isolated we put them next to each other. We assume here that we can treat electrons and holes independently. In this case the excess of electrons in the n-type side will diffuse into the p-type side. This diffusion effect produces a depletion of carriers around the interphase, known as the depletion region, the depletion will be supposed to be complete. Therefore producing an uncompensated charge to both sides of the junction which gives rise to an electric field (assumed to be contained exclusively within the depletion region). This electric field, in turn, generates a drift current opposite to the diffusion current, until an equilibrium between the two processes is reached.

$$I_{\text{diff}} = qD_p\frac{dp}{dx},$$
$$I_{\text{drift}} = qp\mu_p E, \qquad E = -\frac{dV}{dx} \quad (2.4)$$

Taking into account the Einstein relation Dp = kT¹p=q it is straightforward to obtain a relationship between the carrier density to both sides of the junction,

$$\frac{p_{p0}}{p_{n0}} = \frac{n_{n0}}{n_{p0}} = e^{qV_0/kT} \quad (2.5)$$

where V0 = Vn ¡ Vp is the potential drop across the junction generating the drift electric field, also known as built-in or contact potential. By using eqs. (2.2), (2.3) and (2.1) into (2.4) we can write

$$qV_0 = kT\log\frac{N_A N_D}{n_i^2} = E_g + kT\log\frac{N_A N_D}{N_V N_C} \quad (2.6)$$

which relates linearly the contact potential and the gap up to an additive factor depending on temperature, doping and density of states.

3 I-V Characteristic for an ideal diode

We can determine the general shape of the current flow through the junction considering a general argument for the mechanisms implied in the formation of the diode (see for example ref. [1], page 169).

The electrons emitted by the thermionic cathode form a negative *space charge* cloud around the cathode, dense enough that if no electrons are removed by attraction to the anode, the rate of emission is equal to the rate of return. When the anode is made positive, some of the electrons are attracted to it out of the space-charge cloud, and a *thermionic current* results. The amount of this current is given by $I = A\ V^{3/2}$, where V is the voltage from anode to cathode. This is called the Langmuir-Child law, and shows that electric field at the space charge produced by the anode controls the electron current. The cathode emits electrons copiously, so much that there are always enough electrons available to satisfy Langmuir-Child. Of course, at a sufficiently high anode voltage, the current may *saturate*, when all the emitted electrons are attracted to the anode, but this never occurs in normal operation, so small variations in cathode temperature have no effect. The current in a vacuum tube is said to be *space-charge controlled*.

If enough gas is present in the tube, the positive ions can counterbalance the negative electron space charge, robbing the anode of control and greatly increasing the current. Also, the positive ions can take over a large part of the electron emission at the cathode. Such tubes make efficient rectifiers, and the gas pressure can be quite low, as in some rectifiers, or rather high, as in a mercury-arc rectifier. In receiving tubes, positive ion collisions can destroy the delicate, high-efficiency cathode surface. Positive ions also cause small currents to negative electrodes that otherwise might be expected to carry no current at all. For all these reasons, receiving tubes have a high vacuum.

The electric field at the space charge that controls the current does not have to be created by the anode alone. A third electrode, the *grid*, is placed between the cathode and the anode, closer to the cathode. It is made of a spiral of fine wire, filaments used low voltages and high currents, and had a short life because of the high temperatures required for adequate emission. Most radio sets had a rheostat to adjust the filament current properly. Wehnelt emitters would have been quickly destroyed by positive-ion bombardment in these tubes. Apparently by accident, thoriated tungsten wire was used in a trial at the GE factory in Harrison, NJ in 1920 on a UV201 tube. Thoriated tungsten gave 75 mA/W of filament power, while tungsten gave only 1.75. Thoriated-tungsten filaments became popular for receiving tubes, such as the UV201A, which was an improved UV201, around 1924. Its filament required 0.25 A at 5 V, while the UV201's had required 1.0 A. Since tubes were now all *hard*, or high-vacuum tubes, indirectly heated oxide-coated cathodes, which gave copious emission at low temperatures, were used almost exclusively in receiving tubes after 1930. Not only did these cathodes have a long life, but were also *equipotential*, making circuit design simpler. Thoriated tungsten remained for transmitting tubes, where a rugged emitter was necessary because of the higher plate voltages, but even here tungsten was the only suitable choice for really high voltages, to avoid damage from positive-ion bombardment.

The rate of emission of electrons from a heated metal is given by the Richardson-Dushman equation, $i = AT^2e^{-b/T}$ A/cm^2, where T is the absolute temperature in K, and A and b are constants typical of the emitter. For tungsten, A = 60 and b = 52,400K, while average values for an oxide cathode are A=0.01, b = 11,600. The exponential factor has by far the largest influence, so emission increases rapidly with temperature. This makes thermionic cathodes very suitable even for heavy currents. In all tubes, electrons are emitted in far greater numbers than required; most simply return to the cathode.

As it was shown before there are two mechanisms which give rise to the diode conducting properties. On one hand there is the diffusion of electrons from the n-type side into the p-type side. These electrons have to overcome the potential barrier (contact potential) established by the electric field of the junction. This is achieved by a pure thermal activation mechanism

$$n \sim e^{qV_0/kT} \qquad (3.1)$$

and therefore the density of carriers overcoming the potential barrier depend on the height of the barrier (given by the contact potential V0), and the temperature.

This diffusion current is compensated by a drift current originated by the electric feld across the junction. Differently to the diffusion mechanism this drift current is not very much affected by the variations in the contact potential between the two sides of the junction, since the electric feld acts on the minority carriers, and is the small density of them which determines the flow rather than the electric field itself.

We can assume therefore that in equilibrium V = 0 the diffusion and drift currents are compensated, and are both given by I0. When we apply a forward bias on the diode V > 0 the potential barrier is reduced from V0 to V0 ¡ V , and therefore the diffusion current increases by a factor eqV=kT , while the drift current remains unchanged. In the case of applying a reverse bias V < 0 the diffusion current will be reduced by a factor e¡qjV j=kT , while again the drift current is unaffected.

From the previous argument we can establish the I-V characteristic for the diode

$$I = I_0 \left(e^{qV/kT} - 1\right) \quad (3.2)$$

where I0 represents the equilibrium diffusion and drift current for V = 0. Notice that for and infinity reverse bias the diffusion current goes to zero and the only remaining one is the drift, such that I(V = ¡1) = ¡I0, a small finite value known as reverse saturation current. On the other hand, for large forward bias the current across the junction increases exponentially, and it is completely dominated by diffusion.

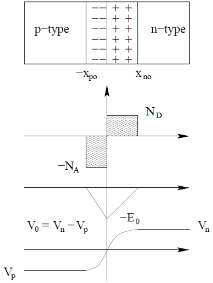

Figure 1. Ideal abrupt p-n junction, charge distribution, profile of the electric field and contact potential across the junction.

3.1 Reverse saturation current I0

Now that we have obtained the general expression for the I-V characteristics of the current across a p-n junction we will turn to compute the value of the reverse saturation current I0 under

"receiving" comes from their use in radio receivers, their principal commercial application, but refers to all small vacuum tubes for general electronic purposes. For the cathode-ray tube and making your own oscilloscope, see The Cathode-Ray Tube.

The electrons move from the cathode (K), the negative electrode, to the anode or *plate* (P), the positive electrode. Conventional current is in the opposite direction. The electrons are liberated at the cathode by heat--thermionic emission--or as a result of bombardment by positive ions, which can cause emission of electrons or even heat the cathode the required amount for thermionic emission. All receiving tubes employ thermionic emission, though we will note certain examples of *cold cathodes* in special cases. These were not usually really cold, but heated by ion bombardment rather than by a current supplied externally. The space in which the electrons move is not completely devoid of gas, so some gas molecules may be *ionized* by collision with speedy electrons, when an electron is knocked off, leaving a *positive ion*. The positive ions move in the opposite direction to the electrons (but their current is in the same direction, since they are of opposite charge). The effect of positive ions in a receiving tube is very small, because of the very high vacuum that is used.

Self-heated electron emitters are called *filaments*. The carbon filaments of the Edison Effect were soon replaced by metallic emitters, usually tantalum or tungsten, which were used by Fleming and de Forest. In Germany, Arthur Wehnelt discovered in 1903 that barium or calcium oxides baked on a platinum base emitted copiously, and used these *oxide-coated* emitters on vacuum rectifiers evolved from discharge tubes, which he patented in 1904. However, the use of *soft* tubes, which contained residual gas, demanded the use of rugged tungsten filaments, which dominated in 1910-20. These

1904 (B.P. 24850), where a filament and plate were arranged in the same envelope in a rather low vacuum, which could be used as a rectifier, or as a rather insensitive radio detector. In 1907, Lee de Forest patented the *triode* (which he called the Audion; the term "triode" was not used until much later, after it threated to become a trade name), in which a third electrode, the grid, was introduced to control the electron stream. This made a more sensitive detector, but the amplifying property was not used at first, and de Forest, who did not understand well what was going on, defended gassy tubes with their gas amplification. The introduction of high vacuum, as well as improved materials and processes, especially metal-to-glass seals, created a very useful amplifying device that allowed great developments in radio, telephony and sound reproduction. Schottky suggested a screen grid between the plate and control grid to make the electron tube useful at higher frequencies in 1919 (and actually made tubes with a second grid, but this was for space-charge control), but this was only realized by Hull and Williams in 1928 in radio receivers. The metal tube was introduced in 1935, but glass envelopes never disappeared and were constantly improved. The final pattern of electron tube was the "miniature" or all-glass type, which became the predominant receiving-type tube after about 1945. Transistors were invented in 1948, and in the next decade were improved to the point where they could take over most of the amplifying applications of electron tubes at much lower cost, and with greater reliability. Electron tubes remain in use as cathode-ray tubes, magnetrons, X-ray tubes, and for handling large powers. They were remarkable devices, using many sophisticated materials and processes, yet were widely available at low cost. We shall look here mainly at examples of *receiving tubes*, the smaller amplifying devices that have been completely replaced by semiconductors in current practice, but nevertheless will deepen our knowledge of electronics, while being fascinating to study. The name

the assumption that it is generated by low level (minority carrier) injection. In this approximation the majority carrier density is unaffected.

As shown in fig. 1 we suppose as a good approximation that we can treat electrons and holes independently, and that the drop of the contact potential is produced within the depletion region.

Assuming that the depletion region is completely depleted of carriers we have two regions with opposite charge, and densities equal to the doping, NA for the p-type side, and ND for the n-type one. This uniform charge produces a linearly spatially distributed electric field and a quadratically distributed contact potential, as show in fig. 1.

Under the previous conditions the current across the junction is established by the minority carrier injection on the other side of the junction. For example, holes will be injected

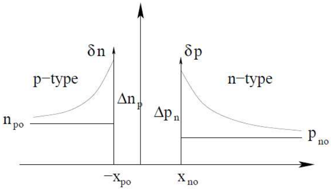

Figure 2. Profile of the low level injection in the p-n abrupt junction.

from the p-type side into the n-type side, and its concentration will decay to the equilibrium n-type hole concentration, pn0, due to recombination of the injected holes, as shown in fig. 2. Then, the concentration of holes in the n-type semiconductor is given by

$$\delta p = \Delta p_n \, e^{-(x-x_{n0})/L_p} \qquad (3.3)$$

and a similar expression for electrons,

$$\delta n = \Delta n_p \, e^{(x+x_{p0})/L_n} \qquad (3.4)$$

where xn0, xp0, Lp, Ln are the border of the depletion region in the n-type and p-type sides and the hole and electron diffusion lengths. ¢pn ´ p(xn0) ¡ pn0 is the excess of holes over the bulk n-type hole concentration at the n-type border of the depletion region. Similarly for ¢np ´ n(¡xp0) ¡ np0 for the p-type side.

The current passing through xn0 is given by the diffusion

$$I_p(x_{n0}) = -qAD_p \left.\frac{d(\delta p)}{dx}\right|_{x_{n0}} \qquad (3.5)$$

A similar expression to that in eq. (2.5) for the relation of carriers to both sides of the junciton under equilibrium conditions can be derived for the case of a forward bias, i.e. V > 0. The effect of a forward bias is to reduce the potential barrier to V0 ¡ V , and therefore the relationship between the holes in both sides of the junction is now given by

$$\frac{p(-x_{p0})}{p(x_{n0})} = e^{q(V_0-V)/kT} \qquad \longrightarrow$$

$$\Delta p_n \equiv p(x_{n0}) - p_{n0} = p_{n0}(e^{qV/kT} - 1) \qquad (3.6)$$

Theory of Vacuum Tubes

J. B. Calvert
Created 17 August 2001
Last revised 8 April 2003

http://mysite.du.edu/~etuttle/electron/elect27.htm#Diodes

Miniature vacuum tubes with cathodes of high-field-emitting carbon nanotubes are currently under study at Agere Systems in Murray Hill, NJ. A triode with amplification factor of 4 has been constructed, with an anode-cathode spacing of 220 μm, and a pentode is planned. Vacuum tubes may return to electronic technology! See *Physics Today*, July 2002, pp. 16-18.

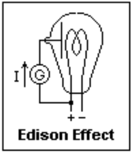

Devices in which a stream of electrons is controlled by electric and magnetic fields have many applications in electronics. Because a vacuum must be provided in the form of an evacuated enclosure in which the electrons can move without collisions with gas molecules, these devices were called *vacuum tubes* or *electron tubes* in the US, and *thermionic valves* in Britain. In 1883, Thomas Edison observed that a current flowed between the filament of an incandescent lamp and a plate in the vacuum near it (see figure at the right), when the plate was connected to the positive end of the filament, but not when the plate was connected to the negative side (the plate was actually between the two legs of the filament). No important application was made of this unexplained *Edison Effect* at the time. In 1899, J. J. Thomson showed that the current was due to a stream of negatively-charged particles, electrons, that could be guided by electric and magnetic fields. Fleming patented the *diode* in

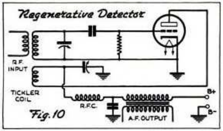

Fig. 10

ranged that part of the amplified signal is fed back to the grid to be reamplified. The feedback is supposed to be just a little too weak to sustain self-excited oscillation, the trick being to adjust the feedback just right.

A modification of this, the super-regenerative circuit, has several interesting and useful features. Though tricky, it's exceedingly sensitive and sharp-tuning. Essentially it is a regenerative detector, but a "quench circuit" consisting of a triode and an inductance-resistance network acts to interrupt its functioning at a supersonic frequency. The theory of its operation is ingenious. An ordinary regenerative detector with a little too much feedback will hesitate for a fraction of a second, then gradually sweep into oscillation. The super-regenerative detector has so much feedback as to cause oscillation, but the quench circuit keeps interrupting it before the oscillation can establish itself. This permits a tremendous amount of reamplification. The circuit is particularly useful on frequencies of 100 megacycles and above, where tube losses tend to be severe, for the principle permits B power to replace such losses.

The problem of demodulating F.M. signals is quite different from that with A.M.

In Fig. 11 the R.F. input is fed to a special intermediate-frequency transformer that has two secondary circuits, S_1 and S_2, each of which drives a conventional diode detector and filter circuit. (Usually one 6H6 dual-diode detector tube is used.) If the intermediate frequency of this receiver is 5.00 megacycles, the primary of the "discriminator transformer" would be tuned to the same value, S_1 tuned to 5.10, and S_2 to 4.90 megacycles. If the signal comes through at 5.00 megacycles, both secondaries and their associated detectors receive an equal voltage; and since the filter circuits are so arranged as to oppose the voltages, they cancel each other and no A.F. is produced. But when the signal frequency varies either up or down, it will approach the resonance of one secondary and move farther away from that of the other. One will get a decreased and the other an increased voltage, and a net voltage will appear across the audio output. Since the voltage across the audio-output terminals will increase as the R.F. signal departs from 5.00 megacycles, audio volume will be determined by the extent of frequency deviation. Audio pitch will likewise be determined by the number of times a second that frequency deviation occurs. The circuit therefore meets the required conditions of F.M. demodulation.

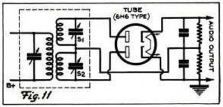

Fig. 11

Lt. M. L. Rupert's "foxhole radio," used on the Italian front, gets reception on a razor blade. O. B. Hanson, chief engineer of N.B.C., improved it by tying pencil lead to the pin point. Wire gauge is not critical, but good headphones, a long aerial, and patience in locating a sensitive spot on the razor blade are all requisite

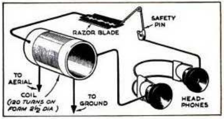

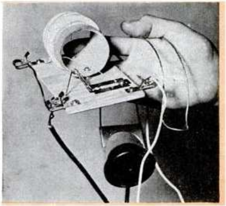

Plugging (3.6) into (3.5) and differentiating we obtain an expression for the current at the border of the depletion region xn0

$$I_p(x_{n0}) = qA\frac{D_p p_{n0}}{L_p}\left(e^{qV/kT} - 1\right) \tag{3.7}$$

and therefore the total current, adding up the electronic part, is given by

$$I = qA\left[\frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n}\right]\left(e^{qV/kT} - 1\right) \tag{3.8}$$

from where we can easily read I0.

After using eqs. (2.2), (2.3) and (2.5) we can rewrite the reverse saturation current as

$$\begin{aligned} I_0 &= qA\left[\frac{D_p}{N_D L_p} + \frac{D_n}{N_A L_n}\right] N_V N_C\, e^{-E_g/kT} \\ &= qA\left[\frac{D_p N_A}{L_p} + \frac{D_n N_D}{L_n}\right] e^{-qV_0/kT} \end{aligned} \tag{3.9}$$

which provide an explicit expotential dependence on the temperature and energy gap, or contact potential, assuming the terms inside the square brackets cancel each other or have a subdominant temperature dependence.

4 Non-ideal mechanisms

There are several aspects which have not been taken into account in the above derivation of the I-V characteristics, and further complicate the understanding of the diode behavior (ref. [1], page 211):

Recombination within the depletion region: We assumed that there was no generation-recombination processes in the depletion region. For material with a small density of carriers the depletion region can be quite large and then the probability of recombination inside the depletion region is higher even dominating the shape of the curve. The recombination rate is proportional to

$$R \sim 1/n_i \sim e^{qV/2kT} \quad (4.1)$$

Generation within the depletion region: For reverse bias what it is important is the rate of generation, since as soon as pairs are generated they are swept away by the enhanced electric field

$$G \sim W \sim \sqrt{V} \quad (4.2)$$

High level injection: As commented before we supposed that the density of majority carriers was not affected. However, as we increase the potential drop a large amount of carriers is taken from one side of the junction to the other, and injection for the majority carriers must be considered. When this is done we can see that the forward bias is dominated by the following form of the current:

$$I \sim e^{qV/2kT} \quad (4.3)$$

Ohmic losses: The potential drop inside the p and n regions, as well as that in the contacts, can be taken into account by introducing a series resistance Rs in the diode, which will influence the measurements as we use an external tension $V = V_{\text{ext}} - IR_s$.

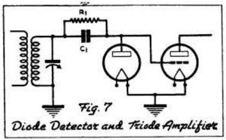

Fig. 7

Diode Detector and Triode Amplifier

The diode detector is commonly used today. It can handle very strong signals without overloading, has excellent fidelity, and is simple and reliable. However, it isn't extremely sensitive, since the filter system imposes a load that draws power from the tuned circuit. This in turn reduces selectivity.

Fig. 7 shows a modified diode detector directly coupled with a triode amplifier. Here the filter circuit operates somewhat differently; when a positive R.F. signal comes along, the 250-mmf. condenser C_1 transmits it freely, applying a positive potential to the diode plate and drawing electrons from the filament until the charge on the condenser is neutralized. This may take less than a ten-thousandth of a microsecond. When the positive peak declines to zero, the electrons are trapped on one plate of the condenser, on the plate of the diode, and on the triode grid. Their only escape is through the high-resistance shunt, and it takes more nearly ten thousand microseconds than a ten thousandth of a microsecond for the charge to leak away. Thus the charge on the condenser can increase at radio frequencies but can decrease only at audio frequencies. Since it is applied directly to the triode grid, the triode plate current will follow A.F. modulation.

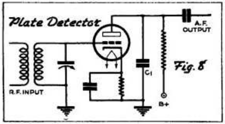

Plate Detector

Fig. 8

The grid of a triode is quite capable of collecting electrons from the filament, and consequently can serve as a diode plate in itself. In the grid-leak detector of years past, a single triode was used to perform this double function. The grid-leak detector was selective, though the fidelity was not high and it tended to overload on strong signals. It is seldom used today.

The circuit of a plate detector is shown in Fig. 8. The high value of the cathode resistance makes the grounded grid strongly negative with respect to the cathode, so much so as to cut off the plate current when there is no signal. When a negative signal reaches the grid, practically no change of plate current occurs, for it's already nearly zero. Bút when a positive signal arrives, the strong bias is decreased and plate current flows. This gives the nonlinear reaction necessary for demodulation. Condenser C_1 shunts the R.F. pulses to the ground. Able to handle fairly heavy signals, the plate detector shown gives good fidelity. It imposes no load on the tuned circuit and hence has good selectivity.

The infinite-impedance detector, so called because it has almost infinite resistance to R.F. between grid and ground, and hence places no load on the tuned circuit, is illustrated in Fig. 9. A positive signal applied to the grid permits the plate current to flow, and this causes a voltage drop through R_1 that drives the cathode positive. Since there is always some current flowing through the tube, the cathode is always somewhat positive with respect to the grid, and no grid current flows. Whatever the positive potential of the grid, the cathode will rise to match it, though it won't follow the negative swings as rapidly because of C_1. The result is that the potential across R_1 exactly follows the peaks of the R.F. signal. There is no amplification, but the detector has excellent fidelity, will handle strong signals, and is much used today.

Fig. 10 is the regenerative detector, of evil memory. It has excellent sensitivity and selectivity, but against these are poor fidelity, extreme unreliability, complexity of operation, and a bad tendency to turn into a transmitter able to send out squeals and howls into neighboring counties. Outlawed on broadcast frequencies now, it is still used by some amateurs on short-wave receivers. Basically it is a grid-leak detector so ar-

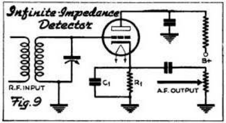

Infinite-Impedance Detector

Fig. 9

AND HOW THEY WORK

derstood phenomenon in which certain metallic oxides exhibit the ability of rectifying A.C. The plate-and-needle detector, essentially a metal needle resting lightly on a metal plate, embodies this principle, as does the razor-blade radio shown on page 209.

One of the most sensitive instruments known in the early days of wireless was the galvanometer. Unfortunately, it requires direct current; if it is fed A.C. of even moderate frequencies, the needle tries to swing one way on positive impulses and the other way on negative ones, with the result that it stands quite still. Before it could detect R.F., it had to be coupled with a rectifier that would intercept half the R.F. wave.

To improve the detector, the British Marconi Company called in Dr. J. A. Fleming, a leading English research physicist. Fleming had already done some notable research for a lamp-bulb manufacturer. The carbon-filament lamps of the time had a tendency to burn out at the positive end of the filament, due to the "Edison Effect." Edison had reported that a hot filament in an evacuated tube would retain positive but not negative charges—in modern terms, that it would give off electrons. Therefore electrons jumped from the negative end to the positive, carrying part of the current and heating the positive end by bombardment.

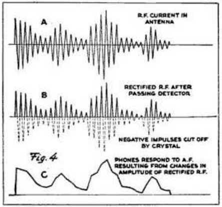
A
R.F. CURRENT IN ANTENNA
B
RECTIFIED R.F. AFTER PASSING DETECTOR
NEGATIVE IMPULSES CUT OFF BY CRYSTAL
Fig. 4
C
PHONES RESPOND TO A.F. RESULTING FROM CHANGES IN AMPLITUDE OF RECTIFIED R.F.

Dr. Fleming's studies of the Edison effect, plus his realization that the galvanometer would make a sensitive detector if it could be used with R.F. currents, gave him the answer—the Fleming Valve. It was a simple diode rectifier, a heated filament and a plate in an evacuated envelope, in series with a galvanometer and the antenna circuit. The diode permitted one half of the waves to kick the moving coil, but blocked the opposite set of half-waves. Before long headphones, taken over from the telephone, were used in place of the galvanometer and found unexpectedly efficient (Fig. 5).

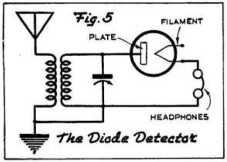

Essentially, a demodulating detector is a nonlinear element—one that will pass more current in one direction than in the other—followed by a filter circuit. (A diode can be made to pass 95 percent of one half-wave and none of the other; a crystal may pass 80 percent of one and 40 percent of the other. So long as conductivity is dissimilar, demodulation will take place.)

In the simple diode detector shown in Fig. 5, the headphones themselves constitute the filter; the phone coils have sufficient inductance and capacitance to make an excellent R.F. trap. In Fig. 6, which shows a more modern version of the diode detector, a resistance-capacitance filter is used instead. The condensers and resistors are of such values that the condensers discharge through the resistors only at comparatively low frequencies—15,000 cycles a second or less—with the result that the ultrarapid R.F. pulses are completely smoothed out and only modulation remains. *[Turn the page]*

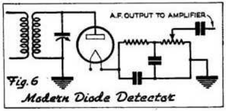

---

5 Experimental Procedure

In order to compare the theory of the ideal diode with the real diode behavior we will measure the current crossing a Germanium transistor ACY20 when a tension is applied across it, in forward and reverse bias.

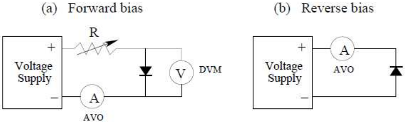

Figure 3. Experimental set-up for the forward and reverse bias configuration of the electric circuits to measure the I-V characteristic of the diode.

As shown in fig. 3a for forward bias an external voltage is applied to the diode and a series varistor, and the voltage drop on the diode, as well as the current, are measured. In fig. 3b the set-up for measuring reverse bias is shown. In this case the varistor is not necessary any more, since the variations in current as voltage is changed are small, and therefore the tension drop can be directly read from the voltage source.

At room temperature we measure the I-V characteristics by controlling the current for forward bias ($I < 40mA$), and taking care of obtaining enough data for small values of the voltage ($0V < V < 0{:}2V$) in order to capture the expected exponential behavior of the characteristic. For reverse bias we can control the voltage itself and stop the reading before reaching the breakdown point.

In order to verify the exponential behavior of the reverse saturation current I0 with temperature (and energy gap Eg or contact potential V0), we heat the sample, while in the reverse bias set-up, for a given value of the applied voltage, V = ¡5V , from room temperature up to a temperature of about t = 80oC.

A new measurement of the I-V characteristic at about t = 80C is performed following the same procedure used for the room temperature case. We should obtain an increment of the current, in both forward and reverse bias, due to the thermal activation of the density of carriers, as predicted by the theory.

6 Experimental Results

We present in this section our results obtained from the previously described experimental set-up for a germanium diode. First of all, in fig. 4 we can see the I-V characteristic for forward and reverse bias, at room temperature (t = 20dC) and t = 80dC. (Notice the different scale used for the positive and negative axes of voltage and current).

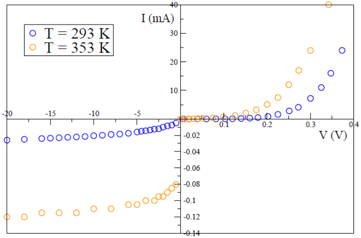

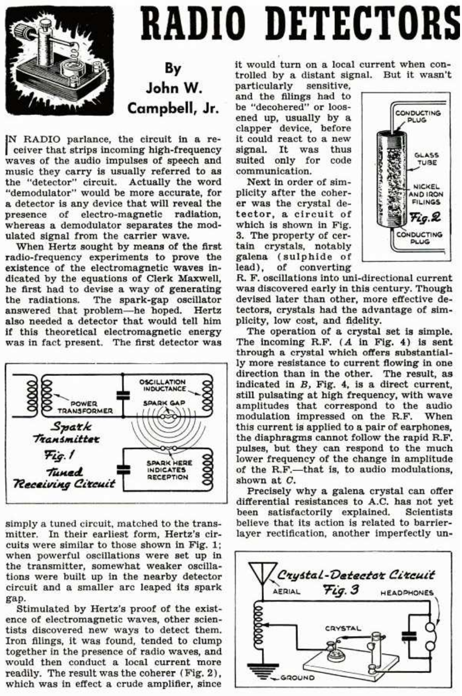

RADIO DETECTORS

By
John W.
Campbell, Jr.

IN RADIO parlance, the circuit in a receiver that strips incoming high-frequency waves of the audio impulses of speech and music they carry is usually referred to as the "detector" circuit. Actually the word "demodulator" would be more accurate, for a detector is any device that will reveal the presence of electro-magnetic radiation, whereas a demodulator separates the modulated signal from the carrier wave.

When Hertz sought by means of the first radio-frequency experiments to prove the existence of the electromagnetic waves indicated by the equations of Clerk Maxwell, he first had to devise a way of generating the radiations. The spark-gap oscillator answered that problem—he hoped. Hertz also needed a detector that would tell him if this theoretical electromagnetic energy was in fact present. The first detector was simply a tuned circuit, matched to the transmitter. In their earliest form, Hertz's circuits were similar to those shown in Fig. 1; when powerful oscillations were set up in the transmitter, somewhat weaker oscillations were built up in the nearby detector circuit and a smaller arc leaped its spark gap.

Stimulated by Hertz's proof of the existence of electromagnetic waves, other scientists discovered new ways to detect them. Iron filings, it was found, tended to clump together in the presence of radio waves, and would then conduct a local current more readily. The result was the coherer (Fig. 2), which was in effect a crude amplifier, since it would turn on a local current when controlled by a distant signal. But it wasn't particularly sensitive, and the filings had to be "decohered" or loosened up, usually by a clapper device, before it could react to a new signal. It was thus suited only for code communication.

Next in order of simplicity after the coherer was the crystal detector, a circuit of which is shown in Fig. 3. The property of certain crystals, notably galena (sulphide of lead), of converting R. F. oscillations into uni-directional current was discovered early in this century. Though devised later than other, more effective detectors, crystals had the advantage of simplicity, low cost, and fidelity.

The operation of a crystal set is simple. The incoming R.F. (*A* in Fig. 4) is sent through a crystal which offers substantially more resistance to current flowing in one direction than in the other. The result, as indicated in *B*, Fig. 4, is a direct current, still pulsating at high frequency, with wave amplitudes that correspond to the audio modulation impressed on the R.F. When this current is applied to a pair of earphones, the diaphragms cannot follow the rapid R.F. pulses, but they can respond to the much lower frequency of the change in amplitude of the R.F.—that is, to audio modulations, shown at *C*.

Precisely why a galena crystal can offer differential resistances to A.C. has not yet been satisfactorily explained. Scientists believe that its action is related to barrier-layer rectification, another imperfectly un-

Figure 4. I-V Characteristics: values for t = 20dC and t = 80dC. Notice that the scale for positive and negative axis is different.

For forward bias we notice an exponential behavior (a more detailed study is presented in fig. 5) of the characteristics, and how the higher temperature characteristic allows for higher values of current in agreement with the assumption of a larger number of carriers available due to thermal activation.

In the reverse bias scenario we can see how the current tends to saturation. A saturation point is not reached probably because of generation current within the depletion region, or other kind of non-ideal mechanism as we will discuss later. Again the higher temperature curve yields a higher current.

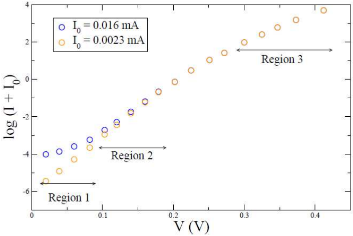

Figure 5. Ideality Factor: measurements in forward bias. Three different regions are identified, corresponding to different dominant mechanisms. Low voltage (region 1), low-intermediate voltage (region 2), and large voltage (region 3) regions are indicated in the figure.

The confirmation of the exponential behavior for the forward bias characteristic is presented in fig. 5. According to eq. (3.2), by plotting log(I + I0) against V we should obtain a straight line if the diode followed the ideal behavior. When non-ideal mechanisms are considered, as those presented in section 4, different regimes should be identified as the slope of the curve changes, since

$$\frac{qV}{kT} \longrightarrow \frac{qV}{AkT} \quad (6.1)$$

where A is known as the ideality factor. Usually A = 1 for the ideal mechanism, corresponding to low injection current, and A = 2 for high injection current (large voltage) or recombination current within the depletion region. In general intermediate values are expected since all the effects work at once.

There is a further uncertainty, since we do not know with precision the value of the reverse saturation current I0. Although in principle we are asked to take the value for a voltage of V = ¡5V , I0 = 0:016mA, we will argue that it is as reasonable to choose as saturation current the value for V = ¡0:05V , I0 = 0:0023mA. Other intermediate values of the saturation current would lead to intermediate results for ideality factors in the low voltage region. Curves for both saturation currents chosen are shown in fig. 5.

We can divide the points in fig. 5 in three regions, low voltage (0V < V < 0:1V , (region 1), low-intermediate voltage (0:1V < V < 0:2V , region 2), and large voltage (V > 0:3V , region 3) regions. By determining the slope of the linear fitting in each region we can obtain different ideality factors for each region and curve (m = q=AkT, see eq. (6.1)). The results for A are presented in table 1.

low band. About 1-2 wire diameter spacing between the two windings is recommended.

experienced when using a conventional inductor. If greater selectivity is needed at the high end of the BC band when using a conventional inductor, antenna coupling must be reduced and/or the diode must be tapped down on the tank to raise the loaded Q. Either approach results in a greater insertion power loss and a weaker or inaudible signal to the phones when tuning stations near the high end of the BC band . The low inductance (parallel connection) of the contra-wound inductor enables a 4 times reduction in bandwidth at 1710 kHz, compared to results with conventional inductor. This reduces the need to tap the diode down on the tank and re-match the antenna when one needs to increase selectivity, as mentioned above.

Note:
One could use two separate conventional non-coupled inductors, one of 250 uH and the other of 62.5 uH, instead of a contra-wound configuration. This is not recommended because the Q of the 62.5 uH inductor will probably be less than that of the 250 uH unit unless it is made physically as large as the contra-wound coil.and employs larger diameter wire. Also, when using the contra-wound approach the hot end of the inductor, when the two coils are connected in parallel, can be in the center of the overall unit, with the outer wire ends of the assembly placed at ground potential. This reduces electric field coupled losses from end mounting brackets and surroundings.

The inductances of the two connection configurations (parallel and series) of a contra-wound coil will depend upon how closely spaced the two windings are placed, but, the ratio of the inductance of the series to that of the parallel connection always remains at 4 times no matter how far or close together the windings are placed. Remember that overall distributed capacity is greater when using the parallel connection in the

---

Region	Io (mA)	A
1	0.016	3.10
2	0.016	1.45
3	0.016	2.58
1+2	0.0023	1.31

Table 1. Ideality factor for the different regions and curves in fig. 5

Our first curve (Io = 0:016mA) presents, in the low voltage region, a much larger value of A than expected from a recombination in the depletion region mechanism. In the low-intermediate region we could suggest a mix presence of recombination and ideal mechanisms. For the high voltage region a high value of A = 2:58 is obtained, which suggests that not only high level mechanisms are present. For slightly smaller values of the voltage drop (0:25V < V < 0:35V) a better agreement with high level injection current can be achived. The large value of A suggests that ohmic losses must be considered, which reduces the net current through the device.

We have considered as well the value I0 = 0:0023mA because at room temperature, kT » 25meV , the saturation of the current in reverse bias should be obtained for values of the reverse voltage as small as V » ¡0:05V . The existence of other effects in reverse bias (ref. [1], fig. 5-37, page 218), as pair generation in the depletion region, enhances the current, which will be misleading when trying to fit the ideal behavior. The smaller value of the saturation current in this case smoothers the curve for low voltages, and we can obtain a nice linear behavior spanning regions 1 and 2, with a value of the ideality factor A = 1:31, which is closer to ideallity. This value can be obtained when considering recombination in the depletion region due to interphase recombination, or defects in excess to what would be ideal.

Eq. (3.9) provides an explicit temperature dependence of the reverse saturation current with temperature, namely, an exponential behavior (provided we assume that the temperature dependence of the terms in square brackets cancels each other, or at least is subdominant). We can use the data of the saturation current, measured at V = ¡5V in the range of temperatures of t = 20d ¡ 80dC to obtain a value for the contact potential.

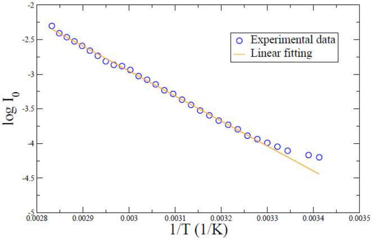

Figure 6. The contact potential is obtained from the slope of the linear fitting. This value is to be compared with the energy gap for Germanium. In the linear fitting the five most right experimental points where dropped.

In fig. 6 we plot the log I0 against 1=T , and from the slope of the linear fitting (m = 3607:9K) the value of the contact potential is V0 = 0:31V . This value of the contact potential is directly related to the value of the gap of the semiconductor

---

The value needed, using a contra-wound approach is 138.6 pF. One can derive, from data values in Figs. 1 - 4 in Article 28, that the Q of the common 365 pF, non-ceramic insulated variable capacitor (capacitor B), at 1710 kHz comes out as follows:

- If one uses a conventional 250 uH inductor tuned by 20 pF stray capacity with 14.65 pF more from the variable capacitor, the capacitor Q comes out at about 460.
- If one uses a contra-wound inductor that has 62.5 uH inductance with the two windings in parallel, tuned by 20 pF stray capacity with 118.6 pF more from variable capacitor B, the Q comes out at about 1770, 3.5 times as great! This translates directly to greater sensitivity and selectivity when using the commonly available 365 pF capacitor.

From Fig. 3 in Article #24 we can see that, at 1710 kHz, the Q of capacitor A, a ceramic-insulated, with silver plated plates capacitor manufactured by Radio Condenser Corporation, or its successor TRW, has a Q of 9800. This is much higher than that of capacitor B when using a conventional 250 uH inductor. Changing to a contra-wound coil while using the easily available capacitor B goes a long way toward a goal of reducing the effect of the variable capacitor on tank Q and loss at the high end of the band.

Less selectivity variation and less insertion power loss: Conventional inductor: The 3 dB down RF bandwidth will vary from 3.69 kHz at 520 kHz to 39.9 kHz at 1710 kHz, a variation of 11.6 times . Contra-wound inductor: The 3 dB down RF bandwidth will vary from 3.69 kHz at 520 kHz to 12.15 kHz at 943 kHz in the low band, and from 3.04 kHz at 943 kHz to 9.99 kHz at 1710 kHz in the high band, an overall variation of 4.00 times. This is about 1/4 of the variation

See 'The contra-wound tank inductor' in Part 3 of Article #26 and the paragraph after Figs. 2 and 3 in Article #29 for descriptions of two different contra-wound configurations.

Discussion:

Let us divide the BC band geometrically into two halves: This gives us 520-943 kHz for the low band and 943-1710 kHz as the high band. Assume, for ease of understanding, that the tank inductor for the conventional approach has an inductance of 250 uH.

Conventional 250 uH inductor: The whole BC band of 520-1710 kHz can be tuned by a capacitance varying from 374.7 to 34.65 pF.

Contra-wound 250/62.5 uH inductor: The low band of 520-943 kHz can be tuned, using the 250 uH series connection, by a capacitance varying from 374.7 to 113.94 pF. The high band of 943-1710 can be tuned, using the 62.5 pF parallel connection, by a capacitance varying from 455.76 to 138.60 pF.

For the purposes of this discussion, let us assume that antenna matching (see Part 2 of Article #22) is always adjusted to reflect a fixed shunt resistance of 230k ohms for driving the diode, over the full BC band. 230k ohms is also the RF input resistance of an ITT FO-215 germanium diode when fed a signal power well below its linear-to-square law crossover-point (see Article #10, points 1, 2 and 3 below Fig.1 in Article #15, Article 17A and Article #22). This setting approximates that for minimum insertion power loss (see Article #28).

Reduction of insertion power loss at the high end of the BC band (1720 kHz): The total tuning capacitance needed when tuning a conventional 250 uH inductor to 1710 kHz is 39.9 pF.

through eq. (2.6), which tells us that the value of the gap is given by the contact potential up to an additive term depending on the temperature, the doping, and the density of states. Therefore we can assume that the values of the gap and contact potential must be within the same order of magnitude, as it is the case since for Germanium Eg = 0:67eV .

As in the case of the ideality factor, here we also have to consider the uncertainty in the saturation current I0. The results for V = ¡5V could be affected by other effects which produce uncontroled deviations of the values of the contact potential and gap.

7 Conclusions

We have studied the behavior of a Germanium p-n junction, developing the simplest theory to explain it, namely, ideal behavior of an abrupt junction with low level injection current. We also outlined several corrections to this theory, which must be considered when analyzing the experimental data.

The I-V characteristic for the germanium diode was presented for room temperature (t = 20dC) and t = 80dC, showing the expected qualitative dependence with the applied potential, exponential increment with forward bias, and small values quite independent of the reverse bias.

A further study of the forward bias regime was presented to try to fit the different mechanisms involved in the transport of carriers through the junction. It was found that the low voltage region is consistent with the ideal behavior plus, possibly, generation within the depletion region. The high voltage region is consistent with a high level injection current mechanism and ohmic losses.

The dependence on temperature of the reverse saturation current I0 (V = ¡5V) yields a value for the contact potential generated by the junction, which is directly related to the energy gap of germanium. It was found that both values are within an order of magnitude, and the doping density and density of states account for the difference between them.

Acknowledgments

I would like to thank Rodrigo Ruiz Campo for his enthusiastic view of laboratory work, and very fruitful discussions. I also want to acknowledge discussions with Susana Fernandez Robledo.

References

[1] B. G. Streetman and S. Banerjee, Solid State Electronic Devices, Prentice Hall, fifth edition, 2000.

[2] S. M. Sze, Semiconductor Devices, Physics and Technology, John Wiley & Sons, 1985.

[3] J. P. McKelvey, Solid State and Semiconductor Physics, Harper & Row/John Weatherhill, 1966.

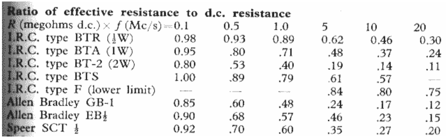

Ratio of effective resistance to d.c. resistance

R (megohms d.c.) × f (Mc/s) =	0.1	0.5	1.0	5	10	20
I.R.C. type BTR (½W)	0.98	0.93	0.89	0.62	0.46	0.30
I.R.C. type BTA (1W)	0.95	.80	.71	.48	.37	.24
I.R.C. type BT-2 (2W)	0.80	.53	.40	.19	.14	.11
I.R.C. type BTS	1.00	.89	.79	.61	.57	—
I.R.C. type F (lower limit)	—	—	—	.84	.80	.75
Allen Bradley GB-1	0.85	.60	.48	.24	.17	.12
Allen Bradley EB½	0.90	.68	.57	.46	.23	.15
Speer SCT ½	0.92	.70	.60	.35	.27	.20

12. The effects from using the contra-wound dual-value inductor configuration in crystal sets as compared to using a conventionally wound inductor, both using capacitive tuning

Some quick facts:

- Crystal sets using a conventional single-valued tank coil usually suffer from poor selectivity and sensitivity at the high end of the BC band.
- Use of both connections of a contra-wound dual-value inductor enables the achievement of much higher selectivity and sensitivity at the high end of the BC band (series connection for the low half and parallel for the high half of the BC band).
- There will be some small reduction in tank Q in the lower half of the BC band. One reason is that distributed capacity is greater in the series-connected contra-coil than in the conventional solenoid (the close-space adjacent ends of the contra-coil windings have 1/2 the tank voltage across them). Tank Q at the high end of the BC band is noticeably improved.
- It is assumed that comparisons between conventional and contra-wound inductors use coils having the same physical dimensions and wire specifications. The inductance of the conventional solenoid is assumed to be about the same as that of the series-connected contra-coil.

A fact of interest that some may not know is this: The AC resistance of carbon composition resistors, and film resistors, to a much lesser degree, decrease with increasing frequency (the Boella Effect). This effect is strongest in high value resistors, above, say, 22k ohms and above 50 MHz (film resistors). The effect is noticeable in 500k and 1 meg units at lower frequencies. Low value resistors having short leads and resistances in the mid 10s to mid hundreds of ohms are quite free of this effect up through many hundreds of MHz. A typical graph of the ratio of AC-to-DC resistance vs frequency, of various values of conventional commercial axial-lead carbon film type resistors, taken from a Brell Components catalog is .

FIG. 6 High Frequency Characteristic

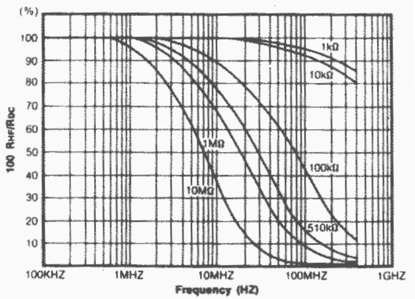

A chart providing similar info on carbon composition resistors, taken from the Radiotron Designer's Handbook, Fourth Edition, page 189 is .

DIODES

Dick Kleijer crystal-radio.eu

http://www.crystal-radio.eu/endiodes.htm

Which diode can I use best in my crystal receiver?

Maybe you think a diode with a voltage drop as low as possible, then also small signals at the detector circuit are detected. But diodes with a low voltage drop, also have a high reverse current (leaking current), this will load the detector circuit heavier, the Q of the detector circuit reduces, and with that also the voltage across the LC circuit. At a lower input voltage the diode will give much more losses, and it can happen that despite the lower voltage drop of the diode, you have less voltage at the load resistor. Besides that, reduction of circuit Q will also gives a less selective receiver.

For every 20 mV less voltage drop, the reverse current will approximately double.

Germanium, silicon, en schottky diodes. Depending on the material they are made from, we can distinguish germanium diodes, silicon diodes and schottky diodes. There are some more types, which are not discussed here.

Silicon diodes have the highest voltage drop (about 0.5 Volt) and are for this reason not very useable for crystal receivers. Unless we use a small DC bias current, which brings the diode already a little bit in conduction.

Germanium diodes have a low voltage drop (about 0.1 - 0.2 Volt) and are often used in crystal receivers. The properties like voltage drop and reverse current can vary a lot between two germanium diodes of the same type. In practice we can best test several germanium diodes in our receiver and then choose the best. The diode resistance RD of germanium diodes is most times rather low, and only useable in crystal receivers with a low Q (low sensitivity and low selectivity). For high performance receivers, we can better use a suitable schottky diode.

Schottky diodes have a voltage drop of about 0.25 Volt. The differences in properties between two diodes of the same type are often small. Schottky diodes with the correct resistance RD are very useable in high quality crystal receivers.

The given voltage drop is normally measured at a forward current of about 1 mA. Also if we measure the voltage drop of a diode with a multimeter, the test current shall be about 1 mA. But also below this voltage drop the diode can conduct current, and can rectify a RF (radio frequency) signal. Only the current through the diode is then much smaller. When receiving very weak stations, the current through the diode can be e.g. only 10 nA. At such a low current, the voltage drop of the diode is also much lower then at 1 mA.

Detected voltage as function of the input voltage. If we rectify a RF signal with a diode we can distinguish two situations.

Situation 1: Rectifying in the linear region

If the input voltage is high enough (well above the voltage drop of the diode at 1 mA), the output voltage of the diode will be about proportional to the input voltage.

So double input voltage, gives about double output voltage.

The output voltage is almost equal to the peak value of the input voltage.

The power losses in the diode are in this region very low compared to the rectified power.

The brand of resistor may be guessed by examining the smoothness and shininess of its surface finish, and looking at each end of the resistor to see where the wire exits. Allen Bradley resistors look the best. They have bright color code colors and a smooth shiny finish. At the wire exit point from the body one can usually see the appearance of a small shiny ring embedded in the plastic. Actually, this is part of the lead, shaped to be the contact electrode. Stackpole resistors look next best. They have somewhat duller colors on the color code and the surface is somewhat rougher and less shiny. The wires exit cleanly from the end of the resistor, no ring is visible. The Speer resistors have the dullest color code colors and a rougher surface than the Stackpole's. They usually look as if they have been wax impregnated. At the axial exit points from the body, a small copper colored dot may be seen next to the wire lead. This is actually the end of the lead, which was folded over and back on itself to form the electrode. The IRC so-called carbon comp. resistors can be identified by the visible 'mold-flash' marks on the body and ends. The colors are good, but the body is rough. Their end surfaces are slightly convex, not planar as in the case of the other resistors.

Remember, these resistors usually made spec. when new, passed incoming inspection and standard aging tests. Unfortunately, no aging tests could be made that covered the span of many decades.

It is interesting to note that the best resistors, from a long term resistance drift point of view turn out to be the AB units. They also cost the most. The Speer units cost the least and the Stackpole's were in between.

Ohmite carbon comp resistors I have seen looked like A-B units.

- Audio transformers having too low a shunt inductance will reduce bass response. When using magnetic headphone elements, this can be partially compensated for by connecting the transformer to the headphones using a suitable capacitor.
- Refer to Articles #2, #3,#5 and #14 for more info. Consider the 'Ulti-Match' by Steve Bringhurst at http://www.crystalradio.net/sound-powered/matching/index.html.

11. Long term resistance drift and frequency dependence of the AC resistance of low power resistors, etc:

From my early experience in the manufacturer of Blonder-Tongue products, the following is some insight relative to run-of -the-mill commercial carbon-composition resistors that we used:

The process used by the resistor manufacturer is an important factor in the determination of long term resistance drift. Allen-Bradley (A-B) used their 'hot-mold' process, producing a more dense product then did the other manufacturers, as far as I know. The value of this carbon comp. resistor drifts the least, as a rule. Stackpole composition. resistors used their 'cold-mold' process and seem to drift more than do the A-B units. Composition carbon resistors mfg. by the Speer company, using their 'cold-mold' process drift more than the Stackpole resistors, as a rule. The IRC resistors that look like carbon comp. units actually are made by another process. They are called metallized resistors. My impression is that their drift is similar to the of Stackpole resistors. I have found that the IRC resistors usually generate much more low frequency noise when passing a DC current than the others. It seems, as a general rule, that the high value resistors drift more, over time, than the low value ones.

Situation 2: Recifying in the square law region

If the input voltage is low, lower then the voltage drop of the diode (at 1 mA) then the situation is completely different.
The input of the diode behaves for the RF signal like a resistor with value RD.
The output of the diode behaves like a DC voltage source in series with a resistor, the value of this resistor is also equal to RD.
The value of the DC voltage source is square law related to the amplitude of the RF input signal.
So double input voltage, gives 4 times as much detected DC voltage at the output
In the square law region the output voltage of the diode will be much lower then the input voltage, the diode gives much power loss between input and output.

The lower the input voltage, the higher the losses.
The higher the input voltage, the lower the diode losses.
When further increasing the diode input voltage, we gradually come into the linear detection region.
When receiving weak stations, detection takes place in the square law region.

Between the linear and square law region, there is a region not linear, and not square law but somewhere in between.
This region is not discussed here.

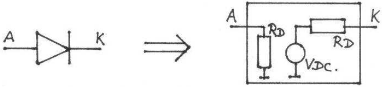

Equivalent circuit of a diode at low input voltages

Via this link you find a measurement on several schottky diodes, which shows detection in the square law region takes place at input voltages below 200 mVpp.

Diode resistance RD.

At zero voltage, diodes have a certain resistance.

This resistance at zero Volt we call RD.

The lower the reverse leaking current of the diode, the higher resistance RD.
When detecting small signals (in the square law region) the input of the diode also behaves like a resistor with value RD.

But how do we know the RD of a diode?

We can calculate it with the formula:
formula 1: RD = 0.000086171 x n x TK / Is
RD = diode resistance at zero Volt (unit: Ohm)
n = ideality factor, the lower this factor the better, between 1.0 and 1.1 is a very good value.
TK = temperature in Kelvin (= temperature in °C + 273)
Is = saturation current (unit: A)
x = multiply

The values of n and Is can (sometimes) be found in the diode datasheet.

In the following table some types of schottky diodes, with the values for n, Is and Rd, the maximum reverse voltage and the diode capacitance at zero voltage.

Here are some practical experimental ways to vary the audio source resistance of a crystal radio set when receiving weak-to-medium-strength signals. A medium strength signal is defined as one at the crossover point between linear to square law operation (LSLCP). See the graphs in Article #15A.

- Change the diode to one having a lower saturation current, such as from a germanium diode (1N34A) to one or several paralleled Schottky diodes such as the Agilent 5082-2835. Schottky diodes described as "zero bias detectors' have a high saturation current and are not suitable for most crystal radio set use. Schottky diodes described as "power rectifiers' usually have a high saturation current as well as a high junction capacitance. A high diode junction capacitance will reduce treble response. Too large a diode RF bypass capacitor in the crystal radio set can also reduce treble response. A side benefit from a change to a diode having a lower saturation current value, on some crystal radio sets is an increase in selectivity. This is because the RF load resistance presented by the diode to the tank is raised when the diode saturation current value is reduced. This reduced loading raises the tank Q and hence, increases selectivity.
- Use an audio transformer between the detector output and the phones. A smaller step-down transformer impedance transformation ratio will raise the transformed diode source resistance seen by the phones. A larger ratio will decrease it.
- If the headphone elements are in series, reconnecting them in parallel will reduce their impedance to 1/4 the previous value. This has the same effect as increasing the effective source resistance driving the headphones. If they are in parallel, series connecting them has the effect of decreasing the effective source resistance.

#9082 are suitable and are available from various distributors such as Alltronics, Digi-Key, etc.

Surface mount diodes manufactured using the SOT-23 package can be handled using Surfboard #6103. Diodes using the smaller SOT-323 package can be handled using Surfboard #330003. This includes many Agilent surface mount diodes useful in crystal radio sets. Packages containing multiple diodes exist that use the SOT-363 six lead package. They can be handled using Surfboard #330006. Agilent produces many of their Schottky diodes in dual, triple and quad form in the SOT-363 package.

It is recommended that anyone considering using Surfboards visit the above mentioned Website and read "Application Notes" and the "How-to Index".

10. How to modify the tone quality delivered by headphones: It is interesting to note that driving magnetic headphone elements with a high source resistance tends to improve the treble and reduce the bass response, compared to the response when the AC source resistance matches the effective impedance of the elements. Conversely, driving the headphones elements from a low resistance source tends to roll off the treble, and relatively speaking, improve the bass. With piezo ceramic or crystal elements, a high source resistance tends to reduce the treble and improve the bass response, compared to the response where the source resistance matches the effective impedance of the elements. A low source resistance tends to reduce the bass and emphasize the treble. Some piezo elements sound scratchy. This condition can be minimized by driving the elements from a lower resistance source.

type diode	n	Is at 25 ºC	RD at 25 ºC	maximum reverse voltage	capacitance	
5082-2835	1.08	22 nA	1260 kΩ	8 Volt	1 pF	datasheet
BAT85	?	?	± 200 kΩ ??	30 Volt	10 pF	datasheet
HSMS 2820	1.08	22 nA	1260 kΩ	15 Volt	1 pF	datasheet
HSMS 2850	1.06	3000 nA	9.07 kΩ	2 Volt	0.3 pF	datasheet
HSMS 2860	1.10	38 nA	743 kΩ	4 Volt	0.3 pF	datasheet

To decrease the value RD, we can connect more diodes in parallel, with two the same diodes parallel the value of RD shall halve.
With 3 diodes parallel, the value of RD shall be divided by 3 etc..

Diode resistance when using bias current.

We can decrease the value of RD by sending a small DC bias current (e.g. 0.1 uA) in forward direction through the diode.
The higher the bias current, the lower RD will be.

With the following formula we can calculate the diode resistance RD, when we make use of a DC bias current.

Formula 2: RD = 0.000086171 x n x TK /(Ib + Is)

RD= diode resistance at certain DC bias current Ib (unit: Ohm)
n= ideality factor of the diode
TK = Temperature in Kelvin (= temperature in ºC + 273)
Ib= DC bias current through the diode in A
Is =saturation current of the diode in A

A diode with a certain RD value at a certain bias current, gives the same receiving performance as a diode without bias current with the same RD.

Influence of temperature on "saturation current: Is"

The saturation current (Is-value) is strongly depending on temperature.
A temperature increase of 1 ºC will increase the Is value by about 7 %.
In datasheets, the Is value is most times given at 25 ºC.
If the diode temperature is not 25 ºC, but another value "T", then we must multiply Is with a factor 1.07^(T-25).
T = diode temperature
^ = raise to the power of

Ideality factor n

The ideality factor n of a diode indicates how good the diode performs with regard to an ideal diode.
A (not existing) ideal diode has a value of n=1.
At low input signals, the maximum available detected output power is proportional to 1/n.
So doubling the n will halve the output power (this only applies at weak signals).

signals, each of power P/2 and that there will be four output components, as stated above. They are:

- •The two second harmonic components (both of the same frequency and phase).
- •The sum frequency component (f1+f2) Hz, which will be of the same frequency and phase as the second harmonic components since f1=f2.
- •The difference frequency component (f1-f2) at a frequency of zero Hz.
- •If we filter the harmonic and sum components as well as the two original signals from the output, only the zero Hz signal will remain; and we call it the detected DC output.

A diode detector can be thought of as a "Black Box". If the DC output impedance of the detector is matched to its load resistor and the AC signal power source of P Watts 'available power' is impedance matched to the input AC impedance of a diode detector, the DC output power can closely approach the 'available power' from the AC source. This gives us another way to look at a detector. It can be considered to be a "Black Box" that changes incident AC power of frequency "f" Hz into output power of frequency zero Hz (DC). This is the detected DC output.

9. Using surface mount components in crystal radio sets:
A convenient way to connect to the tiny leads of small surface-mount diode and IC devices is to first solder them to a "Surfboard". Pigtail leads can them be soldered through holes drilled in the Surfboard conducting races for connection to a circuit.
A surface mount device such as the OPA-349 integrated circuit (Eight lead SOIC package) can be soldered to a surfboard such as that manufactured by Capital Advanced Technologies (http://www.capitaladvanced.com). Their Surfboards #9081 or

voltage (remember the equivalent voltage is 3 times the original source internal voltage). The 1/2 comes from the 2:1 voltage division between the resistance of the equivalent source of 90,000 Ohms and the detector input resistance of 90,000 Ohms. The ratio of the new detector voltage to the old is: 3 times 1/2 divided by 0.9 = 1.67 times. This equates to a 4.44 dB increase in power applied to the detector. If the input signal to the detector is so weak that the detector is operating in the square-law region, the audio output power will increase by 8.88 dB! This is about a doubling of volume.

7. Caution to observe when cutting the leads of a glass Agilent 5082-2835 Schottky diode (or any other glass diode): When it is necessary to cut the leads of a glass packaged diode close to the glass body, use a tool that gives a scissors type of cut. Diagonal cutters give a sudden physical shock to the diode that can damage its electrical performance. This physical shock is greater than one might expect because of the use of plated steel instead of more ductile copper wire. Steel is used, in part, because of its lower heat conductivity, to reduce the possibility of heat damage during soldering.

8. Several different ways to look at a diode detector: A diode detector can be thought of as a mixer, if one thinks of its input signal as consisting of two identical signals of equal power, in phase with each other. It is well known that if a common AM mixer is fed with two signals of frequencies f1 and f2 Hz, most of the output it generates will consist of the second harmonic of each signal and two more signals at other frequencies. One is at the sum frequency (f1+f2) Hz and one at the difference frequency (f1-f2). Additional mixer products can be generated, but they will be weaker than those mentioned and will be neglected in this discussion. In the case of an AM diode detector, we may consider that its input signal of power P Watts is in reality the sum of two equal in-phase

Diode capacitance

Between the two connections of the diode there will be a certain capacitance (capacitor value), when this capacitance is fairly high (e.g. 10 pF) the tuning range at high frequencies is limited.
At increasing reverse voltage across the diode the capacitance will reduce, also the detected voltage in a crystal receiver is such a reverse voltage.

Through this, the frequency of the circuit can shift upwards when receiving strong signals.
On the next page: experiments with a detector unit you find in table 3 a measurement about the frequency shift.

Measuring the Is value of a diode.

We can measure the Is value of a diode as follows:
Send a small current through the diode, the value of this current (ID) must be about 1 μA.
Measure the voltage across the diode (VD).

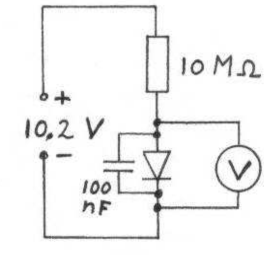

Circuit diagram for measuring the Is value of a diode.
The voltage across the diode is about 0.2 Volt.
The voltage across the resistor is about 10 Volts, so the current is about 1 μA.
The voltmeter must have a resistance of at least 10 MΩ.
The 100 nF capacitor reduces the influence of radio signals and hum on the measurement.

Calculate Is with the formula:

formula 3: Is = ID / (e^ (VD /(0.0257xn))-1)
Is = saturation current of the diode in nA
ID = current through the diode in nA, (1 μA = 1000 nA)
e = base of the natural logarithms, this is about 2.718
^ = raise to the power of
VD = voltage across the diode in Volt
n = ideality factor of the diode, if you don't know this value, take for instance: n= 1.08

An impedance-matched condition occurs when the resistance component of the input impedance of the crystal radio set equals the source resistance component of the impedance of the antenna-ground system. Also, the reactive (inductive or capacitive) component of the impedance of the antenna-ground system must see an opposite reactive (capacitive or inductive) impedance in order to be canceled out. In the impedance-matched condition, all of the maximum available power (See section on "Maximum Available Power" above) intercepted by the antenna-ground system is made available for use in the crystal radio set and none is reflected back towards the antenna to be lost.

Now we are at the point where confusion often exists: The voltage concept vs. the power concept. Let's assume that the diode detector has a RF input resistance of 90,000 Ohms. Assume that the antenna-loaded resonant resistance of the tuned circuit driving it is 10,000 Ohms. If one uses voltage concepts only, one might think that this represents a low loss condition. NOT SO! After all, 9/10 of the actual source voltage is actually applied to the detector. If one impedance matches the 10k ohm source RF resistance to the diode 90k ohm RF resistance via RF impedance step-up transformation (maybe connecting the antenna to a tap on the tuned circuit, and leaving the diode on the top), good things happen. (We will assume here that, in the impedance transformation to follow, the ratio of loaded-to-unloaded Q of the tuned circuits is not changed.) For an impedance match, the tuned circuit resonant resistance should be transformed up by 9 times. If this was done by a separate transformer (for ease of understanding) it would have a turns ratio of 1:3, stepping up the equivalent source voltage by 3 times and changing the equivalent source resistance to 90,000 Ohms. What now? Before matching, the diode got 9/10 of the source voltage applied to it. Now it gets 1/2 the new equivalent source and spreads out as it goes away from the antenna. One can prove that power is radiated by substituting a LED diode for the regular diode, getting physically close enough to the station and then tuning it in. The LED will light up (give off light power), showing that some power is being broadcast and that it can be picked up. Now back at home, if one tunes in the station one gets sound in the headphones. What activates one's hearing system is the power of the perceived sound. BTW, if one gets too much sound power in the ear for a long enough time, the power can be strong enough to break off some of the hair cells in the inner ear and reduce one's hearing sensitivity forever. The theoretical best one can do with a crystal radio set setup is the following: (1) Use an antenna-ground system to pick up as much as possible of the RF power passing through the air in its vicinity . In general, a higher antenna will pick up more power from the passing RF waves than will a lower one. (2) Convert the intelligence carrying AM sideband RF power into audio electrical power. (3) Convert the electrical audio power into sound power and get that power into the ear.

There are power losses at each of the three steps and our job is to minimize them in order to get as much of the sideband RF power passing through the vicinity of the antenna (capture area) changed into audio power for our ears. We want all of the "available power" at the antenna-ground system to be absorbed into the crystal radio set then passed on through it to our headphones as sound. However, some of it will be unavoidably lost in the RF tuned circuit. If the input impedance of the crystal radio set is not correctly matched to the impedance of the antenna, some of the RF power hitting the input to the crystal radio set will be reflected back to the antenna-ground system and be lost.

More information about measuring the Is you can find on the website of Ben Tongue , in his articles number 4 and 16.

I measured the Is value of several diodes, and calculated the diode resistance RD.
Also some European germanium types are measured.
Several diodes are measured of the type OA95 and AA119

Diode	VD (Volt) at 1 µA	Is (nA)	RD (kΩ).
HSMS282K	0.1341	7.9	3428
HSMS282K 2 parallel	0.118	14.5	1867
HSMS286K (1 diode)	0.1116	18.3	1479
5082-2800	0.1871	1.14	23756
5082-2835	0.1464	5.04	5373
5082-2835 2 parallel	0.1289	9.5	2850
BAT 82	0.136	7.3	3710
BAT 85	0.0686	90.8	298
OA81 (germanium)	0.0225	800	34

OA95 #1 (germanium)	0.0272	600	45
OA95 #2	0.0221	821	33
OA95 #3	0.0271	604	45
OA95 #4	0.0304	502	54
AA116 (germanium)	0.0441	256	106
AA119 #1 (germanium)	0.0320	461	59
AA119 #2	0.0363	370	73
AA119 #3	0.0428	272	100

The HSMS282K is the same as the HSMS2820, only the HSMS282K has 2 equal diodes in one package.
The Is value of the HSMS282K, the HSMS286K and the 5082-2835 is lower than the value in the datasheet, this has also been noticed by other people.

Ben Tongue wrote me, that the Is value of the 5082-2835 has been reduced over the years by the manufacturer.
Also temperature has big influence, I measured at 18 ºC, in datasheets the Is value is given at 25 flC.
A increase from 18 flC to 25 flC will increase the Is by 60 %.

power (1.0^2)/Ri to equal the output power [(1.0*2*sqrt2)^2]/Ro, the input RF resistance (Ri) must equal 1/8 Ro. That is, Ri=Ro/8. This illustrates the direct interaction between the RF input resistance and output audio resistive load. At high input power levels selectivity drops substantially if the output resistive audio load value is lowered.
The audio output resistance of the detector approaches 8 times the RF source resistance driving it. This fact is seldom recognized and it may be the cause of some of the problems encountered by those experimenting with doublers.

4) Half-wave voltage doubler operating at a low input signal power level: The detector, in this case, does not operate as a peak detector, and it has significant power loss. At low input signal power levels Ri approaches (0.026*n/Is)/2 ohms and becomes independent of the value of Ro.

The audio output resistance of the detector approaches twice the axis-crossing resistance of the diode.

5) Summary: At high input power levels, and with both input and output matched, power loss in both half wave and half wave voltage doubling detectors approaches zero dB. Sound volume should be the same with either detector. At low input power levels both detectors exhibit substantial power loss. I believe, but have not proven, that at low input power levels the doubler has a higher power loss than the straight half wave detector, and should deliver less volume.

6. Some misconceptions regarding Impedance matching and Crystal Radio Sets: To understand the importance of impedance matching, one must first accept the concept of power. A radio station accepts power from the mains and converts some of it to RF power which is radiated into space. This power leaves the transmitting antenna at the speed of light

1) Conventional half-wave detector operating at a high input signal power level: The detector, in this case, operates as a peak detector. Since it is a passive device, its output power will approximately equal its input power, under impedance-matched conditions. The output DC voltage will approach sqrt2 times the input RMS voltage, since the peak value of a sine wave is sqrt2 times its RMS value. For the input power, (1.0^2)/Ri, to equal the output power, [(1.0*sqrt2)^2]/Ro, the input RF resistance (Ri) must equal 1/2 Ro. That is, Ri=Ro/2. This illustrates the direct interaction between the RF input resistance and output audio resistive load. At high input power levels selectivity drops when the resistive audio output load value is lowered. The audio output resistance of the detector approaches 2 times the RF source resistance driving it. If the diode were an ideal diode, the word "approximately" should be eliminated, and "approaches" should be changed to "becomes" .

2) Conventional half-wave detector operating at a low input signal power level: The detector, in this case, does not operate as a peak detector, and exhibits significant power loss. At low input signal power levels Ri approaches 0.026*n/Is ohms (diode axis-crossing resistance) and becomes independent of the value of Ro.

The audio output resistance of the detector approaches the same value as the axis-crossing resistance (see above).

3) Half-wave voltage doubling detector operating at a high input signal power level: The detector, in this case, operates as a peak detector. Since it is a passive device, its output power will approximately equal its input power, under impedance-matched conditions. The output DC voltage will approach 2.0*sqrt2 times the input RMS voltage, since the peak of a 1.0 volt RMS sine wave is sqrt2 times its RMS value. For the input

Diode forward voltage drop:
David Knight

http://www.g3ynh.info/circuits/Diode_det.pdf

All of the diode detectors described above have at least one diode forward voltage-drop in the path to the measuring device. The diode forward voltage drop, Vf, varies approximately in proportion to the logarithm of the current, If, passing through the diode, and also depends on the temperature of the diode junction. For many types of diode, this behaviour is described to a good approximation by a modified form of the Ebers-Moll equation[12]:

$$Vf = m (kT/q) \ln[(If / Is) +1]$$

where k is Boltzmann's constant (1.380662 10 Joules/Kelvin) , q is the charge of an electron (1.6021892 10 Coulomb), and T is the temperature in Kelvin, i.e., the absolute temperature (flC+273.16). The factor kT/q is sometimes given the symbol VT, and has a value of 25.3mV at 20flC. "m" is a dimensionless correction factor between 1 and 2. "ln" is the natural logarithm (Loge). Is is the junction reverse saturation leakage current.

For small-signal rectifier voltmeters, diodes should be chosen for low forward voltage drop and low junction capacitance. Most sources of information now maintain that the best diodes in this respect are silicon Schottky-barrier (i.e., silicon-metal junction) diodes, such as the 1N5711 (Agilent 5082-2800) [18]; and so to investigate this matter, the forward voltage drop versus current characteristics of a variety of small signal diodes were measured. The results are shown in the graph below:

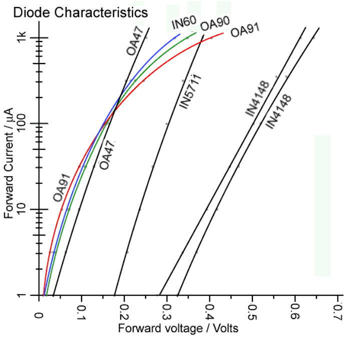

The 1N4148 is a silicon P-N junction diode. The 1N5711 is a silicon Schottky diode. The OA47 is an archaic germanium gold-bonded diode, and the rest are germanium point-contact diodes. Two 1N4148 diodes from different manufacturers were measured merely to illustrate the point that silicon P-N diodes have the highest forward voltage drop and are therefore a poor choice for low voltage detectors. The 1N5711 curve is the average of results from four diodes, all from the same batch, which had practically identical characteristics. The OA47 curve is the average for four diodes from two manufacturers, all having similar characteristics. The OA90, OA91, and 1N60 curves are from single examples, and are therefore not necessarily representative of the type. All data were recorded at an ambient temperature of 21flC.

signals are received, since both input and output ports become impedance matched.

Here is an interesting conceptual view of a high signal level diode detector circuit: Assume that it is driven with a sufficiently high level sine wave voltage so it operates in its peak detection mode, and is loaded with a parallel RC of a sufficiently long time const ant. This detector may be thought of as a low loss impedance transformer with a two-to-one impedance step up from input to output, BUT having an AC input and a DC output, instead of the usual AC input and output. The DC output power will approximately equal the AC input power and the DC output voltage will be about sqrt 2 times the RMS AC input voltage.

5A. A comparison of conventional half-wave and half-wave voltage-doubling detectors: Here is some info that may be of interest re conventional half-wave detectors vs. voltage doubling half-wave detectors when each is terminated with an output load of Ro. For illustration purposes we will assume the input voltage to the detector to be 1.0 volt RMS. The RF input resistance of the detector will be designated as Ri. All diodes have the same Is and n. It is assumed that good diodes such as a 5082-2835 Schottky, ITT FO-215 germanium or other are used. The info relates to the RF input resistance of detectors (it has a large effect upon selectivity) and their output audio resistance. See Point 4 in this Article for info on diode Is and n.

A high input power level is defined as one that is high compared to that at the LSLCP of the detector. A low input power level is defined as on that is low compared to that at the LSLCP of the detector. See "Quick Summary" in Article #15 for info on LSLCP.

resistance. Further, since the detector is now a peak detector, the DC output voltage is the "square root of 2" times larger than of the applied input RF RMS voltage. (It's equal to the peak value of that voltage). These existence of these relationships is necessary so that in an ideal peak detector, the output power will equal the input power (No free lunch). Summary: Output DC voltage equals sqrt2 times input RMS voltage. Since the output power must equal the input power, and power equals voltage squared divided by resistance, the output load resistance must equal two times the source resistance, assuming impedance matched conditions prevail. If we were to adjust the input source resistance to, say 495k ohms (reduce it by sqrt2) and the output load resistance to 990k ohms (increase it by sqrt2) by changing the input and output impedance transformation ratios, the insertion loss would become even lower than before the change and the input and output impedance matches would be very much improved (remember we are now dealing with high signal levels).

A good compromise impedance match, from one point of view, occurs if one sets the RF source resistance to 0.794*Rd and the audio load resistance to 1.26*Rd. With this setup, theoretically, the impedance match at both input and output remains very good over the range of signals from barely readable to strong enough to produce close to peak detection. A measure of impedance match is "Voltage Reflection Coefficient", and in this case it is always better than 18 dB (VSWR better than 1.3). Excess insertion loss is less than 1/3 dB and selectivity is largely independent of signal level. Information presented in Article #28 shows that, if the diode load resistance is made equal to Rd and the RF source resistance is made equal to Rd/2, the weak signal output of the detector will be about 2 dB greater than if both ports are impedance-matched! There is little benefit when strong

---

The data indicate that the IN4148, the 1N5711, and the OA47, all obey a logarithmic V/I relationship rather well, whereas the germanium point-contact characteristics show considerable curvature due to high internal resistance. With regard to the forward drop however, the germanium diodes are all superior to the IN5711 in the 1-100 A range, and the preference for the latter may merely reflect the fact that many semiconductor manufacturers can no longer fabricate germanium. Silicon Schottky diodes, such as the 1N5711 and 1N6263, being essentially UHF devices, have better high-frequency performance than germanium diodes, but germanium diodes work well at VHF and are therefore perfectly adequate for HF applications. Among the germanium diodes, there is little difference between the gold-bonded and standard varieties in the 1 to 100 A range, but the OA47 is the best choice for currents up to 1mA. We should observe however, that detector diodes only conduct on the peaks of the applied waveform, and so the instantaneous current is much higher than the average current, the difference being about an order of magnitude. Therefore, in selecting diodes for average currents in the region of 1 - 100 A, we should consider the steady-state voltage drop in the region 10 A - 1mA; in which case the germanium gold-bonded diode offers the lowest forward-drop without contest. Note however, that one of the consequences of the Ebers-Moll equation is that low forward voltage-drop is associated with high reverse leakage current. If reverse leakage is an issue, then silicon Schottky diodes are to be preferred. To put this issue in perspective however, the reverse leakage current of an OA47 diode was measured as follows:

Vr /Volts	1.0	2.0	3.0	5.0	10.0	15.0	20.0
Ir /uA	1.1	1.3	1.4	2.1	4.1	6.3	8.8

The leakage current is approximately linear in the 5-20V range and may be modelled by assuming a resistor of about 2M in parallel with the diode. Such a defect has little effect on the operation of a detector loaded with a 10-100K resistor. Some final points in favour of the silicon Schottky diodes however, are that germanium diodes show a wider spread of characteristics, and that the OA47 is obsolescent. Hence the Schottky diodes are definitely preferable in applications requiring diode matching, precise calibration, or availability through normal commercial channels. The 1N5711 in particular also, has a very high reverse breakdown voltage for a device of its class, its Vr max of 70V making it suitable for half-wave detectors of up to 24.7V DC ouput. An OA47 half-wave detector has a maximum DC output of 10.6V if Vr max is not to be exceeded.

Detector diode data: Source: refs. [18], [19]

---

having 3 dB less potential weak signal output than one of the diodes by itself.

5. Explanation of why, in a diode detector, and by how much, the RF input resistance and audio output resistances change as a function of input signal power. Consider first, a diode detector that is well impedance-matched both at its input and its output when driven by a very low power RF input signal. There will then exist an appreciable power loss in the detector. (The audio output power will be appreciably less than the input power.). The input and output impedances of the detector will approximately equal each other and approach: Rd = 0.026*n/Is. See part 3 above for a definition of terms. For this illustration, let the diode have an Is of 38 nA and an n of 1.02. Rd will be 700k Ohms. The well impedance-matched condition will hold if the input power is raised from a low value, but only up to a point. After that , the match will start to deteriorate. At an input power about 15 dB above that of the square-law-linear crossover point, the match will have deteriorated to a VSWR of about 1.5:1 (VSWR = Voltage Standing Wave Ratio.). A further increase of input signal power will result in a further increase of VSWR. This means that the input and output resistances of the detector have changed from their previously matched values.

The input resistance of the diode detector decreased from the value obtained in the well impedance matched low power level situation. The output resistance increased. The reason for this change is that a new law now governs input and output resistance when a diode detector is operated at a high enough power level to result in a low detector insertion power loss. It now operates as a peak detector. The rule here is that the CW RF input resistance of a diode peak detector approaches ½ the value of its output load resistance. Also, the audio output resistance approaches 2 times the value of the input AC source crystal radio set operation, but varies with diode current in silicon pn junction and germanium point contact diodes. A way of thinking about n is to consider it as a factor that effectively reduces the applied signal voltage to a diode detector compared to the case of using an ideal diode having an n of 1.0. Less applied signal, of course, results in less detected output.

Here are a few bits of information relative to diodes:
Typically, if a diode is biased at 0.0282*n volts in the forward direction, it will pass a current of 2 times its Is. If it is biased at 0.0182*n volts in the reverse direction, it will pass a current of 0.5 times its Is. If a diode is biased at 0.0616*n volts in the forward direction, it will pass a current of 10 times its Is. If it is biased at -0.0592*n volts, it will pass a current of -0.9 times its Is. These values are predicted from the classic Shockley equation. In the real world, reverse current can depart substantially from values predicted by the equation because of effects not modeled (the reverse current becomes higher). Gold bonded germanium diodes usually depart somewhat from the predicted values when operated in the forward direction. The effect appears as an increase of Is when measurements are made at currents above about 6 times the low-current Is.

Values of Is and n determine the location of the apparent 'knee" on a linear graph of the diode forward current vs. forward voltage. See Article #7. An easy way to estimate the approximate value of Is can be found in Article #4, section 2. A method of measuring Is and n is given in Article #16.

If one connects two identical diodes in parallel, the combo will behave as a single diode having twice the Is, and the same n as one of them. If one connects two identical diodes in series, the combo will behave as a single diode having twice the n and the same Is as one of them. This connection results in a diode

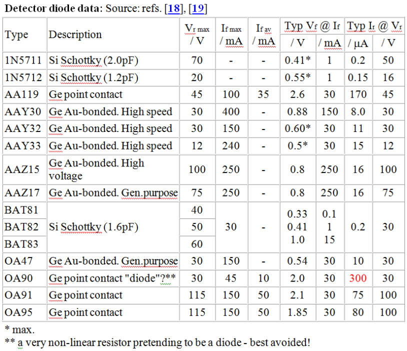

Detector diode data: Source: refs. [18], [19]

Type	Description	$V_{r\,max}$ / V	$I_{f\,max}$ / mA	$I_{f\,av}$ / mA	Typ V_f @ I_f		Typ I_r @ V_r	
					/ V	/ mA	/ µA	/ V
1N5711	Si Schottky (2.0pF)	70	-	-	0.41*	1	0.2	50
1N5712	Si Schottky (1.2pF)	20	-	-	0.55*	1	0.15	16
AA119	Ge point contact	45	100	35	2.6	30	170	45
AAY30	Ge Au-bonded. High speed	30	400	-	0.88	150	8.0	30
AAY32	Ge Au-bonded. High speed	30	150	-	0.60*	30	11	30
AAY33	Ge Au-bonded. High speed	12	240	-	0.5*	30	15	12
AAZ15	Ge Au-bonded. High voltage	100	250	-	0.8	250	16	100
AAZ17	Ge Au-bonded. Gen.purpose	75	250	-	0.8	250	16	75
BAT81	Si Schottky (1.6pF)	40	30	-	0.33 0.41 1.0	0.1 1 15	0.2	30
BAT82		50						
BAT83		60						
OA47	Ge Au-bonded. Gen.purpose	30	150	-	0.54	30	10	30
OA90	Ge point contact "diode"?**	30	45	10	2.0	30	300	30
OA91	Ge point contact	115	150	50	2.1	30	75	100
OA95	Ge point contact	115	150	50	1.85	30	80	100

* max.
** a very non-linear resistor pretending to be a diode - best avoided!

6-8a. Thermionic diodes:
Given that semiconductor diodes are imperfect, some readers may wonder if thermionic diodes (valves) are capable of better performance. The answer however (avoiding expletives) is an an unequivocal 'no!'. The forward voltage of an indirectly heated valve diode in the low-current (space-charge limited) regime (according to Terman [20]) is given by the expression:

$$V_f = [\sqrt[3]{(I_f^2 / k)}] - V_c$$

where k is a constant determined by the geometry of the diode and Vc is called the contact potential, by analogy with the potential developed by a thermocouple. The diode contact potential arises because some of the electrons ejected from the cathode arrive at the anode even when there is no bias, and so the anode becomes negatively charged. This means that the diode must be reverse biased in order to prevent it from conducting, and the amount of bias required varies with the heater temperature and the age of the valve. In precision measurement terms, the contact potential is enormous. A sample of eight double-diode valves of type EB91 (6AL5) showed contact potentials ranging from 0.48 to 0.96V (with a mean of 0.73V) when measured using a voltmeter with 10M input resistance and a stabilised 6.30V DC heater supply (sensible measurement was impossible using a filament transformer connected to the domestic mains supply). A 100 A moving coil meter with an internal resistance of 980 was connected across a diode having an open-circuit contact potential of 0.74V, and registered a zero-bias current of 70 A, i.e., the diode gave a DC output of 4.8 W due to the thermal current. When the meter was padded to 10K , to simulate the diode loading in a realistic detector circuit, the zero-bias current was 22 A, i.e., 22% of full-scale deflection. We may safely conclude that thermionic diodes have no merit whatsoever in small-signal measuring applications, even if they do glow attractively in the dark.

(Ed note: thermionic diodes were the original diode detectors and can STILL make an interesting radio.)

increase in reverse current at a specified voltage. These diodes are called Zener diodes.

Diode Saturation Current is a very important SPICE parameter that, along with the diode Ideality Factor, n determines the actual diode current when it is forward biased by at particular DC Voltage. Id=Is*(e^(Vd/(0.026*n)-1) at room temperature. This expression ignores the effect of the parasitic series resistance of the diode because it has little effect on the operation of crystal radio sets at the low currents usually encountered. Here Id is the diode current, e is the base of the natural logarithms (2.7183...), ^ means raise the preceding symbol to the power of the expression that follows (Sometimes e^ is written 'exp'), * means multiply the preceding and following symbols, VD is the voltage across the diode and n equals the "Ideality factor" of the diode. At low signal levels, most detector diodes have an n of between 1.05 and 1.2). The lower the value of n, the higher will be the weak signal sensitivity. One can see that Is is a scaling factor for the actual curve generated by the factor (e^(VD/(0.026*n)-1).

Diode ideality factor (n): The value of n affects the low signal level sensitivity of a diode detector and its RF and audio resistance values. n can vary between 1.0 and 2.0. The higher the value of n, the worse the low signal level detector sensitivity. The low signal level RF and audio resistances of a diode detector vary directly with the value of n. Schottky diodes usually have a value of n between 1.03 and 1.10. Good germanium diodes have an n of about 1.07 to 1.14 when detecting weak signals. Silicon p-n junction diodes such as the 1N914 have values of n of about 1.8 at low currents and therefore have a lower potential sensitivity as diode detectors than Schottky and germanium point contact diodes. The value of n in Schottky diodes seems to be approximately constant over the full range of currents and voltages encountered in source, so we can say we have a 'no loss' situation. Now, assume that a transformer or other two-port device is inserted between the Vs,Rs source and Ro, and that an output voltage Vo is developed across Ro. The output power is (Vo^2)/Ro. The 'insertion power loss' can now be calculated. It is: 10*log (output power)/(maximum available input power) dB. After substituting terms, the equation becomes: Insertion power loss =10*log [(Vo/Vs)^2)*(4*Rs/Ro)] dB.

If the input voltage is referred to by its peak value (Vsp) as it is in a SPICE simulation, instead of by its RMS value, the equation changes. The RMS voltage of a sine wave is equal to the peak value of that wave divided by the "square root of 2". Since the power equation squares the voltage, the equation for the 'available input power' changes to Pa = (Vsp^2)/(8Rs).

4. Diode Saturation Current and Ideality Factor: Saturation current is abbreviated as Is in all of these articles. Assume that one connects a DC voltage source to a diode with the polarity of the voltage source such as to bias the diode in the back direction. Increase the voltage from zero. If the diode obeys the classic Shockley ideal equation exactly, the current will start increasing, but the increase will flatten out to a value called the saturation current as the voltage is further increased. That is, as the voltage is increased, the current will asymptotically approach the saturation current for that diode. A real world diode has several mechanisms that cause the current to actually keep increasing somewhat and not flatten out as the back direction voltage is further increased. Diode manufacturers characterize this as reverse breakdown and specify that the back current will be less than a specified value, say 10 uA at a specified voltage, say 30 V, called the reverse breakdown voltage. BTW there are other causes of excessive reverse current that are collectively referred to as reverse bias excess leakage current. Some diodes have a sharp, controlled

---

A Procedure for Measuring the Saturation Current and Ideality Factor of a Diode, along with Measurements on various diodes

http://www.bentongue.com/xtalset/16MeaDio/16MeaDio.html
by Ben H. Tongue

Quick Summary: A schematic and operational instructions are given for a device for use in measuring Saturation Current and Ideality Factor of a diode. Measurements of various detector diodes are included.

The Saturation Current and Ideality Coefficient of a diode can be determined by measuring an applied junction voltage along with the associated current flow at two different voltages. These two data pairs are then substituted into the Shockley diode equation to create two simultaneous equations in Is and n, and then solved for Is and n. Since the equations include exponential functions, they can not be solved by ordinary algebra. Numerical methods must be used.

The Shockley diode equation at 25 degrees C. is: Id = Is*(exp(Vd/(0.0256789*n))-1) Amps. Id = Diode Current (amps), Is=Saturation Current (amps), Vd = Diode Voltage, n = Ideality Coefficient. The series resistance Rs of the diode is ignored because the measurement currents are so low that the voltage drop across Rs is negligible. Measurements have shown that Is and n of point contact germanium diodes can vary with current, but are relatively constant, down to very low currents, when the current is under six times Is. Silicon p-n junction diodes exhibit values of Is and n that vary with current. The values for Is and n of Schottky diodes are quite constant over the range of currents used in ordinary crystal radio set reception.

A convenient set of measuring currents is about 6*Is and 3*Is. Substituting Id = 6*Is, then Id = 3*Is into the Shockley and solving for Vd yields: For Id = 6*Is, Vd = 0.05000*n volts. For Id = 3*Is, Vd = 0.03561*n volts. The value of n will probably be between 1.0 and 1.2 for the type of diodes used in crystal radio sets, so use 1.1 in determining the applied voltage to use. Suggested voltages to use are about 0.055 and 0.039 volts, although other values may be used.

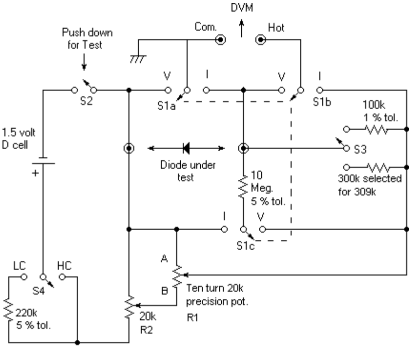

Fig. 1 - Schematic of Device for Measuring Diode Is and n.
Schematic

S1 is a triple pole double throw switch, S2 is a push button momentary-contact SPST switch. DVM is a digital voltmeter with 10 Meg input resistance having a 200 mV range setting. S3 is a range switch that enables greater precision when using a conventional 3 1/2 digit DVM. It is also used when

gain of, say 80 times, the output power is 2.5 X 80=200 mW. 2.5 mW expressed in dBm is +4 about dBm. A power gain of 80 times is about +19 dB. The output power is 4+19=+23 dBm.

3. Maximum Available Power: If one has a voltage source Vs with an inaccessible internal resistance Rs, the load resistance to which the most power (Pa) can be delivered is equal to Rs. Pa is called the 'maximum available power' from the source Vs, Rs. Any load resistance other than one equal to the source resistance, Rs, will absorb less power. This applies whether the voltage is DC or AC (RMS). The formula for power absorbed in a resistance is "voltage-squared divided by resistance". In the impedance matched condition, because of the 2 to 1 voltage division between the source resistance and load resistance, one-half of the internal voltage Vs will be lost across the internal source resistance. The other half will appear across the load resistance. The actual power available to the load will be, as indicated in the preceding relation: Pa = [(Vs/2)^2]/Rs = (Vs^2)/(4*Rs). Again, in the impedance matched condition, the total power delivered to the series combination of source and load resistance is divided up into two halves. One half is unavoidably lost in the internal source resistance. The other half is delivered as "useful output power" to the load resistance.

The 'maximum available power' approach is useful when measuring the insertion power-loss of two-port devices such as transformers, amplifiers and crystal radio sets, which may not exhibit an input or output impedance that is matched to the power source. The input impedance may be, in fact a combination of resistive and reactive components. If the Vs,Rs source is connected to a resistive load (Ro) of value equal to Rs ohms, it will receive and dissipate a power of Pa Watts. This is the maximum available power from the Vs, Rs frequencies. This is the main reason why diodes having large values for Rs and CJ perform poorly at high frequencies.

2. An explanation of the meaning and use of dB and dBm: In the acronym dBm, "d" means one-tenth. "B" refers to the Bel and is named after Alexander Graham Bell. The Bel is used to express the ratio of two powers, say (Output Power)/(Input Power). Let's call this power ratio "(pr)". Mathematically, a power ratio, expressed in Bels, is equal to the logarithm of the ratio of the two powers. B=log (pr). If the two powers are equal, the power ratio expressed in Bels is 0 B. This is because the log of one is zero. Another illustration: Assume that the power ratio is twenty. (Pr)=20. The logarithm of 20 is about 1.3. This power ratio in Bels is 1.3 B. One decibel is equal to 0.1 Bel. That is, 10 dB=1 B. If we express the two power ratios mentioned above (1 and 20) in dB, we get 0 dB and about 13 dB.

So far, we have seen that the decibel is used to express the ratio of two powers, it is not a measure of a power level itself. A convenient way to express an actual power level using dB is to use a standard implied reference power for one of the powers. dBW does this. It expresses the ratio of a power to the reference power (One Watt in this case). dBm uses a reference power of one milliwatt. A power level of, say 100 milliwatts, can be said to be a power level of +20 dBm (twenty dB above one milliwatt). Why? (100 milliwatts)/(1 milliwatt)=100. The logarithm of 100 is 2. 10 times 2 equals 20.

The convenient thing about using dB comes from a property of logarithms: The logarithm of the product of two numbers is equal to the sum of the logarithm of each number, taken separately. An illustration: If one has a power source of, say 2.5 mW and amplifies it through an amplifier having a power

measuring diodes having a high Is. R2 is used for coarse setting of the diode voltage. R1 is a ten turn precision 20k pot such as part # 594-53611203 from Mouser. It is used for fine setting of the diode voltage.

Procedure for Measuring Is and n:

1. Set S3 for 300k for diodes expected to have a low to medium Is. Set S3 to 100k if the diode is expected to have a high Is. S4 to HC and R1 to 1 about turn from point B.
2. Take Data Set #1: Set S1 to V. Push S2 and adjust R2 to obtain a reading of about 0.055 volts. Use R1 to set the voltage to the voltage desired (0.055 volts is suggested). Call this voltage V1. Set S2 to I, read the DVM and call that voltage V2.
3. Take Data set #2: Set S1 to V. Push S2 and adjust R2 to obtain a reading of about 0.039 volts. Use R1 to set the voltage to the voltage desired (0.039 is suggested). Call this voltage V3. Set S2 to I, read the DVM and call that voltage V4.
4. The diode voltage (Vd1) from Data Set #1 is V1. The diode current from Data Set #1 (Id1) is (V2/300,000)-(V1/10,000,000) or (V2/100,000)-(V1/10,000,000) Amps, depending on the setting of S3. The diode voltage (Vd2) from Data Set #2 is V3. The diode current (Id2) is (V4/300,000)-(V3/10,000,000) or (V4/100,000)-(V3/10,000,000) Amps, depending on the setting of S3.
5. The two data sets Vd1, Id1 and Vd2, Id2 must now be entered into two Shockley diode equations (shown above) in order to make two simultaneous equations in Is and n. Solving them will yield values for Is and n, measured at an average current of about 4.25 times Is.

A numerical equation solver can be used to solve the two simultaneous equations for Is and n. One is available in

MathCad. If you have MathCad 5.0 or higher, go to http://www.agilent.com/. Click your way through Communications, Communications Designer Solutions, RF and Microwave, Schottky Diodes, Library, MathCad worksheets and download the file: sch_char.mcd. Execute it in MathCad, then enter your Current and Voltage values: Id1, Vd1 and Id2, Vd2 as I2, V2, I1 and V1. Pull down 'Math' and click 'Calculate Worksheet'. The program calculates Is and n. Since most crystal set operation occurs at currents so low that there is negligible voltage drop across the diodes' parasitic series resistance, there is no need to enter any new numbers for I3, 4, 5 and V3, 4, 5 on the worksheet. The program sch_char.mcd does not work in versions of MathCad earlier than 6. If you have an earlier version of MathCad, and it has a non-linear equation solver, actual entry of the Data Set will have to take place without the convenience of the sch_char program. Those who do not have MathCad but do have Microsoft Windows Word can get an unformatted view of the default data and text provided in the MathCad program by clicking here.

There is currently available on the Web, a program from Polymath Software at: http://www.polymath-software.com/. This program has many capabilities, and among them is a nonlinear equation solving capability. A free demo copy of the latest program is available for download, but is limited to 20 uses. After that, for more usage, you have to buy it.

Some programmable pocket calculators include a nonlinear equation solver. One calculator that has one is the HP 32S Scientific Calculator. A program to solve for n and Is takes only 28 steps of program memory and is here.

Mike Tuggle posted on 'The Crystal Set Radio Club' the following simple procedure for determining Is and n by using a

PRACTICAL CONSIDERATIONS, helpful definitions of terms and useful explanations of some concepts

Ben H Tongue

http://www.klimaco.com/HAMRADIOPAGES/xtal_how_to.htm

1. An explanation as to why some diodes that work well in the Broadcast Band cause low sensitivity and selectivity when used at Short Waves: The parasitic (approximately fixed) series resistance Rs of a diode is in series with the parallel active elements. The nonlinear active elements are the junction resistance Rj, which is a function of current through the diode, and the junction capacitance Cj, which is a function of the voltage across it.

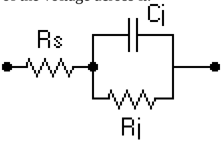

The nonlinear junction resistance effect is what we use to get detection. The nonlinear capacitance effect is used when the diode is designed to be a voltage variable capacitor (a varactor diode).

The parasitic series resistance of some 1N34 diodes can be pretty high, and in series with the junction capacitance, make that capacitance have a rather low Q at high frequencies. This capacitance is, in a crystal radio set, effectively in parallel with the RF tank. The tank usually has a small value tuning capacitor itself, so the overall tank circuit Q is reduced at high resistive loading occurs from the real but usually neglected current passed in the back direction during the RF reverse peak voltage excursions. This causes power loss and is a cause of loss of volume. For a given Is, good germanium diodes are usually superior to Shottkys in this respect.

Best regards,

Ben
eug4not

spreadsheet. 'In lieu of an equation solver package, the Schottky parameters can be solved for by simple trial-and-error. This is easily done with an ordinary spreadsheet, like Excel or Lotus. For the two measurement points, (Id1, Vd1) and (Id2, Vd2), set up the spreadsheet to calculate: Id2[exp(Vd1/0.0257n) - 1] and, Id1[exp(Vd2/0.0257n) - 1]. Then experimentally plug in different trial values of n, until the two expressions become equal. This gives the correct value of n. Now, plug this value of n into: Is = Id1 / [exp(Vd1/0.0257n) - 1] or, Is = Id2 / [exp(Vd2/0.0257n) - 1] to get the correct value of Is.' An Excel spreadsheet constructed as Mike suggested is here. An example from data taken on an Agilent HBAT-5400 is entered, for reference, on line 2. Line 3 may be used for calculations using data from other diodes. Column H automatically calculates a value for Is each time n is changed. All one has to do is enter the values as described above in columns A through E and hit enter.

Caution: If one uses a DVM to measure the forward voltage of a diode having a high saturation current, a problem may occur. If the internal resistance of the DC source supplying the current is too high, a version of the sampling voltage waveform used in the DVM may appear at its terminals and be rectified by the diode, thus causing a false reading. One can easily check for this condition by reducing the DC source voltage to zero, thus leaving only the internal resistance of the source in parallel with the diode, connected across the terminals of the DVM. If the DVM reads more than a tenth of a millivolt or so, the problem may be said to exist. It can usually be corrected by bypassing the diode with a ceramic capacitor of between 1 and 5 nF, preferably, an NPO type. I use a 0.047 uF NPO multi-layer ceramic cap from Mouser Electronics. Connect the capacitor across the diode with very short leads, or this fix may not work.

Tips

* If the Is of the diode under test is too high, 0.055 volts will not be attainable for V1 in step 1. The solution is to set switch S3 to 100k. The calculations for diode current then become: Id1=(V2/100,000)-(V1/10,000,000) Amps and (Id2=V4/100,000)-(V3/10,000,000) Amps.

* If the voltage readings seem to unstable, try placing the measuring setup on a ground plane and connect the common lead of the DVM to it. A sheet of household aluminum can be used for the ground plane. Use shielded cable from the lead from the DVM to the test setup.

* The voltage readings are very sensitive to diode temperature. You can see this easily by grasping the diode body with thumb and forefinger and noting the change in the voltage reading when measuring V1 or V3. Don't take data until the readings stabilize. Saturation current is a strong function of junction temperature. For germanium and the usual (n-doped) Schottky diodes, a temperature increase of 10fl Celsius results in a saturation current increase of about two times. A simple rule is: For each 1fl C. increase in temperature, Is increases by 7.2%. The figures are different for zero-bias-type Schottkys. Here, a 14 degree C. (25 degree F.) change in temperature will result in approximately a two times change in Is.

* Shield glass enclosed diodes from ambient light by placing a cardboard box over the unit. Many diodes have a photo-diode response and will give an output voltage when exposed to light even if no current is applied.

Note: A simplified method of determining the Saturation Current of a diode, if the Ideality Factor is estimated in advance is shown in Section #2 of Article #4.

resonance increases as the coupling between the detector and ATU is reduced - and therefore a diode with a larger rc value is a better match. If this is the case, then for really good radios with very high Q detector tank circuits, two different diodes may give the loudest output - one when coupling is very tight, and a different one when coupling is very loose. (Any comments??)

For my radio used here, the 1N141 diode (thank you Bruce Kizerian!) seems to give about the loudest output for both tight and loose coupling at the low end of the band. But the FO-215's output is very close and has a bit better selectivity. The schottky diodes (1SS98 & 5082-2835) are seen to be useful even for my modest radio if there is plenty of signal that can be traded-off for greatly improved selectivity.

Well, that's my attempt at measuring my crystal radio's audio output. Hope it's of interest. If my measurements or methods are way off base, please let me know.

It sure is a blessing to be retired and to have the luxury of time to play.

73's, Dan

May 23, 2007
Hi Dan,

You have developed some really insightful information.

There is one thing, I believe, that is being overlooked re the low Is (high Rx) Schottky diodes when one is going for maximum volume. They have rather low reverse breakdown voltages so, when a large RF is being detected, a lot of extra

2.) Loose coupling between the detector and ATU (12" separation).

The last plot again shows how the ac "audio" voltage Va and the dc benny voltage Vb vary as different diodes are used in the radio.

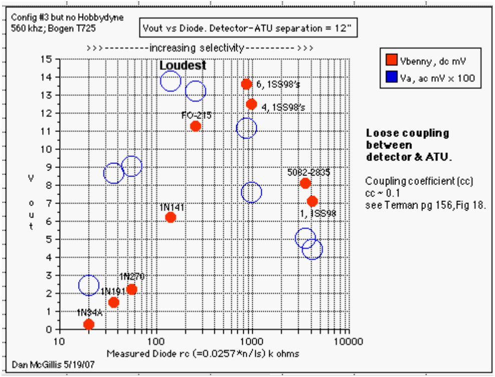

The benny voltage (RED DOTS) seems to peak in the vicinity of the benny resistance as before. However, the peak ac "audio" voltage Va range (BLUE CIRCLES), may be shifting toward higher diode rc values than for the tight coupling case.

My assumption is that the detector tanks' parallel resistance at

---

Summary of measurements on some diodes:

The following charts show typical values for Is and n for diodes that might be used in crystal radio sets. One can see, for any particular diode, that Is and n do not vary by much over a moderate current range. Therefore, they may be considered to be dynamically constant when receiving a signal. Each value of n and Is is calculated from two voltage/current pairs as described above. The diode current (Id) given for each of the n, Is pairs is the geometric mean of the two currents used in the measurement. A Fluke model '89 IV' 4-1/2 digit DVM was used to enable measurements down to as low as 15 nA on some diodes. Noise problems cause some measurement error at low currents. That is the reason for the fluctuations in some of the readings. Values of n very close to 1.0 or below are obvious measurement errors. Those low values for n should have come out somewhat higher and the associated values of Is, also higher.

Note that the germanium diodes show an unexpected tendency to increased values for Is and n at the higher currents. The 1N4148 silicon p-n junction shows the expected increase of Is and n at lower currents. The Schottky diodes seem to have pretty constant values of Is and n across the current ranges measured. Experiments described in Article #27 indicate that the measured values of Is and n for silicon Schottky diodes tested here, when used as detectors, remain at the measured values at rectified currents so low that a voice signal is barely readable. This is not necessarily true for all germanium diodes.

Table 1 - Measured n and Is values for various diodes, over a range of currents (Id), in nA.

1N4148 silicon p-n junction diode			Base-emitter junction of 1N404A Ge transistor			Blue Radio Shack 1N34A Ge diode having no nomenclature			Agilent HBAT-5400 Schottky, high Is version			Infineon BAT62-03W Schottky diode		
Id	**n**	**Is**	**Id**	**n**	**Is**	**Id**	**n**	**Is**	**Id**	**n**	**Is**	**Id**	**n**	**Is**
710k	1.73	1.23				710k	1.71	3500						
			570k	1.04	1800									
350k	1.75	1.45				350k	1.69	3200						
177k	1.87	2.19	179k	1.03	1670	177k	1.61	2550						
88k	1.82	2.26				88k	1.51	1980						
44k	1.80	1.98	56k	1.01	1540	44k	1.39	1470						
22k	1.88	3.00	18.9k	1.01	1580	22k	1.28	1100						
11k	1.89	3.10				11k	1.22	950						
5500	1.93	3.80	5700	1.04	1660	5500	1.14	800	8100	1.15	265			
2760	1.94	3.90	1790	0.98	1730	2760	1.10	750				2600*	1.06	248
1380	2.02	4.90				1380	1.05	680						
690	1.98	4.40	620	0.99	1740	690	1.20	830	990	1.15	248	970	1.04	240
343	2.06	5.30				343	1.01	670	360	1.15	265	341	1.04	236
170	2.18	6.70				170	1.08	720	160	1.15	255	133	1.04	236
									76	1.15	254	87	1.04	236
												59	1.01	228
									40	1.15	261	39	1.06	233

* This Infineon diode has an unusually high series resistance of 130 ohms. The voltage drop across this resistance is low enough in all the measurements to be ignored, except for the highest current one. There, a correction for the voltage drop was made.

Table 2 - Measured n and Is values for various diodes, over a range of currents (Id), in nA.

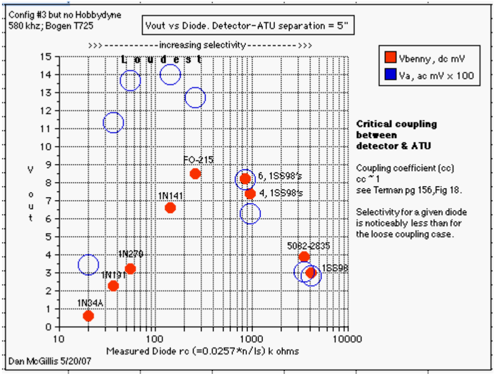

If Va is in fact related to audio power, then the plot indicates that for this radio, the audio power delivered peaks for diodes with critical resistances in the range of say 50k to 200k - a pretty broad range. The dc benny voltage Vb (RED DOTS) however, seems to peak in the vicinity of about 500k - the value of the benny resistance. As others have noted, the selectivity - as judged by listening - continuously improves as diodes with larger and larger rc values are used. By choosing different diodes, you can tradeoff loudness for selectivity. The plot shows the magic of impedance matching. (Thank you Ben Tongue for your articles!) All this is old hat to some folks but for a novice like me, it's neat to see.

The modulation on my talk show test signal seems fairly constant - as long as measurements are avoided during commercials or the occasional music. A spotting radio helps in avoiding high modulation surprises. The ac signal does still bounce around a bit. Taking 20 random voltage readings per data point and averaging helps smooth the results.

The radio used as a test bed is double tuned with 165/46 litz coils close-wound directly on Amidon 0.5”x4” ferrite 61 rods. The coupling between the detector’s tank circuit and the ATU tank is varied by changing the end-to-end separation between the ferrite rods.

1.) Tight (approximately critical) coupling between the detector and ATU.

The first plot shows how the “audio” ac voltage Va (BLUE CIRCLES) varies as the detector diode is changed - with a small 5” separation between the detector and ATU rods.

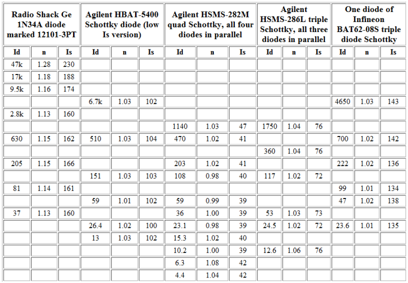

Radio Shack Ge 1N34A diode marked 12101-3PT			Agilent HBAT-5400 Schottky diode (low Is version)			Agilent HSMS-282M quad Schottky, all four diodes in parallel			Agilent HSMS-286L triple Schottky, all three diodes in parallel			One diode of Infineon BAT62-08S triple diode Schottky		
Id	**n**	**Is**	**Id**	**n**	**Is**	**Id**	**n**	**Is**	**Id**	**n**	**Is**	**Id**	**n**	**Is**
47k	1.28	230												
17k	1.18	188												
9.5k	1.16	174												
			6.7k	1.03	102							4650	1.03	143
2.8k	1.13	160												
						1140	1.03	47	1750	1.04	76			
630	1.15	162	510	1.03	104	470	1.02	41				700	1.02	142
									360	1.04	76			
205	1.15	166				203	1.02	41				222	1.02	136
			151	1.03	103	108	0.98	40	117	1.02	72			
81	1.14	161										99	1.01	134
			59	1.01	102	59	0.99	39				47	1.02	138
37	1.13	160				36	1.00	39	53	1.03	73			
			26.4	1.02	100	23.1	0.98	39	24.5	1.02	72	23.6	1.01	135
			13	1.03	102	15.3	1.02	40						
						10.2	1.00	39	12.6	1.06	76			
						6.3	1.08	42						
						4.4	1.04	42						

A rare germanium diode that seems to be ideal for many crystal radio set designs is the FO 215, branded ITT. A search of the Internet has not turned up a manufacturer's datasheet. ITT is not in the germanium diode business anymore, but from the Internet search it appears that the original company was a German company named ITT Intermetall. Some of their semiconductor business became ITT Semiconductors. This was later sold, around 1997 to General Semiconductor Industries. That business was later sold to Vishay. One source indicated that General Instruments was also one of the intermediate owners. Averages of measurements on three samples of the FO 215 are: Is=109 nA and n=1.02. These measurements were made at an average current of about 250 nA. Interesting note: The average Is of the FO 215 diodes is about equal to the geometric mean of that of the Agilent 5082-2835 and a typical 1N34A. I obtained my FO 215 diodes from Mike Peebles at: http://www.peeblesoriginals.com/ .

Article #27 shows detector measurements of how diodes having different values of Is and n perform as weak signal detectors when impedance matched at both input and out put.

#16 Published: 03/28/01; Revised: 02/10/2004

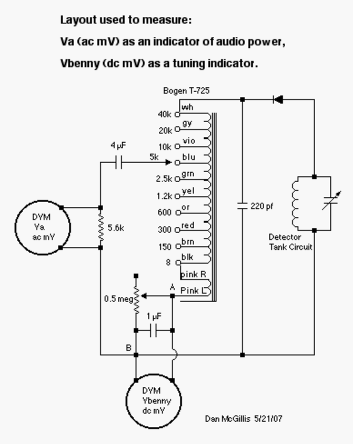
Layout used to measure:
Va (ac mV) as an indicator of audio power,
Vbenny (dc mV) as a tuning indicator.
Bogen T-725
40k
wh
20k
gy
10k
vio
4 µF
5k
blu
2.5k
grn
1.2k
yel
600
or
300
red
150
brn
8
blk
pink R
Pink L
220 pf
DVM
Va
ac mV
5.6k
Detector
Tank Circuit
A
0.5 meg
1 µF
B
DVM
Vbenny
dc mV
Dan McGillis 5/21/07

The ac voltage (Va) across the 5.6k resistor load is assumed to be proportional to audio power (=V^2/R). The DVM used is supposed to have a 10 meg input on the “ac mV” scale. It’s specs say it has a 40 - 400 hz ac response, and an rms output calibrated to a sine wave. Since the audio voltage across the 5.6k is not a sine wave, and has frequency components outside the 40 - 400 hz response range, I can only hope that the DVM’s output will be a relative measure of audio power. The DVM is all I’ve got, so it will have to do.

Audio and Benny Voltage Measurement:

May 21, 2007

A lot of the fun playing with crystal radio for me involves "measuring stuff". The "benny voltage" - the dc voltage across the benny resistor in the matching transformer circuit - is a favorite parameter of mine. It's easy to measure with a DVM and it seems to vary with everything. And it's a very good tuning aid.

However, Dave Schmarder has gently and patiently reminded me on many occasions that it's the audio power delivered to the headphones that's important -- that I should try to measure relative audio power as well as benny voltage when evaluating radio changes. The following is my first attempt. Dave thought it might be of interest to the group.

With a DVM as the only measurement tool, and weak daytime talk stations (non sports) as signal sources, my measurements are low-tech. The circuit used to measure the ac voltage (Va) - that I hope is related to relative audio power - is shown in the attached layout.

Diode Modeling Update

Kevin Smith

http://www.lessmiths.com/~kjsmith/crystal/dmodel.shtml

The following explains my first attempt to compare a measured diode characteristics with calculated plots based on the Shockley Equation.

I have been taking serious look into diode modeling and measurement recently in an attempt to better understand these solid-state successors to the humble crystal and cat's whisker detector. There is much good information on diodes to be found in many excellent web sites of course, but there is nothing quite like actually making the measurements and working with them to get a good feel. This page reports some of my protocols and test setups which I have found to be useful.

Most probably the single best web article for the measurement of diode Is and n parameters is given by Ben Tongue in his Article 16. In this article he describes an interesting circuit for the measurement as well as protocols for making them. For my purposes I did not wish to make this a construction project and felt I might get along with good quality meters and my already-built "Diode Test Jig". Essentially Ben Tongue's method consists of making two precision measurements of Voltage and Current through a diode at small signal levels, essentially about 3 and 6 times Is (sufficiently low that the voltage drop across the series resistance Ro can be ignored). The measurements are then substituted into the Shockley equation Is*{exp[(qe/(n*k*T))*(V-I*Rs)]-1} and solved simultaneously for the two data pairs. Mike Tuggle has provided a nice excel spreadsheet to do the math. This spreadsheet, Cal_n_Is.xls forms the basis of my technique and

I am thus indebted to both Ben Tongue and Mike Tuggle for making this project feasible.

In addition to simply calculating the main parameters Is and n, I also wished to measure enough points to plot a characteristic curve, and to compare the measured curve with one calculated directly from the Shockley equation. My first spreadsheet then combined Tongue/Tuggle calculation protocol and a graphical view of the match between measured and theoretical. I did make a few methodology modifications which I thought/hoped would allow added accuracy:

1) Using the full Shockly equation without simplifying for an assumed 25dC room temperature. While the 25dC assumption is generally good and probably within the measurement error (or perhaps not), including the actual measured temperature eliminates doubts of inaccuracies due to this parameter. With the power of PC's and spreadsheets there is no reason simplify the equation.

2) Used Vd = 0.04 and 0.05 V (adequately close to Ben Tongue's recommended 0.039 and 0.055 V) whenever possible. Note for some diodes with significantly different forward voltage drops, Si diodes in partictular I had to use higher values for Vd.

3) Reporting: I decided that, as there is no unique solution to n and Is, each is dependant on the valuse Vd and Id, I feel that reporting both Vd and Id is necessary for repeatability. Naturally I also include the ambient temperature in the report as well, and

4) I included a calculation for Ro (or Rg if you are using Tongue's reference) as that is the goal of the exercise.
A screen shot shown below:

Effect of Diode Rd on Audio Power

http://theradioboard.com/rb/viewtopic.php?t=330
By Dan McGillis

Introduction:
Sep 13, 2011

Kevin - thanks for posting your work. Really good stuff.
Mike Tuggle & Mike Peebles gave excellent advice on choosing a diode.

Here's a plot of how the choice of diodes - and their associated Rd's - affected the relative loudness of one of my crystal sets.
http://theradioboard.com/rb/viewtopic.php?t=330

As you can see, there is a "best" choice of Rd for this particular set if maximum loudness is the goal. Using diodes with Rd's larger than that needed for maximum loudness gave the crystal set more selectivity (and less loudness).

I should note that you get this kind of curve of loudness vs diode Rd when detecting WEAK signals. For very strong signals, all my diodes gave about the same measured loudness. ie. The choice of diode (& it's Rd) doesn't matter too much for very strong signals.

As a side note, my measured Rd's for a given diode are a bit higher than yours and Mike's because I tend to measure at higher Vd's for convenience. And my measurements are no where near as accurate as yours.

Sure is fun stuff isn't it. 😃

73, Dan

then: $d2I / dV2 = d/dV [U-1] = -1 * U-2 * dU / dV$, where

$dU / dV = (dU / dV) + (dU / dI) * (dI / dV)$, and where

$(dU / dV) = (-1 / Is) * \exp [-q (V - I Rs) / nkT]$

$(dU / dI) = (Rs / Is) * \exp [-q (V - I Rs) / nkT]$

Sze's Gamma:

$\gamma = (d2I / dV2) / (dI / dV)$

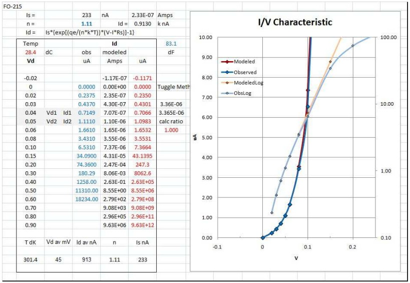

FO-215					
Is =		233	nA	2.33E-07	Amps
n =		1.11	Id =	0.9130	k nA
Id =	Is*{exp[(qe/(n*k*T))*(V-I*Rs)]-1}				
Temp			Id		83.1
28.4	dC	obs	modeled		dF
Vd		uA	Amps	uA	
-0.02			-1.17E-07	-0.1171	
0		0.0000	0.00E+00	0.0000	Tuggle Meth
0.02		0.2375	2.35E-07	0.2350	
0.03		0.4370	4.30E-07	0.4301	3.36E-06
0.04	Vd1 Id1	0.7149	7.07E-07	0.7066	3.365E-06
0.05	Vd2 Id2	1.1110	1.10E-06	1.0983	calc ratio
0.06		1.6610	1.65E-06	1.6532	1.000
0.08		3.4310	3.55E-06	3.5531	
0.10		6.5310	7.37E-06	7.3664	
0.15		34.0900	4.31E-05	43.1395	
0.20		74.3600	2.47E-04	247.3	
0.30		180.29	8.06E-03	8062.6	
0.40		1258.00	2.63E-01	2.63E+05	
0.50		11310.00	8.55E+00	8.55E+06	
0.60		18234.00	2.79E+02	2.79E+08	
0.70			9.08E+03	9.08E+09	
0.80			2.96E+05	2.96E+11	
0.90			9.63E+06	9.63E+12	
T dK	Vd av mV	Id av nA	n	Is nA	
301.4	45	913	1.11	233	

To use the sheet one need only adjust the voltage Vd to that shown in the first column and record the current Id. If only Vd1/Vd2 and Id1/Id2 are measured then the next step is to adjust the value for "n" until the match between the two equations is exact. I even provided a simple ratio calculation to easily test for a match.

While the math is good and results excellent, I quickly found a couple deficiencies in my method,

1) measurements require high precision and its nearly impossible to land EXACTLY at the voltage required. I needed only to get quite close, let the meters stabilize for 2 - 5 minutes, and then record BOTH Vd and Id.

2) the work is rather tedious and, for a large number of diodes it pays dividends to measure only the needed Vd and Id and let the plot aside. My second spreadsheet thus dispenses with the plot. It is with my second spreadsheet that all my data tables and results are posted.

3) better comparisons between diodes can be made if one targets specific Currents rather than Voltages, discussion on this below.

A view of my current data reporting spreadsheet showing ALL input parameters as well as determined values of n, Is, and Rd.
Data input fields in blue: (target Id1 close to 1.0 uA and Id2 close to 0.5 uA).
Calculated fields in red and black.
Adjustment field "n" in green: (adjust value of n until the ratio of the two calculations = exactly 1).

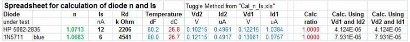

Spreadsheet for calculation of diode n and Is Tuggle Method from "Cal_n_Is.xls"

Diode		n	Is	Rd	Temperature		Vd2	Id2	Vd1	Id1	Calc	Calc. Using	Calc. Using
under test			nA	k Ohm	dF	dC	V	uA	V	uA	ratio	Vd1 and Id2	Vd2 and Id1
HP 5082-2835		1.0713	12	2206	80.2	26.8	0.10215	0.4961	0.12215	1.0384	1.0000	4.124E-05	4.124E-05
1N5711	blue	1.0683	6	4541	80.0	26.7	0.12115	0.4917	0.13981	0.9757	1.0000	7.931E-05	7.931E-05

(Note: the engineers in the crowd will have noticed in the above report than I am carrying a degree of precision not justified by the level of accuracy in my measurements. The final results of n, Is, and Rd should be rounded off to not more than three significant figures.)
A copy of my spreadsheet can be downloaded here..

While measuring a good number of diodes, both Germanium, Schottky, Silicon, and even a few LED's, I noted that Tongue's recommended measurements at Vd = 0.039 and 0.055 V resulted in current reports spread out over more than two orders of magnitude. As the determination of Is and n is not unique but a function of Id, I felt uneasy by this method. While most Germanium diodes I have measured have a fairly narrow range of Forward Voltage drops, (Vf), that of Schottky's can range up to a tenth of a volt. This is the factor responsible for the huge range of measured Id values, see the following plot of I/V characteristics for some selected diodes:

Schottky Law with I*Rs Term

Current, I:

$$I = Is \{ \exp [q (V - I *Rs) / nkT] - 1 \} \quad \text{(implicit expression)}$$

Dynamic Conductance, dI / dV:

rearranging, let: $\exp [q (V - I\,Rs) / nkT] - ((I / Is) + 1) = 0 = F(V, I)$

then: $dI / dV = (- dF / dV) / (dF / dI)$

$(- dF / dV) = \exp [q (V - I\,Rs) / nkT] * (q / nkT)$, and

$(dF / dI) = \exp [q (V - I\,Rs) / nkT] * (-q\,Rs / nkT) - 1 / Is$

then,

$$dI / dV = 1 / \{ (nkT / q\,Is) * (\exp [-q (V - I\,Rs) / nkT] + Rs \}$$

Dynamic Resistance, 1 / (dI / dV):

$$Rd = dV / dI = (nkT / q\,Is) * (\exp [-q (V - I\,Rs) / nkT] + Rs$$

Curvature, d2I / dV2 :

$$d2I / dV2 = d/dV [1 / \{ (nkT / q\,Is) * (\exp [-q (V - I\,Rs) / nkT] + Rs \}]$$

Let: $(nkT / q\,Is) * (\exp [-q (V - I\,Rs) / nkT] + Rs = U$

3. With increasing V, the slope (or conductance) itself increases initially, and then levels off at a constant value of 1 / Rs. (red curve)

4. Curvature, the rate-of-change in slope with increasing V, increases initially to a maximum level. After that it decreases and goes to zero as the straight-line portion of the curve is reached. (blue curve)

5. Sze's sensitivity figure, curvature over slope, starts out constant initially but decreases to zero as curvature goes to zero. (red curve in figure below -- compared to the black curve case with no Rs term)

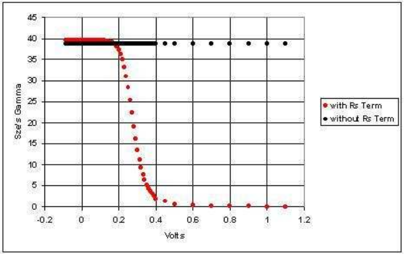

6. Note that Sze's gamma takes a dive near the same voltage where curvature peaks. (??)

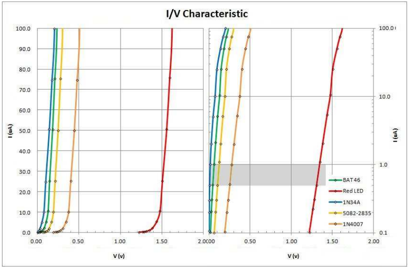

In order to get what I feel is a better comparison between diodes, I have decided to target not some pre-determined voltage, but rather target currents of Id = 0.5 and 1.0 uA. On the above plot on the right (with Log Id vs Vd scale) it will be clear that I am aiming at the same part of the characteristic curve regardless of Vf. Hopefully this will allow good comparison between diodes with rather different Vf, even as far as including Silicon and LED's in my mix. On va voir....

Shockley equation Id = Is*{exp[(qe/(n*k*T))*(V-I*Rs)]-1}
where:
n = ideality factor
Is = Saturation current in Amps
Id = Diode Current in Amps
Vd = Diode Voltage
k = boltzmann = 1.38E-23 J/K
T = temp K = 300
K = dC + 273.15 Kelvin
qe = electron charge = 1.609E-19 cmb

Often simplified to:
Id = Is*(exp(Vd/(0.0256789*n))-1)   [at room temperature]

Diode Resistance Rd = VT * n / Is
Where:
T = 300K
VT = k*T/qe = 0.0256789
SO: Rd = k * T / qe * n / Is

I cannot express enough my indebtedness and thanks to Ben Tongue and Ken Khun for their great work and web documentation in matters Diode!

A photo of the current "bench" setup. Keithey 195A pico-ampmeter + Keithley 192 DMM bench meter.

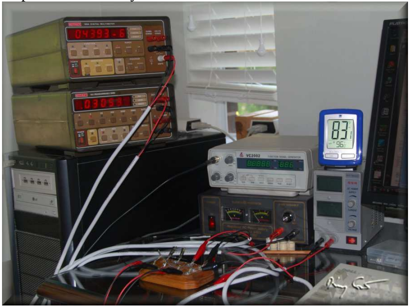

Kevin 09/2011

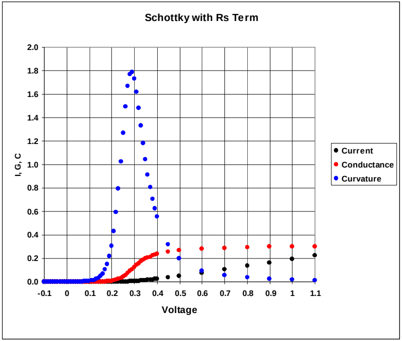

Diode: Is = 1e-7 A.; eta (n) = 1; Ro = 258 k-ohms; Rs = 3.2 ohms; T = 300 K.

1. At low voltages, current rises exponentially at first but then increases linearly as the I*Rs voltage term dominates. (black curve)

2. Initially resistance (1/slope) decreases exponentially from a high value. Then it becomes constant as barrier resistance R becomes vanishingly small and parasitic resistance Rs becomes the dominant resistance. (resistance curve not shown)

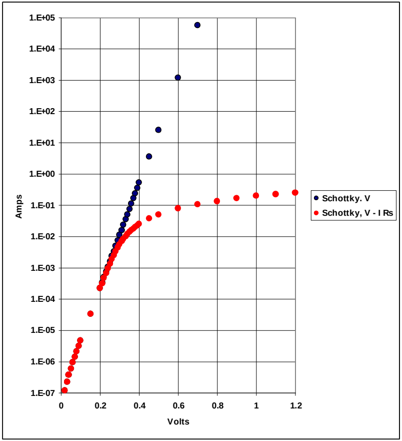

Schottky Behavior with Rs Term

Diode voltage-current behavior is modeled with the Schottky expression that includes the I*Rs term. Below are graphs of some of the resulting quantities.

Spreadsheet for calculation of diode n and Is (@1.0uA)

Diode		n	Is nA	Rd k Ohm
Germanium Diodes				
FO-215	ITT	1.11	175	163
FO-215	ITT	1.10	196	144
1N277	black	1.50	2293	17
1N277	black	1.60	2296	18
1N270	bonafide	1.28	915	36
1N270	blue	1.26	857	38
1N270	blue	1.66	2310	19
D18	russia	1.20	194	160
D18	russia	1.27	193	170
GAZ 51	Tesla	1.14	140	211
GAZ 51	Tesla	1.52	617	64
OA 5	Tesla	1.85	3384	14
OA 5	Tesla	1.48	1996	19
D9E	russia	1.42	2161	17
D9E	russia	1.54	2414	16
1N34A	bonafide	1.82	2153	22
1N34A	bonafide	1.25	1392	23
1N34A	green	1.57	1389	29
1N34A	green	1.30	1185	28
1N34A	green	1.32	1554	22
1N34A 37	orange	1.10	1458	19
1N34A 37	orange	1.36	1370	26
1N34A	red	1.34	833	42
1N34A	red	1.32	993	34
D310	russia	1.01	1215	21

Diode		n	Is	Rd
			nA	k Ohm
GD 402A	russia	1.55	970	41
GD 402A	russia	1.71	1360	33
UK A	russia	1.56	1565	26
UK A	russia	2.00	5766	9
UK B red	russia	1.64	2049	21
UK B red	russia	1.85	3777	13
UK C ora	russia	1.92	3462	14
UK C ora	russia	1.95	3354	15
UK D blk	russia	2.00	4901	10
UK E blk	russia	2.00	5862	9
UK F blk	russia	1.82	3523	13
UK G blue	russia	1.86	3123	15
Average		1.39	1334	54
Schottky Diodes				
HP 5082-2835		1.07	12	2206
1SS98		1.08	12	2302
1SS98		1.06	13	2188
1N5711	blue	1.07	6	4541
1N5711	blue	1.07	6	4816
1N60		1.09	178	158
1N60		1.10	161	176
BAT 46		1.12	141	204
BAT 46		1.18	172	177
1N34A ?	schottky	1.11	317	90
1N34A ?	schottky	1.24	482	66
1N5819		1.19	750	41

Effect of Parasitic Resistance Rs on Diode Current

The common Schottky expression predicts extraordinarily high currents for applied voltages above several hundred millivolts (black curve). For example, at the "knee" voltage, 0.7 volts, predicted current is on the order of 50,000 amps! Of course, this doesn't happen. When one includes the effect of series parasitic resistance Rs, replacing V by (V - I Rs), current is tamed to a more reasonable 100 milliamps in this example.

When I*Rs term is included, I appears on both sides of the Schottky expression and cannot be rearranged into a single I term. It must be solved for by trial-and-error. Excel's SOLVER function can be used, but still, each data point must be solved separately.

Diode: Is = 1e-7 A.; eta (n) = 1; Ro = 258 k-ohms; Rs = 3.2 ohms; T = 300 K.

DC RESISTANCE (Shockley-Law):

$$RDC \;=\; V / I \;=\; (nkT / q{*}I) * \ln [(I / Is) + 1] + Rs$$
$$RDC \;=\; R \;+\; Rs$$

Dynamic resistance Rd is the impedance (resistive) offered by the diode to an alternating (RF) signal when the diode is biased by a DC current I. Rd is the inverse of the slope of the DC I versus V curve at voltage point V (or current I). The dynamic resistance at zero volts DC applied is the characteristic resistance. It is the resistance to be considered when matching the diode to, say, the tank circuit in a crystal set. See illustration below.

DYNAMIC RESISTANCE:

$$Rd \;=\; dV/dI = d/dI\, [(nkT /q) * \ln [(I / Is) \;+ 1]+ I * Rs]$$
$$Rd \;=\; (nkT / q{*}I) * \ln [(I / Is) + 1]$$

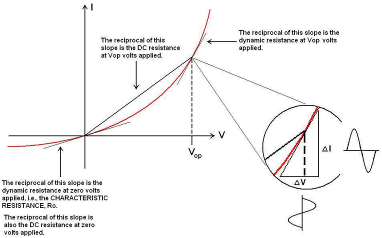
I
The reciprocal of this slope is the DC resistance at Vop volts applied.
The reciprocal of this slope is the dynamic resistance at Vop volts applied.
V
V_{op}
ΔI
ΔV
The reciprocal of this slope is the dynamic resistance at zero volts applied, i.e., the CHARACTERISTIC RESISTANCE, Ro.
The reciprocal of this slope is also the DC resistance at zero volts applied.

1N5819		1.14	807	36
Diode		n	Is nA	Rd k Ohm
1SS16	NEC	6.24	14288	11
1SS16		2.27	4101	14
Average		1.13	276	1137
Silicon Diodes				
1N914		1.95	5	9293
1N914		2.02	7	7618
1N4148		2.00	6	8934
1N4148		2.00	6	9323
1N4007		1.51	1	64311
1N4007		1.51	1	77954
6A10		1.57	2	19275
6A10		1.54	2	20379
1N4736A	Zener	1.15	0	1.91E+09
1N4736A	Zener	1.14	0	2.50E+09
KB 130	russian	1.14	0	5.83E+08
KB 130	russian	1.16	0	4.36E+08
UK H	russian	1.94	8	6263
UK H	russian	1.77	4	12548
UK I	russian	1.78	0	260253
KD 401A	russian	1.51	0	232930
KD 401A	russian	1.38	0	1539619
D 220	russian	1.26	0	322405
D 220	russian	1.24	0	537707

D 223A	russian	1.20	0	10262661
Diode		n	Is nA	Rd k Ohm
D 223A	russian	1.20	0	8565213
UK J	russian	1.46	4	9849
UK J	russian	1.40	2	20013
Average		1.52	2	2.37E+08

Light Emitting Diodes

Red	2.15	0	1.70E+12
Red	2.07	0	5.52E+12
Amber	1.56	0	2.09E+17
Amber	1.62	0	6.46E+16
Yellow	1.59	0	8.03E+17
Yellow	1.38	0	2.87E+20
Green	2.05	0	4.23E+13
Green	2.07	0	4.34E+13
Water Green	3.27	0	4.85E+13
Water Green	1.93	0	5.79E+19
White-Gn	1.44	0	1.05E+20
White-Gn	3.04	0	6.88E+09
Average	2.01	0	3.76E+19

Diode Model

Notes provided to the editor by Mike Tuggle

The diode is modeled as follows:

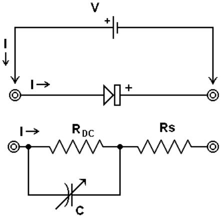

If a DC voltage V is placed across the diode, current I flows through the diode.

RDC is the non-linear, Shockley-law DC resistance that varies with V and I. RDC is determined by diode-specific parameters Is and eta (n) in the exponential expression.

Rs is the fixed parasitic (bulk) resistance that stays constant with V.

C is the junction capacitance that can be ignored when considering just the direct current (DC) behavior of the diode.

CURRENT:

$$I = Is \{\exp [(V - I*Rs) / (nkT)] - 1 \}$$

VOLTAGE:

$$V = (nkT / q) * \ln [(I / Is) + 1] + I * Rs$$

Diode Test

Kevin Smith

http://www.lessmiths.com/~kjsmith/crystal/dtest.shtml

So, what is the best diode out there to use? Do crystals always show poor sensitivity compared to germanium? What about Fleming's marvelous device, the vacuum diode? I start by using a test jig, the design shamelessly modified from Dan Petersen in the Nov 2009 Crystal Set Society Newsletter, (if you aren't a member, go sign up to be.. NOW):

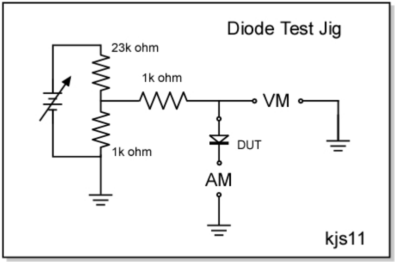

This is a very simple circuit allowing power input, attaching a diode for testing, and sampling both the current and voltage across the diode under test, (DUT). I made a simple modification by adding a voltage divider, about 20:1 at the power input. The power supply can be connected either at the front across the divider, or after the divider in front of the 100k current limiting resistor for higher voltage readings. I get the input voltage from a modest 18V, 2A regulated DC power supply. This unit has both volt and amp controls with digital readouts, nice to use and doesn't take up much room on my desk. Here is a case where I recommend you look at getting a new unit. I looked long and hard on ebay, but found few bargains and few units with the same features. Still, the supply regulator was never designed to adjust the output to precise small values making it somewhat difficult/tedious to adjust. Adding a small rheostat to the test jig may aid for making fine adjustments on the voltage. Another piece of equipment you will need for this test is an amp meter. At first I used a nice small digital meter I found new on ebay for $15 shipping included. It reads amps to a lowest range of 200uA, more than enough precision for the task. For that price get two. Later as I got to making more precision measurements for the determination of diode n and Is I acquired a Keithley 195A bench volt/amp meter with very high input impedance. The DVM is, of course, my trusty Keithley workhorse.

I have received occasional criticism for the simple test jig with the suggestion that proper measurements cannot be taken with the jig as-is. I note that, while the test setup is simple, the instrumentation is high quality and performs the job well. I have had the opportunity to collaborate with Mike Tuggle, comparing our respective measurements. Even my early characteristic curves taken with the small amp meter compare well and provide a good view of the diode and determination of the diode Rs series resistance. The detail n and Is measurements presented here essentially match, though at somewhat less precision, those data from Mike. The reader should have some assurance that the data and measurements found on this page are vetted and reliable.

Easy protocol as follows:

1) Set up the test with the diode and meters clipped to the jig.

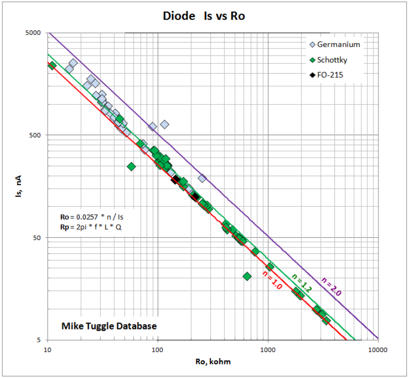

I absolutely agree, your data set on diodes is preeminent for its detail and variety of diodes covered. I'll attach a summary index of all the diodes I've measured (ed note: summary chart above). Each diode has a full multipoint analysis and I-V plot, if you're interested in any of those. Quality varies, as the methodology has improved over the years.

Best regards, -Mike-

Hi, Kevin.

Coincidentally, I have been thinking about the measurement of diode I-V's at higher voltage levels. For an 0.4 volt drop across the 1N34A that we've been considering, the Shockley law current through the diode will be some 312 mA.

To measure that with a 50 k-ohm current sense resistor, we would need to supply 15,600.4 volts -- 0.4 volts across the diode; 15,600 volts across the current sense resistor. Obviously not practical, for my setup anyway.

A much smaller current sense resistor is needed. I'd say about 1 ohm. Then there would be 0.4 volts across the diode and 0.312 volts across the current sense resistor -- or 0.712 volts total. That could be done. Whatever current sense resistor is used, its value needs to be known accurately.

Under these conditions, the DC resistance of the 1N34A is 0.4 volts/0.312 amps = 1.28 ohms. The bulk (parasitic) resistance of the diode would need to be considered. I would suggest multiple measurements regressed by Excel's SOLVER, using the Shockley expression with bulk resistance parameter. Some notes are attached. (p 153).

2) Adjust the power supply until the reading on the DVM meter is where you plan to measure. Adjust the voltage (DVM readout) in steps from 0 to 1.0 Volt in 0.1 V increments and record the current (Amps) in a spreadsheet.
3) Repeat at each tenth volt recording while the spreadsheet makes a sweet graph.

Test Results
First, I offer a quick look at my diode modeling setup. I am currently upping the ante on resolution in order to determine diode parameters Is and n, hoping for success soon! Setup and protocol..

Over a period of a few weeks I ran I-V characterization tests on a good number and variety of diodes, crystals and my two tubes to see how very thing stacks up. The following presents the resulting curves, diode photos, and some puzzling questions/conclusions I churned up in the process. I start off with the realization that when one orders diodes, it is by no means certain what one will get. This seems especially true for the supposedly ubiquitous 1N34A. Unless you see the part number actually on the diode, you probably need to test it to know what it really is. So, lets take a look!

Germanium Diodes

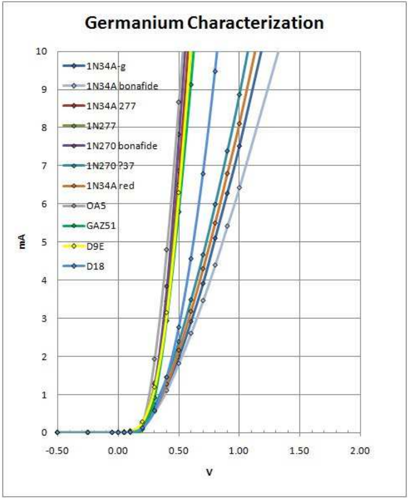

I can get the full 100 mV. With low resistance diodes, the attainable voltage is proportionately less.

My measurement sets typically consist of 20 (10 + and 10 -) V-I measurements. I then regress these measurements against the Shockley equation. (I minimize, using Excel's SOLVER, the sum of squared differences between measured currents and Shockley model currents, by adjusting eta (n) and Is.) Parasitic (bulk) resistance is neglected.

I may be able to get higher currents through a 1N34 by going to a 1.5 volt C cell and shunting across the 5.6 k-ohm voltage divider resistor. I don't want to mess with the 100 k-ohm current sense resistor -- it gets too complicated.

Here's a shot of my test stand:

Mike

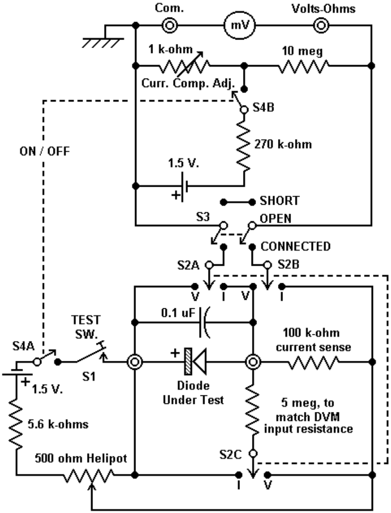
Com.
mV
Volts-Ohms
1 k-ohm
10 meg
Curr. Comp. Adj.
S4B
ON / OFF
270 k-ohm
1.5 V.
+
SHORT
S3
OPEN
CONNECTED
S2A
S2B
V
I
V
I
0.1 uF
TEST
SW.
100 k-ohm
current sense
S4A
+
S1
+
1.5 V.
Diode
Under Test
5 meg, to
match DVM
input resistance
5.6 k-ohms
S2C
500 ohm Helipot
I
V

Test voltage is supplied by a single 1.5 volt AA cell. Design philosophy was to cover voltages (and currents) experienced by diodes in crystal sets. This setup can supply over 100 mV across the diode. The maximum voltage across a diode depends upon the resistance of that diode. With silicon diodes,

1N34A series

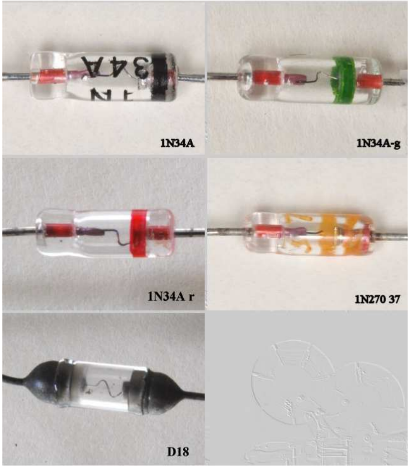
1N34A
1N34A-g
1N34A r
1N270 37
D18

1N270-277 series

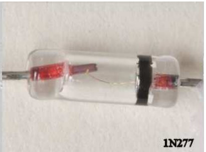

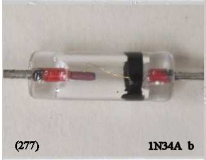

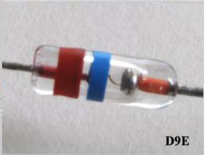

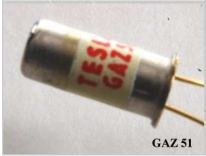

Diode Test Apparatus

Mike Tuggle

Hi, Kevin. My basic diode test circuit is:

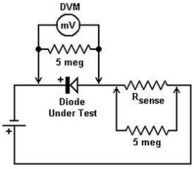
DVM
mV
5 meg
+
Diode
Under Test
Rsense
+
5 meg

Basic Diode Voltage Measurement

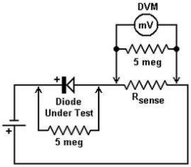
DVM
mV
5 meg
+
Diode
Under Test
Rsense
+
5 meg

Basic Diode Current Measurement

Diode voltage and current measured sequentially by the same DVM in voltage mode. The 5 meg resistor compensates for the 5 meg DVM input resistance. Here's the complete measurement circuit designed with much help from Ben Tongue:

Photo of the lab with my diode test jig, amp meter, and diode vacuum tubes under test.

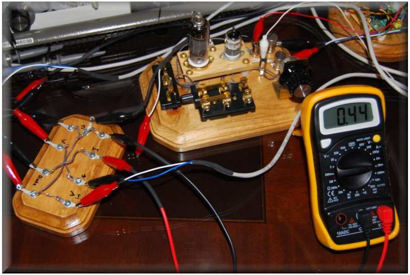

Here we see a set of Germanium diodes and photos of the diodes under test. Immediately you should notice two sets of curves, what I call the "1N34A" set and the "1N270-277" set. In this analysis, the second set appears clearly superior in sensitivity and deservedly so. In the photos it is clear that one bunch of diodes I purchased as 1N34A were in fact better 1N277, the black band and thin gold contact wire as well as common curve unite these. "True" 1N34A's have a more robust contact wire. I also purchased a 1N270 and received pretty orange diodes with the number "37" stamped on it, a fairly robust contact wire and a curve that most closely resembles a 1N34A, go figure. Mercifully, I DID manage to get some bona-fide 1N34A's with the part number on the diode, as well as a single bona-fide ITT 1N270 diode with the part number. These are my base for comparison.

In addition to the standard Germanium Diodes tested above, I also have latched my hands on a few "other" germanium diodes for the fun of testing. These include two vintage Russian types, a D9E (in the 270 class) and a D18 (transitional between 270 and 1N34A). Additionally, I got a few FUZZ diodes popular, I suspect, with the guitar gadget crowd. These are larger packages in metal cases so I cannot see how the internal contact is made other that they are stated to be gold bonded point contact type. Clearly these are highly sensitive diodes in my 270 class, the OA5 looking best of the lot. Interestingly, I recently purchased from good Mr. Peebles a few of his "Holy Grail" ITT FO-215 diodes. My measured characteristic for one of these is a dead laydown on the gold-bonded OA5. From the photos, these are in different packaging with the FO-215 in a traditional glass casing. What makes the FO-215 so great is the fact that its resistance Rd is an interesting 150k ohm or so which matches very well with typical moderate Q tank circuits. Interestingly, I find that an old Russian D18 diode has very similar characteristics to the

FO-215 and so much also be placed into the "Holy Grail" class of germanium diode. Take your pick!

Spreadsheet for calculation of Germanium diode n and Is (modified from Mike Tuggle's spreadsheet)

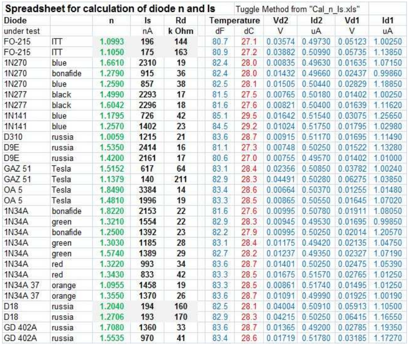

Spreadsheet for calculation of diode n and Is						Tuggle Method from "Cal_n_Is.xls"				
Diode		n	Is	Rd	Temperature		Vd2	Id2	Vd1	Id1
under test			nA	k Ohm	dF	dC	V	uA	V	uA
FO-215	ITT	1.0993	196	144	80.7	27.1	0.03574	0.49730	0.05123	1.00250
FO-215	ITT	1.1050	175	163	80.9	27.2	0.03882	0.50990	0.05735	1.13850
1N270	blue	1.6610	2310	19	82.4	28.0	0.00835	0.49630	0.01635	1.07150
1N270	bonafide	1.2790	915	36	82.4	28.0	0.01432	0.49660	0.02437	0.99860
1N270	blue	1.2590	857	38	82.5	28.1	0.01505	0.50440	0.02829	1.18850
1N277	black	1.4990	2293	17	81.5	27.5	0.00765	0.50180	0.01402	1.00250
1N277	black	1.6042	2296	18	81.6	27.6	0.00821	0.50400	0.01639	1.11620
1N141	blue	1.1795	726	42	85.1	29.5	0.01642	0.51540	0.03075	1.25650
1N141	blue	1.2570	1402	23	84.5	29.2	0.01024	0.51750	0.01795	1.02980
D310	russia	1.0059	1215	21	83.6	28.7	0.00915	0.51170	0.01695	1.11490
D9E	russia	1.5350	2414	16	81.1	27.3	0.00748	0.50250	0.01522	1.13280
D9E	russia	1.4200	2161	17	80.6	27.0	0.00755	0.49570	0.01402	1.01000
GAZ 51	Tesla	1.5152	617	64	83.1	28.4	0.02356	0.50850	0.03782	1.00240
GAZ 51	Tesla	1.1379	140	211	82.9	28.3	0.04491	0.50280	0.06275	1.03850
OA 5	Tesla	1.8490	3384	14	83.4	28.6	0.00664	0.50370	0.01255	1.01480
OA 5	Tesla	1.4810	1996	19	83.3	28.5	0.00865	0.50550	0.01645	1.07020
1N34A	bonafide	1.8220	2153	22	81.6	27.6	0.00995	0.50780	0.01911	1.08050
1N34A	green	1.3210	1554	22	82.9	28.3	0.00945	0.49530	0.01695	0.99850
1N34A	bonafide	1.2500	1392	23	82.2	27.9	0.00995	0.50250	0.02014	1.20570
1N34A	green	1.3030	1185	28	83.1	28.4	0.01175	0.49420	0.02135	1.04750
1N34A	green	1.5740	1389	29	82.7	28.2	0.01237	0.49350	0.02327	1.07190
1N34A	red	1.3220	993	34	83.6	28.7	0.01401	0.50250	0.02475	1.05390
1N34A	red	1.3430	833	42	83.3	28.5	0.01675	0.51570	0.02765	1.01250
1N34A 37	orange	1.0955	1458	19	83.3	28.5	0.00861	0.51740	0.01495	1.01250
1N34A 37	orange	1.3550	1370	26	83.6	28.7	0.01091	0.49990	0.01925	1.00190
D18	russia	1.2040	194	160	82.5	28.1	0.04004	0.50910	0.05913	1.10500
D18	russia	1.2706	193	170	82.9	28.3	0.04215	0.50250	0.06415	1.16550
GD 402A	russia	1.7080	1360	33	83.6	28.7	0.01365	0.49200	0.02785	1.19350
GD 402A	russia	1.5535	970	41	83.4	28.6	0.01719	0.51780	0.03185	1.17270

Averages: n = 1.38 Is = 1315 nA Rd = 52 kOhm

The above spreadsheet is based on measurements of the diodes shown above. For the determination of Is and n, I chose at random two examples from my collection of various diodes, (all germanium in this case) and measured via a modified version of the methodology outlined by Ben Tongue and Mike Tuggle. For this work I have chosen to set Id2 about 0.5uA and Id1 about 1.0uA and then read the needed voltage. This is the inverse of the normal method but the justification is that for

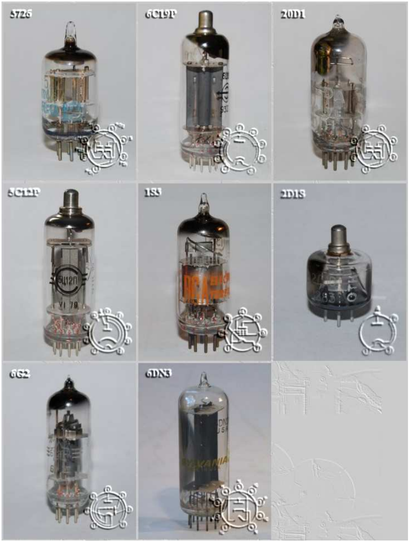

Small currents will continue to flow even with a negative bias to the plate, (Contact potential).

In order to explore this idea, I ran a series of tests with the tubes running at lower operating voltages (effectively diminishing the space charge thus lowering the perveance) as shown:

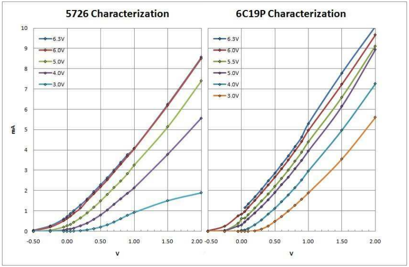

Here you see the impact of lower operating voltages. With the tube operating between 3 and 4 volts the characteristic curves begin to pass through the origin as in regular solid-state diodes and crystals. It is precisely in this voltage range that I found the radio sensitivity, as measured by loudness to my ear, to be most pronounced. It appears that the diode characteristic needs to pass through the origin of the I-V graph for highest sensitivity. This takes me to the 20D1 tube which operates at 9v but only 200mA. This will still take a power supply to use, but the good thing is that the characteristic curve passes through the origin. Running at its design parameters the tube looks a lot like the 6C19P at 4v.

any diode regardless of forward voltage drop the measurement is made at the same part of the LOG I vs V characteristic. This allows comparison between all diodes. I have included the actual room temperature in the calculations in order to eliminate this variable as a source of doubt or error. All the measured parameters as well as the determined results are presented in order to facilitate repeatability.

Schottky Diodes

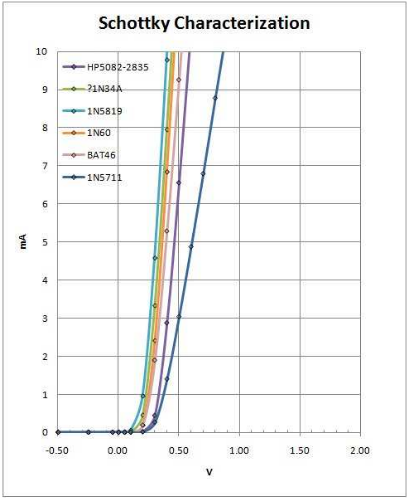

The perveance numbers above range from 0.02 to over 2 mA/V^3/2 (with >2 being good for small-signal detection) and clearly shows my choice of the 6C19P to be an excellent one. I am surprised to see a huge 5.4 for the 6DN3 diode, a color television damping diode. It takes a Novar 9-pin socket and is a large tube so doing more with this tube may have to wait a bit.

Calvert's 2001 published data

Tube	Perveance
6AL5	2.42
6H6	0.50
7Y4	0.58
2X2A	0.017
6V3-A	2.3
6AX4GT	1.42
6AV6	0.085
1A3	0.075

The high-perveance diode tube types I tested show an interesting property in that their characteristic curves do not pass through the origin of the graph. These curves are wholly acceptable as crystal radio detectors although they will require power to operate. What I have noticed with such tubes is that when I turn them off after using them with a radio, on cooling they go through a period of much increased sensitivity (very loud) before fading to nothing. It is as though running them at the full 6.3V lowers their full potential as crystal radio rectifiers. The reason is in the high perveance of the tubes, the anode is already proximate to the space charge before any plate voltage is applied. Effectively, while the plate (anode) "Zero" voltage is measured with respect to ground, the plate itself is still positive with respect to the cathodic space charge.

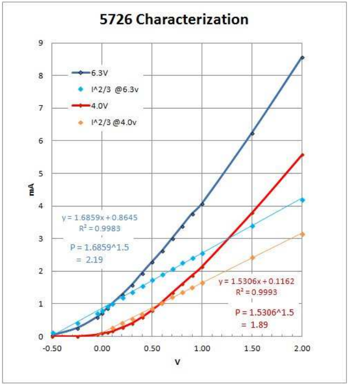

With my measurements calculating the perveance is practical and quickly done, results as follows:

Tube	Perveance	R2
6C19P	2.54	0.997
5726	2.19	0.998
6DN3	5.40	0.997
20D1	2.02	0.985
5C12P	0.32	0.999
1S5	0.063	0.999
2D1S	0.155	0.975
6G2	0.050	0.999

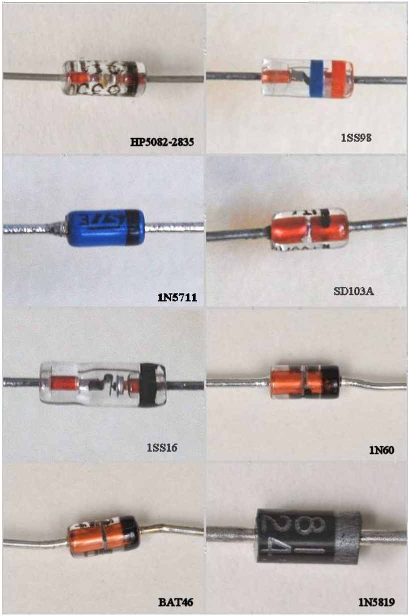
HP5082-2835
1SS98
1N5711
SD103A
1SS16
1N60
BAT46
1N5819

Schottky diodes are very sensitive diodes that work excellently in crystal radios. Their construction and theory are different and I confess to not fully understanding these components. Still, from the characteristic curves, they are excellent! Again, note that you dont always get what you bargain for. Here I found what was supposed to be 1N34A's to be some sort of Schottky of unknown pedigree. The 1N5819, while having the lowest forward voltage drop, has a rather very high Junction Capacitance and will not perform well. Posts on the RaidoBoard Crystal Radio Forum however highly recommend the 1N5711 for crystal sets although the characteristic curve dosnt look that fabulous. Many web pages out there on Shottky diodes, I recommend you do your homework. Of all the schottky's, I note that Ben Tongue recommends most highly the HP5083-2835. The high resistance Rd makes them useful for DX sets with very very high Q tanks. Even so the diodes need to be paralleled with up to 4 or 5 diodes to match correctly the Rd with the tank Rp. I have found these somewhat hard to find and expensive, especially when one requires using several in parallel. I recently measured a few 1SS98 diodes and I discover them to have characteristics extermely similar to the HP and I feel they deserve more attention. Sadly, on ebay they seem to be just as difficult to find and expensive. No free ride!

Spreadsheet below for calculation of Schottky diode n and Is (modified from Mike Tuggle's spreadsheet)

Langmuir-Child Law, the steepness of the I/V characteristic is largely due to the geometry of the tube elements and the volume of electron space-charge between them. JB Calvert's Theory of Vacuum Tubes informs one that this is the property called Perveance in a vacuum diode. Measuring this requires plotting I raised to the 2/3 power against V and taking the slope of the best-fit line through the data. This slope raised to the 3/2 power is the perveance. The following chart illustrates graphically the relation bewtween the tube characteristic and perveance. For the 5726 tube I plot the characteristic in dark blue for the tube heated with the design 6.3 volts. In light blue I have plotted the measurements and raised them to the 2/3 power. One sees immediately that the new curve is linear. The slope of the curve raised to the 3/2 power yields the perveance factor for the tube.

(Note in addition that when running the tube at a more sensitive point with 4 volts only on the heater the perveance is slightly less than spec.).

Vacuum Diodes

When I decided early on to construct a radio based on the vacuum diode (see my Fleming Radio section), I had to find a suitable tube for the project. That began with an extensive search through the datasheets checking vacuum diode properties and reviewing the characteristic curves for a large number of diode tubes. In the above figure I plot the characteristics as best I can on a commpn plot for comparison. I was hoping to find a good candidate with sensitive characteristics and low energy consumption. As you will find, such a beast did not exist. Following this research began a period ot purchase and testing.

I have tested seven different tube types including 6.3V rectifier diodes that take a lot of power to run and are not really suitable for battery use, a 9v dual diode tube (20D1), a pentode/diode tube (1S5) that is designed to run on a battery at 1.5V and 50mA, and a couple miscellaneous but cute tubes. In testing the 1S5, I found the filament never glowed incandescant at 1.5V and, on measuring, was barely sensitive to anything. I cannot imagine this tube would make much of a diode for crystal radio use. I tested different manufacture 1S5 tubes from two different suppliers. No dice. Finally, I tried to push the tube to operate at higher-than-specified voltages. At 4.3V the I-V curve was still very flat, but shifted slightly higher, crossing the Amp axis at about 0.4 mA, (see graph). At 5.5V one of my tubes gave up the ghost and I didnt do more. I pretty much rule the 1S5 out for my crystal radio work. Looks like i'll be needing the power supply when I get around to building/operating that Fleming Radio.

A question remains as to what in fact is different about these tubes to give such different characteristics. While the I/V characteristic of most vacuum diodes pretty much follow the

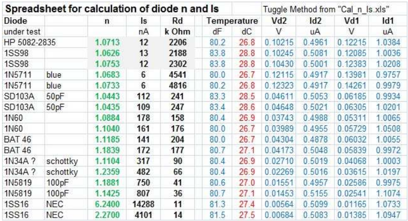

Spreadsheet for calculation of diode n and Is Tuggle Method from "Cal_n_Is.xls"

Diode		n	Is	Rd	Temperature		Vd2	Id2	Vd1	Id1
under test			nA	k Ohm	dF	dC	V	uA	V	uA
HP 5082-2835		1.0713	12	2206	80.2	26.8	0.10215	0.4961	0.12215	1.0384
1SS98		1.0626	13	2188	83.8	28.8	0.10245	0.5081	0.12085	1.0036
1SS98		1.0753	12	2302	83.8	28.8	0.10430	0.5001	0.12383	1.0208
1N5711	blue	1.0683	6	4541	80.0	26.7	0.12115	0.4917	0.13981	0.9757
1N5711	blue	1.0733	6	4816	80.2	26.8	0.12323	0.4917	0.14261	0.9979
SD103A	50pF	1.0443	112	241	83.3	28.5	0.04611	0.5053	0.06185	0.9934
SD103A	50pF	1.0435	109	247	83.4	28.6	0.04648	0.5021	0.06305	1.0201
1N60		1.0884	178	158	80.4	26.9	0.03743	0.4988	0.05311	1.0065
1N60		1.1040	161	176	80.0	26.7	0.03989	0.4955	0.05729	1.0508
BAT 46		1.1185	141	204	80.0	26.7	0.04304	0.4878	0.06032	1.0055
BAT 46		1.1839	172	177	80.7	27.1	0.04173	0.5048	0.05839	0.9972
1N34A ?	schottky	1.1104	317	90	80.4	26.9	0.02710	0.5019	0.04068	1.0003
1N34A ?	schottky	1.2359	482	66	80.4	26.9	0.02269	0.5016	0.03615	1.0197
1N5819	100pF	1.1881	750	41	80.6	27.0	0.01551	0.4957	0.02586	0.9975
1N5819	100pF	1.1425	807	36	80.7	27.1	0.01453	0.5155	0.02541	1.1074
1SS16	NEC	6.2400	14288	11	81.3	27.4	0.00564	0.5099	0.01165	1.0733
1SS16	NEC	2.2700	4101	14	81.5	27.5	0.00684	0.5083	0.01385	1.0947

Averages: n = 1.48 Is = 1275 nA Rd = 1030 kOhm

The above spreadsheet is based on measurements of the diodes shown above. For the determination of Is and n, I chose at random two examples from my collection of various diodes, (all schottky in this case) and measured via a modified version of the methodology outlined by Ben Tongue and Mike Tuggle. For this work I have chosen to set Id2 about 0.5uA and Id1 about 1.0uA and then read the needed voltage. This is the inverse of the normal method but the justification is that for any diode regardless of forward voltage drop the measurement is made at the same part of the LOG I vs V characteristic. This allows comparison between all diodes. I have included the actual room temperature in the calculations in order to eliminate this variable as a source of doubt or error. All the measured parameters as well as the determined results are presented in order to facilitate repeatability.

Silicon Diodes

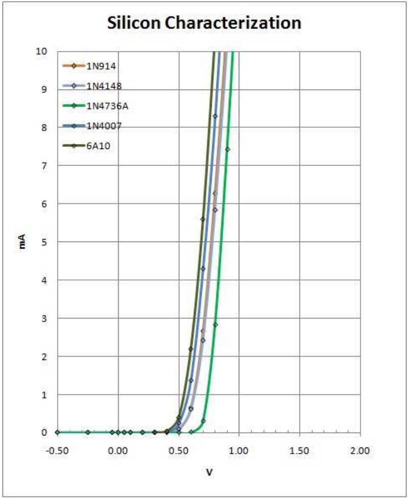

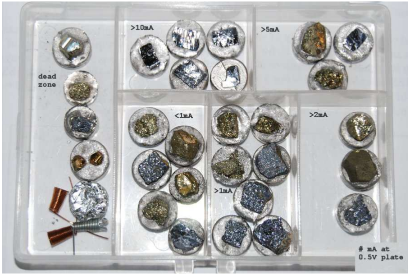

In the photo I indicate groupings based on an easy measure of performance. I note the current in milliamps for each set where the plate voltage is set at 0.5V. The greater the current the better are your chances to get a sensitive crystal, assuming Ohm's Law is not followed! Crystals in the "dead zone" on the left will have their woods metal re-melted for new detector crystals and the bad ones tossed, its tough love for crystals. I find that easily half the potted crystals I make are tossed this way, and only a few can be considered superlative.

Galena" samples (Tintic Utah, or Leadville Colorado) there are numerous hot spots under virtually every place I touch the probe, but in general the sensitivity is good to so so. These crystals are very kind to work with in terms of finding spots and avoiding frustration. Mirror galena on the other hand may have quality hot spots, but any hot spots at all are rare and frustratingly difficult to locate. Here my Philmore detector crystal shines with an almost ideal "Galena" response. To chase down this rabbit I purchased some lovely mirror galena from Sweetwater Missouri. I broke off a few appropriate-size chunks to pot in woods metal and test. At first I was very excited with the high currents I was seeing at moderate voltages. Figured I had struck gold. When these crystals failed miserably to rectify anything in my radios, I re-measured things in both forward and reverse directions. These crystals obey Ohm's law and act like typical resistors, not suitable for radio work at all.

For my pyrite crystals the work has been especially tedious and frustrating. With one of the crystals one "hot spot" alternated, entirely on its own, between hot and bad while I was making the measurements. I would start over and over, sometimes getting interesting readings then suddenly it would drop to low values and I'd start over, back and forth. I present this data as best as I have measured, and I don't intend to go back! You see at least one of the crystals, my "China 1", (from a lead/silver mine in Hunan) gave a sweet classic-looking curve. More to the point, "ideal" curves for natural minerals, are difficult to come by. Most crystals you use will be less than ideal. The good news is that, while listening to your set, poking about for a good spot is far easier than what I have gone through to produce these curves. Your ear will take care of you!

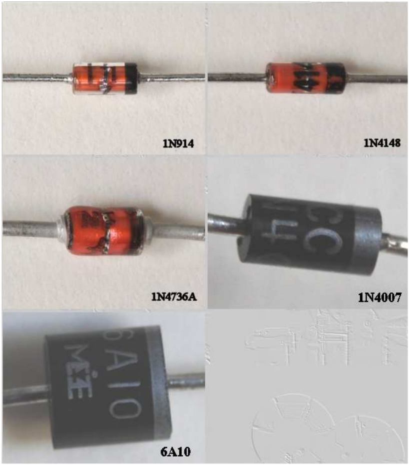

Silicon diodes have good characteristics, but an unacceptably high forward voltage drop making them a very poor choice for crystal radio unless used with bias. The 1N4736A is a Zenner diode.

Spreadsheet for calculation of Silicon diode n and Is (modified from Mike Tuggle's spreadsheet)

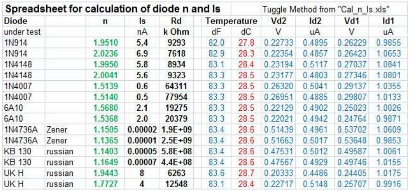

Spreadsheet for calculation of diode n and Is Tuggle Method from "Cal_n_Is.xls"

Diode		n	Is	Rd	Temperature		Vd2	Id2	Vd1	Id1
under test			nA	k Ohm	dF	dC	V	uA	V	uA
1N914		1.9510	5.4	9293	82.0	27.8	0.22733	0.4895	0.26229	0.9855
1N914		2.0236	6.9	7618	82.9	28.3	0.22354	0.4857	0.26423	1.0653
1N4148		1.9950	5.8	8934	83.1	28.4	0.23194	0.5117	0.27037	1.0841
1N4148		2.0041	5.6	9323	83.3	28.5	0.23177	0.4803	0.27346	1.0801
1N4007		1.5139	0.6	64311	83.3	28.5	0.26320	0.5041	0.29137	1.0355
1N4007		1.5140	0.5	77954	83.3	28.5	0.26951	0.4885	0.29807	1.0133
6A10		1.5680	2.1	19275	83.3	28.5	0.22129	0.4902	0.25023	1.0026
6A10		1.5368	2.0	20379	83.3	28.5	0.22021	0.4942	0.24764	0.9871
1N4736A	Zener	1.1505	0.00002	1.9E+09	83.4	28.6	0.51439	0.4961	0.53702	1.0609
1N4736A	Zener	1.1365	0.00001	2.5E+09	83.4	28.6	0.51663	0.5017	0.53648	0.9853
KB 130	russian	1.1403	0.00005	5.8E+08	83.4	28.6	0.47531	0.5012	0.49587	1.0061
KB 130	russian	1.1649	0.00007	4.4E+08	83.4	28.6	0.47567	0.4929	0.49746	1.0155
UK H	russian	1.9443	8	6263	83.6	28.7	0.20333	0.4486	0.24405	1.0175
UK H	russian	1.7727	4	12548	83.1	28.4	0.22717	0.5148	0.25707	0.9916

Averages: n = 1.52 Is = 2 nA Rd = 2.4E8 kOhm

The above spreadsheet is based on measurements of the diodes shown above. For the determination of Is and n, I chose at random two examples from my collection of various diodes, (all silicon in this case) and measured with a modified methodology to that outlined by Ben Tongue and Mike Tuggle. In this case, with the radically different forward voltage drop of these diodes from Germanium or Schottky diodes, I have kept the Id values constant (about Id2=0.5uA and Id1=1.0uA) and varied Vd. I have included the actual room temperature in the calculations in order to eliminate this variable as a source of doubt or error. All the measured parameters as well as the determined results are presented in order to facilitate repeatability.

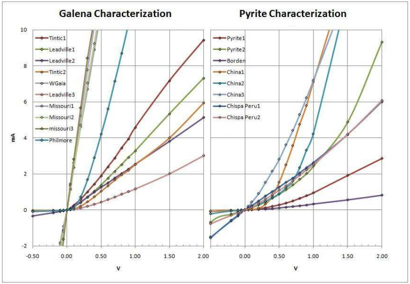

The above curves demonstrate the wide variation in properties and qualities that can be found in natural crystals of galena or pyrite, the two most common and best quality natural stones. In each crystal test I first poked around the crystal for some time to 1) determine a typical sensitivity for the crystal in question and 2) to locate the best hot spot with which to test. This turns out to be a non-trivial exercise on the diode test setup. In a crystal radio one need only listen for the loudest spot. With the test setup one needs test both the forward and reverse current in order to determine if the whisker is on a hot spot or not. Very tedious work! (In retrospect, if I were starting over making a test jig, I would definately add a DPDT switch to readily change between forward and reverse current measurements. I'd probably toss in a rheostat for fine adjustments as well).

For many crystals there are limited number of possible hot spots, but these may be hot indeed. For most of my "Steel

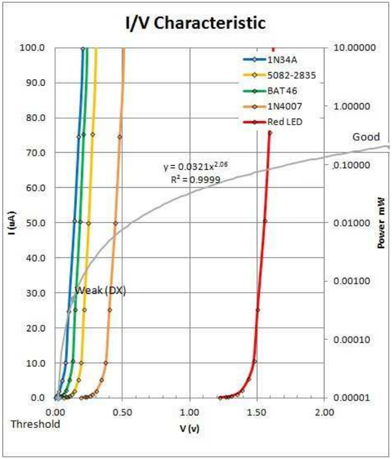

This plot at left takes some actual diode I/V measurements rather than the models presented below and plots them along with the above RF voltage across the LC circuit versus RF power in mW. Here one sees that pushing for very very low Vf (via high Is / low n combinations) pushes the limit of detectable signal power from the antenna. At such minute power levels I imagine the reverse leakage current becomes significant and probably offsets the low operating point gains sought.

Crystal Galena and Pyrite Diodes

Light Emitting Diodes

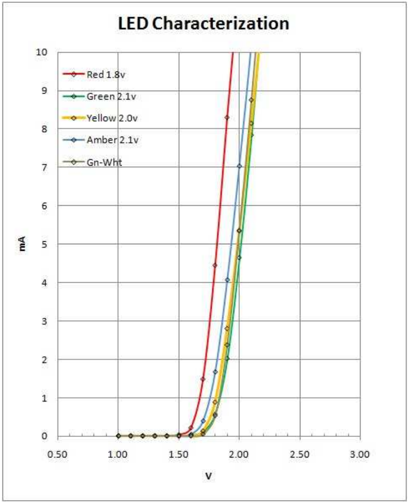

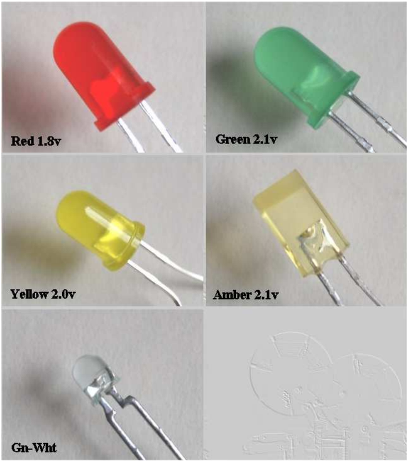

As long as I am measuring various and sundry diodes, I figure I ought to include that most ubiquitous of modern diode, the LED. Found everywhere, these diodes are rapidly becoming the low-energy light source of choice for many lighting applications. I have read occasionally of someone asking whether these ought to be useful for radio applications as well. To this question the answer is generally a resounding "No!". The turn-on voltage is waaay above any reasonable value expected to be delivered by an antenna to a crystal set. Still,

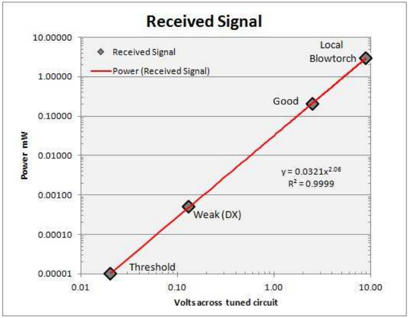

The above is an interesting plot taken from the text discussion in "Crystal Set Analysis" by Berthold Bosch. It presents received signal strength across the tuned circuit for various scenarios from threshold audibility to local blowtorch. This plot should be considered a single example specific to his location and antenna/ground system. In the text he describes his antenna an an inverted L 43m long (140') and 10m high (32'), an excellent antenna most of us do not have the real estate to erect, but offsetting this is a poor ground with Rg = 210ohms. Given the offsetting conditions I would imagine this is a good generic example of what received signal voltage and power levels presented across the LC circuit to the diode are likely to be. Bosch cited 40nV as the threshold of audible detectability (impedance matched conditions with 16kohm RF impedance / 4kohm DC phones) and for this plot I pushed it back to 20nV as it gave a superior regression fit.

Additionally I have plotted RF signal Rs as orange circles from published coil unloaded Q (see my section on Coil Q). To convert from Qu to Ql I have made the assumption that loaded Q is about 1/5 unloaded Q for low-end sets while in performance sets it may approach 1/2.5 the unloaded Q. This is only an estimate then, but starting from actual data. To make the conversion to Rs I simply plugged the data into the standard equation R = 2pi*f*L*Q using a scaling factor to divide down Qu to Ql from high Q (1400/2.5) to low Q (100/5). This provides a nice visual display of the expected range for Rs in many crystal sets. I plot the Rs against the value of Ql used. Note that these data are for the case at about 1MHz and inductances between 200 and 300uH, as published.

Matching the diode to the tank is a matter of finding a diode with both sensitive qualities (low n) and an Rd close to the Rs presented to the diode (as per Ben Tongue's suggestions for Peak and Square Law detection). Those diodes with Is's in the 100 - 250 nA (100-280 kohm for n=1.1) or so range have the possibility to match while connected to the top of the tank coil without the use of Q-lowering taps, while untapped matching to high-end high-performance big-Litz baskets will require diodes with Is's in the 35 - 60 nA (500-800 kohm for n=1.1) range. Most "typical" germanium diodes have high Is values (>500, >60kohm) and will require a tap. Diodes such as FO-215, BAT 46, 1N60 and GAZ51 can be matched to many tanks without taps. For tanks with high quality Litz coils one will be using HP 5082-2835 or 1SS98 diodes, generally 3 to 4 in parallel to lower the Rd to the desired range. (Curiously, I have not found any diodes with Rd values in the 300 to 4500 kohm range.) Note that both resistance and impedance are frequency-specific so a match at one frequency will not remain matched across the broadcast band.

this simple answer avoids the actual question, what in fact does the characteristic curve of a LED really look like? Where is the turn-on voltage with respect to the published junction voltage, (assuming you can find that).

To provide just such a look I visited my local electronics store and bought a small handful of LED's, most with the junction voltage listed and took them home to measure. Typical LED junction voltages seem to range from about 1.8v to 2.1v or more. The turn-on voltages look closer to 1.6v-1.7v. Anyone used to working with Carborundum crystals or silicon diodes will be used to biasing their rectifier to get good sensitivity. These LED's once on have a very sharp rise and with a proper bias should work quite well as detector diodes. As a bonus you will get a sweet glow as well. OK, not as cool as the glow of a vacuum diode, but certainly more sensitive!

Spreadsheet for calculation of Light Emitting Diode n and Is (modified from Mike Tuggle's spreadsheet)

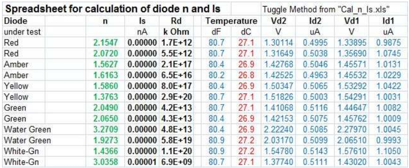

Spreadsheet for calculation of diode n and Is						Tuggle Method from "Cal_n_Is.xls"			
Diode	n	Is	Rd	Temperature		Vd2	Id2	Vd1	Id1
under test		nA	k Ohm	dF	dC	V	uA	V	uA
Red	2.1547	0.00000	1.7E+12	80.7	27.1	1.30114	0.4995	1.33895	0.9875
Red	2.0720	0.00000	5.5E+12	80.7	27.1	1.31649	0.5038	1.35690	1.0745
Amber	1.5627	0.00000	2.1E+17	80.4	26.9	1.42768	0.5046	1.45571	1.0131
Amber	1.6163	0.00000	6.5E+16	80.2	26.8	1.42525	0.4963	1.45532	1.0229
Yellow	1.5860	0.00000	8.0E+17	80.4	26.9	1.50347	0.5065	1.53292	1.0422
Yellow	1.3763	0.00000	2.9E+20	80.7	27.1	1.51826	0.5003	1.54291	1.0031
Green	2.0490	0.00000	4.2E+13	80.7	27.1	1.41068	0.5116	1.44647	1.0082
Green	2.0650	0.00000	4.3E+13	80.4	26.9	1.42153	0.5075	1.45762	1.0009
Water Green	3.2709	0.00000	4.8E+13	80.4	26.9	2.22240	0.5085	2.27970	1.0045
Water Green	1.9273	0.00000	5.8E+19	80.9	27.2	2.03170	0.5099	2.06510	0.9993
White-Gn	1.4366	0.00000	1.1E+20	80.9	27.2	1.54780	0.5143	1.57610	1.1050
White-Gn	3.0358	0.00001	6.9E+09	80.7	27.1	1.37740	0.5111	1.43020	1.0043

Averages: n = 2.01 Is = Tr nA Rd = 3.8E19 kOhm

The above spreadsheet is based on measurements of the diodes shown above. For the determination of Is and n, I chose at random two examples from my collection of various diodes, (all led's in this case) and measured with a modified methodology to that outlined by Ben Tongue and Mike Tuggle. In this case, with the radically different forward voltage drop of these diodes from Germanium or Schottky diodes, I have kept the Id values constant (about Id2=0.5uA and Id1=1.0uA) and varied Vd. I have included the actual room temperature in the calculations in order to eliminate this variable as a source of doubt or error. All the measured parameters as well as the determined results are presented in order to facilitate repeatability.

Why know the values of Is, n, and Rd of your diodes?

I have been attempting to get a handle on the main diode parameters Is, n, and Rd and how they impact the operation of a crystal radio. My method of choice is to use a graphical approach, plotting charts where variations between plotted parameters become apparent and easy to visualize. I have developed an interesting chart from the measurements above that compares many diodes with respect to Is, n, and Rd, and to understand how this fits into a matching situation with a fuzzy indication of typical tank impedance limits.

Ultimately one seeks a diode with an Rd that matches with and conjugate to the impedance it sees from the tank (+antenna). In the following discussions I will start with the cross-plot of Is to Rd and then proceed to a look at how received RF power plays a part in the selection and why low Is and n are desired.

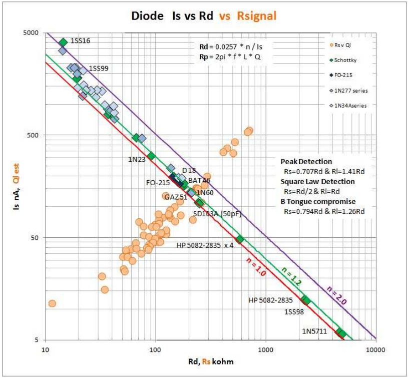

The plot above shows the relationship between diode Is and Rd. Rd is calculated via the equation Rd = VT * n / Is where VT is the thermal voltage = k * T / q = 0.0257V at room temperature and Is and n are from the measured diodes. k is Boltzmann's constant = 1.38E-23 J/K and q is the electron charge = 1.609E-19 coulombs. On such a plot I can show lines of constant n, and plot the values for individual diodes for which I have spent considerable effort to determine the parameters Is, n and Rd (where Is is the diode saturation current, n is the ideality factor, and Rd is the diode resistance). On the plot it should be evident how changes in n or Is impact a diodes' Rd.

KEVIN’S WEBSURFER HANDBOOK IV
FOR CRYSTAL RADIO

ANTENNAS – GROUNDS & TUNING UNITS

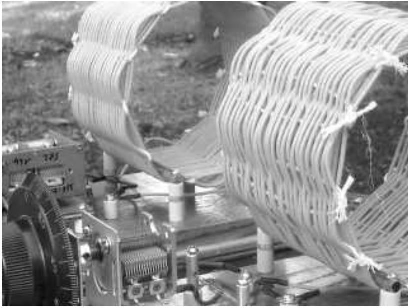

Kevin Smith
2011

Printing / Binding Instructions

1. Choose “fit to page” in print menu
2. Print document double sided on letter size paper
3. Cut the entire printed document in half
4. Fold over making sure the page numbering is continuous
5. For the cover: Print just the first page on card stock paper
   Cut the cover in half as well
6. Assemble the covers on the document
7. Punch the left side for a binding, spiral or comb as desired

http://www.lessmiths.com/~kjsmith/crystal/catalog.shtml

KJ Smith

NOTES

INTRODUCTION

The contents of this fourth handbook concerns itself primarily with antenna and grounds as applied to crystal radio (avoiding all the HF world). Finally, It was inspired by my realization that I really didn't understand antenna tuning as well as I thought I did, and especially I didn't understand the Tuggle Circuit. Very little qualitative discussion exists on the web for this in fact. The references I have rounded up here in this handbook should go a long way towards "rectifying" the situation.

In this booklet I start with a good discussion RF field strength, historically, qualitatively and quantitatively. I then bring in some considerations on grounds and especially a discussion of ground resistances regionally and locally as these are essential to modeling my, or any, antenna prior to designing a proper antenna tuner unit. What the follows are excellent resources helping with the design and understanding the functioning of a good "Tuggle Front End". The handbook will help the beginner to quickly get up to speed and allow the experienced builder to find endless new ideas. This is not a book of Hookups or circuit designs, that is covered in my Catalog of Crystal Hookups, nor is it a tutorial on Crystal Radio, that can be found in my Handbook Volume 1.

Finally, although I intend this handbook to consider the antenna / ground as part of the crystal radio system, it is true that antennas receive potentially harmful spherics as well. Due to this, I have included as a last chapter (but not as an afterthought) a section concerning lightening protection.

Much of the material in this handbook is copyright for which I have not sought permission. Therefore this is not presented for publication or copy, and certainly not for profit. It is only my personal resource. I encourage anyone finding this copy to pursue ON THE WEB the web pages identified within. I include the name of the author and web address of each section. I wish to sincerely thank every author presented for their excellent pages and ask forgiveness for my editing into this handbook.

NOTES

NOTES

CONTENTS:

1 **Field Strength Calculations: A History**
Radio-Time Traveler

9 **RF Basics**
Martin D. Stoehr

19 **RF Field Strength**
Kenneth A. Kuhn

23 **Field Strength Calculations**
Radio-Time Traveler
- 23 Part i: Ground Conductivity
- 27 Part ii: Measurements
- 36 Part iii: Calculating
- 45 Part iv: Field Strength Calculator One

49 **Antennas and Ground System**
Kenneth A. Kuhn

61 **Antenna Mathematical Model**
Kenneth A. Kuhn

79 **Grounds**
Jim Lux

85 **USA Conductivity Map**
US Government

93 **Harris County Soil Resistivity**
Harris County

95 **How to measure A-E Impedance**
Ben H Tongue

101 **Dummy Antenna**
F.E. Terman

102 Antenna Matching
Kenneth A. Kuhn

109 Working With Antenna Impedance
Jack Bryant

113 Experiments with an Antenna Tuner Unit
Dick Kleijer

117 Component Values for an Antenna Tuner
Dick Kleijer

125 Notes on simulations for a Tuggle Front End
Kevin Smith

137 Analysis of the Tuggle Frong End
Ramon Vargas Patron
  137 Part i: Formula Derivation
  142 Part ii: Tuggle Tuner
  158 Part iii: Model

171 How to Hook Up A Variable Capacitor
Darrel Boyd

174 Variable Capacitor Wired Wiperless
Jim KR1S

175 Analyzing Three Antenna Analyzers
Greg Ordy

207 What You Should Know: Lightning Protection
Joseph H. Reisert

214 Notes

References:

1. "The "Grounds" for Lightning and EMP Protection", by Roger Block, published by PolyPhaser Corporation.

Finally, there should be a large diameter (#4 AWG or larger) copper wire connecting the equipment to the Earth ground. The shorter the wire, the better. Additional information on grounding can be found in Reference 1.

Mains and Power Supply Protection

Just because your radio and antenna are miles away from a lightning strike does not mean that you are protected. Lightning often strikes power lines and produces a large voltage surge or spike that can be transmitted for miles on the main power lines.

Therefore, for maximum protection, all power line interfaces should include a transient voltage surge protector. These devices are becoming quite common and inexpensive. Again, there are simple protectors and those that may include additional protection with built-in line inductors. Just make sure that the surge protector is placed between the power lines and the equipment power supply.

Summary

As stated at the beginning of this paper, lightning damage can be extensive and costly. We have tried to provide you with an overview of things to consider when installing a radio and antenna system but only you can determine how much protection is required. Astron Wireless Technologies, Inc. has access to various lightning protection solutions for most applications. Please call for additional information on lightning protection devices and how they can help protect your investment.

Field Strength Calculations: A History

Radio-Time Traveler
http://radio-timetraveller.blogspot.com/2012/12/field-strength-calculations-history.html
16 December, 2012

A previous three-part series on RADIO-TIMETRAVELLER delved into Field Strength Calculations. It covered ground conductivity's effects on signal strength, measurements quantifying signal intensity, and how to use the FCC Groundwave Conductivity Graphs to calculate expected received signal strength. Mathematical formulas, somewhere, produced those graphs. What is their history? Might we use a simplified formula to calculate expected received signal strength for our DX purposes?

Let's continue with the story behind Field Strength Calculations and explore the 50 year quest for accuracy in calculating signal strength by mathematical formula. It is an interesting tale. We will finish with a handy field strength calculator program I wrote using a simplified formula.

When we talk about field strength, we are really talking about radio propagation - the behavior of radio waves when they are transmitted or propagated from one point on the earth to another, or into various parts of the atmosphere. In our formula quest, we will mostly be concerned with those signals that hug the ground, or "ground wave". It may surprise many who are new to the hobby of mediumwave DXing that daytime ground wave range for a mediumwave signal might extend out to as much as several hundred, and in extreme cases, nearly 1000 miles!

Accurate formulas for calculating expected signal strength at mediumwave and longwave frequencies took many years to develop. Radio originally inhabited the longwaves in its infancy. Many of Marconi's early broadcasts, including his 1905-1906 transatlantic tests, were sub-100 KHz. The trend would be decidedly upward in frequency and downward in wavelength.

At the end of World War I, a fierce battle ensued between the US government and the Department of the Navy over control of the airwaves. The Department of Commerce eventually won and became master of the air and the regulatory agency for commercial radio. They started by establishing two broadcast frequencies: 833 KHz (360 meters) and 619 KHz (485 meters). The Federal Radio Commission took charge in 1926, lasting until 1934 when the current Federal Communications Commission was formed.

Throughout the early years of radio, interest mounted to quantitatively determine the service area of broadcast stations. Engineers redoubled their efforts to derive an accurate attenuation formula. The radio world was focused on accuracy of measurements at broadcast frequencies.

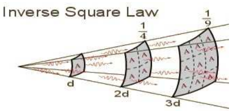

"Accuracy" is the key word here. The Inverse-Square Law, as applied to physics, had been commonly known since Isaac Newton's day in the 1600s. Applied to radio, it stated that the power density of the wave is proportional to the inverse of the square of the distance from a point source. In other words, doubling the distance from a transmitter means that the power density of the radiated wave

range with low VSWR up to about 2 gHz. However, it too must be replaced after a strike although it will not necessarily warn the user by going to a short circuit so preventative maintenance is required.

The quarter wavelength shorting stub is becoming a very popular device, especially above 800 MHz where system bandwidth is generally narrow. It consists of a tuned quarter wavelength shorted coaxial type transmission line that is placed directly across the transmission line. Simple types have a narrow bandwidth, typically 10'-. bandwidth, but are low loss and inexpensive.

The optimum place to locate an input protector is at the entry point to the building where the equipment is located. Don't forget to provide a low impedance ground connection to the protector as described below. Always try to keep the lightning and the protection devices outdoors wherever possible!

Grounding

The most important lightning protection is a good low impedance Earth/ground connection to the associated equipment. The Earth ground connection should be a copper plated rod preferably at least 5-8 feet in length driven into the ground. This ground rod should be located as close to the equipment as possible, typically just outside of a building at the entry point of the antenna feedlines.

Greater protection can be provided by using additional ground rods spaced at least 8 feet from and connected to the original rod. Substituting plumbing, power ground return and other "so called" grounds for a ground rod is definitely not recommended.

will not discharge through the antenna. Furthermore, the boom or mast should be grounded to the mast or tower. More on this shortly. Don't forget to ground guy wires that are used on stabilize towers. They are just as likely to be hit since they extend over a wide area around the tower.

Input Protection

Input protection is typically provided by a lightning or surge protector at the input (or antenna side) of a radio. There are three major types of lightning protection devices for the radio input. They are the spark gap, the gas discharge tube and the quarter-wavelength (1/4) wavelength shorted stub. Each method has its pluses and minuses.

The spark gap is the oldest know lightning protection having probably been invented by Ben Franklin! Basically it consists of two balls or points closely spaced and directly across the transmission line. When a strike occurs, the high voltage present will jump across the points and be conducted to ground. RF transmission devices for 50 Ohm systems such as this have been around since the 1950's.

The shortcoming with this older device is that it may not protect against a weak or lower voltage strike. Adjustment at the factory may produce varying voltage breakdowns. Furthermore, once a strike is taken, the device may fail or short circuit so maintenance is required.

A more recently developed input protection device is the gas tube. It works in a similar manner to the spark gap but can be designed to operate reliably at much lower voltages down to 100 Volts for low power circuits and 250 Volts or higher for higher power applications. Another advantage is device is that it can be designed to operate over a very broadband frequency

at that new location is reduced to one-quarter of its previous value. But did it apply?

"Free-space" formulas calculating signal loss in the vacuum of space or "perfectly conducting earth" using the so-called inverse-square law are indeed accurate for those environments. But the Earth is not a perfect conductor, nor does it represent perfect-world conditions. Free-space formulas alone are not usable for our purposes. You will find many of them on the web, even calculators, purporting to deliver a signal strength solution for a given transmitter-to-receiver distance. They can be ignored as inaccurate. In fact, they are not even close.

Arnold Sommerfeld, 1868-1951

Mathematicians started with a "plane earth" (flat earth) theory when they first envisioned a signal attenuation formula. Brilliant, German-born genius Arnold Sommerfeld, nominated a record 81 times for the Nobel Prize during his lifetime, solved the plane earth general problem by 1909, publishing signal attenuation graphs in 1911. Bruno Rolf, basing his work on Sommerfeld's findings, published more attenuation graphs in 1930, some 21 years later. From this information, the Federal Radio Commission compiled formulas and curves, published in 1931. They were used in hearings and allocation matters at least until 1933. It was just the beginning.

In the intervening years from 1909 to 1930, four more scientists obtained independent solutions of the Sommerfeld problem which agreed with the 1909 solution. That is, except for one difference - an inverted mathematical sign. Apparently none of these authors noticed this discrepancy until the FCC's K.A. Norton, in a letter to the editor of "Nature" in 1935, pointed it out and showed that it was responsible for the anomalies in propagation predicted by the Sommerfeld-Rolf graphs. Norton in 1936 was able to construct a universal curve for prediction of field strength at relatively short distances.

Focusing on the plane earth theory, Sommerfeld expected that the surface or ground wave would be only slightly affected by the curvature of the earth since it is guided around the earth's curve in much the same manner as an electric field can follow around the bend in a wire with a comparatively small loss of energy. This explains the general success of the Sommerfeld plane earth formula at distances far beyond the line of sight. However, two major roadblocks to accuracy still existed.

The first, and most important, was "diffraction". The other, "intermediate distance".

Out beyond what is called the "radio horizon", radio signals undergo atmospheric and ionospheric diffraction, that is, refraction and scattering caused by atmospheric irregularities. This enables AM radio signals in low-noise environments to be received well after the transmitting antenna has dropped below the horizon. It has been shown theoretically that the ground wave attenuation factor at mediumwave frequencies is very little affected by diffraction at distances less than about 55 miles, the approximate "radio horizon" for mediumwave.

DIFFRACTION

---

115 VAC we obtain from a line cord). Let's discuss each of them separately.

Antenna Mounting

It is well known that lightning statistically strikes the highest electrical conductor in an area and then follows the lowest resistance and shortest path to ground. Since antennas are usually mounted in high places, they are very susceptible to lightning strikes. Most antennas have a metallic boom and the elements are often attached directly to the boom so they are a likely target for a lightning strike.

Therefore, the antenna location and how it is mounted is probably the most controversial topic when any discussion of lightning occurs. Ben Franklin gave rise to the theory that the lightning was electricity and found this out when he almost was killed by a lightning strike conducted down the wire holding down his kite. Franklin is sometimes credited with the ball discharger and the pointed rods on houses. As a result, to this day lightning rods with grounding wires are a part of folk law and many are installed on high buildings and homes, especially in areas prone to lightning activity. Properly installed and grounded, these devices surely do work.

Nowadays, new controversy has resulted with the use of spline balls, static dischargers and wicks mounted on antennas and the top of towers. These devices are said to provide a constant discharge thus decreasing the potential for a direct strike. Some users claim a diminished amount of direct hits after installing these devices.

Suffice it is to say, if at all possible, don't mount your antenna on the highest building or tower. Place it a few feet lower and hopefully the fickle lightning bolt, if it generates a direct hit,

The Nature of Lightning

Lightning can form and stay in the upper atmosphere. This is often a beautiful sight to behold as the bolts jump from cloud to cloud. On the other hand, when lightning leaves the clouds and strikes the ground or a tall object, it can inflict instant destruction and even death to those unfortunate to be near the strike.

The energy present in a lightning bolt can be considerable and a direct hit will inflict the maximum damage. It is estimated that a typical bolt may contain a potential of millions (1,000,000's) of volts thus generating currents up to 100,000 amperes! That is very destructive energy. At the same time, the heat in the bolt can have a temperature up to 30,000 K, hot enough to start fires.

A lightning bolt will often drag or jump along the ground. Therefore, there can be considerable damage in a wide area (even hundred's of feet) surrounding the strike. Just ask a dairy farmer what happened to his cows that tried to hide beneath a tree that had a direct lightning strike.

Lightning Protection in General

Lightning protection must be examined from four distinct directions. First off, the place where the antenna is mounted (such as on a tower) is important. Then there must be input protection from the lightning strike itself, typically in the form of a huge and rapid build up of voltage and current at the input to the radio. Thirdly, a proper ground system must be employed to rapidly conduct the lightning bolt energy away from the radio. Finally, protection is required at the output or main power supply such as the line voltage supply (e.g. the

---

Norton, also in 1936, provided curves for greater distances in the diffraction region. These curves, however, were based on an incompletely developed theory. Mathematical solutions were being developed in Europe, and were two years away from completion. Europeans van der Pol and Bremmer published their paper in 1938, offering a more complete solution of the radio diffraction problem for propagation. Never-the-less, the calculation of field strength beyond the radio horizon still proved troublesome, though Norton's remarkable work clarified Sommerfeld's ground wave propagation theory.

The radio horizon at the longer wavelengths, including mediumwave, can be calculated quite simply.

$$\textbf{Radio Horizon} = 8\text{x}10^4/f^{1/3} \text{ (meters)}$$

$$f = \textit{frequency in MHz}$$

For example, the radio horizon for a station transmitting on 600 KHz is about 59 miles.

By 1940 the FCC, through the work of K. A. Norton, had developed a practical method for constructing curves approximately representing the theoretical predictions. The method used the flat earth theory of Sommerfeld out to a distance of about 80 kilometers, and the diffraction theory of van der Pol and Bremmer at relatively great distances, those in excess of 200-300 kilometers depending on frequency and ground constants.

The gap in the curve was still intermediate distances. The Watson transformation, a theory originally described in 1918 by English mathematician G.N. Watson, was an attempt to connect the two. How to incorporate it into the general theory, to calculate the intermediate distances, was still the final problem. In the curves published in 1940, the gap was simply sketched in by a draftsman.

In 1952, George A. Hufford of the National Telecommunications and Information Administration provided a basis for unifying the ground wave prediction methods of Sommerfeld with Watson's diffraction transformation. It had been 43 years since Sommerfeld's 1909 thesis. There was finally light at the end of the tunnel. New curves were added in 1954 for very low conductivity. These were quite accurate, although freehand drawing was still necessary to join the Sommerfeld curve segment to the curve segment calculated for the diffraction field at relatively great distances.

Then in 1958, Hendricus Bremmer, the same Bremmer who in 1938 brought the general solution to the diffraction problem, provided correction terms which completed the search for the practical formula. Engineers could finally calculate ground wave field strength with accuracy. It had been 50 years in the making. The formula was born.

The FCC curves were considered satisfactory for regulatory purposes until it became necessary to convert to metric units toward the end of the 1970s. In a 1979 FCC report, it was recommended that a computer program be written for recalculating the curves using the methods in Bremmer's 1949 book. The program was subsequently used to produce new FCC curves in 1985 which agree within 1 to 2 decibels with the previous curves. However, the 1979 computer program was mathematically deficient in its ability to cover all the range of intermediate distances, and the great distance values it computed were shifted upward to force a match in the middle. FCC curves drawn for the X-band, 1605-1705 KHz, are the most recent. They are the result of precise calculations of field

What You Should Know About Lightning Protection

http://www.astronwireless.com/topic-archives-antenna-lightning-protection.asp

By Joseph H. Reisert

Antenna Lightning Protection

Antenna manufacturers are often asked about lightning and whether their antenna is lightning proof. This is not a simple question to answer. In this application note we will attempt to describe some of the statistical properties of lightning strikes and give some recommendations on how to best protect your installation from damage due to a lightning strike.

It is important to point out at the start that lightning protection is primarily a function of how much time and money you are willing to spend. Obviously, the more expensive the radio and the importance of system connectivity, the more robust your protection should be.

An Introduction to Lightning

Lightning has been around since the beginning of time. In fact, at this very moment, lightning is striking somewhere on the earth. In the lower 48 US states and Canada, lightning is most prevalent in Florida (with an average of 70-100 storm days per year!) and the Rocky Mountains especially in the Colorado area (with an average of 70 storm days per year). In California and the Northern parts of the USA and Canada the likelihood of encountering lightning diminishes to about 10-30 storm days per year. Lightning is most common starting in the spring and ending in the fall with a large peak in the summer but it can occur at any time of the year.

analysis, the author spends some time looking at the accuracy of *low-power SWR testers*. For the Autek RF-1, he reports *good accuracy* for loads from 12.5 to 200 Ohms. He also measures the SWR accuracy of the MFJ-259, and describes it as *very good* across the HF bands.]

Last update: Friday, May 07, 2004 01:40:39 PM

---

strength over the full range of distances of interest, including the previously troublesome intermediate distances.

And thus we have the short version of the history to achieve accuracy in field strength formulas. Stay tuned for the next installment, a program to calculate field strength, based on a simplified formula.

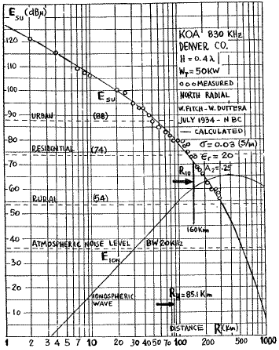

Original measured vs. calculated f/s values for KOA, Denver, 1934 compendium was available. Dan's article, *T-Time for the Analyzers*, is on page 40 of the compendium. Dan has made a number of other contributions in the area of transmission line analysis. Please see his site for more information.

[Epilogue 1 - The ARRL Antenna Compendium #5 has an article entitled: *Measuring Antenna SWR and Impedance*, by Francis Merceret, WB4BBH, on page 148. The author was initially looking into the reliability of SWR measurements. One way to measure SWR is to compute it from the so-called *forward* and *reflected* power. The popular *cross needle* power meter places two meter movements and two indicators in one meter case. One indicates the forward power, and other other the reverse. The intersection of the indicator lines define a point which reveals the SWR, which is indicated on a third scale. I have always found these meters a little difficult to read, and the author indicated that some models do not have high accuracy. Their strong point, in my opinion, is that they can be left in line while transmitting at full power. For purposes such as adjusting an antenna tuner for minimum SWR, these sort of meters are probably very appropriate. I just wouldn't take the SWR value too seriously, especially as it rises. The author also considered an inexpensive noise bridges and one antenna analyzer, the Autek RF-1. His conclusion is: *My results indicate that SWR bridges often do not measure capacitive loads or high SWR values accurately. I found that a noise bridge is more accurate, but that it is also much more difficult to use. The most accurate and easiest instrument to use is a modem, microprocessor-based "RF analyzer"*. When is uses the term *SWR bridges*, he is referring to the power meters that also measure SWR.]

[Epilogue 2 - The ARRL Antenna Compendium #5 has an article entitled: *Baluns in the Real (and Complex) World*, by Frank Witt, AI1H, on page 171. As part of doing some balun measurement discrepancies were due to sloppy measurement techniques.

My Excel spreadsheet data and graphs can be downloaded as Analyzers.xls

Some folks have commented that my results appear to show much more accuracy in these meters than they expected. I think that there are at least two reasons for this. First, it's very easy to introduce errors due to poor measurement technique, primarily excessive lead length. I certainly made several *lousy* measurement runs and had to continue to try to improve the quality of my test fixture. Another reason is that I was measuring what I would call a *shielded system* - a handful of resistors connected to a rather short length of transmission line sitting on my test bench The more common application for an antenna analyzer is to measure, *antennas*. Antennas are designed to pick up signals, and this can include RF from nearby sources. This RF, when presented to the antenna analyzer, can upset the reading. The W8JI calibration page discusses this issue in more detail. A common problem is trying to measure the impedance of a 160 meter vertical when you live a few miles away from an AM broadcast station. I have seen this problem also arise on a 15 meter monoband Yagi. The usual solution is to filter the signal before applying it to the analyzer, keeping the undesired signal out of the analyzer. This must be done carefully, however, since the filter itself can act as an impedance transformer. Some specialized devices exist to filter while not altering the impedance. The MFJ-731 is an example.

ARRL Antenna Compendium #7 has an article by Dan Maguire, AC6LA, which largely duplicates the work on this page. Dan and I worked independently, and we happened to exchange emails after I had finished my page, but before the

---

RF Basics

Martin D. Stoehr
PMTS, ISM-RF Strategic Applications
http://pdfserv.maximintegrated.com/en/an/AN5300.pdf

Introduction

Radio frequency (RF) can be a complex subject to navigate, but it does not have to be. If you are just getting started with radios or maybe you cannot find that old reference book about antenna aperture, this guide can help. It is intended to provide a basic understanding of RF technology, as well act as a quick reference for those who “know their stuff” but may be looking to brush up on that one niche term that they never quite understood. This document is also a useful reference for Maxim’s products and data sheets, an index to deeper analysis found in our application notes, and a general reference for all things RF.

History (How Do We Know What We Know?)

“If I have seen a little further, it is by standing on the shoulders of Giants.”
–Isaac Newton

The wireless radio technology that is so ubiquitous today is relatively new. However, there is a long and rich background that leads to our modern knowledge. The very first investigations of what we now call the RF spectrum came from early experiments in optics, electricity, and magnetism. The behavior of light was studied as far back as ancient Greece by Plato, Euclid, Ptolemy, and many others, eventually leading to Newton in the late 17th Century. From ancient triboelectric materials and chemical batteries, various theories of electricity were eventually developed by Coulomb, Volta, and Gauss. Likewise, lode stones from ancient China birthed early theories for magnetism from Kuo and Gilbert, eventually propelling the investigations of Ampere and again, Gauss.

Before the early 19th Century, electricity and magnetism were seen as separate forces. However, in 1820 Ørsted found that electric currents exerted a force on magnets, and in 1831 Faraday determined that changes in a magnetic field could induce electrical currents. In 1839, further experiments in electricity led Faraday to show that voltaic electricity (chemical battery), static electricity (triboelectric charge), and magnetically induced currents were all manifestations of the same phenomenon. In 1864, Maxwell combined these discoveries in his paper, "A Dynamical Theory of the Electromagnetic Field,"[2] ushering in our modern understanding of the subject:

Gauss's Law relates electric charge to its electric field: $\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}$. The divergence of the electric field is related to the charge density.

Gauss's Law for Magnetism states that magnetic monopoles do not exist: $\nabla \cdot \mathbf{B} = 0$. The divergence of the magnetic field is zero or there is no net magnetic flux entering or leaving a volume.

Faraday's Law of Induction and the Maxwell-Faraday equation state that a changing magnetic field induces an electric field: $\nabla \times \mathbf{E} = -\frac{\partial B}{\partial t}$. The curl of the electric field is related to the change in magnetic flux density.

Ampere's Circuital Law, modified by Maxwell to include displacement current, relates the magnetic field to a current in

This gives me the convenience of the analyzer with a hopefully higher accuracy. I have built several RF networks using this technique, and the results are very accurate.

Conclusion

All of the analyzer plotted results followed the shape of the equivalent computed curve. They are clearly trying to do the job, and in most cases, come darn close. Most of the errors in my data appear in regions of large value swings. This would tend to make the test very sensitive to even slight errors on my part. I tend to consider the regions of the data where the curve is relatively flat over several megahertz. This will reduce the sensitivity to test fixture errors. In most cases, the errors seem to be no more than 5 Ohms. Errors also seem to increase as frequency increases, which could be nothing more than additional testing error.

If I miscalculated the true frequency where the cable was 1/2 wavelength, that would also slightly skew the two curves, creating additional error in the regions of large transitions.

If you use excessive test lead length, you will certainly introduce as much error as any error already in the analyzer. Perhaps the biggest lesson I learned in testing for this page was to keep lead length as short as possible. The next lesson was to use new batteries, or better yet, a beefy power supply with plenty of voltage and current.

I do hope that the next generation of analyzers will accurately resolve the sign of the reactance, and have increased accuracy around zero Ohms. For my recent applications, that would be very helpful. Frankly, I expected the errors to be much greater than I measured. I suspect that many of my unexplained

All of the analyzers are sensitive to low battery voltage. For consistency sake, use an external power supply, or fresh batteries. The MFJ-269 and CIA-HF units have a high current draw, apparently due to the power consumed by the low distortion oscillator. Batteries seem to last substantially longer in the RF-1.

I measured the current drain on the MFJ-269 and the CIA-HF at approximately 225 mA. When the MFJ-269 is placed in UHF mode, the current draw does increase by another 150 mA. If you are going to power one of these analyzers from a *wall wart*, be sure that you have at least a 500 mA supply. Make sure that the external power supply has very low ripple and noise.

Improving Measurement Accuracy

The only technique that I have developed that improves accuracy is in the area of component measurement. I use a more accurate RF bridge to measure a given capacitor and inductor with the highest accuracy that I have available to me. This is a slow process due to the operation of the bridge, and the need for additional compensation computations.

Once I have these measurements, I measure the same components with an analyzer, and determine a simple scaling factor between the two units. I do this for a single amateur band. I then work exclusively with the analyzer, measuring inductors and capacitors with the use of the scaling factors. I move the test leads from the analyzer to the bridge before I make its measurements so that any lead inductance is folded into the scaling factor.

---

a wire: $\acute{\nabla} \times B = \mu_0\left(\acute{J} + \varepsilon_0 \frac{\partial E}{\partial t}\right)$. The curl of the magnetic flux density is related to the current density and the change in the electric field.

These four concepts formed the basis of electrodynamics or modern-day electromagnetic (EM) theory, and are referred to as Maxwell’s Equations. Maxwell had unified the theories of electricity, magnetism, and optics and started the first ventures into electromagnetic-based communication.

The late 19th and early 20th Centuries saw the birth of the electrical and electromagnetic era. Continuing from Maxwell’s work, Hertz, Tesla, and Marconi contributed to EM theory and early forms of practical instruments of communication. In 1887, Hertz showed that an EM wave could travel distances with a basic spark-gap transmitter and spark-gap receivers, and later linked the velocity of those waves to the speed of light. In 1891, Tesla demonstrated wireless power transmission, demonstrated wireless telegraphy in 1893, and filed for the first U.S. patent for a radio in 1897. Likewise, in 1901 and 1902 Marconi began demonstrating trans-Atlantic communication with the first example reaching from England to Newfoundland (nearly 3500km), using a kite-flown antenna.

Just as the theories behind EM communication progressed quickly, radios have also developed at an astounding rate. Wireless telegraphs first appeared in the early 1900s, and developed into the AM radio broadcasts of the 1920s. FM broadcast radio picked up with commercial backing in the 1940s. Satellite communication was quickly adopted after the launch of Sputnik in 1957. With the launch of the Telstar and others in the early ’60s, the use of satellites for relay communication progressed quickly. Through the late ’60 and ’70s, satellites took on larger loads of long-distance communication until submarine cables rebounded with the use of fiber optics in the ’80s. Modern satellites still bear a great burden of media transmission, especially after the emergence of direct broadcast satellite services for television and other broadband media. Ground-based RF communication also progressed from simple numeric pagers in the 1980s, to cellular phones, and eventually the establishment of the ISM bands in 1985 led to our now ubiquitous use of Wi-Fi. These names and technologies mentioned above should be very familiar—these are the giants upon which we stand today.

What Is RF?

RF signals are a form of electromagnetic wave, such as visible light, which make up a portion of the electromagnetic (EM) spectrum. The EM spectrum encompasses all forms of light, which ranges from audible frequencies such as the ubiquitous 60Hz, through the standard radio bands which include AM Radio, FM Radio, TV channels, and other RF bands. The spectrum continues through infrared, visible, and ultra-violet light, to higher forms of EM energy like X-rays, Gama-rays, and cosmic rays.

What we refer to as the Radio or RF spectrum is between the low-frequency waves that we could hear if the EM waves were turned into air pressure waves (20Hz to 20kHz) and the high-frequency EM waves that produce infrared and visible light (1mm to 750nm for IR and 750nm to 390nm for visible (or about 400THz to 770THz)).

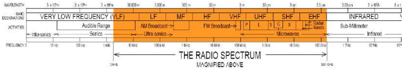
WAVELENGTH
BAND DESIGNATIONS
ACTIVITIES
FREQUENCY
30,000 m
3,000 m
300 m
30 m
3 m
30 cm
3 cm
0.3 cm
0.03 cm
VERY LOW FREQUENCY (VLF)
LF
MF
HF
VHF
UHF
SHF
EHF
INFRARED
Audible Range
AM Broadcast
FM Broadcast
P
L
S
C
X
Radar Bands
Sub-Millimeter
Infra-sonics
Sonics
Ultra-sonics
Microwaves
Infrared
0
10 Hz
100 Hz
1 kHz
10 kHz
100 kHz
1 MHz
10 MHz
100 MHz
1 GHz
10 GHz
100 GHz
1 THz
THE RADIO SPECTRUM
MAGNIFIED ABOVE
3 kHz
300 GHz

Figure 1. Radio frequency spectrum[5]

CIA-HF will not even report capacitance or inductance if the phase angle is -90 or 90. You must use special test leads. The other analyzers do not specify any special test leads.

Pure Capacitive Reactance Test at 3.8 MHz (pf)			
pf	MFJ-269	CIA-HF	RF-1
143	148	157	147
215	220	232	220
430	449	472	438
860	921	972	918
1290	1396	1500	1432

All measurements were made at 3.8 MHz.

The CIA-HF is *picky* about making capacitance and inductance measurements. The impedance angle must be within a certain range that appears to be approximately 70 degrees. I had to increase the resistance of the test leads in order to make the measurements at 143 and 215 pf.

Because of my limited supply of the 430 pf capacitors, and the need to rearrange them in several series/parallel combinations, I had to use about 4 inches of wire to make the connections. I have no doubt that this influenced the readings. The added inductance should lower the capacitive reactance, which increases the pf value. Since the errors all appear to be on the high side, this extra wire may be the issue.

In any case, the accuracy is within 10 percent, and perhaps much better if the lead length influenced the readings. The lead length issue was worse on the CIA-HF because it does not have a ground lug near the UHF connector.

The Effect of Battery Voltage and Current may not know the difference between capacitive and inductive reactance. It knows what to convert to (C or L) based upon the mode you have selected.

Measuring components is also different than general antenna/transmission line impedance measurement because the resistive component is zero, or very nearly zero. All that is being measured is a pure reactance. This is the opposite test to what was performed in the previous section, where pure resistances were measured. It is also the case that typical reactance values may be several hundred Ohms, as opposed to antenna measurements (especially vertical antennas), where resistance and reactance are usually each under 50 Ohms (near resonance).

I again visited the junk box, and found several 430 pf silver-mica capacitors that were labeled as being 1% tolerance. I connected these three parts into several series and parallel combinations in order to create a set of expected values. It is my belief and understanding that these analyzers measure reactance, without directly knowing the type of reactance, and then convert that into units of capacitance of inductance depending upon frequency and the readout (capacitance or inductance) selected by the user. If the reactance accuracy is not a function of the reactance sign, then the accuracy in measuring inductors and capacitors will be the same. In any case, I do not have any *precision inductors*, so I can't make absolute measurements.

The CIA-HF manual suggests that you measure components using test leads that include a 50 Ohm series resistor. In fact, they suggest that the accuracy of the component measurement will be greatest when the phase angle of the impedance is 45 degrees. In the absence of any resistance, the phase angle will be (nearly) -90 or +90 degrees, which is not 45 degrees. The

---

This RF spectrum (shown in Figure 1) is further divided into conventional bands, which are typically classified by their frequency range and broken across decades. For example, the 300MHz to 3GHz range is called the UHF band (designated by the International Telecommunication Union (ITU)). In the UHF, SHF, and EHF bands, organization such as the IEEE and NATO tend to break the bands up into smaller categories.

Table 1. RF Spectrum Bands:

Name	**ELF**	**SLF**	**ULF**	**VLF**	**LF**	**MF**
f (Hz)	$3x10^0$ – $3x10^1$	$3x10^1$ – $3x10^2$	$3x10^2$ – $3x10^3$	$3x10^3$ – $3x10^4$	$3x10^4$ – $3x10^5$	$3x10^5$ – $3x10^6$
λ (m)	10^8 – 10^7	10^7 – 10^6	10^6 – 10^5	10^5 – 10^4	10^4 – 10^3	10^3 - 10^2
Uses	NA	AC power	NA (audible)	Navigation	Maritime	AM radio

Name	**HF**	**VHF**	**UHF**	**SHF**	**EHF**	**Infrared**
f (Hz)	$3x10^6$ – $3x10^7$	$3x10^7$ – $3x10^8$	$3x10^8$ – $3x10^9$	$3x10^9$ – $3x10^{10}$	$3x10^{10}$ – $3x10^{11}$	$3x10^{11}$ – $4x10^{14}$
λ (m)	10^2 – 10^1	10^1 – 10^0	10^0 – 10^{-1}	10^{-1} – 10^{-2}	10^{-2} – 10^{-3}	10^{-3} – $7.5x10^{-7}$
IEEE	HF	VHF	UHF | L | S	S | C | X | Ku | K | Ka	Ka | V | W | mm	
NATO			A – E	F – K	K – M	
Uses	Shortwave, CB	TV, FM radio	ISM, TV, Wi-Fi®	Microwave	Radar	“Light”

In the United States, the Federal Communications Commission (FCC) is the governing body that manages the RF spectrum allocation and permissible uses. The role of the FCC and its foreign equivalents is necessary to provide central organization of this limited resource and to establish a framework that allows for compatible operation of the many and varied radio frequency systems. Without these regulating bodies, anyone would be able to broadcast without regard to frequency, power, bandwidth, or duty cycle—overpowering competitive and noncompetitive uses alike. This could result in monopolizing the airwaves and would possibly interfere with essential forms of command, control, and communication. For more information on how ISM radios are governed by FCC and ETSI regulations, refer to application note 1772, “Where to Go for Regulations Concerning Short-Range Devices (SRD)” and application note 3587, “FCC and ETSI Requirements for Short-Range UHF ASK-Modulated Transmitters.”

RF Glossary

Amplitude and Power

V – Voltage: In RF systems, the voltage of a signal is typically referenced to a 50Ω load.

P – Power: In an RF system, power is typically referenced to a 50Ω load.

$P = VI = \frac{V^2}{R}$ or in a 50Ω system $P = \frac{V^2}{50}$

dB – Decibels: This is a unitless ratio measure (similar to %) typically used in RF systems when discussing power. The ratio “dBm” is more common in RF applications where the “m” refers to using 1mW as the referenced point. The difference between a 1W reference point (dBW) and 1mW reference point (dBm) is 30dB, that is: dBm = dBW + 30dB. When referring to voltage levels, dB is used to represent a ratio such as an output amplitude to an input amplitude.

$$L(dB) = 10 \cdot log_{10}\left(\frac{V_{Out}}{V_{In}}\right)$$

When used in RF applications, the dB is usually a power ratio based on a voltage gain.

$$G(dB) = 10 \cdot log_{10}\left(\frac{50 \cdot {V_{Out}}^2}{50 \cdot {V_{In}}^2}\right) = 20 \cdot log_{10}\left(\frac{V_{Out}}{V_{In}}\right)$$

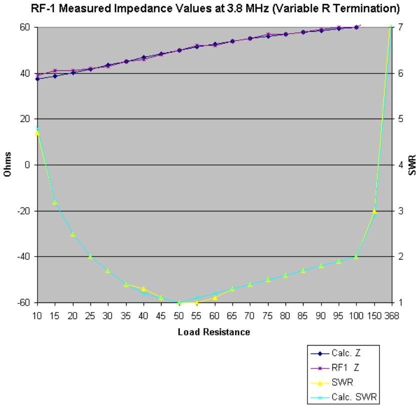

RF-1 Response at 3.8 MHz with a Complex Input

Very good for a tiny little box.

Component Measurements

Another common use of these analyzers is to measure the values of inductors and capacitors. Unlike the typical RLC (resistance, inductance, capacitance) meter, which measures at a fixed frequency, such as 1 MHz, these analyzers will make their measurements at the intended frequency of operation. It is usually the case that the analyzer measures nothing more than reactance, and then converts that into capacitance or inductance using the standard formulas. In fact, the analyzer

Very accurate tracking. Since the frequency is being held at 3.8 MHz, the deviation at higher resistances is not due to frequency. The gap does appear to be real, since the values are rather stable in this region.

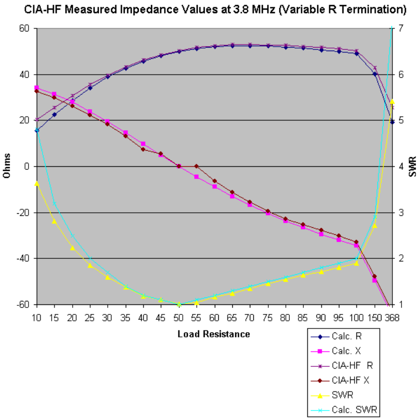

CIA-HF Response at 3.8 MHz with a Complex Input

Something a little strange happened around zero reactance again. Otherwise, very close tracking.

Table 2. Power Levels in Different Units

W	dBW $10 \cdot \log_{10}(P)$	dBm $10 \cdot \log_{10}(1000 \cdot P)$	V √50P (w/ 50Ω)
1.000	0	+30	7.071
0.032	-15	+15	1.257
0.020	-16.990	+13.010	1.000
0.010	-20	+10	0.707
0.003	-25	+5	0.397
0.001	-30	0	0.224
316.2µW	-35	-5	0.126
100µW	-40	-10	0.071
0.1nW	-100	-70	70.71µV
0.1pW	-130	-100	2.236µV
10fW	-140	-110	0.707µV
1fW	-150	-120	0.224µV
4.142E-21 (kT at 300K)	-203.8	-173.8	0.455nV

In addition to dBW and dBm, occasional use of other forms of decibels may appear. In all instances, these additional letters indicate what the base unit of reference may be: dBc is carrier referred, dBi is the gain over an isotropic antenna, and dBd is the gain over a dipole antenna.

Field Strength

V/m – Volts per meter: This is a typical measure used for electrical field strength. It is often more common to see values with higher resolution units such as mV/m or µV/m. These measurements are subject to reference antenna gains at the TX and RX portions of the system, the fields are measured at 3m distance (FCC specified7), and are dependent upon the operating frequency (refer to application note 3815, “Radiated Power and Field Strength from UHF ISM Transmitters” for more information).

FCC field strength: for the ISM bands, the FCC defines a maximum field strength based on a linear extrapolation from the defined band end points of 3.750mV/m at 260MHz to 12.5mV/m at 470MHz7, Section 15.231. To calculate this maximum value for frequencies between 260MHz and 470MHz, use the equation:

$$E = \frac{8.75}{210} f - 7.08\overline{3}$$

Where E is the field strength (in mV/m), and f is the frequency of operation (in MHz). For any frequencies above 470MHz and under 900MHz, the FCC caps the Field Strength at 12.5mV/m. In the 902MHz to 928MHz band, the field strength limit is 500mV/m. For unit translation of E(mV/m) to E(dBuV/m) the equation is: $E = 20 \cdot log_{10}(1000 \cdot E)$

EIRP – Effective isotropic radiated power: This is a term that merges the power generated by a transmitter with the efficiency of an antenna into one term $(EIRP = P_T G_T)$. Typically these two items can be broken into the TX power (PT) and the gain of a transmitting antenna, GT(Θ,Φ). However, the antenna gain can be highly dependent on orientation, being a function of Θ (planar angle) and Φ (elevation angle) as noted. To remove these dependencies, the antenna gain can be simplified to a gain relative to an isotropic radiator, which by definition is uniform for any orientation.

To translate from Field Strength to the EIRP, use the FCC-specified equation:

$$EIRP = 10 \cdot log_{10}(300 E^2)$$

Where EIRP is the effective isotropic radiated power (dBm) and E is the field strength (V/m). This equation assumes the field is calculated at a distance of 3m from the radiated power source.

150	40.22	-49.56	2.9	34	-51	3	42.8	-47.7	2.72	65	3
368	19.16	-64.72	7.2	18	-64	8	25.5	-63.1	5.41	69	7.6

As before, the sign of the reactance was taken from the computed data. For each analyzer I created a graph comparing it to the computed values, for impedance and SWR. The Autek graph, as before, is compared against aggregate impedance, since it does not resolve Z into R and X.

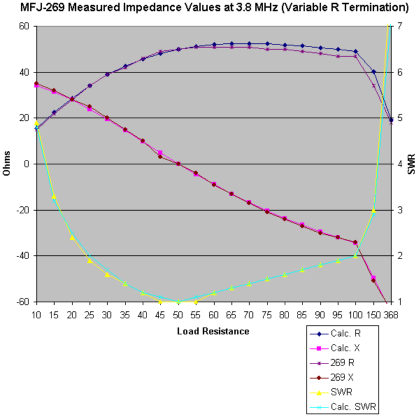

MFJ-269 Response at 3.8 MHz with a Complex Input

25	34.21	23.89	2	34	25	1.9	35.6	22.2	1.86	42	2
30	38.87	19.35	1.7	39	20	1.6	39.5	18.4	1.6	43	1.7
35	42.74	14.55	1.4	42	15	1.4	43.1	13.2	1.38	45	1.4
40	45.83	9.64	1.2	46	10	1.2	46.3	7.4	1.19	46	1.3
45	48.23	4.76	1.1	49	3	1	48.4	5.4	1.11	48	1.1
50	50.00	0.00	1.0	50	0	1	50.2	0	1.01	50	1
55	51.23	-4.57	1.1	51	-4	1	51.8	0	1.05	52	1
60	52	-8.91	1.2	51	-9	1.2	52.5	-6.4	1.17	52	1.1
65	52.38	-13	1.3	51	-13	1.3	53	-11.3	1.25	54	1.3
70	52.45	-16.81	1.4	51	-15	1.4	53.1	-15.4	1.35	55	1.4
75	52.27	-20.34	1.5	50	-21	1.5	52.6	-19.5	1.46	57	1.5
80	51.89	-23.61	1.6	50	-24	1.6	52.6	-22.9	1.55	57	1.6
85	51.35	-26.63	1.7	49	-27	1.7	52.2	-25.3	1.64	58	1.7
90	50.69	-29.41	1.8	48	-30	1.8	51.9	-27.6	1.71	59	1.8
95	49.94	-31.96	1.9	47	-32	1.9	51.2	-30.3	1.81	60	1.9
100	49.12	-34.3	2	47	-34	2	50.4	-32.9	1.9	60	2

Table 3. Field Strength

f(MHz)	FCC Field Strength (mV/m)	Field Strength (dBµV/m)	EIRP (dBm)
260	3.750	71.5	-23.7
300	5.417	74.7	-20.6
315	6.042	75.6	-19.6
330	6.667	76.5	-18.8
434	11.000	80.8	-14.4
435	11.042	80.9	-14.4
470	12.500	81.9	-13.3
[868]	12.500	81.9	-13.3
902	500	114	+18.8
915	500	114	+18.8
928	500	114	+18.8

Per FCC Part 15.231, average field strength limits G_T = 0dB, and d = 3m.

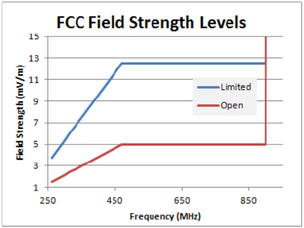

Per FCC Part 15.231, average field strength limits are restricted based on low-rate periodic operation (limited) or higher-rate periodic and "prohibited" uses (open).

The CIA-HF and RF-1 did a good job of measuring pure resistance, especially at values under 150 Ohms. The MFJ-269 always showed nonzero reactance, except at 50 Ohm of resistance. Practically, most antenna work does not require measuring resistance above 150 Ohms, unless open wire feeders are used.

I checked some of the resistances at other frequencies, up to 30 MHz. All of the analyzers were insensitive to frequency, and produced values nearly identical to their values at 3.8 MHz.

I then used the Lowband software to compute the complex impedance at the input of the cable when the load was set to various resistances. I connected a small noninductive trimmer pot directly to a coax connector, and set it to the different resistance values with the aid of my DMM. For each setting, I recorded the values displayed by each analyzer. Here is the data table. All values are measured in units of Ohms, except SWR, which is dimensionless.

Complex Impedance Test at 3.8 MHz (Ohms)											
Set	Lowband Computed			MFJ-269			CIA-HF			RF-1	
R (Load)	R (Input)	X (Input)	SWR	R	X	SWR	R	X	SWR	Z	SWR
10	15.67	34.19	4.8	15	35	4.9	20.3	32.7	3.63	39	4.7
15	22.63	31.5	3.2	22	32	3.3	25.6	29.9	2.81	41	3.2
20	28.75	28	2.5	28	28	2.4	30.8	26.2	2.24	41	2.5

capacitors. We can continue to use our same piece of coax, but vary the pure resistance at the load, creating a range of complex impedance values at the input (analyzer) side. The same software previously used can be reused to compute the expected impedance values. We won't have arbitrary control over our complex impedance values, but we can at least generate values that we have some confidence in.

In the upcoming tests I'm going to hold the frequency constant at 3.8 MHz. For my current application, I'm most interested in measurements in the 160, 80, and 40 meter bands. 3.8 MHz is a representative value somewhat in the middle of the range from 1.8 MHz to 7.3 MHz.

Perhaps the simplest test is to use the analyzer as an RF resistance meter. In other words, directly connect a pure resistance to the meter and see what value is reported. Using a standard DMM, I measured a set of resistors and then tested them with the analyzers. Here are the results.

Pure Resistance Test at 3.8 MHz (Ohms)			
Resistance	59	CIA-HF	RF-1
17.2 Ohms	16, j 2	15.7, j 0	17
25.4 Ohms	24, j 2	24.7, j 0	25
33.5 Ohms	32, j 4	33.6, j 0	33
50.8 Ohms	49, j 0	50.7, j 0	51
99.8 Ohms	97, j 13	98.8, j 0	99
151.2 Ohms	147, j 21	148.8, j 0	148
385.2 Ohms	358, j 116	380.0, j 0	375
680.0 Ohms	586, j 281	648.5, j 0	660

I happened to have 50 and 25 termination loads. These are completely enclosed termination resistors with an integrated BNC connector. Their actual measured values are shown in the table. The 17.2 Ohms resistor is their parallel combination, using BNC "Tee" connectors.

RF Field Strength

by Kenneth A. Kuhn

Jan, 6, 2008 (draft –more to come)

http://www.kennethkuhn.com/students/crystal_radios/rf_field_strength.pdf

Introduction

The process of building a crystal radio begins with the transmitter. Without the transmitter there would be no point in building the radio.

AM Broadcast stations in the United States operate at 10 kHz intervals between 540 kHz and 1,700 kHz using double sideband amplitude modulation with a modulation bandwidth of 5 kHz (10 kHz total channel width) and a transmitted power ranging from around 1 kW to 50 kW. The broadcast antenna is generally some form of a one-quarter to over one-half wave vertical tower (the entire tower is the antenna) with numerous onequarter wave ground radials. A quarter-wave antenna at 540 kHz would be around 139 meters high!

A quarter-wave antenna is attractive because it has a low resistive (i.e. no reactive component) impedance of around 35 ohms. For a 50 kW transmitter the applied voltage to the antenna would be around 1,300 volts rms! This impedance is not the resistance of the tower –that needs to be as low as possible for high efficiency. The impedance of the antenna is the result of the fact that power leaves the antenna as a result of radiated electromagnetic fields –that is the purpose of the antenna.

If the electrical length of the antenna is not a quarter wave then there is a reactive component in the antenna impedance. This reactance is in the way of coupling power to the antenna.

Various tuning schemes can eliminate the net reactance. Antenna designers are interested in maximizing the strength of the electromagnetic field emitted from the antenna. They also have to deal with other limitations such as maximum allowed antenna height, electrical characteristics of the soil around the antenna, and other factors. For these reasons the electrical height of the antenna may deviate from the seemingly ideal one-quarter wave. The result is an antenna that optimizes the broadcast range of the transmitter. Tuning networks can compensate for any antenna reactance.

The signal strength some distance from the transmitting tower is commonly measured in volts per meter. There is a corresponding current measured in amperes per meter obtained by dividing volts per meter by the impedance of free space, 377 ohms. There is no power dissipated in free space as the voltage and current are physically orthogonal. The voltage per meter is a cyclic gradient and is only meaningful over distances that are short (roughly 20 electrical degrees) with respect to a wavelength. Short antenna probes can measure this voltage. Amperes per meter is also a cyclic gradient and is a bit more difficult to visualize as it represents a magnetic field that would exist in response to an actual current. There is no actual current in free space which is an electrical insulator. The magnetic field can be measured using small (with respect to a wavelength) loop antenna probes in which a current is induced.

Thus, electromagnetic radiation is a combination of a cyclic voltage field and an orthogonal cyclic magnetic field. Either alone is just induction as opposed to radiation. As fields spread, the strength of an electromagnetic wave follows a one over distance law. Induction fields follow a one over distance cubed law. Thus, induction fields are useful only for very short distances.

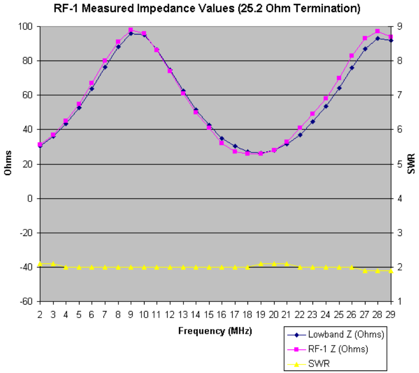

RF-1 Frequency Sweep Data

Since the RF-1 only reports impedance (Z), I took the original Lowband data and converted R and X to Z, using the standard formula (Z = SQRT $(R^2 + X^2)$). This derived data is available on the spreadsheet file which can be downloaded at the end of this page.

Impedance at a Constant Frequency

If we want to hold the analyzer at a constant frequency, but yet test its SWR and impedance accuracy, we simply must present it with a range of impedance inputs. We could certainly wire up combinations of resistors and inductors/capacitors, creating reference standards. This approach is complicated by the difficulty of obtaining accurate reference inductors and

CIA-HF

The following graph displays the recorded data when the frequency was swept from 2 through 29 MHz, and the CIA-HF was connected to the test coax with a 25.2 Ohm resistive load.

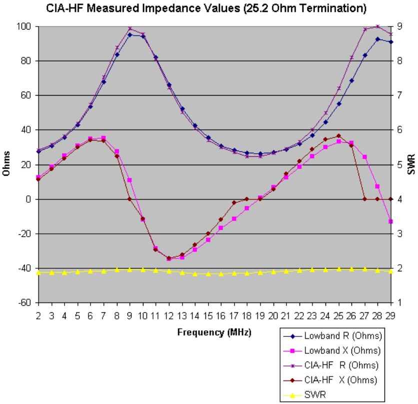

CIA-HF Frequency Sweep Data

The CIA-HF results are remarkably similar to the MFJ-269 results, although something a little strange seems to happen when the reactance is near zero.

RF-1

The following graph displays the recorded data when the frequency was swept from 2 through 29 MHz, and the RF-1 was connected to the test coax with a 25.2 Ohm resistive load.

In order to engineer a crystal radio we need to know the expected electromagnetic field strengths we intend to receive. Those strengths must be above some minimum or we will hear nothing in our headphones. Data is readily available that gives us a reasonable expectation of the amplitude of a radio wave in terms of volts per meter at a given distance from the broadcast antenna. We must understand not to interpret such data too literally as there are always variables related to terrain and various structures that will affect the data –mostly negatively but sometimes positively.

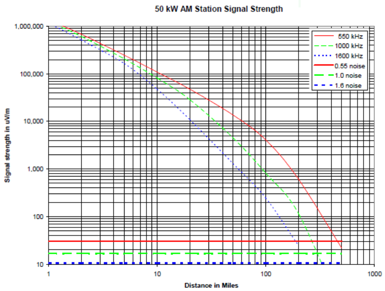

Figure 1: Signal strength with distance adapted from Reference Data for Radio Engineers, Howard W. Sams, Inc. sixth edition, pages 30-3 to 30-5

Figure 1 shows typical signal strengths for a 50 kW station during daylight hours. The voltages will be about one-third for a 5 kW transmitter. The increased drop-off with distance is due to the curvature of the earth and varies at night and with weather conditions so a significantly greater range is possible at times. The typical atmospheric noise level for a 10 kHz bandwidth is also shown although this can vary significantly with location and season. Excellent reception is when the signal strength is 100 times the noise level. Poor, but usable, reception is when the signal strength is around 5 times the noise level. The noise floor is only shown for reference. Crystal radios are typically so insensitive that the minimum usable signal is around 1000 times the noise floor.

MFJ-269

The following graph displays the recorded data when the frequency was swept from 2 through 29 MHz, and the MFJ-269 was connected to the test coax with a 25.2 Ohm resistive load.

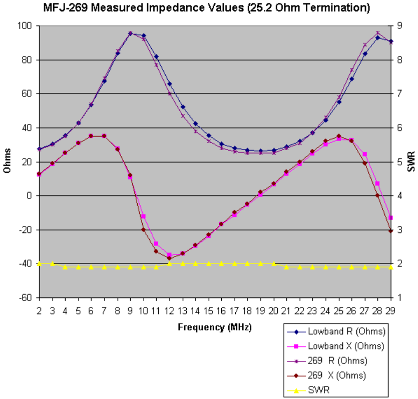

MFJ-269 Frequency Sweep Data

This graph captures the impedance results, as well as the SWR. This is true for all three analyzers. The SWR trace is shown in yellow, and its values are labeled on the right. Ideally, the SWR would be a nearly horizontal line, at the 2.0 value. The MFJ-269 only displays whole number values.

6	54	35	1.9	55	34	1.91	67	2.0
7	69	35	1.9	70.5	33.6	1.92	80	2.0
8	85	27	1.9	87.7	24.6	1.95	91	2.0
9	96	12	1.9	98.6	0	1.96	98	2.0
10	92	-20	1.9	95.7	-11.5	1.95	96	2.0
11	77	-33	1.9	81.4	-29.5	1.94	86	2.0
12	60	-37	2.0	64.6	-34.4	1.91	74	2.0
13	47	-34	2.0	50.5	-32.3	1.88	61	2.0
14	38	-29	2.0	40.9	-26.7	1.84	50	2.0
15	32	-23	2.0	34.1	-20.1	1.84	41	2.0
16	28	-17	2.0	29.8	-12.1	1.83	32	2.0
17	26	-10	2.0	27	-2.3	1.85	27	2.0
18	25	-5	2.0	24.5	0	1.86	26	2.0
19	25	2	2.0	24.5	0	1.87	26	2.1
20	25	7	2.0	26.8	5.7	1.89	28	2.1
21	28	14	1.9	29.1	14.4	1.92	33	2.1
22	31	20	1.9	33.3	21.7	1.93	41	2.0
23	37	26	1.9	40	28.8	1.95	49	2.0
24	46	32	1.9	50	34.4	1.96	58	2.0
25	58	35	1.9	64.2	36.6	1.97	70	2.0
26	74	32	1.9	82.2	30.8	1.98	83	2.0
27	89	19	1.9	98.5	0	1.97	93	1.9
28	96	0	1.9	100	0	1.94	97	1.9
29	90	-21	1.9	95.5	0	1.91	94	1.9

I supplied the reactance sign in the 269 and CIA-HF columns, since the devices do not indicate the sign. I obtained the sign by looking at the computed expected values. The Autek meter only reports the total impedance, not the values of the separate components. The following three subsection graphs compare each analyzer with the computed results from the Lowband software.

Field Strength Calculations

Radio-Time Traveler

PART I: Ground Conductivity

http://radio-timetraveller.blogspot.com/2011/06/field-strength-calculations-ground.html
20 June, 2011

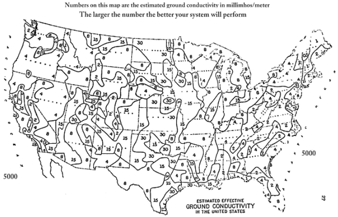
Numbers on this map are the estimated ground conductivity in millimhos/meter
The larger the number the better your system will perform
5000
5000
ESTIMATED EFFECTIVE GROUND CONDUCTIVITY IN THE UNITED STATES

Ground (also called soil) conductivity plays a huge role in how far the mediumwave signal travels during the daytime. Lesser known to many, station frequency is also a factor, and maybe more of one than you would think. Though one can argue successfully that frequency is not a factor in the formula for calculating received signal strength, it indeed becomes relevant as you will see. In this series, let's explore ground conductivity and station frequency and see how they relate to "how far you can hear" on the mediumwave band during the daytime. We will end with a method to calculate approximate field strength for stations of interest.

Many years ago, ground conductivity measurements were compiled into a map titled "Estimated Effective Ground Conductivity in the United States" (Figure M3) by the FCC.

This map is used for the allocations planning for placement of MW stations in the United States. The map presents optimistic ground conductivities and is used when measured conductivity is not available. This information has been used and accepted since it was compiled in 1954.

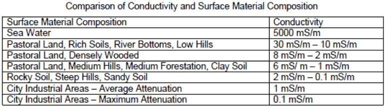

Comparison of Conductivity and Surface Material Composition

Surface Material Composition	Conductivity
Sea Water	5000 mS/m
Pastoral Land, Rich Soils, River Bottoms, Low Hills	30 mS/m – 10 mS/m
Pastoral Land, Densely Wooded	8 mS/m – 2 mS/m
Pastoral Land, Medium Hills, Medium Forestation, Clay Soil	6 mS/ m – 1 mS/m
Rocky Soil, Steep Hills, Sandy Soil	2 mS/m – 0.1 mS/m
City Industrial Areas – Average Attenuation	1 mS/m
City Industrial Areas – Maximum Attenuation	0.1 mS/m

Soil conductance is measured in siemens per meter but most generally shown in millisiemens per meter. This is the mS/m designation you see in the accompanying chart. The siemens (symbolized S) is the Standard International (SI) unit of electrical conductance. The old term for this unit is the mho (ohm spelled backwards). Conductance (mho) is of course the opposite of resistance (ohm). As you can see, salt sea water provides the best conductance by far (5000 mS/m). The higher the value the better. Average soil runs 6-8 mS/m. Find your location on the map and see what your local conductance value is. Notice also that a distant station's receive path may transition across more than one zone.

Indeed, for all pairs of values in the table, you end up with the same SWR. Here are those equations.

$$|\rho| = \sqrt{\frac{\left((R_a - R_0)^2 + X_a^2\right)}{\left((R_a + R_0)^2 + X_a^2\right)}} \qquad EQ.1$$

$$SWR = \frac{(1 + |\rho|)}{(1 - |\rho|)} \qquad EQ.2$$

$$|\rho| = \frac{(SWR - 1)}{(SWR + 1)} \qquad EQ.3$$

Equations Relating Load Impedance to SWR

The third equation relates SWR back to the reflection coefficient (ρ). For an SWR of 2, the reflection coefficient is 0.333333... If you grind through EQ.1 with the values from the previous table, you will always end up with an answer very close to 0.333333... By the way, R_a and X_a are the load resistance and reactance. R_0 is the line impedance (50 Ohms). This equation only holds for low loss lines at low frequencies.

Getting back on track, here is the impedance and SWR data recorded from the three analyzers.

Measured Impedance Values as a Function of Frequency								
	MFJ-269			CIA-HF			RF-1	
Freq. (MHz)	R (Ohms)	X (Ohms)	SWR	R (Ohms)	X (Ohms)	SWR	Z (Ohms)	SWR
2	27	13		28.3	11.1	1.88	31	2.1
3	30	19	2.0	31.7	17.4	1.88	37	2.1
4	35	25	1.9	36.6	23.6	1.87	45	2.0
5	43	31	1.9	43.8	29.8	1.89	55	2.0

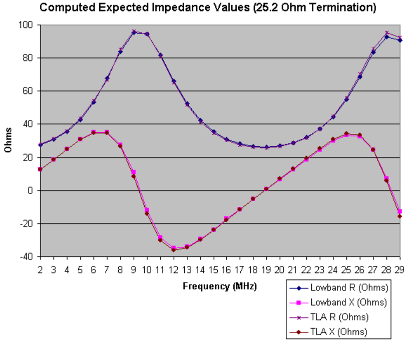

Lowband and TLA Computed Input Impedance

As the table and graph shows, the two different methods of calculating impedance diverge, especially at the extremes, but, in general they closely agree. For the purposes of future graphs related to the 2.0 SWR example, I will only compare against the Lowband data.

Here's a little diversion. You might ask yourself: *how can all of these different pairs of R and X values always result in an SWR of 2.0?* I certainly was interested in that question. The answer can be found on pages 24-6 through 24-9 of the ARRL Antenna Book, 18th Edition. Two formulae are needed. The first (EQ.1) takes the load resistance and reactance, and transmission line resistance, and computes the *magnitude of the reflection coefficient.* The second formula (EQ.2) converts the magnitude of the reflection coefficient into the SWR.

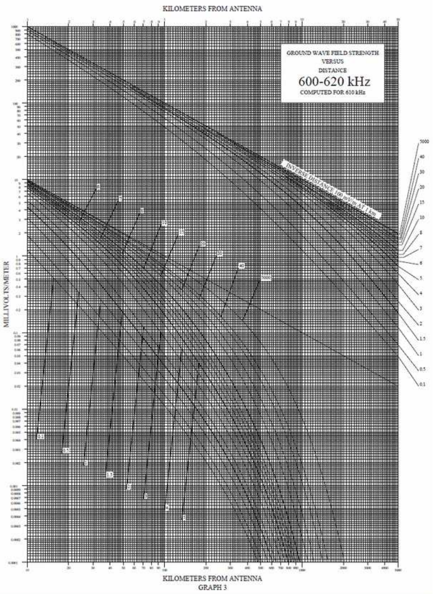

FCC Ground Wave Field Strength Versus Distance Graphs

ftp site for finding all field strength vs distance graphs from the fcc:
ftp://ftp.fcc.gov/pub/Bureaus/MB/Databases/AM_groundwave_graphs/

Equally important, the FCC also produces a series of charts known as the "Ground Wave Field Strength Versus Distance" graphs. These 20 graphs in .PDF form are grouped by mediumwave channels in the 540-1700 KHz range, and allow prediction of received signal strength by cross-referencing the distance to the receiving location with the ground conductivity factor between you and the station. These charts cover soil conductivity ranges of 0.1 mS/m to 5000 mS/m. They are still in use today. The only working link I have for them is:

Be sure to get the .PDF versions.

How well a mediumwave transmitter "gets out" is not only dependent on its power, frequency, and the ground conductivity between it and you, but also on the ground condition at its location. The following is a quote from the Standards of Good Engineering Practice Concerning Standard Broadcast Stations (550-1600 kc.), 1939, and is still relevant today:

"The ideal location of a broadcast transmitter is in a low area of marshy or 'crawfishy' soil or area, which is damp the maximum percentage of time and from which a clear view over the entire center of population may be had... The type and condition of the soil or earth immediately around a site is very important. Important, to an equal extent, is the soil or earth between the site and the principal area to be served. Sandy soil is considered the worst type, with glacial deposits and mineral-ore areas next. Alluvial, marshy areas and salt-water bogs have been found to have the least absorption of the signal."

20	26.99	6.62	1.9	26.66	6.88	1.93
21	28.83	12.66	1.9	28.57	13.03	1.93
22	32.02	18.72	1.9	31.88	19.22	1.93
23	37.07	24.64	1.9	37.06	25.30	1.93
24	44.50	29.91	1.9	44.77	30.76	1.92
25	55.03	33.35	1.9	55.78	34.34	1.92
26	68.73	32.59	1.9	70.24	33.47	1.92
27	83.40	24.30	1.9	85.75	24.42	1.92
28	92.87	7.21	1.9	95.38	5.85	1.92
29	90.87	-13.01	1.9	92.31	-15.64	1.92

The Lowband software only reports SWR to tenths. The SWR does slightly drop under 2.0, but the drop is next to nothing, since the line is short, and it is a low-loss cable. As expected, the resistance cycles close to 25 and 100 Ohms, and the reactance varies between +36 (approximately) and -36 Ohms, switching signs when the resistance value reverses direction. Since I sampled at the arbitrary unit of 1 MHz, the actual value extremes are not exactly captured.

I took the resistance and reactance points and graphed them in Excel. Here is that graph.

two programs did not agree exactly, down to the last digit. I suspect that these programs could be made to produce identical results if you understood how to answer the different sets of setup questions with an eye towards how those questions drove the underlying computations, which should be identical, or at least very close.

The expected impedance values, and SWR, are shown in the following table, as a function of frequency.

Computed Expected Impedance Values as a Function of Frequency						
	Lowband Software			TLA Software		
Freq. (MHz)	R (Ohms)	X (Ohms)	SWR	R (Ohms)	X (Ohms)	SWR
2	27.61	12.42		27.0	12.42	1.97
3	30.69	18.71	2.0	31.08	18.71	1.97
4	35.56	24.95	2.0	36.07	24.95	1.96
5	42.87	30.72	2.0	43.54	30.69	1.96
6	53.44	34.93	1.9	54.32	34.80	1.96
7	67.69	35.17	1.9	66.83	34.80	1.95
8	83.79	27.65	1.9	85.09	26.76	1.95
9	95.31	10.89	1.9	96.35	8.47	1.95
10	94.43	-12.14	1.9	94.68	-14.13	1.95
11	81.95	-28.45	1.9	81.41	-30.19	1.95
12	66.02	-34.77	1.9	65.12	-35.18	1.94
13	52.40	-33.93	1.9	51.45	-34.67	1.94
14	42.42	-29.55	1.9	41.54	-29.98	1.94
15	35.55	-23.81	1.9	34.79	-24.03	1.94
16	30.64	-17.01	1.9	30.36	-17.75	1.94
17	28.19	-11.51	1.9	27.63	-11.49	1.93
18	26.69	-5.42	1.9	26.21	-5.32	1.93
19	26.31	0.60	1.9	25.91	0.78	1.93

All well and good. Our transmitter is well-located, emitting a good signal traveling over perhaps many kilometers or miles to our receiving location. We either hear it or we don't depending on the natural attenuation decay between us and the transmitter. But just how do we predict the outcome? How do we (abstractly) measure a mediumwave signal's strength at the receiving end? We will use these tools and others to find out.

PART II: Measurements
http://radio-timetraveller.blogspot.com/2011/06/field-strength-calculations.html
23 June, 2011

Continuing on with ground conductivity, the higher the conductivity the farther away the station's signal will be copyable. Study the accompanying graph showing the range of a 600 KHz signal at the 50KW power level at various ground conductivities. 50 kilowatts over good ground conductivity of 15 mS/m will produce a very copyable signal at a distance of 375 kilometers (233 miles). At an average to poor ground conductivity of 4 mS/m, coverage is reduced to under 200 kilometers. Excellent ground conductivity of 30 mS/m takes the same signal out to more than 500 kilometers! A purely over-seawater path (5000 mS/m) should result in a fair signal out to about 750 kilometers or some 465 miles.

Also study the graph of the transmitter with only 1KW output at 600 KHz. It gets out quite well considering its output is a mere one-fiftieth of its bigger brother! Its range is perhaps a third of the 50KW transmitter.

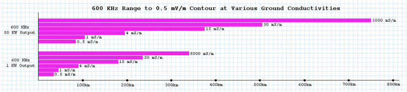

Take another look at the M3 map from the previous part in this series. The country is filled with pockets of different ground conductivities, some large in area, some small. The mid-section of the country has some of the best. A large portion of eastern Colorado and western Kansas are in the 15 mS/m range. Most of Kansas proper is an incredible 30 mS/m. Signals propagating eastward out of Colorado travel a long way. Northern Texas and Oklahoma also have great ground conductivity.

Frequency of operation makes an important difference in how far a signal travels. Propagation distance at the high end of the mediumwave band is less than half the distance of that at the low end for the same received signal strength. A study of the two accompanying graphs comparing the distance coverage of 600 KHz, 1100 KHz, and 1600 KHz signals clearly show this.

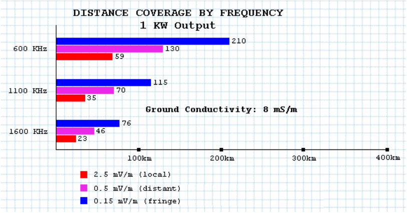

---

parameters. I also terminated the line with a 50 Ohm load and verified that the SWR was 1.0 across the entire HF range. Any errors within the transmission line itself will throw off all of the results.

For the purposes of this page, I am going to assume that my cable is 180 electrical degrees at 18.90 MHz. This value splits the difference between the three meters. If I need to know the electrical length at some other frequency, I will simply scale this ratio. At 3.8 MHz, for example, the electrical length is:

Length(degrees) = (180 degrees / 18.90 MHz) * 3.8 MHz = 36.19 degrees

The lengths will be used to determine the cable impedance transformation at various frequencies.

I connected the terminated cable to each analyzer, and swept the frequency from 2 to 29 MHz, in steps of 1 MHz. I recorded all of the information that each meter provided, placing the data into a Microsoft Excel spreadsheet. I used an external power supply or new battery with each analyzer.

But first, I used the ON4UN Lowband software (*Lowband*) to compute the expected impedance and SWR at the input of the cable. I used the *lossy* cable model (coax1.exe), which requires the specification of cable loss per 100 feet at two frequencies. For this cable, I specified 0.2 dB at 1 MHz, and 0.6 dB at 10 MHz. That data came from a web site that sells that particular cable. I further used the TLA program obtained from my ARRL Antenna Book back cover to compute the same information. Each program takes a different approach to how loss is entered. I did my best to supply the correct loss data to TLA, but this is based upon looking at typical loss curves, and selecting values. As you might imagine, the results of these and again where the SWR stepped up from 1.0 to 1.1. I used the center frequency of the region where the SWR was 1.0. This region was several hundred kilohertz wide. The values I obtained were:

Half Wavelength Frequency of Test Cable	
Analyzer	Frequency (MHz)
MFJ-269	18.750
RF-1	19.045
CIA-HF	19.150

Because of the lack of agreement, I decided to short the load end of the cable and divide the frequency in two in order to treat the line as a 90 degree section. I obtained the following values.

Quarter Wavelength Frequency of Test Cable	
Analyzer	Frequency (MHz) X 2
MFJ-269	18.726
RF-1	19.055
CIA-HF	19.550

It is also possible to compute the half wavelength physical length of the cable using published data. The cable velocity factor was listed as 0.66. The formula is:

f(MHz) = (491.8 / 17.1 feet) X 0.66 VF = 18.98 MHz

The measured frequency where this cable is a half wavelength was approximately 18.90 MHz. The frequency determined by computation was 18.98 MHz, a frequency error of slightly less than 1/2 percent. The published velocity factor closely agreed with the measured factor. I measured another cable with a 10 percent velocity factor error. To be on the safe side, I would reject lines that showed large deviations from published

---

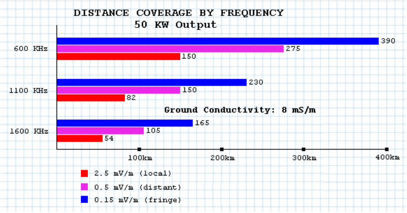

The common unit used in measuring received field strength is volts per meter, or usually, millivolts per meter "mV/m". This is also the FCC requirement. An odd term, millivolts (thousandths of a volt) per meter. We all know what voltage is, but per meter of what, exactly?

Volts per meter expresses the voltage that would be induced in a one meter long wire placed parallel to the lines of flux of the received signal (remember the electrical flux from a mediumwave tower is vertical, the magnetic horizontal). This induced voltage results from the movement of the flux across the wire.

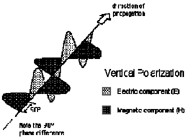

It is important to note that the E (electrical) component of an electromagnetic field is measured in a single dimension. Why? The intensity-versus-distance relation is a straight inverse rule, not the inverse square law commonly used for calculating received power density. In the perfect environment of space, if you double the distance the signal has to travel, the received voltage (in volts/meter or fraction thereof) is halved. Ten times farther away results in 1/10 of the voltage. This is known as field strength. Power density, a different method of measuring received signal strength, follows the inverse square law. That is, the density is related to the inverse of the square of the distance.

In broadcast parlance, millivolts per meter is often referred to in a different context, that being "dBu", or more accurately, dBµV/m. This is decibels above or below 1 millionth of a volt per meter. It is a convenient way to represent field strength, as the decibel is simply a ratio of values. Be sure not to confuse this dBu (lowercase "u") with the Greek "mu" ("µ") on the new model DSP ultralight radios. They are different. These radios actually measure dBµV - voltage across antenna terminals at a certain impedance, not volts per meter.

The FCC offers a conversion calculator to convert from dBu to mV/m and back.

Or, you can figure it yourself by using the following formula:

dBu = 20 * Log(mV/m * 1000)

To reverse the computation, converting dBu back to mV/m:

mV/m = (10 ^ (dBu / 20)) / 1000

But let's put received millivolts per meter into practical application, something a little more understandable. Hatfield & Dawson, Consulting Electrical Engineers out of Seattle, WA have a wealth of interesting mediumwave engineering

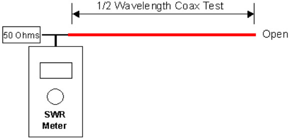

Determining Coax Length via the 1/2 Wavelength Method

At some frequency, this length of coax will be an electrical half wavelength. At that frequency, the open circuit at the far end of the cable will be exactly transferred to the input end. This means that the SWR meter measures nothing except the 50 Ohm resistive load, which has an SWR of 1.0. As the frequency is raised or lowered, the open half wavelength will present an impedance in parallel with the 50 Ohm load, and no matter what value that impedance is, it will result in a change in the SWR. After creating this set up, all you have to do is to sweep the frequency of the meter and look for the SWR dip to 1.0. The only caution is to make sure that your frequency is not so high that you are actually measuring a multiple of a half wavelength.

If you wish to consider the cable to be 1/4 wavelength, then you must short circuit the far end. A 90 degree cable reverses the sense of opens and shorts on its two ends. So, a shorted load appears as an open input, which is the same as the 1/2 wavelength case.

Of course I obtained three different readings from the three different meters. The CIA-HF meter actually reports a single lowest SWR frequency. On the other two meters, I recorded the frequency where the SWR stepped from 1.1 down to 1.0, A quick comment on coaxial cable loss. Cable loss does have the effect of reducing the SWR as the loss increases, either due to a longer cable, or higher frequency. We should, therefore, not demand that the SWR stay at 2.0, especially as the frequency rises. The correct SWR value will be less than 2.0. In my tests I will use calculated models of lossy cable in order to understand with the highest accuracy what our expected results should be.

I decided to duplicate the Caron experiment with my meters. I went to the old coax cabinet, and the first similar length I pulled out was 17.1 feet of RG-213 cable. I happened to have a commercial 25 Ohm termination load that measured at 25.2 Ohms with my 3 1/2 digit digital multimeter (DMM).

Before I started the first SWR measurements, I wanted to determine the electrical length of my sample coax at some frequency. Given that information, a new frequency, the load resistance, and cable attenuation, one can compute the actual impedance which should be observed at the input of the cable. If we are willing to believe those results, then we can generate a set of expected values that the analyzers should report.

The published techniques for determining cable length are usually based upon treating the cable as either a quarter wavelength or half wavelength section at some frequency. Once you have either 90 or 180 degrees at some frequency, you can scale the electrical length to the desired frequency. The following test setup is used.

---

documents for perusal, floating about the web. This document seems to describe it best in terms the layman can understand.

From an engineering report by Hatfield & Dawson, 2002:

"MW radio signal strengths are measured in volts per meter (V/m). The FCC requires that MW radio stations provide a predicted 5 mV/m daytime signal and a 5 mV/m nighttime signal or a Nighttime Interference Free (NIF) signal, whichever is greater, over the city of license."

"MW radio receivers vary greatly in sensitivity and much has been written lately about the poor performance of MW receivers built today. The general practice for broadcaster use for coverage is 2 mV/m for coverage in vehicles, 5 mV/m to 25 mV/m for in home and 25 mV/m in downtown office buildings. On a Walkman-type portable radio, you may need as much as 5 millivolts (5 mV/m) of signal to have static-free reception. The reason for these recommended signal levels is to overcome the effects of interference. Sources of interference include fluorescent lights, computers, TVs, office equipment, overhead power lines, and other appliances operating near a radio that can overload the receiver. In a city core, stations generally need more signal than this because of heavy attenuation inside large steel structures like office buildings. In your home, depending on the location, type of radio, and the utilization of any external antennas, you can have good reception with signals between 1 mV/m and 25 mV/m."

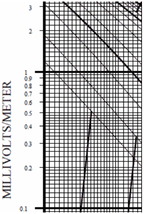

radio-locator.com produces widely used antenna pattern plots, available to interested parties over the web. These depict the expected signal coverage of the mediumwave station. radio-locator divides the signal coverage area into three distinct ranges.

1. Local (red line), the area in which the field strength is 2.5 mV/m or greater, where "....you should be able to receive the radio station on almost any radio with moderately good to very good reception".

2. Distant (purple line), the area between 0.5 and 2.5 mV/m, where "....the signal of the radio station may be weak unless you have a good car radio or a good stereo with a good antenna. You may not be able to receive the station at all on Walkmans or other portable radios".

3. Fringe (blue line), the area between 0.15 and 0.5 mV/m, where "....the station's signal will be very weak. You may be able to receive this station if you have a very good radio with a good antenna, but it's possible that interference from other stations may prevent you from picking up these stations at all." This seems to agree with Hatfield and Dawson's report.

Although the analyzers have evolved from that point, measuring SWR is still one of their most important functions.

As I was going through all of my paper and on-line resources, trying to figure out how to make accurate impedance measurements, I came across an article in ARRL Antenna Compendium #3, by Wilfred Caron, entitled: *The Hybrid Junction Admittance Bridge*. In this article, he uses an evaluation and calibration strategy which is based upon a section of coax cable, and a fixed resistor. He connected a 25 Ohm resistor to the end of a 16 foot length of RG-58 coax. What do we know about that circuit? Ignoring transmission line loss for a moment, we know that the SWR measured at the input of the cable should always be 2.0, at any frequency. In addition, the impedance measured at the cable input will change as a function of frequency, displaying both resistive and reactive values. The resistance will vary between 25 and 100 Ohms, and the reactance will swing both positive (inductive) and negative (capacitive) around zero Ohms. When the impedance is exactly the load impedance, 25 + j 0 Ohms, the line will be an electrical half wavelength, or multiple of it, at the target frequency. All of the impedance values will repeat every half wavelength. Peter Dodd, G3LDO, has used this same approach to evaluate his Autek VA-1 analyzer, and his HP4085A vector impedance meter. Those results are presented on one of his web pages.

So, for the price of a resistor and a short length of coax cable, we can present a range of impedance values to an SWR measuring device that should always read 2.0, at any frequency. In fact as we sweep frequency, the impedance values will change, but always representing a 2.0 SWR. This is conceptually similar to drawing the 2.0 SWR circle on a Smith Chart, and moving around on the circle.

While all of these analyzers can be used as RF signal generators, the CIA-HF is, by far, the most stable of the three. In order to use it as a signal generator, you must set the *width* to 0, so that the unit emits a constant signal. Otherwise, it generates RF periodically, as it takes frequency sweep readings and updates the screen. The only issue is that the analyzer then emits a tone which is in proportion to the SWR. This can be a little annoying, but it is not that loud. The quality of the CIA-HF waveform is also dependent upon battery voltage and current. If you do not provide it with sufficient power, the output waveform begins to resemble a square wave, as opposed to a sine wave.

Autek RF-1

The Autek RF-1 is often noted for its very small size. It's easy to carry around, for example, to the top of an antenna tower. This unit is several years old. Unlike the MFJ and AEA units, which measure both resistance (R) and reactance (X), the RF-1 only measures total impedance (Z). Autek sells an updated model, the VA-1, which does report resistance and reactance. Both units are the least expensive of the alternatives.

My software version number is 9.7.

SWR Measurements

The initial use of the portable antenna analyzer was to measure SWR. SWR meters have been around for a long time, but they tended to be meters which we used in the radio shack, excited with transmitter power, often times hundreds of watts. They were not suited for making adjustments out at the antenna. Early antenna analyzers were nothing more than a low power variable oscillator, frequency counter, SWR sensor/bridge, and a detector meter, all placed in a battery-operated package.

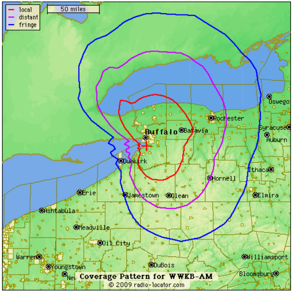

Another important factor in our quest to predict received signal strength is transmitter antenna efficiency. Size does matter. The better the antenna, the better the station gets out. Returning again to Hatfield and Dawson, now discussing mediumwave antennas:

"MW antenna heights are referenced to a wavelength (this only includes the radiating portion of the tower). In MW broadcasting, 5/8 wavelength (or 225°) antennas are more efficient than ¼ wave antennas. A ¼ wavelength (or 90°) antenna is near the lower end of acceptable antenna heights. Antenna heights much below ¼ wavelength are undesirable, as efficiencies decrease dramatically below this height. A 225° antenna provides the maximum coverage and is the theoretical maximum. The efficiency decreases for antennas taller than 225°, which results in reduced coverage. In summary, taller towers are more efficient and shorter towers decrease coverage and are more difficult to work with."

So back to our quest of received signal strength prediction. Few documents are available which demonstrate how to predict "real-world" surface propagation distances at mediumwave frequencies. Many analysts use vanilla formulas for calculating free space path loss using the inverse square law, computing the path loss (in dB) from transmitter to receiver. The path loss, in turn, is used to determine the received field strength value.

Unfortunately, free space path loss assumes a perfect environment, like you would find out in space or over a perfectly conducting earth. Be wary of cute little field strength calculators on the web, their resultant field strength figures are almost an order of magnitude higher than what is seen in real life. Formulas exist which will predict field strength at mediumwave frequencies, but the math is well beyond the average person's ability.

A simpler method exists to ballpark the figure we want. We need but one more piece of information to complete the puzzle. That is the millivolt per meter level at one kilometer from the station transmitter, emanating in your direction. Using that with the FCC's Ground Wave Field Strength Versus Distance graphs and the M3 conductivity map we can estimate the received signal level in mV/m to fairly good accuracy at our location for any station.

The next part of this article will show you how to do that. But let me leave you with this before we go. What information can we derive from the FCC graphs and other station data to maximize our daytime DX distance?

1. The lower frequencies propagate better per unit of distance. Stations near the low end of the broadcast band propagate nearly three times farther than ones at the highest end, for the

Shortly after I purchased my CIA-HF, I learned that the phase angle was considered to be *unreliable* under 8 degrees (-8 to +8 degrees) [on more recent units, this unreliable region has been reduced to 4 degrees). For that reason, I tend to think of the analyzer as not reporting the sign of the reactance, but if you are willing to use the Data Screen, and be careful under 8 degrees, the unit does report indeed the sign.

Here is a screen shot showing the resistance, reactance, and SWR of a dipole antenna fed with a tuner, and adjusted for resonance at 14.200 MHz. The data was transferred over a serial line from the CIA-HF to the computer, and displayed with my viewing software.

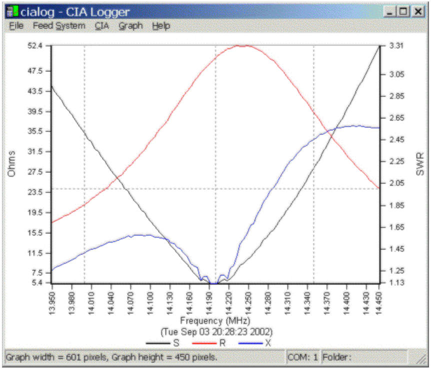
cialog - CIA Logger
File
Feed System
CIA
Graph
Help
Ohms
52.4
47.5
43.5
39.5
35.5
31.5
27.5
23.5
19.5
15.5
11.5
7.5
5.4
SWR
3.31
3.05
2.85
2.65
2.45
2.25
2.05
1.85
1.65
1.45
1.25
1.13
13.950
13.980
14.010
14.040
14.070
14.100
14.130
14.160
14.190
14.220
14.250
14.280
14.310
14.340
14.370
14.400
14.430
14.450
Frequency (MHz)
(Tue Sep 03 20:28:23 2002)
S
R
X
Graph width = 601 pixels, Graph height = 450 pixels.
COM: 1
Folder:

CIA-HF Data Displayed on a Computer

MFJ does make available a calibration procedure for the 259B (I have never asked about the 269). Tom, W8JI, has also placed a calibration procedure on his web site.

AEA CIA-HF

I have one of the early CIA-HF units, serial number 0341. This unit has an LCD screen which can produce actual graphs of measured data, such as SWR, impedance, resistance, and reactance. In addition, the unit has an RS-232 serial interface, which can be used to interface the analyzer with a computer. Several software packages exist. I use one that I wrote myself. Unlike the other analyzers, the CIA-HF takes a few seconds to update the screen. This makes it less useful for use while *tweaking* a control, since it takes a few seconds to see the updated data (there is a special SWR mode that updates quickly, and produces an audio tone in proportion to the SWR). The CIA-HF had several hardware updates which were added in its first year of production. My unit predates all of those upgrades.

My software version number is 1.4.

On this web page, I claim in several places that these analyzers do not report the sign of the reactance. That is certainly true for the MFJ-269 and the RF-1. The CIA-HF is a more complex story. When you are on what is called the *Data Screen*, the sign of the reactance is indicated by the sign of the impedance angle, ranging from -90 to +90 degrees. A negative value indicates capacitive reactance, and a positive value indicates inductive reactance. As best as I can tell, this is the only indication of the reactance sign. In all other data displays, the reactance or phase angle is always shown as an unsigned number. On the Data Screen, the angle has an explicit + or - sign.

---

same power output and same ground conductivity. The lesson: listen to the lower end of the MW band if you want extreme distance in reception.

2. Look for over-water paths. A local example: Ground conductivity around here in the Rochester, NY area is about 8 or a little less, tending towards the 6 mS/m figure judging from reception experience. Without passive loop assistance, normal daytime reception limits seem to max out at about 150 miles. That is, except for two stations: WJR-760, Detroit, MI and CKLW-800, Windsor, Ontario, Canada, near Detroit. These two are out at about 250 miles or about 400 kilometers. Why are they receivable? Most of their signal path is across water over the east-west length of Lake Erie into Rochester. Reception is routine on most simple radios with medium or better sensitivity.

3. Study station antenna patterns. The FCC web site is great for this. Virtually all stations with multiple towers have directional patterns and the FCC makes the pattern plot available in .PDF form. Study the plot and pay particular attention to your azimuth away from the station. Pick stations that push a lot of signal in your direction. They are the best prospects. Example: WXXI-1370 pattern

4. The radio-locator "fringe" 0.15 mV/m boundary seen on their plots is just that - fringe. Many stations are copyable at this boundary and beyond on a sensitive radio. Don't hesitate to try for them. Example: Denver stations KHOW-630 (only 5KW), KKZN-760 (50KW), and KOA-850 (50KW) all have excellent strength east to well past Hays, Kansas (300+ miles, or 483+ km) despite the fringe zone ending miles back towards the Kansas/Colorado state line.

The jury seems to be out on how far ground conductivity plays a role in daytime propagation. Some documents refer to the 55 kilometer figure, others to 150 kilometers. My bet is it is well beyond the 150 km figure as I have routinely heard stations at the 800 kilometer range during high daylight hours at all times of the year.

PART III: Calculating

Radio-Time Traveler

http://radio-timetraveller.blogspot.com/2011/06/field-strength-measurements-calculating.html

25 June, 2011

As stated in the previous post, one more piece of information is required to complete the puzzle of calculating received field strength. That is the millivolt per meter level at 1 kilometer from the station transmitter, eminating in your direction. Notice I said "eminating in your direction". It is not good enough to simply calculate the mV/m level at 1 kilometer for the station's overall power output. Two things must be accounted for that change that result and would make our field strength calculation inaccurate. They are:

1. Antenna efficiency. A mediumwave tower or array of towers will be more or less efficient depending on their radiation length(s). The FCC provides us with a figure called RMS Theoretical for every station's antenna array whether it be one tower or several, measured in mV/m. Reflected in this figure is the efficiency.

2. Pattern gain. Multiple tower arrays inherently have broadcast patterns. Meaning, of course, they aim to broadcast a majority of their signal in a certain direction to cover their market audience and/or avoid co-channel interference with

Meet the Players

Here are the three antenna analyzers that I owned and tested at the time of the creation of this web page. Please click on the picture for a larger view.

Antenna Analyzers: MFJ-269, RF-1, CIA-HF

The [MFJ-269](MFJ-269) looks like the very popular MFJ-259B. There are some differences. It has a button labeled *UHF* which enables the device to operate up through the 440 MHz band. For my purposes, the primary difference is that the 269 uses a 12-bit A/D (Analog to Digital) converter, which is claimed to produce more accurate results. It is my guess that other analyzers use 8-bit A/D converters, since they are less expensive, and often available directly on the microprocessor IC. The 269 is somewhat more expensive than the 259B.

My software version number is 1.24.

results, they must be used in conjunction with a signal generator and a detector/receiver. Depending upon what you are trying to measure, several computations must be performed to arrive at the result. It can sometimes take *minutes* to make a single measurement with an impedance bridge setup. The delay is due to having to *balance* the bridge, which is a process of adjusting two controls until a very sharp null is found. Another alternative, perhaps the best, is a lab quality impedance analyzer. HP has made a number of models over the years that seem to cost well in excess of $10,000. That takes them way out of my hobby budget. So, for the money, these analyzer boxes are a great value, even with limitations and errors. Overall, their accuracy appears to be quite good.
It's also true that I tested exactly one of each of these meters - the one that I own. Each is several years old, and while I don't believe that I have ever dropped or abused them, they have been well used. I assume that there is one or more *calibration adjustments* for each analyzer, and I have no idea if my units are well-calibrated, or out of calibration.

I don't know how representative my units are of each model, or if there have been design improvements over the years that would change the results.

Finally, the battery quality and lead length will influence results. You can obtain better results if you use fresh batteries (or an external power supply) and short leads. At HF frequencies, even a few inches of wire can act as an inductor with several Ohms of reactance. I can't stress this enough. Short leads, better yet, no leads.

another station. Where you are in relation to that pattern is important. If you are in the major lobe of a 50KW station, it may be pumping upwards of 100KW towards you, or more. If you are in a sharp pattern null, it may only be beaming hundreds of watts towards you. The amount of millivolts per meter "facing you" is the important figure. The FCC provides that as well, in their pattern plots.

There is only one case where we will need to do a simple extra calculation to arrive at the full millivolt per meter level for a station. That is for stations with a single tower only. I will explain why in just a minute.

So let's put together what we need. First, pick a station within reasonable distance you think you'd like to log. Note its frequency. Next, gather the following four things:

1. The ground conductivity in mS/m between you and the station. Use the M3 conductivity map. If the station path crosses a couple of zones, estimate the average ground conductivity for the entire path. The resultant figure should fall between 0.1 mS/m and 30 mS/m, or possibly higher if part of the path is over salt seawater.

2. The Ground Wave Field Strength Versus Distance graph for the frequency of the station, one of the 20 graphs published by the FCC. Several frequencies are usually grouped into one graph. The graphs are in .PDF form. Have your .PDF viewer ready.

3. You will need to find your distance to the station in kilometers, and also the reverse bearing from the station back to you. Many calculators exist on the web which will compute this information. The FCC has a good one, be sure to check out their calculator. These calculators require you to know the

latitude and longitude of both your location and the station's location.

www.wikimapia.org is a great way to determine your home latitude and longitude as it has a crosshair defining the center of the map, and thus the latitude and longitude. Move the map to your exact location and read the latitude-longitude in the web browser's address bar.

The station's latitude and longitude can be found in a couple of ways. The FCC's AM Query web page allows us to query the station by call sign. The search output will display basic information like latitude and longitude. Click on the call sign link and you will be taken to the FCC's web page for the facility (station). Example: WHAM-1180 page.

4. The last item. Get the millivolt per meter value at 1 kilometer from the station transmitter, headed your way. The method of locating this figure will depend on whether the station has one antenna tower or multiple towers in its array.

Determine if the station uses a single tower or multiple towers for the service you are interested in. This information can be found on the FCC's web page for the facility (station), as shown just above.

Note that stations may have more than one entry on the page, one for each service they operate under, i.e., UNLIMITED, DAYTIME, NIGHTTIME, CRITICAL HOURS. Be sure you are looking at the correct service. Sometimes stations use a different number of towers for day and night.

mV/m for multiple towers.

The mV/m figure is gotten from the pattern data. It's simple.

W8WWV - Analyzing Three Antenna Analyzers

Greg Ordy
http://www.seed-solutions.com/gregordy/Amateur%20Radio/Experimentation/EvalAnalyzers.htm

Introduction

On this page I would like to present some measurements that I made with three of the popular antenna analyzers. I became interested in their accuracy when I had an application where I wanted to make accurate measurements of relatively low impedance values. Values where the resistance and reactance were close to zero Ohms, certainly under 50 Ohms. I had no reason to doubt any of the measurements, but from time to time I would make measurements with more than one of the units and find that they might differ by up to 10 Ohms. For many applications this difference is not significant, but I decided to see if I could determine how the analyzers compared to reference values, as opposed to each other.

Accuracy and Value

I'm not trying to present a detailed evaluation of these analyzers, or to rank their quality. My measurements and their accuracy are not precise enough for that purpose. While it's clear that these analyzers can report results with error, they are all a tremendous value, and a useful addition to any amateur radio station.

I call them tremendous values because I don't know how to make better measurements without spending substantially more time and/or money. Historically, impedance measurements were made with the *impedance bridge*. While a laboratory quality impedance bridge can produce excellent

Variable Capacitor Wired Wiperless

By Jim, Amateur Radio Station KR1S

http://theradioboard.com/rb/viewtopic.php?t=3262&postdays=0&postorder=asc&start=15

If I'm reading this correctly, a schematic would be:

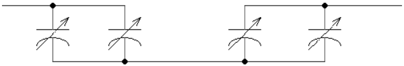

For a four-section, 500-pF per section cap, the total C would be 500 pF.

Multiple tower arrays will give you the option to display the pattern plot. The pattern plot link will be under a heading that looks like this:

Horizontal Pattern at 1 km radius (Sections 73.150 and 73.152):
    Electric Field Strength pattern plot
    Pattern Data for WXXI

Either link will give us the information we need, though the "Electric Field Strength pattern plot" gives a nice graphic pattern plot for the station. Click one of the links.

RMS Theoretical values, and in some cases RMS Standard or even RMS Augmented values will be displayed for each five degrees of compass, 0-360. Find the compass bearing that most closely matches the return bearing from the station to you. We need to record one value only. Preferably, record the RMS Augmented value, if given. If not available, record the RMS Standard value. If not given, record the RMS Theoretical value. These values are in millivolts per meter and are the value of signal level the station presents towards you.

A quick definition of RMS Standard and RMS Augmented values. RMS Standard is essentially the RMS Theoretical value plus 5%. It is a "guard" against interference to other co-channel stations by overstating the RMS Theoretical calculated value. If stations have pattern augmentations, and many do, the RMS Augmented field will be present. Augmentations are enhancements or detractions to the theoretical pattern.

I'll use WXXI-1370 for the example. At night it runs 4 towers and has a roughly figure-8 pattern north-south. WXXI's return bearing to me is 204.7 degrees. Checking the FCC pattern plot for the station, we see that the 205 degree return azimuth presents a facing RMS Theoretical of 350.59 mV/m and a facing RMS Standard of 368.88 mV/m. We will use the RMS Standard value in the final calculation.

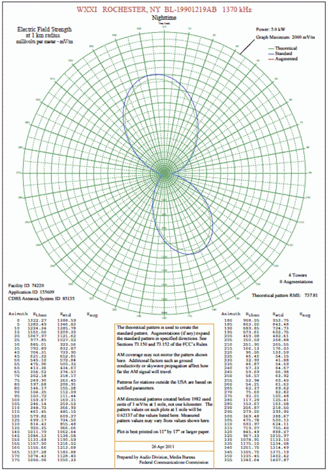

WXXI ROCHESTER, NY BL-19901219AB 1370 kHz

Nighttime

Electric Field Strength
at 1 km radius
millivolts per meter - mV/m

Power: 5.0 kW

Graph Maximum: 2000 mV/m

— Theoretical
— Standard
— Augmented

4 Towers
0 Augmentations

Theoretical pattern RMS: 737.81

Facility ID: 74220
Application ID: 155609
CDBS Antenna System ID: 85135

Azimuth	E_{theo}	E_{std}	E_{aug}
0	1322.27	1388.59	
5	1282.49	1346.83	
10	1224.34	1285.78	
15	1151.50	1209.30	
20	1067.97	1121.62	
25	977.85	1027.02	
30	885.01	929.56	
35	792.89	832.87	
40	704.31	739.90	
45	621.32	652.81	
50	545.10	572.84	
55	475.96	500.32	
60	413.36	434.67	
65	356.02	374.57	
70	302.18	318.17	
75	249.90	263.45	
80	197.68	208.91	
85	146.17	155.28	
90	104.35	112.09	
95	103.72	111.44	
100	159.67	169.31	
105	246.14	259.53	
110	348.73	366.93	
115	461.45	485.10	
120	579.82	609.27	
125	699.17	734.50	
130	814.43	855.48	
135	920.35	966.66	
140	1011.79	1062.64	
145	1084.13	1138.58	
150	1133.68	1190.59	
155	1157.90	1216.02	
160	1155.66	1213.68	
165	1127.28	1183.88	
170	1074.43	1128.40	
175	1000.06	1050.33	
180	908.05	953.75	
185	803.00	843.48	
190	689.85	724.73	
195	573.61	602.75	
200	459.08	482.61	
205	350.59	368.88	
210	251.90	265.55	
215	166.13	176.03	
220	96.06	103.59	
225	46.40	54.15	
230	32.93	41.88	
235	46.87	54.59	
240	57.33	64.67	
245	59.69	66.98	
250	56.57	63.92	
255	52.94	60.40	
260	54.21	61.63	
265	62.37	69.62	
270	75.60	82.82	
275	93.01	100.48	
280	117.29	125.41	
285	153.67	163.08	
290	206.87	218.50	
295	279.00	293.90	
300	369.48	388.67	
305	475.78	500.13	
310	593.97	624.11	
315	719.07	755.40	
320	845.49	888.07	
325	967.33	1015.97	
330	1078.91	1133.10	
335	1175.10	1234.08	
340	1251.70	1314.49	
345	1305.70	1371.19	
350	1335.45	1402.42	
355	1340.64	1407.87	

The theoretical pattern is used to create the standard pattern. Augmentations (if any) expand the standard pattern in specified directions. See Sections 73.150 and 73.152 of the FCC's Rules.

AM coverage may not mirror the pattern shown here. Additional factors such as ground conductivity or skywave propagation affect how far the AM signal will travel.

Patterns for stations outside the USA are based on notified parameters.

AM directional patterns created before 1982 used units of 1 mV/m at 1 mile, not one kilometer. The pattern values on such plots at 1 mile will be 0.62137 of the values listed here. Measured pattern values may vary from values shown here.

Plot is best printed on 11" by 17" or larger paper.

26 Apr 2011

Prepared by Audio Division, Media Bureau
Federal Communications Commission

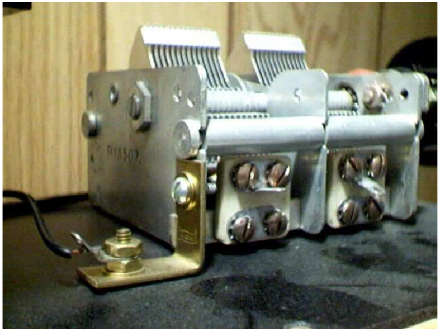

This is my common to the frame or body of the capacitor.

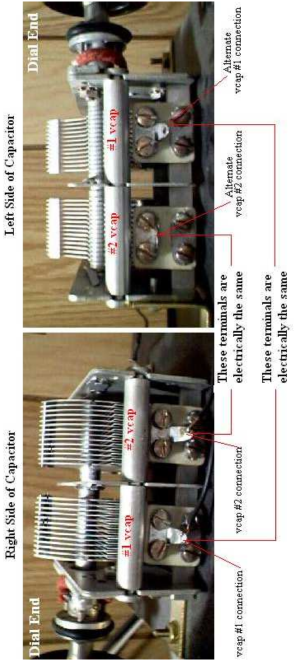
Left Side of Capacitor
Dial End
#1 vcap
#2 vcap
Alternate vcap #1 connection
Alternate vcap #2 connection
These terminals are electrically the same
These terminals are electrically the same
Right Side of Capacitor
#2 vcap
#1 vcap
vcap #2 connection
vcap #1 connection
Dial End

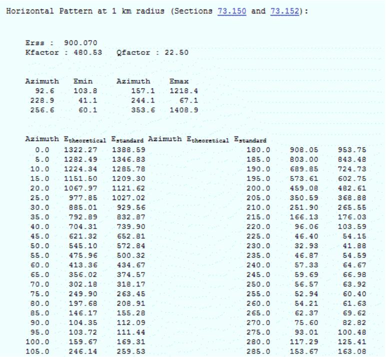

```
Horizontal Pattern at 1 km radius (Sections 73.150 and 73.152):


    Erss :  900.070
    Kfactor : 480.53   Qfactor : 22.50


    Azimuth   Emin     Azimuth   Emax
      92.6    103.8       157.1   1218.4
     228.9     41.1       244.1     67.1
     256.6     60.1       353.6   1408.9


    Azimuth Etheoretical Estandard Azimuth Etheoretical Estandard
      0.0   1322.27   1388.59                          180.0    908.05    953.75
      5.0   1282.49   1346.83                          185.0    803.00    843.48
     10.0   1224.34   1285.78                          190.0    689.85    724.73
     15.0   1151.50   1209.30                          195.0    573.61    602.75
     20.0   1067.97   1121.62                          200.0    459.08    482.61
     25.0    977.85   1027.02                          205.0    350.59    368.88
     30.0    885.01    929.56                          210.0    251.90    265.55
     35.0    792.89    832.87                          215.0    166.13    176.03
     40.0    704.31    739.90                          220.0     96.06    103.59
     45.0    621.32    652.81                          225.0     46.40     54.15
     50.0    545.10    572.84                          230.0     32.93     41.88
     55.0    475.96    500.32                          235.0     46.87     54.59
     60.0    413.36    434.67                          240.0     57.33     64.67
     65.0    356.02    374.57                          245.0     59.69     66.98
     70.0    302.18    318.17                          250.0     56.57     63.92
     75.0    249.90    263.45                          255.0     52.94     60.40
     80.0    197.68    208.91                          260.0     54.21     61.63
     85.0    146.17    155.28                          265.0     62.37     69.62
     90.0    104.35    112.09                          270.0     75.60     82.82
     95.0    103.72    111.44                          275.0     93.01    100.48
    100.0    159.67    169.31                          280.0    117.29    125.41
    105.0    246.14    259.53                          285.0    153.67    163.08
```

mV/m for single tower.

A special case requiring a simple calculation. We will calculate the mV/m figure from the RMS Theoretical value.

Again go to the FCC's web page for the facility (station), as above. The RMS Theoretical value will be on this page.

Note again that stations may have more than one entry on the page, one for each service they operate under, i.e., UNLIMITED, DAYTIME, NIGHTTIME, CRITICAL HOURS. Be sure you are looking at the correct service. Sometimes stations use a different number of towers for day and night.

The FCC computes all RMS Theoretical values from a formula of course. The values are calculated for a distance of 1 kilometer. The FCC formula used generates accurate millivolt per meter values (as published) for multiple tower arrays. Single tower arrays are a special case, however, in that the published mV/m value is always based on a 1 kilowatt output power calculation. Hence, the only published single tower mV/m values we can use are those of 1 KW stations. For all others, we will do a simple calculation to arrive at the correct mV/m value. Proof of this is simple. For example, check the FCC's published figures for my local WHAM-1180 station out of Rochester, NY. This 50KW station shows a calculated RMS Theoretical value at 1 kilometer of only 376.59 mV/m. Now of course this cannot be correct for a 50KW station, as a 1 KW station running a quarter wave (.250 wavelength) monopole has an exact calculated figure of 305.768 mV/m at 1 kilometer.

376.59 mV/m would, however, be correct for a 1 KW station using the same single tower antenna that WHAM uses (a .492 wavelength antenna).

To accurately calculate the mV/m figure for WHAM (or any other single tower station, including those 1 KW stations), the following formula must be applied:

(Power in KW, distance in KM):

mV/m = RMSTheoretical x SQRoot(Power/Distance)

Thus in WHAM's case:

2662.89 = 376.59 * SQRoot(50/1)

WHAM's actual RMS Theoretical value is 2662.89 mV/m. And since it is a single tower antenna having an

---

How to Hook Up A Variable Capacitor: Two Gang Capacitors

by Darrel Boyd

http://www.crystalradio.net/beginners2/caphookup.shtml

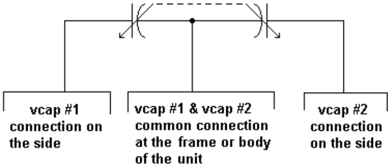

This is the typical wiring for a two gang vcap
Refer to photo below

Next, we shall compute the bandwidths at three frequencies: 530kHz, 1MHz and 1.7MHz. Results are tabulated below.

Freq.(kHz)	BW (kHz)	Q	R_P (kΩ)	A (kΩ)	C (pF)	Ca (pF)	r (Ω)	L_1 (uH)
530	5.004	105.9	354.32	155.50	450	200	30	152
1000	13.47	74.24	558.7	190	100	200	30	152
1700	24.83	68.49	487.072	406.8	31	200	30	152

Results for BW are very close to those obtained from simulation of circuit of Fig.1.b.

Ramon Vargas Patron

rvargas@inictel.gob.pe

Lima-Peru, South America

April 11th 2004

---

omnidirectional pattern, it presents this value of signal in all directions. Use the value you calculate for your single tower station of interest as the mV/m value that the station presents towards you.

Making The Calculation

Now we have all of our information. Let's get busy. We will use the FCC's Ground Wave Field Strength Versus Distance graph to arrive at the received mV/m signal level. Proceed with the following steps.

1. To make them universal, the FCC's Ground Wave Field Strength Versus Distance graphs are based on 100 mV/m levels at 1 kilometer. We simply need to calculate how many 100s our mV/m value is. Just move the decimal point left two places. In WHAM's case, 2662.89 mV/m, 26.6289 (26.6289 x 100 = 2662.89). The multiplier value we will use for WHAM is 26.6289. In WXXI's case, 368.88 mV/m, 3.6888 (3.6888 x 100 = 368.88). The multiplier value we will use for WXXI is 3.6888.

2. Find the station distance in kilometers on the graph, usually at the bottom. The bottom range is 10 to 1000 kilometers. The top range is 0.1 to 50 kilometers.

3. Draw a trace upwards (or downwards if using the top scale) until you hit the ground conductivity value curve that matches the average ground conductivity between you and the station.

4. From the previous point, draw a trace leftward to the scale on the left side of the graph. This is the base millivolt per meter level based on 100 mV/m at 1 kilometer. Multiply this value by your multiplier value. In WHAM's case, multiply times 26.6289. In WXXI's case, multiply times 3.6888. This

resultant value is the received field strength in millivolts per meter at your location.

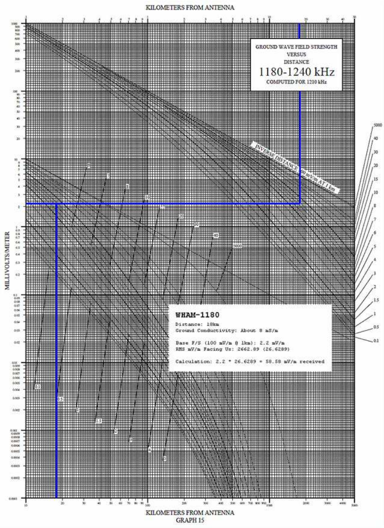

There you have it. You have ballparked the approximate field strength of your station of interest. If done correctly, you should find this in fairly good agreement with V-Soft's figure

---

$$2L_1[C_T + C(1 \pm \omega_r C_T R_e)]\Delta\omega = \pm C_T R_e \left(\frac{Ca}{2Ca + C}\right)$$

Now, if $\omega_r C_T R_e \ll 1$, then:

$$2\Delta\omega = \pm \frac{R_e}{L_1}\left(\frac{C_T}{C_T + C}\right)\left(\frac{Ca}{2Ca + C}\right)$$

this is:

$$\Delta\omega = \pm \frac{R_e}{2L_1}\left(\frac{Ca}{2Ca + C}\right)\left(\frac{Ca}{2Ca + C}\right)$$

$$= \pm \frac{R_{s1}}{L_{eq}}\left(\frac{Ca}{2Ca + C}\right) \qquad \dots(2)$$

From part II of our study we can obtain the following expression for R_{s1}:

$$R_{s1} = r\left(1 + \frac{A}{R_P}\right)$$

Then, the 3dB bandwidth is given by:

$$BW = 2\Delta\omega = 2r\left(1 + \frac{A}{R_P}\right)\left(\frac{1}{L_1\left(2 + \frac{C}{Ca}\right)}\right)\left(\frac{Ca}{2Ca + C}\right) \qquad \dots(3)$$

in radians per second.

or:

$$\frac{\omega L_1}{1-\omega^2 L_1 C} - \frac{1}{\omega C_T} = \pm R_e \qquad \ldots(1)$$

Let ω_r be the resonant frequency and $\omega = \omega_r + \Delta\omega$ the frequency at a –3dB point on the amplitude curve. The left hand member of eq. (1) can be written as:

$$\frac{\omega^2 L_1(C_T+C)-1}{(1-\omega^2 L_1 C)\omega C_T} = \frac{(\omega_r^{\ 2}+2\omega_r\Delta\omega)(C_T+C)L_1-1}{(1-(\omega_r^{\ 2}+2\omega_r\Delta\omega)L_1 C)\omega_r C_T}$$

with the following approximations:

$$(\omega_r+\Delta\omega)C_T \approx \omega_r C_T$$

$$(\omega_r+\Delta\omega)^2 = \omega_r^{\ 2}+2\omega_r\Delta\omega+\Delta\omega^2 \approx \omega_r^{\ 2}+2\omega_r\Delta\omega$$

Then:

$$\frac{\omega_r^{\ 2} L_1(C_T+C)+2\omega_r(C_T+C)L_1\Delta\omega-1}{(1-\omega_r^{\ 2}L_1 C-2\omega_r L_1 C\Delta\omega)\omega_r C_T} = \pm R_e$$

or:

$$\frac{2\omega_r(C_T+C)L_1\Delta\omega}{\left(\dfrac{Ca}{2Ca+C}\right)-2\omega_r L_1 C\Delta\omega} = \pm\omega_r C_T R_e$$

which simplifies to:

if you are near the zipcode point they based their calculation on. With a list of expected receive field strengths for various stations, you can judge the approximate sensitivity of your receiver. After a few times trying this, you will find the calculation to be rather simple to do.

PART IV: Field Strength Calculator One

http://radio-timetraveller.blogspot.com/2012/12/field-strength-calculator-one.html

20 December, 2012

Field Strength Calculator One is a program which will calculate expected received ground wave signal strength at longwave and mediumwave frequencies.

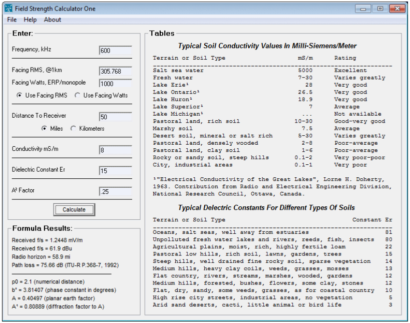

Frequency, kHz 600

Facing RMS, @1km 305.768

Facing Watts, ERP/monopole 1000

● Use Facing RMS ○ Use Facing Watts

Distance To Receiver 50

● Miles ○ Kilometers

Conductivity mS/m 8

Dielectric Constant Er 15

A² Factor .25

Formula Results:

Received f/s = 1.2448 mV/m
Received f/s = 61.9 dBu
Radio horizon = 58.9 mi
Path loss = 75.66 dB (ITU-R P.368-7, 1992)

p0 = 2.1 (numerical distance)
b° = 3.81407 (phase constant in degrees)
A = 0.40497 (planar earth factor)
A¹ = 0.80889 (diffraction factor to A)

Typical Soil Conductivity Values In Milli-Siemens/Meter

Terrain or Soil Type	mS/m	Rating
Salt sea water	5000	Excellent
Fresh water	7-30	Varies greatly
Lake Erie¹	28	Very good
Lake Ontario¹	26.5	Very good
Lake Huron¹	18.9	Very good
Lake Superior¹	7	Average
Lake Michigan¹	...	Not available
Pastoral land, rich soil	10-30	Good-very good
Marshy soil	7.5	Average
Desert soil, mineral or salt rich	5-30	Varies greatly
Pastoral land, densely wooded	2-8	Poor-average
Pastoral land, clay soil	1-6	Poor-average
Rocky or sandy soil, steep hills	0.1-2	Very poor-poor
City, industrial areas	0.1-1	Very poor

¹"Electrical Conductivity of the Great Lakes", Lorne H. Doherty, 1963. Contribution from Radio and Electrical Engineering Division, National Research Council, Ottawa, Canada.

Typical Delectric Constants For Different Types Of Soils

Terrain or Soil Type	Constant Er
Oceans, salt seas, well away from estuaries	81
Unpolluted fresh water lakes and rivers, reeds, fish, insects	80
Agricultural plains, moist, rich, highly fertile loam	22
Pastoral low hills, rich soil, lawns, gardens, trees	15
Steep hills, well drained fine rocky soil, sparse vegetation	14
Medium hills, heavy clay coils, weeds, grasses, mosses	13
Flat country, rivers, streams, marshes, wooded, gardens	12
Medium hills, forested, bushes, flowers, some clay, stones	12
Flat, dry, sandy, some weeds, grasses, as for coastal country	10
High rise city streets, industrial areas, no vegetation	5
Arid sand deserts, cacti, little animal or bird life	3

DOWNLOAD

https://app.box.com/s/g257enuzufin7kincn6y
To download, see the link at the top of the right sidebar under LATEST PROGRAMS. The sidebar at the top right will have the most current link in case the program is updated. The link will change in the case of an update, so I would avoid copying and pasting it into a forum or other web page. Come to the main page of this blog instead.

DISCUSSION

Being a mediumwave DXer and particularly a daytime mediumwave DXer, I wanted a way to determine a "ballpark" signal strength for various stations not only in my immediate area, but out to 100, 200, or even more miles distant. I was unhappy with virtually all the web-based signal strength calculators found on the internet, as they use the vanilla Inverse-Square-Law to calculate signal attenuation. Fine, if you are in an outer space vacuum or on a perfecting conducting surface, but not even close in accuracy for normal people here on Mother Earth.

The few stand-alone programs out there were either wildly expensive, too complicated to use, inaccessible, or plainly won't work on the Windows platform. I set out to accumulate information, formulas, and data to start writing the field strength calculator program. Investigating the history and ferreting out the pertinent information to arrive at a simplified formula that was reasonably accurate took some time.

The result was and is Field Strength Calculator One. It is based on the work of numerous engineers and mathematicians, who, starting about 1909, spent some 50 years developing the extremely complicated formulas to predict accurate signal strength at mediumwave frequencies. The basic, simplified formula has been known since the 1930s, being slightly

---

We shall now proceed to calculate the 3dB bandwidth of circuit of Fig.1.b.
Mesh current is given by:

$$I = \frac{E_a}{R_e + j\left(\frac{\omega L_1}{1-\omega^2 L_1 C} - \frac{1}{\omega C_T}\right)}$$

where:

$$C_T = \frac{CaC}{Ca + C}$$

The amplitude-frequency relationship for I is determined by:

$$I(\omega) = \frac{E_a}{\sqrt{R_e{}^2 + \left(\frac{\omega L_1}{1-\omega^2 L_1 C} - \frac{1}{\omega C_T}\right)^2}}$$

At the -3dB points:

$$I(\omega) = \frac{E_a}{\sqrt{2}R_e}$$

The corresponding frequencies must satisfy the equation:

$$R_e{}^2 + \left(\frac{\omega L_1}{1-\omega^2 L_1 C} - \frac{1}{\omega C_T}\right)^2 = 2R_e{}^2$$

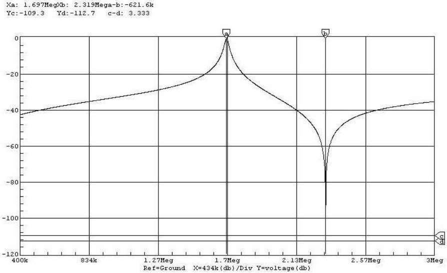

Fig.3.e Frequency response for circuit of Fig.1.b when resonance is adjusted to 1.7MHz

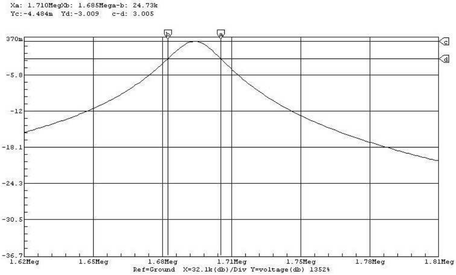

Fig.3.f 3dB bandwidth for a resonance frequency of 1.7MHz is 24.73kHz

3dB bandwidth calculations

modified by various people and agencies since then. It is accurate to within a couple of percent of the big programs that calculate field strength - those using additional input like the transmit and receive array heights above average ground, and the earth's topographic elevation changes along the signal path.

Simplified ground wave electrical field intensity calculations can be made by the introduction of a shadow or diffraction factor in the Sommerfeld-Norton planar earth expression. A mouthful! This simply means that a factor is computed and introduced to account for the additional attenuation caused by wave diffraction out beyond the radio horizon. It permits one to calculate the ground wave E (electrical) field well beyond the geometric and radio horizon, where E field values are close to the atmospheric noise level.

Be sure to read about the history of how this fascinating formula came about in the recent article on RADIO-TIMETRAVELLER: Field Strength Calculations: A History. Many of the terms used in the previous and next paragraphs are explained.

The simplified formula used by Field Strength Calculator One takes into account Sommerfeld's original plane earth theory, modified by diffraction factoring. It uses an exponential function which takes into account the spherical earth diffraction effects, and acts on the planar earth equation even before the radio horizon is reached, so the resultant E field values, as a function of distance produce a continuous curve, thus rounding-in difficult intermediate distances.

The long-accepted concept of "numerical distance" (p0) and "phase angle" (b) are used in all calculations, two variables determined by frequency, distance, and dielectric constants of the ground as a radio conductor. Numerical distance depends not only on frequency and ground constants, but also on the actual distance to the transmitter. Phase angle is the measure of the power factor angle of the earth.

Field Strength Calculator One returns expected received field strength in millivolts per meter and dBu (also known as dBµV/m), based on ground conductivity, earth dielectric and several other input constants. It also displays the distance to the radio horizon and the signal path loss in dB, along with several more technical parameters. The resulting output of Field Strength Calculator One should be accurate in most cases to a couple of percent in the longwave and mediumwave bands. It compares favorably to ITU program GRWAVE and currently available FCC Ground Wave Conductivity graphs.

Field strength calculations by Field Strength Calculator One are based on the works of A.Sommerfeld (1909), B.Rolf (1930), K.A.Norton (1936), H.Bremmer (1949, 1958), NTIA Report 86-203 (1986), ITU-R P.368-7 (1992), and NTIA Report 99-368 (1999).

For further information on how field strength is calculated see the Field Strength Calculations Series previously published on RADIO-TIMETRAVELLER.

INSTALL

An install file is included on the web page:
http://radio-timetraveller.blogspot.com/2012/12/field-strength-calculator-one.html

Included in the .zip is a readme.txt file. Be sure to have a look.

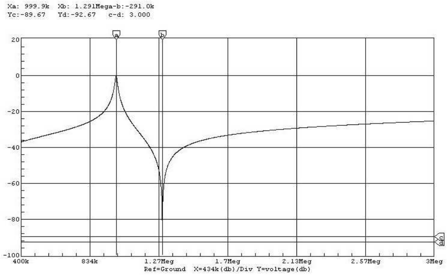

Fig.3.c Frequency response for circuit of Fig.1.b
when resonance is adjusted to 1MHz

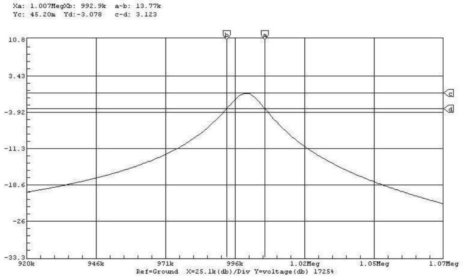

Fig.3.d 3dB bandwidth for a resonance
frequency of 1MHz is 13.77kHz

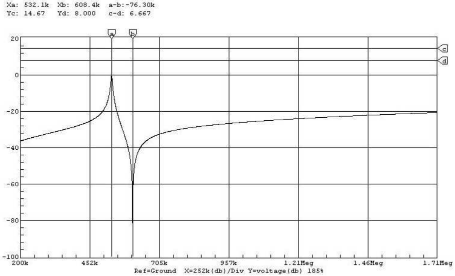

Fig.3.a Frequency response for circuit of Fig. 1.b
when resonance is adjusted to 530kHz

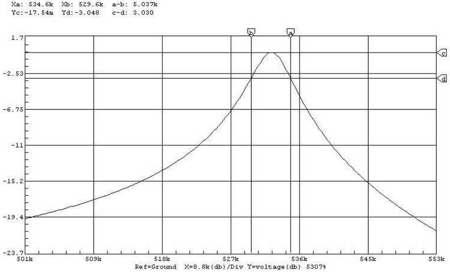

Fig.3.b 3dB bandwidth for a resonance
frequency of 530kHz is 5.037kHz

ANTENNA AND GROUND SYSTEM

by Kenneth A. Kuhn

http://www.kennethkuhn.com/students/crystal_radios/antenna_and_ground_system.pdf

Introduction

This article explains the basics of the antenna and ground system for a crystal radio and the associated mathematical model. The mathematical model tells us how to design the antenna input section of the radio for optimum performance.

The passing RF wave induces an RF voltage across a length of wire referred to as an antenna. The induced voltage is the signal strength in volts/meter multiplied by the electrical (not the physical) height of the antenna in meters. For antennas that are shorter than one-quarter wavelength the electrical height is roughly the wire length (including lead-in wire) in meters. There are very little directivity effects for short antennas so orientation is not really a factor –i.e. the antenna is omnidirectional. The physical height of the antenna above the earth affects the signal amplitude as there is increased attenuation of the signal close to the ground.

The antenna should be as long as practical and as high as practical. A quarter wavelength for 540 kHz is (300,000/4)/540 = 139 meters. A quarter-wavelength at 1700 kHz is 44 meters. For a basic crystal radio there is little to be gained by an antenna longer than about 40 meters or higher than about 8 meters –that is a substantial antenna that few have the room to construct. A minimum antenna might be about 10 meters long and about 3 meters off the ground. Anything in-between can produce acceptable performance. Advanced crystal radio enthusiasts will construct bigger and more advanced antennas but that involves a degree of engineering beyond the scope of this article.

One advanced method that the author has not had the space to try is a true dipole antenna that is one-half wavelength long (in some cases longer). Such an antenna is balanced and so does not depend on a low-impedance ground for operation. The power received by the antenna would be significantly more than that of the simple long wire antennas described in this article. With such an antenna a local station could drive a speaker to significant volume. But, the antenna is only useful over a narrow band of frequencies and is also directive so it must be oriented properly relative to the transmitting antenna.

Figure 1 shows a typical antenna and ground setup for a crystal radio. The following sections discuss each attribute.

Figures 3.a through 3.f show simulation results for circuit of Fig.1.b at resonance frequencies also of 530kHz, 1MHz and 1.7MHz. Again, r = 30 ohms and Ca = 200pF.
Re = $2Rs_1$, and it can be easily shown that:

$$R_{s1} = r\left(1 + \frac{A}{R_p}\right)$$

The Y axis on the graphics represents voltage across Re in decibels, with Ea being a 1 volt-amplitude unmodulated carrier.

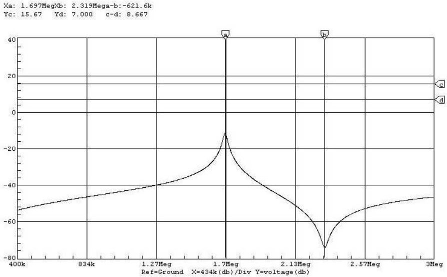

Fig.2.e Frequency response of circuit of Fig.1.a when resonance is adjusted to 1.7MHz

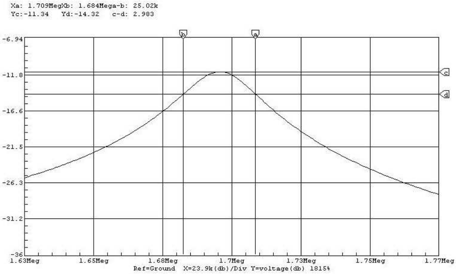

Fig.2.f 3dB bandwidth for a resonance frequency of 1.7MHz is 25.02kHz

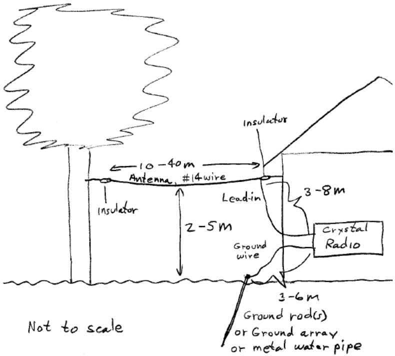
Insulator
10 - 40m
Antenna, #14 wire
3 - 8m
Lead-in
Insulator
2 - 5m
Crystal Radio
Ground wire
3 - 6m
Ground rod(s)
or Ground array
or metal water pipe
Not to scale

Figure 1: Basic Antenna System for a Crystal Radio

Distal support

If available, a tree makes a convenient support for the antenna wire. Otherwise some kind of post will have to be constructed. Never use a utility pole or any other support that is not your property –your potential liabilities could be astronomical if some accident (lightning for example) or failure causes damage to someone else’s property or person. An insulator typically made of ceramic or plastic is used to terminate the antenna wire and to connect to the antenna support usually via some kind of rope. These insulators are readily available from various hobby electronics sources. A large wood screw typically with a hook or eye end driven into a tree or other support is used as an attachment point for the rope. Do not wrap the rope around the tree as that will constrict growth in future years.

Antenna wire

The antenna wire is typically #14 AWG solid or stranded copper wire and may or may not be insulated. Insulation has no effect on reception but is an advantage to reduce corrosion effects. There is no advantage to using heavier wire such as #12 or #10 as the resistance of the wire is negligible compared to everything else and the added weight makes the end supports more challenging. Wire sizes of #16 and #18 also work well but are less strong –but are also lighter and will put less stress on the supports. Wire gauge sizes higher than 18 are not recommended because they are more fragile and are hard to see if they fall –you would like to not run over the wire with a lawn mower –could be very disastrous! Although low resistance is nice, it is a myth to go to great efforts for low resistance in the antenna wire as the resistance of the ground is going to be considerably higher.

The height of the antenna absolutely must be above head level for someone who might be standing on some vehicle (i.e. bicycle or motorcycle) that might possibly travel under the antenna to prevent what could be a horrible or even tragic accident –this means at least 3 meters. The antenna wire should not pass over anything such as a road or other wires such as electrical power such that if the antenna fell that a dangerous situation could result. Within practical limits the higher the antenna the more signal it will pick up. That does not extend indefinitely and there is a point of diminishing returns reached at roughly 5 meters. If it is easy for you to make the antenna higher then do so but there is little point in going to great effort to achieve that as you will only notice a difference in the extreme small signal case.

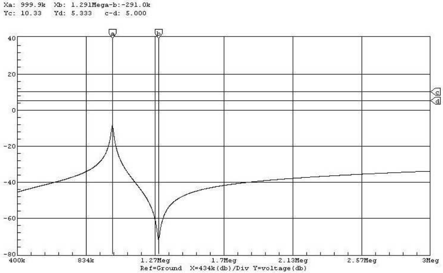

Fig.2.c Frequency response of circuit of Fig.1.a
when resonance is adjusted to 1MHz

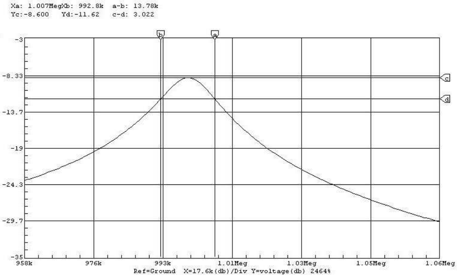

Fig.2.d 3dB Bandwidth for a resonance
frequency of 1MHz is 13.78kHz

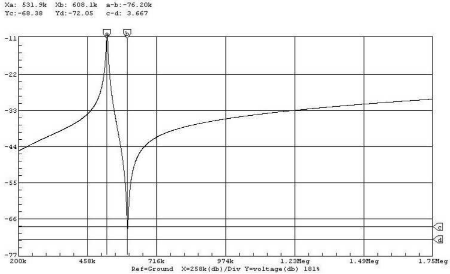

Fig.2.a Frequency response of circuit of Fig.1.a
when resonance is adjusted to 530kHz

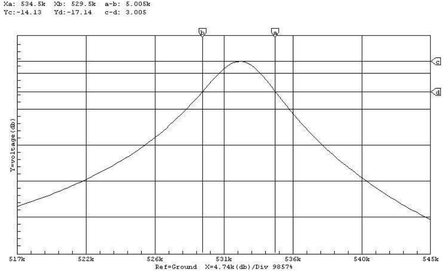

Fig.2.b 3dB bandwidth for a nominal
resonance frequency of 530kHz is 5.005kHz.
Actually, the resonance frequency is 532kHz.

Proximal support

This support is similar to the distal support and may be attached to either the dwelling or a convenient tree that happens to be close. An insulator supports the antenna and rope. The lead-in wire attaches to the antenna at the insulator. It is important to make sure that the support can break away without damaging the structure should something fall on the antenna wire.

Lead-in wire

The lead-in wire can be whatever is convenient and is often #18 wire. Since the antenna system is much shorter than one-quarter wavelength the lead-in counts as part of the antenna length. Thus, if the antenna wire was 25 meters and the lead-in wire was 5 meters, the total antenna length would be 30 meters.

Lightning arrestor

A lightning arrestor is an important part of the antenna system that is located outside the building and usually close to where the antenna and ground wires penetrate. A lightning arrestor consists of a spark gap that will arc when more than a few hundred volts exists across the points. Contrary to the name, a lightning arrestor will not protect you or the dwelling from a direct lighting strike. A lightning arrestor will reduce the probability of damage to your crystal set and anything nearby should lightning strike in the vicinity. A nearby strike can induce many hundreds or even thousands of volts on the antenna that could cause damage or injury. The lightning arrestor will then arc to limit the voltage to typically several tens of volts. You should never operate a crystal radio if there is any possibility of a lightning strike. A number of people install shorting switches or connections to connect the antenna to ground when the radio is not being used. This provides some measure of safety but is not absolute. Nothing can protect you or your dwelling from a direct lightning strike.

Ground wire and ground system

The ground wire is typically #18 copper (because that is a convenient size) and should connect to either a nearby metal water pipe or some ground system –either buried pipe or ground rod(s). It is a myth to use wide braid or other large conductor for the ground wire as the resistance of a short length of #18 is negligibly small in comparison to the earth. It is nice for the wire to be no longer than necessary because inductive reactance in the wire can interfere with the operation of the radio –but that is a small point.

Most difficulties or frustrations with crystal radios can be traced to a poor ground system. Building a good ground system is the most labor intensive and even most expensive part of a crystal radio. If you are lucky then there is a long metal water pipe located just outside the window of your crystal radio. In that special case your ground system is easy and cheap.

If you are not so lucky then one alternative is to drive three or more 2.4 meter (8 foot) ground rods straight into the ground near the house. These should be spaced roughly 2 meters apart. The quality of this ground can vary a lot depending on soil conditions and it is not likely to be as good as that of a water pipe. At the easiest, this is a very labor intensive to even impossible job if the ground is very hard. A heavy duty hammer drill can make the job easier. Otherwise you will need

Simulation results

Figures 2.a through 2.f illustrate simulation results for circuit of Fig.1.a at resonance frequencies of 530kHz, 1MHz and 1.7MHz. Assumed values for r and Ca are 30 ohms and 200pF, respectively. The values for R_T are those obtained when the secondary load R_2 is impedance matched to the primary side (please refer to part II). Notch frequencies occurring above resonance can be observed on the graphics. The Y axis represents voltage across r in decibels with Ea being a 1volt-amplitude unmodulated carrier.

Analysis of the Tuggle Front End – Part III

As a first approximation, the 3dB bandwidth of the antenna-ground-lossy tuner system under matched load conditions can be computed assuming that the L_1-C tank behaves as an equivalent constant inductance Leq in the 3dB passband, this inductance being in series with the rest of the circuit. However, this approach leads to large errors in the results, as suggested by a SPICE circuit simulation.

A precise model for accurate bandwidth computation is shown in Fig.1.a below. R_T is the net RF resistance in parallel with the L_1-C tank at $\omega = \omega_r$, as found in part II of our study. Ground losses Rg and antenna radiation resistance r_a are accounted for by r. However, calculations on this circuit are rather tedious. Simulation shows that the circuit depicted in Fig.1.b can be alternatively employed for bandwidth computation with equivalent results to those given by the circuit of Fig.1.a. Here, Re = $2Rs_1$ (please refer to part II). Circuit calculations in this case are much more simple.

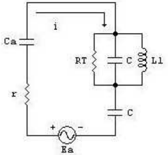

Fig.1.a Equivalent circuit for accurate 3dB bandwidth calculation

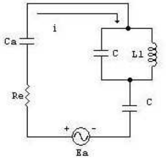

Fig.1.b Alternative circuit for bandwidth calculation as suggested by simulation

---

a very heavy sledge hammer and strong and enduring muscles as well as significant patience.

An alternative is to dig a shallow trench at least 0.2 meters deep and bury a 10 meter or more length of copper pipe. It helps for the soil to be wet down to the level of the pipe when the radio is being used. This can make a decent ground for radio reception although it should in no way be considered a “safety” ground.

The following table provides a general idea of the quality of the ground system.

Ground Quality	Resistance
Excellent	< 10 ohms
Good	10 to 20 ohms
Fair	20 to 40 ohms
Poor	40 to 100 ohms
Bad	> 100 ohms

A good question is how to measure the ground resistance. This is kind of hard to do since you have access to only one wire. Where is the other end that we measure with respect to? The answer is that it is nebulous but there does exist an effective point that completes the circuit for the radio signal. The effective ground resistance can be inferred from measurements using a received signal. The method is discussed in another chapter. Mathematical model of the antenna and ground system

A simple electrical model of a short wire antenna less than one-quarter wavelength is shown in Figure 2. It consists of a voltage source equal to the induced amplitude with a small series resistance and very large series capacitive reactance. A simple electrical model of the ground circuit is a series resistance that may range from single digit ohms (an excellent ground) to several tens of ohms. A poor ground may have a series resistance of over one hundred ohms.

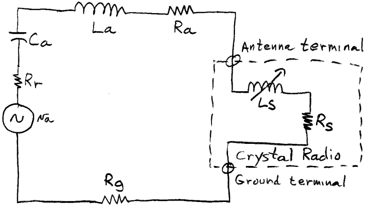

Figure 2: Simple Circuit Model for Antenna and Ground

Signal pickup

The radio frequency field strength is measured in volts/meter (or amperes/meter if you divide volts/meter by the free space impedance of 377 ohms). This is a cyclic voltage gradient. The voltage that will be induced in a conductor that is short in comparison to a quarter wavelength is roughly the RF signal multiplied by the length of the conductor. Proximity of the conductor to other conductive objects (such as trees and the earth) will reduce the induced voltage. This is one reason why it is desirable for the antenna to be as high as practical above the earth. There is no difference between the antenna proper and the lead-in so the total length is what counts. In Figure 2 this voltage is shown as Va.

Comments:

The values obtained for the coupling coefficient K hold for $R_2 = \frac{R_{TNK}}{3}$, as discussed previously. The transformed antenna-ground system resonance resistance, as seen from the secondary, will be equal to R_2, or $\frac{R_{TNK}}{3}$, as we are dealing with maximum power transfer to R_2. Under these conditions, the loaded Q of the secondary circuit will be 1/6 of the unloaded value. For other load conditions, the respective data should be entered into eq.(9).

f = 530 kHz
L_1 = 152 uH
Ca = 200 pF (assumed)
r = 30 ohms (assumed)
A = 1.555 x 10^5 ohms
Q_1Q_2A = 2.938 x 10^{10} ohms
K = 4.165 x 10^{-3}
Check:

$$\frac{R_P}{1+K^2Q_1Q_2} = 82.8 kohms >> \omega_r L_{eq} = 2.214 kohms$$

f = 1 MHz
L_1 = 152 uH
Ca = 200 pF (assumed)
r = 30 ohms (assumed)
A = 1.9 x 10^5 ohms
Q_1Q_2A = 2.334 x 10^{10} ohms
K = 5.663 x 10^{-3}
Check:

$$\frac{R_P}{1+K^2Q_1Q_2} = 113.1 kohms >> \omega_r L_{eq} = 2.387 kohms$$

f = 1.7 MHz
L_1 = 152 uH
Ca = 200 pF (assumed)
r = 30 ohms (assumed)
A = 4.068 x 10^5 ohms
Q_1Q_2A = 1.29 x 10^{10} ohms
K = 8.324 x 10^{-3}

Check:

$$\frac{R_P}{1+K^2Q_1Q_2} = 152 kohms >> \omega_r L_{eq} = 3.498 kohms$$

Radiation resistance

There is an effective radiation resistance associated with the total antenna length. For electrically short antennas this may only be a few ohms but can rise to several tens of ohms as the antenna length approaches one quarter wavelength. In Figure 2 this resistance is shown as rr.

Antenna reactance

An antenna has both a series capacitance and series inductance that makes up a net reactance that varies with wavelength. Capacitive reactance dominates if the antenna is shorter than one-quarter wavelength and inductive reactance dominates between one quarter and one-half wavelength at which capacitive reactance dominates again for the next quarter wavelength, etc. Normal antennas for crystal radios are significantly shorter than one-quarter wavelength so there is a large capacitive reactance in series with the induced voltage. The inductive reactance reduces the net reactance somewhat for short antennas. At quarter-wave resonance the reactance terms cancel and the antenna is purely resistive. In Figure 2 these are shown as Ca and La.

Antenna resistance

The antenna wire has electrical resistance that is mainly due to the skin effect of the conductor since the frequency is high. This resistance is typically a few ohms and is often small in comparison with the ground resistance. In Figure 2 the antenna wire resistance is shown as Ra.

Antenna tuner inductance

The first element typically found in a crystal radio is an adjustable inductor to neutralize the capacitive reactance of the antenna thus maximizing the power transfer from the antenna to the receiver input resistance. This is known as conjugate matching where the sum of the positive inductive reactance of the inductor and negative capacitive reactance of the antenna is zero. In Figure 2 this inductance is shown as Ls for series inductance.

Crystal radio input resistance

The crystal radio generally appears as a pure resistance at resonance with a received signal. For optimum power transfer this resistance should be equal to the sum of the ground plus antenna plus radiation resistance. The ground resistance typically dominates this equation. The input resistance of the radio can be adjusted via taps on the tuning inductor or other matching network. For best performance it is important to match the radio to the antenna and ground system. Failure to do this results in an underperforming radio.

Ground resistance

The ground resistance completes the circuit back to the rather nebulous point that is the effective reference for the radio frequency signal. Ideally, this resistance is less than 10 ohms but typically is in the 10 to 100 ohm range.

A single short wire antenna is unbalanced and a low-impedance ground is required to complete the RF circuit. It is easy to set up a simple wire antenna that can work well but it is a challenge to construct a low-impedance ground. Without a low-impedance ground much of the signal picked up by the

---

Some experimental results

Two coils, L_1 and L_2, were wound on 4.5” diameter styrene forms using 660/46 Litz wire. L_1 measured 152 uH and L_2, 222 uH. A two-gang 475 pF variable capacitor with bakelite insulation was used to tune L_1. L_2 was tuned with a 480 pF variable capacitor with ceramic insulators.
Unloaded Qs for each of the tuned circuits were measured at three frequencies. Accordingly, the corresponding RF losses were calculated. Data is tabulated below.

f	Q_1	Q_{2UL}	Q_2	R_{Pcalc}	$R_{TNKcalc}$	C	C_2
530kHz	700	810	270	354.32 kohms	598.816 kohms	450 pF	406 pF
1 MHz	585	630	210	558.70 kohms	878.766 kohms	100 pF	114 pF
1.7 MHz	300	318	106	487.072 kohms	754.065 kohms	31 pF	39.5 pF

C : two-gang 475 pF variable capacitor with bakelite insulation
C_2 : 480 pF variable capacitor with ceramic insulation
Q_{2UL} : unloaded Q of L_2-C_2 combination
Q_1, Q_2 : defined in the text
R_P, R_{TNK}: defined in the text

Using the tabulated data, values for the optimum coupling coefficient K will be calculated for a working crystal set.

$$R_{s1} = (\omega_r L_{eq})^2 \left(\frac{N^2 L_2}{L_1}\right)\left(\frac{Q_1 Q_2}{R_P}\right)$$

$$= (\omega_r L_{eq})^2 \left(\frac{N^2 L_2}{L_1}\right)\left(\frac{1}{\omega_r L_1}\right)\left(\frac{R_2}{\omega_r L_2}\right)$$

$$= \left(\frac{L_{eq}}{L_1}\right)^2 N^2 R_2$$

$$= \left(\frac{2Ca + C}{Ca}\right)^2 N^2 R_2$$

or:

$$N^2 R_2 = \left(\frac{Ca}{2Ca + C}\right)^2 R_{s1}$$

Substituting this equivalence into eq.(13):

$$P_2 = \frac{{E_a}^2}{8R_{s1}} = P_{MAX}$$

according to eq.(10).

P_{MAX} is then dissipated by N^2R_2 and by consequence, this power is delivered to the secondary load.

antenna will be wasted and the crystal radio will perform poorly if at all. Much frustration from poor performance of a crystal radio set is often caused by a high-impedance ground. An ideal situation is to connect the crystal radio set to a metal water pipe (either old-fashioned galvanized or preferably copper). The long exposure length to the earth produces a low-impedance ground in the low tens of ohms. The impedance is even lower if the soil is very organic and wet from a recent rain. Dry sandy soils are the worst. One common error is connecting to a metal water pipe assuming that there are no plastic pipe lengths involved.

If a metal water pipe is not available then it might be tempting to connect to the ground system of the electrical distribution in the house or building. For safety considerations I do not recommend this. In the vast majority of cases it may be alright but you never know what errors or faults may be in the system that could result in a nasty or lethal surprise. Even if safety were not an issue, because of inductance the impedance of the ground system at radio frequencies will be much higher than the impedance at 60 Hz. So, this practice is dangerous at worst or is not likely to work well at best. There are basically only two choices for constructing your own ground. One is to purchase several standard eight foot ground rods and laboriously install them spaced six or more feet apart. This can make a decent ground but is very difficult (impossible is probably more accurate) without the right equipment for installation. An easier alternative that is fine for RF (although not recommended for 60 Hz systems) is to bury a length of bare copper wire or pipe perhaps six to twelve inches below the surface of the ground. Deeper is better but the incremental improvement is probably not worth the extra labor. The total length should be fifty feet or more and parallel runs spaced several feet apart can be used to accumulate effective length. AM broadcast stations use such a system with many buried radial wires from the tower. Each wire is typically a quarter wavelength and the total wire length of all the radials is often over a mile. Such extreme measures are necessary for a high power transmitter because otherwise many thousands of watts could be wasted. The much more modest system described is fine for home use and a point of diminishing returns is quickly reached.

Figure __ shows the theoretical radiation resistance for a single wire antenna that is shorter than one-half wavelength calculated under the assumption that there are no nearby objects that interfere with the impedance. For most situations the antenna is not very far above the ground and the actual resistance will likely be somewhat less than shown. But this is a reasonable model to use for analysis.

Figure __ shows the theoretical series capacitive reactance

How much signal can be received? With a variety of impedance transformations we can obtain most any voltage or current we want but what really matters is how much power we can receive. It is power that will make audible sounds in our earphones. Assuming that there are no losses and that the load consists of an inductor with the same magnitude of reactance as the antenna capacitive reactance in series with a resistive load equal to the radiation resistance of the antenna, then the power delivered to the resistive load is:

P = (V/m/2 * L)^2 /(40*pi^2*L^2*FMHz^2 /300^2)

= (V/m)^2 *57 / FMHz^2

$$R_{s1} = \left(\omega_r L_{eq}\right)^2 \frac{K^2 Q_1 Q_2}{R_P} \quad \ldots(12)$$

Then:

$$\frac{\omega_r L_{eq} K^2 Q_2}{R_{s1}} = \frac{R_P}{\omega_r L_{eq} Q_1}$$

$$= \frac{L_1}{L_{eq}}$$

$$= \frac{Ca}{2Ca + C}$$

Then, we obtain:

$$E_2{}' = \frac{E_a}{2}\left(\frac{Ca}{2Ca + C}\right)$$

The power dissipated by N^2R_2 is:

$$P_2 = \frac{\left(E_2{}'\right)^2}{2N^2 R_2}$$

$$= \frac{\left(\frac{E_a}{2}\right)^2 \left(\frac{Ca}{2Ca + C}\right)^2}{2N^2 R_2} \quad \ldots(13)$$

Eq.(12) can be written as follows:

$$E_2' = \frac{E_{Leq}}{j\omega_r L_1 + N^2 R_2} N^2 R_2$$

In crystal sets, $N^2 R_2 << \omega_r L_1$, due to the loose coupling between L_1 and L_2. Then:

$$E_2' = \frac{E_{Leq}}{j\omega_r L_1} N^2 R_2$$

$$= \frac{E_{Leq}}{j\omega_r L_1}\left[\frac{(\omega_r L_1)^2}{R_{P1}}\right]$$

$$= \frac{E_{Leq}}{jR_{P1}}\omega_r L_1$$

Bearing in mind eq.(3):

$$E_2' = \frac{E_{Leq}}{j} K^2 Q_2$$

Substituting the value of E_{Leq} given by eq.(11) into the above expression we obtain:

$$E_2' = \frac{E_a}{2R_{s1}}\omega_r L_{eq} K^2 Q_2$$

We can recall that:

MATHEMATICAL MODEL OF A WIRE ANTENNA

by Kenneth A. Kuhn

www.kennethkuhn.com/.../mathematical_model_of_wire_antenna.pdf

Introduction

Although very simple in structure, the mathematical model of a wire antenna is extremely complicated. There are so many variables that it is basically impossible to accurately model. However, a reasonable model can be derived that provides useful information and understanding. That is essential for doing engineering. An imperfect model is far better than no model. We begin with the simplest possible model (that is always a good starting point!). Then, we refine the model to include more advanced concepts. We keep refining until we have accounted for all significant variables that provides useful information.

A wire antenna consists of some length of wire at some height over the earth. Such a structure will have some capacitance to the earth, some inductance, and some resistance. The resistance term is more complicated than just the simple ohmic resistance of the wire as it includes a series term known as radiation resistance. In very simple terms radiation resistance accounts for a flow of energy between the antenna and free space involving time-dependent electromagnetic fields not considered in a simple RLC circuit. Our initial model is based on frequencies much less than that where the antenna is one-quarter wavelength resonant. For typical crystal radio applications this is not a bad model as it is rare to have the luxury of enough space to construct an antenna approaching one-quarter wavelength in the AM broadcast band.

Capacitance

We start with the capacitance of the antenna to the earth. From the basic laws of physics, a capacitance will exist between any two conductors separated by insulation. Here, we are making the rather crude assumption that the earth is a good conductor. That turns out not to be bad as we will see that the reactance of this capacitance at frequencies in the AM broadcast band is large compared to the actual resistance of the earth. A brief but excellent derivation of the capacitance between a cylinder (i.e. the antenna wire) and a plane is provided in Reference 1. The summary result is:

$$\text{Capacitance per meter} = \frac{2 * \pi * \varepsilon 0}{\ln(2 * \text{height} / \text{diameter})} \quad \text{Eq. 1}$$

where height and wire diameter are in meters and ε0 is the permittivity of free space, 8.85 pF/meter. Note that this equation assumes that propagation time over the wire length is negligibly small. This equation also assumes that the earth is a perfect conductor.

Although there is an error when considering the composite antenna, there is no error for an incremental portion that satisfies the assumptions.

A quick example indicates that a 20 meter antenna of #14 wire (1.63 mm dia.) at a height of 3 meters would have a total capacitance of 135 pF at frequencies low enough so that the propagation time over the length is negligibly small. That is not true for broadcast band frequencies and the capacitance will be smaller as will be seen in the plots at the end.

Inductance

$$K = \sqrt{\frac{R_P + A}{Q_1 Q_2 A}} \qquad \ldots(9)$$

which gives the value of the coupling coefficient for maximum RF power transfer to the secondary.

Power calculations

The RF power delivered to the secondary load of Fig.1 will be at a maximum at resonance when eq.(6) is satisfied, this is, when r + $R_s = R_{s1}$. The maximum available power is then:

$$P_{MAX} = \frac{\left(\frac{E_a}{2}\right)^2}{2R_{s1}}$$

$$= \frac{E_a{}^2}{8R_{s1}} \qquad \ldots(10)$$

where E_a is the peak value of the voltage induced in the antenna.

The power delivered to the secondary load R_2 is the same as that dissipated by the coupled resistance N^2R_2. To compute this power we need the voltage across L_1 at resonance. This is the same as the voltage across L_{eq} in Fig.6.a. Then:

$$E_{Leq} = \frac{E_a}{2R_{s1}} j\omega_r L_{eq} \qquad \ldots(11)$$

From Fig.5 we obtain for the voltage across N^2R_2:

$$= L_1\left(2+\frac{C}{Ca}\right) \quad \ldots(7)$$

being

$$\omega_r^{\ 2} = \frac{1}{L_1 C}\left(\frac{Ca+C}{2Ca+C}\right)$$

Then:

$$\left(\omega_r L_{eq}\right)^2 = \frac{1}{L_1 C}\left(\frac{Ca+C}{2Ca+C}\right)L_1^{\ 2}\left(\frac{2Ca+C}{Ca}\right)^2$$

$$= \frac{L_1}{C}\left(2+\frac{C}{Ca}\right)\left(1+\frac{C}{Ca}\right)$$

Eq.(6) is now written as:

$$\frac{L_1}{rC}\left(2+\frac{C}{Ca}\right)\left(1+\frac{C}{Ca}\right) = \frac{R_P}{K^2 Q_1 Q_2 - 1} \quad \ldots(8)$$

Letting $A = \frac{L_1}{rC}\left(2+\frac{C}{Ca}\right)\left(1+\frac{C}{Ca}\right)$, eq.(8) takes the more compact form:

$$A = \frac{R_P}{K^2 Q_1 Q_2 - 1}$$

Solving for K we obtain:

On page 141 of Reference 1 the inductance of a wire over a plane is given as

$$\text{inductance per meter} = \frac{\mu 0}{2 * \pi} * \ln(2*\text{height} / \text{diameter}) \quad \text{Eq. 2}$$

where height and wire diameter are in meters and μ0 is the permeability of free space, 4*π*1E-07 henries/meter. Equation 2 simplifies to

inductance per meter = (0.2 uH) * ln(2*height / diameter) Eq. 3

A quick example indicates that a 20 meter antenna of #14 wire (1.63 mm dia.) at a height of 3 meters would have a total inductance of 33 uH if propagation time is negligibly small. Note that Equation 2 assumes that the current is uniform over the length of the wire and that propagation time is negligibly small. The first part of the assumption can not be true for our antenna as the far end is an open circuit. The equation also assumes that the earth is a perfect conductor. The typical error for our use in antennas for crystal radios falls into the category of not bad. We are not going to worry about it as this is just a starting point and we are going to make refinements. Although there is an error when considering the composite antenna, there is no error for an incremental portion that satisfies the assumptions.

Resistance

The resistance of a wire antenna is the sum of the ohmic resistance including skin effect and the radiation resistance. The resistance of #14 wire including skin effect at a frequency of 1 MHz is around 0.05 ohms per meter. Thus a 20 meter antenna would have a total resistance of 1 ohm. This assumes that the current is uniform over the length which can not really be true since the far end is an open circuit. However, this is not going to be an issue for us as the resistance of the ground will be significantly larger.

The ohmic resistance represents a power loss. Radiation resistance becomes a factor when we consider the coupling of energy between the antenna and electromagnetic fields of free space. The concept is easy to understand in the case of a transmitter which delivers power to the antenna (perhaps significant power in the case of a 50 kW transmitter!) but with efficient design the heating of the antenna is very small –i.e. the power is radiated as an electromagnetic wave instead of being converted to heat –i.e. power is coupled to free space. The transmitter delivers the power to the antenna as some voltage and some in-phase current. Thus, the antenna appears to have a resistance. That resistance is known as the radiation resistance. Unlike ohmic resistance, it does not represent a power loss –it represents a power coupling to free space instead. There is no magic or optimum value for antenna impedance. However, values between single digit ohms and several hundred ohms are the easiest to interface to and so are preferred.

It is not as easy to visualize radiation resistance when the antenna is used for reception. However, the concept of reciprocity applies. Thus, however an antenna appears in the transmitting case, it appears identically in the receiving case. The calculation of radiation resistance is very complicated at the easiest. At this point in our model development we have no way to determine what the radiation resistance of our antenna might be. However, the incremental capacitance and inductance are essential for the next level model.

$$= (\omega_r L_{eq})^2 \left(\frac{1 + K^2 Q_1 Q_2}{R_P} \right)$$

$$= \frac{(\omega_r L_{eq})^2}{R_P} + (\omega_r L_{eq})^2 \frac{K^2 Q_1 Q_2}{R_P}$$

Let $R_s = \frac{(\omega_r L_{eq})^2}{R_P}$ and $R_{s1} = (\omega_r L_{eq})^2 \frac{K^2 Q_1 Q_2}{R_P}$.

R_s is the series term due to R_P and R_{s1}, that coming from the coupled resistance N^2R_2. Now, maximum RF power transfer to N^2R_2 occurs when:

$$r + R_s = R_{s1}$$

or when:

$$r + \frac{(\omega_r L_{eq})^2}{R_P} = (\omega_r L_{eq})^2 \frac{K^2 Q_1 Q_2}{R_P}$$

or

$$\frac{(\omega_r L_{eq})^2}{r} = \frac{R_P}{K^2 Q_1 Q_2 - 1} \qquad \ldots(6)$$

From part I of this study we know that:

$$L_{eq} = \frac{L_1}{1 - \omega_r^2 L_1 C}$$

If $R_T >> \omega_r L_{eq} = \dfrac{\omega_r L_1}{1-\omega_r^{\,2} L_1 C}$ (please refer to part I of this study) we can redraw the equivalent circuit of the tuner at resonance as indicated by Fig.6.a, and, by virtue of the above inequality, the resonant frequency ω_r will still be given by:

$$\omega_r^{\,2} L_1 C\left(\frac{2Ca + C}{Ca + C}\right) = 1 \qquad \ldots(5)$$

Applying a parallel-to-series transformation, the equivalent circuit of Fig.6.b is obtained.

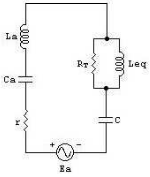

Fig.6.a Equivalent circuit of the tuner when $R_T >> \omega_r Leq$

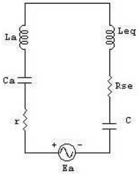

Fig.6.b Equivalent circuit after transformation

The series transformed resistance R_{se} is given by:

$$R_{se} = \frac{(\omega_r L_{eq})^2}{R_T}$$

A finite propagation time model

This model is based on well-developed transmission line theory and treats the wire antenna as a lossy unterminated transmission line where the loss is electromagnetic radiation instead of heat. This model builds on the previous equations for capacitance and inductance and gives us a realistic picture of what goes on. In general, because there are a lot of variables that we are unable to quantify the accuracy of this model is not great. However, the model can be tweaked to provide a good fit to a specific scenario. That is what makes the model useful. It is great at providing us general information for understanding. Understanding what is going on can be more valuable than knowing specific data accurately.

The impedance looking into a transmission line is

$$Z = Z0 * \frac{(ZT/Z0) + \tanh(\alpha + j\theta)}{1 + (ZT/Z0) * \tanh(\alpha + j\theta)} \qquad \text{Eq. 4}$$

where:
Z is in ohms and is in general complex.
Z0 is the characteristic impedance of the line –often taken as real but could be complex.
ZT is the termination impedance at the end of the line and could be complex.
α is the loss over the length of the line.
θis the phase angle on the line and is equal to 2*π*length/wavelength.
j is the square root of -1.

The form of the equation I have chosen to use includes length in the α and θ terms as a matter of convenience. Other forms of the equation have the length as a multiplier on the inside of the tanh term.

If the terminating impedance is an open circuit as it will be for our wire antenna then we can write Equation 4 simply as

Z = Z0 / [tanh(α + jθ)] Eq. 5

For reference:

ex –e-x
tanh(x) = ----------- Eq. 6
ex + e-x

The Euler relations are:

ejθ = cos(θ) + j*sin(θ) Eq. 7

e-jθ = cos(θ) – j*sin(θ) Eq. 8

These relations let us write Equation 5 as

eα+jθ + e-α-jθ
Z = Z0 * ----------------
eα+jθ–e-α-jθ

eα* [cos(θ) + j*sin(θ)] + e-α* [cos(θ) –j*sin(θ)]
= Z0 * ---
eα * [cos(θ) + j*sin(θ)] –e-α* [cos(θ) –j*sin(θ)]

(eα+ e-α)*cos(θ) + j*(eα–e-α)*sin(θ)
= Z0 * -- Eq. 9
(eα–e-α)*cos(θ) –j*(eα+ e-α)*sin(θ)

Let $Q_2 = \frac{R_2}{\omega_r L_2}$. Then:

$$R_{P1} = \frac{\omega_r L_1}{K^2 Q_2} \qquad \ldots(3)$$

Next, we compute the equivalent resistance R_T of the parallel combination of R_P and R_{P1}. It is given by:

$$R_T = \frac{R_P R_{P1}}{R_P + R_{P1}}$$

Substituting R_{P1} by its equivalent given by eq.(3):

$$R_T = \frac{R_P\left(\frac{\omega_r L_1}{K^2 Q_2}\right)}{R_P + \frac{\omega_r L_1}{K^2 Q_2}}$$

$$= \frac{R_P \omega_r L_1}{\omega_r L_1 + K^2 Q_2 R_P}$$

Letting $Q_1 = \frac{R_P}{\omega_r L_1}$ (unloaded Q of L_1-C tank) we obtain:

$$R_T = \frac{R_P}{1 + K^2 Q_1 Q_2} \qquad \ldots(4)$$

Up to this point, in the tuner side we have the equivalent circuit depicted in Fig.5.

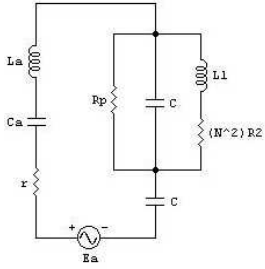

Fig.5 Equivalent circuit of the tuner with coupled secondary load

The series-coupled resistance N^2R_2 can be transformed into a resistance R_{P1} in parallel with R_P. Using the known series-to-parallel "loss resistance" transformation we get:

$$R_{P1} = \frac{(\omega_r L_1)^2}{N^2 R_2} \qquad \ldots(2)$$

Although Equation 9 looks complicated it really consists of just three components and it is simple to substitute values and reduce it to a simple real plus imaginary result.

For reference: Eq. 10

$$\frac{A + jB}{C + jD} = \frac{A + jB}{C + jD} * \frac{C - jD}{C - jD} = \frac{AC + BD}{C2 + D2} + j\frac{BC - AD}{C2 + D2}$$

Equation 10 gives us a basic method to evaluate the division of complex numbers.

When the series and shunt losses are low the characteristic impedance of a transmission line is

Z0 = sqrt(inductance per unit length / capacitance per unit length) Eq. 11

We substitute Equations 1 and 2 into Equation 11 to obtain

$$Z0 = \text{sqrt}\left[\frac{\left[\frac{\mu 0}{2 * \pi} * \ln(2*\text{height} / \text{diameter})\right]}{\left[\frac{2*\pi*\varepsilon 0}{\ln(2*\text{height} / \text{diameter})}\right]}\right] \qquad \text{Eq. 12}$$

which simplifies to

Z0 = 60.0 * ln(2*height / diameter) Eq. 13

Thus, we make use of our previous work on the simple model.

The free-space wavelength in meters for a given frequency in MHz is

$\lambda = 300 / \text{FMHz}$ Eq. 14

The propagation velocity on wire antennas is about 95 percent of free space so we use

$\lambda = 285 / \text{FMHz}$ Eq. 15

$\theta = 2*\pi*\text{FMHz}*\text{length} / 285$ Eq. 16

So far this has not been too complicated. Now, we delve into the challenge of determining the remaining constant, α. This factor, which is a function of frequency, will determine the antenna impedance characteristics. There is no way to calculate this factor. Instead, we will use available information to fit an approximate equation to it. We begin by noting that the free space impedance of a half-wave dipole antenna is 73 ohms. Our wire antenna is only half of a dipole so the free space impedance will be nominally 37 ohms (rounded half of 73) at a frequency where the wire length is an electrical one-quarter wavelength.

However, our antenna is very close to the ground so the result must be modified. Reference 2 shows a theoretical plot of the impedance of a half-wave dipole over a perfectly conducting plane as a function of height. This plot is reproduced below.

In the transformer circuit, the impedance coupled to the primary side consists of a capacitance C_2 / N^2 in parallel with a resistance N^2R_2. Both are in parallel with the magnetizing inductance K^2L_1 (please see Fig.4.a).

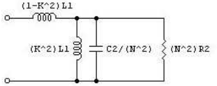

Fig.4.a Equivalent circuit as seen from the primary side

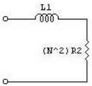

Fig.4.b Reduced equivalent at $\omega = \omega_r$

K^2L_1 and C_2 / N^2 resonate at a frequency

$$\omega_0 = \frac{1}{\sqrt{K^2 L_1 \left(\frac{C_2}{N^2} \right)}}$$

$$= \frac{1}{\sqrt{L_2 C_2}} = \omega_r$$

as

$$N^2 = K^2 \frac{L_1}{L_2}.$$

The equivalent circuit of Fig.4.a reduces to that of Fig.4.b, taking into account that for crystal set use, normally K<<1.

In this case, the net parallel resistance to be coupled to the primary is:

$$R_2 = \frac{R_{TNK}\left(\frac{R_{TNK}}{2}\right)}{R_{TNK} + \left(\frac{R_{TNK}}{2}\right)}$$

$$= \frac{R_{TNK}}{3} \qquad \ldots(1)$$

We shall work on this later.

Some circuit equivalents

In Fig.1, let´s replace L_1 and the coupled secondary circuit by the equivalent shown in Fig.3.a, which in turn can be replaced by the transformer circuit shown in Fig.3.b.

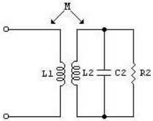

Fig.3.a Magnetically coupled circuits

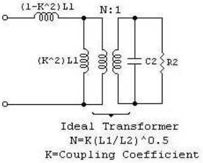

Fig.3.b Equivalent transformer circuit

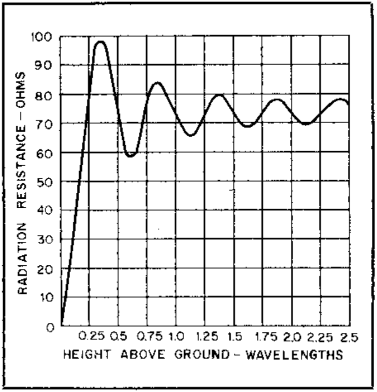

Figure 1: Theoretical impedance of half-wave dipole over perfect ground plane

Reference 2 describes that the impedance of a half-wave dipole over a realistic ground is very little affected by height over about one-half wavelength and the impedance soon attains the 73 ohm free space value.

We will use a simplified exponential to provide us with an approximate impedance of our antenna at the one-quarter wave resonant frequency. We first need a relation that converts the physical height of our antenna to height in wavelengths at the quarter-wave resonant frequency. This is provided by the following relation.

h/λ= height / wavelength = height * FMHz / 300 Eq. 17

$R = 37 * (1 - e-k1*h/\lambda)$ Eq. 18

The k1 value is chosen to be 7 because that provides the best least squares fit over the first 0.2 wavelengths of height. We can write this as

$R = 37 * (1 - e-0.02333*height*FMHz)$ Eq. 19

Equation 19 provides us with a target to set the appropriate value for α. At the frequency where the line length is one-quarter wave resonant, the cosine term in Equation 9 goes to zero and the sine term goes to one. This produces a real result with the imaginary part zero. It should be noted for this case that:

$A = 0$

$B = e\alpha - e-\alpha$

$C = 0$

$D = 2$

Using Equation 10 the above results in a resistance of

$R = Z0 * (e\alpha - e-\alpha) / 2 = Z0 * \sinh(\alpha)$ Eq. 20

We invert Equation 20 to obtain the required value for a for a given value of R as follows

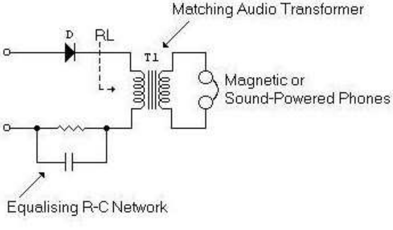

Fig. 2 Load coupled to the Tuggle tuner

Usually, it is assumed that optimum RF power transfer occurs when the antenna-ground system resonance resistance r is matched to the unloaded-secondary resonance parallel resistance R_{TNK}, with the diode detector´s input resistance matched to this combination. Thus,

$$\frac{1}{R_{OPT}} = \frac{1}{R_{TNK}} + \frac{1}{R_{TNK}} + \frac{1}{\left(\frac{R_{TNK}}{2}\right)}$$

or

$$R_{OPT} = \frac{R_{TNK}}{4} .$$

This is the overall parallel RF resistance of the secondary tank under matched conditions, and suggests that the unloaded Q of this tank circuit has been reduced to ¼ of its value.

Analysis of the Tuggle Front End – Part II

We shall now consider the Tuggle tuner delivering power to a load. First, we must account for the parallel RF losses of the unloaded tuned circuits. Let R_P represent the losses of the tank circuit comprising L_1 and C, and R_{TNK} those of L_2 and C_2 (please see Fig. 1).

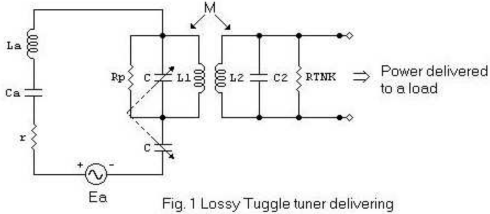

Fig. 1 Lossy Tuggle tuner delivering power to the secondary load

The load consists of a diode detector D in series with an audio load R_L (usually an audio transformer matching a pair of 2k ohms DC resistance magnetic headphones or low-impedance sound powered phones to the detector) and is coupled to the tuner via the magnetic coupling existing between L_1 and L_2, being M the mutual inductance of the coils. The secondary is tuned to the same radian frequency as the primary. An schematic for the load can be seen in Fig. 2.

α= sinh-1(R/Z0) = ln[(R/Z0) + sqrt[(R/Z0)2 + 1)]] Eq. 21

Before continuing, let us review where we are at. We know Z0 from the height and diameter of the antenna wire. We know a target resistance, R, for the antenna impedance at the one-quarter wave resonant frequency based on the height. Equation 21 lets us calculate α for that frequency. We need to expand this so that we can calculate the appropriate α for any frequency.

We make use of the following general knowledge about wire antennas to complete our model. At frequencies well below the one-quarter wave resonance the resistive component of the antenna impedance is practically zero. At frequencies where the antenna is many quarter wavelengths long the resistive component of the antenna impedance converges to Z0. This suggests that we have a basic scale factor on α and a multiplier based on the number of quarter wavelengths are on the line. This leads us to Eq. 22

α= k0 * Z0 * 37 * (1 –e-0.02333*height*FMHz) * (F / Fq)
where:

k0 is a constant to achieve the target quarter-wave resonant impedance
Z0 is the characteristic impedance of the antenna
height is in meters
F is the frequency in MHz

Fq = (71.25 / length) and is the frequency in MHz where the antenna is one-quarter wave resonant

The only unknown in Equation 22 is k0. We calculate k0 when F is equal to Fq. The procedure is as follows:

First we set FMHz = Fq from the above relation

R = 37 * (1 –e-0.02333*height*FMHz) Eq. 19

Z0 = 60.0 * ln(2*height / diameter) Eq. 13

α= sinh-1(R/Z0) = ln[(R/Z0) + sqrt[(R/Z0)2 + 1)]] Eq. 21

k0 = α/ (Z0 * R) Eq. 23

Studies done with this equation indicated very good results as long as the frequency was greater than the quarter-wave resonance. For lower frequencies this model gave very poor results –indicating significantly higher impedance than reality and even a negative slope in some cases. A modifier to the model was needed to highly attenuate αfor low frequencies only. The factor, (1 –1/(1 + 0.14*(F/Fq)2+ 1.0*(F/Fq)3 + 0.79*(F/Fq)4 + 50*(F/Fq)24)), was determined to significantly improve results. The factors were derived from a least squares fit to an alternative equation for short dipoles provided by reference 3 –shown below modified for only half a dipole.

R = 40*π2*(L/λ)2 Eq. 24

where L is the antenna length and λ is the wavelength, both in meters. This equation is reported to give good results up to about 0.2 wavelengths. A plot showing the fitted model to Equation 24 is shown below. Note that the model has an excellent fit to the theoretical "correct" model for wavelengths shorter than 0.2. The "correct" model under predicts the resistance for longer wavelengths and the effect of the 24th power term comes into play to bring the resistance close to the theoretical 37 ohms at 0.25 wavelengths. This is not exactly what goes on but is a simple to work with fit.

The circuit will tune up to:

$$f_{MAX} = f_{MIN}\left(\frac{C_{eqMAX}}{C_{eqMIN}}\right)^{\frac{1}{2}}$$

$$= 2.128MHz$$

If we use for C a variable capacitance with
C_{MIN} = 20 pF and C_{MAX} = 365 pF, then
C_{eqMIN} = 38.18 pF and C_{eqMAX} = 494.20 pF, giving for the required inductance L a value of 182.46 uH. The circuit will tune up to f_{MAX} = 1.906 MHz.

Acknowledgements: Special thanks are given to Ben Tongue for his comments on the manuscript and for encouraging further mathematical analysis of the circuit regarding bandwidth variation with frequency, which will be done shortly.

After some algebraic manipulation we obtain:

$$\omega_r^{\,2} LC\left(\frac{2Ca+C}{Ca+C}\right)=1 \qquad \ldots(3)$$

The equivalent capacitance resonating with L is:

$$C_{eq}=C\left(\frac{2Ca+C}{Ca+C}\right)$$

Clearly, $C_{eq}>C$.

Following is a numerical example illustrating the use of the above results.
Let C be a variable capacitance with
C_{MIN} = 20 pF and C_{MAX} = 475 pF.
Let also C_a be 200 pF.
Then, C_{eq} varies between
C_{eqMIN} = 38.18 pF and C_{eqMAX} = 615.74 pF.

If we wish to tune the MW broadcast band starting at 530 kHz, then the required inductance L will be:

$$L=\frac{1}{\omega_{rMIN}^{\;\;2} C_{eqMAX}}$$

$$=146.45uH$$

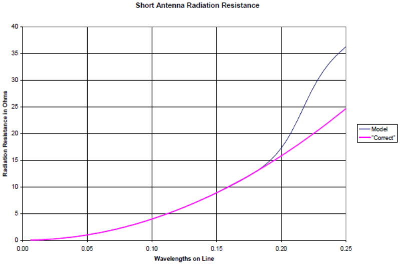

Figure 2: Model fit to short antenna radiation resistance

The complete equation for α is (using n = F/Fq) is shown below. Eq. 25

α= k0*Z0*37*(1 –e-0.02333*height*F)*n*

                    1
(1 - --)
     1 + 0.14*n2 + 1.0*n3 + 0.79*n4 + 50*n24

To calculate the approximate impedance of the antenna for any frequency we first compute the αfactor for the frequency using Equation 25. Then we compute the positive and negative exponentials using α. Next, we compute theta using Equation 16 and then the cosine and sine terms. We substitute these values into Equation 9 and use Equation 10 to solve. This is best done in a spread sheet for a range of frequencies and the results plotted. A wide range plot showing multiple quarter-wave resonances for a wire antenna is shown in Figure 3.

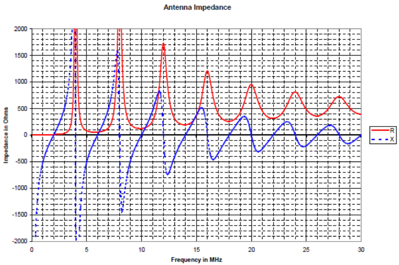

Figure 3: 2 MHz resonate antenna 5 meters off ground

The computed impedance at 2 MHz of this antenna is 7.5 ohms. The impedance at the third harmonic (6 MHz) is 55 ohms –for both resonant frequencies the reactive part is zero. It should be observed that the impedance is at a minimum and purely resistive at odd harmonics of the quarter-wave resonance. The impedance is very high at even harmonics. This figure should be viewed as illustrative rather than accurate. Although the pattern of impedance variations is true, the magnitude could vary considerably depending on actual ground conditions and the proximity of other structures. It is true that the impedance variations become smoother as the number of quarter-waves increases. Only a minimal attempt has been made to correctly model that for this model as the primary interest is in the sub quarter-wave region. More work will be done on the high frequency region at a later date. It is expected that the required effect can be obtained by the appropriate exponent on the n term that immediately follows the exponential in Eq. 25 –but some kind of short power series might be required instead. Example Antennas

This is, when:

…(2)

$$\omega\left(La+\frac{L}{1-\omega^{2}LC}\right)-\frac{1}{\omega\left(\frac{CaC}{Ca+C}\right)}=0$$

which is satisfied at certain radian frequency ω_r.
At this frequency, the L-C tank circuit behaves as an equivalent inductance

$$\frac{L}{1-\omega_r^{\,2}LC}$$

Usually, L is much greater than L_a for antennas used in crystal set work. Then,

$$La<<\frac{L}{1-\omega_r^{\,2}LC}$$

Equation (2) can be written as:

$$\frac{\omega_r L}{1-\omega_r^{\,2}LC}-\frac{1}{\omega_r\left(\frac{CaC}{Ca+C}\right)}=0$$

We can then write:

$$\frac{\omega_r L}{1-\omega_r^{\,2}LC}=\frac{Ca+C}{\omega_r CaC}$$

$$I = \frac{Ea}{r + j\left(\omega La - \frac{1}{\omega Ca}\right) + \bar{Z} - \frac{j}{\omega C}} \qquad \ldots(1)$$

$$\bar{Z} = \frac{1}{\bar{Y}}$$

where:

$$\bar{Y} = \frac{1}{j\omega L} + j\omega C$$

$$= j\left(\omega C - \frac{1}{\omega L}\right)$$

$$= j\left(\frac{\omega^2 LC - 1}{\omega L}\right)$$

$$\therefore \qquad \bar{Z} = \frac{1}{j}\left(\frac{\omega L}{\omega^2 LC - 1}\right)$$

or:

$$\bar{Z} = j\left(\frac{\omega L}{1 - \omega^2 LC}\right)$$

Then, I = I_{MAX} when:

$$\left(\omega La - \frac{1}{\omega Ca}\right) + \left(\frac{\omega L}{1 - \omega^2 LC}\right) - \frac{1}{\omega C} = 0$$

The following plots show the typical impedance for several wire antennas. In each case #14 AWG wire was used and the solid line is the radiation resistance and the dotted line is the reactance. In all cases the radiation resistance is very low. The key feature is the capacitive reactance of the antenna.

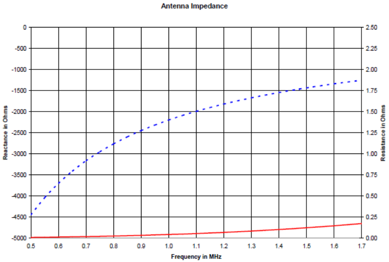

Figure 4: Antenna length = 10 m, height = 3 m

This is about as short as an antenna can be and still be practical. This antenna appears as 73 pF.

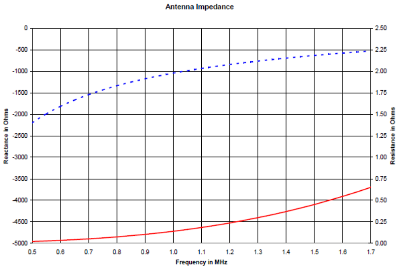

Figure 5: Antenna length = 20 m, height = 3 m

This antenna appears as 145 pF at low frequencies and 184 pF at high frequencies. The change in capacitance with frequency is because the reactance of the series inductance cancels part of the capacitive reactance.

ANALYSIS OF THE TUGGLE FRONT END

By Ramon Vargas Patron

http://www.inictel-uni.edu.pe/index.php?option=com_content&view=article&id=235&Itemid=152

Part I:

This article analyzes the Tuggle tuner, of common use in high-performance DX crystal sets. An equivalent circuit for the antenna-ground system with the tuner connected is shown in Fig. 1 below. It must be recognized that there is some stray capacitance of the rotor and frame of the two-gang variable capacitor to ground. This should be shown as a fixed capacitor across the bottom variable capacitor. Its presence will reduce the maximum frequency to which the circuit will tune. However, in the present analysis this stray capacitance is neglected.

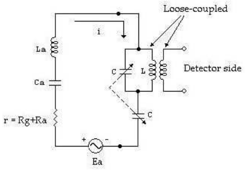

-Ea is the voltage induced in the antenna by the radio wave.
-La is the equivalent series inductance of the antenna.
-Ca is the equivalent series capacitance of the antenna.
-r accounts for the antenna radiation resistance Ra and ground losses Rg.

Fig. 1 The Tuggle front end connected to an antenna-ground system

The mesh current is described by:

25	340	20	165	440	1400	30	50
25	375	20	165	300	1700	26	27

25	220	20	200	700	550	130	335
25	300	20	200	585	1100	35	72
25	375	20	200	300	1700	24	20

50 ohm / 150 uH High resistivity Earth Case sensitivities

Ohm	pF	uH	uH	Qu	kHz	cpl	tank
Rg	Ca	La	L1		f	C2	C1
50	220	20	150	475	550	128	476
50	300	20	150	400	1100	34	108
50	375	20	150	200	1700	24	35
50	220	20	150	700	550	96	490
50	260	20	150	640	750	53	256
50	300	20	150	585	1100	28	113
50	340	20	150	440	1400	22	65
50	375	20	150	300	1700	19	39
50	220	20	125	700	550	110	595
50	300	20	125	585	1100	31	139
50	375	20	125	300	1700	21	49

Kevin Smith
09/2011

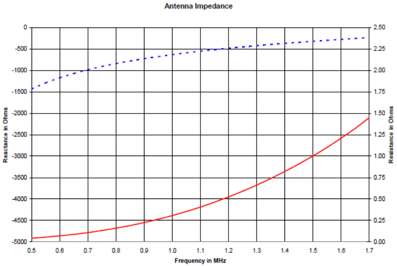

Figure 6: Antenna length = 30 m, height = 3 m

This antenna appears as 220 pF at low frequencies and 375 pF at high frequencies.

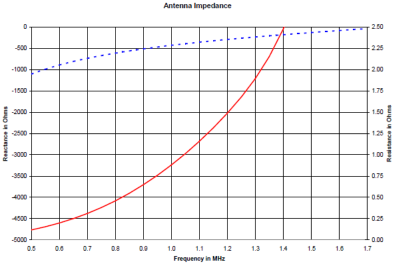

Figure 7: Antenna length = 40 m, height = 5 m

This antenna appears as 290 pF at low frequencies and is almost resonant at 1.7 MHz. This is about as long as an antenna needs to be for excellent reception. Note also that the antenna is higher. The increase in radiation resistance at higher frequencies is an indication that quarter wave resonance is being approached.

References:

1. Theory and Problems in Electromagnetics, Schaum's Outline Series, Joseph A.
Edminister, McGraw-Hill Book Company, 1979, pages 92-93, 141.

2. The ARRL Antenna Book, edited by Gerald L. Hall, The American Radio Relay
League, Inc. Newington, CT, 1984, Fourteenth edition, page 2-20.

3. Antennas and Transmission Lines, John A. Kuecken, MFJ Enterprises, Inc., 1996,
First edition, page 64.

Rg	Ca	La	L1		f	C2	C1
15	220	20	180	475	550	337	328
15	300	20	180	400	1100	61	63
15	375	20	180	200	1700	40	10
15	220	20	180	700	550	220	352
15	260	20	180	640	750	100	176
15	300	20	180	585	1100	49	73
15	340	20	180	440	1400	37	37
15	375	20	180	300	1700	32	17
15	220	20	130	700	550	313	511
15	300	20	130	585	1100	59	110
15	375	20	130	300	1700	38	30
15	220	20	205	700	550	195	302
15	300	20	205	585	1100	45	61
15	375	20	205	300	1700	30	13

25 ohm / 165 uH "Standard" resistivity Earth Case sensitivities

Ohm	pF	uH	uH	Qu	kHz	cpl	tank
Rg	Ca	La	L1		f	C2	C1
25	220	20	165	475	550	213	396
25	300	20	165	400	1100	48	84
25	375	20	165	200	1700	32	22
25	220	20	150	700	550	164	462
25	300	20	150	585	1100	41	102
25	375	20	150	300	1700	27	31
25	220	20	165	700	550	151	416
25	260	20	165	640	750	76	213
25	300	20	165	585	1100	39	91

I strongly urge you to put down the Litz, set aside the silver-plated caps, store the holy-grail diodes and go do some research on your local ground. All the time, expense, and effort on the greatest "state-of-the-art" crystal receiver will be wasted if the ground is not attended to. If the resistivity of your earth is high in ohm-meters, get more metal in the ground or consider a counter-poise. The following table from LEM Instruments shows the relation between earth resistivity in ohm meters and actual earth resistance in ohms as a function of the earthing method used.

Models and Data:

10 ohm / 220 uH Lowest resistivity Earth Case sensitivities

Ohm	pF	uH	uH	Qu	kHz	cpl	tank
Rg	Ca	La	L1		f	C2	C1
10	220	20	220	475	550	442	228
10	300	20	220	400	1100	68	36
10	375	20	220	200	1700	44	-3
10	220	20	220	700	550	273	255
10	260	20	220	640	750	114	122
10	300	20	220	585	1100	55	47
10	340	20	220	440	1400	41	20
10	375	20	220	300	1700	36	5
10	220	20	200	700	550	304	287
10	300	20	200	585	1100	58	54
10	375	20	200	300	1700	38	7

15 ohm / 180 uH Low resistivity Earth Case sensitivities

Ohm	pF	uH	uH	Qu	kHz	cpl	tank

GROUNDS
by Jim Lux
http://home.earthlink.net/~jimlux/hv/grounds.htm

The following discussions are really applicable to line frequencies. For higher frequencies, dielectric loss must be considered, as well as skin effect. For lower frequencies (i.e. DC) consideration should be given to electrolytic effects such as polarization.

The dominant effect for a ground is the current distribution within the earth. The ground rod or wire itself typically has negligible resistance, as does the interface between the rod and soil. As such, the soil conductivity has a very large influence on the ground resistance.

Resistivity of different soils and resistance of a single rod. The rod, in the table below, is a standard rod 5/8 inch in diameter and 10 feet long (16 mm diam by 3m long) Data taken from IEEE Std 142-1991.

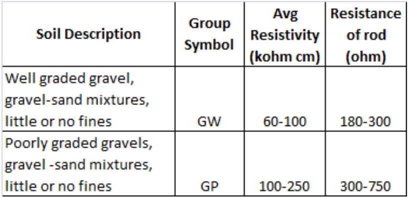

Soil Description	Group Symbol	Avg Resistivity (kohm cm)	Resistance of rod (ohm)
Well graded gravel, gravel-sand mixtures, little or no fines	GW	60-100	180-300
Poorly graded gravels, gravel -sand mixtures, little or no fines	GP	100-250	300-750

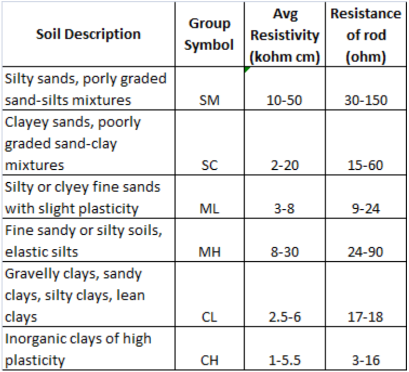

Soil Description	Group Symbol	Avg Resistivity (kohm cm)	Resistance of rod (ohm)
Silty sands, porly graded sand-silts mixtures	SM	10-50	30-150
Clayey sands, poorly graded sand-clay mixtures	SC	2-20	15-60
Silty or clyey fine sands with slight plasticity	ML	3-8	9-24
Fine sandy or silty soils, elastic silts	MH	8-30	24-90
Gravelly clays, sandy clays, silty clays, lean clays	CL	2.5-6	17-18
Inorganic clays of high plasticity	CH	1-5.5	3-16

Note 1) ohm M = kohm cm * 10
2) for the last two, clays, the resistivity is highly dependent on soil moisture.

Effect of Moisture Content on Soil Resistivity

The following table gives resistivity (ohm M) for three types of soil, for moisture contents from 2% to 24% by weight. Data taken from IEEE Std 142-1991.
Note: ohm M = kohm cm * 10

represent increasing earth resistance presented to the ATU from a very low 10 ohms to a fairly high 50 ohms. Actual earths can go up to two orders of magnitude higher. I chose L1 inductances such that the maximum needed capacitance would approach 500pFs, easily found on many variable caps. With respect to the coupling cap (C2, red curve), as the earth resistivity increases the needed capacitance declines. The tank cap value needs to resonate with the inductor and increases with smaller-value inductance. In searching for a 500pF max cap value, the choice of inductance needs to decrease with increasing earth resistance. Parameters other than Rg, L1 and C1-C2 have minor impact on the models.

Conclusions:

From the plots one can readily see that above 10 ohm earth the two variable capacitors on the ATU do not track well. As the earth resistivity increases, the worse the tracking. Most locations are not blessed with a low resistivity earth and this needs to be factored into the ATU design. There is little to be done, changing the design coil inductance will change the needed capacitance on BOTH C1 and C2, (larger L1 leads to smaller C1 and C2 for resonance and vice versa). A possible solution well worth trying is to use a dual-gang capacitor where one section has a different value than the other, the above plots give an easy way to decide the max values needed. Use the smaller gang on the coupling circuit and the larget on the tank. Experience tells us that ganging the capacitors on a "Tuggle" front end works well, but from the models I have to imagine one might squeeze a bit more performance by giving up the convenience of one-dial tuning on the ATU, especially where your earth has a fairly high resistance.

The best design concept will be to know first and critically your earth resistance. If this is just a guessed-at parameter then

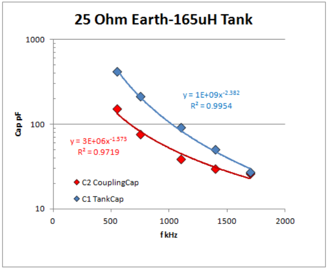

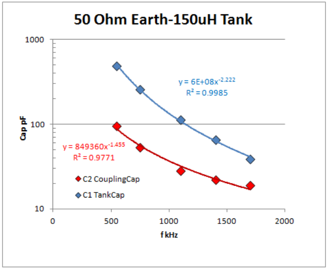

The above plots show the calculated capacitance versus frequency for four different models of Earth Resistivity / Tank Inductance. In each case I maintain the same log capacitance vs frequency scale for easy comparison. The cases modeled

---

Moisture content (% by weight)	Top Soil	Sandy Loam	Red Clay
2		185	
4		60	
6	135	38	
8	90	28	
10	60	22	
12	35	17	180
14	25	14	55
16	20	12	20
18	15	10	14
20	12	9	10
22	10	8	9
24	10	7	8

Effect of Temperature on Soil Resitivity
The following table gives variation in soil resistivity with temperature. The significant transition is at the freezing point, and above that the resitivity drops fairly linearly with temperature. I'm not sure about the double entry for 0 degrees.. perhaps it represents the variability at the transition temperature? Data taken from IEEE Std 142-1991.

Temperature (deg C)	Resistivity (kohm cm)
-5	70
0	30
0 (?)	10
10	8
20	7
30	6
40	5
50	4

Note: ohm M = kohm cm * 10

Calculation of Resistance to Earth
An expression, accurate to 15%, for a single 10ft (3m) rod 5/8" (16mm) in diameter is:
Rground(rod) = rho/335 ohms
where rho is in ohm-cm. (ohm-cm = ohm M *100)

Multiple rods

Multiple ground rods are often used particularly where high currents may be involved. The ground resistance is not simply the resistance of one rod, divided by the number of rods, unless the rods are very far apart. However, even though the ground resistance of the combination may not be all that much better than a single rod, the current is shared among the rods, reducing the current per rod. A guideline from IEEE Std 152-1991 to avoid "smoking rods" is:

Max Current(amps) = 34.8E3 * d * L / sqrt(rho * t)
where:
d is the rod diameter in meters
L is the rod length in meters
rho is the soil resistivity in **ohm meters**

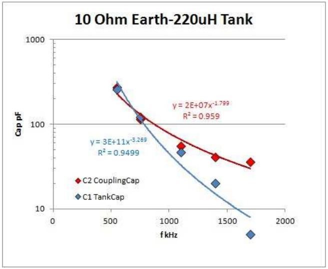

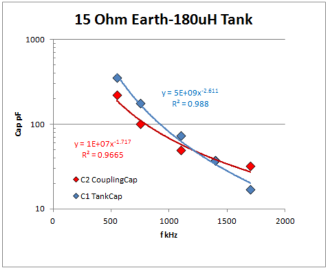

1) Frequency is chosen per your interest, I have taken models at F = 550-1100-1700 kHz.

2) Coil inductance one has control over when winding. I have chosen base values that resonate with about 400pF tank capacitance, with some sensitivities.

3) Unloaded Q I do not have. For modeling I have taken Vargas's measures for a 4.5" coil wound with 660/46 Litz wire. I assume this will be about as good a coil as one can wind. I also made a sensitivity for lower Q. Q is a function of frequency and I have used the appropriate value of Q to match the modeled frequency.

4) Series Resistance is basically what the earth and ground system will deliver plus a small 1-2 ohm contribution from Ra and Rr. I have modeled four cases 10, 15, 25 and 50 ohms. The first two cases reflect my needs on the Gulf Coast with low resistivity soil. The 25 ohm case is a "typical" or "standard" grround. Finally, many will have much higher resistivity soils and you need to know just how difficult things can get!

5) Series capacitance I based for my specific antenna on the model of Ken Kuhn (Mathematical Model of Wire Antenna). The capacitance of a 30m antenna 3m high he calculated to range from 220pF and low frequencies to 375pF and high frequencies. I input the correct capacitance to match the frequency modeled.

6) Series inductance I just input 20 uH every time.

The data and plots for four different scenarios follow:

t is the duration of the current in seconds (and is valid for short times only)

The text of IEEE Std 142-1991 claims that for 1 ft 5/8" rod (32 cm x 16 mm) this expression yields 116A in 2500 ohm cm soil and 58A for 10,000 ohm-cm soil, however I don't get these numbers.

Here's a table of some formulae to calculate ground resistance for combinations of rods. This table is taken from IEEE Std 142-1991, but it's cited there as coming from: Dwight, H.B.,"Calculation of resistance to ground", *AIEE Transactions*, vol 55, Dec 1936, pp 1319-1328.

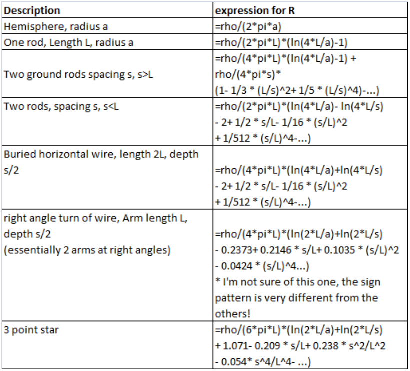

Description	**expression for R**
Hemisphere, radius a	=rho/(2*pi*a)
One rod, Length L, radius a	=rho/(2*pi*L)*(ln(4*L/a)-1)
Two ground rods spacing s, s>L	=rho/(4*pi*L)*(ln(4*L/a)-1) + rho/(4*pi*s)* (1- 1/3 * (L/s)^2+ 1/5 * (L/s)^4)-...)
Two rods, spacing s, s<L	=rho/(2*pi*L)*(ln(4*L/a)- ln(4*L/s) - 2+ 1/2 * s/L- 1/16 * (s/L)^2 + 1/512 * (s/L)^4-...)
Buried horizontal wire, length 2L, depth s/2	=rho/(4*pi*L)*(ln(4*L/a)+ln(4*L/s) - 2+ 1/2 * s/L- 1/16 * (s/L)^2 + 1/512 * (s/L)^4-...)
right angle turn of wire, Arm length L, depth s/2 (essentially 2 arms at right angles)	=rho/(4*pi*L)*(ln(2*L/a)+ln(2*L/s) - 0.2373+ 0.2146 * s/L+ 0.1035 * (s/L)^2 - 0.0424 * (s/L)^4...) * I'm not sure of this one, the sign pattern is very different from the others!
3 point star	=rho/(6*pi*L)*(ln(2*L/a)+ln(2*L/s) + 1.071- 0.209 * s/L+ 0.238 * s^2/L^2 - 0.054* s^4/L^4- ...)

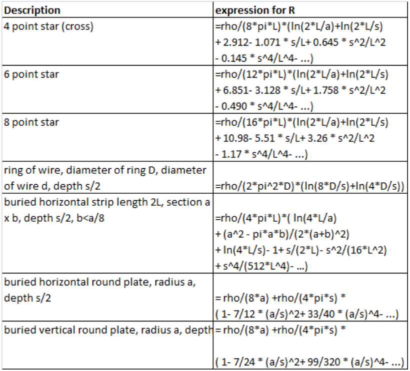

Description	expression for R
4 point star (cross)	=rho/(8*pi*L)*(ln(2*L/a)+ln(2*L/s) + 2.912- 1.071 * s/L+ 0.645 * s^2/L^2 - 0.145 * s^4/L^4- ...)
6 point star	=rho/(12*pi*L)*(ln(2*L/a)+ln(2*L/s) + 6.851- 3.128 * s/L+ 1.758 * s^2/L^2 - 0.490 * s^4/L^4- ...)
8 point star	=rho/(16*pi*L)*(ln(2*L/a)+ln(2*L/s) + 10.98- 5.51 * s/L+ 3.26 * s^2/L^2 - 1.17 * s^4/L^4- ...)
ring of wire, diameter of ring D, diameter of wire d, depth s/2	=rho/(2*pi^2*D)*(ln(8*D/s)+ln(4*D/s))
buried horizontal strip length 2L, section a x b, depth s/2, b<a/8	=rho/(4*pi*L)*(ln(4*L/a) + (a^2 - pi*a*b)/(2*(a+b)^2) + ln(4*L/s)- 1+ s/(2*L)- s^2/(16*L^2) + s^4/(512*L^4)- ...)
buried horizontal round plate, radius a, depth s/2	= rho/(8*a) +rho/(4*pi*s) * (1- 7/12 * (a/s)^2+ 33/40 * (a/s)^4- ...)
buried vertical round plate, radius a, depth	= rho/(8*a) +rho/(4*pi*s) * (1- 7/24 * (a/s)^2+ 99/320 * (a/s)^4- ...)

hv/grounds.htm - 18 March 2003 –
Jim Lux

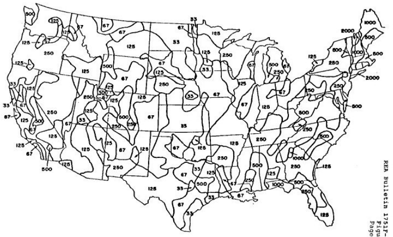

FIGURE 2

Estimated Average Earth Resistivity in U.S. (ohm-meters)

Modeling:

All the above discussion is great but... What is important really? The following section presents the results of applying different parameters to get a feel for the sensitivity of it all. To begin with I need explain my base assumptions. All the models are based on Kleijer's calculation spreadsheet which requires inputs for the following parameters:

Frequency
Coil inductance
LC circuit Q unloaded
* Complex impedance of antenna or
Series Resistance (Rg+Ra+Rr)
Series Capacitance Ca
Series Inductance La

* If the series values are given, the complex impedance is calculated.

of typical ranges in ohm meters for some different soils follows:

Loam	5	-	50	ohm meter
Clay	4	-	100	
Sand/Gravel	50	-	1,000	
Limestone	5	-	10,000	
Sandstone	20	-	2,000	
Granite	1,000	-	2,000	
Slates	600	-	5,000	

Moisture up to about 17% dramatically lowers resistivity, mineral salts are needed and pure water is an insulator, and as the temperature approaches freezing the resistivity also rises dramatically. All this factors into your estimation of Rg. The actual resistivity seen by the circuit depends on the earth resistivity and the type of grounding system you have installed. The more metal in the ground, and deeper, the lower the resistance. Know thy earth! A map of USA soil resistivity follows:

Map of Effective Ground Conductivity in the USA
REA Bulletin 1751F-802
http://www.rurdev.usda.gov/SupportDocuments/UTP_Bulletins_1751F-802.pdf

1. GENERAL

1.1 Ground, is defined as a conducting connection by which a circuit or equipment is connected to the earth. The connection is used for establishing and maintaining the potential of the earth, or approximately that potential, on the circuit or equipment connected to it. A "ground" consists of a grounding conductor, a grounding electrode, a grounding connector which attaches the grounding conductor to the ground electrode, and the soil in contact with the ground electrode.

1.2 Using Grounds in Protection Applications:

1.2.1 For natural phenomena, such as lightning, grounds are used to discharge the system of current before customer or personnel can be injured or vulnerable system components can be damaged.

1.2.2 For potentials due to faults in electric power systems with ground return, grounds aid in ensuring rapid operation of the power system protective relays by providing additional low resistance fault current paths. The low resistance path provides the means for the removal of the potential as rapidly as possible. The ground should drain the potential before personnel are injured or the telephone system damaged.

1.3 Ground Resistance: Ideally, a ground should be of zero ohms resistance. In reality, this value cannot be obtained due to the series resistances shown in Figure 1: Components of Resistance in a Ground Connection. Grounding theory and methods for obtaining a ground of the smallest practical resistance will be discussed in subsequent paragraphs.

2. PHENOMENA AFFECTING GROUND RESISTANCE

2.1 Introduction: A grounding electrode cannot be driven into the soil with the expectation of obtaining a good, low resistance, ground. Many factors, both natural and human, may affect results. Some of the factors include:

Figure 1: Components of Resistance in a Ground Connection

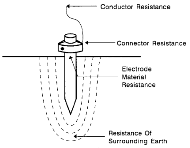
Conductor Resistance
Connector Resistance
Electrode Material Resistance
Resistance Of Surrounding Earth

2.1.1 **Earth Resistivity**: The electrical resistivity of the earth (resistance of the earth to the flow of current) is of major importance. The unit of earth resistivity, the ohm·meter, is defined as the resistance, in ohms, between opposite faces of a cube of earth one cubic meter in volume. An alternative unit of measurement, the ohm·centimeter, is defined as the resistance in ohms, between opposite faces of a one centimeter cube of earth. To convert ohm·meters to ohm·centimeters, multiply by the former by 100.

2.1.1.1 Earth resistivity varies over a considerable range. Within the United States it varies from a few ohm·meters along some coasts to many thousands of ohm·meters in rocky, mountainous country. Figure 2: Estimated Average Earth Resistivity in U.S. provides very general data on average surface earth resistivity throughout the United States.

inductance L1 and capacitance C1. The capacitance and inductance of the ATU is needed to tune out the reactance of the various components of the antenna for which we do not have data. What to do?

Without an antenna analyzer (expensive) one can only estimate the many component values based on published antenna models. Ken Khun's engineering page has a number of nice models and a good discussion of the antenna parameters we will be trying to understand and use. For my personal setup my antenna consists of about 75' of 14awg wire averaging about 10-12' high and with another 25' of lead-in. For such an antenna Khun models a 30m antenna 3m high which approximates my situation closely enough. Wire resistance is negligible as is the radiation resistance which he shows to vary between some 0.1 to 1.5 ohms. Such an antenna is capacitive by nature and will have about 220 - 375pF along with some small 20uH inductance, not too far off a standard "dummy" antenna. This pretty much leaves earth resistance Rg as the main unknown parameter.

For Rg one has the option of punting and taking the "Standard" value of 25 ohms. In my modeling I have found this parameter to be critical and highly sensitive. I do not recommend guessing here, it is recommended one go to the internet page of their state, county, or local government and search for reports on ground or soil resistivity (or conductivity). As this is an important agricultural and engineering parameter, it has been surveyed for most places and should be available with some effort. Effort well rewarded. Earth resistivity for the Texas Gulf Coast is a mercifully low 10 - 15 ohm meters, but for many regions this will not be the case.

Earth resistivity depends on a number of variables including the material, moisture, mineral salt, and temperature. A table and for this I am deeply in his debt. My hat's off to Dick, thanks so much.

The following discussion presents the results of many models cranked through the design calculator. I wished to understand what factors play main roles and which have minor parts.

For a conventional detector-circuit coil one can pretty much control all the main factors, really only the inductance and capacitance of the circuit. The main and only surprise comes in the form of "stray capacitance" resulting from the spacing of the coil wires. This can readily be 1) guessed at and 2) minimized by good coil winding technique. For the antenna circuit by contrast many of the needed parameters, earth resistance, antenna capacitance and inductance, and antenna resistance in the forms of actual wire resistance and radiation resistance are, for most of us, unknown and only guessed at. The following figure illustrates the antenna and ATU with a list of the main "components" that need to be understood and/or modeled.

The Antenna Equivalent Circuit consists of an AC voltage source Va (the signal of interest), in series with some radiation resistance Rr, antenna capacitance Ca, antenna inductance La, antenna resistance Rr, and finally a ground resistance Rg for the return path to complete the circuit. In addition the ATU consists of a coupling capacitance C2, and a tank with

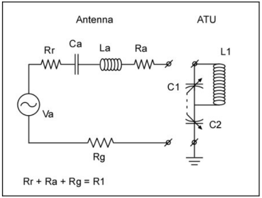

2.1.1.2 In addition to regional variations, earth resistivity may vary widely within very small distances due to local soil conditions. Table I lists typical ranges of earth resistivity for various soil types. This table should be useful in selecting locations at which a ground is to be installed.

TABLE I: RESISTIVITY OF VARIOUS SOILS

SOIL	RESISTIVITY RANGE (ohm·meters)
Loam	5 - 50
Clay	4 - 100
Sand/Gravel	50 - 1,000
Limestone	5 - 10,000
Sandstone	20 - 2,000
Granite	1,000 - 2,000
Slates	600 - 5,000

2.1.2 **Soil Moisture**: Nearly any soil, with a zero moisture content, is an insulator. Fortunately, this condition is rarely encountered except in desert areas or during periods of extreme drought. Figure 3: Typical Variation of Soil Resistivity with Moisture illustrates the typical affect of moisture on soil resistivity. It should be noted that above 17% moisture by weight additional moisture has little effect. Below this figure resistivity rises rapidly until, at 2% moisture it reaches 100 times its value at 17% moisture. Thus, a good ground connection should always be in contact with soil having a ground water content in excess of 17%. Local well drillers

Figure 3: Typical Variation of Soil Resistivity with Moisture

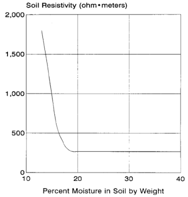

should be able to provide information concerning the depth of the water table in their areas. Water content alone does not provide a good ground in many areas so do not be misled by moisture depth only. (See Soil Mineral Content section to follow).

2.1.3 **Soil Mineral Content**: Water with no mineral salt content is nearly as good an insulator as soil with no moisture content. Figure 4: Typical Effect of Mineral Salt on Earth Resistivity illustrates the effect of mineral salt content on soil resistivity. Soils which lack adequate soluble mineral salts may be encountered from time to time.

Figure 4: Typical Effect of Mineral Salt on Earth Resistivity

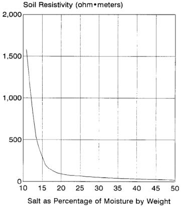

2.1.4 **Temperature**: As the temperature of soil decreases, resistivity increases. When the soil temperature drops below the freezing point of water, resistivity increases rapidly, as shown in Figure 5: Typical Variation of Soil Resistivity with Temperature.

Figure 5: Typical Variation of Soil Resistivity with Temperature

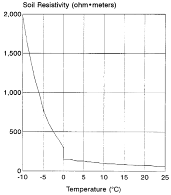

Notes on simulations for a Tuggle Front End

http://www.lessmiths.com/~kjsmith/crystal/atu.shtml

By Kevin Smith

Introduction:

When designing my radios I always make extensive use of Mike Peebles and Dan Petersen's "Professor Coil" spreadsheet. This tool along with the many tutorials and explanations online have vastly eased the design work around homebrew coil winding. We can now model a coil with good accuracy prior to the actual job and thus better match the coil to the other components we plan to include. One quickly becomes aware that most coils for the broadcast band vary around a nominal 220 or so uH. More than this and most variable capacitors will have too much bottom-end capacitance to tune the top of the band. Less than this and you need caps with fairly high values when fully meshed, 500+pF or more.

All the above is pretty straight forward. On the other hand, when I have been reviewing double-tuned set designs, I often note that the coil used for the ATU (antenna tuning unit) will have an inductance quite a bit lower than the main tank coil. Professor Coil doesn’t address this aspect of coil design and I have found little discussion online concerning ATU design. Mostly this seems to be dealt with in passing or as a digression when dealing with other subjects. The most useful web sites for addressing this aspect of crystal set design include Dick Kleijer's excellent work and Ramon Vargas's detailed analysis of the "Tuggle Front End", that’s about it. Neither site discusses the ramifications of varying different parameters such as the Earth Resistance, Coil Inductance, etc although Kleijier's site includes a great calculator page to allow one to ask these questions. In my explorations of ATU design (primarily Tuggle) I have made extensive use of Kleijier'e page

APPENDIX C MEASUREMENT OF SOIL RESISTIVITY

1. GENERAL

1.1 Soil Resistivity Measurements are commonly made with a test instrument that uses the four-terminal fall of potential method. The test instrument has four terminals that are connected to four electrodes arranged at equal distances along a straight line (shown in Figure C1: "Four-Terminal" Method for Measurement of Soil Resistivity). Internally the instrument contains a current circuit and a voltage circuit. The current source can be a handdriven a.c. generator or a voltage reversing vibrator that causes a current to flow between the two outer electrodes (Terminals C1 and C2). A potential is measured between the inner electrodes (Terminals P1 and P2). The voltage and current circuits are coupled within the test set to provide a reading in ohms.

1.2 The Theory for This Measurement was developed by Dr. Frank Wenner of the U.S. Bureau of Standards in 1915 and published in Report No. 258, Bulletin of Bureau of Standards, Vol. 12, No. 3, October 11, 1915, "A Method of Measuring Earth Resistivity." Dr. Wenner established that, if the test electrode depth is small compared to the distance between the electrodes, the following equation applies to determine the average soil resistivity to a depth equal to the distance between the electrodes:

$\rho = 2\pi AR = 6.28AR$
Where: q = Average soil resistivity to depth equal to A, in ohm·centimeters
$\pi = 3.1416$
A = Distance between electrodes, in centimeters

R = Test instrument resistance reading, in ohms
Note: Divide ohm·centimeters by 100 to convert to ohm·meters.

1.3 Ground Test Instruments generally use an alternating voltage source with a frequency not related to power system fundamental frequencies or their harmonics. This avoids the effects of polarization and foreign earth currents which could produce erroneous results.

1.4 Objectives of Soil Resistivity Measurements: The first is to determine the type of earth connection required to provide the objective resistance to earth. The second is to define any geological limitations that might be present, such as a rock layer, that would restrict installation of the grounding system.

2. BASIC SOIL RESISTIVITY MEASUREMENT

2.1 Introduction: The depth to which the average soil resistivity is desired determines the distance (A) between the test electrodes. This distance will typically be the length of the ground electrode to be installed plus the depth below the earth's surface to which it will be driven. A measurement should be taken with test electrode spacings of one-half, one, two and four times the length of the proposed ground electrode. This will identify the presence of large deviations in the soil resistivity. Place four test electrodes along a base line in relation to the proposed vertical ground electrode location as shown in Figure C2: "Four-Terminal" Method for Measuring soil Resistivity. The test electrodes should be driven into the soil to a depth equal to A/20. Depths for test electrodes for various distances (A) are shown in Table C-I.

TABLE C-I

When we disconnect the antenna from the antenna tuner, the resonance frequency of the circuit will increase to value f2.
This is because Cp is no longer part of the tuned circuit.
f2=1/(2.pi.√(L5.C6)
The frequency shift of the circuit is equal to: f2-f

The loaded Q of the circuit
If we connect the antenna (via C4) to the LC circuit, the Q of the circuit will degrease.
The Q we have then is called the Q of the loaded circuit (or the loaded Q).
If the LC circuit is well matched to the antenna, the loaded Q will be halve the value of the unloaded Q.

Step 4
Now we have X8 we can calculate also XP with the formula:
$Xp=(R1^2 + X8^2)/X8$

XP is a capacitor which is parallel to C6.
XP represents a capacitor value of: $Cp=1/(2.pi.f.XP)$, and this is in parallel with C6

Step 5
For resonance of the circuit must apply:
$f=1/(2.pi.\sqrt{(L5.Ctotal)}$

Or:

$Ctotal = C6+CP = (1/(2.pi.f))^2/L5$ if we substract the value of CP, the value of C6 is left.

Step 6
Finally we want to know the value of C4.
We already had X8.

With the formula $XC4=X8+XL2-XC3$ we can calculate the impedance of C4.
And then with $C4= 1/(2.pi.f.XC4)$ calculate the value of C4.
Now we know all values of the matched circuit.

Sometimes, XC4 will have a negative value, in that case it is not possible to get a match by means of a variable capacitor on the place of C4.
As an alternative, we then can replace C4 by a coil with value:
$L4= -XC4/(2.pi.f)$

Frequency shift when disconnecting the antenna

Test Electrode Depths for Various Distances Between Electrodes Distance Between

Electrodes (A)		Test Electrode Depth (B)	
Feet	Meters	Inches	Centimeters
5	1.52	3.0	8
8	2.44	5.0	13
10	3.05	6.0	15
16	4.88	10.0	25
20	6.10	12.0	30
30	9.14	18.0	46
40	12.19	24.0	61
50	15.4	30.0	76

2.2 Performing the Measurement: Connect the leads from the four test electrodes to the proper terminals on the test set, C1, P1, P2 and C2. Complete the measurement as described by the manufacturer of the test equipment. Calculate the soil resistivity by the equation in Paragraph 1.2 and record results.

REA Bulletin 1751F-802
Figures
Page 59

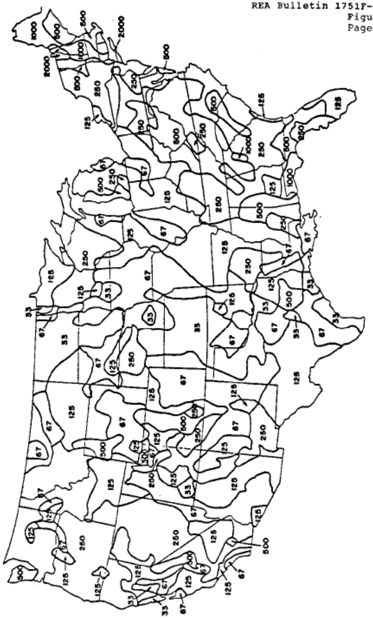

FIGURE 2

Estimated Average Earth Resistivity in U.S. (ohm meters)

Step 2

The antenna can be considered as 3 separate components, R1, Z2 and Z3

We take the sum of Z2, Z3 and variable capacitor Z4, we consider this sum as one capacitor with impedance value Z8.

So $Z8=Z2+Z3+Z4= -JX8$

Z8 must provide the match between resistor R1 and R7

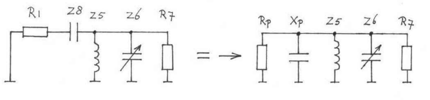

Step 3

Z8 in series with R1 can be converted to a parallel circuit RP, XP met de formulas:

$RP=(R1^2 + X8^2)/R1$ $XP=(R1^2 + X8^2)/X8$ however, we can't use these formulas yet, because X8 not known yet.

For impedance match, parallel resistor RP must be equal to parallel resistor R7.

So: $RP=R7$

From this follows:

$$(R1^2 + X8^2)/R1 = R7$$

$$R1^2 + X8^2 = R7.R1$$

$$X8^2 = R7.R1 - R1^2$$

$$X8= \sqrt{(R7.R1 - R1^2)}$$

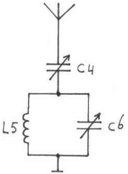

The antenna tuner

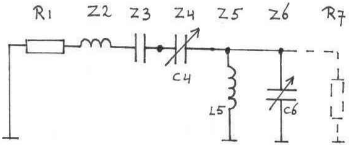

The antenna tuner split into separate impedances.
R1, Z2 and Z3 represents the antenna.
R7 is not a real resistor, but represents the losses in parallel circuit L5, C6

Step 1
Choose a frequency, calculate the complex impedance of the coil: Z5=+J(2.pi.f.L5)
Calculate the parallel resistor (R7) of the circuit L5 C6 with the formula: R7=2.pi.f.L5.Q
Resistor R7 represents the losses occurring in L5 and C6.

Soil Survey of Harris County, Texas
http://texashistory.unt.edu/ark:/67531/metapth130219/m1/148/

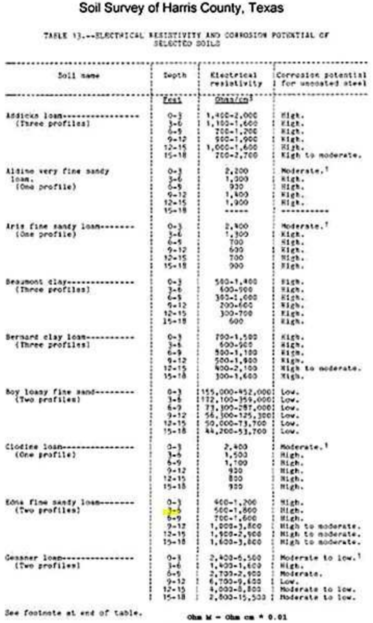

Soil Survey of Harris County, Texas

TABLE 13.--ELECTRICAL RESISTIVITY AND CORROSION POTENTIAL OF SELECTED SOILS

Soil name	Depth	Electrical resistivity	Corrosion potential for uncoated steel
	Feet	Ohm/cm^3	
Addicks loam---------------- (Three profiles)	0-3	1,400-2,000	High.
	3-6	1,100-1,600	High.
	6-9	700-1,200	High.
	9-12	900-1,900	High.
	12-15	1,000-1,600	High.
	15-18	700-2,700	High to moderate.
Aldine very fine sandy loam. (One profile)	0-3	2,200	Moderate.[1]
	3-6	1,000	High.
	6-9	930	High.
	9-12	1,400	High.
	12-15	1,900	High.
	15-18	-----	----------
Aris fine sandy loam-------- (One profile)	0-3	2,400	Moderate.[1]
	3-6	1,300	High.
	6-9	700	High.
	9-12	600	High.
	12-15	700	High.
	15-18	900	High.
Beaumont clay--------------- (Three profiles)	0-3	500-1,400	High.
	3-6	600-900	High.
	6-9	300-1,000	High.
	9-12	200-600	High.
	12-15	300-700	High.
	15-18	600	High.
Bernard clay loam----------- (Three profiles)	0-3	700-1,500	High.
	3-6	600-900	High.
	6-9	500-1,100	High.
	9-12	500-1,900	High.
	12-15	400-2,100	High to moderate.
	15-18	300-1,600	High.
Boy loamy fine sand--------- (Two profiles)	0-3	155,000-452,000	Low.
	3-6	172,100-359,000	Low.
	6-9	73,300-287,000	Low.
	9-12	56,300-125,300	Low.
	12-15	50,000-73,700	Low.
	15-18	44,200-53,700	Low.
Clodine loam---------------- (One profile)	0-3	2,400	Moderate.[1]
	3-6	1,500	High.
	6-9	1,100	High.
	9-12	930	High.
	12-15	800	High.
	15-18	900	High.
Edna fine sandy loam-------- (Two profiles)	0-3	400-1,200	High.
	3-6	500-1,800	High.
	6-9	700-1,600	High.
	9-12	1,000-3,800	High to moderate.
	12-15	1,900-2,900	High to moderate.
	15-18	1,600-3,800	High to moderate.
Gessner loam---------------- (Two profiles)	0-3	2,400-6,500	Moderate to low.[1]
	3-6	1,400-1,600	High.
	6-9	2,700-2,900	Moderate.
	9-12	6,700-9,600	Low.
	12-15	4,000-8,800	Moderate to low.
	15-18	2,800-15,500	Moderate to low.

See footnote at end of table.

Ohm M = Ohm cm * 0.01

Harris County Summary (@ 1 – 2 m depth):

Loam	11 - 16	ohm M
VF Sandy Laom	10	
Sandy Loam	13	
Clay	6 - 9	
Clay Loam	6 - 9	
Fine Sand	11k – 34k	
Loam	15	
Fine Sandy Loam	5 - 16	
Loam	14 - 16	

Equation of single rod resistance of earthing:

$Rrod=\rho/2\pi Lr\ .[\ln(8Lr/d)-1]$

ρ: is the resistivity of the soil in Ω•m

Lr: is the length of the ground rod in m

D: is the diameter of the ground rod in m

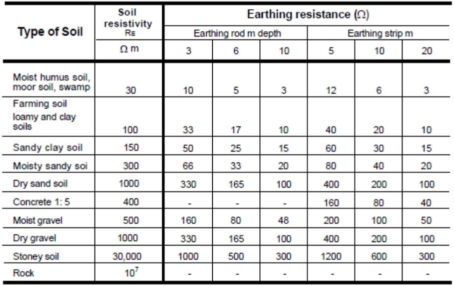

Type of Soil	Soil resistivity R_E	Earthing resistance (Ω)					
		Earthing rod m depth			Earthing strip m		
	Ω m	3	6	10	5	10	20
Moist humus soil, moor soil, swamp	30	10	5	3	12	6	3
Farming soil loamy and clay soils	100	33	17	10	40	20	10
Sandy clay soil	150	50	25	15	60	30	15
Moisty sandy soi	300	66	33	20	80	40	20
Dry sand soil	1000	330	165	100	400	200	100
Concrete 1: 5	400	-	-	-	160	80	40
Moist gravel	500	160	80	48	200	100	50
Dry gravel	1000	330	165	100	400	200	100
Stoney soil	30,000	1000	500	300	1200	600	300
Rock	10^7	-	-	-	-	-	-

LEM Instruments Brochure

=X8 (Ω)
=XP (Ω)
=CP (pF)
=C6+CP (pF)
=XC4 (Ω)
=Frequency without antenna (kHz)

Calculating the component values.

Here is explained what the calculator is calculating.
To calculate the component values, we need the following data:

The frequency.
The induction value of the coil (L5).
The unloaded Q of the LC circuit
The complex impedance of the antenna

The complex impedance of the antenna is build up with a certain resistor R1 which is in series with a coil (L2) and a capacitor (C3).

The values of R1, L2, and C3 are depending on many factors, like frequency, antenna length, height of antenna above ground, etc.

The impedance of the antenna is: Zantenna = R1+Z2+Z3 = R1+J(XL2 - XC3).

With a short antenna (shorter then 1/4 wavelength) XC3 will have a higher value then XL2, the result is a capacitive complex impedance for the antenna, so with a minus sign in front of the J.

Click on "calculate".

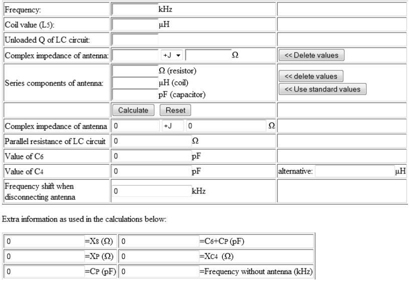

Frequency:	kHz	
Coil value (L_5):	µH	
Unloaded Q of LC circuit:		
Complex impedance of antenna:	+J ▾ Ω	<< Delete values
Series components of antenna:	Ω (resistor) µH (coil) pF (capacitor)	<< delete values << Use standard values
	Calculate Reset	
Complex impedance of antenna	0 +J 0 Ω	
Parallel resistance of LC circuit	0 Ω	
Value of C_6	0 pF	
Value of C_4	0 pF	alternative: µH
Frequency shift when disconnecting antenna	0 kHz	

Extra information as used in the calculations below:

0 =X_S (Ω)	0 =C_6+C_P (pF)
0 =X_P (Ω)	0 =X_{C4} (Ω)
0 =C_P (pF)	0 =Frequency without antenna (kHz)

Frequency: kHz
Coil value (L5): µH
Unloaded Q of LC circuit:
Complex impedance of antenna: Ω
Series components of antenna: Ω (resistor)
µH (coil)
pF (capacitor)

Complex impedance of antenna Ω
Parallel resistance of LC circuit Ω
Value of C6 pF
Value of C4 pF alternative: µH
Frequency shift when
disconnecting antenna kHz

Extra information as used in the calculations below:

How to measure the impedance of an AM-band antenna-ground system, what one can do with the results, along with some measurements

By Ben H. Tongue
http://www.bentongue.com/xtalset/20MeaAGs/20MeaAGs.html

Quick Summary: This Article describes a method to measure the series capacitive and resistive parameters of the impedance of an antenna-ground system vs frequency. Results from measurements on an attic antenna are given.

Schematic using a half wave rectifier Schematic using a full wave rectifier

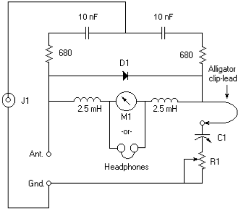

Fig. 1 - Ver. A of Antenna-ground measurement bridge.

The circuit in Fig. 1 was inspired by an Article in The Crystal Set Society Newsletter of Jan 1, 1995. It was written by Edward Richley. He used a 1 MHz crystal oscillator for his source, so had no problem with using a 200 uA meter. I use a sine wave function generator for my RF source, but a radio Service man's oscillator may also be used if it has enough output. Either of these sources cannot supply as much signal as the xtal oscillator, so I had to increase sensitivity. That's what the 2.5 mH chokes and 5 nF caps are for. The 2.5 mH chokes eliminate RF loading by any resistive component of the meter or phones on the diode detector. The 5 nF caps eliminate resistive DC loading on the detector from the two 680 ohm resistors. I lay out the components breadboard style on a nonconductive table to minimize stray capacity, keep connections short, and especially keep the signal source lead of J1 away from the connections to each end of D1. In my setup D1 is a 1N34A, M1 is a 0-20 uA DC meter, R1 is a 75 ohm non-inductive carbon pot and C1 is a two gang variable cap of 365 pF per section. I parallel the two sections when the antenna capacitance is above 365 pF. A lower sensitivity meter can be used than the one used here, at the cost of a requiring a higher applied signal to J1.

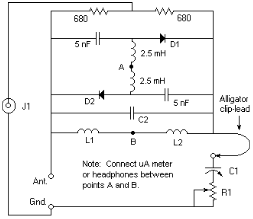

Fig. 2 - Ver. B of Antenna-ground measurement bridge.

If a sensitive enough meter is not available, a pair of high impedance phones (2000 ohms DC resistance) or preferably, a sound powered pair with the elements wired in series can be used. In this case, the generator must have its AM audio modulation turned on at its highest level. A modulation frequency of about 1 kHz is recommended. If the meter is used, do not connect the phones. If phones are used, do not connect a meter.

To use the bridge, tune the generator to a frequency of interest. Adjust C1 and R1 for minimum deflection on M1 or a null of the modulation tone in the phones. Increase the RF signal to J1 as much as possible in order to get the sharpest and most

---

CALCULATING COMPONENT VALUES FOR AN ANTENNA TUNER

Dick Kleijer crystal-radio.eu
http://www.crystal-radio.eu/entunercalc.htm

In this article we are going to match a parallel LC circuit to an antenna.

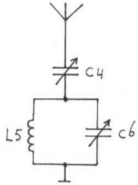

Matching means, connecting LC circuit and antenna in such a way that there is maximum power transfer from antenna to LC circuit.
This article makes use of complex impedances, a explanation about this you can find here (external link).

There are several ways of matching the antenna to the LC circuit, I discuss here only the matching via a variable capacitor between antenna and LC circuit.

With the next calculator, the component values of this antenna tuner can be calculated.
Enter: frequency, value of the coil in the LC circuit (L5) and the unloaded Q of the LC circuit L5, C6
Also enter the complex impedance of the antenna, or the series component values of the antenna.
(when you enter both complex impedance, and series components, the complex impedance is used, and not the series components).
If you don't know the antenna impedance, use the standard values for the series components: 25 Ω, 20 µH en 200 pF.

Now with various settings of C1, the frequency range of the circuit, the voltage across coil L1, and the Q (across coil L1) are measured.

The measured voltages are Volt peak-peak.

The frequency range can be slightly influenced by the measuring amplifier which was connected across coil L1.

precise null. Measure the resistance of R1 with an ohmmeter. Use any desired method to measure C1. I use the cap. measurement range of my Fluke DVM. I'm sure the reader does not need to be reminded that this test involves radiating a weak RF signal from the antenna when making the measurements, so the length of time the generator is on should be kept as short as possible.

Possible issues: More sensitivity is needed or interference from antenna pickup of local stations obscures the bridge null.

If insufficient signal is available from the RF generator to provide satisfactory meter readings, one can use the more sensitive broadband circuit shown in Fig. 2. The values of L1 and L2 are 2.5 mH and C2 is set to zero in the broadband version. A full wave rectifier is used instead of the half wave one used in Fig.1 and it gives about twice the output. One can also change from using 1N34A diodes and try Schottky Zero Bias detector diodes such as the Agilent HSMS-2850 in either circuit. The HSMS-2855 Zero Bias diode is especially suitable for use in the circuit shown in Fig. 2 since it is a package having two independent diodes, one for D1 and the other for D2. One must be cautious when using the HSMS-2855 because the diodes can be damaged by the application of too strong a signal to J1. This can happen if the signal generator signal is very strong when the bridge is greatly unbalanced. It's best to start with a weak signal, balance the bridge, then increase the signal if necessary.

If the signal from the RF generator is not strong enough to override local pickup, thus obscuring the meter null, selectivity may be added to the bridge shown in Fig. 2 by making use of C2 and changing L1 and L2. If L1 and L2 are changed to, say, 10 uH inductors and C2 is made equal to 1200 pF, the bridge will be tuned to about 1 MHz. These changes will reduce the influence of local pickup upon measurement of antenna-ground impedance at 1 MHz. One suitable 10 uH inductor is Mouser's "Fastron" #434-23-100.

If one uses headphones instead of a meter as the null indicator, even greater sensitivity can be achieved by AF modulating the bridge signal generator and connecting a parallel L/C tuned to the modulating frequency of the generator across the phones. This will filter out much of the interfering cross talk from local pickup and pass the modulation tone with little loss. Suggested values are L=47 mH and C=0.5 uF if the modulating frequency used is about 1 kHz. A low cost coil having an inductance of 47 mH and a Q of about 9 at 1 kHz is available from Mouser as a Fastron Plugable Shielded coil, #434-02-473J ($1.20 each). Greater selectivity against cross-talk can be obtained by decreasing the inductance and increasing the capacitor.

I live about 9 miles from WOR and 12 from WABC, both 50 kW stations. 10 volts peak-to peak applied to the bridge overrides the local radio station pickup sufficiently to provide a clear null on the meter when using the circuit shown in Fig. 1 when using a 1N34A diode. A useable null with an applied signal of only 1.5 volts p-p can be obtained when using the circuit in Fig. 2 with zero bias detector diodes, sound-powered phones instead of a meter and the parallel LC filter.

Notes:

* If the RF source has too great a harmonic content, the bridge balance null will become less deep and sharp. That's why I used a sine wave function generator to assure a low harmonic content. If one uses a function generator for pure sine waves, make sure the symmetry control is set for best symmetry (minimum reading on the bridge microampmeter). In April

L12, only the outermost winding is removed to reduce inductance a little bit. This reduced the total wirelenght from 15 to 14.5 meters.

The frame of this antenna unit is made of 8mm polyethene sheet.

Tuning capacitor C2 is driven via a 1:5 vernier drive, so we can tune it very accurate. Tuning capacitor C1 has no vernier drive.

Experiments:

During the measurements the antenna unit is connected to the signalgenerator via a dummy antenna.

The dummy antenna.

Made with a resistor of 50 Ohm, a coil of 20 uH, and a capacitor of 200 pF. The 50 Ohm output resistance of the signal generator is in parallel with the 50 Ohm resistor.

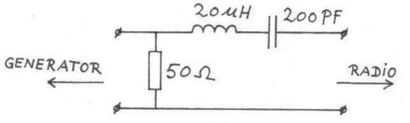

The dummy antenna makes the receiver input is driven with a impedance of 25 Ohm in series with 20 uH in series with 200 pF.. This is about the impedance of a average longwire antenna. With the dummy antenna connected to the generator (but without the receiver connected), the output voltage of the generator is adjusted to 50 mV peak-peak.

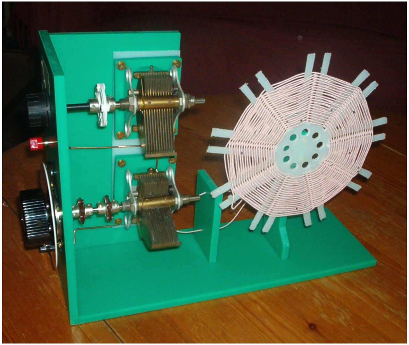

The circuit L1,C2 has a high impedance (e.g. 1 M.Ohm). But the antenna has a low impedance (e.g. 10 Ohm), tuning capacitor C1 forms a impedance match between this high and low impedance.

With a certain value of C1, there wil be maximum power transfer from antenna to the circuit L1,C2. Then there is maximum voltage across the coil L1, and maximum sensitivity of the receiver. At low frequencies we must for instance adjust C1 to 100 pF, and at high frequencies to 20 pF, but these values are depending on the (lenght of) antenna we connect to it. For the circuit Q however, the lower the value of C1, the higher the Q. More information about this, you wil find here.

Coil L1 is wound with litzwire 660x 0.04mm (660/46 AWG), on a polypropylene former. This coil is described here as coil

2004 Tom Polk published a description and schematic for a low distortion medium wave home brew signal generator. It looks very good, and can be found at:
http://www.beecavewoods.com/testequipment/sinewave.html

* If the resistance of a specific antenna-ground system is greater than 100 ohms, use a pot of a higher value than 100 ohms.

* A typical antenna-ground system will show a capacitance of a few hundred pF at the low end of the BC band. Because of the series inductance in the system, the measured capacitance will rise at higher frequencies. At a high enough frequency the system will go into series resonance and the bridge will not be able to be balanced. To measure the system series resistance at or above this resonance, place a hi Q capacitor of, say 100 to 220 pF in series with the antenna. That will raise the resonant frequency sufficiently so that the capacitor-antenna-ground circuit will be capacitive, a null can be obtained and the resistive component determined. An NPO ceramic or mica cap should be OK.

* At my location, detected signals from local strong stations show up as fluctuations at about 15% of full scale on the meter, but are not strong enough to obscure the bridge nulls from of the signal generator's signal.

* Unless the signal generator connected to J1 is battery powered (most aren't), it is important to put a common-mode radio frequency choke in the power line to the generator. I made mine by bundling a length of 18 ga. lamp cord into an 18 turn coil having a 9 inch diameter, and then fitting a male AC plug on one end and a female socket on the other. The turns were kept together using twist ties.

What can one do with the measurement results?

The main practical thing one can do with the bridge is to Measure and Monitor antenna-ground circuit resistance. This resistance comes primarily from the physical ground, not the antenna and ground connecting leads or radiation resistance of the antenna. Any increase in the antenna-ground resistance serves to reduce the signal power available from the antenna. Any decrease, of course increases it. A halving of the antenna-ground system resistance provides a 3 dB increase in available signal power, if one properly rematches to the crystal radio set input circuit.

Measure: One can experiment with different grounds and various ground paralleling schemes to come up with the one that has the lowest resistance. Use of this one will result in maximizing the available signal power (more volume). Experiments using a counterpoise ground can be made.
Monitor: As has been recently been posted on the Yahoo Club: thecrystalsetradioclub, earth ground resistance deteriorates (increases) over time. This results in a gradual decrease in available signal power (less volume). Periodic measurement can alert one if this is happening so steps can be taken to correct the problem.

The other thing one can do, if one is mathematically engineering a crystal radio set, is to use the R and C values as parameters in the design. See Article #22.

Measurement results on an indoor attic antenna system:

My present external (as opposed to loop) antenna is in the attic. The horizontal element used to be made up of 7 twisted strands of #26 copper wire (17 ga.), suspended by strings about 1 1/2 feet below the peak of an asphalt shingled roof. It runs

---

EXPERIMENTS WITH AN ANTENNA TUNER UNIT

Dick Kleijer crystal-radio.eu
http://www.crystal-radio.eu/enantunittest1.htm

Here some measurements done with my "antenna unit1".

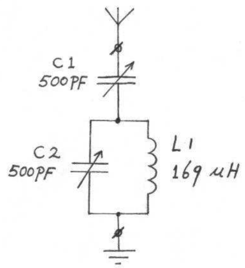

Schematic of the antenna unit 1.

The antenna unit has a tuned circuit L1, C2.

If we only look at L1 and C2, then the tuning range is 550 - 2184 kHz. But if we also connect the antenna and earth, the frequency of the circuit will decrease, so we can also receive the lowest mediumwave frequency of 530 kHz.

Variable capacitor C1 and the antenna and earth are also a part of the tuned circuit, but the antenna and earth also give reduction of circuit Q.

along under the peak and parallel to it for 53 feet. The wire is about 24 feet above ground level. The lead-in, connected to the center of the horizontal wire, runs horizontally, at a right angles for about 9 feet and then drops down vertically to the crystal radio set location, about four feet above ground level. The ground system consists of a connection to the cold water supply in parallel with a connection to the hot water baseboard heating system. To achieve a low inductance ground connection I use 300 ohm TV twinlead, both conductors soldered in parallel, for each lead. The addition of a connection to the AC neutral does not seem to reduce the inductance or resistance of this antenna-ground system. I always suggest trying the addition of a connection to the AC neutral. Sometimes it helps.

The measured antenna-ground system capacitance was 295 pF at 0.5 MHz, 325 at 1.0 MHz, 410 at 1.5 MHz and and 660 at 2.0 MHz initially. The respective series resistances measured: 17, 12, 10 and 14 ohms. The equivalent reactance elements of this antenna are a capacitance of 285 pF in series with an inductance of 12.5 uH. Since my ground is composed of the house cold water supply pipes in parallel with the the hot water baseboard heating system pipes, much of the capacitance from the horizontal attic antenna wire is to them and the roof, not a real resistive earth ground. That, I think explains the low resistance and high capacitance readings. Probably the ground system is acting as a sort of counterpoise.

I decided to see if I could get greater signal pickup by changing to a very crude simulation of a flattop antenna. To do this, I paralleled the antenna wire with a piece of TV twinlead connected to it at each end and at the point of down-lead takeoff. The twinlead was separated from the 7/26 wire by about 2 1/2 feet. The new measured antenna-ground system parameters became: Capacitance: 430 pF at 0.5 MHz, 510 at

1.0 MHz and 860 at 1.5 MHz. The respective series resistance values became: 15, 12 and 11 ohms. The equivalent reactance elements became a capacitance of 405 pF in series with an inductance of 14.2 uH. Signal pickup increased by a negligible 0.8 dB at 710 kHz, and even that was, I'm sure, within experimental error.

One may want to compare these equivalent impedance components with the 'Standard Dummy Antenna', as specified in 1938 by the IRE (Institute of Radio Engineers) in 'Standards on Radio Receivers'. My reference for this is Terman's Radio Engineer's Handbook, first edition, 1943, pp 973 and 974. A rather complex equivalent circuit for the antenna is shown on page 974. It is stated that a simpler alternative network, given in footnote #2 on page 973, can be used when only the BC band is of interest. It consists of the series combination of a 200 pF capacitor, 25 ohm resistor and 20 uH inductor. Terman states that the two antenna equivalent circuits have closely the same impedance characteristics in the BC band. The impedance graph on page 974 and the impedance from the series combination of 200 pF, 25 ohm resistor and 20 uH differ, particularly in the resistive curve in the complex equivalent circuit. The 25 ohm resistance in the simplified circuit is probably taken from the resistance in the complex circuit, at the geometric center of the BC band. This resistance is shown as constant in the simplified circuit, and as a strong function of frequency in the more complex circuit. It is suggested that the complex equivalent circuit is theoretically derived, assuming a perfect ground and therefore does not include the resistance of the ground return path. The ground circuit can easily add 15-50 or more ohms to the circuit.

#20 Published: 11/24/01; Revised 04/16/2004
© 2001-2008 Ben H. Tongue, All rights reserved

of -j14 ohms. The negative sign indicates a capacitor is needed. Here is the value for 1650 kHz:
XC=1/(2*pi*F*C)

Solving for C, we get:
C=1/(2*pi*F*XC)

C=1/(6.28*1650 kHz*14)
C=6,890 pf
This is a huge capacitor and would probably be a combination of fixed and variable capacitors.

Now, referring to the first antenna example before we added more capacitance, we can see what value of reactance is required to bring the antenna into resonance. Since that antenna had an impedance of 20 - j34 ohms, we would need a series inductor to combine with the capacitance of -j34 ohms. This inductance is:

XL= 2*pi*F*L

Solving for L we get: L=XL/(2*pi*F)

L=34/(6.28*1650 kHz) L=3.28 uH

These calculations were all done at the higher end of the broadcast band. When moving down to 530 kHz, the antennas we have been analyzing would have smaller resistance and higher negative reactance.

=-j241 ohms (capacitance has a minus sign)

The value of this complex impedance is:
= 20 + j207 - j241 ohms
= 20 - j34 ohms

As mentioned earlier the use of complex numbers allows us to simply add the resistance and reactance. The sign of the reactive component -j34 is negative and indicates the antenna is capacitive. It is also shorter than the resonant length (at resonance XL + XC = 0).

Now if you increase the capacitance of the antenna to ground with a flat top section, where the series capacitance increases to 500 pF, here is what happens:
XC=1/(2*pi*F*C)
=193 ohms
= -j193 ohms

The value of this new complex impedance is:
=20 + j207 - j193 ohms
=20 + j14 ohms

The sign of the reactive component is now positive, indicating the antenna is inductive, or long compared to resonance.

Use of the complex number notations also lets you figure what value of series reactance you would need to bring the antenna circuit into resonance. You could tune out the reactance and then match the remaining resistance to the input of the crystal set.

Referring to our last set of complex numbers: 20 + j14, you could bring the antenna circuit into resonance with a reactance

From Terman's Radio Engineer's Handbook, first edition, 1943, pp 973 and 974.

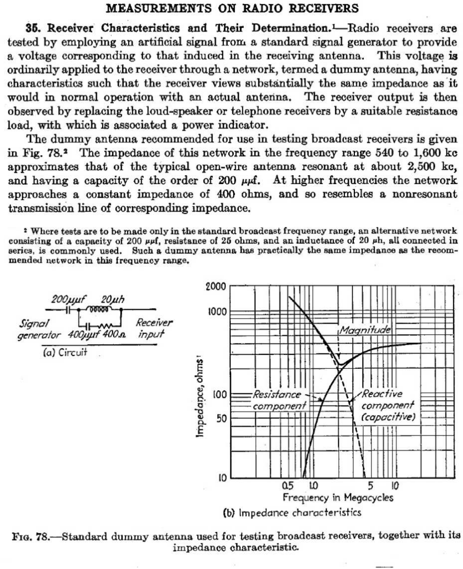
200μμf
20μh
Signal generator
400μμf 400Ω
Receiver input
(a) Circuit
2000
1000
100
50
10
Magnitude
Resistance component
Reactive component (capacitive)
Impedance, ohms
0.5
1.0
5
10
Frequency in Megacycles
(b) Impedance characteristics

MEASUREMENTS ON RADIO RECEIVERS

35. Receiver Characteristics and Their Determination.[1]—Radio receivers are tested by employing an artificial signal from a standard signal generator to provide a voltage corresponding to that induced in the receiving antenna. This voltage is ordinarily applied to the receiver through a network, termed a dummy antenna, having characteristics such that the receiver views substantially the same impedance as it would in normal operation with an actual antenna. The receiver output is then observed by replacing the loud-speaker or telephone receivers by a suitable resistance load, with which is associated a power indicator.

The dummy antenna recommended for use in testing broadcast receivers is given in Fig. 78.[2] The impedance of this network in the frequency range 540 to 1,600 kc approximates that of the typical open-wire antenna resonant at about 2,500 kc, and having a capacity of the order of 200 μμf. At higher frequencies the network approaches a constant impedance of 400 ohms, and so resembles a nonresonant transmission line of corresponding impedance.

[2] Where tests are to be made only in the standard broadcast frequency range, an alternative network consisting of a capacity of 200 μμf, resistance of 25 ohms, and an inductance of 20 μh, all connected in series, is commonly used. Such a dummy antenna has practically the same impedance as the recommended network in this frequency range.

FIG. 78.—Standard dummy antenna used for testing broadcast receivers, together with its impedance characteristic.

Working With Antenna Impedance

Jack Bryant
Birmingham Crystal Radio Group
http://www.crystalradio.us/antennas/impedance.htm

This is an elaboration of a post I made on the Rap 'N Tap Discussion group. The objective was to explain how a multi-wire flat top antenna adds capacitance to the antenna, while making it more inductive. This is an exercise in complex numbers. For example the expression 20 + j40 displays 20 ohms resistance and 40 ohms of inductive reactance. The complex number approach allows us to simply add series resistance, capacitance and inductance. The resistance remains constant when the frequency changes. However, the reactance of an inductor increases with frequency. The reactance of a capacitor decreases with frequency.

For purposes of illustration, assume an antenna with the following parameters for R, L and C in series for F=1650 kHz (our frequency of interest):

Rtotal = 20 ohms (total of radiation resistance and losses, including ground losses, since we are using a Marconi type antenna, i.e., worked against ground.)

L = 20 uh; the reactance is calculated as:
XL= 2*pi*F*L (where 2*pi =6.28, F is in hertz, and L in henries)
=207 ohms
=+j207 ohms(inductance has a plus sign)

C = 400 pF; the reactance is calculated as:
Xc=1/(2*pi*F*C) (Where F is in hertz and C in farads)
=241 ohms

ANTENNA MATCHING

by Kenneth A. Kuhn

www.kennethkuhn.com/students/crystal_radios/antenna_matching.pdf

As discussed in an earlier chapter the ten to thirty meter antenna used for crystal radios has a very low resistance and a high capacitive reactance. The ground resistance discussed previously is typically in the several tens of ohms and is effectively in series with the antenna. As an example, an antenna/ground system may have an impedance of 20 –j1000 ohms at 1 MHz. For maximum power transfer from the antenna to the resonant circuit the input impedance of the set should be a conjugate match –that is have similar resistance but the reactance will be equal in magnitude but opposite in sign. For the example this means that the crystal set should have an input resistance of about 20 ohms and a reactance of about +j1000 ohms (160 uH) at 1 MHz. The positive reactance is obtained by an inductance in series with the antenna circuit. This inductance should be variable to tune out the capacitive reactance of the antenna across the AM broadcast band. Tuning is not sharp as the Q of resonance is low. Table 1 shows some typical values. Note that the inductance tuning range becomes wider for longer antennas since the capacitive reactance drops rapidly as the length approaches one-quarter wavelength at the upper end of the AM band. It is important that this series inductance have very low losses or the advantage of using it will vanish –a lossy inductor could be worse than nothing.

Antenna Length	Typical Antenna XC @ 550 kHz	Typical Antenna XC @ 1.7 MHz	Antenna series Inductor tuning range
10 m -	j4040 ohms -	j1250 ohms	1200 –120 uH
15 -	j2680 -	j780	780 – 70
20 -	j1990 -	j530	580 – 50
25 -	j1580 -	j360	460 – 30
30 -	j1290 -	j240	370 – 20

Table 1: Antenna series inductance tuning range

One effect of not tuning out the reactance of the antenna is that the resonant frequency of the tuned circuit will shift because the antenna becomes a reactive load. One way to know if the series inductance has been tuned to the right value is that the station is received at the calibration point –assuming the radio tuning was calibrated.

The next issue is creating a low input impedance of several tens of ohms. There are two ways to do this and they are essentially the same. One method is wind a turn or so of wire near the ground end of the coil of resonant circuit –one end of the wire goes to the series inductor to the antenna and the other end connects to ground. This small winding transforms the low antenna/ground impedance to a high impedance across the coil. The other method is to make a tap a turn or so above the ground end of the coil of the resonant circuit to accomplish the same effect. If the winding or tap is too few turns then there is an impedance mismatch and a weak signal will result although selectivity will be relatively sharp. If the winding or tap has too many turns then the coil is overloaded by the antenna/ground impedance which also results in weak signals and the selectivity will be broad. The optimum is the point of proper impedance match although it is not very critical. The issue is how to determine the proper point. A second issue is that the required impedance level varies across the AM broadcast band. Thus, the tap should be variable.

If the coil is wound on a ferrite toroid core which enables a high degree of flux coupling from turn to turn then it is fairly easy to calculate at what turn a tap should be for the desired impedance transformation. However, our coil is typically an air-core solenoid which has a complicated flux relationship. Calculation is difficult and very error prone. The best way is to make a variety of taps and measure the impedance using laboratory methods and note the results. I will present the data of just such an experiment on a typical coil for crystal radios when this article continues …

KEVIN'S WEBSURFER HANDBOOK VI
FOR CRYSTAL RADIO

RADIO WAVE PROPAGATION

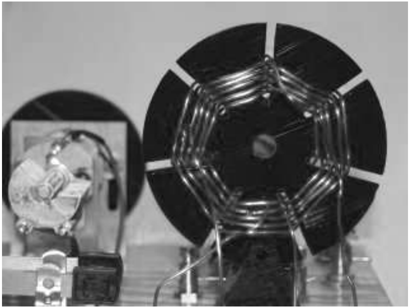

Kevin Smith
2013

Printing / Binding Instructions

1. Choose “fit to page” in print menu
2. Print document double sided on letter size paper
3. Cut the entire printed document in half
4. Fold over making sure the page numbering is continuous
5. For the cover: Print just the first page on card stock paper
   Cut the cover in half as well
6. Assemble the covers on the document
7. Punch the left side for a binding, spiral or comb as desired

http://www.lessmiths.com/~kjsmith/crystal/catalog.shtml

KJ Smith

NOTES:

Introduction

This handbook is primarily devoted to radio wave propagation. This is not a topic specific to crystal radio but to radio generally and here with a mind to LW/MW propagation. For many in the crystal radio hobby, long-distance reception, or DX, is an interesting and sought after goal. Understanding propagation adds to the joy of the hunt. Additionally, as crystal sets have no amplification or signal modification other than rectification, they are more subject to the vagaries of atmospherics and propagation irregularities.

We open with Tom Giella's excellent layman's overview of MF propagation. This discussion reviews medium frequency propagation from a thoroughly modern standpoint, benefitting from land and space-based tools and 100 years of progress in understanding the phenomenon. This is the answer and if propagation is your only interest then you can stop reading after this article. Following this we step back in time to the days of spark transmission and unamplified reception. First Chen-Pang Yang reviews the intellectual background of the scientists and engineers who were involved and the new observations they were working to explain. From here I provide source documents written by the players themselves. These give a feel for the excitement of discovery and advancement that characterize the early days of radio. Finally, I provide a personal study on propagation made in the late 1980's. This gives some personal perspective as well as highlights the interesting kind of projects available to the hobbyist.

Of general note, the web is a marvelous source of data and information. Many long-time crystal set builders, and many others have created dedicated sites to disseminate information and resources, to share their creations and knowledge. I am eternally in your debt. All of the material in this handbook is copyright for which I have not sought permission. Therefore this is not presented for publication or copy. It is only my personal resource. I encourage anyone finding this copy to pursue ON THE WEB the web pages identified within. I include the name of the author and web address of each section. I wish to sincerely thank every author presented for their excellent pages and ask forgiveness for my editing into this handbook.

Kevin Smith
2013

www.lessmiths.com/~kjsmith/crystal/cr0intro.shtml

NOTES:

NOTES:

CONTENTS:

1 MF Radio Wave Propagation Notes
Thomas F. Giella
- 3 Medium Frequency Propagation
- 18 Aurora Oval
- 20 Equatorial Ring Current
- 21 Coronal Mass Ejection
- 23 Coronal Hole
- 25 Solar Filament
- 25 MW vs. Solar Flux and Ap & Kp Indices
- 31 E Valley-F Layer Propagation
- 35 Electron Gyro Frequency Absorption
- 36 MF/HF Radio Signal Propagation Path Skewing
- 38 Geomagnetic/Ionosphere Storm
- 41 Geological/Meteorological Effects
- 44 Polar Cap Absorption
- 44 Sunspot Group
- 46 Short Wave Fadeout
- 47 Solar Flare
- 49 Sporadic-D
- 50 Sporadic-E
- 54 Long Delayed Echo
- 55 Sudden Stratospheric Warming
- 57 D Layer Mid-Winter Anomaly
- 58 F3 Ionosphere Layer
- 58 Gray line Propagation
- 59 Plage

61 The study of Propagation: 1900-1919
Chen-Pang Yeang

97 Electrically-Conducting Strata
Kennelly 1902

99 **Radiotelegraphy**
Marconi 1911

115 **Experiments in Radiotelegraphy**
Austin 1911

165 **Diurnal Variations of Electric Waves**
Eccles 1912

187 **Daylight effect in Radio**
Kennelly 1913

215 **Seasonal Variation in signal strength**
Austin 1915

219 **The Heavyside Layer**
Marchant 1916

231 **Solar Flare detection by Shortwave Radio**
Kevin Smith

1...Introduction

231 1.1...Fun with radio, a hobbyist’s dash
234 1.2...Some Definitions
236 1.3...Short Bibliography

2...Background

237 2.1...Solar flares
239 2.2...Earth's ionosphere
241 2.3...Radio wave propagation

3...Solar flare detection by shortwave radio

244 3.1...Daily record
247 3.2...Seasonal variations
255 3.3...Flare response
260 3.4...Unsettled to active field response

4...Serendipity

264 4.1...Ionization Problem
267 4.2...Poor man's Ionosonde
270 4.3...Density waves

Below is my “ionosond” display for that day.

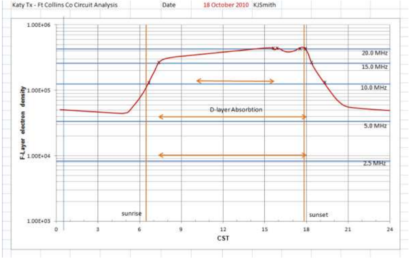

Would that I had such equipment back in 1987. Times have changed and so has the hardware with which we use, but the work goes on!

Kevin Smith
2010

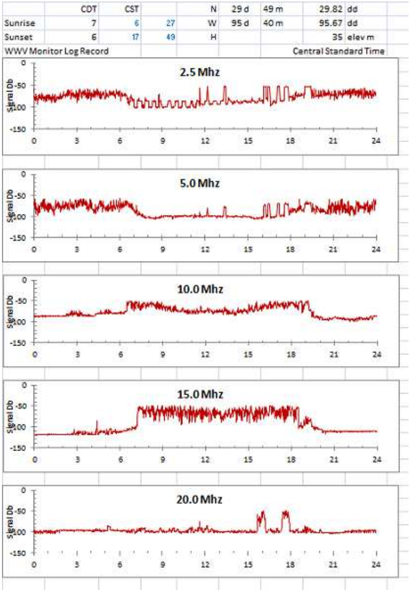

	CDT	CST		N	29 d	49 m	29.82	dd
Sunrise	7	6	27	W	95 d	40 m	95.67	dd
Sunset	6	17	49	H			35	elev m
WWV Monitor Log Record							Central Standard Time	

2.5 Mhz
5.0 Mhz
10.0 Mhz
15.0 Mhz
20.0 Mhz
Signal Db
0
-50
-100
-150
0
3
6
9
12
15
18
21
24

5...General observations on radio propagation

273 5.1...Signal strength by season
276 5.2...Signal strength by solar activity
278 5.3...The "grass", noise revealed
281 6…Postscript 2010

285 Notes

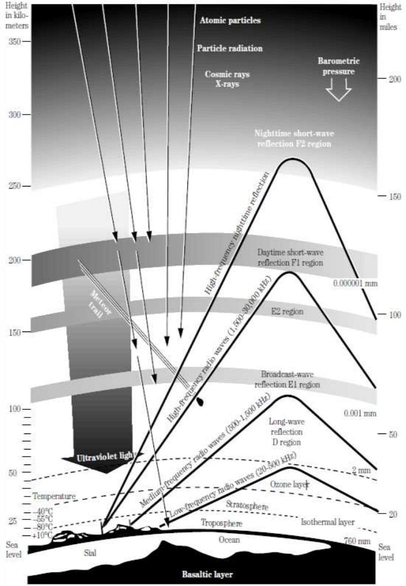
Height in kilo- meters
350
300
250
200
150
100
50
25
Sea level
Height in miles
200
150
100
50
20
Sea level
Atomic particles
Particle radiation
Cosmic rays X-rays
Barometric pressure
Nighttime short-wave reflection F2 region
High-frequency nighttime reflection
Daytime short-wave reflection F1 region
0.000001 mm
E2 region
Meteor trail
High-frequency radio waves (1,500-30,000 kHz)
Broadcast-wave reflection E1 region
Medium-frequency radio waves (500-1,500 kHz)
0.001 mm
Long-wave reflection D region
Ultraviolet light
Low-frequency radio waves (20-500 kHz)
2 mm
Ozone layer
Temperature
−40°C
−55°C
−80°C
+10°C
Stratosphere
Troposphere
Isothermal layer
Ocean
760 mm
Sial
Basaltic layer

2-7 Radio propagation in the ionosphere is affected by a number of different physical factors: cosmic rays, atomic particles, solar radiation.

Joseph J. Carr

Postscript 2010

The above report describes a study I made back in 1987 – 1991. At the time my tools included a good communications receiver (Yaesu 7700) coupled to a Commodore 64 computer. The connection was cobbled together via the game port interface. Looking back, it seems amazing that it worked at all. Today with modern receivers like my Icom R75 we can connect directly to a modern personal computer with the radio's built-in interface. Monitoring several bands becomes an easy pursuit with the correct software.

Below I provide two figures made with my current setup. The date 18 October. The record shows simultaneous monitors on 2.5, 5.0, 10.0, 15.0, and 20.0 MHz WWV. The easiest bands to interpret are for 10 and 15 MHz. For 2.5 and 5 Mhz the record is best interpreted in terms of D-layer absorption. 20 MHz is barely reached except for two small blips at 16 and 17h CST.

small compared to the 15MHz carrier. On top of that is a second interesting beat in the amplitude with 40 to 50 second period or about 0.02Hz (also seen on the power spectrum).

Voila! No noise at all, just signal everywhere.

Supplemental Records

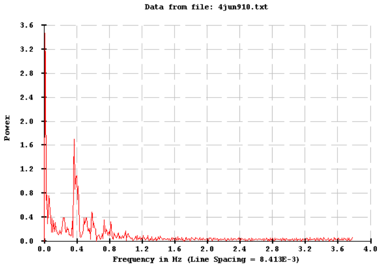

Power spectrum of the data collected on 04 June, 1998. The data was converted to the frequency domain via a Fourier transform.

Fin.

MEDIUM FREQUENCY RADIO WAVE PROPAGATION THEORY NOTES

Layman Level Explanations Of "Seemingly" Mysterious MF Radio Wave Propagation Occurrences

By: Thomas F. Giella, W4HM

http://www.wcflunatall.com/propagation5.htm

0.) Introduction.

Welcome to the "W4HM Medium Frequency Radio Propagation Theory Notes" weblog styled website. Contained within you will find the most comprehensive explanation of medium frequency radio propagation occurrences on the Internet.

In my personal opinion understanding and taking advantage of low/medium/high frequency radio propagation conditions is an integral part of successful DX operation. Therefore this site exists as an educational tool.

This website is permanently under construction as I add new data and research information continuously. If you see any spelling or sentence structure errors I've overlooked "please" feel free to advise me, thanks! Also feel free to ask for inclusion on this website of "legitimate" propagation theory. Remember that the definition of "theory" means that the concept has not or can not be definitively proven in a laboratory setting but can be inferred via systematic study.

These theory notes are primarily applicable to the 300-529 kc long wave aviation and marine navigation beacon band, 600 meter band, 530-1700 kc AM broadcast band, 160 meter amateur band and 120 meter shortwave tropical broadcast band. However much of the content is also applicable to the 3000-30000 kc HF shortwave spectrum.

A while back I became involved in a propagation research project of sorts on 10, 6 and 2 meters. I have set up a propagation beacon currently on 10 meters, the frequency is 28131 kc USB +1500 Hz using 25 watts. Actually there is a group of hams running the propagation beacons in the U.S. and around the world. Basically it's a marriage between digital PSK31 and APRS technology, actually software's run on a computer that is interfaced to a transceiver. Beacon transmission will commence on other HF bands in the future, with the ultimate aim of all HF bands and even 160 meters. It's been very interesting so far watching Es and F2 propagation openings on 10 meters, when conventional propagation wisdom says that the opening should not be occurring. A knowledgeable observer can also pick out Sporadic-E (Es) openings. You can learn more about the concept by clicking here for HF PropNET.

I have attempted to keep the propagation theory explanations in simple to understand layman terms, because long complicated technical explanations can be boring and make one's eyes glaze over. Unfortunately though sometimes while trying to keep things simple, certain definitions, meanings and technical aspects can get watered down or even lost, which tends to open me up to criticism from certain fellow space weather scientists that just don't understand the educational and public relations concept of the keep it simple stupid (KISS) principle. I choose to use W6SAI's (SK) "KISS" method of writing and communicating. I have found that this method works best whether it be in teaching about space or atmospheric weather or any other subject.

seen the characteristic signature of signal fading. There are in fact two periods of fading superimposed, one of about 50 seconds, the second about 5 seconds. The fading has the typical rounded crests and sharp drop to the separating nodes. This pattern results from the beat amplitude of two or more interfering signal frequencies.

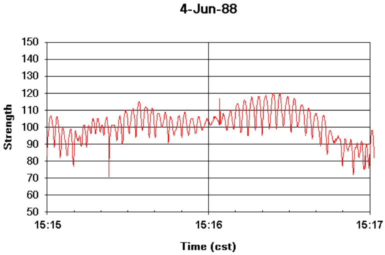

characteristic signature of signal fading

The second chart is from a run of only seven minutes giving about a tenth second interval between readings. There the fading signature shows with about 50 beats over the two-minute interval shown. Fading generally results from the interference of the same signal arriving from two paths at the receiver. A power spectrum of the data shows a strong peak around 0.4 hertz corresponding to the dominant beat (25b/m or 0.4b/s). For a beat to be set up the signals traveling along separate paths, (presumably a single hop and a double hop) they must not merely be arriving slightly out of phase but also with a difference in frequency of 0.4Hz. That shift is extremely

The "Grass", Noise Revealed

Some final words and observations concerning the "grass", the noise on the records with the very high frequency. Hidden within that noise lies some interesting signals and an interesting message. This "noise" seen on the records is in fact just more signal, but at a higher frequency than I have been presenting. The following charts were made on 04 June, 1988. They were taken specifically to look at the high frequency content of the signal. It should be noted that in both charts the signal strength scale remains exactly the same as in all the previous charts so the amplitude of the "noise", the range from crest to node is quite significant when compared to the daily variations being measured.

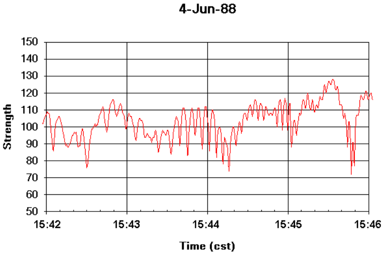

characteristic signature of signal fading

In the above case I gathered a standard 3200 points over 29 minutes with no filter giving a half-second period between points. The chart shows four minutes of data and on it can be

We hams are a curious lot with inquiring minds. A good number of us have a keen interest in low, medium, high and very high frequency radio wave propagation mechanisms and this website conglomeration is directed at this forward looking group.

We also have a segment in our radio service that is basically disinterested in radio propagation and don't feel it necessary to understand it in order to successfully work DX, which is certainly okay. However within this group exists a total lack of understanding concerning the most basic aspects of the subject. Often times I will hear them say, "The band is shifting". This lack of basic knowledge can be traced back to the licensing process where very few questions exist in the exam pools.

Then we have a third and smaller group with gigantic runaway egos that insist that they are omniscient by virtue of their Extra Class license, ARRL DXCC entity totals and "possible" electrical engineering backgrounds. Anal Retentive types?! They spend their time arguing with ignorance (Alchemists) against explanations put forth via this and other scientists, with solid backgrounds in atmospheric and/or space weather physics. You know who you are and should be ashamed of yourselves.

1.) Medium Frequency Radio Wave Propagation Overview-

Popular Myth- We don't understand medium frequency radio wave propagation mechanisms and therefore it can't be forecasted.

Fact- Yes it can and is on a regular basis at W4HM MF/HF/6M Frequency Radio wave Propagation Forecast http://www.solarcycle24.org and MF at http://www.wcflunatall.com/propagation.htm .

a.) Medium frequencies encompass 300 to 3000 kc. The simplest way to look at medium frequencies with respect to propagation issues from a layman's point of view, is to accept the fact that propagation is poor the majority of the time (See definition #6. Electron Gyro Frequency Absorption), especially past approximately 1250 miles (one refraction off of the E layer), with occasional short-lived good periods as far as 3200 miles.

Medium frequency radio waves possess elliptical polarization, with the signal splitting into ordinary and extra-ordinary rays. These rays can propagate in or out of phase, more often out of phase. The out of phase extra-ordinary ray represents a 50% power loss on the receive end of a path.

b.) Why is medium frequency propagation poor the majority of the time? D layer absorption! At daytime the D layer which is at an approximate height of 30-60 miles in the mesosphere, totally absorbs medium frequency RF signals the majority of the time. I say the majority of the time because at higher latitudes, during the winter season and especially at the low part of a sunspot cycle, daytime penetration of RF signals through the weakened D layer and then refraction via the E layer and/or Sporadic E (Es) clouds does occur. Another issue is the fact that the D layer does not totally disappear at night. Many books that deal with wave propagation erroneously state that the D and E layer's disappear after sunset, totally incorrect thanks to Galactic X-Rays, Galactic Cosmic Rays and Lightning.

otherwise. Increased ionization levels increase the signal absorption as well. Also, times of high solar activity bring instabilities in the earth's geomagnetic field. Both of these effects have an important impact on signal strength.

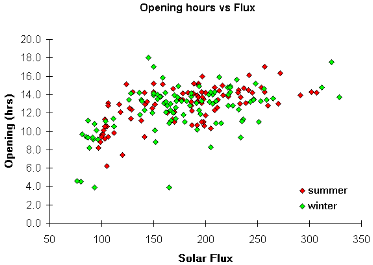

Solar activity (flux) vs length of openings

Solar activity (flux) on the other hand does indeed improve the conditions in terms of the length of the openings, especially for levels of flux above about 120. This is logical when considering the overall increased ionization levels. While the increased ionization may actually hurt signals by increasing absorption and causing greater geomagnetic instability, it does make for longer openings. Next: The "grass", noise revealed..

Signal Strength by Solar Activity

The following charts show the relationship of solar flux to signal strength and to the length of time for an opening for propagation along a circuit. Color coding provides an indication of the variation due to season as seen in the previous section. In this case I have distinguished "summer" from "winter" data on these charts in the following way: Winter = October to March, Summer = April to September.

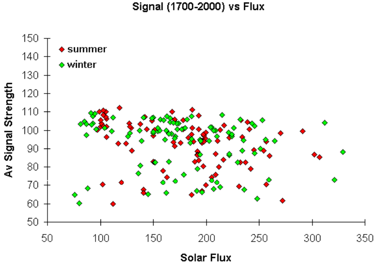

Plot of signal strength averaged over the midday period 1700 - 2000UT

Plot of signal strength averaged over the midday period 1700 - 2000UT (11-13 CST).
For both summer and winter seasons I see a weak but real inverse relationship of decreasing average signal strength with increasing solar flux. It has often been stated that propagation improves during times of high solar activity, this data suggests

c.) Background electromagnetic radiation in the 1 to 10 Angstrom range (Hard X-Rays) is a major source of ionization of the day time D layer, with our Sun as the source of Cosmic Rays, also playing a role.

The following information was contributed by Carl Luetzelschwab K9LA, a scientist and all around great guy with a very good understanding of radio wave propagation.

.....A couple years ago I was playing with Proplab Pro on a one-hop 936km path on 160m during daylight. I plotted absorption versus sunspot number. I expected a nice monotonic increase as the sunspot number increased. But the plot showed that absorption started at about 60dB at zero sunspots and was constant out to a sunspot number of about 50. Then it started climbing, reaching 100dB at a sunspot number of 150. This suggested that there was something other than hard X-rays and Galactic Cosmic Rays as the source of daytime D region absorption. So I dug into Davies 1990 (page 61), Hunsucker and Hargreaves (page 31), and Brekke (page 233). They all seem to point to the Lyman-alpha line of the solar spectrum at 1215 Angstroms ionizing NO as the main source of the quiet daytime D region. So in terms of my absorption versus sunspot number plot, the flat portion up to a sunspot number of 50 is probably due to the Lyman-alpha line ionizing NO. Then above a sunspot number of 50 the hard X-rays start contributing as the Sun becomes more active.....

Carl has produced two really good .pdf files on 160 meter propagation in 2003 and 2004. Read them here: 160 Meter Propagation & Disturbances To Propagation . He also has a propagation website with allot of good information on it at K9LA's Amateur Radio Propagation. In 2009 Carl produced yet another excellent article on 160 meter operation at Is This

Solar Minimum Better or Worse Than the Last Solar Minimum on 160m?

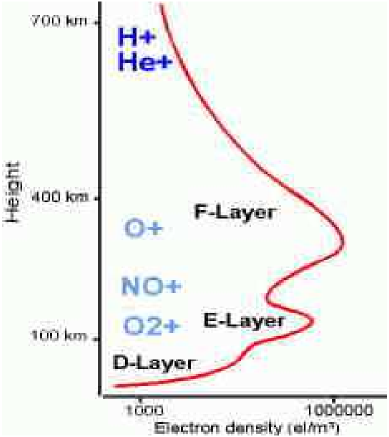
700 km
H+
He+
Height
400 km
F-Layer
O+
NO+
O2+
E-Layer
100 km
D-Layer
1000
1000000
Electron density (el/m³)

Ionosphere Ion Types

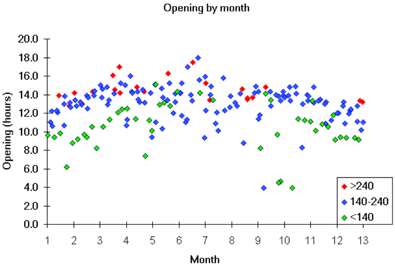

opening hours versus season

When plotting the length of time of the openings by season the situation is reversed with longer openings showing a weak but real preference to the summer season. This results principally from the longer time of ionizing radiation on the ionosphere, but in this case moderate to high levels of solar flux can add one to two hours of additional opening per day. I agree with the oft-stated general conclusion that winter is the best time for long distance communication by HF radio, but given the scatter I would temper that by addressing more closely the question of solar activity. Next: Signal strength by solar activity..

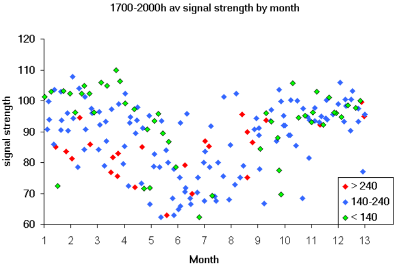

signal strength versus season

This chart, already presented in section 3.2, clearly shows that the summer season does not provide the same signal strength that can be found in winter. In general summer openings have low signal strength and greater variation from day to day. Overall they are less dependable. The chart distinguishes data collected under conditions of different levels of solar flux, (low flux in green to high flux in red). Curiously, there is a clear association between the strongest signals and the lowest levels of solar flux.

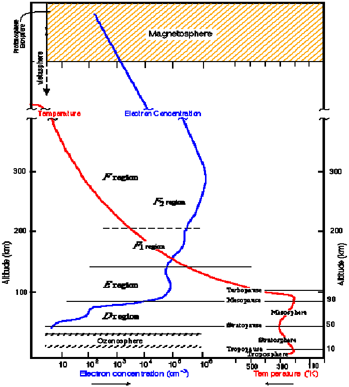
Protonosphere
Exosphere
Metasphere
Magnetosphere
Temperature
Electron Concentration
F region
F2 region
F1 region
E region
D region
Ozonosphere
Turbopause
Mesopause
Mesosphere
Stratopause
Stratosphere
Tropopause
Troposphere
Altitude (km)
300
200
100
90
50
10
10
10^2
10^3
10^4
10^5
10^6
Electron concentration (cm^{-3})
500
300
100
Temperature (°K)

Ionosphere Profile

Also speaking of the ionosphere at and near solar minimum the mid and high latitude D layer of the ionosphere at mesosphere altitude cools approximately 2 degrees C, contracts in thickness and signal absorption in the layer increases.

While I'm visiting the subject of electromagnetic radiation, our Sun emits electromagnetic radiation and matter, as a result of the nuclear fusion process. Electromagnetic radiation at wavelengths of 100 to 1000 Angstroms (Ultraviolet) ionizes the F layer, radiation at 10 to 100 Angstroms (Soft X-rays), as well as Galactic Cosmic Rays ionize the E layer. Galactic X-rays, Galactic Cosmic Rays and Lightning are the reason that the E layer is "always" present at night time, the D layer also. Background electromagnetic radiation in the 1 to 10 Angstrom range (Hard X-Rays) is a major source of ionization of the day time D layer,

Via K7RA's weekly ARRL Propagation Forecast Bulletin #46 published on November 9, 2007:

In last week's bulletin, Carl Luetzelschwab K9LA said the closest measurement we have to radiation that ionizes the F2 region is the GOES X-ray data at 0.1 to 0.8 nm. K9LA says that is not correct- he received an e-mail from Michael Keane, K1MK, with the following information:

"There does exist an instrument that measures solar EUV flux directly. That is the SOHO Solar EUV Monitor (SEM) at http://umtof.umd.edu/semflux. One SEM channel covers solar EUV in the 17-70 nm range. The other channel monitors just the 30.4 nm resonance line of singly ionized helium. In most models, this 30.4 nm line by itself represents 25-50% of the energy input to the thermosphere/ionosphere."

Galactic Cosmic Rays are not rays at all, but particles. They are ionized atoms, atoms with missing electrons ranging from a single proton up to an iron nucleus and beyond but typically protons and alpha particles, which have 2 protons and 2 neutrons. They originate from deep space, being produced by a number of different sources, such as other stars, and more exotic objects, such as supernova, which are exploding stars and their remnants, neutron stars, black holes, and distant galaxies. Cosmic Ray particles travel very close to the speed of light, and are highly energetic.

Signal Strength by Season

After spending some four years collecting and analyzing propagation data, it is reasonable to have a few observations concerning the subject. I have often read general statements concerning propagation versus time of year and solar cycle, so how did this study compare? Keep in mind that the entire data set is from a single circuit, Ft. Collins, Colorado to Houston, Texas, and that it is for a single frequency, 15.000 MHz. While it cannot be generalized to all HF radio, it provides a good test under semi-controlled conditions. Another assumption that is implicit in all this data is that the transmitted power remained constant over the whole period. I will make that assumption while noting that I did not ever confirm it to be true.

The following charts addresses the question of signal strength and opening hours versus season. Color coding provides an indication of the variation due to solar flux, but this will be explored more in the next section.

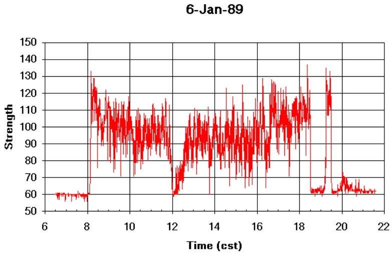

06jan89
Major flare at 1805UT, (1205 CST). Also note the evening signal dropout and return.

While on the subject of distant galactic objects, on 12/27/2004 more than a dozen spacecraft recorded the brightest event from outside the solar system ever observed in the history of astronomy. This gamma and x-ray producing super flare was emitted by a Magnetar star named SGR 1806–20. This star is an estimated 50,000 light years distant in the constellation Sagittarius on the far side of the Milky Way galaxy and obscured behind dense interstellar clouds. A similar event also occurred in 1998.

Upon arrival at Earth the X-rays were powerful enough to increase absorption in the D layer of our ionosphere and create a dayside Sudden Ionosphere Disturbance (SID) and a blackout of radio signals, amazing!!! To read more about this rare event check out this link at: http://skyandtelescope.com/news/article_1464_1.asp and http://www.sciencedaily.com/releases/2006/02/060221084628.htm .

d.) Recently I saw a post on the Topband Reflector lamenting the seemingly unexplainable differences in 160 propagation on certain paths from night-to-night. is there a reasonable explanation? Yes, unfortunately small increases in the density of the night time D layer over short periods of time, caused by smaller solar flares and also the general variability of the solar background X-Ray flux level of greater than A0, can have a profound negative impact on propagation in the form of increased absorption of high and even mid latitude medium frequency signal paths, both on the medium frequency AM broadcast band, 160 and 120 meters. Why? It only takes 10 electron volts (ev) of energy to ionize the atmosphere and 1-10 Angstrom x-ray photons energize the atmosphere at a factor of 100. This translates into D layer absorption of medium frequency signals. The lower half of the medium frequency broadcast is always affected first followed by the upper half of the medium frequency AM broadcast band, then 160 and 120 meters. If you learn nothing else on this website, remember this simple explanation and pass the word.

e.) After much personal observational research over a 35 year period, I've come to the conclusion that high and mid latitude TA and TP propagation paths tend to open up only after an approximate three day period of time passes with an energetic proton event of no greater then (10+0) on the medium frequency AM broadcast band, 160 and 120 meters.

f.) Also there are daily extremes of the background x-ray flux level. So even though the daily average might have been pretty good at say A1.1, the daily "extreme" maximum could have been C1.5, which would have been bad and would have caused a short period of increased D layer absorption.

g.) Though high latitude paths on the day light side of the Earth are primarily effected, night time high latitude paths can also be impacted by higher intensity energetic proton events. This fact is still stubbornly opposed by some otherwise very knowledgeable space weather physicists hung up on high latitude threshold Riometer data tied to Polar Cap Absorption (PCA).

h.) Another wrench in the gears preventing consistent good propagation on medium frequencies is related to Sporadic-D (Ds) absorption. Sporadic-D (Ds) occurrences have an interrelationship with brief but intense Sun based and Galactic Cosmic Rays, extremely large positive cloud to ground lightning strokes and interrelated Elves. Very large bursts of Gamma Rays have also been observed to occur in conjunction with Sprites.

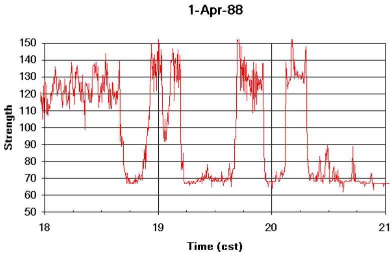

record for 01 April, 1988, detail

The above view focuses in on 18 to 21 CST showing more detail in the features. They clearly represent moments where the ionosphere electron density is passing repeatedly through the critical density for propagation. These features show a sinusoidal variation in electron density with a period between 25 and 30 minutes, (from drop to drop). I interpret the variations as density waves propagating in the ionosphere. A test of this would require one or preferable two other beacons being monitored at a close frequency to see whether the same features appear offset in time. Note also on the first chart that there are two signal loss periods between 9 and 11 CST as well, more evidence of unsettled conditions in the F layer that day.

Similar features appear on other records already presented in this article such as 06 January, 1989.

Density Waves

On a number of records I found interesting features which were not easily explainable. The record below for 01 April, 1988 shows a series of falls and jumps in signal strength between 18 and 21 CST. Similar features can occur on the morning side as well as the evening side, but seldom near midday.

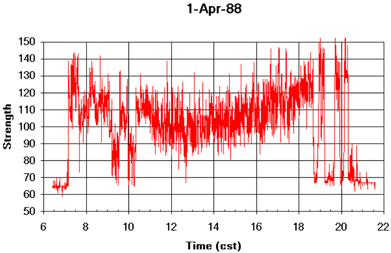

record for 01 April, 1988

i.) Also there is another unavoidable problem, Magneto Ionic Power Coupling. Antenna polarization plays a large role in the success of a long haul DX contact. As a medium frequency RF signal traverses Earth's magnetic lines of force in a perpendicular manner on high and mid latitude paths say between W3 land and SM, higher angle horizontally polarized signals are more readily absorbed than lower angle vertically polarized signals. On other propagation paths on the globe opposite results can be found, i.e., horizontally polarized signals suffer less absorption on a propagation path between VK6 and W6 or S9 and W4.

Magneto Ionic Power Coupling expert NM7M Robert Brown, PhD. has a good educational thread on this bugaboo on the May 2002 Topband Reflector. The thread can read in its entirety by going to this link Topband Reflector May 2002 Archives Layer .

Also an excellent but more technically oriented website covering 160 meter propagation and more is the "HF Propagation Tutorial" by NM7M Bob Brown, Ph.D. and hosted by ON4SKY Thierry Lombry and can be found at: http://www.astrosurf.com/luxorion/qsl-hf-tutorial-nm7m.htm
.

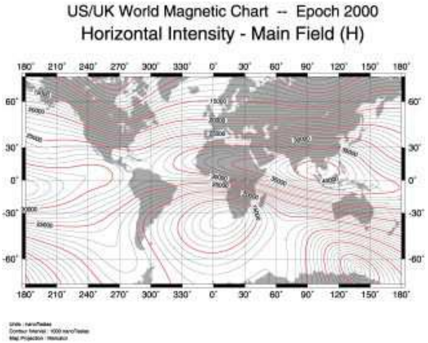
US/UK World Magnetic Chart -- Epoch 2000
Horizontal Intensity - Main Field (H)
180° 210° 240° 270° 300° 330° 0° 30° 60° 90° 120° 150° 180°
60° 30° 0° -30° -60°
Units : nanoTeslas
Contour Interval : 1000 nanoTeslas
Map Projection : Mercator

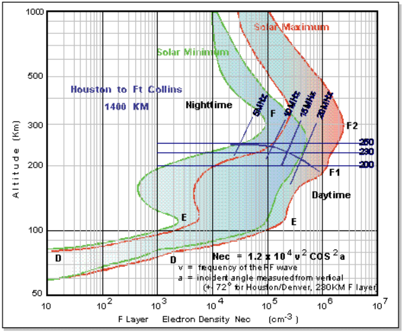

Electron density versus altitude as per section 2.2. This chart shows lines of criticality for different transmission frequencies on the Houston - Ft. Collins circuit. As the ionosphere ionizes in the morning or recombines in the evening, the density passes critical in the altitude range of 220 to 250 Km. Note that the F-Layer density never falls below critical for 5MHz, but does affect 10 through 20MHz. The November 1988 test was somewhere between solar minimum and maximum. (from Tascione, 1988)

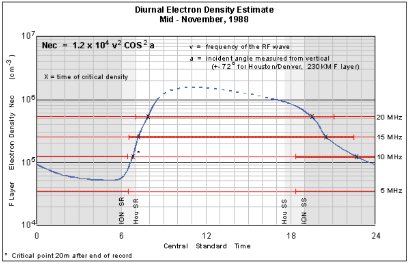

curve of the diurnal change in ionosphere density

The plot shows that the ionosphere goes through a morning period of fairly rapid ionization peaking perhaps around 10am CST (1600UT). Then the process stops and reverses, going through a slower recombination period in the evening and into the night until near ionosphere sunrise the process again reverses and repeats the rapid ionization. This plot in fact compares quite well to published mid-latitude ionosonde data such as Olivier and others, 1988. It is even possible, where circumstances permit, to see more rapid ionospheric variations as described in the following section. Next: Density waves..

Supplemental Records

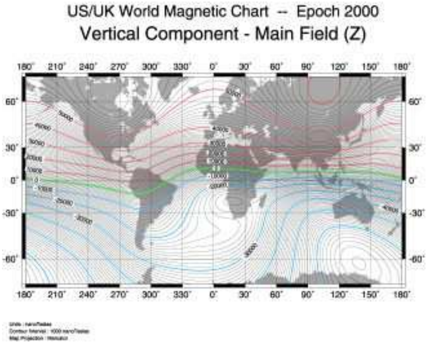
US/UK World Magnetic Chart -- Epoch 2000
Vertical Component - Main Field (Z)
180°
210°
240°
270°
300°
330°
0°
30°
60°
90°
120°
150°
180°
60°
30°
0°
-30°
-60°

Horizontal & Vertical Components Of Earths Geomagnetic Field

US/UK World Magnetic Chart -- Epoch 2000
Declination - Main Field (D)

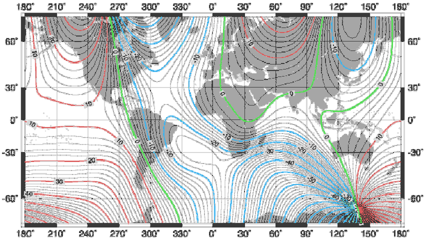
180°
210°
240°
270°
300°
330°
0°
30°
60°
90°
120°
150°
180°
60°
30°
0°
-30°
-60°
10
20
30
40
-10
-20
-30
-40
-50
-60

Units (Declination) : degrees
Contour interval : 2 degrees
Map Projection : Mercator

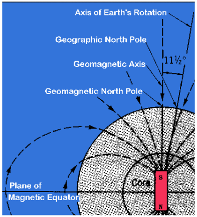
Axis of Earth's Rotation
Geographic North Pole
Geomagnetic Axis
11½°
Geomagnetic North Pole
Plane of
Magnetic Equator
Core
S
N

Earth's Geomagnetic Field

Poor Man's Ionosonde

My reading suggested a solution to the ionization problem described in the previous section. This solution dispensed with the notion that the jump in signal strength was caused by some jump in F layer ionization. It can be found in the relationship between the main parameters affecting propagation: Nec - electron density (cm^-3), v - frequency, and a - incident angle measured from vertical. The expression is as follows:

Nec = 1.2 x 10^4 v^2 cos^2 a

If this expression were true, then it was simple to devise a test to confirm it. If the ionization truly proceeded at a much slower rate, it should affect different radio frequencies at different times, in a regular manner. The beacon, WWV, transmits as previously stated at several frequencies, 2.5, 5.0, 10.0, 15.0, and 20.0 MHz. With this beacon, I could maintain the same basic circuit geometry. In mid November, 1988 I conducted a series of recordings at 5.0, 10.0, 15.0, and 20.0 MHz which, when plugged into the above expression gave me discreet electron densities.

F Elev	Angle a	Frequency			
		5 MHz	10 MHz	15 MHz	20 MHz
250	71	3.34+E4			
240	71		1.24+E5		
230	72			2.59+E5	
220	73				4.26+E5

The signal jump would give me the exact time that the F layer passed the calculated critical density. The following chart is my curve of the diurnal change in density, a sort of "Poor-man's" ionosonde.

electron density, radio frequency, and incident angle which I explained in section 2.3. The answer to the problem presented itself quite simply: My circuit, Ft. Collins to Houston had a fairly fixed geometry (if you assume a constant 230 Km height for the refracting region, close but clearly not constant), and a single radio frequency of 15 MHz. The only variable then is the F layer density which goes critical at about 2.6 x 10^5 cm^-3. Below that density there can be no propagation. As the F layer ionizes, quite slowly perhaps, it passes that critical density and causes the sudden jump in strength seen on my records.

Pretty neat.. If you gather data and think about a thing, you are bound to learn something fascinating. The above relationship quickly suggested to me the experiment described in the following section.

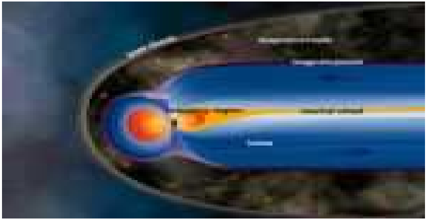

Earth's Magnetosphere

j.) Geological effects such as earthquakes and volcanic eruptions, as well as meteorological effects such as troposphere originating Internal Buoyancy/Gravity Waves (IBGW), stratospheric level Quasi-Biennial Oscillations (QBO) and stratospheric warming (See definition #20 on Stratospheric Warming) have a negative effect on medium frequency RF signals in the form of small to medium increased absorption variations of medium frequency RF signals via the D layer caused by traveling ionosphere disturbances (TID's). Also temperature and moisture discontinuities (frontal inversions) can refract/scatter medium frequency radio signals in unpredictable ways, most notably on high transmitted RF power levels.

k.) The Quasi-Biennial Oscillation (QBO) is a wind shift in the equatorial stratosphere, an oscillation from easterly to westerly and back on the time scale of approximately two years (26 months) and is a source of Internal Buoyancy/Gravity Waves (IBGW's) which create absorptive perturbations in the D and E layers.

l.) A note, the E-valley/F layer ducting propagation mechanism does not exist only during gray line periods. Internal Buoyancy/Gravity Waves (IBGW's) are a source of the ducting mechanism and allow for occurrences of ducting along any propagation path in total darkness. Measurement of the timing of arrival of propagated medium frequency RF signals demonstrates the existence of the ducting mechanism, versus conventional numerous E layer land/ocean surface hops which, would allow for approximately 40 db of attenuation on a North America to Europe propagation path.

Another note! When it comes to 160 meter vertical antenna's you can get a lower take off angle (TOA) from a full 1/4 wave vertical or electrical 1/4 wave tee vertical of 10-20 deg., versus ~30 deg. with the inverted L. However it's a moot point as the night time E layer MUF blocks 160 meter low angle transmitted radio signals from ever reaching the F layer to be propagated. So unlike with high frequency propagation, medium frequency propagation success does not require the lowest of take off angles.

Also higher take off angles of 30-40 deg. via the inverted L are better able to take advantage of the low signal loss E valley-F layer propagation duct mechanism, a form of Chordal Hop propagation.

A solar flux of at least 150 is necessary for routine stable formation of the E Valley/F Layer ducting mechanism. Therefore formation of the duct is less prevalent at the bottom of solar cycle and long haul propagation poorer at solar minimum.

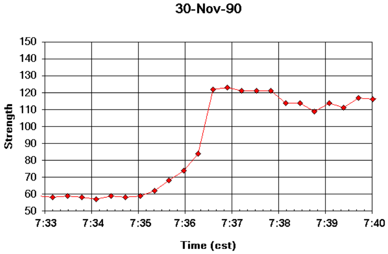

Record for 30 November 1990, detail

At the time I had the idea that the rise in signal strength in the morning, (and drop in the evening) was associated directly with the ionization of the reflecting F layer. As the layer ionized under the rising sun the signal would pick up. The records though could not be interpreted in that manner. As I thought about it, I realized that there were two problems with my theory: First, I noted the curious fact that sunrise on the F layer, and therefore ionization, took place up to one or two hours prior to the morning jump in signal strength (and vise versa in the evening). If the rise in strength was caused by a rise in ionization, why didn't it occur earlier? Secondly, the duration of the jump, sometimes in just seconds was far too rapid. The volume of refracting space in the ionosphere was simply too large to be illuminated by sunrise and ionize so rapidly.

Thinking about this problem sent me to the textbooks and literature. There I learned about the relationship between the

Ionization Problem

Early on in my project I encountered a problem in the interpretation of the data. Most records like the one below for 30 November 1990, showed a very rapid increase in signal strength from near 60 (S0, no reception) to a maximum of 120 or more (S9+, extremely strong).

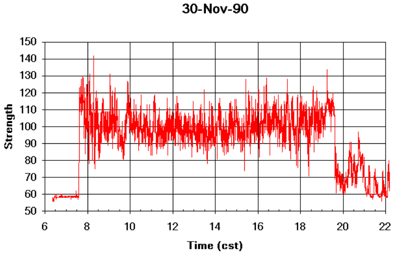

Record for 30 November 1990

The time from nothing to maximum was often under a minute or two, as shown in the detail chart below where in the minute around 7:36 am CST the signal makes its jump. (This also shows some of the tight detail available on the records which looks spiky on the normal 15-hour presentation charts).

m.) Yet another mechanism to deal with that impacts medium frequency radio wave propagation in a negative fashion is the D Layer Mid Winter Anomaly. It is a period of increased medium frequency radio wave absorption at high and mid latitudes occurring in mid winter and is associated with sudden stratospheric warming and the Quasi Biennial Oscillation (QBO).

n.) The HAARP ionosphere research program, earthquakes, volcanic eruptions, thunderstorms, lightning (especially positive cloud to ground strokes), elves, tornadoes, hurricanes and even man made activities such as rocket launches including the space shuttle, are all sources of (IBGW's). Many times I've heard ham's lament that propagation was going to go to crap due to another space shuttle launch, in a sense they are correct.

o.) Another issue facing medium frequency AM broadcast Band DXers and 160 meter operators is lower latitude propagation path absorption due to the Equatorial Ring Current. This phenomenon acts as a repository for precipitated electrons and the end result is unpredictable medium frequency RF signal blockage absorption and refraction. Absorption is similar to higher latitude Aurora absorption.

p.) LF propagation theory is out of my realm from a standpoint of formal education. Alan Melia G3YNK is studying LF propagation and has made some very interesting observations and put forth some fascinating theories.

Here are some interesting website links concerning LF and ELF radio propagation theory.

LF PROPAGATION THEORY INFO BY ALAN MELIA G3YNK (link broken)

PROPAGATION OF LONG RADIO WAVES BY J.A. ADCOCK VK3ACA (link broken)

RADIO WAVES BELOW 22 KC
http://www.vlf.it/

2.) Aurora Oval Blockage, Absorption And Refraction-

The aurora ovals "generally" have a negative impact on medium frequency propagation. If the path over which you are communicating lies along or inside one of the Aurora Ovals, you will experience degraded propagation in one of several forms; strong signal absorption, brief periods of strong signal enhancement, which is mainly caused by tilts in the ionosphere that allow signals to become focused at your location or very erratic signal behavior in the form of strong and rapid fading, etc., caused by a variety of effects such as multi-pathing, anomalous and rapid variations in absorption, non-great-circle propagation, horizontal or side refraction and/or scatter (skewing) due to changes in electron density and polarization changes. (See definition #7. Propagation Path Skewing).

When the Aurora Oval zones are contracted and latitudinally-thin coinciding with low geomagnetic activity, it is possible for a medium-frequency transmitted signal to propagate through the Aurora Oval zone without being heavily absorbed by skirting underneath it.

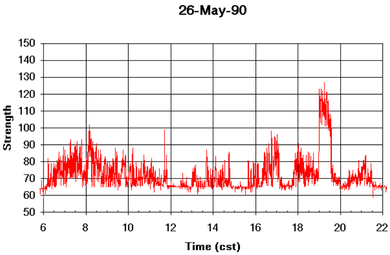

26may90
Major flare and geomagnetic storm, propagation never had much of a chance! Return..

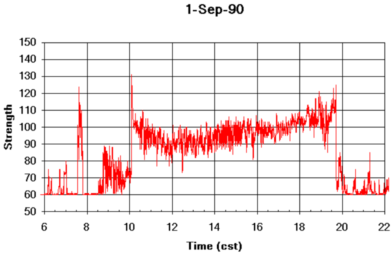

record of 01 September, 1990

This record for 01 September, 1990 indicates that not all geomagnetic activity necessarily destroys propagation. This day with Ap=26 had conditions ranging from quiet to major storm yet the openings remain generally good. Geomagnetic storms in fact can be quite local and do not a priori indicate poor communications. Recall that my recordings only probe a small portion of the ionosphere located about halfway between Ft. Collins, Colorado and Houston, Texas.

Supplemental Records

During periods of very low geomagnetic activity, areas of the Aurora Oval zones may only have a latitudinal thickness of approximately 300 miles. But radio signals reflected from the E layer can travel over distances of as much as 300 to 1250 miles at heights below the ionosphere for low take-off angles of between 10 and 25 degrees. When the geometry is just right, the medium-frequency transmitted signal can literally propagate underneath and through the Aurora Oval zones into the polar ionosphere which is less disturbed and from the polar ionosphere back into the middle latitude ionosphere, without ever coming in contact with the highly absorptive Aurora Ionosphere. This type of propagation is not as rare as you might think and it can provide unusually stable polar region path openings to (TA) Transatlantic and (TP) Transpacific regions. But because the Aurora Oval zone expands and contracts constantly, such conditions often do not last very long. (See definition #3. Equatorial Ring Current).

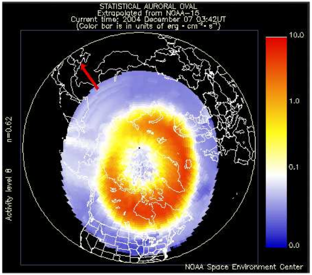
STATISTICAL AURORAL OVAL
Extrapolated from NOAA-15
Current time: 2004 December 07 03:42UT
(Color bar is in units of erg • cm⁻² • s⁻¹)
10.0
1.0
0.1
0.0
n=0.62
Activity level 8
NOAA Space Environment Center

3.) Equatorial Ring Current-

A phenomenon that acts as a repository for precipitated electrons in the vicinity of the magnetic equator. The electrons travel by spiraling around north south magnetic field lines at a frequency called the 'gyro frequency. The end result is lower latitude propagation path medium frequency transmitted RF signal blockage and absorption via the D layer. Absorption is similar to higher latitude Aurora Oval absorption and is inter-related with same.

A reliable gauge for measuring the up to three day lingering post geomagnetic storming medium frequency transmitted RF absorption is the Dst index, measured in nT's. It is an estimated value from Kyoto Japan and is based on a formula. Large negative values after a major geomagnetic storm indicates a high Equatorial Ring Current level. (See definition #2. Aurora Oval Blockage, Absorption And Refraction). Here is a website link to the Kyoto, Japan Dst Index http://swdcwww.kugi.kyoto-u.ac.jp/dst_realtime/presentmonth/index.html and the U.C. Berkeley website link http://sprg.ssl.berkeley.edu/dst_index and a NASA GSFC website link http://sprg.ssl.berkeley.edu/dst_index . (See definition #2. Aurora Oval Blockage, Absorption And Refraction). Another excellent source of a daily Dst figure is at http://www.alan.melia.btinternet.co.uk/latest.htm .

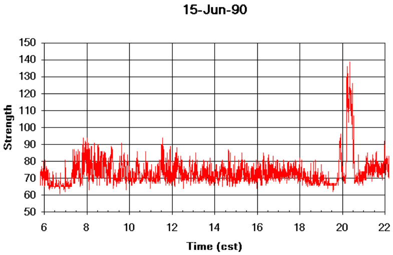

record of 15 June, 1990

Another storm ending on 15 June, 1990 in the summer season has in this record nearly wiped out the opening totally. Ap for the record also a high 47. Similarly, the flare and storm on 26 May, 1990 with an Ap=38 had terrible results for communications.

Unsettled to Active Field Response

Solar activity can cause disruptions of the earth's magnetic field leading to choppy or poor openings. The usual cause of these disruptions is the interaction of charged particles expelled from the sun by flares with the earth's ionosphere. Being charged, they carry magnetic field lines from the sun and distort the local earth field lines. The effects are most pronounced when the sun's magnetic poles are reversed with respect to the earth's. Even when sunspot activity seems low particles can still be ejected through coronal holes and disrupt local conditions.

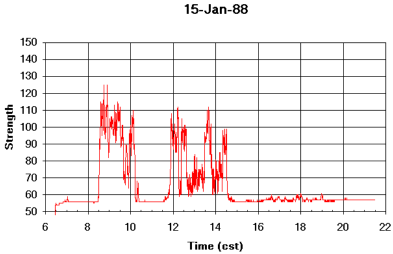

record of 15 January, 1988

This record of 15 January, 1988 was taken on the last day of a major geomagnetic storm. The planetary "Ap" value, (a measure of geomagnetic activity) was a high 45. The openings here are choppy, short lived but still with good signal strengths typical of a winter record.

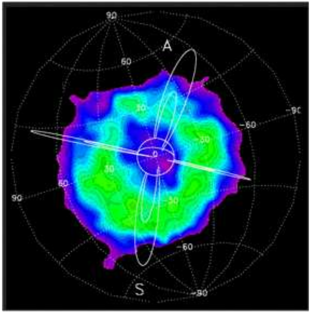

4.) Coronal Mass Ejection (CME)-

A Coronal Mass Ejection is the name given to an ejection of a large amount of matter from the Sun's outer atmosphere or corona. These ejections typically comprise millions of tons of material in the form of charged particles and can be seen because the material reflects sunlight. When one of these ejections is directed towards the Earth (or conversely, directly away from the Earth), it looks like a roughly circular "halo" surrounding the blanked out Sun.

The "Halo CME's" are those CME's which are more likely to impact the Earth than those which are shot out at right angles to the Earth-Sun line. Energetic protons emitted during CME's play a major role in increased day time and night-time D layer absorption of medium frequencies.

Coronal Mass Ejections were once thought to be completely initiated by solar flares. However it is now known that many (CME's) are not associated with Solar Flares but instead with collapsing Solar Filaments. If a (CME) collides with the Earth, it can excite a Geomagnetic Storm if the polarity of the Interplanetary Magnetic Field (IMF) has a negative sign. We must be vigilant in watching for geo-effective (CME's), in order to not be caught by surprise with a seemingly sudden and unexpected Geomagnetic Storm. (See definition #6. Solar Filament). (See definition #11. Geomagnetic/Ionosphere Storm). (See definition #16. Solar Flare).

Coronal Mass Ejections are not random meaningless eruptions but instead a process by which the Sun expels complex magnetic signatures enroute to changing its magnetic polarity or said a different way the swapping of the Sun's magnetic poles. Basically the Sun swapped its magnetic polarity at the peak of present solar cycle 23 somewhere between July 2000 and December 2001. The next polarity swap will occur during solar cycle 24 somewhere around 2012.

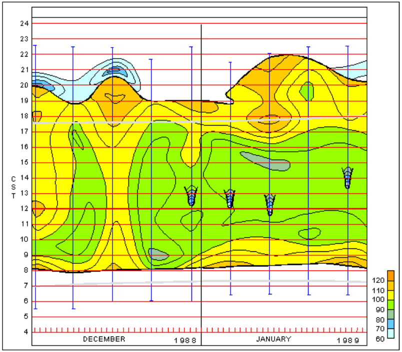

Map for the winter 88/89 which shows the typical short openings and strong signals. Major flares on 30 Dec, 88, 1802 UT (1202 CST); 06 Jan, 90, 1805 UT (1205 CST); 13 Jan, 89, 1734UT (1134 CST); and 27 Jan, 89, 1917UT (1317 CST). Solar flux during the period was understandably high hovering between 154 and 170 in early December, jumping to 180 to 290 through January 89.

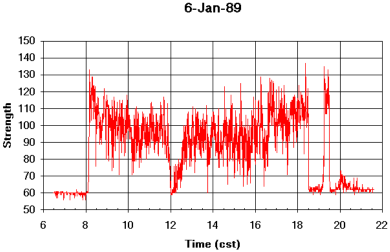

Major flare at 1805UT, (1205 CST). Also note the evening signal dropout and return.

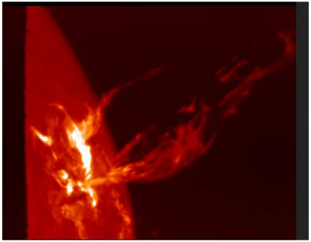

5.) Coronal Hole-

The Corona is not part of the Sun's surface. It is instead part of the Sun's atmosphere, much like Earth's troposphere. Coronal Holes are low density areas associated with open magnetic field lines and are found near the Sun's poles at the bottom of a sunspot cycle and everywhere during a cycle maximum. A Coronal Hole is a dark region where a breakdown in the magnetic field structure in the solar corona has occurred. From these regions stream the high velocity solar wind that is a source of geomagnetic storming on Earth.

Coronal Holes occur most often on the downside of a solar cycle and their absence at the bottom of a solar cycle and at the beginning of the next, allow for the best medium frequency radio propagation conditions. Many think it's the lower solar flux values seen at the bottom of a solar cycle that accounts for improved propagation conditions but it's actually pretty much a lack of Coronal Holes and geomagnetic storming. (See definition #11. Geomagnetic/Ionosphere Storm).

One thing to keep in mind is that the high velocity solar wind stream emanating from a Coronal Hole is a neutral phenomenon with respect to the Bz (magnetic component) of the Interplanetary Magnetic Field (IMF). If the Bz component is negative (southward) prior to arrival of the solar stream, there will exist a tendency to see a larger swing negative after the disturbance arrives. If the Bz component is positive (northward) prior to arrival of the solar stream, there will exist a tendency to see a larger swing positive after the disturbance arrives.

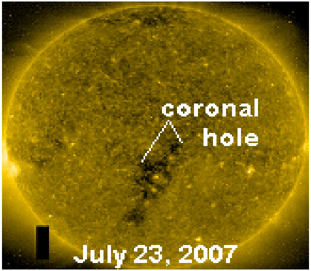

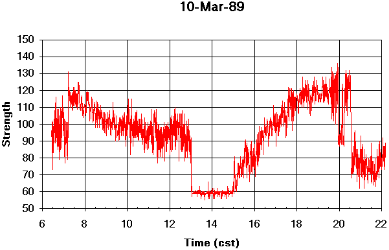

See «http://www.ips.oz.au/background/richard/power_1989.shtmll » for a description of the event and its effects in Quebec. Next: Unsettled to active field response..

Supplemental Records

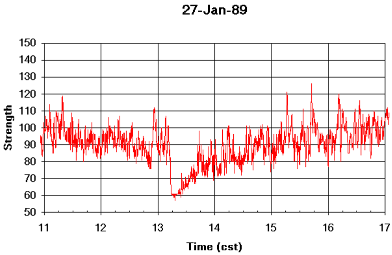

record of 27 January, 1989, detail

The above detail shows more clearly the sudden drop followed be a rise back to the former signal level. This was a pretty typical flare, strong but not truly large. Another example of a typical flare comes from the record of 06 January, 1989. The winter and spring of 1989 was a very active period for flares and I caught them repeatedly on many records as the map for December 89 / January 90 shows. On the 10th of March, 1989 there occurred a truly powerful outburst which blocked radio communications for several hours and the geomagnetic storm that followed three days later was so severe that it knocked out the power grid in Quebec, Canada, and caused wonderful auroral displays as far south as Florida.
record of 10 March, 1989

6.) Solar Filament-

A relatively cool and dense ribbon of gas held together by solar magnetic fields. From Earth they usually appear as relatively dark lines across the face of the Sun. At times the magnetic lines holding the filament open up creating a tremendous eruption similar in size and impact of a Coronal Mass Ejection (CME). (See definition #4. Coronal Mass Ejection). (See definition #11. Geomagnetic/Ionosphere Storm).

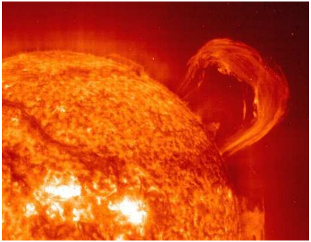

7.) Correlation Of Energetic Protons, Solar Flux and Ap & Kp Indices With Medium Frequencies-

I've been observing energetic proton levels, as well as the Ap & Kp indices for 35 years and see a direct correlation between high energetic proton levels above 10 MeV (10+0) and poor propagation on high and at times mid latitude medium frequency paths at day AND night, where as the A & K indices don't as readily correlate. (See paragraph three of definition #2. Aurora Oval Blockage, Absorption And Refraction) and (definition #7. High Latitude Path Skewing) for a further explanation on the lack of correlation of Ap & Kp indices with medium frequency propagation conditions.

High solar flux values are "incorrectly" considered to be detrimental to medium frequency signals both domestic and TA/TP, as more absorption can be present as the transmitted signal makes two trips through the D layer, near sunrise and sunset. However most medium wave frequency RF signals in excess of 3100 miles are propagated via the E valley/F layer ducting and/or Chordal Hop/Pederson Ray propagation mechanism and a high solar flux value ensures a strong E valley/F layer duct mechanism. Actually a solar flux of at least 150 is needed for a consistent E valley/F layer ducting mode.

The main reason that medium frequency radio propagation "seems to be better" at the bottom of a sunspot cycle is not so much due to lower solar flux levels BUT due to much less geomagnetic activity.

Keep in mind though that the 10.7 cm (2800 MHz) solar flux index is not a "reliable" gauge of ionization in our atmosphere, as the energy of photons at this frequency is to low on the order of one million times. However most are used to solar flux and sunspot number and it's a hard habit to break. A better indicator is the inter-related background X-ray flux. (See definition #1 paragraphs e & f).

An elevated energetic proton flux level greater then (10+0) creates noticeably increased winter time day and year round night time D layer absorption of medium wave frequencies, especially on high latitude propagation paths but it can also

Flare Response

Flares result from the sudden release of tremendous amounts of energy from the sun's surface. Some eight minutes after a flare bursts on the sun, high energy ultraviolet radiation impacts the earth ionizing all the ionosphere layers in the process. While this might be expected to enhance F layer propagation, the D layer density quickly becomes opaque to the signal causing a sudden fadeout. Depending on the duration of the flare, the fadeout will last from minutes to hours and then as the ions recombine there follows a logarithmic rise back to normal. The following record of 27 January shows a classic flare signature against a clean winter record.

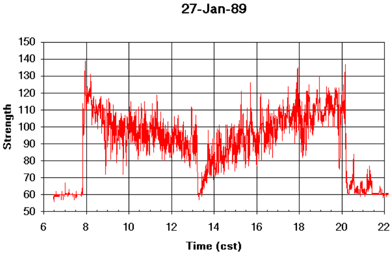

record of 27 January, 1989

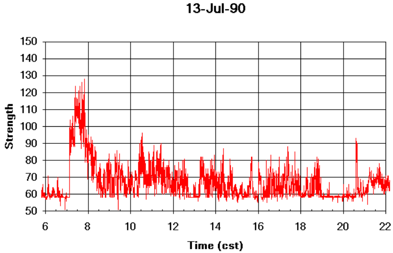

Summertime blues, solar flux good at 153, Ap=10, pretty quiet day but with lousy reception.

negatively impact mid latitudes, depending on the intensity of the event.

Elevated energetic proton events too small to be categorized as a Polar Cap Absorption event (PCA) can still impact high and at times mid latitude medium frequency propagation paths in the form of excessive D layer absorption.

((((Note, high latitude medium frequency radio propagation paths can still be disturbed for days and up to weeks, following the end of an official >10 MeV (10+0) proton event.))))

GENERAL GUIDELINES CONCERNING CORRELATION OF PROPAGATION INDICES TO ACTUAL MF PROPAGATION CONDITIONS-

NOTE!!! The propagation indices "interpretations" are my personal intellectual property. Therefore the propagation indices interpretations contained herein is copyrighted © 1988-2013 by Thomas F. Giella, W4HM, all rights reserved. Reproduction of information herein is allowed without permission in advance as long as proper credit is given.

All 13 of the following indices have to occur as described below in order to see the best global medium frequency (MF) radio wave propagation possible.

1.) Dropping geomagnetic field indices numbers are better, Kp of 0 best.

2.) A daily sunspot number under 100, under 70 best.

3.) A daily sunspot number no higher then the 100 for routine stable formation of the E Valley/F Layer ducting mechanism.

4.) Previous 24 hour Ap index under 10, fewer than 7 for several days consecutively are best.

5.) Previous 3 hours Kp index fewer than 3 for mid latitude paths, fewer than 2 for high latitude paths, 0 for several days consecutively is best.

6.) Energetic proton flux levels no greater than 10 MeV (10+0).

7.) Background x-ray flux levels of A0 for several days consecutively.

8.) No current STRATWARM alert.

9.) Interplanetary magnetic field (IMF) Bz with a (positive number) sign, indicates a lesser chance of high latitude path aurora absorption/unpredictable refraction or scattering of medium frequency RF signals, when the Kp is above 3.

10.) A -10 or better towards a positive number Dst index during the recovery time after a geomagnetic storm, as related to the equatorial ring current. A positive number is best.

11.) Galactic cosmic rays decrease to -3 units below zero and trending towards zero.

12.) Energetic electrons no greater than 2 MeV (2+0).

13. A solar wind speed of less than 300 km/s for several days consecutively.

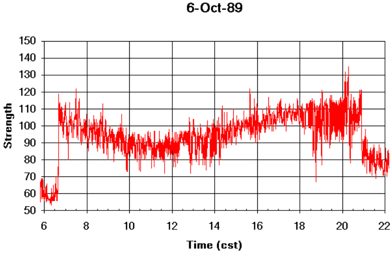

"Ideal" recording with winter response, good opening, midday sag, and long evening period with strong signal reception.

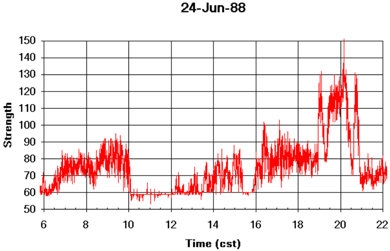

Major flare at 1609UT (1009 CST).

average strength for the midday period 1700 - 2000 UT

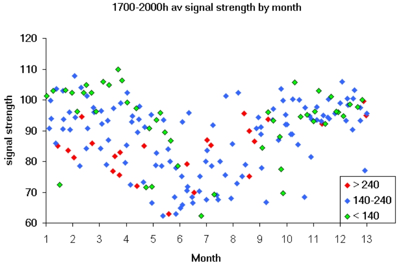

The following section shows the signature of a solar flare since finally, that’s what all this is about!

Supplemental:

GENERAL GUIDELINES CONCERNING CORRELATION OF PROPAGATION INDICES TO ACTUAL HF PROPAGATION CONDITIONS-

NOTE!!! The propagation indices "interpretations" are my personal intellectual property. Therefore the propagation indices interpretations contained herein is copyrighted © 1988-2013 by Thomas F. Giella, W4HM, all rights reserved. Reproduction of information herein is allowed without permission in advance as long as proper credit is given.

All 14 of the following indices have to occur as described below in order to see the best global high frequency (HF) radio wave propagation possible.

1.) Dropping geomagnetic field indices numbers are better, Kp of 0 best.

2.) A daily sunspot number of 150 or higher, 200 or higher best.

3.) A daily sunspot number of greater than 100 for routine stable formation of the E Valley/F Layer ducting mechanism.

4.) Previous 24 hour Ap index under 10, fewer than 7 for several days consecutively are best.

5.) Previous 3 hours Kp index fewer than 3 for mid latitude paths, fewer than 2 for high latitude paths, 0 for several days consecutively is best.

6.) Energetic proton flux levels no greater than 10 MeV (10+0).

7.) Background x-ray flux levels greater than B1 for several days consecutively, greater than C1 best.

8.) No current STRATWARM alert.

9.) Interplanetary magnetic field (IMF) Bz with a (positive number) sign, indicates a lesser chance of high latitude path aurora absorption/unpredictable refraction or scattering of medium frequency RF signals, when the Kp is above 3.

10.) A -10 or better towards a positive number Dst index during the recovery time after a geomagnetic storm, as related to the equatorial ring current. A positive number is best.

11.) Rising positive T index number. The T Index tracks with the F2 layer critical frequency (foF2) and sunspot number (SSN) and indicates the capability of the F2 layer to refract RF signals.

12.) Galactic cosmic rays decrease to -3 units below zero and trending towards zero.

13.) Energetic electron flux levels no greater than 2 Mev (2+0).

14. A solar wind speed of less than 300 km/s for several days consecutively.

8.) E Valley/F Layer Propagation Ducting Mechanism/Chordal Hop Propagation-

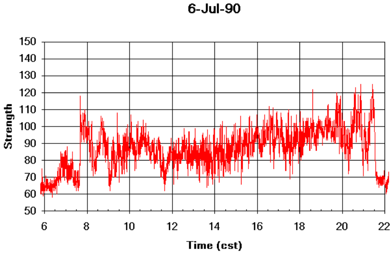

record taken on 06 July, 1990

This chart shows a fairly typical (if there is such a thing) "summer" record taken on 06 July, 1990. The day had strong solar flux at 209, and the geomagnetic field was moderately quiet. Still, the signal was weak most of the day with lots of noise and levels only in the S3 - S5 (80 - 90) range. A week later on 13 July the opening made a brief appearance in the morning and then headed south all day. Good day to go fishing.

I have taken averaged hourly signal strengths for all 200 records and put them on a spreadsheet to make further analyses. The following graph shows the average strength for the midday period 1700 - 2000 UT (10am to 1pm CST) averaged together and plotted by month. That is, all February data regardless of year together, all June and so forth. Although some years may be better overall, you can clearly see the seasonal signature with good autumn to spring signal and poor summer reception.

generally the records are choppy and noisy, fading is a far bigger problem.

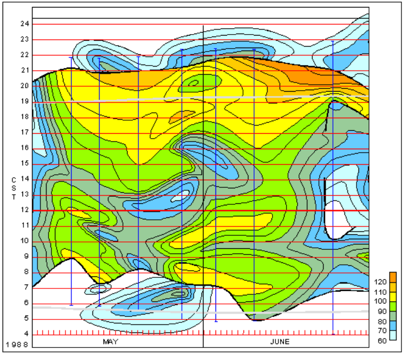

period of June and July, 1988

On 24 June of this period there occurred a major flare at 1609UT (1009CST) which closed the window for several hours.

Antenna polarization plays a large role in the success of a long haul DX contact. As a medium frequency RF signal traverses our planets magnetic lines of force in a perpendicular manner on high and mid latitude paths say between W3 land and SM, higher angle horizontally polarized signals are more readily absorbed then lower angle vertically polarized signals. On other paths on the globe opposite results can be found, i.e. horizontally polarized signals suffer less absorption on a propagation path between VK6 and W4.

You would expect a true long path QSO on 160 to be theoretically possible but improbable on most paths during any season. However a G to VK long path might be possible if the E Valley/F layer ducting propagation mechanism or the Chordal Hop propagation mechanism is involved. A 160 meter signal can traverse a daylight path via these propagation modes if the transmitted signal enters/exits at each end of the path at or near sunrise/sunset when the D layer ionization is weak (ionosphere tilting).

There is an upward tilt of ionosphere layers towards the east at sunrise. As a result, signals coming from the west are refracted upward at steeper angles and are therefore heard better on higher angle antennas. The opposite is true at local sunset.

A note though, the E-valley/F layer ducting propagation mechanism does not exist only during gray line periods. Internal Buoyancy/Gravity Waves (IBGW's) are a source of the ducting mechanism and allow for occurrences of ducting along any propagation path in total darkness. Measurement of the timing of arrival of propagated medium frequency RF signals demonstrates the existence of the ducting mechanism, versus conventional numerous E layer land/ocean surface hops.

The majority of the time medium frequency RF signals in excess of approximately 3200 miles propagate via the E Valley/F Layer propagation mechanism or via the Chordal Hop (mostly on HF near local sunrise and sunset) propagation mechanism. Typically the majority of transmit antenna's radiation must be focused between 40-60 deg. to enter the E Valley/F Layer duct. (See definition #23.) The Gray line/Gray line Propagation).

A solar flux of at least 150 is necessary for routine stable formation of the E Valley/F Layer ducting mechanism. Therefore formation of the duct is less prevalent at the bottom of solar cycle and long haul propagation poorer at solar minimum.

You won't see this in any mainstream books on radio wave propagation but the vast majority of the time MF and HF signals travel around the world not by successive hops off of the F/F2 layer and the surface of the earth, whether it be landmasses and/or the ocean, but either between the E and F/F2 layer. This is especially true on long paths.
Well Known Chordal Hop LP Routes Courtesy Of Larry Duncan K4WLS-

Here are some well known Chordal Hop LP routes from the East Coast and Mid-West:

Late Afternoon, Mid-February to Mid-March - Western Australia and beyond, and Southern Malaysia: Predominately 20M.

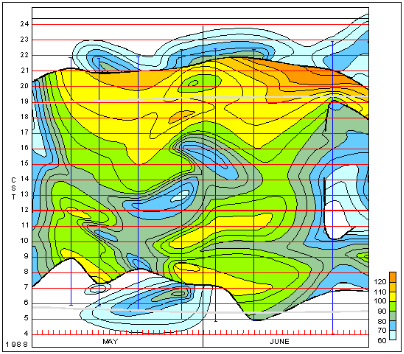

This map shows the period of March and April, 1988. In it you see the strong morning and evening signal (in browns) and the midday sag in yellow to green. On the right the signal for the record of 22 April becomes poor, (blues). This was an active day with geomagnetic disturbances disrupting the propagation. Also in this map you can see typical evening sporadic low-level propagation after the main opening has ended, especially in early March.

In contrast to the above "winter" view, the following map shows the doldrums of summer. It is for the period of May and June, 1988. I characterize the propagation in summer as generally longer duration, but with lower signal strength and much greater day to day variability. It is not unusual to have a good strong opening for an hour or two in the evening, but

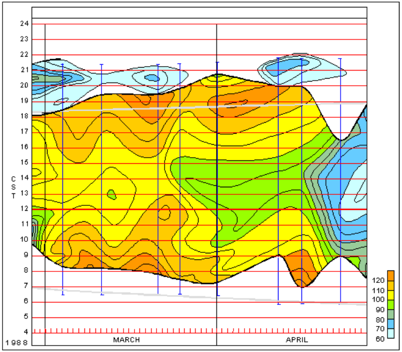

0700-1000 Local, Early to Late Summer - Eastern and Southern Africa, and Indian Ocean: 20, 15, 17, 12, and 10M (17 through 10M depending on Solar Flux).

0800-1000 Local, Early Fall - Western Australia and S.E. Indian Ocean: 20M

0500-0700 Local, * December - Malaysia, Indonesia, and S.E. Asia: 40M

Sunset to 1 Hour Before, Fall to Mid-December - Middle and Eastern Asia: 40M

0700-0800 Local, Mid-December - Middle East: Predominately 20M

Sunset to 1 Hour Before, Mid-December - Northern Middle and Eastern Asia: 20 and 40M (20M depending on Solar Flux).

* Sporadically as late as early March

If one is lucky enough to be on the receive end of a ducted medium frequency signal due to an IBGW or two, a change in the vertical and/or horizontal electron gradient will allow the RF to drop out of the duct at your QTH. Galactic Cosmic Rays also play a role in where an RF signal drops out of the duct.

A note, high solar activity in the form of increased ionization created by ultraviolet and X-ray radiation, can fill in the E Valley/F Layer ducting region with medium frequency absorptive ionization and interfere with the E Valley/F Layer ducting mechanism. In a sense the E/F layer duct is shut down and the medium frequency RF signal can only propagate between the E layer and land/ocean surface, at a higher angle and with more signal loss. This closing of the duct can be reciprocal on each end of the propagation path or one way only. (((((When closing of the duct occurs the advantage of a low angle vertical radiator is lost, with a higher takeoff angle horizontal dipole making the contact still possible, albeit maybe weaker.)))))

Medium frequency radio waves possess elliptical polarization, with the signal splitting into ordinary and extra-ordinary rays. These rays can propagate in or out of phase, more often out of phase. The out of phase extra-ordinary ray represents a 50% power loss on the receive end of a propagation path.

As follows is a recent experience I had in Florida with this propagation mode on 160 meters.
I began listening for DX on 160 meters at 5:00 pm EST this evening February 05-06, 2006. I was watching the OH2AQ spot and the stations in ME to VA were working stations in Europe and Africa. I could hear the stateside stations real well but not a peep out of any DX.

Then right at my local sunset which was at 6:15 pm EST it was like flipping a switch as all the DX stations just showed up. It was a classic example of the E Valley-F Layer ducting mechanism propagation mode with the duct opening up right over Florida as the ionized layers changed height with the arrival of the gray line terminator.

Once the DX showed up I heard oodles of CW DX stations including VQ9LA and 6W/G4WFQ who were 55 on the receive loop. MM0SJH, G3FPQ and I7RIZ showed up on phone between 1841 and 1849 kc and they were 57. I didn't bother to work anything this time as the only countries I heard

Seasonal Variations

The record presented in the previous section represented a nearly "ideal" winter record. I found that winter openings were generally dependable with a distinct jumps in the morning and drop in the evening, and good strong reception all day. 06 October, 1989 was another good day. It is possible to "map" the data from week to week in order to visualize how the propagation openings change with time. The following chart shows a map of signal strength contoured against time. Local Central Standard Time is on the Y-axis and calendar time on the X-axis. Each day starts at the bottom and moves up, the next day starting again at the bottom next to (to the right of) the prior one so the calendar flows to the right. Additionally, the Gray curves running along the upper and lower portions of the map indicate local Houston sunrise and sunset respectively.

is an asymmetry to the opening as well with a longer period of strong signal in the evening than in the morning. I attribute this to the geometry of the Houston-Ft. Collins circuit with Ft. Collins being to the west. By the time the F layer ionizes in the morning at a point halfway between Houston and Ft. Collins, the lower D layer is already ionized, conversely, in the evening there is still a couple of good hours remaining for the F layer after the D has dissipated.

Base level, (S0) on this record is 58, the solar flux that day was quite high at 239 and the geomagnetic field was quiet. The resulting record is pretty close to ideal which is actually quite unusual. In all, the project ran for nearly four years from just after sunspot minimum to past sunspot maximum and resulted in nearly 200 weekly records. The next section provides additional views giving a flavor of the variability possible, and why.

that I've never worked were VQ9LA and 6W/G4WFQ and I couldn't break the pileups with 100 watts.

9.) Electron Gyro Frequency Absorption-

Unfortunately medium frequencies fall within or very near the electron gyro-frequency which is in the approximate range of 630 to 1630 kHz and of course the AM broadcast band and 160 meter band is very close to these electron gyro frequencies. There is a direct correlation between the strength of Earth's magnetic field lines and electron gyro frequencies.

Basically, the electron gyro frequency is a measure of the interaction between an electron in the Earth's atmosphere and the Earth's magnetic field. The closer a transmitted medium frequency carrier or sideband wave frequency is to the electron gyro frequency, the more energy that is absorbed by the gyro (spinning) electrons from that carrier wave frequency. This is especially true for medium frequency signals traveling perpendicular to the Earth's magnetic field, meaning high latitude NW and NE propagation paths. Unfortunately this form of medium frequency signal absorption is ALWAYS present.

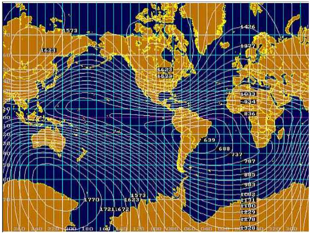

Electron Gyro-frequency Map From Prop lab Pro Software

10.) Medium & High Frequency Radio Signal Propagation Path Skewing-

Medium & High frequency radio signal propagation path skewing occurs due to changes in the "horizontal" electron gradient. Put in simple layman's terms the transmitted RF signal will "always" seek to propagate along the path with least absorption, which almost always means via a darkness path.

As an example a medium frequency signal (say 1830 kc) transmitted from Norway to New England, which is via a polar great circle path, will be directly absorbed most of the time by the Aurora Oval, with the remaining signal skirting south and then west on the darkness path via the E layer, arriving in New England from say the SE rather then the expected NE path.

Another example is a high frequency (say 28400 kc) signal transmitted from Florida that arrives from the SE at Clipperton Island (FO/TX5), which is located in the Eastern North Pacific Ocean near 11 degrees north latitude and 110 degrees west

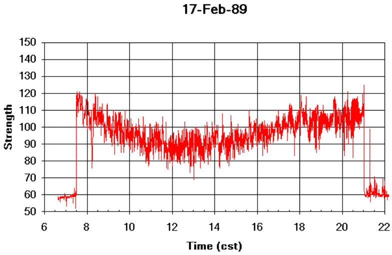

The above record shows a fairly "classic" response for a 15 hour run. In all my runs regardless of duration I collected 3200 samples only. That was an expedient to simplify the programming, keep the record within the memory limits of a Commodore 64, and have a simple multiple of the high-resolution screen's 320 pixel width. For a nominal 15 hour run that translates to about one record each 17 seconds which is more than sufficient for the project. In addition, I actually sampled continuously between saves and read the average of many readings for each saved record. Even then, there is considerable fading and short second to minute scale noise which accounts for the "grassy" surface.

Note also two important features of the record, the very sharp discontinuities at the beginning and end of the "opening", and the general sag or lower strength of the signal at mid-day. More on the discontinuities later, the sag itself is due to increased absorption by the lower D layer at it reached full ionization near midday and dissipates by the afternoon. There

Daily Record

My study utilizes the WWV time standards beacon at Ft. Collins, Colorado. That transmitter emits at 2.500, 5.000, 10.000, 15.000, and 20.000 MHz. I monitor the 15 MHz signal which provides a strong signal at my station in Houston, Texas. 10MHz openings last longer into the night but would be useless for flare detecting at that time as the sun has set! I use a Yaesu FRG7700 receiver which is not terribly expensive and extremely stable once it has warmed up.

The interface between the receiver and computer uses a voltage to ohms converter which proved very acceptable but, being a resistance device, was susceptible to temperature induced drift. In Houston that was generally only a problem in the spring or fall where large changes in day to day temperature are typical in the Gulf of Mexico's battle against the continent for dominance of the climate of the region. I made occasional calibrations as necessary to the interface and kept the base level within a narrow range. The following list gives the correspondence between the digital signal strength readings presented on the charts and the radio's S-meter output:

S-meter	Computer
S0	60
S3	79
S5	90
S7	101
S9	112
+10db	132
+20db	157

record of 17 feb, 1989

longitude. This relatively short path contact gives the appearance of the double hop Sporadic E (Es) propagation mode, which frequently occurs during a low point in a solar cycle, when actually it is a skewed F2 layer propagation mode.

A general east-west band of high (25000-35000 kc) maximum usable frequencies (MUF's) exist north and south of the geomagnetic equator http://www.spacew.com/www/realtime.gif . These bands allow for the existence of the north-south propagation mode called Trans Equatorial Propagation (TEP). In any event the contact between Florida and Clipperton Island was made via the horizontal gradients that existed in the band of high MUF's north of the geomagnetic equator.

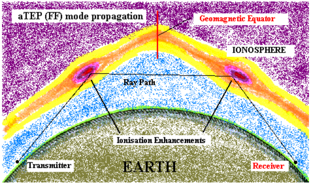

Side View Of High MUF Bands North & South Of The Geomagnetic Equator

By the way skewed propagation paths are the norm rather than the exception on medium frequencies, especially past approximately 3200 miles.

11.) Geomagnetic/Ionosphere Storm-

A worldwide disturbance of the Earth's magnetosphere and or ionosphere, induced by direct connection to the Sun's Interplanetary Magnetic Field (IMF), distinct from regular diurnal variations. Basically it's a precipitation of electrons trapped within our magnetosphere, as the electrons collide. The end result is a reduction of the MUF of the F2 layer. (See definition #3. Equatorial Ring Current). (See definition #4. Coronal Mass Ejection). (See definition #5. Coronal Hole). (See definition #6. Solar Filament).

Geomagnetic Storm Levels

Planetary K Indices	Geomagnetic Storm Level
K = 5	G1 Minor
K = 6	G2 Moderate
K = 7	G3 Strong
K = 8	G4 Severe
K= 9	G5 Extreme

Active K = 4
Unsettled K = 3
Quiet K= 0, 1, 2

A= 100-400 Severe
A= 50-99 Major
A= 30-49 Minor
A= 16-29 Active
A= 8-15 Unsettled
A= 0-7 Quiet

K- 0= A- 0
K- 1= A- 3
K- 2= A- 7
K- 3= A- 15
K- 4= A- 27
K- 5= A- 48
K- 6= A- 80
K- 7= A- 140
K- 8= A- 240
K- 9= A- 400

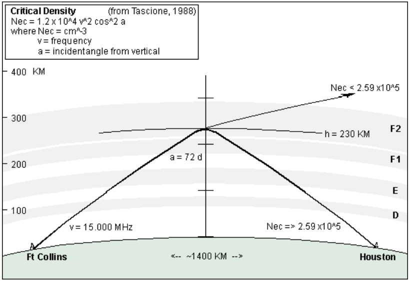
Critical Density (from Tascione, 1988)
Nec = 1.2 x 10^4 v^2 cos^2 a
where Nec = cm^-3
v = frequency
a = incidentangle from vertical
400 KM
300
200
100
Nec < 2.59 x10^5
h = 230 KM
F2
F1
E
D
a = 72 d
v = 15.000 MHz
Nec => 2.59 x10^5
Ft Collins
<-- ~1400 KM -->
Houston

Graphical presentation of the relationships between the frequency, ionosphere electron density, and propagation incident angle measured from vertical are given in the expression:
Nec = 1.2 x 10^4 v^2 cos^2 a ...(from Tascione, 1988)

For the given circuit parameters:
v = 15.000 MHz
a = 72 d ...(h = 230 KM)

Then the Critical ionosphere density from the formula is
Nec = 2.59 x 10^5 cm^-3.

For densities lower than that value the signal will pass out of the ionosphere into space. For densities above that value the signal will continue to propagate, but with increasing absorption on this circuit.

"critical" density needed to return reflections. The relationship between the three parameters (Nec = electron density (cm^-3), v = frequency, and a = incident angle measured from vertical) is given by the following expression, (Tascione, 1988):

Nec = 1.2 x 10^4 v^2 cos^2 a

It should be noted that this expression is for the critical situation only. Where the angle is approximately known based on the circuit geometry and the frequency is known precisely, the density calculated will be the lowest density that, below which, no return (reception) will be possible. A radio wave of a given frequency therefore will continue to travel upwards into an ionosphere layer until it reaches the elevation where the density rises to the "critical" level (Nec) and the wave will be turned back to earth.

For most HF radio communication the main reflecting layer is the F layer situated between 200 to 300 KM altitude. The D and E layers are generally too weakly ionized to be of much importance except during the day over shorter distances. Ionization, although falling, persists through much of the night in the F layer as well. Intense ionization of the F layer causes it to warm and split into a second, higher, level during the day, returning to a single layer at night. There are also seasonal variations on the theme. The E layer, about 100-150 KM up, can persist into the night to some extent and in addition there are occasional "clouds" or masses of E level ionization which can drift causing sporadic radio openings when the right geometries exist. The D layer at about 80 to 90 KM is only useful for propagation during the day recombining quickly after sunset. In the negative sense, it absorbs signals passing through it to or from the F layer.

Solar Radiation Storm Levels

Flux Level of > 10 MeV Particles	Solar Radiation Storm Level
10	S1 Minor
102	S2 Moderate
103	S3 Strong
104	S4 Severe
105	S5 Extreme

Medium Frequency Radio Blackout Levels

Peak X-Ray Level and Flux	Radio Blackout Level
M1 and (10-5)	R1 Minor
M5 and (5 x 10-5)	R2 Moderate
X1 and (10-4)	R3 Strong
X10 and (10-3)	R4 Severe
X20 and (2 x 10-3)	R5 Extreme

((((Note! Unfortunately elevated Kp indices of as little as a 3 will create absorptive conditions for medium frequency signal propagation on higher propagation paths)))).

Initial phase of a geomagnetic storm is that period when there may be an increase of the middle latitude horizontal intensity.

Main phase of a geomagnetic storm is that period when the horizontal magnetic field at middle latitudes is generally decreasing.

Recovery phase of a geomagnetic storm is that period when the depressed northward field component returns to normal levels.

By the way effects of the solar wind on the magnetosphere decreases as we approach the Summer/Winter solstice and increase at the Fall/Spring Equinox. Why? Basically it's the orientation of Earth's magnetic field with respect to the Interplanetary Magnetic Field within the Solar Wind. When solar material and shock waves reach Earth their effects may be enhanced or dampened depending on the angle at which they arrive. http://science.nasa.gov/headlines/y2001/ast26oct_1.htm?list101234.

The Wang-Sheeley Interplanetary Magnetic Field (IMF) Model is used to predict Sun's IMF polarity. When the polarity of the IMF is negative a visible mid latitude Aurora display is likely as a Coronal Mass Ejection (CME) strikes the Earth's magnetic field.

Radio Wave Propagation

Because the different levels in the ionosphere are electrically conducting, they interact with electromagnetic radio waves bending them in much the same manner that glass or water bends light. The refractive index of a layer in the ionosphere depends on its electron density and has an effect on waves according to their wavelength. Levels with higher electron densities (higher altitudes where recombination is slower) have lower refractive indices. As a wave passes up from below into such layers they get refracted, or bent back down towards the lower electron density layers. The amount of bending depends on the incident angle and the frequency of the wave. Higher frequencies are less affected. Finally, in passing through conducting layers, energy in the wave vibrates the electrons thus transferring some energy into the layer and being partially absorbed in the process.

For much of the radio spectrum the frequencies are simply too high to be refracted back to the earth given the range electron densities which exist and propagation angles required. This is not true in the HF band of frequencies between a few hundred hertz and about 30 megahertz. There a delicate balance exists where, depending on the natural variations in density of the ionosphere, radio waves may be either refracted back to the earth or passed out into space. For the case of vertical incidence and a given ionospheric density there is a critical frequency above which no reflection will be received. Inversely, you can say that for a given frequency there is a

The early study of the ionosphere involved measuring its density variations and electrical properties. Most work was done in conjunction with radio studies and radio waves remain the primary way to probe its structure. Studies focus on its variability and the nature and causes of disturbances. In the process much has been learned about the solar-terrestrial connection. With a communications receiver and an inquiring mind, anyone can reach above the sky and touch the near-space environment.

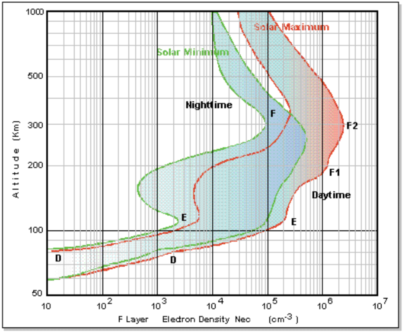

Daytime to nighttime electron concentrations which constitute the midlatitude Ionosphere. Graph with variation between solar minimum and solar maximum curves. (from Tascione, 1988)

12.) Geological/Meteorological Effects On Medium Frequency Propagation-

Geological effects such as earthquakes and volcanic eruptions, as well as meteorological effects such as Troposphere originating Internal Buoyancy/Gravity Waves (IBGW's), Stratosphere level Quasi Biennial Oscillations (QBO) and warming (STRATWARM) have a negative effect on medium frequency RF signals in the form of small to medium increased absorption variations of medium frequency RF signals via the D layer, due to traveling ionosphere disturbances (TID's).

Also temperature and moisture discontinuities involved with cold frontal inversions and air mass triple points involved with extra-tropical low pressure systems can refract, diffract or scatter medium frequency radio signals in unpredictable ways, most notably on high transmitted RF power levels. This is another concept that a fellow Physicist and expert in optics took me to task over.

As far as medium frequency refraction it's more significant at say 3000 kc, then 1850 kc or 1500 kc. But it's also more noticeable with higher transmitted RF powers, i.e. WSAI 1530 50 KW and even more so with BSKA 1521 KC 1000 KW and now defunct 2000 kc region 100 KW marine stations.

We know that the medium frequency spectrum is defined as 300-3000 kc but the differences in refractive properties between 300 and 3000 is very significant At 3000 kc refraction is a good description, on 160 scattering, at 300 kc diffraction.

Using the strictest definition of RF refraction, its effect on 160 meters is small but it has been measured by government researchers as significant enough to impact 160 but near the air mass triple point. In my opinion scattering is actually the more consistent propagation medium for 160 meters along a cold front, away from the extra-tropical cyclone center.

However the temperature and moisture discontinuities in the vicinity of a triple point air mass structure such as seen with a mature extra-tropical cyclone is very complex and fluid. The NW quadrant of the extra-tropical cyclone is the location that the original government researchers identified as the region of existence for the complex temperature/moisture discontinuity structure that allows for refraction of RF signals as low as 1500 kc. I have not been successful at garnering data from the federal government that can be released to the general public. NOAA has been similarly stymied and therefore is now conducting similar research.

The QBO is a wind shift in the equatorial stratosphere, an oscillation from easterly to westerly and back on the time scale of approximately two years (26 months) and is a source of Internal Buoyancy/Gravity Waves (IBGW) which create absorptive perturbations in the D and E layers and even possibly the F 1/2 layer. A note, the E-valley/Flayer ducting propagation mechanism does not exist only during gray line periods. Internal Buoyancy/Gravity Waves (IBGW's) are a source of the ducting mechanism and allow for occurrences of ducting along any propagation path in total darkness. Measurement of the timing of arrival of propagated medium frequency RF signals demonstrates the existence of the ducting mechanism, versus conventional numerous E layer land/ocean surface hops.

The HAARP ionosphere program, earthquakes, volcanic eruptions, thunderstorms, lightning (especially positive cloud to ground strokes), elves, tornadoes and hurricanes and even

Earth's Ionosphere

The density and composition of the earth's atmosphere changes with height above the surface. From the sun, solar ultraviolet radiation impacts the upper atmosphere, ionizing the atoms there in the process of being absorbed. As the density of the atmosphere falls to very low levels the time it takes for ionized electrons and their atomic nuclei to recombine becomes sufficiently long to form a persistent ionized zone. This zone of ionized particles is called the earth's ionosphere. Due to the chemistry of the atmosphere, density variations, and varying energy levels in the ionizing ultraviolet radiation, the ionosphere forms distinct zones or layers which have been given the rather unimaginative names of D, E, F1 and F2 with increasing height above the earth.

With the complex interplay of the different factors involved, the ionosphere varies considerable with time. Regular variations form from the diurnal rotation of the earth bringing different parts of the atmosphere under the radiation effects of the sun, and from seasonal changes, and from the regular change in solar energy associated with the sunspot cycle. Other more random variations result from interactions of the earth's magnetic field with the charged particles in the ionosphere, and from disturbances to the magnetic field by solar activity. Its behavior also varies by latitude, with polar, mid-latitude, and low-latitude regions all behaving in characteristic ways.

visible and radio wavelengths. Visible-light observers generally monitor the sun in Hydrogen-alpha light where the flares shine brightest. Radio observation has traditionally utilized the very low frequency band. According to Mr. Hudak, (Sky and Telescope, 1984) this method requires a radio-quiet observing site, something simply not available to most amateurs today. This report describes observations using his method of shortwave fadeout to detect flares. In addition, the data gathered reveal several other interesting phenomena which I describe as well.

man made activities such as rocket launches including the space shuttle, are all sources of (IBGW's).

Many times I've heard ham's lament that propagation was going to go to crap due to another NASA Space Shuttle launch, in a sense they are correct.

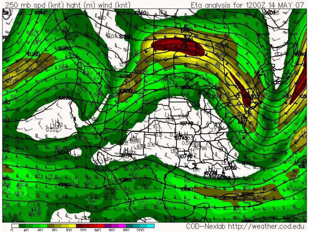
250 mb spd (knt) hght (m) wind (knt)
Eta analysis for 1200Z 14 MAY 07
COD–Nexlab http://weather.cod.edu

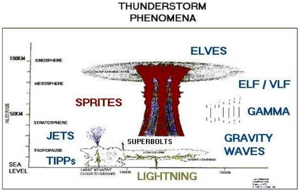
THUNDERSTORM
PHENOMENA
100KM
IONOSPHERE
MESOSPHERE
ALTITUDE
50KM
STRATOSPHERE
TROPOPAUSE
SEA
LEVEL
ELVES
ELF / VLF
SPRITES
GAMMA
JETS
SUPERBOLTS
GRAVITY
WAVES
TIPPs
LIGHTNING

Troposphere Height Jet Streams And Lightning/Thunderstorms And Their Interaction With The Ionosphere

-----

13.) Polar Cap Absorption (PCA)-

An anomalous condition of the polar Ionosphere whereby medium frequency (300-3000 kc) radio waves are absorbed, and LF and VLF (3-300 kHz) radio waves are wave guided at lower altitudes than normal. In practice, the absorption is inferred from the proton flux at energies greater than 10 MeV (10+0), so that PCA's, Polar Radio Blackouts and Proton Events are interrelated and often simultaneous.

((((NOTE!!! high latitude radio propagation paths may still be disturbed for days, up to weeks, following the end of an official proton event.)))) This fact is still stubbornly opposed by some otherwise very knowledgeable space weather physicists, hung up on threshold Riometer readings.

-----

14. Sunspot Group-

Sunspot groups are bipolar magnetic concentration regions on the photosphere of the Sun where magnetic field strengths many thousands of times stronger than the Earth's magnetic field reside. Sunspots appear as dark spots on the surface of the Sun because temperatures in the dark centers of sunspots drop to approximately 3700 K compared to 5700 K for the surrounding photosphere. The difference in temperature makes the spots appear darker than elsewhere. Sunspots typically last for several days to several weeks. They are seen to rotate around the sun, since they are on the surface, and the sun rotates fully every 27.5 days.

Sunspot groups have a magnetic classification as follows:

Solar Flares

August 7th, 1972

The study of solar flares is driven by both practical as well as scientific reasons. Flares are explosive eruptions of energy on the surface of the sun. These releases are the most powerful form of solar activity and occasionally have severe impact on terrestrial communications and power distribution. Tight magnetic fields associated with active regions on the sun form sunspots. These magnetic fields become twisted and kinked, storing tremendous amounts of energy. The process that causes the flare energy release is not well understood, and no predictive model exists which allows scientists to know in advance when or where a flare will erupt.

The energy released in a flare occurs over the full spectrum of wavelengths from the radio to visible, to ultraviolet and x-ray. By far the major release of energy is in the shorter wavelength ultraviolet and x-ray regions. Ultraviolet energy radiates away from the sun at the speed of light and, upon impacting the earth's ionosphere 8 minutes later, rapidly ionizes all levels disrupting communication. Charged particles released in the outburst take longer, arriving about 1 to 4 days after a major flare. They interact with the earth's magnetic field causing further communications disruptions and auroral displays.

Fortunately, flares last only a short time, up to several hours in their extended, or "main" phase after the initial burst. Given their terrestrial effects and short duration, flares have long been a welcome challenge to amateurs observing at both

Short Bibliography

Birney, D.S., 1969, Modern Astronomy, Allyn and Bacon, Inc. Boston, 338p.

Eddy, J.A., 1979, A New Sun: The Solar Results from Skylab; Edited by R. Ise., NASA Special Publication - 402, 198p.

Hudak, J., 1984, Shortwave Detection of Solar Flares; in Amateur Astronomers conducted by J. W. Briggs, Sky and Telescope, November 1984, pp 452-454.

Jacobs, G., and Cohen, T.J., 1979, The Shortwave Propagation Handbook, Principles, Theory, Prediction, CQ Technical Series, CQ Publishing, Hicksville NY., 150p.

Maloney, F.P., 1986, Computer CB Scanner; Computer Digest insert in Radio Electronics, v.57, n9, pp 7-9.

Olivier, W.L., and others, 1988, Ionosphere incoherent Scatter Measurements with the Middle and Upper Atmosphere Radar: Observations During the Large Magnetic Storm of February 6-8, 1986, Jour of Geophys. Res., v.93, n.a12, pp.14,694-14,655., Dec. 1988.

Tascione, T.F., 1988, Introduction to the Space Environment; Orbit Book Co., Malarbar Florida, 116p.

A - Alpha (a single polarity spot)
B - Beta (a bipolar spot configuration)
G - Gamma (an atypical mixture of polarities)
B-G - Beta-Gamma (a mixture of polarities in a dominantly bipolar configuration)
D - Delta (an opposite polarity umbrae within single penumbra)
B-D - Beta with a Delta configuration
B-G-D - Beta-Gamma with a Delta configuration

Sunspots usually come in groups with two opposing sets of spots. Whether two or twenty sunspots exist in a particular group they are counted as one sunspot group and numbered, such as 10500. That number would signify sunspot group number 10500, with the number counting system beginning in 1972 if my memory serves me correctly.

One set of sunspots will have a positive or north magnetic field while the other set will have a negative or south magnetic field. See image below.

The magnetic field is strongest in the darker parts of the sunspots called the umbra and weaker and more horizontal in the lighter part called the penumbra. The twisted magnetic fields associated with sunspot groups are one source of the solar flares, coronal mass ejections and geomagnetic storms that wreak havoc with the ionosphere here on Earth.

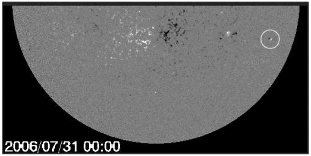

Magnetogram Image

The current system of counting sun spots hails from a previous era when direct observation of sun spots was inherently inaccurate. The sunspot number is derived by counting 10 points for each sunspot group and then adding one point for each spot. So if a sunspot group contains 1 individual sunspot the official count becomes 11. 4 individual sunspots in a sunspot group equals 44 sunspots. (See definition #24. Plage).

15.) Short Wave Fadeout (SWF)-

During a Sudden Ionosphere Disturbance (SID), which is tied to a Solar Flare and Coronal Mass Ejection, abrupt increased ionization of the D layer results in reduced to total absorption of medium frequency circuits which are refracted by the Ionosphere on the sun lit hemisphere of the earth. This is known as a Shortwave Fadeout (SWF).

Solar flares produce copious amounts of electromagnetic radiation including energetic protons which increase the ionization of the daytime D layer. Medium frequency communication depends on the refraction of signals from the

A indices (a, A, ap, and Ap)
Indices derived from the K index but converted to a linear scale as follows:

K	0	1	2	3	4	5	6	7	8	9
a	0	3	7	15	27	48	80	140	240	400

The Ap, or planetary A index is an average daily report which I have commonly used to indicate geomagnetic conditions. The equivalent values are defined thus:

Ap	Condition
0-7	Quiet
8-15	Unsettled
16-29	Active
30-49	Minor Geomagnetic Storm
50-100	Major Geomagnetic Storm
>100	Severe Geomagnetic Storm

Some Definitions

Sunspot Number
This is a measure of the number and area of both individual sunspots and sunspot groups. Naturally this is based on visual observations from a number of observatories, the data reduced and averaged. Sunspot numbers were introduced in 1848 by the Swiss astronomer J.R. Wolf and have been calculated, with caution, back to around 1610 making this the longest continuous record of solar activity available. Sunspot numbers range from lows near 0 to highs over 200 during maximums of solar activity.

Solar Flux
In many ways similar to the sunspot number, and well correlated, the solar flux index is a measure of solar radio flux at a frequency of 2800MHz, or 10.7 cm as it is commonly called. This measure was introduced in 1947 at Ottawa, Canada and has obvious advantages over the sunspot number in that it does not rely on visual, often subjective observations.

K indices (K, Kp)
Quasi-logarithmic index of geomagnetic activity relative to quiet levels for a local recording station. These measurements are taken over a 3-hour period and reported on a scale of 0 to 9. Planetary (Kp) values are determined from data from 12 to 13 stations worldwide. This indice was begun in 1949 at the Institut für Geophysic, Göttingen University, Germany.

higher E and F2 layers and these signals must travel through the D layer at least twice.

Lower frequencies are affected first and higher frequencies last. The stronger the event, the stronger the ionization of the D layer, the higher the frequency effected via absorption.

Daytime E layer propagation of the medium frequency AM broadcast band and 160 meters (See definition #1. Overview) usually only occurs during the winter season and especially at higher latitudes with a lower sun angle, also at the bottom of a sunspot cycle, therefore Swift's rarely are rarely noticed. The 80/75 meter and 40 meter bands are most noticeably affected, with the higher bands least affected.

An SWF can last from several hours on the lower frequencies to minutes on the higher frequencies. (See definition #16. Solar Flare).

16.) Solar Flare-

A day side Earth bound solar filament and/or approximate C5 class or higher solar flare can move the proton flux >10 MeV (10+0) and initiate large scale high latitude propagation path absorption but even smaller C4 class flares and weaker are the culprit behind hour-to-hour and night-to-night variations in signal strength on the MF AM broadcast band and 160 meters, both stateside and DX. This transfer of increased density and RF signal absorption from the day-side D layer to night-side of the ionosphere occurs through high level neutral winds.

X-Ray Class Solar Flare. The rank of a solar flare based on its X-ray energy output. Flares are classified by the order of magnitude of the peak burst intensity (I) measured at the earth in the 1 to 10 angstrom band as follows:
Class (in Watt/sq. Meter)
B- I less than (l.t.) 10.0E-06
C- 10.0E-06 l.e.= I l.t.= 10.0E-05
M- 10.0E-05 l.e.= I l.t.= 10.0E-04
X- I g.e.= 10.0E-04

Background radiation in the 1 to 10 Angstrom range (Hard X-Ray's), as well as Solar and Galactic Cosmic Rays and ionization of Nitric Oxide (NO) in our atmosphere is the source of ionization of the D layer.

Basically a C-class solar flare possesses energy 1/10 the level of an M- class solar flare and an M-class solar flare possesses energy 1/10 the level on an X-class solar flare. (See definition #15. Shortwave Fadeout).).

Solar flares are not random meaningless explosions but instead a process inter-related with Coronal Mass Ejections (CME's), by which the Sun expels complex magnetic signatures enroute to changing its magnetic polarity or said a different way the swapping of the Sun's magnetic poles. Basically the Sun swapped it magnetic polarity at the peak of present Solar Cycle 23 somewhere between July 2000 and December 2001. The next polarity swap will occur during Solar Cycle 24 somewhere around 2010-2011.

A true confession, while I can ponder for hours the possible physical reasons behind a discontinuity in the data, I personally get little pleasure from soldering two wires together. With modern Pentium PC computers its probably best just to purchase a manufactured A/D circuit and concentrate on the science. My own C64 retired soon after the end of this study in 1991.

described a scanner controller project for use with a Commodore 64. Two circuits were presented of which, one was an interface between the radio automatic gain control (AGC) and a C64 computer "paddle" port. I suspected that the unit would work as well with my Yaesu 7700 shortwave receiver and there followed correspondence with Dr. Maloney. He was a great help and in fact he had the same radio. The circuit had one major drawback from my perspective, it was battery operated and I was proposing to make regular observations of 15 plus hours duration. Finally the solution came in the form of a modification by my clever brother in Seattle allowing the interface to run off the 11vDC power supply or the radio itself.

With the hardware in place after my years of waiting, I was ready to go to work. September 1987 was only a year past the minimum between sunspot cycles 21 and 22 and the solar activity was still fairly low, but picking up rapidly. My goals included both detecting flares and a more general one of observing radio propagation variation from minimum to maximum in the sunspot cycle, a project therefore lasting several years. In actual fact I gathered data weekly from September 1987 until July of 1991. Each Friday early I would switch on the computer and radio and, with a homebrew Basic program, monitored WWV's signal at 15MHz.

A technical note on the interface. The device works by taking the variable voltage output of the receiver's AGC circuit, accessible easily from the rear of the unit, and converting it to a variable resistance. The Commodore 64 computer game port could read both joysticks and the even-then obsolete game "paddles" which worked on resistance. The computer read the resistance and converted it to digital values, a built-in analog to digital interface.

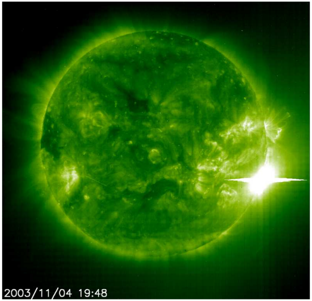

Image Of An X45 Class Solar Flare Click To Enlarge

Click for X45 Super Solar Flare Movie
http://www.wcflunatall.com/x28superflare031104.gif

17.) Sporadic-D (Ds) Absorption & Wave Guiding-

Sporadic-D (Ds) occurrences have an inter-relationship with brief but intense Sun based and Galactic X-rays and Galactic Cosmic Rays, huge positive cloud to ground lightning strokes and interrelated Elves and Sprites. Very large bursts of Gamma Rays have also been observed to occur in conjunction with Sprites.

Sporadic-D (Ds) absorption occurs both at day and night. Much of the night time occurrence of Sporadic-D (Ds) absorption is often masked by lightning QRN, as well as a lack of radio operation during thunderstorm events, due to the lightning strike hazard and also due to the operator not being able to recognize the mode due to unfamiliarity with it. It's doubtful that you will read about the Sporadic-D (Ds) phenomena anywhere else other then on this website.

While on the topic of lightning and propagation, an ionized lightning channel which normally has a maximum diameter of approximately a silver dollar, can reflect RF much like meteor trails do. I've personally noticed it on the 70 cm band, as a single propagation burst lasting 1/4 to 1/2 second. (See definition #20. D Layer Mid Winter Absorption Anomaly).

18.) Sporadic-E (Es) Absorption, Blocking & Refraction-

Just as the E layer is the main refraction medium for medium frequency (300-3000 kc) signal propagation within approximately 5000 km/3100 mi, so is a Sporadic-E (Es) cloud. Sporadic-E (Es) clouds occur at approximately 100 km/60 miles in altitude and generally move from ESE to WNW.

Like Stratosphere level warming and Troposphere level temperature and moisture discontinuities, Sporadic-E (Es) clouds can depending on the circumstances absorb, block or refract medium, high and very high frequency RF signals in an unpredictable manner.

The main source for "high latitude" Sporadic E (Es) clouds is geomagnetic storming induced radio aurora activity.

The main source for "mid latitude" Sporadic-E (Es) clouds is wind shear produced by internal buoyancy/gravity waves

Solar Flare detection by Shortwave Radio

Kevin Smith

http://www.lessmiths.com/~kjsmith/radio/r0ctnts.shtml

Introduction:
Fun with radio, a hobbyists' dash into the ionosphere..

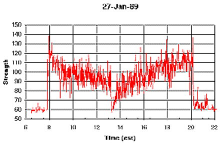

This is the tale of a project long in the making, nurtured over three years of thinking and patience, of persistent attention, and finally of research and discovery. I hope it will fascinate and inspire the reader as much as it did myself.

The object of this study was to observe solar flares via shortwave radio. The technique, described in an article by Mr. John Hudak, (Sky and Telescope, Nov. 1984), involves monitoring the shortwave signal of a distant beacon and looking for a characteristic fadeout signature indicative of a solar flare. The equipment utilized in the article included a shortwave receiver and a chart recorder. I was interested in this project and had the necessary receiver, but no chart recorder so things merely stewed in my mind. It would not be easy or practical to purchase a chart recorder for a single purpose. Instead I considered ways of utilizing a Commodore 64 computer for the purpose of recording the data.

Things stood that way for three years until one day while shopping with my family I picked up a Radio Electronics magazine quite on a lark. Reading it later at home I was quite stunned to find an article by Dr. Frank Maloney which

(IBGW's), that create traveling ionosphere disturbances (TID's), most of which are produced by severe thunderstorm cell complexes with overshooting tops that penetrate into the Stratosphere. Another tie in between Sporadic-E (Es) and a severe thunderstorm is the Elve.

The main sources for "low latitude" Sporadic-E (Es) clouds is wind shear produced by internal buoyancy/gravity waves (IBGW's), that create traveling ionosphere disturbances, most of which are produced by severe thunderstorm cell complexes tied to tropical cyclones. High electron content in the Equatorial Ring Current also plays a role.

The forecasting of Sporadic-E (Es) clouds has long been considered to be impossible. However it is possible to identify certain troposphere level meteorological conditions that can lead to the formation of Sporadic E (Es) clouds. One is as mentioned above the severe thunderstorm cell complex.

Sporadic-E (Es) clouds have been observed to initially occur within approximately 150 km/90 mi to the right of a severe thunderstorm cell complex in the northern hemisphere, with the opposite being observed in the southern hemisphere. To complicate matters is the fact that Sporadic-E (Es) clouds that initially form to the right of a severe thunderstorm complex in the northern hemisphere, then move from ESE-WNW and end up to the left of the severe thunderstorm complex in the northern hemisphere. So one has to look for Sporadic-E (Es) clouds on either side of a severe thunderstorm cell complex. Things get even more complicated when two severe thunderstorm cell complexes exist approximately 1000- 2000 miles apart.

Not all thunderstorm cell complexes reach severe levels and not all severe thunderstorm cell complexes produce Sporadic-E (Es). This is where knowledge in troposphere physics and weather analyses/forecasting is necessary. Coincidentally I have a B.S. in Meteorology and an M.S. in Space Plasma Physics and am qualified to identify which severe thunderstorm cell complexes are most likely to produce Sporadic-E (Es) clouds.

Some of the key elements in identifying which severe thunderstorm cell complexes have the potential to produce Sporadic-E (Es) via wind shear, from internal buoyancy/gravity waves, that produce traveling ionosphere disturbances include:

1.) Negative tilted mid and upper level long wave troughs.

2.) Approximate 150 knot/170 mph jet stream jet maxes that produce divergence and therefore create a sucking vacuum effect above thunderstorm cells, that assist thunderstorm cells in reaching and penetrating the Tropopause into the Stratosphere.

3.) 500 mb temperatures of -20 deg. C or colder, which produce numerous positive and negative lightning bolts and inter-related Sprites and Elves.

DISCUSSION

Leonard F. Fuller (by letter): Operators using continuous waves have frequently observed that altho the received night signals may be stronger on the shorter waves, fading is so much more frequent on these that they are often commercially inferior to longer waves even for night work.

This may be explained by the following:

If the mean of the observations on curve Figure 13 of the American Institute of Electrical Engineers' paper mentioned by Professor Marchant are plotted, a curve of received watts is obtained following approximately the curve of energy radiated from the transmitting antenna.

This is as it should be and is the reason for the first of the practical observations of operators. The second may be explained by the following consideration of the theory of Fgure 3 of Professor Marchant's paper.

As he points out, interference bands, in the sense in which they are usually understood in light phenomena, do not exist if we consider the path of transmission parallel to the earth surface as a whole or even locally at *A* and *B*. Inasmuch as the possible number of Heaviside cloud arrangements are infinite, there are possibilities of an infinite number of points similar to *A* and *B*. Thus regions of weak or strong signals may be entirely irregular as to dimensions and spacing.

These regions are produced by the superposition of two waves of the same frequency, *out of phase*, rather than by the combination of two waves of different frequency. The relative amount of signal fading or amplification at a given point is therefore dependent upon the angular phase displacement of the two interfering waves. This displacement is dependent upon the actual mechanical dimensions of the convolutions on the under side of the Heaviside layer. Assuming the contours of this layer constantly changing, it is obvious that the longer the wave length the less the phase displacement of the interfering waves for a given change in the Heaviside layer. Thus fading may be less troublesome on the longer waves and their commercial value enhanced thereby.

Fuller's results now confirm this explanation, tho it is necessary in addition, in the San Francisco-Honolulu tests, to assume refraction. It is to be hoped that experiments may be continued, since, besides giving a great deal of information on radio telegraphic technology, such tests may enable us to gain further knowledge of the nature of the upper atmosphere, at altitudes higher than those at which balloon observations are possible. The author wishes to express his indebtedness to Professor Wilberforce for his co-operation.

SUMMARY: Variations in received signal strength are ascribed partly to the existence of a "Heaviside cloud" layer consisting of masses of ionized gas at considerable heights. A partial bibliography of the subject is given.

The theory of interference caused by a spherical reflector is given, and results of receiving experiments due to Mr. Leonard Fuller, are studied in the light of the derived theory. It is shown that the Heaviside layer is probably quite irregular, and that refraction of the traveling waves is probably existent

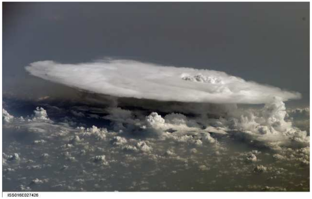

ISS016E027426

Overshooting Top Is The Puffy Blister Just
To The Right Of Center Of Flat Ice Anvil Top

4.) Approximate 150-175 knot/172-200 mph updrafts within thunderstorm cells complexes that create overshooting tops (see photograph above) that penetrate the Tropopause into the Stratosphere (See definition #20 on Stratospheric Warming), launching upwardly propagating internal buoyancy/gravity waves, which create traveling ionosphere disturbances and then wind shear.

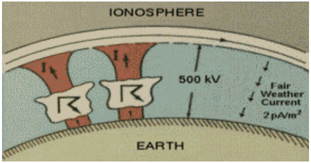

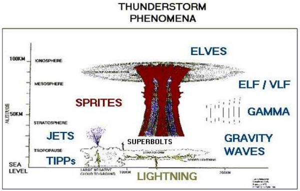

Lightning And Thunderstorms And Their Interaction With The Ionosphere

-----

19.) Long Delayed Echo (LDE)-

A fairly common propagation mechanism by which an RF transmitted signal returns to the sender within 1.25-5 seconds and in rare cases of up to 30 seconds. Research in the 1980's with HF OTHR discovered one propagation mechanism which involves ducting of the transmitted signal in the E-valley-F layer duct region of the ionosphere. A signal traveling along a magnetic field line much like a lightning induced whistler is another possibility.

The best time to observe an LDE is during the Fall/Spring equinox period when conditions are more balanced in the ionosphere. LDE's are very noticeable on amateur and SW broadcast signals between 17-28 mc with a peak near the maximum usable frequency (MUF). As recently as fall 2003 I did my own brief experiments using Morse code (CW) on the 15 meters band. I personally observed LDE's of my own

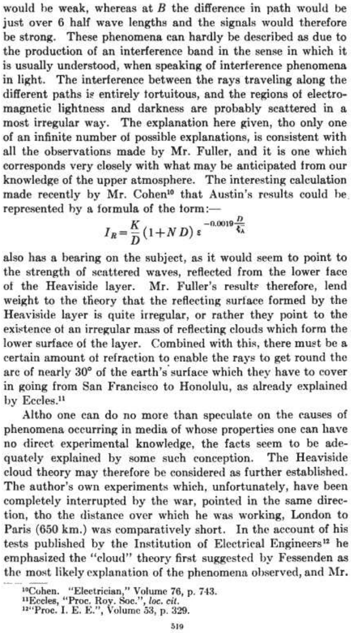

would be weak, whereas at B the difference in path would be just over 6 half wave lengths and the signals would therefore be strong. These phenomena can hardly be described as due to the production of an interference band in the sense in which it is usually understood, when speaking of interference phenomena in light. The interference between the rays traveling along the different paths is entirely tortuitous, and the regions of electromagnetic lightness and darkness are probably scattered in a most irregular way. The explanation here given, tho only one of an infinite number of possible explanations, is consistent with all the observations made by Mr. Fuller, and it is one which corresponds very closely with what may be anticipated from our knowledge of the upper atmosphere. The interesting calculation made recently by Mr. Cohen[10] that Austin's results could be represented by a formula of the form:—

$$I_R = \frac{K}{D}(1+N\,D)\,\varepsilon^{-0.0019\frac{D}{\sqrt{\lambda}}}$$

also has a bearing on the subject, as it would seem to point to the strength of scattered waves, reflected from the lower face of the Heaviside layer. Mr. Fuller's results therefore, lend weight to the theory that the reflecting surface formed by the Heaviside layer is quite irregular, or rather they point to the existence of an irregular mass of reflecting clouds which form the lower surface of the layer. Combined with this, there must be a certain amount of refraction to enable the rays to get round the arc of nearly 30° of the earth's surface which they have to cover in going from San Francisco to Honolulu, as already explained by Eccles.[11]

Altho one can do no more than speculate on the causes of phenomena occurring in media of whose properties one can have no direct experimental knowledge, the facts seem to be adequately explained by some such conception. The Heaviside cloud theory may therefore be considered as further established. The author's own experiments which, unfortunately, have been completely interrupted by the war, pointed in the same direction, tho the distance over which he was working, London to Paris (650 km.) was comparatively short. In the account of his tests published by the Institution of Electrical Engineers[12] he emphasized the "cloud" theory first suggested by Fessenden as the most likely explanation of the phenomena observed, and Mr.

---

[10] Cohen. "Electrician," Volume 76, p. 743.
[11] Eccles, "Proc. Roy. Soc.", *loc. cit.*
[12] "Proc. I. E. E.", Volume 53, p. 329.

in Figure 3, interference might take place between the waves reflected from two regions *C E* and *D F*. If one assumes a reflecting surface of such a nature[8] that the angle of incidence of a ray on it changes considerably for slightly differing altitudes as indicated in the figure, i. e., if the reflecting surface is irregular, a series of rays may be reflected at different angles by such a cloud at such points as *C E* (closely adjacent) in directions *C A* and *E B*, and it is easy to see that the difference in length of the path from *C* to *A* direct, and by the path *C D A* may be an odd number of half wave lengths, while the difference in length of path from *E* to *B* direct, and by the path *E F B* may be an even number of half wave lengths.[9] If this is the case, the distance between the two positions *A* and *B* at which the two waves would reinforce and neutralize each other respectively, may vary within wide limits. The diagram has been drawn to correspond, as nearly as possible, with the conditions given in Mr. Fuller's paper in which he has two stations 9 miles apart. The difference in length of the two paths to *A* is about 30 kilometers and to *B* about 26 kilometers. For a wave length of 7,500 meters, the signals at *A* would be strong, since the difference in path is 8 half wave lengths, at *B* the signals would be weak, since the difference in path is nearly 7 half wave lengths. For a wave length of about 8,500 meters the difference in path to *A* would be just under 7 half wave lengths so that signals

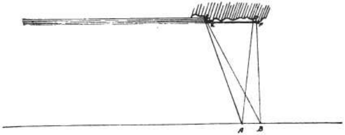

FIGURE 3

---

[8] It is perhaps misleading to speak of angle of incidence when the dimensions of the reflecting surface are of the same order of magnitude as the wave length, the incident radiation is really scattered.

[9] Altho the diagram has been drawn with the two interfering rays in one plane, it is clear that the rays may be reflected from other directions not coplanar.

transmitted signal of approximately 1.5-3 seconds and I could hear a mushy kind of Doppler shift on my returned signal frequency.

Claims of very long delayed echo's (VLDE) on the order of hours a4nd even days have been reported since the beginning of radio. Time periods of this magnitude would point to the "seeming possibility" of a refracting ionosphere type medium outside of Earth's own ionosphere, possibly somewhere past Pluto in the Oort Cloud. However no evidence so far has been found of such a medium and 99% of reported VLDE's are "probably" hoaxes.

http://heim.ifi.uio.no:80/~sverre/LDE
http://www.qslnet.de/member/la3za/prop

20.) Sudden Stratospheric Warming (STRATWARM ALERT)-

Sudden Stratosphere Warming is a major temperature change of the Winter time Polar and middle atmosphere from the Tropopause, where the Troposphere transitions into the Stratosphere to the base D layer of the ionosphere, which is at Mesosphere level. The warming lasts for many days at a time and is characterized by a warming of the Stratosphere temperature by some tens of degrees (temperature inversion), in unison with adjacent Troposphere cooling.

Another way to explain Stratosphere Warming is a major disturbance of the Winter time Polar middle atmosphere from the lower Stratosphere to the Mesosphere, resulting from a breakdown of the single Arctic Circumpolar Vortex into two circulation cells. Air trapped in the vortexes is mixed by the new meridional flow and is exposed to sunlight. Solar Lyman Alpha ionizes the Nitric Oxide (NO) gasses, resulting in an increase in electron density and producing strong medium frequency signal absorption at D layer height.

A little related Troposphere level Meteorology:

Interrelated with the splitting and shifting of the Arctic Circumpolar Vortex, is a Troposphere level negative North Atlantic Oscillation (NAO) and Pacific-North America Anomaly (PNA), mid and upper air height anomaly pattern. This equates to a large high pressure ridge in Western North America extending northward all the way into the Yukon region of Canada and a deep trough in the Eastern North America, from the eastern U.S. extending down into the Yucatan region of Mexico, with a second ridge in the western North Atlantic Ocean. This pattern is also called a dual blocking ridge and taps Siberian Arctic air, sending it across the North Pole into the eastern 2/3's of Canada and the U.S. providing for very cold surface temperatures.

As the Stratosphere lies below the Ionosphere, which is at Mesosphere and Thermosphere height, you would not expect to see Stratosphere Warming effect medium frequency propagation in any way BUT medium frequency signals do propagate off of Troposphere temperature inversions and moisture discontinuities and a temperature inversion is involved with Stratosphere Warming. So it's probable that a medium frequency signal could do any number of things when scattering off of a temperature inversion at any height. Unfortunately though some otherwise very knowledgeable Physicists stubbornly resist this concept.

Curve, Figure, Number	Date, 1914	San Francisco Time	Difference in length of path of interfering waves.	Condition of intervening space
5	Mar. 8	3 p.m.— 3.54 p.m.	28 km.	all light
6	" 15	10.30 a.m.— 12 p.m.	21 "	" "
7	" 15	3 p.m.— 4 p.m.	35 "	" "
8	Apr. 1	3 a.m.— 4 a.m.	28 "	all dark
10	May 17	11.30 a.m.— 12 a.m.	15 "	all light
11	" 24	10.25 a.m.—11.25 a.m.	36 "	" "
12	" 24	11.25 a.m.— 12 a.m.	14 "	" "
13	June 13	2.45 a.m.— 3.30 a.m.	20 "	all dark
14	" 22	10.30 a.m.—11.15 a.m.	average 15 km.	all light
15	" 28	10.45 a.m.—11.30 a.m.	36 "	" "

There is, of course, a difference between the day and night records in that the variations at night, due to altering wave length, are usually greater than they are in the day time, thus showing that the interfering waves are more nearly equal in intensity at night than they are by day. This is clearly shown in Figure 13 of Mr. Fuller's paper. The morning records give large variations also, and it would appear that altho the intervening space is fully lit, night conditions of semi-transparency still supervene. One of the most interesting results mentioned at the end of Mr. Fuller's paper is that he finds that the character of the land or water between the stations appears to make very little difference in long distance transmission, i. e., that the signals over land are as strong as they are over water. This seems almost conclusive proof that the observed signals are due to rays which are almost entirely refracted and reflected. The result, of course, is at variance with observations with shorter wave lengths over smaller distances. In dealing with interference phenomena it may be assumed therefore that the two rays travel round the earth thru the narrow passage formed by the earth and the Heaviside layer, being refracted as they pass, as was explained by Dr. Eccles, and being also reflected from the lower surface of the Heaviside layer.

If one assumes a ray refracted so as to follow nearly the curvature of the earth, which strikes a reflecting surface as shown

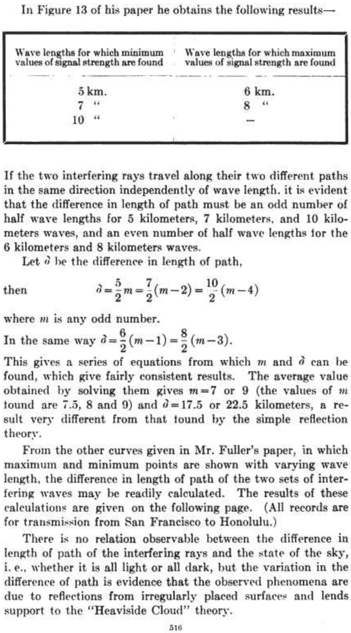

In Figure 13 of his paper he obtains the following results—

Wave lengths for which minimum values of signal strength are found	Wave lengths for which maximum values of signal strength are found
5 km.	6 km.
7 "	8 "
10 "	–

If the two interfering rays travel along their two different paths in the same direction independently of wave length, it is evident that the difference in length of path must be an odd number of half wave lengths for 5 kilometers, 7 kilometers, and 10 kilometers waves, and an even number of half wave lengths for the 6 kilometers and 8 kilometers waves.

Let δ be the difference in length of path,

then $$\delta = \frac{5}{2} m = \frac{7}{2}(m-2) = \frac{10}{2}(m-4)$$

where m is any odd number.

In the same way $\delta = \frac{6}{2}(m-1) = \frac{8}{2}(m-3)$.

This gives a series of equations from which m and δ can be found, which give fairly consistent results. The average value obtained by solving them gives $m=7$ or 9 (the values of m found are 7.5, 8 and 9) and $\delta = 17.5$ or 22.5 kilometers, a result very different from that found by the simple reflection theory.

From the other curves given in Mr. Fuller's paper, in which maximum and minimum points are shown with varying wave length, the difference in length of path of the two sets of interfering waves may be readily calculated. The results of these calculations are given on the following page. (All records are for transmission from San Francisco to Honolulu.)

There is no relation observable between the difference in length of path of the interfering rays and the state of the sky, i. e., whether it is all light or all dark, but the variation in the difference of path is evidence that the observed phenomena are due to reflections from irregularly placed surfaces and lends support to the "Heaviside Cloud" theory.

516

You can almost always correlate the coldest weather occurrences with poor medium frequency signal propagation conditions.

Also Stratospheric Warming (STRATWARM) has a negative effect on medium frequency propagation, due to increasing medium frequency radio wave absorption by the D layer, via upward propagating Internal Buoyancy/Gravity Waves (IBGW's).

This phenomenon also occurs in Southern Hemisphere Winter but is less pronounced.

U. Of Berlin Germany Stratospheric Research Group Layer
http://www.geo.fu-berlin.de/en/met/ag/strat/index.html

21.) D Layer Mid Winter Absorption Anomaly-

A period of increased medium frequency radio wave absorption at high and mid latitudes occurring in mid winter and is associated with sudden stratospheric warming and the Quasi Biennial Oscillation (QBO). If you look in your radio logs for 160 meters you will notice that most of your good DX contacts are in the fall and spring. This is due to the D Layer Mid Winter Absorption Anomaly. (See definition #17. Sporadic-D (Ds) Absorption & Wave Guiding). (See definition #20. Sudden Stratospheric Warming (STRATWARM ALERT).

Weather In The Upper Atmosphere
http://www.albany.edu/faculty/rgk/atm101/weather.htm

22.) F3 Ionosphere Layer-

A PDF Article Via IPS Australia About The Long Suspected But Only Recently Verified F3 Ionosphere Layer
http://www.wcflunatall.com/F3layer.pdf

The F3 layer primarily exists only in the vicinity of the Earth's magnetic equator. This may represent part of an explanation for (TEP) Trans Equatorial Propagation.

23.) The Gray line/Gray line Propagation-

A general east-west transition between daytime and nighttime (twilight) where enhanced propagation conditions "may" occur. Near local sunrise the absorptive D layer has yet to become illuminated by the Sun, though the higher in altitude F/F2 layer has. Inversely near local sunset the absorptive D layer is losing illumination by the Sun, though the higher in altitude F/F2 layer still is. There is also a strengthening and weakening process in the E layer, as well as angle tilts and altitude changes in the D, E and F layers.

This process can allow for enhanced propagation conditions within the general north-south gray line corridor. It is most pronounced on 30, 40 and 60 meters and less so on 80 and 160 meters. Actually most gray line propagation on 160 meters and to a lesser extent on 80 meters is perpendicular (right angles) to the corridor. In my professional observation the gray line propagation enhancement process is still not totally understood and it's benefit exaggerated to almost mythical proportion.

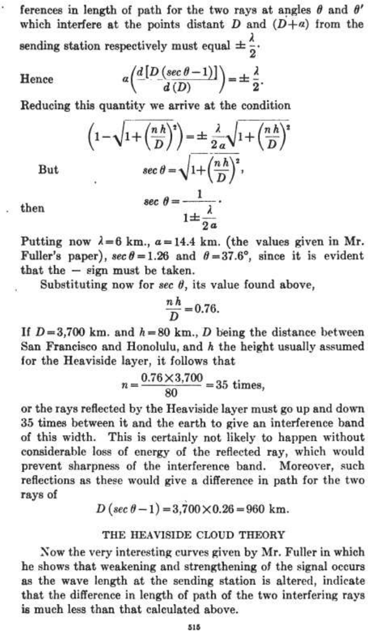

ferences in length of path for the two rays at angles θ and θ' which interfere at the points distant D and $(D+a)$ from the sending station respectively must equal $\pm\frac{\lambda}{2}$.

Hence $$a\left(\frac{d\left[D\,(sec\,\theta-1)\right]}{d\,(D)}\right)=\pm\frac{\lambda}{2}.$$

Reducing this quantity we arrive at the condition

$$\left(1-\sqrt{1+\left(\frac{n\,h}{D}\right)^2}\right)=\pm\frac{\lambda}{2\,a}\sqrt{1+\left(\frac{n\,h}{D}\right)^2}$$

But $$sec\,\theta=\sqrt{1+\left(\frac{n\,h}{D}\right)^2},$$

then $$sec\,\theta=\frac{1}{1\pm\frac{\lambda}{2\,a}}.$$

Putting now $\lambda=6$ km., $a=14.4$ km. (the values given in Mr. Fuller's paper), $sec\,\theta=1.26$ and $\theta=37.6°$, since it is evident that the $-$ sign must be taken.

Substituting now for $sec\,\theta$, its value found above,

$$\frac{n\,h}{D}=0.76.$$

If $D=3{,}700$ km. and $h=80$ km., D being the distance between San Francisco and Honolulu, and h the height usually assumed for the Heaviside layer, it follows that

$$n=\frac{0.76\times3{,}700}{80}=35 \text{ times},$$

or the rays reflected by the Heaviside layer must go up and down 35 times between it and the earth to give an interference band of this width. This is certainly not likely to happen without considerable loss of energy of the reflected ray, which would prevent sharpness of the interference band. Moreover, such reflections as these would give a difference in path for the two rays of

$$D\,(sec\,\theta-1)=3{,}700\times0.26=960 \text{ km.}$$

THE HEAVISIDE CLOUD THEORY

Now the very interesting curves given by Mr. Fuller in which he shows that weakening and strengthening of the signal occurs as the wave length at the sending station is altered, indicate that the difference in length of path of the two interfering rays is much less than that calculated above.

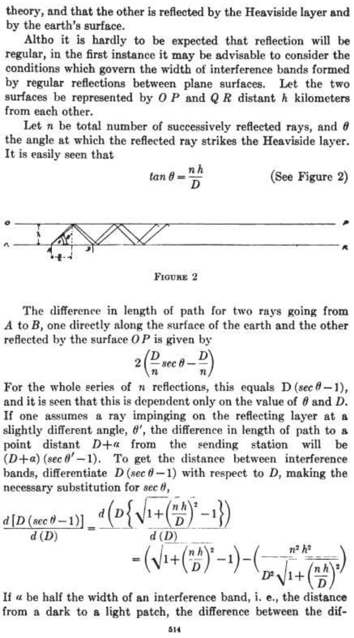

theory, and that the other is reflected by the Heaviside layer and by the earth's surface.

Altho it is hardly to be expected that reflection will be regular, in the first instance it may be advisable to consider the conditions which govern the width of interference bands formed by regular reflections between plane surfaces. Let the two surfaces be represented by $O\,P$ and $Q\,R$ distant h kilometers from each other.

Let n be total number of successively reflected rays, and θ the angle at which the reflected ray strikes the Heaviside layer. It is easily seen that

$$tan\,\theta = \frac{n\,h}{D} \qquad \text{(See Figure 2)}$$

FIGURE 2

The difference in length of path for two rays going from A to B, one directly along the surface of the earth and the other reflected by the surface $O\,P$ is given by

$$2\left(\frac{D}{n}\,sec\,\theta - \frac{D}{n}\right)$$

For the whole series of n reflections, this equals $D\,(sec\,\theta - 1)$, and it is seen that this is dependent only on the value of θ and D. If one assumes a ray impinging on the reflecting layer at a slightly different angle, θ', the difference in length of path to a point distant $D+a$ from the sending station will be $(D+a)\,(sec\,\theta' - 1)$. To get the distance between interference bands, differentiate $D\,(sec\,\theta - 1)$ with respect to D, making the necessary substitution for $sec\,\theta$,

$$\frac{d\,[D\,(sec\,\theta - 1)]}{d\,(D)} = \frac{d\left(D\left\{\sqrt{1+\left(\frac{n\,h}{D}\right)^2} - 1\right\}\right)}{d\,(D)}$$
$$= \left(\sqrt{1+\left(\frac{n\,h}{D}\right)^2} - 1\right) - \left(\frac{n^2\,h^2}{D^2\sqrt{1+\left(\frac{n\,h}{D}\right)^2}}\right)$$

If a be half the width of an interference band, i. e., the distance from a dark to a light patch, the difference between the dif-

514

(See definition #8.) E Valley/F Layer Propagation Ducting Mechanism/Chordal Hop Propagation).

24.) Plage-

A patchy H-alpha brightening on the solar disk commonly found in or near active regions of which can last for several days or so. A Plage is irregular in shape and variable in brightness and marks areas of nearly vertical emerging or reconnecting magnetic field lines. Often times a sunspot group will emerge from a Plage. (See definition #14. Sunspot Group).

Note! I make no claim of ownership of the images displayed on this website. Also I have attempted to keep the propagation theory explanations in simple to understand layman terms, because long complicated technical explanations can be boring and make one's eyes glaze over. Unfortunately though sometimes while trying to keep things simple, certain definitions, meanings and technical aspects can get watered down or even lost.

Therefore use these definitions at your own risk with no guarantee or warranty implied. The explanations contained herein is my intellectual property and copyrighted © 1988-2013 by Thomas F. Giella, W4HM, all rights reserved. Reproduction of information herein is allowed without advanced permission as long as proper credit is given.

Thomas F. Giella W4HM
Retired Meteorologist & Plasma Physicist
Lakeland, FL, USA at
w4hm at tampabay dot rr dot com

of the atmosphere which act as reflectors and refractors for the waves that are used for the transmission of radio signals.

Other facts bearing on the presence of this layer have been dealt with by Dr. Eccles in a paper published by the Royal Society.[7] It will not be necessary to reproduce the argument he uses to prove its existence, that may be assumed. It is the object of this paper to discuss to what extent the Heaviside layer can explain the phenomena described by Mr. Fuller.

THEORY OF INTERFERENCE BANDS ON A SPHERICAL REFLECTOR

In the first place, the conditions governing the production of interference bands on a spherical surface with a surrounding envelope, also of spherical form, whose internal surface acts as a reflector, may be discussed. It is usually assumed that the height at which the Heaviside layer becomes sufficiently conducting to act as a reflector is about 50 miles (80 km.). This is shown to scale in Figure 1, the points *A* and *B* corresponding respectively with San Francisco and Honolulu. It is clear that the passage between this layer and the earth is very much in the nature of a narrow crevasse between two parallel surfaces or reflecting mirrors.

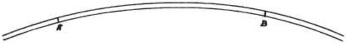

FIGURE 1

It is sufficient, as a first approximation, to consider, therefore the formation of interference bands by reflection from a pair of flat parallel surfaces. The distance between the bands of electromagnetic lightness and darkness is determined by the fact that the difference between the paths of the rays reaching these points by two alternative routes is half a wave length. If the difference in distance along the two paths is a multiple of a wave length the point is one of brightness; if it is an odd number of half wave lengths the point is dark.

It may be assumed in the first instance that one ray travels along the surface of the earth, as supposed in Sommerfeld's to that already available for the discussion of this subject, and it may be useful, therefore, to consider them in their bearing on the existence and probable nature of what, in this country, is generally called the "Heaviside layer."

---

[7] Eccles, "Proc. Roy. Soc.," A, Volume 87, pp. 79–99.

Tho it is usually called by this name, Professor Fleming[4] observed recently that it was Sir James Dewar, who was one of the first to draw attention to its existence. In a lecture to the Royal Institution in 1902[5], when discussing the constitution of the atmosphere Dewar pointed out that there were really two parts to it: the lower part in which atmospheric currents circulated, and in which the constituents were similar to those of the atmosphere at lower levels, and the upper part in which the distribution of gases was governed by their density. In two lectures delivered recently to the Royal Institution[6], Professor Fleming has discussed the formation of this upper ionized layer and the causes which produce it. He points out that, in order to produce ionization in such a gas as oxygen, by light radiation, it is necessary to have a wave length of the order of 1,500 to 1,800 ångstrom units (10^{-7} mm.), that is, light which is far beyond the ultra violet end of the spectrum. If such light really produces any ionization, then it is to be expected that the ionization would be reduced at night; and, therefore, that signals might be expected to vary in strength at night, if these ionized gases are the cause of signal variation. Professor Fleming suggests, however, that at heights of the order of 60 miles (100 km.), where the ordinary constituents of the atmosphere disappear and are replaced by hydrogen and helium and possibly other lighter gases, the most likely agency in producing ionization is the solar dust projected from the sun and transmitted to the earth thru the agency of light pressure. This explanation of the production of an upper ionized layer of gas is verified by the fact that the time interval elapsing between the passage of a sun spot across the solar meridian and the corresponding magnetic storm, as shown by Arrhenius, is about 45 hours, a figure which agrees fairly closely with the time Professor Fleming calculates that a particle of 1,200 ångstrom units diameter would take to pass from the sun to the earth. Whatever the cause which produces this layer, there is little doubt that such a layer exists, not necessarily in the form of a shell concentric with the earth, with fairly flat surfaces, but more likely in the form of large masses of gases in the upper regions

---

[4] "Electrician," Volume 75, p. 348.

[5] "Proc. Royal Institution," Volume 17, p. 223.

[6] *Loc. cit.*

CHEN-PANG YEANG[*]

The study of long-distance radio-wave propagation, 1900-1919

GUGLIELMO MARCONI'S TESTS of the trans-Atlantic wireless telegraphy made sensational news at the turn of the 20th century. Since the early 1890s, the Italian-born inventor-entrepreneur had established radio communication links spanning as far as the English Channel but what he did in 1901 was unprecedented. He attempted to exchange telegraphic signals between Britain and North America without submarine cables or other mediation. Marconi's moment came on December 12. Sitting in a station on a hill in Newfoundland, Canada, Marconi and his assistant George Kemp heard regular sharp clicks from earphones connected to the receiving apparatus. They had heard a wireless message transmitted from the high antenna tower energized by a spark-gap circuit located at Poldhu, England. The New York Times quickly featured Marconi's story.[1]

---

[*] Dibner Institute for the History of Science and Technology, and Program in Science, Technology, and Society, Massachusetts Institute of Technology, 38 Memorial Drive, Cambridge, MA 02139 (cpyeang@mit.edu).

I thank Jed Buchwald, Evelyn Hammond, Hermann Haus, David Kaiser, Rob Martello, David Mindell, George Smith for reading and commenting on various versions of this draft. I am also grateful for the help of the following people in accessing primary sources: Kun-Hau Ding, Brendan Foley, Wilhelm Füßl and the staff of Deutsches Museum Archives (Munich), the United States National Archives (Washington, D.C.), Institute of Electrical Engineers Archives (London), MIT Institute Archives (Cambridge, MA), Harvard University Archives (Cambridge, MA), and the library of the Air Force Research Laboratory (Hanscom, MA). A part of the content of this paper was presented at the Joint Atlantic Seminar in the History of Physical Sciences at the Chemical Heritage Foundation, Philadelphia, on September 28-29, 2001.

The following abbreviations are used: *AP*, *Annalen der Physik;* DM, Deutches Museum Archiv, Munich; (http://www.lrz-muenchen.de/~Sommerfeld/ for the archives Sommerfeld documents); *Obit*, Royal Society of London, *Obituary notices of fellows*; *PRS*, Royal Society of London, *Proceedings*.

[1] Orrien E. Dunlap, *Marconi, the man and his wireless* (New York, 1937), 87-102, and Degna Marconi, *My father Marconi* (New York, 1962), 111-120.

HSPS, Volume 33, Part 2, pages 369-404. ISSN 0890-9997. ©2003 by The Regents of the University of California. All rights reserved. Send requests for permission to reprint to Rights and Permissions, University of California Press, 2000 Center St., Ste. 303, Berkeley, CA 94704-1223.

The trans-Atlantic wireless test not only caught the attention of the general public, corporate capitalism, and the electrical industry for its technological implications, but also raised a curious scientific question. To establish a link across a distance as long as one sixth of the earth's perimeter, the wireless waves had to travel along a path conforming to the curved surface of the earth. Why do wireless waves, which behave much like optical and acoustic waves, not follow rectilinear trajectories? Why does the curvature of the earth not block them?

This paper shows how these problems were turned into mathematical representations, and how the empirical observations made by industrial practitioners evolved toward quantitative experimental data. It explores how engineering technologies affected experimental investigations, how traditions of applied mathematicians shaped numerical problems and approaches, and how theoretical tendencies of late-19th-century microphysics directed the formation of physical models. It also discusses how the problem of long-distance transmission emerged as a conjunction of distinct collective cultures, and how these cultures meshed with broader social contexts.

Two alternative theories of the peculiarly long transmission were proposed: the wave propagating along the surface results from diffraction by the body of the earth; and the wave bounces back and forth between the earth's surface and a conducting layer in the upper atmosphere. While theorists discussed these alternatives, experimenters worked on constructing empirical quantitative relations among the physical variables involved in long-distance transmission.

Three communities of researchers may be identified. European mathematical physicists and mathematicians worked on diffraction theories and developed mathematics to convert the diffracted fields into numerically tractable forms. Anglo-American electrical engineers and experimental physicists focused on atmospheric reflection and explored the conducting properties of the air. American wireless telegraphers established the empirical formula governing the relation between received wave intensity and distance for given wavelengths. It proved difficult to decide between the two theoretical models. The quantitative predictions of the diffraction theorists disagreed with the empirical formula. The reflection theorists could explain the bending of the waves and static noise qualitatively, but could not make quantitative predictions. The indeterminacy ended in 1919 when the English mathematician George Neville Watson developed a new mathematical technique for the diffraction theories and used it to demonstrate the superiority of the reflection approach.

The three different practices were significantly shaped by distinct intellectual traditions, purposes, and research styles. The long-distance radio experiments related to pragmatic instrumentality in wireless engineering: to test the validity of high-power transmitters and sensitive radio detectors for the first long-range radio station of the U.S. Navy. The mathematics of series or integral approximations used in the diffraction theories incorporated then recently developed complex-variable analysis in mathematics. The reflection model, which began as a speculation among electrical researchers, became part of the Maxwellian-microphysics agenda.

Proceedings of the Institute of Radio Engineers: 1916

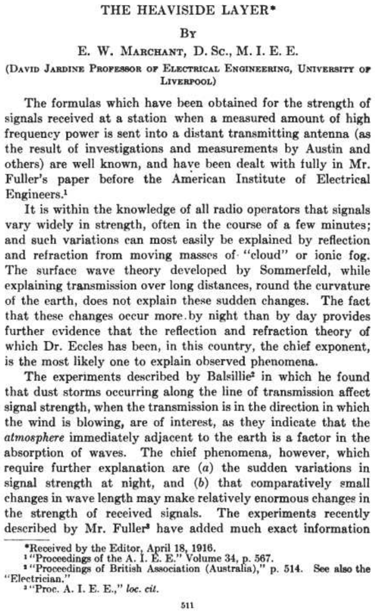

THE HEAVISIDE LAYER*

By

E. W. MARCHANT, D. Sc., M. I. E. E.

(DAVID JARDINE PROFESSOR OF ELECTRICAL ENGINEERING, UNIVERSITY OF LIVERPOOL)

The formulas which have been obtained for the strength of signals received at a station when a measured amount of high frequency power is sent into a distant transmitting antenna (as the result of investigations and measurements by Austin and others) are well known, and have been dealt with fully in Mr. Fuller's paper before the American Institute of Electrical Engineers.[1]

It is within the knowledge of all radio operators that signals vary widely in strength, often in the course of a few minutes; and such variations can most easily be explained by reflection and refraction from moving masses of "cloud" or ionic fog. The surface wave theory developed by Sommerfeld, while explaining transmission over long distances, round the curvature of the earth, does not explain these sudden changes. The fact that these changes occur more by night than by day provides further evidence that the reflection and refraction theory of which Dr. Eccles has been, in this country, the chief exponent, is the most likely one to explain observed phenomena.

The experiments described by Balsillie[2] in which he found that dust storms occurring along the line of transmission affect signal strength, when the transmission is in the direction in which the wind is blowing, are of interest, as they indicate that the *atmosphere* immediately adjacent to the earth is a factor in the absorption of waves. The chief phenomena, however, which require further explanation are (*a*) the sudden variations in signal strength at night, and (*b*) that comparatively small changes in wave length may make relatively enormous changes in the strength of received signals. The experiments recently described by Mr. Fuller[3] have added much exact information

---

*Received by the Editor, April 18, 1916.

[1] "Proceedings of the A. I. E. E." Volume 34, p. 567.

[2] "Proceedings of British Association (Australia)," p. 514. See also the "Electrician."

[3] "Proc. A. I. E. E.," *loc. cit.*

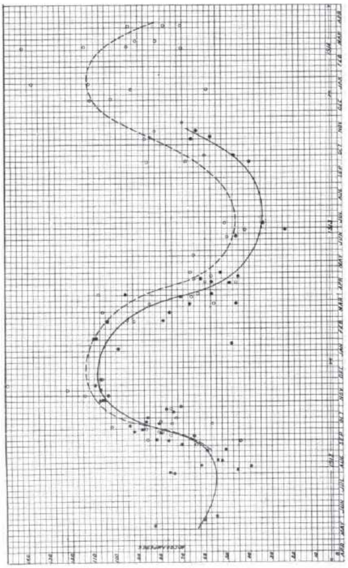

STRENGTH OF SIGNALS RECEIVED AT WASHINGTON
o = SIGNAL FROM PHILADELPHIA
• = SIGNAL FROM NORFOLK

These communities worked toward different goals, legitimized different methodologies, and gave priority to different technical and social concerns.

Diffraction theorists and the reflection modelers did not compete directly; their practices, and those of the experimenters, exhibited more complementary than competitive elements. Although the three communities were tied together by the same question (the possibility of long-distance radio-wave transmission), the most important aspect of their intertwined history was not the mutually exclusive and competitive answers they reached, but their contributions of different pieces of knowledge eventually subsumed in the solution.

The three communities' essential differences and the intellectual content of their interactions may be understood by analyzing their epistemic status—their judgments about what was known, what was important to know, and possible to know. Here reference to the teachings of Pierre Duhem and Sylvain Bromberg are in order. Duhem distinguished two aims of a physical theory: [to explain] a group of laws experimentally established, and "to *summarize* and *classify logically* a group of experimental laws without claiming to explain these laws."[2] Duhem's mathematical calculus for organizing scientific knowledge included quantitative and logical reasoning, numerical predictions, and recursive revision. Similarly, Bromberger distinguished theory as an intellectual device that could systematically generate answers to questions from theory as an answer to a why question.[3] Following Duhem's and Bromberger's theories of epistemic status, analyzing what is known, important to know, and possible to know amounts to deciding whether, for the historical actors, a piece of knowledge is explanatory or representational, an answer to a question or a part of a question-answering device.

The essential distinction between the diffraction theorists and the reflection theorists concerned the kinds of questions asked and the kinds of knowledge required, not the mutual exclusiveness of their physical models. The diffraction theorists developed rigorous mathematical theories to represent a physical model that aimed to account for only one wireless phenomenon—long-distance propagation. For this single phenomenon, they could mobilize quantitative reasoning and consolidate numerical predictions from mathematical theories. The reflection theorists constructed elaborate physical models to account for several significant wireless phenomena—not only long-distance transmission but also the diurnal and seasonal variations of signals and atmospheric noise. But their mathematical tools lagged much behind the diffraction theorists'. The diffraction theorists asked questions about mathematical tractability. The reflection theorists asked questions about the causes of various puzzling observations. Their theories were mainly explanations. But until a late stage they did not contain a systematic means for giving qualitative or quantitative predictions.

---

[2] Pierre Duhem, *The aim and structure of physical theory* (Princeton, 1982), 7.

[3] Sylvain Bromberger, "A theory about the theory of theory and about the theory of theories," in Bromberger, *On what we know we don't know: Explanation, theory, linguistics, and how questions shape them* (Chicago, 1992), 52-74.

The epistemic approach helps to understand the activities of the experimenters as well. The questions they asked were practical—they wanted to know the performances of transmitters and receivers under operating conditions in order that the U.S. Navy's first long-distance wireless would function properly. They were not motivated by theory-driven questions to conduct experiments. Likewise, their measured results served as reference for engineering specifications rather than as tests of theories. Because, however, the physics-trained experimenters believed that representing the measured results in terms of a mathematical formula, which was much more convenient than raw numbers for the theorists to work with, served engineering, they did provide material for checking theoretical predictions.

1. EMPIRICAL OBSERVATIONS

Guglielmo Marconi invented a transmitter and receiver for wireless telegraphy in the 1890s. Oliver Lodge in Britain, Alexander Popoff in Russia, Edouard Ducretet in France, and Ferdinand Braun in Germany made similar inventions about the same time.[4] The European and American governments, commercial and industrial enterprises, and inventors, engineers, and scientists quickly recognized the potential of these devices. Numerous sophisticated novel technologies of wireless telegraphy were developed between 1895 and 1920.

To the engineering communities, the primary concern with wireless telegraphy was the improvement of transmitters, receivers, and techniques of measurements.[5] They were not as successful in modeling the process of wave transmission as in engineering transmitters and receivers. They had too little information about wave propagation in the real world for which they intended the devices. To proceed, they needed to establish the empirical phenomena of wireless wave transmission. That task began only in the first few years of the 20th century.

The empirical problem may be stated as follows: under what environmental or instrumental conditions does the propagated electromagnetic energy experience a certain amount of change. The relevant major phenomena discovered before 1910 consisted of:

- Marconi's achievement of wireless communications over one sixth of the earth.
- The effect of the terrain. The maximum effective transmission distance of a wireless wave is the longer the greater the ground conductivity; thus a wave over sea usually propagates farther than one over land. The British naval officer Henry Bradwardine Jackson deduced the effect from experiments conducted on his own while serving on British warships between 1899 and 1902.[6]

[4] Hugh G.J. Aitken, *Syntony and spark: The origins of radio* (New York, 1976), 198.

[5] E.g., Jonathan Adolf Wilhelm Zenneck, *Wireless telegraphy* (New York, 1915); John Ambrose Fleming, *The principles of electric wave telegraphy and telephony* (London, 1916).

[6] Henry B. Jackson, "On the phenomena affecting the transmission of electric waves over the surface of the sea and the earth," *PRS, 70* (1902), 254-272.

SUMMARY: The strength of received signals from two stations was measured at the Bureau of Standards over a period of about two years. The transmitting wave length was 1,000 meters, spark frequency, 1,000, and sending antenna current about 10 amperes for each of the transmitters. Their distances were respectively 185 and 235 kilometers. The curves giving variation in intensity of received signals are shown and discussed.

DISCUSSION

Robert H. Marriott: It will be found interesting and instructive to compare what Dr. Austin has found with the results I described in my paper on "Radio Range Variation" before the Institute, (PROCEEDINGS OF THE INSTITUTE OF RADIO ENGINEERS, Volume 2, Number 1, page 37) and especially with chart 3, Figures 1 and 12 of that paper.

Alfred N. Goldsmith: It is evident from Dr. Austin's results that for the particular stations under consideration, the day of best transmission is close to January 1st, and the day of most difficult transmission to July 15th. The average ratio of received energy in winter to received energy in summer (for the extreme cases) is found to be 6.3. However, this last result is not very accurate, since the individual values of the ratio lie between 3.9 and 10.

was 69 ohms. The figure shows a well marked difference between the summer and winter intensities, but the great variation among the individual values makes it difficult to draw quantitative conclusions; observations on succeeding days in several instances differing from each other in a ratio of more than two to one, while the errors of observation are certainly less than 10 per cent. Rough curves have been drawn among the individual points of observation, indicating the general course of the changes. The Philadelphia values in general lie higher than the Norfolk values, with the exception of those taken in the Autumn of 1912 before certain changes were made in the Philadelphia antenna which appear to have increased its efficiency. No observations were taken in Norfolk after November, 1913, as changes in that station made it impossible properly to make comparison between the observations before and after that time. Notwithstanding the irregularities among the observations, a few facts appear fairly certain: The seasonal variations seem to be different in different years, the minimum of 1912 being higher than that of 1913. The rise in the curves in the Autumn of 1912 appears to be steeper than that of 1913, the practical maximum being attained by November 1st in 1912 and not until the middle of December in 1913. It has not been found possible definitely to connect the strength of signal with the changes in foliage conditions, altho it is possible that this is an important factor in the variations. Contrary to the ideas previously held, there seems to be no very marked connection between rainfall and the transmission of the signals. This was especially noticeable in the Autumn of 1912, when after a dry period, rain set in and fell heavily for four days. This, however, caused no certain increase in the strength of the received signals.

This preliminary series of observations shows that for a thoro study of the subject it will be necessary to observe at least twice a week, and preferably every day, for a long period of time. From these observations it will then be possible to derive average values from which the general course of the phenomena can be deduced with some degree of accuracy. It may then be possible by comparison with the curves of meteorological and magnetic phenomena to find relations which will help to explain the seasonal changes, and also the irregularities among the single observations.

Most of the observations have been taken by my assistant: H. J. Meneratti, Chief Electrician, U. S. N

- A wave goes further in dry than in humid air (Jackson).
- The maximum effective distance at night exceeds that in daytime (Marconi).[7]
- Noise, "static," or "stray wave" is more serious at night than in daytime (Jackson), during the summer than during the winter (Jackson), and at low than at high latitudes (Popoff, Feriyi, and Turpain).[8]
- The received energy generally increases with the length of the wave and the arrangement of the antenna.
- The radiation pattern of an antenna system can be different at different directions. A vertical aerial receives maximum energy when tilted along the transmission direction; several aerials with identical phase delay can give a strongly directional transmission (Marconi, Zenneck, Sigsfeld, Braun).[9]

Until 1905, the antenna consisted of a long vertical conducting wire connected at its lower end to a spark gap. The amplitude of received current is independent of the direction of the receiver's location with respect to the antenna, provided the receiver is on the ground. In 1906, Marconi suggested that a directional effect antenna could be obtained with a horizontal antenna.[10] The received current attains the maximum when the transmitter-receiver direction lies along the direction of the transmitter or the receiver antenna. Typically, reception is a minimum at angles of 110° or 250°. The minimum loci and the shape of the radiation pattern vary with antenna parameters. Figure 1 illustrates two of the radiation patterns Marconi measured.

Many of the empirical observations were qualitative descriptions that made either ontological assertions identifying physical factors relevant to the propagation of electromagnetic waves or quasi-quantitative comparative statements linking propagation with increase or decrease of a physical quantity. Quantitative descriptions of the empirical observations existed as tabulated data, not mathematical formulas. However far this phenomenological knowledge was from ideal scientific evidence, theorists had no choice but to use them as empirical evidence for their theories. Rigorous and controlled quantitative experiments on long-distance wave transmission remained to be done.

---

7. Dunlap (ref. 1), 122-127; Guglielmo Marconi, "A note on the effect of daylight upon the propagation of electromagnetic impulses over long distances," *PRS, 70* (1902), 344-347.
8. See Fleming (ref. 5), 851-852.
9. Guglielmo Marconi, "On methods whereby the radiation of electric waves may be mainly confined to certain directions, and whereby the receptivity of a receiver may be restricted to electric waves emanating from certain directions," *PRS, 77* (1906), 413-421; Ferdinand Braun, "On directed wireless telegraphy," *Electrician, 57* (1906), 222-224, 244-248. Cf. Friedrich Kurylo and Charles Susskind, *Ferdinand Braun: A life of the Nobel prizewinner and inventor of the cathode-ray oscilloscope* (Cambridge, 1981), 134, 143, 170.
10. Marconi (ref. 9).

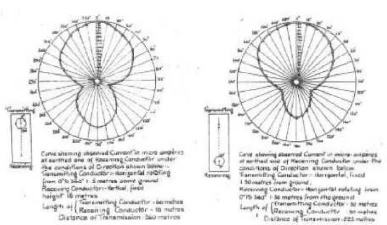
Transmitting
Receiving
Curve showing observed Current in micro amperes at earthed end of Receiving Conductor under the conditions of Direction shown below:-
Transmitting Conductor:- Horizontal rotating from 0° to 360° 1.5 metres above ground
Receiving Conductor:- Vertical, fixed height 18 metres
Length of Transmitting Conductor - 60 metres
Receiving Conductor - 18 metres
Distance of Transmission - 360 metres
Transmitting
Receiving
Curve showing observed Current in micro amperes at earthed end of Receiving Conductor under the conditions of Direction shown below
Transmitting Conductor - Horizontal, fixed 1.50 metres from ground.
Receiving Conductor - Horizontal rotating from 0° to 360° 1.50 metres from the ground
Length of Transmitting Conductor - 30 metres
Receiving Conductor - 30 metres
Distance of Transmission - 225 metres

FIG. 1 The radiation patterns of Marconi's directive antenna. Left panel, the transmitter's antenna rotates from 0° to 360°; right panel, the receiver's antenna rotates from 0° to 360°, Marconi (ref. 9), figs. 2, 4.

2. SURFACE DIFFRACTION THEORIES

The British initiative

To the community of physicists in the late 19th century, the phenomenon of long-distance wireless wave transmission was not altogether unfamiliar. The physical picture associated with this phenomenon could be found in the physical optics of diffraction and in acoustic scattering theory. Lord Rayleigh had developed a set of analytic techniques to deal with the acoustic problem.[11]

Between 1901 and 1919, two groups of mathematical physicists and mathematicians developed Rayleigh's approach for application to radio-wave transmission. One was a British group at Cambridge University plus the French mathematician Henri Poincaré. The other group was led by the mathematical physicist Arnold Sommerfeld at the University of Munich and the German electrical engineer Jonathan Zenneck.

Hector Munro Macdonald received a bachelor's degree at the University of Aberdeen before moving to Cambridge University, where he became fourth Wrangler in the Mathematical Tripos of 1889. His research agenda was set by the Adams Prize problem of 1901: to describe the propagation of electromagnetic waves under a number of boundary conditions with simple geometry. Published in 1902 as *Electric waves*, Macdonald's prize-winning essay applied an electromagnetic theory based on an energy expression to study the effect of an antenna on its electrical oscillating frequencies, and to solve the diffraction field at the edge of a perfectly

---

[11] John William Strutt, *The theory of sound* (New York, 1945).

SEASONAL VARIATION IN THE STRENGTH OF RADIOTELEGRAPHIC SIGNALS*

By

Louis W. Austin, Ph.D.

(*Director of the United States Naval Radiotelegraphic Laboratory*)

In 1912, experiments were begun at the Bureau of Standard on the measurement of the strength of the receiving antenna current produced by signals sent from the radio stations in the Philadelphia and Norfolk Navy Yards. The object of the experiments was the determination of the variation in the strength of the signals at different times of the year.

It had been known qualitatively that the winter signals in general were stronger than those of summer, especially when the transmission took place overland. The reason ordinarily given for this was the absorption of the waves during the summer, due to the vegetation.

The conditions of the experiments were as follows: The sending wave length was 1,000 meters, and the spark frequency was approximately 1,000 per second, the sending antenna current being kept not far from 10 amperes, and care being taken that waves of only one frequency were emitted. The height to the center of capacity of the Philadelphia antenna was 39 meters, and of the Norfolk antenna 52 meters. The antenna at the Bureau of Standards is a harp 55 meters high, having an effective height to the center of capacity of 30 meters. The capacity is 0.0014 microfarad. The distance from the Bureau of Standards to the Philadelphia station is 185 kilometers, and to the Norfolk station 235 kilometers. The method of measuring the received antenna current has been described in another place.†

The observations are shown in the accompanying figure. The ordinates represent microamperes of received current reduced to a constant sending antenna current of 10 amperes. The total receiving antenna resistance, including that of coupling, conductingprism.[12] When the news of Marconi's trans-Atlantic transmission spread in 1901, Macdonald was ready to work out a theoretical account of the phenomenon based on diffraction model.

---

* Delivered before The Institute of Radio Engineers, New York, December 2, 1914.

† Bulletin, Bureau of Standards 7, p. 295, 1910. Reprint No. 157.

Macdonald's paper of 1903 begins with a simple model that included all the supposedly relevant physical characteristics of wireless wave transmission.[13] The transmission antenna appears as a vertically polarized point current source, the so-called "Hertzian dipole," above the ground. Macdonald treated the atmosphere as free space with uniform dielectric constant and permeability and zero conductivity, and the earth as a perfect conductor (figure 2). The model captured some reality; but also erred seriously in giving the atmosphere no role (different from free space) in wave transmission. All diffraction theorists after Macdonald began their theoretical work by imposing the assumption.[14]

To calculate the intensity of the electric and magnetic fields at any point on the earth, Macdonald solved Maxwell's equations for the electromagnetic fields. He wrote down the wave equation for the azimuthal component of the magnetic field intensity in spherical coordinates (the axis of the azimuth is the radius connecting the earth's center and the dipole's location). Following Rayleigh's *Theory of sound*, Macdonald expressed the solution as a series of Bessel-Hankel functions and Legendre polynomials.[15] He determined the coefficients of the terms in this series from the conditions that the field intensity be infinite at the location of the Hertzian dipole and the tangential component of the electric field at the surface of the perfect conducting earth should vanish.[16]

The diffraction theorists in the first twenty years of the 20th century agreed with all Macdonald's arguments up to this point. They disagreed about how to approximate the analytical solution in a numerically tractable form. Macdonald noticed that the problem had exactly the same form as one dealt with in Rayleigh's *Theory of sound*. Following Rayleigh, Macdonald exploited the asymptotic properties of Hankel functions to establish that when the wavelength λ is much smaller than the radius of the earth a, the field intensity obeys a simple relation. The ratio of the electric field at the sphere's surface at the separation angle θ (the angle between the oscillator P and the point of observation C as seen from the earth's center) to the electric field at $\theta=0°$ is $1-\cos\chi$, χ being the angle between the observer and the earth's center as seen from the dipole.[17]

This overly succinct result implies that the electric field produced by a Hertzian dipole does not vanish at any point on the earth, including the diametrically opposite point, $\chi = 0°$. It means that the earth never casts any shadow on the propa-

---

[12] *Obit., 11* (1935), 553-555.

[13] Hector Munro Macdonald, "The bending of electric waves round a conducting obstacle," *PRS, 71* (1903), 251-258.

[14] Some of them later assigned the earth a finite conductivity and dielectric constant, but others continued to work on the perfect-conductor case.

[15] Strutt (ref. 11), chapt. 17.

[16] Ibid, 253.

[17] Ibid, 255.

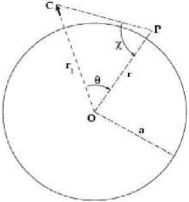

FIG. 2 Spherical boundary condition of the diffraction theory. The sphere represents the earth, the small arrow at C the transmission antenna modeled as a vertically polarized Hertzian dipole; the observer is at P.

gation of the electromagnetic field. The field can travel anywhere on the earth. Thus the magic of trans-Atlantic wireless transmission: the field diffracts across the surface of the earth. Macdonald considered this conclusion to be a complement to Rayleigh's discovery in acoustics.[18]

Unfortunately, Macdonald's approximation scheme had a serious problem, which Rayleigh pointed out. Rayleigh claimed that a shadowless wireless wave could not move around the earth since it would lack analogy to the optical case. The wavelength of the wireless wave (less than 50 kilometers) had about the same ratio to the radius of the earth as the wavelength of visible light had to one inch; but the light shining on a conducting ball one inch in radius does not creep around the ball to illuminate its rear surface. Furthermore, Rayleigh points out that Macdonald's asymptotic approximation did not in fact hold when the wavelength is much smaller than the radius of the sphere owing to a property of the Bessel functions that need not be detailed here.[19] Henri Poincaré also pointed out the difficulty with Macdonald's asymptotic approximation.[20]

In the following decade, the mathematical problem of how to approximate the exact solution of the diffracted field given by Macdonald was widely discussed among mathematicians and mathematical physicists, who proposed new approaches to the notorious Bessel functions. None of these efforts survived criticism. Discus-

---

18. Ibid, 328.
19. John William Strutt, "On the bending of waves around a spherical obstacle," *PRS, 72* (1904), 40-41.
20. Henri Poincaré, "Sur la diffraction des ondes electriques: Àpropos d'un article de M. Macdonald," *PRS, 72* (1904), 42-52.

H. E. HALLBORG: We should try more transmission experiments between stations lying north and south, so as to be able to compare ranges with those lying east and west. It has already been noted by ship stations that the transmission of signals in a north and south line is superior to that in an east and west line.

GUY HILL: Referring to the surface waves, Marconi has used wires on the ground for long distance reception, and some time ago Fessenden received messages over 600 miles on ground antennae.

A. E. KENNELLY: We may assume, in the present state of the theory of radio-telegraphic received signals, that the voltage of a signal received from a given steady wave-train is directly proportional to the maximum height of the receiving antenna above the plane of the equivalent perfect ground surface. The electric energy of the signal, however, is of course proportional to the received voltage and the received quantity. The quantity probably depends on the extent of wave surface area on the wave-front intercepted by the antenna. If this way of considering the matter is correct, a very low antenna of great length might give as strong a receiving signal as a high antenna of small width or surface area.

ALFRED N. GOLDSMITH: There seems to be a non-reciprocity of sending and receiving properties, on which Lord Rayleigh has already commented. A high antenna is necessary for transmission, but since reception of messages is largely accomplished on surface waves (at least for long distances) a low receiving antenna suffices.

CHARLES A. LE QUESNE, JR.: In a recent number of the Telephone and Telegraph Age I find a reference to some experiments by Austin Curtis on an effect of moonlight on reception of signals. The effects presented were similar to those for sunlight.

ALFRED N. GOLDSMITH: These experiments were originally disclosed in The London Electrician for March 21, 1913, (page 1104) and May 2, 1913, (page 143), and they have been considered by Dr. Eccles in the same periodical for March 28, 1913, (page 1144).

antennae, λ the wave length, and s, the distance between the stations.

$$I_2 = \frac{4.25\ I_1\ h_1\ h_2}{s}\ \varepsilon^{-\frac{\alpha s}{\sqrt{\lambda}}}$$

where α is a constant, equal approximately to 0.0015.

These experiments covered several types of antennae, and distances up to 1,000 miles (1,600 km.) with various antenna currents. The current for audibility (thru an equivalent resistance of 25 ohms) was taken as 10 microamperes. It appears that we need only 5 microamperes, or with the heterodyne receiver less than 2.5 microamperes. Making this change in the equation above, the results had in the recent Arlington-Salem tests offer good substantiation of the relation. With a sending antenna effectively 450 feet (138 meters) high, and a receiving antenna of 130 feet (40 meters), groups of thirty-word messages have been consistently received by daylight up to a distance of 2,383 miles (3,830 km.). From this it can be shown by a graphical solution of Austin-Cohen equation that the daylight range of an Arlington-type station to a similar station is 2,920 miles (4,700 km.) at a wave length of 4,000 meters; and rises to 3,400 miles (5,500 km.) at a wave length of 10,000 meters. (Further details of these heterodyne experiments, and the Arlington-Salem tests are given in the latter portion of the next paper in this issue of the Proceedings. Editor.)

I wish to endorse Professor Kennelly's suggestions. I trust that the observations will cover all parts of the twenty-four hour day, and that transmitting station records, as recommended by Mr. Weagant, may be secured.

LLOYD ESPENSCHIED: Such experiments as just proposed can be best carried on by the Navy. They would materially add to the efficiency of the Weather Bureau's work, if properly planned.

CAPT. F. J. BEHR (Coast Artillery Corps, U. S. A.): This subject is one of great interest to us; particularly the relative advantages of "damped" and "undamped" waves. These meetings of THE INSTITUTE OF RADIO ENGINEERS are doing yeoman service in calling attention to these points and the proper method of investigating them.

sion of the problem was dominated by Cambridge-trained people—and by Henri Poincaré.

Poincaré tried to work out an adequate and rigorous approximate solution for the mathematical problem formulated by Macdonald. He worked at it from 1909 until his death in 1912, publishing nine papers in all, including a 100-page monograph published in 1910.[21] Poincaré's strategy was to convert the infinite series obtained by analytically solving the wave equation into a definite integral and to employ Cauchy's residue theorem to evaluate it. Like Macdonald, Poincaré expressed the analytical solution of the wave equation in terms of a series of spherical harmonics constituted of Bessel functions and Legendre polynomials. In the case of a dipole on the ground, the infinite series can be converted into a sum of integrals associated with the poles of the spherical harmonics. When the size of the conductor is much larger than the wavelength, the term corresponding to the pole with the smallest imaginary part dominates the other terms. Poincaré proved that the dominant-pole contribution to the field intensity on the spherical surface is proportional to $\exp[-\rho(ka)^{1/3}\theta]$, where θ is the angle defined in figure 2, $k=2\pi/\lambda$ is the earth's radius and ρ is an unspecified constant. The field intensity produced by a Hertzian dipole on a spherical conductor has the form of exponential decay with respect to the separation angle.[22]

Poincaré's work involved esoteric theories of Bessel functions and complicated mathematical manipulations. But its conclusion had a clear and straightforward physical meaning: the diffraction field on a conducting sphere decays exponentially: the larger the wavelength, the longer the transmission distance. Four years after Poincaré published his important formula, a debate broke out over whether the field's decay rate varied with $\lambda^{1/3}$, as required by his theory, or with $\lambda^{1/2}$, as revealed from experiment. But Poincaré's result also was incomplete. It did not specify the numerical values of the decay rate γ and the amplitude constant. Without these values, the mathematical theory of diffraction could not produce quantitative results strictly comparable to the experimental data. The task of obtaining the missing numerical value of the decay rate fell to John William Nicholson.

Nicholson, Cambridge Wrangler began to publish on the diffraction problem in 1910. He opened with a criticism of Poincaré's method.[23] Poincaré had not carried out the asymptotic approximation of the Bessel functions correctly, and his procedure for converting the infinite series to an integral was not adequately rigorous. Nonetheless, like Poincaré Nicholson converted the series into an integral and obtained its approximate value by the contribution from the dominant pole. He examined the pole structures of the functions and concluded that for *ka*>>1, the magnetic field intensity is dominated by the contribution from the zero of the de-

---

21. Henri Poincaré, "Sur la diffraction des ondes hertziennes," Circolo matematico di palermo, *Rendiconti, 29* (1910), 169-259.

22. Ibid, 201.

23. John William Nicholson, "On the bending of electric waves round the earth," *Philosophical magazine, 19* (1910), 276.

rivative of the Hankel function with the smallest imaginary part. The imaginary part of this zero has the form of $\rho(ka)^{1/3}$. Thus the approximate field is proportional to $\exp[-\rho(ka)^{1/3}\theta]$.

This conclusion is the same as Poincaré's. In addition, Nicholson solved for the numerical value of the attenuation coefficient ρ and obtained the value 0.696. Having this numerical information, Nicholson constructed tables providing quantitative predictions of the diffracted electromagnetic waves around the earth.

No quantitative experimental data on wireless wave transmission were available when Nicholson finished his calculation. Still he could make a definitive and surprising statement: "diffraction must be a relatively insignificant agency in the success of experiments such as those of Marconi."[24] Nicholson's confidence came from a dramatic discrepancy between the numerical scale of the diffraction theory and that of Marconi's experiment. The exponential decay at $0.696(ka)^{1/3}$ made the diffracted field diminish much faster than it should have done if diffraction was responsible for long-distance wave transmission. Although Nicholson continued to work on the diffraction theory for a few years, he no longer believed that it could explain long-distance transmission. He thought that reflections from the upper atmosphere offered a more plausible theory.

Nicholson's conclusion did not destroy the diffraction theory. His mathematical solution to Macdonald's predicament was not the final one. Other mathematicians developed different methods that generated numerical results different from his. Also, while the British-French diffraction theorists all stuck to the assumption that the earth can be modeled as a perfect conductor, which guides the diffracted field around the earth, the German diffraction theorists did not focus only on this case.

German follow-up

In Germany, a group of diffraction theorists worked on the problem of wireless-wave transmission over ground. They believed that finite ground conductivity played a role in several wireless phenomena. They assumed that a "surface wave" would rise above the ground and creep along the ground surface. The finite ground conductivity could modify the polarization of the wave. The German diffraction theorists during the 1900s and the 1910s included an electrical engineer, Jonathan Zenneck, at Braunschweig, the theoretical physicist Arnold Sommerfeld, and his protégés at the University of Munich.

In 1889, while pursuing his doctoral degree at the University of Tübingen, Zenneck became assistant of Ferdinand Braun, a pioneer of radio research in Germany. During their work together (1892 to 1906) Braun's research group competed fiercely with Marconi's team. In 1899, Braun launched experiments on long-distance wireless telegraphy in the Cuxhaven region near Hamburg. Zenneck was

---

[24] John Nicholson, "On the bending of electric waves round a large sphere: III," *Philosophical magazine, 21* (1911), 67-68.

---

their use for a stated amount of energy. But it is as yet an obscure point.

JOHN L. HOGAN, JR.: The point that Mr. Hill suggests regarding the prominence of reflection (and interference) effects with sustained waves is well substantiated by some observations presented by Dr. de Forest before this INSTITUTE. (Proceeding of the Institute of Radio Engineers, Vol. 1, Number 1, page 37.) He shows that if continuous waves pass from a transmitting station to a receiving station by two different routes, one of which is direct and the other of which is caused by reflection of waves which strike elevated cloud layers, very marked interference effects will be produced at certain frequencies. That is, at definite wave lengths the signals will be either markedly weakened or strengthened. Change of wave length may bring the signals to normal strength.

It is possible that these selective absorption effects are phenomena based on slow resonance, and that therefore they should be more marked with sustained than with damped waves. Instances of marked reflective amplification may be responsible for the transmission of long-distance signals which have brought forward the contention that the range to be attained by the use of sustained waves is greater for equal output than with damped waves. I have discussed this claim with Dr. Austin while in Washington, and he appeared very sceptical concerning it. We felt that we had not enough data to warrant acceptance of it and that there should be no great difference in the transmission of sustained waves as compared with those of low decrement.

It seems to be a wonderful confirmation of Professor Kennelly's hypothesis that even such observations as those taken between Clifden and Glace Bay (where signals were graded in strength by simple aural classification as "very strong," "strong," "moderate," "weak," and "very weak"), are in such good accord with the theoretical conclusions.

In connection with abnormal daylight absorption treated by Dr. Kennelly, it is interesting to note that the value and relation for normal daylight absorption have been partially confirmed.

In 1911, Drs. Austin and Cohen, on the basis of experimental work performed from Brant Rock, gave the following law for the received antenna current I_2 in terms of I_1, the transmitting antenna current, h_1 and h_2, the heights of transmitting and receiving

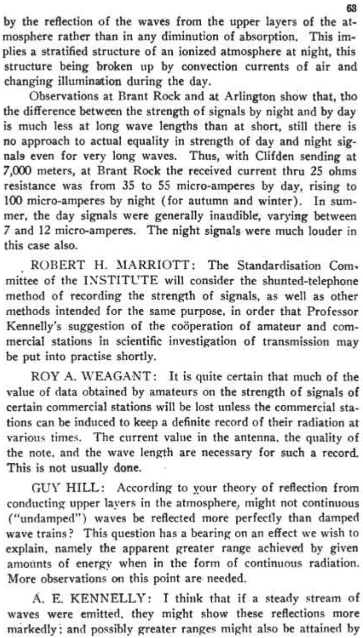

by the reflection of the waves from the upper layers of the atmosphere rather than in any diminution of absorption. This implies a stratified structure of an ionized atmosphere at night, this structure being broken up by convection currents of air and changing illumination during the day.

Observations at Brant Rock and at Arlington show that, tho the difference between the strength of signals by night and by day is much less at long wave lengths than at short, still there is no approach to actual equality in strength of day and night signals even for very long waves. Thus, with Clifden sending at 7,000 meters, at Brant Rock the received current thru 25 ohms resistance was from 35 to 55 micro-amperes by day, rising to 100 micro-amperes by night (for autumn and winter). In summer, the day signals were generally inaudible, varying between 7 and 12 micro-amperes. The night signals were much louder in this case also.

ROBERT H. MARRIOTT: The Standardisation Committee of the INSTITUTE will consider the shunted-telephone method of recording the strength of signals, as well as other methods intended for the same purpose, in order that Professor Kennelly's suggestion of the coöperation of amateur and commercial stations in scientific investigation of transmission may be put into practise shortly.

ROY A. WEAGANT: It is quite certain that much of the value of data obtained by amateurs on the strength of signals of certain commercial stations will be lost unless the commercial stations can be induced to keep a definite record of their radiation at various times. The current value in the antenna, the quality of the note, and the wave length are necessary for such a record. This is not usually done.

GUY HILL: According to your theory of reflection from conducting upper layers in the atmosphere, might not continuous ("undamped") waves be reflected more perfectly than damped wave trains? This question has a bearing on an effect we wish to explain, namely the apparent greater range achieved by given amounts of energy when in the form of continuous radiation. More observations on this point are needed.

A. E. KENNELLY: I think that if a steady stream of waves were emitted, they might show these reflections more markedly; and possibly greater ranges might also be attained by

deeply involved in the Cuxhaven project. By 1905, Zenneck had become an authoritative figure in the community of telegraphy engineer. In 1906, he became a professor at the Braunschweig Technical University.[25]

Zenneck's first work of interest here concerned wave propagation along an infinite planar interface between the air and a conductor (figure 3).[26] The novelty of Zenneck's approach in comparison with the British theorists did not lie only in the geometry. In contrast to previous work, he did not use any information about the dipole oscillator to solve the overall field generated by the source and shaped by the boundary condition. He supposed a particular form for the electric and magnetic fields and confirmed that it satisfied Maxwell's equations and the boundary condition. This form was a plane wave with field components containing a factor $\exp[i\,(\omega t+sx)]$, where ω is the angular frequency and s is the wave number in the horizontal direction. Plugging these expressions into Maxwell's equations and the boundary condition, he found that the field quantities above the surface are proportional to $\exp[i(\nu t+sx-r_0z)]$, and those below the surface to $\exp[i(\nu t+sx+r_1z)]$, where $r_0= (k_0{}^2-s^2)^{1/2}$ and $r_1= (k_1{}^2-s^2)^{1/2}$, and k_0 ands k_1 are the wave numbers in the air and on the ground, respectively. The values of s, r_1, and r_0, all obtained by solving simple algebraic equations, are determined by the dielectric constants and conductivities of the air and the ground. They were complex numbers.

The physical implications of Zenneck's field solution are extraordinary. First, unlike the optical plane waves in free space, Zenneck's wave not only propagates but also attenuates along both the *x* and *z* directions. In addition, the polarization of Zenneck's wave is determined by the ground conductivity and dielectric constant. In contrast to free-space plane waves, the polarization of Zenneck's waves cannot be chosen freely.

When the ground conductivity is finite (in the scale of earth, stone, or sand conductivity), the polarization direction of Zenneck's wave inclines along the direction from which the wave comes. Zenneck pointed out that Marconi's experimental results on the directive antenna agreed with this finding. Marconi discovered that the receiving antenna receives maximum power when the aerial inclines along the line of sight between the transmitter and the receiver. According to Zenneck's theory, the finite ground conductivity causes the direction of polarization of the propagating wireless wave to incline toward the direction of propagation; the antenna has maximum efficiency to convert the field into an oscillating current when it aligns with the polarization direction of the field (figure 4).

The attenuation of Zenneck's wave along the propagation direction reaches its maximum at a finite ground conductivity. When the conductivity is either zero or infinite, the attenuation is zero. A ground with high resistance (a dielectric material) could support long-distance propagation. Zenneck also showed that the at-

---

[25] Kurylo and Susskind (ref. 9), 73-74, 130-173.

[26] Jonathan A.W. Zenneck, "Über die Fortpflanzung ebener elektromagnetischer Wellen längs einer ebenen Leiterfläche und ihre Beziehung zur drahtlosen Telegraphie," *AP, 23* (1907), 846-866.

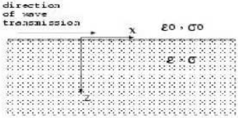

FIG. 3 Zenneck's boundary condition.

tenuation decreases with wavelength and is much more serious along the vertical than along the propagation direction. The field intensity decreases to $1/e$ of the original intensity at $z=0$ in some centimeters along the vertical and in some kilometers along the horizontal. Most of the energy is concentrated near the air-ground boundary; Zenneck's wave is a "surface wave."

Zenneck's work provided a novel insight. Unlike the British and French diffraction theorists, who held that the shape of the earth enabled long-distance wave propagation, Zenneck suggested that the ground resistance also played a critical role. Nonetheless, his paper of 1907 could not be considered as a complete work. "Zenneck's wave" is only one possible solution of the Maxwell's equations and the boundary conditions, not necessarily the solution describing physical reality. In contrast to the British diffraction theory, Zenneck's approach did not specify the source that generated his wave; we not only have no idea how to generate it, we do not know whether it can be generated at all! Enter Arnold Sommerfeld.

Sommerfeld had worked on optical diffraction before he took up radio waves. His novel approach of converting the solution of a differential equation with a proper boundary condition into a closed-form complex integral suitable for numerical evaluation would become a hallmark of his work.[27] He published his first paper on the Hertzian waves in 1899.[28] It studied the propagation of electromagnetic waves along a conducting wire. Sommerfeld demonstrated that as current flows in the wire, the Hertzian wave it produces also propagates along the wire. Once electrical phenomena were conceived in terms of waves in the aether, there was no essential difference between the wired and the wireless. Since the energy transfer associated with the flow of an electric current in a wire could be understood as the propagation of an aetherial wave along the wire, it was reasonable to understand a wave propagating above ground in terms of a flow of energy guided by the ground.

---

[27] Paul Forman, "Arnold Sommerfeld," in *Dictionary of scientific biography, 12* (New York, 1975), 526-529; Arnold Sommerfeld, "Autobiographische Skizze," in Sommerfeld, *Gesammelte Schriften, 4* (Braunschweig, 1968), 673-682.

[28] Arnold Sommerfeld, "Über die Fortpflanzung elektrodynamischer Wellen längs eines Drahtes," *AP, 67* (1899), 233-290.

(22) J. A. Fleming, Nature, Vol. 90, Oct. 31 and Nov. 7, 1912, pages 262 and 292, respectively.
(23) J. Zenneck, Leitfaden der Drahtlosen Telegraphie, Stuttgart, 1913.
(24) J. E. Ives, Philosophical Magazine, May, 1913.
(25) L. W. Austin, Journal Washington Academy of Sciences, June 4th, 1913.

Dr. L. W. AUSTIN (by letter): An idea, held some time ago, that the difference between the strength of signals by day and those by night was due to the ionization of the air around the sending antenna, caused by the ultra-violet light from the sun, has been entirely abandoned.

An alternative explanation, that the increased strength of signals at night was due to diminished conductivity in the upper conducting layers of the atmosphere at night, seems improbable in view of data in the possession of the U. S. Navy Department.

This data shows that: (1) In certain regions and at certain wave lengths the ground absorption is greater than over equal stretches of salt water in the ratio of as much as twenty to one. Yet signals are sometimes received in such regions at night with the same strength as if there were no absorption at all. The sunlight can hardly affect ground absorption. (2) When working with sustained waves from arc radio-telegraph sets, the strength of signals may be great on certain wave lengths and very weak at other slightly different wave lengths. Thus a change of two or three per cent. in wave length produces enormous changes in intensity of signals. The probable explanation of this effect is, in accordance with Dr. de Forest's suggestion, (Proc. Inst. Radio Engineers, Vol. I, No. 1, page 37, 1913) that interference between a set of waves travelling along the earth and another set which have been reflected from conducting layers of the upper atmosphere takes place. There is no doubt as to the existence of this effect. The probable reasons for the failure to observe it with spark sets is partly because of the greater changes of wave lengths employed with such apparatus, and partly because the shortness of the wave trains does not permit the direct and reflected waves to overlap and interfere for any considerable number of wave lengths.

The most probable explanation of the increased strength of night signals is to be found in an increase of energy received

BIBLIOGRAPHY.

(1) M. Abraham, Physikalische Zeitschrift, Vol. 2, 1901, page 329.

(2) J. Zenneck, Annalen der Physik, Vol. 23, 1907, page 846.

(3) F. Hack, Annalen der Physik, Vol. 27, 1908, page 43.

(4) P. Epstein, Jahrbuch der Drahtlosen Telegraphie, etc., Vol. 4, 1910, page 176.

(5) A. Blondel, Comptes Rendus du Congrès de Nantes, 1898, page 212.

(6) E. Lecher, Physikalische Zeitschrift, Vol. 3, 1901, page 273.

(7) K. Uller, Beiträge zur Theorie der Elektromagnetischen Strahlung, Rostock, 1903.

(8) K. Uller, Jahrbuch der Drahtlosen Telegraphie, etc., Vol. 2, 1908, page 8.

(9) A. Sommerfeld, Annalen der Physik, Vol. 28, 1909, page 665.

(10) H. March, Annalen der Physik, Vol. 37, 1912, page 29.

(11) H. Poincaré, Comptes Rendus, Vol. 154, 1912, page 795, or Jahrbuch der Drahtlosen Telegraphic, etc., Vol. 3, 1910, page 445.

(12) H. Poincaré, Rendiconti Palermo, Vol. 29, 1910, page 1.

(13) J. W. Nicholson, Philosophical Magazine, April, 1910.

(14) W. v. Rybcnski, Annalen der Physik, Vol. 6, 1913, page 191.

(15) L. W. Austin, Bulletin Bureau of Standards, Vol. 7, 1911, page 315.

(16) Brylinski, Bulletin de la Socité internationale des Electriciens, June, 1906, page 291.

(17) J. A. Fleming, Nature, Nov. 7, 1912, page 294.

(18) J. A. Fleming, The Marconigraph, Vol. II, October, 1912, page 270.

(19) L. de Forest, Proceedings Institute of Radio Engineers, Vol. 1, No. 1, Jan., 1913, page 42.

(20) W. H. Eccles, Proceedings Royal Society, Vol. 87A, 1912, page 79.

(21) M. Dieckmann, Experimentelle Untersuchungen aus dem Grenzgebiet zwischen Drahtloser Telegraphie und Luftelektrizität, Berlin, 1912.

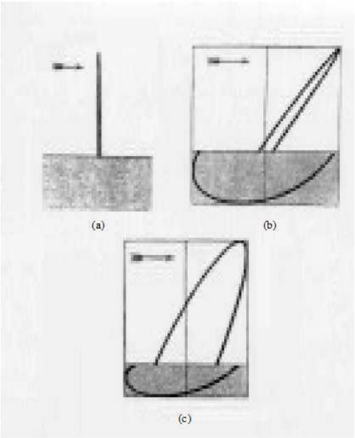

FIG. 4 The polarization diagrams of Zenneck's waves. (a) large ground conductivity; (b) small ground conductivity and dielectric constant; (c) moderate ground conductivity and small ground dielectric constant; The arrows denote the direction of wave transmission. Zenneck (ref. 26), figs. 4-6.

Sommerfeld recognized that he could resolve the questions how or whether Zenneck's surface wave could be produced by the mathematical techniques he had developed in the mid-1890s to tackle diffraction problems and wired waves. In 1909 Sommerfeld showed that Zenneck's "surface wave" could be generated by a vertically polarized Hertzian dipole oscillator located above the flat conducting surface.[29]

---

[29] Arnold Sommerfeld, "Über die Ausbreitung der Wellen in der drahtlosen Telegraphie," *AP*, *28* (1909), 665-736.

Sommerfeld's problem has the same geometry condition as Zenneck's: an infinite flat interface separating the conducting material below and the air above. But different from Zenneck, he put a vertical Hertzian dipole on the ground as the source of waves. To solve the problem with his Hertzian radiator, Sommerfeld expanded the Hertzian potential in terms of a set of basis functions more convenient for mathematical manipulations than Macdonald's. However, Sommerfeld's expansion is not a discrete sum of spherical harmonics with various half-integer orders. As was typical of his mathematical approach in the 1890s, it was an integral expansion involving only the Bessel function of order 0. The different choices of expansion led to an essential difference between the British-French and the German theories of diffraction. Sommerfeld noticed that any cylindrical wave in the form of $CJ_0(q\,r)\exp[(q^2-k^2)^{1/2}z]$, is a solution of the wave equation where q is a free parameter and J_0, is the 0th order Bessel function. Sommerfeld's solution of the Hertzian potential has the form of an integral expansion of cylindrical waves over q.

To evaluate this integral, Sommerfeld used Cauchy's residue theorem to convert the integral over the entire real axis into a contour integral over the entire complex plane. The integral came out as the sum of three terms; one of which had the same form as Zenneck's surface wave. Furthermore, the other terms became insignificant at long distance. In short, Sommerfeld proved that Zenneck's surface wave is the asymptotic solution of the diffracted field produced by a vertically polarized Hertzian dipole sitting just above the flat boundary surface.

The German diffraction theory differed from the British-French rival by indirectly showing that not only the geometric shape of the boundary surface but also the resistance of the ground affect the transmission distance. Moreover, Sommerfeld's approach to the diffraction problems created a unique tradition of practices. Expanding the field with respect to an integral of cylindrical waves differed significantly from the common technique that expanded the field over a sum of discrete-order spherical harmonics. Sommerfeld knew that his approach engendered many problems, which he gave to his students at the University of Munich. Herman William March wrote a thesis in 1911 extending Sommerfeld's integral approach to the diffracted wave along a spherical conductor. Witold von Rybczynski wrote on a similar topic in 1913.

3. ATMOSPHERIC REFLECTION HYPOTHESES

A few very important phenomena could not be explained by a surface wave that depends solely upon the characteristics of the ground. The moisture effect, the daylight effect, and "stray rays" all strongly suggested that the propagation of an electromagnetic wave has to do with air as well as ground. This consideration lay behind the theory that extended wave transmission arises from reflection of electromagnetic waves by an electrically conducting layer in the upper atmosphere. Around 1902, Oliver Heaviside, Arthur Edwin Kennelly, André Blondel, Henri

(a) Long waves, because their concentration factor is larger, that is, they follow the earth more closely. In fact, in

(b) Any conditions where surface waves predominate, e. g., over water.

The increased dielectric constant of moisture laden air may account in part for certain well-defined interference effects, as explained by Dr. Lee de Forest (19).

Dr. W. H. Eccles (20) gives a theory of wave propagation in ionised air which, under certain conditions, leads to conclusions directly conflicting with those of Fleming. Thus, according to Eccles, as a result of ionisation of the upper layers of the atmosphere, the wave front may be tilted forward so as to follow the earth's surface more closely; in fact, ionisation might thus assist long distance transmission. There are, then, two opposing effects, the relative preponderance of which determines whether the wave front is tilted forward or backward.

(a) It is tilted backward if, because of ionisation there is a deposition of water on the ionic nuclei, with a consequent increase in the dielectric constant.

(b) It is tilted forward if the presence of ions of molecular dimensions increases the velocity of the waves.

Further experiment on these points is highly desirable.

It will, however, be noted, that regardless of whether effect (a) or (b) predominates, Professor Kennelly's explanation of the sunrise and sunset effects as due to reflection at the boundary surface between ionised and un-ionised air still holds. Effect (a) would provide a more nearly complete reflection.

It has been further suggested that the antenna may be actually discharged by ultra-violet light in sunlight falling on it (Hallwachs effect). This, and other interesting points relative to the influence of atmospheric conditions on transmission and reception are fully treated in a recent publication by M. Dieckmann (21). The attention of the experimenter in this field is directed particularly to Professor Fleming's valuable paper before the British Association (22) and to the chapter on electric wave transmission in Professor Zenneck's "Leitfaden der Drahtlosen Telegraphie" (23) to both of which sources I desire to acknowledge my indebtedness for suggesting some of the material given in this discussion.

A complete theory of propagation of electromagnetic waves on a sphere of material of *finite* conductivity has not yet been developed.

There is another source of loss of energy in transmission which is generally not considered, namely, the loss thru ground currents in the immediate neighbourhood of the antenna. Inasmuch as considerable absorption of energy undoubtedly takes place in this vicinity, some interest is attached to conclusions drawn by Brylinski (16) relative to the resistance to alternating current of a homogeneous plane conductor of indefinite extent, the current flow being parallel to the plane. (It has been shown experimentally that this is very nearly the case for actual transmitting stations). Brylinski shows that the resistance increases with the specific resistance of the material, with the frequency of the current, and with the damping of the current. Remembering that the nearest approach to an ideal Hertzian doublet, radiating without loss in the surrounding medium would be attained by having a perfect conducting mass extending thru the equatorial plane of the doublet, and that this implies that the ground resistance to an infinite distance is zero, we find that long distance transmission is best attained by the use of sea water grounds, long waves, and continuous radiation of energy. The first two of these conclusions are in accord with the facts, the third point is one requiring further investigation.

Returning to the effect of atmospheric ionisation on transmission, an ingenious explanation of daylight absorption has been suggested by J. A. Fleming (17), (18). If the upper layers of the atmosphere are strongly ionised by sunlight, the ions produced may act as nuclei for the condensation of water vapor. As a result of the presence of the water drops, and the high dielectric constant of water (namely, 80) the upper layers of air have a higher dielectric constant after exposure to sunlight. Fleming experimentally found that the dielectric constant of steam-laden air varied between 1.004 and 1.026. Therefore the electric waves will travel more slowly in the upper layers of air than in the lower, and the wave front will be tilted backward relative to the direction of transmission. In consequence of this tilting back of the wave surface, the entire wave may pass directly over the receiving station. The effect is quite similar to diminishing the concentration factor. On Fleming's hypothesis, daylight absorption should be least for

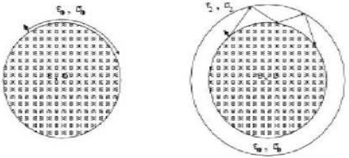
ε_0, σ_0
ε_2, σ_2
ε_0, σ_0

FIG. 5 The surface-diffraction model (left panel) versus the atmospheric-reflection model (right panel); the thicker arrow gives the direction of the Hertzian oscillator, the lighter arrows the direction of wave propagation.

Poincaré, and Charles-Edouard Guillaume all considered the possibility of atmospheric reflection.[30] But only Heaviside and Kennelly published their ideas.

Oliver Heaviside, a self-educated theorist unaffiliated with any academic institute, developed a mathematical model for telegraph-signal transmission, a new formulation of Maxwell's theory, and the technique of operational calculus for solving linear ordinary differential equations.[31] He never studied wireless telegraphy seriously. But his work on wired signal transmission played a significant heuristic role in his thought on the transmission of wireless waves.

In 1901, Heaviside was studying the electromagnetic field patterns produced by an electrical signal transmitted through a coaxial cable. He obtained the electromagnetic field by solving Maxwell's equations for the boundary condition specified. Additionally, he made various geometric metamorphoses of the coaxial cable to other mathematically tractable structures, for example, transformation of the cylindrical condition to a point source sitting on top of a large hemisphere surrounded by another hemisphere.[32] He found that this structure can support the transmission of waves along the curved surfaces and conjectured that the same mechanism accounted for Marconi's wireless signals across the Atlantic Ocean.

He published this idea in 1902, in a short paragraph in an introductory article on telegraphy in the *Encyclopedia Britannica*.[33] In it Heaviside observed that, because both the sea and the land have non-vanishing conductivities, wireless waves could travel along the earth's surface in the same manner as electromagnetic waves

30. Alexander Russell, "The Kennelly-Heaviside layer," *Nature* (24 Oct 1925), 609.

31. Paul Nahin, *Oliver Heaviside, sage in solitude: The life, work, and times of an electrical genius of the Victorian age* (New York, 1987); Ido Yavetz, *From obscurity to enigma: The work of Oliver Heaviside, 1872-1889* (Basel, 1995).

32. Nahin (ref. 31), 279-281.

33. Oliver Heaviside, "The theory of electric telegraphy," *Encyclopedia Britannica* (10th edn., 1902), *33* (1902), 215.

travel along a conducting telegraph wire. He added the possibility that traveling waves are bounded not only by the earth's surface but also by a conducting layer in the upper atmosphere. That was all. He did not discuss the source of the conducting layer, empirical evidence for it, or the behavior of waves propagating between the concentric spherical conductors.

The other man holding the hypothesis of an atmospheric reflection layer was Arthur Edwin Kennelly. Born in India, he left the University of London, where he was a student, to become a full-time telegrapher. After working in the Eastern Telegraph Company for years, he joined Thomas Edison's West Orange laboratory in 1887. The experience transformed him from an operation engineer to a researcher on new electrical technologies. He became a professor of electrical engineering at Harvard University in 1902, and stayed there until 1913.[34] His principal interests were electric circuits and power systems and his contribution to wireless telegraphy was comparatively minor. His primary contribution was popularizing the theories and technologies of wireless telegraphy to the public by providing simple physical explanations of the phenomena encountered in wireless practices. He emphasized physical intuition rather than complicated mathematical theory.

In 1902 Kennelly published a paper with a model similar to Heaviside's.[35] Unlike Heaviside, however, Kennelly explained the existence of the atmospheric conducting layer. The explanation rested on J.J. Thomson's discovery that the air has an electric conductivity when thinned. The more dilute the air, the higher the conductivity. Kennelly evoked the standard dependence of atmospheric pressure on height and Thomson's extrapolated experimental formula to deduce that at the height of 80 kilometers the air conductivity is 20 times that of sea water! After securing the causal explanation of the atmospheric conducting layer, Kennelly argued that if the propagation space is confined by the atmospheric conducting layer and the earth surface, then the wave's energy density diverges as fast as a cylindrical wave's. In the absence of an atmospheric conducting layer, the energy density diverges as fast as a spherical wave's. The argument concerning energy divergence plus the atmospheric reflection model account for a higher field intensity over a given distance.

Kennelly's paper in 1902 gave a causal explanation of the existence of the conducting layer and a more exact estimate of its height. But he did not connect the hypothesis with observed phenomena. Nor did he devise a quantitative theory of wave transmission from Maxwell's equations and the given boundary condition, as the diffraction theorists had done.

---

[34] Karl Willy Wagner, "Arthur Edwin Kennelly, zu seinem 70. Geburtstage," *Elektrische Nachrichten Technik, 8: 12* (1931), 1.

[35] Arthur E. Kennelly, "On the elevation of the electrically-conducting strata of the earth's atmosphere," *Electrical world and engineer*, 15 Mar 1902; reprinted in "Kennelly-Heaviside ionized layer—a classic of science," *Science news letter, 17* (18 Jan 1930), 45.

---

papers relate to the superiority of long wave transmission, and transmission over water. A further aspect of the problem is considered in them, namely the effect of the curvature of the conducting sheet or ground.

It will be remembered that Sommerfeld regarded this conducting sheet as plane, and it would certainly seem that the transmission of electromagnetic waves over an earth quadrant despite the curvature requires explanation. The papers just cited supply this explanation. They show that

(a) For a spherical, perfectly conducting ground, the energy of the wave decreases not as $\frac{1}{S}$ but as $\frac{1}{S}\sqrt{\frac{\theta}{\sin\theta}}$ (6)

where θ is the angular separation of the points considered measured along a great circle of the earth (neglecting "scattering").

It follows, therefore, that the amplitude at any distance from the antenna will be greater than for a plane ground, and may even rise to considerable values at the antipodes. There is, however, a second consideration, namely, the failure of the energy to fully follow the earth's surface and its consequent re-radiation or "scattering." It results in the introduction of what I shall call the "concentration factor."*

The theory shows that for a spherical, perfectly conducting ground

(b) The energy of the wave, taking account of scattering or re-radiation of energy from the surface wave, is obtained by multiplying the value obtained under statement (a) above by the concentration factor:

$$\varepsilon^{-0.0019\frac{s}{\sqrt[3]{\lambda}}} \qquad (7)$$

where λ is the wave length (wave length and distance expressed in kilometers).

The calculated values of the concentration factor for a distance of 5,000 km. is 0.0025 for a wave length of 4,000 meters and 0.0086 for a wave length of 8,000 meters. The advantage of using long waves is again apparent.

The expression for the energy of the wave obtained according to statement (b) above is in reasonable agreement with the results obtained experimentally by Dr. Louis W. Austin (15).

---

*Usually called "Zerstreuungs Faktor," or "scattering factor."

media (that is, in space, and in the partially conducting sheet).

The relative importance of the space and surface waves is found to be determined by the value of ρ. The value of ρ for sea water is 0.03, for pure water 30, for wet earth 6.5, and for dry soil 300 (for a wave length of 2,000 meters at a distance of 2,500 km.) It is seen that the numerical distance increases with the real distance, diminishes with increased wave length, and is less for equal distances over sea water than over land. Furthermore, Sommerfeld's analysis shows that the assumption of perfect conductivity of the ground is allowable only for short distances over sea water. The distance from the antenna at which the surface or space waves predominate is determined as follows:

(a) For small values of the numerical distance, the space waves predominate.
(b) For larger values of the numerical distance, surface waves predominate.
(c) For very large values of the numerical distance, the space waves may again predominate, but this last effect may be neutralized by the effect of the curvature of the earth, and is probably not important in practise.

The surface wave is the more desirable one for long distance reception because it decreases far more slowly with distance than the space wave. The numerical distance should therefore be kept small so that the surface waves soon predominate. This may be secured by the following means:

(a) Increasing the wave length.
(b) Increasing the conductivity of the ground.
(c). Increasing the dielectric constant of the ground.

These conclusions are well borne out in practise. Sommerfeld has also suggested that daylight absorption is due to higher conductivity of the air causing an increase in U_1, and thereby increasing the numerical distance. In support of this, he mentions that Ebert found that the conductivity of air at a height of 2,500 meters was 23 times greater than at the earth's surface. It is, however, the opinion of other investigators, notably Messrs. Zenneck and Pierce, that the ionisation of the air due to sunlight is entirely insufficient to account for the magnitude of the observed effects.

Sommerfeld's work has been carried further by H. March (10), H. Poincaré (11), (12), J. W. Nicholson (13), and W. V. Rybczynski (14). The practical conclusions to be drawn from all these

---

The experiments done by Louis Austin's team at the U.S. Naval Wireless Telegraphic Laboratory in 1910 were thought to be the only high-quality evidence for long-distance wireless-wave transmission. The mathematical formula synthesized from the experimental results, known as the "Austin-Cohen formula," turned out to be the empirical law governing the relationship between the field intensity and the large transmission distance. Nevertheless, the broad acceptance of the empirical formula made people overlook the fact that the Navy's "experiments" were actually organized not for scientifically investigating long-distance wave transmission phenomena, but for testing the equipment of the Navy's recently installed high-power wireless-telegraphy station.

In 1899 the Bureau of Equipment of the U.S. Navy established a Radio Division to look into replacing flag-and-light signaling with wireless. The plan failed owing to cultural gaps and the associated organizational inertia. Experienced seamen rejected the new gadgets without performance records. Engineering officers living in a conservative military culture clashed with the wireless inventors who preferred flexibility and novelty. The Bureau of Equipment did not have enough power to overcome the opposition. The Navy ended up on-board radio sets rarely used by the combat units and without a system of standardized training in operational and maintenance procedures.[36]

The opposition had much to say for its position. The wireless technology of the early 20th century was unreliable. The mainstream transmitter (spark-gap discharger) suffered from its highly damped oscillation. The mainstream receiver (coherer) could not follow faithfully the continuous variation of a signal and performed unstably. To incorporate the radio into operation, the Navy required reliable data and reliable instruments for which they could develop standard operational-maintenance procedures and systematic procurement schemes. It needed a research establishment for equipment tests, accurate measurements, and technical evaluation of wireless technologies.

The head of the Radio Division, Cleland Davis, and the director of the U.S. National Bureau of Standards Samuel Wesley Stratton, agreed to station the laboratory be stationed within the existing organizational structure of the Bureau. The Naval Wireless Telegraphic Laboratory thus was born.[37] Its first head, Louis Winslow Austin, a physicist then working at the Bureau, had obtained a doctoral degree at the University of Strassburg. After graduation, he worked at the Reichsanstalt, the German model for the Bureau of Standards, which he joined on his return to the U.S. in 1904. He was transferred to the Navy Department in 1908 in order to head the Wireless Telegraphic Laboratory. He started with an assistant, George H. Clark, and a few part-time technicians.

---

[36] Susan Douglas, "Technological innovation and organizational change: The Navy's adoption of radio, 1899-1919," in Merritt Roe Smith, ed., *Military enterprise and technology change: Perspectives on the American experience* (Cambridge, 1985), 117-173; Linwood S. Howeth, *History of communications-electronics in the United States Navy* (Washington, D.C., 1963), chapts. 12, 13.

[37] Louis W. Austin, "The work of the U.S. Naval Radio-Telegraphic Laboratory," American Society of Naval Engineers *Journal, 24* (1912), 122-141.

In 1908, the Navy issued a tender for its first high-power wireless station in Arlington, Virginia. It specified that the transmitter should be capable of sending messages at all times and at all seasons to a radius of 3000 miles in any navigable direction. The National Electric Signaling Company (NESCO) obtained the contract in 1909. The transmitter would be similar to the 100 kilowatt Fessenden synchronous rotary spark discharger then installed at NESCO's experimental wireless station at Brant Rock, Massachusetts. Both the Navy and NESCO knew that this machine could not meet the long-distance specification in the contract. Nonetheless, the Navy awarded the contract with NESCO in the belief that the Fessenden discharger was the best available. To legitimate the compromise, the Navy required further technical tests and measurements of the existing system at Brant Rock while the construction for the Arlington station proceeded.[38] The task of testing and measuring the 100-kw rotary spark discharger at the Brant Rock station was assigned to the Naval Wireless Telegraphic Laboratory. During the summer and autumn of 1909, the electricians of the Laboratory conducted preliminary measurements on the wireless sets at Brant Rock and on communications from there to the scout cruisers *Birmingham* and *Salem*.[39]

From subsequent field tests done in July, 1910 Austin synthesized the first empirical formula governing the relation between distance and radiation intensity for long-range wave transmission. That realized the major aim of the tests: to evaluate the communication qualities of particular wireless sets, namely the Fessenden rotary-spark transmitters, not to produce an empirical law for the science of wave propagation. Austin paid attention to particular instrumentality rather than general regularity.

The experiments of 1910 were not the first attempts to obtain quantitative relations between received antenna current and distance. William Duddell and J.E. Taylor in Britain and Camille Tissot in France did similar experiments in 1904 and 1906. Within a range of 50 miles, they discovered that the current is inversely proportional to distance.[40] Austin's tests ran to 1000 miles. *Birmingham* sailed about 1200 miles south from Brant Rock, *Salem* about 450 miles southeast; through the voyages the wireless sets at Brant Rock and on the cruisers regularly transmitted and received signals from one another, both day and night. The whole experimen-

---

[38] Hugh G.J. Aitken, *The continuous wave: Technology and American radio, 1900-1932* (Princeton, 1985), 88.

[39] Austin (ref. 52), 125, 147-153, and Louis W. Austin,"Some quantitative experiments in long-distance radiotelegraphy," Bureau of Standards, *Bulletin, 7:3* (1911), 315-363.

[40] William Duddell and J.E. Taylor, "Wireless telegraphy measurements,"*Electrician, 55* (1905), 258-261, and Camille Tissot,"Note on the use of the bolometer as a detector of electric waves,"*Electrician, 56* (1906), 848-849.

part in radio telegraphy are A. Blondel (5), E. Lecher (6), and K. Uller (7), (8). However, it remained for A. Sommerfeld in 1909 (9) to give a broad theoretical treatment of the spread of electromagnetic waves over a partially conducting sheet placed, as before, in the equatorial plane of the oscillator; the sheet being plane. Because of their importance, Sommerfeld's explanations and his conclusions will be considered in some detail.

He first discusses the fundamental energy distinction between space and surface waves. Space waves, such as electromagnetic (Hertzian) waves and sound waves, spread three-dimensionally. Their energy (per unit area of wave front) is therefore proportional to the inverse square of the distance from the source. Surface waves, such as ripples on water, electric waves on wires, and elastic waves of surface distortion in solids, decay according to a different law of energy diminution. Thus, for waves spreading circularly over water, and neglecting dissipative absorption in the medium, the energy varies inversely with the distance; not as the inverse square.

Sommerfeld defines a quantity U in his paper as follows:*

$$U^2 = \frac{K \mu n^2 + j \mu \sigma n}{c^2} \tag{3}$$

where K is the dielectric constant, μ is the permeability,
σ is the conductivity, j is $\sqrt{-1}$,
n is the frequency, c is the velocity of light.

From U_1 and U_2, the values of U for air and the conducting sheet respectively, he builds up a new quantity ρ, which he calls the "numerical distance." If S is the actual distance, ρ is defined by the following equation:

$$\rho = \frac{U_1^4}{U_2^4} \cdot \frac{U_1^2 - U_2^2}{U_1^2} \cdot \frac{U_1 s}{2} \tag{4}$$

Sommerfeld sets up the partial differential equation for his problem, introduces the necessary boundary conditions, and, after a series of elaborate analytical transformations, arrives at the following value for Π, the vector potential at any point:

$$\Pi = P + Q_1 + Q_2 \tag{5}$$

Herein P represents that portion of Π. which is due to a true surface wave, and Q_1 and Q_2 represent space waves in the two

---

*The quantity U is Sommerfeld's "k". The notation here employed is, where possible, that recommended by the Standardisation Committee of THE INSTITUTE OF RADIO ENGINEERS.

(b) The electric and magnetic field intensities are in phase.

It will be noted that these conditions apply only when the equatorial plane is a *perfect* conductor. This is nearly the case for sea water, but not at all for dry land. J. Zenneck considered this problem in detail for the case of an imperfect conducting plane. (2) For the case of the partial conductor he found

(a) The lines of magnetic force are parallel to the plane, the lines of electric force are inclined in the direction of motion of the wave.

(b) The electric force has a horizontal component which is nearly in phase with the magnetic force.

(c) The vertical component of the electric force and the horizontal component of the electric force are out of phase with each other.

There is a resultant rotary electric field at the surface of the plane.

Zenneck gives the following values for the field forces at a distance from an antenna of height h, of form factor α, in which flows a current I_m, and from which are radiated waves of length λ:

$$\text{Electric Force} = 4\pi \frac{\alpha h}{\lambda} I_m \frac{\varepsilon^{-\beta s}}{s} 3.(10)^{10} \quad \text{C. G. S. Units} \quad (1)$$

$$\text{Magnetic Force} = 4\pi \frac{\alpha h}{\lambda} I_m \frac{\varepsilon^{-\beta s}}{s} \quad (2)$$

where β is a constant. It will be seen that the values of these forces are dependant on the wave length.

The effects of the conductivity of the ground plane and of underground water or rock have been exhaustively studied by F. Hack (3) and P. Epstein (4) has given figures for the electric lines of force of actual waves (of length of 2,000 meters) passing over soil of high conductivity, and at considerable distances from the antenna.

Of particular interest in these latter papers is the proof of the existence of a varying horizontal component of the electric force on the wave front. Such electric forces exist on wires along which guided or surface electric waves are passing, and the question naturally arises whether we are not dealing with surface waves if the case of terrestrial radio transmission. Among the earlier investigators of such guided waves and of their probable

tal process, including instrument calibration and maintenance, on-site measurements, and data analysis, was executed by electricians hired by the Naval Wireless Telegraphic Laboratory and engineers of NESCO.[41]

Austin carefully documented the instrument specifications in the report he wrote for the Bureau's *Bulletin*. The transmitters were Fessenden synchronous rotary-spark dischargers, powered at 100 kw at the Brant Rock station and 2 kw on the cruisers. The broadcasts occurred at 1000 meters and 3750 meters. When the separation between the transmitter and the receiver exceeded 100 miles, operators used charcopyrite zinkite rectifiers with galvanometers, and shunted telephone circuits, to detect the antenna current. Both kinds of sensors had high sensitivity for weak current.

During the voyage the electricians noticed several unusual phenomena. After the departure of the ships, the signals became too weak for detection by the crystal-rectifier detector. Except for a few signals taken in the first two days, all the data were taken by the shunted telephone. In addition, the electricians observed that the signals received at night were significantly more erratic than those received during the day. The signal level at night was usually stronger, but it had more fluctuation and experienced more disturbance. The nine-day experimental campaign produced five tables of measured data (for example, see figure 6).

The Austin-Cohen formula

Austin at first tried to fit the daytime data with Duddell's formula that made received antenna current inversely proportional to distance. The inverse law gave much higher values than the measured data beyond 200 miles. Then he tried the ad hoc assumption that the received current experiences an additional exponential factor owing to atmospheric absorption. The NESCO engineer Louis Cohen discovered that a fixed damping coefficient A could reproduce all the data for the same wavelength; to capture the data better for both wavelengths, he set $A = \alpha\lambda^{-1/2}$. Thus the exponential factor is $\exp[\alpha\lambda^{-1/2}d]/d$. They specified that $\alpha = 0.0015$, when the distance d and wavelength λ are expressed in kilometers.[42]

The Laboratory also sought an empirical law governing the transmitted antenna current and the heights of the transmitting and receiving antennas. Experiments conducted between Brant Rock and Washington, D.C. after the sea voyage in mid-July 1910 showed that the receiving antenna's current was proportional to the transmitting antenna current and to the heights of both antennas, and inversely proportional to the wavelength (regardless of the exponential decay factor). Combining these relations, Austin obtained the famous empirical "Austin-Cohen formula:"

---

[41] Austin (ref. 39), 320-330.

[42] Ibid, 326-327.

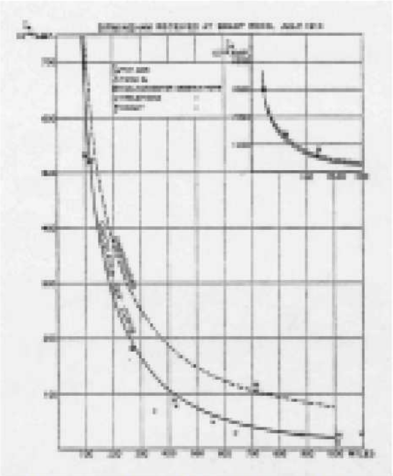

FIG. 6 Birmingham's signal as and received at Brant Rock (July 1910); 3750 meters. The vertical axis represents the received antenna current, the horizontal axis the distance. The solid curve is calculated from the Austin-Cohen formula, the dashed curve from the inverse law "N" denotes data from night-time measurements. Austin (ref. 39), fig. 4.

where I_r is the current received through an antenna with an equivalent resistance of 25 ohm; I_s, the transmitting antenna current; h_1, the height of the transmitting antenna; h_2, the height of the receiving antenna; λ the wavelength; d the distance; and $\alpha=0.0015$; all the lengths being in kilometers and the current in amperes. According to Austin, the formula is "an equation which will cover the normal day received current over salt water through 25 ohm for two stations with flat-top

up this question experimentally, and should endeavour to lay a firm foundation for a theory of long-distance radio-telegraphy by accumulating information on the important question of the electric conductivity and ionisation of our atmosphere at various heights above the earth's surface. Something might be done by the use of a Gerdieu's ionimeter in dirigible balloons or in aeroplanes.

University College, London; June 24th, 1913.

ALFRED N. GOLDSMITH: There are a number of effects in the transmission and reception of signals by radio communication which are as yet only partly explained. Those most closely related to the present paper may be classified as follows:

(1) The superiority of transmission over water as compared with transmission over land.

(2) The superiority (under certain conditions) of transmission on high wave lengths as compared with transmission on short wave lengths.

(3) The clinging of electromagnetic waves to the earth's surface, in spite of the considerable curvature of the globe.

(4) The superiority of transmission by night as compared with transmission by day.

(5) The "daylight effect," or change in intensity of received signals at or near sunrise and sunset.

Before classifying the proposed explanations of these effects, we shall consider briefly the bibliography of the propagation of electro-magnetic waves thru space and over the surfaces of more or less perfect conductors. The classical papers of Hertz and his immediate followers have rendered wave transmission thru space alone thoroly clear. The next problem in order of complexity is the radiation from a Hertzian doublet (dumb-bell oscillator) thru the equatorial plane of which a perfect conductor of indefinite extent stretches. Such a plane conductor may be regarded as a first approximation to the "ground" of a radio-telegraphic station.

M. Abraham, in 1901, gave the mathematical solution of this case (1) and its physical interpretation. Considering the wave at a distance from the oscillator, he found that

(a) The lines of electric force were perpendicular to the conducting plane, the lines of magnetic force parallel to it.

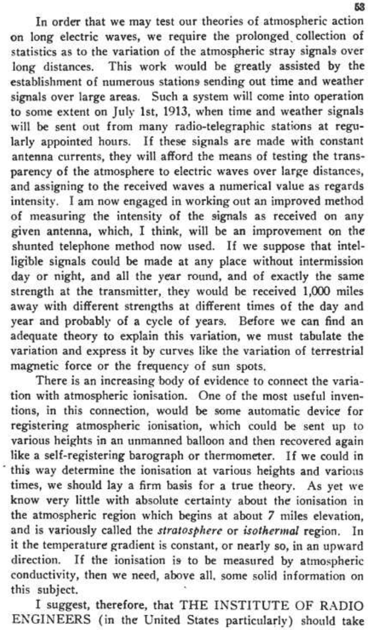

In order that we may test our theories of atmospheric action on long electric waves, we require the prolonged collection of statistics as to the variation of the atmospheric stray signals over long distances. This work would be greatly assisted by the establishment of numerous stations sending out time and weather signals over large areas. Such a system will come into operation to some extent on July 1st, 1913, when time and weather signals will be sent out from many radio-telegraphic stations at regularly appointed hours. If these signals are made with constant antenna currents, they will afford the means of testing the transparency of the atmosphere to electric waves over large distances, and assigning to the received waves a numerical value as regards intensity. I am now engaged in working out an improved method of measuring the intensity of the signals as received on any given antenna, which, I think, will be an improvement on the shunted telephone method now used. If we suppose that intelligible signals could be made at any place without intermission day or night, and all the year round, and of exactly the same strength at the transmitter, they would be received 1,000 miles away with different strengths at different times of the day and year and probably of a cycle of years. Before we can find an adequate theory to explain this variation, we must tabulate the variation and express it by curves like the variation of terrestrial magnetic force or the frequency of sun spots.

There is an increasing body of evidence to connect the variation with atmospheric ionisation. One of the most useful inventions, in this connection, would be some automatic device for registering atmospheric ionisation, which could be sent up to various heights in an unmanned balloon and then recovered again like a self-registering barograph or thermometer. If we could in this way determine the ionisation at various heights and various times, we should lay a firm basis for a true theory. As yet we know very little with absolute certainty about the ionisation in the atmospheric region which begins at about 7 miles elevation, and is variously called the *stratosphere* or *isothermal* region. In it the temperature gradient is constant, or nearly so, in an upward direction. If the ionisation is to be measured by atmospheric conductivity, then we need, above all, some solid information on this subject.

I suggest, therefore, that THE INSTITUTE OF RADIO ENGINEERS (in the United States particularly) should take

---

antennas of any height, with any value of sending current and any wave length, provided the sending station is so coupled as to give but one wave length." It had the further limitation that it had been tested on only the Fessenden rotary spark charger. The coefficient 4.25 in equation (1) is particularly dependent on the instrument. Austin pointed out it applied strictly only to the antennas used in the experiments of 1910.[43]

The Austin-Cohen formula became the most important, and perhaps the only quantitative, empirical basis for all theories of long-distance wireless wave transmission. Before its publication in 1911, scientists worried whether their theories were consistent with physical intuition or wireless know-how, after 1911, they worried whether their numerical results fit Austin's quantitative data or the Austin-Cohen formula. That altered the epistemic situation significantly.

5. DIFFRACTION THEORIES AGAIN

The fit between diffraction theories and the Austin-Cohen formula was far from satisfactory. The theorists adopted several strategies to save the day. The German physicist Witold von Rybczynski showed that his calculations agreed with selected data from Austin's report better than the Austin-Cohen formula did. The English physicist Augustus Edward Hugh Love questioned whether the detector in Austin's experiments had functioned properly. Meanwhile, the experimenters strengthened their position by confirming the Austin-Cohen formula. The test campaign made by the U.S. Naval Radio Telegraphic Laboratory in 1913, which was planned to choose between the rotary-spark discharger and the arc oscillator as the future standard transmission technology, gave an opportunity to produce more measured data. The Austin-Cohen formula survived the tests.

$$I_r = 4.25 \frac{{}_s h_1 h_2}{d\lambda} \exp\left[\frac{-\alpha d}{\tilde{A}\lambda}\right] \tag{1}$$

The Germans

Sommerfeld assigned the problem of wireless-wave diffraction along the earth surface to an American doctoral student, Hermann William March, who completed the project in 1911.[44] March expressed the fields as the spatial derivatives of a Hertzian potential and the general solution of the wave equation as an expansion in spherical harmonics. The British theorists took such an expansion as a sum over a discrete index. In contrast, March expanded the solution as an integral over a continuous index. He matched the integral with the spherical boundary condition to determine the functional form of the integrand, which turned out to be inversely proportional to the derivative of the Hankel function with argument *ka*. To evaluate the integral, March developed the Hankel function for large *ka* asymptotically in the same way as Macdonald had done. Then the integral could be calculated analytically. In March's final result, the Hertzian potential is proportional to $\exp[-ika\theta]/(\theta\sin\theta)^{1/2}$ for large *ka* (θ is the usual angle of separation between the transmitter and the receiver). March's solution did not have any exponential decay with respect to θ like $\exp[-\rho(ka)^{1/3}\theta]$; the exponent $\exp[-ika\theta]$ is a sinusoidal function of θ.

---

[43] Ibid, 340-341.

[44] Hermann William March, "Über die Ausbreitung der Wellen der drahtlosen Telegraphie auf der Erdkugel," *AP, 37* (1912), 29-50.

March's result, though fresh, had the defect of decaying much more slowly with distance than British theorists predicted and wireless practitioners observed. March's mathematical problem was identical to Macdonald's: in integrating the derivative of a Hankel function for large *ka*, the asymptotic expansion cannot be performed in the usual way. The publication of a new version of this error energized Poincaré, who pointed out the illegitimate approximation in a letter to Sommerfeld.[45] In a short published note on the matter, Poincaré observed that March's predictions disagreed with Austin's experimental data.[46]

Sommerfeld put Rybczinski in the problem. Following Poincaré, Rybczinski took the dominant contribution of the integral from the pole with the smallest imaginary part.[47] He replaced March's integral with another similar in functional form; but the value of Rybczinski's integrand at the projected point of the dominant pole on the real axis equalled that of Poincaré's integrand at the same point. He thus retained the virtues of both March's and Poincaré's approximations: the new integrand was more precise than March's in the region that gives the dominant contribution, and much easier to integrate analytically than Poincaré's. Rybczinski obtained a radiation intensity proportional to $\exp[-0.33(ka)^{1/3}\theta]/(\theta\sin\theta)^{1/2}$.

Rybczinski showed more persuasively than March that the diffraction theory could produce an exponentially damped wave along a large spherical conductor. Rybczinski's factor $\exp[-0.33(ka)^{1/3}\theta]$ differed from Nicholson's factor $\exp[-0.7(ka)^{1/3}\theta]$, and both formulas disagreed with Austin-Cohen's formula in their dependence on wavelength. The diffraction theories stuck with decay rates inversely proportional to $\lambda^{1/3}$, whereas the empirical regularity required $\lambda^{1/2}$. Nicholson's formula decays significantly faster than Rybczinski's, the Austin-Cohen formula significantly slower. As we know, Rybczinski justified his form by appealing to a selected set of data from Austin's paper of 1911 (seven daytime and four nighttime data points for $\lambda = 3750$ meters between 400 and 1000 miles). For these cases, Nicholson's predictions were too low, the Austin-Cohen's too high.

Rybczinski's theory met with approval by some wireless telegraphers, especially the German school. They did not think that he had given a complete physical

---

[45] Henri Poincaré to Arnold Sommerfeld, 1 Jan 1912 (DM, HS 1977-28/A,266).

[46] Henri Poincaré, "Sur la diffraction des ondes hertziennes," Académie des Sciences, Paris, *Comptes rendus, 154* (1912), 795-797.

[47] Witold von Rybczinski, "Über die Ausbreitung der Wellen in der drahtlosen Telegraphie auf der Erdkugel," *AP, 41* (1913), 191-208.

---

probably be willing to assist by furnishing a continuous record of the current and voltage in their sending antenna.

SUMMARY: The influence of solar radiation on radio transmission is discussed. The changes of intensity of signals near sunrise and sunset are explained by reflecting effects which may be expected at the boundary surface or "shadow wall" between darkness (air of small conductivity) and illumination (ionised air of marked conductivity). The theory and recorded observations are found to be in reasonable agreement.

It is proposed that amateur radio telegraphists shall cooperate with THE INSTITUTE OF RADIO ENGINEERS in gathering data on the strength of received signals under various conditions.

DISCUSSION.

DR. J. A. FLEMING (by letter): The subject of Dr. Kennelly's Paper is one which continues to attract great attention from radio-telegraphists on account of its practical importance. Unfortunately, it is a large scale effect, and one not very amenable to laboratory experiments. Much, however, may be done by systematic observations of stray atmospheric waves and on the signals sent out from large stations.

At the British Association Meeting last year, at Dundee, I had the honor of opening a discussion on "Some of the Unsolved Problems of Radiotelegraphy," and at the conclusion of that paper made the suggestion that the British Association should appoint a Radio Telegraphic Committee to bring conjoint investigation to bear on some of these questions. That suggestion was adopted, and the Committee appointed with Sir Oliver Lodge as Chairman. The Committee has already held one meeting. I hope that, as one result, an attempt will be made to organize systematic observations on the number and intensity of atmospheric stray waves during the hours of day and night, and over large areas. This is a work in which we may enlist the assistance of amateurs, and it has the advantage that it is entirely receptive work, and involves no production of waves or disturbance of the ether. Much of the amateur work hitherto done has been merely playing at radio-telegraphy, and has had no other result except to disturb commercial work.

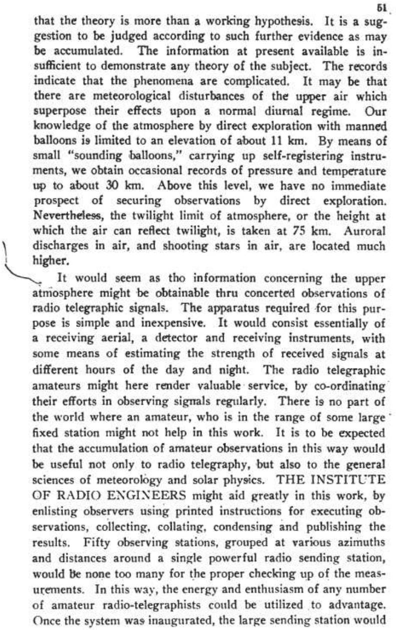

that the theory is more than a working hypothesis. It is a suggestion to be judged according to such further evidence as may be accumulated. The information at present available is insufficient to demonstrate any theory of the subject. The records indicate that the phenomena are complicated. It may be that there are meteorological disturbances of the upper air which superpose their effects upon a normal diurnal regime. Our knowledge of the atmosphere by direct exploration with manned balloons is limited to an elevation of about 11 km. By means of small "sounding balloons," carrying up self-registering instruments, we obtain occasional records of pressure and temperature up to about 30 km. Above this level, we have no immediate prospect of securing observations by direct exploration. Nevertheless, the twilight limit of atmosphere, or the height at which the air can reflect twilight, is taken at 75 km. Auroral discharges in air, and shooting stars in air, are located much higher.

It would seem as tho information concerning the upper atmosphere might be obtainable thru concerted observations of radio telegraphic signals. The apparatus required for this purpose is simple and inexpensive. It would consist essentially of a receiving aerial, a detector and receiving instruments, with some means of estimating the strength of received signals at different hours of the day and night. The radio telegraphic amateurs might here render valuable service, by co-ordinating their efforts in observing signals regularly. There is no part of the world where an amateur, who is in the range of some large fixed station might not help in this work. It is to be expected that the accumulation of amateur observations in this way would be useful not only to radio telegraphy, but also to the general sciences of meteorology and solar physics. THE INSTITUTE OF RADIO ENGINEERS might aid greatly in this work, by enlisting observers using printed instructions for executing observations, collecting, collating, condensing and publishing the results. Fifty observing stations, grouped at various azimuths and distances around a single powerful radio sending station, would be none too many for the proper checking up of the measurements. In this way, the energy and enthusiasm of any number of amateur radio-telegraphists could be utilized to advantage. Once the system was inaugurated, the large sending station would

picture of long-distance wireless-wave transmission. Trans-Atlantic radio communications was still not fully explained. But at a minimum Rybczinski's theory gave an approximately correct picture of wave propagation along the earth without atmospheric reflection. Sommerfeld wrote to Wilhelm Wien in November, 1913. that Rybczinski's theory was analytically right and gave a rate of attenuation that matched experiment to some extent. But (Sommerfeld continued) reflection from conducting layers in the upper atmosphere probably played the major role for very long-distance transmission.[48]

Rybczinski's work did not fare as well outside as inside the Sommerfeld school. His ad hoc procedure of approximation lacked mathematical rigor. He did not select the most reliable data. His seven points, taken from the voyage of 1909, were obtained with instruments still under adjustment and under severe weather conditions. And four of his data points came from measurements at night time, when wireless signals are notoriously unstable. In a paper written in 1914, Austin criticized Rybczinski's data points for these reasons.[49]

A new experiment

The results from the test voyages of 1910 did not significantly hinder the promise of the Fessenden rotary-spark discharger. The Arlington station went on line in February, 1913. The Laboratory tested the transmission equipment. The U.S.S. *Salem* proceeded to Gibraltar, 6400 km Washington, D.C.[50] A critical task of the voyage was to compare the signal from the Fessenden rotary-spark transmitter with that from a new technology—the arc transmitter.

The arc transmitter used the negative resistance created by an electric arc in a low-density gas. Because the value of the negative resistance was precise, the electric arc could be used to design an oscillating circuit with a sharp resonance spectrum. The acceptance test for the Arlington station offered a good opportunity to carry out systematic measurements of the arc transmitter's operational conditions. *Salem* simultaneously received signals from the 100 kw Fessenden transmitter. The round trip to Gibraltar took about six weeks. The arc out-performed the spark.[51] Austin: "At distance over 1000 miles the arc waves appear to begin to show advantages over the spark waves."[52]

The Gibraltar voyage also offered an opportunity to test the Austin-Cohen formula under a set of physical conditions different from those of the 1910 tests: a new (arc) transmitter, a different set of wavelengths (3800 and 2000 meters), and a much longer maximum distance (3500 miles).

---

[48] Arnold Sommerfeld to Wilhelm Wien, 29 Nov 1913 (DM, NL 56, 010).

[49] Louis W. Austin,"Quantitative experiments in radiotelegraphic transmission," Bureau of Standards, *Bulletin, 11* (1914), 69-86.

[50] Howeth (ref. 36), 178-183.

[51] Austin (ref. 49); John L. Hogan, "Quantitative results of recent radio-telegraphic tests between Arlington, VA., and U.S.S. `Salem'," Electrician, 63 (1913), 720-723.

[52] Austin to the Chief of the Bureau of Steam Engineering, 3 Apr 1913, in RG 19, Bureau of Ships, E 988, 841(24), Box 1926, National Archives, Washington, D.C.

Austin and John L. Hogan, an engineer at NESCO, found that the day-time data in 1913 agreed well with the empirical formula. Hogan called the agreement "exceptionally close."[53] Austin went further. He compared Rybczinski's formula, the Austin-Cohen formula, and the new experimental data. Most data points fell closer to the Austin-Cohen formula than to Rybczinski's. Austin again: "There can be no doubt from these results that the theoretical equation [Rybczinski's] gives values too low to be reconciled with the observations, but that they are in very fair agreement with the semi-empirical equation."[54]

Still, something might be retained from the diffraction approach. Austin wrote Zenneck:[55]

> I am becoming quite convinced that the theoretical transmission formula, given in your book, represents approximately the weakest signals observed; while our Navy formula gives a fairly good average. Although I have taken a great many observations, I am still somewhat doubtful regarding the power to which the wave length should be raised in the exponential term. The observations are exceedingly discordant, apparently due to selective reflection.

The British

In January, 1914, Macdonald proposed a new method to approximate the infinite series of his diffraction theory.[56] He introduced a new series, much easier to sum up, that approximated the original series quite well where the spherical harmonic of $n+1/2$ equals ka. Macdonald argued that the new sum well approximated the old one because the dominant contribution of the sum came from this neighborhood. There he replaced the Hankel functions of order $n+1/2$ and its derivatives by Hankel functions of order 1/3 and 2/3. He evaluated the integral via Cauchy's residue theorem. When ka is large, the pole of the Hankel function of order 2/3 with minimum imaginary part dominates. Macdonald found that the resultant field intensity had an exponential decay in the form of $\exp[-\beta(ka)^{1/3}\sin(\theta/2)]$. The functional form of Macdonald's exponential decay differed from both Nicholson's and Rybczinski's factors, $\exp[-\rho(ka)^{1/3}\theta]$.

Enter Augustus Edward Hough Love, still another Cambridge-trained mathematician (second Wrangler, 1885). At the time he contributed to the theory of wireless telegraphy he was a professor of Natural Philosophy at Oxford.[57] In a paper published in 1915, Love gave a comprehensive overview of the research

---

[53] Hogan (ref. 51), 721.

[54] Austin (ref. 49), 77-79.

[55] Austin to Zenneck, 14 Sep 1916 (DM, NL 053).

[56] Hector M. Macdonald, "The transmission of electric waves around the earth's surface," PRS, 90 (1914), 50-61.

[57] *Obit.*, *3* (1939-41), 469-470.

black dots. There is a reinforcement after the dip, and perhaps also before it. On the other hand, at sunset, no dip is indicated, such as the foregoing theory would indicate.

Figure 7 is taken from observations by Messrs. Dolbear and Proctor (8) in March, 1911. Here Glace Bay is again the sending station; while there were two independent receiving stations, one at Somerville, and the other at Revere; both suburbs of Boston. The diagram, taken from the published article, purports to give the general average of all the observations made at both receiving stations, at a period of the year near the vernal equinox. It will be seen that there is a dip both at mid-sunrise and at mid-sunset, with reinforcements both before and after each dip. There is a nocturnal maximum and a daylight reduction. The results set forth in this particular diagram are in closer accordance with the hypothesis here put forward than almost any others, but no explanation was offered or theory advanced in the article, by its writers.

Figure 8 gives some observations reported from Madrid, and Barcelona in Spain, and also Ceuta, in Africa, nearly opposite to Gibraltar. Here the nocturnal maximum is very short. There is a sunrise and sunset dip, with reinforcement before and after.

Figure 9 gives a published series of observations reported from Clifden, Ireland, as to the diurnal strength of signals from Glace Bay, N. S. One column gives the diurnal chart for each day in April, 1911. Here the agreement with the theory is not so good. There is usually, but not always, a dip at mid-sunrise and mid-sunset. Sometimes the sunrise dip is missing, and sometimes the sunset dip. There is often a reinforcement in the signals before and after a dip; but in many instances such a reinforcement is not indicated.

The records appear to have been made in all cases by shunting the receiving telephone with non-inductive resistance down to the point of inaudibility. The strength of signals is then estimated from the conductance of the limiting shunt.

It appears, therefore, that there is sufficient warrant from the observations at hand in giving the theory here suggested for the sunset and sunrise dips further consideration. It is not claimed

---

(8) The Effects of Sunlight on the Transmission of Wireless Signals, by B. L. Dolbear and J. A. Proctor, Electrical World, N. Y., Vol. 58, No. 6, Aug. 5th, 1911, pages 321-323.

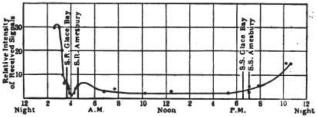

Fig. 6. Observations taken by Mr. Pickard on the relative intensity of signals received at different hours of day and night.

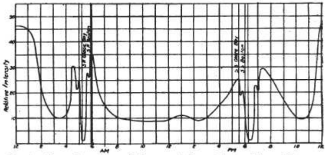

Fig. 7. General average of all curves taken at Somerville and Revere March, 1911.

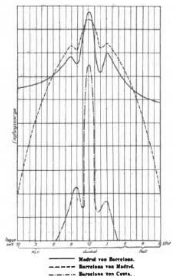

Figure 8

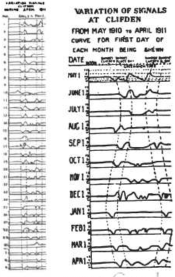

Figure 9

Google

status of long-distance wireless-wave transmission; computed the numerical values of the diffraction series with an approach different from all previous ones; compared his calculations with the values obtained from Macodnald's new formula, found them in agreement, and declared Macdonald's theory the best of the lot; and reinterpreted the long-distance experimental data by correcting the asserted relation between the audibility factor, the results from the shunt-telephone measurements, and the antennacurrent.[58] That improved the fit with his (and hence Macdonald's) predictions.

Love's method of approximation was numerical. Like Nicholson and Macdonald, Love approximated the terms in the neighborhood of $n+1/2 = ka$. Then he computes the numerical values of a sufficient number of terms, added them up, and compared the numerical results with those he obtained from Macdonald's method at several separation angles.[59]

The Austin-Cohen formula made the field intensity proportional to $[1/\theta]\exp[-9.6\lambda^{-1/2}\theta]$, which was quite different from all the diffraction formulae, including Macdonald's.[60] Love saved the phenomena by reinterpreting the measured data. Love noticed that Hogan used a "law of device" to convert the data from the measured audibility factor to the antenna current (he made the audibility factor proportional to the square of the antenna current). However, from Austin's 1911 report, this device law did not always hold. Austin's measurements for calibration for weak signals suggested that the audibility factor was proportional to the antenna current. Considering that most data were taken when signals were weak, Love argued that the proper device law was the direct proportion. He modified Hogan's data accordingly and thus gave empirical support to the diffraction theory.

Were Love's conclusions justified? No! His numerical method was based on assumptions about infinite series identical to Macdonald's; it was not a surprise that the two methods gave similar predictions. The real issue, whether the underlying assumptions employed for the purpose of approximation were legitimate, remained unsettled. Moreover, Love's reinterpretation of the data was prblematic. He relied on regularities measured in 1911. Austin reported the same experimental results as Hogan did in 1913, but he did not discover any significant deviation of the measured data from the Austin-Cohen formula.

6. ATMOSPHERIC REFLECTION AGAIN

Atmospheric reflection produced much less literature after 1911 than surface diffraction theories. The wireless communities found it convenient to interpret radio phenomena in terms of atmospheric effects, but much more difficult to elaborate a reflection model via experimental or theoretical means. Still, atmospheric reflection theorists managed to go one step further than Heaviside and Kennelly. They wished to understand why static noise is greater at night, and why the transmission efficiency of the wireless signals experiences a diurnal change. Having different research agendas and satisfied with qualitative explanations they did not compete directly with the mathematical physicists working on diffraction theories. Rather, they drew on the intellectual tradition of microphysics, which studied electromagnetic wave propagations in various media in order to reveal the internal molecular structures of materials.[61]

---

[58] Augustus E.H. Love, "The transmission of electric waves over the surface of the earth," Royal Society of London, *Philosophical transactions A, 215* (1915), 105-131.

[59] Ibid, 116-123.

[60] Ibid, 127.

William Henry Eccles, who earned a bachelor's degree in physics from the University of London in physics in 1898, joined Marconi's research team on wireless telegraphy in 1899. It was as a professor at the University of London, where he moved in 1910, that he taught the general wireless communities the empirical ground of the atmospheric-reflection hypothesis.[62]

Eccles discovered that if static was detected at one station, then it was very likely to be detected simultaneously at others some distance away. He deduced that the effect involved a long-distance mechanism, perhaps the discharge of atmospheric electricity at hundreds or thousands of miles away from the receiver stations.[63] In a paper published in 1912, he attempted to explain the cause of diurnal variations of the static intensity. Since trans-Atlantic telegraphic signals and static can be transmitted over long distances, they might propagate in a similar manner.[64] A correct physical model for long-distance wave propagation might explain the data on trans-Atlantic wireless telegraphy and on static.

In Eccles's model, the earth is surrounded by a permanent conducting (Heaviside) layer in the upper atmosphere, and another, concentric layer between the Heaviside and the earth. This new layer corresponds to a region of air with gradually changing physical properties. The ultraviolet component of sunlight ionizes the air in this region.[65] Since sunlight alternates with penetration into this region, the number of charged particles per unit volume in it increases with height.

Eccles evoked a simple microphysical model to describe wave propagation in this medium. When an electric field is applied, the ions move; Eccles deduced the average induced ionized current from Newton's second law of motion. Incorporating the induced current into Maxwell's equations, he expressed the refractive index of the medium in terms of the ions' number density, mass, and charge. From this simple Maxwellian theory, Eccles demonstrated that the phase velocity of an electromagnetic wave increases with the number density of ions.

---

[61] Jed Buchwald, *From Maxwell to microphysics: Aspects of electromagnetic theory in the last quarter of the nineteenth century* (Chicago, 1985).

[62] *Obit., 17* (1971), 195-196.

[63] William H. Eccles and H. Morris Airey, "Note on the electrical waves occurring in nature," *PRS, 85* (1911), 145-150.

[64] William H. Eccles, "On the diurnal variations of the electric waves occurring in nature, and on the propagation of electric waves round the bend of the earth," *PRS, 87* (1912), 79-99.

[65] Ibid, 88-89.

---

(5) When both stations are in full sunlight, their signals are subjected to daylight attenuation, by diffused conduction in the upper air.

(6) When the sunset wall gets behind the eastern station, as at 17, 17′, the wall will temporarily serve as a reflector behind the latter, and strengthen the signals.

(7) When the sunset wall intervenes between the stations, the wall will act as an intercepting barrier, reflecting waves back, and markedly, weakening the signals.

(8) When the sunset wall gets behind the western station, there will again be a temporary increase of signals by reflection. After this the conditions should approach those of permanent shadow, or night time.

In one revolution of the earth, therefore, we should expect to find maximum strength on full night shadow, a lowered strength in full daylight, and a marked weakening of the signals, with the wall between the stations, either at mid-sunrise, or at mid-sunset. (6) Each of these dips in the strength of signals should be both preceded and followed by a brief interval of stronger signals, due to partial reflection, in the same way as a rough or imperfect sheet reflector behind a lamp intensifies its rays.

On the other hand stations on the same parallel of longitude should have no dip on signals on the equinoxes; but should have a dip at sunrise and sunset with the sun near the solstice. (December and June.)

Stations north and south might expect a longer signaling range than those east and west owing to the aid of partial reflections along the shadow wall.

We may now compare the foregoing deductions with recorded observations. Figure 6 is taken from observations by Mr. G. W. Pickard in 1909, (7) as published in Figure 97 of Prof. G. W. Pierce's book on Wireless Telegraphy. It will be observed that there is a marked dip in the intensity of signals received at Amesbury, near Boston, Massachusetts, when sunrise was about midway between Amesbury and the sending station at Glace Bay, Nova Scotia. Both of these stations are indicated in Figure 4 by

---

(6) This provisional theory of the sunrise and sunset dips was first put forward by the writer at the Radio-Telegraphic Discussion of the Dundee meeting of the British Association, September, 1912.

(7) "Principles of Wireless Telegraphy," G. W. Pierce, New York, 1910, page 135.

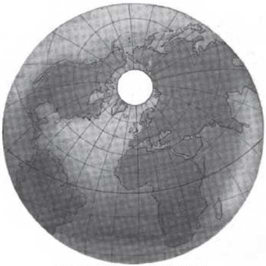

Figure 4. Projection of the Globe

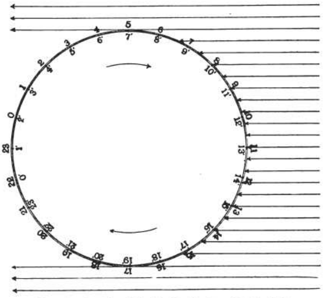

Figure 5. Rotation of the Earth with Regard to Sunlight as Viewed from South Pole.

Digitized by

Eccles explained three distinct wireless phenomena on his model: the possibility of wave transmission along the surface of the earth, the radiation directivity of Marconi's tilted aerial, and the diurnal variation of static. The theory of ionic refraction suggested that the larger the density of ions, the higher the phase velocity. From Eccles's physical model, the ion density increases with height, as, therefore, does the wave's phase velocity. According to Snell's law, an upgoing radio wave would be gradually drawn downward by the refraction in the gradually varying medium. The wave path would curve. If its curvature equalled that of the earth, the wave could run naturally along the surface without any action of the ground material. Unfortunately the quantitative prediction from ionic refraction disagreed with the Austin-Cohen formula.[66]

Eccles' theory suggested that the amount of refraction decreases with frequency. When the frequency falls low enough, the curvature of the refracted wave path becomes larger than the earth's curvature. In this case, the wave path is obliquely incident to the ground and the direction of wave polarization no longer remains vertical. To match the polarization to the maximum extent, the receiving antenna must be tilted toward the direction of transmitter. Thus Eccles reached the same conclusion as Zenneck's without using the "surface wave."

The ionized layer results from sunlight and thus does not exist at night. In England, the major source of long-distance static comes from Africa, where thunderstorms and other electrically disturbing weather processes are more severe. In the daytime, long-distance static is controlled by refraction in the ionized layer in the middle and lower atmosphere. During the night, long-distance static is directed by reflection from the conducting Heaviside layer in the upper atmosphere. The waves refracted through the absorptive ionized atmosphere suffer more energy dissipation than the waves reflected by a conducting surface. Hence static during the daytime is weaker than that at night.

Eccles's atmospheric refraction theory contained an essential difference from Heaviside's. In Heaviside's model, radio waves are guided by the concentric spheres of the earth and the permanent conducting layer in the upper atmosphere in accordance with the conductivities of the air and the ground. In Eccles's model, radio waves are directed by the refractive condition of the ionized atmosphere and do not respond to the condition of the ground. For Eccles, ionic refraction alone explains the bending of the wave propagating direction.[67]

Eccle's work of 1912 brought the atmospheric reflection theory from a tentative hypothesis to a sophisticated model. Heaviside and Kennelly attended to only one fact. In contrast, Eccles covered several apparently unrelated phenomena—long-distance wave transmission, tilted polarization, and diurnal variation of static. Eccles's model was essentially a theory of the qualitative behavior of wireless phenomena. Only partial and preliminary results were achieved by efforts to describe the wave-propagation characteristics mathematically, and these were not consistent with the empirical regularity (the Austin-Cohen formula).

---

[66] Ibid, 91.

[67] Eccles to Heaviside, 27 Nov 1912, Heaviside Papers, UK0108 SC MSS 005/I/6/10, Institute of Electrical Engineers Archives, London.

The English physicist John Ambrose Fleming followed up Eccles's work by trying to account for the diurnal variations of wireless signals. But the static patterns seemed too complex for a consistent explanation. In the annual meeting of the British Association for the Advancement of Science (BA) in 1912, Fleming had organized a discussion on "The scientific theory and outstanding problems of wireless telegraphy" and suggested that the BA form a committee to guide and formulate research on them. The committee was formed and planned systematic observations on atmospheric stray rays. It did not achieve significant results.[68]

Kennelly attended Fleming's radio-telegraphic session at the BA meeting in 1912. Kennelly agreed that the boundary between the sunshine and the shadow regions formed a reflecting surface critical to the diurnal variation observed in wireless telegraphy. He proposed a model for the variation of wireless signals around twilight. Two wireless stations sit in an east-west direction and the transitional band moves gradually toward the east station. Before the band crosses when both stations are on the same side of the blocking curtain formed at the transitional band, it does not affect signal transmission between them. When the band moves behind the east station (shortly before the east station's sunrise), the blocking curtain functions as a reflecting surface to bounce the overshoot waves from the west station back to the east station. Hence the received signal strength there is intensified. After sunrise at the east station, the band moves between the stations, wave transmission is blocked, and the received signal strength diminishes significantly. When the sunrise just passes the west station, the band functions again as a reflecting surface behind the west station to enhance the received signal strength there. As the band moves farther away, the reflective enhancement also wanes. The same pattern happens around sunset. Kennelly's mechanism predicts that a wireless signal transmitted along an east-west direction has a maximum just before the twilight of the east station and right after the twilight of the west station, and a minimum in between. The pattern of the signal variation predicted from Kennelly's model agreed with observations made at wireless stations at Nova Scotia and Amesbury, Massachusetts.[69]

These investigations typified the work of the atmospheric-reflection theorists. They paid much more attention to the diurnal variation of wireless energy than to mathematical relation between the signal intensity and the distance, and they emphasized qualitative characteristics rather than quantitative information in the experimental data. Their theoretical work involved sophisticated model-building and simple mathematical theories of wave propagation based on microphysics.

Toward 1920, several young experimental physicists well trained in applied mathematics began to appreciate the physical importance of the ionic-atmosphere

---

[68] Fleming (ref. 5), 860.

[69] Arthur E. Kennelly, "The daylight effect in radio telegraphy," Institute of Radio Engineers, *Proceedings, 1:3* (1913), 12.

---

toward G in the diagram. All that can be asserted definitely is that there is a greater probability of a transition layer, or partially reflecting layer, being formed between sunshine and shadow at sunrise and sunset, than at any place in the sunlit region of daylight. The ionised layer boundary cannot coincide with the shadow boundary AB, but will lag behind it, and will rise more nearly vertically, owing to the effects of atmospheric absorption, and of building up with time. We may call this hypothetical transition layer, or boundary between daylight and darkness, the "shadow wall" accompanying sunrise or sunset.

In Figure 4 we have a stereographic projection of part of the northern hemisphere. The lines of longitude at hourly intervals may be considered as representing the positions occupied by either the sunrise or sunset shadow walls, at successive hourly intervals at the equinox, when the sun appears on the equator. The shadow, considered as an imperfect electromagnetic mirror, may only extend upwards say from the 30 km. level to the 130 km. level; or may only occupy a height of say 100 km. in all; but, if it exists, it extends northwards and southwards for thousands of kilometers. In summer, the wall would slant from S. E. to N. W., and in winter from S. W. to N. E., across the globe.

If the boundary surface between day and night; or the shadow wall, possesses roughly reflecting or scattering influence on electromagnetic waves, we should expect to find the following series of phenomena in relation to two stations east and west, on or near the same parallel of latitude, as indicated in Figure 5; where the observer is supposed to be at the earth's south pole looking at the rotating pole beneath him.

(1) When both stations are in full shadow, as at 20, 20′; 21, 21′; 22, 22′; 23, 23′; 0, 0′; 1, 1′; 2, 2′; 3, 3′; 4, 4′; the signals exchanged should be normal, in the absence of thunderstorms or meteorological disturbances.

(2) When, shortly before dawn, at the eastern station, the shadow wall gets behind that station, it should act as a partial reflector to that station, and intensify the signals.

(3) When the shadow wall advances to a point between the stations, as at 6, 6′, the wall should act as a partial barrier between them and weaken the signals.

(4) When the wall reaches a point a little beyond the western station, as at 7, 7′, it should act as a temporary reflector to the latter, and temporarily strengthen the signals.

vation bb′, 50 km. above the sea level GG′, the air pressure and density are 0.05 per cent of those on the sea, while 99.95% of the air lies below this level. Again, at an elevation of approximately 60 km., the density has fallen to 0.01%, and 99.99% of the total atmosphere lies beneath. In each 10 km., the air pressure is falling to one-fifth. A vertical sunbeam, reaching the level BB′, has passed thru 0.01% of the total atmosphere. The successive levels AA′, BB′, are lines of equal penetration, and there should be equal intensities of ionisation over each such level; but there should not be any layer of sudden transition.

Figure 2 represents the same atmosphere conditions as in Figure 1; except that the sun's rays are supposed to be entering the atmosphere at an inclination of 60° with the zenith, as indicated by the arrows on either side. Along these inclined paths, the sunbeam will encounter approximately twice as much air between any two given elevations as in Figure 1. Consequently, at the 50 km. elevation, the sunbeams have traversed 0.1% of a vertical atmosphere, instead of only 0.05%. The levels AA′, BB′, CC′ of 0.001%, 0.01% and 0.1% respectively are all raised about 5 km. with respect to those on Figure 1. Otherwise, there is very little change between the conditions of penetration by rays from an overhead sun at the equator, and those from the sun at either 8 A. M. or 4 P. M. There should be no sudden transition layer or surface of discontinuity in ionisation, in either case.

When, however, the sun's rays are striking tangentially over a place on the globe as in Figure 3, there tends to be a transition along the line AB, between air in the shadow and air in the sunshine. If the condition represented is that of sunrise, then the air still in shadow is presumably air that has become neutralized during night, with a relatively low conductivity. The illuminated air on the other hand is rapidly becoming ionised and more conductive. While, therefore, we cannot expect the moving shadow plane to be a sharply defined surface separating ionised from neutral air, we might reasonably expect a roughly defined bounding surface, such as might produce some diffuse reflection of electromagnetic waves. No attempt is made to indicate the lines of equal penetration or ionisation, owing to the complexity of the actions. It is known that there is a very appreciable refraction of the beams of light. We should also expect absorption to take place at different elevations, and inoisation to increase to some extent with time. The shadow boundary AB steadily advances

---

model. They tried to bring together the work of all three communities. One of them was Balthasar van der Pol, who worked at the Cavendish Laboratory between 1917 and 1919. In 1918, he tackled the discrepancies among different diffraction formulas (Nicholson, Rybczinski, Macdonald). He reasoned that if he could show that one of them was mathematically rigorous, he could demonstrate that diffraction could not account for the Austin-Cohen formula. The mathematics required was too difficult to him. He turned to a Cambridge-trained mathematician, George Neville Watson, for help.

7. WATSON'S WORK

Like Macdonald, Nicholson, and Love, Watson excelled at the Cambridge training in mathematics, becoming Senior Wrangler, Smith's Prize man, and a Fellow of Trinity College. In 1918, he took up the professorship of mathematics at the University of Birmingham. Unlike the diffraction theorists, Watson was almost detached from physics at Cambridge. He specialized in complex-variable theory applied to Bessel functions. He was also interested in theories of approximation, numbers and computability.[70] He was just the man to solve van der Pol's problem.

Papers of 1918 and 1919

In 1918, Watson gave a rigorous mathematical proof that the field intensity diffracted along the curvature of a large conducting sphere has an exponential factor whose decay rate is proportional to $\lambda^{-1/3}$ rather than $\lambda^{-1/2}$ (as required by the Austin-Cohen formula). In 1919, he showed that the field intensity diffracted in a space bounded by a large conducting sphere and a conducting surface exterior and concentric to the sphere has the $\lambda^{-1/2}$ dependence.[71] The atmospheric reflection theory can explain the empirical regularity.

Watson found the rigorous solution of the diffraction problem without the mathematical problems of previous diffraction theories by working on the Hertzian potential instead of the magnetic field intensity. True to the Cambridge approach, he expanded the Hertzian potential in terms of a discrete sum of spherical harmonics instead of an integral. The result was an infinite series different from Macdonald's, but with a similar angular dependence and pole structure.

Watson's great innovation was to convert the series expansion of the Hertzian potential into an integral expansion without infinities. He interpreted all terms in the diffraction series as residues of a complex function associated with poles on the real axis and so expressed the diffraction series as a contour integral in the complex plane. This procedure, later known as "Watson's transformation," converted the Hertzian potential from a series to a complex integral. This integral as evaluated by Cauchy's residue theorem contained an exponential decay $\exp[-23.94\lambda^{-1/3}\theta]$, quite close to Nicholson's $\exp[-23.8\lambda^{-1/3}\theta]$.[72] Watson confirmed that the intensity of the field diffracted along the earth's surface was significantly weaker than the Austin-Cohen formula required. The exponential decay of the one contains $\lambda^{-1/3}$, of the other to $\lambda^{1/2}$. The diffraction theory was mathematically consistent, but empirically inadequate.

---

[70] *Obit.*, *12* (1966), 521-522.

[71] George N. Watson, "The diffraction of electric waves by the earth," Royal Society of London, Proceedings, 95 (1918-1919), 83-99; and "The transmission of electric waves round the earth," ibid., 546-563.

Watson located the problem in the physical model. All the diffraction theorists assumed that the earth's surface alone diffracts the field radiated by the dipole oscillator. Since the surface diffraction alone cannot account for the empirical observations, the upper reflective regions of the atmosphere might play the dominant role at long distances. Could diffraction theory incorporate the physical assumption of an atmospheric reflective layer? Watson took up the question in his paper of 1919. He took the earth to be a conducting sphere on which the Hertzian dipole sits, and the atmospheric reflective layer to be concentric with the earth.[73]

To evaluate the Hertzian potential in this new boundary condition, Watson expanded the Hertzian potential in a series of spherical harmonics, applied "Watson's transformation" to convert the series into a complex integral, analyzed the pole structure of the integrand, and evaluated the integral in terms of these poles. He discovered that when both the inner and outer conductors are perfect, the field intensity is a superposition of oscillatory modes periodic with the distance, that is, that the field does not decay at all! When both the inner and outer conductors are good but imperfect, the field has an exponential decay proportional to $\lambda^{-1/2}$. By adjusting the conductivity of the ionized layer, Watson could match the numerical value of the decay rate of his theory and the Austin-Cohen formula.

Reception

Watson established a conjunction of mathematical representations, explanatory models, and experiments in the study of long-distance wave transmission. He applied the complex-variable techniques of the diffraction theories to the physical model entertained by the atmospheric-reflection theorists to derive quantitative results consistent with the empirical formula obtained from long-distance experiments. By connecting the mathematical representation with the formulated experimental data, Watson gave the atmospheric-reflection theory the promise of becoming a question-answering device operated and using standard mathematical techniques. When it left his hands, Watson's theory lacked verisimilitude. The physical model of a homogeneous and sharp conducting boundary was an over idealization. Watson did not incorporate a vertical atmospheric profile with gradually varying refractive indices. His theory could not incorporate Eccles's ion-refraction theory in a profound way. Nor was Watson's theory able to marshal em-

---

[72] Watson, "diffraction" (ref. 71), 97.

[73] Watson, "transmission" (ref. 71), 547.

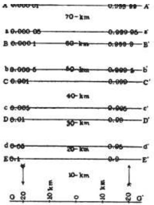

Figure 1. Air-Density and Penetration at Different Atmospheric Levels. Sun's Rays Vertical

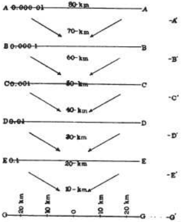

Figure 2. Penetration of Inclined Solar Rays at Different Atmospheric Levels

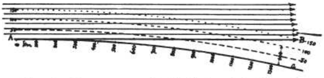

Figure 3. Diagram representing the Surface of the Globe with Three Atmospheric Levels 50, 100 and 150 Km. above the sea. Rays of Sunlight following Parallel Straight Lines are Tangential to the Globe at A and are Penetrating the Atmosphere above the Line AB which Marks the Bounding Surface between Light and Shadow. The Sun's Parallax and Atmospheric Refraction are

Digitized by Google

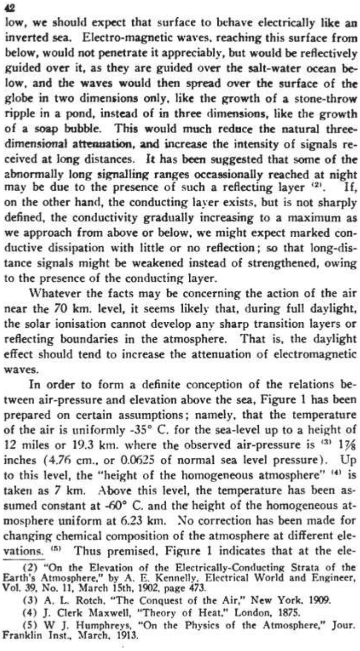

low, we should expect that surface to behave electrically like an inverted sea. Electro-magnetic waves, reaching this surface from below, would not penetrate it appreciably, but would be reflectively guided over it, as they are guided over the salt-water ocean below, and the waves would then spread over the surface of the globe in two dimensions only, like the growth of a stone-throw ripple in a pond, instead of in three dimensions, like the growth of a soap bubble. This would much reduce the natural three-dimensional attenuation, and increase the intensity of signals received at long distances. It has been suggested that some of the abnormally long signalling ranges occassionally reached at night may be due to the presence of such a reflecting layer (2). If, on the other hand, the conducting layer exists, but is not sharply defined, the conductivity gradually increasing to a maximum as we approach from above or below, we might expect marked conductive dissipation with little or no reflection; so that long-distance signals might be weakened instead of strengthened, owing to the presence of the conducting layer.

Whatever the facts may be concerning the action of the air near the 70 km. level, it seems likely that, during full daylight, the solar ionisation cannot develop any sharp transition layers or reflecting boundaries in the atmosphere. That is, the daylight effect should tend to increase the attenuation of electromagnetic waves.

In order to form a definite conception of the relations between air-pressure and elevation above the sea, Figure 1 has been prepared on certain assumptions; namely, that the temperature of the air is uniformly -35° C. for the sea-level up to a height of 12 miles or 19.3 km. where the observed air-pressure is (3) 1⅞ inches (4.76 cm., or 0.0625 of normal sea level pressure). Up to this level, the "height of the homogeneous atmosphere" (4) is taken as 7 km. Above this level, the temperature has been assumed constant at -60° C. and the height of the homogeneous atmosphere uniform at 6.23 km. No correction has been made for changing chemical composition of the atmosphere at different elevations. (5) Thus premised, Figure 1 indicates that at the ele-

---

(2) "On the Elevation of the Electrically-Conducting Strata of the Earth's Atmosphere," by A. E. Kennelly, Electrical World and Engineer, Vol. 39, No. 11, March 15th, 1902, page 473.

(3) A. L. Rotch, "The Conquest of the Air," New York, 1909.

(4) J. Clerk Maxwell, "Theory of Heat," London, 1875.

(5) W J. Humphreys, "On the Physics of the Atmosphere," Jour. Franklin Inst., March, 1913.

pirical evidence beyond the Austin-Cohen formula, for example, the diurnal variation of static.

Appreciation of Watson's contributions came slowly. The first few published papers to mention his work all emphasized the mathematical rather than the empirical implication. Van der Pol stressed Watson's mathematical contribution to clarifying the controversy involving different approximation methods for wave diffraction above a spherical conductor, rather than Watson's theoretical prediction of the Austin-Cohen formula.[74] Macdonald commented on Watson's work without addressing empirical adequacy at all. He concentrated on extending Watson's approach to the general case where single-frequency time-dependence does not hold.[75] Sommerfeld's student Otto Laporte used Watson's results to reconcile the British representation of the diffracted field in terms of a series with the German representation in terms of an integral, not for their own sake.[76]

Most scientists who elaborated the atmospheric-reflection model in the 1920s, studied wave propagation through ionic media. They downplayed Watson's old-fashioned model (the sharply defined "Kennelly-Heaviside layer") and counter-intuitive approach (the complex-variable theory of diffraction). In a classic paper on the subject written in 1924, Joseph Larmor mentioned Watson only at the end of a list of diffraction theorists beginning with Macdonald, and then as the man who demonstrated that surface diffraction cannot account for long-distance wave transmission around the earth.[77] Among these initial responses to Watson's papers, one thing stands out: after Watson had proved the mathematical rigor of Nicholson's predictions, no one attempted to reconcile the experimental data with the surface diffraction theory. Watson's paper in 1918 tilled the diffraction theory as a physical model for long-distance wave transmission.

Watson's fortunes improved in the late 1920s. In 1928, G.W. Kenrick of the University of Pennsylvania reinvestigated the mathematical and empirical significance of Watson's theory. Kenrick pointed out that much work had recently been done to explain short wave transmission by reflection and refraction of electric waves, but "less attention has been given to modifications produced in the classical Hertzian solution for the field at a distant point due to an oscillating doublet."[78] He called for a reexamination of Watson's work in the interest of short-wave analysis. Kenrick calculated the electromagnetic field radiated by a Hertzian dipole using the same boundary condition that Watson had. Instead of working on the residue waves, he computed the multiple reflected rays bouncing through the space between the earth and the atmospheric layer, and summed all the reflective terms. He thus reproduced Watson's mathematical formula and hence the Austin-Cohen formula. Watson's work became significant not only for disproving the diffraction model but also for its quantitative predictions of observed phenomena.

---

[74] Balthasar van der Pol, "On the propagation of electromagnetic waves round the earth," *Philosophical magazine, 38* (1919), 365-380.

[75] Hector M. Macdonald, "The transmission of electric waves around the earth's surface," *PRS, 98* (1920), 216-222, 409-410; 108 (1925), 52-76. "On the determination of the directions of the forces in wireless waves at the earth's surface," *PRS, 107* (1925), 587-601.

[76] Otto Laporte, "Zur Theorie der Ausbreitung elektromagnetischer Wellen auf der Erdkugel," *AP, 70* (1923), 595-616.

[77] Joseph Larmor, "Why wireless electric rays can bend round the earth," *Philosophical magazine, 48* (1924), 1026-1036.

[78] Gleason Willis Kenrick, "Radio transmission formulae," *Physical review, 31* (1928), 1040-1050, on 1040.

The radio scientists and engineers of the 1930s worked intensely on wave propagation under various scenarios specified by atmospheric and terrestrial conditions. A theory of pure refraction without taking into account the effect of the earth was no longer adequate. Their theoretical exercise dealt with wave diffraction above a ground with different possible geometric and material specifications in a heterogeneous atmosphere. "Watson's transformation" proved a very useful technique for analyzing these problems. In 1937, a Harvard professor of physics, Harry R. Mimno reviewed the literature on the physics of the ionosphere.[79] He highlighted the significance of the discrepancy between the $\lambda^{-1/3}$ dependence predicted by all the diffraction theorists before Watson and the $\lambda^{-1/2}$ dependence given by the Austin-Cohen formula, and he stressed Watson's contribution in providing a theoretical account of the empirical $\lambda^{-1/2}$ dependence. Van der Pol's student H. Bremmer wrote that "the pioneering work clearing the way for further investigations was done by Watson in 1918. By a transformation with the aid of an integral in the complex plane, this author succeeded in transforming the rigorous series of zonal harmonics into a new series converging rapidly enough to be of use in the radio problem. As a matter of fact, almost all of the later literature is based upon this transformation of Watson."[80]

Often in the history of science two mutually exclusive theories compete for the answers to the same set of questions. In the case considered here, two mutually exclusive theories address different types of questions. The essential difference between the surface diffraction theories and the atmospheric reflection theories was their distinct epistemic status rather than their different physical models. The surface diffraction theorists began by investigating the possibility of long-distance radio-wave transmission. To answer this question, they constructed a straightforward physical model and attempted to develop a rigorous mathematical solution of the problem described by the physical model. But they soon discovered a mathematical difficulty in obtaining an accurate approximate solution. They switched to a mathematical question—what is an accurate approximation of the diffracted field intensity above a large conducting sphere? Almost all their effort involved proper approximations of the analytic form of the diffracted field.

The atmospheric-reflection theorists asked for causal explanations for a broader realm of wireless phenomena, including long-distance transmission, diurnal variation of signal strength, static, the effect of weather fluctuation, and the directive

---

79. Harry R. Mimno, "The physics of the ionosphere," Review of modern *physics, 9:1* (1937), 1-43.

80. H. Bremmer, *Terrestrial radio waves: Theory of propagation* (New York, 1949), 7.

Commencing, say, at zero, with sufficient height; it might increase to a maximum at a moderate height, and then dwindle down to a minimum near the earth's surface. If the sunlight ionisation, instead of varying gradually in this way terminated suddenly, so that, at some particular elevation a bounding surface formed with non-conducting air on one side, and conducting air on the other; then this boundary surface might be expected to develop strong reflecting properties, on the principle that wave disturbances are subject to reflection at surfaces of discontinuity. Thus clouds, or diffused masses of water vapor, reflect both sound and light. A travelling compression wave, or sound wave in air, reflects light. Any change in a medium for wave transmission, occurring at a surface, is known to set up a reflection. If the change is sudden and well-marked, so that the bounding surface is sharp, the reflection will be definite and powerful. If the change is gradual, and by easy transition; so that the bounding surface is not clearly defined, the reflection will be diffusely scattered and weak.

Consequently, if the ionisation of the air developed a sharp transition layer, or succession of layers, we might expect reflection to occur at and from such layers, with a lessening of attenuation; whereas, if the ionisation were gradually varying from layer to layer, with no clearly marked transition, there would be mere dissipation of energy by conduction or scattering without any gain by reflection, thereby increasing the attenuation.

It was pointed out by Dr. J. J. Thomson (1) that rarefied air at a pressure of 0.01 mm. of mercury in a glass chamber devoid of metallic electrodes, conducts electricity in the laboratory as well as an aqueous solution of sulfuric acid. At an elevation of about 70 km. (43.5 miles) above the sea level, and a uniform temperature of -60° C., such an air density may be expected to exist. If this free rarefied air conducts electricity in response to the feeble electric intensities of radio telegraphy, as well as it does in vacuum tubes to the more powerful intensities used in the laboratory; then, whether the sun is shining on this or not, we should expect a conductivity in it of the same order of magnitude as in ocean water. If such a conducting layer developed suddenly at a certain elevation, so that a definite bounding surface separated the conducting air above from non-conducting air be-

---

(1) J. J. Thomson, "Recent Researches in Electricity and Magnetism," 1893, page 101.

beyond, the violet, when the sun's rays fall perpendicularly as at the tropical noon-day. At morning or evening, when the sun's rays pass aslant thru much greater distances of air before reaching the ground, the violet, and even the blue rays largely disappear, leaving a predominance of red in the light that remains, and thus producing the ruddy hues of dawn or evening landscapes.

If the upper regions of the atmosphere are appreciably ionised by full and sustained solar radiation during daylight hours we may consider that these regions are thereby rendered partially conducting. That is to say, instead of being a perfect insulator, like free space or un-ionised air, ionised air has a certain small conductivity. This would involve a loss of energy in any electromagnetic waves traversing it, which, in turn, would involve additional attenuation of such waves. Moreover, if the ionisation-conductivity were not uniform but developed in clouds or patches, there would be scattering as well as absorption of energy.

It seems therefore possible to explain the weakening effect of broad sunlight upon radio-transmission signals by attributing conductivity, distributed either uniformly or non-uniformly thru the sunlit upper atmosphere, where the ultra-violet waves are likely to be more intense than in the region near the ground. We have no direct evidence, however, as to whether such ionisation-conductivity is quantitatively sufficient to account for the observed effects. It has been pointed out by Zenneck that the observed conductivity of air near the earth's surface for continuous current is far too small to account for the effects in question; but we have no experimental evidence as to what the conductivity may be at high atmospheric levels to alternating electric intensities.

If we assume, for simplicity, a tropical sun sending its rays perpendicularly down thru normally distributed air towards the earth, the degree of ionisation should be uniform over any surface situated at a uniform level. That is, the ultra-violet radiation would be most intense at a great height, and gradually weaken by absorption as it penetrated downwards. On the other hand, the number of air molecules per c.c.; i. e., the air density, would be relatively very small at a great height, and would increase exponentially with the downward penetration. In any one horizontal layer of air, the number of free ions might be assumed uniform.

antenna pattern. To answer these questions, they constructed elaborate physical models. They were partially successful in offering reasonable causal explanations to observed wireless phenomena. But they failed to develop a mathematical theory for systematic quantitative predictions before 1910, succeeded to a very limited extent after Eccles, and only began to make steady progress after Watson. The reflection theorists had difficulty in formulating answers to numerical-prediction questions. Thus the diffraction theorists and the reflection theorists carried out different agendas: the former tried to resolve a mathematical problem of approximation, the latter aimed to explain newly discovered phenomena. This fact explains why atmospheric-reflection theories rarely engaged in any public debate with the surface-diffraction theorists, why some mathematical physicists kept working on diffraction theories regardless of their dubious empirical adequacy, and why an experimenter strongly supporting the reflection models still gave credit to the diffraction models. The diffraction theories and the reflection theories were not competing worldviews.

From the epistemic viewpoint, long-distance wave-transmission research left unexpected legacies to all the participating communities. They started from what they wanted to know, and found what they did not expect to learn. The diffraction theorists found that approximation can be a critical issue in physical problems in which the analytic solution cannot give meaningful quantitative information and direct numerical computation is intractable. The new condition that the wavelength was much shorter than the scatterer's dimension forced them to develop a repertoire of advanced mathematical techniques other than those developed for acoustic scattering for dealing with the approximations of series or integrals. Sommerfeld's integral and Watson's transformation initiated studies of mathematical questions that would become classical problems in mathematical physics: a vertical or horizontal dipole oscillator above or below a homogeneous or layered horizontal plane, a vertical or horizontal dipole oscillator above or below a homogeneous or heterogeneous sphere within a homogeneous or a concentrically layered medium, and so on. These problems did not necessarily correspond to real physical situations. But the mathematicians contentedly investigated the complex-variable techniques to solve the problems for their own sake. The mathematical theory of complex series and integrals prospered from 1930s to 1950s largely owing to the heritage of the German and the British diffraction theorists in the 1900s and 1910s.[81]

The atmospheric-reflection theorists started from the puzzles associated with radio operations, but ended up with a new science of the atmosphere. Their models of the ionosphere were connected with contemporary electron theories of matter. Thus the study of atmospheric effects on radio became incorporated into one of the largest intellectual movements of physics in the early 20th century: the rise of atomic and molecular physics. The radio scientists benefitted from this incorporation: they developed various Maxwellian theories of electromagnetic wave propagation in ionized media. Moreover, the reflection theorists discovered that the radio could be an experimental means to explore atmospheric phenomena. Edward Appleton's experiments of the 1920s, which won him a Nobel Prize, gained "direct" evidence about the physical condition of the ionosphere from radio-interference methods. Meteorology and planetary science became accessible not only through the traditional means of natural history, but also through experiments made possible by efforts of the reflection theorists.[82]

---

[81] Alfredo Baños, *Dipole radiation in the presence of a conducting half-space* (Oxford, 1966).

The original goal of the American wireless experimenters was to test equipment for the first long-distance radio station of the U.S. Navy. They had a practical engineering problem to solve, but in pursuing it made an important contribution to pure science. The Navy's questionable decision to settle with NESCO's transmitter, the missionary agenda of the Bureau of Standards, and Austin's technical training in German experimental science incorporated meticulous instrumental design systematic operational procedure, and mathematical representation of the data into the investigation. The outcome of their experiments was an empirical law that served as the only quantitative check on theories of long-distance radio-wave transmission in the 1910s. The U.S. Navy did not gain directly from scientific studies of wave transmission, but from the experimental results: it learned that the arc transmitter performed better than the spark-gap transmitter at long distances. The Austin-Cohen formula provided a primary guide for wireless engineers to design long-distance long-wave wireless stations throughout the 1910s. Engineers relied on this simple mathematical relation to select the wavelength and antenna height of a transmitting station for a signal-strength requirement at a given distance. The other face of the empirical law was an engineering formula.

These drastically diverse post-developments enforce the lesson that our three technical communities, working against with different intellectual backgrounds and toward different goals, created a unified study of long-distance radio-wave transmission. They did so by contributing different elements: one offered empirical evidence, another a physical model, and a third a mathematical tool. This episode might be interpreted as confirmation of Peter Galison's estimates of the strength of the constraints on a scientist or engineer within his own communal tradition or of Andrew Pickering's emphasis on the importance of the contingencies a scientist or engineer might confront.[83] More significantly, it warns us not to assume that the intellectual work of different technical communities apparently concerned with the same general phenomena were alternative answers to the same questions. They might have had their eyes on different, and different kinds of questions.

---

[82] Peter Galison and Alexi Assmus, "Artificial clouds, real particles," in David Gooding, Trevor Pinch, and Simon Schaffer, eds., *The use of experiment: Studies in the natural sciences* (Cambridge, 1989), 225-274.

[83] Peter Galison, "Context and constraints," and Andrew Pickering,"Beyond constraint: The temporality of practice and the historicity of knowledge," in Jed Buchwald, ed., *Scientific practice: Theories and stories of doing physics* (Chicago, 1995), 13-55.

Proceedings of the Institute of Radio Engineers: 1913

THE DAYLIGHT EFFECT IN RADIO TELEGRAPHY.

By A. E. Kennelly.

(Professor of Electrical Engineering, Harvard University.)

It is now generally admitted that the range of radio-transmission of signals is materially influenced by solar radiation; not only in regard to false signals or "X's"; but also in regard to the attenuation of the transmitted electro-magnetic waves.

This attenuating influence of solar radiation on the transmitted waves ordinarily consists of (1) a nearly steady action during the daytime, together with (2) certain marked disturbances occurring near sunrise, or sunset, or both.

In regard to the first or steady effect, we may consider that during the day, whatever the weather may be; i. e., the conditions of wind, temperature, pressure, cloudiness or precipitation in the first few kilometers of air nearest to the ground surface, the sun's rays are steadily falling upon the upper layers of the air; where the air density is relatively very low. It is known from physical laboratory experiments, that ultra-violet light, passing thru attenuated air, ionises it; or decomposes electrically neutral air molecules into positive and negative constituents, the energy of decomposition being absorbed from the radiation. If the ultra-violet radiation is then withdrawn, these constitutents attract each other and recombine, perhaps converting the energy of recombinaion into heat energy or molecular oscillations. For a given intensity of received radiation of assigned wave-length in the ultra-violet region of the spectrum, we may suppose that there exists, in the final state, a certain corresponding number of free electrons per unit of air-volume. It is also reasonable to consider that after the ultra-violet rays in the sunlight have penetrated deeply into the air, they become: (1) scattered and diffusely reflected by the air-molecules, thereby giving us the blue color of the clear sky, and (2) absorbed in decomposition and ionisation of the air-molecules. Consequently, but little ultra-violet light from the sun reaches the ground, after passing thru the atmosphere. The solar spectrum at the ground, or ocean level, may be considered as terminating near to and only a little

CHEN-PANG YEANG

The study of long-distance radio-wave propagation, 1900-1919

ABSTRACT:

At the beginning of the 20th century, scientists and engineers were puzzled by the fact that the long wireless waves could propagate along the earth's curvature without being blocked by the earth. Two explanatory theories were suggested: that the waves are diffracted along the earth's surface and that the waves are reflected back and forth between the earth and a conducting atmosphere. The surface diffraction theory, first proposed by Hector Munro Macdonald in 1901, was continuously elaborated by the British and German mathematical physicists. But its predictions were not consistent with the empirical Austin-Cohen formula obtained from the U.S. Navy's long-distance experiments. The atmospheric reflection theory, first proposed by Arthur Kennelly and Oliver Heaviside in 1901/2, was more commonly believed to be the correct physical model. Yet it had problems yielding quantitative predictions because of its lack of mathematical development. In 1919, the English mathematician, George Neville Watson, developed a mathematical theory of atmospheric reflection that generated predictions consistent with the Austin-Cohen formula based on the analytic techniques established by the surface diffraction theorists.

the one that occurs most commonly—the phenomena vary greatly from day to day—is that one appearing when the sun is setting at a place about half-way between the stations. The minimum is fairly well marked in the sunrise curve shown in fig. 8. The reflections that are so pronounced a feature of Marconi's long-distance observations are not nearly so evident over short distances, according to the author's experience. But fig. 9 shows a case where the variations soon after sunset at Clifden were very decided. The curve does not do full justice to the phenomena, however. As a matter of fact the chief variations were so rapid and so wide that there was not time to measure them; indeed, on occasion, the changes in intensity are startling in their amplitude and swiftness.

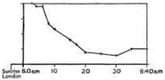

Fig. 8.—Intensity of Signals, January 25, 1912.

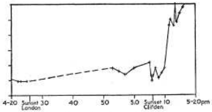

Fig. 9.—Intensity of Signals, January 20, 1912.

It will be noticed that all the curves exhibit a great difference between the strength of the night signals and that of the day signals, although the distance is only 440 miles. Yet it is known that the same signals are heard at Glace Bay as strongly in the day as in the night. We gather from this that the daytime trajectory of the radiation passes well above places relatively near to the sending station and descends again, after overtaking the curvature of the earth, at the greater distance. It is well to recall here that in his recent experiments on the reception of signals from Clifden at distances up to 6000 miles, Marconi found the signals readable only at night at greater distances than 4000 miles—which seems to indicate that the trajectories of the rays in the daytime are such as to bring down within the distance named practically all the radiation starting, at all elevations, from the antenna.

The author desires to tender his thanks to the Government Grant Committee of the Royal Society for a grant in aid of the observations described in this communication.

through the atmosphere can give rise to somewhat regular reflections, so that when it passes over and behind the sending station it changes from being a hindrance to being a help in signalling. This view contains nothing that is fundamentally inadmissible. But the reflecting process was observed to be better with short waves than with long; perhaps the following considerations may in some degree account for this. First, it is known from general electromagnetic principles that when a wave crosses layers of changing refractive index there is a reflected wave propagated backwards, and this reflected wave is the more intense the greater the change in index. Second, let us assume that the surfaces of equal ionisation rise from the day level to the night level in a long slope extending over, perhaps, a hundred miles from east to west through the twilight belt, being rather broken of course, by irregularities in the changing ionisation. On account of the broken character of the belt reflection from the sloping surfaces will be irregular. But it will be more irregular for the more refrangible radiation, that is to say, will be more irregular for waves of low frequency than for waves of higher frequency. Perhaps with this may be conjoined the fact that the frictional absorption suffered by the longer wave is greater than that suffered by the shorter.

The phenomena just discussed are to some extent noticeable over relatively short distances. The following curves are drawn from observations of the intensity of signals from Clifden as heard at the author's laboratory in London, the measurements being made by balancing the intensity of the Clifden signals against the adjustable intensity of locally produced artificial signals of about the same acoustic frequency. On the curves the intensities are plotted in arbitrary units as ordinates, with the times of measurement as abscissæ. The observation had to be snatched, so to speak, at the moments when the station happened to despatch a message, and the points of observation are therefore often rather irregularly distributed. Fig. 7 shows two remarkable minima, which are produced, presumably, by the presence of the ionic curtain between the stations. Of these two minima,

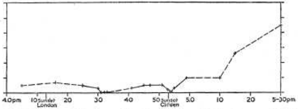

Fig. 7.—Intensity of Signals reaching London from Clifden, January 12, 1912.

ELECTRICAL WORLD AND ENGINEER.
MARCH 15, 1902. 473

On the Elevation of the Electrically-Conducting Strata of the Earth's Atmosphere.

By A. E. Kennelly.

ACCORDING to the measurements of Professor J. J. Thomson ("Recent Researches in Electricity and Magnetism," p. 101), air at a pressure of 1-100 mm. of mercury has a conductivity for alternating currents approximately equal to that of a 25 per cent aqueous solution of sulphuric acid. The latter is known to be roughly 1 mho-per-centimeter, so that a centimeter cube would have a resistance of about one ohm. Consequently, air at ordinary temperatures, and at a rarefaction 76,000 times greater than that at sea level, has a conductivity some 20 times greater than that of ocean water, although about 600,000 times less than that of copper.

If we apply the ordinary formula for finding the elevation corresponding to a given air-rarefaction, we find that if the air had a uniform temperature of 0 deg. C., the height of this stratum of air with a rarefaction of 76,000, would be

18.39 log 76,000 kilometers above the sea.
or 89.77 kilometers,
or 55.77 miles.

If the air had a uniform temperature of —50 degs. C. this elevation would be reduced 18.3 per cent, or to 73.3 kilometers (45.5 miles). The temperature of the earth's atmosphere has only been measured within a range of a very few kilometers above the surface of the sea, and consequently the materials are not at hand for any precise calculation of the height of electrically conducting strata. It may be safe to infer, however, that at an elevation of about 80 kilometers, or 50 miles, a rarefaction exists which, at ordinary temperatures, accompanies a conductivity to low-frequency alternating currents about 20 times as great as that of ocean water.

There is well-known evidence that the waves of wireless telegraphy, propagated through the ether and atmosphere over the surface of the ocean, are reflected by that electrically-conducting surface. On waves that are transmitted but a few miles the upper conducting strata of the atmosphere may have but little influence. On waves that are transmitted, however, to distances that are large by comparison with 50 miles, it seems likely that the waves may also find an upper reflecting surface in the conducting rarefied strata of the air. It seems reasonable to infer that electromagnetic disturbances emitted from a wireless sending antennæ spread horizontally outwards, and also upwards, until the conducting strata of the atmosphere are encountered, after which the waves will move horizontally outwards in a 50-mile layer between the electrically-reflecting surface of the ocean beneath, and an electrically-reflecting surface, or successive series of surfaces, in the rarefied air above.

If this reasoning is correct, the curvature of the earth plays no significant part in the phenomena, and beyond a radius of, say, 100 miles from the transmitter, the waves are propagated with uniform attenuation cylindrically, as though in two-dimensional space. The problem of long-distance wireless wave transmission would then be reduced to the relatively simple condition of propagation in a plane, beyond a certain radius from the transmitting station. Outside this radius the voluminal energy of the waves would diminish in simple proportion to the distance, neglecting absorption losses at the upper and lower reflecting surfaces, so that at twice the distance the energy per square meter of wave front would be halved. In the absence of such an upper reflecting surface, the attenuation would be considerably greater. As soon as long-distance wireless waves come under the sway of accurate measurement, we may hope to find, from the observed attenuations, data for computing the electrical conditions of the upper atmosphere. If the attenuation is found to be nearly in simple proportion to the distance, it would seem that the existence of the upper reflecting-surface could be regarded as demonstrated.

---

two hours. It then begins to strengthen, and finally reaches a maximum—sometimes a very high one—at the time of sunset at Glace Bay. During the night the signals are very variable in strength, varying from very weak to very strong. Shortly before sunrise at Clifden the signals grow stronger, and reach a high maximum shortly after sunrise; they now dwindle to a marked minimum about two hours after, and then they return gradually to their normal day strength. The ratio of the intensity of the signals during the twilight maximum to the average intensity throughout the day is much greater for a wave 5000 metres long (frequency 60,000 per second) than for a wave 7000 metres long (frequency 43,000 per second), and the long wave signals are uniformly stronger during the day than those of the shorter wave.

Some of these observed facts can be understood by the aid of the hypothesis of ionic refraction. We have only to lay down the principle that the aggregate curvature of the trajectory of the longer waves is nearer to the curvature of the earth than that of the shorter waves, or, in other words, that the daylight trajectory of the longer waves is more suited than that of the shorter waves to the distance between the Irish and Canadian stations. Again, the minimum that occurs at about two hours after sunset at Clifden is readily explained by the conception, already discussed, that in the twilight regions the recombination of ions has consequences equivalent to a somewhat opaque curtain hanging from the top to the bottom of the middle atmosphere. Moreover, at two hours after sunset at Clifden the sun is setting at a place between the stations about 650 miles from Glace Bay. At this place the horizontal plane of Glace Bay passes between 50 and 60 miles overhead. If now we assume that the height of the curtain of irregularly ionised air is of the order 50 miles, thus making no allowance whatever for the bending of the rays from Glace Bay in their progress below that level, we see that the signals transmitted to Clifden are weakest when the curtain comes on the horizon of the sending station. If, on the other hand, we allow for the likelihood of the rays following considerably bent paths even in the lower middle atmosphere, we must take Marconi's results as showing that the ionic curtain reaches effectively to much lower atmospheric levels than 50 miles. All of this applies, *mutatis mutandis*, to the morning minimum produced by the sunrise belt passing between the stations.

In regard to the remaining point quoted from Dr. Marconi, namely, the strong maximum in signals to Clifden at about sunset at Glace Bay, and before sunrise at Clifden, there is more difficulty in finding an explanation. It would seem that the heterogeneous ionisation following the twilight carried on with radiation that had started at considerable upward inclination, and would therefore be accomplished with difficulty. The wave-fronts arriving at the receiving station would also be tilted forward considerably, and consequently the horizontal component of the electric field of the waves might approach the magnitude of the vertical component. In this case an inclined antenna would be a better receiver than a vertical one.

In various parts of the world it has been found that stations on the opposite sides of a mountain chain can communicate in the night with ease, though only with great difficulty, if at all, in the day. This is especially the case if a short wave is in use, and such a pair of stations can sometimes establish day communication by adopting a longer wave-length. It is, in fact, now common knowledge that for communication across hilly country in the daytime, a long wave—a thousand metres or more—should be used. The explanation is obvious on the hypotheses developed above. The rays, starting with sufficient elevation from a sending station in the plains, travel in straight lines through the lower atmosphere past the mountain tops, and then, reaching the middle atmosphere, are deflected downward by refraction in the ionised air. Short waves are refracted much less than long waves, and are therefore not bent so fully into the lower atmosphere as are the long waves. Indeed, the short waves may be entirely lost, and the long waves be bent down abundantly and come to earth again on the far side of the mountains. In the night, however, the ionisation of the middle atmosphere has disappeared, the Heaviside layer is open, and waves of all frequencies are reflected down to earth again. Another fact that emphasises the existence of elevated trajectories is afforded by the experiences of the Alpine receiving stations. These stations commonly receive signals from great distances in all directions—from stations in all parts of Europe and from ships on the Atlantic—so that it has been said that "the Alps attract signals." Stations in the plains do not get these distant signals nearly so often. The fact is that the high mountain stations have, of course, a much better chance of lying on the trajectories of the waves, or, as suggested by Larmor, in a slightly different connection, of "tapping a stronger stratum of radiation."

Recently, in an evening discourse at the Royal Institution, Marconi has described* the striking effects of sunrise and sunset on the strength of signals received from across the Atlantic Ocean. He stated that the intensity of the signals received at Clifden, Ireland, from Glace Bay, Canada, remains fairly steady during the day, but shortly after sunset at Clifden it becomes gradually weaker, and reaches a minimum in about

---

* June 2, 1911.

RADIOTELEGRAPHY.[1]

[With 1 plate.]

By Commendatore G. Marconi, LL.D., D.Sc.

The practical application of electric waves to the purposes of wireless telegraphic transmission over long distances has continued to extend to a remarkable degree during the last few years, and many of the difficulties, which at the outset appeared almost insurmountable, have been gradually overcome, chiefly through the improved knowledge which we have obtained in regard to the subject generally and to the principles involved.

The experiments which I have been fortunate enough to be able to carry out, on a much larger scale than can be done in ordinary laboratories, have made possible the investigation of phenomena often novel and certainly unexpected.

Although we have—or believe we have—all the data necessary for the satisfactory production and reception of electric waves, we are yet far from possessing any very exact knowledge concerning the conditions governing the transmission of these waves through space, especially over what may be termed long distances. Although it is now perfectly easy to design, construct, and operate stations capable of satisfactory commercial working over distances up to 2,500 miles, no really clear explanation has yet been given of many absolutely authenticated facts concerning these waves. Some of these hitherto apparent anomalies I shall mention briefly in passing.

Why is it that when using short waves the distances covered at night are usually enormously greater than those traversed in the day time, while when using much longer waves the range of transmission by day and night is about equal and sometimes even greater by day?

What explanation has been given of the fact that the night distances obtainable in a north-southerly direction are so much greater than those which can be effected in an east-westerly one?

Why is it that mountains and land generally should greatly obstruct the propagation of short waves when sunlight is present and not during the hours of darkness?

---

[1] Reprinted by permission from author's separate of Proceedings of the Royal Institution. Read before Royal Institution of Great Britain at weekly evening meeting, Friday, June 2, 1911.

The general principles on which practical radiotelegraphy is based are now so well known that I need only refer to them in the briefest possible manner.

Wireless telegraphy, which was made possible by the fields of research thrown open by the work of Faraday, Maxwell, and Hertz, is operated by electric waves, which are created by alternating currents of very high frequency, induced in suitably placed elevated wires or capacity areas. These waves are received or picked up at a distant station on other elevated conductors tuned to the period of the waves, and the latter are revealed to our senses by means of appropriate detectors.

My original system as used in 1896 consisted of the arrangement shown diagrammatically in figure 1, where an elevated or vertical wire was employed. This wire sometimes terminated in a capacity or was connected to earth through a spark gap.

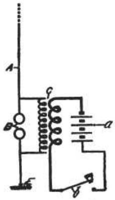

Fig. 1.

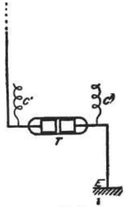

Fig. 2.

By using an induction coil or other source of sufficiently high tension electricity sparks were made to jump across the gap; this gave rise to oscillations of high frequency in the elevated conductor and earth, with the result that energy in the form of electric waves was radiated through space.

At the receiving station (fig. 2) these waves induced oscillatory currents in a conductor containing a detector, in the form of a coherer, which was usually placed between the elevated conductor and earth.

Although this arrangement was extraordinarily efficient in regard to the radiation of electrical energy, it had numerous drawbacks.

The electrical capacity of the system was very small, with the result that the small amount of energy in the aerial was thrown into space in an exceedingly short period of time. In other words, the energy, instead of giving rise to a train of waves, was all dissipated after only a few oscillations, and, consequently, anything approaching good tuning between the transmitter and receiver was found to be unobtainable in practice.

reach the earth again in appreciable amount. This would probably not have happened with longer waves.

As for the night signals, both long and short waves are propagated through the lower and middle atmosphere in straight lines to great heights and reflected at the Heaviside layer, and then they descend to earth again, having suffered comparatively little absorption. The waves may be imagined to creep round this electrical vault of the atmosphere in a manner somewhat analogous to the creeping of sound round a whispering gallery, being plentifully scattered downward in their progress by the irregularities in the reflecting surface, or, to put it another way, we may imagine that a transmitting station "lights up the sky," in an electrical sense, for many degrees below its horizon.

Since in the observations quoted the daylight signals were perceptible at 700 miles, where the horizontal plane of the sender crosses the observer's vertical at a height of 60 miles, we may conclude that the trajectory of the radiation directed horizontally from the sender does not reach higher than 30 miles in the daytime. It may be mentioned at this point that Marconi originally suggested that the phenomenon might be due to a possible discharging action of sunlight on the sending antenna, and J. J. Thomson considered it rather due to the absorption of energy by the ionised air in the immediate neighbourhood of the antenna, but both these explanations ought to make the contrast between day and night signals the same for short distances as for long, which is not the case.

The above considerations suggest that there should exist a best frequency for signalling over great distances. Now the radiation from a Hertzian oscillator is most intense in its equatorial plane, and therefore, from a vertical linear earthed antenna, is most intense in the horizontal plane of the sending station. Hence we conclude that the best frequency is that for which, in a given ionic condition of the atmosphere, the trajectory of the radiation which starts nearly horizontally returns to the earth's surface near the receiving station. Marconi has stated* that a wave-length of 5000 metres is almost always better than one of 4000 metres for Transatlantic signalling—though, he remarks, the shorter wave-length is better than the longer occasionally. In this connection it may be pointed out that, if the transmission of signals be attempted with an exceedingly long wave-length, the aggregate curvature produced by the ionic refraction in the day might be sharper than the curvature of the earth. This would cause the nearly horizontal radiation to be turned down to the earth within a relatively short distance from the radiator. In that case reception at a distance would be exhibit a different frequency; or, in other words, there should be a distinct best frequency at which to adjust the receiving apparatus at each of these stations. This is on account of the difference of trajectory that difference of frequency brings. There is but little experimental evidence bearing on this point, though what there is favours the assumptions; but clearly if dispersion do occur in the sunlit middle atmosphere, and do not occur at night, the weakness of the day strays is fully accounted for without invoking the assistance of absorption.

---

* Nobel Lecture, December, 1909.

Though the hypothesis of propagation round the earth by refraction in the ionised middle atmosphere has now been applied to the problem that prompted it, namely, the explanation of the minimum phenomenon of natural electric waves, yet it seems desirable to enquire how the hypothesis comports itself towards the known facts and properties of the artificial electric waves used in signalling.

The two assumptions on which the discussion has so far been built are, first, that there exists in the atmosphere a permanently conducting upper layer which is somewhat sharply defined, and which therefore reflects waves of every frequency—we may call it Heaviside's reflecting layer; and, second, that in the day (and only to a slight and erratic extent in the night) the atmosphere below this reflecting layer is ionised in nearly horizontal strata, the ionisation diminishing as the earth's surface is approached, with the result that electric waves are given a bent trajectory and the Heaviside layer put out of action. In using these assumptions in what follows, the atmosphere will be supposed at rest.

One of the most important of the facts known concerning the transmission of artificial waves is the difference between day and night signals discovered by Marconi* in 1902 during a voyage from England to New York. He found that there was little difference between day and night signals at distances less than 500 miles from the sending station, but that the day signals were unreadable at distances of 800 miles and more, while the night signals were readable up to distances of 2000 miles. This is possibly due to the same causes as the weakening of the day strays relatively to the night strays, but is most probably due to the failure of the heterogeneously ionised air to bend the waves sufficiently to fit the convexity of the earth. Thus, in explaining the daylight effect observed first by Marconi in 1902, it is only necessary to suppose that the relatively short waves then in use travelled to great heights in the atmosphere on account of the smallness of the curvature of their trajectory, and were not refracted sufficiently to

---

* Marconi, 'Roy. Soc. Proc.,' June, 1902.

Many mechanical analogies could be quoted which show that in order to obtain syntony the operating energy must be supplied in the form of a sufficient number of small oscillations or impulses properly timed. Acoustics furnish us with numerous examples of this fact, such as the resonance produced by the well-known tuning fork experiment.

Other illustrations of this principle may be given; e. g., if we have to set a heavy pendulum in motion by means of small thrusts or impulses, the latter must be timed to the period of the pendulum, as otherwise its oscillations would not acquire any appreciable amplitude.

In 1900 I first adopted the arrangement which is now in general use, and which consists (as shown in fig. 3) of the inductive association of the elevated radiating wire with a condenser circuit which may be used to store up a considerable amount of electrical energy and impart it at a slow rate to the radiating wire.

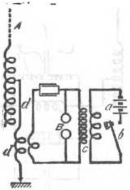

Fig. 3.

As is now well known, the oscillations in a condenser circuit can be made to persist for what is electrically a long period of time, and it can be arranged moreover that by means of suitable aerials or antennæ these oscillations are radiated into space in the form of a series of waves, which through their cumulative effect are eminently suitable for enabling good tuning and syntony to be obtained between the transmitter and receiver.

The circuits, consisting of the condenser circuit and the elevated aerial or radiating circuit, were more or less closely coupled to each other. By adjusting the inductance in the elevated conductor, and by the employment of the right value of capacity or inductance required in the condenser circuit, the two circuits were brought into electrical resonance, a condition which I first pointed out as being essential in order to obtain efficient radiation and good tuning.

The receiver (as shown in fig. 4) also consists of an elevated conductor or aerial connected to earth or capacity through an oscillating transformer. The latter also contains the condenser and detector, the circuits being made to have approximately the same electrical time period as that of the transmitter circuits.

At the long distance station situated at Clifden, in Ireland, the arrangement which has given the best results is based substantially upon my syntonic system of 1900, to which have been added numerous improvements.

An important innovation from a practical point of view was the adoption at Clifden and Glace Bay of air condensers, composed of insulated metallic plates suspended in air at ordinary pressure. In this manner we greatly reduce the loss of energy which would take place in consequence of dielectric hysteresis were a glass or solid dielectric employed. A very considerable economy in working also results from the absence of dielectric breakages, for, should the potential be so raised as to even produce a discharge from plate to plate across the condenser, this does not permanently affect the value of the dielectric, as air is self-healing and one of the few commodities which can be replaced at a minimum of cost.

Various arrangements have been tried and tested for obtaining continuous or very prolonged trains of waves, but it has been my experience that, when utilizing the best receivers at present available, it is neither economical nor efficient to attempt to make the waves too continuous. Much better results are obtained when groups of waves (fig. 5) are emitted at regular intervals in such manner that their cumulative effect produces a clear musical note in the receiver, which is tuned not only to the periodicity of the electric waves transmitted but also to their group frequency.

In this manner the receiver may be doubly tuned, with the result that a far greater selectivity can be obtained than by the employment of wave tuning alone.

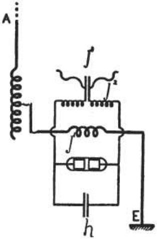

FIG. 4.

In fact, it is quite easy to pick up simultaneously different messages transmitted on the same wave length, but syntonized to different group frequencies.

As far as wave tuning goes, very good results—almost as good as are obtainable by means of continuous oscillations—can be achieved with groups of waves, the decrement of which is in each group 0.03 or 0.04, which means that about 30 or 40 useful oscillations are radiated before their amplitude has become too small to perceptibly affect the receiver.

The condenser circuit at Clifden has a decrement of from 0.015 to 0.03 for fairly long waves.

This persistency of the oscillations has been obtained by the employment of the system shown in figure 6, which I first described in a patent taken out in September, 1907. This method eliminates

---

is sometimes a complete zero lasting for only two minutes or less. Assuming that in this case the source of the natural electric waves and the receiving station lie both on the great circle of twilight, we deduce that the twilight band is at least 30 miles wide, this being the distance the earth rotates eastward in two minutes. In addition, the same assumption would indicate that the principal source of the strays during November, 1909 and 1911, lay in the direction of the eastern portion of the Atlas Mountains.

Again, the observations have shown that the chief twilight minima occur about 10 minutes before sunrise and about 10 minutes after sunset during the same periods. This fact is accounted for by the consideration that the time of sunrise in the middle atmosphere is a little earlier and the time of sunset a little later than at the surface of the earth. In this connection it should be noticed that the electrically effective sunrise at a point in the middle atmosphere, as measured by the ionising power of the sunlight reaching the point, is not coincident in time with the sunrise at the same point as indicated by mere luminosity, that is, with the geometrical sunrise. The rays heralding the sunrise in the latter and ordinary sense must possess very little ionising power, for the reason that in passing the earth tangentially they have traversed so long a path in the lower atmosphere as to have lost their ionising radiation. Thus an observation of the time interval of the stray minimum before geometrical sunrise and after sunset does not determine the height of the electrically disturbed regions. Besides this, the exact time of the stray minimum will be affected to some extent by the obliquity of the ionic curtain to the line of propagation of the waves.

During the day the electric waves travel in the relatively narrow shell of dielectric between some stratum in the middle atmosphere and the surface of the earth. At night they travel in the much wider shell of dielectric between the assumed high conducting layer and the earth. In England, in winter, the day strays are much weaker than the night strays. From this we might conclude either that the aggregate absorption in the thin shell of dielectric is greater than in the deeper night shell, or that the ionisation of the middle atmosphere during the day is sufficiently non-uniform to hinder the propagation of the waves. But another factor must be recognised. The electric disturbance produced by a lightning discharge is doubtless impulsive in character, and is probably either a solitary wave or a very short train of waves. The study of the refraction of such disturbances leads to some well-known theoretical difficulties, but if we assume that the wave undergoes dispersion during its progress through the ionised middle atmosphere, then the disturbance arriving at a given receiving station should exhibit a fairly definite frequency, and that arriving at a station at a different distance should still atmosphere be drawn round the globe they will be nearest the earth at places where the sun is on the meridian, and will rise away from the earth somewhat sharply at places where the sun is rising or setting. The regions in which the change from the day level to the night level takes place form a great circular band round the globe and inclined to the meridians at an angle depending on the season. This region of the atmosphere, since it is perpetually moving with the sun, will be in a highly disturbed electrical condition. Formation of ions is actively proceeding in one half of the great circle and recombination in the other, and these processes doubtless take place somewhat irregularly even in a still atmosphere—with the result that patches or banks of ionised air, analogous to the banks of fog met at sea—will transiently constitute this band in the middle atmosphere. The effect of such patches of variously ionised air on electric waves propagated through the region is, in view of the connection between the velocity of the waves and the concentration of the ions, certain to be difficult. The scattering by repeated refractions will tend to make the region impenetrable to waves directed through it. Hence it may be expected that the regularity of the propagation through the steadily ionised horizontal strata of the daytime will be greatly disturbed by the twilight transitional banks and patches, with the ultimate consequence that the sounds heard in the receiving apparatus will be greatly weakened.

The author's experience up to the present indicates that the existence or non-existence of clouds in the vicinity of the receiving station has but little influence on the intensity and character of the stray minima, or, for that matter (provided the day is not brilliantly clear), on signals received from any distance and any point of the compass. Whence we may conclude that the irregularly ionised band is situated above the ordinary cloud level. The twilight transitional region may therefore be regarded as a sort of curtain enringing the earth and occupying the middle atmosphere and not the lower. Thus it can affect only the trajectories of waves travelling from great distances. The weakening of such long-distance waves will probably be greater or less according as they have to penetrate the curtain more or less obliquely. In the case of the natural electric waves received by an antenna in England during the autumn and winter the origin of the waves must in general lie to the south. It is reasonable to suppose that tropical Africa will supply most of them. In that case, the twilight transitional band must have a very great and a relatively short-lived influence on the intensity of the strays heard in the telephones, for the path of the waves from the suggested source to the receiving station is nearly coincident with the twilight band. These suppositions accord precisely with the observed facts. The minimum

---

almost completely the spark gap and its consequent resistance, which, as is well known, is the principal cause of the damping or decay of the waves in the usual transmitting circuit.

The apparatus shown in figure 6 consists of a metal disk *a*, having copper studs firmly fixed at regular intervals in its periphery and placed transversely to its plane. This disk is caused to rotate very rapidly between two other disks, *b*, by means of a rapidly revolving electric motor or steam turbine. These side disks are also made to slowly turn round in a plane at right angles to that of the middle disk. The connections are as illustrated in the figure. The studs are of such length as to just touch the side disks in passing, and thereby bridge the gap between the latter.

With the frequency employed at Clifden, namely, 45,000, when a potential of 15,000 volts is used on the condenser, the spark gap is practically closed during the time in which one complete oscillation only is taking place, when the peripherical speed of the disk is about 600 feet a second. The result is that the primary circuit can continue oscillating without material loss by resistance in the spark gap. Of course the number of oscillations which can take place is governed by the breadth or thickness of the side disks, the primary circuit being abruptly opened as soon as the studs attached to the middle disk leave the side disks.

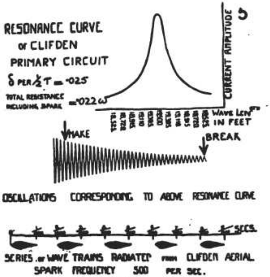

Fig. 5.

This sudden opening of the primary circuit tends to immediately quench any oscillations which may still persist in the condenser circuit; and this fact carries with it a further and not inconsiderable advantage, for if the coupling of the condenser circuit to the aerial is of a suitable value the energy of the primary will have practically all passed to the aerial circuit during the period of time in which the

primary condenser circuit is closed by the stud filling the gap between the side disks; but after this the opening of the gap at the disks prevents the energy returning to the condenser circuit from the aerial, as would happen were the ordinary spark gap employed. In this manner the usual reaction which would take place between the aerial and the condenser circuit can be obviated, with the result that with this type of discharger and with a suitable degree of coupling the energy is radiated from the aerial in the form of a pure wave, the loss from the spark gap resistance being reduced to a minimum.

I am able to show a resonance curve taken at Clifden which was obtained from the oscillations in the primary alone (fig. 5).

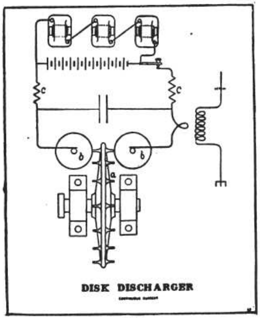

Fig. 6.

An interesting feature of the Clifden plant, especially from a practical and engineering point of view, is the regular employment of high-tension direct current for charging the condenser. Continuous current at a potential which is capable of being raised to 20,000 volts is obtained by means of special direct-current generators; these machines charge a storage battery consisting of 6,000 cells, all connected in series, and it may be pointed out that this battery is the largest of its kind in existence. The capacity of each cell is 40 ampere-hours. When employing the cells alone the working voltage is from 11,000 to 12,000 volts, and when both the direct-current generators and the battery are used together the potential may be raised to 15,000 volts through utilizing the gassing voltage of the storage cells.

For a considerable portion of the day the storage battery alone is employed, with a result that for 16 hours out of the 24 no running

---

0·01. This in turn indicates that the number of ions per cubic centimetre should be about 160,000 when the wave-length is 2000 metres, and about 16,000 when the wave-length is 6000 metres. Such ionic concentrations are not improbable.

It may be objected that there has as yet been no experimental corroboration of this concentration of the energy of the waves into a comparatively thin stratum near the surface of the earth. But as a fact no measurements have as yet been carried out over great distances on the variation of intensity of signals with distance and under unvarying atmospheric conditions; and clearly the measurements that have been made over short distances—which all support the inverse square law as was to be expected—cannot have any bearing on refractions that take place in high layers. When measurements of intensity over distances of 1000 to 2000 miles become available it may be expected that the inverse square law will hold for a distance of three or four hundred miles, and after that a law indicating rather less divergence may hold for several hundred miles more, if the frequency is low enough for the ionic refraction to produce bending at least as great as the convexity of the globe.*

We now proceed to the explanation of the stray minima described earlier. In the first place it is evident that if the surfaces of equal ionisation in a

---

* Since the above was written, an account by L. W. Austin ('Washington Bureau of Standards Bulletin,' October, 1911) of some new measurements of the intensity of signals has come to hand, which appears to strengthen the author's position greatly, so far as the measurements go. The inverse square law for the divergence of energy shows that $I \propto x^{-1}$, where I is the intensity of the current received on an antenna, and x is the distance of the sending station. If the waves were travelling in free space of electrical constitution like that of our lower atmosphere, the absorption would demand a formula of the type

$$I \propto x^{-1}\epsilon^{-\alpha x},$$

where α is independent of the wave-length. Also it has been shown above that if the waves were travelling in free space filled with air highly ionised, the absorption would require a formula of the type

$$I \propto x^{-1}\epsilon^{-\alpha x\lambda^2},$$

where α is again independent of the wave-length. Now, the observations quoted support an empirical formula very different from either of these, namely,

$$I \propto x^{-1}\epsilon^{-\alpha x/\sqrt{\lambda}}.$$

A formula involving the wave-length in this manner is not suggested, and cannot even be accounted for, by absorption in an ionised atmosphere or in a badly conducting surface such as that of land or sea; but it is clearly in rough general accord with the law, developed in this paper, that the desired bending of the rays is better with long waves than with short waves; or, to put it in another way, the loss of radiation by failure to turn the curve of the earth is greater with short waves than with long. The measurements made up to the present by Austin and his collaborators have extended to distances of only 900 miles, involving only very slight bending, and, besides, have not been very numerous.

ions moving up and negative ions moving down, with the result that some parts of the middle atmosphere may remain ionised after the sun has set. If we suppose, however, that the ions do for the most part recombine, then the effect of the change from day to night is to remove a veil, as it were, of ionised air from between the upper conducting layer and the earth.

Since the velocity of the waves in the sunlit middle atmosphere is greater the higher the level at which they are travelling, a ray of electric radiation starting from a point of the earth's surface in a direction inclined slightly upward will pursue a straight path in the lower atmosphere and a slightly bent path, with its concavity downwards, in the middle atmosphere, thus following to a greater or less extent the curvature of the earth. If its curvature in the middle atmosphere is on the average greater than that of the earth—and not otherwise—the ray will be turned down to the lower atmosphere and will again traverse a straight line. In other words, the wave-fronts will be tilted forward as they travel, in a manner quite analogous to the refraction of sound in air when the temperature varies upward. This bending of the rays may be given for shortness the name ionic refraction; it would be very rapid in layers where γ approached unity. Thus the radiations diverging in all directions from a lightning stroke or from a wireless telegraph antenna become confined in the day between the conducting surface of the earth and a certain level in the middle atmosphere. Even the rays that start horizontally from a place on the earth's surface must, owing to the earth's curvature, reach, within 300 miles of the source, to heights where the air may be expected to be strongly ionised, and must then suffer refraction downwards. Since the quantity γ is inversely proportional to the frequency of the transmitted waves, the limiting height of penetration of the waves is smaller the lower the frequency, and therefore low-frequency waves become concentrated nearer to the earth's surface than do higher frequency waves.

The curvature of the trajectory of waves travelling at a height z is $dv/(vdz)$, v being the velocity at the place. But when γ is small $v = (1+\frac{1}{2}\gamma)/\sqrt{(\mu\kappa)}$, and therefore the curvature is $\frac{1}{2}d\gamma/dz$ approximately. If we assume a condition of things in which the radius of curvature of rays at all heights is $r+z$, where r is the radius of the earth, and that $\gamma = 0$ at $z = 0$, we find

$$\gamma = 2\log\frac{r+z}{r} = 2\frac{z}{r}, \text{ approximately.}$$

Though, in fact, the bending of rays in the lower atmosphere is probably not so great as this, the equation indicates that the order of magnitude of γ required in the middle atmosphere at, say, a height of 20 miles is about

---

machinery need be used for operating the station, with the single exception of the small motor revolving the disk.

The potential to which the condenser is charged reaches 18,000 volts when that of the battery or generators is 12,000. This potential is obtained in consequence of the rise of potential at the condenser plates, brought about by the rush of current through the choking or inductance coils at each charge. These coils are placed between the battery or generator and the condenser *c*, figure 6.

No practical difficulty has been encountered either at Clifden or Glace Bay in regard to the insulation and maintenance of these high-tension storage batteries. Satisfactory insulation has been obtained by dividing the battery into small sets of cells placed on separate stands. These stands are suspended on insulators attached to girders fixed in the ceiling of the battery room. A system of switches, which can all be operated electrically and simultaneously, divides the battery into sections, the potential of each section being low enough to enable the cells to be handled without inconvenience or risk.

The arrangement of aerial adopted at Clifden and Glace Bay is shown in figure 7. This system, which is based on the result of tests which I first described before the Royal Society in June, 1906,[1] not only makes it possible to efficiently radiate and receive waves of any desired length, but it also tends to confine the main portion of the radiation to any desired direction. The limitation of transmission to one direction is not very sharply defined, but nevertheless the results obtained are exceedingly useful for practical working.

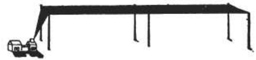

FIG. 7.

In a similar manner, by means of these horizontal wires, it is possible to define the bearing or direction of a sending station, and also limit the receptivity of the receiver to waves arriving from a given direction.

The commercial working of radiotelegraphy and the widespread application of the system on shore and afloat in nearly all parts of the world has greatly facilitated the marshaling of facts and the observation of effects. Many of these, as I have already stated, still await a satisfactory explanation.

A curious result which I first noticed over nine years ago in long-distance tests carried out on the steamship *Philadelphia*, and which still remains an important feature in long-distance space telegraphy,

---

[1] "On methods whereby the radiation of electric waves may be mainly confined, etc." Proc. Roy. Soc., A, vol. 77, p. 413.

is the detrimental effect produced by daylight on the propagation of electric waves over great distances.

The generally accepted hypothesis of the cause of this absorption of electric waves in sunlight is founded on the belief that the absorption is due to the ionization of the gaseous molecules of the air affected by the ultra violet light, and as the ultra violet rays which emanate from the sun are largely absorbed in the upper atmosphere of the earth, it is probable that that portion of the earth's atmosphere which is facing the sun will contain more ions or electrons than that which is in darkness, and therefore, as Sir J. J. Thomson has shown,[1] this illuminated or ionized air will absorb some of the energy of the electric waves.

The wave length of the oscillations employed has much to do with this interesting phenomenon, long waves being subject to the effect of daylight to a very much lesser degree than are short waves.

Although certain physicists thought some years ago that the daylight effect should be more marked on long waves than on short, the reverse has been my experience; indeed, in some transatlantic experiments, in which waves about 8,000 meters long were used, the energy received by day at the distant receiving station was usually greater than that obtained at night.

Recent observation, however, reveals the interesting fact that the effects vary greatly with the direction in which transmission is taking place, the results obtained when transmitting in a northerly and southerly direction being often altogether different from those observed in the easterly and westerly one.

Research in regard to the changes in the strength of the received radiations which are employed for telegraphy across the Atlantic has been recently greatly facilitated by the use of sensitive galvanometers, by means of which the strength of the received signals can be measured with a fair degree of accuracy.

In regard to moderate power stations such as are employed on ships, and which, in compliance with the international convention, use wave lengths of 300 and 600 meters, the distance over which communication can be effected during daytime is generally about the same, whatever the bearing of the ships to each other or to the land stations —whilst at night interesting and apparently curious results are obtained. Ships over 1,000 miles away, off the south of Spain or round the coast of Italy, can almost always communicate during the hours of darkness with the post-office stations situated on the coasts of England and Ireland, whilst the same ships, when at a similar distance on the Atlantic to the westward of these islands and on the usual track between England and America, can hardly ever communi-

---

[1] Philosophical Magazine, ser. 6, vol. 4, p. 253.

---

molecular size, and the other is that these same agencies also produce, by direct action on the gases of the atmosphere, condensation nuclei consisting of solid or liquid compounds which are not electrically charged when first formed. Evidently, the condensation nuclei can have only slight influence on the value of the quantity γ as compared with the electrical carriers of molecular size. These heavy ions, or condensation nuclei, doubtless frequently become charged by attaching one or more of the light ions, which has the effect of putting such ions out of action for our purposes. The lighter ions are probably in a majority in the higher parts of the middle atmosphere, and the heavy ions in a majority in the lower parts, and also in the lower atmosphere. It may be mentioned that the heavier ions found in the lower atmosphere, whether consequences of solar radiation or not, escape being counted by the kind of apparatus usually used in measurements of atmospheric electricity, on account of their immobility.

So far as the quantity γ is concerned the principal difference between day and night conditions, and between the conditions at different times of the day, is due to the variations of the number of ions per cubic centimetre provided by solar radiation. It is not possible to be precise on this matter, even in the lower atmosphere. But, broadly, it is clear that the value of n in the lower and middle atmospheres must vary considerably with the obliquity of the sun's rays at the place, that is to say, with the season and the time of day, and also must vary profoundly from daylight to darkness. About the time of sunrise at any particular place the process of ionisation by the solar radiation will be occurring at all heights of the atmosphere over an area extending many miles to east and west of the place; at sunset recombination of the ions will occur through a similar space. Of course the layers nearest the earth will be least disturbed in electrical constitution, mainly for the reason that the sun's rays must have been robbed of much of their ionising powers by the time they reach low levels; but it is known from direct observations at various levels up to heights of several miles that the influence of the sun is quite perceptible. But, in fact, it is easily seen from the formula deduced above for the absorption per wave-length, that the normal ionisation observed in the lower atmosphere produces inappreciable absorption of the waves at any time of day over terrestrial distances.

Our knowledge of the conditions ruling in the hours of darkness is even less precise than that of the day conditions. The very rapid rate of recombination of ions when the ionising agent is removed points to the possibility of the middle atmosphere being perfectly free from ions during darkness. But it is probable that there occurs during the day a great sifting of oppositely charged ions under the operation of the earth's vertical electric field, positive

In air under standard conditions $d\xi/dt = 1\cdot5$ cm./sec. in a field having a gradient of 1 volt per centimetre. Take $e = 1\cdot0 \times 10^{-20}$ in electromagnetic units, then $f = 7 \times 10^{-13}$, roughly. Again taking $p = 10^7$ and $m = 2 \times 10^{-21}$ in grammes, then $mp = 2 \times 10^{-14}$. Thus for waves of the order of frequency 1,000,000 per second, the two terms in the denominator of the quantity γ are of about the same order of importance at low levels in the atmosphere. For waves of lower frequency the term f^2 is much more important than the term m^2p^2, which may then be neglected in the denominator of γ. At higher levels, on the contrary, the term f^2 probably becomes negligible in comparison with m^2p^2 since the value of f is known to fall off much faster than that of m as the rarefaction increases. Thus, at high levels, we have

$$\gamma = 4\pi ne^2/\kappa mp^2, \text{ approximately.}$$

At low levels the value of γ works out as $0\cdot55 \times 10^{-14} \times n$ with the numbers already assumed. The number of ions per cubic centimetre at sea level is often given as between 1000 and 10,000. Thus γ is quite negligible compared with unity if this estimate of n and that of f are valid. At high levels, using the last equation, we find $\gamma = 0\cdot6 \times 10^{-11} \times n$ for $p = 10^7$ (corresponding to a wave-length of nearly 200 metres) and $\gamma = 0\cdot6 \times 10^{-9} \times n$ for $p = 10^6$ (corresponding to a wave-length of nearly 2000 metres in ordinary air). It is more than probable, however, that at moderately high levels, where the air is rather rarefied—for example, at a height of 20 miles the pressure is, on the theory of convective equilibrium, about 1/100 of the pressure at sea level—the ion is of much smaller mass, say 100 times smaller, than is assumed above, and this would make the last figure become $\gamma = 0\cdot6 \times 10^{-7} \times n$ for $p = 10^6$.

For convenience in discussion, the portion of the atmosphere below the permanently conducting layer and throughout which the equation

$$\gamma = 4\pi ne^2/\kappa mp^2$$

holds good, that is the portion of the atmosphere which is ionised strongly and directly by the sun, will be called the middle atmosphere. The part below this will be called the lower atmosphere, and here the equation

$$\gamma = 4\pi ne^2 m/\kappa f^2$$

is probably appropriate while n has low values. Of the middle atmosphere and of the upper atmosphere we know nothing directly. Perhaps the best information available is that contained in the memoirs of P. Lenard and C. Ramsauer,* who showed that the ultra-violet light of the sun will produce in air two effects of interest in the present connection. One effect is that the ultra-violet light, and possibly the cathode rays, of the sun produce electrical carriers of

---

* 'Heidelberger Akademie Sitzungsber.,' 1910—11.

cate with these shore stations unless by means of specially powerful instruments.

It is also to be noticed that in order to reach ships in the Mediterranean the electric waves have to pass over a large portion of Europe and, in many cases, over the Alps. Such long stretches of land, especially when including very high mountains, constitute, as is well known, an insurmountable barrier to the propagation of short waves during the daytime. Although no such obstacles lie between the English and Irish stations and ships in the North Atlantic en route for North America, a night transmission of 1,000 miles is there of exceptionally rare occurrence. The same effects generally are noticeable when ships are communicating with stations situated on the Atlantic coast of America.

Although high power stations are now used for communicating across the Atlantic Ocean, and messages can be sent by day as well as by night, there still exist periods of fairly regular daily occurrence during which the strength of the received signals is at a minimum.

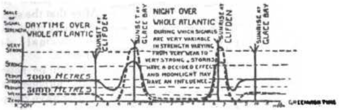

FIG. 8.

Thus in the morning and the evening, when, in consequence of the difference in longitude, daylight or darkness extends only part of the way across the ocean, the received signals are at their weakest. It would almost appear as if electric waves, in passing from dark space to illuminated space and vice versa, were reflected and refracted in such a manner as to be diverted from the normal path.

Later results, however, seem to indicate that it is unlikely that this difficulty would be experienced in telegraphing over equal distances north and south on about the same meridian, as, in this case, the passage from daylight to darkness would occur more rapidly over the whole distance between the two stations.

I have here some diagrams which have been carefully prepared by Mr. H. J. Round. These show the average daily variation of the signals received at Clifden from Glace Bay.

The curves traced on the diagram (fig. 8) show the usual variation in the strength of these transatlantic signals on two wave lengths —one of 7,000 meters and the other of 5,000 meters.

The strength of the received waves remains as a rule steady during daytime.

Shortly after sunset at Clifden they become gradually weaker, and about two hours later they are at their weakest. They then begin to strengthen again, and reach a very high maximum at about the time of sunset at Glace Bay.

They then gradually return to about normal strength, but through the night they are very variable. Shortly before sunrise at Clifden the signals commence to strengthen steadily, and reach another high maximum shortly after sunrise at Clifden. The received energy then steadily decreases again until it reaches a very marked minimum, a short time before sunrise at Glace Bay. After that the signals gradually come back to normal day strength.

It can be noticed that, although the shorter wave gives on the average weaker signals, its maximum and minimum variations of strength very sensibly exceed that of the longer waves.

Figure 9 shows the variations at Clifden during periods of 24 hours, commencing at 12 noon throughout the month of April, 1911, the vertical dotted lines representing sunset and sunrise at Glace Bay and Clifden.

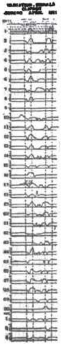

Fig. 9.

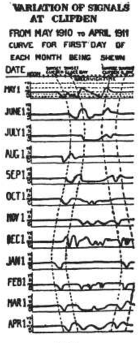

Fig. 10.

Figure 10 shows the curve for the first day of each month for one year, from May, 1910, to April, 1911.

I carried out a series of tests over longer distances than had ever been previously attempted, in September and October of last year, between the stations of Clifden and Glace Bay, and a receiving station placed on the Italian Steamship *Principessa Mafalda*, in the course of a voyage from Italy to Argentina (pl. 1, fig. 1).

Eliminating β and ζ among these three equations, we have

$$\frac{\partial^2 Z}{\partial x^2} = \mu\kappa \frac{\partial^2 Z}{dt^2} + \frac{4\pi\mu ne^2}{mip+f}\frac{\partial Z}{\partial t}.$$

There is a solution of the form

$$Z = \epsilon^{-lx+ip(t-x/v)}$$

for waves of frequency $p/2\pi$, if the velocity

$$v = \frac{mp}{\gamma f}\sqrt{\left(\frac{2(1-\gamma)}{\mu\kappa}\right)}\left\{\sqrt{\left[1+\left(\frac{\gamma f}{mp(1-\gamma)}\right)^2\right]}-1\right\}^{\frac{1}{2}}$$

and the absorption factor $l = \frac{\mu\kappa\gamma f}{2m}v.$

Here γ has been put for the quantity

$$\frac{4\pi ne^2 m}{\kappa(m^2p^2+f^2)},$$

which, it will be shown, is usually smaller than unity. In addition, the quantity $\gamma f/mp(1-\gamma)$ is usually very small compared with unity, hence, approximately,

$$v = \frac{1}{\sqrt{(\mu\kappa)}}(1+\tfrac{1}{2}\gamma), \qquad l = \frac{\sqrt{(\mu\kappa)}f}{2m}\gamma(1+\tfrac{1}{2}\gamma);$$

or the absorption coefficient per wave-length is

$$l' = \frac{\pi}{1-\gamma}\frac{\gamma f}{mp}.$$

It should be mentioned that, in forming the equations, it has been implicitly assumed that the ions are so heavy that they acquire only small velocities and make very small excursions under the action of the waves. That is to say, the ions are supposed to be collections of molecules, or, at the smallest, single molecules. The absorption in the case of very small ions, that is, electrons, has been worked out by J. J. Thomson,* and involves different considerations from those appropriate here.

Rough estimates of the various magnitudes involved in the last equations may be obtained by using the results of laboratory experiments on ionised gases, though, unfortunately, there are as yet but very few data available on the ionisation of air by solar radiation. First notice that the friction coefficient f can be estimated from experiments on the terminal velocities of ions in various gases. From the equation of motion of an ion we see that its terminal velocity under a steadily applied electric field Z is

$$\frac{d\zeta}{dt} = \frac{e}{f}Z.$$

* 'Phil. Mag.,' August, 1902, p. 253.

other recorded facts of long-distance transmission be explained. The writer has therefore investigated another, and closely related, possibility, which, it turns out, throws light on the causes of the stray minimum, as well as on many of the observed facts of long-distance transmission. To these new considerations we now turn.

The hypothesis to be introduced is based on the influence of the ionisation of the air on the propagation of electric waves through it. It is well known that, under normal conditions, the air at the sea-level is only slightly ionised even in strong sunshine, and that at a height of a few miles above the earth's surface the ionisation is, according to observations from balloons, sometimes 20 times as great as at the surface. Higher still the ionisation doubtless increases further, on account of the more and more intense ionising action of the solar radiation, which, it is plain, must be greater in these higher and rarer regions than in the dense regions below. No law can be legitimately assigned for the computation of this gradual transition from low conductivity to high conductivity, yet, for the purpose in hand, it is necessary to form some idea of the effect of this heterogeneity on the propagation of electric waves.

It is first necessary to examine the effect of charged ions of molecular mass on the velocity of the waves, and here we may follow, with suitable modifications, the standard methods applied to the study of optical dispersion in media containing electrons. Let e be the charge, m the mass, of each ion, and let n be the number of ions per cubic centimetre at a point whose co-ordinates are x, y, z, referred to a right-handed rectangular frame of axes with the axis of z vertical. Suppose that electric waves are advancing through the medium in the positive direction of the x axis, and with electric force Z, magnetic force β, at the point x, y, z. Then if μ be the permeability of the medium and κ the dielectric constant of the unionised air,

$$\frac{\partial Z}{\partial x} = \mu \frac{d\beta}{dt} \quad \text{and} \quad \frac{\partial \beta}{\partial x} = \kappa \frac{dZ}{dt} + 4\pi ne \frac{d\zeta}{dt},$$

where ζ indicates the displacement of each ion from its original position produced by the waves. The equation of motion of an ion is

$$m \frac{d^2\zeta}{dt^2} + f \frac{d\zeta}{dt} = eZ,$$

where f is a frictional constant of viscous type.

If the time factor in Z be ϵ^{ipt}, this equation becomes

$$\frac{d\zeta}{dt} = \frac{eZ}{mip+f}.$$

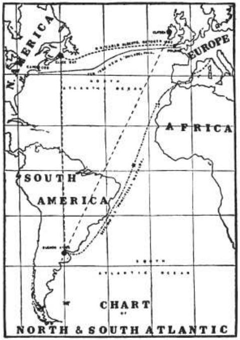

1. Long Distance Wireless Tests in 1910.

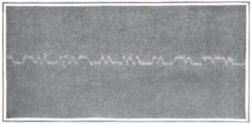

2. Record of Wireless Signals.

During these tests the receiving wire was supported by means of a kite, as was done in my early transatlantic tests of 1901, the height of the kite varying from about 1,000 to 3,000 feet. Signals and messages were obtained without difficulty, by day as well as by night, up to a distance of 4,000 statute miles from Clifden.

Beyond that distance reception could only be carried out during nighttime. At Buenos Aires, over 6,000 miles from Clifden, the night signals from both Clifden and Glace Bay were generally good, but their strength suffered some variations.

It is rather remarkable that the radiations from Clifden should have been detected at Buenos Aires so clearly at nighttime and not at all during the day, whilst in Canada the signals coming from Clifden (2,400 miles distant) are no stronger during the night than they are by day.

Further tests have been carried out recently for the Italian Government between a station situated at Massaua in East Africa and Coltano in Italy. Considerable interest attached to these experiments, in view of the fact that the line connecting the two stations passes over exceedingly dry country and across vast stretches of desert, including parts of Abyssinia, the Soudan, and the Libyan Desert. The distance between the two stations is about 2,600 miles.

The wave length of the sending station in Africa was too small to allow of transmission being effected during daytime, but the results obtained during the hours of darkness were exceedingly good, the received signals being quite steady and readable.

The improvements introduced at Clifden and Glace Bay have had the result of greatly minimizing the interference to which wireless transmission over long distances was particularly exposed in the early days.

The signals arriving at Clifden from Canada are as a rule easily read through any ordinary electrical atmospheric disturbance. This strengthening of the received signals has moreover made possible the use of recording instruments, which will not only give a fixed record of the received messages, but are also capable of being operated at a much higher rate of speed than could ever be obtained by means of an operator reading by sound or sight. The record of the signals is obtained by means of photography in the following manner: A sensitive Einthoven string galvanometer is connected to the magnetic detector or valve receiver, and the deflections of its filament caused by the incoming signals are projected and photographically fixed on a sensitive strip, which is moved along at a suitable speed (pl. 1, fig. 2). On some of these records, which I am able to show, it is interesting to note the characteristic marks and signs produced amongst the signals by natural electric waves or other electrical disturbances of

---

diffraction has been thoroughly settled by the investigations of Rayleigh, Poincaré, Macdonald, Nicholson, and others, whose work has shown irrefutably that the energy propagated round a quarter of the globe by the process of diffraction would be utterly inappreciable. Heaviside has suggested that, in the type of waves used in wireless telegraphy, the Faraday lines of electric force are attached, as it were, to the earth, and slide along its surface, and that therefore they cannot leave it; but this view, in fact, throws us back on the diffraction result. Again, in 1906, Zenneck added to this suggestion a consideration based on a well-known illustration of Poynting's theorem. In his original memoir* on the propagation of energy in the electromagnetic field, Poynting gives as an illustration of the theorem the case of a wire possessing resistance and carrying an electric current. Heaviside,† referring to this illustration, showed that the moving electric field cannot be purely radial; in other words, the Faraday lines belonging to the flowing electricity must lean forward as they move along the wire—so that the energy vector shall have a component perpendicular to the surface of the wire. This particular case has been worked out in great detail since the date of the memoir, but Zenneck's application was to the simple case of the propagation of a plane wave over a badly conducting solid, such as the earth, with a plane boundary. His suggestion is, briefly, that the tilting forward of the wave fronts near the conductor will lead to a general slow turning downward of the waves towards the earth, so that a large proportion of the energy of the waves will be deflected round the bend of the earth instead of being propagated linearly into space. This suggestion does not in reality assist the pure diffraction theory at all.

Another hypothesis was put forward by Heaviside in 1900, when he suggested that the attenuated gases of the upper atmosphere might provide a conducting surface concentric with the earth, between which and the surface of the earth itself the waves might spread with two-dimensional divergence. This hypothesis has not yet been supported or denied by any trustworthy experiments or observations. Such experiments or observations will necessarily have to be made on waves that have travelled long distances, for the upper layers of the atmosphere cannot be greatly concerned in short-distance transmission. But it will be perceived, from the remarks already made on the minima of natural electric waves, that this upper conducting layer might supply the explanation of the phenomenon, and thereby gain some support. On examination, however, it appears that the phenomenon cannot be explained by means of the bare hypothesis, and still less can the practically as perfect in strong sunlight as in the dark. On all these counts, therefore, the supposition that the minima phenomena under discussion can be explained by ionisation of the lower layers of the atmosphere must be discarded as untenable.

---

* 'Phil. Trans.,' Jan. 10, 1884.

† 'Electrical Papers,' vol. 2, p. 94.

We must therefore turn to the possibility of alterations in the resistance of the ground round the antenna, alterations of such a nature that the waves are absorbed much more freely during a certain few minutes of twilight than before or after. Here we may at the same time consider a subordinate possibility, namely, that the resistance of the ground within a few yards of the earth plates of the antenna might vary through a minimum value during twilight. If this really happened there would be extra dissipation near the earth plate of the energy of the oscillations excited on the antenna by the incoming waves. Both of these views have been negatived by the author's observations on signals coming from short distances, which show that signals originating at distances of less than 50 miles do not undergo appreciable diminution in strength, even on days when the stray minimum is well marked. But, to make certain of the non-existence of the subordinate possibility just referred to, a long series of determinations was made, during the past autumn, of the high-frequency resistance of the antenna and its earth connection at various frequencies and at various times of the day; and especially at the twilight periods. These measurements were carried out by two distinct and trustworthy methods. The details of the determinations need not be given here; it is sufficient to state that the results leave no doubt that the local earth resistance does not vary in the rapid manner, nor to the wide extent, that is required to explain the twilight minima of the strays.

Another point that supports the view that the stray minimum is not produced by strictly local causes may be stated here. It is very noticeable that the twilight minima are easier to observe when the receiving apparatus is adjusted for long waves (say 6000 metres) than when it is adjusted for short waves (say 1000 metres). This is of obvious significance when taken with the fact that signal waves of great length are most suitable for communication over great distances. Now the most striking difference between signals received from long and from short distances is that in the former case the curvature of the earth may exert appreciable, perhaps very great, influence on the signal intensity.

This immediately raises the question of the manner of propagation of long electric waves round the earth. The fact is now thoroughly established that signal-bearing waves can travel a quarter of the way round the globe and still be easily perceptible. That this is not a manifestation of ordinary

---

the atmosphere, which, on account of their doubtful origin, have been called "X's."

Although the mathematical theory of electric wave propagation through space was worked out by Clerk Maxwell more than 50 years ago, and notwithstanding all the experimental evidence obtained in laboratories concerning the nature of these waves, yet so far we understand but incompletely the true fundamental principles concerning the manner of propagation of the waves on which wireless telegraph transmission is based. For example, in the early days of wireless telegraphy it was generally believed that the curvature of the earth would constitute an insurmountable obstacle to the transmission of electric waves between widely separated points. For a considerable time not sufficient account was taken of the probable effect of the earth connection, especially in regard to the transmission of oscillations over long distances.

Physicists seemed to consider for a long time that wireless telegraphy was solely dependent on the effects of free Hertzian radiation through space, and it was years before the probable effect of the conductivity of the earth was considered and discussed.

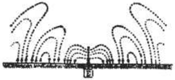

FIG. 11.

Lord Rayleigh, in referring to transatlantic radiotelegraphy, stated in a paper read before the Royal Society in May, 1903, that the results which I had obtained in signaling across the Atlantic suggested "a more decided bending or diffraction of the waves round the protuberant earth than had been expected," and further said that it imparted a great interest to the theoretical problem.[1] Prof. Fleming, in his book on electric wave telegraphy, gives diagrams showing what may be taken to be a diagrammatic representation of the detachment of semiloops of electric strain from a simple vertical wire (fig. 11).

As will be seen, these waves do not propagate in the same manner as does free radiation from a classical Hertzian oscillator, but instead glide along the surface of the earth.

Prof. Zenneck[2] has carefully examined the effect of earthed receiving and transmitting aerials, and has endeavored to show mathematically that when the lines of electrical force, constituting a wave front, pass along a surface of low specific inductive capacity—such as the earth—they become inclined forward, their lower ends being retarded by the resistance of the conductor, to which they are attached. It therefore would seem that wireless telegraphy as at present practiced is, to some extent at least, dependent on the conductivity of the earth, and that the difference in operation across long distances of sea compared to over land is sufficiently explained by the fact that sea water is a much better conductor than is land.

---

[1] Proc. Roy. Soc., vol. 72, p. 40.

[2] "Annalen der Physik," vol. 23, p. 846, "Physikalische Zeitschrift," 1906, pp. 50, 553.

The importance or utility of the earth connection has been sometimes questioned, but in my opinion no practical system of wireless telegraphy exists where the instruments are not in some manner connected to earth. By connection to earth I do not necessarily mean an ordinary metallic connection as used for wire telegraphs. The earth wire may have a condenser in series with it, or it may be connected to what is really equivalent, a capacity area placed close to the surface of the ground. It is now perfectly well known that a condenser, if large enough, does not prevent the passage of high-frequency oscillations, and therefore in this case, when a so-called balancing capacity is used, the antenna is for all practical purposes connected to earth.

I am also of opinion that there is absolutely no foundation in the statement which has recently been repeated to the effect that an earth connection is detrimental to good tuning, provided of course that the earth is good.

Certainly, in consequence of its resistance, what electricians call a bad earth will damp out the oscillations, and in that way make tuning difficult; but no such effect is noticed when employing an efficient earth connection.

In conclusion, I believe that I am not any too bold when I say that wireless telegraphy is tending to revolutionize our means of communication from place to place on the earth's surface. For example, commercial messages containing a total of 812,200 words were sent and received between Clifden and Glace Bay from May 1, 1910, to the end of April, 1911; wireless telegraphy has already furnished means of communication between ships and the shore where communication was before practically impossible. The fact that a system of imperial wireless telegraphy is to be discussed by the imperial conference, now holding its meetings in London, shows the supremely important position which radiotelegraphy over long distances has assumed in the short space of one decade. Its importance from a commercial, naval, and military point of view has increased very greatly during the last few years as a consequence of the innumerable stations which have been erected, or are now in course of construction, on various coasts, in inland regions, and on board ships in all parts of the world. Notwithstanding this multiplicity of stations and their almost constant operation, I can say from practical experience that mutual interference between properly equipped and efficiently tuned instru-

---

may be regarded as very small in itself. But, in the present case, the result is so completely inexplicable by the ordinary conceptions of the propagation of electric waves through the atmosphere, that we are compelled by its refusal to fit into the accepted scheme of things to attempt an extension of that scheme. Now, in searching for an explanation of these twilight minima, we have to notice that there are two main alternative possibilities. In the first place, the atmospheric discharges that produce the strays may themselves, for some reason, become temporarily infrequent at twilight; and, in the second place, the space through which the waves travel may become temporarily less easily traversed in twilight. The former alternative, when taken in conjunction with the author's experience over nearly 20° of longitude that it is the receiving station's local time which is concerned, implies that the bulk of the strays received at a given station are produced by atmospheric discharges occurring in regions of the atmosphere that have the same sunset and sunrise as that station; or, in other words, implies that the strays observed have their origin at places on (roughly) the meridian of the receiving station. Evidence has been already quoted, however, to show that the larger proportion of the strays observed at any station usually originate at a great distance from the station; hence the alternative under discussion leads to the unlikely conclusion that every station receives its strays from atmospheric discharges occurring at a great distance along its own meridian, and from that place solely.

We turn, therefore, to the other alternative, that the propagation of electric waves towards any place is hindered by some unknown effect of twilight. That this latter alternative is really the correct one is strongly indicated by observations (to be described later) on artificial electric waves coming from considerable distances.

Electric waves travelling near the earth's surface might conceivably be affected by the presence or absence of the sun at a point of observation in two obvious ways: First, light may ionise the air locally in sufficient degree to cause considerable absorption of the energy of the waves; and, secondly, sunlight might affect the electrical resistance of the ground within, say, 50 miles round the antenna, and thus have an effect on the absorption of the waves when they run over this region. Considering the former hypothesis first, the atmospheric absorption by ionisation due to solar radiation must, if it exists, be greater in strong sunshine than during twilight, which is contrary to observation. In point of fact, the absorption arrived at in this manner is, as will be shown incidentally later, much too small at any time of day to produce the effects actually observed. Moreover, it has long been known that the propagation of electric waves over short distances is aggregate area of the representative marks in any convenient intervals of time. The value of this time integral taken over, say, two minutes, is treated as the ordinate corresponding to the middle moment of the interval, which is taken as abscissa and a smooth curve drawn. But an even rougher method, consisting of counting the number of marks in each two minutes' interval, and using these numbers as ordinates corresponding to the mid-times, may suffice in many cases. Of course, a theoretically more perfect procedure would be to pass the detector currents through a sensitive galvanometer possessing a heavy moving system. A time record of the deflections indicated by the instrument would, at first sight, give a better quantitative result than the hand-made record just described; but actual trials show that there is an increase of accuracy only when the strays are very numerous, more numerous, in fact, than on the average occasion. Adopting, then, the method already described, we obtain curves such as that of fig. 6. These observations were made at the author's laboratory in London.

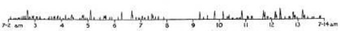

FIG. 2.—Strays, November 14, 1909. Sun rises at 7.17 A.M.

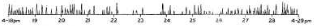

FIG. 3.—Strays, November 13, 1911. Sun sets at 4.13 P.M.

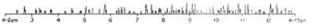

FIG. 4.—Strays, November 23, 1911. Sun sets at 4.0 P.M.

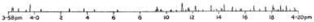

FIG. 5.—Strays. December 9, 1911. Sun sets at 3.50 P.M.

The scientific value or importance of an isolated result like the present one

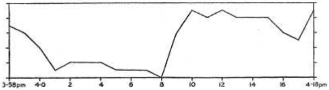

FIG. 6.—Integral Intensity Curve, Sunset, December 9, 1911.

ments has so far been almost entirely absent. Some interference does without doubt take place between ships, in consequence of the fact that the two wave lengths adopted in accordance with the rules laid down by the international convention, are not sufficient for the proper handling of the very large amount of messages transmitted from the ever increasing number of ships fitted with wireless telegraphy. A considerable advantage would be obtained by the utilization of a third and longer wave to be employed exclusively for communication over long distances.

In regard to the high-power transatlantic stations, the facility with which interference has been prevented has to some extent exceeded my expectations. At the receiving station situated at a distance of only 8 miles from the powerful sender at Clifden, during a recent demonstration arranged for the Admiralty, messages could be received from Glace Bay without any interference from Clifden when this latter station was transmitting at full power on a wave length differing only 25 per cent from the wave radiated from Glace Bay, the ratio between the maximum recorded range of Clifden and 8 miles being in the proportion of 750 to 1.

Arrangements are being made to permanently send and receive simultaneously at these stations, which, when completed, will constitute in effect the duplexing of radiotelegraphic communication between Ireland and Canada.

The result which I have last referred to also goes to show that it would be practicable to operate at one time, on slightly different wave lengths, a great number of long-distance stations situated in England and Ireland without danger of mutual interference.

The extended use of wireless telegraphy is principally dependent on the ease with which a number of stations can be efficiently worked in the vicinity of each other.

Considering that the wave lengths at present in use range from 200 to 23,000 feet, and moreover that wave group tuning and directive systems are now available, it is not difficult to foresee that this comparatively new method of communication is destined to fill a position of the greatest importance in facilitating communication throughout the world.

Apart from long-distance work, the practical value of wireless telegraphy may perhaps be divided into two parts, (1) when used for transmission over sea and (2) when used over land.

Many countries, including Italy, Canada, and Spain, have already supplemented their ordinary telegraph systems by wireless-telegraphy installations, but some time must pass before this method of communication will be very largely used for inland purposes in Europe generally, owing to the efficient network of land lines already existing which render further means of communication unnecessary; and

therefore it is probable that, at any rate for the present, the main use of radiotelegraphy will be confined to extra-European countries, in some of which climatic conditions and other causes absolutely prohibit the efficient maintenance of land-line telegraphy. A proof of this has been afforded by the success which has attended the working of the stations recently erected in Brazil on the upper Amazon.

By the majority of people the most marvelous side of wireless telegraphy is perhaps considered to be its use at sea. Up to the time of its introduction, ships at any appreciable distance from land had no means of getting in touch with the shore throughout the whole duration of their voyage. But those who now make long sea journeys are no longer cut off from the rest of the world; business men can continue to correspond at reasonable rates with their offices in America or Europe; ordinary social messages can be exchanged between passengers and their friends on shore; a daily newspaper is published on board most of the principal liners, giving the chief news of the day. Wireless telegraphy has on more than one occasion proved an invaluable aid to the course of justice—a well-known instance of which is the arrest, which took place recently through its agency, of a notorious criminal when about to land in Canada.

The chief benefit, however, of radiotelegraphy lies in the facility which it affords to ships in distress of communicating their plight to neighboring vessels or coast stations; that it is now considered indispensable for this reason is shown by the fact that several governments have passed a law making a wireless-telegraph installation a compulsory part of the equipment of all passenger boats entering their ports.

---

connected to an earth plate. The inductance was usually about 6×10^6 cm., and the natural period of the antenna was about 50,000 per second—which corresponds to a wave-length of about 6000 metres. This frequency was selected merely because it was of the same order of magnitude as that of the Marconi Transatlantic stations, signals from which were also being measured from time to time. The antenna inductance engaged inductively with another coil forming part of a secondary circuit that could be tuned to the antenna by aid of a variable condenser in it. A detector and telephone were connected with the circuit in the ordinary way. The Pickard zincite-chalcopyrite detector was used.

With such apparatus several ways of making and recording observations present themselves. The easiest and most obvious method consists in listening at sunrise and sunset to the noises made in the telephone by the natural electric wave trains; other methods involve the use of a galvanometer in place of a telephone receiver. The observations in this paper were practically all made with the telephone. A careful listener will find that on a typical morning the following phenomena appear:—First, starting to listen about half an hour before sunrise, the strays heard in the telephone are loud and numerous and much as they have been all night; then about 15 minutes before sunrise a change sets in, the strays get weaker and fewer rather quickly, till at about 10 minutes before sunrise a distinct lull occurs, of perhaps a minute's duration. At this period there is sometimes complete silence. Then the strays begin to appear again, and within 10 minutes of the lull they have settled down to the steady stream proper to the daytime. These day strays are weaker and fewer than the night strays, except on rare occasions. The lull is sometimes very pronounced, and at other times there is no lull at all. It is usually more marked at sunset than at sunrise.

The simplest way of representing these events, just as they are heard in the telephone, is by a hand-written record of the sounds. With practice, it becomes easy to make pencil marks on paper ruled in convenient units of time in such a way that the height of the mark represents the intensity of the sound, and the general shape of the mark the duration and character of the sound. Some of the records that have been obtained in this way are reproduced in figs. 2 to 5. These are selected out of a large number of records as examples of the different kinds of minima obtained at both sunrise and sunset. It will be noticed that at sunset the minimum is about 10 minutes after the calendar time of setting. Records such as these can be made to yield quantitative results of sufficient accuracy for the discussion of so irregular a phenomenon, by plotting rough estimates of the time integral of the intensity of the strays, that is, by estimating the radius of two or three hundred miles of the receiving station.) During the winter months, on the contrary, the number and intensity of the strays are relatively regular. The study of the phenomena belonging to the strays of distant origin may clearly be more favourably pursued in the winter than in summer, since the confusing feature of local lightning discharge is absent in winter. Besides the seasonal variations in the number and the intensity of the strays there is at every station a well-marked diurnal variation. Leaving out the irregularities due to local storms, we may say that the strays are in general more frequent and numerous during the night hours than during the day hours. The variations may be represented graphically as a curve in any of the methods indicated below and then they appear as shown in fig. 1, which may be regarded as a typical 24 hours' continuous record of the integral of number and intensity. These diurnal variations have not yet been investigated thoroughly. The author has not been able to find any account of observations in which the effects of local storms have been eliminated.

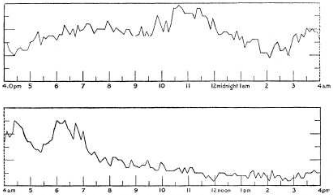

Fig. 1.—Twenty-four hours' Continuous Record of Integral Intensity of Strays, November 14 and 15, 1910.

From what has been said it is plain that the most interesting parts of the diurnal curve are those at which day and night conditions meet and change one into the other. To investigate these important parts of the curve the ordinary apparatus of a wireless telegraph station may be utilised. The apparatus employed by the author consisted of an antenna of 12 wires sloping from south to north at an angle of about 45° from a height of 170 feet, connected at the bottom to a coil of variable inductance, which was in turn

---

Some Quantitative Experiments in Long-Distance Radiotelegraphy

By: L.W. Austin, 1911

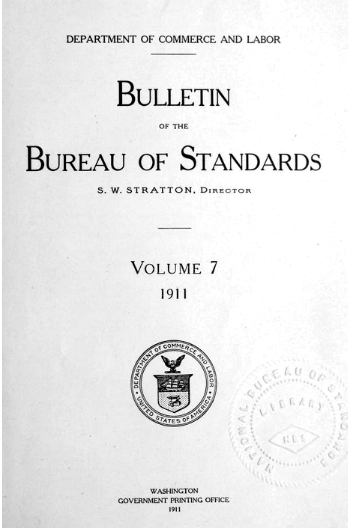

DEPARTMENT OF COMMERCE AND LABOR

BULLETIN
OF THE
BUREAU OF STANDARDS

S. W. STRATTON, DIRECTOR

VOLUME 7
1911

WASHINGTON
GOVERNMENT PRINTING OFFICE
1911

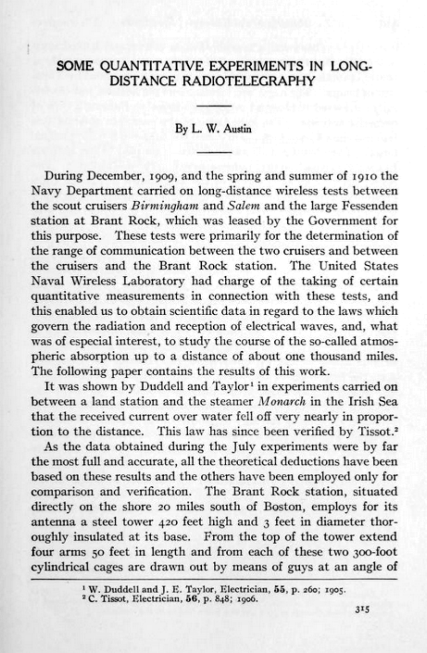

SOME QUANTITATIVE EXPERIMENTS IN LONG-DISTANCE RADIOTELEGRAPHY

By L. W. Austin

During December, 1909, and the spring and summer of 1910 the Navy Department carried on long-distance wireless tests between the scout cruisers *Birmingham* and *Salem* and the large Fessenden station at Brant Rock, which was leased by the Government for this purpose. These tests were primarily for the determination of the range of communication between the two cruisers and between the cruisers and the Brant Rock station. The United States Naval Wireless Laboratory had charge of the taking of certain quantitative measurements in connection with these tests, and this enabled us to obtain scientific data in regard to the laws which govern the radiation and reception of electrical waves, and, what was of especial interest, to study the course of the so-called atmospheric absorption up to a distance of about one thousand miles. The following paper contains the results of this work.

It was shown by Duddell and Taylor[1] in experiments carried on between a land station and the steamer *Monarch* in the Irish Sea that the received current over water fell off very nearly in proportion to the distance. This law has since been verified by Tissot.[2]

As the data obtained during the July experiments were by far the most full and accurate, all the theoretical deductions have been based on these results and the others have been employed only for comparison and verification. The Brant Rock station, situated directly on the shore 20 miles south of Boston, employs for its antenna a steel tower 420 feet high and 3 feet in diameter thoroughly insulated at its base. From the top of the tower extend four arms 50 feet in length and from each of these two 300-foot cylindrical cages are drawn out by means of guys at an angle of

---

[1] W. Duddell and J. E. Taylor, Electrician, **55**, p. 260; 1905.
[2] C. Tissot, Electrician, **56**, p. 848; 1906.

Proceedings of the Royal Society of London 1912

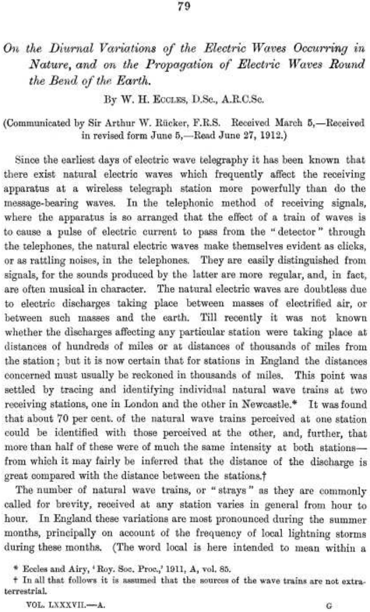

On the Diurnal Variations of the Electric Waves Occurring in Nature, and on the Propagation of Electric Waves Round the Bend of the Earth.

By W. H. Eccles, D.Sc., A.R.C.Sc.

(Communicated by Sir Arthur W. Rücker, F.R.S. Received March 5,—Received in revised form June 5,—Read June 27, 1912.)

Since the earliest days of electric wave telegraphy it has been known that there exist natural electric waves which frequently affect the receiving apparatus at a wireless telegraph station more powerfully than do the message-bearing waves. In the telephonic method of receiving signals, where the apparatus is so arranged that the effect of a train of waves is to cause a pulse of electric current to pass from the "detector" through the telephones, the natural electric waves make themselves evident as clicks, or as rattling noises, in the telephones. They are easily distinguished from signals, for the sounds produced by the latter are more regular, and, in fact, are often musical in character. The natural electric waves are doubtless due to electric discharges taking place between masses of electrified air, or between such masses and the earth. Till recently it was not known whether the discharges affecting any particular station were taking place at distances of hundreds of miles or at distances of thousands of miles from the station; but it is now certain that for stations in England the distances concerned must usually be reckoned in thousands of miles. This point was settled by tracing and identifying individual natural wave trains at two receiving stations, one in London and the other in Newcastle.* It was found that about 70 per cent. of the natural wave trains perceived at one station could be identified with those perceived at the other, and, further, that more than half of these were of much the same intensity at both stations—from which it may fairly be inferred that the distance of the discharge is great compared with the distance between the stations.†

The number of natural wave trains, or "strays" as they are commonly called for brevity, received at any station varies in general from hour to hour. In England these variations are most pronounced during the summer months, principally on account of the frequency of local lightning storms during these months. (The word local is here intended to mean within a

---

* Eccles and Airy, 'Roy. Soc. Proc.,' 1911, A, vol. 85.

† In all that follows it is assumed that the sources of the wave trains are not extra-terrestrial.

7 to 30 amperes antenna heights 37 to 130 feet, wave lengths from 300 to 3750 meters, and distances up to 1000 nautical miles.

In regard to the value of the day absorption it is only possible to say that the above expression is satisfied within the limits of error of the observations. It is quite possible that when observations are made at distances of 2000 to 3000 miles, the value of the absorption coefficient will have to be corrected by 10 or even 20 per cent, as this amount of error could exist without discovery at the distances covered in these experiments. It is also possible that the square root law relating the absorption with the wave length is only an approximation.

U. S. Naval Wireless Telegraphic Laboratory,
Washington, February 1, 1911.

---

about 45°. This forms a system of eight conductors spaced symmetrically about the tower. These cages are approximately 4 feet in diameter and consist of four wires each, separated by a system of hoops. The cages are insulated at the bottom and electrically connected to the steel tower at its top, thus forming a type of umbrella antenna. The inductance of the complete antenna system amounts to 0.055 millihenry, and the capacity is 0.0073 microfarad. The closed circuit was inductively coupled to the antenna just loosely enough so that but one wave length[3] was radiated. The closed circuit condensers were of the well-known Fessenden compressed-air type, and for a wave length of 3750 meters the capacity was 0.18 microfarad. The spark gap was of the Fessenden synchronous rotary type with two fixed electrodes and a system of rotary electrodes mounted on the shaft of the generator. The generator was rated at 100 kilowatts at 500 cycles, thus producing 1000 sparks per second. It was not possible to run the generator at full power with the steam engine available at the time of the experiments, the actual power expended being 50 to 60 kilowatts.

The scout cruisers were provided with flat top antennas supported by steel masts surmounted by wooden topmasts. The antenna dimensions were: Length 116 feet, breadth 40 feet, number of longitudinal wires 14, cross wires 15. During the December test, the height above the water was approximately 112 feet, but this height was increased in the three later tests to 130 feet. The capacity was about 0.0018 microfarad and the inductance 0.038 millihenry. During the May experiments the capacity of the *Birmingham* was increased by two 70-foot cages forward and two 80-foot cages aft set at an angle of approximately 30° with the vertical. These cages were similar to those forming the umbrella of the Brant Rock tower. This increased the capacity to about 0.0025. In the midst of the July tests, cages were added to the *Salem* also, but the change resulted in no observable variation in the intensity of the signals, the increase in capacity being apparently counterbalanced by the decrease in effective height. The closed circuit was coupled to the antenna, so that but one wave length was radiated, and the closed circuit capacity was 0.036 microfarad at 3750 meters, and 0.018 microfarad at 1000 meters wave length. The motor generators were 500-cycle machines rated at 10 kilowatts, and the spark gap was of the same type as that used at Brant Rock but of smaller dimensions.

---

[3] Experience has shown that when the closed circuit is properly designed—that is, without undue waste of energy from brushing, etc.—the greatest range is obtained by a coupling loose enough to give but one wave length in the antenna.

JULY EXPERIMENTS

I_s. Sending antenna current in amperes.
I_R. Received antenna current through 25 ohms.
N. Night observations.
K. Received antenna current at 1 kilometer distance.
d. Distance in kilometers.
h. Height of flat top antenna.
α. Day absorption coefficient = 0.0015.
λ. Wave length.
H. Stat. Heavy static = atmospheric discharges preventing reception of signals.
Miles. Nautical mile = 1.85 kilometers.
For good communication. $I_R = 40 \times 10^{-6}$ amperes through 25 ohms = 4×10^{-8} watts.
For audible signals. $I_R = 10 \times 10^{-6}$ amperes through 25 ohms = 2.5×10^{-9} watts.

Preliminary experiments were begun early in July, the scout cruisers lying at anchor in the harbor at Provincetown, 22 miles from Brant Rock. The path of the waves between the ships and the shore station, except for a narrow strip of sand hills bounding Provincetown Harbor, lay entirely over water. Later, when the ships put to sea, it was found that this strip of land made no appreciable difference in the received signals. Careful measurements were made on the intensity of the received antenna current at Brant Rock for the two wave lengths 1 000 meters and 3 750 meters. This current was measured either by means of a hot-wire ammeter of 15 ohms resistance, provided with a mirror for mirror and scale readings,[4] or by a tellurium constantan thermo-element,[5] also of 15 ohms resistance. The hot wire ammeter gave a scale deflection of about 0.2 millimeter for 1 milliampere oscillatory current in the antenna and 9.4 millimeters for 10 milliamperes. One millimeter galvanometer deflection with the thermo-element was equivalent to 263 microamperes, and the deflections

---

[4] L. W. Austin, Electrical World, **49**, p. 308; 1907.

[5] This Bulletin, **7**, p. 304.

---

heights of from 30 to 80 feet and wave lengths from approximately 1500 to 4000 meters.

(*c*) Taking account of the influence of antenna height and wave length equation (2) may be extended and a general day transmission formula written as follows

$$(3) \quad I_R = 4.25 \frac{I_s\, h_1\, h_2}{\lambda d} e^{-\frac{0.0015\, d}{\sqrt{\lambda}}}$$

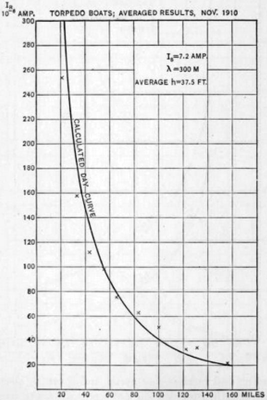

Fig. 28

where the currents are given in amperes and all lengths in kilometers.[19] This formula has been tested for sending currents from

---

[19] From this it would appear that it is advisable to rate stations according to the magnitude of the antenna current, or perhaps better, according to the product of the current into the height.

Variations also appear to occur during the daytime, but these are probably in general small.

(*b*) The received antenna currents between two stations with salt water between are proportional to the product of the heights[18] of the sending and receiving antennas and inversely proportional to the wave length, provided the antenna resistances remain constant. These experiments were carried on with flat-top antenna

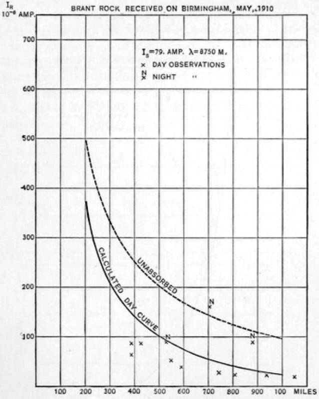

Fig. 27

---

[18] The experiments indicate that for the greatest efficiency of a flat-top antenna the vertical lead wires should be bunched so as to reduce their capacity as much as possible and concentrate the capacity at the greatest height.

---

were, as nearly as could be determined, directly proportional to the square of the oscillatory current. For the 1000-meter wave length a variable air condenser was placed in series with the Brant Rock antenna, and in this way it was possible to adjust the antenna resistance for both wave lengths to about 25 ohms. From a number of observations it was determined that the received current at Brant Rock for a wave length of 1000 meters was 10 500 microamperes for the *Birmingham*, with a sending antenna current of 33 amperes,[6] and 11 000 microamperes for the *Salem*, corresponding to a sending antenna current of 27 amperes. For a wave length of 3750 meters the received currents were 3200 microamperes for the *Birmingham* and 4100 microamperes for the *Salem*, corresponding to sending antenna currents of 27 and 24 amperes.

After the ships had attained distances too great to permit the reading of the received signals at Brant Rock by means of thermoelements, a chalcopyrite zinkite rectifier with galvanometer was employed whenever the atmospheric disturbances permitted. This was connected in a secondary circuit coupled to the antenna and calibrated by means of a thermoelement in the antenna and an exciting buzzer circuit which could be tuned to the wave length used.[7]

Throughout the experiments, both at Brant Rock and on shipboard, shunted telephone readings were taken on the incoming signals. The detectors used for the shunt readings were of the free wire electrolytic type and the head telephones connected in series were each of 600 ohms resistance,[8] the shunt being placed across one of these according to the Fessenden method. The circuits for the galvanometer and for the shunt readings at Brant Rock are shown in Fig. 1.

An extended comparison of antenna currents through circuit A and the shunt readings as taken in circuit B was made, and the

---

[6] It will be noticed that while the radiation current of the *Birmingham* is considerably greater than that of the *Salem*, the strength of received signal is slightly less. It is possible that the apparently greater radiation of the *Birmingham* was due to a deck insulator which gave trouble throughout the July test. It is not improbable, therefore, that the actual antenna current was not greater than that of the *Salem*. The mean of the readings of the two ships, 30 amperes at a wave length of 1000 meters, has been taken as a basis for calculation.

[7] See this Bulletin, **7**, p. 295.

[8] The inductive resistance of each telephone used in calculating the shunt ratio was 2000 ohms.

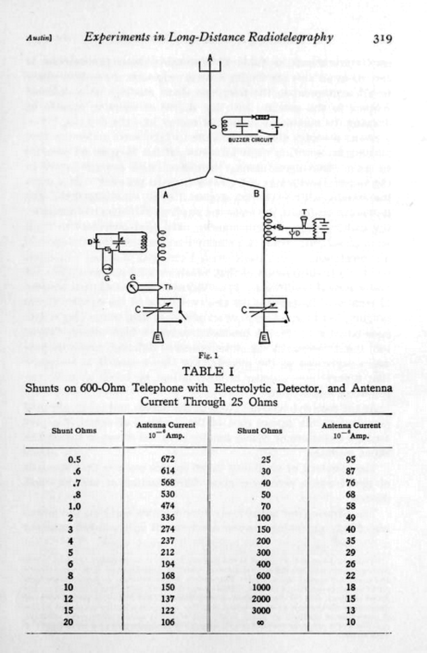

A
BUZZER CIRCUIT
A
B
T
D
D
G
G
Th
C
C
E
E

Fig. 1

TABLE I

Shunts on 600-Ohm Telephone with Electrolytic Detector, and Antenna Current Through 25 Ohms

Shunt Ohms	Antenna Current 10^{-6} Amp.	Shunt Ohms	Antenna Current 10^{-6} Amp.
0.5	672	25	95
.6	614	30	87
.7	568	40	76
.8	530	50	68
1.0	474	70	58
2	336	100	49
3	274	150	40
4	237	200	35
5	212	300	29
6	194	400	26
8	168	600	22
10	150	1000	18
12	137	2000	15
15	122	3000	13
20	106	∞	10

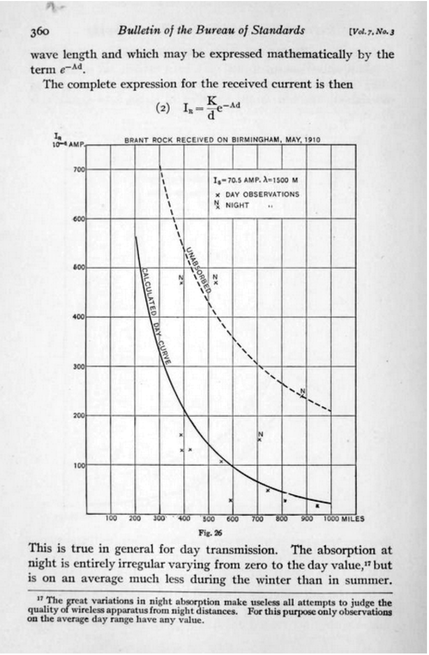

wave length and which may be expressed mathematically by the term e^{-Ad}.

The complete expression for the received current is then

$$(2)\quad I_R = \frac{K}{d} e^{-Ad}$$

I_R 10^{-6} AMP.
BRANT ROCK RECEIVED ON BIRMINGHAM, MAY, 1910
700
600
500
400
300
200
100
I_s = 70.5 AMP. λ = 1500 M
× DAY OBSERVATIONS
N× NIGHT "
UNABSORBED
CALCULATED DAY CURVE
N
100
200
300
400
500
600
700
800
900
1000 MILES

Fig. 26

This is true in general for day transmission. The absorption at night is entirely irregular varying from zero to the day value,[17] but is on an average much less during the winter than in summer.

---

[17] The great variations in night absorption make useless all attempts to judge the quality of wireless apparatus from night distances. For this purpose only observations on the average day range have any value.

SUMMARY AND CONCLUSIONS

Quantitative measurements have been carried out in long distance wireless telegraphy up to 1000 miles for the purpose of determining the law of the variation of strength of signal with distance.

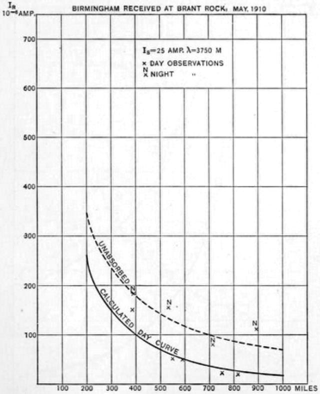

Fig. 25

Supplementary work was also done on the effect of height of antenna and wave length on sending and receiving.

The results are given briefly as follows:

(*a*) Over salt water the electrical waves decrease in intensity in proportion to the distance as found by Duddell and Taylor. In addition they are subject to an absorption which varies with the

---

results are shown in Table I. The table shows the amount of current in an antenna having a total resistance of 25 ohms which would correspond to the telephone shunt readings on a detector coupled to the antenna with the degree of coupling capable of drawing the maximum amount of energy into the detector.[9]

Shunt readings are very difficult to take even under the best circumstances. In a quiet laboratory it has been found possible to make them agree among themselves with average errors in the single observations not greater than 10 per cent. In a wireless station, with its varying degree of noise, especially if the station is on shipboard, the errors are much greater than this, amounting under ordinary circumstances with good operators to from 20 to 40 per cent; while in a stormy time, or when the atmospheric discharges are heavy, the observed strength of signal will often be only a small fraction of that which would have been observed under normal conditions. It is evident that most of these sources of error tend to decrease the observed value of the signal. Occasionally, however, a loose contact or a loose plug where plug resistance boxes are used, will produce abnormally high values. Then, too, the difference in ear sensitiveness of different operators produces variations in the record. For this reason it is necessary that a very large number of observations be taken to make the results of much value.

As the received currents given in the curves and tables are proportional to the square root of the loudness of signal, the percentage errors are of course smaller than the errors in taking the shunt readings.

The standard of audibility taken in this work is that strength of signal which permits a clear differentiation of the dots and dashes.

The *Birmingham* sailed from Provincetown early on the morning of July 14, taking a course nearly south until she had attained a distance of about 1200 miles from Brant Rock. The *Salem*

---

[9] The actual amount of current in the antenna for the same intensity of signal in the detector varies with the effective resistance, which in turn depends on the wave length and amount of tuning inductance. The energy in the detector, required to produce given signals is constant, however. This energy is what is actually measured in taking shunt readings, and for convenience of comparison its square root is put in terms of micro amperes through 25 ohms which was the resistance actually employed in the thermoelement readings at Brant Rock, and which is approximately the effective resistance of a properly coupled ship's antenna of moderate size at a wave length of 1000 meters.

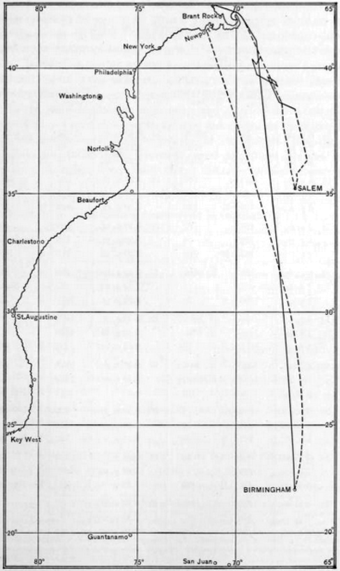
80°
75°
70°
65°
40°
35°
30°
25°
20°
Brant Rock
Newport
New York
Philadelphia
Washington
Norfolk
Beaufort
Charleston
St. Augustine
Key West
Guantanamo
San Juan
SALEM
BIRMINGHAM

Fig. 2

---

curve as calculated from Table XVI for antenna heights of 37.5 feet, a sending current of 7.2 amperes, and a wave length of 300 meters. The agreement is remarkable, considering the fact that the table from which the theoretical curve was calculated was based on data derived from ships with 130-foot masts working at a wave length of 1000 meters and employing a spark frequency of 1000 per second. These results show, as has often been observed in our laboratory, how little the ratio of sending currents and received signals depends on the spark frequency, the telephones apparently never showing the added sensitiveness for high frequencies exhibited in the case of sinusoidal alternating currents.

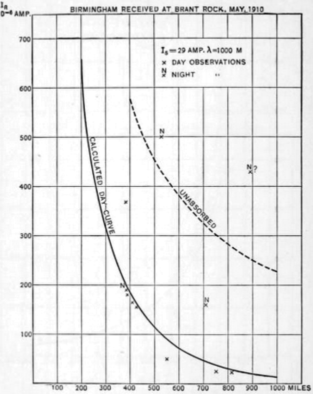

Fig. 24

mately 1.12. The telephones on both boats were of approximately the same sensitiveness and were of 2000 ohms resistance with an inductive resistance under working conditions of about 4000 ohms.

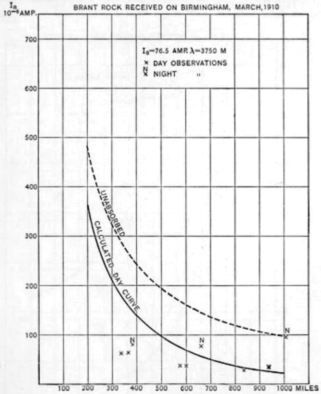

Fig. 23

The observations of these tests, shown in Table XVII, were taken with great care, the engines being stopped during the time of observation. Each one is the mean of a number of readings.

In the last column are given the received currents averaged for both boats corrected to a sending current of 7.2 amperes and to the readings of the electrolytic detector. In Fig. 28 is shown the

---

TABLE II

Birmingham Received at Brant Rock

July		$\lambda=1000$ M. $I_s=33$ amp.				July		$\lambda=3750$ M. $I_s=27$ amp.			
		Distance, Nautical Miles		Received Current 10^{-6} amp.				Distance, Nautical Miles		Received Current 10^{-6} amp.	
14	8 a. m.	22		10500	Galvanometer	14	8.30 a. m.	22		3000	Galvanometer
	2 p. m.	50		4600			3.30 p. m.	60		1370	
	8 p. m.	110	N	2100			9.30 p. m.	120	N	520	
15	8 a. m.	256		520		15	9.30 a. m.	266		180	
	2 p. m.	335		157			3.30 p. m.	345		67	
	8 p. m.	415	N	336			9.30 p. m.	425	N	77	
16	8 a. m.	550		77		16	9.30 a. m.	560		49	
	2 p. m.	626		54			3.30 p. m.	636		H. Stat.	
	8 p. m.	702	N	210			9.30 p. m.	712	N	106	
17	8 a. m.	854		H. Stat.		17	9.30 a. m.	864		26	
	2 p. m.	927					3.30 p. m.	937		Stat.	
	8 p. m.	1000	N	19			9.30 p. m.	1010	N	10	
18	8 a. m.	1170		H. Stat.		18	9.30 a. m.	1170			
	2 p. m.	1185		H. Stat.			3.30 p. m.	1185			
	8 p. m.	1200	N	51			9.30 p. m.	1200	N	44	
19	8 a. m.	1152		H. Stat.		19	9.30 a. m.	1152		26	
	2 p. m.	1103		H. Stat.			3.30 p. m.	1103			
	8 p. m.	1035	N	210			9.30	1025	N	Not send	
20	8 a. m.	890		35		20	9.30 a. m.	880		26	
	2 p. m.	841		29			3.30 p. m.	831			
	8 p. m.	772	N	240			9.30 p. m.	762	N	Not send	
21	8 a. m.	648		26		21	9.30 a. m.	638		26	
	2 p. m.	559					3.30 p. m.	549			
	8 p. m.	470	N	150			9.30	460	N	Not send	
22	8 a. m.	320		210		22	9.30 a. m.				
	2 p. m.	240					3.30 p. m.				
	8 p. m.	190					9.30 p. m.				

followed on the morning of July 15 and proceeded slowly to a point about 450 miles from Brant Rock. The courses of the two ships are shown in figure 2. Regular sending and receiving periods were observed at the three stations. At Brant Rock the signals were received as far as possible on the thermoelement and a few readings were taken with the rectifier, but while the preceding week had been almost completely free from atmospheric disturbances, unfortunately, soon after the ships sailed, strong disturbances set in which precluded the possibility of galvanometer measurements, so that for the rest of the experiments all the readings had to be taken by the shunted telephone method. On the ships no attempt was made to use deflection methods, although

TABLE III

Brant Rock Received on Birmingham

July		$\lambda=1500$ m, $I_s=56$ amp.		July		$\lambda=3750$ m, $I_s=69$ amp.	
		Distance, Nautical Miles	Received Current 10^{-6} amp.			Distance, Nautical Miles	Received Current 10^{-6} amp.
15	10.15 a. m.	290	336	15	10.45 a. m.	295	120
	10.15 p. m.	445	N 274		10.45 p. m.	450	N 63
16	10.15 a. m.	580	58	16	10.45 a. m.	585	
	10.15 p. m.	727	N 160		10.45 p. m.	732	N 77
17	10.15 a. m.	885	28	17	10.45 a. m.	890	
	10.15 p. m.	1040	N 106		10.45 p. m.	1045	N 37
18	10.15 a. m.	1180	Stat.	18	10.45 a. m.	1185	
	10.15 p. m.	1215	N 130		10.45 p. m.	1215	N 110
19	10.15 a. m.	1115	Stat.	19	10.45 a. m.	1110	30
	10.15 p. m.	1005	N 106		10.45 p. m.	1000	N 212
20	10.15 a. m.	850	37	20	10.45 a. m.	845	28
	10.15 p. m.	730	N 336		10.45 p. m.	725	Not send
21	10.15 a. m.	600	150	21	10.45 a. m.	595	51
	10.15 p. m.	435	N 194		10.45 p. m.	430	N 54
22	10.15 a. m.	275	390	22	10.45 a. m.	270	Not send
	10.15 p. m.	115	N 336		10.45 p. m.	110	Not send

TABLE XVII

	Distance, Nautical Miles	Stringham Received on Bailey $\lambda=340$ m			Bailey Received on Stringham $\lambda=280$ m			I_R Average of Received Currents Corrected to Electrolytic and to $I_S=7.2$ amp. 10^{-6} amp.
		I_S amp.	Received Current 10^{-6} amp.		I_S amp.	Received Current 10^{-6} amp.		
			Observed	Corrected to $I_S=7.2$ amp.		Observed	Corrected to $I_S=7.2$ amp.	
Nov. 3, 1910	21.5	5.45	170	225	5.4	....	....	253
	33	6.00	132	159	5.1	87	123	158
	43	6.14	68	80	5.1	86	121	112
	55	6.00	58	70	5.1	76	107	99.0
Nov. 7,	65	6.27	55	63	7.5	75	72	75.5
	83	6.54	51	56	7.5	57	55	62.2
	100	6.00	39	47	7.2	45	45	51.6
Nov. 8,	122	6.55	23.9	26	7.0	32	33	33.0
	131	6.27	24.5	28	7.0	33	34	34.8
	156	6.82	Interference		7.0	18	18.6	20.9

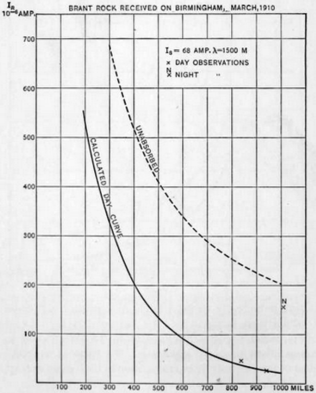

Fig. 22

The antenna of the *Stringham* consisted of four wires 2 feet apart, 135 feet long, and 48.5 feet above the water. Its capacity was 0.00134 microfarad. The closed circuit condenser was the same as that of the *Bailey*. The length of the *Stringham's* radiated wave was 340 meters. The receiving apparatus of both boats was of the latest inductively coupled type, and perikon detectors were used which were standardized by comparison with electrolytic detectors, which experiments have shown to have a constant sensitiveness when properly adjusted. The ratio of current sensitiveness of the electrolytic to the perikon in both cases was approxi-

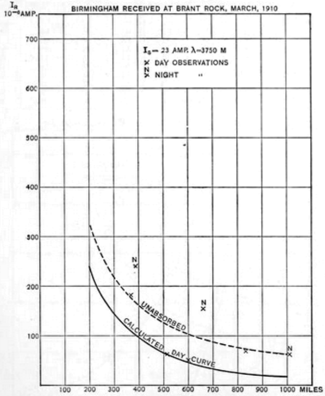

Fig. 21

---

it is possible that satisfactory measurements could have been taken by means of marine galvanometers. Tables II to VI contain the data of the experiments, the received signals being reduced to terms of current through 25 ohms antenna resistance.

TABLE IV

Salem Received at Brant Rock

July	λ=1000 m I_s=27 amp.		July	λ=3750 m I_s=23 amp.	
	Distance, Nautical Miles	Received Current 10^{-6} amp.		Distance, Nautical Miles	Received Current 10^{-6} amp.
			15 6 a. m.	22	4400
15 5 a. m.	22	11000	9 a. m.	52	1420
8 a. m.	45	5000	3 p. m.	70	1250
2 p. m.	60	3600	9 p. m.	117	N * 613
8 p. m.	117	N 2170			
16 8 a. m.	125	2250	16 9 a. m.	130	693
2 p. m.	105	2275	3 p. m.	110	* 567
8 p. m.	137	N 1328	9 p. m.	140	N * 474
17 8 a. m.	207	* 1062	17 9 a. m.	210	* 274
2 p. m.	190	646	3 p. m.	190	* 500
8 p. m.	196	N 400	9 p. m.	196	N * 179
18 8 a. m.	249	580	18 9 a. m.	250	* 237
2 p. m.	270	520	3 p. m.	275	* 274
8 p. m.	290	N 410	9 p. m.	295	N * 194
19 8 a. m.	366	200	19 9 a. m.	370	* 106
2 p. m.	394	182	3 p. m.	400	* 115
8 p. m.	425	N * 672	9 p. m.	430	Not send
20 8 a. m.	322	325	20 9 a. m.	315	Not send
2 p. m.	254	582	3 p. m.	245	Not send
8 p. m	181	N Stat.	9 p. m.	175	Not send

* Telephone readings

Wherever possible the readings taken by the deflection method are used. During the experiments the atmospheric disturbances were heavy most of the time, especially at night and at the longer wave length. These frequently entirely prevented the reception of the signals as indicated by the blanks in the received current column. No table is given for Brant Rock received on the *Salem*, because during a great part of the time the signals were too strong for accurate shunt readings. No tables are given for the work between the two cruisers at a wave length of 3 750 meters, as their antennas seemed too small to work successfully at this wave length and only a few unsatisfactory shunt readings were obtained.

TABLE V

"Salem" Received on "Birmingham"

July		$\lambda=1000$ m $I_s=27$ Amp.		
		Distance, Nautical Miles		Received Current 10^{-6} Amp.
15	8 a. m.	253		170
	2 p. m.	274		196
	8 p. m.	312	N	120
16	8 a. m.	420		87
	2 p. m.	521		54
	8 p. m.	574	N	87
17	8 a. m.	661		21
	2 p. m.	750		20
	8 p. m.	840	N	30
18	8 a. m.	960		Stat.
	2 p. m.	972		Stat.
	8 p. m.	960	N	46
19	8 a. m.	828		10
	2 p. m.	720		26
	8 p. m.	618	N	46
20	8 a. m.	589		43
	2 p. m.	566		54
	8 p. m.	555	N	69
21	8 a. m.	666		46
	2 p. m.			
	8 a. m.			

TABLE VI

"Birmingham" Received on "Salem"

July		$\lambda=1000$ m $I_s=33$		
		Distance, Nautical Mile		Received Current 10^{-6} Amp.
15	8.30 a. m.	256		160
	2.30 p. m.	278		150
	8.30 p. m.	315	N	105
16	8.30 a. m.	425		60
	2.30 p. m.	526		43
	8.30 p. m.	579	N	63
17	8.30 a. m.	666		H. Stat.
	2.30 p. m.	755		H. Stat.
	8.30 p. m.	845	N	38
18	8.30 a. m.	960		10
	2.30 p. m.	972		10
	8.30 p. m.	960	N	10
19	8.30 a. m.	823		10
	2.30 p. m.	715		Stat.
	8.30 p. m.	613	N	77
20	8.30 a. m.	584		38
	2.30 p. m.	561		38
	8.30 p. m.	555		Messages
21	8 30 a. m.	661		38

For determining the law of the decrease of the intensity of the signal with the distance a smooth curve was drawn through the observed day readings and points on this were taken as the

---

with small 120-cycle directly connected sending sets which, however, could be coupled loosely enough so that but one wave length was radiated. The flat-top antenna of the *Bailey* consisted of four wires 2 feet apart, 80 feet long, 32 feet above the water. The antenna capacity was 0.000875 microfarad; the closed circuit condenser consisted of Leyden jars in air having a capacity of 0.015 microfarad. The length of the wave radiated by the *Bailey* was 280 meters.

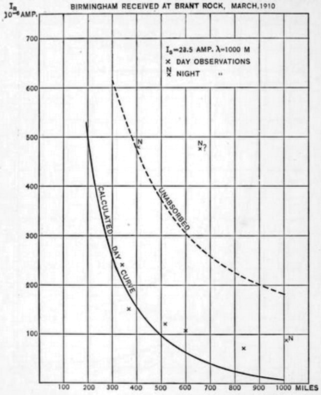

Fig. 20

of long range night communication must be considered exceptional so late in the year, though perhaps it would not be so considered in midwinter.

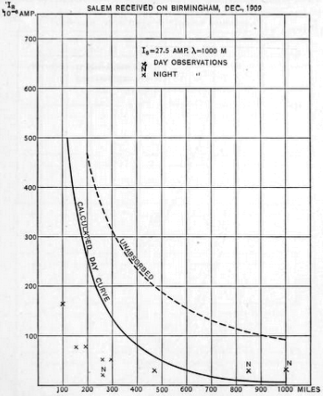

Fig. 19

TORPEDO-BOAT TESTS

In order to test the validity of equation (3) for very short wave lengths and small antennas, the torpedo boats *Stringham* and *Bailey* were detailed to carry out a series of tests in Chesapeake Bay during the month of November, 1910. The two boats were fitted

---

TABLE VII

Birmingham Received at Brant Rock, July

		$I_S=33$ amp. $\lambda=1000$ m. $K=475000\ 10^{-6}$ amp.			$I_S=27$ amp. $\lambda=3750$ m. $K=138000\ 10^{-6}$ amp.		
Nautical Miles	d Km	$\frac{K}{d}$ 10^{-6} amp.	I_R 10^{-6} amp.		$\frac{K}{d}$ 10^{-6} amp.	I_R 10^{-6} amp.	
			Obs.	Cal.		Obs.	Cal.
20	37	12830	11500	12200	3730	3500	3620
50	92.5	5135	4600	4460	1492	1400	1385
100	185	2565	2100	1940	746	700	645
200	370	1283	800	737	373	260	280
300	555	856	260	370	249	120	161
400	740	642	130	212	187	60	105
500	925	514	80	128	149	40	73
600	1110	428	50	81.5	124	30	52
800	1480	321	35	35	93	24	29
1000	1850	257	20	15.4	75	22	18

TABLE VIII

Brant Rock Received on Birmingham, July

		$I_S=56$ amp. $\lambda=1500$ m $K=310000.10^{-6}$ amp.			$I_S=69$ amp. $\lambda=3750$ m $K=160000.10^{-6}$ amp.		
Nautical Miles	d Km	$\frac{K}{d}$ 10^{-6} amp.	I_R 10^{-6} amp.		$\frac{K}{d}$ 10^{-6} amp.	I_R 10^{-6} amp.	
			Obs.	Cal.		Obs.	Cal.
20	37	8380		8000	4325		4200
50	92.5	3350		2980	1730		1610
100	185	1675		1340	865		750
200	370	838		533	432		324
300	555	558	320	283	288	120	187
400	740	419	160	170	216	85	121
500	925	335	80	108	173	65	84.5
600	1110	279	52	72.5	144	52	60.8
800	1480	209	38	34.7	108	40	34.1
1000	1850	167	25	17.1	87	30	20.6

observed values for purposes of calculation. It was assumed from the results of Duddell and Taylor and Tissot that the received currents would be inversely proportional to the distance provided no absorption existed. The observed curve indicated that this was approximately true up to a point between 100 and 200 miles, but beyond this point the currents evidently dropped much more rapidly. The simplest assumption in regard to absorption is that it is proportional to the distance. Joining this to the Duddell and Taylor law we have as an expression for the received current

$$I_R = \frac{K}{d} e^{-Ad} \qquad (1)$$

where d is the distance, K the received current at unit distance, e the base of the natural logarithms, and A a constant. Dr. Louis Cohen of the National Electric Signaling Co., while testing the validity of this formula made the discovery that A was inversely proportional to the square root of the wave length within the limits of accuracy of the observations. Then, writing $A = \frac{\alpha}{\sqrt{\lambda}}$ the expression becomes

$$I_R = \frac{K}{d} e^{-\frac{\alpha d}{\sqrt{\lambda}}} \qquad (2)$$

α is the absorption coefficient and in these experiments equals 0.0015, the distance and wave length being expressed in kilometers.

Tables VII to XI contain a comparison of the calculated values with those taken from the smoothed observation curves. It is seen that the agreement is exceedingly good in the case of the signals between the *Birmingham* and the *Salem* and those of the *Salem* received at Brant Rock. In the case of the signals between Brant Rock and the *Birmingham*, however, the observed values fall somewhat below the calculated, from 300 to 600 miles. As the observed curve has the same form both for the signals received on the *Birmingham* and those received at Brant Rock the divergence is probably not accidental, and is probably due to a temporary increase of absorption accompanying a marked change in the weather at about this time. The effect is not observed in the *Salem* signals received at Brant Rock because the *Salem* at this time had not attained a great enough distance to make the absorption play an important part.

---

and the short wave at 1305 miles. Night signals from Brant Rock were first received at 2776 miles and were heard without interruption from that point on. Both waves from the *Birmingham* were received faintly in the daytime at Brant Rock at 2090 miles, and the long wave again at 1315 miles, no other day signals being received until the distance was reduced to 815 miles. Night signals were received by Brant Rock from the *Birmingham* first at 2418 miles and continued nightly for the rest of the voyage, though very irregular in intensity. This remarkable continuation

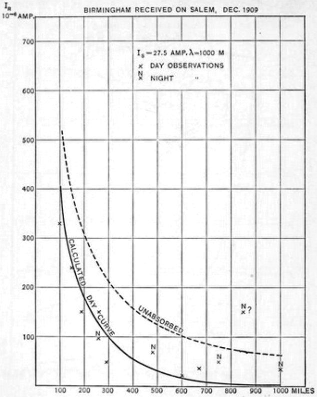

Fig. 18

errors of observation, or possibly to some kind of reflection from the upper atmosphere.[16] Beyond the range of the curves abnormal signals were again observed, faint day signals of the Brant Rock long wave being received on the *Birmingham* at 2330 miles,

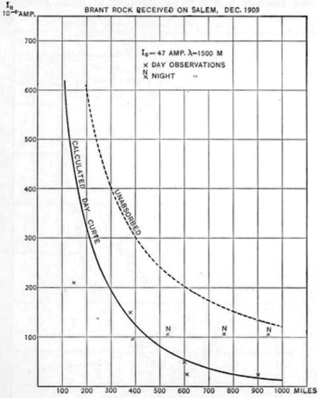

Fig. 17

---

[16] There can be no doubt that at times something of the nature of a reflection of signals from the upper atmosphere takes place. During the past winter the Norfolk station sending at a wave length of 1000 meters and with an antenna current of 10 or 12 amperes and antenna height below 200 feet was distinctly heard at night at the Mare Island Navy Yard in San Francisco, a distance of approximately 2000 nautical miles across the continent. Although it is possible that over sea water a clearing up of the atmospheric absorption would have explained this phenomenon, when we take into account the ground absorption which is probably independent of day and night and generally much greater than the atmospheric absorption, this explanation becomes untenable.

TABLE IX

"Salem" Received at Brant Rock, July

Nautical Miles	d Km	I_S = 27 Amp. λ = 1000 m K = 490000 10^{-6} Amp.			I_S = 23 Amp. λ = 3750 m K = 170000 . 10^{-6} Amp.		
		$\frac{K}{d}$ 10^{-6} Amp.	I_R 10^{-6} Amp.		$\frac{K}{d}$ 10^{-6} Amp.	I_R 10^{-6} Amp.	
			Obs.	Cal.		Obs.	Cal.
20	37	13250	12000	12500	4595	4500	4460
50	92.5	5300	4700	4610	1838	1800	1710
100	185	2650	2200	2000	919	850	800
150	278	1760	1200	1170	612	500	495
200	370	1325	750	760	459	330	344
300	555	883	370	382	306	190	198
400	740	662	200	218	230	115	129
500	925	530	130	132	184	88	90

TABLE X

Salem Received on Birmingham, July

Nautical Miles	d Km	I_s = 27 amp. λ = 1000 m. K = 230000 10^{-6} amp.		
		$\frac{K}{d}$ 10^{-6} amp.	I_R 10^{-6} amp.	
			Obs.	Cal.
20	37	6220		5880
50	92.5	2486		2160
100	185	1243		941
200	370	622	325	357
300	555	414	165	179
400	740	311	95	102
500	925	249	65	62
600	1110	207	45	39.4
800	1480	155	19	17.1
1000	1850	124	10	7.5

TABLE XI

Birmingham Received on Salem, July

Nautical Miles	d Km	I_s = 33 amp. λ = 1000 m. K = 185000 10^{-6} amp.		
		$\frac{K}{d}$ 10^{-6} amp.	I_R 10^{-6} amp.	
			Obs.	Cal.
20	37	5000		4730
50	92.5	2000		1740
100	185	1000		757
200	370	500	260	287
300	555	333	130	144
400	740	250	75	82.4
500	925	200	50	49.9
600	1110	167	35	31.7
800	1480	125	16	13.8
1000	1850	100	10	6.2

The curves of figures 3 to 10 give the results in graphic form. The dotted line gives the strength of signal which would have been received if the K/d law had obtained; that is, if there had been no absorption. The continuous curve gives the theoretical day values as calculated from equation 2, while the individual observations are represented by crosses. Observations taken by the deflection method are indicated by circles around the crosses, while night observations are accompanied by the letter N. It is seen that the day observations correspond approximately to the values of the calculated curve, but that the night signals are entirely irregular,

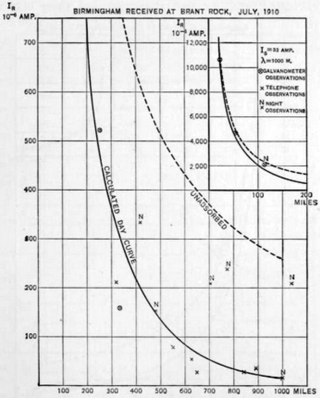

Fig. 3

Birmingham at 2550 miles, while on the same night both waves from the *Birmingham* were received at Brant Rock.

MAY EXPERIMENTS

During the month of May the *Birmingham* returned from Liberia to Hampton Roads, and the results of the telephone observations taken during this voyage are shown in figures 24 to 27.

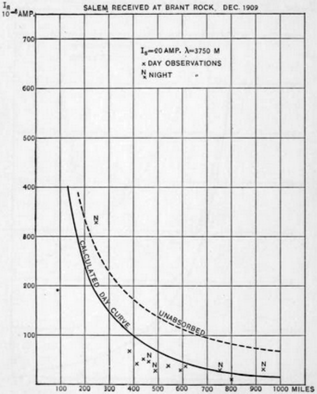

Fig. 16

Within the range covered by the curves the day observations are not as irregular as during the March voyage, although, as was also the case in March, some of the night observations lie above the curve calculated for zero absorption. As has already been said in describing the July experiments, these may be due to

In some of the curves it is seen that the observations are exceedingly irregular, the day signals showing as much variation as the night signals during the July tests. In addition to the observations shown on the curves, faint day signals from the *Birmingham* were received at Brant Rock at a distance of 1325 miles, and again at 1720 miles, while the Brant Rock short wave was received faintly on the *Birmingham* at 1325 miles. All of these observations are beyond the range of calculated day audibility and show that there was an abnormally low absorption on these days.

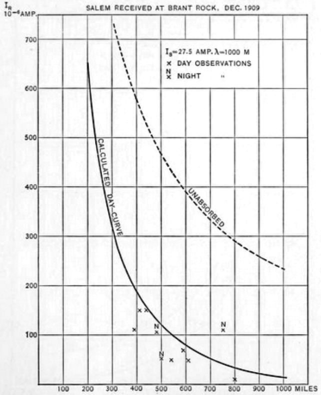

Fig. 15

Night signals during this voyage were also heard at times at great distances, the Brant Rock short wave being received on the

---

being in general stronger than the day signals, as was first observed by Marconi. Sometimes they lie close to the K/d curve, indicating that the absorption has disappeared, while at others they are practically of the same strength as the day readings. In a very few cases, night signals were observed considerably stronger than the calculated value for zero absorption, but these may very probably be due to errors of observation, since it is not observed that the remarkable strength of signal is reciprocal between the two stations. If the observations are genuine, however, it would perhaps indicate some kind of reflection from the upper layers of the atmosphere.

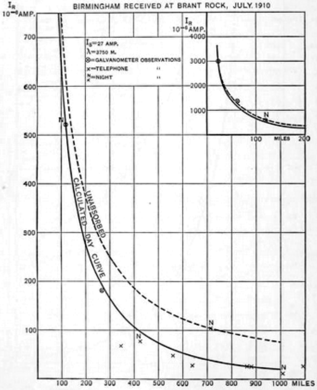

Fig. 4

THE CAUSE OF THE ABSORPTION

According to the calculations of Zenneck,[10] the conductivity of air at moderate heights can not explain the magnitude of the observed absorption, neither can the sea-water absorption be of the proper magnitude, according to the same author. But as the wave front at any considerable distance from the sending station must extend far into the upper layers of the atmosphere, it does not seem improbable that this is the region of absorption. If

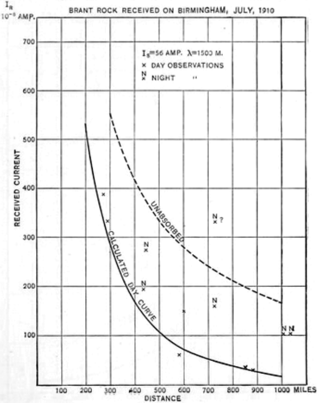

Fig. 5

---

[10] Zenneck, Ann. d. Phys., 23, p. 846; 1907.

The arrangement of the apparatus on the cruisers was highly unsatisfactory, largely on account of lack of space. In addition to this the weather was extremely tempestuous much of the time. In the light of all this, it is not surprising that the observed received currents fell below the calculated values.

The results are shown in the curves of Figs. 12 to 19.[15]

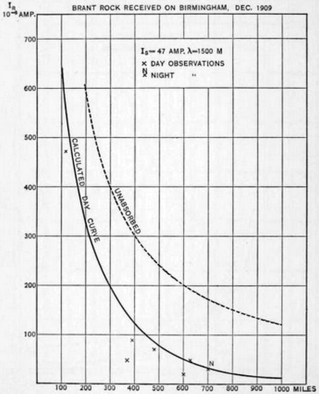

Fig. 14

MARCH EXPERIMENTS

In March the *Birmingham* was sent to Liberia, and during the voyage current measurements were made on the ship and at Brant Rock. The results are shown in Figs. 20 to 23.

---

[15] Neither of the cruisers was able to take any satisfactory measurements on the 3750-meter wave.

all this into account, the results of these earlier tests can not by any means be considered to have the same weight as those in July. But as it is desirable to obtain all the light possible on the validity of the transmission formula, the following observations are given.

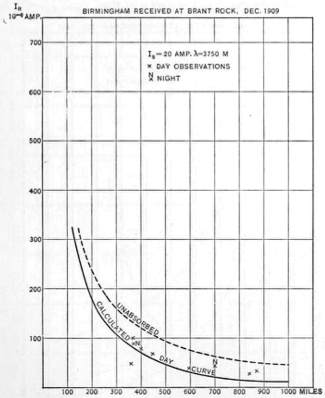

Fig. 13

DECEMBER EXPERIMENTS

The December experiments were made shortly after the installation of the sets on the cruisers, the *Birmingham* taking an easterly course from Provincetown, while the *Salem* sailed to the southward. At the greatest distance attained the two ships were distant from each other and from Brant Rock approximately 1000 miles.

---

the conductivity is increased by the sun's rays at these heights, this would also explain the differences in the strength of the day and night signals. The observations would indicate, if this explanation is true, that the excessive ionization may, especially in summer, persist through many nights. The general experience of wireless telegraphy would indicate that during the colder months the absorption dies out more uniformly at night. The day absorption appears from the data obtainable to be fairly uniform

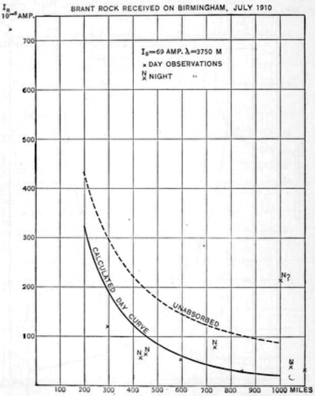

Fig. 6

throughout the year, at least in the portion of the ocean covered by our observations, although there are undoubtedly variations at times, as has been already mentioned in regard to the *Birmingham*-Brant Rock signals. There are well-authenticated instances when, for a day or two, the absorption seems to nearly disappear in the daytime. A case in point is that of the long-distance day signals between the *Birmingham* and Brant Rock on the return voyage from Liberia mentioned on page 351.

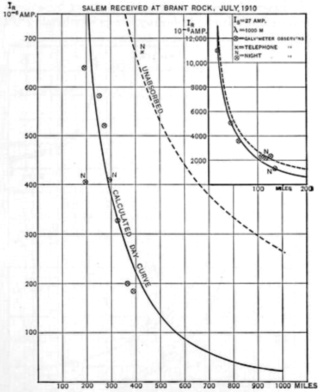

Fig. 7

OTHER LONG-DISTANCE WORK

In all of the experiments with the cruisers excepting those carried on during the month of July the calculated received currents at unit distance were estimated from the ratio of the sending currents to those observed at the corresponding stations in July, account of course being also taken of the change in antenna heights on the cruisers from 112 feet to 130 feet after the close of the December tests. This calculation is somewhat arbitrary, since there were many changes in the arrangement of the apparatus which very probably had an influence on the efficiency of the sets. Taking

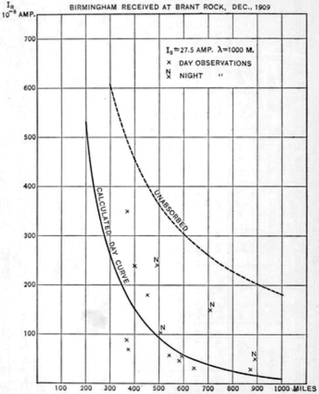

Fig. 12

TABLE XVI-A

Sending Currents Required to Produce Received Currents as Given in Table XVI for Various Antenna Heights

$h_1=h_2$	I_S amp.		$h_1=450$ ft.			$h_1=200$ ft.			$h_1=130$ ft.			$h_1=100$ ft.	
			h_2	I_S amp.		h_2	I_S amp.		h_2	I_S amp.		h_2	I_S amp.
32.5 ft.	480		32.5 ft.	34.5		32.5 ft.	78		32.5 ft.	120		32.5 ft.	156
65	120		65	17.3		65	39		65	60		65	78
100	51		100	11.3		100	25.3		100	39		130	39
130	30		130	8.7		130	19.5		160	24.4		160	31.8
160	19.8		160	7.1		160	15.8		200	19.5		200	25.4
200	12.7		200	5.6		300	8.5		300	13		300	16.9
300	5.66		300	3.76		450	5.65		450	8.7		450	11.3
450	2.5												

SUPPLEMENTARY MEASUREMENTS

PROPORTIONALITY OF SENDING AND RECEIVING CURRENTS

It was thought possible that when the antennas were used with the highest power, especially at the longer wave lengths, there might be losses of energy, either by brushing or by leakage over the insulators, which might destroy the proportionality between

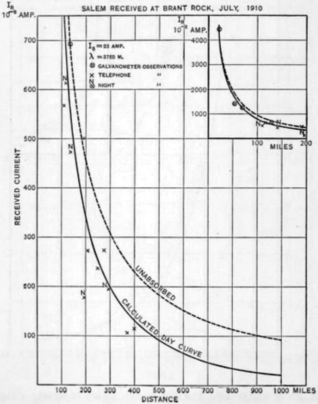

Fig. 8

sending antenna current and received antenna current. To settle this question, the station at Brant Rock sent with varying powers at a wave length of 3750 meters. The signals were received on the antenna at the Bureau of Standards at Washington, D. C., 380 miles away. The receiving antenna was of the harp form, 180 feet high at top, 40 feet high at bottom, and 18 feet wide, with 8 wires. The capacity was approximately 0.0012 microfarad. A chalcopyrite zinkite rectifier, with a sensitive galvanometer, was connected in a tuned secondary circuit coupled to the antenna inductance so as to give maximum deflection, the circuits being those shown under circuit A in Fig. 1. The results are shown in the curve of Fig. 11. The galvanometer deflections are proportional to the square of the current, hence the square roots of the deflections expressed in microamperes of received antenna current are compared with the sending antenna current at Brant Rock.

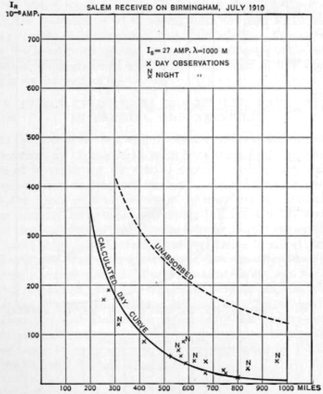

Fig. 9.

TABLE XVI

Currents in Microamperes Received Through 25 Ohms for a Sending Current of 30 Amperes and Sending and Receiving Flat Top Antenna Heights of 130 Feet, over Salt Water

[Calculated from Eq. 3]

Nautical Miles	d Km	λ=300 m		λ=600 m		λ=1000 m		λ=1500 m		λ=2500 m		λ=3750 m		λ=6000 m	
		K=6.67.10⁵		K=3.33.10⁵		K=2.10⁵		K=1.33.10⁵		K=8.10⁴		K=5.33.10⁴		K=3.33.10⁴	
		$\frac{K}{d}$	I_R	$\frac{K}{d}$	I_R	$\frac{K}{d}$	I_R	$\frac{K}{d}$	I_R	$\frac{K}{d}$	I_R	$\frac{K}{d}$	I_R	$\frac{K}{d}$	I_R
20	37	18000	16300	9000	8350	5400	5100	3590	3430	2160	2080	1440	1400	900	880
50	92.5	7200	5600	3600	3020	2160	1880	1435	1282	864	790	576	536	360	340
100	185	3600	2170	1800	1270	1080	817	720	573	431	361	280	249	180	161
200	370	1800	650	900	433	540	310	359	228	216	152	144	108	90	71.8
300	556	1200	263	600	205	360	156	240	122	144	85	96	62.4	60	42.7
400	740	900	116	450	107	270	89	180	73	108	53.6	72	40.8	45	28.6
500	925	720	55.4	360	60.2	216	53.8	144	46.3	86.4	35.9	57.6	28.2	36	20.5
600	1110	600	27.6	300	34.8	180	34.2	120	31.1	72	25.3	48	20.4	30	15.2
800	1480	450	7.72	225	12.8	135	14.7	90	14.8	54	13.1	36	11.5	22.5	9.06
1000	1850	360	2.27	180	4.97	108	6.7	72	7.5	43.1	7.42	28.8	6.77	18	5.80
1200	2220	300	0.66	150	2.0	90	3.24	60	3.95	36	4.36	24	4.34	15	3.89
1500	2780	240	0.12	120	0.54	72	1.10	48	1.58	28.8	2.04	19.2	2.25	12	2.18
2000	3700	180	0.0081	90	0.069	54	0.205	36	0.385	21.6	0.64	14.4	0.817	9	0.928
2500	4630	144	....	72	0.0086	43.2	0.038	24.8	0.098	17.3	0.213	11.5	0.318	7.2	0.418
3000	5560	120	....	60	0.0010	36	0.0079	24	0.031	14.4	0.073	9.1	0.129	6.0	0.198

currents and four wave lengths. The table shows the great advantage of long wave lengths for very distant stations. It is seen that for good communication at a distance of 2500 miles not less than 240 amperes[13] must be used at a wave length of 6000 meters, though one-fourth of this would be audible.

For the purpose of easily calculating the probable day working distance for various sending currents, heights, and wave lengths according to equation (3) Tables XVI and XVI-A are given. Table XVI gives the current which may be expected to be received through 25 ohms for an antenna height at both stations of 130 feet and a sending antenna current of 30 amperes. Table XVI-A gives the sending currents which will be required for various heights of the sending and receiving antennas h_1 and h_2 to give the values of the received currents as shown in Table XVI. To determine the current which will be received for any antenna height, wave length, and distance for a given sending current, multiply the value, given in Table XVI for the given distance and wave length, by the given current, divided by the value of I_s in Table XVI-A for the given antenna heights.[14] In Table XVI it may be assumed that 40×10^{-6} ampere insures good communication and that 10×10^{-6} ampere is just audible.

---

[13] From the experience of the Marconi transatlantic stations it would appear possible that for the same antenna current the distances given in the table could be obtained with a smaller antenna height provided a bent antenna were used.

[14] Example: If $h_1=200$ feet, $h_2=130$ feet, $\lambda=600$ meters, distance=500 miles, and $I_S=12$ amperes, then $I_R=60.2 \times 12/19.5=37.0 \times 10^{-6}$ ampere. The above tables have been given in this convenient form for calculation largely with the hope that they will be tested by the experience of the various wireless companies. It is very common in wireless stations to have an antenna meter for the measurement of the sending current. When this is used with a shunt, the multiplying value of the shunt for the given frequency may easily be obtained by connecting a like unshunted hot-wire meter in the antenna in series with the regular instrument and sending with reduced power so that the needle of the unshunted meter just remains on the scale. The ratio of the readings of the two meters then gives the multiplying value. When only one meter is available an approximate calibration may be obtained by sending with reduced power and observing the reading of the meter without the shunt and then with the shunt connected. Of course this calibration must be carried out with the same wave length used in regular work. The range of good daylight communication with another flat-top antenna of known height over sea water can be compared with the values in the table for the given wave length, distance, and sending current, provided the sending station radiates but a single wave length. If the equation is correct, the tables should indicate a current of from 20 to 40 microamperes.

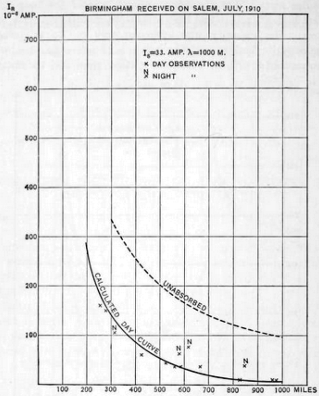

Fig. 10

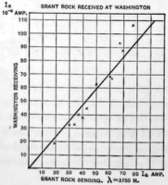

Fig. 11

The proportionality is fairly good. It is possible that in the changes of adjustment of the circuits at Brant Rock the tuning became somewhat deranged, or that the change of the coupling at Brant Rock produced a change in the damping of the waves, which was not compensated for by a change in the coupling at Washington. A few observations were taken at Brant Rock on signals sent from the *Birmingham* when coming into Newport, R. I., both at wave lengths of 1000 meters and 3750 meters. From these it was concluded that there was no marked overloading of the sending antenna, although there were slight indications in this direction at the wave length of 3750 meters.

THE EFFECT OF ANTENNA HEIGHT AND WAVE LENGTH IN SENDING AND RECEIVING

At the close of the test with the scout cruisers, a flat top antenna was erected at Plymouth, 11 miles from Brant Rock, for the purpose of investigating the relationship of height of antenna and wave length to sending and receiving. Except for a narrow sand spit the whole path of the waves between the two stations lay over water. The antenna was swung between two noninsulated steel masts in order to reproduce as fully as possible the conditions on board ship. The Plymouth antenna had a length of 110 feet and consisted of 11 wires 2 feet apart. The capacity was approximately 0.001 microfarad. The Plymouth station was furnished with a 2-kilowatt, 500-cycle Fessenden sending set, and a somewhat larger 500-cycle set used at Brant Rock. The currents were measured at both stations by means of 15-ohm thermoelements with galvanometers as described on page 317. The experiments were begun by my assistants, but were completed by Dr. Cohen of the National Electric Signalling Co.

SENDING

Table XII gives the results of observations taken by Dr. Cohen and his assistants, Plymouth being the sending station and Brant Rock the receiving. This table is for the most part self-explanatory. The deflections of the galvanometer connected to the thermoelement are proportional to the square of the current, so that the square roots of these deflections are taken as proportional to the current itself. The deflections are in arbitrary units. For

much as 30 per cent under ordinary conditions. This difference in the value of the constant does not, however, make a very great difference in the range of communication, although for short distances it would make a considerable difference in the loudness of signal. To show the variation in range of communication with different antenna currents, Table XIV is given, showing the range of communication between two flat top antennas 130 feet high at a wave length of 1000 meters for various values of the sending antenna current, assuming that 40×10^{-6} ampere assures good communication. A quarter of the indicated current would produce audible signals under favorable conditions. The table shows how very slowly the range increases with increasing sending current. Increasing the current from 30 to 60 amperes increases the distance only from 565 miles to about 725 miles, while even with 10 amperes, corresponding to a moderately efficient 2-kilowatt set, 345 miles is easily reached.

In columns 3 and 5 are given the distances attainable for regular communication and audibility on the assumption that there is no atmospheric absorption; that is, the extreme range of night communication.[12]

TABLE XV

Good Working Distance and Sending Current for Two Stations with Flat Top Antennas 450 Feet High

Nautical Miles	$\lambda=1000$ m	$\lambda=2500$ m	$\lambda=3750$ m	$\lambda=6000$ m
1000	15 amp.	13.5 amp.	15 amp.	17 amp.
1250	38	27	27	30
1500	91	49	44	46
1750	200	95	77	74
2000	490	155	122	105
2250		245	200	160
2500		470	314	235
2750			500	335
3000			775	500

In Table XV is given the estimated day range of two large stations with flat-top antennas 450 feet high for various antenna

---

[12] Excluding the possibility of the effects of reflection from the upper atmosphere.

We may write this equation as follows:

$$I_R = 4.25 \frac{I_s h_1 h_2}{\lambda d} e^{-\frac{\alpha d}{\sqrt{\lambda}}} \qquad (3)$$

where I_R represents the current received through 25 ohms resistance, I_s the sending current, h_1 the height of one antenna, h_2 that of the other, d, the distance between the stations, and λ the wave length, where all the lengths are taken in kilometers and the currents in amperes. α is the absorption coefficient, which in our experiments was found to be equal to 0.0015. This disregarding the absorption term $e^{-\frac{\alpha d}{\sqrt{\lambda}}}$, corresponds in form to the Herzian equation for the electromotive force in a vertical resonator at a distance from the oscillator.

TABLE XIV

Calculated Relation between Antenna Current and Distance for Two Ships with Antenna Heights of 130 Feet

$\lambda=1000$ m

Antenna Current I_S	Working Distance 40.10^{-6} amp.		Extreme Distance of Audibility 10.10^{-6} amp.	
	Day	Night (Zero Absorption)	Day	Night (Zero Absorption)
1 amp.	75 miles	90 miles	200 miles	360 miles
2	135	180	300	720
3	180	270	375	1080
5	235	450	475	1800
7	280	630	550	2520
10	345	900	630	3600
15	420	1350	725	5400
20	475	1800	790	7200
25	525	2250	840	9000
30	565	2700	900	10800
40	630	3600	970	14400
50	685	4500	1025	18000
60	725	5400	1150	21600

In our formula the constant 4.25 applies strictly only to ship antennas with the same losses due to masts, rigging, etc., both in sending and receiving, as found on the scout cruisers. On other ships the value of the constant might probably differ by as

---

comparison, the square roots of the deflections were reduced to a uniform sending antenna current. These values are given under I_R in the next to the last column, while the last column gives this quantity divided by the height. Although there is considerable

TABLE XII

Plymouth Received at Brant Rock

$\lambda=3900$ m

Antenna Plymouth h	I_S	Deflection Receiving D	$I_R = \sqrt{D}$ reduced to $I_S=4.7$ amp.	$\frac{I_R}{h}$
80 ft.	4.7 amp.	62.3	7.9	0.099
70	4.6	34	5.95	.085
60	4.7	30.3	5.5	.092
50	4.85	24.1	4.76	.095
40	4.9	15.6	3.79	.095
30	5.1	7.4	2.51	.084

$\lambda=1585$ m[1]

h	I_S	D	I_R ($I_S=5.5$ amp.)	$\frac{I_R}{h}$
70 ft.	5.5 amp.	292	17.1	0.244
60	5.5	198	14.1	.235
50	5.65	110.7	10.2	.204
40	6.05	84	8.32	.207
30	6.3	62	6.88	.229

h=80 ft.
35 ohms added in receiving antenna

λ	I_S	D	I_R ($I_S=4.2$ amp.)	$I_R \times \lambda$
3900 m.	4.2 amp.	4.0	2.0	7800
2980	4.5	8.3	2.69	8030
2400	4.75	13.6	3.26	7830
1985	5.1	26.3	4.22	8400

Shorter wave lengths were not used on account of the long natural wave length of the Brant Rock station.

variation in the values of this column it is evident that there is no systematic variation, and for the individual wave lengths we must consider that the received current is proportional to the height of the sending flat-top antenna. The receiving antenna was tuned by means of a variable inductance in series with the thermoelement.

In the last section of Table XII the results show the variation with wave length, and in this case the equivalent received current is multiplied by the wave length and the product is seen to be approximately constant. Thirty-five ohms was introduced into the receiving antenna to nullify the changes in the resistance of the inductance during the process of tuning. This table shows that the received antenna current is proportional to the height of the sending attenna and inversely proportional to the wave length.

RECEIVING

Table XIII gives similar data in regard to the Plymouth antenna used for receiving, the sending station being Brant Rock. This table shows that the same relationship holds for receiving as for sending; that is, for a constant sending antenna height and current the received current is proportional to the height of the receiving antenna and inversely proportional to the wave length.[11]

While it appears that the sending and receiving relations for a flat-top antenna are reciprocal, this is apparently not the case for an umbrella antenna. It is seen from the observations that the scout cruisers, sending with 10 kilowatts, were received at Brant Rock with almost the same strength as that with which Brant Rock, sending with 60 kilowatts, was received on the cruisers. Comparing the strength of the sending and receiving currents, it appears that the umbrella antenna at Brant Rock for sending is only equivalent to a flat-top antenna 170 feet high, which is not far from the height of the lower ends of the umbrella, while for receiving its equivalent height is much greater. A few experiments have also been made on an umbrella and flat top at Plymouth, which also showed that an umbrella is a better receiver than radiator.

---

[11] This will not strictly hold in the case where the radiation resistance of the receiving antenna is comparable with the rest of the antenna resistance. See R. Ruedenberg, Ann. d. Phys., 25, p. 446; 1908.

TABLE XIII

Brant Rock Received at Plymouth

$\lambda = 4000$ m

Antenna Plymouth h	I_S	Deflection Receiving D	$I_R = \sqrt{D}$ reduced to $I_S = 7.0$ amp.	$\frac{I_R}{h}$
80 ft.	7.0 amp.	8.4	2.90	0 0362
70	10.0	12.4	2.46	.0352
60	9.4	8.0	2.10	.0350
50	9.75	6.0	1.76	.0352

$\lambda = 1980$

h	I_S	D	I_R ($I_S = 7.4$ amp.)	$\frac{I_R}{h}$
80 ft.	7.4 amp.	26.0	5.1	0.0637
70	7.6	24.0	4.77	.0682
60	11.1	41.1	4.28	.0713
50	10.9	22.6	3.22	.0644
40	10.4	15.5	2.80	.0700

h = 70 ft.
35 ohms added in receiving antenna

λ	I_S	D	I_R ($I_S = 10$ amp.)	$I_R \times \lambda$
4000 m	10.0 amp.	4.3	2.07	8300
3550	9.1	4.8	2.41	8560
2980	8.6	6.0	2.85	8500
2510	9.9	12.9	3.63	9110
1984	7.6	10.0	4.16	8240

Uniting the experimental data contained in the last two tables with the data obtained from the experiments with the cruisers, we may write an equation which will cover the normal day received current over salt water through 25 ohms for two stations with flat-top antennas of any height, with any value of sending current and any wave length, provided the sending station is so coupled as to give but one wave length.

KEVIN'S WEBSURFER HANDBOOK VIII
FOR CRYSTAL RADIO

RESONANCE and RESONANT CIRCUITS

Kevin Smith
2012

Printing / Binding Instructions

1. Choose “fit to page” in print menu
2. Print document double sided on letter size paper
3. Cut the entire printed document in half
4. Fold over making sure the page numbering is continuous
5. For the cover: Print just the first page on card stock paper
   Cut the cover in half as well
6. Assemble the covers on the document
7. Punch the left side for a binding, spiral or comb as desired

http://www.lessmiths.com/~kjsmith/crystal/catalog.shtml

KJ Smith

NOTES:

INTRODUCTION

This handbook explores radio in general and crystal radio from the standpoint of RF Resonance. I again begin things with a brief walk down memory lane with an introduction to the subject of resonance and coupled circuits by Dr Hoag written in 1942. This represents an early, but already well-developed notion on the theory and application of resonance to radio apparatus.

The next few articles explore general theory concerning resonance with a strong emphasis on the mathematical base in order to help the reader better apply these concepts in theory own work, especially with respect to setting up excel spreadsheets. Note that while I like the pretty plots from SPICE simulations, I see this as a black box because authors almost never present the equations that make the plots possible, they are canned inside the program, bad form. I feel that, if I can plot it myself, I must be understanding it pretty well.

Then included are general discussion articles on resonance and coupled circuits. Most crystal radios of any pretentions to good performance will be coupled with separate antenna tuning. These articles should help explore this aspect of your radio!

I finally follow with an article on impedance matching and resonance as these are closely-related and need to be thought of together when designing, or at least understanding your radio.

Of general note, the web is a marvelous source of data and information. Many long-time crystal set builders, and many others have created dedicated sites to disseminate information and resources, to share their creations and knowledge. I am eternally in your debt. All of the material in this handbook is copyright for which I have not sought permission. Therefore this is not presented for publication or copy. It is only my personal resource. I encourage anyone finding this copy to pursue ON THE WEB the web pages identified within. I include the name of the author and web address of each section. I wish to sincerely thank every author presented for their excellent pages and ask forgiveness for my editing into this handbook.

Kevin Smith
2012

www.lessmiths.com/~kjsmith/crystal/cr0intro.shtml

NOTES:

as shown in Fig. 24*b*. It will be noted that the use of link coupling has the effect of reducing the equivalent primary and secondary inductance, and hence of raising the resonant frequencies. In other respects, the use of link coupling gives the same behavior as obtainable from ordinary inductive coupling.

6. Coupled Circuits with Resonant Primary and Secondary Tuned to Slightly Different Frequencies.—When two circuits resonant at slightly different frequencies are coupled together, the behavior depends upon the coefficient of coupling and the relative and absolute circuit Q's. When $Q_p = Q_s$, the response curve of secondary current (or voltage) has a shape of the same type as would be obtained if the two circuits were both resonant at the same frequency and coupled with an increased coefficient of coupling. In other words, the effect of detuning the circuits for the case of equal Q's is equivalent, as far as shape is concerned, to increasing the effective coupling. The only difference is that detuning causes the absolute magnitude of the response curve to be less than when the same shape is obtained without detuning.

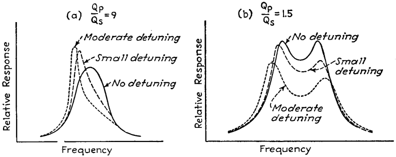

FIG. 25.—Curves illustrating the effect produced on the shape of the response curve by detuning primary and secondary circuits when the primary and secondary circuits do not have identical Q's.

Analysis shows that when $Q_p = Q_s$ the shape with detuning is as though the circuits were both tuned to the same frequency and coupled by an amount[1]

$$\left.\begin{array}{r}\text{Equivalent coupling}\\ \text{corresponding to}\\ \text{detuned operation}\end{array}\right\} = \sqrt{k^2 + \left(\frac{\Delta}{f_0}\right)^2} \qquad (43)$$

where k = actual coefficient of coupling.
Δ = difference between resonant frequency of primary and secondary circuits.
f_0 = frequency midway between primary and secondary resonant frequencies.

In the more general case where the circuit Q's are not the same, the secondary response curve is no longer symmetrical about the mean frequency, as shown in Fig. 25. In cases where unsymmetrical peaks are obtained, the low-frequency peak will be depressed when the secondary is tuned to a higher frequency than the primary and the secondary Q is higher, or when the secondary is tuned to a lower frequency and the secondary Q is the lower. Otherwise the high frequency peak will be depressed.

Detuning produces much more effect upon the shape of the primary current curve than upon the secondary response, and if the detuning is at all large the primary

---

[1] See Harold A. Wheeler and J. Kelly Johnson, High Fidelity Receivers with Expanding Selectors, *Proc. I.R.E.*, Vol. 23, p. 594, June, 1935; or C. B. Aiken, *loc. cit.*
Equation (43) involves the assumption that Δ/f_0 is small compared with unity.

CONTENTS:

2 **Resonant Circuits**
J Barton Hoag 1942

15 **In the beginning... Crystal Radio**
L Van Warren

59 **Resonant Circuits**
Mouloud Rahmani

75 **Resonance**
Tony R. Kuphaldt
75 6.1 An electric pendulum
83 6.2 Simple parallel (tank circuit) resonance
90 6.3 Simple series resonance
96 6.4 Applications of resonance
98 6.5 Resonance in series-parallel circuits
112 6.6 Q and bandwidth of a resonant circuit
112 6.6.1 Series resonant circuits
112 6.6.2 Parallel resonant circuits
121 6.7 Contributors

123 **Resonant Circuits**
DesignSoft, Inc

145 **Resonance in a series RLC circuit**
Kshitij Education India

149 **Resonance in RLC Circuits**
Boundless.com

151 **RLC Parallel Circuit**
HyperPhysics

155 **The Parallel RLC Circuit**
AspenCore

169 **Mutual Inductance of Two Coils**
Wayne Storr

175 **Mutual Inductance**
Dr. David F. Cattell

181 **Variometer Design**
Claudio Girardi

185 **Air-Core Transformers**
Reuben Lee

191 **Two inductively coupled RLC circuits**
Ben-Gurion University

197 **Mutual inductance**
BGU Lab-7 download

207 **Using a Bandspread Capacitor**
Robert Weaver

221 **Terman Inductive Coupling Model**
Mike Tuggle

231 **A graphical look at Resonance**
Kevin Smith

247 **Inductively Coupled Circuits (REH)**
F.E. Terman, 1943

263 **Notes**

---

circuit, as shown in Fig. 17. In many cases where coupled resonant circuits are employed, the excitation of the primary circuit is obtained by applying the voltage in parallel with the primary circuit through a resistance, as shown in Fig. 23*a*. This arrangement can be termed *parallel excitation*, and corresponds to the case when the coupled circuits are excited from a vacuum tube.

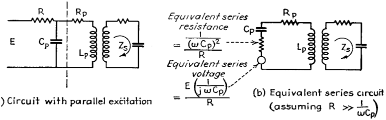
R
Rp
E
Cp
Lp
Zs
Equivalent series resistance
$= \frac{\frac{1}{(\omega C_p)^2}}{R}$
Cp
Rp
Lp
Zs
Equivalent series voltage
$= \frac{E\left(\frac{1}{j\omega C_p}\right)}{R}$

(a) Circuit with parallel excitation

(b) Equivalent series circuit (assuming $R \gg \frac{1}{\omega C_p}$)

FIG. 23.—Actual and equivalent circuits for the case of parallel excitation of two coupled circuits.

The parallel excited circuit of Fig. 23*a* can be reduced by Thévenin's theorem to the equivalent arrangement shown in Fig. 23*b*. This is accomplished by considering that the portion of the circuit to the left of the dotted line is the equivalent generator circuit supplying the remainder of the circuit. The principal effect of parallel excitation is to introduce an added resistance in series with the primary circuit that becomes greater the lower the series resistance R used in the parallel feed. The equivalent series voltage also varies somewhat with frequency, but in a limited range, such as the region in which the resonance phenomena occur, can be considered as being substantially constant.

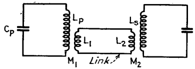

(a) Actual circuit

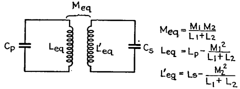
Meq
Cp
Leq
L'eq
Cs
$M_{eq} = \frac{M_1 M_2}{L_1 + L_2}$
$L_{eq} = L_p - \frac{M_1^2}{L_1 + L_2}$
$L'_{eq} = L_s - \frac{M_2^2}{L_1 + L_2}$

(b) Equivalent circuit (assuming zero link losses)

FIG. 24.—Actual and equivalent circuits for link coupling.

Link-coupled Circuits.—Two resonant circuits are often coupled together by a link circuit, as shown in Fig. 24*a*. When this arrangement is reduced to an inductively coupled circuit by the method of Par. 8, the result for negligible losses in the link is

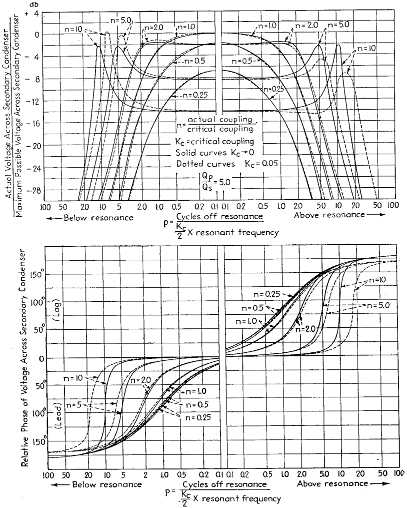

FIG. 22b.—Universal curves giving the phase and relative magnitude of the voltage across the secondary condenser for the case of two coupled circuits resonant at the same frequency and having a Q ratio of 5.

Parallel Feed in the Primary Circuit.—The analysis and discussion given above have all been for the case where the voltage was applied in series with the primary

---

http://www.tubebooks.org/Books/hoag.pdf

BASIC RADIO

The Essentials of

ELECTRON TUBES AND THEIR CIRCUITS

By

J. BARTON HOAG, PH.D.

PROFESSOR (T), WITH THE RANK OF LIEUTENANT COMMANDER, U.S.C.G.
Head of the Department of Science, U. S. Coast Guard Academy
Fellow of the American Physical Society
Member of the Institute of Radio Engineers
FORMERLY, *Assistant Professor, Dept. of Physics, Univ. of Chicago*

FIFTH PRINTING

NEW YORK
D. VAN NOSTRAND COMPANY, INC.
250 FOURTH AVENUE
1942

CHAPTER 6

RESONANT CIRCUITS

6.1 Series Resonance. When two bodies, or two currents, or two voltages, oscillate at the same frequency they are said to be in resonance.

A series *LCR* circuit can be built which resonates to a generator of given frequency. Then the current which flows through the circuit will have its greatest effective value. This is accomplished by choosing the inductive and capacitive reactances equal to each other. From the equations for the reactances, it is easily proven that the resonant frequency is given by

$$\boxed{f_r = \frac{1}{2\pi\sqrt{LC}} \times 10^9}$$

where f_r is in cycles per second, $2\pi = 6.28$, $10^6 = 1$ million, L is the inductance in microhenries (μh), and C is the capacitance in micromicrofarads ($\mu\mu$f). This is one of the most important equations in radio. Note that the resistance of the circuit does not appear in the equation. At resonance, the only opposition to the flow of the current is that due to the resistance, i.e., $I_r = E/R$ and not $I = E/Z$.

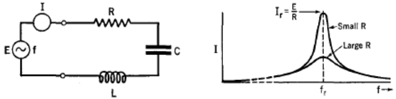

FIG. 6 A. A series circuit and its resonance curve

The current in the "tuned" or resonant series circuit is indicated at the peak of the "resonance" curve of Fig. 6 A. Note that if the total resistance of the circuit is large, the curve is broader and flatter, and vice versa.

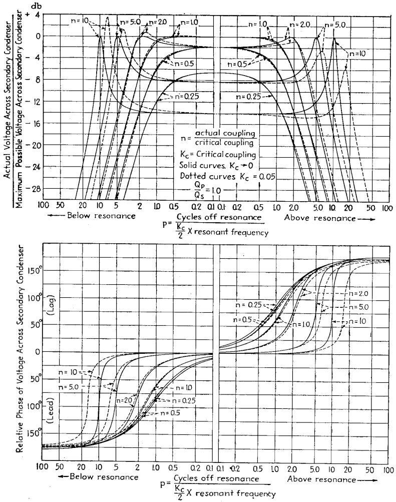

FIG. 22a.—Universal curves giving the phase and relative magnitude of the voltage across the secondary condenser for the case of two coupled circuits resonant at the same frequency and having a Q ratio of unity.

many cycles as do the peaks, this can be readily obtained by neglecting the circuit resistance and using Eq. (34).

When $Q_p = Q_s$, the value of E_c/E approximates the maximum possible response that can be obtained, as given by Eq. (38). On the other hand, when $Q_p \neq Q_s$, the peaks always have a height less than the maximum possible response in the secondary.

The ratio of the mean of the responses at the two peaks to the response at resonance is plotted in Fig. 21 for a number of Q ratios.

Universal Response Curves.—Equations (33) and (38) can be combined to give

$$\frac{\left\{\begin{matrix}\text{Actual voltage across}\\ \text{secondary condenser}\end{matrix}\right\}}{\left\{\begin{matrix}\text{Maximum possible volt-}\\ \text{age across secondary}\\ \text{condenser with}\\ \text{optimum coupling}\end{matrix}\right\}} = \frac{2n}{\left(1+\frac{pk_c}{2}\right)^2\left[(n^2+1)-(mp)^2+j(mp)\left(\sqrt{\frac{Q_p}{Q_s}}+\sqrt{\frac{Q_s}{Q_p}}\right)\right]} \tag{42a}$$

where $n = \dfrac{k}{k_c} = \dfrac{\text{actual coefficient of coupling}}{\text{critical coefficient of coupling}}$.

$$p = \frac{\text{cycles off resonance}}{\left(\frac{k_c}{2}\right) \times \text{resonant frequency}}.$$

$$m = \frac{1+\frac{pk_c}{4}}{\left(1+\frac{pk_c}{2}\right)^2}.$$

Universal curves based on Eq. (42*a*) are given in Figs. 22*a* and 22*b* for $k_c = 0$, and $k_c = 0.05$. These curves, especially when considered in conjunction with Figs. 18, 19, and 21, provide a complete and accurate picture of the response that can be expected when two resonance circuits are tuned to the same frequency and coupled together.

Figure 22 shows that when both primary and secondary circuits have the same Q, the response curve separates into double peaks whenever the coefficient of coupling exceeds the critical value, and these peaks maintain a substantially constant height irrespective of their location. On the other hand, when the two circuits have different Q's, then if the coupling exceeds the critical value by only a small amount, the response curve still has a single peak, but with reduced response at resonance. Double humps do not occur until the coupling exceeds the critical value by an amount that depends upon the ratio Q_p/Q_s, and the peaks become slightly lower as the coupling increases.

Practical Calculation of Coupled Systems Involving Two Resonant Circuits Tuned to the Same Frequency.—When it is desired to obtain the response curve of a given system, the relative shape of the curve of voltage developed across the secondary condenser can be determined with an accuracy sufficient for all ordinary purposes by the use of the universal curves of Fig. 22, interpolating between these in accordance with the Q ratio involved. The absolute magnitude of the curve can then be obtained by calculating the response at the resonant frequency with the aid of Eq. (35).

An alternative procedure for the case where the curve shows two peaks is to calculate the response at resonance by Eq. (35), determine the location and height of the peaks by Figs. 19 and 21, or Eqs. (40) and (42), and then to note that the response beyond the peaks falls to the response at resonance when the frequency is $\sqrt{2}$ times as far from resonance as the coupling peaks. This information is sufficient to define the shape of the double peaked curve in the vicinity of resonance. If quantitative information is desired at frequencies differing from resonance by at least twice as

---

6.2 Parallel Resonant Circuits. Fig. 6 B shows a parallel circuit and its impedance curves. The curves are of the same shape as the current curves of a series circuit. Note that the impedance across the terminals of a parallel circuit is a maximum at resonance,[1] whereas it is a minimum for the series circuit. The current in the *LCR* circuit itself is often very large at resonance and can cause considerable heating.

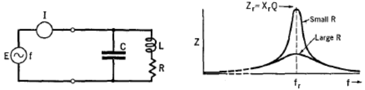

Fig. 6 B. A parallel circuit and its impedance curve

The impedance of a parallel circuit at resonance is given by

$$Z_r = \frac{(2\pi f_r L)^2}{R} = (2\pi f_r L)Q = X_r Q = \frac{L}{RC}$$

if all the resistance is in the coil (which is very nearly true in practice). The meaning of Q is given in Section 6.4. The resonant frequency f_r is given by the same equation as for a series circuit provided R is small in comparison with X. This is often true in radio frequency circuits.

6.3 Selectivity. The sharpness of the resonance curves is greater when the resistances are smaller. Sharpness of tuning is also called "selectivity." It shows the ability of the circuit to discriminate between voltages of different frequencies.

6.4 Q of a Circuit. As current surges back and forth in a resonant circuit, the electrical energy is alternately shifted back and forth between the magnetic field of the coil and the electric field of the condenser, losing a certain proportion of the total into the resistance at each alternation. In order to maintain oscillations, the generator must supply a small amount of energy each cycle. This is analogous to the small push we give to a swing; only a little push is needed at the end of each swing to keep it going. Electrically, the interest lies in the comparison between the total energy in the circuit to that which is dissipated (and must be re-supplied by the generator). This ratio is called the "Q" (quality) of the circuit. In practice, nearly all of a circuit's resistance is in the coil. It may then be shown that

---

[1] There are really three different ways of defining resonance in the case of a parallel circuit, but, if the resistance is low, the resonant frequency in all three cases is practically identical.

$$\boxed{Q = \frac{2\pi fL}{R} = \frac{X_L}{R}}$$

Practical radio frequency coils have Q's from 50 to 200 and occasionally 500. Audio frequency coils range from 1 to 10 and occasionally are as low as 1/2. The higher the Q, the better.

The rate at which current dies down in a resonant circuit, after the source has been disconnected, is called the damping of the circuit, whose measure is called " the decrement " δ. $Q = \pi/\delta$. A high Q circuit is lightly damped, has a small decrement, a sharp resonance peak, and a high selectivity.

6.5 The L to C Ratio. In order that a circuit shall resonate to a given frequency f_r, the product LC must have a definite value given by $LC = 1/(2\pi f_r)^2$. This does not tell us, however, whether the product shall be made up of a large L and a small C or vice versa. Within limits, an increase in the number of turns of wire in a coil increases its reactance faster than its resistance. Hence, for circuits alone, or for those connected to high load resistances, such as a vacuum tube, the coils should be made with a relatively large inductance, i.e., the L/C ratio should be large.

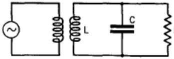

FIG. 6 C. If R is small, the LC circuit will be heavily loaded and will have a low Q. Making the ratio L/C small helps to keep Q large

When the load on a resonant circuit (R in Fig. 6 C) is small, say only a few thousand ohms, as in transmitters and induction heaters, a majority of the energy loss takes place in the load. The coil's resistance plays only a negligible role. In this case, it can be shown that L should be comparatively small and C large if Q is to be satisfactorily high.

6.6 Resonant Voltages. At resonance, the voltage across the inductance is numerically equal to that across the condenser, although they are always of opposite polarity. This voltage, at resonance, is equal to Q times the voltage in series with the circuit. The condenser must be built with an insulator which will not be punctured by this voltage.

6.7 Crystal Resonators. A properly cut and ground piece of quartz, located between two metal plates and placed in series with a high frequency generator, has a natural vibrational period dependent largely upon its thickness. It is equivalent to a series resonant circuit of very high Q.

Furthermore, in the very common case when the coefficient of coupling is small, Eq. (41a) becomes

$$\frac{\text{Frequency at peak of secondary voltage}}{\text{Resonant frequency of tuned circuits}} = 1 \pm \frac{k}{2} \qquad (41b)$$

In practical work, one is commonly interested in the frequency band over which the response in the secondary circuit equals or exceeds the response at resonance. This band is illustrated schematically in Fig. 20, and can be shown to be equal to $\sqrt{2}$ times the width of the frequency band between coupling peaks.[1]

The height of the peaks of response in the secondary circuit can be obtained by substituting the value of γ_p from Eq. (40) into Eq. (33). This gives

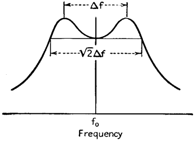

FIG. 20.—Relationship between band width, and width between secondary peaks existing when two circuits resonant at the same frequency are coupled together.

Value of $\frac{E_c}{E}$ at peaks

$$= \frac{1}{\gamma_p^2}\sqrt{\frac{L_s}{L_p}}\frac{\sqrt{Q_pQ_s}}{\left[\sqrt{\frac{Q_p}{Q_s}} + \sqrt{\frac{Q_s}{Q_p}}\right]\left[1 - \frac{k_c^2}{4k^2}\left(\sqrt{\frac{Q_s}{Q_p}} - \sqrt{\frac{Q_p}{Q_s}}\right)^2\right]^{1/2}} \qquad (42)$$

It will be noted that the lower frequency peak is the highest peak. However, since γ_p differs only slightly from unity in ordinary cases, it is customary to consider that the peaks are identical.

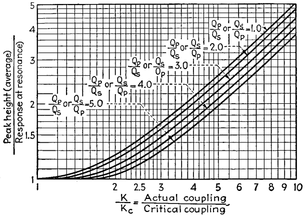

FIG. 21.—Universal curves giving the ratio of actual peak height to the response at resonance as a function of the coefficient of coupling for two coupled circuits resonant at the same frequency.

---

[1] See Aiken, *loc. cit.*

is given in Fig. 19 for various conditions, in terms of a parameter *g* given by the lower part of Fig. 19.

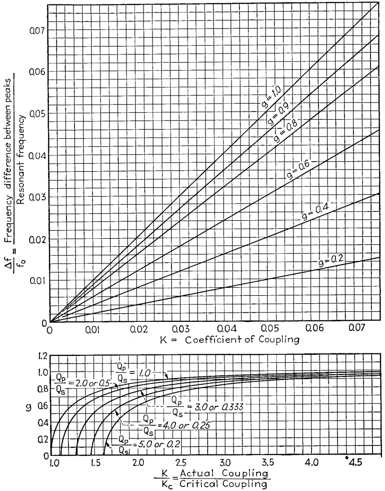

FIG. 19.—Curves from which the separation between peaks can be determined for the case of two coupled circuits resonant at the same frequency.

When the circuit *Q*'s are not too unequal, and the coefficient of coupling greatly exceeds the critical value, Eq. (40) reduces to

$$\frac{\text{Frequency at peak of secondary voltage}}{\text{Resonant frequency of tuned circuits}} = \frac{1}{\sqrt{1 \pm k}} \qquad (41a)$$

CHAPTER 7

COUPLED CIRCUITS

7.1 Introduction. It is sometimes desirable to transfer electrical energy from one circuit to a neighboring circuit. This can be accomplished, as in Fig. 7 A, by means of a transformer, wherein the source of energy is in the primary circuit and the load is in the secondary circuit. There are other coupling methods which employ a link circuit between the primary and the secondary. In still other types of coupling, the primary and the secondary circuits are actually connected together, using a coil or a condenser or a resistance, as the element common to both primary and secondary circuits.

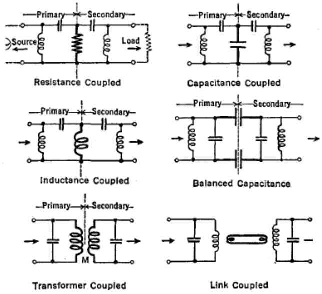
Primary
Secondary
Source
Load
Resistance Coupled
Primary
Secondary
Capacitance Coupled
Primary
Secondary
Inductance Coupled
Primary
Secondary
Balanced Capacitance
Primary
Secondary
M
Transformer Coupled
Link Coupled

FIG. 7 A. Typical coupling methods

The *transformer*, as shown in Fig. 7 A, differs in its action from the iron core transformer discussed in the chapter on a.c. circuits, in that only a few of the lines of force of the primary cut the wires of the secondary coil. The simple voltage step-up and impedance ratios applicable to the iron core case no longer hold true. The extent to which the lines of force of the primary cut the secondary is measurable, and is called the *mutual inductance*. Its symbol is M and its unit is the *henry*. When the secondary coil is placed close to the primary and on the same axis, the *coupling* will be great and M will be large. If the secondary coil is rotated, or is moved farther away from the primary, the voltages set up in it will be smaller, the coupling will be less, and M will represent a smaller number.

The coefficient of coupling is defined by the equation,

$$k = M/\sqrt{L_1L_2},$$

where the L's are the self-inductances of the primary and secondary circuits. If all of the lines of force from the primary were to cut the secondary, k would be equal to unity; the coefficient would be 100 per cent. This is the "tightest" possible coupling.

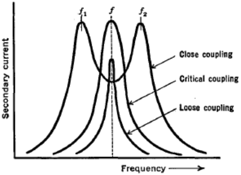

FIG. 7 B. Transformer coupling. The primary and secondary circuits were independently tuned to the same frequency (f)

When the circuits are tuned to the same frequency and are so related that but little energy is transferred from the primary to the secondary, they are said to be *loose coupled*. As the coupling is increased, the secondary increasingly "loads" the primary; and also, the primary increasingly loads the secondary. This loading action broadens the resonance curves of both circuits. At *critical coupling,* maximum energy is transferred and the resonance curve has quite a broad peak, as in Fig. 7 B. If the circuits are still closer coupled, the energy transferred from one circuit to the other decreases, and the resonance curve will have two peaks, one on either side of the single

---

This coupling k_c is often termed the *critical coupling*. The corresponding mutual inductance is

$$\omega M = \sqrt{R_pR_s} \tag{37}$$

With the mutual inductance for critical coupling, the resistance coupled into the primary circuit at resonance by the presence of the secondary equals the primary resistance. This gives the maximum possible secondary current. The resulting maximum possible voltage across the secondary condenser is

$$\text{Maximum possible } \frac{E_c}{E} = \frac{1}{2k_c}\sqrt{\frac{L_s}{L_p}} = \frac{\sqrt{Q_pQ_s}}{2}\sqrt{\frac{L_s}{L_p}} \tag{38}$$

where k_c is the critical coupling as given by Eq. (36). The voltage developed across the condenser at resonance with other couplings is less than value given by Eq. (38) by an amount depending upon the ratio of actual to critical coupling, as shown in Fig. 18.

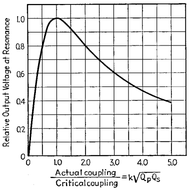

FIG. 18.—Relative output voltage obtained at resonance as a function of coefficient of coupling in the case of two circuits tuned to the same frequency and coupled together.

Conditions for Existence of Double Peak —Height, Location, and Width of Peaks.—The curve of response in the secondary circuit exhibits double peaks whenever the coefficient of coupling is sufficient to cause the right-hand side of Eq. (33) to show two maxima. This occurs when[1]

$$\frac{\text{Actual coefficient of coupling}}{\text{Critical coefficient of coupling}} \geqslant \left[\frac{1}{2}\left(\frac{Q_p}{Q_s} + \frac{Q_s}{Q_p}\right)\right]^{1/2} \tag{39}$$

It will be noted that when $Q_p = Q_s$, double peaks occur whenever the coupling exceeds the critical value, but if the Q's are different, double peaks do not occur until the coupling exceeds the critical value by an amount determined by the Q ratio.

When double humps occur, the location of the peaks in relation to the resonant frequency depends upon the actual coupling, the critical coupling, and the Q ratio. Analysis based upon Eq. (33) shows that to an accuracy sufficient for all ordinary purposes the location of the peaks is given by[2]

$$\left.\frac{\text{Frequency at peak of secondary voltage}}{\text{Resonant frequency of tuned circuits}}\right\} = \gamma_p = \frac{1}{\sqrt{1 \pm k\left[1 - \frac{k_c^2}{2k^2}\left(\frac{Q_p}{Q_s} + \frac{Q_s}{Q_p}\right)\right]^{1/2}}} \tag{40}$$

where k_c is the critical coefficient of coupling as given by Eq. (36) and the remaining notation is the same as above. The ratio represented by the spacing between coupling peaks to the resonant frequency, as a function of the coefficient of coupling,

---

[1] Equation (39) assumes that the γ^2 multiplying the denominator of Eq. (33) is constant. The approximation that results is so small as to be of negligible consequence, but results in enormous simplification of the equations.

[2] Eq. (40) involves the same approximation as does Eq. (39), but again the error introduced is negligible and the simplification considerable.

The curve of secondary current is then somewhat broader in the immediate vicinity of the resonant frequency than the resonance curve of the secondary circuit, and the primary curve shows double peaks. With still greater coupling, the double humps in the primary current curve become more prominent, and the peaks spread farther apart. At the same time the secondary current curve begins to show double humps, with the peaks becoming progressively more pronounced and farther apart as the coefficient of coupling is increased.[1]

Basic Equations.—In practical work, the most important property of the coupled system under consideration is the ratio of voltage developed across the condenser in the secondary circuit to the voltage applied in series with the primary circuit. A useful expression for this important ratio is obtained by rearranging Eqs. (22) and (23) to give the relation[2]

$$\frac{\text{Voltage across secondary condenser}}{\text{Voltage applied in series with primary}} = \frac{E_c}{E}$$

$$= \frac{-1}{\gamma^2}\sqrt{\frac{L_s}{L_p}}\,\frac{k}{\left[k^2 + \frac{1}{Q_pQ_s} - \left(1 - \frac{1}{\gamma^2}\right)^2 + j\left(1 - \frac{1}{\gamma^2}\right)\left(\frac{1}{Q_p} + \frac{1}{Q_s}\right)\right]} \tag{33}$$

where $\omega = 2\pi$ times frequency.

$\omega_0 = 2\pi$ times resonant frequency.

$\gamma = \frac{\omega}{\omega_0} = \frac{\text{actual frequency}}{\text{resonant frequency}}$.

E = voltage applied in series with primary.

$Q_p = \omega L_p/R_p$ for primary circuit.

$Q_s = \omega L_s/R_s$ for secondary circuit.

$k = M/\sqrt{L_pL_s}$ = coefficient of coupling.

L_p = total inductance of primary circuit.

L_s = total inductance of secondary circuit.

This equation involves no approximations, other than assuming that the circuits are resonant at the same frequency.

In the special case where the circuit losses are zero, or where the frequency differs sufficiently from the resonance frequency to allow the losses to be neglected, Eq. (33) takes the form

$$\frac{E_c}{E}\text{ for zero losses} = \frac{-1}{\gamma^2}\sqrt{\frac{L_s}{L_p}}\,\frac{k}{\left[k^2 - \left(1 - \frac{1}{\gamma^2}\right)^2\right]} \tag{34}$$

Effect of Coupling on Response at Resonance—Critical Coupling.—The response at resonance is obtained by substituting $\gamma = 1$ in Eq. (33), which gives

$$\frac{E_c}{E}\text{ at resonance} = -\sqrt{\frac{L_s}{L_p}}\,\frac{k}{\left[k^2 + \frac{1}{Q_pQ_s}\right]} \tag{35}$$

This expression is maximum when the coefficient of coupling has the value

$$k_c = \frac{1}{\sqrt{Q_pQ_s}} \tag{36}$$

---

[1] When the circuit Q's are not equal, double humps do not appear in the curve of secondary current until the coupling is somewhat greater than the critical value. This is illustrated in the universal curves of Fig. 22.

[2] For derivation, see Terman, *op. cit.*, p. 81.

---

frequency to which the circuits were originally tuned. The closer the coupling, the greater the frequency separation of these peaks.

If the two circuits are not exactly tuned to the same frequency, their effect upon each other is not only to increase the effective resistance (the Q of the circuit decreases) but also to reduce the self-inductance. Hence the peak of the resonance curve shifts to a different frequency.

7.2 The Effect of " Neighboring Bodies." The secondary circuit need not necessarily consist of a coil, condenser, and resistance, as in the cases we have just discussed, but can consist of any metallic, and even a dielectric body, in the neighborhood of the circuit containing the alternating current. A piece of metal placed inside a coil will have small *eddy currents* induced in it. These currents, in turn, react upon the primary circuit to increase its resistance, i.e., lower its Q, and *decrease* its inductance and hence increase its resonant frequency. Because of the increased resistance, the generator must supply an additional amount of power (current squared, times resistance). Thus, if we use the symbol $R_1{}^1$ for the *effective resistance* of a circuit (its actual resistance plus that which is " reflected " from the load), we may write

$$\underbrace{I_1^2R_1{}^1}_{\text{Total Heat}} = \underbrace{I_1^2R_1}_{\text{Prim. Heat}} + \underbrace{I_2^2R_2}_{\text{Sec. Heat}},$$

where the term on the left side represents the total power which the source must supply. The first term on the right-hand side of the equation represents the heat developed in the primary circuit, and the second term represents the heat developed in the secondary, or load — be it a circuit or a metallic body or an insulator. Now, this state of affairs may be good or it may be bad. For example, it is " good " when the circuit is used as an *induction heater,* which consists of a radio frequency generator with a tuned circuit, the coil of which consists of a few turns of heavy copper wire. The coil is placed around a body which is to be heated; such, for example, as the metal electrodes sealed inside a vacuum tube during its process of manufacture, or of a human body in which it is desired to develop an " artificial fever." The eddy currents set up in the conductive parts of the body inside the coil cause the body to heat up and, in some cases, to an extent such as to melt it completely.

If an insulator is placed inside the coil, the high frequency electrostatic fields cause the electrical charges in its atoms and molecules to oscillate back and forth about their normal positions. This requires energy which can only come from the primary circuit. This causes the dielectric to heat up, and is spoken of as *dielectric loss*. That energy should be absorbed from a circuit by a neighboring dielectric is, in general, undesirable. The resistance of the primary circuit is effectively increased by these losses.

We may generalize the concept of resistance by defining it in the following manner:

$$\text{Generalized Resistance} = \frac{\text{Watts Lost}}{\text{Current Squared}},$$

where the "watts lost" include the heat losses in the circuit itself and in neighboring bodies (as ohmic, eddy current, or dielectric heating); in general, lost in any form whatsoever from the source of power.

7.3 Shielding. A metal shield can be used to prevent coupling between two circuits. Capacitive or electrostatic coupling can be prevented by shielding either the primary or secondary, or both, circuits with an enclosing metal container. The shield should be grounded and be made of material of low resistance. Also, metal shields can be used at radio frequencies to prevent magnetic coupling. Here, the eddy currents in the shield have magnetic fields which oppose the original field and more or less completely keep it out of the inside of the shield. The shielding effect is greater at the higher frequencies, is greater for more conductive materials, and also depends upon the thickness of the shielding material. At low frequencies, the eddy currents are so feeble that this method is not satisfactory. The best that can be done in this case is to surround the circuit with a complete shield of soft iron. This will partially divert the magnetic fields.

A shield changes the resonant frequency and the Q of the circuit being shielded. In shielding a coil, the spacing between the sides of the coil and the shield should be equal to at least half the diameter of the coil, and the distance of the shield from the end of the coil should be not less than the diameter of the coil. Copper and aluminum are satisfactory metals for shields.

7.4 Introduction to Filters. Various combinations of coils and condensers are used:

1. To separate currents of different frequencies from each other.
2. As a coupling unit between two circuits whose impedances are not equal to each other.

---

5. Two Resonant Circuits Tuned to the Same Frequency and Coupled Together.[1] When two circuits resonant at the same frequency are coupled together, the resulting behavior depends very largely upon the coupling, as shown in Fig. 17. When the coefficient of coupling is small, the curve of primary current as a function of frequency approximates closely the series resonance curve of the primary circuit considered alone. The secondary current at the same time is small and varies with frequency according to a curve having a shape approximating the product of the resonance curves of primary and secondary circuits taken alone. As the coupling is increased, the curve of primary current becomes broader and the peak value of primary current is reduced. At the same time, the secondary current becomes larger and the sharpness of the secondary current curve is reduced. These trends continue until the coupling is such that the resistance that the secondary circuit couples into the primary at resonance is equal to the primary resistance. This coupling is called the *critical coupling*, and causes the secondary current to have the maximum possible value that it can attain.

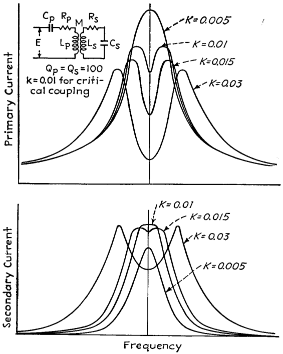

Fig. 17.—Curves showing variation of primary and secondary current with frequency for different coefficients of coupling when the primary and secondary are separately tuned to the same frequency.

---

[1] A very complete discussion of this subject is given by C. B. Aiken, Two-mesh Tuned Coupled Circuit Filters, *Proc. I.R.E.*, Vol. 25, p. 230, February; p. 672, June, 1937. See also E. S. Purington, Single and Coupled-circuit Systems, *Proc. I.R.E.*, Vol. 18, p. 983, June, 1930.

When ωL_p is not negligible compared with the primary resistance, the response curve still has the shape of a resonance curve, but the peak is shifted to a slightly higher frequency. This shift in location of the peak is due to the fact, expressed by Eq. (26), that maximum secondary current occurs when secondary is sufficiently detuned that its coupled reactance neutralizes the reactance of the primary. When the primary reactance is not negligible, this detuning is appreciable. The coupling required for maximum secondary current is also greater as the ratio $\omega L_p/R_p$ is increased.

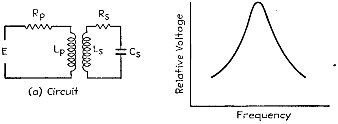

(a) Circuit

(b) Curve of Voltage Across C_S

FIG. 15.—Inductively coupled circuit with tuned secondary and untuned primary.

Analysis of the output of coupled systems with untuned primary and tuned secondary can be most conveniently carried out by using the equivalent secondary circuit of Fig. 12*d*. Such equivalent circuits are shown in Fig. 16. These show how the association of the primary with the resonant secondary has the effect of increasing the effective secondary resistance and hence of lowering the effective Q, and also that an inductive primary neutralizes some of the secondary inductance and so raises the resonant frequency.

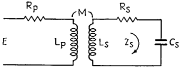

Actual circuit

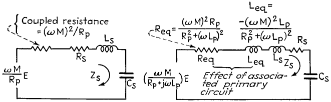
Coupled resistance = (ωM)²/Rp
Ls
Rs
ωM/Rp E
Zs
Cs
Req = (ωM)²Rp / (Rp² + (ωLp)²)
Leq = −(ωM)²Lp / (Rp² + (ωLp)²)
Rs
Req
Leq
Ls
Zs
(ωM / (Rp + jωLp)) E
Effect of associated primary circuit
Cs

Equivalent secondary circuit

When $R_p \gg \omega L_p$

When ωL_p not negligible

FIG. 16.—Equivalent secondary circuit for case of inductive coupling with tuned secondary and untuned primary.

---

3. To shift the phase between voltage and current.
4. To alter the magnitude of voltage or current.

Figure 7 C shows some of the filters with which we are already familiar, together with their curves of output current at different input frequencies, their basic equations, and their names. See also Figs. 5 A and 5 B. More complicated combinations will serve better in separating neighboring frequencies and in more completely suppressing others. In other words, there are filters which have a *sharper cutoff*.

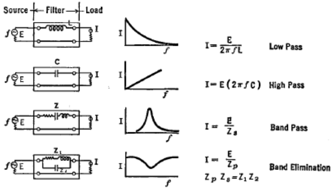

FIG. 7 C. Crude filters. Low-pass types use series inductance (top figure). High-pass use series capacitance. Band-pass use series *LC* circuits. Band-elimination use parallel *LC* circuits (bottom figure)

Of the host of possible *LC* combinations, a few simplified forms have been developed under the following assumptions and simplifications:

1. The resistances of all parts of the filter circuits are to be kept as low as possible. We shall assume that the resistances are zero. Such filters are called " ideal " or *non-dissipative.*
2. There shall be no batteries or other sources of electromotive force, nor vacuum tubes with their attendant amplification, in the filter. Such filters are said to be *passive.*
3. The inductances of all coils which contain iron cores shall be assumed to be the same for any of the currents which flow through them. Filters constructed with such inductances are said to be *linear.*
4. There shall be no magnetic coupling of the lines of force from any of the coils to any other coils in the circuits.

5. The output or load circuit shall consist of pure resistance (no coils or condensers), whose value shall be the same as that of the generator or input circuit.

It is possible to describe the characteristic of filters in terms of the output current as in Fig. 7 C, or in terms of the *ratio* of output to input currents, or the ratio of the output voltage across the load to the input voltage, or the ratio of the power output to the power input, or, as is more common and as we shall now do, in terms of the *attenuation* of the filter. The attenuation is expressed in decibels (Sec. 2.6) and is proportional to the logarithm of the ratio of the output to the input currents or voltages or powers. It is a measure of the losses which take place in the filter at the different frequencies.

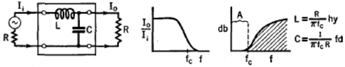

FIG. 7 D. An L-section, low-pass filter. The lower frequencies in the range "A" pass through, while those of higher value are more or less attenuated or lost in the filter

7.5 Low-Pass Filters. The circuit of Fig. 7 D is called an L-type, low-pass filter. It contains a series inductance and a shunt capacitance. From the attenuation (db.) curve in this figure, it can be seen that the choking action of the coil, aided by the bypass action of the condenser, has materially sharpened the curve over that of a single series inductance (Fig. 7 C), especially in the neighborhood of the *cutoff frequency*, f_c. In the equations given in Fig. 7 D, L is in henries, C is in

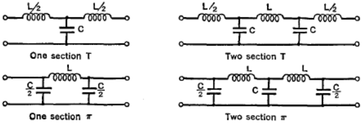

FIG. 7 E. — T- and π-type, low-pass filters

---

resistance of the tube. When the secondary is a resonant circuit, the curve of coupled impedance $(\omega M)^2/Z_s$ with variation in frequency has substantially the same shape and characteristics as the parallel impedance curve of the secondary, but the absolute magnitude is determined by the mutual inductance.

The important characteristic of this circuit is the curve of voltage developed across the secondary condenser as a function of frequency under conditions where $R_p >> \omega L_p$. Such a characteristic is shown in Fig. 15*b*, and has substantially the same shape as a resonance curve with the peak at the resonant frequency of the secondary and an effective Q somewhat lower than the actual Q of the secondary circuit.

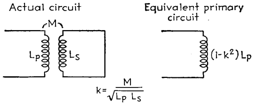

(a) Untuned reactive secondary with zero losses

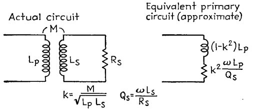
Actual circuit
Equivalent primary circuit (approximate)
M
Lp
Ls
Rs
(1-k²)Lp
$k^2 \frac{\omega L_p}{Q_s}$
$k = \frac{M}{\sqrt{L_p L_s}}$
$Q_s = \frac{\omega L_s}{R_s}$

(b) Untuned secondary with small losses

FIG. 14.—Inductively coupled circuits with untuned secondaries.

On the assumption that $R_p >> \omega L_p$, the response at resonance is

$$\frac{\left\{\begin{matrix}\text{Voltage across secondary} \\ \text{condenser at resonance}\end{matrix}\right\}}{\text{Voltage applied to primary}} = \frac{\omega M}{R_p R_s + (\omega M)^2} \times \omega L_s \tag{30}$$

With given primary and secondary circuits, the response at resonance is maximum when the coupled resistance equals the primary resistance, *i.e.*, when

$$\omega M = \sqrt{R_p R_s} \tag{31}$$

With greater or less coupling, the response is less.

The relationship between the effective Q of the response curve and the actual Q of the secondary circuit is, still assuming that $R_p >> \omega L_p$

$$\frac{\text{Effective } Q \text{ of amplification curve}}{\text{Actual } Q \text{ of tuned circuit}} = \frac{1}{1 + \frac{(\omega M)^2/R_s}{R_p}} \tag{32}$$

The effective Q of the response curve approaches the actual Q of the secondary circuit when the coupling is very small, and is exactly one-half of the actual Q when the coupling corresponds to the value giving maximum response.

Inductive Coupling Considered as a Transformer.—Two coils inductively coupled represent the general case of a transformer, and can be reduced to the equivalent transformer circuit of Fig. 13. Here the primary and secondary inductances are divided into leakage and closely coupled components. The leakage inductances correspond to magnetic flux lines that produce no linkages to any secondary circuit, while the coupled inductances are taken as having unity coefficient of coupling.

In a conventional transformer with iron core, the coefficient of coupling between primary and secondary is so large that the leakage inductances represent only a very small fraction of the total circuit inductances. Under these conditions it is convenient to analyze the behavior in terms of the turn-ratio, leakage inductance, etc., as is commonly done with 60-cycle power transformers. However, in the more general case where the coefficient of coupling may be small, most of the inductance is leakage, and turn ratio has relatively little significance. It is then preferable to use the general method of analysis given above.

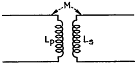

(a) Actual circuit

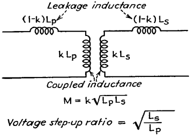

(b) Same circuit shown as transformer

FIG. 13.—Equivalent circuit of a transformer expressed in terms of primary and secondary inductances and coefficient of coupling.

4. Analysis of Some Simple Coupled Circuits. *Untuned Secondary Consisting of a Pure Reactance.*—This case corresponds to a coil serving as the primary, with a metal mass of low resistivity serving as the secondary. In this case, which is illustrated in Fig. 14*a*, one has

$$\text{Coupled reactance} = -\frac{(\omega M)^2}{\omega L_s} = -k^2\omega L_p \qquad (29)$$

where k is the coefficient of coupling between the coil and the secondary. The presence of the secondary reduces the inductance as viewed from the coil terminals to an equivalent value $(1 - k^2)L_p$.

It will be noted that the amount of reduction in inductance is determined solely by the coefficient of coupling and not by the number of primary turns, etc. It will be also observed that the effect of the secondary decreases rapidly as the coefficient of coupling becomes small.

Untuned Secondary Having Both Resistance and Reactance.—This corresponds to the practical case of a coil in a shield, or near a metal panel or other mass, with the resistance of the metal secondary taken into account. For such a secondary, the coupled impedance as given by Eqs. (24) and (25) consists of a resistance and a negative reactance. The effect of the secondary upon the primary is hence to lower the equivalent inductance and increase the equivalent resistance of the primary coil as viewed from its terminals.

Under practical conditions, a secondary consisting of a shield or other metal object is usually of copper or aluminum, so that the secondary reactance is considerably greater than the secondary resistance. Under these conditions, the equivalent primary circuit has the values and form given in Fig. 14*b*.

Tuned Secondary, Untuned Primary.—This circuit, illustrated in Fig. 15*a*, is of importance because it represents the equivalent circuit of the transformer-coupled tuned radio-frequency amplifier, with the primary resistance R_p representing the plate

---

farads, R is in ohms, and f_c is in cycles per second. Note that the generator and the load both have the same resistance.

Referring to Fig. 7 E, we see that several " T "-sections or " π "-sections can be connected in series with each other to form two, three, or more section filters. The cutoff becomes increasingly sharp the more sections used. Then, of two frequencies near f_c, the lower one will get through and the higher one will not. The same design equations are used as for the L-section low-pass filter. Note that, in Fig. 7 E, only one-half the inductance or one-half the capacitance is used in certain places while the full value is used at others.

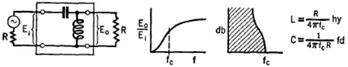

FIG. 7 F. An L-section, high-pass filter. L is in henries, C is in farads, R is in ohms and is the same at the input and output. The cutoff frequency, f_c, is in cycles per second

7.6 High-Pass Filters. Whereas the low-pass filters of the preceding section permitted the passage, more or less, of all frequencies below the cutoff value, and attenuated or suppressed, more or less, all frequencies above this value, high-pass filters do just the reverse. A single L-section high-pass filter is shown in Fig. 7 F and is seen to con-

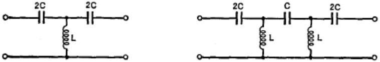

One section T　　Two section T

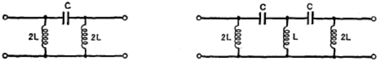

One section π　　Two section π

FIG. 7 G.—T- and π-type, high-pass filters

sist of a series condenser, whose reactance is large for low frequencies and small for higher frequencies, and a shunt inductance, which readily bypasses low frequencies but forces the higher frequencies to continue on out to the load R.

Figure 7 G shows T- and π-type, one- and two-section high-pass filters which offer greater discrimination to frequencies near the cutoff. The design equations are the same as for the L-section.

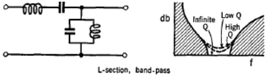

L-section, band-pass

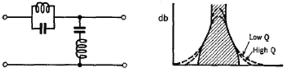

L-section, band-elimination

FIG. 7 H. Simple band-pass and band-elimination filters

7.7 Band-Pass and Band-Elimination Filters. It is possible to transmit a certain range of frequencies through a *band-pass filter*, and more or less completely suppress all lower and all higher frequencies. A simple filter of this type, together with its attenuation curve, is given in Fig. 7 H.

It is also possible to construct filters which will, more or less, transmit all frequencies except those which lie within a certain range. A simple form of *band-elimination* or *band-suppression filter* is also given in Fig. 7 H, together with its attenuation curve.

With multi-section filters of these types, the separation of adjacent frequencies can be made much sharper than with the simple L-section. On the other hand, in practice, the coils are not pure inductances but have some resistance. Now Q is our measure of the inductive-ness versus resistance of a coil, i.e., $Q = 2\pi fL/R$. In Fig. 7 H, the dotted lines show in what manner the attenuation is affected when low- and high-Q coils are used. Similar unsharpening effects occur in practice with all filters.

---

circuit leads to the same results as the first form, since the extra voltage that is considered as acting in the second case is just compensated for by the voltage drop in the extra impedance considered as existing in the secondary circuit.

The resistance and reactance components of the coupled impedance are

$$\text{Resistance component} = \frac{(\omega M)^2}{R_s^2 + X_s^2} R_s \tag{24}$$

$$\text{Reactance component} = -j \frac{(\omega M)^2}{R_s^2 + X_s^2} X_s \tag{25}$$

where R_s and X_s are, respectively, the resistance and reactance components of the secondary impedance Z_s. The effect that the secondary has on the primary circuit is exactly as though the resistance and reactance given by Eqs. (24) and (25) had been inserted in series with the primary circuit. The energy and reactive volt-amperes consumed by the primary current flowing through this hypothetical resistance and reactance represent the energy and reactive volt-amperes that are transferred to the secondary circuit.

Conditions for Maximum Secondary Current.[1]—In dealing with coupled circuits, it is desirable to keep in mind the adjustments giving maximum secondary current. The behavior in the more important cases is summarized below for the case of two circuits inductively coupled, with resistance and reactance in each circuit.

Case 1. Maximum Secondary Current with Variation of Secondary Reactance.—When frequency, mutual inductance, primary impedance, and secondary resistance are constant, the secondary current is maximum when

$$X_s = \frac{(\omega M)^2}{R_p^2 + X_p^2} X_p \tag{26}$$

where X_p and X_s = reactance of primary and secondary circuits, respectively.
R_p and R_s = resistance of primary and secondary circuits, respectively.
ω = 2π times frequency.
M = mutual inductance.

Case 2. Maximum Secondary Current with Variation of Primary Reactance.—If the primary rather than secondary reactance is varied, maximum secondary current is obtained when

$$X_p = \frac{(\omega M)^2}{R_s^2 + X_s^2} X_s \tag{27}$$

Case 3. Maximum Possible Secondary Current.—The secondary current has its maximum possible value when the adjustments are such that the resistance and reactance coupled into the primary circuit by the secondary are equal to the resistance of the primary circuit and the negative of the primary reactance, respectively; *i.e.*, when

$$\begin{aligned} \frac{(\omega M)^2}{R_s^2 + X_s^2} R_s &= R_p \\ -\frac{(\omega M)^2}{R_s^2 + X_s^2} X_s &= -X_p \end{aligned} \tag{28}$$

It will be noted that in order to obtain maximum possible secondary current, it is necessary that two independent variables be adjusted. One of these is commonly the mutual inductance; the other, the reactance of one or both circuits. Unless the mutual inductance satisfies the relation $\omega M \geqq \sqrt{R_pR_s}$, it is impossible to satisfy the condition for maximum possible secondary current.

---

[1] For a more detailed discussion, see G. W. Pierce, "Electric Oscillations and Electric Waves," Chap. XI, McGraw-Hill, New York, 1920.

circuit as a result of a voltage applied to that circuit produces magnetic flux that induces a voltage in the coupled circuit, resulting in induced currents and a transfer of energy from the first or primary circuit to the coupled or secondary circuit.

Analysis of Inductively Coupled Circuits.—The behavior of inductively coupled circuits can be determined by making use of the following principles:[1] *First,* as far as the primary circuit is concerned, the effect of the presence of the coupled secondary circuit is exactly as though an impedance $(\omega M)^2/Z_s$ had been added in series with the primary, where

M = mutual inductance.
$\omega = 2\pi f$.
Z_s = series impedance of secondary circuit when considered by itself.

Second, the voltage induced in the secondary circuit by a primary current of I_p has a magnitude of ωMI_p and lags the current that produces it by 90°. *Third,* the secondary current is exactly the same current that would flow if the induced voltage were applied in series with the secondary and if the primary were absent.

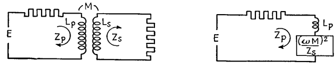

(a) Inductively coupled circuit (b) Equivalent primary circuit

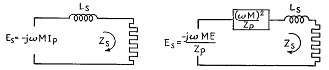

(c) Equivalent secondary circuit (d) Alternative form of equivalent secondary circuit

FIG. 12.—Inductively coupled circuit and also equivalent primary and secondary circuits.

The preceding analysis shows that the actions taking place in the primary and secondary circuit are as though one had the equivalent circuits shown in Figs. 12*b* and 12*c*. Here the primary circuit consists of the actual primary impedance Z_p in series with an additional impedance $(\omega M)^2/Z_s$ that takes into account the effect the secondary produces on the primary current. This extra impedance $(\omega M)^2/Z_s$ is termed the *coupled impedance.* The equivalent secondary circuit of Fig. 12*c* consists of the voltage $-j\omega MI_p$ induced in the secondary by the primary current, in series with the secondary impedance Z_s as shown.

An alternative form of equivalent secondary circuit is given in Fig. 12*d*. This is obtained by applying Thévenin's theorem and noting that when the secondary is opened, the open circuit voltage that appears is $-j\omega MI_p$, while when the applied voltage is short-circuited, the equivalent impedance as viewed from the opened secondary terminals is $Z_s + (\omega M)^2/Z_p$. This second form of the equivalent secondary

---

[1] These rules for handling coupled circuits are direct consequences of the mesh equations, which for this case can be written as

$$E = I_pZ_p + j\omega MI_s \tag{22}$$
$$\text{Induced voltage} = -j\omega MI_p = I_sZ_s \tag{23}$$

It is well known that a quartz slab cut from its crystal in proper fashion and mounted between two metal plates acts like a circuit of very high Q. Hence, in circuits where the frequency of the currents is comparable with the natural vibrational frequency of the quartz, it is possible to obtain very sharp cutoff filters of all types.[1]

---

[1] Details are given by W. P. Mason and R. A. Sykes, in *The Bell System Technical Journal,* Volume XIX, page 221, April 1940.

Radio Engineers Handbook
F.E. Terman, 1943
Section 3: Circuit Theory
pp 148 - 163
Inductively Coupled Circuits

3. Inductively Coupled Circuits.—When mutual inductance exists between coils that are in separate circuits, these circuits are said to be inductively coupled. The effect of the mutual inductance is to make possible the transfer of energy from one circuit to the other by transformer action; *i.e.*, the alternating current flowing in one

Lee, Reuben: 1947.
Electronic Transformers and Circuits
http://www.vias.org/eltransformers/lee_electronic_transformers_07b_22.html

Storr, Wayne
Mutual Inductance of Two Coils
http://www.electronics-tutorials.ws/inductor/mutual-inductance.html

Tuggle, R.M.
Notes: Terman Inductive Coupling Model:
Inductively Coupled Circuits -- Terman, p. 149, ff.

Van Warren, L, AE5CC
Crystal Radio
http://www.wdv.com/Electronics/SoftRadio/BlazingFastSDR/BFSDR-Chapter1.docx

Weaver, Bob: 2016
Bandspread Calculations - Part 3
http://electronbunker.ca/eb/Bandspreading_3.html

kjs 8/2016

In The Beginning… Crystal Radio

By L. Van Warren MS CS, AE - AE5CC
All Rights Reserved
www.wdv.com/.../SoftRadio/BlazingFastSDR/BFSDR-Chapter1.docx

“Every day sees humanity more victorious in the struggle with space and time.”
– Guglielmo Marconi

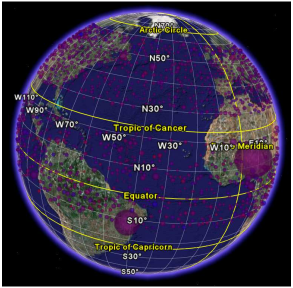

Figure 1: Earth Footprints of Celestial Radio Sources

Introduction

There are lots of books, articles and websites describing Software Defined Radio (SDR). My goal, in this introductory work, is to give you blazing fast access to a working set of concepts you can use to decide when and how SDR will be useful to you. It will start simply and build essential ideas step-by-step. This book has two goals. The first is to provide a working overview of SDR. The second is to make hardware and software prototyping easier for the uninitiated.

This will not be a mathematically intensive development but rather a plug and "play" approach. Each chapter will start with interactive simulation and end with real devices - devices you can explore and interconnect. The interested reader should visit the references provided in the final chapter to clarify the more sophisticated ideas. Running each simulation is easy and highly recommended.

The book is divided roughly in half. In the first chapters, essential radio hardware issues will be discussed. For the foreseeable future SDR has not eclipsed the entire radio. Front-end RF hardware is still required to gather, sample and downconvert the signal. In the latter chapters we will transition to software-based concerns, while keeping an eye on hardware and instrumentation that will make our lives easier and our understanding more complete.

To demonstrate hardware concepts, we will be using a set of PostCardKits™. The pattern is this. We will use simulation to understand the theory behind each PostCardKit™. Then we mix and match the postcards to configure different kinds of radios. Pretty fun and exciting! Later we will mix and match software blocks to accomplish the same objective.

http://www.electronics-tutorials.ws/accircuits/parallel-circuit.html

Ben Gurion University
Mutual Inductance
Pdf downloaded from BGU Lab 7 web page
http://www.ee.bgu.ac.il/~intrlab/lab_number_7/

Ben Gurion University
EE Lab 7
http://www.ee.bgu.ac.il/~intrlab/lab_number_7/Two%20inductively%20coupled%20RLC%20circuits.pdf

Casio Computer Co., Ltd, 2016
Impedance of R, C and L in parallel Calculator
http://keisan.casio.com/exec/system/1258032708

DesignSoft, Inc.
Resonant Circuits
http://www.tina.com/English/tina/course/28resonant/resonant.htm

Georgia State U: HyperPhysics - Electricity and Magnetism
RLC Parallel Circuit
http://hyperphysics.phy-astr.gsu.edu/hbase/electric/rlcpar.html

Hoag, J Barton, 1942
Basic Radio
http://www.tubebooks.org/Books/hoag.pdf

Kuphaldt, Tony R.
Resonance
http://www.allaboutcircuits.com/vol_2/chpt_6/1.html

where Ze is the equivalent primary impedance
Z1 is the source impedance (primary) and
Z2 is the secondary load.

From Reuben Lee, 1947.
http://www.vias.org/eltransformers/lee_electronic_transformers_07b_22.html

Equivalent primary impedance:
Ze = Z1 + (XM^2 / Z2)

where Ze is the equivalent primary impedance
Z1 is the primary circuit impedance and
Z2 is the secondary circuit impedance and
XM = jwM

From Bob Weaver: 2016
http://electronbunker.ca/eb/Bandspreading_3.html

Net impedance:
Ze = Zca || (Z1 + ((wM)^2 / (Z2 + Zcb))

where Zca Primary circuit Z capacitor = -j / wCa
Zcb Secondary circuit Z capacitor = -j / wCb
Z1 Primary circuit Z inductor = j wL1
Z2 Secondary circuit Z inductor = j wL2

Bibliography:

AspenCore , 2016
The Parallel RLC Circuit

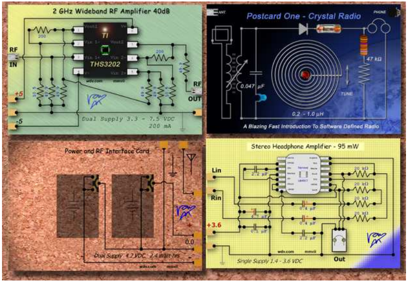
2 GHz Wideband RF Amplifier 40dB
THS3202
RF IN
RF OUT
+5
-5
Dual Supply 3.3 - 7.5 VDC 200 mA
Postcard One - Crystal Radio
0.047 µF
0.2 - 1.0 µH
TUNE
47 kΩ
A Blazing Fast Introduction To Software Defined Radio
Power and RF Interface Card
Dual Supply 4.2VDC 2.4 Watt hrs
Stereo Headphone Amplifier - 95 mW
Lin
Rin
+3.6
20 kΩ
Out
Single Supply 1.4 - 3.6 VDC
wdv.com

Figure 2: Mix and Match Postcards for Chapter 1

In the second chapter on amplifiers the crystal radio card is improved using an audio amplifier PostCardKit™, an RF amplifier PostCardKit™, and two kinds of power cards. One power card features rechargeable lithium batteries; the other uses solar cells for recharging and direct power. How green is that?! PostCardKits™ are flat, lead-free evaluation cards, printed on high quality paper with conductive ink. PostCardKits™ can be stamped and mailed, or mailed in envelopes to maintain pristine appearance. The first chapter is introduced with a crystal radio set. This card functions without any external or battery power. It receives AM radio stations. A first attemp on a file card pulled in stations from Asia and Central America.

You can hear fainter stations if you add an audio amplifier card. You can receive more stations if you add the RF amplifier. These additional cards require power. We will reuse the audio, RF amplifier and power cards in later chapters in novel ways. For example, the audio card is a stereo amplifier used for a special kind of station hunting called binaural radio.

Laterwe will develop radio software on a PC. Towards the end, we will extend the power of the hardware and software and reach for the stars.

The first PostCardKit™, Crystal Radio, utilizes a germanium diode for signal detection. It demonstrates the simplest effective combination of discrete components. It consists of an inductor, a capacitor, a resistor, a germanium diode detector and a piezoelectric crystal earphone.

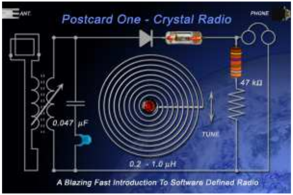
ANT
Postcard One - Crystal Radio
PHONE
47 kΩ
0.047 µF
TUNE
0.2 - 1.0 µH
A Blazing Fast Introduction To Software Defined Radio

Figure 3: Crystal Radio PostCardKit™

Don't let our simple start fool you; We will be moving many these functions into software and Software Defined Radio (SDR) can do sophisticated things.

$(A + jB) \;/\; (C + jD)$

1)

$$\frac{(A+jB)}{(C+jD)} * \frac{(C+jD)}{(C+jD)} = \frac{AC + jAD + jBC + j^2BD}{C^2 + D^2}$$

Remember that $j^2 = -1$, so

2)

$$\frac{(A+jB)}{(C+jD)} * \frac{(C+jD)}{(C+jD)} = \frac{AC + jAD + jBC - BD}{C^2 + D^2}$$

3)

$$\frac{(A+jB)}{(C+jD)} * \frac{(C+jD)}{(C+jD)} = \mathbf{\frac{AC - BD}{C^2 + D^2} + j\,\frac{AD + BC}{C^2 + D^2}}$$

Parallel Combination

$$\mathbf{Zeq = (Z1*Z2) / (Z1+Z2)}$$

Appendix 2:

Coupled Impedance Formulae:

From Ben Gurion University EE Lab 7
http://www.ee.bgu.ac.il/~intrlab/lab_number_7/Two%20inductively%20coupled%20RLC%20circuits.pdf

Equivalent primary impedance:
$$\mathbf{Ze = Z1 - ZM^2 / Z2}$$

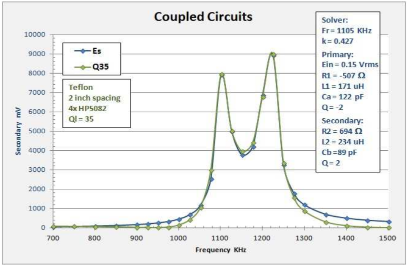

Just for fun, I let solver run with k, Ca, E, Cb, R1 and R2. Adding R1 to solver's program really allowed it to nail the data with a close-fit model. The price? Look at the resistivity values it chose. R1 = -439 ohms and R2 = 719 ohms. Well, at least R2 is not negative! These results may be telling me something important, but as yet I am too dense to understand. Or, they may be just what they appear, totally bogus.

Appendix 1:

Big Fun with Complex Algebra

Multiplication
(A + jB) * (C + jD) = **(AC – BD) + j(AD + BC)**

Division
multiply both top and bottom by the conjugate of the bottom.

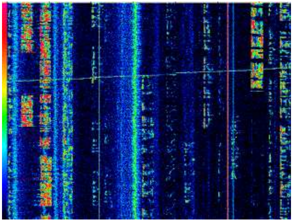

Figure 4: Visualization of Radio Spectrum

The Crystal Radio PostCardKit™

The crystal radio is the simplest of all radios. In World War II, Allied GI's used paper clips set against rusty razor blades to form crude diode junction receivers that the Nazi forces could not detect. These were dubbed "Foxhole receivers". Crystal radios have a long and colorful history documented in Wikipedia and various radio collections documented on the web.

Attach the earphone provided to the jack in the upper left corner of the postcard. You probably won't hear anything unless you live close to a powerful AM radio station. An antenna and a ground will improve your reach considerably. Just as a picture in the dark cannot be seen, a radio without an antenna cannot be heard. Lighting is half of art. The antenna is half of radio. You can learn more about antennas in the ARRL Antenna Book. It is highly recommended.

Putting Up the Litz

There is one errand to run before heading back to the easy chair. It is essential to route the Litz wire provided around a wall or ceiling to create an antenna. With antennas, bigger is usually better. I use a fold of masking tape to make a tiny hangar that holds the antenna on the wall. You can stick a clear pushpin through the tape to secure the antenna. Suspend the wire from the four corners of the room so it is up and out of the way. The wire provided is fine, so it is a quite aesthetic. When you are done, tin or sand the ends of the wire so that all the strands are conducting and install them in the connector provided. Now you have an antenna.

This square wire loop is a versatile omnidirectional antenna. If you wrap the antenna more than once around the room, the inductance will increase and the resonant frequency will drop according to:

$$f_{resonant} = \frac{1}{2\pi\sqrt{LC}}$$

Formula: Resonant Frequency

This formula informs us that those stations you manage to receive will be lower in the band, and lower in frequency. Start with one trip around the average sized room. Take your time getting this antenna right, it will serve you well. If you live on top of a hill, you will get better reception, but since radio waves bounce off the ionosphere, you will usually hear something unless you live in a salt mine. Using the clip provided, attach your antenna to the upper left hand corner your PostCardKit™ by the ANT. symbol. That was the hard part. You will also need a good ground. Grounding is

happens at the -3dB point. My simulation suggests a resistive loss of 18 ohms and reproduced the Q = 83. Pretty interesting. But, before congratulating myself too quickly...

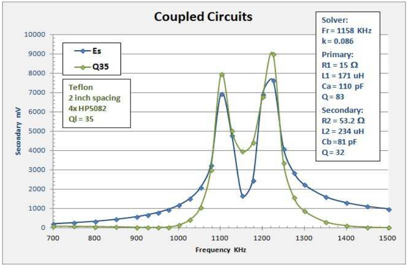

Here I look at the over coupled case. As before, I am just unable to reproduce the steep flanks of my measured data with such a simple coupled circuit model. At a 2 inch separation, and looking only at the lower peak (recall my re-peaking), I had measured a Q = 35. The model does a fairly poor job of simulating this circuit. Note that in my simulation, I kept the known values of inductance for the two coils. Additionally I found that solver gave unacceptable values for R1 and so I pegged it at 15 ohms. I had solver look for k, Ca, E, Cb, and R2.

The first thing to note on the above plot is that the over-coupled curve is not centered with respect to the loosely coupled curves. This is a result of my A) not knowing at the time about exactly how the resonance curve splits with over coupling and B) my technique of re-peaking the set before each set of measurements. It is obvious that after moving the ATU to 2 inches separation, on re-peaking the output, I must have subtracted capacitance shifting the curve upward in frequency until the I got a peak at my desired measurement frequency of 1100 KHz. I should have made my resonance study earlier!

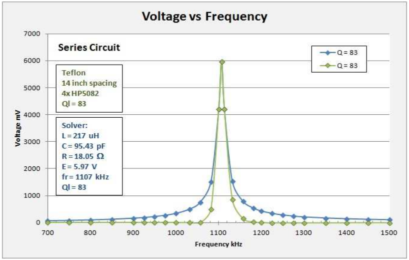

Looking now at just the measured data from the 14 inch spacing and comparing it to a curve generated with my series resonance spreadsheet. The data in green is my radio which, at this spacing gave me a set loaded Q = 83. I used excel's Solver routine on the central three datapoints to search for model parameters L, C and R. The steep flanks of the measured data were impossible to reproduce, but at least I could see what

discussed in an essential book on radio: The ARRL Radio Amateur's Handbook.

Figure 5: Ideal Reception

Using the Crystal Radio

Now that the radio has a good antenna, you should be able to hear more stations. You can tune the radio with the knob in the center. Remarkably, it needs no power! You might want to keep a logbook of the signals you hear, the time of day along with any tweaks you have made to the radio or antenna. Low frequency signals travel better at night than in the daytime. Some high frequency signals are the opposite. Where do we look when an aircraft is lost? The radio logbook.

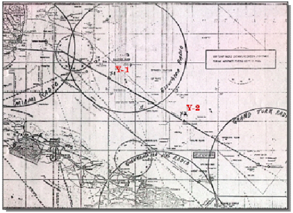

Figure 6 – Keeping a log enables discovery.

After you log a few entries it will be time to improve the radio. We will do that with the amplifiers mentioned above. Now for a little about how the crystal radio works.

Basic Parts

Picking up the card and shining the light on it you will notice it contains but five parts! The radio contains a diode, a resistor, an inductor, and two capacitors. Here is the schematic for the crystal radio including the Tidy TINA meter for RF power gain. The meter is used to optimize performance – it doesn't appear in the final circuit.

should give pause to note just how little signal one deals with in crystal sets. Note that critical coupling is when the secondary current attains its greatest value, not necessarily where the double hump disappears. Kc = 1/sqrt(Qp*Qs).

A copy of the coupled circuit spreadsheet can be found here: Coupled_kjs.xlsx

What else can one do with these spreadsheets? One of my motivations in building spreadsheets is that it allows me to compare measurements from actual radios to their theoretical counterparts. This allows me to better understand my sets and approach a best guess on certain parameters I might not otherwise be able to measure. The following few plots take data from a double tuned set that I have made frequency versus DC out at three differing coupling distances, 2 inches between coils (over coupled), 8 inches (presumed near critical), and 14 inches (under coupled).

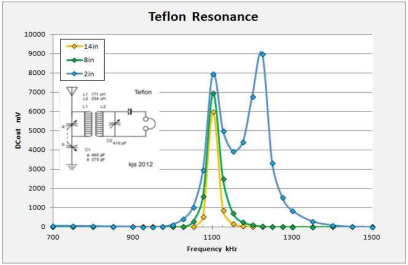

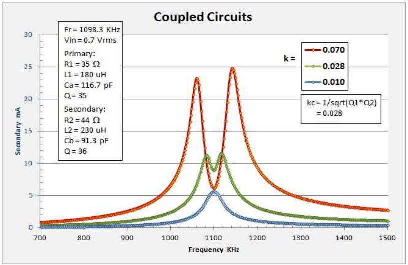

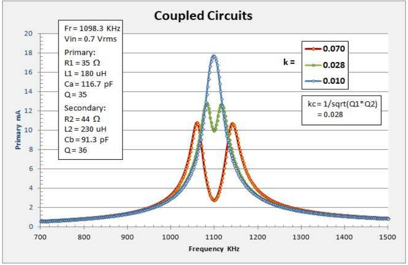

In each plot I show curves with coupling factors of 0.070 (over-coupled), 0.028 (critically coupled), and 0.010 (loosely coupled). I find it amazing just how quickly the induced secondary current falls off with lowered coupling. These plots look quite different from most "generic" cartoon plots and

---

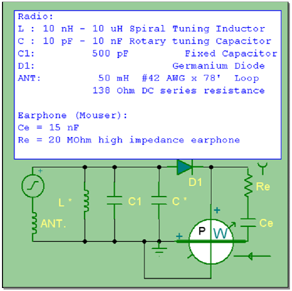

```
Radio:
L : 10 nH - 10 uH Spiral Tuning Inductor
C : 10 pF - 10 nF Rotary tuning Capacitor
C1:          500 pF            Fixed Capacitor
D1:                         Germanium Diode
ANT:          50 mH  #42 AWG x 78'  Loop
             138 Ohm DC series resistance

Earphone (Mouser):
Ce = 15 nF
Re = 20 MOhm high impedance earphone
```

Example 0: Crystal Radio Schematic and Values

Just as in the Richter scale of earthquakes and the Fujitsu scale of tornadoes, we use a logarithmic scale when comparing the intensity of radio signals. This scale is measured in decibels (dB). This makes for much more reasonable comparisons. If two signals differ by a factor of two, they are about 3 dB apart. If they differ by a factor of four, they are 6 dB, and so on. Logarithmic scales turn multiplication into addition. This is useful when we want to talk about very large or small numbers. To convert a factor of 1000 to dB, you first count zeros to get 3. This corresponds to log(1000) = 3. Then you multiply by 10. 3 x 10 = 30 dB. So if two signals differ in power by a factor of 1000, then they are 30 dB apart:

In short dB = 10log(P), where P is power. Can you feel the power?

Tip - Use dB to compare the power of radio signals.

The resistor Re and capacitor Ce simulate the earphone. To really understand the crystal radio, we must understand the principles of the parts. If you are already an expert skip this quick review, but you might want to glance at the gain curves for voltage, current and power.

Resistors (units: Ohms) dissipate energy as heat. They impede the flow of electrical current, causing a voltage drop across the terminal ends. I once asked my dad if it wouldn’t be better if a circuit had no resistors at all because of this energy loss. He said “No” and then paused for a moment and said, “Yes”. The voltage drop E across a resistor is R times the current I, using Ohm’s law. You can think of an Ohm of resistance as the volt of force required to make an ampere of current flow.

$$\boxed{E=R\times I}$$

Formula: Ohm’s Law

Power (Watts) is voltage times current. Is your resistor rated for the power passing through it? Touch it and see, but don’t get burned.

$$\boxed{P=E\times I}$$

Formula: Power

With Ohms Law and Power, you can derive six others! Two other handy resistor formulas:

Xs = wLs - 1 / (wCs)
Rcoupl = (wM)^2 * Rs / (Rs^2 + Xs^2)
Xcoupl = -1 * (wM)^2 * Xs / (Rs^2 + Xs^2)
Real = Rp + Rcoupl Imag = wLp + Xcoupl - 1/(wCp)
w = 2pi * frequency

Where the subscripts "s" means secondary circuit and "p" means primary circuit. M is the Mutual Induction of the two circuits defined as: M = sqrt(Lp*Ls) and is measured in Heneries. The Coupling Coefficient k indicates the amount of inductive coupling. This is a fractional value between 0 (no coupling) and 1 (perfect coupling). The Coupling Factor for the circuit then is defined as follows: M = k * sqrt(Lp*Ls).

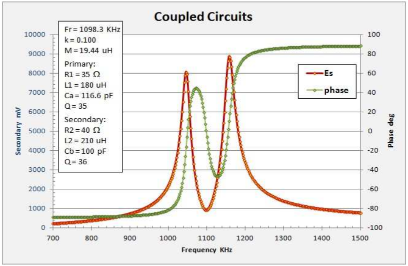

The plot above presents a generic coupled circuit with loose coupling (k = 0.1). I show the current induced in the secondary and the phase of the circuit impedance. Below are two interesting plots with the secondary current and primary current each at three different coupling factors.

value, reactance consists of inductive XL and capacitive XC parts. The parts are not necessarily the same and so adding them together does not necessarily equal zero. Note that, at resonance, they are equal in magnitude but opposite in sign (by convention, XC is negative) and they do add to zero. Where they have differing magnitudes, the combined impedance Z is the vector sum of the resistance and total reactance (XL + -XC). Z is the hypotenuse of a right triangle with an interior angle phi. This is the phase of the circuit. Where the phase is negative the circuit is capacitive, where the phase is positive the circuit is inductive. Where XL-XC is very large with respect to R, the phase approaches (but never reaches) plus or minus 90 degrees.

A copy of the parallel circuit spreadsheet can be found here: Parallel_kjs.xlsx

Time to move to the big leagues with coupling of two circuits. Here I am afraid my skills prove inadequate and I have turned to Mike Tuggle for help and, ultimately, a spreadsheet to make the needed calculations. For this I wish to publically acknowledge and thank Mike. Also to apologize to him for the significant liberties I have taken with said spreadsheet.

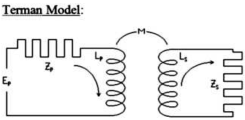

The model is based on Terman 1947 p149, shown at left. Mike builds a number of equations, the essential ones of which I used as follows:

Add two resistors in series to obtain the equivalent resistance:

$$\boxed{R_{series} = R_1 + R_2}$$

Formula: Series Equivalent Resistance

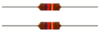

Use the product over sum for resistors in parallel:

$$\boxed{R_{parallel} = \frac{R_1 R_2}{R_1 + R_2}}$$

Formula: Parallel Equivalent Resistance

The current flow in circuit loops and the voltage drop across circuit elements can be computed using Kirchoff's Laws and Thevenin Equivalent circuits. The programming of these laws is already done for you in a tidy program called TINA-TI, a free download from the TI web site. I highly recommend it. Here is a classic voltage divider, simulated in TI's TINA-SPICE

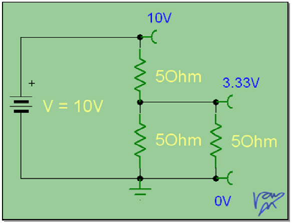

Circuit 1: Classic Voltage Divider Solved in Tidy TINA.

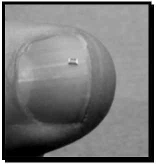

- Ordy

Figure 7: No room for color codes on surface mount resistors!

w = 2pi * frequency

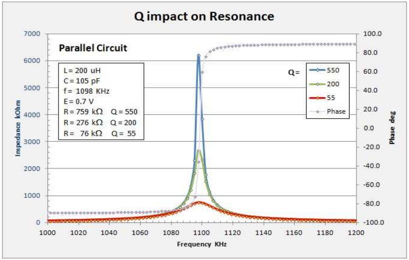

This plot above for a parallel circuit looks, as it ought, just like that for the series circuit with the same values of L and C. Note though that parallel R is significantly higher than in the series case. Parallel resistance is not a measure of the circuit loss but rather how "tight" it is, (spoken as a non-engineer). You should see that Q values now go up with increasing value of parallel resistance.

Additionally in this plot I include the phase of the impedance (for the Q = 550 case). The diagram at left indicates why it is necessary to keep track of the real (resistance R) and imaginary (reactance X) parts of the equations. While resistance is a "real"

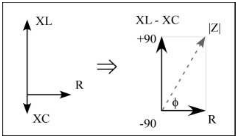

plot on the chart the current versus frequency for the same tank with three variations of resistivity which gives three different values of Q. It is immediately clear that keeping resistance in the circuit low good of the quality of the set.

A copy of the series circuit excel spreadsheet can be found here: Series_kjs.xlsx

Having proved that I can simulate a simple series circuit it is time to take it up a notch with a PARALLEL circuit. Analysis of a parallel is inherently more complicated as it involves imaginary, or complex numbers. It is necessary to keep close track of the real and imaginary parts of the calculations. Also, the math dealing with complex numbers is different, especially the rules for multiplication and division. See my Appendix at the bottom of this page.

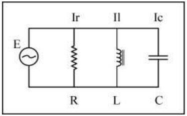

I take the tack of starting with the impedances (Z) of the components and combining them in parallel. The symbol "||" means to combine in parallel. Formulas for this computation as follows:

Zs = Zc || Y
R || Zl
Y = (R * Zl) / (R + Zl)
Zs = (Zc * Y) / (Zc + Y)

Where:
Zl = jwL and
Zc = -j / wC

Notes on Resistors.
1) Always measure the value of a resistor before using it in a prototype circuit. Make sure your volt-ohm meter has a fresh battery.

2) In radio sections that operate at high frequencies we want resistors whose value does not vary with frequency. Thin-film and metal film resistors are preferred to wirewound resistors, which are really just lossy miniature inductors! Speaking of inductors…

Inductors (units: Henries)
store energy as a magnetic field. They are usually coils of wire or other conductive material in various shapes. Inductors have a direct current (DC) response and an alternating current (AC) response. These responses can be steady state or transient. Let's throw the switch!

When the switch is closed on the circuit below, an equal and opposite voltage is "induced" in the inductor. This is induced voltage is called "back EMF". After several time constants, the circuit reaches its "steady state". The magnetic field is established and this opposing voltage disappears. If the switch is opened, the magnetic field collapses and sparks can ensue! There was an old saying, "nature abhors a vacuum". Magnetic fields don't like suddenly open switches. This is an important principle when working with sensitive semiconductor components. Fried!

To track voltage in Tidy TINA we add a pin connection . To track current we add an arrow connection seen at the bottom of the circuit. By convention positive current flows from positive to negative. Electron vacancies or “holes” move in this direction, but real electrons flow the other way. Thanks Ben Franklin!

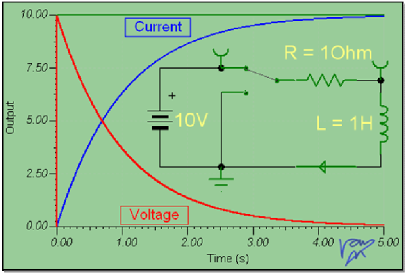

Circuit 2: Inductor Transient DC Response

Tidy TINA shows us two curves if we request a Transient Analysis. The top curve shows the current in the circuit. Since a coil is a conductor, keeping the switch on drains the battery. The bottom curve shows the induced voltage as it decays over time. This transient DC response ends with the steady-state DC response. What about AC, the stuff of which radio signals are made?

Consider the same circuit as before, but this time, we replace the DC battery with an AC signal generator and simplify the circuit to obtain:

Starting with a series RLC circuit. This circuit is straightforward and the math is pretty easy. The formula for the current through a SERIES circuit is:

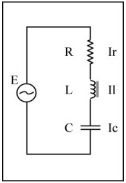

I = V / sqrt[R^2 + (XL - XC)^2]

Since reactances XL = 2pifL and XC = 1/2pifC, the above formula becomes:

I = V / sqrt[R^2 + (2pifL - 1/(2pifC))^2]

The resonance response with input values often discovered in a crystal set, L = 200 uH, C = 105 pF, and R = 6.9 ohms peaks at a frequency of 1098 KHz with a Q factor of 200.

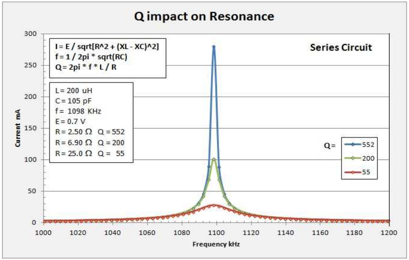

The above plot shows first that it is indeed straightforward to evaluate a circuit in excel, a series circuit anyway. Further, it gives an indication as to what visualizations can be made. I increases with frequency while capacitive reactance decreases. Resonance is that frequency where the two reactance's cancel each other leaving only resistance in the circuit. A little fun derivation may be in order.

Looking at reactance's where f = frequency, L = inductance, and C = capacitance:

XL = 2pifL ... Inductive reactance
XC = 1/2pifC ... Capacitive reactance, so:
2pifL = 1/2pifC ... condition of resonance

taking f out of the denominator:
2pif^2L = 1/2piC

divide both sides by 2piL
f^2 = 1/2^2pi^2LC

take the square root of each side
f = sqrt(1) / sqrt((2pi)^2LC)

and simplifying...

f = 1 / 2pi sqrt(LC)

This is the general formula for resonance.

In all the plots from here on I will show sweeps of frequency with impedance or current as the dependant variable. The point of resonance on each plot should be apparent as that point where the dependant variable reaches its maximum value.

The AC signal causes the inductor's magnetic field to repeatedly collapse and expand in alternating directions. Ohm's law is constantly running, but now there is a delay. This delay is caused by the union of magnetic field workers whose boss is Maxwell and whose contact is binding. Forget that. Remember this. Voltage E Leads Current I in an inductor.

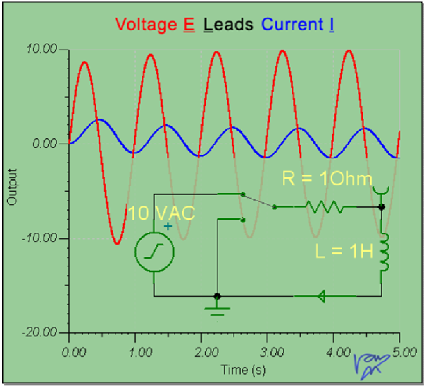

Circuit 3: Inductive AC Response

ELI the ICE man reminds us that voltage leads current by 90□ in an inductor. This is called phase shift. Radio is all about keeping track of phase. We draw voltage in red and current in blue on the same graph so we can see their relationship in time. But what about frequency?

Frequency Response: What Happens When Inductance and Resistance Change?
Now we can play with the inductor and resistor values and see what happens in our circuit. We will measure this by comparing the gain of various configurations. Gain is the amplitude of the signal in a circuit. Gain comes in three flavors, voltage gain, current gain and power gain. Power gain is voltage gain times current gain. We want to know how the gain changes as we change the frequency of our input signal. We can determine this quickly with Tidy TINA. First, we fix the resistor at 1 Ohms and set the inductor to 1 microHenries. Then we ask TINA to compute the AC Transfer Characteristics. Voila! We get a graph that yields major insight.

A graphical look at Resonance

Kevin Smith

http://www.lessmiths.com/~kjsmith/crystal/resonance.shtml

I write this page in order to review, in my own mind mostly, those things which go into radio frequency resonance. My goal was to cast into spreadsheet form the formulas and equations that determine resonance for simple circuits that one often encounters in crystal radio.

On may ask why create spreadsheets to do what SPICE or Mathematica can easily do already. As I am addressing a crowd who enjoys building their own radios when such can easily and cheaply be had at a local store, I think the answer to my rhetorical question is clear. But more than just "doing it myself", there are good reasons to do so. I have read numerous textbook and online discussions of resonance. Most will feature "resonance curves" either generic cartoons or SPICE plots with no indication of the actual formulas that created the plot. I start to think of SPICE as a kind of black box that magically produces interesting results even to those with little to no understanding of the math inside.

Additionally, those occasional plots with actual scales are for values and magnitudes of R, L, and C that have little to do with medium wave broadcast band work. In order to set or adjust my expectations, I prefer to see things in the scales I typically deal with. Finally, while SPICE and other software may be essential to evaluate complicated circuits, those typically found in crystal radio are really not beyond spreadsheet work.

Something about resonance. An RLC circuit contains both inductive L and capacitive C reactance. Inductive reactance

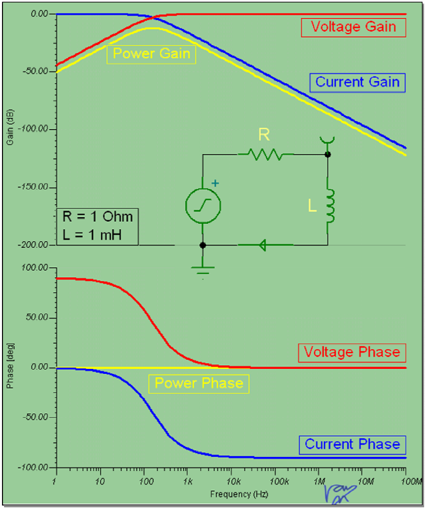

Circuit 4: One milliHenry Inductor Frequency Response

Next we want to know what happens if the inductance changes, say, to a thousandth of its value. That would be 1 microHenry (uH). Again, TINA computes the AC transfer characteristic, sweeping the frequency from 1 to 100 MegaHertz. This feels more like radio! Out pops our next graph. Decreasing the inductance has shifted all our gain curves to higher frequencies.

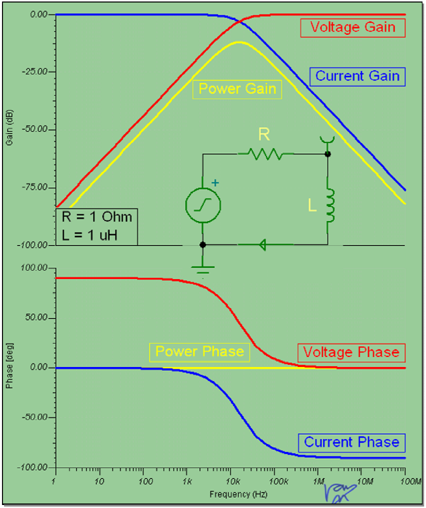

Circuit 5: One microHenry Inductor Frequency Response

Maximum possible secondary current and maximum possible voltage across C_S, when:

$$\mathbf{E_C / E_P = (1/ 2k_c) * \sqrt{(L_S L_P)} = [\sqrt{(Q_P Q_S)} / 2] * \sqrt{(L_S L_P)}}$$

- goes to zero at unity coupling
- equals 3.01 dB (50 percent loss) at critical coupling
- becomes infinite at zero coupling.

If primary power held constant by adjusting the primary voltage as coupling is varied, then secondary voltage, current, power and power transfer all increase with increasing coupling. There are no maxima at critical coupling -- just the characteristic 3 dB power transfer loss.

Coupled Resonant Circuits, from Terman, p. 154, ff.

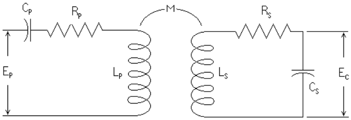

The coupling coefficient is:

$$\mathbf{k = M / \sqrt{(L_P L_S)}}$$

Voltage ratio:

$$\mathbf{E_C / E_P = -\sqrt{(L_S L_P)} * k / [k^2 + 1/(Q_P Q_S)]}$$

Voltage ratio is maximum when:

$$\mathbf{k = k_c = 1 / \sqrt{(Q_P Q_S)}}, \quad \text{and}$$

$$\boldsymbol{\omega} \mathbf{M = \sqrt{(R_P R_S)}}$$

Now we can run different cases for hours, and trust me, I have. The trick is to focus on essential relationships. What happens if we change the resistor value but not the inductor?

Let's return the inductor to 1 milliHenry and change the resistance from 1 to 100 Ohms. What happens? How does increased input resistance affect frequency response?

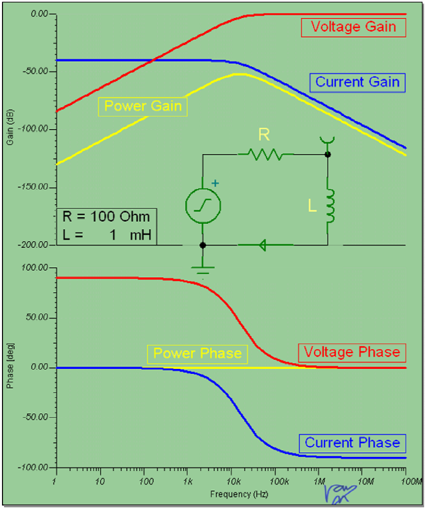

Circuit 6: Inductive Frequency Response – Increased Series Resistance

If we compare Circuit 4 and Circuit 6, there is a loss in current gain as a direct consequence of the resistor. That makes sense. So to first order, we observe that with respect to voltage we have a high pass filter – so named because high frequencies are

---

Resistance and reactance components of the coupled impedance (see 1.) are:

Resistance component: $$\frac{\mathbf{(\omega M)^2}}{\mathbf{R_S^2 + X_S^2}} * \mathbf{R_S}$$

Reactance component: $$\frac{\mathbf{-j\ (\omega M)^2}}{\mathbf{R_S^2 + X_S^2}} * \mathbf{X_S}$$

R_S and X_S are resistance and reactance components of Z_S.

The effect of the secondary on the primary is exactly as if these components had been inserted, in series, into the primary circuit. The energy and volt-amperes consumed by the primary current flowing through this hypothetical resistance and reactance represent the energy and reactive volt-amperes that are transferred to the secondary circuit. (Alternatively, $\mathbf{Z_P + (\omega M)^2 / Z_S}$ can be calculated directly using standard complex algebra.)

When primary and secondary are both resonant at the same frequency, X_S equals zero, the reactance component equals zero, and coupled impedance equals the resistance component, $\mathbf{(\omega M)^2 / R_S}$.

If, further, the coupling is critical, the coupled impedance is simply R_P.

If input voltage is held constant and coupling is varied, then primary and secondary currents and powers and secondary voltage all change with coupling.
Secondary voltage, current and power all reach maximum values when coupling is critical. Power transfer loss (from primary to secondary):

Z_S = series impedance of secondary when considered by itself

2. The voltage induced into the secondary by primary current, I_P, has a magnitude, ω M I_P, and lags the current that produces it by 90 degrees. The equivalent secondary circuit is:

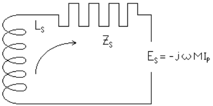

(I_P is the primary current after the coupled impedance from the secondary has been factored in.)

3. The secondary current is exactly the same current that would flow if the induced voltage, E_S, were applied in series with the secondary and with the primary absent.

An alternate form of the equivalent secondary circuit is:

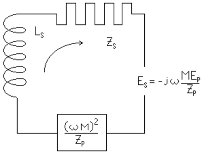

(Z_P is the primary series impedance before the coupled impedance from the secondary is factored in.)

---

passed and low frequencies are blocked. With respect to current, we have a low¬¬-¬¬¬¬pass filter, and with respect to power, we have a band-pass filter. Interesting, no?

Inductor Calculations

Inductors are like resistors when it comes to equivalent circuits.

Adding two inductors in series gives the equivalent inductance of the pair:

$$\boxed{L_{series} = L_1 + L_2}$$

Formula: Series Equivalent Inductance

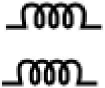

Use the product over sum for inductors in parallel:

$$\boxed{L_{parallel} = \frac{L_1 L_2}{L_1 + L_2}}$$

Formula: Parallel Equivalent Inductance

Transformers are inductors that are magnetically coupled by their proximity to each other. We will discuss them in more detail later.

Inductors have a kind of imaginary AC resistance called inductive reactance that has units of Ohms.

$$\boxed{X_L = 2\pi f L}$$

Formula: Inductive Reactance

Inductive circuits have a time constant that we alluded to above. This is the time it takes for the current to build up to 63.2% of its steady state value. The units are seconds.

$$\boxed{t_L = \frac{L}{R}}$$

Formula: Inductive Time Constant

By choosing the right value of inductors, we can tailor the frequencies we block or pass using "analog filtering". More on that and its upscale digital cousin in a moment. Take a break. Don't become incapacitated!

Capacitors (Farads)

Chapter 2-3 Coupled Circuits

Inductively Coupled Circuits -- from Terman, p. 149, ff. See also Landee, Davis and Albrecht, Section 13.

Terman Model:

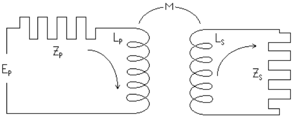

Pertinent mesh equations:

$$\mathbf{E_P = I_P Z_P + j\omega M I_S}$$
$$\mathbf{E_S = -j\omega M I_P = I_S Z_S}$$ induced voltage in secondary

1. The effect of the presence of the secondary, coupled to the primary, is to add a series impedance, $\mathbf{(\omega M)^2 / Z_S}$, to the primary.

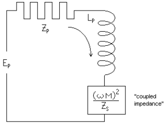

M = mutual inductance
$\omega = 2\pi * \text{frequency}$

- the coupled impedance components are added to the primary circuit impedance components to give the complex impedance of the primary.

- magnitude of the total primary current

- the current flow divides going through "Lp", thus the actual current through the L-R branch (through L) is proportional to the ratio of the L-R branch admittance magnitude to the overall "Lp" admittance magnitude:

L-R branch current = total primary current * | Y (L-R branch) | / | Y (Co branch) + Y (L-R branch) | (Note, the L-R branch current can actually exceed the total primary current due to the "tank" effect of the Co || (L-R) network.)

- the voltage induced into the secondary Es is calculated (step 2, above)

The voltage |V2| across load resistor Rload is Es times the ratio Rload / |total sec. impedance|.

---

store energy as an electric field. They consist of plates of foil separated by an insulating or dielectric material. Like their inductive counterparts, capacitors have a direct current (DC) response and an alternating current (AC) response.

When the switch is closed, there is a surge of current until charge accumulates on the plates of the capacitor. After several time constants, the circuit reaches “steady state”. The electric field is established and the current surge disappears. If the switch is opened, nothing happens but the capacitor remains fully charged! A large capacitor can shock you!

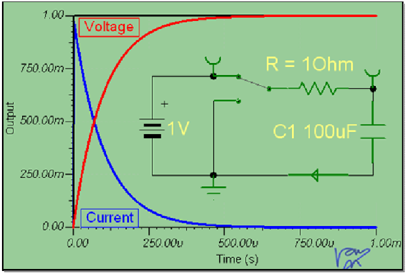

Circuit 7: Capacitor Voltage and Current Vs. Time, Transient DC Response

This simulation uses a 1-Farad capacitor, which is physically large, about the size of a large soup can. In radio, we typically work with much smaller values as we shall soon see. The principles and response curves are similar; the time constants are much shorter. Remember these units and abbreviations; you will use them often, especially nano and pico.

Unit	Abbrev.	"Of a Farad"	Multiplier	Comment
Farad	F	1	1	Huge!
milliFarad	mF	1 thousandth	10-3	Big!
microFarad	µF	1 millionth	10-6	Pwr. Sup.
nanoFarad	nF	1 billionth	10-9	Various
picoFarad	pF	1 trillionth	10-12	RF freq.

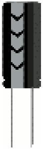

Table 1: Unit Prefixes, Abbreviations and Multipliers

Consider the same circuit as above, but we replace the DC battery with an AC signal generator like so.

The complex impedance and admittance of primary and secondary coils "Lp" and "Ls" are:

$$\mathbf{Z = R / [(1 - \omega^2 LCo)^2 + \omega^2 Co^2R^2]}$$

$$\mathbf{+\ j * \omega [L(1 - \omega^2 LCo) - CoR^2] / [(1 - \omega^2 LCo)^2 + \omega^2 Co^2 R^2]}$$

$$\mathbf{Y = R / [R^2 + \omega^2 L^2]\ -\ j * \omega[L(1 - \omega^2 LCo) - CoR^2] / [R^2 + \omega^2 L^2]}$$

The **spreadsheet** requires the following **input data**:

L, **Co** and, at each frequency, **R** for coils "Lp" and "Ls." (L's are true inductances.)

Coupling coefficient **k** or mutual inductance M.

Residual circuit resistances (if any) **Rpri** and **Rsec**.

The spreadsheet proceeds by calculating

- the complex impedance and admittance of the primary coil "Lp"

- the admittance calculation is checked by calculating the admittances of the parallel legs in "Lp": the Co leg and the series L-R leg.

- the complex impedance of secondary coil "Ls"

Resistance and reactance components of the **coupled impedance** are:

Resistance component: $$\frac{(\omega M)^2}{R_S^2 + X_S^2} * R_S$$

Reactance component: $$\frac{-j\ (\omega M)^2}{R_S^2 + X_S^2} * X_S$$

$\mathbf{R_S}$ and $\mathbf{X_S}$ are the resistance and reactance components of $\mathbf{Z_S}$.

The effect of the secondary on the primary is exactly as if these components had been inserted, in series, into the primary circuit. The energy and volt-amperes consumed by the primary current flowing through this hypothetical resistance and reactance represent the energy and reactive volt-amperes that are transferred to the secondary circuit. (Alternatively, $Z_P + (\omega M)^2 / Z_S$ can be calculated directly using standard complex algebra.)

2. The voltage **Es** induced into the secondary by primary current, $\mathbf{I_P}$, has a magnitude, $\boldsymbol{\omega}$ **M** $\mathbf{I_P}$, and lags the current that produces it by 90 degrees. $\mathbf{I_P}$ is the primary current after the coupled impedance from the secondary has been factored in.

3. The secondary current is exactly the same current that would flow if the induced voltage were applied in series with the secondary and with the primary absent.

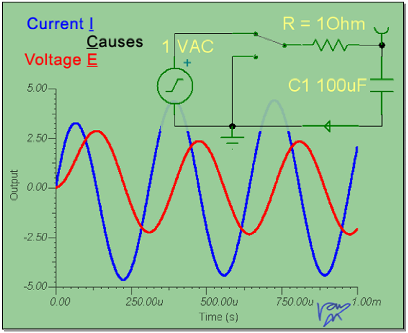

Circuit 8: Capacitive AC Response

The AC signal causes the capacitor’s electric field to repeatedly collapse and expand in alternating directions. Ohm’s law is constantly running, and again there is a phase delay. Current Causes Voltage in a capacitor. The word “Causes” is just a hack so that we remember the C for Capacitor in the famous ELI-the-ICE-man phrase that reminds us that voltage leads current in inductors and current leads voltage in capacitors.

Frequency Response: What Happens When Capacitance and Resistance Change?

Just like before we play with capacitance and resistor values to see what happens in our circuit.

Again, we will measure this by comparing the gain of various configurations. First, we fix the resistor at 1 Ohms and set the capacitor to 100 uF. Then we ask TINA to compute the AC Transfer Characteristic. A graph again provides insight:

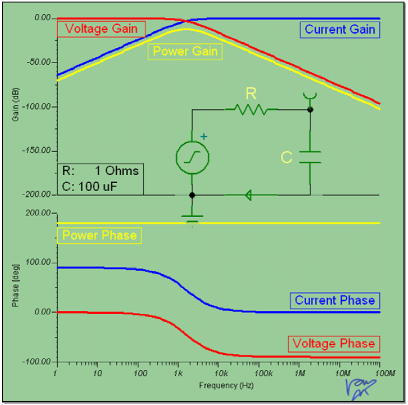

Circuit 9: 100 uF Capacitor Frequency Response

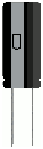

As before we want to know what happens if the capacitance changes, this time to a hundredth of its value. We set the capacitor to 1 uF (1 microFarad). Again, TINA computes the AC transfer

Terman Inductive Coupling Model

Mike Tuggle
(Personal Comm)

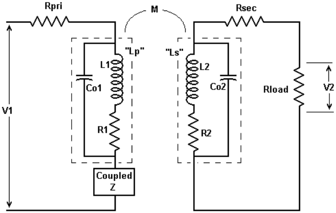

While the primary and secondary circuit impedances involve the Co || (L-R) networks "Lp" and "Ls", we assume that the coupling between circuits is a function solely of true inductances L1 and L2 and their mutual inductance M.

The **basis of the calculation** is drawn from F.E.Terman, Radio Engineers' Handbook, 1st ed., 1943, p. 148 ff:

1. The effect of the presence of the secondary, coupled to the primary, is to add a series impedance, **coupled Z =** $(\omega M)^2 / Z_S$, to the primary, where:

M = mutual inductance
$\omega = 2\pi * \text{frequency}$
Z_S = series impedance of secondary when considered by itself

v3 = jω M23 i2

The voltage v1 induced in L1 will be:

v1 = jω M12 i2

Since voltages in series must add up, the voltage across the entire coil v1 will be:

v1 = v2 + v3

Putting these voltages in terms of i2 from the previous relationships, we get

jωM12 i2 = jωL2 i2 + jωM23 i2

Taking out the common factor jωi2, this simplifies to

M12 = L2 + M23

This page last updated: January 3, 2016
Copyright 2009, 2016, Robert Weaver

characteristic, sweeping the frequency from 1 to 100 MegaHertz. That radio feeling is coming on strong.

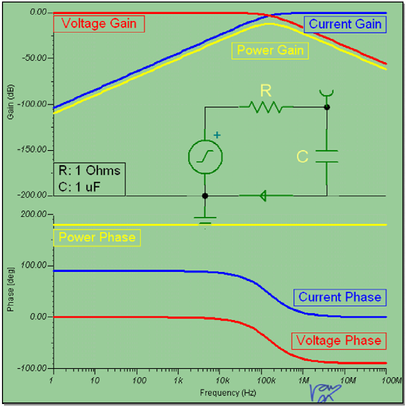

Circuit 10: 1 uF Capacitor Frequency Response

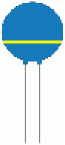

Are you starting to see a pattern? What happens if we change the value of the resistor but not the capacitor? Let’s return the capacitor to 100 uF and change the resistance from 1 to 100 Ohms.

What happens? How does increased series resistance affect circuit frequency response?

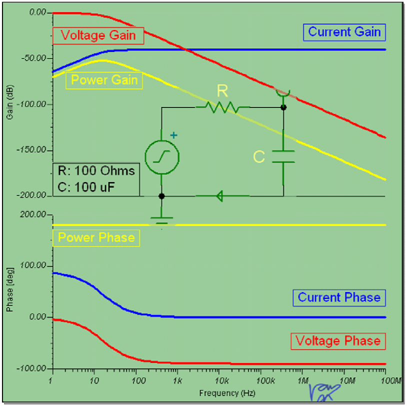

Circuit 11: Capacitive Frequency Response – Increased Series Resistance

If we compare Circuit 9 and Circuit 11, there are two effects of keeping the same capacitor and increasing the resistance. Current gain decreases. That makes sense. The second effect is to shift the curves to the left. It looks like we increased the capacitance, but we didn't.

---

components. To develop formulae to find the required component values is beyond the scope of this discussion. In this case we are better off availing ourselves of a spreadsheet using these formulae, and then plugging in various values until we have the desired frequency range. If it turns out that the practical values of inductor and bandspread capacitor we have to work with don't give a narrow enough frequency range, we can of course, reduce the range of the bandspread capacitor as we did in Parts 1 and 2 by adding a padder and/or trimmer.

Acknowledgement

I would like to express my thanks to Mike Tuggle who provided valuable feedback and also performed SPICE simulations verifying the formulae.

Appendix

Derivation of Mutual Inductance Between a Coil and a Section of the Same Coil

Using the same notation as used previously, L1 is the inductance of the entire coil, L2 is the inductance of the bottom section, and L3 is the inductance of the top section.

If an AC voltage source v2 is impressed across L2, the current i2 will be:

$$\mathbf{i2 = v2/(j\omega L2)}$$

and hence

$$\mathbf{v2 = j\omega L2\ i2}$$

The voltage v3 induced in L3 will be:

$$(1-k^2)L_1L_2C_AC_B\omega^4 - L_1C_A\omega^2 - L_2C_B\omega^2 + 1 = 0 \tag{31}$$

where the coupling coefficient k is given by:

$$k = \frac{L_1 + L_2 - L_3}{2\sqrt{L_1L_2}}$$

and then (31) can be solved for frequency using the usual quadratic formula:

$$x = \frac{-b - \sqrt{b^2 - 4ac}}{2a}$$

where:

$x = \omega^2$

$a = (1-k^2)L_1L_2C_AC_B$

$b = -(L_1C_A + L_2C_B)$

$c = 1$

Having calculated x from the above, the resonant frequency F will be

$$F = \frac{\sqrt{x}}{2\pi}$$

where F is in Hz, and the other component values are given in Farads and Henries.

For convenience, the calculation may be done using picofarads and microhenries, but the resulting frequency will then be in Gigahertz, and therefore will need to be multiplied by 1,000,000 to get kHz. Hence:

$$F = \frac{10^6\sqrt{x}}{2\pi}$$

In contrast to Parts 1 and 2, where we selected an arbitrary frequency range and then developed formulae to find the necessary components values, here we have developed a formula to find the resonant frequency from a given set of

---

We observe that our capacitor drains more slowly when the resistance is higher. In an opposite sense to inductors, capacitors are a low-pass filter with respect to voltage and a high-¬¬¬¬pass filter with respect to current. With respect to power, we have a band-pass filter as before.

Capacitor Calculations

Capacitors are the opposite of inductors and resistors when it comes to equivalent circuits.

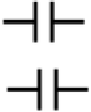

Because it looks like increasing plate area, adding two capacitor values gives the PARALLEL equivalent capacitance:

$$\boxed{C_{parallel} = C_1 + C_2}$$

Formula: Parallel Equivalent Capacitance

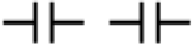

Use the product over sum for capacitors in SERIES:

$$\boxed{C_{series} = \frac{C_1 C_2}{C_1 + C_2}}$$

Formula: Parallel Equivalent Inductance

There isn’t the capacitive equivalent of a transformer.

Capacitive circuits have a time constant. This is the time it takes for the voltage to build up to 63.2% of its steady state value. The units are seconds.

$$\boxed{t_C = RC}$$

Formula: Capacitive Time Constant

Capacitors also have a kind of imaginary AC resistance called capacitive reactance that has units of Ohms.

$$\boxed{X_L = \frac{1}{2\pi f C}}$$

Formula: Capacitive Reactance

That’s it for capacitance right now. Consult the references in the last chapter if you want to delve in deeper than time allows here.

Summary – Capacitance and Inductance:

The figure below summarizes what we have just discovered by direct simulation. Inductors and Capacitors are the inverses of each other. This is an idea as deep as the electron itself.The left

aiding or opposing, and the sign and value of M12 will be self-evident.

If we are dealing with the single tapped winding of the circuit in figure 1, then we can directly measure the values of L1, L2 and L3.

Then the mutual inductance between L2 and L3, is

$$M_{23} = \frac{L_1 - L_2 - L_3}{2}$$

And it can be shown (see Appendix below) that the mutual inductance between L1 and L2 is

$$M_{12} = M_{23} + L_2$$

Hence:

$$M_{12} = \frac{L_1 + L_2 - L_3}{2}$$

Now that we have determined M12 by one of the above methods, the coupling coefficient is then given by:

$$k_{12} = \frac{M_{12}}{\sqrt{L_1 L_2}}$$

or combining these two formula, we get:

$$k_{12} = \frac{L_1 + L_2 - L_3}{2\sqrt{L_1 L_2}}$$

This now gives us all the information required to calculate resonant frequency of the bandspread tank circuit.

Summary

The relationship between component values, and resonant frequency of a tapped coil bandspread circuit has been given by formula (31):

$$x = \omega^2$$
$$a = (1 - k^2) L_1 L_2 C_A C_B$$
$$b = -(L_1 C_A + L_2 C_B)$$
$$c = 1$$

Note, that a quadratic equation will generally have two solutions because the square root portion can be either positive or negative. For normal situations, the form given above will provide the correct frequency. The other result is equally valid mathematically, and will give a much higher resonant frequency which we may interpret as a parasitic. Therefore, it may be prudent to ensure that related circuit components are selected to prevent operation at undesired frequencies.

Now, we can plug the component values into the above formula and solve for ω2, and consequently solve for ω and frequency.

The only value that will not be readily known is the coupling coefficient. However it can be calculated from the other coil parameters.

If we are dealing with the isolated windings of the circuit in figure 2, then we can measure the inductance of the two windings separately, and then connect them in series, and measure them again. Using the following formula, the mutual inductance can be calculated:

$$L_{SERIES} = L_1 + L_2 \pm 2M_{12}$$

Hence:

$$M_{12} = \pm \frac{L_{SERIES} - L_1 - L_2}{2}$$

The contribution due to M12 will be positive or negative depending on whether the polarity of the series connections are

---

column shows a capacitive circuit and the right column shows its inductive counterpart. The component values are summarized in the lower left corner of each diagram. As before voltage gain is red, current gain is blue, and power gain is yellow.

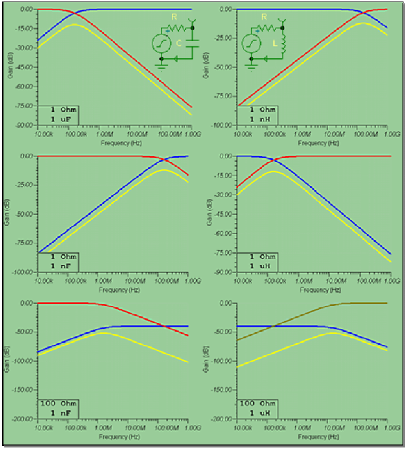

Figure 8: Side-By-Side Comparison - Capacitance and Inductance

RLC Behavior – Parallel Case

Consider an RLC circuit where the inductor and capacitor are in parallel. The following figures catalog voltage, current and power gain as we vary component values. You can reproduce these results in Tidy TINA and see which values result in which curves, a worthwhile bit of fun. The flat line in each image is the gain for the source, set to 1 milliVolt to simulate a strong radio station.

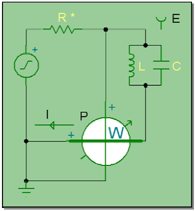

Circuit 12: RLC Circuit – Parallel LC

$$Z_N = Z_{CA} \,||\, \frac{(k^2-1)\omega^2 L_1 L_2 + j\omega L_1 Z_{CB}}{j\omega L_2 + Z_{CB}} \tag{26}$$

Next, we invert the formula so that we can combine the parallel impedances:

$$\frac{1}{Z_N} = \frac{1}{Z_{CA}} + \frac{j\omega L_2 + Z_{CB}}{(k^2-1)\omega^2 L_1 L_2 + j\omega L_1 Z_{CB}} \tag{27}$$

Also note that the reciprocal of the impedance is the admittance, and at parallel resonance the admittance 1/ZN (neglecting any resistive components) will be zero. Hence:

$$\frac{1}{Z_{CA}} + \frac{j\omega L_2 + Z_{CB}}{(k^2-1)\omega^2 L_1 L_2 + j\omega L_1 Z_{CB}} = 0 \tag{28}$$

Capacitive reactance is given by:

ZC = – j/(ωC) = 1/(jωC)

Substituting this for the ZCx terms in the formula:

$$j\omega C_A + \frac{j\omega L_2 + \frac{1}{j\omega C_B}}{(k^2-1)\omega^2 L_1 L_2 + \frac{L_1}{C_B}} = 0 \tag{29}$$

Multiplying everything by the denominator of the fraction:

$$j\omega C_A((k^2-1)\omega^2 L_1 L_2 + \frac{L_1}{C_B}) + j\omega L_2 + \frac{1}{j\omega C_B} = 0 \tag{30}$$

and then multiplying by jωCB we get a relatively simple formula:

$$(1-k^2)L_1 L_2 C_A C_B \omega^4 - L_1 C_A \omega^2 - L_2 C_B \omega^2 + 1 = 0 \tag{31}$$

Since there are no ω or ω3 terms we can solve this as yet another quadratic equation:

$$x = \frac{-b - \sqrt{b^2 - 4ac}}{2a}$$

where

Note that when we convert from impedance to capacitance, 1/n2 becomes n2. Also, since n=N2/N1, and we assume that because N2 represents the tapped part of the coil in the original circuit, it is smaller than N1. Hence, n will be less than 1. Therefore, when CB is transferred to the primary side, its effective value becomes smaller.

We have demonstrated that if the coupling coefficient is 1, then the equivalent circuit is reduced to the secondary impedance multiplied by the turns ratio and transferred to the primary in parallel with the primary inductance and capacitance, which agrees with traditional transformer theory. The capacitor CB on the secondary side can be scaled by multiplying by n2 and placed in parallel with CA on the primary side. However, it's worth noting one important fact. In traditional transformer theory, the self inductance of the primary is generally considered much higher than the transferred impedance from the secondary side, and is often neglected. However, we can't neglect it in our case, since it forms the inductive part of the resonant circuit.

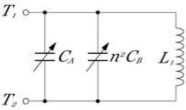

Figure 5

Since we have carried the effect of the coupling coefficient k through to formula (21), we can go back a few steps and see the effect of a coupling coefficient less than 1.

Starting with (21) again:

$$Z_N = Z_{CA} \,||\, \frac{-\omega^2 L_1 L_2 + j\omega L_1 Z_{CB} + \omega^2 k^2 L_1 L_2}{j\omega L_2 + Z_{CB}} \qquad (21)$$

First we combine the common ω2L1L2 terms

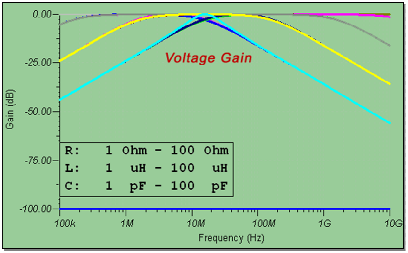

Figure 9: RLC Circuit – Parallel LC – Voltage Gain

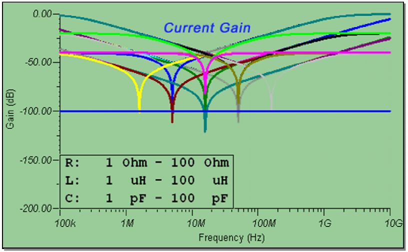

Figure 10: RLC Circuit – Parallel LC – Current Gain

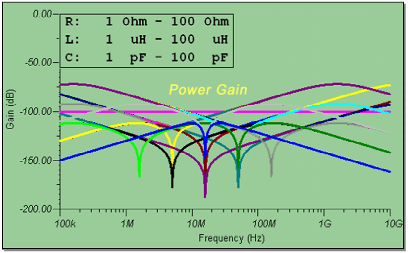

Figure 11: RLC Circuit – Parallel LC – Power Gain

RLC Behavior – Series Case

Consider an RLC circuit where the inductor and capacitor are in series. Note the difference in the response curves in series versus parallel components. You can right click a specific curve in Tidy TINA to discover the RLC values that gave rise to it. One thing you will notice is that while the parallel case is a band-stop filter for RF power, the series case is a band-pass in most cases. The peakiness of the filter is the Q or Quality Factor of the resonant circuit. More on that later. Notice that series resistance hurts performance of the band-pass filter, turning it into a band-stop filter! Not good for tuning in your favorite crystal radio station. Again, we set the source to 1 mV to simulate a strong station. We will make those conditions more severe later.

Taking everything on the right to a common denominator:

$$Z_N = Z_{CA} \,||\, \left(\frac{j\omega L_1(j\omega L_2+Z_{CB})}{j\omega L_2+Z_{CB}} + \frac{\omega^2 k^2 L_1 L_2}{j\omega L_2+Z_{CB}}\right)$$

$$Z_N = Z_{CA} \,||\, \left(\frac{j\omega L_1(j\omega L_2+Z_{CB})+\omega^2 k^2 L_1 L_2}{j\omega L_2+Z_{CB}}\right) \quad (20)$$

Remembering that **j2=-1**:

$$Z_N = Z_{CA} \,||\, \frac{-\omega^2 L_1 L_2 + j\omega L_1 Z_{CB} + \omega^2 k^2 L_1 L_2}{j\omega L_2+Z_{CB}} \quad (21)$$

Notice that in the numerator, the first and last terms will cancel each other out when k=1. Let's assume for the moment that this is the case. Then we are left with:

$$Z_N = Z_{CA} \,||\, \frac{j\omega L_1 Z_{CB}}{j\omega L_2+Z_{CB}} \quad (22)$$

Also when k=1, the n2 relationship exists between inductance values as previously mentioned:

L2/L1 =(N2/N1)2=n2

where we define n as the turns ratio. Substituting L2/n2 for L1

$$Z_N = Z_{CA} \,||\, \frac{1}{n^2}\frac{j\omega L_2 Z_{CB}}{j\omega L_2+Z_{CB}} \quad (23)$$

Notice that we now have a fraction in the form of AB/(A+B) which is the formula used when combining parallel impedances. Therefore, this is equivalent to having A and B in parallel. Hence:

$$Z_N = Z_{CA} \,||\, \frac{1}{n^2} Z_{CB} \,||\, \frac{1}{n^2} j\omega L_2 \quad (24)$$

or

$$Z_N = Z_{CA} \,||\, \frac{1}{n^2} Z_{CB} \,||\, j\omega L_1 \quad (25)$$

So, for the case of k=1 (which would generally apply in the situation of a coil on a ferromagnetic core), we can redraw our circuit as shown in figure 5.

From this we can define Z', the coupled impedance as:

$$Z' = \frac{(\omega M)^2}{Z_2+Z_{CB}} \qquad (16)$$

Using equation (16) we can simplify the circuit of figure 3 into a single loop as shown in figure 4.

In summary then, the effect that the current flowing in the secondary circuit has on the primary circuit can be treated as another voltage drop VZ' due to an impedance Z' applied to the primary circuit as shown in figure 4. Here, Z' is placed in series with the other primary circuit components, and completely accounts for the effect due to the secondary circuit. We have now reduced the original two loop circuit to a single loop circuit.

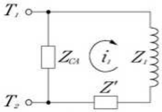

Figure 4

The net impedance ZN of the primary circuit is:

$$Z_N = Z_{CA} \mid\mid (Z_1 + \frac{(\omega M)^2}{Z_2+Z_{CB}}) \qquad (17)$$

where || indicates components combined in parallel.

Substituting (11) into (17) we put M in terms of L1 L2 and k.

$$Z_N = Z_{CA} \mid\mid (Z_1 + \frac{\omega^2 k^2 L_1 L_2}{Z_2+Z_{CB}}) \qquad (18)$$

If we neglect any resistance in the circuit, then the impedances are purely reactive. The impedance of an inductor is given as:

$$Z=j\omega L$$

Hence:

$$Z_N = Z_{CA} \mid\mid (j\omega L_1 + \frac{\omega^2 k^2 L_1 L_2}{j\omega L_2+Z_{CB}}) \qquad (19)$$

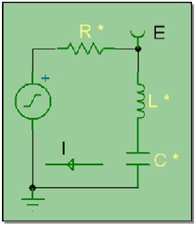

Circuit 13: RLC Circuit – Series LC

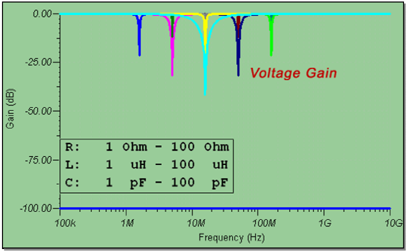

Figure 12: RLC Circuit – Series LC – Voltage Gain

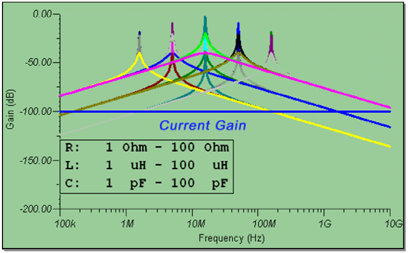

Figure 13: RLC Circuit – Series LC – Current Gain

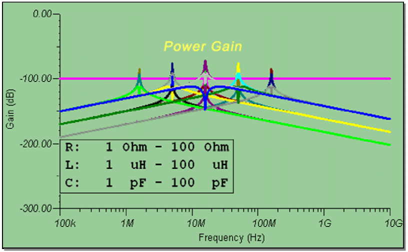

Figure 14: RLC Circuit – Series LC – Power Gain

Now, let us assume that a sinusoidal alternating voltage VT is applied across terminals T1 and T2. This causes an alternating current i1 to flow in the primary circuit. This is illustrated in figure 3, which shows the circuit components as impedances.

From the previous definition of mutual inductance, then this alternating current induces a voltage VZ2 in the secondary circuit across Z2, which in turn causes a current i2 to flow in the secondary circuit.

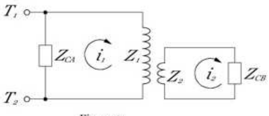

Figure 3

Now, current i2 in the secondary also induces a voltage VZ1 back in the primary across Z1.

The voltage and current relationships are given by the following formulae:

$$V_T = i_1 Z_1 + j\omega M i_2 \tag{12}$$

$$i_2 = \frac{V_{Z2}}{Z_2+Z_{CB}} \tag{13}$$

$$V_{L2} = -j\omega M i_1 \tag{14}$$

The second term in equation (12) shows that the effect of the secondary circuit corresponds to a voltage drop. We can substitute in equations (13) and (14) to get this voltage drop in terms of i1.

$$V_T = i_1 Z_1 + \frac{(j\omega M)(-j\omega M) i_1}{Z_2+Z_{CB}}$$

or

$$V_T = i_1 Z_1 + \frac{(\omega M)^2}{Z_2+Z_{CB}} i_1 \tag{15}$$

L2/L1 =(N2/(N2 +N3))2=(N2/N1)2

etc.

However, except as specifically noted below we will not make any assumptions about the value of the coupling coefficient.

Note that L1 is not equal to L2+L3 (except when k=0). The inductance of the entire coil is, in fact, given by:

L1=L2+L3+2M23

We will leave the circuit of Figure 1 for the time being, and deal with an equivalent coupled circuit consisting of completely separate windings as shown in Figure 2.

In this circuit, we will still assume that even though the coils are separate, they are still inductively coupled. Notice also, we keep the same terminology between figures 1 and 2 such that CA is still connected across L1 and CB is still connected across L2. We will also assume that L1 has the same number of turns in both figures and L2 has the same number of turns in both figures. Consequently, for the purpose of this analysis, both circuits will behave essentially the same. The reason for separating the bandspread section of the circuit into an electrically isolated winding is to make the circuit analysis a bit simpler to follow.

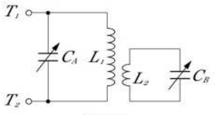

Figure 2

Diodes
do not store energy like capacitors and inductors. They are one-way valves for the flow of current. They are arguably the most important single component in radio because of the multiple purposes they serve. Diodes are semiconductors consisting of a P-N junction doped to attain desireable characteristics. Like their siblings, diodes have a direct current (DC) response and an alternating current (AC) response. These responses can be steady state or transient. Let’s throw the switch!

Notice that the circuit below uses two single pole double throw switches connected so that we can switch the polarity on the diode. We retain the series resistor as a current-limiting resistor, although in a real circuit, say with a light-emitting diode, the value would be considerably higher, between 500 and 2000 Ohms to prevent the diode from burning out.

In this simulation we assume the diode can take whatever the flow of current is and we observe the transient DC response for two cases. The first case when both switches are down, corresponds to the normal polarity of DC voltage seen in previous examples. The diode is positioned so that this is a forward voltage corresponding to the direction in which the diode allows current to flow. The current curve is blue and the forward voltage curve is red. Notice that the diode takes only a few picoseconds for the diode to switch on. The time it takes a diode to turn on is an imporant parameter of the diode, especially for radio work.

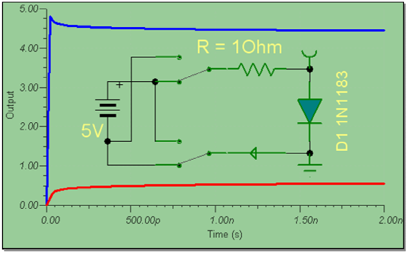

Circuit 14: Diode Transient DC Response – Forward Current

Now let's reverse the position of both switches to simulate flipping a double pole double throw (DPDT) switch. This reverses the polarity of the battery. What do think the curves will look like?

$$V_2 = \omega M i_1 \qquad (10)$$

and conversely

$$V_1 = \omega M i_2$$

where the subscripts 1 and 2 designate the parameters of the first and second coil respectively. M is the mutual inductance between the two coils, and the voltages and currents are sinusoidal at an angular frequency ω=2πf. (ω is the lowercase Greek letter omega.) If the two coils are perfectly coupled, then

$$M = \sqrt{L_1 L_2}$$

If they are not perfectly coupled then we introduce a coupling coefficient into the relationship thus:

$$M = k\sqrt{L_1 L_2} \qquad (11)$$

where k ranges between zero (no coupling at all) and one (perfect coupling).

From Figure 1:

L3 is the inductance of the upper section of the coil, and has N3 turns

L2 is the inductance of the bottom section of the coil, and has N2 turns

L1 is the total inductance of the whole coil and has N1=N2+N3 turns

If the coupling coefficient is one, then there is a direct relationship between number of turns and inductance, and the following relationships apply:

L3/L2=(N3/N2)2

linear relationship between frequency and tuning capacitance, as the following graph illustrates:

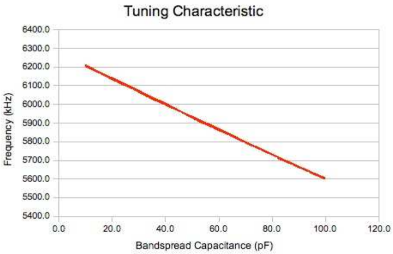

In this analysis we will treat the coil as an autotransformer keeping in mind that traditional transformer theory diverges somewhat from loosely coupled coil theory. With transformer theory, the coupling coefficient between windings is nearly unity, which allows for a few simplifications. Therefore, in this analysis, we will need to adjust the model to account for the fact that the sections of the coil are loosely coupled. (This work is based largely on the information provided in the classic texts: 'Radio Engineering' and 'Radio Engineer's Handbook,' both by Frederick E. Terman.)

When two coils are inductively coupled, then a current in one coil will induce a magnetic flux in the other coil and vice versa. When the current in one coil changes, the flux changes, and the changing flux will induce a voltage in the other coil according to this relationship:

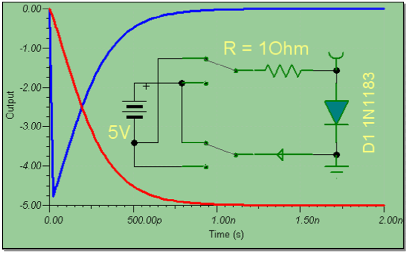

Circuit 14: Diode Transient DC Response – Reverse Current

This case shows the diode in the direction it does not want to conduct. There is a momentary surge of current until the diode turns off. Note that it takes this diode longer to turn off than it does to turn on – in the simulation at least. About a nanosecond. Do these curves remind you of anything familiar?

Now let's replace the DC battery with an AC signal generator and simplify the circuit to. We don't need the DPDT switch, because the AC signal generator is doing that for us. The simplified circuit looks like this:

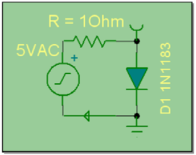

Circuit 15: Diode AC Circuit

Let’s run the signal generator at a low frequency, say 60 Hz. This is the frequently encountered in power supplies running from wall current in the US after a step-down transformer. We obtain the expected and classic waveform for half-wave rectification of an AC signal.

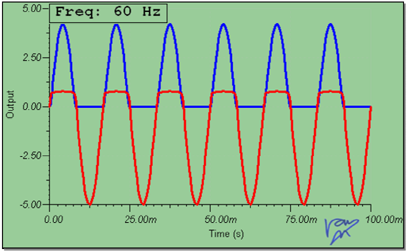

Figure 15: Diode 60 Hz Frequency Response

Now let’s run the frequency up to the high end of the audio sampling spectrum, say 44 kHz. Notice that we start

---

Using a Bandspread Capacitor

Robert Weaver

http://electronbunker.ca/eb/Bandspreading_3.html

Connected to a Tap on the Tank Coil

Here in Part 3, we take off in another direction. We are going to analyze a method of bandspreading where the bandspread capacitor is connected to a tap on the tank inductor, and a bandset capacitor is connected across the entire coil. This configuration is shown in Figure 1.

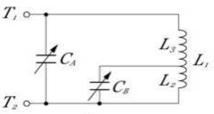

Figure 1

This is a system that has been used on both commercial and homebrew shortwave receivers dating back to the 1930’s. The circuit is as shown in figure 1. In this circuit, the main “Bandset” variable capacitor is CA, and may be either fixed or variable, while the “Bandspread” capacitor CB is variable.

In essence, the goal is to find a relationship between circuit impedance across terminals T1 & T2 and frequency for given values of capacitance and inductance.

Before we go any further, we may ask ourselves why we want to do something that looks quite a bit more complicated than the bandspread arrangements we have already looked at. The answer, as we will discover, is that this gives a remarkably is no longer maximum at resonance. As we have already noted, at a frequency slightly below resonance the coupled impedance is *inductive*, whereas the impedance of the series tuned primary is *capacitive*. Therefore there are two frequencies, one slightly above the resonant frequency and one slightly below where the coupled reactance tunes the primary reactance to give minimum total primary impedance and therefore maximum primary current. This double hump in the primary current tends to flatten the peak of the secondary current response curve since the emf induced into the secondary at the resonant frequency is not quite as great as at the two hump frequencies either side of resonance.

If the tuned circuits are overcoupled, the increase in coupled impedance moves the primary current humps further apart. This in turn causes such a decrease in primary current at resonance that the secondary current also starts to show a double humped curve as shown in Fig. 19.15. In practice a coefficient of coupling of 1.5 times the critical coefficient of coupling produces such a slight dip in the secondary current at resonance that the resulting secondary current resonance curve has a very desirable flat top with steep skirts.

---

encountering some switching noise as we approach the switching speed of the diode.

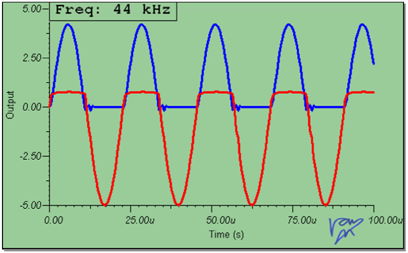

Figure 16: Diode 44 kHz Frequency Response

Finally let’s run the diode at a frequency we might encounter in our crystal radio, say the middle of the AM band:

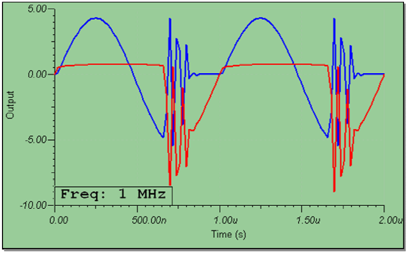

Figure 17: 1N1183 Diode 1 MHz Frequency Response

Now our signal is dominated by switching noise. This particular diode, can't switch fast enough to rectify the signal. Notice the ringing. The only way to see it is to integrate using the Gear method with a 6th order integration. Changing the diode to a faster 1N4150 largely eliminates the ringing.

---

a **critical coupling** for a given pair of tuned circuits at which maximum energy transfer from primary to secondary takes place.

At resonance, $Z_p = R_p$

and the coupled $Z = \dfrac{(\omega M)^2}{R_s}$

Therefore for critical coupling,

$$M^2 = R_pR_s/\omega^2$$

But $M^2 = K^2L_pL_s$ (19.32)

Therefore $K^2 = \dfrac{R_p}{\omega L_p} \times \dfrac{R_s}{\omega L_s}$

from which $K_c = \dfrac{1}{\sqrt{Q_pQ_s}}$ (19.40)

With critical coupling, the secondary current attains its greatest value. But with the coupled impedance rising at resonance, the primary current

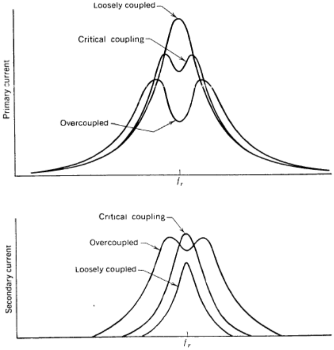

Fig. 19.15. Effect of coefficient of coupling on the resonance curves of a tuned transformer.

Considering now the effect of tuning the secondary circuit, the total secondary circuit impedance [$\mathbf{Z}_s + \mathbf{Z}_L$ in Equation (19.39)] will be minimum and equal to R_s at resonance. Therefore the secondary circuit behaves like a series resonant circuit. Since $(\omega M)^2$ is essentially constant over the small range of frequencies near resonance, when $\mathbf{Z}_s + \mathbf{Z}_L$ becomes a *minimum*, the coupled impedance becomes a *maximum* as shown in Fig. 19.14. And when the source frequency is slightly lower than the resonant frequency, X_C, is greater than X_L, and the secondary impedance becomes capacitive. But dividing a *capacitive* impedance into $(\omega M)^2$ with its 0° angle results in an *inductive* coupled impedance. Therefore, as far as the signal source is concerned, the secondary behaves as if it were a *parallel* resonant circuit in *series* with the primary winding.

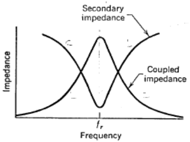

Fig. 19.14. Effect of a tuned secondary circuit.

The primary impedance by itself in the circuit of Fig. 19.13 is attempting to become a minimum at resonance in order to allow maximum primary current to flow. But the coupled impedance is attempting to raise the primary impedance at resonance, thus limiting the maximum primary current. The extent to which the coupled secondary tuned circuit affects the primary resonance curve depends on the degree of coupling between the coils. When the coupling is very loose, the mutual inductance is very small and even at resonance the coupled impedance is *smaller* than the resistance of the primary circuit. Under these circumstances, the only effect that the coupled secondary has on the shape of the primary current resonance curve is to limit its peak value slightly. The secondary current tends to have the usual series resonance curve. But with loose coupling, since the primary current rises to a maximum at resonance and since the secondary induced emf depends on the primary current, the secondary resonance curve is much sharper than that of a single tuned circuit having the same Q.*

We will recall from earlier studies that maximum transfer of energy occurs when the load resistance is equal to the resistance of the source. Applying this to magnetically coupled tuned circuits, maximum energy transfer will take place when the coupled resistance is equal to the primary circuit resistance. Since the coupled resistance depends on the mutual inductance, which in turn depends on the coefficient of coupling, there is

---

* The bandwidth is approximately $\Delta f \approx Kf_r$.

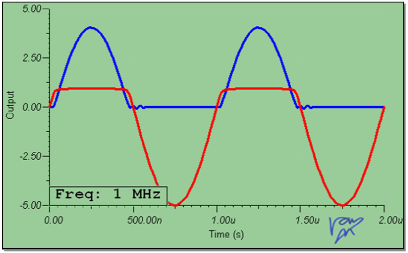

Figure 18: 1N4150 Diode 1 MHz Frequency Response

Diode Calculations

For low frequency applications diodes can be stacked in series.

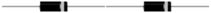

Connecting two diodes in series doubles the peak reverse voltage (PRV) rating:

$$\boxed{PRV_{series} = PRV_1 + PRV_2}$$

Formula: Series Equivalent Inductance

Connecting two diodes in parallel doubles the current rating:

$$\boxed{I_{parallel} = I_1 + I_2}$$

Formula: Series Equivalent Inductance

It the first case it is necessary connect a high value resistor across eacg diode to minimize transients and equalize slight differences in the characteristics of the diode. One rule of thumb is to multiply the PRV of the diode by 400. In the second a low-value resistor, usually less than an ohm, is connected in series with the pair of diodes.

Summary

This concludes chapter one. We have seen a simple radio, a crystal radio, and how each of the parts work. Now we will look in detail at fundamentals of software defined radio, including software and hardware.

---

the vector division, dividing an impedance with a + angle into $(\omega M)^2$ with its 0° angle results in a *capacitive* coupled impedance. Since $\boldsymbol{Z}_p$ is largely *inductive*, the *capacitive* impedance coupled into the primary circuit via the mutual inductance adds the coupled resistance component in series with the primary resistance component but *reduces* the total primary reactance. As a result, the total impedance becomes smaller as the secondary is loaded with a resistance, thus allowing more primary current to flow in order to transfer energy to the secondary circuit. This then checks with our first approach to transformer action in which we discovered that primary current must increase when a transformer is loaded in order to keep the amplitude of the mutual flux sine wave constant.

Example 8: If the resistance of the windings can be neglected, determine the total input impedance to the transformer of Examples 6 and 7, (a) with the secondary open circuit, and (b) with a 50 ohm resistance connected to the secondary.

Solution:

(a) $$\boldsymbol{Z}_{in} = \boldsymbol{Z}_p = \omega L_p = 377 \times 5 = +j1885 \text{ ohms}$$

(b) $$\text{Coupled } \boldsymbol{Z} = \frac{(\omega M)^2}{\boldsymbol{Z}_s + \boldsymbol{Z}_L} = \frac{(377 \times 1.67)^2}{+j(377 \times 0.8) + 50}$$

$$= \frac{396000\angle 0^\circ}{304\angle 80.5^\circ} = 1300\angle -80.5^\circ = 213 - j1284 \text{ ohms}$$

$$\therefore \quad \boldsymbol{Z}_{in} = (+j1885) + (213 - j1284) = 213 + j601$$
$$= \mathbf{636 \text{ ohms}\angle +70.5^\circ}$$

7. Tuned Transformers

The most important application of loose coupling is the use of tuned transformers in radio circuitry. If we consider the source of emf to be a constant voltage source as in Fig. 19.13, the primary winding is designed to form a series resonant circuit with the capacitive reactance of C_p equal to the inductive reactance of the primary winding at the desired resonant frequency. This will result in maximum current in the primary winding at resonance (neglecting for the moment any coupled impedance due to secondary current).*

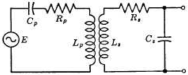

Fig. 19.13. A tuned transformer.

---

* If a vacuum-tube amplifier is used to feed the primary of a tuned transformer, the vacuum tube represents a constant *current* source providing a signal current equal to the mutual conductance of the tube times the signal input voltage to its grid. Hence C_p must then be connected in *parallel* with the primary winding to obtain maximum primary current (resonant rise of current) at resonance.

form of equivalent circuit for this transformer, representing the self-induced emf by an inductance symbol and the mutually induced emf by a generator symbol. To satisfy Lenz's law, the polarity of the generator must be such that it opposes changes in $\mathbf{I}_1$ due to the applied emf $\mathbf{E}_1$. Hence, changing the polarity of the generator symbol in the secondary circuit will *not* change the polarity of the voltage $\mathbf{I}_s(+j\omega M)$ in the primary circuit. Consequently, Kirchhoff's voltage law equation for the primary loop becomes

$$\mathbf{E}_p = \mathbf{I}_p(R_p + j\omega L_p) + \mathbf{I}_s(+j\omega M)$$

or

$$\mathbf{E}_p = \mathbf{I}_p\mathbf{Z}_p + \mathbf{I}_s(+j\omega M) \tag{19.34}$$

where $\mathbf{Z}_p$ is the open-circuit impedance of the primary circuit by itself, and ωM is the mutual reactance of the two windings. Note that Lenz's law gives us the same direction for $\mathbf{I}_2$ that we chose for our generalized coupling network. [Compare Equation (19.34) with Equation (19.5).] Hence, we write the loop equation for the secondary circuit of Fig. 19.12(c) as

$$0 = \mathbf{I}_s(\mathbf{Z}_L + R_s + j\omega L_s) + \mathbf{I}_p(+j\omega M)$$

or

$$0 = \mathbf{I}_s(\mathbf{Z}_L + \mathbf{Z}_s) + \mathbf{I}_p(+j\omega M) \tag{19.35}$$

where $\mathbf{Z}_s$ is the open-circuit impedance of the secondary winding by itself. [Compare Equations (19.35) and (19.14).] From Equation (19.35)

$$\mathbf{I}_s = \frac{-\mathbf{I}_p(+j\omega M)}{\mathbf{Z}_s + \mathbf{Z}_L} \tag{19.36}$$

Substituting in Equation (19.34),

$$\mathbf{E}_p = \mathbf{I}_p\mathbf{Z}_p - \mathbf{I}_p\frac{(+j\omega M)^2}{\mathbf{Z}_s + \mathbf{Z}_L} \tag{19.37}$$

Equation (19.37) is the equivalent of Equation (19.16). In this case we can go one step further by substituting -1 for j^2. Thus

$$\mathbf{E}_p = \mathbf{I}_p\mathbf{Z}_p + \mathbf{I}_p\frac{(\omega M)^2}{\mathbf{Z}_s + \mathbf{Z}_L} \tag{19.38}$$

and dividing through by $\mathbf{I}_p$,

$$\mathbf{Z}_{in} = \mathbf{Z}_p + \frac{(\omega M)^2}{\mathbf{Z}_s + \mathbf{Z}_L} \tag{19.39}$$

where $(\omega M)^2/(\mathbf{Z}_s + \mathbf{Z}_L)$ is the **coupled impedance** *for transformer coupling.* Note that when we substitute -1 for j^2, $(\omega M)^2$ has now become an impedance with a 0° angle.

We can check Equation (19.39) by considering the effect of loading the secondary winding of a pair of magnetically coupled coils. If the secondary is left open-circuit, $\mathbf{Z}_L$ is infinitely large and, therefore, the coupled impedance in Equation (19.39) becomes zero. Therefore, the total impedance is simply the primary impedance alone as we would expect. If we connect a resistance across the secondary winding, $\mathbf{Z}_L$ will have a 0° angle. Since $\mathbf{Z}_s$ is largely the inductive reactance of the secondary winding, the total secondary circuit impedance is inductive. Therefore, when we carry out

---

Resonant Circuits
Mouloud Rahmani
http://mouloudrahmani.com/Electrical/RFMicrowave/ResonantCircuits.html

The resonant circuit is used in electronics systems to selectively pass a certain frequency or group of frequencies from a source to a load while attenuating all other frequencies outside of this passband.

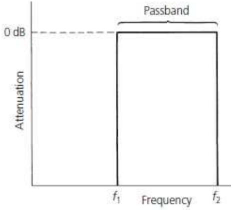

FIG. 2-1. The perfect filter response.

Bandwidth

The bandwidth of any resonant circuit is most commonly defined as being the difference between the upper and lower frequency (f2 −f1) of the circuit at which its amplitude response is 3 dB below the passband response. It is often called the half-power bandwidth.

Q

The ratio of the center frequency of the resonant circuit to its bandwidth is defined as the circuit Q.

Q = fe/(f2 − f1)

Shape Factor

The shape factor of a resonant circuit is typically defined as being the ratio of the 60-dB bandwidth to the 3-dB bandwidth of the resonant circuit. Shape factor is simply a degree of measure of the steepness of the skirts. The smaller the number, the steeper are the response skirts.

Insertion Loss

Whenever a component or group of components is inserted between a generator and its load, some of the signal from the generator is absorbed in those components due to their inherent resistive losses. Thus, not as much of the transmitted signal is transferred to the load as when the load is connected directly to the generator. (No impedance matching function is being performed.) The attenuation that results is called insertion loss and it is usually expressed in decibels (dB).

The voltage division rule (illustrated in Fig. 2-4) states that whenever a shunt element of impedance Zp is placed across the output of a generator with an internal resistance Rs, the maximum output voltage available from this circuit is:

Vout =(Vin) Zp/(Rs + Zp)

of $\boldsymbol{E}_1$ in Fig. 19.12, we can reverse the phase of the output voltage by reversing the direction of one of the windings, or simply by reversing the leads to one of the windings. Where it is necessary to keep track of this phase relationship in a circuit diagram, we mark one end of each winding with a dot as shown in Fig. 19.12(a) and (b). If, at a certain moment, changing mutual flux induces an instantaneous emf into the primary winding with a polarity such that the dotted end of the winding is positive with respect to the undotted end, then the same mutual flux change must induce an instantaneous emf into the secondary winding with a polarity such that the dotted end of the secondary winding is positive with respect to the undotted end.

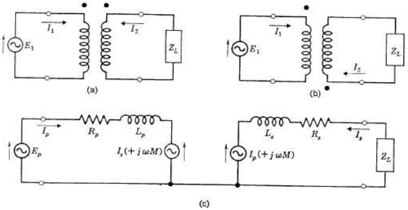

Fig. 19.12. An equivalent circuit of an air-core transformer.

Although reversing the secondary polarity will reverse the current through the external circuit, note that $\boldsymbol{I}_2$ in Fig. 19.12(b) still must flow into the dot end of the secondary winding just as it does in Fig. 19.12(a). Hence, the primary circuit is not affected by any phase reversal we obtain by reversing the secondary leads. In selecting the direction for these current arrows, we must remember that the *same* mutual flux (which is produced by these currents) induces emf's into both primary and secondary windings. So if we draw $\boldsymbol{I}_1$ pointing into the dot end of the primary winding, we must also draw $\boldsymbol{I}_2$ pointing into the dot end of the secondary winding.

When we write the Kirchhoff's voltage law (loop) equation for the primary loop, there are three voltages which add up (vectorially) to equal $\boldsymbol{E}_1$: an IR drop across the resistance of the primary circuit, a *self*-induced emf due to $\boldsymbol{I}_1$ flowing through the primary winding, and a *mutually* induced emf due to $\boldsymbol{I}_2$ flowing in the secondary circuit. Figure 19.12(c) shows one

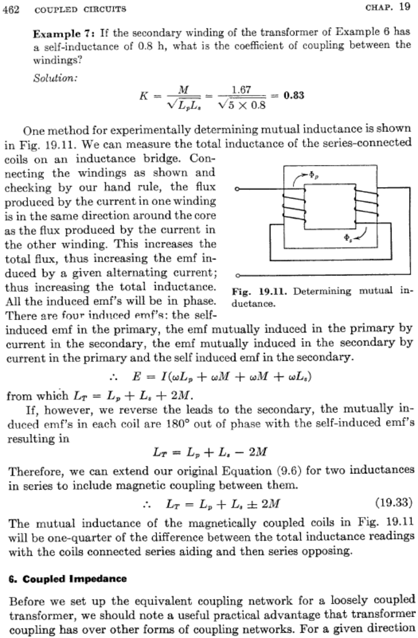

Example 7: If the secondary winding of the transformer of Example 6 has a self-inductance of 0.8 h, what is the coefficient of coupling between the windings?

Solution:

$$K = \frac{M}{\sqrt{L_pL_s}} = \frac{1.67}{\sqrt{5 \times 0.8}} = \mathbf{0.83}$$

One method for experimentally determining mutual inductance is shown in Fig. 19.11. We can measure the total inductance of the series-connected coils on an inductance bridge. Connecting the windings as shown and checking by our hand rule, the flux produced by the current in one winding is in the same direction around the core as the flux produced by the current in the other winding. This increases the total flux, thus increasing the emf induced by a given alternating current; thus increasing the total inductance. All the induced emf's will be in phase. There are four induced emf's: the self-induced emf in the primary, the emf mutually induced in the primary by current in the secondary, the emf mutually induced in the secondary by current in the primary and the self induced emf in the secondary.

Fig. 19.11. Determining mutual inductance.

$$\therefore \quad E = I(\omega L_p + \omega M + \omega M + \omega L_s)$$

from which $L_T = L_p + L_s + 2M$.

If, however, we reverse the leads to the secondary, the mutually induced emf's in each coil are 180° out of phase with the self-induced emf's resulting in

$$L_T = L_p + L_s - 2M$$

Therefore, we can extend our original Equation (9.6) for two inductances in series to include magnetic coupling between them.

$$\therefore \quad L_T = L_p + L_s \pm 2M \tag{19.33}$$

The mutual inductance of the magnetically coupled coils in Fig. 19.11 will be one-quarter of the difference between the total inductance readings with the coils connected series aiding and then series opposing.

6. Coupled Impedance

Before we set up the equivalent coupling network for a loosely coupled transformer, we should note a useful practical advantage that transformer coupling has over other forms of coupling networks. For a given direction

---

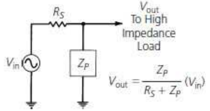

$$V_{out} = \frac{Z_P}{R_S + Z_P}(V_{in})$$

FIG. 2-4. Voltage division rule.

If Zp is a frequencydependent impedance, such as a capacitive or inductive reactance, then Vout will also be frequency dependent and the ratio of Vout to Vin, which is the gain (or, in this case, loss) of the circuit, will also be frequency dependent.

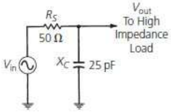

(A) *Simple circuit*

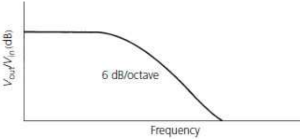

(B) *Response curve*

FIG. 2-5. Frequency response of a simple RC (resistor-capacitor) low-pass filter.

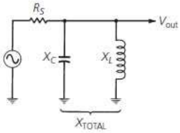

FIG. 2-7. Resonant circuit with two reactive components.

And $$M = L_p \frac{E_s}{E_p} = 5 \times \frac{40}{120} = \mathbf{1.67\ h}$$

Writing Equation (19.28) in terms of the instantaneous emf induced into the secondary,

$$e_s = M \frac{di_p}{dt} \qquad (19.28)$$

And since we are concerned only with sine-wave primary currents,

$$i_p = I_m \sin \omega t \quad \text{and} \quad \frac{di_p}{dt} = \omega I_m \cos \omega t$$

Hence, $$e_s = \omega M I_m \cos \omega t \qquad (19.31)$$

Since the instantaneous emf in Equation (19.31) will be at its peak value when $\cos \omega t = 1$,

$$E_m = 2\pi f M I_m$$

from which $$X_m = \frac{E_{m(\text{sec})}}{I_{m(\text{pri})}} = 2\pi f M \qquad (19.30)$$

Combining Equations (18.1) and (19.28),

$$\frac{N\Phi_m}{t} = \frac{MI_{p_m}}{t}$$

But from Equation (19.27), $\Phi_m = K\Phi_p$

$$\therefore \quad MI_{p_m} = N_s K \Phi_p \qquad (1)$$

If we now reverse the windings and use the original secondary winding as the primary and vice versa, it follows that

$$MI_{s_m} = N_p K \Phi_s \qquad (2)$$

Multiplying Equations (1) and (2),

$$M^2 = K^2 \left(\frac{N_p \Phi_p}{I_{p_m}}\right) \left(\frac{N_s \Phi_s}{I_{s_m}}\right) \qquad (3)$$

But from the definition of self-inductance,

$$E_{av} = \frac{LI_m}{t}$$

and from the definition of the weber,

$$E_{av} = \frac{N\Phi}{t}$$

$$\therefore \quad L = \frac{N\Phi}{I_m}$$

Substituting in Equation (3) gives

$$M = K\sqrt{L_p L_s} \qquad (19.32)$$

5. Mutual Inductance

To develop an equivalent coupling network for an air-core transformer, we require one parameter which is common to both primary and secondary circuits. This parameter is the **mutual inductance** of the pair of coils. In Chapter 9, we found that a coil has a *self*-inductance of one henry when current in that coil *changing* at the rate of one ampere per second induces a cemf of one volt in the coil. Applying the same line of thought to the process of mutual induction, we can say that

> **A pair of magnetically-coupled coils has a mutual inductance of one henry when current changing at the rate of one ampere per second in one coil induces an average emf of one volt in the other coil.**
>
> **The letter symbol for mutual inductance is M.**

Therefore, by definition, $$E_{s_{av}} = \frac{MI_{p_m}}{t} \qquad (19.28)$$

Since the instantaneous current must rise from zero to maximum in one quarter of a cycle of a sine wave,

$$E_{s_{av}} = \frac{MI_{p_m}}{1/4f} = 4fMI_{p_m}$$

But $$E_{av} = \frac{2}{\pi}E_m \qquad (11.15)$$

Therefore $$\frac{2}{\pi}E_{s_m} = 4fMI_{p_m}$$

from which $$E_s = 2\pi fMI_p \qquad (19.29)$$

Rearranging Equation (19.29) gives

$$\frac{E_s}{I_p} = 2\pi fM = X_M \qquad (19.30)$$

where X_M is called the **mutual reactance** *of the magnetically coupled windings.*

> **Example 6:** 120 v, 60 c is applied to the primary of a transformer whose primary inductance is 5 h. The open-circuit secondary voltage is 40 v. Neglecting losses, what is the mutual inductance between the two windings?
>
> *Solution:*
>
> $$E_s = \omega MI_p$$
>
> But $$I_p = \frac{E_p}{\omega L_p}$$
>
> $$\therefore \quad E_s = \frac{M}{L_p}E_p$$

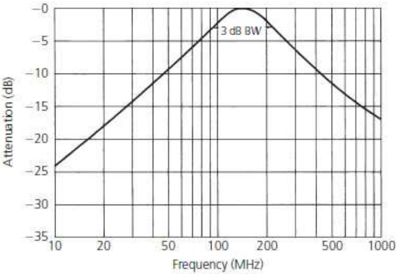

FIG. 2-8. Frequency response of an LC (inductor–capacitor) resonant circuit.

Notice, in Fig. 2-8, that as we near the resonant frequency of the tuned circuit, the slope of the resonance curve increases to 12 dB/octave. This is due to the fact that we now have two significant reactances present and each one is changing at the rate of 6 dB/octave and sloping in opposite directions.

As we move away from resonance in either direction, however, the curve again settles to a 6-dB/octave slope because, again, only one reactance becomes significant.

Loaded Q

The Q of a resonant circuit was defined earlier to be equal to the ratio of the center frequency of the circuit to its 3-dB bandwidth. This “circuit Q,” , is often given the label loaded Q because it describes the passband characteristics of the resonant circuit under actual in-circuit or loaded conditions. The loaded Q of a resonant circuit is dependent upon three main factors.

1 The source resistance (Rs).

2 The load resistance (RL).

3 The component Q.

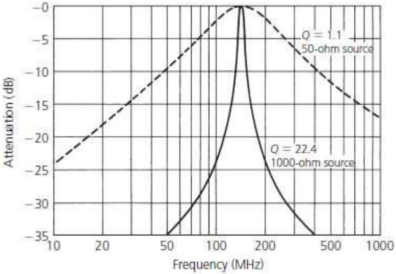

FIG. 2-10. The effect of R_s and R_L on loaded Q.

Raising the source impedance will increase the Q of our resonant circuit.

Mutual Inductance
From an unknown text
Pdf downloaded from BGU Lab 7 web page

http://www.ee.bgu.ac.il/~intrlab/lab_number_7/

http://www.ee.bgu.ac.il/~intrlab/lab_number_7/Mutual%20Inductance%20and%20coupled%20RLC%20circuits.pdf

If an external load were attached to the resonant circuit, the effect would be to broaden or “de-Q” the response curve to a degree that depends on the value of the load resistance.

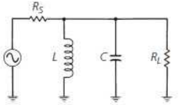

(A) *Resonant circuit with an external load*

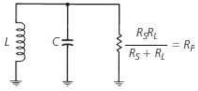

(B) *Equivalent circuit for Q calculations*

FIG. 2-11. The equivalent parallel impedance across a resonant circuit.

The resonant circuit sees an equivalent resistance of Rs in parallel with RL, as its true load. This total external resistance is, by definition, smaller in value than either Rs or RL, and the loaded Q must decrease. If we put this observation in equation form, it becomes (assuming lossless components):
Q = Rp/Xp

This illustrates that a decrease in Rp will decrease the Q of the resonant circuit and an increase in Rp will increase the circuit Q, and it also illustrates another very important point. The same effect can be obtained by keeping Rp constant and varying Xp.

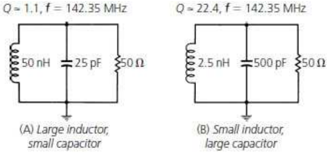

(A) Large inductor, small capacitor

(B) Small inductor, large capacitor

FIG. 2-12. Effect of Q vs. X_p at 142.35 MHz.

In most cases, we only need to involve the Q of the inductor in loaded-Q calculations. The Q of most capacitors is quite high over their useful frequency range, and the equivalent shunt resistance they present to the circuit is also quite high and can usually be neglected. Care must be taken, however, to ensure that this is indeed the case.

Impedance Transformation

Low values of source and load impedance tend to load a given resonant circuit down and, thus, tend to decrease its loaded Q and increase its bandwidth. This makes it very difficult to design a simple LC high-Q resonant circuit for use between two very low values of source and load resistance.

One method of getting around this potential design problem is to make use of one of the impedance transforming circuits shown in Fig. 2-18.

---

dip much between the peaks and the response is close to what is ideally required. The bandwidth can be estimated as BW = kfo, where k is the coefficient of coupling and fo is the resonant frequency of each circuit.

Figure 3 shows an example of secondary circuit response Vc(f) (voltage measured on capacitor) for different coupling coefficients.

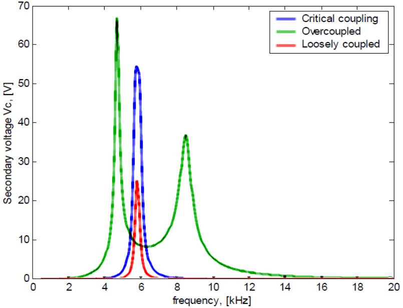

Fig 3

(R = 250 Ohm, C = 5 nF, L = 0.15 H, Vin =5V) inductance (Ze ≈ Z1 + ω2M2·jωC), and an inductance as a capacitance (Ze ≈ Z1 + ω2M2/ jωL).

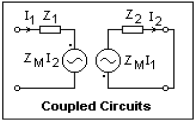

Fig 2

At resonance, the reflected impedance is resistive, and acts to lower the Q of the primary, and thereby to reduce the output. This is counteracted by the increased coupling, which increases the output. The lower Q gives a wider passband. At frequencies lower than exact resonance, the reflected impedance is inductive, which adds to the inductance of the primary and resonates at a lower frequency, producing a peak in the output. At frequencies higher than exact resonance, the reflected impedance is capacitive, which cancels part of the inductance and causes the circuit to resonate at a higher frequency, producing the other peak.

As the coupling is reduced, the response becomes single-peaked at critical coupling, and then decreases as the coupling is made even looser.

The critical coupling coefficient is given by $k_c = \frac{1}{\sqrt{Q_1 \cdot Q_2}}$.

in terms of the Q's of the individual tuned circuits.

Frequently it is assumed that optimum coupling occurs for k ≈ 1.5kc. In this case, the response is double-peaked, but does not

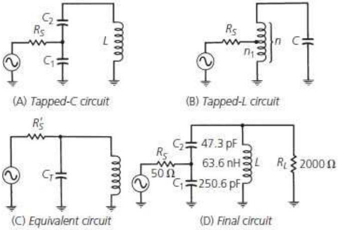

(A) Tapped-C circuit (B) Tapped-L circuit

(C) Equivalent circuit (D) Final circuit

FIG. 2-18. Two methods used to perform an impedance transformation.

For the tapped- C transformer (Fig. 2-18A), we use the formula:

R's = R's (1 + C1/ C2)^2

The equivalent capacitance (CT) that will resonate with the inductor is equal to C1 in series with C2, or:

CT = C1C2/(C1 + C2)

For the tapped-L network of Fig. 2-18B, we use the following formula:

R's = R's (1 + n/ n1)^2

Coupling of Resonant Circuits

In many applications where steep passband skirts and small shape factors are needed, a single resonant circuit might not be sufficient. In situations such as this, individual resonant circuits are often coupled together to produce more attenuation at certain frequencies than would normally be available with a single resonator.

The coupling mechanism that is used is generally chosen specifically for each application, as each type of coupling has its own peculiar characteristics that must be dealt with. The most common forms of coupling are: capacitive, inductive, transformer (mutual), and active (transistor).

Capacitive Coupling

If capacitor C12 of Fig. 2-19 is too large, too much coupling occurs and the frequency response broadens drastically with two response peaks in the filter's passband. If capacitor C12 is too small, not enough signal energy is passed from one resonant circuit to the other and the insertion loss can increase to an unacceptable level. The compromise solution to these two extremes is the point of critical coupling, where we obtain a reasonable bandwidth and the lowest possible insertion loss and, consequently, a maximum transfer of signal power.

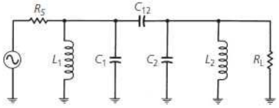

FIG. 2-19. Capacitive coupling.

Thus, the solution from which the frequency response can be obtained is:

$$I_1 = \frac{aV_0}{a^2 - b^2} \text{ and } I_2 = -\frac{bV_0}{a^2 - b^2} \qquad (4)$$

Resonance occurs at the 2 frequencies given by the following equations:

$$\omega_1^2 = \frac{1}{LC \cdot \left[1 - \frac{M}{L}\right]} \text{ and } \omega_2^2 = \frac{1}{LC \cdot \left[1 + \frac{M}{L}\right]}$$

Here, the coupling coefficient k = M/L (for L1=L2).

The behavior of the circuit can be understood qualitatively on the basis of the reflected impedance (or coupled impedance).

A transformer (or inductively coupled circuit) is said to "reflect" impedance in the secondary into the primary circuit. Consider the coupled circuits shown in Fig 2. The positive direction of the currents is chosen into the polarity mark on the generator representing the induced voltages, so that Kirchhoff's equations are

$$\mathbf{Z}_1\mathbf{I}_1 + \mathbf{Z}_M\mathbf{I}_2 = \mathbf{V}_1, \text{ and } \mathbf{Z}_M\mathbf{I}_1 + \mathbf{Z}_2\mathbf{I}_2 = 0. \text{ (see Eq.2)}$$

ZM is the mutual impedance jωM, Z1 includes the source impedance, and Z2 the secondary load. These equations may be solved for the equivalent primary impedance

$$\mathbf{Z}_e = \mathbf{V}_1/\mathbf{I}_1 = \mathbf{Z}_1 - \mathbf{Z}_M^2/\mathbf{Z}_2. \text{ (from eqs. (2) and (4))}$$

The reflected impedance is then ω2M2/Z2. Note that a resistance is reflected as a resistance, a capacitance as an

The coupling constant is independent of the number of turns in a coil. The number of turns in a coil determines the magnetic field, which will be produced for a given current. The coupling constant is concerned with how the lines of magnetic force produced by one coil interact with another coil, and hence the coupling constant between two air spaced coils depends only on their physical size and disposition in space. Hence to obtain the best coupling between primary and secondary in an air-cored transformer we can only change the size and spatial relationships of the coils.

Kirchhoff's voltage law equations for the primary and secondary loops are given by

$$\left[R + j\omega L + \frac{1}{j\omega C}\right] \cdot I_1 + j\omega M \cdot I_2 = V_0 \qquad (2)$$

$$j\omega M \cdot I_1 + \left[R + j\omega L + \frac{1}{j\omega C}\right] \cdot I_2 = 0$$

(it is assumed that R1+RL1 = R2+RL2 =R and L1 = L2).
We can write these eqs. in the matrix form as follows

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix} \cdot \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} V_0 \\ 0 \end{bmatrix} \qquad (3)$$

where, $a = \left[R + j\omega L + \frac{1}{j\omega C}\right]$ and $b = j\omega M$.

Following Cramer's rule $I_1 = \frac{\Delta_1}{\Delta}$ and $I_2 = \frac{\Delta_2}{\Delta}$,

where $\Delta_1 = \begin{vmatrix} V_0 & b \\ 0 & a \end{vmatrix}$, $\Delta_2 = \begin{vmatrix} a & V_0 \\ b & 0 \end{vmatrix}$, and $\Delta = a^2 - b^2$.

There are instances in which overcoupling or undercoupling might serve a useful purpose in a design, such as in tailoring a specific frequency response that a critically coupled filter cannot provide.

In this section, we will only concern ourselves with critical coupling as it pertains to resonant circuit design.

The loaded Q of a critically coupled two-resonator circuit is approximately equal to 0.707 times the loaded Q of one of its resonators. Therefore, the 3-dB bandwidth of a two-resonator circuit is actually wider than that of one of its resonators. Thus, the main purpose of the two-resonator passively coupled filter is not to provide a narrower 3-dB bandwidth, but to increase the steepness of the stopband skirts and, thus, to reach an ultimate attenuation much faster than a single resonator could.

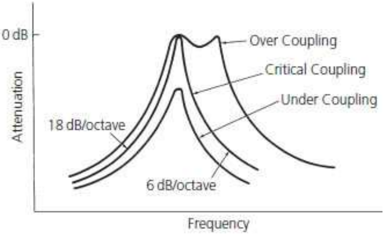

FIG. 2-20. The effects of various values of capacitive coupling on passband response.

Notice that even for the critically coupled case, the response curve is not symmetric around the center frequency but is skewed somewhat. The lower frequency portion of the response plummets down at the rate of 18-dB/octave while the upper slope decreases at only 6-dB/octave.

Below resonance, we have the circuit of Fig. 2-22A. The reactance of the two resonant-circuit capacitors (Fig. 2-19) has increased, and the reactance of the two inductors has decreased to the point that only the inductor is seen as a shunt element and the capacitors can be ignored. This leaves three reactive components and each contributes 6-dB/octave to the response.

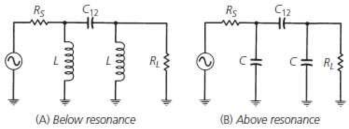

(A) Below resonance (B) Above resonance

FIG. 2-22. Equivalent circuit of capacitively coupled resonant circuits.

Above resonance, the equivalent circuit approaches the configuration of Fig. 2-22B. Here the inductive reactance has increased above the capacitive reactance to the point where the inductive reactance can be ignored as a shunt element. We now have an arrangement of three capacitors that effectively look like a single shunt capacitor and yields a slope of 6-dB/octave.

Inductive Coupling

Two inductively coupled RLC circuits

Ben-Gurion University

http://www.ee.bgu.ac.il/~intrlab/lab_number_7/Two%20inductively%20coupled%20RLC%20circuits.pdf

Two inductively coupled RLC circuits are shown in Figure 1. Having 2 circuits gives 2 resonant frequencies whose separation depends on the value of the mutual inductance M (the ratio of the voltage in the secondary to the rate of change of primary current with time, and the unit is the henry. This has a reactance at the operating frequency Xm = ωM).

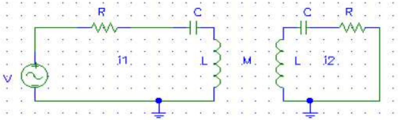

The mutual inductance coupling between primary and secondary can be related to their self-inductance by means of the coupling constant k:

$$k = \frac{M}{\sqrt{L_1 \cdot L_2}}$$

Notice, that since k is defining the relationship between magnetic flux linkages in the circuit, it can never be greater than 1. A value of 1 means that all the flux produced by the primary is linked with the secondary and vice versa. A value of k greater than 1 would mean that more than all of the flux produced by the primary is linked with the secondary.

control and high Q are essential to good response and selectivity.

If the primary circuit is made to resonate at a different frequency from the secondary, audio response is much worse, and considerable distortion is likely. Moreover, the response at mean frequency is less than it would be if the circuits were properly tuned. Air-core transformers are usually made adjustable for tuning and coupling.

Two types of inductively coupled resonant circuits are shown in Fig. 2-23. One type (Fig. 2-23A) uses a series inductor or coil to transfer energy from the first resonator to the next, and the other type (Fig. 2-23B) uses transformer coupling for the same purpose.

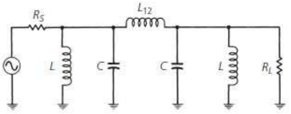

(A) *Series inductor*

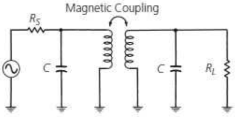

(B) *Transformer*

FIG. 2-23. Inductive coupling.

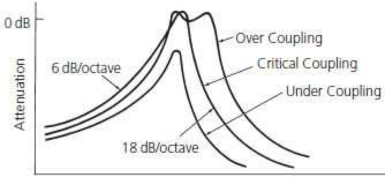

(A) *Inductive coupling*

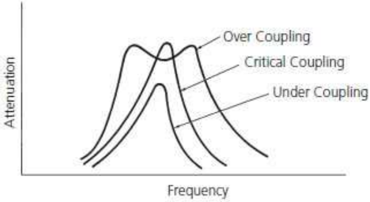

(B) *Transformer coupling*

FIG. 2-24. The effects of various values of inductive coupling on passband response.

different values of coupling. If the value of coupling is such that

$$X_M = \sqrt{R_1 R_2}$$

we obtain a condition similar to that of equation 103, in which the maximum power or current is produced in the secondary circuit. Maximum current through condenser C2 gives maximum voltage E2. This value of coupling is known as the critical value. Smaller coefficient of coupling gives a smaller maximum value of E2. Greater coefficient of coupling results in a "double hump" as shown in Fig. 173.

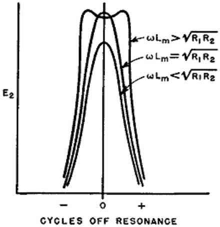

Fig. 173. Response curves for circuit of Fig. 172.

The heights of resonant peaks and frequency distance between peaks depend upon circuit Q and coefficient of coupling k. The double hump curve of Fig. 173 is desirable because, with modulated waves, frequencies in adjacent channels are rejected; yet very little attenuation is offered to audio frequencies which effectively add or subtract from the carrier frequency normally corresponding to resonance. Close tuning

$$k = \frac{L_m}{\sqrt{L_1 L_2}} = \frac{L_m}{\sqrt{\left(L_m + \frac{L_s}{2}\right)^2}}$$ [104a]

$$= \frac{1}{1 + (L_s/2L_m)}$$

If Lm >> Ls

$$k \approx 1 - \frac{L_s}{2L_m}$$ [104b]

Equations 104 (a) and (b) are useful in estimating approximate transformer band width.

A tuned air-core transformer often used in receivers is shown in Fig. 172.

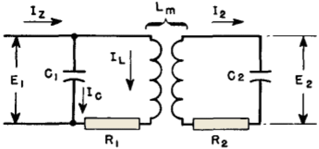

Fig. 172. Tuned air-core transformer.

Here a sinusoidal voltage E1 may be impressed on the primary circuit by a vacuum tube amplifier. Resistances R1 and R2 are usually the inevitable resistance of coils, but occasionally resistance is added to change the circuit response. The value of voltage E2 obtained from this circuit depends on the impressed frequency; in Fig. 173 it is shown for resonance at three

---

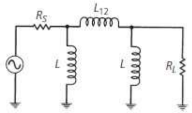

(A) *Below resonance*

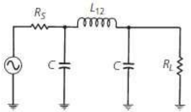

(B) *Above resonance*

FIG. 2-25. Equivalent circuit of inductively coupled resonant circuits.

states that the mutual reactance XM is the geometric mean between the two values of resistance.

The ratio of mutual inductance to the geometric mean of the primary and secondary self-inductances is the coupling coefficient:

$$k = L_m/\sqrt{(L_1 L_2)} \qquad [104]$$

The value of k is never greater than unity, even when coils are interleaved to the maximum possible extent. Values of k down to 0.01 or lower are common at high frequencies.

Coupling coefficient is related to untuned transformer open- and short-circuit reactance by means of the transformer equivalent circuit shown in Fig. 107(a), p. 147. Assume that the transformer has a 1:1 ratio, and leakage inductance is equally divided between primary and secondary windings. Then if L1 and L2 are the self-inductances of primary and secondary, respectively, Ls is the total leakage inductance (measured in the primary with secondary short-circuited), and Lm the mutual inductance,

$$L_1 = \frac{X_P + X_N}{2\pi f} = \frac{L_s}{2} + L_m$$
$$L_2 = \frac{X_S + X_N}{2\pi f} = \frac{L_s}{2} + L_m$$

From equation 104,

$$0 = Z_2 I_2 + j\omega L_m I_1 \qquad [101]$$

where ω = 2π times operating frequency, and Lm is the mutual inductance between the primary and secondary coils.

From equation 101 we see that the voltage in the secondary coil is numerically equal to ωLmI1, the product of primary current and mutual reactance at the frequency of applied voltage E1. The equivalent impedance of the circuit of Fig. 171 when referred to the primary side is given by

$$Z' = Z_1 + (X_M{}^2/Z_2) \qquad [102]$$

where $X_M = j\omega L_m$.

In the above formulas, the impedances Z1, Z2, and Z' are complex quantities whose real and imaginary terms depend upon the values of resistance, inductance, and capacitance in the circuit. One common practical case arises when the primary resistance is zero, or virtually zero, and the secondary coil is tuned to resonance so that Z2 is a pure resistance R2. Under these conditions, equation 102 reduces to

$$R' = X_M{}^2/R_2 \qquad [103]$$

where R' is the equivalent resistance in the primary.

Equation 103 gives the value of mutual inductance required for coupling a resistance R2 so that it will appear like resistance R' with a maximum power transfer between the two coils, and

---

RESONANCE

Tony R. Kuphaldt

http://www.allaboutcircuits.com/vol_2/chpt_6/1.html

6.1 An electric pendulum

Capacitors store energy in the form of an electric field, and electrically manifest that stored energy as a potential: static voltage. Inductors store energy in the form of a magnetic field, and electrically manifest that stored energy as a kinetic motion of electrons: current. Capacitors and inductors are flip-sides of the same reactive coin, storing and releasing energy in complementary modes. When these two types of reactive components are directly connected together, their complementary tendencies to store energy will produce an unusual result.

If either the capacitor or inductor starts out in a charged state, the two components will exchange energy between them, back and forth, creating their own AC voltage and current cycles. If we assume that both components are subjected to a sudden application of voltage (say, from a momentarily connected battery), the capacitor will very quickly charge and the inductor will oppose change in current, leaving the capacitor in the charged state and the inductor in the discharged state: (Figure 6.1)

The capacitor will begin to discharge, its voltage decreasing. Meanwhile, the inductor will begin to build up a “charge” in the form of a magnetic field as current increases in the circuit: (Figure 6.2)

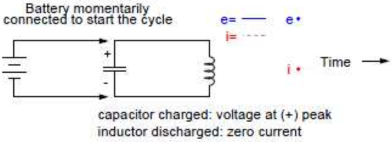
Battery momentarily
connected to start the cycle
e=
i=
e
i
Time
+
-
capacitor charged: voltage at (+) peak
inductor discharged: zero current

Figure 6.1: Capacitor charged: voltage at (+) peak, inductor discharged: zero current.

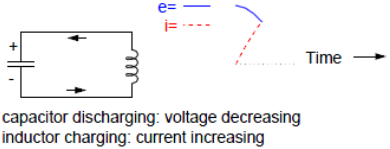
e=
i=
Time
+
-
capacitor discharging: voltage decreasing
inductor charging: current increasing

Figure 6.2: Capacitor discharging: voltage decreasing, Inductor charging: current increasing. The inductor, still charging, will keep electrons flowing in the circuit until the capacitor hasbeen completely discharged, leaving zero voltage across it: (Figure 6.3)

Air-Core Transformers

Reuben Lee

http://www.vias.org/eltransformers/lee_electronic_transformers_07b_22.html

Air-Core Transformers

Transformers considered hitherto have had iron or ferrite cores. A class of transformers is widely used in radio-frequency circuits without cores or with small slugs of powdered iron. In a transformer with an iron core, the exciting current required for inducing the secondary voltage is a small percentage of the load component of current. In an air-core transformer all the current is exciting current and induces a secondary voltage proportional to the mutual inductance.

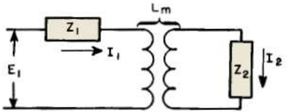
Lm
Z1
I1
E1
Z2
I2

Fig. 171. General case of inductive coupling.

Consider the circuit of Fig. 171 in which Z1 is complex and includes the self-inductance of the primary coil. Likewise, secondary impedance Z2 is complex and includes the self-inductance of the secondary coil. With a sinusoidal voltage applied, Kirchhoff's laws give the following:

$$E_1 = Z_1 I_1 + j\omega L_m I_2 \qquad [100]$$

F.E. Terman, "**Radio Engineers' Handbook**," London, McGraw-Hill, 1st ed., Sep. 1950. P.71

http://www.itermoionici.it/letteratura_files/Radio-Engineers-Handbook.pdf

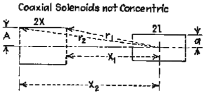

Fig. 42.—Coaxial solenoids not concentric.

16. Mutual Inductance between Single-layer Solenoids. *Coaxial Coils Not Concentric.*—When the configuration is as shown in Fig. 42, the mutual inductance is given by

$$M = 0.02505 \frac{a^2A^2n_1n_2}{4lx}(K_1k_1 + K_3k_3 + K_5k_5) \text{ microhenrys} \tag{85}$$

where a = smaller radius, measured from the axis of the coil to the center of the wire, in.

A = larger radius, measured in the same way, in.

$2l$ = length of coil of smaller radius = number of turns times pitch of winding, in.

$2x$ = length of coil of larger radius, measured in the same way, in.

n_1 and n_2 = total number of turns on the two coils.

$$K_1 = \frac{2}{A^2}\left(\frac{x_2}{r_2} - \frac{x_1}{r_1}\right), \qquad k_1 = 2l$$

$$x_1 = D - x, \qquad r_1 = \sqrt{x_1^2 + A^2}$$

$$x_2 = D + x, \qquad r_2 = \sqrt{x_2^2 + A^2}$$

$$K_3 = \frac{1}{2}\left(\frac{x_1}{r_1^5} - \frac{x_2}{r_2^5}\right), \qquad k_3 = a^2l\left(3 - \frac{4l^2}{a^2}\right)$$

$$K_5 = -\frac{A^2}{8}\left[\frac{x_1}{r_1^9}\left(3 - \frac{4x_1^2}{A^2}\right) - \frac{x_2}{r_2^9}\left(3 - \frac{4x_2^2}{A^2}\right)\right]$$

$$k_5 = a^4l\left(\frac{5}{2} - 10\frac{l^2}{a^2} + 4\frac{l^2}{a^4}\right)$$

D = axial distance between centers of coils, in.

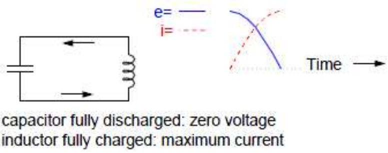
e=
i=
Time
capacitor fully discharged: zero voltage
inductor fully charged: maximum current

Figure 6.3: Capacitor fully discharged: zero voltage, inductor fully charged: maximum current. The inductor will maintain current flow even with no voltage applied. In fact, it will generate a voltage (like a battery) in order to keep current in the same direction. The capacitor, being the recipient of this current, will begin to accumulate a charge in the opposite polarity as before: (Figure 6.4)

When the inductor is finally depleted of its energy reserve and the electrons come to a halt, the capacitor will have reached full (voltage) charge in the opposite polarity as when it started: (Figure 6.5) Now we’re at a condition very similar to where we started: the capacitor at full charge and zero current in the circuit. The capacitor, as before, will begin to discharge through the inductor, causing an increase in current (in the opposite direction as before) and a decrease in voltage as it depletes its own energy reserve: (Figure 6.6)

Eventually the capacitor will discharge to zero volts, leaving the inductor fully charged with

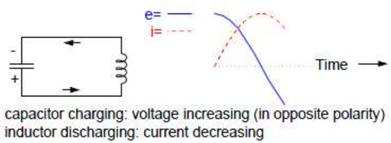
e=
i=
-
+
Time
capacitor charging: voltage increasing (in opposite polarity)
inductor discharging: current decreasing

Figure 6.4: Capacitor charging: voltage increasing (in opposite polarity), inductor discharging: current decreasing.

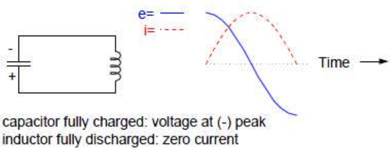
e=
i=
-
+
Time
capacitor fully charged: voltage at (-) peak
inductor fully discharged: zero current

Figure 6.5: Capacitor fully charged: voltage at (-) peak, inductor fully discharged: zero current.

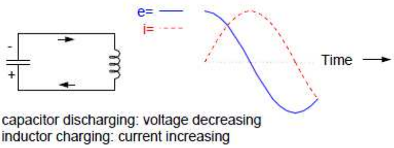
e=
i=
-
+
Time
capacitor discharging: voltage decreasing
inductor charging: current increasing

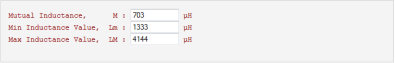

Mutual Inductance, M : 703 µH
Min Inductance Value, Lm : 1333 µH
Max Inductance Value, LM : 4144 µH

When the coils can not rotate but only move along their axis (coaxial coils), assuming a distance D between the two coils centres, we have:

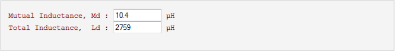

Mutual Inductance, Md : 10.4 µH
Total Inductance, Ld : 2759 µH

The total inductance value for this case is computed assuming a positive mutual inductance (coils not rotated 180°).

References:

[1] R. Lundin, "A Handbook Formula for the Inductance of a Single-Layer Circular Coil," Proc. IEEE, vol. 73, no. 9, pp. 1428-1429, Sep. 1985.
[2] F.E. Terman, "Radio Engineers' Handbook," London, McGraw-Hill, 1st ed., Sep. 1950.

rotating a coil 180° gives a variation of the total inductance of 4Lm.

Two coaxial coils If the two coaxial coils are moved along their axis instead (D<>0), the mutual inductance decreaes, eventually reaching zero at infinity, but does not change sign. This means that there is less variation in the total inductance just moving the coils with respect to rotating them.

The form below computes the main parameters for the two types of variometer described above. The self-inductances are calculated using the formulas in [1], while the mutual inductances use the formulas from [2], with some minor corrections.

These formulas agree very well with the results of some electromagnetic simulations I have done, see the page for details.

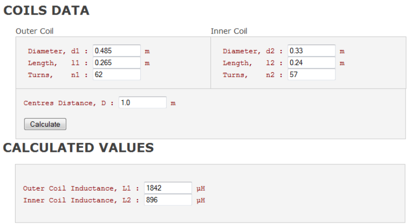
COILS DATA

Outer Coil

Diameter, d1 : 0.485 m
Length, l1 : 0.265 m
Turns, n1 : 62

Inner Coil

Diameter, d2 : 0.33 m
Length, l2 : 0.24 m
Turns, n2 : 57

Centres Distance, D : 1.0 m

CALCULATED VALUES

Outer Coil Inductance, L1 : 1842 µH
Inner Coil Inductance, L2 : 896 µH

In a typical variometer the two coils are concentric (i.e. D=0) and the overall inductance is varied rotating the inner coil; in this case we have:

Figure 6.6: Capacitor discharging: voltage decreasing, inductor charging: current increasing. full current through it:

(Figure 6.7)

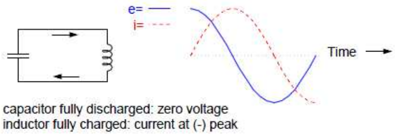
e=
i=
Time
capacitor fully discharged: zero voltage
inductor fully charged: current at (-) peak

Figure 6.7: Capacitor fully discharged: zero voltage, inductor fully charged: current at (-) peak.

The inductor, desiring tomaintain current in the same direction, will act like a source again, generating a voltage like a battery to continue the flow. In doing so, the capacitor will begin to charge up and the current will decrease in magnitude: (Figure 6.8)

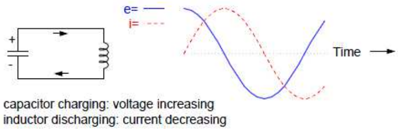
e=
i=
+
-
Time
capacitor charging: voltage increasing
inductor discharging: current decreasing

Figure 6.8: Capacitor charging: voltage increasing, inductor discharging: current decreasing.

Eventually the capacitor will become fully charged again as the inductor expends all of its energy reserves trying to maintain current. The voltage will once again be at its positive peak and the current at zero. This completes one full cycle of the energy exchange between the capacitor and inductor: (Figure 6.9)

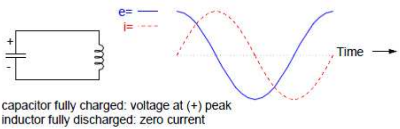

Figure 6.9: Capacitor fully charged: voltage at (+) peak, inductor fully discharged: zero current.

This oscillation will continue with steadily decreasing amplitude due to power losses from stray resistances in the circuit, until the process stops altogether. Overall, this behavior is akin to that of a pendulum: as the pendulum mass swings back and forth, there is a transformation of energy taking place from kinetic (motion) to potential (height), in a similar fashion to the way energy is transferred in the capacitor/inductor circuit back and forth in the alternating forms of current (kinetic motion of electrons) and voltage (potential electric energy).

At the peak height of each swing of a pendulum, the mass briefly stops and switches directions. It is at this point that potential energy (height) is at a maximum and kinetic energy (motion) is at zero. As the mass swings back the other way, it passes quickly through a point where the string is pointed

---

Variometer Design

Claudio Girardi
http://www.qsl.net/in3otd/variodes.html

Variometer Design

A variometer consists usually of two coaxial coils, connected in series, where the coils relative position can be varied in some way. If L1 and L2 are the self-inductances of the first and second coil and Lm is the mutual inductance between the two, the total inductance can be written as Ltot=L1+L2±2Lm. The mutual inductance is defined as the flux linked by the turns of an inductance when the other carries a unit current; this of course depends not only on the coil length and diameter but also on their relative position.
For two coaxial coils, like the ones in the picture, the mutual inductance is at maximum when they are also concentrical, i.e. D=0.

If the two concentrical coils are rotated, so that their axes are not parallel any more, the mutual inductance decreases, reaching zero (almost) when the angle is 90°. Continuing the rotation beyond 90° the mutual inductance increases again, but this time with the opposite sign. So, according to the above formula,

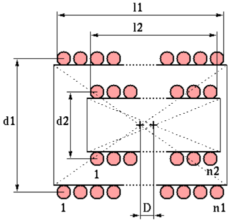

straight down. At this point, potential energy (height) is at zero and kinetic energy (motion) is at maximum. Like the circuit, a pendulum's back-and-forth oscillation will continue with a steadily dampened amplitude, the result of air friction (resistance) dissipating energy. Also like the circuit, the pendulum's position and velocity measurements trace two sine waves (90 degrees out of phase) over time: (Figure 6.10)

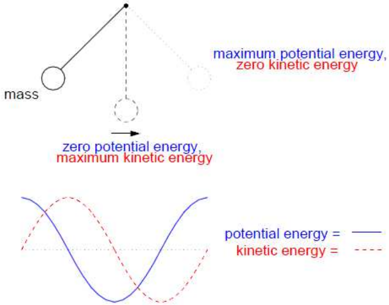
maximum potential energy,
zero kinetic energy
mass
zero potential energy,
maximum kinetic energy
potential energy = ——
kinetic energy = - - - -

Figure 6.10: Pendelum transfers energy between kinetic and potential energy as it swings low to high.

In physics, this kind of natural sine-wave oscillation for a mechanical system is called Simple Harmonic Motion (often abbreviated as "SHM"). The same underlying principles govern both the oscillation of a capacitor/inductor circuit and the action of a pendulum, hence the similarity in effect. It is an interesting property of any pendulum that its periodic time is governed by the length of the string holding the mass, and not the weight of the mass itself. That is why a pendulum will keep swinging at the same frequency as the oscillations decrease in amplitude. The oscillation rate is independent of the amount of energy stored in it.

The same is true for the capacitor/inductor circuit. The rate of oscillation is strictly dependent on the sizes of the capacitor and inductor, not on the amount of voltage (or current) at each respective peak in the waves. The ability for such a circuit to store energy in the form of oscillating voltage and current has earned it the name tank circuit. Its property of maintaining a single, natural frequency regardless of how much or little energy is actually being stored in it gives it special significance in electric circuit design.

However, this tendency to oscillate, or resonate, at a particular frequency is not limited to circuits exclusively designed for that purpose. In fact, nearly any AC circuit with a combination of capacitance and inductance (commonly called an “LC circuit”) will tend to manifest unusual effects when the AC power source frequency approaches that natural frequency. This is true regardless of the circuit’s intended purpose.

If the power supply frequency for a circuit exactly matches the natural frequency of the circuit’s LC combination, the circuit is said to be in a state of resonance. The unusual effects will reach maximum in this condition of resonance. For this reason, we need to be able to predict what the resonant frequency will be for various combinations of L and C, and be aware of what the effects of resonance are.

$$\Phi = \mu_0 \frac{NI}{l}A. \;\; L = \frac{N\Phi}{I} = \frac{N}{\cancel{I}}\mu_0 \frac{N\cancel{I}}{l}A$$

$$L = \mu_0 \frac{N^2}{l}A \;\text{ or }\; \boxed{L = \mu_0 n^2 A l} \;\text{ where } n = \frac{N}{l}.$$

(Note how L is independent of the current I.)

be turned into an equation by introducing a constant. Call this constant L, the ***self-inductance*** (or simply ***inductance***) of the coil:

$$N\Phi = LI \quad \text{or} \quad L = \frac{N\Phi}{I}$$

As with mutual inductance, the unit of self-inductance is the henry.

The self-induced emf can now be calculated using Faraday’s law:

$$\mathcal{E} = -N\frac{\Delta\Phi}{\Delta t} = -\frac{\Delta(N\Phi)}{\Delta t} = -\frac{\Delta(LI)}{\Delta t} = -L\frac{\Delta I}{\Delta t}$$

$$\boxed{\mathcal{E} = -L\frac{\Delta I}{\Delta t}}$$

The above formula is the ***emf due to self-induction***.

Example

Find the formula for the self-inductance of a solenoid of N turns, length l, and cross-sectional area A.

Assume that the solenoid carries a current I. Then the magnetic flux in the solenoid is

---

- REVIEW:

- A capacitor and inductor directly connected together form something called a tank circuit, which oscillates (or resonates) at one particular frequency. At that frequency, energy is alternately shuffled between the capacitor and the inductor in the form of alternating voltage and current 90 degrees out of phase with each other.

- When the power supply frequency for an AC circuit exactly matches that circuit’s natural oscillation frequency as set by the L and C components, a condition of resonance will have been reached.

6.2 Simple parallel (tank circuit) resonance

A condition of resonance will be experienced in a tank circuit (Figure 6.11) when the reactance of the capacitor and inductor are equal to each other. Because inductive reactance increases with increasing frequency and capacitive reactance decreases with increasing frequency, there will only be one frequency where these two reactances will be equal.

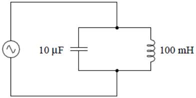

Figure 6.11: Simple parallel resonant circuit (tank circuit).

In the above circuit, we have a 10 µF capacitor and a 100 mH inductor. Since we know the equations for determining the reactance of each at a given frequency, and we're looking for that point where the two reactances are equal to each other, we can set the two reactance formulae equal to each other and solve for frequency algebraically:

$$X_L = 2\pi fL \qquad X_C = \frac{1}{2\pi fC}$$

. . . setting the two equal to each other, representing a condition of equal reactance (resonance) . . .

$$2\pi fL = \frac{1}{2\pi fC}$$

Multiplying both sides by f eliminates the f term in the denominator of the fraction . . .

$$2\pi f^2L = \frac{1}{2\pi C}$$

Dividing both sides by 2pL leaves f2 by itself on the left-hand side of the equation . . .

$$f^2 = \frac{1}{2\pi 2\pi LC}$$

Taking the square root of both sides of the equation leaves f by itself on the left side . . .

$$f = \frac{\sqrt{1}}{\sqrt{2\pi 2\pi LC}}$$

. . . simplifying . . .

The apparatus used in Experiment EM-11B consists of two coaxial solenoids. A solenoid is essentially just a coil of wire. For a long, tightly-wound solenoid of *n* turns per unit length carrying current *I* the magnetic field over its cross-section is nearly constant and given by $B = \mu_0 nI$. Assume that the two solenoids have the same cross-sectional area *A*. Find a formula for the mutual inductance of the solenoids.

The magnetic flux in the primary coil is

$\Phi_1 = \mu_0 \frac{N_1 I_1}{l_1} A$ where l_1 is the length of the primary coil.

$$M = \frac{N_2\Phi_2}{I_1} \text{ But } \Phi_2 = \Phi_1 : \ M = \frac{N_2\Phi_2}{I_1} = \frac{N_2}{\cancel{I_1}}\mu_0 \frac{N_1 \cancel{I_1}}{l_1} A; \quad \boxed{M = \mu_0 \frac{N_1 N_2}{l_1} A}$$

(Note how M is independent of the current I_1.)

Self-Inductance

A current-carrying coil produces a magnetic field that links its own turns. If the current in the coil changes the amount of magnetic flux linking the coil changes and, by Faraday's law, an emf is produced in the coil. This emf is called a *self-induced* emf.

Let the coil have N turns. Assume that the same amount of magnetic flux Φ links each turn of the coil. The net flux linking the coil is then $N\Phi$. This net flux is proportional to the magnetic field, which, in turn, is proportional to the current I in the coil. Thus we can write $N\Phi \propto I$. This proportionality can

Let the primary coil have N_1 turns and the secondary coil have N_2 turns. Assume that the same amount of magnetic flux Φ_2 from the primary coil links each turn of the secondary coil. The net flux linking the secondary coil is then $N_2\Phi_2$. This net flux is proportional to the magnetic field, which, in turn, is proportional to the current I_1 in the primary coil. Thus we can write $N_2\Phi_2 \propto I_1$. This proportionality can be turned into an equation by introducing a constant. Call this constant M, the ***mutual inductance*** of the two coils:

$$N_2\Phi_2 = MI_1 \quad \text{or} \quad M = \frac{N_2\Phi_2}{I_1}$$

the unit of inductance is $\dfrac{\text{wb}}{\text{A}} = henry\ (\text{H})$ named after Joseph Henry.

The emf induced in the secondary coil may now be calculated using Faraday's law:

$$\mathcal{E}_2 = -N_2\frac{\Delta\Phi_2}{\Delta t} = -\frac{\Delta\left(N_2\Phi_2\right)}{\Delta t} = -\frac{\Delta\left(MI_1\right)}{\Delta t} = -M\frac{\Delta I_1}{\Delta t}$$

$$\boxed{\mathcal{E}_2 = -M\frac{\Delta I_1}{\Delta t}}$$

The above formula is the ***emf due to mutual induction***.

Example

$$\boxed{f = \frac{1}{2\pi\sqrt{LC}}}$$

So there we have it: a formula to tell us the resonant frequency of a tank circuit, given the values of inductance (L) in Henrys and capacitance (C) in Farads. Plugging in the values of L and C in our example circuit, we arrive at a resonant frequency of 159.155 Hz.

What happens at resonance is quite interesting. With capacitive and inductive reactances equal to each other, the total impedance increases to infinity, meaning that the tank circuit draws no current from the AC power source! We can calculate the individual impedances of the 10 µF capacitor and the 100mH inductor and work through the parallel impedance formula to demonstrate this mathematically:

$$X_L = 2\pi fL$$

$$X_L = (2)(\pi)(159.155\ \text{Hz})(100\ \text{mH})$$

$$X_L = 100\ \Omega$$

$$X_C = \frac{1}{2\pi fC}$$

$$X_C = \frac{1}{(2)(\pi)(159.155\ \text{Hz})(10\ \mu\text{F})}$$

$$X_C = 100\ \Omega$$

As you might have guessed, I chose these component values to give resonance impedances that were easy to work with (100 even). Now, we use the parallel impedance formula to see what happens to total Z:

$$Z_{\text{parallel}} = \frac{1}{\dfrac{1}{Z_L} + \dfrac{1}{Z_C}}$$

$$Z_{\text{parallel}} = \frac{1}{\dfrac{1}{100\ \Omega \angle 90^\circ} + \dfrac{1}{100\ \Omega \angle -90^\circ}}$$

$$Z_{\text{parallel}} = \frac{1}{0.01 \angle -90^\circ + 0.01 \angle 90^\circ}$$

$$Z_{\text{parallel}} = \frac{1}{0} \qquad \textit{Undefined!}$$

We can't divide any number by zero and arrive at a meaningful result, but we can say that the result approaches a value of infinity as the two parallel impedances get closer to each other. What this means in practical terms is that, the total impedance of a tank circuit is infinite (behaving as an open circuit) at resonance. We can plot the consequences of this over a wide power supply frequency range with a short SPICE simulation: (Figure 6.12)

The 1 pico-ohm (1 p) resistor is placed in this SPICE analysis to overcome a limitation of SPICE: namely, that it cannot analyze a circuit containing a direct inductor-voltage source loop. (Figure 6.12) A very low resistance value was chosen so as to have minimal effect on circuit behavior.

Mutual Inductance

Dr. David F. Cattell
faculty.ccp.edu/faculty/dcattell/Sp12/.../Mutual%20Inductance.doc

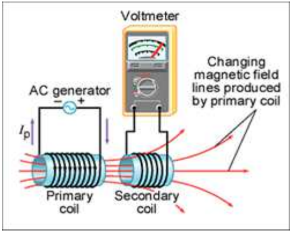

Suppose we hook up an AC generator to a solenoid so that the wire in the solenoid carries AC. Call this solenoid the primary *coil*. Next place a second solenoid connected to an AC voltmeter near the primary coil so that it is coaxial with the primary coil. Call this second solenoid the *secondary coil*. See the figure at the right.

The alternating current in the primary coil produces an alternating magnetic field whose lines of flux *link* the secondary coil (like thread passing through the eye of a needle). Hence the secondary coil encloses a *changing* magnetic field. By Faraday's law of induction this changing magnetic flux induces an emf in the secondary coil. This effect in which changing current in one circuit induces an emf in another circuit is called ***mutual induction***.

the resulting mutual inductance will be equal to the geometric mean of the two individual inductances of the coils.

Also when the inductances of the two coils are the same and equal, L 1 is equal to L 2, the mutual inductance that exists between the two coils will equal the value of one single coil as the square root of two equal values is the same as one single value as shown.

$$M = \sqrt{L_1 L_2} = L$$

Mutual Inductance Example No1

Two inductors whose self-inductances are given as 75mH and 55mH respectively, are positioned next to each other on a common magnetic core so that 75% of the lines of flux from the first coil are cutting the second coil. Calculate the total mutual inductance that exists between the two coils.

$$M = k\sqrt{L_1 L_2}$$

$$M = 0.75\sqrt{75\text{mH x }55\text{mH}} = 48.2\text{mH}$$

Mutual Inductance Example No2

When two coils having inductances of 5H and 4H respectively were wound uniformly onto a non-magnetic core, it was found that their mutual inductance was 1.5H. Calculate the coupling coefficient that exists between.

$$k = \frac{M}{\sqrt{L_1 L_2}} = \frac{1.5}{\sqrt{5\times 4}} = 0.335 = 33.5\%$$

This SPICE simulation plots circuit current over a frequency range of 100 to 200 Hz in twenty even steps (100 and 200 Hz inclusive). Current magnitude on the graph increases from

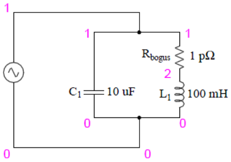

Figure 6.12: Resonant circuit sutitable for SPICE simulation.

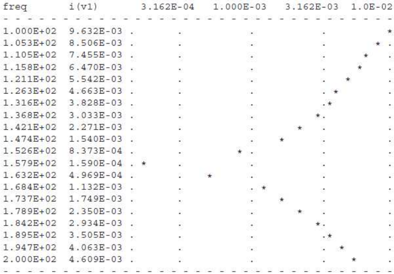

```
freq       i(v1)      3.162E-04   1.000E-03   3.162E-03  1.0E-02
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
1.000E+02  9.632E-03 .       .           .           .          *
1.053E+02  8.506E-03 .       .           .           .        * .
1.105E+02  7.455E-03 .       .           .           .      *   .
1.158E+02  6.470E-03 .       .           .           .     *    .
1.211E+02  5.542E-03 .       .           .           .   *      .
1.263E+02  4.663E-03 .       .           .           . *        .
1.316E+02  3.828E-03 .       .           .           .*         .
1.368E+02  3.033E-03 .       .           .          *.          .
1.421E+02  2.271E-03 .       .           .       *   .          .
1.474E+02  1.540E-03 .       .           .    *      .          .
1.526E+02  8.373E-04 .       .         * .           .          .
1.579E+02  1.590E-04 . *     .           .           .          .
1.632E+02  4.969E-04 .       .    *      .           .          .
1.684E+02  1.132E-03 .       .           . *         .          .
1.737E+02  1.749E-03 .       .           .    *      .          .
1.789E+02  2.350E-03 .       .           .       *   .          .
1.842E+02  2.934E-03 .       .           .          *.          .
1.895E+02  3.505E-03 .       .           .           .*         .
1.947E+02  4.063E-03 .       .           .           .  *       .
2.000E+02  4.609E-03 .       .           .           .    *     .
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

```
tank circuit frequency sweep
```

```
v1 1 0 ac 1 sin
c1 1 0 10u
*  rbogus  is  necessary  to  eliminate  a
direct loop
*  between  v1  and  l1,  which  SPICE  can’t
handle
rbogus 1 2 1e-12
l1 2 0 100m
.ac lin 20 100 200
.plot ac i(v1)
.end
```

left to right, while frequency increases from top to bottom. The current in this circuit takes a sharp dip around the analysis point of 157.9 Hz, which is the closest analysis point to our predicted resonance frequency of 159.155 Hz. It is at this point that total current from the power source falls to zero.

The plot above is produced fromthe above spice circuit file (*.cir), the command (.plot) in the last line producing the text plot on any printer or terminal. A better looking plot is produced by the “nutmeg” graphical post-processor, part of the spice package. The above spice (*.cir) does not require the plot (.plot) command, though it does no harm. The following commands produce the plot below: (Figure 6.13)

```
spice -b -r resonant.raw resonant.cir
(  -b  batch  mode,  -r  raw  file,  input  is
resonant.cir)
nutmeg resonant.raw
```

From the nutmeg prompt:

```
>setplot ac1 (setplot {enter} for list of
plots)
>display (for list of signals)
>plot mag(v1#branch)
```

However, the above equation assumes zero flux leakage and 100% magnetic coupling between the two coils, L 1 and L 2. In reality there will always be some loss due to leakage and position, so the magnetic coupling between the two coils can never reach or exceed 100%, but can become very close to this value in some special inductive coils.

If some of the total magnetic flux links with the two coils, this amount of flux linkage can be defined as a fraction of the total possible flux linkage between the coils. This fractional value is called the coefficient of coupling and is given the letter k.

Coupling Coefficient

Generally, the amount of inductive coupling that exists between the two coils is expressed as a fractional number between 0 and 1 instead of a percentage (%) value, where 0 indicates zero or no inductive coupling, and 1 indicating full or maximum inductive coupling.

In other words, if k = 1 the two coils are perfectly coupled, if k > 0.5 the two coils are said to be tightly coupled and if k < 0.5 the two coils are said to be loosely coupled. Then the equation above which assumes a perfect coupling can be modified to take into account this coefficient of coupling, k and is given as:

Coupling Factor Between Coils

$$k = \frac{M}{\sqrt{L_1 L_2}} \quad \text{or} \quad M = k\sqrt{L_1 L_2}$$

When the coefficient of coupling, k is equal to 1, (unity) such that all the lines of flux of one coil cuts all of the turns of the second coil, that is the two coils are tightly coupled together,

$$M_{12} = \frac{N_2 \Phi_{12}}{I_1}$$

Likewise, the flux linking coil one, L1 when a current flows around coil two, L2 is exactly the same as the flux linking coil two when the same current flows around coil one above, then the mutual inductance of coil one with respect of coil two is defined as M21. This mutual inductance is true irrespective of the size, number of turns, relative position or orientation of the two coils. Because of this, we can write the mutual inductance between the two coils as: M12 = M21 = M.

Then we can see that self inductance characterises an inductor as a single circuit element, while mutual inductance signifies some form of magnetic coupling between two inductors or coils, depending on their distance and arrangement, an hopefully we remember from our tutorials on Electromagnets that the self inductance of each individual coil is given as:

$$L_1 = \frac{\mu_o \mu_r N_1^2 A}{\ell} \text{ and } L_2 = \frac{\mu_o \mu_r N_2^2 A}{\ell}$$

By cross-multiplying the two equations above, the mutual inductance, M that exists between the two coils can be expressed in terms of the self inductance of each coil.

$$M^2 = L_1 L_2$$

giving us a final and more common expression for the mutual inductance between the two coils of:

$$M = \sqrt{L_1 L_2} \ \text{H}$$

Mutual Inductance Between Coils

```
(magnitude of complex current vector
v1#branch)
```

Incidentally, the graph output produced by this SPICE computer analysis is more generally known as a Bode plot. Such graphs plot amplitude or phase shift on one axis and frequency on the other. The steepness of a Bode plot curve characterizes a circuit's "frequency response," or how sensitive it is to changes in frequency.

• REVIEW:

• Resonance occurs when capacitive and inductive reactances are equal to each other.

• For a tank circuit with no resistance (R), resonant frequency can be calculated with the following formula:

$$f_{resonant} = \frac{1}{2\pi \sqrt{LC}}$$

• The total impedance of a parallel LC circuit approaches infinity as the power supply frequency approaches resonance.

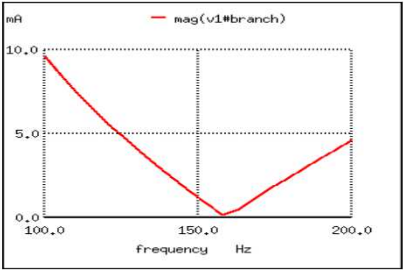

Figure 6.13: Nutmeg produces plot of current I(v1) for parallel resonant circuit.

- A Bode plot is a graph plotting waveform amplitude or phase on one axis and frequency on the other.

6.3 Simple series resonance

A similar effect happens in series inductive/capacitive circuits. (Figure 6.14) When a state of resonance is reached (capacitive and inductive reactances equal), the two impedances cancel each other out and the total impedance drops to zero!

---

wound one on top of the other over a common soft iron core unity coupling is said to exist between them as any losses due to the leakage of flux will be extremely small. Then assuming a perfect flux linkage between the two coils the mutual inductance that exists between them can be given as.

$$M = \frac{\mu_o \mu_r N_1 N_2 A}{\ell}$$

Where:

- μo is the permeability of free space (4.π.10-7)
- μr is the relative permeability of the soft iron core
- N is in the number of coil turns
- A is in the cross-sectional area in m2
- l is the coils length in meters

Mutual Induction

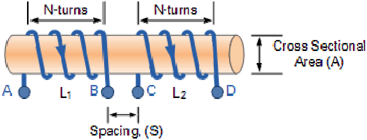

Here the current flowing in coil one, L1 sets up a magnetic field around itself with some of these magnetic field lines passing through coil two, L2 giving us mutual inductance. Coil one has a current of I1 and N1 turns while, coil two has N2 turns. Therefore, the mutual inductance, M12 of coil two that exists with respect to coil one depends on their position with respect to each other and is given as:

coil turns of the second coil inducing a relatively large emf and therefore producing a large mutual inductance value.

Likewise, if the two coils are farther apart from each other or at different angles, the amount of induced magnetic flux from the first coil into the second will be weaker producing a much smaller induced emf and therefore a much smaller mutual inductance value. So the effect of mutual inductance is very much dependant upon the relative positions or spacing, (S) of the two coils and this is demonstrated below.

Mutual Inductance between Coils

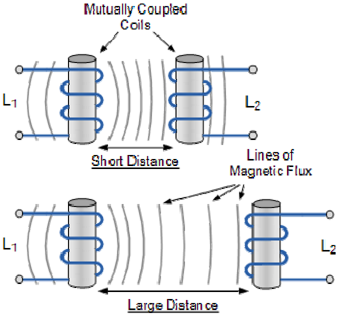

The mutual inductance that exists between the two coils can be greatly increased by positioning them on a common soft iron core or by increasing the number of turns of either coil as would be found in a transformer. If the two coils are tightly

---

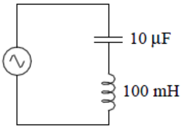

Figure 6.14: Simple series resonant circuit.

At 159.155 Hz:

$Z_L = 0 + j100\ \Omega$ $Z_C = 0 - j100\ \Omega$

$$Z_{series} = Z_L + Z_C$$

$$Z_{series} = (0 + j100\ \Omega) + (0 - j100\ \Omega)$$

$$Z_{series} = 0\ \Omega$$

With the total series impedance equal to 0 at the resonant frequency of 159.155 Hz, the result is a short circuit across the AC power source at resonance. In the circuit drawn above, this would not be good. I'll add a small resistor (Figure 6.15) in series along with the capacitor and the inductor to keep the maximum circuit current somewhat limited, and perform another SPICE analysis over the same range of frequencies: (Figure 6.16)

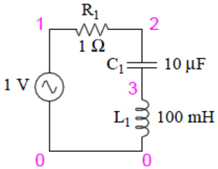

Figure 6.15: Series resonant circuit suitable for SPICE.

```
series lc circuit
v1 1 0 ac 1 sin
r1 1 2 1
c1 2 3 10u
l1 3 0 100m
.ac lin 20 100 200
.plot ac i(v1)
.end
```

As before, circuit current amplitude increases from bottom to top, while frequency increases from left to right. (Figure 6.16) The peak is still seen to be at the plotted frequency point of 157.9 Hz, the closest analyzed point to our predicted resonance point of 159.155 Hz. This would suggest that our resonant frequency formula holds as true for simple series LC circuits as it does for simple parallel LC circuits, which is the case:

Mutual Inductance of Two Coils

Wayne Storr
http://www.electronics-tutorials.ws/inductor/mutual-inductance.html

Mutual Inductance of Two Coils

In the previous tutorial we saw that an inductor generates an induced emf within itself as a result of the changing magnetic field around its own turns, and when this emf is induced in the same circuit in which the current is changing this effect is called Self-induction, (L). However, when the emf is induced into an adjacent coil situated within the same magnetic field, the emf is said to be induced magnetically, inductively or by Mutual induction, symbol (M). Then when two or more coils are magnetically linked together by a common magnetic flux they are said to have the property of Mutual Inductance.

Mutual Inductance is the basic operating principal of transformers, motors, generators and any other electrical component that interacts with another magnetic field. Then we can define mutual induction as the current flowing in one coil induces an emf in an adjacent coil. But mutual inductance can also be a bad thing as "stray" or "leakage" inductance from a coil can interfere with the operation of another adjacent component by means of electromagnetic induction, so some form of electrical screening to a ground potential may be required.

The amount of mutual inductance that links one coil to another depends very much on the relative positioning of the two coils. If one coil is positioned next to the other coil so that their physical distance apart is small, then nearly nearly all of the magnetic flux generated by the first coil will interact with the a frequency point where their two reactive components cancel each other out influencing the characteristics of the circuit. Both circuits have a resonant frequency point.

The difference this time however, is that a parallel resonance circuit is influenced by the currents flowing through each parallel branch within the parallel LC tank circuit. A tank circuit is a parallel combination of L and C that is used in filter networks to either select or reject AC frequencies. Consider the parallel RLC circuit below.

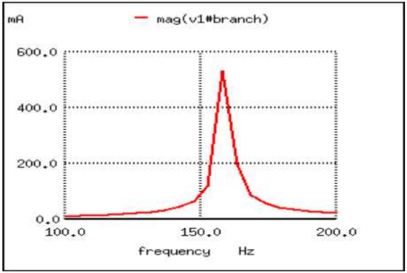

Figure 6.16: Series resonant circuit plot of current I(v1).

$$f_{resonant} = \frac{1}{2\pi \sqrt{LC}}$$

A word of caution is in order with series LC resonant circuits: because of the high currents which may be present in a series LC circuit at resonance, it is possible to produce dangerously high voltage drops across the capacitor and the inductor, as each component possesses significant impedance. We can edit the SPICE netlist in the above example to include a plot of voltage across the capacitor and inductor to demonstrate what happens: (Figure 6.17)

```
series lc circuit
v1 1 0 ac 1 sin
r1 1 2 1
```

```
c1 2 3 10u
l1 3 0 100m
.ac lin 20 100 200
.plot ac i(v1) v(2,3) v(3)
.end
```

According to SPICE, voltage across the capacitor and inductor reach a peak somewhere around 70 volts! This is quite impressive for a power supply that only generates 1 volt. Needless to say, caution is in order when experimenting with circuits such as this. This SPICE voltage is lower than the expected value due to the small (20) number of steps in the AC analysis statement (.ac lin 20 100 200). What is the expected value?

Given: $f_r = 159.155$ Hz, L = 100mH, R = 1

$X_L = 2\pi fL = 2\pi(159.155)(100mH) = j100\Omega$

$X_C = 1/(2\pi fC) = 1/(2\pi(159.155)(10\mu F)) = -j100\Omega$

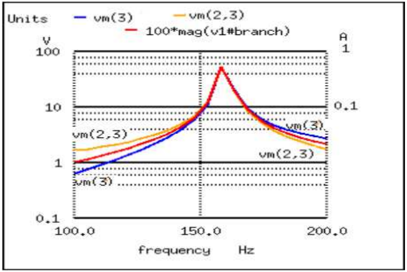

G + jB with the duality between the two complex impedances being defined as:

Series Circuit	Parallel Circuit
Voltage, (V)	Current, (I)
Resistance, (R)	Conductance, (G)
Reactance, (X)	Susceptance, (B)
Impedance, (Z)	Admittance, (Y)

As susceptance is the reciprocal of reactance, in an inductive circuit, inductive susceptance, BL will be negative in value and in a capacitive circuit, capacitive susceptance, BC will be positive in value. The exact opposite to XL and XC respectively.

We have seen so far that series and parallel RLC circuits contain both capacitive reactance and inductive reactance within the same circuit. If we vary the frequency across these circuits there must become a point where the capacitive reactance value equals that of the inductive reactance and therefore, XC = XL. The frequency point at which this occurs is called resonance and in the next tutorial we will look at series resonance and how its presence alters the characteristics of the circuit.

Parallel Resonance

The Parallel Resonance Circuit

In many ways a parallel resonance circuit is exactly the same as the series resonance circuit we looked at in the previous tutorial. Both are 3-element networks that contain two reactive components making them a second-order circuit, both are influenced by variations in the supply frequency and both have

Current and Admittance Triangles

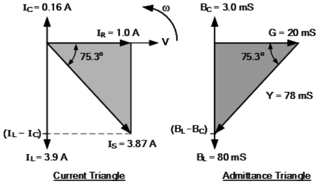

Parallel RLC Circuit Summary

In a parallel RLC circuit containing a resistor, an inductor and a capacitor the circuit current IS is the phasor sum made up of three components, IR, IL and IC with the supply voltage common to all three. Since the supply voltage is common to all three components it is used as the horizontal reference when constructing a current triangle.
Parallel RLC networks can be analysed using vector diagrams just the same as with series RLC circuits. However, the analysis of parallel RLC circuits is a little more mathematically difficult than for series RLC circuits when it contains two or more current branches. So an AC parallel circuit can be easily analysed using the reciprocal of impedance called Admittance.

Admittance is the reciprocal of impedance given the symbol, Y. Like impedance, it is a complex quantity consisting of a real part and an imaginary part. The real part is the reciprocal of resistance and is called Conductance, symbol Y while the imaginary part is the reciprocal of reactance and is called Susceptance, symbol B and expressed in complex form as: Y =

Figure 6.17: Plot of Vc=V(2,3) 70 V peak, VL=v(3) 70 V peak, I=I(V1#branch) 0.532 A peak

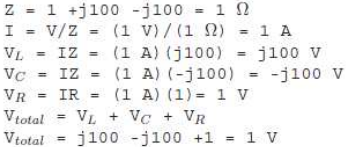

$$\begin{aligned}
Z &= 1 + j100 - j100 = 1\ \Omega \\
I &= V/Z = (1\ V)/(1\ \Omega) = 1\ A \\
V_L &= IZ = (1\ A)(j100) = j100\ V \\
V_C &= IZ = (1\ A)(-j100) = -j100\ V \\
V_R &= IR = (1\ A)(1) = 1\ V \\
V_{total} &= V_L + V_C + V_R \\
V_{total} &= j100 - j100 + 1 = 1\ V
\end{aligned}$$

The expected values for capacitor and inductor voltage are 100 V. This voltage will stress these components to that level and they must be rated accordingly. However, these voltages are out of phase and cancel yielding a total voltage across all three components of only 1 V, the applied voltage. The ratio of the capacitor (or inductor) voltage to the applied voltage is the “Q” factor.

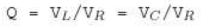

$$Q = V_L/V_R = V_C/V_R$$

- REVIEW:

- The total impedance of a series LC circuit approaches zero as the power supply frequency approaches resonance.

- The same formula for determining resonant frequency in a simple tank circuit applies to simple series circuits as well.

- Extremely high voltages can be formed across the individual components of series LC circuits at resonance, due to high current flows and substantial individual component impedances.

6.4 Applications of resonance

So far, the phenomenon of resonance appears to be a useless curiosity, or at most a nuisance to be avoided (especially if series resonance makes for a short-circuit across our AC voltage source!). However, this is not the case. Resonance is a very valuable property of reactive AC circuits, employed in a variety of applications. One use for resonance is to establish a condition of stable frequency in circuits designed to produce AC signals. Usually, a parallel (tank) circuit is used for this purpose, with the capacitor and inductor directly connected together, exchanging energy between each other. Just as a pendulum can be used to stabilize the frequency of a clock mechanism's oscillations, so can a tank circuit be used to stabilize the electrical frequency of an AC oscillator circuit. As was noted before, the frequency set by the tank circuit is solely dependent upon the values of L and C, and not on the magnitudes of voltage or current present in the oscillations: (Figure 6.18)

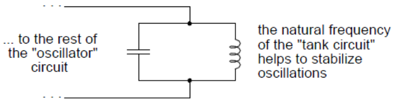
... to the rest of the "oscillator" circuit
the natural frequency of the "tank circuit" helps to stabilize oscillations

Figure 6.18: Resonant circuit serves as stable frequency source.

Another use for resonance is in applications where the effects of greatly increased or decreased impedance at a particular frequency is desired. A resonant circuit can be used to "block"

7). Total supply current, (IS):

$$I_S = \sqrt{I_R^2 + (I_L - I_C)^2} = \sqrt{1^2 + (3.9 - 0.16)^2} = 3.87\,(A)$$

8). Conductance, (G):

$$G = \frac{1}{R} = \frac{1}{50} = 0.02S \text{ or } 20mS$$

9). Inductive Susceptance, (BL):

$$B_L = \frac{1}{X_L} = \frac{1}{12.6} = 0.08S \text{ or } 80mS$$

10). Capacitive Susceptance, (BC):

$$B_C = \frac{1}{X_C} = \frac{1}{318.3} = 0.003S \text{ or } 3mS$$

11). Admittance, (Y):

$$Y = \frac{1}{Z} = \frac{1}{12.7} = 0.078S \text{ or } 78mS$$

12). Phase Angle, (φ) between the resultant current and the supply voltage:

$$\cos\phi = \frac{G}{Y} = \frac{20mS}{78mS} = 0.256$$

$$\phi = \cos^{-1}0.256 = 75.3^{o} \;\; (\text{lag})$$

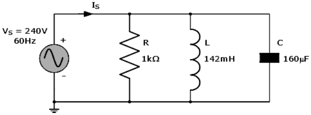

Is
Vs = 240V
60Hz
+
-
R
1kΩ
L
142mH
C
160µF

1). Inductive Reactance, (XL):

$$X_L = \omega L = 2\pi f L = 2\pi.100.0.02 = 12.6\Omega$$

2). Capacitive Reactance, (XC):

$$X_C = \frac{1}{\omega C} = \frac{1}{2\pi f C} = \frac{1}{2\pi.100.5\times 10^{-6}} = 318.3\Omega$$

3). Impedance, (Z):

$$Z = \frac{1}{\sqrt{\left(\frac{1}{R}\right)^2 + \left(\frac{1}{X_C} - \frac{1}{X_L}\right)^2}} = \frac{1}{\sqrt{\left(\frac{1}{50}\right)^2 + \left(\frac{1}{318.3} - \frac{1}{12.6}\right)^2}}$$

$$Z = \frac{1}{\sqrt{0.0004 + 0.0058}} = \frac{1}{0.0788} = 12.7\Omega$$

4). Current through resistance, R (IR):

$$I_R = \frac{V}{R} = \frac{50}{50} = 1.0\,(A)$$

5). Current through inductor, L (IL):

$$I_L = \frac{V}{X_L} = \frac{50}{12.6} = 3.9\,(A)$$

6). Current through capacitor, C (IC):

$$I_C = \frac{V}{X_C} = \frac{50}{318.3} = 0.16\,(A)$$

(present high impedance toward) a frequency or range of frequencies, thus acting as a sort of frequency "filter" to strain certain frequencies out of a mix of others. In fact, these particular circuits are called filters, and their design constitutes a discipline of study all by itself: (Figure 6.19)

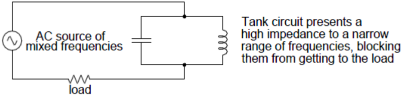

AC source of
mixed frequencies
Tank circuit presents a
high impedance to a narrow
range of frequencies, blocking
them from getting to the load
load

Figure 6.19: Resonant circuit serves as filter.

In essence, this is how analog radio receiver tuner circuits work to filter, or select, one station frequency out of the mix of different radio station frequency signals intercepted by the antenna.

• REVIEW:

• Resonance can be employed to maintain AC circuit oscillations at a constant frequency, just as a pendulum can be used to maintain constant oscillation speed in a timekeeping mechanism.

• Resonance can be exploited for its impedance properties: either dramatically increasing or decreasing impedance for certain frequencies. Circuits designed to screen certain frequencies out of a mix of different frequencies are called filters.

6.5 Resonance in series-parallel circuits

In simple reactive circuits with little or no resistance, the effects of radically altered impedance will manifest at the resonance frequency predicted by the equation given earlier. In a parallel (tank) LC circuit, this means infinite impedance at resonance. In a series LC circuit, it means zero impedance at resonance:

However, as soon as significant levels of resistance are introduced into most LC circuits, this simple calculation for resonance becomes invalid. We'll take a look at several LC circuits with added resistance, using the same values for capacitance and inductance as before: 10 μF and 100 mH, respectively. According to our simple equation, the resonant frequency should be 159.155 Hz. Watch, though, where current reaches maximum or minimum in the following SPICE analyses:

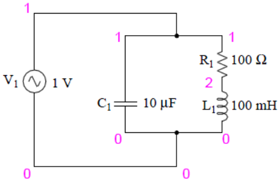

Figure 6.20: Parallel LC circuit with resistance in series with L.

Here, an extra resistor (Rbogus) (Figure 6.22)is necessary to prevent SPICE from encountering trouble in analysis. SPICE

Admittance: $$Y = \sqrt{\left(\frac{1}{R}\right)^2 + \left(\frac{1}{(\omega L)} - (\omega C)\right)^2}$$

Impedance: $$Z = \frac{1}{Y} = \frac{1}{\sqrt{\left(\frac{1}{R}\right)^2 + \left(\frac{1}{(\omega L)} - (\omega C)\right)^2}}$$

Giving us a power factor angle of:

$$\cos\phi = \frac{G}{Y} \qquad \phi = \cos^{-1}\left(\frac{G}{Y}\right)$$

or

$$\tan\phi = \frac{B}{G} \qquad \phi = \tan^{-1}\left(\frac{B}{G}\right)$$

As the admittance, Y of a parallel RLC circuit is a complex quantity, the admittance corresponding to the general form of impedance Z = R + jX for series circuits will be written as Y = G - jB for parallel circuits where the real part G is the conductance and the imaginary part jB is the susceptance. In polar form this will be given as:

Example No1

A 50Ω resistor, a 20mH coil and a 5uF capacitor are all connected in parallel across a 50V, 100Hz supply. Calculate the total current drawn from the supply, the current for each branch, the total impedance of the circuit and the phase angle. Also construct the current and admittance triangles representing the circuit.

Parallel RLC Circuit

susceptance, B. This makes it possible to construct an admittance triangle that has a horizontal conductance axis, G and a vertical susceptance axis, jB as shown.

Admittance Triangle for a Parallel RLC Circuit

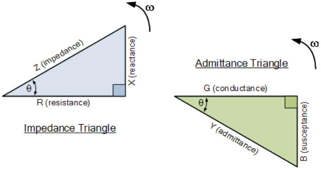
ω
Z (impedance)
X (reactance)
θ
R (resistance)
Impedance Triangle
ω
Admittance Triangle
G (conductance)
θ
Y (admittance)
B (susceptance)

Now that we have an admittance triangle, we can use Pythagoras to calculate the magnitudes of all three sides as well as the phase angle as shown.

from Pythagoras,

$$Y = \sqrt{G^2 + \left(B_L - B_C\right)^2}$$

where: $Y = \frac{1}{Z}$ $G = \frac{1}{R}$

$$B_L = \frac{1}{\omega L} \qquad B_C = \omega C$$

Then we can define both the admittance of the circuit and the impedance with respect to admittance as:

---

can’t handle an inductor connected directly in parallel with any voltage source or any other inductor, so the addition of a series resistor is necessary to “break

```
resonant circuit
v1 1 0 ac 1 sin
c1 1 0 10u
r1 1 2 100
l1 2 0 100m
.ac lin 20 100 200
.plot ac i(v1)
.end
```

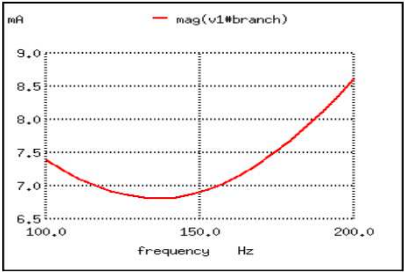

Figure 6.21: Resistance in series with L produces minimum current at 136.8 Hz instead of calculated 159.2 Hz

Minimum current at 136.8 Hz instead of 159.2 Hz!

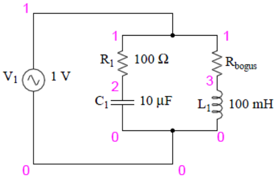

Figure 6.22: Parallel LC with resistance in serieis with C.

up" the voltage source/inductor loop that would otherwise be formed. This resistor is chosen to be a very low value for minimum impact on the circuit's behavior.

```
resonant circuit
v1 1 0 ac 1 sin
r1 1 2 100
c1 2 0 10u
rbogus 1 3 1e-12
l1 3 0 100m
.ac lin 20 100 400
.plot ac i(v1)
.end
```

Minimum current at roughly 180 Hz instead of 159.2 Hz!

difference between the voltage and the current. The admittance of a parallel circuit is the ratio of phasor current to phasor voltage with the angle of the admittance being the negative to that of impedance.

Conductance (G) :

Conductance is the reciprocal of resistance, R and is given the symbol G.

$$G = \frac{1}{R}\ [S]$$

Conductance is defined as the ease at which a resistor (or a set of resistors) allows current to flow when a voltage, either AC or DC is applied.

Susceptance (B) :

Susceptance is the reciprocal of reactance, X and is given the symbol B.

$$B_L = \frac{1}{X_L}\ [S]$$

$$B_C = \frac{1}{X_C}\ [S]$$

In AC circuits susceptance is defined as the ease at which a reactance (or a set of reactances) allows current to flow when a voltage is applied. Susceptance has the opposite sign to reactance so capacitive susceptance BC is positive, +ve in value and inductive susceptance BL is negative, -ve in value.

In AC series circuits the opposition to current flow is impedance, Z which has two components, resistance R and reactance, X and from these two coponents we can construct an impedance triangle. Similarly, in a parallel RLC circuit, admittance, Y also has two components, conductance, G and

$$\frac{1}{Z_T} = Y_T = Y_1 + Y_2 + Y_3 + Y_4 + \ldots\ldots\text{etc}$$

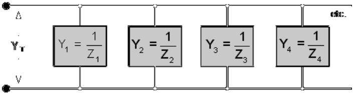

The new unit for admittance is the Siemens, abbreviated as S, (old unit mho's ℧, ohm's in reverse). Admittances are added together in parallel branches, whereas impedances are added together in series branches. But if we can have a reciprocal of impedance, we can also have a reciprocal of resistance and reactance as impedance consists of two components, R and X. Then the reciprocal of resistance is called Conductance and the reciprocal of reactance is called Susceptance.

Conductance, Admittance and Susceptance

The units used for conductance, admittance and susceptance are all the same namely Siemens (S), which can also be thought of as the reciprocal of Ohms or ohm-1, but the symbol used for each element is different and in a pure component this is given as:

Admittance (Y) :

Admittance is the reciprocal of impedance, Z and is given the symbol Y.

$$Y = \frac{1}{Z}\ [S]$$

In AC circuits admittance is defined as the ease at which a circuit composed of resistances and reactances allows current to flow when a voltage is applied taking into account the phase

---

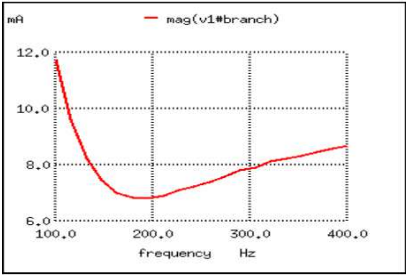

Figure 6.23: Resistance in series with C shifts minimum current from calculated 159.2 Hz to roughly 180 Hz.

Switching our attention to series LC circuits, (Figure 6.24) we experiment with placing significant resistances in parallel with either L or C. In the following series circuit examples, a 1 resistor (R1) is placed in series with the inductor and capacitor to limit total current at resonance. The “extra” resistance inserted to influence resonant frequency effects is the 100 resistor, R2. The results are shown in (Figure 6.25).

And finally, a series LC circuit with the significant resistance in parallel with the capacitor. (Figure 6.26) The shifted resonance is shown in (Figure 6.27)

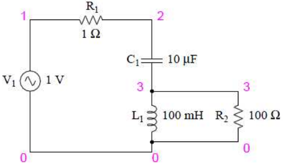

Figure 6.24: Series LC resonant circuit with resistance in parallel with L

```
resonant circuit
v1 1 0 ac 1 sin
r1 1 2 1
c1 2 3 10u
l1 3 0 100m
r2 3 0 100
.ac lin 20 100 400
.plot ac i(v1)
.end
```

Maximum current at roughly 178.9 Hz instead of 159.2 Hz!

Impedance of a Parallel RLC Circuit

$$R = \frac{V}{I_R} \quad X_L = \frac{V}{I_L} \quad X_C = \frac{V}{I_C}$$

$$Z = \frac{1}{\sqrt{\left(\frac{1}{R}\right)^2 + \left(\frac{1}{X_L} - \frac{1}{X_C}\right)^2}}$$

$$\therefore \frac{1}{Z} = \sqrt{\left(\frac{1}{R}\right)^2 + \left(\frac{1}{X_L} - \frac{1}{X_C}\right)^2}$$

You will notice that the final equation for a parallel RLC circuit produces complex impedances for each parallel branch as each element becomes the reciprocal of impedance, (1/Z) with the reciprocal of impedance being called Admittance. In parallel AC circuits it is more convenient to use admittance, symbol (Y) to solve complex branch impedances especially when two or more parallel branch impedances are involved (helps with the math's). The total admittance of the circuit can simply be found by the addition of the parallel admittances. Then the total impedance, ZT of the circuit will therefore be 1/YT Siemens as shown.

Admittance of a Parallel RLC Circuit

Since the voltage across the circuit is common to all three circuit elements, the current through each branch can be found using Kirchoff's Current Law, (KCL). Kirchoff's current law or junction law states that "the total current entering a junction or node is exactly equal to the current leaving that node", so the currents entering and leaving node "A" above are given as:

$$\text{KCL:}\ I_S - I_R - I_L - I_C = 0$$

$$I_S - \frac{V}{R} - \frac{1}{L}\int v\,dt - C\frac{dv}{dt} = 0$$

Taking the derivative, dividing through the above equation by C and rearranging gives us the following Second-order equation for the circuit current. It becomes a second-order equation because there are two reactive elements in the circuit, the inductor and the capacitor.

$$I_S - \frac{d^2V}{dt^2} - \frac{dV}{RCdt} - \frac{V}{LC} = 0$$

$$\therefore I_{S(t)} = \frac{d^2V}{dt^2} + \frac{dV}{dt}\frac{1}{RC} + \frac{1}{LC}V$$

The opposition to current flow in this type of AC circuit is made up of three components: XL XC and R and the combination of these three gives the circuit impedance, Z. We know from above that the voltage has the same amplitude and phase in all the components of a parallel RLC circuit. Then the impedance across each component can also be described mathematically according to the current flowing through, and the voltage across each element as.

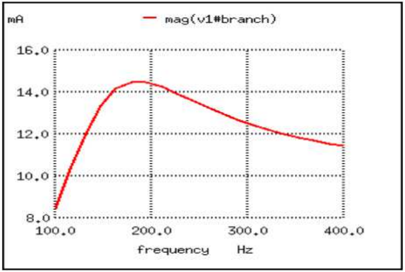

Figure 6.25: Series resonant circuit with resistance in parallel with L shifts maximum currentfrom 159.2 Hz to roughly 180 Hz.

```
resonant circuit
v1 1 0 ac 1 sin
r1 1 2 1
c1 2 3 10u
r2 2 3 100
l1 3 0 100m
.ac lin 20 100 200
.plot ac i(v1)
.end
```

Maximum current at 136.8 Hz instead of 159.2 Hz!

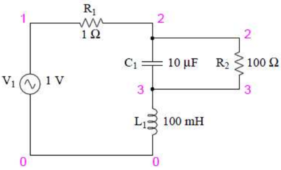

Figure 6.26: Series LC resonant circuit with rsistance in parallel with C.

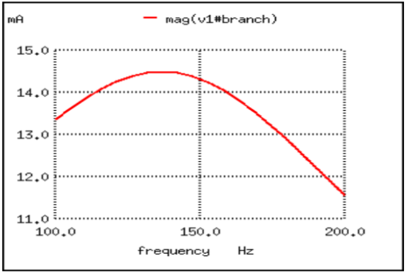

Figure 6.27: Resistance in parallel with C in series resonant circuit shifts curreent maximum from calculated 159.2 Hz to about 136.8 Hz

---

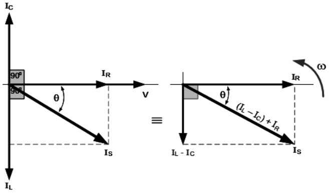

We can see from the phasor diagram on the right hand side above that the current vectors produce a rectangular triangle, comprising of hypotenuse IS, horizontal axis IR and vertical axis IL - IC Hopefully you will notice then, that this forms a Current Triangle and we can therefore use Pythagoras's theorem on this current triangle to mathematically obtain the magnitude of the branch currents along the x-axis and y-axis and then determine the total current IS of these components as shown.

Current Triangle for a Parallel RLC Circuit

$$I_S^2 = I_R^2 + \left(I_L - I_C\right)^2$$

$$I_S = \sqrt{I_R^2 + \left(I_L - I_C\right)^2}$$

$$\therefore\ I_S = \sqrt{\left(\frac{V}{R}\right)^2 + \left(\frac{V}{X_L} - \frac{V}{X_C}\right)^2} = \frac{V}{Z}$$

where: $I_R = \frac{V}{R},\quad I_L = \frac{V}{X_L},\quad I_C = \frac{V}{X_C}$

In the above parallel RLC circuit, we can see that the supply voltage, VS is common to all three components whilst the supply current IS consists of three parts. The current flowing through the resistor, IR, the current flowing through the inductor, IL and the current flowing through the capacitor, IC.

But the current flowing through each branch and therefore each component will be different to each other and to the supply current, IS. The total current drawn from the supply will not be the mathematical sum of the three individual branch currents but their vector sum.

Like the series RLC circuit, we can solve this circuit using the phasor or vector method but this time the vector diagram will have the voltage as its reference with the three current vectors plotted with respect to the voltage. The phasor diagram for a parallel RLC circuit is produced by combining together the three individual phasors for each component and adding the currents vectorially.

Since the voltage across the circuit is common to all three circuit elements we can use this as the reference vector with the three current vectors drawn relative to this at their corresponding angles. The resulting vector IS is obtained by adding together two of the vectors, IL and IC and then adding this sum to the remaining vector IR. The resulting angle obtained between V and IS will be the circuits phase angle as shown below.

Phasor Diagram for a Parallel RLC Circuit

The tendency for added resistance to skew the point at which impedance reaches a maximum or minimum in an LC circuit is called antiresonance. The astute observer will notice a pattern between the four SPICE examples given above, in terms of how resistance affects the resonant peak of a circuit:

- Parallel (“tank”) LC circuit:

- R in series with L: resonant frequency shifted down

- R in series with C: resonant frequency shifted up

- Series LC circuit:

- R in parallel with L: resonant frequency shifted up

- R in parallel with C: resonant frequency shifted down

Again, this illustrates the complementary nature of capacitors and inductors: how resistance in series with one creates an antiresonance effect equivalent to resistance in parallel with the other. If you look even closer to the four SPICE examples given, you’ll see that the frequencies are shifted by the same amount, and that the shape of the complementary graphs are mirror-images of each other!

Antiresonance is an effect that resonant circuit designers must be aware of. The equations for determining antiresonance “shift” are complex, and will not be covered in this brief lesson. It should suffice the beginning student of electronics to understand that the effect exists, and what its general tendencies are.

Added resistance in an LC circuit is no academic matter. While it is possible to manufacture capacitors with negligible unwanted resistances, inductors are typically plagued with substantial amounts of resistance due to the long lengths of wire used in their construction. What is more, the resistance of wire tends to increase as frequency goes up, due to a strange phenomenon known as the skin effect where AC current tends to be excluded from travel through the very center of a wire, thereby reducing the wire’s effective cross-sectional area. Thus, inductors not only have resistance, but changing, frequency-dependent resistance at that.

As if the resistance of an inductor’s wire weren’t enough to cause problems, we also have to contend with the “core losses” of iron-core inductors, which manifest themselves as added resistance in the circuit. Since iron is a conductor of electricity as well as a conductor of magnetic flux, changing flux produced by alternating current through the coil will tend to induce electric currents in the core itself (eddy currents). This effect can be thought of as though the iron core of the transformer were a sort of secondary transformer coil powering a resistive load: the less-than-perfect conductivity of the iron metal. This effects can be minimized with laminated cores, good core design and high-grade materials, but never completely eliminated.

One notable exception to the rule of circuit resistance causing a resonant frequency shift is the case of series resistor-inductor-capacitor (“RLC”) circuits. So long as all components are connected in series with each other, the resonant frequency of the circuit will be unaffected by the resistance. (Figure 6.28) The resulting plot is shown in (Figure 6.29).

The Parallel RLC Circuit

© 2016 by AspenCore

http://www.electronics-tutorials.ws/accircuits/parallel-circuit.html

The Parallel Circuit

The Parallel RLC Circuit is the exact opposite to the series circuit we looked at in the previous tutorial although some of the previous concepts and equations still apply. However, the analysis of parallel RLC circuits can a little more mathematically difficult than for series RLC circuits so in this tutorial about parallel RLC circuits only pure components are assumed in this tutorial to keep things simple.

This time instead of the current being common to the circuit components, the applied voltage is now common to all so we need to find the individual branch currents through each element. The total impedance, Z of a parallel RLC circuit is calculated using the current of the circuit similar to that for a DC parallel circuit, the difference this time is that admittance is used instead of impedance. Consider the parallel RLC circuit below.

Parallel RLC Circuit

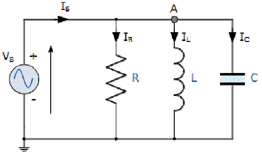

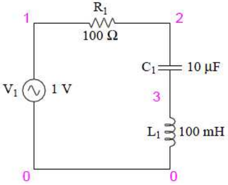
R1
1
2
100 Ω
C1
10 µF
V1
1 V
3
L1
100 mH
0
0

Figure 6.28: Series LC with resistance in series.

```
series rlc circuit
v1 1 0 ac 1 sin
r1 1 2 100
c1 2 3 10u
l1 3 0 100m
.ac lin 20 100 200
.plot ac i(v1)
.end
```

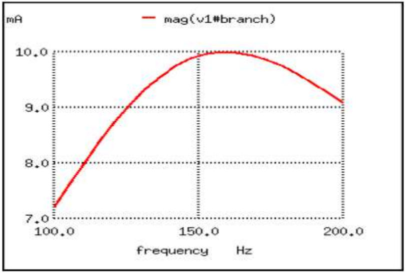

Figure 6.29: Resistance in series resonant circuit leaves current maximum at calculated 159.2 Hz, broadening the curve.

Maximum current at 159.2 Hz once again!

Note that the peak of the current graph (Figure 6.29) has not changed from the earlier series LC circuit (the one with the 1 token resistance in it), even though the resistance is now 100 times greater. The only thing that has changed is the "sharpness" of the curve. Obviously, this circuit does not resonate as strongly as one with less series resistance (it is said to be "less selective"), but at least it has the same natural frequency!

It is noteworthy that antiresonance has the effect of dampening the oscillations of freerunning LC circuits such as tank circuits. In the beginning of this chapter we saw how a capacitor and inductor connected directly together would act something like a pendulum, exchanging voltage and current peaks just like a

$$R_{eq} = \frac{\left(R_L R_C + \frac{L}{C}\right)\left(R_L + R_C\right) + \left(\omega L R_C - \frac{R_L}{\omega C}\right)\left(\omega L - \frac{1}{\omega C}\right)}{\left(R_L + R_C\right)^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}$$

$$X_{eq} = \frac{\left(\omega L R_C - \frac{R_L}{\omega C}\right)\left(R_L + R_C\right) - \left(R_L R_C + \frac{L}{C}\right)\left(\omega L - \frac{1}{\omega C}\right)}{\left(R_L + R_C\right)^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}$$

When the complex impedances of the branches of the parallel RLC circuit are combined, the equivalent impedance is of the form:

$$Z_{equiv} = \frac{Z_L Z_C}{Z_L + Z_C} = \frac{(R_L + j\omega L)\left(R_C - \frac{j}{\omega C}\right)}{(R_L + R_C) + j\left(\omega L - \frac{1}{\omega C}\right)}$$

When this expression is rationalized and put in the standard form:

$$Z_{eq} = R_{eq} + jX_{eq} = |Z|e^{j\phi}$$

then the impedance in ohms and the phase can be determined. By setting the = 0, the resonant frequency can be calculated. The expressions for these calculations are quite lengthy.

RLC Parallel Expressions

The complex impedances of the parallel RLC circuit takes the form:

$$Z_{equiv} = \frac{Z_L Z_C}{Z_L + Z_C} = \frac{(R_L + j\omega L)\left(R_C - \frac{j}{\omega C}\right)}{(R_L + R_C) + j\left(\omega L - \frac{1}{\omega C}\right)}$$

$$Z_{eq} = R_{eq} + jX_{eq} = |Z|e^{j\phi}$$

When rationalized, and the components have the form:

pendulum exchanges kinetic and potential energy. In a perfect tank circuit (no resistance), this oscillation would continue forever, just as a frictionless pendulum would continue to swing at its resonant frequency forever. But frictionless machines are difficult to find in the real world, and so are lossless tank circuits. Energy lost through resistance (or inductor core losses or radiated electromagnetic waves or . . .) in a tank circuit will cause the oscillations to decay in amplitude until they are no more. If enough energy losses are present in a tank circuit, it will fail to resonate at all.

Antiresonance's dampening effect is more than just a curiosity: it can be used quite effectively to eliminate unwanted oscillations in circuits containing stray inductances and/or capacitances, as almost all circuits do. Take note of the following L/R time delay circuit: (Figure 6.30)

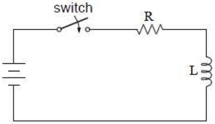

Figure 6.30: L/R time delay circuit

The idea of this circuit is simple: to "charge" the inductor when the switch is closed. The rate of inductor charging will be set by the ratio L/R, which is the time constant of the circuit in seconds. However, if you were to build such a circuit, you might find unexpected oscillations (AC) of voltage across the inductor when the switch is closed. (Figure 6.31)Why is this?

There’s no capacitor in the circuit, so how can we have resonant oscillation with just an inductor, resistor, and battery?

All inductors contain a certain amount of stray capacitance due to turn-to-turn and turn-to-core insulation gaps. Also, the placement of circuit conductors may create stray capacitance. While clean circuit layout is important in eliminating much of this stray capacitance, there will always be some that you cannot eliminate. If this causes resonant problems (unwanted AC oscillations), added resistance may be a way to combat it. If resistor R is large enough, it will cause a condition of antiresonance, dissipating enough energy to prohibit the inductance and stray capacitance from sustaining oscillations for very long.

Interestingly enough, the principle of employing resistance to eliminate unwanted resonance is one frequently used in the design of mechanical systems, where any moving object with mass is a potential resonator. A very common application of this is the use of shock absorbers in automobiles. Without shock absorbers, cars would bounce wildly at their resonant frequency after hitting any bump in the road. The shock absorber’s job is to introduce a strong antiresonant effect by dissipating energy hydraulically (in the same way that a resistor dissipates energy electrically).

RLC Parallel Circuit
HyperPhysics***** Electricity and Magnetism
http://hyperphysics.phy-astr.gsu.edu/hbase/electric/rlcpar.html

RLC Parallel Circuit

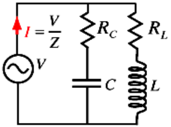

Finding the impedance of a parallel RLC circuit is considerably more difficult than finding the series RLC impedance. This is because each branch has a phase angle and they cannot be combined in a simple way. The impedance of the parallel branches combine in the same way that parallel resistors combine:

$$\frac{1}{Z_{equiv}} = \frac{1}{Z_L} + \frac{1}{Z_C} \quad \text{so} \quad Z_{equiv} = \frac{Z_L Z_C}{Z_L + Z_C}$$

But although the branch impedance magnitudes can be calculated from:

$$Z_L = \sqrt{R_L^2 + \omega^2 L^2} \quad \text{and} \quad Z_C = \sqrt{R_C^2 + \frac{1}{\omega^2 C^2}}$$

they cannot be directly combined as suggested by the expression above because they are different in phase - like vectors in different directions cannot be added directly. This dilemma is most easily solved by the complex impedance method.

RLC Parallel: Complex Impedance Method

$$I_{rms} = \frac{V_{rms}}{\sqrt{R^2+(X_L-X_C)^2}},$$

where I_{rms} and V_{rms} are rms current and voltage, respectively. The reactances vary with frequency ν, with X_L large at high frequencies and X_C large at low frequencies given as:

$$X_L = 2\pi\nu L, X_C = \frac{1}{2\pi\nu C}.$$

At some intermediate frequency ν_0, the reactances will be equal and cancel, giving Z=R —this is a minimum value for impedance, and a maximum value for I_{rms} results. We can get an expression for by taking X_L=X_C. Substituting the definitions of X_L and X_C yields:

$$\nu_0 = \frac{1}{2\pi\sqrt{LC}}.$$

ν_0 is the resonant frequency of an RLC series circuit. This is also the natural frequency at which the circuit would oscillate if not driven by the voltage source. At ν_0, the effects of the inductor and capacitor cancel, so that Z=R, and I_{rms} is a maximum. Resonance in AC circuits is analogous to mechanical resonance, where resonance is defined as a forced oscillation (in this case, forced by the voltage source) at the natural frequency of the system.

The receiver in a radio is an RLC circuit that oscillates best at its . A variable capacitor is often used to adjust the resonance frequency to receive a desired frequency and to reject others. is a graph of current as a function of frequency, illustrating a resonant peak in I_{rms} at $\nu_0 = f_0$. The two curves are for two different circuits, which differ only in the amount of resistance in them. The peak is lower and broader for the higher-resistance circuit. Thus higher-resistance circuits do not resonate as strongly, nor would they be as selective in, for example, a radio receiver.

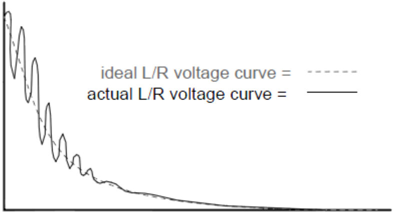

Figure 6.31: Inductor ringing due to resonance with stray capacitance.

- REVIEW:

- Added resistance to an LC circuit can cause a condition known as antiresonance, where the peak impedance effects happen at frequencies other than that which gives equal capacitive and inductive reactances.

- Resistance inherent in real-world inductors can contribute greatly to conditions of antiresonance. One source of such resistance is the skin effect, caused by the exclusion of AC current from the center of conductors. Another source is that of core losses in iron-core inductors.

- In a simple series LC circuit containing resistance (an “RLC” circuit), resistance does not produce antiresonance. Resonance still occurs when capacitive and inductive reactances are equal.

6.6 Q and bandwidth of a resonant circuit

The Q, quality factor, of a resonant circuit is a measure of the "goodness" or quality of a resonant circuit. A higher value for this figure of merit correspondes to a more narrow bandwith, which is desirable in many applications. More formally, Q is the ration of power stored to power dissipated in the circuit reactance and resistance, respectively:

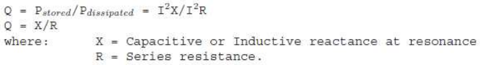

$Q = P_{stored}/P_{dissipated} = I^2X/I^2R$

$Q = X/R$

where: X = Capacitive or Inductive reactance at resonance
R = Series resistance.

This formula is applicable to series resonant circuits, and also parallel resonant circuits if the resistance is in series with the inductor. This is the case in practical applications, as we are mostly concerned with the resistance of the inductor limiting the Q. Note: Some text may show X and R interchanged in the "Q" formula for a parallel resonant circuit. This is correct for a large value of R in parallel with C and L. Our formula is correct for a small R in series with L

A practical application of "Q" is that voltage across L or C in a series resonant circuit is Q times total applied voltage. In a parallel resonant circuit, current through L or C is Q times the total applied current.

6.6.1 Series resonant circuits

A series resonant circuit looks like a resistance at the resonant frequency. (Figure 6.32) Since the definition of resonance is XL=XC, the reactive components cancel, leaving only the resistance to contribute to the impedance. The impedance is also at a minimum at resonance. (Figure 6.33) Below the resonant frequency, the series resonant circuit looks capacitive

---

Resonance in RLC Circuits

https://www.boundless.com/physics/textbooks/boundless-physics-textbook/induction-ac-circuits-and-electrical-technologies-22/ac-circuits-162/resonance-in-rlc-circuits-586-3035/

Resonance is the tendency of a system to oscillate with greater amplitude at some frequencies—in an RLC series circuit, it occurs at $\nu_0 = \frac{1}{2\pi\sqrt{LC}}$.

\

Resonance is the tendency of a system to oscillate with greater amplitude at some frequencies than at others. Frequencies at which the response amplitude is a relative maximum are known as the system's resonance frequencies. To study the resonance in an RLC circuit, as illustrated below, we can see how the circuit behaves as a function of the frequency of the driving voltage source.

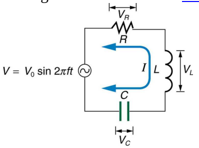

Combining Ohm's law, $I_{rms}=V_{rms}/Z$, and the expression for impedance Z from

$Z = \sqrt{R^2 + (X_L - X_C)^2}$ gives

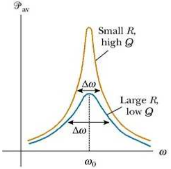

Figure 33.16 Average power versus frequency for a series *RLC* circuit. The width $\Delta\omega$ of each curve is measured between the two points where the power is half its maximum value. The power is a maximum at the resonance frequency ω_0.

The receiving circuit of a radio is an important application of a resonant circuit. One tunes the radio to a particular station (which transmits a specific electromagnetic wave or signal) by varying a capacitor, which changes the resonant frequency of the receiving circuit. When the resonance frequency of the circuit matches that of the incoming electromagnetic wave, the current in the receiving circuit increases. This signal caused by the incoming wave is then amplified and fed to a speaker. Because many signals are often present over a range of frequencies, it is important to design a high-Q circuit to eliminate unwanted signals. In this manner, stations whose frequencies are near but not equal to the resonance frequency give signals at the receiver that are negligibly small relative to the signal that matches the resonance frequency.

---

since the impedance of the capacitor increases to a value greater than the decreasing inducitve reactance, leaving a net capacitive value. Above resonance, the inductive rectance increases, capacitive reactance decreases, leaving a net inductive component.

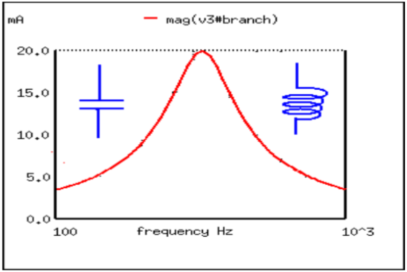

Figure 6.32: At resonance the series resonant circuit appears purely resistive. Below resonance it looks capacitive. Above resonance it appears inductive.

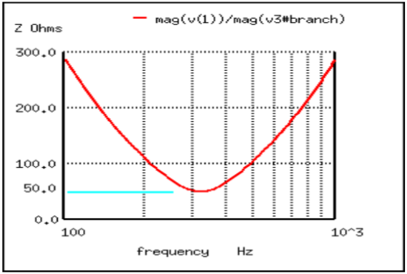

Figure 6.33: Impedance is at a minumum at resonance in a series resonant circuit.

Current is maximum at resonance, impedance at a minumum. Current is set by the value of the resistance. Above or below resonance, impedance increases.

The resonant current peak may be changed by varying the series resistor, which changes the Q. (Figure 6.34) This also affects the broadness of the curve. A low resistance, high Q circuit has a narrow bandwidth, as compared to a high resistance, low Q circuit. Bandwidth in terms of Q and resonant frequency:

Because XL = ωL, XC = 1 /ωC and ω02 = 1/LC, we can express the term (XL - XC)2 as

$$(X_L - X_C)^2 = \left(\omega L - \frac{1}{\omega C}\right)^2 = \frac{L^2}{\omega^2}\,(\omega^2 - {\omega_0}^2)^2 \qquad (33.34)$$

Using this result in Equation 33.34 gives

$$\mathcal{P}_{\mathrm{av}} = \frac{(\Delta V_{\mathrm{rms}})^2\, R\omega^2}{R^2\omega^2 + L^2(\omega^2 - {\omega_0}^2)^2} \qquad (33.35)$$

This expression shows that at resonance, when ω = ω0, the average power is a maximum and has the value (ΔVrms)2/R. Figure 33.15b is a plot of average power versus frequency for two values of R in a series RLC circuit. As the resistance is made smaller, the curve becomes sharper in the vicinity of the resonance frequency. This curve sharpness is usually described by a dimensionless parameter known as the quality factor, denoted by Q:

$$Q = \frac{\omega_0}{\Delta\omega}$$

where Δω is the width of the curve measured between the two values of ω for Pav which has half its maximum value, called the half-power points (see Fig. 33.15b.) It is left as a problem (Problem 70) to show that the width at the half-power points has the value Δω = R/L, so

$$Q = \frac{\omega_0 L}{R}$$

The curves plotted in Figure 33.16 show that a high-Q circuit responds to only a very narrow range of frequencies, whereas a low-Q circuit can detect a much broader range of frequencies. Typical values of Q in electronic circuits range from 10 to 100.

curves correspond to three values of R. Note that in each case the current reaches its maximum value at the resonance frequency ω0 . Furthermore, the curves become narrower and taller as the resistance decreases. By inspecting Equation 33.32, we must conclude that, when R = 0, the current becomes infinite at resonance. Although the equation predicts this, real circuits always have some resistance, which limits the value of the current.

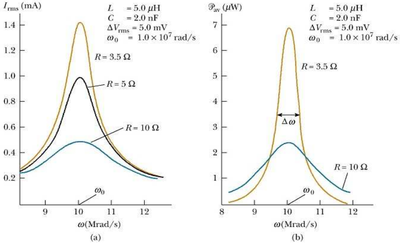

Figure 33.15 (a) The rms current versus frequency for a series RLC circuit, for three values of R. The current reaches its maximum value at the resonance frequency ω0 . (b) Average power versus frequency for the series RLC circuit, for two values of R.

It is also interesting to calculate the average power as a function of frequency for a series RLC circuit. Using Equations 33.30, 33.31, and 33.23, we find that

$$\mathcal{P}_{av} = I_{rms}^2 R = \frac{(\Delta V_{rms})^2}{Z^2} R = \frac{(\Delta V_{rms})^2 R}{R^2 + (X_L - X_C)^2}$$

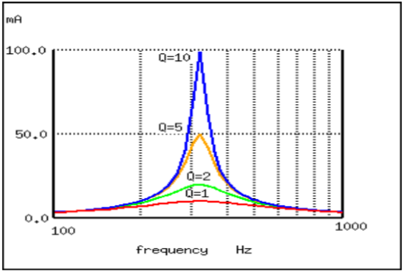

Figure 6.34: A high Q resonant circuit has a narrow bandwidth as compared to a low Q

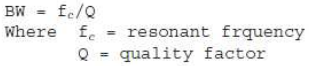
BW = f_c/Q
Where f_c = resonant frquency
Q = quality factor

Bandwidth is measured between the 0.707 current amplitude points. The 0.707 current points correspond to the half power points since P = I2R, (0.707)2 = (0.5). (Figure 6.35)

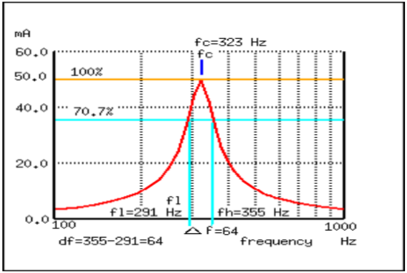

Figure 6.35: Bandwidth, · f ismeasured between the 70.7% amplitude points of series resonant circuit.

BW = Δf = $f_h - f_l$ = f_c/Q

Where f_h = high band edge, f_l = low band edge

f_l = f_c - $\Delta f/2$

f_h = f_c + $\Delta f/2$

Where f_c = center frequency (resonant frequency)

In Figure 6.35, the 100% current point is 50 mA. The 70.7% level is 0707(50 mA)=35.4 mA. The upper and lower band edges read from the curve are 291 Hz for fl and 355 Hz for fh. The bandwidth is 64 Hz, and the half power points are ± 32 Hz of the center resonant frequency:

Resonance in a series RLC circuit

© 2011-2016 Kshitij Education India

http://www.kshitij-iitjee.com/Resonance-in-a-series-RLC-circuit

A series RLC circuit is said to be in resonance when the current has its maximum value. In general, the rms current can be written

$$I_{\text{rms}} = \frac{\Delta V_{\text{rms}}}{Z} \qquad (33.31)$$

where Z is the impedance. Substituting the expression for Z from Equation 33.23 into 33.31 gives

$$I_{\text{rms}} = \frac{\Delta V_{\text{rms}}}{\sqrt{R^2 + (X_L - X_C)^2}} \qquad (33.32)$$

Because the impedance depends on the frequency of the source, the current in the RLC circuit also depends on the frequency. The frequency ω0 at which XL – XC = 0 is called the resonance frequency of the circuit. To find ω0 , we use the condition XL = XC , from which we obtain ω0L = 1/ω0C, or

$$\omega_0 = \frac{1}{\sqrt{LC}} \qquad (33.33)$$

Note that this frequency also corresponds to the natural frequency of oscillation of an LC circuit. Therefore, the current in a series RLC circuit reaches its maximum value when the frequency of the applied voltage matches the natural oscillator frequency—which depends only on L and C. Furthermore, at this frequency the current is in phase with the applied voltage.

A plot of rms current versus frequency for a series RLC circuit is shown in Figure 33.15a. The data assume a constant ΔVrms = 5.0 mV, that L = 5.0 μH, and that C = 2.0 nF. The three

$$BW = \Delta f = f_h - f_l = 355-291 = 64$$

$$f_l = f_c - \Delta f/2 = 323-32 = 291$$

$$f_h = f_c + \Delta f/2 = 323+32 = 355$$

Since BW = fc/Q:

$$Q = f_c/BW = (323\ Hz)/(64\ Hz) = 5$$

6.6.2 Parallel resonant circuits

A parallel resonant circuit is resistive at the resonant frequency. (Figure 6.36) At resonance XL=XC, the reactive components cancel. The impedance is maximum at resonance. (Fig6.6. ure 6.37) Below the resonant frequency, the parallel resonant circuit looks inductive since the impedance of the inductor is lower, drawing the larger proportion of current. Above resonance, the capacitive rectance decreases, drawing the larger current, thus, taking on a capacitive characteristic.

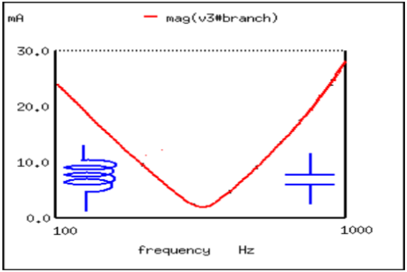

Figure 6.36: A parallel resonant circuit is resistive at resonance, inductive below resonance, capacitive above resonance.

Impedance is maximum at resonance in a parallel resonant circuit, but decreases above or below resonance. Voltage is at a peak at resonance since voltage is proportional to impedance (E=IZ). (Figure 6.37)

A low Q due to a high resistance in series with the inductor produces a low peak on a broad response curve for a parallel resonant circuit. (Figure 6.38) conversely, a high Q is due to a low resistance in series with the inductor. This produces a higher peak in the narrower response curve. The high Q is achieved by winding the inductor with larger diameter (smaller gague), lower resistance wire.

The bandwidth of the parallel resonant response curve is measured between the half power points. This corresponds to the 70.7% voltage points since power is proportional to E2. ((0.707)2=0.50) Since voltage is proportional to impedance, we may use the impedance curve. (Figure 6.39)

In Figure 6.39, the 100% impedance point is 500

. The 70.7% level is 0707(500)=354

. The upper and lower band edges read from the curve are 281 Hz for fl and 343 Hz for fh. The bandwidth is 62 Hz, and the half power points are ± 31 Hz of the center resonant frequency:

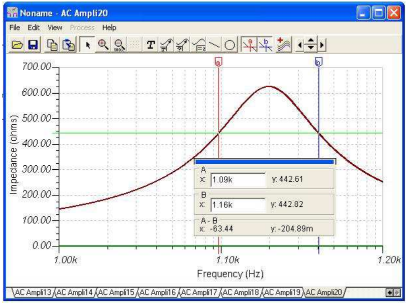
Noname - AC Ampli20
File Edit View Process Help
a
b
700.00
600.00
500.00
400.00
300.00
200.00
100.00
0.00
Impedance (ohms)
1.00k
1.10k
1.20k
Frequency (Hz)
AC Ampli13
AC Ampli14
AC Ampli15
AC Ampli16
AC Ampli17
AC Ampli18
AC Ampli19
AC Ampli20

A
x: 1.09k y: 442.61

B
x: 1.16k y: 442.82

A - B
x: -63.44 y: -204.89m

The difference of the A-B cursors is 63.44Hz, which is in very good agreement with the theoretical 63.8Hz result even taking the inaccuracy of the graphic procedure into consideration.

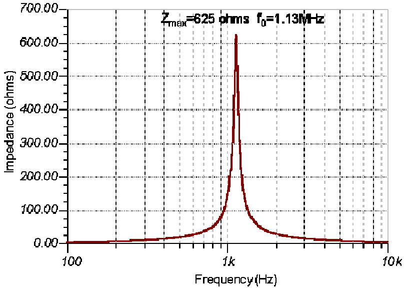

Finally let's examine the bandwidth of this circuit.

The calculated value:

$$\Delta f = \frac{f_0}{Q_0} = \frac{1130}{17.7} = 63.8Hz$$

Lets confirm it graphically using the diagram.

Zmax = 625 ohm. The impedance limits that define the cutoff frequencies are:

$$Z_1 = Z_2 = \frac{Z_{max}}{\sqrt{2}} = 442\Omega$$

$BW = \Delta f = f_h - f_l = 343-281 = 62$
$f_l = f_c - \Delta f/2 = 312-31 = 281$
$f_h = f_c + \Delta f/2 = 312+31 = 343$

$Q = f_c/BW = (312\ Hz)/(62\ Hz) = 5$

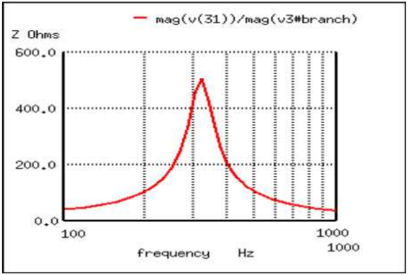

Figure 6.37: Parallel resonant circuit: Impedance peaks at resonance.

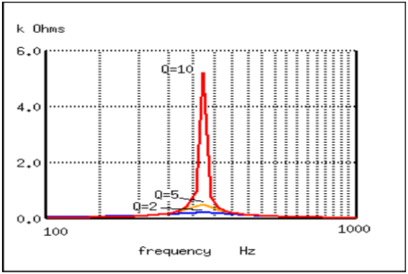

Figure 6.38: Parallel resonant response varies with Q.

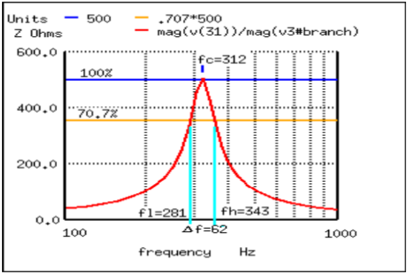

The equivalent parallel resistance: Req = Qo2 RL = 625 ohm

The equivalent parallel circuit:

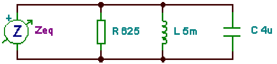

The impedance diagram:

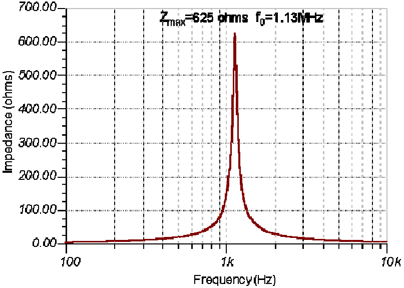

Finally, if we use copy and paste to see both curves on one diagram, we get the following picture where the two curves coincide.

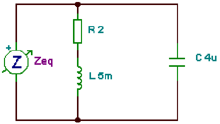

The resonant (Thomson) frequency:

$$\omega_o = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{5*4*10^{-9}}} = \frac{10^4}{\sqrt{2}}$$

The impedance diagram is the following:

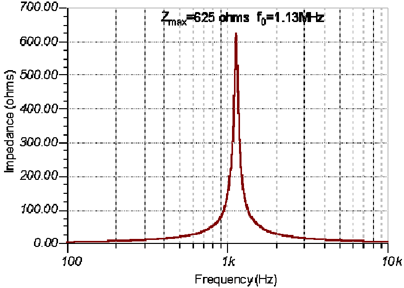

$$Q_o = \frac{\omega_o L}{R_L} = \frac{10^4}{\sqrt{2}} \frac{5*10^{-3}}{2} = \frac{50}{2*\sqrt{2}} = 12.5*\sqrt{2} = 17.7$$

Figure 6.39: Bandwidth, · f is measured between the 70.7% impedance points of a parallel resonant circuit.

6.7 Contributors

Contributors to this chapter are listed in chronological order of their contributions, from most recent to first. See Appendix 2 (Contributor List) for dates and contact information. Jason Starck (June 2000): HTML document formatting, which led to a much better looking second edition.

The equivalent impedance:

$$Z_{eq} = \frac{\frac{1}{j\omega C}(R + j\omega L)}{\frac{1}{j\omega C} + (R + j\omega L)} = \frac{R + j \ \omega \ L}{1 - \omega^2 LC + j\omega \ C \ R}$$

Let's examine this impedance at the resonant frequency where 1-w02LC=0

We will also assume that the quality factor Qo = woL/ RL>>1.

$$Z_{eq} = \frac{R_L + j\omega_o L}{j\omega_o CR} = R_L \frac{1 + jQ}{j\omega_o CR} \cong \frac{R\omega_o L}{\omega_o CR}$$

At the resonant frequency

Since at resonant frequencyw0L= 1/w0C

Zeq=Qo2 RL

Since in the pure parallel resonant circuit at the resonant frequency Zeq = R, the real parallel resonant circuit can be replaced by a pure parallel resonant circuit, where:

R = Qo2 RL

Example 3

Compare the impedance diagrams of a real parallel and its equivalent pure parallel resonance circuit.

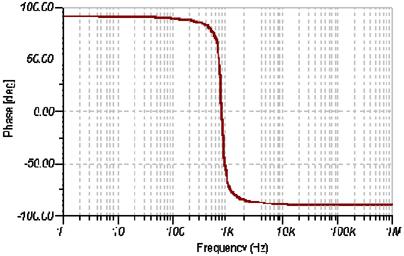

The "pure" parallel circuit above was very easy to examine since all components were in parallel. This is especially important when the circuit is connected to other parts.

However in this circuit, the series loss resistance of the coil was not considered.

Now let's examine the following so called "real parallel resonant circuit," with the series loss resistance of the coil present and learn how we can transform it into a "pure" parallel circuit.

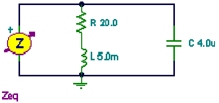

RESONANT CIRCUITS

Copyright © 2012 DesignSoft, Inc.

http://www.tina.com/English/tina/course/28resonant/resonant.htm

Circuits containing R, L, C elements often have special characteristics useful in many applications. Because their frequency characteristics (impedance, voltage, or current vs. frequency) may have a sharp maximum or minimum at certain frequencies these circuits are very important in the operation of television receivers, radio receivers, and transmitters. In this chapter we will present the different types, models and formulas of typical resonant circuits.

SERIES RESONANCE

A typical series resonant circuit is shown in the figure below.

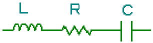

The total impedance:

$$Z = j\omega L + R + \frac{1}{j\omega C} = R + j\left(\omega L - \frac{1}{\omega C}\right)$$

In many cases, R represents the loss resistance of the inductor, which in the case of air core coils simply means the resistance of the winding. The resistances associated with the capacitor are often negligible.

The impedances of the capacitor and inductor are imaginary and have opposite sign. At the frequency w0 L= 1/w0C, the total imaginary part is zero and therefore the total impedance is R, having a minimum at the w0frequency. This frequency is called the series resonant frequency.

The typical impedance characteristic of the circuit is shown in the figure below.

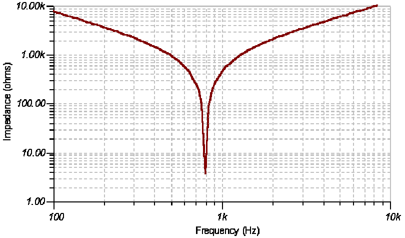

From the w0L= 1/w0Cequation, the angular frequency of the series resonance: or for the frequency in Hz:

$$T_0 = \frac{2\pi}{\omega_0} = 2\pi\sqrt{LC}.$$

f0 =

This is the so-called Thomson formula.

If R is small compared to the XL, XC reactance around the resonant frequency, the impedance changes sharply at the series resonant frequencyIn this case we say that the circuit has good selectivity.

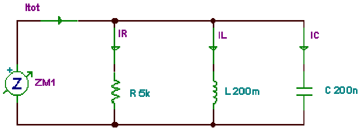

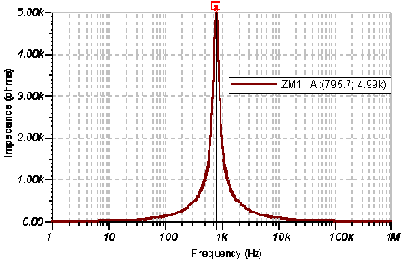

<

Find the resonant frequency and the resonant quality factor of a pure parallel resonance circuit where R = 5 kohm, L = 0.2 H, C = 200 nF.

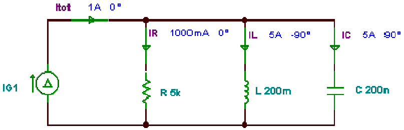

The resonant frequency:

$$\omega_0 = \sqrt{\frac{1}{LC}} = \sqrt{\frac{1}{0.2*0.2*10^{-6}}} = 5000\, {}^{rad}\!/_{s}$$

$$f_0 = \frac{\omega_0}{2\pi} = 795.8\,Hz$$

and the resonant quality factor:

$$Q = \frac{B_C}{G} = \frac{\omega C}{G} = \omega CR = R\sqrt{\frac{C}{L}} = 5*10^3 * \sqrt{\frac{0.2*10^{-6}}{0.2}} = 5$$

Incidentally, this quality factor is equal to IL /IR at the resonant frequency.

Now let us draw the impedance diagram of the circuit:

The simplest way is to replace the current source by an impedance meter and run an AC Transfer analysis.

The selectivity can be measured by the quality factor Q If the angular frequency in the formula equals the angular frequency of resonance, we get the resonant quality factor There is a more general definition of the quality factor:

$$Q = 2\pi \frac{\text{energy stored in inductor}}{\text{the energy turned to heat during one period}} = \frac{I^2 \omega L}{I^2 R} = \frac{\omega L}{R}$$

The voltage across the inductor or capacitor can be much higher then the voltage of the total circuit. At the resonant frequency the total impedance of the circuit is:

Z=R

Assuming that the current through the circuit is I, total voltage on the circuit is

Vtot=I*R

However the voltage on the inductor and the capacitor

$$V_L = V_C = I\omega_0 L = \frac{I}{\omega_0 C}$$

Therefore

$$\frac{V_L}{V_{tot}} = \frac{V_C}{V_{tot}} = Q_0$$

This means at the resonant frequency the voltages on the inductor and the capacitor are Q0 times greater than the total voltage of the resonant circuit.

The typical run of the VL, VC voltages is shown in the figure below.

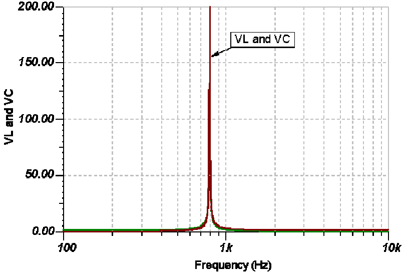

Let's demonstrate this via a concrete example.

Example 1

Find the frequency of resonance (f0) and the resonant quality factor (Q0) in the series circuit below, if C=200nF, L=0.2H, R=200 ohms, and R=5 ohms. Draw the phasor diagram and the frequency response of the voltages.

voltage and admittance changes sharply around the resonant frequency. In this case we say the circuit has good selectivity.

Selectivity can be measured by the quality factor Q

$$Q = \frac{B_C}{G} = \frac{\omega C}{G} = \omega CR$$

When the angular frequency equals the angular frequency of resonance, we get the resonant quality factor

$$Q_0 = \frac{B_C}{G} = \frac{\omega_0 C}{G} = \omega_0 CR = R\sqrt{\frac{C}{L}}$$

There is also a more general definition of the quality factor:

$$Q = 2\pi \ \frac{\text{energy stored in capacitance}}{\text{the energy turned to heat during one period}} = \frac{V^2 \omega_0 C}{V^2 G} = \omega_0 C$$

Another important property of the parallel resonant circuit is its bandwidth. The bandwidth is the difference between the two cutoff frequencies, where the impedance drops from its maximum value to $\frac{1}{\sqrt{2}} = 0.707 \quad (-3dB)\ of\ldots$ the maximum.

It can be shown that the Δf bandwidth is determined by the following simple formula:

$$\Delta f = \frac{f_0}{Q_0}$$

This formula is also applicable for series resonant circuits.

Let us demonstrate the theory through some examples.

Example 2

linear impedance axis is shown below. Note that viewed with this axis, the impedance appears to be changing even more rapidly near resonance.

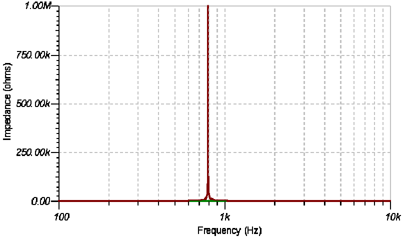

The susceptances of the inductance and capacitance are equal but of opposite sign at resonance: BL = BC, 1/w0L = w0C, hence the angular frequency of the parallel resonance:

$$\omega_0 = \sqrt{\frac{1}{LC}}$$

determined again by the Thomson formula.

Solving for the resonant frequency in Hz:

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{1}{LC}}$$

At this frequency the admittance Y = 1/R = G and is at its minimum (i.e., the impedance is maximum). The currents through the inductance and capacitance can be much higher then the current of the total circuit. If R is relatively large, the

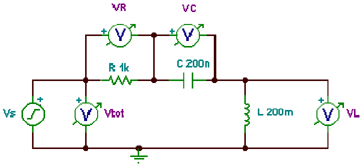

$$f_0 = \frac{\omega_0}{2\pi} = 795.8Hz$$

For R=200 ohms

This is a quite low value for practical resonant circuits, which normally have quality factors over 100. We have used a low value to more easily demonstrate the operation on a phasor diagram.

The current at the resonance frequency I=Vs/R=5m>

The voltages at current of 5mA: VR = Vs =1 V

meanwhile: VL = VC = I*w0L = 5*10-3 *5000*0.2 = 5V

The ratio between VL, VC,and Vs is equal to the quality factor!

Now let's see the phasor diagram by calling it from the AC Analysis menu of TINA.

We used the Auto Label tool of the diagram window to annotate the picture.

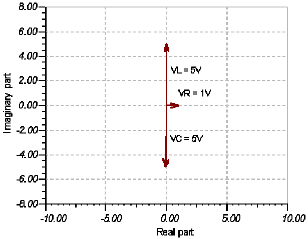

The phasor diagram nicely shows how the voltages of the capacitor and inductor cancel each other at the resonance frequency.

Now let's see VLand VCversus frequency.

see below the loss resistance of the inductor can be transformed into this resistor.

The total admittance:

$$Y=\frac{1}{R}+j\omega C+\frac{1}{j\omega L}=R+j(\omega C-\frac{1}{\omega L})$$

The admittances (called susceptances) of the capacitor and inductor are imaginary and have opposite sign. At the frequency w0C= 1/w0Lthe total imaginary part is zero, so the total admittance is 1/R-its minimum value and the total impedance has its maximum value. This frequency is called the parallel resonant frequency.

The total impedance characteristic of the pure parallel resonant circuit is shown in the figure below:

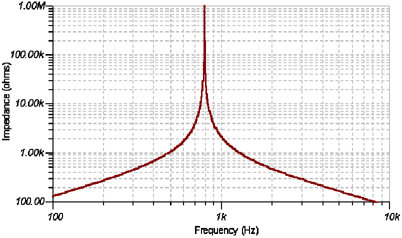

Note that the impedance changes very rapidly around the resonance frequency, even though we used a logarithmic impedance axis for better resolution. The same curve with a

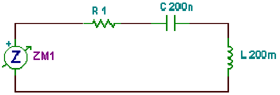

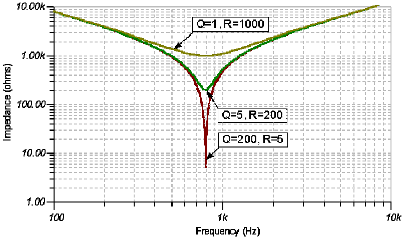

PARALLEL RESONANCE

The pure parallel resonant circuit is shown in the figure below.

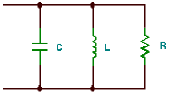

If we neglect the loss resistance of the inductor, R represents the leakage resistance of the capacitor. However, as we will

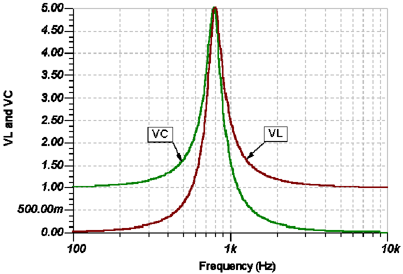

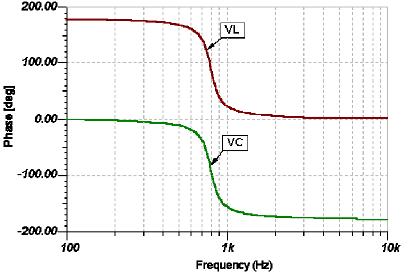

Note that VL starts from zero voltage (because its reactance is zero at zero frequency) while VC starts from 1 V (because its reactance is infinite at zero frequency). Similarly VL tends to 1V and VCto 0V at high frequencies.

Now for R=5 ohms the quality factor is much greater:

This is a relatively high quality factor, close to the practical achievable values.

The current at the resonance frequency I=Vs/R=0.2A

meanwhile: VL = VC = I*w0L = 0.2*5000*0.2 = 200

Again the ratio between the voltages equals the quality factor!

Now let's draw just VL and VC voltages versus frequency. On the phasor diagram, VR would be too small compared to VLand VC

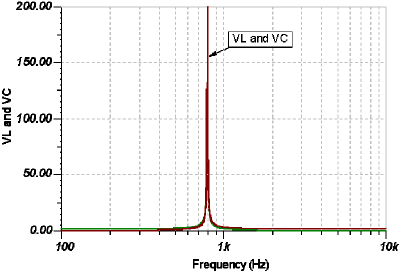

As we can see, the curve is very sharp and we needed to plot 10,000 points to get the maximum value accurately. Using a narrower bandwidth on the linear scale on the frequency axis, we get the more detailed curve below.

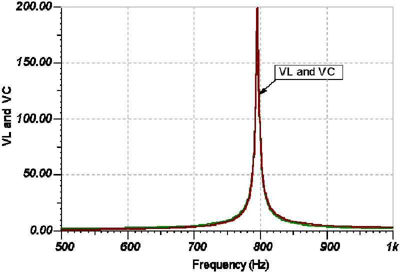

Finally let's see the impedance characteristic of the circuit: for different quality factors.

The figure below was created using TINA by replacing the voltage generator by an impedance meter. Also, set up a parameter stepping list for R = 5, 200, and 1000 ohms. To set up parameter stepping, select Control Object from the Analysis menu, move the cursor (which has changed into a resistor symbol) to the resistor on the schematic, and click with the left mouse button. To set a logarithmic scale on the Impedance axis, we have double-clicked on the vertical axis and set Scale to Logarithmic and the limits to 1 and 10k.

KEVIN’S WEBSURFER HANDBOOK VII
FOR CRYSTAL RADIO

DEDICATED ENTIRELY to VACUUM DIODES

Kevin Smith
2012

Printing / Binding Instructions

1. Choose “fit to page” in print menu
2. Print document double sided on letter size paper
3. Cut the entire printed document in half
4. Fold over making sure the page numbering is continuous
5. For the cover: Print just the first page on card stock paper
   Cut the cover in half as well
6. Assemble the covers on the document
7. Punch the left side for a binding, spiral or comb as desired

http://www.lessmiths.com/~kjsmith/crystal/catalog.shtml

KJ Smith

NOTES

INTRODUCTION

The contents of this latest handbook, volume 7 already, is concerned exclusively with the subject of vacuum diodes. Crystal radio is already a hobby dominated by an interest in vintage and antiquated technology. Vacuum tubes are yet another aspect of vintage radio technology. Interestingly, the vacuum diode was invented and patented prior to the crystal detector, one year prior it is true, but the priority is there.

As a crystal radio hobbyist, my interest is in diodes and rectification, not amplification so no triode (or higher) type tubes are on the menu. My personal interest in tubes stems from a childhood of listening to classic tube radios. I had one on a small shelf next to my bed and would often lay awake in the dark listening to music (turned down low so my parents would not hear) and look at the red glow from the inside of the set. There was something reassuring about the glow, you knew the radio would work as long as the glow was there. Agreed, having to wait a minute or so while the radio "warmed up" before hearing anything could be at times frustrating. When your friends called and said to "quick, turn on the radio and listen to this...", well, it was probably already too late. I also recall the first transistor radio and how cool it was to have the sound come on instantly. Still, what goes around, comes around. I recently was showing my adult daughter a cool old tube radio I had purchased. I turned it on and, as usual nothing... We talked on a bit about the hobby and stuff when, as if by magic, the sound from the radio behind us came on. She, astonished, exclaimed that’s "So cool!". Novelty is always fun. I note that she didn't go out to buy herself a tube iphone.

As tubes pretty much went the way of the horse and buggy with the introduction of transistors, most, virtually all good technical information on vacuum tubes is vintage in nature. As such, I have conceived this handbook as a historical walk down memory lane as well as technical reference for those, like myself, who wish to add vacuum tubes to their crystal radio projects and need specific technical help. I open with the words of Fleming himself describing the valves' pedigree and his personal involvement as the inventor of the vacuum valve. From there I move quickly to useful references and analysis from the 1940's, the heyday of the vacuum era. For much this I am MOST indebted to Pete Millett's tubebooks.org for scanned reference works on tubes, (http://www.tubebooks.org/). I have only included those portions of documents directly concerned with diode theory and operation. For your interest in triodes and amplification, I encourage you to visit Pete yourself.

I also include a good recent treatise on vacuum tubes by Calvert in 2003, (also included in my Handbook III on Diodes). And finally some of my own measurements, graphs, and notes on a selection of vacuum tubes I have acquired and played with.

While much of the material in this handbook is way out of copyright, some is and I have not sought permission. Therefore this is not presented for publication or copy, and certainly not for profit. It is only my personal resource. I encourage anyone finding this copy to pursue ON THE WEB the web pages identified, especially tubebooks.org. I wish to sincerely thank every author presented for their excellent pages and ask forgiveness for my editing into this handbook.

Kevin Smith
2012

www.lessmiths.com/~kjsmith/crystal/cr0intro.shtml

NOTES

CONTENTS:

1 **The Thermionic Valve: Historical Introduction**
JA Fleming

47 **Inside the Vacuum Tube: Chapter 6: The Diode**
JF Rider

79 **Principles of Electron Tubes: Section 2-10: Therionic Emission**
HJ Reich

107 **RCA Vacuum Tube Design: Analysis of Rectifier Operation**
OH Schade

147 **Characteristic curves of electron tubes**
PHYWE SYSTEME GMBH

157 **Diode Characteristics**
George Grammer

167 **Vacuum Diode**
Competition Science Vision

179 **Valve Technology**
Graham Dixey

187 **Basic Theory and Application: Chapter 2.3, DC Resistance of diode**
Army Technical 1952

193 **Theory of Vacuum Tubes**
JB Calvert 2003

237 **Diode Testing: Vacuum Diodes**
KJ Smith 2012

253 **Vacuum diode Models & PSpice Simulations**
Stefano Perugini

271 Notes

expressed its real potential. The evidence is given by modern formulation designs wich, often join even in concepts going back to the dawning of electronics (as well as harmonic cancellation, feed-forward, tansformer coupling, choke input filter etc.), can produce sonic results higher than the crowd of Williamson-like amplifiers.

References:

[1] Charles Rydel SIMULATION OF ELECTRON TUBES WITH SPICE", AES preprint 3887 (G-2), 98th AES Convention, Paris 1995;

[2] Scott Reynolds VACUUM TUBE MODELS FOR PSPICE SIMULATIONS, GLASS AUDIO n. 4, 1993

[3] W. Marshall Leach, Jr SPICE MODELS FOR VACUUM TUBE AMPLIFIERS, JAES March 1995

[4] Norman Koren IMPROVED VT MODELS FOR SPICE SIMULATION, GLASS AUDIO, N. 5 1996

[5] F. Langford-Smith Radiotron Designer's Handbook, 4th Ed. 1953

[6] Piotr Mikolajczyk Bohdan Paszkowski Universal Electronic VadeMecum

Condenser input filter		
	MEASUREMENT RESULTS	SIMULATION RESULTS
Vout	146.62Vcc	144.74Vcc
Repetitive peak current	70.59mA	68.21mA
Ripple Percentage	4.09%	4.14%
Choke input filter		
	MEASUREMENT RESULTS	SIMULATIONS RESULTS
Vout	101.320Vcc	101.116Vcc
Repetitive peak current	14.02mA	13.66ma
Ripple Percentage	0.52%	0.75%

Tab. 6;

Vin1, Vin2@120Vrms;
R1= R2= 6.8 ohm;
L1= 30H in choke input filter power supply;
C1= 8.10 mcrF;
Rl= 10.16K;
12X4 vacuum-diode.

The 12X4 SPICE model whose parameter have brought in line 1 of Table 3 have been drawn just from the real diode used for the construction of two power supplies (here explained the reason for the apparent redundance in Tabb. 1, 2, 3). So here I won't make a comparison on the grounds of a model drawn from the average characteristics, but with reference to a Sylvania's 12X4 sample. The utmost closeness of the numerical results in Table 6, makes me rather optimist about the validity of the implemented model; to the small differences that you can recognize, partly contribute also the errors of the measurement process in real world.

Conclusion

I hope this article can add a small wedge to the big mosaic of the circuit simulation applied to vacuum tube amplifiers. I believe the thermionic technology applied to the audio has not yet totally

---

THE
THERMIONIC VALVE

AND ITS DEVELOPMENTS IN RADIOTELEGRAPHY AND TELEPHONY

BY

J. A. FLEMING, M.A., D.Sc., F.R.S.,

University Professor of Electrical Engineering in the University of London, Member of the Institution of Electrical Engineers, Member of the Royal Institution of Great Britain, etc., etc.

AUTHOR OF
"THE PRINCIPLES OF ELECTRIC WAVE TELEGRAPHY AND TELEPHONY," "AN ELEMENTARY MANUAL OF RADIOTELEGRAPHY AND RADIOTELEPHONY," "THE WIRELESS TELEGRAPHIST'S POCKET BOOK OF NOTES, FORMULÆ AND CALCULATIONS," "THE WONDERS OF WIRELESS TELEGRAPHY," ETC., ETC.

LONDON
THE WIRELESS PRESS, LTD.
12 & 13 HENRIETTA STREET, STRAND, W.C.2
SYDNEY, N.S.W.: 97 CLARENCE STREET
MELBOURNE: 422/4 LITTLE COLLINS STREET

NEW YORK
WIRELESS PRESS INC.
25 ELM STREET

ROME
AGENZIA RADIOTELEGRAFICA ITALIANA
VIA GREGORIANA, 36

MADRID
LA PRENSA RADIOTELEGRAFICA
43 CALLE DE ALCALA

ROTTERDAM
NEDERLANDSCH PERSBUREAU RADIO
BOOMPJES 23A

1919

UNIV. OF

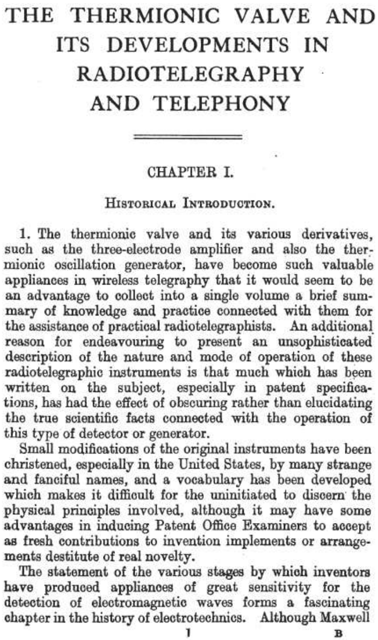

THE THERMIONIC VALVE AND ITS DEVELOPMENTS IN RADIOTELEGRAPHY AND TELEPHONY

CHAPTER I.

HISTORICAL INTRODUCTION.

1. The thermionic valve and its various derivatives, such as the three-electrode amplifier and also the thermionic oscillation generator, have become such valuable appliances in wireless telegraphy that it would seem to be an advantage to collect into a single volume a brief summary of knowledge and practice connected with them for the assistance of practical radiotelegraphists. An additional reason for endeavouring to present an unsophisticated description of the nature and mode of operation of these radiotelegraphic instruments is that much which has been written on the subject, especially in patent specifications, has had the effect of obscuring rather than elucidating the true scientific facts connected with the operation of this type of detector or generator.

Small modifications of the original instruments have been christened, especially in the United States, by many strange and fanciful names, and a vocabulary has been developed which makes it difficult for the uninitiated to discern the physical principles involved, although it may have some advantages in inducing Patent Office Examiners to accept as fresh contributions to invention implements or arrangements destitute of real novelty.

The statement of the various stages by which inventors have produced appliances of great sensitivity for the detection of electromagnetic waves forms a fascinating chapter in the history of electrotechnics. Although Maxwell

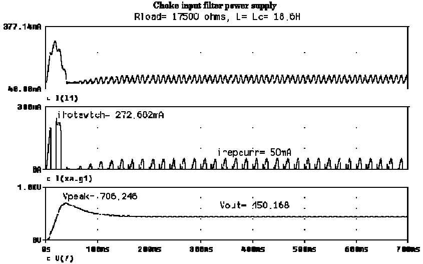

Fig. 11;

As you can observe in Fig. 9 it is rather easy to overcome the limits related to the impulsive currents also with small input condensers therefore their capacitive value has to be valued very attentively. Besides Fig. 11 reveals that the swinging choke calculus is not very accurate.

An actual comparison

In Table 6 I present the results of two simple power supplies with condenser and choke input filter respectively. Schematics are the same of Figs. 9 and 10 but with the following differences:

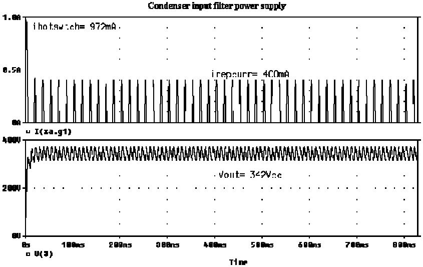

Fig. 9;

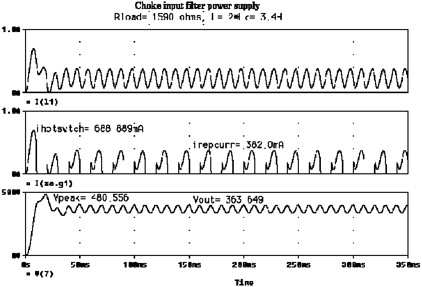

Fig. 10;

enunciated in 1865 his great epoch-making theory that electromagnetic effects must be propagated through dielectrics as a wave motion, and surmised that such electromagnetic radiation was identical in nature with light, the acceptance of his views was retarded by the want of experimental proof that such waves could be created and detected by purely electromagnetic methods. Hertz's invention of the ring resonator with micro-spark gap supplied the first means of detecting these space waves, which G. F. Fitzgerald had previously suggested might be created by the oscillatory discharge of a Leyden jar. The employment of Maxwell's electromagnetic waves to effect radiotelegraphy, involving the propagation of such waves over distances reckoned in miles, was fundamentally dependent on Marconi's key invention of the aerial wire or antenna, but it necessitated as well the possession of far more sensitive and certain means of detecting these waves than the ring resonator of Hertz.

In radiotelegraphy we make use of electromagnetic waves propagated over the earth's surface to set up in a receiving aerial wire feeble electric oscillations which are a copy on a very reduced scale of the strong oscillations created in the sending aerial wire. In modern radiotelegraphy the oscillations are generated in the aerials either continuously or else in uniformly time-spaced trains, the train or group frequency agreeing with that frequency for which the Bell telephone receiver, acting on the human ear, is most sensitive —viz., about 300-500 per second. These oscillations are then cut up into short and long periods or groups of trains to create the Morse-code signals.

In the ordinary language of wireless telegraphy we call any instrument a *detector* which is employed to detect or make evident by audible or visible means these groups or intermittencies of very feeble electric oscillations in the receiving aerial wire and associated circuits which convey the intelligible signals.

The signal-making instrument may be some form of telegraphic printer, such as the Kelvin syphon recorder or the Morse inker, or the more sensitive Einthoven galvanometer with photographic tape record. On the other hand it may be a Bell telephone receiver, making an audible sound and appealing therefore directly to the ear. In

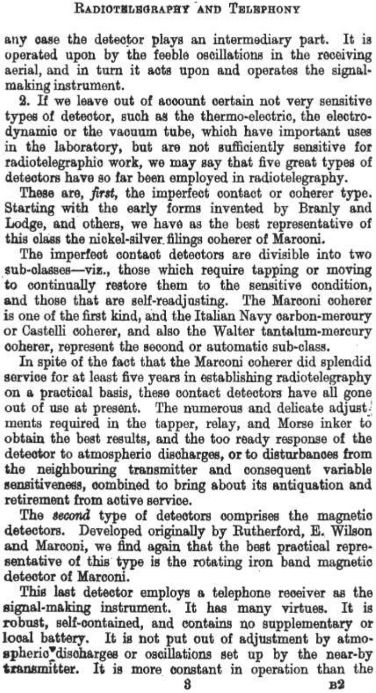

RADIOTELEGRAPHY AND TELEPHONY

any case the detector plays an intermediary part. It is operated upon by the feeble oscillations in the receiving aerial, and in turn it acts upon and operates the signal-making instrument.

2. If we leave out of account certain not very sensitive types of detector, such as the thermo-electric, the electro-dynamic or the vacuum tube, which have important uses in the laboratory, but are not sufficiently sensitive for radiotelegraphic work, we may say that five great types of detectors have so far been employed in radiotelegraphy.

These are, *first*, the imperfect contact or coherer type. Starting with the early forms invented by Branly and Lodge, and others, we have as the best representative of this class the nickel-silver filings coherer of Marconi.

The imperfect contact detectors are divisible into two sub-classes—viz., those which require tapping or moving to continually restore them to the sensitive condition, and those that are self-readjusting. The Marconi coherer is one of the first kind, and the Italian Navy carbon-mercury or Castelli coherer, and also the Walter tantalum-mercury coherer, represent the second or automatic sub-class.

In spite of the fact that the Marconi coherer did splendid service for at least five years in establishing radiotelegraphy on a practical basis, these contact detectors have all gone out of use at present. The numerous and delicate adjustments required in the tapper, relay, and Morse inker to obtain the best results, and the too ready response of the detector to atmospheric discharges, or to disturbances from the neighbouring transmitter and consequent variable sensitiveness, combined to bring about its antiquation and retirement from active service.

The *second* type of detectors comprises the magnetic detectors. Developed originally by Rutherford, E. Wilson and Marconi, we find again that the best practical representative of this type is the rotating iron band magnetic detector of Marconi.

This last detector employs a telephone receiver as the signal-making instrument. It has many virtues. It is robust, self-contained, and contains no supplementary or local battery. It is not put out of adjustment by atmospheric discharges or oscillations set up by the near-by transmitter. It is more constant in operation than the

3 B2

Rs= total effective plate supply impedance per plate.

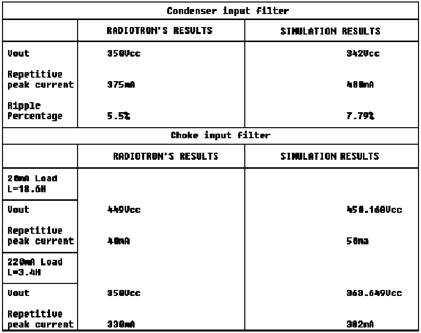

Condenser input filter		
	RADIOTRON'S RESULTS	SIMULATION RESULTS
Vout	350Vcc	342Vcc
Repetitive peak current	375mA	400mA
Ripple Percentage	5.5%	7.79%
Choke input filter		
	RADIOTRON'S RESULTS	SIMULATION RESULTS
20mA Load L=18.6H		
Vout	449Vcc	450.160Vcc
Repetitive peak current	40mA	50ma
220mA Load L=3.4H		
Vout	350Vcc	363.649Vcc
Repetitive peak current	330mA	302mA

Tab. 5;

In Table 5 I show a comparison between manual calculations (as brought in the Radiotron's) and simulations. The more meaningful graphic representations of such simulations are brought in Figg 9, 10, 11.

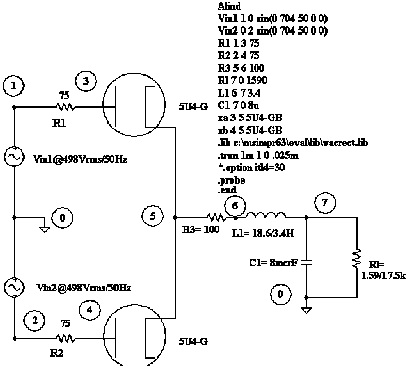

```
Alind
Vin1 1 0 sin(0 704 50 0 0)
Vin2 0 2 sin(0 704 50 0 0)
R1 1 3 75
R2 2 4 75
R3 5 6 100
Rl 7 0 1590
L1 6 7 3.4
C1 7 0 8u
xa 3 5 5U4-GB
xb 4 5 5U4-GB
.lib c:\msimpr63\eval\lib\vacrect.lib
.tran 1m 1 0 .025m
*.option itl4=30
.probe
.end
```

Fig. 8;

In Figg. 7 and 8 I have brought two hypotheses of power supply with condenser and choke input filter respectively together with their SPICE codes related to present examples beginning from the pages 1174 and 1183 of the Radiotron Designer's Handbook. The resistances R1 and R2 are the total effective plate supply impedance per plate and they represent the impedances brought to every secondary winding according to the formula:

Rs= Rsec + N2 * Rpri;

where:

N= Voltage ratio of transformer at no load (primary to half secondary in case of full-wave rectification);
Rpri= Resistance of primary winding in ohms;
Rsec= Resistance of secondary windind in ohms (or half secondary in case of full-wave rectification;

---

capricious coherer, but it is not one of the most sensitive detectors. Nevertheless, for military or ship work it is still much valued, although not very light in weight.

The *third* class of detector comprises the electrolytic detectors in which the electrical oscillations to be detected are caused to alter the conductivity of an electrolytic cell by varying the ionic films on the electrodes. These detectors have never come into very extensive use in Great Britain.

The *fourth* type of detector, and until lately probably the most extensively used, are the rectifying contact or crystal detectors. Starting with Dunwoody's discovery of the rectifying property of certain crystals of carborundum and with the discoveries of Pickard and G. W. Pierce on the rectifying power of contacts between crystals of zincite and chalcopyrite and molybdenite and copper, there has been an immense amount of research on this type of detector, which depends either upon a true unilateral conductivity in the mass or at the contact or else upon sudden changes in curvature in the volt-ampere or characteristic curve.

The simplest mode of use of such crystal detectors is as rectifiers of the oscillation trains into gushes of electricity in the same direction, thus enabling trains of electrical oscillations to affect a telephone receiver. These detectors are rather easily put out of adjustment by strong atmospheric discharges, or by strong oscillations set up by the proximity of the transmitter. They therefore require somewhat frequent readjustment of contact to obtain the best results. Nevertheless, their sensitiveness, simplicity and cheapness made them very quickly a favourite detector. Their introduction was followed at once by an outburst of irresponsible radiotelegraphy at the hands of innumerable electrical amateurs and students which required the firm intervention of National and International legislation to keep it within bounds of reason and safety.

The *fifth* class of detector comprises those to the consideration of which this book is limited—viz., the thermionic detectors. These depend ultimately upon the emission from incandescent bodies of ions or electrons. They are or can be made highly sensitive and yet are not affected injuriously by atmospherics or by neighbouring electric sparks. They are easy of adjustment and always ready for use. The original of all these detectors is the now well-known Fleming

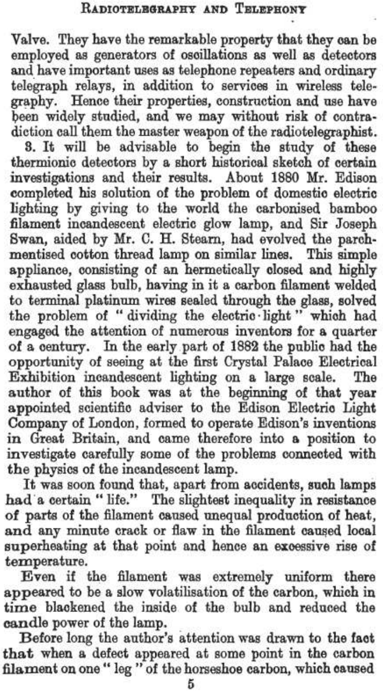

Valve. They have the remarkable property that they can be employed as generators of oscillations as well as detectors and have important uses as telephone repeaters and ordinary telegraph relays, in addition to services in wireless telegraphy. Hence their properties, construction and use have been widely studied, and we may without risk of contradiction call them the master weapon of the radiotelegraphist.

3. It will be advisable to begin the study of these thermionic detectors by a short historical sketch of certain investigations and their results. About 1880 Mr. Edison completed his solution of the problem of domestic electric lighting by giving to the world the carbonised bamboo filament incandescent electric glow lamp, and Sir Joseph Swan, aided by Mr. C. H. Stearn, had evolved the parchmentised cotton thread lamp on similar lines. This simple appliance, consisting of an hermetically closed and highly exhausted glass bulb, having in it a carbon filament welded to terminal platinum wires sealed through the glass, solved the problem of "dividing the electric light" which had engaged the attention of numerous inventors for a quarter of a century. In the early part of 1882 the public had the opportunity of seeing at the first Crystal Palace Electrical Exhibition incandescent lighting on a large scale. The author of this book was at the beginning of that year appointed scientific adviser to the Edison Electric Light Company of London, formed to operate Edison's inventions in Great Britain, and came therefore into a position to investigate carefully some of the problems connected with the physics of the incandescent lamp.

It was soon found that, apart from accidents, such lamps had a certain "life." The slightest inequality in resistance of parts of the filament caused unequal production of heat, and any minute crack or flaw in the filament caused local superheating at that point and hence an excessive rise of temperature.

Even if the filament was extremely uniform there appeared to be a slow volatilisation of the carbon, which in time blackened the inside of the bulb and reduced the candle power of the lamp.

Before long the author's attention was drawn to the fact that when a defect appeared at some point in the carbon filament on one "leg" of the horseshoe carbon, which caused

calculations and formulae just finalized to the solution of actual situation. Other texts I know lack suitable depth or face the subject in a too academic way not enough finalized to the solution of real problems. On the contrary in the Radiotron's by fast procedures it is possible plan power supplies with both condenser input filter and choke input one. Unfortunately the accuracy, especially in the evaluation of the impulsive currents, is not high but this is due to the use of purposely approximate formulae to have agile calculation that at the time were made by the hand. Besides the book introduces two levels of approximation if you have complete data or not.

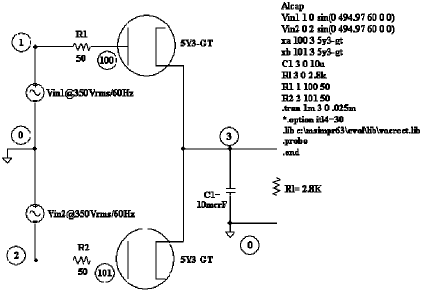

```
Alcap
Vin1 1 0 sin(0 494.97 60 0 0)
Vin2 0 2 sin(0 494.97 60 0 0)
xa 100 3 5y3-gt
xb 101 3 5y3-gt
C1 3 0 10u
Rl 3 0 2.8k
R1 1 100 50
R2 2 101 50
.tran 1m 3 0 .025m
*.option itl4=30
.lib c:\msimpr63\eval\lib\vacrect.lib
.probe
.end
```

Fig. 7;

Now you can use the data in Table 3 to build the SPICE model of the diode. After the excellent articles by Raynolds [2], Marshall [3] and Koren [4], I don't think you will find any difficulty to understand the code in Table 4. The only required operation is to use the information in Table 3 and complete the code in Table 4 on the grounds of the selected diode.

```
.SUBCKT tubename P K
+ PARAMS Ka=       Kb=
+ A=       Eps=
E1              1  0   VALUE = {Ka + Kb * V(P,K)}
RE1   1  0    1G
E2              2   0   VALUE = {V(P,K) + EPS}
RE2   2   0    1G
G1             P     K       VALUE = {V(1)/2 * (PWR(V(2), A) + PWRS(V(2), A))}
RPK      P   K    1MEG
* CPK   P   K     .5n
.ENDS
```

(*) CPK can replace RPK in case of serious convergence problems.
Tab. 4

How use this models

The first reason that can push you to look for an accurate model of vacuum diode is that, you can determine, when the device is used as a rectifier, a series of parameters whose exact evaluation can be either difficult or inaccurate as for istance the repetitive peak current, the hot-switching current, the output impedance etc. Besides the SPICE model you need further informations to place the device in safety areas. These informations are graphically available in Data-Sheets Rating Charts. In my opinion the Radiotron Designer's Handbook magistrally deals with the subject as it faces the whole problem list merely in engineering terms and therefore with

a great rise of temperature at that point, the blackening of the interior of the bulb was less or absent along a certain line in such fashion as to show that the scattered carbon molecules had been shot off in straight lines from the point of excessive heating. Thus was formed what the author termed *a molecular shadow* on the bulb (see Fig. 1). If, for instance, a scent-sprayer is filled with ink and the spray blown on to a sheet of paper the latter will be darkened uniformly by minute drops of ink spray. If, however, a wire or rod is held in the path of the spray the paper will be protected from blackening along a region which corresponds to the shadow of the rod (see Fig. 2). Hence when we find in old carbon filament incandescent lamps a horseshoe filament broken at one spot and the bulb darkened all over the inside with the exception of a white or undarkened line on the bulb corresponding to the shadow of the unbroken part of the filament relatively to the point of rupture or overheating we can conclude that the chief part of the scattering of the carbon particles has taken place in straight lines proceeding from the point in the filament at which it has ultimately burnt through. This phenomenon of "molecular shadows" in incandescent lamps was described by the author in two papers read to the Physical Society of London in 1883 and 1885.[1]

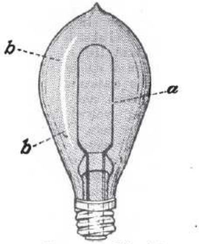

FIG. 1.—Incandescent Electric Lamp, showing the glass bulb blackened by the deposit of carbon or metal due to molecular scattering from point *a* on the filament, a shadow line of no deposit being formed at *b*, due to shielding effect of one leg of the horseshoe filament.

This effect clearly showed that from the incandescent filament there was a projection of matter in straight lines which was deposited on the interior of the bulb, except in those parts shielded by the "shadow" of the rest of the filament, thus leaving a white or unblackened line. This effect is obviously due to the fact that the residual air in the bulb is at such a rarefied condition that the mean free path of a molecule is comparable in length with the diameter of the bulb.

---

[1] See *Proc. Phys. Soc. Lond.*, vol. v., p. 283, 1883, "On a Phenomenon of Molecular Radiation in Incandescent Lamps," by J. A. Fleming; also *Proc. Phys. Soc. Lond.*, vol. vii., p. 178, 1885, "On Molecular Shadows in Incandescent Lamps," by J. A. Fleming.

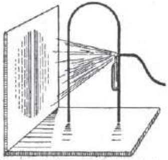

FIG. 2.—A "spray shadow" of a rod thrown on a cardboard screen by ink spray to illustrate the mode of formation of a molecular shadow in a glow lamp.

In air at ordinary temperature and pressure the mean free path of an oxygen or nitrogen molecule is about four one-millionths of an inch, but when the pressure is reduced in a bulb to one-millionth of an atmosphere it becomes increased to about four inches. This means that a molecule of carbon projected in any way from the filament might travel on an average four inches before colliding with other residual air molecules and being deflected from its path.

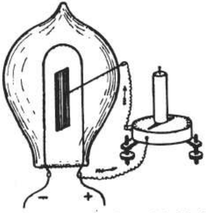

FIG. 3.—Incandescent Lamp with a Metal Plate sealed into the bulb and a galvanometer connected between the terminal of the plate and positive terminal of the filament to show the "Edison effect."

The next important observation on this subject was made by Mr. Edison in 1883.[1] He was apparently examining the phenomena involved when carbon filament lamps are running at a high efficiency, and for that purpose he sealed into the bulb of a glow lamp a metal plate placed between the legs of the horseshoe-shaped carbon filament, the

---

[1] See *Engineering*, December 12th, 1884, p. 553.

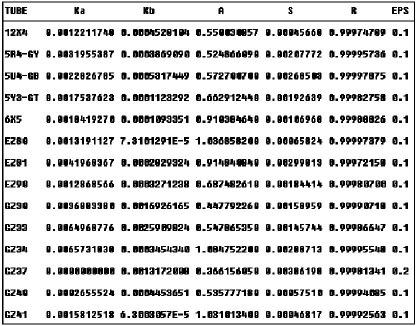

TUBE	Ka	Kb	A	S	R	EPS
12X4	0.0012211740	0.0004520104	0.550030857	0.00045660	0.99974709	0.1
5R4-GY	0.0031955387	0.0003869090	0.524866090	0.00207772	0.99995736	0.1
5U4-GB	0.0022026705	0.0005317449	0.572700700	0.00268503	0.99997075	0.1
5Y3-GT	0.0017537623	0.0001123292	0.662912440	0.00192639	0.99982758	0.1
6X5	0.0010419270	0.0001093351	0.910304640	0.00106960	0.99900026	0.1
EZ80	0.0013191127	7.3101291E-5	1.036858200	0.00065824	0.99997379	0.1
EZ81	0.0041960367	0.0002029324	0.914040040	0.00299013	0.99972150	0.1
EZ90	0.0012868566	0.0003271238	0.687482610	0.00184414	0.99980700	0.1
GZ30	0.0096009980	0.0016926165	0.447792260	0.00158959	0.99990710	0.1
GZ33	0.0064960776	0.0025909024	0.547065350	0.00145744	0.99986647	0.1
GZ34	0.0065731030	0.0003454340	1.004752200	0.00200713	0.99995540	0.1
GZ37	0.0000000000	0.0013172000	0.366156050	0.00306190	0.99981341	0.2
GZ40	0.0002655524	0.0004453651	0.535777180	0.00057510	0.99994085	0.1
GZ41	0.0015812518	6.3603057E-5	1.031013400	0.00046817	0.99992563	0.1

Tab. 3; Results for $I_p = (K_a + K_b \cdot V_p) \cdot (V_p + EPS)^A$ model

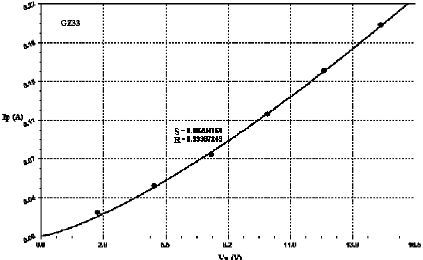

Fig. 6;

Finally you can get an even better fitting if you use the following equation:

$$Ip= (Ka+Kb*Vp)*(Vp+EPS)A \quad (3).$$

The equation (3) differs from (2) because a linear variation to the perveance has been attributed with respect to Vp. The application of LSM with linear regression leads to the data in Table 3. You can get further short improvement margins using more sophisticated mathematical models although this injuries simplicity and brings larger problems of convergence in simulations.

plate being carried on an insulated wire sealed through the bulb (see Fig. 3). He found that, when the filament was rendered incandescent by a direct current of electricity, a galvanometer connected between the middle plate and the positive terminal of the filament indicated a current of a few milliamperes, but little or no current when connected between the negative terminal and the middle plate (see Fig. 4). He did not furnish any explanation of this effect at that time, but it became known then and after as the "Edison effect" in glow lamps.

Mr. Edison gave to Sir William Preece in October, 1884, certain incandescent lamps made with metal plates sealed into their bulbs or into tubular extensions of them, and in the following year, in March, 1885, Preece described to the Royal Society of London experiments which he had made with these lamps.[1] Preece made quantitative measurements of the current flowing through the galvanometer when connected in between the positive terminal of the carbon filament and the middle plate at various voltages applied to the filament. He found that this current was independent of the nature of the metal of which the middle plate was made, but that it increased very rapidly with increase in the potential difference of the filament terminals. He found that for a given lamp voltage the Edison effect varied with the position of the plate in the bulb, but that it was still sensible when this plate was placed at the closed end of a tube opening out of the bulb, provided this tube was straight, but the current was found to vanish if that side tube was bent up at right angles (see Figs. 5 and 6).

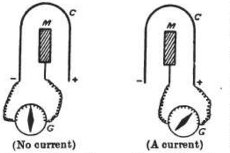

(No current) (A current)

FIG. 4.—Mode of connection of the galvanometer, *G*, to metal plate, *M*, and to the terminals of the carbon filament of the lamp (+and−) in the experiment of the "Edison effect."

---

[1] See *Proc. Roy. Soc. Lond.*, vol. xxxviii., p. 219, 1885, "On a Peculiar Behaviour of Glow Lamps when Raised to High Incandescence," by W. H. Preece.

Preece considered that this Edison effect was connected with the discharge or projection of carbon molecules or particles from the filament in straight lines, but he gave

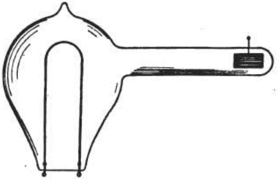

FIG. 5.—Collecting Plate placed at the end of a side tube opening out of the bulb of an incandescent lamp.

no full theory of the phenomenon, nor did he make any application of the facts. A year or two later the author of this book took up the investigation, being convinced that much yet remained to be discovered about it which had not been unravelled by Edison or Preece.

A number of special carbon filament glow lamps with middle plates were made for the author at the Lamp Factory of the Edison and Swan United Electric Light Company and researches were begun with them, and the results communicated to the Royal Society of London in 1889.[1] and in a Friday evening

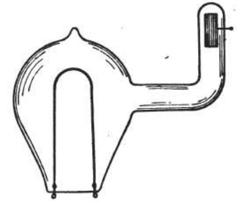

FIG. 6.—Collecting Plate placed at the end of an elbow tube opening out of the bulb of an incandescent lamp.

---

[1] See *Proc. Roy. Soc. Lond.*, vol. xlvii., p. 118, 1890, "On Electric Discharge between Electrodes at Different Temperatures in Air and in High Vacua," by J. A. Fleming.

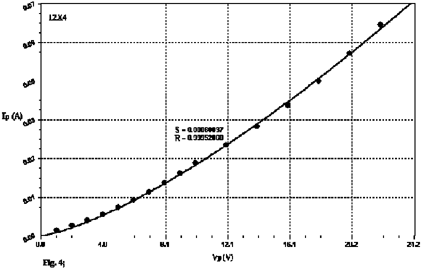

Fig. 4;

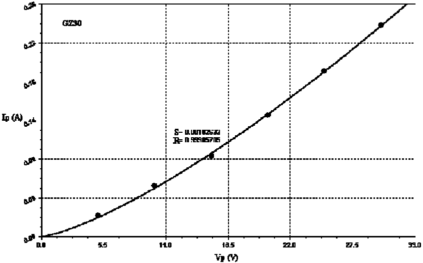

Fig. 5;

TUBE	K	A	S	R	EPS
12X4	0.008775603	1.3693599	0.00060837	0.99952060	0.1
5R4-GY	0.000673836	1.4214962	0.00324640	0.99988644	0.2
5U4-GB	0.007051655	1.5203189	0.00358406	0.99995741	0.2
5Y3-GT	0.000326542	1.4606772	0.00273940	0.99960771	0.2
6X5	0.000649553	1.5190353	0.00183847	0.99962271	0.2
CZ80	0.000549820	1.5792666	0.00144462	0.99974752	0.2
F781	0.001864868	1.4770666	0.00549127	0.99883198	0.008
EZ90	0.000575947	1.5583772	0.00177745	0.99976094	0.1
GZ30	0.002570010	1.3436975	0.00102633	0.99985705	0.1
GZ33	0.005317957	1.3349462	0.00204164	0.99967243	0.1
GZ34	0.002943400	1.5969530	0.00668603	0.99970154	0.1
GZ37	0.001325449	1.3640025	0.00304527	0.99981502	0.01
GZ40	0.000485414	1.5155771	0.00055090	0.99993797	0.1
GZ41	0.000915285	1.4108217	0.00074455	0.99976487	0.1

Tab. 2; Results for the $Ip= K*(Vp+EPS)^A$

to apply LSM to the set of experimental data. The results of this procedure have quoted in Table 2. EPS represents a parameter that can be manually fixed (typical values are 0.1, 0.2) used to guarantee a better convergence of the algorithm. If you compare the Correlation Coefficient of the Table 1 and 2, you can note a real amelioration, underlined subsequently by the new graphic representations of the tubes 12X4, GZ30, GZ33 related in Figg. 4, 5, 6. Futhermore the examination of Table 2 reveals that the real diode is subjugated by the 3/2 power law only in few cases.

discourse on February 14th, 1890, to the Royal Institution.[1]

One of the first new facts the author discovered was that the "Edison effect" was greatly diminished or entirely annulled by surrounding the negative leg of the carbon filament by a glass or metal tube, or by interposing a mica shield between the negative leg of the horseshoe-shaped carbon filament and that side of the middle or insulated plate facing the negative leg (see Figs. 7, 8 and 9). By negative leg is meant that side of the carbon horseshoe loop in connection with the negative pole of the battery which heats the filament. It was not annulled or much diminished when the tube surrounded the positive leg.

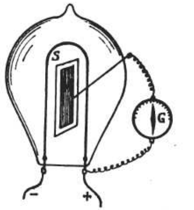

FIG. 7.—Incandescent Lamp having a mica shield placed between the metal plate, *M*, and the negative leg of carbon loop filament, thereby diminishing the "Edison effect."

The author confirmed Preece's observation that the effect

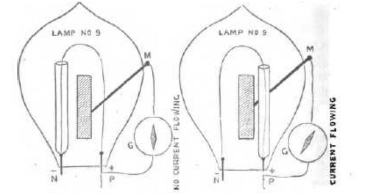

FIG. 8.—Glass or Metal Tube surrounding one leg of horseshoe carbon filament in a glow lamp to show the annulment of the "Edison effect" when it surrounds the negative leg.

---

[1] "Problems in the Physics of an Electric Lamp." *Proceedings of the Royal Institution*, February 14th, 1890.

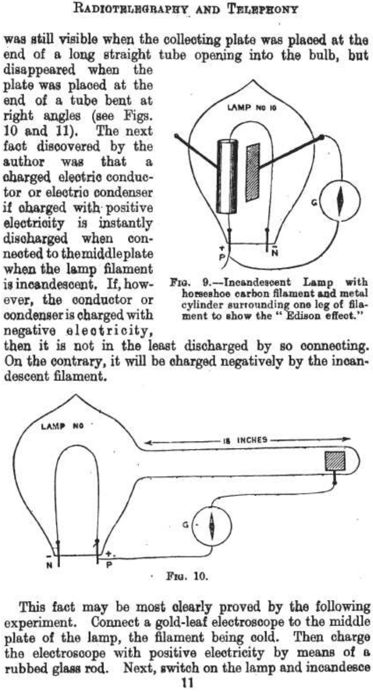
LAMP NO 10
G
+
P
N
LAMP NO
18 INCHES
G
N
+
P

was still visible when the collecting plate was placed at the end of a long straight tube opening into the bulb, but disappeared when the plate was placed at the end of a tube bent at right angles (see Figs. 10 and 11). The next fact discovered by the author was that a charged electric conductor or electric condenser if charged with positive electricity is instantly discharged when connected to the middle plate when the lamp filament is incandescent. If, however, the conductor or condenser is charged with negative electricity, then it is not in the least discharged by so connecting. On the contrary, it will be charged negatively by the incandescent filament.

FIG. 9.—Incandescent Lamp with horseshoe carbon filament and metal cylinder surrounding one leg of filament to show the "Edison effect."

FIG. 10.

This fact may be most clearly proved by the following experiment. Connect a gold-leaf electroscope to the middle plate of the lamp, the filament being cold. Then charge the electroscope with positive electricity by means of a rubbed glass rod. Next, switch on the lamp and incandesce

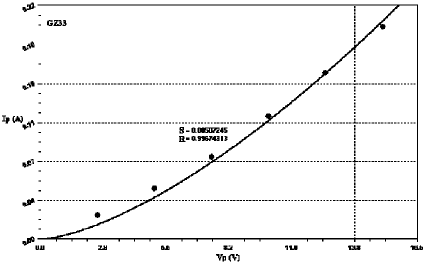

Fig. 3;

Improvement

You can already get a large improvement by varying the exponent of the equation (1). This leads to the following expression:

Ip= K*VpA.

Now use this equation, but modified in the following way:

Ip= K*(Vp+EPS)A (2)

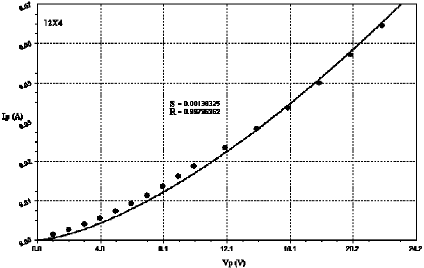

Fig. 1;

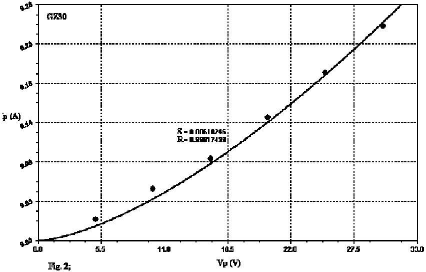

Fig. 2;

the filament. The electroscope instantly loses its positive charge. Then repeat the experiment, charging the gold leaves negatively by means of a rod of sealing wax, and it will be found that switching on the lamp does not discharge the electroscope. These experiments show that the incandescent filament is giving off negative electricity in some manner.

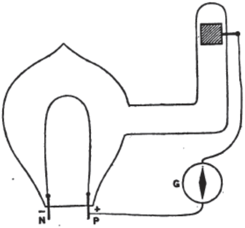

FIG. 11.

The next experiment was arranged to prove that the space between the incandescent filament and the middle plate was conductive for negative electricity from the filament to the plate, but not in the opposite direction. For this purpose a single cell of a battery was joined in series with a galvanometer and one

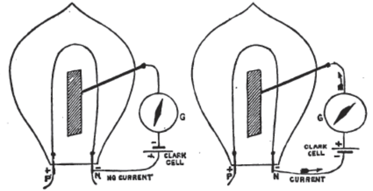

FIG. 12.—Incandescent Lamp with metal plate in the bulb and galvanometer, *G*, connecting plate and negative terminal. Clark cell inserted in plate circuit to show the escape of negative electricity from the incandescent filament.

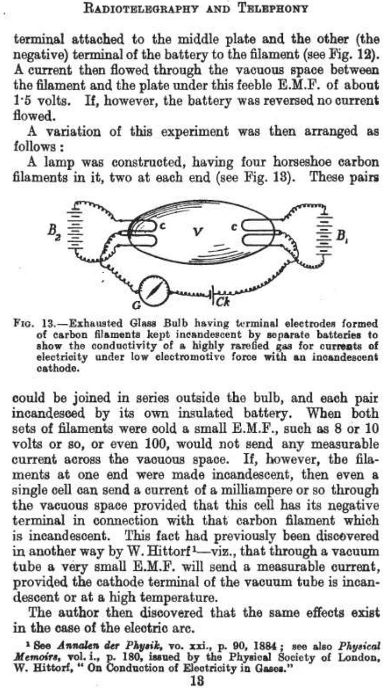

terminal attached to the middle plate and the other (the negative) terminal of the battery to the filament (see Fig. 12). A current then flowed through the vacuous space between the filament and the plate under this feeble E.M.F. of about 1·5 volts. If, however, the battery was reversed no current flowed.

A variation of this experiment was then arranged as follows:

A lamp was constructed, having four horseshoe carbon filaments in it, two at each end (see Fig. 13). These pairs could be joined in series outside the bulb, and each pair incandesced by its own insulated battery. When both sets of filaments were cold a small E.M.F., such as 8 or 10 volts or so, or even 100, would not send any measurable current across the vacuous space. If, however, the filaments at one end were made incandescent, then even a single cell can send a current of a milliampere or so through the vacuous space provided that this cell has its negative terminal in connection with that carbon filament which is incandescent. This fact had previously been discovered in another way by W. Hittorf[1]—viz., that through a vacuum tube a very small E.M.F. will send a measurable current, provided the cathode terminal of the vacuum tube is incandescent or at a high temperature.

FIG. 13.—Exhausted Glass Bulb having terminal electrodes formed of carbon filaments kept incandescent by separate batteries to show the conductivity of a highly rarefied gas for currents of electricity under low electromotive force with an incandescent cathode.

The author then discovered that the same effects exist in the case of the electric arc.

---

[1] See *Annalen der Physik*, vo. xxi., p. 90, 1884; see also *Physical Memoirs*, vol. i., p. 180, issued by the Physical Society of London, W. Hittorf, "On Conduction of Electricity in Gases."

You can see the perveance as the digital imprint or the genetic code of the diode because in (1) it constitues the only parameter able to differentiate the multiplicity of the diodes.

Besides the Child-Langmuir's Law is a phenomenological equation, that is derived on physical considerations grounds.

If we use this equation to apply the Least Square Method (LSM) with linear regression to the set of experimental data extracted by the Data-Sheets it is difficult to get a good fitting, as you can see by examining Table 1. In this table S and R represent the Standard Error and the Correlation Coefficient respectively. A perfect fitting is gotten when R is equal to 1. The good values of the Correlation Coefficient R brought in Table 1 must not deceive you, since R is a paramenter of global evaluation and lower values to 0.998 cannot be thought satisfactory as the graphs of the Figg. 1, 2, 3 related to the diodes 12X4, GZ30, GZ33 show, in wich the model introduces, locally, marked deviations. In this graphs the "small black balls" represent the average data extracted from Data-Sheets. In my opinion this model well represent only tubes like 5U4 and GZ40: you can think of them as perfect diodes (although not ideal !). In all the other cases the results can be improved so to get nearest simulations to the reality. The apparent redundance in Table 1 and those following where is present either 12X4 or EZ90/6X4 will be clarified later.

In the "sacred texts" the equation brought for describing the operation of the diode is the well note Child-Langmuir's Law:

Ip= K*Vp1.5 (1);
where:

Ip is the current that flows in the diode;
Vp is the anode to cathode voltage;
K is a costant said perveance.

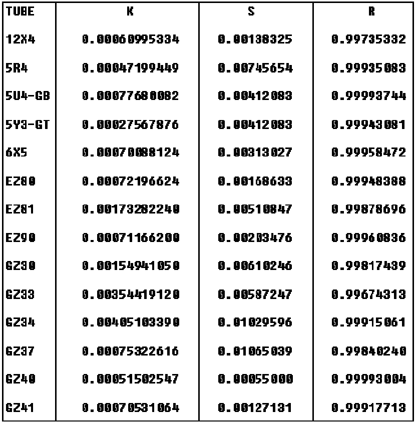

TUBE	K	S	R
12X4	0.00060995334	0.00138325	0.99735332
5R4	0.00047199449	0.00745654	0.99935083
5U4-GB	0.00077680082	0.00412083	0.99993744
5Y3-GT	0.00027567876	0.00412083	0.99943081
6X5	0.00070088124	0.00313027	0.99958472
EZ80	0.00072196624	0.00168633	0.99948388
EZ81	0.00173282240	0.00510847	0.99878696
EZ90	0.00071166200	0.00203476	0.99960836
GZ30	0.00154941050	0.00610246	0.99817439
GZ33	0.00354419120	0.00587247	0.99674313
GZ34	0.00405103390	0.01029596	0.99915061
GZ37	0.00075322616	0.01065039	0.99840240
GZ40	0.00051502547	0.00055000	0.99993004
GZ41	0.00070531064	0.00127131	0.99917713

Tab. 1; Results for the Ip=k*(Vp)$^{1.5}$ model

If a direct current electric arc of some length is formed between carbon electrodes, and if a third carbon electrode is so placed as just to dip into the flame of the arc, or if the arc is deflected against it by a magnet, it will be found that a galvanometer or ammeter connected between the positive carbon and the third or lateral carbon indicates a current which may even be strong enough to ring an electric bell. If, however, the ammeter or bell is connected between the negative carbon of the arc and the middle carbon, then no current passes (see Fig. 14).

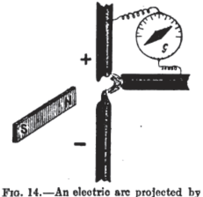

FIG. 14.—An electric arc projected by means of a magnet against a third carbon to show the "Edison effect" in the case of an electric arc.

Also, it can be shown by the same means that the space between the negative carbon of the arc and the middle carbon possesses a unilateral conductivity and will permit negative electricity from a battery to pass from this negative carbon to the middle carbon, but not in the opposite direction (see Fig. 15).

Finally, the author discovered that the same phenomenon can be demonstrated in the air at ordinary pressure. For if a thick carbon filament is rendered incandescent in air by passing a strong current through it, and if an insulated metal plate is held near to it, a galvanometer joined in between this metal

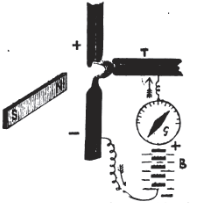

FIG. 15.—An experiment with the electric arc having a third carbon to show that a current of negative electricity can pass under low E.M.F. between the negative terminal of the arc and the third carbon electrode.

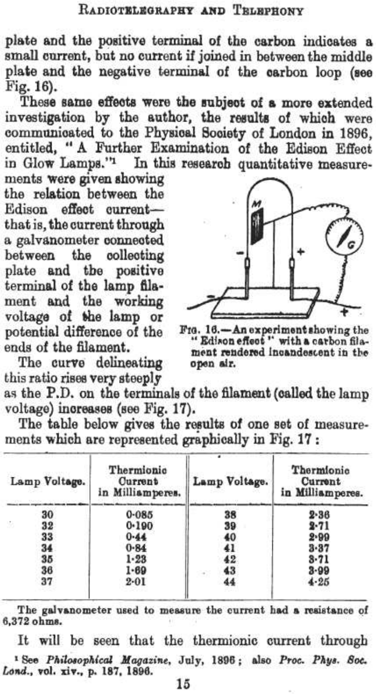

plate and the positive terminal of the carbon indicates a small current, but no current if joined in between the middle plate and the negative terminal of the carbon loop (see Fig. 16).

These same effects were the subject of a more extended investigation by the author, the results of which were communicated to the Physical Society of London in 1896, entitled, "A Further Examination of the Edison Effect in Glow Lamps."[1] In this research quantitative measurements were given showing the relation between the Edison effect current—that is, the current through a galvanometer connected between the collecting plate and the positive terminal of the lamp filament and the working voltage of the lamp or potential difference of the ends of the filament.

FIG. 16.—An experiment showing the "Edison effect" with a carbon filament rendered incandescent in the open air.

The curve delineating this ratio rises very steeply as the P.D. on the terminals of the filament (called the lamp voltage) increases (see Fig. 17).

The table below gives the results of one set of measurements which are represented graphically in Fig. 17:

Lamp Voltage.	Thermionic Current in Milliamperes.	Lamp Voltage.	Thermionic Current in Milliamperes.
30	0·085	38	2·36
32	0·190	39	2·71
33	0·44	40	2·99
34	0·84	41	3·37
35	1·23	42	3·71
36	1·69	43	3·99
37	2·01	44	4·25

The galvanometer used to measure the current had a resistance of 6,372 ohms.

It will be seen that the thermionic current through

---

[1] See *Philosophical Magazine*, July, 1896; also *Proc. Phys. Soc. Lond.*, vol. xiv., p. 187, 1896.

Vacuum diode Models & PSpice Simulations

Stefano Perugini
http://digilander.iol.it/paeng/vacuum_diode_models.htm

This article appeared originally in Glass Audio, Vol.10,Nr. 4, 1998.

The renewed interest in the thermionic technology applied to the audio is pushing the most sensitive designers to use, widely, the modern tools of circuit simulation SPICE-oriented, in the convincement that the best sonic results to be gotten with a deep circuit optimization and not simply making use of the most expensive component. An ulterior big advantage is given by the facility wich is possible to experiment unusual and effective solutions saving up time and money contemporarily.

In this article I focus the attention in the creation of a very simple but accurate vacuum diode model in the conviction that the correct realization of an audio amplifier cannot be separated from the correct realization of associate power-supply. Moreover many people are convinced that the best sonic results could be gotten just turning to the vacuum diode as rectifier element. Nevertheless the use of such devices presents some problems; in fact it is necessary to check by calculations often dull and not very careful, several parameters as the repetitive peak current, the hot-switching current, the maximum inverse voltage and so on, otherwise it is possible to lead the device to a premature death. On the contrary if you resort to the circuit simulation, on the condition that you have accurate models, you can overcome the optimization phase in the virtual world, totally, and go out only to effect the inevitable micro-adaptations and therefore build the real object.
At the end of this article I will show the goodness of the mathematical model used by comparing it with a real situation, that is comparing the results gotten by measurements on a simple power supply phisically realized, with the results gotten by the same simulated power supply.

START

the galvanometer, when connected between the collecting plate and the positive terminal of the filament, rises to several milliamperes in value. If the galvanometer is

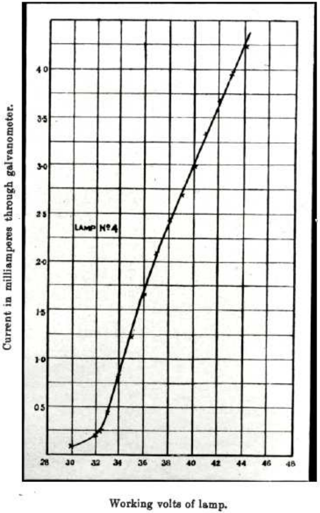

FIG. 17.

connected between the collecting plate and the negative terminal only a very small current of, perhaps, less than

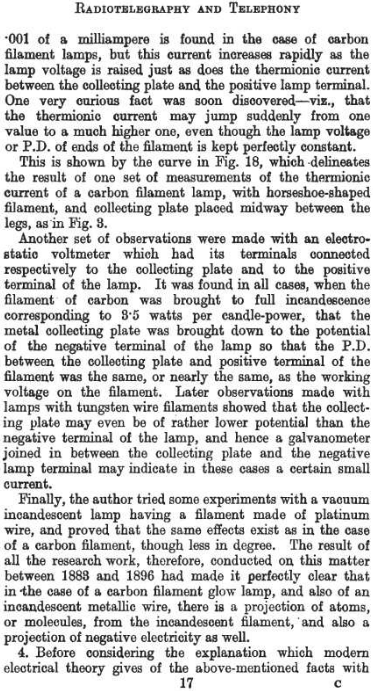

·001 of a milliampere is found in the case of carbon filament lamps, but this current increases rapidly as the lamp voltage is raised just as does the thermionic current between the collecting plate and the positive lamp terminal. One very curious fact was soon discovered—viz., that the thermionic current may jump suddenly from one value to a much higher one, even though the lamp voltage or P.D. of ends of the filament is kept perfectly constant.

This is shown by the curve in Fig. 18, which delineates the result of one set of measurements of the thermionic current of a carbon filament lamp, with horseshoe-shaped filament, and collecting plate placed midway between the legs, as in Fig. 3.

Another set of observations were made with an electrostatic voltmeter which had its terminals connected respectively to the collecting plate and to the positive terminal of the lamp. It was found in all cases, when the filament of carbon was brought to full incandescence corresponding to 3·5 watts per candle-power, that the metal collecting plate was brought down to the potential of the negative terminal of the lamp so that the P.D. between the collecting plate and positive terminal of the filament was the same, or nearly the same, as the working voltage on the filament. Later observations made with lamps with tungsten wire filaments showed that the collecting plate may even be of rather lower potential than the negative terminal of the lamp, and hence a galvanometer joined in between the collecting plate and the negative lamp terminal may indicate in these cases a certain small current.

Finally, the author tried some experiments with a vacuum incandescent lamp having a filament made of platinum wire, and proved that the same effects exist as in the case of a carbon filament, though less in degree. The result of all the research work, therefore, conducted on this matter between 1883 and 1896 had made it perfectly clear that in the case of a carbon filament glow lamp, and also of an incandescent metallic wire, there is a projection of atoms, or molecules, from the incandescent filament, and also a projection of negative electricity as well.

4. Before considering the explanation which modern electrical theory gives of the above-mentioned facts with

Photo of the test setup holding the tubes under test.

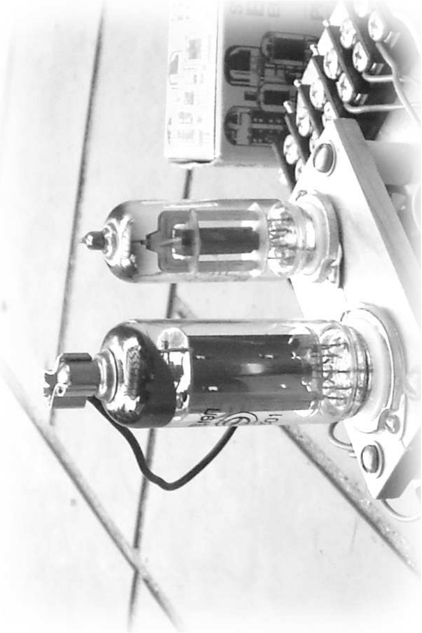

5726

6C19P

20D1

5C12P

1S3

2D1S

6G2

6DN3

regard to the incandescent lamp, we must briefly review certain earlier researches in connection with the loss of electricity from incandescent bodies.

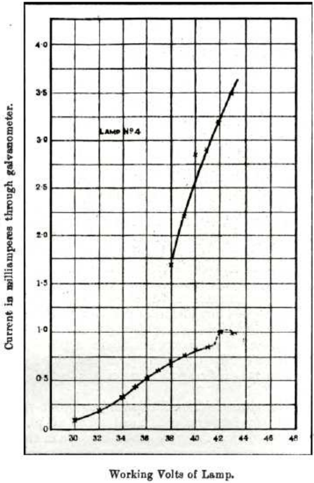

FIG. 18.

It had become known to the investigators of electrical phenomena in the middle of the eighteenth century that heat and flame cause a leakage of electricity from charged conductors.

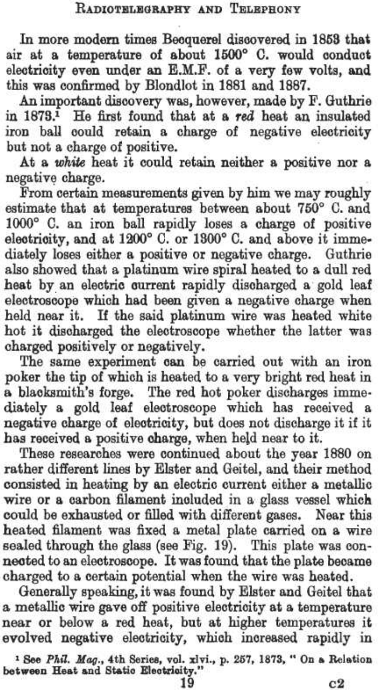

In more modern times Becquerel discovered in 1853 that air at a temperature of about 1500° C. would conduct electricity even under an E.M.F. of a very few volts, and this was confirmed by Blondlot in 1881 and 1887.

An important discovery was, however, made by F. Guthrie in 1873.[1] He first found that at a *red* heat an insulated iron ball could retain a charge of negative electricity but not a charge of positive.

At a *white* heat it could retain neither a positive nor a negative charge.

From certain measurements given by him we may roughly estimate that at temperatures between about 750° C. and 1000° C. an iron ball rapidly loses a charge of positive electricity, and at 1200° C. or 1300° C. and above it immediately loses either a positive or negative charge. Guthrie also showed that a platinum wire spiral heated to a dull red heat by an electric current rapidly discharged a gold leaf electroscope which had been given a negative charge when held near it. If the said platinum wire was heated white hot it discharged the electroscope whether the latter was charged positively or negatively.

The same experiment can be carried out with an iron poker the tip of which is heated to a very bright red heat in a blacksmith's forge. The red hot poker discharges immediately a gold leaf electroscope which has received a negative charge of electricity, but does not discharge it if it has received a positive charge, when held near to it.

These researches were continued about the year 1880 on rather different lines by Elster and Geitel, and their method consisted in heating by an electric current either a metallic wire or a carbon filament included in a glass vessel which could be exhausted or filled with different gases. Near this heated filament was fixed a metal plate carried on a wire sealed through the glass (see Fig. 19). This plate was connected to an electroscope. It was found that the plate became charged to a certain potential when the wire was heated.

Generally speaking, it was found by Elster and Geitel that a metallic wire gave off positive electricity at a temperature near or below a red heat, but at higher temperatures it evolved negative electricity, which increased rapidly in

---

[1] See *Phil. Mag.*, 4th Series, vol. xlvi., p. 257, 1873, "On a Relation between Heat and Static Electricity."

at 20mV input (2.0Vpp on signal generator), and a monotonic increase in output power at 6mV input (0.2Vpp on the generator). I was seeking a peak output power at around 4.0V on the heater so this result is disturbing. I can hear it, but cannot measure it.

Biasing the plate positive will have the effect of "pulling" the space charge towards the plate increasing the current while a negative plate bias will have the opposite effect. With 6.0V heater it will take nearly a negative half volt before the current stops. This is death to sensitivity in a crystal set. Operating the heater at 4.0V provides the most efficient rectification.

Note also that high perveance is not merely a matter of placing the plate as close as possible to the cathode. This may cause an excessive zero-bias current to overcome. The surface-area of both cathode and plate are also critical with large areas providing high current transfer. Vacuum tubes such as pentodes with an included diode plate are poor rectifiers having very low perveance. This is due to the small plate stuck in with all the other hardware in the tube. Get a "fit for purpose" diode for best results.

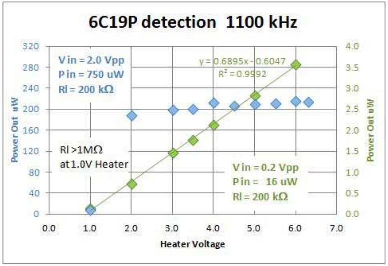

Note that while the above model as well as hearing (a kind of "hearing is beleiving" thing) supports the 4.0V heater voltage as a sweet spot for highest rectification efficiency, I have as yet not been able to measure this. Using my tube-driven Telefunken set (see description on my main radio page) I made a series of output signal power measurements at various tube heater voltages. I we see no relation between matched output power and the tube heater voltage

quantity as the temperature was raised, and therefore gave a negative charge to the insulated plate. Carbon filaments, however, were found to give off negative electricity at all temperatures.

As some references later on will be made to the published papers of Elster and Geitel we give here the titles (in English) and references to the principal papers which concern us in this subject.

(1) "On the Electricity of Flames," *Wiedemann's Annualen*, vol. xvi., p. 193 (1882).

(2) "On the Generation of Electricity by the Contact of Gases and Incandescent Bodies," *Wied. Ann.*, vol. xix., p. 588 (1883).

(3) "On the Unipolar Conductivity of Heated Gases," *Wied. Ann.*, vol. xxvi., p. 1 (1885).

This paper deals with the unsymmetrical conductivity of flames.

(4) "On the Electrification of Rarefied Gases and Electrically Heated Wires," *Wied. Ann.*, vol. xxxi., p. 109 (1887).

In this paper Elster and Geitel describe experiments made with an exhausted glass bulb having stretched across it a platinum wire which could be heated electrically. Above the wire was a metal plate carried on a platinum wire sealed through the glass (see Fig. 19). The bulb could be exhausted or filled with various gases. There is no mention of the use of a carbon filament.

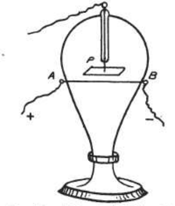

FIG. 19.—Exhausted bulb with incandescent wire, *AB*, in it, and metal plate, *P*, used by Elster and Geitel to show the emission of electricity from incandescent wires.

(5) "On the Generation of Electricity by Contact of Rarefied Gas and Electrically Heated Wires," *Wied. Ann.*, vol. xxxvii., p. 319 (1889).

In this paper (dated March, 1889) Elster and Geitel mention the unilateral conductivity for direct currents of rarefied gases in tubes with one hot and one cold electrode.

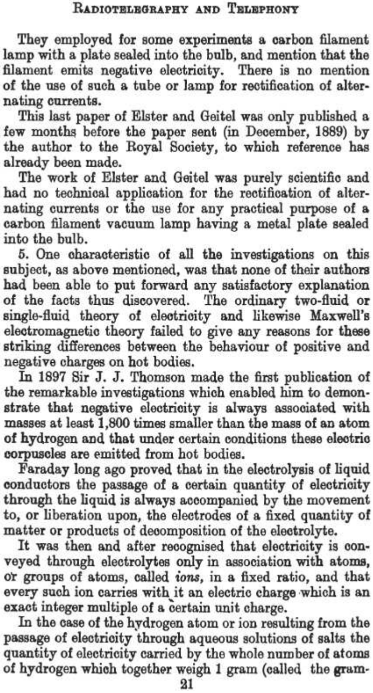

They employed for some experiments a carbon filament lamp with a plate sealed into the bulb, and mention that the filament emits negative electricity. There is no mention of the use of such a tube or lamp for rectification of alternating currents.

This last paper of Elster and Geitel was only published a few months before the paper sent (in December, 1889) by the author to the Royal Society, to which reference has already been made.

The work of Elster and Geitel was purely scientific and had no technical application for the rectification of alternating currents or the use for any practical purpose of a carbon filament vacuum lamp having a metal plate sealed into the bulb.

5. One characteristic of all the investigations on this subject, as above mentioned, was that none of their authors had been able to put forward any satisfactory explanation of the facts thus discovered. The ordinary two-fluid or single-fluid theory of electricity and likewise Maxwell's electromagnetic theory failed to give any reasons for these striking differences between the behaviour of positive and negative charges on hot bodies.

In 1897 Sir J. J. Thomson made the first publication of the remarkable investigations which enabled him to demonstrate that negative electricity is always associated with masses at least 1,800 times smaller than the mass of an atom of hydrogen and that under certain conditions these electric corpuscles are emitted from hot bodies.

Faraday long ago proved that in the electrolysis of liquid conductors the passage of a certain quantity of electricity through the liquid is always accompanied by the movement to, or liberation upon, the electrodes of a fixed quantity of matter or products of decomposition of the electrolyte.

It was then and after recognised that electricity is conveyed through electrolytes only in association with atoms, or groups of atoms, called *ions*, in a fixed ratio, and that every such ion carries with it an electric charge which is an exact integer multiple of a certain unit charge.

In the case of the hydrogen atom or ion resulting from the passage of electricity through aqueous solutions of salts the quantity of electricity carried by the whole number of atoms of hydrogen which together weigh 1 gram (called the gram-

A conceptual look at the impact of heater temperature on the space charge distribution in the annulus between cathode and anode. I start by drawing a qualitative plot of charge density versus distance from the cathode. This is not how such plots are done in the literature but it is a nice visualization. The distribution of charge is a statistical function of heater temperature and plate bias. For the zero-bias case I show qualitative distributions for three heater voltages, 3.0, 4.0, and 6.0 volts. On my chart of current/voltage I find no current flows with 3.0 volts on the heater. Current just begins to be measurable at 4.0V and at 6.0V I found a significant 0.83mA of current indicating that a good portion of the space charge is not impinging on the plate even with zero bias.

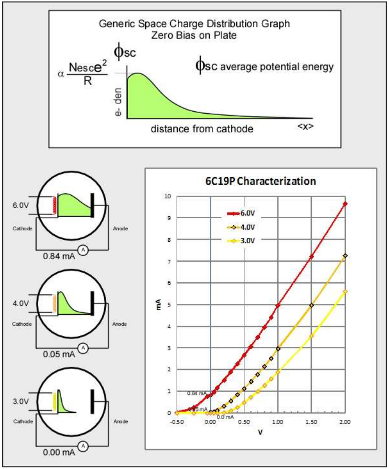

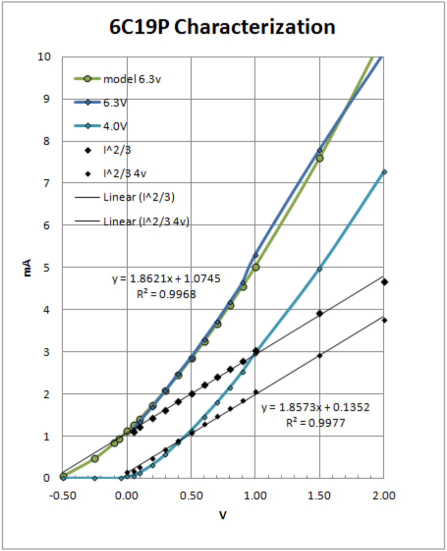

Perveance calculation for the 6C19P diode taken at 6.3 v and 4.0 v heater voltages.

atom of hydrogen) is 96,500 coulombs, or very close to it. Hence, since one coulomb is equal to one-tenth of an electromagnetic unit (E.M.U.) of electrical quantity, we can say that for the hydrogen ion or atom, the weight of which is taken as unity, the ratio of charge (*e*) to mass (*m*) is 9,650.

According to Faraday's Law of Electrolysis the same quantity of electricity passed through various electrolytic cells will liberate on the electrodes masses of ions which are in the ratio of the chemical equivalents of these substances.

Exhaustive experiments have shown that a current of one ampere flowing for one second—that is, one coulomb—deposits in a silver voltameter 0·001118 gram of silver on the cathode. The atomic weight of silver is 108 and its valency is unity. Hence 108 is its chemical equivalent, that of hydrogen being unity. Accordingly a current of one ampere liberates from a water voltameter in one second a mass of hydrogen equal to 0·00001035 gram. In other words, 96,500 coulombs, or 9,650 E.M.U. of electric quantity, liberates one gram of hydrogen and 108 grams of silver at the cathodes of electrolytic cells.

Now 3×10^{10} units of electric quantity in electrostatic measure (E.S.U.) are equivalent to the electromagnetic unit. Accordingly $2895 \times 10^{11} = 3 \times 10^{10} \times 9{,}650$ E.S.U. will liberate at 0° C. and 760 mm. pressure 11,200 cubic centimetres of hydrogen gas which weighs one gram.

Modern researches on the kinetic theory of gases have shown that in one cubic centimetre of any gas at 0° C. and 760 mm. there are most probably $2{\cdot}7 \times 10^{19}$ molecules. As the molecule of hydrogen contains two atoms there are $11{,}200 \times 5{\cdot}4 \times 10^{19}$ atoms of H. liberated by the passage of $2{,}895 \times 10^{11}$ electrostatic units of electricity. Hence each hydrogen ion or atom must be associated with

$$\frac{9{,}650 \times 3 \times 10^{10}}{11{,}200 \times 5{\cdot}4 \times 10^{19}} = 4{\cdot}77 \times 10^{-10}$$

electrostatic units (E.S.U.) of electric quantity.

This charge of the hydrogen ion is denoted by *e* and is a negative charge. Hence we have

$$e = 4{\cdot}77 \times 10^{-10} \text{ electrostatic units.}$$
$$= 1{\cdot}59 \times 10^{-20} \text{ electromagnetic units.}$$

This very small quantity of electricity is Nature's unit of electricity and it is the quantity of negative electricity

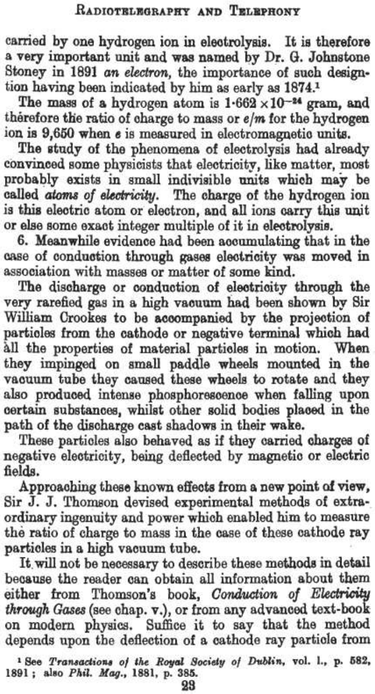

carried by one hydrogen ion in electrolysis. It is therefore a very important unit and was named by Dr. G. Johnstone Stoney in 1891 *an electron*, the importance of such designation having been indicated by him as early as 1874.[1]

The mass of a hydrogen atom is $1{\cdot}662\times10^{-24}$ gram, and therefore the ratio of charge to mass or *e*/*m* for the hydrogen ion is 9,650 when *e* is measured in electromagnetic units.

The study of the phenomena of electrolysis had already convinced some physicists that electricity, like matter, most probably exists in small indivisible units which may be called *atoms of electricity*. The charge of the hydrogen ion is this electric atom or electron, and all ions carry this unit or else some exact integer multiple of it in electrolysis.

6. Meanwhile evidence had been accumulating that in the case of conduction through gases electricity was moved in association with masses or matter of some kind.

The discharge or conduction of electricity through the very rarefied gas in a high vacuum had been shown by Sir William Crookes to be accompanied by the projection of particles from the cathode or negative terminal which had all the properties of material particles in motion. When they impinged on small paddle wheels mounted in the vacuum tube they caused these wheels to rotate and they also produced intense phosphorescence when falling upon certain substances, whilst other solid bodies placed in the path of the discharge cast shadows in their wake.

These particles also behaved as if they carried charges of negative electricity, being deflected by magnetic or electric fields.

Approaching these known effects from a new point of view, Sir J. J. Thomson devised experimental methods of extraordinary ingenuity and power which enabled him to measure the ratio of charge to mass in the case of these cathode ray particles in a high vacuum tube.

It will not be necessary to describe these methods in detail because the reader can obtain all information about them either from Thomson's book, *Conduction of Electricity through Gases* (see chap. v.), or from any advanced text-book on modern physics. Suffice it to say that the method depends upon the deflection of a cathode ray particle from

---

[1] See *Transactions of the Royal Society of Dublin*, vol. l., p. 582, 1891; also *Phil. Mag.*, 1881, p. 385.

6C19P measurements and 6.3 and 4.0 v heater voltage:

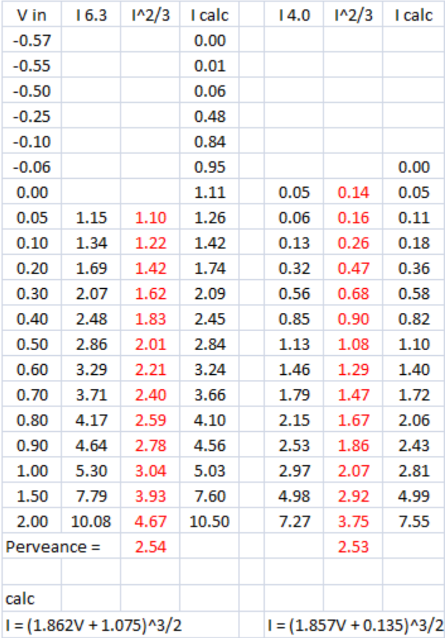

V in	I 6.3	I^2/3	I calc		I 4.0	I^2/3	I calc
-0.57			0.00				
-0.55			0.01				
-0.50			0.06				
-0.25			0.48				
-0.10			0.84				
-0.06			0.95				0.00
0.00			1.11		0.05	0.14	0.05
0.05	1.15	1.10	1.26		0.06	0.16	0.11
0.10	1.34	1.22	1.42		0.13	0.26	0.18
0.20	1.69	1.42	1.74		0.32	0.47	0.36
0.30	2.07	1.62	2.09		0.56	0.68	0.58
0.40	2.48	1.83	2.45		0.85	0.90	0.82
0.50	2.86	2.01	2.84		1.13	1.08	1.10
0.60	3.29	2.21	3.24		1.46	1.29	1.40
0.70	3.71	2.40	3.66		1.79	1.47	1.72
0.80	4.17	2.59	4.10		2.15	1.67	2.06
0.90	4.64	2.78	4.56		2.53	1.86	2.43
1.00	5.30	3.04	5.03		2.97	2.07	2.81
1.50	7.79	3.93	7.60		4.98	2.92	4.99
2.00	10.08	4.67	10.50		7.27	3.75	7.55
Perveance =		2.54				2.53	
calc							
I = (1.862V + 1.075)^3/2					I = (1.857V + 0.135)^3/2		

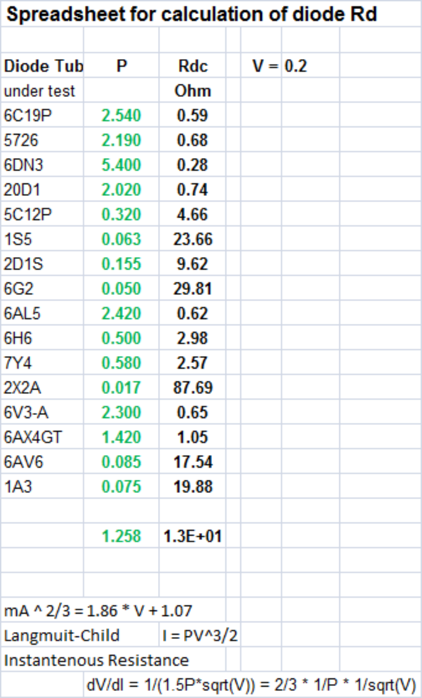

Spreadsheet for calculation of diode Rd

Diode Tub under test	P	Rdc Ohm		V = 0.2
6C19P	2.540	0.59		
5726	2.190	0.68		
6DN3	5.400	0.28		
20D1	2.020	0.74		
5C12P	0.320	4.66		
1S5	0.063	23.66		
2D1S	0.155	9.62		
6G2	0.050	29.81		
6AL5	2.420	0.62		
6H6	0.500	2.98		
7Y4	0.580	2.57		
2X2A	0.017	87.69		
6V3-A	2.300	0.65		
6AX4GT	1.420	1.05		
6AV6	0.085	17.54		
1A3	0.075	19.88		
	1.258	1.3E+01		
mA ^ 2/3 = 1.86 * V + 1.07				
Langmuit-Child		I = PV^3/2		
Instantenous Resistance				
	dV/dI = 1/(1.5P*sqrt(V)) = 2/3 * 1/P * 1/sqrt(V)			

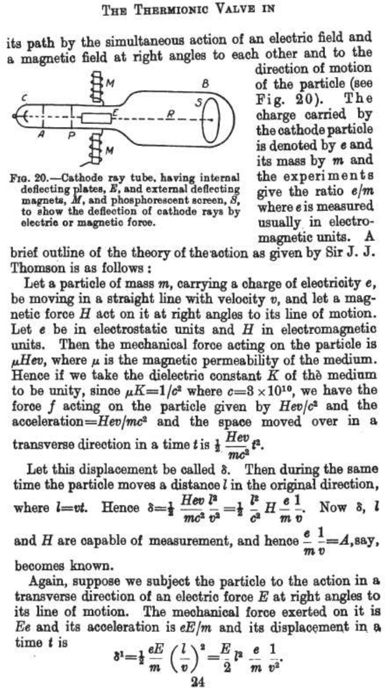

its path by the simultaneous action of an electric field and a magnetic field at right angles to each other and to the direction of motion of the particle (see Fig. 20). The charge carried by the cathode particle is denoted by e and its mass by m and the experiments give the ratio e/m where e is measured usually in electromagnetic units. A brief outline of the theory of the action as given by Sir J. J. Thomson is as follows:

FIG. 20.—Cathode ray tube, having internal deflecting plates, E, and external deflecting magnets, M, and phosphorescent screen, S, to show the deflection of cathode rays by electric or magnetic force.

Let a particle of mass m, carrying a charge of electricity e, be moving in a straight line with velocity v, and let a magnetic force H act on it at right angles to its line of motion. Let e be in electrostatic units and H in electromagnetic units. Then the mechanical force acting on the particle is μHev, where μ is the magnetic permeability of the medium. Hence if we take the dielectric constant K of the medium to be unity, since $\mu K=1/c^2$ where $c=3\times10^{10}$, we have the force f acting on the particle given by Hev/c^2 and the acceleration$=Hev/mc^2$ and the space moved over in a transverse direction in a time t is $\frac{1}{2}\frac{Hev}{mc^2}t^2$.

Let this displacement be called δ. Then during the same time the particle moves a distance l in the original direction, where $l=vt$. Hence $\delta=\frac{1}{2}\frac{Hev}{mc^2}\frac{l^2}{v^2}=\frac{1}{2}\frac{l^2}{c^2}H\frac{e}{m}\frac{1}{v}$. Now δ, l and H are capable of measurement, and hence $\frac{e}{m}\frac{1}{v}=A$, say, becomes known.

Again, suppose we subject the particle to the action in a transverse direction of an electric force E at right angles to its line of motion. The mechanical force exerted on it is Ee and its acceleration is eE/m and its displacement in a time t is

$$\delta^1=\frac{1}{2}\frac{eE}{m}\left(\frac{l}{v}\right)^2=\frac{E}{2}l^2\frac{e}{m}\frac{1}{v^2}.$$

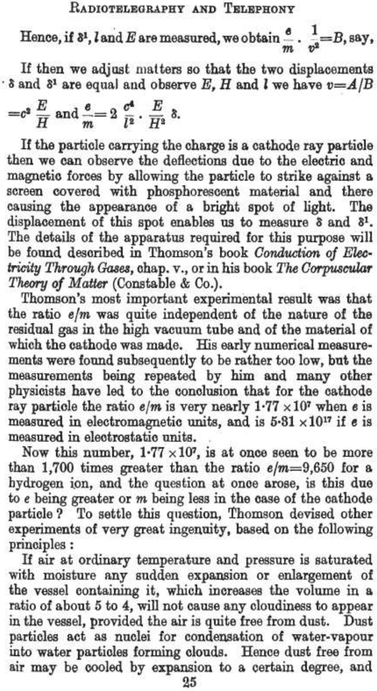

RADIOTELEGRAPHY AND TELEPHONY

Hence, if δ^1, l and E are measured, we obtain $\frac{e}{m} \cdot \frac{1}{v^2} = B$, say,

If then we adjust matters so that the two displacements δ and δ^1 are equal and observe E, H and l we have $v = A/B = c^2 \frac{E}{H}$ and $\frac{e}{m} = 2 \frac{c^4}{l^2} \cdot \frac{E}{H^2} \delta$.

If the particle carrying the charge is a cathode ray particle then we can observe the deflections due to the electric and magnetic forces by allowing the particle to strike against a screen covered with phosphorescent material and there causing the appearance of a bright spot of light. The displacement of this spot enables us to measure δ and δ^1. The details of the apparatus required for this purpose will be found described in Thomson's book *Conduction of Electricity Through Gases*, chap. v., or in his book *The Corpuscular Theory of Matter* (Constable & Co.).

Thomson's most important experimental result was that the ratio e/m was quite independent of the nature of the residual gas in the high vacuum tube and of the material of which the cathode was made. His early numerical measurements were found subsequently to be rather too low, but the measurements being repeated by him and many other physicists have led to the conclusion that for the cathode ray particle the ratio e/m is very nearly $1{\cdot}77 \times 10^7$ when e is measured in electromagnetic units, and is $5{\cdot}31 \times 10^{17}$ if e is measured in electrostatic units.

Now this number, $1{\cdot}77 \times 10^7$, is at once seen to be more than 1,700 times greater than the ratio $e/m = 9{,}650$ for a hydrogen ion, and the question at once arose, is this due to e being greater or m being less in the case of the cathode particle? To settle this question, Thomson devised other experiments of very great ingenuity, based on the following principles:

If air at ordinary temperature and pressure is saturated with moisture any sudden expansion or enlargement of the vessel containing it, which increases the volume in a ratio of about 5 to 4, will not cause any cloudiness to appear in the vessel, provided the air is quite free from dust. Dust particles act as nuclei for condensation of water-vapour into water particles forming clouds. Hence dust free from air may be cooled by expansion to a certain degree, and

25

Here you see the impact of lower operating voltages. With the tube operating between 3 and 4 volts the characteristic curves begin to pass through the origin as in regular solid-state diodes and crystals. It is precisely in this voltage range that I found the radio sensitivity, as measured by loudness to my ear, to be most pronounced. It appears that the diode characteristic needs to pass through the origin of the I-V graph for highest sensitivity. This takes me to the 20D1 tube which operates at 9v but only 200mA. This will still take a power supply to use, but the good thing is that the characteristic curve passes through the origin. Running at its design parameters the tube looks a lot like the 6C19P at 4v.

Note on Perveance

Just a few charts and spreadsheets on my calculations for diode valve perveance. I was very impressed at just how accurately the Langmuir-Child relation was able to predict the modeled results. This enabled me to reduce the observed data to a simple equation that shows the characteristic of the diode over a wide range of voltage values.

My attempts at calculating the diode resistance proves more troublesome. I can easily calculate the diode DC resistance which is low over the operating ranges. The AC resistance will wait until I know more. Stay tuned!

6AX4GT	1.42
6AV6	0.085
1A3	0.075

The high-perveance diode tube types I tested show an interesting property in that their characteristic curves do not pass through the origin of the graph. These curves are wholly acceptable as crystal radio detectors although they will require power to operate. What I have noticed with such tubes is that when I turn them off after using them with a radio, on cooling they go through a period of much increased sensitivity (very loud) before fading to nothing. It is as though running them at the full 6.3V lowers their full potential as crystal radio rectifiers. The reason is in the high perveance of the tubes, the anode is already proximate to the space charge before any plate voltage is applied. Effectively, while the plate (anode) "Zero" voltage is measured with respect to ground, the plate itself is still positive with respect to the cathodic space charge. Small currents will continue to flow even with a negative bias to the plate, (Contact potential).

In order to explore this idea, I ran a series of tests with the tubes running at lower operating voltages (effectively diminishing the space charge thus lowering the perveance) as shown:

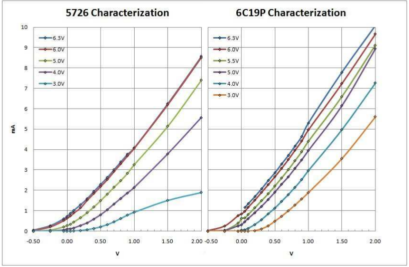

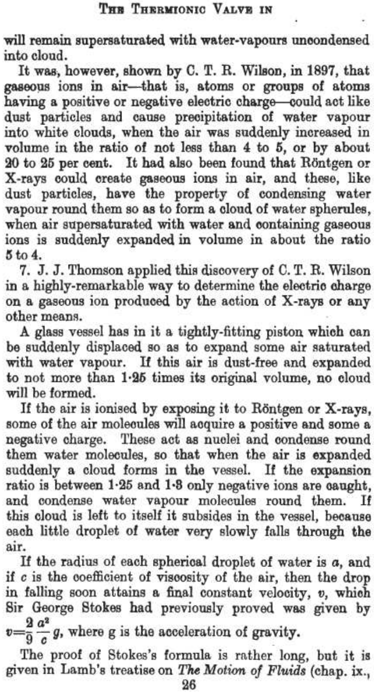

will remain supersaturated with water-vapours uncondensed into cloud.

It was, however, shown by C. T. R. Wilson, in 1897, that gaseous ions in air—that is, atoms or groups of atoms having a positive or negative electric charge—could act like dust particles and cause precipitation of water vapour into white clouds, when the air was suddenly increased in volume in the ratio of not less than 4 to 5, or by about 20 to 25 per cent. It had also been found that Röntgen or X-rays could create gaseous ions in air, and these, like dust particles, have the property of condensing water vapour round them so as to form a cloud of water spherules, when air supersaturated with water and containing gaseous ions is suddenly expanded in volume in about the ratio 5 to 4.

7. J. J. Thomson applied this discovery of C. T. R. Wilson in a highly-remarkable way to determine the electric charge on a gaseous ion produced by the action of X-rays or any other means.

A glass vessel has in it a tightly-fitting piston which can be suddenly displaced so as to expand some air saturated with water vapour. If this air is dust-free and expanded to not more than 1·25 times its original volume, no cloud will be formed.

If the air is ionised by exposing it to Röntgen or X-rays, some of the air molecules will acquire a positive and some a negative charge. These act as nuclei and condense round them water molecules, so that when the air is expanded suddenly a cloud forms in the vessel. If the expansion ratio is between 1·25 and 1·3 only negative ions are caught, and condense water vapour molecules round them. If this cloud is left to itself it subsides in the vessel, because each little droplet of water very slowly falls through the air.

If the radius of each spherical droplet of water is a, and if c is the coefficient of viscosity of the air, then the drop in falling soon attains a final constant velocity, v, which Sir George Stokes had previously proved was given by $v=\frac{2}{9}\frac{a^2}{c}g$, where g is the acceleration of gravity.

The proof of Stokes's formula is rather long, but it is given in Lamb's treatise on *The Motion of Fluids* (chap. ix.,

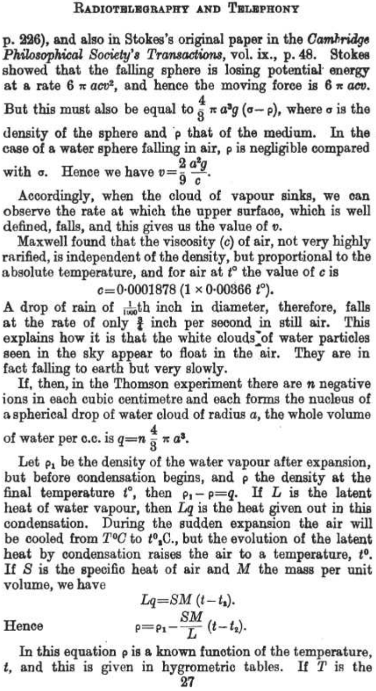

p. 226), and also in Stokes's original paper in the *Cambridge Philosophical Society's Transactions*, vol. ix., p. 48. Stokes showed that the falling sphere is losing potential energy at a rate $6\pi acv^2$, and hence the moving force is $6\pi acv$. But this must also be equal to $\frac{4}{3}\pi a^3 g(\sigma-\rho)$, where σ is the density of the sphere and ρ that of the medium. In the case of a water sphere falling in air, ρ is negligible compared with σ. Hence we have $v=\frac{2}{9}\frac{a^2 g}{c}$.

Accordingly, when the cloud of vapour sinks, we can observe the rate at which the upper surface, which is well defined, falls, and this gives us the value of v.

Maxwell found that the viscosity (c) of air, not very highly rarified, is independent of the density, but proportional to the absolute temperature, and for air at t° the value of c is

$$c=0{\cdot}0001878\,(1\times 0{\cdot}00366\,t^\circ).$$

A drop of rain of $\frac{1}{1000}$th inch in diameter, therefore, falls at the rate of only $\frac{3}{4}$ inch per second in still air. This explains how it is that the white clouds of water particles seen in the sky appear to float in the air. They are in fact falling to earth but very slowly.

If, then, in the Thomson experiment there are n negative ions in each cubic centimetre and each forms the nucleus of a spherical drop of water cloud of radius a, the whole volume of water per c.c. is $q=n\frac{4}{3}\pi a^3$.

Let ρ_1 be the density of the water vapour after expansion, but before condensation begins, and ρ the density at the final temperature t°, then $\rho_1-\rho=q$. If L is the latent heat of water vapour, then Lq is the heat given out in this condensation. During the sudden expansion the air will be cooled from $T^\circ C$ to $t^\circ{}_2$C., but the evolution of the latent heat by condensation raises the air to a temperature, t°. If S is the specific heat of air and M the mass per unit volume, we have

$$Lq=SM\,(t-t_2).$$

Hence

$$\rho=\rho_1-\frac{SM}{L}\,(t-t_2).$$

In this equation ρ is a known function of the temperature, t, and this is given in hygrometric tables. If T is the

A question remains as to what in fact is different about these tubes to give such different characteristics. While the I/V characteristic of most vacuum diodes pretty much follow the Langmuir-Child Law, the steepness of the I/V characteristic is largely due to the geometry of the tube elements and the volume of electron space-charge between them. JB Calvert's Theory of Vacuum Tubes informs one that this is the property called Perveance in a vacuum diode. Measuring this requires plotting I raised to the 2/3 power against V and taking the slope of the best-fit line through the data. This slope raised to the 3/2 power is the perveance. With my measurements this was practical and quickly done, results as follows:

Tube	Perveance	R2	
6C19P	2.54	0.997	
5726	2.19	0.998	
6DN3	5.40	0.997	
20D1	2.02	0.985	(Langmuir-Child?)
5C12P	0.32	0.999	
1S5	0.063	0.999	
2D1S	0.155	0.975	(Langmuir-Child?)
6G2	0.050	0.999	

The perveance numbers above range from 0.02 to over 2 mA/V^3/2 (with >2 being good for small-signal detection) and clearly shows my choice of the 6C19P to be an excellent one. I am surprised to see a huge 5.4 for the 6DN3 diode, a color television damping diode. It takes a Novar 9-pin socket and is a large tube so doing more with this tube may have to wait a bit.

Calvert's 2001 published data

Tube	Perveance
6AL5	2.42
6H6	0.50
7Y4	0.58
2X2A	0.017
6V3-A	2.3

comparison. I was hoping to find a good candidate with sensitive characteristics and low energy consumption. As you will find, such a beast did not exist. Following this research began a period of purchase and testing.

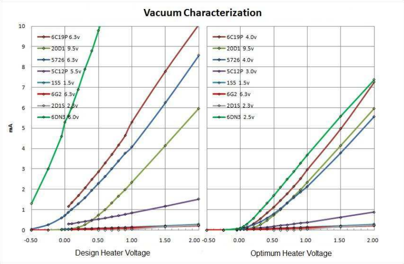

I have finally found, purchased and tested eight different tube types (plot above) including 6.3V rectifier diodes that take a lot of power to run and are not really suitable for battery use, a 9v dual diode tube (20D1), a pentode/diode tube (1S5) that is designed to run on a battery at 1.5V and 50mA, and a couple miscellaneous but cute tubes. In testing the 1S5, I found the filament never glowed incandescent at 1.5V and, on measuring, was barely sensitive to anything. I cannot imagine this tube would make much of a diode for crystal radio use. I tested different manufacture 1S5 tubes from two different suppliers. No dice. Finally, I tried to push the tube to operate at higher-than-specified voltages. At 4.3V the I-V curve was still very flat, but shifted slightly higher, crossing the Amp axis at about 0.4 mA, (see graph). At 5.5V one of my tubes gave up the ghost and I didnt do more. I pretty much rule the 1S5 out for my crystal radio work. Looks like i'll be needing the power supply when I get around to building/operating that Fleming Radio.

temperature at the beginning of the experiment, and P the pressure of the gas in millimetres of mercury, then since 1 c.c. of air at 0^0C and 760 m.m. weighs 0·00129 gram, we have

$$M=\frac{\cdot00129}{x}\cdot\frac{273}{273+T}\cdot\frac{P}{760},$$

where x is the ratio of final to the initial volume of the gas.

Also, $$\rho_1=\frac{\rho^1}{x}$$

where ρ^1 is the density of water vapour at T^0C before expansion. This can be obtained from the hygrometric tables.

Now, in adiabatic or sudden expansion of air the relation between volume v and pressure p is given by the equation $pv^{1\cdot41}=a$ *constant*; whilst for slow expansion we have $pv=R\theta$, where R is the so-called gas constant, and θ is the absolute temperature. Therefore, for adiabatic expansion,

$$v^{0\cdot41}\times\theta=a\ \textit{constant}$$

or $$\log_e\frac{\theta_1}{\theta_2}=0\cdot41\log\frac{v_1}{v_2}$$

or $$\log_e\frac{273+T}{273+t_2}=0\cdot41\log x.$$

From this equation we can find t_2, when x and T are given. Then we have:

$$\rho=\frac{\rho^1}{x}-\frac{SM}{L}(t-t_2)$$

and we can substitute the value of t_2 and x from the previous equation. This last equation can be solved by substituting various values for t, and finding the corresponding value of ρ, until we reach values of t and ρ, which agree with the corresponding values in the hygrometric tables.

Having then found ρ and ρ_1 we know q, and also having found a from Stokes formula we find $n=3q/4\pi a^3$, which gives us the number of drops and therefore of negative ions in the cloud.

By this very ingenious method Thomson weighed the cloud of water spherules and counted the drops of water forming the cloud and therefore determined the number of ions round which the drops were formed. He had then to

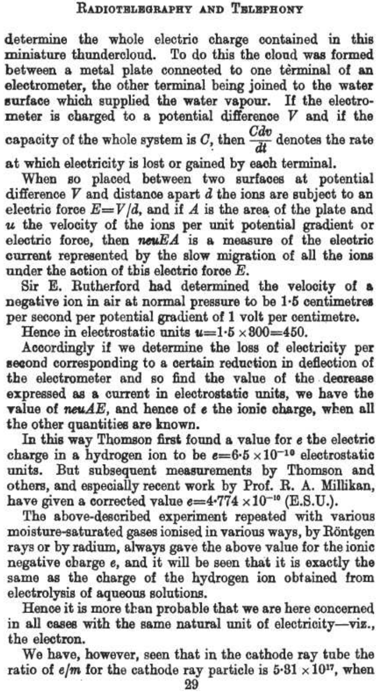

determine the whole electric charge contained in this miniature thundercloud. To do this the cloud was formed between a metal plate connected to one terminal of an electrometer, the other terminal being joined to the water surface which supplied the water vapour. If the electrometer is charged to a potential difference V and if the capacity of the whole system is C, then $\frac{Cdv}{dt}$ denotes the rate at which electricity is lost or gained by each terminal.

When so placed between two surfaces at potential difference V and distance apart d the ions are subject to an electric force $E=V/d$, and if A is the area of the plate and u the velocity of the ions per unit potential gradient or electric force, then $neuEA$ is a measure of the electric current represented by the slow migration of all the ions under the action of this electric force E.

Sir E. Rutherford had determined the velocity of a negative ion in air at normal pressure to be 1·5 centimetres per second per potential gradient of 1 volt per centimetre.

Hence in electrostatic units $u=1{\cdot}5\times 300=450$.

Accordingly if we determine the loss of electricity per second corresponding to a certain reduction in deflection of the electrometer and so find the value of the decrease expressed as a current in electrostatic units, we have the value of $neuAE$, and hence of e the ionic charge, when all the other quantities are known.

In this way Thomson first found a value for e the electric charge in a hydrogen ion to be $e=6{\cdot}5\times 10^{-10}$ electrostatic units. But subsequent measurements by Thomson and others, and especially recent work by Prof. R. A. Millikan, have given a corrected value $e=4{\cdot}774\times 10^{-10}$ (E.S.U.).

The above-described experiment repeated with various moisture-saturated gases ionised in various ways, by Röntgen rays or by radium, always gave the above value for the ionic negative charge e, and it will be seen that it is exactly the same as the charge of the hydrogen ion obtained from electrolysis of aqueous solutions.

Hence it is more than probable that we are here concerned in all cases with the same natural unit of electricity—viz., the electron.

We have, however, seen that in the cathode ray tube the ratio of e/m for the cathode ray particle is $5{\cdot}31\times 10^{17}$, when

Vacuum Diodes

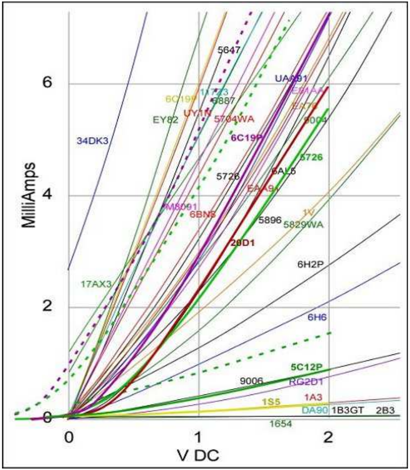

When I decided early on to construct a radio based on the vacuum diode (see my Fleming Radio section), I had to find a suitable tube for the project. That began with an extensive search through the datasheets checking vacuum diode properties and reviewing the characteristic curves for a large number of diode tubes. Datasheet characteristic curves, if they exist at all, are notoriously difficult to make comparisons between. In the above figure I plot as best I can the V/I characteristics at low signal levels on a common plot for units with the same features. Still, the supply regulator was never designed to adjust the output to precise small values making it somewhat difficult/tedious to adjust. Adding a small rheostat to the test jig may aid for making fine adjustments on the voltage. Another piece of equipment you will need for this test is an amp meter. I use a nice small digital meter I found new on ebay for $15 shipping included. It reads amps to a lowest range of 200uA, more than enough precision for the task. For that price get two. The DVM is, of course, my trusty Keithley workhorse. Easy protocol as follows:

1) Set up the test with the diode and meters clipped to the jig.

2) Adjust the power supply until the reading on the DVM meter is where you plan to measure. Adjust the voltage (DVM readout) in steps from 0 to 1.0 Volt in 0.1 V increments and record the current (Amps) in a spreadsheet.

3) Repeat at each tenth volt recording while the spreadsheet makes a sweet graph.

Test Results

First, I offer a quick look at my diode modeling setup. I am currently upping the ante on resolution in order to determine diode parameters Is and n, hoping for success soon! Setup and protocol..

Over a period of a few weeks I ran I-V characterization tests on a good number and variety of diodes, crystals and my two tubes to see how very thing stacks up. The following presents the resulting curves, diode photos, and some puzzling questions/conclusions I churned up in the process. I start off with the realization that when one orders diodes, it is by no means certain what one will get. This seems especially true for the supposedly ubiquitous 1N34A. Unless you see the part number actually on the diode, you probably need to test it to know what it really is. So, lets take a look!

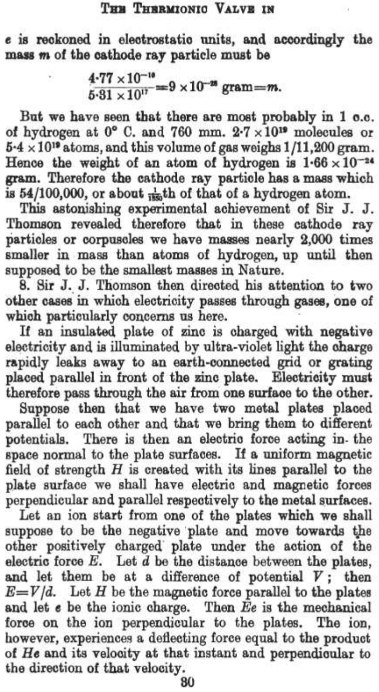

e is reckoned in electrostatic units, and accordingly the mass *m* of the cathode ray particle must be

$$\frac{4\cdot77\times10^{-10}}{5\cdot31\times10^{17}}=9\times10^{-28}\text{ gram}=m.$$

But we have seen that there are most probably in 1 c.c. of hydrogen at 0° C. and 760 mm. $2\cdot7\times10^{19}$ molecules or $5\cdot4\times10^{19}$ atoms, and this volume of gas weighs 1/11,200 gram. Hence the weight of an atom of hydrogen is $1\cdot66\times10^{-24}$ gram. Therefore the cathode ray particle has a mass which is 54/100,000, or about $\frac{1}{1850}$th of that of a hydrogen atom.

This astonishing experimental achievement of Sir J. J. Thomson revealed therefore that in these cathode ray particles or corpuscles we have masses nearly 2,000 times smaller in mass than atoms of hydrogen, up until then supposed to be the smallest masses in Nature.

8. Sir J. J. Thomson then directed his attention to two other cases in which electricity passes through gases, one of which particularly concerns us here.

If an insulated plate of zinc is charged with negative electricity and is illuminated by ultra-violet light the charge rapidly leaks away to an earth-connected grid or grating placed parallel in front of the zinc plate. Electricity must therefore pass through the air from one surface to the other.

Suppose then that we have two metal plates placed parallel to each other and that we bring them to different potentials. There is then an electric force acting in the space normal to the plate surfaces. If a uniform magnetic field of strength H is created with its lines parallel to the plate surface we shall have electric and magnetic forces perpendicular and parallel respectively to the metal surfaces.

Let an ion start from one of the plates which we shall suppose to be the negative plate and move towards the other positively charged plate under the action of the electric force E. Let d be the distance between the plates, and let them be at a difference of potential V; then $E=V/d$. Let H be the magnetic force parallel to the plates and let e be the ionic charge. Then Ee is the mechanical force on the ion perpendicular to the plates. The ion, however, experiences a deflecting force equal to the product of He and its velocity at that instant and perpendicular to the direction of that velocity.

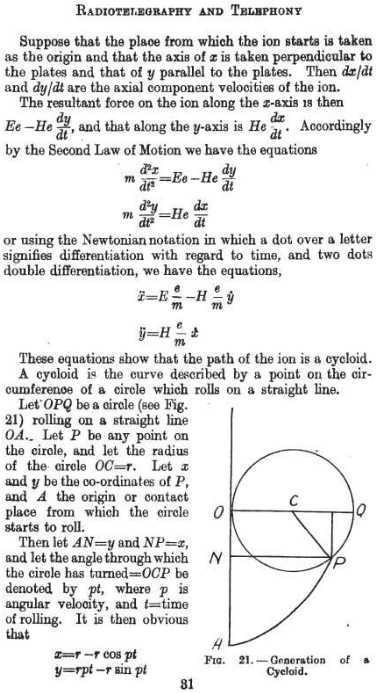

Suppose that the place from which the ion starts is taken as the origin and that the axis of x is taken perpendicular to the plates and that of y parallel to the plates. Then dx/dt and dy/dt are the axial component velocities of the ion.

The resultant force on the ion along the x-axis is then $Ee - He\frac{dy}{dt}$, and that along the y-axis is $He\frac{dx}{dt}$. Accordingly by the Second Law of Motion we have the equations

$$m\frac{d^2x}{dt^2}=Ee-He\frac{dy}{dt}$$

$$m\frac{d^2y}{dt^2}=He\frac{dx}{dt}$$

or using the Newtonian notation in which a dot over a letter signifies differentiation with regard to time, and two dots double differentiation, we have the equations,

$$\ddot{x}=E\frac{e}{m}-H\frac{e}{m}\dot{y}$$

$$\ddot{y}=H\frac{e}{m}\dot{x}$$

These equations show that the path of the ion is a cycloid.

A cycloid is the curve described by a point on the circumference of a circle which rolls on a straight line.

Let OPQ be a circle (see Fig. 21) rolling on a straight line OA. Let P be any point on the circle, and let the radius of the circle $OC=r$. Let x and y be the co-ordinates of P, and A the origin or contact place from which the circle starts to roll.

Then let $AN=y$ and $NP=x$, and let the angle through which the circle has turned $=OCP$ be denoted by pt, where p is angular velocity, and t=time of rolling. It is then obvious that

$$x=r-r\cos pt$$
$$y=rpt-r\sin pt$$

FIG. 21.—Generation of a Cycloid.

Diode Testing

Kevin Smith
http://www.lessmiths.com/~kjsmith/crystal/dtest.shtml

So, what is the best diode out there to use? Do crystals always show poor sensitivity compared to germanium? What about Fleming's marvelous device, the vacuum diode? I start by using a test jig, the design shamelessly modified from Dan Petersen in the Nov 2009 Crystal Set Society Newsletter, (if you aren't a member, go sign up to be.. NOW):

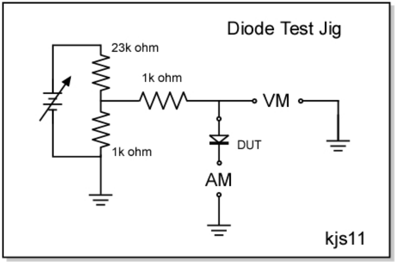

This is a very simple circuit allowing power input, attaching a diode for testing, and sampling both the curent and voltage across the diode under test, (DUT). I made a simple modification by adding a voltage divider, about 20:1 at the power input. The power supply can be connected either at the front across the divider, or after the divider in front of the 100k current limiting resistor for higher voltage readings. I get the input voltage from a modest 18V, 2A regulated DC power supply. This unit has both volt and amp controls with digital readouts, nice to use and dosnt take up much room on my desk. Here is a case where I recommend you look at getting a new unit. I looked long and hard on ebay, but found few bargins and few

Differentiate these equations twice with regard to t, and we obtain

$$\ddot{x}=rp^2-p\dot{y}, \quad \ddot{y}=p\dot{x}.$$

If we compare these two last equations with the above equations of motion of the ion, we see at once they are identical provided we put $p=H\frac{e}{m}$ and $r=\frac{E}{H^2}\frac{m}{e}$. This proves that the path of the ion must be a cycloid.

Hence, if the ion starts from one plate under the action of the two forces due to the electric and magnetic force, it will curl round in a cycloidal path. If, then, the distance between the two plates exceeds twice the radius of the circle, or $\frac{2E}{H^2}\frac{m}{e}$, the particle will never reach the second plate.

Thomson applied the above theorem to the case of the ions emitted by an incandescent carbon filament. He stretched a filament close in front of a metal plate, and placed at a certain distance, d, from it a second parallel metal plate. These plates and filaments were included in a glass vessel very highly exhausted. The filament was connected to the plate just in contact with it, and a difference of potential, V, made between the plates.

Hence the electric force $E=V/d$. Coils of wire were placed outside the tube, and through them a current was sent, so as to make a magnetic force, H, of known and adjustable value parallel to the plates. When the carbon filament was made incandescent it emits, as we have seen, negative ions, and by adjusting the magnetic field it was possible just to prevent these ions reaching the second plate. From measurements of H, V, and d it was then possible to solve the equation $d=\frac{2E}{H^2}\frac{m}{e}$, so as to obtain the value of e/m.

Thomson's experiments made by the above method in 1899 led to a value of e/m for the negative ions emitted by an incandescent carbon filament which was not very different from that obtained from experiments on the cathode ray corpuscle, but in absolute value were too low. More recent measurements by Professor O. W. Richardson gave values $e/m=1{\cdot}45\times10^7$ and $1{\cdot}49\times10^7$, and more recent work still by Bestelmeyer, who used an improved method, has given the

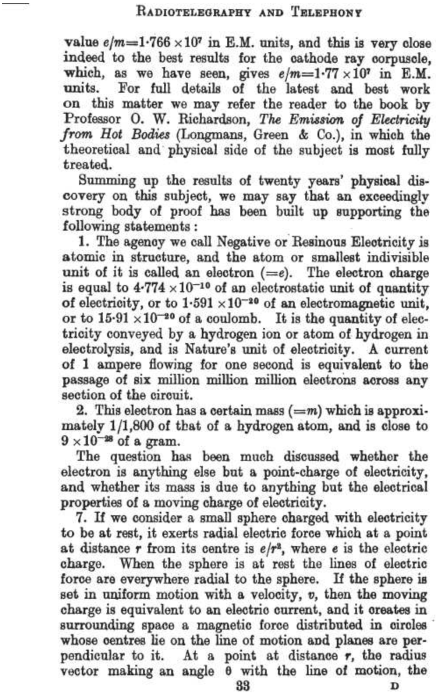

value $e/m = 1{\cdot}766 \times 10^7$ in E.M. units, and this is very close indeed to the best results for the cathode ray corpuscle, which, as we have seen, gives $e/m = 1{\cdot}77 \times 10^7$ in E.M. units. For full details of the latest and best work on this matter we may refer the reader to the book by Professor O. W. Richardson, *The Emission of Electricity from Hot Bodies* (Longmans, Green & Co.), in which the theoretical and physical side of the subject is most fully treated.

Summing up the results of twenty years' physical discovery on this subject, we may say that an exceedingly strong body of proof has been built up supporting the following statements:

1. The agency we call Negative or Resinous Electricity is atomic in structure, and the atom or smallest indivisible unit of it is called an electron ($=e$). The electron charge is equal to $4{\cdot}774 \times 10^{-10}$ of an electrostatic unit of quantity of electricity, or to $1{\cdot}591 \times 10^{-20}$ of an electromagnetic unit, or to $15{\cdot}91 \times 10^{-20}$ of a coulomb. It is the quantity of electricity conveyed by a hydrogen ion or atom of hydrogen in electrolysis, and is Nature's unit of electricity. A current of 1 ampere flowing for one second is equivalent to the passage of six million million million electrons across any section of the circuit.

2. This electron has a certain mass ($=m$) which is approximately 1/1,800 of that of a hydrogen atom, and is close to 9×10^{-28} of a gram.

The question has been much discussed whether the electron is anything else but a point-charge of electricity, and whether its mass is due to anything but the electrical properties of a moving charge of electricity.

7. If we consider a small sphere charged with electricity to be at rest, it exerts radial electric force which at a point at distance r from its centre is e/r^2, where e is the electric charge. When the sphere is at rest the lines of electric force are everywhere radial to the sphere. If the sphere is set in uniform motion with a velocity, v, then the moving charge is equivalent to an electric current, and it creates in surrounding space a magnetic force distributed in circles whose centres lie on the line of motion and planes are perpendicular to it. At a point at distance r, the radius vector making an angle θ with the line of motion, the

Article #27 shows detector measurements of how diodes having different values of Is and n perform as weak signal detectors when impedance matched at both input and out put.

#16 Published: 03/28/01; Revised: 02/10/2004

The center tap of the input transformer is not used in this circuit. The triode has grid leak bias. It is not possible to measure this bias with a DMM because of the large grid resistor, but the plate current is only about 280 μA. The diode and triode are referred to the negative end of the filament, but cathode bias is provided for the power pentode. The bias in operation is about -9.3 V, giving a plate current of about 5 mA and a screen current of about 1 mA. The output transformer can be any transformer on hand. I used the P-T31 that has been used in other places, which matches 8Ω to 5 kΩ. The load resistance should preferably be about 12 kΩ. However, it drove a good speaker at considerable volume with a rather small input. A three-tube battery radio could be made with this tube, an IF amplifier stage, and a pentagrid converter, a tube type which will be treated next.

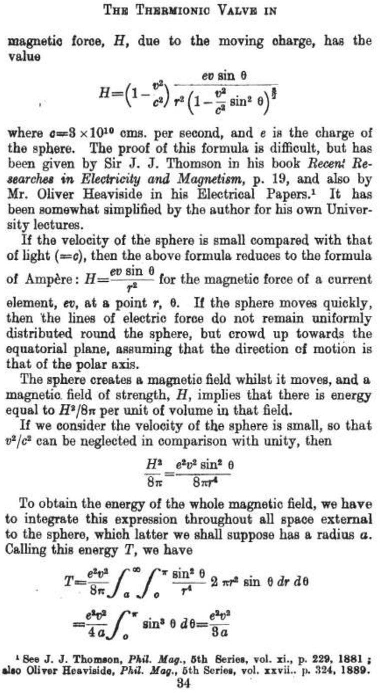

magnetic force, H, due to the moving charge, has the value

$$H=\left(1-\frac{v^2}{c^2}\right)\frac{ev\sin\theta}{r^2\left(1-\frac{v^2}{c^2}\sin^2\theta\right)^{\frac{3}{2}}}$$

where $c=3\times10^{10}$ cms. per second, and e is the charge of the sphere. The proof of this formula is difficult, but has been given by Sir J. J. Thomson in his book *Recent Researches in Electricity and Magnetism*, p. 19, and also by Mr. Oliver Heaviside in his Electrical Papers.[1] It has been somewhat simplified by the author for his own University lectures.

If the velocity of the sphere is small compared with that of light ($=c$), then the above formula reduces to the formula of Ampère: $H=\frac{ev\sin\theta}{r^2}$ for the magnetic force of a current element, ev, at a point r, θ. If the sphere moves quickly, then the lines of electric force do not remain uniformly distributed round the sphere, but crowd up towards the equatorial plane, assuming that the direction of motion is that of the polar axis.

The sphere creates a magnetic field whilst it moves, and a magnetic field of strength, H, implies that there is energy equal to $H^2/8\pi$ per unit of volume in that field.

If we consider the velocity of the sphere is small, so that v^2/c^2 can be neglected in comparison with unity, then

$$\frac{H^2}{8\pi}=\frac{e^2v^2\sin^2\theta}{8\pi r^4}$$

To obtain the energy of the whole magnetic field, we have to integrate this expression throughout all space external to the sphere, which latter we shall suppose has a radius a. Calling this energy T, we have

$$T=\frac{e^2v^2}{8\pi}\int_a^\infty\int_0^\pi\frac{\sin^2\theta}{r^4}\,2\pi r^2\sin\theta\,dr\,d\theta$$

$$=\frac{e^2v^2}{4a}\int_0^\pi\sin^3\theta\,d\theta=\frac{e^2v^2}{3a}$$

---

[1] See J. J. Thomson, *Phil. Mag.*, 5th Series, vol. xi., p. 229, 1881; also Oliver Heaviside, *Phil. Mag.*, 5th Series, vol. xxvii., p. 324, 1889.

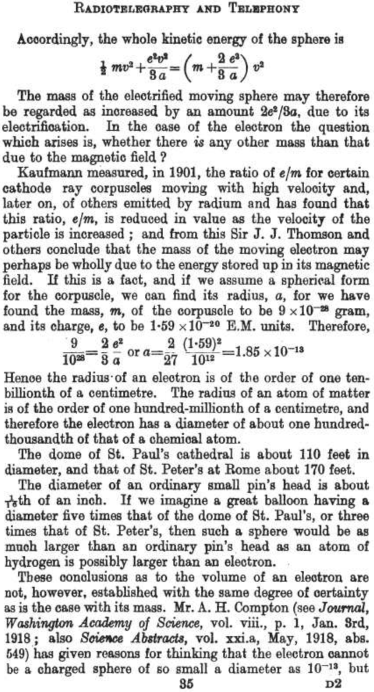

Accordingly, the whole kinetic energy of the sphere is

$$\tfrac{1}{2}\,mv^2+\frac{e^2v^2}{3\,a}=\left(m+\frac{2\,e^2}{3\,a}\right)v^2$$

The mass of the electrified moving sphere may therefore be regarded as increased by an amount $2e^2/3a$, due to its electrification. In the case of the electron the question which arises is, whether there *is* any other mass than that due to the magnetic field ?

Kaufmann measured, in 1901, the ratio of e/m for certain cathode ray corpuscles moving with high velocity and, later on, of others emitted by radium and has found that this ratio, e/m, is reduced in value as the velocity of the particle is increased ; and from this Sir J. J. Thomson and others conclude that the mass of the moving electron may perhaps be wholly due to the energy stored up in its magnetic field. If this is a fact, and if we assume a spherical form for the corpuscle, we can find its radius, a, for we have found the mass, m, of the corpuscle to be 9×10^{-28} gram, and its charge, e, to be $1{\cdot}59\times10^{-20}$ E.M. units. Therefore,

$$\frac{9}{10^{28}}=\frac{2}{3}\frac{e^2}{a}\text{ or }a=\frac{2}{27}\,\frac{(1{\cdot}59)^2}{10^{12}}=1{\cdot}85\times10^{-13}$$

Hence the radius of an electron is of the order of one ten-billionth of a centimetre. The radius of an atom of matter is of the order of one hundred-millionth of a centimetre, and therefore the electron has a diameter of about one hundred-thousandth of that of a chemical atom.

The dome of St. Paul's cathedral is about 110 feet in diameter, and that of St. Peter's at Rome about 170 feet.

The diameter of an ordinary small pin's head is about $\frac{1}{16}$th of an inch. If we imagine a great balloon having a diameter five times that of the dome of St. Paul's, or three times that of St. Peter's, then such a sphere would be as much larger than an ordinary pin's head as an atom of hydrogen is possibly larger than an electron.

These conclusions as to the volume of an electron are not, however, established with the same degree of certainty as is the case with its mass. Mr. A. H. Compton (see *Journal, Washington Academy of Science*, vol. viii., p. 1, Jan. 3rd, 1918 ; also *Science Abstracts*, vol. xxi.a, May, 1918, abs. 549) has given reasons for thinking that the electron cannot be a charged sphere of so small a diameter as 10^{-13}, but

connection. Diode bias is not suitable for high-mu triodes such as the 6AV6 or 6AT6, since they are very sensitive to the bias level. The 6R7 is a medium-mu triode (μ = 16). The 6SR7 is a single-ended equivalent, and will also work in this circuit. The signal generator should be adjusted to peak the output, at 455 kHz. I obtained an audio output of 13 V peak-to-peak with a 30% modulated AM signal of 0.2 V peak-to-peak, for a gain of about 289. The tuned circuit provides a large part of this gain.

The 6AQ6, 6AT6 and 6BF6 are all miniature dual-diode triodes with the same basing as the 6AV6, the 6BF6 being medium-mu and more appropriate for diode bias and transformer coupling. The 6SQ7, 6SR7 and 6ST7 are similar octal tubes, again all with the same basing. This was a very popular and useful tube type, with many options available.

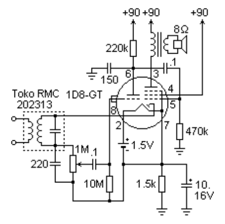

Detector - Audo Amplifier

The 1D8-GT diode-triode-pentode was designed to provide a complete audio system for a battery-powered radio. The circuit is shown at the right. The filament takes 0.1 A at 1.4 V, easily supplied by a D cell. The B+ supply is 90 V, for which batteries were available, consisting of 60 small cells. This circuit has a drain of less than 7 mA, which is quite acceptable, and hardly more than a transistor radio with the same audio output. The filament is hard to see when the tube is in operation. It can just be seen glowing orange in the diode portion of the tube. On one side is the diode with the triode on top; on the other is the pentode, all of which are clearly displayed. The tube cost $3.10, an excellent price considering all the experiments that can be done with it.

signal carrier frequency until the tuned circuit resonates; in my case, this was 500 kHz. For an input signal from the signal generator of 0.18V peak-to-peak (modulation amplitude 0.04V), I got 26V peak-to-peak at the plate of the 6AT6. The RF ripple was negligible, but there was some 60Hz pickup, not surprising with a high-gain breadboarded circuit out in the open. It was interesting to show the AM signal and the audio output simultaneously on the oscilloscope.

The circuit will function without the tuned circuit, as you can easily demonstrate. The signal generator output must then be increased, because the resonance gain is considerable. The circuit will also function with both diode plates connected to one side of the secondary, and the 1M potentiometer and filter capacitor to the other side. In this case, the ripple frequency will be halved, but the output voltage will be doubled. There is not a great advantage in using a full-wave detector. Look at the DC voltage across the potentiometer as the input amplitude is changed. This voltage, negative with respect to ground, can be used for AVC, automatic volume control, as is discussed elsewhere. With the full-wave detector, it varied from -0.5V for 10V p-p audio output to -1.0V for 30V audio output. In practice, the AVC voltage is further filtered to remove the audio.

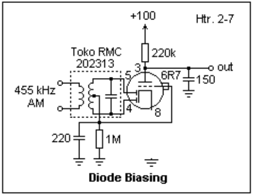

Diode Biasing

The rectified signal voltage is of the correct polarity to provide grid bias for the amplifier tube in this circuit. Therefore, the grid of the triode can be connected directly to some point on the diode load resistor that provides the proper bias, in place of the volume control potentiometer and coupling capacitor shown above. This arrangement is called *diode bias*, illustrated by the circuit at the left. AVC is essential when diode bias is used, to keep the bias voltage in the proper range. The audio signal, of course, rides on the bias

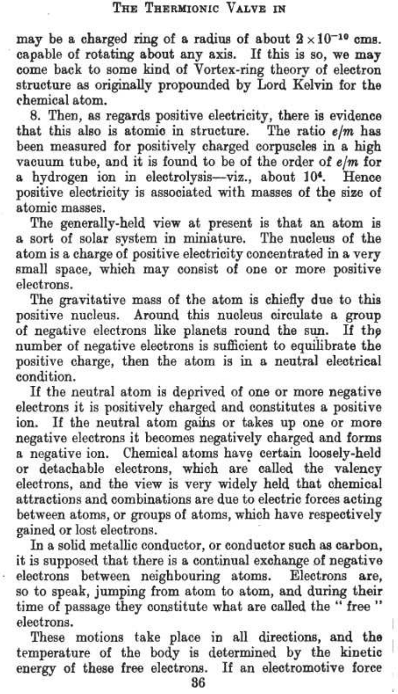

may be a charged ring of a radius of about 2×10^{-10} cms. capable of rotating about any axis. If this is so, we may come back to some kind of Vortex-ring theory of electron structure as originally propounded by Lord Kelvin for the chemical atom.

8. Then, as regards positive electricity, there is evidence that this also is atomic in structure. The ratio e/m has been measured for positively charged corpuscles in a high vacuum tube, and it is found to be of the order of e/m for a hydrogen ion in electrolysis—viz., about 10^4. Hence positive electricity is associated with masses of the size of atomic masses.

The generally-held view at present is that an atom is a sort of solar system in miniature. The nucleus of the atom is a charge of positive electricity concentrated in a very small space, which may consist of one or more positive electrons.

The gravitative mass of the atom is chiefly due to this positive nucleus. Around this nucleus circulate a group of negative electrons like planets round the sun. If the number of negative electrons is sufficient to equilibrate the positive charge, then the atom is in a neutral electrical condition.

If the neutral atom is deprived of one or more negative electrons it is positively charged and constitutes a positive ion. If the neutral atom gains or takes up one or more negative electrons it becomes negatively charged and forms a negative ion. Chemical atoms have certain loosely-held or detachable electrons, which are called the valency electrons, and the view is very widely held that chemical attractions and combinations are due to electric forces acting between atoms, or groups of atoms, which have respectively gained or lost electrons.

In a solid metallic conductor, or conductor such as carbon, it is supposed that there is a continual exchange of negative electrons between neighbouring atoms. Electrons are, so to speak, jumping from atom to atom, and during their time of passage they constitute what are called the "free" electrons.

These motions take place in all directions, and the temperature of the body is determined by the kinetic energy of these free electrons. If an electromotive force

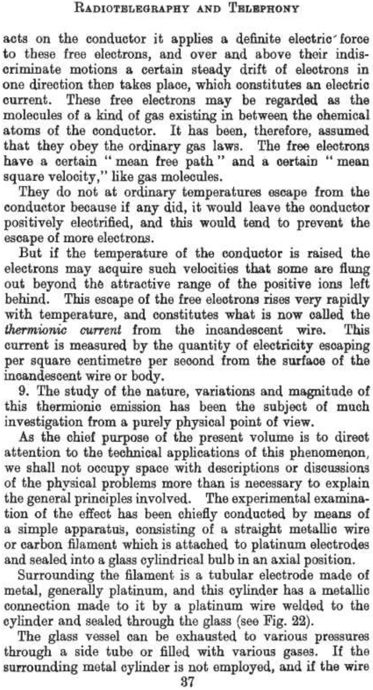

acts on the conductor it applies a definite electric force to these free electrons, and over and above their indiscriminate motions a certain steady drift of electrons in one direction then takes place, which constitutes an electric current. These free electrons may be regarded as the molecules of a kind of gas existing in between the chemical atoms of the conductor. It has been, therefore, assumed that they obey the ordinary gas laws. The free electrons have a certain "mean free path" and a certain "mean square velocity," like gas molecules.

They do not at ordinary temperatures escape from the conductor because if any did, it would leave the conductor positively electrified, and this would tend to prevent the escape of more electrons.

But if the temperature of the conductor is raised the electrons may acquire such velocities that some are flung out beyond the attractive range of the positive ions left behind. This escape of the free electrons rises very rapidly with temperature, and constitutes what is now called the *thermionic current* from the incandescent wire. This current is measured by the quantity of electricity escaping per square centimetre per second from the surface of the incandescent wire or body.

9. The study of the nature, variations and magnitude of this thermionic emission has been the subject of much investigation from a purely physical point of view.

As the chief purpose of the present volume is to direct attention to the technical applications of this phenomenon, we shall not occupy space with descriptions or discussions of the physical problems more than is necessary to explain the general principles involved. The experimental examination of the effect has been chiefly conducted by means of a simple apparatus, consisting of a straight metallic wire or carbon filament which is attached to platinum electrodes and sealed into a glass cylindrical bulb in an axial position.

Surrounding the filament is a tubular electrode made of metal, generally platinum, and this cylinder has a metallic connection made to it by a platinum wire welded to the cylinder and sealed through the glass (see Fig. 22).

The glass vessel can be exhausted to various pressures through a side tube or filled with various gases. If the surrounding metal cylinder is not employed, and if the wire

and maximum plate dissipation of 0.5 W. The maximum cathode current is not specified, but should not exceed a few milliamperes. The tube has an unusually high plate resistance, and the amplification factor is over 100. It is a good triode for a voltage amplifier. My example gave $g_m = 1.3$ mS, $r_p = 67$k, $\mu = 90$ at about 500 μA plate current, and $V_g = -1.0$ V. The transconductance varies strongly with plate current, reaching 2.2 mS at 1 mA, where $\mu = 116$.

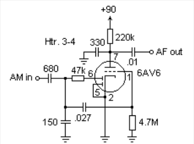

Diode Detector and AF Amplifier

A detector-amplifier circuit is shown at the left. This circuit is fed by a signal generator at AM in, and the audio output is taken from AF out. The amplifier uses grid-leak bias, which effectively clamps the peak of the audio wave at 0 V. The stronger the input signal, the more negative the grid bias, which makes the most of what is there. With an AM input of peak value 0.95 V and modulation amplitude 0.37 V, the peak to peak output was 3 V, for a gain of -8, which is quite good. I used a radio frequency of 2 MHz, and modulation of 1 kHz. At lower radio frequencies, the filtering of the output becomes progressively worse; this could be optimized if desired. This detector gives good fidelity with high sensitivity.

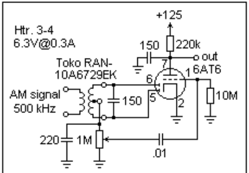

Full-Wave Diode Detector and Amplifier

A full-wave diode detector is shown at the right, using transformer input. The 6AT6 is similar to the 6AV6, but its mu is 70 instead of 100. The transformer is a Toko RAN-10A6729EK, available from Digi-Key. It is specified as 0.63 mH, and for 200 kHz. Adjust the bias detector, however, the FET is suitable, and a BJT biased near cutoff makes an excellent detector, as we saw in the page on Amplitude Modulation.

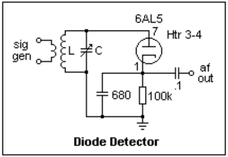

Diode Detector

A diode detector, using a standalone diode, is shown at the left. The output is the same amplitude as the modulation on the carrier that is input, very closely, so the detector could not do any better. Often, the diode was included in the first AF amplifier tube, sharing the same cathode, and dual diodes were also available for full-wave detection, which doubles the output. If you have a tuned circuit at this point, it would be easy to test a diode detector, and high voltages are not required, just the heater power. A solid-state diode would work the same way, but the 0.7 V drop would make it less sensitive, and it would not even respond to weaker signals. Using a Ge diode (1N34A) would help, but not eliminate, the problem. The thermionic diode is ideal in this application.

The typical superheterodyne receiver used a diode detector, followed by an audio voltage amplifier. These functions were usually combined in a single envelope, the diodes and the amplifying triode or pentode sharing the same cathode. A very good tube of this type was the 6AV6, whose structure is shown at the right. The diode plates are at the top of the envelope, with the cathode easily seen between them. The IF output was usually from an air-core, slug-tuned transformer, making it easy to use full-wave rectification. Only one of the diodes could be used for half-wave rectification, which was usually quite adequate. The triode is a high-mu triode with a maximum plate voltage of 300 V,

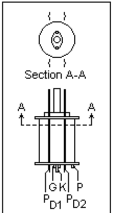

6AV6 Structure

---

or filament is heated by an electric current passed through it, then the ions or electrons given off by the wire would accumulate to a certain extent in the space, vacuous or gas-filled, round the wire and create a *space electric charge* which would soon prevent any further emission. This explanation of the limitation to the increase of the thermionic emission was first given by Mr. C. D. Child in 1911 (see *Physical Review*, vol. xxxii., p. 498, 1911).

By employing the metal cylinder we can make a definite difference of potential between the heated wire and the cold cylinder, and, therefore, an electric force so directed as to urge the ions from the wire to the embracing cylinder. The arrangement has also the convenience that we can determine the temperature of the incandescent wire by measuring its resistence or ratio of volt-drop down the filament to current through it, or else by making the filament one arm of a Wheatstone's Bridge arrangement.

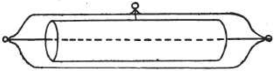

FIG. 22.—Exhausted tube having in it a metal cylinder and concentric incandescent tungsten wire to show the thermionic emission from the wire.

If we have previously determined the resistence of the filament at several standard temperatures we can by interpolation or extrapolution determine its temperature corresponding to other resistances.

Such an apparatus is a convenient modification of the lamps used by the author or by Sir William Preece in examining the Edison effect.

As soon as experiments began to be made with the above apparatus it was found that the observed effects were complicated and depended upon a great many factors, such as (1) the material of the wire or filament heated, (2) its temperature, (3) the gases or impurities occluded in it, (4) the nature of the gas surrounding it, (5) the pressure of this gas, and (6) the previous treatment of this wire as regards heating and exhaustion.

Broadly speaking, we may say that when the pressure of

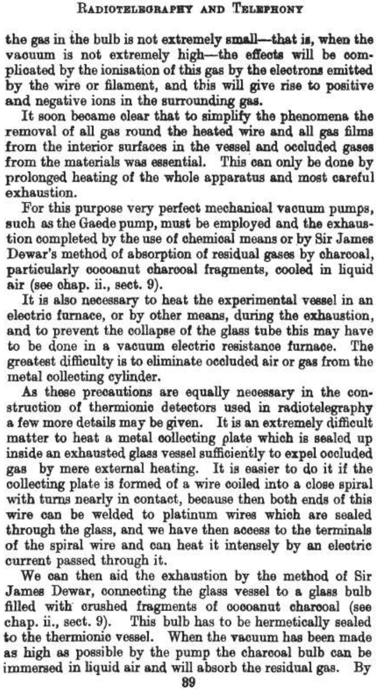

the gas in the bulb is not extremely small—that is, when the vacuum is not extremely high—the effects will be complicated by the ionisation of this gas by the electrons emitted by the wire or filament, and this will give rise to positive and negative ions in the surrounding gas.

It soon became clear that to simplify the phenomena the removal of all gas round the heated wire and all gas films from the interior surfaces in the vessel and occluded gases from the materials was essential. This can only be done by prolonged heating of the whole apparatus and most careful exhaustion.

For this purpose very perfect mechanical vacuum pumps, such as the Gaede pump, must be employed and the exhaustion completed by the use of chemical means or by Sir James Dewar's method of absorption of residual gases by charcoal, particularly cocoanut charcoal fragments, cooled in liquid air (see chap. ii., sect. 9).

It is also necessary to heat the experimental vessel in an electric furnace, or by other means, during the exhaustion, and to prevent the collapse of the glass tube this may have to be done in a vacuum electric resistance furnace. The greatest difficulty is to eliminate occluded air or gas from the metal collecting cylinder.

As these precautions are equally necessary in the construction of thermionic detectors used in radiotelegraphy a few more details may be given. It is an extremely difficult matter to heat a metal collecting plate which is sealed up inside an exhausted glass vessel sufficiently to expel occluded gas by mere external heating. It is easier to do it if the collecting plate is formed of a wire coiled into a close spiral with turns nearly in contact, because then both ends of this wire can be welded to platinum wires which are sealed through the glass, and we have then access to the terminals of the spiral wire and can heat it intensely by an electric current passed through it.

We can then aid the exhaustion by the method of Sir James Dewar, connecting the glass vessel to a glass bulb filled with crushed fragments of cocoanut charcoal (see chap. ii., sect. 9). This bulb has to be hermetically sealed to the thermionic vessel. When the vacuum has been made as high as possible by the pump the charcoal bulb can be immersed in liquid air and will absorb the residual gas. By

(from 1M to 3M) or grid capacitor (100 pF to 220 pF). The grid bias with no signal was about -0.5 V with a 1M grid resistor.

The circuit works by using the small grid current when the grid is slightly negative. This grid current increases rapidly and very nonlinearly, so square-law detection is possible. The grid resistor establishes a quiescent point around which the signal oscillates. The rf part is bypassed through the grid capacitor, while the af part remains and causes a voltage drop across the grid resistor, which is then amplified in the usual way. With a 100k plate resistor, the output af voltage could approach 1 V peak-to-peak. Another stage of audio amplification would give a quite satisfactory result, but I was hoping for one tube. (The 12AX7 has two triodes in the same envelope--if such a tube were used, the receiver would be, strictly speaking, single tube.) With most other kinds of detectors, such a hope would be quite vain, it must be admitted.

If the input signal amplitude is increased, the grid bias starts to decrease because the grid becomes positive on the peaks, and the detection becomes a little more linear. If the grid is biased to cutoff with a C supply, then there is clear linear detection, since the signal is effectively rectified. Such a circuit is called a grid-bias detector, and the signal amplitude must be restricted so that the grid does not go positive, when distortion would result.

You may notice in the tube manuals that a "maximum grid resistance" is given for a tube; for the 6J5, it is 1M. The reason for this is positive ion current, which may drive the grid more positive in a resistance-coupled amplifier, disturbing the bias. This is not a maximum in the same sense as the maximum plate dissipation or maximum cathode current are, since damage to the tube is not in question. There is no trouble in the grid leak detector, and values of several megohms are permissible.

There is no analogy to the grid-leak detector with FET's, which otherwise behave pretty much like vacuum tubes, because the leakage current is far too small for the desired behavior. As a gate-actual circuits that used a 1V6, but the amplifier shown at the left may be studied. It gave a gain of -14 with a plate current of 357 µA, so the effective transconductance was 298 µS. I substituted a 100k plate resistor, but the gain only increased marginally, to about -17. By experimenting, you may be able to get more gain from the tube. At low plate currents, the amplifier oscillated in an odd manner, and there were other peculiarities seen while fiddling around, but I did not take the time to track them down.

The Grid-Leak and Diode Detectors

Detectors are discussed in Amplitude Modulation, where the two types of square-law and linear are explained, and the grid-leak detector is mentioned. A circuit using a 6J5 triode is shown at the right. The audio output transformer T can be replaced with a 10k or 100k resistor, and the output observed with an oscilloscope. I could find only a 1k to 8Ω transformer, really designed for use with transistors. The circuit works, but I was not impressed by its output. Asking for loudspeaker output from a detector is being very optimistic, but I had some hopes. Perhaps a more suitable transformer would give better results.

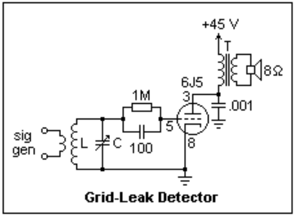

Grid-Leak Detector

The tuned circuit can be one from a crystal set, with a loopstick and variable capacitor. The signal generator is coupled by about 9 turns slipped over the loopstick, providing an AM signal about 30% modulated with 1 kHz, in the broadcast band. The output was roughly constant, no matter what the plate voltage (45 to 105 V), input amplitude, value of grid resistor

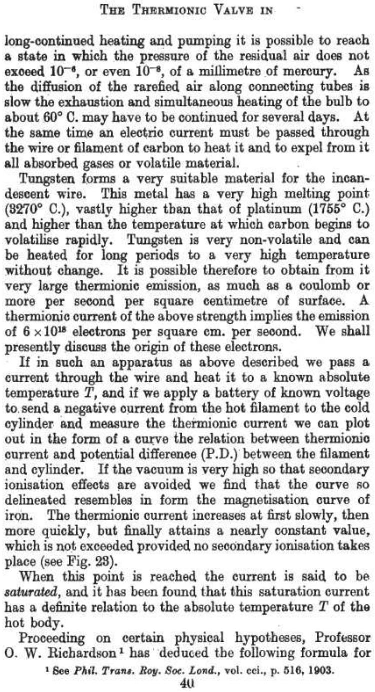

long-continued heating and pumping it is possible to reach a state in which the pressure of the residual air does not exceed 10^{-6}, or even 10^{-8}, of a millimetre of mercury. As the diffusion of the rarefied air along connecting tubes is slow the exhaustion and simultaneous heating of the bulb to about 60° C. may have to be continued for several days. At the same time an electric current must be passed through the wire or filament of carbon to heat it and to expel from it all absorbed gases or volatile material.

Tungsten forms a very suitable material for the incandescent wire. This metal has a very high melting point (3270° C.), vastly higher than that of platinum (1755° C.) and higher than the temperature at which carbon begins to volatilise rapidly. Tungsten is very non-volatile and can be heated for long periods to a very high temperature without change. It is possible therefore to obtain from it very large thermionic emission, as much as a coulomb or more per second per square centimetre of surface. A thermionic current of the above strength implies the emission of 6×10^{18} electrons per square cm. per second. We shall presently discuss the origin of these electrons.

If in such an apparatus as above described we pass a current through the wire and heat it to a known absolute temperature T, and if we apply a battery of known voltage to send a negative current from the hot filament to the cold cylinder and measure the thermionic current we can plot out in the form of a curve the relation between thermionic current and potential difference (P.D.) between the filament and cylinder. If the vacuum is very high so that secondary ionisation effects are avoided we find that the curve so delineated resembles in form the magnetisation curve of iron. The thermionic current increases at first slowly, then more quickly, but finally attains a nearly constant value, which is not exceeded provided no secondary ionisation takes place (see Fig. 23).

When this point is reached the current is said to be *saturated*, and it has been found that this saturation current has a definite relation to the absolute temperature T of the hot body.

Proceeding on certain physical hypotheses, Professor O. W. Richardson[1] has deduced the following formula for

---

[1] See *Phil. Trans. Roy. Soc. Lond.*, vol. cci., p. 516, 1903.

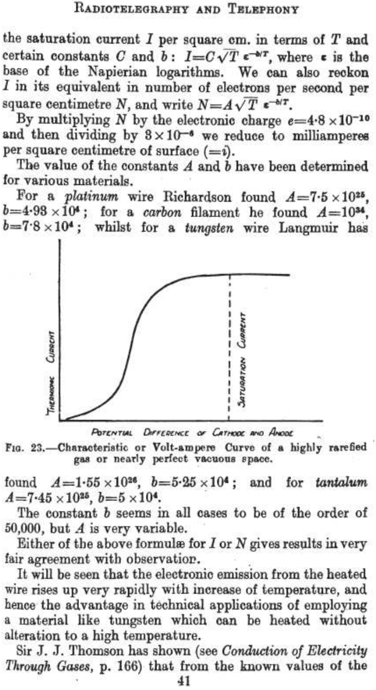

the saturation current I per square cm. in terms of T and certain constants C and b: $I=C\sqrt{T}\,\epsilon^{-b/T}$, where ϵ is the base of the Napierian logarithms. We can also reckon I in its equivalent in number of electrons per second per square centimetre N, and write $N=A\sqrt{T}\,\epsilon^{-b/T}$.

By multiplying N by the electronic charge $e=4{\cdot}8\times10^{-10}$ and then dividing by 3×10^{-6} we reduce to milliamperes per square centimetre of surface $(=i)$.

The value of the constants A and b have been determined for various materials.

For a *platinum* wire Richardson found $A=7{\cdot}5\times10^{25}$, $b=4{\cdot}93\times10^{4}$; for a *carbon* filament he found $A=10^{34}$, $b=7{\cdot}8\times10^{4}$; whilst for a *tungsten* wire Langmuir has found $A=1{\cdot}55\times10^{26}$, $b=5{\cdot}25\times10^{4}$; and for *tantalum* $A=7{\cdot}45\times10^{25}$, $b=5\times10^{4}$.

FIG. 23.—Characteristic or Volt-ampere Curve of a highly rarefied gas or nearly perfect vacuous space.

The constant b seems in all cases to be of the order of 50,000, but A is very variable.

Either of the above formulæ for I or N gives results in very fair agreement with observation.

It will be seen that the electronic emission from the heated wire rises up very rapidly with increase of temperature, and hence the advantage in technical applications of employing a material like tungsten which can be heated without alteration to a high temperature.

Sir J. J. Thomson has shown (see *Conduction of Electricity Through Gases*, p. 166) that from the known values of the

transconductance varies considerably with plate current, which gives rise to distortion for large signal amplitudes.

These tubes make vacuum-tube experiments very easy without requiring high-voltage supplies, only what is found in a normal transistor lab.

Sub-Miniature Tubes

The smallest regular electron tubes were the *subminiature* tubes, designed for battery-powered portable apparatus, from hearing aids to radiosonde transmitters. Typically, the lead wires were brought out through the press seal without a base, but there was also a subminiature base used for a few tubes. As an example, you can work with the 1V6, a pentode-triode with a filament taking 40 mA at 1.25V (the filament supply can be a D cell). The tube is 40mm x 10mm x 7mm, approximately. The maximum plate voltage is 45V, and typical plate currents are less than 1 mA.

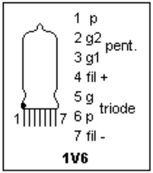

1V6

A sketch of the 1V6 and its connections is shown at the right. On the press seal, there is a red dot that identifies what I have called pin 1. The filament is located at the center of the tube, with the pentode on one side and the triode on the other. The triode shows a μ of about 10, r_p = 18.6 kΩ, and g_m = 440 μS. The pentode has g_m as high as 694 μS above 500 μA plate current, but it drops considerably at lower plate currents. The two devices can be tested as usual, taking care not to exceed 45V or 1 mA. Note that the pentode and the triode share the same filamentary cathode.

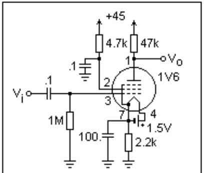

1V6 Amplifier

Unfortunately, I do not have any

charge grid tetrodes, behaving like triodes. The other tubes are of conventional construction, in general having low plate currents, less than 1 mA in many cases. These include twin-diode triodes (12AE6, 12FK6, 12AJ6), sharp cutoff pentodes (12AF6, 12BL6, 12CX6, 12EK6), remote cutoff pentodes (12CN5, 12DZ6), twin-diode remote cutoff pentode (12F8), pentagrid converter (12AD6), pentagrid (variable gain) amplifier 12EG6, and a twin-diode power screen-grid tetrode, 12J8. The last tube can deliver 20 mW output power, half of what the 12DL8 can do.

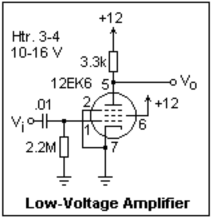

Low-Voltage Amplifier

The 12EK6 (also known as 12DZ6 and 12EA6) is a screen-grid pentode for low-voltage use. A voltage amplifier using this tube is shown at the left. This tube was generally used as an IF or RF amplifier, so the load would have been a resonant circuit, not a resistor, giving much higher gain. However, the resistive load is easy to experiment with. There is no voltage to waste for cathode bias, so grid-leak bias is used. When a signal is applied, the bias will adjust itself so that the top of the input waveform is clamped to 0 V. Vary the input amplitude to verify this. This circuit works with an input of up to about 2 V peak-to-peak, perhaps best for an input of around 1.0 V p-p. The output voltage is then about 5.8 V p-p, the plate potential is 6.0 V, plate current 1.8 mA, and gain about -5.5, corresponding to a transconductance of 1.8 mS, about right from measurements of the characteristics and the position of the operating point in this case. The transfer characteristic is rather curved, leading to considerable distortion for larger amplitudes. Nevertheless, it is remarkable that the circuit works so well at such a low voltage.

As a triode (plate, suppressor and screen connected together), the 12EK6 has $\mu = 8.5$, $g_m = 6.3$ mS and $r_p = 1.35$ kΩ. The

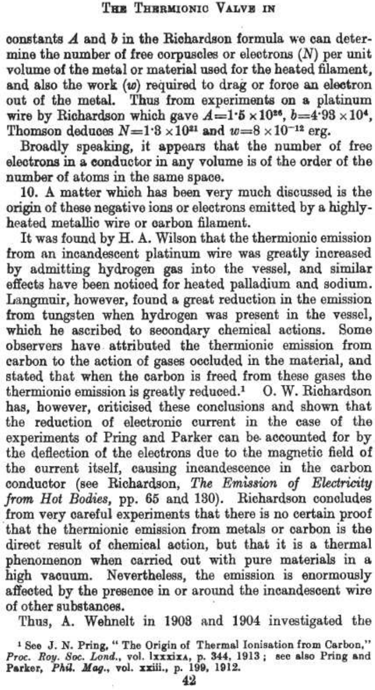

constants A and b in the Richardson formula we can determine the number of free corpuscles or electrons (N) per unit volume of the metal or material used for the heated filament, and also the work (w) required to drag or force an electron out of the metal. Thus from experiments on a platinum wire by Richardson which gave $A=1{\cdot}5\times10^{26}$, $b=4{\cdot}93\times10^{4}$, Thomson deduces $N=1{\cdot}3\times10^{21}$ and $w=8\times10^{-12}$ erg.

Broadly speaking, it appears that the number of free electrons in a conductor in any volume is of the order of the number of atoms in the same space.

10. A matter which has been very much discussed is the origin of these negative ions or electrons emitted by a highly-heated metallic wire or carbon filament.

It was found by H. A. Wilson that the thermionic emission from an incandescent platinum wire was greatly increased by admitting hydrogen gas into the vessel, and similar effects have been noticed for heated palladium and sodium. Langmuir, however, found a great reduction in the emission from tungsten when hydrogen was present in the vessel, which he ascribed to secondary chemical actions. Some observers have attributed the thermionic emission from carbon to the action of gases occluded in the material, and stated that when the carbon is freed from these gases the thermionic emission is greatly reduced.[1] O. W. Richardson has, however, criticised these conclusions and shown that the reduction of electronic current in the case of the experiments of Pring and Parker can be accounted for by the deflection of the electrons due to the magnetic field of the current itself, causing incandescence in the carbon conductor (see Richardson, *The Emission of Electricity from Hot Bodies*, pp. 65 and 130). Richardson concludes from very careful experiments that there is no certain proof that the thermionic emission from metals or carbon is the direct result of chemical action, but that it is a thermal phenomenon when carried out with pure materials in a high vacuum. Nevertheless, the emission is enormously affected by the presence in or around the incandescent wire of other substances.

Thus, A. Wehnelt in 1903 and 1904 investigated the

---

[1] See J. N. Pring, "The Origin of Thermal Ionisation from Carbon," *Proc. Roy. Soc. Lond.*, vol. lxxxixA, p. 344, 1913; see also Pring and Parker, *Phil. Mag.*, vol. xxiii., p. 199, 1912.

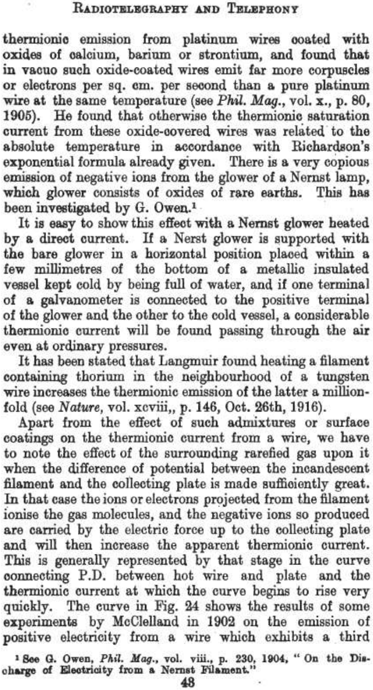

thermionic emission from platinum wires coated with oxides of calcium, barium or strontium, and found that in vacuo such oxide-coated wires emit far more corpuscles or electrons per sq. cm. per second than a pure platinum wire at the same temperature (see *Phil. Mag.*, vol. x., p. 80, 1905). He found that otherwise the thermionic saturation current from these oxide-covered wires was related to the absolute temperature in accordance with Richardson's exponential formula already given. There is a very copious emission of negative ions from the glower of a Nernst lamp, which glower consists of oxides of rare earths. This has been investigated by G. Owen.[1]

It is easy to show this effect with a Nernst glower heated by a direct current. If a Nerst glower is supported with the bare glower in a horizontal position placed within a few millimetres of the bottom of a metallic insulated vessel kept cold by being full of water, and if one terminal of a galvanometer is connected to the positive terminal of the glower and the other to the cold vessel, a considerable thermionic current will be found passing through the air even at ordinary pressures.

It has been stated that Langmuir found heating a filament containing thorium in the neighbourhood of a tungsten wire increases the thermionic emission of the latter a millionfold (see *Nature*, vol. xcviii,, p. 146, Oct. 26th, 1916).

Apart from the effect of such admixtures or surface coatings on the thermionic current from a wire, we have to note the effect of the surrounding rarefied gas upon it when the difference of potential between the incandescent filament and the collecting plate is made sufficiently great. In that case the ions or electrons projected from the filament ionise the gas molecules, and the negative ions so produced are carried by the electric force up to the collecting plate and will then increase the apparent thermionic current. This is generally represented by that stage in the curve connecting P.D. between hot wire and plate and the thermionic current at which the curve begins to rise very quickly. The curve in Fig. 24 shows the results of some experiments by McClelland in 1902 on the emission of positive electricity from a wire which exhibits a third

---

[1] See G. Owen, *Phil. Mag.*, vol. viii., p. 230, 1904, "On the Discharge of Electricity from a Nernst Filament."

The screen grid potential is above the plate's in this circuit. Pentodes offer the capacity for controlling the plate current through the screen voltage, so that the plate voltage can be what is required. For higher voltages, the screen grid is normally at the same or lower potential as the plate, on the average. The grid voltage cannot be measured accurately with a DMM or scope, since the bias conditions will be disturbed. However, you will find that the amplifier works very well. Measure the gain the usual way, with a function generator and oscilloscope.

Radios for motor cars had 6 V DC available, which was right for 6.3 V heaters, but could not be used for the B supply--even now, 6 V is inconveniently low for transistors. The usual solution was a *vibrator* supply. Contacts on the magnetically-driven vibrator armature converted 6 V DC to alternating current, which was stepped up by a transformer and then rectified to DC again (sometimes by contacts on the vibrator) for the B supply. Vibrators were electrically very noisy, and had short lives, but were widely used until solid-state equivalents became available. They came in cylindrical metal cans, and looked like electrolytic capacitors. The 0Z4, mentioned elsewhere, was a rectifer specifically for vibrator supplies. Another possibility was the *dynamotor*, a motor-generator with a single rotating armature and brushes, which could supply more power than a vibrator and was more reliable. However, they were too expensive for general commercial use, though widely used in the military.

When 12 V became the automotive standard, vibrator supplies or their solid-state equivalents, could still be used. Tubes with 12.6 V heaters were already common. Many had the heater center-tapped, so they could be used equally well on 6.3 V. However, 12 V is high enough to be used directly as the plate supply if the tubes are specially designed. A range of tubes was produced that could be used on 12 V only, without any high voltage at all. We have already discussed the 12DL8 and 12K5 space-charge-grid tetrodes, and how the space-charge grid allows larger plate currents for small plate voltages. There were approximately 15 tube types designed for 12V use, of which only three, the 12DL8, 12DS7 and 12K5 are space-

110V, but this can be exceeded a little in testing. I measured the characteristics for plate voltages of 125V and less, at grid voltages of 0.0, -0.5 and -1.0V (relative to the negative end of the filament). The plate current was always less than 1 mA. My results were $\mu = 62$, $g_m = 340\ \mu S$, and $r_p = 182k\Omega$. These values are close to the published values. The characteristics are noticeably curved in this region, so the parameters will vary with plate current. The apparently low value of transconductance is quite reasonable for the small plate currents at which the tube is used.

The amplifier shown gave a voltage gain of G - -40, with an input of 0.4V peak-to-peak. The plate current was about 100 μA, which made the plate voltage 77V and the grid bias -0.26V. The bandwidth of the amplifier was quite good, roughly from 10Hz to 50kHz. It should be noted that the total power drain of this circuit is no more than 83 mW, of which most is the filament power. This is very economical for a normal-sized tube. The tube remains quite cool in service, incidentally.

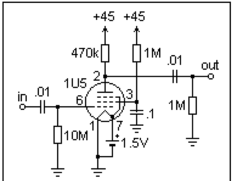

Pentode Voltage Amplifier

A pentode voltage amplifier with a gain of about -33 is shown at the right, using a 1U5, a sharp-cutoff pentode which comes in a 7-pin miniature package. Note the polarity of the filament supply, which is important, since it supplies a little grid bias as well. Since the voltages and currents are low, you can use the same 1/4W resistors used with transistors. The capacitors must have an appropriate voltage rating, of course. The 10M resistor is a "grid leak" that will prevent the grid from going positive. It charges on the positive excursions of the grid to provide whatever bias is necessary. The plate current was 61 μA, and the screen grid current was 20 μA.

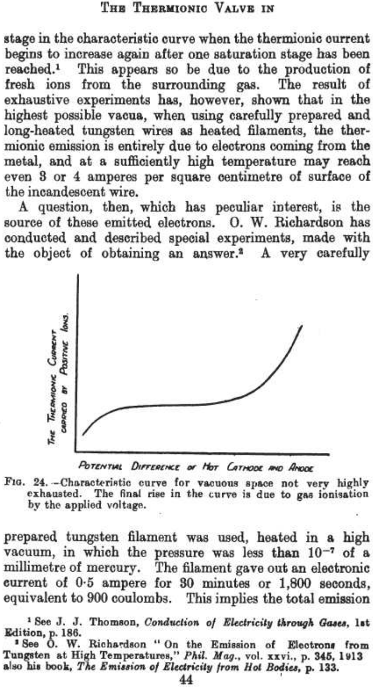

stage in the characteristic curve when the thermionic current begins to increase again after one saturation stage has been reached.[1] This appears so be due to the production of fresh ions from the surrounding gas. The result of exhaustive experiments has, however, shown that in the highest possible vacua, when using carefully prepared and long-heated tungsten wires as heated filaments, the thermionic emission is entirely due to electrons coming from the metal, and at a sufficiently high temperature may reach even 3 or 4 amperes per square centimetre of surface of the incandescent wire.

A question, then, which has peculiar interest, is the source of these emitted electrons. O. W. Richardson has conducted and described special experiments, made with the object of obtaining an answer.[2] A very carefully prepared tungsten filament was used, heated in a high vacuum, in which the pressure was less than 10^{-7} of a millimetre of mercury. The filament gave out an electronic current of 0·5 ampere for 30 minutes or 1,800 seconds, equivalent to 900 coulombs. This implies the total emission

The Thermionic Current carried by Positive Ions.
Potential Difference of Hot Cathode and Anode
FIG. 24.—Characteristic curve for vacuous space not very highly exhausted. The final rise in the curve is due to gas ionisation by the applied voltage.

---

[1] See J. J. Thomson, *Conduction of Electricity through Gases*, 1st Edition, p. 186.

[2] See O. W. Richardson "On the Emission of Electrons from Tungsten at High Temperatures," *Phil. Mag.*, vol. xxvi., p. 345, 1913 also his book, *The Emission of Electricity from Hot Bodies*, p. 133.

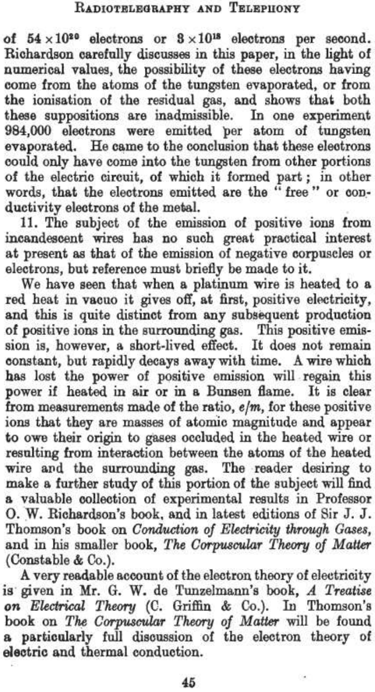

RADIOTELEGRAPHY AND TELEPHONY

of 54×10^{20} electrons or 3×10^{18} electrons per second. Richardson carefully discusses in this paper, in the light of numerical values, the possibility of these electrons having come from the atoms of the tungsten evaporated, or from the ionisation of the residual gas, and shows that both these suppositions are inadmissible. In one experiment 984,000 electrons were emitted per atom of tungsten evaporated. He came to the conclusion that these electrons could only have come into the tungsten from other portions of the electric circuit, of which it formed part; in other words, that the electrons emitted are the "free" or conductivity electrons of the metal.

11. The subject of the emission of positive ions from incandescent wires has no such great practical interest at present as that of the emission of negative corpuscles or electrons, but reference must briefly be made to it.

We have seen that when a platinum wire is heated to a red heat in vacuo it gives off, at first, positive electricity, and this is quite distinct from any subsequent production of positive ions in the surrounding gas. This positive emission is, however, a short-lived effect. It does not remain constant, but rapidly decays away with time. A wire which has lost the power of positive emission will regain this power if heated in air or in a Bunsen flame. It is clear from measurements made of the ratio, *e/m*, for these positive ions that they are masses of atomic magnitude and appear to owe their origin to gases occluded in the heated wire or resulting from interaction between the atoms of the heated wire and the surrounding gas. The reader desiring to make a further study of this portion of the subject will find a valuable collection of experimental results in Professor O. W. Richardson's book, and in latest editions of Sir J. J. Thomson's book on *Conduction of Electricity through Gases*, and in his smaller book, *The Corpuscular Theory of Matter* (Constable & Co.).

A very readable account of the electron theory of electricity is given in Mr. G. W. de Tunzelmann's book, *A Treatise on Electrical Theory* (C. Griffin & Co.). In Thomson's book on *The Corpuscular Theory of Matter* will be found a particularly full discussion of the electron theory of electric and thermal conduction.

45

Battery Tubes

In the early days, nearly all radios (and other electronic apparatus) were powered by batteries, usually primary cells. With the rise of central power distribution, apparatus could be powered (and storage batteries charged) from the AC line, which was a practically unlimited, cheap source so far as radios were concerned. Transformers made high-voltage power supplies easy to build. Low-voltage filaments were largely replaced by 6.3V heaters and unipotential cathodes since power was now cheap. However, there still remained a demand for portable radios beyond the reach of power lines, and these were necessarily powered from batteries. The emphasis here had to be on small power drain, since battery power is expensive. Typical batteries were 1.5V and 3.0 V for filaments ("A" batteries), and 45V or 90V for plate supplies ("B" batteries), both the usual Leclanche "dry" cells. Filaments were generally used, because of their much greater efficiency in mA per watt. Typical battery tubes had filaments taking only 50 mA at 1.4V, only 70 mW. Plate currents were not high, since a few mA would be quite sufficient. For experiments with battery tubes, a D cell in a holder is an adequate filament supply, good for about 100 hours of service. Even an AA cell will do for experiments.

The 1LH4 is a loktal battery triode, which is shown in a voltage amplifier at the left. A similar octal tube is the 1H5-GT, and either will do in this experiment (but the basing is different!). A diode plate is included, connected to pin 4, and using the negative side of the filament as its cathode. The diode is not used in this circuit. The maximum plate voltage is

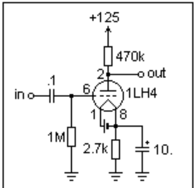

Voltage Amplifier

DMM, it would show 0 V and, of course, no current would flow. However, the Fermi levels of all three metals must line up for equilibrium, making a level surface on which the electrons will not tend to roll one way or the other. It is clear that in this state of equilibrium, the surfaces of metals 1 and 2 will not be at the same electrostatic potential. An oxide-coated cathode has a work function of about 1.0 V, while a nickel plate or grid has a work function of about 5.0 V. When the Fermi levels align, the cathode is 4.0 V positive with respect to the other electrode.

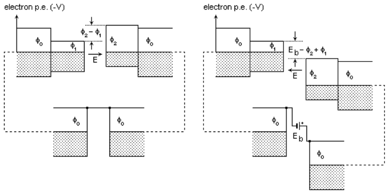

Effects of Contact Potential

At the right, we have connected a source of emf E_b into conductor 0. A source of emf maintains a difference in the Fermi levels equal to its voltage. With the polarity shown, the cathode 1 has now become negative with respect to the anode 2, by an amount equal to the applied voltage less the difference in work functions. If the device is a thermionic diode, current can now flow from cathode to anode. The actual plate voltage is about 4 V less than the meter says, because of the contact potential. In practice, things are a bit more complicated, because of space-charge effects. When you measure the V-I characteristic of a diode, there appears to be a small positive plate voltage when the measured voltage is zero, and a little current flows. The reason is that the electrons are emitted with some kinetic energy, and a negative space charge is created around the cathode. The actual emission to the anode comes from the minimum of this potential well, which is several volts negative, and so the plate may be slightly more positive than this minimum, in spite of contact potentials.

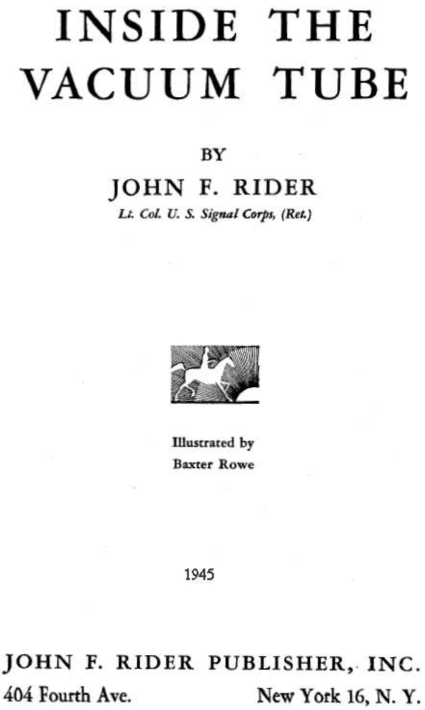

INSIDE THE VACUUM TUBE

BY

JOHN F. RIDER

Lt. Col. U. S. Signal Corps, (Ret.)

Illustrated by
Baxter Rowe

1945

JOHN F. RIDER PUBLISHER, INC.
404 Fourth Ave. New York 16, N. Y.

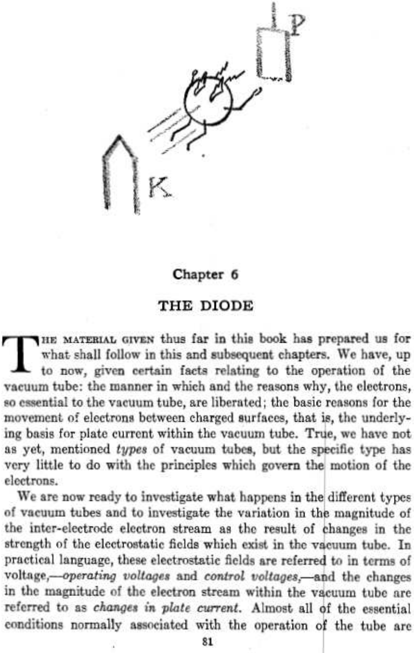

Chapter 6

THE DIODE

THE MATERIAL GIVEN thus far in this book has prepared us for what shall follow in this and subsequent chapters. We have, up to now, given certain facts relating to the operation of the vacuum tube: the manner in which and the reasons why, the electrons, so essential to the vacuum tube, are liberated; the basic reasons for the movement of electrons between charged surfaces, that is, the underlying basis for plate current within the vacuum tube. True, we have not as yet, mentioned *types* of vacuum tubes, but the specific type has very little to do with the principles which govern the motion of the electrons.

We are now ready to investigate what happens in the different types of vacuum tubes and to investigate the variation in the magnitude of the inter-electrode electron stream as the result of changes in the strength of the electrostatic fields which exist in the vacuum tube. In practical language, these electrostatic fields are referred to in terms of voltage,—*operating voltages* and *control voltages*,—and the changes in the magnitude of the electron stream within the vacuum tube are referred to as *changes in plate current.* Almost all of the essential conditions normally associated with the operation of the tube are

81

free electrons, but differ by up to a few volts from the true potential differences.

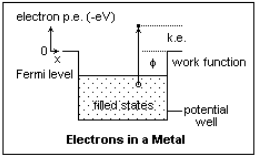

Electrons in a Metal

To understand this, let's review electrons in metals. One or more of the outer electrons of each metal atom is free to wander in the field of the ion cores, effectively binding the metal together. The positive ions create a potential well, as shown in the figure, and the electrons occupy the states of lowest energy. Note that electron energy increases upwards; since electrons are negatively charged, this means that voltage gets more negative upwards. There are not many more states than electrons, so they fill up the states up to some energy called the Fermi energy at 0 K. At ordinary temperatures, there is little difference, since the thermal energy is small compared to the kinetic energy of the electrons near the Fermi level.

If an electron is given enough energy, by thermal agitation or absorption of light, it may be separated completely from the metal and wander freely. The amount of energy necessary to just get outside is called the *work function*. Anything more goes into kinetic energy of the electron. The free electron can now be accelerated or decelerated by electric fields outside the metal in space. Work functions measured by thermionic, photoelectric and contact potential measurements agree roughly, but they depend sensitively on the surface preparation.

What happens in a vacuum tube is illustrated in the figure below. Suppose metal 1 is the cathode and metal 2 is the anode or grid, and the connections are made by a third metal 0 (there could be several such metals, but they all will act like metal 0). At the left, metals 1 and 2 are connected by a wire of metal 0. If metal 0 includes a should be rated at 2 A or more. Increase the filament voltage gradually, looking for the glow. There will be no plate current until the filament current reaches about 1.3 A, but it increases very rapidly beyond this point. The filament glows brilliantly, like an incandescent lamp, since its operating temperature is about 2400K, not the 900K of an oxide-coated filament. The filament current should not be allowed to exceed 1.6 A. If the power supply has current limiting, it can be useful here. By setting the plate voltage at near 200 V, you can see the saturation current as a function of filament current.

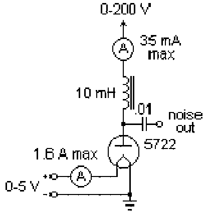

Testing the 5722

For two or more reasonable values of the saturation current, say 5 mA, 12 mA and 20 mA, record the current as a function of plate voltage and plot your results. For $I_f = 1.5$ A, the plate current saturated for about 50 V on the plate, approaching a value of about 12 mA. It is easy to find out what plate voltage to use to ensure saturation when making shot noise in this way. It is very difficult to make noise measurements in the usual breadboarding environment. I thought it just possible to have seen some on my 100 MHz scope with a plate current of 20 mA, without amplification. See the page on Noise for more discussion of noise measurements.

Contact Potential Effects

This is a good place to mention contact potentials, since you will probably run into their effects when experimenting with low-voltage tubes. Because of contact potentials, the voltages you measure with a DMM are not the actual voltages between electrodes, as seen by the

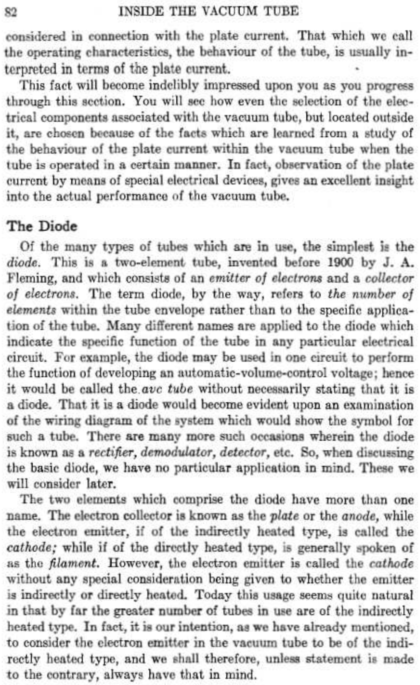

82 INSIDE THE VACUUM TUBE

considered in connection with the plate current. That which we call the operating characteristics, the behaviour of the tube, is usually interpreted in terms of the plate current.

This fact will become indelibly impressed upon you as you progress through this section. You will see how even the selection of the electrical components associated with the vacuum tube, but located outside it, are chosen because of the facts which are learned from a study of the behaviour of the plate current within the vacuum tube when the tube is operated in a certain manner. In fact, observation of the plate current by means of special electrical devices, gives an excellent insight into the actual performance of the vacuum tube.

The Diode

Of the many types of tubes which are in use, the simplest is the *diode*. This is a two-element tube, invented before 1900 by J. A. Fleming, and which consists of an *emitter of electrons* and a *collector of electrons*. The term diode, by the way, refers to *the number of elements* within the tube envelope rather than to the specific application of the tube. Many different names are applied to the diode which indicate the specific function of the tube in any particular electrical circuit. For example, the diode may be used in one circuit to perform the function of developing an automatic-volume-control voltage; hence it would be called the *avc tube* without necessarily stating that it is a diode. That it is a diode would become evident upon an examination of the wiring diagram of the system which would show the symbol for such a tube. There are many more such occasions wherein the diode is known as a *rectifier, demodulator, detector*, etc. So, when discussing the basic diode, we have no particular application in mind. These we will consider later.

The two elements which comprise the diode have more than one name. The electron collector is known as the *plate* or the *anode*, while the electron emitter, if of the indirectly heated type, is called the *cathode;* while if of the directly heated type, is generally spoken of as the *filament*. However, the electron emitter is called the *cathode* without any special consideration being given to whether the emitter is indirectly or directly heated. Today this usage seems quite natural in that by far the greater number of tubes in use are of the indirectly heated type. In fact, it is our intention, as we have already mentioned, to consider the electron emitter in the vacuum tube to be of the indirectly heated type, and we shall therefore, unless statement is made to the contrary, always have that in mind.

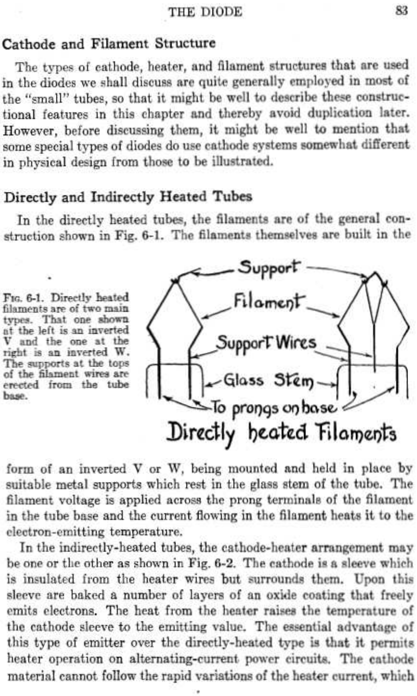

Cathode and Filament Structure

The types of cathode, heater, and filament structures that are used in the diodes we shall discuss are quite generally employed in most of the "small" tubes, so that it might be well to describe these constructional features in this chapter and thereby avoid duplication later. However, before discussing them, it might be well to mention that some special types of diodes do use cathode systems somewhat different in physical design from those to be illustrated.

Directly and Indirectly Heated Tubes

In the directly heated tubes, the filaments are of the general construction shown in Fig. 6-1. The filaments themselves are built in the

Fig. 6-1. Directly heated filaments are of two main types. That one shown at the left is an inverted V and the one at the right is an inverted W. The supports at the tops of the filament wires are erected from the tube base.

form of an inverted V or W, being mounted and held in place by suitable metal supports which rest in the glass stem of the tube. The filament voltage is applied across the prong terminals of the filament in the tube base and the current flowing in the filament heats it to the electron-emitting temperature.

In the indirectly-heated tubes, the cathode-heater arrangement may be one or the other as shown in Fig. 6-2. The cathode is a sleeve which is insulated from the heater wires but surrounds them. Upon this sleeve are baked a number of layers of an oxide coating that freely emits electrons. The heat from the heater raises the temperature of the cathode sleeve to the emitting value. The essential advantage of this type of emitter over the directly-heated type is that it permits heater operation on alternating-current power circuits. The cathode material cannot follow the rapid variations of the heater current, which

A special kind of diode should be mentioned here, because experiments with it are quite interesting. It is the *noise diode*, intended for the specific purpose of producing wide-band RF noise through the *shot effect*. Shot effect noise is fluctuations in the anode current due to the random collection of electrons. We have already mentioned that the anode current is controlled by the space charge around the filament. It was discovered, to some surprise, that this correlated successive electrons so that they were emitted regularly to maintain a constant current, and therefore the shot effect was nearly completely eliminated. That is, a normal diode has no shot effect noise in its plate current.

The noise diode is designed so that at reasonable plate voltages, all electrons emitted by the filament are immediately drawn to the plate without forming much of a space charge. Since the electrons are emitted randomly, the anode current will show the full shot effect noise. This is done by purposely making the filament to have low emission. To do this, a tungsten filament is used. Noise diodes give us the opportunity to observe a tungsten filament, as well as temperature saturation.

An available noise diode is the 5722, whose basing is shown at the right. The 7-pin miniature tube was made as late as 1977, and now costs about $14, which is probably not much more than when it was new. The maximum plate voltage is given as 200 V, and the maximum plate current as 35 mA, so apparently the plate can dissipate 7 W. The plate has wings that make a good dissipation probable.

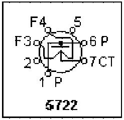

A circuit for testing the 5722 is shown at the left. Note that an RF choke is put in the plate lead to act as a load for the current fluctuations. This choke should be rated for the plate current employed. I connected a variable DC supply to the filament as shown, to pins 3 and 4, leaving the center tap alone. This supply

2, 7 and 9. This tube is designed for the rugged service of a television *damper diode*. During horizontal retrace, the damper diode conducts, charging the boost capacitor while absorbing the large inductive kick. The peak inverse voltage is 6000 V, the peak current 800 mA, and the average current 135 mA. The large-diameter cathode tube and long plate imply a large perveance, which, in fact, is about 2.3 mA/$V^{1.5}$. This tube happens to be very cheap, but would serve as an excellent half-wave rectifier for practically any purpose. There are other damper diodes, such as the 6W4 and the 6AX4GT (perveance 1.42), that would have similar characteristics.

As an example of the small signal diodes that are often combined with a triode or pentode in the same envelope, and share the same cathode, the 6AV6 or 6AT6 furnish good examples. The 6AV6 has its heater at pins 3-4, cathode at pin 2, and the signal diode plates at pins 5 and 6. The maximum current for each diode is 1 mA. I connected the two plates together for measurement, and took the current up to 3 mA, for which a plate voltage of 6.4 V was required. The curve of I against $V^{1.5}$ sagged a little at low currents, but the upper part was quite linear, showing a perveance of 0.085 mA/$V^{1.5}$ for one plate. The incremental resistance was 4.55kΩ, and V/I was 5.05kΩ at 1 mA. The current for one plate obeyed the formula $I = 0.15 + 0.085V^{1.5}$ mA. In this tube (and similar ones) the plates are flat, one on each side of the cathode.

The 1A3 seems to be the smallest signal diode of all. It was designed for portable measuring apparatus. The heater takes 0.15A at 1.4V (a D cell), connected to pins 1 and 7 of the 7-pin miniature envelope. The cathode is at pin 3, the anode at pins 2 and 6. The peak inverse voltage is 330V max., the maximum plate current 5 mA, and the average plate current 0.5 mA DC. Maximum heater-cathode potential is 140V. The anode is only a few millimeters high; most of the envelope contains only vacuum. The measured perveance was 0.075 mA/$V^{1.5}$.

The Noise Diode

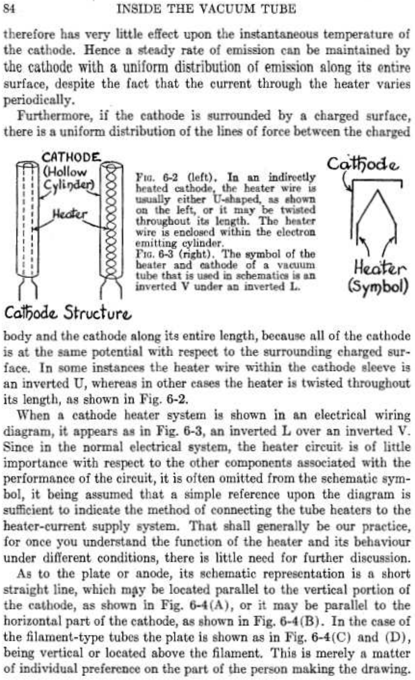

84 INSIDE THE VACUUM TUBE

therefore has very little effect upon the instantaneous temperature of the cathode. Hence a steady rate of emission can be maintained by the cathode with a uniform distribution of emission along its entire surface, despite the fact that the current through the heater varies periodically.

Furthermore, if the cathode is surrounded by a charged surface, there is a uniform distribution of the lines of force between the charged

FIG. 6-2 (left). In an indirectly heated cathode, the heater wire is usually either U-shaped, as shown on the left, or it may be twisted throughout its length. The heater wire is enclosed within the electron emitting cylinder.
FIG. 6-3 (right). The symbol of the heater and cathode of a vacuum tube that is used in schematics is an inverted V under an inverted L.

body and the cathode along its entire length, because all of the cathode is at the same potential with respect to the surrounding charged surface. In some instances the heater wire within the cathode sleeve is an inverted U, whereas in other cases the heater is twisted throughout its length, as shown in Fig. 6-2.

When a cathode heater system is shown in an electrical wiring diagram, it appears as in Fig. 6-3, an inverted L over an inverted V. Since in the normal electrical system, the heater circuit is of little importance with respect to the other components associated with the performance of the circuit, it is often omitted from the schematic symbol, it being assumed that a simple reference upon the diagram is sufficient to indicate the method of connecting the tube heaters to the heater-current supply system. That shall generally be our practice, for once you understand the function of the heater and its behaviour under different conditions, there is little need for further discussion.

As to the plate or anode, its schematic representation is a short straight line, which may be located parallel to the vertical portion of the cathode, as shown in Fig. 6-4(A), or it may be parallel to the horizontal part of the cathode, as shown in Fig. 6-4(B). In the case of the filament-type tubes the plate is shown as in Fig. 6-4(C) and (D), being vertical or located above the filament. This is merely a matter of individual preference on the part of the person making the drawing.

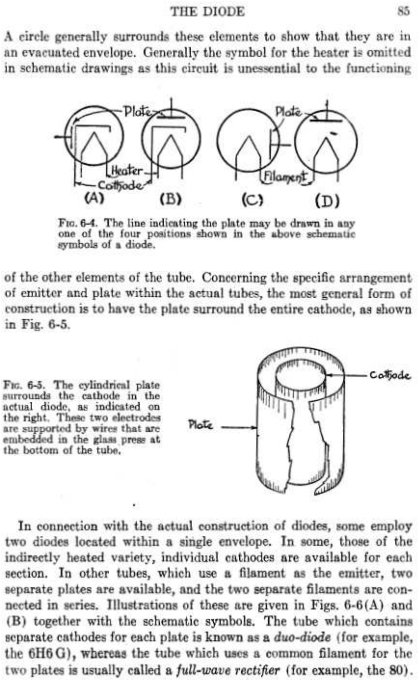
Plate
Plate
Heater
Cathode
Filament
(A)
(B)
(C)
(D)
Cathode
Plate

A circle generally surrounds these elements to show that they are in an evacuated envelope. Generally the symbol for the heater is omitted in schematic drawings as this circuit is unessential to the functioning of the other elements of the tube. Concerning the specific arrangement of emitter and plate within the actual tubes, the most general form of construction is to have the plate surround the entire cathode, as shown in Fig. 6-5.

FIG. 6-4. The line indicating the plate may be drawn in any one of the four positions shown in the above schematic symbols of a diode.

FIG. 6-5. The cylindrical plate surrounds the cathode in the actual diode, as indicated on the right. These two electrodes are supported by wires that are embedded in the glass press at the bottom of the tube.

In connection with the actual construction of diodes, some employ two diodes located within a single envelope. In some, those of the indirectly heated variety, individual cathodes are available for each section. In other tubes, which use a filament as the emitter, two separate plates are available, and the two separate filaments are connected in series. Illustrations of these are given in Figs. 6-6(A) and (B) together with the schematic symbols. The tube which contains separate cathodes for each plate is known as a *duo-diode* (for example, the 6H6 G), whereas the tube which uses a common filament for the two plates is usually called a *full-wave rectifier* (for example, the 80).

The 7Y4 is a typical small full-wave rectifier with an indirectly-heated cathode, like the more common 6X4 (miniature) and 6X5 (octal). This "Loktal" tube is inexpensive. Many of the common rectifier diodes are rather costly, for the curious reasons associated with the current tube market. The heater, taking 6.3V at 0.5A, is connected to pins 1-8 (as with all Loktal tubes). The cathode is pin 7, and the plates are pins 3 and 6. The peak inverse voltage is 1250 V, the peak current 180 mA, and the average dc current 70 mA. The heater-cathode voltage should not exceed 450 V. Measure the plate voltage for currents up to, say, 50 mA, and plot the results as for the 6H6. Again, we find a straight line and a perveance of 0.58 mA/V$^{1.5}$. Note that the plate voltage varies considerably as the current changes, from 4 V at 7 mA, to 16 V at 40 mA. Compare these voltages with those for a mercury-vapor *phanotron* as discussed in the next section. The 7Z4 is a somewhat larger full-wave rectifier (with perveance 0.40), the Loktal equivalent to the types 80 or 5Y3 that are now much more expensive.

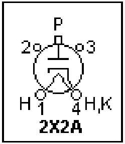

An excellent diode for observing the Langmuir-Child law is the 2X2A. This tube has a 4-pin base like the 82 phanotron discussed below, and the large, bell-like anode is brought out to a cap at the top of the ST envelope. The oxide-coated cathode thimble is easily seen. The heater takes 2.5V at 1.75A, so it can use the same transformer as the type 82. The rated DC current is 7.5 mA, and the maximum voltage is 4500V. A plate voltage of about 60V is needed to reach 7.5 mA plate current, so measurements can be made over a wide range of voltages. Plot your results as $I^{2/3}$ vs. V. A straight line will be found, that intercepts the V axis at -1.2V. The perveance of the 2X2 is found to be 0.0165 mA/V$^{3/2}$. The unusually low value is due to the large cathode-anode spacing.

The 6V3-A is a strange miniature tube with a cap on top that is the cathode connection. Its heater, connected to pins 4 and 5 of the 9-pin miniature base, takes 1.75 A at 6.3 V. The plate is connected to pins

V can be resisted, and the DC plate current should not exceed 9 mA. Peak currents can go up to 45 mA if necessary, however. I measured the perveance as 2.42 mA/$V^{1.5}$, for one plate, a large value. The 6AL5 gives 9 mA with a plate voltage of only about 2.5 V! The heater, connected to H-H, pins 3 and 4, takes 0.3 A at 6.3 V.

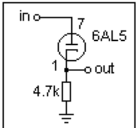

Signal Rectifier

Try the 6AL5 in the circuit shown at the left, which is a basic signal rectifier with a 4.7k load resistor. Feed it with the signal generator, and compare the output and input with the oscilloscope. Try input peak-to-peak voltages of only 2 V or so. You will notice that there is no "diode drop" with the 6AL5--it acts like a perfect diode, rectifying down to small voltages. We know how to do this with a semiconductor diode and an op-amp, but here it's done quite simply. The 6AL5 has an incremental resistance of only about 237 Ω, and is nearly linear. It is easy to run a plate voltage versus plate current curve with a low-voltage power supply. Keep the load resistor, and subtract the voltages at plate and cathode to find the plate voltage.

The 6H6 is an octal dual signal diode like the 6AL5, in a unique small metal envelope. The heater is connected to pins 2-7, the cathodes to 4 and 8, the plates to 3 and 5. 3 and 4 are one diode, 8 and 5 the other, and completely independent. It can be used for any reasonable service, such as AM detection, as a full-wave rectifier, or as a voltage doubler, so long as the current per plate is 8 mA or lower, and inverse voltages do not exceed 420 V. The voltage between heater and cathodes should not exceed 330 V. Measure the plate current as a function of the plate voltage up to 10 mA (the plate voltage will be about 7 V), and plot the current against the 3/2 power of the voltage. I obtained a rather straight line, showing agreement with Langmuir-Child, with a permeance of 0.5 mA/$V^{1.5}$. At 8 mA, the incremental resistance was 590Ω, and V/I = 785Ω. The 12H6 and 7H6 are similar tubes with different heater ratings and basing.

The justification for such a distinction between the two tubes is due solely to construction and not utility, for the duo-diode can be arranged for use as a full-wave rectifier without any trouble. However, the fact that separate cathodes are available in the duo-diode gives it certain operating capabilities which are not possessed by the tube which has the two separate plates, but a single source of electrons. In the case of the duo-diode, there may be only a single heater winding serving the two cathodes.

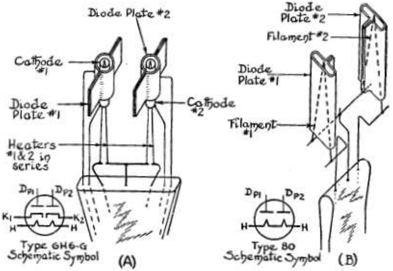

FIG. 6-6. The duo-diode tube shown in (A) has a cathode for each of its two plates and the 80 type rectifier shown in (B) has but a single continuous filament for both of its two plates.

The presence of more than one diode structure in an envelope does not in any way alter the considerations surrounding the basic two-element tube, and this applies equally well to the tube with the two plates but a single filament. Whatever is said about the single diode applies equally well to the other section. The same is true if the two plates of the two diodes are connected in parallel by some electrical connection and two cathodes are operated as one. The only difference resulting from such an arrangement is a change in the characteristics of the tube, but not in the basic principles of operation. Just what we mean by this in actual numerical values will be shown later in this chapter.

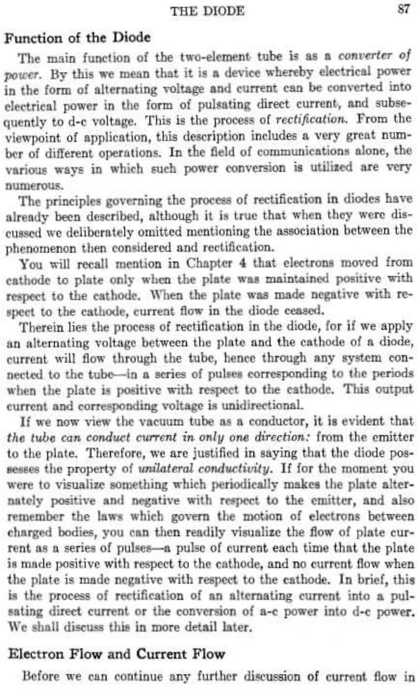

THE DIODE 87

Function of the Diode

The main function of the two-element tube is as a *converter of power*. By this we mean that it is a device whereby electrical power in the form of alternating voltage and current can be converted into electrical power in the form of pulsating direct current, and subsequently to d-c voltage. This is the process of *rectification*. From the viewpoint of application, this description includes a very great number of different operations. In the field of communications alone, the various ways in which such power conversion is utilized are very numerous.

The principles governing the process of rectification in diodes have already been described, although it is true that when they were discussed we deliberately omitted mentioning the association between the phenomenon then considered and rectification.

You will recall mention in Chapter 4 that electrons moved from cathode to plate only when the plate was maintained positive with respect to the cathode. When the plate was made negative with respect to the cathode, current flow in the diode ceased.

Therein lies the process of rectification in the diode, for if we apply an alternating voltage between the plate and the cathode of a diode, current will flow through the tube, hence through any system connected to the tube—in a series of pulses corresponding to the periods when the plate is positive with respect to the cathode. This output current and corresponding voltage is unidirectional.

If we now view the vacuum tube as a conductor, it is evident that *the tube can conduct current in only one direction:* from the emitter to the plate. Therefore, we are justified in saying that the diode possesses the property of *unilateral conductivity.* If for the moment you were to visualize something which periodically makes the plate alternately positive and negative with respect to the emitter, and also remember the laws which govern the motion of electrons between charged bodies, you can then readily visualize the flow of plate current as a series of pulses—a pulse of current each time that the plate is made positive with respect to the cathode, and no current flow when the plate is made negative with respect to the cathode. In brief, this is the process of rectification of an alternating current into a pulsating direct current or the conversion of a-c power into d-c power. We shall discuss this in more detail later.

Electron Flow and Current Flow

Before we can continue any further discussion of current flow in

and other interesting things. They do work rather well, and it is good to make their acquaintance.

The forward voltage (in the direction of current flow) of a diode is always relatively low, less than 15 V or so. The plate current is roughly proportional to the 3/2 power of the anode-cathode voltage (Langmuir-Child law), and the proportionality factor is called the *perveance*. The perveance depends on the geometry of the tube, increasing with larger area and closer spacing. It's remarkable that most diodes agree with Langmuir-Childs so well, in spite of different geometries. Since the voltages are low, contact potentials may affect your measurements. Contact potentials are discussed below in the section on low-voltage tubes. The easiest way to find the perveance is to plot $I^{2/3}$ against V, and to draw the best straight line. The intercept gives the value of the "true" zero plate voltage, and the slope, raised to the 3/2 power, is the perveance. Perveances range from 0.02 to 2.4 $mA/V^{3/2}$ for a representative assortment of 12 diodes of all types. There is no turn-on voltage drop for a thermionic diode, as there is for a silicon diode. Conduction begins immediately when the plate is positive with respect to the cathode, and stops immediately when the plate goes negative. It is easy to measure the V-I characteristic of a diode with a low-voltage DC supply, a voltmeter and an ammeter. I use a 100Ω resistor in series to make adjustment easier and safer. Thermionic diodes are not as easy to destroy as semiconductor diodes, and will take a good deal of abuse.

The 6AL5 dual diode, whose basing is shown at the right (7-pin miniature socket), is a typical signal diode. IS is an internal shield between the diodes. The two diodes and the shield are easily seen through the glass envelope, and you should notice how close the plates are to the cathodes. The close spacing means a large perveance, so only small plate voltages are required. Don't connect this tube directly across high voltages! A peak inverse voltage of 330

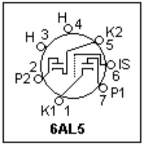

Diodes

Thermionic diodes, like semiconductor diodes, are divided into *signal* diodes that handle small currents at low voltages, and *rectifier* diodes that handle large currents, often with large inverse voltages. A diode has an electron-emitting cathode and an electron-receiving anode or plate. The arrangments of cathodes and plates in commercial tubes, and what they are called, are shown in the figure. Signal diodes are also often added to a triode or pentode, sharing the same cathode and with one or two plates. Current flows only from plate to cathode, and this unidirectional conduction is the purpose of a diode. Diodes cannot amplify.

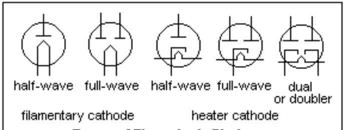

Types of Thermionic Diodes

Signal diodes always have indirectly-heated cathodes, so they are easy to use. It is only necessary to make sure that the heater-cathode voltage does not exceed specified limits, usually a few hundred volts. Rectifier diodes often have filamentary oxide-coated cathodes, since these cathodes are more efficient when large currents are needed, requiring less power. We are considering only vacuum diodes, *kenotrons*, in this section. Thermionic gas diodes, or *phanotrons*, will be treated below, since they have rather different properties.

Thermionic diodes have now been completely superseded by semiconductor diodes, largely for economic reasons, physical size and the need for a filament supply. A silicon diode capable of carrying 1 A is available for $0.04 or so, and takes up very little room. However, diodes can teach us a lot about thermionic emission

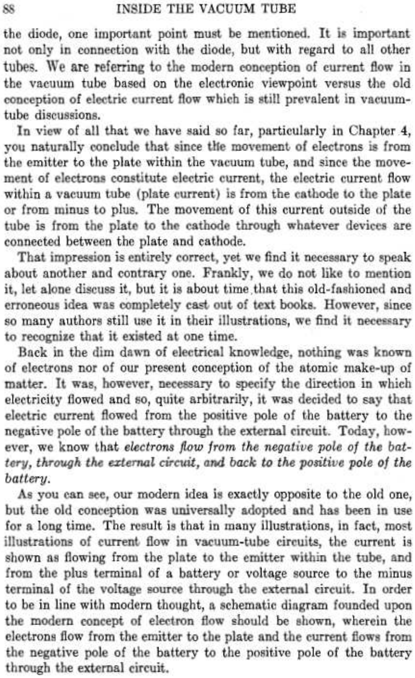

the diode, one important point must be mentioned. It is important not only in connection with the diode, but with regard to all other tubes. We are referring to the modern conception of current flow in the vacuum tube based on the electronic viewpoint versus the old conception of electric current flow which is still prevalent in vacuum-tube discussions.

In view of all that we have said so far, particularly in Chapter 4, you naturally conclude that since the movement of electrons is from the emitter to the plate within the vacuum tube, and since the movement of electrons constitute electric current, the electric current flow within a vacuum tube (plate current) is from the cathode to the plate or from minus to plus. The movement of this current outside of the tube is from the plate to the cathode through whatever devices are connected between the plate and cathode.

That impression is entirely correct, yet we find it necessary to speak about another and contrary one. Frankly, we do not like to mention it, let alone discuss it, but it is about time that this old-fashioned and erroneous idea was completely cast out of text books. However, since so many authors still use it in their illustrations, we find it necessary to recognize that it existed at one time.

Back in the dim dawn of electrical knowledge, nothing was known of electrons nor of our present conception of the atomic make-up of matter. It was, however, necessary to specify the direction in which electricity flowed and so, quite arbitrarily, it was decided to say that electric current flowed from the positive pole of the battery to the negative pole of the battery through the external circuit. Today, however, we know that *electrons flow from the negative pole of the battery, through the external circuit, and back to the positive pole of the battery.*

As you can see, our modern idea is exactly opposite to the old one, but the old conception was universally adopted and has been in use for a long time. The result is that in many illustrations, in fact, most illustrations of current flow in vacuum-tube circuits, the current is shown as flowing from the plate to the emitter within the tube, and from the plus terminal of a battery or voltage source to the minus terminal of the voltage source through the external circuit. In order to be in line with modern thought, a schematic diagram founded upon the modern concept of electron flow should be shown, wherein the electrons flow from the emitter to the plate and the current flows from the negative pole of the battery to the positive pole of the battery through the external circuit.

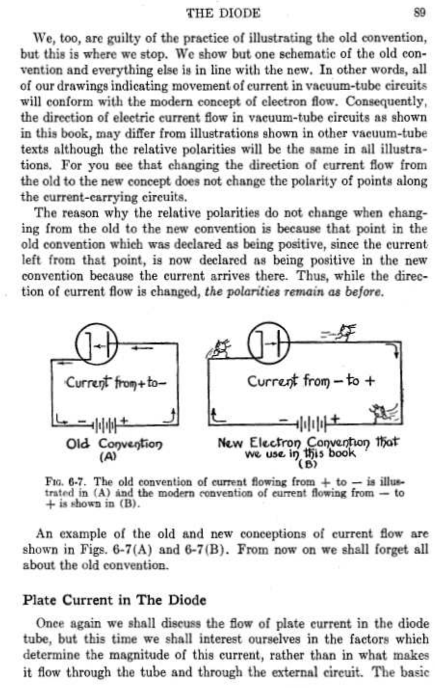

We, too, are guilty of the practice of illustrating the old convention, but this is where we stop. We show but one schematic of the old convention and everything else is in line with the new. In other words, all of our drawings indicating movement of current in vacuum-tube circuits will conform with the modern concept of electron flow. Consequently, the direction of electric current flow in vacuum-tube circuits as shown in this book, may differ from illustrations shown in other vacuum-tube texts although the relative polarities will be the same in all illustrations. For you see that changing the direction of current flow from the old to the new concept does not change the polarity of points along the current-carrying circuits.

The reason why the relative polarities do not change when changing from the old to the new convention is because that point in the old convention which was declared as being positive, since the current left from that point, is now declared as being positive in the new convention because the current arrives there. Thus, while the direction of current flow is changed, *the polarities remain as before.*

FIG. 6-7. The old convention of current flowing from + to − is illustrated in (A) and the modern convention of current flowing from − to + is shown in (B).

An example of the old and new conceptions of current flow are shown in Figs. 6-7(A) and 6-7(B). From now on we shall forget all about the old convention.

Plate Current in The Diode

Once again we shall discuss the flow of plate current in the diode tube, but this time we shall interest ourselves in the factors which determine the magnitude of this current, rather than in what makes it flow through the tube and through the external circuit. The basic

A tube designated simply 6N7 will be a metal-envelope octal tube with a 6.3V heater. A 6N7GT will have a cylindrical glass envelope. A 6N7G would have a shouldered glass envelope of the graceful shape designated ST. The electrical characteristics of such tubes were the same, whatever the envelope shape.

A very important part of vacuum-tube technology was bringing the metal leads through the glass envelope. Coefficients of expansion must be exactly matched, and the seal must be strong. Originally, tubes had bases (usually Bakelite) to support the contact pins mechanically, taking the strain off the pressed-glass seal, which was made of lead glass. Around 1935, the metal envelope was developed, but there was still a base. The all-glass "miniature" tube was made possible by the "button seal" that supported the contact pins mechanically as well as bringing them through the glass, allowing the base to be eliminated and tube size to be reduced. The insides, or "cage," was the same size as in previous tubes, however. It is supported on its leads, which are welded to the contact pins before the envelope is fused in place and evacuated. The button seal is also used, in a larger form, on tubes designated by GB at the end of the type designation, and by Loktals. The final step in manufacture was "flashing" the getter, usually barium or magnesium, to perfect the vacuum by adsorption of any remaining gases, leaving a shiny coating. This was generally done by heating a loop inductively by RF from outside.

7, and often to pins 3-4 on 7-pin, or 4-5 on 9-pin, miniature tubes. Sometimes halves of the heater can be connected in series or parallel, for two different voltages. Sockets were originally mounted in holes punched in aluminum chassis, secured by locking rings or by screws and nuts with a mounting plate. The chassis was, not surprisingly, the ground or common.

The "Loktal" tube was an excellent idea that was never universally adopted, mainly because miniature tubes took over in the 1950's. Since loktal was a trade name, RCA used "lock-in" instead, and you sometimes see "loctal." The loktal tube has an 8-pin button-seal (like the seal on miniature and octal GTB tubes). A natural metal base (of some aluminum alloy, apparently) shields the base of the tube and has a central pin with a circumferential locking groove. The pins project only 6 mm, and are 1.4 mm in diameter, much smaller than octal pins, so the locking action guarantees that the tube will stay in the socket in spite of the small pins. The tubes are roughly the same size as an octal GT tube. Most are one size, but a few power tubes have a slightly longer envelope. There are no grid caps on any Loktal tube, and the heater connections are always to pins 1 and 8. Among the thoughtful features of loktal design, the type number appears in a hexagon on the top of the tube where it is visible from above, not on the side as on octal tubes. There is a dimple on the base corresponding to the key of the central pin, making the tube easy to orient for insertion. It seems that a lot of getter was used, so the tops of the envelope appear heavily silvered. The available types are only those used in AM and FM receivers. There are, nevertheless, enough types for a broad variety of experiments, and the prices are not excessive, so you may want to standardize on Loktals. Type numbers beginning with 7 have 6.3 V heaters, while type numbers beginning with 14 have 12.6 V heaters. There are some 7xx and 14xx tubes that are not Loktal, and some tubes that actually take a 7 V heater supply. One loktal rectifier, the 5AZ4 (a 5Y3 equivalent), has a 5 V filament. Loktal tubes designed specifically for battery-powered equipment had 1.4V filaments. The type numbers began with "1L." There were also rectifier and beam power loktals with 35, 50 and 70-volt heaters for AC/DC sets with series heater connections.

reasons for the flow of current between the elements of the tube have already been discussed in Chapter 4, so that we need merely illustrate the similarity between the diode and the two-element structures previously illustrated. This is done in Figs. 6-8(A), (B), and (C). Since we shall speak about the flow of current through the circuit, a meter M is included in each illustration.

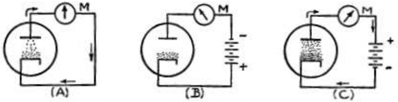

FIG. 6-8. High-speed electrons arriving at the plate form a small plate current, even though no potential is applied to the plate; this is shown in (A). No plate current results when the potential on the plate is negative, as shown in (B), and a current flows that is greater than that of (A) when a positive potential is applied to the plate; this is shown in (C).

In each of these illustrations, the cathode is the equivalent of the emitter discussed in Chapter 4, and the plate is the equivalent of the positively charged surface. The battery shown in Fig. 6-8(C) is the voltage source used in Chapter 4 which made the charged surface positive with respect to the plate. In the case of Fig. 6-8(B), the battery makes the plate negative with respect to the cathode.

The dots shown between the cathode and the plate in Figs. 6-8(A), 6-8(B), and 6-8(C), correspond to the space charge discussed in Chapter 4. In Fig. 6-8(C), the plate voltage is of such magnitude that a flow of plate current occurs. In Fig. 6-8(B), the plate having been made negative with respect to the cathode, the plate current is zero, and the space charge can be seen concentrated in the neighborhood of the cathode just as described in Chapter 4.

We now come to Fig. 6-8(A), which has not been mentioned. Basically, the conditions existing in this circuit correspond to what has been described in Chapter 4 but, as you can see, the battery has been omitted and the plate is joined to the cathode. What happens in this circuit? In the first place, since the plate and the cathode are joined and no voltage source is connected between the plate and cathode, presumably no difference of potential exists between these two elements. Since the two are joined, whatever the voltage may be at the

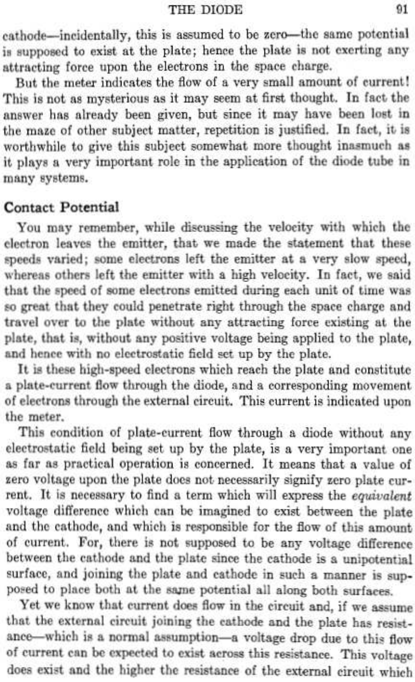

THE DIODE 91

cathode—incidentally, this is assumed to be zero—the same potential is supposed to exist at the plate; hence the plate is not exerting any attracting force upon the electrons in the space charge.

But the meter indicates the flow of a very small amount of current! This is not as mysterious as it may seem at first thought. In fact the answer has already been given, but since it may have been lost in the maze of other subject matter, repetition is justified. In fact, it is worthwhile to give this subject somewhat more thought inasmuch as it plays a very important role in the application of the diode tube in many systems.

Contact Potential

You may remember, while discussing the velocity with which the electron leaves the emitter, that we made the statement that these speeds varied; some electrons left the emitter at a very slow speed, whereas others left the emitter with a high velocity. In fact, we said that the speed of some electrons emitted during each unit of time was so great that they could penetrate right through the space charge and travel over to the plate without any attracting force existing at the plate, that is, without any positive voltage being applied to the plate, and hence with no electrostatic field set up by the plate.

It is these high-speed electrons which reach the plate and constitute a plate-current flow through the diode, and a corresponding movement of electrons through the external circuit. This current is indicated upon the meter.

This condition of plate-current flow through a diode without any electrostatic field being set up by the plate, is a very important one as far as practical operation is concerned. It means that a value of zero voltage upon the plate does not necessarily signify zero plate current. It is necessary to find a term which will express the *equivalent* voltage difference which can be imagined to exist between the plate and the cathode, and which is responsible for the flow of this amount of current. For, there is not supposed to be any voltage difference between the cathode and the plate since the cathode is a unipotential surface, and joining the plate and cathode in such a manner is supposed to place both at the same potential all along both surfaces.

Yet we know that current does flow in the circuit and, if we assume that the external circuit joining the cathode and the plate has resistance—which is a normal assumption—a voltage drop due to this flow of current can be expected to exist across this resistance. This voltage does exist and the higher the resistance of the external circuit which

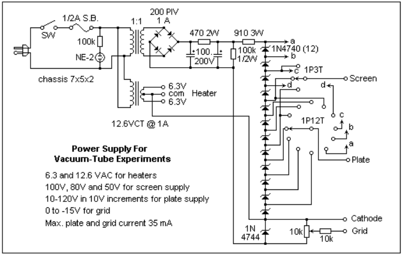

Power Supply For Vacuum-Tube Experiments

6.3 and 12.6 VAC for heaters
100V, 80V and 50V for screen supply
10-120V in 10V increments for plate supply
0 to -15V for grid
Max. plate and grid current 35 mA

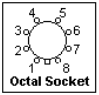

Vacuum tubes come with metal or glass envelopes, and in latter days with either the familiar octal arrangment of 8 pins, or as miniature glass tubes with 7 or 9 pins. There were earlier bases with four, five or six pins. Later, 12-pin miniature "compactron" or "duodecal" tubes were used in TV sets. Miniature tubes were not miniature, simply tubes with a button seal and all-glass envelope closely fitting a normal-sized cage. Subminiature tubes were actually miniature. Sometimes connections to grid, plate or (rarely) cathode were made to caps at the top of the tubes. In small tubes, these caps have a diameter of 1/4". The pins are numbered consectively clockwise, starting from the left of the index key for the octal, or to the left of the wider space, for the miniature, always looking at the bottom of the tube. This is shown for the octal base at the left. Pin numbers are given in the circuit schematics here. Most sockets have pin numbers marked. You will need to get sockets for the tubes you study, one for each type of socket. Solder wires to the tab at each pin that can be inserted in the solderless breadboard. I use the resistor color code for the pin numbers. A convenient octal socket fixture is available that comes with screw terminals for making connections. It was intended for relays, but is very useful for tube experiments. Heater connections for octal tubes are typically (not always!) to pins 2 and put in a box with an on-off switch and convenient terminals. Ground the heater supply (at a center tap if one is provided) to the B+ ground, to avoid excess voltages between heater and cathode. If you have a 12.6V CT (center-tapped) secondary, you can supply both 6.3 and 12.6 V heaters. Many 12.6V miniature tubes can also be connected for 6.3V. Tubes whose designations begin with "1" have filaments that can be supplied from a single D cell. Obtain a holder for the cell so connections are easy. 6.3 V was chosen to be compatible with 6 V car batteries, but the supply is usually AC. Many rectifier diodes use a 5 V filament or heater supply, apparently for historical reasons.

A "C" supply, for the grid bias in measuring characteristics, can be any isolated low-voltage supply of say, 15V, and a potentiometer can be used to pick off a variable voltage, since little current is involved. A separate high-voltage supply for screen grids may also be convenient, though it is easy to pick off the necessary voltages with a Zener or a VR tube from the main B+ supply. This cannot be done, of course, if the B+ voltage is adjusted using a variable transformer.

An all-in-one economical supply for vacuum-tube measurements is shown below. It uses an inexpensive isolation transformer from All Electronics, and can be made for about $30.00. The most expensive single part is the aluminum chassis. The grid potentiometer could be a precision 10-turn pot, but this would be expensive, and an ordinary carbon or plastic potentiometer (1/2 W or better) will be satisfactory. The maximum plate voltage of 120V and maximum plate current of 35 mA is adequate for many measurements. If you use a three-wire line cord, ground the chassis to the green wire. If you use only a two-wire line cord, it is probably better not to ground the chassis.

connects the plate to the cathode, the more easily it can be measured. As to just what you would call this voltage, the best that we can do is to refer to the name by which it is known, namely, *"contact potential."*

The use of this term to identify this voltage is not wholly correct, yet, for want of a better name, it is used in commercial tube literature, and we shall therefore use it on those grounds.

Determination of Diode Behaviour

All of this information about the flow of current in a diode, with a positive, a negative, and zero voltage at the plate, merely supplements the facts given in Chapter. 4. What we wish to determine is the behaviour of the diode when those conditions which cause the flow of plate current are varied. Only then can we establish facts relative to the manner in which the diode is used.

What are the conditions which determine the amount of plate-current flow in the diode? Since the magnitude of plate current is partly dependent upon the emission of electrons from the cathode or filament of the tube, the temperature of the emitter is a factor. In accordance with what was said in Chapter 4, the electrostatic field set up at the plate influences the effectiveness of the space charge, and hence affects the plate current. Consequently, we see that the value of the plate voltage has an effect upon the amount of plate current flowing through the tube.

We must therefore establish characteristic curves, or behaviour curves, showing the variation in plate current with emitter temperature and also characteristic curves showing the variation in plate current with plate voltage. These are the two basic characteristic curves of the diode, although, strange as it may seem, that which can be considered the more basic, namely, the emitter temperature—plate current characteristic curve, is infrequently used. In fact, tube manufacturers do not usually show such curves in their tube specifications, the reason being that tubes are designed to be used at a specific value of emitter temperature, as determined by the voltage applied across the filament or heater.

Emitter Temperature—Plate Current Characteristic

To establish the manner in which the plate current of the diode varies with the emitter temperature, it is necessary to set up a circuit using a diode wherein it is possible to create various emitter temperatures and to maintain the voltage applied to the plate at a constant

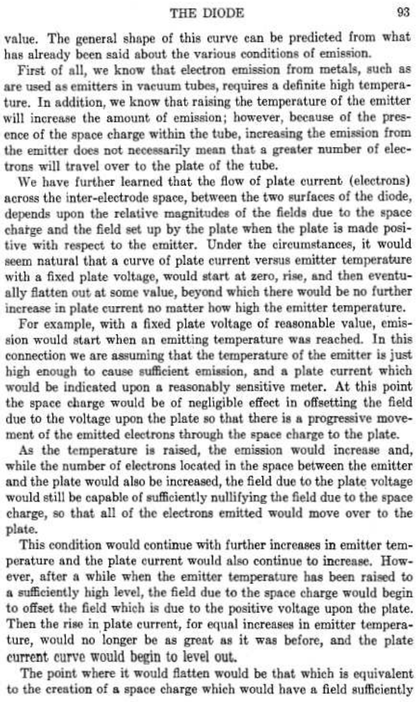

value. The general shape of this curve can be predicted from what has already been said about the various conditions of emission.

First of all, we know that electron emission from metals, such as are used as emitters in vacuum tubes, requires a definite high temperature. In addition, we know that raising the temperature of the emitter will increase the amount of emission; however, because of the presence of the space charge within the tube, increasing the emission from the emitter does not necessarily mean that a greater number of electrons will travel over to the plate of the tube.

We have further learned that the flow of plate current (electrons) across the inter-electrode space, between the two surfaces of the diode, depends upon the relative magnitudes of the fields due to the space charge and the field set up by the plate when the plate is made positive with respect to the emitter. Under the circumstances, it would seem natural that a curve of plate current versus emitter temperature with a fixed plate voltage, would start at zero, rise, and then eventually flatten out at some value, beyond which there would be no further increase in plate current no matter how high the emitter temperature.

For example, with a fixed plate voltage of reasonable value, emission would start when an emitting temperature was reached. In this connection we are assuming that the temperature of the emitter is just high enough to cause sufficient emission, and a plate current which would be indicated upon a reasonably sensitive meter. At this point the space charge would be of negligible effect in offsetting the field due to the voltage upon the plate so that there is a progressive movement of the emitted electrons through the space charge to the plate.

As the temperature is raised, the emission would increase and, while the number of electrons located in the space between the emitter and the plate would also be increased, the field due to the plate voltage would still be capable of sufficiently nullifying the field due to the space charge, so that all of the electrons emitted would move over to the plate.

This condition would continue with further increases in emitter temperature and the plate current would also continue to increase. However, after a while when the emitter temperature has been raised to a sufficiently high level, the field due to the space charge would begin to offset the field which is due to the positive voltage upon the plate. Then the rise in plate current, for equal increases in emitter temperature, would no longer be as great as it was before, and the plate current curve would begin to level out.

The point where it would flatten would be that which is equivalent to the creation of a space charge which would have a field sufficiently

satisfactory for anything but power amplifiers. If supplied from a variable transformer, it becomes a variable supply for all voltages from 0 to 300. Note very carefully that one side of the supply is connected to the AC line, and this must be the grounded side, for your safety, and to avoid ground loops. You cannot ground the positive terminal of this supply to get a negative voltage supply (for use as a C supply, for instance). An isolation transformer, if you have one, would eliminate this hazard. If you don't have an isolation transformer, use a polarized plug to guarantee that the white wire is connected to the circuit ground. If you have a good ground, consider the old trick of connecting only one wire in the power cord, and using the ground to complete the circuit. It is best to observe the power ratings of the resistors and the voltage ratings of the capacitors. This circuit has been tested, except for the fuse. If the 0.5A slow-blow fuse fails, try a 1.0A. This fuse is to turn things off if a capacitor fails; nothing valuable is protected here, but it saves mess.

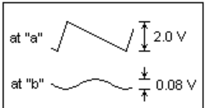

Effect of Ripple Filter

The RC ripple filter is worth the expense. Waveforms are shown at the left. The waveform at node "a" is the familiar one for a "tank" capacitor, and the ripple is fairly large. Since the impedance of a 100 µF capacitor is only 26Ω at 60 Hz, the ripple is reduced by a factor of almost 25. At 300V output and a load of 12 mA, the ripple is less than 0.1V, a very satisfactory result. Note that all that is left in the ripple is the 60 Hz component. The filter would work even better on a full-wave rectifier, but here it is very satisfactory, better and more economical than larger capacitors. Of course, a filter choke could be used for an even better result and less voltage drop, but this would double the cost of the supply.

You will also need a heater transformer, which can be quite small if supplying only one tube that requires 0.3A. The transformer can be loads, should not be used, since they are autotransformers and do not provide isolation. They are, in fact, quite dangerous things, and should be used with great care. The supply was built in a 5" x 7" x 3" aluminum box, with an octal socket for the VR tube on top. The socket can be left vacant when the fixed voltage output is not required.

The voltage regulator requires a certain minimum current (about 5 mA) to function properly. If you are only drawing a few milliamperes from the supply, connect a 12k bleeder resistor across the output. Otherwise, the regulator will not adjust down to the lower voltages. Or, 220Ω and 10k fixed resistors, and a 15k pot, could be used at the voltage regulator, which would draw the necessary minimum current. The VR tube can be replaced by a high-voltage Zener diode.

A 25W isolation transformer is available at the date of writing from All Electronics (See the Your Laboratory page for a link) for $4.50. This transformer is surplus from the Power One firm, and is an excellent value. Solder a jumper between tabs 1 and 3, and another between tabs 2 and 4. The 120V input is connected between 1-3 and 2-4. The output tabs are marked B. This transformer would work well in the circuit above, or it could be put in a box and wired with line cord and output receptacle as a general isolation transformer. It should supply 200 mA without trouble, ample for our purposes.

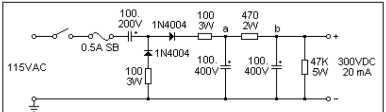

Voltage-Doubler B+ Supply

An idea for an inexpensive B+ supply is shown at the right. The greatest expense is for the capacitors, which will cost about $15. It is based on a half-wave voltage doubler, and gives 300V for a 115V rms input. It cannot supply large currents, but is perfectly

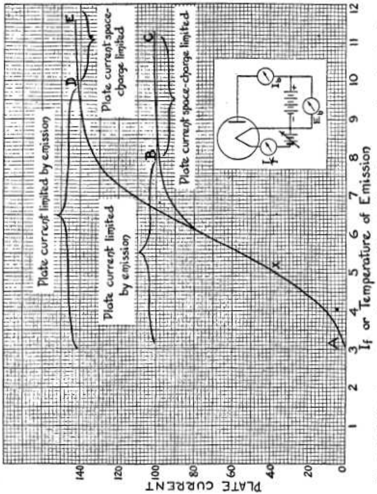

FIG. 6-9. Emitter temperature—plate current characteristic. These curves show that the plate current does not increase appreciably above a certain value even though the temperature of the emitter is increased.

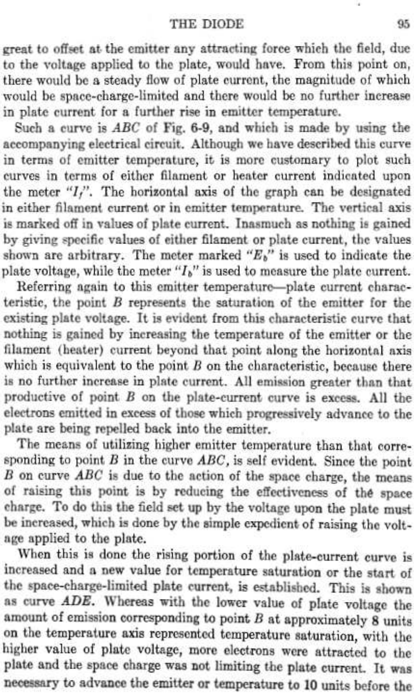

great to offset at the emitter any attracting force which the field, due to the voltage applied to the plate, would have. From this point on, there would be a steady flow of plate current, the magnitude of which would be space-charge-limited and there would be no further increase in plate current for a further rise in emitter temperature.

Such a curve is *ABC* of Fig. 6-9, and which is made by using the accompanying electrical circuit. Although we have described this curve in terms of emitter temperature, it is more customary to plot such curves in terms of either filament or heater current indicated upon the meter "I_f". The horizontal axis of the graph can be designated in either filament current or in emitter temperature. The vertical axis is marked off in values of plate current. Inasmuch as nothing is gained by giving specific values of either filament or plate current, the values shown are arbitrary. The meter marked "E_b" is used to indicate the plate voltage, while the meter "I_b" is used to measure the plate current.

Referring again to this emitter temperature—plate current characteristic, the point *B* represents the saturation of the emitter for the existing plate voltage. It is evident from this characteristic curve that nothing is gained by increasing the temperature of the emitter or the filament (heater) current beyond that point along the horizontal axis which is equivalent to the point *B* on the characteristic, because there is no further increase in plate current. All emission greater than that productive of point *B* on the plate-current curve is excess. All the electrons emitted in excess of those which progressively advance to the plate are being repelled back into the emitter.

The means of utilizing higher emitter temperature than that corresponding to point *B* in the curve *ABC*, is self evident. Since the point *B* on curve *ABC* is due to the action of the space charge, the means of raising this point is by reducing the effectiveness of the space charge. To do this the field set up by the voltage upon the plate must be increased, which is done by the simple expedient of raising the voltage applied to the plate.

When this is done the rising portion of the plate-current curve is increased and a new value for temperature saturation or the start of the space-charge-limited plate current, is established. This is shown as curve *ADE*. Whereas with the lower value of plate voltage the amount of emission corresponding to point *B* at approximately 8 units on the temperature axis represented temperature saturation, with the higher value of plate voltage, more electrons were attracted to the plate and the space charge was not limiting the plate current. It was necessary to advance the emitter or temperature to 10 units before the

well. All the usual resistors and potentiometers are not afraid of 150V, so long as power ratings are observed. 1W and 1/2W resistors may be required in some places. Capacitors must be able to stand the voltages across them; many of those used with transistors will not be adequate. Keep a separate kit of capacitors rated at 100 V and above for this work. High-voltage capacitors are not needed everywhere, only in the plate circuits and for coupling from plate circuits.

The circuit of the laboratory B+ supply that I use for vacuum tubes is shown at the right, and the supply itself is shown in the photograph below. It provides a regulated variable 60-150 V output, and a regulated fixed voltage output (for screen supply) created by a VR tube. VR tubes can be exchanged for different voltages. The MagneTek N-51X 115 V-115 V isolation transformer is available from Antique Electronics (see references), and a cheaper one from All Electronics. The transformer secondary has a DC resistance of 22Ω, which limits the surge current satisfactorily without having to add a series resistor. By no means eliminate the isolation transformer and use the 120 V household supply directly, because of the ground hazard. A variable transformer (Variac) is an *autotransformer* that does not isolate the output from the power line ground. I earnestly recommend that you do not work on AC circuits without isolating them from the service ground. The 110/220 adapters commonly available in 50W and 300W sizes, used for shavers and other small

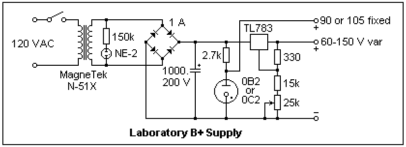

Laboratory B+ Supply

through the discharge in the "normal glow" region, in which the glow does not completely cover the cathode, and expands to accommodate more current. Tubes were manufactured for voltages of 75, 90, 105 and 150 that were used like Zener diodes, handling from 5 to 40 mA. There is more information on glow discharges in Relaxation Oscillators, and on Zeners in Voltage Regulators. VR tubes are treated here because of their association with vacuum tubes, and the higher voltages involved.

Experiments

Vacuum tubes generally operate at higher voltages than transistor circuits. Like transistors, vacuum tubes are happier at higher voltages, which for receiving-type tubes, typically would be, say, 200 to 250V. It was once quite common to make DC power supplies for such voltages, using a transformer with a center-tapped secondary (say 200-0-200V), and a rectifier with double anodes and a common cathode, feeding a filter consisting of capacitors of 8 or 16 μF, and a choke of 10 H or so. It was not convenient to make a bridge rectifier with vacuum diodes (three separate filament transformers are necessary), so full-wave rectification with a center-tapped secondary was usual. These days it is rather difficult (and expensive) to acquire all these things, with the possible exception of the capacitors, which are now available up to millifarads at voltages up to 450 V.

The voltages normally used with receiving tubes are not high enough to be really dangerous, though a shock will not improve your day. If you are eager for new experiences, I can save you the trouble of finding out by saying that a DC shock is kind of like a hammer blow, not the zap of an AC shock, and does not paralyze, as an AC shock does. Shocks are given by current, not simple contact, so a good and old rule is to work with one hand in your pocket around high voltage. Always turn things off before making any adjustments or changes, of course, and be neat. Avoid touching bare metal. With these precautions and normal care, you will be fine. The solderless breadboard, DMM and oscilloscope can handle these voltages quite

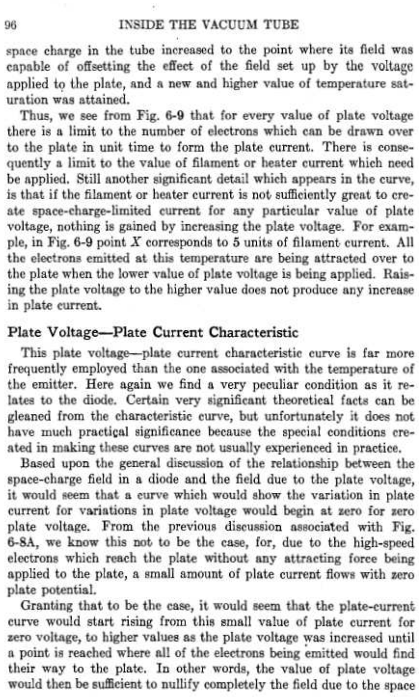

space charge in the tube increased to the point where its field was capable of offsetting the effect of the field set up by the voltage applied to the plate, and a new and higher value of temperature saturation was attained.

Thus, we see from Fig. 6-9 that for every value of plate voltage there is a limit to the number of electrons which can be drawn over to the plate in unit time to form the plate current. There is consequently a limit to the value of filament or heater current which need be applied. Still another significant detail which appears in the curve, is that if the filament or heater current is not sufficiently great to create space-charge-limited current for any particular value of plate voltage, nothing is gained by increasing the plate voltage. For example, in Fig. 6-9 point *X* corresponds to 5 units of filament current. All the electrons emitted at this temperature are being attracted over to the plate when the lower value of plate voltage is being applied. Raising the plate voltage to the higher value does not produce any increase in plate current.

Plate Voltage—Plate Current Characteristic

This plate voltage—plate current characteristic curve is far more frequently employed than the one associated with the temperature of the emitter. Here again we find a very peculiar condition as it relates to the diode. Certain very significant theoretical facts can be gleaned from the characteristic curve, but unfortunately it does not have much practical significance because the special conditions created in making these curves are not usually experienced in practice.

Based upon the general discussion of the relationship between the space-charge field in a diode and the field due to the plate voltage, it would seem that a curve which would show the variation in plate current for variations in plate voltage would begin at zero for zero plate voltage. From the previous discussion associated with Fig. 6-8A, we know this not to be the case, for, due to the high-speed electrons which reach the plate without any attracting force being applied to the plate, a small amount of plate current flows with zero plate potential.

Granting that to be the case, it would seem that the plate-current curve would start rising from this small value of plate current for zero voltage, to higher values as the plate voltage was increased until a point is reached where all of the electrons being emitted would find their way to the plate. In other words, the value of plate voltage would then be sufficient to nullify completely the field due to the space

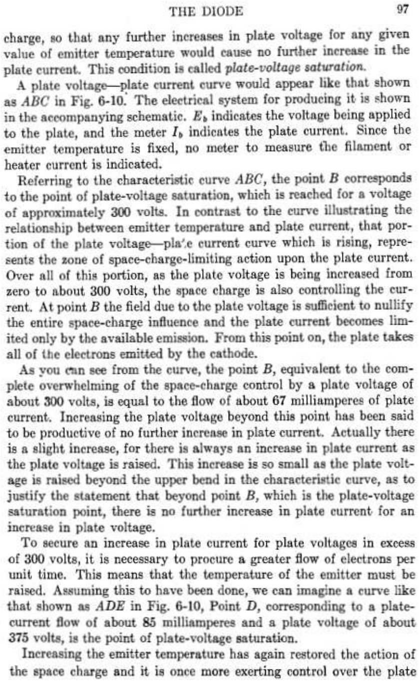

THE DIODE 97

charge, so that any further increases in plate voltage for any given value of emitter temperature would cause no further increase in the plate current. This condition is called *plate-voltage saturation.*

A plate voltage—plate current curve would appear like that shown as *ABC* in Fig. 6-10. The electrical system for producing it is shown in the accompanying schematic. E_b indicates the voltage being applied to the plate, and the meter I_b indicates the plate current. Since the emitter temperature is fixed, no meter to measure the filament or heater current is indicated.

Referring to the characteristic curve *ABC*, the point *B* corresponds to the point of plate-voltage saturation, which is reached for a voltage of approximately 300 volts. In contrast to the curve illustrating the relationship between emitter temperature and plate current, that portion of the plate voltage—plate current curve which is rising, represents the zone of space-charge-limiting action upon the plate current. Over all of this portion, as the plate voltage is being increased from zero to about 300 volts, the space charge is also controlling the current. At point *B* the field due to the plate voltage is sufficient to nullify the entire space-charge influence and the plate current becomes limited only by the available emission. From this point on, the plate takes all of the electrons emitted by the cathode.

As you can see from the curve, the point *B*, equivalent to the complete overwhelming of the space-charge control by a plate voltage of about 300 volts, is equal to the flow of about 67 milliamperes of plate current. Increasing the plate voltage beyond this point has been said to be productive of no further increase in plate current. Actually there is a slight increase, for there is always an increase in plate current as the plate voltage is raised. This increase is so small as the plate voltage is raised beyond the upper bend in the characteristic curve, as to justify the statement that beyond point *B*, which is the plate-voltage saturation point, there is no further increase in plate current for an increase in plate voltage.

To secure an increase in plate current for plate voltages in excess of 300 volts, it is necessary to procure a greater flow of electrons per unit time. This means that the temperature of the emitter must be raised. Assuming this to have been done, we can imagine a curve like that shown as *ADE* in Fig. 6-10, Point *D*, corresponding to a plate-current flow of about 85 milliamperes and a plate voltage of about 375 volts, is the point of plate-voltage saturation.

Increasing the emitter temperature has again restored the action of the space charge and it is once more exerting control over the plate

gain in response to increased negative grid bias, which was used for AGC (automatic gain control) in IF amplifiers. The 6SK7 was a very popular remote cutoff pentode used as an RF and IF amplifier. The 6SJ7, on the other hand, was a *sharp-cutoff* pentode, used as an audio voltage amplifier. The amplification factor has little significance with pentodes, as with transistors, and transconductance is the important parameter. The screen grid also acted as an electrostatic shield between control grid and plate, reducing the Miller capacitance to extremely small values, 0.003 pF in the 6SK7. If the screen and suppressor grid are connected to the plate, the pentode operates as a triode.

A very curious and ingenious kind of tube was the *electron-ray* tube, used on receivers to give a visual indication of the accuracy of tuning to a station. Don't confuse it with the *cathode-ray* tube that uses a guided electron beam for oscilloscopes and TV receivers. It showed a luminous disk, with a dark sector. The dark sector was made as small as possible to achieve accurate tuning. It worked from the AGC (automatic gain control) voltage of the receiver. This is a feedback signal that tries to keep the signal amplitude constant at the output of the intermediate frequency amplifiers, increasing the gain for weak signals and decreasing it for strong. It is usually a negative voltage produced by rectifying the IF output. The tube has a thermionic cathode and a conical anode or *target* covered with cathodoluminescent phosphor (like a CRT), which glows from the 3 or 4 mA of plate current that flows when it is across 125V or more (up to 250). Control electrodes, of which there are two on opposite sides of the 6AF6, make two dark sectors that are widest when at 0V, and narrow as the control voltage approaches the target voltage. The control voltage is typically provided by the plate of a triode controlled by the AVC, such that full negative AVC cuts off the triode and makes the sector as small as possible. All this was cheaper and more graphic than a pointer meter.

Another kind of tube that we'll look at here is the glow-tube voltage regulator. The voltage across a glow discharge depends on the gas and the cathode material, and is almost idependent of the current

Although tetrodes worked as expected, they had a serious defect. It happens that speedy electrons colliding with the plate knock out *secondary* electrons. In a triode, these are rapidly sucked back to the positive plate, and the same happened in a tetrode when the plate potential was higher than the screen grid potential. In normal operation however, especially with large voltage gain, the plate voltage has a large swing, and can become less positive than the screen grid. Now all these secondary electrons (and some of the primary ones, too) are attracted to the screen grid, and there is a definite sag in the characteristic in this region. To prevent this, it is necessary to establish an electric field at the plate that is always directed toward the plate, to suppress the escape of secondary electrons. This is provided by a third grid, the *suppressor grid*, which is usually connected to the cathode. The tube with three grids: control, screen and suppressor, or grids 1, 2 and 3, is called a *pentode*, which turns out to be a superior voltage amplifier, fully equivalent to a transistor. A typical small pentode, the 6SJ7, has a plate resistance of over a megohm, and a transconductance of 1.6 mS.

An ingenious modification of the pentode has electrodes that shape and concentrate the electron beam instead of a suppressor grid, the negative space charge of the electrons doing the same work. These are called *beam power* tubes, and were good for power work, as the name indicates. A typical example, the 6L6, had a transconductance as high as 6.0 mS, and the smaller 6V6 about 4.0 mS. Both types were widely used for high-fidelity audio amplifiers, and tube amplifiers still have proponents. The same tubes were used in small amateur radio transmitters, which shows the versatility of vacuum tubes.

Receiving pentodes were also classified as *sharp cutoff* or *remote cutoff*, an example of designing tubes to fit their applications. A remote cutoff pentode had a grid with variable spacing, so that areas of wider spacing let electrons through when the grid was made more negative, when areas of smaller spacing were cut off. This effectively reduced the transconductance of the tube, decreasing its

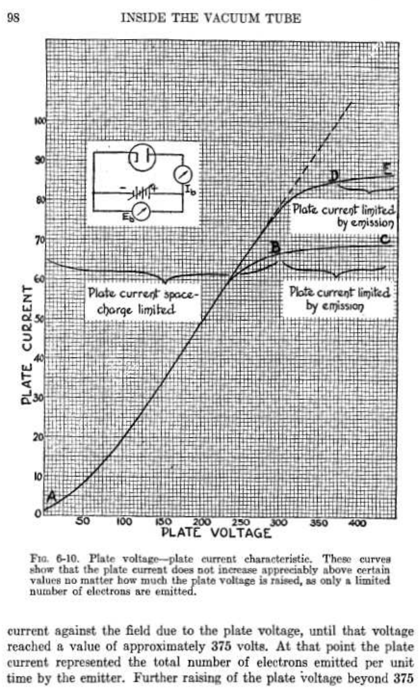

98 INSIDE THE VACUUM TUBE

PLATE CURRENT
PLATE VOLTAGE
0
10
20
30
40
50
60
70
80
90
100
50
100
150
200
250
300
350
400
A
B
C
D
E
I_b
E_b
Plate current limited by emission
Plate current space-charge limited
Plate current limited by emission

FIG. 6-10. Plate voltage—plate current characteristic. These curves show that the plate current does not increase appreciably above certain values no matter how much the plate voltage is raised, as only a limited number of electrons are emitted.

current against the field due to the plate voltage, until that voltage reached a value of approximately **375** volts. At that point the plate current represented the total number of electrons emitted per unit time by the emitter. Further raising of the plate voltage beyond **375**

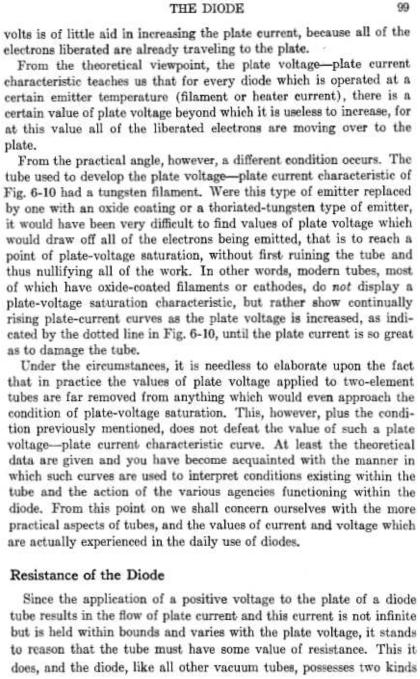

volts is of little aid in increasing the plate current, because all of the electrons liberated are already traveling to the plate.

From the theoretical viewpoint, the plate voltage—plate current characteristic teaches us that for every diode which is operated at a certain emitter temperature (filament or heater current), there is a certain value of plate voltage beyond which it is useless to increase, for at this value all of the liberated electrons are moving over to the plate.

From the practical angle, however, a different condition occurs. The tube used to develop the plate voltage—plate current characteristic of Fig. 6-10 had a tungsten filament. Were this type of emitter replaced by one with an oxide coating or a thoriated-tungsten type of emitter, it would have been very difficult to find values of plate voltage which would draw off all of the electrons being emitted, that is to reach a point of plate-voltage saturation, without first ruining the tube and thus nullifying all of the work. In other words, modern tubes, most of which have oxide-coated filaments or cathodes, do *not* display a plate-voltage saturation characteristic, but rather show continually rising plate-current curves as the plate voltage is increased, as indicated by the dotted line in Fig. 6-10, until the plate current is so great as to damage the tube.

Under the circumstances, it is needless to elaborate upon the fact that in practice the values of plate voltage applied to two-element tubes are far removed from anything which would even approach the condition of plate-voltage saturation. This, however, plus the condition previously mentioned, does not defeat the value of such a plate voltage—plate current characteristic curve. At least the theoretical data are given and you have become acquainted with the manner in which such curves are used to interpret conditions existing within the tube and the action of the various agencies functioning within the diode. From this point on we shall concern ourselves with the more practical aspects of tubes, and the values of current and voltage which are actually experienced in the daily use of diodes.

Resistance of the Diode

Since the application of a positive voltage to the plate of a diode tube results in the flow of plate current and this current is not infinite but is held within bounds and varies with the plate voltage, it stands to reason that the tube must have some value of resistance. This it does, and the diode, like all other vacuum tubes, possesses two kinds

It's easy to see that the maximum voltage gain achievable with a triode is μ, if the load resistance is much higher than the plate resistance. The rule we derived for the gain of a transistor amplifier as the ratio of the collector and emitter resistances holds here as well, expressed by $\mu = g_m r_p$. The quantity analogous to r_e is $1/g_m$, which is 333Ω for the 6J5. Mu is a rather modest number, so triodes are not good for high voltage gain. They make very good power amplifiers, however, since large currents can be controlled. Using the Thévenin source, the gain of a usual common-cathode amplifier (analogous to a common-emitter amplifier) is simply a voltage divider problem. There must be a plate resistor in series with the plate, or else the voltage would never change, but it should be as large as possible, and has only a small effect on the gain. We shall examine these circuits in detail in the experiments, but this provides the background.

In order to provide higher voltage gain, the plate resistance must be reduced somehow. We recall that with a transistor, the analogous collector resistance was very large, and there was no problem with voltage gain. The plate resistance is the result of the effect of plate voltage on the space charge. This effect is not necessary for control, which is provided by the control grid, so what we need is to eliminate the effect of the plate voltage on the space charge. This is done by introducing another grid, the *screen grid* between the control grid and the plate. If this grid is held at a constant potential, the space charge is "screened" from the effects of changes in plate voltage. The screen grid is usually bypassed to ground by a capacitor, whose reactance at the lower corner frequency should be smaller than the resistance connecting the screen grid to B+. Some, but few, electrons are removed by the screen grid, since it is again a coil of fine wire. With this change, the plate characteristics become (nearly) horizontal lines, as for the transistor, and the plate resistance becomes large, approaching a megohm. The resulting tube is called a *screen-grid tetrode*.

voltage electrolytic. Only part of the cathode resistor can be bypassed if some feedback is desired.

A typical triode, of which the 6J5 is chosen here as the example, has a $\mu = 20$, $r_p = 6.7k$ and $g_m = 3.0$ mS. Of course, the relation $\mu = g_p m_p$ always holds. The value of transconductance may seem small, compared to a transistor, but it should be remembered that it refers to a high-voltage plate circuit, and that the input impedance is infinite, so the power amplification is extremely large. The 6J5 is called a "medium-mu" triode. The similar 6SL7, a "high mu" triode, has $\mu = 70$, $r_p = 44k$ and $g_m = 1.6$ mS. The 6SL7 is a *dual* triode, two independent valves sharing the same heater. The dual version of the 6J5 is the 6SN7.

The maximum plate voltage of the 6J5 is 300V, and the maximum plate current is 20 mA. Its maximum *plate dissipation* is 2.5W (product of average plate current and average plate voltage). This gives an idea of the ratings of receiving tubes used as voltage amplifiers. As power amplifiers, the allowable plate currents can be quite a bit larger, and hundreds of watts output is possible with relatively small tubes. The interelectrode capacitances of the 6J5 are on the order of 3.4 pF, and are significant at high frequencies.

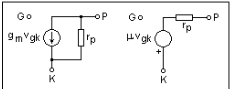

Triode Equivalent Circuits

The small-signal equivalent circuits for the triode are shown at the right. The Norton source circuit is exactly the one for the FET. In the case of the triode, the plate resistance is always important, and cannot be neglected. The other circuit is just the Thévenin source corresponding to the Norton source. It shows the significance of the amplification factor, and is useful for triodes because of the rather small plate resistance. We did not find the Th%eacute;venin source for transistors very useful, and did not introduce a parameter analagous to the amplification factor for this reason.

---

of opposition to the flow of current: *d-c plate resistance and a-c plate resistance.* Of these two classifications, the latter is by far the most important.

D-C Plate Resistance of the Diode

As the name indicates, the d-c plate resistance of the diode is that opposition to the flow of current which is offered by the tube when a d-c voltage is applied to the plate. It is that quantity which can be determined by the application of Ohm's law for resistance, employing the applied d-c plate voltage and the measured value of d-c plate current as the known quantities in the equation.

For example, Fig. 6-11 illustrates a typical plate voltage—plate current characteristic for a typical 6H6 duo-diode tube, using just one pair of elements. By one pair of elements is meant one cathode and its associated plate. The plate voltage applied to the tube is shown upon the horizontal axis of the graph and the resultant plate current is shown upon the vertical axis. While this is the same as is found in the plate voltage—plate current characteristic curve illustrated in Fig. 6-10, the general shape of the two characteristic curves is quite different. This comes about as the result of the special conditions mentioned in connection with Fig. 6-10. The curve shown in Fig. 6-11 is a closer approach to the actual curve with which you will work.

You will remember that we qualified the conditions indicated in the curve of Fig. 6-10 to the extent that the plate-voltage saturation points were not to be expected when using oxide-coated filament or cathode type tubes. The 6H6 is an oxide-coated tube and it is evident in Fig. 6-11 that plate-voltage saturation is a long way off. In fact, judging by the values of plate current which prevail for the comparatively low values of plate voltage, it is doubtful if the plate-voltage saturation point could ever be reached. More than likely the tube would be damaged long before that value could be reached by the application of sufficient plate voltage.

With a comparatively low value of plate voltage, say 28 volts, applied to the plate, the formidable value of about 66.4 milliamperes of current already flows in the tube. And judging by the slope of the plate-current curve, the steep portion is yet to come, that is, if we increased the plate voltage above 32 volts. There is little doubt that were this plate voltage increased to say 50 or 60 volts, the plate current would rise to such a value as to destroy the tube. This does not mean that diodes are not used at plate-voltage values greater than 32 volts. Much higher values of plate voltage are employed, but when

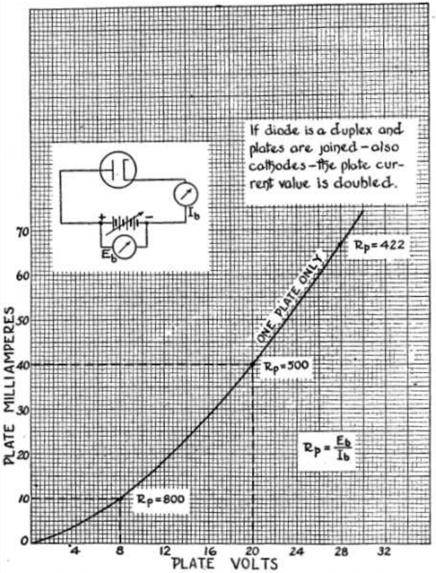

FIG. 6-11. Plate voltage—plate current characteristic for a typical 6H6 diode tube.

A vacuum tube carrying a current I with a plate voltage V dissipates power VI, just as if it were a resistor. However, the process is different. In a resistor, an electron gives up small amounts of energy to the lattice as it is accelerated and then is scattered. In a vacuum tube, the electron acquires a kinetic energy as it is accelerated, which it gives up all at once when it collides with the plate. It is not really correct to ascribe this to the "plate resistance," as some texts do, which is an incremental ratio. Since the plate is in a vacuum, the resulting heat can only be radiated or conducted down the supports. Much of the radiated heat is infrared, which is absorbed by the glass tube envelope. Note how plates are blackened to raise their emissivity and often provided with fins. Really large tubes had plates externally exposed capable of air or water cooling with elaborate seals to the glass parts. *Plate dissipation* is always a limiting factor in power applications.

The line marked *load line* shows the difference between the supply voltage V_{bb} and the voltage drop in a resistance R_L in series with the tube at any plate current, giving the plate voltage directly. The series resistance is mainly a plate resistor, but if there is a cathode resistor (for purposes of biasing) it should be included. There are generally different load lines for static (DC) and dynamic (AC) operation. As the grid voltage is varied, the plate current and voltage vary along the load line. The quiescent or operating point can be selected at some point along the DC load line, and so the DC grid bias can be found. This grid bias can be obtained from a C battery or equivalent, or from a cathode resistor, just as an emitter resistor is used with a transistor.

A cathode bias resistor is often bypassed by a capacitor if its negative feedback effect is not desired in dynamic operation. The reactance of the capacitor at the lower corner frequency should be equal to the resistance looking into the cathode (normally $1/g_m$ in parallel with the cathode resistor). The size of the cathode resistor has only a small effect on the size of the bypass capacitor. The capacitor never has a large voltage across it, and can be a low-types can have quite different (though similar) characteristics, so characteristic curves are much more important.

Let's begin with idealized plate characteristics for a triode, shown at the right. These are curved lines, but we represent them by straight lines for ease of understanding the various slopes and distances involved, which will be constant. Actual characteristics are not really far from straight lines, anyway. There is one curve for each grid voltage represented, which differ by a constant amount, here 2V. The horizontal distance between them, represented by the line labelled μ, is the amount the plate voltage must increase to hold the plate current constant; it represents the relative influence of plate and grid on plate current, and is the *amplification factor*. The vertical distance between them shows how much the plate current changes for a change in grid voltage. The ratio is a conductance, called the *transconductance*, denoted by g_m, measured in siemens (mho). The slope of the curves is the ratio g_m/μ, called the *plate conductance* g_p. With vacuum tubes, the reciprocal r_p was always used, called the *plate resistance*. Since actual characteristics are curved, these quantities vary for different currents and voltages.

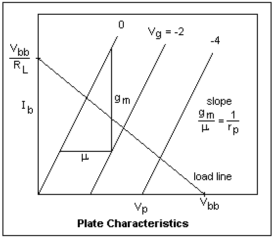

Plate Characteristics

What we desire to represent is a function of two independent variables $I_p(V_p,V_g)$, which can be represented as a surface in three dimensions. Our various characteristic curves are orthogonal views along one or another of the axes. There are three such views possible, each directly related to one of the three parameters, of which we generally use only two, the ones mentioned here.

that is done other elements which tend to keep the plate current to a safe value are also used in the circuit. Such additions will be discussed later.

As we discussed earlier in this text, the plate-current curve of a diode, like that of other vacuum tubes, is of great importance. In this particular instance it serves to furnish information relative to the opposition offered by the tube to the flow of plate current; in other words, it furnishes information about the d-c resistance of the tube. For example, with **8** volts d-c E_b applied to the plate, the plate current I_b, as shown upon the vertical axis, is **10** milliamperes or **0.01** ampere. According to the Ohm's law, these values of plate voltage and plate current represent a d-c plate resistance of

$$R_p = \frac{E_b}{I_b} = \frac{8}{.01} = 800 \text{ ohms}$$

With **20** volts d-c applied to the plate of the tube, we note from the curve that the plate current is **40** milliamperes or **0.04** ampere. This corresponds to a d-c resistance equal to **20/.04 = 500** ohms. With **28** volts d-c applied to the plate, the plate current according to the curve is **66.4** milliamperes or **0.064** ampere and the equivalent d-c resistance is **28/.0664 = 422** ohms.

Examining these three values of d-c plate resistance for three values of applied d-c plate voltage, we note a peculiar condition. The opposition which *the diode offers to the flow of the plate current is not constant*, as we are ordinarily accustomed to experiencing in conventional d-c systems. Said differently, *the resistance offered by the diode to the flow of the plate current is not linear*. Judging by the relationship between the plate voltage, plate current, and the equivalent d-c resistance, it appears that the resistance of the diode *decreases* as the plate voltage is increased, and *increases* as the plate voltage is decreased. Were the resistance of the diode linear over the entire range of plate voltage, the plate-current curve would be a straight line instead of the curved line appearing in Fig. 6-11. Such a straight line would indicate that the d-c plate resistance remained constant over whatever range of plate voltage and plate current is embraced by the upper and lower limits of the straight line.

Of further interest in connection with the plate voltage—plate current characteristic given in Fig. 6-11, is the condition created when the two plates of this duo-diode tube are joined in parallel and the two cathodes are connected in parallel. The plate current for such parallel connection is equal to twice the amount that would occur if the same value of plate voltage were applied to a single pair of elements. Thus,

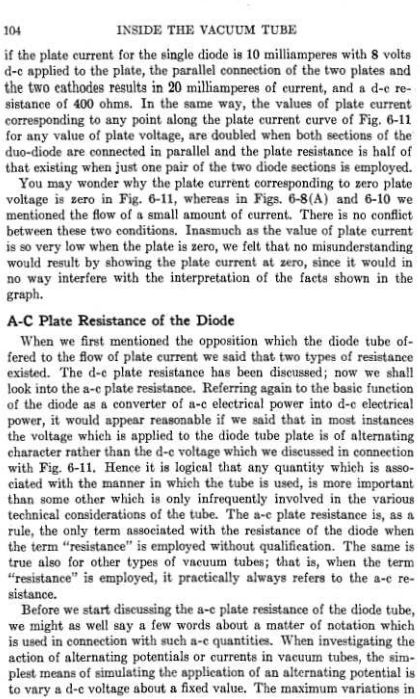

if the plate current for the single diode is 10 milliamperes with 8 volts d-c applied to the plate, the parallel connection of the two plates and the two cathodes results in 20 milliamperes of current, and a d-c resistance of 400 ohms. In the same way, the values of plate current corresponding to any point along the plate current curve of Fig. 6-11 for any value of plate voltage, are doubled when both sections of the duo-diode are connected in parallel and the plate resistance is half of that existing when just one pair of the two diode sections is employed.

You may wonder why the plate current corresponding to zero plate voltage is zero in Fig. 6-11, whereas in Figs. 6-8(A) and 6-10 we mentioned the flow of a small amount of current. There is no conflict between these two conditions. Inasmuch as the value of plate current is so very low when the plate is zero, we felt that no misunderstanding would result by showing the plate current at zero, since it would in no way interfere with the interpretation of the facts shown in the graph.

A-C Plate Resistance of the Diode

When we first mentioned the opposition which the diode tube offered to the flow of plate current we said that two types of resistance existed. The d-c plate resistance has been discussed; now we shall look into the a-c plate resistance. Referring again to the basic function of the diode as a converter of a-c electrical power into d-c electrical power, it would appear reasonable if we said that in most instances the voltage which is applied to the diode tube plate is of alternating character rather than the d-c voltage which we discussed in connection with Fig. 6-11. Hence it is logical that any quantity which is associated with the manner in which the tube is used, is more important than some other which is only infrequently involved in the various technical considerations of the tube. The a-c plate resistance is, as a rule, the only term associated with the resistance of the diode when the term "resistance" is employed without qualification. The same is true also for other types of vacuum tubes; that is, when the term "resistance" is employed, it practically always refers to the a-c resistance.

Before we start discussing the a-c plate resistance of the diode tube, we might as well say a few words about a matter of notation which is used in connection with such a-c quantities. When investigating the action of alternating potentials or currents in vacuum tubes, the simplest means of simulating the application of an alternating potential is to vary a d-c voltage about a fixed value. The maximum variations in

was arbitrary. Filaments were customarily used, especially in power and rectifier tubes, because they gave more current per watt of heating power. The indirectly heated, equipotential cathode that could be supplied by AC rather than by battery power was widely used after 1930.

A tube with just cathode and anode is called a *diode*, a term that has survived into the semiconductor age. Diodes were used for power or signal rectification, just like their semiconductor relatives. A "full-wave" diode has two anodes. When a control grid is provided, the tube is called a *triode*, and is used for amplification. Let's first study the peculiar circuit behavior of triodes, which will lead us to the reason for the addition of more grids, and the creation of the *pentode*, which turns out to act very much like a transistor.

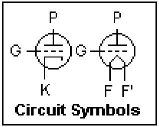

Circuit symbols for a triode are shown at the right, and other tube symbols are derived from it. The connections are plate P, grid G, and cathode K or filament F, F'. These are analogous to the collector, base and emitter of a transistor, with the same polarities and direction of current flow as an NPN transistor. The circle is a part of the symbol. Grid connections can be to the right or left of the symbol, as convenient. A gas tube is indicated by a dot in the lower right-hand part of the circle. The heater of an indirectly heated cathode is usually not shown. A cold cathode (operating by positive-ion bombardment) is shown as a small circle. There are examples of these symbols below.

The important variables are the independent variables V_p and V_g, the plate and grid voltages (with respect to the cathode), and the dependent variable I_p, the plate current. A plot of I_p against V_p for a fixed value of V_g is called a *plate characteristic*, and a plot of I_b against V_g for a fixed value of V_p is called a *transfer characteristic*. From a family of either characteristics, the complete circuit behavior of the tube can be predicted. Unlike transistors, tubes of different usually). Inside is a tungsten heater wire insulated from the cathode with BeO or alundum (aluminum oxide) ceramic insulation. These cathodes must be heated only to about 850K (a dull red) to emit electrons in the amount necessary. Most receiving tubes require 6.3V or 12.6V for their heaters, at about 0.30A or 0.15A, respectively. Every tube type is identified by a type number, such as 6J5, where the first number indicates the heater voltage. 6 means 6.3V, 12 means 12.6V. Heaters are very forgiving of variations in voltage, but it is best to try to use the recommended voltages. AC is generally used, supplied from a small transformer. It is necessary to make sure that the difference in voltage between heater and cathode does not exceed 90V, so the heater AC supply should be grounded. It is usual to ground the center tap on the transformer for this purpose. Tubes for battery radios have plain filaments that are both heater and emitter, and must be supplied with DC. Their type numbers begin with "1" and are intended to be used with a standard dry cell of 1.5V. Larger tubes for high voltages have to use thoriated-tungsten filaments at 2000K,a bright yellow, to avoid damage from positive-ion bombardment, small as it may be. Heaters are sometimes called filaments, and the heater supply the filament supply, out of linguistic inertia.

Before the 1930's, each manufacturer used arbitrary designations for his tubes, and there was no uniformity or system. Tubes were first systematically identified in the U.S. by three-digit numbers, where the first digit denoted the manufacturer. For example, a type x10 was a power triode, a type x36 a screen-grid tetrode, where x was the manufacturer's number (usually omitted). Later, a new system was introduced where the first digit gave the filament or heater voltage, and the last digit gave the number of functional electrodes. This scheme was introduced by the RMA (Radio Manufacturer's Association) in 1934. A letter between these digits was assigned in order of introduction. For example, a 2A3 was a power triode with a 2.5 V filament (this popular tube is, remarkably, still in use for hi-fi amplifiers because of its low distortion!). This system was not comprehensive enough, and in the final system, the first number designated the heater voltage, but the remainder of the designation

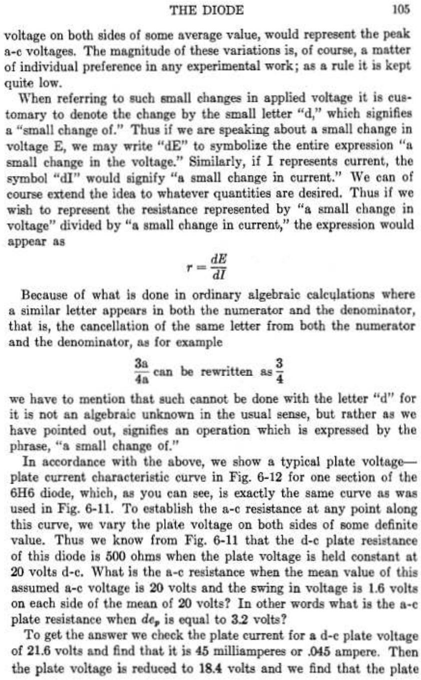

voltage on both sides of some average value, would represent the peak a-c voltages. The magnitude of these variations is, of course, a matter of individual preference in any experimental work; as a rule it is kept quite low.

When referring to such small changes in applied voltage it is customary to denote the change by the small letter "d," which signifies a "small change of." Thus if we are speaking about a small change in voltage E, we may write "dE" to symbolize the entire expression "a small change in the voltage." Similarly, if I represents current, the symbol "dI" would signify "a small change in current." We can of course extend the idea to whatever quantities are desired. Thus if we wish to represent the resistance represented by "a small change in voltage" divided by "a small change in current," the expression would appear as

$$r = \frac{dE}{dI}$$

Because of what is done in ordinary algebraic calculations where a similar letter appears in both the numerator and the denominator, that is, the cancellation of the same letter from both the numerator and the denominator, as for example

$$\frac{3a}{4a} \text{ can be rewritten as } \frac{3}{4}$$

we have to mention that such cannot be done with the letter "d" for it is not an algebraic unknown in the usual sense, but rather as we have pointed out, signifies an operation which is expressed by the phrase, "a small change of."

In accordance with the above, we show a typical plate voltage—plate current characteristic curve in Fig. 6-12 for one section of the 6H6 diode, which, as you can see, is exactly the same curve as was used in Fig. 6-11. To establish the a-c resistance at any point along this curve, we vary the plate voltage on both sides of some definite value. Thus we know from Fig. 6-11 that the d-c plate resistance of this diode is 500 ohms when the plate voltage is held constant at 20 volts d-c. What is the a-c resistance when the mean value of this assumed a-c voltage is 20 volts and the swing in voltage is 1.6 volts on each side of the mean of 20 volts? In other words what is the a-c plate resistance when de_p is equal to 3.2 volts?

To get the answer we check the plate current for a d-c plate voltage of 21.6 volts and find that it is 45 milliamperes or .045 ampere. Then the plate voltage is reduced to 18.4 volts and we find that the plate

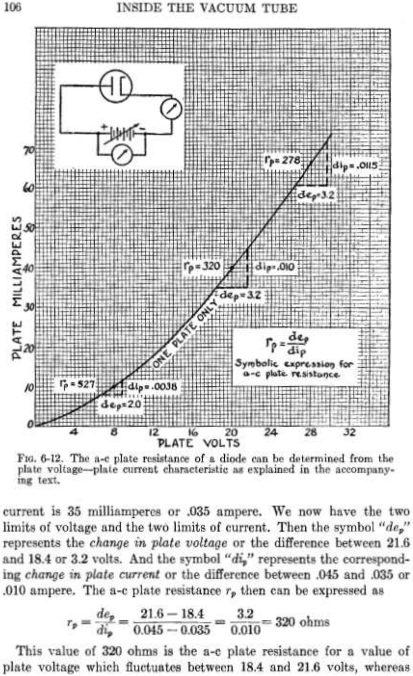

PLATE MILLIAMPERES
PLATE VOLTS
0
10
20
30
40
50
60
70
4
8
12
16
20
24
28
32
$r_p = 278$
$di_p = .0115$
$de_p = 3.2$
$r_p = 320$
$di_p = .010$
$de_p = 3.2$
$r_p = 527$
$di_p = .0038$
$de_p = 2.0$
ONE PLATE ONLY
$r_p = \frac{de_p}{di_p}$
Symbolic expression for a-c plate resistance

FIG. 6-12. The a-c plate resistance of a diode can be determined from the plate voltage—plate current characteristic as explained in the accompanying text.

current is 35 milliamperes or .035 ampere. We now have the two limits of voltage and the two limits of current. Then the symbol "de_p" represents the *change in plate voltage* or the difference between 21.6 and 18.4 or 3.2 volts. And the symbol "di_p" represents the corresponding *change in plate current* or the difference between .045 and .035 or .010 ampere. The a-c plate resistance r_p then can be expressed as

$$r_p = \frac{de_p}{di_p} = \frac{21.6 - 18.4}{0.045 - 0.035} = \frac{3.2}{0.010} = 320 \text{ ohms}$$

This value of 320 ohms is the a-c plate resistance for a value of plate voltage which fluctuates between 18.4 and 21.6 volts, whereas

can take over a large part of the electron emission at the cathode. Such tubes make efficient rectifiers, and the gas pressure can be quite low, as in some rectifiers, or rather high, as in a mercury-arc rectifier. In receiving tubes, positive ion collisions can destroy the delicate, high-efficiency cathode surface. Positive ions also cause small currents to negative electrodes that otherwise might be expected to carry no current at all. For all these reasons, receiving tubes have a high vacuum.

The electric field at the space charge that controls the current does not have to be created by the anode alone. A third electrode, the *grid*, is placed between the cathode and the anode, closer to the cathode. It is made of a spiral of fine wire, so electrons can pass through without hindrance. When it is made negative, it opposes the effect of the anode in creating an electric field, but does not attract any electrons, and so draws no current (except for the positive-ion current mentioned above). If it is made sufficiently negative, it can *cut off* the plate current entirely. If it is made positive, it can enhance the plate current, but then draws some *grid current* itself. The grid provides a sensitive control, using negligible power, of the large plate current, so the vacuum tube is a powerful amplifying device.

Early radio sets were battery-powered (domestic electrification was in its infancy, and absent in rural areas, when radio began), and a convention was established for identifying the batteries required. The filaments required low voltages at high currents (2W or more each), and their supply was called the A battery. The plates required high voltages at small currents, perhaps 90V at a few mA, and their supply was called the B battery. Grid bias was required, to hold the grids negative, demanding low voltages at small currents, and the corresponding battery was the C battery. The notations A and C were later little-used (except for actual battery radios), but the plate supppy came to be generally known as B+, and the letter B appears in subscripts of quantities referring to the plate circuit.

The cathodes of receiving tubes consist of a sleeve of nickel alloy coated with a compound of alkaline-earth oxides (Ba and Sr, did these cathodes have a long life, but were also *equipotential*, making circuit design simpler. Thoriated tungsten remained for transmitting tubes, where a rugged emitter was necessary because of the higher plate voltages, but even here tungsten was the only suitable choice for really high voltages, to avoid damage from positive-ion bombardment.

The rate of emission of electrons from a heated metal is given by the Richardson-Dushman equation, $i = AT^2e^{-b/T}$ A/cm^2, where T is the absolute temperature in K, and A and b are constants typical of the emitter. For tungsten, A = 60 and b = 52,400K, while average values for an oxide cathode are A=0.01, b = 11,600. The exponential factor has by far the largest influence, so emission increases rapidly with temperature. This makes thermionic cathodes very suitable even for heavy currents. In all tubes, electrons are emitted in far greater numbers than required; most simply return to the cathode.

The electrons emitted by the thermionic cathode form a negative *space charge* cloud around the cathode, dense enough that if no electrons are removed by attraction to the anode, the rate of emission is equal to the rate of return. When the anode is made positive, some of the electrons are attracted to it out of the space-charge cloud, and a *thermionic current* results. The amount of this current is given by $I = A\ V^{3/2}$, where V is the voltage from anode to cathode. This is called the Langmuir-Child law, and shows that electric field at the space charge produced by the anode controls the electron current. The cathode emits electrons copiously, so much that there are always enough electrons available to satisfy Langmuir-Child. Of course, at a sufficiently high anode voltage, the current may *saturate*, when all the emitted electrons are attracted to the anode, but this never occurs in normal operation, so small variations in cathode temperature have no effect. The current in a vacuum tube is said to be *space-charge controlled*.

If enough gas is present in the tube, the positive ions can counterbalance the negative electron space charge, robbing the anode of control and greatly increasing the current. Also, the positive ions

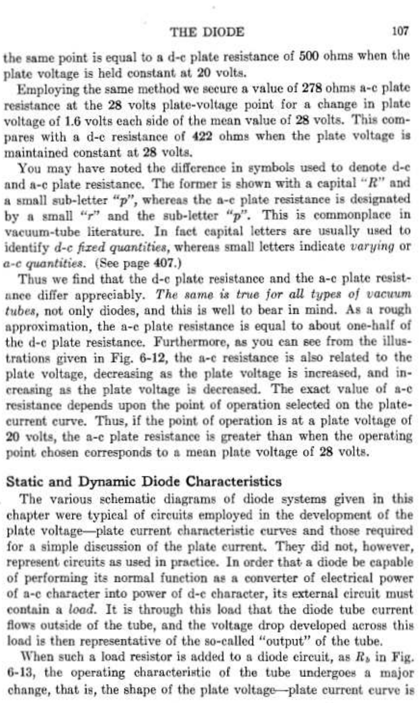

the same point is equal to a d-c plate resistance of 500 ohms when the plate voltage is held constant at 20 volts.

Employing the same method we secure a value of 278 ohms a-c plate resistance at the 28 volts plate-voltage point for a change in plate voltage of 1.6 volts each side of the mean value of 28 volts. This compares with a d-c resistance of 422 ohms when the plate voltage is maintained constant at 28 volts.

You may have noted the difference in symbols used to denote d-c and a-c plate resistance. The former is shown with a capital "*R*" and a small sub-letter "*p*", whereas the a-c plate resistance is designated by a small "*r*" and the sub-letter "*p*". This is commonplace in vacuum-tube literature. In fact capital letters are usually used to identify *d-c fixed quantities*, whereas small letters indicate *varying* or *a-c quantities*. (See page 407.)

Thus we find that the d-c plate resistance and the a-c plate resistance differ appreciably. *The same is true for all types of vacuum tubes*, not only diodes, and this is well to bear in mind. As a rough approximation, the a-c plate resistance is equal to about one-half of the d-c plate resistance. Furthermore, as you can see from the illustrations given in Fig. 6-12, the a-c resistance is also related to the plate voltage, decreasing as the plate voltage is increased, and increasing as the plate voltage is decreased. The exact value of a-c resistance depends upon the point of operation selected on the plate-current curve. Thus, if the point of operation is at a plate voltage of 20 volts, the a-c plate resistance is greater than when the operating point chosen corresponds to a mean plate voltage of 28 volts.

Static and Dynamic Diode Characteristics

The various schematic diagrams of diode systems given in this chapter were typical of circuits employed in the development of the plate voltage—plate current characteristic curves and those required for a simple discussion of the plate current. They did not, however, represent circuits as used in practice. In order that a diode be capable of performing its normal function as a converter of electrical power of a-c character into power of d-c character, its external circuit must contain a *load*. It is through this load that the diode tube current flows outside of the tube, and the voltage drop developed across this load is then representative of the so-called "output" of the tube.

When such a load resistor is added to a diode circuit, as R_b in Fig. 6-13, the operating characteristic of the tube undergoes a major change, that is, the shape of the plate voltage—plate current curve is

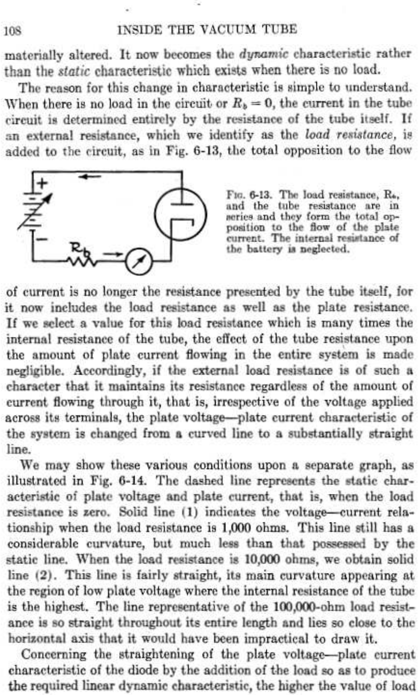

materially altered. It now becomes the *dynamic* characteristic rather than the *static* characteristic which exists when there is no load.

The reason for this change in characteristic is simple to understand. When there is no load in the circuit or $R_b = 0$, the current in the tube circuit is determined entirely by the resistance of the tube itself. If an external resistance, which we identify as the *load resistance*, is added to the circuit, as in Fig. 6-13, the total opposition to the flow of current is no longer the resistance presented by the tube itself, for it now includes the load resistance as well as the plate resistance. If we select a value for this load resistance which is many times the internal resistance of the tube, the effect of the tube resistance upon the amount of plate current flowing in the entire system is made negligible. Accordingly, if the external load resistance is of such a character that it maintains its resistance regardless of the amount of current flowing through it, that is, irrespective of the voltage applied across its terminals, the plate voltage—plate current characteristic of the system is changed from a curved line to a substantially straight line.

FIG. 6-13. The load resistance, R_b, and the tube resistance are in series and they form the total opposition to the flow of the plate current. The internal resistance of the battery is neglected.

We may show these various conditions upon a separate graph, as illustrated in Fig. 6-14. The dashed line represents the static characteristic of plate voltage and plate current, that is, when the load resistance is zero. Solid line (1) indicates the voltage—current relationship when the load resistance is 1,000 ohms. This line still has a considerable curvature, but much less than that possessed by the static line. When the load resistance is 10,000 ohms, we obtain solid line (2). This line is fairly straight, its main curvature appearing at the region of low plate voltage where the internal resistance of the tube is the highest. The line representative of the 100,000-ohm load resistance is so straight throughout its entire length and lies so close to the horizontal axis that it would have been impractical to draw it.

Concerning the straightening of the plate voltage—plate current characteristic of the diode by the addition of the load so as to produce the required linear dynamic characteristic, the higher the value of load

positive ions, which can cause emission of electrons or even heat the cathode the required amount for thermionic emission. All receiving tubes employ thermionic emission, though we will note certain examples of *cold cathodes* in special cases. These were not usually really cold, but heated by ion bombardment rather than by a current supplied externally. The space in which the electrons move is not completely devoid of gas, so some gas molecules may be *ionized* by collision with speedy electrons, when an electron is knocked off, leaving a *positive ion*. The positive ions move in the opposite direction to the electrons (but their current is in the same direction, since they are of opposite charge). The effect of positive ions in a receiving tube is very small, because of the very high vacuum that is used.

Self-heated electron emitters are called *filaments*. The carbon filaments of the Edison Effect were soon replaced by metallic emitters, usually tantalum or tungsten, which were used by Fleming and de Forest. In Germany, Arthur Wehnelt discovered in 1903 that barium or calcium oxides baked on a platinum base emitted copiously, and used these *oxide-coated* emitters on vacuum rectifiers evolved from discharge tubes, which he patented in 1904. However, the use of *soft* tubes, which contained residual gas, demanded the use of rugged tungsten filaments, which dominated in 1910-20. These filaments used low voltages and high currents, and had a short life because of the high temperatures required for adequate emission. Most radio sets had a rheostat to adjust the filament current properly. Wehnelt emitters would have been quickly destroyed by positive-ion bombardment in these tubes. Apparently by accident, thoriated tungsten wire was used in a trial at the GE factory in Harrison, NJ in 1920 on a UV201 tube. Thoriated tungsten gave 75 mA/W of filament power, while tungsten gave only 1.75. Thoriated-tungsten filaments became popular for receiving tubes, such as the UV201A, which was an improved UV201, around 1924. Its filament required 0.25 A at 5 V, while the UV201's had required 1.0 A. Since tubes were now all *hard*, or high-vacuum tubes, indirectly heated oxide-coated cathodes, which gave copious emission at low temperatures, were used almost exclusively in receiving tubes after 1930. Not only "triode" was not used until much later, after it threated to become a trade name), in which a third electrode, the grid, was introduced to control the electron stream. This made a more sensitive detector, but the amplifying property was not used at first, and de Forest, who did not understand well what was going on, defended gassy tubes with their gas amplification. The introduction of high vacuum, as well as improved materials and processes, especially metal-to-glass seals, created a very useful amplifying device that allowed great developments in radio, telephony and sound reproduction. Schottky suggested a screen grid between the plate and control grid to make the electron tube useful at higher frequencies in 1919 (and actually made tubes with a second grid, but this was for space-charge control), but this was only realized by Hull and Williams in 1928 in radio receivers. The metal tube was introduced in 1935, but glass envelopes never disappeared and were constantly improved. The final pattern of electron tube was the "miniature" or all-glass type, which became the predominant receiving-type tube after about 1945. Transistors were invented in 1948, and in the next decade were improved to the point where they could take over most of the amplifying applications of electron tubes at much lower cost, and with greater reliability. Electron tubes remain in use as cathode-ray tubes, magnetrons, X-ray tubes, and for handling large powers. They were remarkable devices, using many sophisticated materials and processes, yet were widely available at low cost. We shall look here mainly at examples of *receiving tubes*, the smaller amplifying devices that have been completely replaced by semiconductors in current practice, but nevertheless will deepen our knowledge of electronics, while being fascinating to study. The name "receiving" comes from their use in radio receivers, their principal commercial application, but refers to all small vacuum tubes for general electronic purposes. For the cathode-ray tube and making your own oscilloscope, see The Cathode-Ray Tube.

The electrons move from the cathode (K), the negative electrode, to the anode or *plate* (P), the positive electrode. Conventional current is in the opposite direction. The electrons are liberated at the cathode by heat--thermionic emission--or as a result of bombardment by

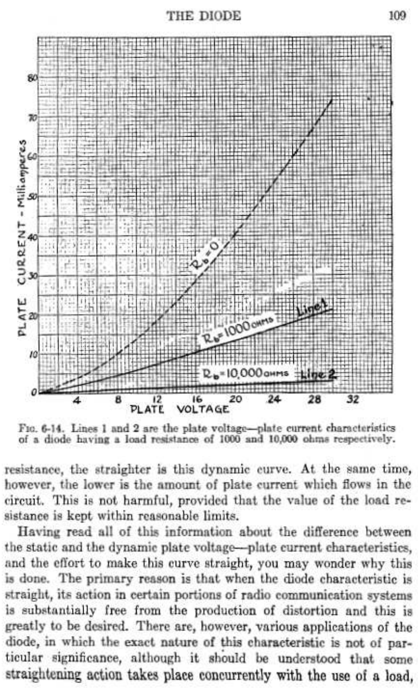

PLATE CURRENT - Milliamperes
80
70
60
50
40
30
20
10
0
4
8
12
16
20
24
28
32
PLATE VOLTAGE
R_b=0
R_b=1000 OHMS
Line 1
R_b=10,000 OHMS
Line 2

Fig. 6-14. Lines 1 and 2 are the plate voltage—plate current characteristics of a diode having a load resistance of 1000 and 10,000 ohms respectively.

resistance, the straighter is this dynamic curve. At the same time, however, the lower is the amount of plate current which flows in the circuit. This is not harmful, provided that the value of the load resistance is kept within reasonable limits.

Having read all of this information about the difference between the static and the dynamic plate voltage—plate current characteristics, and the effort to make this curve straight, you may wonder why this is done. The primary reason is that when the diode characteristic is straight, its action in certain portions of radio communication systems is substantially free from the production of distortion and this is greatly to be desired. There are, however, various applications of the diode, in which the exact nature of this characteristic is not of particular significance, although it should be understood that some straightening action takes place concurrently with the use of a load, and some sort of a load is required in order to make the diode of practical value.

It is also of importance to realize that the higher the load resistance used in a diode circuit, the greater is the permissible voltage which may be applied to the plate without fear of causing the flow of such high values of current that will damage the tube. You can see from the plate current values shown in Figs. 6-11 and 6-12 that the application of several hundred volts to the plate will cause the flow of such very high values of plate current as to damage the tube, unless high values of load resistance are used, which reduce the plate current to permissible values.

A-C Applied to the Diode Plate

It is necessary to speak briefly about the application of an alternating voltage to the plate of the diode. The reason is that everything we have said so far has shown d-c voltage applied to the plate, although we mentioned that the basic function of the diode was as a converter of a-c power into d-c power. The characteristics we have mentioned and developed by means of d-c voltages upon the plate are those which exist when a-c voltages are applied to the plate, for, in

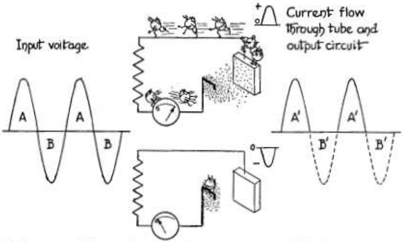
+
0
Current flow through tube and output circuit
Input voltage
A
A
B
B
A'
A'
B'
B'
0
-

Plate current flows during positive alternation and plate is positive with respect to cathode. Plate current does not flow when plate is negative with respect to the cathode. Thus output current and voltage is unidirectional although input is alternating.

Theory of Vacuum Tubes
J. B. Calvert
Created 17 August 2001
Last revised 8 April 2003

http://mysite.du.edu/~etuttle/electron/elect27.htm#Diodes

Miniature vacuum tubes with cathodes of high-field-emitting carbon nanotubes are currently under study at Agere Systems in Murray Hill, NJ. A triode with amplification factor of 4 has been constructed, with an anode-cathode spacing of 220 μm, and a pentode is planned. Vacuum tubes may return to electronic technology! See *Physics Today*, July 2002, pp. 16-18.

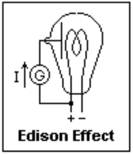

Devices in which a stream of electrons is controlled by electric and magnetic fields have many applications in electronics. Because a vacuum must be provided in the form of an evacuated enclosure in which the electrons can move without collisions with gas molecules, these devices were called *vacuum tubes* or *electron tubes* in the US, and *thermionic valves* in Britain. In 1883, Thomas Edison observed that a current flowed between the filament of an incandescent lamp and a plate in the vacuum near it (see figure at the right), when the plate was connected to the positive end of the filament, but not when the plate was connected to the negative side (the plate was actually between the two legs of the filament). No important application was made of this unexplained *Edison Effect* at the time. In 1899, J. J. Thomson showed that the current was due to a stream of negatively-charged particles, electrons, that could be guided by electric and magnetic fields. Fleming patented the *diode* in 1904 (B.P. 24850), where a filament and plate were arranged in the same envelope in a rather low vacuum, which could be used as a rectifier, or as a rather insensitive radio detector. In 1907, Lee de Forest patented the *triode* (which he called the Audion; the term *d*. A linear dynamic characteristic is desired for the diode, principally because it assures proportionality between applied plate voltage and output current. This affords freedom from distortion in many circuits that use the diode, although, in other applications, distortion is of secondary importance. The most important consideration is the use of a load; a *load* is necessary if a diode is to be of practical use.

27. Uses

a. Since current flow in the diode is unidirectional, one of its most prominent uses is as a rectifier or converter of alternating current into direct current. In figure 36, an alternating voltage is used as the plate voltage. During the positive alternations of the input voltage, the plate is positive relative to the cathode, and plate current i_p flows through the tube and load resistance R_L. The current develops a voltage drop across the load. At every instant the output voltage (in volts) is:

$$e \text{ output} = (i_p \times R_L) - (i_p \times r_p)$$

where R_L is the load resistance in ohms
- r_p is the a-c resistance of the diode in ohms
- i_p is the instantaneous current in amperes.

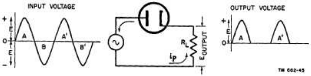

Figure 36. Diagram illustrating unidirectional current flow in diode.

b. If load resistance R_L is very much greater than r_p, the internal tube resistance, the effect of the latter is negligible and i_p times r_p of the preceding formula can be neglected.

c. Concerning the action in figure 36, it can be seen that the changes in current amplitude follow the instantaneous changes in applied plate voltage. This becomes evident when the input plate-voltage alternations, A and A', are compared with the output voltage pulsations, A and A'. These pulsations have the same shape as the two positive alternations of the applied voltage. The absence of conduction in the tube during the two negative alternations, B and B', of the input voltage is caused by the diode plate being negative with relation to the cathode during these periods. If the input voltage has a frequency of 60 cps (cycles per second), the output current pulses will occur 60 times per second, each lasting for a period equal to one-half of each input cycle.

d. Both single- and dual-section diodes have many other uses, developed later in this manual. All of these depend on the principle of unilateral conduction through the tube.

28. Types

High-vacuum diode tubes are available in various types (fig. 37), differing principally in physical dimensions and shape. They differ internally as well but not in the fundamental arrangement of the electrodes. The physical dimensions are related to the magnitudes of voltages applied to the plate and the currents flowing through the tube during operation. Tubes used in circuits which handle signal voltages of relatively low values are known as signal diodes. Another general category is that of power diodes or vacuum-tube rectifiers which handle high values of voltage and current. Each of these types is intended for special application and they are not interchangeable. Signal diodes are restricted to signal circuits. Power diodes generally are used in power circuits, except where certain signal systems entail relatively high voltages, and circuit operation demands an action similar to rectification.

29. Electron Transit Time

a. The time required for an electron to traverse the distance from the cathode to the plate of a diode, or other kinds of electron tubes, is

the final analysis, the tube operates only on one half of the a-c voltage wave, that is, during that time when the plate is positive with respect to the cathode. Therefore, the conditions which exist in the tube during that positive half cycle can be simulated by the application of d-c voltages.

b. With no load in the circuit, or when R_L equals 0, the resistance in the circuit is the resistance of the tube; also, the small resistances of the battery and meter are present, but these can be neglected. But when an external resistance, the *load resistance,* is added (fig. 34), the total opposition to plate-current flow includes that of the tube itself and also that of the load. When the load resistance is many times the value of the internal resistance of the tube, the tube resistance offers only a negligible percentage of the opposition to plate-current flow. Consequently, if the external load resistance maintains its value regardless of the amount of current flowing through it, the plate-current plate-voltage characteristic of the system is changed from a curved line into a substantially straight line.

c. These various conditions can be noted on a separate graph (fig. 35). The dashed line represents the static characteristic—that is, when there is no load resistance. Solid line 1 illustrates the voltage-current relationship when the load resistance is 1,000 ohms. This line possesses some curvature, but it is considerably less than that of the static line. Solid line 2 is obtained with a load resistance of 10,000 ohms. This line is almost straight; some curvature appears in the region of low plate voltage where the internal resistance of the tube is highest. A line representing a value of 100,000 ohms is so straight throughout its length that it lies too close to the horizontal axis, and it cannot be drawn clearly on the graph. Therefore, the higher the load resistor value, the straighter is the dynamic curve, and the lower is the amount of plate current flowing in the circuit. This is not a disadvantage, provided the load resistance is maintained within reasonable limits.

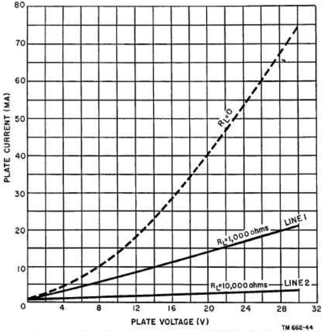

Figure 35. Plate-current plate-voltage characteristic curves of diode having different load values.

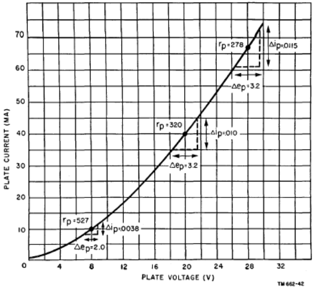

Figure 33. Plate-current plate-voltage characteristic which can be used to determine the a-c plate resistance of a diode.

lected on the characteristic curve. Consequently, if the point of operation is at a plate voltage of 20 volts, the a-c plate resistance is greater than when the operating point chosen corresponds to a mean plate voltage of 28 volts.

26. Static and Dynamic Diode Characteristics

a. The discussion of diodes thus far has been for static conditions and not for actual operating conditions. For a diode, or any tube, to be able to perform its normal function, its external circuit must contain a *load*. It is through this load that the diode current flows outside the tube, and the *voltage drop developed across this load then represents the output of the tube.* With such a load, represented by a resistance R_L (fig. 34), the operating characteristic of the tube is changed materially. The plate-current plate-voltage curve is altered noticeably, and it represents the *dynamic* characteristic rather than the *static* characteristic which applies when there is no load.

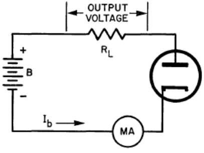

Figure 34. Adding a load resistor, R_L, to a diode circuit to obtain an output voltage.

PRINCIPLES
OF
ELECTRON TUBES

BY

HERBERT J. REICH, Ph.D.
Professor of Electrical Engineering
University of Illinois

First Edition

McGRAW-HILL BOOK COMPANY, Inc.
NEW YORK AND LONDON
1941

CHAPTER 2

THERMIONIC EMISSION; THE HIGH-VACUUM THERMIONIC DIODE

The operation of the majority of electron tubes is dependent upon thermionic emission. The theory of thermionic emission is therefore of great importance in the study of electron tubes. It is the purpose of this chapter to discuss the basic principles of thermionic emission, the construction of practical emitters, and the flow of electron space current in high-vacuum tubes containing two electrodes.

2-1. Theory of Thermionic Emission.—Richardson's theory of the emission of electrons from hot bodies is in many respects analogous to the kinetic theory of vaporization.[1] Heat possessed by a metal is believed to be stored not only in the kinetic energy of random motion of atoms and molecules, but also in the kinetic energy of free electrons. As a result of collisions between electrons or between electrons and atoms or molecules, the speed and direction of motion of a given electron are constantly changing. At any instant there will be, close to the surface of the metal, some electrons that have a component of velocity toward the surface of such magnitude that the corresponding kinetic energy is equal to or exceeds the electron affinity. If these electrons reach the surface without colliding with atoms, they will pass through the surface. At room temperatures the number of electrons meeting this requirement is extremely small, and hence no thermionic emission is detectable. As the temperature of the emitter is increased, however, the average velocity of the free electrons increases. The number having velocity components toward the surface sufficient to allow them to escape, therefore, increases with temperature. Reduction of electron affinity reduces the velocity toward the surface necessary to allow an electron to escape, and so increases the number

---

[1] Richardson, O. W., *Proc. Cambridge Phil. Soc.*, **11**, 286 (1901); "Emission of Electricity from Hot Bodies," rev. ed., Longmans, Green & Company, New York, 1921.

---

when one pair of the two diode portions is used. This will be found to be true in all multiple-section diodes.

h. The absence of plate-current saturation points similar to those appearing in figures 29 and 30 is due to the use of oxide-coated emitters in this tube. An application as low as 32 volts to the plate results in a fairly high value of plate current, almost 70 ma. Judging by the shape of the characteristic, the saturation point is a long way off. The tube probably would be damaged long before the plate-current curve would flatten because of the application of sufficient plate voltage to attract all the emitted electrons. This does not mean that diodes are not used at plate voltages exceeding 32 volts; they are used with very high voltages, many tens of thousands of volts, but when so used, diodes are specially designed.

25. A-c Plate Resistance of Diode

a. The a-c plate resistance of the diode is defined as the resistance of the path between cathode and plate to the flow of *an alternating current* inside the tube. Such is the definition meant when the term *resistance* is used without qualification for the diode and all other electron tubes. As will be seen later, it is an important term and is closely related to operating conditions.

b. The a-c plate resistance is the ratio of a small change in plate voltage by the corresponding change in plate current. Expressed in an equation

$$r_p = \frac{\triangle e_p}{\triangle i_p}$$

in which r_p is the a-c plate resistance in ohms, $\triangle e_p$ represents a small change in plate voltage, and $\triangle i_p$ represents a corresponding small change in plate current caused by the plate voltage change. The Greek letter $\triangle$ signifies a *small change of*.

c. Figure 33 shows a plate-current plate-voltage characteristic curve identical to the one in B of figure 32 and obtained from the circuit in A of figure 32, where one section of a 6H6 diode is used. To establish the *a-c resistance* at any point along the curve, the plate voltage is varied on both sides of it at some definite value. B of figure 32 shows that the d-c plate resistance, R_p for this diode is 500 ohms when the plate voltage is held constant at 20 volts dc. Suppose it is desired to find the a-c resistance when the mean value of this assumed a-c voltage is 20 volts, and the variation or *swing* is arbitrarily chosen at 1.6 volts on either side of the 20-volt point; that is, when $\triangle e_p$ equals 3.2 volts (fig. 33).

d. To compute r_p, the change in plate current for the given change in d-c plate voltage first must be found from the curve. At 21.6 (20 plus 1.6) volts, the plate current is 45 ma or .045 ampere. At 18.4 (20 minus 1.6) volts the plate current is 35 ma or .035 ampere. These provide the two limits of voltage and current. Consequently,

$$r_p = \frac{\triangle e_p}{\triangle i_p} = \frac{21.6 - 18.4}{.045 - .035} = \frac{3.2}{.010} - 320 \text{ ohms.}$$

e. The value of 320 ohms given above is the a-c plate resistance for a plate voltage which varies between 18.4 and 21.6 volts, whereas the same point is equal to a d-c plate resistance of 500 ohms when the plate voltage is held constant at 20 volts. Using the same method, 278 ohms a-c plate resistance is obtained at the 28-volt plate-voltage point for a change in plate voltage of 1.6 volts each side of the mean value of 28 volts, whereas the d-c resistance is 422 ohms when the plate voltage is maintained constant at 28 volts. Similarly, r_p is shown equal to 527 ohms when the 8-volt plate-voltage point varies *1* volt on either side of the 8-volt mean value. If 1.6 volts is used in the latter case instead of 1 volt, r_p still remains relatively close to 527 ohms. However, it must be kept in mind that the smaller the change in plate voltage, the more accurate are the results. This is because of the nonlinearity of the characteristic curve about this small change. In the former two cases, the curve is fairly linear and 1.6 volts are used arbitrarily. Normally, for linear operation, the operating point for this curve is chosen with an r_p equal to about 320 ohms.

f. It can be seen that an appreciable difference exists between d-c plate resistance and a-c plate resistance, the latter being approximately one-half of the former. This is, in general, true for all types of vacuum tubes. Furthermore, as figure 33 shows, the a-c resistance also is related to the plate voltage, *decreasing* as plate voltage is *increased*, and *increasing* as plate voltage is *decreased*. The *exact value of a-c resistance depends on the point of operation se-*

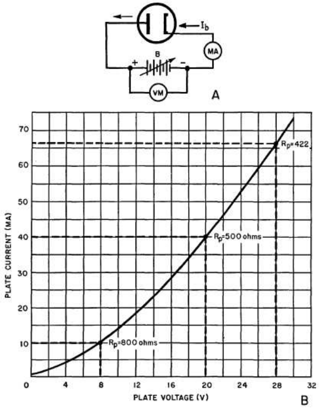

TM 662-41

Figure 32. Plate-current plate-voltage characteristic for type 6H6 tube, using only one diode section.

as the plate voltage increases, and *increases* as the plate voltage is decreased. It has a nonlinear behavior. If the relationship of plate voltage and plate current were linear over the entire range of plate voltage, the plate-current curve would be a straight line instead of having a curved characteristic. A straight line would indicate that the d-c plate resistance remained constant over the full range of plate voltage.

g. Another significant fact in connection with the $E_b = I_b$ characteristic curve is the condition resulting when the two sections of this duo-diode are connected in parallel, and the two cathodes also are joined in parallel. The plate current in this case becomes twice the amount obtained when the same value of plate voltage is applied to a single pair of elements. Accordingly, the values at any point along the plate-current curve are doubled when both sections of the duo-diode are paralleled. The plate resistance, therefore, becomes half of that existing

---

having the requisite velocity. Thermionic emission of electrons therefore increases with increase of temperature and with reduction of electron affinity. Measurable emission is observed at temperatures above 1000°K.

Electrons that escape will have resultant velocities made up of the excess perpendicular to the surface, plus the original components parallel to the surface, which are not altered by the surface forces. If the emitted electrons are not drawn away by an external field, they will form a space charge, the individual particles of which are moving about with random velocities. Because the initial average normal velocity of the electrons after emission is away from the surface and because of the mutual repulsion of like charges, electrons drift away from the surface. Collisions between electrons cause some of them to acquire velocity components toward the emitter, where they may reenter the surface with a gain of kinetic energy equal to the electron affinity. Another factor responsible for the return of electrons to the emitter is the electrostatic field set up by the negative space charge and, if the emitter is insulated, by the positive charge that it acquires as the result of loss of electrons. This field increases with the density of space charge, and equilibrium is established when only enough electrons can move away from the surface to supply the loss by diffusion of the space charge. If diffusion can then be prevented, just as many electrons return to the metal in unit time as leave it. Figure 2-1 gives a rough picture of the electron distribution under equilibrium conditions.

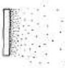

FIG. 2-1.—Distribution of electrons near an emitting surface.

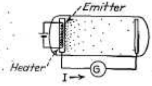

FIG. 2-2.—Flow of anode current as the result of diffusion of electrons from cathode to anode without the application of anode voltage (Edison effect).

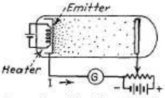

FIG. 2-3.—Use of anode voltage to increase anode current.

If a second, cold electrode is placed near the emitting surface in vacuum and connected to the emitter through a galvanometer, as shown in Fig. 2-2, the meter will indicate the small current resulting from the drift of electrons from the emitter to the second electrode. These electrons return to the emitter through the galvanometer and prevent the emitter from becoming positively charged. This phenomenon, first observed by Edison, is called the *Edison effect*. When the second electrode is made positive with respect to the emitter by the addition of a battery, as shown in Fig. 2-3, the current is increased. As the voltage is gradually raised, it is found that at any emitter temperature there is a more or less definite voltage beyond which the current is nearly constant, all emitted electrons being drawn to the collector. This current is called the *saturation current*, and the corresponding voltage, the *saturation voltage*. The lack of increase of current beyond saturation voltage is spoken of as *voltage saturation*. Saturation current varies with the temperature and electron affinity of the emitter. The negative emitter in Fig. 2-3 is called the *cathode;* and the positive collector, the *anode* or *plate*. An electron tube containing only a cathode and an anode is called a *diode*. Although the electrons move from cathode to anode, the current, according to convention, is said to flow from anode to cathode within the tube.

2-2. Richardson's Equation.—By means of classical kinetic theory, and by thermodynamic theory, Richardson derived two slightly different equations for saturation current as a function of temperature.[1] It is not possible experimentally to determine which of Richardson's equations is correct, but this was later done theoretically by M. v. Laue, S. Dushman, and A. Sommerfeld. The equation that is now believed to be correct is

$$I_s = AT^2\epsilon^{-w/kT} \qquad (2\text{-}1)$$

in which I_s is the saturation current per unit area of emitter, T is the absolute temperature, w is the electron affinity of the emitter, k is Boltzmann's universal gas constant, and A is a constant, probably universal for pure metals. The value of k is 8.63×10^{-5} volt/deg and the theoretical value of A for pure metals is 60.2. The form of the curve that represents Richardson's equation, shown in Fig. 2-4, is determined practically entirely by the exponential factor.

It is important to note that Richardson's equation holds only for the saturation current and that the anode voltage must,

---

[1] RICHARDSON, *loc. cit.*

---

considerably larger, since the plate current is larger, than that of the tungsten types. Oxide-coated emitters release so many electrons that it is practically impossible to find plate voltages which will draw off all of these electrons without ruining the tube. Consequently, most modern tubes which use oxide-coated cathodes do not display a plate-voltage saturation characteristic, but rather show continually rising plate-current curves as the plate voltage is increased as shown in figure 31. Note from the curves that as the operating temperature of the cathode increases, the plate current increases. In addition, this increase will have a more noticeable effect on the oxide-coated cathodes since the solid and dashed lines are much farther apart in the higher operating temperatures than in the lower ones.

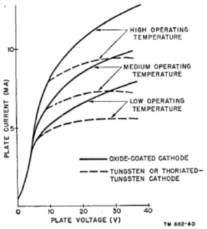

Figure 31. Plate-current emitter-temperature characteristic curves of diode with different types of emitters.

24. D-c Plate Resistance of Diode

a. Since the application of a positive voltage to the plate of a diode results in the flow of plate current, and since some form of control of the magnitude of plate current for different values of voltage exists in the tube, it is entirely proper to view the current control mechanism, whatever may be its nature, as being the equivalent of a resistance. Electrical fundamentals teach that the control of current in a system issues from some form of opposition to the current flow present in the system.

b. Such opposition exists in all electron tubes, among which is the diode. It stems from many factors, such as the spacing between the electrodes, their physical size, the conditions of emission, the conditions of the space charge, and, in general, the energy wasted while the electrons are traversing the space within the tube. Tube resistance is of two kinds: *d-c plate resistance* and *a-c plate resistance*, each of which will be explained. The former usually is symbolized as R_p and the latter as r_p.

c. The d-c plate resistance of the diode is that opposition to plate current flow offered by the tube when a d-c voltage is applied to the plate. It can be calculated from Ohm's law, $R=E/I$, with the plate voltage and the plate current inserted as the known factors in the equation.

d. Referring to figure 32, B shows a typical plate-current plate-voltage characteristic for a type 6H6 duo-diode tube, using only one pair of the two pairs of elements contained inside the envelope—that is one cathode and its associated plate. With 8 volts dc applied to the plate, in A, the plate current, I_b, is seen to be 10 ma, or .01 ampere. (The notation E_b is used as the varying constant voltage between the plate and cathode. This may or may not be the same value as E_{bb}.) Applying Ohm's law for d-c circuits, these values of plate voltage and plate current represent a d-c plate resistance of

$$R_p - \frac{E_b}{I_b} - \frac{8}{.01} = 800 \text{ ohms.}$$

e. Similarly, with 20 volts dc applied to the plate, the curve shows a plate current of 40 ma, or .04 ampere. This corresponds to a d-c resistance of $20/.04 = 500$ ohms. With 28 volts dc applied to the plate, the plate current on the curve is 66.4 ma or .0664 ampere, and the d-c resistance is $28/.0664 - 422$ ohms.

f. Examination of these figures, and also of the shape of the characteristic curve, shows that the resistance offered by the diode to the flow of plate current is *not* constant, as is ordinarily the case when resistance is present in a conventional d-c circuit. The characteristic shows that the resistance of the diode *decreases* therefore, be high enough at all times so that all emitted electrons are drawn to the anode. If the anode voltage is fixed at some value E_1, while the temperature is raised, then at some temperature the current will begin to be limited by space charge in a manner similar to that when there is no accelerating voltage. Further increases of temperature will then have no effect upon the current. This temperature is called the *saturation temperature*, and the failure of the current to increase at higher temperature is spoken of as *temperature saturation*. If the emitter were homogeneous and the electrostatic field constant over the surface of the emitter, the advent of saturation would be abrupt, as indicated at point A. Actually, because of variations of temperature and electron affinity and of electrostatic field, saturation does not take place over the whole cathode surface at the same temperature, and so experimentally determined curves bend over gradually, as shown by the dotted curves. At higher anode voltage E_2, saturation occurs at a higher temperature. If the voltage is increased with temperature, then the current will continue to rise with temperature until the temperature becomes sufficiently high to vaporize the emitter.

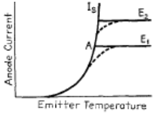

FIG. 2-4.—Curves of anode current *vs.* emitter temperature at two values of anode voltage. The left-hand curve, I_s, represents saturation emission current.

Since only those electrons which have relatively high energies can escape from the metal, thermionic emission necessarily results in the reduction of the average energy of the remaining electrons and molecules, and hence of temperature of the emitter. Heat must be supplied continuously to the emitter in order to prevent its temperature from falling as the result of emission. The cooling effect of emission current is plainly visible in filaments in which the emission current is comparable with the heating current, as in the type 30 tube.

Richardson's equation shows that the emission current which can be obtained at any temperature varies inversely with the electron affinity. Because of the exponential form of the equation, small changes in temperature or electron affinity result in large changes of emission current. A 15 per cent reduction of electron affinity produces an eight- or tenfold increase of emission

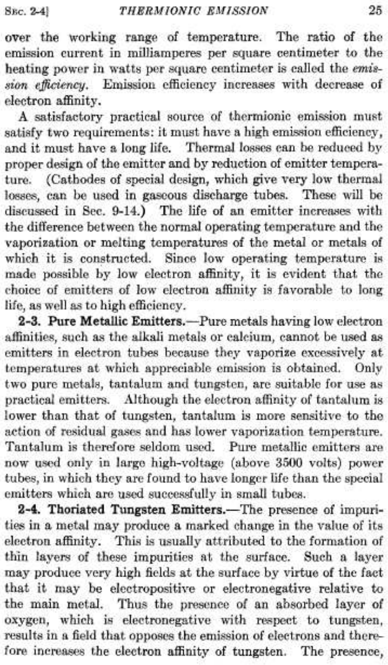

SEC. 2-4] *THERMIONIC EMISSION* 25

over the working range of temperature. The ratio of the emission current in milliamperes per square centimeter to the heating power in watts per square centimeter is called the *emission efficiency*. Emission efficiency increases with decrease of electron affinity.

A satisfactory practical source of thermionic emission must satisfy two requirements: it must have a high emission efficiency, and it must have a long life. Thermal losses can be reduced by proper design of the emitter and by reduction of emitter temperature. (Cathodes of special design, which give very low thermal losses, can be used in gaseous discharge tubes. These will be discussed in Sec. 9-14.) The life of an emitter increases with the difference between the normal operating temperature and the vaporization or melting temperatures of the metal or metals of which it is constructed. Since low operating temperature is made possible by low electron affinity, it is evident that the choice of emitters of low electron affinity is favorable to long life, as well as to high efficiency.

2-3. Pure Metallic Emitters.—Pure metals having low electron affinities, such as the alkali metals or calcium, cannot be used as emitters in electron tubes because they vaporize excessively at temperatures at which appreciable emission is obtained. Only two pure metals, tantalum and tungsten, are suitable for use as practical emitters. Although the electron affinity of tantalum is lower than that of tungsten, tantalum is more sensitive to the action of residual gases and has lower vaporization temperature. Tantalum is therefore seldom used. Pure metallic emitters are now used only in large high-voltage (above 3500 volts) power tubes, in which they are found to have longer life than the special emitters which are used successfully in small tubes.

2-4. Thoriated Tungsten Emitters.—The presence of impurities in a metal may produce a marked change in the value of its electron affinity. This is usually attributed to the formation of thin layers of these impurities at the surface. Such a layer may produce very high fields at the surface by virtue of the fact that it may be electropositive or electronegative relative to the main metal. Thus the presence of an absorbed layer of oxygen, which is electronegative with respect to tungsten, results in a field that opposes the emission of electrons and therefore increases the electron affinity of tungsten. The presence,

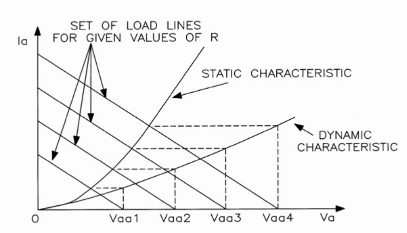

Obtaining the dynamic characteristic for a diode valve and its load.

The points on the dynamic characteristic are obtained by projecting, horizontally, the intersection of a load line and the static characteristic until it in turn intersects a vertical line drawn from a supply voltage value. Since the dynamic characteristic is drawn for a range of values of the supply voltage, this implies that the latter is varying, in other words it is an alternating supply rather than DC, as is the case in rectification.

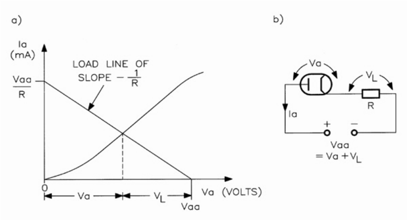

(a) The load line for a diode valve. (b) the diode in series with a linear resistive load.

The second image shows the diode in series with its load and defines the voltages and currents in question. A 'load line' of slope -l/R is drawn between the two points: Ia = Vaa/R Va = 0 and Ia = 0, Va = Vaa. As in solid state practice, the end points of a load line define zero conduction and maximum conduction, the latter being dependent upon the values of supply voltage and load resistance. Any other points on the load line imply intermediate levels of conduction. By dropping a construction line from the intersection of the load line and the static characteristic, we can see how the total supply voltage Vaa is divided into the two separate voltages Va (the voltage across the diode) and Vl (the voltage across the load).

The Dynamic Characteristic

By taking a number of different values of supply voltage Vaa (as would happen if the supply was alternating, for example) and assuming a constant value for the load Rl then, by drawing a separate load line for each value of supply voltage, the dynamic characteristic can be obtained, as shown below.

on the other hand, of a monatomic layer of thorium atoms or ions on the surface of tungsten reduces its electron affinity remarkably. It is of interest to note that the electron affinity of thoriated tungsten may be even lower than that of pure thorium (see Table 1-II).

The reduction of the electron affinity of tungsten as the result of introduction of small amounts of thorium was first observed by Langmuir in 1914 in the course of a study of the properties of tungsten filaments.[1] Thorium oxide is introduced into tungsten during the process of manufacture in order to improve its physical properties. Subsequent high temperature converts a portion of the thorium oxide into metallic thorium, which diffuses to the surface. Investigations by Dushman, Becker, and others[2] indicated that the lowest value of w and the highest value of the constant A in Richardson's equation are obtained when the tungsten is completely, or perhaps very nearly, covered with a single layer of thorium atoms.

Thoriated tungsten shows no increase of emission over that of pure tungsten until it is *activated*. The activation process is performed after evacuation of the tube. It consists first in "flashing" the emitter for a few moments at a temperature of 2500 to 2800°K. This high temperature reduces some of the thorium oxide to thorium. The temperature is then kept for some minutes at about 2200°K, which allows the metallic thorium to diffuse to the surface. The best value of diffusing temperature is determined by the rates of diffusion of thorium to the surface, and of evaporation from the surface. If the temperature is too high, the evaporation exceeds the diffusion, resulting in deactivation. The emitter is normally operated at temperatures that do not exceed 2000°K, which is sufficiently low so that evaporation of thorium from the surface is negligible. If the emitter is accidentally operated at such a high temperature that the whole supply of thorium diffuses to the surface and evaporates, it can be reactivated by repeating the original activation process. This may be done several times before all the thorium oxide is used up.

---

[1] LANGMUIR, I., *Phys. Rev.*, **4**, 544 (1914).

[2] DUSHMAN, S., and EWALD, J., *Phys. Rev.*, **29**, 857 (1927); BECKER, J. A., *Trans. Am. Electrochem. Soc.*, **55**, 153 (1929).

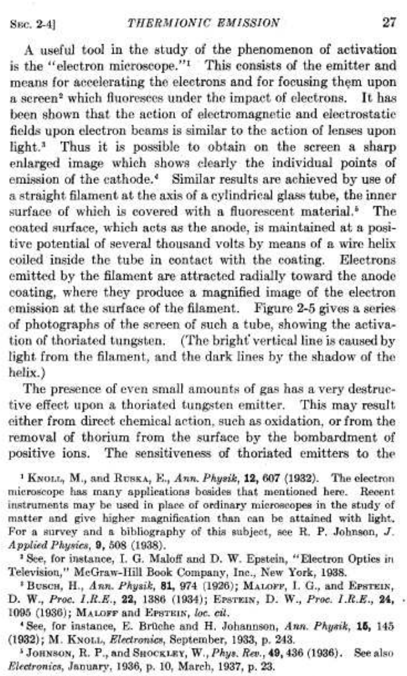

A useful tool in the study of the phenomenon of activation is the "electron microscope."[1] This consists of the emitter and means for accelerating the electrons and for focusing them upon a screen[2] which fluoresces under the impact of electrons. It has been shown that the action of electromagnetic and electrostatic fields upon electron beams is similar to the action of lenses upon light.[3] Thus it is possible to obtain on the screen a sharp enlarged image which shows clearly the individual points of emission of the cathode.[4] Similar results are achieved by use of a straight filament at the axis of a cylindrical glass tube, the inner surface of which is covered with a fluorescent material.[5] The coated surface, which acts as the anode, is maintained at a positive potential of several thousand volts by means of a wire helix coiled inside the tube in contact with the coating. Electrons emitted by the filament are attracted radially toward the anode coating, where they produce a magnified image of the electron emission at the surface of the filament. Figure 2-5 gives a series of photographs of the screen of such a tube, showing the activation of thoriated tungsten. (The bright vertical line is caused by light from the filament, and the dark lines by the shadow of the helix.)

The presence of even small amounts of gas has a very destructive effect upon a thoriated tungsten emitter. This may result either from direct chemical action, such as oxidation, or from the removal of thorium from the surface by the bombardment of positive ions. The sensitiveness of thoriated emitters to the

---

[1] KNOLL, M., and RUSKA, E., *Ann. Physik*, **12**, 607 (1932). The electron microscope has many applications besides that mentioned here. Recent instruments may be used in place of ordinary microscopes in the study of matter and give higher magnification than can be attained with light. For a survey and a bibliography of this subject, see R. P. Johnson, *J. Applied Physics*, **9**, 508 (1938).

[2] See, for instance, I. G. Maloff and D. W. Epstein, "Electron Optics in Television," McGraw-Hill Book Company, Inc., New York, 1938.

[3] BUSCH, H., *Ann. Physik*, **81**, 974 (1926); MALOFF, I. G., and EPSTEIN, D. W., *Proc. I.R.E.*, **22**, 1386 (1934); EPSTEIN, D. W., *Proc. I.R.E.*, **24**, 1095 (1936); MALOFF and EPSTEIN, *loc. cit.*

[4] See, for instance, E. Brüche and H. Johannson, *Ann. Physik*, **15**, 145 (1932); M. KNOLL, *Electronics*, September, 1933, p. 243.

[5] JOHNSON, R. P., and SHOCKLEY, W., *Phys. Rev.*, **49**, 436 (1936). See also *Electronics*, January, 1936, p. 10, March, 1937, p. 23.

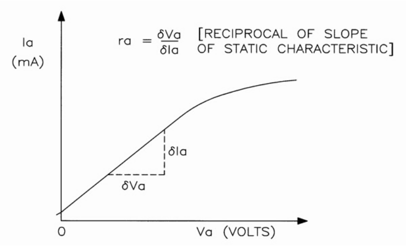

and is the value of resistance obtained by dividing a small change in anode voltage by the corresponding change in anode current. It is therefore the reciprocal of the slope of the static characteristic, and varies with the operating point, although fairly constant over much of the space charge limited region. This is a real value of resistance, since it represents the opposition of the valve to alternating quantities.

Series Circuit Operation

It is usual to operate a diode valve, which is clearly a non-linear device, in series with a resistive load, the latter being a linear device. It is possible to predict how the voltages and current in the circuit will vary by using a graphical construction.

Above three curves have been drawn for different values of cathode temperature, although in practice, as explained earlier, the cathode is held at a constant temperature.

It is interesting to note that:

A) The current is not exactly zero when the anode voltage is zero, but has a value (Iao) of a few micro-amperes. This is known as the 'splash current' and is the result of a few high energy electrons that manage to cross the inter-electrode gap even without an attracting potential.

B) In the space-charge limited region, the characteristic is nearly linear (actually following the 'three-halves' power law: Ia is proportional to Va-3/2).

C) In the temperature limited region there is little change in la even though there are large changes in Va. This is because the anode is collecting electrons at the same rate as they are being emitted by the cathode.

D) No significant current flows when the anode is negative with respect to the cathode.

The Anode Slope Resistance ra

It is worth introducing this parameter at this time since it is one that we shall make use of later in discussing the performance of more complex valves. It is defined as shown below,
Defining anode slope resistance for a diode valve.

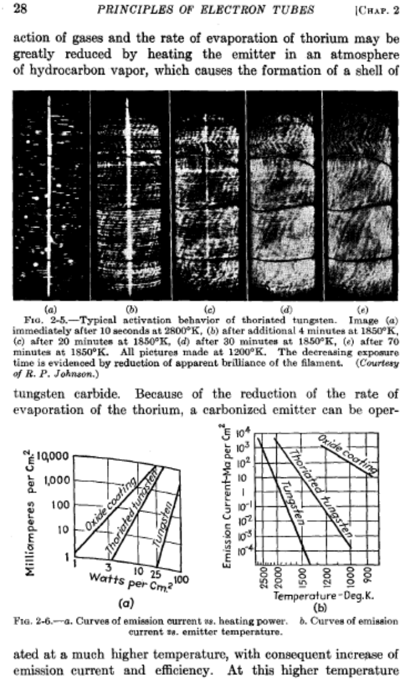

action of gases and the rate of evaporation of thorium may be greatly reduced by heating the emitter in an atmosphere of hydrocarbon vapor, which causes the formation of a shell of

FIG. 2-5.—Typical activation behavior of thoriated tungsten. Image (*a*) immediately after 10 seconds at 2800°K, (*b*) after additional 4 minutes at 1850°K, (*c*) after 20 minutes at 1850°K, (*d*) after 30 minutes at 1850°K, (*e*) after 70 minutes at 1850°K. All pictures made at 1200°K. The decreasing exposure time is evidenced by reduction of apparent brilliance of the filament. (*Courtesy of R. P. Johnson.*)

tungsten carbide. Because of the reduction of the rate of evaporation of the thorium, a carbonized emitter can be oper-

Milliamperes per Cm.²
10,000
1,000
100
10
1
Oxide coating
Thoriated tungsten
Tungsten
1
3
10
25
100
Watts per Cm.²
(a)
Emission Current-Ma per Cm²
10⁴
10³
10²
10
1
10⁻¹
10⁻²
10⁻³
10⁻⁴
Oxide coating
Thoriated tungsten
Tungsten
2500
2000
1500
1200
1000
900
Temperature - Deg. K.
(b)

FIG. 2-6.—*a*. Curves of emission current *vs.* heating power. *b*. Curves of emission current *vs.* emitter temperature.

ated at a much higher temperature, with consequent increase of emission current and efficiency. At this higher temperature

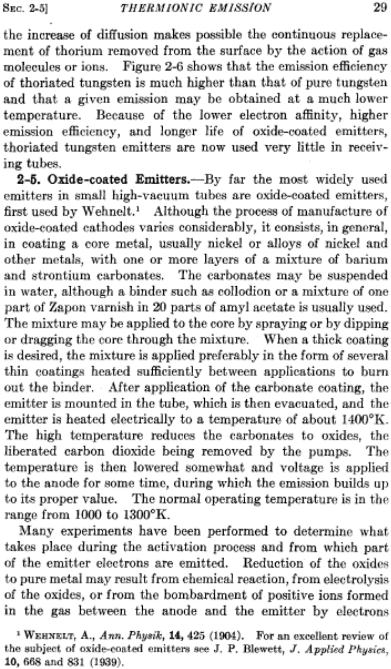

the increase of diffusion makes possible the continuous replacement of thorium removed from the surface by the action of gas molecules or ions. Figure 2-6 shows that the emission efficiency of thoriated tungsten is much higher than that of pure tungsten and that a given emission may be obtained at a much lower temperature. Because of the lower electron affinity, higher emission efficiency, and longer life of oxide-coated emitters, thoriated tungsten emitters are now used very little in receiving tubes.

2-5. Oxide-coated Emitters.—By far the most widely used emitters in small high-vacuum tubes are oxide-coated emitters, first used by Wehnelt.[1] Although the process of manufacture of oxide-coated cathodes varies considerably, it consists, in general, in coating a core metal, usually nickel or alloys of nickel and other metals, with one or more layers of a mixture of barium and strontium carbonates. The carbonates may be suspended in water, although a binder such as collodion or a mixture of one part of Zapon varnish in 20 parts of amyl acetate is usually used. The mixture may be applied to the core by spraying or by dipping or dragging the core through the mixture. When a thick coating is desired, the mixture is applied preferably in the form of several thin coatings heated sufficiently between applications to burn out the binder. After application of the carbonate coating, the emitter is mounted in the tube, which is then evacuated, and the emitter is heated electrically to a temperature of about 1400°K. The high temperature reduces the carbonates to oxides, the liberated carbon dioxide being removed by the pumps. The temperature is then lowered somewhat and voltage is applied to the anode for some time, during which the emission builds up to its proper value. The normal operating temperature is in the range from 1000 to 1300°K.

Many experiments have been performed to determine what takes place during the activation process and from which part of the emitter electrons are emitted. Reduction of the oxides to pure metal may result from chemical reaction, from electrolysis of the oxides, or from the bombardment of positive ions formed in the gas between the anode and the emitter by electrons

---

[1] WEHNELT, A., *Ann. Physik*, **14,** 425 (1904). For an excellent review of the subject of oxide-coated emitters see J. P. Blewett, *J. Applied Physics*, **10,** 668 and 831 (1939).

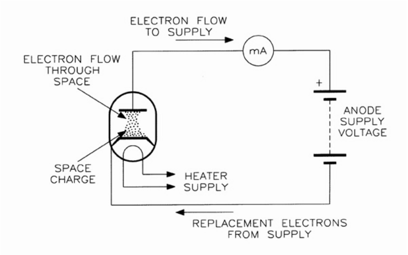

The flow of current in a diode valve.

Diode Static Characteristics

We now start getting into the ways in which specifications for thermionic devices are presented. For the diode, these illustrate clearly the dependence of anode current upon anode voltage.

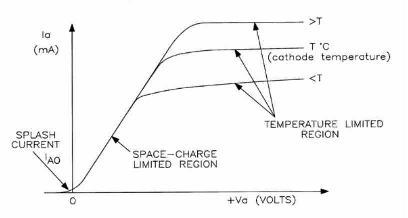

The static characteristic of a diode valve.

Since electrons are negative charged particles, they will only be attracted to the anode if this is given a positive potential with respect to the cathode. This explains why the valve only conducts in one direction, from cathode to anode, and not vice versa. It also explains the choice of the word 'valve' to describe the device, since a valve is, by definition, a one-way device. The Americans, however, never cottoned on to this terminology and always refer to them as 'vacuum tubes'.

The magnitude of the current flowing in a diode depends upon the number of electrons emitted and the magnitude of the voltage applied to the anode (known as the anode voltage Va). The amount of electron emission depends upon the temperature of the cathode, which is fixed by the voltage supply to the heater, this being a constant value. The only true variable is, therefore, the anode voltage. The action of the latter in controlling the anode current can be explained as follows.

As we now know from the foregoing, the cathode is normally surrounded by a cloud of electrons known as the space charge. With zero anode voltage there is no current flow, and there is a state of equilibrium between the electrons being emitted and those falling back onto the cathode's surface, The application of a small positive voltage to the anode causes some of the space charge electrons to be attracted to the anode, resulting in a small anode current flow. The gaps created by these electrons leaving the space charge are filled by further emission from the cathode. Electrons arriving at the anode flow to the positive supply terminal, while at the same time an equal number of electrons leave the negative supply terminal for the cathode. This gives rise to a continuous current flow around the circuit, which may be detected by an ammeter placed in, say, the anode lead. The picture below shows an illustration of this.

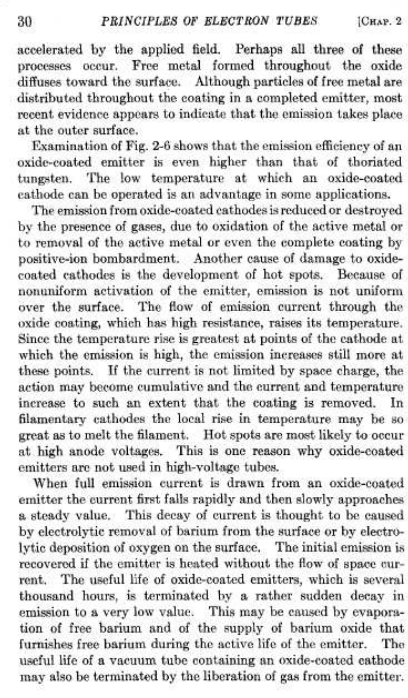

accelerated by the applied field. Perhaps all three of these processes occur. Free metal formed throughout the oxide diffuses toward the surface. Although particles of free metal are distributed throughout the coating in a completed emitter, most recent evidence appears to indicate that the emission takes place at the outer surface.

Examination of Fig. 2-6 shows that the emission efficiency of an oxide-coated emitter is even higher than that of thoriated tungsten. The low temperature at which an oxide-coated cathode can be operated is an advantage in some applications.

The emission from oxide-coated cathodes is reduced or destroyed by the presence of gases, due to oxidation of the active metal or to removal of the active metal or even the complete coating by positive-ion bombardment. Another cause of damage to oxide-coated cathodes is the development of hot spots. Because of nonuniform activation of the emitter, emission is not uniform over the surface. The flow of emission current through the oxide coating, which has high resistance, raises its temperature. Since the temperature rise is greatest at points of the cathode at which the emission is high, the emission increases still more at these points. If the current is not limited by space charge, the action may become cumulative and the current and temperature increase to such an extent that the coating is removed. In filamentary cathodes the local rise in temperature may be so great as to melt the filament. Hot spots are most likely to occur at high anode voltages. This is one reason why oxide-coated emitters are not used in high-voltage tubes.

When full emission current is drawn from an oxide-coated emitter the current first falls rapidly and then slowly approaches a steady value. This decay of current is thought to be caused by electrolytic removal of barium from the surface or by electrolytic deposition of oxygen on the surface. The initial emission is recovered if the emitter is heated without the flow of space current. The useful life of oxide-coated emitters, which is several thousand hours, is terminated by a rather sudden decay in emission to a very low value. This may be caused by evaporation of free barium and of the supply of barium oxide that furnishes free barium during the active life of the emitter. The useful life of a vacuum tube containing an oxide-coated cathode may also be terminated by the liberation of gas from the emitter.

2-6. Cesiated Tungsten Emitters.—A fourth type of emitter, not used commercially in thermionic tubes, is produced by depositing a monatomic layer of cesium on tungsten. Because the ionizing potential of cesium vapor is less than the electron affinity of tungsten, the tungsten removes an electron from a cesium atom which strikes it, leaving a positive ion which is held to the tungsten surface by the resulting electrostatic field. The force of adhesion is even greater if the tungsten is first covered with a monatomic layer of oxygen, which is electronegative with regard to tungsten. The strong electrostatic field between the cesium ions and the tungsten or oxygen reduces the electron affinity to the comparatively low value of 0.7 electron volt or less. Because cesium melts at a temperature only slightly above room temperature, the cesium vapor is obtained by merely introducing a small amount of cesium into the evacuated tube, the subsequent vaporization being sufficient to coat the filament.

The low electron affinity of the tungsten-oxygen-cesium emitter makes possible high emission currents at a temperature of only 1000°K. This type of emitter has several disadvantages, however, which make it impractical for use in commercial tubes. As the result of the high vapor pressure of cesium at operating temperatures of the tube, the characteristics of the tube are influenced by tube temperature. Too high temperature vaporizes the cesium, causing temporary reduction in emission, or even removal of the oxygen layer with permanent reduction of emission. Except at very low anode voltages, ionization of the cesium vapor occurs, resulting in fluctuations of anode current. The presence of positive ions is also detrimental to the action of amplifier tubes for other reasons, which will be discussed (Secs. 2-8, 6-5, 10-16).

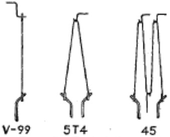

FIG. 2-7.—Typical filamentary cathodes. (*Courtesy of Radio Corporation of America.*)

2-7. Mechanical Construction of Cathodes.—Cathodes used in high-vacuum thermionic tubes are divided into two general classes, filamentary and indirectly heated. Figure 2-7 shows the form of typical filamentary cathodes. Early vacuum tubes used only filamentary cathodes. When filamentary cathodes are operated on alternating current, the stray alternating

---

http://www.r-type.org/articles/art-010b.htm

Valve Technology - A Practical Guide
A series of articles from 1993 by Graham Dixey C.Eng., MIEE republished by kind permission of Maplin Magazine.

The Diode Valve

The diode valve is so called because it has just two electrodes – the cathode and the anode.

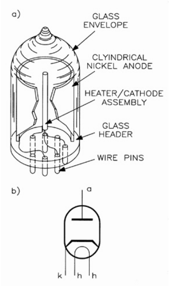

(a) Construction of a modern diode valve (indirectly heated type), (b) circuit symbol for a diode valve.

These correspond to the two electrodes of the original diode valve mentioned above, the cathode being the electrode that is heated and emits electrons, and the anode being the electrode that collects the electrons (notice also that these terms have passed forward into semiconductor phraseology - cathode, anode, emitter, collector!).

electrostatic field and the alternating voltage across the filament cause an alternating component of plate current that may be objectionable in amplifiers in which several tubes are used in succession. This difficulty led to the development of the indirectly heated, or *heater-type*, cathode. In addition to the unipotential emitting surface and freedom from large stray fields, the heater-type cathode has the advantage that a single source of power may be used to heat a number of cathodes between which a difference of potential must exist.

Indirectly heated cathodes used in receiving tubes consist of an oxide-coated cylindrical sleeve, usually of nickel, within which is some form of heater. The most common types of heaters are illustrated in Fig. 2-8. The 5Z4 and 25A6 heater coils are helically wound. That of the 6K7 type of cathode is wound in a reverse helix. After being wound and formed, the coils are coated with a refractory insulating material and inserted into the sleeve. The heater of the 25L6 type of cathode is covered with a refractory insulating coating of sufficient adherence to permit the wire to be bent into the desired shape after it has been coated. Because of the small magnetic field produced by the 6K7 type of heater, there is very little 60-cycle plate-current variation, or "hum," when the heater is operated on alternating current. The return type of helical heater exemplified by the 25A6 heater and the folded type of heater used in the 25L6 cathode make possible the use of enough wire for 25-volt operation. The advantage of the 25L6 construction is its low cost. More complicated cathode structures used in arc-discharge tubes will be discussed in Chap. 9.

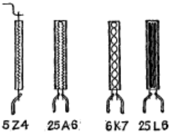

FIG. 2-8.—Structure of typical heater-type cathodes. (*Courtesy of Radio Corporation of America.*)

2-8. Effects of Gas upon Emission and Space Currents.—The deleterious effects of gas upon emitters of various types as the result of chemical action, absorption of thin layers of gas upon the surface, and positive-ion bombardment have already been mentioned. If the anode voltage is sufficiently high to produce ionization of the gas, other effects become apparent. If anode current is at first limited by space charge, the appearance of positive ions tends to neutralize the negative space charge surrounding the filament, thus increasing the anode current. If the voltage is high enough to give saturation current initially, then increase of current occurs because the electrons and ions produced by bombardment of neutral gas molecules by the emitted electrons add to the current. Unfortunately this increase of current is likely to be accompanied by a number of undesirable effects. Currents through ionized gases usually fluctuate, resulting in "noise" in tubes used for amplification. The relatively low velocity of positive and negative ions produces a lag in the response of current to changes of voltage. Variations of gas pressure resulting from changes of temperature or from the absorption or emission of gas from the walls and electrodes may cause the characteristics of the tube to vary. Finally, in a gassy tube, positive-ion current flows to an electrode to which a negative voltage is applied. When the anode current is controlled by means of a negative voltage applied to an electrode through a high resistance, the flow of positive-ion current through the resistance may cause an objectionable voltage drop and, under certain circumstances, may even result in damage to the tube (see Secs. 6-5, 9-17, and 10-16).

In the manufacture of high-vacuum tubes, many precautions are taken to ensure the removal of gas from walls and electrodes. The electrodes are thoroughly cleaned and are then heated for several minutes in an atmosphere of hydrogen, which removes oxygen and water vapor. After the tube is assembled and connected to the pumps, the electrodes are heated to about 800 or 1000°C by high-frequency induction in order to remove other occluded gases. Residual gas is removed by the use of *getters*, which are active chemical substances such as barium, magnesium, aluminum, and tantalum, having the property of combining with gases when they are vaporized. In glass tubes a small amount of the getter is mounted in such a position that it will be heated and vaporized or "flashed" during the inductive heating of the elements. By proper location of the getter, the vapor can be prevented from condensing in places where it might cause electrical leakage or undesirable primary or secondary emission. The effectiveness of the getter results not only from its chemical combination with gases during flashing, but also from subsequent absorption of gases by getter that has condensed on

---

2. A rectifier is always followed by electronic circuits called filter circuits which allow only D.C. to pass through them and by pass the A.C. Thus a rectifier-filter combination gives a D.C. output.

Applications of Diode

The two main applications of diode are :

(a) As a rectifier (b) As a detector

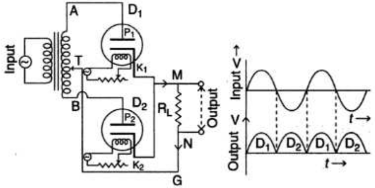

(i) Circuit (ii) Input-Output waveforms

positive. Therefore, the second diode conducts but the first remains passive. Thus, the two diodes conduct alternately. But in both half cycles of A.C., the current in load resistance R_L flows in the same direction. So, we get continuous D.C. at output. For full-wave rectifier,

Note :

$$I_{D.C.} = \frac{2I_0}{\pi}$$

$$I_{rms} = \frac{I_0}{\sqrt{2}}$$

$$\eta_{max} = 81{\cdot}2\%$$

Ripple factor, $r = 0{\cdot}48$

The output of rectifier is fluctuating or pulsating. To make it smooth, filters are used.

Notes : 1. The output obtained from a half wave or full wave rectifier is not a pure D.C. but it is a pulsating D.C. (mixture of A.C. and D.C.)

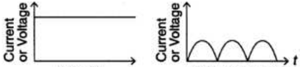

(i) (D.C.) (ii) Pulsating D.C.

---

the walls of the tube. The action of the getter during flashing is increased by ionization of the gas by means of voltages applied between the electrodes or by the radio-frequency field of the induction heater. To ensure the removal of gas from the walls, tubes are baked during the process of manufacture. On machines that exhaust and seal the tubes separately, the bulbs are heated in ovens during exhaustion. On "Sealex" machines, the tubes are sealed and exhausted on the same machine, the heat from sealing being used to drive gases from the bulbs during exhaustion.

In tubes with metal envelopes, the shielding action of the shell makes it impossible to heat the electrodes and the getter by induction. The electrodes may be heated by radiation from the shell, which is heated by gas flames. Although the getter may be fastened to the inside of the shell and vaporized by heating the shell locally, another method has been developed that requires less critical control.[1] A short length of tantalum ribbon, which connects the shell to its terminal pin in the base, is coated with a mixture of barium and strontium carbonates. While the tube is on the pumps, the temperature of the tantalum wire is raised electrically to about 1100°C. This converts the carbonates into oxides. After the tube has been sealed off, the tantalum wire is heated to a temperature in excess of 1200°C. This causes the tantalum to reduce the oxides to pure metallic barium and strontium, which vaporize. Since the vapor moves in straight lines, it can be directed as desired by means of shields and by proper location of the tantalum wire. This type of getter is called *batalum* [See (11) in Fig. 3-11*b*].

2-9. Limitation of Anode Current by Space Charge.—The effect of space charge in limiting space current and the increase of anode current resulting from an accelerating anode potential have already been mentioned in connection with the theory of thermionic emission. Before proceeding to a discussion of the quantitative relation between anode current and anode potential in a two-element tube, it is of interest to discuss further the physical picture underlying the phenomenon. The behavior of the emitted thermionic electrons is complicated by their initial velocities. For this reason it is best first to formulate a theory on the assumption that the initial velocities are zero and then, when they are taken into consideration, to see in what manner the results should be altered. For the present, therefore, initial velocities will be assumed to be zero. It will be further assumed that both cathode and anode are homogeneous, constant-potential, parallel planes of large area, and hence that the electric field over the surface of the cathode may be assumed to be uniform.

---

[1] LEDERER, E. A., and WAMSLEY, D. H., *RCA Rev.*, **2**, 117 (1937). This article also discusses gettering methods used in glass tubes.

Electrons that leave the cathode constitute a space charge that exerts a retarding field at the cathode. The net field at the surface of the cathode is the difference between this retarding field and the accelerating field produced by the positive voltage of the anode. The number of electrons in the space, and hence the retarding component of field at the cathode, increases with the anode current. When the positive anode voltage is applied, the anode current builds up with great rapidity to such a value that the average retarding field at the cathode caused by the space charge is equal to the accelerating field caused by the anode voltage, making the average field zero at the cathode. Increase of emission then does not raise the anode current, as the additional emitted electrons merely reenter the cathode. If it were possible in some manner to increase the density of space charge by increasing the current beyond this equilibrium value, or if the anode voltage were reduced slightly, then the net field at the cathode would be a retarding one. For an instant, all emitted electrons would be prevented from moving away from the cathode, and the current and space-charge density would be automatically reduced to a value that would again make the average field at the cathode zero. An increase of anode voltage causes the accelerating field to exceed the retarding field. The number of electrons moving to the anode then increases until the retarding field again equals the accelerating field.

At first thought it may not seem plausible that there can be a steady flow of electrons to the anode when both the velocities of emitted electrons and the average electrostatic field are zero at the cathode. It is only the *time average* field, however, that is zero at any point on the cathode. The *instantaneous* field at any point may vary in a random manner between positive and negative values. Immediately after one or more electrons have entered some point of the anode, the net field at a corresponding

---

Operation

- A.C. input is applied the plate 'P' becomes positive and negative alternately.
- For the positive half cycle of input, the plate is positive with respect of cathode. So that diode conducts and plate current flows through the tube load R_L and secondary. So output occurs across load. R_L as shown in fig. (ii).

Important Points

- During – ve half cycle of A.C. input, →Plate is negative with respect to cathode. So the diode does not conduct, and no voltage appears across output.
- For half wave rectifier; (I_0 is peak value of current)
  - → Average value of current $I_{D.C.} = I_0/\pi$
  - → Root mean square value of current $I_{rms} = I_0/2$
  - → Maximum efficiency $\eta_{max} = 40{\cdot}6\%$
  - → Ripple factor $r = 1{\cdot}21$

(B) **Full wave rectifier**—It converts full A.C. into D.C. In this rectifier two diodes are used which conduct alternately. The output is obtained across the load resistance R_L. Here in first half cycle of A.C., plate of first diode remains positive and plate of second diode remains negative. Hence, the first diode conducts and the second does not. In second half cycle of A.C., plate of the first diode becomes negative and that of second diode becomes

Rectifier

Rectifier is a device which converts an A.C. into a D.C. The process is called **Rectification**. Rectifiers are of two types :

1. **Half wave rectifier**—Which conducts only during + ve half cycle of input A.C.

2. **Full wave rectifier**—Which conducts during full cycle of input A.C.

(A) **Half wave rectifier** (in half wave rectifier only one diode valve is used)

Circuit

- The input A.C. voltage to be rectified is applied to the primary of a transformer.
- The secondary of transformer is connected in series with load R_L between plate and cathode of diode (as shown in fig.).

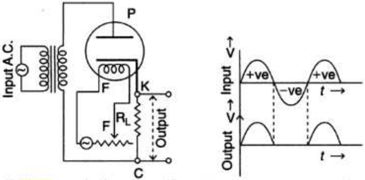

(i) Diode as a half-wave rectifier (ii) Input and output waveforms

---

point of the cathode may be positive, causing one or more electrons to move away from the cathode. These electrons produce a retarding field behind them which prevents the departure of more electrons from that point until the entrance of other electrons into the anode again results in an accelerating field. Many electrons are entering and leaving the space at any instant, so that the field fluctuations are rapid and haphazard.

2-10. Child's Law.—The foregoing descriptive explanation shows that, if an ample supply of electrons is available at the cathode, the anode current in a diode varies with the voltage applied between the anode and cathode. A mathematical analysis of this phenomenon was first made by Child.[1] The equation relating the anode current and voltage is called *Child's law*. The general derivation for electrodes of any size and shape is too difficult to yield a useful equation, and so only the relatively simple cases such as those applying to plane parallel electrodes of large area and to long concentric cylinders are ordinarily considered. In deriving Child's law for plane parallel electrodes the following assumptions are made:

1. The cathode temperature is high enough at all points so that more electrons are emitted than are drawn to the anode; *i.e.*, the current is limited by space charge.
2. The cathode and anode are parallel plates whose area is large as compared to their spacing; *i.e.*, the electrostatic field is uniform over the surface of any plane parallel to the electrodes.
3. The surfaces of the anode and cathode are equipotential surfaces.
4. The space between the cathode and the anode is sufficiently free of gas so that electrons do not lose energy by collision with gas molecules in moving from the cathode to the anode.
5. Emitted electrons have zero initial velocity after emission.

Under these assumptions the following three equations may be written:

$$\frac{\partial^2 V}{\partial x^2} = 4\pi\rho \qquad (2\text{-}2)$$

$$\mathcal{E}V = \tfrac{1}{2}m_e v^2 \qquad (2\text{-}3)$$

$$\rho v A = i_b \qquad (2\text{-}4)$$

in which V is the potential, relative to the cathode, at a distance x from the cathode; ρ is the density of electron space charge at a distance x from the cathode; $\mathcal{E}$ and m_e are the charge and mass, respectively, of an electron; v is the velocity dx/dt of an electron at a distance x from the cathode; i_b is the anode current; and A is the area of the electrodes.

---

[1] CHILD, C. D., *Phys. Rev.*, **32**, 498 (1911).

Equation (2-2) combines in symbolic form the definitions of potential difference and electric field. It is a special form of Poisson's equation, one of the most important fundamental laws of electrostatics, and may be derived directly from Gauss's law[1] (see Sec. 1-10). Equation (2-3) states that the energy gained by an electron in moving from the cathode to a distance x from the cathode under the influence of the electric field appears entirely in the form of kinetic energy. Equation (2-4) is a symbolic formulation of the definition of the magnitude of an electric current as the rate of flow of charge.

In the solution of the simultaneous differential Eqs. (2-2), (2-3), and (2-4), the following boundary conditions must be applied: At the cathode, the potential V, the average electric field $\partial V/\partial x$, and the velocity v are zero. At the anode, where x is equal to the cathode-to-anode spacing d, the potential is equal to e_b, the applied anode voltage. Solution of the equations and substitution of numerical values of $\mathcal{E}$ and m_e give the following equation for the anode current of a diode:[2]

$$i_b = 2.34 \times 10^{-6} \frac{A e_b^{3/2}}{d^2} \quad \text{amp} \qquad (2\text{-}5)$$

By combining Eq. (2-5) with Eqs. (2-2), (2-3), and (2-4), theoretical expressions may be derived for density of space charge, electron velocity, electric field strength, and potential as functions of distance from the cathode. Curves derived from these are shown in Fig. 2-9.

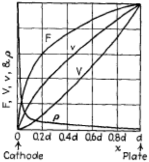

Fig. 2-9.—Variation of field strength F, potential V, electron velocity v, and space-charge density ρ with distance x from plane cathode. Zero initial velocity. Arbitrary units.

Child's law for concentric cylinders whose length is large as compared to their spacing is

---

[1] Page, L., and Adams, N. I., "Principles of Electricity," p. 83, D. Van Nostrand Company, Inc., New York, 1931.

[2] Page and Adams, *op. cit.*, p. 297.

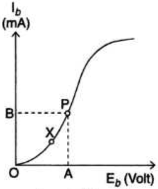

as D.C. plate resistance. The D.C. plate resistance R_p is given by

$$R_p = \frac{OA}{OB}$$

Note : Since plate characteristic curve is not a straight line—Therefore, D.C. plate resistance is variable. Hence, D.C. plate resistance may be calculated at the actual operating point.

(ii) **A. C. plate resistance**—The ratio of a small change in plate voltage across a diode to the resulting change in plate current is known as A.C. plate resistance.

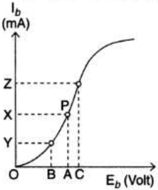

$$r_p = \frac{\Delta V}{\Delta I}$$

Note : The A.C. plate resistance at operating point P can be found by considering small equal changes of plate voltage on either side of the operating point (*i.e.*, AB = A.C.)

$$\text{Change in plate voltage} = BC$$

$$\text{Change in plate current} = YZ$$

$$\therefore \text{ A.C. plate resistance at } p, \; r_p = \frac{BC}{YZ}$$

2. At $V_p = V_1$, almost all the electrons constituting space charge are collected by plate and space charge is eliminated. At this point rate of collection of electrons reaches its maximum value and it equals the rate of emission of electrons by cathode. Therefore, if plate voltage is further increased plate current becomes constant.
3. The only way to increase the plate current is to increase the rate of emission, *i.e.*, to increase the temperature of cathode. Since the plate current, in this region, may be limited by temperature of cathode, this is called **temperature limited region**. In this region plate current is independent of plate voltage and varies with temperature T of cathode as

$$I = cT^2 e^{-b/T} \qquad \ldots\text{(ii)}$$

where *c* and *b* are constant for cathode.

Equation (ii) is simplified form of Richardson equation for this case.

Plate resistance of diode—The plate current (I_p) varies as the plate voltage (V_p) is changed, therefore, a diode offers internal resistance which is known as its plate resistance.

Note : Mainly negative space charge is responsible for the plate resistance of diode. As it is a non-ohmic resistance its value is not constant and it differs at different operating points.

Types of Plate Resistance

(i) **D.C. plate resistance**—The ratio of total D.C. plate voltage across diode to the resulting current is known

---

$$i_b = 14.68 \times 10^{-6} \frac{h e_b^{3/2}}{br} \quad \text{amp} \qquad (2\text{-}6)$$

in which *r* is the radius of the anode, *h* is the length of the electrodes, and *b* is a factor whose value depends upon the ratio of the radius of the anode to that of the cathode. *b* has the approximate value $\frac{1}{4}$ for a ratio 2, $\frac{1}{2}$ for a ratio 3, and 0.9 for a ratio 8. If the plate diameter is large as compared to that of the cathode, Eq. (2-6) reduces to the approximate form

$$i_b = 14.7 \times 10^{-6} \frac{h e_b^{3/2}}{r} \quad \text{amp} \qquad (2\text{-}7)$$

Equations (2-5), (2-6), and (2-7) show the importance of close spacing between cathode and anode if large currents are desired at small anode voltages.

2-11. Deviations from Child's Law Observed in Practical Diodes.—Deviations from Child's law result from the failure of practical diodes to satisfy the assumptions made in its derivation. Since the temperature of the cathode is fixed by considerations of emission efficiency and life, there is always a saturation voltage above which the current is not limited by space charge but by filament emission. If other assumptions were satisfied, the saturation voltage would be quite definite and the current-voltage curve would be as shown by the dotted lines of Fig. 2-10. Because of variations in temperature, electron affinity, and field strength over the cathode surface, the anode voltage at which voltage saturation takes place is not the same for all points of the cathode. The curve of anode current *vs.* anode voltage therefore bends over gradually, as shown by the full line of Fig. 2-10. Above saturation the current is not entirely constant but

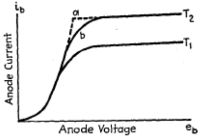

FIG. 2-10.—Curves of anode current *vs.* anode voltage at two values of emitter temperature.

continues to rise somewhat with anode voltage. This is explained by reduction of electron affinity with increase of external field (see Sec. 1-12), and lack of homogeneity of the surface of the emitter. The effect is particularly noticeable with oxide-coated cathodes.

The assumptions of uniform field and equipotential cathode are satisfied fairly closely in heater-type diodes with cylindrical plates. The voltage drop in filamentary cathodes may be shown to change Child's $\frac{3}{2}$-power law into a $\frac{5}{2}$-power law at anode voltages relative to the negative end of the filament that are less than the voltage of the positive end of the filament. This tends to make the lower part of the i_b-e_b curve steeper. The exact effect upon the curve of failure to satisfy the assumption of uniform field is complicated and impossible to predict completely.

2-12. Effect of Initial Velocities of Emitted Electrons.—Modified forms of Child's law which take into consideration the initial velocities of emitted electrons have been derived by Schottky, Langmuir, and others.[1] For the purpose of this book, a qualitative explanation of the effect of initial velocities is sufficient. Let it first be supposed that the electrons emerge with zero velocity. Under equilibrium conditions the space current and the space-charge density assume such values that the average field at the cathode is zero. The field and potential distributions in the interelectrode space are as shown in Fig. 2-9. Now let the emitted electrons suddenly have initial velocities that, for the sake of simplicity, are assumed to be the same for all electrons. Electrons that, without initial velocity, would have reentered the cathode now move toward the anode in spite of the fact that the average field is zero. As a result, the current and the space-charge density increase. The retarding field of the space charge now exceeds the accelerating field of the anode, giving a net retarding field at the cathode surface which slows up the electrons in the vicinity of the cathode. Equilibrium results when the retarding field in the vicinity of the cathode is sufficiently high so that the electrons are brought to rest in a plane a short distance *s* from the cathode. The average field in

---

[1] Schottky, W., *Physik. Z.*, **15**, 526 (1914); *Ann. Physik*, **44**, 1011 (1914). Langmuir, I., *Phys. Rev.*, **21**, 419 (1923). Davisson, C., *Phys. Rev.*, **25**, 808 (1925).

We note from fig. (i) that curves coincide at low voltage (in space charge limited region) but saturation current increases as shown by portion CD of characteristics.

(2) **Low voltage characteristics**—The plate current is practically zero at zero plate voltage (however, plate current is not exactly zero). If we measure plate current using a micro-ampere plate current of the order of few micro-ampere occurs. This is due to "the fact that even at zero plate voltage a few electrons may have sufficient K.E. to reach the plate and constitute a plate current of few micro-ampere." From the fig. (ii) the plate current becomes exactly zero at a particular negative value V_C of plate voltage, called **cut-off voltage.** The max. kinetic energy of emitted electron is related to V_C as :

$$K_{max} = e\,V_C$$

Important Points

1. If we increase plate voltage from zero volt to a value V_1 in steps, more and more electrons are attracted from space charge to plate (though at the same time same no. of electrons are emitted from cathode so that space charge is maintained) so plate current increases. As the plate current is limited by space charge, this region of curve is called **space charge limited region**. In this region, plate current is related to plate voltage by equation.

$$I_p = K \, . \, V_p^{3/2} \qquad \text{...(i)}$$

where K is a constant.

Equation (i) is known as **Child-Langmuir's law.**

V-I characteristics of Diode

(1) **High voltage characteristics**—Fig. (i) shows high voltage plate characteristics when plate current is measured in (mA). Keeping temperature of cathode constant, say (T_1) (*i.e.* filament current constant) when plate voltage increased from zero volt in steps and corresponding value of plate current i_p measured in m.A. the variation occurs as shown in curve OAB. If the cathode temperature is increased from T_1 to T_2.

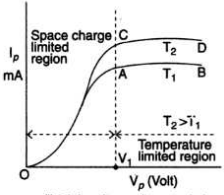

(i) High voltage characteristics

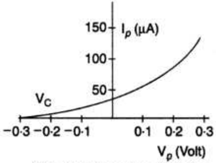

(ii) Low voltage characteristics

Fig. : V-I characteristics of diode.

---

this plane is zero, but instantaneous fluctuations allow just enough electrons to pass to give the required anode current. The behavior is similar to that which would obtain if the initial velocity were zero and the cathode were moved toward the anode by the distance s. Because of the random distribution of electron velocities the phenomenon is actually more complicated than this simplified picture indicates. The simple theory shows, however, that the effect of initial velocities is to increase the anode current corresponding to any anode voltage and is therefore equivalent to that of a small increase of anode voltage. Because the electrons emerge with velocities of the order of a volt or less, the effect is appreciable only for low anode voltages. The field and potential distributions throughout the interelectrode space are plotted in Fig. 2-11. That the potential must pass through a minimum where the field is zero follows from the fact that the field at any point may be expressed as the space derivative of the potential at that point.

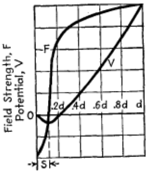

FIG. 2-11.—Variation of field strength and potential with distance from cathode for plane parallel electrodes of separation d. Initial velocities of emitted electrons considered.

The lower part of the i_b-e_b curve of Fig. 2-10 is raised slightly as the result of initial velocities of emitted electrons, and a small negative voltage must be applied in order to reduce the anode current to zero. Theoretical equations relating anode current and voltage at negative anode voltages have been derived by Schottky[1] and Davisson.[2] Because of their complicated form and because of failure to satisfy in practice the assumptions made in their derivation, these are seldom of great practical value. At negative anode voltages that are high enough to reduce the anode current to the order of 50 μa or less, the anode current of diodes with unipotential cathodes follows closely the empirical relation

$$i_b = k_1\epsilon^{k_2e_b} \qquad (2\text{-}8)$$

in which k_1 and k_2 are constants for a given tube. The current departs materially from this exponential law as the negative anode voltage is reduced in the vicinity of zero voltage, particularly at high cathode temperatures. Experimental curves corresponding to Eq. (2-8) were first obtained by Germer.[1]

---

[1] SCHOTTKY, *loc. cit.;* LANGMUIR, *loc. cit.*

[2] DAVISSON, *loc. cit.*

Unless the plate of a highly evacuated diode becomes hot enough to emit electrons, increase of negative anode voltage beyond the value that reduces the current to zero has no further effect upon the anode current. The fact that anode current flows in one direction only is employed in the application of diodes to detection and to power rectification, which will be discussed in Chaps. 7 and 11.

2-13. Relation of Richardson's and Child's Laws.—It should be noted that Richardson's equation and Child's law apply to two different conditions of operation of two-element tubes. Richardson's equation holds only under voltage saturation, whereas Child's law applies only under temperature saturation. In most applications, vacuum tubes are used in such a manner that temperature saturation prevails.

2-14. Shot Effect.—The random motion of electrons causes rapid variations of the number of electrons that pass from the cathode to the anode in unit time, and thus produces fluctuations of anode current. This phenomenon, which may be readily detected by the use of sufficient amplification, is called the *shot effect*. It is one of the factors that limit amplification by vacuum tubes (see Sec. 6-15).

Heating of the Plate.—The kinetic energy acquired by electrons in moving from the cathode to the plate is converted into heat when the electrons strike the plate. The average current that a vacuum tube can pass is limited by the temperature of the plate at which absorbed gas is driven out of the plate or electron emission takes place from the plate. The power that is converted into heat at the plate is equal to the time integral of the product of the plate current and plate voltage. Radiation of heat from the plate is increased by blackening its outer surface.

2-15. Classification of Tubes.—Electron tubes may be classified in a number of ways. These classifications include those based upon the process involved in the emission of electrons from the cathode, the degree of evacuation, the number of elec-

---

[1] Germer, L. H., *Phys. Rev.*, **25**, 795 (1925).

---

plate current. These electrons flow through the external circuit and finally return to the cathode, thus making up the supply of electrons lost by emission. On increasing plate potential more electron will gain sufficient kinetic energy so as to reach anode, hence the plate current increases.

Important Points to Note

- The current flows in the diode only when plate is made positive relative to cathode.
- No current can flow when plate is negative relative to cathode.
- Within a diode electrons can flow only from cathode to plate.
- Due to unidirectional conduction property the diode acts like a valve.
- The diode automatically starts conduction when the plate is positive or stops conduction when the plate is negative (due this property) the diode may act as a rectifier, converting A.C. into D.C.)

Characteristics of Diode

Characteristics of vacuum diode can be best studied by finding the relation between plate voltage and plate current for a given cathode temperature.

The circuit for determining the plate characteristic of an indirectly heated vacuum diode is shown in fig.

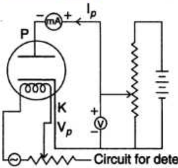

Circuit for determining plate characteristics of diode

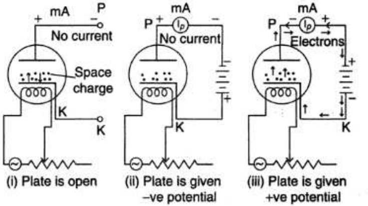
mA
P
+
−
No current
Space charge
K
K
mA
P
+
−
I_p
No current
K
mA
P
−
+
I_p
Electrons
K

(i) Plate is open (ii) Plate is given −ve potential (iii) Plate is given +ve potential

(1) **Anode (plate) at zero potential relative to cathode**—This situation is shown in fig. (i)—The emitted electrons do not have sufficient kinetic energy so as to reach the anode. However, a few electrons may reach the anode on account of their kinetic energy, constituting negligible current. The emitted electrons accumulate near the cathode and form a cloud of electrons. This is known as space charge. At a certain stage the number of electrons forming the space charge becomes constant for a given operating temperature. This space charge becomes a source of electrons that can be attracted to the plate, if it is at a positive potential.

(2) **Anode (plate) at negative potential relative to cathode**—This situation is shown in fig. (ii). The emitted electrons are repelled back due to retarding potential of anode. For a particular negative potential of anode plate current may be zero.

(3) **Anode (plate) at positive potential relative to cathode**—This situation is shown in fig. (iii). The electrons constituting the space charge are attracted to the plate. This flow of electrons from cathode to plate is known as

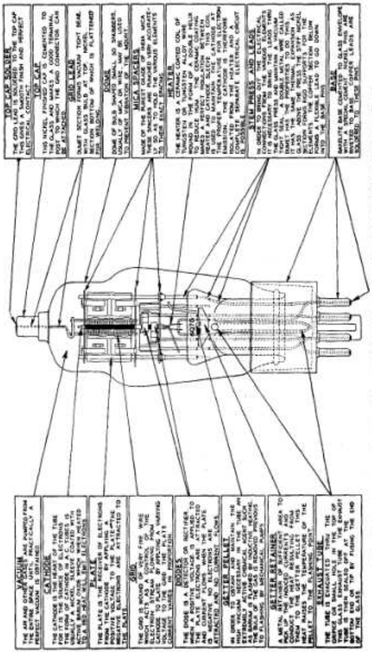
TOP CAP SOLDER
TOP CAP
GRID LEAD
DOME
MICA SPACERS
HEATER
STEM PRESS AND LEADS
BASE
VACUUM
CATHODE
PLATE
GRID
DIODES
GETTER PELLET
GETTER RETAINER
EXHAUST TUBE

FIG. 2-12.—Structure of a typical glass receiving tube. (*Courtesy of Ken-rad Tube and Lamp Corp.*)

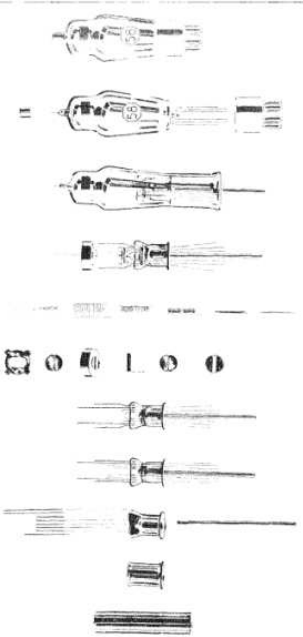

FIG. 2-13.—Assembly of a typical glass receiving tube. (Courtesy of Radio Corporation of America.)

Competition Science Vision:
C.S.V./January/2007/1432

Vacuum Diode or Diode Valve

A vacuum diode consists of two electrodes : a cathode; and an anode enclosed in an evacuated glass tube. Its operation is based upon thermionic emission, it is also called thermionic diode.

Anode (or plate)—The anode is a hollow cylinder made of Ni or molybdenum and surrounds the cathode.

Cathode—A cathode may be directly heated or indirectly heated type. Usually an indirectly heated cathode is provided (except in high power application).

As shown in fig. (i) an indirectly heated cathode is in the form of nickel cylinder coated with oxides of barium or strontium. Inside the cathode, a heater filament of tungsten is inserted.

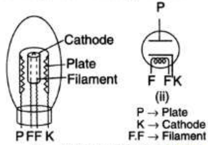

(i) Construction of diode
(ii) Symbol of diode

Fig. (ii) shows it symbol :

Operation of diode—When electric current is passed through the diode it emits a large number of electrons. These emitted electrons may be accelerated or retarded by applying positive or negative potential to anode relative to cathode.

the mutual conductance to decrease, which means that the amplification obtainable from the tube also decreases since amplification is proportional to mutual conductance, other things being equal. On the other hand, the plate resistance increases with increasing negative grid bias. As a check on the accuracy of measurement, the three curves should satisfy the relationship

$$g_m = \frac{\mu}{r_p}$$

within reasonable limits of accuracy, for any given value of grid bias.

If published average curves for the type of tube measured are available, it will be of interest to compare them to the curves determined experimentally. Exact duplication of the published curves is not to be expected, of course, because of slight variations in manufacture.

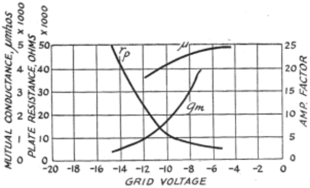

Fig 9.

trodes, and the type of application. Included in the first classification are the thermionic tube and the phototube. A *thermionic tube* is an electron tube in which the electron or ion emission is produced by the heating of an electrode. A *phototube* is an electron tube in which electron emission is produced directly by radiation falling upon an electrode. According to degree of evacuation, electron tubes are classified as high-vacuum tubes

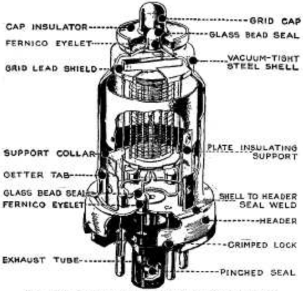
CAP INSULATOR
FERNICO EYELET
GRID LEAD SHIELD
SUPPORT COLLAR
GETTER TAB
GLASS BEAD SEAL
FERNICO EYELET
EXHAUST TUBE
GRID CAP
GLASS BEAD SEAL
VACUUM-TIGHT STEEL SHELL
PLATE INSULATING SUPPORT
SHELL TO HEADER SEAL WELD
HEADER
CRIMPED LOCK
PINCHED SEAL

FIG. 2-14.—Structure of typical metal receiving tube.

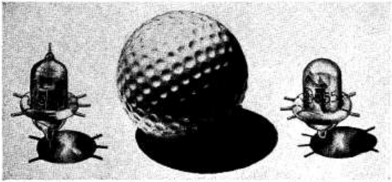

FIG. 2-15.—Size of typical "acorn" tubes shown in comparison with a golf ball. (*Courtesy of Radio Corporation of America.*)

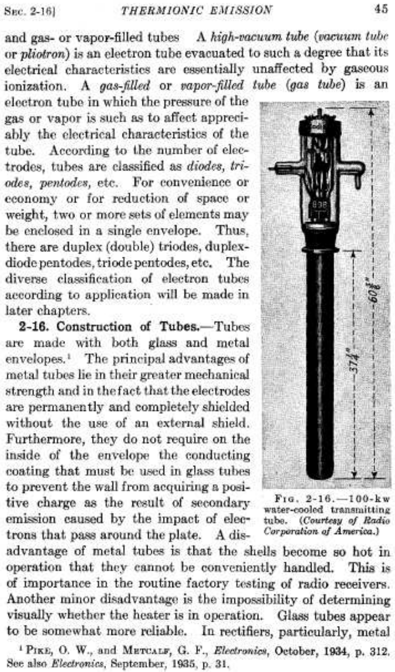

and gas- or vapor-filled tubes A *high-vacuum tube* (*vacuum tube* or *pliotron*) is an electron tube evacuated to such a degree that its electrical characteristics are essentially unaffected by gaseous ionization. A *gas-filled* or *vapor-filled tube* (*gas tube*) is an electron tube in which the pressure of the gas or vapor is such as to affect appreciably the electrical characteristics of the tube. According to the number of electrodes, tubes are classified as *diodes, triodes, pentodes*, etc. For convenience or economy or for reduction of space or weight, two or more sets of elements may be enclosed in a single envelope. Thus, there are duplex (double) triodes, duplex-diode pentodes, triode pentodes, etc. The diverse classification of electron tubes according to application will be made in later chapters.

2-16. Construction of Tubes.—Tubes are made with both glass and metal envelopes.[1] The principal advantages of metal tubes lie in their greater mechanical strength and in the fact that the electrodes are permanently and completely shielded without the use of an external shield. Furthermore, they do not require on the inside of the envelope the conducting coating that must be used in glass tubes to prevent the wall from acquiring a positive charge as the result of secondary emission caused by the impact of electrons that pass around the plate. A disadvantage of metal tubes is that the shells become so hot in operation that they cannot be conveniently handled. This is of importance in the routine factory testing of radio receivers. Another minor disadvantage is the impossibility of determining visually whether the heater is in operation. Glass tubes appear to be somewhat more reliable. In rectifiers, particularly, metal

FIG. 2-16.—100-kw water-cooled transmitting tube. (*Courtesy of Radio Corporation of America.*)

---

[1] PIKE, O. W., and METCALF, G. F., *Electronics*, October, 1934, p. 312. See also *Electronics*, September, 1935, p. 31.

The values of amplification factor, µ, plate resistance, rp, and mutual conductance, gm, can be measured from these three sets of curves. The mutual conductance, ΔIpΔEg can be found from the curves of Fig. 6 since these curves show the relationship between grid voltage and plate current. The plate resistance, ΔIp/ΔEp, can be measured from the curves of Fig.: 7, which relate plate current to plate voltage for various values of grid bias, while the amplification factor ΔIpΔEg, can be taken from the curves of Fig. 8. The method of making these measurements is described in the introduction to this installment. Since these "constants" are a function of three variables a large number of graphs would be required to give their behavior even partially completely, but one special case is shown in Fig. 9. This graph shows the variation in µ, rp and gm as a function of grid bias when the plate voltage is held constant at 250 volts, the normal rated operating voltage for the tube, and is a plot of values measured at 250-volt points on each of the three sets of curves in Figs. 6, 7 and 8. It is plain that the amplification factor changes relatively little compared to the changes in the other two quantities. Increasing negative grid bias causes

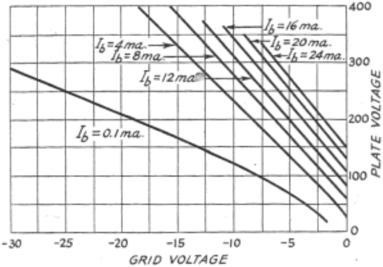

Fig. 8

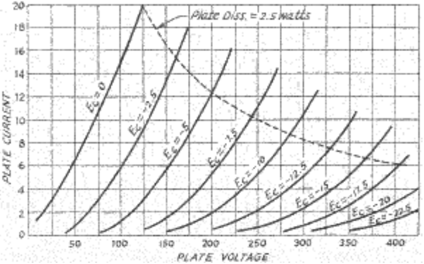

Fig. 7

The "plate family," shown plotted from experimental data in Fig. 7, is obtained by holding the grid bias constant at selected values and measuring the plate current as the plate voltage is varied. These curves show the same general tendency to bend when the plate current is near cut-off, and to straighten out at higher values of plate current. The plate family is frequently more useful than the set of grid voltage-plate current curves represented by Fig. 6.

When the remaining quantity, plate current, is held constant while the grid voltage is varied (the plate voltage being adjusted for each value of grid bias to give the selected value of plate current) the set of curves shown in Fig. 8 results, again plotted from experimental data on a 6J5. These "constant current" curves show the relative effect of grid voltage and plate voltage on plate current. The curves are nearly straight lines for all except very small values of plate current, showing that the amplification factor is practically constant for a given plate-current value regardless of the plate and grid voltages. The fact that, with the exception of the curve for a plate current of 0.1 milliampere, the curves are very nearly parallel indicates that the amplification factor also is nearly independent of the plate current so long as the latter is not near the cut-off point.

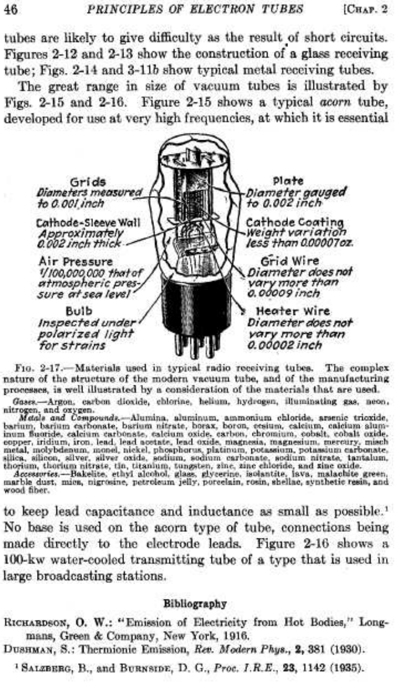

46 PRINCIPLES OF ELECTRON TUBES [CHAP. 2

tubes are likely to give difficulty as the result of short circuits. Figures 2-12 and 2-13 show the construction of a glass receiving tube; Figs. 2-14 and 3-11*b* show typical metal receiving tubes.

The great range in size of vacuum tubes is illustrated by Figs. 2-15 and 2-16. Figure 2-15 shows a typical *acorn* tube, developed for use at very high frequencies, at which it is essential

Grids
Diameters measured to 0.001 inch
Cathode-Sleeve Wall
Approximately 0.002 inch thick
Air Pressure
1/100,000,000 that of atmospheric pressure at sea level
Bulb
Inspected under polarized light for strains
Plate
Diameter gauged to 0.002 inch
Cathode Coating
Weight variation less than 0.00007 oz.
Grid Wire
Diameter does not vary more than 0.00009 inch
Heater Wire
Diameter does not vary more than 0.00002 inch

FIG. 2-17.—Materials used in typical radio receiving tubes. The complex nature of the structure of the modern vacuum tube, and of the manufacturing processes, is well illustrated by a consideration of the materials that are used.

Gases.—Argon, carbon dioxide, chlorine, helium, hydrogen, illuminating gas, neon, nitrogen, and oxygen.

Metals and Compounds.—Alumina, aluminum, ammonium chloride, arsenic trioxide, barium, barium carbonate, barium nitrate, borax, boron, cesium, calcium, calcium aluminum fluoride, calcium carbonate, calcium oxide, carbon, chromium, cobalt, cobalt oxide, copper, iridium, iron, lead, lead acetate, lead oxide, magnesia, magnesium, mercury, misch metal, molybdenum, monel, nickel, phosphorus, platinum, potassium, potassium carbonate, silica, silicon, silver, silver oxide, sodium, sodium carbonate, sodium nitrate, tantalum, thorium, thorium nitrate, tin, titanium, tungsten, zinc, zinc chloride, and zinc oxide.

Accessories.—Bakelite, ethyl alcohol, glass, glycerine, isolantite, lava, malachite green, marble dust, mica, nigrosine, petroleum jelly, porcelain, rosin, shellac, synthetic resin, and wood fiber.

to keep lead capacitance and inductance as small as possible.[1] No base is used on the acorn type of tube, connections being made directly to the electrode leads. Figure 2-16 shows a 100-kw water-cooled transmitting tube of a type that is used in large broadcasting stations.

Bibliography

RICHARDSON, O. W.: "Emission of Electricity from Hot Bodies," Longmans, Green & Company, New York, 1916.

DUSHMAN, S.: Thermionic Emission, *Rev. Modern Phys.*, **2**, 381 (1930).

[1] SALZBERG, B., and BURNSIDE, D. G., *Proc. I.R.E.*, **23**, 1142 (1935).

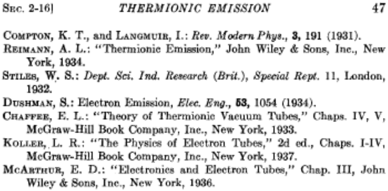

COMPTON, K. T., and LANGMUIR, I.: *Rev. Modern Phys.*, **3**, 191 (1931).

REIMANN, A. L.: "Thermionic Emission," John Wiley & Sons, Inc., New York, 1934.

STILES, W. S.: *Dept. Sci. Ind. Research (Brit.), Special Rept.* 11, London, 1932.

DUSHMAN, S.: Electron Emission, *Elec. Eng.*, **53**, 1054 (1934).

CHAFFEE, E. L.: "Theory of Thermionic Vacuum Tubes," Chaps. IV, V, McGraw-Hill Book Company, Inc., New York, 1933.

KOLLER, L. R.: "The Physics of Electron Tubes," 2d ed., Chaps. I-IV, McGraw-Hill Book Company, Inc., New York, 1937.

MCARTHUR, E. D.: "Electronics and Electron Tubes," Chap. III, John Wiley & Sons, Inc., New York, 1936.

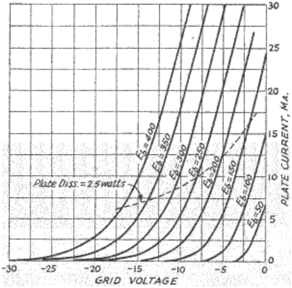

Fig. 6

enough points being taken at each plate voltage to permit smooth curves to be drawn. Notice that for each value of plate voltage the curve bends at the higher values of negative grid voltage (as the plate current decreases toward the cutoff point) but that the curvature decreases as the grid bias becomes less negative. The curves eventually straighten out and become practically parallel, and the distances between the 50-volt intervals also approach equality. The dashed line shows the value of plate current at which the plate dissipation (plate voltage multiplied by plate current) is equal to the maximum rated value for the tube; above this line the plate dissipation is exceeded.

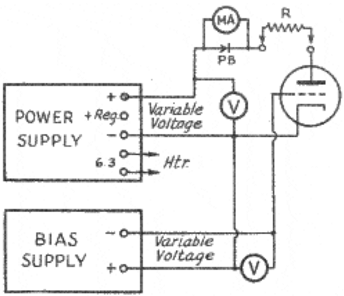

Fig. 5

The resistor R shown in Fig. 5 is not needed in this experiment, so the push-button may be connected directly to the plate.

Procedure: The object of the experiment is to determine the relationship between plate voltage, plate current and grid voltage of a small triode. One quantity is held constant throughout a run, the second is varied, and corresponding measurements of the third are made. A receiving triode such as the 6J5 is suitable. Three sets of characteristics can be taken; the first, with the plate voltage held fixed while the behavior of plate current with varying grid voltage is observed, is called the "grid voltage plate current" characteristic. When a series of such data is taken with several fixed values of plate voltage, a "family" of curves results. A typical grid-voltage plate-current family taken in this way on a 6J5 is shown in Fig. 6. The plate voltage was set at 50- volt intervals from 50 to 400 volts (the maximum output voltage of the power supply described in August QST),

---

RCA - Vacuum Tube Design - 1940

Lecture 19 ANALYSIS OF RECTIFIER OPERATION

0. H. Schade

INTRODUCTION

The design of diodes for rectification of a-power requires unusual care even though their design is simple from the standpoint of the number of electrodes involved. The cathode must be capable of supplying the average current and peak current which will be encountered in ser-vice. The plate must operate at a safe temperature, which in turn depends on plate dissipation, the physical dimensions of the plate, and the thermionic and thermal emissivity of the plate material. Another factor to be considered is the purveyance of the diode, i.e., the steepness of its voltage-current characteristic. Purveyance affects both cathode— and plate-design because it controls current density and voltage gradient at the effective surfaces of both for a given cathode area. Then, too, there is the requirement for adequate insulation. Perveance and insulation requirements depend on the type of service in which the diode is to be used.

It is apparent that an analysis of diode-operating conditions in rectifier circuits is essential in order to arrive at optimum design specifications in each particular case, because the required data bear no simple relationship to the values of rectified current and voltage.

PRINCIPLES OF RECTIFICATION

1. General

Rectification is a process of synchronized switching. The basic rectifier circuit consists of one synchronized switch in series with a single-phase source of single frequency and a resistance load. The switch connection between load terminals and source is closed when source and load terminals have the same polarity, and is open during the time of opposite polarity. The load current consists of half-wave pulses. This simple circuit is unsuitable for most practical purposes, because it does not furnish a smooth load current.

The current may be smoothed by two methods:(a) by increasing the number of phases, and (b)by inserting reactive elements into the circuit. The phase number is limited to two for radio receivers. The circuit analysis which follows later en will treat single- and double-phase rectifier circuits with reactive circuit elements.

Switching in reactive circuits gives rise to “transients." Current and voltage can, there-fore, not be computed according to steady-state methods.

The diode functions as a self-timing electronic switch. It closes the circuit when the plate becomes positive with respect to cathode and opens the circuit at the instant when the plate current becomes zero.

The diode has an internal resistance which is a function of current. When analyzing rectifier circuits, it is convenient to treat the internal resistance of the diode rectifier as an element, separated from the "switch action" of the diode. Fig. 1 illustrates the three circuit elements so obtained. The diode characteristic is the geometric sum of these characteristics. The resistance r^ is effective only when the switch is closed, i.e., during the conduction period of the diode. The effective diode resistance must, therefore, be measured or evaluated

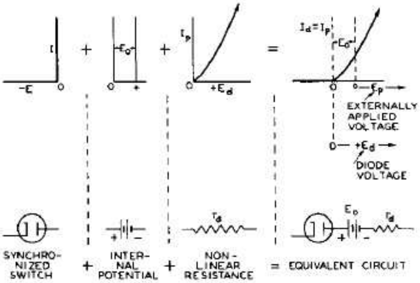

Fig 1

small fraction of a milliampere - it can be neglected in most applications of the tube. However, in flowing through an external load resistance of high value a volt or two may be developed across the load, which may need to be taken into account in some cases.

Triode Static Characteristics

Experiment 22

Apparatus: The set-up for this experiment is shown in Fig. 5. Insofar as the plate circuit of the triode is concerned. the arrangement is practically the same as that used for diode measurements, :Fig. 3, except that it is possible to measure plate voltage with the test instrument rather than the v.t. voltmeter. This is because larger plate voltage steps may be used so that a high range (500 volts or the nearest provided on the test instrument), which will have a resistance of a half megohm or so, will give sufficient accuracy for all measurement. The bias supply is incorporated in the

set-up to provide variable grid bias, and its voltage output also may be measured by the test instrument on the condition that the voltmeter resistance is 25,000 ohms or so (25-volt scale). Be sure that the positive output terminal of the bias supply is connected to the grounded side of the 115-volt line, using the lamp provided for checking as described in July QST. In using a single instrument in place of the three indicated, the push-button should be closed each time the plate voltage is measured so that the voltage will be that existing when plate current flows.

the behavior of the tube when an alternating voltage is applied is of more interest, in which case the a.c, plate resistance, or resistance effective to small changes in applied voltage, is important. The value of this plate resistance is found as described in the introduction to this installment.

When a load resistance is inserted in the plate circuit the linearity of the circuit consisting of the resistance and the tube is better than that of the tube alone. This improvement, which increases as the load resistance is increased, is because the load resistor tends to reduce the effect of variations in the resistance of the tube. For example, if the resistance of the tube varies between 1000 and 3000 ohms with a

certain range of applied voltage the resistance change is 2000 ohms, or an increase of 200%, using the smaller number as a base, If a 10,000-ohm resistor is connected in series, the minimum resistance becomes 11,000 ohms and the maximum resistance 13,000 ohms, so that the increase in resistance is now only 2000/11,000, or 18%. With 100,000 ohms in series, the increase is from 101,000 to 103,000 ohms, so that the percentage increase is now 2%. In the curves of Fig. 3 the addition of the load resistance makes all the points fall on a line which is practically straight except at the low voltage end where the tube resistance has its highest value. The higher the load resistance the less marked does this slight curvature become.

In taking data it will be observed that a small current flows in the plate circuit even at zero plate voltage.

This current is the result of the fact that some electrons are emitted from the cathode with sufficient velocity to reach the plate even though there is no positive charge on the plate to attract them. For complete cut-off of plate current

it would be necessary to make the plate a volt or two negative with respect to the cathode, thus repelling these high energy electrons from the plate. Since the current in any case is very small - a very

within conduction-time limits. Consider a switch in series with a fixed resistance and any number of other circuit elements connected to a battery of fixed voltage. The d-c current and rms current which flow in this circuit will depend on the time intervals during which the switch is closed and open; the resistance value is not obtainable from these current values and the battery voltage. The correct value is obtained only when the current and voltage drop in the resistance are measured during the time angle φ (Fig. 2) when the switch is closed.

The method of analysis of rectifier circuits to be discussed in this lecture is based on the principle that the non-linear effective resistance of the diode may be replaced analytically by an equivalent fixed resistance which will give a diode current equal to that obtained with the actual non-linear diode resistance. The correct value to be used for the equivalent fixed resistance depends upon whether we are analyzing for peak diode current, average diode current, or rms diode current. As will be seen later, the relations among these three equivalent resistances depend upon the circuit under consideration.

2. Definitions of Resistance Values

The instantaneous resistance (r^) of a diode is the ratio of the instantaneous plate voltage e^, to the instantaneous plate current, ip, at any point on the characteristic measured from the operating point (see Fig. 1). It is expressed by the equation

$$r_d = \frac{e_d}{i_p}$$

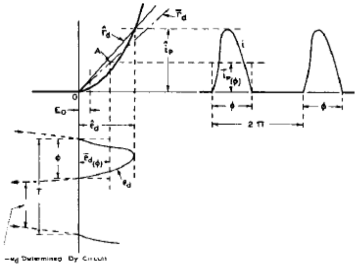

Fig 2

The operating point (0) of a diode is a fixed-point on the characteristic, marked by beginning and end of the conduction time. It is, there-fore, the cut-off point Id = 0 and Ed = 0, as shown in Fig. 1. The operating point is independent of the waveform and of the conduction time φ (see Fig.2).

The peak resistance (¾) is a specific value of the instantaneous resistance and is defined as

$$\hat{r}_d = \frac{\hat{e}_d}{\hat{i}_p} \quad \text{(See Fig.2)}$$

Peak voltage e<j and peak current ip are measured from the operating point 0.
The equivalent average resistance (rd) is de-fined on the basis of circuit performance as a resistance value determining the magnitude of the average current in the circuit. The value r^ is, therefore, the ratio of the average voltage drop β(}(φ) in the diode during

necessary to make provision for closing the plate circuit of the v.t.v.m. when the meter is being used elsewhere.

The observed data should be plotted in the fashion shown in Fig. 4, which gives characteristic curves taken on a 6H6. With no load the current is quite high, reaching 10 milliamperes with about 7.5 volts applied. Other types of tubes may give considerably different plate current values without load, but should approximate the load curves given since the current which flows at a given voltage is principally determined by the load resistance rather than the tube. As is to be expected, the current decreases, at a given applied voltage, as the load resistance is increased.

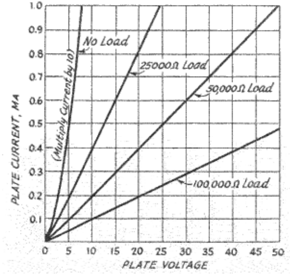

Fig 4

If the no-load curve is inspected carefully, it will be observed that it is not a straight line, particularly near the low-voltage end. The lamp in Exp. 10 was another example of a non-linear circuit, although for a different reason. In the present case, the nonlinearity arises from the fact that the number of electrons drawn to the plate is not strictly proportional to the voltage applied between plate and cathode, The d.c. resistance of the diode at any voltage is equal to that voltage divided by the current which it forces through the tube. In practice

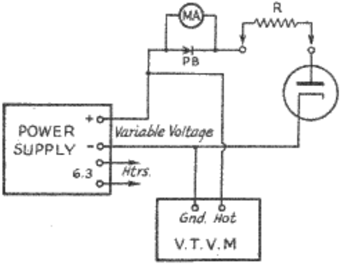

Fig 3.

Procedure: The object of the experiment is to plot characteristic curves, plate voltage vs. plate current for the tube alone (static characteristic) and with various values of load resistance in series with the plate circuit (dynamic characteristics). Starting at zero plate voltage, increase the plate voltage in small steps, taking plate current readings at each voltage step. With no load resistor in the circuit, take readings at intervals of voltage which will give current intervals of about 1 milliampere so that enough points will be secured to give a smooth curve when the points are plotted. In the case of the 6H6 tube, using one plate and cathode only, one-volt intervals are suitable. Proceed similarly when the load resistance is inserted in the circuit; in this case larger voltage intervals (5-volt steps, for instance) can be used

In using the single test set for all measurements. the pushbutton should be closed while the voltage measurement is being made so that the voltage can be adjusted to the proper value with plate current flowing. If the plate circuit is not closed at the time the voltage is adjusted, the voltage will drop when the milliammeter is connected in the plate circuit of the tube to measure plate current. It is not

conduction time to the average current ip(<p) during conduction time, or

$$\bar{r}_d = \frac{\bar{e}_{d(\varphi)}}{\bar{i}_{p(\varphi)}}$$

The curved diode characteristic is thus replaced by an equivalent linear characteristic having lie slope ?d and intersecting the average point A, as shown in Fig. 2. The coordinates ed(q>) and ip (φ) of the average point depend on the shape of voltage and current within the time angle φ.

The analysis of rectifier circuits shows that the shape of the current pulse in actual circuits varies considerably between different circuit types. The equivalent rms resistance (|r^I) is de-fined as the resistance in which the power loss Pd is equal to the plate dissipation of the diode when the same value of rms current II^I flows in the resistance as in the diode circuit. It is expressed by the equation:

$$|r_d| = \frac{P_d}{|I_d|^2}$$

3. Measurement of Equivalent Diode Resistances The equivalent resistance values of diodes can be measured by direct substitution under actual operating conditions. The circuit arrangement is shown in Fig. 3. Because the diode under test must be replaced as a whole by an adjustable resistance of known value, a second switch (a mercury-vapor diode identified in the figure as the ideal diode) with negligible resistance must be inserted in order to preserve the switch-action in the circuit.

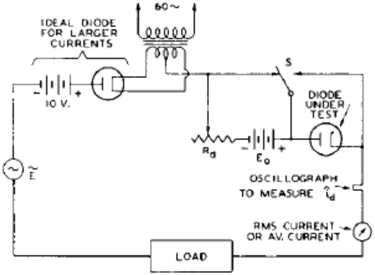
60~
IDEAL DIODE FOR LARGER CURRENTS
10 V.
S
DIODE UNDER TEST
Rd
E0
~E
OSCILLOGRAPH TO MEASURE îd
RMS CURRENT OR AV. CURRENT
LOAD

Fig. 3 - Half-wave circuit for measuring (d)\> rd, and |rd|.

When a measurement is being made, the resistor Rd is varied until the particular voltage or cur-rent under observation remains unchanged for both positions of the switch S. We observe (1)that it is impossible to find one single value of 1¾ which will duplicate conditions of the actual tube circuit, i.e., give the same values of peak, average, and rms current in the circuit;(2) that the ratio of these three "equivalent" resistance values of the diode varies for different combinations of circuit elements; and (3)that equivalent average or rms substitution resistances may have different values when the average cr rms current in the diode circuit is adjusted to the same value in different types of rectifier circuits. The "peak resistance" value obtained for a given peak current is found to be substantially independent of the type of circuit.

$$\hat{r}_d = \bar{r}_d = |r_d|$$

Diode Characteristics

A Course in Radio Fundamentals
Lessons in Radio Theory for the Amateur
By George Grammer, W1DF
http://www.rfcafe.com/references/qst/radio-fundamentals-vacuum-tubes-sep-1942-qst.htm

Experiment 21

Apparatus: This experiment uses the plate power supply, tube board, test set, vacuum-tube voltmeter, and three 1- watt resistors, 25,000, 50,000 and 100,000 ohms. The circuit arrangement is shown in Fig. 3. Measurements must be made of the voltage applied to the tube and the current flowing in its plate-cathode circuit; the single test instrument can be used for both purposes by being shifted back and forth for each pair of readings. However, the small current consumed by the instrument when used as a voltmeter will cause the actual output voltage to be lower when the voltage is being measured than when the instrument is shifted to read plate current. Unless a separate voltmeter which can be left permanently in the circuit is available, it is advisable to use the v.t. voltmeter, thus avoiding the loading effect. The test instrument is therefore shifted between the plate circuit of the tube being tested and the plate circuit of the voltmeter tube.

The tube to be tested may be a 6H6, the diode section of a combination diode-amplifier tube, or simply a small triode such as the 6J5 with the grid and plate connected together to act as a single plate.

Λ. Waveforms and Equivalent Resistance Ratios for Practical Circuit Calculations

The form of the current pulse in practical rectifier circuits is determined by the power factor of the load circuit and the phase number. Practical circuits may be divided into two main groups: (a) circuits with choke-input filter; and(b) circuits with condenser-input filter. The current pulse in choke-input circuits has a rectangular form on which is superimposed one vapor diodes.

Table I

CONDUCTION TIME ANGLE φ	WAVE SHAPE	$\frac{\bar{i}_{p_\varphi}}{\hat{i}_p}$	$\frac{\lvert i_p\rvert_\varphi}{\hat{i}_p}$	3/2-POWER RECTIFIER CHARACTERISTIC			RECTANGULAR CHARACTERISTIC		
				$\frac{\bar{e}_{d_\varphi}}{\hat{e}_d}$	$\frac{\bar{r}_d}{\hat{r}_d}$	$\frac{\lvert\bar{r}_d\rvert}{\hat{r}_d}$	$\frac{\bar{e}_{d_\varphi}}{\hat{e}_d}$	$\frac{\bar{r}_d}{\hat{r}_d}$	$\frac{\lvert\bar{r}_d\rvert}{\hat{r}_d}$
CONDENSER-INPUT CIRCUITS									
≦ 20°	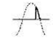	0.500	0.577	0.593	1.185	1.120	1.0	2.00	1.500
90° & 180°	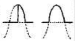	0.637	0.707	0.715	1.120	1.057	1.0	1.57	1.272
130°	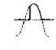	0.725	0.780	0.787	1.085	1.030	1.0	1.38	1.190
CHOKE-INPUT CIRCUITS									
180°	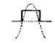	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

In this table, the designation the rms value of the current i-ρΐφ represents the rms during the conduction time. It follows that the relation

$$\hat{r}_d = 0.88\ \bar{r}_d = 0.93\ |r_d|$$

is representative for the group of condenser-input circuits containing high-vacuum diodes, and holds within ± 5 per cent over the entire range of variation in wave shape. The actual error In circuit calculations is smaller, as the diode resistance is only part of the total series resistance in the circuit.

EMISSION AND SATURATION OFOXIDE-COATED CATHODES

1. Oxide-Coating Considerations

The normal operating range of diodes (including instantaneous peak values) is below the saturation potential because the plate dissipation rises rapidly to dangerous values if this potential is exceeded. Saturation is definitely recognized in diodes with tungsten or thoriated-tungsten cathodes as it does not depend on the time of measurement, provided the plate dissipation is not excessive. The characteristic of such diodes is single-valued even in the saturated range, i.e., the range in which the same value of current is obtained at a given voltage whether the voltage has been increased or decreased to the particular value.

Oxide-coated cathodes have single-valued characteristics up to saturation values only under certain conditions. The cathode coating has resistance and capacitance, both of which are a function of temperature, current, and the degree of "activation," and vary during the life of the tube. The conception of a "coating impedance” is essential in explaining many peculiar effects observed when testing and operating oxide-coated diodes in rectifier circuits and will be treated briefly.

The cathode coating is usually a mixture of carbonates (barium, strontium, calcium) which are "broken down" into oxides, metal atoms, and gases during the exhaust process. A highly emit-ting monatomic layer of barium on oxygen is formed on the surface of

If we substitute the values obtained in the Barkhausen equation we obtain

DSRi = 0.02*5.7*9.4 = 1.07 ~ 1.

Notes

Strictly speaking the anode voltage UA is always obtained from the applied voltage minus the difference between the electron work function voltages at the cathode phiC and the anode phiA:

UA = U – (phiA – phiC) .

Secondary phenomena caused by the Maxwellian velocity distribution of the electrons are also neglected.

The open-loop voltage gain μ = D^-1 is often quoted instead of the inverse amplification factor D.

When grid and anode are connected together, the greater part of the current flows through the grid.

B1 = 8.36 V^-1 (150 mA)
B2 = 9.84 V^-1 (140 mA)
B3 = 11.24 V^-1 (130 mA)

and hence the cathode temperatures T:

T1 = 1390 K (150 mA)
T2 = 1180 K (140 mA)
T3 = 1030 K (130 mA)

The most probable velocities at different cathode temperatures are then:

vp (T1) = 204 · 10^3 m s^-1
vp (T2) = 189 · 10^3 m s^-1
vp (T3) = 177 · 10^3 m s^-1

2. Fig. 6 shows the triode control characteristics measured. If we take working point A (UA, IA, UG) in the linear portion of the family characteristics we can determine all the characteristics of the tube. For A (100 V, 20 mA, 1 V) we obtain

S = 10 mA/V

for the “slope” S (gradient of the curve at point A).

The inverse amplification factor D is obtained by going from the 150 V to the 50 V characteristic through A and parallel to the UG-axis and reading off

– dUG/dUA , from which D = 0.018.

To obtain the tube resistance Ri, we go from the 50 V to the 150 V characteristic through A and parallel to the IA-axis and read off

dUA/dIA, from which Ri = 5.85 kohms.

the coating, which, when heated, supplies the electron cloud forming the space charge above the coating surface (see Fig. 4) ·The» coating itself consists of non-conducting oxide clumps or crystals (shown as shaded areas)interposed with metal atoms and ions (circles),which axe produced by high cathode temperature and electrolysis during the activation and aging process of the cathode.

The distances between these metal atoms or atom groups are very small. The current flow from the base material through the coating is effected by relay-emission from atom to atom or particle to particle under the influence of electrostatic potentials. The coating is not con-ducting when cold. Electronic conduction through the coating is high when many metal relay chains, not broken by insulating oxides, have been formed,

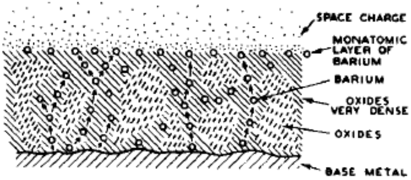

Fig. 4 - Structure of cathode coating.

and also when the electron emission is raised as a result of loosening the atomic structure by increased temperatures. Hence, the coating conductance may be represented by a large number of extremely close-spaced diodes in series-parallel arrangement, as shown in Fig. 4a

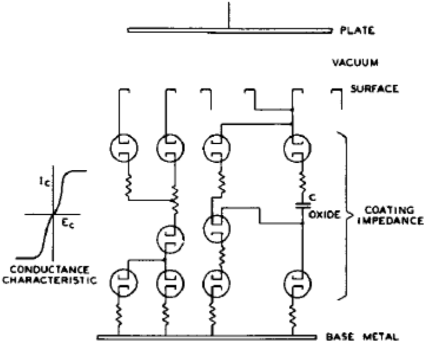

Fig. 4a - Representative diode network.

The coating activation is good when a fully emitting surface layer can be saturated without limitation of electron flow in the coating. This condition is indicated by a single-valued characteristic up to and beyond the saturation potential as shown in Fig. 5. In general, single-valued characteristics are obtained if the surface saturates before the coating conductance be-comes limited. (See temperature—limited conditions in Fig. 5).

The conductance of the coating may, however, limit the electron flow before the surface emission is saturated, and cause a peculiar voltage-current characteristic. Consider a high plate

where No is the total number of electrons and m (the restmass of the electron) is 9.11*10E-31 kg.

From this we obtain the most probable velocity

vp = sqrt4/pi * vp

and the mean velocity

vm = sqrt4/pi * vp

By fitting to exponential curves, using the expression

IA = A * e^B*UA

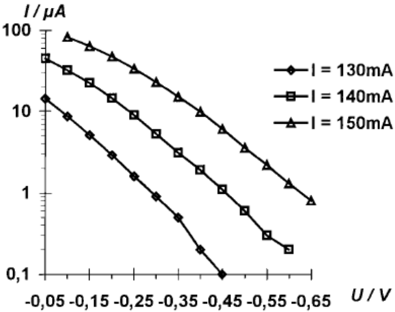

Fig. 5: Initial current of the diode at different heater currents.

we obtain from the measured values in Fig. 5 the exponents B, as follows:

The gradient of the IA/UG characteristic at working point A is called the “slope” S or “mutual conductance”:

$S = dIA/dUG$ at constant UA .

We obtain likewise the tube resistance Ri at working point A and at constant grid voltage:

$Ri = dUA/dIA$ at constant UG .

These three variables are interrelated by the Barkhausen equation (the “tube equation”):

$DSRi = 1$

1. If we plot the anode current against the (negative) and voltage semi-logarithmically in the initial current range of the diode we obtain a straight line of slope

$b = – e/kTc$ (cf. Fig. 5)

in accordance with

$\ln IA = \ln Io – |eUA/kTc|$,

where e = the electron charge = 1.60*10E-19 As, and k = the Boltzmann constant = 1.38*10E-23 VAsk^-1.

From this, we can calculate the cathode temperature

$Tc = – e/kB$.

For a Maxwellian distribution the electron velocites n are calculated in accordance with

$1/No\ dN/dv = 4/\sqrt{\pi}\ (m/2kT)^{3/2} * v^2 * \exp -mv^2/2kT$

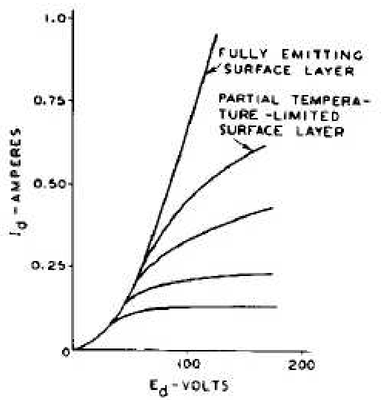

Fig. 5 - Single-valued saturation characteristic of diode with high coating conductance.

voltage suddenly applied over a periodic switch to a diode as in the circuit of Fig. 6. If the coating is not limiting, the current obtained is that at a point P on the corresponding diode characteristic. Hence, the current waveform in the circuit is as shown in Fig. 6a.

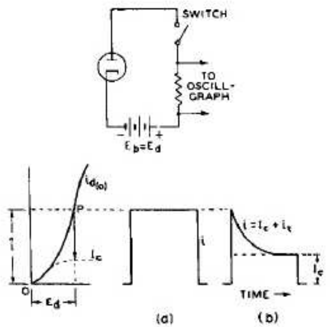

If the surface emission is assumed to be unchanged, but all free electrons have been moved. The oxide groups thus act like condensers in parallel to the coating resistance but with the peculiarity that their charge may be limited by hypothetical series diodes.

The coating resistance is extremely low* below saturation, due to the small spacing and high gradient in the "coating-diodes" but becomes in-finite when the conduction current is saturated; hence, the charging current must flow in the plate circuit (external) of the diode. The total plate current is, therefore, the sum of the conduction current Ic and the "transient" electron displacement current. The "coating transient” decays to zero like normal transients at a rate depending on the actual shunt conductance value and the total series resistance in the circuit(Fig. 6b). The decay can be changed by adding external resistance in the plate circuit. When the surface emission is good, i.e., as long as the total vacuum-space plate current is space-charge limited, the current will rise initially to the value

---

The emitted electrons first form round the cathode an electron cloud from which slow-moving electrons rebound and can thus return to the cathode. If we apply a positive voltage to the anode, some electrons are drawn out and the cloud becomes less dense as the anode voltage increases. In this part of the curve (space-charge region) the equation:

$IA = P \cdot UA^{3/2}$,

applies, where P = the tube constant.

As the anode voltage is increased still further, in the end all the electrons emitted from the cathode are collected, the space charge disappears and the tube reaches saturation. Changing the anode voltage no longer brings about a corresponding change in the anode current.

If we insert another grid-like elektrode between cathode and anode we obtain a three-electrode valve, or triode. The anode current can be controlled by a voltage applied between the grid and the cathode. But as it is controlled also (but to a lesser extent) by the anode voltage UA, the two voltages combine to give a resultant control voltage

$Uc = UG + D \cdot UA$.

The so-called “inverse amplification factor” D is a constant at a given working point A and is defined by

$D = dUG/dUA$ at constant IA .

Since electron tubes generally work in the space-charge region, the anode current is essentially expressed by the space-charge formula

$IA = P\,(UG + D \cdot UA)^{3/2}$

Theory and evaluation

We distinguish three different regions in the current/voltage characteristic of a vacuum diode: the initial current, space charage and saturation regions.

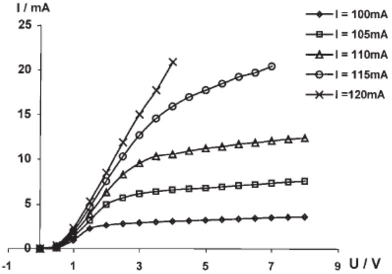

Fig. 4: Current/voltage characteristic of the diode for different heater currents.

A current still flows through the vacuum diode when the anode is negative with respect to the cathode $(UA < 0)$. This current, which is known as the “initial current”, bears the following relationship to the (negative) anode voltage:

$$IA = I0 \cdot \exp - |eUA/kT|.$$

The electrons contributing to this current still have, as the result of their Maxwellian velocity distribution, sufficient kinetic energy after they have left the cathode to surmount the anode field and reach the anode.

(point P) determined by the applied potential, but will then decay to the saturation value determined by the coating conductance.

The condition of oxide-coated cathodes can, therefore, not be judged alone by their capability of furnishing high peak currents, but the rate of change in current flow, and hence the current waveform, must also be carefully considered, because the diode characteristic may not be single-valued. Fig. 7 shows characteristics which are not single-valued. These were taken within 1/120 of a second (l/2 of a 60-cyclesine wave) with a cathode-ray curve tracer.

2. Current Overload and Sputter

The degree of activation is not stable during the life of the cathode. Coating conductance and surface emission change. Factors affecting the change are the coating substances, the evaporation of barium depending on the base material, and the operating conditions to which the cathode is subjected. This life history of the cathode is the basis on which current ratings are established. Rectifier tubes especially are subject to severe operating conditions. If a diode is operated with too high a current in a rectifier circuit and its surface emission is decreased to the saturation value, then the tube voltage drop will increase rapidly, and cause excessive plate dissipation and destruction of the tube. Should the coating conductance in this diode decrease to a value which limits the demanded current, power is dissipated in the now-saturated "coating-diodes," with the result that the coating voltage drop and coating temperature are raised. The temperature rise may cause reactivation but also may become cumulative and melt the coating mate—

* Its magnitude depends on the number of series diodes and, hence, on the barium content and thickness of the coating material.

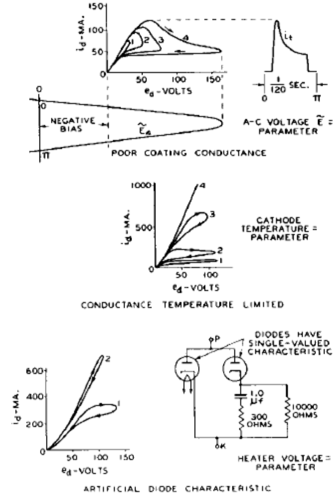

Fig. 7

Vapor or gas discharges may result from saturated coatings and cause breakdowns during the inverse voltage cycle. These breakdowns are known as "sputter," and destroy the cathode. Sputter of this type will occur more readily in diodes with long cylindrical electrodes and close spacing than in diodes with flat cathodes of narrow width and parallel—plane plates, because the activation gas can diffuse faster in the latter. A second type of sputter is caused by the in-tense electrostatic field to which projecting “high spots" on the

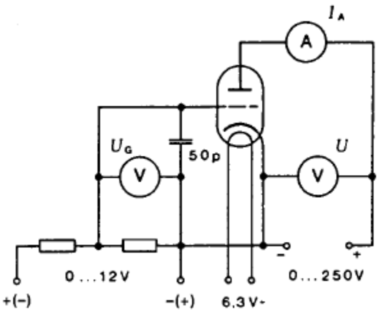

Fig. 3: Circuit for measuring the characteristic curves of a triode.

2. Construct the circuit shown in Fig. 3. As the tubes may oscillate in the VHF range if the wiring is not altogether satisfactory, connect a small capacitor between grid and cathode to short-circuit the high frequency.

Measure the anode current as a function of the grid voltage, both positive and negative, at various anode voltages (e. g. 50, 75, 100, 125, 150 V). Stop taking measurements as soon as the anode current exceeds 30 mA.

Because of the voltage drop across the internal resistance Ri of the ammeter, the anode voltage is represented by
$UA = U - IA \cdot Ri$

Plot the IA/UG-characteristics (control characteristics) on a graph.

Plot the initial current characteristics at different heater currents also. To do this, reverse the polarity on the voltage source and the voltmeter multimeter and measure the anode current with the microammeter (multimeter).

Because of the voltage drop across the microammeter (multimeter) the anode current UA is:

$UA = – (|U| + |IA \cdot Ri|)$

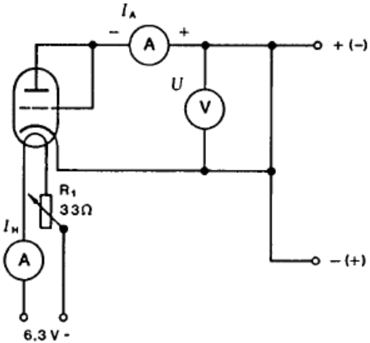

Fig. 2: Circuit for measuring the characteristics of a diode.

---

plate or cathode are subjected. These high spots are formed by loose carbon on the plate or large cathode—coating particles. The resulting current concentration causes these spots to vaporize with the result that an arc may be started.

3. Hot-Cathode Mercury-Vapor Diodes

The ionization potential Ej. of mercury vapor is a function of the gas pressure and temperature. It is approximately 10 volts in the RCA-83and similar tubes. A small electron current be-gins to flow at Ep = 0 (see Fig. 8), and causes ionization of the mercury vapor at Ep = Ej_. This action decreases the variational diode resistance Γρ to a very low value. The ionization becomes cumulative at a certain current value (rp = 0 at40 milliamperes in Fig. 8a), and causes a discontinuity in the characteristic. Hence, it is not single-valued within a certain voltage range. Beyond this range (see Fig. 8b), the slope (rp) of the characteristic becomes again positive until saturation of the emitter is reached.

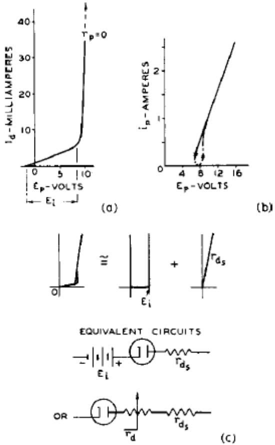

Fig. 8

For circuit analysis, the mercury-vapor diode may be replaced by a bucking battery having the voltage Ej_ and a fixed resistance,

$$r_{d_s} \cong \frac{\hat{e}_d - E_i}{\hat{i}}$$

as shown in Fig. 8c; or the diode characteristic may be replaced by an ideal rectangular characteristic and its equivalent resistance values and the series resistance rd^ as shown.

Characteristic curves of electron tubes (diode, triode)

PHYWE series of publications • Laboratory Experiments • Physics • PHYWE SYSTEME GMBH • 37070 Göttingen, Germany
http://www.nikhef.nl/~h73/kn1c/praktikum/phywe/LEP/Experim/4_1_10.pdf

Principle and task

The IA/UA-characteristic curve of a diode is recorded at different heater currents and the cathode temperature and electron velocity determined therefrom. Mutual conductance, inverse amplification factor and anode resistance are determined from the characteristic curve of a triode.

Problems

1. To measure the anode current of a diode as a function of the anode voltage at different heater currents and to plot it on a graph. To calculate the cathode temperatures and electron velocities from the initial current characteristics.

2. To record the anode current of a triode as a function of the grid voltage at different anode voltages. To determine the mutual conductance, inverse amplification factor and internal resistance of the tube at a working point in the linear portion of the characteristic.

Set-up and procedure

1. Construct the circuit shown in Fig. 2. Connecting the anode and the grid gives tube EC 92 (triode) the properties of a diode.
The heater current and thus the cathode temperature can be set with the potentiometer R1.

Measure the anode current IA as a function fo the anode voltage UA at different heater currents (150–110 mA) and plot the results on a graph. The voltage drop across the internal resistance Ri of the ammeter (multimeter) must be taken into account, so that UA = U–IA · Ri. Stop taking measurements
when IA becomes greater than 30 mA.

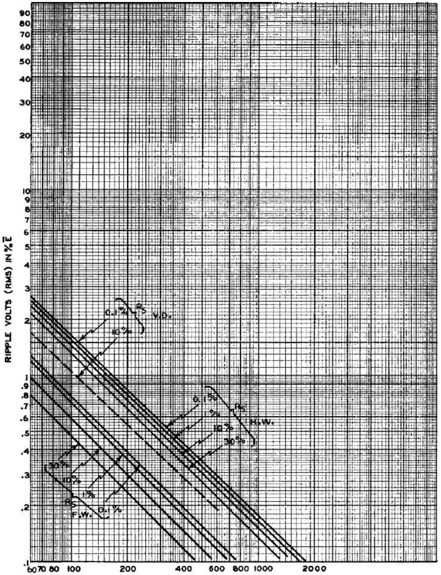

The first representation is advantageous when making approximate calculations, as the equivalent diode resistance is then the substantially fixed value rdg. The value rds is in the order of 4 ohms. The low series resistance and the small constant voltage drop Ei are distinct ad-vantages for choke-input filters, as they cause very good regulation; the low resistance, how-ever, will give rise to enormously high storting transients in condenser-input circuits, in case all other series resistances are also small. Mercury-vapor diodes as well as high-perveance (close-spaced), high-vacuum diodes having oxide cathodes should, therefore, be protected against transient-current overloads when they are started in low-resistance circuits to prevent destruction of the cathode coating. The destruction of the coating in mercury-vapor diodes is apparently caused by concentration of current to small sections cf the coating surface and not by heat dissipation in the coating.

CIRCUIT ANALYSIS

The rectifier diode is a switch operated in synchronism with the applied a-c frequency. The diode alternately opens and closes the circuit. This action causes a series of transients in re-active circuits. According to the decay time of the transients, fundamental rectifier circuits may be classified into two principal groups: (l)circuits with repeating transients in which the energy stored in reactive elements decreases to zero between conduction periods of the diode; and (2) circuits with chain transients in which(a) the magnetic energy stored in the inductance of the circuit remains above zero value, and (b)the electric energy stored in the capacitance of the circuit remains above zero value. The much used "choke-input" and "condenser-input" circuits fall under the second group.

The complete analysis of rectifier-circuit operation requires more time than this lecture provides. It is based on the fact that the total current in a circuit is the sum of all steady-state currents and transient currents within the time between two switching operations. We will analyze briefly the operation in two important circuits, i.e., the full—wave, choke-input circuit, and the condenser-input circuit.

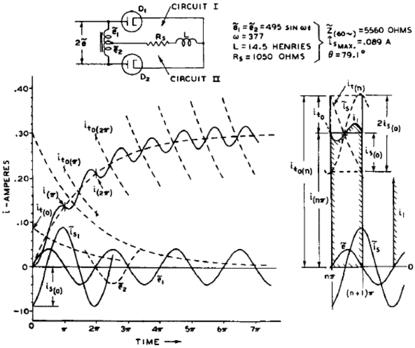
CIRCUIT I
CIRCUIT II
D1
D2
2e
e1
e2
Rs
L
e1 = e2 = 495 SIN ωt
ω = 377
L = 14.5 HENRIES
Rs = 1050 OHMS
Z(60~) = 5560 OHMS
is MAX. = .089 A
θ = 79.1°
i - AMPERES
.40
.30
.20
.10
0
-.10
TIME
0
π
2π
3π
4π
5π
6π
7π
it0(2π)
it0(π)
i(π)
i(2π)
it(0)
is1
is(0)
e2
e1
it(n)
is
i1
it0
2is(0)
is(0)
it0(n)
i(nπ)
e
is
i1
nπ
(n+1)π
0

Fig. 9 - Full-wave choke-input circuit and its operation.

1. The Full-Wave Choke-Input Circuit

Circuit and operation are shown in Fig. 9. The construction is made by considering first one of the diodes shorted to obtain the phase relation of the a-c voltage e, and the steady-state current is, as shown. If we assume that the diode closes the circuit at the time e" = 0, a transient it with the initial value it0= -is(0) will flow in the circuit. The total current, i, is the sum of the currents is + 1¾. It starts, therefore, at zero and rises as shown until the second switching operation occurs at the commutation time t =η when the second diode receives a positive plate voltage. The total current, i,

RMS RIPPLE VOLTAGE OF CONDENSER INPUT LOAD CIRCUITS (lN ?°E) FOR VARIOUS VALUES OF Rs (GIVEN IN ?o R|_)
(MEASURED VALUES}

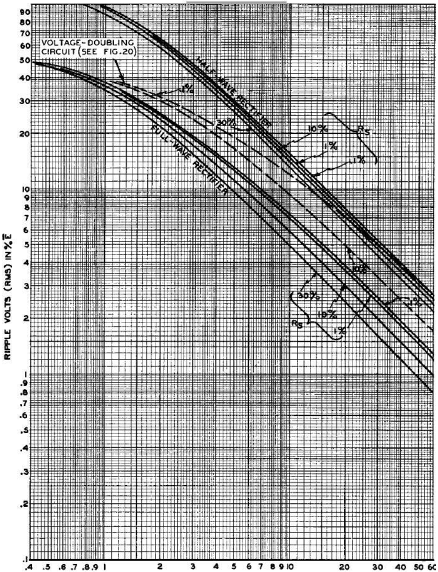

RELATIONS OF PEAK, AVERAGE, AND RMS DIODE CURRENT

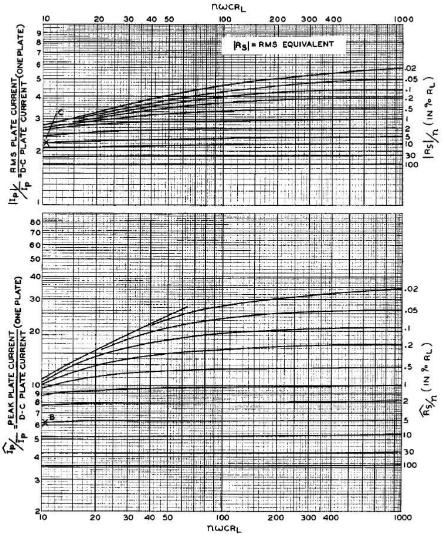

after t = η is again the sum of currents xQ+ i-^, butte initial value if Q is increased by the value ï(π) still flowing in the circuit.

The current it0 increases, therefore, at every new switching time until the decay of the transient it(n) > during the time t = π, is numerically equal to the steady-state current rise 2Ïβ(0),as shown in Fig. 9* This leads to the equation for the final operating current at the n™ com-mutation time:

$$i_{n(\pi)} = - \tilde{i}_{s(o)} \frac{1 + \varepsilon^{-\frac{R}{2fL}}}{1 - \varepsilon^{-\frac{R}{2fL}}}$$

A broken line is shown connecting all commutation-current values. This line represents closely the average current I in the common circuit branch. The final average current ig, during the conduction time, and _the average current Iin the load resistance R, have the same value and are given by equation (7)_. The average plate current per diode is Ip =0.5I, since each diode conducts alternately, and passes a current pulse shown by the shaded area in Fig. 9. With the numerical values of the construction substituted in equation (7) and setting

$$\varepsilon^{-\frac{R}{2fL}} = 0.547$$

we obtain

$$i_{(n\pi)} = \bar{I} = 0.298 \text{ ampere.}$$

The described solution is illustrative in understanding starting and operating conditions of the choke-input circuit, but its accuracy is limited in cases of low power factor (large L or small R), since the decay terms in equation (7)become very small. The construction, however, shows the exact phase relations of current and voltage. The oscillogram in Fig. 10 was taken with the circuit values indicated in Fig. 9.

a. Critical Inductance — The analysis has shown that the transient currents in the particular case never decay to zero and are "chain transients.^ In practical circuits, the d-c load resistance R is shunted by a capacitance C. This does not affect the circuit performance, provided R does not increase beyond a critical value Rcr. Expressing this statement symbolically, we have

$$\bar{R} \leqq \bar{R}_{cr} = 1.5\ Z_{(2f)}$$

in which Z(2f) is the impedance of one branch circuit at double line-frequency. For large values of C,

$$Z_{(2f)} \cong 4\pi f L$$

$$\bar{R} \leqq 6\pi f L \quad \text{or} \quad L \geqq \frac{\bar{R}}{6\pi f}$$

$$L \geqq \frac{0.91\ |\tilde{E}|}{6\pi f\ \bar{I}}$$

where

$|\tilde{E}|$ = rms voltage of one-half the secondary winding

and

$\bar{I}$ = d-c load current.

The relation of equation (9) is characterized by the fact that the current i in the final operating condition (Fig. 9) has an a-c component which is so large that 1 becomes zero at one instant. Any further increase of this component caused by

discharged in series with the other condenser during the time of non-conduction of its associated diode. The analysis of operation is made according to the method discussed but will not be treated.

Table II

TYPE OF CONDENSER-INPUT CIRCUIT	nojCR	A η R	00	φ°	JL max	id	Upl ïρ
Half-Wave η - 1	0. 5	0	26. 5	153. 5	0. 335	3. 33	1. 69
	1.	0	45. 0	134. 0	0. 384	3. 68	1. 81
	2.	0	63. 4	111. 6	0. 486	4. 61	2. 00
	2. 26	0	66. 15	106. 4	0. 503	4-91	2. 02
	4-	0	75-9	87. 1	0. 623	6. 60	2. 24
	8.	0	82. 9	65. 1	0. 742	9. 86	2. 60
	16.	0	86.4	48. 6	0. 862	13-92	3. 00
	32.	0	88. 2	35-3	0. 930	19-90	3. 51
	64.	0	89. 1	25. 1	0. 996	27. 5 ?	4. 16
	2.	0. 10		121.	0. 434	4-48	1. 9
	2. 26	0.147	50.	123.	0.428	4. 42	1. 88
	4.	0. 05	65. I	99-3	0. 632	5. 28	2. 1
	4.	0. 10	56.	108. 4	0. 537	5. 14	2. 0
Full-Wave η - 2	1.	0	26. 5	142. 5	0.644	3. 47	1. 75
	2.	0	45. 0	121. 0	0. 678	4. 17	1. 90
	4-	0	63-4	92. 6	0. 740	6. 06	2. 17
	4-52	0	66. 15	86. 8	0. 744	6. 55	2. 24
	8.	0	75. 9	67. 0	0. 816	9-30	2. 55
	16.	0	83. 0	49. 0	0. 885	13. 74	3. 00
	32.	0	86.4	35. 6	0. 945	19-70	3. 50
	64.	0	88. 2	25. 4	0. 999	27. 1 ?	4-15
	4.	0.05		104.	0. 671	5-43	2. 05
	4-52	0. 0735	50.	105.	0. 636	5-35	2. 04
	8.	0. 05	56.	90.	0. 710	6. 20	2. 20
	30. 2	0.10	17. 9	100. 6	0.646	5-39	2. 08

the reversed problem to determine the magnitude of the applied voltage necessary to give a certain average voltage output for a given load. The series-resistance value Rg includes the equivalent average resistance rd

of one diode and the power-transformer resistances as reflected into one secondary winding. The characteristics were plotted from accurately measured values as their complete calculation required too much time. The measurements were made cti circuits of negligible inductive reactance. Series resistance values were determined accurately by the method shown in Fig. 2. Table II gives a number of calculated values which show the accuracy of the curves to be approximately 5 per cent or bet-ter.

In compiling the data for the current-ratio characteristics in Fig. 21, it was found that the three rectifier-circuit types could be shown by single family after a "charge factor" η was added to the_ product of the circuit constants ωl and to Rg as shown in Table II. The factor η is unity for the half-wave circuit. For the full-wave circuit, η is 2 because the condenser C is charged twice during one cycle. For the voltage-doubling circuit, η is 1/2 because the two condensers require together twice the charge to deliver the same average current at double voltage. The values in the table indicate that the factor η is actually not a constant. The mean value of the current ratios does, however, not depart more than approximately 5 per cent from the true value, the error being a maximum in the steep portion of the curves and decreasing to zero at both ends. The upper section of Fig. 21 shows the ratio of rms current to average current per diode plate. This family is of special interest in the design of power transformers and for computation of diode plate dissipation.

Fig. 22 shows the rms value of the ripple voltage across R^ in per cent of the average voltage.

The voltage-doubling circuit shown with the other two condenser-input circuits in Fig. 17 maybe regarded in principle as a series connection of two half-wave rectifier circuits. Each con-denser is charged separately during conduction time of one diode, but is

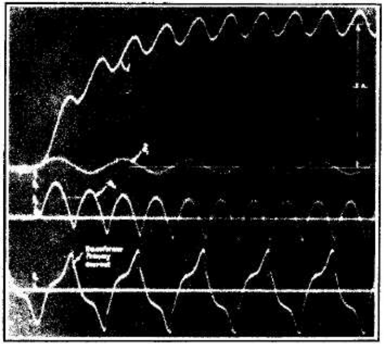

Fig. 10 - Oscillogram taken with circuit values in Fig. 9.

decreasing L still further will, therefore, open the diode circuit and hence interrupt chain-cur-rent operation. The circuit then becomes a type of condenser-input circuit and the output voltage E rises beyond the value given by the equivalent circuit for the chain-current operating range.

b-_Equivalent Circuit _for__ the Chain-^ur-rent_0^erating_Range — Commutation of the cur- rent in the chain-current range occurs at the instant when the supply voltage passes through zero. The voltage energizing the common branch of choke-input rectifier circuits has the form of a_ commutated sine wave for any load value R = Rcr· Since the circuit is never interrupted, currents and voltages may be considered steady-state components in a circuit energized by the equivalent commutated sine voltage as shown in Fig. 11a. The single generator in Fig. 11a may

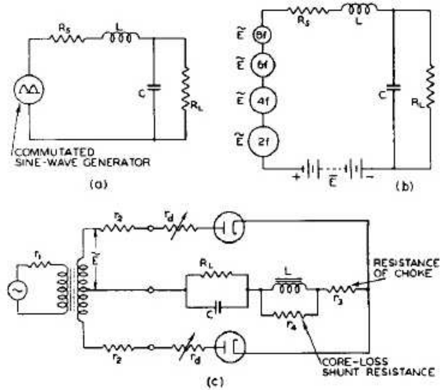
R_S
L
C
R_L
COMMUTATED SINE-WAVE GENERATOR
(a)
R_S
L
Ẽ 8f
Ẽ 6f
Ẽ 4f
Ẽ 2f
C
R_L
+
Ē
−
(b)
r_2
r_d
r_1
Ẽ
R_L
C
L
r_3
r_4
RESISTANCE OF CHOKE
r_2
r_d
CORE-LOSS SHUNT RESISTANCE
(c)

be replaced as shown in Fig. lib by a battery and a series of sine-wave generators having frequencies and amplitudes, as given by the following equation of the commutated sine wave.

$$\tilde{e} = \frac{2\,\tilde{e}_{max}}{\pi}\left[1 - \frac{2\cos 2f}{1\times 3} - \frac{2\cos 4f}{3\times 5} - \frac{2\cos 6f}{5\times 7} - \ldots\ldots\right]$$

The d—c component or battery voltage is the total average voltage E discussed in the preceding sections .
Some useful relations of voltage components are:

RELATION OF APPLIED AC VOLTAGE TO DC OUTPUT VOLTAGEIN VOLTAGE-DOUBLING CIRCUIT
in farads ,rl in ohms

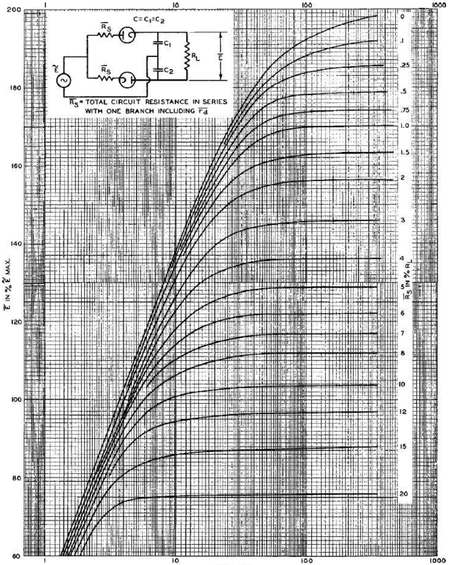

Fig. 20

RELATION OF APPLIED AC VOLTAGE TO DC OUTPUT VOLTAGEIN FULL-WAVE RECTIFIER CIRCUIT

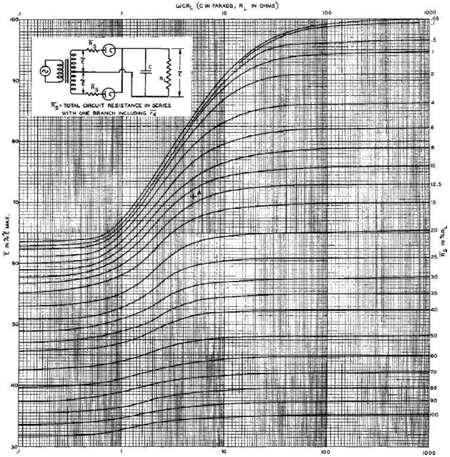

Line voltage induced in one-half of the secondary winding (rms)	$\lvert\tilde{E}\rvert = 1.1\ \bar{E}$
Total average voltage	$\bar{E} = \begin{cases} 0.90\ \lvert\tilde{E}\rvert \\ 0.637\ \tilde{e}_{max} \end{cases}$
Voltage of frequency 2f (rms)	$\lvert\tilde{E}\rvert_{2f} = \begin{cases} 0.424\ \lvert\tilde{E}\rvert \\ 0.471\ \bar{E} \end{cases}$
Voltage of frequency 4f (rms)	$\lvert\tilde{E}\rvert_{4f} = \begin{cases} 0.085\ \lvert\tilde{E}\rvert \\ 0.0945\ \bar{E} \end{cases}$
Total choke voltage (rms)	$\lvert E\rvert_L = \begin{cases} \sqrt{\lvert\tilde{E}\rvert^2 - \bar{E}^2} \\ 0.482\ \bar{E} \end{cases}$

The current components in the common circuit branch are calculated from the above volt-ages divided by the impedance of one branch-circuit at the particular frequency. Because the current is commutated every half-cycle of the line frequency from one to the other branch-circuit, the average current in each diode circuit is one-half of the total average current; and rms values of currents or current components in each branch-circuit are obtained by multiplying the rms current values in the common circuit branch by 1/ /2. The peak current in each diode circuit has the same value as in the common circuit branch.

Average load current $\bar{I} = \dfrac{\bar{E}}{R_s + R_L}$

Average plate current (per diode) $\bar{I}_p = 0.5\ \bar{I}$

Double-frequency current (rms) in common circuit branch $|\tilde{I}|_{2f} = \dfrac{|\tilde{E}|_{2f}}{Z_{(2f)}}$

Total current (rms) in common circuit branch $|I|_L = \sqrt{\bar{I}^2 + |I|^2_{2f}}$

Rms diode current or rms current per transformer winding $|I|_d = \dfrac{|I|_L}{\sqrt{2}}$

Peak diode current $\hat{i}_d = \bar{I} + (|\tilde{I}|_{2f} \times \sqrt{2})$

The total power dissipated in diode and load circuits of the practical secondary circuit shown in Fig. 11c is the sum of the power losses in the circuit resistances. In equation form, it is

Total Power = Series-Resistance Loss
+ Choke-Core Loss
+ D-C Power in Load

The plate dissipation per diode is given by:

$$P_d = 0.5\ |I|^2_L \times |r_d|$$

$$P_d = 0.5\ |I|^2_L \times \frac{\bar{e}_d}{\bar{I}}$$

RELATION OF APPLIED AC VOLTAGE TO DC OUTPUT VOLTAGE IN HALF-WAVE RECTIFIER CIRCUIT. UJCRL
(c IN FARADS , RL IN HMS)

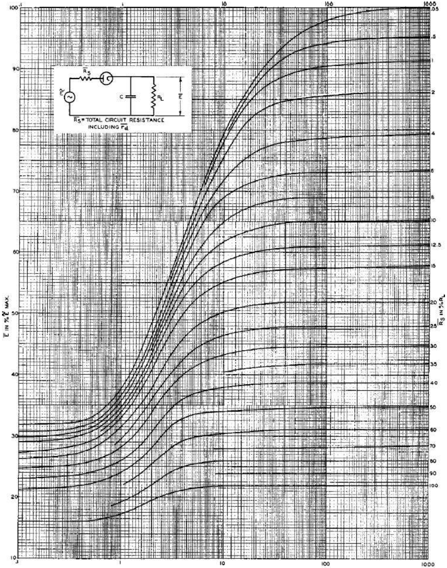

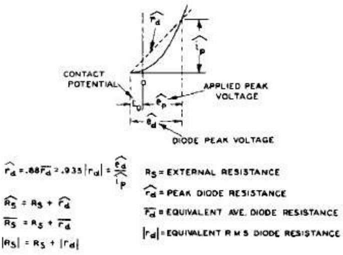
$\widehat{r_d}$
$\widehat{i_p}$
CONTACT POTENTIAL
0
APPLIED PEAK VOLTAGE
E_0
$\widehat{e_p}$
$\widehat{e_d}$
DIODE PEAK VOLTAGE
$\widehat{r_d} = .88\overline{r_d} \approx .935|r_d| = \widehat{e_d} / \widehat{i_p}$
$\widehat{R_S} = R_S + \widehat{r_d}$
$\overline{R_S} = R_S + \overline{r_d}$
$|R_S| = R_S + |r_d|$
R_S = EXTERNAL RESISTANCE
$\widehat{r_d}$ = PEAK DIODE RESISTANCE
$\overline{r_d}$ = EQUIVALENT AVE. DIODE RESISTANCE
$|r_d|$ = EQUIVALENT R M S DIODE RESISTANCE

where ed is the diode voltage taken from the static_diode characteristic at the output-current value I.

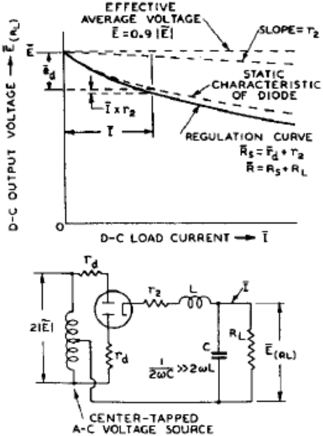
EFFECTIVE AVERAGE VOLTAGE
$\overline{E} = 0.9\,|\widetilde{E}|$
SLOPE= r_2
$\overline{E}$
$\overline{e}_d$
STATIC CHARACTERISTIC OF DIODE
$\overline{I} \times r_2$
$\overline{I}$
REGULATION CURVE
$\overline{R}_S = \overline{r}_d + r_2$
$\overline{R} = R_S + R_L$
D-C OUTPUT VOLTAGE → $\overline{E}_{(R_L)}$
0
D-C LOAD CURRENT → $\overline{I}$
r_d
r_2
L
$\overline{I}$
$2|\widetilde{E}|$
r_d
R_L
C
$\frac{1}{2\omega C} \gg 2\omega L$
$\overline{E}_{(R_L)}$
CENTER-TAPPED A-C VOLTAGE SOURCE

The regulation of choke-input circuits is determined by the_ total series resistance Rg, since the voltage E in the circuit is constant in the useful working range for an energizing-c voltage of constant value._ Thus, the regulation curve has the slope Rg (see Fig. 12),which includes the diode resistance.

The curve is correct for constant voltage e and beyond the critical-current value. In practical circuits, the voltage source e has a certain equivalent resistance, which must be added to r2. The equivalent internal resistance of the rectifier circuit as a d-c supply source is the slope of the regulation curve at the current value under consideration. This value should be used for steady-output conditions only, since the reactance in the load circuit cause transients at the instant of sudden load changes.

2. The Condenser-Input Circuit

In rectifier circuits with shunt-condenser-input loads, the condenser is alternately charged and discharged. In the final state of operation, charge and discharge are balanced. The graphic analysis of such circuits is comparatively simple and readily followed. Formulas for the calculation of specific circuit conditions are easily derived from the constructions.

a._Circuits. _Without_ Series Resistance_ —
The graphic analysis of a half-wave rectifier circuit without series resistance (Rs) is illustrated in Fig. 13· Steady-state voltage e and current is are constructed on the assumption that the diode is shorted. The steady-state condenser voltage e'c coincides with e" because Rg = 0.
The diode timing is as follows: The diode opens the circuit at point 0 when the diode cur-rent becomes zero. The condenser voltage at this instant has the value

$$e_{C(o)} = \tilde{e}_{max} \sin \Theta$$

Since the condenser-discharge circuit consists of C and l, the condenser voltage decays exponentially as shown. At point C it has become equal to the energizing voltage e'. The diode be-comes conducting and closes the circuit. Because there is no potential difference between the steady-state voltages e and ec at this or any instant, the condenser does not receive a transient charge. The current, therefore, rises instantly to the steady-state value as shown on the i s curve and then decreases until it is zero at point 0.

(sine wave) for half-wave, full—wave, and voltage-doubling circuits. They permit the solution of

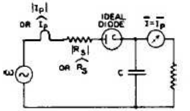

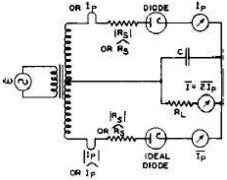

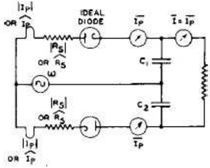

these transients analytically but the method is too involved to discuss in this lecture. The transients e-t and 1¾ in Fig. 15a prevent voltage and current from following the steady-state wave-forms. When Rg is large, they do not decay to zero within one cycle and, therefore, require additional steps in the graphic solution. The oscillogram shown In Fig. 16 checks the graphical construction of Fig. 15·

Fig. 16 — Oscillogram verifying graphical construction of Fig. 15·

c Generalized _operation__ Characteristics^{Steady-State^ Operation) — It has been shown that the conduction angle φ is a function of the circuit constants in condenser-input circuits. The section of the energizing voltage e utilized during conduction time has, therefore, no fixed value as in choke-input circuits where φ = 180°and where the voltage e during φ is a half-sine wave. It is, therefore, not possible to derive general equivalent circuit for condenser-input circuits which contains a voltage source of fixed wave shape and magnitude.

Steady-state conditions as well as transients are controlled by the circuit constants, which are contained in the product wl, The angle φ depends on the relative magnitudes of R^ and Rg and is, therefore, described in general is also the ratio s /^l^3^1011"1· General curve families may thus _be evaluated which show the dependent variables E, i, and I in terms of ratio versus the independent variable coCR^ various parameter values Rg/R^. The series resistance Rg includes the equivalent diode resistance which is evaluated by means of equation (6), because the current wave is periodic in the final operating state. The reasoning leading to equation(6) is not applicable to a single transient, as obtained for starting conditions of rectifier circuits.

Generalized characteristics have been evaluated for the three types of circuits shown in Fig. 17. The characteristics in Figs. 18, 19, and20 show the average voltage E across the load resistance Rl as a function of co CRt and Rg in per cent of the applied a-c peak voltage

The timing of the full-wave circuit in Fig.1Λ is quite similar. The time for the condenser discharge through l is reduced since e0 meets the positive half-cycle 6¾ and thus closes the circuit through D2· Point C in Fig. 14. is located at a higher value of e than in Fig. 13· The conduction angle φ is consequently reduced al-though C, R l, and Θ have the same values in both circuits. The average current in the full-wave circuit is, therefore, smaller than twice that of the half-wave circuit.

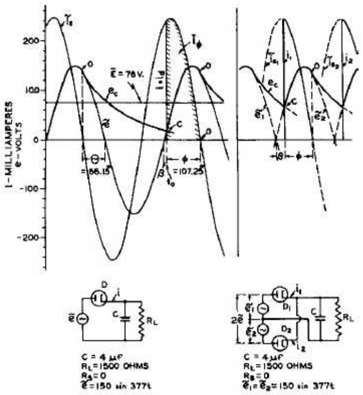

Fig. 13 (left) - Half-wave, condenser-input circuit without series resistance.

Fig. 14 (right) - Full-wave, condenser-input circuit without series resistance.

b. Circuits With_ Series _ Resistance —

In circuits with series resistance, the steady-state condenser voltage ec does not coincide with the supply voltage e, as illustrated in Figs. 15a and15b. Phase displacement and magnitudes of cur-rent and voltage under steady-state conditions are required for analysis of the circuit and are computed in the conventional manner. The parallel circuit C II Rj, is converted into an equivalent series circuit to determine the angles Θ and©' lay which Γ8 is leading e and ec, respectively. The steady—state condenser voltage Za in the parallel circuit equals the voltage across the equivalent circuit as shown by the vector diagram in Fig. 15b.

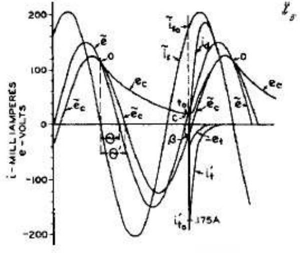

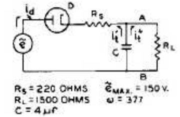
i_d
D
R_S
A
i'_t
i_t
C
R_L
B
ẽ
R_S = 220 OHMS
R_L = 1500 OHMS
C = 4 μf
ẽ_MAX. = 150 V.
ω = 377

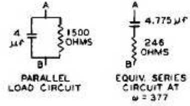
A
4 μf
1500 OHMS
B
PARALLEL LOAD CIRCUIT
A
4.775 μf
246 OHMS
B
EQUIV. SERIES CIRCUIT AT ω = 377

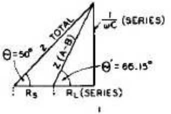
Θ=50°
Z TOTAL
Z(A-B)
1/ωC (SERIES)
Θ' = 66.15°
R_S
R_L (SERIES)
I

The diode opens the circuit at the instant id = 0. For circuit constants as in Fig. 15, the diode current id substantially equals Γ3 at the time of circuit interruption because the transient component i^' of the current has decayed to a negligible value. Point 0 is thus easily located. In circuits with large series resistance, however, id = 0 does not coincide with ïβ=0 due to slow decay of the transient. In both cases the condenser voltage each (0) equals the voltage (o) at the time 0, because id = 0 and consequently there is no potential difference on Rg. The condenser voltage decays exponentially ai Rj, from its Initial value at 0, as discussed for circuits with Rg = 0, and meets the supply voltage e" against point C. At this instant (tQ), the diode closes the circuit. Current and voltage, however, do not rise to their steady-state values as in circuits with Rg = 0, because the steady-state volt-age e"c(o) differs from the line voltage e(0) by the amount e"c(0) = ^s(o)x%· This demanded potential change on the capacitance causes transients. It is possible to determine the value of

Inventing the tech
the world loves

Learn more about 5G NR

Qualcomm

Advertisement Advertisement

JC Bose patents radio wave detector, September 30, 1901

Jessica MacNeil -September 30, 2017

1 Comments

Share G+ Tweet Like 18

James Clerk Maxwell's **equations** predicted the existence of electromagnetic waves of diverse wavelengths in the 1860s, and in 1888 Oliver Lodge and Heinrich Hertz verified it, along wires and in free space. After reading about their experiments, JC Bose, a professor of physics at Presidency College in Calcutta, India, reduced the waves to about 5mm to study them.

He developed the use of galena (lead sulfide) crystals contacted by a metal point for detecting millimeter electromagnetic waves, and filed patent 755840, for a "**Detector for electrical disturbances**" in 1901. It was the first device to use a semiconductor junction to detect radio waves.

Bose created a portable device to study the optical properties of short waves, which incorporated the earliest waveguide and horn antenna. It would be more than 30 years before more research and analysis was done with the now widely-used horn antennas. His use of dielectric lenses, polarizers, prisms, and semiconductors at frequencies as high as 60 GHz, were also well ahead of his time.

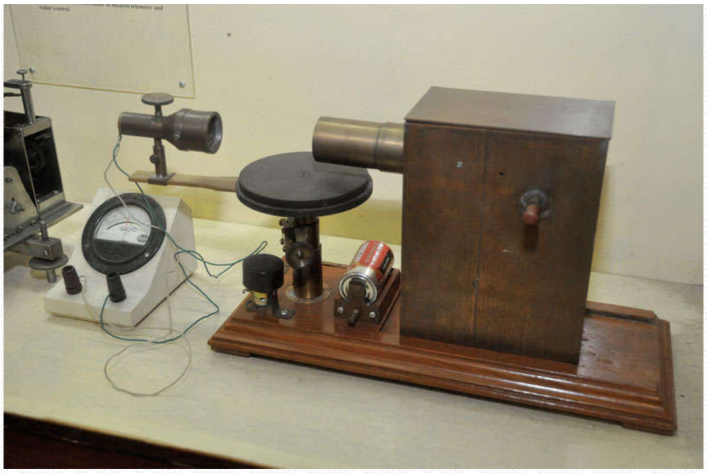

Bose's 60-GHz microwave apparatus at the Bose Institute in Kolkata, India.
Source: Wikipedia user Gangulybiswarup

In a 1895 demonstration, he used electromagnetic waves to ignite gunpowder and ring a bell from a distance, reported to be nearly a mile. In 1897 Bose presented his microwave experiments at the Royal Institution in London (see the photo and diagram below).

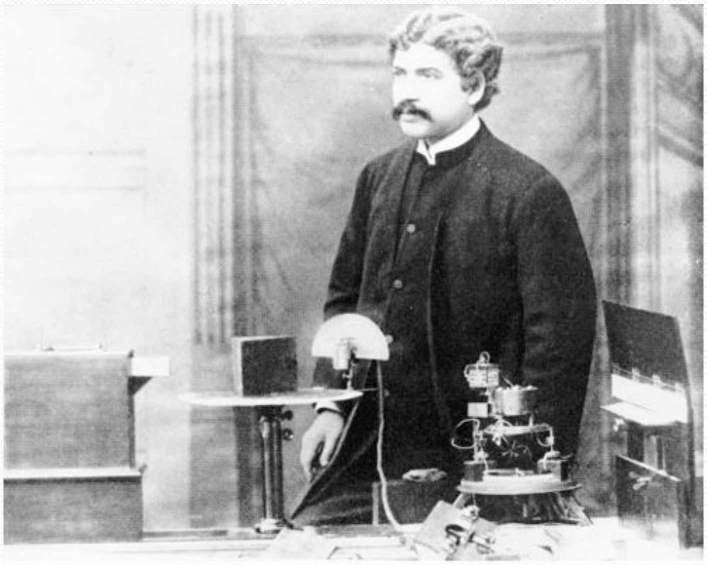

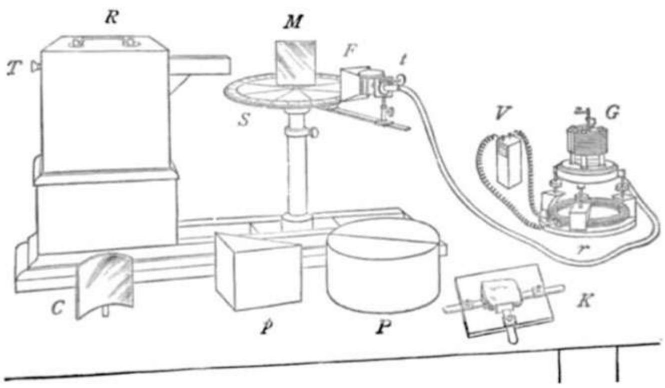

R, radiator ; S, spectrometer-circle ; M, plane mirror ; C, cylindrical mirror ; p, totally reflecting prism ; P, semi-cylinders ; K, crystal-holder ; F, collecting funnel attached to the spiral spring receiver ; t, tangent screw, by which the receiver is rotated ; V, voltaic cell ; r, circular rheostat ; G, galvanometer.

He founded the Bose Institute in 1917 in Calcutta where some of his original instruments and devices are displayed. Bose spent much of his later career studying the effects of electromagnetic radiation on plants.

Much of Bose's component technology would eventually be used to develop microwave radio transmission. For his pioneering work in quasi-optic millimeter wave research, IEEE has called him the father of radio science. According to a 1997 *Microwave Symposium Digest* publication, "He developed an elegant millimeter wave spark transmitter, self recovering coherer detector, wire grid polarizer, cylindrical diffraction grating, dielectric lens and prism, rectangular waveguide, horn antenna and microwave absorber, for the studies of reflection, refraction, absorption and polarization of millimeter waves and its application to wireless remote control for firing a gun."

References

1. **The work of Jagadis Chandra Bose: 100 years of mm-wave research**
2. **Semiconductor Rectifiers Patented as "Cat's Whisker" Detectors**, Computer History Museum

Also see:

- **Popov demonstrates radio receiver, May 7, 1895**
- **Tesla gives 1st public demonstration of radio, March 1, 1893**
- **Teardown: Maxwell's equations**

---

*For more moments in tech history, see **this blog**. EDN strives to be historically accurate with these postings. Should you see an error, please **notify us**.*

***Editor's note**: This article was originally posted on September 30, 2013 and edited on September 30, 2017.*

Share G+ Tweet Like 18

SPONSORED CONTENT

fido ALLIANCE

FIDO isn't just the name of your neighbor's dog anymore. Join Infineon for this live webinar as they discuss how hardware-based multi-factor authentication using Fast IDentity Online (FIDO) improves protection from hackers for your online accounts. Register Now!

Sponsored By Infineon

1 Comments

Write a Comment

Submit Comment

To comment
please Log In

Jiri Polivka

J.C.Bose was really a wonderful genius! He was ahead of time with most of his ideas.
The sad part is that he worked in India, then a British colony. This fact caused that his wonderful work was "out of sight" in Europe nd the U.S. Then most of his ideas if ever found were stolen and copied. Others gained money while he used his modest wealth to establish the first college in India where math and physics were taught.

Oct 1, 2015 2:49 AM EDT

👍1 | 👎0

Reply

EDN MOMENT
Apple III introduced, May 19, 1980

Click here to be ahead in applications, webinars and videos for 5G.
5G
ROHDE&SCHWARZ

MOST RECENT

Apple III introduced, May 19, 1980

Apollo 10 launches, May 18, 1969

Gears are discovered on the Antikythera mechanism, May 17, 1902

Sketchpad creator Ivan Sutherland is born, May 16, 1938

Britain drops its first H-bomb, May 15, 1957

Skylab launches into space, May 14, 1973

Edison's 1st test of electric railway, May 13, 1880

NASA regroups after Challenger disaster, May 12, 1986

Coincident-current magnetic storage is patented, May 11, 1951

Transcontinental railroad is completed, May 10, 1869

Inventing the tech the world loves

Learn more about 5G NR

Qualcomm

Most Popular Most Commented

That 60W-equivalent LED: What you don’t know, and what no one will tell you…

The Picard Loop

6 famous people you may not know are engineers

10 tips for a successful engineering resume

Electronic products from hell

Why does this year have an extra second?

Da Vinci unsuccessfully tests a flying machine, January 3, 1496

The 5 greatest engineers of all time

Talking about Design Ideas

ICs of the 1930s & 1940s

FEATURED RESOURCES

>> Know the Difference Between an Inverter, Converter, Transformer and Rectifier

>>LTC3246 – Low EMI, 500mA Buck-Boost Charge Pump

ASPENCORE DOUBLE SUMMITS 全球双峰会
Help shape the connected future with tech industry leaders and supply chain stakeholders.
2018 ASPENCORE Double Summits
Shenzhen, China, Nov. 8–9 | Two events. Two Days.
JOIN THE GLOBAL CONVERSATION.

Subscribe to RSS: or

DESIGN CENTERS

5G
Analog
Automotive
Components & Packaging
Consumer
DIY
IC Design
LEDs
Medical
PCB
Power Management
Sensors
Systems Design
Test & Measurement

MORE EDN

Blogs
Design Ideas
Tech Papers
Courses
Webinars
EDN TV
Events
About Us

ASPENCORE NETWORK

EBN | EDN | EE Times | EEWeb | Electronic Products | Electronics-Tutorials | Embedded | Planet Analog | ElectroSchematics | Power Electronics News | TechOnline | Datasheets.com | Embedded Control Europe | Embedded Know How | Embedded News | IOT Design Zone | Motor Control Design | Electronics Know How | IoT Times |

GLOBAL NETWORK

EE Times Asia | EE Times China | EE Times India | EE Times Japan | EE Times Taiwan | EDN Asia | EDN China | EDN Taiwan | EDN Japan | ESM China |

Working With Us: About | Contact Us | Media Kits

Terms of Service | Privacy Statement | Copyright ©2018 AspenCore All Rights Reserved

Kit Instruction Manual

Traveler II
&
Traveler II SE

Traveler II
NT
D
ES
Antenna
Coupling
Detector
AM Crystal Radio
Crystal Radio Supply

!

Crystal Radio Supply

http://www.crystalradiosupply.com

Table of Contents

Introduction .... Page 3-4

Care of the Front Panel.. Page 5

Assembly Instructions .... Page 6-13

Basic Tuning Procedures.. Page 14

Typical Antenna Installation.. Page 14

Schematic .... Page 15

Crystal Radios, Kits, Parts, and Supplies

Crystal Radio Supply

http://www.crystalradiosupply.com

Introducing the Traveler II AM Band Crystal Radio

Traveler II

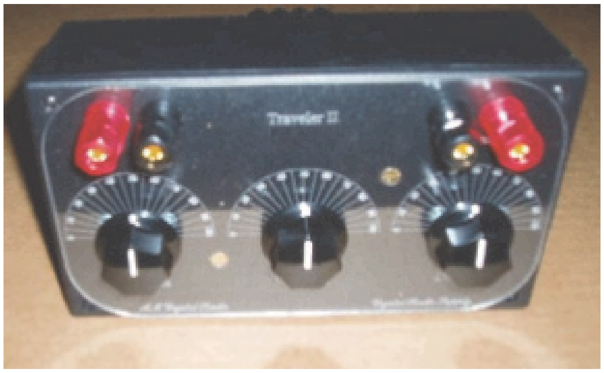

Traveler II
Special Edition

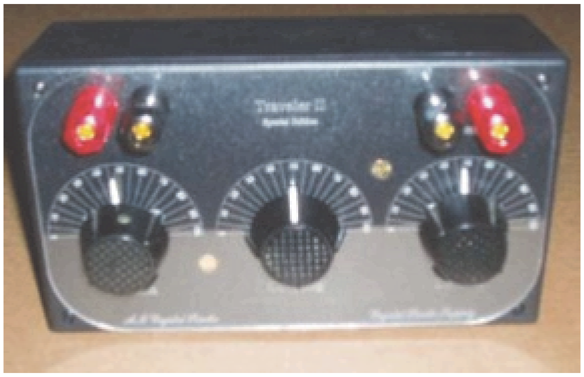

The Traveler II and the Traveler II SE are electronically the same radio. The only mechanical differences between the two sets are:

1. The Special Edition model uses an 8:1 planetary drive capacitor for the detector tuner, whereas the Traveler II uses a direct drive unit.

2. The Special Edition model uses a 6:1 dual-gang planetary drive capacitor for the antenna tuner, whereas the Traveler II uses a direct drive dual-gang unit

3. The Special Edition unit uses dial pointers and more expensive new old stock (NOS) knobs to take advantage of the additional band spread of the planetary drive capacitors.

Both Sets have the following features:

1. For portability, they are built in a tough ABS box that measures approximately 7-7/8" x 4-1/4" x 2-1/2". An assembled unit weighs approximately 1 pound and 10 ounces.

2. A cast acrylic dial plate that is reverse engraved (engraved on the back side of the dial plate so that there is no ink to wear off.

3. The high quality plated binding posts will accept many ways of making connections, including holes that will accept pins from the old-style magnetic headphones.

4. The 1/4" phone jack automatically switches the audio output from the binding post circuit, containing a 47 K resistor, to the phone jack when the headphones are plugged in. The binding posts work great for very high impedance headphones, such as the included ceramic earphone, while the phone jack can be used for lower impedance headphones or even a speaker – for those extra strong signals.

5. These sets use high Q ferrite Litz coils that are wound with 175/46 Litz wire and have a quality factor of 620 @ 790 KHz

6. The Traveler uses the very popular Tuggle front-end circuit that is valued for its fine selectivity.

7. Variable inductive coupling is provided, a feature seldom found on a receiver of this size.

8. Crystal Radio Supply's exclusive 2:1 reduction gear box provides a full 180 degrees of variable coupling for a 90 degree change in coil orientation. It is made from high quality Delrin gears and a marine-grade HDPE housing with all-brass hardware and a spring-temper bronze brake. This provides precise tuning as well as resetting of your inductive coupling when you return to a station.

Care of the Front Panel

The Front Panel on your Traveler II Crystal Radio Set is made from High Quality "Cast Acrylic". This grade of Acrylic is superior over other grades of acrylic and will still look beautiful after many years of service if the following simple precautions are observed:

1. Do not use paper towels to clean your front panel as this may cause scratches.

2. Use a soft cotton fabric such as that from a cotton T-shirt.

3. To remove fingerprints and smudges, clean your panel with isopropyl alcohol (rubbing alcohol) that can be found at most pharmacies.

4. Do not use detergents.

5. Do not touch the front of your panel; only handle it by the edges.

Assembly Instructions for the Traveler II Crystal Radio

Capacitor identification and preparation

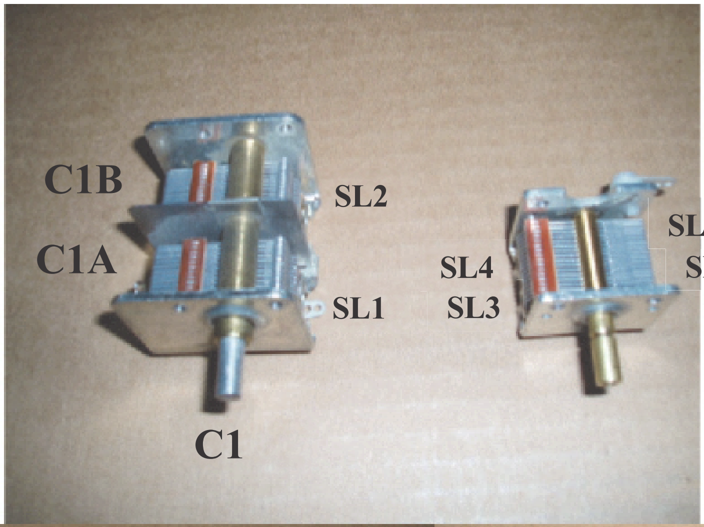

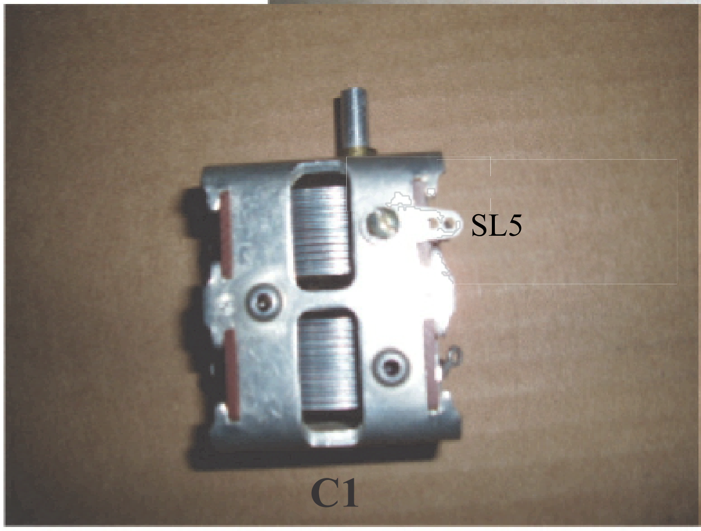

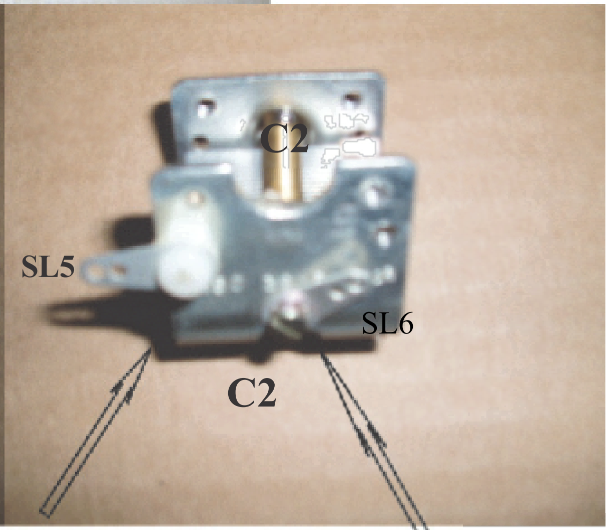

Step 1.
On the bottom of C1, install a #6 solder lug using a #6 fillister head screw. Fillister head screws are the ones with the thick round heads.

Step 2.
On the back of C2, mount a #6 solder lug on top of a 1/4" long nylon spacer using a 3/8" long #6 nylon screw.

Step 3.
In the bottom center hole on the back of C2, install a #6 solder lug using a #6 fillister head screw. Bend the solder lug up slightly as shown.

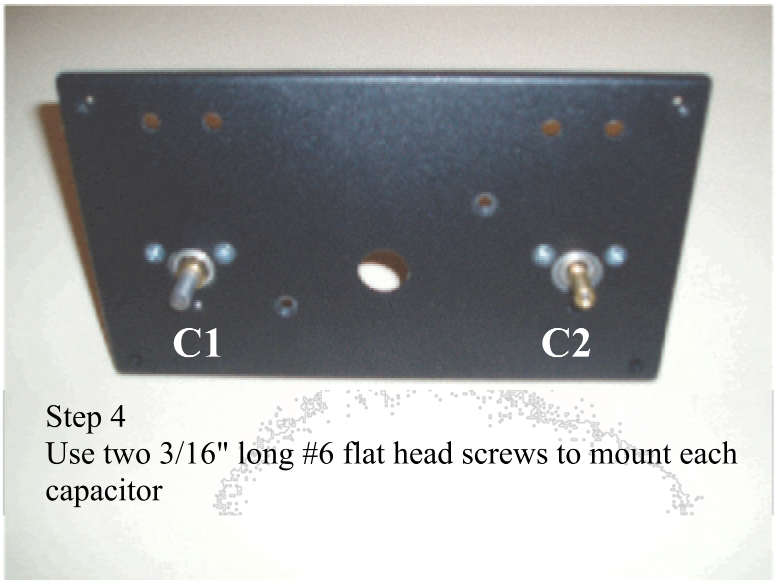

Step 4
Use two 3/16" long #6 flat head screws to mount each capacitor

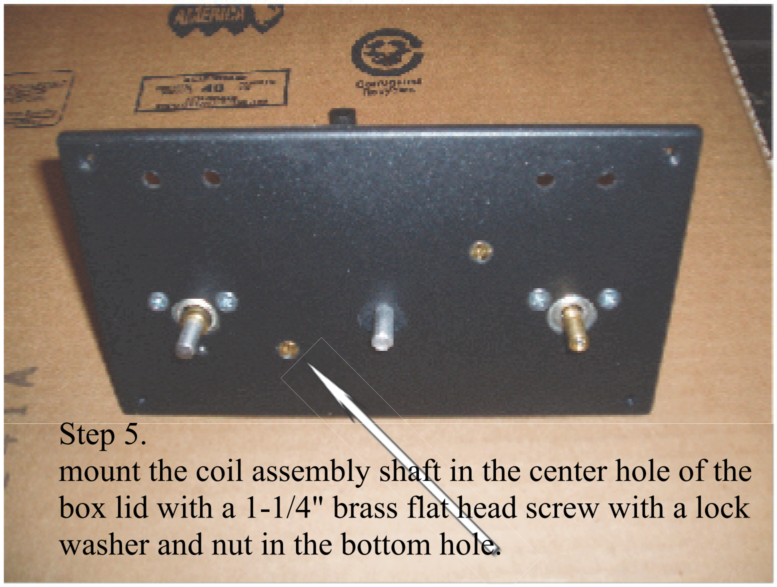

Step 5.
mount the coil assembly shaft in the center hole of the box lid with a 1-1/4" brass flat head screw with a lock washer and nut in the bottom hole.

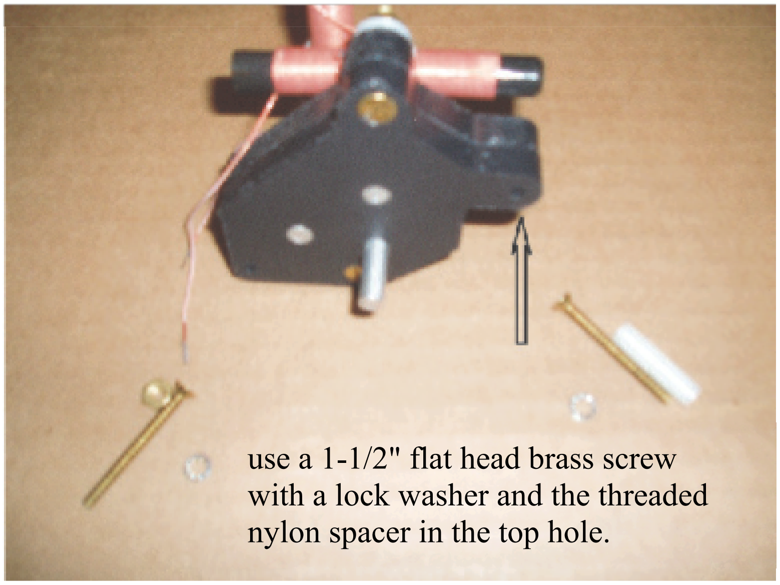

use a 1-1/2" flat head brass screw with a lock washer and the threaded nylon spacer in the top hole.

In the steps below, when handling the dial plate do not touch the face. Only handle by the edges

Set the lid assembly on top of the box while you perform the next step.

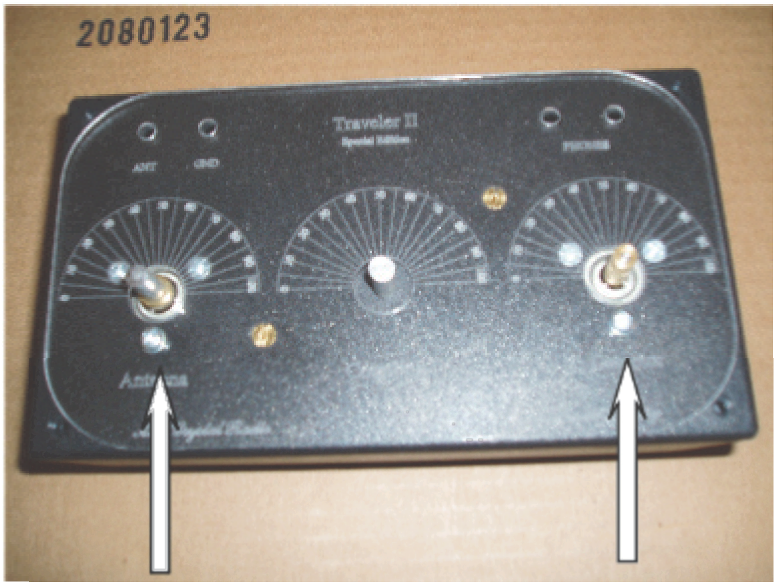

Step 6.
Place the dial plate on the box lid as shown and start two 1/4" #6 round head screws in the bottom capacitor holes. These screws will secure the bottom of your dial plate to the box lid. Leave these screws loose for now; do not tighten.

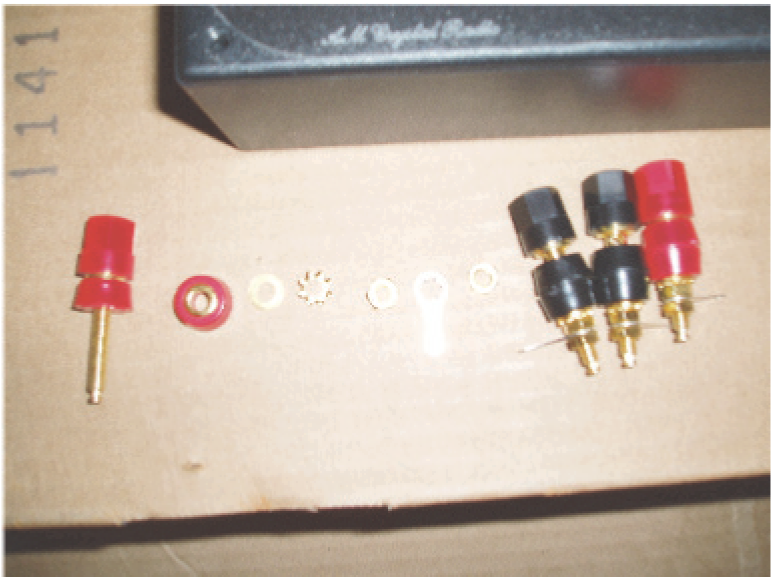

Step 7.
Install the binding posts as shown on page 1. Observe the dissembled binding posts above. The large section goes on the top side of the lid and all the other parts are installed on the bottom of the lid in the order shown from left to right. Do not install the last two parts on each binding post until after the next step. (The solder lug and one of the nuts.)

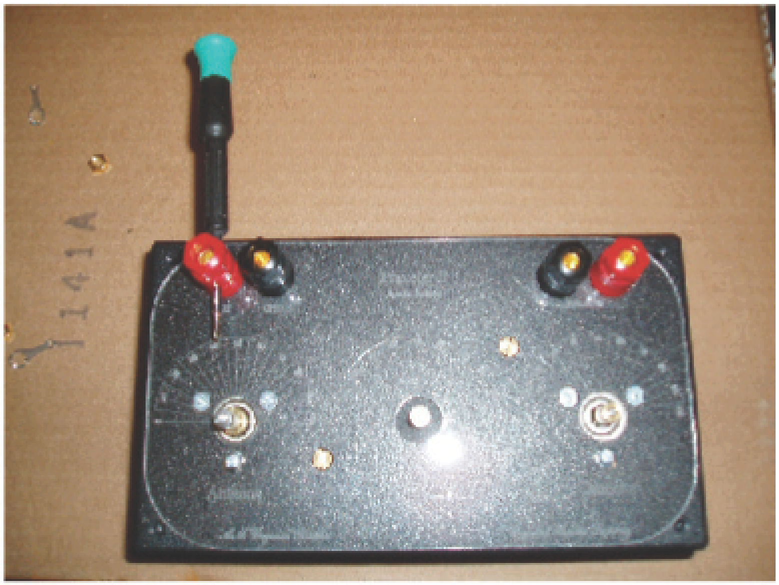

Step 8.
As shown above insert a small screwdriver thru the holes in the binding posts and hold them in the above vertical alignment while you tighten the nut on the bottom of each binding post.

You can also tighten the bottom capacitor screws at this time. Be careful and do not over-tighten as you may damage the dial plate.

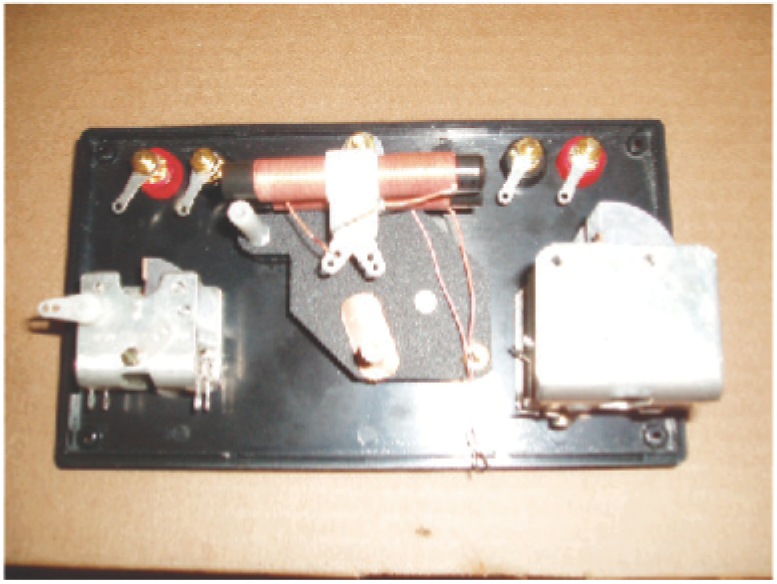

Step 9.
Now install a #10 solder lug and brass nut on each binding post and tighten them down with the solder lugs pointed in the directions shown above.

Installation of knobs and pointers.

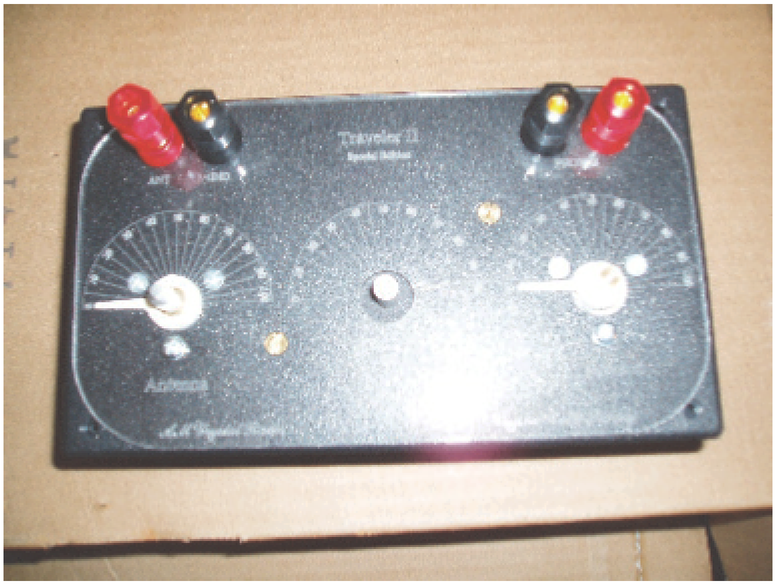

Step 10.
Make sure the two coils are parallel to each other then install the "Coupling" knob with the line pointed at the “100” mark on the dial.

Step 11.
Fully close the tuning capacitors C1 and C2 and for the Traveler II SE, install the dial pointers as shown, pointed at the zero mark on the dial plate. (If you have the Traveler II, instead of the dial pointers, install the knobs with the white line aligned to the zero marks.)

Do not over tighten the dial pointers. Only tighten them enough to keep them from slipping.

Once the dial pointers are installed, Traveler II SE owners can install their tuning knobs. (These knobs have no white mark.)

The knobs and binding post will help protect the dial panel during the remainder of the assembly process. However, it is a good idea to lay the assembly on a soft cloth such as a cotton T-shirt.

If you do use a cloth under the assembly, be careful and do not let it catch fire from soldering.

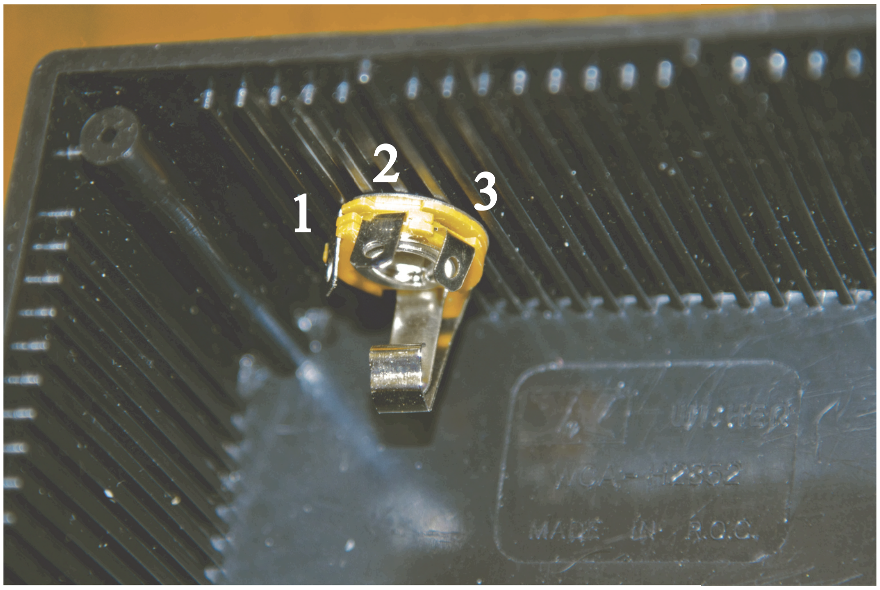

Step 12.

Install the phone jack in the ABS box as shown with the long part of the phone jack on the bottom. Secure with the hardware provided.

Step 13.

Prepare two lengths of #24 gauge TEST LEAD wire by cutting each to the length of 3-1/4" and by removing 1/4" of insulation from each end.

Prepare the following lengths of #26 gauge solid-strand hookup wire by cutting to the desired length and by removing 1/4" of insulation from each end.

2 wires	4" long
1 wire	4-1/2" long
2 wires	6" long
1 wire	6-1/2" long

This end of the coils have lead #2

This end of the coils have lead #1

The top coil is L2
The bottom coil is L1

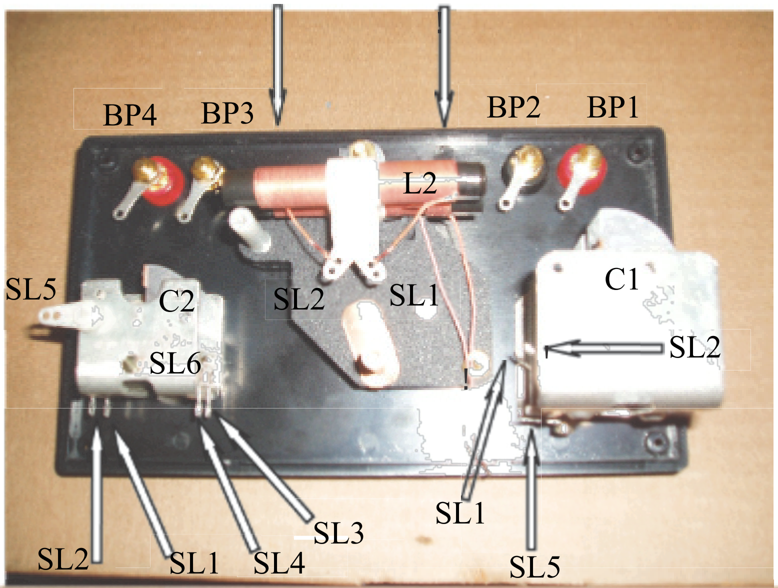

Refer to page 6 for capacitors solder lug locations

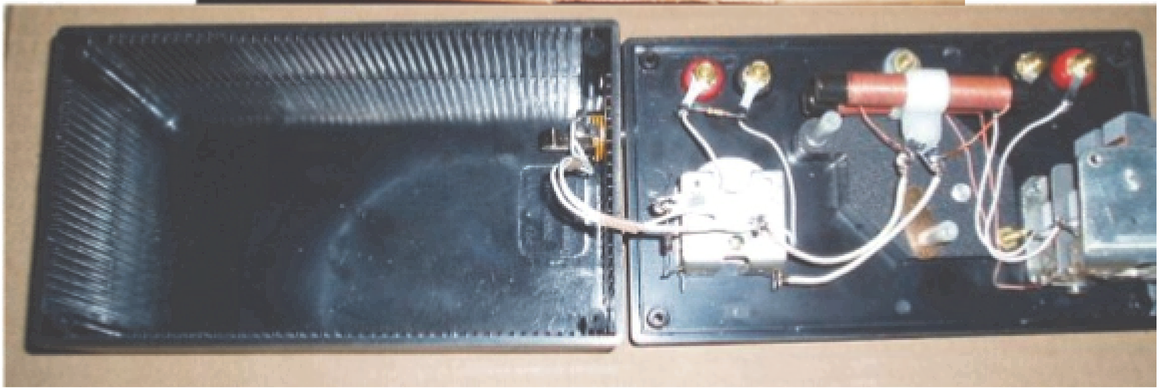

Step 14. Check off when each number is finished.

Do not bend the soldered ends of the wires of coils L1 and L2, as they are brittle and may break. Just place them in the solder lug holes and allow the solder to hold them in place.

1. Make sure lead #1 of coil L2 is attached to the top hole of SL1 that is mounted on the coil tuning shaft.

2. Make sure lead #2 of coil L2 is attached to the top hole of SL2 that is mounted on the coil tuning shaft.

3. Connect one end of a 3-1/4" long Test Lead Wire to SL4 of capacitor C2 and the other end to SL1 on the coil tuning shaft. Solder both ends including coil lead #1 from coil L2.

4. Connect one end of a 3-1/4" Test Lead wire to SL6 on the back of capacitor C2. Do not solder. Connect the other end of this wire to SL2 that is mounted on the coil tuning shaft and solder this connection along with the coil lead #2 from coil L2.

5. Connect coil lead #1 from coil L1 to solder lug #5 of capacitor C1 and solder.

6. Connect coil lead #2 from coil L1 to solder lug SL2 on capacitor C1, but do not solder at this time.

7. Connect a 4" solid strand hookup wire to SL1 of capacitor C1 and connect the other end to the solder lug on binding post #2 and solder and solder both ends.

8. Connect one end of a 4-1/2" long hookup wire to SL2 of C1 and solder two wires. Connect the other end to the solder Lug on binding #1 and solder.

9. Connect one end of a 4" hookup wire to SL6 of C2 and the other end to the solder lug of binding post #3. Do not solder at this time.

10. Connect one end of the 47 K resistor to the solder lug on binding post #3 and solder. Connect other end to the solder lug on binding post #4. Do not solder.

11. Connect one end of a 6-1/2" hookup wire to the solder lug of binding post #4 and solder two wires. Connect the end to solder lug #3 of the phone jack and solder.

12. Connect one end of a 6" hookup wire to solder lug #6 of C2 and solder three wires. Connect the other end to solder lug #1 on the phone jack and solder.

13. Connect one end of a 6" hookup wire to solder lug #2 on the phone jack and solder. Connect the other end to the outside hole of SL5 on C2. Do not solder.

14. Connect the banded end of the 1N34A diode to the bent up solder lug SL2 of C2 and the other end to the inside hole in SL5 and solder both ends.

Congratulations! You have just finished wiring up your crystal radio. Just mount the lid on the box with the four black screws and you are done and ready to fire it up. You will be amazed at its accurate tuning of the full broadcast band and its outstanding ability to separate adjacent stations using various degrees of coupling and fine-tuning of the antenna and detector capacitors.

Basic Tuning Procedures

!

To tune across the band, set the coupling at about 50 and set both capacitors at zero.

Slowly rotate both capacitors simultaneously keeping both pointers at approximately the same place on the dial. When a station is found, adjust both the capacitors and the inductive coupling for best reception.

The less coupling, the better the selectivity. Just take your time and you will master the tuning quickly and will be amazed at the stations that you can dig out of nowhere.

The Traveler II is a quality hi-performance crystal radio. However, without a good outdoor antenna and ground system you will never realize it's full potential. The ceramic earphone that came with the Traveler II is very sensitive. However they are cheap little things and do not last very long. So if you intend to use them, spares are recommended.

It is recommended that you purchase a pair of high-impedance headphones so that you can use both ears. A unit with an impedance of 2000 ohms or more is recommended. Normally headphones of the magnetic type can be purchased used or NOS.

Typical Antenna Installation

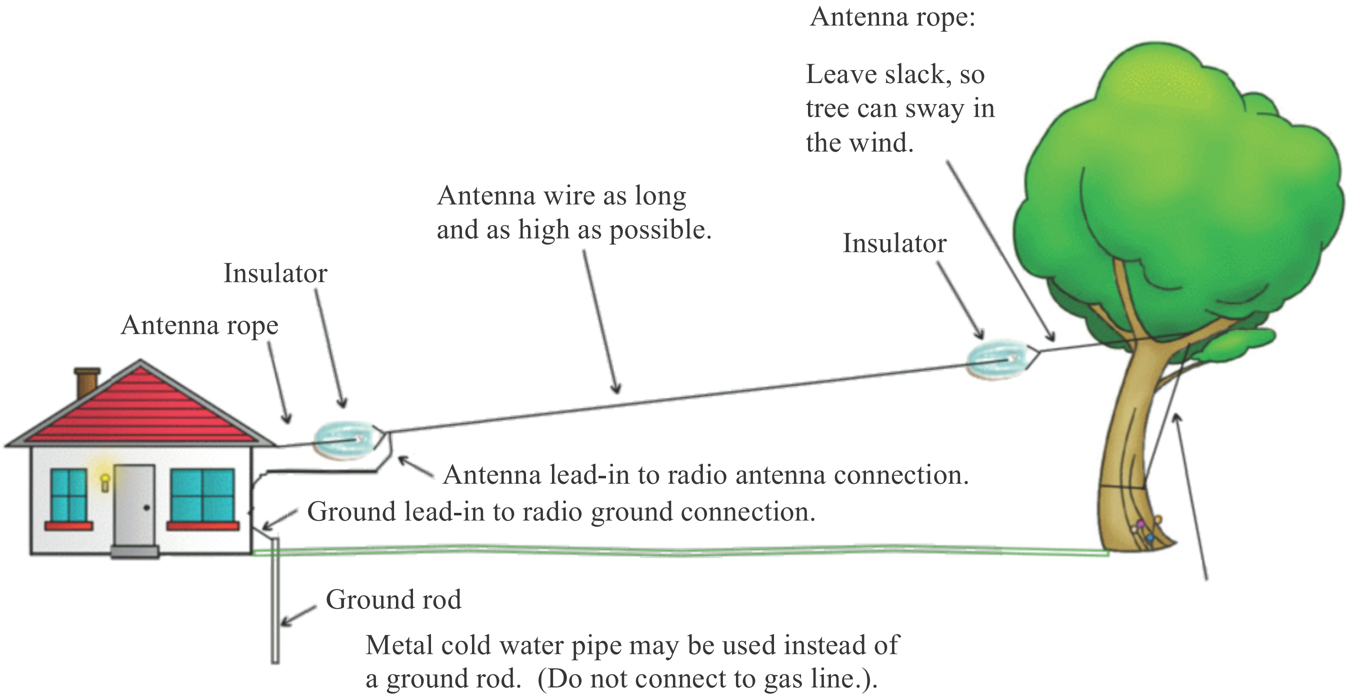
Antenna rope:
Leave slack, so tree can sway in the wind.
Antenna wire as long and as high as possible.
Insulator
Insulator
Antenna rope
Antenna lead-in to radio antenna connection.
Ground lead-in to radio ground connection.
Ground rod
Metal cold water pipe may be used instead of a ground rod. (Do not connect to gas line.).

For safety purposes, a lightning arrestor must be installed in the lead-in wire. If not, remove the lead-in wire from the house when not in use or at least during thunderstorms.

To increase your antenna's length, you can extend it to additional insulated supports – it need not run in a straight line. The length of your antenna greatly affects the performance of your set. A 100 ft or more antenna is considered very good. However, everyone does not have room for a long antenna. In most cases, you just put up the best antenna you can and that is the performance that you get. If you can only get up a short antenna, you can help match it to the radio by adding a coil of wire to the lead-in wire.

!

Traveler II & Traveler II SE Schematic

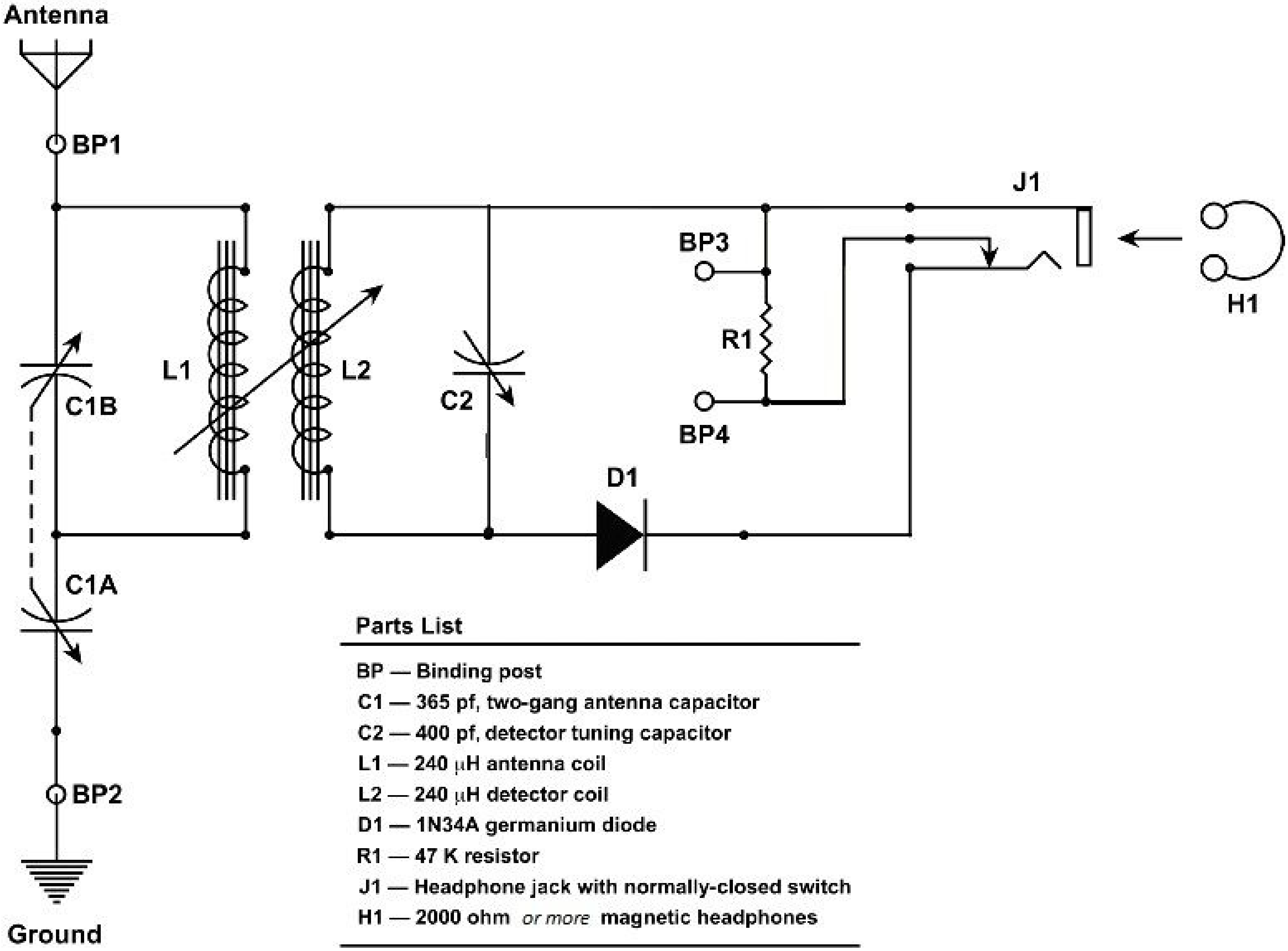

Parts List

BP — Binding post
C1 — 365 pf, two-gang antenna capacitor
C2 — 400 pf, detector tuning capacitor
L1 — 240 μH antenna coil
L2 — 240 μH detector coil
D1 — 1N34A germanium diode
R1 — 47 K resistor
J1 — Headphone jack with normally-closed switch
H1 — 2000 ohm *or more* magnetic headphones

Hi-impedance headphones, such as the included ceramic earphone, can be connected across BP3 and BP4 where R1 provides better sound quality. This resistor can be omitted or altered in value to optimize performance with other headphones.

A note before getting into the crystal radio story.

After I experimented with the Heathkit CR1 crystal radio and built the radio described below, I got out my old regenerative radio I made in 1958 and was once again astounded by how much better it performed than even the CR1. I have now come to the conclusion that anybody who wants to build an ultra-simple radio really should consider my Armstrong "Crystal" Radio first.

Recently added to my website

If you are looking to build a little radio that is as simple and as easy to build as a basic crystal radio, but performs many, many times better than even the best and most complex crystal radio, I highly recommend building my Armstrong "Crystal" Radio. There is a link to the Armstrong Radio project at the very bottom of the following crystal radio story. Please scroll down for the link, but I invite you to look over the pictures and headings to see if there isn't something that might interest you.

If you are determined to build a tiny crystal radio with no amplification or are just interested in reading about my crystal radio and how I built it for its entertainment and especially for its educational value, please read on.

A miniature, but selective and high performance
CRYSTAL RADIO KIT

The description of a project intended for advanced science students and hobbyists and containing personal and controversial opinions and thoughts that you can skip

by John Fuhring

My tiny, but selective and sensitive crystal radio.

Introduction

Not long ago I rebuilt and put into operation a Heathkit CR1 crystal radio. I hadn't built a crystal radio since I put together my first radio when I was 10 years old, nor had I used a crystal set since I was 13. During the period of the mid to late 1950s, I owned at least three crystal radios of varying performance. With the Heathkit, I now had a crystal radio that performed far better than any of the radios that I remember using as a kid and once again I was struck by the magic of building and operating such an utterly simple and basic radio receiver. Suddenly it occurred to me that it might be fun to try my hand at designing and building my own crystal set from scratch. Besides that, a former grammar school teacher I'm friends with suggested to me that school aged kids would have a lot of fun learning about radio theory if they could build a working crystal radio and he asked me to come up with such a radio.. Of course, this strongly reminded me of how, in 1955, my aunt gave me crystal radio kit to build and how that experience lead to my lifelong interest in radio and my career in electronics. The memories flooded back and awoke in me that desire I think we all have and that is to pass along to others those things that have helped us.

Well, I have given a lot of thought to crystal radios lately and I have concluded that the ultra-simple, very poorly performing crystal radios that most teachers have their classes construct are just not worth building. Rather than try to force kids who are just not interested in how their technological world actually works -- and after all, most of us only want to use technology, not design or even understand it -- I think that building a crystal set should be limited to only those students who show a genuine interest in this sort of thing. In my opinion, the only students who should build something like this are the kids I call a school's "the scientific elite" (but who others would call "the nerds"). I think that building a crystal radio should be done as an individual science project, but not as a class project. However, and this is most important, any radio anybody builds should be worth building and should at least approach the performance of the Heathkit CR1.

It soon became my goal to design a low cost, but good performing little radio that would not only be easy and practical to build, but would be something a young person would be proud of and want to keep and use. Of course, this applies to adults too, even old geezers like me. So, the following paragraphs will describe the little radio I have come up with. Please be aware that because the little radio is very compact, it requires some very careful, precision soldering using a dangerously hot soldering iron. Additionally, winding 7 feet of wire on each of two toroid forms is tedious and requires manual dexterity and skill. Because of these difficulties, I am very much of the opinion that this little radio is not a project for very young (under 12) kids. You must use your own judgment regarding whether or not this project is practical for you or your student(s) to build.

Design considerations for a simple but good performing crystal radio

Before I begin the technical phase of this article, please note that I performed several experiments with different kinds of wire and different detectors before I finalized the design that I will be presenting. What I'm presenting here is how and with what materials and components I think a builder should use for optimum performance.

Here then are my design criteria:

Any crystal radio worth building should have good selectivity. That means that you should be able to hear a station without the audio of an adjacent station in there too. Of course, the selectivity of any crystal radio is such that, if you listen closely, you will always hear a powerful nearby station in with your station, but that unwanted signal should be very faint.

Any crystal radio worth building should have good sensitivity. That means that you should be able to hear even relatively distant and weak stations. Of course, crystal radios have no amplification and depend entirely on capturing enough radio energy to operate the earphone to produce sound waves. The radio must have efficient tuning coils, tuning condensers and a efficient detector so that the tiny amounts of energy that is picked up by the antenna is not wasted or lost in the radio any more than absolutely necessary. If the radio is designed properly with low-loss components, the human ear, with its amazing ability to hear even very low energy sound waves, will do the rest.

Any crystal radio worth building should be in a pleasing enclosure or chassis that makes mounting the components easy, makes connecting to the antenna and ground easy, makes it easy to plug in the earphones and makes operating the tuning controls convenient. When the project is complete, the builder should have a radio that looks like a radio and something (s)he is proud of and proud to show to others.

Certainly, the Heathkit CR1 fits these criteria extremely well and I think that the little radio I will present on this page fits the criteria well too. In some respects, if I may say so myself, my radio fits some of them even better than the CR1.

An update as of June 21, 2017

After exchanging some email with a reader, I have revisited my little project and have rethought some things. I then went in and made some simple modifications and I am extremely pleased with the results. The modifications work so well, I am redrawing the schematics and the improved version of this radio is what I will present.

Building the radio

List of parts:

1) Two each, T80-15 toroid cores. **
2) Two each, dual section 335 X 335 PF variable tuning capacitors (670 PF total each). ***

3) Two each, knob and shaft extension for the above tuning capacitors.
4) Four each, 2.5 X 5 MM countersink screws to mount the tuning capacitors.
5) One 1N34A diode.
6) One piezoelectric earphone (or high Z magnetic earphones) with a miniature phone plug.
7) One panel mount miniature phone jack (to match item 6 above).
8) One, panel mount RCA type jack.
9) One RCA type plug with item 10 wire leads soldered to it to match item 8 above.
10) Two each, 6" 20 Ga. colored (red & black) insulated wire leads for the antenna and ground. (Use black for ground)
11) Two each, small alligator clips. Solder to the end of the above insulated wires for antenna and ground connections.
12) Two pieces of 40/46 litz wire each 7 feet long. No. 30 (or larger) magnet wire can be used but not recommended.
13) One 6" section No. 24 ga magnet wire for the interconnecting link. Litz wire can be used, but stiffer wire may be preferred.
14) One 24,000 ohm (24K) resistor. A 1/4 or 1/8 watt resistor is recommended to more easily fit in the limited space.
15) One L1 coil made of 6 turns of litz wire wrapped around a wooden match stick. Cut off unused wood.
16) One 12 coil made of 12 turns of litz wire wrapped around a wooden match stick. Cut off unused wood.
17) A single pole double pole miniature switch.
18) A small tube of RTV type adhesive. Do not use an adhesive that becomes hard because you may want to change coils.
19) Metal or plastic box to mount the components.
20) Antenna 60 + feet or as long as practical and mounted as high up as practical.

** Use only toroids made with the -15 material. Other cores will greatly complicate winding and/or degrade performance. Smaller cores may be used, but the number of turns will have to be adjusted.
*** You can use smaller values for finer tuning. For large chassis projects, conventional 360 PF air variable condensers are excellent and will probably give you higher 'Q' and lower signal losses. Ganging the two sections of the miniature condensers as I have done gives a greater tuning and antenna trim range. Since the tuning of a crystal radio is so broad, condensers ganged this way are still easy to adjust for maximum signal and minimum interference from stations off to the side.

Total cost of parts will be about $25 -$30.

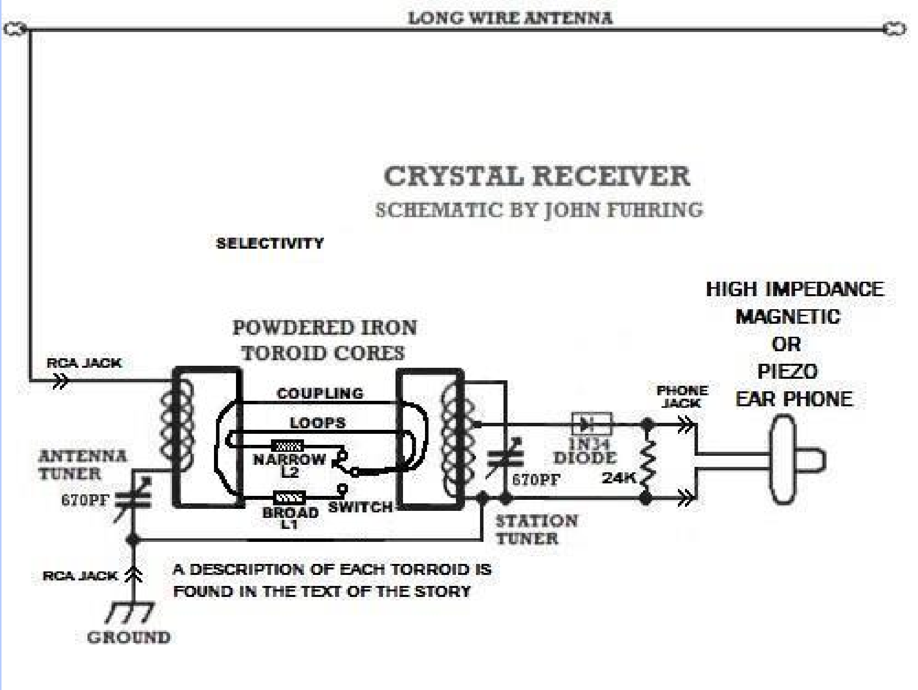
LONG WIRE ANTENNA
CRYSTAL RECEIVER
SCHEMATIC BY JOHN FUHRING
SELECTIVITY
HIGH IMPEDANCE
MAGNETIC
OR
PIEZO
EAR PHONE
POWDERED IRON
TOROID CORES
RCA JACK
COUPLING
LOOPS
PHONE JACK
NARROW
L2
1N34
DIODE
ANTENNA
TUNER
670PF
BROAD
L1
SWITCH
670PF
24K
STATION
TUNER
RCA JACK
A DESCRIPTION OF EACH TORROID IS
FOUND IN THE TEXT OF THE STORY
GROUND

Crystal radio schematic diagram.

Notes

I originally wound the coils with No. 30 magnet wire, but was dissatisfied with the selectivity and sensitivity. I had never used the stuff before, but I knew that a special kind of "woven" multi-strand wire worked best for radio coils operating at broadcast frequencies. Of course, I'm talking about litz (German for woven) wire that reduces something called the "skin effect." Because of the magnetic field that is generated by AC flowing in a conductor, electron flow is forced out to the "skin" or outermost part of the conductor. This is an undesirable feature of conductors carrying AC current but at 60 Hz it is hardly noticeable. The skin effect gets worse the higher in frequency you go and at radio frequencies, it is really bad causing high resistance in coils and ruining their 'Q' or "quality factor." Litz wire, because it is made of several wires woven together, noticeably increases a coil's 'Q' and allows it to tune much sharper. Litz wire also reduces a coil's resistive losses, thereby allowing the radio to pick up faint signals. I rewound both coils with litz wire and as expected, noticed a big improvement in both selectivity and sensitivity.

Next I experimented with detector diodes to see which kind gave the better sensitivity. Originally the radio used a BAT-46 Schottky diode for its detector. I even experimented with biasing the Schottky for additional sensitivity, but in the end, I replaced the Schottky detector with an ordinary 1N34A point contact diode. In my radio with its particular configuration, the old technology 1N34A diode gives superior performance.

I further experimented with with the the radio's selectivity by varying the coupling between the two toroid coils. With toroids, it is impossible to get a coupling lighter than a single turn unless you put a coil, a capacitor or a resistor in series with the coupling loop. To get the maximum selectivity out of one of these radios, it is necessary to have fairly light coupling between the antenna and the tuning toroid coils so at first I tried a resistor and it worked - sort of. When I put a little coil of Litz wire (wrapped around a bit of wooden match stick) in series with a single linking coil, I noticed a profound improvement to the overall selectivity of the set, but not much loss of sensitivity. A second coil was put in with twice as many turns on its matchstick and one or the other is selected by a little toggle switch. When this "larger" coil is selected, it very noticeably improves selectivity over the smaller coil, but it does cut down a little on the radio's sensitivity.

The station (detector) tuning coil is tapped 30 turns from the top so that 2,000 ohm magnetic headphones can be used without loss of selectivity.

In my radio's present configuration, its selectivity now matches the Heathkit CR1, but although very good, it isn't quite as sensitive.

Suggest layout and building instructions.

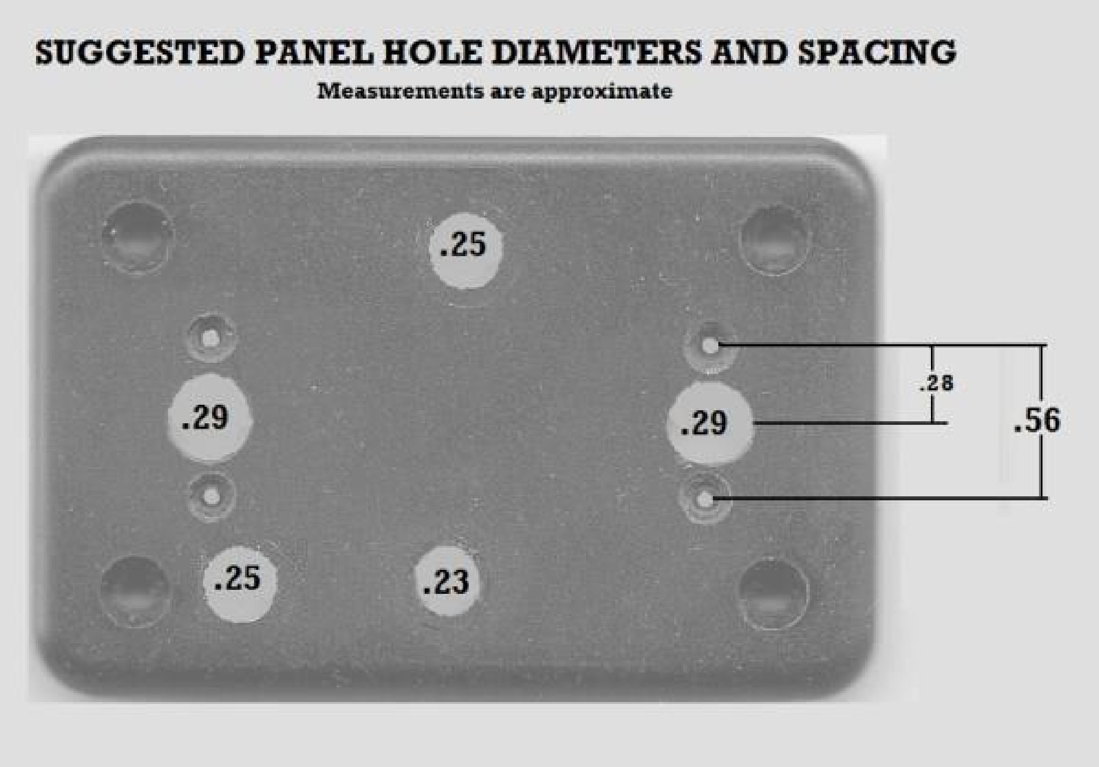
SUGGESTED PANEL HOLE DIAMETERS AND SPACING
Measurements are approximate
.25
.29
.29
.28
.56
.25
.23

All measurements are in inches.

Steps in building the radio

First, shop around and pick out a suitable chassis box to put the radio in. A small box such as shown above will accommodate the miniature tuning condensers and the jacks for the earphone and antenna/ground RCA jack, however the parts will be crowded as shown in the pictures. A larger box will make assembly and soldering somewhat easier. A plastic box will make drilling and mounting the components extremely easy. A small wooden box that could be lacquered and polished might be very handsome.

The two capacitor extension shafts should be filed down for an overall length of 0.4 inches with the mounting screw cut to 0.48 inches. You can leave them their original length, but the knobs won't be flush with the panel.

This radio uses two T80-15 toroid cores that are each wound with 7 feet of 40/46 litz wire. Three turns on each side form a coupling link between the toroids. The station tuner coil is tapped 30 turns from the top. At what will be the top of this coil and with an inch of wire sticking out, square knot tie the wire to the toroid (to hold it in place) and then begin winding the coil while counting each turn. At 30 turns, stop winding and create a tap. Tapping is accomplished by simply twisting the wire for about a half inch. Carefully start winding again and continue to wind until all the wire is used up (about 70 more turns). When finished winding, the twisted wire at the tap and the wires at the ends of the coils are heated with a soldering iron until the insulation melts and then "tin" the wires with hot solder.

Note that the diagram that accompanies the tuning condensers shows which of the long metal tabs are connected to the stator (ground) and the rotor sides of the units. These condensers have 4 trimmers built into them, but we will not use any trimmers so ignore (or cut off) their tabs.

With the photos and schematic as a guide, perform the following:

1) For the antenna tuning condenser, bend the rotor tabs over at a 90 degree angle and solder them together.
2) For the station tuning condenser, bend the rotor tabs over at a 90 degree angle and solder them together.
3) Extend the ground tabs (stators) of the two condensers until they touch and then solder them together. This is the ground bus.
4) Solder short pieces of bare wire from the outer shells of the RCA and phono jacks to this bus.
5) Trim the leads to fit and solder one side of the 24K resistor to this bus.
6) Place the other lead of the 24K resistor into the mounting hole of the ungrounded side of the phone jack.
7) Trim the leads of the detector diode to fit and place one lead in the same hole of the phone jack with the 24K resistor.

8) Solder the detector diode and 24K resistor leads from steps 6&7 to the phone jack.
9) Place the station tuning coil on top of the station tuning condenser and perform the following:
10) Solder the top of the station tuning coil to the station tuning condenser's rotor tabs.
11) Solder the bottom of the station tuning coil to the ground bus.
12) Place the antenna tuning coin on top of the antenna tuning condenser and perform the following:
13) Solder the top of the antenna coil to the ungrounded lug of the RCA jack.
14) Solder the bottom of the antenna coil to the rotors of the antenna tuning condenser.
15) Glue the two toroids to the tops of their tuning condensers as shown. Use a silicone rubber adhesive and allow it to set.
16) L1 With a 4 inch section of litz. wire, wind 6 turns on a wooden match stick and glue them down. Leave one end with a long pig-tail and the other end a little shorter. Trim off the excess wood.
17) L2 With a 4 inch section of litz. wire, wind 12 turns on a wooden match stick and glue them down. Leave one end with a long pig-tail and the other end a little shorter. Trim off the excess wood.
18) Solder the short end of L1 to the "wide" terminal of the switch. No stripping is necessary with litz wire.
19) Solder the short end of L2 to the "wide" terminal of the switch. No stripping is necessary with litz wire.
20) Thread the two L1 and L2 long wires through both toroids as shown and bring them out together to the center terminal of the switch. Twist them together and solder them on. Remove excess wire.
18) Solder the other side of the detector diode to the tap of the station tuning coil as shown.
19) Solder two 3-6 inch pigtail leads on to the RCA plug for the antenna (red) and ground (black) and solder alligator clips to their ends.
20) Connect the alligator clips to a long wire antenna and suitable ground, plug in the earphone and tune to a local station. Recheck wiring if you fail to detect a station.

Refer to the schematic diagram and the photos below to make sure you have wired everything correctly. Remember that a schematic diagram is not exactly as a radio is wired, but is electrically correct and actual wiring must make all the connections (and no others) as shown. As long as the leads are kept short and if all the connections are made properly and there are no shorts to other parts of the circuit, the radio will work It is up to you to figure out the best routing and lead lengths for all the connections shown in the schematic. Keep in mind that external connections such as to the earphone, the antenna and the ground go to the earphone jack and the antenna jack respectively. If you are not familiar with schematic diagrams and how to wire something up using such a diagram, this is an excellent project for you to start learning how this is done.

Finally, the RCA jack and the phono jack normally uses the outside shell connected to ground so that when they are bolted to a metal panel, they are at the right potential. If you are using an insulated panel to mount these jacks, it doesn't matter if the outside shell is grounded, but I recommended that you connect it this way. You may notice that the detector diode is shown in the schematic as "forward biased," however, it makes no difference (in this case only) which way you connect the diode.

Tools you will need

1) A low wattage soldering iron. A temperature controlled soldering station is ideal, but an ordinary iron will work fine.
2) A very fine needle nose plier or (better yet) a fine hemostat.
3) A small wire cutter.
4) Drill, bits and countersinking tool for cutting holes in the panel.
5) Screwdrivers to fit the various screws.
6) Pliers or wrenches for tightening the jack's mounting nuts.

Do not over-tighten screws or nuts.

Photos of the completed project in its tiny box

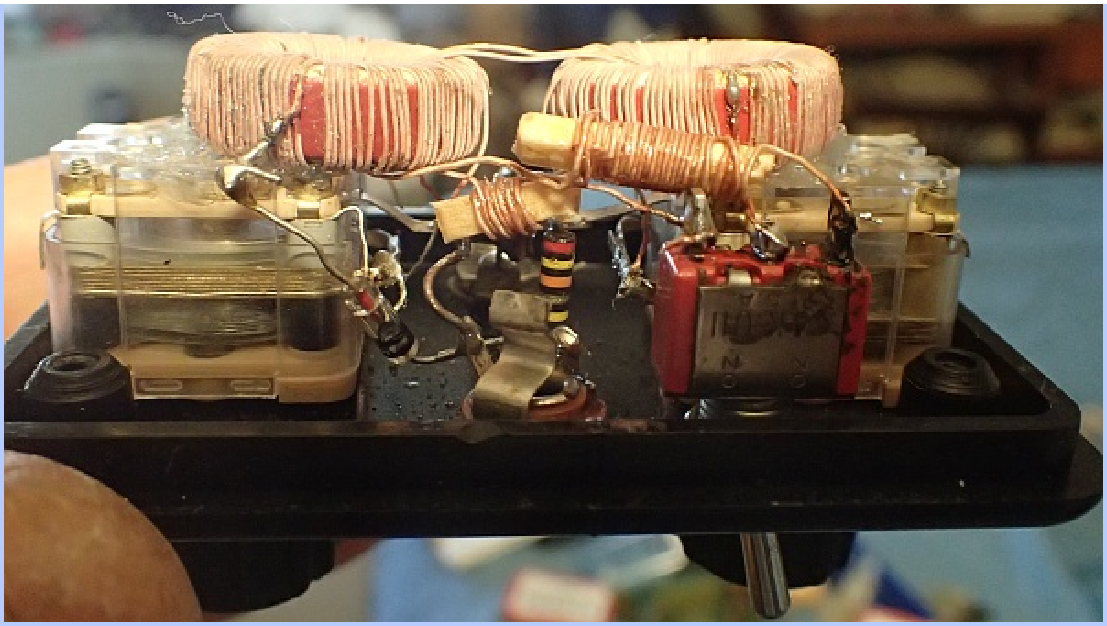

Plastic lid with all components mounted. The toroids are mounted with a drop of RTV.
Silicone rubber mounts the toroids quite securely, but allows for future changes.
Miniature tuning condensers have dual sections, each 335 PF and are ganged for 670 PF.
The long ground tabs of the two condensers are soldered together and make a very convenient ground bus that the other components can attach to.
Also shown here are L1 and L2 which are turns of litz wire wrapped on match sticks.
This little radio will tune to the LF band below the broadcast band, but I doubt you will find anything down there to listen to.

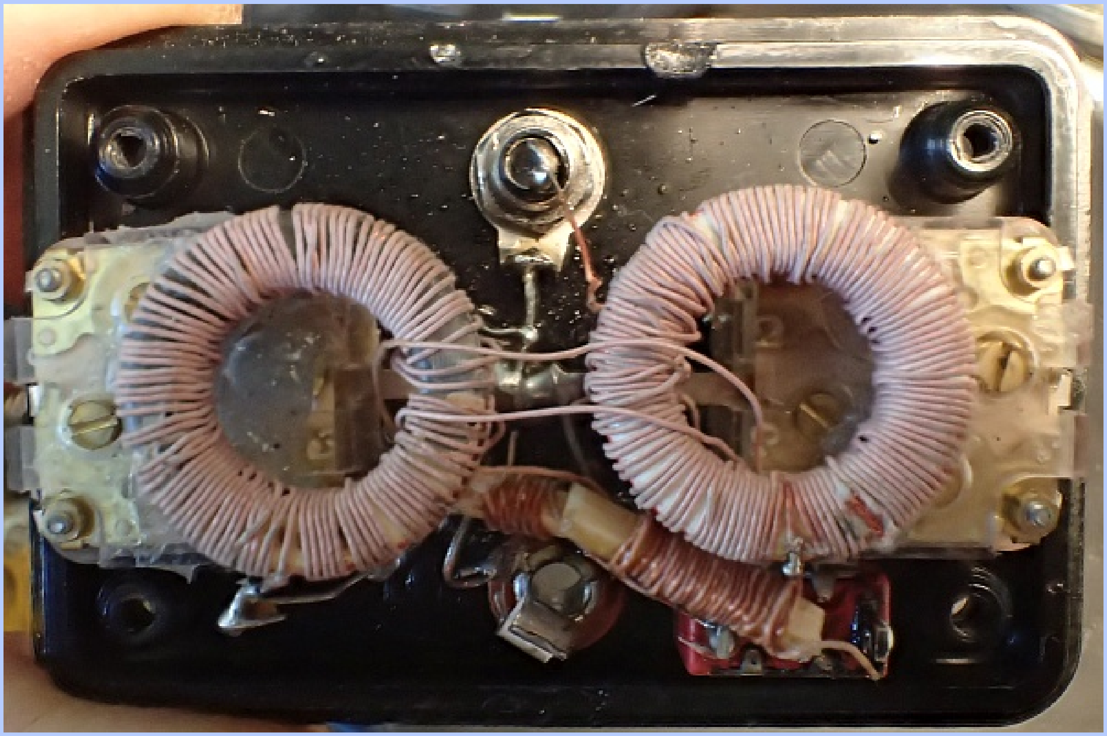

Looking down with the phone jack at the center bottom and the RCA antenna jack on the top.
Notice a ground bus runs in the center connecting the two condensers and that the other components are soldered to it.

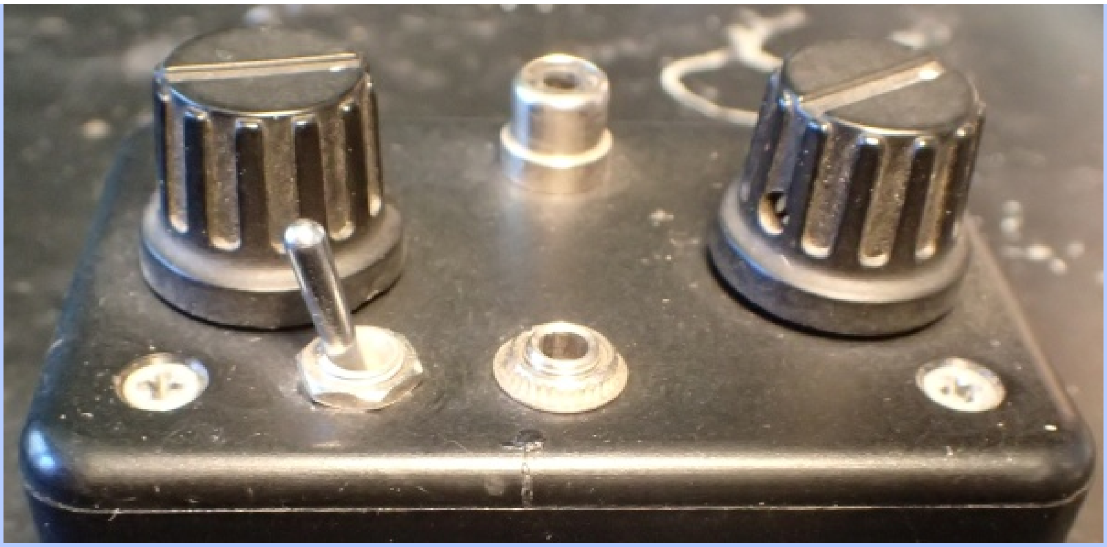

The front panel. Antenna tuning on the left, station tuning on the right.
The toggle switch controls wide or narrow selectivity tuning.

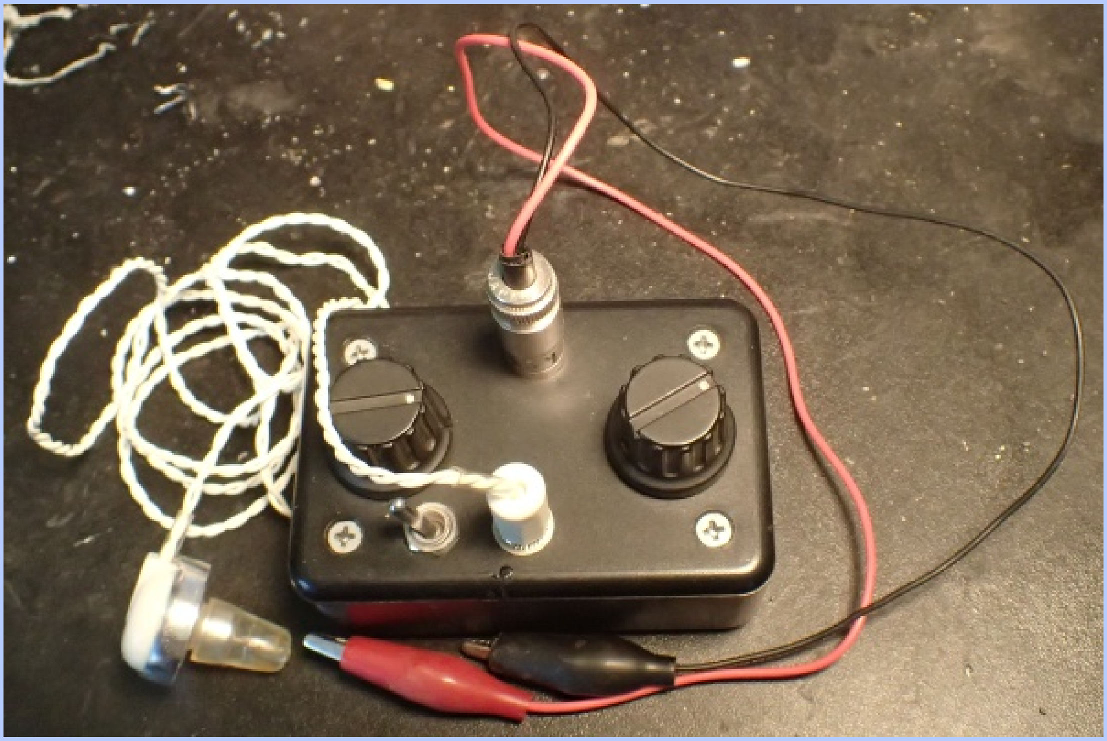

The completed project with a RCA plug attaching the antenna and ground.
I think you will agree that this is a very neat little package.
The tiny size belies its rather excellent performance.

ANT.
GND.
PHONES
Heathkit
CRYSTAL RECEIVER
DET.

Here's a picture of both my radio and the
Heathkit CR1 showing their relative sizes
This is an old photo before the wide/narrow
selectivity switch was installed.

The way this radio works is as basic as it is possible to get

and here's how it all works:

The setting:
Imagine that there is a radio station 10 miles away and that it that is broadcasting radio waves at 1,400 KHz.
Imagine that our radio station is broadcasting hundreds of watts of energy into space and that if we could capture just a tiny amount of this energy, it would be possible to hear what was being broadcasted.
Imagine that we have a crystal radio like the one above connected to a long wire antenna.

(1) The passing radio wave from our station induces a tiny high frequency voltage on our antenna wire. The antenna is connected to the antenna coil (coil on the left of the diagram) with its antenna tuning capacitor to the left and below it.
.

(2) When the antenna capacitor is tuned just right so that the inductance of the coil is matched by the capacitance of the tuning capacitor, a tiny current from the antenna flows in the antenna coil to produce an oscillating (or varying) magnetic field inside the toroid core on the left. By carefully adjusting the antenna tuning capacitor, this oscillating magnetic field will be right in tune with 1,400 KHz and 1,400 KHz only. This is important because it helps separate our 1,400 KHz signal from all the others on the AM band.
.

(3) The oscillating magnetic field created by currents in the antenna coil causes a voltage to be developed in the turns of the link coil through the principle of mutual inductance. The oscillating voltage in the link creates a oscillating magnetic field in the right side core the core that has the station tuning coil wound on it. The link isolates the two circuits so that unwanted signals are not coupled between them.
.

(4) The 1,400 KHz signal that has been coupled to the station tuning coil is contained in a "tank circuit" that is made up of the station tuning coil and the station tuning capacitor (shown to the coil's right).
.

(5) By carefully adjusting the station tuning capacitor so that its capacitance exactly matches the coil's inductance, a so-called tank circuit is formed. This tank circuit "resonates" or "tunes in" the 1,400 KHz signal and that means that it further separates our 1,400 KHz signal from all the other nearby signals. We now have just the radio signal we want, but we need to do more because it is impossible for our ears to hear radio frequency signals.
.

(6) To "detect" the presence of the radio frequency signal that is resonating in the tank circuit, the station tuning coil is wired to one side of the 1N34 detector diode.

.

(7) The 1N34 diode chops off the negative part of the radio frequency wave. The detector needs to do that or the negative part of the wave would cancel out the positive part of the wave and nothing would be heard.

.

(8) When the diode detector chops off part of the radio frequency signal, direct current audio frequency undulations are all that are left over from the original radio wave. These direct current audio frequency undulations correspond to the sound being transmitted by the station and when the wires of a very sensitive earphone are connected to the detector, these undulations of direct current cause the mechanism inside the earphone to vibrate in step with the undulations and it is those vibrations that we hear as sound.

Here is a graphical representation of what's going on when a radio frequency signal is detected (turned into sound):

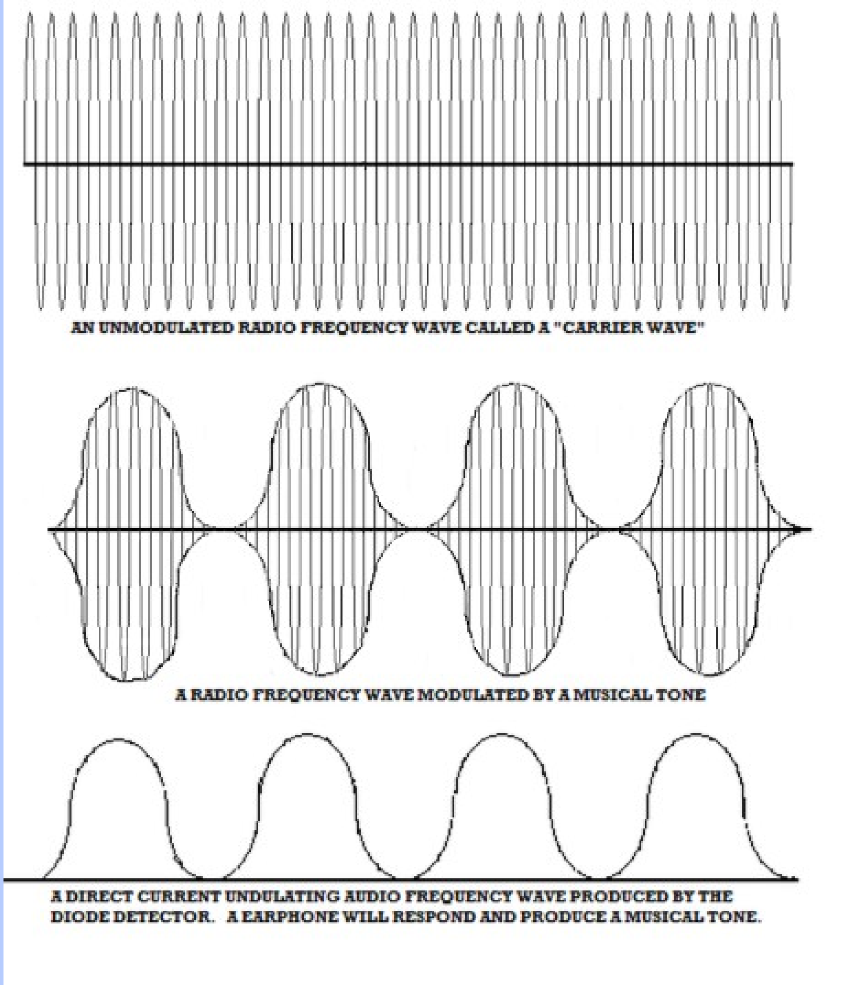

AN UNMODULATED RADIO FREQUENCY WAVE CALLED A "CARRIER WAVE"

A RADIO FREQUENCY WAVE MODULATED BY A MUSICAL TONE

A DIRECT CURRENT UNDULATING AUDIO FREQUENCY WAVE PRODUCED BY THE DIODE DETECTOR. A EARPHONE WILL RESPOND AND PRODUCE A MUSICAL TONE.

Graphs of the various waveforms seen at the tuner stage (top two) and at the detector stage. Unmodulated carrier waves are also called "continuous waves" and can't be heard by a crystal detector except as a kind of a very faint humming noise and a quiet spot on the tuning dial. AM radio stations try to avoid sending out unmodulated carrier waves (called "dead air") and so this waveform is rarely seen. More common is the waveform showing lower frequency sound waves superimposed on the carrier wave.

The thick black line is where the voltage is zero as a carrier wave crosses from positive to negative. As shown here, the negative going parts of the wave are stopped by the detector diode so that positive to zero to positive undulations remain. If you reverse the diode, the direct current undulations will be negative to zero to negative, but the same musical tone will be heard in the earphone.

Crystal radios use only one half of a radio frequency's wave because full wave rectification is unnecessarily complicated and results in reducing the selectivity with little or no increase in sensitivity.

The importance of a good antenna and ground system

These radios are not connected to your house's AC power and they use no batteries. A crystal radio has no amplifying circuits and yet they output sound and sound is energy. Where do they get the energy to make sound? Do they somehow use mysterious and occult "free energy from space" or something goofy like that? The energy needed to run these radios doesn't cost you anything directly, but it isn't free or mysterious. The energy is from a tiny fraction left over from the hundreds of watts of radio frequency energy a station broadcasts out to space. The vast, vast majority of the energy a broadcast station radiates out of its antenna system is lost, but a tiny fraction of it may be picked up by an antenna at somebody's house.

Antennas for listening to AM stations can range from tiny "loop" antennas that are extremely inefficient, but good enough for amplified radios, all the way to elaborate "long wire" systems dozens and dozens of feet long and high up in the air. The general rule was discovered by Marconi over a hundred years ago that the longer and higher the antenna, the better it is at capturing some of that energy that is passing by. To capture enough passing energy so that it can be processed by the crystal radio's components and then come out as sound energy strong enough to vibrate the mechanism inside the earphone, the antenna needs to be as long and as high as is practical. Practical is what you have room for and how high you can safely go. Practical antennas can be strung between poles located at the ends of a roof or from the house out to a distant tree or pole. It is true that antennas longer than 60-100 feet and higher than 25 feet do capture more energy, but things soon get very difficult and for very little extra gain at distances and heights beyond what is practical.

In addition to a really good long wire antenna (as long as is practical as mentioned), this radio must have a good ground. Right now, I am using my house's wiring as a ground and it seems to work well. Water pipes, when available to connect to, are supposed to work even better. When using house wiring, it would be a deadly mistake to connect to anything but the AC ground, so be careful.

Tuning this radio for maximum sensitivity and selectivity

The trick to tuning this radio is to first locate a strong local station of known frequency and tune it in with both the antenna tuner and the station tuning controls. Both controls should be used to peak the signal for maximum loudness. Once you have an idea where on the dial your station is located, move both controls to about where you guess the weaker station you want to listen to might be. Once there, slowly and carefully adjust the station tuning until you hear something and then peak it up with the antenna tuning control. You may have to go back and forth a few times to get maximum volume with minimum interference from nearby and stronger stations.

The best time of the day to have fun with these kinds of radios

During the daytime and during the summer months, about all that can be received with a little set like this, at least around here, is the trash in English that the local broadcasters pander to and a couple of Mexican language and music stations, but in the evening, electrified layers way up high in the outer fringes of the atmosphere allow radio waves to be bent back down to earth so that interesting stations worth listening to, but which are 200 to 300 (sometimes even more) miles away may be heard loudly and clearly. This long distance "refraction" or bending of the radio waves back to earth (a process called "skip") makes it possible to hear these distant stations, but it isn't like listening to the station next door. This "skipping" of radio waves off electrically charged layers in the upper atmosphere makes these signals subject to all kinds of things including interference from even more distant stations, fading in and out, some strange sounding distortions and sometimes there is even an echo effect. These so-called "propagation effects" can be annoying, but if you just wait a few seconds or minute or so, the signal generally returns.

I think that listening for yourself, to hear what Nature is doing to these signals way up in the sky, is part of what makes listening to the Broadcast Band at night so interesting. There is a certain randomness to this skip that is explained to some degree by solar astronomy and physics. Scientists and engineers all over the world closely monitor our local star and measure it for electrical and magnetic activity. Reports and prediction tables are compiled by various scientific agencies and all this stuff is available on the Internet. When you can test these predictions for yourself by listening for "skip," you are participating in a scientific observation in your own little way.

Building and operating a crystal radio gives insights into science and technology

Listening to far distant stations with a crystal radio like this one shows you how ships like the Titanic could communicate with other ships and to shore stations hundreds of miles away with nothing more than crystal radio receivers. Speaking of the Titanic, two thirds of all the people on her died and only about 700 were rescued, but without early radios, very much like this crystal set, everybody would have died. In the 100 years since the Titanic disaster, just imagine how radio has connected ships, airplanes, space missions and your cellphone with the rest of the world.

If you make one of these radios and actually put it to use, you will see how, with a good antenna and good sharp tuning circuits, individual radio signals can be picked out from the mass of signals out there and a station you might want to listen to can be heard from a long way off under the right conditions. For me, this is a very wonderful thing, especially when I consider that these radios do not amplify their signals, but use the tiny amounts of electrical energy that is picked up by the antenna to make sound. Such is the extreme sensitivity of human hearing and the amazing efficiency of the circuits, the detector crystal and the earphone, that microwatts of power, originally coming from a station hundreds of miles away and landing on my antenna wire, can be heard clearly.

How far we have come

When you think about it, it seems that such basic radio technology represented by this little crystal radio is quite old, but in fact, the first Fleming Valve and crystal radios were invented just a little over 100 years ago. When my mother was a girl, people were building crystal sets so that they could listen to the wonderful new radio broadcasts from the big cities. In a relatively short span of time, the time since my grand parents were young in the 1890s to my parent's time, to this very moment, consider just how far we have come with radio, television, GPS, cell phones, satellite broadcasting, wireless Internet, interplanetary missions to space and so much other allied radio technology.

A totally irrelevant comment

I know that the following is irrelevant to this article, but I'd just like to say that compared to the one tube (valve) Armstrong regenerative radio I built in 1958 while in the 8th grade (and without any adult help or supervision and the story of which is found elsewhere on my website), both the Heathkit and this little crystal radio really suck. It is no wonder that the crystal radio was quickly superseded by valve radios (starting about 1919 and using Armstrong's designs) even though the early vacuum tubes (valves) were extremely expensive, ate up expensive batteries and quickly burned out. Yes, even by 1920 the crystal radio was technically obsolete and their owners started to become just a tiny fraction of the listening public that they are today. Nevertheless, crystal radios did not completely disappear. They have remained popular all these years and years and years because these little radios are easy and cheap to build and they have provided endless hours of fun and precious education to uncounted numbers of people over the decades.

Why you might want to have second thoughts about building this radio

(These are my personal opinions that you are not obliged to agree with and you may skip down if you wish)

There is something that should be considered when thinking about having your young people build any kind of AM radio and what I'm referring to is the unpleasant realization that, in many places in the U.S., there is absolutely nothing on AM radio that is (in my opinion) good for young, developing minds to listen to. Many local broadcasters pander to the most disgusting right-wing propaganda and Fundamentalist religious crap. Anyway, it is my opinion that persons wishing to direct young people into building a simple AM radio should carefully monitor their local AM broadcasting for the kind of trash I'm taking about. If the local airwaves are full of this trash, they must ask themselves: would building this kind of a project likely help or harm their young people? I think building an AM radio can have the potential to do more harm than good if your area is dominated (as mine is) with this sort of trash broadcasting. Now, on the other hand, if you actually WANT your kids to be exposed to religious or political propaganda before before their minds are sufficiently developed to resist brainwashing, then ignore this warning.

Of course, this applies to adults wishing to build one of these radios too. Is it really worth the expense and especially the effort to build a crystal radio when there is nothing to listen to? I did it strictly for the technical challenge, but certainly not to listen to local entertainment.

All this aside, if you decide build this radio or (better yet) my Armstrong "crystal" radio, I think you will have a lot of fun and will learn a lot. Just remember that for about the same money you can get a MP3 player with an FM radio and 8 gigs of memory. While the MP3 player is a better deal, listening to it won't teach you anything about electronics. However, the horrible trash that's on AM radio has a well deserved reputation for making its listeners stupid and crass while listening to classical music through your MP3 player just might make you smarter and highly cultured, but it's your choice.

Other considerations

Finally, there are some serious downsides to building this radio you should be aware of. One downside is the extreme tediousness of winding the toroid coils. Let's face it, few of us have toroid winding machines and putting 100 turns (about 7 feet of wire) on one of these cores is time consuming and boring with many a tangled wire. Mounting the tiny components wasn't all that difficult because the plastic box lid was very easy to drill, but it was fine work and time consuming. The most serious downside is the very fine and very careful soldering that must be done to connect up all the very tiny parts. If I would have chosen a larger box to put everything in, the soldering wouldn't have have required so much patience and expertise, but it would still be, in my opinion, beyond the skill of most middle school (and younger) kids to do except under the most close and individual supervision especially when you consider that the young person would be handling a hot soldering iron that could burn him/her badly if mishandled.

A final word

I have to admit that with the exception of the Heathkit CR1 radio, this is the best crystal radio I have ever made or used and I'd say the project was a success. I'm sure that people will tell me that there are some ultra high performance crystal radios out there that have very sophisticated low loss coils, high quality tuning capacitors and superior detectors with that could outperform this and the CR1, but I have no desire to go to the trouble to try to build them.

By the way, if you build a crystal radio based on what has been presented here, please drop me a line and describe it to me and, if you can, attach some pictures. I'd sure appreciate it if you would tell me how much you like the radio and how it performs for you. You can write me directly at my geojohn.org mail address. Be sure to look over my story of the Armstrong "Crystal" Radio before you decide to build this or any crystal radio.
Good luck.

The End

Having arrived this far, obviously you have a superior attention span and reading ability that far exceeds that of the majority of web users. I highly value the opinion of people such as yourself, so I ask you to briefly tell me:

Did you enjoy this article or were you disappointed?

Please visit my guest book and tell me before you leave my website.

Or you can **contact me by E-mail** directly if you have any questions or comments.

If you liked this article, perhaps you would like to read about the radio my radio is based on.

The high performance Heathkit CR1 crystal radio

Crystal Radio Kit

model VEC-121K

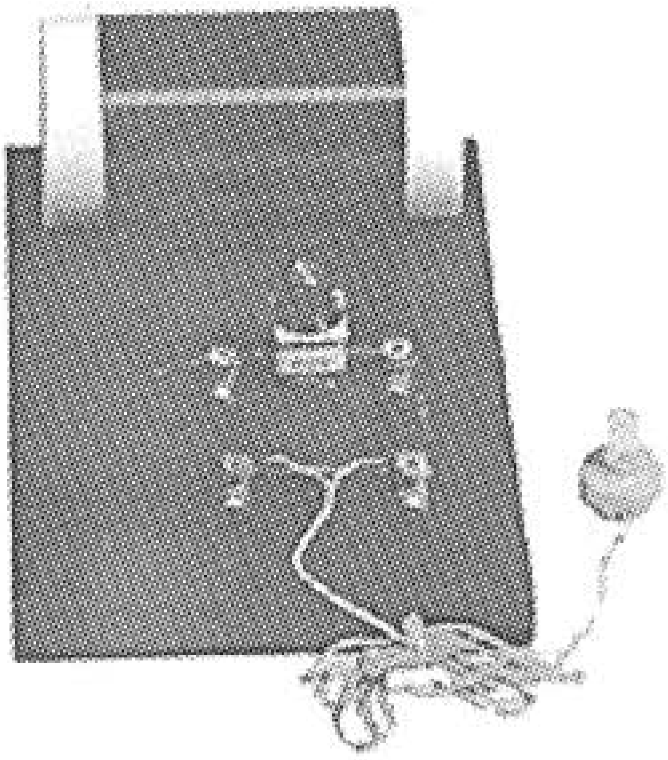

Relive the experience of early radio pioneers. Wire up the earliest radio circuit and listen to the magic of radio that needs no power. This is a lot more than just a nostalgic novelty -- this little baby really works!

- **No soldering needed! Great educational project for kids!**
- **You wind your own inductor and wire up the earliest radio circuit just like the early radio pioneers.**
- **This humble crystal radio set uses a crystal diode detector designed nearly 50 years ago -- you'll get performance that our earliest radio pioneers would envy.**
- **Put up an antenna, connect a ground, and you'll be amazed at the stations you'll pull in and hear.**
- **Includes all hardware and parts needed to build a working crystal radio set plus wire for an antenna and a sensitive earphone.**

Simple: No soldering. Requires a small screwdriver, wire-strippers, cutters. You wind your own coil. Doesn't require test equipment for final adjustment of tuning.

VECTRONICS

High-performance electronic kits . . . fun to build and use!

IMPORTANT WARRANTY INFORMATION! PLEASE READ

Return Policy on Kits When *Not* Purchased Directly From Vectronics: Before continuing any further with your VEC kit check with your Dealer about their return policy. If your Dealer allows returns, your kit must be returned *before* you begin construction.

Return Policy on Kits When Purchased Directly From Vectronics: Your VEC kit may be returned to the factory *in its pre-assembled condition only*. The reason for this stipulation is, once you begin installing and soldering parts, you essentially take over the role of the device's manufacturer. From this point on, neither Vectronics nor its dealers can reasonably be held accountable for the quality or the outcome of your work. Because of this, Vectronics cannot accept return of any kit-in-progress or completed work as a warranty item for any reason whatsoever. If you are a new or inexperienced kit builder, we urge you to read the manual carefully and determine whether or not you're ready to take on the job. If you wish to change your mind and return your kit, you may--but you must do it *before* you begin construction, and within ten (10) working days of the time it arrives.

Vectronics Warrants: Your kit contains each item specified in the parts list.

Missing Parts: If you determine, during your pre-construction inventory, that any part is missing, please contact Vectronics and we'll send the missing item to you free of charge. However, *before* you contact Vectronics, *please look carefully* to confirm you haven't misread the marking on one of the other items provided with the kit. Also, make certain an alternative part hasn't been substituted for the item you're missing. If a specific part is no longer available, or if Engineering has determined that an alternative component is more suitable, Vectronics reserves the right to make substitutions at any time. In most cases, these changes will be clearly noted in an addendum to the manual.

Defective Parts: Today's electronic parts are physically and electrically resilient, and defective components are rare. However, if you discover an item during your pre-construction inventory that's obviously broken or unserviceable, we'll replace it. Just return the part to Vectronics at the address below accompanied with an explanation. Upon receipt, we'll test it. If it's defective and appears unused, we'll ship you a new one right away at no charge.

Missing or Defective Parts After You Begin Assembly: Parts and materials lost or damaged *after construction begins* are not covered under the terms of this warranty. However, most parts supplied with VEC kits are relatively inexpensive and Vectronics can replace them for a reasonable charge. Simply contact the factory with a complete description. We'll process your order quickly and get you back on track.

Factory Repair After You Begin Assembly: *Kits-in progress and completed kits are specifically excluded from coverage by the Vectronics warranty.* However, as a service to customers, technicians are available to evaluate and repair malfunctioning kits for a minimum service fee of $18.00 (½ hour rate) plus $7.00 shipping and handling (prices subject to change). To qualify for repair service, your kit must be fully completed, unmodified, and the printed circuit board assembled using rosin-core solder. In the event your repair will require more than an hour to fix (or $36.00, subject to change), our technicians will contact you in advance by telephone before performing the work. Defective units should be shipped prepaid to:

Vectronics
300 Industrial Park Road
Starkville, MS 39759

When shipping, pack your kit well and include the minimum payment plus shipping and handling charges ($25.00 total). No work can be performed without pre-payment. Also, provide a valid UPS return address and a day time phone number where you may be reached.

INTRODUCTION

Relive those bygone days of radio--a time almost 90 years ago, when early radio pioneers experimented with crystal radio receivers! Wind your own "inductor" coil, wire up the earliest radio circuit; and, there's no soldering needed! Everything is assembled on a pine "breadboard", just like great, great granddad did when he was a youngster. Learn about basic radio theory. Put up an antenna (wire included), connect a ground, and experience the magic of a radio that needs no power. A special high-sensitivity earphone allows you to hear AM Broadcast signals from many miles away; this little baby really performs. The VEC-121K is a fascinating project for youngsters of all ages!

The VEC-121K features a modern "fixed detector", eliminating the need for frequent and frustrating catwhisker adjustments. A large, Hi-Q antenna coil, and a modern miniature Hi-Q tuning capacitor, are used for best sensitivity and selectivity. The high-impedance parallel-resonant tuning circuit allows a short receiving antenna to deliver maximum signal to the set.

TOOLS AND SUPPLIES

Having the right tool to do a job always makes things go better. There is no soldering required, and only a few basic hand tools are needed to assemble your VEC-121K crystal radio set.

This is a list of the tools you will need:

SAFETY EYEGLASSES or GOGGLES! Working with wires, tools and radio parts is a lot of fun, but please, *safety first*! Your local hardware store carries inexpensive safety goggles. Before you are going to be cutting or working with wires, hammering, handling any chemicals, or using any tools, put on your safety eye wear first!

1. **Phillips screwdriver.** There are eight Phillips head screws included in the kit, and we will use this tool when it is time to mount parts on the wood board. Try to find a screwdriver with a *shaft* (the distance between the tip and start of the handle) that is at least 3" long.
2. **Long-nose pliers.** These pliers have a long and tapered nose and are used by electricians. The long slender nose makes it easy to hold or bend a wire that is being worked on.
3. **Wire cutters.** These special pliers have sharp cutters to cut wires. Some long-nose pliers styles feature built-in wire cutters. Wire cutters are sometimes called "nippy cutters" or "diagonal wire cutters"—depending on how they're made.

4. **Ruler or tape measure.** A small 12" ruler or flexible tape measure will do.
5. **Small penknife.** A penknife, or small Exacto™ knife, is needed to scrape off wire insulation. If you aren't old enough to use a knife, a small section of sandpaper will do.
6. **Small Hammer.** It will be used with a finishing nail to mark where the screw holes are placed on the pine board.
7. **Small-Screwdriver or Hex-Tool**. Size to fit the set screw in the tuning knob--depending on which type of knob is supplied with your kit.

Radio Shack carries many inexpensive small hand tools for electronics work.

Here are some items that will come handy:

1. **Masking tape.**
2. **Cyanoacrylate adhesive.** This is the fancy name for contact cement, which is sold as "Super Glue™" or "Crazy Glue™". You'll use the contact cement to hold the coil windings in place. These adhesives are very strong, and can instantly glue your fingers together if you are not careful! If you are not old enough to work with contact cements, have someone who is older help you when gluing is directed in the instructions. The glues used to assemble plastic model airplanes are safer to work with and may also be used instead of super glue. However, they will require a longer time to dry. Always have an adult advise and supervise your use of any chemicals!

WOODWORKING AND FINISHING (OPTIONAL)

Perhaps this is your first kit building experience, or your first homemade receiver project. If so, your VEC-121K crystal receiver will be a source of pride for many years to come. Before assembling this kit, you might wish to consider sanding and staining the pine board—this will bring out the beauty of the wood, and make your project look as if it was done by an "*old pro*". Many early radios were built in what was called "breadboard" fashion. The radio parts were laid out on a small section of wood board.

Early radio parts were very expensive, even a simple crystal radio set could cost a week's pay! Most experimenters had to improvise. Often an old breadboard (used to roll out bread dough) was used as a base to mount radio parts. If you were well-to-do, your "breadboard" was probably a fine piece of mahogany or walnut board!

Sanding: Before applying a stain finish, carefully sand the pine board using 200 grit sandpaper. When sanding the surfaces, always sand in the direction of the grain to avoid scratching the wood. The two board edges that expose the end grains of the wood will be rather rough to the touch, and will require some extra sanding to produce a smooth even surface. You may have to use a 150 grit in the beginning on the end grains. Sanding is best done outdoors. Wear a disposable paper dust mask while sanding.

Finishing: Your local hardware store stocks small cans of woodworker stains. Either a walnut or mahogany wood stain finish is a good choice. Both oil-based and water-based stains are available. We suggest using a water-based stain for this project—it is easier to work with and easier to clean up afterwards. You will need a clean cloth, or *staining cloth*, to apply the stain. Follow the manufacturer's directions very carefully, and read the safety information on the label. Use disposable latex gloves when handling stain. If you're too young to safely work with wood stains, ask an adult for assistance with this portion of the project. *Always handle any chemicals with great care!*

Important Note: Once the stain is dry to the touch, the assembly work may begin. Before attaching the tuning capacitor with the double-sided adhesive tape, the board should be allowed to dry for a few days. This will allow the stain to fully *cure,* or dry, so it does not interfere with the tape adhesive.

BEFORE YOU START BUILDING

Your work area: Use a good workbench with good lighting so you can see what you are doing. At times we will need to set items aside to allow glue to dry; or you will need to stop working on your project for other activities. Your work area should be in a place where your project does not interfere with others.

Keep your work area clean and free of clutter. Kit parts should be kept in the box until you need to use them. When you are finished for the day, return the tools to their proper place. Put unused kit parts back in the box so you will not lose them. If you are old enough to use glues and wood stains, you may have younger brothers and sisters who aren't! Keep chemicals out of the reach of younger children.

Following directions: Kits are built in "steps". These steps follow a logical order, that means following the instructions will make your kit building experience easier and ensures that your project will work *first time* when finished.

You will notice a check box ☐ as each instruction gives you a simple task, or direction, to follow. When the task is finished, make a small check mark in the box to show that part has been completed. This will allow you to keep track of where you left off when returning from a break.

For example, a direction may read:

☐ Tighten the screw until snug, this will make a good electrical connection between the Fahnstock clip and wire.

When finished following the directions, "check off" that the step has been done.

☑ Tighten the screw until snug, this will make a good electrical connection between the Fahnstock clip and wire.

Don't rush!

Have you ever noticed that when you try to rush doing something, it never turns out right? Don't try to build the crystal set receiver in one day! The time spent building a kit is most of the fun. We will be "breaking down" the kit assembly into several phases. For example, the first phase is preparing the "breadboard". The next kit phase involves winding the coil.

Working with wire:

Unspooling: Your kit includes a 100-foot roll of 18 AWG enamel-coated wire. This wire is used to wind the tuning coil, for your receiver's antenna, and is also to be used for making connections between the radio components. Do not remove the wire tie from the roll of wire until you are ready to use it. There is a right and wrong way to remove wire from the spool! Never allow the wire to uncoil, like a Slinky™ toy! Doing so will cause the wire to twist and make kinks when you try to use it. Instead, hold the spool of wire in both hands, and carefully unroll the wire from the spool. Re-tape the unused portion when finished.

Removing insulation: Electrical connections are made by removing the insulation from the enamel wire. When you need to make an electrical connection, you will be instructed to remove the wire insulation. We will tell you where the insulation should be removed, and how much to remove. This can be done by using a small piece of 150 grit sandpaper. Fold the sandpaper in half so the "sandy" sides face together. Place the wire into the sandpaper fold, and slide the sandpaper back-and-forth over the wire while applying a gentle pressure. Turn the sandpaper every so often to fully remove all of the insulation. Gently scraping the insulation with a small penknife will also work. Lay the wire on a piece of scrap board, and draw the knife blade over the insulation to scrape it away. *Always scrape in a direction away from your body!*

All of the insulation must be removed to make a good electrical connection. The bare area of the copper wire should be bright and shiny when you are finished.

Making connections:

The directions will tell you to make one of two types of connections: *"End"* connections or *"through"* connections.

End Connections: End connections are made when the end of a wire is connected. To make an end connection, you will have to remove about 1" of insulation from the wire and form the end into a "fishhook" shape, as shown below.

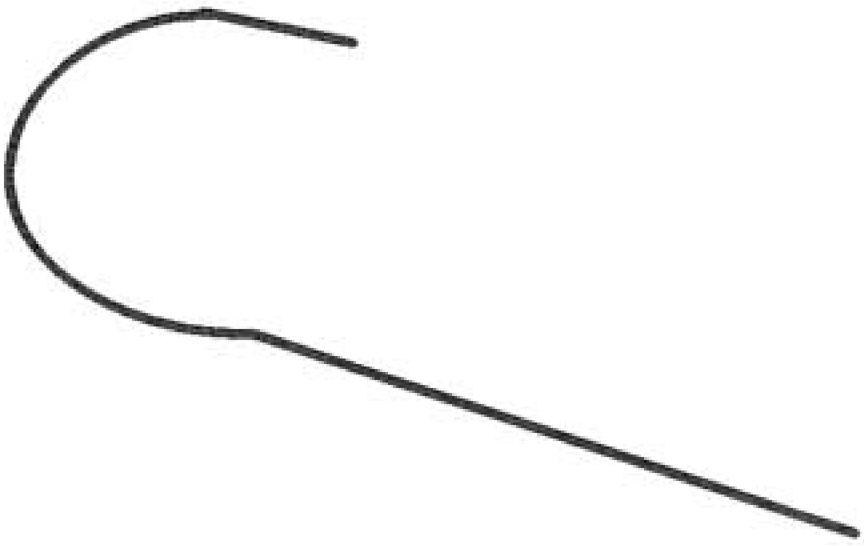

In the following drawing, a wire end connection is used to make an electrical connection to one of the Fahnstock clips mounted on the pine board. Note we show the wire as a "dotted line" where it passes *under* the Fahnstock clip. When the Phillips screw is tightened, a good electrical connection is made. The drawing shows how the wire hook is placed between the Fahnstock clip and board.

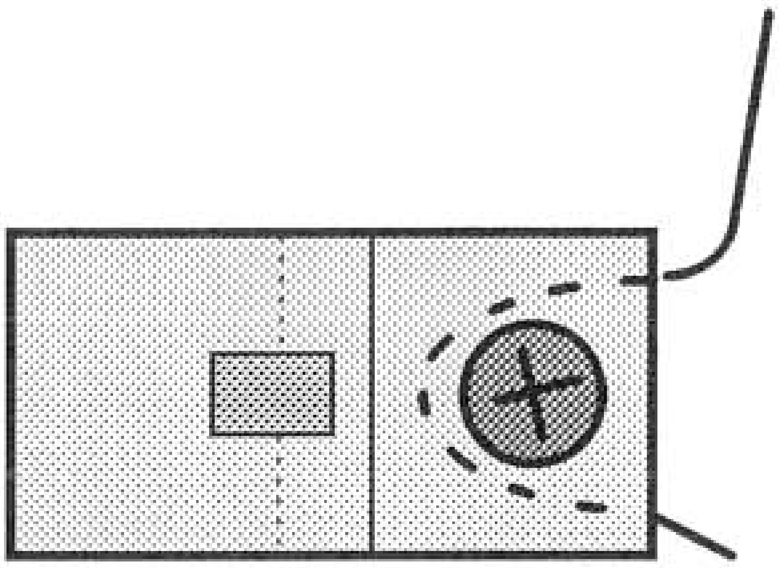

Through connections are used when the wire goes to a electrical connection at a Fahnstock clip, and then continues on to make one or more connections to other points. The directions will tell where to make a through connection on a wire, and how much insulation needs to be removed.

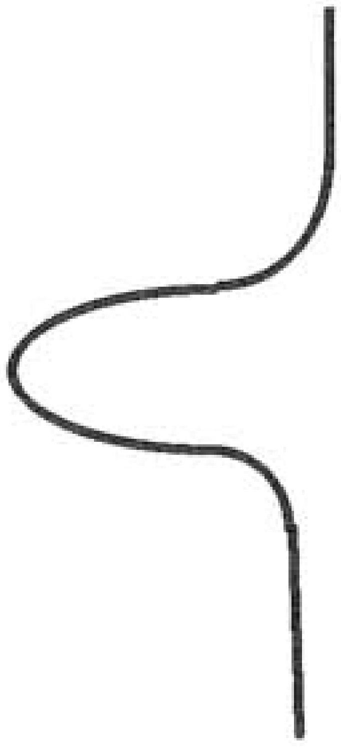

The wire insulation is removed where shown in the directions, and made into a half loop as shown above.

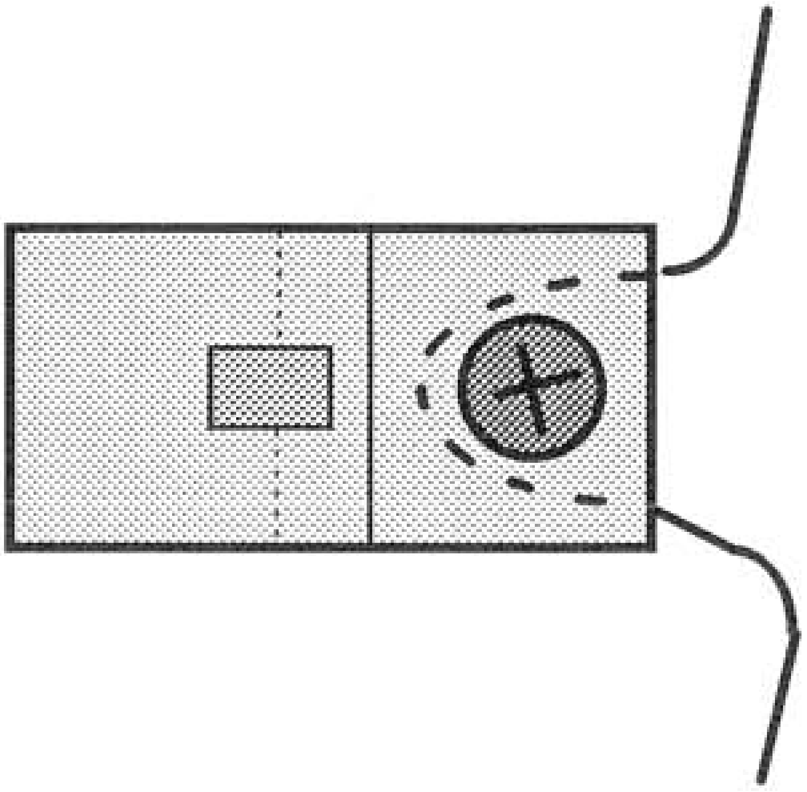

The above drawing shows a through connection being made to a Fahnstock clip. The dotted line shows where the wire runs under the clip body.

PARTS LIST

You are just about ready to begin your kit building adventure! But first, let's take a few minutes to make sure everything needed is present and accounted for. If any parts are missing or damaged, you'll need to refer to the directions given in the Vectronics kit warranty section of the manual.

☑	Qty	Part Description
☐	8	Phillips head screws
☐	1	5 inch length of white 2-1/2" diameter PVC plastic pipe
☐	1	100 foot spool of 18 AWG enamel coated wire.
☐	6	Fahnstock clips.
☐	1	Tuning capacitor shaft extension
☐	1	Tuning capacitor
☐	1	Tuning knob
☐	1	Double-sided foam tape
☐	1	Earphone
☐	1	Pine board , 3/4" high by 5-1/4" deep by 6" wide
☐	1	Insulated stranded wire, 7 feet long
☐	1	27K resistor (red-violet-orange-gold)
☐	1	1N34A germanium diode

Fahnstock clips

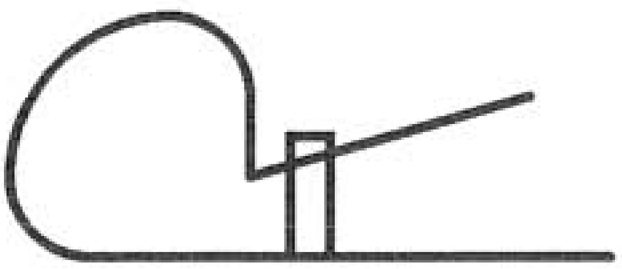

Tuning capacitor.

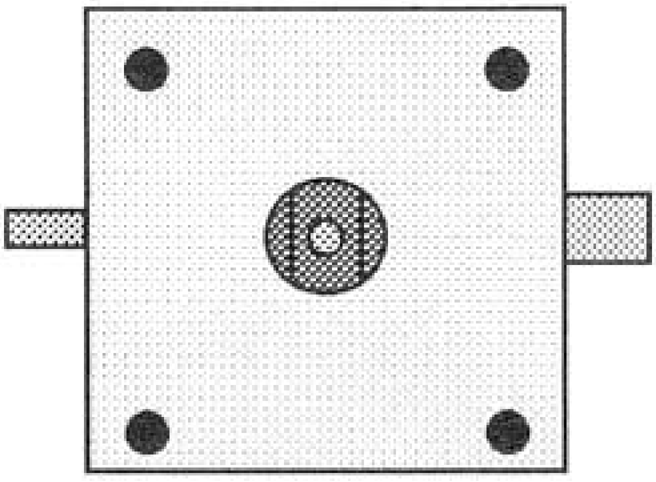

Resistor:

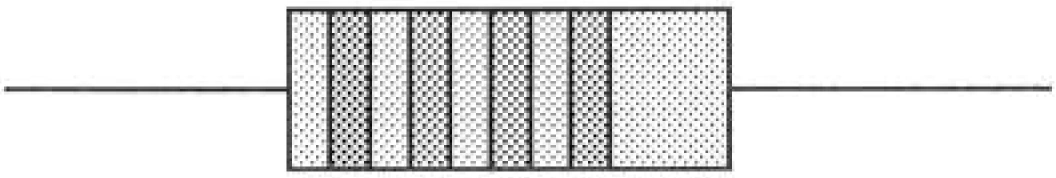

Germanium diode:

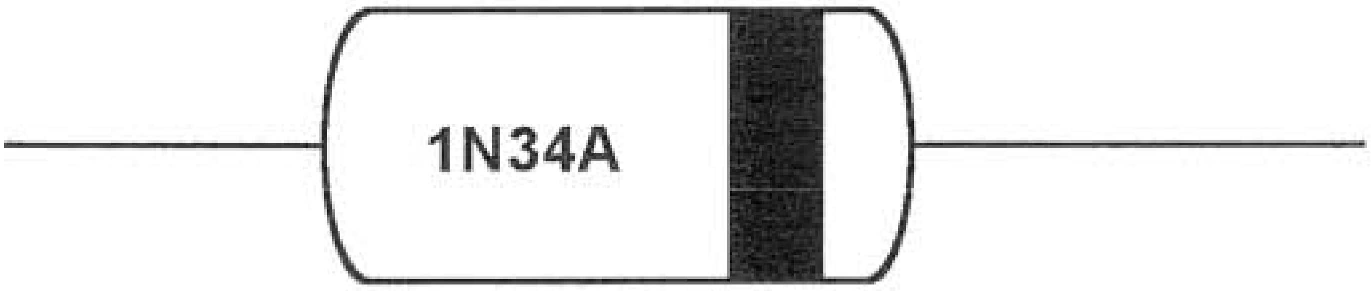

STEP-BY-STEP CONSTRUCTION

Before starting any construction, please read the Vectronics warranty. Kits that have been started *cannot* be returned for credit. Make sure this kit is within your skill level before starting assembly!

Phase 1: Breadboard preparation

☐ Find the pine board. Look it over carefully, and select which side you wish to use for mounting the parts. Chose the side that has the fewest imperfections, and has the nicest looking grain pattern.

☐ Locate the drawing that shows the screw hole pattern for the pine board.

☐ Place the screw hole pattern guide over the board, be sure the board edges line up with the outline on the drawing. Use masking tape to hold the pattern to the board.

☐ Using a finishing nail and small hammer, make a small punch mark at each of the eight screw hole locations. These marks will show you where the screw holes are located on the board. *Don't drive the nail into wood!* We need only make small indentations, about 1/8", to mark the screw locations, and to provide pilot holes for the wood screws.

☐ Remove the paper pattern guide.

☐ Find one of the eight Phillips head screws and a Phillips head screwdriver.

☐ At each of the eight screw hole locations marked by the finishing nail, screw the Phillips head screw fully into the board, stopping when its head reaches the board surface. Make sure the screw remains perfectly straight, and not tilted, when doing each of the eight holes.

☐ Return the screw to the parts bag.

☐ Find the 5-inch length of PVC pipe. The tuning coil will be wound on this section of plastic pipe. It is called a *coil form*.

☐ There are six holes drilled through the pipe, look for the two outer, and larger sized, mounting holes.

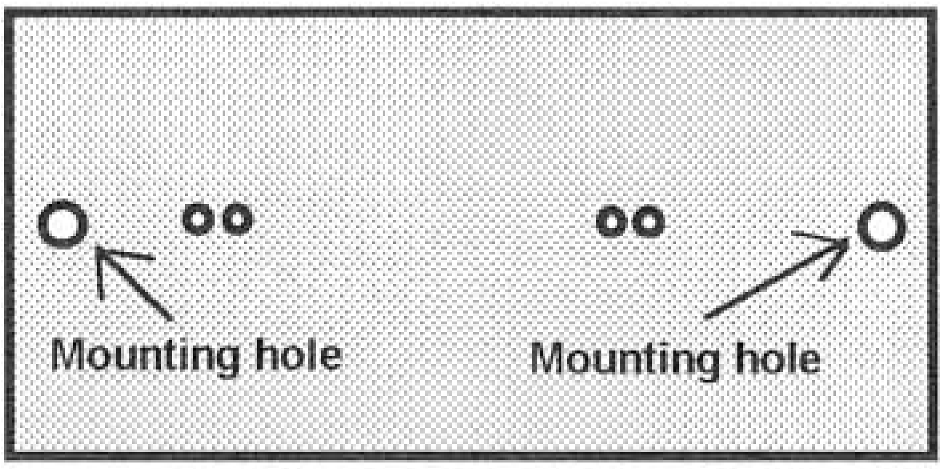

☐ Now, make sure that the two screw pilot holes on the pine board for mounting the coil form are correct. Refer to the drawings in the manual that show the placement for the Fahnstock clips and coil form. Temporarily mount the coil in its proper location using two Phillips head screws. You do not need to fully tighten the screws, we are just checking to see that things fit properly.

Important Note: You will need to use a screwdriver with a long shaft (about three inches) for these two screws because of the large diameter of the coil form.

☐ If the coil form mounting holes align properly with the breadboard's mounting holes, you may remove the coil and return the screws to the parts bag. If things don't match up just right, you will have to mark the correct position for the breadboard pilot hole with a pencil, and remove the coil. Make a new pilot hole by using the hammer and finishing nail to make a small indentation to start the screw. Run a screw into the pilot hole until the head reaches the board surface. Remove the screw and return it to the parts bag.

This completes Phase 1. There should be eight screw holes in the pine breadboard at this point. If so, the mounting holes have been properly located, marked, and the screw holes are ready for parts installation!

If you haven't already done so, this is an excellent time to stain the board (if you are planning to do so).

We are ready to begin Phase 2.

Phase 2: Winding the tuning coil

This portion of the assembly involves winding about 66 turns of 18 AWG enamel coated copper wire over a 2½" area of the PVC plastic coil form. This will take time, patience, planning and practice! Here are a few suggestions to help you along.

You will be directed to unwind a length of wire from the 100-foot coil of enamel wire in several of the following steps. Always unspool the wire, never allow it to uncoil in Slinky™ toy fashion. The unspooled length of wire should be laid out in a straight line—*do not* allow it to simply pile onto the floor, or you will end up with a hopeless tangle of wire. After unwinding a length of wire, use two strips of masking tape to secure the remaining wire on the coil.

The winding on the coil should be smooth and even—with no gaps or kinks. If the copper wire kinks during handling, smooth it out as best you can before trying to wind it. We are going to wind a single-layer coil, this is called a *solenoid* winding. The winding is done in several steps. You will be instructed to stop winding after a certain amount of wire is wound. If you have a friend who is willing to help, ask him to give you a hand winding the coil. Four hands can be very helpful!

☐ Remove several 5-inch lengths of masking tape from the roll of tape. Keep them within arm's reach while winding the coil.

☐ Find the PVC coil form and the 100-foot hank of coil wire.

☐ Unspool about 15' of wire from the 100-foot coil of wire. Leave wire attached to the spool—*do not cut it!* Tape the remaining wire so it will not unwind and become tangled.

☐ First, secure the wire at the starting point of the winding. Insert about 16" of wire into the starting hole, and loop it back out of the adjacent wire hole and pull tightly. Either end of the form may be used as the starting point. This will secure the wire. The illustration below shows how to do this:

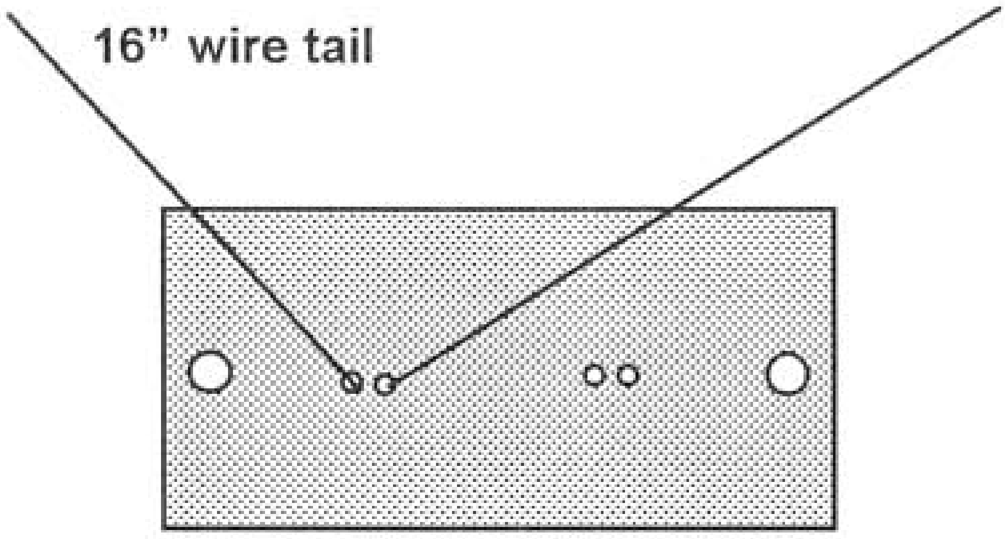

☐ Tuck the 16" wire tail back inside of the coil form, so it will be out of the way during winding.

☐ Hold the coil form in your left hand. The wire coming from the 100-foot coil should be between you and the form. Insert the middle and index finger of your left hand into the coil form. Place your left thumb over the wire at the starting hole.

☐ Use your right hand to help turn the PVC coil form while winding the coil. This takes a little practice, but you will quickly master the technique once you start.

☐ Wind several turns of wire by turning the form. Use your left hand thumb to guide the wire, and keep enough pressure on the wire with your thumb so that the windings are tight, even, and spaced together.

☐ After several turns are completed, examine your work. (Keep your thumb in place so the wire doesn't loosen on the coil form!) The wire windings should be even and spaced so the wire of each turn is firmly pressed against the next. The windings should be tight. There should be no gaps between the wire turns. You can squeeze the windings together to remove small gaps.

☐ The first several turns of the winding form the foundation for the rest of the windings. If you didn't quite get it right the first time, unwind the coil and try again.

☐ Once you are satisfied with your work, grab a piece of masking tape with your right hand, and tape across the windings to hold them securely. *Note:* Tape across and over the windings—the tape should follow the 5" width of the coil.

☐ Using its dispenser tip, put a small drop of contact adhesive on the winding at its beginning and end point. Smear any excess back over adjacent windings using the dispenser tip. Draw a very light bead of adhesive across the full width of the winding. Very little adhesive is needed for these tasks! Let the coil sit until the contact cement has fully cured! (You may use the glue sold for assembling plastic models for these tasks. Remember, it will take several hours for the glue to harden.)

☐ Now, finish winding the tuning coil. It takes about 3 feet of wire to do four full turns. It takes about 66 turns of wire to fully wind the coil over the 2.5" coil span; so you will use about 52 feet of wire for the entire winding. Don't bother trying to count turns as you go. A few extra or missing turns will not affect the performance of the crystal set. What is important is to have the winding cover the full 2.5" inch span, without gaps.

☐ Unspool another 20 feet of 18 AWG enamel coated wire. Remove the masking tape from the coil form. Continue winding until the 20 feet of wire is used. Tape the new windings to hold them in place.

☐ Again, inspect the new windings you have just completed. Are they tightly wound, and free of gaps between adjacent windings? Squeeze the newer windings towards the starting windings to close up any gaps. When you're satisfied, use masking tape to secure the turns, and then fix the windings in place using contact cement or model glue.

Important Note: As you continue adding windings, you will reach a point where your left hand fingers will no longer reach inside of the PVC coil form. At this point, you will need both hands on the form as you work, and you will be using both thumbs to guide the wire and keep the proper tension.

☐ Continue by unspooling 20 foot lengths of wire at time, winding the coil until you reach the "finish" point on the coil form. Securely tape the coil windings, and glue the new windings in place.

☐ At the end of the winding, mark off an additional 16" of wire, and cut the wire at this point.

☐ Insert the 16" inch wire tail into the "finish" wire hole in the PVC coil form. Bring the wire back out through the adjacent hole and pull with enough force so the wire is snug. Secure the end point of the winding with a very small drop of contact cement.

☐ Fish the wire end for the starting point of the winding out of the form. Your coil should now resemble the illustration below:

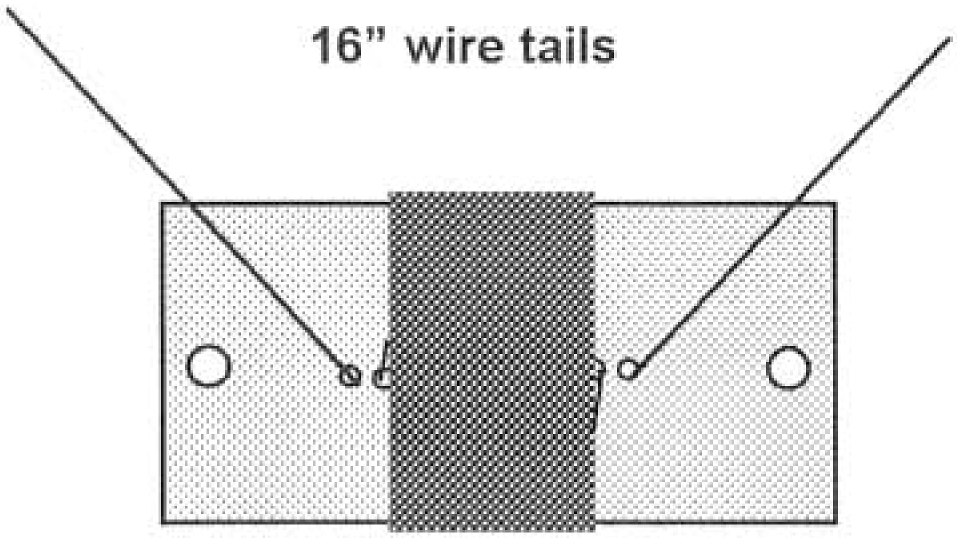

☐ Rewind any excess wire back onto the supply spool.

Congratulations! You have just wound an inductor that has about 217-uH of inductance! Put the coil aside for now, you'll be mounting it on the breadboard soon.

Phase 3: Mounting the parts

☐ Locate the six Fahnstock clips, and six of the eight Phillips screws supplied with the kit.

☐ Refer to the positioning guide below for the following steps.

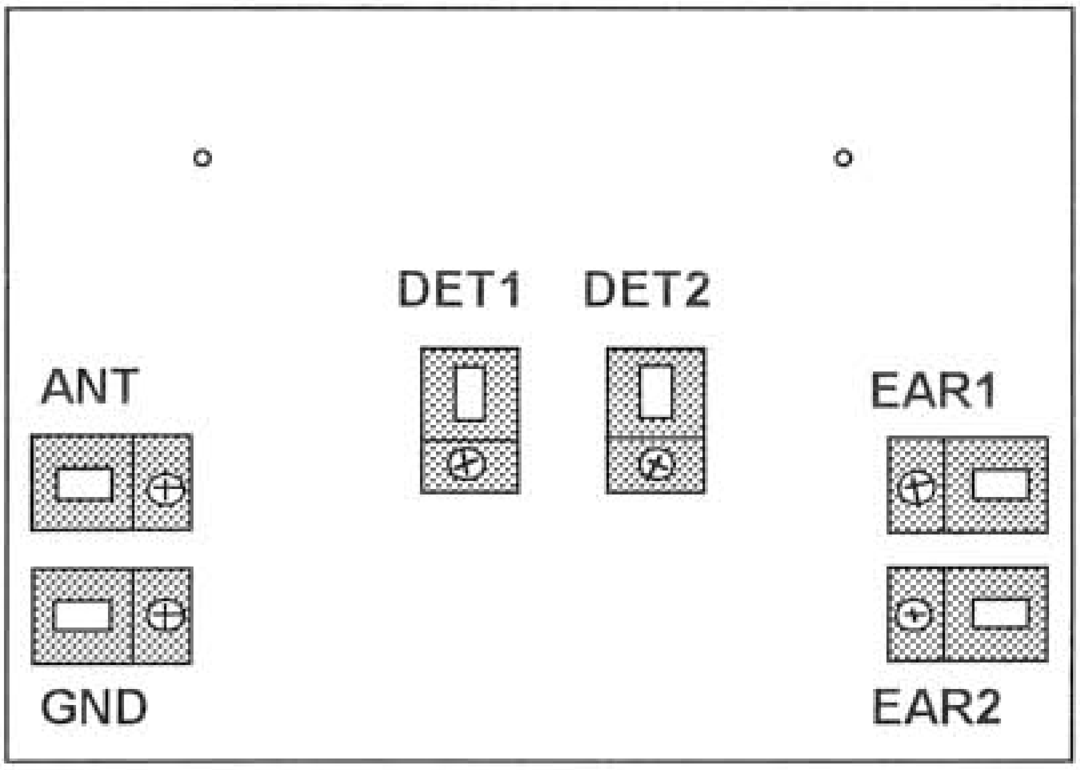

In the following several steps, we will mount the six Fahnstock clips.

☐ Place a Fahnstock clip at the point shown as *GND* in the drawing. Align the clip so its rounded over edge faces the board edge.

☐ Secure the clip to the board using one of the Phillips head screws. Tighten the screw until snug, so it holds the clip firmly. *Do not over tighten.*

For the remaining steps, align the Fashnstock clips as shown in the drawing.

☐ Mount a second clip at the *ANT* position on the breadboard using a Phillips screw.

☐ Mount a clip using a Phillips screw at the *DET1* position on the breadboard.

☐ Mount a clip using a Phillips screw at the *DET2* position on the breadboard.

☐ Mount a clip using a Phillips screw at the *EAR2* position on the breadboard.

☐ Mount a clip using a Phillips screw at the *EAR1* position on the breadboard.

The six Fahnstock clips should all be mounted at this point.

☐ Find the coil and the remaining two Phillips head screws.

☐ Mount the coil to the board using the two Phillips head mounting screws. Before tightening the screws, make sure the two 16" coil lead wires are oriented towards the front of the breadboard. You may now tighten the screws so the coil is mounted snugly against the board. *NOTE: Over-tightening the screws may damage the coil form, or cause the screws to strip the pine board.*

Your crystal receiver should resemble the following drawing:

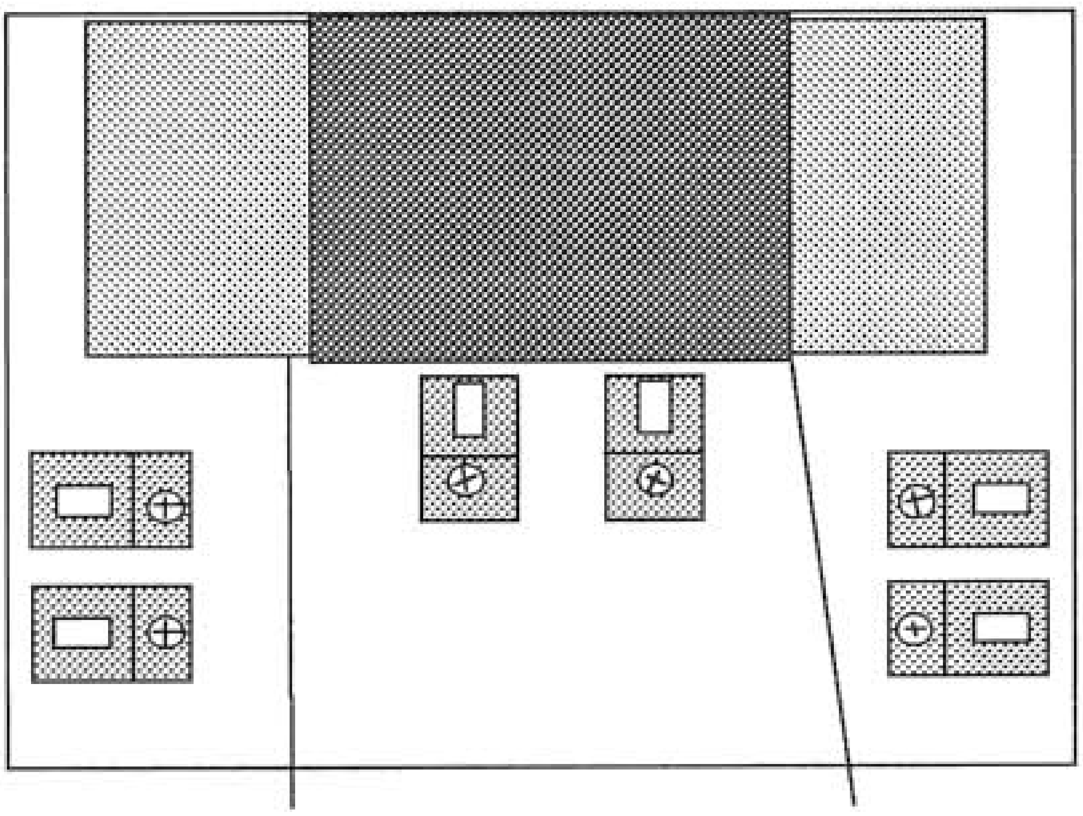

You have finished Phase 3 of your crystal set receiver project. In Phase 4, we'll be doing the wiring between the crystal set's components.

Phase 4: Wiring the crystal set

In phase 4, most of the wiring will involve using the two 16" coil wires. For clarity, we are only showing the components we are going to be working with.

You need to know where the wire leads need to be sanded to make connections. Route the coil wire, as shown in the following diagram, by loosening the Phillips screws mounting the two Fahnstock clips labeled ANT and DET1 (as shown):

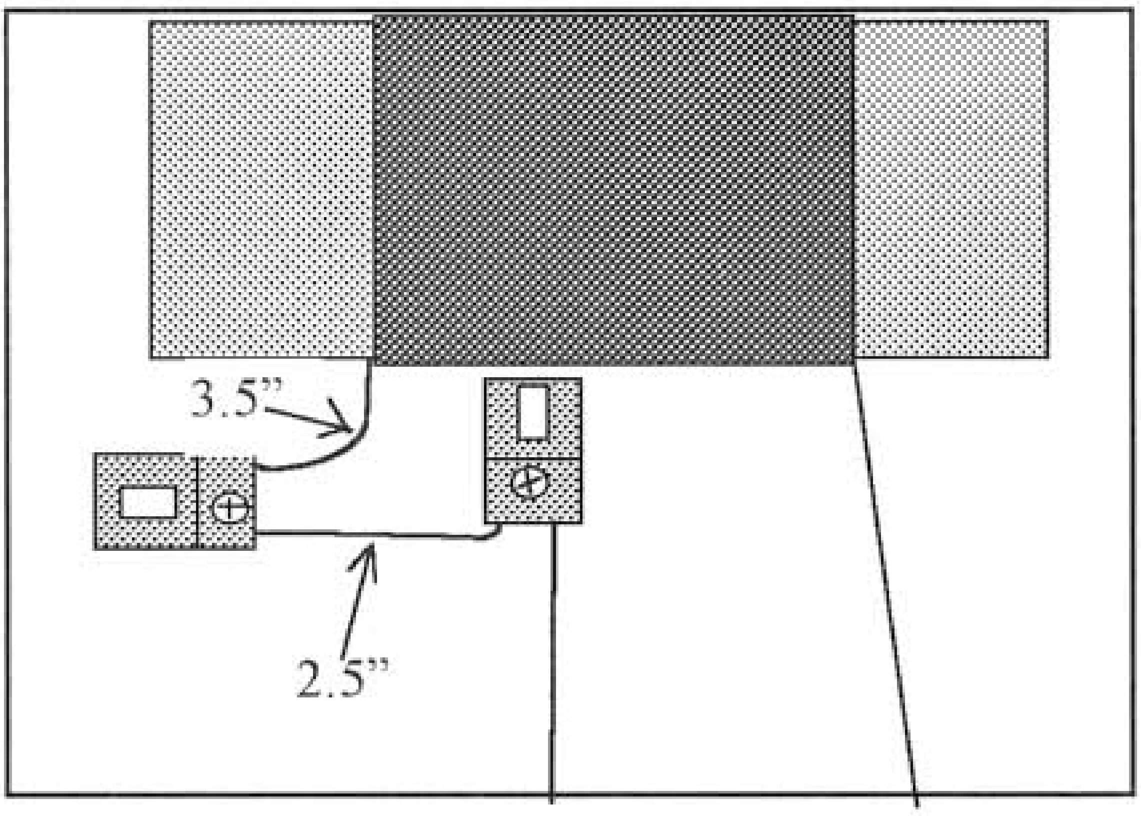

☐ Measure and mark the coil wire lead at the left-hand side of the crystal set at 3.5 inches. The measurement is taken from the point where the wire exits the form.

☐ Remove about 1" of insulation at the 3.5" mark.

☐ Form the wire so the area stripped of insulation forms a *through connection*.

☐ Loosen the Phillips head screw for the *ANT* Fahnstock clip.

☐ Route the *through-connection* loop so it passes between the *ANT* Fahnstock clip and breadboard.

☐ Tighten the *ANT* clip mounting screw so it is snug.

☐ Measure and mark the wire lead emerging from the antenna clip at 2.5".

☐ At the 2.5" point, remove 1" of wire insulation, and form a second *through-connection* half loop.

☐ Loosen the Phillips screw mounting the *DET1* Fahnstock clip.

☐ Route the second through connection so it passes under the *DET1* Fahnstock, loops around the screw shank, and exits as shown in the drawings.

☐ Tighten the Phillips screw for the *DET1* clip so it is snug.

☐ Several inches of wire should remain from the *DET1* clip connection, this will be used later. ***Do not cut the wire until told to do so!***

In the next few steps you will be working with the right-hand coil wire. Refer to the drawing below for the following steps:

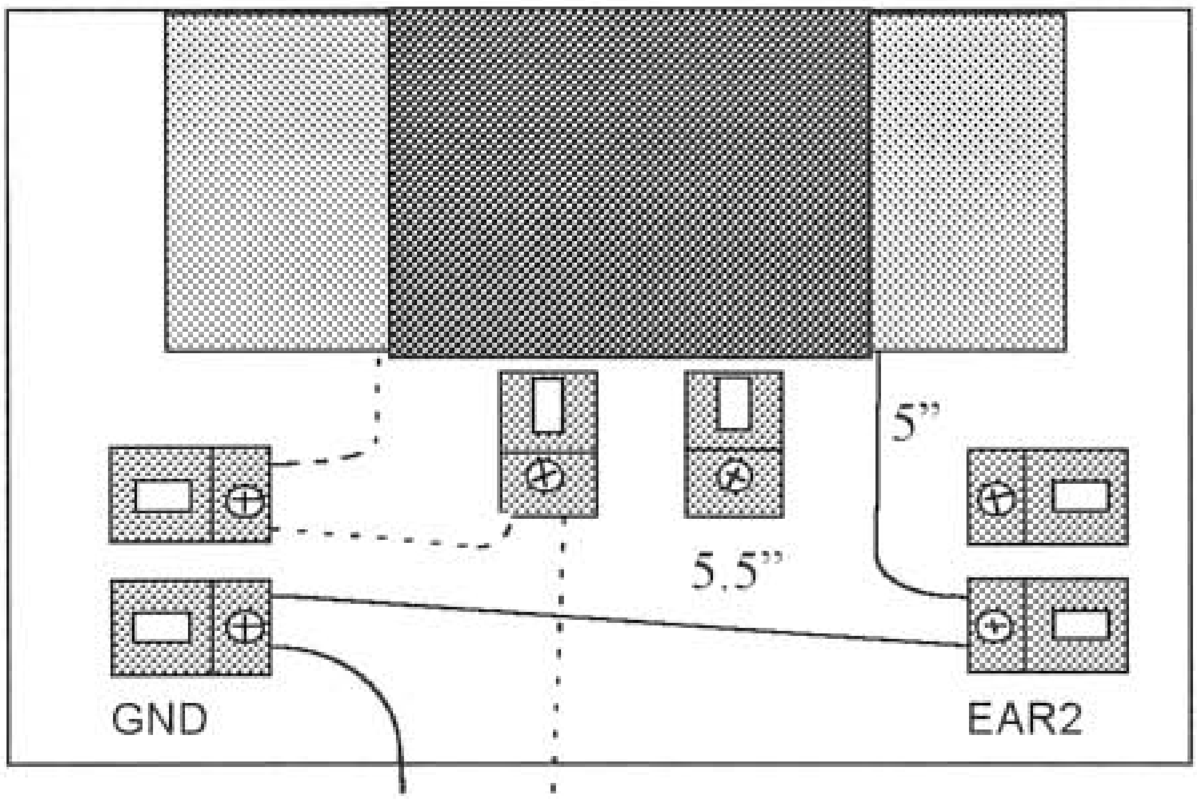

Important Note: The "dotted" lines show wiring done in previous steps.

☐ Mark the left-hand coil wire at 5" from where it exits the coil.

☐ At the 5" point, remove 1" of insulation from the wire.

☐ Form the area stripped of insulation into a *through-connection* half loop.

☐ Loosen the Fahnstock clip mounting screw for the *EAR2* clip.

☐ Pass the *through-connection* between the *EAR2* clip and breadboard, with loop going around the Phillips screw shank, as shown above.

☐ Tighten the mounting screw for the *EAR2* clip until snug.

☐ Measure the wire exiting from the *EAR2* connection, and at the 5.5" point remove 1" of insulation from the wire.

☐ Form the area of wire stripped of insulation into a second *through-connection* half loop.

☐ Loosen the Phillips screw that mounts the Fahnstock clip used for the *GND* connection.

☐ Pass the *through-connection* between the *GND* clip and breadboard, with the loop going around the Phillips screw shank, as shown above.

☐ Tighten the GND Phillips mounting screw until snug.

☐ Several inches of wire should remain from the *GND* connection. ***Do not cut this wire until instructed to do so!***

This completes Phase 4. You've come a long way since first opening the kit. Your little crystal receiving set is really starting to look like something.

Phase 5: Final wiring

Only a few more steps remain, and the set will be finished.

Refer to the drawing below for the last wiring steps.

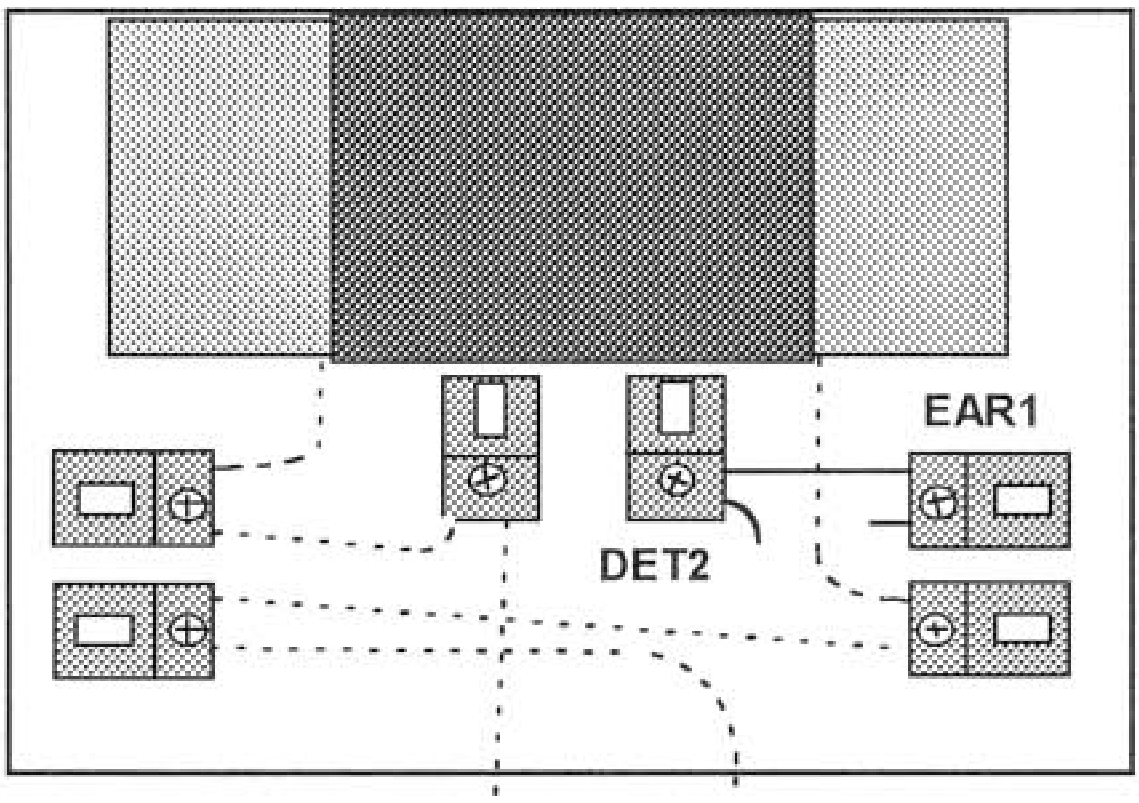

☐ Cut a 4½" length of 18 AWG enamel wire from the wire spool in the kit.

☐ Remove 1" of insulation from each end.

☐ Form each end of the wire into an *end* connection.

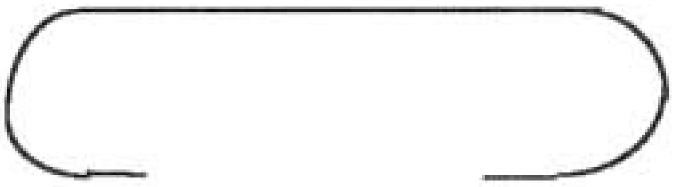

☐ Loosen the screws mounting the Fahnstock clips *DET2* and *EAR1*.

☐ Take one of the wire *end* connections and place between the *DET2* clip and breadboard.

☐ Tighten the mounting screw for the *DET2* clip until snug.

☐ The remaining *end* connection is placed between the *EAR1* clip and breadboard.

☐ Tighten the mounting screw for clip *EAR1* until snug.

☐ The wire should be routed so it crosses over the wire from the right end of the coil without the wires touching each other.

☐ Find the 27K ohm resistor (it has the Red-Violet-Orange-Gold colored bands) and form the leads as shown below:

Important Note: It makes no difference which lead is used for either connection.

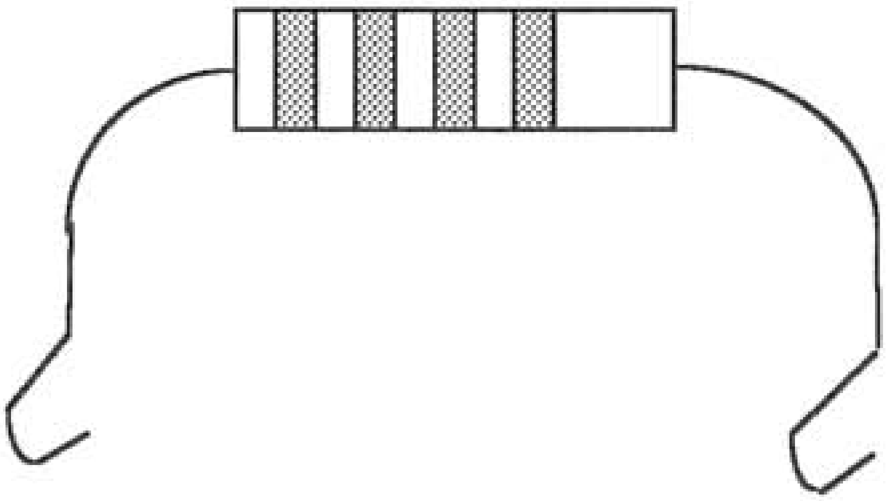

☐ The ends of the resistor leads are formed into little half loop hooks.

☐ Loosen the two Phillips screws for *EAR1* and *EAR2* clips.

☐ Put one of the half loop connections between the *top* of the clip body, and the *bottom* of the screw head for the *EAR1* clip. Adjust the size of the loop so it fits snugly around the shank of the screw.

☐ Tighten the mounting screw for clip *EAR1* until snug. The resistor lead should remain firmly attached.

☐ Place the remaining loop between the screw head and body of the clip *EAR2*. Re-form the loop so it fits snugly around the screw body.

☐ Tighten the screw for clip *EAR2* until snug. The resistor lead should remain firmly attached.

☐ Find the germanium detector diode. Form the leads as shown below.

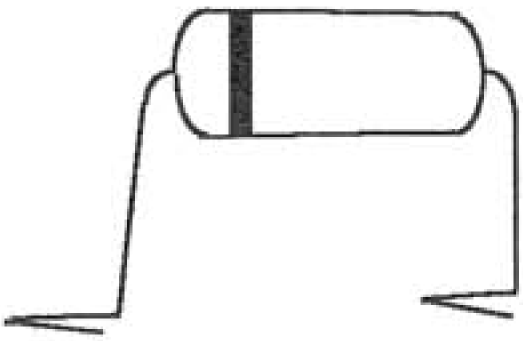

☐ A 3/8" length of each of the diode's lead ends is sharply folded back on itself. This is done because the diode will be mounted in the *DET1* and *DET2* Fahnstock clips. The diode lead wires are a very small diameter, and may not be held tightly enough by the clip unless the leads are made a bit thicker.

☐ Follow the diagram below, and clip the diode's anode lead into the *DET1* Fahnstock clip, and the cathode lead (the lead with the band close to it) into the *DET2* Fahnstock clip. Gently pressing down on the clip opens the jaw, permitting the wire lead to be inserted.

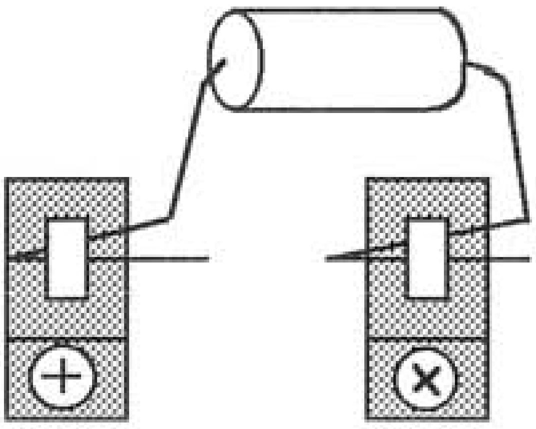

Find the tuning capacitor. Refer to the drawing below for the following steps.

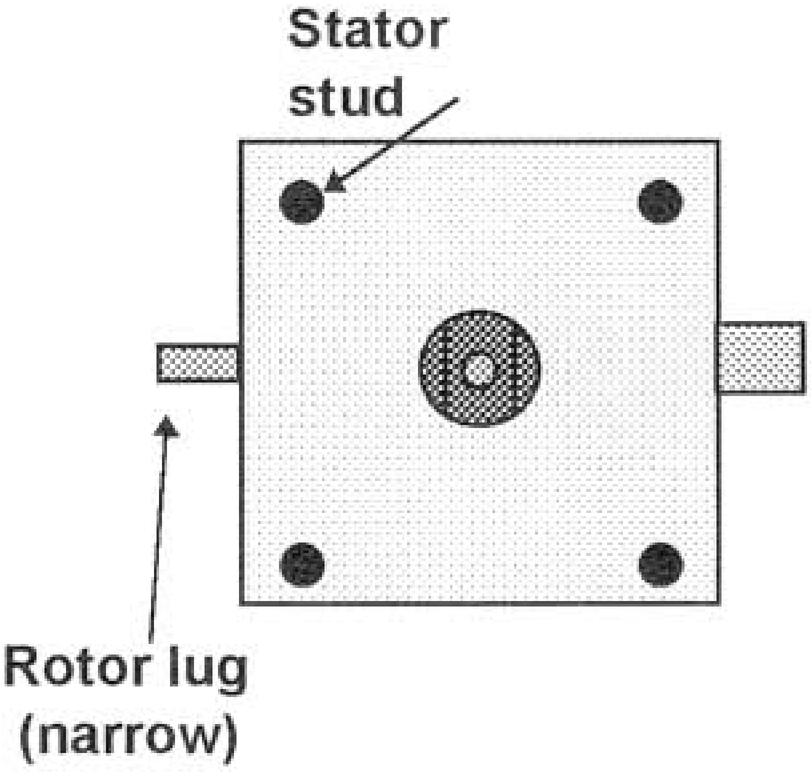

This is the correct orientation for installing the capacitor. Note that the narrower rotor lug must be at the left hand side!

Variable capacitors have two sets of plates. One set of plates remain stationary, and are called the "*stator*" plates. The tuning shaft is attached to another set of plates that rotate as the shaft is turned. They are called the "*rotor*" plates. As the shaft is turned, the plates are either meshed or unmeshed, changing the capacitance value. This is how the tuning capacitor works.

The tuning capacitor will be mounted at the front center of the breadboard using the double-sided foam tape. The remaining enamel wire lead coming from the *GND* Fahnstock clip will be connected to the narrow *rotor lug* terminal on the tuning capacitor. The remaining enamel wire lead from the *DET1* Fahnstock clip will be connected to the round *stator stud* on the front of the capacitor.

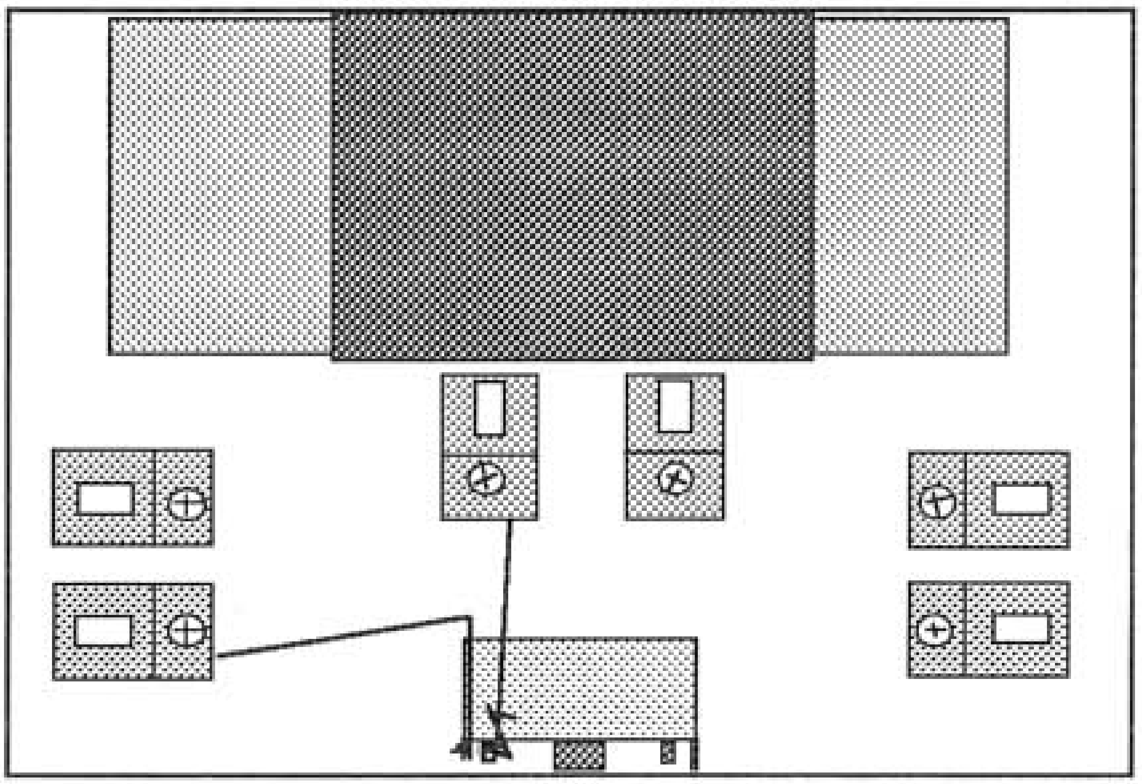

The above drawing shows how the wires will be routed to the tuning capacitor connections.

Connecting the stator lead

- ☐ Temporarily place the tuning capacitor as shown in the drawing. ***Be sure the narrowest stator lug is at the left hand side!***
- ☐ Cut the free enamel wire lead from the *GND* clip to a length of five inches.
- ☐ At the end of the wire, remove 2" of insulation.
- ☐ Route the enamel wire from the *GND* clip so it reaches to the capacitor body. At that point, the wire goes up to the middle of the capacitor, and then makes another sharp bend towards the front of the board.

Study how the capacitor is made. If you look into the capacitor through its clear plastic housing, you can see how the narrow rotor lug continues to the rear of the capacitor. At that point, there is a tiny opening in the plastic capacitor housing.

- ☐ Take the wire from the *GND* clip and insert it into this capacitor opening. Carefully feed the wire so it passes through the capacitor body (it will be between the plastic housing and narrow metal strap) and finally emerges at the front end of the capacitor.
- ☐ Continue feeding the wire until 1" of bare enamel wire clears the front opening in the capacitor.
- ☐ Using your long nose pliers, carefully wrap the enamel wire around the narrow rotor lug until two or three full turns are made. Trim off the excess wire.

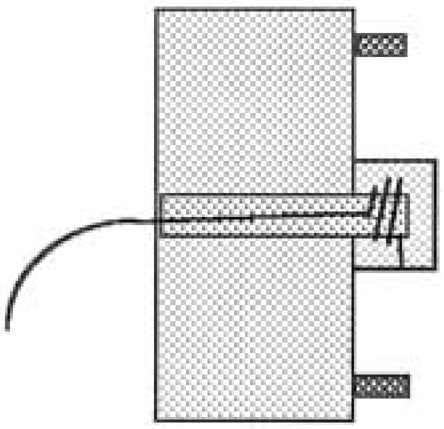

- ☐ Gently, but firmly, squeeze the turns wrapped around the rotor lug, using the long nose pliers, until they are compressed tightly--making a good, tight electrical connection.

Important Note: If for some reason the narrow rotor lug is damaged while trying to attach the lead, use the wider rotor lug as an alternate connection point.

Making the stator connection

☐ Temporarily place the capacitor at its mounting location.

☐ Route the free enamel wire coming from the *DET1* Fahnstock clip so it follows the breadboard, and comes up and over the top of the capacitor until it reaches the stator stud. (Refer to the drawings to identify the proper stud.)

☐ Trim the wire coming from the DET1 clip to a length of 4½".

☐ Remove 1½" of insulation from the wire end.

☐ Wind the wire over the stud to form three *tightly* wound turns. See the following drawing. Use enough wire so when finished you have a small ½" pigtail remaining.

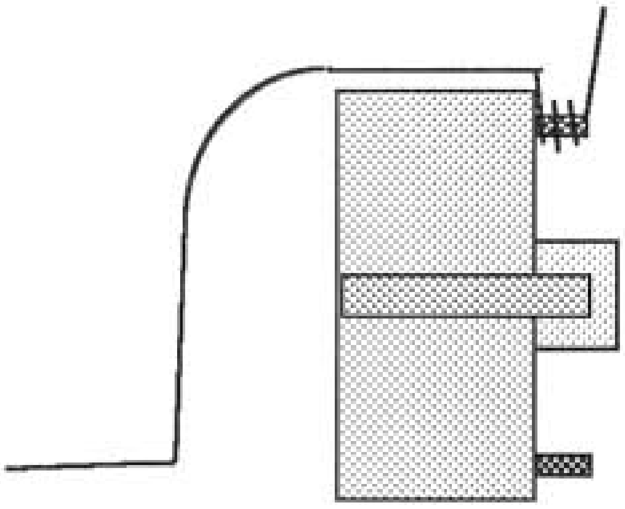

☐ You need a good, tight electrical connection. Slide the turns off of the stud—note how they make a little coil. Gently tighten the coil by twisting the coil so it is smaller.

☐ The coil should slide back on with some difficulty—otherwise the turns are to loose.

☐ To make an even tighter connection, take the pigtail, and wrap it back over the wire in a half-loop.

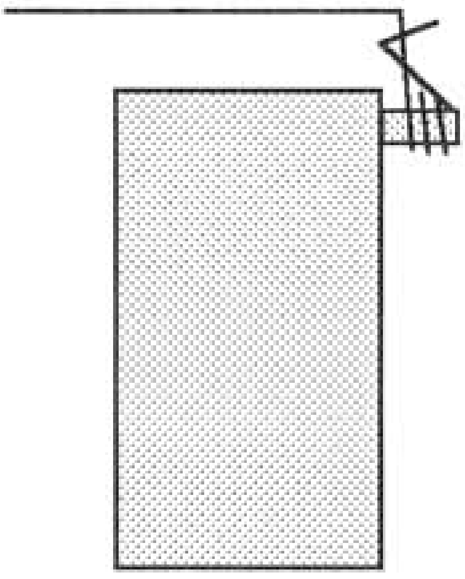

☐ Using the long nose pliers, gently twist the half loop to take up wire slack and tighten the wire wraps made over the stud.

Important Note: If the stator stud is damaged beyond use, the stud located directly beneath it may be used as a backup. The two right hand studs may not be used.

☐ Place the capacitor at its mounting location at the center front of the board. The body of the capacitor should sit about 1/8" back from the board edge. Use a ruler to center the capacitor body at 3" from either outside board edge.

☐ Mark the capacitor body outline by tracing its outline onto the breadboard with a pencil.

☐ Find the double-sided foam tape.

☐ Remove one of the protective covers from one side of the tape. Place the sticky side over the capacitor outline on the breadboard. Firmly press the tape to set the adhesive.

☐ Remove the protective cover over the top of the tape.

☐ While being careful to keep the capacitor centered on the board, and parallel to the front edge of the board, press it to the foam tape. Apply gentle, but firm, pressure to set the adhesive.

☐ Find the ¼" diameter aluminum tuning shaft extension. Align the threaded screw end with the opening in the capacitor tuning shaft. Screw the extension into the tuning capacitor.

☐ Unscrew the tuning shaft about 1 or 2 turns, to open a gap between the two shafts.

☐ Using the dispenser, allow a small drop of contact adhesive to flow into the gap and onto the threaded portion of the extension. *Quickly* screw the extension back into the tuning capacitor to close the gap. If you are using model airplane cement, unscrew the shaft extension, and lightly smear a small amount of glue over the threads. Quickly screw the extension back into the capacitor shaft.

☐ Allow time for the glue to harden.

☐ Using a small screwdriver or hex tool, install the tuning knob on the shaft extension.

This completes the construction of the VEC-121K crystal set receiver.

TESTING AND ALIGNMENT

1. Turn the capacitor shaft in a counter clockwise direction until the stop is reached.
2. Loosen the tuning knob set screw, and align the tuning indicator on the knob so it corresponds to the 9 o'clock position.
3. Tighten the set screw.
4. Verify that the knob indicator properly corresponds to the capacitor stops at the 9 and 3 o'clock positions.

OPERATING INSTRUCTIONS

1. Use the hookup shown in the following diagram to make the antenna (aerial), ground, and earphone connections.

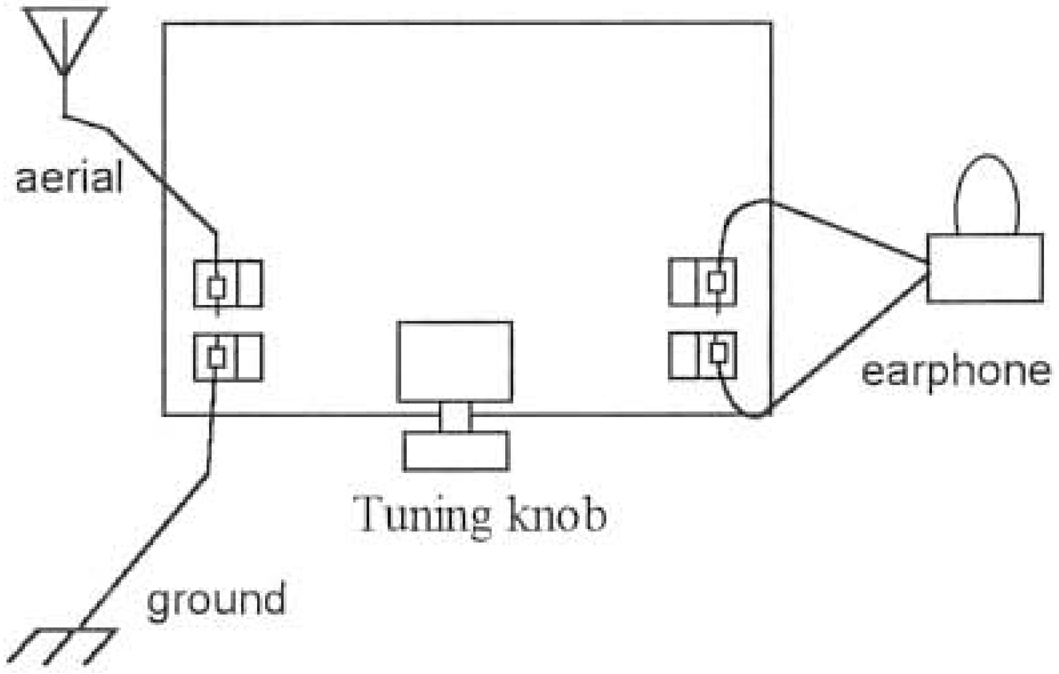

2. Connect the earphone wires to the Fahnstock clips at the right hand side of the radio.

3. Connect a good ground connection to the crystal receiver. Use the 7 foot length of stranded hookup wire supplied with the kit to make the ground connection between a good earth ground and the Fahnstock clip on the receiver.
4. Connect a long wire antenna to the Fahnstock clip on the receiver.
5. Carefully tune the tuning knob while listening for signals in the earphone. You should be able to copy at least one, and maybe several, stations without difficulty.

What can I hear? You should be able to hear local stations without much difficulty. The crystal set tunes the entire broadcast band from 530 kHz to 1700 kHz. The antenna and ground system are the keys to success when using crystal sets. Where you live affects how well you receive signals. If you live in a salt marsh, or along the shore, you may experience exceptional reception. The fellow living in an arid area with poor soil moisture may have much poorer reception.

When to listen? You should hear local stations at any time of the day. At dusk, radio signals propagate over greater distances. Listening just before the sun goes down may yield some pleasant surprises! Also, as the seasons change, you may find you can hear different distant stations at sundown. Keep a simple log of new stations when you receive them. Note the time, call letters and frequency. It's always fun to add a "new one" to the list.

Erecting a good antenna

Your radio will work only as good as the antenna and ground it is connected to do. The best antenna for a crystal set is located outdoors, it is between 100 and 30 feet long, and is mounted high and in the clear. Your kit includes enough wire to make an outdoor 40 foot wire antenna and short lead-in wire to the crystal set receiver.

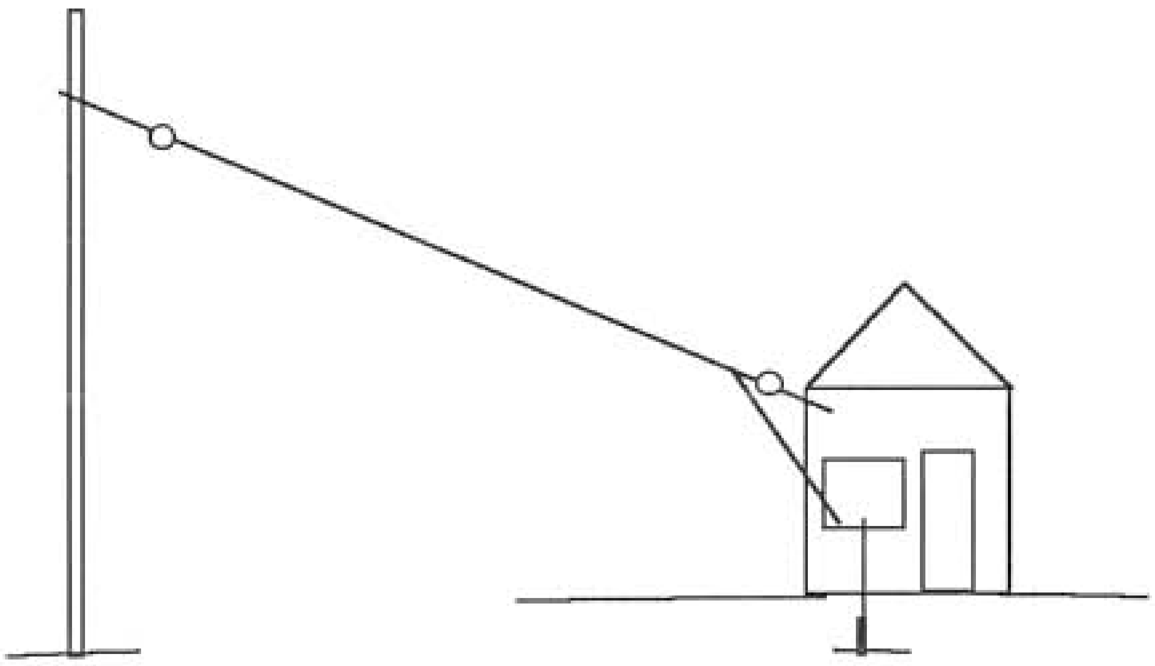

This diagram shows an aerial going between a pole (or tall tree trunk) and location where the receiver is located. A small length of lead-wire connects the antenna to the crystal set inside of the house. (Make sure the lead-in wire makes good electrical connection to the antenna wire!) Note the insulators between the antenna wire and the supporting structures. A good earth ground is provided by a short ground rod driven into the earth outside of the window where the crystal set is located. You may make your own insulators from scraps of 1" plastic pipe, or small sections of 1" by 6" Lexan™ or Plexiglas™ . Drill holes at the ends of the insulators for tie-points for the antenna wire and support rope. Clothesline is a good way to tie the insulator to the supporting structures.

The antenna wire should not be allowed to touch leaves or rub against branches. Doing so will affect your reception. Keep your antenna wire away from power lines.

Finding the elusive ground

Several feet of insulated stranded hookup wire is included in your kit. This is provided to connect a *ground* to the crystal set. But, what is a good *ground* and how can you make or find one?

Ground rods. Many times a crystal set will work quite well without a ground connection, especially if a very good antenna is being used. The most effective ground is a 10 foot ground rod driven into the earth. If you have a TV antenna on your home, it should be connected to a ground rod for lightning protection. If it is near your operating position, you may be able to tie a wire between your crystal set and the TV antenna ground rod. Even four or five feet length of old copper pipe driven into the earth will do in a pinch! But, be sure you use some sandpaper to remove any corrosion on the pipe that would prevent making a good electrical connection to your ground wire lead in.

Water pipe grounds. If you live in an older home that uses metal instead of plastic plumbing, you have a good ground point at the nearest water pipe or radiator. Connect a short wire between the crystal set ground and the water pipe. Remember that the pipe must be free of corrosion or paint to make a good electrical connection.

Counterpoise grounds. The reason the crystal set needs a ground is because a single wire antenna is really only one-half of a complete antenna! The ground makes up for the missing half of the antenna, and gives the wire antenna something to "work against". If you have 50 or 60 feet of scrap wire laying about, you can make what is called a "counterpoise" to serve as a ground for the crystal set. Simply run the wire out on the ground in a straight line. The wire needn't be buried, it will "couple" to the earth due to "stray capacitance".

Connect one end of the counterpoise wire to the ground Fahnstock on the crystal set, and you're in business!

IN CASE OF DIFFICULTY

This crystal set almost *has* to work, if it was built according to the directions. If you can not hear any stations at all, go back over the assembly instructions and try to find any errors that you might have made during construction. Perhaps you missed removing the insulation from the wire at a connection. Or, a wire is in the wrong place. Have a friend check your work. It's very easy to look over a mistake, and a fresh pair of eyes will often quickly find mistakes that elude your detection. Factory repair is available for assembled kits that do not work. Consult the warranty for information regarding factory service.

Here's a quick check list of common problems.

The set receives several signals at once. This is normal for crystal sets. They don't have the ability to separate signals that very are strong or close in frequency.

The tuning control has no effect. Check that the connections to the tuning capacitor are tight, and that the proper connection points to the tuning capacitor were used.

Signals are very weak, or I can only hear one station. Again, this may be normal, if there are no strong signals in your area. Check your ground and antenna connections. A longer antenna or better ground may improve things.

Set works intermittently. This sounds like a poor connection. Check that the Fahnstock clips are making good connections to the antenna, ground and earphone wires. If a clip won't grip the wire properly, remove the wire and bend the clip upwards slightly. This should restore the clasping action of the clip's jaw.

There is a local station at the high or low end of the dial that I can't hear. The VEC-121K is designed to tune the entire broadcast band. However, there will be certain antenna lengths that will behave in unexpected manners at certain frequencies. Also, because of the way the antenna and ground are coupled to the receiver tuned circuits, they will also have some effect on the tuning range limits.

THEORY OF OPERATION AND SPECIFICATIONS

The tuning capacitor and coil form a parallel resonant circuit that tunes from about 530-kHz to over 1.6 MHz. The antenna and ground are directly applied across the resonant circuit. Thus, the set is optimized for use with short, high-impedance antennas under 100' long. The parallel resonant circuit acts like a very low impedance path to signals off resonance. (Signals that are not tuned in see a "short circuit".) At resonance, the parallel circuit becomes a very high impedance (offers little resistance to a signal it is tuned too).

Signals that are "tuned in" are "rectified" by the germanium detector diode. The radio signal is simply a high-frequency AC voltage. During rectification, better called detection, the signal is broken down into two components: a DC voltage and a AC signal that corresponds to the information being carried by the Amplitude Modulated signal. The AC signal maybe voice, music, or whatever program material the radio station is broadcasting.

Because the audio recovered from the radio station AM carrier is a very, very, small voltage, the ear piece must be very sensitive. The high-impedance crystal earpiece supplied with the VEC-121K meets those requirements. Because the crystal earpiece has almost no resistance, a 27K ohm resistor is placed across the earphone clips to provide a path for the detector diode current.

Tuning Range	530-1700 KHz (typical)
Detector diode	Fixed germanium, 1N34A or equiv.
Headphone impedance	2000-ohms or higher
Mode	AM receiver

SCHEMATIC

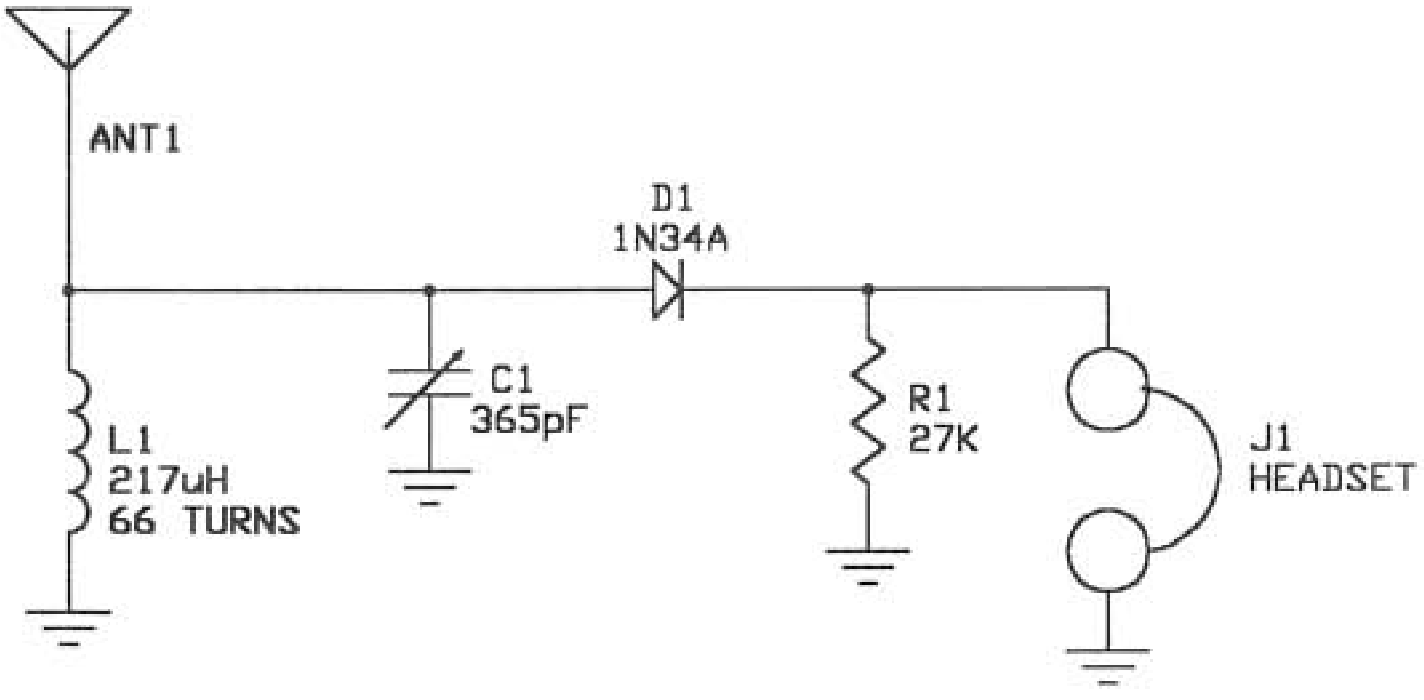

VEC-121K

6 50619 00023 8

High-performance electronic kits . . . fun to build and use!

Kit building is a super fun way to spend a quiet evening or weekend. You'll find it extremely satisfying to build your own electronic equipment. You'll have a useful electronic gadget that you can show off once you're through. You'll cherish it for years because *you built it yourself.* From shortwave converters to aircraft receivers and ham radio kits to an old fashioned crystal radio kit, you'll find many fun items in the VECTRONICS kit line for you.

VECTRONICS kits work! They're created by engineers who are hobbyists-at-heart to give you what you want -- a professional product at a hobby price. Each kit features a professional quality epoxy glass PC board with solder mask and screen printed component legend, simple step-by-step instruction manual and the highest quality components. Kit assembly is easy, and they work the first time.

Don't forget about our custom cabinets -- they turn your kit into a fabulous show piece that your friends won't believe that *you* built.

With VECTRONICS kits you get satisfaction, relaxation, and a super-fun product you'll be proud to use . . . *because you built it yourself!*

VECTRONICS has a worldwide reputation of building the finest quality amateur radio products made. You can trust our 30 years of experience to deliver super quality, high-performance kits.

All VECTRONICS electronic hobby kits are designed and kitted in the USA . . . *and built by you!*

OTHER *VECTRONICS* hobby KITS:

VEC-101K
Shortwave Converter
VEC-131K
Aircraft Receiver
VEC-201K
CW Keyer
VEC-221K
CW Memory Keyer
VEC-412K
Fast Battery Charger
VEC-422K
SCA Decoder
VEC-820K
CW Filter
VEC-821K
Super CW Filter

VEC-830K
Super SSB Audio Filter
VEC-841K
Tunable CW Audio Filter
VEC-920K
20 M QRP Amplifier
VEC-930K
30 M QRP Amplifier
VEC-940K
40 M QRP Amplifier
VEC-980K
80 M QRP Amplifier
VEC-1002K
2 Meter Receiver
VEC-1006K
6 Meter Receiver

VEC-1010K
10 Meter Receiver
VEC-1120K
20 Meter Receiver
VEC-1130K
30 Meter Receiver
VEC-1140K
40 Meter Receiver
VEC-1180K
80 Meter Receiver
VEC-1202K
2 Meter FM Transmitter
VEC-1220K
20 Meter Transmitter
VEC-1230K
30 Meter Transmitter

VEC-1240K
40 Meter Transmitter
VEC-1280K
80 Meter Transmitter
VEC-1290K
AM Radio Transmitter
VEC-1294K
TV Transmitter
VEC-1402K
2 Meter Preamplifier
VEC-1422K
220 MHz Preamplifier
VEC-1444K
440 MHz Preamplifier
VEC-1402DK
Super 2 Meter Preamp

A LADYBIRD BOOK

making a Transistor Radio

A crystal set – the tuned circuit

A crystal set, the simplest form of radio receiver, is the first stage of our radio. It can operate entirely on its own as a radio receiver, even if no further stages are added.

As we have read on page 6, an aerial picks up all the radio signals (the radio frequency carrier waves) in its path. But we need only signals of a certain frequency—those sent out by a particular transmitter. We therefore need some sort of filter—to allow only one radio station to be received at once. A *tuned circuit* is this filter, because it can be tuned to any frequency we want, and can select one of the many radio frequency carrier waves reaching the aerial.

A tuned circuit consists of a *coil* of wire and a *tuning capacitor*. On the opposite page you can see a circuit diagram for a tuned circuit, and beneath are shown the actual components—a *coil* and *tuning capacitor*.

The particular radio frequency waves which a tuned circuit selects depends on the number of turns of wire on the coil and the 'value' of the capacitor. Since it is inconvenient to keep changing the number of turns of wire on the coil, the capacitor is made with movable plates so that the tuning can be varied and the signals we want selected.

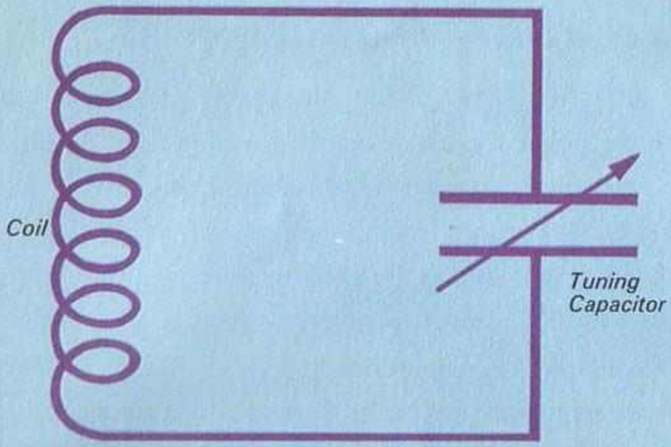
Coil
Tuning
Capacitor

A TUNED CIRCUIT

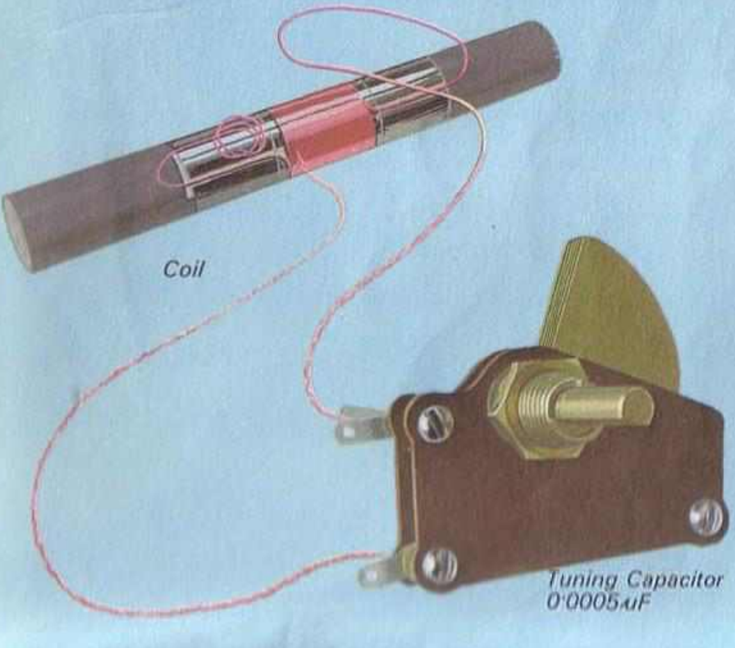
Coil
Tuning Capacitor
0·0005μF

A crystal set – the crystal diode

After the tuned circuit has selected the particular carrier wave frequency required, the audio frequency signal which it carries must be extracted from it. The *crystal* (sometimes referred to as a *diode*) is a simple device that does this by allowing current to pass only one way. It is used in the crystal set as a 'detector'. The action is a complex one, but we can simply say that it detects from the radio frequency carrier wave the *electrical* waves of the same frequency as sound (*audio*) waves and which a loudspeaker or headphone can convert into *sound*.

On the opposite page you will see a circuit diagram showing the tuned circuit mentioned on the previous page but with the symbols for the diode and the earpiece included. With the addition of the aerial and earth symbols, we now have a circuit diagram for a complete working radio set.

Also illustrated are a typical earpiece and a crystal diode.

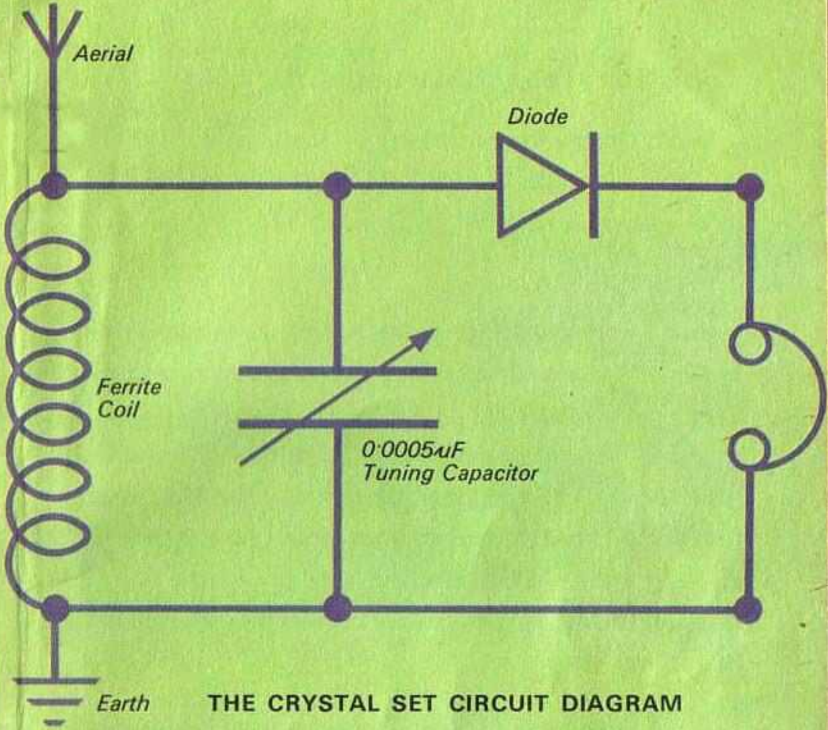
Aerial
Diode
Ferrite
Coil
0·0005µF
Tuning Capacitor
Earth

THE CRYSTAL SET CIRCUIT DIAGRAM

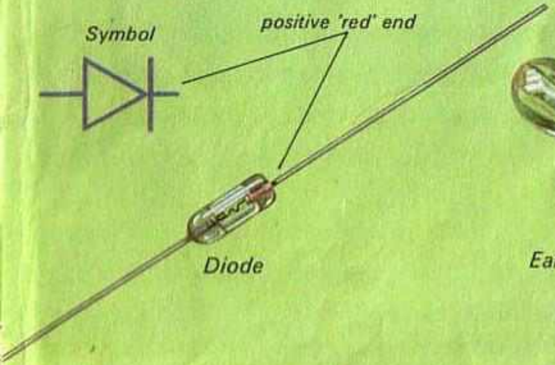
Symbol
positive 'red' end
Diode

Tools for radio construction

Simple radio construction, of the sort described in this book, requires few tools. Most of these can be found in a normal household toolbox or are inexpensive to buy.

During the building of the radio, wires have to be bent into place and sometimes cut to the correct length. Some components are joined by P.V.C. covered copper wire which has to be cut to size and its covering removed from the ends to make the connection. The most useful tool for this is a pair of pointed-nose pliers. These will also strip the plastic from the wire before the connections are made. A pair of wire cutters may also be used to cut and strip wire.

It is possible to buy an inexpensive little tool called a 'wire-cutter and stripper' to do both jobs. A pocket knife is useful for scraping the ends of the wire to clean them before the connections are made. Sandpaper or the rough edge of a matchbox can also be used to clean wires before they are connected. A medium-sized screwdriver will be needed for making the connections.

This radio is to be built on a wooden base, so a few simple woodworking tools—handsaw, gimlet, brace and bit—are also required.

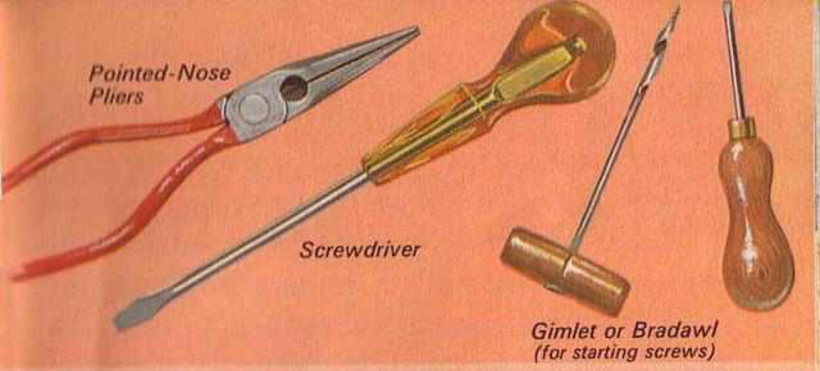
Pointed-Nose Pliers
Screwdriver
Gimlet or Bradawl
(for starting screws)

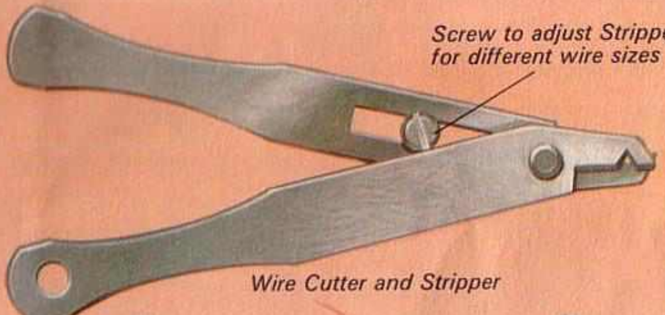
Screw to adjust Stripper for different wire sizes

Wire Cutter and Stripper

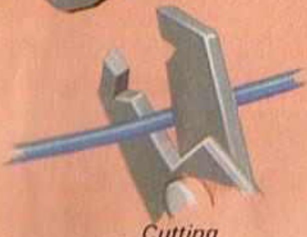

Cutting

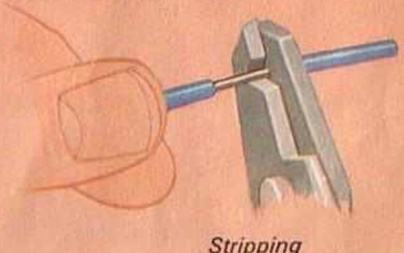

Stripping

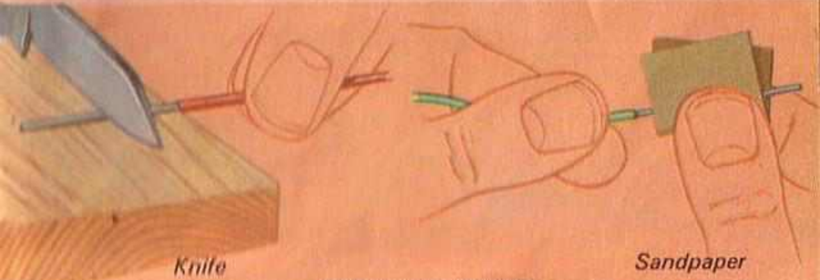

Knife

Sandpaper

CLEANING WIRES

Fixing the components

In electronic construction it is usual to *solder* the components into their places in the circuit. This radio is designed for the beginner, so a non-soldering method of connection is used. For those who wish to solder the connections, the technique is explained on the inside back covers of this book.

A simple method of connecting the components uses woodscrews and washers. The screws are No. 6 brass countersunk, half inch screws and the washers are No. 6 brass screwcups. Two dozen of each are required. All are obtainable from your local ironmonger.

The method is explained in the diagram opposite, but remember the following tips: (1) Make sure the wire to be connected is scraped clean until it shines. (2) Where several wires are trapped under one washer, do not finally tighten the screw until all the wires are in place.

It is possible to test a connection with the simple battery and bulb arrangement in the diagram. Touch the two leads onto some of the bare wire either side of the connection. If the connection is good, the bulb will light up.

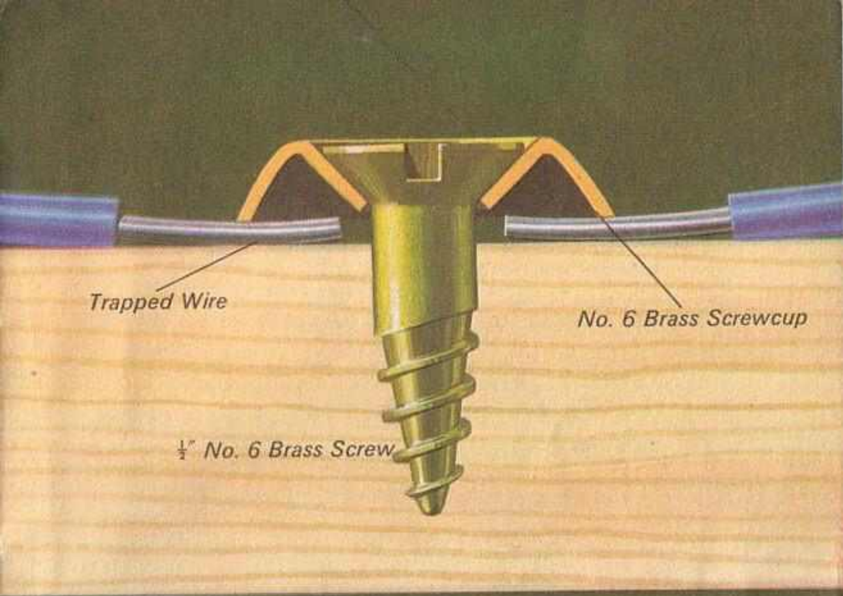
Trapped Wire
No. 6 Brass Screwcup
½" No. 6 Brass Screw

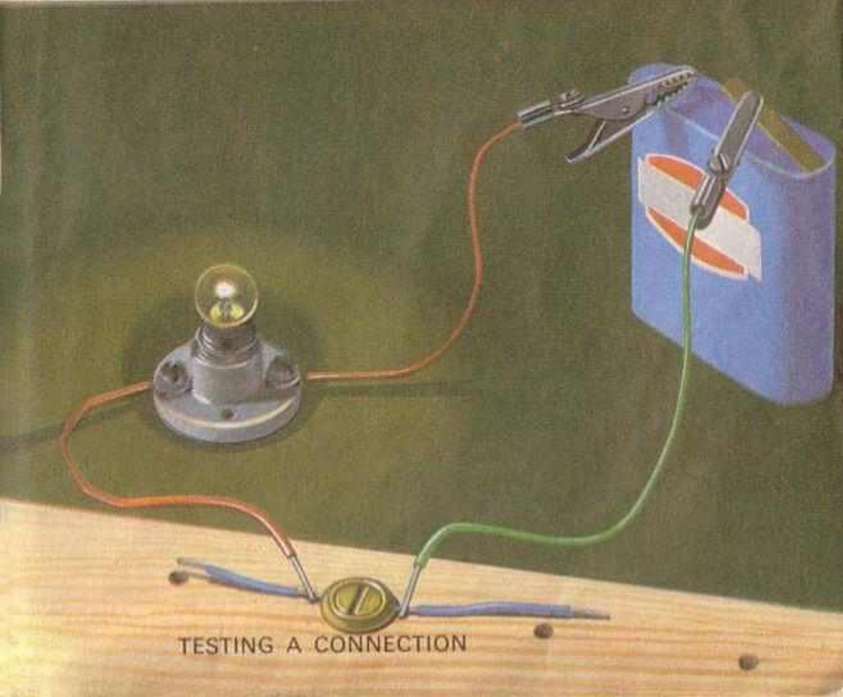
TESTING A CONNECTION

Making the circuit board

In the early days of radio, receivers were made on a baseboard sometimes called a 'bread board'. We are using a form of bread board mounting.

Softwood is used for the base. Mount the screws and screwcups, used for connections, at one inch intervals in two rows of eight. The rows should be an inch-and-a-half apart. This is shown in the drawing facing this page. The two large holes are made with a brace and bit and are for mounting the control components.

The accuracy of the measurements is not vitally important, but it would be helpful to follow those given as closely as possible. Not all the screws are used in building the first part of the radio; the additional screws are needed later as the radio is completed. Although this board is very simple to make, it is important to try to get the screws in straight so that the screwcups press firmly onto the wires.

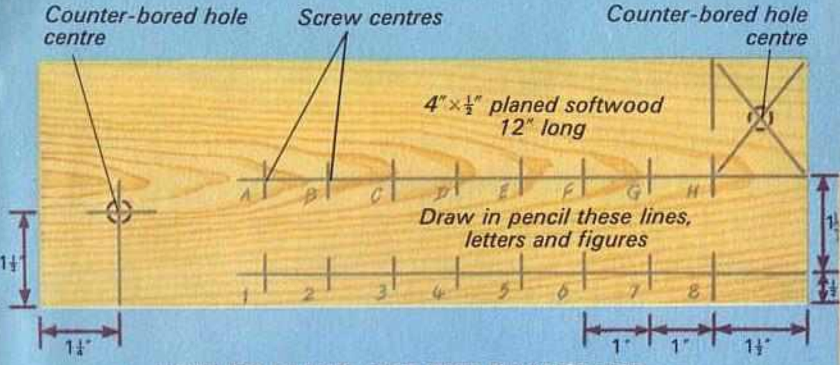
Counter-bored hole centre
Screw centres
Counter-bored hole centre
4″×½″ planed softwood 12″ long
A B C D E F G H
Draw in pencil these lines, letters and figures
1 2 3 4 5 6 7 8
1½″
1¼″
1″
1″
1½″
1½″
½″

HOW TO MARK OUT THE BASEBOARD

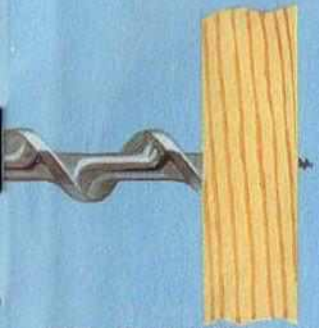

With a ⅜″ bit, drill from the marked side until the point of the bit just appears on the other side. (Do not drill right through)

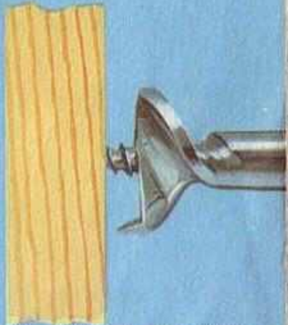

With a ¾″ bit, drill from the other side, half way through the board

MAKING THE HOLES

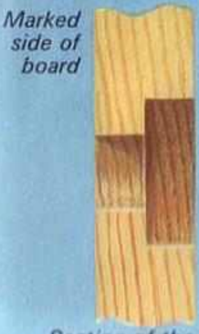
Marked side of board

Section of the counter-bored holes

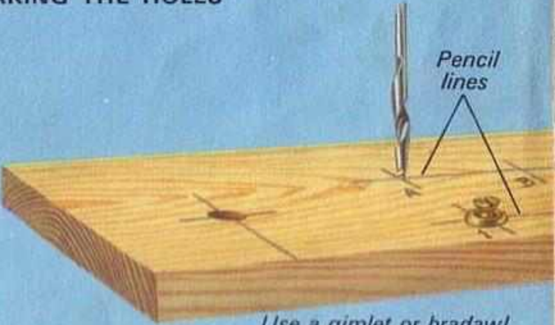
Pencil lines

Use a gimlet or bradawl to make holes for the screws. (Do not fully tighten the screws at this stage)

Contents

The world of radio 4
Radio reception 6
Reading circuit diagrams 8
A crystal set—the tuned circuit 10
A crystal set—the crystal diode 12
Tools for radio construction 14
Fixing the components 16
Making the circuit board 18
Winding the coil 20
Building the crystal set 22
The crystal set—its aerial and earth 24
An experimental P.O.W. radio 26
The transistor 28
Adding a transistor 30
Resistors 32
Biasing the transistor 34
Adding a second transistor 36
The second transistor stage 38
Adding a loudspeaker 40
Regeneration 42
Adding regeneration 44
Wiring the regeneration stage 46
Using the radio 48
The completed radio 50

Winding the coil

Before the crystal set can be built, we must wind a tuning coil for the tuned circuit. This is done on a piece of ⅜ inch diameter ferrite rod. Ferrite is iron dust stuck together. Usually ferrite rods are 6 inches long, so cut a notch all round the centre and smartly snap the rod in half (Figs. 1 and 2).

Wrap about 1½ inches of sticky tape around the centre of the rod on which to wind the coil (Fig. 3).

Obtain some 36 s.w.g. (standard wire gauge) enamelled copper wire from your radio or electrical dealer. Leaving about 3 inches of wire, secure one end of the wire with sticky tape (Fig. 4) and wind 5 turns, side by side, onto the rod (Fig. 5). Pull out a 3 inch loop and twist the wire until it is tight up to the rod (Fig. 6). Then wind a further 45 turns, side by side, securing the final end with tape. The whole coil can be covered with tape to keep the windings secure.

Remember to leave about 3 inches of spare wire at both ends and at the twisted wire (the 'tapping' point). Later the enamel must be scraped off so that connections can be made.

The illustrations opposite will guide you through each stage.

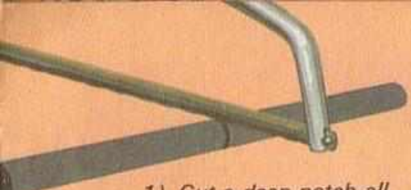

1) Cut a deep notch all the way around the rod

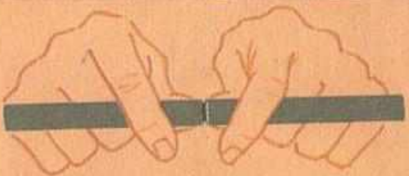

2) Snap the rod by holding close to the notch with forefingers and thumbs only (otherwise it may splinter)

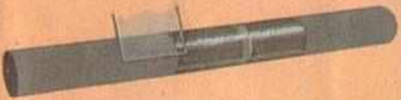

3) Wind the tape in separate parallel bands, slightly overlapping. (If spirally bound the tape will wrinkle)

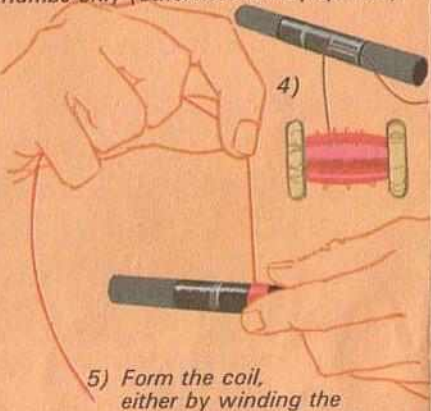

5) Form the coil, either by winding the wire or rotating the rod, whichever you find easier

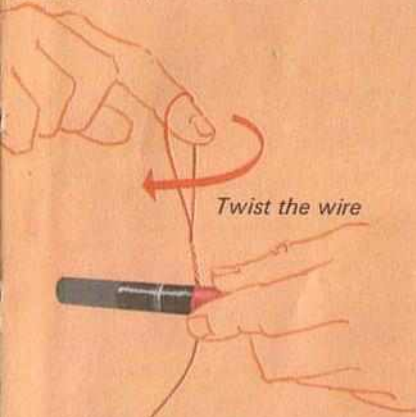

6) Twist the tapping point with a finger in the loop, holding the rod still with the wire trapped under finger and thumb

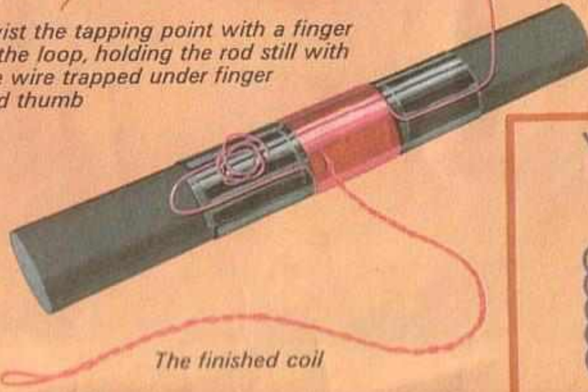

The finished coil

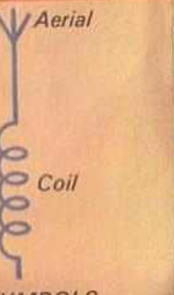

SYMBOLS

Building the crystal set

Having made the coil, the ferrite rod can be mounted on the baseboard by twisting a loop of stiff wire around each end and screwing the wire under two screwcups, as shown opposite.

Next the tuned circuit can be completed by the addition of the tuning capacitor. You can buy one from your local radio dealer and you should ask for a 0·0005μF solid dielectric type. The illustration opposite shows how it is mounted on the baseboard, the capacitor spindle passing through the ⅜ inch hole. The wiring connections are made to the two terminal nuts on the back of the capacitor. The connecting wire used should be PVC covered copper wire, the ends being bared before they are trapped under the screwcups.

The illustration also shows how the diode is included in the circuit by means of the screwcups. The earpiece is also shown inserted. When purchasing this, you should ask for a crystal earpiece. A pair of high resistance headphones could be used in place of an earpiece, and may give even better results. *Resistance* is discussed on page 32, in the section headed 'Resistors'.

On the following two pages there is some useful information about the aerial and earth.

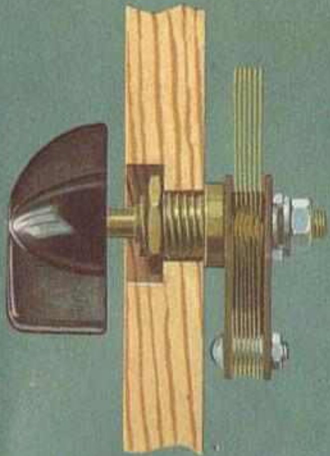

HOW THE CAPACITOR IS MOUNTED THROUGH THE BOARD

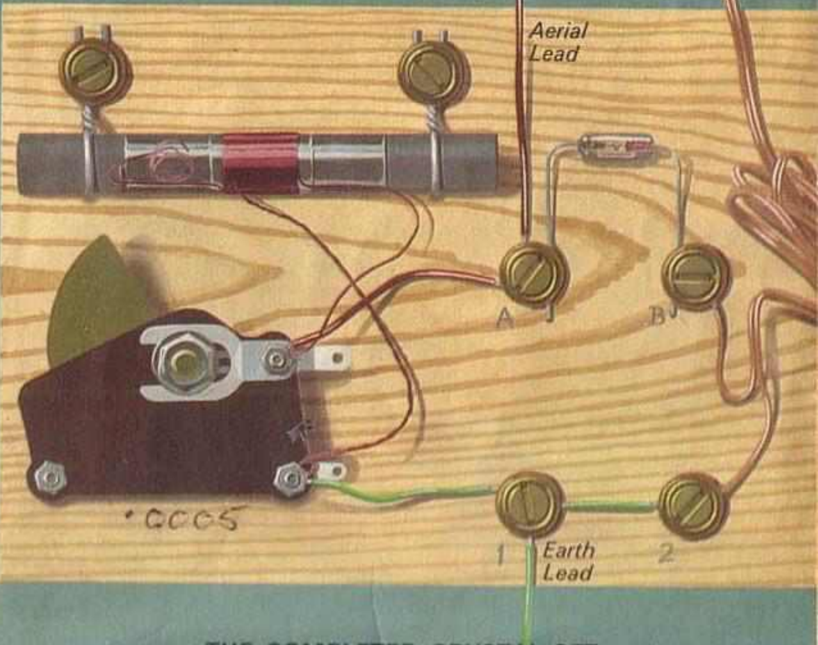
Aerial Lead
A
B
·0005
1
Earth Lead
2

THE COMPLETED CRYSTAL SET

The crystal set – its aerial and earth

The aerial is the part of the receiver which 'picks up' the radio signals. Unlike our crystal set, many modern transistor radios are sufficiently sensitive to pick up enough signal from the ferrite core of the tuning coil. Although the final radio which we will make will not need an aerial and earth, the crystal set relies on the electrical current set up between a good aerial and earth to work well.

A suitable aerial can be made from inexpensive PVC covered copper wire. The idea is to suspend as much of this wire as possible, as high as possible. The far end can be attached to a tree, the top of a pole or even the eaves of a friendly neighbour's house, but this can be dangerous and is best done by an adult or under the supervision of an adult. The wire should be fastened onto a large screw eye or hook. The near end of the wire may be taken in the top of an upstairs window and connected to the aerial connection on our radio.

For the aerial to be effective, the other side of the tuned circuit has to be connected to earth. The simplest method is to attach the earth lead to the nearest household water pipe. These pipes lead to the ground. If your water pipes are plastic, make your own earth by driving about a yard of copper pipe into soft earth and connecting the earth wire to the top.

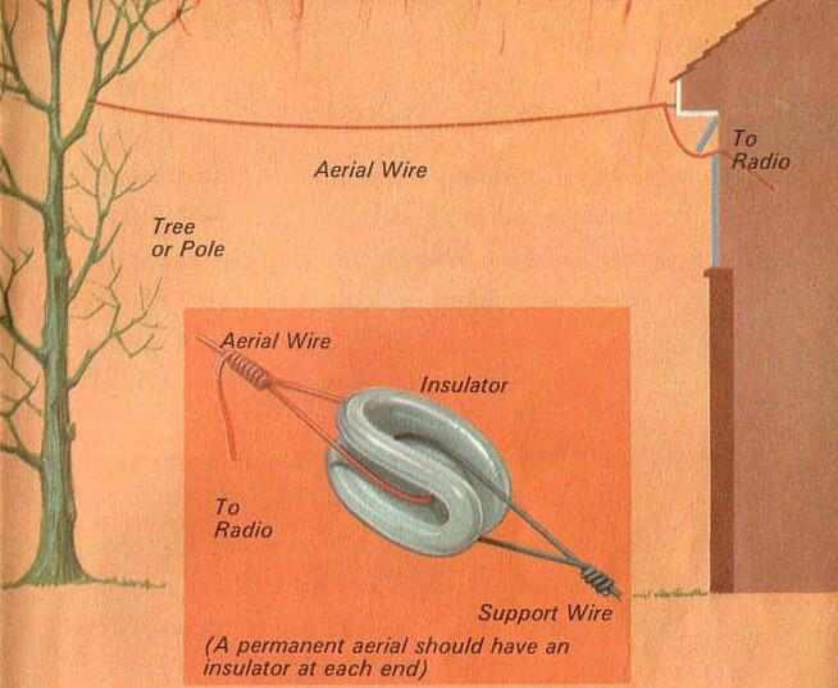
Aerial Wire
To Radio
Tree or Pole
Aerial Wire
Insulator
To Radio
Support Wire

(A permanent aerial should have an insulator at each end)

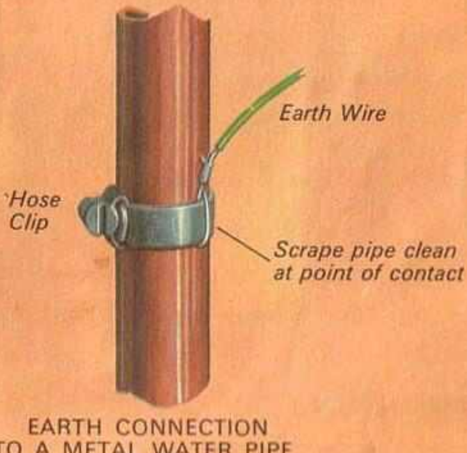
Earth Wire
Hose Clip
Scrape pipe clean at point of contact

EARTH CONNECTION TO A METAL WATER PIPE

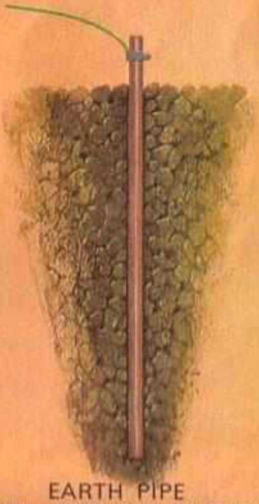
EARTH PIPE DRIVEN INTO THE GROUND

An experimental P.O.W. radio

During the last war, some prisoners in prison camps used home-made radios to keep in touch with events in their native lands. Often, all the components had to be made from items picked up within the camp. Diodes were almost impossible to obtain, so several ingenious home-made alternatives were tried.

The diagram shows one simple method of improvising a diode. Take a small piece of wood, about 1 inch thick, and a piece of washed coke about the size of a pea. Mount the coke on the wood as shown, using a screw and screwcup, and taking a lead from beneath the screwcup. Wind about a foot of steel wire—a length of piano or guitar wire is ideal—around a number 4 or 6 knitting needle, leaving about 1 inch on each end. The wire should now be slightly springy. Fasten one end to the board with a screw and screwcup, taking another lead from this screwcup. Stretch the wire until the free end lightly touches the coke.

Remove the diode from the crystal set and screw the wires in its place. It should be possible to hear some local stations by adjusting the tuning capacitor. If nothing is heard, keep moving the position of the wire on the coke until a signal is heard.

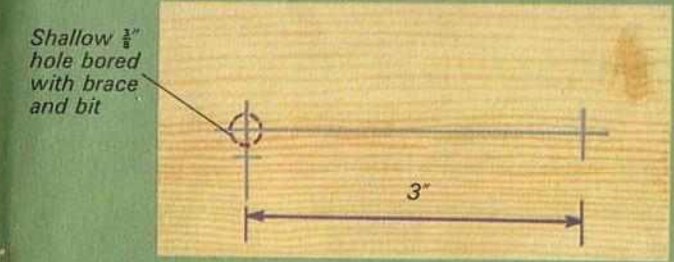
Shallow ⅛" hole bored with brace and bit
3"

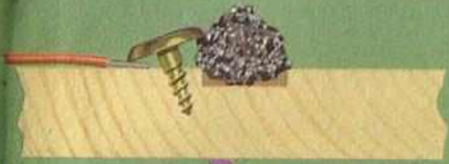

The screw is at a slight angle so as to grip both coke and lead securely

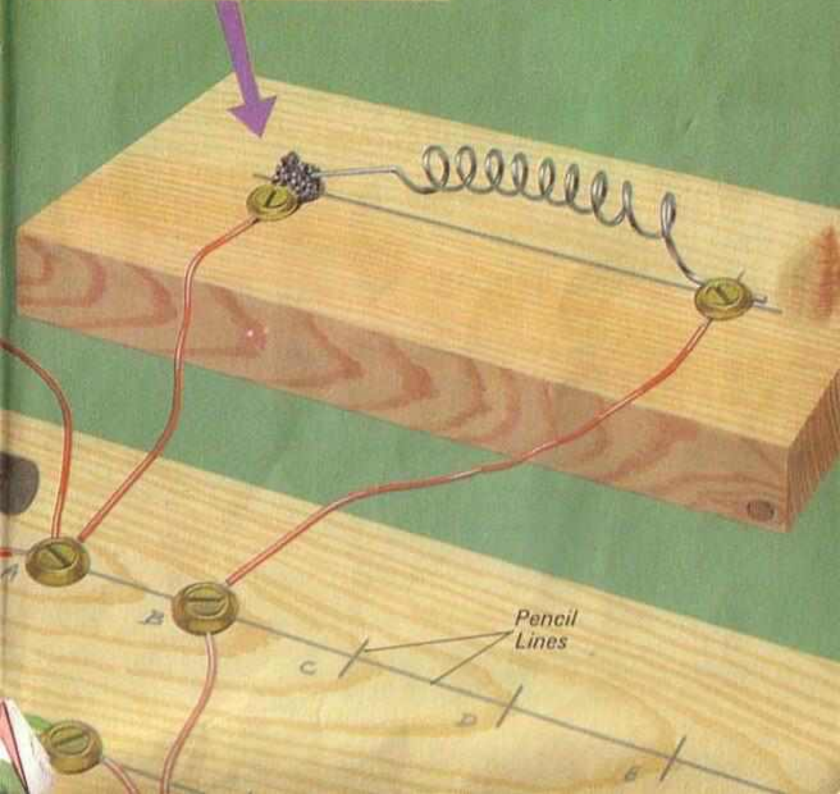
A
B
C
D
E
Pencil Lines

The transistor

To most people the word TRANSISTOR means a small radio set, but a transistor is really a single electronic component. The transistor is one of the most important inventions of the last 25 years. It has helped to make electronic equipment smaller and less dependent upon high voltages.

A transistor is like two diodes working together inside one case. Most transistors have three leads, emitter, collector and base. A transistor can be damaged unless the correct leads are used when connecting it into a circuit. The base lead is usually the middle wire and the collector lead is set slightly apart from the other two wires. The transistors we are to use also have a red spot painted on the casing to indicate the collector lead.

We shall be using transistors as amplifiers; that is, to make electrical signals stronger. Transistors are designed so that a small, changing current between the base and emitter produces a large current change between the collector and the emitter. So small signals applied to the base will appear amplified (or made bigger) at the collector. Our first transistor will be used to make louder the sound coming from the crystal set.

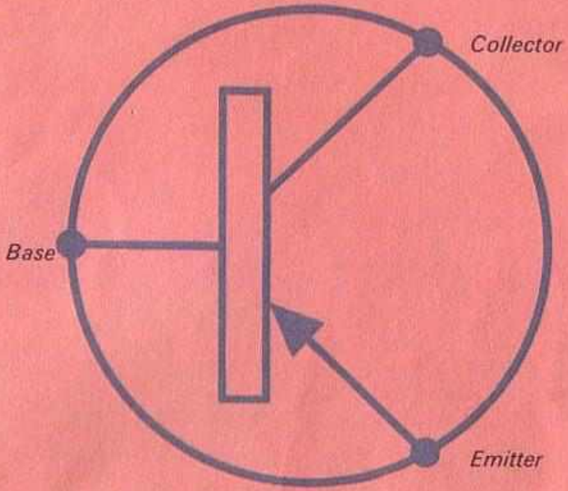
Collector
Base
Emitter

THE SYMBOL FOR A TRANSISTOR

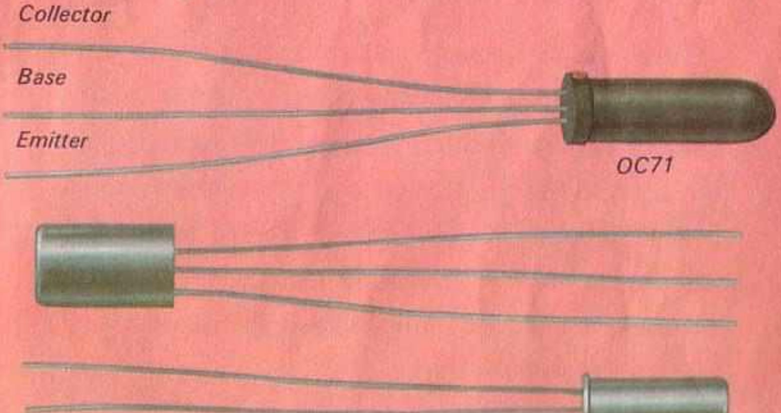
Collector
Base
Emitter
OC71

TRANSISTORS

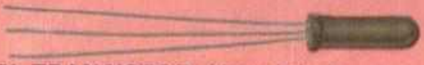

OC71 TRANSISTOR *(Actual size)*

A Ladybird 'How to make it' book

making a Transistor Radio

by G. C. DOBBS

with illustrations by B. H. ROBINSON

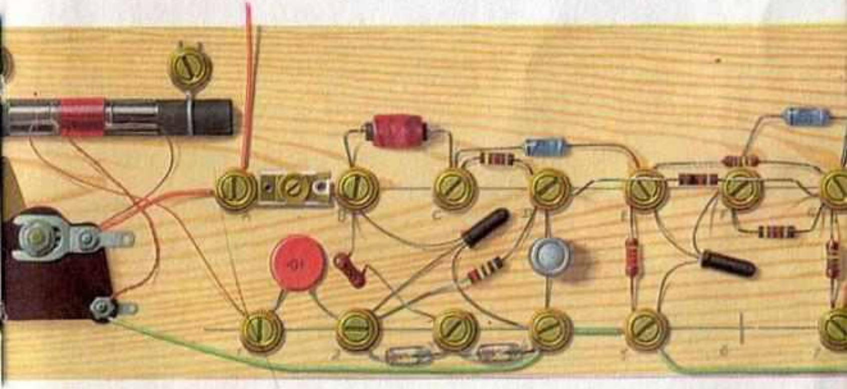

© Wills & Hepworth Ltd *1972*

Printed in England

Adding a transistor

The circuit diagram shows how a simple transistor amplifier is inserted where previously we had the earpiece. Compare this diagram with the circuit diagram of the crystal set on page 23.

The transistor used is an OC71, and other extra components for this stage are a 2·7K resistor and a 9-volt battery. The battery is the small transistor radio type PP3, for which can be obtained an inexpensive snap connector. Resistors are discussed on the next page. This resistor forms an output 'load' for the collector, and the earpiece is connected across the resistor.

Take care to connect up the circuit as shown in the drawing. The connections of the transistor are important. Join the emitter to screw 5, the base to E and the collector to F. The earpiece is removed and reconnected to F and G. Connect the red battery connector lead to screw 5, and the black lead to screw G. The radio is switched on by pushing the connector into place on top of the battery.

This one transistor should make a big improvement to the crystal set; the same stations will be heard but they will be much louder.

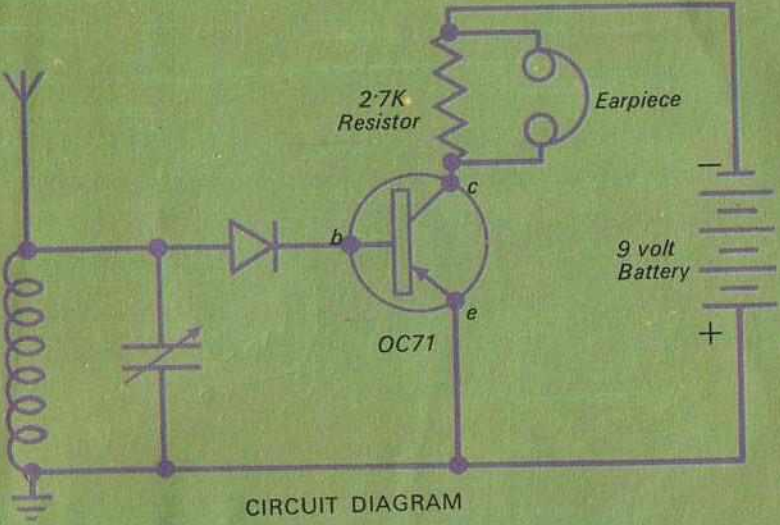
2·7K
Resistor
Earpiece
c
b
e
OC71
−
9 volt
Battery
+

CIRCUIT DIAGRAM

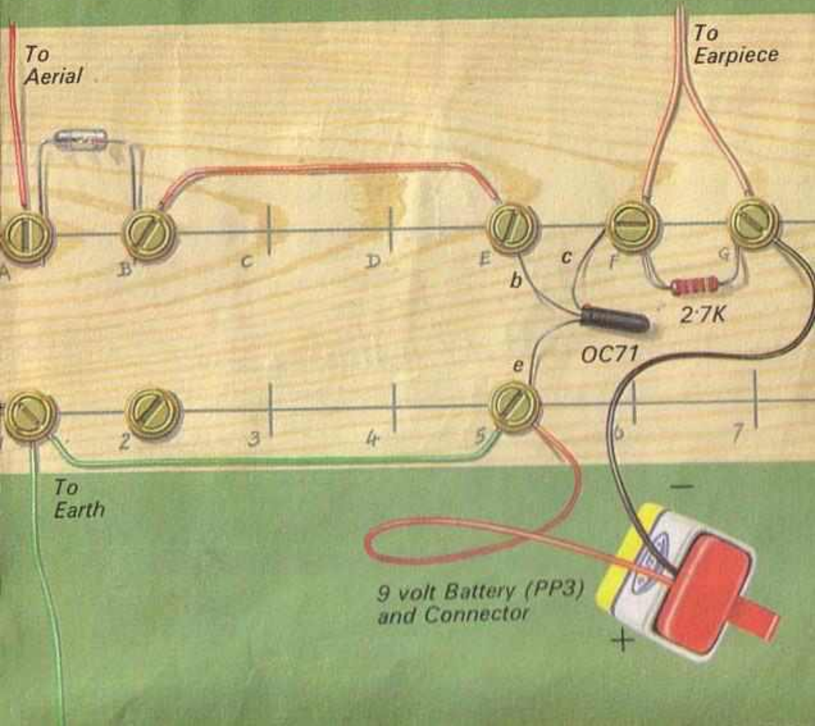
To
Aerial
To
Earpiece
A
B
C
D
E
F
G
c
b
2·7K
OC71
e
2
3
4
5
6
7
To
Earth
−
9 volt Battery (PP3)
and Connector
+

Resistors

Resistors are very common components in electronic construction. Look at most pieces of equipment and you will see several small, wire-ended tubes with brightly coloured bands. These are called resistors because they 'resist' the flow of electrical current in the circuit. This is done to adjust the voltages and flow of current at various places in a circuit.

The little tubes are filled with a carbon composition, and the coloured bands give the *value* of the resistor. This value is called the *resistance* and is measured in *ohms*. When one knows the colour code opposite, it is easy to work out the resistance from the coloured bands. The tolerance of the resistor, that is, how close one can expect it to be to the value, is sometimes indicated with another coloured band. Our radio can use resistors of any tolerance, so this band may be ignored.

The symbol for resistance is the Greek letter omega Ω. Quite often, very high resistances are used—sometimes thousands or millions of ohms. To reduce the number of noughts, we call thousands K Ω (Kilo ohms) and millions M Ω (or Mega ohms). The ohm sign is usually omitted from circuit diagrams when the Kilo (K) and Mega (M) signs are used.

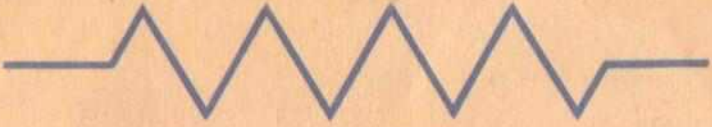

THE SYMBOL FOR A RESISTOR

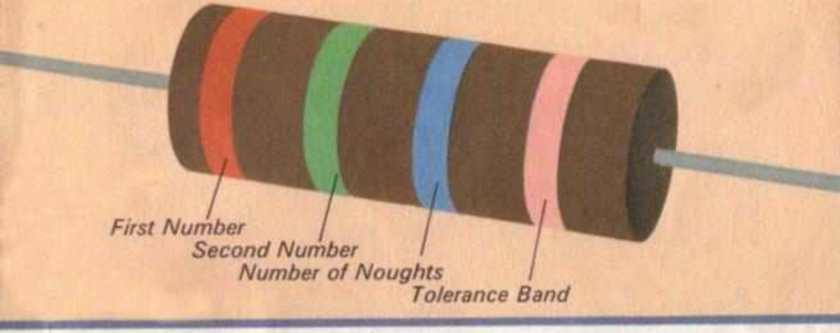
First Number
Second Number
Number of Noughts
Tolerance Band

RESISTOR COLOUR SCOPE

BLACK	0 (or 0 noughts)	GREEN	5 (or 5 noughts)
BROWN	1 (or 1 nought)	BLUE	6 (or 6 noughts)
RED	2 (or 2 noughts)	VIOLET	7 (or 7 noughts)
ORANGE	3 33000 (or 3 noughts)	GREY	8 (or 8 noughts)
YELLOW	4 (or 4 noughts)	WHITE	9 (or 9 noughts)

TOLERANCE COLOURS

GOLD 5%	SILVER 10%	SALMON 20%

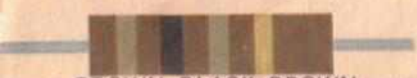

BROWN, BLACK, BROWN
=100

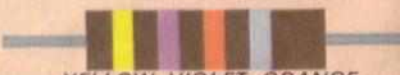

YELLOW, VIOLET, ORANGE
=47000 or 47k

Biasing the transistor

The transistor amplifier we have built is the simplest arrangement possible. If a transistor is to be used efficiently, one has to supply the correct voltages to its various parts; this is called *biasing*. In our simple arrangement the signal from the diode provided a voltage to *bias* the base of the transistor. Resistors can be added to make the biasing correct.

The circuit shows how the transistor is biased with two resistors. (33K Ω—orange, orange, orange, and 150K Ω—brown, green, yellow.) They give the correct voltages between the collector, base and emitter. Take care to connect the resistors in the places shown or the effect will be ruined.

This addition to the radio will not increase the volume in any noticeable way, but the biasing is essential before the next stage is added.

The wire ends of resistors, like any other components, can become dirty and greasy, so scrape the wires clean before screwing them into place.

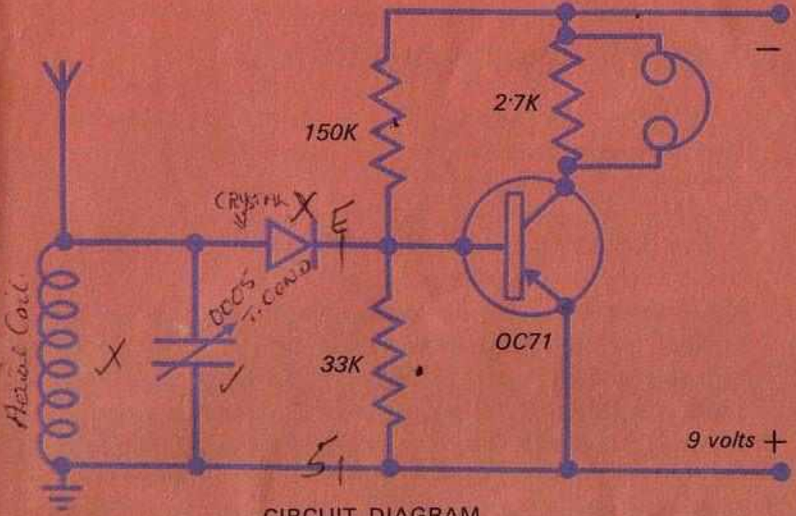
150K
2·7K
CRYSTAL
X
E
·0005 T. COND
Aerial Coil
X
33K
OC71
S1
9 volts +
−

CIRCUIT DIAGRAM

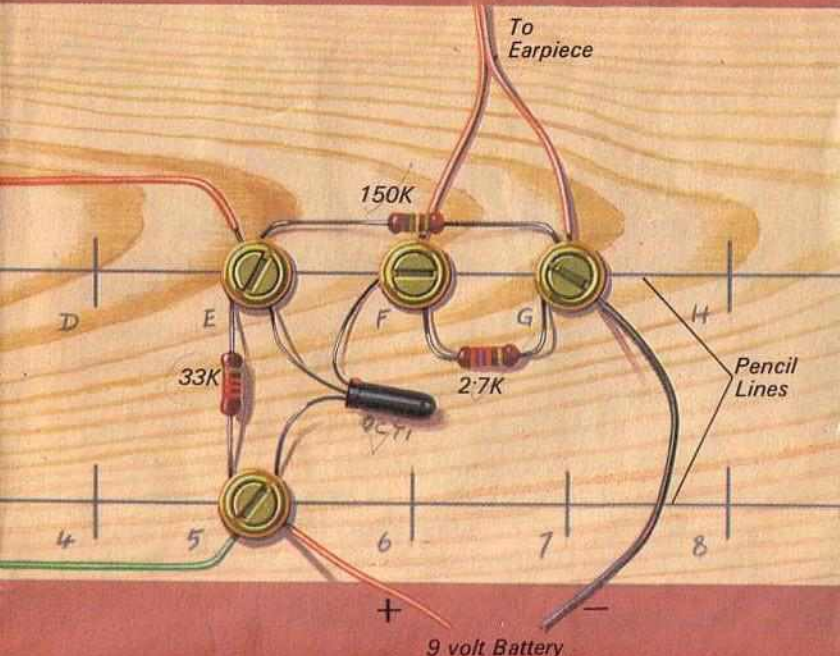
To Earpiece
150K
D
E
F
G
H
33K
2·7K
OC71
Pencil Lines
4
5
6
7
8
+
−
9 volt Battery

Adding a second transistor

Before we can add another transistor to our radio, two new components must be introduced—a *potentiometer* and a *capacitor*.

The potentiometer is a variable resistance. Look at the drawing on the opposite page and you will see that it has three contacts. The outer two are each end of a circular carbon track. The centre tag is joined to a movable contact which slides over the track when the shaft is rotated. By moving the shaft, which is usually fitted with a plastic knob, any part of the resistance can be used and tapped off through the centre tag. Potentiometers are often used, as in this case, as volume controls.

A *capacitor* is a component which allows the signal to pass but not the fixed voltages used to power a circuit. It consists of two plates which are separated by insulating material; the circuit symbol shows this. Electrolytic capacitors must be connected into a circuit the right way round. One end is marked either with a red band or a + sign, and this side goes to the positive side of the circuit. The negative (—) wire is connected to the metal can which forms the body of the capacitor, and which may be marked with a black band. Sometimes this is not obvious, but follow the circuit and layout diagrams with care to avoid mistakes. Other fixed capacitors can be connected either way round: we will meet these later. The value of capacity is written in microfarads, (μF or sometimes MFD). We require 10μF at a working voltage of 16-volts.

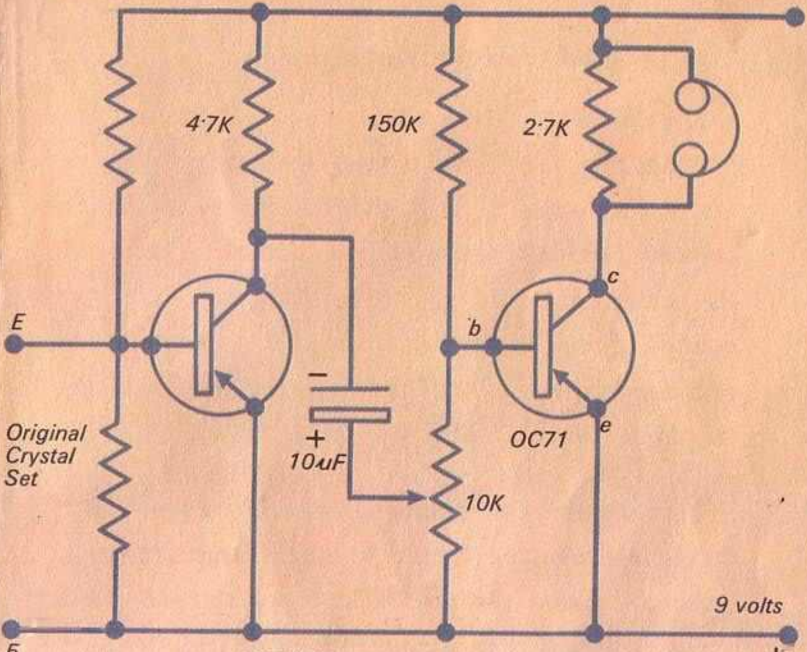
4·7K
150K
2·7K
E
b
c
e
Original Crystal Set
OC71
−
+
10µF
10K
9 volts
5
+

CIRCUIT DIAGRAM

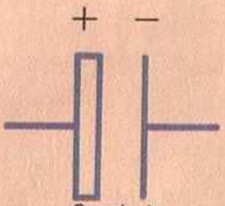

Symbol

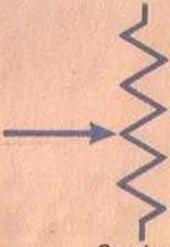

Symbol

ELECTROLYTIC CAPACITOR

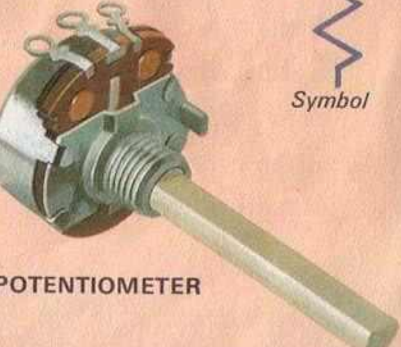

POTENTIOMETER

The second transistor stage

The circuit for the second stage is shown on the previous page. The signal is taken from the collector of our first transistor, through the capacitor to the sliding contact of the volume control potentiometer. The nearer the centre contact is adjusted to the base side of the control, the more signal is allowed to go to the base of the second transistor. The greater the signal reaching this transistor, the louder will be its output.

The potentiometer is mounted in the spare ⅜ inch hole. Its tags are too close together to use the large screwcups, so small screws are used. These are screwed down through the holes in the potentiometer tags, trapping the leads under the tags.

Remove the 2·7K resistor from F and G and connect in the new components carefully, following the layout diagram facing this page. As you work, compare the layout diagram, connection by connection, with the circuit diagram. This additional transistor will greatly increase the volume, enough to drive a loudspeaker, which is the next addition.

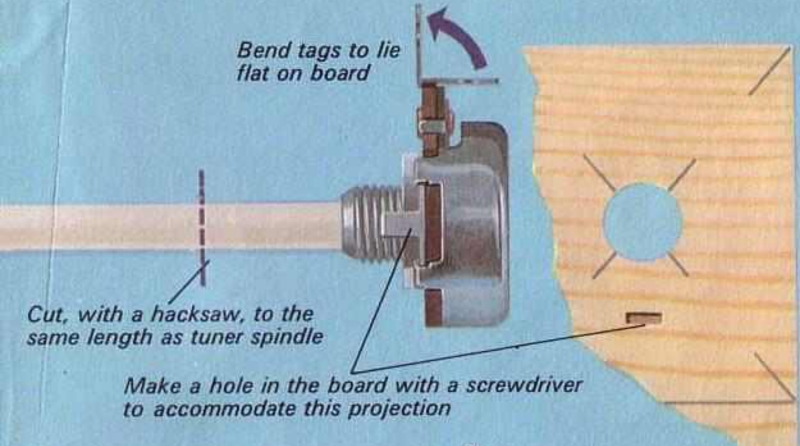

Bend tags to lie flat on board

Cut, with a hacksaw, to the same length as tuner spindle

Make a hole in the board with a screwdriver to accommodate this projection

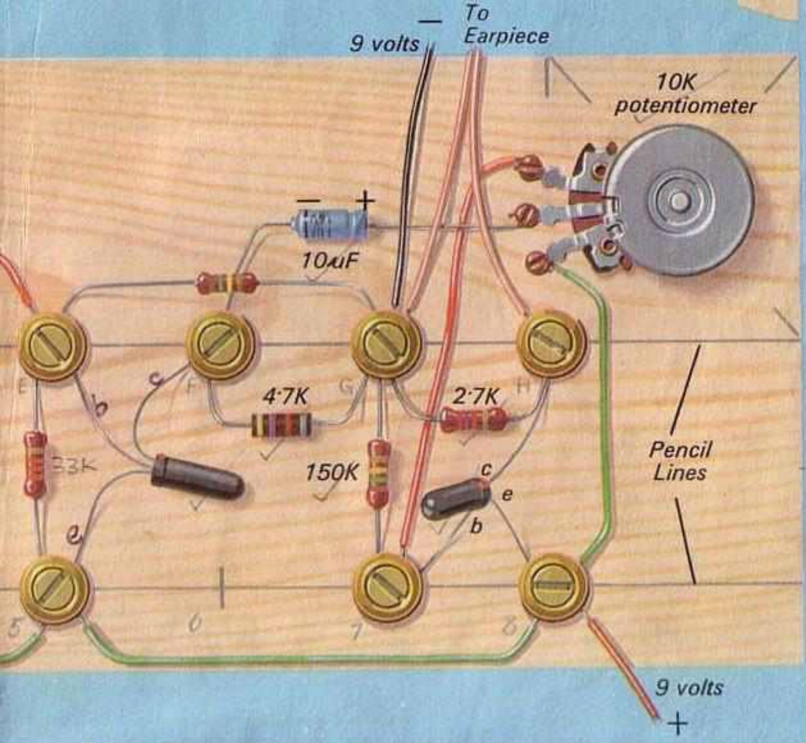

9 volts
—
To Earpiece
10K potentiometer
—
+
10µF
E
F
G
H
4·7K
2·7K
33K
150K
b
c
e
c
e
b
Pencil Lines
5
6
7
8
9 volts
+

The world of radio

Today, radio is part of our everyday life, yet only at the beginning of this century it was a miracle. It is still something to marvel at—and a fascinating subject to study.

With this book you will learn, stage by stage, how to build more than one simple receiver. Everything you should know is explained simply and illustrated clearly, and you will have the pleasure of constructing something marvellous that really works.

All the components used are easily available, usually from your local radio dealer. Most large towns have shops specialising in stocking components, but it is also very easy to obtain any you need by post. All the popular electronic construction magazines contain advertisements for mail order supplies. Many of these suppliers provide catalogues which list all the components you will need.

0 7214 0324 7

Adding a loudspeaker

Using our radio with an earpiece means, of course, that only one person at a time can hear the programme. A loudspeaker changes electrical waves into sound waves in the same way that an earpiece does, but it has a large cardboard cone so that several people can hear the sound. Loudspeakers can have very large cones, but a smaller cone speaker with a diameter of 3 inches is ideal for our purpose.

Loudspeakers are made with inputs of varying resistances, and our radio requires a 3 or 5 Ω type. Unfortunately the loudspeaker cannot directly replace the earpiece because its resistance is much lower than that of the earpiece, so it has to be matched into the circuit with a *transformer*. A transformer can change a current of one voltage to a smaller current of a higher voltage or to a larger current of a smaller voltage. It consists of a 'primary' and 'secondary' winding of fine wire round a metal 'former'. The transformer we require is the Eagle LT700. This transformer has rather long leads, and these may be shortened by winding them tightly round a knitting needle. This makes spiral leads which can then be pulled out to reach the screws.

Screw the transformer onto the base, as shown; remove the 2·7K resistor across G and H and connect the correct leads (colour coded) to the screwcups. Two more screwcups are added at the end of the board to connect the output leads to the loudspeaker, using two small crocodile clips.

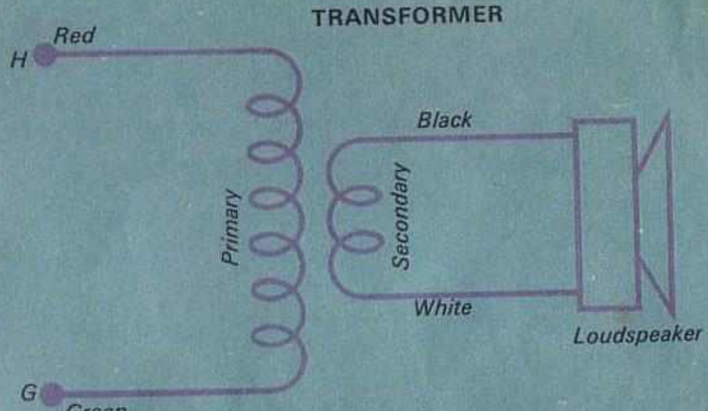
TRANSFORMER
Red
H
Black
Primary
Secondary
White
Loudspeaker
G
Green

CIRCUIT DIAGRAM

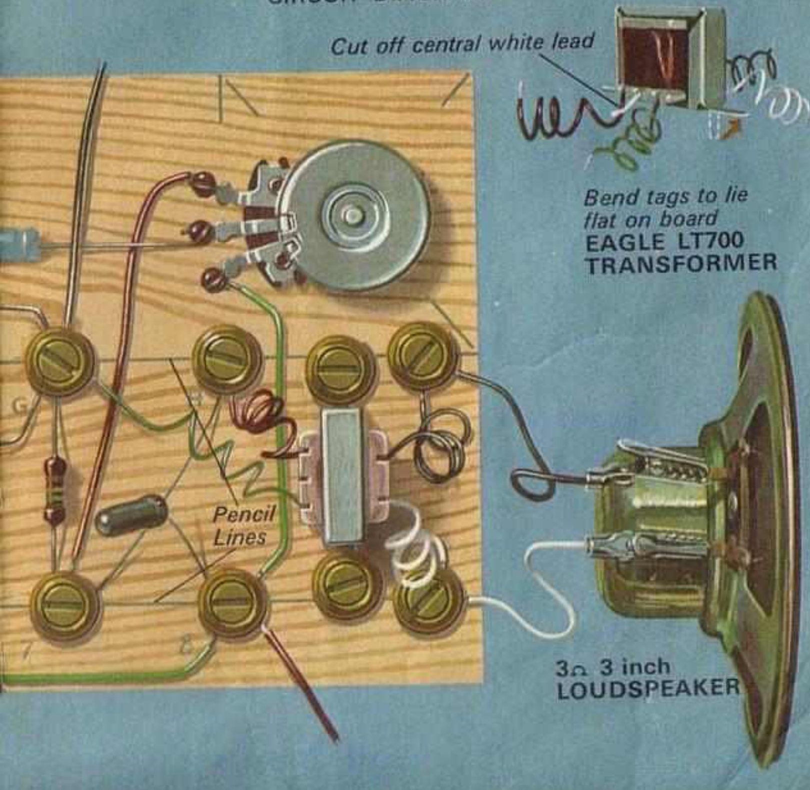
Cut off central white lead
Bend tags to lie flat on board
EAGLE LT700 TRANSFORMER
G
H
Pencil Lines
7
8
3Ω 3 inch LOUDSPEAKER

Regeneration

We shall now replace our simple crystal set stage with a transistor stage, using *regeneration*. Regeneration will enable our first transistor in the radio to be used three times.

The circuit shows the regeneration stage. The signal from the aerial is tuned by the coil and variable capacitor, the orange marked signal then passes through capacitor C1 to be amplified by the transistor and appear at a *radio frequency choke* (R.F.C.). Part of this signal, marked in blue, passes through C2 to a pair of diodes which change the radio signal into an audio signal, which is amplified again by the transistor. The signal, now marked red, passes through C3 back to the coil. The part of the coil tapped between the transistor collector (through C1) and earth, allows the signal to reach the collector again.

This means that some of the signal at R.F.C. has been amplified three times, so this is a very economical use of one transistor. Passing signals back again like this is called *feedback*. If the feedback is too great, the circuit will produce a howling sound, so C3 is an adjustable capacitor to control the amount of feedback. The final signal is taken from R.F.C. through capacitor C4.

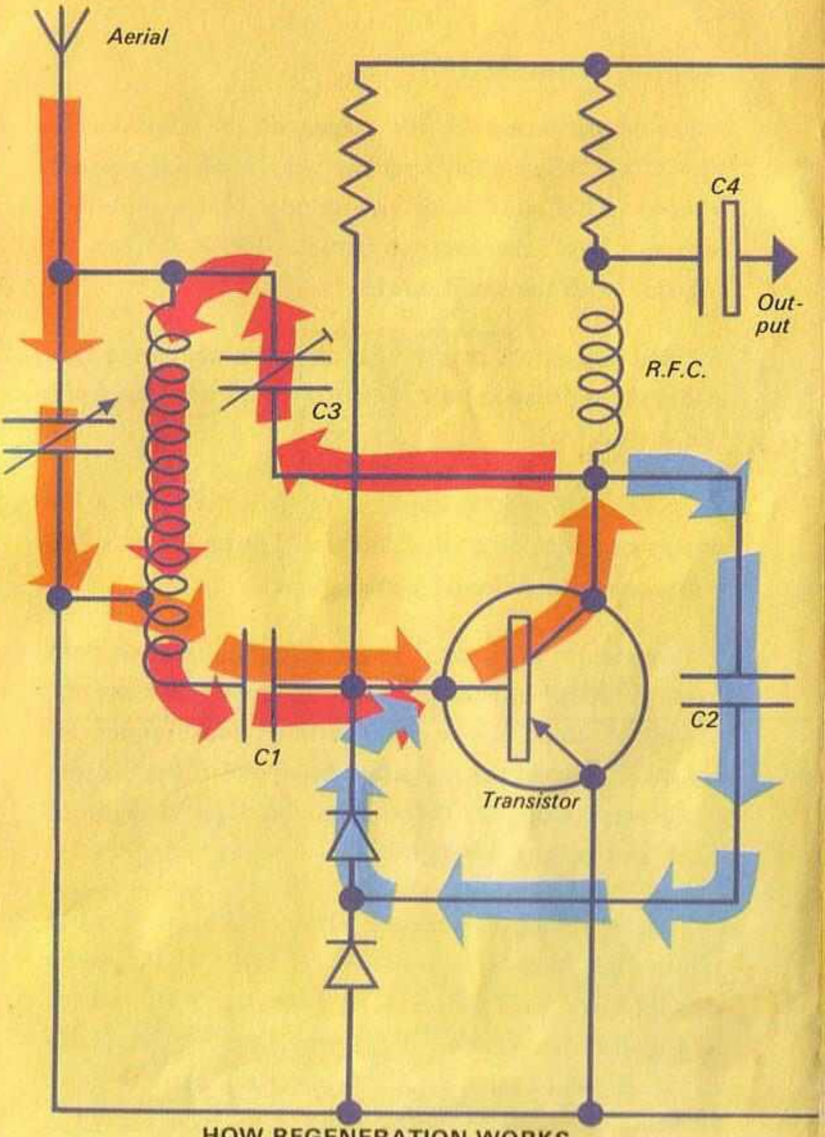
Aerial
C4
Out-
put
R.F.C.
C3
C1
C2
Transistor

HOW REGENERATION WORKS

In circuit diagrams, wires which cross one another, (but do NOT touch and are NOT connected), are sometimes shown like this + (as in the diagram) or like this —⌒—. Shown like this —●— always means they are connected.

Adding regeneration

The circuit shows the new stage with the components added. The adjustable capacitor is a 10 pF (picafarad) *trimmer*. This consists of two copper plates mounted on a pot base; the distance between the plates can be adjusted with the small screw.

Fixed capacitors may be flat or cylindrical with two leads; the type we require, *ceramic capacitors*, may look like a disc.

The *radio frequency choke* is really a coil with a lot of turns of wire on a small former. The choke acts like a resistance to radio-frequency waves.

Two diodes are needed to provide the feedback path from the 500pF capacitor. If the top diode were omitted, the signal would not be changed into audio frequencies before it reached the transistor base, and if the bottom diode were omitted, the signal would go straight to earth and be lost.

The 1K resistor lowers the battery voltage needed to power this stage. It is possible for some of the audio signal to feed back to this stage through this 1K resistor and distort the signal. This is prevented by the 100μF capacitor which takes any stray signals to earth without affecting the battery voltage.

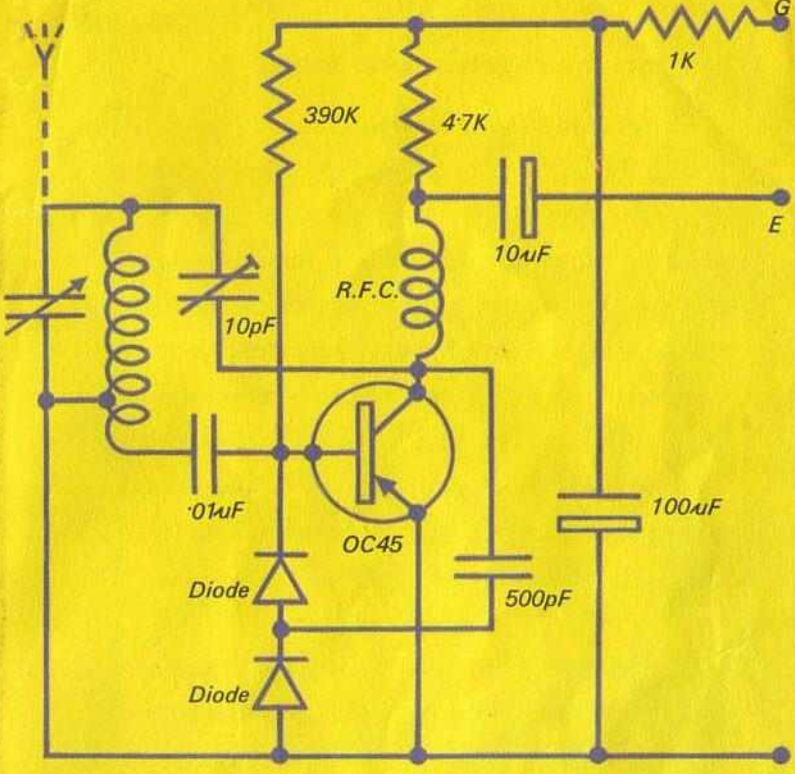
G
1K
390K
4·7K
E
10µF
R.F.C.
10pF
·01µF
OC45
100µF
Diode
500pF
Diode

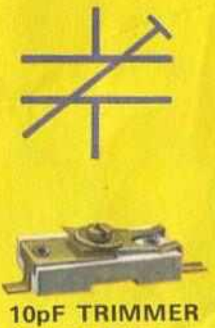
10pF TRIMMER
postage stamp
type'

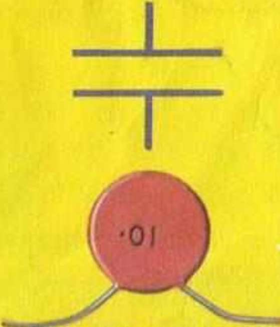
·01
DISC CERAMIC CAPACITOR

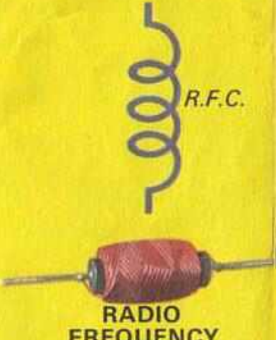
R.F.C.
RADIO
FREQUENCY
CHOKE

Wiring the regeneration stage

The drawing shows how this stage is wired onto the board. Before wiring the stage, all the components from the crystal stage, up to screws E and 5, are removed except for the coil and tuning capacitor. The diode is removed to become one of the two diodes in the new stage. Follow the drawing carefully, adding one component at a time and checking its position on the circuit diagram. There are many bare wires which come close to each other, and great care must be taken to ensure they do not touch each other. The 1K resistor must not touch screws E and F.

The trimmer is trapped by its tags under screws A and B; if it is too long or too short, the screws may be moved. All three connections to the coil are now used, and if the wires are not placed in the correct positions, regeneration will not occur. The drawing shows how to place these three wires. The 10μF capacitor takes the signal from the new stage to the base of the next OC71 transistor. The advantage of this new stage is that our radio will have a greater output and no longer need a large aerial and an earth. Regeneration is tricky to get used to, but once mastered, the results are surprising.

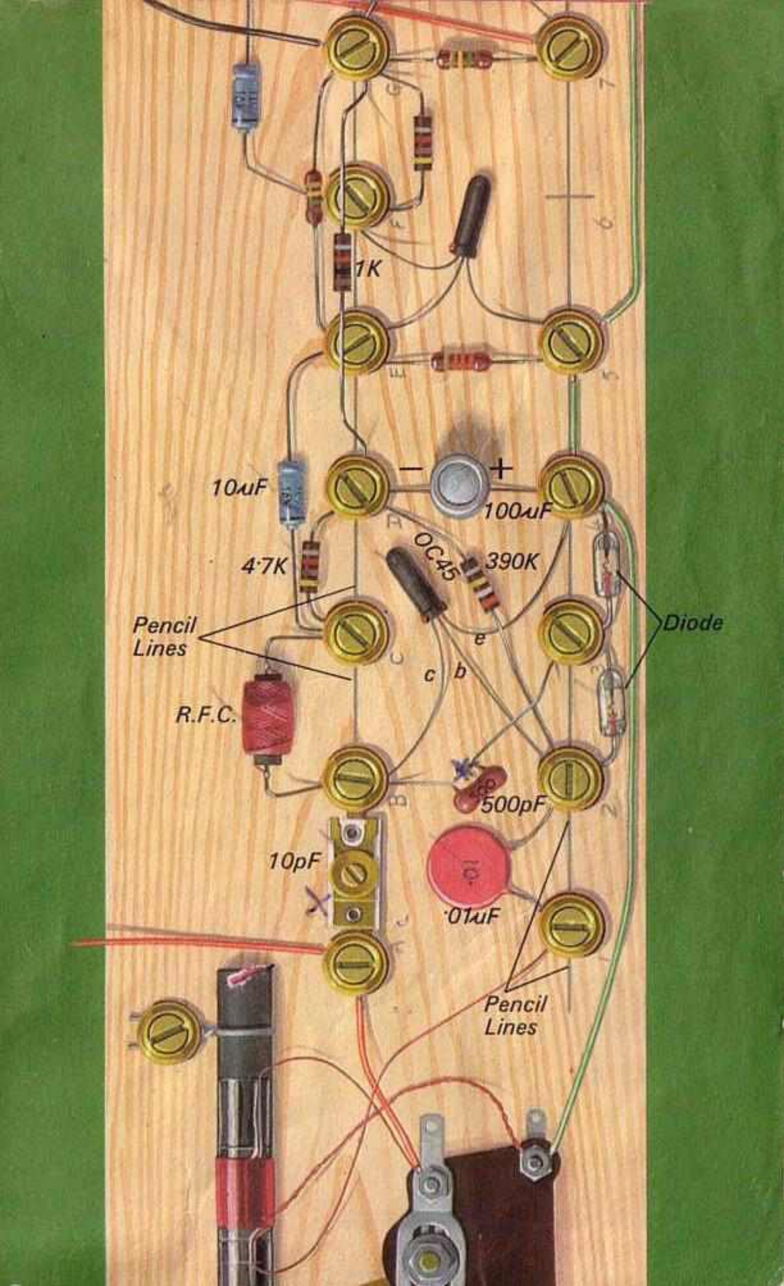
1K
−
+
10µF
100µF
OC45
390K
4·7K
Pencil Lines
Diode
e
c
b
R.F.C.
500pF
10pF
·01µF
Pencil Lines

Using the radio

Operating a regenerative radio requires care at first because the trimmer has to be adjusted to suit the strength of each signal. Begin by connecting about 3 yards of wire to the radio as an aerial; no earth is required. Set the tuning capacitor about midway and screw in the trimmer until the radio begins to howl. Unscrew the trimmer until the noise just stops, and try to tune in a station. The trimmer should then be adjusted until it is just short of the point when the radio howls. The general method for receiving any station is to tune it in with the tuning capacitor and to adjust the feedback trimmer until the howling just disappears.

The feedback not only makes the radio more sensitive but also the tuning becomes sharper, so the tuning knob must be rotated very slowly. This sensitivity makes the radio ideal for the medium wave band, which is rather overcrowded. The tuning should be sharp enough to tune in individual stations without the background sounds of other stations. In areas of good signal strength, it will be possible to use the radio without an aerial at all, merely relying upon the signal picked up by the ferrite rod coil.

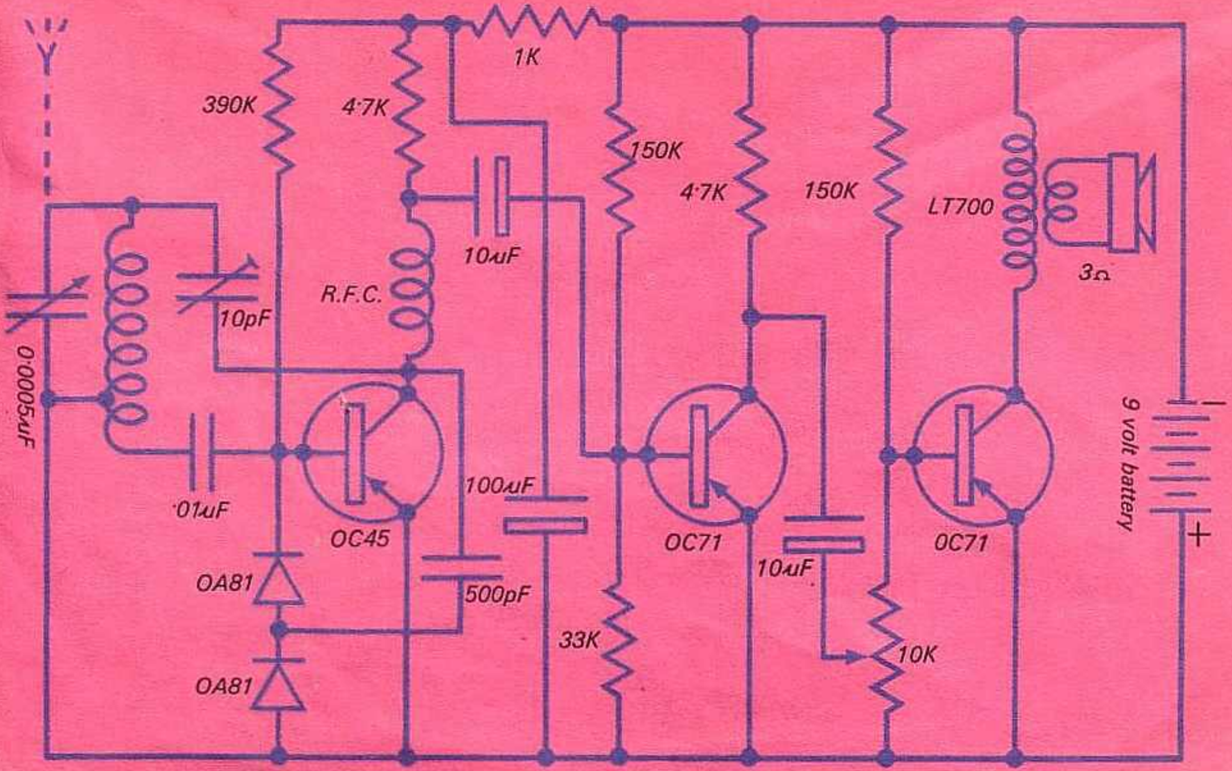
1K
390K
4·7K
150K
4·7K
150K
LT700
10µF
3Ω
10pF
R.F.C.
0·0005µF
·01µF
OC45
100µF
OC71
OC71
9 volt battery
−
+
OA81
500pF
10µF
33K
10K
OA81

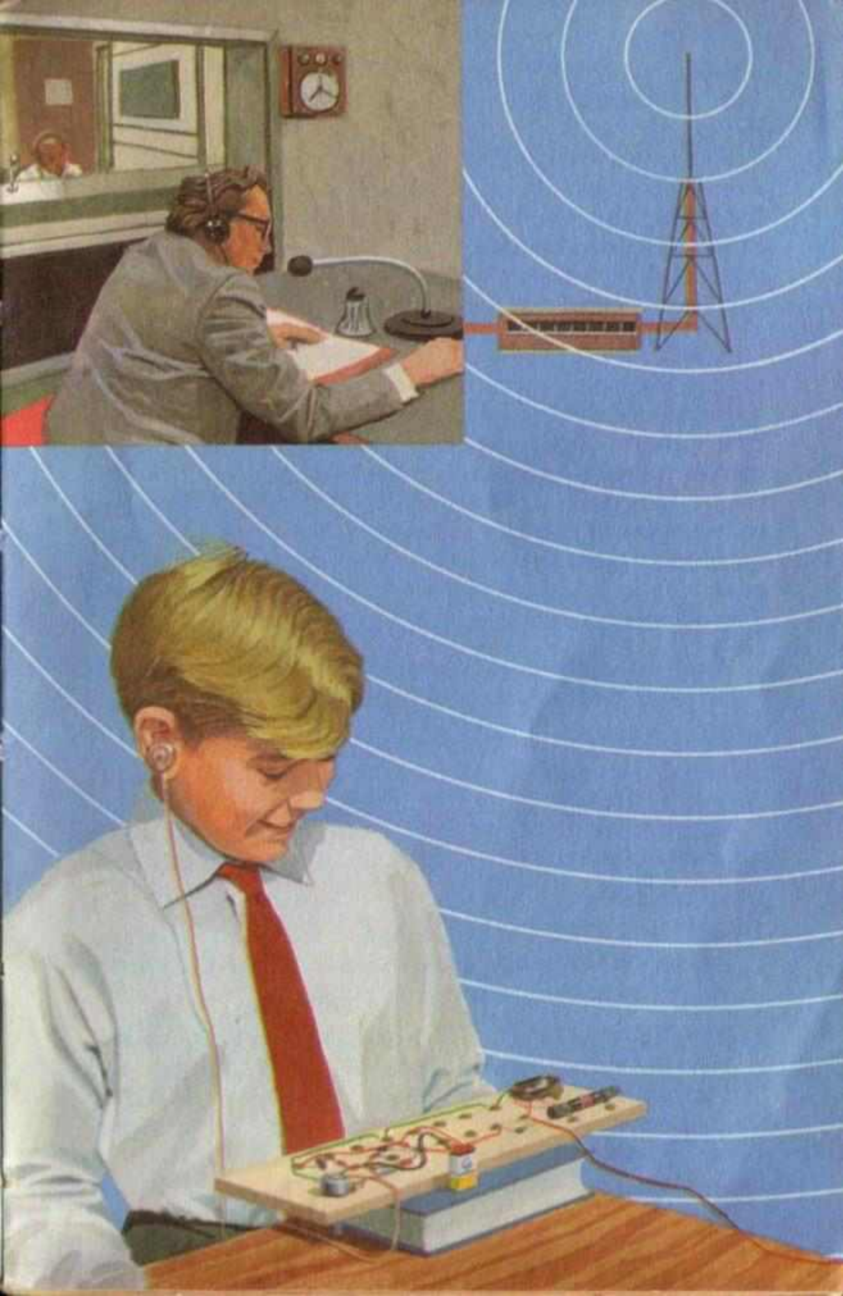

The completed radio

Now that the radio is completed, you may wish to make the circuit board into a simple case. The circuit board is used as a front panel, and end and bottom panels are added.

These can be made from similar wood to the circuit board (4 inch by ½ inch planed softwood). The bottom panel is the same size as the circuit board and the measurements for the end panels are shown opposite. The end panel next to the output transformer may be used to hold the loudspeaker. A series of ⅜ inch holes are drilled with a brace and bit to allow the sound to leave the case, and the speaker is secured with screws and cups. The panels are joined with panel pins and glue.

The battery may be firmly mounted above screws D and E on the circuit board, using 2½ inches of ¼ inch elastic and two screws and cups. Two smart plastic pointer knobs and a tuning scale complete the front panel. The tuning scale is a paper gummed label which can be marked with the positions of the main stations.

If you have followed these pages carefully, you will now have a useful working radio. You will also have shared the thrill of thousands of other radio constructors, listening to sounds coming from a radio receiver which is your own work.

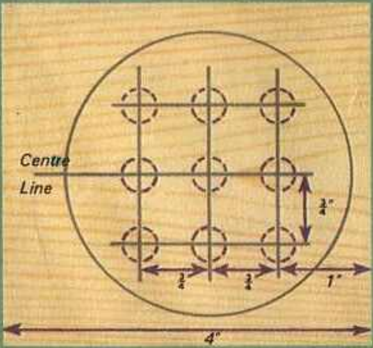
Centre
Line
$\frac{3}{4}$"
$\frac{3}{4}$
$\frac{3}{4}$
1"
4"

How to mark out the position of the Loudspeaker and $\frac{1}{8}$" holes

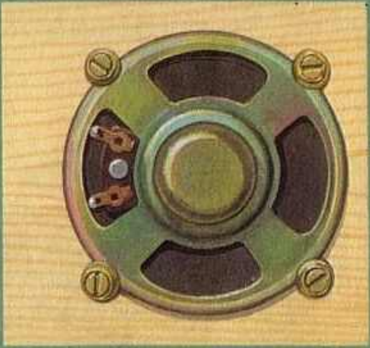

The Loudspeaker is mounted over the circle with 4 screws and cups

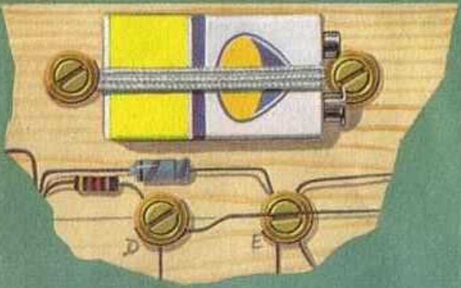

The Battery is mounted above screws D and E

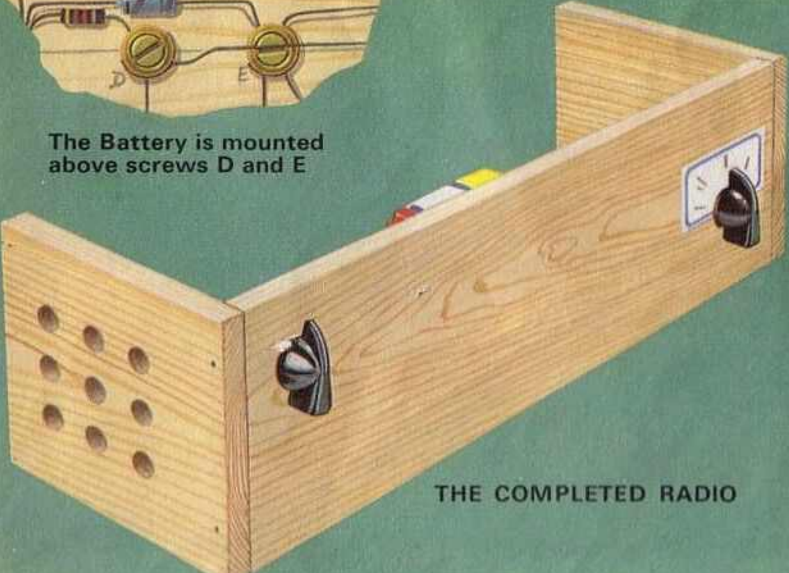

THE COMPLETED RADIO

Radio reception

Radio waves cannot be seen or heard. They are *electro-magnetic* waves, and they flow backwards and forwards (*oscillate*). Each backward and forward movement is known as a *cycle*. The rate of this backward and forward movement (*oscillation*) is known as the frequency of the wave.

A radio wave is also a *carrier* wave. It can 'carry' an electro-magnetic wave of a frequency exactly similar to the sound waves of the speech or music which is being transmitted. This frequency, which is imposed on the carrier wave, is known as an *audio* frequency. It is this electro-magnetic *audio frequency* that a radio receiver translates into sound via the loudspeaker. For a more detailed explanation of this, you might like to read pages 20-26 of the Ladybird book: 'How it Works—Television'.

When radio waves carrying an audio frequency arrive at a receiver, three things happen:

1. *TUNING*. The radio frequency carrier wave is picked up by the aerial. Along this same aerial pass the radio frequency carrier waves from many other transmitting stations, but the signal we want from a particular transmitter must be selected (tuned) from them.

2. *DETECTION*. The next stage is the detection stage, in which the electro-magnetic audio frequency wave is detected from the carrier wave.

3. *AMPLIFICATION*. The electro-magnetic audio signal wave is amplified, or made bigger, and translated from an electrical signal into SOUND by the loud-speaker, from which we hear a reproduction of the transmitted sound.

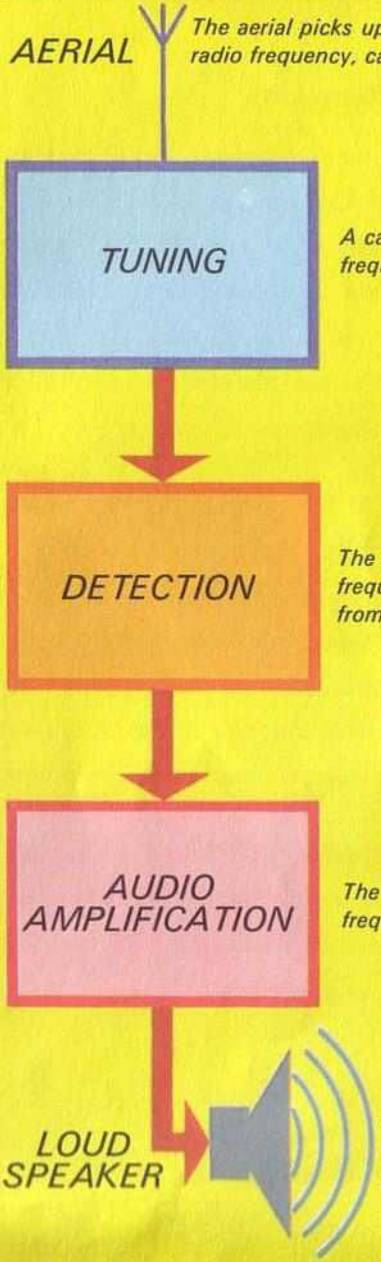
AERIAL
TUNING
DETECTION
AUDIO AMPLIFICATION
LOUD SPEAKER

The aerial picks up electro-magnetic, radio frequency, carrier waves

A carrier wave of a particular frequency is selected

The electro-magnetic, audio frequency wave is detected from the carrier wave

The electro-magnetic, audio frequency wave is amplified

The electro-magnetic audio frequency wave is converted into a sound wave

Reading circuit diagrams

Before building a piece of electronic equipment, a constructor needs to know how to wire up the various components. This can be done with the help of a diagram showing how the components are joined. Electronic engineers have developed a 'shorthand' method of showing the various components and their joining wires. This is the *circuit diagram*.

A circuit diagram is like a map. All the electronic components have symbols, and the connecting wires are shown as straight lines. The positions of the components on the circuit diagram do not normally correspond to their *actual* positions in the completed equipment. Sometimes a *layout diagram* is also used to show where to position the components on a base board.

Many symbols are used in circuit diagrams and we will learn them as we meet them in the course of our construction. To make construction easier we will use both circuit and layout diagrams.

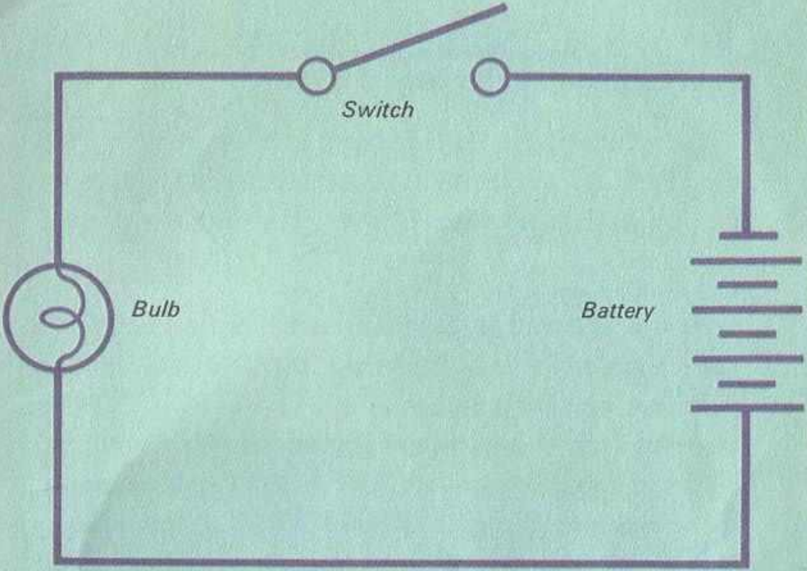
Switch
Bulb
Battery

CIRCUIT DIAGRAM

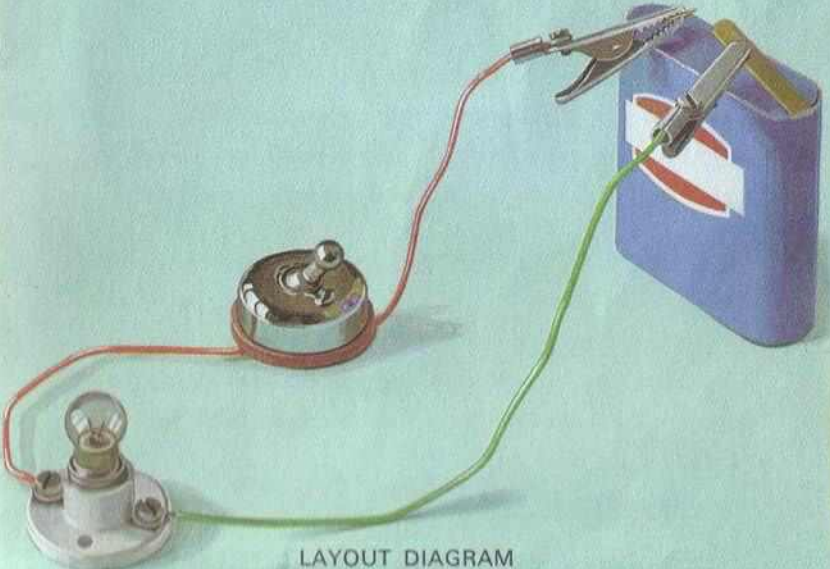

LAYOUT DIAGRAM

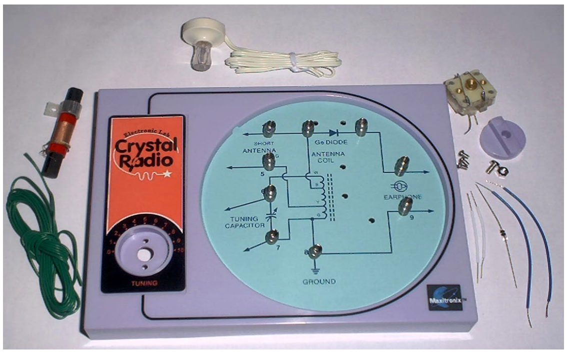

CRYSTAL RADIO
MX-901

PARTS LIST

- Antenna Coil
- Plastic Case with Labeled Panel
- D iode
- Earphone
- Anten na Coil Cl¡p
- Tunrng Capac tor Knob
- Nut, (1 ea)
- Screw smali, {3 ea.)
- Screw, medium (1 ea.)
- Spíng Terminals (9 ea.)
- Tuning Capacilor
- Con necting Wrres:
- B ue. Long (2 ea.)
- Whlte. Short (2 ea.)
- Antenna Wire, Green (3 N,4eters)

TOOL YOU MAY NEED

- Long-nose pliers
- Wire,cutters
- Ph llipslype screwdrlver
- Pencil or ballpoint pen

INTRODUCTION

With your "Crystal" Fadio Kit you can tune in AfV radro stattons without using batteries or AC powerl This device works on the same principle as the early crystal rad¡o, but the crysta has been replaced with a more reliable modern device called a d¡ode. Your C'ysta, Bad o Kit 'coü.es ro p.evious experience with electronics. You can build and operate it by fo lowing the simple instructions in this manual.

ASSEMBLING YOUR CRYSTAL RADIO

Follow the instructions in the order given, beginning wlth Step 1 .

Notes:

Don't rush. Take your time and enjoy bu¡ld ng your KIl,
You can check off each step as you finish it to besure yo{r don't forget an},thing.a
Allthrough these instructions you're asked to make various conr'reclrons. Secure electrical connections are very mponant In eiectronic devices. One poor connect,on can keep your crystal set from
Be sure all your connections are good ones
To con¡ect a wjre 10 the sprjng terminals:
A - Bend ihe spring to one side with your f nger.
B - Insert the end of the wire into one of the gaps i n : t r e spring.
C - Re ease the sp|ng. I t wi i hold the wire tirrn ly.

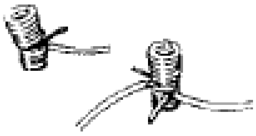

Be sure you connect clean. bare wire to each terminal. Some of the wires have insulation that has been stripped of the ends. Other wires have been "t¡nned" (coated with solder) at the ends so that they can make a good connection.
Nake sure lhat you insert the bare end of the wire into the spring terminal Be careful to keep the nsu,ated part out of the spring.
If a w¡re has a shiny t¡nned part at the end, Insert the shiny part into the spring.
Step - 1 Using your pencrl or pen, punch out the nine larger holes and the tour srnaller holes in the panel.
Step - 2 Copy the term¡nal numbers lrom the frontof the panel to the corresponding positions on the back side
Step - 3 Install the spring terminals :
A - Press the smaller end of each sprlng terminal into one of the punched holes.

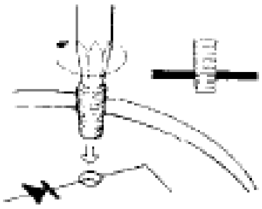

B - Use the pointed end of a pencil to tw¡st each term nal firmly into place.
Step - 4 Mount lhe diode as follows:
A - Use your long-nose pliers to bend each of the diode's tvr'o wires as shown.
B - Inserthe wire from the diode's stnped (cathode)end thro!gh lhe hole nearest Terminal 3.
C - Insert tnc ,/rire lrom lhe diode's other (anode) end through the small hole nearesl Terminal 2.
D - Attach the wlre from the diode's strlped end to Terminal 3
E - Attach the wire from the d¡ode's other end to Terminal 2.

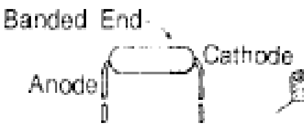

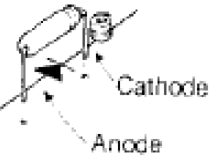

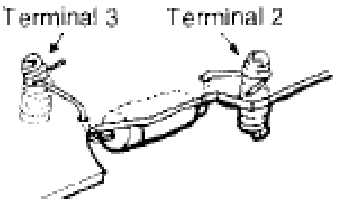

Note: A diode allows currento pass through it in only one direciron, from anode to cathode. fhis ability to block alternating current and pass direct current makes it useful rn many types of electron¡c c¡rcu¡ts.

Slep – 5 Mount th tuning capacrlor :

A - Position the capacrtor beneath the panel opening that is just above the word TUNING.
B - The metal shaft ol the capacitor should extend through the openrng.
C - Faslen the capacitor to the panel using the two srna screws.

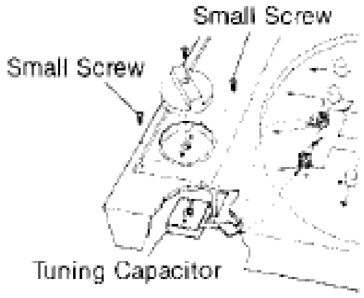

Step - 6 Install the luning capacitor's knob :
A - Turn the tuning capacitor's shaft all the way to the left.
B - Piace the tunjng knob on the upper end of the shaft
C - Llne up the dot on the knob wlth the 0 on lhe paner.
D - Faster :he knob in place using the reinaining smalll screw.

Step - 7 Connect the tuning capacitor's wires:
A - Connect the wire from the tuning capacitor's langet silver strip to Terminal 6.
B - Connect the wire from the tuning capacltor's sho,lcr silver strip to Termina 7.

Slep - 8 Install the antenna coil :
A - Position the antenna coil on the panel near the three rernaining small punched holes.
B - lnsert the antenna coil's green and yellow wires through the hole nearest the Y on the diagram.
C - insen the black and white wires through the hole nearest the W on the diagram.
D - Fasten the antenna coil in place wiih lhe nylon clip, medium screw, and nut, using the last of the small punched holes.

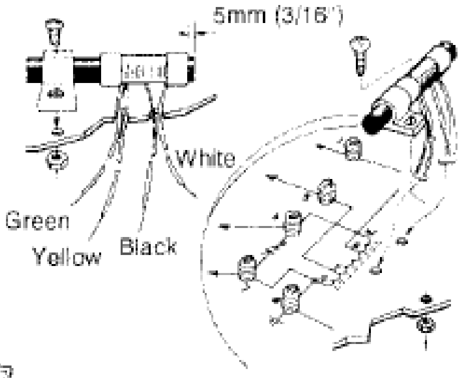

Step -9 Connect the antenna coil wires:

A - Connecr the black wire to Terminal 2.
B - Connect the yel¡ow wjre to Terminal 5
C - Connect the while wire to Terminal 6.
D - Connect the green wire to Terminal 8.

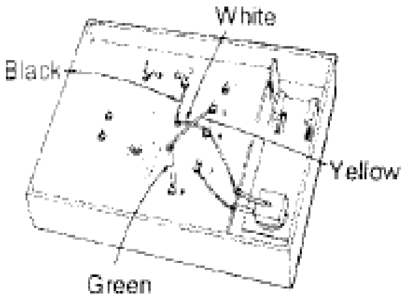

Step - 10 Instafl the blue and white connecting wires:

These connecting wrres can be instailed either on the front or the back of the panel. We suggest that you pul thern on the front, so that you can see how the wiring relates to the diagram on the panel.
A - Connect a short wire between Termrnals I and 2
B - Connect a lo¡g wlre between Terminals 3 and 4.
C - Connect a short wire belween Terminals 7 and 8 .

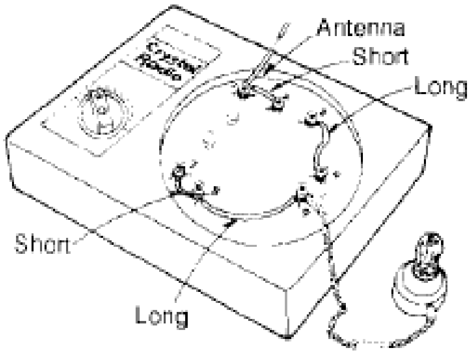

Step - 11 Use your wire cutters to neatly trim the ends of all the w¡res.
Step - 12 Connect the earphone wires to Termrnals 4andL

Step - 13 Install your ground wire:
The ground wire is an lmportant part ot your crystal radjo. Conneci it as caretully as you have connected ail the other parts.
A - From the co¡l of green wire, cut off about 5 feet (150crn)
B - Remove about 6 inches (15cm) of insuiation from one end of the wrre.
C – Find a metal cold water pipe
D - Scrape any paint or d¡rt from the pipe untll you see bright metal al the way around the pipe.
E - Wrap the bare end of the wire several ttmes around the Scraped part of the pipe and twrst the w¡re tightly.
F - Remove % Inch (7mm) of insulatron irom the other end of the ground wire
G -Connect this end of the ground wire to Terminal B of your radio.

Step - 14 Install your antenna:
Your antenna ¡s a very impoftant part of your crystal radio. Connect it as carefully as you have connected all the other parts.

A - Flemove % inch (7mm) of Insulation lrom one end of the remaining length of green wire.
B - Connect this end to Terminal 1 oi your radio.
C - Extend the anlenna w re horizontaily to its full length.

You have oow completed the building of your crystal Iadio. If yoLr have followed the instructions carefully, you should be able to tune n one or more AM,4 radio statrons right away.

USING THE CRYSTAL RADIO

A – Place the earphone in your ear.
B - Adjust the tuning capacitor's knob for the Clearest reception of a local Station.
C - Remember that your crystal radio has no amplifier, so the stations you receive might not be as loudi as if you were listening to a modern radio.

Notes:
- If you ive rn a city, you might be able to receive several stat¡ons.
- Il you do not live near a radlo station, you might need an outside antenna (not supplied), 35 io 85 feet (101o 25m) long, connected to Terminal 5.

lf your radio does not receive stations well, try connecting the antenna w¡re to term nal 6 instead of terrninal 1 or 5. If reception is still poor, change the antenna coil connection as follows :

White wire from terminal 6 to terminal 2
Black wire frorn terminal 2 to terminal 6

HOW YOUR CRYSTAL RADIO WORKS

All radio broadcast stations do basically the same thing. Thcy combine sound, or audio waves with a radio carries wave. The carrier wave travels great distances. and carries with its information about the strength and pitch of the sound waves

One method of combining the radio carrier wave and the audio wave is called Amplitude Nodulation (AM)

Your AM radio antenna prcks up tne carner wave seni by the radio statlon.

Your crystal radio then does three inlportant thlgs (with sorne help from you) to allow you to hear the radio program. These are:
- Tuning
- Detect¡on
- Changing of an electrical current to sound waves

TUNING

When you turn the tuning capacitor's knob, you are adjusting a circuit formed by the tuning capacitor and the antenna coil. This circult allows only one radio station's carrier wave at a time to enter your radio.
By tuning this circuit, you se{ect the carr er wave lo the station you want to hear'

DETECTION

The d¡ode gets rid oi the carrier wave by sendlng it to the ground through your cold water pipe At the same time. it allows a tiny electric current that represents the sound (audio) inlormation to go to the earphone.

CHANGING ELECTRICITY INTO SOUND

When the electric current reaches the earphone, It causes a small piece of ceramic material to vibrate. The movement of this ceramic to vibrates the crating the sound waves that you hear.

Your crystal radio works very much the sarme way as the early ones. Of course, the “old-time” crystal sets didn't have diodes. They used a piece of galena crystal (lead ore) held in contact with a fine wire called a “cat's whisker.”

This earlier type of radio, like yours, had no amplification. Then as now, more complicated radios amplified the audio waves to vibrate a loudspeaker and produce ouder sound.

All radios slill use the basjc combination of (1) a tuned circuit that selects the carrier wave sent by tne radio station, (2) a detector that separates the sound inforrnatror from the carrier wave and (3) an earphone or loudspea ker yo let you to hear the sound

SCHEMATIC DIAGRAM

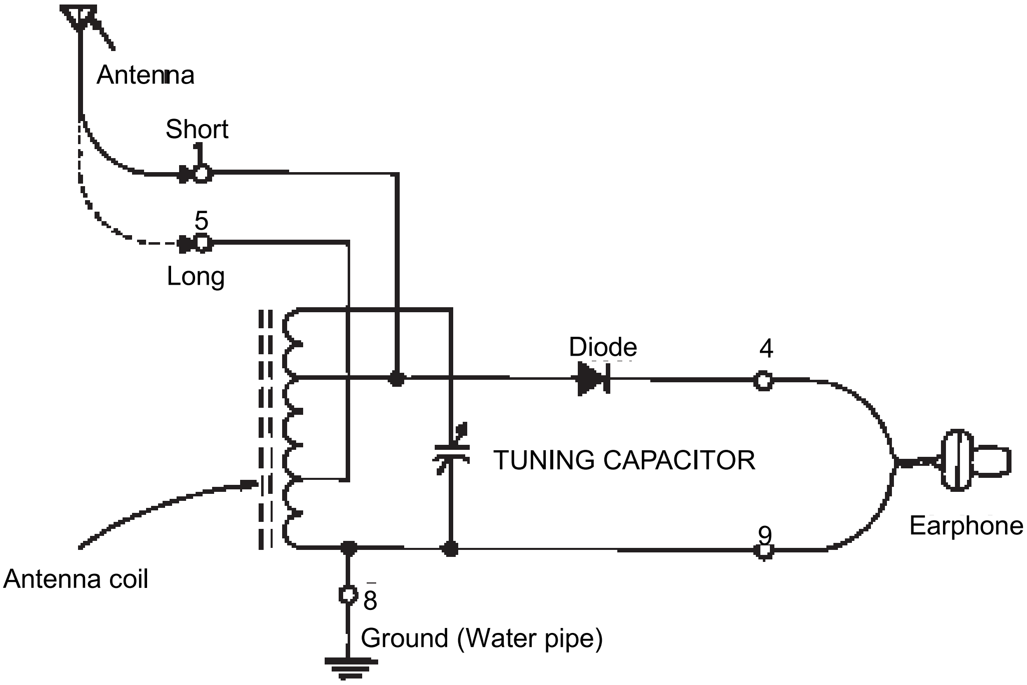
Antenna
Short
1
5
Long
Diode
4
TUNING CAPACITOR
Earphone
9
Antenna coil
8
Ground (Water pipe)

KIT RECOMMENDED FOR CHILDREN From 8 YEARS OLD, ACCOMPANIED BY ADULTS

A Pretty Good Crystal Set – Mark II

By Al Klase, N3FRQ, http://www.skywaves.ar88.net/

This is a revised version of the original New Jersey Antique Radio Club PGXS with minor changes to improve performance with short antennas in strong-signal areas. This radio will give a good account of itself on the AM broadcast band.

OK! It's time to build a crystal set. This time it's going to be a good one.

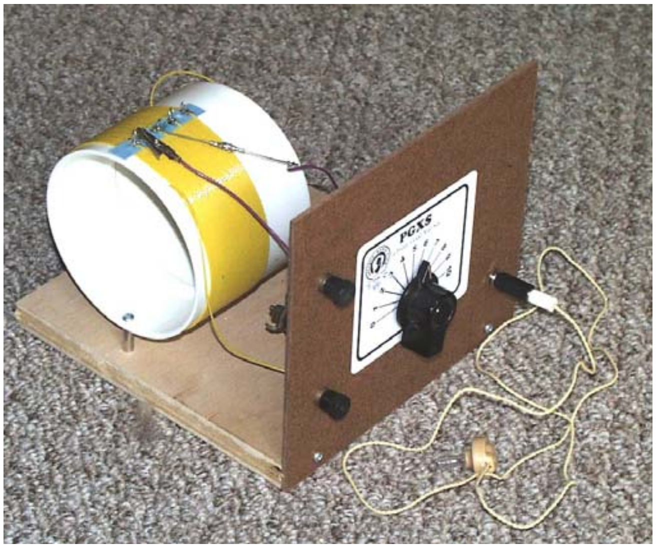

Countless millions of crystal radios have been built over the last 100 years, but the performance of most of these leaves a lot to be desired. The design of this radio goes back to about 1990 when I was a den leader for my son's Cub-Scout pack. I wanted the guys to have the experience of listening to a radio they had built with their own hands. However, we were in the deep suburbs of Philadelphia (actually half way to Reading). There were no strong local stations, so an average crystal set combined with the sort of makeshift antenna most families would erect, wasn't going to get the job done. The result was the Den Two Crystal Radio. (See "Crystal Set Projects," The Xtal Set Society, 1997.) This radio was based on a spider-web coil that 10-year-old hands could wind with a little supervision. It was sensitive enough to hear distant sky-wave stations at night, and selective enough that you could usually listen to just one station. Seven were constructed and at least six of those went into service at home. Everyone was duly impressed. We even had reports of hearing Radio Japan! (That was from Ontario, however.) We'll use the same circuit with a better solenoid-style coil.

Selecting Components

First, let's deal with the headset. Look around for a traditional pair of 2000-ohm 'phones. That's the DC resistance, they'll have an AC impedance of about 10,000 ohms. Test the 'phones by putting them on your head, holding on to one of the tips and touching the other to an electrical ground such as a water pipe or the screw on an outlet plate. You should hear a click or maybe a hum. If not, they probably won't be sensitive enough for our purposes. Check the DC resistance with an ohm meter, it should be in the range of 1-5 K-ohms. Sometimes the cords are open.

An acceptable alternative to a vintage headset is one of the crystal (ceramic) ear plugs available from Mouser.com for about two bucks. It's best to install a 50-100K resistor in parallel with these capacitive devices to provide a DC load for the detector to eliminate audio distortion.

Modern low-impedance headsets with a matching transformer might work if you have a lot of signal strength, but most are designed to sacrifice sensitivity for flat frequency response.

Next comes the detector. Stick to a germanium diode or the base-emitter junction of a germanium transistor for openers. A digital voltmeter in the diode position will indicate the forward voltage drop. Germanium devices will show 0.2-0.3 while silicon will indicate 0.5-0.7 volts. You can use a modern silicon device, but you'll be giving up a lot of sensitivity. The basic germanium diode is the 1N34A, but the part number doesn't mean much. They vary widely. Dump out the junk box, try them in a radio, and pick the best one. You can move on to mineral detectors and cat's whiskers for a little nostalgia after you have a working radio.

Now you'll need an antenna, the higher, longer, and more in the open the better. Sixty feet long and forty feet high would be excellent, but wire in the attic will probably suffice, especially if you're a "city mouse." A wire to a hose clamp around a water pipe or a connection to the screw on an outlet plate (AC safety ground) will complete the antenna-ground system.

Start Simply

Now it's time to build the simplest possible radio: Clip-lead the diode across the headset and then connect one side to the antenna and the other side to ground. You should hear something, quite possibly three stations at once. This will verify that you have a usable headset and detector, and that you have enough antenna for the available signals.

Construction of the PGXS

To make this kludge into a serviceable radio we'll add a resonant "tank" circuit consisting of a coil and a variable capacitor to allow us to select the station we want to hear. The coil has multiple taps so it can serve as an autotransformer to achieve a good impedance match, allowing optimal energy transfer, between the antenna-ground system and the tank circuit as well as between the tank and the detector and it's load, the headset.

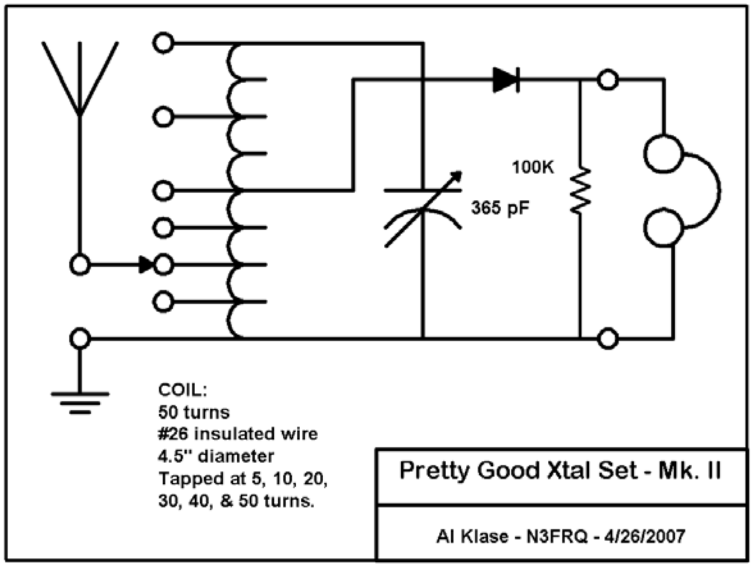

The target value for the cap is 365 pico-farads (mmF or pF), but a 500 pF unit as found in some of the old battery sets will work nearly as well. You could also use the large half of the two-gang cap from a junked all-American-five radio, etc.

The coil is wound on a 4-inch ABS pipe coupling from the “home center.” Wire is generally expensive and hard to buy retail. You can use #20 insulated bell wire, also from the “home center” because it is easily available. The down side is that it’s twisted pair that needs to be untwisted, and they make you buy 500 feet, for about 30 bucks, when you only need 85 feet for the project. Any plastic insulated wire between #26 and #20 will do, but you may need to adjust the total number of turns after you get the radio working to cover both ends of the broadcast band. The best option, for maximum performance, is silver-plated Teflon-insulated hook-up wire if you have any in the junk box, but even common vinyl-insulated tinned stranded wire will work just fine. Close-wound enameled magnet wire is a bad bet, as eddy currents in the closely spaced turns cause considerable loss.

Winding the coil: Tie one end of an 85-foot length of wire to a solid support. Drill two small holes about 3/8 of an inch from the end of the coil form. Keep the drill handy. Put your wire cutter and a strip of thin cardboard about 3/8 inch wide and 3 inches long in your shirt pocket. Lace the end of the wire into one hole and out the other, leaving about a foot protruding from the form. Grasp the coil form firmly and pull the wire taught against the tied off end. Wind 5 turns on the coil keeping the wire close-wound and tight. Place the end of the cardboard strip under the 5th turn. Wind four more turns, sneaking them under the strip. Wind the 10th turn over the strip. Continue winding placing the 20th, 30th, 40th, and 50th turns on top of the cardboard. Holding the wire firmly in place with your left thumb, clip the wire leaving about a foot to terminate the coil. Drill two more holes in the form just beyond where the 50th turn crosses the cardboard, and lace the end of the wire through them.

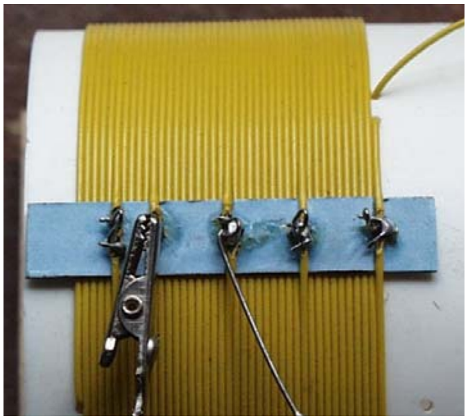

Use a utility knife to whittle away the insulation where the turns cross the cardboard. Wrap and solder a piece of bare wire at each of these sites to form the coil taps.

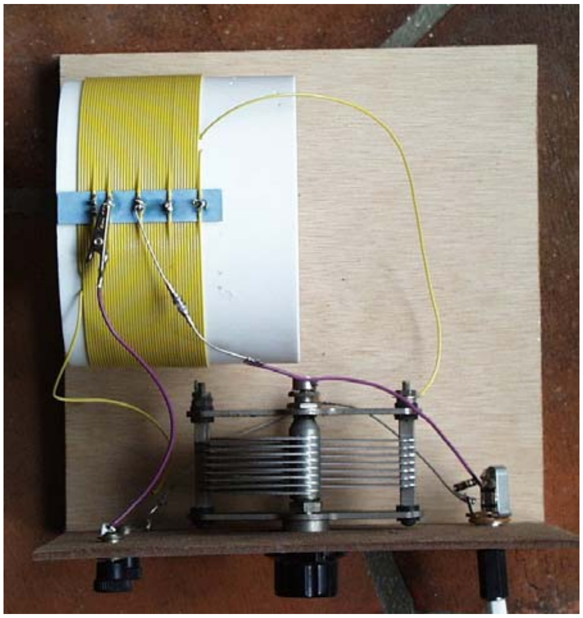

"Beadboard" the radio with clipleads or tack-solder joints to make sure everything works before committing to a physical design.

The radio in the picture is built on an 8-inch square piece of ¾" plywood with an 8 x 6-inch front panel of tempered hardboard. An alligator clip is used to connect the selected coil tap to the antenna terminal. The 100K-ohm diode load resistor is installed on the ¼-inch phone jack. An adaptor is used to plug in the ceramic earpiece. Print out and paste the dial scale to the front panel. A large knob with a pointer is the way to go. You can install a rotary switch for the coil taps if you want to get fancy. Layout is not critical, but keep the coil in the approximate orientation shown so it can be coupled to the antenna tuner to be described in "Phase II."

Operation

Connect the antenna, ground, and headset. Put the alligator clip on the middle tap where the diode connects. Tune in a station. Move to another tap, and retune looking for the best performance. The optimum coil tap will be dependent on the frequency your tuning.

In strong signal areas, you can just connect your ground wire to the top (right most) tap. The headphone cord and your body will serve as the antenna.

A reasonably high 50-75- foot outdoor antenna should allow you to hear night-time stations hundreds of miles away.

Some Helpful Links

The author's website: **http://www.skywaves.ar88.net/**

The Radio Technology Museum: **http://www.rtm.ar88.net/**

The New Jersey Antique Radio Club: **http://www.njarc.org/**

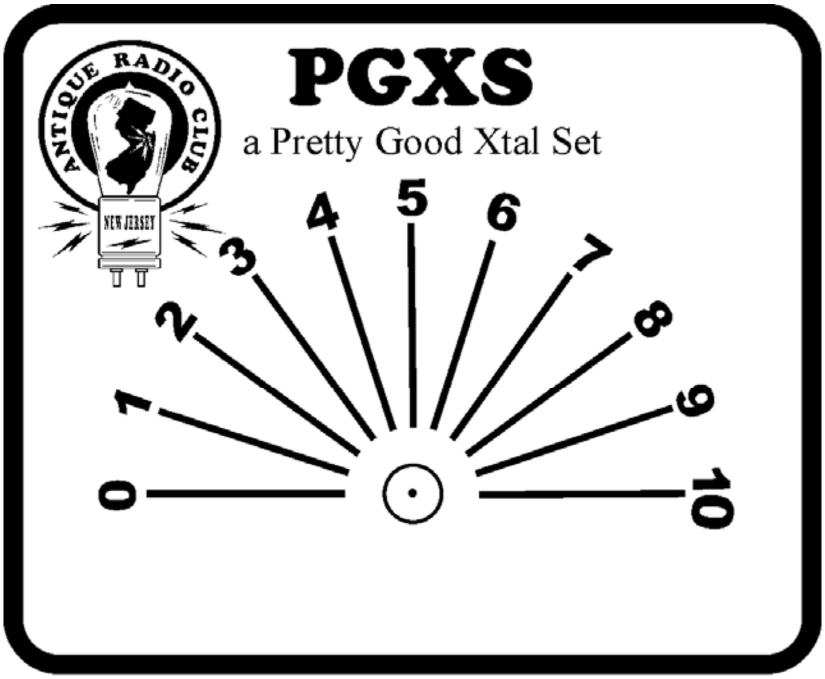
ANTIQUE RADIO CLUB
NEW JERSEY
PGXS
a Pretty Good Xtal Set
0
1
2
3
4
5
6
7
8
9
10

Figure 1 - Dial Artwork. Print on label stock.

A Pretty Good Crystal Set

By Al Klase, N3FRQ, http://www.skywaves.ar88.net/

OK! It's time to build a crystal set. This time it's going to be a good one.

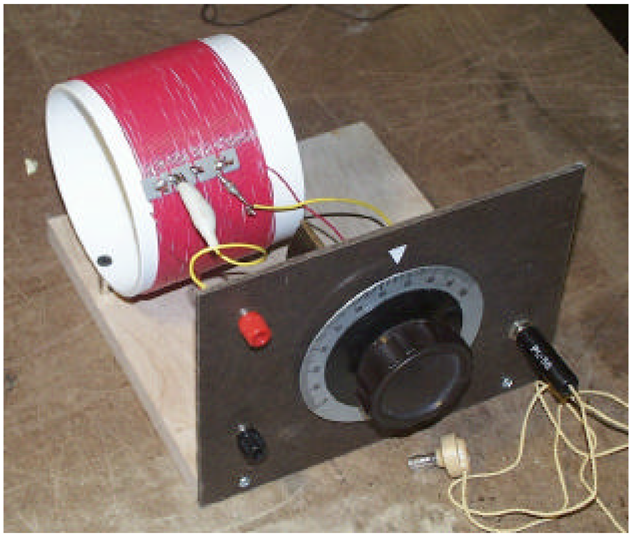

Countless millions of crystal radios have been built over the last 100 years, but the performance of most of these leaves a lot to be desired. The design of this radio goes back to about 1990 when I was a den leader for my son's Cub-Scout pack. I wanted the guys to have the experience of listening to a radio they had built with their own hands. However, we were in the deep suburbs of Philadelphia (actually half way to Reading). There were no strong local stations, so an average crystal set combined with the sort of makeshift antenna most families would erect, wasn't going to get the job done. The result was the Den Two Crystal Radio. (See "Crystal Set Projects," The Xtal Set Society, 1997.) This radio was based on a spider-web coil that 10-year-old hands could wind with a little supervision. It was sensitive enough to hear distant sky-wave stations at night, and selective enough that you could usually listen to just one station. Seven were constructed and at least six of those went into service at home. Everyone was duly impressed. We even had reports of hearing Radio Japan! (That was from Ontario, however.) We'll use the same circuit with a better solenoid-style coil.

First, lets deal with the headset. Look around for a traditional pair of 2000-ohm 'phones. That's the DC resistance, they'll have an AC impedance of about 10,000 ohms. Test the 'phones by putting them on your head, holding on to one of the tips and touching the other to an electrical ground such as a water pipe or the screw on an outlet plate. You should hear a click or maybe a hum. If not, they probably won't be sensitive enough for our purposes. Check the DC resistance with an ohm meter, it should be in the range of 1-5 K-ohms. Sometimes the cords are open.

An acceptable alternative to a vintage headset is one of the crystal (ceramic) ear plugs available from Mouser.com for about two bucks. It's best to install a 50-100K resistor in parallel with these 'phones to provide a DC load for the detector to eliminate audio distortion.

Modern low-impedance headsets with a matching transformer might work if you have a lot of signal strength, but most are designed to sacrifice sensitivity for flat frequency response.

Next, comes the detector. Stick to a germanium diode or the base-emitter junction of a germanium transistor for openers. A digital voltmeter in the diode position will indicate the forward voltage drop. Germanium devices will show 0.2-0.3 while silicon will indicate 0.5-0.7 volts. You can use a modern silicon device, but you'll be giving up a lot of sensitivity. The basic germanium diode is the 1N34A, but the part number doesn't mean much. They vary widely. Dump out the junk box, try them in a radio, and pick the best one. You can move on to mineral detectors and cat's whiskers after you have a working radio.

Now you'll need an antenna, the higher, longer, and more in the open the better. Sixty feet long and forty feet high would be excellent, but wire in the attic will probably suffice, especially if you're a "city mouse." A wire to a hose clamp around a water pipe or a connection to the screw on an outlet plate (AC safety ground) will complete the antenna-ground system.

Now it's time to build the simplest possible radio: Clip-lead the diode across the headset and then connect one side to the antenna and the other side to ground. You should hear something, quite possibly three stations at once.

To make this kludge into a serviceable radio we'll add a tuned "tank" circuit consisting of a coil and a variable capacitor. The target value for the cap is 365 pico-farads (mmF or pF), but a 500 pF unit as found in some of the old battery sets will work nearly as well. You could also use the large half of the two-gang cap from a junked all-American five radio, etc.

The coil is wound on a 4-inch styrene pipe coupling from the "home center." Wire is generally expensive and hard to buy retail. I've specified #20 insulated bell wire, also from the "home center" because it is easily available. The down side is that it's twisted pair that needs to be unwound, and they make you buy 500 feet, for about 30 bucks, when you only need 85 feet for the project. Any plastic insulated wire between #26 and #20 will do, but you'll probably need to adjust the total number of turns after you get the radio going to cover both ends of the broadcast band. The best option, for maximum performance, is silver-plated Teflon-insulated hook-up wire if you have any in the junk box, but even common vinyl-insulated tinned stranded wire will work all right. Close-wound magnet wire is a bad bet, as eddy currents in the adjacent turns cause considerable loss.

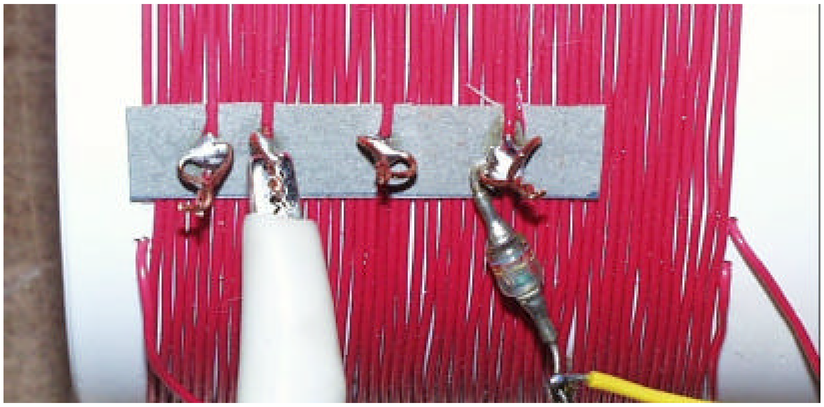

Winding the coil: Tie one end of an 85-foot length of wire to a solid support. Drill two small holes 3/8 of an inch from the end of the coil form. Keep the drill and wire cutters handy. Put a strip of thin cardboard about 3/8 inch wide and 3 inches long in your shirt pocket. Lace the end of the wire into one hole and out the other, leaving about a foot protruding from the form. Grasp the coil form firmly and pull the wire taught against the tied off end. Wind 5 turns on the coil keeping the wrie close-wound and tight. Place the end of the cardboard strip under the 5th turn. Wind four more turns, sneaking them under the strip. Wind the 10th turn over the strip. Continue winding placing the 20th and 30th turns on top of the cardboard. Wind 18 more turns for a total of 48. Clip the wire leaving about a foot to terminate the coil., drill two more holes in the form, and lace the end of the wire through them.

Use a utility knife to whittle away the insulation where the 5th, 10th, 20th, and 30th turns cross the cardboard. Wrap and solder a piece of bare wire at each of these sites to form the coil taps.

“Beadboard” the radio with clipleads or tack-solder joints to make sure everything works before committing to a physical design. The radio in the picture is built on an 8-inch square piece of ¾” plywood with an 8 x 6-inch front panel of tempered hardboard. Layout is not critical, but keep the coil in the approximate orientation shown so it can be coupled to the antenna tuner to be described in “PGXS – Phase II.”

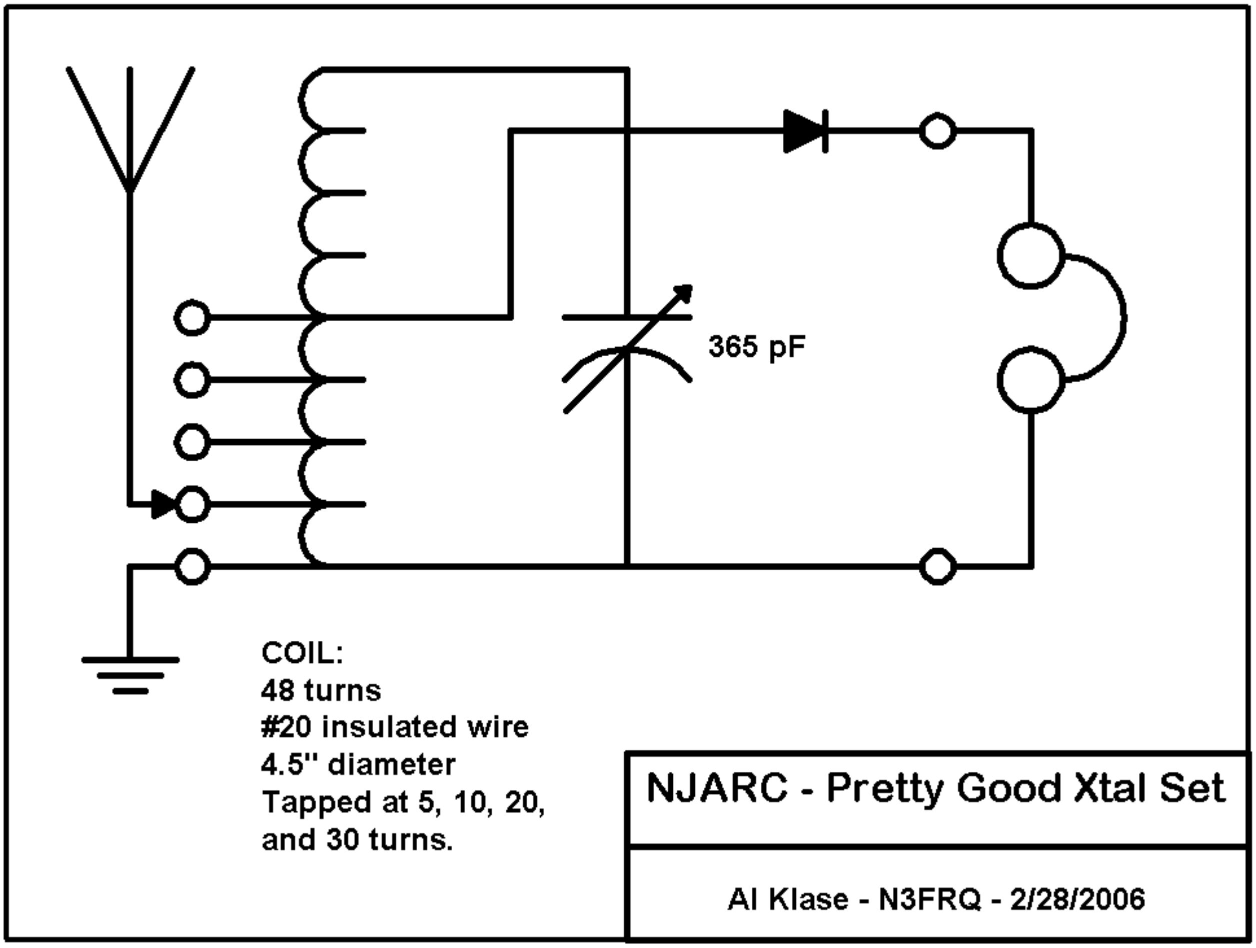
365 pF
COIL:
48 turns
#20 insulated wire
4.5" diameter
Tapped at 5, 10, 20,
and 30 turns.
NJARC - Pretty Good Xtal Set
Al Klase - N3FRQ - 2/28/2006

RCA RAB Receiver

An Exercise in No-Holds-Barred Receiver Design Ca. 1931

This is an ongoing investigation. Any information on the history of this radio would be greatly appreciated. This information is from the instruction manual dated 30 June 1931.

The United States Navy became one of the early adopters of long-range short-wave communications as a result of positive results obtained during fleet maneuvers in 1925. Early receiving sets were TRF-autodynes, but by the early 1930's the state of the art had progressed far enough to allow the design of a truly effective super-heterodyne for the HF range. The RAB was probably the first mass-produced short-wave superhet, and had a number of advanced features that would not appear in commercial sets for years.

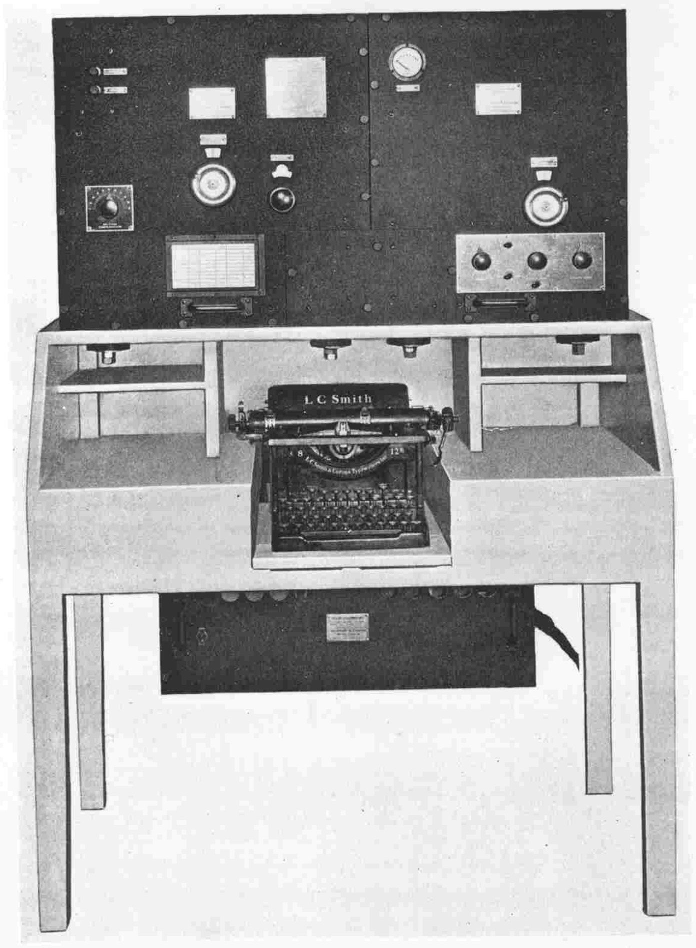

Frontispiece—Typical Installation Model RAB Equipment

Model RAB Radio Receiving Equipment

- Range 1,000 - 30,000 KC in eight band-switched ranges
- 2 Tuned RF stages for effective image rejection
- 2 IF stages with 8 tuned circuits for a modicum of selectivity.
- IF frequencies: 600, 1450, 3250, 7200 KC The higher IF's ease the image problem at higher frequencies, and widen the bandwidth to compensate for transmitter instability.
- IF frequency depends on band in use
- Remote-cutoff tubes in RF and IF for excellent manual gain control
- Regulated PS for LO and BFO
- Audio filtering for AM and CW
- Audio limiter (AVC)
- Weight: 455 lb (plus table and typewriter)

RF Tuner CRV-4552
(rear view)

RF and oscillator coils are mounted on the top and bottom of the main chassis.

Band selection is accomplished by a sliding assembly containing the five-gang tuning capacitor and tubes. This carriage is moved by a lead screw actuated by the milling-machine-like knob on the front panel.

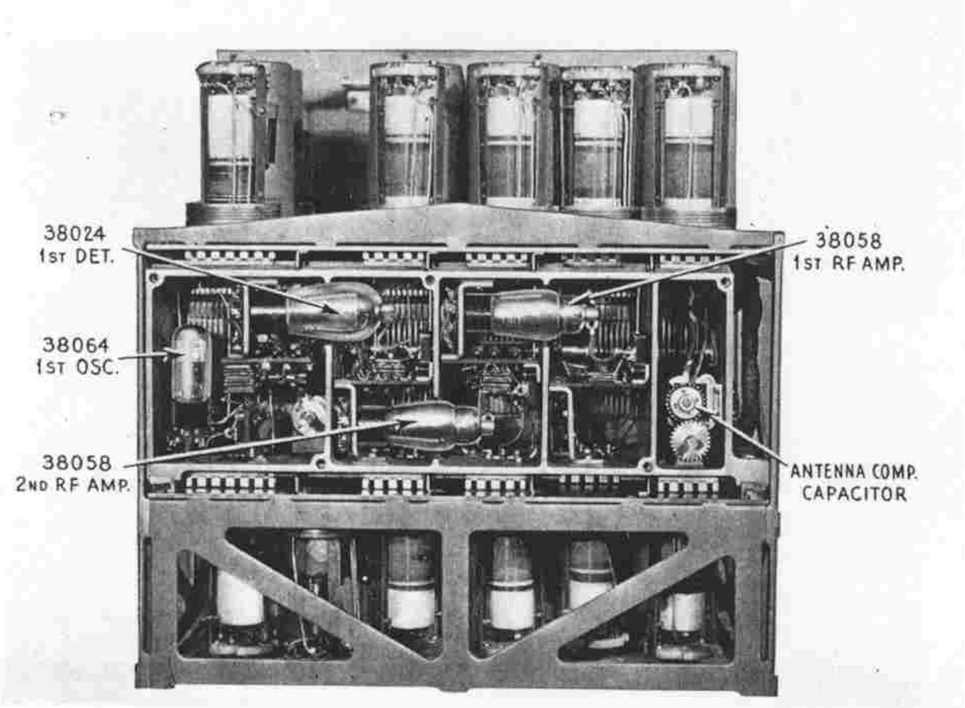
38024
1ST DET.
38058
1ST RF AMP.
38064
1ST OSC.
38058
2ND RF AMP.
ANTENNA COMP.
CAPACITOR

Figure 3—Radio Frequency Tuner CRV-4552
(Rear View Chassis—Carriage Cover Plate Removed)

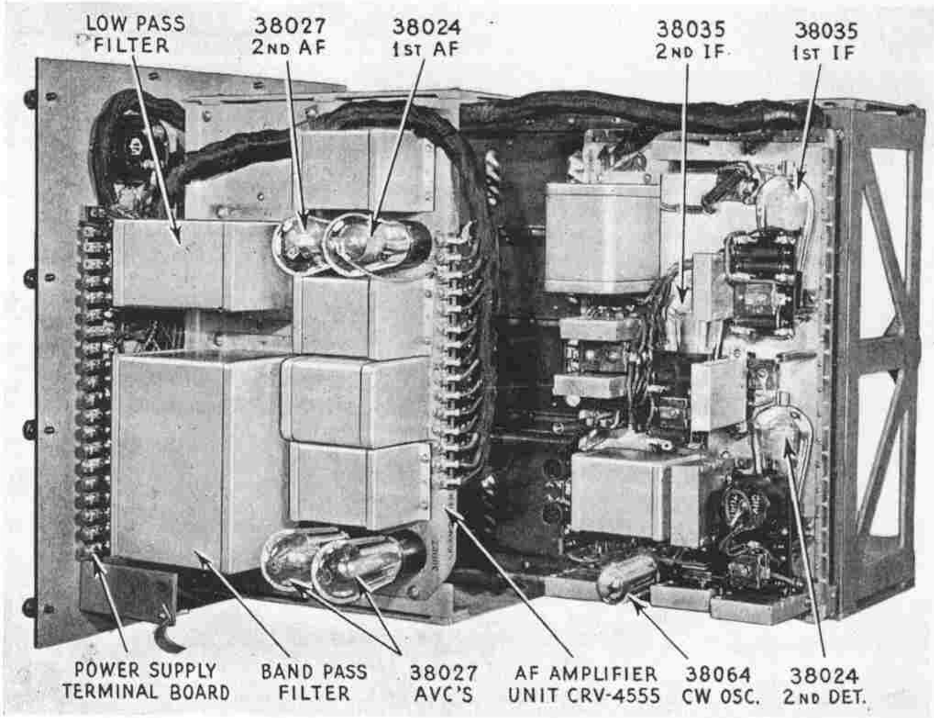
LOW PASS
FILTER
38027
2ND AF
38024
1ST AF
38035
2ND IF
38035
1ST IF
POWER SUPPLY
TERMINAL BOARD
BAND PASS
FILTER
38027
AVC'S
AF AMPLIFIER
UNIT CRV-4555
38064
CW OSC.
38024
2ND DET.

Figure 4—IF and AF Amplifier CRV-4553
(Top View Chassis—Carriage Cover Removed)

IF and AF Amplifier CRV-4553

Four sets of IF transformers and BFO coils are mounted on the underside. IF frequency selection is by way of a sliding chassis containing the tubes and support circuits.

RAB_schematic.gif (188KB)

RAB Manual (11MB)

BENJAMIN AND DAVID GOLDENBERG 'S VERY-FINE

OLD-TIME CRYSTAL RADIO

Parts List:

Item	Qty.
PVC pipe, 1" outside diameter (3/4” size) x 4" long	1
Aluminum or steel sheet metal, 0.5” wide x 5” long	1
Coil wire (#24 magnet wire)	36 ft
Hook-up wire (#22)	25 ft
Crystal earphone (GC electronics, cat. no. 30-5302)	1
Jack for the earphone plug	1
Crystal diode (1N34A)	1
Capacitor, 470 picofarad	1
#8 sheet metal screws	8
8-32 machine screw, 1" long	1
8-32 nuts	2
flat washers	6
Ring terminal	1
Board, 3/4" thick, approximately 7" x 9"	1

What else you need:

Vise-Grip pliers
Wire cutters
Soldering iron and solder
Drill, with 1/8", 3/16” and 3/8” bits
File or motorized grinder (e.g. Moto-Tool)
Screwdriver
Scotch tape

Instructions

I. CONSTRUCTION OF SPECIAL PARTS.

The following steps involve making a few special parts for the radio. Some of these steps are most easily performed using power tools and some involve soldering. This work should be done by someone old enough and knowledgeable enough to handle these tools safely. Once these parts are made, however, the rest of the project requires only hand tools and can be completed by nearly anyone.

1. Coil Form

The coil form is a piece of PVC pipe, 1” in diameter (outside) and 4” long. (For reasons fully understood only by plumbers, pipe of this diameter is referred to as 3/4”.)

To facilitate mounting the coil to the board, drill two sets of holes, on opposite sides of the form. On one side of the form, drill two holes, 3/16” in diameter and 3/8” from the ends. Directly opposite from these holes, drill two more holes 3/8” in diameter. Screws will be placed through the smaller holes, while the larger ones make it easy to reach the screws with a screwdriver.

2. Slider

Drill two 3/16” holes in the piece of sheet metal, each 1/4” from an end. After the holes are drilled, place a bend in the metal 3/4” from each end, as shown in the figure on page 3. Use a vise or pliers to bend the metal. Use a file to smooth all of the edges and corners of the metal.

CAUTION: Drilling sheet metal can be very dangerous! If the bit catches on the metal, you can have a very sharp edge flying through your hands. Hold the metal firmly with a pair of pliers (preferably Vice-Grip type), not with your hands! Also, place a piece of wood under the metal as you drill. This will help keep the metal from tearing and grabbing the bit as the passes through.

3. Pointed screw for slider

Using either a file or a grinder, grind a point on the 8-32 machine screw. The point doesn’t need to be very sharp; about like the tip of a medium ball-point pen is ideal.

4. Earphone jack with wires.

Cut two pieces of hook-up wire 2” long. Remove 1/4” of insulation form one end of each wire and solder the wires to the terminals of the earphone jack. Remove 1/2” of insulation from the other ends of the wires.

Note: Headphone jacks often have more than two terminals. Make sure that the terminals you solder to are the ones that will be connected to the tip and side of the plug when it is inserted.

5. Ring terminal with two wires.

Cut two pieces of hook-up wire, one 7” long and the other 2”. Remove 1/4” of insulation from one end of each wire. Solder the two wires to the ring terminal. Remove 1/2” of insulation from the other ends of the wires.

6. Board

Tape the layout template (the last page of this document) to a board, approximately 7” x 9”, and drill a 1/8" hole at each position marked with a cross.

II. ASSEMBLY

1. Wind the coil.

The coil is made by wrapping the #24 magnet wire onto the coil form, which is made from a piece of PVC pipe. The magnet wire is the wire that looks like it has no plastic insulation on it. But, it is actually coated with a thin layer of enamel. This coating must be removed from the ends of the wire before making any connections.

The best way to wind the coil is with two people, one to hold the wire taught (the holder) and the other to wind it onto the coil form (the wrapper). First, thread the end of the coil wire from the outside of the form through one of the large holes and out the end of the form. Leave about 4" of wire going out the end. Hold onto the other end of the wire and begin wrapping it around the form, starting right next to the hole. After you have wound a few turns on, push the wire together so it is nice and neat and put a piece of Scotch tape over it. The coil should look like this:

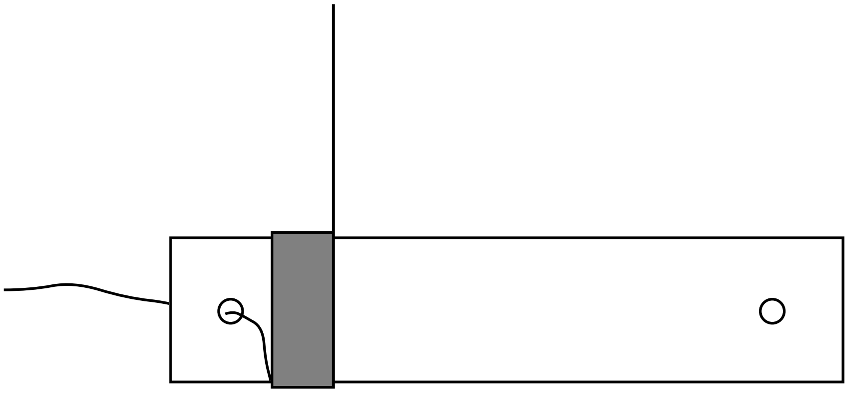

Now, the "holder" should hold the spool of wire loosely while the "wrapper" walks across the room unwinding the wire. Then, the holder should hold the spool very tightly so that the wrapper can wind the wire onto the pipe evenly. Each turn should be wrapped on right next to the previous turn, but shouldn't overlap. As the wire is wrapped on, the wrapper slowly walks across the room keeping the wire tight. When the wrapper reaches the holder, hold the wire that has been wrapped onto the form tightly, while the holder lets the wire on the spool unwrap. The wrapper then walks back across the room unwinding the wire from the spool, and the same procedure is repeated.

When the coil is within about 1/8" of the holes at the other end of the coil, stop winding. Put a piece of Scotch tape on the end of the wrapping, and cut the wire a few inches from the end of the coil. Thread the end of the wire through the large hole in the form.

2. Attach the coil to the board.

Use two of the #8 sheet metal screws to attach the coil to the board, using holes B and C that you drilled. Place the screws through the small holes in the coil form from the inside. You can then use the large holes on the other side of the form to reach the screws with a screwdriver. Make sure the coil is attached snugly to the board.

3. Assemble the slider.

Find the slider, the 8-32 screw and the two 8-32 nuts. Thread one of the nuts onto the screw until it is about 3/8" from the head of the screw. Put the screw through one of the holes in the slider and thread the other nut on, so that the slider looks like this from the side:

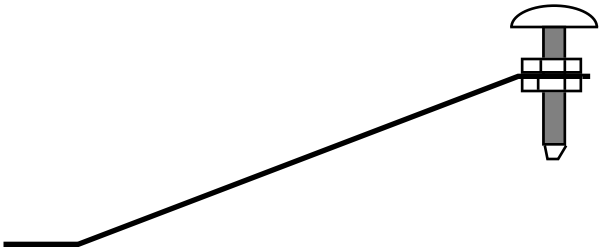

Tighten the second nut against the slider and the first nut with a wrench.

4. Attach the slider to the board.

Find a sheet metal screw, a washer and the ring terminal with two wires attached to it. Put the screw through the ring terminal, the washer and finally the open hole in the slider assembly (from the top of the slider as shown above). Attach the entire assembly to the board with the screw at hole E. Tighten the screw so that it is snug but still allows the slider to move back and forth. The pointed end of the machine screw at the other end of the slider should press firmly against the wire on the coil. If necessary bend the slider so that it makes firm contact with the coil but can still move.

Move the slider back and forth on the coil to scrape the enamel insulation from the coil wire where the pointed end of the machine screw touches it. The screw should be able to make an electrical contact with the coil wire at any position along the length of the coil.

5. Put the remaining screws in the board.

Screw the remaining sheet metal screws into holes A, D, F, G and H, with a washer on each screw. Do not tighten the screws yet.

6. Wire the components of the radio together.

Use the circuit drawing as a guide. For each connection, make sure that the insulation is removed from the end of the wire. Wrap the bare ends of the wires around the screw between the screw head and the washer underneath.

a. Wrap the two wires from the earphone jack around screws F and H. Do not tighten these screws yet.

b. Take the short wire from the slider and wrap the bare end around screw F (which should already be attached to one of the earphone jack wires). Do not tighten the screw yet.

c. Take the long wire from the slider and wrap the bare end around screw A. Do not tighten this screw yet.

d. Find the crystal diode, the small cylinder with two bare wires attached to it. Wrap one of the wires around screw H (attached to the other earphone jack wire), and tighten the screw.

e. Take the other wire from the diode and wrap it around screw G. Do not tighten this screw yet.

f. Find the capacitor, the small disc with two bare wires attached to it. Wrap one of the wires around screw F and tighten the screw.

g. Wrap the other wire from the capacitor around screw G, but do not tighten this screw yet.

h. Cut a piece of hook up wire about 6" long and remove about 1/2" of insulation from each end. Wrap one end around screw G and tighten the screw. Wrap the other end of the wire around screw D, but do not tighten this screw yet.

i. Take the wire from the right hand end of the coil and remove the insulation from about 1/2" from the end of the wire. To remove the insulation, either lightly scrape the wire with a knife or use a small piece of sandpaper. Wrap the bare end of the wire around screw D, but do not tighten the screw yet.

7. Connect the radio to ground.

In order to work, the crystal radio must be wired to a good ground connection. Often, the easiest way to make a ground connection is to use the grounding circuit for the house's electrical system. Cut a piece of hook-up wire long enough to reach from the radio to a nearby electrical outlet. Remove 1/2" of insulation from each end of the wire. Wrap one end around screw A of the radio and tighten the screw. On the electrical outlet, loosen the screw that holds the plate on the box. Wrap the other end of the wire around this screw and tighten it.

CAUTION: Be sure not to let the wire or anything else touch the 120 V socket!

Alternatively, you can use a cold water pipe to connect the radio to ground. Remove a few inches of insulation from the wire from screw A of the radio and wrap the bare end around a cold water pipe.

8. Connect an antenna to the radio.

Use the rest of the hook-up wire to make an antenna. Remove 1/2" of insulation from one end of the wire and wrap it around screw D. Tighten the screw. Unwind the wire from the spool and try to place it as high as possible in the room.

In areas with many powerful radio stations nearby, an inside antenna should be adequate to receive a few stations. If you have trouble hearing anything, it may help to set up a long outside antenna. Radio Shack sells a kit with the necessary wire and other parts (cat. no. 278-758).

9. Try the radio!

Plug the earphone into the jack. Listen to the earphone and slowly move the slider across the coil. Try to find the position where a station is loudest. You should be able to hear a few different stations with the slider at different positions.

10. If the radio doesn't work, check the following:

a. Make sure all of the connections are correct and that the insulation has been removed from the wires where they are wrapped around screws. Also check that there aren't any short circuits between the screws or bare wires.

b. Make sure that you scraped the insulation from the coil wire with the slider so that the slider can make good contact with the coil.

c. Make sure that you have a good ground connection. Try a different electrical outlet or a cold water pipe.

d. Place the antenna higher up or try a longer wire.

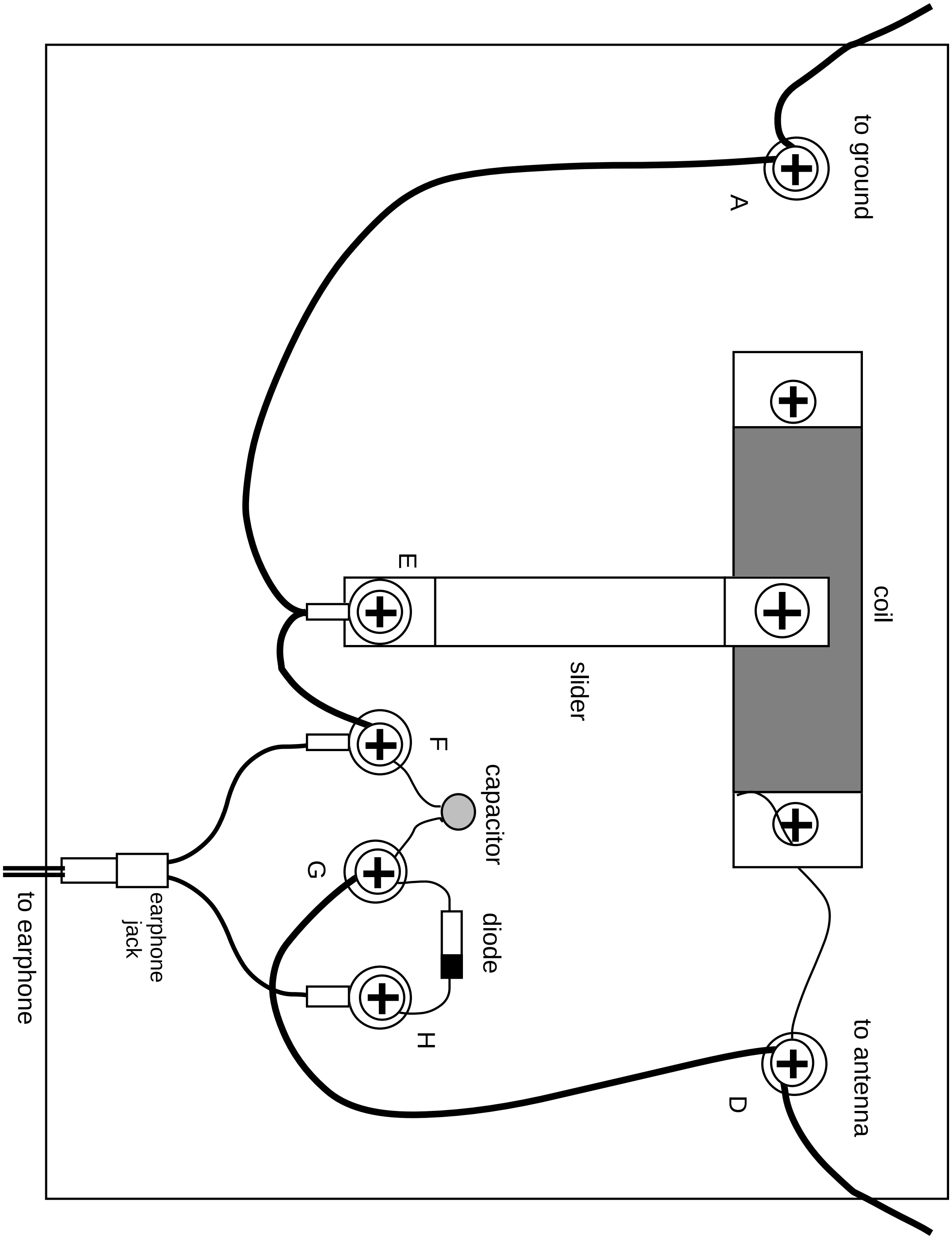
to ground
A
coil
E
slider
F
capacitor
G
diode
H
to antenna
D
earphone jack
to earphone

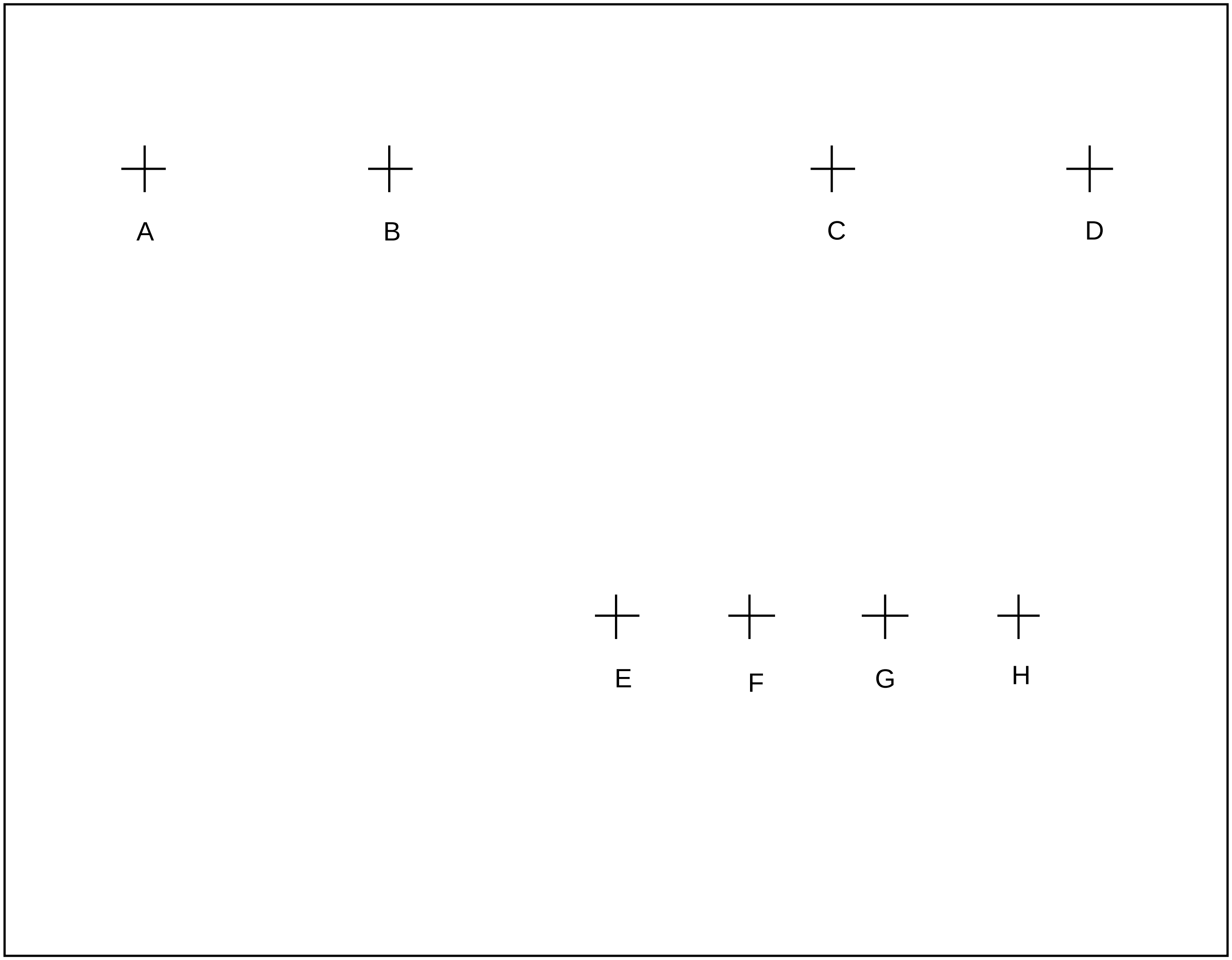
A
B
C
D
E
F
G
H

Superreaction receiver with ECC81 Regenerative RX with 12AT7

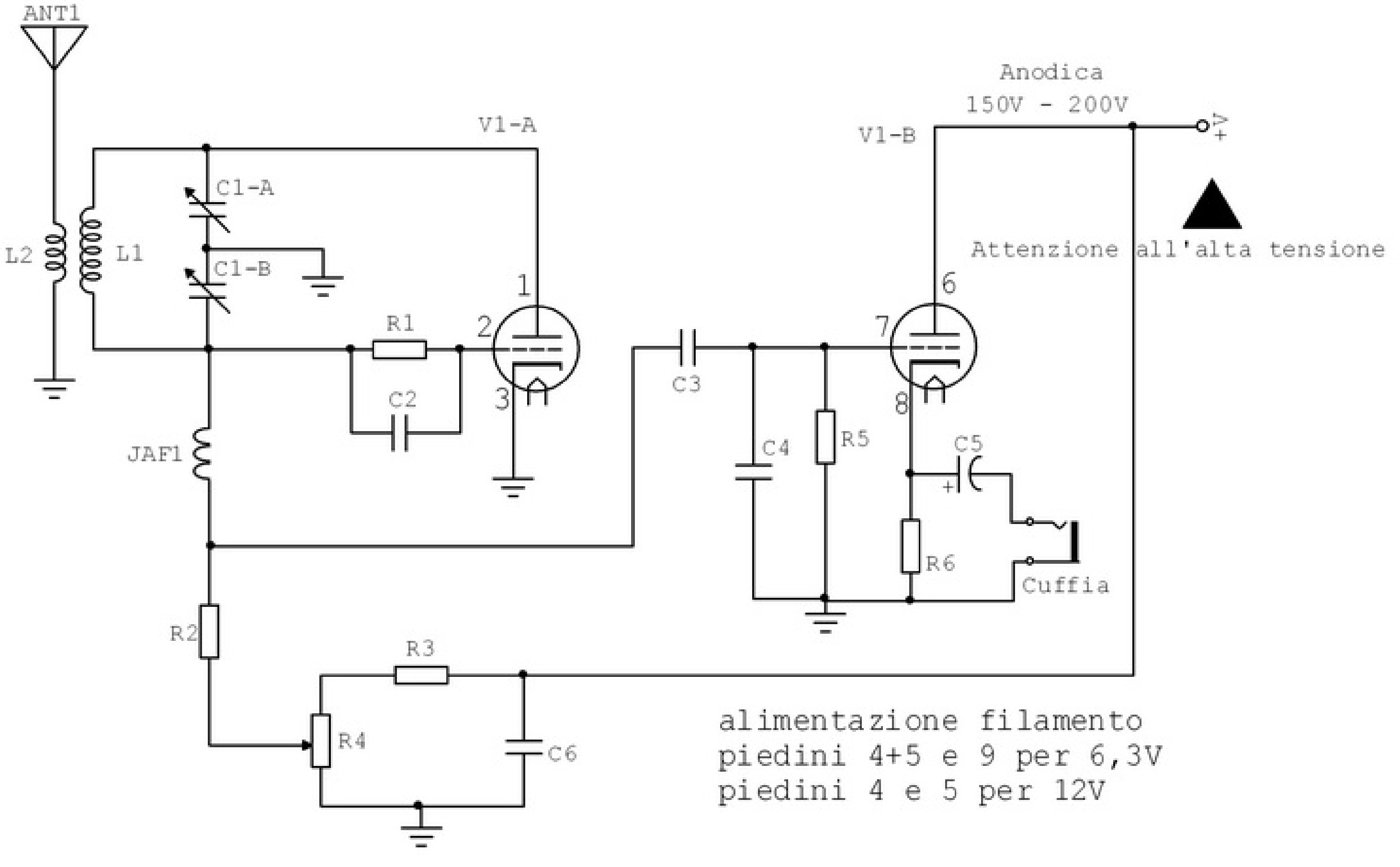

alimentazione filamento
piedini 4+5 e 9 per 6,3V
piedini 4 e 5 per 12V

Ricevitore FM in super reazione con ECC81-12AT7

Building a receiver like this nowadays may seem like suicide.

Why embark on a critical realization and doubtful functioning from the point of view of quality, when it would be enough to buy one of the many monolithic chips all do in which virtually attack the antenna, headphones and battery, and you have better results than this.

Yeah, why ?!

Do you want me to list them point by point or do you imagine them?
I know, I know the answers you would give me, colorful, funny, offensive and elusive, accusatory and vaunting.
The truth is that I needed to get involved.

Everyone talked, talked, talked about the social crisis in which we find ourselves, everyone pretended nothing, everyone put the ham on the eyes and the prepaid TV in the ears so that via cell phone we were more stoned than what in the end not we would always be lost in futile artfully created problems to mislead us from the real terror we are slipping into.
With only one salary and a wonderful family, I did what the state does not pass in the mind to do, cut the unnecessary expenses and, among these, the first to be eliminated, were the outlays for the purchase of radio, books and magazines.
At the beginning it was particularly frustrating, I do not deny it, I felt like a failure, but then I made myself a reason and I realized that, in the bad, I had found a lifeline to cling to.
I rediscovered the radios and books that surrounded me, I realized that I no longer considered them, occupied as

I was to accumulate as much as possible, to fill the void that left you the fact of not feeling part of a community indeed, a group of people who now thought only of their business.
So I went to the basement to dust the dust cover but, above all, to open all the shoe boxes full of disassembled components over the years and start their cataloging in order to build, build anything, equipment, joke, havoc, provided I felt alive again by putting myself to the test.

In all this psychological, psychiatric, sociological chaos, my friend Stefano Leopardi arrived and offered me an advanced radio editing course at his school, for the afternoon activities chosen by the boys. Advanced course because the base we had already done :-)
The project was ambitious, to build to the boys, after the galena in OM amplified with two transistors, a valve receiver but with the reception in MF.
After considering the few circuits that are on the net and on the books, the choice fell on the famous receiver of prof. Pegna and that you can find **here** . The first step was to find the material and to test the circuit, eliminating immediately the uninsulated power supply from the network and using a New Electronics kit for the supply of valve devices. The second step was the appointment at school, on the 29th of December 2012, during which we handled the welder and prepared the first prototype, working immediately even if with some difficulty.

I admit, a bit 'skeptical I was, not convinced of the potential of the circuit, not convinced of using high voltage to boys and, above all, not convinced that it worked just so much so that, three consecutive times, we misfired for assembly errors.
But then, after listening to the first notes and voices, and having found that, after all, the project was not so utopian, I went on to build my own version, to make possible problems that could have occurred to the boys, and that's where the vicissitudes started.

Do not expect me a mathematical treatment of superreaction, they practically do not exist (at least in the books and magazines in my possession). The only deals in Italian on the web, which have some validity, **here** and **here** . The disquisition is interesting and detailed, in an attempt to extrapolate the functioning of the FM superreaction, but does not draw a spider from the hole, in the sense that there are no certainties about the real mode of detection of frequency modulation.

Another help could come in an empirical form, using a whole radio frequency instrumentation, in order to have the possibility of achieving maximum performance using the equipment and the empirical replacement of the components, logically knowing why and what to replace. In one way or another, I wandered into the limbo of the evolved kit builder, that is, that freak being that is neither a designer, nor a graduate, nor a rich one who can afford the purchase of what you want , that is of that Sunday animal that has survived since the radio-electronic culture has had light.

The devices in super-action are suitable for the reception of very high frequency signals and with excellent sensitivity, despite requiring a small number of components and allowing the detection of signals both amplitude modulated and frequency modulated.
They are easy to build and are very simple, but require a set-up that is often laborious and that makes mass production virtually impossible, as each receiver is a world apart.
They usually consist of a detector valve followed by one or two audio amplifying valves (the same is true for transistor circuitry).
The sensitivity of the detector valve in reaction is extremely high but it is for a very short time since immediately after the valve starts to oscillate making us hear the typical whistle in the speaker and starting to transmit disturbing the equipment of neighbors.
It has therefore been thought to prevent the valve from entering oscillation, by making its operation intermittent around the point of maximum sensitivity of the valve, that is to keep the operation of the valve suspended between the moment of triggering of the oscillations and the moment immediately before, turning off the oscillations every time they exceed a certain threshold.
If the intermittences were 100 per second, reception would be impossible because the resulting signal would be a dark buzz; as the intermittences increase, the buzzing gives way to a sharp whistle and then to a rustle, so it is from there that we try to create intermittences that are outside the upper hearing threshold of the human ear.
This intermittence, is called the *extinction frequency,* and in general, to obtain the above result, a frequency of 100Khz is applied to the detector, so that each positive half-cycle corresponds to the valve operation and to each negative half-cycle it corresponds to the valve locking same.
The drawback of this type of reception consists in the fact that at each operating interval many cycles of the incoming signal must correspond. If the shutdown frequency is 100KHz it will be impossible to receive long waves

or medium waves, and if you want to receive ultra-short waves, receiving 100MHz signals would have 1000 cycles of incoming signal for each cycle.

Our circuit is of the self-oscillating type, the left triode in the electric scheme acts as a grid triode detector and oscillating triode. The antenna group, in addition to performing the usual functions of tunable range, constitutes a Colpitts oscillator. The JAF1-C3-C4 group constitutes the cell that together with the triode generates the extinction frequency; the two resulting signals will subsequently be revealed in grid flaws.
The potentiometer R4, by varying the anodic voltage of the triode, varies its operation (switch off and oscillation). The left-hand triode is a simple low-frequency amplifier triode once the signal has been detected, in which the R5-R6-C5 components constitute the components that correctly polarize the trip. Capacitor C6 filters the spurious signals into radiofrequency to ground.

THE CONSTRUCTION AND THE LIST OF THE SAINTS

What you see is the first version of the receiver we have talked about so far.
It is built in an absolutely Spartan way, taking into account only the length of the wires and the terminals of the components. The power supply is essential but also spartan, without frills, even the filament is powered by the 9V transformer instead of the necessary 12V, as, under load, the transformer pulls out 12V instead of 9V mah, the mysteries of the copper electron !

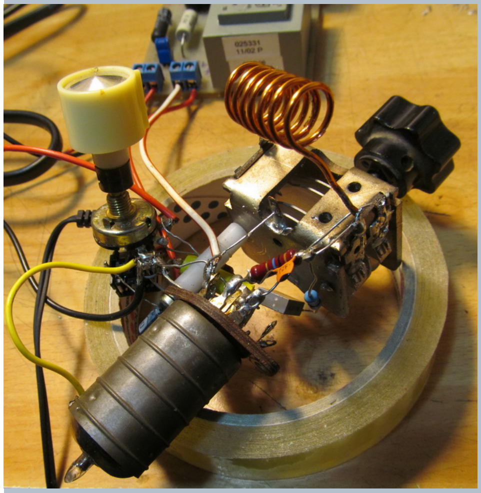

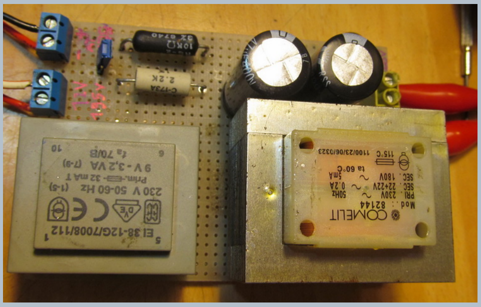

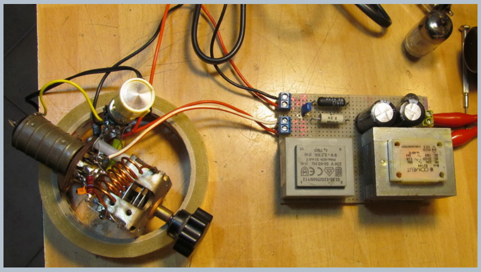

In the first tests done with Stefano, they powered the radio at 150V of anodic, the power supply used in my case has two voltages, 170V and 200V, does not change practically anything to the ear deafly humanoid like mine, if I had the instrumentation, maybe some more info I gave it to you .... The only noteworthy difference is the displacement of the point where the reaction intervenes. You can even go down on 130V, but the more you go down, the more the reception, the reaction capacity and the output level decrease.
Turned on the circuit, I bowed to His Lordship Superreaction: although I found a house point (the usual closet) stingy with radio waves, I could easily receive many commercial stations with a good volume on the headphones, holding the command of the reaction just the typical rustling started, from nothingness to chaos in a light touch. Compared to the original scheme, the value of the double variable capacitor and the number of turns of the chord and oscillation coil had changed. In this situation there was a lot of buzzing in the background, a remnant of the alternate, although tolerable and there was an accentuated hand effect. There was no antenna and no loose coupling, just as the welder had created. Computer speakers, even if powered continuously by batteries, at maximum volume,
In all this beehive nest, antithesis of the measurement laboratory, listen to the number of stations that you receive and, in general, the quality of the stations received in function of the simplicity of the assembled circuit. In the evening, with the headphones, as the external noise and interference diminished, the audio quality increased.

After these comforting results and after verifying the problems that the boys might have encountered, I decided that I would have prepared a small wooden frame at least to give some solidity to the structure and a better arrangement of the components. Actually I just picked up the receiver and screwed it to the wooden front panel, putting the power supply on the insulating surface, always made of wood, so as not to risk unnecessary short circuits.
The result?
Disappointing.
The reception was no longer the same, the signal seemed to have given way to the noise, although the test locations were always the same, the reception seemed silenced, even changing the valve with other functions, the result did not change. Then I started to change the assembly plan, changing the arrangement of the components, moving away, approaching the distanced and approachable, but the only result I got was the displacement of the receivable band (I have also come to change the speakers. ...).
You can listen to the result of the second experiment and watch it in this video.

Disappointed and disheartened, I went to Stefano's house one Saturday to help me untangle the skein and learn something and instead, after a good coffee, my receiver started playing almost like a commercial superheterodyne: clear and strong sound, even with the addition of an external antenna, which considerably increased the audio output (which I had not been able to do at home because it was replaced by a strong buzz that obscured any transmission). In addition we had connected the valve amplifier that was the other part of the school project and the sound was powerful and full-bodied.
Perhaps the receiver had had a mystical transmutation in road transport or had been blessed by an intergalactic

wave coming from the XTY56 sector of the Bravosian galaxy or, simply, it was the usual c ... of the Sunday arrangist that in the presence of the supreme master bowed to the knowledge ....

I can not explain it, coming home from Stefano I felt relieved, all the hours lost to make changes and attempts, had led to the creation of a spectacular receiver, the results far above my expectations and the very good sound quality.

I returned home and, after falling asleep, I put on the headphones and I tried again the reception and .... I took the calendar and I enumerated in alphabetical order the sanctified progeny of past centuries. Joke, some dirty word I pulled it down, but nothing striking because if I had really named all those saints, you know the money I would have to pay in copyright ......

Nothing had changed, everything as before, but at least I had the certainty and awareness of what the good Ravalico and good Brazioli had been preaching for years now: every receiver in superreaction is a world in itself, both on the constructive side and in the context in which it is used.

The test was exhaustive, at Stefano's house it was received that it was a marvel, in my house a crap.

Since I had nothing to do, I changed the circuit again, which by now was taking on the definitive aspect and I started the tests again. The following is the result.

During the tests I notice a strange behavior: just before the sudden reaction comes into operation, there is a zone of rotation of the potentiometer in which there seems to be a clean radio reception from the noise. Since the grid voltage oscillated around 27V, I decided to extend this range of variability by replacing the 100Kohm potentiometer with a 47Koohm resistor series and a 22Kohm linear variation potentiometer. In this way, that point of radio grace was perfectly manageable and easily adjustable.

The receiver began to have its own soul and life.

I get inspired by optimism and I also try to insert the antenna winding. Now the generation of the buzz is not as unmanageable as before, I make a thousand attempts between number of turns and positioning of the loop and arrival at the optimum, using two coils of the same diameter of the antenna coil with one end to ground and the other connected to a stylus antenna. Now the signal is strong enough and the sound is clean enough. I block the antenna coil with the cyano-acrylic glue and I relax by scrolling through the frequency modulation range, enjoying the long-awaited success and letting myself be lulled by the silence that only night can give you.

This is the final result, divided into two videos due to Crozza's intervention in Sanremo.

This time I let the video talk rather than the images, I inserted the photos of the initial realization and I finish by inserting the photos of the final version.
As usual, I did not pay attention to the aesthetics, but rather to the final result I had proposed. I am certain, because experienced, that the receiver built in this way will give its best under all circumstances, as improved in the worst reception conditions, those of my house.
I do not want to, an excellent project is the combination of all the aspects that compose it, I always defect in the aesthetic one but I will improve, one day.
The complete project involves the construction of the power supply and the low-frequency power stage, but these topics will see the light editor later. **BUILDING TIPS, WORKING EXPERIENCES AND FINAL IMPRESSIONS** Where do we start? I do not know, I go in random order, just like memories come back to me. **Defects and advice.** The circuit has its own heart and, therefore, its own independent life. Mine will be different from yours and the constructive or implemented strategies that you will adopt will be similar, but it is said that they reach the same result.

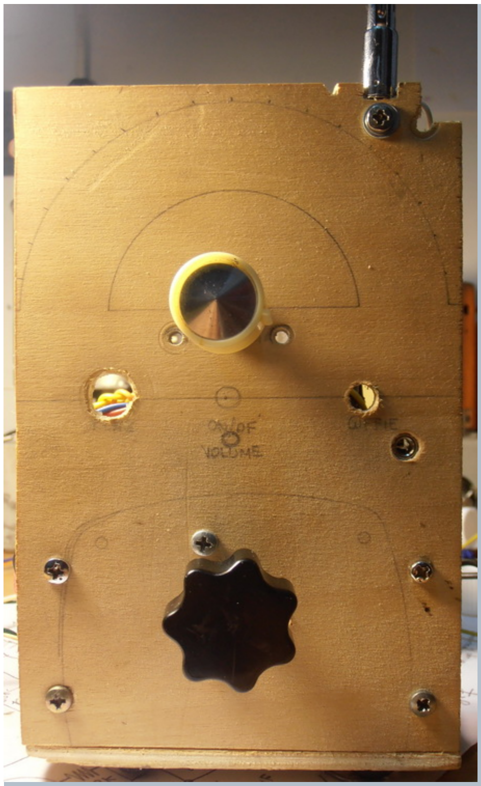

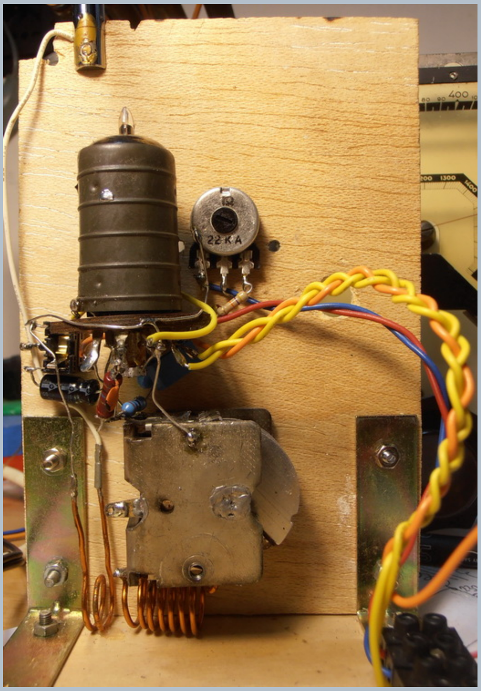

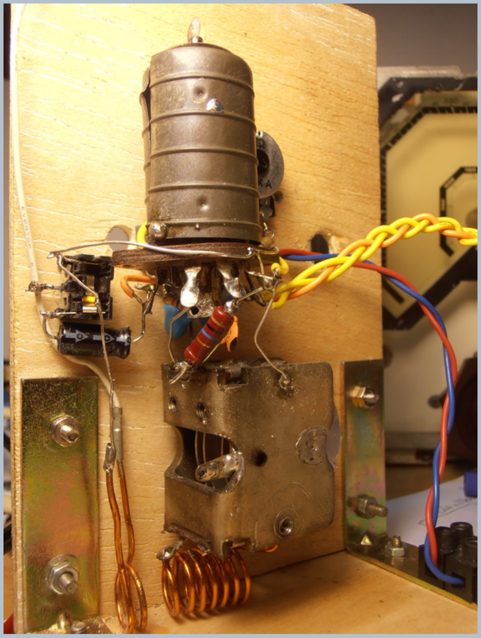

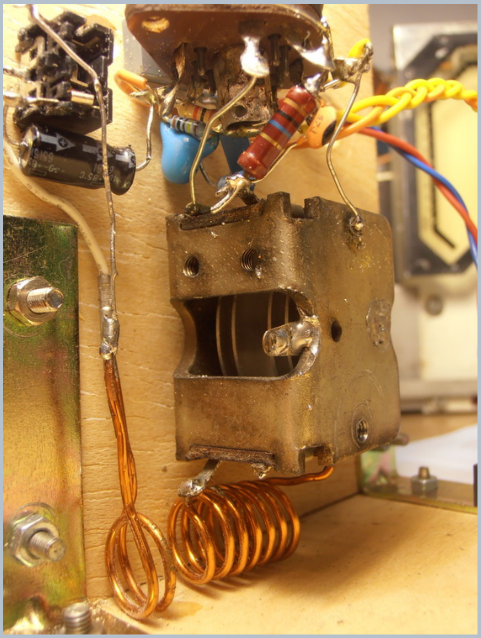

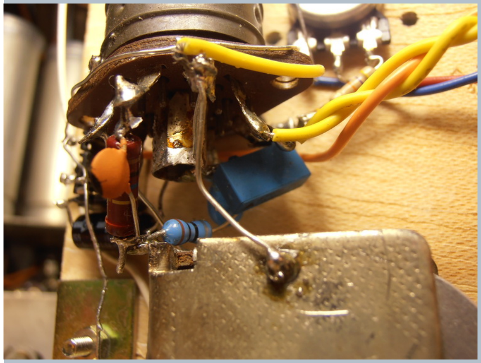

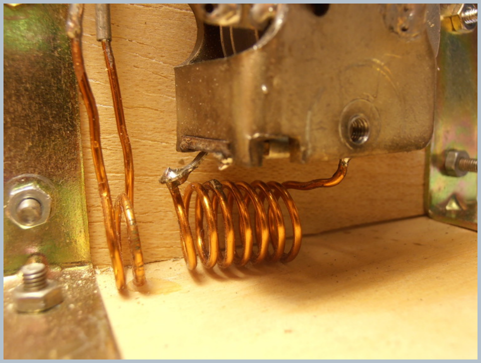

Do not use long, parallel or 90 ° links as well as hieroglyphs: high frequency editing has few basic rules and these are some of those. If you can, use rigid wire not interwoven as those normally used in electronics, for example the resistance terminals are excellent, use links like those, sell them already made but you can also find them in vintage wires peeling. You can use the copper wire of the transformers that you will have scrapped, but you will

have the hassle of the peeling and the subsequent welding before putting them in place.
As usual, it is a must to use what you have at home, especially in this period of crisis (and then become an anti-consumerism philosophy).

The circuit is not very selective, not because of me, but because of how it was conceived.
Trying to get a high sensitivity has had to give way to selectivity. In OM and OC low there is the problem that between one channel and the other there are 9-10Khz with a signal of amplitude however very strong of the order of mV. However, having to do with the high frequencies, the view is reversed, we must go to look for very weak signals, from 5uV up but with bandwidths of separation between one channel and another of about 200KHz. So when you have two stations nearby, or you hear them together, or the most powerful will overwhelm the weaker. You can maneuver the variable capacitor to try to get the best reception, perhaps using a manopolone or reducing the rotation (in this case it would go against the cheapness of the realization if you buy a knob specially gutted, try to invent a tuning system with lanyard or use a large knob, maybe the cap of one liter bottles of fruit juice, in short, the choice that is most congenial to you).

You'll probably need a valve shield. It is a choice that you will have to face straight away, otherwise you will have to lose the track again. The screen isolates the valve from the hand effect and the residual humming but becomes part of the tuning and reaction circuit. It is one of the faults of the circuit: the construction, the components, the connections, everything becomes part of the reception and the reaction, first you understand this and first you will be able to understand the behavior of the receiver, without the need for mathematical formulas and instruments scientific.

The capacitor used in this receiver, is a double variable in air 8-19pF then, thinking about the small capacities involved,

In the absence of a tuned signal, a loud background noise is heard, which disappears almost completely when a sufficiently strong station is tuned in, noise that still annoys, especially in prolonged listening or in headphones.

The receiver, as a reaction, is an excellent transmitter on the tuned frequency: you must then duly measure the reaction in order to avoid that the roommates of the building, aware of your passion, waiting for you outside the door and beat you to the sound of music.

It is possible that the hand effect is present, an effect that leads your hand to be part of the tuning circuit and that leads to the variation of the tuned frequency according to the distance of your hand from the variable. You can avoid this problem by using the aluminum front panel, to the ugly use aluminum to cook in film, try what happens if you put it to ground or not. In the final configuration the hand effect was not there while it was present in the initial one.

Give the valve a few minutes to warm up and arrive at an equilibrium temperature with the environment in which it is located; this will lead to no frequency shift during listening.

For all the observations reported so far, we understand how the use of these receivers is limited to amateur receivers or particular applications, but despite everything is a configuration that gives enormous satisfaction to those who venture into the world of VHF with the superreation for the first time: it is a great bench on which to experience and learn new concepts.

Highs

After having exposed all the problems that may arise, I guess I'll find mica so many advantages ....

The receiver is built with very few components, without coils with strange initials, with pieces that you can easily find in your drawer or at fairs (I'm talking about the condenser and the valve). The receiver is cheap, the valve, fortunately, is not widely used by audiophiles, so its price does not suffer the mood of the market. The variable condenser Stefano paid him at the fair € 2.50 ....

The receiver, once you have worked on it, it works really well.

You can use the headphones that you use for media players, so you can purchase them at little cost, or you can connect external speakers for computers, even if the sound output will not be particularly powerful. When you go to the usual Chinese store, check the specifications that the headphones are from 32ohm impedance, otherwise a low and distorted sound output.

The receiver allows you to experience in high frequency and allow you to learn new things, which is not a little. It is possible to receive a wide range of frequencies, modulated both in amplitude and in frequency, giving you the possibility to discover amateur ranges otherwise out of reach.

The power supply can be obtained using two transformers sold in Chinese shops, those with primary to 220V and multiple secondary, from 1.5V to 12V. I bought two for other purposes, from 1A and I paid € 4.5 each, if you take those from 500mA cost you € 1 less. If you use the 6V or 12V outputs you can also feed the filament in series or in parallel. The anodica requires at most 5mA (remember primary-secondary-secondary-primary).

In short, I do not have other things to say only that, as you have heard, the receiver works really well (and unfortunate not to have registered at Stefano's house ...).

Last-minute tips

I recommend using the earphones: the closer they approach the 64ohm required impedance, the greater the volume in headphones and the better the sound quality. I have tried 7 of earphones and only two have given the desired results, even if no one takes me out of my head that at most are 32ohm headphones. As mentioned previously, an incorrect and low output impedance leads to increased distortion and headphone output volume.

The impedance L1 regulates the tuned band and the type of oscillations. With the double variable capacitor in my possession that does not exceed 20 + 20pF and the copper wire of that diameter, 7 turns were needed to have the whole FM band receivable with only a few degrees of rotation of the lost capacitor. Depending on the capacitor in your possession you have to jig on the turns and their distance to have the whole band and avoid the overlap between the stations and the same applies to the diameter of the wire you have, must still be close to that indicated in the original scheme. If the thread was unglazed but silvered, things should be better.

The coil L2 you have to try with the assembly that you will have: you will have to test the distance from the L1 for which you have greater coupling and therefore output signal strength and minimum residual hum. Test.

As for the JAF1, try all the reels you have, as long as the value is between 5 and 10 micro Henry. The higher the value, the less stations you receive, the more difficult it will be to react. At the moment there is a commercial 10uH reel mounted, I had mounted one when I went to the house of Stefano that his digital inductance measured 7uH, commercials less than 10uH I had to try.

I tried to mount an ECC83 but the result was not exciting. He behaved well, some stations felt limpid, but the management of the reaction became impossible and the point of work of grace that I had found was no longer found. You probably have to rebuild the circuit on that valve, you can try if you do not already have it in the drawer and you do not want to spend the money.

Depending on whether you use the headphones or the external speakers of the computer, you will have to manage the socket for female audio jack: if you use the computer speakers, the head that goes to ground connect it with the control unit of the female jack.

Depending on whether you use the headphones or the external speakers of the computer, you will have to manage the socket for female audio jack: if you use the computer speakers, the head that goes to ground connect it with the control unit of the female jack, the output will be monophonic but the output power and quality will increase significantly.

Remember to connect the screen to ground and place the wire so as to have as little buzz as possible.

In the receiver mounted by me, three operating modes can be identified.

Before arriving at the threshold where the reaction is abruptly inserted, there is a state of grace that allows you to listen to the stations without the shadow of rustle, as you can hear from the last video. In this state the outgoing sound volume, although still manageable through the potentiometer, does not have a particularly high level.

When it comes to the actual reaction, noise and volume increase significantly. Once the station is centered, the noise will disappear almost completely but it will still be present. The higher the level of the signal that we will receive in the antenna, the less will be the noise in the headphones.

There is the third and last state, in which, continuing to increase the reaction, it will be possible to tune a station as if it were at the center of the minus flex of a sine wave. By turning the potentiometer slightly to the right and slightly to the left, you will hear a whistle soaked in rustle; in the background to the centered station you will always find the usual rustle. Generally, this behavior occurs when a very weak station is received.

List of components

V1 = valve ECC81 or 12AT7
R1 = 10Mohm
R2 = 100Kohm
R3 = 47Khohm
R4 = potentiometer 22Kohm linear
R5 = 1Mohm
R6 = 1Kohm
C1 = variable capacitor in air 20 + 20pF
C2 = ceramic 47pF
C3 = 10nF minimum 100V
C4 = 2.2nF
C5 = 25uF 16V electrolytic
C6 = 10nF 250V or more
L1 = 7 spiers spaced 1.5mm, diameter 1.4mm thread on 1cm support
L2 = 2 spaced spaced 1.5mm, diameter 1.4mm thread on 1cm support
JAF1 = 5-10uH Noval
valve plinth with screen support or take the screen separately.
Stereo jack to be connected with the two earphones in series (do not connect the central mass terminal).
2 knobs for tuning and reaction.
mammoths for connections.
Directional stylus antenna.
For resistances to the maximum use from half a watt but also 1 / 4W should not give problems.

It is time to conclude.
I set up the article as if it were a conversation between friends. I hope that the time taken to read the article and watch the videos was pleasant and relaxing.
See you next time.
Carlo.

PS I remembered, when I inserted the biography, that the third electronic kit that I bought in the distant youth, was just LX441 present in the number of Nuova Elettronica 74. I remember that I was very bad at the end of the editing because I only heard rustle and I could not to capture no station, even in FM. If only someone had taught me how superreaction was managed and how was it to be received, I would probably be another man today, better? worst? boh !!!! I tried to find it in the cellar, but I could not find it, who knows what happened to it, I would have liked to lose my hands today ......

BIBLIOGRAPHY

- Internet, the web of spider webs and Santo Google, the spiders' spider .
- **The site of prof. Pegna** .
- **The Luciano Marroccu site** .
- **Sophie's Radio** (forum and A transistor for FM).
- New Electronics magazine 74 January 1981 .
- Radio Libro - Ravalico 8th Edition - Chapter 25 from page 486 to page 499.
- Two numbers of Practical System of the 60s by the legendary Gianni Brazioli.
- Radiotechnics and Terman Electronics, from page 11 to page 13.
- Manual of Radiotechnical Engineering of the Terman, from pag.787 to pag.791.

WWW.ELECTRONICS-TUTORIALS.COM

Like Share 16 people like this. Sign Up to see what your friends like.

G+

•NEW! ‣ - Amazon Electronic Component Packs. ***Check out the [Amazon Electronic Component Packs page](#).***

WHAT ARE THE BASICS OF CRYSTAL RADIO RECEIVERS?

Here is the schematic diagram for a very basic crystal radio set without any particular embellishments. This basic old time radio uses no power other than that provided by the transmitting antenna from the radio station. Free power from the sky eh!. It truly is a marvel!.

As a matter of interest, some amateurs and experimenters living in the vicinity of high powered AM Radio transmitters have literally used crystal radio sets to trickle charge their batteries. Now THAT is real COOL! If the demand exists we will do it too!.

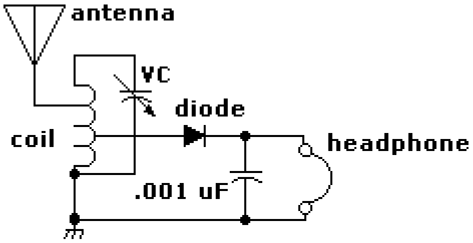

Figure 1 - schematic of a basic crystal radio set

NOTE: For people unfamiliar with metric measurements used here 25.4 mm equals 1"

This circuit consists of an **inductor** (also called a coil), a **variable capacitor** (used to be called a variable condenser), a germanium **diode** (formerly called a crystal), a filtering capacitor and finally very high impedance headphones.

The inductor has taps on it:

(a) one to connect the antenna to.
(b) one lower down to connect the detector diode to.

These taps are often best determined by experimentation

The variable capacitor is usually connected across the whole of the inductor to form a **tuned circuit** for our crystal radio set.

Note: the bottom junction of the inductor and variable capacitor is connected to ground / earth - **literally**.

Earth connection for crystal radio set

For the crystal radio set circuit to perform at all well you will need a very good earth connection. A length of old steel water pipe driven **deep** into damp earth is ideal. The electrical earth or water service at home is **NOT** a good or safe earth for this project, in fact it is **VERY DANGEROUS**. - see **earth dangers**.

The funny looking symbol connected to the top of the inductor or coil and variable capacitor denotes an antenna. The higher and longer (50' or 17 metres) this antenna is, the better the likely reception. Reality says of course you can erect an antenna in the space available with the material available. Consider running the antenna around the top of your fence line if is not metallic.

Safety comes first with our crystal radio set

KEEP WELL AWAY FROM POWER LINES AND POWER SOURCES - **REPEAT AFTER ME** - I PROMISE TO KEEP WELL AWAY FROM POWER LINES AND POWER SOURCES.

Always remember the motto "safety first in everything". At this site here I stress that all the time - "safety first" and I can never stress it enough. See my important page about **earth dangers**

Antenna for crystal radio set

Remember we're using the antenna (aerial) to get free power here and we want to capture as much as possible from the sky for our crystal radio set.

Perhaps you might like to review some **antenna basics** before proceeding.

Technical Note: Radio signals (waves) such as we encounter in the am radio band have two halves. One half travels across the surface of the earth at the speed of light through people, buildings and other objects. The other half, a mirror image, travels beneath the surface of the earth. This radio wave has a definite length. Its length is the speed of light divided by the frequency. For frequencies in the megahertz region this approximates to 300 divided by the frequency in megahertz. These radio waves we want to detect with our crystal radio set

An example which we use later is the **am radio band** of 530 Khz to 1650 Khz. These frequencies when converted to .53 Mhz and 1.65 Mhz respectively give wavelengths of 566 metres and 182 metres.

Now some critical aspects:

Variable Capacitor: Here you should try and salvage one from an old time radio. Just make sure the radio is not in the "vintage" class and now worth hundreds of dollars. The ideal ones are nearly impossible to buy now. Unless you have access to measuring equipment you will have to guess the capacitance. A lot of old time radios had capacitors which tuned about 15 pF to 365 pF.

Inductor or Coil: Here we are normally dealing with an air cored inductor wound on some suitable **non metallic** form. Some examples are - cardboard toilet roll former, a short length of PVC tubing 40 mm, 50 mm, 90 mm or 100 mm diameter.

Diode: In lieu of the old crystal detector we use a germaniumdiode of the 1N34 or OA90 type. Do NOT use a 1N914 type.

Fixed Capacitor: This is for filtering and may be .001 uF, 1 NF or 1000 pF type (all those values are the same - just expressed in different units). Something near in the same values is quite O.K.

Headphones: This is by far the hardest part to obtain. The type used for hi-fi **WILL NOT** work here. Ideally you need high impedance 2,000 ohm types but these are nearly impossible to find. You can sometimes buy 1,000 ohm crystal earpieces (no **not** the usual transistor radio type). The headphone is a high impedance load for the crystal set and as we are working on free power from the air we can't load it down. If you used the 8 ohm hi-fi type it would be like trying to run your electric toaster off a very small battery. The power is not available. Remember we're using free power from the sky.

Further reading on **Headsets for Crystal Radios** by Alan R. Klase.

Frequency range: The frequency range of a set like this is mainly determined by the square root of the *ratio of Maximum Capacitance to Minimum Capacitance of the variable capacitor.* In the example above this would be the square root of 365 divided by 15 or about 4.9:1. In the real world there are other capacitance's in the circuit that you can't see (called "strays") and they affect the minimum value. In practice I would expect a real range of about 3:1

A.M. Radio band: Depending upon your country this would cover about 530 Khz to about 1650 Khz. This is slightly more than a 3:1 ratio.

Let us assume for discussion our capacitor was as described above. And for convenience we have strays of about 25 pF (typical). If we add this to both the Max and Min capacitance we get 390 pF Max and 40 pF Min - a ratio of 9.75 the square root of which is 3.122. Happily this also happens to be nearly the ratio of Freq $_{max}$ divided by Freq $_{min}$ or 1650 Khz / 530 Khz. How about that?

Of course these variable capacitors were deliberately manufactured that way for that very reason!

Well now we have to resonate our capacitor with the inductor. What inductance do we need?. The easiest formula is this - take the frequency in Megahertz (i.e. freq in Khz divided by 1000) and square it. Using our two frequencies 530 Khz and 1650 Khz we would have to square both .530 Mhz and 1.65 Mhz which of course become 0.2809 and 2.7225

And you thought maths at school was/is a waste of time!.

These two figures are divided into the standard figure of 25330.3 respectively. The 530 Khz one (0.2809) would give us LC 90175 and the 1650 Khz one (2.7225) will give us LC 9304. Now when tuning a crystal set the more capacitance you use (plates meshed further in) the lower the frequency you receive. In our case we are using a max of 365 pF + 25 pF strays or 390 pF total. At our lowest freq of 530 Khz the LC (L means inductance and C means capacitance) is 90175 and when divided by our C of 390 pF we get a needed inductance of 231.2 uH (uH means micro henry, a unit of inductance).

Similarly if we go to the highest frequency of 1650 Khz our required LC is 9304 and when divided by our minimum capacitance's (reduce the capacitance by withdrawing plates from one another and go higher in frequency) of 15 pF + 25 pF strays or 40 pF we get a required inductance of _______ uH?. Go on work it out. And remember we are only interested in whole numbers because we have made a number of approximations here.

Now how do we get a needed inductance of about **230 uH?**.

Gee whizz not more maths headed our way!

The formula for inductance - using toilet rolls, PVC pipe etc. can be well approximated by:

$$\text{Inductance L} = \frac{0.394 * r^{2} * N^{2}}{(9 * r) + (10 * \text{Len})}$$

Here:

r = radius of the coil i.e. form diameter (in cm.) divided by 2
Len = length of the coil - again in cm.
L = inductance in uH.
* = multiply by

Now assume we have used a piece of PVC pipe of about 104 mm (10.4 cm) diameter. This has a radius of 5.2 cm. Let us wind a coil consisting of 57 turns (with taps at about every 5 turns) over a length of 104 mm or 10.4 cm. For the Mum's and Dad's who work in old measurements that's 4". This is a piece of standard 4" PVC sewer tube.

Politely ask any plumber on any building site for a scrap off-cut about 150 mm (6") long. In another life I'm also a plumber and we like to help polite people.

Let's put those numbers into the our formula above and do some sums:

0.394 * 5.2 * 5.2 * 57 * 57 = 34614

AND (9 * 5.2) + (10 * 10.4) = 46.8 + 104 = 150.8

AND 34614 divided by 150.8 = about 230 uH

Boy am I a clever kid or what. Did you work through that on paper all by yourself? I do hope so.

Why and how taps every 5 turns?. The why first - in the beginning I said connections from the antenna (aerial for old timers) and diode would need to be experimented with. With taps you can select different spots along the inductor.

How?. Easy! Wind your inductor as follows:

You will need a length of enamelled copper winding wire (about 24 gauge or 0.5 mm dia) about 20 metres long - yes you will need just about all of it. You will also need some matches (used are O.K.). Assuming you have a 150 mm piece of sewer tube or similar, drill a small hole approx 25 mm (1") from one end. Drill a similar hole 104 mm further away toward the opposite end (or even 20 mm in from the opposite end). Basically the idea is for the holes to be nearly 10.4 cm apart. or *the length of the coil you are winding*.

Push through about 15 cm (150 mm) of wire and wind the short end through a couple of times to form an anchor (knot) at that point. Then holding the wire tightly wind 5 complete turns spaced about three wire diameters apart from each other (you really need a helper here). Under the 5th turn slip a very short length of match, continue winding very tightly placing a match piece under every 5th turn until you have wound 57 complete turns. Tie a knot similar to the other end and again leave about 150 mm left over. **CAREFULLY** with a sharp knife scrape the enamel insulation away on all 5th turns as well as about 5 mm from the end of the two end pieces.

Presto! there you have a 57 turn inductor of about 230 uH with accessible taps at about every 5th turn.

How it works: (simple version)

Well we have the antenna. This picks up the radio signals from the sky and with our real good earth we hopefully get lots and lots of signals.

The inductor combined with different settings of the variable capacitor selects the frequency you want to listen to and discards all others.

This selected signal goes through the germanium diode and is rectified. This means we are left with the audio frequencies a human being can hear (music - words). This audio has been stripped off the radio signal by the diode. The small .001 uF capacitor gets rid of any of the radio signal left. We can then listen to these very tiny audio signals through our headphones or crystal earpiece.

How it works: (complicated version)

Your long wire antenna and earth wire will pick up thousands and thousands of radio signals from the air. The exact setting of your variable capacitor will determine a set amount of capacitance and this in conjunction with the fixed inductance of your coil we have a **resonant circuit**.

As an example, in my locality we have a very strong AM station at 630 Khz. Using the formula above for **resonant frequency** we find the LC combination (inductance times capacitance) required to receive this signal is:

$25330.3 / (0.63)^2 = 63820$

Therefore if we had a fixed inductance of 230 uH then our variable capacitor (including strays) would need to be set at the point where it exhibits 63820 / 230 **OR** 277 pF. In other words a capacitance of 277 pF in conjunction with an inductance of 230 uH resonates at 630 Khz. We can rather generalise here and say, although not strictly accurate, that this coil / capacitor combination will **ACCEPT** only signals at 630 Khz and **REJECT** all others. This is the way we select our particular frequency from among the thousands and thousands of other signals on our antenna.

Our signal at 630 Khz has been "modulated" at the radio station with audio frequencies (speech and music). Because it is **A**mplitude **M**odulated it is called **A.M.**

This means the original 630 Khz signal strength (amplitude) at the transmitter is expanded and contracted by the audio frequencies.

After being selected by our coil / capacitor combination the 630 Khz signal is passed to the germanium diode. As the signal passes through the diode only one half of the signal is used and the fixed .001 uF capacitor dumps the 630 Khz signal to ground leaving a very tiny audio frequency which is an exact replica of the audio frequencies (speech and music) as they originated in the radio station's studio.

This very tiny audio signal then drives our crystal earpiece or high impedance headphone so we can hear the words and music. If it were amplified many, many times over it would then have the power to drive a loudspeaker. The speaker cone goes in and out at the same audio rate as the original 630 Khz signal is modulated. The moving loudspeaker cone creates a tiny sound pressure level in the room to which your ear is able to respond.

Radio Kits for Beginners and Youngsters of ALL Ages!

My special amateur radio friend Bruce Kizerian, KK7ZZ who has worked tirelessly for years helping the young and not so young get effortlessly into radio now has a site devoted to "Radio Kits for Beginners and Youngsters of ALL Ages!".

There has been a crying need for these kits and Bruce now fills that need.

Check out Bruce's **"Radio Kits for Beginners and Youngsters of ALL Ages!"**

NEW! - "High Sensitivity Crystal Set"

Build a “crystal radio ” by using a new zero-voltage-threshold MOSFET. By Bob Culter, N7FKI as printed in QST [Amateur Radio Magazine] January 2007

Check out the PDF file of the whole article - **"http://www.arrl.org/files/file/Technology/tis/info/pdf/culter.pdf"** 467 kB. **PS:** The article can be more than somewhat highly technical for newcomers though.

NEXT:

Put it all together and reload this page by typing in the URL - www.electronics-tutorials.com/receivers/crystal-radio-set.htm and **tell me how your crystal set works** for you.

If you are real keen you can also check out the **crystal set society's** home page.

Another excellent article (in pdf format) is from the ARRL. From QST December, 1997 **The Crystal Radio**

Fox Hole Radio

Another VERY interesting project is the **Fox Hole Radio**. Building a foxhole radio is rewarding and the basic setup is very simple. It is, however, difficult to adjust, and it may take several

attempts to find a proper razor blade for the detector. This is a project that requires patience and much trial and error, but it will pay off once it begins to work. It will help to be versed in the construction and operation of crystal sets before building one. These sets are extremely simple in construction, but tuning and modification require some basic understanding of theory, as well as practice. All sets presented here are based on old articles, notes, and people's recollections. There are fairly major variations in design and materials among these plans. It must be remembered that these were improvised under often adverse conditions; there was no "standard" design. With this in mind, take this entire article as a whole, and use it a bit here, a bit there, to build towards a Fox Hole Radio design that works best using modern materials.
See: **Fox Hole Radio**

A Candle powered Radio

Yep a World Book Science Project from a looooonnnng time ago.
See: **Candle powered Radio**

A BOOK I THOROUGHLY RECOMMEND FOR YOUNG PEOPLE

Radio Rescue
by Lynne Barasch

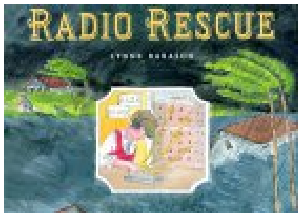

From Booklist
Gr. 2-5. In 1923, a 10-year-old New Yorker became the youngest licensed amateur wireless radio operator. In a clear, first-person voice, the boy describes what Morse code is, how operators used wireless radios to send and receive signals, and his experiences setting up an in-house transmitting station. The boy finds fame when he picks up a signal from Florida hurricane victims and radios for help, and the story concludes with a reproduced newspaper photograph and the full Morse code alphabet. An introductory note reveals that the story is based on the author's father and gives readers background information on the era's telephones and the alternative that the wireless offered, including examples of how Morse code has been used throughout its history. In a well-designed mix of insets, brief sketches, and full-page drawings, the author's uncluttered color cartoons do an excellent job of illustrating the technology and the code, at the same time creating likable, expressive characters. Tehnology-minded children may see parallels between the freedom and excitement of the wireless and the Internet in this engaging read. *Gillian Engberg Copyright © American Library Association. All rights reserved*

ORDER - U.S.A. - Radio Rescue now!

ORDER - Canada - Radio Rescue

ORDER - U.K. - Radio Rescue - ORDER NOW U.K.

ANOTHER EXCELLENT SCIENCE FAIR PUBLICATION

Janice Vancleave's Guide to More of the Best Science Fair Projects by Janice Pratt Vancleave - 160 pages

This book by Janice Vancleave I can recommend to concerned parents and proves a very inexpensive source of the "Best Science Fair Projects". This book aims to help children to have fun and get winning results as well. Advice is given about choosing, developing and displaying a winning project, along with dozens of science fair experiments.

ORDER - U.S.A. - Janice Vancleave's Guide to More of the Best Science Fair Projects - Janice Vancleave
ORDER - U.K. - Janice Vancleave's Guide to More of the Best Science Fair Projects - Janice Vancleave
ORDER - La FRANCE - Janice Vancleave's Guide to More of the Best Science Fair Projects - Janice Vancleave

ORDER - DEUTSCHLAND - Janice Vancleave's Guide to More of the Best Science Fair Projects - Janice Vancleave

---

Google Custom Search

Google Custom Search

Got a question on this topic?

If you are involved in electronics then consider joining our **"electronics Questions and Answers"** news group to ask your question there as well as sharing your thorny questions and answers. Help out your colleagues!.

The absolute fastest way to get your question answered and yes, I **DO** read most posts.

This is a mutual help group with a very professional air about it. I've learn't things. It is an excellent learning resource for lurkers as well as active contributors.

RELATED TOPICS on crystal radio set

Check out Bruce's **"Radio Kits for Beginners and Youngsters of ALL Ages!"**

Fox Hole Radio

antenna basics

capacitance

diodes

inductance

resonant frequency

The Skywaves High-Performance Crystal Set

am radio receivers

radio receiver basics

tuned radio frequency TRF receivers

regenerative radio receivers

superhetrodyne radio receivers

fm radio receivers

Click here
to rate
this page!

TELL A FRIEND
WORD OF MOUTH WORKS!

Please send me your valuable comments and suggestions! Tell your friends, tell a news group, tell the world!

YOU ARE HERE: HOME > RECEIVERS > CRYSTAL RADIO SET

the author Ian C. Purdie, VK2TIP of www.electronics-tutorials.com asserts the moral right to be identified as the author of this web site and all contents herein. Copyright © 2000 - 2001, all rights reserved. See **copying and links**. These electronic tutorials are provided for individual private use and the author assumes no liability whatsoever for the application, use, misuse, of any of these projects or electronics tutorials that may result in the direct or indirect damage or loss that comes from these projects or tutorials. All materials are provided for free private and public use. Commercial use prohibited without prior written permission from www.electronics-tutorials.com.

Copyright © 2000 - 2001, all rights reserved. URL - https://www.electronics-tutorials.com/receivers/crystal-radio-set.htm

Updated 27th April, 2001

Contact VK2TIP

WWW.ELECTRONICS-TUTORIALS.COM

Recommend 47

Share

G+1 2

•**NEW!** ‣ - Amazon Electronic Component Packs. ***Check out the Amazon Electronic Component Packs page.***

WHAT ARE THE BASICS OF CRYSTAL RADIO RECEIVERS?

Here is the schematic diagram for a very basic crystal radio set without any particular embellishments. This basic old time radio uses no power other than that provided by the transmitting antenna from the radio station. Free power from the sky eh!. It truly is a marvel!.

As a matter of interest, some amateurs and experimenters living in the vicinity of high powered AM Radio transmitters have literally used crystal radio sets to trickle charge their batteries. Now THAT is real COOL! If the demand exists we will do it too!.

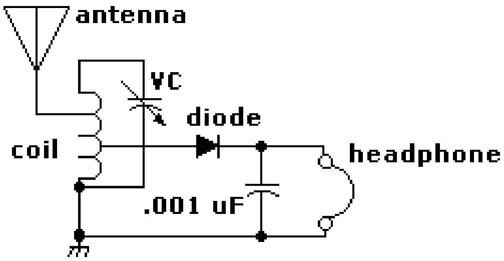

Figure 1 - schematic of a basic crystal radio set

NOTE: For people unfamiliar with metric measurements used here 25.4 mm equals 1"

This circuit consists of an **inductor** (also called a coil), a **variable capacitor** (used to be called a variable condenser), a germanium **diode** (formerly called a crystal), a filtering capacitor and finally very high impedance headphones.

The inductor has taps on it:

(a) one to connect the antenna to.
(b) one lower down to connect the detector diode to.

These taps are often best determined by experimentation

The variable capacitor is usually connected across the whole of the inductor to form a **tuned circuit** for our crystal radio set.

Note: the bottom junction of the inductor and variable capacitor is connected to ground / earth - **literally**.

Earth connection for crystal radio set

For the crystal radio set circuit to perform at all well you will need a very good earth connection. A length of old steel water pipe driven **deep** into damp earth is ideal. The electrical earth or water service at home is **NOT** a good or safe earth for this project, in fact it is **VERY DANGEROUS**. - see **earth dangers**.

The funny looking symbol connected to the top of the inductor or coil and variable capacitor denotes an antenna. The higher and longer (50' or 17 metres) this antenna is, the better the likely reception. Reality says of course you can erect an antenna in the space available with the material available. Consider running the antenna around the top of your fence line if is not metallic.

Safety comes first with our crystal radio set

KEEP WELL AWAY FROM POWER LINES AND POWER SOURCES - **REPEAT AFTER ME** - I PROMISE TO KEEP WELL AWAY FROM POWER LINES AND POWER SOURCES.

Always remember the motto "safety first in everything". At this site here I stress that all the time - "safety first" and I can never stress it enough. See my important page about **earth dangers**

Antenna for crystal radio set

Remember we're using the antenna (aerial) to get free power here and we want to capture as much as possible from the sky for our crystal radio set.

Perhaps you might like to review some **antenna basics** before proceeding.

Technical Note: Radio signals (waves) such as we encounter in the am radio band have two halves. One half travels across the surface of the earth at the speed of light through people, buildings and other objects. The other half, a mirror image, travels beneath the surface of the earth. This radio wave has a definite length. Its length is the speed of light divided by the frequency. For frequencies in the megahertz region this approximates to 300 divided by the frequency in megahertz. These radio waves we want to detect with our crystal radio set

An example which we use later is the **am radio band** of 530 Khz to 1650 Khz. These frequencies when converted to .53 Mhz and 1.65 Mhz respectively give wavelengths of 566 metres and 182 metres.

Now some critical aspects:

Variable Capacitor: Here you should try and salvage one from an old time radio. Just make sure the radio is not in the "vintage" class and now worth hundreds of dollars. The ideal ones are nearly impossible to buy now. Unless you have access to measuring equipment you will have to guess the capacitance. A lot of old time radios had capacitors which tuned about 15 pF to 365 pF.

Inductor or Coil: Here we are normally dealing with an air cored inductor wound on some suitable **non metallic** form. Some examples are - cardboard toilet roll former, a short length of PVC tubing 40 mm, 50 mm, 90 mm or 100 mm diameter.

Diode: In lieu of the old crystal detector we use a germaniumdiode of the 1N34 or OA90 type. Do NOT use a 1N914 type.

Fixed Capacitor: This is for filtering and may be .001 uF, 1 NF or 1000 pF type (all those values are the same - just expressed in different units). Something near in the same values is quite O.K.

Headphones: This is by far the hardest part to obtain. The type used for hi-fi **WILL NOT** work here. Ideally you need high impedance 2,000 ohm types but these are nearly impossible to find. You can sometimes buy 1,000 ohm crystal earpieces (no **not** the usual transistor radio type). The headphone is a high impedance load for the crystal set and as we are working on free power from the air we can't load it down. If you used the 8 ohm hi-fi type it would be like trying to run your electric toaster off a very small battery. The power is not available. Remember we're using free power from the sky.

Further reading on **Headsets for Crystal Radios** by Alan R. Klase.

Frequency range: The frequency range of a set like this is mainly determined by the square root of the *ratio of Maximum Capacitance to Minimum Capacitance of the variable capacitor.* In the example above this would be the square root of 365 divided by 15 or about 4.9:1. In the real world there are other capacitance's in the circuit that you can't see (called "strays") and they affect the minimum value. In practice I would expect a real range of about 3:1

A.M. Radio band: Depending upon your country this would cover about 530 Khz to about 1650 Khz. This is slightly more than a 3:1 ratio.

Let us assume for discussion our capacitor was as described above. And for convenience we have strays of about 25 pF (typical). If we add this to both the Max and Min capacitance we get 390 pF Max and 40 pF Min - a ratio of 9.75 the square root of which is 3.122. Happily this also happens to be nearly the ratio of Freq $_{max}$ divided by Freq $_{min}$ or 1650 Khz / 530 Khz. How about that?

Of course these variable capacitors were deliberately manufactured that way for that very reason!

Well now we have to resonate our capacitor with the inductor. What inductance do we need?. The easiest formula is this - take the frequency in Megahertz (i.e. freq in Khz divided by 1000) and square it. Using our two frequencies 530 Khz and 1650 Khz we would have to square both .530 Mhz and 1.65 Mhz which of course become 0.2809 and 2.7225

And you thought maths at school was/is a waste of time!.

These two figures are divided into the standard figure of 25330.3 respectively. The 530 Khz one (0.2809) would give us LC 90175 and the 1650 Khz one (2.7225) will give us LC 9304. Now when tuning a crystal set the more capacitance you use (plates meshed further in) the lower the frequency you receive. In our case we are using a max of 365 pF + 25 pF strays or 390 pF total. At our lowest freq of 530 Khz the LC (L means inductance and C means capacitance) is 90175 and when divided by our C of 390 pF we get a needed inductance of 231.2 uH (uH means micro henry, a unit of inductance).

Similarly if we go to the highest frequency of 1650 Khz our required LC is 9304 and when divided by our minimum capacitance's (reduce the capacitance by withdrawing plates from one another and go higher in frequency) of 15 pF + 25 pF strays or 40 pF we get a required inductance of _______ uH?. Go on work it out. And remember we are only interested in whole numbers because we have made a number of approximations here.

Now how do we get a needed inductance of about **230 uH?**.

Gee whizz not more maths headed our way!

The formula for inductance - using toilet rolls, PVC pipe etc. can be well approximated by:

$$\text{Inductance } L = \frac{0.394 * r^2 * N^2}{(9 * r) + (10 * Len)}$$

Here:

r = radius of the coil i.e. form diameter (in cm.) divided by 2
Len = length of the coil - again in cm.
L = inductance in uH.
* = multiply by

Now assume we have used a piece of PVC pipe of about 104 mm (10.4 cm) diameter. This has a radius of 5.2 cm. Let us wind a coil consisting of 57 turns (with taps at about every 5 turns) over a length of 104 mm or 10.4 cm. For the Mum's and Dad's who work in old measurements that's 4". This is a piece of standard 4" PVC sewer tube.

Politely ask any plumber on any building site for a scrap off-cut about 150 mm (6") long. In another life I'm also a plumber and we like to help polite people.

Let's put those numbers into the our formula above and do some sums:

0.394 * 5.2 * 5.2 * 57 * 57 = 34614

AND (9 * 5.2) + (10 * 10.4) = 46.8 + 104 = 150.8

AND 34614 divided by 150.8 = about 230 uH

Boy am I a clever kid or what. Did you work through that on paper all by yourself? I do hope so.

Why and how taps every 5 turns?. The why first - in the beginning I said connections from the antenna (aerial for old timers) and diode would need to be experimented with. With taps you can select different spots along the inductor.

How?. Easy! Wind your inductor as follows:

You will need a length of enamelled copper winding wire (about 24 gauge or 0.5 mm dia) about 20 metres long - yes you will need just about all of it. You will also need some matches (used are O.K.). Assuming you have a 150 mm piece of sewer tube or similar, drill a small hole approx 25 mm (1") from one end. Drill a similar hole 104 mm further away toward the opposite end (or even 20 mm in from the opposite end). Basically the idea is for the holes to be nearly 10.4 cm apart. or *the length of the coil you are winding*.

Push through about 15 cm (150 mm) of wire and wind the short end through a couple of times to form an anchor (knot) at that point. Then holding the wire tightly wind 5 complete turns spaced about three wire diameters apart from each other (you really need a helper here). Under the 5th turn slip a very short length of match, continue winding very tightly placing a match piece under every 5th turn until you have wound 57 complete turns. Tie a knot similar to the other end and again leave about 150 mm left over. **CAREFULLY** with a sharp knife scrape the enamel insulation away on all 5th turns as well as about 5 mm from the end of the two end pieces.

Presto! there you have a 57 turn inductor of about 230 uH with accessible taps at about every 5th turn.

How it works: (simple version)

Well we have the antenna. This picks up the radio signals from the sky and with our real good earth we hopefully get lots and lots of signals.

The inductor combined with different settings of the variable capacitor selects the frequency you want to listen to and discards all others.

This selected signal goes through the germanium diode and is rectified. This means we are left with the audio frequencies a human being can hear (music - words). This audio has been stripped off the radio signal by the diode. The small .001 uF capacitor gets rid of any of the radio signal left. We can then listen to these very tiny audio signals through our headphones or crystal earpiece.

How it works: (complicated version)

Your long wire antenna and earth wire will pick up thousands and thousands of radio signals from the air. The exact setting of your variable capacitor will determine a set amount of capacitance and this in conjunction with the fixed inductance of your coil we have a **resonant circuit**.

As an example, in my locality we have a very strong AM station at 630 Khz. Using the formula above for **resonant frequency** we find the LC combination (inductance times capacitance) required to receive this signal is:

$25330.3 / (0.63)^{2} = 63820$

Therefore if we had a fixed inductance of 230 uH then our variable capacitor (including strays) would need to be set at the point where it exhibits 63820 / 230 **OR** 277 pF. In other words a capacitance of 277 pF in conjunction with an inductance of 230 uH resonates at 630 Khz. We can rather generalise here and say, although not strictly accurate, that this coil / capacitor combination will **ACCEPT** only signals at 630 Khz and **REJECT** all others. This is the way we select our particular frequency from among the thousands and thousands of other signals on our antenna.

Our signal at 630 Khz has been "modulated" at the radio station with audio frequencies (speech and music). Because it is **A**mplitude **M**odulated it is called **A.M.**

This means the original 630 Khz signal strength (amplitude) at the transmitter is expanded and contracted by the audio frequencies.

After being selected by our coil / capacitor combination the 630 Khz signal is passed to the germanium diode. As the signal passes through the diode only one half of the signal is used and the fixed .001 uF capacitor dumps the 630 Khz signal to ground leaving a very tiny audio frequency which is an exact replica of the audio frequencies (speech and music) as they originated in the radio station's studio.

This very tiny audio signal then drives our crystal earpiece or high impedance headphone so we can hear the words and music. If it were amplified many, many times over it would then have the power to drive a loudspeaker. The speaker cone goes in and out at the same audio rate as the original 630 Khz signal is modulated. The moving loudspeaker cone creates a tiny sound pressure level in the room to which your ear is able to respond.

Radio Kits for Beginners and Youngsters of ALL Ages!

My special amateur radio friend Bruce Kizerian, KK7ZZ who has worked tirelessly for years helping the young and not so young get effortlessly into radio now has a site devoted to "Radio Kits for Beginners and Youngsters of ALL Ages!".

There has been a crying need for these kits and Bruce now fills that need.

Check out Bruce's **"Radio Kits for Beginners and Youngsters of ALL Ages!"**

NEW! - "High Sensitivity Crystal Set"

Build a “crystal radio ” by using a new zero-voltage-threshold MOSFET. By Bob Culter, N7FKI as printed in QST [Amateur Radio Magazine] January 2007

Check out the PDF file of the whole article - **"http://www.arrl.org/files/file/Technology/tis/info/pdf/culter.pdf"** 467 kB. **PS:** The article can be more than somewhat highly technical for newcomers though.

NEXT:

Put it all together and reload this page by typing in the URL - www.electronics-tutorials.com/receivers/crystal-radio-set.htm and **tell me how your crystal set works** for you.

If you are real keen you can also check out the **crystal set society's** home page.

Another excellent article (in pdf format) is from the ARRL. From QST December, 1997 **The Crystal Radio**

Fox Hole Radio

Another VERY interesting project is the **Fox Hole Radio**. Building a foxhole radio is rewarding and the basic setup is very simple. It is, however, difficult to adjust, and it may take several attempts to find a proper razor blade for the detector. This is a project that requires patience and much trial and error, but it will pay off once it begins to work. It will help to be versed in the construction and operation of crystal sets before building one. These sets are extremely simple in construction, but tuning and modification require some basic understanding of theory, as well as practice. All sets presented here are based on old articles, notes, and people's recollections. There are fairly major variations in design and materials among these plans. It must be remembered that these were improvised under often adverse conditions; there was no "standard" design. With this in mind, take this entire article as a whole, and use it a bit here, a bit there, to build towards a Fox Hole Radio design that works best using modern materials.
See: **Fox Hole Radio**

A Candle powered Radio

Yep a World Book Science Project from a looooonnnng time ago.
See: **Candle powered Radio**

A BOOK I THOROUGHLY RECOMMEND FOR YOUNG PEOPLE

Radio Rescue
by Lynne Barasch

From Booklist
Gr. 2-5. In 1923, a 10-year-old New Yorker became the youngest licensed amateur wireless radio operator. In a clear, first-person voice, the boy describes what Morse code is, how operators used wireless radios to send and receive signals, and his experiences setting up an in-house transmitting station. The boy finds fame when he picks up a signal from Florida hurricane victims and radios for help, and the story concludes with a reproduced newspaper photograph and the full Morse code alphabet. An introductory note

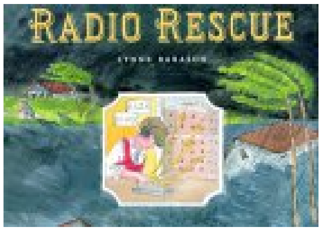

reveals that the story is based on the author's father and gives readers background information on the era's telephones and the alternative that the wireless offered, including examples of how Morse code has been used throughout its history. In a well-designed mix of insets, brief sketches, and full-page drawings, the author's uncluttered color cartoons do an excellent job of illustrating the technology and the code, at the same time creating likable, expressive characters. Tehnology-minded children may see parallels between the freedom and excitement of the wireless and the Internet in this engaging read. *Gillian Engberg Copyright © American Library Association. All rights reserved*

ORDER - U.S.A. - Radio Rescue now!

ORDER - Canada - Radio Rescue

ORDER - U.K. - Radio Rescue - ORDER NOW U.K.

ANOTHER EXCELLENT SCIENCE FAIR PUBLICATION

Janice Vancleave's Guide to More of the Best Science Fair Projects by Janice Pratt Vancleave - 160 pages

This book by Janice Vancleave I can recommend to concerned parents and proves a very inexpensive source of the "Best Science Fair Projects". This book aims to help children to have fun and get winning results as well. Advice is given about choosing, developing and displaying a winning project, along with dozens of science fair experiments.

ORDER - U.S.A. - Janice Vancleave's Guide to More of the Best Science Fair Projects - Janice Vancleave
ORDER - U.K. - Janice Vancleave's Guide to More of the Best Science Fair Projects - Janice Vancleave
ORDER - La FRANCE - Janice Vancleave's Guide to More of the Best Science Fair Projects - Janice Vancleave
ORDER - DEUTSCHLAND - Janice Vancleave's Guide to More of the Best Science Fair Projects - Janice Vancleave

---

Search

Google™ Custom Search

Got a question on this topic?

If you are involved in electronics then consider joining our **"electronics Questions and Answers"** news group to ask your question there as well as sharing your thorny questions and answers. Help out your colleagues!.

The absolute fastest way to get your question answered and yes, I **DO** read most posts.

This is a mutual help group with a very professional air about it. I've learn't things. It is an excellent learning resource for lurkers as well as active contributors.

RELATED TOPICS on crystal radio set

Check out Bruce's **"Radio Kits for Beginners and Youngsters of ALL Ages!"**

Fox Hole Radio

antenna basics

capacitance

diodes

inductance

resonant frequency

The Skywaves High-Performance Crystal Set

am radio receivers

radio receiver basics

tuned radio frequency TRF receivers

regenerative radio receivers

superhetrodyne radio receivers

fm radio receivers

Click here
to rate
this page!

TELL A FRIEND
WORD OF MOUTH WORKS!

Please send me your valuable comments and suggestions! Tell your friends, tell a news group, tell the world!

YOU ARE HERE: HOME > RECEIVERS > CRYSTAL RADIO SET

the author Ian C. Purdie, VK2TIP of www.electronics-tutorials.com asserts the moral right to be identified as the author of this web site and all contents herein. Copyright © 2000 - 2001, all rights reserved. See **copying and links**. These electronic tutorials are provided for individual private use and the author assumes no liability whatsoever for the application, use, misuse, of any of these projects or electronics tutorials that may result in the direct or indirect damage or loss that comes from these projects or tutorials. All materials are provided for free private and public use.
Commercial use prohibited without prior written permission from www.electronics-tutorials.com.

Copyright © 2000 - 2001, all rights reserved. URL - http://www.electronics-tutorials.com/receivers/crystal-radio-set.htm

Updated 27th April, 2001

Contact VK2TIP

BUILD A FREE WEBSITE OF YOUR OWN ON Angelfire
Click here to find a Relay For Life near you.
RELAY FOR LIFE American Cancer Society

undefined

Home · Contact · Free Radiosets · How Earphones Work · Aerials & Earths **Crystal Radio Sets on the Net!**

HOME

GETTING STARTED
Introduction
How Earphones Work
Aerial & Earth Wires

RADIO KITS
The Nic Nac Radio
MK 484 TRF Receiver

COMING SOON!
Book of Radio Kits

CONTACT US
E-mail
Pre-order our Book
On-line Form

Fun with Radio and Electronics!

Print out this page and have a go at building a crystal radio set yourself!*

INTRODUCTION

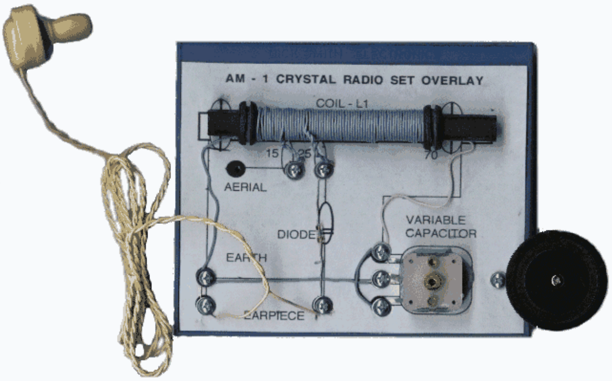

If you're stuck inside on a Saturday afternoon with a junk box full of electronic parts and nothing much to do, you can spend an hour or so and build this very simple but effective Crystal Radio receiver. Often, the satisfaction of building these simple radios is lost amidst the latest computer program or the latest digital marvel chip!

THEORY

Briefly, a Crystal Radio receiver derives it energy from. the radiated power from the AM radio stations in your area, This energy travels in waves, and is induced into your Aerial/Earth wire system (more about safe Aerial and Earth wires later on) and then fed into the coil/capacitor circuit. This coil/capacitor combination is known as a "tank" circuit, because it. stores radiated energy. The coil stores this energy in a pulsating electro-magnetic field, while the. variable capacitor stores energy in 'an electro-static field between its fixed and moving sets of plates.

The Ferrite Rod acts as a bar magnet for radio waves. concentrating them inside the turns of the coil. When.you turn the knob of the variable capacitor you change the value of thisdevice. Changing the value of the capacitor will vary the frequency at which"the Crystal Radio receives. When you tune into a local AM radio station, the tank circuit experiences what is called "resonance". This is like a musician using a tuning fork to tune up the strings of his instrument. Each string will 'resonate' with the chosen tuning fork when they are both pitched at the same frequency.

!n the same way, when a radio station and the Crystal Radio are tuned into the same frequency, resonancetakes place. This state of resonance allows for maximum signal transfer between transmitter and receiver. The selected frequency is then fed into the Germanium Diode detector. This corn portent separates (rectifies) the radio station's "carrier wave" from the program signal (voice and music). The rectified signal is thenfed in the Crystal Earpiece where it is converted into sound waves that we can hear.

OVERLAY FOR CRYSTAL SET

(Refer to printable document for accuracy)

Parts List Dick Smith Electronics	
Ferrite Rod	R5107
Diode OA91	Z3040
VC1 220pf	R2970
Earpiece	C2765
Coil wire (3m)	W2123
Aeria/Earth wire	W2260
8 x screws	H1634
8 x washer	H1488
1 x b'board	H4050

CONSTRUCTION

Our radio set can be made up on a square of wood or plastic around 10cm by 12cm by 1cm. If you or a parent can solder, then feel free to drive nickel plated nails into the wooden baseboard. Then you can "tin" these with a bit of solder and use each nail head as a tie point to interconnect the component leads. **BE VERY CAREFUL** with soldering tools - wear eye safety glasses at all times. Alternatively, you can use a screw/washer combination with the DSE plastic baseboard. These should be the #4x9mm screws and the washers around 5mm in diameter (see parts list). If you decide to solder, take special care with soldering the Diode and the solder tags from the Variable Capacitor. The Diode should only be soldered by using a heat sink (a simple device that will draw any excess heat away from the Diode itself) and this can be a crocodile clip attached to either end of the Diode's leadouts (close to the glass bulb) during soldering. The caseing of the Variable Capacitor is made of plastic, and any excessive heat applied during soldering may well melt a part of this fragile case.

You can use other coils in this simple Crystal Radio design. You can either wind an air-cored coil (if you can't get hold of a Ferrite Rod easily) as shown below (comprised of the same number of turns -70) wound on a paper roll tube complete with taps (at 15 turns and 25 turns) or you can retrieve a factory made coil from an old, disused transistor radio. This coil actually has two windings, and you can tune the Crystal Radio by wiring a 120pF capacitor (instead of the variable capacitor) across the main winding (usually 'white and black wires) and sliding the square Ferrite Rod up and down inside the coil former to tune in local stations (see diagram below).

AERIALS AND EARTHS

The best Aerial is one that is at least two metres high off the ground (higher if possible) and at least ten metres long. You MUST stay well clear of overhead power and telephone cables for safety's sake. The safest Earth wire is one that's connected to a cold water pipe that is embedded in moist soil. A garden tap fitting just outside a window is ideal. Do NOT use the earth pin of a 240 volt power outlet, nor a hot water pipe, nor a gas pipe, as **these are all unsafe** to use as Earth connections, and may well result in electrical shocks or explosions occuring.

DO NOT operate this Crystal Radio receiver during an electrical thunderstorm.

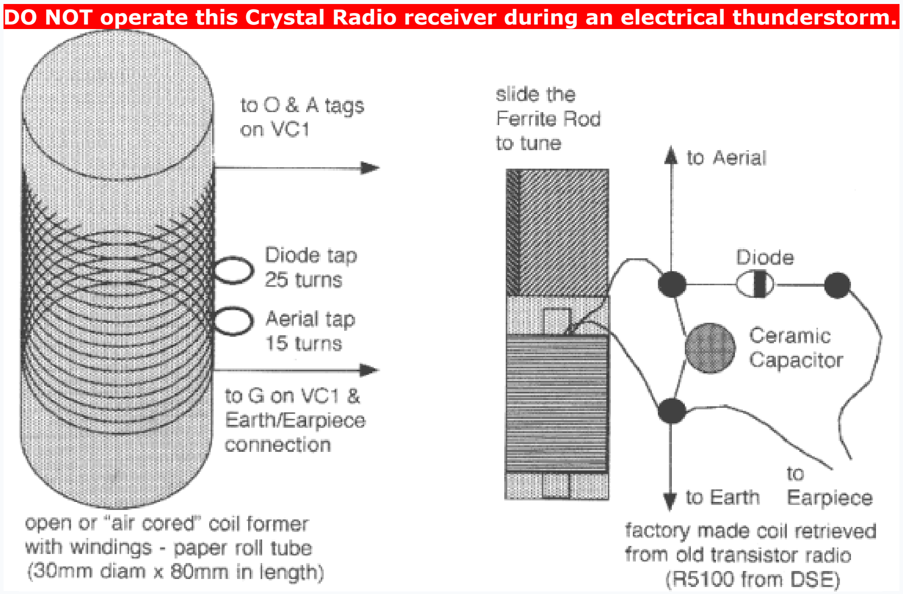
to O & A tags
on VC1
Diode tap
25 turns
Aerial tap
15 turns
to G on VC1 &
Earth/Earpiece
connection
open or "air cored" coil former
with windings - paper roll tube
(30mm diam x 80mm in length)
slide the
Ferrite Rod
to tune
to Aerial
Diode
Ceramic
Capacitor
to Earth
to
Earpiece
factory made coil retrieved
from old transistor radio
(R5100 from DSE)

design by **the manikin**

Copyright © **A. Hellier** 1998 **Disclaimer**.

CANCER DOESN'T COME WITH INSTRUCTIONS.
LIVESTRONG
LEARN HOW WE CAN HELP

BUILD A FREE WEBSITE ON Angelfire
American Diabetes Association. diabetes

BUILD A FREE WEBSITE OF YOUR OWN ON Angelfire
Click here to find a Relay For Life near you.
RELAY FOR LIFE American Cancer Society

undefined

Home · Contact · Free Radiosets · How Earphones Work · Aerials & Earths
Crystal Radio Sets on the Net!

HOME
GETTING STARTED
Introduction
How Earphones Work
Aerial & Earth Wires
RADIO KITS
The Nic Nac Radio
MK 484 TRF Receiver
COMING SOON!
Book of Radio Kits
CONTACT US
E-mail
Pre-order our Book
On-line Form

The Nic Nac Crystal Radio

Copyright © **Austin Hellier** 12/02/2001

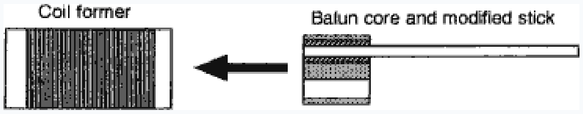

How It Works

Radio signals are travelling through the atmosphere all the time. This simple radio receiver has been designed to capture them, and turn them into sound waves, so that you can hear what's going on. The Aerial and Earth wires form a return signal path from the AM transmitter, for these signals, and they are then fed into the coil and capacitor. This arrangement is known as a 'tank' circuit, because it stores radio energy, much in the same way your family car's tank stores petrol. The combination of a certain sized capacitor (68pF) and the amount of turns on the coil (around 120 turns) determines the frequency you will receive at. The Diode changes the signal from a radio one, to an audio one, and the Crystal Earphone changes the audio signal into sound waves that you can hear. the resistor adds stability by helping to prevent the Earphone from occasional 'dropout'.

Building The Radio

You'll need all the components listed below and a Tic Tac lolly box. Start by mounting the earphone jack into the lid of the box and then solder up each of the components into a tidy version of the 'rat's nest' construction method. Make sure you use a crocodile clip as a heat sink for the diode when soldering - they're pretty sensitive devices to heat. Then take the coil and remove the ferrite rod from inside it. Take the ferrite balun core and place one end of the plastic stick (obtained by cutting off both ends of a cotton bud) and glue it inside the hole, with its end level with the end of the balun core.

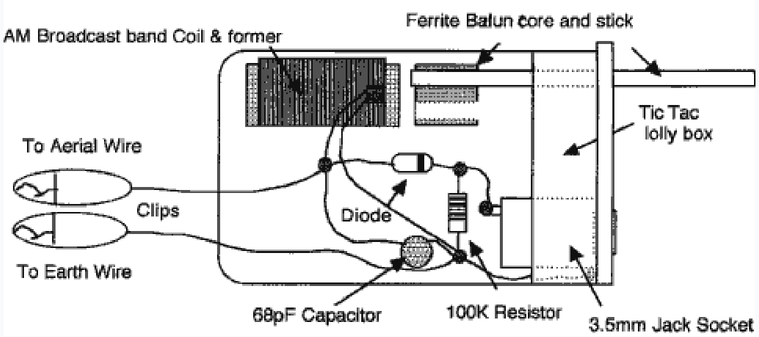

Diagram 2. Place balun inside coil former and use the stick to slide it up and down inside the coil to tune the Nic Nac Radio.

Experimenting With The Radio

You can perform a number of experiments with your new radio receiver, so try the following experiments and have lots of fun while you do:

1. Try hooking up your Aerial wire first, without the Earth wire, to see if you're close enough to a radio station, not to need the Earth wire.

2. Try using different things for an Aerial, instead of the long wire - e.g.: a wire clothes line, a Hills Hoist clothes line, a metal window frame, a metal security door, a cyclone fence, bedsprings etc.

3. If you live really close to a radio transmitter, try rotating the Crystal Radio without an Aerial or Earth wire, to see if you can pick up these local stations.

I hope that you have as much fun making the **Nic Nac Radio**, as I did in designing it.

design by **the manikin**

Copyright © **A. Hellier** 1998 **Disclaimer**.

CANCER DOESN'T
COME WITH
INSTRUCTIONS.
LIVESTRONG
LEARN HOW WE CAN HELP

BUILD A FREE WEBSITE ON
Angelfire

American Diabetes Association® diabetes®

BUILD A FREE WEBSITE OF YOUR OWN ON Angelfire
UGG
SHOP NOW

undefined

The Nic Nac Crystal Radio

Copyright © **Austin Hellier** 12/02/2001

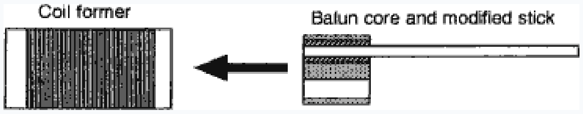

How It Works

Radio signals are travelling through the atmosphere all the time. This simple radio receiver has been designed to capture them, and turn them into sound waves, so that you can hear what's going on. The Aerial and Earth wires form a return signal path from the AM transmitter, for these signals, and they are then fed into the coil and capacitor. This arrangement is known as a 'tank' circuit, because it stores radio energy, much in the same way your family car's tank stores petrol. The combination of a certain sized capacitor (68pF) and the amount of turns on the coil (around 120 turns) determines the frequency you will receive at. The Diode changes the signal from a radio one, to an audio one, and the Crystal Earphone changes the audio signal into sound waves that you can hear. the resistor adds stability by helping to prevent the Earphone from occasional 'dropout'.

Building The Radio

You'll need all the components listed below and a Tic Tac lolly box. Start by mounting the earphone jack into the lid of the box and then solder up each of the components into a tidy version of the 'rat's nest' construction method. Make sure you use a crocodile clip as a heat sink for the diode when soldering - they're pretty sensitive devices to heat. Then take the coil and remove the ferrite rod from inside it. Take the ferrite balun core and place one end of the plastic stick (obtained by cutting off both ends of a cotton bud) and glue it inside the hole, with its end level with the end of the balun core.

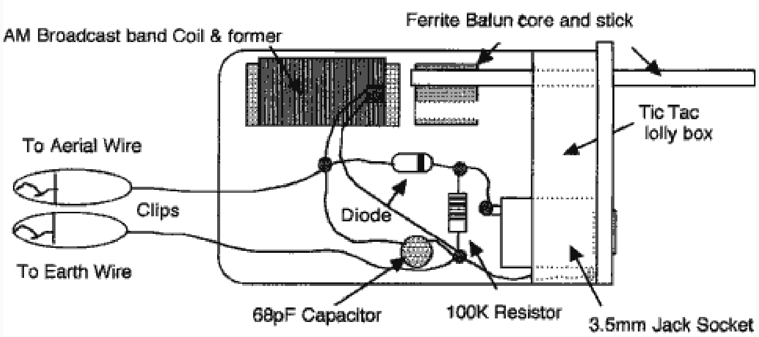

Diagram 2. Place balun inside coil former and use the stick to slide it up and down inside the coil to tune the Nic Nac Radio.

Experimenting With The Radio

You can perform a number of experiments with your new radio receiver, so try the following experiments and have lots of fun while you do:

1. Try hooking up your Aerial wire first, without the Earth wire, to see if you're close enough to a radio station, not to need the Earth wire.

2. Try using different things for an Aerial, instead of the long wire - e.g.: a wire clothes line, a Hills Hoist clothes line, a metal window frame, a metal security door, a cyclone fence, bedsprings etc.

3. If you live really close to a radio transmitter, try rotating the Crystal Radio without an Aerial or Earth wire, to see if you can pick up these local stations.

I hope that you have as much fun making the **Nic Nac Radio**, as I did in designing it.

design by **the manikin**

Copyright © **A. Hellier** 1998 **Disclaimer**.

WordPress
It's what we do

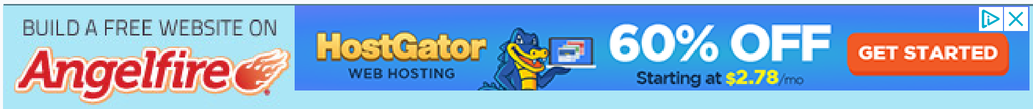

BUILD A FREE WEBSITE ON Angelfire
HostGator WEB HOSTING 60% OFF Starting at $2.78/mo GET STARTED

n°2 - 1954

IL FIDO

Un eccellente ricevitore a galena

Forse parte dei nostri lettori sorrideranno a leggere questo titolo pensando che la galena è un apparecchio ormai passato di moda, e che ad altro non serva che a figurare in un museo o in una esposizione retrospettiva di radio. Questa è una supposizione sbagliata, infatti la prima cosa che fanno coloro che desiderano entrare nel campo radio, è quella di affrontare la costruzione di un ricevitore a galena, poi quella di un ricevitore ad una valvola e li vediamo poi in un anno o due, diventare perfetti radiotecnici. Non solo, ma tutti coloro che desiderano avere il loro piccolo ricevitore da usare magari a letto, cosa costruiscono? un apparecchio a galena.

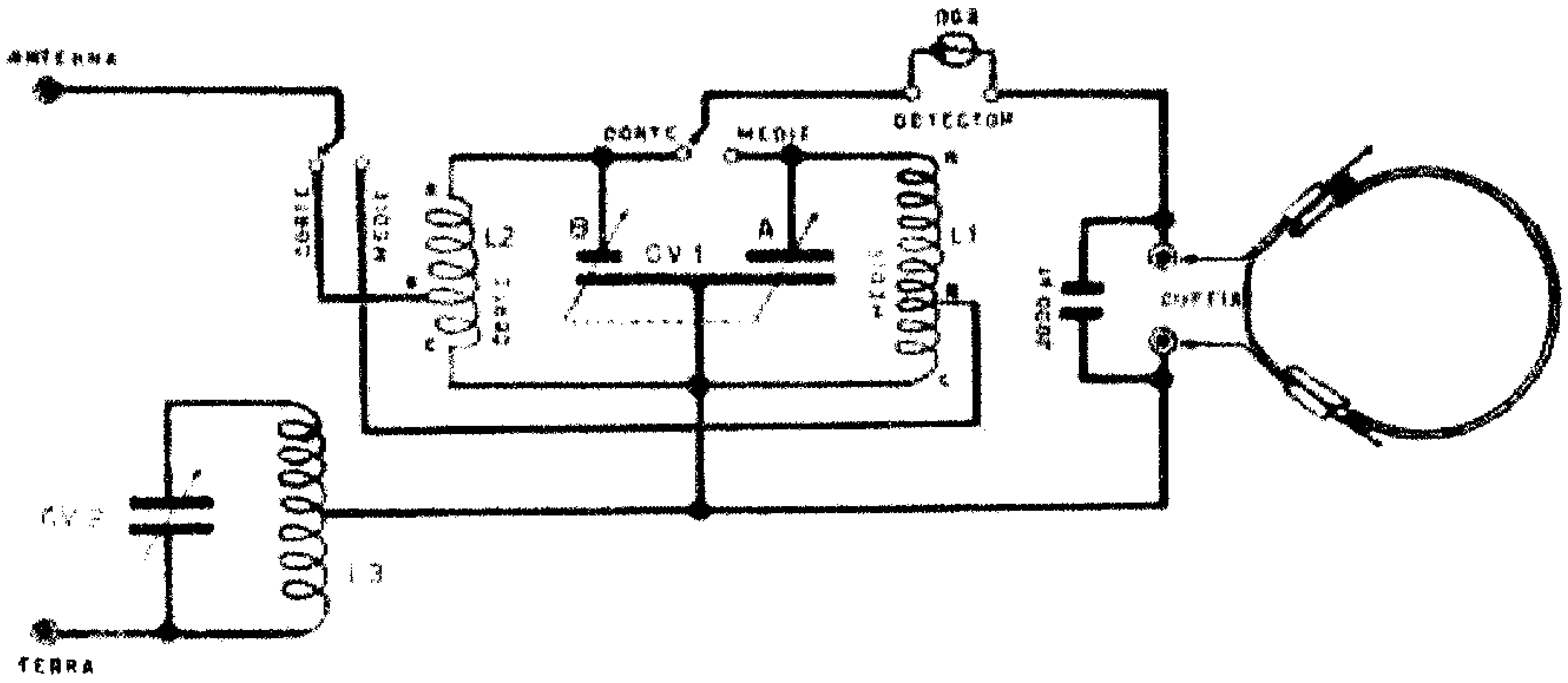

Vi è poi chi deve dimorare per vario tempo in un ospedale, chi non può disporre di una qualsiasi sorgente di alimentazione per far funzionare un ricevitore a valvola, ed infine tutti coloro che non possono, causa il prezzo elevato, costruire o acquistare un apparecchio normale, trovano nel ricevitore a galena il loro unico mezzo per godere di una bella trasmissione. Possiamo quindi affermare che il ricevitore a galena, non è abbandonato, ed anzi ora che i diodi di germanio semplificano ancora di più il circuito, e la loro sensibilità è maggiore, accrescono le possibilità e i desideri di molti lettori, che sempre ci scrivono perchè pubblichiamo ricevitori a galena o a diodi di germanio.

Come funziona un ricevitore a galena

Bisogna ben comprendere che una radio a galena, contrariamente alla radio a valvole termoioniche, non ha nessun dispositivo amplificatore; la sola antenna captando le emissioni le invia ad un circuito oscillante dove, per mezzo del piedistallo, viene rettificato. Occorre però, oltre a captare l'emmissione desiderata, selezionarla ed escluderla da tutte le altre. Questa selezione si opera sempre per mezzo di un condensatore variabile e di una bobina d'antenna. Quando si hanno due o tre stazioni locali, è necessario ottenere una maggior selettività e questa può essere data solo da un buon circuito filtro composto da una seconda bobina e da un secondo condensatore variabile.

La galena o il diodo di germaio, posto come rettificatore, ha il compito di lasciare passare le alternanze di un solo senso; sono queste alternate che mettono poi in movimento le membrane della cuffia riproducendo fedelmente ogni suono. Che occorre fare per ricevere meglio le stazioni trasmittenti con una radio a galena?

Prima cosa: per ricevere bene, occorre installare un buon captatore di onde, vale a dire una buona antenna e preparare una buona presa di terra, poichè l'una non vale se non è buona l'altra. Poi occorrono dei buoni collegamenti, corti e perfettamente saldati specialmente, come nel nostro caso, se si desidera ricevere le onde corte. Infine occorre una buona cuffia abbastanza sensibile. Il « Fido » che proponiamo è il ricevitore che abbiamo studiato e provato per Voi perciò il risultato è sicuro e difficilmente si avranno insuccessi.

Il sistema di accordo è semplice e dopo tante prove è stato l'unico che ci ha dato il massimo di sensibilità. La selettività, grazie al circuito e filtro L3 e C2, è eccellente, tanto da poter selezionare completamente anche più di una stazione locale. Veniamo dunque a parlare un po' della presa d' antenna sulle bobine L1 e L2. La presa da noi effettuata è quella che ci ha dato migliori risultati. Chi avrà un po' di tempo disponibile, oppure chi trova diletto a sperimentare potrà provare di spostare la presa B cercando di ottenere, con la antenna usata, un rendimento migliore. Qualche lieve miglioramento si avrà pure modificando leggermente il numero delle spire delle bobine per le onde medie. Infatti coloro che desiderano, o che sono costretti ad ascoltare le stazioni tra i 400 e 500 metri, otterranno, con qualche spira in più nella bobina L1, un aumento di selettività e di potenza Questo ricevitore, oltre alla gamma delle onde medie, è pure previsto per la ricezione della gamma delle onde corte. Il passaggio delle onde viene effettuato per mezzo di una spinetta a banana, e due boccole da galena. Questo semplifica di molto il ricevitore rendendolo così adatto pure al principiante. Anche per l'antenna si usa il solito sistema delle due boccole da galena e della spinetta a banana. Per cambiare le onde è sufficiente inserire la spina dell'antenna sulla boccola che va alla presa media o corta della bobina sintonia cotemporaneamente alla presa del rivelatore a galena. Per ascoltare le onde medie l'antenna andrà inserita sulla presa B della bobina L1, e la galena sulla presa A sempre della bobina L1. Mentre per le onde corte, l'antenna andrà inserita sulla presa B della bobina L2 e la galena sulla presa A della bobina L 2. Per non usare due variabili, necessari per ascoltare le due gamme d'onda, abbiamo usato un variabile a due sezioni, uno con una capacità di 465 pF CV 1 A e uno con la capacità di 100 pF CV1B. Con questo sistema abbiamo ridotto, oltre allo spazio, il prezzo, poichè il variabile menzionato costa solo 100 lire in più di quanto costa un solo variabile. Quello che noi abbiamo adottato costa infatti L. 850. Dopo la parte Alta Frequenza composta dalla bobina e dai variabili di sintonia abbiamo il rivelatore a cristallo, che può essere convenientemente sostituito, ottenendo una maggiore sensibilità ed eliminando nello stesso tempo lo sgradevole inconveniente di cercare il punto sensibile del detector, da un diodo di germanio. Per chi ancora non lo sapesse nel diodo di germaio non c'è bisogno di cercare nessun punto sensibile così in luogo del detector sarà molto più conveniente usare un diodo di germaio tipo DG2 - OA5O - IN34 ecc. Per il diodo di germanio, non è necessario pure nessuna tensione di alimentazione per cui può essere perfettamente inserito nel nostro schema, apportandovi numerosi vantaggi. Al detector segue la cuffia di ricezione, la cui resistenza può essere compresa fra i 1000 e 4000 ohm. Questa cuffia viene shuntata da un condensatore da 2000 pF destinato a facilitare il passaggio della corrente Alta Frequenza che esiste dopo la rivelazione.

Terminata l'esposizione teo-

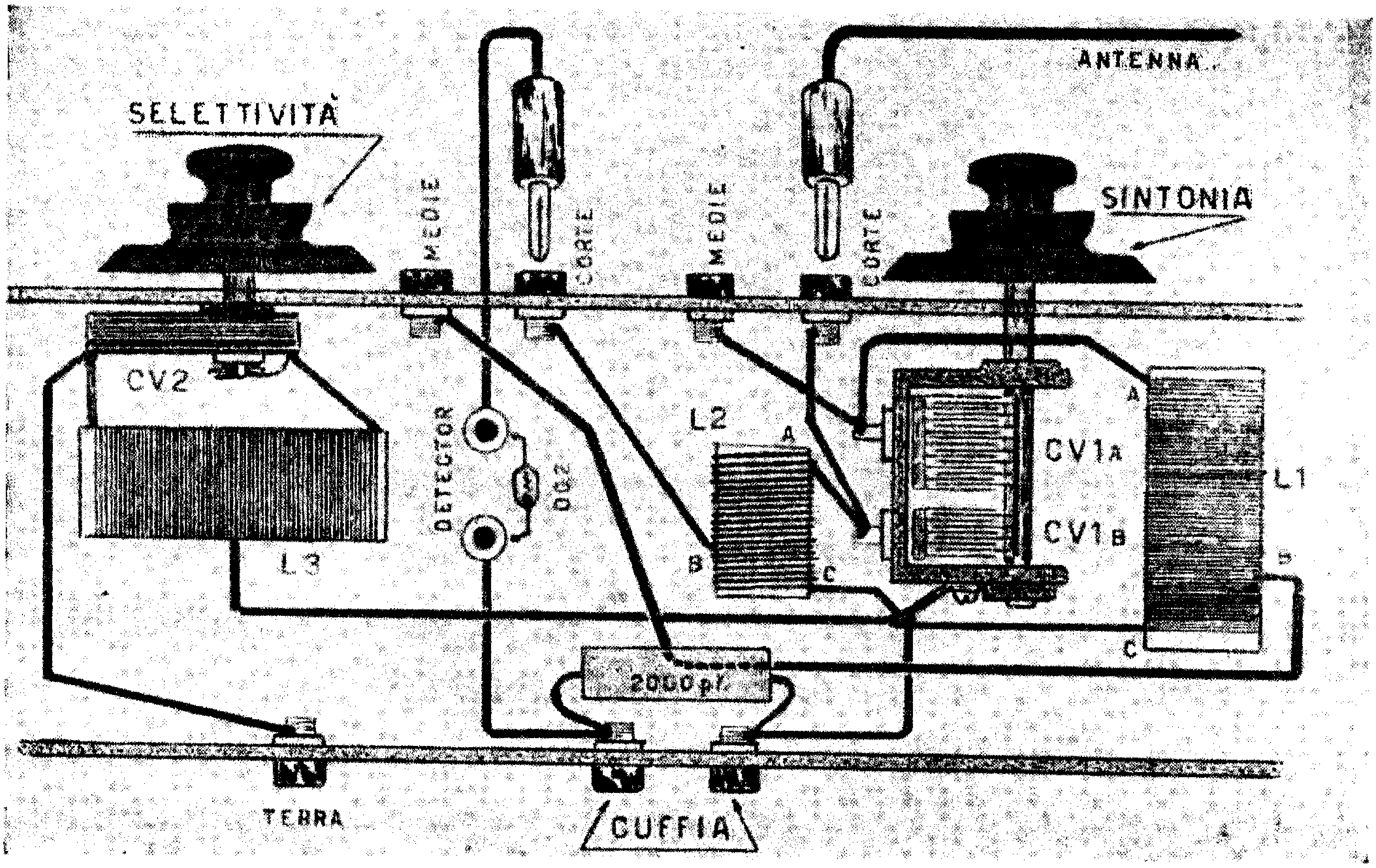

rica di questo nostro piccolo Fido. Ecco come noi andremo a realizzarlo praticamente. Su di un piccolo chassis delle dimensioni di centimetri 18 x 8 costruito in legno o faesite troverà posto tutto l'occorrente per la realizzazione. A destra monteremo il variabile doppio CV1A e CV1B a sinistra il variabile CV2. Nel centro del pannello, fisseremo pure le quattro boccole che ci serviranno per effettuare il cambio d'onda. A destra avremo le due boccole per l'antenna e a sinistra le due boccole per il detector. Uno sguardo al disegno pratico di montaggio chiarirà in modo più evidente la disposizione dei vari pezzi. Nel retro dello chassis fisseremo tre boccole, due serviranno per la cuffia e una per la presa di terra. Chi userà un detector a galena dovrà pure usare due boccole per quello; mentre chi userà il diodo di germanio potrà addirittura fissare internamente il diodo, sotto le due boccole dove andrebbe inserito il detector. Tutti i collegamenti andranno effettuati con fili di rame ricoperto di cotone. Tutte le giunte andranno saldate, con stagno e pasta salda; escludere nel modo più assoluto acido muriatico, che corroderebbe i fili. Per le bobine, si userà del tubetto bachelizzato (oppure di legno o di cartone) di diametro di 2 cm. Per L1 avvolgeremo 90 spire di filo da 0,18 mm. tipo smaltato. L'avvolgimento andrà effettuato a spire affiancate, la presa B per la bobina L1 andrà effettuata alla 30ª spira del lato C. Questa bobina servirà per le onde medie. Per le onde corte useremo una seconda bobina (L2). Le spire necessarie per costruire L2 sono 18 con presa B alla 6ª spira (p. dal lato di C) filo da 0,6 mm. coperto di cotone, su un tubetto dal diametro di 2 cm.

L' avvolgimento sarà effettuato a spire affiancate. Per L3 si userà sempre un tubetto del diametro di cm. 2. Su di esso si avvolgeranno 90 spire con presa centrale. Il filo sarà come per L1 di 0,18 mm. smaltato, spire affiancate.

Per i variabili si userà, come precedentemente detto, uno doppio ad aria da 465 pF e 100 pF (CV1A e CV1B) ed uno a mica o ad aria da 500 pf (CV2).

Per CV1, la sezione grande, cioè la parteA, verrà usata per le Onde Medie mentre la parte B, cioè quella con meno lamelle, per le Onde Corte. La massa metallica del variabile, cioè tutto il cestello che sostiene il variabile, servirà per saldarci il lato C delle due bobine (L1 e L2). Siccome molte volte questa presa, sul cestello metallico del variabile, non è sempre presente si potrà facilmente ottenerla avvitando in un qualsiasi foro sempre presente sul cestello, (carcassa) una vite, con una piccola linguetta di ottone per presa.

Il variabile CV1 sarà poi fissato con viti a ferro sullo chassis.

CV1 servirà per la sintonia, cioè per la ricerca delle stazioni trasmittenti, mentre CV2 servirà per la selettività, cioè per escludere una stazione disturbatrice.

Il diodo può essere inserito nel circuito da qualsiasi lato, poichè non c'è qui senso particolare. Appena montato tutto il circuito, se si è usato un diodo di germanio, il ricevitore funzionerà immediatamente, mentre se useremo un detector a galena, occorrerà cercare il punto sensibile del cristallo; e, a volte, per questa ricerca è necessario armarsi di pazienza. Per questo consigliamo tutti i giovani di abbandonare la galena e di usare in sua vece il diodo che non presenta, come la galena, i difetti sopra accennati. Occorre tener ben presente che non installando una buona antenna difficilmete si otterranno dei buoni risultati sulle Onde Corte. Così per le onde non in tutte le zone si potranno ascoltare programmi a tutte le ore. Così in certi luoghi ci sarà chi ascolterà bene le Onde Corte supponiamo alla mattina. Ci sarà chi, invece, la massima ricezione, sulle Onde Corte, la avrà la sera. Assicuriamo invece che le Onde Medie saranno ascoltate sempre e ottimamente a tutte le ore e con una ottima selettività.

PREZZO DEI COMPONENTI

Diodi di germanio .	L.	750
Variabili CV1 a due sezioni	»	850
Variabile CV2 a mica	»	250
7 boccole e 2 banane	»	150
Cuffia	»	1200
3 bobine	»	250

Fate conoscere ai vostri amici

SISTEMA PRATICO

Vi ringrazieranno e.. vi ringrazieremo pure noi.

BANCO DE CIRCUITOS

VOLUME 6

100 CIRCUITOS DE RÁDIOS E TRANSMISSORES

NCB

BANCO DE CIRCUITOS

Volume 6

100 CIRCUITOS DE RÁDIOS E TRANSMISSORES

Instituto NCB
www.newtoncbraga.com.br
leitor@newtoncbraga.com.br

BANCO DE CIRCUITOS – V.6 - 100 CIRCUITOS DE RÁDIOS E TRANSMISSORES

Autor: Newton C. Braga
São Paulo - Brasil - 2013

Palavras-chave: Eletrônica - Engenharia Eletrônica - Componentes – Reparação – Radio – Transmissores - RF

Copyright by
INTITUTO NEWTON C BRAGA.
1ª edição

Todos os direitos reservados. Proibida a reprodução total ou parcial, por qualquer meio ou processo, especialmente por sistemas gráficos, microfílmicos, fotográficos, reprográficos, fonográficos, videográficos, atualmente existentes ou que venham a ser inventados. Vedada a memorização e/ou a recuperação total ou parcial em qualquer parte da obra em qualquer programa juscibernético atualmente em uso ou que venha a ser desenvolvido ou implantado no futuro. Essas proibições aplicam-se também às características gráficas da obra e à sua editoração. A violação dos direitos autorais é punível como crime (art. 184 e parágrafos, do Código Penal, cf. Lei nº 6.895, de 17/12/80) com pena de prisão e multa, conjuntamente com busca e apreensão e indenização diversas (artigos 122, 123, 124, 126 da Lei nº 5.988, de 14/12/73, Lei dos Direitos Autorais).

Diretor responsável: Newton C. Braga
Diagramação e Coordenação: Renato Paiotti

Índice

Apresentação..8

Introdução..10

Receptores..12
1. Rádio Elementar..12
2. Rádio de Cristal..14
3. Receptor de Cristal de 1926...............................15
4. Radio de Galena de 1926 (2)..............................17
5. Rádio de Cristal (2)...19
6. Rádio de Cristal (3)...20
7. Micro Rádio AM...21
8. Rádio AM de 1 Transistor...................................22
9. Rádio de Cristal (3) 23
10. Rádio AM com FET..24
11. Rádio AM com FET (2).....................................25
12. Receptor AM Experimental...............................26
13. Micro Rádio AM...27
14. Rádio Solar..28
15. Rádio com Energia Grátis.................................29
16. Receptor AM com FET (3)................................30
17. Rádio AM Regenerativo com FET......................31
18. Rádio Alternativo com o CK722.........................32
19. Rádio AM de Dois Transistores..........................33
20. Receptor AM com Pilha Galvano Telúrica............34
21. Simples Rádio AM Experimental........................35
22. Rádio AM de 1 Transistor..................................36
23. Simples Receptor de VHF..................................37
24. Rádio AM de Dois Transistores..........................38
25. Rádio AM Simples de 2 Transistores...................39
26. Rádio AM Simples de 2 transistores (2)...............40
27. Rádio AM Simples de 2 transistores (3)...............41
28. Rádio AM simples com ZN414...........................42

29. Receptor ZN414..43
30.Receptor de Ondas médias (dois transistores).......44
31. Receptor Experimental de FM...........................45
32. Sintonizador AM..46
33. Receptor Regenerativo com FET........................47
34. Rádio AM de Dupla Sintonia 48
35. Rádio OM e OC com FET..................................49
36. Rádio AM Experimental...................................50
37. Rádio AM CMOS...51
38. Rádio AM Simples 52
39. Receptor com Diodo Tunnel..............................53
40. Receptor AM Simples.....................................54
41. Rádio AM com 2 FETs.....................................55
42. Rádio AM com 2 FETs (2).................................56
43. Rádio de Cristal...57
44. Rádio AM Mecânico..58
45. Receptor Regenerativo de Ondas Médias e Longas 59
46. Rádio ZN414...60
47. Rádio ZN414 (2)..61
48. Rádio Solar...62
49. Receptor AM Reflex..63
50. Rádio AM de Dois Transistores.........................64
51. Rádio AM de 2 Transistores (2).........................65
52. Receptor de FM de 2 Transistores......................66
53. Rádio AM de 2 Transistores (3).........................67
54. Rádio com o 741...68
55. Receptor AM Solar..69
56. Rádio AM de dois transistores (4)......................70

Transmissores **71**
57. Micro Transmissor de FM.................................72
58. Mini Transmissor AM.......................................73
59. Pequeno Transmissor de AM.............................74
60. Micro Transmissor de FM (2).............................75
61. Transmissor de Onda Contínua (CW)..................76

62. Microfone sem fio de FM....................................77
63. Transmissor de Ondas Médias............................78
64. Microfone sem fio AM..79
65. Microfone sem fio AM (2)...................................80
66. Microfone sem fio AM (3)...................................81
67. Micro Transmissor de AM 82
68. Transmissor de FM Alimentado por AC.................83
69. Micro Transmissor de FM (4)..............................84
70. Micro Transmissor de FM (5)..............................85
71. Transmissor de FM 86
72. Micro Transmissor de Ondas Curtas....................87
73. Microfone sem fio AM (4)..................................88
74. Microfone sem fio AM (5)...................................89
75. Microfone sem fio de FM (2)..............................90
76. Transmissor de FM com Integrado 91
77. Rádio AM de Três Transistores............................92
78. Micro Espião de FM...93
79. Transmissor gara 7 MHz.....................................94
80. Transmissor de tom de 30 a 100 MHz..................95
81. Transmissor sem bobinas...................................96
82. Transmissor de FM Com Diodo Tunnel..................97
83. Transmissor de FM com Diodo Tunnel (2).............98
84. Transmissor de Bips Sinalizador.........................99
85. Transmissor Sinalizador de 49 MHz...................100
86. Micro Transmissor Modulado de Ondas Curtas.....101
87. Micro Transmissor de FM Para 220 V.................102
88. Transmissor de Bips..103
89. Microfone sem fio AM (6).................................104

Circuitos Diversos..105
90. Verificador de Transmissores AM......................105
91. Detector de RF..106
92. Detector Ativo de RF..107
93. Localizador de Transmissores...........................108
94. Amplificador de OM...109
95. Reforçador AM..110

96. Reforçador AM (2)..111
97. Medidor de Intensidade de Campo Para 27 MHz. .112
98. Amplificador de Antena..................................113
99. Pré-Seletor de Ondas Médias............................114
100. Super Booster de OM.....................................115

PROBLEMAS CAUSADOS POR TRANSMISSORES DE FM............116

QUANDO UMA MONTAGEM NÃO FUNCIONA............................119

O RÁDIO..133

Apresentação

Durante nossa longa carreira como escritor de artigos e livros técnicos, por diversas vezes abordamos o tema "coletânea de circuitos", incluindo também informações. Assim, anteriormente, abordando este tema, publicamos as séries "Circuitos e Informações" (7 volumes) e "Circuitos e Soluções" (5 volumes) contendo centenas de circuitos úteis e informações técnicas de todos os tipos.

As séries se esgotaram, o tempo passou, mas os leitores ainda nos cobram algo semelhante atualizado e que possa ser usado ainda em projetos de todos os tipos. De fato, circuitos básicos usando componentes discretos comuns, de transistores a circuitos integrados, são ainda amplamente usados como solução simples para problemas imediatos, parte de projetos mais avançados e até com finalidade didática atendendo à solicitação de um professor que necessita de uma aplicação para uma teoria. Assim, voltamos agora com esta série, mas com uma estrutura diferenciada, novos projetos e nova abordagem. O diferencial na abordagem será dividir os diversos volumes da série por temas. Assim, no nosso primeiro volume tivemos circuitos de áudio, depois circuitos de fontes e seguindo a série neste sexto teremos circuitos de rádios, transmissores aplicações semelhantes como boosters e filtros.

Em nosso estoque de circuitos, coletados de todas as fontes possíveis, já temos mais de 5000 deles, muitos dos quais podendo ser acessados de forma dispersa no site. A vantagem de se ter estes circuitos organizados em volumes, além do acesso em qualquer parte, está na fácil localização de um circuito. As informações, por outro lado, serão agregadas aos circuitos, com links internos, o que só é possível numa publicação digital. A maioria destes circuitos, colhidos em publicações que, em alguns casos, pode não ser muito atuais, recebe um tratamento especial com comentários, sugestões e atualizações que viabilizam sua execução mesmo em nossos dias. Enfim, com esta série, damos aos leitores a oportunidade de ter em seus tablets, Iphones, Ipads, PCs, notebooks e outras mídias uma fonte de consulta de grande importância tanto para seu trabalho, como para seus estudos ou simples como hobby.

Newton C. Braga

Introdução

Depois do sucesso do Banco de Circuitos no meu site e das coleções esgotadas de Circuitos e Informações e Circuitos e Soluções, levo aos meus leitores uma coletânea de circuitos selecionada de minha enorme coleção disponível. Durante minha vida toda colecionei praticamente todas as revistas técnicas de eletrônica estrangeiras, dos Estados Unidos, França, Espanha, Itália, Alemanha, Argentina e até mesmo do Japão, possuindo assim um enorme acervo técnico.

Não posso reproduzir os artigos que descrevem os projetos que saem nessas revistas, por motivos ditados pela lei dos direitos autorais, mas a mesma lei permite que eu utilize uma figura do texto, com citação, comentando seu conteúdo para efeito de informação ou complementação de um conteúdo maior. É exatamente isto que faço na minha seção no site e também disponibilizo neste livro. Estou selecionando os principais circuitos destas publicações, verificando quais ainda podem ser montados em nossos dias, com a eventual indicação de componentes equivalentes, fazendo alterações que julgo necessárias e disponibilizando-os aos nossos leitores. Para o site já existem mais de 5 000 circuitos, no momento que escrevo este livro, mas a quantidade aumenta dia a dia.

Frequente o site, que ele poderá lhe ajudar a encontrar aquela configuração que você precisa para seu projeto. Os 100 circuitos selecionados para a primeira edição desta série e depois mais 100 para a segunda e mais 100 para cada uma seguinte, são apenas uma pequena amostra do que você vai encontrar no site. Para esta edição escolhemos 100 circuitos de rádios e transmissores experimentais com componentes comuns e de fácil obtenção na maioria dos casos.

Volumes Anteriores:
Volume 1 - 100 Circuitos de áudio
Volume 2 – 100 Circuitos de fontes
Volume 3 – 100 Circuitos osciladores
Volume 4 - 100 Circuitos de potência
Volume 5 - 100 Circuitos com LEDs e Displays

Banco de Circuitos – Volume 6
Rádios e Transmissores

Neste volume teremos uma coletânea de circuitos de receptores de rádio elementares para as faixas de ondas médias, curtas, FM e VHF usando componentes comuns, além de transmissores de baixa potência para as mesmas faixas dos receptores. Lembramos que a operação de transmissores com alcances maiores e potências mais altas que os indicados neste livro está sujeita a restrições legais. Assim, a finalidade dos circuitos apresentados é experimental e didática.

Receptores

O desempenho dos receptores descritos (sensibilidade e seletividade) tanto depende do grau de sofisticação dos circuitos, lembrando que a maioria é experimental, como da habilidade do próprio montador e das condições locais de recepção. As bobinas, em especial são componentes críticos, assim como a antena. Em muitos casos, a localização em ambientes com muitas interferências e ruídos também afeta o desempenho destes circuitos.

1. Rádio Elementar

Este circuito foi publicado numa Poptronics do ano 2000, mas pode ser montado com facilidade ainda hoje. A bobina consta de 20 + 80 espiras de fio 28 ou próximo, num bastão de ferrite de 1 cm de diâmetro por 20 ou mais centímetros de comprimento. A posição da tomada pode ser alterada para se modificar a seletividade e a sensibilidade. O fone deve ser de alta impedância ou piezoelétrico com um resistor de 100k em paralelo. Os capacitores de sintonia podem ser obtidos de rádios AM fora de uso. A antena consiste num fio estendido de pelo menos 5 metros de comprimento. Veja que este circuito tem dupla sintonia e exige uma antena bem longa para funcionar bem. A ligação à terra, que pode ser feita num objeto grande de metal, é importante para o melhor desempenho.

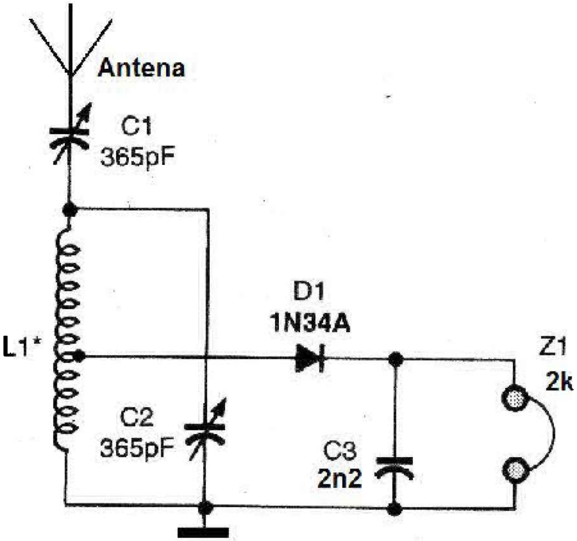
Antena
C1
365pF
L1*
D1
1N34A
Z1
2k
C2
365pF
C3
2n2

2. *Rádio de Cristal*

Este circuito usa um diodo de germânio como detector e o fone pode ser magnético de alta impedância ou ainda de cristal com um resistor de 100k em paralelo. L1 consta de 100 espiras de fio 28 num tubo de PVC ou papelão de 2,5 cm de diâmetro e CV é um capacitor variável de rádio AM comum. A antena deve ser longa, com pelo menos 5 m de comprimento, e a conexão à terra é importante. O circuito é do livro Fun Projects for the Experimenter de Newton C. Braga, publicado nos Estados Unidos.

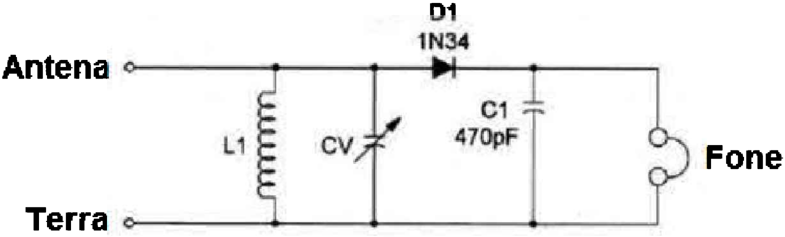

3. Receptor de Cristal de 1926

Este circuito foi encontrado numa publicação alemã de 1926. Trata-se de um receptor de cristal ou galena (D). O cristal (D) pode ser substituído por um diodo de germânio. O fone deve ser magnético, ou se for de cristal deve ter um resistor de 22 k em paralelo. L1 é formada por 10 espiras sobre L2, e L2 por 65 espiras de fio de 0,5 mm (pode ser o 28) num tubo de papelão de 8 cm de comprimento e 1 polegada de diâmetro. C2 era especificado na época em centímetros (!), mas um capacitor de 1 nF cerâmico serve. Da mesma forma, C1 pode ser um capacitor variável de rádio transistorizado antigo. A antena deve ser longa. Na figura a página em alemão da publicação original.

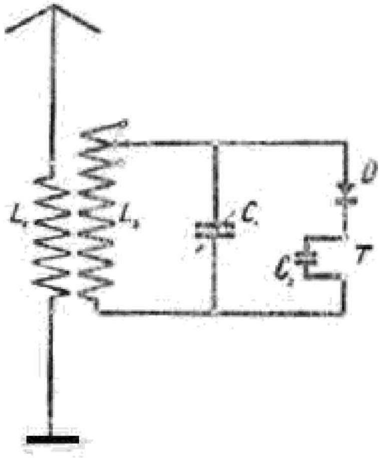

Kristalldetektor-Empfänger mit aperiodischer Antennenkopplung.

- L 1 10–12 Windungen 0,7 mm umsp. Draht über oder neben L 2 gewickelt
- L 2 65 Windungen 0,5 mm umsp. Draht auf Spulenkörper von 8 cm Durchmesser. Anzapfungen bei der 25., 35., 45., 55. Windung.
- C 1 Drehkondensator 500—1000 cm
- D Kristalldetektor
- C 2 Blockkondensator 2000 cm
- T Telefonsteckbuchsen

Durch die aperiodische Antennenkopplung wird eine große Abstimmschärfe erzielt. Diese Kopplungsart ist besonders bei langer Antenne vorteilhaft. Bei kurzer Antenne ist ein Empfänger nach Schaltbild 5 vorzuziehen.

Die Grobabstimmung erfolgt durch den Stufenschalter, die Feinabstimmung mittels des Drehkondensators.

Will man die Antenne abstimmen, so schalte man eine Schiebespule von etwa 40 Windungen auf Spulenkörper von 6 cm Durchmesser zwischen Antenne und Empfänger.

4. Radio de Galena de 1926 (2)

Esta raridade é de uma publicação de 1926. O detector de galena (D) pode ser substituído numa versão moderna por um diodo 1N34 ou equivalente. A sintonia é em série e C1 pode ter 1 nF já que na época os capacitores de construção caseira eram especificados em centímetros. O capacitor variável pode ser obtido num rádio AM antigo. A bobina pode ser enrolada com 60-80 espiras de fio 28 num tubo de 2,5 cm de diâmetro (papelão ou PVC). Abaixo a página original da publicação em 1926.

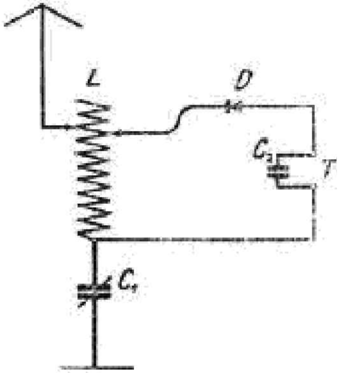

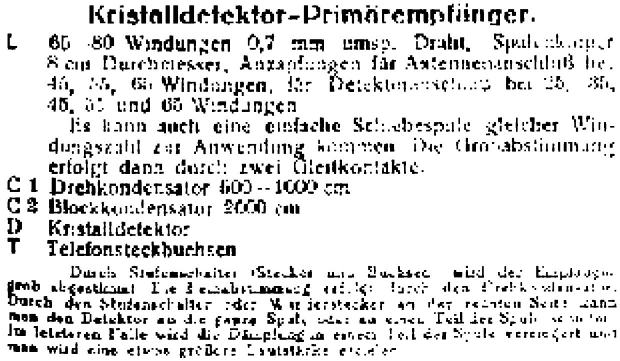

Kristalldetektor-Primärempfänger.

L 60–80 Windungen 0,7 mm umsp. Draht, Spulenkörper 8 cm Durchmesser, Anzapfungen für Antennenanschluß bei 45, 55, 65 Windungen, für Detektoranschluß bei 25, 35, 45, 55 und 65 Windungen.

Es kann auch eine einfache Schiebespule gleicher Windungszahl zur Anwendung kommen. Die Grobabstimmung erfolgt dann durch zwei Gleitkontakte.

C 1 Drehkondensator 600–1000 cm
C 2 Blockkondensator 2000 cm
D Kristalldetektor
T Telefonsteckbuchsen

Durch Stufenschalter (Stecker und Buchsen) wird der Empfänger grob abgestimmt. Die Feinabstimmung erfolgt durch den Drehkondensator. Durch den Stufenschalter oder Wanderstecker an der rechten Seite kann man den Detektor an die ganze Spule oder an einen Teil der Spule schalten. Im letzteren Falle wird die Dämpfung in einem Teil der Spule verringert und man wird eine etwas größere Lautstärke erzielen.

5. Rádio de Cristal (2)

Este circuito tradicional de rádio de cristal foi obtido numa documentação de 1990. O diodo pode ser o 1N34 ou qualquer outro de germânio que ainda possa ser encontrado. A bobina é formada por 100 espiras de fio 28 num bastão de ferrite e o capacitor variável pode ser obtido num rádio antigo de ondas médias. O fone é de cristal ou piezoelétrico e a antena deve ser bem longa, com pelo menos 5 metros de comprimento. Recomenda-se uma conexão à terra para maior desempenho.

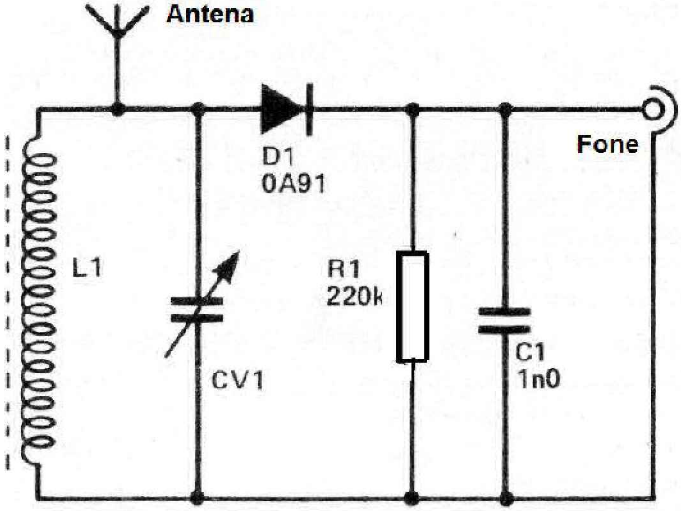

6. Rádio de Cristal (3)

A bobina consiste em 100 espiras de fio 28 num tubo de madeira ou PVC de 2 cm a 2,5 cm de diâmetro com tomada na 50ª espira. O diodo pode ser o 1N34 ou qualquer um de germânio. A antena deve ser longa, com pelo menos 10 metros de comprimento, C1 é um capacitor variável de rádio AM, C2 de 10 nF e C3 de 100 nF. O fone deve ser de cristal ou cerâmico. A ligação à terra é importante. R1 é de 10 k ohms.

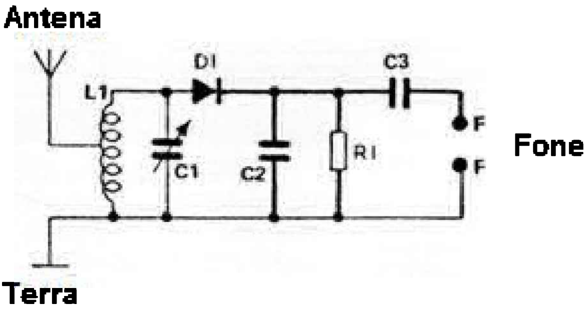

7. Micro Rádio AM

Este simples receptor capta as estações locais de AM com reprodução num pequeno alto-falante. A antena deve ter pelo menos uns 3 metros de comprimento. A bobina é formada por 30 + 50 espiras de fio 28 num bastão de ferrite. A ligação à terra é importante e o transformador deve ser obtido num rádio transistorizado antigo. Pode ser usado qualquer transistor de uso geral em lugar do original. O receptor é do livro Fun Projects for the Experimenter de Newton C. Braga, publicado nos Estados Unidos.

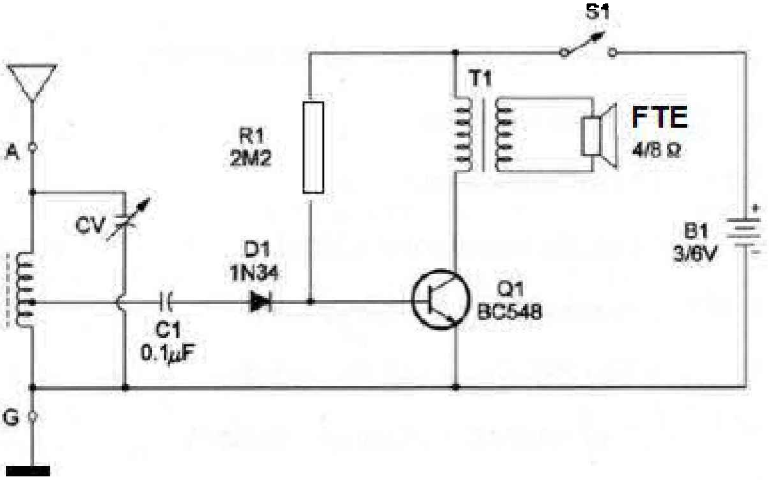

8. Rádio AM de 1 Transistor

Este circuito de rádio AM experimental foi encontrado numa revista inglesa de 1979. Ele pode ser montado com um BC548, alimentando um fone de cristal ou piezoelétrico. A antena deve ser longa e a ligação à terra é importante. A bobina consta de 10 + 20 espiras de fio 28 num bastão de ferrite para L1, e L2 consiste em mais 70 espiras do mesmo fio no mesmo bastão. O capacitor variável pode ser obtido de um rádio AM fora de uso. S1 seleciona a faixa de sintonia e S2 liga ou desliga o aparelho. SK1 é uma entrada opcional para a antena, conforme a faixa sintonizada. A alimentação é de 6 a 9 V e o consumo muito baixo.

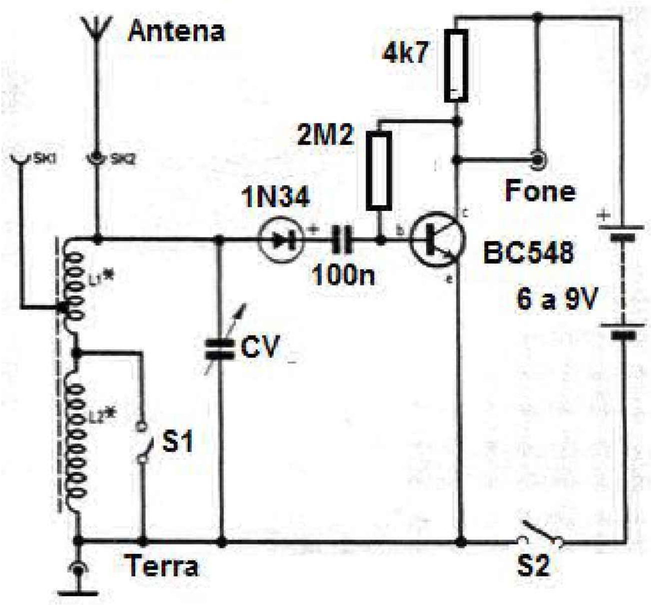

9. Rádio de Cristal (3)

Este receptor foi obtido numa revista de 1961. A revista não mais é publicada, mas o circuito pode ser montado utilizando-se qualquer diodo de germânio, fone de cristal ou de alta impedância e para a bobina, 100 espiras de fio 28 num bastão de ferrite. A antena deve ser longa e a ligação à terra é importante. C2 é um capacitor variável obtido de um rádio de ondas médias fora de uso e C3 é de 1nF, assim como C1, os quais devem ser cerâmicos. O diodo pode ser de qualquer tipo de germânio como o 1N34 ou 1N60. Para a conexão da antena e massa (terra) é usada uma ponte de dois terminais com parafusos.

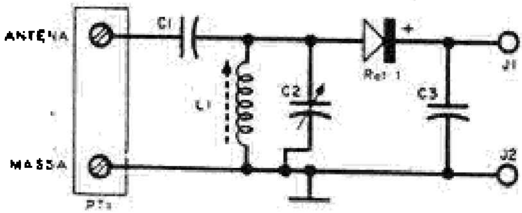

10. Rádio AM com FET

Este circuito foi publicado numa Poptronics de janeiro de 2000, mas pode ser montado com facilidade ainda hoje. A bobina consta de 20 espiras para L1 e 80 espiras de fio 28 ou próximo, para L2, as duas num bastão de ferrite de 1 cm de diâmetro por 20 ou mais centímetros de comprimento. O fone deve ser de alta impedância ou piezoelétrico com um resistor de 10k em paralelo. Os capacitores de sintonia podem ser obtidos de rádios AM fora de uso. A antena 1 consiste num fio estendido de pelo menos 5 metros de comprimento e a Antena 2, mais curta. O FET pode ser o BF245.

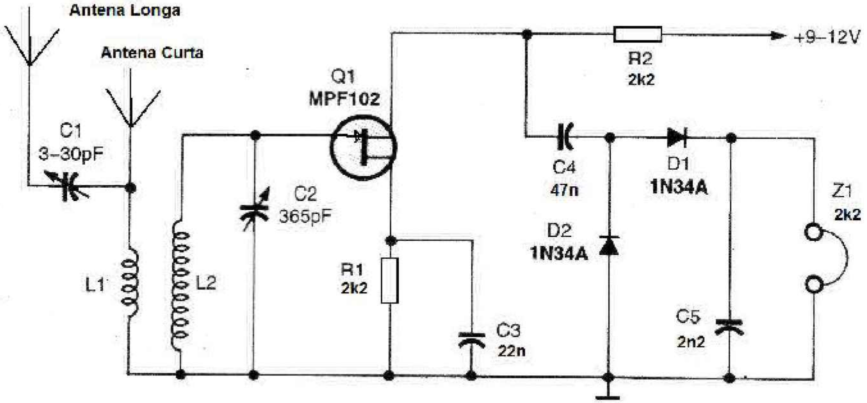
Antena Longa
Antena Curta
C1
3-30pF
L1
L2
C2
365pF
Q1
MPF102
R1
2k2
C3
22n
R2
2k2
+9-12V
C4
47n
D1
1N34A
D2
1N34A
C5
2n2
Z1
2k2

11. Rádio AM com FET (2)

Este circuito é de uma publicação americana de 1987 chamada Electronics Handbook. O circuito ainda pode ser montado com facilidade, pois todos os componentes são comuns. O transistor pode ser o BF245 ou MPF102 (FET). L2 é formada por 100 espiras de fio 28 num bastão de ferrite e L2, por 20 espiras do mesmo fio no mesmo bastão. C1 pode ser um capacitor de sintonia de um rádio AM comum. Os fones podem ser piezoelétricos ou magnéticos de 1k ou mais de impedância. O controle de volume é opcional, podendo o fone ser ligado diretamente ao capacitor C3.

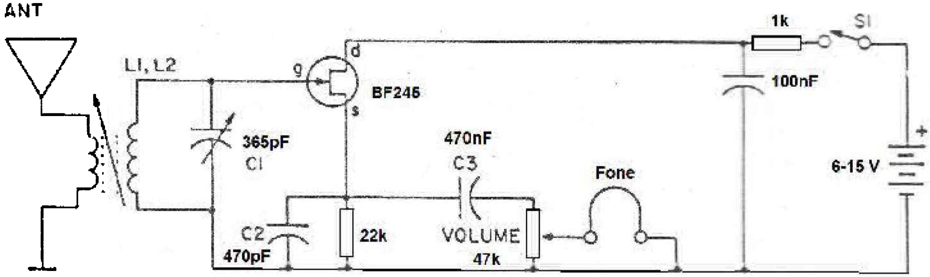
ANT
L1, L2
g
d
s
BF246
365pF
C1
C2
470pF
22k
470nF
C3
VOLUME
47k
Fone
1k
S1
100nF
+
6-15 V

12. Receptor AM Experimental

Este receptor foi obtido numa publicação de 1970, podendo ser montado com transistores mais modernos como o BC558, já que no original era utilizado um transistor de germânio. A bobina é formada por 15 + 20 + 70 espiras de fio 28 num bastão de ferrite e o capacitor C pode ser retirado de um rádio antigo. XRF é de 10 a 100 mH e os diodos são 1N34 ou equivalentes. O fone deve ter impedância entre 600 e 5000 ohms ou ser usado um fone de cristal com um resistor de 2k2 em paralelo. Os pontos da bobina são para a conexão alternativa da antena, conforme a estação captada. Observe a polaridade da alimentação que também pode ser de -6 V.

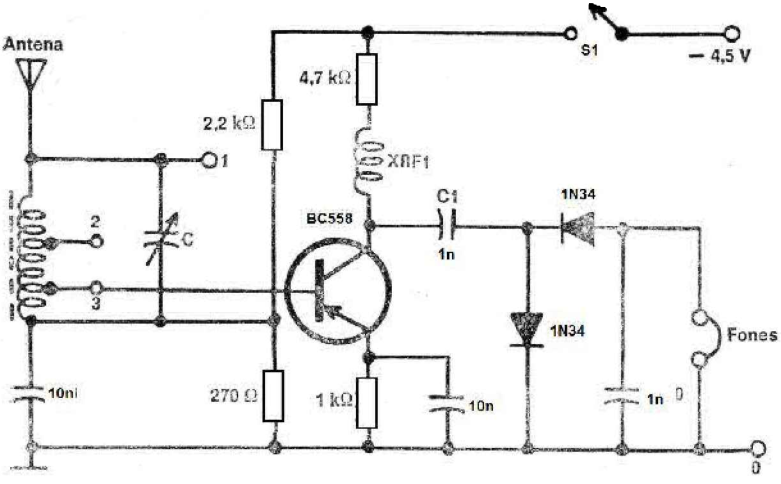
Antena
S1
– 4,5 V
4,7 kΩ
2,2 kΩ
1
XRF1
C1
1N34
2
C
BC558
1n
3
1N34
Fones
10n
270 Ω
1 kΩ
10n
1n
0

13. Micro Rádio AM

Este pequeno rádio experimental sintoniza as estações AM locais proporcionando escuta em fone de 1k a 2k. Pode ser usado fone de cristal ou um transdutor piezoelétrico, bastando que seja ligado um resistor de 2k2 a 4k7 em paralelo. A bobina consta de 20 + 80 espiras de fio 28 num bastão de ferrite e CV é capacitor variável obtido de um rádio AM fora de uso. A alimentação pode ser feita com duas ou quatro pilhas e a antena consiste num pedaço de fio de 3 a 10 metros de comprimento. A conexão à terra, num objeto de metal de grande porte, é importante para se obter maior sensibilidade. O circuito também funciona com transistores PNP, bastando inverter a polaridade de alimentação.

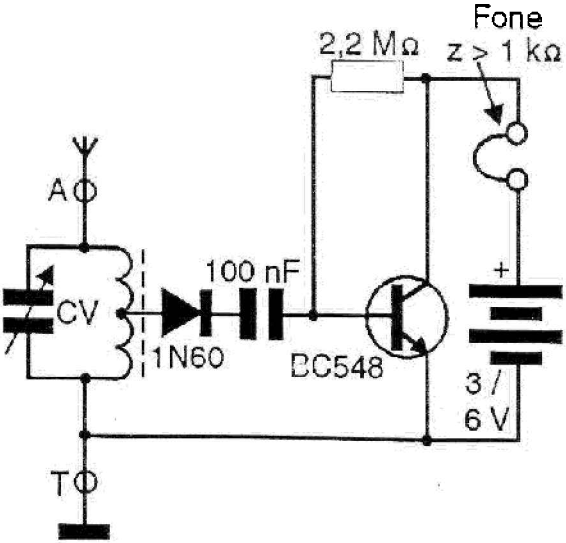

14. Rádio Solar

Este simples receptor AM alimentado por energia solar foi obtido numa Popular Electronics de 1974. O transistor pode ser o BC558, mas transistores de germânio obtidos de rádios antigos proporcionam melhores resultados com células de menor rendimento. A célula solar pode ser de qualquer tipo de 1,5 a 6 V, por exemplo, do tipo encontrado em calculadoras e outros aplicativos. A bobina consiste em 30 + 50 espiras de fio 28 num bastão de ferrite e o capacitor variável é do tipo encontrado em rádios AM antigos. A antena deve ser longa e o fone magnético de pelo menos 600 ohms. Para fones de cristal ou piezoelétricos ligue em paralelo um resistor de 2k2. A chave SPT comuta a alimentação para duas pilhas comuns. O consumo do circuito é muito baixo e a ligação à terra é importante para maior rendimento.

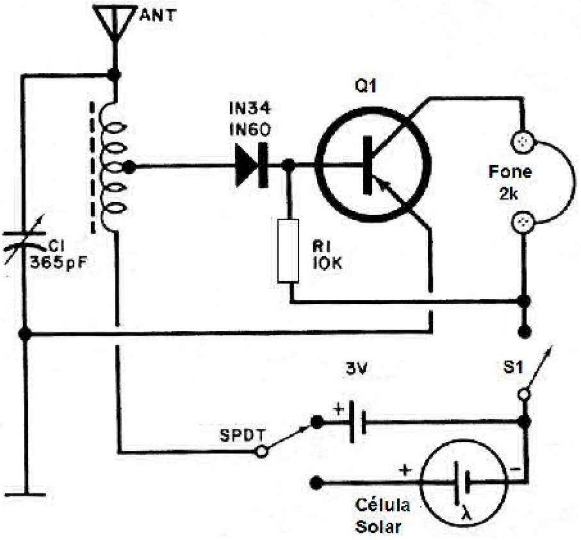

15. Rádio com Energia Grátis

O transistor deste rádio é alimentado com a própria energia do sinal, produzindo sua amplificação. Deve-se dar preferência a um transistor de germânio antigo NPN, pois é mais sensível. L1 consta de 80 espiras de fio 28 num bastão de ferrite, a antena deve ser bem longa e o capacitor variável pode ser aproveitado de um rádio fora de uso. A ligação à terra é importante para se ter o melhor desempenho. Dada a baixa tensão que se obtém na prática, os transistores de germânio proporcionam melhores resultados. Consiga um num velho rádio transistorizado fora de uso. Tipos como o 2SB75 são bastante interessantes. A antena deve ser bem longa, pois ela capta também a energia que o receptor precisa para funcionar. O circuito é do livro Fun Projects for the Experimenter de Newton C. Braga, publicado nos Estados Unidos.

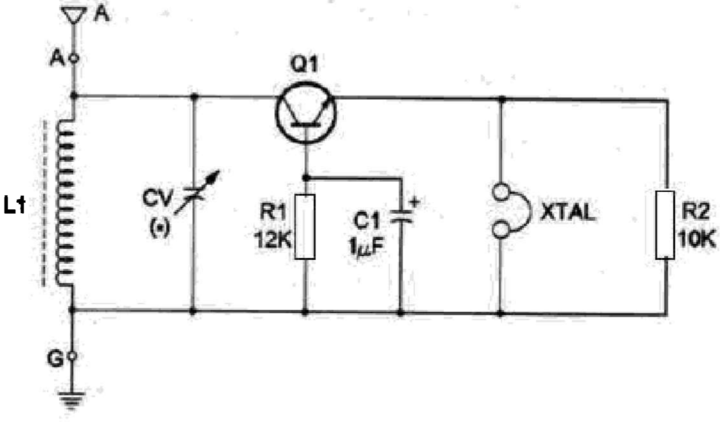

16. Receptor AM com FET (3)

Este circuito foi encontrado numa Elementary Electronics de 1979. Ele pode ser montado com facilidade utilizando-se um JFET BF245 ou equivalente. L1 é uma bobina formada por 100 espiras de fio 28 num bastão de ferrite e L2 um choque de RF. O fone deve ser do tipo magnético de alta impedância ou então piezoelétrico com um resistor de 22 k em paralelo. C7 controla o ganho de RF e C3 faz a sintonia. L2 pode ser um choque comercial de 2,2 a 4,7 mH, já que o valor 2,4 mH indicado no original não é fácil de obter. O circuito sintoniza a faixa de ondas médias, mas pode ser alterado para a faixa de ondas curtas com menos espiras na bobina.

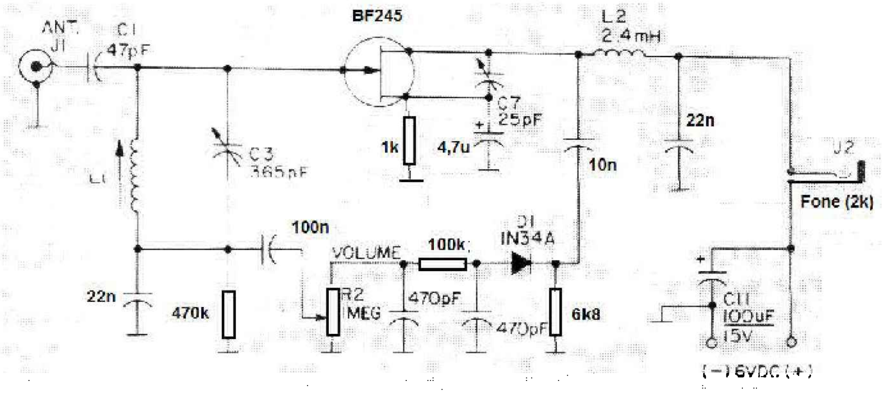
ANT.
J1
C1
47pF
BF245
L2
2.4mH
C7
25pF
1k
4,7u
10n
22n
J2
Fone (2k)
C3
365pF
L1
100n
VOLUME
100k
D1
IN34A
22n
470k
R2
MEG
470pF
470pF
6k8
C11
100uF
15V
(−) 6VDC (+)

17. Rádio AM Regenerativo com FET

Este circuito foi publicado numa Poptronics de janeiro de 2000, mas pode ser montado com facilidade ainda hoje. A bobina L2 consta de 20 + 80 espiras de fio 28 ou próximo, num bastão de ferrite de 1 cm de diâmetro por 20 ou mais centímetros de comprimento ou mesmo num tubo de 2 cm sem núcleo. L1 é formada por 20 espiras do mesmo fio sobre L1. O fone deve ser de alta impedância ou piezoelétrico com um resistor de 10k em paralelo. Os capacitores de sintonia podem ser obtidos de rádios AM fora de uso. A antena consiste num fio estendido de pelo menos 5 metros de comprimento. O FET pode ser o BF245. Se a sensibilidade for reduzida, inverta a ligação de L1.

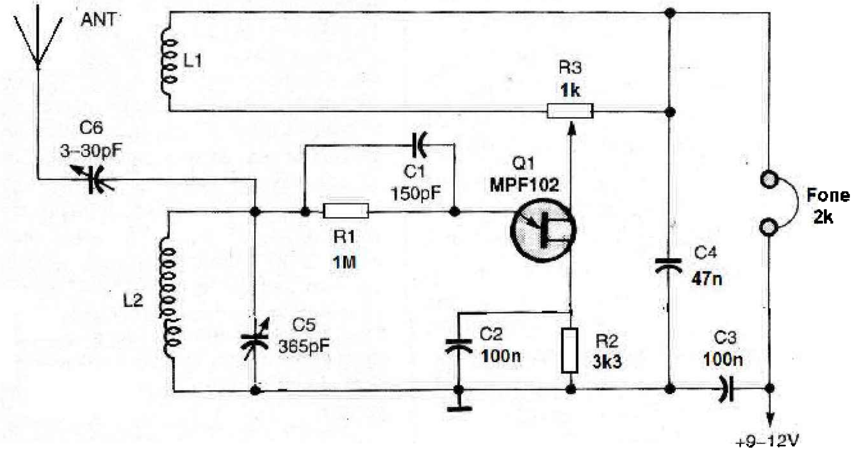
ANT
L1
R3
1k
C6
3-30pF
C1
150pF
Q1
MPF102
Fone
2k
R1
1M
C4
47n
L2
C5
365pF
C2
100n
R2
3k3
C3
100n
+9–12V

18. Rádio Alternativo com o CK722

Este circuito é de uma revista Radio Electronics de dezembro de 1954. Nele se mostra numa montagem experimental como o transistor, que então era novidade, pode ser usado num receptor capaz de funcionar com tensões a partir de 1,5 V, por exemplo, de uma pilha feita com moedas ou uma célula solar. O transformador é do tipo então encontrado em rádios transistorizados e o circuito deve funcionar com qualquer transistor PNP de germânio de uso geral. A bobina consta de 30 + 60 espiras de fio 28 num bastão de ferrite ou num tubo de papelão de 1 a 2 cm de diâmetro. O circuito capta as estações AM locais. O fone é de baixa impedância, podendo até ser usado um pequeno alto-falante se a antena for longa. No original foi usada uma bobina osciladora para a válvula 6SA7.

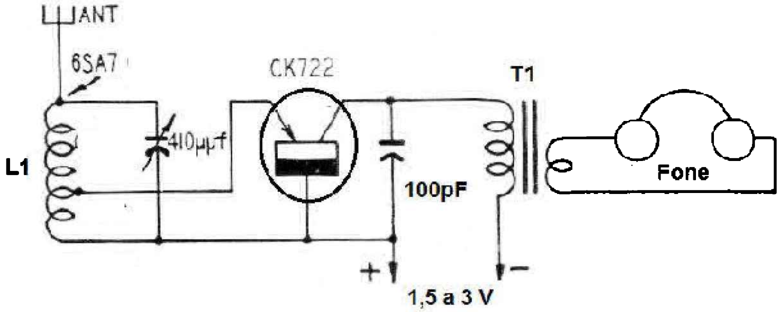

19. Rádio AM de Dois Transistores

Este simples rádio AM foi encontrado numa publicação de 1975, mas pode ser montado com transistores BC548. O fone também pode ser uma cápsula cerâmica piezoelétrica. O circuito pode ser alimentado com 3 ou mesmo 4 pilhas. L1 consiste em 80 espiras de fio 28 num bastão de ferrite e L2 é formada por 20 espiras do mesmo fio no mesmo núcleo. Uma antena pode ser ligada a L1. Uma antena pode ser ligada ao circuito em L1 para maior sensibilidade.

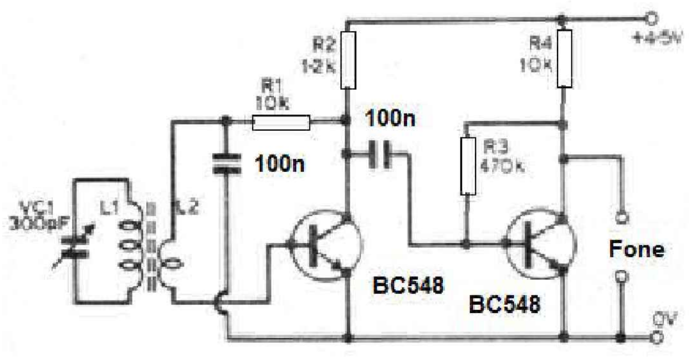
R2
1·2k
R1
10k
100n
100n
R4
10k
+4·5V
R3
470k
VC1
300pF
L1
L2
BC548
BC548
Fone
0V

20. Receptor AM com Pilha Galvano Telúrica

Este circuito, de uma publicação da década de 70, utiliza um tipo de alimentação diferente que já havíamos sugerido em artigo anterior na revista Eletrônica Popular. Duas placas de metais diferentes enterradas formam uma pilha telúrica capaz de gerar perto de 1 V para alimentar o circuito. A bobina é enrolada em ferrite com fio 28 e o transistor para melhores resultados deve ser de germânio. O capacitor variável pode ser aproveitado de um velho rádio AM e a antena deve ser longa. As placas devem ter de 10 x 10 a 30 x 30 cm de dimensões com a separação indicada na figura. O circuito capta estações da faixa de ondas médias e longas (em uso para poucas aplicações em nosso país). Pode ser usado um fone piezoelétrico com um resistor de 2k2 em paralelo.

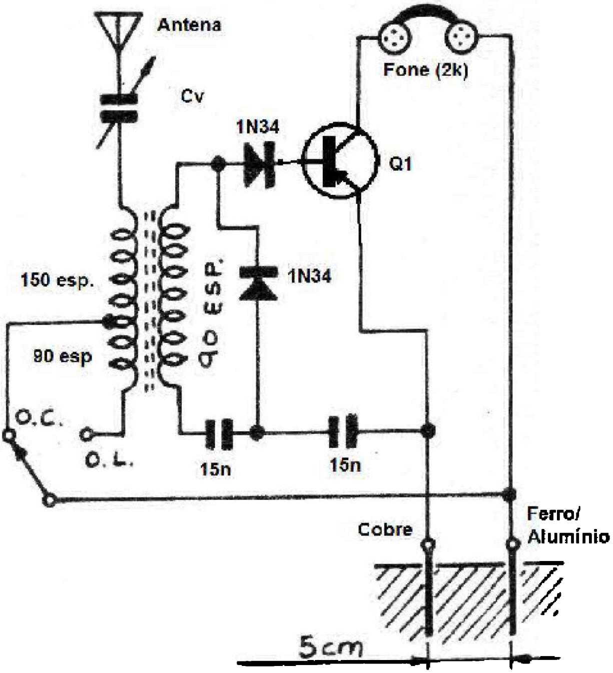

21. Simples Rádio AM Experimental

Este circuito capta as estações locais de ondas médias com escuta numa cápsula piezoelétrica ou fone de cristal. O transistor pode ser um PNP de germânio ou silício como o 2SB75, BC558, etc. CV é um capacitor variável retirado de um velho rádio de ondas médias transistorizado. L1 consta de 20 ou 30 espiras de fio 28 mais 80 espiras num bastão de ferrite. A antena é um pedaço de fio de 3 a 20 metros esticado e a ligação à terra é importante para se obter mais sensibilidade. O circuito é do livro Science Fair Projects de Newton C. Braga, publicado nos Estados Unidos.

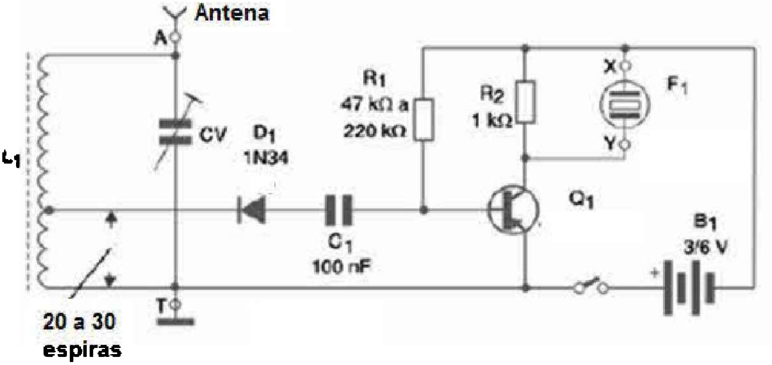
Antena
A
L1
CV
D1
1N34
R1
47 kΩ a
220 kΩ
R2
1 kΩ
X
Y
F1
Q1
C1
100 nF
B1
3/6 V
+
T
20 a 30
espiras

22. Rádio AM de 1 Transistor

Este rádio experimental para a faixa de ondas médias foi encontrado numa Poptronics de novembro de 2001. O transistor também pode ser um BC548 e o fone deve ser magnético de alta impedância, com pelo menos 2k ohms de impedância. Pode ser usado um fone piezoelétrico com um resistor de 4k7 em paralelo. A bobina consta de 30 + 70 espiras de fio 28 enroladas num bastão de ferrite e a antena deve ter pelo menos uns 5 metros de comprimento. Com um diodo de germânio o receptor será mais sensível. Uma ligação à terra feita no negativo da alimentação ajuda a obter mais sensibilidade. O resistor R3 pode ser alterado no sentido de se obter melhor rendimento.

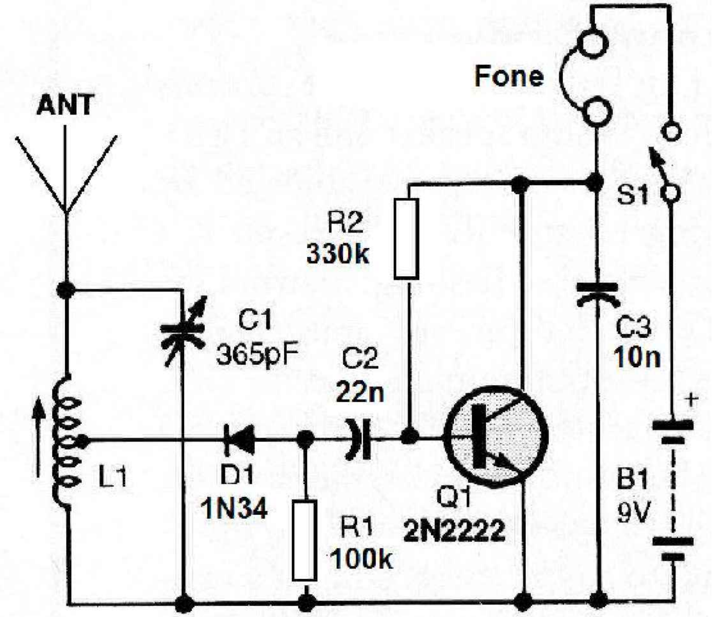

23. Simples Receptor de VHF

Se você mora perto de um aeroporto ou outro local em que opere um transmissor de VHF na faixa de 50 a 150 MHz, pode captar seus sinais com este simples receptor. Este circuito foi indicado em artigo da revista americana Electronics Hobbyst do inverno de 1971. A revista não existe mais, mas o projeto pode ser montado com a utilização de equivalentes para o diodo e transistores. Para o diodo pode ser usado o 1N34, para Q1 e Q2 BC558. Dada a baixa tensão de alimentação, o circuito funcionará melhor com transistores de germânio antigos, como os originalmente indicados. O fone pode ser magnético do tipo de 2k, muito difícil de obter em nossos dias ou uma cápsula piezoelétrica de telefone em paralelo com um resistor de 2k2. A alimentação vem de uma bateria de 9 V. Os sinais devem ser fortes o bastante para levar o transistor à condução. A antena é telescópica de pelo menos 2 metros de comprimento. A bobina é formada por 8 espiras de fio comum sem núcleo e diâmetro de 1 cm, com tomadas de 2 em 2 espiras. Com uma antena maior e uma conexão à terra, o circuito terá melhor rendimento.

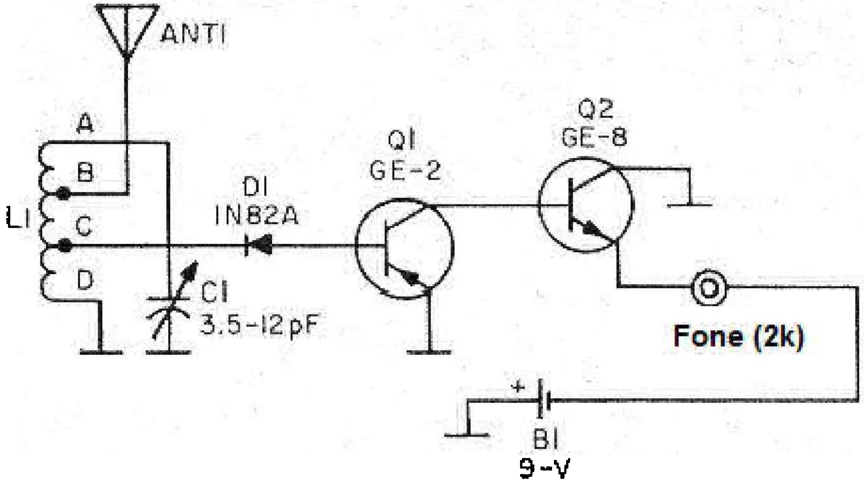

24. Rádio AM de Dois Transistores

Encontramos este circuito numa revista Hobby Electronics de fevereiro de 1981. No entanto, pelos componentes que utiliza, o circuito é perfeitamente viável ainda hoje. Os transistores podem ser os BC549 e o diodo 1N34. A bobina é enrolada num bastão de ferrite. O primário, ligado a VC1 consiste em 80 espiras de fio 28 AWG e o secundário 20 espiras. O núcleo é um bastão de ferrite de 12 a 20 cm de comprimento. O fone pode ser de cristal ou piezoelétrico de alta impedância. O consumo é muito baixo, o que significa que uma bateria de 9 V deve durar muito. O receptor não necessita de antena.

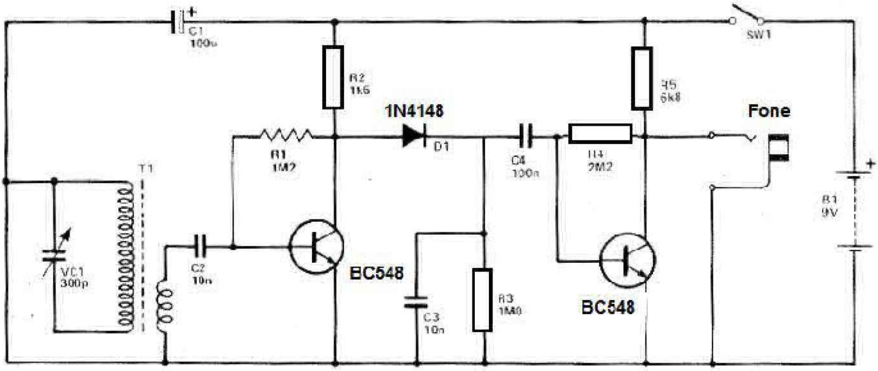
C1
100u
R2
1k5
1N4148
D1
R1
1M2
T1
VC1
300p
C2
10n
BC548
C3
10n
R3
1M0
C4
100n
R4
2M2
R5
6k8
SW1
Fone
B1
9V
+
BC548

25. Rádio AM Simples de 2 Transistores

Originalmente, este circuito é de uma revista inglesa, mas encontrei-o na Revista Radio Chassis Television da Argentina, numa edição de 1978. A bobina consta de 80 a 100 voltas de fio 28 num bastão de ferrite e na saída tanto podemos ligar um fone de cristal como a entrada de um amplificador. Para estações mais fracas use uma antena. A ligação do variável na bobina da forma indica resulta numa baixa seletividade. Veja o circuito 26 para maior seletividade. O equivalente ao BC108, usado originalmente é o BC548 já indicado no diagrama.

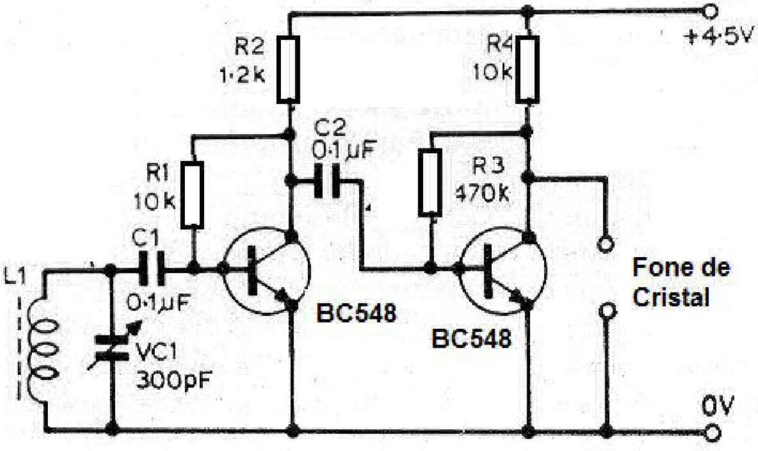
R2
1·2k
R4
10k
+4·5V
C2
0·1µF
R1
10k
R3
470k
C1
L1
0·1µF
BC548
BC548
Fone de Cristal
VC1
300pF
0V

26. Rádio AM Simples de 2 transistores (2)

Originalmente, este circuito é de uma revista inglesa, mas encontrei-o na Revista Radio Chassis Television da Argentina, numa edição de 1978. A bobina L1 é formada por 10 espiras de fio 28 sobre L2 e L2 consta de 80 a 100 voltas de fio 28 num bastão de ferrite. Na saída tanto podemos ligar um fone de cristal como a entrada de um amplificador. Para estações mais fracas use uma antena. A ligação do variável na bobina da forma indica resulta numa melhor seletividade em relação ao projeto 25. Use os BC548 em lugar dos originalmente recomendados que eram os BC108. A antena pode ser ligada à base do transistor.

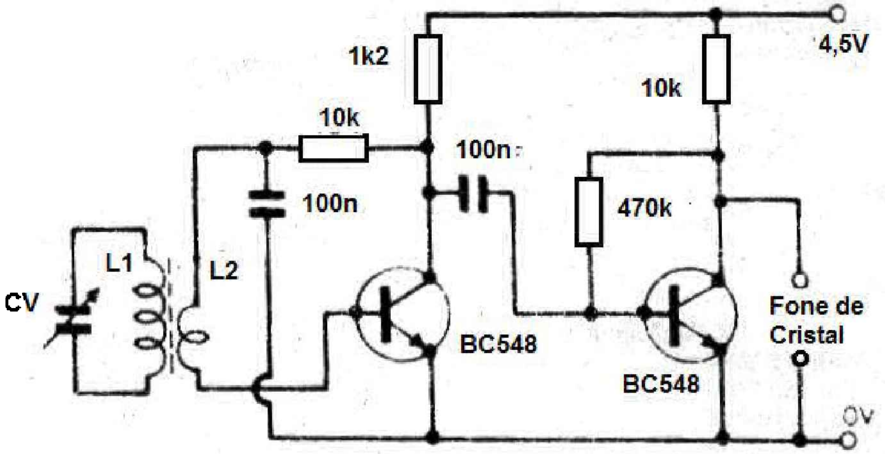

27. Rádio AM Simples de 2 transistores (3)

Originalmente, este circuito é de uma revista inglesa, mas encontrei-o na Revista Radio Chassis Television da Argentina, numa edição de 1978. L2 consta de 80 a 100 voltas de fio 28 num bastão de ferrite. Na saída tanto podemos ligar um fone de cristal como a entrada de um amplificador. Para estações mais fracas use uma antena. A antena pode ser ligada à base do primeiro transistor. O uso de três transistores num circuito com realimentação proporciona grande sensibilidade e seletividade. Use os BC548 em lugar dos BC108 que foram indicados na publicação original.

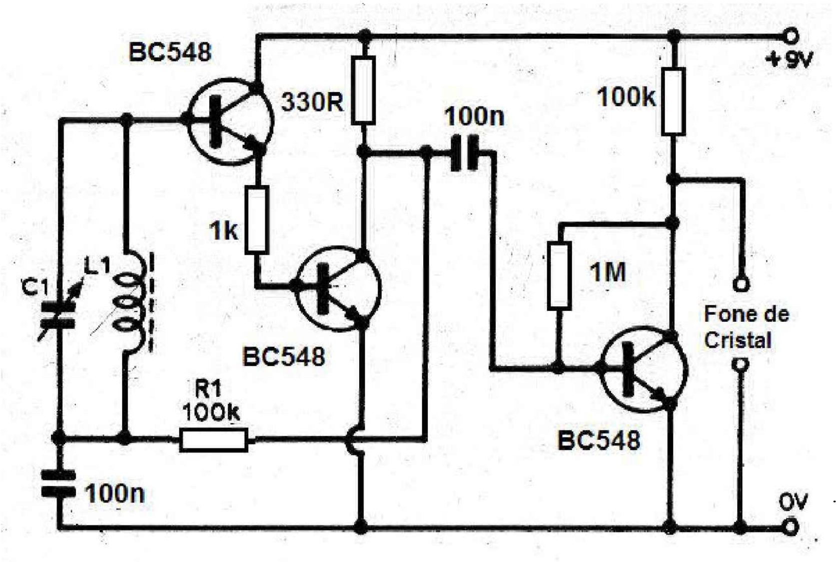
BC548
330R
100n
100k
+9V
1k
C1
L1
1M
Fone de Cristal
BC548
R1 100k
BC548
100n
0V

28. Rádio AM simples com ZN414

Originalmente, este circuito é de uma revista inglesa, mas encontrei-o na Revista Radio Chassis Television da Argentina, numa edição de 1978. L1 consta de 80 a 100 voltas de fio 28 num bastão de ferrite. Na saída tanto podemos ligar um fone de cristal como a entrada de um amplificador. Para estações mais fracas use uma antena. O integrado é o elemento critico do circuito, já que é difícil de obter atualmente. O capacitor de sintonia CV pode ser obtido em qualquer rádio de ondas médias fora de uso. Use o BC548 em lugar do transistor original que é o BC108, já trocado no nosso desenho. Observamos que o circuito integrado ZN414 é raro em nosso mercado, por não ser mais fabricado.

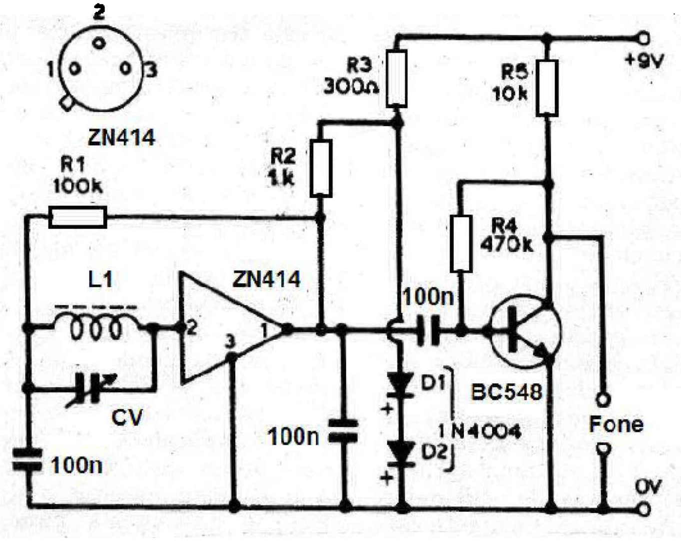
2
1
3
ZN414
R1
100k
L1
CV
100n
ZN414
2
3
1
R2
1k
R3
300Ω
100n
100n
D1
D2
1N4004
BC548
R4
470k
R5
10k
+9V
Fone
0V

29. Receptor ZN414

Este circuito foi obtido numa Hands On Electronics, revista americana de janeiro de 1987. O ZN414 não é muito fácil de obter, mas se o leitor conseguir pode montar um excelente radinho de ondas médias para escuta em fone. A alimentação é feita por uma pilha e a bobina consta de 80 a 100 voltas de fio de 32 a 28 num bastão de ferrite. O capacitor de sintonia pode ser obtido num radinho transistor fora de uso. Este circuito tem uma etapa amplificadora que pode ser elaborada com os transistores BC558 e BC548.

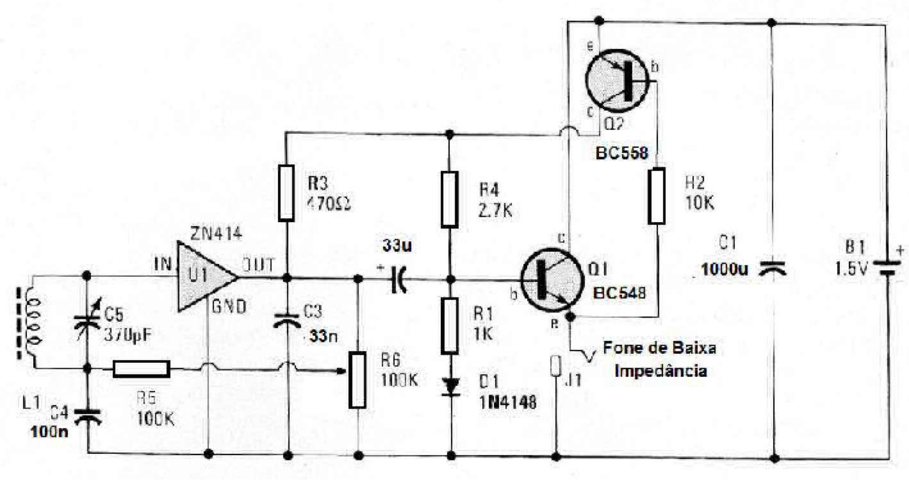
ZN414
IN
U1
OUT
GND
R3
470Ω
R4
2.7K
33u
+
C3
33n
C5
370pF
L1
C4
100n
R5
100K
R6
100K
R1
1K
D1
1N4148
b
c
e
Q1
BC548
Q2
BC558
R2
10K
C1
1000u
B1
1.5V
+
J1
Fone de Baixa Impedância

30.Receptor de Ondas médias (dois transistores)

Este simples receptor de ondas médias foi encontrado numa revista Técnica Prática da Itália do ano de 1965. O projeto é ainda atual se trocarmos os transistores pelos BC558, como já indicados na figura. O transformador de saída é obtido de um rádio transistorizado fora de uso. A bobina é formada por 75 espiras de fio 28 num tubo de 2 cm de diâmetro. As tomadas são feitas de 15 em 15 espiras. A antena deve ser longa e a ligação à terra é importante. CV pode ser obtido num rádio AM fora de uso. O circuito admite diversas melhorias. Uma delas consiste em se ligar de L1/C1 Para terra com um resistor de 47k em paralelo com um capacitor de 100 nF.

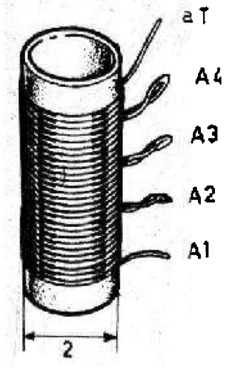

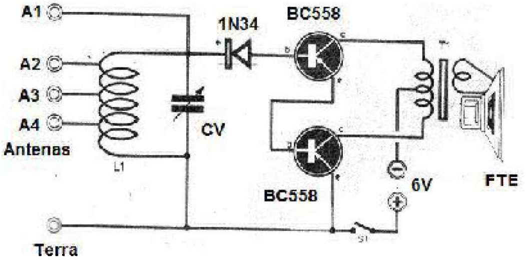

31. Receptor Experimental de FM

Encontrei este circuito numa revista de 1975, mas ele pode ser montado com facilidade utilizando-se um transistor BF494 ou BC548 em lugar do original. A saída deve ser ligada a um fone de cristal ou à entrada de um amplificador. A antena pode ser do tipo telescópica ou uma antena externa de TV. L1 tem 8 espiras de fio 28 numa forma de 1 cm e L2 de 2 a 4 espiras do mesmo fio ao lado de L1. Os choques L3 e L4 são de 1 mH a 2,2 mH. O circuito funciona com 9 V e a sintonia ajustada ao mesmo tempo em C1 e C2, que podem ser pequenos capacitores variáveis obtidos de rádios FM fora de uso.

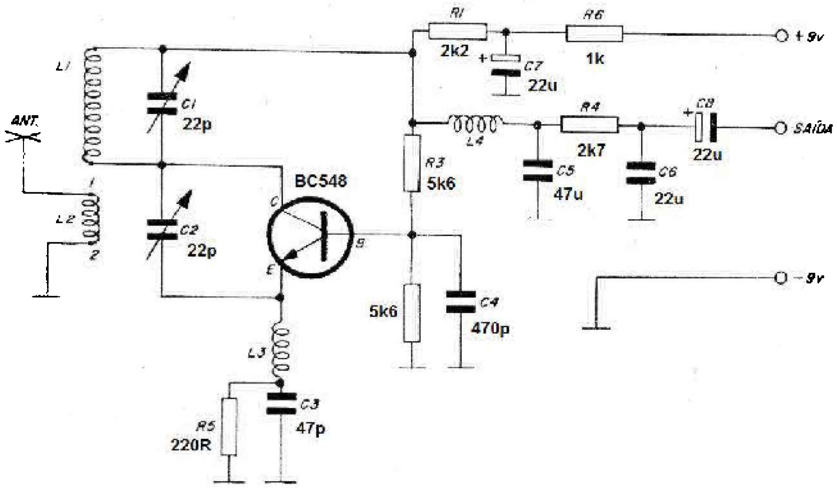
R1
2k2
R6
1k
+9v
+
C7
22u
L1
C1
22p
ANT.
L4
R4
2k7
+
C8
22u
SAÍDA
R3
5k6
C5
47u
C6
22u
BC548
C
B
E
1
L2
2
C2
22p
5k6
C4
470p
-9v
L3
R5
220R
C3
47p

32. Sintonizador AM

Encontrei este artigo de meu amigo A. Fanzeres numa revista uruguaia Corriente Alterna de julho de 1972. O artigo descreve um simples sintonizador de AM que pode ser ligado na entrada de qualquer amplificador de áudio para receber as estações AM locais. O artigo dá os detalhes da construção da bobina de antena. Mantivemos o formato original pelo valor histórico e como homenagem ao nosso amigo, já falecido, um grande nome da eletrônica em nosso país e no exterior.

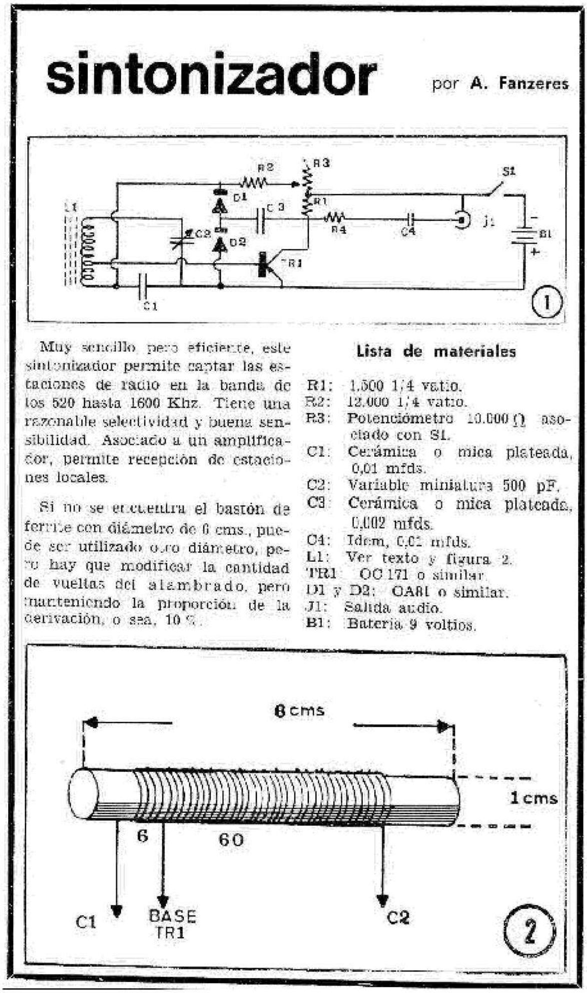

sintonizador

por **A. Fanzeres**

R2
R3
D1
C3
R1
S1
L1
C2
D2
R4
C4
J1
B1
+
TR1
C1
1

Muy sencillo pero eficiente, este sintonizador permite captar las estaciones de radio en la banda de los 520 hasta 1600 Khz. Tiene una razonable selectividad y buena sensibilidad. Asociado a un amplificador, permite recepción de estaciones locales.

Si no se encuentra el bastón de ferrite con diámetro de 6 cms., puede ser utilizado otro diámetro, pero hay que modificar la cantidad de vueltas del alambrado, pero manteniendo la proporción de la derivación, o sea, 10 %.

Lista de materiales

R1: 1.500 1/4 vatio.
R2: 12.000 1/4 vatio.
R3: Potenciómetro 10.000 Ω asociado con S1.
C1: Cerámica o mica plateada, 0,01 mfds.
C2: Variable miniatura 500 pF.
C3: Cerámica o mica plateada, 0,002 mfds.
C4: Idem, 0,02 mfds.
L1: Ver texto y figura 2.
TR1 OC 171 o similar.
D1 y D2: OA81 o similar.
J1: Salida audio.
B1: Bateria 9 voltios.

33. Receptor Regenerativo com FET

Este circuito foi obtido numa revista inglesa de 1974. O circuito pode ser montado com facilidade com qualquer JFET como o BF245 e TR2 pode ser um BC548. A saída é para fone de cristal ou piezoelétrico. L1 é formada por 100 espiras de fio 28 num bastão de ferrite de 12 a 20 cm de comprimento e L2 por 5 espiras do mesmo fio. Inverta a ligação de L2 se o circuito apresentar baixo rendimento. C1 pode ser retirado de um velho rádio transistor fora de uso.

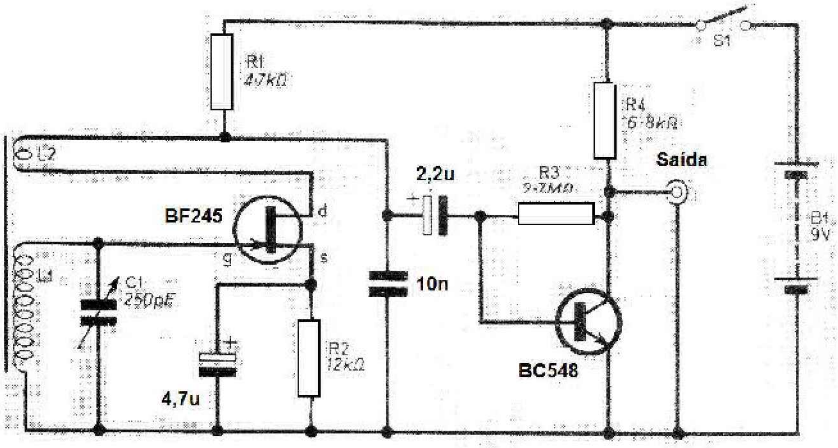
R1
47kΩ
L2
BF245
d
g
s
L1
C1
250pF
4,7u
R2
12kΩ
10n
2,2u
R3
2·2MΩ
R4
6·8kΩ
Saída
S1
B1
9V
BC548

34. Rádio AM de Dupla Sintonia

Encontrei este circuito numa publicação de 1964, mas ele pode ainda ser montado. Com transistores de germânio obtém-se maior sensibilidade, mas ele também funciona com os BC558 e BC548 (PNP e NPN). Os dois circuitos de sintonia permitem obter mais seletividade. As bobinas são formadas por 100 espiras de fio 28 num bastão de ferrite e os capacitores variáveis obtidos de rádios transistores fora de uso. O fone deve ter 1k ou mais ou então ser um de cristal com um resistor de 4k7 em paralelo. O diodo pode ser o 1N34 e a antena deve ter pelo menos 3 metros de comprimento. Uma conexão à terra dota o circuito de maior sensibilidade.

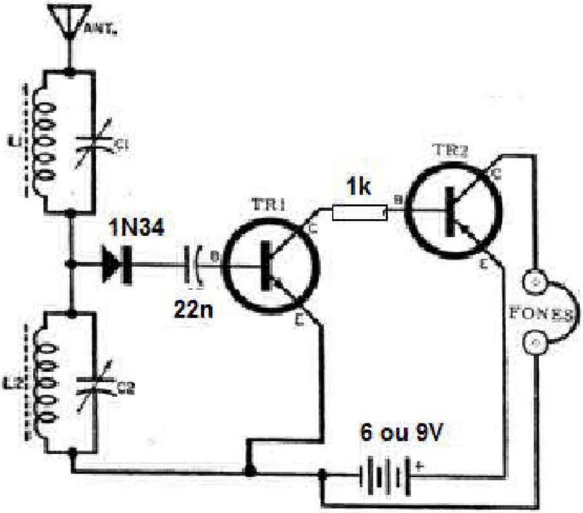

35. Rádio OM e OC com FET

Este receptor experimental foi encontrado numa publicação inglesa de 1979. Ele pode receber sinais de ondas médias até a faixa de ondas curtas até uns 30 MHz, dependendo da bobina. Essa bobina deve ter de 20 a 100 espiras de fio 28 num bastão de ferrite com tomadas obtidas em posições diversas, de modo a se obter maior rendimento. O fone deve ser de alta impedância ou então piezoelétrico com um resistor de 4k7 em paralelo. Os FETs podem ser os BF245 ou MPF102.

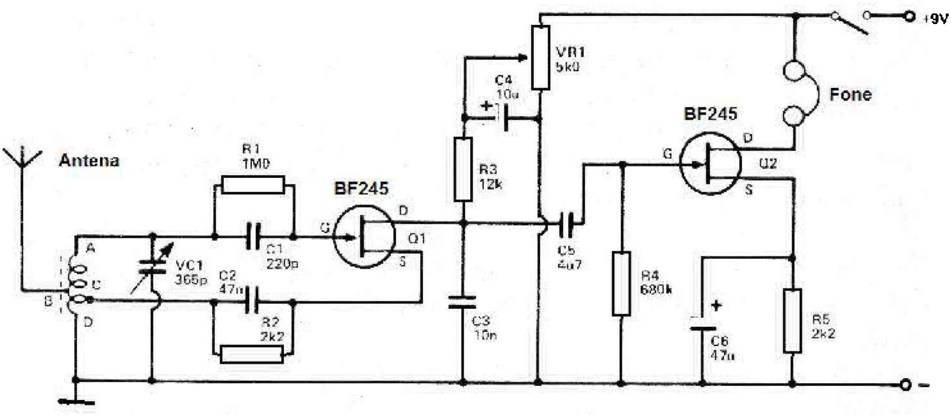
+9V
VR1
5k0
C4
10u
+
Fone
BF245
G
D
Q2
S
Antena
R1
1M0
BF245
D
G
Q1
S
R3
12k
A
VC1
365p
C1
220p
C2
47n
B
C
D
R2
2k2
C3
10n
C5
4u7
R4
680k
+
C6
47u
R5
2k2
-

36. Rádio AM Experimental

Este circuito foi obtido numa revista de 1964 e tem algumas características interessantes como a polarização de entrada do primeiro transistor feita pelo próprio sinal. Deve-se tomar cuidado para não inverter o diodo. O resistor deve ser obtido experimentalmente para maior rendimento do resistor na faixa de 1k a 10k. A bobina consta de 100 espiras de fio 28 num bastão de ferrite e os transistores podem ser os BC558. O transformador é do tipo de saída para rádios transistorizados. O circuito funciona com 6 a 8 V de alimentação. O capacitor de sintonia pode ser obtido num rádio transistorizado AM fora de uso. A antena deve ser longa e uma boa conexão à terra é importante para se obter maior sensibilidade, já que um bom sinal é necessário para se excitar o alto-falante.

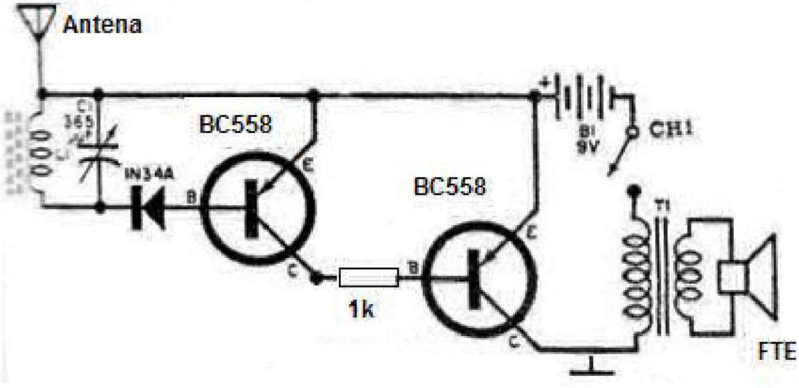

37. Rádio AM CMOS

Encontrei este circuito numa revista americana Hands On de 1986. Esta revista americana não mais existe, mas o circuito pode ainda ser montado com facilidade, pois seus componentes são comuns. O circuito fornece saída para fone ou para um amplificador e deve ser usada uma boa antena. A bobina L1 é formada por 80 a 100 voltas de fio AWG 28 num bastão de ferrite de 10 a 15 cm de comprimento e L2 é formada por 10 a 12 espiras do mesmo fio sobre L1. C2 é um capacitor variável obtido de um rádio AM comum. A alimentação pode ser feita com tensões de 9 a 12 V e o potenciômetro ajusta o ponto de funcionamento. C7 pode ser necessário para se obter melhor desempenho.

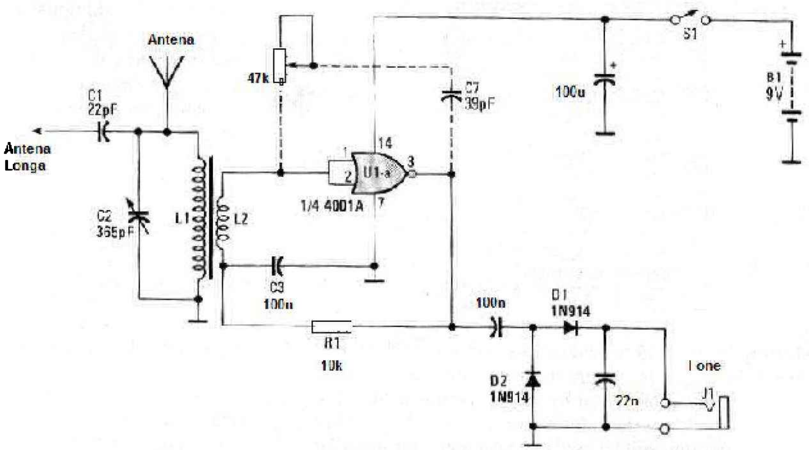
Antena
C1
22pF
Antena
Longa
C2
365pF
L1
L2
47k
1
2
U1-a
3
14
7
1/4 4001A
C7
39pF
C3
100n
100u
S1
+
B1
9V
R1
10k
100n
D1
1N914
D2
1N914
22n
Fone
J1

38. Rádio AM Simples

Este sensível Rádio AM não precisa de antena externa, podendo captar com facilidade as estações locais. A escuta é em fone de ouvido que tanto pode ser do tipo de cristal como um transdutor piezoelétrico do tipo encontrado em telefones. O circuito é alimentado por uma bateria de 9 V e tem um consumo muito baixo. A bobina L1 é formada por 100 espiras de fio esmaltado 26 ou 28 AWG num bastão de ferrite de 0,8 a 1 cm de diâmetro e de 12 a 20 cm de comprimento. Os bastões maiores proporcionarão maior sensibilidade.

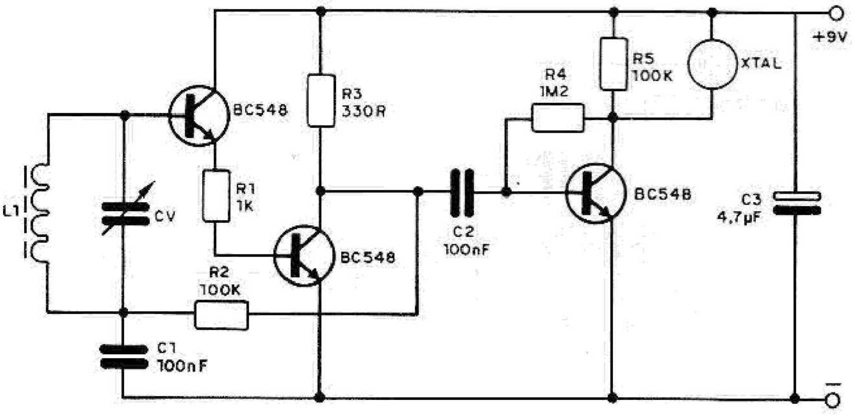

L1
CV
BC548
R1
1K
R2
100K
C1
100nF
R3
330R
BC548
C2
100nF
R4
1M2
R5
100K
XTAL
BC548
C3
4,7μF
+9V
−

39. Receptor com Diodo Tunnel

Este circuito é de uma publicação de 1964. O circuito aproveita as características de resistência negativa do componente para detectar sinais de rádio. A bobina consta de 100 espiras de fio 28 num tubo de PVC de 2 cm de diâmetro para operação na faixa de ondas médias. O potenciômetro ajusta o ponto de funcionamento do circuito. A saída pode ser aplicada a um fone ou a um amplificador de áudio. A alimentação (B1) consiste numa pilha de 1,3 V a 1,5 V. A antena deve ser longa e a conexão à terra é importante para se obter maior sensibilidade.

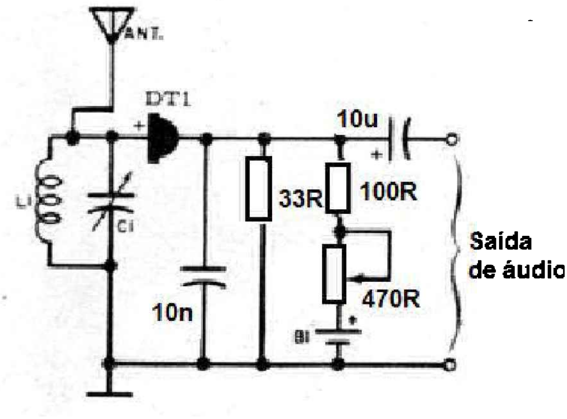

40. Receptor AM Simples

O circuito apresentado foi obtido numa Radio Elettronica italiana de 1976, mas pode ser montado com facilidade utilizando-se transistores BC548, troca que já fizemos ao redesenhar o esquema. A saída deve ser feita em fone magnético de alta impedância ou piezoelétrico com um resistor de 2k2 em paralelo. O circuito é alimentado por bateria de 9 V. L1 consta de 30 + 70 espiras de fio 28 num bastão de ferrite e L2, 20 espiras sobre L1. A antena deve ser longa e uma conexão à terra ajuda a obter melhor desempenho. Os capacitores variáveis podem ser obtidos em antigos rádios de ondas médias.

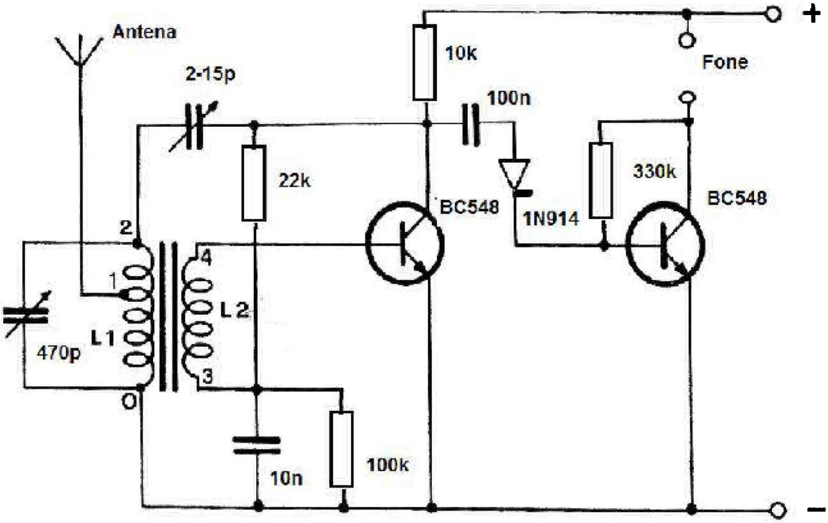
Antena
2-15p
10k
100n
Fone
+
22k
BC548
1N914
330k
BC548
2
1
4
L2
L1
470p
3
0
10n
100k
–

41. Rádio AM com 2 FETs

Este circuito foi publicado numa Poptronics de janeiro de 2000, mas pode ser montado com facilidade ainda hoje. A bobina L2 consta de 20 + 80 espiras de fio 28 ou próximo, num bastão de ferrite de 1 cm de diâmetro por 20 ou mais centímetros de comprimento ou mesmo num tubo de 2 cm sem núcleo. L1 é formada por 20 espiras do mesmo fio sobre L1. O fone deve ser de alta impedância ou piezoelétrico com um resistor de 10k em paralelo. Os capacitores de sintonia podem ser obtidos de rádios AM fora de uso. A antena 1 consiste num fio estendido de pelo menos 5 metros de comprimento e a Antena 2, mais curta. Os FETs podem ser os BF245. Uma conexão à terra é importante par se obter melhor recepção.

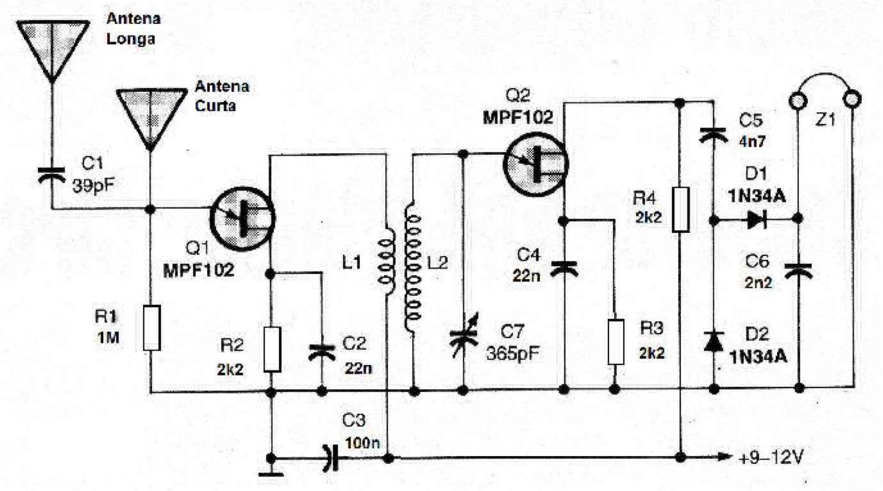
Antena Longa
Antena Curta
C1 39pF
Q1 MPF102
R1 1M
R2 2k2
C2 22n
L1
L2
C3 100n
Q2 MPF102
C4 22n
C7 365pF
R3 2k2
R4 2k2
C5 4n7
D1 1N34A
C6 2n2
D2 1N34A
Z1
+9–12V

42. Rádio AM com 2 FETs (2)

Este circuito foi publicado numa Poptronics de janeiro de 2000, mas pode ser montado com facilidade ainda hoje. A bobina L2 consta de 20 + 80 espiras de fio 28 ou próximo, num bastão de ferrite de 1 cm de diâmetro por 20 ou mais centímetros de comprimento ou mesmo num tubo de 2 cm sem núcleo. L1 é formada por 20 espiras do mesmo fio sobre L1. O fone deve ser de alta impedância ou piezoelétrico com um resistor de 10k em paralelo. Os capacitores de sintonia podem ser obtidos de rádios AM fora de uso. A antena 1 consiste num fio estendido de pelo menos 5 metros de comprimento e a Antena 2 é mais curta. O FET pode ser o BF245.

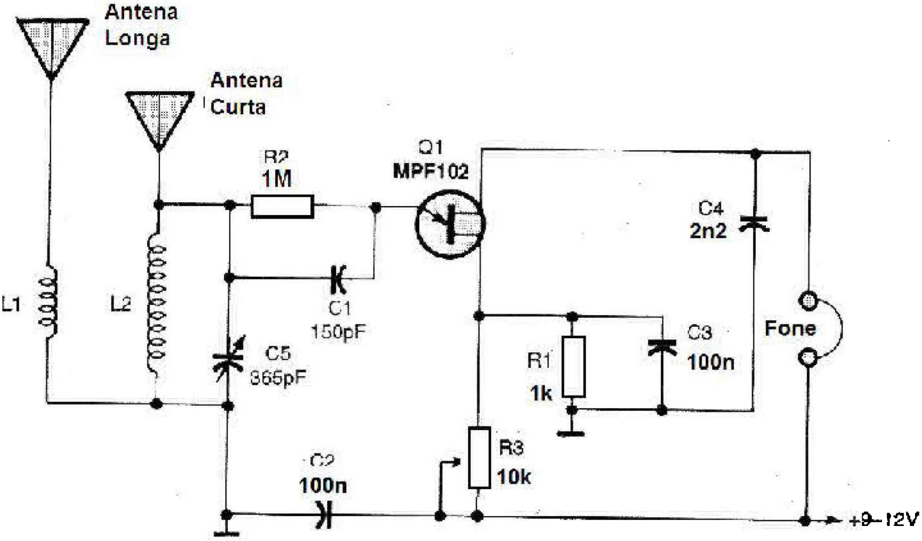
Antena Longa
Antena Curta
R2 1M
Q1 MPF102
C4 2n2
L1
L2
C1 150pF
C5 365pF
R1 1k
C3 100n
Fone
C2 100n
R3 10k
+9-12V

43. Rádio de Cristal

Este circuito saiu no Electronics Handobook Vol XIII de 1993, consistindo num rádio de galena ou cristal com diodo 1N34 e duas etapas de amplificação com FET. Uma delas é o amplificador de RF ou de antena com um transistor na configuração de gate comum de modo a baixar a impedância de entrada e a outra é o amplificador de áudio. Com isso o circuito pode ser alimentado a uma antena acoplada por cabo de 50 ohms. A saída da segunda etapa pode ser ligada a um amplificador de áudio externo. A bobina tem primário de 30 espiras de fio 28 ou 30 num cabo de vassoura ou tubo de PVC de 2,5 cm de diâmetro. Seu secundário, ligado ao variável consta de 100 espiras do mesmo fio. O primário deve ser enrolado sobre o secundário ou ao lado. O capacitor variável pode ser obtido num velho rádio de válvulas. O circuito tem muito baixo consumo, podendo ser alimentado por uma bateria de 9 V. O MPF102 pode ser substituído pelo BF245.

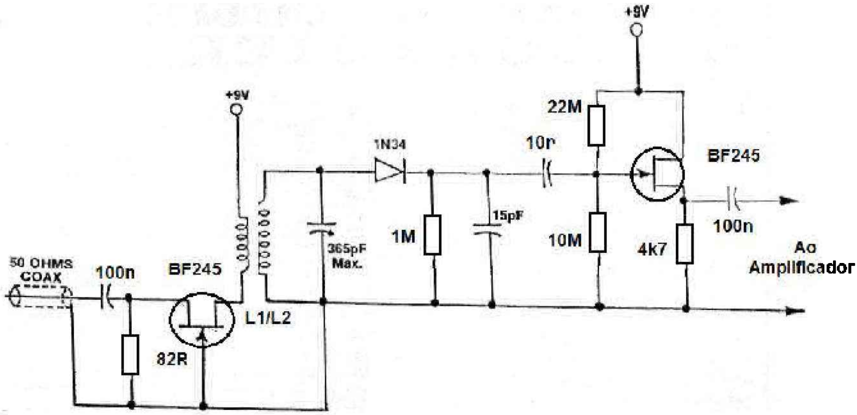

44. Rádio AM Mecânico

Este circuito foi encontrado numa revista inglesa de 1999, mas pode ser montado com facilidade com transistores BC548 ou qualquer NPN de uso geral. O circuito funciona com a energia gerada por um pequeno motor de corrente contínua que opera como dínamo. A energia carrega um capacitor e é usada para alimentar o rádio. L1 consta de 100 espiras de fio 28 num bastão de ferrite e CV é um capacitor variável obtido de um rádio AM fora de uso. Este projeto é interessante para demonstrar o funcionamento de fontes alternativas de energia, como o dínamo. O fone de cristal pode ser uma cápsula cerâmica piezoelétrica. O leitor deve fazer experiências com os motores que conseguir, pois o rendimento varia bastante de tipo para tipo.

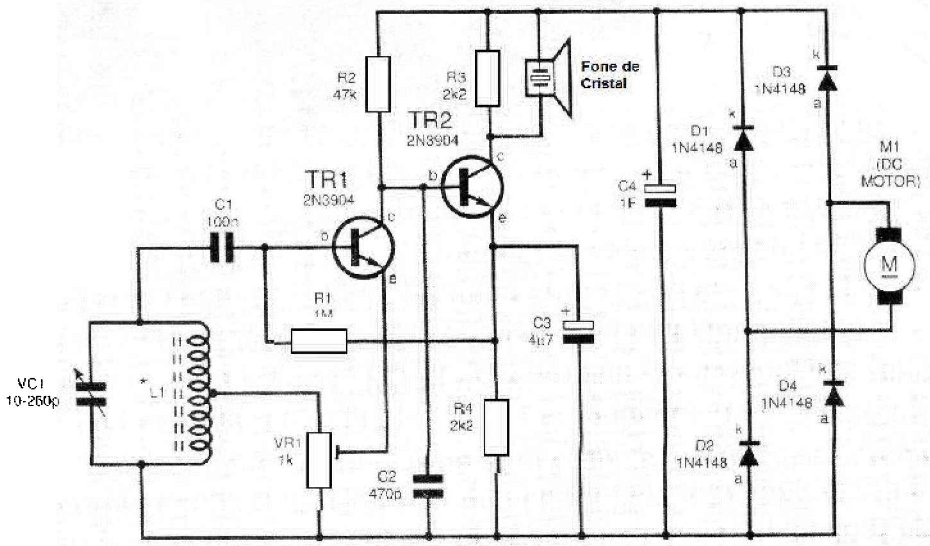
R2
47k
R3
2k2
Fone de Cristal
D3
1N4148
TR2
2N3904
D1
1N4148
M1
(DC MOTOR)
TR1
2N3904
C1
100n
C4
1F
C3
4u7
R1
1M
M
VC1
10-260p
L1
VR1
1k
R4
2k2
C2
470p
D4
1N4148
D2
1N4148
b
c
e
k
a

45. Receptor Regenerativo de Ondas Médias e Longas

Encontramos este circuito numa Popular Electronics de maio de 1990. Com dois transistores e um fone de ouvido de alta impedância é possível receber as estações locais de ondas médias. Q1 e Q2 podem ser os BC548. L2 consta de 40 + 40 espiras de fio 28 num tubo de plástico ou papelão de 2 a 2,5 cm de diâmetro e L1 20 espiras do mesmo fio no mesmo tubo, ao lado de L2. O variável C1 pode ser obtido num rádio transistorizado fora de uso.

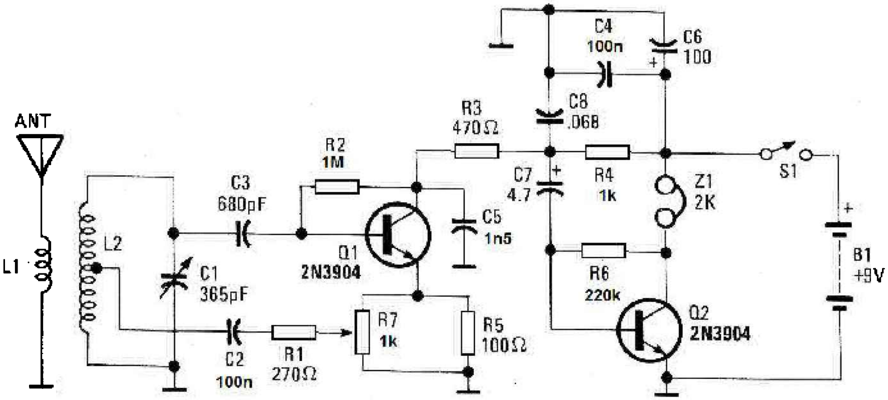
ANT
L1
L2
C1
365pF
C3
680pF
C2
100n
R1
270Ω
R2
1M
Q1
2N3904
R7
1k
R5
100Ω
C5
1n5
R3
470Ω
C8
.068
C4
100n
C6
100
C7
4.7
R4
1k
Z1
2K
R6
220k
Q2
2N3904
S1
B1
+9V

46. Rádio ZN414

Este circuito foi obtido numa Hands On Electronics, revista americana de janeiro de 1987. O ZN414 não é muito fácil de obter, mas se o leitor conseguir pode montar um excelente radinho de ondas médias para escuta em fone ou ainda um sintonizador aplicando a saída de áudio a um amplificador. A alimentação é feita por uma pilha e a bobina consta de 80 a 100 voltas de fio de 32 a 28 num bastão de ferrite. O capacitor de sintonia pode ser obtido num radinho transistor fora de uso. O circuito tem baixo consumo, podendo funcionar com fontes alternativas de energia como células solares e pilhas experimentais.

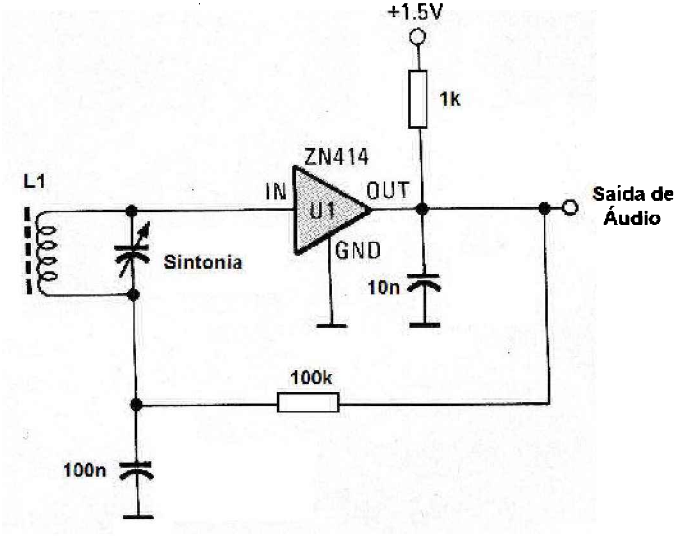

47. Rádio ZN414 (2)

Este circuito foi obtido numa Hands On Electronics, revista americana que não mais existe, edição de janeiro de 1987. O ZN414 não é muito fácil de obter, mas se o leitor conseguir pode montar um excelente radinho de ondas médias para escuta em fone. A alimentação é feita por uma pilha e a bobina consta de 80 a 100 voltas de fio de 32 a 28 num bastão de ferrite. O capacitor de sintonia pode ser obtido num radinho transistor fora de uso. Este circuito possui um controle de ganho e o resistor de 27 ohms nem sempre é necessário. O potenciômetro funciona como controle de ganho e, como o consumo é muito baixo, também pode ser usada na alimentação uma célula solar ou fonte alternativa.

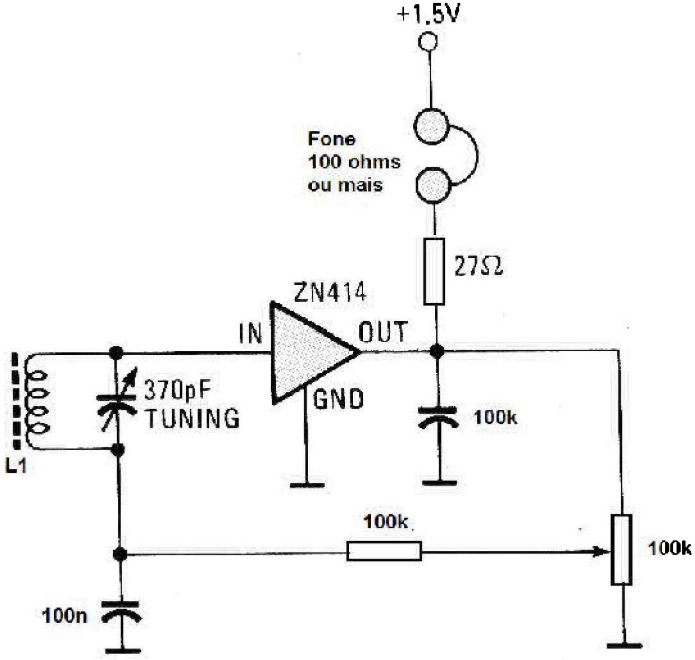

48. Rádio Solar

Encontramos este circuito numa revista inglesa de 1983, mas ele pode ser montado com facilidade com o transistor BC548 em lugar do original, troca que já fizemos, e se o circuito integrado for encontrado. Na verdade, o circuito integrado em questão, que consiste num receptor completo é difícil de se obter atualmente no nosso mercado. A bobina é formada por 80 a 100 espiras de fio 28 num bastão de ferrite. Pode ser usada uma célula solar de calculadora ou ainda outras fontes alternativas de energia.

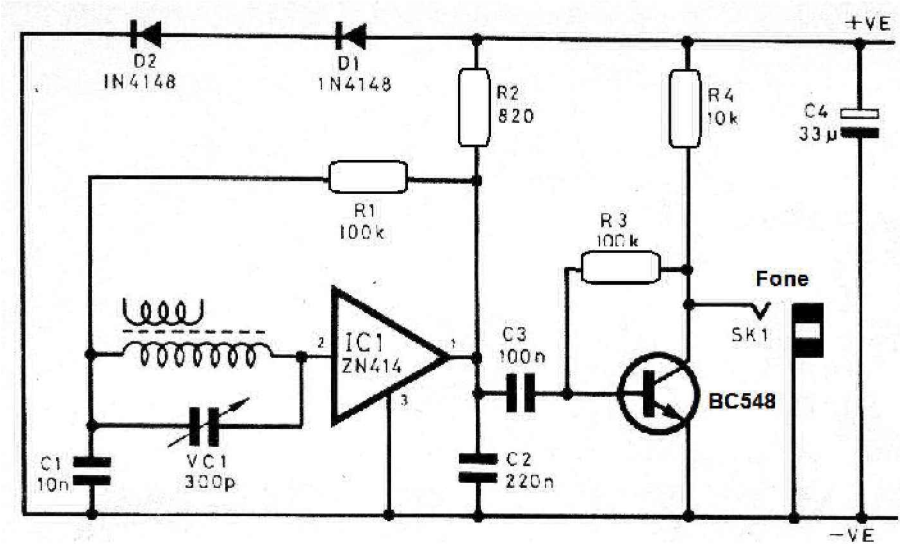
+VE
D2
1N4148
D1
1N4148
R2
820
R4
10k
C4
33µ
R1
100k
R3
100k
Fone
SK1
2
IC1
ZN414
1
3
C3
100n
BC548
C1
10n
VC1
300p
C2
220n
−VE

49. Receptor AM Reflex

Num receptor reflex o sinal de áudio detectado é aplicado de volta à entrada do circuito, de modo a receber nova amplificação. Neste receptor L1 é formada por 10 espiras de fio 28 num bastão de ferrite e L2 por 80 + 10 espiras do mesmo fio no mesmo bastão. O transformador de saída pode ser obtido de rádios transistorizados fora de uso e os transistores podem ser os BC548. O circuito foi obtido numa publicação internacional dos anos 80.

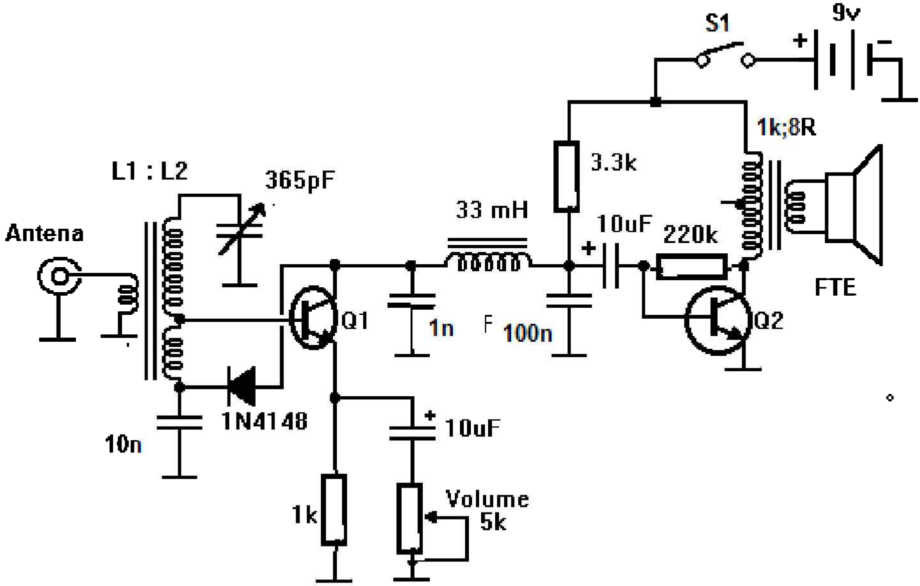

S1
9v
+
-
1k;8R
3.3k
L1 : L2
365pF
33 mH
Antena
10uF
220k
+
FTE
Q1
1n
F
100n
Q2
1N4148
10n
+
10uF
1k
Volume
5k

50. Rádio AM de Dois Transistores

Este circuito foi obtido na publicação 101 Electronics Projects de 1979. O circuito pode ser montado facilmente, pois todos os componentes são comuns. Este receptor capta as estações locais de ondas médias. O transformador pode ser obtido de um rádio AM antigo, assim como a bobina. Esta bobina também pode ser enrolada num bastão de ferrite sendo formada por 80 espiras de fio 28. O circuito também funciona com alimentação de 3 a 9 V. Os transistores podem ser o BC548 para Q1 e BC558 para Q2, já que os tipos indicados na publicação original não são fáceis de obter atualmente. A troca já foi feta ao redesenharmos o circuito. Uma boa antena externa e uma conexão à terra são importantes para se obter o melhor desempenho do circuito.

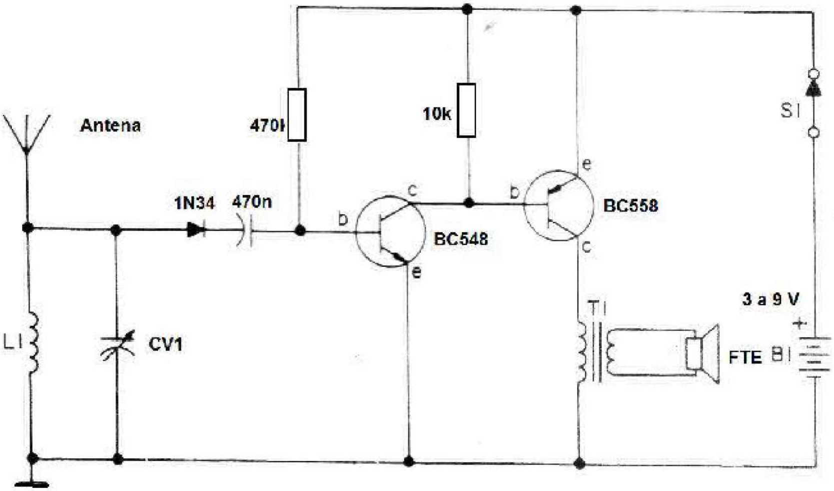

51. Rádio AM de 2 Transistores (2)

Este simples circuito de rádio AM de dois transistores foi obtido numa publicação argentina dos anos 1970. O circuito funciona com tensões de 3 a 12 V e a bobina consiste em 20 + 60 espiras de fio 28 num tubinho de papelão de 1 cm de diâmetro. O capacitor variável pode ser obtido de um rádio fora de uso e o fone deve ter de 600 ohms a 2k. Na sua falta use um transdutor piezoelétrico com um resistor de 2k2 em paralelo. Os transistores podem ser os BC548.

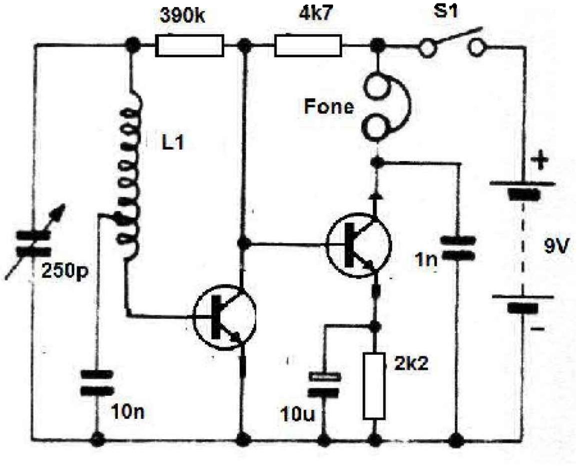

52.Receptor de FM de 2 Transistores

Este circuito foi obtido numa revista inglesa do ano 2000. Trata-se de um simples receptor de FM cujo sinal deve ser aplicado a um fone piezoelétrico ou à entrada de um amplificador. L consta de 4 espiras de fio 28 com diâmetro de 1 cm sem núcleo e C1 pode ser um capacitor variável retirado de rádio FM. A antena pode ser um pedaço de fio rígido de 40 cm a 1 metro. Podem ser usados os transistores BF494. Deve ser usada uma pequena antena telescópica. O sinal de áudio deve ser aplicado à entrada de um bom amplificador.

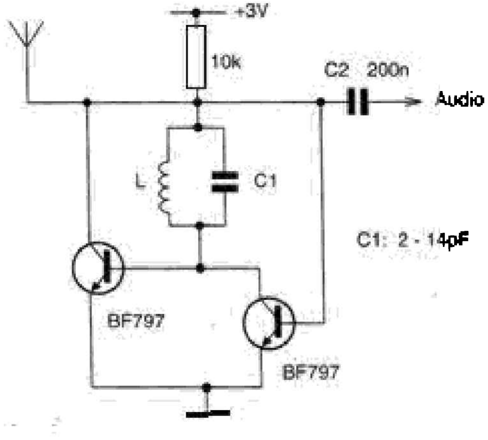

53.Rádio AM de 2 Transistores (3)

Este receptor foi encontrado numa revista Radio Elettronica de 1976. A revista italiana não mais existe , mas o receptor pode ser montado com transistores BC548 e 2N2222. Em lugar do 2N2222 também podemos usar o BD135. A bobina consta de 100 espiras de fio num bastão de ferrite com tomada na 30ª espira. A alimentação pode ser feita com bateria ou pilhas e tensões entre 6 e 12 V. Recomenda-se uma boa antena e a ligação à terra aumenta a sensibilidade. A tomada pode ser alterada para se modificar a sensibilidade e seletividade.

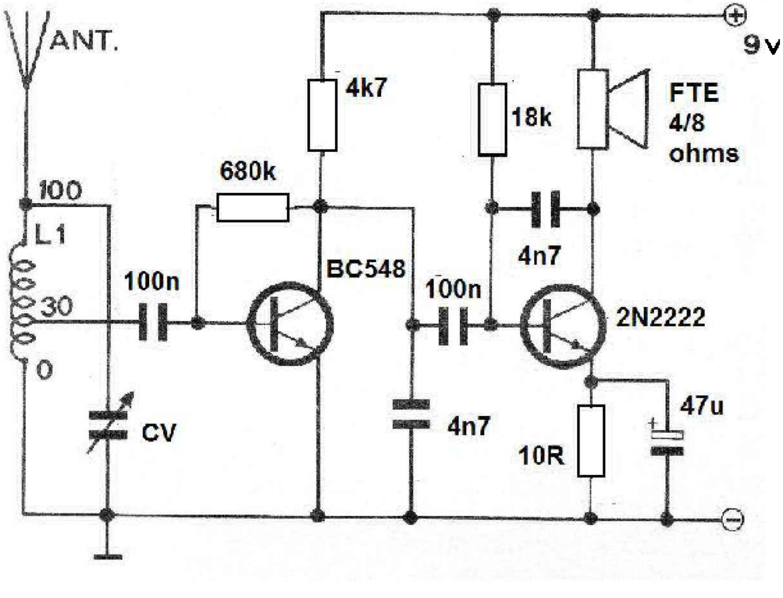
ANT.
100
L1
30
0
CV
100n
680k
4k7
BC548
100n
4n7
18k
4n7
2N2222
FTE
4/8
ohms
9v
47u
10R
+
-

54. Rádio com o 741

Este circuito é de uma publicação americana de 1987 chamada Electronics Handbook. O circuito pode ser montado com facilidade, pois todos os componentes são comuns. L1 consta de 100 espiras de fio 28 num tubo de 2 a 2,5 cm de diâmetro e C1 consiste num capacitor variável obtido de um rádio transistorizado fora de uso. O transformador também pode ser aproveitado de um radio fora de uso. O potenciômetro de 27k (opcional) funciona como controle de tom.

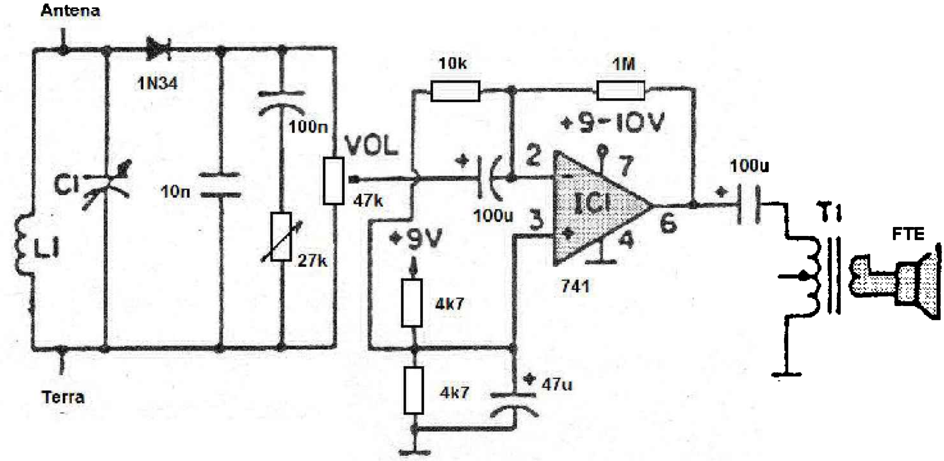
Antena
1N34
CI
L1
10n
100n
27k
VOL
47k
10k
1M
+9-10V
2
7
100u
3
IC1
6
4
741
+9V
4k7
4k7
47u
100u
T1
FTE
Terra

55. Receptor AM Solar

O circuito que apresentamos foi encontrado numa revista inglesa de 1974. Salvo o circuito integrado ZN414 que é pouco comum, os demais componentes podem ser obtidos com facilidade, pois o transistor pode ser trocado pelo BC548. O fone deve ser magnético de alta impedância ou piezoelétrico com um resistor de 4k7 em paralelo. L1 é formada por 100 espiras de fio 28 num bastão de ferrite e o capacitor variável obtido de um velho rádio AM fora de uso. As células solares podem ser obtidas de calculadoras que as utilizem, ou podem ser de outro tipo. O circuito funcionará com tensões de 1,3 V a 6 V.

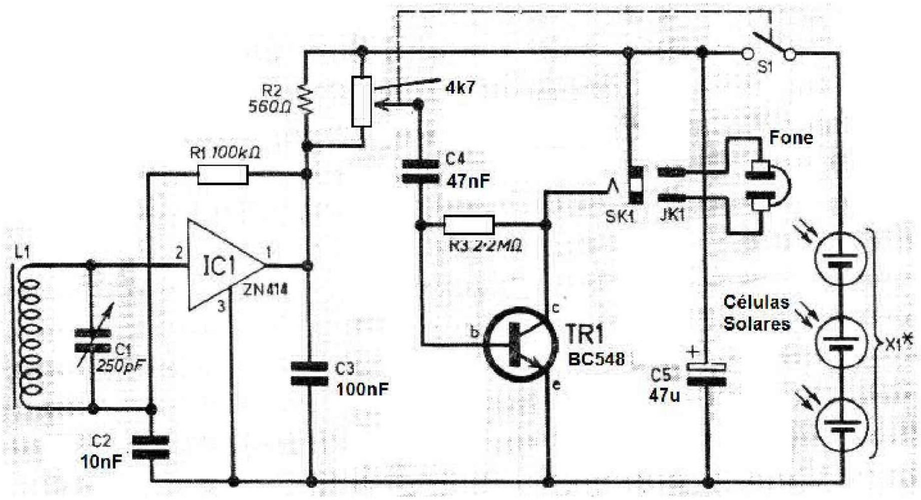
S1
R2
560Ω
4k7
R1 100kΩ
C4
47nF
Fone
SK1
JK1
R3 2·2MΩ
L1
2
IC1
1
ZN414
3
C1
250pF
C3
100nF
b
c
e
TR1
BC548
Células
Solares
+
C5
47u
X1*
C2
10nF

56. Rádio AM de dois transistores (4)

Este simples radinho AM não precisa de antena externa para captar as estações locais. Ele foi encontrado numa documentação técnica de 1979, podendo ser montado com os transistores BC548 ou equivalentes mais modernos dos tipos indicados no original, que eram os BC108. A bobina é enrolada num bastão de ferrite de 20 a 25 cm de comprimento e o fio pode ser o 22 AWG. A tomada é feita na oitava espira, como indicado no diagrama. O fone deve ser magnético de baixa impedância. Se for cerâmico ou de cristal ligue em paralelo um resistor de 1k a 2,2 k. O diodo pode ser o 1N34. O consumo é baixo.

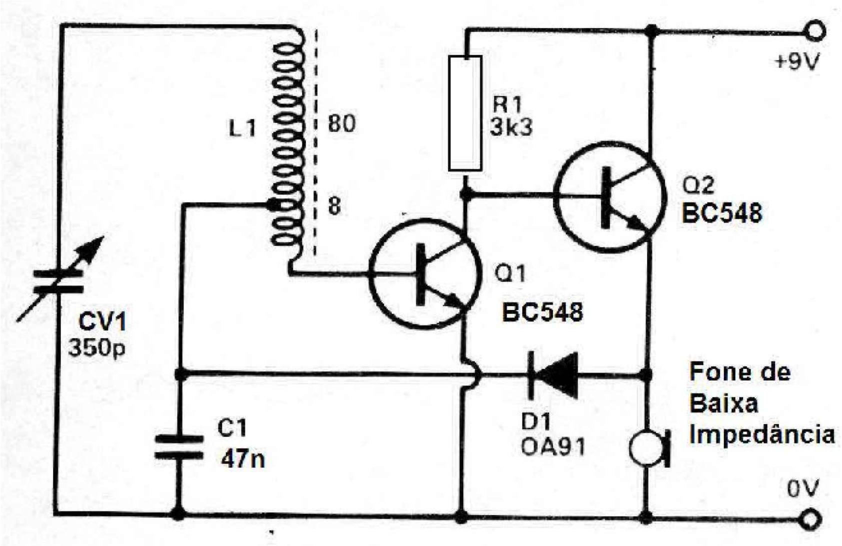
L1
80
8
R1
3k3
+9V
Q2
BC548
Q1
BC548
CV1
350p
C1
47n
D1
OA91
Fone de
Baixa
Impedância
0V

Transmissores

Os transmissores que descrevemos são todos do tipo experimental simples de baixa potência, já que, para operar transmissores potentes é preciso ter permissão especial. Na montagem de transmissores é muito importante que os capacitores das etapas de RF sejam cerâmicos e especial cuidado deve ser tomado com a confecção das bobinas.

57. Micro Transmissor de FM

Este circuito é utilizado com microfones de baixa impedância como, por exemplo, pequenos alto-falantes, cápsulas magnéticas de fones até 600 ohms. O circuito é de uma publicação de 1989, mas pode ser montado com um transistor BF494. A bobina consta de 4 espiras de fio 24 ou 26 em forma de 1 cm sem núcleo. A antena pode ter de 10 a 40 cm de comprimento e o alcance pode chegar a algumas centenas de metros. Os capacitores devem ser cerâmicos. O trimmer não é crítico admitindo valores com capacitâncias máximas até 33 pF.

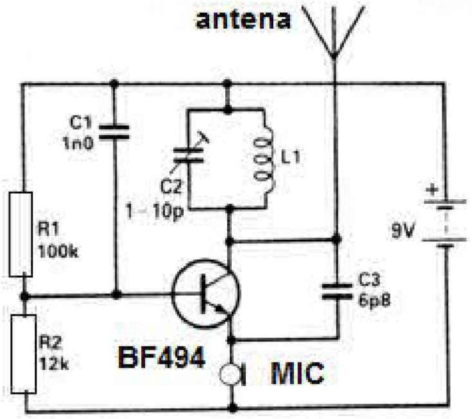

58. Mini Transmissor AM

Este circuito foi encontrado numa revista de 1964, que não existe mais. Trata-se de um pequeno transmissor ou microfone sem fio que envia seus sinais a um rádio AM a alguns metros de distância. O microfone deve ser obrigatoriamente de carvão ou magnético de baixa impedância. As bobinas são formadas por 40 + 40 espiras de fio 28 num bastão de ferrite e a antena é telescópica de 40 a 120 cm de comprimento. A alimentação é feita com 6 ou 9 V e o transistor pode ser o BC558. R1 é de 10k, R2 de 22k, C1 de 10 uF e 2 um variável de rádio AM antigo. Uma alteração consiste em se ligar o resistor de 4k7 e o capacitor de 10u à terra e usar um microfone piezoelétrico em paralelo com estes componentes.

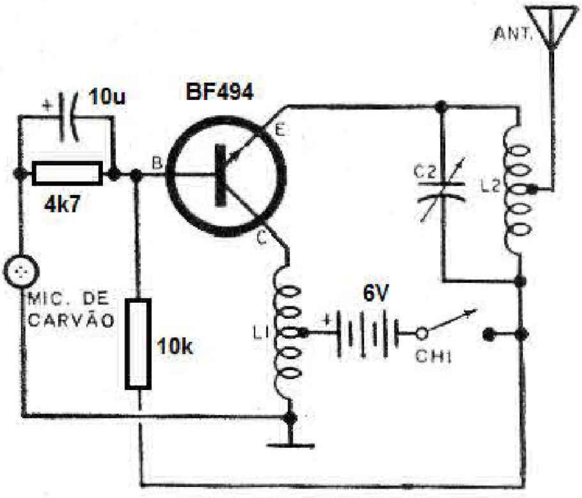

59. Pequeno Transmissor de AM

Este pequeno transmissor AM experimental ou microfone sem fio foi obtido numa antiga publicação latina, dos anos 70. Seu alcance é da ordem de alguns metros com uma pequena antena e o microfone pode ser de qualquer tipo. A alimentação é feita com bateria especial de 15 V para maior alcance, mas ele funcionará com 9 V utilizando-se um transistor BC558. L1 consta de 40 + 40 espiras de fio 28 e O capacitor de 2-20 pF ou mais pode ser um pequeno variável para se obter sintonia, o que no circuito original não existe.

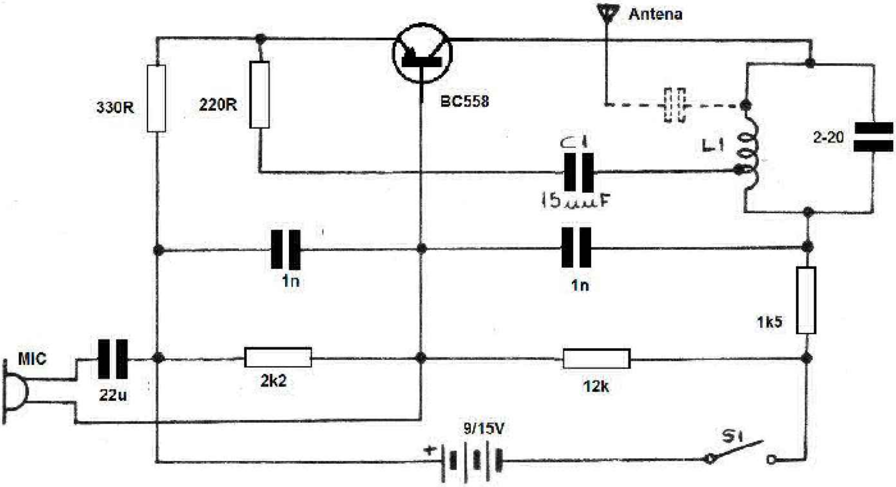
Antena
330R
220R
BC558
C1
15μμF
L1
2-20
1n
1n
1k5
MIC
22u
2k2
12k
9/15V
S1
+

60. Micro Transmissor de FM (2)

Este circuito transmite sinais para um receptor de FM a uma distância de até algumas dezenas de metros. A bobina é formada por 2 + 2 espiras de fio 22 ou 24 em forma de 1 cm sem núcleo. CV é um trimmer de 2-20 pF ou próximo disso e a antena tem de 10 a 30 cm de comprimento. O alcance depende da tensão de alimentação. O circuito é do livro Fun Projects for the Experimenter.

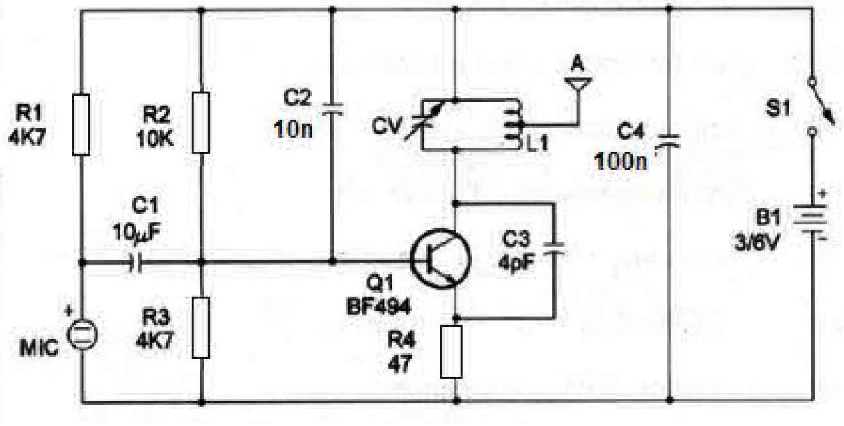
R1
4K7
R2
10K
C2
10n
CV
A
L1
C4
100n
S1
C1
10µF
B1
3/6V
C3
4pF
Q1
BF494
MIC
R3
4K7
R4
47

61. Transmissor de Onda Contínua (CW)

Este circuito foi indicado em artigo da revista americana Electronics Hobbyst do inverno de 1971. A revista não existe mais, mas o projeto pode ser montado com as modificações feitas e a utilização de um transistor BC548 ou BF494. Para a faixa de ondas curtas de 3 a 10 MHz enrole 30 espiras de fio 28 AWG num bastão de ferrite de 10 a 15 cm de comprimento com tomada na 10ª espira a partir do ladro de terra (R1). A antena pode ser telescópica e a alimentação tanto feita com 4 pilhas como bateria de 9 V. C2 é um variável comum tirado de um rádio AM. Os capacitores devem ser cerâmicos. O manipulador para envio de sinais telegráficos pode ser intercalado à fonte de alimentação.

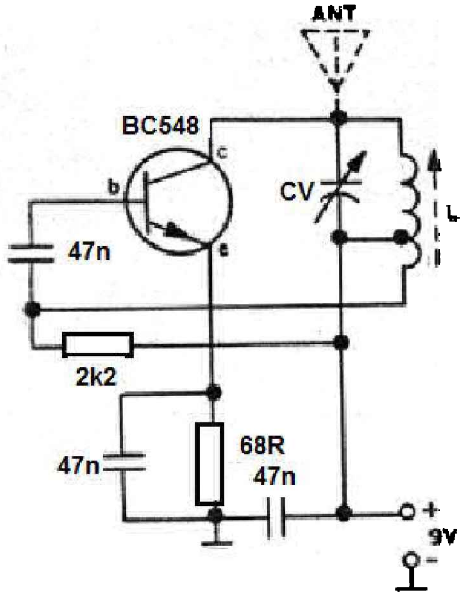

62. Microfone sem fio de FM

Este circuito foi indicado em artigo da revista americana Electronics Hobbyst do inverno de 1971. A revista não existe mais, mas o projeto pode ser montado com a utilização de transistores equivalentes. Para Q1 podemos usar o BC558 e para Q2 o BC558 e se não oscilar qualquer PNP de RF. Os capacitores menores devem ser cerâmicos e a antena deve ter 15 cm ou mais. O microfone é piezoelétrico, mas pode ser usado um eletreto com um resistor de 10k ao positivo da alimentação.

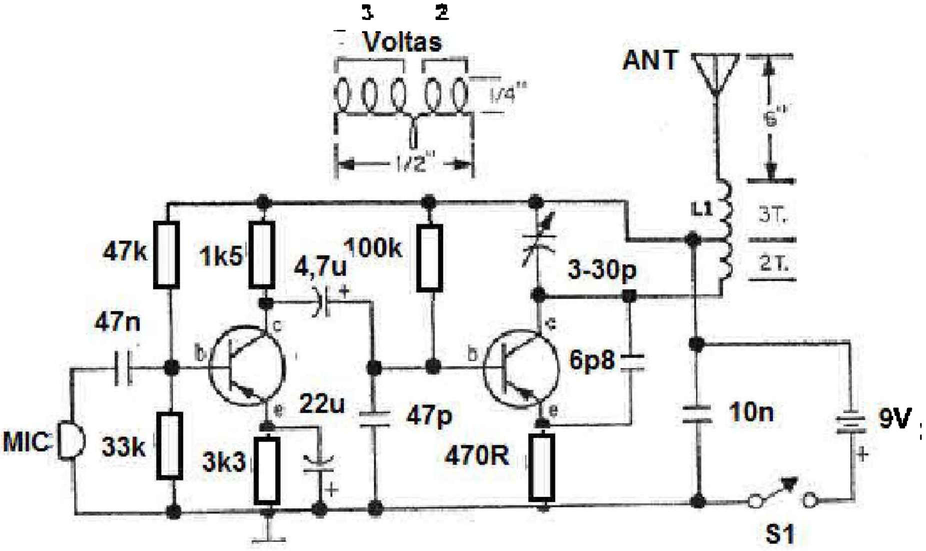
3
2
Voltas
1/4"
1/2"
ANT
6"
L1
3T.
2T.
47k
1k5
4,7u
100k
3-30p
47n
c
b
b
c
6p8
22u
47p
e
e
10n
9V
MIC
33k
3k3
470R
S1

63. Transmissor de Ondas Médias

Este circuito foi obtido na revista italiana Radio Elettronica de agosto de 1974. A revista não mais existe, mas o circuito pode ser montado mesmo em nossos dias com transistores equivalentes como os BC548 para TR1 e TR2, e BD136 para TR3. L1 é formada por 60 + 20 espiras de fio 28 num bastão de ferrite. Alterações nos resistores de polarização podem ser necesssárias para melhor desempenho. A antena pode ser telescópica com pelo menos 60 cm de comprimento. O alcance, dada a faixa de operação, é pequeno, da ordem de alguns metros apenas.

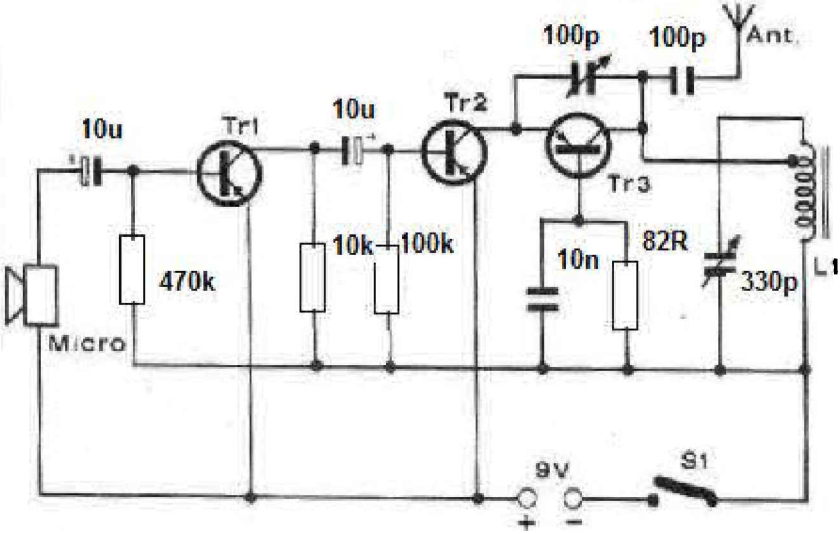
100p
100p
Ant.
10u
Tr1
10u
Tr2
Tr3
470k
10k
100k
10n
82R
330p
L1
Micro
9V
S1
+
-

64. Microfone sem fio AM

Este artigo saiu numa Popular Electronics de julho de 1965. O artigo é atual exceto pelo transistor e também pelo seu alcance muito pequeno. O transistor pode ser um BC558. A bobina consiste em 50 + 50 espiras de fio 28 AWG num bastão de ferrite de 10 a 15 cm de comprimento. O microfone deve ser piezoelétrico. A alimentação pode ser feita com tensões de 6 a 9 V, obtidas de 4 pilhas pequenas ou uma bateria. Em lugar do capacitor fixo de 100 pF pode ser usado um capacitor variável para facilitar a sintonia, pois no original o ajuste é feito apertando-se e afastando-se as espiras da bobina. Uma antena no coletor do transistor aumenta o alcance. Alterando-se a bobina o circuito opera também na faixa de ondas curtas.

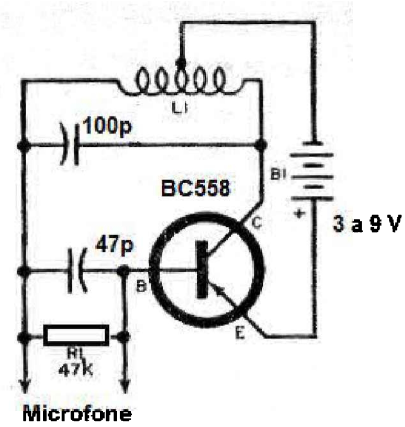

65. Microfone sem fio AM (2)

Este circuito foi encontrado numa revista de 1963, mas ainda pode ser montado com facilidade com alguns componentes atuais. Os transistores podem ser os BC558 e o transformador é obtido de um rádio transistorizado fora de uso. A bobina consta de 40 + 40 espiras de fio 28 num bastão de ferrite de 12 cm de comprimento aproximadamente. A antena tem de 30 a 50 cm e o alcance é de alguns metros na faixa de 550 a 1600 kHz. O resistor R3 de 100R a 1k deve ser obtido experimentalmente, para o melhor desempenho. Os capacitores devem ser cerâmicos.

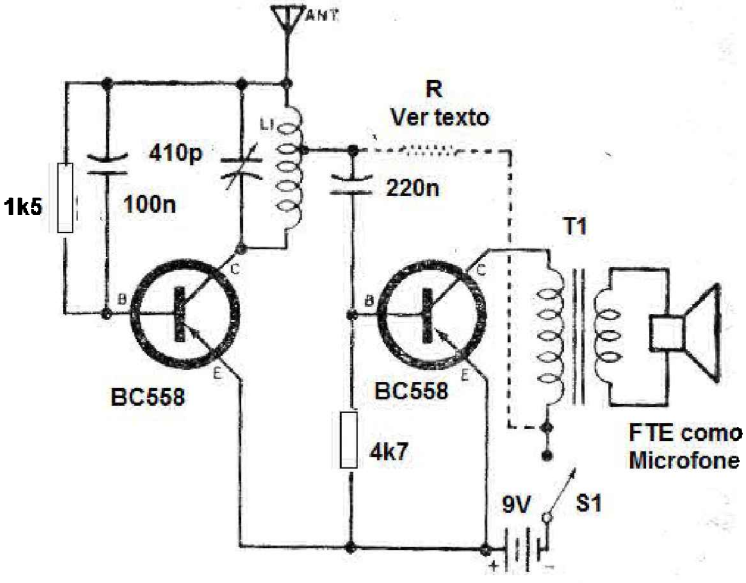
ANT
R
Ver texto
L1
410p
1k5
100n
220n
T1
B
C
E
BC558
B
C
E
BC558
4k7
FTE como
Microfone
9V
S1
+
-

66. Microfone sem fio AM (3)

O microfone deste transmissor AM de curto alcance deve ser magnético de baixa impedância (um pequeno alto-falante) ou de carvão. L1 é formada por 80 espiras de fio 28 num bastão de ferrite e CV é um capacitor variável de antigo radio AM. O transistor pode ser o BC558 e a antena telescópica de 40 cm a 1,5 m de comprimento. A faixa de frequências está entre 500 e 1600 kHz. A bobina Poe ser alterada para operação na faixa de ondas curtas.

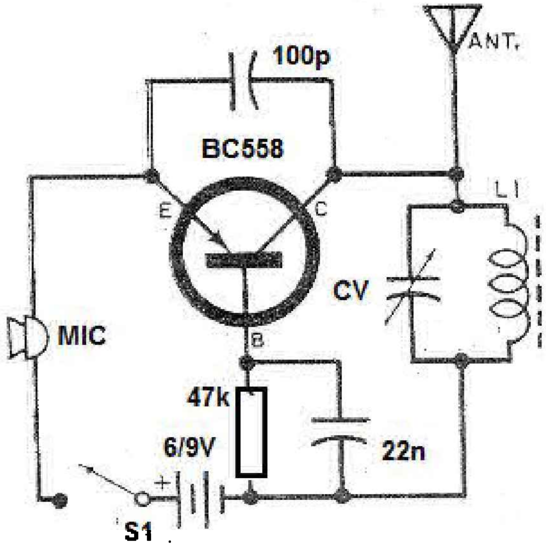

67. Micro Transmissor de AM

Este circuito foi obtido numa tradução de revista americana de 1965. A bobina é formada por 40 + 40 espiras de fio 28 num tubo de papelão que desliza sobre bastão de ferrite para ajuste de frequência. O microfone pode ser um pequeno alto-falante e a antena é do tipo telescópico. A alimentação pode ser feita com tensões de 6 a 9 V. O transistor pode ser o BC548 e a faixa de frequências de operação está entre 550 kHz e 1600 kHz. O alcance é de alguns metros. A bobina pode ser fixa, trocando-se o capacitor de 220 pF por um variável comum de rádio de ondas médias.

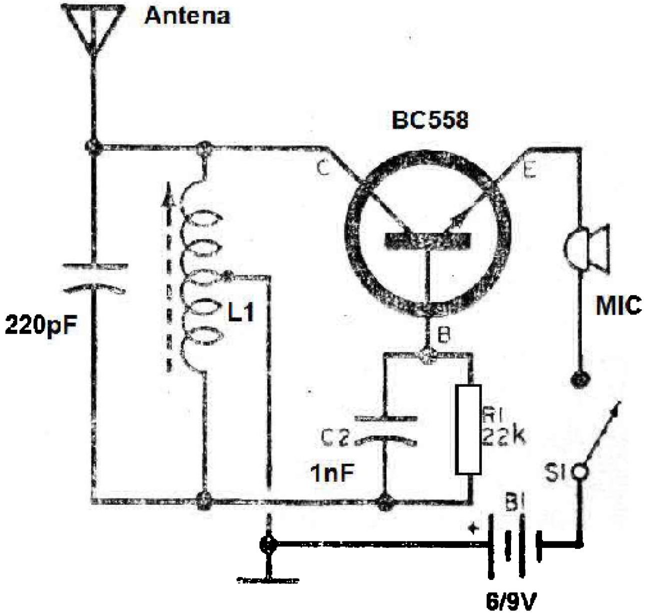

68. Transmissor de FM Alimentado por AC

Este circuito, de nossa autoria, publicado numa antiga edição de década de 90, é alimentado diretamente pela rede de energia. Os sons captados pelo microfone são enviados para um receptor a uma distância de algumas dezenas de metros e recebidos num rádio comum de FM. O circuito tem boa sensibilidade e baixo consumo. A bobina é formada por 4 espiras de fio rígido fino com diâmetro de 1 cm sem núcleo e a antena um pedaço de fio rígido de 12 a 20 cm de comprimento. Os capacitores devem ser cerâmicos, exceto C1 que é de poliéster para 400 V e C2 eletrolítico. O valor entre parênteses é para a rede de 220 V.

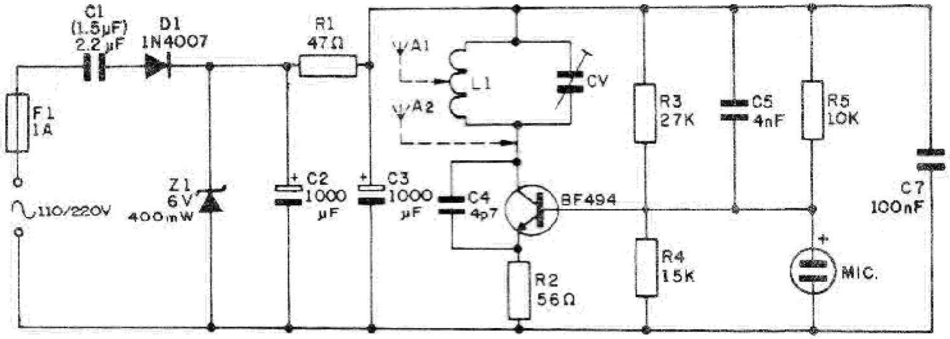
C1
(1.5µF)
2.2µF
D1
1N4007
R1
47Ω
A1
A2
L1
CV
R3
27K
C5
4nF
R5
10K
F1
1A
110/220V
Z1
6V
400mW
C2
1000
µF
C3
1000
µF
C4
4p7
BF494
C7
100nF
R2
56Ω
R4
15K
MIC.

69. Micro Transmissor de FM (4)

Este circuito é utilizado com microfones de alta impedância como, por exemplo, cápsulas de cerâmica ou cristal. O circuito é de uma publicação de 1989, mas pode ser montado com um transistor BF494. A bobina consta de 4 espiras de fio 24 ou 26 em forma de 1 cm sem núcleo. A antena pode ter de 10 a 30 cm de comprimento e o alcance pode chegar a algumas centenas de metros. Os capacitores devem ser cerâmicos. O circuito também funciona bem com alimentação d 3 ou 6 V, mas com menor alcance. Para microfone de eletreto, ligue entre o positivo do eletreto e da alimentação um resistor de 4k7. C2 é um trimmer comum de 10 a 30 pF de capacitância máxima.

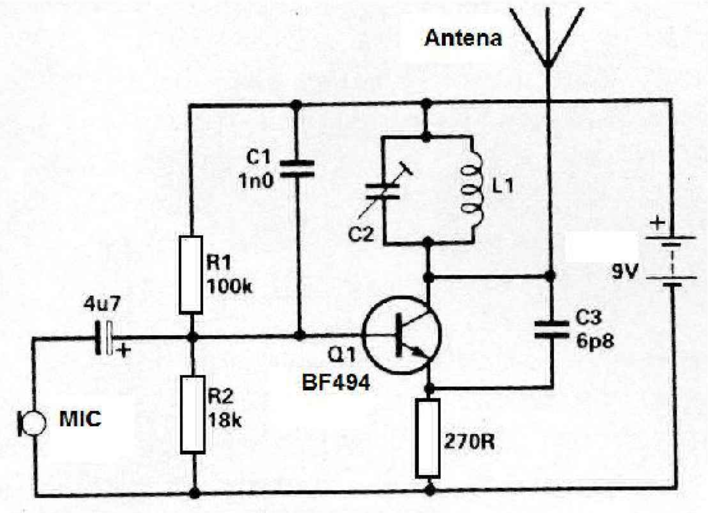

70. Micro Transmissor de FM (5)

Este circuito é utilizado com microfones de alta impedância como, por exemplo, cápsulas de cerâmica ou cristal. O circuito é de uma publicação de 1989, mas pode ser montado com um transistor BF494 e um BC548. A bobina consta de 7 espiras de fio 24 ou 26 em forma de 0,5 cm com núcleo ajustável. A antena pode ter de 10 a 30 cm de comprimento e o alcance pode chegar a algumas dezenas de metros. Os capacitores devem ser cerâmicos. A alimentação original é com uma pilha botão d 1,3 V (mercúrio) ou 1,5 V (comum), mas o circuito também funciona com alimentação de 1,5 V ou 3 V de pilhas AAA ou AA.

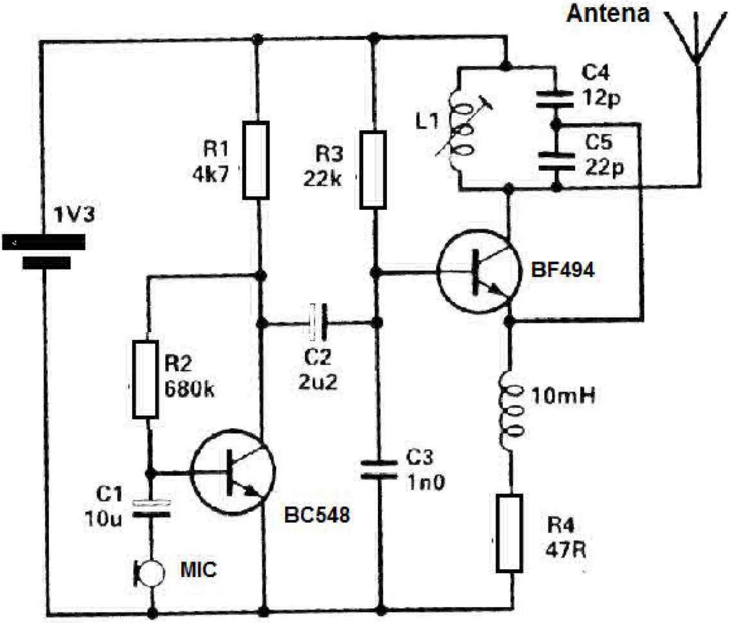
Antena
C4
12p
C5
22p
L1
R1
4k7
R3
22k
1V3
BF494
R2
680k
C2
2u2
10mH
C1
10u
BC548
C3
1n0
MIC
R4
47R

71. Transmissor de FM

O alcance deste pequeno transmissor de FM pode superar os 100 metros. L1 tem 4 espiras de fio 20 a 22 sem núcleo em forma de 1 cm. A antena é um fio rígido de 10 cm a 40 cm. Q1 pode ser um BC548 e Q2 um BF494 ou 2N2222. O microfone é de eletreto. O circuito também funciona com 6 V. C5 pode ser qualquer trimmer comum de 20 a 50 pF de capacitância máxima. Os capacitores menores devem ser todos cerâmicos, com especial atenção para C6 cujo valor pode necessitar de alteração na faixa de 4,7 pF a 10 pF.

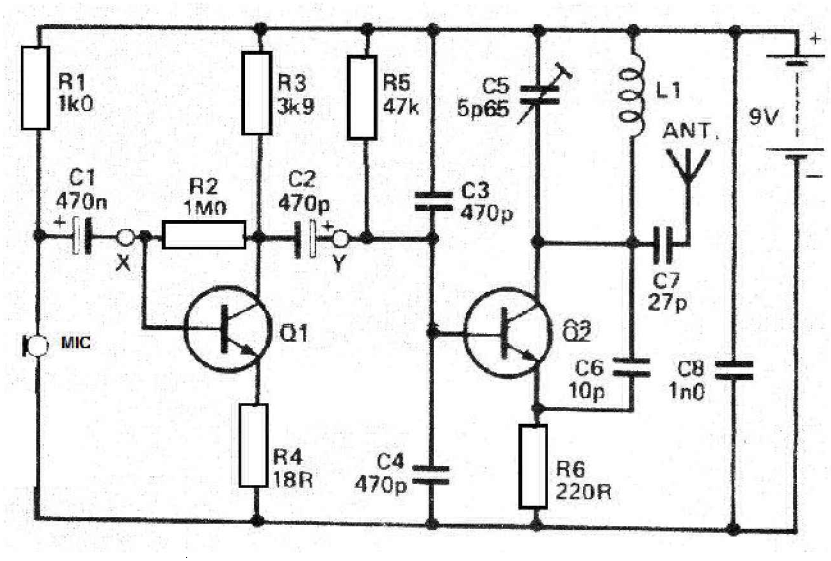
R1
1k0
R3
3k9
R5
47k
C5
5p65
L1
ANT.
9V
+
−
C1
470n
R2
1M0
C2
470p
C3
470p
X
Y
C7
27p
MIC
Q1
Q2
C6
10p
C8
1n0
R4
18R
C4
470p
R6
220R

72.Micro Transmissor de Ondas Curtas

Este transmissor telegráfico de ondas curtas pode enviar seus sinais a um receptor a algumas dezenas de metros. L1 depende da faixa podendo ter de 10 + 10 a 20 + 20 espiras de fio 28 num bastão de ferrite para a faixa de 5 a 15 MHz. CV é um capacitor variável de rádio AM e a antena pode ter de 50 cm a 2 metros de comprimento. A alimentação pode ser feita com tensões de 6 a 9 V. Para 12 V use um transistor BD135. Os capacitores devem ser cerâmicos. O circuito é de um livro do autor publicado nos Estados Unidos.

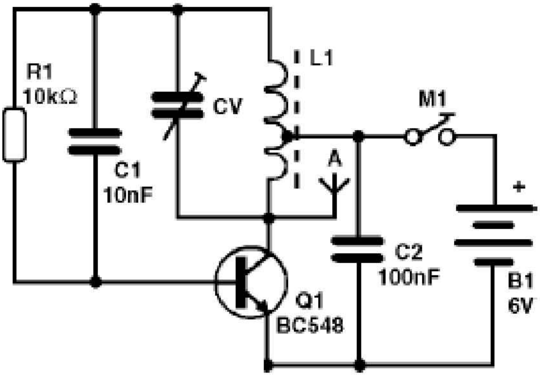

73. Microfone sem fio AM (4)

Este circuito foi obtido numa publicação de 1965, mas ainda pode ser montado com um transistor BC558 em lugar do original que era de germânio. L1 consta de 50 + 50 espiras de fio 28 num tubo de papelão que pode ser deslizado sobre um bastão de ferrite para ajuste da frequência. A antena é telescópica de 30 a 100 cm, e o microfone do tipo piezoelétrico. O alcance é de alguns metros.

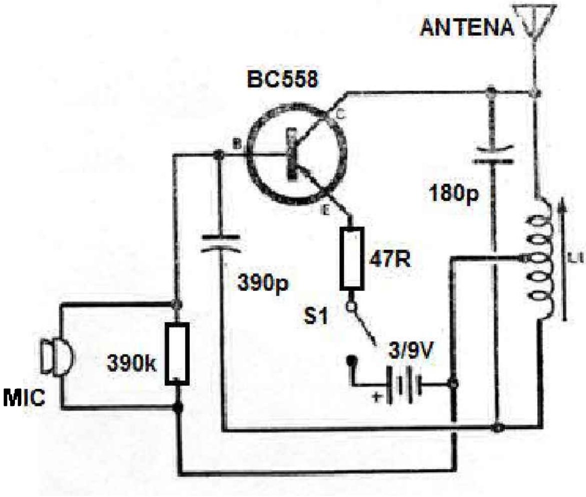

74. Microfone sem fio AM (5)

Este pequeno transmissor AM de curto alcance foi obtido numa publicação argentina dos anos 1970. O circuito tem alguns metros de alcance, utilizando uma pequena antena telescópica e o microfone deve ser piezoelétrico (cerâmico). A bobina é formada por 40 + 40 espiras de fio 28 num bastão de ferrite e pode ser usado um transistor BC558 em lugar do original. A alimentação deve ser feita com 6 ou 9 V e a sintonia é feita deslocando-se o núcleo de ferrite dentro da bobina, que deve ser enrolada num tubinho de papelão. Pode-se ter a sintonia num capacitor variável de rádio AM ligado em paralelo, se a bobina for fixada no núcleo de ferrite.

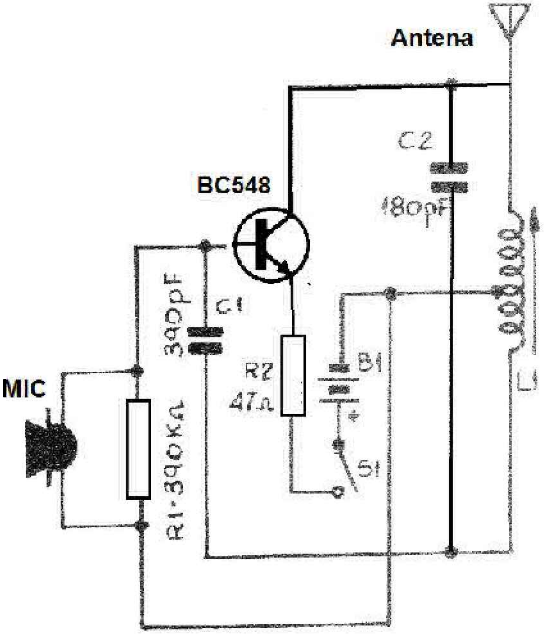

75. Microfone sem fio de FM (2)

Este excelente circuito de microfone sem fio para a faixa de FM modulado por varicap saiu na revista Electronics Now &Radio Electronics de fevereiro de 1993, podendo ser montado com facilidade com um transistor 2N2222 ou BF494 em lugar do original. O microfone pode ser piezoelétrico e a bobina é formada por 4 espiras de fio 22 a 26 em forma de 1 cm sem núcleo. A antena tem de 10 a 30 cm de comprimento. Varicaps da série BB podem ser usados em lugar do original, que era de tipo pouco comum no nosso mercado de componentes. A troca já foifeita quando redesenhamos o circuito.

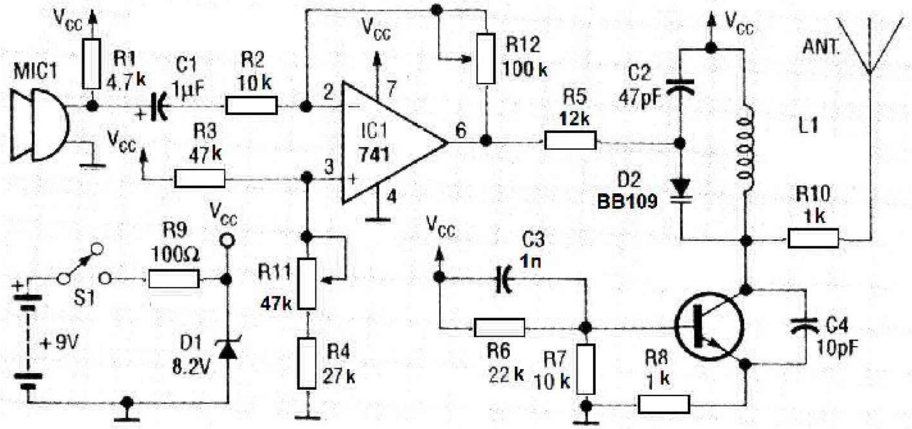
MIC1
Vcc
R1 4.7k
C1 1µF
R2 10k
R3 47k
Vcc
R12 100k
2
3
7
6
4
IC1 741
R11 47k
R4 27k
R9 100Ω
Vcc
S1
+9V
D1 8.2V
R5 12k
C2 47pF
D2 BB109
L1
ANT.
R10 1k
C3 1n
Vcc
R6 22k
R7 10k
R8 1k
C4 10pF

76. Transmissor de FM com Integrado

Este transmissor usa um amplificador operacional 741 na modulação. O microfone pode ser qualquer tipo e uma antena de 10 a 20 cm pode ser ligada ao coletor do transistor L1 tem 4 espiras de fio 22 a 26 em forma de 1 cm sem núcleo e o trimmer é comum. O circuito pode ser alimentado por tensões de 6 a 9 V e seu alcance é da ordem de algumas dezenas de metros.

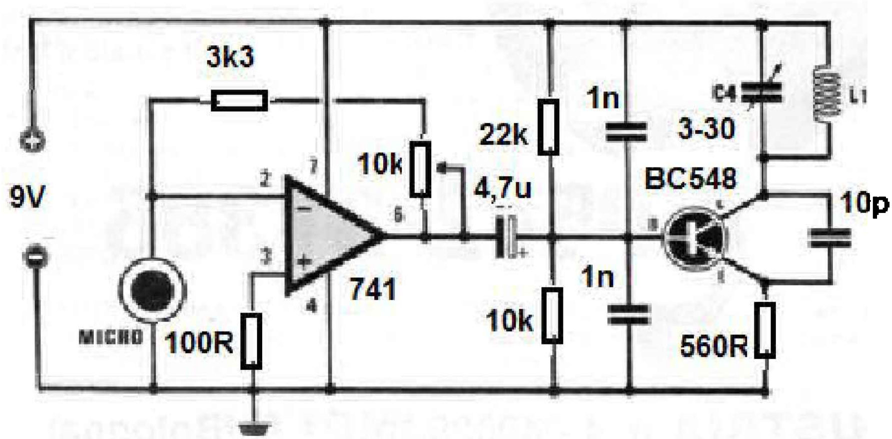
3k3
9V
10k
2
7
6
3
4
741
MICRO
100R
22k
1n
4,7u
10k
1n
C4
3-30
L1
BC548
10p
560R

77.Rádio AM de Três Transistores

Este simples rádio AM foi encontrado numa publicação de 1975, mas pode ser montado com transistores BC548 em lugar dos originais que eram BC108. O fone também pode ser uma cápsula cerâmica. O circuito pode ser alimentado com 3 ou mesmo 4 pilhas. L1 consiste em 80 espiras de fio 28 num bastão de ferrite. Uma antena pode ser ligada a L1. C1 é um capacitor variável obtido num rádio AM antigo. A antena pode ser ligada à base do transistor.

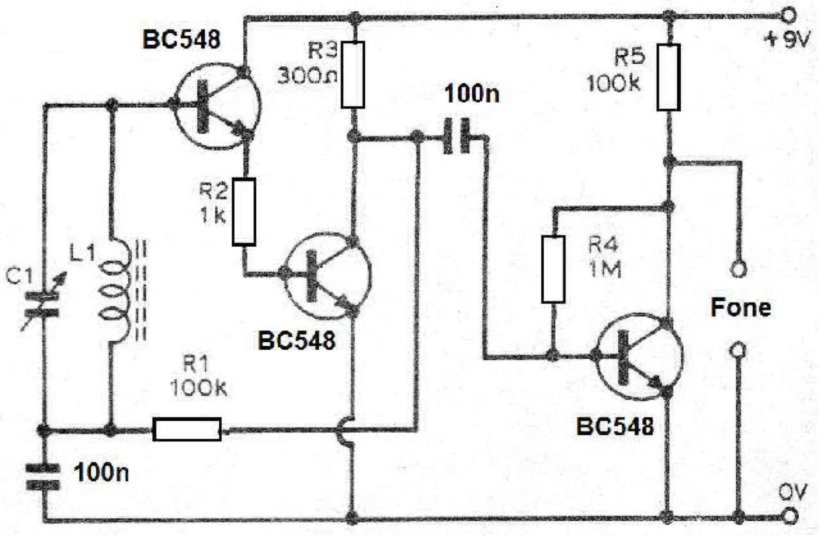
BC548
R3
300Ω
100n
R5
100k
+9V
R2
1k
R4
1M
L1
C1
BC548
Fone
R1
100k
BC548
100n
0V

78. Micro Espião de FM

Este circuito foi encontrado numa Radio Elettronica (Itália) de abril de 1974. O circuito pode ser montado ainda hoje com o transistor 2N2222. Somente o circuito integrado original é mais difícil de ser obtido. O alcance é da ordem de algumas dezenas de metros e a bonina consiste em 4 espiras de fio 26 ou 28 com diâmetro de 1 cm sem núcleo. A antena deve ter de 10 a 40 cm se for usada, devendo ser ligada ao coletor de Q1. O problema maior deste circuito é encontrar o circuito integrado TAA320, que não mais é fabricado. Uma possibilidade consiste em substituir esta etapa por um estágio de um único transistor com mudanças no circuito, devido ao acoplamento do microfone.

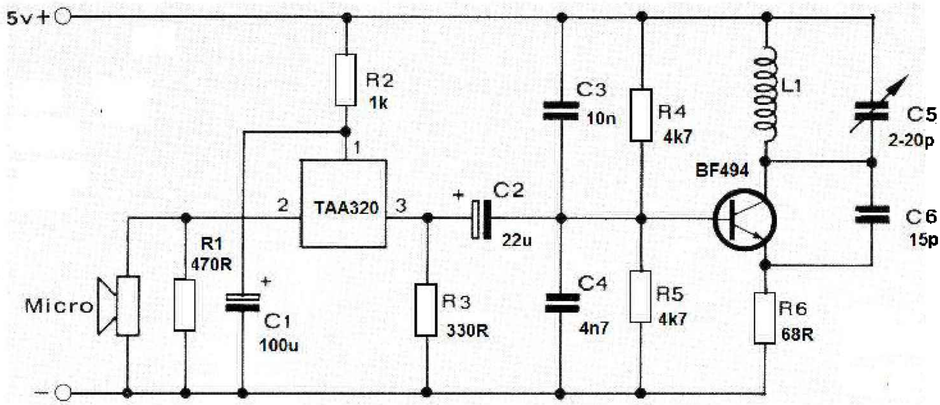
5v+
R2
1k
1
2
TAA320
3
C3
10n
R4
4k7
L1
C5
2-20p
BF494
+
C2
22u
C6
15p
R1
470R
+
Micro
C1
100u
R3
330R
C4
4n7
R5
4k7
R6
68R
−

79. Transmissor para 7 MHz

Este transmissor da Radiotron foi publicado num livro de 1953. O circuito pode ser implementado com um BC558 e tem um alcance que depende da antena. Os capacitores devem ser cerâmicos e a bobina é formada por 14 espiras de fio 28 num tubo de 2 cm sem núcleo. A bobina que acopla a antena tem 4 espiras do mesmo fio.

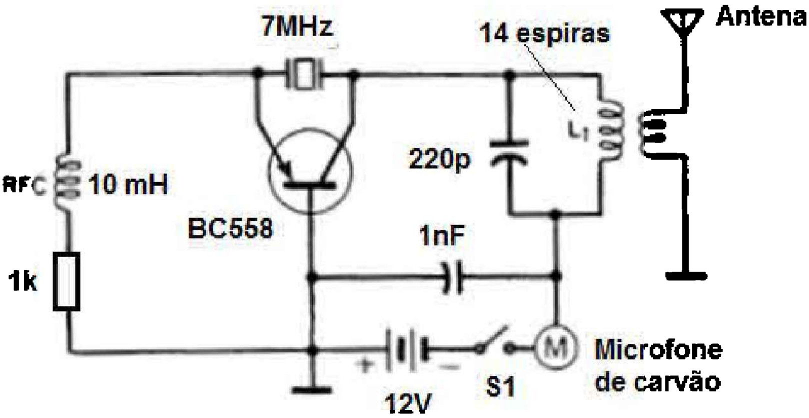
7MHz
14 espiras
Antena
L1
220p
RFC
10 mH
BC558
1nF
1k
+
-
12V
S1
M
Microfone
de carvão

80. Transmissor de tom de 30 a 100 MHz

Este circuito pode ser usado para transmitir sinais modulados em tom para controles remotos ou ainda pulsos num sistema de link de dados. O alcance pode chegar aos 100 metros com uma antena telescópica de 30 cm. A bobina é feita com fio 22 em forma de 1 cm de diâmetro sem núcleo. CV é um trimmer de 2-20 pF, aproximadamente, e todos os capacitores devem ser cerâmicos. O circuito é de um livro de Newton C. Braga publicado nos Estados Unidos.

Faixa	Espiras
30 to 50 MHz	10
50 to 70 MHz	6
70 to 90 MHz	4
90 to 100 MHz	3

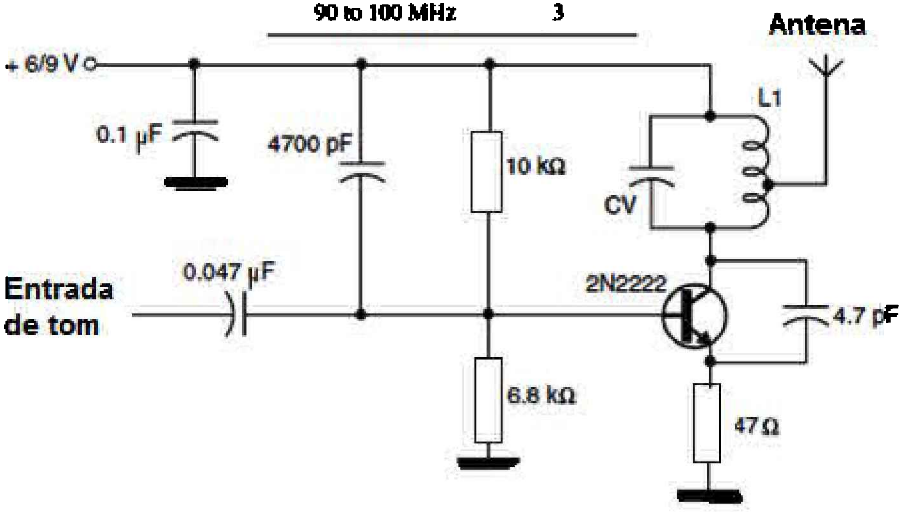
Antena
+ 6/9 V
L1
0.1 µF
4700 pF
10 kΩ
CV
Entrada de tom
0.047 µF
2N2222
4.7 pF
6.8 kΩ
47 Ω

81. Transmissor sem bobinas

Na verdade este circuito transmite um sinal rico em harmônicas sem sintonia alguma podendo ser captado num receptor de ondas médias ou curtas nas proximidades em diversos pontos de seu dial. O circuito é experimental devendo usar uma pequena antena, no máximo de 1 metro. Sua finalidade é apenas demonstrar a produção de sinais de rádio por um circuito oscilador. Este circuito é do livro Science Fair Projects de Newton C. Braga.

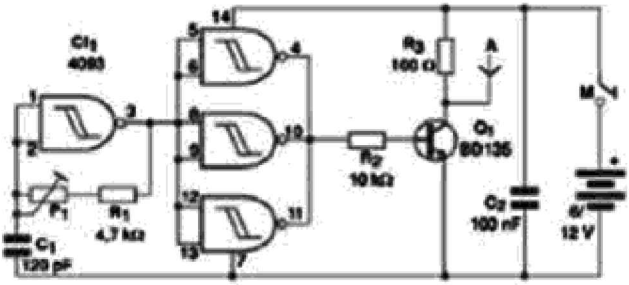
CI1
4093
1
2
3
14
5
6
4
8
9
10
12
13
11
7
P1
R1
4,7 kΩ
C1
120 pF
R2
10 kΩ
R3
100 Ω
A
Q1
BD135
C2
100 nF
M
+
6/
12 V

82. Transmissor de FM Com Diodo Tunnel

Este micro transmissor de FM utiliza um diodo tunnel como oscilador. Encontrei-o numa publicação inglesa de 1989. Veja que a alimentação é feita com uma bateria de mercúrio de 1,3 V. O circuito também funcionará com uma pilha botão ou AAA de 1,5 V. O microfone é uma cápsula cerâmica ultraminiatura ou mesmo magnética. O alcance é de apenas alguns metros, mas o circuito pode ter uma montagem extremamente pequena. L1 consiste em 5 espiras de fio 30 em forma de 0,5 cm sem núcleo.

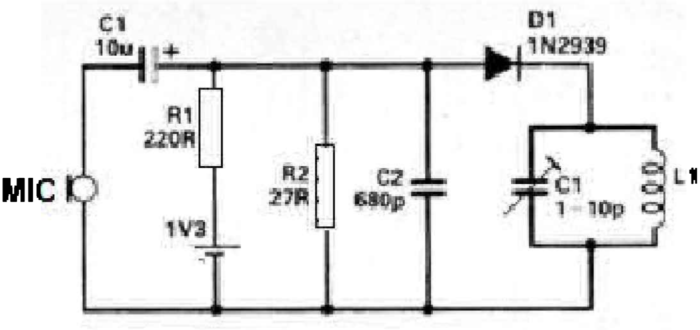

83. Transmissor de FM com Diodo Tunnel (2)

Este micro transmissor de FM utiliza um diodo tunnel como oscilador. Encontrei-o numa publicação inglesa de 1989. Veja que a alimentação é feita com uma bateria de mercúrio de 1,3 V, mas pode ser usada uma pilha botão ou AAA. O microfone é uma cápsula cerâmica ultraminiatura ou mesmo magnética. O alcance é de apenas alguns metros, mas o circuito pode ter uma montagem extremamente pequena, servindo como espião ou localizador. L1 consiste em 5 espiras de fio 30 em forma de 0,5 cm sem núcleo. CV é um trimmer com capacitância máxima d 10 a 33 pF. O transistor pode ser o BC548 ou outro de uso geral.

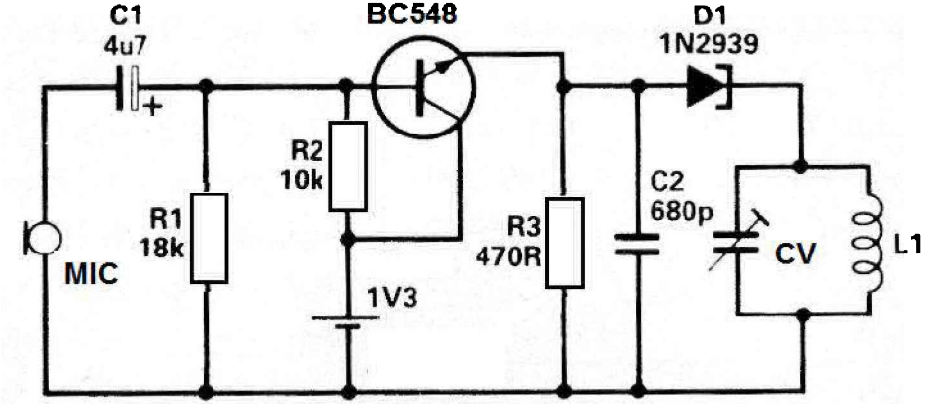

84. Transmissor de Bips Sinalizador

Este transmissor opera na faixa de FM emitindo bips que podem ser captados num rádio comum a uma distância até mais de 100 metros em campo aberto. A antena é telescópica de 20 a 80 cm e a bobina é formada por 2 + 2 espiras de fio 28 numa forma sem núcleo de 1 cm de diâmetro. CV é um trimmer comum de 10 a 40 pF de capacitância máxima. O circuito é do livro Science Fair Projects de Newton C. Braga.

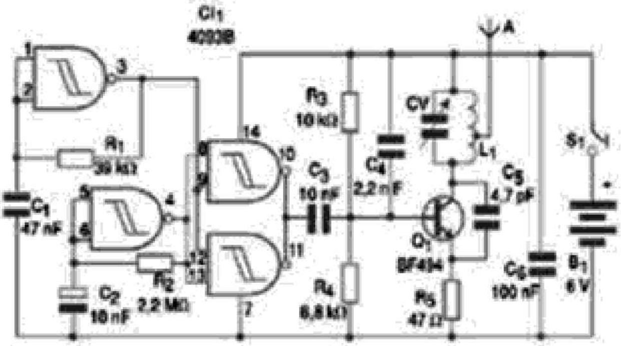

CI1
4093B
1
2
3
R1
39 kΩ
C1
47 nF
5
6
4
C2
10 nF
R2
2,2 MΩ
14
8
9
10
12
13
11
7
R3
10 kΩ
C3
10 nF
R4
6,8 kΩ
C4
2,2 nF
CV
L1
A
C5
4,7 pF
Q1
BF494
R5
47 Ω
C6
100 nF
S1
+
B1
6 V

85. Transmissor Sinalizador de 49 MHz

Este circuito emite um sinal contínuo na frequência de 49 MHz ou outra conforme o cristal utilizado. Ele pode ser utilizado em veículos ou objeto para sua localização. A antena é telescópica e a autonomia depende da bateria. Para maior autonomia use pilhas em lugar de bateria de 9 V. O alcance depende também da sensibilidade do receptor. O trimmer é de 30 ou 40 pF de capacitância máxima.

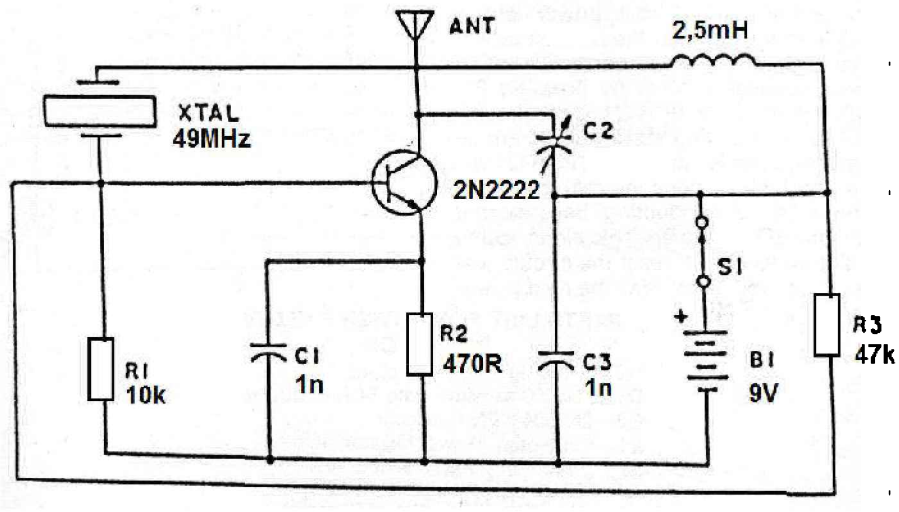
ANT
2,5mH
XTAL
49MHz
C2
2N2222
S1
+
R1
10k
C1
1n
R2
470R
C3
1n
B1
9V
R3
47k

86. *Micro Transmissor Modulado de Ondas Curtas*

Este transmissor de ondas curtas pode enviar seus sinais a um receptor a algumas dezenas de metros. L1 depende da faixa podendo ter de 10 + 10 a 20 + 20 espiras de fio 28 num bastão de ferrite para a faixa de 5 a 15 MHz. CV é um capacitor variável de rádio AM e a antena pode ter de 50 cm a 2 metros de comprimento. A alimentação pode ser feita com tensões de 6 a 9 V. Para 12 V use um transistor BD135. Os capacitores devem ser cerâmicos.

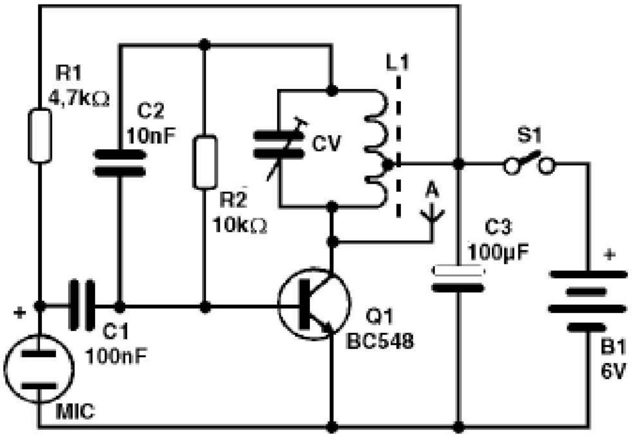
R1
4,7kΩ
C2
10nF
L1
CV
S1
R2
10kΩ
A
C3
100µF
+
+
C1
100nF
Q1
BC548
B1
6V
MIC

87. Micro Transmissor de FM Para 220 V

Este circuito é de uma revista italiana de 1993, mas pode ser montado com facilidade ainda hoje. Para 110 V, basta usar um capacitor de poliéster de 560 nF em lugar do C1 original de 330 nF. Uma pequena antena pode ser ligada ao coletor de TR2. Os transistores podem ser o BC548 para TR1 e BF494 para TR2. L1 tem 4 espiras de fio 22 a 26 em forma de 1 cm sem núcleo. O microfone é de eletreto de 3 terminais.

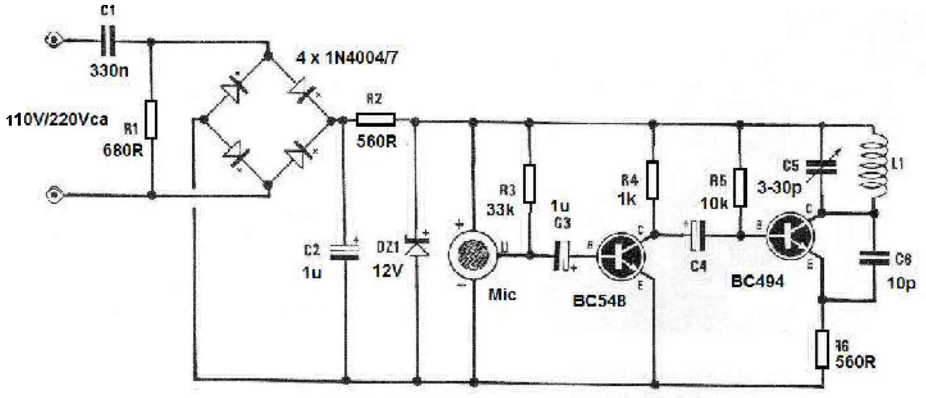
C1
330n
110V/220Vca
R1
680R
4 x 1N4004/7
R2
560R
C2
1u
DZ1
12V
Mic
R3
33k
1u
C3
R4
1k
BC548
R5
10k
C4
C5
3-30p
L1
BC494
C6
10p
R6
560R

88. Transmissor de Bips

Este transmissor emite bips que podem ser captados num receptor de FM a uma distância de até algumas centenas de metros. Para a faixa de FM L1 consta de 4 espiras de fio 22 ou 24 em forma de 1 cm sem núcleo e CV é um trimmer de aproximadamente 2 -20 pF.Os capacitores devem ser cerâmicos e a antena deve ter de 15 cm a 40 cm de comprimento. O circuito é do livro Fun Projects for the Experimenter de Newton C. Braga.

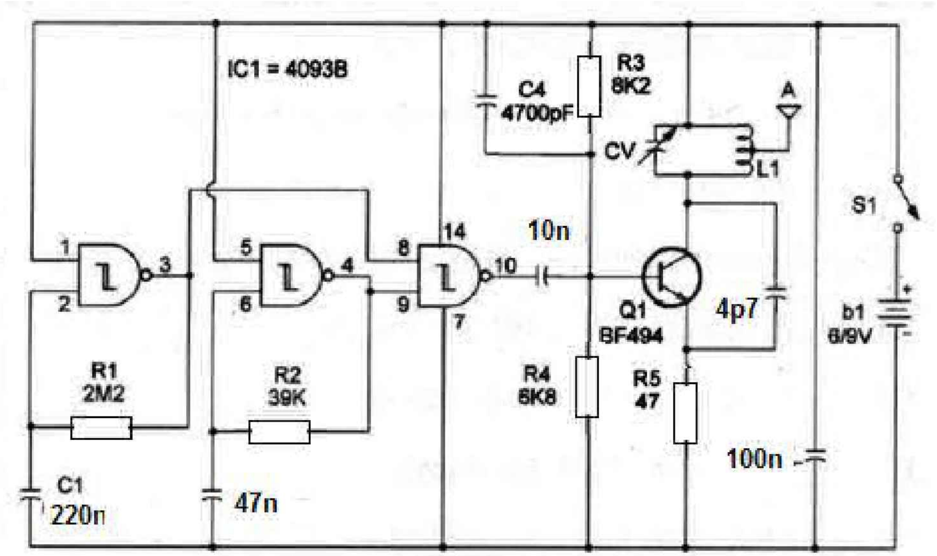
IC1 = 4093B
C4
4700pF
R3
8K2
CV
A
L1
S1
1
2
3
5
6
4
8
9
14
10
7
10n
Q1
BF494
4p7
b1
6/9V
R1
2M2
R2
39K
R4
6K8
R5
47
100n
C1
220n
47n

89. Microfone sem fio AM (6)

Encontrei este circuito numa publicação de 1964. O circuito ainda pode ser montado com facilidade com um transistor BC558 em lugar do original. O alcance é de alguns metros com uma antena telescópica de 40 a 100 cm. A bobina L2 consiste em 80 espiras de fio 28 num tubo de papelão que desliza sobre um bastão de ferrite para o ajuste de frequência. L1 é um choque de 2,5 mH. O microfone deve ser de baixa impedância, por exemplo, um pequeno alto-falante. A alimentação é feita por uma tensão de 3 a 9 V. A frequência de operação vai de 550 a 1600 kHz. O trimpot de ajuste é de 47k (será conveniente ligar um resistor de 4k7 em série, não dado na publicação original) e C1 é de 10 nF. C2 é de 470 pF.

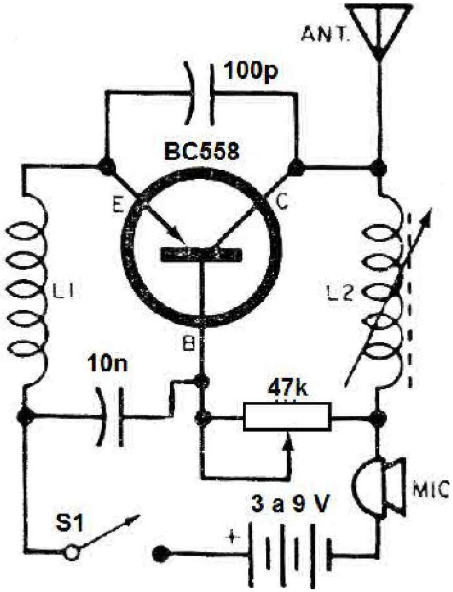

Circuitos Diversos

Damos a seguir uma seleção de circuitos que são utilizados tanto com receptores como transmissores. Neste grupo incluímos os reforçadores de sinais, comprovadores de transmissores, detectores de transmissores e muito mais.

90. Verificador de Transmissores AM

Este circuito foi indicado em artigo da revista americana Electronics Hobbyst do inverno de 1971. A revista não existe mais, mas o projeto pode ser montado com facilidade, pois os componentes são comuns. O fone pode ser de cristal ou uma cápsula piezoelétrica. A bobina consiste em 3 voltas de qualquer fio numa forma de 3 a 4 cm sem núcleo que deve ser aproximada da bobina de saída do transmissor em teste. O diodo pode ser o originalmente indicado 1N60 ou equivalente como o 1N34. O circuito também funciona com transmissores da faixa de ondas curtas e mesmo osciladores modulados.

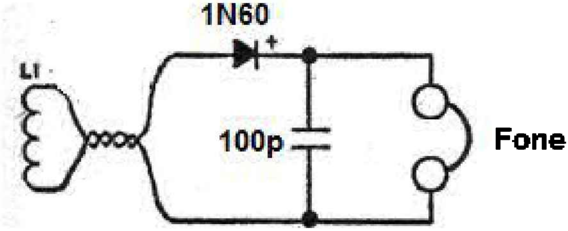

91.Detector de RF

Na verdade, trata-se de um indicador de intensidade de campo com um amplificador operacional equivalente ao 741. Este circuito foi encontrado numa Popular Electronics de 1991, mas pode ser montado com facilidade ainda hoje. A antena é telescópica e o choque pode ser enrolado com 50 espiras de fio 32 num bastão de ferrite.

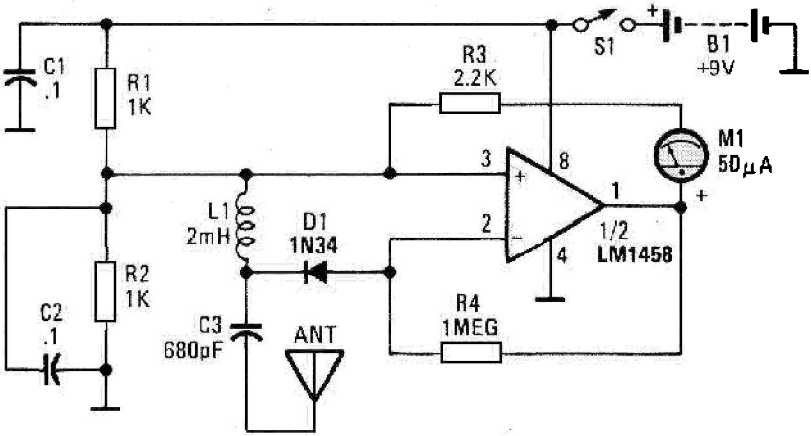
C1
.1
R1
1K
R2
1K
C2
.1
L1
2mH
D1
1N34
C3
680pF
ANT
R3
2.2K
R4
1MEG
3
2
8
4
1
1/2
LM1458
M1
50µA
S1
B1
+9V

92. Detector Ativo de RF

Este circuito foi encontrado numa revista Poptronics de dezembro de 2001. O circuito consiste num detector de RF com amplificação. L1 deve ter indutância de acordo com a frequência do sinal que está sendo detectado. O transistor também pode ser o BF245 e os capacitores devem ser cerâmicos. O consumo do circuito é baixo, da ordem de poucos miliampères.

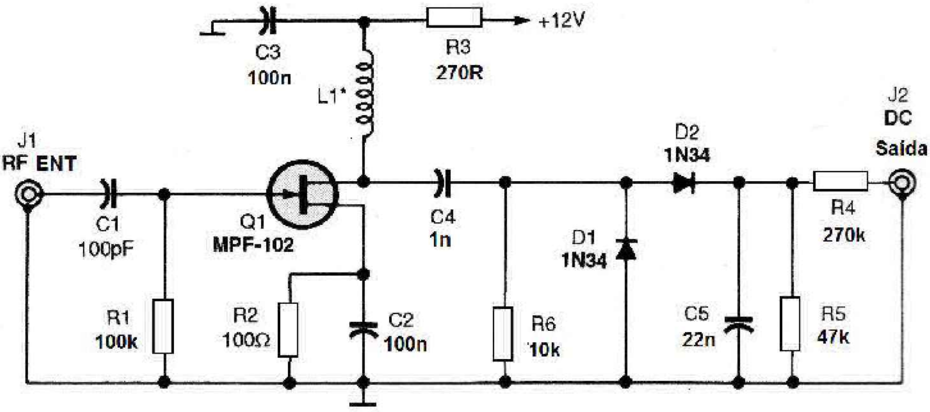
C3
100n
R3
270R
+12V
L1*
J1
RF ENT
C1
100pF
Q1
MPF-102
C4
1n
D2
1N34
J2
DC
Saída
R4
270k
D1
1N34
R1
100k
R2
100Ω
C2
100n
R6
10k
C5
22n
R5
47k

93. Localizador de Transmissores

Com este circuito detector de sinais de rádio é possível localizar transmissores ocultos ou fontes de interferências. O circuito é da revista Hands-On Electronics dos Estados Unidos, edição de fevereiro de 1987. A revista não existe mais, mas o circuito pode ser montado como par de transistores BC548 e BC558. A antena é telescópica e buzzer é opcional podendo ser substituído por um miliamperímetro em série com um trimpot de 10k ohms.

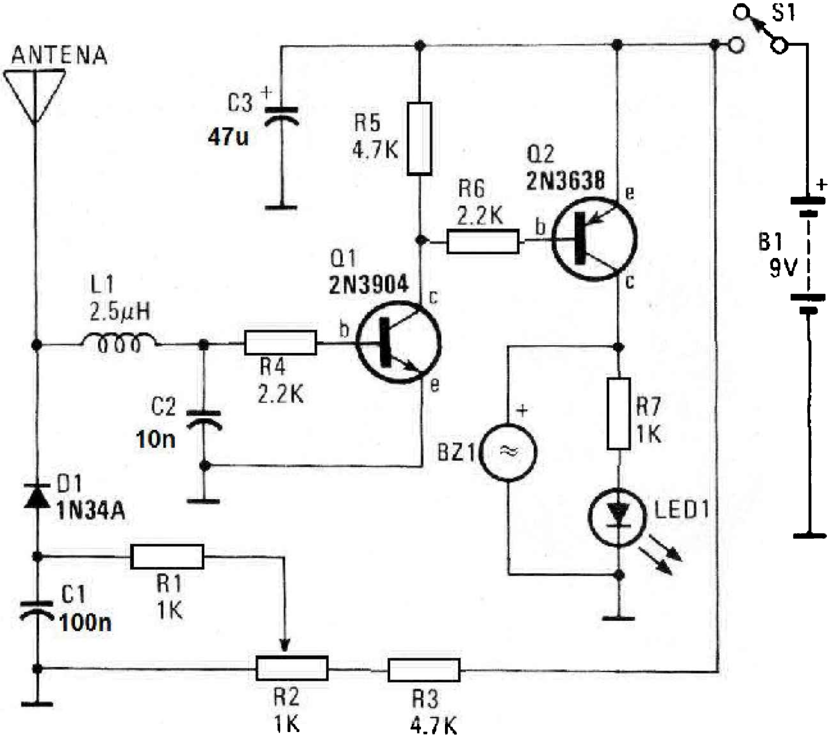
ANTENA
S1
C3 +
47u
R5
4.7K
Q2
2N3638
R6
2.2K
e
b
c
+
B1
9V
Q1
2N3904
c
b
e
L1
2.5µH
R4
2.2K
C2
10n
+
BZ1
R7
1K
D1
1N34A
LED1
C1
100n
R1
1K
R2
1K
R3
4.7K

94. Amplificador de OM

Este circuito foi encontrado numa Radio Electronics de Março de 1987. A revista não existe mais, mas o circuito pode ser montado com BF245 para o FET, BC558 para Q2 e 2N2222 para Q3. O circuito amplifica sinais na faixa de 500 kHz até mais de 5 MHz. A alimentação pode vir de fonte simples, pois o consumo é muito baixo.

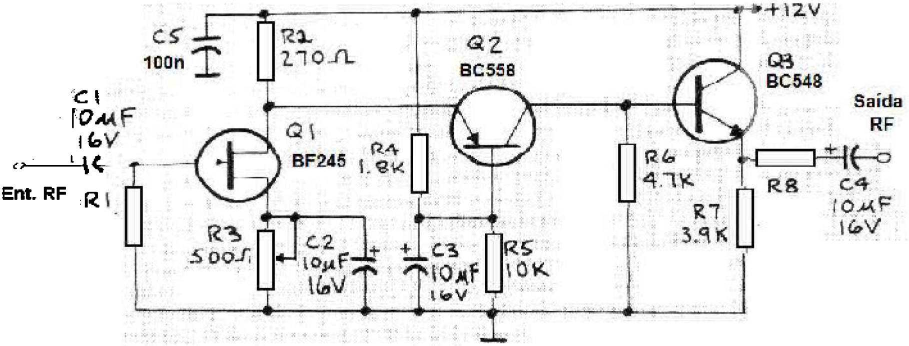
+12V
C5
100n
R2
270Ω
Q2
BC558
Q3
BC548
Saída
RF
C1
10μF
16V
Q1
BF245
R4
1.8K
R6
4.7K
Ent. RF
R1
R8
C4
10μF
16V
R7
3.9K
R3
500Ω
C2
10μF
16V
C3
10μF
16V
R5
10K

95. Reforçador AM

Este reforçador, originalmente projetado para rádios AM de carro, foi encontrado numa revista argentina dos anos 1970. Ele pode ser montado com facilidade com transistores BC548 e é intercalado entre a antena e a entrada de antena do rádio. Sua alimentação de 12 V vem do próprio carro e seu consumo é muito baixo.

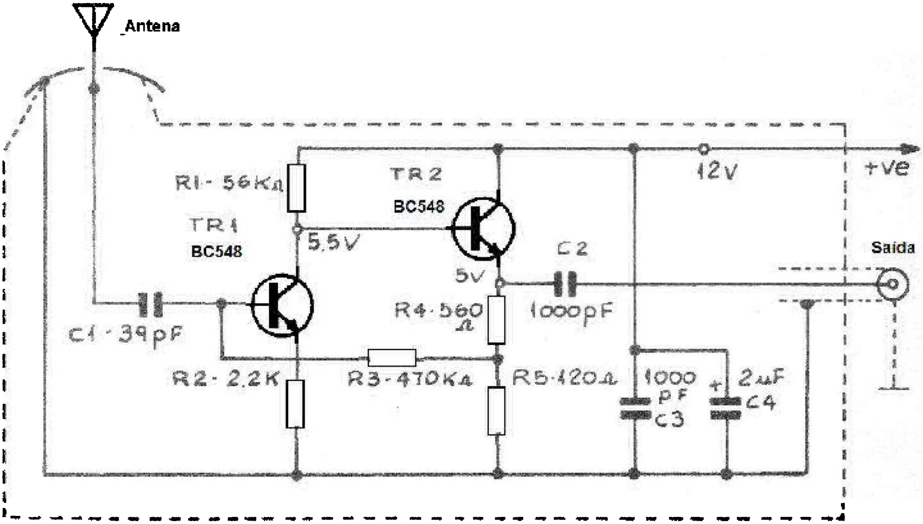
Antena
R1 - 56KΩ
TR2
BC548
TR1
BC548
5.5V
5V
12V
+ve
C2
1000pF
Saida
C1 - 39pF
R4 - 560Ω
R2 - 2.2K
R3 - 470KΩ
R5 - 120Ω
1000 pF C3
2uF C4
+

96.Reforçador AM (2)

Este circuito booster, encontrado numa publicação argentina da década de 1970, foi originalmente projetado para reforçar sinais de AM em rádios de carro. O circuito é do tipo aperiódico, ou seja, sem sintonia, reforçando sinais de toda a faixa. A alimentação é de 9 V e o consumo muito baixo. O transistor original de RF de germânio pode ser substituído pelo BF494 ou BF495.Os cabos de entrada e saída devem ser blindados.

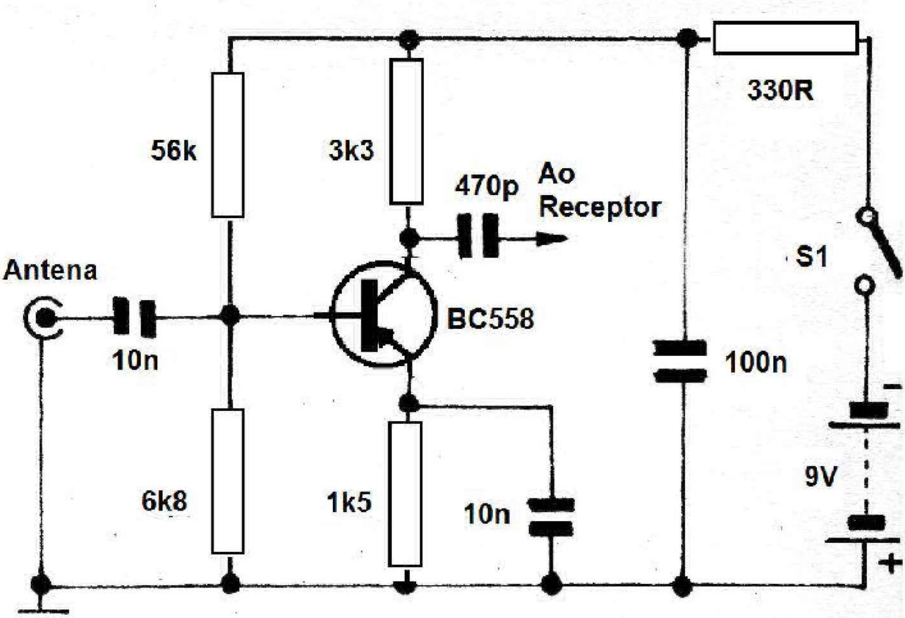
56k
3k3
470p
Ao Receptor
330R
S1
Antena
10n
BC558
100n
9V
6k8
1k5
10n
-
+

97. Medidor de Intensidade de Campo Para 27 MHz

Originalmente este circuito foi projetado para medir a intensidade de sinal de transmissores da faixa do cidadão, 27 MHz ou 11 metros, como também é conhecida. Mas, com alteração da bobina, pode ser usado com transmissores de outras faixas. O circuito foi obtido numa publicação argentina dos anos 1970. A bobina é formada por 4 mais 8 voltas de fio 26 num tubo de 1 cm de diâmetro e o trimmer é comum podendo também ser usado um variável de FM. O transistor pode ser o BC558 em lugar dos originais de germânio e a alimentação é feita com uma ou duas pilhas pequenas. O diodo pode também ser o 1N34.

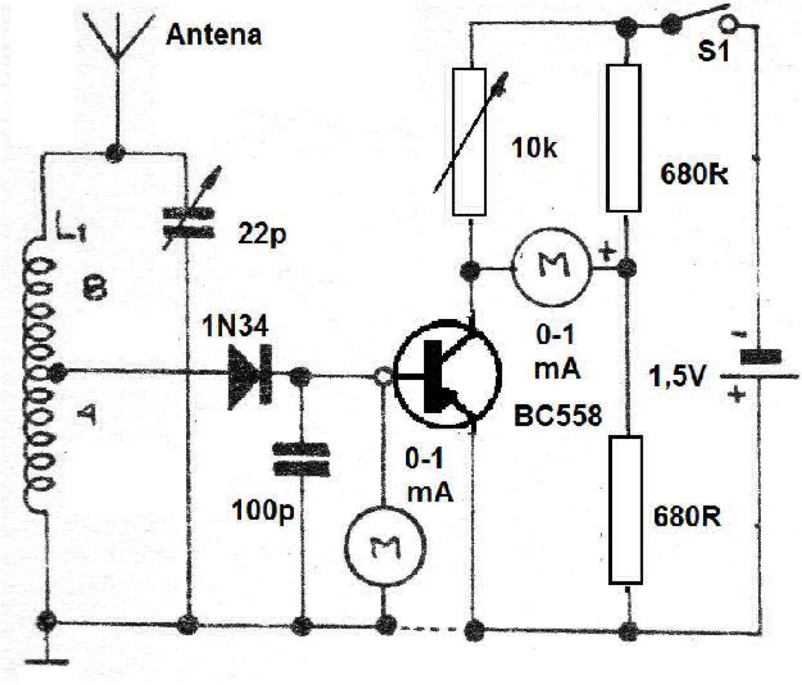

98. Amplificador de Antena

Intercalado entre uma antena externa e a entrada de antena de um receptor este circuito proporciona melhor recepção. O circuito serve para receptores de ondas médias e curtas e com um pouco menos de ganho, para receptores de FM. As conexões de entrada e saída devem ser blindadas. Este circuito é de um livro de Newton C. Braga.

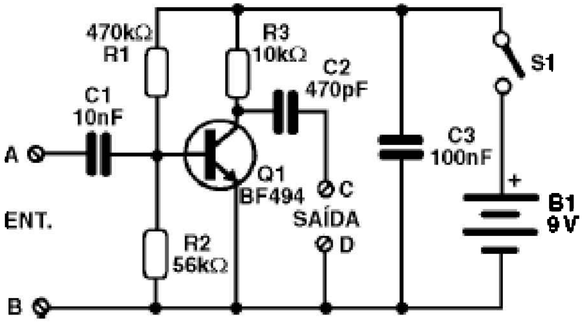
470kΩ
R1
R3
10kΩ
C2
470pF
S1
C1
10nF
A
C3
100nF
Q1
BF494
C
+
B1
9V
ENT.
SAÍDA
R2
56kΩ
D
B

99. Pré-Seletor de Ondas Médias

Muitos ainda gostam de ouvir ondas médias, mas residem em locais de difícil recepção com muitas interferências e baixa intensidade de sinal o que exige o uso de antenas externas. A recepção nestes locais pode ser melhorada por um pré-seletor. A bobina consta de 100 espiras de fio 28 num tubo de papelão ou PVC de 1 polegada com tomadas de 20 em 20 espiras. O capacitor variável é de um rádio antigo de ondas médias. O circuito é ligado entre a antena e terra do receptor. Com uma boa antena externa e um receptor sensível, mais a ajuda deste circuito, estações muito distantes de ondas médias podem ser captadas, principalmente à noite. O circuito é do Fun Projects for the Experimenter, livro de Newton C. Braga.

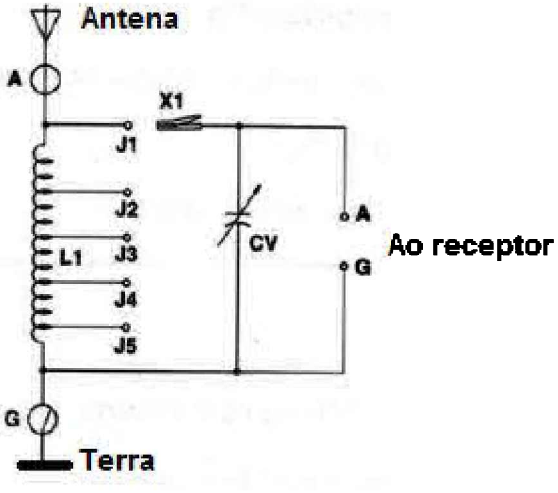

100. Super Booster de OM

Este circuito é de uma Elementary Electronics de 1977. Ele pode ser montado com certa facilidade trocando-se Q2 por um BF494. Q1 pode ser substituído por qualquer MOSFET de comporta dupla. L1 consiste em 100 espiras de fio 28 num bastão de ferrite e L2 20 espiras sobre L1. O capacitor variável pode ser obtido de qualquer rádio AM fora de uso. A saída do circuito é ligada a umas 4 espiras de fio que envolvem a caixa do rádio AM, formando a bobina L2. O consumo é muito baixo, podendo ser usada uma bateria na sua alimentação.

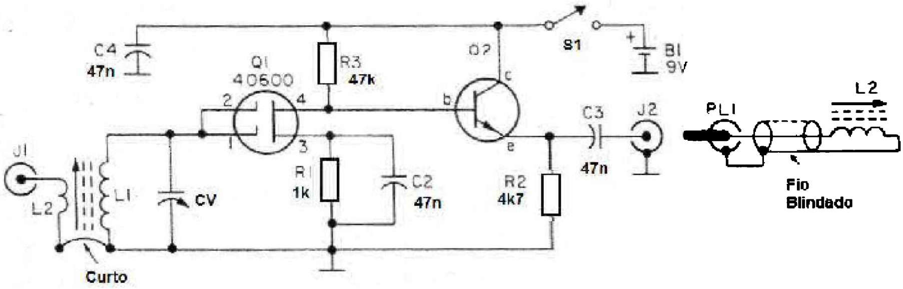
C4
47n
Q1
40600
2
4
1
3
R3
47k
Q2
b
c
e
S1
+
B1
9V
C3
47n
J2
PL1
L2
Fio
Blindado
J1
L2
L1
CV
R1
1k
C2
47n
R2
4k7
Curto

PROBLEMAS CAUSADOS POR TRANSMISSORES DE FM

Recentemente aeroportos de grandes cidades como São Paulo, tiveram sérios problemas de interferências de emissoras clandestinas de FM, que afetando a operação dos sistemas de navegação fizeram com que pousos e manobras importantes de aproximação de aeronaves fossem abortadas.

Uma ação drástica contra essas emissoras está sendo feita pelas autoridades, pois o que está em jogo é a segurança do tráfego aéreo. É melhor prevenir um acidente grave agora do somente tomar as medidas cabíveis depois que acontecer algo.

O Problema

Com o falso rótulo de "Emissoras Comunutárias", muitas pessoas e entidades têm posto no ar emissoras de FM de potências muito acima do que a idéia da legislação inicial que existia sobre o assunto permitia.

Assim, os 5 W de potência máxima não são respeitados da mesma forma que a altura máxima da torres. Temos visto transmissores de até 500 W (!) com torres que chegam a 40 metros, na rota de aproximação de aeronaves no Aeroporto de Guarulhos!

Mas, não é somente a potência que causa problemas. Os transmissores usados não têm qualquer controle de qualidade, normalmente são montagens caseiras precárias que acabam por ter seu sinal espalhando por um espectro que se estende muito além da faixa de FM.

É justamente aí que temos o ponto crítico. A faixa destinada às comunicações entre torre e aeronaves e o sistema de balizamento dos aeroportos opera na faixa que vai dos 108 MHz a 136 MHz, ou seja, justamente acima da faixa de FM.

Assim, um transmissor de FM descalibrado, e não precisa ser muito potente, pode ter sinais espúrios na faixa de aviação, interferindo dos equipamentos de aproximação e comunicação das aeronaves, conforme mostra a figura 1.

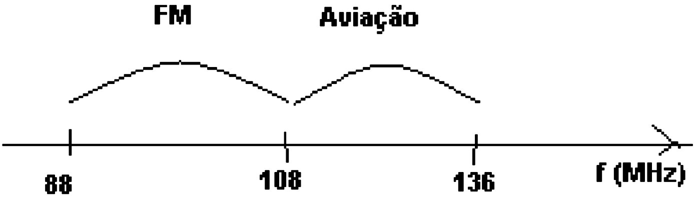

E, para que um transmissor emita sinais espúrios e harmônicas não é preciso muito. Basta que os circuitos das etapas osciladora e de amplificação não sejam bem ajustados para que isso ocorra. A não existência de filtros de saída que eliminem harmônicas e espúrios também causa problemas e o próprio sistema de antenas e cabos mal dimensionados também é causa de problemas.

Montando Transmissores

Muitas vezes publicamos transmissores de FM, alguns até com potências elevadas. Sempre, entretanto, alertamos os leitores para as restrições legais que existem quanto ao seu uso, e principalmente os cuidados que devem ocorrer com sua instalação e montagem.

Em suma, montar e operar transmissores não são brinquedos. Justifica-se a utilização de um transmissor com alguma potência numa localidade isolada, como uma vila ou pequena cidade em que o serviço não existe.

Justifica-se a instalação de um pequeno transmissor num clube ou escola, desde que isso não afete qualquer tipo de serviço de comunicação que eventualmente possa existir no local.

Na verdade, essa é a filosofia básica da Emissora Comunitária que seria prestar um serviço no local em que ele não é disponível. No entanto, no momento em que esse tipo de equipamento pode causar problemas, e principalmente colocar em risco vidas humanas, sua presença não tem qualquer justificativa.

Instalar e operar indevidamente transmissores de FM em locais densamente povoados, podendo também causar interferências em TV e outros serviços, perto de aeroportos onde podem interferir nos equipamentos das aeronaves, é algo que deve ser combatido rigorosamente.

Transmissor sem Alcance

Montei um transmissor de FM, mas o seu alcance não vai além de uns 5 m. Me afasto do receptor de FM e o sinal some. O que está acontecendo?

Provavelmente você está sintonizando um sinal espúrio ou harmônica. Explicamos. O circuito de muitos pequenos transmissores produz um sinal mais forte na frequência para o qual está sintonizado e outros mais fracos, denominados espúrios ou harmônicas em frequências diversas. Estes sinais, por serem mais fracos, não têm alcance.

Se você em lugar de sintonizar o sinal principal do transmissor, sintonizar um destes sinais espúrios, o alcance será muito pequeno e ele some. O que você deve fazer é sintonizar novamente o transmissor e procurar o sinal mais forte. Deve fazer várias tentativas até encontrar este sinal que é o fundamental. Isso também ocorre com transmissores de controle remoto e outros.

QUANDO UMA MONTAGEM NÃO FUNCIONA

Dicas Para os Montadores

A experiência do autor de mais de 30 anos de publicação de montagens, além das aulas que dá mostram que os problemas de funcionamento dos aparelhos montados por leitores e alunos têm sempre as mesmas origens. É claro que não descartamos os casos em que os próprios diagramas ou placas podem conter erros introduzidos no processo de desenho e que às vezes escapam de nossa revisão.

Temos sempre o cuidado de montar os aparelhos publicados, muitos dos quais muitas vezes não funcionam sozinhos, pois suas configurações podem fazer parte de mais um projeto, o que nem sempre é compreendido por quem vê algo montado, aparentemente "igualzinho" ao original, mas que não funciona. Se isso acontecer, o que fazer? O que pode ter ocorrido?

Analisemos os casos mais comuns:

Confira sempre sua montagem pelo diagrama e depois pelo desenho da placa de circuito impresso.

Nos projetos que damos em nossas publicações e no site é comum colocarmos junto ao diagrama os valores dos componentes que são repetidos na lista de material com informações adicionais como tensão de trabalho para os capacitores, dissipação e código de cores para os resistores, etc.

Assim, se houver uma discrepância entre os valores o leitor pode ter um ponto de partida para uma eventual falha. Veja depois se os resistores e capacitores estão com os valores corretos na montagem.

É muito comum que um resistor de 1k seja colocado em lugar de 10k numa montagem. Marrom, preto, vermelho pode ser facilmente confundido com marrom, preto, laranja, principalmente se o vermelho do primeiro estiver desbotado. Para os capacitores, o problema é mais grave devido aos códigos utilizados.

É comum que se troque os valores de componentes como, por exemplo usar 4k7 = 4 700 pF em lugar de 4K7 = 4,7 pF numa montagem, o que a leva a não funcionar. Na figura 1 mostramos alguns códigos de capacitores cerâmicos que causam confusões.

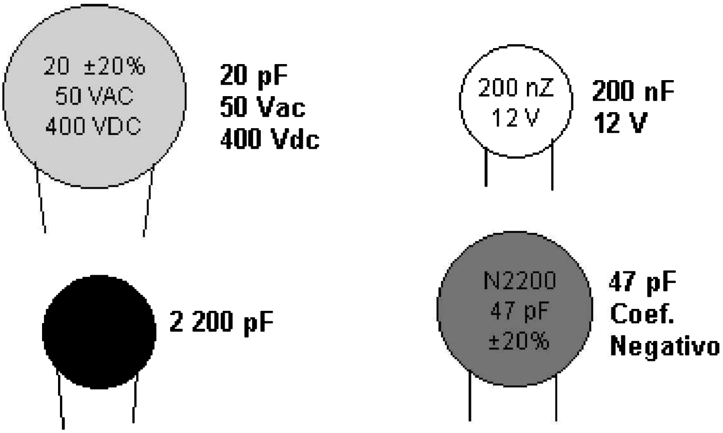

Figura 1 – Códigos de capacitores cerâmicos.

Conferir com uma lente os valores dos componentes numa montagem é fundamental, antes mesmo de se alimentar o circuito para um teste de funcionamento.

Tipos e especificações de componentes

O tipo de componente indicado para uma aplicação é fundamental para o seu funcionamento, principalmente nos casos mais críticos. Isso ocorre, por exemplo, nos circuitos de altas freqüências (transmissores, receptores, osciladores, etc.) onde existem funções em que o uso de capacitores cerâmicos é obrigatório. Se, na função indicada para um capacitor cerâmico for utilizado um de poliéster, mesmo que de mesmo valor que o pedido, o circuito poderá não funcionar.

Para os transistores, diodos, SCRs e outros semicondutores em alguns casos pode ser indicado um sufixo que significa uma característica adicional. Por exemplo, para os transistores a última letra pode significa tanto o ganho como a tensão de trabalho, e ela deve ser observada quando se adquirir o componente. Por exemplo, o TIP31A é diferente do TIP31 porque o "A" significa que este transistor suporta uma tensão maior, conforme mostra a figura 2.

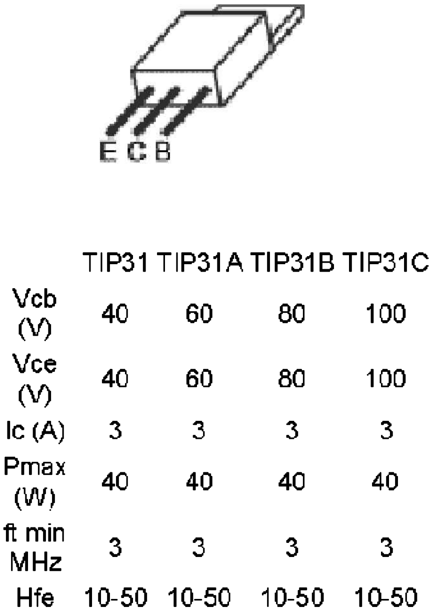

	TIP31	TIP31A	TIP31B	TIP31C
Vcb (V)	40	60	80	100
Vce (V)	40	60	80	100
Ic (A)	3	3	3	3
Pmax (W)	40	40	40	40
ft min MHz	3	3	3	3
Hfe	10-50	10-50	10-50	10-50

Figura 2 – O sufixo está associado às características do componente.

Da mesma forma, o BC548B tem um ganho maior que o BC548.

Se usarmos um BC548 comum onde se recomenda um BC548B poderemos ter problemas de funcionamento no aparelho que está sendo montado. Existem ainda os casos em que são especificadas tensões e corrente mínimas. Para os capacitores é comum indicarmos a tensão de trabalho. Na figura 3 temos alguns capacitores com as indicações de suas tensões de trabalho.

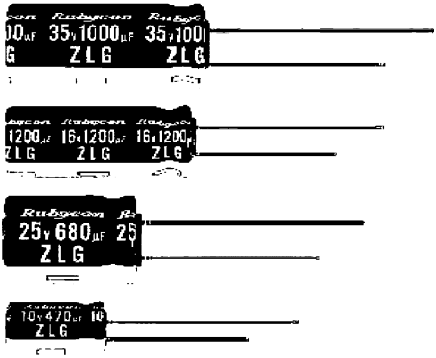

Figura 3 – Eletrolíticos com tensões de trabalho indicadas.

Esta é a tensão máxima que eles suportam e não a tensão que neles aparece quando em funcionamento. Assim, se num circuito temos tensões até uns 12 V, é comum utilizarmos capacitores de pelo menos 16 V (nos locais em que os 12 V podem aparecer), dando assim uma tolerância para funcionamento seguro.

Se um capacitor de menor tensão for utilizado, ele pode ter seu dielétrico rompido "queimando" e impedindo o funcionamento do aparelho. Para as correntes vale o mesmo. Num circuito em que tenhamos uma corrente menor do que 1 A podemos utilizar um diodo que suporte 1 A ou mais. Também é preciso estar atento às dissipações, ou seja, à quantidade de calor que o componente pode dissipar.

Se utilizarmos um resistor de 1/8 W onde se exige um de 1 W, ele esquentará demais e poderá queimar. Os resistores têm suas dissipações associadas ao seu tamanho, conforme mostra a figura 4.

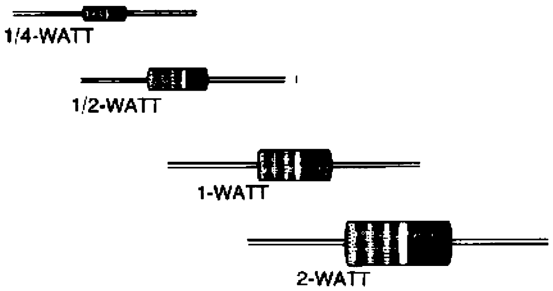

Figura 4 – Resistores maiores dissipam mais potência

Nesta relação devemos ainda incluir os componentes que devem ser "montados" em casa. É o caso das bobinas que são elementos extremamente críticos nos transmissores e receptores. Se o número de espiras, separação entre espiras, diâmetro e características do núcleo não forem obedecidas o circuito não vai oscilar na freqüência correta.

Posição de componentes polarizados

Diodos, transistores, capacitores eletrolíticos, SCRs e circuitos integrados possuem posições corretas para sua montagem. Se forem invertidos o aparelho certamente não funcionará e mais do que isso: o próprio componente pode queimar. Normalmente, nos desenhos das placas de circuito impresso são indicadas as posições dos componentes, mas para os casos em que o próprio leitor projeta a placa, ele deve conhecer o formato e a disposição dos terminais para saber como posicionar cada componente. Na figura 5 temos o modo como as posições dos componentes polarizados são indicadas.

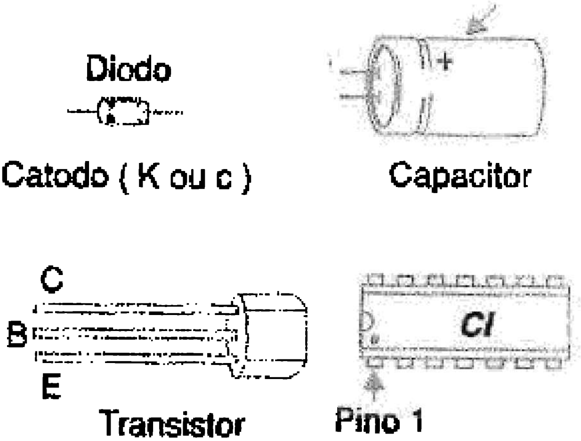

Figura 5 – Indicações de componentes polarizados ou com posição certa de montagem.

Soldas frias

Esta é, sem dúvida alguma, a maior causa de não funcionamento de aparelhos eletrônicos. Chamamos de "solda fria" aquela que "não pega". Aparentemente o componente está soldado, mas se examinarmos com uma lente veremos que ele está solto, que a solda não aderiu aos seus terminais ou à placa de circuito impresso. Resultado: a corrente não pode passar e com isso o aparelho não funciona.

O caso mais comum ocorre quando soldamos fios esmaltados de bobinas numa placa. Estes fios precisam ser raspados antes de serem soldados, pois a fina capa de esmalte que os recobre impede a aderência da solda. Na figura 6 temos alguns casos de soldas frias.

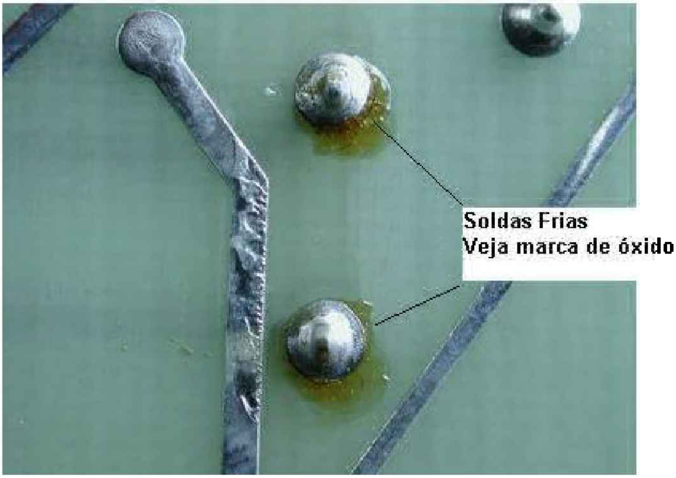

Figura 6 – Soldas frias (temperatura insuficiente do soldador ou placa suja)

Desenho de placa

Ao se elaborar uma placa de circuito impresso, diversos problemas podem ocorrer.

O primeiro é a interrupção e curto entre trilhas quando a placa é corroída. As interrupções podem ser reparadas com uma ponte de soldas e os curtos eliminados com um estilete. Na figura 7 temos algumas imperfeições na elaboração de placas de circuito impresso.

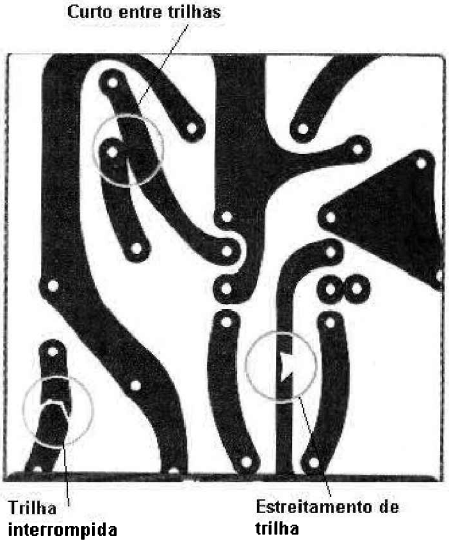

Figura 7 – Imperfeições de uma placa de circuito impresso.

De qualquer forma, antes de fazer a montagem, examine com uma lente todas as trilhas da placa para verificar se estão perfeitas. Outro problema que pode ocorrer é o projeto mal feito de uma placa. Trilhas "esquecidas" podem significar a falta de uma conexão entre componentes que vai impedir o funcionamento do aparelho.

Confira a montagem da placa com o diagrama e isso pelo menos duas vezes!

Existem também as montagens que são críticas quanto a disposição e largura das trilhas. As trilhas que trabalham com sinais de altas freqüências devem ser curtas e não podem ter curvas em ângulos retos (quinas). As trilhas que trabalham com correntes intensas devem ser largas.

Adota-se a regra geral de que para cada ampère de corrente a trilha deve ter 1 mm de largura. Numa fonte de 3 A, as trilhas que conduzem a corrente principal deve ter pelo menos 3 mm de largura.

Tolerâncias

Este é o fator que influi enormemente no funcionamento de certos aparelhos, principalmente os mais críticos. O controle sobre a ação das tolerâncias dos componentes é algo bastante difícil de fazer, principalmente quando os componentes utilizados são adquiridos em lojas comuns ou aproveitados de outros equipamentos.

O que ocorre é que, quando se projeta um circuito ou se faz a montagem de um protótipo utilizam-se componentes comuns que podem ter tolerâncias elevadas.

Por exemplo, os resistores podem ter tolerâncias de 5% a 20% enquanto que os capacitores eletrolíticos chegam a ter tolerâncias de até 40%. Isso significa que se somarmos os desvios dos valores de todos os resistores de um circuito, se por uma casualidade todos tiverem desvios no mesmo sentido o circuito poderá ser levado a uma condição em que o funcionamento se torna impossível.

Para os transistores o problema pode ser ainda mais grave. Um transistor BC548 pode ter ganhos entre 110 e 800. Se, casualmente o montador de um projeto experimental utilizar um transistor que esteja com ganho 500, portanto dentro da faixa e conseguir fazê-lo funcionar de um modo precário.

Outro montador que utilize o mesmo transistor, mas que esteja com ganho 200 (dentro ainda da faixa), não conseguirá fazer o mesmo circuito funcionar!

Existe ainda um fator agravante para o uso de transistores: um transistor de uso geral pode ser "recarimbado". Existem fornecedores desonestos que compram lotes de transistores que não passam nos testes de fábrica por estarem com suas características abaixo das exigidas e os recarimbam com designações de tipos de uso geral como BC548, 2N2222, 2N3904, etc. e os vendem.

Nas aplicações menos críticas eles funcionam, mas se for exigido um ganho maior o comprador "dança". O correto (normalmente fazemos isso nos nossos projetos) é calcular o circuito para o transistor na condição mínima de ganho.

De qualquer forma, diante de um circuito que não oscile, que tenha baixo ganho ou ainda apresente instabilidades uma possibilidade de se sanar o problema é a troca do transistor, mesmo que por um igual mas que pode estar com maior ganho...

Muitos multímetros possuem teste de transistores com indicação do ganho e existem também os provadores de transistores. Na figura 8 temos um provador de transistores.

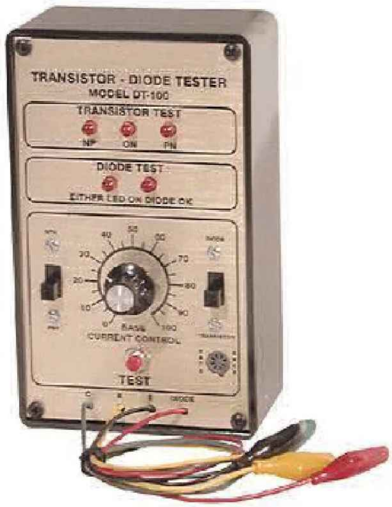
TRANSISTOR - DIODE TESTER
MODEL DT-100
TRANSISTOR TEST
DIODE TEST
EITHER LED ON DIODE OK
BASE CURRENT CONTROL
TEST

Figura 8 – Um teste de transistor e diodos de baixo custo.

Será interessante ter um deles.

Fornecedor confiável, componentes confiáveis

Evidentemente, problemas de fornecimento de componentes com características diferentes das exigidas, de má procedência ou mesmo "recarimbados", podem ser evitados com um fornecedor de confiança.

Ajustes

Muitos projetos precisam de ajustes críticos para serem colocados no ponto de funcionamento. Se bem que em revistas destinadas a leitores que não tenham recursos para aquisição de instrumentos complicados a maioria dos ajustes possa ser feita sem instrumentos ou no máximo com um multímetro, existem certos cuidados a serem considerados.

O principal problema que ocorre com ajustes estão nos circuitos osciladores e de alta freqüência. É comum que o montador ajuste um circuito não para sua freqüência correta mas para uma freqüência múltipla, ou seja, uma harmônica. Quando isso acontece o circuito funciona, mas seu rendimento é bem menor que o esperado.

Num transmissor, por exemplo, o alcance ficará reduzido e o montador terá dificuldades em saber por quê. Num reconhecedor de tom de um controle remoto, por exemplo, o circuito se torna instável e começa a responder comandos indevidos, conforme mostra a figura 9.

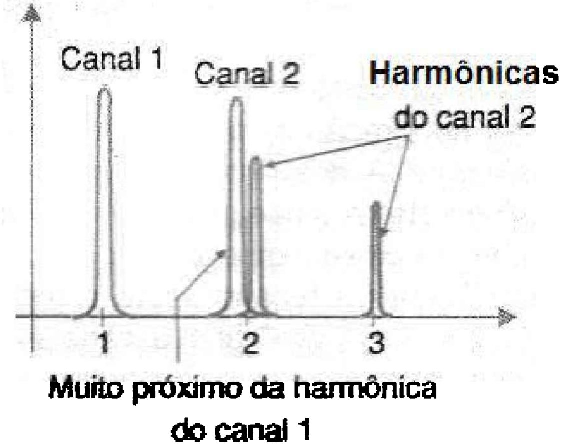

Figura 9 – Ajustes indevidos em reconhecedores de tom.

Finalmente temos até a perda de seletividade dos circuitos que começam a "misturar" estações. O melhor procedimento neste caso é tentar o reajuste em outra posição do elemento de controle (trimmer, trimpot, etc.).

Configurações semelhantes

Os circuitos eletrônicos seguem configurações que têm sempre a mesma disposição de componentes. Transistores, por exemplo, sempre têm elementos de carga (resistores, bobinas, etc.), resistores de polarização e capacitores de acoplamento e desacoplamentos, conforme mostra a figura 10.

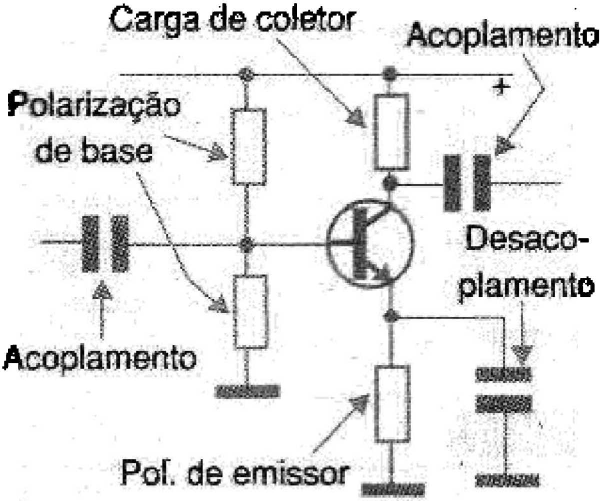

Figura 10 – Configuração comum.

Procure um circuito semelhante e veja se os valores estão compatíveis, se é daquele modo que a ligação deve ser feita, a polaridade da alimentação, etc. Se a configuração empregar um circuito integrado popular como o 555, verifique em outros artigos ou esquemas se as ligações são semelhantes para aquela função. Elas podem servir de referência, inclusive dando uma ordem de grandeza dos componentes utilizados.

Desconfiômetro

Uma outra "ferramenta" que é muito importante quando se monta algum circuito eletrônico e ele não funciona é o "desconfiômetro".

Se alguma coisa vai mal, desconfie de tudo.

Lembre-se que, pela Lei de Murphy, justamente aquilo que tem a menor possibilidade de falhar ou de estar errado é justamente o que está causando o problema na sua montagem...

Verifique tudo! Mesmo o que você tem certeza de que não é a causa do problema...

Procurando ajuda

Evidentemente, o primeiro lugar em que o leitor pensa para poder pedir ajuda é a nossa equipe técnica. No entanto, para que possamos ajuda sem ver a montagem (o que para nós é impossível – pois não temos condições de examinar um a um os projetos que eventualmente nos sejam enviados) é preciso que o leitor saiba explicar o que fez.

Muitos leitores nos enviam longas cartas descrevendo de forma imprecisa o que acontece com seus aparelhos o que dificulta bastante a possibilidade de darmos qualquer ajuda. Para que o leitor tenha sua consulta efetivamente respondida e de forma que possa ajudar proceda da seguinte forma:

Seja objetivo na sua consulta – ela deve ser curta e direta para facilitar nosso exame.

Indique sempre em que ponto do site se encontra o artigo que montou. De preferência dê o nome (ART003, PN002, etc.)

Explique de forma clara o que acontece com ele: não funciona, não dá ajuste, não tem alcance.

Indique se usou algum componente com especificação diferente da indicada na lista de materiais.

Indique se alterou valores de componentes

Se algum componente aquecer ou queimar indique qual.

Se tem um multímetro indique as tensões que pode medir nos principais pontos do circuito.

É importante também alertar que a nossa correspondência é volumosa.

Assim, não espere que uma consulta enviada pela manhã já tenha sua resposta à tarde. Precisamos de tempo para analisar o problema, localizar o artigo e, é lógico atender os leitores que estão na sua frente.

Também precisamos informar que não temos condições de alterar projetos publicados para atender às necessidades do leitor. "Como modificar o amplificador X, artigo Y para que ele se transforme numa máquina de fazer pipocas?"

Também não podemos dar opiniões sobre equipamentos comerciais ou indicar alterações. Isso ocorre porque alterações e modificações nem sempre são simples e possíveis. Mesmo quando são possíveis não podemos indicar alguma coisa ao leitor sem experimentar ou ter certeza de que funcione!

O RÁDIO

Houve época em que toda a eletrônica se traduzia numa palavra ”rádio”. De fato, naqueles tempos, o único aparelho que podia ser denominado eletrônico, e de uso comum, era o rádio. Presente em muitos lares permitia o “milagre”de ouvir vozes de pessoas distantes.

Com a evolução da eletrônica, novos equipamentos surgiram como os amplificadores para fonógrafos, gravadores, a TV e depois tudo que hoje conhecemos, uma infinidade de aplicativos a nossa disposição.

Assim, se quisermos ter uma visão da história da eletrônica, nada melhor do que usar o rádio como referência. A história do rádio está intimamente ligada a história da evolução de novas tecnologias.

Neste artigo tratamos um pouco do rádio, explicando sua origem, os diversos tipos de circuitos e algumas características importantes para projetos. Daremos até alguns circuitos práticos simples para os que desejarem fazer experimentos, ou demonstrações com finalidades didáticas.

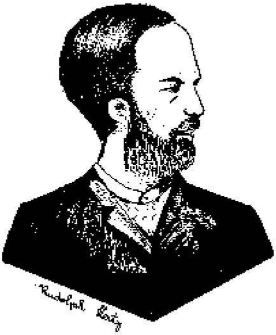

Rudolf Hertz (desenho de Dyrce Braga)

O Início

Maxwell, através de equações matemáticas, previu a existência de ondas eletromagnéticas, da mesma natureza que as ondas de luz, que podiam se propagar no espaço com a velocidade próxima de 300 000 km/s, mas de freqüência muito mais baixa.

Lendo os trabalhos de Maxwell, um cientista alemão chamado H. R. Hertz, em 1887, realizou as primeiras experiências práticas que provaram a existência de tais ondas.

A prova foi simples, consistindo na montagem do primeiro transmissor do primeiro receptor de rádio. Produzindo faíscas elétricas num canto do laboratório, estas faíscas geravam ondas eletromagnéticas que, provocavam o aparecimento de outras num dispositivo colocado do outro lado do laboratório.

Esta experiência também serviu para a elaboração da primeira antena que consistia em placas de metal.

Guglielmo Marconi, lendo os trabalhos publicados por Hertz, relatando suas experiências, percebeu que estas ondas poderiam, ser usadas para levar mensagens, trabalhando a partir de então no aperfeiçoamento da descoberta.

Marconi desenvolveu então dispositivos capazes de gerar e receber ondas eletromagnéticas.

A partir de 1895, com apenas 21 anos de idade, Marconi deu início a uma série de experiências que culminaram com a transmissão com êxito de mensagens a navios situados a 20 km de distância da costa.

Depois, Marconi criou um sistema capaz de transmitir direcionalmente os sinais eletromagnéticos e, finalmente em 1901, um sistema capaz de transmitir sinais entre a Inglaterra e a Terra Nova, sendo essa a primeira transmissão através do Atlântico.

A primeira transmissão de radiodifusão documentada ocorreu em 1906, tendo sido realizada por R. Fessenden nos Estados Unidos.

Na véspera de natal de 1906, usando um alternador de radiofreqüência, que operava em 50 kHz, construído pela General Electric, foi realizada a primeira transmissão de radiodifusão.

O microfone usado para modular este primeiro transmissor era ligado diretamente à antena, controlando assim toda a potência aplicada. O resultado era que o microfone precisava ser refrigerado a água, tal o calor desenvolvido.

Landel de Moura

Da mesma forma que existe uma grande controvérsia em relação à invenção do avião, caso em que os americanos defendem os irmão Wright e nós Santos Dumont, podemos dizer que para o rádio existe também uma boa discussão sobre o assunto.

Os russos defendem Popov enquanto que "oficialmente", Marconi é o inventor do rádio. No entanto, existem provas de que muitos pesquisadores transmitiram e receberam ondas de rádio antes de Marconi.

É o caso do padre brasileiro Roberto Landel de Moura que enviou sinais de rádio entre pontos diferentes da cidade de São Paulo, antes de Marconi e também teria transmitido a voz e imagens na mesma época sem o devido reconhecimento.

Roberto Landel de Moura – Inventor do rádio

Somente agora suas descobertas estão sendo revistas com a atribuição do devido valor que possuem. Sugerimos aos leitores interessados que visitem o site de Luiz Netto sobre Landel de Moura em http://www.rlandell.hpg.ig.com.br/ ou ainda que leiam o livro de Hamilton Almeida, Padre Landel de Moura: um herói sem glória. O brasileiro que inventou o rádio, a TV, o teletipo...” (Editora Record).

Os Circuitos Receptores

A partir do primeiro receptor de Hertz, que consistia simplesmente numa espira com duas esferas separadas por uma distância muito pequena, onde era possível observar as faíscas, foi criado um dispositivo denominado “coesor”. O coesor era um tubinho cheio de limalha de metal, onde a presença de sinais de rádio fazia saltar faíscas, tornando-o condutor.

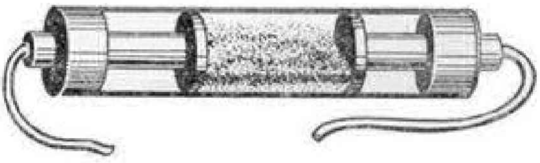

Figura 1 – o coesor de Branley, primeiro detector de sinais de rádio

Depois disso, diversos dispositivos foram criados para detectar as ondas de rádio, chegando às técnicas avançadas que hoje são utilizadas.

Rádio de Galena

Um dos primeiros tipos de rádio que existiu não utilizada nenhum dos modernos dispositivos eletrônicos que conhecemos como transistores ou circuitos integrados. Estes rádios tinham uma estrutura bastante simples, conforme mostra a figura 2.

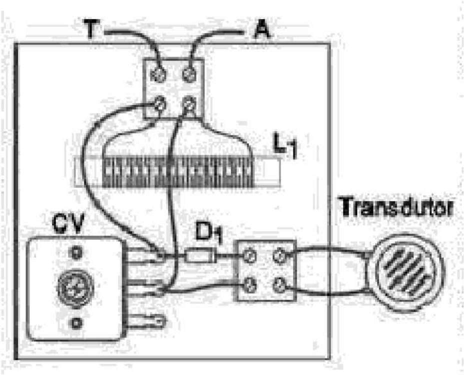

Figura 2 – Rádio de galena simples usando um diodo (D1) em lugar do cristal de galena.

Uma enorme antena (A), consistindo em fio estendido com comprimento de 10 a 50 metros de comprimento, captava o máximo de energia das ondas eletromagnéticas emitidas pela estação. Estas ondas induziam na antena correntes de alta freqüência que então eram levadas a um circuito seletor. O seletor mais simples que podemos descrever consiste numa bobina e num capacitor variável, ligados em paralelo, conforme mostra a figura 3.

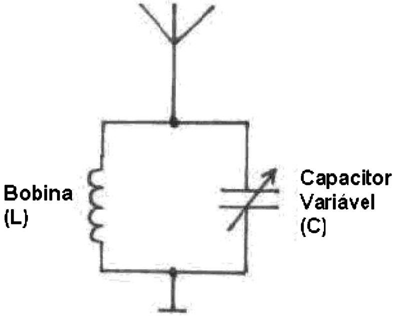

Figura 3 – Circuito de sintonia formado por uma bobina e um capacitor

Este circuito até hoje é usado na maioria dos receptores de rádio. O número de voltas de fio da bobina e a quantidade de placas do capacitor determinam a faixa de freqüências das estações que podem ser selecionadas.

Para o caso das estações de AM (amplitude modulada) de ondas médias onde as estações têm freqüência entre 530 e 1605 kHz,usando capacitores variáveis entre 270 e 465 pF, o número de voltas da bobina estará tipicamente entre 80 e 120.

Este circuito tem a propriedade de deixar passar para a terra as correntes de todas as freqüências captadas, exceto as da freqüência da estação que desejamos ouvir.

O sinal separado, da estação selecionada, é então levado a um detector de envoltória ou simplesmente detector. O detector nada mais é do que um retificador que conduz a corrente num único sentido de modo que, através de uma filtragem (filtro passa baixas), possamos separar a corrente de baixa freqüência da modulação (som) da corrente de alta freqüência que a transporta (portadora).

Isso ocorre porque o som do microfone na emissora é aplicado à "onda" que lhe transporta. A onda eletromagnética é, portanto apenas um meio de transporte para os sinais de freqüências menores que correspondem aos sons, conforme mostra a figura 4.

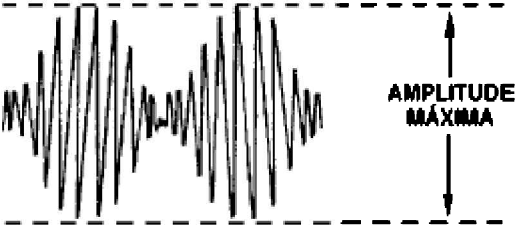

Figura 4 – Sinal de alta frequência transportando sinal de áudio (AM)

Atualmente temos componentes próprios que podem ser usados como detectores, como os diodos semicondutores. No entanto, antigamente, as coisas eram mais difíceis.

O detector tinha de ser fabricado com um cristal de galena, uma espécie de óxido de chumbo, que era montado numa base de material condutor. Um fio extremamente fino, chamado "bigode de gato", era usado para encontrar os pontos sensíveis do cristal.

O operador do rádio deveria, com muita calma e habilidade, encostar o bigode de gato em diversos pontos do cristal, até encontrar o "ponto sensível" que possibilitaria de detecção dos sinais de rádio. Sem dúvida, ouvir rádioi exigia habilidade naqueles tempos!

A corrente que temos depois do diodo já é de baixa freqüência, devendo ser filtrada antes de ser levada a um fone. O fone nada mais é do que um reprodutor de som.

Formado por uma bobina de fio fino, com a passagem da corrente é criado um campo magnético que, atuando sobre uma placa de metal chamada diafragma, a faz vibrar. Estas vibrações são transferidas para o ar na forma do som originalmente enviado a partir da estação, conforme mostra a figura 5.

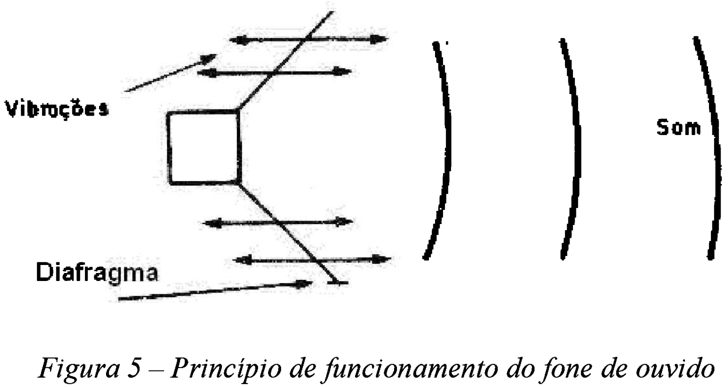

Figura 5 – Princípio de funcionamento do fone de ouvido

Veja que o princípio de funcionamento desse receptor é muito simples, mas apresenta muitos inconvenientes: toda a energia que vai para o fone deve ser captada pela antena. Assim, o volume do som depende da eficiência a antena na captação dos sinais.

Mesmo estações fortes ou próximas só podem ser ouvidas com o fone muito próximo do ouvido. O uso de um alto-falante é praticamente inviável com esse tipo de receptor.

Para os leitores que desejam experimentar com esse tipo de receptor, damos um circuito prático, onde a única modernização é a troca do cristal de galena (difícil de obter) por um diodo de germânio de qualquer tipo.

Esse tipo de componente já tem um "bigode de gato" interno soldado no ponto de máxima sensibilidade, evitando assim o inconveniente de se precisar encontrar o ponto sensível. O circuito é mostrado na figura 6.

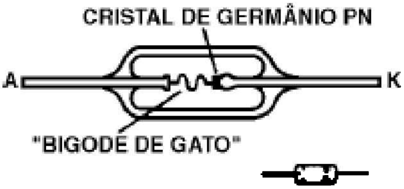

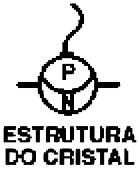

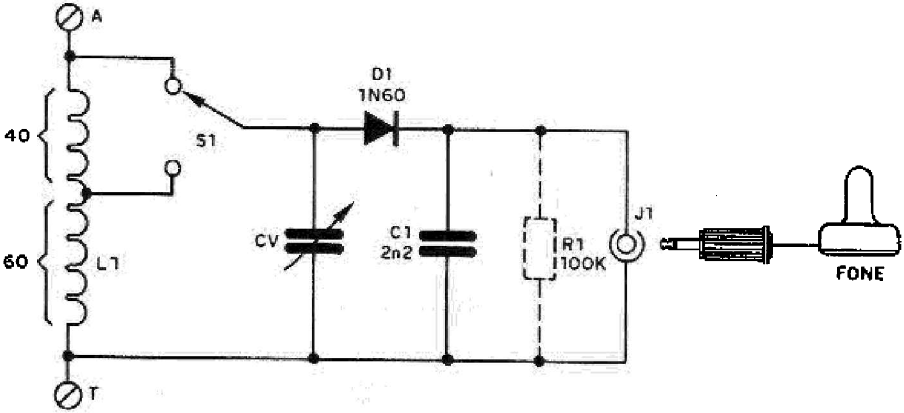

Figura 6 – O diodo de germânio e um rádio de galena
(veja mais em ART031 no site www.newtoncbraga.com.br)

A bobina é enrolada num cabo de vassoura ou tubo de PVC de aproximadamente 2,5 cm de diâmetro. Ela consta de 100 voltas de fio esmaltado 28 ou 26 (40 + 60). O capacitor variável CV é aproveitado de qualquer rádio antigo fora de uso e o fone deve ser de cristal ou magnético de alta impedância.

Sugerimos o uso de uma cápsula piezoelétrica. Não devem ser usados fones de walkman ou radinhos modernos que são de baixa impedância.

A antena deve ter pelo menos 10 metros de comprimento, consistindo num fio comum encapado. A ligação à terra pode ser feita num cano de água ou qualquer objeto de metal acessível de grande tamanho como a esquadria de uma janela.

A Válvula, o Transistor e a Amplificação Direta

Thomas Alva Edison havia descoberto um importante efeito em suas lâmpadas: quando uma placa adicional era colocada no interior da lâmpada, uma corrente podia ser detectada nesse elemento. Essa corrente, conforme se verificou posteriormente, era formada por elétrons que deixavam o filamento aquecido, conforme mostra a figura 7.

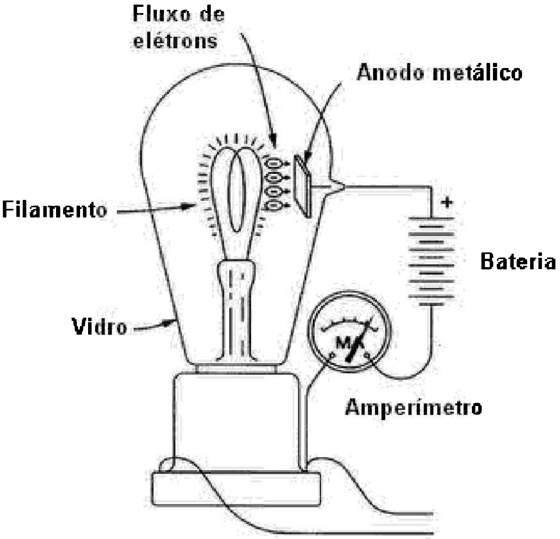

Figura 7 – A válvula diodo

J. A. Fleming, um inglês que visitava Edison justamente quando ele fez essa descoberta, observou que o dispositivo em questão funcionava como um diodo, deixando a corrente circular apenas num sentido.

Posteriormente, trabalhando nessa mesma válvula, o americano Lee de Forest, em 1906, colocou uma espécie de grade entre o filamento aquecido e a placa. Aplicando tensões nesta grade ele podia controlar a corrente que circulava entre o filamento e a placa, conforme mostra a figura 8.

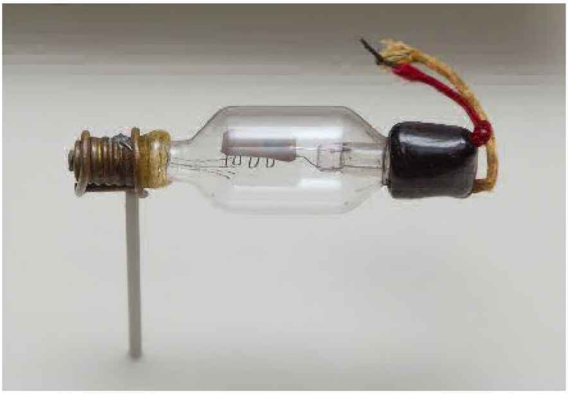

Figura 8 – Uma válvula triodo de 1906

Essa configuração, denominada válvula triodo, poderia ser usada num circuito conforme mostra a figura 9, para amplificar um sinal de rádio detectado e dessa forma obter muito maior volume num fone de ouvido. Um rádio de maior sensibilidade poderia ser obtido com essa configuração.

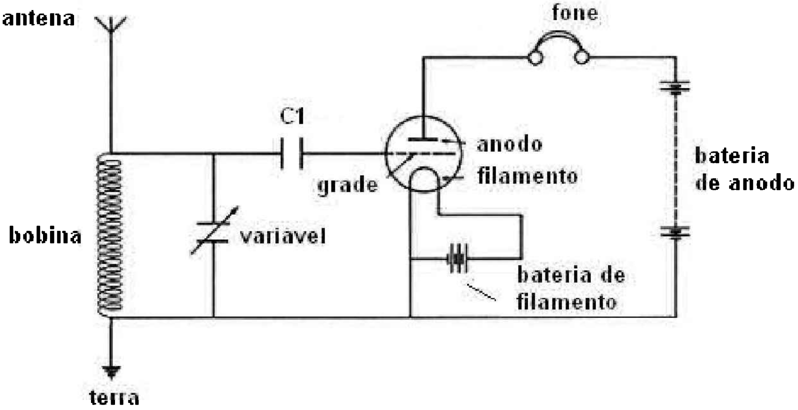
antena
fone
C1
anodo
filamento
grade
bateria
de anodo
bobina
variavel
bateria de
filamento
terra

Figura 9 – Rádio com válvula triodo (audion)

Esse tipo de rádio, em que o sinal detectado (retirado do diodo) era aplicado diretamente a uma válvula amplificadora, recebeu o nome de receptor de "amplificação direta".

O transistor, que foi inventado somente em 1948, na verdade funciona como um amplificador de sinais semelhante a válvula triodo, com a vantagem de que não usa um filamento.

Pelo fato das correntes circularem num transistor através de um meio sólido (o material semicondutor), dizemos que se trata de um dispositivo de estado sólido, que substituem as válvulas com vantagens, quer seja pelo seu tamanho diminuto, quer seja pela pouca energia que precisam para funcionar.

Na figura 10 temos um receptor de amplificação direta, equivalente à configuração valvulada que mostramos na figura 9, mas que usa um transistor.

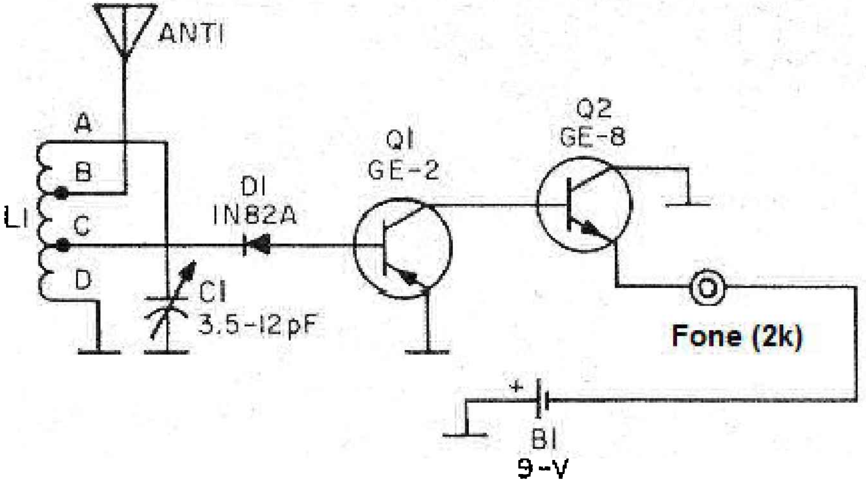

Se o leitor quiser montá-lo, a bobina é formada por 80 espiras de fio esmaltado 28 ou mais fino, com tomada na 30ª e 50ª espira, e o capacitor variável pode ser aproveitado de qualquer rádio antigo.

O diodo é de germânio e o fone deve ser de alta impedância ou de cristal com um resistor em paralelo. A antena deve ter pelo menos 5 metros de comprimento.

Entretanto, já nos tempos das válvulas verificou-se que a recepção dos sinais poderia ser melhorada através do uso de diversos artifícios, que também podem ser usados em circuitos transistorizados (com finalidades experimentais).

No caso dos transistores, o que ocorre é que quando eles surgiram, as técnicas de recepção já estavam suficientemente desenvolvidas para se adotar a melhor. Assim, as outras ficaram apenas como curiosidades históricas.

Receptores Reflex

Verifica-se que uma válvula triodo tanto podem amplificar os sinais detectados como os próprios sinais de alta frequência captados pela antena. Num receptor reflex utiliza-se a mesma válvula para amplificar duas vezes o sinal: na primeira vez o sinal amplificado é o de alta frequência, vindo do circuito de sintonia.

Logo depois da amplificação, este sinal é detectado, voltando então a componente de baixa frequência à mesma válvula, onde recebe nova amplificação. Daí ele pode então ser aplicado ao fone de ouvido. Na figura 11 temos um circuito de um receptor desse tipo.

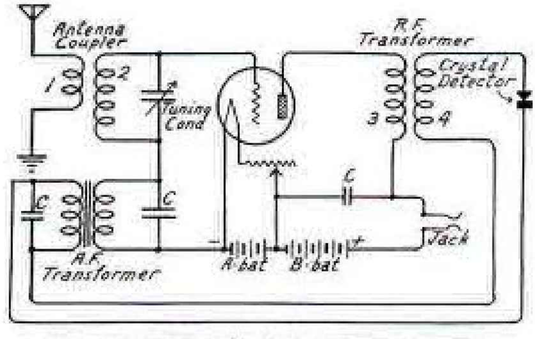
Antenna Coupler
1
2
Tuning Cond
R.F. Transformer
Crystal Detector
3
4
C
C
C
A.F. Transformer
A bat
B bat
Jack

Figura 11 – Receptor reflex valvulado

O mesmo circuito na versão transistorizada é mostrado na figura 12.

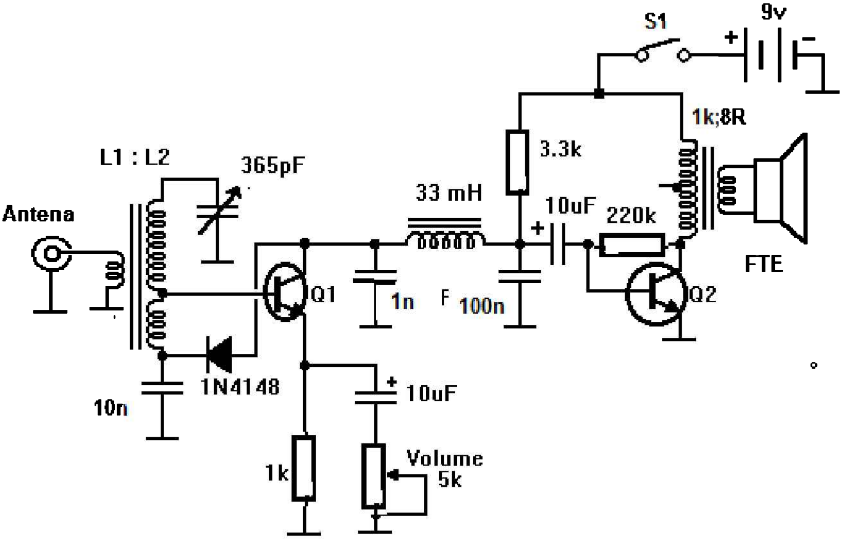
S1
9v
1k;8R
3.3k
L1 : L2
365pF
33 mH
10uF
220k
Antena
Q1
1n
F
100n
Q2
FTE
10n
1N4148
10uF
1k
Volume
5k

Figura 12 – Um receptor reflex transistorizado.

Receptores Regenerativos

Outra técnica interessante usada na recepção de sinais de rádio consiste na regeneração. O que se faz é amplificar o sinal uma vez e depois "jogá-lo" de volta à entrada do mesmo amplificador, para que ele receba nova amplificação. Uma mesma válvula (ou transistor) tem então sua capacidade de amplificação multiplicada, com resultados bastante interessantes. Na figura 13 temos um rádio desse tipo usando uma válvula.

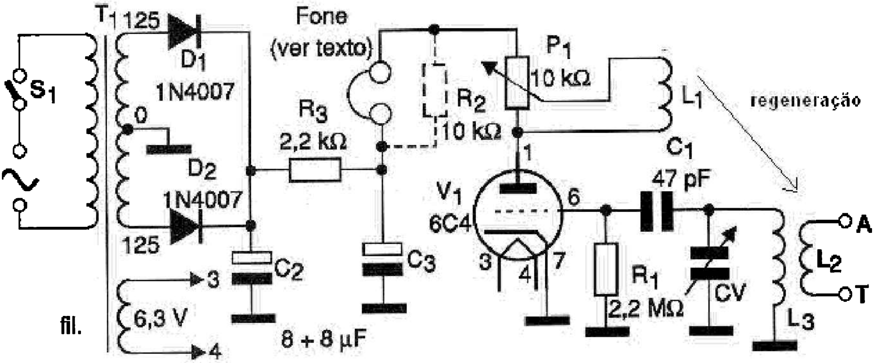
T1
125
D1
1N4007
S1
0
D2
1N4007
125
3
6,3 V
4
fil.
C2
8 + 8 µF
Fone
(ver texto)
R3
2,2 kΩ
C3
R2
10 kΩ
P1
10 kΩ
L1
regeneração
C1
47 pF
V1
6C4
1
6
3
4
7
R1
2,2 MΩ
CV
L3
L2
A
T

Figura 13 – Um receptor regenerativo valvulado (V008 no site do autor www.newtoncbraga.com.br)

A bobina L3 sintoniza o sinal, enquanto que L1 é usada para "jogar" de volta o sinal amplificado. Este tipo de circuito tem uma característica interessante: se a regeneração for muito forte, ou seja, muito do sinal "for jogado de volta" na entrada, ele entra em oscilação, ou seja, o sinal começa a "rodar" pelo circuito num processo que impede a audição de qualquer estação, aparecendo apenas um forte apito no fone.

Existe então um controle que "dosa" a regeneração de modo que ela seja máxima, mas não o suficiente para ocorrer a oscilação. Isso o operador pode perceber facilmente, pois no momento em que o ponto crítico é ultrapassado, o fone "apita".

Na figura 14 temos um circuito de um receptor transistorizado regenerativo.

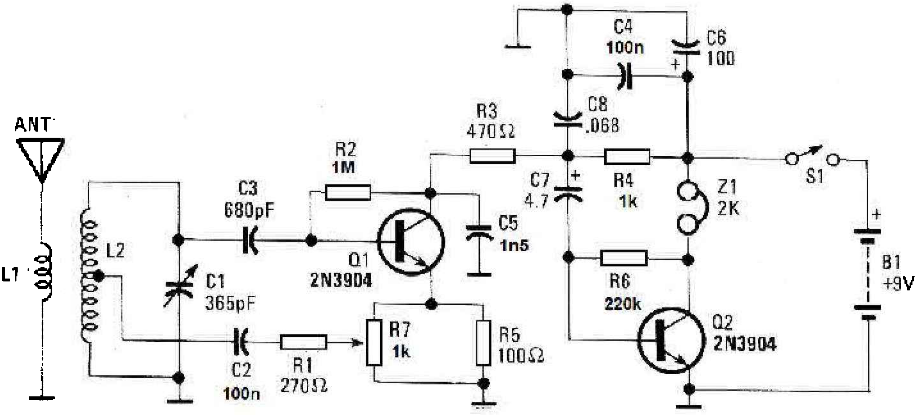
ANT
L1
L2
C1
365pF
C3
680pF
R2
1M
Q1
2N3904
C2
100n
R1
270Ω
R7
1k
R5
100Ω
C5
1n5
R3
470Ω
C8
.068
C4
100n
C6
100
C7
4.7
R4
1k
Z1
2K
R6
220k
Q2
2N3904
S1
B1
+9V

Figura 14 – R7 controla a regeneração

A bobina L2 consiste em 20 + 80 espiras de fio 28 num bastão de ferrite e L120 espiras sobre L2. Os transistores podem ser os BC548

Diversas Etapas

É claro que, em lugar de se usar apenas uma válvula ou transistor para amplificar o sinal, é possível usar diversos. O sinal passa de um para o outro sendo amplificado, até atingir a intensidade desejada.

Com um número suficiente de válvulas ou transistores pode-se chegar a intensidade suficiente para excitar um alto-falante. Na figura 15 temos circuitos de rádios com diversas etapas, usando transistores (mais podem ser encontrados na parte prática deste livro). Evidentemente são circuitos que têm apenas finalidade didática.

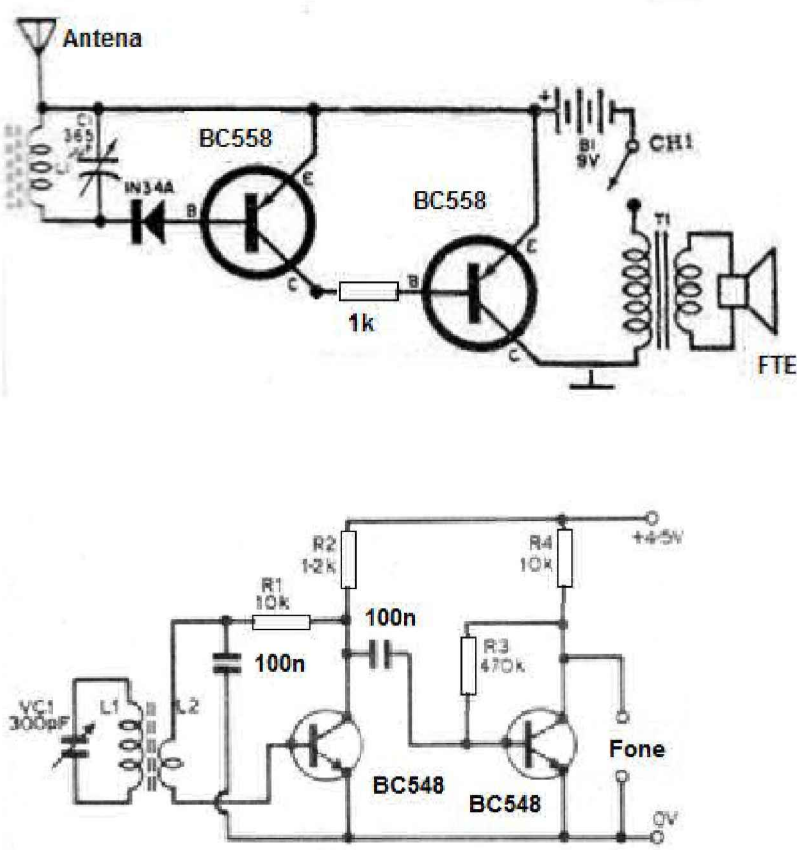
Antena
C1
365
pF
L1
IN34A
BC558
B
E
C
1k
BC558
B
E
C
+
B1
9V
CH1
T1
FTE
R2
1·2k
R4
10k
+4·5V
R1
10k
100n
100n
R3
470k
VC1
300pF
L1
L2
BC548
BC548
Fone
0V

Figura 15 – Rádios de mais de uma etapa

Tente montar o segundo circuito transistorizado. A bobina L1 tem 80 voltas de fio esmaltado 28 e L2 é formada por 20 espiras, tudo num bastão de ferrite de 210 a 15 cm. O capacitor variável é do tipo usado em rádios AM. O fone deve ser piezoelétrico de alta impedância e B1 é formada por 4 pilhas pequenas.

A antena pode ser um pedaço de fio estendido de 3 a 5 metros de comprimento, já que temos muito maior sensibilidade do que nos receptores dados anteriormente. Essa antena pode ser ligada à base do primeiro transistor.

Mas, mesmo com diversas etapas, os rádios ainda não podem ser considerados ideais, pois apresentam problemas como:

Instabilidade: muitas etapas tendem a gerar oscilações por sinais que voltam às etapas de entrada. Apitos e ruídos podem ocorrer.

Falta de seletividade: o rádio não consegue separar estações que transmitem em freqüências próximas.

Para solucionar estes problemas existem outras técnicas que apareceram com o tempo, algumas das quais muito engenhosas.

Neutrodinos, Sincrodinos e Outros

Nesta fase da história do rádio, passou-se a ter uma preocupação maior com a estabilidade dos circuitos. Assim, um primeiro passo foi o que levou aos receptores denominados neutrodinos. Um processo de neutralização evitava que ocorressem oscilações por realimentação entre as diversas etapas do circuito.

No receptor sincrodino, o circuito gerava um sinal sincronizado com o da estação, dando assim maior estabilidade ao circuito. No entanto, esse rádios logo foram superados por uma categoria de circuito que existe até hoje: o superheteródino.

O Superheteródino

Como conseguir aliar alta sensibilidade a uma ótima seletividade e tudo isso num circuito de grande estabilidade? A resposta para este problema está no circuito superheteródino.

Rádios de todos os tipos e até mesmo receptores de TV e telecomunicações modernos se baseiam nesta configuração que se revela satisfatória para a maioria dos casos em que a recepção dos sinais deva ser feita de forma estável e sensível.

Na figura 16 temos a estrutura em blocos de um receptor superheteródino, a partir da qual procuraremos explicar o funcionamento da maneira mais simples possível.

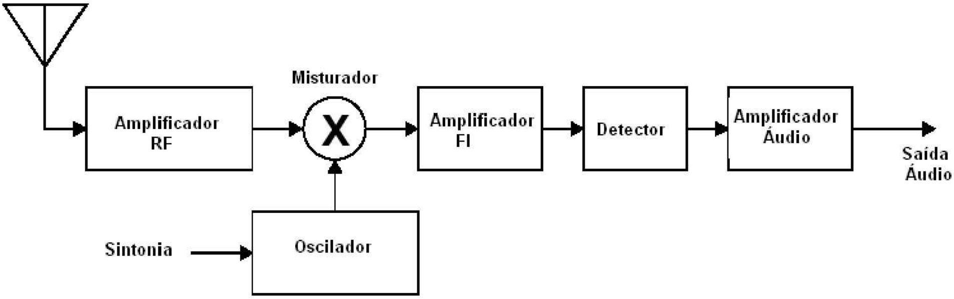

Figura 16 – Receptor superheteródino em blocos

Os sinais são captados pela antena e levado a um circuito de sintonia, onde o sinal da estação que se deseja ouvir é separado. O primeiro passo desse sinal é ser processado por uma etapa chamada misturadora.

A função da etapa misturadora ou misturador é simplesmente combinar o sinal da estação selecionada com um sinal que é gerado no próprio receptor. Este sinal deve ter uma frequência, que no caso do AM, deve ser 455 kHz maior que o sinal captado e no caso do FM, 10,7 MHz maior.

A mistura dos dois sinais provoca um fenômeno denominado batimento ou heterodinagem: obtemos na saída da etapa dois sinais que correspondem à soma e diferença de suas frequências. Como a diferença é fixa, 455 kHz ou 10,7 MHz, as etapas seguintes podem perfeitamente trabalhar com frequências fixas, não se necessitando mais de circuitos de sintonia variável.

Nos receptores de AM temos então etapas amplificadoras seguintes operando em 455 kHz e nos de FM, operando em 10,7 MHz. Estas etapas são denominadas "de FI" ou "Frequência Intermediária".

O sinal de FI que leva então a informação do sinal original captado, tendo apenas uma frequência, pode ser amplificado por uma ou duas etapas adicionais até ficar suficientemente intenso para poder ser detectado.

A detecção é feita então da maneira convencional: nos rádios AM usamos um diodo que detecta a envolvente, ou seja, separa a modulação do sinal original de sua portadora, conforme mostra a figura 17.

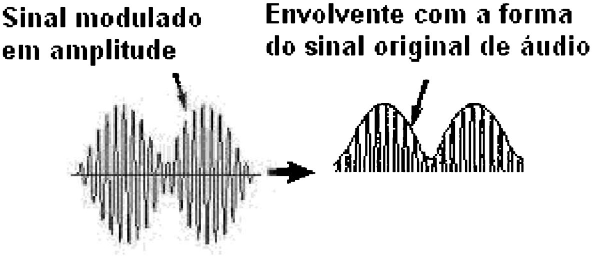
Sinal modulado em amplitude
Envolvente com a forma do sinal original de áudio

Figura 17 – Detecção do sinal modulado em amplitude

Nos rádios de FM utiliza-se um circuito denominado discriminador, pois a modulação é feita de modo diferente. De qualquer forma, a partir dessa etapa temos no circuito apenas sinais de baixas frequências ou áudio que correspondem aos sons originais emitidos.

Podemos então amplificá-los ainda mais, mas agora usando um amplificador comum de áudio. O volume ou potência do som que obteremos na saída vai depender das características deste amplificador.

Nos rádios de FM estéreo, existe ainda um circuito adicional que é o decodificador (multiplex) que faz a separação dos sons dos dois canais que são enviados a dois amplificadores diferentes.

Assim, para rádios portáteis ou do tipo walkman, temos pequenos amplificadores com potências da ordem de milésimos de watt (mW), para os rádios maiores e de carro, as potências aumentam para vários watts e finalmente nos grandes sons, essa potência podem superar os 100 W. Na figura 18 temos um circuito típico de receptor superheteródino.

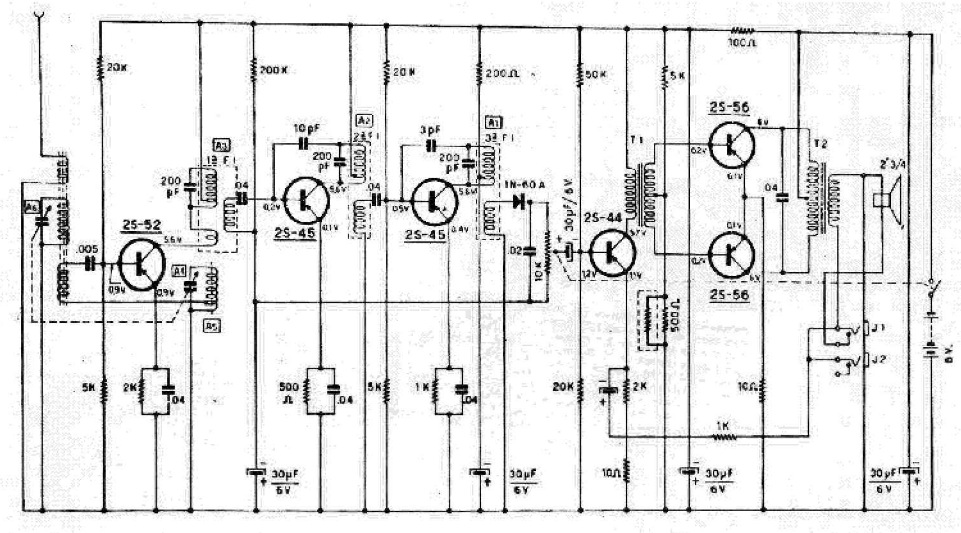
20K
200 K
20 K
200 Ω
50 K
5 K
100 Ω
A3
1ª F I
200 pF
.04
A2
2ª F I
10 pF
200 pF
A1
3ª F I
3 pF
200 pF
IN-60 A
30 µF / 6V
25-52
.005
0,9V
0,9V
5,6V
A6
A4
A5
2S-45
0,2V
4,1V
5,6V
.04
0,5V
2S-45
0,4V
5,8V
.02
10 K
2S-44
12V
1,1V
5,7V
T1
2S-56
6V
0,2V
6,1V
.04
6,1V
0,2V
6V
2S-56
T2
2" 3/4
500 Ω
J1
J2
6V
5K
2K
.04
500 Ω
.04
5K
1K
.04
20K
2K
10 Ω
1K
30 µF / 6V
30 µF / 6V
10 Ω
30 µF / 6V
30 µF / 6V

Figura 18 – Receptor Spica dos anos 50 e 60.

O leitor não deve tentar montar esse rádio. O que ocorre é que as técnicas industriais que levam à produção em massa fazem uso de circuito integrados que unem todas as funções em chips (circuitos integrados) e eles são fabricados por máquinas automáticas que barateiam sua produção.

Conclusão

Circuitos integrados podem reunir milhares ou milhões de componentes possibilitando a elaboração de receptores extremamente complexos e eficientes. As novas técnicas que vão possibilitar a transmissão da informação na forma digital para o rádio comum já estão a caminho.

Assim, tudo o que vimos passará definitivamente para a história, uma história muito interessante que alguns de nós viveram uma boa parte, testemunhando até a época em que o rádio era considerado uma das maravilhas da tecnologia e a TV ainda era uma simples curiosidade acessível a poucos.

NEWTON C. BRAGA

BANCO DE CIRCUITOS

VOLUME 31

100 CIRCUITOS DE RÁDIOS E TRANSMISSORES-2

NCB

BANCO DE CIRCUITOS
100 CIRCUITOS DE RÁDIOS E TRANSMISSORES - 2

São Paulo - 2016
100 CIRCUITOS DE RÁDIOS E TRANSMISSORES - 2
Autor: Newton C. Braga
São Paulo - Brasil - 2016
Palavras-chave: Eletrônica - Engenharia Eletrônica - Componentes – Transmissor - Rádio

Copyright by
INTITUTO NEWTON C BRAGA.
1ª edição

Todos os direitos reservados. Proibida a reprodução total ou parcial, por qualquer meio ou processo, especialmente por sistemas gráficos, microfílmicos, fotográficos, reprográficos, fonográficos, videográficos, atualmente existentes ou que venham a ser inventados. Vedada a memorização e/ou a recuperação total ou parcial em qualquer parte da obra em qualquer programa juscibernético atualmente em uso ou que venha a ser desenvolvido ou implantado no futuro. Essas proibições aplicam-se também às características gráficas da obra e à sua editoração. A violação dos direitos autorais é punível como crime (art. 184 e parágrafos, do Código Penal, cf. Lei nº 6.895, de 17/12/80) com pena de prisão e multa, conjuntamente com busca e apreensão e indenização diversas (artigos 122, 123, 124, 126 da Lei nº 5.988, de 14/12/73, Lei dos Direitos Autorais).

Diretor responsável: Newton C. Braga
Diagramação e Coordenação: Renato Paiotti
Instituto NCB
www.newtoncbraga.com.br
leitor@newtoncbraga.com.br

ÍNDICE

APRESENTAÇÃO

INTRODUÇÃO

- 1 - RÁDIO DE CRISTAL
- 2 - RÁDIO DE CRISTAL (2)
- 3 - RÁDIO SEM FONTE DE ENERGIA
- 4 - GALENA TRADICIONAL
- 5 - GALENA COMPLETO DE 1922
- 6 - GALENA COM CAPACITOR VARIÁVEL
- 7 - RECEPTOR A CRISTAL POTENTE
- 8 - GALENA DE 1926
- 9 - RÁDIO COM DETECTOR DE PERNA DE RÃ
- 10 - RÁDIO DE 1 TRANSISTOR DE 1957
- 11 - SINTONIZADOR DE AM
- 12 - RECEPTOR SEM ALIMENTAÇÃO
- 13 - MICRO RECEPTOR DE AM
- 14 - MINI RECEPTOR DE FM
- 15 - RÁDIO COM ENERGIA ALTERNATIVA
- 16 - RECEPTOR DE VLF
- 17 - GALENA DE ONDAS MÉDIAS E CURTAS
- 18 - RÁDIO ELEMENTAR DE 1 TRANSISTOR
- 19 - RECEPTOR ELEMENTAR DE ONDAS MÉDIAS E CURTAS
- 20 - RADIO SIMPLES DE ONDAS MÉDIAS E CURTAS
- 21 - SUPER REGENERATIVO A CRISTAL
- 22 - RECEPTOR AM DE DOIS TRANSISTORES
- 23 - RÁDIO AM COM O CK722
- 24 - RECEPTOR REGENERATIVO COM O CK722
- 25 - RECEPTOR AM DE DOIS TRANSISTORES
- 26 - RÁDIO AM DE DOIS TRANSISTORES
- 27 - RÁDIO AM SENSÍVEL
- 28 - RÁDIO DE DOIS TRANSISTORES
- 29 - RÁDIO AM COM DETECÇÃO EM PUSH PULL
- 30 - RÁDIO OM E OC COM 1 TRANSISTOR
- 31 - RECEPTOR PARA 11 METROS
- 32 - RÁDIO CMOS
- 33 - RECEPTOR REFLEX DE TRÊS TRANSISTORES
- 34 - RÁDIO OM E OC SEM ALIMENTAÇÃO
- 35 - RECEPTOR CMOS AM
- 36 - RÁDIO AM COM O LM3909
- 37 - RÁDIO AM DE 3 TRANSISTORES
- 38 - RECEPTOR REGENERATIVO DE ONDAS MÉDIAS
- 39 - RECEPTOR SUPER HETERÓDINO DE 1957
- 40 - RÁDIO COMERCIAL DA DÉCADA DE 70
- 41 - RECEPTOR AM PARA FONE SEM FIO
- 42 - RECEPTOR AM ESTÉREO COM O MC13024
- 43 - RECEPTOR PARA 11 METROS
- 44 - RÁDIO AM COM O TBA820
- 45 - TRANSMISSOR OM EXPERIMENTAL
- 46 - TRANSMISSOR DE CENTELHA
- 47 - MICROFONE SEM FIO AM
- 48 - TRANSMISSOR DE ARCO
- 49 - TRANSMISSOR ALIMENTADO POR ENERGIA SOLAR
- 50 - MICRO TRANSMISSOR CW DE ONDAS CURTAS
- 51 - TRANSMISSOR DE 400 UW
- 52 - TRANSMISSOR AM CK722

53 - TRANSMISSOR CW AM DE CURTO ALCANCE
54 - TRANSMISSOR CW PARA ONDAS MÉDIAS
55 - TRANSMISSOR TTL SEM BOBINAS
56 - TRANSMISSOR TTL SEM BOBINAS (2)
57 - TRANSMISSOR SOLAR
58 - TRANSMISSOR PARA 27 MHZ
59 - TRANSMISSOR FONOGRÁFICO
60 - TRANSMISSOR ULTRA-MINIATURIZADO
61 - TRANSMISSOR DE ONDAS CURTAS XTAL
62 - TRANSMISSOR AM DE 4 W
63 - TRANSMISSOR DE TELEMETRIA
64 - TRANSMISSOR PARA 15 METROS
65 - TRANSMISSOR PARA ESCUTA CLANDESTINA
66 - TRANSMISSOR ESPIÃO DE FM
67 - TRANSMISSOR DE FM
68 - TRANSMISSOR DE FM EXPERIMENTAL
69 - TRANSMISSOR DE 27 MHZ
70 - TRANSMISSOR ESPIÃO COM DIODO TUNNEL
71 - TRANSMISSOR DE FM COM FET
72 - MICROFONE SEM FIO
73 - TRANSMISSOR DE FM
74 - TRANSMISSOR DE VÍDEO
75 - TRANSCEPTOR PARA A FAIXA DOS 11 METROS
76 - TRANSMISSOR DE FM COM FET (9V)
77 - TRANSMISSOR COM DIODO TUNNEL (1)
78 - TRANSMISSOR COM DIODO TUNNEL (2)
79 - TRANSMISSOR ESPIÃO COM DIODO TUNNEL
80 - TRANSMISSOR ESPIÃO COM DIODO TUNNEL (2)
81 - TRANSMISSOR ESPIÃO DE FM COM VARICAP
82 - TRANSMISSOR ESPIÃO ULTRA MINIATURA
83 - MICRO TRANSMISSOR DE FM DE 1 PILHA
84 - TRANSMISSOR DE FM DE 200 MW
85 - TRANSMISSOR DE FM COM DOIS FETS
86 - TRANSMISSOR DE FM DE 300 MW COM FET
87 - TRANSMISSOR ESPIÃO
88 - TRANSMISSOR INTEGRADO MC833
89 - TRANSCEPTOR EXPERIMENTAL
90 - GRID DIP METER
91 - GATE DIPPER
92 - MONITOR CW
93 - OSCILADOR DE 500 MHZ
94 - INJETOR FM
95 - OSCILADOR DE 200 MHZ
96 - ETAPA DE ÁUDIO PARA RÁDIOS TRANSISTORIZADOS
97 - OSCILADOR DE RF LM3909
98 - MODULADOR BALANCEADO MC1556G
99 - DECODIFICADOR QAM PARA AM ESTÉREO
100 - TRANSMISSOR INTEGRADO SEM BOBINAS
A QUALIDADE DOS RÁDIOS AM
AJUSTE DE RÁDIOS AM
AJUSTANDO TRANSMISSORES

Apresentação

Durante nossa longa carreira como escritor de artigos técnicos, incluindo também informações, abordamos o termo Coletânea de Circuitos. Assim, há diversos anos passados, abordando este tema, publicamos as séries “Circuitos e Informações” (7 volumes) e “Circuitos e Soluções” (5 volumes) contendo centenas de circuitos úteis e informações técnicas de todos os tipos. As séries se esgotaram, o tempo passou, mas os leitores ainda nos cobram algo semelhante atualizado e que possa ser usado ainda em projetos de todos os tipos. De fato, circuitos básicos usando componentes discretos comuns, de transistores a circuitos integrados, são ainda amplamente usados como solução simples para problemas imediatos, parte de projetos mais avançados e até com finalidade didática atendendo à solicitação de um professor que necessita de uma aplicação para uma teoria,ou de um aluno que precisa fazer um trabalho. Assim, voltamos com esta série, mas com uma estrutura diferenciada, novos projetos e nova abordagem. O diferencial na abordagem será dividir os diversos volumes da série por temas. Assim, no nosso primeiro volume tivemos circuitos de áudio, depois circuitos de fontes, no terceiro, circuitos osciladores, e assim por diante. Chegamos ao trigésimo primeiro volume com uma seleção de circuitos de Rádios e Transmissores, a segunda da série, pois na primeira não conseguimos colocar todos os circuitos que temos. Na verdade como temos uma enorme quantidade de circuitos em nosso estoque, mais de 9000 deles e a cada dia este estoque aumenta, já fizemos 1 volume anterior com circuitos de transmissores e receptores, mas temos muito mais. Muitos dos circuitos apresentados podem ser acessados de forma dispersa no site do autor e também em outros livros. A vantagem de se ter estes circuitos organizados em volumes, além do acesso em qualquer parte, está na fácil localização de um circuito de determinado tipo. Teremos ainda artigos complementares de utilidade, relacionados com os temas abordados. A maioria destes circuitos, colhidos em publicações que, em alguns casos, podem não ser muito atuais, recebe um tratamento especial com comentários, sugestões e atualizações que viabilizam sua execução mesmo em nossos dias. Enfim, com esta série, damos aos leitores a oportunidade de ter em seus tablets, iPhones, iPads, PCs, notebooks e outras mídias uma fonte de consulta de grande importância tanto para seu trabalho, como para seus estudos ou simples como hobby.

Newton C. Braga

Introdução

Depois do sucesso da seção Banco de Circuitos no meu site e das coleções esgotadas de Circuitos e Informações e Circuitos e Soluções, levo aos meus leitores uma coletânea de circuitos selecionados de minha enorme coleção de documentos técnicos e livros. Durante minha vida toda colecionei praticamente todas as revistas técnicas de eletrônica estrangeiras, dos Estados Unidos, França, Espanha, Itália, Alemanha, Argentina e até mesmo do Japão, possuindo assim um enorme acervo técnico. Não posso reproduzir os artigos completos que descrevem os projetos que saem nessas revistas, por motivos ditados pela lei dos direitos autorais, mas a mesma lei permite que eu utilize uma figura do texto, com citação, comentando seu conteúdo para efeito de informação ou complementação de um conteúdo maior. É exatamente isso que faço na minha seção no site e também disponibilizo neste livro. Estou selecionando os principais circuitos destas publicações, verificando quais ainda podem ser montados em nossos dias, com a eventual indicação de componentes equivalentes, fazendo alterações que julgo necessárias e disponibilizando-os aos nossos leitores. Para o site já existem mais de 9000 circuitos, no momento que escrevo este livro, mas a quantidade aumenta dia a dia. Acesse o site, que ele poderá lhe ajudar a encontrar aquela configuração que você precisa para seu projeto. Os 100 circuitos selecionados para esta edição da série são apenas uma pequena amostra do que você vai encontrar no site. Para esta edição escolhemos mais uma remessa com 100 circuitos de receptores e transmissores, tais como rádios AM, transmissores de AM, FM e CW e muito mais. Estes circuitos, na maioria dos casos, podem ser montados com facilidade, já que na maioria usam componentes comuns. Fica por conta do leitor a obtenção das partes necessárias, caso deseje ter uma montagem prática.

Newton C. Braga

Volumes Anteriores:
Volume 1 - 100 Circuitos de áudio
Volume 2 – 100 Circuitos de fontes
Volume 3 – 100 Circuitos osciladores
Volume 4 - 100 Circuitos de potência
Volume 5 – 100 Circuitos com LEDs
Volume 6 – 100 Circuitos de rádios e transmissores
Volume 7 – 100 Circuitos de Filtros
Volume 8 – 100 Circuitos de Alarmes e Sensores
Volume 9 – 100 Circuitos de Testes e Instrumentação
Volume 10 – 100 Circuitos de Tempo
Volume 11 – 100 Circuitos com Operacionais
Volume 12 – 100 Circuitos de Áudio 2
Volume 13 – 100 Circuitos com FETs
Volume 14 – 100 Circuitos Diversos
Volume 15 – 100 Circuitos com LEDs e Displays
Volume 16 – 100 Circuitos de Potência 2
Volume 17 – 100 Circuitos Automotivos
Volume 18 – 100 Circuitos de Efeitos de Luz e Som

Volume 19 – 100 Circuitos Fotoelétricos
Volume 20 – 100 Circuitos de Fontes 2
Volume 21 – 100 Circuitos de Osciladores (2)
Volume 22 – 100 Circuitos Ecológicos e Para Saúde
Volume 23 – 100 Circuitos com Válvulas
Volume 24 – 100 Circuitos de Jogos e Recreação
Volume 25 – 100 Circuitos TTL
Volume 26 – 100 Circuitos CMOS
Volume 27 – 100 Circuitos Para o Lar
Volume 28 - 100 Circuitos de Automação e Controle
Volume 29 – 100 Circuitos com LEDs e Displays
Volume 30 – 100 Circuitos com Shields para Arduino
Outros livros recomendados do mesmo autor:
- Como Testar Componentes em quatro volumes
- Curso de Eletrônica Digital – Vol. 1 e 2
- Como Fazer Montagens
- Os segredos no Uso do Multímetro
- Curso de Eletrônica – Eletrônica Básica
- Curso de Eletrônica – Eletrônica Analógica

Conserte Tudo

1 - Rádio de Cristal

Este circuito tradicional de rádio de cristal, mas que tem a versão moderna com diodo foi encontrado numa revista italiana Sperimentare de 1977. O fone também pode ser de cristal e a bobina consta de 100 espiras de fio 28 num tubo que pode deslizar sobre um bastão de ferrite para fazer a sintonia. A antena deve ser bem longa. O diodo pode ser o 1N34 ou 1N60 também.

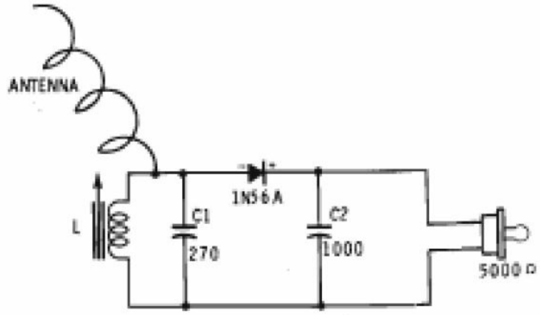

2 - Rádio de Cristal (2)

Este circuito é de um manual de transistores da Raytheon de 1957. O diodo pode ser de germânio de uso geral como o 1N34 e a bobina enrolada com 30 + 70 espiras de fio 28 num bastão de ferrite. A antena deve ser longa e o fone deve ser magnético de alta impedância. O variável pode ser obtido num rádio AM comum.

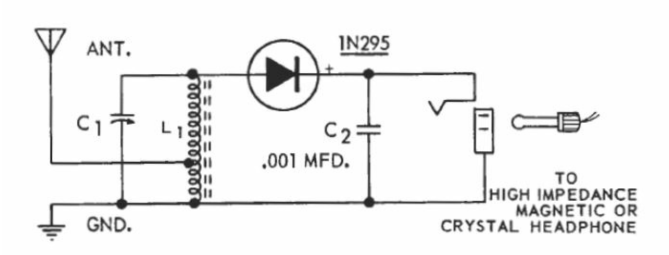
ANT.
1N295
+
C1
L1
C2
.001 MFD.
GND.
TO
HIGH IMPEDANCE
MAGNETIC OR
CRYSTAL HEADPHONE

C1 – 365 MmF, tuning capacitor
C2 – 0.001 Mfd, disc ceramic capacitor
L1 – Tapped antenna coil (Miller no. 2000 or 2004 or Meissner no. 14-9003)
Diode – 1 Raytheon 1N295.

3 - Rádio sem Fonte de Energia

Este receptor usa a própria tensão induzida pelo sinal da estação para alimentar o amplificador com 1 transistor. O transistor é o CK722, o primeiro transistor comercial do mundo, lançado pela Raytheon na década de 50. O fone deve ser de alta impedância e as bobinas têm as seguintes características: L1: 30 espiras de fio 28, L2 60 + 40 espiras de fio 28, as duas no mesmo bastão de ferrite. Qualquer transistor PNP de germânio pode ser usado.

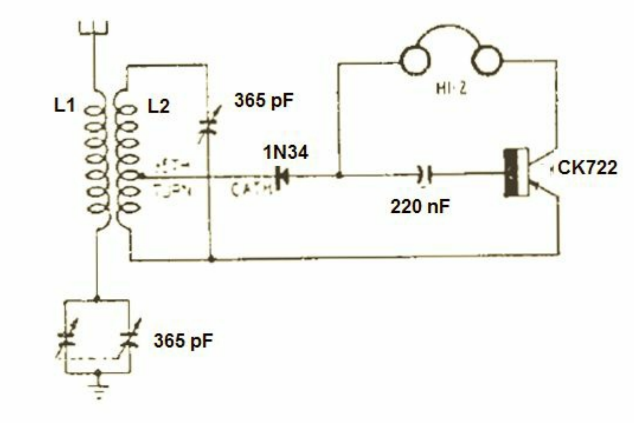

4 - Galena Tradicional

Este circuito é de uma publicação do início do século XX. O detector é um cristal de galena e a bobina formada por 100 espiras de fio 28 com a parte em que desliza o cursor raspada. A forma é de 3 a 5 cm. A antena deve ser longa e o fone magnético de alta impedância

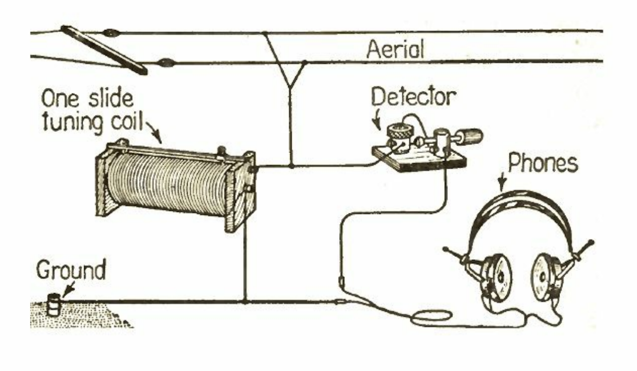
Aerial
One slide
tuning coil
Detector
Phones
Ground

5 - Galena Completo de 1922

Os planos de construção deste receptor de galena de 1922 foram encontrados numa publicação de Hugo Gernsback. Na figura temos todos os detalhes construtivos do rádio de ondas médias da época. O fone deve ser de alta impedância (acima de 2k) e a antena muito longa. O cristal de galena pode ser substituído por um diodo 1N34 ou equivalente.

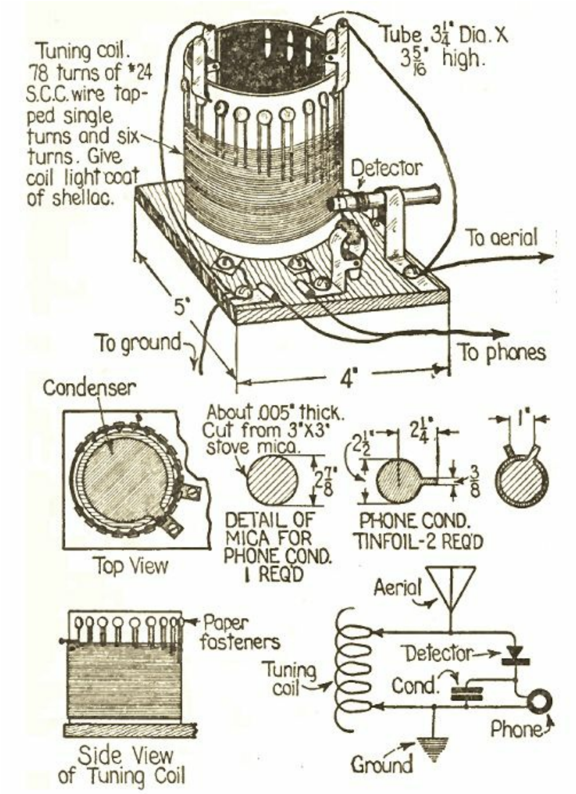
Tuning coil. 78 turns of #24 S.C.C. wire tapped single turns and six turns. Give coil light coat of shellac.
Tube 3¼" Dia. X 3 5/16" high.
Detector
To aerial
5"
To ground
To phones
4"
Condenser
Top View
About .005" thick. Cut from 3"X3" stove mica.
2 7/8"
DETAIL OF MICA FOR PHONE COND. 1 REQ'D
2½"
2¼"
3/8
PHONE COND. TINFOIL-2 REQ'D
1"
Aerial
Paper fasteners
Tuning coil
Detector
Cond.
Phone
Ground
Side View of Tuning Coil

6 - Galena com Capacitor Variável

Este rádio de galena usa um capacitor variável de construção caseira. O circuito é do início do século XX. A bobina é formada por 100 espiras de fio 28 ou próximo num tubo de PVC ou papelão de 2,5 cm de diâmetro. Pode ser usado um capacitor variável de rádio AM comum a e antena deve ser bem longa. A ligação à terra é importante.

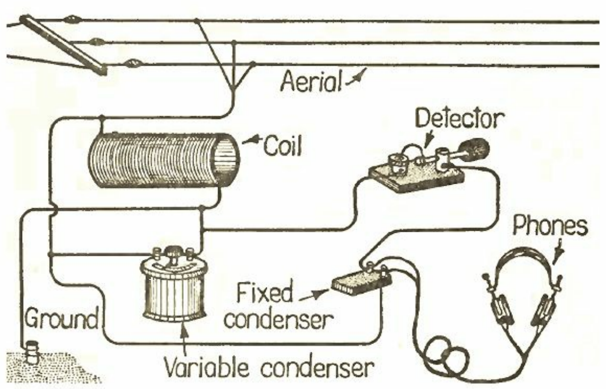
Aerial
Detector
Coil
Phones
Fixed condenser
Ground
Variable condenser

7 - Receptor a Cristal Potente

O circuito mostrado na figura foi encontrado numa publicação de 1964. Os diodos podem ser de germânio e L2 têm 100 4espiras de fio 28 num bastão de ferrite. L1 é enrolada sobre L2 com 30 espiras. L3 é formada por 22 voltas de fio 22 no mesmo bastão de ferrite em que é enrolada L3. Os diodos podem ser 1N34 e o fone é cerâmico ou magnético de alta impedância. A antena deve ser longa.

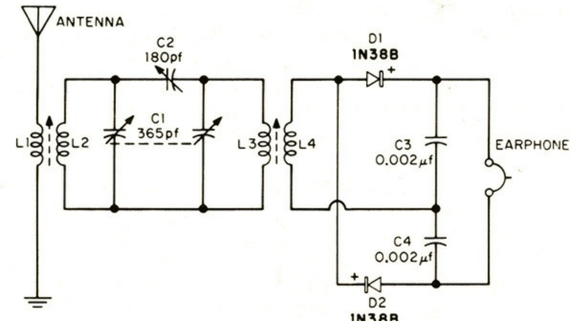
ANTENNA
C2
180pf
C1
365pf
L1
L2
L3
L4
D1
1N38B
C3
0.002μf
C4
0.002μf
EARPHONE
D2
1N38B

8 - Galena de 1926

Encontramos este circuito na revista La Radio per Tutti da Itália de 1926. A bobina pode ser elaborada com 100 espiras fio 28 num tubo de papelão ou PVC de 2 a 2,5 cm e com tomadas de 10 em 10 ou de 15 em 15 espiras. A antena deve ser longa e o cristal de galena pode ser substituído por um diodo de germânio. O fone deve ser de alta impedância. O capacitor C pode ser de 2n2 cerâmico.

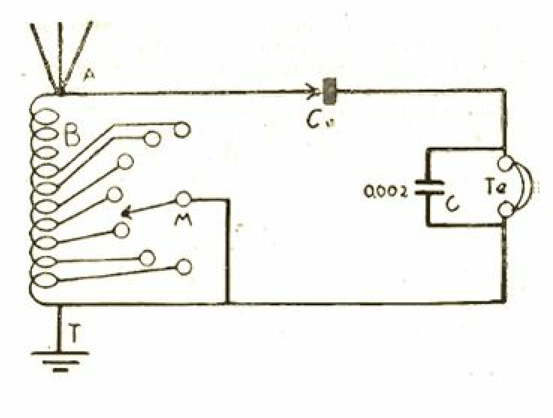

9 - Rádio com Detector de Perna de Rã

Sim, é isso mesmo: perna de rã. Naquela época os experimentos de Galvani com "eletricidade animal" faziam crer que a eletricidade que se manifestava nos seres vivos era diferente da eletricidade natural e muitos experimentos foram feitos com a excitabilidade das pernas de rã. O circuito mostrado é de uma publicação de 1921 e mostra um detector de sinais de rádio usando uma perna de rã.

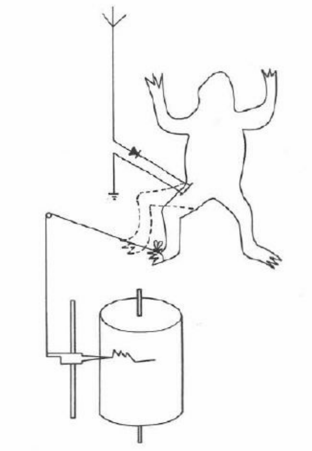

10 - Rádio de 1 Transistor de 1957

Encontramos este rádio AM num manual de aplicações de transistores de 1957, publicado pela Raytheon. O transistor pode ser um equivalente moderno como o BC558 e a bobina tem 20 + 80 espiras de fio 28 num bastão de ferrite. O fone deve ser de alta impedância.

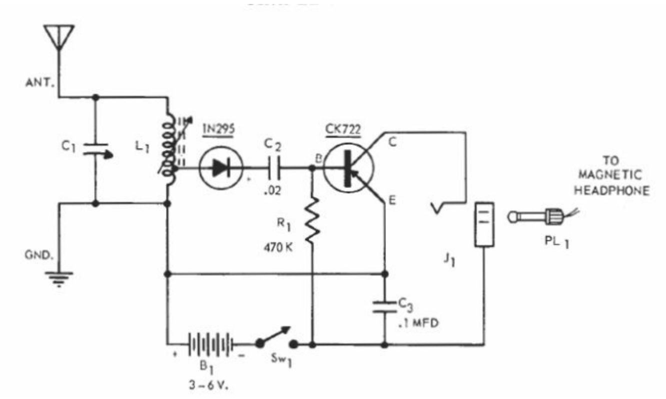
ANT.
C1
L1
1N295
C2
.02
B
CK722
C
E
R1
470 K
TO MAGNETIC HEADPHONE
PL 1
J1
GND.
C3
.1 MFD
+
B1
3 - 6 V.
-
Sw1

R_1 — 470K, ½ watt carbon resistor.
C_1 — 365 Mmf. tuning capacitor.
C_2 — .02 Mfd. disc ceramic capacitor.
C_3 — 0.1 Mfd., 200 volt paper capacitor.
L_1 — Tapped antenna coil (*Miller* No. 2004 or *Meissner* 14-9019 or *Argonne* No. AR-68).
Sw_1 — SPST toggle or slide switch.
B_1 — 3 to 6 volt battery (two to four penlight cells in series).
J_1 — Open circuit 'phone jack.
PL_1 — Standard 'phone plug.
DIODE — (1) *Raytheon* 1N295.
TRANSISTOR — (1) *Raytheon* CK722.

11 - Sintonizador de AM

Este circuito foi encontrado num livro americano de 1964. Com ele podemos sintonizar estações de ondas médias e aplicar o sinal a um amplificador de áudio. L1 é formada por 100 espiras de fio 28 num bastão de ferrite e o transistor pode ser o BC558. A alimentação pode ser feita por tensões de 3 a 9 V.

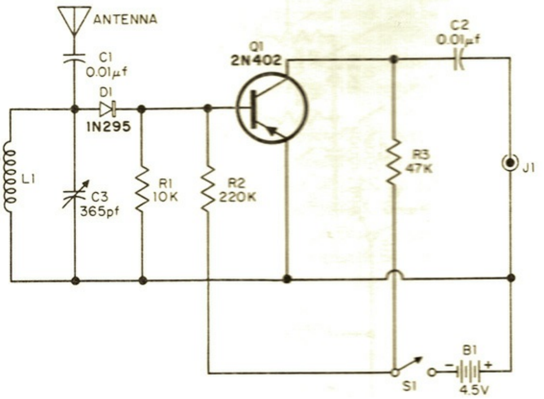
ANTENNA
C1
0.01μf
D1
1N295
Q1
2N402
C2
0.01μf
L1
C3
365pf
R1
10K
R2
220K
R3
47K
J1
B1
−
+
4.5V
S1

12 - Receptor Sem Alimentação

No receptor apresentado, de uma documentação de 1964, um transistor é usado como detector, funcionando sem alimentação. A antena deve ser muito longa e L1 formada por 100 espiras de fio 28 num bastão de ferrite. L2 é formada por 30 espiras do mesmo fio sobre L1. O fone é de alta impedância. Qualquer transistor pode ser usado, mas melhores resultados são obtidos com tipos antigos de germânio.

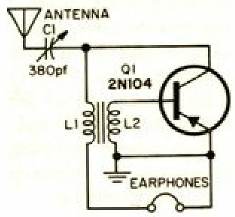

13 - Micro Receptor de AM

Este circuito experimental que preferivelmente deve usar transistor de germânio foi encontrado numa publicação de 1964. A bobina é formada por 30 + 70 espiras de fio 28 num bastão de ferrite. A antena deve ser longa e o circuito funciona com células fotoelétricas comuns, como as obtidas em velhas calculadoras. O fone deve ser magnético de alta impedância.

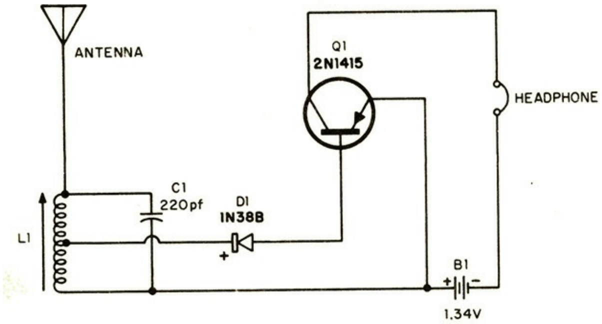
ANTENNA
Q1
2N1415
HEADPHONE
C1
220pf
D1
1N38B
L1
+
B1
+
−
1,34V

14 - Mini Receptor de FM

Encontramos este circuito de receptor super-regenerativo de FM numa publicação americana de 1964. O transistor original pode ser substituído por NPN BF494 com a inversão da fonte de alimentação. O fone deve ser magnético de alta impedância e os capacitores cerâmicos. A bobina L1 consta de 4 espiras de fio 26 ou 28 em forma de 1 cm sem núcleo e RFC1 é um choque de 10uH.

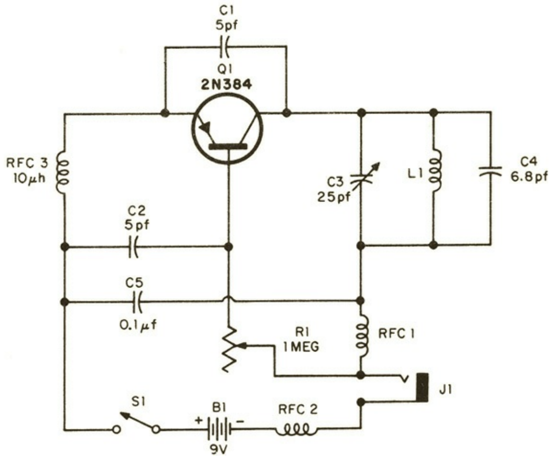
C1
5pf
Q1
2N384
RFC 3
10μh
C2
5pf
C3
25pf
L1
C4
6.8pf
C5
0.1μf
R1
1MEG
RFC 1
J1
S1
B1
+
-
9V
RFC 2

15 - Rádio Com Energia Alternativa

Duas barras de metal numa solução de água e sal fornecem a energia para alimentar este rádio AM, encontrado numa publicação de 1964. L2 é formado por 100 espiras de fio 28 num bastão de ferrite. L1 tem 10 espiras do mesmo fio num bastão de ferrite. O transistor pode ser um BC558, mas o circuito funcionará melhor com transistores de germânio. O Fone deve ser magnético de alta impedância.

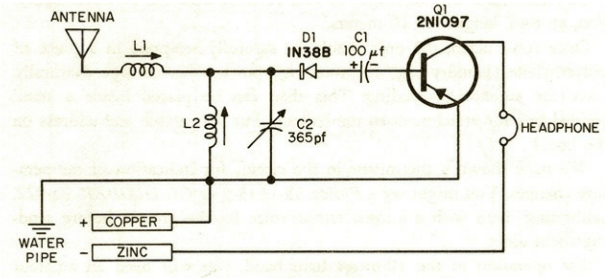
ANTENNA
L1
D1
IN38B
C1
100 µf
+
−
Q1
2N1097
L2
C2
365pf
HEADPHONE
+
COPPER
WATER
PIPE
−
ZINC

16 - Receptor de VLF

Este receptor experimental, encontrado num livro de 1964, se destina à recepção de sinais entre 4 e 16 kHz. A antena de quadro L1 consiste em 200 voltas de fio 26 ou 28 num quadro com pelo menos 1 m de diâmetro. Os transistores podem ser os BC558 e a saída deve ser aplicada a um fone ou amplificador.

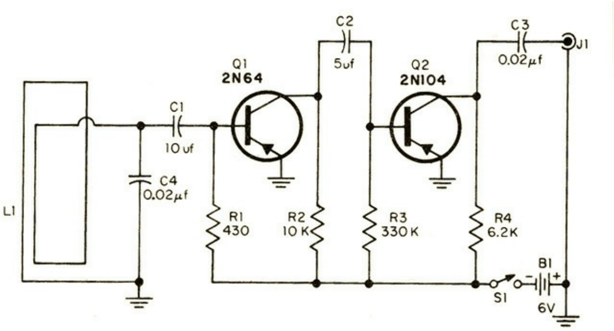
C2
5uf
C3
0.02μf
J1
Q1
2N64
Q2
2N104
C1
10 uf
C4
0.02μf
L1
R1
430
R2
10 K
R3
330 K
R4
6.2K
B1
S1
6V

17 - Galena de Ondas Médias e Curtas

Este receptor com diodo detector de germânio para a faixa de ondas médias e curtas foi encontrado numa publicação dos anos 80. A bobina de ondas curtas consiste em 30 + 30 espiras de fio 28 num tubo de 2,5 cm de diâmetro e a bobina de ondas médias tem 100 espiras do mesmo fio no mesmo tubo. A antena deve ser bem longa.

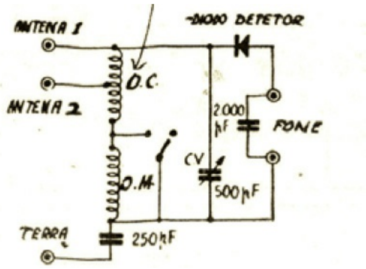
ANTENA 1
-DIODO DETETOR
O.C.
ANTENA 2
2000 pF
FONE
CV
O.M.
500 pF
TERRA
250 pF

18 - Rádio Elementar de 1 Transistor

Encontramos este receptor para ondas médias e curtas numa publicação dos anos 80. A bobina para ondas médias consiste em 100 espiras de fio 28 num tubo de 1 polegada e a bobina de ondas curtas em 30 espiras do mesmo fio no mesmo tubo. O circuito originalmente usa transistor de germânio, mas equivalente de silício pode ser empregado. O fone é magnético de alta impedância.

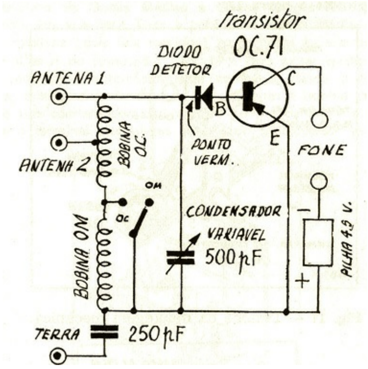
transistor
OC.71
DIODO
DETETOR
ANTENA 1
C
B
E
ANTENA 2
BOBINA O.C.
PONTO VERM.
FONE
OM
OC
CONDENSADOR
VARIAVEL
500 pF
BOBINA OM
−
+
PILHA 4,5 V.
TERRA
250 pF

19 - Receptor Elementar de Ondas Médias e Curtas

Este receptor à cristal com um diodo detector usa uma bobina para receber ondas médias e curtas. Encontramos este circuito numa documentação dos anos 80. A bobina ½ tem 40 espiras de fio 28 num tubo de 2,5 cm (PVC). A bobina 2,3,4 tem 30+50 espiras do mesmo fio no mesmo tubo. O fone pode ser cerâmico ou de cristal.

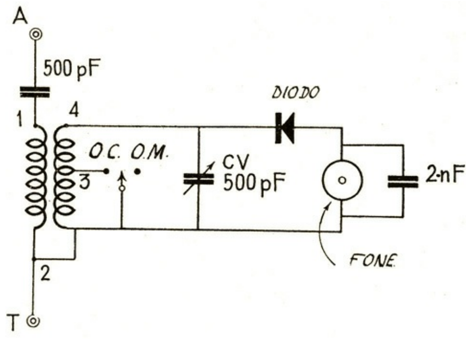
A
500 pF
1
4
DIODO
O.C. O.M.
3
CV
500 pF
2nF
FONE
2
T

20 - Radio Simples de Ondas Médias e Curtas

Este rádio para duas faixas de ondas, curtas e médias, foi encontrado numa publicação dos anos 70. A bobina de ondas curtas tem 15 + 5 espiras de fio 28 num tubo de PVC de 1 polegada e a bobina de ondas médias tem 20 + 80 espiras do mesmo fio num tubo igual. A antena deve ser longa e a ligação à terra é importante.

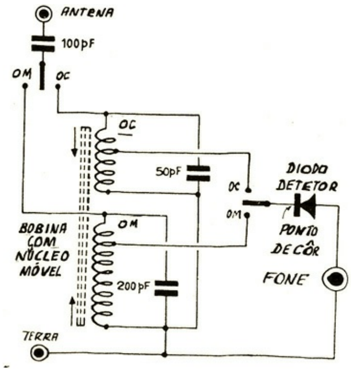
ANTENA
100pF
OM
OC
OC
50pF
OC
OM
DIODO DETETOR
PONTO DE CÔR
OM
BOBINA COM NÚCLEO MÓVEL
200pF
FONE
TERRA

21 - Super Regenerativo a Cristal

Este interessante circuito tem sua frequência sincronizada por um cristal de 10 MHz, por exemplo. L1 determina a faixa de frequências que o circuito sintoniza. Os transistores podem ser os 2N2222. O circuito é de uma documentação de telecomunicações de 1972. O amplificador de áudio pode ser o LM386 obtendo-se saída em alto-falante.

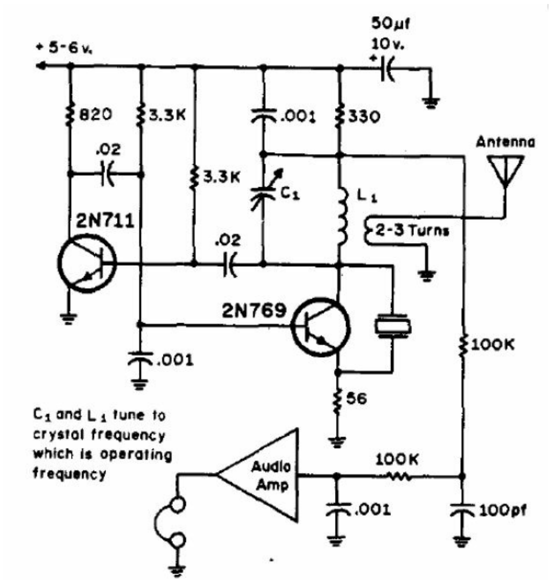
+5-6 v.
50μf
10 v.
820
3.3K
.001
330
Antenna
.02
3.3K
C₁
L₁
2-3 Turns
2N711
.02
2N769
100K
.001
56
C₁ and L₁ tune to crystal frequency which is operating frequency
Audio Amp
100K
.001
100pf

22 - Receptor AM de Dois Transistores

Este circuito usa originalmente transistores de germânio tendo sido obtido numa publicação de 1960. No entanto, o circuito pode ser elaborado com transistores de silício e a alimentação feita com dois conjuntos de 3 V (duas pilhas). A antena depende da intensidade dos sinais e a bobina consiste em 100 espiras de fio 28 num bastão de ferrite. O variável pode ser obtido de um rádio de ondas médias abandonado.

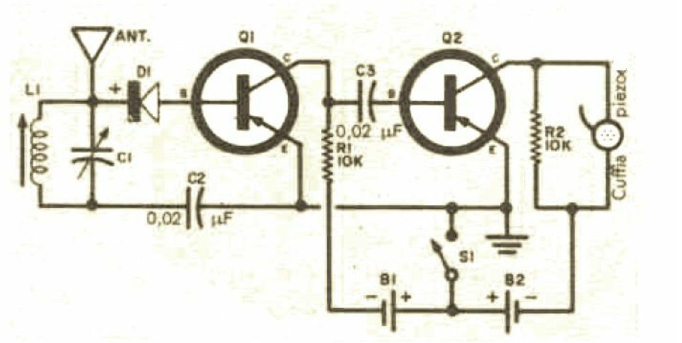
ANT.
Q1
Q2
L1
+
D1
B
C
E
C3
0,02 µF
R1
10K
C1
C2
0,02 µF
R2
10K
piezo
Cuffia
S1
B1
− +
B2
+ −

23 - Rádio AM com o CK722

O CK722 foi o primeiro transistor comercial do mundo, lançado na década de 50. O projeto abaixo é do próprio manual da Raytheon, que o lançou. Podem ser usados transistores mais modernos. O rádio é para a faixa de AM. As bobinas constam de 80 + 30 espiras de fio 28 em bastão de ferrite. Na época era indicada uma bobina comercial pronta.

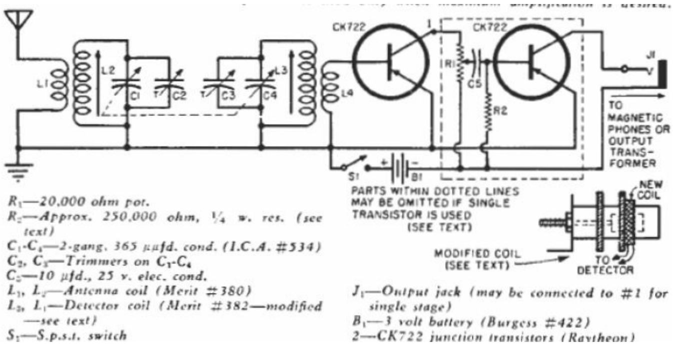
L1
L2
C1
C2
C3
C4
L3
L4
CK722
1
CK722
R1
C5
R2
J1
TO MAGNETIC PHONES OR OUTPUT TRANS-FORMER
S1
B1
PARTS WITHIN DOTTED LINES MAY BE OMITTED IF SINGLE TRANSISTOR IS USED (SEE TEXT)
NEW COIL
MODIFIED COIL (SEE TEXT)
TO DETECTOR

R_1—20,000 ohm pot.
R_2—Approx. 250,000 ohm, ¼ w. res. (see text)
C_1-C_4—2-gang, 365 μμfd. cond. (I.C.A. #534)
C_2, C_3—Trimmers on C_1-C_4
C_5—10 μfd., 25 v. elec. cond.
L_1, L_2—Antenna coil (Merit #380)
L_3, L_4—Detector coil (Merit #382—modified —see text)
S_1—S.p.s.t. switch
J_1—Output jack (may be connected to #1 for single stage)
B_1—3 volt battery (Burgess #422)
2—CK722 junction transistors (Raytheon)

24 - Receptor Regenerativo com o CK722

O CK722 foi o primeiro transistor comercial do mundo, lançado pela Raytheon na década de 50. Este circuito é de um manual de aplicações da Raytheon da época. Podem ser usados transistores equivalentes mais modernos. O fone é de alta impedância e a bobina consta de 80 + 30 espiras de fio 28 num bastão de ferrite.

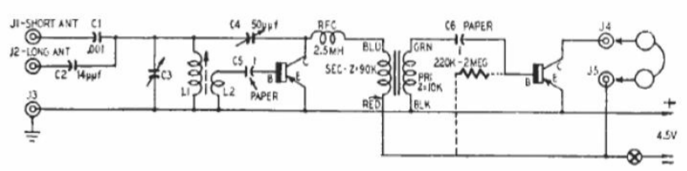
J1-SHORT ANT
C1
.001
J2-LONG ANT
C2
14μμf
J3
C3
L1
L2
C4
50μμf
C5
PAPER
B
E
C
RFC
2.5MH
BLU
SEC-Z=90K
RED
GRN
PRI
Z=10K
BLK
C6 PAPER
220K-2MEG
B
E
C
J4
J5
+
4.5V

25 - Receptor AM de Dois Transistores

Este receptor foi encontrado numa publicação de 1968, mas pode ser montado com um BC548 e um BC558 em lugar dos transistores originais. O fone deve ser de alta impedância. L1 tem um enrolamento com 20+80 espiras de fio 28 num tubo de 2,5 cm e outro enrolamento de 20 espiras. A antena deve ser longa.

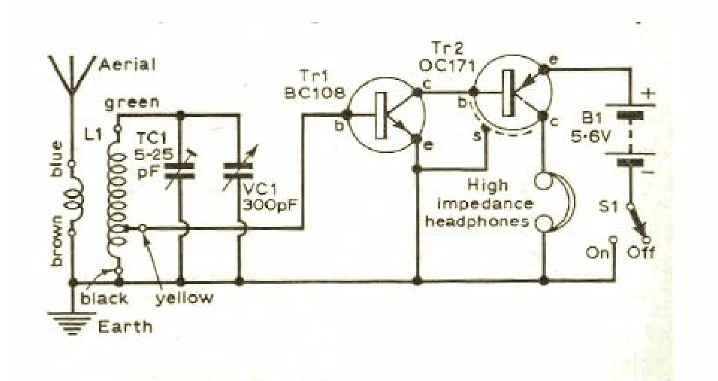
Aerial
green
L1
blue
brown
TC1
5-25
pF
VC1
300pF
Tr1
BC108
b
c
e
Tr2
OC171
b
s
e
c
High
impedance
headphones
B1
5·6V
+
−
S1
On
Off
black
yellow
Earth

26 - Rádio AM de Dois Transistores

Encontramos este rádio AM numa publicação da Raytheon de 1957. Os transistores podem ser tipos equivalentes modernos como os BC548 e o diodo pode ser o 1N34. A bobina tem 30 + 70 espiras de fio 28 num bastão de ferrite a antena deve ser longa. O fone deve ser magnético de alta impedância.

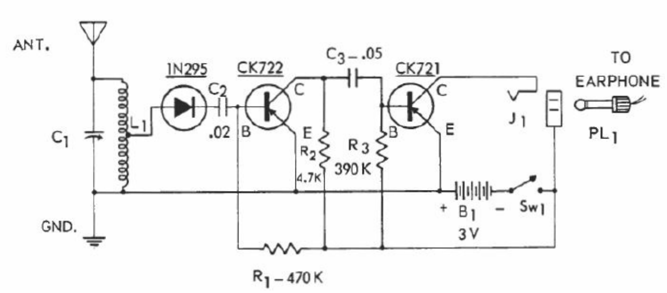
ANT.
1N295
CK722
C_3 – .05
CK721
TO EARPHONE
C_2
C
C
L_1
J_1
C_1
.02
B
E
B
E
PL_1
R_2
4.7K
R_3
390 K
+
B_1
–
Sw_1
3 V
GND.
R_1 – 470 K

27 - Rádio AM Sensível

Este circuito é de uma publicação de 1977 caracterizando-se por um ajuste do ponto de polarização do diodo (que pode ser de germânio ou silício) de modo a se obter maior sensibilidade. A bobina consta de 100 espiras de fio 28 num bastão de ferrite e o capacitor de sintonia é obtido de um rádio transistorizado fora de uso. O fone deve ser de cristal.

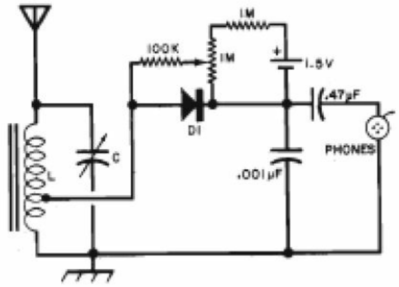

28 - Rádio de Dois Transistores

Este rádio AM para as estações locais é de uma publicação de 1964. Podemos montá-lo com o uso de componentes equivalentes mais modernos. A bobina é formada por 80 espiras de fio 28 num bastão de ferrite e os transistores podem ser BC558. O fone é de alta impedância ou de cristal.

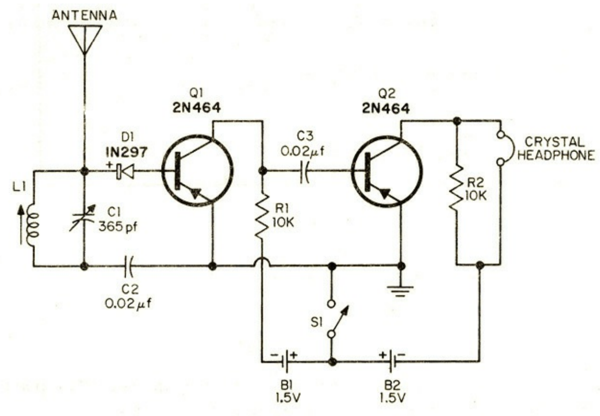
ANTENNA
Q1
2N464
D1
IN297
L1
C1
365 pf
C2
0.02μf
C3
0.02μf
R1
10K
Q2
2N464
R2
10K
CRYSTAL
HEADPHONE
S1
B1
1.5V
B2
1.5V

29 - Rádio AM com Detecção em Push Pull

Esta configuração diferente de rádio AM com detecção e amplificação em push pull proporciona bom rendimento para as estações locais. O circuito é de uma documentação de 1964 e os transistores podem ser o BC558. Os diodos podem ser do tipo 1N34 ou equivalente, e o transformador é do tipo de saída encontrado em rádios transistorizados. A alimentação também pode ser feita com 6 V e L1 é formada por 80 a 100 voltas de fio esmaltado 28 num bastão de ferrite.

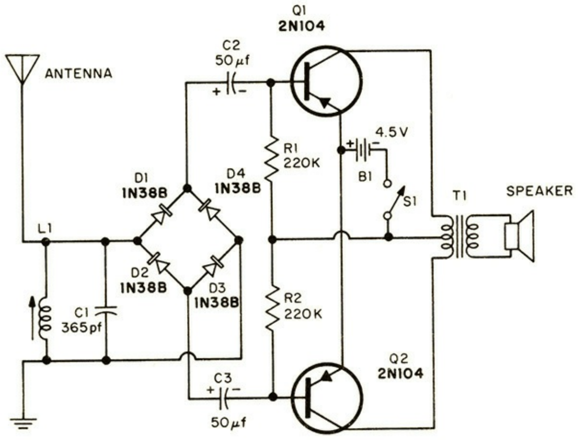
Q1
2N104
C2
50μf
+
−
ANTENNA
4.5V
+
R1
220K
B1
S1
D1
1N38B
D4
1N38B
T1
SPEAKER
L1
D2
1N38B
D3
1N38B
C1
365pf
R2
220K
Q2
2N104
C3
+
−
50μf

30 - Rádio OM e OC com 1 Transistor

Este receptor tradicional com dois jogos de bobinas foi encontrado numa documentação dos anos 70. O circuito funciona tanto com transistores de germânio como silício e o fone deve ser de alta impedância. Para ondas médias enrole 20 + 80 espiras de fio 28 num tubo de PVC de 1 polegada e para ondas curtas 5 + 265 espiras do mesmo fio.

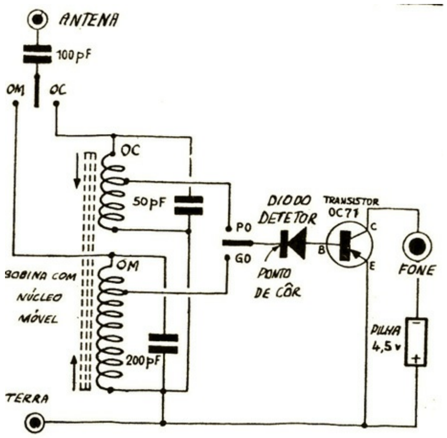
ANTENA
100pF
OM
OC
OC
50pF
PO
GO
DIODO DETETOR
PONTO DE CÔR
TRANSISTOR OC71
B
C
E
FONE
BOBINA COM NÚCLEO MÓVEL
OM
200pF
PILHA 4,5 v
−
+
TERRA

31 - Receptor para 11 Metros

Este receptor recebe os sinais da faixa dos 27 MHz ou 11 metros, podendo ser usado como base para um intercomunicador ou ainda um sistema de escuta. A bobina é formada por 12 espiras de fio 26 ou 28 num bastão de ferrite de 0,5 cm. O capacitor é um variável de rádio de ondas curtas. A escuta pode ser em fone ou em um amplificador.

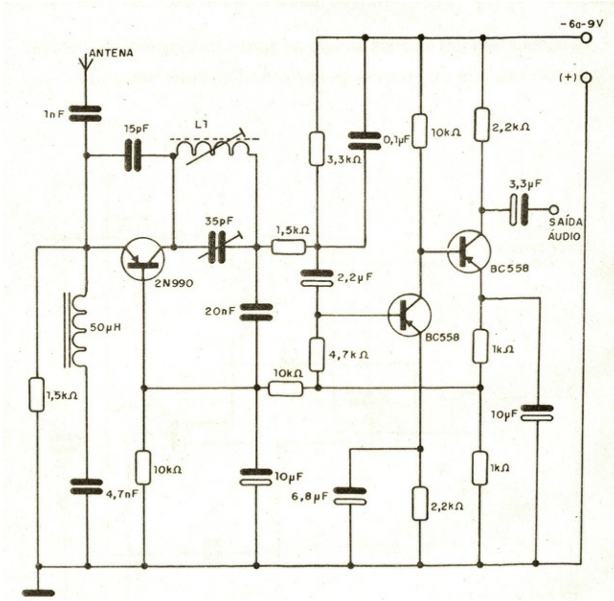
-6a-9V
(+)
ANTENA
1nF
15pF
L1
3,3kΩ
0,1µF
10kΩ
2,2kΩ
3,3µF
SAÍDA ÁUDIO
35pF
1,5kΩ
2N990
BC558
2,2µF
20nF
50µH
BC558
4,7kΩ
1kΩ
10kΩ
1,5kΩ
10µF
10kΩ
10µF
6,8µF
4,7nF
1kΩ
2,2kΩ

32 - Rádio CMOS

L1 é formada por umas 80 espiras de fio 28 em um bastão de ferrite para a captação da faixa de ondas médias. O circuito pode ser alimentado por tensões de 5 a 12 V e a saída pode ser feita numa cápsula piezoelétrica. O circuito é de uma documentação de 1978.

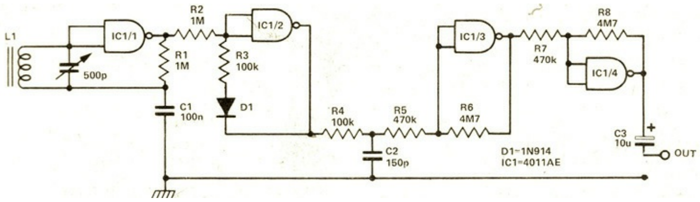
L1
500p
IC1/1
R2
1M
R1
1M
C1
100n
IC1/2
R3
100k
D1
R4
100k
C2
150p
R5
470k
R6
4M7
IC1/3
R7
470k
R8
4M7
IC1/4
C3
10u
+
OUT
D1-1N914
IC1=4011AE

33 - Receptor Reflex de Três Transistores

Este receptor reflex para três faixas de ondas foi encontrado numa documentação dos anos 70. Os transistores originais são de germânio, mas o circuito pode ser adaptado para usar transistores modernos. As bobinas dependerão das faixas de onda que se deseja captar.

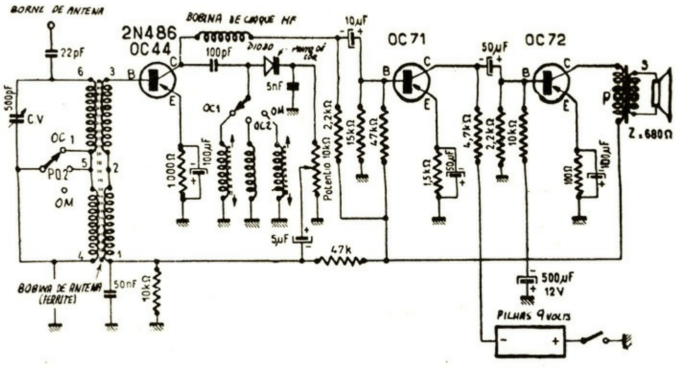
BORNE DE ANTENA
22pF
500pF
C.V
6
3
5
2
4
1
OC 1
PO2
OM
BOBINA DE ANTENA (FERRITE)
50nF
10kΩ
2N486
OC44
B
C
E
BOBINA DE CHOQUE HF
100pF
DIODO
PONTO DE COR
5nF
OC1
OC2
OM
1000Ω
100µF
5µF
Potenc. 10kΩ
2,2kΩ
10µF
15kΩ
47kΩ
47k
OC71
1,5kΩ
50µF
4,7kΩ
2,2kΩ
10kΩ
50µF
OC72
100Ω
100µF
500µF
12V
PILHAS 9 VOLTS
P
S
Z.680Ω

34 - Rádio OM e OC sem Alimentação

Encontramos este interessante circuito numa documentação dos anos 80. A bobina com os terminais 1 e 2 consta de 30 espiras de fio 28 num tubo de PVC de 1 polegada. A bobina com os terminais 3, 4 e 5 tem 30 + 70 espiras de fio 28 no mesmo tubo. O fone deve ser magnético de alta impedância. A antena deve ser longa e a ligação à terra é importante. Melhores resultados são obtidos com transistores antigos de germânio.

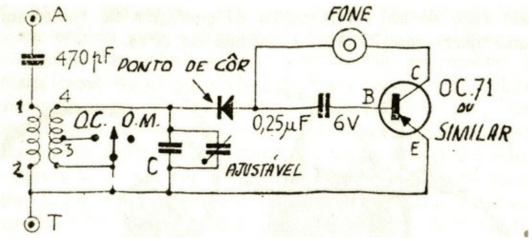
A
470pF
PONTO DE CÔR
FONE
1
4
Q.C.
O.M.
3
2
C
AJUSTÁVEL
0,25µF
6V
B
C
OC.71
ou
SIMILAR
E
T

35 - Receptor CMOS AM

Este circuito é de uma revista americana dos anos 70. L1 consiste em 80 espiras de fio 28 num bastão de ferrite e o capacitor de sintonia é obtido de um rádio AM fora de uso. T1 pode ser eliminado se na saída usarmos um fone de alta impedância ou piezoelétrico. O circuito integrado é o 4007.

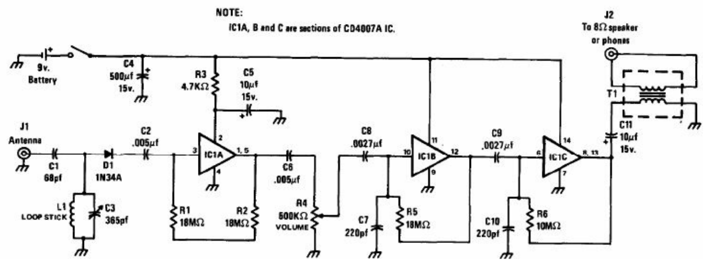
NOTE:
IC1A, B and C are sections of CD4007A IC.
J2
To 8Ω speaker or phones
9v.
Battery
C4
500μf
15v.
R3
4.7KΩ
C5
10μf
15v.
T1
J1
Antenna
C1
68pf
D1
1N34A
C2
.005μf
IC1A
3
2
1, 5
4
C6
.005μf
C8
.0027μf
IC1B
10
11
12
9
C9
.0027μf
IC1C
6
14
8, 13
7
C11
10μf
15v.
L1
LOOP STICK
C3
365pf
R1
18MΩ
R2
18MΩ
R4
500KΩ
VOLUME
C7
220pf
R5
18MΩ
C10
220pf
R6
10MΩ

36 - Rádio AM com o LM3909

O circuito apresentado foi obtido numa documentação técnica da National Semiconductor. A bobina consiste em 20 + 80 espiras de fio esmaltado 28 em bastão de ferrite de 15 a 20 cm de comprimento e o capacitor variável pode ser obtido de um rádio AM comum.

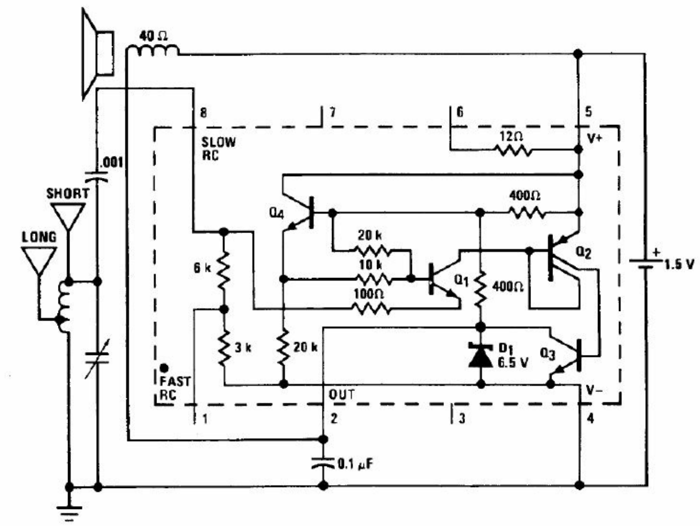
40 Ω
8
7
6
5
SLOW
RC
12Ω
V+
.001
SHORT
LONG
Q4
400Ω
20 k
10 k
100Ω
Q1
400Ω
Q2
+
1.5 V
6 k
3 k
20 k
D1
6.5 V
Q3
FAST
RC
OUT
V−
1
2
3
4
0.1 μF

37 - Rádio AM de 3 Transistores

Este receptor é de uma antiga revista inglesa de 1965. Os transistores usados são de germânio, não mais comuns no nosso mercado, mas adaptações podem ser feitas para funcionamento com transistores atuais de silício como os BC558. L1 consiste em 50 voltas de fio 28 num bastão de ferrite e L2 em 12 voltas do mesmo fio. O circuito terá mais sensibilidade com transistores originais de germânio. O fone é piezoelétrico ou de cristal. Uma antena pode ser ligada à bobina.

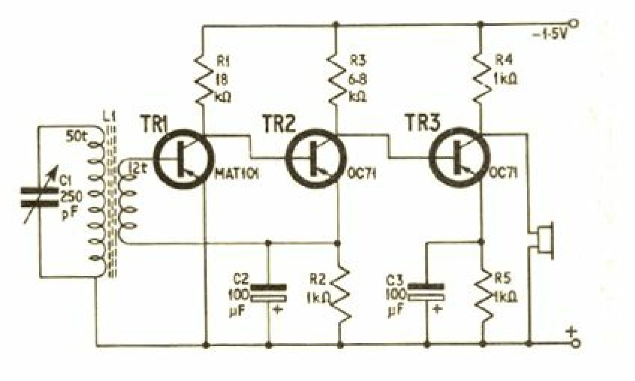
-1·5V
R1
18
kΩ
R3
6·8
kΩ
R4
1kΩ
L1
50t
TR1
TR2
TR3
12t
MAT101
OC71
OC71
C1
250
pF
C2
100
μF
+
R2
1kΩ
C3
100
μF
+
R5
1kΩ
+

38 - Receptor Regenerativo de Ondas Médias

Este rádio de ondas médias com transistores de germânio é de uma publicação da Raytheon de 1957. Transistores equivalentes modernos podem ser experimentados e o fone de ouvido deve ser de alta impedância. L1 tem 30 + 70 espiras de fio 28 e L2 10 espiras do mesmo fio no mesmo bastão. A antena deve ser longa.

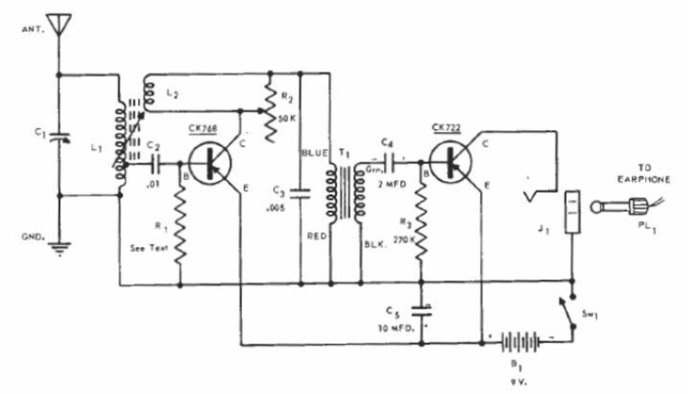
ANT.
GND.
C_1
L_1
L_2
C_2
.01
R_1
See Text
CK768
B
C
E
R_2
50K
BLUE
RED
C_3
.005
T_1
C_4
Grn.
2 MFD
BLK.
R_3
270K
CK722
B
C
E
C_5
10 MFD.
J_1
TO EARPHONE
PL_1
Sw_1
B_1
9 V.

R_1 — ½ watt carbon resistor (see text).
R_2 — 50K potentiometer, with switch.
R_3 — 270K, ½ watt carbon resistor.
C_1 — 365 Mmf. tuning capacitor.
C_2 — .01 Mfd., disc ceramic capacitor.
C_3 — .005 Mfd., disc ceramic capacitor.
C_4 — 2 Mfd., 8 volt electrolytic capacitator.
C_5 — 10 Mfd., 10 volt electrolytic capacitor.
L_1 — Tapped antenna coil (*Meissner* No. 14-9015 or *Argonne* No. AR70).
L_2 — Feedback coil (see text).
T_1 — Interstage transistor transformer (*Argonne* No. AR-109 or *Thordarson* No. TR-7.
Sw_1 — SPST switch (on R_2).
J_1 — Subminiature open circuit jack.

39 - Receptor Super Heteródino de 1957

Como as bobinas e transformadores usados neste circuito são críticos, não sendo mais fabricado, o esquema serve apenas de orientação ou curiosidade para um trabalho. O receptor para a faixa de ondas médias foi encontrado num manual da Raytheon da época.

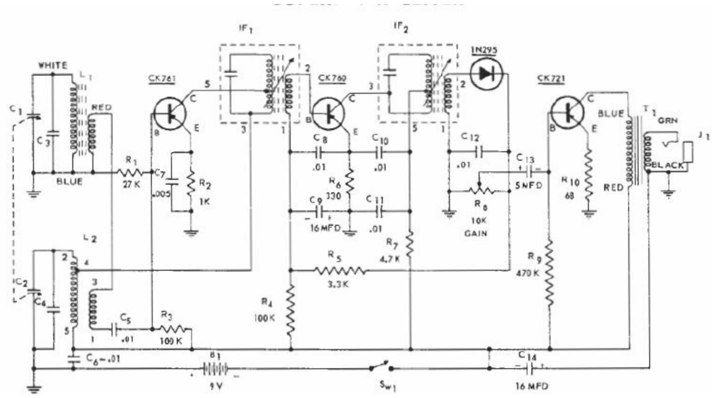
IF1
IF2
1N295
WHITE
L1
CK761
CK760
CK721
C1
C3
RED
BLUE
R1
27 K
C7
.005
R2
1K
C8
.01
C10
.01
R6
330
C9
16 MFD
C11
.01
C12
.01
C13
5 MFD
R8
10K
GAIN
R10
68
BLUE
T1
GRN
J1
BLACK
RED
L2
C2
C4
C5
.01
R3
100 K
R4
100 K
R5
3.3K
R7
4.7 K
R9
470 K
C6 -.01
B1
9 V
Sw1
C14
16 MFD
B
C
E
1
2
3
4
5

40 - Rádio Comercial da Década de 70

Este circuito de 6 transistores tinha excelente sensibilidade usando transistores de germânio. A etapa em push-pull usava dois transformadores, sendo um drive e outro de saída como encontramos normalmente nos receptores da época. O receptor é do tipo super-heteródino.

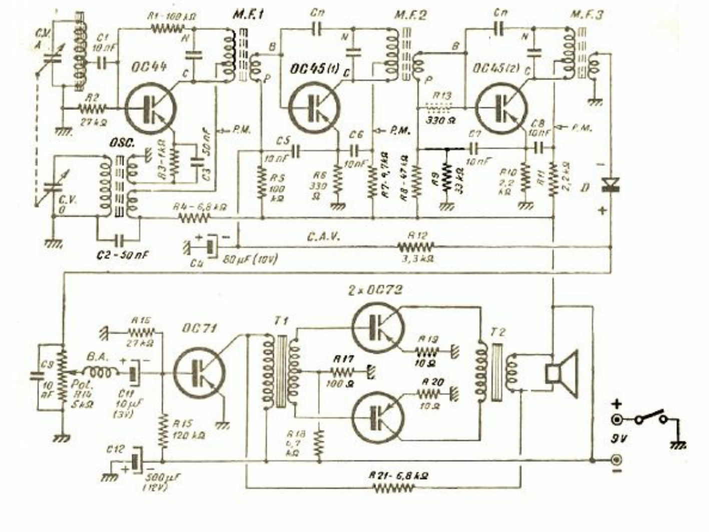
C.V. A
C1 10 nF
R1 - 100 kΩ
M.F.1
N
C
B
P
OC44
R2 27 kΩ
OSC.
R3 - 1 kΩ
C3 50 nF
P.M.
C.V. 0
R4 - 6,8 kΩ
C2 - 50 nF
Cn
M.F.2
OC45 (1)
C5 10 nF
C6 10 nF
R5 100 kΩ
R6 330 Ω
R7 - 4,7 kΩ
R8 - 47 kΩ
Cn
M.F.3
OC45 (2)
R13 330 Ω
C7 10 nF
C8 10 nF
R9 33 kΩ
R10 2,2 kΩ
R11 2,2 kΩ
D
+
-
C4
80 µF (10V)
C.A.V.
R12 3,3 kΩ
2 x OC72
C9 10 nF
Pot. R14 5 kΩ
B.A.
R16 27 kΩ
C11 10 µF (3V)
OC71
R15 120 kΩ
C12
500 µF (12V)
T1
R17 100 Ω
R18 4,7 kΩ
R19 10 Ω
R20 10 Ω
T2
R21 - 6,8 kΩ
9V

41 - Receptor AM Para Fone Sem Fio

Este simples receptor para sinais de um transmissor AM modulado com áudio de um televisor, por exemplo, alimenta um fone de ouvido de alta impedância. O circuito é do manual de aplicações do CK722 da década de 50 da Raytheon.

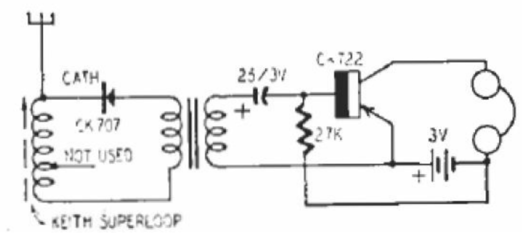
CATH
CK 707
NOT USED
KEITH SUPERLOOP
25/3V
+
CK722
27K
3V
+

42 - Receptor AM Estéreo com o MC13024

O sistema estéreo AM QAM funciona experimentalmente com algumas estações, mas não deve ser adotado, pois o rádio digital deve ocupar seu lugar. O circuito apresentado é do Linear Databook da Motorola de 1990.

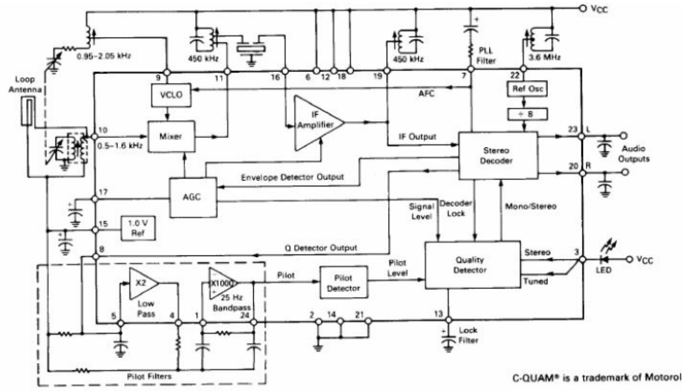
VCC
0.95–2.05 kHz
450 kHz
450 kHz
PLL Filter
3.6 MHz
Loop Antenna
9
11
16
6
12
18
19
7
22
VCLO
AFC
Ref Osc
÷ 8
IF Amplifier
10
0.5–1.6 kHz
Mixer
IF Output
Stereo Decoder
23
L
20
R
Audio Outputs
Envelope Detector Output
AGC
17
15
1.0 V Ref
Signal Level
Decoder Lock
Mono/Stereo
Q Detector Output
8
Quality Detector
Stereo
Tuned
3
LED
VCC
X2
Low Pass
X1000
25 Hz Bandpass
Pilot
Pilot Detector
Pilot Level
5
4
1
24
2
14
21
13
Lock Filter
Pilot Filters

C-QUAM® is a trademark of Motorola.

43 - Receptor Para 11 Metros

Este receptor recebe os sinais da faixa de PX ou Rádio Cidadão em 27 MHz. O circuito é de uma publicação americana dos anos 60 e o componente crítico é a bobina. A bobina é formada por 8 + 8 espiras de fio 22 em forma de 1,5 cm sem núcleo. O transformador T1 é de rádios transistorizados (driver) e a saída é em fone de alta impedância.

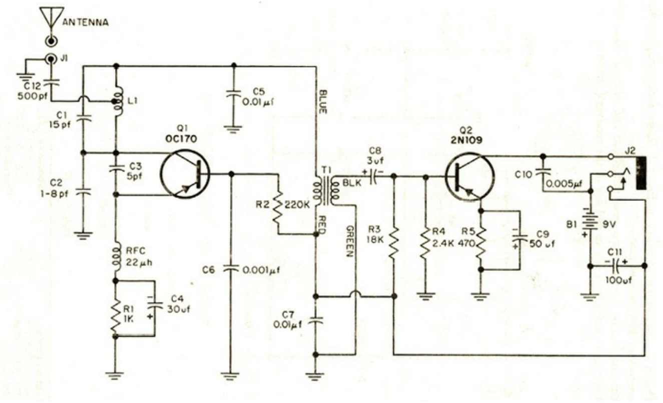
ANTENNA
J1
C12
500pf
L1
C1
15pf
C5
0.01µf
BLUE
Q1
OC170
C3
5pf
C2
1-8pf
T1
BLK
C8
3uf
Q2
2N109
J2
C10
0.005µf
R2
220K
RED
GREEN
R3
18K
R4
2.4K
R5
470
C9
50uf
B1
9V
C11
100uf
RFC
22µh
C6
0.001µf
R1
1K
C4
30uf
C7
0.01µf

44 - Rádio AM com o TBA820

Este receptor de AM de baixo custo é do manual de circuitos integrados lineares da SGS. O fio da bobina pode ser de 26 a 30 e o capacitor variável pode ser obtido num rádio AM fora de uso. Pode ser usada antena externa para maior sensibilidade e o diodo pode ser o 1N34.

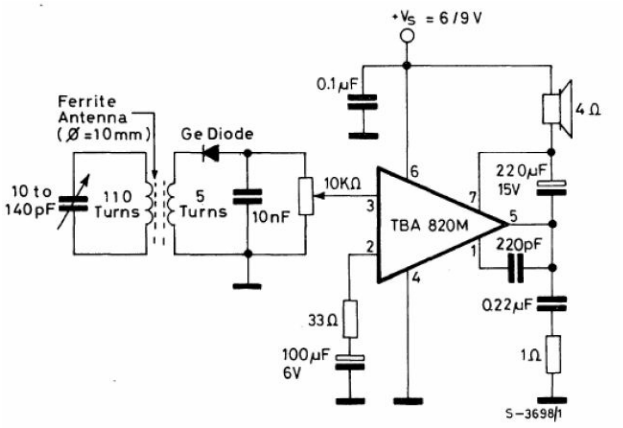
+Vs = 6/9 V
0.1μF
Ferrite Antenna (Ø=10mm)
Ge Diode
10 to 140pF
110 Turns
5 Turns
10nF
10KΩ
TBA 820M
6
3
2
4
7
5
1
4Ω
220μF 15V
220pF
0.22μF
1Ω
33Ω
100μF 6V
S-3698/1

45 - Transmissor OM Experimental

Este circuito foi obtido numa revista Italiana Sperimentare de 1975. A bobina consta de 100 espiras de fio 28 num bastão de ferrite e os transistores podem ser equivalentes mais modernos como os BC548. A sintonia pode ser feita em C5, trocando por um capacitor variável de um velho rádio de ondas médias. A antena telescópica pode ser ligada ao coletor de TR3.

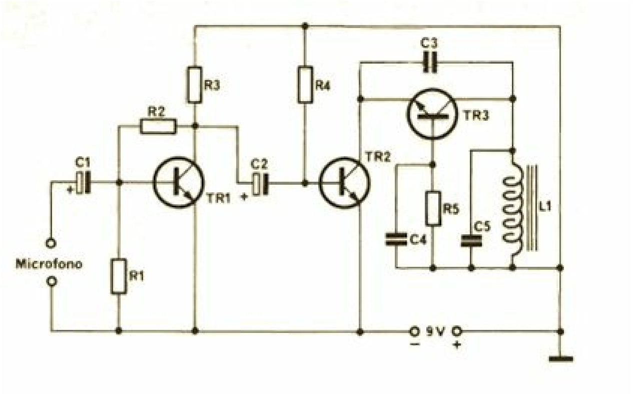

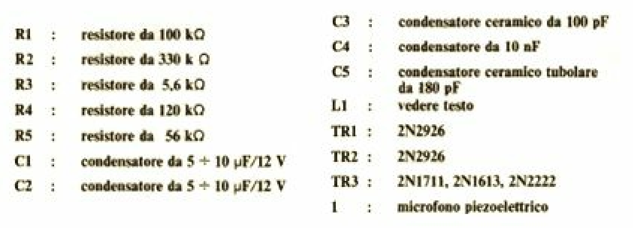

R1	:	resistore da 100 kΩ
R2	:	resistore da 330 k Ω
R3	:	resistore da 5,6 kΩ
R4	:	resistore da 120 kΩ
R5	:	resistore da 56 kΩ
C1	:	condensatore da 5 ÷ 10 μF/12 V
C2	:	condensatore da 5 ÷ 10 μF/12 V
C3	:	condensatore ceramico da 100 pF
C4	:	condensatore da 10 nF
C5	:	condensatore ceramico tubolare da 180 pF
L1	:	vedere testo
TR1	:	2N2926
TR2	:	2N2926
TR3	:	2N1711, 2N1613, 2N2222
1	:	microfono piezoelettrico

46 - Transmissor de Centelha

Este circuito é de uma publicação de Hugo Gernsback de 1922. O circuito gera centelhas que produzem ondas determinadas pela bobina TC. SC é um transformador de alta tensão podendo ser usada uma bobina de ignição. A alimentação pode ser feita com tensões de 6 a 12 V.

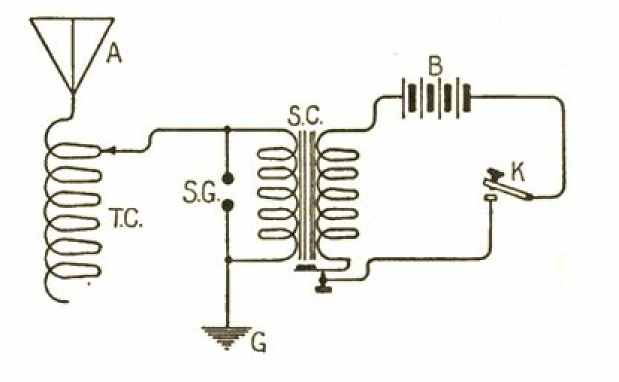

47 - Microfone sem Fio AM

Encontramos este circuito de pequeno alcance para a faixa de ondas médias numa publicação de 1964. L1 consta de 80 espiras de fio 28 num bastão de ferrite e sobre ela é enrolada L2 com 20 espiras do mesmo fio. O transistor pode ser o BC558 e a antena é telescópica de 30 a 90 cm de comprimento. O microfone deve ser de carvão. Se for usado cerâmico, ligue em paralelo um resistor de 10 k.

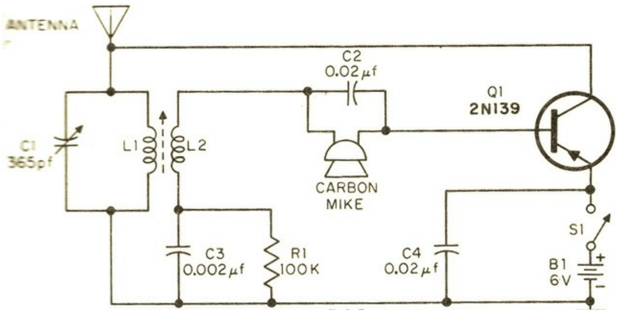
ANTENNA
C2
0.02μf
Q1
2N139
C1
365pf
L1
L2
CARBON
MIKE
S1
C3
0.002μf
R1
100K
C4
0.02μf
B1
6V
+
-

48 - Transmissor de Arco

Este primitivo transmissor do início do século 20 usa um arco-voltáico para gerar a RF aplicada à antena. O microfone deve ser do tipo de carvão e muito bem isolado, pois está submetido diretamente a alta tensão.

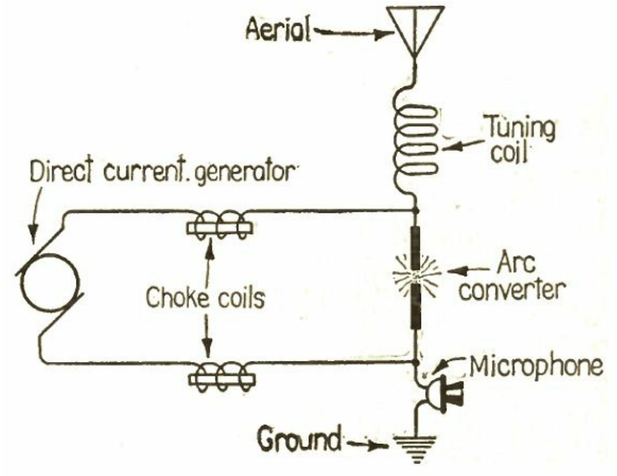
Aerial
Tuning coil
Direct current generator
Arc converter
Choke coils
Microphone
Ground

49 - Transmissor Alimentado por Energia Solar

Este pequeno transmissor envia os sinais para a faixa dos 7 MHz, mas pode ser modificado para operar em outras frequências. A bobina L1 é formada por 25 voltas de fio 28 num bastão de ferrite e o transistor pode ser o BC558. O circuito é de uma documentação de 1964.

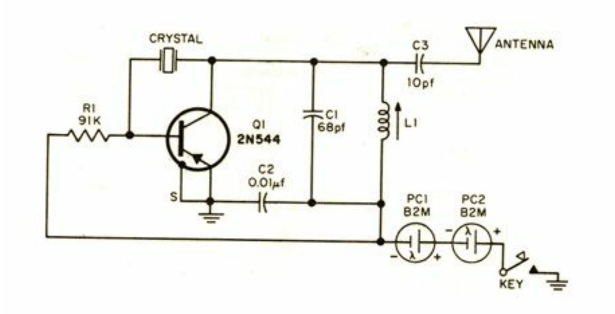
CRYSTAL
C3
10pf
ANTENNA
R1
91K
Q1
2N544
C1
68pf
L1
C2
0.01μf
S
PC1
B2M
PC2
B2M
KEY

50 - Micro Transmissor CW de Ondas Curtas

Este micro-transmissor opera numa frequência entre 3,5 MHz e 10 MHz determinada pelo cristal. O transistor pode ser um BC558 ou mesmo um NPN como o BF494 se a alimentação for invertida. A antena é do tipo telescópico com 1 m a 3 m de comprimento.

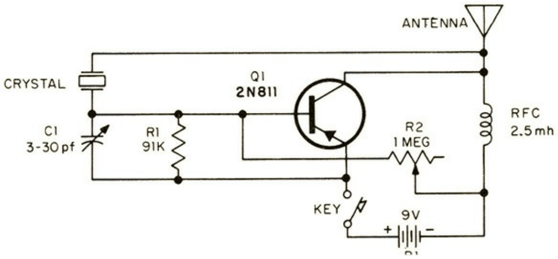
ANTENNA
CRYSTAL
Q1
2N811
C1
3-30pf
R1
91K
R2
1 MEG
RFC
2.5mh
KEY
9V
+
-
B1

51 - Transmissor de 400 uW

Este transmissor, cuja frequência é determinada pelo cristal, pode emitir sinais desde a faixa de ondas médias até a faixa de ondas curtas, com alcance de alguns metros. Os transistores podem ser os BF494 ou equivalentes, e a antena pode ter comprimentos de 1 a 5 metros. A alimentação é feita por pilhas comuns.

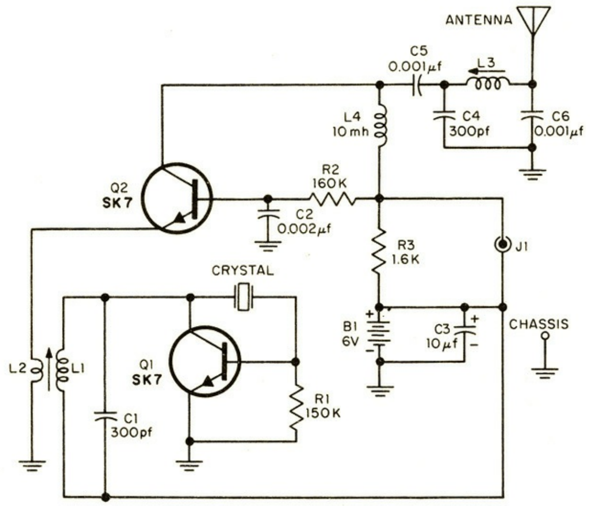
ANTENNA
C5
0.001µf
L3
L4
10 mh
C4
300pf
C6
0.001µf
Q2
SK7
R2
160K
C2
0.002µf
R3
1.6K
J1
CRYSTAL
B1
6V
C3
10µf
CHASSIS
L2
L1
Q1
SK7
R1
150K
C1
300pf

52 - Transmissor AM CK722

Este transmissor de curto alcance usa o primeiro transistor comercial do mundo lançado na década de 50 pela Raytheon. A antena deve ter pelo menos 1 m de comprimento e a entrada de áudio pode ser feita por um microfone de baixa impedância. A bobina consiste em 100 espiras de 28 num bastão de ferrite. O ajuste da posição do bastão determina a frequência.

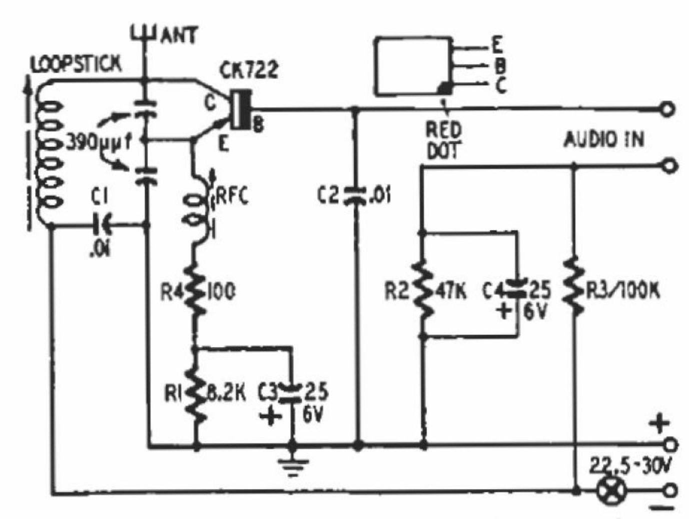
ANT
LOOPSTICK
CK722
C
B
E
390μμf
C1
.01
RFC
R4
100
R1
8.2K
C3
25
6V
+
C2
.01
E
B
C
RED
DOT
AUDIO IN
R2
47K
C4
25
6V
+
R3/100K
+
22.5-30V
−

53 - Transmissor CW AM de Curto Alcance

Este transmissor, encontrado numa documentação da Raytheon de 1957, tem um alcance de alguns metros. L1 pode ter 30 + 50 espiras de fio 28 e L2 20 espiras sobre L1 com o mesmo fio. CT pode ser um capacitor variável para se sintonizar o transmissor.

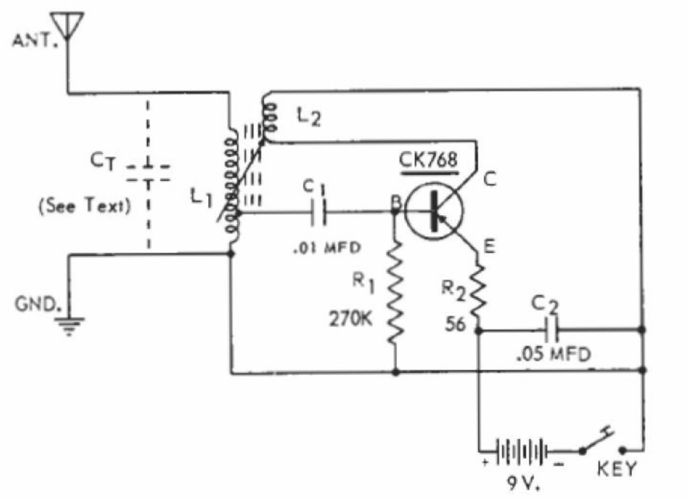
ANT.
L2
CK768
C
CT
(See Text)
L1
C1
B
.01 MFD
E
R1
270K
R2
56
C2
.05 MFD
GND.
+
9 V.
−
KEY

54 - Transmissor CW Para Ondas Médias

Este transmissor experimental para a faixa de ondas médias tem um alcance de alguns metros e a frequência é ajustada num capacitor variável de rádio antigo. O transistor pode ser o BC548 e a bobina L1 é formada por 50 + 50 espiras de fio 28 num bastão de ferrite. L2 consiste em 12 espiras do mesmo fio sobre L1. Se não oscilar inverta uma das bobinas.

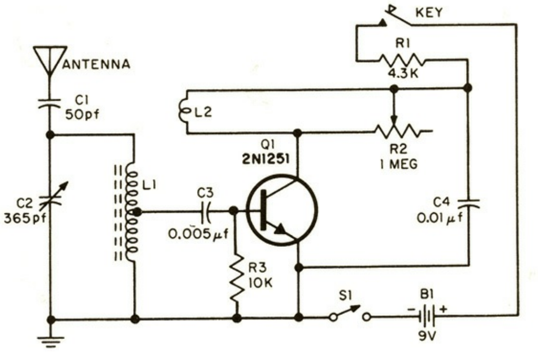
KEY
R1
4.3K
ANTENNA
C1
50pf
L2
R2
1 MEG
Q1
2N1251
L1
C2
365pf
C3
0.005μf
C4
0.01μf
R3
10K
S1
B1
9V

55 - Transmissor TTL sem Bobinas

Este circuito envia seus sinais à receptores de ondas curtas e até FM fora de estações colocados nas proximidades. Como seu alcance é muito pequeno, ele serve para demonstrações e para escuta através de paredes. A alimentação deve ser feita com tensões de 4,5 a 5 V.

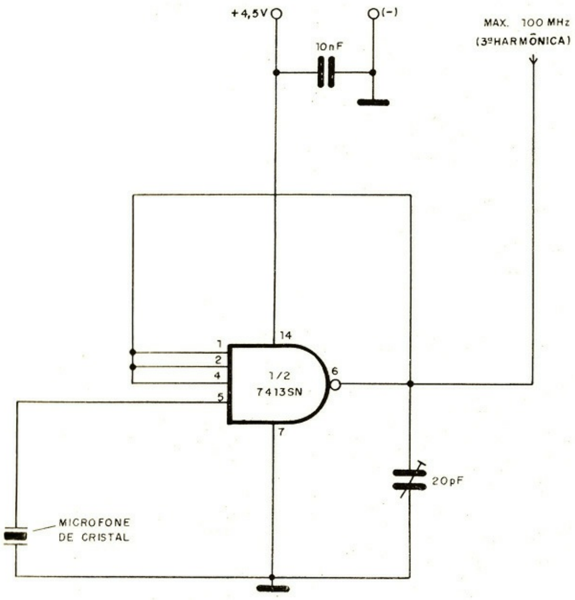
+4,5V
(-)
10nF
MAX. 100 MHz
(3ºHARMÔNICA)
14
1
2
4
5
1/2
7413SN
6
7
20pF
MICROFONE
DE CRISTAL

56 - Transmissor TTL Sem Bobinas (2)

Este transmissor experimental de curto alcance emite um sinal com muitas harmônicas, sem modulação. Ele pode ser usado como um mini-localizador e a alimentação deve ser feita com tensões de 4,5 a 5 V. A antena é um fio de 50 cm a 2 metros de comprimento.

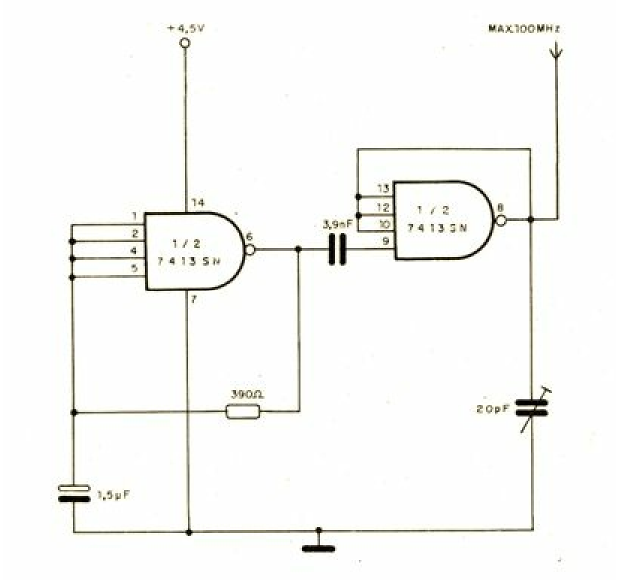
+4,5V
MAX100MHz
14
1
2
4
5
1 / 2
7 4 1 3 S N
6
7
3,9nF
13
12
10
9
1 / 2
7 4 1 3 S N
8
390Ω
20pF
1,5μF

57 - Transmissor Solar

Este pequeno transmissor de FM pode ser alimentado por células solares na faixa de 1,2 V a 6 V. Seu alcance depende das condições de propagação do local chegando a algumas dezenas de metros. A bobina é formada por 1 + 1 + 4 espiras de fio 28 ou 26 numa forma de 0,5 cm com núcleo ajustável. A antena pode ter de 15 a 30 cm de comprimento.

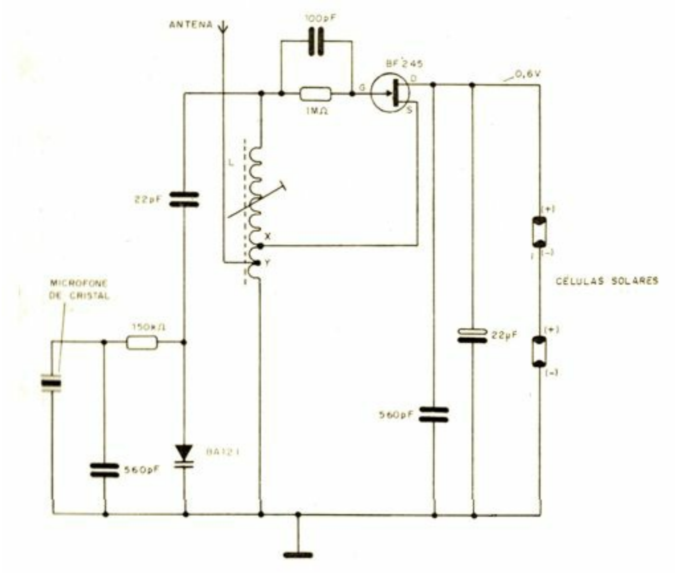
ANTENA
100pF
BF 245
G
D
S
1MΩ
0,6V
L
X
Y
22pF
(+)
(-)
CÉLULAS SOLARES
(+)
(-)
22μF
MICROFONE DE CRISTAL
150kΩ
560pF
BA121
560pF

58 - Transmissor para 27 MHz

Este circuito foi encontrado numa publicação japonesa dos anos 70. Os transistores podem ser equivalentes como o 2N2222 para o oscilador e BD135 para a etapa de potência a bobina é formada por 20 espiras de fio 28 em forma sem núcleo de 1 cm de diâmetro.

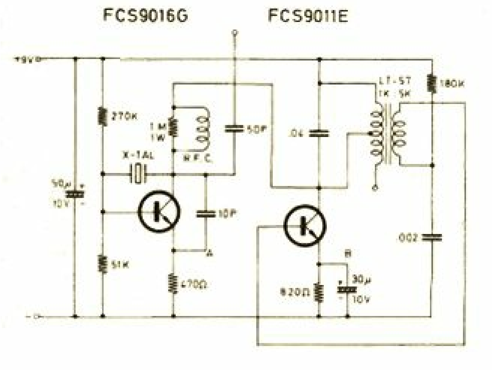
FCS9016G
FCS9011E
+9V
270K
1M
1W
R.F.C.
50P
.04
LT-57
1K : 5K
180K
X-1AL
50µ
10V
10P
A
B
51K
470Ω
820Ω
30µ
10V
.002
-

59 - Transmissor Fonográfico

Os sons de um toca-discos com fonocaptor de cristal ou cerâmico são transmitidos para um rádio de ondas médias nas proximidades. O circuito é de 1964 podendo ser implementado com um transistor BC558. Os capacitores são cerâmicos e L1 é formada por 80 espiras de fio 28 num bastão de ferrite. A antena é telescópica ou um pedaço de fio de 50 cm a 4 metros.

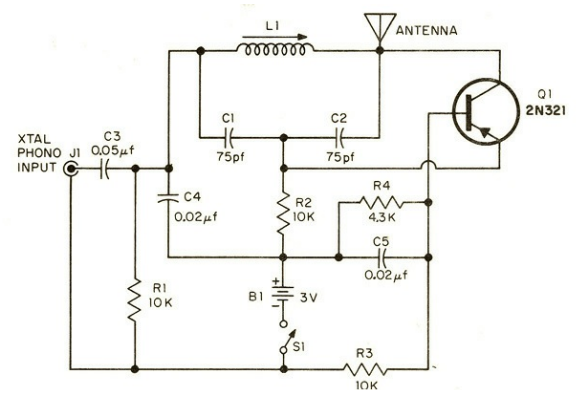
L1
ANTENNA
Q1
2N321
C1
75pf
C2
75pf
XTAL
PHONO
INPUT
J1
C3
0.05µf
C4
0.02µf
R2
10K
R4
4.3K
C5
0.02µf
R1
10K
B1
3V
S1
R3
10K

60 - Transmissor Ultra-Miniaturizado

Se montado com componentes SMD este transmissor ficará tão pequeno que poderá ser instalado dentro de uma caneta comum. A bobina consta de 6 espiras de fio 28 numa forma de 0,5 cm sem núcleo. A sintonia é feita apertando-se e distendendo-se as espiras. A alimentação é feita por uma pilha tipo botão. O circuito é de uma publicação sobre espionagem de 1993.

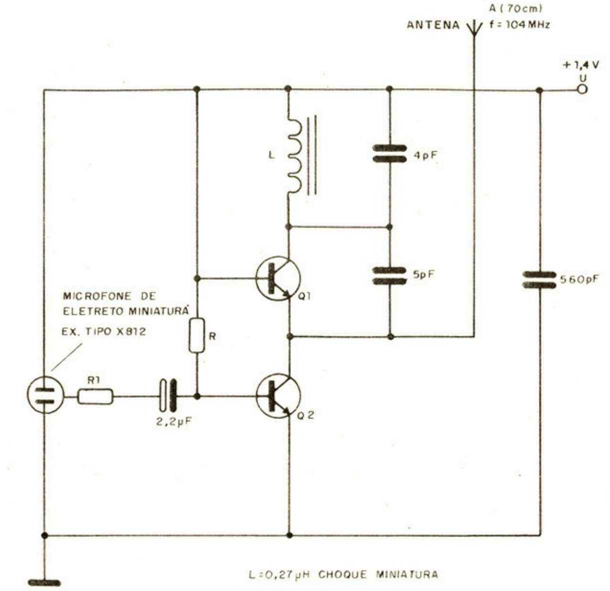
A (70cm)
ANTENA
f = 104 MHz
+1,4 V
U
L
4pF
5pF
560pF
Q1
Q2
R
R1
2,2µF
MICROFONE DE
ELETRETO MINIATURA
EX. TIPO X812

L=0,27µH CHOQUE MINIATURA

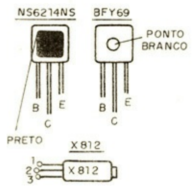
NS6214NS
BFY69
PONTO
BRANCO
B
C
E
B
C
E
PRETO
X812
1
2
3
X812

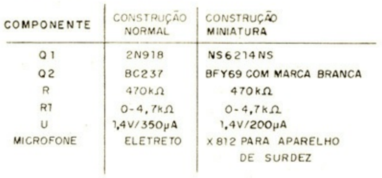

COMPONENTE	CONSTRUÇÃO NORMAL	CONSTRUÇÃO MINIATURA
Q1	2N918	NS6214NS
Q2	BC237	BFY69 COM MARCA BRANCA
R	470kΩ	470kΩ
R1	0-4,7kΩ	0-4,7kΩ
U	1,4V/350µA	1,4V/200µA
MICROFONE	ELETRETO	X812 PARA APARELHO DE SURDEZ

61 - Transmissor de Ondas Curtas XTAL

Este transmissor opera na faixa dos 80 metros, em torno de 3,5 MHz, com valor exato da frequência dado pelo cristal. L1 consta de 15 espiras de fio 28 em um bastão de ferrite de 0,8 cm de diâmetro. A tomada na bobina é central. Os capacitores devem ser cerâmicos exceto o de 2,2 uF. A antena deve ter 1,5 m.

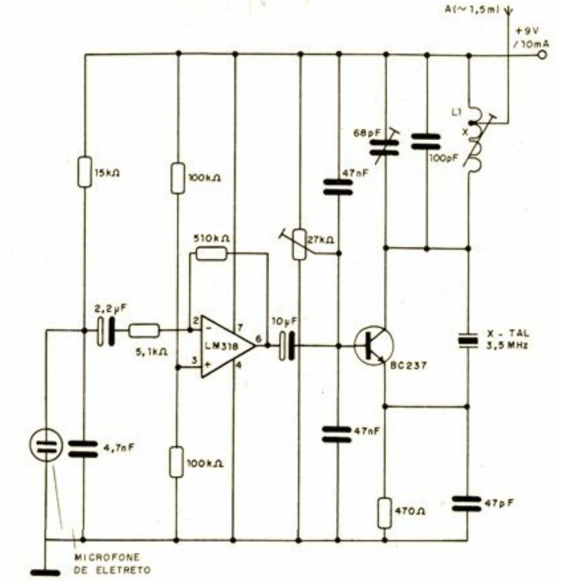
A(~1,5m)
+9V
/10mA
L1
X
68pF
100pF
47nF
15kΩ
100kΩ
510kΩ
27kΩ
2,2µF
5,1kΩ
2
3
7
6
4
LM318
10µF
BC237
X - TAL
3,5 MHz
47nF
4,7nF
100kΩ
470Ω
47pF
MICROFONE
DE ELETRETO

62 - Transmissor AM de 4 W

As bobinas deste circuito vão determinar a frequência de transmissão deste aparelho. O circuito é do manual de referência de transistores da Mullard de 1960 e faz uso de transistores de germânio.

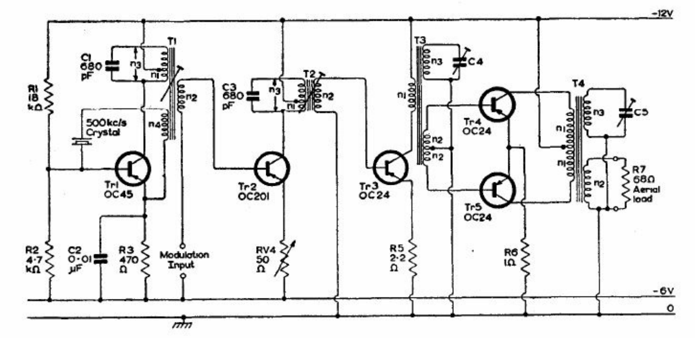
-12V
T1
C1
680
pF
n3
n1
n2
R1
18
kΩ
500kc/s
Crystal
n4
Tr1
OC45
C3
680
pF
T2
n3
n1
n2
Tr2
OC201
T3
n3
C4
n1
n2
n2
Tr3
OC24
Tr4
OC24
Tr5
OC24
T4
n3
C5
n1
n1
n2
R7
68Ω
Aerial
load
R2
4·7
kΩ
C2
0·01
μF
R3
470
Ω
Modulation
Input
RV4
50
Ω
R5
2·2
Ω
R6
1Ω
-6V
0

63 - Transmissor de Telemetria

Este circuito emite sinais numa frequência proporcional à resistência do sensor, um NTC, por exemplo. A frequência é determinada pelo cristal que pode estar entre 3 e 15 MHz tipicamente. O transistor de RF pode ser o BD136 e os demais BC558.

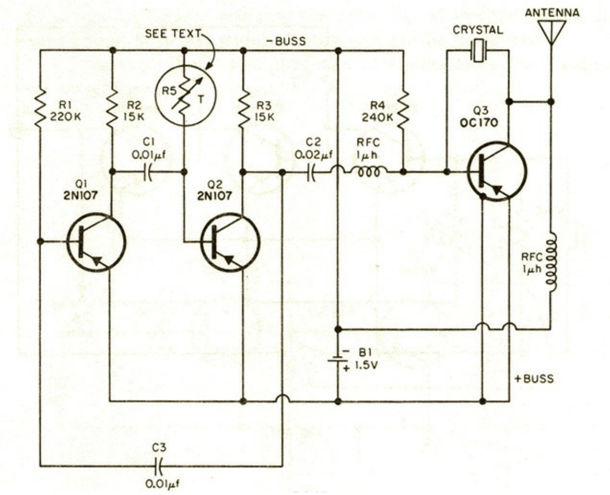
SEE TEXT
-BUSS
CRYSTAL
ANTENNA
R1
220K
R2
15K
R5
T
R3
15K
R4
240K
Q3
OC170
C1
0.01µf
C2
0.02µf
RFC
1µh
Q1
2N107
Q2
2N107
RFC
1µh
B1
1.5V
+BUSS
C3
0.01µf

64 - Transmissor para 15 Metros

Este transmissor para a faixa dos 21 MHz é controlado por um cristal. Encontramos o circuito numa documentação antiga para radioamadores. O cristal não precisa oscilar na frequência fundamental e a bobina L1 é formada por 17 espiras de fio 24 ou 22 numa forma de 2 cm com tomada central. O microfone deve ser de carvão. O circuito pode ser elaborado com um BC558 ou, invertendo-se a alimentação, com um 2N2222.

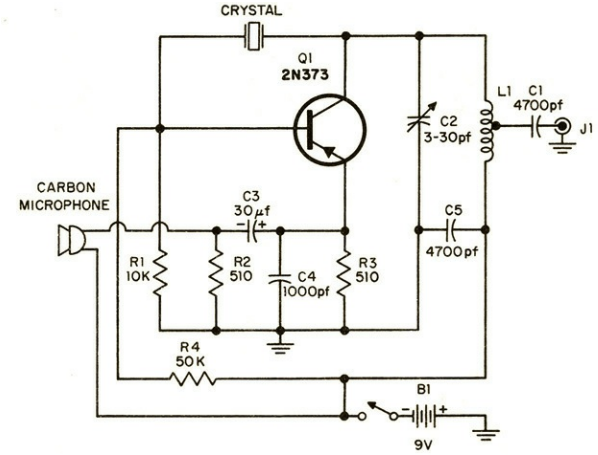
CRYSTAL
Q1
2N373
L1
C1
4700pf
J1
C2
3-30pf
CARBON
MICROPHONE
C3
30μf
C5
4700pf
R1
10K
R2
510
C4
1000pf
R3
510
R4
50K
B1
9V

65 - Transmissor Para Escuta Clandestina

O circuito apresentado emite sinais na faixa de FM, os quais podem ser captados a uma distância de algumas dezenas de metros. Os capacitores devem ser cerâmicos e o microfone deve ser de cristal ou piezoelétrico. O circuito é de uma documentação de 1993. Detalhes das bobinas estão no próprio texto.

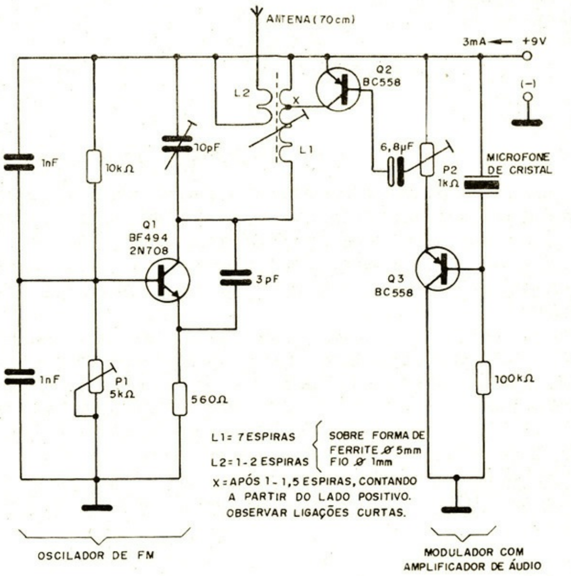
ANTENA (70cm)
3mA
+9V
(−)
Q2
BC558
L2
X
L1
10pF
1nF
10kΩ
6,8μF
P2
1kΩ
MICROFONE
DE CRISTAL
Q1
BF494
2N708
3pF
Q3
BC558
1nF
P1
5kΩ
560Ω
100kΩ
OSCILADOR DE FM
MODULADOR COM
AMPLIFICADOR DE ÁUDIO

L1= 7ESPIRAS
L2=1-2 ESPIRAS
SOBRE FORMA DE
FERRITE Ø 5mm
FIO Ø 1mm
X = APÓS 1 - 1,5 ESPIRAS, CONTANDO
A PARTIR DO LADO POSITIVO.
OBSERVAR LIGAÇÕES CURTAS.

66 - Transmissor Espião de FM

Este circuito transmite sinais para um receptor de FM colocado a uma distância até uns 200 m. A antena deve ter de 15 a 40 cm e os capacitores devem ser cerâmicos. L1 tem 4 espiras de fio 24 a 28 numa forma de 1 cm e L2 consiste em 3 espiras sobre L1.

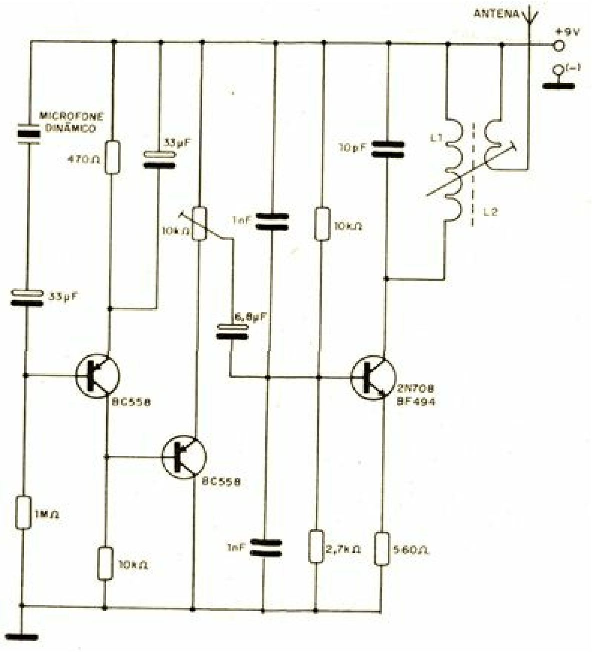
ANTENA
+9V
(-)
MICROFONE DINÂMICO
470Ω
33µF
10kΩ
1nF
10kΩ
10pF
L1
L2
33µF
6,8µF
BC558
BC558
2N708
BF494
1MΩ
10kΩ
1nF
2,7kΩ
560Ω

67 - Transmissor de FM

Este pequeno transmissor telefônico utiliza uma cápsula telefônica como microfone. A bobina L1 tem 4 a 5 espiras de fio 28 em forma de 1 cm. A alimentação pode ser feita com tensões de 6 a 9 V e o alcance é de algumas dezenas de metros. Os transistores admitem equivalentes. O circuito é de uma Nuova Elettronica de 1984

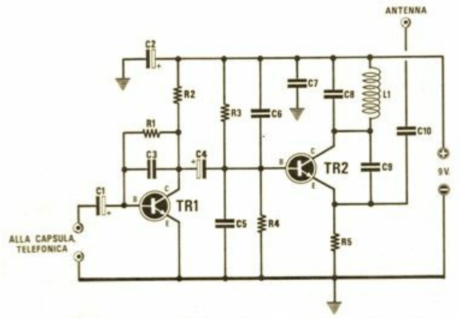
ANTENNA
C2
R2
R3
C6
C7
C8
L1
C10
R1
C3
C4
TR2
C9
9V.
C1
TR1
C5
R4
R5
ALLA CAPSULA TELEFONICA

ELENCO COMPONENTI

R1 = 470.000 ohm 1/4 watt
R2 = 12.000 ohm 1/4 watt
R3 = 12.000 ohm 1/4 watt
R4 = 12.000 ohm 1/4 watt
R5 = 270 ohm 1/4 watt
C1 = 10 mF elettrolitico 16 volt
C2 = 47 mF elettrolitico 16 volt
C3 = 330 pF a disco
C4 = 10 mF elettrolitico 16 volt
C5 = 470 pF a disco
C6 = 470 pF a disco
C7 = 10.000 pF a disco
C8 = 22 pF a disco
C9 = 2,2 pF a disco
C10 = 10 pF a disco
TR1 = transistor NPN tipo BC237
TR2 = transistor NPN tipo 2N915
L1 = vedi testo
Capsula telefonica

68 - Transmissor de FM Experimental

Encontramos este circuito numa publicação francesa dos anos 80. A bobina L1 e L2 são formadas por 4 espiras de fio 26 em forma de 1 cm de diâmetro com tomada central. T2 pode ser um 2N2222 e T3 um 2N2218 ou outro de maior potência. O circuito deve ser alimentado por fonte, dado seu consumo, ou ainda por pilhas grandes.

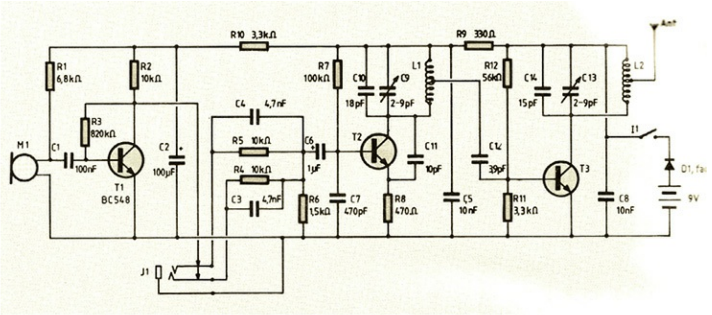
R10 3,3kΩ
R9 330Ω
Ant
R1 6,8kΩ
R2 10kΩ
R7 100kΩ
C10 18pF
C9 2-9pF
L1
R12 56kΩ
C14 15pF
C13 2-9pF
L2
C4 4,7nF
R3 820kΩ
M1
C1 100nF
T1 BC548
C2 100µF
R5 10kΩ
C6 1µF
T2
C11 10pF
C14 39pF
T3
I1
D1, fa
R4 10kΩ
C3 4,7nF
R6 1,5kΩ
C7 470pF
R8 470Ω
C5 10nF
R11 3,3kΩ
C8 10nF
9V
J1

69 - Transmissor de 27 MHz

Este transmissor foi encontrado numa documentação de 1993, mas pode ser montado com operacionais como o 741. O próprio transistor pode ser o 2N2218. O alcance depende da sensibilidade do receptor e das condições de propagação.

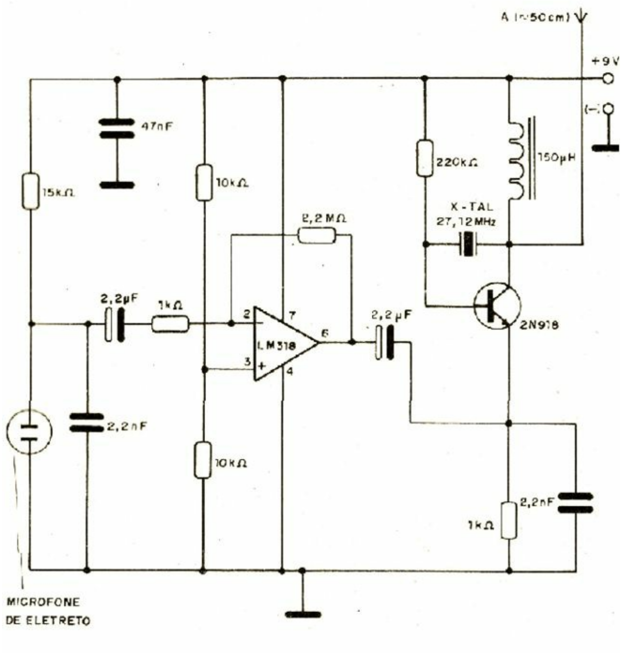
A (≈50cm)
+9V
(–)
47nF
15kΩ
10kΩ
220kΩ
150µH
X-TAL
27,12MHz
2,2MΩ
2,2µF
1kΩ
2
7
6
3
4
LM318
2,2µF
2N918
2,2nF
10kΩ
2,2nF
1kΩ
MICROFONE
DE ELETRETO

70 - Transmissor Espião com Diodo Tunnel

Este transmissor emite sinais na faixa de FM, podendo alcançar alguns metros de distância. O ajuste da modulação é feito em P1 enquanto que P2 ajusta o ponto de oscilação do circuito. A antena deve ter de 20 a 50 cm de comprimento e os capacitores menores são cerâmicos. Qualquer diodo tunnel pode ser usado. L1 e L2 como no projeto 79.

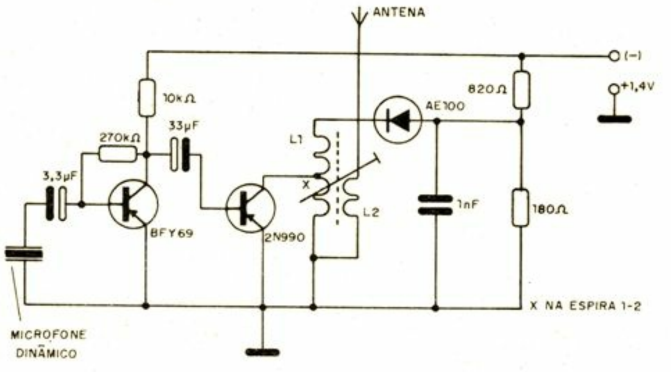
ANTENA
(-)
+1,4V
10kΩ
820Ω
AE100
270kΩ
33µF
L1
3,3µF
X
L2
1nF
180Ω
BFY69
2N990
X NA ESPIRA 1-2
MICROFONE
DINÂMICO

71 - Transmissor de FM com FET

Este pequeno transmissor de FM usa um FET comum para gerar sinais que podem ser captados a uma distância de algumas dezenas de metros. A bobina consta de 1 + 1 + 4 espiras de fio 26 ou 28 em forma com núcleo de 0,5 cm. A frequência é ajustada pela movimentação do núcleo. Os capacitores devem ser cerâmicos.

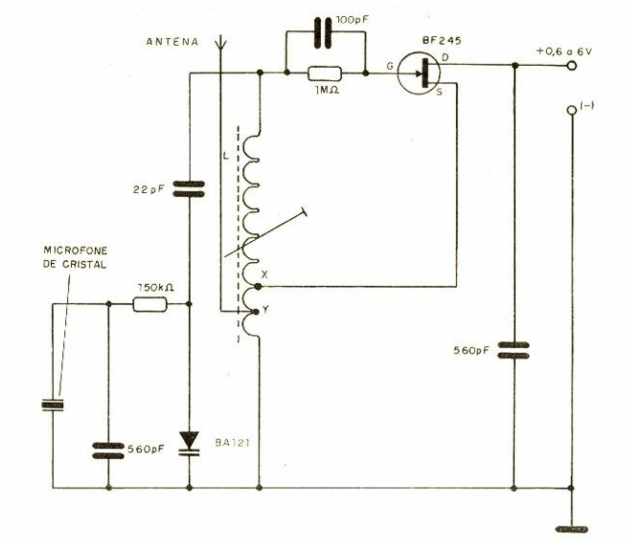
100pF
ANTENA
BF245
G
D
S
+0,6 a 6V
1MΩ
(-)
L
22pF
MICROFONE
DE CRISTAL
150kΩ
X
Y
560pF
560pF
BA121

72 - Microfone sem Fio

Este circuito, encontrado numa documentação de 1981, pode operar tanto na faixa de ondas médias como curtas, dependendo da bobina. Para ondas médias enrole 30 + 70 espiras num tubinho que pode deslizar sobre um núcleo de ferrite para a sintonia. O transistor pode ser um BC558.

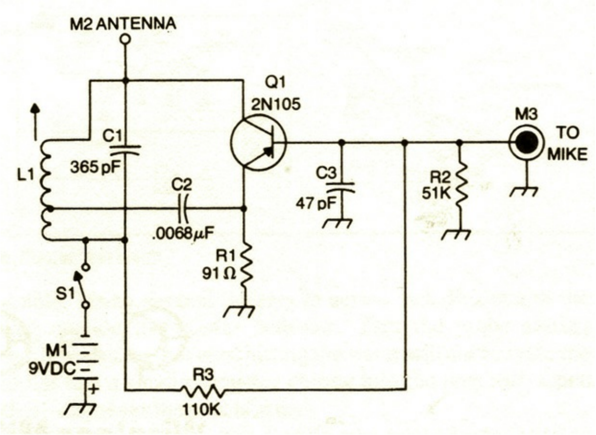
M2 ANTENNA
Q1
2N105
M3
TO
MIKE
L1
C1
365 pF
C2
.0068µF
C3
47 pF
R2
51K
R1
91 Ω
S1
M1
9VDC
+
R3
110K

73 - Transmissor de FM

Este circuito foi obtido numa publicação de 1964, mas pode ser adaptado para funcionar com componentes modernos. T1 é um transformador de áudio de rádios transistorizados. L1 consta de 4 espiras de fio 26 ou 24 em forma de 1 cm sem núcleo e L2 3 espiras do mesmo fio sobre L1. Q2 pode ser um BF494 e Q1 um BC548 invertendo a polaridade da fonte de alimentação. A alimentação pode ser feita com tensões de 6 a 9 V. C1 é de 47 pF cerâmico.

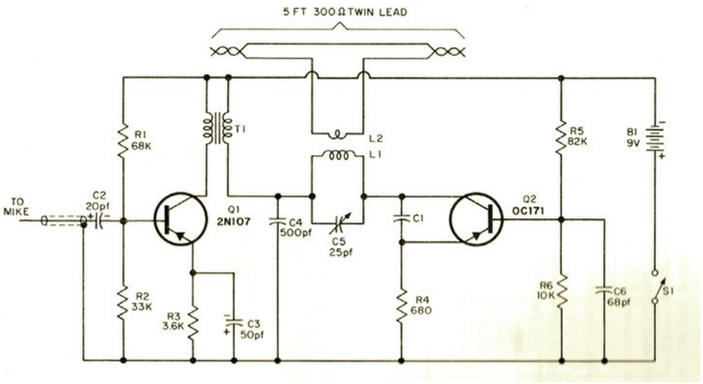
5 FT 300Ω TWIN LEAD
R1
68K
T1
L2
L1
R5
82K
B1
9V
TO
MIKE
C2
20pf
Q1
2N107
C4
500pf
C5
25pf
C1
Q2
OC171
R2
33K
R3
3.6K
C3
50pf
R4
680
R6
10K
C6
68pf
S1

74 - Transmissor de Vídeo

Para os canais baixos (2 a 6) a bobina é formada por 5 espiras de fio 20 ou 22 numa forma de 1 cm sem núcleo. Para os canais altos (7 a 13) a bobina é formada por 2 espiras do mesmo fio na mesma forma. A antena deve ter de 10 a 30 cm e os capacitores devem ser todos cerâmicos. O sinal de vídeo composto tem a modulação ajustada no trimpot. O alcance é de 20 a 50 metros.

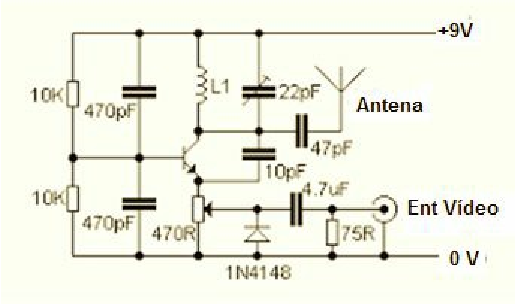
+9V
10K
470pF
L1
22pF
Antena
47pF
10pF
4.7uF
Ent Vídeo
10K
470pF
470R
75R
0 V
1N4148

75 - Transceptor Para a Faixa dos 11 metros

Este pequeno transceptor portátil para a faixa dos 11 metros (27 MHz) foi encontrado no manual de transistores da RCA de 1964. A frequência é determinada pelo cristal. Os transformadores são componentes críticos deste projeto. Transistores equivalentes podem ser empregados.

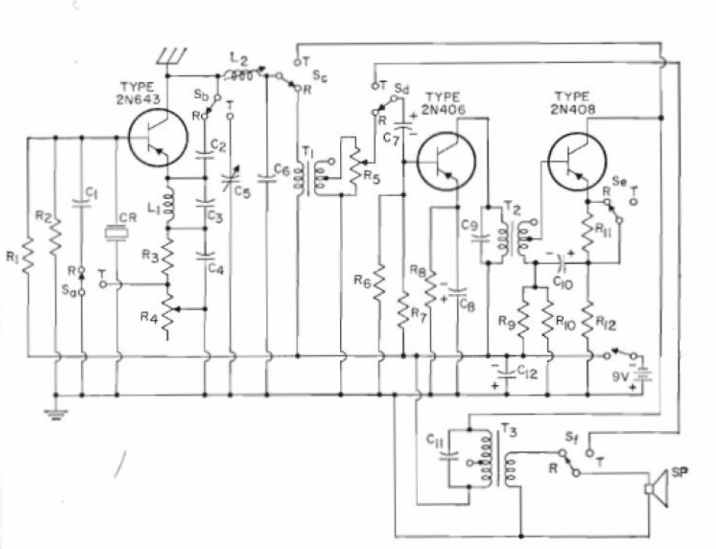

C_1 C_6 C_9 = 0.001 μf, ceramic disc
C_2 C_3 = 27 pf, mica
C_4 = 0.02 μf, ceramic disc
C_5 = 3-35 pf, trimmer
C_7 = 10 μf, electrolytic, 3 v.
C_8 C_{10} = 30 μf, electrolytic, 3 v.
C_{11} = 0.2 μf, ceramic disc
C_{12} = 200 μf, electrolytic, 10 v.
CR = crystal, 27.12 Mc (series resonant mode)
L_1 = 25 μh, radio-frequency choke
L_2 = 9 turns No. 24 enam. close-wound on ¼-inch form, ferrite slug
R_1 = 22000 ohms, 0.5 watt
R_2 = 2200 ohms, 0.5 watt
R_3 = 240 ohms, 0.5 watt
R_4 = regeneration-control potentiometer, 1000 ohms
R_5 = volume-control potentiometer, 5000 ohms
R_6 = 6800 ohms, 0.5 watt
R_7 = 56000 ohms, 0.5 watt
R_8 = 1000 ohms, 0.5 watt
R_9 = 10000 ohms, 0.5 watt
R_{10} = 560 ohms, 0.5 watt
R_{11} = 15 ohms, 0.5 watt
R_{12} = 27 ohms, 0.5 watt
S = receive-transmit switch, six-pole two-position
Sp = speaker, 14-ohm voice coil
T_1 = transformer; primary 10000 ohms; secondary 1000 ohms, center-tapped (one-half secondary used)
T_2 = transformer; primary 20000 ohms; secondary 800 ohms, center-tapped (one-half secondary used)
T_3 = transformer; primary 650 ohms, center-tapped; secondary 16 ohms

76 - Transmissor de FM com FET (9V)

Este pequeno transmissor de FM com algumas dezenas de metros de alcance, é alimentado por bateria de 9 V. A bobina consta de 1 + 1 + 4 espiras de fio 26 ou 28 em núcleo de ferrite ajustável de 0,5 cm. A antena deve ter de 15 a 50 cm de comprimento. A modulação vem de um microfone cerâmico.

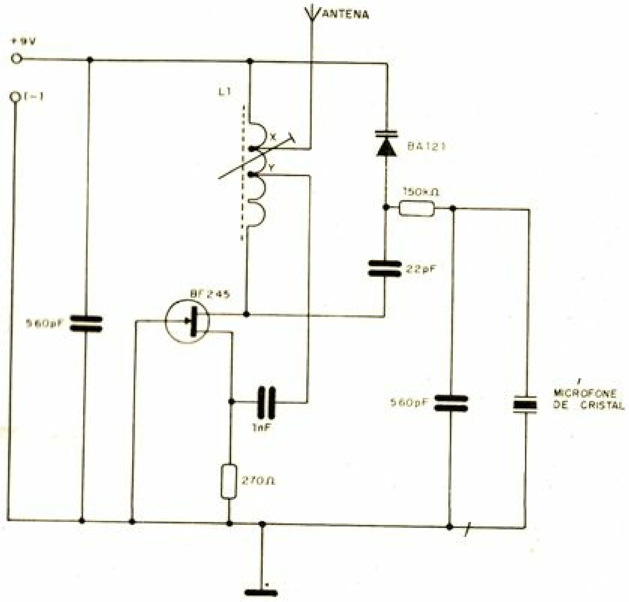
ANTENA
+9V
(-)
L1
X
Y
BA121
150kΩ
22pF
BF245
560pF
1nF
270Ω
560pF
MICROFONE
DE CRISTAL

77 - Transmissor com Diodo Tunnel (1)

Este circuito de um pequeno transmissor foi encontrado numa documentação técnica de 1964. O diodo tunnel é o componente crítico podendo ser testados equivalentes. O circuito opera na faixa dos 50 MHz (6 metros). A bobina L1 é formada por 4 espiras de fio 16 em forma de1,5 cm sem núcleo e L2 por 1 ou 2 espiras sobre L1.

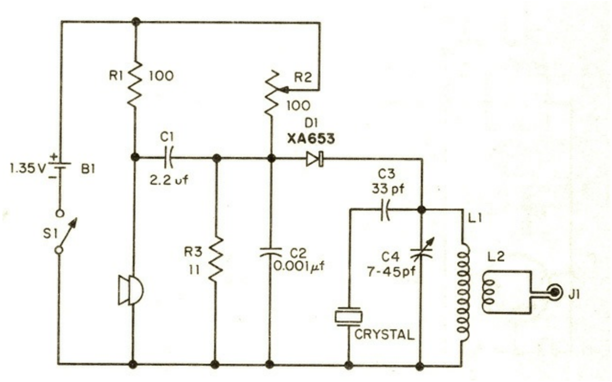
R1
100
R2
100
D1
XA653
C1
2.2 uf
1.35V
+
-
B1
S1
C3
33 pf
L1
R3
11
C2
0.001µf
C4
7-45pf
L2
J1
CRYSTAL

78 - Transmissor com Diodo Tunnel (2)

Encontramos este circuito experimental numa documentação de 1964. O diodo tunnel admite equivalente e o cristal é conforme a frequência desejada. L1 é formada por 18 voltas de fio 28 ou 30 em bastão de ferrite e L2 é formada por 5 voltas do mesmo fio sobre L1. O transformador tem secundário de 6 + 6 V com 50 mA ou mais. O potenciômetro deve ser de fio.

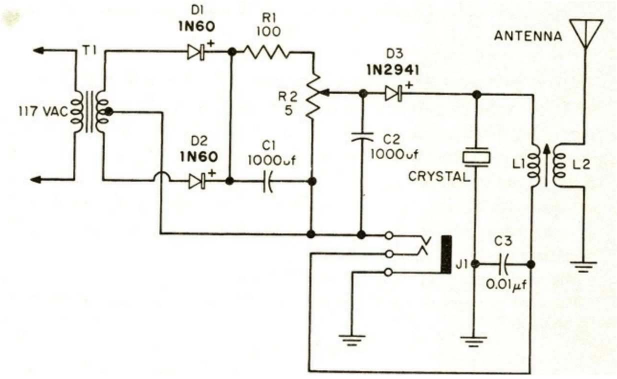
D1
1N60
R1
100
ANTENNA
T1
D3
1N2941
R2
5
117 VAC
C2
1000uf
D2
1N60
C1
1000uf
L1
L2
CRYSTAL
C3
J1
0.01μf

79 - Transmissor Espião com Diodo Tunnel

Este transmissor emite sinais na faixa de FM, podendo alcançar alguns metros de distância. O ajuste da modulação é feito em P1 enquanto que P2 ajusta o ponto de oscilação do circuito. A antena deve ter de 20 a 50 cm de comprimento e os capacitores menores são cerâmicos. Qualquer diodo tunnel pode ser usado.

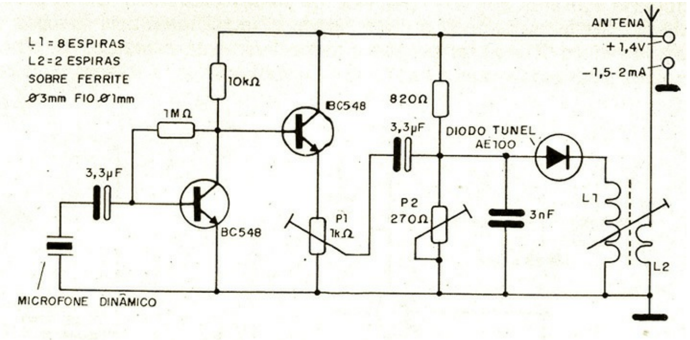
L1 = 8 ESPIRAS
L2 = 2 ESPIRAS
SOBRE FERRITE
Ø3mm FIO Ø1mm
1MΩ
10kΩ
BC548
3,3μF
BC548
P1
1kΩ
820Ω
3,3μF
DIODO TUNEL
AE100
P2
270Ω
3nF
L1
L2
ANTENA
+1,4V
-1,5-2mA
MICROFONE DINÂMICO

80 - Transmissor Espião com Diodo Tunnel (2)

Este transmissor emite sinais na faixa de FM, podendo alcançar alguns metros de distância. O ajuste da modulação é feito em P1 enquanto que P2 ajusta o ponto de oscilação do circuito. A antena deve ter de 20 a 50 cm de comprimento e os capacitores menores são cerâmicos. Qualquer diodo tunnel pode ser usado.

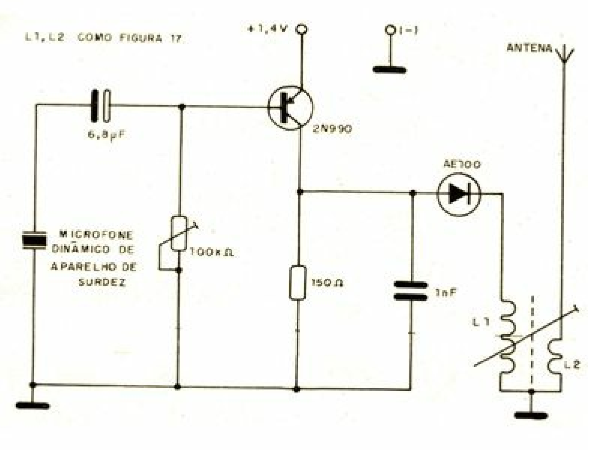
L1, L2 COMO FIGURA 17
+1,4 V
(-)
ANTENA
6,8 pF
2N990
AE100
MICROFONE DINÂMICO DE APARELHO DE SURDEZ
100 kΩ
150 Ω
1 nF
L1
L2

81 - Transmissor Espião de FM com Varicap

Encontramos este circuito num pequeno manual de espionagem de 1993. Os componentes são atuais e as bobinas fáceis de enrolar. L1 consta de 4 espiras de fio 26 ou 28 em forma de 1 cm sem núcleo e L2 mais 3 espiras sobre L1. A sintonia pode ser feita com o deslocamento do núcleo ou apertando-se e distendendo-se as espiras da bobina.

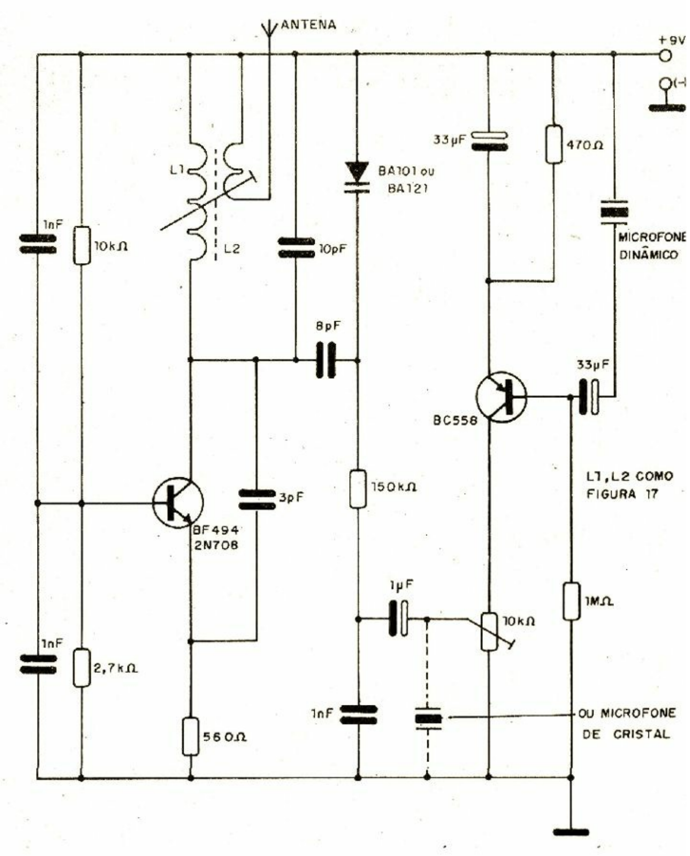
ANTENA
+9V
1nF
10kΩ
L1
L2
10pF
BA101 ou
BA121
33μF
470Ω
MICROFONE
DINÂMICO
8pF
33μF
BC558
3pF
BF494
2N708
150kΩ
L1,L2 COMO
FIGURA 17
1μF
10kΩ
1MΩ
1nF
2,7kΩ
1nF
560Ω
OU MICROFONE
DE CRISTAL

82 - Transmissor Espião Ultra Miniatura

Este transmissor pode ser montado numa caixa extremamente pequena, principalmente se forem utilizados componentes SMD. A alimentação é feita com uma pilha botão e o alcance pode superar os 10 metros, dependendo da sensibilidade do receptor. L1 consta de 4 a 6 espiras de fio 26 ou 28 em forma de 1 cm sem núcleo ou com pequeno núcleo de ferrite e tomada na terceira espira para a antena.

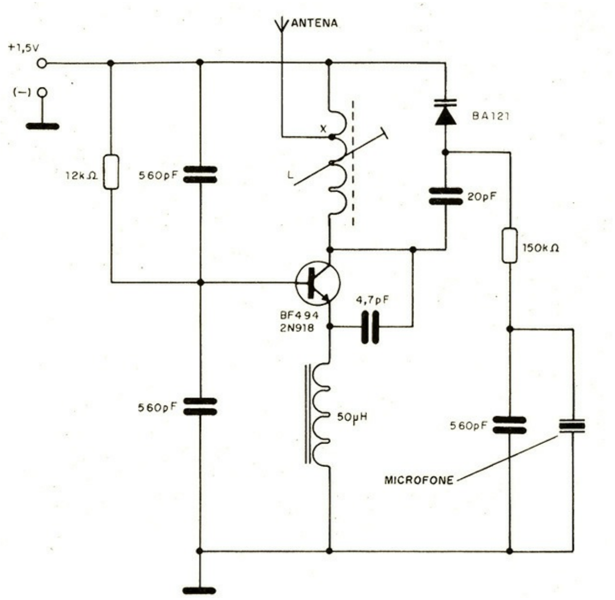
ANTENA
+1,5V
(–)
BA121
X
L
12kΩ
560pF
20pF
150kΩ
4,7pF
BF494
2N918
560pF
50μH
560pF
MICROFONE

83 - Micro Transmissor de FM de 1 Pilha

Este circuito pode ser alimentado por uma pilha apenas, seja ela comum ou do tipo botão. A bobina consta de 4 + 2 espiras de fio 28 ou 30 em núcleo de ferrite ajustável. A antena tem de 15 a 30 cm de comprimento e o alcance é de algumas dezenas de metros.

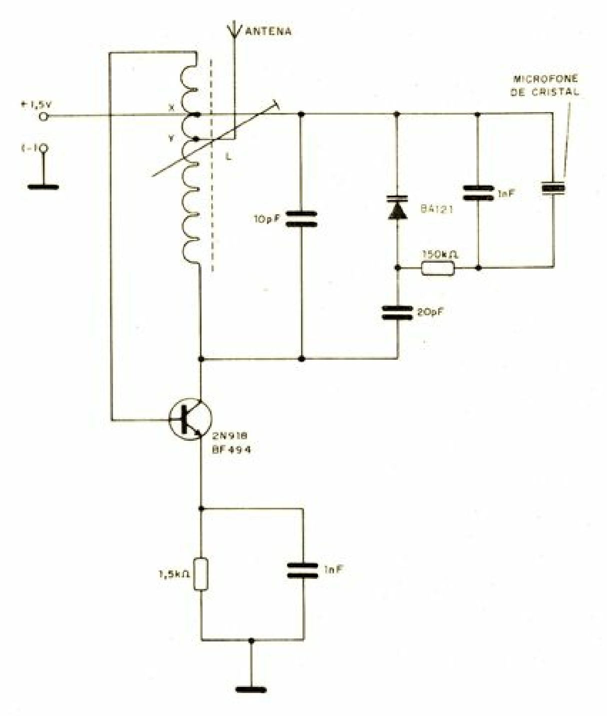
ANTENA
+1,5V
(-)
X
Y
L
MICROFONE DE CRISTAL
10pF
BA121
1nF
150kΩ
20pF
2N918
BF494
1,5kΩ
1nF

84 - Transmissor de FM de 200 mW

Este transmissor tem um alcance de algumas centenas de metros dependendo das condições do local. A bobina consta de 1 + 1 + 4 ou 5 espiras de fio 28 ou 26 em forma de 0,5 cm com núcleo ajustável. A alimentação é feita com 9 V e a antena de tem de 60 cm a 1 m de comprimento.

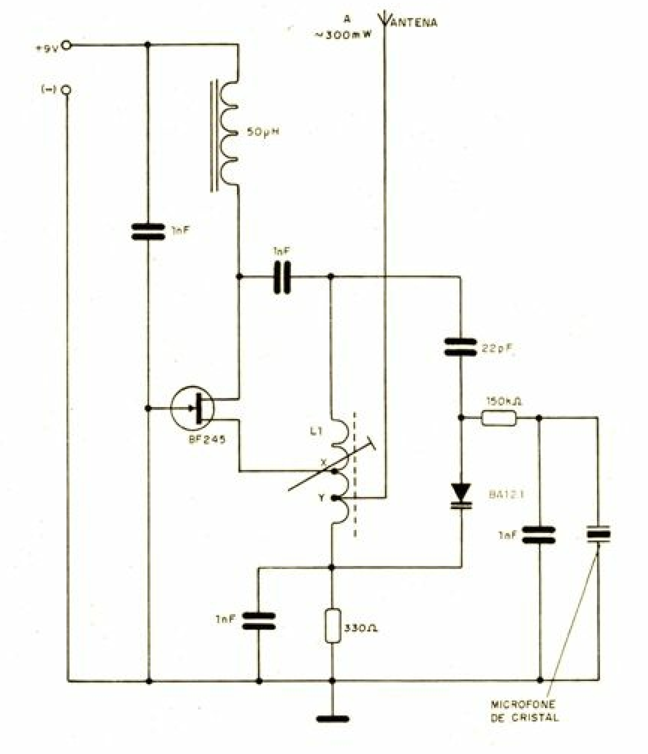
A
~300mW
ANTENA
+9V
(–)
50μH
1nF
1nF
22pF
150kΩ
BF245
L1
X
Y
BA121
1nF
1nF
330Ω
MICROFONE
DE CRISTAL

85 - Transmissor de FM com Dois FETs

O circuito mostrado é de uma documentação de 1993. Ele usa dois FETs numa configuração osciladora que opera em torno de 100 MHz. A bobina é formada por 4 a 6 espiras de fio 28 ou 26 em forma de 0,5 cm com núcleo ajustável. A antena tem de 30 a 80 cm de comprimento.

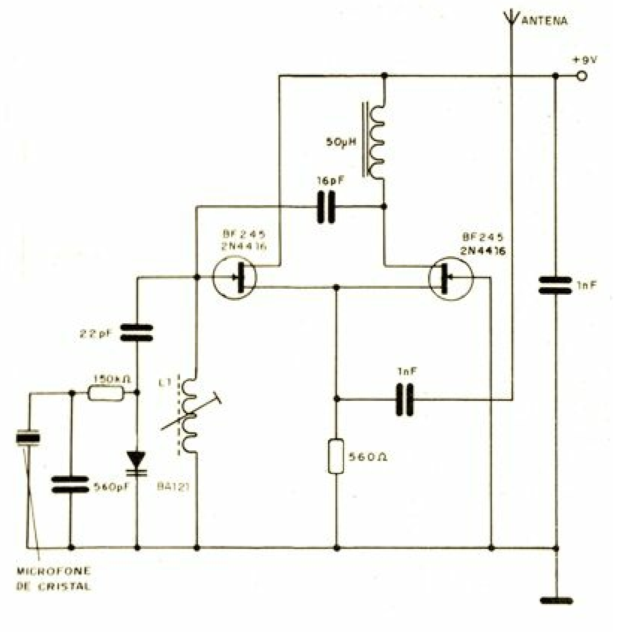
ANTENA
+9V
50µH
16pF
BF245
2N4416
BF245
2N4416
1nF
22pF
150kΩ
L1
1nF
560Ω
560pF
BA121
MICROFONE
DE CRISTAL

86 - Transmissor de FM de 300 mW com FET

Este potente transmissor de FM com FET pode alcançar mais de 1 km em condições favoráveis. L1 consta de 7 espiras, com tomada na terceira espira, de fio 28 ou 26 em forma com núcleo de ferrite ajustável de 0,8 cm e L2 por 3 espiras sobre L1. Os capacitores devem ser cerâmicos e a alimentação feita com tensões de 4,5 a 12 V.

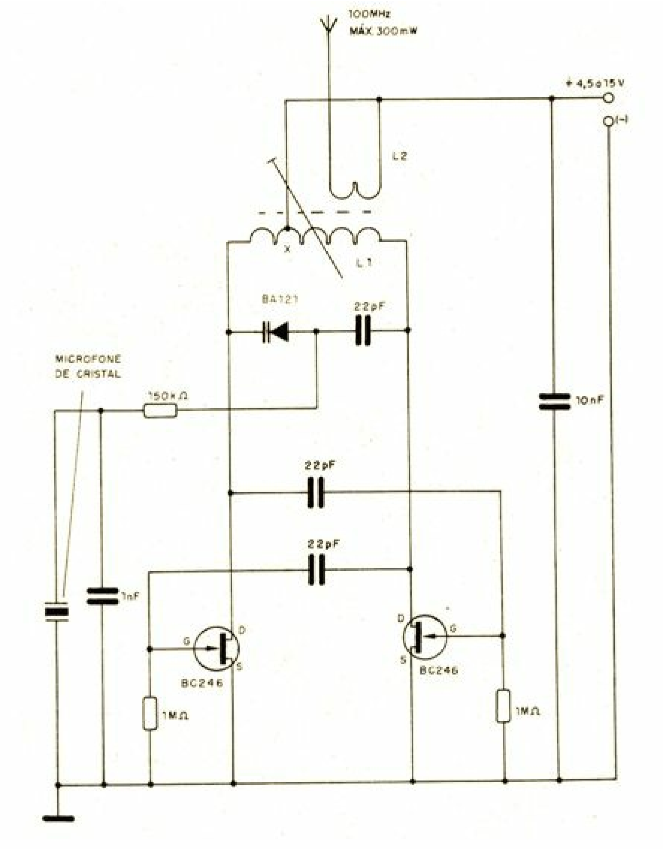
100MHz
MÁX 300mW
+ 4,5 a 15 V
(–)
L2
X
L1
BA121
22pF
MICROFONE
DE CRISTAL
150kΩ
10nF
22pF
22pF
1nF
G
D
S
BC246
1MΩ
D
G
S
BC246
1MΩ

87 - Transmissor Espião

Este transmissor emite sinais na faixa de FM, podendo alcançar alguns metros de distância. O ajuste da modulação é feito em P1 enquanto que P2 ajusta o ponto de oscilação do circuito. A antena deve ter de 20 a 50 cm de comprimento e os capacitores menores são cerâmicos. Qualquer diodo tunnel pode ser usado. L1 e L2 como no projeto 79.

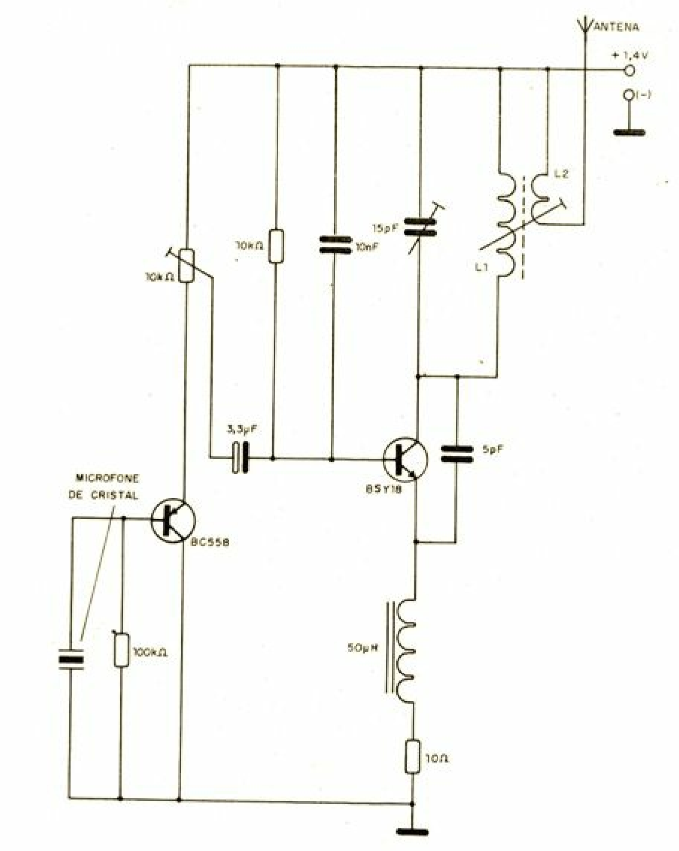
ANTENA
+ 1,4 V
(-)
L2
15pF
10kΩ
10nF
10kΩ
L1
3,3µF
5pF
MICROFONE DE CRISTAL
BSY18
BC558
50µH
100kΩ
10Ω

88 - Transmissor Integrado MC833

Encontramos este circuito transmissor com um único chip no Linear Databook da Motorola de 1990. O circuito é alimentado por uma bateria de 9 V e o cristal deve ser de acordo com a frequência transmitida. Detalhes das bobinas são dados junto ao diagrama.

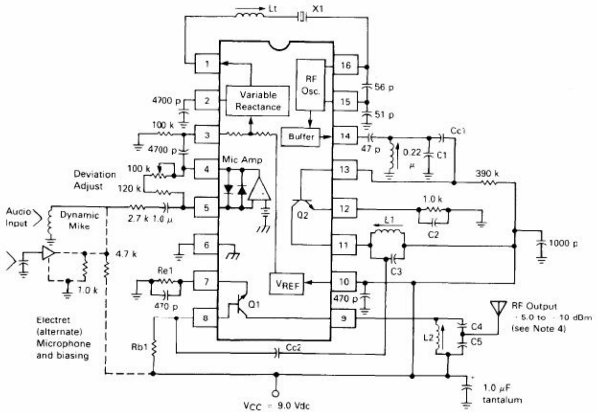
Lt
X1
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Variable Reactance
RF Osc.
Buffer
Mic Amp
VREF
Q1
Q2
4700 p
100 k
4700 p
100 k
120 k
Deviation Adjust
Audio Input
Dynamic Mike
2.7 k
1.0 μ
4.7 k
1.0 k
Re1
470 p
Electret (alternate) Microphone and biasing
Rb1
Cc2
56 p
51 p
47 p
0.22 μ
C1
Cc1
390 k
1.0 k
C2
L1
C3
1000 p
470 p
RF Output
5.0 to 10 dBm
(see Note 4)
L2
C4
C5
1.0 μF tantalum
VCC = 9.0 Vdc

NOTES:

1. Components versus output frequency:

Output RF	X1 (MHz)	Lt (μH)	L1 (μH)	L2 (μH)	Re1	Rb1	Cc1	Cc2	C1	C2	C3	C4	C5
49.7 MHz	16.5667	3.3–4.7	0.22	0.22	330	390 k	33 p	33 p	33 p	470 p	33 p	47 p	220 p
76 MHz	12.6	5.1	0.22	0.22	150	300 k	68 p	10 p	68 p	470 p	12 p	20 p	120 p
144.6 MHz	12.05	5.6	0.15	0.10	150	220 k	47 p	10 p	68 p	1000 p	18 p	12 p	33 p

89 - Transceptor Experimental

Este circuito de pequeno transceptor foi encontrado numa documentação dos anos 70. Para a faixa de ondas curtas o alcance é da ordem de alguns metros ou até uns 20 metros se usada a faixa indicada. A bobina (A) tem 30 espiras de fio 28, a bobina B tem 15 espiras e C tem 10 espiras num bastão de ferrite. O sentido dos enrolamentos é importante. Os transistores podem ser os BC548 e BC558.

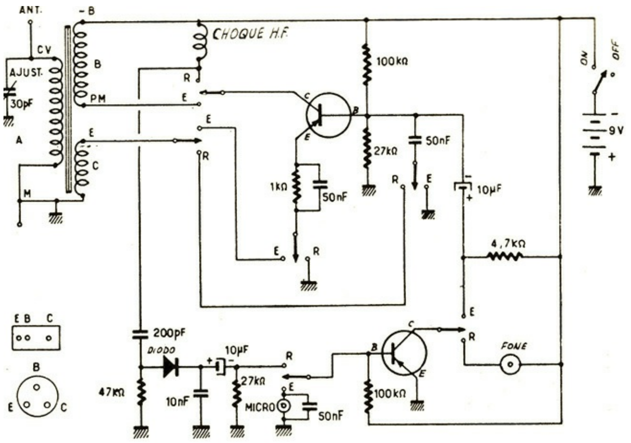
ANT.
- B
CV
AJUST.
30pF
A
B
PM
E
C
M
CHOQUE H.F.
R
E
E
R
C
B
E
100kΩ
27kΩ
50nF
1kΩ
50nF
R
E
E
R
10µF
ON
OFF
9V
4,7kΩ
E
R
E B C
B
E
C
200pF
DIODO
10µF
R
47kΩ
10nF
27kΩ
E
MICRO
50nF
B
C
E
100kΩ
FONE

90 - Grid Dip Meter

Este circuito é de uma publicação americana de 1973. Podemos implementá-lo com o NF494 para o transistor de RF e BC548 para os demais. A lâmpada pode ser trocada por um LED em série com um resistor de 220 ohms. A bobina depende da frequência que se deseja detectar. O diodo é de germânio como o 1N34. O circuito opera até uma frequência de algumas dezenas de megahertz.

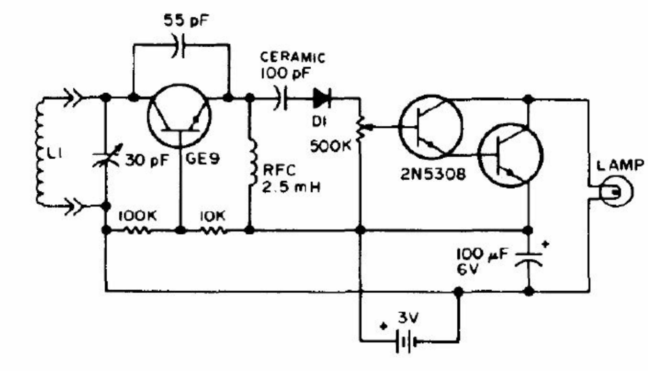
55 pF
CERAMIC
100 pF
D1
500K
L1
30 pF
GE9
RFC
2.5 mH
2N5308
LAMP
100K
10K
100 µF
6V
+
+
3V

91 - Gate Dipper

Esta é uma versão com MOSFET de um grid-dip meter para frequências entre 1,8 MHz e 150 MHz, dependendo apenas da bobina utilizada. O FET pode ser o MPF102 ou BF245 e o instrumento pode ser a escala mais baixa de correntes de um multímetro comum. O circuito é de uma publicação de 1977.

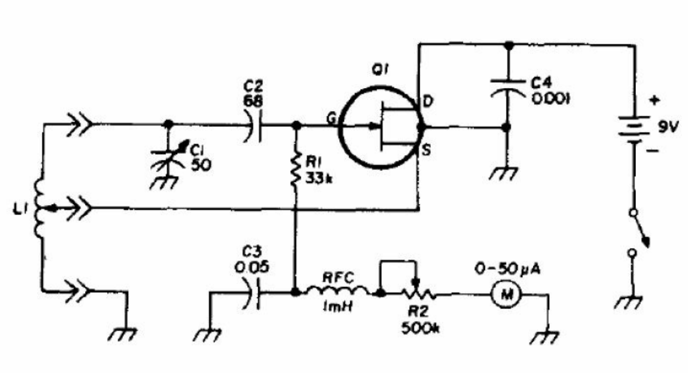
Q1
C2
68
G
D
S
C4
0.001
+
9V
−
C1
50
R1
33k
L1
C3
0.05
RFC
1mH
R2
500k
0-50μA
M

92 - Monitor CW

Este é um circuito usado por radioamadores para monitorar as transmissões em telegrafia. Colocado junto ao transmissor, acoplado a bobina de saída, ele produz sons ao se manipular um sinal. O circuito não precisa de alimentação, pois ela vem do próprio sinal monitorado.

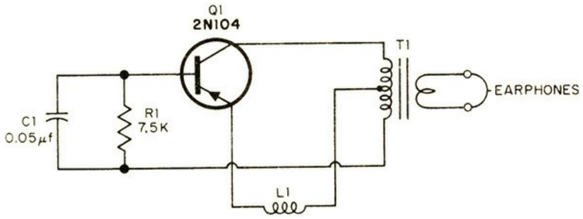
Q1
2N104
T1
EARPHONES
C1
0.05μf
R1
7.5K
L1

93 - Oscilador de 500 MHz

Este circuito pode ser usado como base para um transmissor de UHF ou ainda um equipamento de prova. O circuito é crítico e os capacitores devem ser cerâmicos. A bobina osciladora nada mais é do que um pedaço de cabo coaxial no comprimento que resulte a frequência desejada.

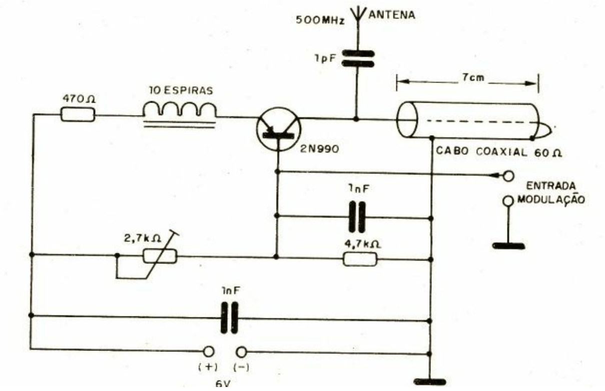
500MHz
ANTENA
1pF
7cm
470Ω
10 ESPIRAS
2N990
CABO COAXIAL 60Ω
ENTRADA
MODULAÇÃO
1nF
2,7kΩ
4,7kΩ
1nF
(+)
(−)
6V

94 - Injetor FM

O oscilador apresentado gera um sinal de teste na faixa de FM, para a verificação de receptores. A frequência depende de L1 que pode ter de 3 a 4 espiras de fio 26 ou 28 numa forma de 1 cm de diâmetro sem núcleo, Os capacitores menores devem ser cerâmicos. Encontramos este circuito numa documentação de 1964. Use um PNP de RF ou inverta a alimentação e use um BF494.

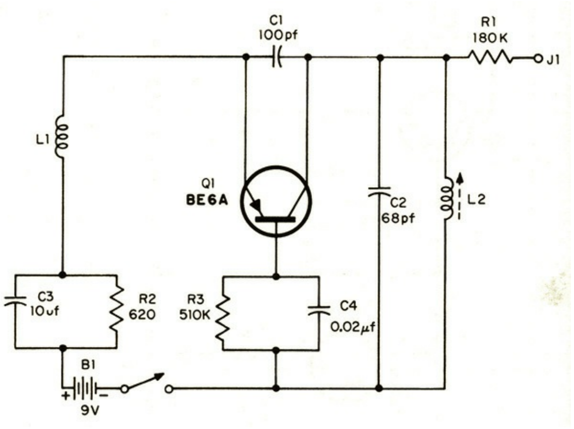
C1
100pf
R1
180K
J1
L1
Q1
BE6A
C2
68pf
L2
C3
10uf
R2
620
R3
510K
C4
0.02μf
B1
+
−
9V

95 - Oscilador de 200 MHz

As bobinas deste circuito são elos feitos com fio comum nas dimensões indicadas. O circuito produz um sinal em torno de 200 MHz e aceita modulação externa. Os transistores podem ser substituídos por equivalentes modernos que oscilem bem na faixa de VHF.

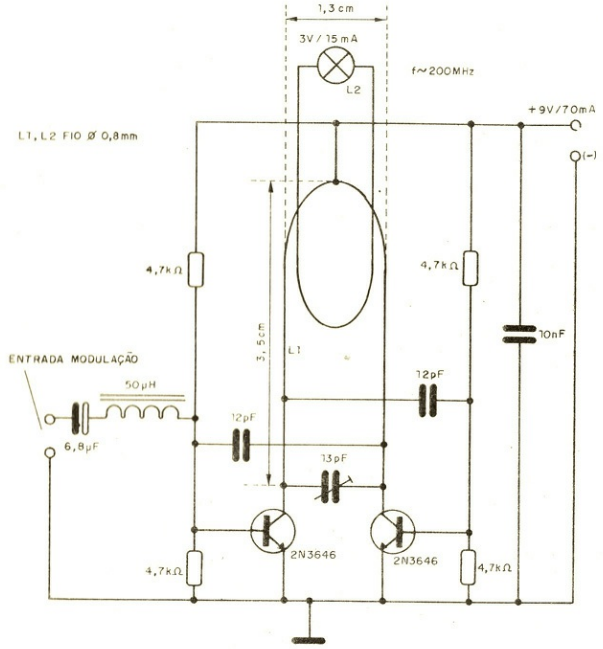
1,3 cm
3V / 15 mA
L2
f~200MHz
+9V/70mA
(-)
L1, L2 FIO Ø 0,8mm
4,7kΩ
4,7kΩ
3,5cm
L1
10nF
ENTRADA MODULAÇÃO
50μH
12pF
12pF
6,8μF
13pF
2N3646
2N3646
4,7kΩ
4,7kΩ

96 - Etapa de Áudio para Rádios Transistorizados

Esta etapa é encontrada em muitos receptores dos anos 70, podendo ser utilizada em rádios e outros aparelhos experimentais. Na versão original são usados transistores de germânio que podem ser encontrados em antigos rádios, dos quais, os transformadores também podem ser aproveitados.

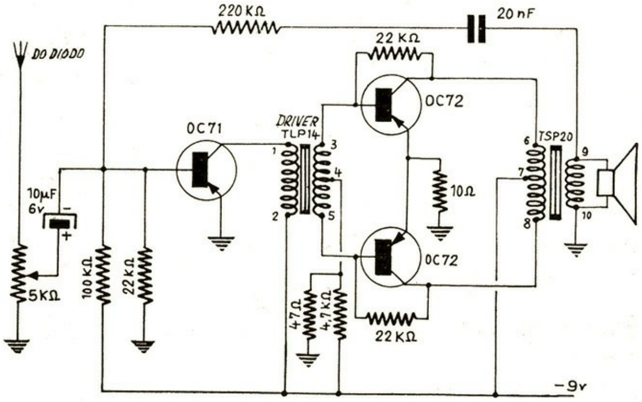
220 KΩ
20 nF
DO DIODO
22 KΩ
OC72
OC71
DRIVER
TLP14
1
2
3
4
5
10Ω
TSP20
6
7
8
9
10
10µF
6v
-
+
5KΩ
100 KΩ
22 KΩ
OC72
47Ω
4,7 KΩ
22 KΩ
-9v

97 - Oscilador de RF LM3909

Encontramos este circuito numa documentação da National Semconductor. Para a faixa de ondas médias a bobina consiste em 50+50 espiras de fio 28 num bastão de ferrite.

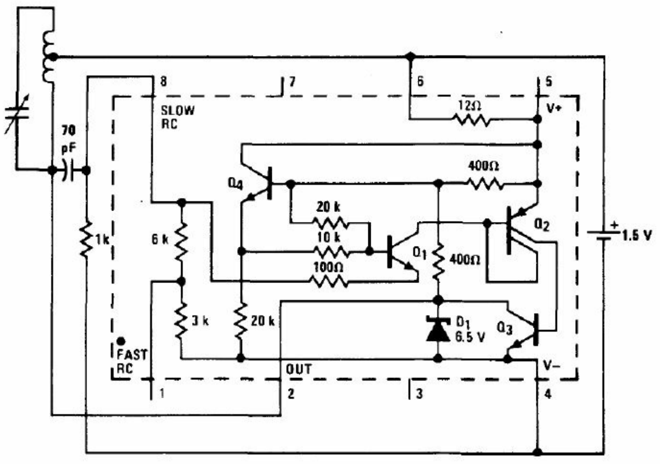
8
7
6
5
SLOW
RC
12Ω
V+
70
pF
400Ω
Q4
20 k
10 k
100Ω
Q1
400Ω
Q2
+
1.5 V
1k
6 k
3 k
20 k
D1
6.5 V
Q3
FAST
RC
OUT
V–
1
2
3
4

98 - Modulador Balanceado MC1556G

Encontramos este circuito no Linear Databook da Motorola de 1990. O componente semicondutor usado não é mais comum em nossos dias.

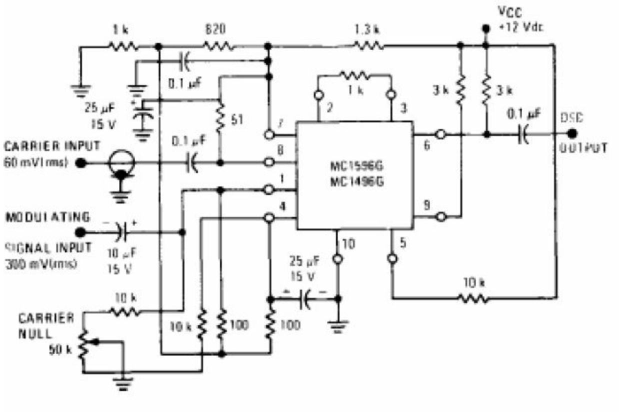
VCC
+12 Vdc
1 k
820
1.3 k
0.1 µF
25 µF
15 V
51
1 k
2
3
3 k
3 k
0.1 µF
DSC
OUTPUT
7
CARRIER INPUT
60 mV(rms)
0.1 µF
8
6
MC1596G
MC1496G
1
MODULATING
SIGNAL INPUT
300 mV(rms)
4
9
10 µF
15 V
10
5
25 µF
15 V
10 k
10 k
CARRIER
NULL
50 k
10 k
100
100

99 - Decodificador QAM Para AM Estéreo

Se bem que o sistema seja experimental e não deve ser adotado pela vinda do rádio digital, para os que desejam saber como é um circuito decodificador encontramos um no Linear Databook da Motorola de 1990.

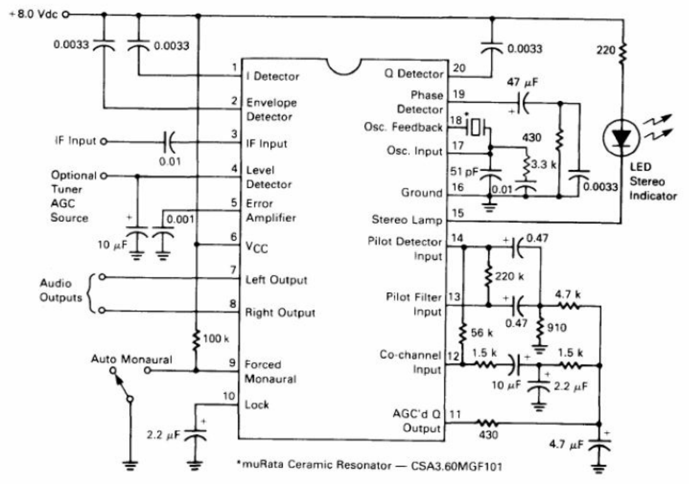
+8.0 Vdc
0.0033
0.0033
1 I Detector
2 Envelope Detector
3 IF Input
4 Level Detector
5 Error Amplifier
6 VCC
7 Left Output
8 Right Output
9 Forced Monaural
10 Lock
IF Input
0.01
Optional Tuner AGC Source
10 µF
0.001
Audio Outputs
100 k
Auto Monaural
2.2 µF
20 Q Detector
19 Phase Detector
18 Osc. Feedback
17 Osc. Input
16 Ground
15 Stereo Lamp
14 Pilot Detector Input
13 Pilot Filter Input
12 Co-channel Input
11 AGC'd Q Output
0.0033
47 µF
220
430
51 pF
3.3 k
0.01
0.0033
LED Stereo Indicator
0.47
220 k
4.7 k
0.47
910
56 k
1.5 k
1.5 k
10 µF
2.2 µF
430
4.7 µF
*muRata Ceramic Resonator — CSA3.60MGF101

100 - Transmissor Integrado Sem Bobinas

O TAA131 usado neste circuito não é muito fácil de encontrar atualmente, mas pode ser usado um amplificador de áudio de baixa potência equivalente em seu lugar. O transmissor não utiliza bobinas, tem curto alcance e sua faixa de sintonia, bastante precária é muito ampla com a emissão de muitas harmônicas. Trata-se de circuito experimental.

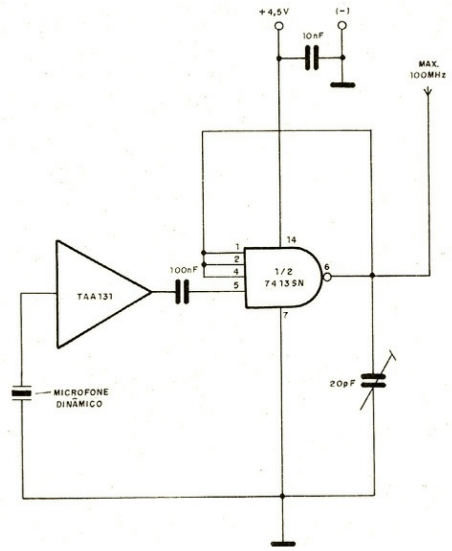
+4,5V
(-)
10nF
MAX.
100MHz
14
1
2
4
5
1/2
7413SN
6
7
100nF
TAA131
20pF
MICROFONE
DINÂMICO

A Qualidade dos Rádios AM

Com o final das transmissões da faixa de ondas médias terminando, com a migração das emissoras para a faixa de FM, devemos fazer algumas considerações sobre a tecnologia antiga do AM que dominou por muitos anos a radiodifusão e as novas tecnologias que estão chegando como, por exemplo, o rádio digital.

Sempre gostei de ouvir rádios AM, pois é ainda nesta modalidade de transmissão que temos a maioria das transmissões de jogos de futebol e também noticiários.

Se bem que as notícias possam ser vistas no computador, ainda é agradável para muitos não precisar ficar olhando para a tela, pois desejam descansar, numa rede, por exemplo, ou simplesmente sem se preocupar com nada.

No interior, as emissoras de AM ainda têm bastante força, pois em alguns casos, elas chegam nos sítios e fazendas onda a internet ainda não chegou, consistindo sua audição num hábito das pessoas mais velhas.

No entanto, mesmo podendo encontrar rádios AM à venda, tem nos preocupado a perda de qualidade desses aparelhos.

Antigamente, era possível comprar um bom rádio AM com excelente sensibilidade e seletividade capaz de sintonizar emissoras distantes ou fracas sem problemas.

Tais rádios utilizavam a tecnologia do super-heteródino com diversos circuitos sintonizados (FIs) capazes de aumentar a seletividade e com muitas etapas de amplificação, tinham excelente sensibilidade.

Rádios como o Transglobe, Moto-Rádio eram comuns e os preferidos por quem desejava uma boa recepção.

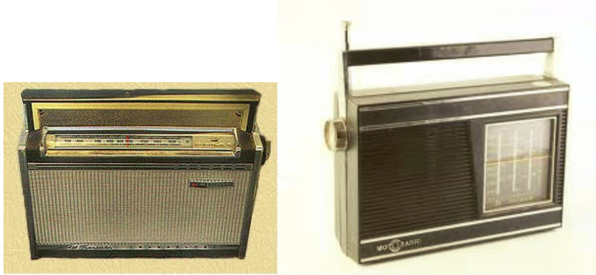

FIGURA 1 – RÁDIOS TRANSGLOBE DA PHILCO E MOTO-RADIO AM

Hoje, o que temos são rádios AM de procedência chinesa que, para maior economia, usam chips únicos com amplificadores de RF e Áudio simples, para reduzir custos e etapas de amplificação direta.

FIGURA 2- RADIO AM DE BAIXA SELETIVIDADE E SENSIBILIDADE

Essas etapas simplesmente detectam o sinal de áudio captado pela antena de ferrite sem que eles passem por outros circuitos sintonizados.

O resultado disso é uma perda de seletividade e de sensibilidade que afetam em muito a qualidade de recepção.

A faixa de ondas curtas de tais rádios é péssima, e a de ondas médias muito ruins quando se deseja sintonizar uma estação mais fraca ou mesmo separar estações próximas.

A não disponibilidade de bons receptores é talvez uma das causas da alegação de que o AM "está sujeito a muitas interferências e ruídos", que está levando justamente a migração para o FM.

Sim, está sujeito a mais interferências e ruídos, mas isso foi minimizado durante muito tempo com bons receptores.

Fiz uma experiência muito simples, para mostrar isso.

Liguei meu Moto-Rádio antigo perto de meu computador e mesmo assim, pude ouvir com clareza a maioria das estações de AM que costumo sintonizar, facilmente "driblando" as interferências que estavam presentes.

No entanto, meu rádio chinês de baixo custo não conseguiu sintonizar sequer a mais forte das estações que costumo ouvir.

Sim, vamos mudar para o FM e depois para o digital, mas não há necessidade de desativar o AM que ainda pode ser recebido em muitos locais, sem problemas, desde que estejam disponíveis bons receptores...

AJUSTE DE RÁDIOS AM

Muitos ainda possuem rádios e FM com circuitos tradicionais (super-heteródinos) com boninas de FI, osciladora e antena que precisam estar corretamente ajustadas para se obter um bom funcionamento. Os técnicos que recebem estes receptores eventualmente, dependendo do componente trocado, precisam refazer o ajuste. Damos a seguir os procedimentos possíveis:

Procedimento 1:
O primeiro procedimento de ajuste que vamos dar é “de ouvido”, sendo indicado para o caso do leitor não possuir um gerador de sinais.

a) Sintonize uma estação fraca no meio da escala (em torno de 900 a 1000 kHz).

b) ajuste os núcleos dos transformadores de FI, começando pelo terceiro (que está ligado ao detector) de modo a obter o melhor nível de sinal.

c) Os ajustes devem ser feitos com ferramentas não magnéticas (plástico ou madeira) conforme mostra a figura 1.

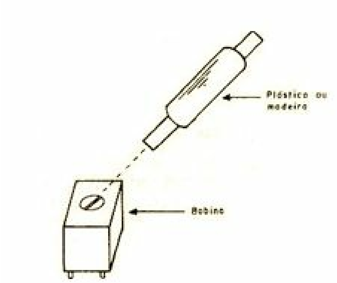

FIGURA 1 – USANDO FERRAMENTAS NÃO METÁLICAS PARA O AJUSTE.

Use o diagrama do receptor para identificar os transformadores de FI do receptor, lembrando que o vermelho não deve ser tocado, pois é a bobina osciladora.

d) Procure agora sintonizar uma estação no extremo superior da faixa de ondas médias (entre 1300 e 1600 kHz). Procure uma estação de frequência conhecida, de preferência.

e) Mova o trimmer do oscilador junto ao variável até que, retocando a sintonia você consiga levar a estação sintonizada para a frequência correta, se ela estiver incorreta.

Na figura 2 mostramos a localização deste trimmer (veja que o trimmer do oscilador “muda” a frequência do sinal sintonizado, enquanto que o outro trimmer do variável, que é de sintonia, atua sobre a intensidade do sinal).

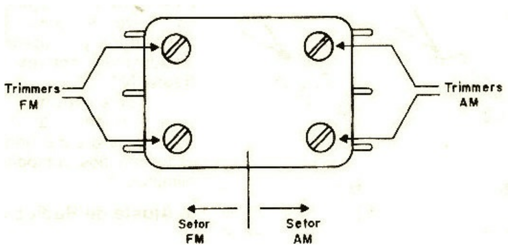

FIGURA 2 – LOCALIZAÇÃO DO TRIMMER DE AJUSTE DO OSCILADOR.

f) Sintonize agora uma estação de frequência conhecida no extremo inferior da faixa de ondas médias, entre 520 e 700 kHz.

g) Ajuste o núcleo da bobina osciladora (vermelha) para que o sinal seja captado no ponto correto da escala, se estiver fora dela (indicação errada). Isso será necessário, por exemplo, se a estação transmitir em 550 kHz e você a captar em 600 kHz.

h) Sintonize agora novamente uma estação no extremo superior da faixa e ajuste o trimmer do capacitor variável para que seu sinal seja captado com máxima intensidade.

i) Volte a sintonia para a estação do extremo inferior e mova o núcleo da bobina de antena (de núcleo de ferrite) até obter o melhor sinal. Este ajuste é feito deslocando-se levemente a bobina sobre o bastão de ferrite conforme mostra a figura 3.

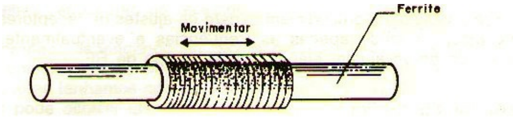

FIGURA 3 – AJUSTANDO A BOBINA DE ANTENA PELA POSIÇÃO DO NÚCLEO DE FERRITE.

j) refaça todos os ajustes algumas vezes até obter o melhor rendimento.

k) Fixe a posição da bobina de antena no núcleo de ferrite pingando algumas gotas de cera de vela.

Procedimento 2:

O procedimento que vamos ver agora faz uso de instrumentos, no caso o gerador de sinais ou RF que produza sinais na faixa de 100 kHz a 10 MHz pelo menos. Ajustes das FIs:

a) Coloque o gerador de sinais em 455 kHz e aplique seu sinal ao receptor de uma das formas indicadas no procedimento anterior.

b) Ajuste os núcleos das bobinas com uma ferramenta não magnética até obter o máximo de sinal no alto-falante (maior nível de apito).

c) Refaça os ajustes várias vezes, pois eles são interdependentes.

Ajustes de RF:

a) Certifique-se que o receptor a ser calibrado está com a chave seletora na posição de ondas médias (OM) de 520 ou 550 a 1600 ou 1630 kHz.

b) Conecte o gerador ao receptor usando para isso a antena ou ainda um elo de acoplamento, conforme mostra a figura 4.

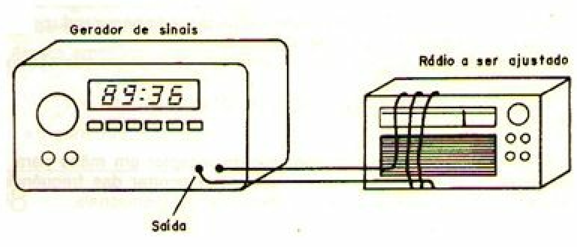

FIGURA 4 – ACOPLANDO O GERADOR DE SINAIS OU FUNÇÕES AO RECEPTOR QUE DEVE SER AJUSTADO.

c) Coloque a sintonia no extremo superior da faixa (1600 kHz).

d) Ajuste o gerador para fornecer um sinal modulado na frequência máxima sintonizada pelo receptor.

e) Ligue o receptor e retoque o trimmer do oscilador (junto ao variável) até ouvir o sinal do gerador. Ajuste sua intensidade para que o sinal não sature o receptor, mas seja apenas audível, pois devemos nos orientar pelo aumento de sua intensidade na reprodução.

f) Coloque gora a sintonia do receptor no extremo inferior da faixa de ondas médias, em torno de 540 kHz.

g) Coloque o gerador de sinais para produzir um sinal na mesma frequência em que o receptor está sintonizado.

h) Ajuste o núcleo da bobina osciladora até que o sinal seja captado.
Reduza a intensidade do sinal do gerador se ele estiver se tornando muito forte.

i) Reajuste o núcleo da bobina osciladora até obter a máxima intensidade, agora com um nível de sinal menor do gerador.

j) Ajuste a sintonia do receptor para o extremo superior da faixa, em torno de 1450 kHz.

k) Coloque o gerador na mesma frequência. Ajuste levemente a sintonia do receptor até que o sinal seja localizado. Se estiver muito longe da frequência desejada refaça os ajustes iniciais até conseguir isso.

l) Ajuste o trimmer de antena para que haja maior intensidade de sinal.

m) Coloque o gerador de sinais para produzir um sinal em 700 kHz.

n) Ajuste a sintonia do receptor para a mesma frequência do sinal. Se esse sinal não for captado, desloque a bobina de antena sobre o núcleo de ferrite até que ele seja sintonizado.

o) Obtida a sintonia fixe a bobina com um pouco de cera de vela.

p) Refaça os ajustes pelo menos duas vezes, pois eles são interdependentes.

AJUSTANDO TRANSMISSORES

O máximo rendimento de um transmissor, e também a menor possibilidade de se irradiar sinais indesejáveis, só se consegue com um ajuste perfeito de todos os seus circuitos. No entanto, não basta ter um transmissor e se fazer sua conexão a uma boa antena para se garantir que o bom funcionamento esteja garantido. Como ajustar um transmissor e como obter o máximo de seu rendimento é o que veremos agora.

Um transmissor ideal, corretamente ajustado, deve emitir sinais numa única frequência e com a máxima intensidade possível. Porém, na prática, não é isso o que ocorre.

Os circuitos não são perfeitos, gerando sinais espúrios e harmônicas; o acoplamento a antena e o próprio cabo não são perfeitos, gerando reflexões responsáveis por ondas estacionárias e a própria antena não é perfeita, não irradiando 100% da energia que recebe de um transmissor.

Tudo isso faz com que o sinal que realmente sai de uma estação fique reduzido em diversas proporções, levando transmissores de igual potência, com o mesmo circuito, quando instalados a apresentarem desempenhos completamente diferentes.

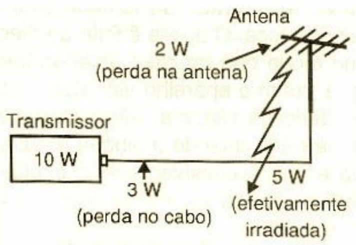

Fig. 1 - Perdas podem afetar o rendimento de uma estação transmissora.

Até a própria topografia do local em que o transmissor é instalado pode influir no seu desempenho, dependendo de sua faixa de frequências, com a dificuldade que os sinais podem encontrar para passar por determinados obstáculos ou se propagar em função de um terreno de baixa condutividade elétrica.

Colocar um transmissor em condições de funcionar com o máximo rendimento não é simples, exigindo-se um profundo conhecimento técnico de quem vai fazer isso.

Para cada modalidade de operação, para cada tipo de instalação, para cada tipo de antena existem procedimentos típicos que vão desde o ajustes dos circuitos, até a instalação dos cabos e posicionamento da antena.

Neste artigo falaremos especificamente dos ajustes dos circuitos, mostrando como devem ser feitos.

MODULAÇÃO

Os transmissores mais comuns usados por radioamadores, em transmissores de radiodifusão e em alguns serviços públicos operam com modulação em amplitude ou modulação em frequência (AM ou FM).

Na modulação em amplitude, a intensidade do sinal varia com o sinal de áudio que deve ser transmitido, ou ainda o sinal de vídeo se for uma emissão de TV analógica, conforme mostra a figura 2.

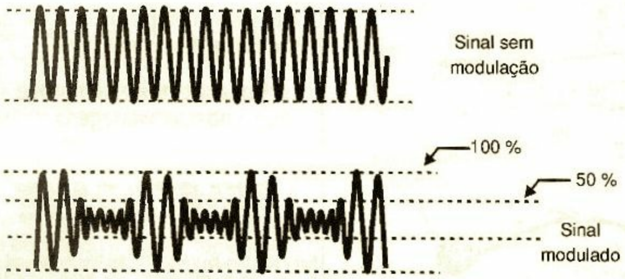

Fig. 2 - Sinal modulado em amplitude.

Nesta figura mostramos um sinal sem modulação, ou seja, um sinal portador de alta frequência puro ou sem modulação e um sinal modulado.
Veja que o sinal de áudio que vai modular a portadora e que portanto corresponde à informação transmitida faz com que a amplitude do sinal de alta frequência varie de 50% sua intensidade.
Nos pontos de menor amplitude temos 50% da intensidade dos pontos de maior amplitude. Dizemos, neste caso, que este sinal tem uma modulação de 50%.
Para que uma transmissão tenha o máximo rendimento, o ideal é que ela tenha uma porcentagem de modulação maior, ou seja, 100%.
Isso significa que, nos pontos de menor intensidade o sinal praticamente deve ser cortado quando nos pontos de maior intensidade temos a amplitude máxima do sinal que o transmissor pode gerar, conforme mostra a figura 3.

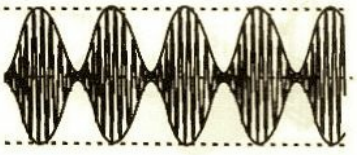

Fig. 3 - Sinal com 100% de modulação.

Tudo seria simples no ajuste se aplicando uma boa potência a um transmissor conseguíssemos chegar aos 100% de modulação com facilidade. Mas, não é isso o que ocorre.
Se a potência do sinal modulador for muito pequena teremos uma modulação de menos de 100% e o rendimento do transmissor não será dos melhores.
No entanto, se excedermos a potência que o circuito precisa para modulação, ocorre uma sobremodulação, ou seja, mais de 100% de modulação, conforme mostra a figura 4.

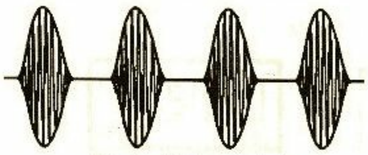

Fig. 4 - Sinal com mais
de 100% de modulação sobremodulação.

O problema da sobremodulação não é apenas a distorção que ocorre com a informação que o sinal carrega, mas também a produção de sinais espúrios que causam fortes interferências.
Parte da potência do transmissor é então desviada para estes sinais e com isso o rendimento do equipamento na frequência que se deseja transmitir cai.
Como medir a porcentagem de modulação?

OSCILOSCÓPIO

O osciloscópio é um instrumento de grande utilidade também no ajuste de transmissores.

Com ele podemos ver a porcentagem de modulação de um sinal modulado em amplitude com facilidade, conforme mostra a figura 5.

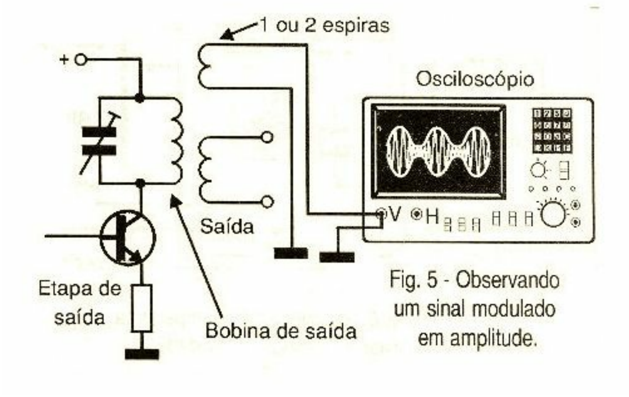

Fig. 5 - Observando um sinal modulado em amplitude.

Basta fazer um elo de captação de sinal e colocá-lo junto à bobina do tanque de saída do transmissor, conforme mostra a figura, e aplicar um sinal de áudio senoidal na entrada do transmissor, por exemplo 1 kHz.

Podemos visualizar na tela do osciloscópio o sinal modulado e fazer os ajustes no sentido de obter 100% de modulação.

Uma outra maneira de se medir a porcentagem de modulação de um transmissor é com o arranjo mostrado na figura 6.

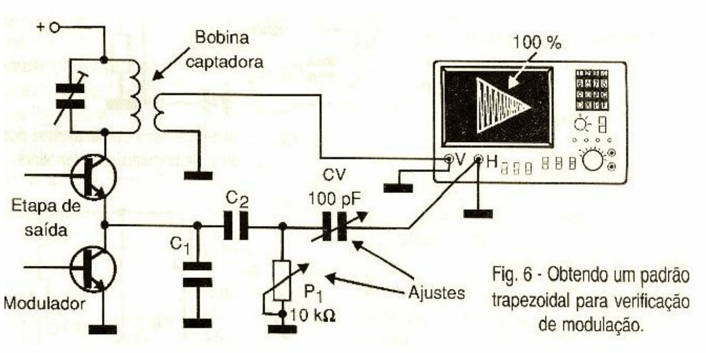

Fig. 6 - Obtendo um padrão trapezoidal para verificação de modulação.

Neste arranjo, a varredura horizontal é desativada, ou seja, colocada na posição de "varredura externa" ou EXT,.

O sinal de varredura vai então ser fornecido pelo transformador de modulação do próprio transmissor conforme indicado no mesmo diagrama.

O sinal do transmissor será então aplicado à entrada vertical do osciloscópio sendo retirado da saída do transmissor com a ajuda de um elo de captação.

Este elo pode ser formado por algumas espiras de fio comum colocado próximo da bobina tanque de saída do transmissor.

Veja que o transmissor deve estar conectado a uma carga fantasma nestes testes já que sem antena corre-se o perigo de provocar a queima dos transistores de saída.

As formas de imagem obtidas com este arranjo são mostradas na figura 7, indicando as porcentagens de modulação.

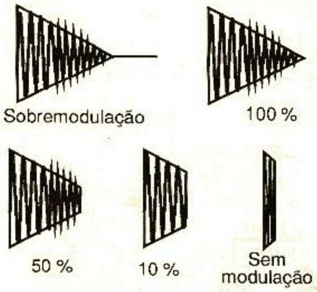

Fig. 7 - Formas de visualização para diferentes porcentagens de modulação.

A porcentagem de modulação será obtida dividindo-se o comprimento relativo da base menor do trapézio pelo comprimento da base maior.
Se o padrão trapezoidal tiver deformações como as mostradas na figura 8, isso indica que os circuitos de RF podem ter problemas de linearidade.

Fig. 8 - Deformações que indicam problemas de circuito e não de ajuste.

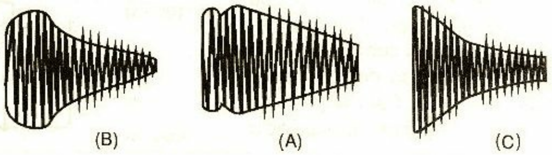

A presença de barras verticais no padrão mostrado, por exemplo indica que a etapa de potência do circuito não está devidamente neutralizada com tendência à oscilação.

AJUSTE DE TRANSMISSORES DE FM

A verificação do funcionamento dos transmissores de FM é um pouco mais complexa, exigindo métodos diferentes.
Duas quantidades devem ser analisadas quando examinamos o funcionamento de um transmissor de FM: desvio de frequência e linearidade.

a) verificando a frequência

A frequência de um transmissor de FM, sem modulação, deve ser estável e mantida num valor determinado, mesmo quando as condições ambientes variam (temperatura, umidade, etc.).
Os transmissores comerciais, para que isso realmente ocorra, são controlados por cristais, os quais são mantidos em câmaras térmicas, conforme mostra a figura 9.

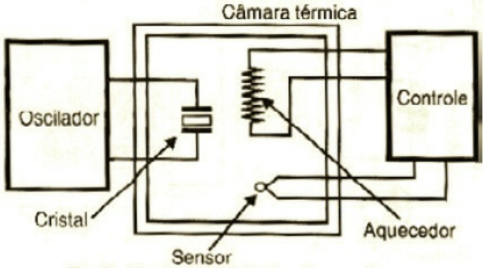

Fig. 9 - Mantendo constante a temperatura do cristal para maior estabilidade.

Nestas câmaras a temperatura é mantida constante graças a um sensor e a um elemento de aquecimento.

O ajuste da frequência é simples, já que hoje pode-se contar com frequencímetros de precisão que medem as frequências produzidas pelos circuitos com facilidade.

Mesmo os frequencímetros mais baratos podem ser usados com a ajuda de um prescaler.

Este dispositivo, conforme mostra a figura 10, divide a frequência de um sinal por um valor inteiro (10 por exemplo) possibilitando assim o uso de frequencímetros de menor alcance na medida de frequências mais altas.

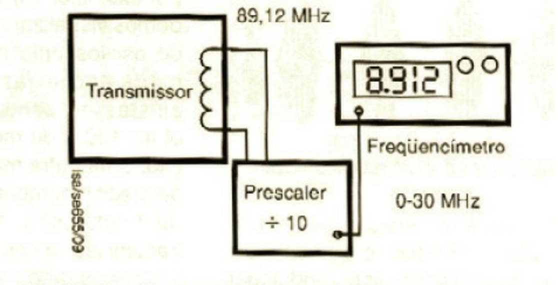

Fig. 10 - Estendendo o alcance de um fenômeno com um *prescale*

b) Desvio

Um sinal modulado em frequência tem sua frequência variando com a informação (áudio, por exemplo), conforme mostra a figura 11.

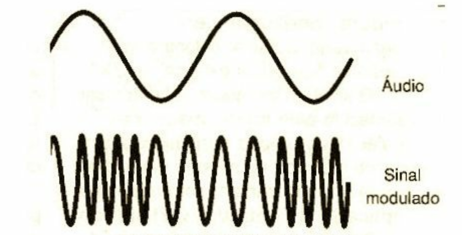

Fig. 11 - Sinal modulado frequência.

Isso significa que, com a máxima intensidade de modulação, este sinal deve ter sua frequência variando em torno do valor central entre dois valores muito bem determinados.

Veja que não é a frequência do sinal que determina a frequência que o sinal de alta frequência se desloca mas sim sua amplitude.

Esta largura de faixa é que vai determinar a eficiência na recuperação da informação nos circuitos discriminadores do receptor.

Se o transmissor estiver mal calibrado, com um desvio muito pequeno do sinal, a tensão na saída do discriminador será pequena e o sinal de áudio recuperado fraco.

Do mesmo modo, se o desvio for muito grande, o sinal "satura" o discriminador e a informação não é recuperada da forma desejada, conforme mostra a figura 12.

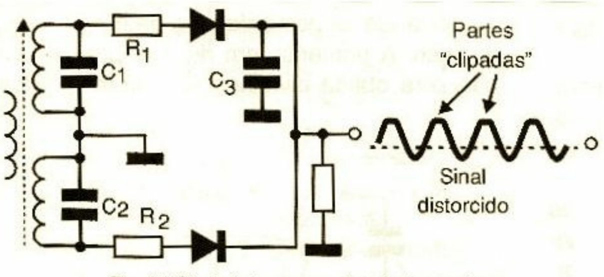

Fig. 12 Sinais intensos ou desajustes podem afetar a demodulação de um sinal.

Na figura 13 temos um circuito simples que permite a medida do desvio de frequência de um sinal modulado em frequência.

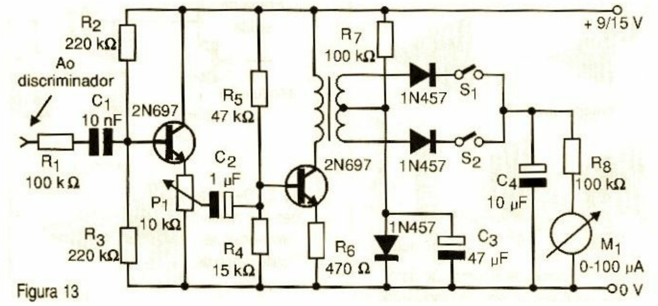
R2
220 kΩ
Ao discriminador
C1
10 nF
R1
100 kΩ
2N697
P1
10 kΩ
R3
220 kΩ
R5
47 kΩ
C2
1 µF
R4
15 kΩ
2N697
R6
470 Ω
R7
100 kΩ
1N457
S1
1N457
S2
1N457
C3
47 µF
C4
10 µF
R8
100 kΩ
M1
0-100 µA
+ 9/15 V
0 V

Figura 13

Este circuito foi originalmente sugerido no Radio Amateur Handbook (edição de 1987) e os transistores usados podem ser substituídos por equivalentes como os BC548 ou equivalentes/

O circuito é ligado na etapa discriminadora de qualquer receptor de FM antes do circuito de ênfase.

O transformador usado pode ser qualquer um do tipo encontrado em velhos rádios transistorizados com impedâncias de 100 a 1000 ohms de primário e 8 ohms de secundário.

O que se faz então é sintonizar o sinal de um transmissor de FM e ligar este circuito observando a indicação do instrumento.

O instrumento deve ser ajustado com base num sinal que tenha uma modulação-padrão.

Normalmente esta modulação é feita com um sinal que provoque um desvio de 15 kHz de frequência na frequência do sinal central.

NEWTON C. BRAGA

BANCO DE CIRCUITOS

VOLUME 30

100 CIRCUITOS DE SHIELDS PARA ARDUINO

NCB

100 circuitos de shields para arduino

Braga, Newton C.

9788565050791

131 páginas

Compre agora e leia

Para esta edição escolhemos mais uma remessa com 100 circuitos que podem ser usados para interfacear com microcontroladores tendo como ponto de partida o Arduino. No entanto, eles também são compatíveis com outros microcontroladores como PICs, MSP430. E muitos outros, em alguns casos exigindo apenas modificações ou alterações de valores de componentes. Estes circuitos, denominados Shields, podem ser tanto usados para transferir os sinais de um microcontrolador a um circuito externo, como podem ser os próprios circuitos externos usados pelo microcontroladores. Eles também podem ser usados para transferir sinais de sensores e outras fontes para os microcontroladores. Todos eles usam componentes comuns.

Compre agora e leia

NEWTON C. BRAGA

CIRCUIT BENCH

VOLUME 30

100 CIRCUITS

SHIELDS FOR ARDUINO

NCB

Circuit Bench - 100 Shields For Arduino - Vol. 30

Braga, Newton C.

9788565050807
131 páginas

Compre agora e leia

We can say that in this serie we will give to the readers the opportunity to have in their tablets, iPhones, iPads and PCs a powerful source of ideas for projects and informartions. Microcrocontrollers such as Arduino, MSP430, PICs and others can´t source a large amount of current to loads like motors, relays and lamps. They also can´t work with signals sourced by some types of sensors plugged to their inputs. In these cases they need special ads, circuits to allow the use of power loads and sensor. These circuits are called shields. This book is a collection of 100 circuits of shields including drive to high current loads, motors, sensor, to produce audio signals and much more.

Compre agora e leia

NEWTON C. BRAGA

BRINCADEIRAS & EXPERIÊNCIAS COM ELETRÔNICA

VOLUME 1

SUPER HEAVY DUTY
TOSHIBA
EXPERIÊNCIAS E BRINCADEIRAS COM
ELETRÔNI
1º VOLUME
PARA PRINCIPIANTES HOBISTAS E ESTUDANTES
EXPERIÊNCIAS
BRINCADEIRA
ELE
NCB

Brincadeiras e Experiências com Eletrônica - volume 1

Braga, Newton C.

9788565050845

100 páginas

[Compre agora e leia](#)

Em 1976 publicávamos nosso primeiro livro, uma coletânea de projetos simples para amadores, iniciantes e estudantes, que chamamos de Experiências e Brincadeiras com Eletrônica. O livro fez um sucesso incrível com a venda de dezenas de milhares de exemplares. O grande sucesso da época pode ser constatado ainda hoje quando encontramos professores universitários, engenheiros em cargos de chefia de grandes empresas, profissionais donos de grandes empresas que nos falam, com satisfação, que graças a este livro e a esta série eles se interessaram por eletrônica, seguindo então suas carreiras de sucesso. Ainda hoje, encontramos profissionais que guardam suas edições como verdadeiras relíquias, ou tesouros de valor incalculável, pois elas representam muito em sua vida, na verdade, o ponto de partida de sua vida profissional. Muitos, ao nos encontrar, já não tendo suas edições nos perguntam se não temos "guardada no fundo do baú" uma edição antiga para lhes ceder. Infelizmente, as que temos também são guardadas a sete-chaves, pelo seu valor e justamente pegando uma delas, a primeira, resolvemos atender, não só os que desejam ter em mãos esta relíquia, tanto na versão impressa como virtual, para recordação dos "bons tempos" de inicio de carreira, como desejam algo mais: iniciar seus filhos e netos nesta maravilhosa ciência que é a eletrônica. Assim fizemos um novo livro baseados naquele, uma edição inicial (e depois virão as outras) em que usamos os projetos originais, inserimos notas ou comentários que visam facilitar quem deseja repetir aquelas montagens em nossas dias ou iniciar seus filhos, netos ou jovens de uma sala de aula num curso de iniciação ou num clube de eletrônica. Enfim, uma nova edição baseada num livro histórico com a abordagem que tanto sucesso fez na época.

[Compre agora e leia](#)

NEWTON C. BRAGA

BANCO DE CIRCUITOS

VOLUME 30

100 CIRCUITOS DE SHIELDS PARA ARDUINO

NCB

100 circuitos de shields para arduino (español)

Braga, Newton C.

9788565050814

131 páginas

Compre agora e leia

Con esta serie, damos a los lectores la oportunidad de tomar sus tablets, iPhones, iPads, PCs, notebooks y otros medios de comunicación una fuente de consulta de gran importancia tanto por su trabajo y para sus estudios o para un simple hobby. Los 100 circuitos seleccionados para esta edición de la serie son sólo una pequeña muestra de lo que encontrará en el sitio. Para esta edición hemos elegido otro envío con 100 circuitos que pueden utilizarse para el Arduino y otros microcontroladores como punto de partida. Sin embargo, también son compatibles con otros microcontroladores tales como fotos, MSP430, PIC y muchos otros, en algunos casos requiriendo sólo modificaciones o actualizaciones a los valores de los componentes. Estos circuitos, llamados shields, pueden ser utilizados para transferir señales de un microcontrolador a un circuito externo, como sus propios circuitos externos utilizados para microcontroladores. También puede utilizarse para transferir señales de sensores y otras fuentes para los microcontroladores. Todos utilizan componentes comunes y en muchos casos aprovechados de aparatos fuera de uso.

Compre agora e leia

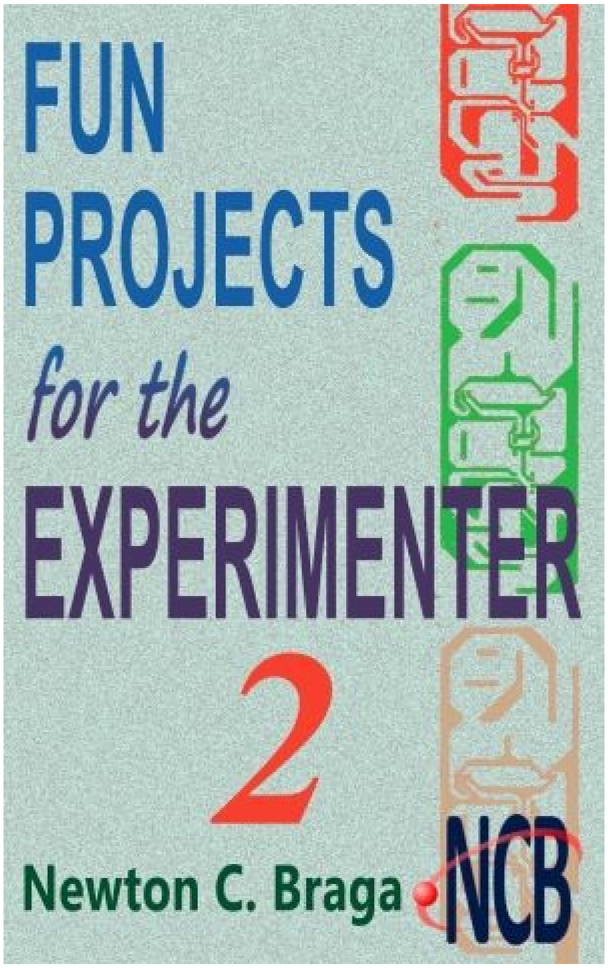

FUN
PROJECTS
for the
EXPERIMENTER
2

Newton C. Braga

Fun Projects for the Experimenter - volume 2

Braga, Newton C.

9788565050753

185 páginas

Compre agora e leia

During more than 30 years, as a collaborator with American, European and Latin American electronics magazines (*), has published a large assortment of practical circuits using common parts. In 1999 he included the first selection in a volume published by Prompt Publications in USA. The idea was to proceed with the series, publishing many volumes more. But, Prompt closed his activities and the idea was forgotten although the first volume became a best seller. Now with his own publishing house (NCB Publications) the author returned with the idea of make many volumes more of the series. So, the second volume is here proceeding with the same idea: give simple projects to the experimenters who want learn electronics using common parts and with no need of special knowledge about electronics. So, as in the first volume, many of the projects collected by the author are included in this volume, most of which you can build in one evening. The projects range from fun types through practical types to amusement types. Of course, there are other devices that can be used to teach you something about circuits and components. An important feature of theses projects are the ideas to Explore, intended for students looking for projects in science or to use in practical research. This ideal can be complemented by our book Science Fair and Technology Education Projects, also published in English by the author. We can consider this book as a source book of the easiest and fun-to-make of hundreds of projects created and published by the author during his life. (see more about Newton C. Braga in "about the author" in his site).

Compre agora e leia

USERS

¡CONTIENE EJEMPLOS PRÁCTICOS
PARA PONER MANOS A LA OBRA!

ELECTRÓNICA PRÁCTICA

APRENDA A ANALIZAR, SIMULAR Y CONSTRUIR CIRCUITOS

AUTORANGING
0 2 3.1
mV
HOLD
RANGE
REL△
Hz%
SELECT
Hz
4000μADC
400mADC
10ADC
OFF
Ω
VDC
OFF
10A
COM
μA Ω
mA V Hz

ADEMÁS

- CONCEPTOS BÁSICOS DE ELECTRÓNICA
- USO DEL MULTÍMETRO Y EL OSCILOSCOPIO
- EL TALLER IDEAL Y LAS HERRAMIENTAS NECESARIAS
- DISEÑO Y ARMADO DE UNA FUENTE DE ALIMENTACIÓN

COLECCIÓN USERS ELECTRÓNICA

www.FreeLibros.me

FUNCIONAMIENTO, PROGRAMACIÓN Y APLICACIONES PRÁCTICAS

USERS
¡CONTIENE EJEMPLOS PRÁCTICOS PARA PONERSE MANOS A LA OBRA!
MICRO CONTROLADORES
FUNCIONAMIENTO, PROGRAMACIÓN Y APLICACIONES PRÁCTICAS
ADEMÁS
FAMILIAS LÓGICAS
MEMORIAS: TECNOLOGÍAS Y UTILIZACIÓN
COMPUERTAS LÓGICAS, REGISTROS DE DESPLAZAMIENTO
LÓGICA DIGITAL: ÁLGEBRA Y SISTEMAS DE NUMERACIÓN
COLECCIÓN USERS ELECTRÓNICA
MICROCONTROLADORES

» HARDWARE
» 192 PÁGINAS
» ISBN 978-987-1773-22-0

Esta obra presenta las diferencias entre los sistemas analógicos y los digitales. Además, analizaremos los conceptos y los dispositivos fundamentales en los que se basa la electrónica digital, ingresaremos en el mundo de las memorias y aprenderemos cómo programar microcontroladores.

SOBRE LA COLECCIÓN: USERS ELECTRÓNICA

- Aprendizaje guiado mediante explicaciones claras y concisas
- Proyectos prácticos basados en necesidades reales
- Consejos de los profesionales
- Producciones fotográficas profesionales
- Infografías y procedimientos paso a paso

LLEGAMOS A TODO EL MUNDO VÍA OCA * Y DHL**

* SÓLO VÁLIDO EN LA REPÚBLICA ARGENTINA // ** VÁLIDO EN TODO EL MUNDO EXCEPTO ARGENTINA

usershop.redusers.com // usershop@redusers.com

www.FreeLibros.me

LIBRO Electronica 1 VIOLETA - COLECCION FLUO 15x19 - BOMBO RT - Feb 10.indd 1 08/04/2011 13:11:51

ELECTRÓNICA PRÁCTICA

AUTORANGING
023.1
mV
HOLD
RANGE
RELΔ
Hz %
SELECT
Hz
4000µADC
AC
400mADC
AC
10ADC
AC
OFF
Ω
VDC
AC
OFF
10A
COM
µA Ω
mA V Hz

www.FreeLibros.me

TÍTULO: Electrónica práctica
COLECCIÓN: desde Cero
FORMATO: 15 X 19 cm
PÁGINAS: 192

Copyright © MMXI. Es una publicación de Fox Andina en coedición con DALAGA S.A. Hecho el depósito que marca la ley 11723. Todos los derechos reservados. Esta publicación no puede ser reproducida ni en todo ni en parte, por ningún medio actual o futuro sin el permiso previo y por escrito de Fox Andina S.A. Su infracción está penada por las leyes 11723 y 25446. La editorial no asume responsabilidad alguna por cualquier consecuencia derivada de la fabricación, funcionamiento y/o utilización de los servicios y productos que se describen y/o analizan. Todas las marcas mencionadas en este libro son propiedad exclusiva de sus respectivos dueños. Impreso en Argentina. Libro de edición argentina. Primera impresión realizada en Sevagraf, Costa Rica 5226, Grand Bourg, Malvinas Argentinas, Pcia. de Buenos Aires en VII, MMXI.

ISBN 978-987-1773-21-3

Electrónica práctica / coordinado por Daniel Benchimol.
- 1a ed. - Buenos Aires: Fox Andina; Dalaga, 2011.
v. 17, 192 p. ; 19x15 cm. - (Desde cero; 17)
ISBN 978-987-1773-21-3

1. Informática. I. Daniel Benchimol, coord.
CDD 005.3

www.FreeLibros.me

Prólogo **al contenido**

La electrónica comenzó con el diodo de vacío, cuando John Ambrose Fleming colocó una lámina dentro de una bombilla para evitar el ennegrecimiento en la ampolla de vidrio. Pasó más de un siglo de esta invención y la electrónica evolucionó a pasos agigantados, hasta llegar a la actual era digital.

Ingresar en el universo de la Electrónica Digital es descubrir cómo se componen y funcionan muchos de los dispositivos tecnológicos que se utilizan en la actualidad, tanto en el ámbito hogareño como en el industrial y en el de entretenimiento. El bajo costo de aplicación, la versatilidad y la eficiencia de los sistemas digitales permiten su aplicación en las computadoras, en los equipos de telefonía, en los de comunicaciones y en una larga lista de dispositivos de alta complejidad. Estos aspectos son solo algunos de los motivos por los cuales invitamos a los lectores a recorrer este apasionante mundo.

Tanto el aficionado a la Electrónica como el hobbista –aquél que tiene algunas nociones teóricas y prácticas en sistemas digitales– encontrarán, en este libro, las herramientas para realizar sus primeros proyectos.

El material teórico y práctico que abarca esta obra se alinea en un solo sentido: explicar todos los fundamentos de la Electrónica Digital y su aplicación real y práctica. Por medio de un despliegue editorial de alto impacto visual y de ágil lectura, el lector encontrará una respuesta para cada una de las preguntas que se plantee.

www.FreeLibros.me

El libro **de un vistazo**

Este libro está enfocado a aficionados de la electrónica, hobbistas y a todo aquel que tiene algunas nociones teóricas y prácticas en sistemas digitales. A lo largo de la obra, veremos cómo aplicar los conocimientos adquiridos mediante proyectos prácticos..

CAPÍTULO 1
ELECTRÓNICA DIGITAL

Conoceremos las primeras nociones de electrónica digital, cuáles son sus alcances y los límites que debemos respetar. Veremos las herramientas más importantes y qué instrumentos no nos pueden faltar.

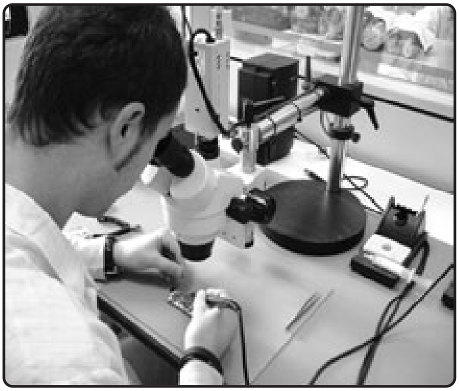

CAPÍTULO 2
CONSTRUCCIÓN DE CIRCUITOS

Comprender las bases de la composición de los circuitos digitales resulta de suma importancia para poder abordar, más adelante, los procesos complejos.

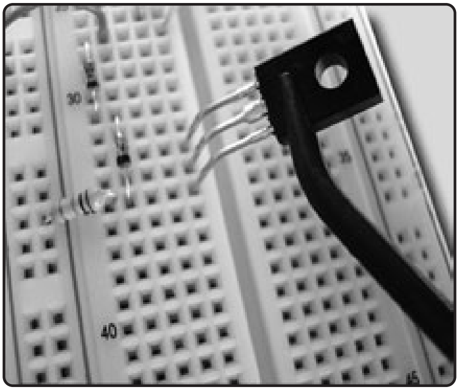

CAPÍTULO 3
INSTRUMENTOS DE MEDICIÓN

En este capítulo, explicaremos cuáles son las medidas que podemos realizar sobre un circuito, cuál es la manera apropiada de hacerlo y qué instrumentos necesitamos.

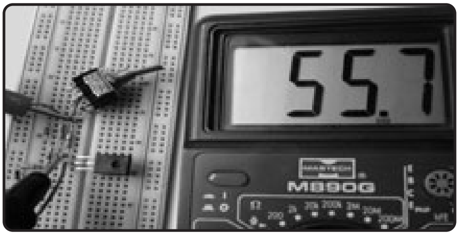

www.FreeLibros.me

CAPÍTULO 4
DISEÑO DE CIRCUITOS IMPRESOS

Conoceremos las características, funciones y limitaciones de algunos paquetes de diseño para la creación, edición y gestión de proyectos electrónicos.

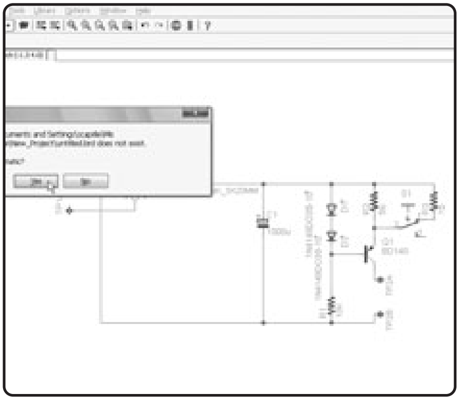

CAPÍTULO 5
SIMULACIÓN DE CIRCUITOS EN LA PC

En esta clase explicaremos qué es la simulación de circuitos, su utilidad en el proceso de diseño y cómo simular un circuito, utilizando el software ISIS.

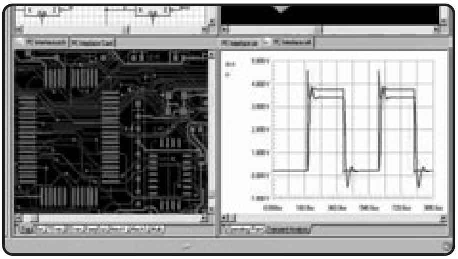

CAPÍTULO 6
FUENTES DE ALIMENTACIÓN

La fuente de alimentación es la encargada de proveer la tensión adecuada para el correcto funcionamiento de un circuito. Conoceremos sus características y funciones.

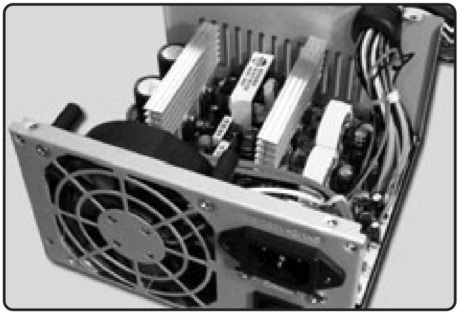

APÉNDICE
EL MERCADO LABORAL

En este apéndice, describiremos el amplio abanico laboral sobre el cual se desarrollaran las personas con conocimientos en electrónica.

SERVICIOS
AL LECTOR

En este último apartado, encontraremos un índice temática que nos ayudará a encontrar de forma más rápida y precisa los principales conceptos de la obra.

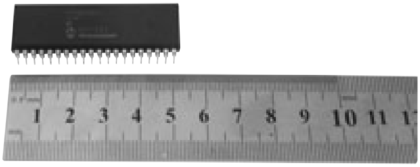

www.FreeLibros.me

Contenido **del libro**

Prólogo al contenido 003
El libro de un vistazo 004
Introducción a Sitios web 010

CAPÍTULO 1
ELECTRÓNICA DIGITAL 011

La electrónica digital 012
El concepto 013
- Los componentes 014
El taller 016
- Sobre la seguridad 017
Las herramientas 020
- Soldador y accesorios 022
- Los instrumentos 023
- Accesorios y complementos 025
Multiple choice 026

CAPÍTULO 2
CONSTRUCCIÓN DE CIRCUITOS 027

Primeros pasos 028
Circuitos electrónicos 028
- Los componentes 028
- El esquemático 029
- El PCB 029
- La transferencia del trazado a la placa 030
- La perforación y el estañado 030
El protoboard 031
- Topología 031
- Accesorios útiles 032
- Limitaciones 033
Uso del protoboard 033
- Alimentación del circuito 034
- Colocación de componentes 034
- Consejos finales 035
Cómo trabajar con el protoboard 036
Circuito impreso universal 041
- UPCB 045
- Montaje del circuito 045
Circuito impreso en detalle 046
- Método de Fabricación casera 046
- Fabricación profesional 046
- Construcción de un circuito impreso 047
Soldadura para electrónica 047
- Material para soldar 052
- Proceso de soldado 052
- Desoldado 053
Multiple choice 056

CAPÍTULO 3
INSTRUMENTOS DE MEDICIÓN 057

Instrumentos de medición 058
Instrumentos analógicos 058

www.FreeLibros.me

Preliminares_Electronica.qxp 04/07/2011 16:01 PÆgina 7

- Alcance de tensión del instrumento analógico 058
- Cifra en Ohm/Volt 059
- Infografía 3: Instrumentos de medición 060
- Amperímetro analógico 062
- Índice de clase de un instrumento analógico 062
- El multímetro digital 063
- Principio de funcionamiento 063
- Uso del multímetro digital 064

Cargador de baterías Ni-Cd 067
- Comprobaciones de continuidad 068
- Parámetros de componentes 068

Mediciones con el circuito alimentado 070
- Medición de la corriente de carga Ic 071
- Medición sobre la batería durante la carga 072

El osciloscopio 072
- Mediciones con osciloscopio 072
- Tipos de osciloscopios 072
- Osciloscopio analógico 072
- Osciloscopio de almacenamiento digital (DSO) 072
- Osciloscopio de fósforo digital (DPO) 074

Funcionamiento del osciloscopio 074
- Ancho de banda 075
- Disparo 075
- Los ojos del osciloscopio 075
- Calibración 076
- Clips de masa 077
- Mediciones de formas de onda 078

Medición de tensión y frecuencia 079
- Medición de tensión 079
- Medición de frecuencia 079

Multiple choice 080

CAPÍTULO 4
DISEÑO DE CIRCUITOS IMPRESOS 081

Diseño de circuitos impresos en la PC 082
- Método de diseño tradicional 082
- Entornos de diseño CAD 084
- El módulo de captura de esquemáticos 085
- El módulo de CAD 086
- El control de errores 086
- Cadsoft EAGLE 087

Interfaz de los módulos 088
- Componentes disponibles 089
- Diseño de un circuito esquemático 089

Creación de librerías 091

Generación del circuito impreso 093
- Función de ruteo automático 094
- Manejo de capas 095
- Cómo utilizar CadSoftEAGLE 095
- Diseño de una PCB 095

Limitaciones de la versión freeware 104
- Reglas de ruteo 105

Sistema métrico y mils 106
- Tipos de encapsulados más comunes 107

Multiple choice 108

CAPÍTULO 5
SIMULACIÓN DE CIRCUITOS EN LA PC 109

Diseño y simulación 110
- Proceso de diseño 110
- Ventajas de la simulación de circuitos 112

www.FreeLibros.me

- Modelado de circuitos electrónicos 112

Simulación lógica funcional 113
- La lista de red 114
- Programas de simulación 115
- Proteus 115
- OrCAD 116
- Altium Designer 117

Software ISIS 118
- Entorno ISIS 119

Área de trabajo 120
- Barras de acceso rápido 120
- Desplazamiento por el área de trabajo 122

Diseño en ISIS 124
- Librerías de componentes 124
- Inserción de componentes 125

Diseño de un circuito 126
- Inserción de masa y alimentación 127
- Cableado de dispositivos 128
- Instrumentos de simulación 128
- Simulador del cargador de baterías 129

Simulación de código con ISIS 142

Limitaciones de las versiones 143

Multiple choice 144

CAPÍTULO 6
FUENTES DE ALIMENTACIÓN 145

Las fuentes de alimentación 146
- Especificaciones de las fuentes 148
- Protección 149

Fuentes lineales y conmutadas 149
- La fuente lineal 150
- La fuente conmutada 150

Comparación entre fuentes lineales y conmutadas 152

Etapas de una fuente de alimentación lineal 153
- El transformador 153
- ¿Cómo elegir? 154
- El rectificador 154
- El rectificador de media onda 155
- Los rectificadores de onda completa 156
- Rectificador puente de Graetz 157
- Filtros de salida 158

Diseño de una fuente de alimentación 162
- Tensión de rizado o ripple 163

Los reguladores de voltaje 165
- Reguladores fijos 165
- Los reguladores variables 168
- Fuente de alimentación para el taller 169

Multiple choice 176

APÉNDICE
EL MERCADO LABORAL 177

El mercado laboral 178
- Integración 180

SERVICIOS
AL LECTOR 181

Índice temático 182

Catálogo 185

www.FreeLibros.me

RedUSERS
MEJORA TU PC

Desarrollos temáticos en profundidad

Libros.

Coleccionables.

Cursos intensivos con multimedia

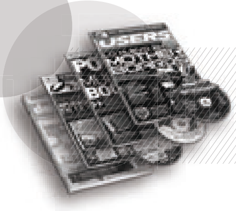

Capacitación dinámica

Revistas.

Sitios Web.

Noticias al día, downloads, comunidad

Información actualizada al instante

Newsletters.

La red de productos sobre tecnología más importante del mundo de habla hispana.

redusers.com

www.FreeLibros.me

Introducción a Electrónica Práctica

Este libro, realizado por expertos, ofrece un extenso y muy completo recorrido por el maravilloso mundo de los sistemas digitales, donde tanto el ensayo como el error son claves fundamentales para perfeccionarse.

A lo largo de los primeros capítulos, conoceremos cuál es el alcance de la Electrónica y sus aspectos prácticos. Veremos en detalle las herramientas e instrumentos infaltables para un taller que se especialice en este área. Además, nos familiarizaremos con la construcción de circuitos electrónicos mediante la utilización de un protoboard.

Una vez que conozcamos las primeras nociones sobre el tema, nos iniciaremos en el mundo de la instrumentación electrónica: aprenderemos a usar el voltímetro, el amperímetro y los multímetros digitales, para medir los parámetros más importantes de un circuito. Siguiendo el orden de aprendizaje, conoceremos el proceso de simulación de circuitos en la PC con ISIS y su utilidad en proyectos electrónicos.

Además, desarrollaremos cuáles son los programas y las funciones para simular el comportamiento real de un circuito electrónico. Por otro lado, conoceremos las funciones y las características de las fuentes de alimentación, su utilidad y los diferentes tipos existentes. Aprenderemos cada una de las etapas de su circuito y los parámetros para diseñar.

Encontrarán, en esta obra, las herramientas para realizar sus primeros proyectos. Es una excelente oportunidad para que transitemos juntos este sorprendente recorrido donde se combinan aspectos vinculados con la programación, el hardware y la Robótica, entre muchos otros. A lo largo de todo el libro, encontrará un completo desarrollo teórico y podrá llevar a cabo proyectos de aplicación real.

www.FreeLibros.me

Capítulo 1

Electrónica Digital

En este primer capítulo, haremos una introducción a los principales conceptos de Electrónica Digital.

RedUSERS.com

www.FreeLibros.me

Electrónica **Digital**

En la actualidad, y desde hace muchos años, la **Electrónica Digital** ha estado presente en infinidad de aparatos domésticos, industriales y militares (**Figura 1**). Nació al mismo tiempo que la **Electrónica Analógica**, pero tomó especial significación desde el momento en que se empleó para el diseño y la construcción de **sistemas básicos de cómputo**, el **control industrial**, los **autómatas programables**, los **microcontroladores**, los **microprocesadores** y los **computadores personales**. Gracias a que la Electrónica Digital considera solo dos valores –concepto que veremos más adelante–, es mucho más simple procesar y manipular las señales de forma digital que hacerlo de manera analógica. Esta ventaja se produce a partir de que esta última resulta inexacta y costosa, considerando la gran cantidad de componentes que necesita para su fabricación: **transistores** y **resistores electrónicos**.

> La Electrónica Digital ha estado presente en infinidad de aparatos domésticos, industriales y militares

FIGURA 1. **Actualmente, la electrónica brinda un abanico de posibilidades laborales muy diversas.**

SEÑALES DIGITALES

Si bien podemos manejar señales digitales con un transistor que trabaja de forma lineal o analógica, lo cierto es que no tiene ningún sentido hacerlo, ya que no nos interesan los valores intermedios, sino solo los extremos mínimos y máximos de una señal.

www.FreeLibros.me

EL CONCEPTO

La Electrónica Analógica trata con variables (**tensión**, **corriente** y **potencia**, entre otras) que varían de forma continua con el tiempo. Como ejemplo podemos citar una señal de audio o la tensión que proporciona un sensor de temperatura. Todas estas señales están compuestas por múltiples valores de tensión y la Electrónica Analógica considera todos estos valores. Como contrapartida, en la Electrónica Digital, las variables pueden tomar solo **dos valores** o **estados** posibles y bien definidos. Dado que estamos tratando el aspecto eléctrico del concepto, esos dos valores pueden ser **0 V** y **5 V** o, de forma binaria, **0** y **1**. La elección del valor de **5 V** no es caprichosa o antojadiza. Es debido a que la mayoría de los sistemas digitales trabajan con una tensión de **5 V**. Sin embargo, en la actualidad, podemos encontrar a menudo sistemas que operan con tensiones de **3,3 V** y aún menores, dependiendo del circuito.

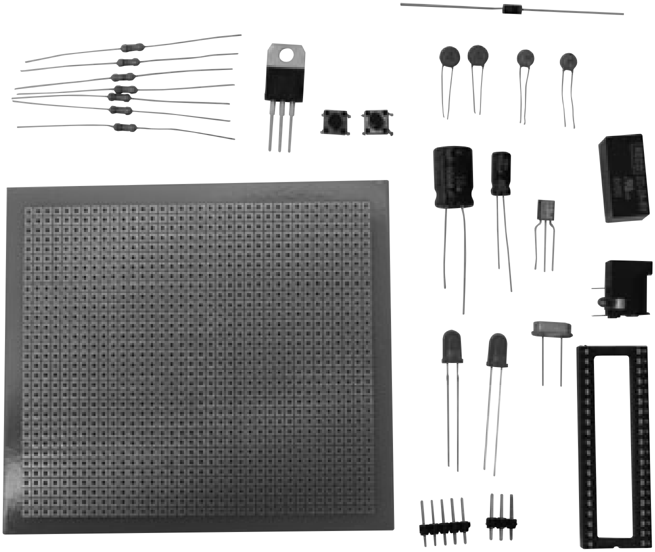

ANALÓGICA Y DIGITAL

Un circuito analógico es mucho más costoso de diseñar y construir. Es por este motivo que el concepto de Electrónica Digital cobra sentido y evoluciona constantemente. En este libro, analizaremos todas sus ventajas y desventajas.

www.FreeLibros.me

Términos fundamentales

ELECTRÓNICA ANALÓGICA
Trata con variables, como la tensión, la corriente y la potencia, que varían de forma continua con el tiempo.

ELECTRÓNICA DIGITAL
Las variables pueden tomar solo dos valores o estados posibles y definidos.

DAC
Es un circuito cuya función es convertir un valor digital en uno analógico.

ADC
Es un circuito cuya función consiste en convertir un valor analógico en digital.

DISCRETOS
Son componentes primitivos, como las resistencias, diodos, capacitores y transistores, entre otros.

TRANSISTOR
Dispositivo electrónico semiconductor que cumple funciones de amplificador, oscilador, conmutador o rectificador.

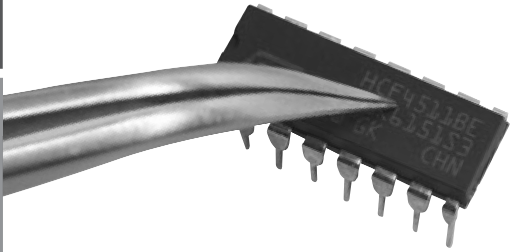

LOS COMPONENTES

En Electrónica –tanto Analógica como Digital– se utilizan exactamente los mismos **componentes discretos**. Cuando decimos discretos, queremos significar que se trata de **componentes primitivos**, como las resistencias, los diodos (**Figura 2**), los capacitores y los transistores, entre otros. Por ejemplo, un circuito integrado no es un componente discreto, ya que contiene un circuito en su interior que hace un trabajo determinado por el fabricante. Sin embargo, el circuito integrado está construido en base a varios componentes discretos interconectados entre sí.

Es entonces que nos preguntamos cuál es la diferencia, ya que se utilizan los mismos componentes, tanto para Electrónica Analógica como para Digital. La diferencia se halla en la manera de hacer funcionar estos componentes. Observemos lo siguiente:

Los componentes discretos son: resistores, diodos, capacitores y transistores

www.FreeLibros.me

ELECTR_Cap1_xxx_xxx_ajustado.qxp 04/07/2011 16:30 PÆgina 15

un transistor puede funcionar como una resistencia variable que opera como un interruptor eléctrico que anula el paso por completo de la corriente o la deja pasar. Estos dos estados del transistor se los conoce como estado de **bloqueo** (cuando está abierto) y estado de **saturación** (cuando se encuentra cerrado).

Cuando un transistor funciona solo con los estados de bloqueo y saturación (como si fuese un interruptor) podremos decir que está operando de **manera digital**, ya que no considera los valores intermedios. Un ejemplo de ello puede ser un **interruptor**, que conecta o desconecta una bombilla de luz. Si, en cambio, hacemos funcionar un transistor de manera tal que tenga en cuenta los estados o valores intermedios, el transistor estará operando de **manera lineal** o analógica. Como ejemplo, podemos citar el **regulador de luminosidad** o **dimmer** con que cuentan algunos hogares, que permite graduar la intensidad de luz deseada emitida por la bombilla.

A lo largo de todo el libro, se estudiarán las **compuertas digitales**. Si bien comercialmente se adquieren bajo la denominación de **Circuito Integrado** (**CI**), interiormente son transistores que operan de forma digital, que es lo mismo que decir en estado de bloqueo y saturación.

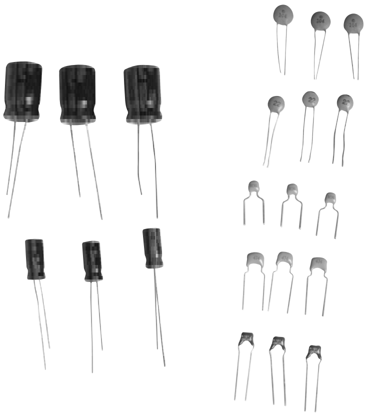

FIGURA 2.
El diodo permite la circulación de corriente eléctrica en un sentido. El más común es el diodo semiconductor.

DESCARGA A TIERRA

Es fundamental conectar a tierra todas las clavijas centrales de los tomacorrientes y de la pulsera de descarga estática. Si no disponemos de una conexión a tierra en la red domiciliaria, podemos utilizar una jabalina enterrada, siguiendo las especificaciones del fabricante.

www.FreeLibros.me

El **taller**

El taller es el espacio en el cual efectuaremos todas las pruebas y tareas de Electrónica (**Figura 3**). Por este motivo, es necesario equiparlo adecuadamente. Veamos qué necesitamos. Debe proveer una adecuada ventilación **para evacuar los gases emitidos** por desengrasantes, tóxicos, productos químicos empleados en la fabricación de circuitos impresos y los compuestos de soldadura. Para lograr una buena ventilación, es necesario tener extractores de aire para asegurarnos la renovación constante de oxígeno.

EL BANCO DE TRABAJO

El banco o **mesa de trabajo** debe tener una altura adecuada, el estándar es de **80 centímetros**. Si bien el uso de **cajones** no es aconsejable ya que entorpecen nuestro movimiento, pueden aceptarse si los tiene solo a los costados. Como manipularemos artefactos eléctricos, el banco debe ser de **madera** o un **compuesto aislante de fibra**. Siempre es recomendable adherir, en toda su superficie, una lámina de goma lisa de un espesor de **2 milímetros**.

El banco debe ser muy robusto, ya que permanentemente soportará el peso de los equipos de medición, herramientas y, ocasionalmente, aparatos muy pesados, que contengan transformadores o baterías. Sobre su **fondo** podremos agregar una alzada o estante para colocar equipos y componentes sin ocupar espacio de trabajo en su superficie.

Es importante incorporarle una **lámpara con lupa** (lente de aumento) de brazo flexible, para poder ajustar con la mano la posición exacta del circuito que necesitemos. Debemos proveerlo de cuatro o más **tomacorrientes** de **tres clavijas**

FIGURA 3. El ambiente donde desarrollemos nuestra tarea debe ser un espacio bien iluminado en general, independientemente de la iluminación puntual sobre el banco de trabajo.

www.FreeLibros.me

El banco o mesa de trabajo debe tener una altura adecuada, el estándar es de 80 centímetros

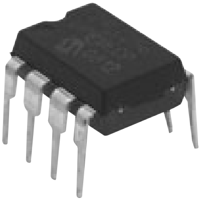

con conexión a tierra a cada lado (izquierdo y derecho). Es sumamente relevante contar con un **interruptor térmico bipolar** para desconectar eléctricamente el banco en caso de ser necesario.

SOBRE LA SEGURIDAD

El primero de los equipos de seguridad que debemos instalar en el taller es el interruptor diferencial o disyuntor. A pesar de tomar todos los recaudos posibles para mantenernos aislados de tierra, existe una situación de extrema peligrosidad y es nuestra conexión a la **pulsera antiestática**. No olvidemos que esta pulsera nos conecta a tierra, todo lo contrario a lo que evitamos con el piso aislante de goma y los recaudos de aislamiento del banco de trabajo.

En caso de que, accidentalmente, toquemos con nuestra mano el terminal positivo o "vivo" de la red de canalización, quedaremos con nuestro cuerpo conectado directamente a la red eléctrica. Si bien el interruptor diferencial actuaría en consecuencia, es preferible no llegar a esta situación. Para ello, evitemos colocarnos la pulsera antiestática, a menos que sea indispensable.

Una solución definitiva a este problema es aislar eléctricamente el taller por completo o, en su defecto, el banco de trabajo de la red eléctrica externa. Para hacerlo, podemos colocar un **transformador** con relación **1:1** y de la potencia adecuada.

Por ejemplo, si la **tensión nominal es de 110 V** (depende del país), el transformador deberá ser de **110 V** de entrada y **110 V** de salida, conectando esta última al banco de trabajo. Si tenemos previsto conectar aparatos de gran consumo, debemos emplear un transformador con una potencia mínima de **1 KVA** (**VA: VoltAmper - K: Kilo**) o superior.

PULSERA ANTIESTÁTICA

Consiste en una cinta, con un abrojo para ajustarla a la muñeca, conectada a un cable de toma de tierra que permite descargar cualquier acumulación de electricidad estática en nuestro cuerpo. Lleva una resistencia de **1 MOhm**.

www.FreeLibros.me

INFOGRAFÍA 1: EL TALLER ELECTRÓNICO

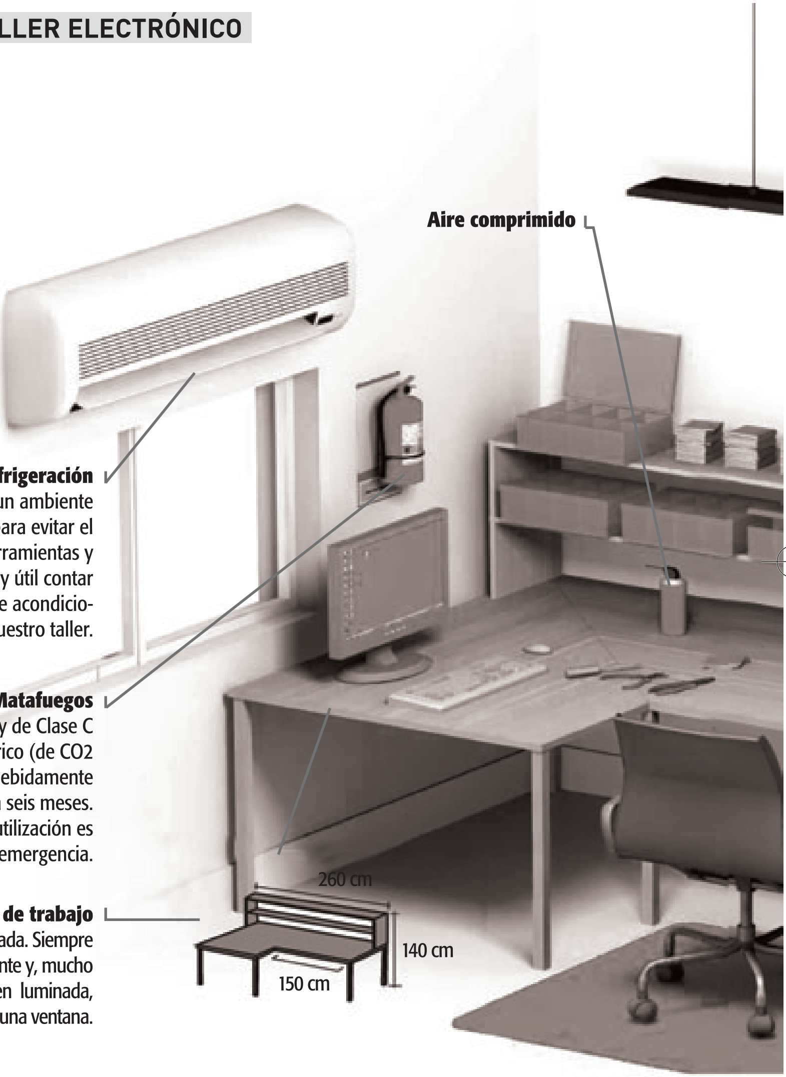

Refrigeración
Conviene trabajar en un ambiente cómodo y seco para evitar el deterioro de herramientas y componentes. Es muy útil contar con un equipo de aire acondicionado en nuestro taller.

Matafuegos
Matafuegos siempre a mano y de Clase C para incendios de origen eléctrico (de CO2 o de tipo Polvo ABC), debidamente cargados y controlados cada seis meses. Darle una correcta utilización es fundamental en situaciones de emergencia.

Mesa de trabajo
Cómoda, amplia, limpia y ordenada. Siempre de madera o de material aislante y, mucho mejor, con estantes. Debe estar bien luminada, en lo posible, cerca de una ventana.

www.FreeLibros.me

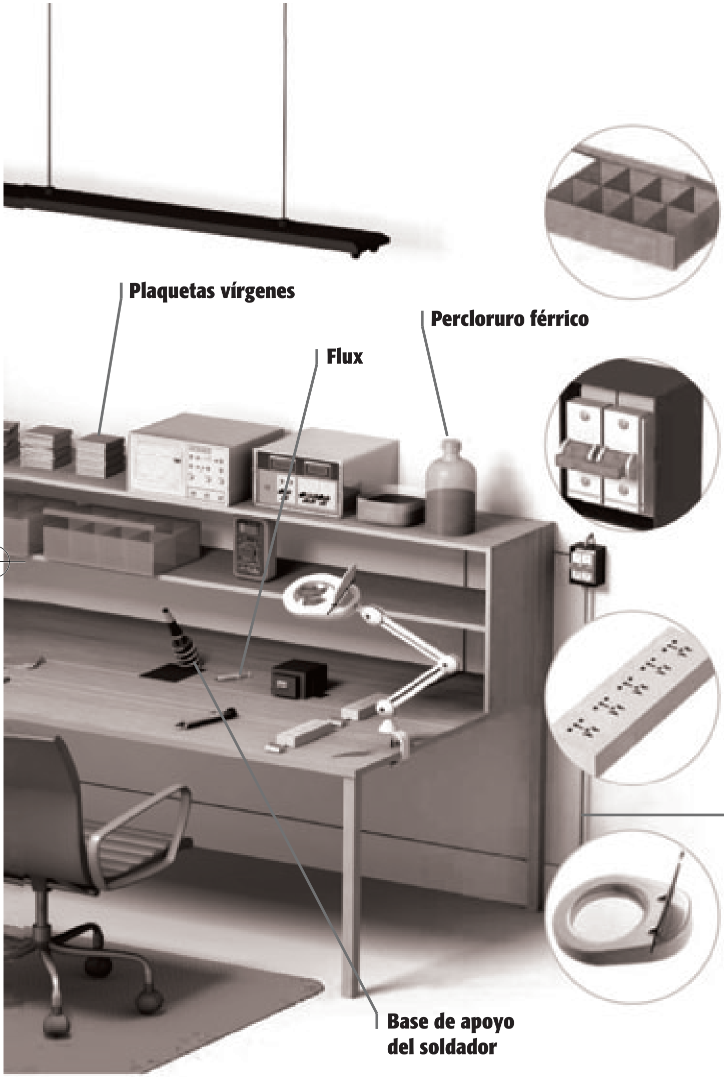

Gaveteros
Estos cómodos clasificadores de elementos nos permitirán mantener ordenado el taller, teniendo siempre a disposición los componentes en forma ágil y rápida.

Disyuntor
Es imprescindible contar con un disyuntor en la entrada de la casa y, al menos, un interruptor termomagnético individual, en la sección del taller, para trabajar con riesgo mínimo. El dimensionado debe realizarlo personal matriculado.

Tomacorrientes
Se necesitan una puesta a tierra general en buen estado y tomacorrientes de 3 bocas en la instalación. Las zapatillas se ubicarán de manera que permitan fácil acceso, evitando cruces de cables.

Jabalina a tierra

Lupa articulada
Muy útil para tareas de soldado through-hole y superficial. Se utiliza para observar el estado final de Zlas soldaduras y para facilitar las maniobras sobre los diferentes circuitos electrónicos.

www.FreeLibros.me

Claves de seguridad

RED ELÉCTRICA

Debemos asegurarnos de que la red domiciliaria se encuentre en perfectas condiciones de funcionamiento. De ello dependerá nuestra seguridad.

TOMACORRIENTES

Es preciso tener una buena cantidad de tomacorrientes en cada uno de los rincones del taller. De este modo, podremos conectar varios dispositivos sin necesidad de alargues.

DESCARGA A TIERRA

En la actualidad, un taller de Electrónica que no posea un sistema de descarga a tierra es impensable; de él dependerá nuestra seguridad personal.

MATAFUEGO

Es natural que en un taller de Electrónica se genere algún cortocircuito y, para evitar incendios, es recomendable tener a mano un matafuego para fuegos clases A B C.

El taller es el espacio de trabajo en el que pasaremos gran parte del tiempo. Es por eso que debemos mantenerlo en óptimas condiciones operativas y de seguridad. Un espacio incómodo impide el desarrollo de tareas prolongadas (**Figura 4**).

Las herramientas

El taller puede convertirse en un verdadero laboratorio electrónico, siempre que tengamos los instrumentos y las herramientas adecuadas. Veamos cada herramienta en detalle (**Figura 5**).

- **Alicate**: el alicate común produce un corte en forma de "**V**" y está indicado para cortar cables o terminales gruesos. Tiene que ser de mango aislado pero no necesariamente grueso, ya que no lo usaremos en líneas energizadas, sino en alambres de conexión sin energía domiciliaria.
- **Pinza de fuerza**: es una pinza convencional con mango aislado, utilizada para sujetar tuercas cuando montamos disipadores de calor y, también, otros elementos.
- **Pinza de punta**: puede ser de punta recta u oblicua y se emplea para sujetar componentes o doblar terminales, según sea necesario.

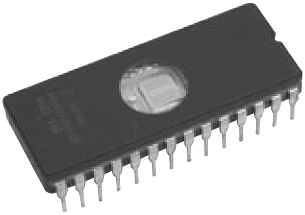

www.FreeLibros.me

FIGURA 4. **Un taller inseguro pone en riesgo nuestra salud y los componentes que debemos reparar.**

- **Pinzas Bruselas**: son pequeñas pinzas de punta que se usan para manipular pequeños componentes, por ejemplo, para introducir los alambres en un protoboard o tablero de experimentación.
- **Destornilladores**: aunque resulte obvio, los destornilladores en Electrónica cumplen un papel fundamental para ajustar controles electrónicos y sujetar o retirar tornillos. Recomendamos adquirir un kit para Electrónica de destornilladores pequeños, con punta plana y en cruz.

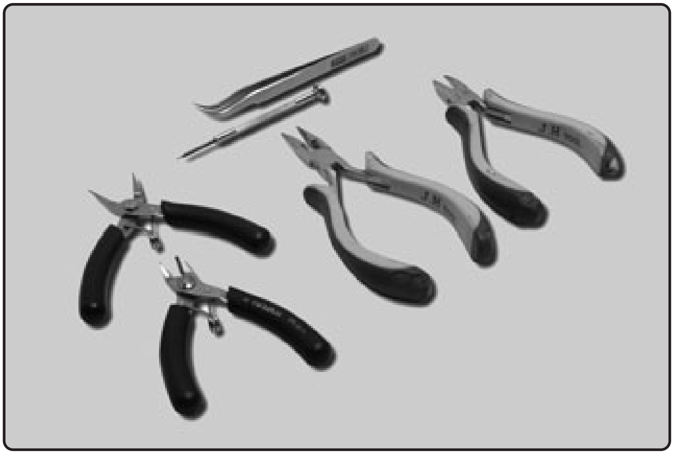

FIGURA 5. **El taller ideal se construye progresivamente. Primero, se adquieren las herramientas más elementales; luego, las más complejas y, finalmente, los instrumentos.**

www.FreeLibros.me

- **Fuente de alimentación**: si bien no es un instrumento de medición, la mencionamos aquí porque es indispensable para efectuar todos los experimentos o prácticas en nuestro taller. Una pequeña fuente que suministre dos tensiones fijas y reguladas será más que suficiente (**Figura 6**).

SOLDADOR Y ACCESORIOS

Además de las herramientas mencionadas, necesitamos varios accesorios para reparaciones electrónicas:

- **El soldador** es fundamental para los trabajos electrónicos, ya que todas las conexiones son unidas por una aleación de estaño y plomo o soldadura libre de plomo (**Figura 7**). Es importante el uso de un soporte para poder depositarlo sobre el banco caliente sin riesgos. Algunos soportes, además, cuentan con una espuma renovable de limpieza para la punta del soldador, algo muy útil debido a los residuos que se acumulan en ese sector, que pueden entorpecer la correcta soldadura.

- **Estaño o soldante**: es un alambre de un diámetro de entre **0,5 mm** y **1 mm**, compuesto de una aleación de **60%** de **plomo** y **40%** de **estaño**. Además, en su núcleo cuenta con una pasta a base de resina natural que ayuda a la soldadura. Requiere almacenamiento refrigerado.

Necesitamos también ciertos accesorios en el taller de reparaciones electrónicas

- **Pasta para soldar o "flux"**: es un compuesto en forma de pasta o líquido, que ayuda a soldar/desoldar debido a sus propiedades limpiadoras, ya que baja el punto de fusión del estaño. Es ampliamente utilizado en la soldadura de componentes **SMD**.
- **Bomba succionadora de estaño**: es una pequeña bomba de vacío accionada a mano que, al momento de dispararla sobre la soldadura caliente, la aspira y la remueve. Es indispensable para desoldar circuitos integrados.
- **Cinta desoldante**: es una malla o entretejido de hilos de cobre muy delgados, que están tratados con un producto similar a la pasta para soldar. Para desoldar, apoyamos la punta de la cinta sobre la soldadura que deseamos eliminar, ubicamos el soldador encima de la cinta y dejamos que el calor la traspase hasta fundir la soldadura. Esta se licuará y será absorbida por la malla de cobre. También es fundamental para desoldar circuitos integrados o efectuar una limpieza de los puntos de soldadura de los circuitos impresos y demás componentes.

ALICATE DE CORTE AL RAS

Se utiliza para cortar los terminales sobrantes en una tarjeta de circuito impreso, ya que sus hojas de corte tienen un lado en ángulo y otro totalmente plano, que puede lograr un corte realmente al ras de la tarjeta. No debemos emplearlo para cables o terminales gruesos.

www.FreeLibros.me

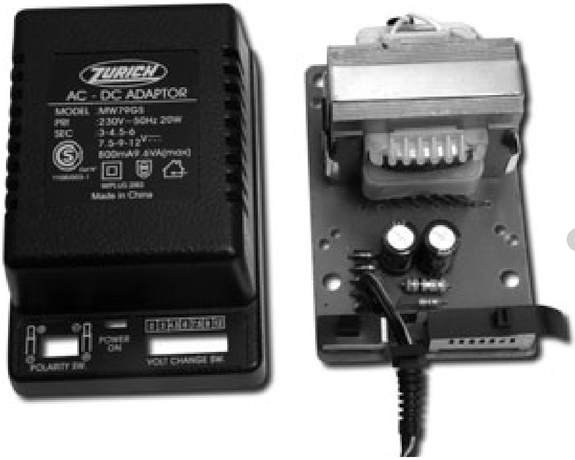
ZURICH
AC - DC ADAPTOR
MODEL MW79GS
PRI 230V~50Hz 20W
SEC 3-4.5-6
7.5-9-12V
800mA
Made in China
POLARITY SW
POWER ON
VOLT CHANGE SW

FIGURA 6.
Las tensiones recomendadas son de 5 V **y** 12 V **de** 1 Amper **o superior.**

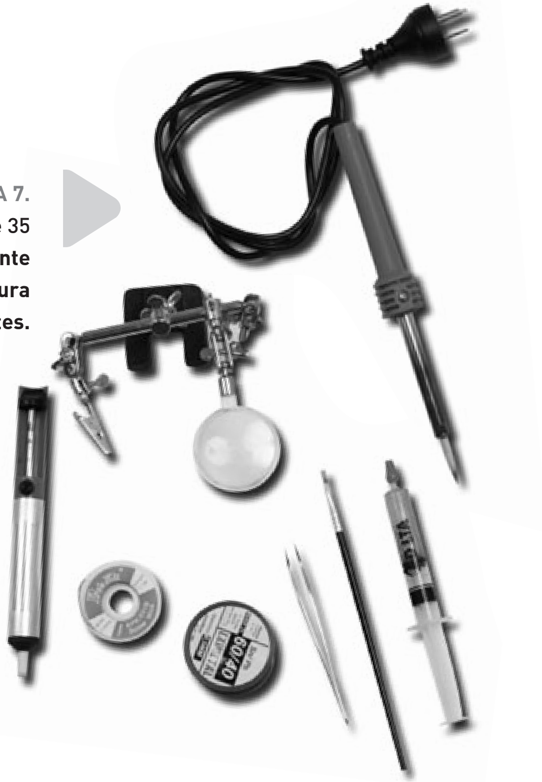

FIGURA 7.
Recomendamos un soldador del tipo lápiz de 35 **o** 40 **Watts de potencia porque resulta suficiente para la mayoría de los trabajos de soldadura de componentes y partes.**

LOS INSTRUMENTOS

Además de las herramientas y los accesorios mencionados anteriormente, el taller debe contar con los instrumentos imprescindibles (**Figura 8**), que detallaremos a continuación:

- **Multímetro Digital**: también conocido como **DMM** (*Digital MultiMeter*), es el principal instrumento de medición de todo taller electrónico. Por medio del DMM, podremos medir tensiones, corrientes y resistencias, además de cortocircuitos.

www.FreeLibros.me

Algunos modelos tienen un zumbador que emite un sonido cuando se tocan sus puntas, y es muy útil para seguir trazas en un **PCB** (*Printed Circuit Board*) o cortocircuitos. Los modelos más sofisticados disponen de una sonda para medir temperatura en determinadas partes de un circuito. Otros de los agregados son medidores de capacidad en condensadores o capacitores, medidores de transistores bipolares y **FET** (*Field Effect Transistor* - Transistor de Efecto de Campo). En el **Capítulo 2** estudiaremos el uso del DMM.

- **Sonda o punta lógica**: una punta lógica permite visualizar, por medio de **LEDs** (**Diodo Emisor de Luz**), los estados lógicos en un punto del circuito. Por ejemplo, si en el punto de medición hay **0 V**, se iluminará el led que indica un **cero lógico** (**0**). En cambio, si hay **5 V**, se iluminará solo el led que indica el estado que denominaremos binario uno (**1**).
- **Osciloscopio**: el osciloscopio es un instrumento de medición y visualización de señales en el transcurso del tiempo.

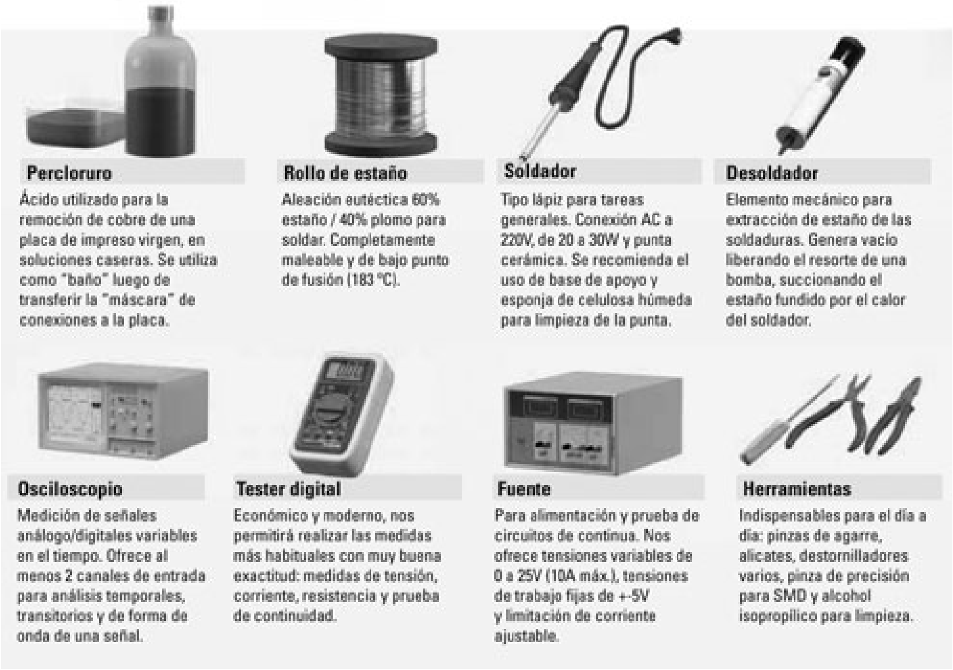

Percloruro
Ácido utilizado para la remoción de cobre de una placa de impreso virgen, en soluciones caseras. Se utiliza como "baño" luego de transferir la "máscara" de conexiones a la placa.

Rollo de estaño
Aleación eutéctica 60% estaño / 40% plomo para soldar. Completamente maleable y de bajo punto de fusión (183 °C).

Soldador
Tipo lápiz para tareas generales. Conexión AC a 220V, de 20 a 30W y punta cerámica. Se recomienda el uso de base de apoyo y esponja de celulosa húmeda para limpieza de la punta.

Desoldador
Elemento mecánico para extracción de estaño de las soldaduras. Genera vacío liberando el resorte de una bomba, succionando el estaño fundido por el calor del soldador.

Osciloscopio
Medición de señales análogo/digitales variables en el tiempo. Ofrece al menos 2 canales de entrada para análisis temporales, transitorios y de forma de onda de una señal.

Tester digital
Económico y moderno, nos permitirá realizar las medidas más habituales con muy buena exactitud: medidas de tensión, corriente, resistencia y prueba de continuidad.

Fuente
Para alimentación y prueba de circuitos de continua. Nos ofrece tensiones variables de 0 a 25V (10A máx.), tensiones de trabajo fijas de +-5V y limitación de corriente ajustable.

Herramientas
Indispensables para el día a día: pinzas de agarre, alicates, destornilladores varios, pinza de precisión para SMD y alcohol isopropílico para limpieza.

FIGURA 8. **El taller puede convertirse en un verdadero laboratorio electrónico, siempre que tengamos los instrumentos y las herramientas adecuados.**

www.FreeLibros.me

ACCESORIOS Y COMPLEMENTOS

Además de las herramientas, instrumentos y productos mencionados hasta el momento, un taller se completa con los siguientes accesorios:

- **Protoboard**: es una tarjeta con perforaciones a lo largo de toda la superficie, que permite construir un circuito electrónico sin soldadura. Las conexiones entre los componentes se realizan con alambre del tipo convencional.
- **Circuito impreso**: es una placa de PCB perforada en toda su superficie, que posibilita introducir los componentes y soldarlos. Las conexiones se efectúan manualmente.
- **Resistores**: recomendamos tener todas de 1/4 de Watt y una cantidad de **10** por cada valor: **100 ohms** – **330 Ohms** - **470Ω** - **1 KÙ** - **2,2 KΩ** - **3,3 KΩ** - **4,7 KΩ** - **10 KΩ** - **22 KΩ** - **47 KΩ** - **100 KΩ**.
- **Capacitores cerámicos**: Todos con aislamiento para **25 V** o más y una cantidad de **10** por cada valor: **.01 µF - 1 µf.** (**µ** significa micro y la letra **F** es la unidad de medida, el **Faradio**). Por lo tanto, **µF** significa **microfaradio**. Cuando se indica un punto delante de la capacidad, como **.01**, en realidad es para resumir, queriendo significar **0,01**.
- **Capacitores electrolíticos**: todos con un aislamiento de **25 V** o más y un cantidad de **10** por cada valor: **1 µF** – **10 µF** - **100 µF** - **470 µF** - **1000 µF**.

> El analizador es una sonda que permite ver, en pantalla, varias señales al mismo tiempo

- **Transistores**: el transistor BC548 es de uso muy frecuente en Electrónica Digital. Recomendamos tener una cantidad de 10 unidades disponibles.
- **Diodos**: el diodo de conmutación por excelencia es el 1N4148. Con una decena de ellos, será suficiente.
- **Reguladores**: en este caso, solo será necesario contar con **5** unidades de cada tipo: **7805** - **7812**.
- **Zócalos Dip**: los zócalos permiten insertar, sin soldadura, integrados, memorias o microcontroladores. Cantidad: **5** de cada tipo: **4x2** - **7x2** - **8x2** - **9x2**.
- **Micro Switch**: pequeños interruptores que tienen el mismo tamaño y distribución de contactos que un circuito integrado ordinario y convencional.
- **Led**: indispensable como indicador lumínico. Cantidad: varios colores, de varios diámetros.
- **Indispensables**: Fusibles de **1 Ampe**r y **20 mm** de longitud - forro o **spaghetti termo contraible** - cables de varios colores - alambres para insertar en protoboard.

MÁS SOBRE SOLDADORES

Algunos soldadores tienen un regulador de temperatura, aunque no es indispensable para los trabajos comunes del taller. No aconsejamos el uso de los llamados soldadores instantáneos, ya que la punta no es lo suficientemente fina y precisa para trabajar en electrónica.

www.FreeLibros.me

Multiple **choice**

1 ¿Cuál de las siguientes herramientas es un alambre compuesto por una aleación 60% plomo y 40% estaño?

a- Pasta de soldar o "flux".
b- Cinta desoldante.
c- Estaño o soldante.
d- Multímetro.

2 ¿Cuál de las siguientes herramientas es una malla o entretejido de hilos de cobre muy delgados?

a- Pasta de soldar o "flux".
b- Cinta desoldante.
c- Estaño o soldante.
d- Multímetro.

3 ¿Cuál de las siguientes herramientas es un compuesto en forma de pasta o líquido, que ayuda a soldar/desoldar?

a- Pasta de soldar o "flux".
b- Multímetro.
c- Bomba succionadora de estaño.
d- Osciloscopio.

4 ¿Cuál de las siguientes herramientas es una pequeña bomba de vacío accionada a mano?

a- Estaño o soldante.
b- Multímetro.
c- Bomba succionadora de estaño.
d- Osciloscopio.

5 ¿Cuál de los siguientes es el principal instrumento de medición de todo taller electrónico?

a- Estaño o soldante.
b- Multímetro.
c- Bomba succionadora de estaño.
d- Osciloscopio.

6 ¿Cuál de los siguientes es un instrumento de medición y visualización de señales en el transcurso del tiempo?

a- Estaño o soldante.
b- Multímetro.
c- Bomba succionadora de estaño.
d- Osciloscopio.

Respuestas: 1 c, 2 b, 3 a, 4 c, 5 b, 6 d.

www.FreeLibros.me

Capítulo 2

Construcción de circuitos

En este segundo capítulo, analizaremos las bases de la composición de los circuitos digitales.

RedUSERS.com

www.FreeLibros.me

Primeros **pasos**

Para iniciarnos en el universo de la **Electrónica Digital** es necesario conocer los principios elementales del funcionamiento de los circuitos. Es por este motivo que comenzaremos a dar los primeros pasos en la construcción de un circuito sobre la base de un **protoboard**, plantilla que ofrece la ventaja de trabajar sin soldaduras.

Con el transcurso de las páginas avanzaremos sobre las características de los **circuitos universales**, conoceremos sus ventajas, limitaciones y utilización práctica. Además, en este segundo capítulo, subiremos la apuesta para detallar de forma teórica y práctica el proceso de **construcción** de un **circuito impreso casero**. Sobre el final de este capítulo, conoceremos todos los secretos sobre la **soldadura** para electrónica y cuáles son los **elementos** necesarios para realizarla correctamente.

Circuitos **electrónicos**

Un circuito electrónico es un conjunto de componentes eléctricos o electrónicos, interconectados por medio de hilos conductores, con el objetivo de generar, transportar o procesar una señal eléctrica.

Existen diferentes métodos de construcción de circuitos electrónicos, dependiendo de su complejidad, el volumen de producción y el tipo de componentes utilizados.

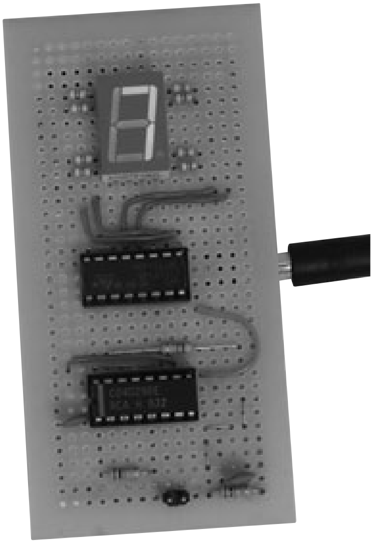

LOS COMPONENTES

Dentro de un circuito podemos encontrar una gran variedad de componentes, como **fuentes de señal** (sensores y micrófonos), **fuentes de alimentación** (baterías o la red domiciliaria), **dispositivos eléctricos** (bobinas, lámparas, resistencias y capacitores), **componentes electrónicos** (diodos, transistores, circuitos integrados) y **dispositivos mecánicos** (llaves y conmutadores), (**Figura 1**).

> El "esquemático" es la representación gráfica de un circuito electrónico

www.FreeLibros.me

Según el modo en que se monta un circuito, un componente puede ser de tecnología **through hole** (a través de orificio) o de **montaje superficial**.

EL ESQUEMÁTICO

Antes de comenzar con la construcción de cualquier circuito, debemos contar con un **esquema circuital**. El esquemático es una representación gráfica de un circuito electrónico. Muestra los diferentes componentes con su simbología asociada y las conexiones entre los mismos. Podemos dibujarlo a mano alzada o utilizar un **software** de captura de esquemático, como **Eagle** u **Orcad**. En este libro, utilizaremos el software Eagle.

La creación de un esquema circuital es clave, ya que cualquier error en la construcción derivará en el mal funcionamiento del circuito. Por esta razón recomendamos dibujar el esquemático con un software para tal fin, lo que nos permitirá también, en la mayoría de los programas disponibles, realizar una simulación del comportamiento real del circuito, predecir sus características antes de construirlo y generar el circuito impreso.

EL PCB

Una vez que tenemos el esquemático debemos construir físicamente el circuito. Generalmente, los circuitos se montan mediante un **circuito impreso** o **PCB** (*Printed Circuit Board*). Es una tarjeta plástica que conecta eléctricamente los componentes del circuito a través de pistas de cobre laminadas sobre un **sustrato no conductor** (fenólico o epoxi). Actúa también como soporte de todo el circuito.

A partir de un esquemático, el diseño del circuito impreso puede hacerse con un software como los mencionados anteriormente; estos poseen muchísimas librerías con las formas físicas de los componentes (**footprints**) para facilitar el diseño. El trazado de las pistas (**ruteo**) puede realizarse de forma manual o automática.

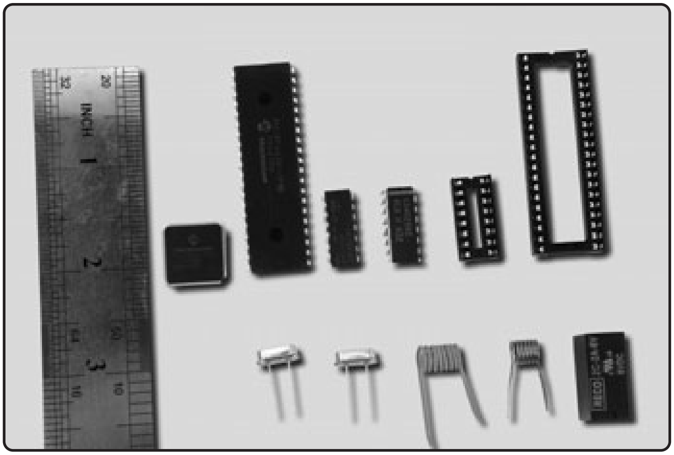

FIGURA 1.
Es importante que sepamos reconocer los componentes más utilizados y sus características para poder construir nuestros propios circuitos. Por ejemplo, un diodo emisor de luz posee polaridad, y no es lo mismo conectarlo hacia un lado o hacia otro.

www.FreeLibros.me

LA TRANSFERENCIA DEL TRAZADO A LA PLACA

Cuando el diseño está terminado, necesitamos pasarlo a la placa. Para ello existen diversos métodos, pero casi todos se hacen a partir de una lámina de cobre que cubre completamente el sustrato (placa virgen), donde luego se quita el cobre indeseado para dejar los trazados diseñados (**Figura 2**). Este proceso se logra utilizando primero una máscara de trazado, que se obtiene al aplicar tintas (serigrafía) o mediante un proceso de fotograbado.

También existen métodos donde se utiliza una fresa mecánica o hasta un láser para eliminar el cobre residual. Sin embargo, el método más accesible es el de la transferencia del diseño a la placa a través de calor. Para ello, se imprime el trazado en un **material termosensible** como el papel de ilustración.

LA PERFORACIÓN Y EL ESTAÑADO

Los orificios para colocar los componentes se realizan, a nivel industrial, con un taladro controlado por computadora. Nosotros podremos utilizar una

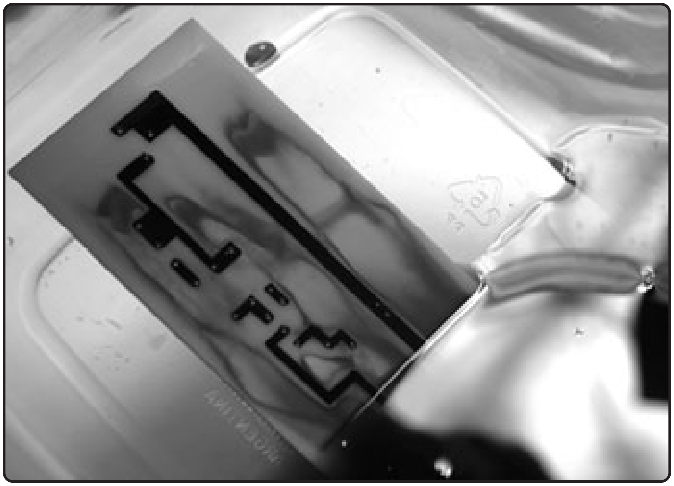

FIGURA 2. **Una** máscara temporal de tinta **aplicada a la placa virgen. De esta forma, queda preparada para el ataque químico. El** percloruro férrico **es el elegido para realizar esta acción, al eliminar el** cobre indeseado**.**

MÚLTIPLES CAPAS

Un **PCB** puede estar formado por múltiples capas conductoras (hasta **dieciséis**). La mayoría de sus componentes son de montaje superficial y son colocados por un robot denominado **pick and place**. Para nosotros, un PCB de una o dos capas será más que suficiente.

www.FreeLibros.me

agujereadora de banco o un **mini torno**. Para finalizar con la construcción del circuito, los componentes se insertan en los orificios (si son de montaje through hole) o apoyan sobre las pistas de la tarjeta, si son de montaje superficial. Luego se sueldan con estaño, ya sea con un soldador manual o mediante una máquina de soldadura por ola (en caso de grandes volúmenes de producción). Con el circuito armado, ya estamos en condiciones de comprobar su funcionamiento.

Los orificios para colocar los componentes se realizan, a nivel industrial, con un taladro controlado por computadora

El **protoboard**

El protoboard o tarjeta de proyecto es una placa plástica con orificios metalizados y conexiones eléctricas preestablecidas que se utiliza como banco de pruebas para la realización de circuitos electrónicos sencillos. Es económico y su mayor ventaja es que no requiere de soldaduras para interconectar los componentes, los cuales son simplemente insertados en los orificios para tal fin. La disposición de sus conexiones internas hace posible el montaje temporal de cualquier circuito (**Figura 3**).

TOPOLOGÍA

El espaciado de los orificios de la tarjeta es generalmente de **2,54** mm, una medida estándar en el

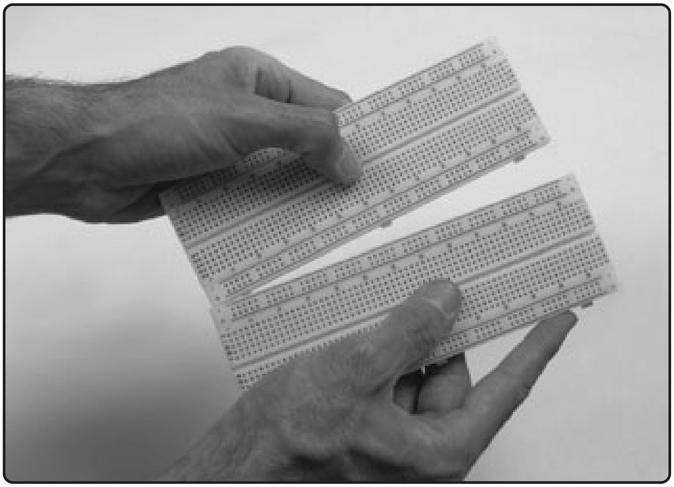

FIGURA 3. Protoboard **de tamaño estándar; son** 62 **columnas, donde cada una tiene** dos secciones **de orificios (de** A a E **y de** F a G**). Es posible unir dos o más protoboards mediante los enganches ubicados sobre uno de los bordes.**

www.FreeLibros.me

mundo de la electrónica. Podemos distinguir en el protoboard, seis secciones de orificios separadas entre sí por un material aislante. Las secciones **1** y **5**, marcadas en **rojo**, tienen continuidad horizontal y se utilizan como una de las **líneas de alimentación del circuito** (**Vcc**). En general se conectan entre sí externamente para disponer de ellas a ambos lados de la tarjeta, al igual que las secciones **2** y **6**, marcadas en azul, que constituyen la otra línea de alimentación, es decir, la **masa circuital** o **retorno de corriente** (**Gnd**).

Las secciones **3** y **4** están compuestas por columnas de cinco orificios cada una y poseen continuidad vertical, haciendo posible la formación de nodos en el circuito. Cada columna se encuentra eléctricamente aislada de las columnas adyacentes.

El **canal o surco central** del protoboard se utiliza para insertar los circuitos integrados con encapsulado tipo **DIP** (*Dual In-line Package*), cuya separación de pines es justamente la misma que la del protoboard. Este hecho los convierte en los circuitos integrados más prácticos a la hora de construir prototipos.

Es importante aclarar que para colocar el protoboard en un lugar fijo, como por ejemplo un tablero, la placa trae un **adhesivo doble faz**. Si no deseamos pegarlo en ningún lugar, le pegamos una plancha de aluminio que también viene incluida, para que no moleste la cinta doble faz.

ACCESORIOS ÚTILES

Para el armado de los circuitos en el protoboard, recomendamos algunos accesorios que nos facilitarán la tarea. Podemos realizar las conexiones entre puntos del circuito mediante un cable **unifilar**, es decir, un alambre constituido por una sola pieza, como lo son el **cable UTP** y el **multipar calibre 20** o **22**. Estos poseen el diámetro adecuado para su inserción en el protoboard. Podemos utilizar el alambre sobrante de las patas

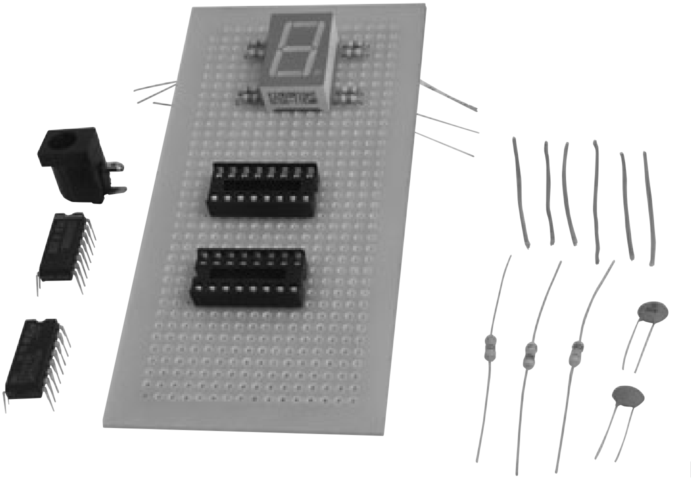

www.FreeLibros.me

de resistencias y capacitores solo para conexiones cortas, ya que no es un conductor aislado.

Debemos saber que existen componentes electrónicos que no pueden ser colocados directamente sobre el protoboard, como los potenciómetros e interruptores. En estos casos soldaremos el cable unifilar a los pines de los componentes para poder colocarlos en nuestro prototipo.

LIMITACIONES

Si bien el protoboard es una herramienta útil a la hora de construir prototipos, tiene ciertas limitaciones. En cuanto a especificaciones de potencia, el límite máximo es de **5 Watts**, lo que equivale a manejar **1 A** (Amper) en **5 V** (Volts) o **0,4 A** en **12 V**. Si sobrepasamos este límite. las conexiones internas de la tarjeta podrían dañarse y el plástico se derretirá.

El protoboard tiene una pobre funcionalidad para circuitos donde intervienen señales de radiofrecuencia debido a sus características de capacitancia: **2** a **30 pF** (pico Faradios) por punto de contacto. Por esta razón, recomendamos usarlo en aplicaciones que trabajen a frecuencias menores a **20 MHz** (Mega Hertz). El valor de capacitancia expresa la habilidad de un capacitor para almacenar carga eléctrica. La unidad de capacitancia es el Faradio.

El valor de capacitancia expresa la habilidad de un capacitor para almacenar carga eléctrica; su unidad de medida es el Faradio

Otra de sus limitaciones es que no nos permite insertar componentes de montaje superficial directamente. Para poder hacerlo, debemos comprar adaptadores, muchas veces costosos y difíciles de conseguir, o lanzarnos a la compleja tarea de construirlos nosotros mismos.

Uso **del protoboard**

Si queremos armar un circuito electrónico en el protoboard, debemos proceder ordenadamente para obtener los resultados esperados. Para ello, necesitamos contar con ciertos materiales. Primero, tenemos que disponer de un diagrama esquemático donde se encuentra el diseño del circuito. Debemos tener a mano todos los componentes electrónicos que

FALSOS CONTACTOS

Es alta la probabilidad de encontrarnos durante el armado del prototipo con falsos contactos o cables sueltos, que no se observan a simple vista. Debemos tener paciencia y ser organizados al montar un circuito en el protoboard para evitarnos posibles dolores de cabeza.

www.FreeLibros.me

forman el diseño. Necesitamos cables unifilares calibre 20 o 22 para realizar las conexiones. Además, precisamos algunas herramientas, como por ejemplo una pinza, un alicate y un cúter, que nos servirán para trabajar los puentes de cable y colocar los componentes. Por último, nos será útil el uso de un multímetro para evaluar el funcionamiento del circuito.

ALIMENTACIÓN DEL CIRCUITO

En los dos bordes de mayor longitud del protoboard se hallan las líneas o buses de alimentación. En color **rojo** tenemos la línea de tensión de alimentación (Vcc), y en **azul**, la de masa de circuito (Gnd). Es útil muchas veces hacer un puente entre ambos Vcc y otro puente entre ambos Gnd. En algunos protoboards, estas líneas están divididas a la mitad en un mismo extremo y es conveniente conectarlas también. De esta manera tendremos energía a ambos lados y a lo largo de la tarjeta cuando conectemos los bornes de la fuente de alimentación a estas líneas.

COLOCACIÓN DE COMPONENTES

Los primeros componentes que tenemos que colocar son los circuitos integrados o chips. Recordemos que el protoboard dispone de un surco o canal central para tal fin. El chip debe quedar sobre el canal central y paralelo al mismo, con sus pines insertados en los orificios que bordean el surco. Así nos aseguramos de que no exista un cortocircuito entre los pines del circuito integrado.

Observando el chip desde arriba, vemos que tiene una pequeña muesca. A la izquierda de ella, se encuentra el pin número 1. Debemos consultar la hoja de datos del componente (buscando por Internet) para conocer la función de cada uno de sus pines y no correr el riesgo de dañar el circuito integrado por una conexión incorrecta. Además, tenemos que colocar todos los chips en la misma dirección para simplificar el montaje del circuito (**Figuras 4** y **5**).

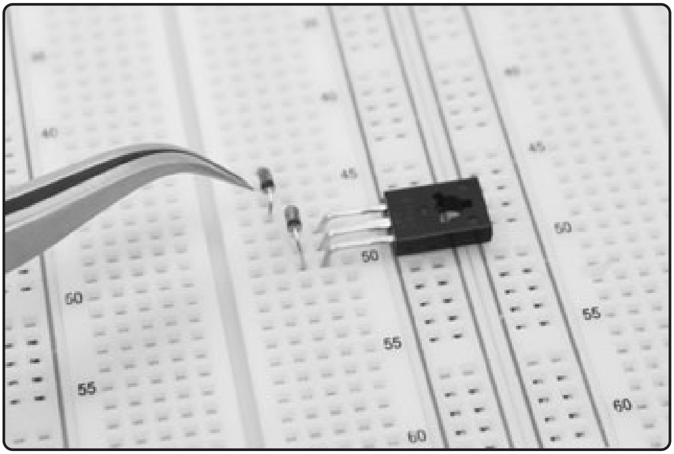

FIGURA 4. **Una de las ventajas que ofrece el protoboard es la posibilidad de insertar los cables sin necesidad de utilizar soldaduras.**

www.FreeLibros.me

En los dos bordes de mayor longitud del protoboard se hallan las líneas o buses de alimentación

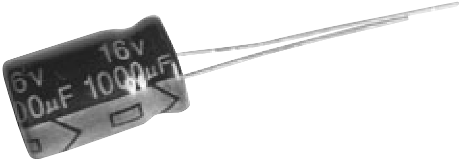

Para colocar resistencias, capacitores, diodos y transistores fácilmente, podemos usar una pinza con la que doblar sus patas. También podemos acortárselas utilizando un cúter y evitaremos así que queden demasiado elevados por encima del protoboard.

Debemos tener en cuenta que hay componentes que poseen polaridad. Esto quiere decir que tienen una pata etiquetada como [+] (positiva) y otra como [–] (negativa). Las resistencias no tienen esta característica y las podemos conectar de cualquier manera. No olvidemos considerar los rangos de operación de cada componente, es decir, las especificaciones de potencia, tensión y corriente máxima.

CONSEJOS FINALES

Daremos algunos consejos para montar un circuito electrónico en el protoboard y no fracasar en el intento. Debemos planear la distribución de los componentes en la tarjeta para que no se produzcan concentraciones de los mismos en una zona. Es preciso asegurarnos de interconectar los componentes correctamente.

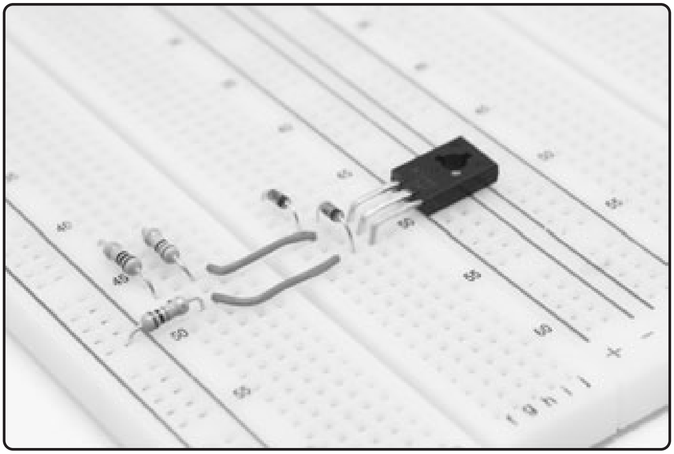

FIGURA 5.
Aquí podemos observar que tenemos energía a ambos lados de la tarjeta gracias a los puentes de cable que conectan estas líneas.

www.FreeLibros.me

Un consejo que puede parecer obvio, pero en el que queremos hacer hincapié, es que siempre tenemos que guiarnos por el esquemático para realizar el montaje. Es importante ir tachando las conexiones en el diagrama a medida que vamos realizándolas. También es conveniente que utilicemos cables de la menor longitud posible para disminuir los problemas de ruido.

En lo posible, los cabes deben estar aislados para evitar cortocircuitos con las patas de los componentes y con otros cables. Tratemos de utilizar cables de diferentes colores en zonas donde se concentran muchas conexiones. En caso de mal funcionamiento del circuito, revisemos primero las conexiones de alimentación y, luego, los falsos contactos en el resto del circuito. Si no podemos solucionar el problema, montemos el circuito en otra zona del protoboard.

En caso de mal funcionamiento del circuito, revisemos primero las conexiones de alimentación y, luego, los falsos contactos

Cómo trabajar con el protoboard

Realizaremos el montaje de un sencillo **cargador de baterías de níquel-cadmio** basado en un **transistor modelo BD140**. Este actúa como una fuente de corriente constante para entregar la carga necesaria a la batería. Para que el transistor funcione de esta manera, utilizaremos diodos de tipo **1N4148** y, con el fin de fijar las corrientes del circuito, emplearemos resistencias de **10k**, **56**, y **15 Ohms**. También incluiremos una **llave** que nos dará la posibilidad de alternar entre una carga rápida o lenta. Debemos disponer del esquemático del diseño y comprender cómo se interconectan los componentes electrónicos (**Figura 6**).

Vemos en la parte izquierda del diagrama un transformador conectado a la línea domiciliaria, un puente rectificador de diodos (**4x1N4001**), un fusible de **0,5 A** y un capacitor electrolítico de **1000 uF**. Estos componentes constituyen la fuente de alimentación del circuito (**Figura 7**). Debemos identificar cada uno de ellos. En la imagen vemos las **resistencias** de diferentes valores con su **código de colores**, el transistor, los **diodos 1N4148** y el **interruptor**. Para facilitar el montaje del circuito, utilizaremos una fuente de alimentación de banco de pruebas, como podemos ver en el **Paso a paso 1**.

www.FreeLibros.me

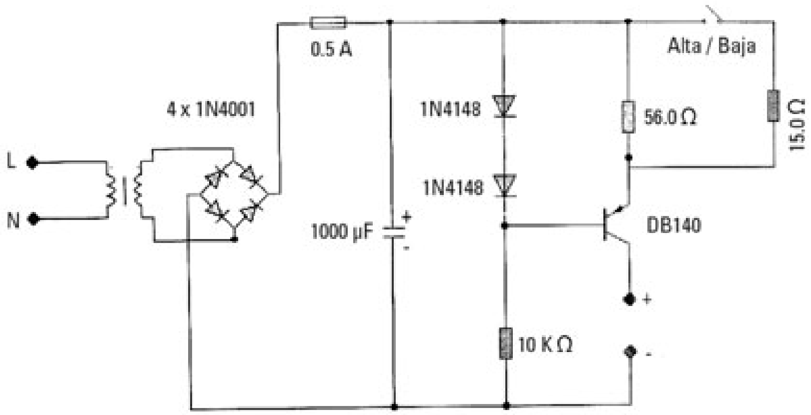
0.5 A
Alta / Baja
4 x 1N4001
1N4148
56.0 Ω
15.0 Ω
L
N
1N4148
1000 µF
+
-
DB140
+
10 K Ω
-

FIGURA 6. **Debemos recordar que, como estamos en etapa de experimentación, los reemplazaremos con el uso de una fuente de alimentación de pruebas.**

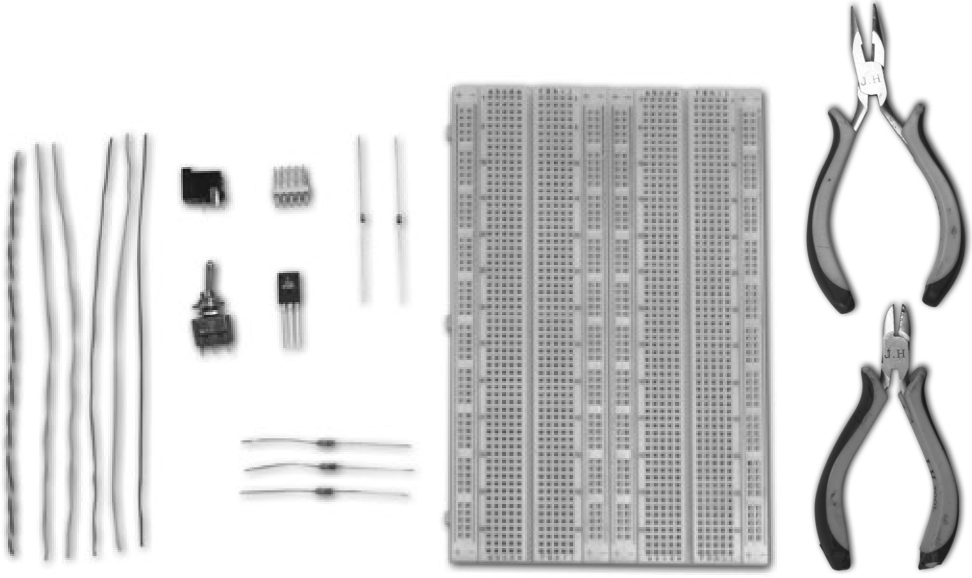

FIGURA 7. **Entre los componentes que necesitamos para trabajar, también debemos disponer de los** cables unifilares **para hacer los** puentes de conexión **y de las** pinzas para corte**.**

www.FreeLibros.me

PASO A PASO /1
Montaje del circuito

1

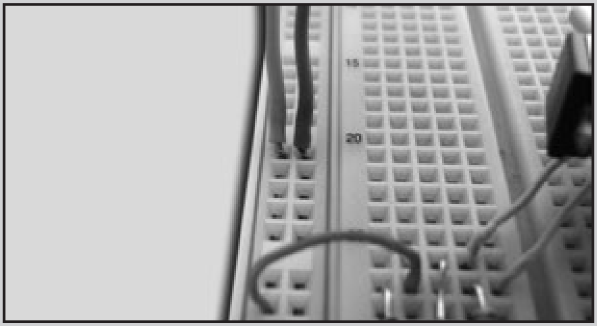

Conecte cada línea de alimentación con su opuesta mediante puentes de cable, respetando los colores: **rojo** para la **alimentación positiva** (**Vcc**) y **azul** para la **negativa** (**Gnd**). Tiene que conectar otro cable para cada línea con uno de sus extremos al aire para engancharlo con las pinzas cocodrilo de la fuente de alimentación. Para evitar cortocircuitos, **Vcc** debe quedar del lado izquierdo y **Gnd** del lado derecho, o viceversa.

2

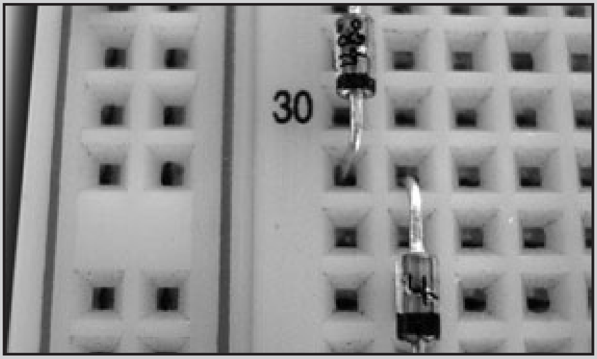

Ahora coloque los dos **diodos 1N4148**, cuyas caídas de tensión hacen que el transistor funcione como fuente de corriente constante. Estos van conectados en serie como se ve en la imagen. El **ánodo** del primer diodo va conectado a la línea de alimentación positiva. El **cátodo** se conecta a la misma columna que el ánodo del segundo diodo.

www.FreeLibros.me

PASO A PASO /1 (cont.)

3

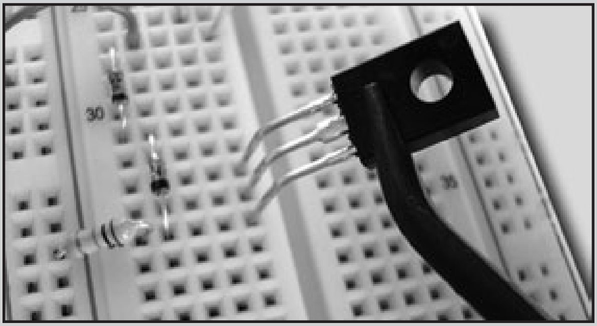

El **transistor BD140** es de tipo **PNP**, lo que significa que la corriente de emisor tiene sentido entrante, mientras que las de colector y base, sentido saliente. Observando el transistor del lado del orificio con unas **pequeñas marcas**, de izquierda a derecha, se observa: **emisor**, **colector** (**centro**) y **base**. Debe entonces conectar la base del transistor al cátodo del segundo diodo, y el emisor y el colector, a distintas columnas del protoboard.

4

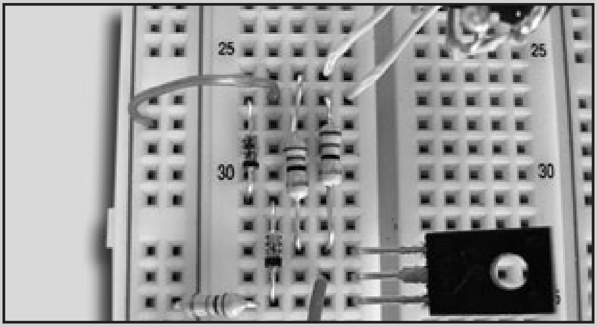

Tome la resistencia de valor **10k Ohms** (marrón, negro, naranja) y coloque una de sus patas en la misma columna que la base del transistor del paso anterior. Conecte la otra al **bus de alimentación** marcado con azul (Gnd). Inserte la resistencia de valor **56 Ohms** (verde, azul, negro) de manera que uno de sus extremos quede conectado al bus de alimentación positivo (Vcc) y, el otro, a la columna donde se encuentra el emisor del transistor.

www.FreeLibros.me

PASO A PASO /1 (cont.)

5

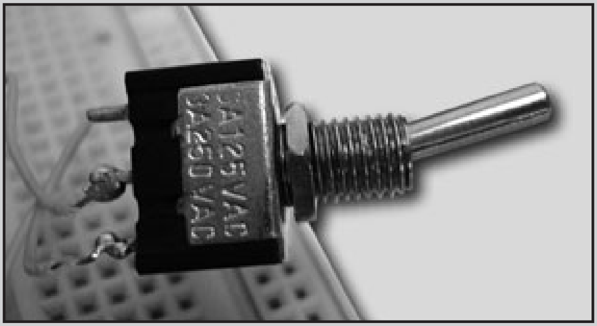

Para colocar la llave de selección debe soldarle un cable en cada uno de los extremos. Uno de los cables debe ir conectado a Vcc, y el otro, a una columna libre del protoboard. Coloque entonces la última resistencia, la de **15 Ohms**. Inserte una de sus patas en la columna donde conectó el extremo de la llave que no está unido a Vcc, y la otra, en la columna donde se encuentra el emisor del transistor.

6

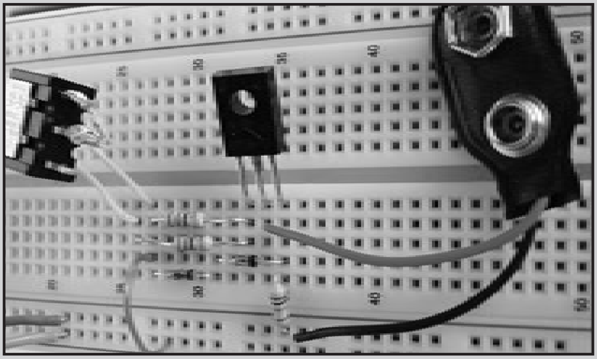

Ahora tiene que conectar el porta batería de manera que el extremo positivo quede unido al **colector del transistor** (pata central), y el extremo negativo, a la **masa circuital Gnd** (la línea de alimentación azul). Luego coloque la batería en el porta batería. Una vez que realice esta acción, el circuito está listo para ser alimentado.

www.FreeLibros.me

ELECTR_Cap2_xxx_xxx_ajustado.qxp 04/07/2011 16:32 PÆgina 41

Circuito impreso **universal**

El circuito impreso universal o **UPCB** (*Universal Printed Circuit Board*) es una tarjeta de uso general para construir prototipos electrónicos permanentes utilizando soldadura (**Figura 8**, **9**, **10 y 11**).

Entonces, no necesitaremos recurrir al diseño y fabricación de un circuito impreso específico, con lo cual nos ahorraremos un tiempo considerable. Al igual que en el protoboard, el espaciado de los orificios es de **2,54 mm** (0,1 pulgadas). Su precio es accesible, y viene en diferentes tamaños, configuraciones y calidades.

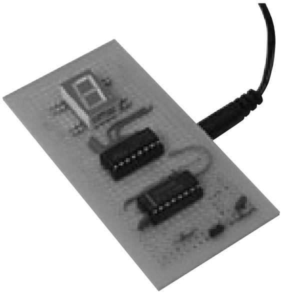

PERFBOARD

El circuito impreso universal más básico es el **perfboard,** una placa compuesta de **fenólico** o **epoxi**. Consta de varias columnas de orificios metalizados aislados eléctricamente. Los componentes se insertan en estos agujeros, pero para hacer las conexiones debemos utilizar puentes de cable o estaño. Este tipo de placa se utiliza para la realización de proyectos no profesionales y los diseños resultantes no son muy prolijos. No es recomendable si queremos hacer el mismo circuito más de una vez.

STRIPBOARD

Una versión más avanzada del perfboard es el **stripboard**, también llamado **veroboard**, la marca comercial más popular. En este caso las columnas de orificios (tiras) están conectadas eléctricamente.

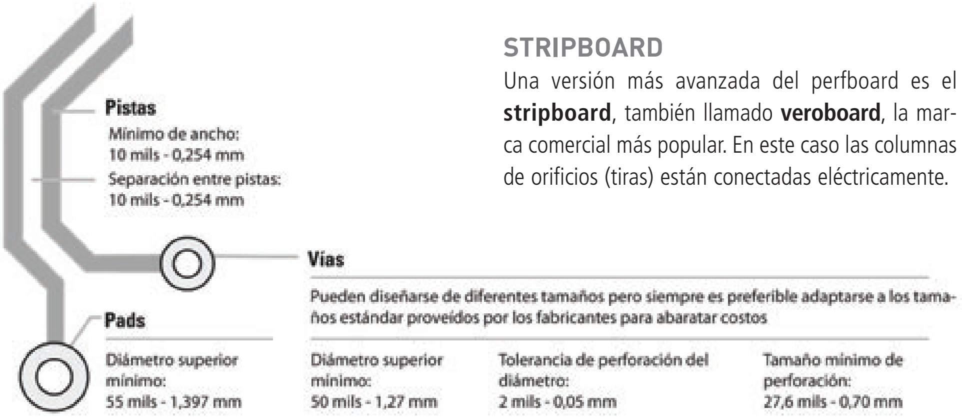
Pistas
Mínimo de ancho:
10 mils - 0,254 mm
Separación entre pistas:
10 mils - 0,254 mm
Vías
Pueden diseñarse de diferentes tamaños pero siempre es preferible adaptarse a los tamaños estándar proveídos por los fabricantes para abaratar costos
Pads
Diámetro superior mínimo:
55 mils - 1,397 mm
Diámetro superior mínimo:
50 mils - 1,27 mm
Tolerancia de perforación del diámetro:
2 mils - 0,05 mm
Tamaño mínimo de perforación:
27,6 mils - 0,70 mm

FIGURA 8. **Vemos el ejemplo de especificaciones de un fabricante. Existe la posibilidad de trabajar con pistas más finas y vías más pequeñas.**

www.FreeLibros.me

ELECTR_Cap2_xxx_xxx_ajustado.qxp 04/07/2011 16:32 PÆgina 42

INFOGRAFÍA 2: EL CIRCUITO IMPRESO (PCB)

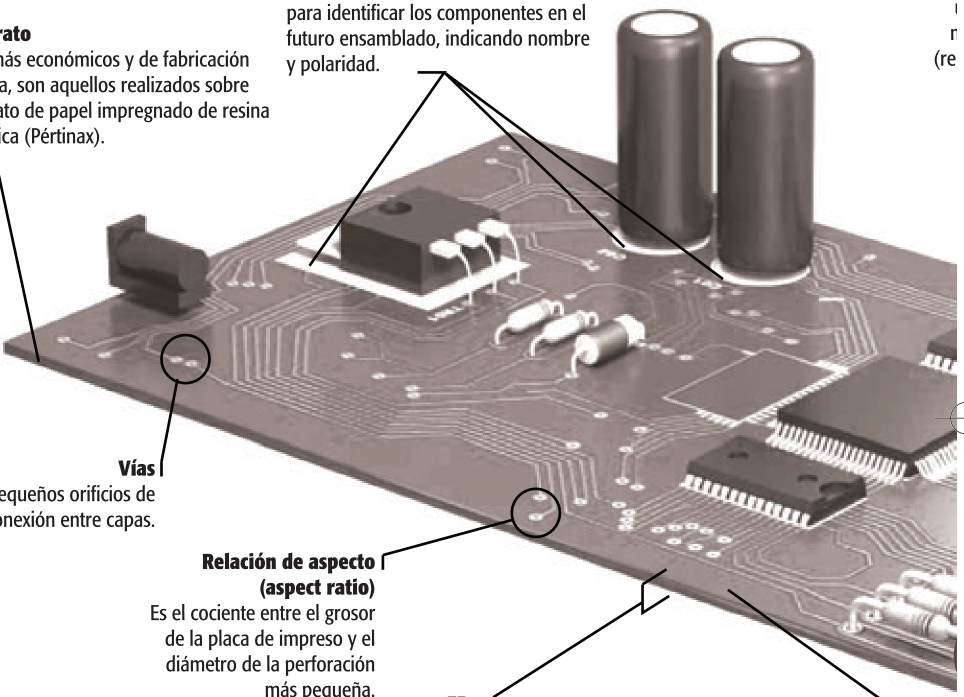

Silkscreen
Grabado de colores (generalmente blanco) sobre la superficie de un impreso, utilizado para identificar los componentes en el futuro ensamblado, indicando nombre y polaridad.

de
tes.
(re

Sustrato
Los más económicos y de fabricación casera, son aquellos realizados sobre sustrato de papel impregnado de resina fenólica (Pértinax).

Vías
Son pequeños orificios de conexión entre capas.

Relación de aspecto (aspect ratio)
Es el cociente entre el grosor de la placa de impreso y el diámetro de la perforación más pequeña.

FR-4
Existen también los llamados impresos "doble capa" –double layer-. Son generalmente de sustrato de fibra de vidrio del tipo FR-4 ("Flame retardant" : retardante de llama de factor 4) .

www.FreeLibros.me

Pad
Área de material conductor en un PCB designado para el montaje de componentes. Puede ser una patita de un footprint de un componente SMD o una perforación metalizada para montar un componente (resistencia, capacitor, etc.) ó un integrado through-hole.

Corona (annular ring)
Es el anillo circular remanente luego de la perforación de un pad. Es de sustrato conductivo y superficie metalizado.

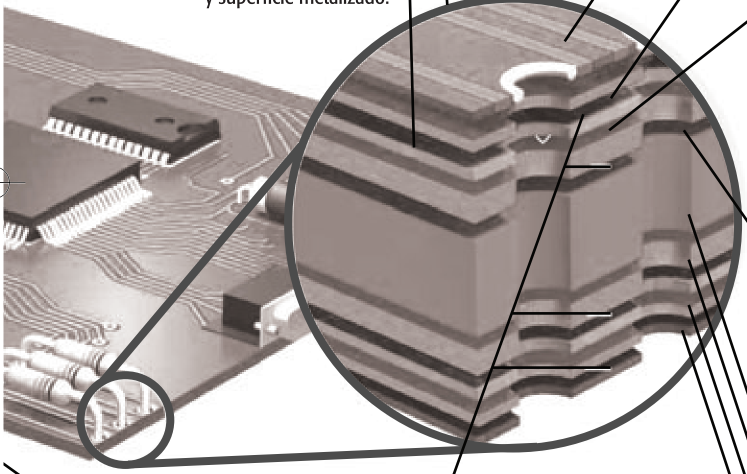

Foots prints
Huellas de montaje para componentes de montaje superficial (SMT)

Signal layer
Generalmente esta capa externa se usa para ruteado general.

Prepreg
Es una delgada capa de fibra de vidrio impregnada de resina epoxi, aplicada entre 2 capas internas o “cores” de un PCB multilayer. Funciona como pegamento en el proceso de “laminación” del fabricante.

Internal layer
Capa interna de un PCB multicapa. En impresos de 4 capas se las destina a planos de tierra o alimentación, aunque puede ser usada para ruteado general.

Core

Internal layer

Prepreg

Signal layer

Lado de componente (component side)
El lado o cara de un circuito en donde se montarán la mayoría de los componentes.

Four-Layer PCB (PCB 4 capas):
Es un impreso que cuenta con 4 capas de interconexión. Generalmente se utilizan 2 para ruteado general (las 2 externas) y 2 para planos de tierra y alimentación respectivamente (las internas).

www.FreeLibros.me

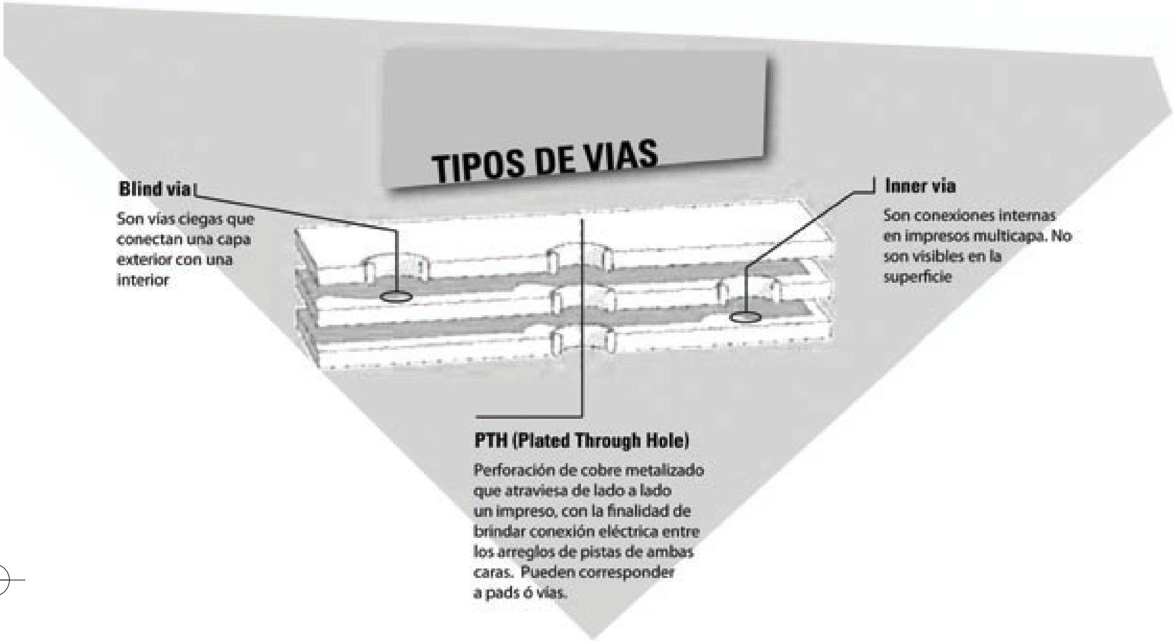

Blind via

Son vías ciegas que conectan una capa exterior con una interior

Inner via

Son conexiones internas en impresos multicapa. No son visibles en la superficie

PTH (Plated Through Hole)

Perforación de cobre metalizado que atraviesa de lado a lado un impreso, con la finalidad de brindar conexión eléctrica entre los arreglos de pistas de ambas caras. Pueden corresponder a pads ó vías.

FIGURA 9. El PCB está formado por una capa de cobre montada sobre un sustrato aislante.

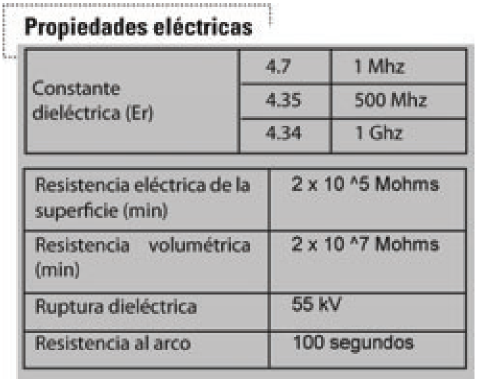

Propiedades eléctricas

Constante dieléctrica (Er)	4.7	1 Mhz
	4.35	500 Mhz
	4.34	1 Ghz

Resistencia eléctrica de la superficie (min)	2 x 10 ^5 Mohms
Resistencia volumétrica (min)	2 x 10 ^7 Mohms
Ruptura dieléctrica	55 kV
Resistencia al arco	100 segundos

FIGURA 10. En esta tabla podemos analizar sus propiedades eléctricas.

Es importante que diseñemos la distribución y el trazado antes de montar el circuito. Una vez hecho esto, debemos montar los circuitos integrados para que queden perpendiculares a las columnas de conducción. Luego, cortamos la conducción de las tiras de orificios de manera que los pines enfrentados del integrado no se encuentren unidos. Habiendo tomado esta precaución, aprovechamos las tiras de orificios para conectar el resto de los componentes. Al igual que en el protoboard, los puentes de cable también serán útiles.

Es mucho más fácil hacer una réplica del circuito en el stripboard que en el perfboard. La única

www.FreeLibros.me

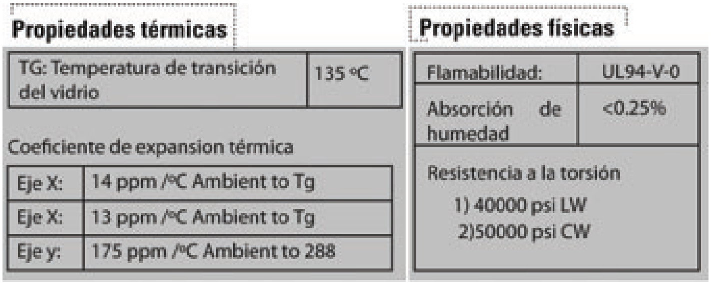

Propiedades térmicas

TG: Temperatura de transición del vidrio	135 ºC

Coeficiente de expansion térmica

Eje X:	14 ppm /ºC Ambient to Tg
Eje X:	13 ppm /ºC Ambient to Tg
Eje y:	175 ppm /ºC Ambient to 288

Propiedades físicas

Flamabilidad:	UL94-V-0
Absorción de humedad	<0.25%
Resistencia a la torsión 1) 40000 psi LW 2)50000 psi CW	

FIGURA 11.
El circuito impreso es una placa especialmente diseñada para el montaje de los componentes.

desventaja es que es menos modificable, ya que cortamos algunas columnas de orificios para adecuarlas a nuestro diseño.

UPCB

Es el circuito impreso universal propiamente dicho, lo más parecido a un PCB. Cuenta con el surco central necesario para montar circuitos integrados de encapsulado DIP (*Dual In Package*). También incluye las líneas o buses en los bordes para alimentar el circuito electrónico. Permite una gran variedad de conexiones, ya que también podemos utilizar puentes de cables en él. Es la opción de mejor calidad, pero también la más costosa. Posee una máscara que recubre las áreas que no necesitan de soldadura y que protege contra la oxidación de la placa y los posibles cortocircuitos. Si operamos con cuidado, lograremos una muy prolija presentación del circuito utilizando el UPCB.

MONTAJE DEL CIRCUITO

La construcción del diseño en el circuito impreso universal depende en gran medida del formato elegido. Debemos familiarizarnos profundamente con la forma y la estructura de la configuración de la placa que vamos a utilizar. Necesitamos disponer de las mismas herramientas que en el caso del protoboard, y sumar el soldador para electrónica (o cautín) y el estaño.

Partiendo de un esquema circuital, será conveniente dibujar primero la distribución de los componentes y su conexión en un papel, para después lanzarnos a la tarea de colocarlos y soldarlos a la placa. Recordemos que al no tratarse de un protoboard, cada vez que decidamos quitar un componente soldado, el cobre de las pistas se verá debilitado, y podrá despegarse de la tarjeta si es que realizamos el desmontaje sucesivas veces. Si al finalizar la construcción del circuito dejamos una parte de la tarjeta sin utilizar, podemos cortarla con un cúter o sierra. De esta forma minimizaremos el tamaño del prototipo y aprovecharemos la placa sobrante para armar otro circuito. Si deseamos producir en serie el diseño probado en el circuito impreso universal, deberemos realizar un PCB específico.

El UPCB posee una máscara que protege a la placa de la oxidación y de los cortocircuitos

www.FreeLibros.me

Circuito impreso **en detalle**

Una vez hecho el diseño del circuito impreso, es necesaria la construcción del mismo. Esta es una tarea que requiere ciertos cuidados, ya que una plaqueta mal fabricada puede poner en peligro el funcionamiento de nuestro proyecto.

MÉTODO DE FABRICACIÓN CASERA

Para impresos que no tengan un alto grado de complejidad, podemos optar por la construcción casera. Se parte de una plaqueta virgen, la cual consta de una base aislante llamada sustrato, sobre la cual se encuentra adherida una fina placa de cobre. El objetivo es obtener nuestro circuito sobre dicha placa, eliminando el cobre sobrante. Para lograr esto tenemos que dibujar a mano sobre la capa de cobre el circuito a obtener, utilizando un marcador indeleble. Luego, para eliminar el cobre sobrante se sumerge la plaqueta en una solución de percloruro férrico.

FABRICACIÓN PROFESIONAL

Para plaquetas con alto grado de dificultad o para producción a gran escala, la construcción de circuitos impresos en forma profesional es la más indicada. De esta manera podemos obtener en forma automatizada circuitos multicapas, los cuales poseen varias capas conductoras. El proceso de fabricación de estos impresos comienza por las capas internas, las cuales poseen láminas de cobre en ambos lados.
Se aplica en estas un film fotosensible para que, mediante fotoexposición y posterior revelado, el diseño del circuito impreso quede grabado sobre las láminas de cobre. Posteriormente se elimina el cobre sobrante, es decir, el que no se encuentra grabado, y los restos de film.

Luego, se realiza una inspección óptica automática, en la cual se compara el impreso obtenido con el diseño original para detectar diferencias. Las placas que no pasan el test son descartadas. Pasado el test, se agregan las capas siguientes mediante prensado y se vuelve a repetir el proceso. La interconexión entre los circuitos correspondientes a cada capa del impreso se realiza mediante orificios llamados vías.

La construcción casera solo es viable para impresos simples, sin demasiada complejidad

OTRO MÉTODO DE FABRICACIÓN

Otro método de fabricación es realizar el diseño en una PC e imprimirlo en un papel vegetal con una impresora láser. Ese dibujo se pasará a la plaqueta virgen por medio del planchado del papel impreso sobre la capa de cobre.

www.FreeLibros.me

CONSTRUCCIÓN DE UN CIRCUITO IMPRESO

En este apartado aprenderemos paso a paso cómo realizar la construcción de un circuito impreso correspondiente a un **cargador de baterías de Ni-Cd**. Los materiales que necesitamos son: **plaqueta virgen de pertinax** o **epoxi** (de **10 cm** x **10 cm**), **percloruro férrico**, **lana de acero**, **guantes de látex, alcohol fino**, **marcador indeleble**, **multímetro**, **alicate**, **agujereadora** y **mecha de acero rápido** de **1 mm**. Con respecto a los componentes electrónicos, necesitamos: **2** diodos **1N4148**, una resistencia de **10 KΩ**, una resistencia de **56Ω**, una resistencia de **15Ω**, un transistor **BD140** y una llave simple. Para soldar los componentes a la plaqueta, necesitaremos un soldador de **30 W** y **estaño** (**Paso a paso 2**).

Soldadura
para electrónica

El soldador es una de las herramientas básicas de todo electrónico profesional o hobbista. Con él podremos realizar las uniones entre los componentes electrónicos y las pistas de cobre de los circuitos impresos. Para estos casos, los soldadores empleados son de baja potencia, alrededor de **30 W**, para evitar el deterioro de los componentes electrónicos en el proceso de soldado.

El soldador está formado por una resistencia eléctrica en su interior, una **punta de cobre** y un **mango aislante**. Al conectar el cable de alimentación, la resistencia eléctrica se calienta, transmitiendo ese calor a la punta de cobre. Dado que se trata de una herramienta que alcanza temperaturas elevadas, es conveniente el empleo de un soporte adecuado para el mismo. Existen diferentes tipos de soldadores, pero los que se utilizan en electrónica son los llamados “lápiz”, nombre que se les da debido a su forma. La potencia estos instrumentos es de alrededor de 30 W.

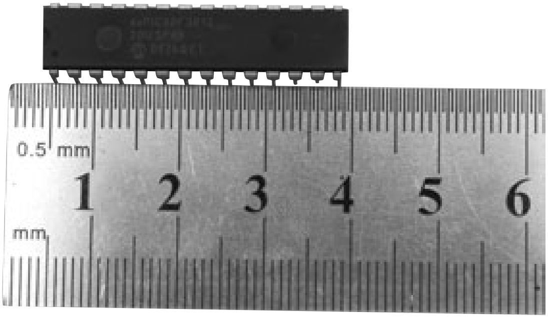

TIPOS DE SOLDADORES

Existen diferentes clases de soldadores. Los más comunes son los de tipo lápiz y pistola. El segundo se calienta por medio de una corriente que pasa por él. Es útil para trabajos esporádicos porque se calienta instántaneamente.

www.FreeLibros.me

PASO A PASO /2
Construcción de circuito impreso

1

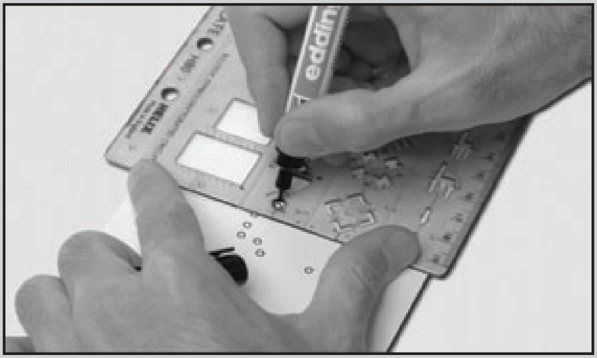

Lo primero que debe hacer es diseñar el circuito que quiere desarrollar. Para ello tiene que utilizar un marcador indeleble y una plantilla para diseño electrónico (*electronic template*). Simplemente, debe trazar las líneas que corresponden a las pistas y los orificios, que es donde colocará los componentes.

2

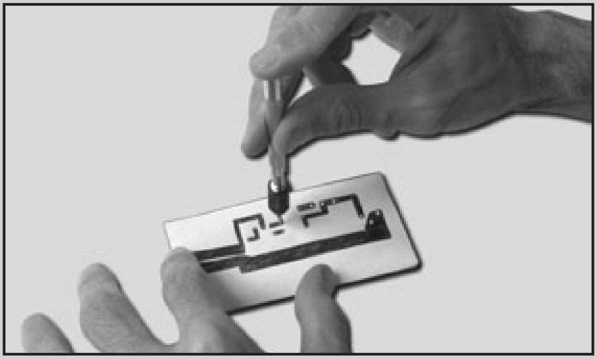

Superponga el diagrama del circuito sobre la plaqueta para realizar las perforaciones donde posteriormente se soldarán los componentes electrónicos. Utilice una perforadora con mecha de **1 mm**. Es una tarea delicada, ya que corre el riesgo de que se levante alguna isla de la plaqueta. Las islas son las áreas de cobre donde van soldados los componentes y donde se debe perforar.

www.FreeLibros.me

PASO A PASO /2 (cont.)

3

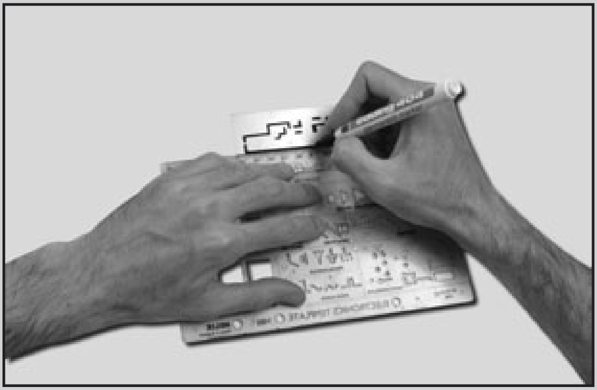

Una vez que tiene los orificios creados, copie el diseño del circuito (**PCB**) de la figura sobre la capa de cobre. Esto se hace con el marcador indeleble. Recuerde que todo lo que esté cubierto con la tinta del marcador indeleble no será atacado por la solución de percloruro férrico.

4

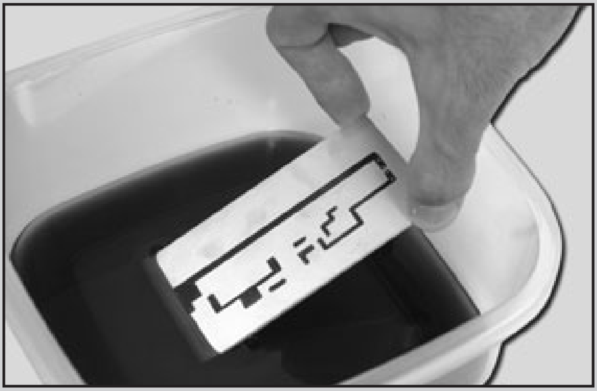

Vierta **percloruro férrico** en un recipiente plástico o de vidrio y sumerja la plaqueta en él durante **15 minutos** aproximadamente. Pasado el tiempo, retire la plaqueta y verifique que no queden sectores de cobre más que los correspondientes al circuito.

www.FreeLibros.me

PASO A PASO /2 (cont.)

5

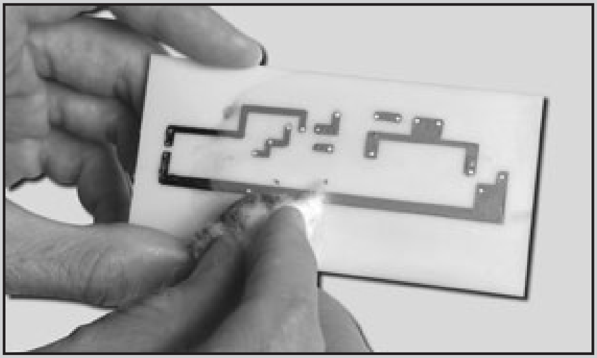

Retire la plaqueta del recipiente y lávela varios minutos con agua. Recuerde que el **percloruro férrico** es un ácido, con lo cual deberá tener mucho cuidado al manipularlo. Seque la plaqueta y retire la tinta del marcador indeleble con una lana de acero. Luego, limpie la superficie con papel tissue humedecido en alcohol fino.

6

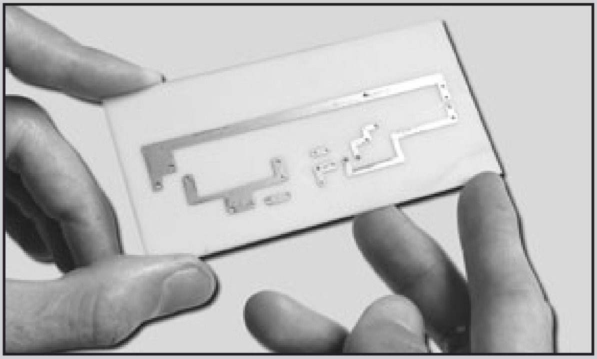

Una vez que tiene la plaqueta lista, debe verificar que las pistas hayan quedado correctamente formadas, del mismo modo que las islas. Para una revisión más exhaustiva utilice un multímetro en la función de continuidad o midiendo resistencia. En el caso de que exista un corto entre dos pistas, será necesario eliminar el cobre excedente con un cúter o herramienta similar.

www.FreeLibros.me

PASO A PASO /2 (cont.)

7

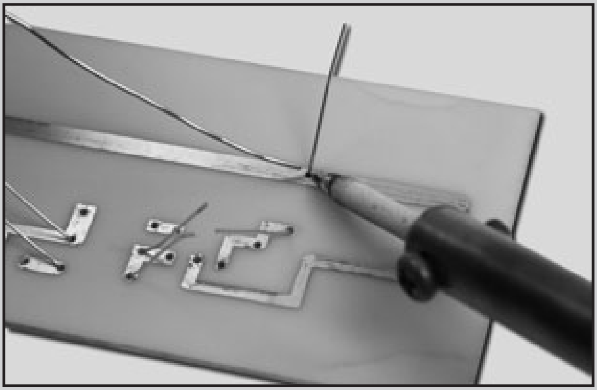

Con la plaqueta en condiciones, coloque los componentes más pequeños, como las resistencias y los diodos. Luego dé vuelta la plaqueta y aplique la soldadura. Siempre es recomendable que el soldador tenga una buena temperatura para trabajar mejor con la aleación de estaño.

8

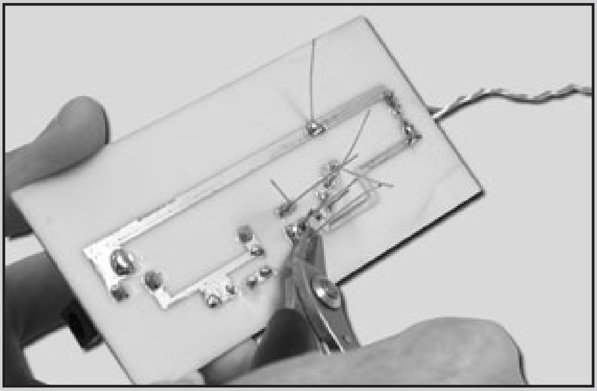

Luego, siga con los componentes más grandes, como los capacitores, transistores, integrados y llaves. Esta práctica facilitará el proceso de soldado. Finalmente, debe cortar con el alicate el alambre excedente de los terminales de los componentes.

www.FreeLibros.me

MATERIAL PARA SOLDAR

El estaño es el elemento que se utiliza para realizar la soldadura. En realidad, no se trata de estaño solamente, sino de una aleación de **60% de estaño** y **40% de plomo**. Además de estos dos materiales, posee una resina, la cual permite realizar una buena soldadura, protegiendo a las superficies de la temperatura del soldador. Físicamente es un alambre de contextura blanda. Se vende en carretes y existen distintos espesores, como de **0,5 mm**, **1 mm**, etc., dependiendo del tamaño del área a soldar.

PROCESO DE SOLDADO

El proceso de soldado consiste en unir, tanto mecánica como electrónicamente, un componente electrónico con la correspondiente pista de cobre, empleando un soldador y estaño.
Una vez que el soldador alcanzó la temperatura de operación, se deberá tener la precaución de que la punta del instrumento esté limpia. Para ello, se puede emplear un cepillo o un trozo de tela de jean. Nunca se deben usar materiales que rayen la punta. Las partes que se vayan a soldar deben estar limpias y libres de impurezas. Una buena práctica es estañar previamente las partes a soldar.

Los soldadores empleados son de baja potencia y manejan cerca de 30 Watts

Ahora sí estamos listos para soldar. Para hacerlo, tomamos con una mano el soldador y, con otra, un trozo de estaño. Calentamos entre 2 y 3 segundos el área a soldar y sin quitar el soldador aplicamos el alambre de estaño en la zona. Rápidamente quitamos el estaño y por último el soldador, con el fin de permitir que se enfríe la soldadura. No debemos forzar el enfriamiento, ya que esto puede derivar en una mala soldadura.

Dependiendo del tamaño del área a soldar, variará la cantidad de estaño a utilizar. Si aplicamos poca cantidad, si bien conducirá la electricidad, la soldadura no va a tener una adecuada resistencia mecánica, con lo cual, puede desoldarse fácilmente. Si aplicamos estaño en exceso, corremos el riesgo de poner en corto las pistas del circuito impreso al desbordarse el estaño. En general, con unos pocos milímetros es suficiente (**Figura 12, 13, 14 y 15**).

FIGURA 12.
Soldar es sencillo, pero requiere de experiencia. Es importante cumplir con las buenas prácticas, como mantener limpia la punta del soldador, aplicar el calor adecuado y usar la correcta cantidad de estaño.

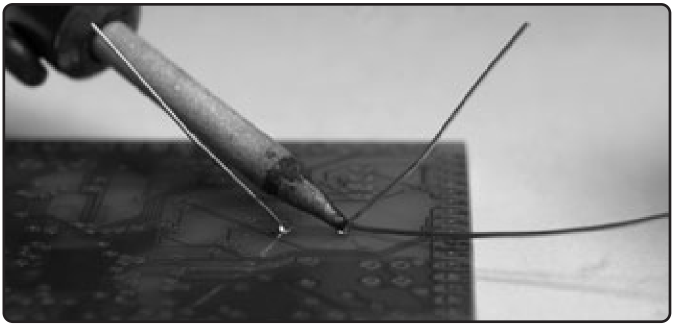

www.FreeLibros.me

Se debe aplicar el calor justo y necesario, ya que si nos excedemos, es probable que se dañen los componentes o que se levanten las pistas del circuito impreso. Por otro lado, si aplicamos poco calor, se puede realizar una soldadura "fría", la cual a simple vista es opaca, y tendrá poca resistencia mecánica y baja conductividad eléctrica.

DESOLDADO

Existen varios métodos para desoldar componentes electrónicos. Uno de ellos es mediante el uso de un **desoldador** (**chupador de estaño**). Este dispositivo es básicamente una bomba de succión, formada en general por un cilindro de aluminio el cual tiene en su interior un émbolo accionado por un resorte.

Posee una punta por la cual se succiona el estaño y un pulsador que libera el resorte que desplaza el émbolo y produce la succión.

El procedimiento de uso es bastante simple. Primero se prepara el desoldador, accionando el émbolo. Luego se apoya la punta del desoldador sobre el estaño que necesita ser quitado. Seguidamente, se apoya la punta del soldador sobre el área a ser desoldada. (**Figura 16**)

Luego de unos pocos segundos, cuando se funde el estaño, se acciona el pulsador del desoldador, para producir la succión del estaño. Este proceso se deberá repetir si aún quedan restos de estaño.

En este capítulo, aprendimos cómo realizar circuitos impresos y sus características. Además hicimos las primeras conexiones en el protoboard, valiéndonos de las ventajas que aporta trabajar sobre este. También, analizamos todo sobre el proceso de soldado.

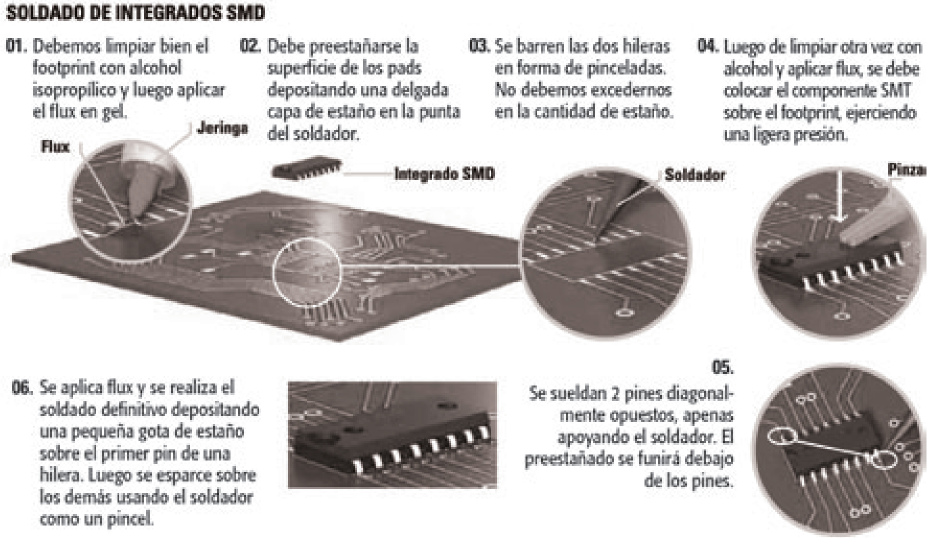

SOLDADO DE INTEGRADOS SMD

01. Debemos limpiar bien el footprint con alcohol isopropílico y luego aplicar el flux en gel.

02. Debe preestañarse la superficie de los pads depositando una delgada capa de estaño en la punta del soldador.

03. Se barren las dos hileras en forma de pinceladas. No debemos excedernos en la cantidad de estaño.

04. Luego de limpiar otra vez con alcohol y aplicar flux, se debe colocar el componente SMT sobre el footprint, ejerciendo una ligera presión.

05. Se sueldan 2 pines diagonalmente opuestos, apenas apoyando el soldador. El preestañado se funirá debajo de los pines.

06. Se aplica flux y se realiza el soldado definitivo depositando una pequeña gota de estaño sobre el primer pin de una hilera. Luego se esparce sobre los demás usando el soldador como un pincel.

FIGURA 13. En esta infografía, encontramos las diferentes técnicas de desoldado.

www.FreeLibros.me

ELECTR_Cap2_xxx_xxx_ajustado.qxp 04/07/2011 16:32 PÆgina 54

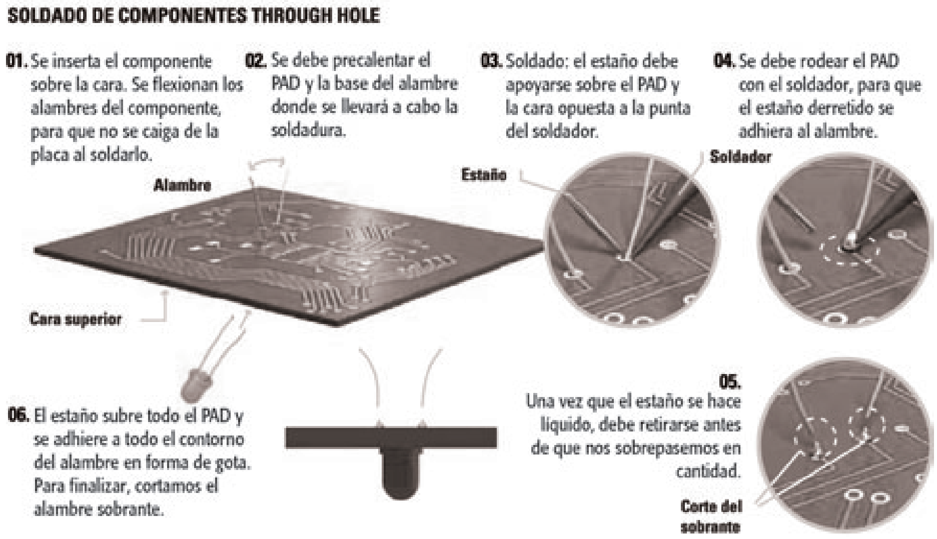

SOLDADO DE COMPONENTES THROUGH HOLE

01. Se inserta el componente sobre la cara. Se flexionan los alambres del componente, para que no se caiga de la placa al soldarlo.

02. Se debe precalentar el PAD y la base del alambre donde se llevará a cabo la soldadura.

03. Soldado: el estaño debe apoyarse sobre el PAD y la cara opuesta a la punta del soldador.

04. Se debe rodear el PAD con el soldador, para que el estaño derretido se adhiera al alambre.

05. Una vez que el estaño se hace líquido, debe retirarse antes de que nos sobrepasemos en cantidad.

06. El estaño subre todo el PAD y se adhiere a todo el contorno del alambre en forma de gota. Para finalizar, cortamos el alambre sobrante.

FIGURA 14. **En estos pasos, analizamos cómo realizar el soldado de componentes through hole.**

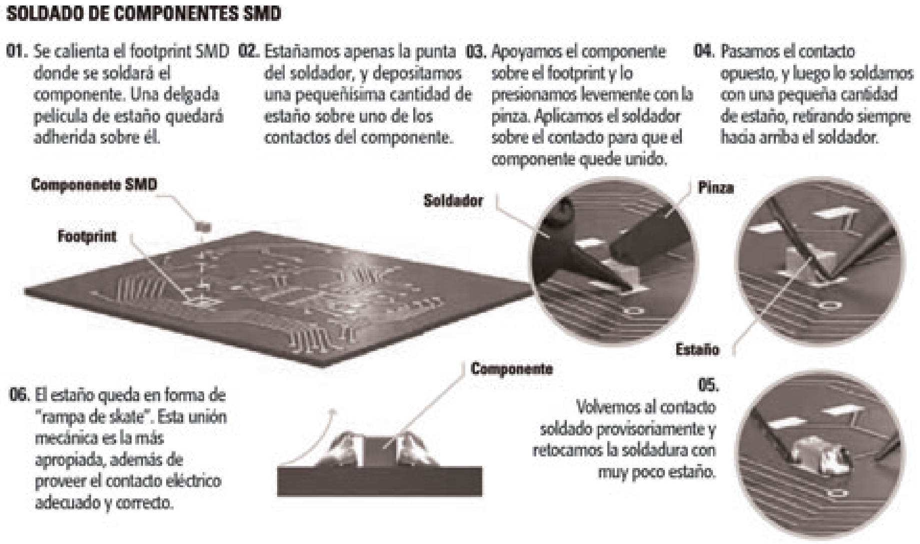

SOLDADO DE COMPONENTES SMD

01. Se calienta el footprint SMD donde se soldará el componente. Una delgada película de estaño quedará adherida sobre él.

02. Estañamos apenas la punta del soldador, y depositamos una pequeñísima cantidad de estaño sobre uno de los contactos del componente.

03. Apoyamos el componente sobre el footprint y lo presionamos levemente con la pinza. Aplicamos el soldador sobre el contacto para que el componente quede unido.

04. Pasamos el contacto opuesto, y luego lo soldamos con una pequeña cantidad de estaño, retirando siempre hacia arriba el soldador.

05. Volvemos al contacto soldado provisoriamente y retocamos la soldadura con muy poco estaño.

06. El estaño queda en forma de "rampa de skate". Esta unión mecánica es la más apropiada, además de proveer el contacto eléctrico adecuado y correcto.

FIGURA 15. **Al seguir este procedimiento, aprenderemos a soldar componentes SMD.**

www.FreeLibros.me

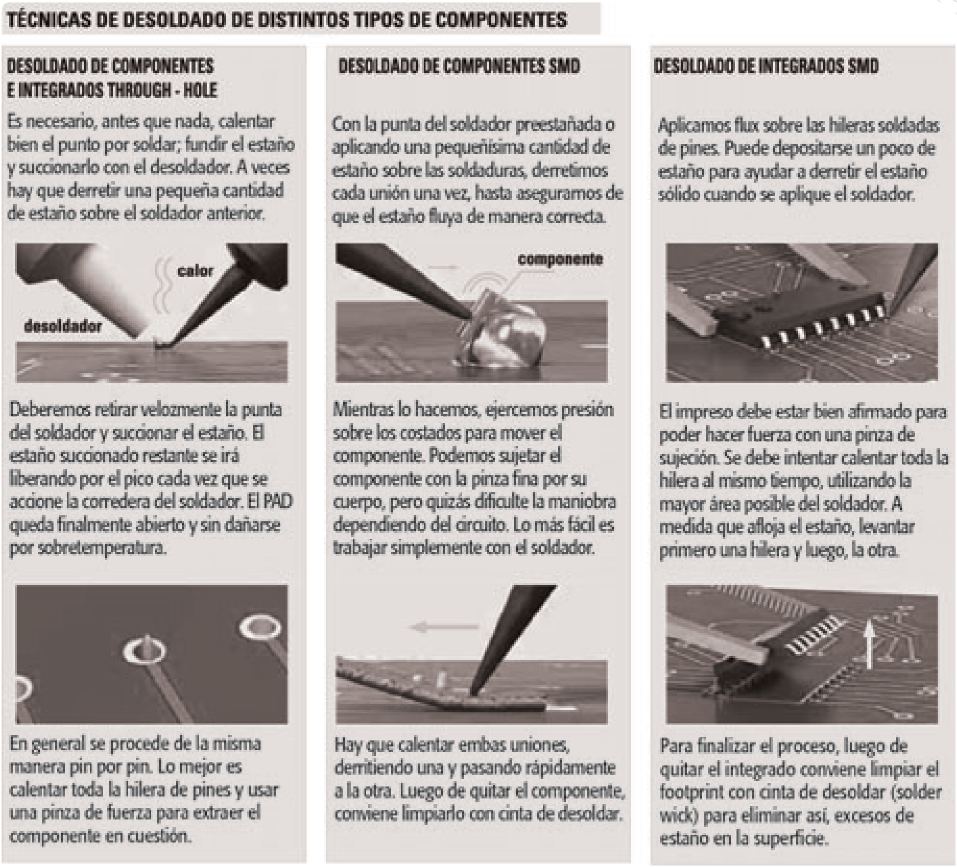

TÉCNICAS DE DESOLDADO DE DISTINTOS TIPOS DE COMPONENTES

DESOLDADO DE COMPONENTES E INTEGRADOS THROUGH - HOLE

Es necesario, antes que nada, calentar bien el punto por soldar; fundir el estaño y succionarlo con el desoldador. A veces hay que derretir una pequeña cantidad de estaño sobre el soldador anterior.

Deberemos retirar velozmente la punta del soldador y succionar el estaño. El estaño succionado restante se irá liberando por el pico cada vez que se accione la corredera del soldador. El PAD queda finalmente abierto y sin dañarse por sobretemperatura.

En general se procede de la misma manera pin por pin. Lo mejor es calentar toda la hilera de pines y usar una pinza de fuerza para extraer el componente en cuestión.

DESOLDADO DE COMPONENTES SMD

Con la punta del soldador preestañada o aplicando una pequeñísima cantidad de estaño sobre las soldaduras, derretimos cada unión una vez, hasta asegurarnos de que el estaño fluya de manera correcta.

Mientras lo hacemos, ejercemos presión sobre los costados para mover el componente. Podemos sujetar el componente con la pinza fina por su cuerpo, pero quizás dificulte la maniobra dependiendo del circuito. Lo más fácil es trabajar simplemente con el soldador.

Hay que calentar embas uniones, derritiendo una y pasando rápidamente a la otra. Luego de quitar el componente, conviene limpiarlo con cinta de desoldar.

DESOLDADO DE INTEGRADOS SMD

Aplicamos flux sobre las hileras soldadas de pines. Puede depositarse un poco de estaño para ayudar a derretir el estaño sólido cuando se aplique el soldador.

El impreso debe estar bien afirmado para poder hacer fuerza con una pinza de sujeción. Se debe intentar calentar toda la hilera al mismo tiempo, utilizando la mayor área posible del soldador. A medida que afloja el estaño, levantar primero una hilera y luego, la otra.

Para finalizar el proceso, luego de quitar el integrado conviene limpiar el footprint con cinta de desoldar (solder wick) para eliminar así, excesos de estaño en la superficie.

FIGURA 16. **Vemos de qué manera realizar el soldado de integrados SMD.**

HISTORIA DEL PROCESO DEL SOLDADO

El primer proceso de soldadura fue la **soldadura de fragua**, que predominó hasta el siglo XIX. Se desarrollaron luego la **soldadura por arco** y la **soldadura a gas**, seguidas por la **soldadura por resistencia**. A principios del siglo XX, las técnicas de soldado se multiplicaron.

www.FreeLibros.me

Multiple choice

1 **¿Cuáles de las siguientes vías son conexiones internas en impresos multicapa?**

a- Blind vía.
b- Inner via.
c- Foot via.
d- Core via.

2 **¿Cuáles de las siguientes son vías ciegas que conectan una capa exterior con una interior?**

a- Blind vía.
b- Inner via.
c- Foot via.
d- Core via.

3 **¿Cómo se llama el área de material conductor en un PCB designado para el montaje de componentes?**

a- Internal layer.
b- Four Layer PCB.
c- Corona.
d- Pad.

4 **¿Cómo se llama la capa interna de un PCB multicapa?**

a- Internal layer.
b- Four Layer PCB.
c- Corona.
d- Pad.

5 **¿De qué manera se denomina el impreso que cuenta con cuatro capas de interconexión?**

a- Internal layer.
b- Four Layer PCB.
c- Corona.
d- Pad.

6 **¿Cómo se denomina el anillo circular remanente luego de la perforación de un pad?**

a- Internal layer.
b- Four Layer PCB.
c- Corona.
d- Core.

Respuestas: 1 b, 2 a, 3 d, 4 a, 5 b, 6 c.

www.FreeLibros.me

Capítulo 3

Instrumentos de medición

F3
CH1
MENU
CH2
MENU
F4
VOLTS / DIV
VOLTS / DIV
5mV
CH 2

En este tercer capítulo, explicaremos cuáles son las medidas que podemos realizar sobre un circuito.

RedUSERS.com

www.FreeLibros.me

Instrumentos de medición

En esta capítulo explicaremos cuáles son las medidas que podemos realizar sobre un circuito, cuál es la manera apropiada de hacerlo y qué instrumentos necesitamos.

Tomar una medida es comparar dos cantidades de una determinada magnitud física. Para esto, podemos contar con instrumentos digitales o analógicos. Los primeros –por definición de la norma **IEC 485** (*International Electrotecnic Comission*)– son aquellos donde la indicación aparece en forma numérica.

Los segundos –por definición de la norma **IEC 51**– son aquellos en los cuales la indicación resulta de relacionar la posición de un índice respecto de una escala graduada. Lo que diferencia a un instrumento analógico de uno digital son los procesos que se realizan en la adquisición de la señal a medir. Para el caso de estos últimos, los procedimientos involucran técnicas de conversión analógico-digital de la señal de entrada. Los analógicos, en cambio, proporcionan una salida de naturaleza continua, ante una entrada del mismo tipo.

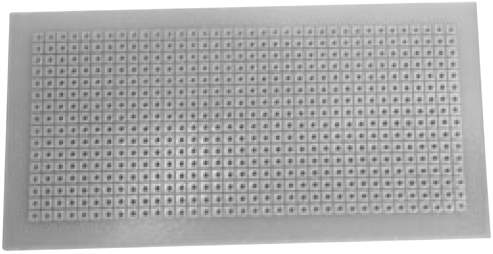

Instrumentos analógicos

En este apartado, conoceremos las nociones elementales inherentes a los instrumentos analógicos. Analizaremos sus parámetros y especificaciones más relevantes. La presencia de una escala graduada y de un índice o aguja para la indicación de resultados son datos más que suficientes para reconocer, en forma inmediata, un instrumento analógico. El corazón de todo instrumento de medición analógico en continua es el elemento de imán permanente y bobina móvil, también conocido como **D´Arsonval**.

Los analógicos son instrumentos en los que se establece un campo magnético en el espacio existente (entrehierro) entre la bobina móvil y el imán permanente (**Figura 1**). La bobina móvil es solidaria con el índice (aguja de indicación) que se desplaza sobre la escala graduada. Dicho campo magnético tiene dirección radial y uniforme. Esto conlleva a una relación matemática entre la **cupla** (par de fuerzas que genera el movimiento del índice) y la **señal** que se detecta.

ALCANCE DE TENSIÓN DEL INSTRUMENTO ANALÓGICO

Describiremos a continuación cómo se utiliza en los voltímetros analógicos, el instrumento de D´Arsonval como núcleo. Si llamamos **Im** a la intensidad de corriente máxima que circula por la bobina del galvanómetro, y **Rg** a la resistencia eléctrica que presenta dicha bobina, al aplicar la ley de Ohm, podemos determinar el alcance de tensión **Um** del instrumento básico. Se denomina alcance de un instrumento al máximo valor de medida –valor de fondo de escala–

www.FreeLibros.me

que puede indicar bajo una determinada configuración. Ahora bien, ¿qué pasaría si necesitáramos realizar medidas de tensiones mayores a las que permite el instrumento básico de D´Arsonval?

Para resolver esto, es necesario intercalar otra resistencia **Rs** (denominada resistencia multiplicadora) en serie con la bobina. También tendremos que aplicar una tensión **Us** mayor, para establecer una corriente **Im** que produzca la misma deflexión del índice. De este modo, hemos cambiado el alcance de tensión del instrumento de **Um** a **Us**.

Todo voltímetro analógico de continua incorpora el instrumento de D´Arsonval como núcleo

CIFRA EN OHM/VOLT

Cuando se habla de **cifra en Ohm/Volt**, se hace referencia a la indicación de la sensibilidad del instrumento. Recordemos que **Im** es la intensidad de

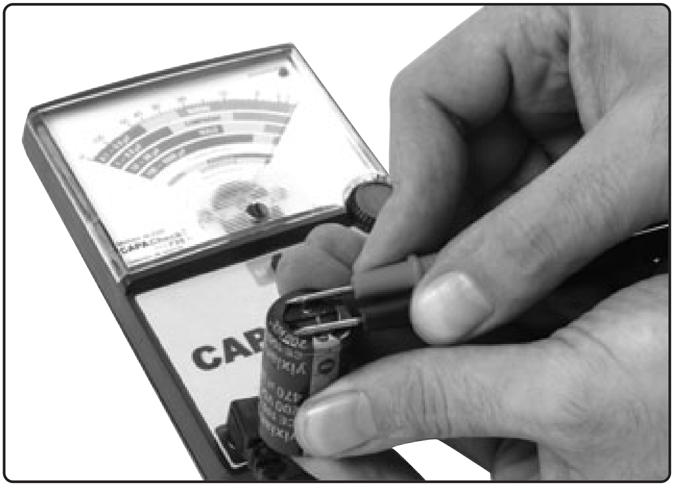

FIGURA 1.
Observamos un multímetro analógico. En su frente se destacan los valores de medición y el switch para alterna y continua.

INSTRUMENTO DE D´ARSONVAL

Es un miliamperímetro cuya indicación es sensible al valor medio de la corriente que circula por su bobina. Al ejercer una cupla antagónica de igual magnitud a la de la excitación, se consigue detener el índice sobre la escala graduada.

www.FreeLibros.me

ELECTR_Cap3_xxx_xxx_ajustado.qxp 04/07/2011 16:33 PÆgina 60

INFOGRAFÍA 3: INSTRUMENTOS DE MEDICIÓN

Fuente variable regulada
Instrumento que se utiliza para la prueba y excitación de circuitos de continua. En general estos instrumentos, nos brindan tensiones de trabajo de +-5 V, +-12 V y una fuente continua de 0 a 25 V con salidas desde 5 hasta 10 A. Poseen sistemas de limitación de corriente. Si aumenta la corriente hasta el límite configurado, la fuente cortará automáticamente.

Generador de señales
Fuente de señales de distintas formas de onda, amplitud, frecuencia y ciclo de trabajo. Nos permite excitar circuitos de prueba de manera sencilla, sin la necesidad de construir un circuito oscilador.

Te
Per
alte
Es
inte
cor
pru
gar

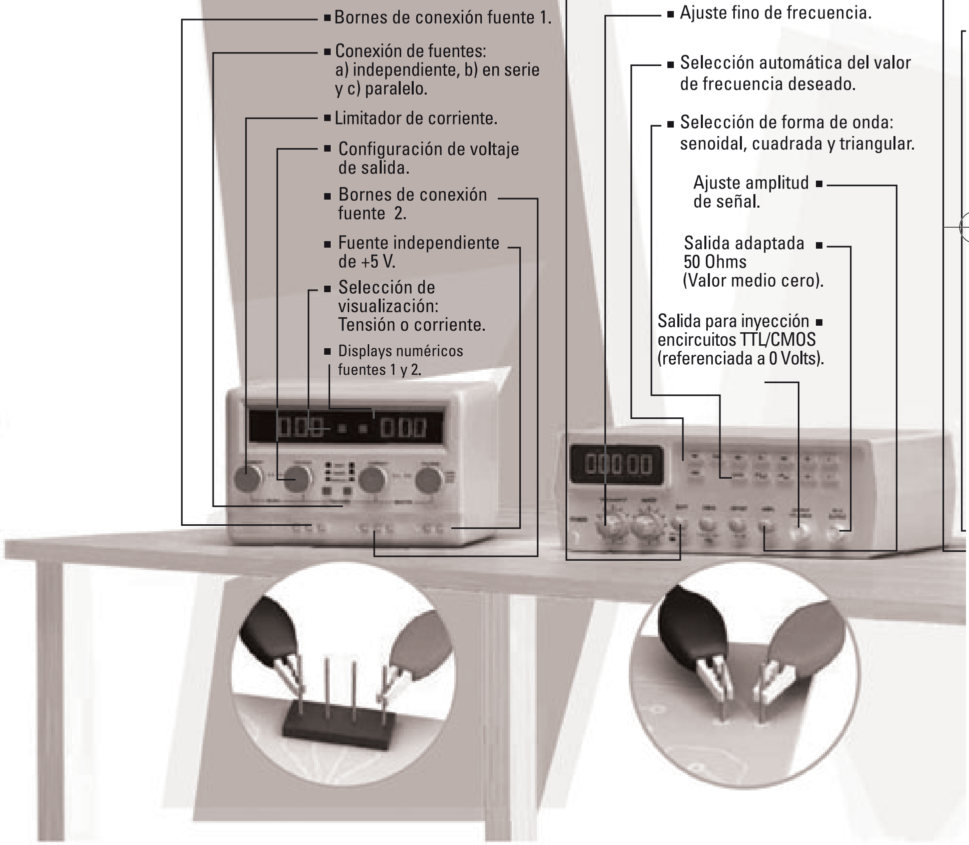

Bornes de conexión fuente 1.
Conexión de fuentes: a) independiente, b) en serie y c) paralelo.
Limitador de corriente.
Configuración de voltaje de salida.
Bornes de conexión fuente 2.
Fuente independiente de +5 V.
Selección de visualización: Tensión o corriente.
Displays numéricos fuentes 1 y 2.
Ajuste del ciclo de trabajo.
Ajuste fino de frecuencia.
Selección automática del valor de frecuencia deseado.
Selección de forma de onda: senoidal, cuadrada y triangular.
Ajuste amplitud de señal.
Salida adaptada 50 Ohms (Valor medio cero).
Salida para inyección encircuitos TTL/CMOS (referenciada a 0 Volts).

www.FreeLibros.me

ELECTR_Cap3_xxx_xxx_ajustado.qxp 04/07/2011 16:33 PÆgina 61

Tester digital o multímetro
Permite realizar las medidas más habituales, en alterna y en continua, con muy buena exactitud. Es posible medir diferencias de tensión, intensidad de corriente, resistencia y comprobar continuidad eléctrica. Incorpora funciones de prueba de diodos, capacímetros y medición de ganancia hfe de transistores bipolares.

Tester digital o multímetro
Permite realizar las medidas más habituales, en alterna y en continua, con muy buena exactitud. Es posible medir diferencias de tensión, intensidad de corriente, resistencia y comprobar continuidad eléctrica. Incorpora funciones de prueba de diodos, capacímetros y medición de ganancia hfe de transistores bipolares.

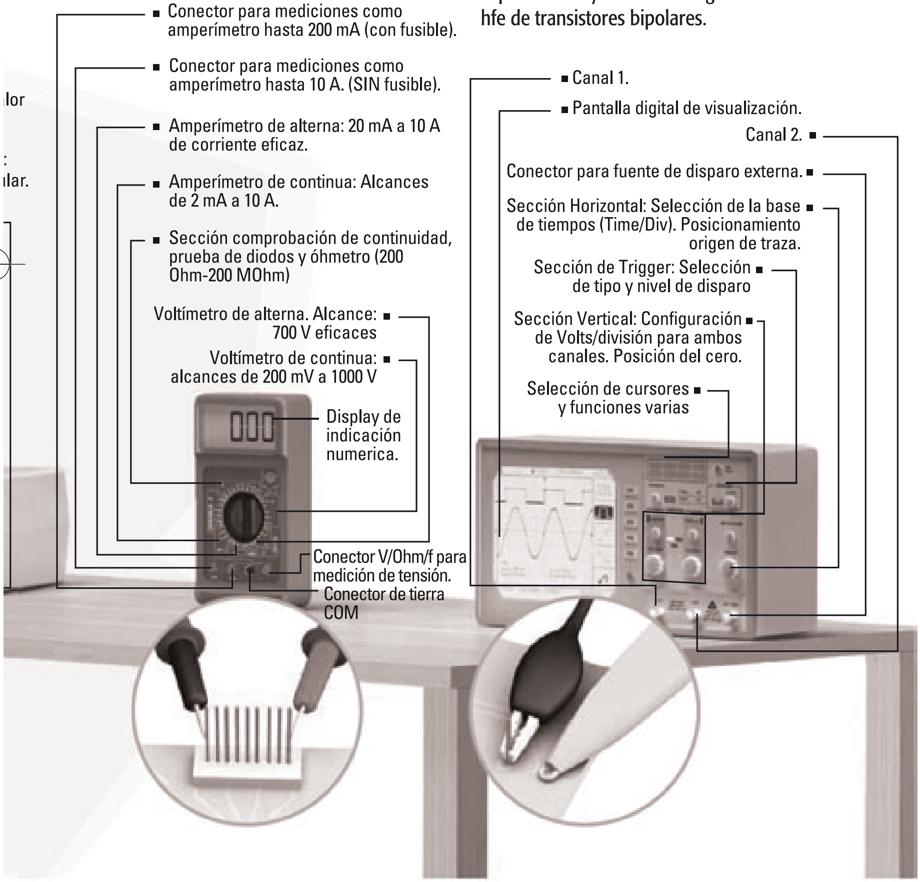
lor
:
lar.
Conector para mediciones como amperímetro hasta 200 mA (con fusible).
Conector para mediciones como amperímetro hasta 10 A. (SIN fusible).
Amperímetro de alterna: 20 mA a 10 A de corriente eficaz.
Amperímetro de continua: Alcances de 2 mA a 10 A.
Sección comprobación de continuidad, prueba de diodos y óhmetro (200 Ohm-200 MOhm)
Voltímetro de alterna. Alcance: 700 V eficaces
Voltímetro de continua: alcances de 200 mV a 1000 V
Display de indicación numerica.
000
Conector V/Ohm/f para medición de tensión.
Conector de tierra COM
Canal 1.
Pantalla digital de visualización.
Canal 2.
Conector para fuente de disparo externa.
Sección Horizontal: Selección de la base de tiempos (Time/Div). Posicionamiento origen de traza.
Sección de Trigger: Selección de tipo y nivel de disparo
Sección Vertical: Configuración de Volts/división para ambos canales. Posición del cero.
Selección de cursores y funciones varias

www.FreeLibros.me

corriente máxima admisible por el instrumento de **D´Arsonval**. Para el voltímetro se cumple la siguiente ecuación: **Im = Um / Rg = Us / Ri = cte,** donde **Ri** es la resistencia interna del voltímetro para un determinado alcance.

Llamaremos resistencia interna de un voltímetro a la suma de las resistencias **Rs** (multiplicadora) y **Rg** (del devanado de la bobina): Ri = Rs + Rg.

Al cociente **Ri / Us** se lo denomina **cifra en Ohm/Volt** y es una indicación de la sensibilidad del instrumento. A mayor cifra en Ohm/Volt, estaremos en presencia de un instrumento de mayor calidad.

AMPERÍMETRO ANALÓGICO

Un razonamiento similar al anterior se emplea para utilizar el instrumento de D´Arsonval cómo **amperímetro** de distintos alcances. En estos casos, se hace uso de una resistencia derivadora **Rshunt** (denominada resistencia de Shunt) en paralelo con el instrumento básico, conformando un divisor de corriente. Ante la excitación de una corriente **Is**, la resistencia derivadora Rshunt se elige de manera que, a través de la bobina de resistencia **Rg**, circule **Im**, por la simple aplicación de la **Ley de Ohm**.

Otro índice o parámetro de fabricación característico de los instrumentos analógicos es el denominado índice de clase C

ÍNDICE DE CLASE DE UN INSTRUMENTO ANALÓGICO

Otro índice o parámetro de fabricación característico de los instrumentos analógicos es el denominado **índice de clase C**, el cual nos permite calcular el error intrínseco del instrumento. Se define así al error que se comete sistemáticamente por la utilización del instrumento en la obtención de una medida. Es propio del instrumento. Dado el índice de **clase C**, el error intrínseco (**Ei**) del instrumento se obtiene al despejar la siguiente ecuación: **Ei = C x (Valor de fondo de escala)** (**Figura 2**).

FIGURA 2. **Los instrumentos analógicos son precisos, pero debemos saber que siempre hay un margen de error en la medición.**

www.FreeLibros.me

EL MULTÍMETRO DIGITAL

Se denomina multímetro a aquel instrumento que nos permite configurarlo como **voltímetro**, **amperímetro** y **óhmetro** (instrumento de medición de valores de resistencia). Las versiones digitales incorporan funciones de **capacímetro**, **frecuencímetro** (hasta 20 **KHz**), de medición de **hfe** (ganancia), de transistores y de termocuplas, para medir temperaturas sobre superficies. A estas funciones extra se les suman las utilidades de comprobación de **continuidad eléctrica** y **prueba de diodos**.

Podemos destacar también que con multímetros digitales es posible realizar medidas, no solo de **voltajes** y de **corrientes continuas**, sino también de **señales alternas** que se ajusten a los alcances del instrumento (**Figura 3**). Algunos multímetros más completos incorporan la función "**autorango**", es decir, el instrumento selecciona automáticamente el rango, sin necesidad de intervención de la persona.

PRINCIPIO DE FUNCIONAMIENTO

Un multímetro digital está formado por distintos bloques funcionales que conforman las etapas de tratamiento de la señal a medir. La primera etapa está compuesta por un amplificador atenuador construido a partir de amplificadores operacionales.

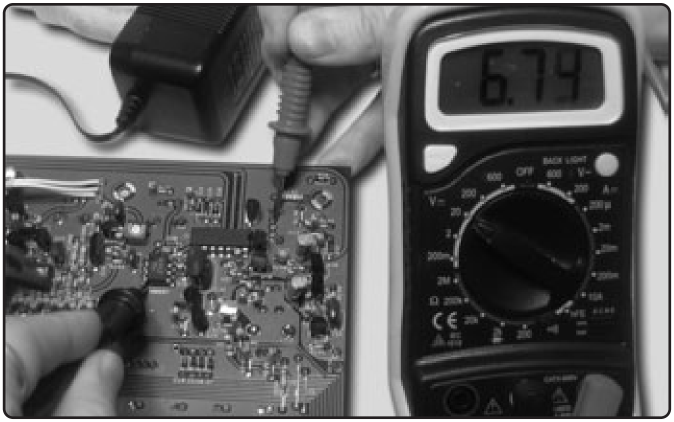

FIGURA 3.
Un multímetro digital **nos permitirá realizar la mayoría de las mediciones necesarias en la práctica: medidas de** tensión AC y DC, corriente AC y DC, capacidad, resistencia y comprobación de continuidad.

CONSEJO PARA INSTRUMENTOS ANALÓGICOS

Siempre conviene medir lo más próximo posible al fondo de escala, para que el error relativo tenga menor peso. El error relativo mínimo que podemos obtener es el que se comete cuando se mide a fondo de escala.

www.FreeLibros.me

El bloque atenuador, además de actuar como selector de rango, proporciona una elevada resistencia de entrada, para minimizar el error de inserción en las medidas de voltaje (**Figura 4**).

El bloque siguiente corresponde a un acondicionador de señal, que adapta la salida del amplificador atenuador al rango de trabajo del conversor A/D (generalmente, de **12 bits** y del tipo doble-rampa). Para mediciones de alterna, se deriva la entrada a circuitos rectificadores que convierten la señal a un valor de continua. Lo que resta, simplemente, es la etapa de procesamiento de la salida digital del conversor A/D, que permite la visualización del resultado en el display numérico (**Figura 5**).

USO DEL MULTÍMETRO DIGITAL

Aprenderemos a realizar dos mediciones sobre un circuito: la medida de diferencia de tensión entre dos puntos y la intensidad de corriente por una rama.

El multímetro es el instrumento que permite realizar la mayoría de las mediciones de un circuito

Antes de empezar, realizaremos unas comprobaciones de rutina. Primero, nos aseguraremos de que el **tester** tenga buena carga de batería. Algunos testers muestran el nivel de batería en el display. Pero también podemos darnos cuenta de que la batería está baja, porque el pitido que genera el instrumento –cuando medimos continuidad– es débil.

Verificaremos, también, que el **fusible** de protección del multímetro no se encuentre cortado. Generalmente, el fusible queda expuesto cuando retiramos la tapa trasera del tester. Además, controlaremos que las puntas y los cables de las puntas estén en buenas condiciones. Por último, al realizar las medidas, tomaremos la precaución de no tocar con nuestros dedos las partes metálicas de las puntas de medición; Las tomaremos siempre por su mango plástico, para no afectar las medidas y también por seguridad. De todas maneras, en este caso, los valores de tensión y corriente manejados son completamente

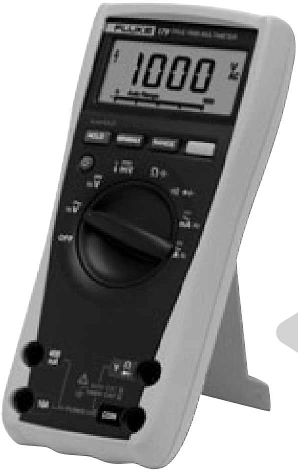

FIGURA 4.
Observamos la rueda de selección de un multímetro digital con los alcances y tipos de medición que se pueden realizar.

www.FreeLibros.me

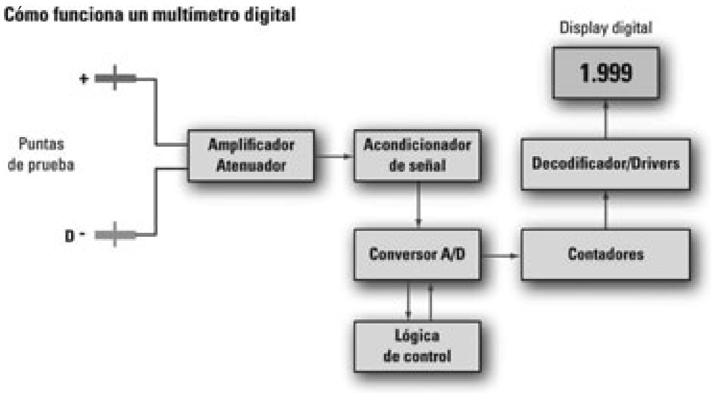
Cómo funciona un multimetro digital
Display digital
1.999
+
Puntas de prueba
D -
Amplificador Atenuador
Acondicionador de señal
Decodificador/Drivers
Conversor A/D
Contadores
Lógica de control

FIGURA 5. **Este es el esquema constructivo interno de un multímetro digital. Se observan todas las etapas involucradas hasta la visualización de la medida en el display numérico.**

inocuos. El circuito en cuestión se puede armar sencillamente con una **batería** de **9 V** y dos **resistencias comerciales** de **3,3 KOhm**.

Para medir una diferencia de tensión, ubicamos en el circuito los bornes de la **resistencia R2**. Debemos asegurarnos, previo a la medición, de que el alcance seleccionado sea el adecuado, para no dañar al tester y poder registrar la medida. La tensión a medir será aproximadamente de **4,5 V**. Luego, el alcance que se deberá seleccionar con la rueda será el de **20 V** para la "sección de continua".

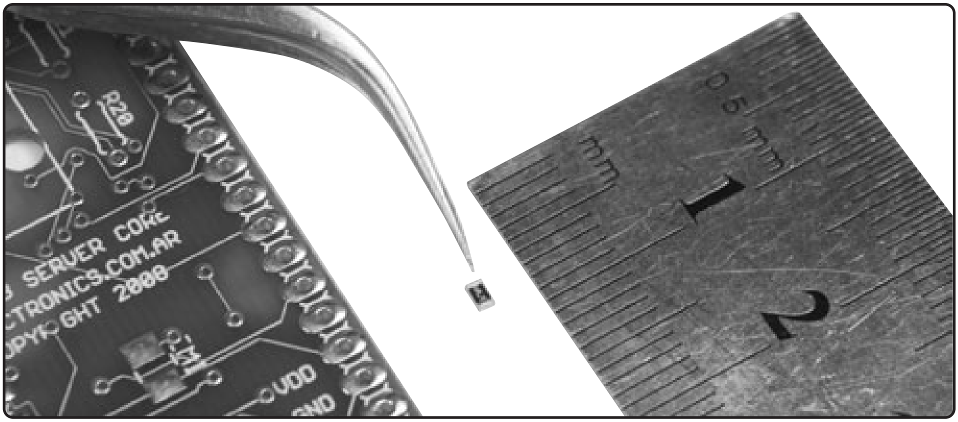
R28
SERVER CORE
CTRONICS.COM.AR
OPYRIGHT 2008
VDD
GND
0.5 mm
1
2

www.FreeLibros.me

Para medir una diferencia de **tensión positiva**, debemos ubicar la **punta roja** del tester digital sobre el borne superior de la resistencia **R2**. Debido al sentido de la circulación de corriente, este borne será el de valor tensión positivo. La masa o retorno del multímetro (**COM**) se ubicará sobre el otro borne, que corresponde a la tierra del circuito, es decir, nos "colgamos" en paralelo de la resistencia **R2** (**Figura 6**).

El tester configurado como voltímetro en su condición ideal actúa como un circuito abierto. Presenta, de este modo, una impedancia infinita cuando se conectan las puntas en paralelo con los bornes a medir. Esto no es así en la realidad. La resistencia interna (Ri) del tester tiene un valor definido que derivará corriente para sí, disminuirá la caída sobre R2 y afectará, así, la medida.

A esto se lo llama error de inserción. Por suerte, se los puede calcular en forma exacta. Para nuestro caso, hacemos las cuentas como se ve en la **Figura 7**.

Para realizar una medición de corriente, ahora configuraremos el multímetro como amperímetro. La primera diferencia con respecto a la medición de voltaje es que el tester deberá colocarse en serie con la corriente a medir. Así, nos vemos obligados a abrir el circuito para intercalar el multímetro. La corriente por medir será de **1,36 mA**.

Luego, el alcance a seleccionar será el de **2 mA** para la sección de corriente continua. Para efectuar la medición, debemos cambiar la posición del cable de **punta rojo** (**vivo**). Ahora, se ubicará en la entrada (**mA**) del tester, protegida por el fusible de **200 mA**.

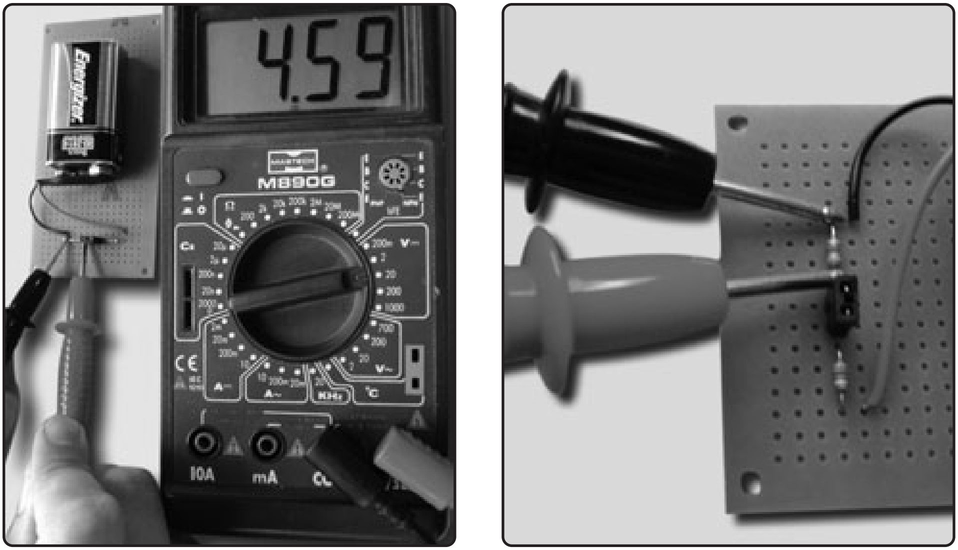

FIGURA 6. Al colocar las puntas al revés, la medida arrojará un valor negativo en el multímetro.

www.FreeLibros.me

Algunos testers muestran el nivel de batería en el display

Las puntas del tester deben colocarse de manera tal que la roja se conecte con la entrada de corriente a medir, y la negra, con el punto de salida. Registraremos así un valor positivo de corriente. Si invirtiéramos las puntas, el signo de la indicación en el display cambiaría. El amperímetro ideal es visto como un cortocircuito (**resistencia cero**) por el circuito. Pero la resistencia interna del amperímetro, aunque muy pequeña, se suma a la equivalente del circuito, y se opone también al paso de la corriente. Por lo tanto, la corriente medida tiene un valor menor al real. El cálculo de dicho error de inserción se ve en la **Figura 8**.

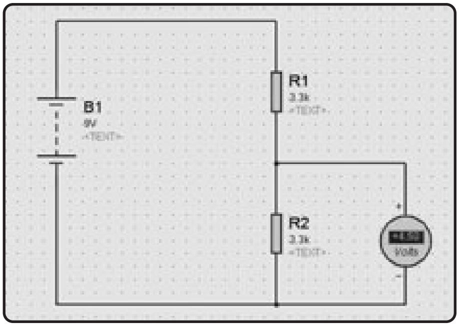

FIGURA 7.
Rinterna voltímetro = 1 Mohm
Vverdadero = 9 V * [(3.3 k + 3.3 k) / 3.3 k] = 4.5 V
Vmedida = [9 V / (R1 + R2 // Ri)] * (R2 // Ri) = [9 V / (3.3 k + 3.289 k)] * (3.289 k) = 4.4925
%error = (Verdadero–Vmedido)/Verdadero = 0.75% Menor al 1%.!!!

Cargador de baterías Ni-Cd

En este apartado realizaremos la comprobación de continuidad, y la medición de parámetros de componentes, voltajes y corrientes en funcionamiento normal. Haremos las mediciones pertinentes para la comprobación del funcionamiento del circuito cargador de baterías de Ni-Cd. Para tal tarea, utilizaremos un multímetro digital típico, de **3 ½ dígitos**. Todas las mediciones que llevaremos a cabo serán de continua, ya que el circuito no posee tensiones ni corrientes de alterna, salvo en el circuito de fuente, que no analizaremos aquí.

Primero, haremos ciertas comprobaciones sin alimentación, relacionadas con la seguridad eléctrica,

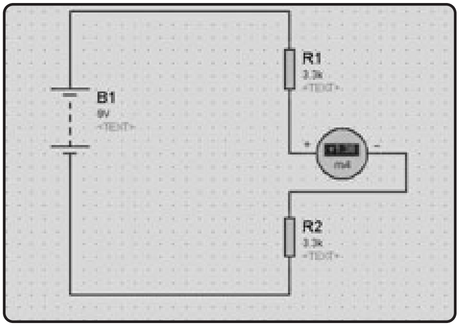

FIGURA 8.
Rinterna amperímetro = 100 Ohm (ejemplo)
Iverdadera = 9 V * [(3.3 k + 3.3 k)] = 1.3636 A
Imedida = [9 V / (R1 + R2 + Ri)] = 9 V / (3.3 k + 3.3 k + 100 Ohm) = 1.3433
%error = (Iverdadera–Imedida)/Iverdadera = 1.49 %

www.FreeLibros.me

ELECTR_Cap3_xxx_xxx_ajustado.qxp 04/07/2011 16:33 PÆgina 68

conexionado y buen funcionamiento de los componentes del circuito. Las otras mediciones se harán con el circuito cargador alimentado a **12 V de continua** (**Figura 9**).

No importa el orden con que se conecten las puntas sobre el circuito para medir continuidad

COMPROBACIONES DE CONTINUIDAD

Realizaremos una comprobación de rutina, para la cual se empleará el multímetro en el modo **detección de continuidad**. En este modo, no importa la posición de las puntas. Mientras haya conexión, la alarma sonora se activará indicando que existe continuidad eléctrica. La punta de medición roja debe conectarse al borne **V/Ohm** del multímetro y, la negra, a la tierra **(COM)**. La comprobación consiste en asegurarnos de que haya continuidad eléctrica en los bornes que deben estar conectados a la masa del circuito.

Tomando como referencia la **Figura 9**, hacemos base en **C** (GND a la entrada del circuito) con cualquier punta, y recorremos los puntos **D** (borne a masa de la resistencia de 10k) y **F** (borne negativo de la batería sobre el circuito). La alarma sonora indicará la existencia de continuidad eléctrica.

PARÁMETROS DE COMPONENTES

Ahora, vamos a medir ciertos parámetros de los componentes utilizados, que nos indicarán si se encuentran en buen estado. Salvo la medición del **hfe** (ganancia) del transistor **PNP**, las otras dos se pueden realizar con el circuito ensamblado. Las mediciones son las siguientes:

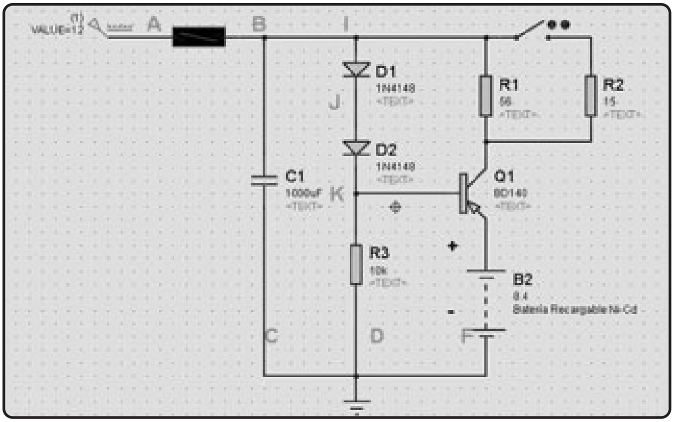

FIGURA 9. **Observamos el diseño del circuito donde se marcan los puntos de referencia para las mediciones.**

www.FreeLibros.me

- Medición del umbral de los diodos (I y J de la **Figura 9**, en la página anterior).
- Medición de las resistencias de selección de velocidad de carga.
- Medición del hfe (ganancia) del transistor PNP.

La utilidad de comprobación de diodos se encuentra casi siempre en la misma posición de la rueda que la detección de continuidad. Al posicionar la punta roja **V/Ohm** sobre el ánodo del diodo (**I**) y la punta **COM** sobre el cátodo (**J**), observaremos, en pantalla del multímetro, un valor en **mVolts**, que indicará la tensión umbral del diodo (**cercano a los 600 mV**).

Si invertimos las puntas de medición, el tester deberá indicar infinito, mostrando el número 1 en el bit más significativo. Esto es así ya que, ahora, la corriente que inyecta el tester para realizar la medición se encuentra con el diodo en inversa. Luego, el multímetro ve un circuito abierto y la medición no se efectúa por no poder establecerse la corriente.

A continuación, pasaremos a medir las resistencias de conmutación, cuyos valores se encuentran calculados para el establecimiento de dos intensidades de corriente fijas: **15 mA** y **45 mA**. Para llevar a cabo estas mediciones, emplearemos el multímetro digital como **óhmetro**. Siempre con la punta roja en el borne **V/Ohm** y la negra en la tierra común **COM**, medimos a ambos extremos de las resistencias del circuito. Es suficiente la selección del alcance más bajo de **200 Ohm**. La medición del hfe del transistor bipolar es muy sencilla de realizar, aunque no todos los multímetros incorporan esa funcionalidad.

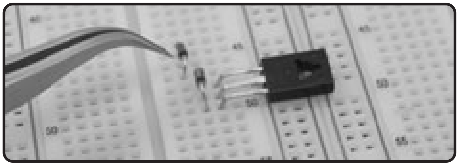

Una vez insertado el componente en el zócalo, se visualiza en pantalla la ganancia del transistor. El **BD140** es un transistor bipolar PNP, por lo que debe insertarse en el zócalo correspondiente. Esta es una función que mide el parámetro **hfe** de transistores bipolares. No debe usarse para hacer medidas de transistores del tipo **FET**, ya que puede dañarse (**Figura 10**).

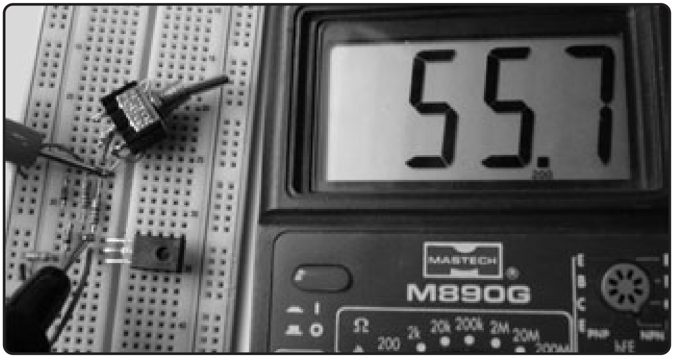

FIGURA 10.
Con el circuito sin alimentar, es posible utilizar el multímetro como óhmetro y estimar el valor de las resistencias.

www.FreeLibros.me

Mediciones con el circuito **alimentado**

Hasta el momento, aprendimos a tomar medidas. Ahora, realizaremos algunas mediciones elementales pero, en esta ocasión, lo haremos con el circuito alimentado con 12 V.

Realizaremos una medición de voltaje sobre los diodos **D1 y D2**. Una vez que se han realizado todas las mediciones anteriores, podemos proseguir a alimentar el circuito con los **12 V de continua**. Mediremos, ahora, la caída de tensión de "**on**" de ambos diodos en conducción. Estos diodos tienen como función polarizar, en forma permanente, la juntura **Emisor-Base** del transistor **PNP**.

Como todos sabemos, establecida **IB**, **IC** queda unívocamente determinada a través de la ganancia **hfe** del transistor de **unión bipola**r: **IC = hfe x IB**. Vamos a verificar, mediante dos rápidas medidas, que efectivamente encontremos 1,2 V sobre los diodos, polarizando la juntura del PNP. También veremos qué valor de tensión encontramos sobre la **juntura E-B**.

Como primer paso, seleccionaremos el alcance de **2 V** del instrumento, corriendo la rueda de selección hacia la posición indicada. Luego, colocaremos la punta roja sobre el **ánodo (I)** del diodo 1. Efectuaremos la medida, posicionando ahora la **punta COM** sobre el **cátodo** (**K**) correspondiente al diodo 2. Como era previsto, la medida arrojará un valor de tensión fijo cercano a los **+1,2 V**.

Para medir la corriente, ponemos el cable rojo en el borne (mA) del tester

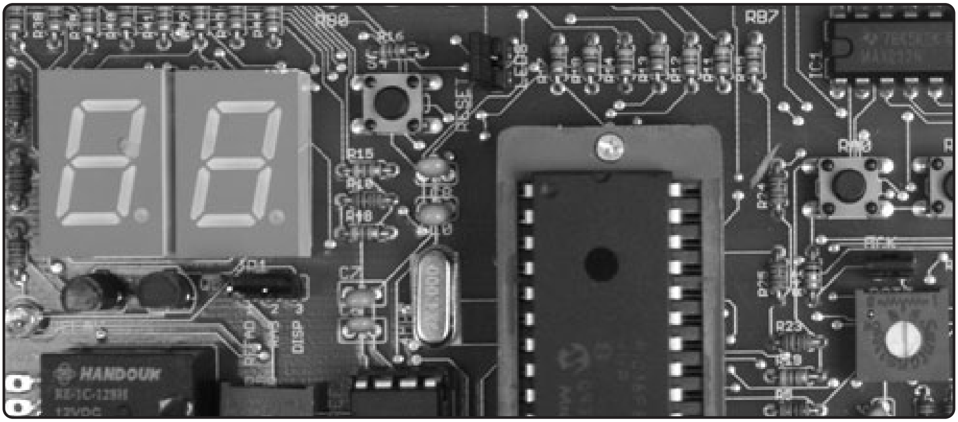

www.FreeLibros.me

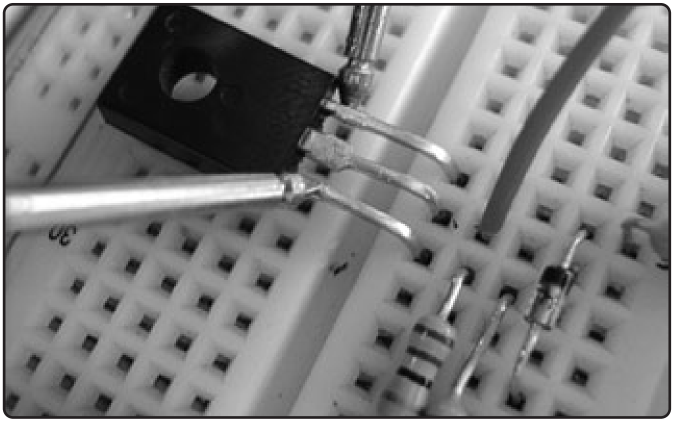

FIGURA 11.
Los diodos polarizan la juntura del transistor en forma permanente. Con esta configuración del tester, podremos medir la caída de tensión sobre la juntura E-B (emisor-base).

Si procedemos de la misma manera, pero ahora ubicamos las puntas para medir la diferencia de **tensión E-B** (**emisor-base**), comprobamos una caída cercana a **1 V**. Notemos que colocamos las puntas siguiendo el sentido de la corriente de base, es decir, saliendo de la misma. Por eso se ubica la punta negra **COM (-)** del tester sobre la base y la roja sobre el emisor que está a un potencial superior (**Figura 11**).

MEDICIÓN DE LA CORRIENTE DE CARGA IC

A continuación, pasemos a la medición de la corriente **Ic** de carga del circuito, que es la correspondiente al colector del transistor **PNP** y es la que se utilizará para cargar la batería cuando se la conecte en los bornes indicados. Como el transistor **BD140** se comporta como una fuente de corriente, siempre entregará por su colector una intensidad de carga constante, independientemente de la tensión que exista en bornes de la batería. Por cálculo, se estima que esta será de **15 mA** con la llave abierta y de **45 mA** con la llave en cortocircuito.

Ahora, necesitaremos dos alcances de continua para completar las mediciones.

Para realizar una medición de corriente, debemos desconectar el cable de punta roja del borne **(V/Ω/f)** del multímetro y posicionarlo en el borne marcado como **(mA)**, por el que pueden circular hasta **200 mA** sin destruir el fusible de protección. Si necesitáramos medir una corriente mayor a este valor, colocaríamos el cable de punta roja en el borne marcado **20 A**, lo que indica la máxima corriente que soporta el instrumento en esa conexión que, por otra parte, no está protegida por fusible.

Debido a que la resistencia interna **Ri** del tester en modo amperímetro es muy pequeña, no debe apreciarse un cambio radical en la corriente de colector o de carga. Procediendo así, tomamos la lectura del instrumento para ambas posiciones de la llave y usamos el alcance apropiado para cada una. Vale destacar que debemos posicionar las puntas del multímetro de manera que coincida la punta roja

www.FreeLibros.me

con el borne de entrada de corriente (**+**), y el negro COM con el de salida (**-**).

MEDICIÓN SOBRE LA BATERÍA DURANTE LA CARGA

Por último, debemos monitorear el valor de tensión en carga que toma la batería a medida que transcurre el tiempo. Para ello, posicionaremos el tester en paralelo con la batería en carga cada 15 minutos.

El tiempo de carga dependerá, principalmente, del parámetro **mA/h** de la batería en cuestión, pudiendo llegar a varias horas. Espaciaremos la toma de muestras en el tiempo, para no reiterar demasiado las mediciones. Siempre tenemos que comprobar que la tensión registrada no se escape del valor nominal **+10%** que establecimos como límite. Cuando detectemos que el valor de tensión alcanzó el **10%** del valor nominal de entrega de tensión de la batería, desconectaremos el circuito y consideraremos que la carga se ha realizado.

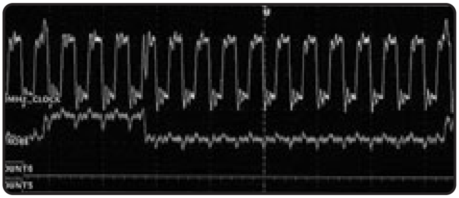

Para medir la corriente, debemos desconectar el cable rojo del borne (V/Ω/f) **y colocarlo en el borne** (mA)

El **osciloscopio**

Es uno de los instrumentos de medición más utilizados por los profesionales. Veamos de qué se trata y cuáles son las claves que necesitamos saber para emplearlo. El osciloscopio es un instrumento que permite reproducir gráficamente las variaciones de señales de tensión a través del tiempo, en una pantalla. De este modo, es posible analizar, con gran detalle, distintas formas de onda y comprender el funcionamiento de un circuito (**Figura 12**). Este elemento es una herramienta esencial en cualquier laboratorio o taller de electrónica y en otras áreas también, por ejemplo, la militar y la médica.

MEDICIONES CON OSCILOSCOPIO

El concepto básico de este instrumento es que representa formas de onda en dos dimensiones.

OSCILOSCOPIO DE MUESTREO

Es un instrumento utilizado para analizar señales periódicas de muy alta frecuencia, mayor aún que el ancho de banda (osciloscopio analógico) o la tasa de muestreo (osciloscopio digital), y de muy baja amplitud, generalmente, de hasta 3 V pico a pico.

www.FreeLibros.me

F3
F4
CH1 MENU
CH2 MENU
VOLTS / DIV
VOLTS / DIV
5V
5mV
mV
CH 2

FIGURA 12.
Es posible utilizar el osciloscopio para analizar distintas formas de onda individualmente, o también para observar y comparar dos o más, y ver el modo en el que interactúan entre sí.

El eje vertical, o **eje Y**, se usa en general para representar la tensión de entrada, mientras que el horizontal, o **eje X**, se emplea usualmente como eje de tiempo. Así, tendremos en pantalla una representación temporal de la señal de tensión aplicada a la entrada. En los osciloscopios duales y en los multicanales, es factible representar dos o más señales en pantalla en forma simultánea, lo cual resulta práctico para comparar distintas formas de onda y analizar su interacción.

Aunque la representación temporal de formas de onda es la aplicación más común en los osciloscopios, también es posible representar de manera gráfica el comportamiento de una variable respecto de otra, o funciones paramétricas. Un aplicación de este último caso son las **figuras de Lissajous**, mediante las cuales podemos determinar gráficamente algunos parámetros y relaciones de dos formas de onda de manera simultánea.

TIPOS DE OSCILOSCOPIOS

Más allá de que la filosofía de funcionamiento de estos instrumentos es la misma, existen diferentes tipos, lo que nos permite clasificarlos en categorías. Una de las más importantes es si son analógicos o digitales. A su vez, dentro de estos, tenemos subcategorías, como por ejemplo, los osciloscopios con persistencia y los de muestreo.

OSCILOSCOPIO ANALÓGICO

Utiliza el principio de deflexión electrostática de un haz de electrones, para crear las gráficas en un Tubo de Rayos Catódicos (TRC). Aquí, las tensiones aplicadas a las placas deflectoras X e Y producen un punto que se mueve en la pantalla. En el eje horizontal, esto es controlado por una base de tiempos, mientras que, en el eje vertical, la deflexión es proporcional a la amplitud de la señal de entrada.

OSCILOSCOPIO DE ALMACENAMIENTO DIGITAL (DSO)

Es la forma convencional de los osciloscopios digitales y emplea pantallas del tipo raster, como las usadas en los monitores de PC para representar las gráficas. El equipo convierte la señal de entrada a un formato digital, por medio de técnicas de conversión

www.FreeLibros.me

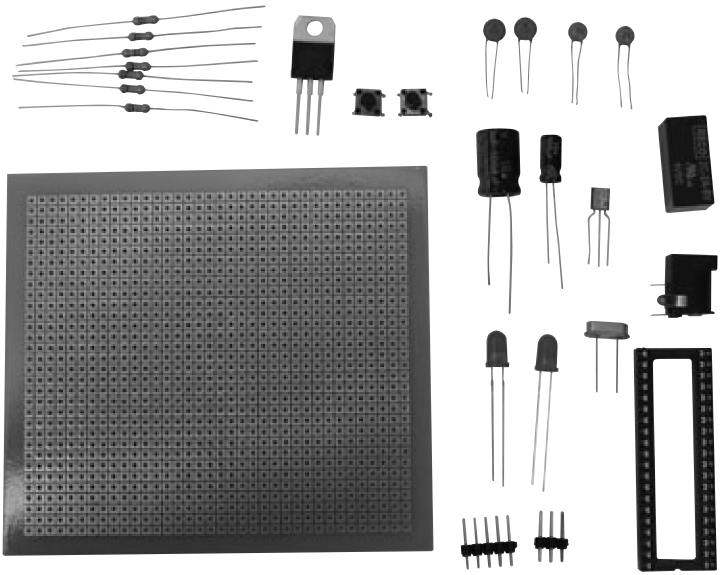

analógico-digital, para luego guardar esta información y procesar la señal. La limitante está en la velocidad de la conversión.

OSCILOSCOPIO DE FÓSFORO DIGITAL (DPO)

Con poco más de una década en el mercado, es una variación muy versátil del DSO, que utiliza una arquitectura de procesamiento en paralelo para permitir el análisis de señales que resultarían prácticamente imposibles con un DSO. Se especializa en el muestreo de fenómenos transitorios que ocurren en el inicio de los sistemas digitales (pulsos, glitches y errores de transición).

Las puntas de prueba son las encargadas de entregarnos la mejor señal

Funcionamiento del osciloscopio

Hasta el momento hemos conocido las características principales del osciloscopio. En este apartado, veremos los principios elementales de su funcionamiento. En el osciloscopio, una señal con forma de rampa proveniente de la base de tiempos, produce

www.FreeLibros.me

el movimiento del haz en el sentido horizontal. La señal a medir, convenientemente acondicionada, ocasiona el movimiento en sentido vertical. La base de tiempos se sincroniza con un instante de la señal a medir y está calibrada de modo de hacer corresponder el desplazamiento del haz con una medida temporal y visualizar la señal en la pantalla.

ANCHO DE BANDA

En un **osciloscopio analógico**, la respuesta en frecuencia del amplificador vertical ocasiona una atenuación gradual de las señales de alta frecuencia. Por ejemplo, un osciloscopio de 20 MHz puede visualizar correctamente una señal senoidal de 20 MHz, pero puede mostrar señales de mayor frecuencia, solo que serán atenuadas por el amplificador vertical.

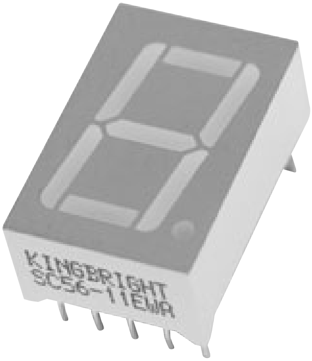

En el caso de un **osciloscopio digital**, existe una frecuencia de muestreo, que es la que determina la cantidad de muestras de la señal que el osciloscopio toma para reconstruirla. El fabricante especifica además el ancho de banda de forma similar al osciloscopio analógico. Para poder visualizar de manera correcta una señal, debemos garantizar que sus componentes más importantes entren en el ancho de banda que puede mostrar el osciloscopio.

DISPARO

La base de tiempos debe sincronizarse con la señal a visualizar para que siempre arranque la visualización en el mismo instante. Para esto, el osciloscopio incorpora un circuito de comparación que genera un pulso de sincronismo interno basado en la observación del nivel de la señal y el sentido (ascendente o descendente) en el que lo cruza. Esta comparación puede hacerse además con una señal externa. Existen, también, otros modos de trabajo en los que se fuerza un disparo aún sin señal, para ver señales sin variaciones, o se lo demora por un cierto tiempo, para evitar redisparos no deseados en señales complejas.

LOS OJOS DEL OSCILOSCOPIO

Las puntas de prueba son las encargadas de entregarnos la mejor señal. Veamos, entonces, cuáles son las características que deben tener para arrojar resultados precisos.

KARL FERDINAND BRAUN

El físico alemán Karl Ferdinand Braun fue el inventor del osciloscopio de rayos catódicos. Lo realizó en el año 1897, y también descubrió que ciertos cristales actuaban como rectificadores, convirtiendo la corriente alterna en continua.

www.FreeLibros.me

Muchas veces, perdemos de vista la importancia que tienen las puntas de prueba, cuando evaluamos integralmente las condiciones de medición.

Debemos tener presente que ellas conforman los "ojos" de nuestro osciloscopio. Si no son capaces de entregarnos la mejor calidad de señal posible, resulta poco probable que obtengamos buenos resultados de la señal bajo prueba. Idealmente, las puntas deberían ser "transparentes" para el equipo, queriendo significar con esto que la señal presente en el punto de prueba del circuito es "vista" por el osciloscopio, sin ningún tipo de interferencia o distorsión, tanto para el instrumento como para el circuito bajo prueba.

CALIBRACIÓN

Aunque no siempre lo hacemos, para garantizar una buena medición es fundamental efectuar una calibración de las puntas antes de ser usadas. De este modo, obtendremos una respuesta plana dentro del espectro de frecuencias (es decir, todo el espectro es tratado de igual forma).

Para ello, es posible utilizar la señal de calibración del osciloscopio, midiéndola con la punta a calibrar. Esta es una onda cuadrada de una amplitud y frecuencia estable y conocida. Debemos ajustar el pequeño potenciómetro incluido en la punta de prueba, hasta que la señal se vea perfectamente cuadrada: sus bordes no deben aparecer redondeados o con sobreimpulsos. Las puntas de prueba pueden tener una atenuación fija, o contener una llave conmutadora para seleccionarla. Estas atenuaciones vienen prefijadas en valores **x1** (sin atenuación) y **x10** (con atenuación). Veremos, aquí, la manera en que estas señales son presentadas en la pantalla de nuestro osciloscopio y la forma de interpretarlas.

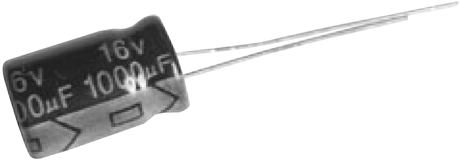

La atenuación x10 de las puntas lógicas se emplea para disminuir el efecto de carga de la punta sobre el circuito

Los osciloscopios digitales tienen características que ayudan a la interpretación de la señal, tal como la indicación automática de la frecuencia y el nivel de tensión de la señal mostrada.

Primero, observemos cómo está configurada la pantalla del instrumento. Se ve que está dividida, formando una retícula de **10x8** cm, con dos ejes centrales subdivididos cada **2 mm**. Los circuitos internos del osciloscopio están calibrados de manera tal que la representación gráfica de la amplitud y el tiempo de la señal en pantalla pueden leerse directamente a través de la retícula, contando simplemente la cantidad de divisiones que ocupa.

Esto es posible porque los mandos de amplitud (**eje vertical**) y base de tiempos (**eje horizontal**) están

www.FreeLibros.me

ajustados a ganancias prefijadas, relativas a cada división. Están expresadas en V por división (**V/div**) y, en segundos por división (**s/div**), respectivamente. Algunos osciloscopios incluyen también, en uno de los costados de la pantalla, una escala expresada en porcentaje. Particularmente, se indican los valores **0%**, **10%**, **90%** y **100%**. Estas marcas están para facilitar una lectura directa del tiempo de los flancos de subida y bajada de una determinada señal. Recordemos que estos parámetros se miden, justamente, entre el **10%** y el **90%** del valor pico a pico de la señal.

Entonces, si ajustamos la amplitud de la señal para que su valor pico a pico coincida con las marcas **0%** y **100%**, podremos obtener una lectura del tiempo de dichos parámetros directamente desde donde la señal cruza las marcas **10%** y **90%**. Solamente debemos contar las divisiones que ocupa y multiplicar por la base de tiempo seleccionada (**Figura 13**).

Cualquiera sea el tipo de osciloscopio que utilicemos, debemos tener presente que se trata de un instrumento de precisión que tiene sus limitaciones. Si bien para cada área de especialización existe un osciloscopio que se adapta mejor, la mayoría de las aplicaciones no profesionales pueden resolverse mediante uno analógico de **20 MHz**, siendo un **DSO** de **100 MHz** apropiado para gran cantidad de tareas profesionales.

CLIPS DE MASA

Las puntas del osciloscopio vienen con un clip de puesta a masa que es removible. Este será conectado a un punto de prueba del circuito para proveer la referencia común a ambos. En muchos casos no existe mayor inconveniente, pero cuando trabajamos con potencias o señales de muy alta frecuencia, deberemos tener el menor camino posible desde la señal hasta la punta de prueba, por lo que el clip de masa se conecta directamente al punto de masa más próximo al elemento a medir. Esto nos asegurará una muy buena performance desde el punto de vista eléctrico.

Cabe destacar que la masa del osciloscopio está vinculada al terminal de tierra del enchufe, y suele ser además un punto de puesta a tierra.

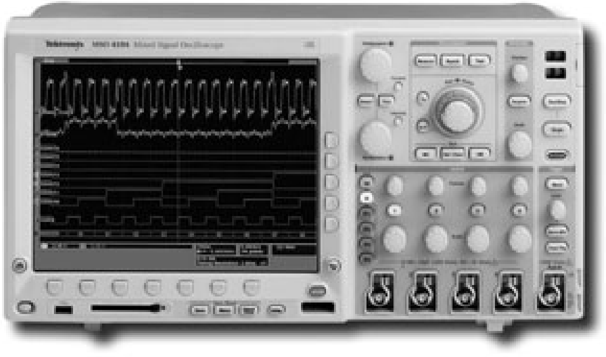

FIGURA 13.
Los circuitos internos del osciloscopio están calibrados de forma tal que la gráfica mostrada puede ser interpretada a través de las divisiones y marcas que posee la pantalla.

www.FreeLibros.me

MEDICIONES DE FORMAS DE ONDA

Toda señal que se repite en el tiempo es periódica. Posee parámetros que la definen y que son inherentes a su naturaleza: la amplitud, la frecuencia o el período y su fase. Veamos cada una de ellas:

- **Amplitud**: es la diferencia entre el valor máximo de la señal (positivo o negativo) y el valor tomado como referencia (masa del circuito). Su unidad es el Volt (**V**) o el Ampere (**A**), dependiendo de si estamos analizando formas de onda de tensión o de corriente, respectivamente.
- **Período y frecuencia**: la frecuencia de una señal se expresa con la letra **f** y la unidad que la representa es el Hertz (**Hz**). Indica la cantidad de ciclos u oscilaciones de una señal periódica que ocurren en un intervalo de tiempo de un segundo. El período se expresa con la letra **T** y su unidad es el segundo (**s**). Existe una relación unívoca entre ambos los parámetros: la frecuencia es la inversa del período. A continuación vemos la fase.
- **Fase**: este parámetro se define para señales periódicas, en donde un ciclo completo de la señal corresponde a 360° (**deg**) o 2 π radianes (**rad**). La fase de una señal en un determinado instante es la fracción de ciclo transcurrido desde su inicio. Cuando comparamos dos señales periódicas idénticas, puede ocurrir que no coincida en el tiempo el paso por dos puntos equivalentes de ambas señales.

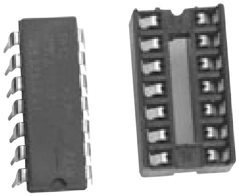

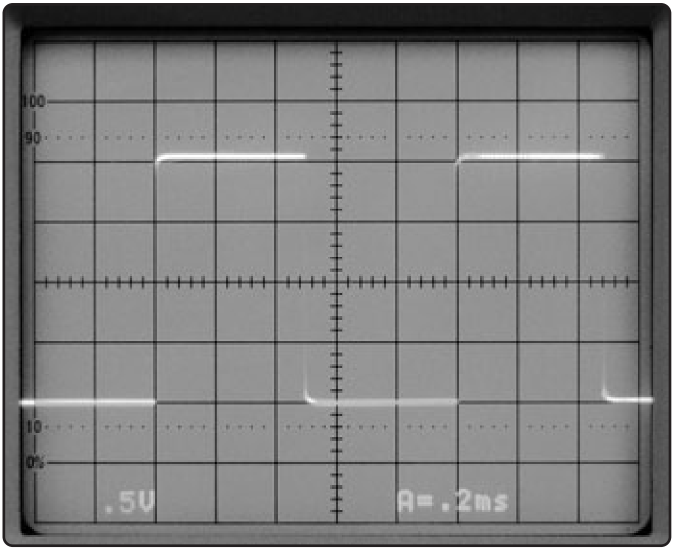

FIGURA 14.
Señal cuadrada bajo medición. Se utiliza la retícula calibrada para medir la amplitud y la frecuencia o período de la señal.

www.FreeLibros.me

ELECTR_Cap3_xxx_xxx_ajustado.qxp 04/07/2011 16:33 PÆgina 79

Medición de tensión y frecuencia

La medición de estos parámetros de la señal bajo análisis puede efectuarse a través de la retícula que posee la pantalla del osciloscopio. Hoy en día, la mayoría de los osciloscopios incorporan diversas funciones que nos ayudan a obtener una lectura más rápida y precisa de la señal en pantalla. Nos muestran información de la diferencia de tensión o de tiempo entre dos puntos, y otros datos relevantes de la señal (**Figura 14**).

MEDICIÓN DE TENSIÓN

En primer lugar, ajustamos la posición vertical del canal hasta que el piso o el techo de la señal coincidan con alguna división horizontal de la retícula, y contamos cuántas divisiones verticales ocupa.

A este valor, lo multiplicamos por el valor de ganancia del canal seleccionado y obtenemos, así, el valor de tensión pico a pico de la señal. En nuestro caso, la señal ocupa 4 divisiones verticales exactas, y como la ganancia del canal está en 0,5 V/div, tenemos:

$$V = 4\text{div} \,.\, 0{,}5\,\frac{V}{\text{div}} = 2\ \text{Volt}_{\text{picon pico}}$$

La amplitud es la diferencia entre el valor máximo de la señal y el valor tomado como referencia de masa

MEDICIÓN DE FRECUENCIA

Procedemos de forma similar a la medición de tensión. Ajustamos la posición horizontal del canal hasta que uno de los flancos (puede ser el ascendente o el descendente) coincida con alguna división vertical de la retícula. Luego, contamos cuántas divisiones horizontales ocupa un período de la señal.

A este valor, lo multiplicamos por la base de tiempo seleccionada y obtenemos, así, el período. Si hacemos la inversa del período, podemos inferir el valor de la frecuencia de la señal. En nuestro caso, la señal ocupa 5 divisiones horizontales exactas y, como la base de tiempo del canal está en 0,2 ms/div, tenemos:

$$T = 5\text{div} \,.\, 0{,}2\,\frac{\text{ms}}{\text{div}} = 1\ \text{ms}$$

$$\Rightarrow\ f = \frac{1}{T} = \frac{1}{1\,.\,10^{-3}\,\text{s}} = 1000\,\frac{1}{\text{s}} = 1\text{kHz}$$

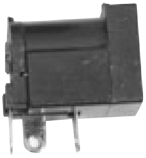

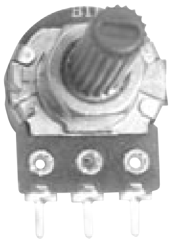

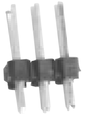

www.FreeLibros.me

Multiple **choice**

1 ¿Cómo se llama el instrumento que se utiliza para la prueba y excitación de circuitos de corriente continua?

a- Osciloscopio digital.
b- Tester digital o multímetro.
c- Fuente variable regulada.
d- Generador de señales.

2 ¿Cómo se llama el instrumento que nos permite excitar circuitos de prueba de manera sencilla, sin la necesidad de construir un circuito oscilador?

a- Osciloscopio digital.
b- Tester digital o multímetro.
c- Fuente variable regulada.
d- Generador de señales.

3 ¿Cómo se llama el instrumento que permite realizar las medidas más habituales, en alterna y en continua, con muy buena exactitud?

a- Osciloscopio digital.
b- Tester digital o multímetro.
c- Fuente variable regulada.
d- Generador de señales.

4 ¿Cómo se denomina el instrumento que nos permitirá analizar de manera directa, forma de onda, amplitud y frecuencia de señales periódicas?

a- Osciloscopio digital.
b- Tester digital o multímetro.
c- Fuente variable regulada.
d- Generador de señales.

5 ¿Cuál es la forma convencional de los osciloscopios digitales, que emplea pantallas del tipo raster?

a- Osciloscopio de fósforo digital.
b- Osciloscopio analógico.
c- Osciloscopio de almacenamiento digital.
d- Ninguno de los anteriores.

6 ¿Cuál es el osciloscopio que utiliza el principio de deflexión electrostática de un haz de electrones, para crear las gráficas en un Tubo de Rayos Catódicos?

a- Osciloscopio de fósforo digital.
b- Osciloscopio analógico.
c- Osciloscopio de almacenamiento digital.
d- Ninguno de los anteriores.

Respuestas: 1 c, 2 d, 3 b, 4 a, 5 c, 6 b.

www.FreeLibros.me

Capítulo 4

Diseño de circuitos impresos

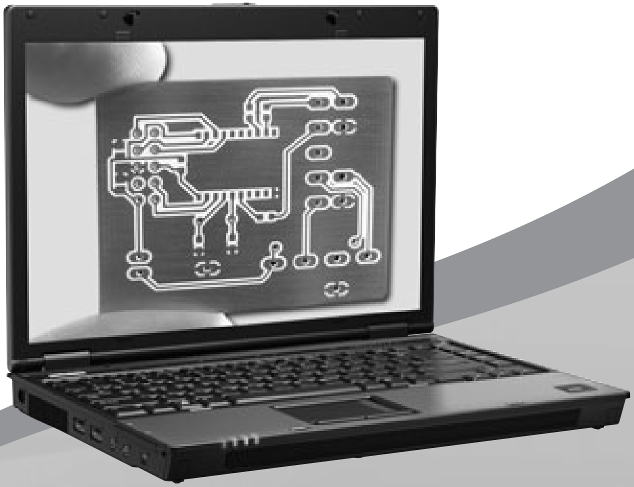

Conoceremos las características
del software para la creación, edición
y gestión de proyectos electrónicos.

RedUSERS.com

www.FreeLibros.me

Diseño de circuitos **impresos en la PC**

Conoceremos las características, funciones y limitaciones de algunos de los paquetes de diseño asistido por computadora para la creación, edición y gestión de proyectos electrónicos.

En esta clase, abordaremos los conceptos teóricos y prácticos que nos introducirán por completo en el mundo de la creación de diseños de circuitos impresos. Además, conoceremos las ventajas que existen sobre el método tradicional de trabajo, y los diferentes módulos, herramientas y áreas de trabajo que forman parte del paquete **CadSoft EAGLE**. Haremos hincapié en las bibliotecas de símbolos que vienen incluidas, en la edición y creación de nuevos componentes y en la generación de **PCBs** mediante métodos manuales, automáticos y mixtos. A modo informativo, conoceremos las limitaciones de la versión gratuita, el uso de reglas de ruteo y los diseños mediante capas. Culminaremos esta clase con un proyecto que nos guiará paso a paso hacia la creación de una placa correspondiente a un cargador de baterías Ni-Cd.

Un circuito impreso sirve como soporte físico y base de conexión para los componentes electrónicos

Recordemos que un circuito impreso es una placa o plaqueta de material aislante, generalmente construida en Pertinax o material Epoxi. Posee, en una o en ambas de sus caras, láminas de cobre delgado que conectan los pines de los diferentes componentes que conforman un circuito electrónico. El montaje de los componentes se realiza a través de orificios practicados en la placa o, en el caso de la tecnología de montaje superficial, soldados directamente sobre las láminas conductoras (**Figura 1**).

En la actualidad, el diseño de placas de circuitos impresos avanzó a la par del crecimiento de la informática. Debido al grado de complejidad de los circuitos electrónicos, es necesario utilizar herramientas dedicadas que brinden asistencia completa durante el proceso de diseño. Para dichos fines, surgen los programas de **CAD** (*Computer Aided Design*, **Diseño Asistido por Computadora**), que permiten realizar la captura, es decir, llevar a nuestra pantalla el circuito esquemático y generar los archivos de soporte necesarios para finalizar con el proceso de creación del PCB (**Figura 2**).

MÉTODO DE DISEÑO TRADICIONAL

Tradicionalmente, realizar el diseño de una placa de circuito impreso era un proceso tedioso y artesanal. Era necesario disponer de paciencia, mucha atención y pericia para lograr un diseño sin errores y que cumpliera una serie de reglas acordes al esquema elegido para armar.

Para trabajar, se utilizaban elementos de dibujo técnico, como reglas, escuadras, plantillas, lápices y fibras. Primero, había que llevar a cabo un boceto o borrador para poder darle forma, y conocer la ubicación y el

www.FreeLibros.me

ELECTR_Cap4_xxx_xxx_ajustado.qxp 04/07/2011 16:39 PÆgina 83

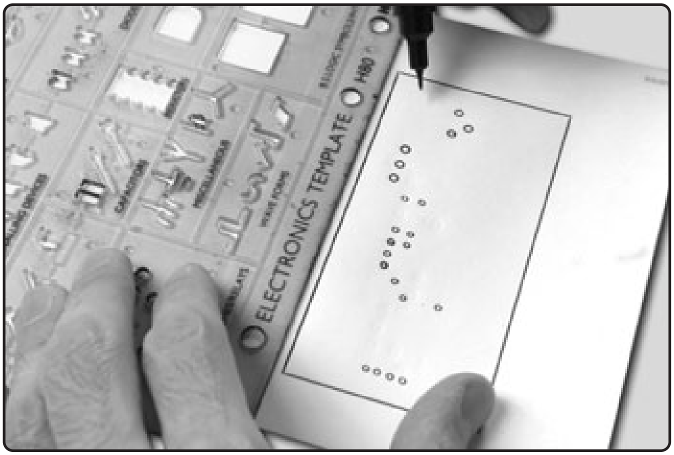
ELECTRONICS TEMPLATE

FIGURA 1.
Observamos el método manual de creación de un diseño de circuito impreso. Para esta labor, se utilizaban herramientas comunes de dibujo técnico y el trabajo era artesanal.

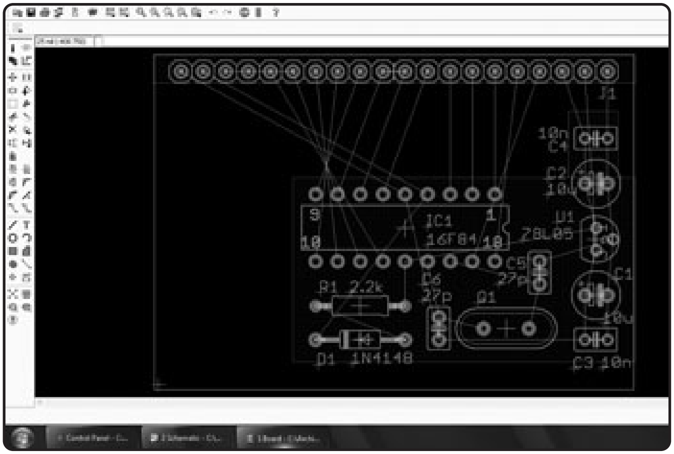
J1
10n
C4
C2
10u
IC1
16F84
9
1
10
18
U1
78L05
C5
27p
C1
10u
R1 2.2k
C6
27p
Q1
D1 1N4148
C3 10n

FIGURA 2.
Diseño de un prototipo con microcontrolador, realizado en una placa de circuito impreso de una sola cara utilizando el aplicativo Board**, incluido en el paquete CadSoft EAGLE.**

tamaño de los componentes. El segundo paso consistía en diagramar las conexiones eléctricas. Una vez concluido esto, se pasaba el dibujo en limpio, preferentemente con papel vegetal y tinta china.

También se utilizaba un método en el cual se recortaban y pegaban trozos de pistas, curvas y terminales, previamente diseñados sobre mylar. Esto llevaba el nombre de su mayor fabricante: Bishop.

www.FreeLibros.me

Para terminar, el arte era transferido por métodos fotográficos al proceso de fabricación final (fotográfico, serigráfico).

ENTORNOS DE DISEÑO CAD

Los programas de diseño de circuitos impresos están conformados por un entorno que incluye varios módulos. Uno muy importante es el que permite realizar la captura del circuito esquemático. En los casos donde este no se incluya, es primordial que el entorno permita importar las listas de conexiones generadas a partir de archivos esquemáticos creados con otra aplicación de la misma empresa o de terceros.
Si no tenemos ninguna opción de captura o importación de esquemáticos, nos veremos obligados a dibujar la placa, colocando los componentes de forma manual y uniendo los respectivos terminales. Este método es, en esencia, similar al manual, con la diferencia de que lograremos mayor calidad de impresión y definición de los trazos o pistas de conexión. Hay que tomar ciertos recaudos a la hora de trabajar con esta clase de software, ya que generalmente no incluyen ninguna opción de chequeo de errores ni reglas de diseño.

Los entornos profesionales de última generación, como **Altium Designer** y **Cadence OrCAD**, incluyen todo lo necesario, y todavía más (**Figuras 3** y **4**).

Los programas de diseño de circuitos impresos están conformados por un entorno que incluye varios módulos

Las diferencias entre el diseño manual y el diseño asistido por computadora son varias y muy importantes. La reducción del tiempo empleado en el diseño y la calidad final son las primeras en destacarse. Sin embargo, la optimización toma importancia a la hora de realizar el posicionamiento y el conexionado de los componentes. Esto reduce drásticamente los errores que se pueden llegar a producir.

Otras características que incorporan las aplicaciones de CAD son las que permiten la industrialización o producción en serie. Es posible exportar diferentes archivos para operar máquinas de taladrado y de corte por control numérico.

EL MÓDULO DE CAPTURA DE ESQUEMÁTICOS

Este es el módulo encargado de llevar a cabo la captura del diagrama eléctrico, es decir, el proceso de

SOBRE LOS ENTORNOS DE TRABAJO

Los entornos ideales para trabajar son los que permiten dibujar el diagrama esquemático (proceso de captura), generar los listados de conexiones y de componentes utilizados y, por último, exportar todo al módulo PCB.

www.FreeLibros.me

dibujo del circuito. Para ello, el software proporciona herramientas y funciones que permiten seleccionar los símbolos de los diferentes componentes que forman parte de él, realizar las conexiones eléctricas, asignar valores y nombres a los objetos utilizados, aplicar comandos de edición como borrar, copiar y pegar, e indicar conectores, fuentes de alimentación y puntos de prueba. Además, una vez concluido el proceso de captura, es posible exportar diferentes archivos, en general de texto en **formato ASCII**, con

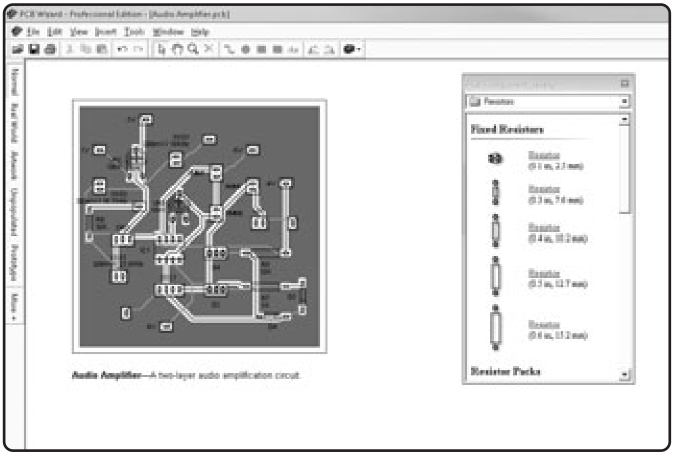
PCB Wizard - Professional Edition - [Audio Amplifier.pcb]
File Edit View Insert Tools Window Help
Fixed Resistors
Resistor Packs
Audio Amplifier—A two-layer audio amplification circuit.

FIGURA 3.
Entorno de diseño PCB Wizard. Permite diseñar PCBs de hasta dos caras. Es ideal para usar en ambientes educativos o para aficionados que realizan placas simples.

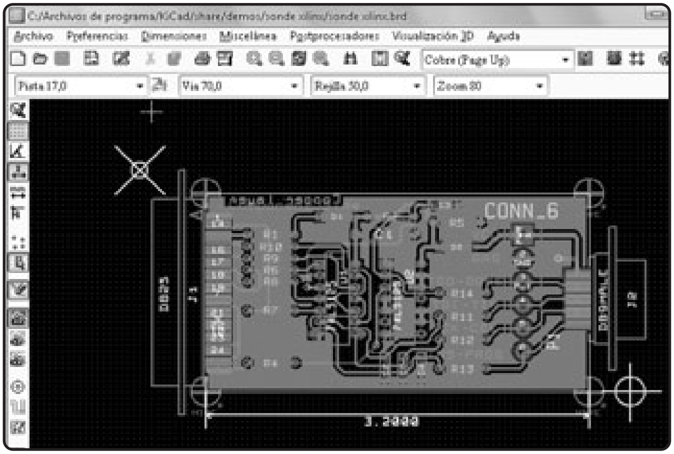
C:/Archivos de programa/KiCad/share/demos/sonde xilinx/sonde xilinx.brd
Archivo Preferencias Dimensiones Misceláneas Postprocesadores Visualización 3D Ayuda
Cobre (Page Up)
Pista 17,0
Via 70,0
Rejilla 30,0
Zoom 80
CONN_6
3.2000

FIGURA 4.
Entorno de diseño KiCAD. Es un completo set de herramientas de software libre, aplicado a la creación, edición y gestión de proyectos de diseño.

www.FreeLibros.me

datos útiles para otros programas de diseño. Usualmente, es posible exportar listados de materiales, listados con indicación de todas las conexiones y listados con errores, entre otros.

EL MÓDULO DE CAD

Una vez que el esquema está concluido, se debe pasar a la parte de diseño. Cada paquete de software de diseño de PCB cuenta con el **editor CAD**. Este módulo permite importar el diagrama capturado, convertir los símbolos en encapsulados y realizar las pistas de conexión. Este proceso puede ser manual, automático o una mezcla de ambos (**Figura 5**).

> Es preciso que el software de diseño cuente con un sistema de control que evite errores

EL CONTROL DE ERRORES

Todo buen programa de diseño de circuitos impresos permite controlar posibles errores cometidos durante el proceso de captura o de creación del PCB.

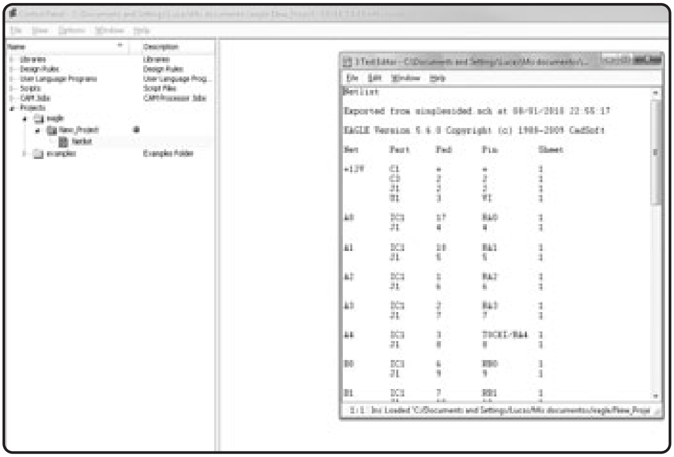

FIGURA 5.
El archivo netlist **realizado con** CadSoft EAGLE. **En él, se encuentran detallados todos los componentes que forman el circuito, con sus valores, pines y nodos de conexión.**

ARCHIVOS NETLIST

Es una lista de los nodos (nets). Un nodo se define como un punto de conexión común entre dos o más elementos. Según el archivo netlist, este puede incluir información variada, como el formato físico de los trazos de circuito impreso y las propiedades de cada componente.

www.FreeLibros.me

Por ejemplo, conexiones hechas entre los polos positivo y negativo de la fuente de alimentación o pines sin conexión eléctrica (**Figura 6**). Generalmente, se encuentran disponibles dos tipos de verificación: **ERC** (*Electrical Rules Check* - **Chequeo de Reglas Eléctricas**) y **DRC** (*Design Rule Checking* - **Chequeo de Reglas de Diseño**).

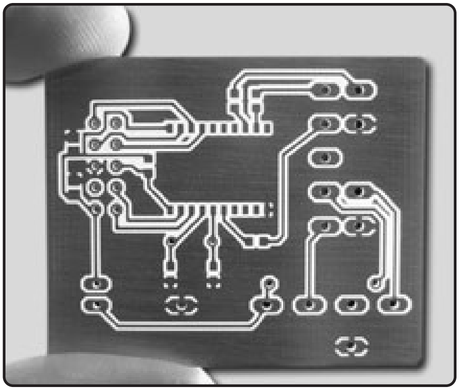

FIGURA 6. **Una vez diseñado y probado en la computadora, el circuito deja de ser virtual para pasar a ser real.**

CADSOFT EAGLE

Para iniciarnos en el manejo, la creación y la gestión de diseños simples, podemos descargar desde la página web **www.cadsoftusa.com** una versión gratuita para aplicaciones no comerciales. Una vez descargada e instalada, nos encontraremos con una potente aplicación de diseño.

El software posee un **Panel de control** (*Control Panel*) desde donde se gestionan todos los proyectos. Desde allí, se pueden generar o editar circuitos esquemáticos (archivos **.SCH**), **diseños de circuitos impresos** (archivos **.PCB**), librerías de componentes (archivos **.LBR**), scripts (archivos **.SCR**) y programas en lenguaje de usuario (archivos **.ULP**).

El Panel de control utiliza la vista en estructura de árbol para navegar a través de los archivos del sistema: librerías, scripts, proyectos, programas escritos en lenguaje de usuario y trabajos **CAM** (*Computer Aided Manufacturing* - **Fabricación asistida por computadora**). A la derecha de la estructura, se visualiza la descripción del ítem seleccionado (**Figura 7**).

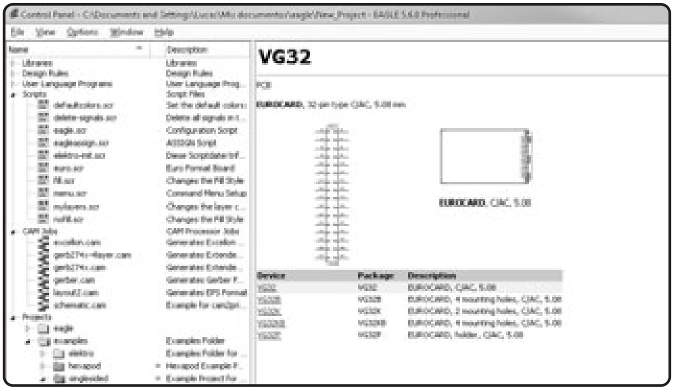
Control Panel - C:\Documents and Settings\Lucas\Mis documentos\eagle\New_Project - EAGLE 5.6.0 Professional
File
View
Options
Window
Help
VG32
PCB
EUROCARD, CJAC, 5.08

Name	Description
Libraries	Libraries
Design Rules	Design Rules
User Language Programs	User Language Prog...
Scripts	Script Files
defaultcolors.scr	Set the default colors
delete-signals.scr	Delete all signals in t...
eagle.scr	Configuration Script
eagleassign.scr	ASSIGN Script
elektro-init.scr	Diese Scriptdatei trif...
euro.scr	Euro Format Board
fill.scr	Changes the Fill Style
menu.scr	Command Menu Setup
mylayers.scr	Changes the layer c...
nofill.scr	Changes the Fill Style
CAM Jobs	CAM Processor Jobs
excellon.cam	Generates Excellon ...
gerb274x-4layer.cam	Generates Extende...
gerb274x.cam	Generates Extende...
gerber.cam	Generates Gerber F...
layout2.cam	Generates EPS Format
schematic.cam	Example for cam2pri...
Projects	
eagle	
examples	Examples Folder
elektro	Examples Folder for ...
hexapod	Hexapod Example F...
singlesided	Example Project For ...

EUROCARD, 32-pin type CJAC, 5.08 mm

Device	Package	Description
VG32	VG32	EUROCARD, CJAC, 5.08
VG32B	VG32B	EUROCARD, 4 mounting holes, CJAC, 5.08
VG32X	VG32X	EUROCARD, 2 mounting holes, CJAC, 5.08
VG32XB	VG32XB	EUROCARD, 4 mounting holes, CJAC, 5.08
VG32F	VG32F	EUROCARD, holder, CJAC, 5.08

FIGURA 7. **El** Panel de control **de** EAGLE **permite gestionar todos los archivos generados por sus distintos módulos, gracias al uso de una estructura de árbol.**

www.FreeLibros.me

Además del panel de control, **EAGLE** dispone de varios subprogramas o módulos para realizar cada tarea. Por ejemplo, **Schematic**, para la edición y captura de diagramas esquemáticos; **Board**, una herramienta de trabajo sobre circuitos impresos, y **Library**, un completo editor de librerías (**Figura 8**).

También posee un editor de textos para crear los archivos script o programas **ULP** y un procesador **CAM**. Este último módulo permite generar los datos necesarios para realizar la fabricación de las placas de circuito impreso. Puede producir archivos, por ejemplo, para controlar estaciones de taladrado, impresoras, plotters (máquina impresora lineal) y máquinas postscript.

Interfaz de los módulos

El usuario de EAGLE dispone de una barra de **Menú** donde se encuentran las funciones de cada módulo: utilidades de **zoom**, acceso a programas ULP y acceso directo a las otras aplicaciones del entorno. Debajo de esta barra de menú encontramos una barra de **Parámetros**. En este espacio, aparecerán distintas funciones y propiedades de acuerdo con la herramienta elegida. Una de las barras más importantes es **Herramientas**. Desde ella, es posible acceder a los botones que nos permitirán realizar y editar los esquemas y diseños de circuito impreso.

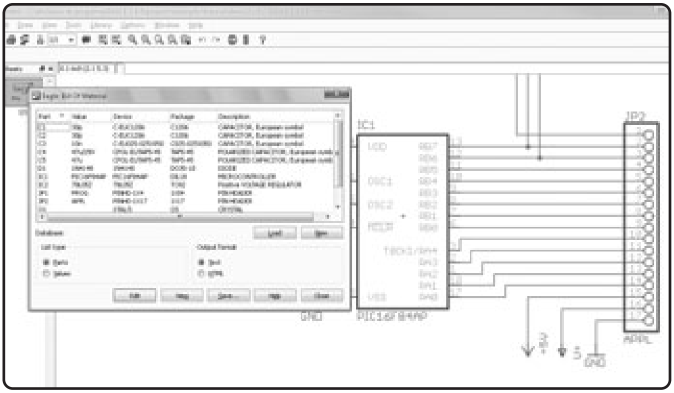

FIGURA 8. El módulo de captura Schematic **ejecutando una aplicación realizada en el** lenguaje ULP**. En este caso, se trata de un generador de listados de componentes con posibilidades de exportación a formato** HTML**.**

SOBRE ULP

EAGLE incorpora un lenguaje de programación llamado **ULP**. Se trata de una herramienta flexible, que permite tener un completo acceso a las estructuras de datos internas y escribir pequeños programas para añadir funcionalidades que resuelvan todo tipo de situaciones.

www.FreeLibros.me

ELECTR_Cap4_xxx_xxx_ajustado.qxp 04/07/2011 16:39 PÆgina 89

También es posible acceder a todos los comandos mediante órdenes de texto (con la línea de comandos). Asimismo, se encuentra disponible una **barra de Estado** que muestra informaciones y descripciones útiles sobre los comandos que están siendo utilizados. Además de las diferentes barras, cada módulo integra un **área de trabajo**. Dicha sección es el lugar de edición de los esquemas o librerías. Está delimitada por un rectángulo que equivale a la clásica hoja de papel en donde, de forma manual, antiguamente se dibujaban los circuitos.

COMPONENTES DISPONIBLES

EAGLE cuenta con una extensa librería de componentes por defecto, que se hallan organizados por fabricantes de dispositivos electrónicos, familias de productos y funciones (**Figura 9**). Estos símbolos generalmente se encuentran normalizados, es decir que respetan las medidas y los formatos físicos de los componentes, pero CadSoft no asegura que esto sea realmente así.

El usuario de EAGLE dispone de una barra de Menú donde se encuentran las funciones de cada módulo

Es un punto a tener en cuenta, ya que puede haber diferencias de tamaño con la vida real. También es posible agregar nuevas librerías desde el sitio de Internet de CadSoft, en la sección de descargas **www.cadsoftusa.com**, o utilizar el editor de librerías para crear nuevos componentes.

DISEÑO DE UN CIRCUITO ESQUEMÁTICO

El primer paso a seguir para crear un circuito impreso es la captura del diagrama esquemático. Como ya sabemos, EAGLE dispone de la herramienta

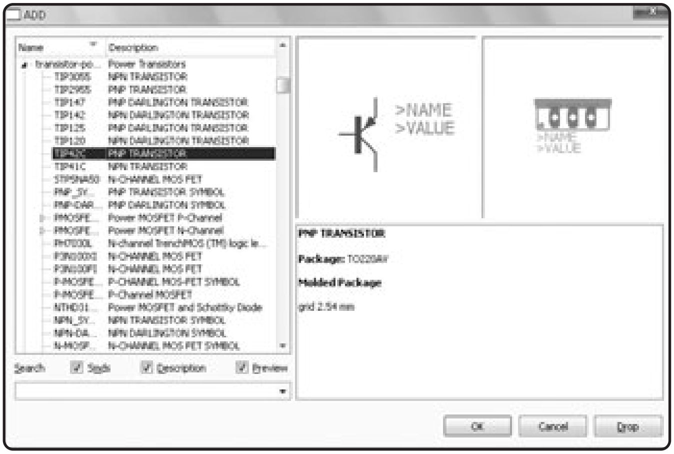

Name	Description
transistor-po...	Power Transistors
TIP3055	NPN TRANSISTOR
TIP2955	PNP TRANSISTOR
TIP147	PNP DARLINGTON TRANSISTOR
TIP142	NPN DARLINGTON TRANSISTOR
TIP125	PNP DARLINGTON TRANSISTOR
TIP120	NPN DARLINGTON TRANSISTOR
TIP42C	PNP TRANSISTOR
TIP41C	NPN TRANSISTOR
STP5NA50	N-CHANNEL MOS FET
PNP_SY...	PNP TRANSISTOR SYMBOL
PNP-DAR...	PNP DARLINGTON SYMBOL
PMOSFE...	Power MOSFET P-Channel
PMOSFE...	Power MOSFET N-Channel
PH7030L	N-channel TrenchMOS (TM) logic le...
P3N100XI	N-CHANNEL MOS FET
P3N100FI	N-CHANNEL MOS FET
P-MOSFE...	P-CHANNEL MOS-FET SYMBOL
P-MOSFE...	P-Channel MOSFET
NTHD31...	Power MOSFET and Schottky Diode
NPN_SY...	NPN TRANSISTOR SYMBOL
NPN-DA...	NPN DARLINGTON SYMBOL
N-MOSF...	N-CHANNEL MOS FET SYMBOL

PNP TRANSISTOR

Package: TO220AV

Molded Package

grid 2.54 mm

FIGURA 9.

Explorando las librerías. Al utilizar la función Add**, se ejecuta un diálogo que permite escoger un componente. Podemos ver el nombre, la descripción, el símbolo eléctrico, y el correspondiente encapsulado.**

www.FreeLibros.me

`Schematic` para realizar el dibujo. La abrimos desde **`Control Panel/File/New/Schematic`**. A continuación, se desplegará en pantalla la ventana del Editor de esquemas, en la cual podemos empezar a dibujar nuestro circuito.

El primer paso es escoger los componentes que utilizaremos. En la barra de **`Herramientas`**, disponemos del botón **`Add`**, que permite explorar las librerías, elegir uno a uno los símbolos y arrastrarlos hasta el área de dibujo (**Figura 10**).

El segundo paso consiste en ubicar los componentes en el lugar apropiado mediante el botón **`Move`**. También puede ser necesario emplear el botón **`Group`**, en caso de querer agrupar varios componentes para moverlos juntos. Para rotar objetos o cambiar su orientación, contamos con los botones **`Mirror`** y **`Rotate`**.

Una vez ubicados los componentes, deberemos llevar a cabo las conexiones eléctricas entre ellos. Escogemos de la barra de **`Herramientas`** el botón **`Net`**; en este modo nos posicionamos en cada pin de cada componente con el puntero del mouse y, al hacer clic, aparecerá el alambre que permitirá realizar la conexión. Este paso deberemos repetirlo hasta concluir con todos los componentes.

Por último, será necesario nombrar cada componente y configurar el valor correspondiente. Con los botones **`Name`** y **`Value`**, respectivamente, podremos realizar estas acciones.

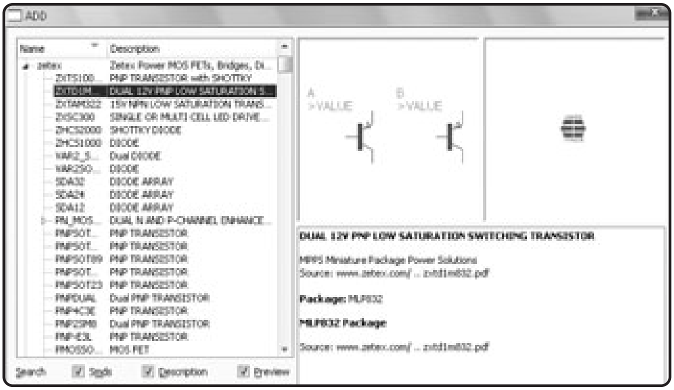

Name	Description
zetex	Zetex Power MOS FETs, Bridges, Di...
ZXTS100...	PNP TRANSISTOR with SHOTTKY
ZXTD1M...	DUAL 12V PNP LOW SATURATION S...
ZXTAM322	15V NPN LOW SATURATION TRANS...
ZXSC300	SINGLE OR MULTI CELL LED DRIVE...
ZHCS2000	SHOTTKY DIODE
ZHCS1000	DIODE
VAR2_5...	Dual DIODE
VAR250...	DIODE
SDA32	DIODE ARRAY
SDA24	DIODE ARRAY
SDA12	DIODE ARRAY
PN_MOS...	DUAL N AND P-CHANNEL ENHANCE...
PNPSOT...	PNP TRANSISTOR
PNPSOT...	PNP TRANSISTOR
PNPSOT89	PNP TRANSISTOR
PNPSOT...	PNP TRANSISTOR
PNPSOT23	PNP TRANSISTOR
PNPDUAL	Dual PNP TRANSISTOR
PNP4C3E	PNP TRANSISTOR
PNP2SM8	Dual PNP TRANSISTOR
PNP-E3L	PNP TRANSISTOR
PMOSSO...	MOS FET

DUAL 12V PNP LOW SATURATION SWITCHING TRANSISTOR

MPPS Miniature Package Power Solutions
Source: www.zetex.com/ ... zxtd1m832.pdf

Package: MLP832

MLP832 Package

Source: www.zetex.com/ ... zxtd1m832.pdf

FIGURA 10.
A través del botón `Add` **accedemos a la ventana de** `Selección de componentes`**. En ella podemos navegar una por una las librerías instaladas en el sistema, ver los componentes y escoger el adecuado para utilizar en nuestro esquema.**

VERIFICACIÓN DE ERRORES

EAGLE cuenta con una poderosa herramienta para detección de errores de diseño y eléctricos. Tenemos dos botones: **`ERC`** permite aplicar las reglas de diseño eléctrico, y el segundo botón, **`Errors`**, se utiliza para mostrar los errores encontrados.

www.FreeLibros.me

Creación **de librerías**

Para acceder al Editor de librerías, tenemos que dirigirnos a **`Control Panel/File/New/Library`** y se desplegará en nuestra pantalla la ventana del Editor. Lo primero que debemos hacer es crear una nueva librería vacía: vamos a **`New/File`**. Un punto a considerar es que dentro de una librería se guardan los dispositivos, y que cada uno de estos posee un símbolo y un encapsulado. En la **`barra de Herramientas`** disponemos de los botones que permiten acceder a la creación y edición de cada dispositivo, con su símbolo y su encapsulado.

Mediante el botón **`Symbol`**, tendremos la posibilidad de crear el símbolo que representará al dispositivo en el módulo de captura del esquemático. Lo pulsamos, le damos un nombre y, al presionar el botón **`Ok`**, podremos empezar a editarlo. Una vez creado el símbolo, deberemos agregar y posicionar los pines que forman parte de él. Para ello, nos valemos del botón **`Pin`**.

Los pines poseen propiedades que podemos modificar si hacemos clic con el botón derecho del mouse sobre cada uno. Es posible editar el nombre, la posición, el ángulo de inclinación, la dirección, la función y la longitud, entre varias funciones más.

Con los pines creados, posicionados y nombrados, podremos dibujar el componente

Una vez creados todos los pines, es necesario nombrarlos. Si bien esta opción se encuentra en el cuadro de propiedades, también disponemos del botón **`Name`** para realizarla. Al seleccionar y hacer clic sobre cada pin se nos habilitará dicha opción. La elección del nombre del pin generalmente tiene que ver con la función que cumple y es recomendable respetar este método.

Con los pines creados, posicionados y nombrados, podremos dibujar el componente. En la barra de **`Herramientas`**, pulsamos el botón **`Wire`**, seleccionamos la capa **94Symbols** y le damos forma al símbolo.

También es posible agregar texto con nombres y valores. Disponemos de la herramienta **`Text`** para hacerlo. Hay que tener en cuenta la capa en la que estamos trabajando, ya que cada tipo de texto posee una específica, como por ejemplo, **95Names**, **96Values**, **97Info**, entre otras.

¿QUÉ ES UN PAD?

Se llama **PAD** a la isla de soldadura o superficie de contacto entre los pines de un componente electrónico y las láminas de cobre. EAGLE permite configurar varios parámetros de los pads, como formato, diámetro, tamaño del orificio y ángulo de orientación.

www.FreeLibros.me

El segundo paso en la creación de una nueva librería es la realización del encapsulado (**Figura 11**). Este tipo de símbolo es la representación de cada componente que se utilizará en el **Editor de PCB** (**Board**) y reemplaza a cada símbolo eléctrico empleado en los diagramas esquemáticos.

En la barra de **Herramientas**, se encuentra disponible un botón llamado **Package** que, al ser presionado, desplegará una ventana donde podremos nombrar y crear el símbolo del encapsulado. Acto seguido, podremos agregar los pads correspondientes, editar sus propiedades y, por último, dibujar el encapsulado mediante la herramienta **Wire**.

Para finalizar con la creación de nuestra librería, tendremos que llevar a cabo un dispositivo nuevo. Presionamos el botón **Device** para que se despliegue la ventana en la cual podremos darle un nombre y confirmar su creación.

En la barra de **Herramientas**, encontramos el botón **Add**, que permite añadir un símbolo a nuestro dispositivo. Al pulsarlo, se desplegará una ventana donde veremos el símbolo creado anteriormente. Una vez finalizado este proceso, deberemos agregar el encapsulado correspondiente. Pulsamos el botón **New** para que se despliegue la ventana de diálogo con el encapsulado creado anteriormente.

Con el Editor de librerías**, es posible crear y editar librerías de componentes, símbolos y encapsulados.**

El último paso es hacer coincidir los pines del símbolo con los pads del encapsulado. Para esto, disponemos del botón **Connect**. Deberemos realizar estos pasos repetidas veces, hasta que estén todos los pines conectados. También es posible agregar una descripción general, que será visualizada en el explorador de librerías cada vez que busquemos un componente. Es recomendable completar este campo.

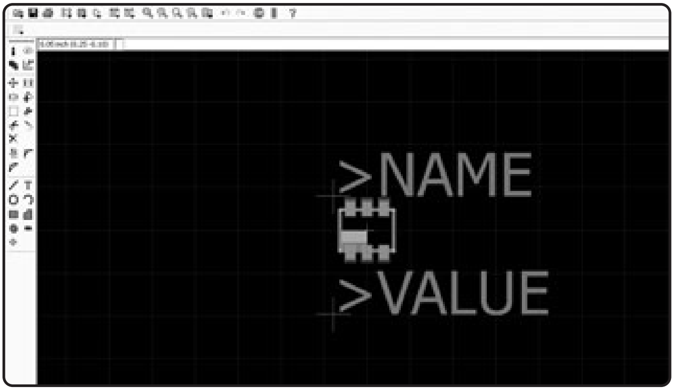

FIGURA 11. Mediante el Editor de encapsulados **es posible crear o modificar el encapsulado correspondiente.**

www.FreeLibros.me

Generación del **circuito impreso**

El editor de PCB puede lanzarse tanto desde el **`Panel de Control`** como desde el editor de esquemáticos, al pulsar el botón **`Board`**. En este último caso, una ventana nos preguntará si deseamos crear un nuevo diseño de circuito impreso a partir del diagrama esquemático. Veremos más adelante, al crear el PCB para nuestro cargador de baterías Ni-Cd, que aparecerán todos los componentes utilizados en el esquema, desorganizados y fuera del espacio definido para la placa (indicado por un rectángulo de color blanco). Las líneas de color amarillo indican las conexiones eléctricas entre ellos.

Podemos seleccionar los componentes por medio del botón **`Group`** y realizar cualquier acción, por ejemplo, moverlos con el botón **`Move`**.

Hay que tener en cuenta que, cuando el diseño parte de una captura de esquemático (realizada con el módulo **`Schematic`**), el programa no permite la inclusión de nuevos componentes ni conexiones que no figuren en el esquema. En el caso de tener que realizar modificaciones, habrá que hacerlas desde el módulo de captura.

Para transformar las líneas de conexión (*ratsnest* – **nido de ratas**) en pistas de circuito impreso, tarea que recibe el nombre de ruteo, disponemos del botón **`Route`**, que nos permitirá ir dibujando las pistas una a una. Solo basta con hacer clic con el puntero del mouse sobre uno de los pads a rutear y mover la pista a nuestra conveniencia. Al llegar al pad de destino, volvemos a hacer clic y el trazo queda terminado. Si nos equivocamos en el trazado, podremos volver a la pista anterior mediante el botón **`Ripup`**.

Al utilizar el botón **`Route`**, se despliega una nueva barra de Herramientas que permite configurar varios parámetros disponibles para las pistas. Es posible elegir el ángulo de los codos, el tipo de codo (rectangular o curvo), y hasta el ancho de las pistas. Por último, tenemos la posibilidad de agregar texto en cualquiera de las caras mediante el botón **`Text`**, y corregir aspectos de la serigrafía a través de **`Smash`** (**Figura 12**).

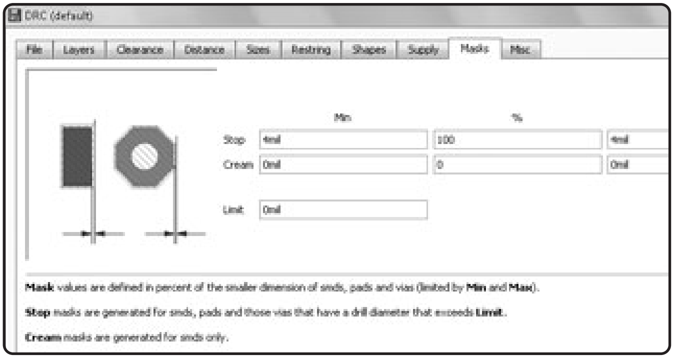

Stop 4mil 100 4mil
Cream 0mil 0 0mil
Limit 0mil

Mask values are defined in percent of the smaller dimension of smds, pads and vias (limited by **Min** and **Max**).

Stop masks are generated for smds, pads and those vias that have a drill diameter that exceeds **Limit**.

Cream masks are generated for smds only.

FIGURA 12.
Configuración de las reglas de diseño (DRC). Con este diálogo, es posible definir todos los parámetros referentes al diseño: tamaños, separaciones, distancias y terminación de las figuras son algunos de los disponibles.

www.FreeLibros.me

FUNCIÓN DE RUTEO AUTOMÁTICO

Esto permite que EAGLE realice el trabajo de ruteo de forma automática. Es posible utilizar el autoruteo en cualquier momento, sin importar la existencia de pistas trazadas de forma manual. Si existen, no se verán afectadas.

Al invocar el modo de autoruteo mediante el botón **`Auto`** de la barra de **`Herramientas`**, se despliega en pantalla una ventana con varias pestañas de configuración. En la pestaña **`General`**, es posible definir la cantidad de capas a emplear. En la versión gratuita, el máximo es de dos capas: superior (*Top*) e inferior (*Bottom*). En el caso de utilizar una única capa, también se puede decidir si será la que pertenece al lado de los componentes o al de las soldaduras. En esta pestaña, disponemos de dos opciones y cada una posee una lista desplegable que permite elegir uno de los parámetros disponibles (**"-"**, "**|**", "**/**", "\", "*****" y "**`N/A`**"). La opción "**`N/A`**" posibilita desactivar una de las capas, mientras que los otros símbolos indican la dirección que predomina en los trazos de cada capa: horizontal, vertical, diagonales y automático.

Una vez hecha la configuración, presionamos el botón **`Ok`** y veremos cómo EAGLE empieza a buscar todas las combinaciones posibles de trazado de pistas para realizar el ruteo. Dependiendo de la complejidad del esquema, será el tiempo que tarde en lograr el objetivo (**Figuras 13** y **14**).

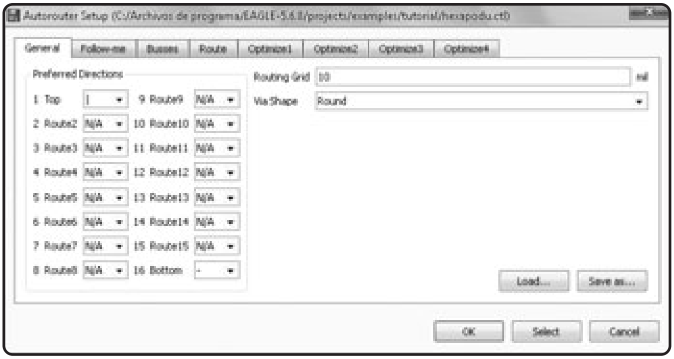

1 Top |
2 Route2 N/A
3 Route3 N/A
4 Route4 N/A
5 Route5 N/A
6 Route6 N/A
7 Route7 N/A
8 Route8 N/A
9 Route9 N/A
10 Route10 N/A
11 Route11 N/A
12 Route12 N/A
13 Route13 N/A
14 Route14 N/A
15 Route15 N/A
16 Bottom -
Routing Grid 10 mil
Via Shape Round

FIGURA 13. **Ventana de configuración del sistema de autoruteo. Si bien cuenta con varias pestañas con opciones, podemos lograr resultados satisfactorios si escogemos algunos parámetros únicamente desde el menú** `General`**.**

REGLAS DE DISEÑO (DRC)

Es muy importante utilizar y configurar las reglas de diseño (DRC). Mediante un correcto uso de esta función, es posible reducir los errores, como por ejemplo, pistas solapadas, violaciones de distancias mínimas y objetos fuera del área de trabajo.

www.FreeLibros.me

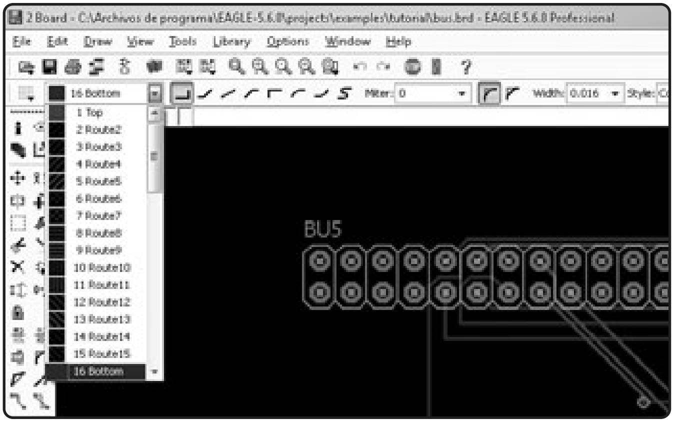
2 Board - C:\Archivos de programa\EAGLE-5.6.0\projects\examples\tutorial\bus.brd - EAGLE 5.6.0 Professional

16 Bottom

Miter: 0

Width: 0.016

FIGURA 14.
Placa generada mediante la función de autoruteo utilizando dos capas, superior e inferior. Para realizar las conexiones entre pistas de diferentes capas, se emplean las `Vías`**.**

MANEJO DE CAPAS

EAGLE utiliza diferentes capas para trabajar con los objetos que forman parte de un esquema. El resultado final es parte de una combinación de las capas.

Por ejemplo, para imprimir el arte que será transferido a la lámina de cobre de una placa de una sola cara se deberán activar las capas **`Bottom`** y **`Pad`** con el botón **`Display`**. Es posible realizar diferentes combinaciones para los trabajos de impresión o para limpiar el área de trabajo cuando los diseños son complejos y se dificulta seguir al ratsnest.

CÓMO UTILIZAR CADSOFT EAGLE

La idea de este **Paso a paso 1** es familiarizarnos con las opciones principales del programa, para poder orientarnos en la utilización de sus aplicaciones fundamentales. Como hemos visto en páginas anteriores, hay dos secciones importantes dentro del programa. Una de ellas es la que emplearemos para la confección del esquema del circuito (modo **`Schematic`**). Con este modo, seleccionaremos los componentes necesarios e indicaremos las conexiones entre ellos. La otra sección de relevancia es la que nos servirá para realizar el esquema de la placa (modo **`Board`**). Con este modo, trazaremos las rutas del circuito que unirán los componentes. Finalmente, podremos imprimir el esquema del circuito PCB.

DISEÑO DE UN PCB

Veremos un ejemplo de cómo utilizar EAGLE para la construcción de un esquema de circuito sencillo y práctico, correspondiente al conocido cargador de baterías (**Paso a paso 2**).

Realizaremos el diseño de una placa de circuito impreso (**PCB** - *Printed Circuit Board*) para el cargador de pilas Ni-Cd, que desarrollamos en el **Capítulo 2**. Para ello, en primer lugar, llevaremos a cabo el esquemático y, luego, lo transferiremos al módulo de diseño de PCB. Ubicaremos los componentes, haremos algunos trazados en forma manual y, finalmente, pediremos a EAGLE que realice el ruteo en forma automática.

www.FreeLibros.me

ELECTR_Cap4_xxx_xxx_ajustado.qxp 04/07/2011 16:39 PÆgina 96

PASO A PASO /1
Cómo utilizar CadSoft EAGLE

1

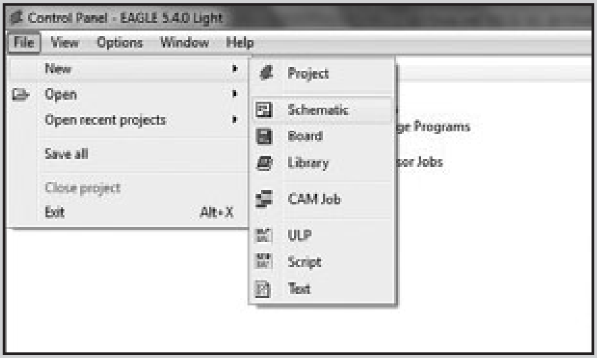
Control Panel - EAGLE 5.4.0 Light
File
View
Options
Window
Help
New
Open
Open recent projects
Save all
Close project
Exit
Alt+X
Project
Schematic
Board
Library
CAM Job
ULP
Script
Text
ge Programs
sor Jobs

Cuando ejecute el programa CadSoft EAGLE, aparecerán diversas opciones en la barra superior de la pantalla. Por el momento, vaya a la opción **`File/New/Schematic`**, que servirá para crear un nuevo esquema de circuito. Si desea abrir un esquema que haya guardado anteriormente, vaya a **`File/Open/Schematic`** y busque el archivo.

2

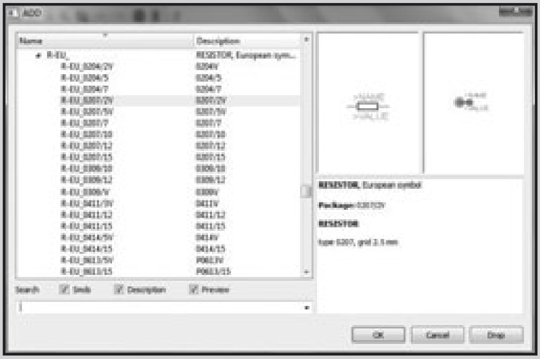
ADD
Name
Description
R-EU_
RESISTOR, European sym...
RESISTOR, European symbol
Search
Smds
Description
Preview
OK
Cancel
Drop

Verá un panel blanco que corresponde a la ventana principal y la barra de herramientas a la izquierda de la pantalla. Por ahora, usará la herramienta **`Add`**, que permitirá agregar nuevos componentes. Al hacer clic con el mouse sobre esa opción, se abrirá una ventana con una amplia lista de componentes. Seleccione el adecuado y haga clic en **`OK`**.

www.FreeLibros.me

ELECTR_Cap4_xxx_xxx_ajustado.qxp 04/07/2011 16:39 PÆgina 97

PASO A PASO /1 (cont.)

3

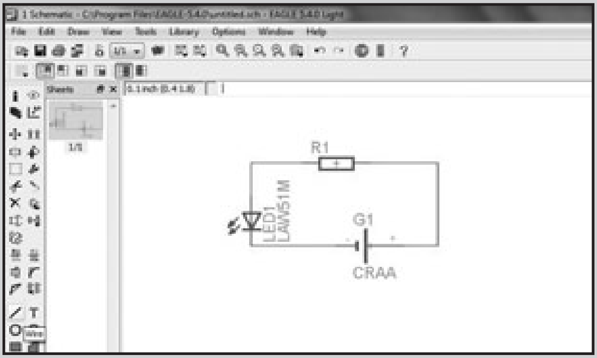
1 Schematic - C:\Program Files\EAGLE-5.4.0\untitled.sch - EAGLE 5.4.0 Light
File
Edit
Draw
View
Tools
Library
Options
Window
Help
Sheets
1/1
0.1 inch (0.4 1.8)
R1
LED1
LAW51M
G1
CRAA

Use la herramienta **`Net`** para trazar las líneas. La herramienta **`Rotate`** posibilitará rotar cada componente cada vez que sea necesario. Cuando termine, haga clic en la opción **`Board`**, ubicada en una de las barras superiores. Verá un mensaje que le preguntará si desea crear el diseño de la placa a partir del esquema del circuito. Haga clic en **`Yes`**.

4

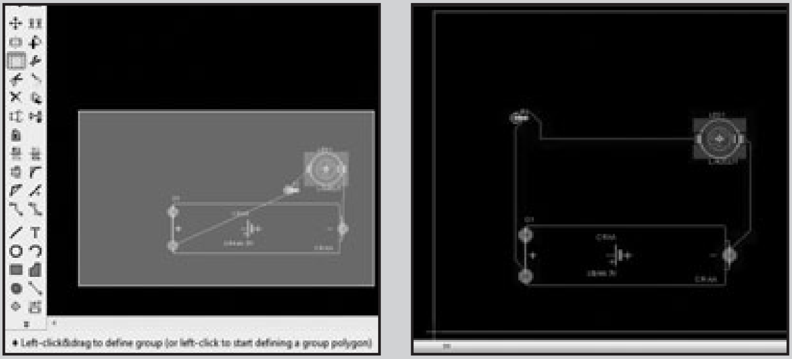

Left-click&drag to define group (or left-click to start defining a group polygon)

Tiene tres elementos: el panel principal de color negro, el dibujo del circuito y el sombreado de color blanco, que indica la zona delimitada para la placa, donde debe mover los componentes. Utilice **`Group`** para sombrear el circuito y presione **`Move`** para mover los objetos. Pulse la tecla **`CTRL`** y haga clic con el botón derecho para arrastrarlo hacia el panel principal.

www.FreeLibros.me

ELECTR_Cap4_xxx_xxx_ajustado.qxp 04/07/2011 16:39 PÆgina 98

PASO A PASO /1 (cont.)

5

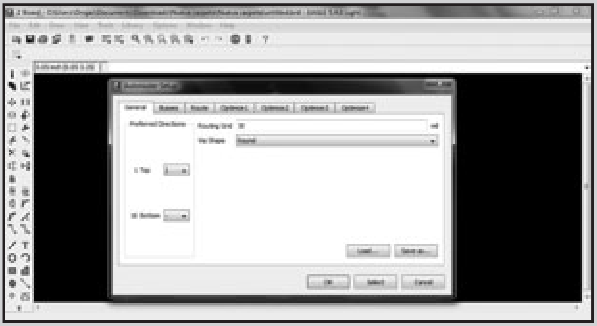

En la barra superior, seleccione **Tools/Auto.** Aparecerá un cuadro con las opciones del autoruteo. Haga clic en **Ok** y verá cómo el circuito se ordena de una forma particular que, por ahora, deberá considerarse como la más conveniente. Puede grabar el esquema de la placa desde **File/Save** y elegir su ubicación en el disco duro.

PASO A PASO /2
Diseño de un PCB para el circuito cargador de baterías Ni-Cd

1

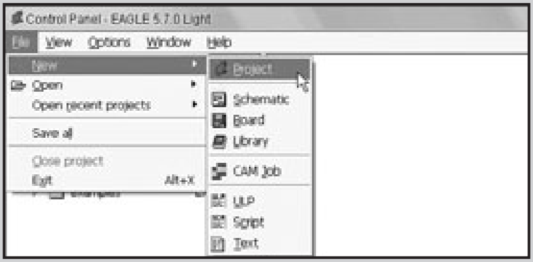
Control Panel - EAGLE 5.7.0 Light
File
View
Options
Window
Help
New
Open
Open recent projects
Save all
Close project
Exit
Alt+X
Project
Schematic
Board
Library
CAM Job
ULP
Script
Text

Abra un nuevo proyecto para comenzar con la creación del circuito. En la barra superior, haga clic en **File**. Se desplegará un submenú con opciones. Seleccione **New** y, luego, **Project**.

www.FreeLibros.me

ELECTR_Cap4_xxx_xxx_ajustado.qxp 04/07/2011 16:39 PÆgina 99

PASO A PASO /2 (cont.)

2

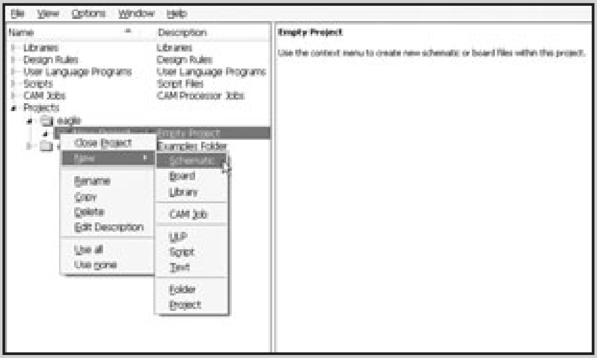
File View Options Window Help
Close Project
New
Rename
Copy
Delete
Edit Description
Use all
Use none
Empty Project
Examples Folder
Schematic
Board
Library
CAM Job
ULP
Script
Text
Folder
Project
eagle

Name	Description
Libraries	Libraries
Design Rules	Design Rules
User Language Programs	User Language Programs
Scripts	Script Files
CAM Jobs	CAM Processor Jobs
Projects	

Empty Project

Use the context menu to create new schematic or board files within this project.

Una vez creado el nuevo proyecto, el cursor se posicionará sobre la carpeta en cuestión. Puede cambiarle el nombre o eliminarlo. Haga clic con el botón derecho del mouse sobre el proyecto y se desplegará un menú de diálogo. Seleccione la opción **New** y, luego, **Schematic**, que le servirá para crear un nuevo esquema del circuito.

3

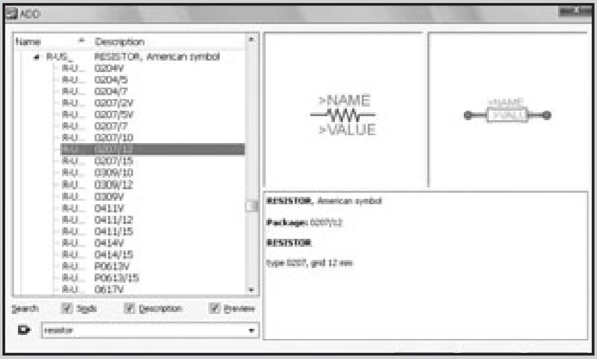
ADD
>NAME
>VALUE
Search
☑ Smds
☑ Description
☑ Preview

Name	Description
R-US_	RESISTOR, American symbol
R-U...	0204V
R-U...	0204/5
R-U...	0204/7
R-U...	0207/2V
R-U...	0207/5V
R-U...	0207/7
R-U...	0207/10
[illegible]	[illegible]
R-U...	0207/15
R-U...	0309/10
R-U...	0309/12
R-U...	0309V
R-U...	0411V
R-U...	0411/12
R-U...	0411/15
R-U...	0414V
R-U...	0414/15
R-U...	P0613V
R-U...	P0613/15
R-U...	0617V

resistor

RESISTOR, American symbol

Package: 0207/12

RESISTOR

type 0207, grid 12 mm

Ahora, seleccione los componentes necesarios para el circuito. Como vio anteriormente, debe agregarlos mediante la opción **Add**. Tiene que adicionar tres resistores, entonces, busque por resistor y seleccione aquél que se acerca mejor a lo que tiene, en este caso, en **R-US, R-US_207/12**. Repita esta operación para el capacitor electrolítico y los diodos **1N4148**.

www.FreeLibros.me

ELECTR_Cap4_xxx_xxx_ajustado.qxp 04/07/2011 16:39 PÆgina 100

PASO A PASO /2 (cont.)

4

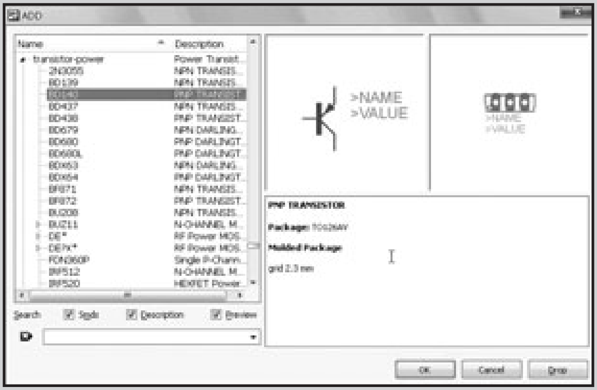

Name	Description
transistor-power	Power Transist...
2N3055	NPN TRANSIS...
BD139	NPN TRANSIS...
BD140	PNP TRANSIST...
BD437	NPN TRANSIS...
BD438	PNP TRANSIST...
BD679	NPN DARLING...
BD680	PNP DARLINGT...
BD680L	PNP DARLINGT...
BDX53	NPN DARLING...
BDX54	PNP DARLINGT...
BF871	NPN TRANSIS...
BF872	PNP TRANSIST...
BU208	NPN TRANSIS...
BUZ11	N-CHANNEL M...
DE*	RF Power MOS...
DEIX*	RF Power MOS...
FDN360P	Single P-Chann...
IRF512	N-CHANNEL M...
IRF520	HEXFET Power...

PNP TRANSISTOR

Package: TO126AV

Molded Package

grid 2.3 mm

A continuación, necesita seleccionar el transistor BD140, que encontrará dentro de la sección de transistores de potencia. Para seleccionarlo, tiene que seguir el mismo procedimiento del paso anterior. Recuerde que la opción **`Add`** le permite adicionar el componente elegido.

5

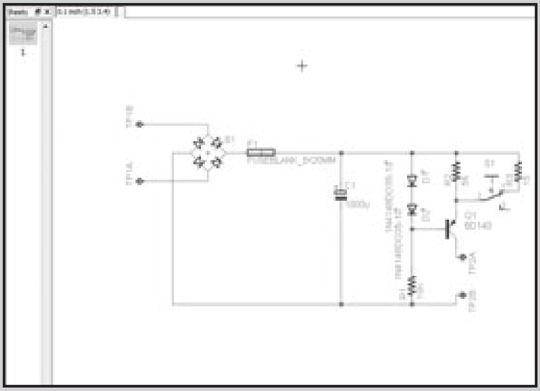

Realice el esquema del circuito. Para girar los componentes, utilice la herramienta **`Rotate`**, ubicada en la barra de la izquierda de la pantalla. Para trazar las líneas use la opción **`Net`**; de este modo, unirá cada uno de los componentes, formando la configuración deseada del circuito.

www.FreeLibros.me

PASO A PASO /2 (cont.)

6

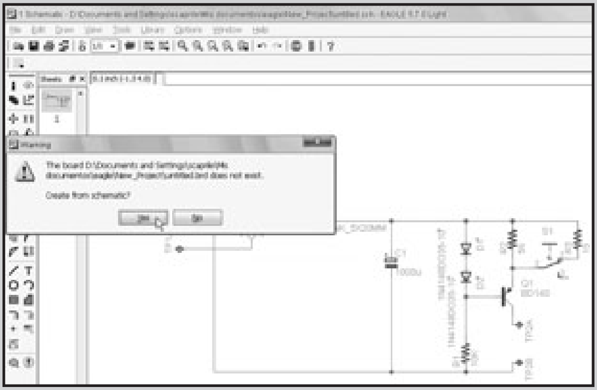

Una vez que haya terminado el dibujo del esquema, proceda seleccionar la opción **`Board`**, ubicada en la barra debajo del menú. Aparecerá un mensaje informativo que le preguntará si desea generar una placa a partir del esquemático del circuito. Haga clic en **`Yes`**.

7

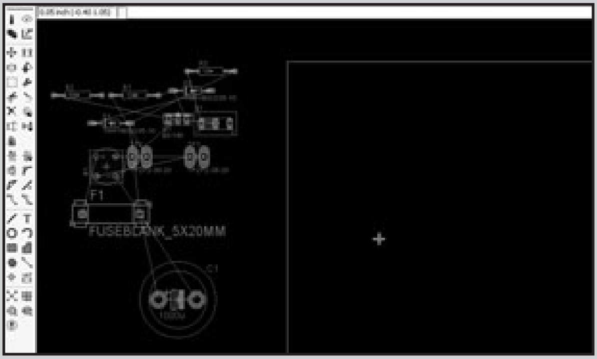

Ahora, se encuentra en el modo **`Board`**. En él, podrá observar, por un lado, los componentes y, por el otro, la placa a diseñar. Lo que tiene que hacer es mover los componentes dentro de la placa. Para ello, utilice la herramienta **`Move`**. Las líneas amarillas indican las conexiones que se transformarán en pistas (ratsnest).

www.FreeLibros.me

ELECTR_Cap4_xxx_xxx_ajustado.qxp 04/07/2011 16:39 PÆgina 102

PASO A PASO /2 (cont.)

8

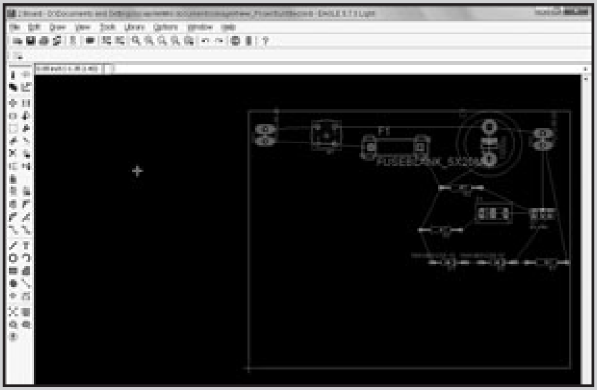

Queda acomodar de manera óptima los componentes. El posicionamiento deberá responder a la manera natural de seguir las pistas, lo cual se ve en el ratsnest, agrupando de forma de minimizar sus vueltas y recorridos. Luego, podrá comenzar a realizar el ruteo.

9

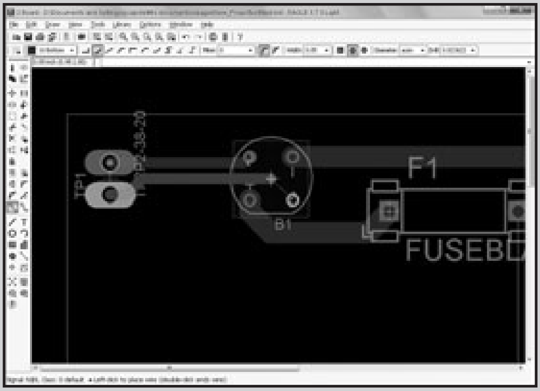

Con la herramienta **`Route`**, haga clic sobre las líneas amarillas, que servirán como guías para el trazado de las pistas. El zoom puede resultar útil para observar mejor la tarea. Debe rutear con pistas gruesas aquéllas que, por ejemplo, transportan mayor corriente.

www.FreeLibros.me

PASO A PASO /2 (cont.)

10

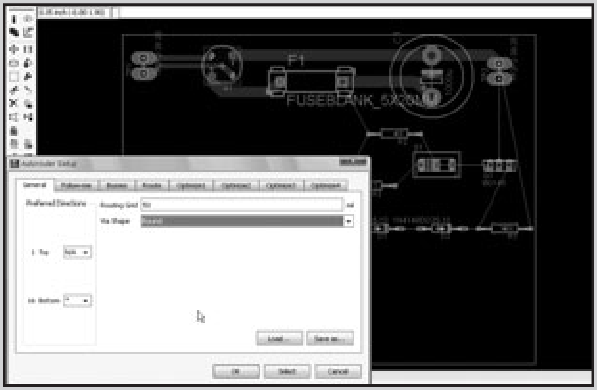

F1
FUSEBLANK_5X20M
Autorouter Setup
General
Preferred Directions
Routing Grid
Via Shape
Round
1 Top
16 Bottom
Load...
Save as...
OK
Select
Cancel

Observará, aquí, algunas de las pistas ruteadas manualmente. A continuación, invoque el autorouter **`Auto`** mediante la barra de herramientas a la izquierda y, en sus opciones, indique que solo use el lado soldadura.

11

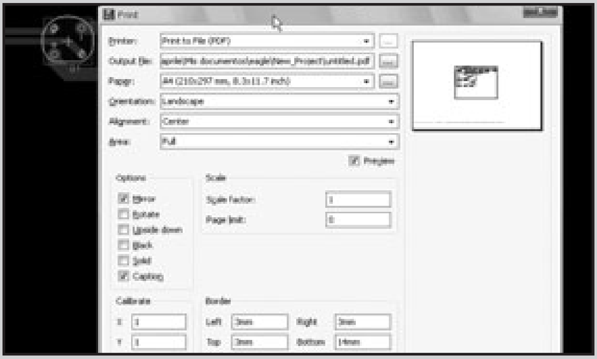

Print
Printer:
Print to File (PDF)
Output file:
Paper:
A4 (210x297 mm, 8.3x11.7 inch)
Orientation:
Landscape
Alignment:
Center
Area:
Full
☑ Preview
Options
☑ Mirror
☐ Rotate
☐ Upside down
☐ Black
☐ Solid
☑ Caption
Scale
Scale factor:
1
Page limit:
0
Calibrate
X
1
Y
1
Border
Left
3mm
Right
3mm
Top
3mm
Bottom
14mm

Teniendo todo listo para imprimir, vaya a la barra de herramientas superior y elija **`File`**/**`Print setup`**. Se abrirá una ventana en donde podrá modificar el tamaño de la hoja y otras propiedades de impresión. También tiene la posibilidad de elegir la posición en la hoja y de cambiar el color de las pistas dibujadas. Cuando esté todo listo, vaya a **`File`**/**`Print`**.

www.FreeLibros.me

Limitaciones de la **versión freeware**

La versión freeware de EAGLE es suficiente para la confección de un circuito PCB, sin embargo, no tiene toda la potencia de la versión comercial.

Explicamos, anteriormente, cuáles son los usos que podemos darle al programa EAGLE en su versión freeware, es decir, sin costo y con uso ilimitado. En general, cuenta casi con la misma funcionalidad que la versión de pago, salvo algunas excepciones que veremos a continuación.

- El área de las placas que puede ser utilizada es de solo 100 x 80 milímetros.
- Únicamente puede crearse una sola hoja dentro del editor de esquema del circuito.
- Solamente podemos emplear dos capas para el diseño de la placa, denominadas top y bottom. No podemos usar capas internas.
- No incluye la posibilidad de soporte técnico vía telefónica o fax.

De todas maneras, para nuestros propósitos, la versión freeware es muy completa, ya que nos permitirá abordar toda clase de proyectos a nivel educativo y nos será más que suficiente para estudiar los contenidos que se desarrollan en esta obra.

La versión freeware del programa solamente se puede utilizar con fines educativos, no comerciales

Además de la versión freeware, el programa ofrece una profesional, sin limitaciones y con soporte permanente. Es importante destacar que este programa no tiene una versión en castellano, solamente se lo puede conseguir en inglés. Sin embargo, cuenta con la ventaja de ser compatible con varios sistemas operativos:

- **Microsoft**: Windows 2000, XP, Vista y Windows 7.
- **Linux**: solo compatible con Linux Kernel versión 2.6.
- **Mac**: Mac OS X 10.4

Para más información acerca del programa, podemos visitar el sitio web del fabricante del software (**www.cadsoftusa.com**), donde tendremos la posibilidad de profundizar nuestros conocimientos.

ACTUALIZACIONES

Una de las ventajas que ofrece la versión profesional de EAGLE es la posibilidad de realizar actualizaciones desde su web oficial, en **www.cadsoftusa.com/upload.htm**. Esto es importante, ya que existen sitios no oficiales que ofrecen este programa sin el soporte necesario.

www.FreeLibros.me

También podemos descargarlo desde la sección de Download dentro de la página y acceder a los foros de usuarios del programa, que nos serán de suma utilidad ante cualquier duda o problema que tengamos (**Figura 15**). La dirección del sitio es **www.eaglecentral.ca/forums**.

REGLAS DE RUTEO

La finalidad de las reglas de ruteo tiene que ver con la minimización del área utilizada, la disminución de posibilidades de interferencias y de ruidos, y una apropiada distribución de los componentes. Veamos algunas reglas elementales (**Figura 16**).

- **Posición de capacitores de desacople**: deben colocarse lo más cerca posible de la alimentación de los circuitos integrados que desean desacoplarse.
- **Separación mínima de pistas**: si se trabaja con tensiones bajas, el mínimo debe ser de 0,3 mm, pero tiene que aumentarse para tensiones mayores. Si más de dos pistas son paralelas entre sí, su separación debe ser uniforme.
- **Separación entre los bordes de la placa y las pistas**: como mínimo, de entre 2 y 3 mm.
- **Ancho de las pistas**: entre otros factores, dependerá de la corriente que pasará por ellas. A mayor corriente, mayor ancho de pistas. Las pistas de alimentación deberán tener un mínimo de anchura de 1 o 2 mm, independientemente de la corriente que consuma el circuito.
- **Largo de las pistas**: tenemos que tratar que sean lo más cortas posibles, con el fin de hacer un diseño simple y reducir el tamaño final del PCB.
- **Trazado de intersecciones**: debe evitarse que, en la intersección de dos pistas, se formen ángulos de 90 grados. Lo ideal es 45º.
- **Conectores de entrada y salida**: tienen que colocarse alejados entre sí, en lo posible, ubicados en extremos opuestos, para evitar que se produzca un acoplamiento entre ellos.
- **Líneas de masa**: en caso de que el circuito cuente con una parte analógica y otra digital, debe haber dos líneas de masa diferentes para cada una, las cuales se unirán en un solo punto.

EAGLE permite configurar diversos parámetros para las reglas de ruteo y puede testear su cumplimiento mediante la opción **[DRC]** (*Design Rule Check*).

CadSoft online
Newcomers
Info
Tour
Freeware
Non-Profit
Dealers
Prices
Order
EAGLE Users
New in V5
FAQ
Support
Download
Upload
Forum
Simulation
Prototyping
Boardhouses
Training
Books
Miscellaneous
Awards
Order online

Welcome to CadSoft Computer!

EAGLE

Version 5.7

Now with interactive Follow-me Router!

Take our Guided Tour and discover EAGLE. You will be surprised!

You can test drive EAGLE with our Freeware version!

Call our US office toll free at 1-800-858-8355.

FIGURA 15.

El programa no tiene una versión en castellano, pero sí encontraremos versiones disponibles para diferentes sistemas operativos.

www.FreeLibros.me

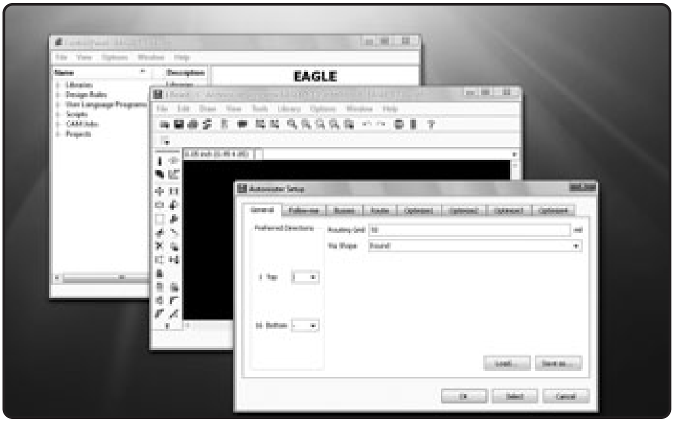

FIGURA 16. **Panel con diversas opciones para configurar el ruteo automático. Los parámetros pueden guardarse en un archivo y, posteriormente, cargarse desde allí.**

Sistema métrico y mils

Haremos una comparación entre dos sistemas de unidades, que nos servirán como herramientas para distintas mediciones de las partes de un circuito.

El sistema métrico es el que normalmente utilizamos porque la mayoría de los países lo eligió para representar medidas de longitud. Utiliza el metro como unidad y es parte de un sistema de unidades mucho más complejo denominado Sistema Internacional de Unidades. Un metro es la longitud de una barra llamada **metro patrón**, que se guarda en la Oficina de Pesos y Medidas en París. Hoy en día, hay formas más modernas de definir el metro, y la más actual es "la distancia que recorre la luz en el vacío durante un intervalo de 1/299792458 de segundo", lo cual se desprende de que la velocidad de la luz ha sido determinada en 299792458 m/s y elimina la necesidad de mantener dicha barra a temperatura constante. Un milímetro es la milésima parte de un metro (mili metro).

AUTORUTEO O RUTEO AUTOMÁTICO

Una de las opciones interesantes es la del autoruteo. Una vez que nosotros hayamos completado el diseño del circuito en la placa, esta opción nos optimizará la posición de los diversos componentes en base a los parámetros establecidos previamente como reglas de ruteo.

www.FreeLibros.me

Hablar de mils significa referirse a la milésima de una pulgada

Muchos de los países de habla inglesa utilizan el llamado sistema anglosajón o imperial. En este sistema, una unidad de longitud es la pulgada, equivalente a 25,4 mm. Cuando hablamos de **mils**, nos estamos refiriendo a la milésima de pulgada, algo así como el equivalente del milímetro en el sistema imperial. Un mil es, entonces, 0,0254 milímetros o 25,4 micrones. Por las actuales dimensiones de los componentes electrónicos y como muchos de los encapsulados tienen sus dimensiones dadas por el sistema imperial, el mil resulta mucho más conveniente que otras alternativas. La equivalencia sería:
1 mil = $25{,}4\times10^{-6}$ m = 1×10^{-3} inches (pulgadas) = $25{,}4\times10^{-3}$ mm = 25,4 µ

Recordemos que, en notación científica, el exponente del número 10 determinará los lugares que se correrá la coma. Si es negativo, hacia la izquierda y, si es positivo, hacia la derecha.

TIPOS DE ENCAPSULADOS MÁS COMUNES

La palabra encapsulado hace referencia a la cavidad en la cual podemos encerrar algún objeto. En nuestro caso, lo que encerramos son componentes electrónicos. Si bien el encierro es mayormente hermético, permitimos una forma de conexión con el exterior, para poder acceder a este y hacer que cumpla su función en nuestro circuito. La finalidad de los encapsulados es: protección del componente, accesibilidad, comodidad y facilidad para el montaje y desmontaje, y disipación de calor, entre otras.

Algunos de los encapsulados más comunes para circuitos integrados son los siguientes:

- **DIP (*Dual Inline Package*)**: es un encapsulado que cuenta con dos filas de pines (o patitas), las cuales se insertan dentro de un zócalo de la placa del circuito. El integrado es más pequeño que el encapsulado exterior. Si bien esto puede ser visto como una desventaja en términos de tamaño ocupado, es cierto, también, que facilita su instalación en una placa.
- **SOP (*Small Outline Package*)**: se diferencian de los DIP en la terminación de los pines y en su menor tamaño. En los SOP, los pines están doblados de forma tal que no se incrustan en un zócalo, sino que se son soldados en la placa, lo cual implica una dificultad mayor para removerlos o instalarlos.
- **PLCC (*Plastic Leaded Chip Carrier*) o QFJ (*Quad-Flat-J-Leg Chipcarrier*):** ambos nombres refieren al mismo tipo de encapsulado. Su forma puede ser cuadrada o rectangular y los pines se disponen a sus cuatro lados. La ventaja principal de estos encapsulados es que se aprovechan los cuatro costados para realizar conexiones.

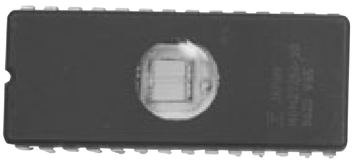

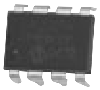

www.FreeLibros.me

Multiple **choice**

1 ¿Cuál es el primer paso que debemos seguir para crear un circuito impreso?

a- Agrupar los elementos.
b- Generar las conexiones.
c- Capturar el diagrama esquemático.
d- Ubicar los componentes.

2 ¿Cómo se llama la herramienta para dibujar nuestro circuito?

a- **Schematic**.
b- **Move**.
c- **Net**.
d- **Erc**.

3 ¿Qué botón nos permite explorar las librerías?

a- **Add**.
b- **Move**.
c- **Net**.
d- **Erc**.

4 ¿Con qué botón se pueden ubicar los componentes en el lugar apropiado?

a- **Add**.
b- **Move**.
c- **Net**.
d- **Erc**.

5 ¿Qué botón nos permite aplicar las reglas de diseño eléctrico?

a- **Add**.
b- **Move**.
c- **Net**.
d- **Erc**.

6 ¿Con qué botón realizamos las conexiones eléctricas entre los elementos?

a- **Add**.
b- **Move**.
c- **Net**.
d- **Erc**.

Respuestas: 1 c, 2 a, 3 a, 4 b, 5 d, 6 c.

www.FreeLibros.me

Capítulo 5

Simulación de circuitos en la PC

En este capítulo, explicaremos qué es la simulación de circuitos y su utilidad en el proceso de diseño.

RedUSERS.com

www.FreeLibros.me

Diseño y simulación

La simulación de circuitos permite evaluar su comportamiento real y predecir sus características

En este capítulo explicaremos qué es la simulación de circuitos, su utilidad en el proceso de diseño y cómo simular un circuito visto en el **Capítulo 2**, utilizando el software ISIS. Un paso necesario en el proceso de diseño de un circuito es asegurarnos de que este se comporte de acuerdo con las especificaciones dadas. Estas pruebas se llevan a cabo modificando la señal de entrada o variando las condiciones de operación, como la temperatura.

Todo el proceso requiere de un tiempo considerable y es costoso, ya que involucra diversos equipos de laboratorio. Afortunadamente, podemos realizar estas pruebas usando la simulación de circuitos por computadora. De este modo, se reducen los costos y el tiempo empleado en las pruebas. Para nuestro caso, se trata, más que nada, de validar, mediante la simulación, el funcionamiento elemental del circuito para, luego, abocarse a la construcción de un prototipo.

Veremos las herramientas de simulación de circuitos electrónicos que nos permitirán observar su funcionamiento sin construirlo físicamente. Los circuitos electrónicos se vuelven cada día más complejos, pero también más eficientes. El proceso de diseño puede llegar a ser muy complicado. En la actualidad, existen sistemas electrónicos que poseen miles y hasta millones de transistores, por lo que el análisis de su funcionamiento sería imposible de realizar manualmente. Por esta razón, se crearon herramientas de simulación de circuitos por computadora, tales como SPICE (*Simulation Program with Integrated Circuit Emphasis*), que permite al diseñador simular el comportamiento real de dichos circuitos y predecir sus características. Gracias a ellas, se acortan drásticamente los tiempos de diseño y se disminuyen los costos (**Figura 1**).

PROCESO DE DISEÑO

El proceso se inicia con el enunciado del problema y las especificaciones, etapa en la cual el diseñador comienza con el desarrollo de una solución abstracta que se transformará en un circuito. El diseño se evalúa mediante un modelo que simula su operación, para verificar la respuesta del circuito frente a diversos estímulos de entrada. El modelo se simula y se revisa las veces que sean necesarias, hasta obtener los resultados deseados. También se evalúan las diferentes opciones de diseño y la manera en que cada una incide en los costos.

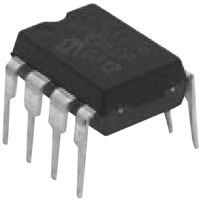

www.FreeLibros.me

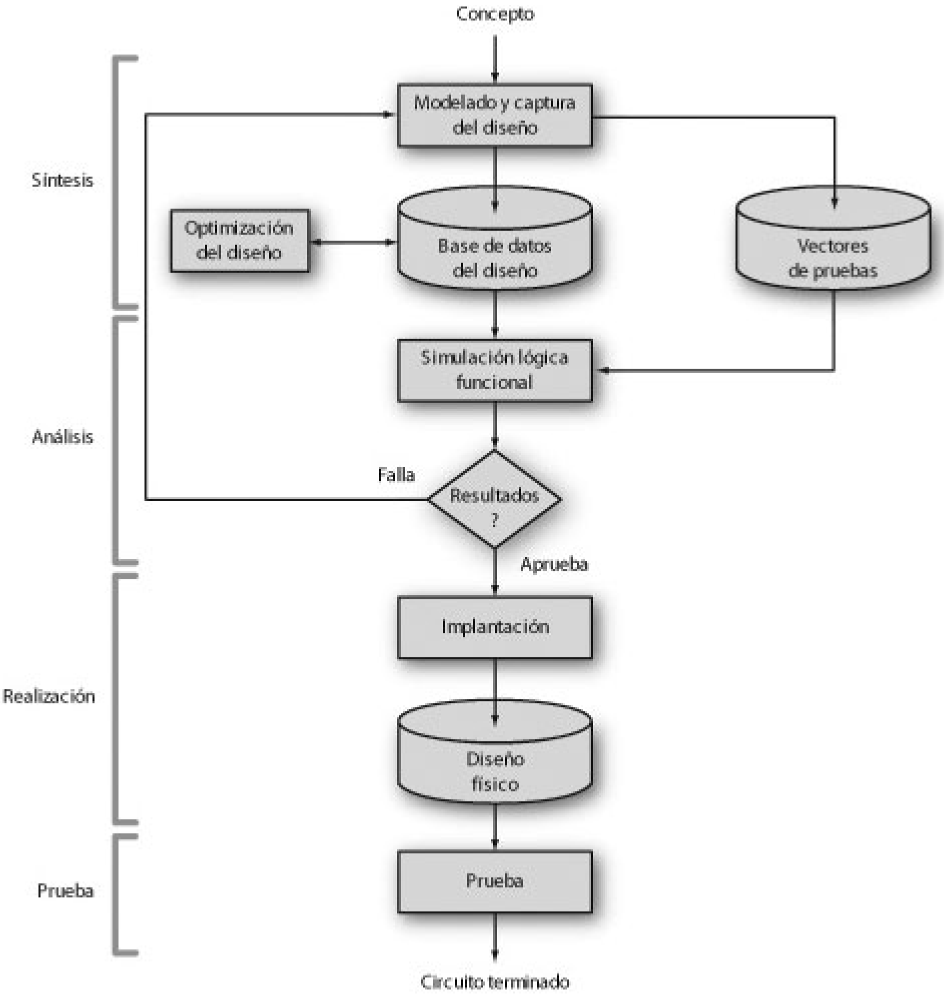
Concepto
Modelado y captura del diseño
Síntesis
Optimización del diseño
Base de datos del diseño
Vectores de pruebas
Simulación lógica funcional
Análisis
Falla
Resultados ?
Aprueba
Implantación
Realización
Diseño físico
Prueba
Prueba
Circuito terminado

FIGURA 1. **Diagrama en bloques del proceso de diseño de un circuito electrónico. Consta, básicamente, de cuatro etapas:** síntesis, análisis**,** realización **y** prueba. **Vemos que la simulación del circuito se encuentra dentro de la etapa de análisis.**

www.FreeLibros.me

Una vez validado el diseño, se desarrolla e implementa el diseño físico. Por último, se prueba el circuito terminado, y los resultados de la prueba se comparan con el comportamiento modelado para detectar posibles fallas (**Figura 2**).

VENTAJAS DE LA SIMULACIÓN DE CIRCUITOS

La simulación de circuitos por computadora nos permitirá realizar las siguientes acciones:

- En primer lugar, podemos observar el funcionamiento del circuito antes de construirlo.
- Realizar mediciones difíciles (debido al ruido eléctrico), no factibles (por no disponer del equipamiento adecuado) o inapropiadas (el circuito de prueba podría dañarse).
- Emplear componentes ideales para realizar una aproximación a un circuito electrónico, proporcionando una respuesta más simple de calcular.
- Cambiar los parámetros de los modelos de los dispositivos semiconductores para efectuar simulaciones en distintas condiciones, como análisis de corriente continua o alterna.
- Por último, podemos observar la dependencia de la temperatura, la generación de ruido, el efecto de capacitancias intrínsecas y las propiedades físicas del disposi-tivo por simular.

MODELADO DE CIRCUITOS ELECTRÓNICOS

Los programas de simulación de circuitos electrónicos utilizan modelos matemáticos para predecir el comportamiento real de un dispositivo o circuito.

El modelo es la representación de un sistema con un grado de abstracción determinado, en tanto que la abstracción es el proceso por el cual se aísla conceptualmente una propiedad de un objeto: comportamiento, entrada/salida y estructura.

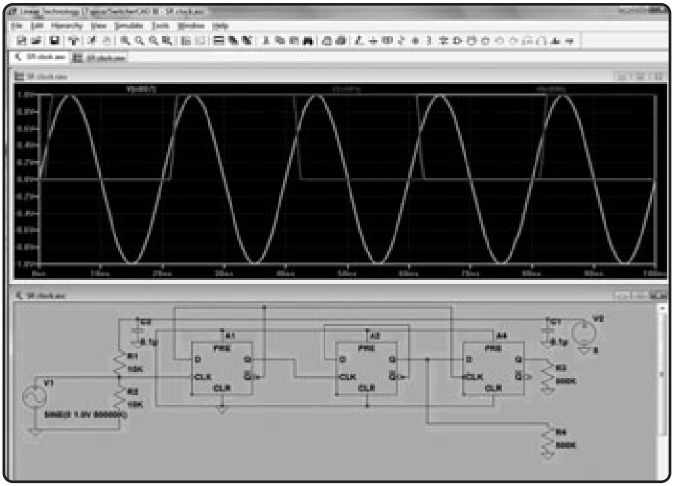

FIGURA 2.
Captura de un software de simulación. Cuando no disponemos de las herramientas de medición, como el osciloscopio o el generador de funciones, la simulación resulta de gran utilidad.

www.FreeLibros.me

> El modelo se simula y se revisa las veces que sean necesarias

Para realizar el modelo matemático, un dispositivo es reemplazado por un bloque que contiene expresiones matemáticas capaces de describir su funcionamiento. Estas pueden estar, por ejemplo, en función de la corriente o de las tensiones del circuito. Un modelo matemático muy utilizado es el de la función transferencia, en el cual, a través de un cociente, se relaciona la respuesta de un sistema a una señal de entrada. A partir de los modelos matemáticos, las suposiciones en el comportamiento de cada componente y sus interconexiones (nodos), el programa realiza la simulación (**Figura 3**).

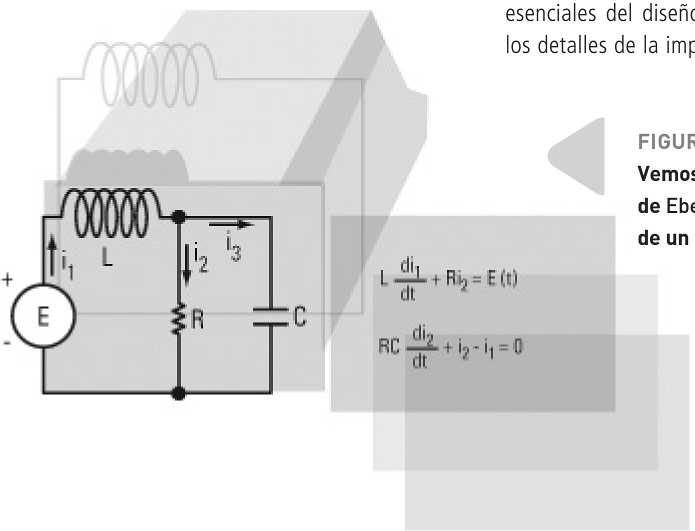

Simulación lógica **funcional**

En este apartado veremos en qué consiste el proceso de simulación y cómo el software emplea los modelos matemáticos de los componentes para realizarlo. La simulación de un circuito electrónico tiene tres propósitos: la verificación lógica, el análisis de desempeño y el desarrollo de pruebas. El proceso de diseño comienza con un modelo abstracto que solo describe el comportamiento o la función deseada del circuito, sin especificar sus componentes reales. En esta parte se utiliza la simulación funcional, que es el proceso de simular la relación entre las entradas y las salidas del circuito sin hacer referencia a la implementación, es decir, reflejando solo la funcionalidad de sus módulos. Esto permite deducir las características esenciales del diseño, sin tener que ocuparse de los detalles de la implementación.

FIGURA 3.
Vemos aquí el modelo matemático de Ebers Moll **y sus ecuaciones de un** transistor TBJ **de** unión NPN.

www.FreeLibros.me

LA LISTA DE RED

En los programas de captura de esquemático, el modelo del circuito es tomado por el simulador en forma de lista de red (*netlist*). La **netlist** es un archivo de texto generado por el software de captura de esquemático que posee información relativa a los dispositivos integrantes del diseño y sus interconexiones.

La netlist se basa en el proceso de reemplazo de todos los módulos del circuito no primitivos, por las redes de elementos primitivos que representan. Se denomina componente primitivo a aquel que no se forma con elementos menores. De esta manera, cada dispositivo se representa por un símbolo gráfico, que se usará para dibujar el esquema, y un modelo asociado, que describe su comportamiento (**Figuras 4** y **5**).

La simulación lógica consiste en simular los componentes del circuito que se expresan en ecuaciones

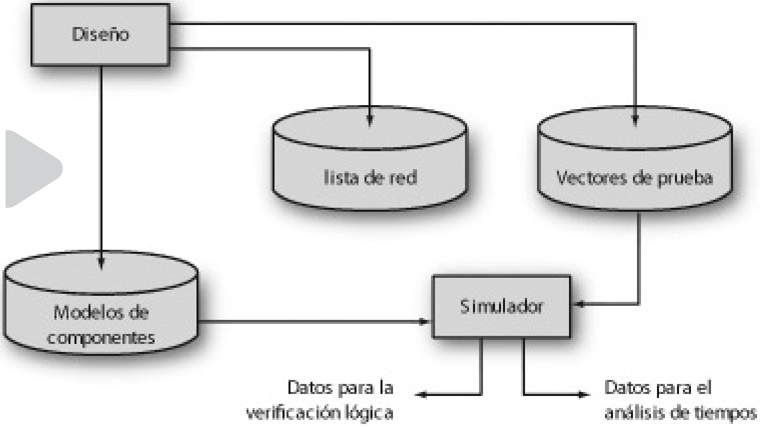

FIGURA 4. **Un diagrama en bloques del entorno del proceso de simulación lógica.**

SIMULACIÓN DE COMPONENTES DIGITALES

En los circuitos que involucran componentes digitales, la información se utiliza para realizar la simulación lógica. Esta simula los componentes del circuito que se expresan en términos de ecuaciones lógicas, utilizando los modelos obtenidos en las librerías del software.

www.FreeLibros.me

ELECTR_Cap5_xxx_xxx.qxp 04/07/2011 16:40 Página 115

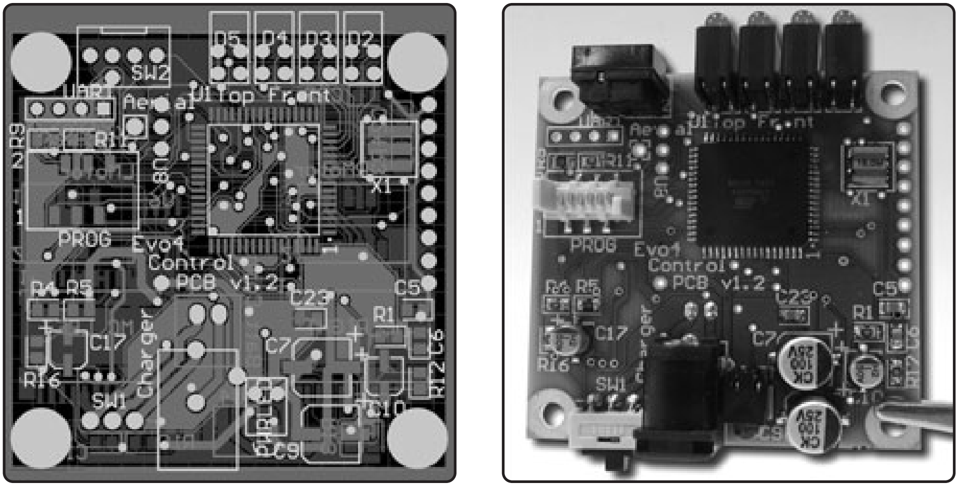
D5
D4
D3
D2
SW2
UART
Aerial
U1Top Front
R9
R11
X1
PROG
Evo4
Control
PCB v1.2
R4
R5
C5
C23
R1
C17
C7
R16
Charger
C6
R17
SW1
C10
C9
CK 100 25V

FIGURA 5. **En estas dos imágenes, vemos el modelo virtual y físico del mismo circuito.**

PROGRAMAS DE SIMULACIÓN

Los programas de simulación de circuitos para PC parten de un software de captura de esquemas. Se trata de una herramienta gráfica interactiva con la que un diseñador dibuja un diagrama de circuito que debe ser procesado por el motor de simulación. Poseen librerías de los modelos matemáticos que describen a los componentes. Veamos los más populares.

PROTEUS

Es un entorno integrado para el diseño y el desarrollo de circuitos electrónicos. Es de licencia propietaria, pero posee una versión de prueba con limitaciones. Para la elaboración de esquemáticos, cuenta con el software ISIS, que incorpora una librería de más de 6000 modelos de dispositivos digitales y analógicos (**Figura 6**). Una vez dibujado el diagrama, podemos hacer la simulación a través del módulo **ProSPICE**, basado en el conocido estándar de simulación **SPICE** (**Figura 7**). A diferencia de los tradicionales simuladores de circuitos basados en gráficos, ProSPICE ofrece una completa animación interactiva. Permite emplear componentes interactivos tales como LEDs, displays de siete segmentos y motores.

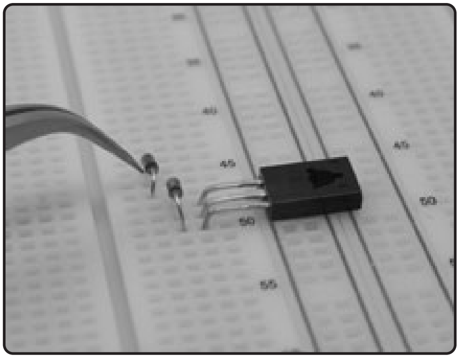
40
45
50
55
40
45
50

www.FreeLibros.me

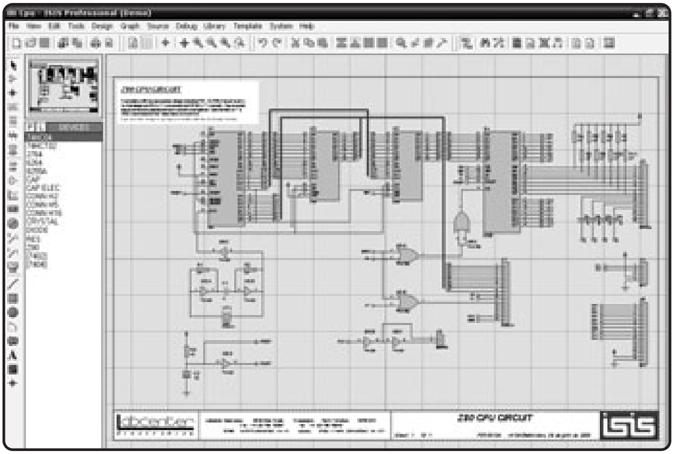
File View Edit Tools Design Graph Source Debug Library Template System Help
Z80 CPU CIRCUIT
isis

FIGURA 6.
Captura de ISIS**, el editor de esquemáticos de** Proteus**. En cuanto a la edición, el usuario puede seleccionar el ancho de las líneas, los estilos de relleno, los colores y las fuentes.**

Incluye, además, varios instrumentos virtuales: amperímetros, voltímetros, osciloscopios y generadores de señales, que funcionan de manera similar a los que se pueden encontrar en el mundo real.

Para esquemas electrónicos grandes, ProSPICE permite la simulación parcial de un diseño, para no tener que dividir y probar un esquema por partes. Una característica que se destaca en él es la inclusión de modelos de simulación **VSM** (*Virtual System Modeling*). Esta técnica permite la simulación completa de esquemas electrónicos con microcontroladores, incluida toda la electrónica asociada y con velocidades cercanas al tiempo real.

ORCAD

El paquete de diseño y desarrollo de circuitos electrónicos del software **OrCAD** incluye el módulo **PSPICE A/D** para los procesos de simulación. Es también de licencia propietaria. Mediante el programa **OrCAD Capture** es posible dibujar el esquemático del circuito para luego simularlo en **PSPICE A/D** (**Figura 8**). Incluye simulación de componentes, tanto electrónicos como digitales. El software de captura de esquemático posee un servicio opcional llamado **CIS** (*Component Information System*), el cual otorga acceso a las librerías de componentes de las compañías de semiconductores más importantes.

El módulo PSPICE A/D posee una gran cantidad de modelos, tales como **IGBTs** (*Insulated Gate Bipolar Transistor*), conversores A/D y D/C y moduladores de ancho de pulso. Además, permite diseñar y generar modelos de transformadores e inductores. Tiene integración con el software de simulación **Simulink**, de **Matlab**.

El módulo PSPICE A/D posee una gran cantidad de modelos

www.FreeLibros.me

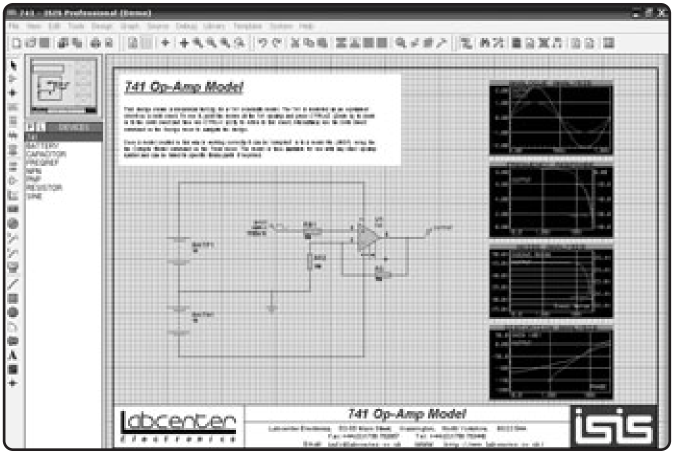
741 Op-Amp Model
labcenter electronics
741 Op-Amp Model
isis

FIGURA 7.
Captura de una simulación de ProSPICE, **de** Proteus. **El programa brinda la posibilidad de utilizar modelos de componentes suministrados por los fabricantes.**

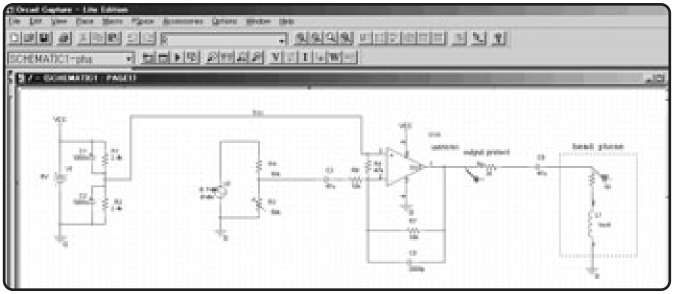
SCHEMATIC1-pfha
SCHEMATIC1 : PAGE1
VCC
output protect
head phone

FIGURA 8.
Captura del programa PSPICE A/D, **que forma parte del paquete** OrCAD. **Colocando voltímetros virtuales sobre el esquemático, visualizamos las señales involucradas.**

ALTIUM DESIGNER

Es un conjunto de programas para el diseño electrónico en todas sus fases. No se trata de un grupo de paquetes sueltos, como una suite, sino de un solo programa. Su versión **Device Intelligence** incluye un simulador **SPICE** compatible con los modelos

OTROS PROGRAMAS I

Además de los mencionados, existen otros programas de simulación. Uno de ellos es **GEDA**, para Linux, de licencia GNU, que actualmente no cuenta con compatibilidad para Windows. Utiliza un motor de simulación llamado **NGSPICE**.

www.FreeLibros.me

habitualmente proporcionados por el fabricante. Permite simular cualquier tipo de dispositivos, incluyendo microcontroladores, aunque está orientado al desarrollo y la simulación de **FPGAs** (*Field Programmable Gate Array*). Es de licencia propietaria y no cuenta con soporte en español (**Figura 9**).

ISIS es un software para el diseño y la simulación de circuitos impresos por computadora

Software **ISIS**

En este apartado describiremos el uso del módulo ISIS, correspondiente a la versión **Proteus 6**, en la confección de esquemas circuitales. El software ISIS forma parte del paquete de productos de software Proteus, de la empresa Labcenter Electronics.

Se trata de una aplicación CAD (*Computer Aided Design*) que ofrece una solución completa para el diseño, la simulación y la elaboración de circuitos impresos. El paquete Proteus está compuesto de los siguientes módulos:

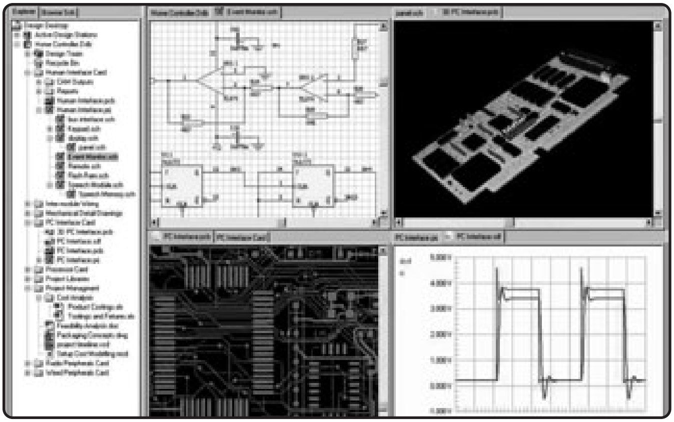

FIGURA 9.

Captura del software Altium Designer. **Por un lado, vemos el esquemático y las señales involucradas. Por el otro, las pistas de** PCB **y la** placa **en tres dimensiones.**

OTROS PROGRAMAS II

Electronic Workbench Multisim, de la empresa **National Instruments**, es muy popular. Su estética se asemeja a la de un laboratorio virtual y se utiliza frecuentemente como una herramienta para la enseñanza y el aprendizaje de la teoría de circuitos electrónicos.

www.FreeLibros.me

ELECTR_Cap5_xxx_xxx.qxp 04/07/2011 16:40 PÆgina 119

- **ISIS** (***Intelligent Schematic Input System***): diseña los esquemáticos de los circuitos.
- **VSM** (***Virtual System Modeling***): es el módulo de simulación funcional de los circuitos.
- **ARES** (***Advanced Routing Modeling***): se ocupa del diseño de los circuitos impresos (PCBs).

ENTORNO ISIS

El entorno ISIS es una herramienta gráfica de diseño de esquemáticos para circuitos impresos, que también incorpora funciones de simulación. Como podemos imaginar, ya sea para simular con el entorno **VSM** o para confeccionar un impreso con **Ares**, primero necesitamos crear un esquemático a partir de los componentes seleccionados para el diseño.

También es posible crear nuestros propios componentes, asignando pines y seleccionando encapsulados para el diseño de impresos, así como su modelización para simulación (**Figura 10**).

Otra característica importante es que se cuenta con la posibilidad de gestionar proyectos jerárquicos. Es decir, que un componente creado puede representar un circuito completo en su interior, conformando un proyecto con varios niveles jerárquicos. Además, admite anidar múltiples hojas de esquemáticos. Resulta cómodo trabajar de esta manera con proyectos de mayor complejidad. Por último, con ISIS podemos simular nuestros diseños en tiempo real mediante su menú de funciones **Debug**, de la barra de herramientas, o desde la barra de simulación ubicada en la parte inferior de la pantalla.

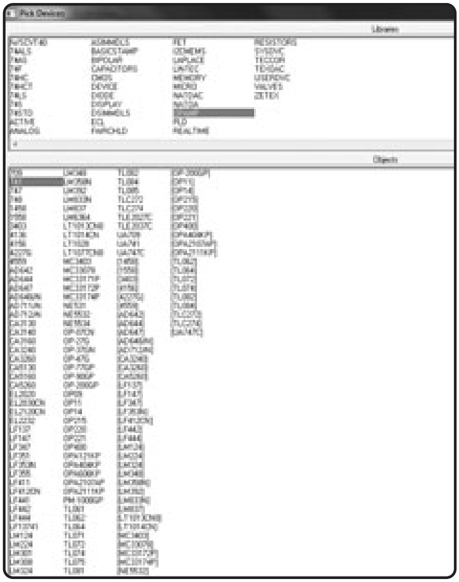

FIGURA 10. Vemos aquí el conjunto de librerías que ofrece la suite para simulación y confección de impresos.

LOS COMPONENTES DE ISIS

Encontramos varios tipos de fuentes de señal, conectores, switches, llaves, relés y demás elementos necesarios en la confección de un circuito. Elementos como voltímetros, amperímetros, osciloscopios y simuladores de interfaces nos brindan una excelente versatilidad.

www.FreeLibros.me

ELECTR_Cap5_xxx_xxx.qxp 04/07/2011 16:40 PÆgina 120

Área **de trabajo**

Vamos a conocer a fondo cómo configurar y movernos en nuestra área de trabajo, la distribución de las barras de herramientas y las ventanas auxiliares para el diseño. Una vez que abrimos ISIS, nos encontramos con la hoja de diseño, que será nuestro lugar de trabajo. Notamos, en primer lugar, la presencia de una **grilla,** cuyo espaciado es configurable, y que nos permitirá trabajar con un "paso" más fino o más grueso a la hora de mover componentes o trazar líneas de conexión. Desde el menú principal podemos realizar las siguientes configuraciones:

- Configuración del tamaño del área de trabajo: **`System/Set sheet sizes`**.
- Cambiar el espaciado de la grilla: **`View/Snap`**.
- Activar/desactivar la visualización de las distintas barras: **`View/Toolbars`**.

Las barras lateral y superior poseen los accesos a las diferentes herramientas para agilizar el trabajo

BARRAS DE ACCESO RÁPIDO

En la parte superior de la pantalla están el **`Menú principal`** y las **`Barras de acceso rápido`** para funciones de visualización, impresión y herramientas varias. Sobre la izquierda, se ubican las ventanas y las barras de botones para el diseño. Las barras de botones y ventanas ubicadas a la izquierda de la pantalla nos ofrecen varias funciones importantes para confeccionar circuitos.

Las **`Barra de funciones`** y **`ventanas`** se ubican en la parte superior de la hoja de trabajo, y contienen funciones generales, opciones y herramientas de proyecto. Veamos en detalle cada una de estas opciones en las **Guías Visuales 1 y 2**.

SOBRE NETLIST

La **Netlist** es el archivo con el detalle de conexión entre los componentes de un diseño que Ares debe interpretar. Su información ayuda con el diseño del PCB al unir automáticamente los componentes del esquemático confeccionado en ISIS, a través de líneas de guía virtuales.

www.FreeLibros.me

ELECTR_Cap5_xxx_xxx.qxp 04/07/2011 16:40 PÆgina 121

GUÍA VISUAL /1

Barra de funciones y ventanas

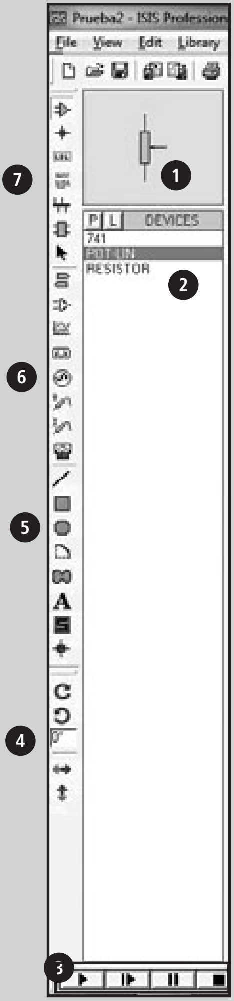

① **`Ventana de vista completa`**: desde aquí podremos visualizar la totalidad del proyecto y movernos sobre toda la hoja, haciendo clic sobre el punto donde queremos centrar la vista.

② **`Ventana de componentes y librerías`**: en ella aparecen las opciones correspondientes a cada menú de trabajo. Por ejemplo, el menú **`Component`** explora las librerías y muestra los dispositivos del proyecto.

③ **`Barra de simulación`**: contiene las teclas **`Run`**, **`Step`**, **`Pause`** y **`Stop`** para la simulación temporal de los circuitos.

④ **`Barra de rotación de componentes`**: se utiliza para modificar la orientación de los componentes con el fin de facilitar su posicionamiento e inserción en los circuitos que se diseñen. Podemos rotar a derecha e izquierda y espejar los dispositivos en forma vertical y horizontal.

⑤ **`Barra de dibujo`**: permite agregar figuras 2D a nuestros diagramas. También podemos incluir texto con los distintos estilos de formato asignados para referencias, indicadores y dispositivos de las librerías.

⑥ **`Barra de diseño electrónico`**: está destinada a la inserción de dispositivos, tales como generadores de señal de todo tipo, instrumentos de medición para simulación en tiempo real, puntas de medición de corriente/tensión y terminales de conexión entre distintas hojas de trabajo, entre otras.

⑦ **`Barra de modos de trabajo`**: esta barra contiene algunas de las opciones más utilizadas. El botón **`Component`**, para listado de los componentes del proyecto y la exploración de librerías; la opción **`Wire Label`**, para darles nombre a las líneas de conexión; **`Buses Mode`**, para la inserción de buses, y el modo instantáneo de configuración de los dispositivos.

www.FreeLibros.me

GUÍA VISUAL /2
Barras de acceso rápido

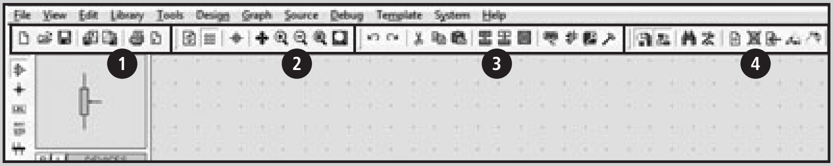
File View Edit Library Tools Design Graph Source Debug Template System Help
1
2
3
4

1. **`Barra de manejo de archivos`**: aquí podemos encontrar las funciones típicas para crear un nuevo proyecto, guardar información, importar y exportar, y las opciones de impresión.
2. **`Barra de visualización`**: ofrece las opciones de zoom, centrado de la hoja de trabajo, vista de hoja completa y selección del área por visualizar. También podemos habilitar/deshabilitar la grilla y el botón de actualización de vista.
3. **`Barra de edición`**: incluye las funciones de **`Undo`** y **`Redo`**, **`Copy`**, **`Cut`** y **`Paste`** de elementos individuales. También, las opciones de copiado, movimiento, desbloqueo y eliminación de objetos agrupados (opciones tagged). Además, vemos la función de conversión de un grupo en elemento de librería y el botón **`Pick/Update Device Symbol`**, para búsqueda de componentes.
4. **`Barra de proyecto`**: incluye el botón de búsqueda de dispositivos por propiedades internas (**`Matched property`**) y todas las opciones relacionadas con proyectos multihoja y jerárquicos: retorno y cambio de hoja de diseño, movimiento por la estructura jerárquica y eliminación de hojas. También están los botones de acceso rápido a la confección del reporte de materiales (**`Bill of materials`**), chequeo de conexión eléctrica (**`Electrical rules check`**) y generación de la Netlist para el diseño de impreso en Ares.

DESPLAZAMIENTO POR EL ÁREA DE TRABAJO

Para movernos sobre el área de trabajo disponemos de varias opciones, que detallamos a continuación:

- **`Herramienta Re*centre Display`**
  Es el botón marcado con una **X** que se encuentra en la barra de acceso rápido. Al presionarlo, el cursor toma la forma de una mira telescópica. Moviéndonos hasta el lugar donde centraremos la vista, presionamos el botón izquierdo para realizar el trabajo.

- **`Herramienta Zoom`**
  Utilizando el scroll del mouse, podemos acercar o alejar la vista de la región centrada en la posición del puntero. También podemos hacerlo con la

www.FreeLibros.me

opción **Zoom to Area**, de la solapa **View** del menú principal, o desde la barra de acceso rápido, con la opción **View selected Area**.

- **Ventana de vista completa**
  La ventana de vista completa muestra, en forma minimizada, el aspecto de todo el diseño. Haciendo clic con el puntero del mouse sobre la zona donde queremos centrar la visualización, lograremos movernos por la hoja. El recuadro verde muestra la zona visible en la hoja de trabajo. Esta puede modificarse haciendo zoom con el scroll del mouse o mediante **View selected Area**, de la barra de visualización, en la parte superior (**Figura 11**).

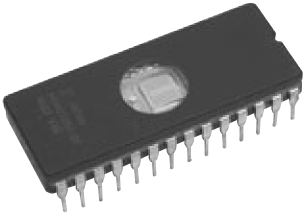

- **Tecla Shift y mouse**
  Si combinamos esa tecla con el movimiento del mouse, lograremos un desplazamiento muy veloz. Esta combinación funciona cuando tenemos maximizada la visualización y el área de trabajo no entra en la pantalla del monitor. Es la forma más rápida de movernos por el proyecto.

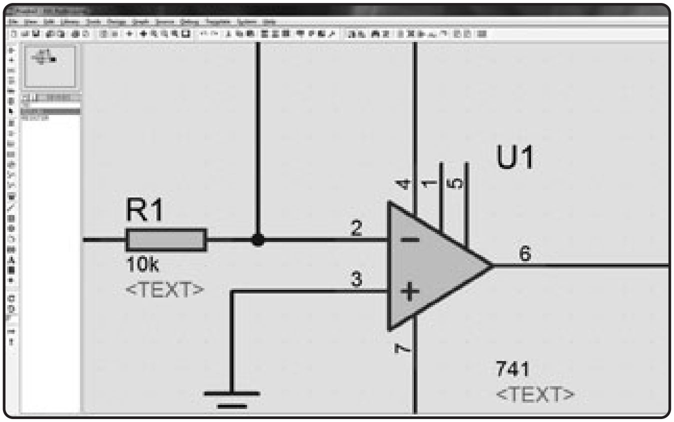

FIGURA 11. Observamos un recuadro de color verde en la ventana de vista completa, que demarca la zona visualizada en la hoja de trabajo.

MEDICIÓN DE DISTANCIAS

Es posible medir distancias relocalizando el origen de la hoja por medio del botón Enable/Disable Manual Origin Specification, de la barra de visualización de acceso rápido ubicada en la parte superior de la hoja.

www.FreeLibros.me

ELECTR_Cap5_xxx_xxx.qxp 04/07/2011 16:40 PÆgina 124

Diseño **en ISIS**

Luego de explorar el área de trabajo, pasaremos a explicar el proceso de inserción de componentes desde las librerías de ISIS y las distintas formas de hacerlo.

Este programa cuenta con una gran librería de componentes donde, además de la forma esquemática tradicional, se muestra la forma física o footprint del elemento seleccionado. Además, ISIS ofrece la aplicación **ProSPICE,** que permite el uso de instrumentos virtuales para la visualización del comportamiento de las señales de los circuitos.

Para la construcción del esquemático del amplificador inversor, vamos a seleccionar los componentes dentro de las librerías de ISIS. Luego, los colocamos en la ventana de trabajo y realizamos la unión entre ellos por medio de las pistas. Finalmente, dejamos conectado un **osciloscopio** para visualizar la señal de salida.

LIBRERÍAS DE COMPONENTES

ISIS nos ofrece unas completas librerías, no solo para confección de impresos, sino también para simulación (**Figura 12**). Entre ellas, podemos encontrar componentes estándar, como resistencias, capacitores, transistores, diodos, etc. Además, se incluyen diversos integrados clasificados por fabricante: microcontroladores, memorias, amplificadores operacionales, **ICs** varios y **CPLDs**, entre otros.

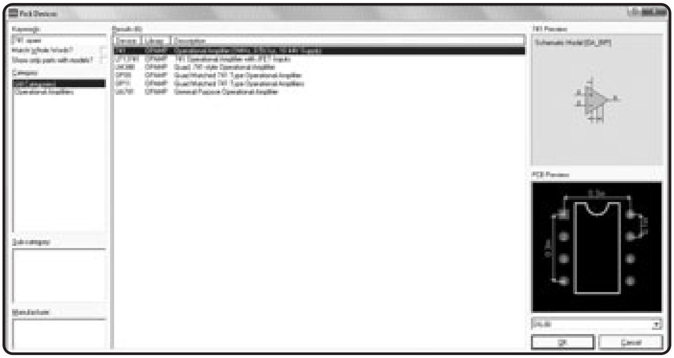

FIGURA 12.
A la derecha se observa la ventana de librerías, el esquemático del amplificador 741 **y su correspondiente** footprint **para el diseño de impresos.**

ESQUEMÁTICO Y FOOTPRINT

El esquemático hace referencia al símbolo normalizado que se utiliza para representar gráficamente componentes electrónicos: el footprint del componente electrónico, la forma física que este ocupará en una placa electrónica, y la disposición de perforaciones y pads.

www.FreeLibros.me

> ProSPICE permite el uso de instrumentos virtuales para visualizar las señales de los circuitos

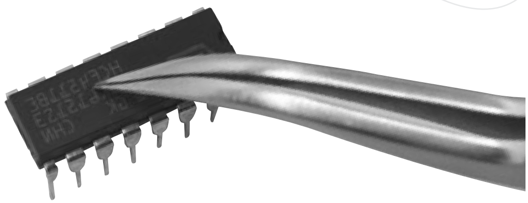

INSERCIÓN DE COMPONENTES

Desde la barra de modos de trabajo, presionamos el primer botón, que corresponde a la opción **`Component`**. Haciendo clic sobre **`Pick Device`**, se abrirá el menú de librerías, con los distintos componentes disponibles.

Allí podemos seleccionar una categoría del menú **`Library`**. Una vez hecho esto, desde la lista **`Objects`** es posible elegir cuáles de los componentes que forman parte de ella utilizaremos.

Otra manera de acceder a las librerías es mediante el botón **`Pick or update device/symbol`** o desde el menú principal, desplegando la solapa **`Library`**, donde se ubica el comando denominado **`Pick Device/Symbol`** (**Figura 13**).

La ventana que se despliega, en ambos casos, corresponde a un buscador de texto por cadena de caracteres. Se debe seleccionar el tipo de componente o dispositivo por buscar y el modo de comparación: coincidencia exacta, solo del comienzo, que contenga el texto, solo fin del texto, etc. Luego, escribiendo en el campo el nombre del componente o texto correspondiente, las coincidencias aparecerán listadas en la ventana. Como en el caso anterior, al hacer doble clic sobre un componente, este se añadirá a la lista del proyecto.

Cerrando las ventanas de selección y volviendo a la hoja de trabajo, solo resta insertar el componente desde la lista de proyecto que fuimos armando. Para hacerlo, desde la lista de proyecto de la ventana de componentes, seleccionamos el dispositivo que

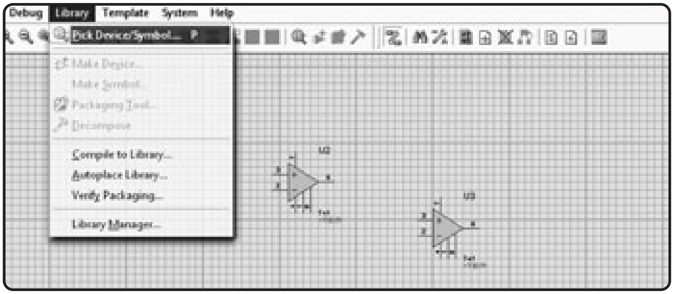

FIGURA 13.
Otra forma de realizar la búsqueda e inserción de componentes es accediendo al menú `Library/Pick Device Symbol`**.**

www.FreeLibros.me

queremos insertar usando el botón izquierdo del mouse. Lo veremos en la ventana principal, pero para insertarlo en la hoja, debemos hacer clic con el botón izquierdo del mouse en algún punto del área de trabajo.

Para eliminar un componente, hacemos doble clic con el botón derecho del mouse. Con el primer clic, el componente se selecciona y se marca en rojo. Si volvemos a hacer otro clic derecho, este se quitará de la hoja, pero no de la lista de proyecto.

Por último, contamos con la posibilidad de rotar el componente y espejarlo a nuestra comodidad, utilizando las opciones de la barra de rotación, sobre la izquierda de la pantalla. Observamos cinco posibilidades: rotar 90º a la derecha, rotar 90º a la izquierda, rotar un ángulo específico, reflejar horizontalmente y reflejar verticalmente.

Podemos utilizarlas antes de posicionar el componente sobre la hoja o una vez que este ya fue insertado, marcándolo en rojo con el botón derecho del mouse y presionando el botón de la rotación deseada (**Figura 14).**

Diseño de un circuito

Estamos en condiciones de diseñar nuestro primer circuito con ISIS. Para el ejemplo, elegimos una etapa de ganancia a amplificador operacional con un sencillo LM741.

Utilizando los conceptos anteriormente descriptos, presionamos el botón **`Components`** y, luego, con el botón **`P`** de la herramienta **`Pick`**, seleccionamos los dispositivos que usaremos en el proyecto. Estos son:

- El **componente 741** de la librería **OPAMP**.
- Un resistor de la librería **ASSIMDLS**.
- Un potenciómetro **POT-LIN** de la librería **Active**.

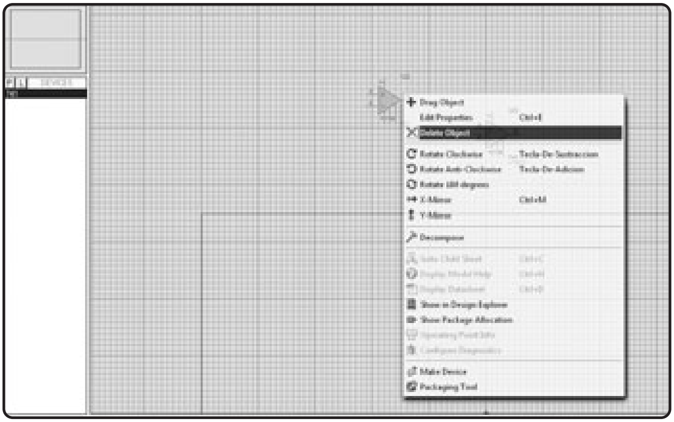
P
L
DEVICES
Drag Object
Edit Properties
Ctrl+E
Delete Object
Rotate Clockwise
Tecla De Sustraccion
Rotate Anti-Clockwise
Tecla De Adicion
Rotate 180 degrees
X-Mirror
Ctrl+M
Y-Mirror
Decompose
Show in Design Explorer
Show Package Allocation
Make Device
Packaging Tool

FIGURA 14.
Con el botón derecho del mouse, es posible seleccionar los componentes para realizar distintas acciones. En este caso, elegimos `Delete Objet`**, para eliminarlo.**

www.FreeLibros.me

ELECTR_Cap5_xxx_xxx.qxp 04/07/2011 16:40 PÆgina 127

Primero debemos configurar nuestros componentes. En este caso, asignamos al resistor y al potenciómetro un valor de **10 K?**. Para hacerlo, podemos utilizar el botón de configuración instantánea de la barra de modos de trabajo. También podemos presionar el botón derecho del mouse sobre el cuerpo del componente y, luego, el botón izquierdo sobre el mismo marcado de rojo.

INSERCIÓN DE MASA Y ALIMENTACIÓN

Desde la barra de diseño electrónico, presionamos el botón **`Inter-sheet Terminal`**. Sobre la ventana de componentes, aparecerán varias opciones, entre ellas, el conector de **`Ground`**. Sobre la ventana, seleccionamos esta opción con el botón izquierdo del mouse. Por último, con el puntero del mouse nos dirigimos hacia algún lugar de la hoja de trabajo y, presionando el botón izquierdo del mouse, insertamos la masa del circuito.

Necesitamos también alimentación para el circuito y para simular la entrada de señal que queremos amplificar/atenuar. Sobre la misma barra de diseño electrónico, ahora presionamos el botón **`Generator`**. Seleccionando la opción **`DC`** y presionando el botón izquierdo del mouse sobre la hoja de trabajo, insertamos dos fuentes de continua al proyecto. Para la señal de entrada, utilizamos el dispositivo **`SINE`**, que es una fuente sinusoidal (**Figura 15**).

Resta configurar cada uno de los dispositivos. Una vez más, marcándolos de rojo con el botón derecho del mouse, abrimos una ventana de configuración. Para las fuentes DC, configuramos los valores de tensión 5 V y -5 V para los pines del LM741 (7) y (4), respectivamente. La fuente sinusoidal tendrá una amplitud de 2 V y una frecuencia de 1 KHz (**Figura 16**).

> Necesitamos también alimentación para el circuito y para simular la entrada de señal

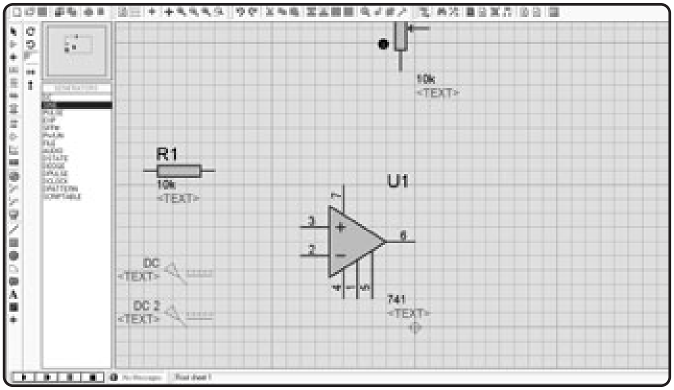

FIGURA 15.
Observamos las dos fuentes de alimentación de continua para el proyecto, denominadas DC1 y DC2.

www.FreeLibros.me

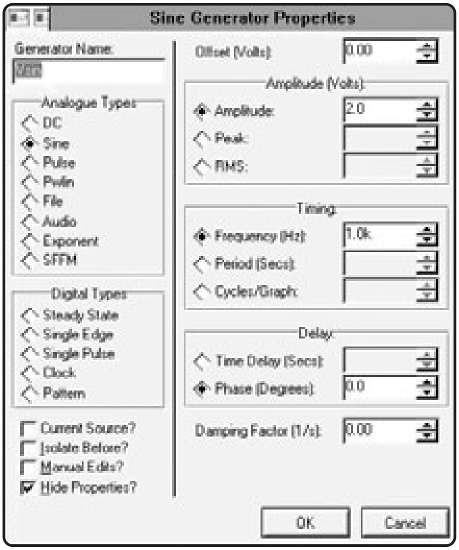

Generator Name:

Analogue Types
○ DC
● Sine
○ Pulse
○ Pwlin
○ File
○ Audio
○ Exponent
○ SFFM

Digital Types
○ Steady State
○ Single Edge
○ Single Pulse
○ Clock
○ Pattern

☐ Current Source?
☐ Isolate Before?
☐ Manual Edits?
☑ Hide Properties?

Offset (Volts): 0.00

Amplitude (Volts):
● Amplitude: 2.0
○ Peak:
○ RMS:

Timing:
● Frequency (Hz): 1.0k
○ Period (Secs):
○ Cycles/Graph:

Delay:
○ Time Delay (Secs):
● Phase (Degrees): 0.0

Damping Factor (1/s): 0.00

FIGURA 16. Aquí se muestra la ventana de configuración de la fuente sinusoidal elegida para la entrada de señal del amplificador operacional. Esta tendrá una amplitud de 2 V **y una frecuencia de** 1 KHz.

CABLEADO DE DISPOSITIVOS

Ahora solo resta unir los componentes para conformar el circuito. Para hacerlo, trazaremos las líneas de conexión con el botón izquierdo del mouse, siempre partiendo de un pin.

Vamos a notar que el cursor cambia a una cruz cuando nos ubicamos con el puntero encima de un pin. Presionando una sola vez el botón izquierdo del mouse, comenzamos a trazar las líneas, desde ese punto hacia el pin destino del circuito. Volviendo a presionar el botón izquierdo sobre el pin destino, se realizará el trazado automático de la línea, siguiendo el camino más corto, pero sin pasar por encima de ningún componente u otra traza. Continuamos así hasta completar las conexiones y lograr que el circuito quede completamente armado.

Los instrumentos virtuales se seleccionan desde la opción INSTRUMENTS

INSTRUMENTOS DE SIMULACIÓN

Realizaremos una simulación de funcionamiento en tiempo real muy sencilla. Para este fin, utilizaremos un osciloscopio que debemos seleccionar desde el botón `Instruments`, de la barra de diseño electrónico (**Figura 17**).

Una vez insertado el osciloscopio, como cualquier otro dispositivo, conectamos su **Canal A** a la salida del amplificador operacional, es decir, en el pin (6). Luego, presionamos el botón **`Run`** de la barra de simulación y corremos nuestro diseño. Automáticamente se abrirá una ventana con la pantalla del osciloscopio virtual. Allí configuramos la sección vertical en 1 V/división, y la horizontal, en 500 us/división.
Manejando el potenciómetro insertado, veremos cómo amplificar la señal para valores mayores a 10 KOhm y atenuarla para valores inferiores, siguiendo la relación de amplificación: * RB / RA. Para este caso, RB = RV1 y RA=R2. Finalmente, presionando el botón **`Stop`** de la barra de simulación, la detenemos (**Figura 18**).

www.FreeLibros.me

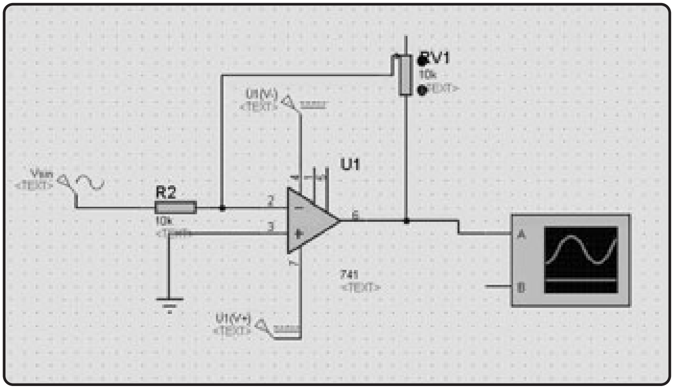

FIGURA 17.
Este es el circuito completo que debemos seguir para confeccionar el ejemplo.

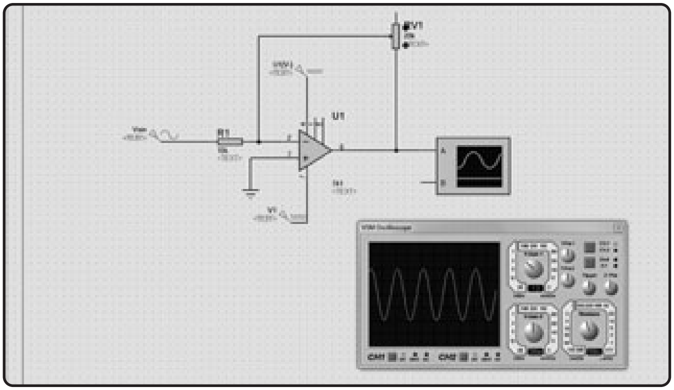

FIGURA 18.
Vemos aquí la simulación correspondiente al instrumento Osciloscopio de ISIS. Observamos cómo la señal de salida se amplifica para valores del potenciómetro mayores a 10 K, mediante la relación (- RV1 / R2).

SIMULADOR DEL CARGADOR DE BATERÍAS

En el **Paso a paso 1** aprenderemos paso a paso cómo realizar la simulación del cargador de baterías que se detalló en el **Capítulo 1**. Daremos nuestros primeros pasos en la simulación de circuitos electrónicos. Este proceso puede realizarse en ISIS, sin necesidad de abrir otro programa. Para efectuar este procedimiento, construiremos el esquemático del cargador de baterías, luego colocaremos los instrumentos de medición para ver su funcionamiento y, finalmente, realizaremos la simulación del circuito, para ver la curva de salida correspondiente.

www.FreeLibros.me

ELECTR_Cap5_xxx_xxx.qxp 04/07/2011 16:40 PÆgina 130

PASO A PASO /1
Simulación del cargador de baterías

1

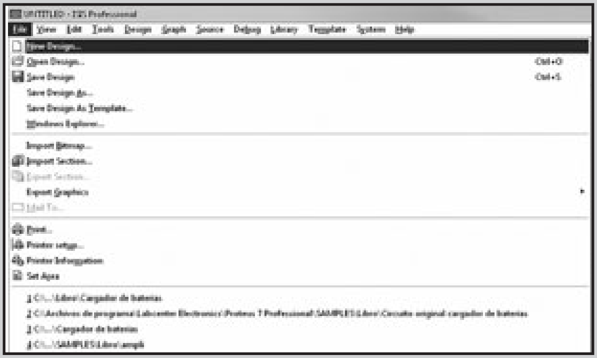
UNTITLED - ISIS Professional
File
View
Edit
Tools
Design
Graph
Source
Debug
Library
Template
System
Help
New Design...
Open Design...
Ctrl+O
Save Design
Ctrl+S
Save Design As...
Save Design As Template...
Windows Explorer...
Import Bitmap...
Import Section...
Export Section...
Export Graphics
Mail To...
Print...
Printer setup...
Printer Information
Set Area
1 C:\...\Libro\Cargador de baterias
2 C:\Archivos de programa\Labcenter Electronics\Proteus 7 Professional\SAMPLES\Libro\Circuito original cargador de baterias
3 C:\...\Cargador de baterias
4 C:\...\SAMPLES\Libro\ampli

El primer paso es crear un nuevo diseño. Puede realizarlo desde la pestaña **`File`**, ubicada en la Barra de menús, o también mediante la herramienta **`New File`**/**`From Default Template`**, que se encuentra en la Barra de archivos. En esta oportunidad, lo hará desde la opción **`File`**/**`New Design`**.

2

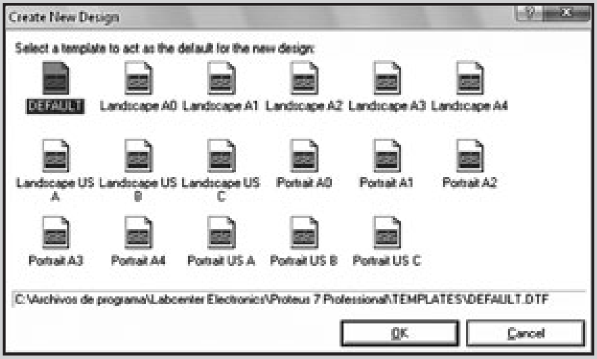
Create New Design
Select a template to act as the default for the new design:
DEFAULT
Landscape A0
Landscape A1
Landscape A2
Landscape A3
Landscape A4
Landscape US A
Landscape US B
Landscape US C
Portrait A0
Portrait A1
Portrait A2
Portrait A3
Portrait A4
Portrait US A
Portrait US B
Portrait US C
C:\Archivos de programa\Labcenter Electronics\Proteus 7 Professional\TEMPLATES\DEFAULT.DTF
OK
Cancel

Al entrar en ese menú, se abre la ventana emergente **`Create New Design`**, donde se ofrecen varias plantillas para realizar el diseño. Estas se utilizan para hacer presentaciones de informes impresos con rótulos, como nombre de los archivos, día de presentación, nombre del diseño, autor, revisión y hora. Para su caso seleccione **`Default`** y presione **`Ok`**.

www.FreeLibros.me

PASO A PASO /1 (cont.)

3

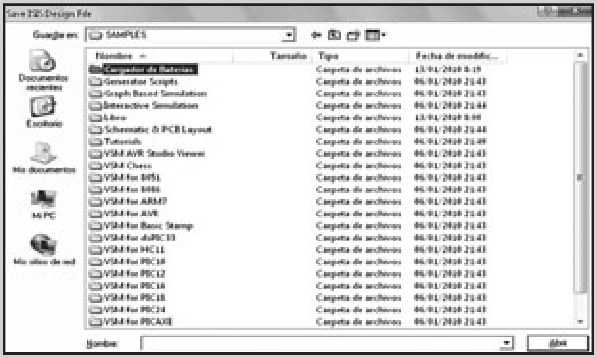
Save ISIS Design File
Guardar en:
SAMPLES
Nombre
Tamaño
Tipo
Fecha de modific...
Documentos recientes
Escritorio
Mis documentos
Mi PC
Mis sitios de red
Nombre:
Abrir

Una vez que ha generado un nuevo diseño, es recomendable crear una carpeta con el nombre del proyecto. Puede recurrir al comando `Save Design`, en la Barra de menús. Para este ejemplo, use `File`/`Save Design`. Aparece la ventana emergente `Save ISIS Design File`, que tiene las herramientas necesarias para crear la carpeta donde alojará el archivo de diseño. Cree la carpeta que se llamará **Cargador de baterías** y, dentro de ella, guarde el diseño; presione **ok**.

4

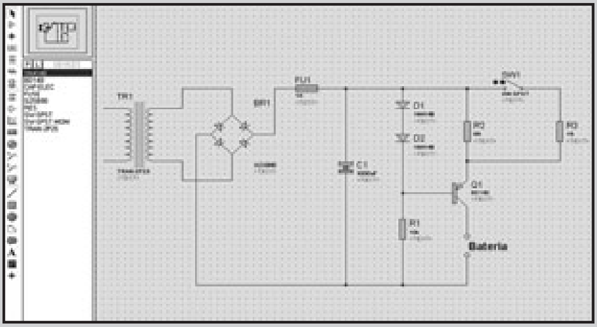
TR1
BR1
FU1
SW1
D1
D2
C1
R1
R2
R3
Q1
Batería

En este paso comparará el circuito original con el que va a simular. En el que simulará, no precisa el fusible, ya que en este caso no habrá ningún exceso de corriente. También va a reemplazar el puente rectificador y a colocar una fuente de corriente continua, lo que le permitirá ver el funcionamiento del cargador.

www.FreeLibros.me

PASO A PASO /1 (cont.)

5

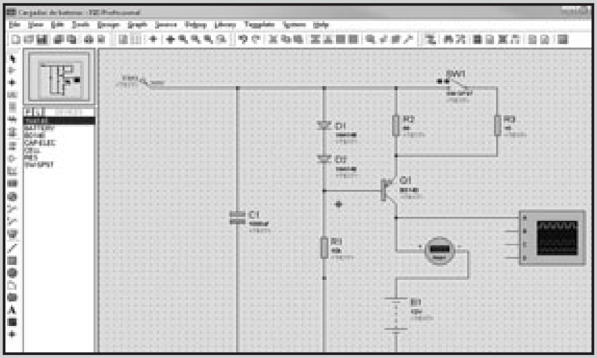

El circuito por simular no está conectado a la red domiciliaria, de modo que puede reemplazar el transformador, el circuito rectificador y el fusible. El puente rectificador se sustituye por una fuente de continua de **24 V**. Coloque una batería de **12 V** y saque el fusible, ya que no tiene modelo de simulación. Agregue el amperímetro y coloque un osciloscopio para visualizar la forma de la señal.

6

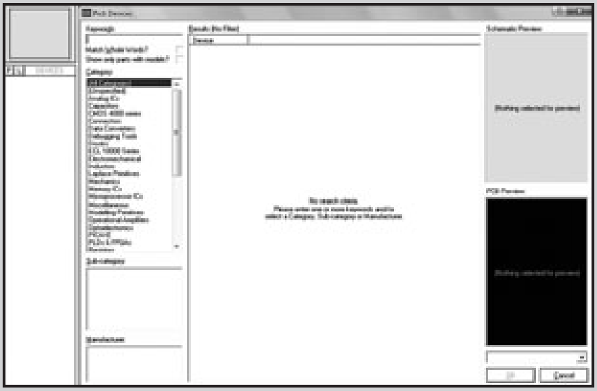

Debe ubicar los componentes que desea volcar. Presione el botón **`P`** para que se despliegue la ventana **`Pick Devices`**, en la cual tiene que buscar los distintos componentes del cargador de baterías.

www.FreeLibros.me

PASO A PASO /1 (cont.)

7

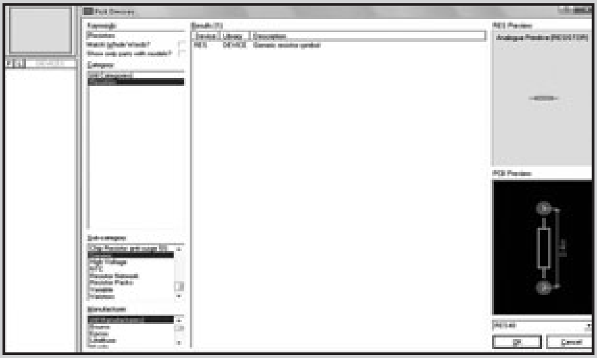

Ahora pondrá los resistores que lleva el circuito. Para hacerlo, coloque la clave **Resistor** en el cuadro de diálogo **`Keywords`** de la ventana **`Pick Devices`**. En **`Pick Devices`**/**`Category`** seleccione **`Resistors`**. En **`Sub-category`** elija **`Generic`**, y en **`Manufacturer`** deje la opción **`All Manufacturers`**. Presione **`Ok`** para colocar el componente en la ventana de trabajo.

8

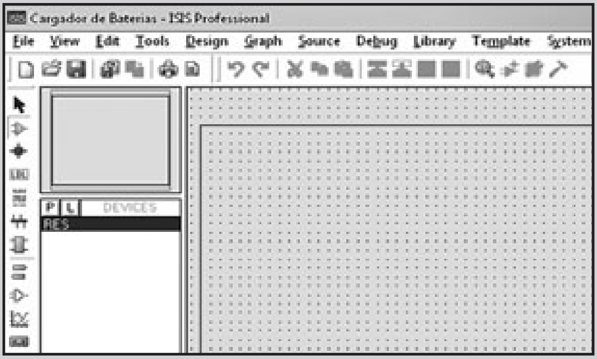
Cargador de Baterías - ISIS Professional
File
View
Edit
Tools
Design
Graph
Source
Debug
Library
Template
System
P
L
DEVICES
RES

El circuito que va a simular tiene tres resistores. Una vez que seleccione un componente –en este caso, los resistores–, se lo carga en la ventana **`Componentes y librerías`**, con lo cual no es necesario volver a elegirlo. Para colocar los resistores que faltan, seleccione **`Res`** en la misma ventana.

www.FreeLibros.me

PASO A PASO /1 (cont.)

9

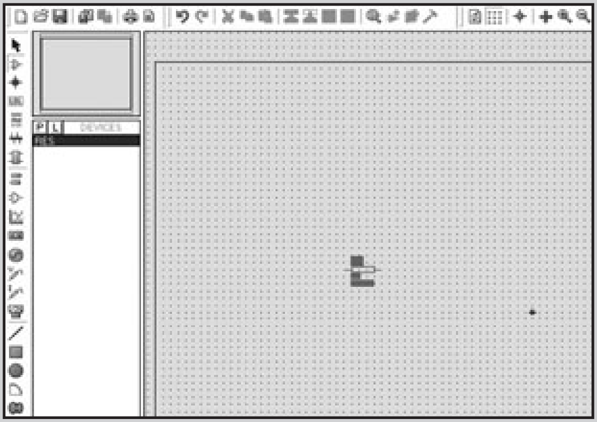
P
L
DEVICES
RES

Haga clic en la ventana principal de trabajo para ubicar el componente. Desplácese con él por la ventana y, una vez que elija dónde ponerlo, haga otro clic para fijarlo.

10

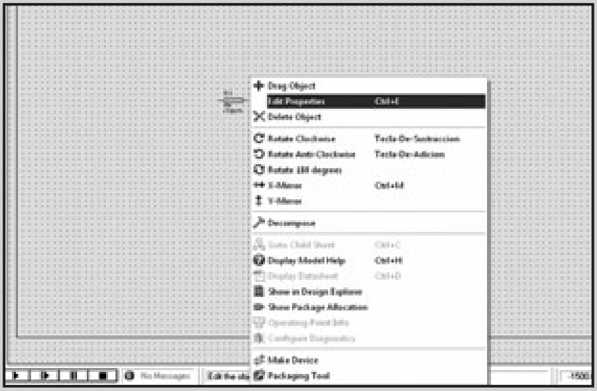
Drag Object
Edit Properties Ctrl+E
Delete Object
Rotate Clockwise Tecla De Sustraccion
Rotate Anti-Clockwise Tecla De Adicion
Rotate 180 degrees
X-Mirror Ctrl+M
Y-Mirror
Decompose
Goto Child Sheet Ctrl+C
Display Model Help Ctrl+H
Display Datasheet Ctrl+D
Show in Design Explorer
Show Package Allocation
Operating Point Info
Configure Diagnostics
Make Device
Packaging Tool
No Messages
-1500

Para darles valores a las resistencias, elija el icono **`Selection Mode`**, en la ventana **`Edición`**. Luego, posiciónese sobre los resistores y haga clic con el botón derecho del mouse. Se abre una ventana emergente donde tiene que elegir **`Edit Properties`**.

www.FreeLibros.me

PASO A PASO /1 (cont.)

11

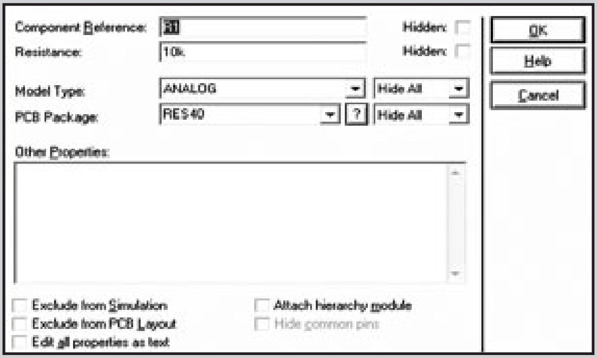

Component Reference: R1 Hidden: ☐
Resistance: 10k Hidden: ☐
Model Type: ANALOG Hide All
PCB Package: RES40 ? Hide All
Other Properties:

☐ Exclude from Simulation ☐ Attach hierarchy module
☐ Exclude from PCB Layout ☐ Hide common pins
☐ Edit all properties as text

OK
Help
Cancel

Aparece la ventana **`Edit Component`**, en la cual puede dar valores a los componentes, como nombre, valor de resistencia, tipo de modelo, tipo de PCB, y otras propiedades. Configure el primer resistor: en la ventana **`Component Reference`** coloque la denominación **`R1`**, y en **`Resistance`** ingrese el valor **10K**. Para **R2** coloque una **`Resistance`** de **66**, y para **`R3`**, de **15**.

12

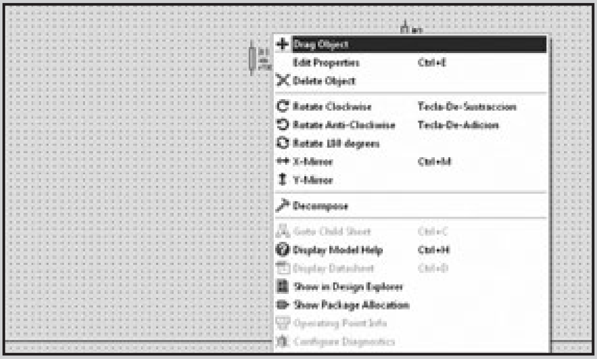

Drag Object
Edit Properties Ctrl+E
Delete Object
Rotate Clockwise Tecla-De-Sustraccion
Rotate Anti-Clockwise Tecla-De-Adicion
Rotate 180 degrees
X-Mirror Ctrl+M
Y-Mirror
Decompose
Goto Child Sheet Ctrl+C
Display Model Help Ctrl+H
Display Datasheet Ctrl+D
Show in Design Explorer
Show Package Allocation
Operating Point Info
Configure Diagnostics

Si necesita mover la resistencia o cualquier otro componente, pase a **`Selection Mode`**, en la ventana **`Edición`**. Luego, haga clic derecho sobre el resistor y, en la ventana emergente, seleccione **`Drag Object`**. Esto permitirá mover el objeto a la posición deseada.

www.FreeLibros.me

PASO A PASO /1 (cont.)

13

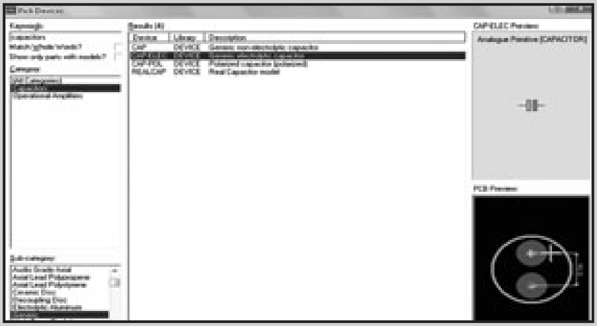

Ahora va a colocar el capacitor, para lo cual debe realizar los mismos pasos que con el resistor. Cuando busque en **`Pick Devices`**, coloque la clave **`Capacitor`** en **`Keywords`**. En **`Sub-category`** ponga **`Generic`** y, en **`Results`**, elija **`CAP-ELEC.`** Luego, presione **`Ok`**. El valor de capacidad será de **1000 ì F** y los pasos para asignarlo son los mismos vistos para darle valores al resistor.

14

El circuito tiene dos diodos **1N4148**. Para encontrarlos, vaya a la ventana **`Componentes y librerías`**, y seleccione **`Pick From Libraries`**. En el cuadro de diálogo **`Keywords`** coloque la clave **1N4148**, y en **`Results`** seleccione **1N4148**. Presione **`Ok`**. Regrese a la ventana para seleccionar el diodo y colocarlo en el área de trabajo.

15

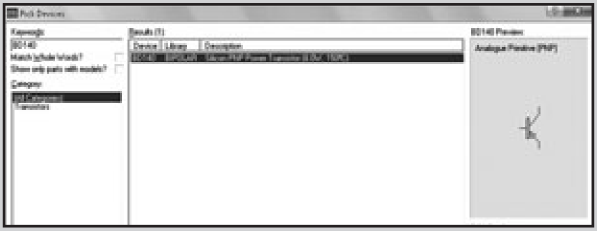

Ubique el transistor BD140, para lo cual tiene que elegir **`Pick From Libraries`** en la ventana **`Componentes y librerías`**. En el cuadro de diálogo **`Keywords`** coloque la clave BD140; en **`Results`** seleccione **`BD140`** y presione **`Ok`**. Coloque el transistor en la ventana de trabajo utilizando los comandos de rotación y reflexión.

www.FreeLibros.me

PASO A PASO /1 (cont.)

16

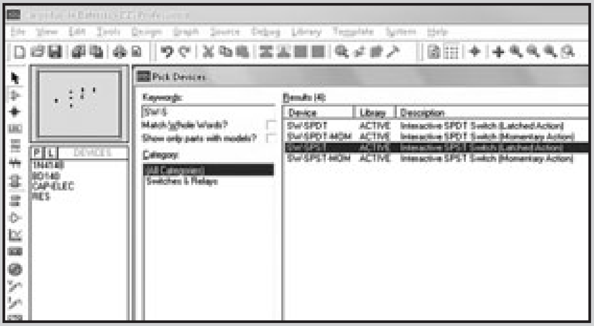
Pick Devices
Keywords:
Match Whole Words?
Show only parts with models?
Category:
(All Categories)
Switches & Relays
P L DEVICES
1N4148
BD140
CAP-ELEC
RES

Results (4):

Device	Library	Description
SW-SPDT	ACTIVE	Interactive SPDT Switch (Latched Action)
SW-SPDT-MOM	ACTIVE	Interactive SPDT Switch (Momentary Action)
SW-SPST	ACTIVE	Interactive SPST Switch (Latched Action)
SW-SPST-MOM	ACTIVE	Interactive SPST Switch (Momentary Action)

Es el momento de colocar el switch. Para hacerlo, diríjase a **`Pick Devices`**/**`Keywords`** y coloque la clave **`sw-spst`**. Debe seleccionar el que tiene la descripción **`Latched Action`** y no el **`Moment Action`**. La diferencia entre ellos es que el primero tiene los botones de bajada y subida, mientras que el segundo tiene un solo botón para ambas tareas.

17

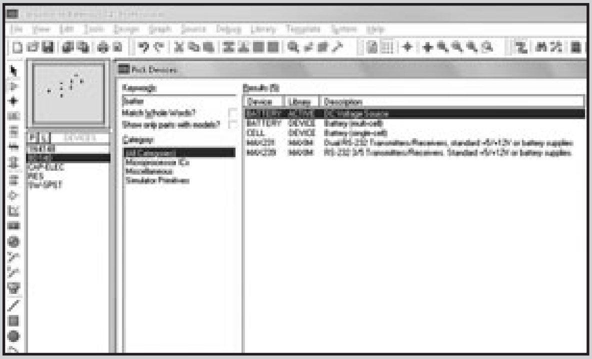
Pick Devices
Keywords:
Match Whole Words?
Show only parts with models?
Category:
(All Categories)
Microprocessor ICs
Miscellaneous
Simulator Primitives
P L DEVICES
1N4148
BD140
CAP-ELEC
RES
SW-SPST

Results (5):

Device	Library	Description
BATTERY	ACTIVE	DC Voltage Source
BATTERY	DEVICE	Battery (multi-cell)
CELL	DEVICE	Battery (single-cell)
MAX231	MAXIM	Dual RS-232 Transmitters/Receivers, standard +5/+12V or battery supplies
MAX239	MAXIM	RS-232 3/5 Transmitters/Receivers. Standard +5/+12V or battery supplies

Ahora pase a la batería. Para encontrarla, vaya a **`Pick From Libraries`**/**`Pick Devices`**. En el cuadro de diálogo **`Keywords`** coloque la clave **`Battery`**; en **`Results`** seleccione **`Battery`** con la descripción **`DC Voltage Source`** y presione **`Ok`**. La batería es **`DC Voltage Source`**, ya que permite cargar parámetros de baterías que hay en el mercado.

www.FreeLibros.me

PASO A PASO /1 (cont.)

18

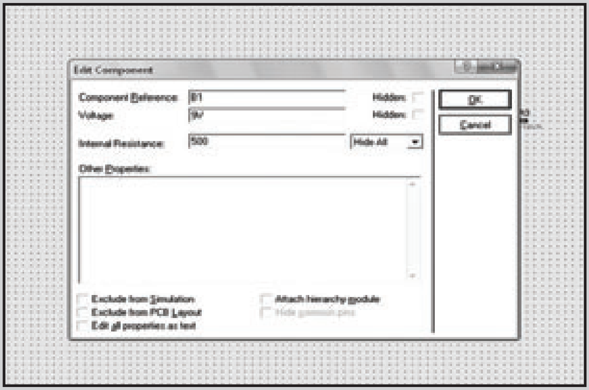
Edit Component
Hidden:
Hidden:
OK
Cancel
Hide All
Other Properties:

Component Reference: B1
Voltage: 9V
Internal Resistance: 500

☐ Exclude from Simulation
☐ Exclude from PCB Layout
☐ Edit all properties as text
☐ Attach hierarchy module
☐ Hide common pins

Para configurar la batería, diríjase a la ventana emergente **Edit Component**. En la casilla **Voltage** coloque (9V); en **Internal Resistance** escriba **500**, sin necesidad de colocar la unidad, ya que, por defecto, se toma ohmio (Ω); y, finalmente, pulse **Ok**.

19

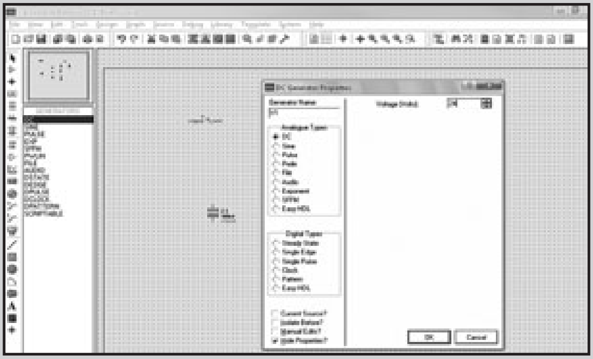
DC Generator Properties
Generator Name:
Voltage (Volts):
OK
Cancel

Generator Name: V1
Voltage (Volts): 24

Ahora coloque la fuente de alimentación, para lo cual debe seleccionar **Generator Mode** en la ventana **Edición**. Finalmente elija **DC** en la ventana **Componentes y librerías**. Para configurar la fuente, vaya a **Edit Component**, en **Generator Name** coloque **V1** y en **Voltage** opte por **24**. Presione **Ok**.

www.FreeLibros.me

PASO A PASO /1 (cont.)

20

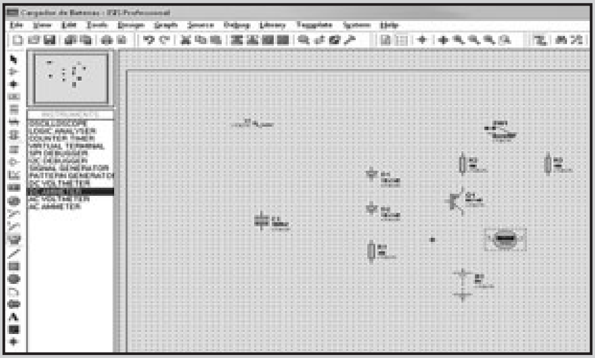

Para medir la corriente que cargará la batería, se necesita un amperímetro que se colocará entre esta y el transistor. Cuando se cierra el interruptor, el circuito trabaja con una corriente mínima de carga de **20 mA**, que se visualizará en el amperímetro, mientras que cuando se abre, esta es próxima a 0 mA. Para encontrar el amperímetro vaya a **`Virtual Instruments Mode`** de la ventana **`Edición`** y, en **`Componentes y librerías`**, elija **`DC ammeter`**.

21

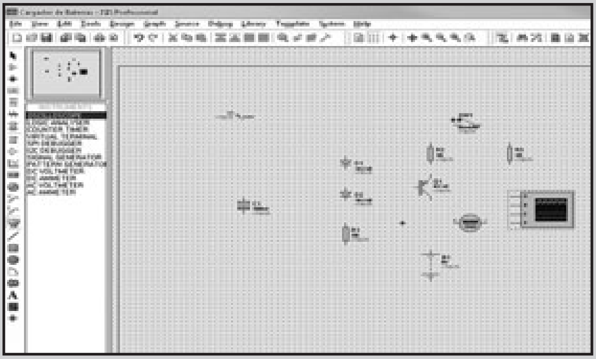

Para observar cómo la batería se va cargando en función del tiempo, utilice el osciloscopio, que debe buscar en **`Virtual Instruments Mode`**. Conéctelo en uno de los pines del circuito, y una el pin del colector del transistor **BD140** al **canal A** de ese elemento.

www.FreeLibros.me

PASO A PASO /1 (cont.)

22

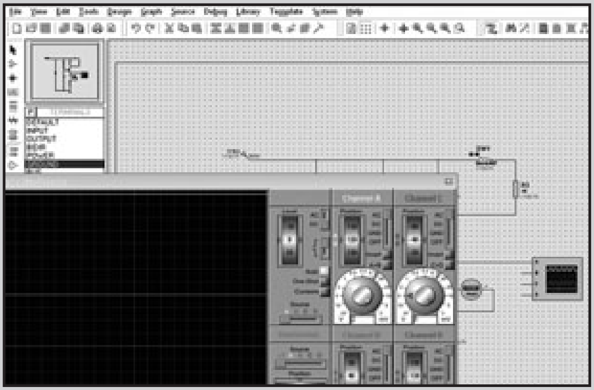

Antes de realizar las conexiones eléctricas, coloque la masa, que está en **`Terminals Mode`**, de la ventana **`Edición`**. Finalmente, seleccione **`Ground`** y póngala en la ventana de trabajo.

23

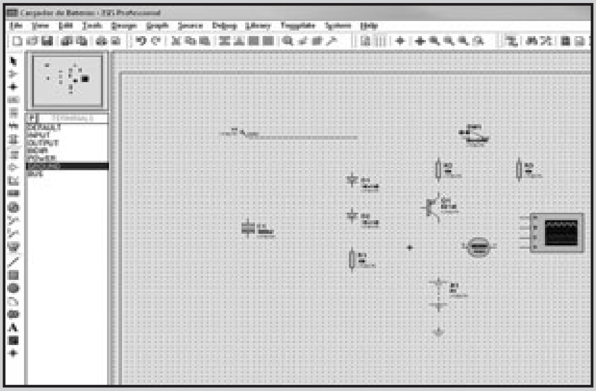

Para conectar los componentes eléctricamente, debe posicionarse en los extremos de ellos. Esto provocará que el cursor en formato de lápiz blanco pase a color verde. Entonces, haga clic sobre cada uno de los extremos que desea interconectar. Para suspender el proceso, pulse **`Esc`**.

www.FreeLibros.me

PASO A PASO /1 (cont.)

24

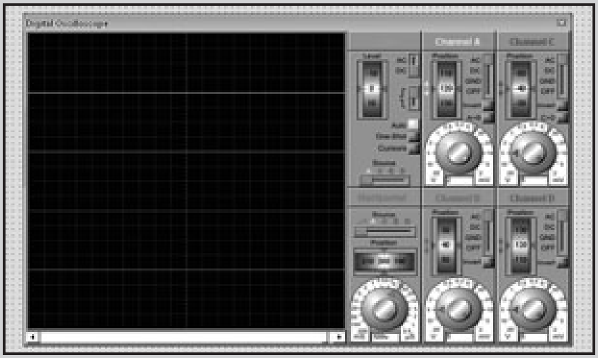
Digital Oscilloscope
Channel A
Channel C
Channel B
Channel D
Horizontal
Level
Position
AC
DC
GND
OFF
Auto
One-Shot
Cursors
Source

Los comandos de simulación se encuentran en la barra del mismo nombre. Esta tiene cuatro comandos: **`Play`**, **`Step`**, **`Pause`** y **`Stop`**. **`Play`** corre la simulación sin pausas, en tanto que **`Step`** trabaja paso a paso en tramos de tiempo predefinidos y permite el uso de herramientas de depuración. Presione **`Play`** para comenzar con la simulación.

25

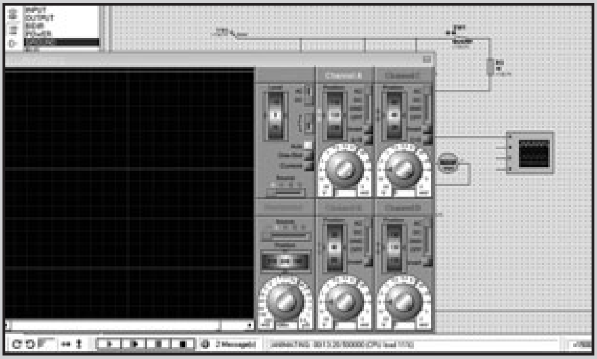
INPUT
OUTPUT
BIDIR
POWER
Channel A
Channel C
Channel B
Channel D
Horizontal

El osciloscopio virtual funciona de modo similar al real. Es por eso que va a configurar el atenuador vertical o amplitud y la base de tiempo horizontal. La amplitud es independiente para cada canal; el A tiene el nombre **Channel A** y dentro de él debe elegir el tipo de señal de entrada que quiera medir. En este caso déjelo en **`AC`**, que es la opción predefinida.

www.FreeLibros.me

Simulación de código con ISIS

Con Proteus podemos simular código ensamblado de microprocesadores para visualizar el funcionamiento de un microcontrolador mientras se ejecuta el código correspondiente.

La simulación de código ensamblado es una de las herramientas que han facilitado el trabajo con los microcontroladores, ya que permite probar su comportamiento antes de implementar estos elementos en la protoboard o en la placa. Este proceso nos ahorra tiempo de desarrollo porque podemos depurar el circuito sin temor a equivocarnos, considerando que todo el desarrollo y la prueba de diseño se realiza en un programa de computadora. Recordemos que VSM (*Virtual System Modeling*) es capaz de trabajar con archivos de código Assembler de algunos de los siguientes fabricantes de microprocesadores: AVR, PIC, 8051, HC11, ARM/LPC200 y Basic Stamp. Además, integra la simulación de circuitos electrónicos tradicionales con modelos de animación de componentes electrónicos y microprocesadores que conforman el circuito.

> La simulación de código ensamblado permite depurar el circuito antes de su implementación real

Es importante destacar que la potencia de Proteus radica en que puede simular en un solo software todo el hardware necesario para un proyecto con microcontroladores. Por lo tanto, no solo posibilita el procesamiento del código Assembler, sino que también nos permite trabajar en tiempo real con su propio hardware, usando Display LCD, LEDs, teclados, puertos RS232, etc.

Todas estas opciones están incorporadas en ISIS a través de **Source** y **Debug** (**Figura 19**), ubicadas en

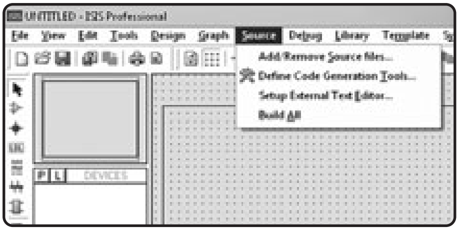
UNTITLED - ISIS Professional
File
View
Edit
Tools
Design
Graph
Source
Debug
Library
Template
Add/Remove Source files...
Define Code Generation Tools...
Setup External Text Editor...
Build All
DEVICES

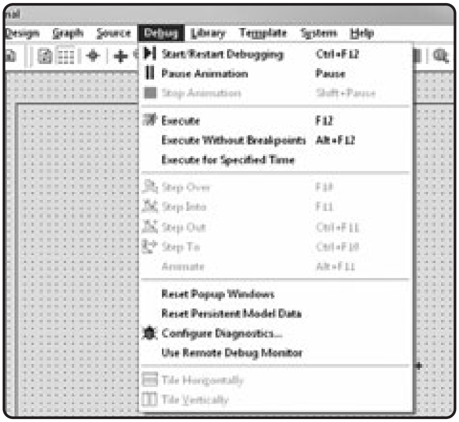
Design
Graph
Source
Debug
Library
Template
System
Help
Start/Restart Debugging Ctrl+F12
Pause Animation Pause
Stop Animation Shift+Pause
Execute F12
Execute Without Breakpoints Alt+F12
Execute for Specified Time
Step Over F10
Step Into F11
Step Out Ctrl+F11
Step To Ctrl+F10
Animate Alt+F11
Reset Popup Windows
Reset Persistent Model Data
Configure Diagnostics...
Use Remote Debug Monitor
Tile Horizontally
Tile Vertically

FIGURA 19. Ubicación de Source **para poder leer el código** Assembler**, y de** Debug **para ver el funcionamiento del programa y los distintos registros del microprocesador.**

www.FreeLibros.me

la barra de **Menú**. En la primera se incluyen las funciones que nos permiten cargar el código, el tipo de microcontrolador y la herramienta generadora de código. También se pueden cargar las herramientas destinadas a hacer el ensamble entre el código realizado y el microcontrolador, para su posterior compilación. En **Debug**, en cambio, se ofrecen las herramientas que nos permiten ver los registros del microcontrolador, el código fuente, la memoria de datos y la memoria EPROM, entre otras características que analizaremos más adelante.

Limitaciones de las versiones

Proteus ofrece diferentes versiones que presentan algunas limitaciones. Veamos cuáles son y sus características. La versión de Proteus que se distribuye sin costo es la denominada Lite, que incluye ISIS y Ares. El primero sirve para dibujar los esquemáticos de los circuitos, por lo que contiene todas las herramientas para tal fin; en tanto que el segundo se usa para desarrollar las placas electrónicas, con lo cual incorpora toda la librería de componentes y sus encapsulados con limitaciones en autoruteador de placas.

En Proteus es posible trabajar con instrumentos virtuales gracias a la implementación de ProSPICE Lite. Esta aplicación se utiliza para simular instrumentos que podemos tener en un taller, como el voltímetro, el amperímetro y el osciloscopio, entre otros. Veamos a continuación las diferencias principales entre la versión Profesional y la versión Lite de este programa:

- No hay integración entre ISIS y Ares en la versión Lite del software.
- ISIS Lite no implementa el verificador de las reglas de conexiones eléctricas (ERC).
- En la versión **Ares de Proteus Lite** no se pueden generar archivos Gerber o Excellon, necesarios para la producción de circuitos impresos.
- La versión de **Proteus Lite ProSPICE** no tiene limitaciones en cuanto a la complejidad de los circuitos que se analizan.
- Proteus Lite no incluye modelos VSM, sino que los necesarios deben comprarse aparte. A su vez, la simulación VSM está limitada a los ejemplos que vienen en esta versión. Es posible realizar cambios al código fuente suministrado en los ejemplos de Proteus Lite, el cual no debe superar 1k.

Más allá de estas restricciones, la versión Lite es una excelente herramienta para complementar nuestro estudio de la electrónica tanto analógica como digital.

MÁS SOBRE PROTEUS

En lo que respecta a la parte analógica, Proteus tiene toda la potencia necesaria para analizar circuitos con componentes pasivos, como resistores, capacitores e inductores; y activos, como diodos, transistores y amplificador operacional.

www.FreeLibros.me

Multiple **choice**

1 ¿Cómo se llama el encargado de diseñar los esquemáticos de los circuitos?

a- ISIS.
b- VSM.
c- ARES.
d- Ninguno de los anteriores.

2 ¿Cómo se llama el módulo de simulación funcional de los circuitos?

a- ISIS.
b- VSM.
c- ARES.
d- Ninguno de los anteriores.

3 ¿Cuál es el módulo Se ocupa del diseño de los circuitos impresos?

a- ISIS.
b- VSM.
c- ARES.
d- Ninguno de los anteriores.

4 ¿Cómo se llama la herramienta que nos permite agregar figuras 2d a nuestros diagramas?

a- Ventana de componentes y librerías.
b- Barra de dibujo.
c- Barra de simulación.
d- Barra de modos de trabajo.

5 ¿Cómo se llama la herramienta que contiene los para la simulación temporal de los circuitos electrónicos?

a- Ventana de componentes y librerías.
b- Barra de dibujo.
c- Barra de simulación.
d- Barra de modos de trabajo.

6 ¿Cómo se llama la herramienta que contiene las opciones correspondientes a cada menú de trabajo?

a- Ventana de componentes y librerías.
b- Barra de dibujo.
c- Barra de simulación.
d- Barra de modos de trabajo.

Respuestas: 1 a, 2 b, 3 c, 4 b, 5 c, 6 a.

www.FreeLibros.me

Capítulo 6

Fuentes de alimentación

Estudiaremos qué es y para qué sirve una fuente de alimentación y cuáles son los diferentes tipos.

www.FreeLibros.me

Las fuentes de alimentación

La fuente de alimentación es la encargada de proveer la tensión adecuada para el correcto funcionamiento de un circuito. Con diseño simple o complejo; reguladas, estabilizadas, lineales, conmutadas, fijas o variables; con circuitos integrados o con componentes discretos, las fuentes de alimentación están presentes en todos los aparatos electrónicos.
Algunas veces menospreciadas por su simpleza aparente, todo taller de este rubro debe contar con una. En particular, los modelos de laboratorio brindan tensión estabilizada y regulada, junto con protección y regulación de corriente. Con ellos, es posible limitar la corriente máxima que se le entregará al circuito, de modo que, aun bajo una falla grave, este se mantenga dentro de condiciones seguras para él y para los usuarios. En estas páginas veremos qué es y para qué sirve una fuente de alimentación, y cuáles son los diferentes tipos y especificaciones para alimentar cualquier circuito electrónico. Las fuentes son necesarias para alimentar los sistemas electrónicos. Es por eso que resulta clave conocerlas a fondo (**Figura 1**)

La energía eléctrica que llega a través de la red domiciliaria lo hace en forma de **corriente alterna**. Esto se debe a que los procesos de generación, transmisión y conversión de electricidad son más simples y eficientes en esta modalidad. En algunos países, se provee de una tensión alterna de **50 Hz**, cuyo valor eficaz es de **220 V**. En otros, se trabaja a **60 Hz** y **110 V**. Sin embargo, casi todos los circuitos electrónicos requieren una tensión

Las fuentes son necesarias para alimentar los sistemas electrónicos

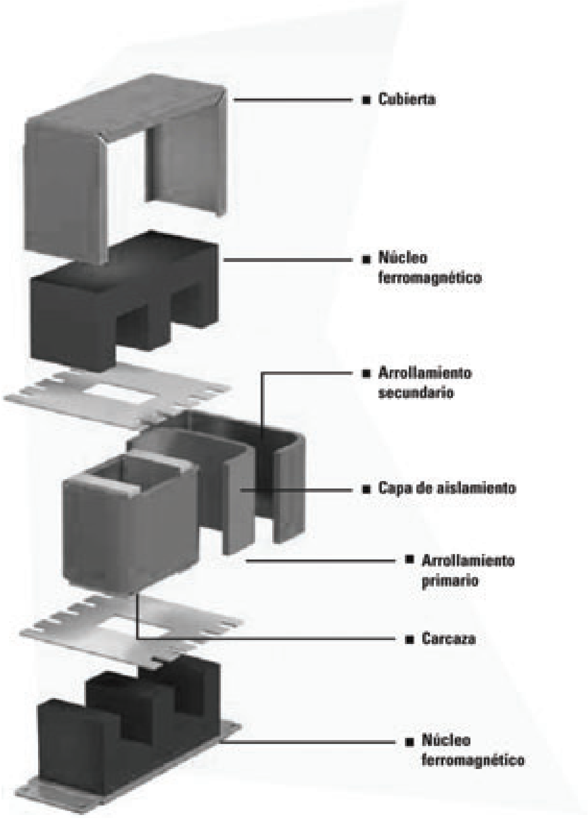

FIGURA 1. **Una fuente convierte energía** AC **en energía** DC. **Las tensiones** DC **logradas deben ser filtradas y reguladas.**

www.FreeLibros.me

continua para funcionar. La fuente de alimentación es, entonces, el dispositivo capaz de convertir la tensión alterna de la red de suministro en tensión continua. Las fuentes se clasifican, básicamente, en **lineales** y **conmutadas**, aunque también existen otras variantes. A continuación, haremos una introducción al tema para, luego, entrar en mayor detalle (**Figura 2**).

- **Fuentes lineales**: tienen un diseño simple, y su circuito posee las siguientes etapas: transformador, rectificador, filtro, regulación y salida. Más adelante estudiaremos en profundidad este tipo de fuentes.
- **Fuentes conmutadas** (switching): utilizan circuitos convertidores de nivel de tensión formados por bobinas y capacitores. La regulación se produce a través de dispositivos de estado sólido, como transistores y **MOSFETs**, entre otros, que operan como llaves. Una fuente de computadora es un ejemplo de fuente conmutada (**Figura 3**).

Una fuente convierte la tensión alterna de la red de suministro en tensión continua

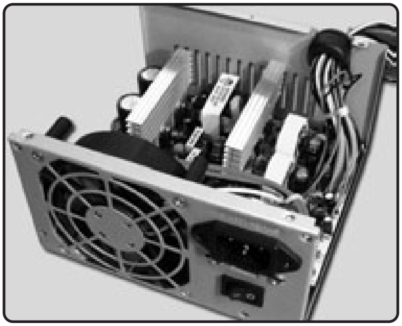

FIGURA 2. **Una fuente de alimentación de laboratorio o banco de pruebas. Tiene la ventaja de poder funcionar como fuente de tensión o de corriente.**

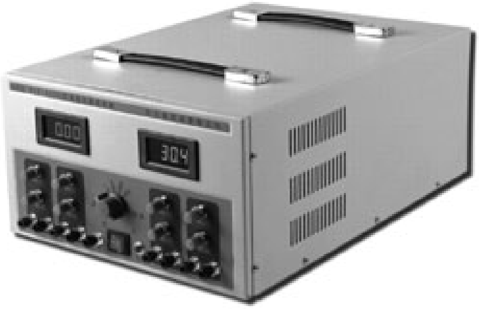

FIGURA 3. **Una fuente de PC de tipo** ATX. **Es conmutada, y su circuito resulta bastante más complejo que el de una lineal.**

BATERÍAS

Son una fuente de alimentación lineal formada por celdas electroquímicas que almacenan la energía eléctrica suministrada para, luego, entregarla al dispositivo en cuestión. Los tipos de baterías más habituales son la de plomo y ácido sulfúrico y la de Ni-Cd.

www.FreeLibros.me

ESPECIFICACIONES DE LAS FUENTES

La primera especificación de una fuente es su tensión de salida, que suele ser de 12 V. La segunda, y prácticamente la más importante, es la corriente máxima de salida. Para entender esta característica, debemos pensar que el circuito o dispositivo electrónico que utilizará la fuente "demanda" la corriente necesaria para funcionar, y la fuente se la entrega. Si utilizamos una fuente que puede entregar una **corriente mayor** a la que el dispositivo precisa, éste funcionará correctamente. En cambio, si es menor, el funcionamiento del circuito se verá afectado, lo que se conoce como **sobrecarga**. La corriente máxima de salida multiplicada por el voltaje de salida da como resultado la potencia máxima que la fuente puede entregar, medida en Watts. Podemos ver otras especificaciones importantes en la **Tabla 1**.

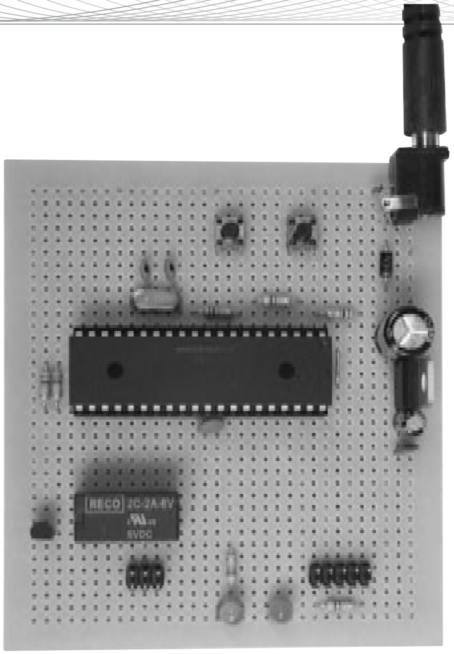

ESPECIFICACIÓN	DETALLES
`Ripple`	Es la ondulación residual que permanece luego del proceso de rectificación y filtrado.
`Regulación de línea`	Mide la variación en la salida nominal de la fuente cuando ocurre una variación en la tensión de entrada. Puede deberse a un cambio en la tensión alterna de la red o al ripple, en caso de un filtrado deficiente.
`Regulación de carga`	Mide la variación en la salida nominal de la fuente al estar sometida a variaciones en la carga conectada a ella.
`Estabilidad en temperatura`	Mide la variación de la tensión continua de salida respecto a las variaciones de temperatura, ya sea del ambiente o debido al calentamiento interno de la fuente.
`Eficiencia`	Es una medida de cuánta energía de la que utiliza la fuente para cumplir su función es realmente entregada a la carga. El resto se pierde en forma de calor.

TABLA 1. **Detallamos aquí otras especificaciones de las fuentes.**

www.FreeLibros.me

PROTECCIÓN

En caso de sobrecarga o cortocircuito, en circuitos sencillos, es común el empleo de un fusible: un finísimo cable dentro de un pequeño tubo de vidrio o plástico que interrumpe el funcionamiento de la fuente, ya que se funde si fluye una corriente peligrosa para la integridad del equipo. En aparatos más complejos, se utiliza un circuito de protección contra sobrecarga, que se activa solo si se intenta exceder el límite de corriente permitido. Una configuración frecuente es la de limitación de corriente **foldback**. Aquí, al ocurrir la sobrecarga, se disminuyen tanto la corriente como el voltaje de salida, con el fin de reducir la disipación de potencia del elemento regulador y de los otros componentes del circuito de la fuente.

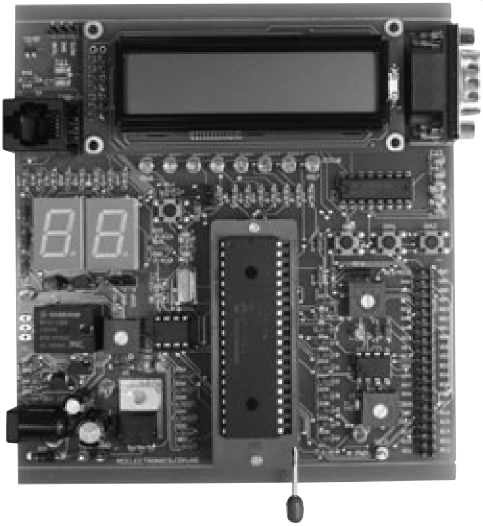

Fuentes lineales y conmutadas

Analizaremos y compararemos los dos tipos de fuentes de alimentación, que se clasifican de acuerdo con la manera de transformar una tensión alterna en otra continua.

Las fuentes de alimentación son necesarias para alimentar casi todos los sistemas electrónicos

Las fuentes de alimentación son necesarias para alimentar casi todos los sistemas electrónicos, tanto digitales como analógicos. La mayoría se diseña para cumplir con alguno o con todos los requerimientos siguientes:

- **Regulación de salida**: la tensión de salida debe permanecer constante dentro de los límites prefijados. Las posibles causas de variación son los cambios en el voltaje de entrada o en la carga manejada por la salida de la fuente.
- **Aislación galvánica**: la salida debe estar eléctricamente aislada de la entrada. De este modo, se evita la interacción entre la salida de la fuente y la tensión de la red de distribución, y se aumenta la seguridad eléctrica.
- **Múltiples salidas**: generalmente se prefiere tener múltiples salidas (positivas y negativas) que difieran en cuanto a especificaciones de voltaje y corriente.

www.FreeLibros.me

Mod_Cap6_xxx_xxx.qxp 04/07/2011 16:41 PAgina 150

LA FUENTE LINEAL

Para llevar la tensión de línea al rango de tensión de salida deseado, y obtener aislamiento galvánico, la fuente lineal utiliza un transformador de **50 Hz** (**Figura 4**). La tensión alterna de salida del transformador ingresa en un circuito rectificador formado por diodos, para convertirse en una tensión continua pulsante. Luego, un circuito que actúa como filtro aplana la señal para que esta ingrese (o no) en un **regulador de tensión**. Este último proporciona una tensión continua y disminuye las variaciones que posee la señal que sale del filtro (ripple). Estas variaciones están directamente relacionadas con la frecuencia de **50 Hz** de la señal de línea. La regulación del voltaje de salida se logra a través de un sistema de control por realimentación; es decir, la salida es registrada y comparada con una referencia, que es, en general, el valor de tensión deseado. Según el resultado de esta comparación, el regulador ajusta la tensión de salida.

La aislación galvánica evita la interacción entre la salida de la fuente y la tensión de línea

LA FUENTE CONMUTADA

A diferencia de la fuente lineal, la conmutada utiliza las etapas en el siguiente orden: rectificador, conmutador, transformador, un segundo rectificador, filtro, regulador y salida. La tensión alterna de entrada se convierte en una tensión continua no regulada, empleando un rectificador similar al de la fuente lineal (**Figura 5**).

La transformación de tensión continua de un nivel a otro se logra por medio de circuitos convertidores como el **flyback**. Estos poseen dispositivos de estado sólido, como transistores y MOSFETs, entre otros, que operan como llaves, completamente encendidos o apagados.

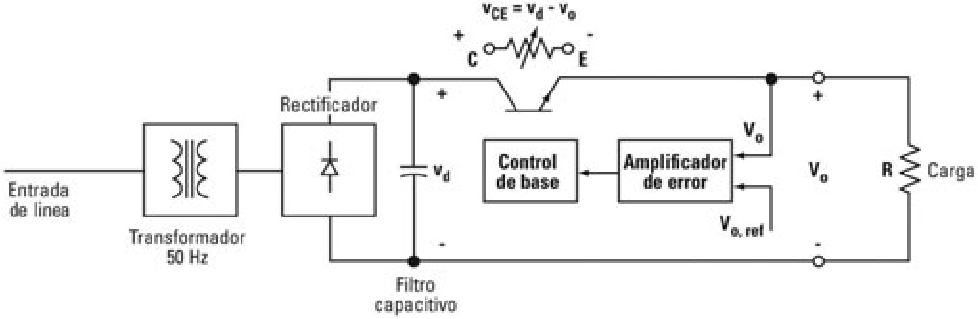

FIGURA 4. **Diagrama en bloques de una fuente lineal. Vemos cómo el transistor controla la tensión de salida de la fuente a través de una realimentación. Dependiendo de la salida, el transistor actúa como regulador ajustando su resistencia.**

www.FreeLibros.me

Mod_Cap6_xxx_xxx.qxp 04/07/2011 16:41 PÁgina 151

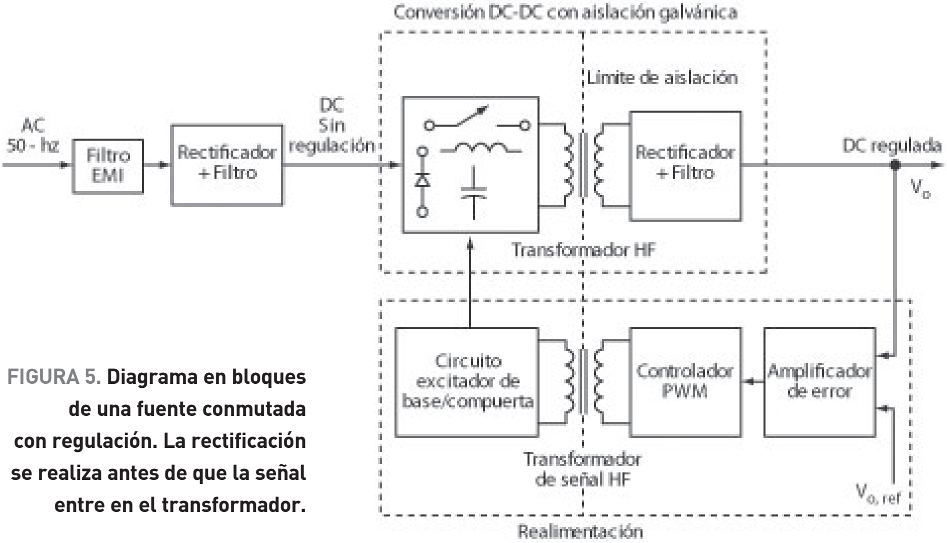
Conversión DC-DC con aislación galvánica
Límite de aislación
AC
50 - hz
Filtro
EMI
Rectificador
+ Filtro
DC
Sin
regulación
Rectificador
+ Filtro
DC regulada
V_o
Transformador HF
Circuito
excitador de
base/compuerta
Controlador
PWM
Amplificador
de error
Transformador
de señal HF
$V_{o, ref}$
Realimentación

FIGURA 5. Diagrama en bloques de una fuente conmutada con regulación. La rectificación se realiza antes de que la señal entre en el transformador.

Para obtener este comportamiento, se utiliza la técnica de **PWM** (*Pulse Width Modulation*), que consiste en conmutar una llave a una frecuencia constante (por lo tanto, a un período constante) y ajustar la fracción del período en que la llave permanece encendida, para lograr, así, controlar el voltaje promedio de salida.

Para obtener el **aislamiento galvánico**, los circuitos convertidores recurren a transformadores de alta frecuencia, ya que la conmutación se realiza a este nivel. La etapa final del convertidor consiste en otro rectificador, que, junto con un circuito integrador, formado por dispositivos que almacenan energía –como bobinas y capacitores–, produce una tensión continua a la salida.

Al igual que la fuente lineal, la regulación del voltaje de salida se consigue a través de una realimentación, porque la salida es registrada y comparada con una referencia, que es el valor de tensión deseado. Según el resultado de esta comparación, el controlador PWM ajusta la fracción de período que la llave permanece encendida para estabilizar la salida de la fuente.

16V
1000µF
6V
00µF

www.FreeLibros.me

Comparación entre **fuentes lineales y conmutadas**

Ahora que ya hemos estudiado el funcionamiento de ambos tipos de fuentes, analicemos sus ventajas y desventajas.

Las ventajas de las fuentes lineales tienen que ver con los siguientes aspectos:

- Poseen un diseño simple, por lo que es viable construirlas de forma casera.
- Para aplicaciones que manejan potencias menores a 10 W, el costo de los componentes es menor que en las fuentes conmutadas.
- Tienen mejor inmunidad a las radiaciones electromagnéticas.
- No son un generador importante de EMI (*Electro Magnetic Interference*), por lo que no afectan el funcionamiento de otros equipos.

Como contrapartida, las **desventajas de las fuentes lineales** son:

- Poseen muy baja eficiencia en aplicaciones normales (entre 30% y 60%). Los valores se incrementan a medida que aumentemos la potencia de la fuente.
- Solo funcionan como reductor de tensión, dado que cada regulador lineal tiene una única salida.

Por otra parte, las **fuentes conmutadas** tienen sus ventajas, que se relacionan con los siguientes aspectos:

- Su eficiencia se encuentra entre un 70% y un 90%, porque el transistor de potencia funciona como llave y no trabaja en su zona activa.
- El transformador de aislación es de alta frecuencia (mayor a 15 KHz), por lo que su tamaño y peso se ven disminuidos de manera considerable.
- Pueden funcionar como reductor, elevador o inversor de tensión con múltiples salidas, porque la tensión de entrada es conmutada y ubicada en un elemento magnético. De este modo, permiten modificar la relación de transformación entre entrada y salida.

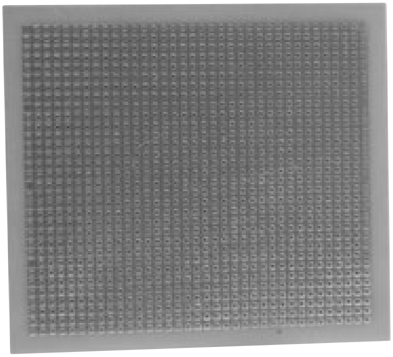

LA ELECCIÓN ADECUADA

Si queremos utilizar una fuente de pequeña potencia y bajos niveles de ruido, y además estamos interesados en realizarla nosotros mismos, la lineal es la mejor opción. Pero, si preferimos la eficiencia y la reducción de espacio, optaremos por una fuente conmutada.

www.FreeLibros.me

Finalmente, describiremos algunas desventajas de este tipo de fuentes:

- Tienen un diseño complejo, que requiere de amplios conocimientos y experiencia en electrónica de potencia. Su diseño, construcción y puesta a punto pueden demandar mucho tiempo.
- El ruido es mayor que en las fuentes lineales. Tanto la entrada como la salida producen radiaciones electromagnéticas, por lo que precisan filtros EMI.

Los aspectos detallados hasta el momento son las principales ventajas y desventajas de estas fuentes. Podemos completar esta información con el material complementario de la obra.

Etapas de una fuente de alimentación lineal

Hasta el momento hemos visto las características de una fuente de alimentación lineal. En estas páginas, estudiaremos en detalle sus tres primeras etapas: transformador, rectificador y filtro, a través de los cuales la tensión alterna de línea se convierte en continua.

EL TRANSFORMADOR

En una fuente lineal, el transformador se encarga de reducir la tensión alterna de la red de distribución a un nivel aceptable para el circuito rectificador. Además, proporciona aislación galvánica. Su funcionamiento se basa en el fenómeno de inducción electromagnética. Está compuesto por dos bobinas devanadas en un núcleo de hierro. La de entrada se denomina devanado primario, y la de salida, devanado secundario. El cociente entre los voltajes en las terminales de ambas bobinas (relación de transformación) es directamente proporcional al cociente de las vueltas de los devanados de las bobinas. Esto significa, por ejemplo, que la tensión de línea de 220 V se verá reducida a 20 V si el número de vueltas del secundario es la décima parte del número de vueltas del primario. Como la potencia se mantiene, la corriente que pasa por el primario será la décima parte de lo que circule por el secundario, según la carga. Para fuentes lineales, el transformador más frecuente es el que posee un núcleo de tipo E-I. Este se emplea en aplicaciones en las cuales bajar los costos es más importante que minimizar el flujo magnético que puede interferir en el funcionamiento del resto del circuito. Para disminuir este efecto se utilizan transformadores toroidales, en los que el núcleo tiene forma de anillo, sobre el cual se bobinan primario y secundario. El flujo magnético queda confinado en el núcleo, y las interferencias son mínimas.

El transformador reduce la tensión alterna de la red a un nivel aceptable para el circuito rectificador

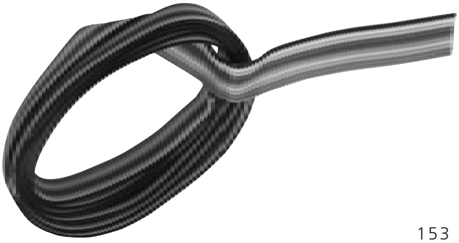

www.FreeLibros.me

¿CÓMO ELEGIR?

La elección de un transformador a partir de un voltaje y una corriente de salida dados es, en principio, simple, pero en la práctica nos llevará a un proceso de prueba y error (**Figura 6**).

Si queremos calcular la tensión de salida del transformador necesaria para una tensión de fuente adecuada, debemos tener en cuenta la caída de voltaje que existe en el rectificador de potencia. Los rectificadores tienen un valor de referencia a una determinada corriente; para circuitos chicos se puede tomar 1 V como referencia por cada diodo. Hay que buscar una relación de transformación tal, que la tensión de salida del transformador se encuentre por encima de la tensión de salida de la fuente. Un valor mayor que 3 V será suficiente. Si vamos a utilizar un regulador en la salida, no es conveniente excedernos con esta diferencia de tensiones, ya que cuanto mayor sea esta, mejor será la disipación en el regulador.

Un detalle para tener en cuenta es que el fabricante nos proporciona el dato de tensión de salida del transformador en valor eficaz, por lo que debemos multiplicarlo por la raíz cuadrada de 2 (relación entre valor pico y valor eficaz), con el fin de obtener el valor pico de alterna.

EL RECTIFICADOR

Se denomina rectificador a cualquier dispositivo eléctrico que presente una resistencia pequeña a la corriente en un sentido, y una resistencia grande en el opuesto. Este dispositivo es capaz de convertir una forma de onda de entrada sinusoidal, cuyo valor medio es cero, en una forma de onda continua pulsante, con su correspondiente valor medio. Para la rectificación se utilizan uno o más diodos. Estos semiconductores permiten el paso de una corriente en un sentido y lo bloquean en el otro. Estudiaremos aquí las configuraciones de rectificación más habituales.

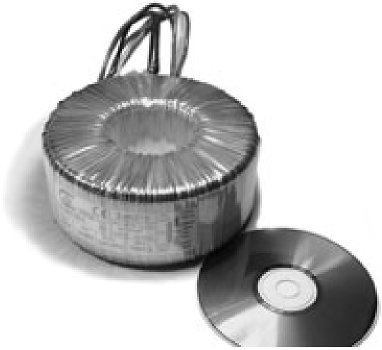

FIGURA 6.
Un transformador toroidal. En este caso, el flujo magnético queda concentrado uniformemente en el núcleo, lo que minimiza las interferencias al resto del circuito.

FUENTES SIN TRANSFORMADOR

Estas proporcionan una alternativa de bajo costo. Su funcionamiento se basa en el principio de impedancia: utilizan un capacitor para reducir el voltaje de entrada, y para proporcionar una tensión de salida continua emplean un diodo zener.

www.FreeLibros.me

EL RECTIFICADOR DE MEDIA ONDA

Esta configuración de rectificación consiste en interponer un diodo entre la fuente de tensión (el secundario del transformador) y la carga que representa el circuito por alimentar (**Figura 7**). Cuando la tensión de la fuente está en su semiciclo positivo, el sentido de la corriente es favorable y esta circula. Si consideramos que no existe caída de tensión en el diodo, la forma de onda de la tensión en la carga es igual a la del secundario durante todo el semiciclo. En cambio, si la tensión de fuente se encuentra en su semiciclo negativo, el diodo bloquea la corriente y, entonces, el voltaje en la carga es nulo.

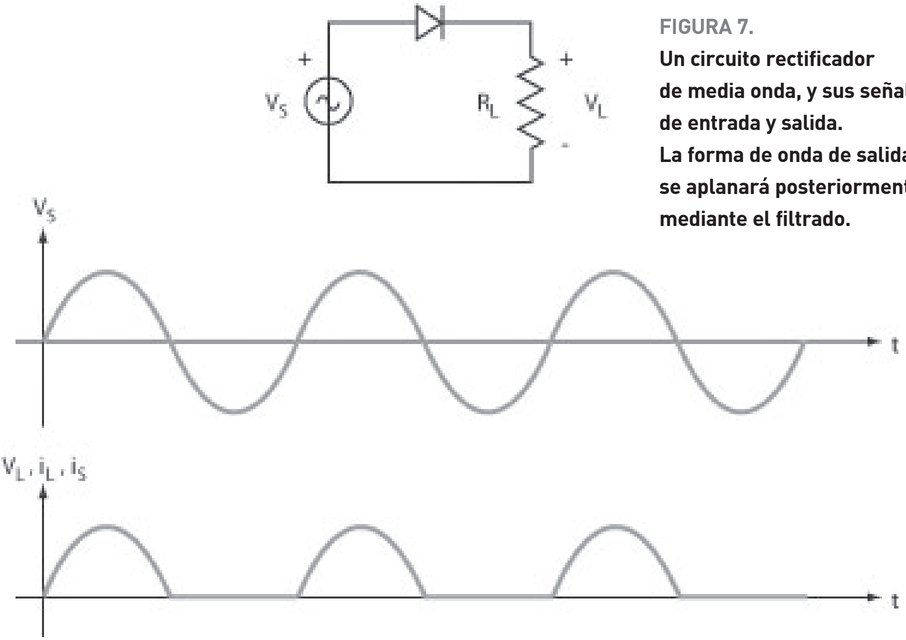

FIGURA 7.
Un circuito rectificador de media onda, y sus señales de entrada y salida. La forma de onda de salida se aplanará posteriormente mediante el filtrado.

DESVENTAJAS DEL RECTIFICADOR DE MEDIA ONDA

El rectificador de media onda no permite utilizar toda la energía disponible, debido a que, en los semiciclos negativos de la tensión del secundario del transformador, el diodo bloquea la corriente y no llega energía a la carga.

www.FreeLibros.me

LOS RECTIFICADORES DE ONDA COMPLETA

Existen dos tipos de rectificadores de onda completa. El primero se logra utilizando dos fuentes en contrafase en vez de una sola, y colocando en cada una de ellas un rectificador de media onda. Se emplea un transformador con doble devanado en el secundario, tomándose el punto medio como masa común (**Figura 8**).

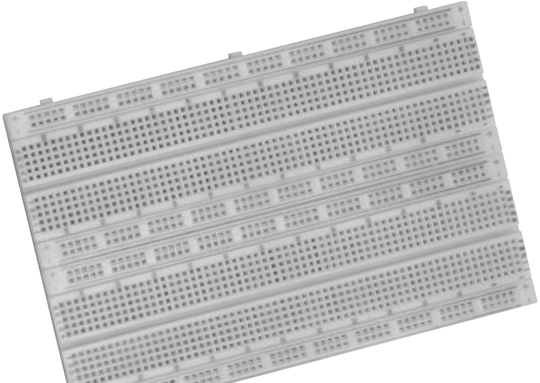

> El rectificador de onda completa del transformador de punto medio emplea dos fuentes en contrafase en vez de una sola

Cuando la señal de alterna de entrada está en su semiciclo positivo, el diodo **D1** conduce, y el **D2**, no. Entonces, la tensión de entrada queda aplicada directamente a la carga a través de **D1** y con una polaridad positiva. Cuando la forma de onda de entrada se encuentra en su semiciclo negativo, estará en el semiciclo positivo en el bobinado en contrafase; entonces conduce **D2**, por lo cual otra vez se aplica una tensión de fuente positiva a la carga. En este caso no existe magnetización neta permanente en el transformador y, además, se aprovecha la totalidad de la forma de onda de entrada.

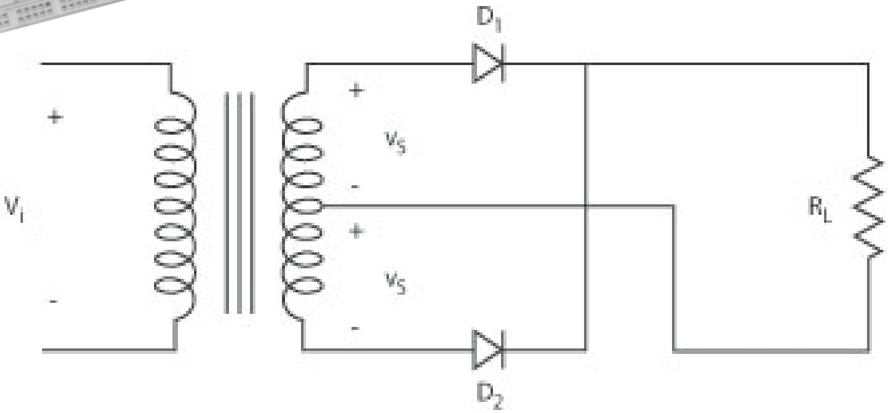

FIGURA 8. Un rectificador de onda completa con un transformador con doble devanado secundario, también llamado transformador de punto medio.

www.FreeLibros.me

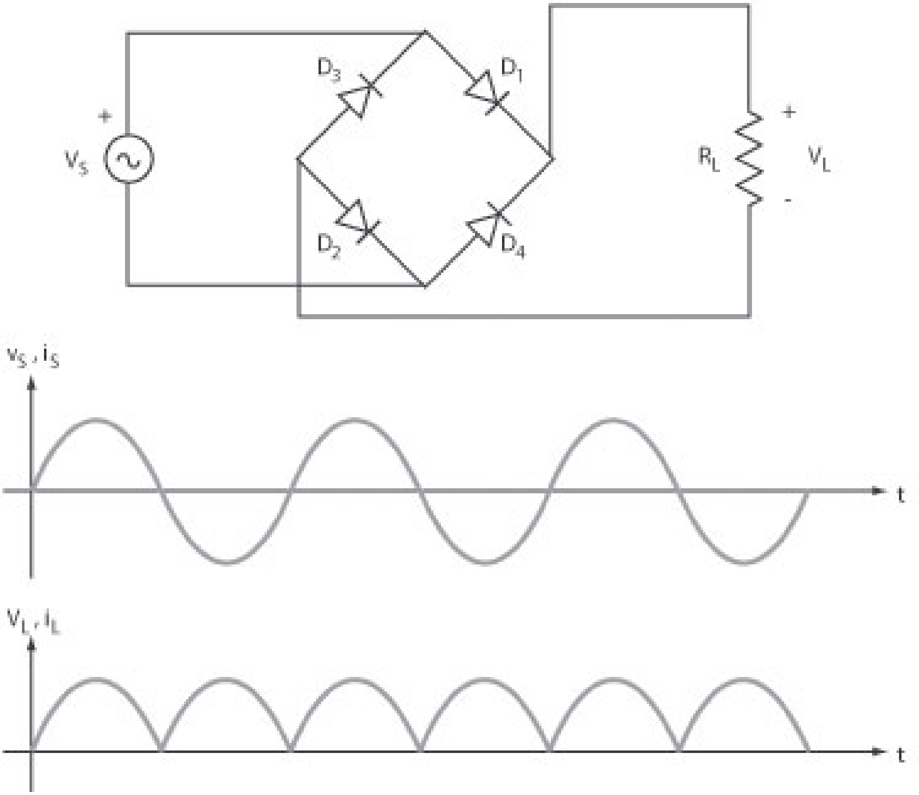

FIGURA 9. **Un circuito rectificador de onda completa tipo puente. Vemos cómo, al ser un rectificador de onda completa, se aprovecha toda la energía disponible.**

RECTIFICADOR PUENTE DE GRAETZ

Se compone de un arreglo de cuatro diodos intercalados entre la fuente de tensión y la carga (**Figura 9**). Cuando la salida del secundario es mayor que 0, los diodos **D1** y **D2** quedan polarizados en forma directa. Por lo tanto, permiten la conducción de corriente, mientras que los diodos **D3** y **D4** no conducen. Si los tomamos como elementos ideales y, por lo tanto, despreciando la caída de tensión entre sus terminales, la forma de onda de la tensión en la carga es igual a la del secundario durante todo el semiciclo. Cuando la entrada pasa a su semiciclo negativo y, por lo tanto, es ahora menor que 0, **D3** y **D4** son los que conducen, en tanto que **D1** y **D2** bloquean el paso de corriente a través de sus terminales. Como resultado, el secundario queda "conectado al revés", y la forma de onda de la tensión en la carga es la del secundario invertida, durante todo el semiciclo.

www.FreeLibros.me

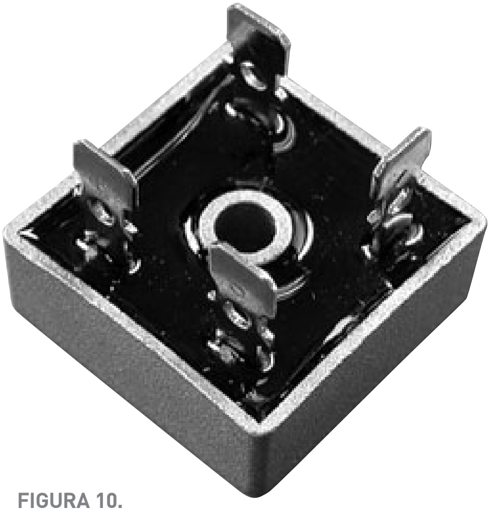

FIGURA 10.
Un puente rectificador de diodos integrado. Los cuatro diodos vienen dentro de él y así se logra una disminución en el tamaño del puente. Además, este tipo de encapsulado mejora la disipación de los diodos.

Una ventaja de los rectificadores tipo puente es que utilizan transformadores con devanado simple en el secundario, de modo que resultarán más económicos que un rectificador con transformador de doble devanado (**Figura 10**). En cuanto a las especificaciones de los diodos, estos deben ser capaces de disipar la potencia máxima necesaria y un margen de seguridad. Los diodos rectificadores se clasifican según su corriente promedio y el voltaje pico inverso. El valor medio de la corriente por el diodo es la mitad del de la carga, porque conduce sobre uno de los semiciclos solamente. La tensión inversa máxima sobre el diodo es lo que hay en el capacitor: valor de pico, más el pico inverso en el transformador; es decir, dos veces el valor de pico.

FILTROS DE SALIDA

Las formas de onda de salida de los circuitos rectificadores necesitan un filtrado para alimentar un circuito electrónico. Para este fin, se utiliza un circuito integrador que funciona como si fuera un filtro **pasabajos** para eliminar el ripple (**Figura 11**). Por lo general, este filtrado se lleva a cabo colocando un capacitor en paralelo con la carga. El funcionamiento de este sistema se basa en que el capacitor almacena energía durante el período de

El filtrado a la salida de la fuente se lleva a cabo, por lo general, colocando un capacitor en paralelo con la carga

UN DATO IMPORTANTE

Para el caso del rectificador de onda completa, el funcionamiento es similar al de media onda, con la diferencia de que la caída de la tensión del capacitor debido a su descarga está con el pico negativo rectificado, donde comienza a cargarse otra vez.

www.FreeLibros.me

conducción del diodo, y la entrega a la carga en el período de no conducción. De esta manera, se prolonga el tiempo durante el cual circula corriente por la carga, y se disminuye notablemente la componente de alterna residual (ripple). Para explicar el filtrado, supongamos un rectificador de media onda con un capacitor en paralelo con la carga. Vamos a considerar que la resistencia de carga es infinita. El capacitor se cargará a la tensión máxima del secundario del transformador, y mantendrá esta tensión porque no existe ningún camino por el que pueda descargarse, ya que el diodo impide la circulación de corriente en sentido inverso. De este modo, el filtrado es perfecto.

Si suponemos que la carga es finita, el capacitor se descargará a través de ella, porque ahora cuenta con un camino para hacerlo. Cuanto mayor sea el valor de capacidad del capacitor, menor será su caída de tensión por la descarga (disminuye el ripple). Cuanto menor sea la resistencia de carga, mayor será esta caída (aumenta el ripple) **Figura 12**.

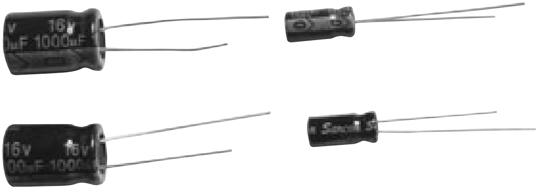

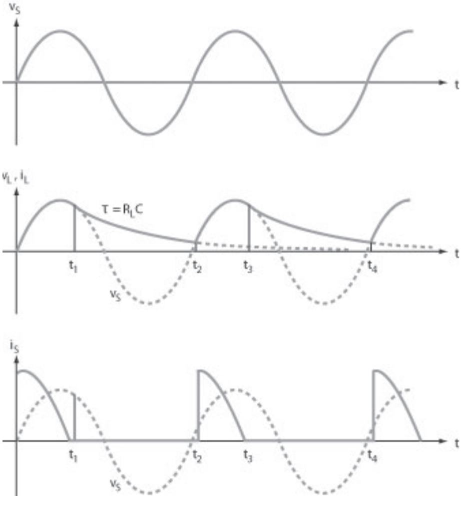

FIGURA 11.

Señal de entrada a un rectificador con filtro de capacitor, tensión, y corriente por la carga y por el diodo. La constante de descarga es ? =R x C.

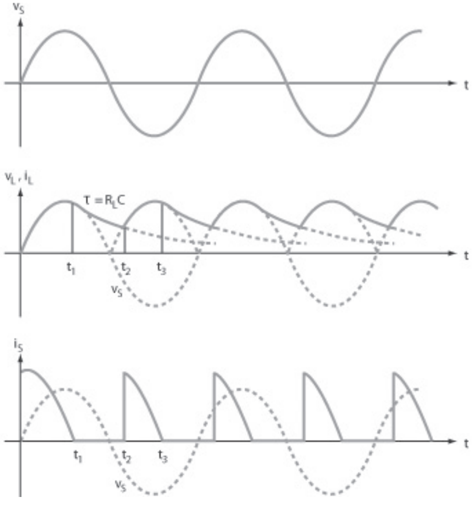

FIGURA 12.

En el diagrama podemos visualizar las formas de onda de entrada y salida de un rectificador de onda completa con filtro.

www.FreeLibros.me

INFOGRAFÍA 1

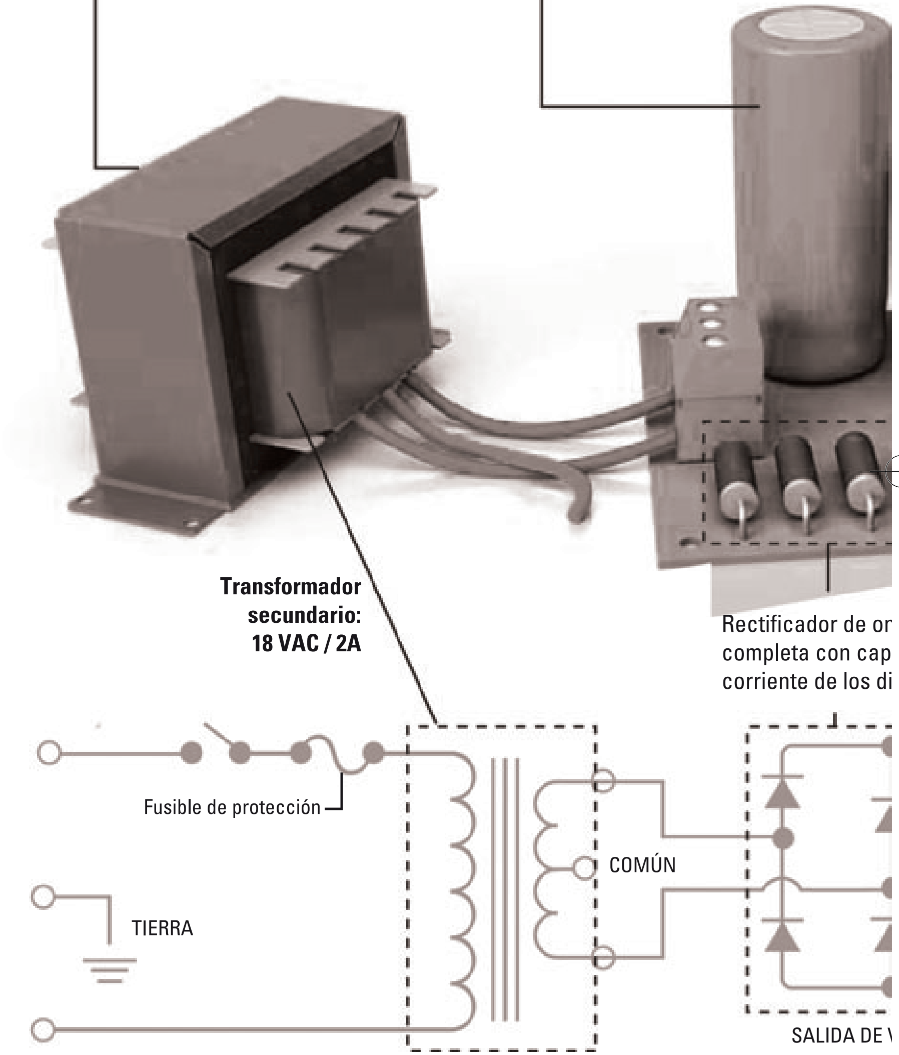

Primario del transformador 220 / 110 VAC

Capacitor electrolítico de filtrad
3300 uf / 35 V (max). Siempre elegir
tensión de aplicación máxima el dobl
de la salida del puente rectificador.

Transformador secundario: 18 VAC / 2A

Rectificador de or
completa con cap
corriente de los di

Fusible de protección
COMÚN
TIERRA
SALIDA DE

www.FreeLibros.me

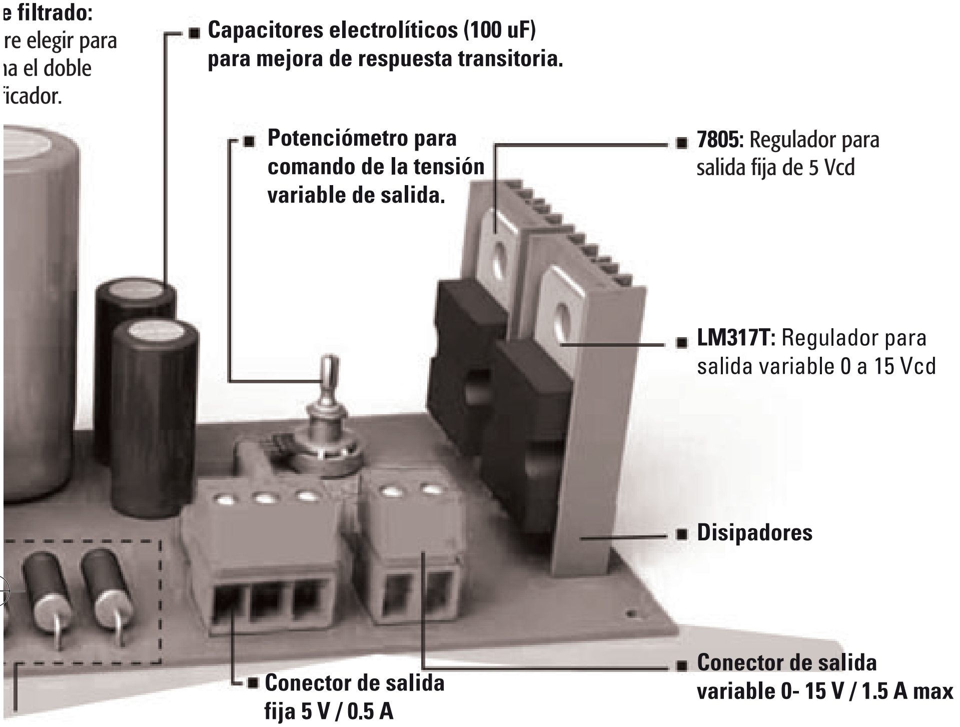
e filtrado:
re elegir para
a el doble
ficador.
Capacitores electrolíticos (100 uF) para mejora de respuesta transitoria.
Potenciómetro para comando de la tensión variable de salida.
7805: Regulador para salida fija de 5 Vcd
LM317T: Regulador para salida variable 0 a 15 Vcd
Disipadores
Conector de salida fija 5 V / 0.5 A
Conector de salida variable 0- 15 V / 1.5 A max

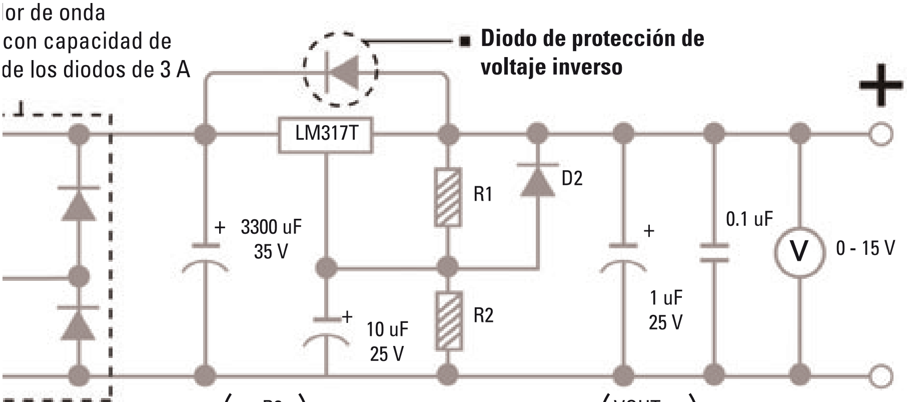
lor de onda
con capacidad de
de los diodos de 3 A
Diodo de protección de voltaje inverso
+
LM317T
R1
D2
R2
+ 3300 uF 35 V
+ 10 uF 25 V
+ 1 uF 25 V
0.1 uF
V
0 - 15 V
−

LIDA DE VOLTAJE= $1.25 \times \left(1 + \frac{R2}{R1}\right) + 1\ (ADJ \times R2)$ $\approx R2 = \left(\frac{VOUT}{R1} - 1\right) \times R1$

www.FreeLibros.me

Diseño de una fuente de alimentación

En estas páginas conoceremos cuáles son las características que debemos tener en cuenta para diseñar una fuente de alimentación.

Tomemos como ejemplo un circuito que consume una corriente de **800 mA** (0,8 A) y debe ser alimentado con una tensión de **12 V**. Con estos valores, podemos calcular la potencia consumida. Recordemos que potencia **W (Watt) = I (Ampere) x E (Volt)**. Por lo tanto, el consumo será de **0,8 A x 12 V = 9,6 W**.

Nuestro transformador deberá suministrar una potencia mayor; por razones de confiabilidad, recomendamos que sea del doble.

Para pequeñas potencias, también podemos seleccionar el transformador especificando la corriente. En nuestro caso, uno de **2 A** sería más que suficiente; **2 A x 12 V = 24 W**.

Con respecto a la tensión, si bien necesitamos **12 V**, debemos dejar cierto margen para que el circuito regulador pueda operar. La tensión rectificada será el valor de pico de la tensión alterna; es decir, el producto de su valor eficaz (**12 V**) por la raíz cuadrada de **2**, casi **17 V**. A esta cifra debemos restarle la caída de tensión en los rectificadores.

Si utilizamos un rectificador en puente de onda completa (cuatro diodos), en cada ciclo de la tensión alterna estarán involucrados dos diodos. La caída de tensión por diodo será cercana a **1,2 V**. Tenemos, entonces: **17 V - 2,4 V = 14,6 V**. Esta será la tensión resultante cuando la fuente consuma cierta cantidad de corriente. Dado que debemos regular este valor a **12 V**, utilizaremos un regulador fijo de tipo **7812**.

Según las características del regulador, se necesita una tensión de entrada mínima de **14,6 V**, que asegure una regulación de la tensión de salida. Debemos dejar un poco de margen y aumentar la tensión de salida del transformador; uno de **14 V** y **20 W** será adecuado.

Los rectificadores deberán soportar la corriente media. Como utilizamos un rectificador en puente, cada diodo conduce la mitad del tiempo, y la corriente media es la mitad de la que pide la carga (**0,4 A**). Entonces, para el rectificador en puente vamos a elegir cuatro diodos de **1 A** como mínimo. El **1N4002** es perfecto para nuestros propósitos, ya que maneja **1 A** y una tensión de pico inversa de **100 V**.

UN DATO IMPORTANTE

Sobredimensionar un transformador nunca está de más. Siempre debemos considerar la peor de las situaciones, que es el trabajo continuo y al máximo de su potencia. Conviene duplicar la corriente necesaria para dejar un margen de seguridad.

www.FreeLibros.me

Mod_Cap6_xxx_xxx.qxp 04/07/2011 16:41 Página 163

TENSIÓN DE RIZADO O RIPPLE

A la salida del rectificador disponemos de una tensión positiva pulsante, no apta para aplicar a ningún circuito electrónico. El capacitor de filtro se cargará con los picos o pulsos de la tensión de salida del rectificador, y mantendrá esa tensión durante la ausencia de estos. De este modo, obtenemos una tensión continua pura, que es la que finalmente aplicaremos al regulador de tensión (**Figura 13**).

> Por cada ampere de consumo se requieren aproximadamente 2000 μF para el capacitor de filtro

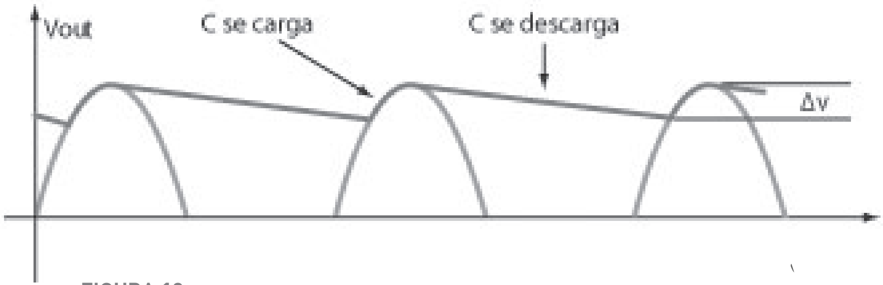

FIGURA 13.
Vout es el voltaje en el capacitor. Este se carga en el flanco creciente del ciclo hasta llegar al pico; a partir de allí, se descarga en el circuito. Circuito rectificador de media onda.

Ahora bien, debido a que un capacitor mantiene su carga durante un tiempo determinado, su capacidad debe ser lo suficientemente alta para que la tensión no decaiga entre los picos de la tensión pulsante de la salida del rectificador. Esto es lo que se conoce como rizado o ripple, y es la caída de tensión (**ΔV**) entre los "valles" de los picos de la tensión pulsante (**Figura 14**). Si conectamos un osciloscopio a una tensión con ripple, veremos una línea con pequeñas deformaciones repetitivas.

www.FreeLibros.me

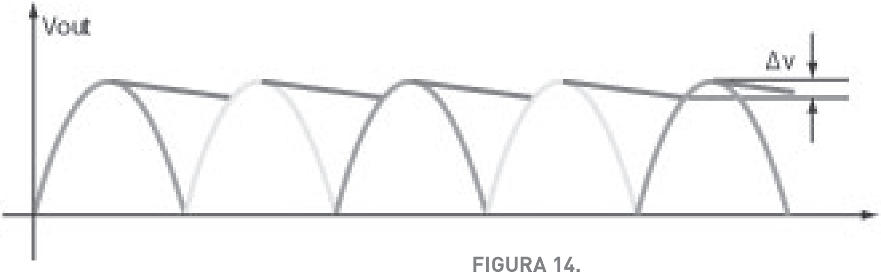

FIGURA 14.
En esta imagen vemos los beneficios de la rectificación de onda completa, que, a diferencia de la de media onda, tiene menor ripple porque los ciclos están juntos entre sí y el capacitor se recarga en la mitad de tiempo.

Algunos circuitos son más inmunes que otros al rizado, pero éste sería un factor inadmisible en un amplificador de audio, ya que escucharíamos de fondo un leve zumbido producido por el ripple de la fuente.

El cálculo del capacitor se realiza estimando cuánto se puede dejar caer la tensión (**ΔV**) en el tiempo que tarda en recargarse (**Δt**), dada la corriente (**I**) que pide el circuito: **C = I x Δt / ΔV**. Una regla empírica dice que, para onda completa, por cada ampere de consumo se requieren unos **2000 μF**. La tensión que debe soportar el capacitor tiene que ser superior a la tensión rectificada; se recomienda que sea cercana al doble.
Volviendo al ejemplo anterior, un capacitor de **2200 μF** y **25 V** será suficiente, ya que consumimos casi **1 A** (**0,8 A**), pero debido a que disponemos de un transformador de 2 A, podríamos aprovechar mejor la corriente extra y colocar un capacitor de **4700 μF** y **25 V**.

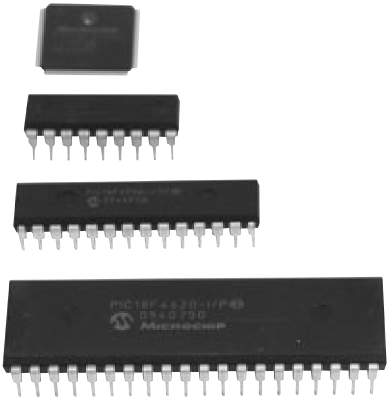

MÁS SOBRE EL RIPPLE

Con la rectificación de onda completa, en el mismo tiempo tenemos dos pulsos de la corriente rectificada en contraste con el rectificador de media onda, que sólo proporciona uno. Esto significa que, con el mismo capacitor en onda completa, habrá menor ripple.

www.FreeLibros.me

Los reguladores de voltaje

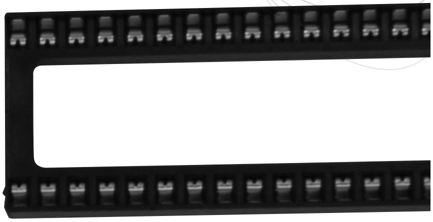

Son los dispositivos encargados de mantener un voltaje estable a lo largo del tiempo, independientemente del consumo y de los cambios de voltaje a su entrada.

En electrónica hay muchas maneras de hacer lo mismo, pero se opta siempre por la solución más segura y económica sobre la base de las necesidades del circuito.

REGULADORES FIJOS

Existen dos tipos de reguladores: en paralelo y en serie. Los primeros son utilizados cuando la demanda de corriente es baja. En ellos se emplean diodos **zener**, que tienen la propiedad de conducir cierta corriente cuando están polarizados de forma inversa y se supera determinado nivel de voltaje (**Figura 15**).

Pongamos como ejemplo un diodo zener de **5,1 V**. Si aplicamos entre ánodo y cátodo un voltaje que supere los **5,1 V**, el diodo comenzará a conducir corriente, manteniendo entre sus terminales una tensión cercana a **5,1 V**. Debido a que estos diodos tienen un límite en la conducción de corriente, se les agrega en serie con la tensión una resistencia limitadora de corriente, que se calcula considerando la corriente que requiere el circuito al cual se conecta, más la necesaria por el diodo para mantener la tensión de zener.

Existen dos tipos de reguladores: en paralelo y en serie

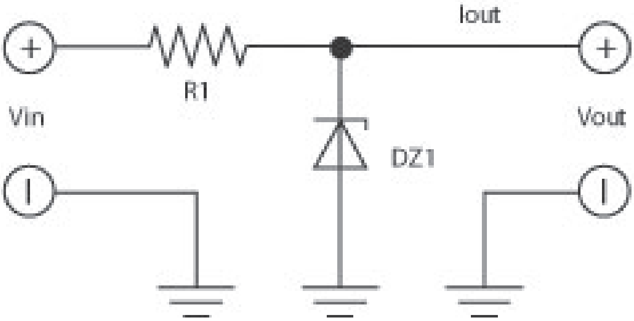

FIGURA 15.
El diodo zener deja pasar cierta corriente por sí mismo para que las variaciones de Vin (entrada) **y la corriente de carga** Iout (consumo) **no modifiquen el voltaje** Vout.

www.FreeLibros.me

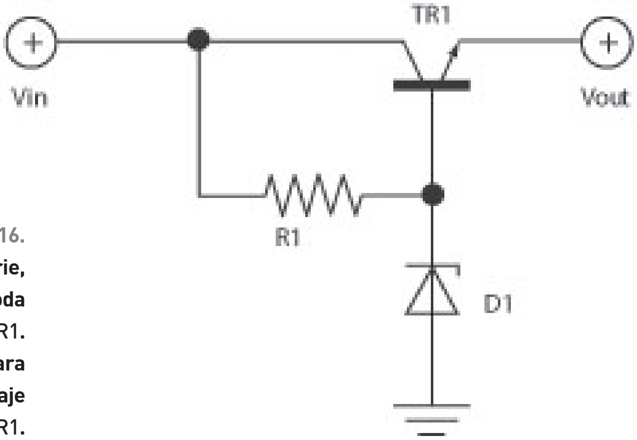

FIGURA 16.
En los reguladores serie, el componente que soporta toda la corriente es el transistor TR1**. El diodo** zener **es usado para dar una referencia de voltaje a la base de** TR1**.**

Por ejemplo, precisamos alimentar un circuito con **5,1 V** que consume una corriente de **20 mA**, y disponemos de una tensión de **14 V**. Escogemos el diodo **zener 1N4733**, que es de **5,1 V** y es capaz de disipar **1 W**. El fabricante dice que requiere, como mínimo, una corriente de **1 mA** para regular. Para asegurarnos la regulación, debemos considerar las peores condiciones, y ellas son: que la tensión de entrada disminuya (por ejemplo, que baje a **12 V**) y que el circuito consuma, por momentos, el doble de corriente (**40 mA**). A partir de estos valores calculamos la resistencia serie:

[Voltaje de entrada (el más bajo) - **voltaje del zener] / [corriente de consumo del circuito** (la máxima) **+ corriente mínima necesaria por el zener]=**

(12 V - 5,1 V) / (0,040 A + 0,001 A) = 168 Ohms

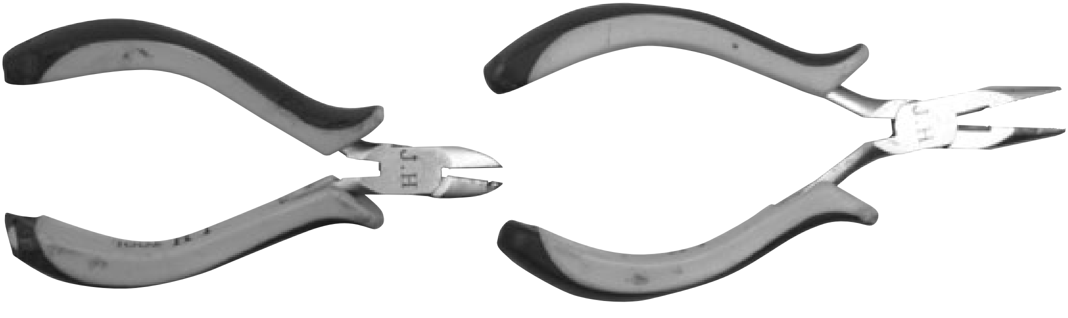

www.FreeLibros.me

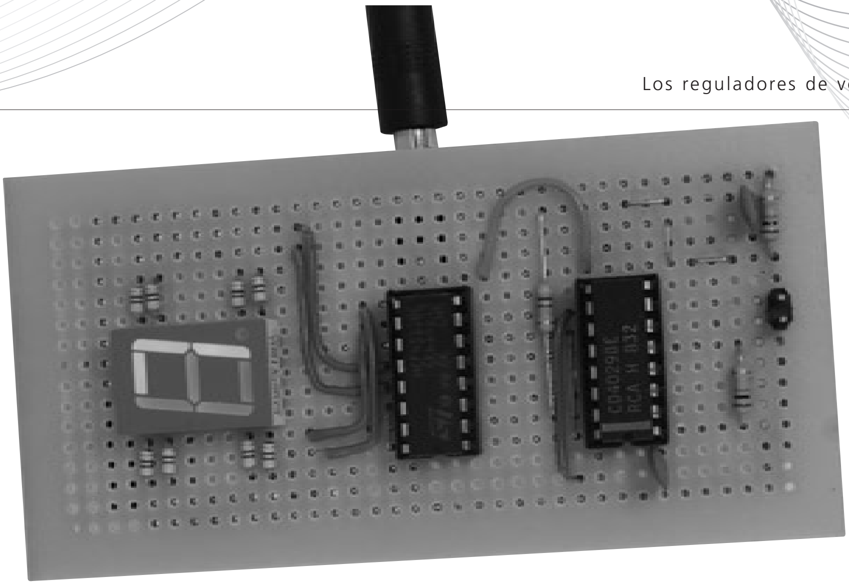

Ahora supongamos que desconectamos la carga (el circuito) del diodo zener. Entonces, toda la corriente circulará por el diodo y, en el peor caso, la máxima tensión de entrada será de **(14 V - 5,1 V) / 168 Ohms = 53 mA**, correspondiente a una potencia de **5,1 V x 53 mA = 270 mW**, dentro de los parámetros de operación del diodo.

En circuitos donde se requiere mayor corriente, se apela a la regulación en serie. En este caso, se utilizan transistores conectados en serie con la tensión, y se polarizan con un elemento de referencia fija como el diodo zener.

Por último, están los reguladores integrados, como el **7805**, que en su interior tiene un circuito similar al regulador serie, pero con compensaciones por temperatura y protecciones contra cortocircuitos. Utilizan tres terminales: entrada, masa y salida.

En circuitos donde se requiere mayor corriente, se apela a la regulación en serie

Para el ejemplo anterior de la fuente de **12 V**, podremos utilizar un regulador **78T12**, que soporta hasta **3 A**. De este modo, aprovecharemos el máximo potencial del transformador.

SOBRE ZENER

Existen diodos zener de cientos de Watts. Sin embargo, su uso está restringido a ciertos circuitos de control de potencia. En nuestro caso, sólo los usaremos cuando necesitemos mantener estable un voltaje de baja corriente en determinado punto de un circuito.

www.FreeLibros.me

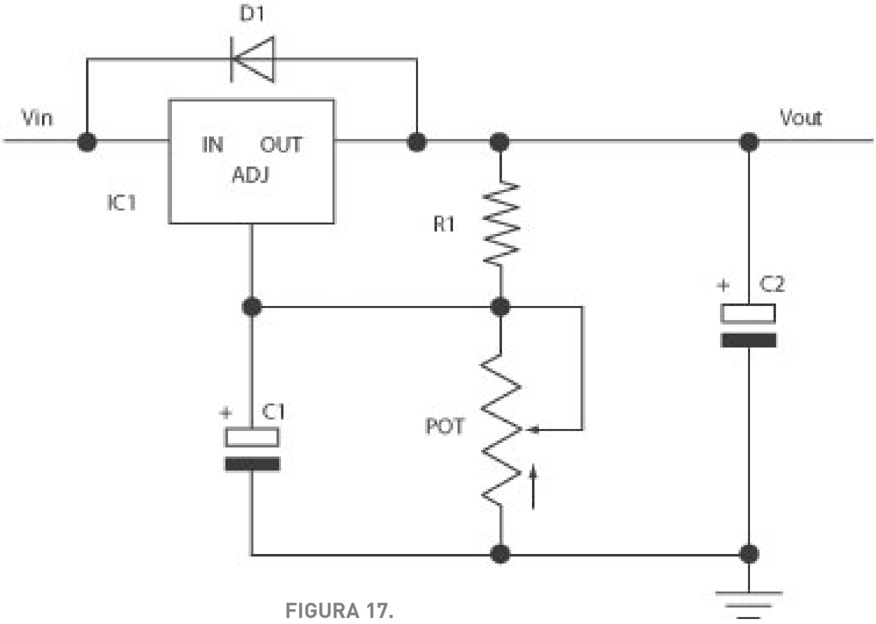

FIGURA 17.
El terminal de ajuste (adj) del LM317 **se ocupa de regular su tensión de salida.**

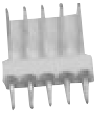

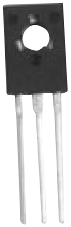

LOS REGULADORES VARIABLES

Como su nombre lo indica, estos reguladores permiten obtener una gama continua de distintos voltajes. Gracias a ellos es posible construir una fuente de alimentación universal. Como en el caso de los reguladores fijos, los reguladores variables pueden construirse a partir de componentes discretos o por medio de circuitos integrados (**CI**) especializados. A continuación, veremos un regulador variable construido en torno al **CI LM317** (**Figura 17**).

El **CI LM317** cuenta con tres terminales: uno es utilizado como entrada de la tensión sin regular, y otro, como salida regulada. Un tercer terminal es el encargado de controlar la tensión de salida. En la hoja de datos del fabricante encontraremos los múltiples usos de este versátil regulador.

www.FreeLibros.me

FUENTE DE ALIMENTACIÓN PARA EL TALLER

Es indispensable contar con una fuente de alimentación para realizar todas las prácticas del curso y, además, utilizarla en todos los montajes que necesitemos probar.

La siguiente fuente de alimentación está diseñada para que sea un elemento de gran valor para el técnico. Fue concebida teniendo en cuenta los múltiples voltajes que se utilizan habitualmente en electrónica digital. No desaprovecharemos la oportunidad de construir una fuente que sirva para la mayoría de los circuitos digitales. Las tensiones de salida son **3.3 V**, **5 V**, **12 V** y una corriente de **1 A**.

Como en cada uno de los proyectos, lo primero que debemos tener en claro es el esquemático que especifica detalladamente las características y componentes de la fuente de alimentación que vamos a desarrollar (**Paso a paso 1**). Tomando como ruta este esquemático, evitaremos cometer errores (**Figura 18**).

En primer lugar debemos contar con todos los elementos necesarios para armar la fuente de alimentación. En nuestro caso aplicaremos los reguladores estudiados. Para los **12 V** y **5 V** emplearemos reguladores fijos **7812** y **7805** respectivamente. Para los **3,3 V** utilizaremos el versátil regulador variable **LM317** y ajustaremos el **preset** (potenciómetro de ajuste) para **3,3 V** de salida.

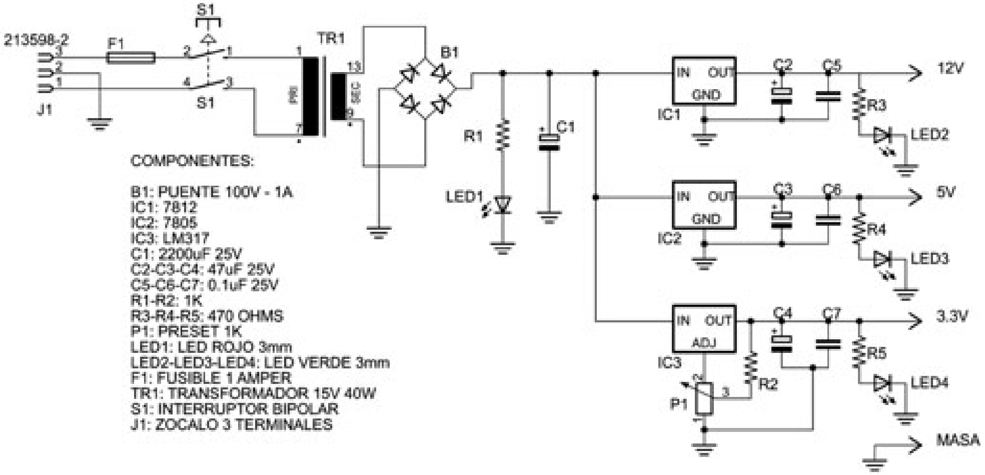

COMPONENTES:

B1: PUENTE 100V - 1A
IC1: 7812
IC2: 7805
IC3: LM317
C1: 2200uF 25V
C2-C3-C4: 47uF 25V
C5-C6-C7: 0.1uF 25V
R1-R2: 1K
R3-R4-R5: 470 OHMS
P1: PRESET 1K
LED1: LED ROJO 3mm
LED2-LED3-LED4: LED VERDE 3mm
F1: FUSIBLE 1 AMPER
TR1: TRANSFORMADOR 15V 40W
S1: INTERRUPTOR BIPOLAR
J1: ZOCALO 3 TERMINALES

FIGURA 18.

El costo de realizar una fuente de un voltaje múltiple es casi el mismo que el de una fuente de voltaje único.

www.FreeLibros.me

PASO A PASO /1
Fuente de alimentación para el taller

1

El primer paso es definir el transformador necesario. Para este fin, basta con saber cuál será el voltaje máximo que pretende extraer de la fuente. En este caso, será de **12 V** y, según lo visto en párrafos anteriores, el transformador debe proveer **14** o **15 V** de salida. Como hay varias salidas, use uno de **40 W** aproximadamente.

2

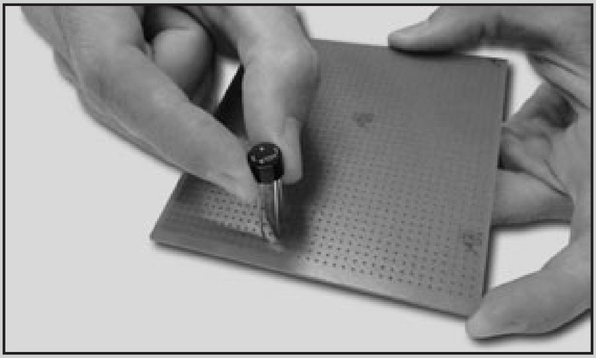

Comience por colocar el puente rectificador en la tarjeta de circuito impreso universal. Recuerde la polaridad indicada en la superficie del encapsulado, entendiendo que los símbolos ~ son las entradas de corriente alterna; **+** y **-** indican positivo y negativo respectivamente.

www.FreeLibros.me

PASO A PASO /1 (cont.)

3

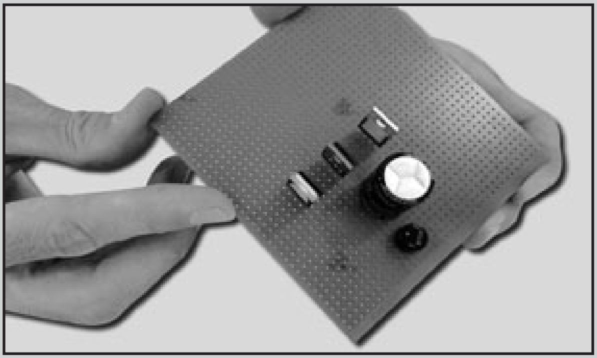

Ahora coloque el capacitor principal de filtro con el terminal positivo orientado hacia el mismo terminal de salida del puente rectificador. Inserte los reguladores fijos **7812** y **7805**, como indica la imagen. De esta manera, las conexiones que deberá efectuar serán las más cortas. Introduzca el **LM317** en el PCB según la fotografía.

4

Es momento de girar el PCB y comenzar con la soldadura de los componentes. Suelde uno a uno los terminales hasta finalizar con todos. En este momento, puede comenzar con la conexión entre ellos y, en algunos casos, podrá emplear los mismos terminales de los componentes como conexiones, soldándolos en todo su recorrido para evitar desconexiones por vibraciones.

www.FreeLibros.me

PASO A PASO /1 (cont.)

5

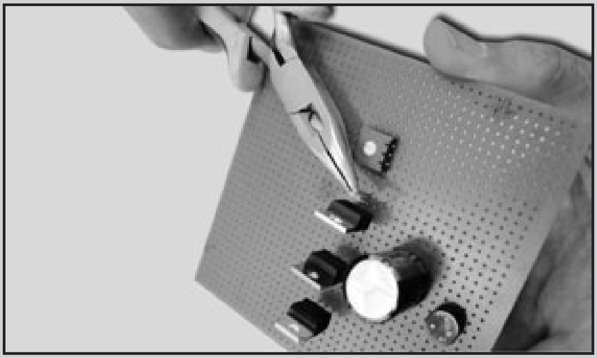

Ahora introduzca el preset y la resistencia de **1K**. Gire nuevamente el PCB y suelde estos componentes. Realice las conexiones del preset y la resistencia según el esquemático. No confunda los terminales del **LM317**, ya que el primero de la izquierda es el de ajuste, el del medio es la salida, y el de la derecha, la entrada.

6

Coloque ahora los capacitores de desacople de salida, que consisten en uno de **10 µF 25 V** y otro de **0.1 µF**. Preste especial atención a la polaridad de estos. Por norma, los terminales más largos en los capacitores electrolíticos corresponden al terminal (+). Lo mismo ocurre con los LEDs.

www.FreeLibros.me

PASO A PASO /1 (cont.)

7

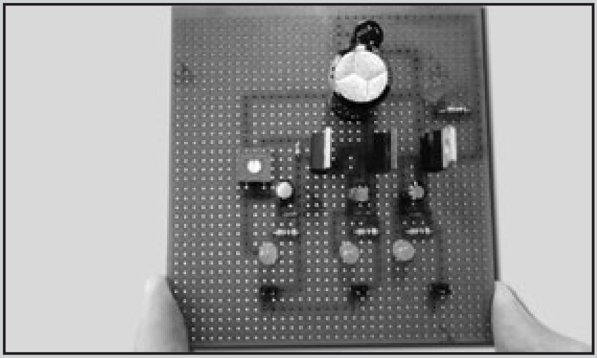

Ahora agregue las resistencias limitadoras de corriente para los LEDs. La de **1 K** es solo para el LED conectado a la salida del rectificador y, el resto de **470 Ohms**, para las salidas reguladas. Gire el PCB y suéldelas. Luego, agregue los LEDs dejándolos a una distancia de **1 cm** del PCB para que sean más visibles.

8

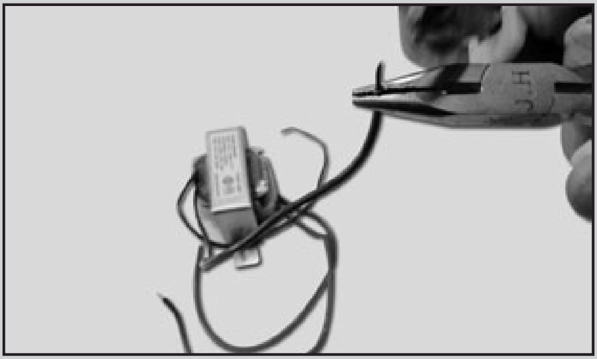

Conecte el transformadora a un interruptor y al cable para conectar a la toma de red eléctrica. Puede agregar un fusible en caso de cortocircuito. Agregue unos pines terminales al PCB para conectar las salidas de tensión a los bornes de un gabinete que escoja para la fuente. Finalice todas las conexiones según el esquemático.

www.FreeLibros.me

PASO A PASO /1 (cont.)

9

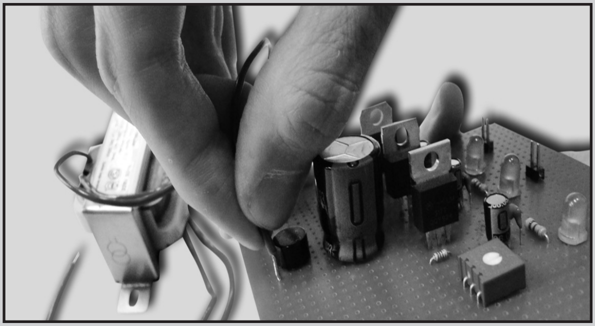

Conecte el secundario del transformador al PCB. Verifique con la ayuda de un multímetro configurado para medir continuidad, si no existe algún cortocircuito y si las conexiones están correctas. Preste especial atención al primario del transformador, que sea efectivamente el primario de este y no el secundario. Tome todos los recaudos antes de que su ansiedad lo lleve a conectar precipitadamente la fuente.

10

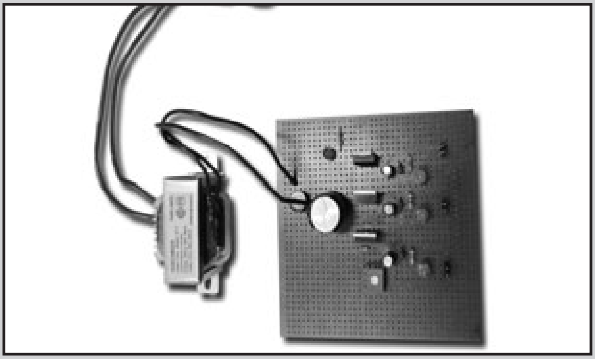

Ahora tiene la fuente armada, los componentes montados al PCB y el transformador con sus cables rojos listos para enchufar a la red eléctrica. Luego, solo tiene que encender el interruptor. Si nota algo fuera de lo común, como humo o un ruido anormal, desconecte inmediatamente todo y revise con detenimiento todo el circuito. Si todo funciona bien, deberá ver iluminarse los LEDs.

www.FreeLibros.me

PASO A PASO /1 (cont.)

11

Conecte el multímetro a la salida de **3,3 V** para medir voltaje y, con un destornillador, ajuste el preset para obtener exactamente **3,3 V** en la pantalla del multímetro. Verifique que las salidas de **5** y **12 V** indiquen esos voltajes o aproximados **+/-0,1 V**.

En este último capítulo, estudiamos las características principales de las fuentes lineales y conmutadas conectadas a la red domiciliaria. Además analizamos los bloques constitutivos de una fuente de alimentación y sus características fundamentales.
Realizamos el diseño de una fuente de alimentación simple, partiendo de las características de operación de un circuito electrónico. Vimos el armado de una fuente de taller que nos proveerá de las tensiones más comunes para trabajar con proyectos de Electrónica Digital.

www.FreeLibros.me

Mod_Cap6_xxx_xxx.qxp 04/07/2011 16:41 PÁgina 176

Multiple **choice**

1 ¿Cómo se denomina la ondulación residual que permanece luego del proceso de rectificación y filtrado?

a- Regulación de carga.
b- Estabilidad.
c- Eficiencia.
d- Ripple.

2 ¿Cómo se denomina lo que mide la variación de línea en la salida nominal de la fuente cuando ocurre una variación en la tensión de entrada?

a- Regulación de línea.
b- Regulación de carga.
c- Estabilidad.
d- Eficiencia.

3 ¿Cómo se denomina la variación de carga en la salida nominal de la fuente al estar sometida a variaciones en la carga conectada a ella?

a- Regulación de línea.
b- Regulación de carga.
c- Estabilidad.
d- Eficiencia.

4 ¿Cómo se denomina la variación de la tensión continua de salida respecto a las variaciones de temperatura, ya sea del ambiente o debido al calentamiento interno de la fuente?

a- Regulación de línea.
b- Regulación de carga.
c- Estabilidad.
d- Eficiencia.

5 ¿Cómo se llama la medida de cuánta energía de la que utiliza la fuente para cumplir su función es realmente entregada a la carga?

a- Regulación de línea.
b- Regulación de carga.
c- Estabilidad.
d- Eficiencia.

6 ¿Cómo es el costo de las fuentes de alimentación respecto de las que si tienen?

a- Más del doble.
b- Menos del doble.
c- Igual.
d- Menor.

Respuestas: 1 d, 2 a, 3 b, 4 c, 5 d, 6 d.

www.FreeLibros.me

Apendice_Electronica.qxp 04/07/2011 16:42 PÆgina 177

Apéndice

El mercado laboral

En este apéndice,
veremos las principales salidas laborales
que nos brinda la Electrónica Digital.

www.FreeLibros.me

El mercado **laboral**

La electrónica brinda en la actualidad un abanico de posibilidades laborales muy diverso. El alcance del mercado laboral hoy en día es inabarcable, ya que siempre existirá la industria que necesite digitalizar sus máquinas, desarrollar nuevas creaciones electrónicas o adaptar distintas tecnologías entre sí. Cada día es más frecuente encontrarnos en la industria o en el hogar con máquinas y dispositivos conectados a Internet. En el sector agropecuario se han realizado importantes adelantos, añadiendo unidades de control electrónico para optimizar el desempeño de las máquinas de siembra y cosecha. Muchas de estas innovaciones han sido creadas por hobbistas o técnicos electrónicos. Si bien el mercado actual y la masificación de los dispositivos electrónicos redujeron considerablemente sus costos y, en muchos casos, los hicieron descartables, existirán siempre aplicaciones que justifiquen la intervención de un técnico electrónico. Estas aplicaciones se cuentan por miles y son las que, a modo de resumen, expondremos en detalle a continuación.

La electrónica brinda en la actualidad un abanico de posibilidades laborales muy diverso

- **Ecología**: incluimos este sector ya que, con el correr de los años, será excluyente el reemplazo de antiguos sistemas eléctricos/electrónicos por sus semejantes actuales de menor consumo. En este sentido, podremos incursionar en el desarrollo de sistemas de reemplazo con los componentes actuales del mercado.

LA ELECTRÓNICA EN LA PRÁCTICA

Se efectúan diferentes tareas en la práctica diaria de la electrónica. Estas son: diagnosticar el funcionamiento de circuitos integrados, reparación y reemplazo de los pequeños componentes que forman parte de los circuitos estudiados en este libro.

www.FreeLibros.me

- **Sector industrial**: toda industria en el mundo repara y renueva anualmente sus máquinas y equipos. Estos incluyen sistemas electrónicos de medición, sensores y actuadores.
- **Sector Informático**: dispositivos de hardware de PC, sistemas UPS (*Uninterruptible Power Supply* - **Sistema Ininterrumpido de Energía**), estabilizadores de tensión, control remoto Bluetooth, placas de capturas analógicas-digitales, entre otras opciones. Todos ellos son susceptibles de ser reparados o modificados mediante la Electrónica.
- **Telemetría y Posicionamiento**: el control a distancia realizó grandes avances en las últimas décadas. Desde controles térmicos, de suelo, posicionamiento por medio de GPS (*Global Positioning System* - **Sistema de Posicionamiento Global**) o de riego remoto y mediciones de alturas en ríos. Todos estos equipos podrán ser reparados e incluso construidos y desarrollados desde cero.
- **Electromedicina**: se construyen y reparan permanentemente sistemas de baja complejidad, como incubadoras, esterilizadores y nebulizadores.
- **Seguridad**: en el campo de la seguridad, el técnico puede volcarse al diagnóstico y la reparación de alarmas, enlaces con el propietario, enlaces policiales y monitoreo.

Lo detallado hasta el momento es solo una muestra de las posibilidades laborales y de creaciones electrónicas posibles. Algunas áreas requerirán hacer un estudio particular, pero los principios electrónicos serán siempre los mismos. Por ejemplo,

Integrar significa, en términos electrónicos, vincular o unir dos o más dispositivos electrónicos entre sí

SECTOR AUTOMOTRIZ

Además de todos los sectores que detallamos, no debemos olvidarnos de la industria automotriz, en donde los sistemas electrónicos tomaron gran relevancia debido a las múltiples modificaciones que se realizaron en términos de controles de manejo, confort y motorización.

www.FreeLibros.me

Apendice_Electronica.qxp 04/07/2011 16:42 PÆgina 180

una compuerta siempre será una compuerta, pero dependiendo del uso que le demos, será la acción que realice en cada caso.

INTEGRACIÓN

Integrar significa, en términos electrónicos, vincular o unir dos o más dispositivos electrónicos entre sí, pero que cumplen funciones perfectamente definidas por separado. Hoy en día, existen miles de circuitos electrónicos con aplicaciones específicas a precios muy económicos. En muchas ocasiones, será conveniente integrar estos circuitos a nuestras necesidades en lugar de desarrollarlos y fabricarlos desde cero, ya que seguramente el costo y el tiempo resultante serán inaceptables.

Como ejemplo, supongamos que disponemos de un sistema de control de temperatura en un horno, que activa una **alarma visual** (luz) cada vez que se supera una determinada temperatura. Nosotros deseamos agregarle otro sistema para que, además, envíe un mensaje de texto **SMS** a un teléfono móvil cuando esta condición ocurra. Para ello, deberemos integrar, al sistema del horno, nuestro dispositivo SMS que permita la conexión a Internet y pueda enviar el mensaje a través de la prestadora de telefonía móvil, que dependerá de cada país en particular.

www.FreeLibros.me

Servicios **al lector**

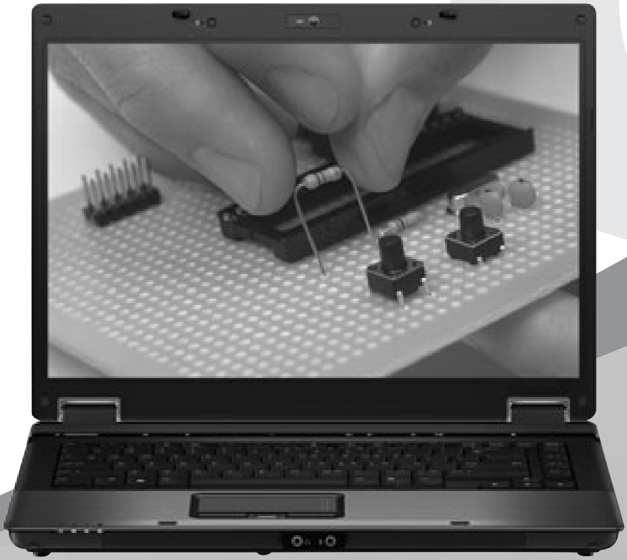

Encontraremos información adicional relacionada con el contenido que servirá para complementar lo aprendido.

RedUSERS.com

www.FreeLibros.me

Índice **temático**

A

Aislación galvánica 147
Alicate de corte al ras 20
Alicate 18
Altium designer 87
Amperímetro analógico 60
Amplitud 76
Ancho de banda 73
Ares 89

B

Banco de trabajo 16/17
Barras de acceso rápido 92
Barras de funciones y ventanas 91
Baterías 145
Bomba succionadora de estaño 20

C

Cable UTP 30
Cableado de dispositivos 98
Cajones 16
Calibración 74/75
Canal 30
Cargador de baterías Ni-Cd 65/66/67
Chupador de estaño 51
Cinta desoldante 20
Circuito impreso en detalle 44/45
Circuito impreso universal 39/42/43
Clips de masa 75
Componentes discretos 14
Componentes primitivos 14
Comprobaciones de seguridad 66
Control de errores 84/85
Creación de librerías 89/90

D

D'Arsonval 56/57
Descarga a tierra 15/18
Desoldado 51
Destornillador 19
Diseño de una PCB 93
Disparo 73

E

Eficiencia 146
Encapsulados 105
Esquema circuital 27
Estabilidad de temperatura 146
Estaño 20/50

F

Falsos contactos 31
Fase 76
Figuras de Lissajous 71
Flux 20
Flyback 148/149
Footprints 27
Fuentes conmutadas 145/148
Fuentes lineales 145/148
Función de ruteo automático 92

www.FreeLibros.me

G

GEDA 87
Herramienta Re*centre Display 92
Herramienta Zoom 92

I

IEC 485 56
Índice de clase 60
Instrumentos de simulación 98
Interfaz de los módulos 86/
Interruptor térmico bipolar 17

K

Karl Ferninand Braun 73

L

Lámpara con lupa 16
Líneas de alimentación 30

M

Manejo de capas 93
Masa circuital 30
Matafuego 18
Medición de frecuencia 77
Medición de tensión 77
Mediciones de forma de onda 76
Módulo de CAD 84
Multímetro digital 61/62/63/64/65

O

Orcad 86
Osciloscopio analógico 71
Osciloscopio de almacenamiento digital 71
Osciloscopio de fosforo digital 72
Osciloscopio 70/71/72/73/74/75/76

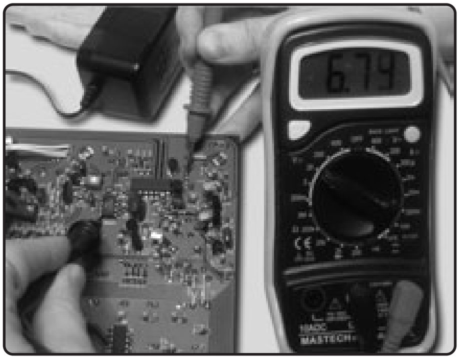

P

Pads 39
Pasta para soldar 20
Perfboard 39
Pinza bruselas 19
Pinza de fuerza 18
Pinza de punta 18
Pistas 39
Pspice A/D 86
Pulsera antiestática 17
Punta lógica 22
PWM 149

www.FreeLibros.me

N

Rectificador de media onda 153
Rectificador de onda completa 154
Rectificador puente de Graetz 155
Rectificador 152
Reglas de ruteo 103
Regulación de carga 146
Regulación de línea 146
Regulación de salida 147
Retorno de corriente 30
Ripple 146/161/162
Ruteo 27

S

Señales digitales 12
Simulación de código 112/113
Simulador del cargador de baterías 99/100/101/102/103/107/108/109/110/111
Simulink 86
Sistema métrico 104/105
Soldadura para electrónica 45/46/47/48/49
Soldante 20
Sonda 22
Stripboard 39
Surco central 30

T

Taller 16/17/18
Toma-corriente 17/18

U

ULP 86
UPCB 43

V

Ventana de vista completa 92
Vías 39/42
Virtual System Modeling 86

www.FreeLibros.me

CLAVES PARA COMPRAR
UN LIBRO DE COMPUTACIÓN

1 SOBRE EL AUTOR Y LA EDITORIAL

Revise que haya un cuadro "sobre el autor", en el que se informe sobre su experiencia en el tema. En cuanto a la editorial, es conveniente que sea especializada en computación.

2 PRESTE ATENCIÓN AL DISEÑO

Compruebe que el libro tenga guías visuales, explicaciones paso a paso, recuadros con información adicional y gran cantidad de pantallas. Su lectura será más ágil y atractiva que la de un libro de puro texto.

3 COMPARE PRECIOS

Suele haber grandes diferencias de precio entre libros del mismo tema; si no tiene el valor en tapa, pregunte y compare.

4 ¿TIENE VALORES AGREGADOS?

Desde un sitio exclusivo en la Red, un Servicio de Atención al Lector, la posibilidad de leer el sumario en la Web para evaluar con tranquilidad la compra, y hasta la presencia de adecuados índices temáticos, todo suma al valor de un buen libro.

5 VERIFIQUE EL IDIOMA

No solo el del texto; también revise que las pantallas incluidas en el libro estén en el mismo idioma del programa que usted utiliza.

usershop.redusers.com
VISITE NUESTRO SITIO WEB

- » Vea información más detallada sobre cada libro de este catálogo.
- » Obtenga un capítulo gratuito para evaluar la posible compra de un ejemplar.
- » Conozca qué opinaron otros lectores.
- » Compre los libros sin moverse de su casa y con importantes descuentos.
- » Publique su comentario sobre el libro que leyó.
- » Manténgase informado acerca de las últimas novedades y los próximos lanzamientos.

TAMBIÉN PUEDE CONSEGUIR NUESTROS LIBROS EN KIOSCOS O PUESTOS DE PERIÓDICOS, LIBRERÍAS, CADENAS COMERCIALES, SUPERMERCADOS Y CASAS DE COMPUTACIÓN.

LLEGAMOS A TODO EL MUNDO VÍA OCA * Y DHL **

* SOLO VÁLIDO EN LA REPÚBLICA ARGENTINA // ** VÁLIDO EN TODO EL MUNDO EXCEPTO ARGENTINA

usershop.redusers.com // usershop@redusers.com

www.FreeLibros.me

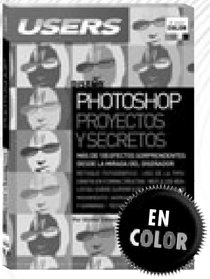
USERS
PHOTOSHOP
PROYECTOS
Y SECRETOS
EN COLOR

Photoshop: proyectos y secretos

En esta obra aprenderemos a utilizar Photoshop, desde la original mirada de la autora. Con el foco puesto en la comunicación visual, a lo largo del libro adquiriremos conocimientos teóricos, al mismo tiempo que avanzaremos sobre la práctica, con todos los efectos y herramientas que ofrece el programa.

→ COLECCIÓN: MANUALES USERS
→ 320 páginas / 978-987-1773-25-1

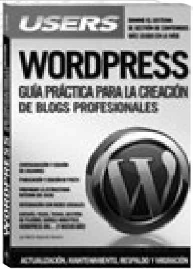
USERS
WORDPRESS
GUÍA PRÁCTICA PARA LA CREACIÓN DE BLOGS PROFESIONALES

WordPress

Este manual está dirigido a todos aquellos que quieran presentar sus contenidos o los de sus clientes a través de WordPress. En sus páginas el autor nos enseñará desde cómo llevar adelante la administración del blog hasta las posibilidades de interacción con las redes sociales.

→ COLECCIÓN: MANUALES USERS
→ 352 páginas / 978-987-1773-18-3

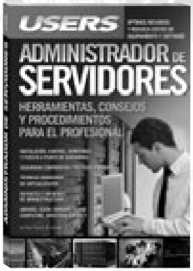
USERS
ADMINISTRADOR DE SERVIDORES
HERRAMIENTAS, CONSEJOS Y PROCEDIMIENTOS PARA EL PROFESIONAL

Administrador de servidores

Este libro es la puerta de acceso para ingresar en el apasionante mundo de los servidores. Aprenderemos desde los primeros pasos sobre la instalación, configuración, seguridad y virtualización; todo para cumplir el objetivo final de tener el control de los servidores en la palma de nuestras manos.

→ COLECCIÓN: MANUALES USERS
→ 352 páginas / ISBN 978-987-1773-19-0

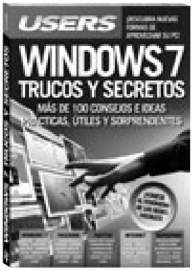
USERS
WINDOWS 7
TRUCOS Y SECRETOS
MÁS DE 100 CONSEJOS E IDEAS PRÁCTICAS, ÚTILES Y SORPRENDENTES

Windows 7: Trucos y secretos

Este libro está dirigido a todos aquellos que quieran sacar el máximo provecho de Windows 7, las redes sociales y los dispositivos ultraportátiles del momento. A lo largo de sus páginas, el lector podrá adentrarse en estas tecnologías mediante trucos inéditos y consejos asombrosos.

→ COLECCIÓN: MANUALES USERS
→ 352 páginas / ISBN 978-987-1773-17-6

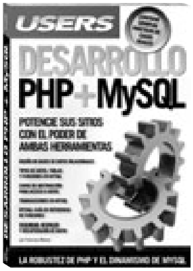
USERS
DESARROLLO
PHP + MySQL
POTENCIE SUS SITIOS CON EL PODER DE AMBAS HERRAMIENTAS

Desarrollo PHP + MySQL

Este libro presenta la fusión de dos de las herramientas más populares para el desarrollo de aplicaciones web de la actualidad: PHP y MySQL. En sus páginas, el autor nos enseñará las funciones del lenguaje, de modo de tener un acercamiento progresivo, y aplicar lo aprendido en nuestros propios desarrollos.

→ COLECCIÓN: MANUALES USERS
→ 432 páginas / ISBN 978-987-1773-16-9

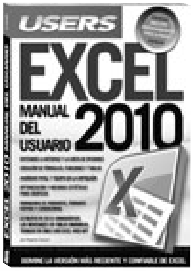
USERS
EXCEL
MANUAL DEL USUARIO
2010

Excel 2010

Este manual resulta ideal para quienes se inician en el uso de Excel, así como también para los usuarios que quieran conocer las nuevas herramientas que ofrece la versión 2010. La autora nos enseñará desde cómo ingresar y proteger datos hasta la forma de imprimir ahorrando papel y tiempo.

→ COLECCIÓN: MANUALES USERS
→ 352 páginas / ISBN 978-987-1773-15-2

www.FreeLibros.me

¡Léalo antes Gratis!

En nuestro sitio, obtenga GRATIS un capítulo del libro de su elección antes de comprarlo.

USERS
TÉCNICO HARDWARE
EN COLOR

Técnico Hardware

Esta obra es fundamental para ganar autonomía al momento de reparar la PC. Aprenderemos a diagnósticar y solucionar las fallas, así como a prevenirlas a través del mantenimiento adecuado, todo explicado en un lenguaje práctico y sencillo.

→ COLECCIÓN: MANUALES USERS
→ 320 páginas / ISBN 978-987-1773-14-5

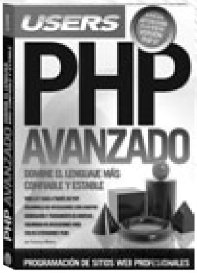
USERS
PHP
AVANZADO
PROGRAMACIÓN DE SITIOS WEB PROFESIONALES

PHP Avanzado

Este libro brinda todas las herramientas necesarias para acercar al trabajo diario del desarrollador los avances más importantes incorporados en PHP 6. En sus páginas, repasaremos todas las técnicas actuales para potenciar el desarrollo de sitios web.

→ COLECCIÓN: MANUALES USERS
→ 400 páginas / ISBN 978-987-1773-07-7

USERS
AutoCAD
GUÍA PRÁCTICA DE APRENDIZAJE

AutoCAD

Este manual nos presenta un recorrido exhaustivo por el programa más difundido en dibujo asistido por computadora a nivel mundial, en su versión 2010. En sus páginas, aprenderemos desde cómo trabajar con dibujos predeterminados hasta la realización de objetos 3D.

→ COLECCIÓN: MANUALES USERS
→ 384 páginas / ISBN 978-987-1773-06-0

USERS
WINDOWS
SE7EN
AVANZADO

Windows Seven Avanzado

Esta obra nos presenta un recorrido exhaustivo que nos permitirá acceder a un nuevo nivel de complejidad en el uso de Windows 7. Todas las herramientas son desarrolladas con el objetivo de acompañar al lector en el camino para ser un usuario experto.

→ COLECCIÓN: MANUALES USERS
→ 352 páginas / ISBN 978-987-1773-08-4

USERS
PHOTOSHOP
TÉCNICAS DE MANIPULACIÓN Y RETOQUE
EN COLOR

Photoshop

En este libro aprenderemos sobre las más novedosas técnicas de edición de imágenes en Photoshop. El autor nos presenta de manera clara y práctica todos los conceptos necesarios, desde la captura digital hasta las más avanzadas técnicas de retoque.

→ COLECCIÓN: MANUALES USERS
→ 320 páginas / ISBN 978-987-1773-05-3

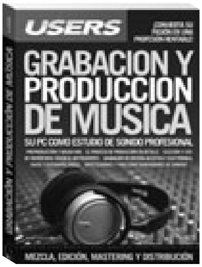
USERS
GRABACION Y PRODUCCION DE MUSICA
SU PC COMO ESTUDIO DE SONIDO PROFESIONAL
MEZCLA, EDICIÓN, MASTERING Y DISTRIBUCIÓN

Grabación y producción de música

En este libro repasaremos todos los aspectos del complejo mundo de la producción musical. Desde las cuestiones para tener en cuenta al momento de la composición, hasta la mezcla y el masterizado, así como la distribución final del producto.

→ COLECCIÓN: MANUALES USERS
→ 320 páginas / ISBN 978-987-1773-04-6

✉ usershop@redusers.com

www.FreeLibros.me

usershop.redusers.com

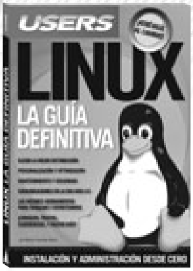

Linux

Este libro es una completa guía para migrar e iniciarse en el fascinante mundo del software libre. En su interior, el lector conocerá las características de Linux, desde su instalación hasta las opciones de entretenimiento, con todas las ventajas de seguridad que ofrece el sistema.

→ COLECCIÓN: MANUALES USERS
→ 320 páginas / ISBN 978-987-26013-8-6

Premiere + After Effects

Esta obra nos presenta un recorrido detallado por las aplicaciones audiovisuales de Adobe: Premiere Pro, After Effects y Soundbooth. Todas las técnicas de los profesionales, desde la captura de video hasta la creación de efectos, explicadas de forma teórica y práctica.

→ COLECCIÓN: MANUALES USERS
→ 320 páginas / ISBN 978-987-26013-9-3

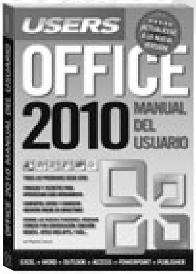

Office 2010

En este libro aprenderemos a utilizar todas las aplicaciones de la suite, en su versión 2010. Además, su autora nos mostrará las novedades más importantes, desde los minigráficos de Excel hasta Office Web Apps, todo presentado en un libro único.

→ COLECCIÓN: MANUALES USERS
→ 352 páginas / ISBN 978-987-26013-6-2

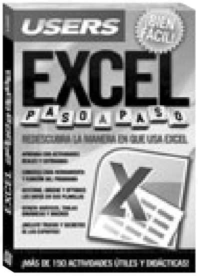

Excel Paso a Paso

En esta obra encontraremos una increíble selección de proyectos pensada para aprender mediante la práctica la forma de agilizar todas las tareas diarias. Todas las actividades son desarrolladas en procedimientos paso a paso de una manera didáctica y fácil de comprender.

→ COLECCIÓN: PASO A PASO
→ 320 páginas / ISBN 978-987-26013-4-8

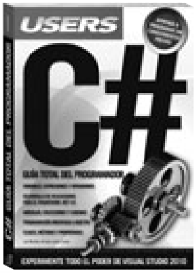

C#

Este libro es un completo curso de programación con C# actualizado a la versión 4.0. Ideal tanto para quienes desean migrar a este potente lenguaje, como para quienes quieran aprender a programar desde cero en Visual Studio 2010.

→ COLECCIÓN: MANUALES USERS
→ 400 páginas / ISBN 978-987-26013-5-5

200 Respuestas Seguridad

Esta obra es una guía básica que responde, en forma visual y práctica, a todas las preguntas que necesitamos contestar para conseguir un equipo seguro. Definiciones, consejos, claves y secretos, explicados de manera clara, sencilla y didáctica.

→ COLECCIÓN: MANUALES USERS
→ 320 páginas / ISBN 978-987-26013-1-7

www.FreeLibros.me

¡Léalo antes Gratis!

En nuestro sitio, obtenga GRATIS un capítulo del libro de su elección antes de comprarlo.

USERS
FUNCIONES EN EXCEL
GUÍA PRÁCTICA DE USO Y APLICACIÓN

Funciones en Excel

Este libro es una guía práctica de uso y aplicación de todas las funciones de la planilla de cálculo de Microsoft. Desde las funciones de siempre hasta las más complejas, todas presentadas a través de ejemplos prácticos y reales.

→ COLECCIÓN: 200 RESPUESTAS
→ **368** páginas / ISBN 978-987-26013-0-0

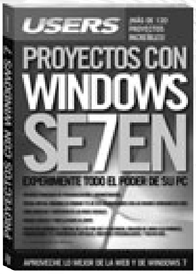
USERS
PROYECTOS CON WINDOWS SE7EN
EXPERIMENTE TODO EL PODER DE SU PC

Proyectos con Windows 7

En esta obra aprenderemos cómo aprovechar al máximo todas las ventajas que ofrece la PC. Desde cómo participar en las redes sociales hasta las formas de montar una oficina virtual, todo presentado en 120 proyectos únicos.

→ COLECCIÓN: MANUALES USERS
→ **352** páginas / ISBN 978-987-663-036-8

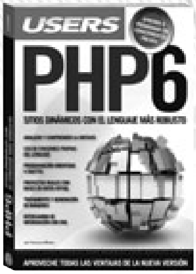

PHP 6

Este libro es un completo curso de programación en PHP en su versión 6.0. Un lenguaje que se destaca tanto por su versatilidad como por el respaldo de una amplia comunidad de desarrolladores, que lo convierten en un punto de partida ideal para quienes comienzan a programar.

→ COLECCIÓN: MANUALES USERS
→ **368** páginas / ISBN 978-987-663-039-9

USERS
200 Respuestas
Blogs
¡No requiere conocimientos previos!

200 Respuestas: Blogs

Esta obra es una completa guía que responde a las preguntas más frecuentes de la gente sobre la forma de publicación más poderosa de la Web 2.0. Definiciones, consejos, claves y secretos, explicados de manera clara, sencilla y didáctica.

→ COLECCIÓN: 200 RESPUESTAS
→ **320** páginas / ISBN 978-987-663-037-5

USERS
HARDWARE
PASO A PASO
EN COLOR

Hardware paso a paso

En este libro encontraremos una increíble selección de actividades que abarcan todos los aspectos del hardware. Desde la actualización de la PC hasta el overclocking de sus componentes, todo en una presentación nunca antes vista, realizada íntegramente con procedimientos paso a paso.

→ COLECCIÓN: PASO A PASO
→ **320** páginas / ISBN 978-987-663-034-4

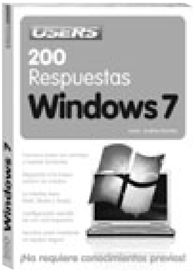
USERS
200 Respuestas
Windows 7
¡No requiere conocimientos previos!

200 Respuestas: Windows 7

Esta obra es una guía básica que responde, en forma visual y práctica, a todas las preguntas que necesitamos conocer para dominar la última versión del sistema operativo de Microsoft. Definiciones, consejos, claves y secretos, explicados de manera clara, sencilla y didáctica.

→ COLECCIÓN: 200 RESPUESTAS
→ **320** páginas / ISBN 978-987-663-035-1

✉ usershop@redusers.com

www.FreeLibros.me

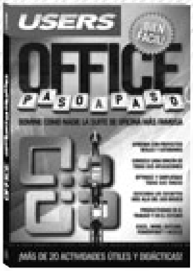
USERS
OFFICE
PASO A PASO

Office paso a paso

Este libro presenta una increíble colección de proyectos basados en la suite de oficina más usada en el mundo. Todas las actividades son desarrolladas con procedimientos paso a paso de una manera didáctica y fácil de comprender.

→ COLECCIÓN: PASO A PASO
→ 320 páginas / ISBN 978-987-663-030-6

USERS
101 SECRETOS DE HARDWARE
LA MEJOR GUÍA DE TRUCOS Y CONSEJOS
EN COLOR

101 Secretos de Hardware

Esta obra es la mejor guía visual y práctica sobre hardware del momento. En su interior encontraremos los consejos de los expertos sobre las nuevas tecnologías, las soluciones a los problemas más frecuentes, cómo hacer overclocking, modding, y muchos más trucos y secretos.

→ COLECCIÓN: MANUALES USERS
→ 352 páginas / ISBN 978-987-663-029-0

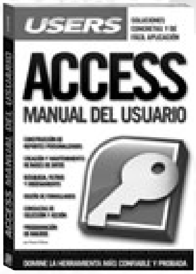
USERS
ACCESS
MANUAL DEL USUARIO

Access

Este manual nos introduce de lleno en el mundo de Access para aprender a crear y administrar bases de datos de forma profesional. Todos los secretos de una de las principales aplicaciones de Office, explicados de forma didáctica y sencilla.

→ COLECCIÓN: MANUALES USERS
→ 320 páginas / ISBN 978-987-663-025-2

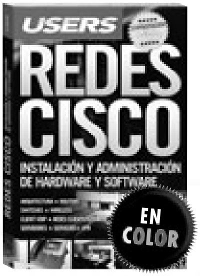
USERS
REDES CISCO
INSTALACIÓN Y ADMINISTRACIÓN DE HARDWARE Y SOFTWARE
EN COLOR

Redes Cisco

Este libro permitirá al lector adquirir todos los conocimientos necesarios para planificar, instalar y administrar redes de computadoras. Todas las tecnologías y servicios Cisco, desarrollados de manera visual y práctica en una obra única.

→ COLECCIÓN: MANUALES USERS
→ 320 páginas / ISBN 978-987-663-024-5

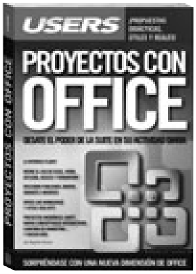
USERS
PROYECTOS CON OFFICE

Proyectos con Office

Esta obra nos enseña a usar las principales herramientas de Office a través de proyectos didácticos y útiles. En cada capítulo encontraremos la mejor manera de llevar adelante todas las actividades del hogar, la escuela y el trabajo.

→ COLECCIÓN: MANUALES USERS
→ 352 páginas / ISBN 978-987-663-023-8

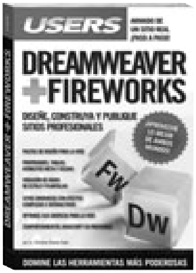
USERS
DREAMWEAVER + FIREWORKS
DISEÑE, CONSTRUYA Y PUBLIQUE SITIOS PROFESIONALES
Fw
Dw

Dreamweaver y Fireworks

Esta obra nos presenta las dos herramientas más poderosas para la creación de sitios web profesionales de la actualidad. A través de procedimientos paso a paso, nos muestra cómo armar un sitio real con Dreamweaver y Fireworks sin necesidad de conocimientos previos.

→ COLECCIÓN: MANUALES USERS
→ 320 páginas / ISBN 978-987-663-022-1

www.FreeLibros.me

¡Léalo antes Gratis!

En nuestro sitio, obtenga GRATIS un capítulo del libro de su elección antes de comprarlo.

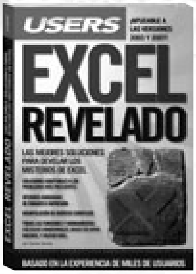

Excel revelado

Este manual contiene una selección de más de 150 consultas de usuarios de Excel y todas las respuestas de Claudio Sánchez, un reconocido experto en la famosa planilla de cálculo. Todos los problemas encuentran su solución en esta obra imperdible.

→ COLECCIÓN: MANUALES USERS
→ 336 páginas / ISBN 978-987-663-021-4

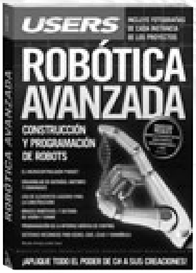

Robótica avanzada

Esta obra nos permitirá ingresar al fascinante mundo de la robótica. Desde el ensamblaje de las partes hasta su puesta en marcha, todo el proceso está expuesto de forma didáctica y sencilla para así crear nuestros propios robots avanzados.

→ COLECCIÓN: MANUALES USERS
→ 352 páginas / ISBN 978-987-663-020-7

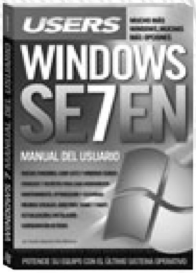

Windows 7

En este libro encontraremos las claves y los secretos destinados a optimizar el uso de nuestra PC tanto en el trabajo como en el hogar. Aprenderemos a llevar adelante una instalación exitosa y a utilizar todas las nuevas herramientas que incluye esta versión.

→ COLECCIÓN: MANUALES USERS
→ 320 páginas / ISBN 978-987-663-015-3

De Windows a Linux

Esta obra nos introduce en el apasionante mundo del software libre a través de una completa guía de migración, que parte desde el sistema operativo más conocido: Windows. Aprenderemos cómo realizar gratuitamente aquellas tareas que antes hacíamos con software pago.

→ COLECCIÓN: MANUALES USERS
→ 336 páginas / ISBN 978-987-663-013-9

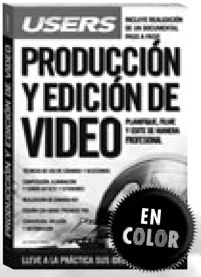

Producción y edición de video

Un libro ideal para quienes deseen realizar producciones audiovisuales con bajo presupuesto. Tanto estudiantes como profesionales encontrarán cómo adquirir las habilidades necesarias para obtener una salida laboral con una creciente demanda en el mercado.

→ COLECCIÓN: MANUALES USERS
→ 336 páginas / ISBN 978-987-663-012-2

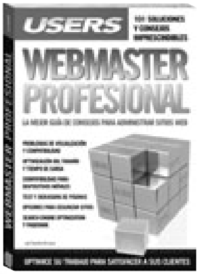

Webmaster Profesional

Esta obra explica cómo superar los problemas más frecuentes y complejos que enfrenta todo administrador de sitios web. Ideal para quienes necesiten conocer las tendencias actuales y las tecnologías en desarrollo que son materia obligada para dominar la Web 2.0.

→ COLECCIÓN: MANUALES USERS
→ 336 páginas / ISBN 978-987-663-011-5

www.FreeLibros.me

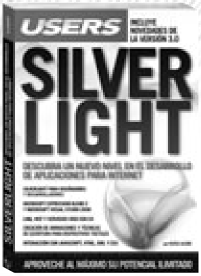
USERS
SILVER LIGHT
APROVECHE AL MÁXIMO SU POTENCIAL ILIMITADO

Silverlight

Este manual nos introduce en un nuevo nivel en el desarrollo de aplicaciones interactivas a través de Silverlight, la opción multiplataforma de Microsoft. Quien consiga dominarlo creará aplicaciones visualmente impresionantes, acordes a los tiempos de la incipiente Web 3.0.

→ COLECCIÓN: MANUALES USERS
→ 352 páginas / ISBN 978-987-663-010-8

USERS
FLASH EXTREMO
TODO EL PODER DE FLASH CS4 Y ACTIONSCRIPT 3.0
Fl
DESARROLLO DE APLICACIONES DE ALTO IMPACTO

Flash Extremo

Este libro nos permitirá aprender a fondo Flash CS4 y ActionScript 3.0 para crear aplicaciones web y de escritorio. Una obra imperdible sobre uno de los recursos más empleados en la industria multimedia que nos permitirá estar a la vanguardia del desarrollo.

→ COLECCIÓN: MANUALES USERS
→ 320 páginas / ISBN 978-987-663-009-2

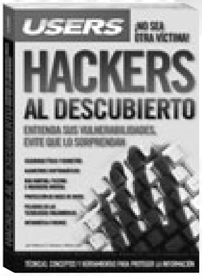
USERS
¡NO SEA OTRA VÍCTIMA!
HACKERS AL DESCUBIERTO

Hackers al descubierto

Esta obra presenta un panorama de las principales técnicas y herramientas utilizadas por los hackers, y de los conceptos necesarios para entender su manera de pensar, prevenir sus ataques y estar preparados ante las amenazas más frecuentes.

→ COLECCIÓN: MANUALES USERS
→ 352 páginas / ISBN 978-987-663-008-5

USERS
SOLUCIONES, TRUCOS Y SECRETOS
VISTA AVANZADO
EXPERIMENTA TODO SU PODER
NO SEA UN SIMPLE USUARIO, ¡DOMÍNELO!

Vista avanzado

Este manual es una pieza imprescindible para convertirnos en administradores expertos de este popular sistema operativo. En sus páginas haremos un recorrido por las herramientas fundamentales para tener máximo control sobre todo lo que sucede en nuestra PC.

→ COLECCIÓN: MANUALES USERS
→ 352 páginas / ISBN 978-987-663-007-8

USERS
101 SECRETOS DE EXCEL
LA MEJOR GUÍA DE TRUCOS Y CONSEJOS
TOP SECRET
DOMINE LA HERRAMIENTA MÁS PODEROSA DE OFFICE

101 Secretos de Excel

Una obra absolutamente increíble, con los mejores 101 secretos para dominar el programa más importante de Office. En sus páginas encontraremos un material sin desperdicios que nos permitirá realizar las tareas más complejas de manera sencilla.

→ COLECCIÓN: MANUALES USERS
→ 336 páginas / ISBN 978-987-663-005-4

USERS
NO REQUIERE CONOCIMIENTOS PREVIOS
ELECTRÓNICA & MICROCONTROLADORES PIC
GUÍA PRÁCTICA DE PROGRAMACIÓN
INCLUYE PROYECTOS REALES

Electrónica & microcontroladores PIC

Una obra ideal para quienes desean aprovechar al máximo las aplicaciones prácticas de los microcontroladores PIC y entender su funcionamiento. Un material con procedimientos paso a paso y guías visuales, para crear proyectos sin límites.

→ COLECCIÓN: MANUALES USERS
→ 368 páginas / ISBN 978-987-663-002-3

www.FreeLibros.me

CATALOGO FINAL 15x19 BN.indd 8 24/6/2011 10:52:39

USERS PRESENTA...

¡EL PRIMER EBOOK USERS!

Sí, ya podés leer Hackers al descubierto
en tu PC, notebook, Amazon
Kindle, iPad, en el celular...

USERS
¡NO SEA OTRA VÍCTIMA!
HACKERS
AL DESCUBIERTO
ENTIENDA SUS VULNERABILIDADES, EVITE QUE LO SORPRENDAN
SEGURIDAD FÍSICA Y BIOMETRÍA
ALGORITMOS CRIPTOGRÁFICOS
BUG HUNTING, FUZZING E INGENIERÍA INVERSA
PROTECCIÓN DE BASES DE DATOS
PELIGROS EN LAS TECNOLOGÍAS INALÁMBRICAS
INFORMÁTICA FORENSE
por Federico G. Pacheco y Héctor Jara
TÉCNICAS, CONCEPTOS Y HERRAMIENTAS PARA PROTEGER LA INFORMACIÓN

CONSEGUILO
DESDE CUALQUIER
PARTE DEL MUNDO

A UN PRECIO
INCREÍBLE

¿QUÉ ESTÁS
ESPERANDO?

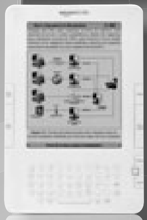

HACKERS
AL DESCUBIERTO

¡LEELO
DONDE
QUIERAS!

INGRESA YA A USERSHOP.REDUSERS.COM Y ENTERATE MÁS

www.FreeLibros.me

COLECCION FLUO - BOMBO EBOOK ULTIMA PAGINA.indd 1 08/11/2010 15:58:22

CONÉCTESE
CON LOS MEJORES
LIBROS DE COMPUTACIÓN

LLEGAMOS A TODO EL MUNDO
VÍA OCA* Y DHL**

usershop.redusers.com
usershop@redusers.com

* SÓLO VÁLIDO EN LA REPÚBLICA ARGENTINA // ** VÁLIDO EN TODO EL MUNDO EXCEPTO ARGENTINA

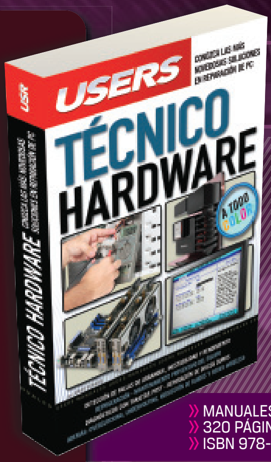
USERS
TÉCNICO HARDWARE
CONOZCA LAS MÁS NOVEDOSAS SOLUCIONES EN REPARACIÓN DE PC
A TODO COLOR

CONOZCA LAS MÁS NOVEDOSAS SOLUCIONES EN REPARACIÓN DE PC

» MANUALES USERS
» 320 PÁGINAS
» ISBN 978-987-1773-14-5

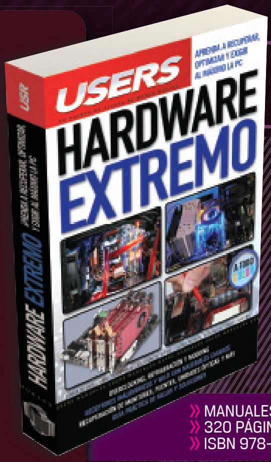
USERS
HARDWARE EXTREMO
APRENDA A RECUPERAR, OPTIMIZAR Y EXIGIR AL MÁXIMO LA PC
A TODO COLOR

EL LIBRO DE HARDWARE IDEAL PARA QUIENES NO ADMITEN LOS LÍMITES

» MANUALES USERS
» 320 PÁGINAS
» ISBN 978-987-1347-90-2

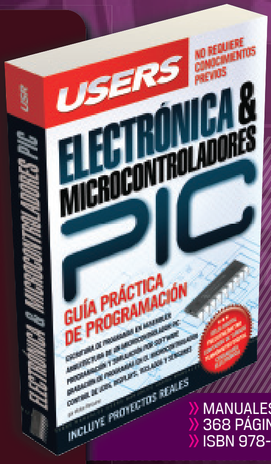
USERS
ELECTRÓNICA & MICROCONTROLADORES PIC
NO REQUIERE CONOCIMIENTOS PREVIOS
GUÍA PRÁCTICA DE PROGRAMACIÓN
INCLUYE PROYECTOS REALES

PROGRAME APLICACIONES Y APRENDA LOS MÚLTIPLES USOS DE LOS PICS

» MANUALES USERS
» 368 PÁGINAS
» ISBN 978-987-663-002-3

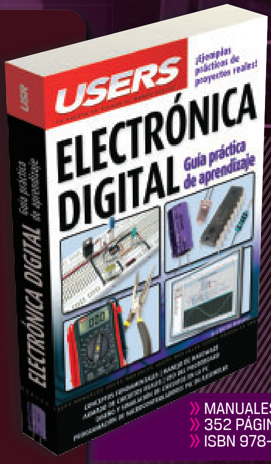
USERS
ELECTRÓNICA DIGITAL
Guía práctica de aprendizaje
¡Ejemplos prácticos de proyectos reales!

LOS MEJORES PROYECTOS PARA ARMAR CON SUS PROPIAS MANOS

» MANUALES USERS
» 352 PÁGINAS
» ISBN 978-987-1347-73-5

www.FreeLibros.me

COLECCION FLUO - BOMBO RCT - Base Editable - 15x19 - Dic 09.indd 1 07/04/2011 17:50:40

ELECTRÓNICA PRÁCTICA

En este libro recorreremos los conceptos básicos de la electrónica y sus aspectos prácticos. Entraremos en el taller y veremos en detalle las herramientas y los instrumentos imprescindibles. Además, conoceremos en profundidad el análisis, la construcción y la simulación de circuitos en la PC.

DENTRO DEL LIBRO ENCONTRARÁ

El taller electrónico ■ Las herramientas ■ Circuitos electrónicos ■ Protoboard ■ Construcción de circuitos impresos ■ Técnicas de soldado ■ Instrumentos de medición ■ Uso del multímetro digital ■ El osciloscopio ■ Diseño de circuitos impresos ■ Diseño y simulación ■ Simulación del cargador de baterías ■ Diseño de una fuente de alimentación ■ Los reguladores de voltaje

USERS
NETWORKING C
MICROCONTROLADO
USERS
PROYECTOS CO
MICROCONTROLADO
USERS
MICRO CONTROLADORES

ADEMÁS

MICROCONTROLADORES

Funcionamiento, programación
y aplicaciones prácticas

PROYECTOS CON MICROCONTROLADORES

Aprenda a desarrollar sus propias
aplicaciones

NETWORKING CON MICROCONTROLADORES

Descubra cómo acceder remotamente
a sus equipos

SOBRE LA COLECCIÓN: ELECTRÓNICA

■ Aprendizaje guiado mediante explicaciones claras y concisas ■ Proyectos prácticos basados en necesidades reales ■ Consejos de los profesionales ■ Infografías y procedimientos paso a paso ■ Producciones fotográficas profesionales

ELECTRONICS FROM SCRATCH

In this book we will learn the basics of Electronics as well as the practical aspects to bear in mind when dealing with analysis, simulation and construction of electronic circuits. Furthermore, we will be trained on how to use tools and electronic measuring instruments.

NIVEL DE USUARIO

BÁSICO	INTERMEDIO	AVANZADO	EXPERTO
🟩			

ISBN 978-987-1773-21-3

RedUSERS.com

Nuestro sitio reúne a la mayor comunidad de tecnología en América Latina. Aquí podrá comunicarse con lectores, editores y autores, y acceder a noticias, foros y blogs constantemente actualizados. Además, podrá descargar material adicional de los libros y capítulos gratuitos, o conocer nuestras otras publicaciones y acceder a comprarlas desde cualquier parte del mundo.

Si desea más información sobre el libro: Servicio de atención al lector **usershop@redusers.com**

www.FreeLibros.me

.

Beaver Baby Grand
Beaver Machine and Tool Company produced several variations of this crystal radio during the 1920s, and also apparently supplied the radio to India Ivory Company. One of the smallest radios at the time at 3-3/4 x 2 x 3/4 inches, it was advertised as The Beaver Baby Grand Vest Pocket Radio Receiving Set.

These are shown here in the order they may have been produced based on a progression of part upgrades.

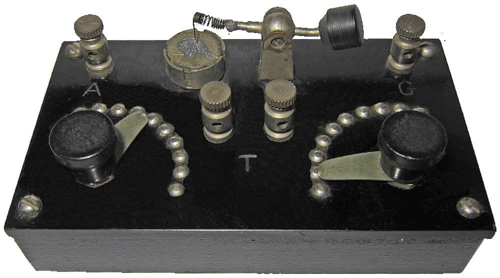

Type 1
Black painted wood case, flat top knobs, small binding posts, simple engraved panel, and no ID of any kind.

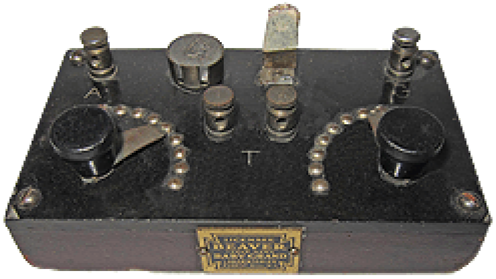

Type 2
Reddish brown colored wood, flat top knobs, small binding posts, simple engraved panel, metal tag on cabinet. (detector arm is missing)

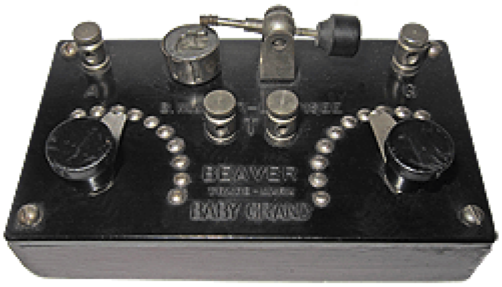

Type 3
Reddish brown colored wood, flat top knobs, small binding posts, raised lettering on panel, no metal tag.

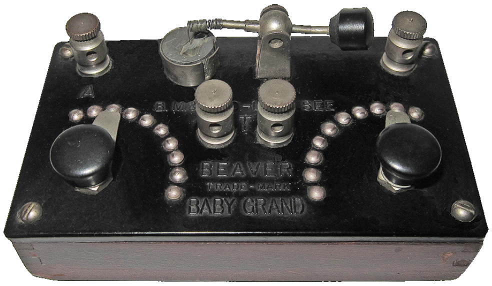

Type 4
Reddish brown colored wood, dome top knobs, large binding posts, raised lettering on panel, no metal tag.

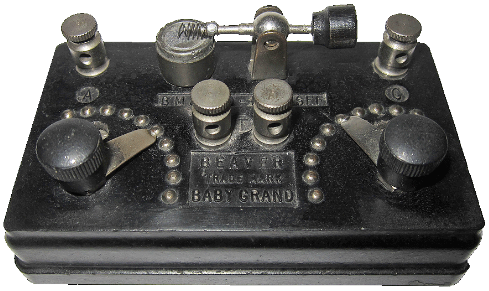

Type 5
Hard rubber cabinet, ribbed knobs, large binding posts, lettering molded into panel. A cover on the bottom protects the coil inside.

Bethlehem Radio Corporation, of Bethlehem, Pennsylvania, made these two metal crystal radios during the 1920s, and may have also made other radio equipment under the BERACO name.

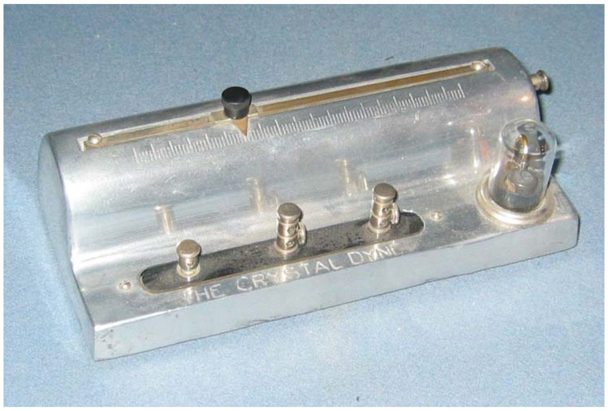

Bethlehem 'The Crystal Dyne'
It has a cast aluminum case, a glass-enclosed crystal detector, and a slide tuner. It is about 8-1/2 inches long x 4 inches deep.

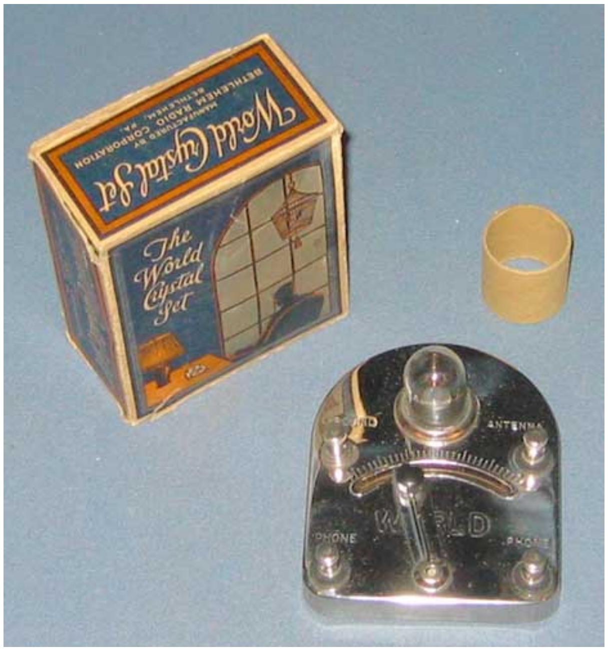

Bethlehem 'World Crystal Set'
It has a nickle-plated metal case, a glass-enclosed crystal detector, and a slide tuner. It measures about 5 x 4 inches. It is shown with its original box. These were also sold painted in various colors including red and green.

BMS Pandora Crystal Set
The Brooklyn Metal Stamping Company, of Brooklyn, New York, manufactured a number of crystal radios in the 1920s, including the 'Pandora' which came in two styles and 4 or 5 colors. This example has the coil raised up on a small pedestal and has the original Russett Bronze paint. It is in mint condition except the detector arm is missing. It is shown with its original instruction sheet and box.

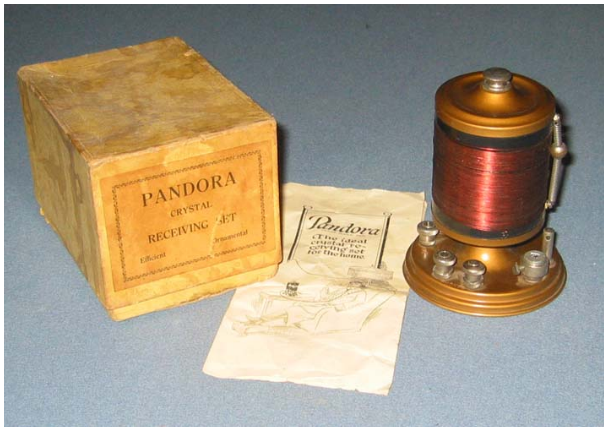

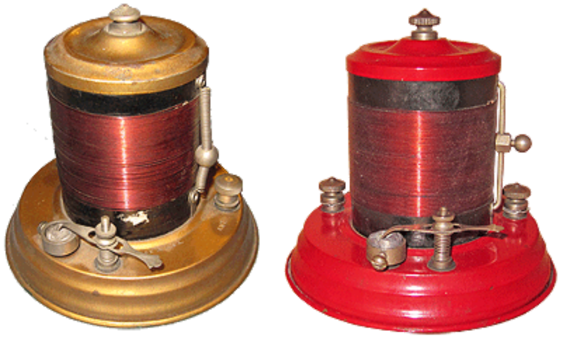

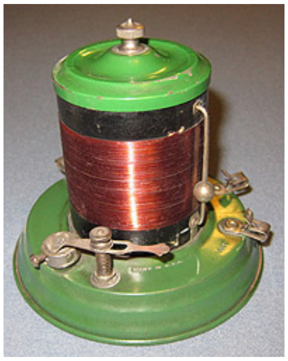

BMS Pandora Crystal Sets
These BMS Pandora sets are styled different from the one above, without the pedestal. Note that the red and green examples have a solid slider bar while the bronze one has a spring, the bronze and green examples have a ball slider while the red one has a slider with a knob.

Also, the green example has fahenstock clip connectors while the others have binding posts. The Russett Bronze, red, and Medium Green colors are all original paint.

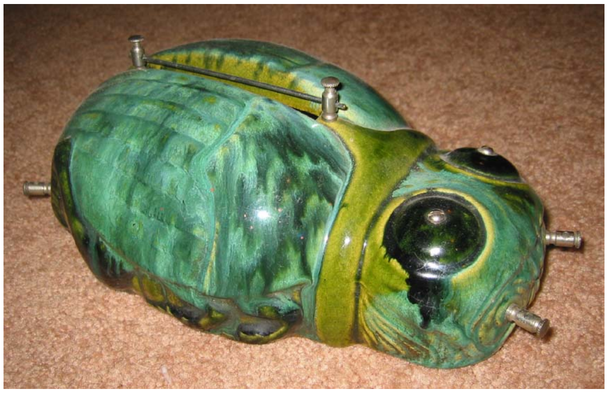

Brush-McCoy 'Mystic'
In 1927 the Brush-McCoy Pottery Company decided to produce a unique pottery item that would appeal to radio fans all over the world, who were then called "Radio Bugs." It was a ceramic bug 9" long containing a crystal radio set. These were called "MYSTIC" Radio Sets. The bug itself is pottery. An article that was in The Zanesville Signal said, "The insulating properties of this material are so high that exceptional reception results."

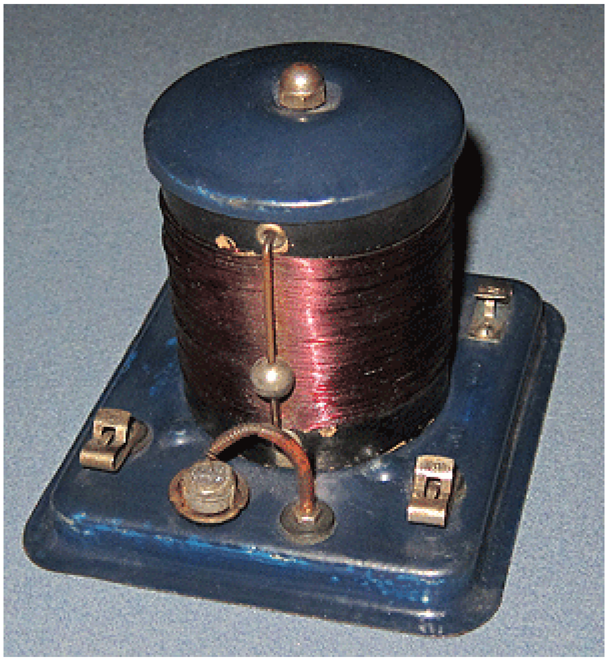

Cannon Ball 'Baby Grand'
1920s crystal radio with metal base.

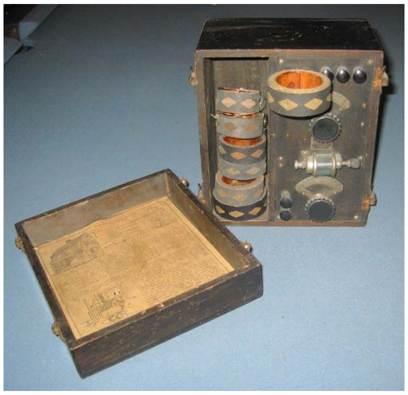

DeForest Everyman
This Deforest Everyman Radiophone Receiver, Type DT-600, was made in the 1920s and used different plug-in coils to receive different wavelengths. It has a glass-enclosed crystal detector.

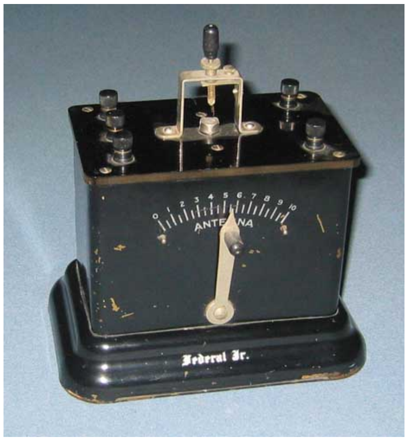

Federal 'Junior'
This Federal Jr. crystal radio was made in the 1920s. The cabinet is metal with a bakelite top. A slider control on one side is labeled 'Antenna', and a similar control on the other side is labeled 'Detector'.

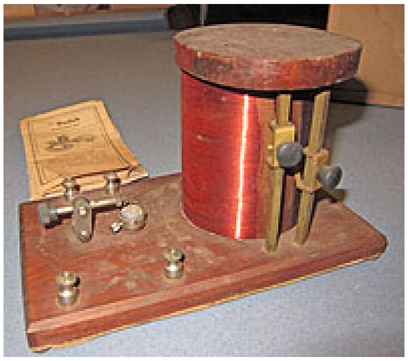

Fletcher 'Radak Wireless Set'
Made and distributed by Fletcher Electric Company at 34 West 36th Street in New York City in the early 1920s. The original operating instructions call this 'The Hand Camera of Wireless! A snappy little Radio Set! Small and light as a postcard camera'.

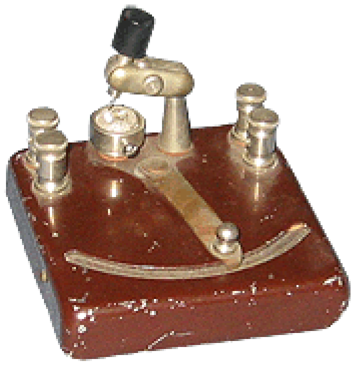

Flivver
This little Flivver crystal radio was made in the 1920s. It has a metal cabinet and is identified as a Flivver on the underside.

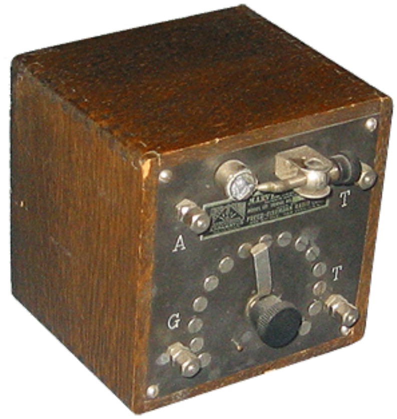

Freed Eisemann 'Marvel 101'
Freed Eisemann manufactured this crystal radio in the 1920s and housed it in a nice oak dovetailed cabinet.

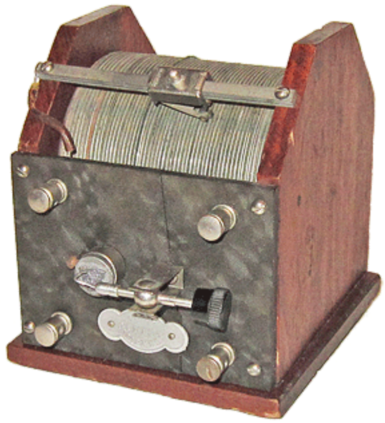

Gillet Marvel
1920s crystal radio

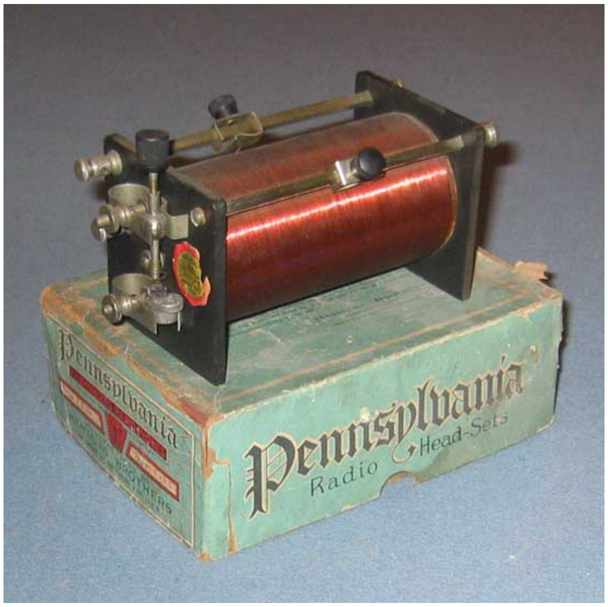

Hamburg Brothers - Pennsylvania
This little crystal radio was sold about 1926 by Hamburg Brothers of Pittsburgh, Pennsylvania, and was endorsed by Radio News magazine. It is sitting on a box of Hamburg Bros. 'Pennsylvania' headphones.

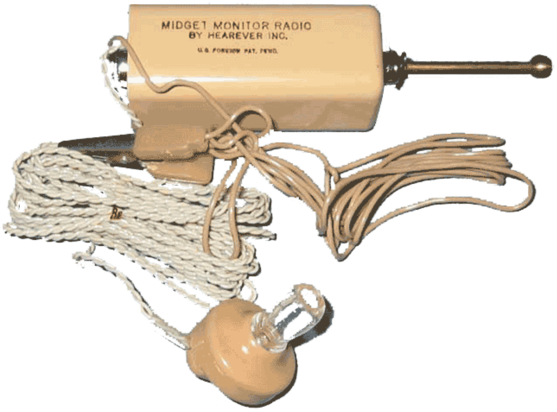

Hearever 'Midget Monitor'
This tiny crystal set was made in the 1950s or 60s and is built into a standard radio IF coil can. Sliding the antenna rod up and down tunes a station. This radio is mint in the original box, and is marked 'MIDGET MONITOR RADIO BY HEAREVER INC.' The instruction sheet gives the company address as: MIDGET MONITOR RADIO, Dept. 150, 623 Marquette Ave., Minneapolis 1, Minn.

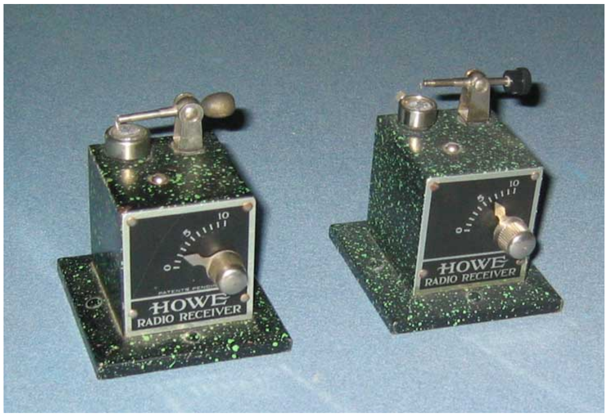

Howe: Early, Late
Two 1920s Howe crystal radios. The left one is the earlier version - it has 'Patents Pending' printed on the front plate, the detector arm has a wooden handle, and the crystal cup is larger.

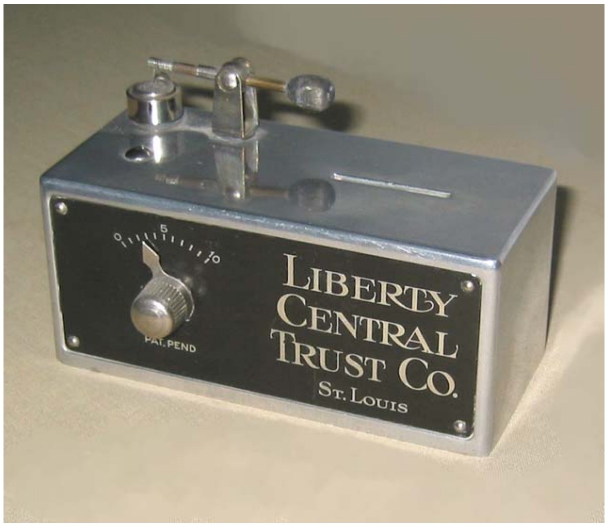

Howe - Liberty Central Trust Co.
This is a bank advertising the Liberty Central Trust

Company of St. Louis Missouri, and it has a Howe crystal radio built into it!

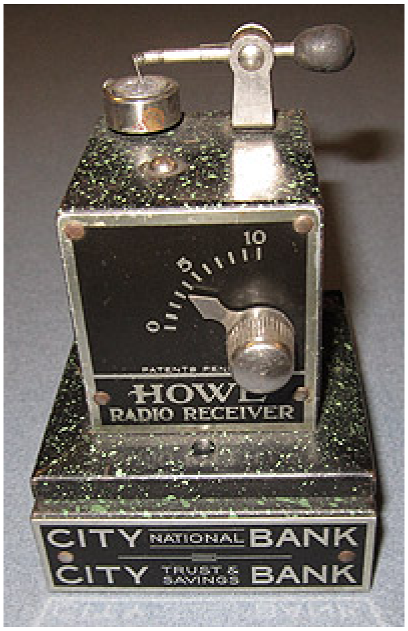

Howe - National City Bank
This is a bank advertising the National City Bank, and it has a Howe crystal radio built into it!

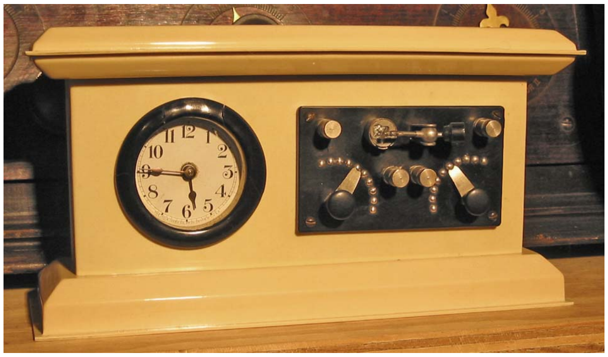

India Ivory
A rare 1920's crystal radio, only documented with a brief reference in the Maurice Siever's Crystal Set books, and a photo in Morgan McMahon's Vintage Radio book, and 1 of only 3 I've seen in collections. This off-white celluloid clock case is inscribed 'Radio Clock Case, India Ivory Co., Pat. Apl'lied For'. The paper face of the wind-up clock is marked 'India Ivory Company, Providence, R.I., U.S.A.' The radio itself looks exactly like an early version of the 1920's Beaver Baby Grand chassis.

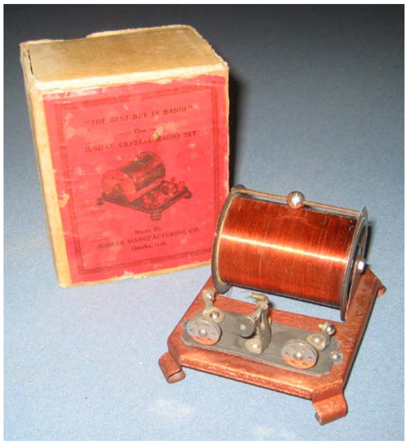

Jubilee
Jubilee Manufacturing Company, which operated in Omaha, Nebraska from 1905 until about 1988, produced this Jubilee Crystal Radio Set in the later 1920's. The red label on the original box proclaims this to be 'The Best Buy in Radio'. Unique connectors allowed up to six pair of headphones to easily be connected to the radio!

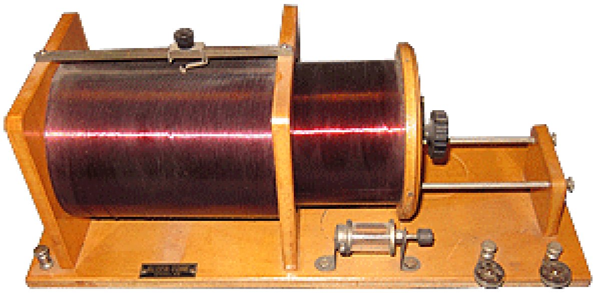

Kleer-Tone Model 100

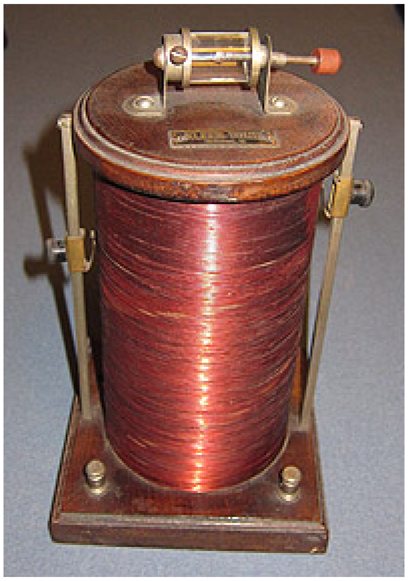

Kleer-Tone Model B

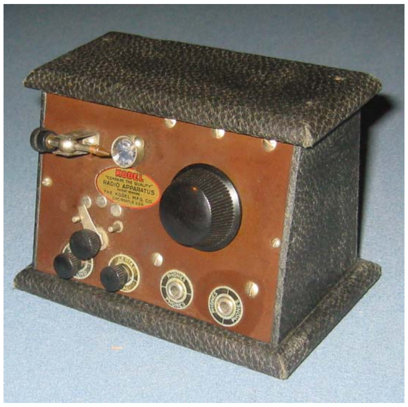

Kodel Model S-1
Kodel's smallest and lowest priced radio, this originally sold for $5.

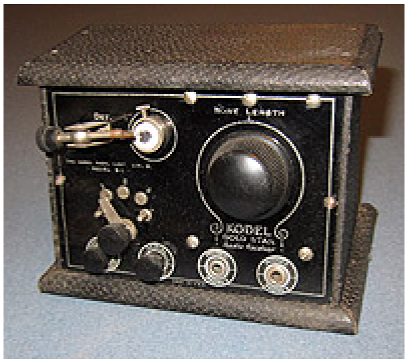
WAVE LENGTH
KODEL
GOLD STAR

Kodel Gold Star Model S-1
Kodel's Gold Star model S-1 is identical to the other version except for the front panel design.

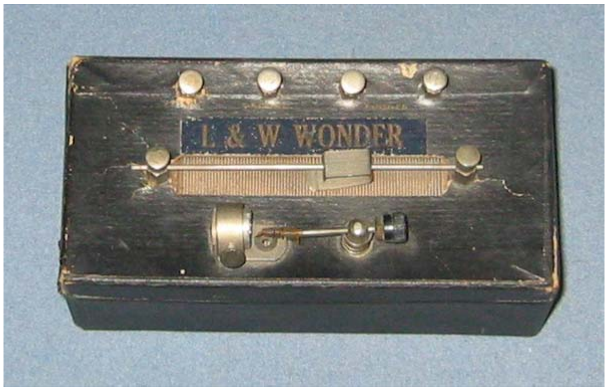

L&W Wonder
A cheaply made crystal radio in a cardboard box cabinet.

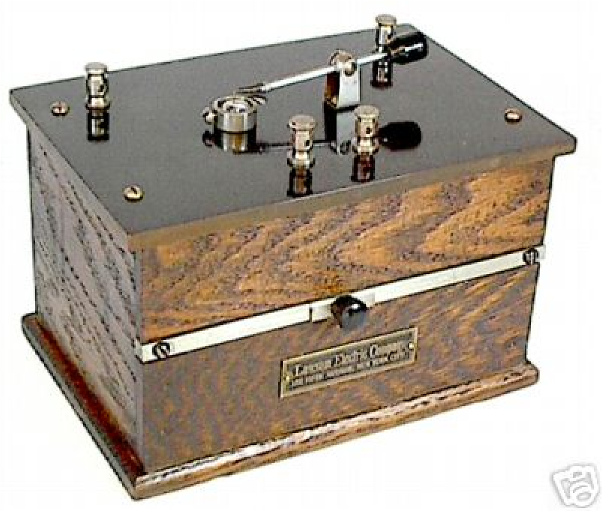
Lawsam Electric Company

Lawsam Electric Company
This 1920s crystal radio was manufactred by Lawsam Electric Company at 122 Fifth Avenue, in New York City. The recessed slider on the front of the cabinet is unique.

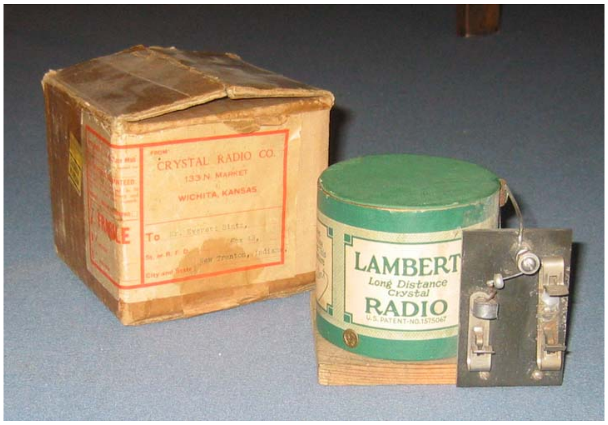
CRYSTAL RADIO CO.
133 N. MARKET
WICHITA, KANSAS
LAMBERT
Long Distance
Crystal
RADIO

Leon Lambert
This old crystal radio is shown with its original box. It has a very crudely designed detector that looks homemade, but this is original from the factory. It is identical to the wooden version, except with a less expensive cabinet.

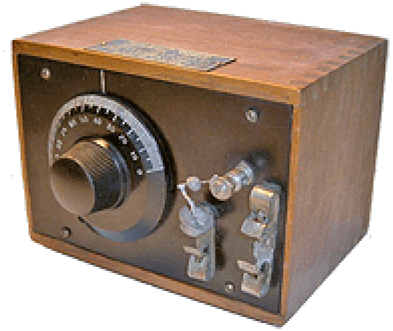

Leon Lambert
This crystal radio has a wood cabinet with a brass tag on top saying 'Lambert Radio, Leon Lambert Mfg. Co., Wichta, Kansas'. It is the deluxe version of the green Leon Lambert radio.

Lombard Multiphone
An unusual old crystal radio in a masonite cabinet

painted crackle green with cowboys on the sides. A Lombard detector is mounted below the tuning dial and there are connections for six pairs of headphones! Operating instructions are on the back side and earphone instructions are on the bottom of the cabinet.

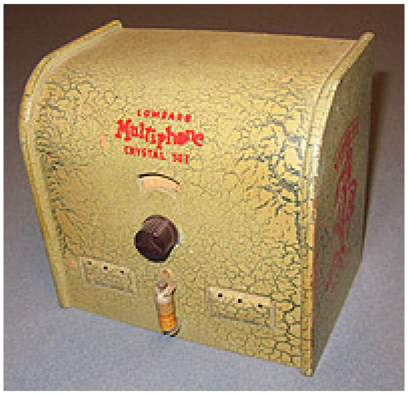
LOMBARD
Multiphone
CRYSTAL SET

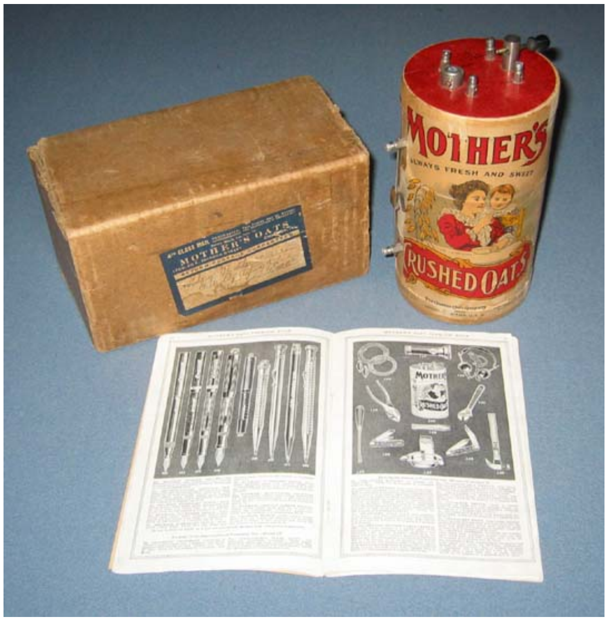
MOTHER'S
ALWAYS FRESH AND SWEET
CRUSHED OATS

Marquette 'Mother's Oats'
This 1924 Mother's Oats premium catalog offered this crystal radio for 110 coupons, or 5 coupons and $1.00. The radio was made by Marquette Radio Corporation. It is shown with its original box.

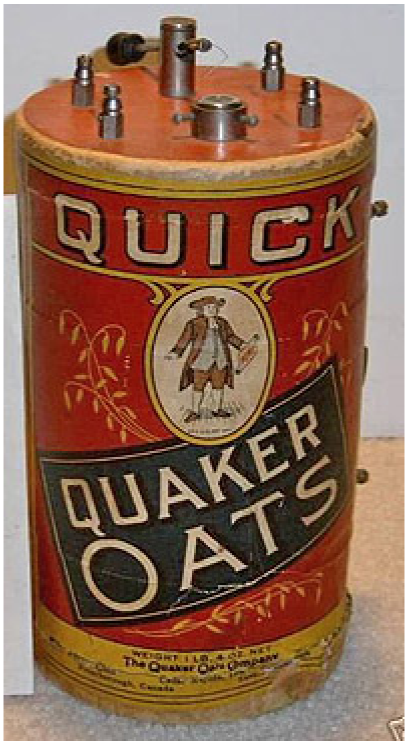
QUICK
QUAKER
OATS
WEIGHT 1 LB. 4 OZ. NET
The Quaker Oats Company

Marquette 'Quaker Oats'
This early 1920s radio is very similar to the Mothers Oats set. It was also made by Marquette Radio Corporation and was a coupon premium.

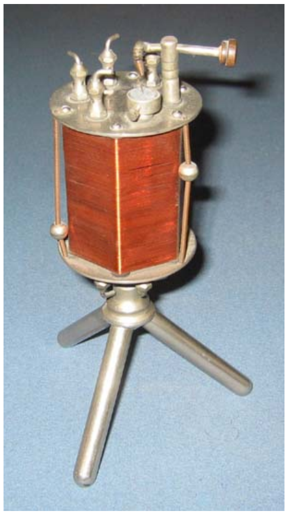

Martian 'Big 4'
This is a unique and well-made crystal radio mounted on a tripod base. It has two sliders, a crystal detector,

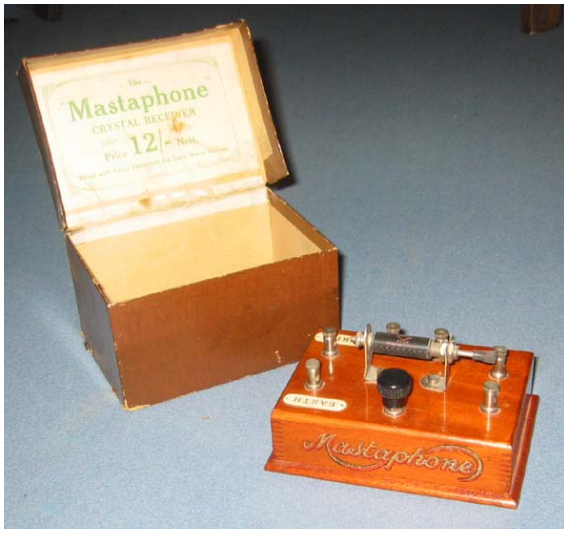
The
Mastaphone
CRYSTAL RECEIVER
Price 12/- Nett
EARTH
Mastaphone

Mastaphone
This crystal radio was made in Great Britain during the

and enough binding posts to connect four sets of headphones. It was made in Newark, New Jersey, and stands 9 inches tall.

1920s, and is shown with its original box. It has a spring-loaded enclosed cats whisker detector.

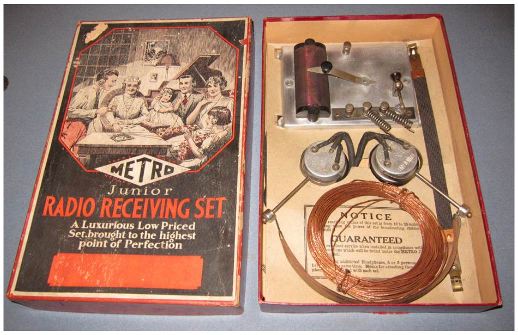
METRO
Junior
RADIO RECEIVING SET
A Luxurious Low Priced Set brought to the highest point of Perfection
NOTICE
GUARANTEED

Metro Junior
A 1920s crystal radio made by Metro Electrical Company in Newark, New Jersey. The same radio was also marketed under other names (see Pal). This is shown in its original box.
The base is aluminum and measures 6 x 5 x 1/2 inch.

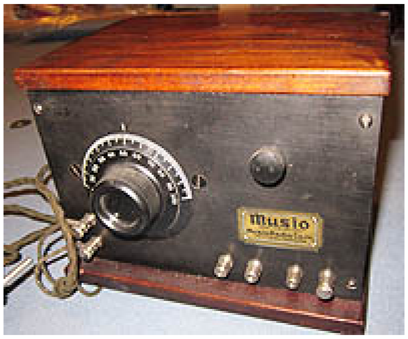

Musio
A 1920s crystal radio made by Musio Radio Company. The small knob controls the internal catwhisker. The lid is held on only by spring clips so is often missing. A pair of Musio headphones completes the set.
It measures 8-1/4 x 6-1/4 x 6 inches.

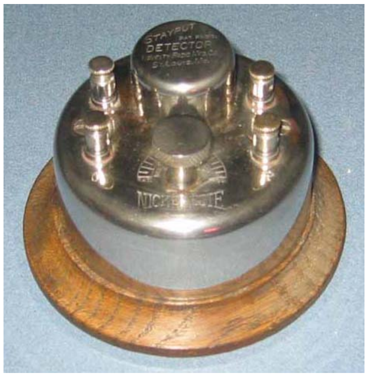
STAYPUT
DETECTOR
NICKELETTE

Novelty Radio Mfg. Co. 'Nickelette'
A 1920s crystal radio marked 'Trade Mark NICKELETTE'. The detector is under the metal cover marked 'STAYPUT DETECTOR, NOVELTY RADIO MFG. CO., ST. LOUIS, MO, PAT PND'G'. You need to remove the metal cover to adjust the wire cats whisker. Directions are on a paper label on the bottom of the radio.

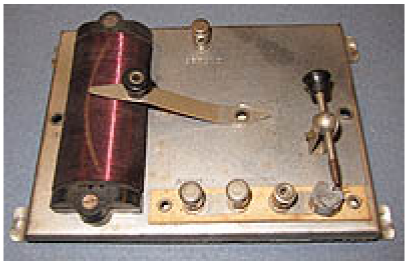

Pal
A 1920s crystal radio made by Pal Radio Company in Jersey City, New Jersey. The same radio was also marketed under other names (see Metro).
The base is aluminum, has a cardboard instruction card riveted underneath, and measures 6 x 5 x 1/2 inch.

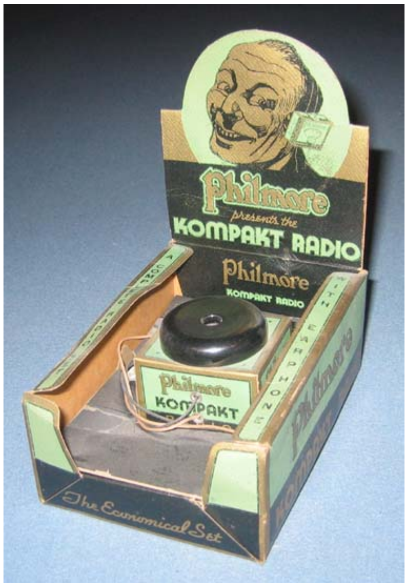
Philmore
presents the
KOMPAKT RADIO
Philmore
KOMPAKT RADIO
Philmore
KOMPAKT
The Economical Set

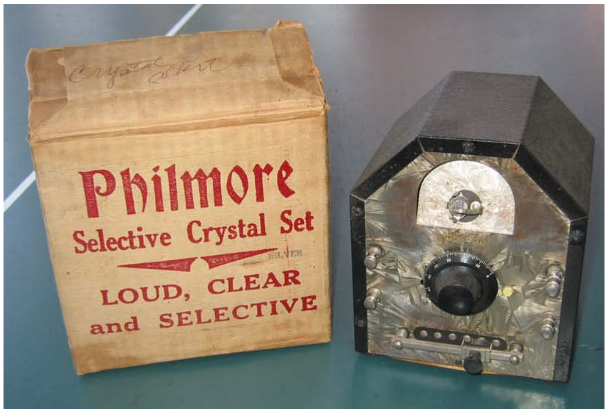
Crystal set
Philmore
Selective Crystal Set
SILVER
LOUD, CLEAR
and SELECTIVE

Philmore Selective
One of Philmore's earliest models. These came in various colors - the box is stamped 'SILVER', indicating the color of the front panel on this one. It is shown with its original box.

Philmore Kompakt
Philmore made many types of crystal radios, and this is one of their more unusual models. The radio is a small square box with an earpiece on it, a tuning knob, antenna and ground wires, and with a detachable headband that holds the unit against your ear. It is shown with its original counter display box.

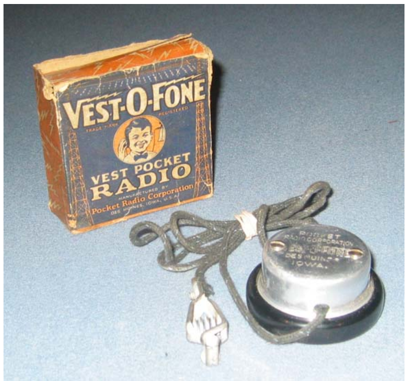
VEST-O-FONE
VEST POCKET RADIO
Pocket Radio Corporation

Pocket Radio: Vest-O-Fone
This is a crystal radio made by Pocket Radio Corporation at 720 Grand Avenue in Des Moines, Iowa, in the 1920s. It is basically a single headphone with a fixed cats-whisker crystal detector built into the headphone. It is shown with its original box.

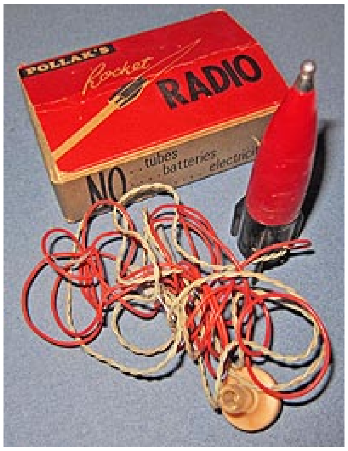
POLLAK'S
Rocket
RADIO
NO ...tubes ...batteries ...electric

Pollak's Rocket Radio
Marked on the box as 'Made in Japan for Pollak's, San Francisco 3, Calif.' This is a cheap little plastic set with an antenna that pulls out to tune it.

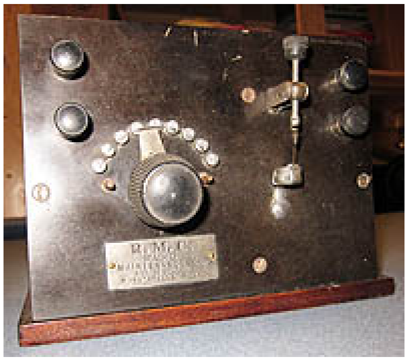

Radio Maintenance Company
A 1920s crystal radio made by Radio Maintenance Company (R.M.C.) in Minneapolis, at 31st and Snelling Avenue.

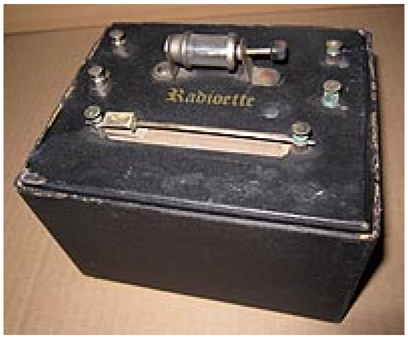

Radioette
A cheaply made crystal radio in a cardboard box cabinet.

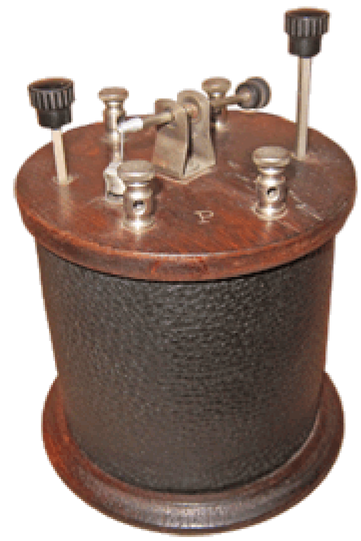

MINNEAPOLIS
MINN.

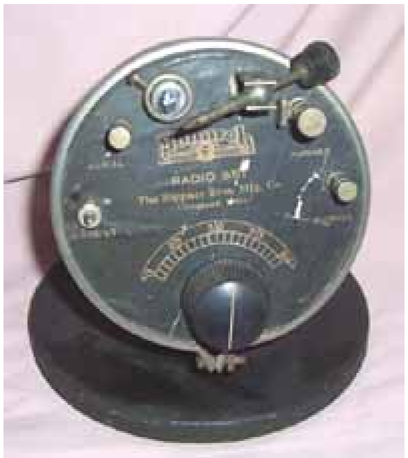

Rippner 'Monarch Radio Set'
A 1920s crystal radio made by The Rippner Bros. Mfg.

RAO 'Double-Slider'
A 1920s crystal radio made by RAO Manufacturing Company in Minneapolis. Two knobs on top move up and down to tune hidden slider contacts inside the coil. 'RAO' is stamped into the wood on the bottom. RAO Manufacturing Company still exists and is now located in Fridley, Minnesota.

Co., Cleveland, Ohio. This one has the rarer deep blue color can.

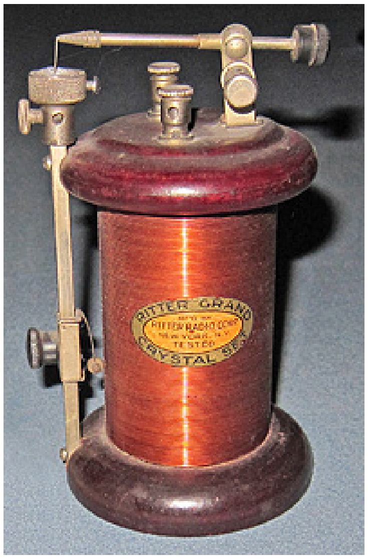
RITTER GRAND
RITTER RADIO CORP
NEW YORK N.Y.
TESTED
CRYSTAL SET

Ritter Grand Crystal Set
A 1920s crystal radio made by Ritter Radio Corporation at 232 Canal Street in New York City.

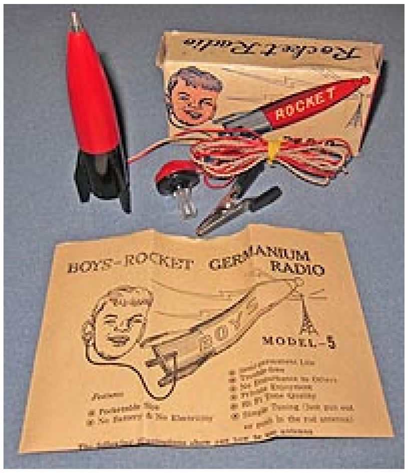
Rocket Radio
ROCKET
BOYS-ROCKET GERMANIUM RADIO
BOYS
MODEL-5

Rocket Radio - Model 5
Marked on the box as 'Made in Japan'. This is a cheap little plastic set with an antenna that pulls out to tune it.

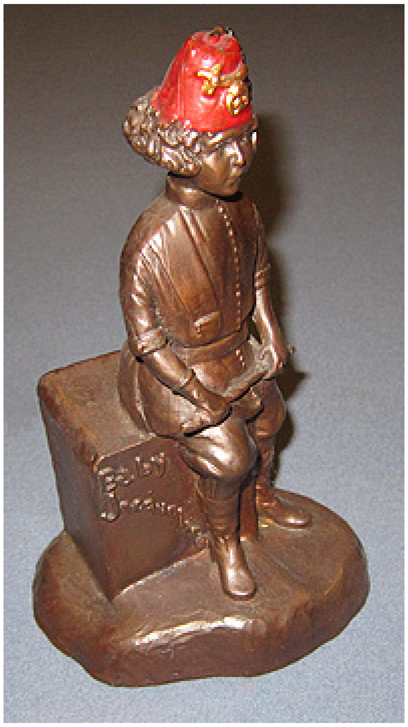
Baby Jacqueline

Shriner 'Baby Jacqueline'
Mid-1920s crystal radio in 9-1/2 inch tall cast plaster figure of 'Baby Jacqueline' wearing a shriner fez. The radio chassis is very similar to the one in the 1925 Shriner Convention radio. It is unknown who Baby Jacqueline was or exactly when and why this radio was made.

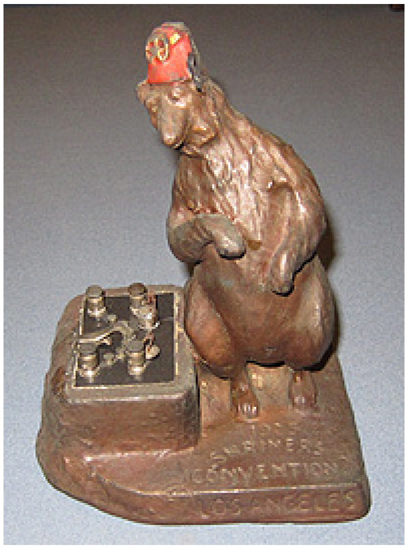
SHRINERS
CONVENTION
LOS ANGELES

Shriner 1925 Convention radio
A crystal radio is embedded in this 9-inch tall cast plaster figure of a bear wearing a shriner fez. Molded into the base is '1925 SHRINERS CONVENTION, LOS ANGELES'. The Al Malaikah Temple in Los Angeles hosted the national shriner convention in 1925 and numerous souveniers were available. It is unknown if this radio was just a souvenier or if it was some sort of prize or award. The radio chassis is very similar to the one in the Shriner 'Baby Jacqueline' radio.

Spencer Pocket Radio
This old crystal radio is still in mint condition.

Spencer Pocket Radio
This crystal radio is in excellent condition, with its original box and operating instructions.

Wezco 'DOL'
This crystal radio is shown with its original box.

portable
DX
pedestrian mo
contests
vk3ye dot com
CW
antennas
shortwave
CB
videos
HF

Return to VK3YE amateur radio projects

A simple crystal set for free power radio

Of any electronic project, the crystal set would have to rate as one of the most popular. Many amateurs are on the air today because of their early construction of a crystal set. Most practical electronic books for beginners include at least one crystal set project. Unfortunately, some of these circuits take simplicity too far and deliver mediocre performance, often by omitting key components such as the tuning capacitor, or failing to provide coil taps.

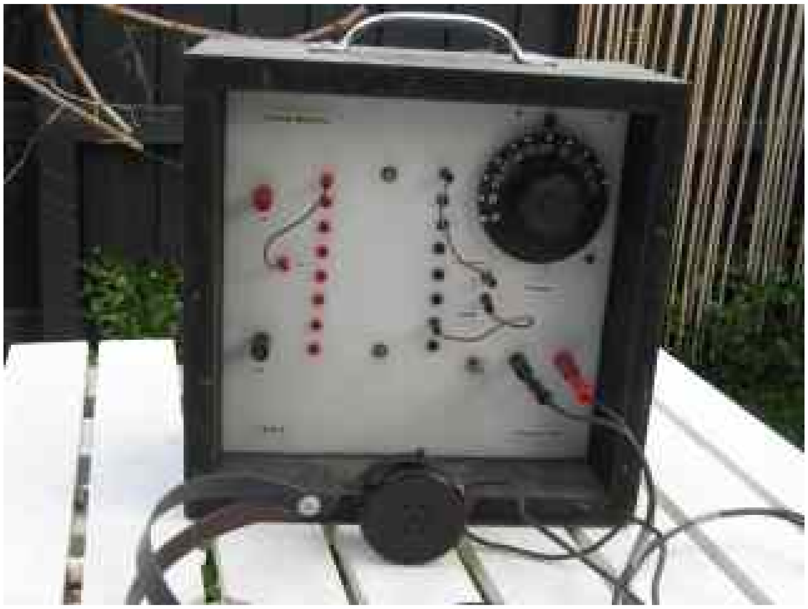

This article describes a crystal set of medium complexity. It features coil taps for the antenna and diode to make it useful for both country and metropolitan listeners. The taps allow the set to cover 160 metres if desired. All parts are easily obtainable, making it a good choice for the beginner. The endless possibilities for experimentation also make crystal sets interesting novelty projects for experienced constructors. The schematic is shown here:

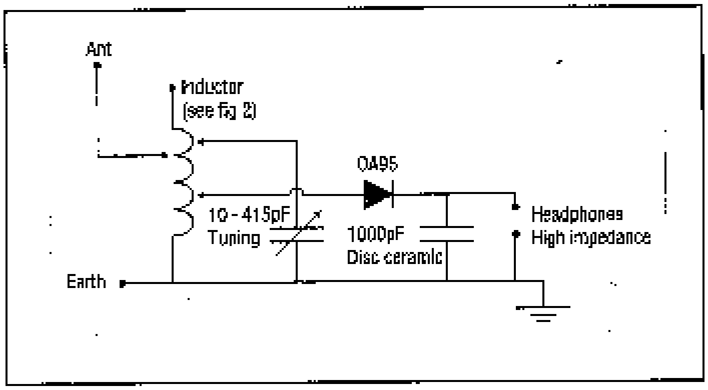
Ant
Inductor
(see fig 2)
OA95
10 - 415pF
Tuning
1000pF
Disc ceramic
Headphones
High impedance
Earth

VK3YE Crystal Set

Schematic & Layout diagrams

Press Back button to return to main article

Figure One: Schematic diagram of crystal set

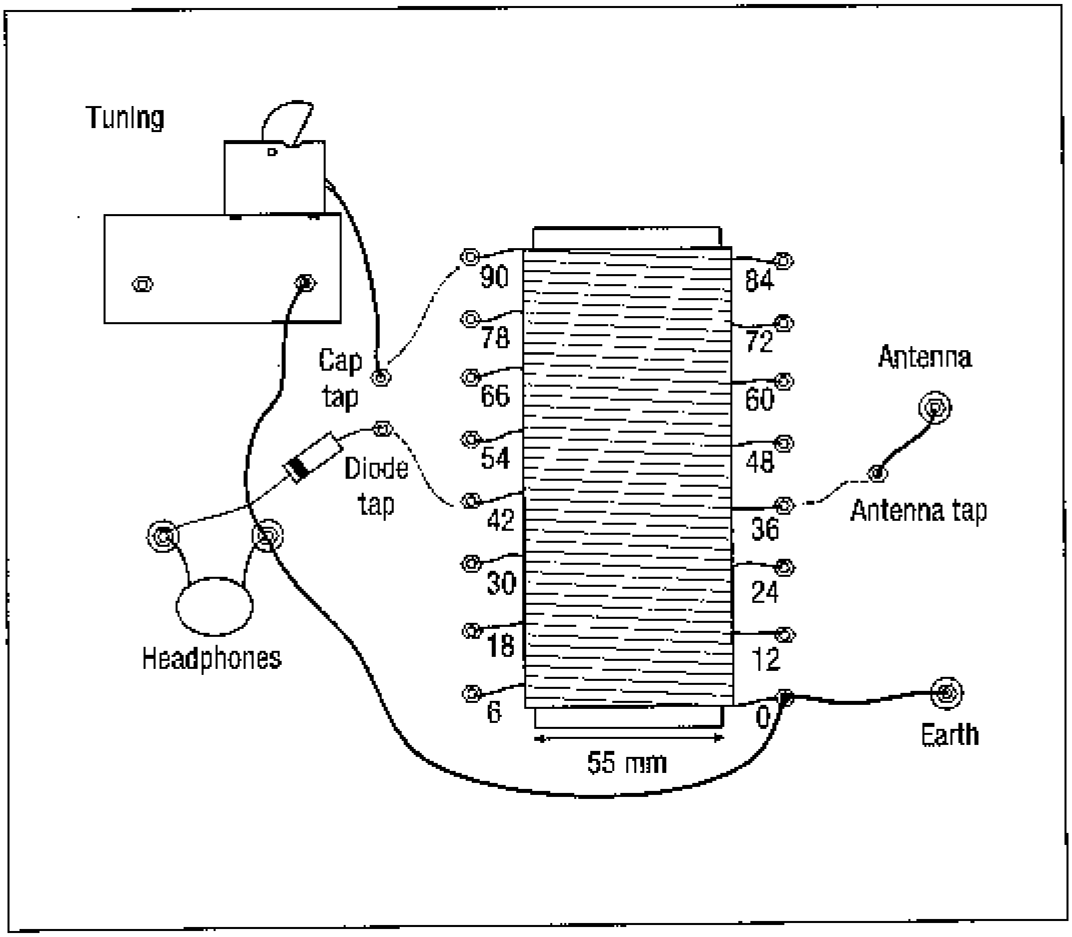
Tuning
Cap
tap
Diode
tap
Headphones
90
84
78
72
66
60
54
48
42
36
30
24
18
12
6
0
Antenna
Antenna tap
Earth
55 mm

Figure Two: Rear view of front panel, showing coil details

(from Amateur Radio February 2001, page 20)

Obtaining the parts

Tuning capacitor

A large air-spaced type, covering about 10 to 415 picofarads is preferred. These capacitors were common in valve and early transistor radios and often appear at hamfests. Their long shafts make it easy to attach large tuning knobs. When purchasing one see that the shaft turns freely, but is not loose. Ensure that the plates are straight and do not touch when meshed - use a multimeter (preferably with audible continuity function) to test this. Avoid capacitors with 3/8 inch shafts unless nothing else is available â€“ knobs for these are not obtainable, and an old valve radio dial drum will need to be used instead.

If a large capacitor is unavailable, a small plastic dielectric unit is suitable (such as stocked by Jaycar and online). The lower maximum capacitance (160 pF) means that more coil turns are required to provide coverage of the lower end of the broadcast band. This can be partially overcome by connecting the 60 and 160 pF sections in parallel (link the 'A' and 'O' tabs). The main disadvantage of these capacitors is their short shafts, which makes it harder to attach most types of knobs.

Vernier Drive and dial

The use of a vernier reduction drive is not necessary. However, its inclusion makes tuning easier, particularly on the higher frequencies. These are rare new but sometimes come up at hamfests. If your drive lacks a dial, one can be fashioned from a plastic or metal disc, such as a jar lid or salvaged computer hard disc. Glue the dial directly to the skirt of the tuning knob if you lack a vernier drive.

Diode

This is the most easily obtainable and cheapest component in the project. A germanium diode, such as a 1N60, 0A90, 0A91, 0A95 or 1N34A will be suitable. The purists still make their own diode detector with a 'cats whisker' and lump of galena, but modern diodes provide more stable and repeatable performance.

Headphones

The very old high impedance headphones are required for this circuit. A minimum of two kilohms is suggested. Medium impedance headphones (approx 600 ohms) will also work, but are less sensitive.

High impedance headphones have become difficult to obtain. Alternatives include:

1. Old-style telephone earpiece. Quite sensitive. Found in the old Telecom corded phones.

2. Crystal earpiece. These are sensitive, easy to obtain and inexpensive. You may need to connect a 100k - 470k resistor in parallel for it to work properly.

3. Piezo transducer. Believe it or not, these actually will work as an earphone. Some sizes even fit snugly in the outer ear in a similar manner to modern earpieces, such as used with mobile phones. The main drawback with transducers is their peaky audio response. Â In some cases it may be necessary to wire a 100k - 470k ohm resistor across the earphone connections for correct operation.

4. 1k to 8 ohm audio transformer and standard low-impedance headphones. Works well, but not as sensitive as a crystal earpiece.

5. Cheating! Use a transistor or IC amplifier kit to run a speaker. This approach eliminates the 'free radio' advantage of the crystal set, but provides louder reception in weak signal areas and allows speaker listening.

Coil and Coil former

This needs to be a cylinder about 55 millimetres in diameter and 150 mm long. The length needs to be long enough to accommodate all ninety coil turns used, with enough left over for mounting to the front panel. Plastic pipe, shampoo container or similar will suffice. Though enamelled copper wire can be used for the winding, the prototype used thin plastic-covered stranded insulated wire.

Front panel

All parts are mounted on a 6mm-thick polyethylene chopping board, which forms the front panel. A hacksaw was used to cut the panel to fit inside the wooden case. Use the thinnest chopping board available so that the many screw-mounted sockets used can be fastened properly. The front panel pictured was cut to 240 mm square.

Case and handle

Use non-metallic material for the enclosure. The box used in the prototype was originally a speaker bought cheaply at a school fete. The lid (which held the speaker) was removed, and the rest of the box painted. The top carry handle is optional and came from a hardware store.

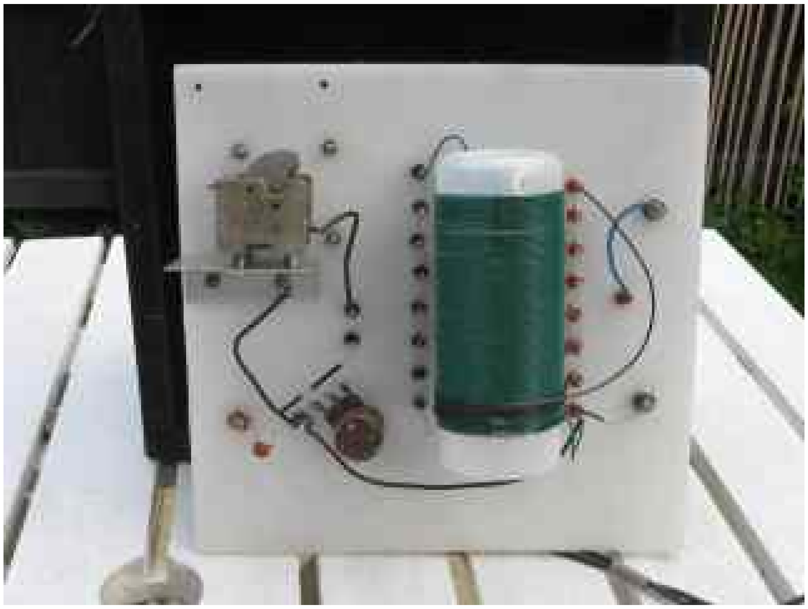

Construction

Commence construction once all components have been obtained. Plan how the parts will fit behind the front panel. The diagram and pictures above show the arrangement used in the prototype. The coil is fastened with stand-offs and the variable capacitor is screwed to an aluminium L-shaped bracket. 4mm binding posts with banana sockets are used for the antenna and headphone connections, and 2mm micro sockets for the coil tapping points.

Start by winding the coil. This consists of ninety turns of thin stranded insulated wire close wound on a plastic tube approximately 55 millimetres in diameter. A large number of tapping points are provided so that the user can vary the set's frequency coverage, and antenna and diode coupling. This makes it possible to obtain the best compromise between volume and selectivity for a particular station.

Figure Two shows the coil construction. Start from the earth end (identified as '0' in the diagram). Make two holes in the former to anchor the end of the wire. Wind six turns and then an extra half-turn. With a knife remove about 1cm of insulation, taking care not to cut the wire. Form the bare wire into a loop and lightly coat with solder. Do not apply excessive heat - the wire insulation easily melts. Wind another five and a half turns and make another tap. Repeat for the remainder of the coil until approximately ninety turns have been wound. Add more turns and taps if using a smaller variable capacitor than specified. Again make two small holes in the former to anchor the wire.

Place the completed coil aside and start work on the front panel. Mount the 4mm banana binding post terminals for the antenna, earth and headphones, as shown in Figure Two. Drill holes and mount the 2mm terminals for the coil taps and the antenna, diode, variable capacitor taps. The tuning capacitor can also be fastened at this time.

Two sets of screws and spacers can be used to mount the coil to the rear of the front panel. A 10mm separation between the coil and the panel is adequate. Solder in the various components and connecting wires as per Fig 2. Use insulated wire for the connections between the sockets and to the variable capacitor. Tinned copper wire can be used for the short links between the coil taps and the 2mm sockets. Use insulated wire for the three jumper leads. The jumpers should be sufficiently long to be able to make connections with all taps along the coil.

This completes the construction. The panel can now be inserted into the box. In the unit pictured, the front panel is recessed â€“ this protects the banana sockets and dial and makes the set more rugged. It also allows attachment of a hinged lid if required.

Parts List

* 10 â€“ 415 pF variable capacitor x1 (see text)

* 0.001 uF disc ceramic capacitor x1

* 1N60 germanium diode x1

* Vernier dial or drive x1 (optional)

* 2mm micro socket x19

* 2mm micro plug x6

* Banana socket (red) x2

* Banana socket (black) x2

* Insulated wire 20m

* Tinned copper or bell wire 1m

Other items: case and handle; polyethylene chopping board; Coil former â€“ 55mm dia, 150 mm long; screws, nuts, washers and spacers; mounting bracket for variable capacitor.

How it works

To receive signals, a radio circuit must perform three functions: selection, detection and reproduction.

The inductor and variable capacitor form a tuned circuit. The role of the tuned circuit is to select one of the many signals present at the antenna. The size of the inductor and capacitor determines the frequencies that can be tuned. The capacitor is made variable to allow the full range of AM broadcast band frequencies (531 to 1602 kHz) to be received.

The diode detector converts the selected radio frequency signal to an electrical current varying at audio frequencies.

The headphones convert this audio frequency energy to sound. The principle is similar to a relay â€“ the signal cause current to flow in a winding that forms an electromagnet. The magnetism generated vibrates the metal diaphragm, thus creating sound. Crystal earpieces perform the same function, but rely on the piezo-electric effect.

Unlike in a conventional radio, which uses amplifying devices such as transistors and integrated circuits, crystal sets are powered by the signal from the incoming station, so no batteries is required. If provided with an efficient antenna and earth, crystal sets can receive signals thousands of kilometres away.

Antenna and earth

A crystal set requires a wire antenna to operate properly. The longer and higher it is the better. A length of at least 10 metres in urban areas, and 20 â€“ 30 metres elsewhere should provide reception in most cases. The antenna should always be installed away from power lines for safety reasons and to reduce interference pick-up. An existing amateur or TV antenna can also be effective, especially if the coaxial feedline is used as part of the antenna. This is achieved by connecting both the outside and the inside of the coaxial connector to the receiver's antenna terminal.

An earth provides stronger signals, and is essential in remote areas. In homes with copper water pipes, this can simply be a lead to the nearest cold water tap. In newer homes, where plastic pipes are used, an outside ground stake can be used instead.

For long distance reception (hundreds or thousands of kilometres) more than usual effort needs to be taken when installing the aerial and earth. Reference One suggests a length of about 100 metres and a height of at least 12 metres. A series of buried radials is suggested for the earth, rather than the water pipe suggested above.

Operation

Connect antenna, earth and earphones. Install the three jumper leads. Set the capacitor tap to near the top of the coil (either the 78th or 90th turn) and the diode and antenna taps to approximately midway along the coil.

In a quiet room, adjust the tuning control and listen for a station. If several stations are audible, move the diode or antenna taps nearer the earth end (lower numbered turns) of the coil. This increases the set's selectivity and makes it possible to separate stations. In a capital city it should be possible to separate at least nine or ten stations. Optimum tap settings vary across the broadcast band â€“ lower frequency stations are often best received with higher tap settings. In rural areas volume is normally more important than selectivity, so the taps can be moved near the top of the coil.

Reception of AM operators on the 1.8 MHz (160 metre) amateur band is possible by moving the capacitor tap lower down the coil, to the 54th or 66th turn. Performance will be well down on a superhet or regenerative receiver, and SSB signals cannot be resolved. Whether you hear amateurs or not depends on your antenna system and the extent of activity from nearby operators. Here in Melbourne 160 metre AM activity includes the Monday to Saturday AM morning net starting at 11am.

In many areas there are narrowcast stations between the top of the official AM broadcast band and 160 metres. Because of their low power these stations will be weaker than the mainstream broadcasters. However these stations are excellent tests of your receiver and antenna system.

Video demonstrations of this project

Conclusion

A crystal set of moderate complexity has been described. It is the minimum required to provide good reception of local stations in urban and rural areas. However numerous refinements to increase sensitivity, selectivity or audio output can be made. These include:

1. Double tuned circuits (with variable coupling between them) to improve selectivity

2. Use of a tuned trap to null out interfering signals

3. Attention to the construction of coils to provide the highest possible Q

4. Addition of an impedance matching network to provide efficient power transfer between the antenna and the tuned circuit

5. Use of a large loop antenna for the coil to allow reception of signals without an external antenna and nulling of unwanted signals

6. Voltage doubler diode detector circuit using two diodes to increase volume

7. Use of DC bias (from a DC voltage applied to the diode) or RF bias (from a locally generated RF signal on the receiving frequency) to improve sensitivity, or, in the case of the latter, to provide CW and SSB reception.

8. Use of a Q multiplier to increase sensitivity and allow CW and SSB reception.

Should you decide to experiment with these changes, it would be desirable to keep this set as a reference and build a second receiver as a test bed for the experiments.

Obtaining the parts

Suitable parts were discussed in detail above. Many if not all can be bought online, with examples presented below.

Variable capacitor and knob

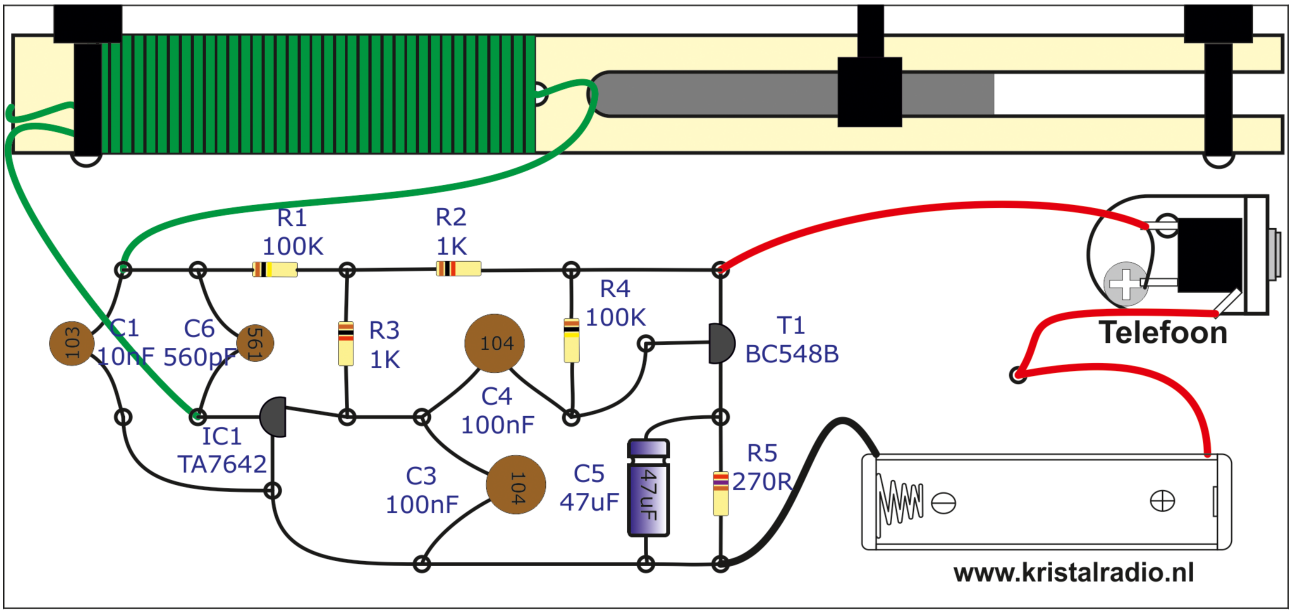
R1
100K
R2
1K
103
C1
10nF
C6
560pF
561
R3
1K
104
R4
100K
T1
BC548B
C4
100nF
IC1
TA7642
C3
100nF
104
C5
47uF
47uF
R5
270R
Telefoon
www.kristalradio.nl

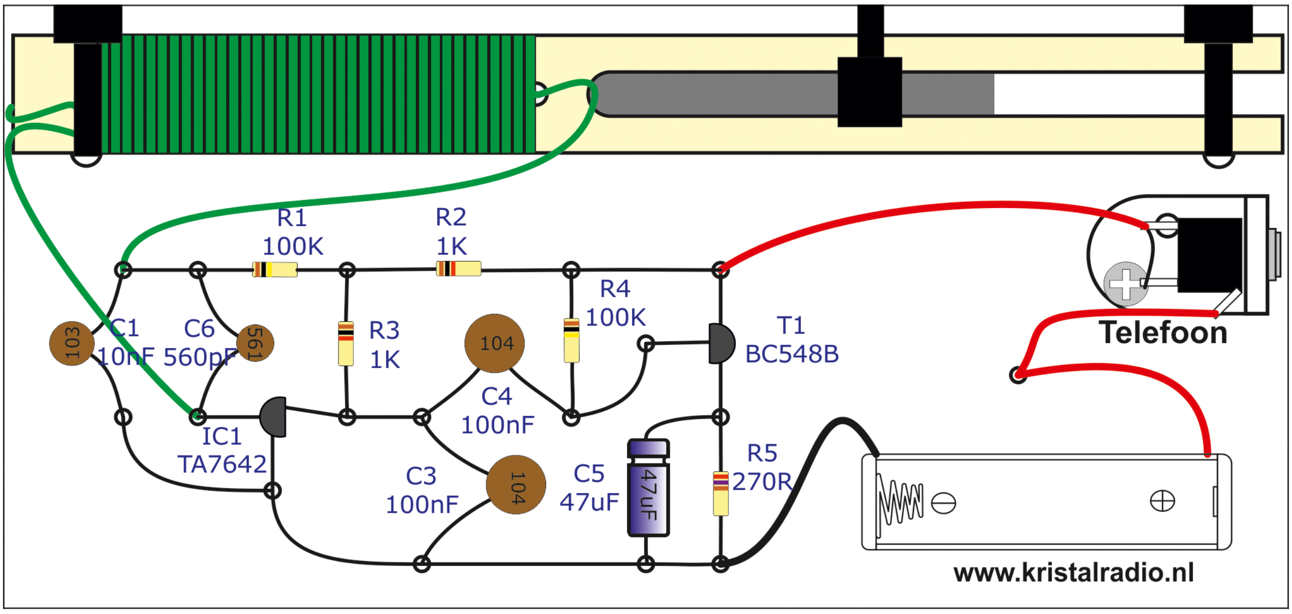
R1
100K
R2
1K
103
C1
10nF
C6
560pF
561
R3
1K
104
R4
100K
T1
BC548B
C4
100nF
IC1
TA7642
C3
100nF
104
C5
47uF
47uF
R5
270R
Telefoon
www.kristalradio.nl

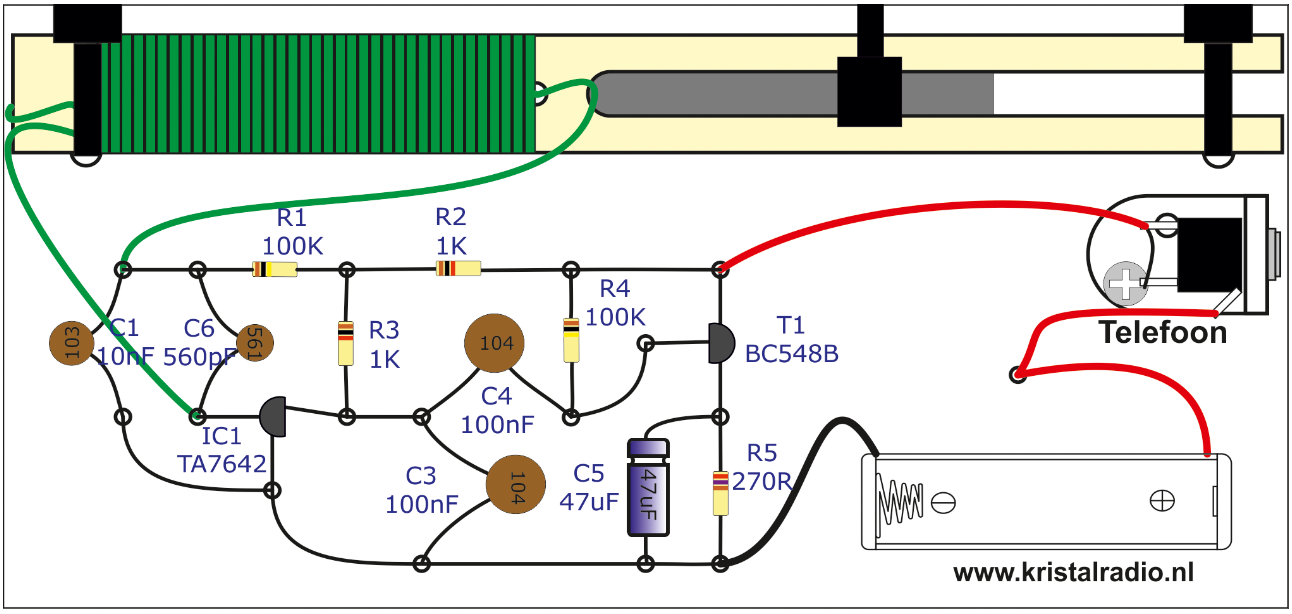
R1
100K
R2
1K
103
C1
10nF
C6
560pF
561
R3
1K
104
R4
100K
T1
BC548B
C4
100nF
IC1
TA7642
C3
100nF
104
C5
47uF
47uF
R5
270R
Telefoon
www.kristalradio.nl

A very selective and high performance crystal radio
My Heathkit model CR-1

A comprehensive story containing historical, educational, technical and biographical elements & opinions

by

John Fuhring

If you found this page while looking for a simple, but great performing little radio that you can build yourself, you might be more interested in reading about

An Armstrong "Crystal" Radio

Rebuilding my Heathkit CR1 and getting it to work

I can not really remember when or how I came by this little radio, but it was many years ago. I seem to recall that I picked it up cheap at a thrift store as a non-functional item that had been poorly or incompletely put together. I remember that some of the screws and a connection post was missing and the parts weren't soldered in properly or correctly. To tell the truth, I didn't know how the radio was supposed to be wired and I must not have been very interested in getting it working, so it sat in a cabinet for several years until a year or so ago.

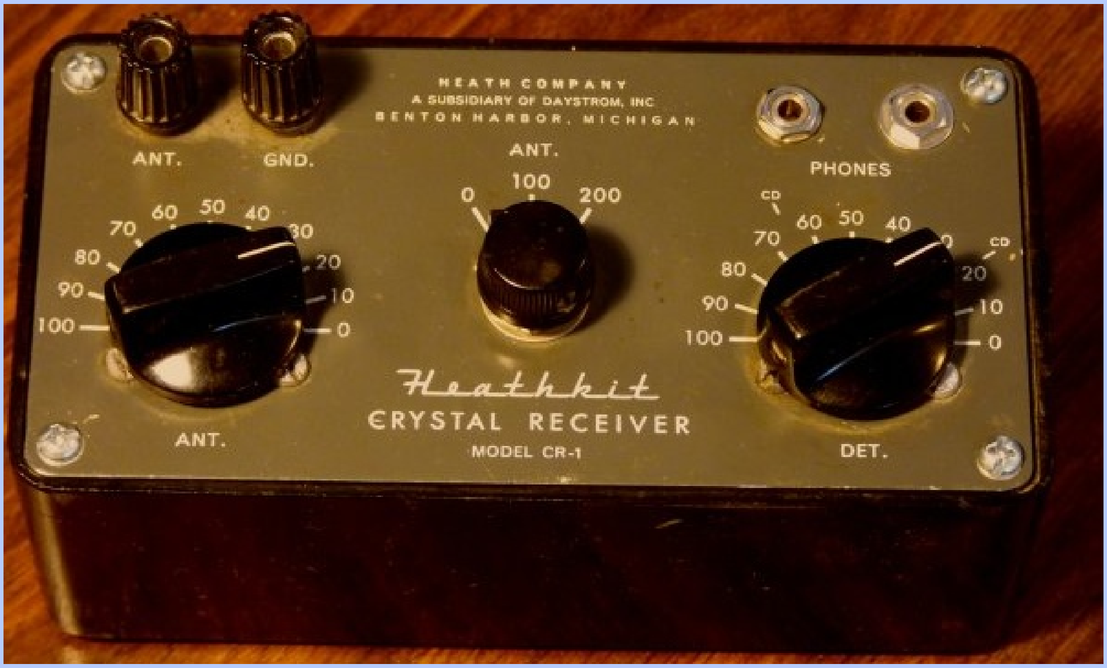

My CR-1 crystal radio modified for two earphone jacks.

About a year ago I came across the little radio and decided that I would figure out how it was supposed to work and then I'd put it together correctly. The first thing I did was put a sub-miniature phone jack in place of the missing binder post since my piezo earphone has a sub-miniature plug on it and besides, I didn't have a matching binding post to replace the missing one. I took off the grounded binding post and put a screw to cover the hole. Next I soldered the antenna tuner coil to its variable capacitor and wired the two fixed capacitors to the selector switch. I then soldered the station tuning coil side to its variable capacitor and from the tap on the station tuner coil, I soldered a 1N34 germanium detector diode to the phone jack. By the way, Heathkit called the tuning capacitor connected to the station tuning coil the "detector tuner" and on the panel it is labeled as "DET" for detector.

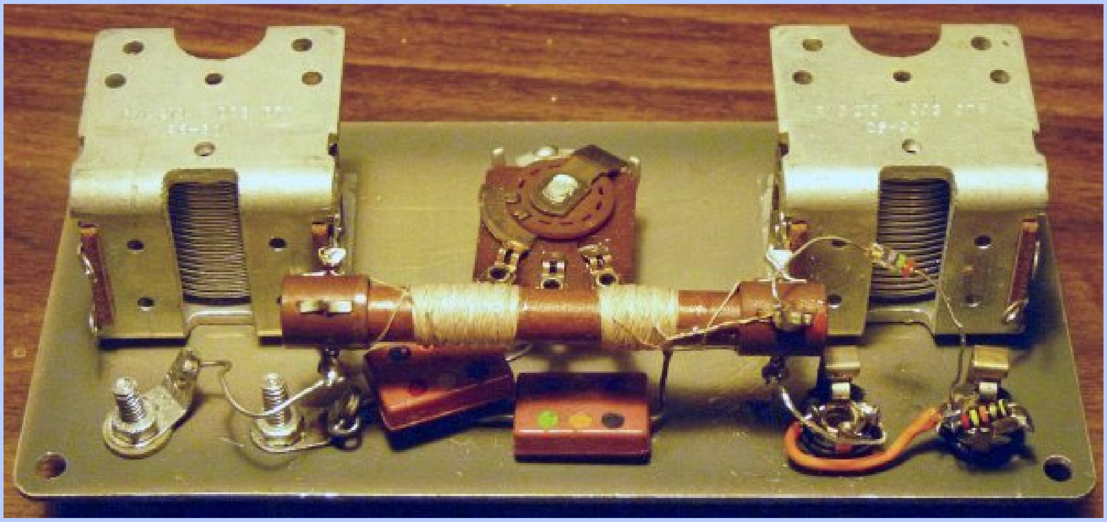

Back side showing the radio's components.
The antenna and the station tuning coils are wound on the phenolic tube. The antenna coil is to the left of center. A powdered iron (ferrite) slug is inside the tube between the two coils.

When I finished soldering everything in correctly, I connected the little set to my antenna & ground and soon I could clearly hear my local AM stations, but I noticed that the tuning dial seemed to be way off. At the very top of the AM dial (around 1600) I know there is a very strong Mexican station, but I couldn't go high enough to tune it in. To match the dial with the AM band, I took a thin wooden stick and gently pushed on the ferrite slug in the coil form until it was nearly centered between the two sets of coils. By moving the slug, I was able to match the dial to about where I knew my local stations were and was able to hear the Mexican station near the top of the dial. The slug is held inside the tube by wax and if you are careful and apply slow, even pressure, the slug will move to where you want it.

I now had a crystal radio put together as it should be and it worked -- sort of. Below is the schematic diagram I developed for the little radio. I'm kind of big on schematic diagrams, but I wasn't always so.

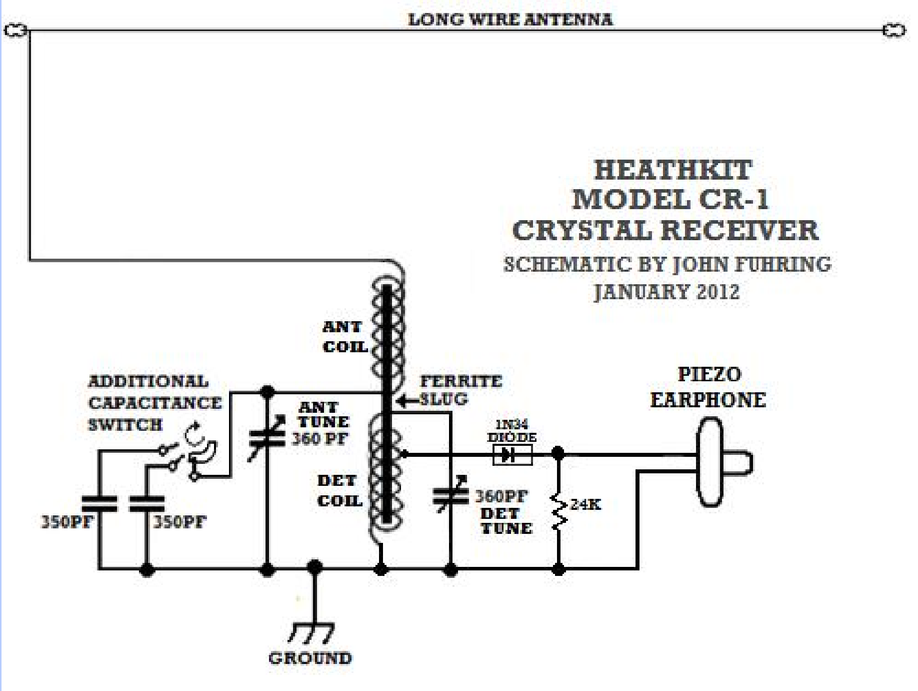
LONG WIRE ANTENNA
HEATHKIT
MODEL CR-1
CRYSTAL RECEIVER
SCHEMATIC BY JOHN FUHRING
JANUARY 2012
ANT
COIL
ADDITIONAL
CAPACITANCE
SWITCH
ANT
TUNE
360 PF
FERRITE
SLUG
PIEZO
EARPHONE
1N34
DIODE
DET
COIL
360PF
DET
TUNE
24K
350PF
350PF
GROUND

My schematic diagram of the Heathkit model CR-1 crystal radio.

In many respects, this design operates on similar principles to the "Trench Radios" of WW 1. Military radios of that day, such as the US

Army's BC-14/SCR-54, also used antenna tuning, light coupling between stages, a tuned diode section and they operated on the same range of frequencies, 550-1600 KHz.

I understand that these kits were very popular with Boy Scouts and were marketed as a kind of a "Survivor's Radio" during the worst parts of the "Cold War" when my government's Civil Defense planners expected an Atomic War with the USSR to break out at any time. With only a few CONELRAD stations on the air and with us few survivors trapped in "bomb shelters" for months and years, with our entire civilization and economy wrecked, still, news and information would be available to any doomed citizen having one of these radios, even though there was no electricity and the battery operated radios had all gone dead. Doesn't that sound cozy? Doesn't that sound like fun? Well, at least we would be "better off dead than red" and, being dead, we'd not have to "live under godless communism."

Oh, weren't the "Good Old Days" just wonderful?!?

At the time I first got this little radio working, all I had was a short outside antenna that I used with my one-tube regenerative radio and my EC-1 shortwave radios. To tell the truth, when I connected the Heathkit crystal radio up to my rather poor antenna, I was very much UN-impressed with this radio. Compared to my Armstrong regenerative radio (written about elsewhere on this website), it performed very poorly and all I could receive was my four really strong local AM stations. You know, there is nothing on my local AM radio except all this mean, downright goofy, unpatriotic right-wing hate talk or the even more disgusting and ignorant Fundamentalist preaching, none of which I can stand to listen to, but there is a Mexican music and local talk station that wouldn't be bad if I understood a word of Spanish. With nothing to listen to, I lost interest and did nothing more with the radio.

Surprising performance from this kind of radio

A few weeks ago I vastly improved my outside antenna and had some fun using it to do some AM DX listening with my 1958 Armstrong one tube regenerative radio. One evening, a few days ago, the AM signals were coming in from San Francisco, Los Angeles and Reno very strongly and so I started wondering if the crystal radio could pick them up too. I was very skeptical because the crystal radio had never performed all that well, but now I had a better antenna and maybe things would be different.

When I connected the crystal radio to my long wire antenna, I was pleased to discover that both the station tuning AND the antenna tuning sections now worked together to make the tuning very sharp. Before, with the short antenna, the antenna tuning section didn't seem to do anything and I was wondering what it was put in there for, but now it worked and I was extremely impressed with how selective the radio could be for a crystal set. Strong stations that had once blended together now could be separated and the strong stations didn't bleed on to the weaker stations. So, with the radio working better than at any time since I rebuilt it, I tuned to about where KGO up in San Francisco should be (middle of the dial at 810 KHz). Wow! by carefully adjusting both tuners, I could hear KGO loud and clear and without other stations overlapping. By the way, KGO is about the last station on AM that hasn't been taken over by the right-wing propagandists or America's own brand of the Talaban and they feature real people talking about real things that matter to people and not the steady diet of political and religious garbage that dominates nearly all of AM these days. Sadly, when KGO goes silent, I'm sure someday it will, AM radio will be well and truly dead.

Listening to this simple little radio pull in distant stations like that was a kind of "radio communications epiphany" for me. I always wondered how ships like the Titanic could communicate with other ships and to shore stations with nothing more than Fleming Valve or crystal radio receivers. My experience with this little radio has shown me how, with a good antenna and good sharp tuning circuits, individual radio signals can be picked out from a mass of signals out there and can be heard from a long way off. For me, this is a very wonderful thing, especially when I consider that these radios do not amplify their signals, but use the tiny electrical energy that is picked up by the antenna to make sound. Such is the extreme sensitivity of human hearing and the amazing efficiency of the circuits, detector crystal and earphones that microwatts of power, originally coming from a station hundreds of miles away and landing on my antenna wire, can be heard clearly.

The way this radio works is as basic as it is possible to get

and here's how it all works:

Imagine that there is a radio station 10 miles away and that it that is broadcasting radio waves at 1,400 KHz. Imagine that our radio station is broadcasting hundreds of watts of energy into space and that if we could capture just a tiny amount of this energy, we would be able to hear what was being broadcasted. Imagine that we have a crystal radio like the one above connected to a long wire antenna.

(1) The passing radio wave from our station induces a tiny high frequency voltage on our antenna wire. The antenna is connected to the antenna coil (coil on the top of the diagram) with its antenna tuning capacitor to the left and below it.

(2) When the antenna capacitor is tuned just right so that the inductance of the coil is matched by the capacitance of the tuning capacitor, a tiny current from the antenna flows in the antenna coil to produce an oscillating (or varying) magnetic field inside the antenna coil. By carefully adjusting the antenna tuning capacitor, this oscillating magnetic field will be right in tune with 1,400 KHz and 1,400 KHz only. This is important because it helps seperate our 1,400 KHz signal from all the others on the AM band.

(3) The oscillating magnetic field inside the antenna coil is coupled over to the other coil (the station tuning coil) through the well known principle of "Mutual Inductance" discovered by Michael Faraday in the 1830s.

(4) The 1,400 KHz signal that has been coupled to the station tuning coil is contained in a "tank circuit" that is made up of the station tuning coil and the station tuning capacitor (shown to the coil's right).

(5) By carfully adjusting the station tuning capacitor so that its capacitance exactly matches the coil's inductance, a so-called tank circuit is formed. This tank circuit "resonates" or "tunes in" the 1,400 KHz signal and that means that it further seperates our 1,400 KHz signal from all the other nearby signals. We now have just the radio signal we want, but we need to do more because it is impossible for our ears to hear radio frequency signals.

(6)
To "detect" the presence of the radio frequency signal that is resonating in the tank circuit, some of the tank's energy is tapped off near the bottom of the station tuning coil and it is wired to one side of the 1N34 detector diode.

(7)
The 1N34 diode chops off the negative part of the radio frequency wave. The detector needs to do that or the negative part of the wave would cancel out the positive part of the wave and nothing would be heard.

(8)
When the diode detector chops off part of the radio frequency signal, direct current audio frequency undulations are all that are left over from the original radio wave. These direct current audio frequency undulations correspond to the sound being transmitted by the station and when the wires of a very sensitive earphone are connected to the detector, these undulations of direct current cause the mechanism inside the earphone to vibrate in step with the undulations and it is those vibrations that we hear as sound.

Here is a graphical representation of what's going on when a radio frequency signal is detected (turned into sound):

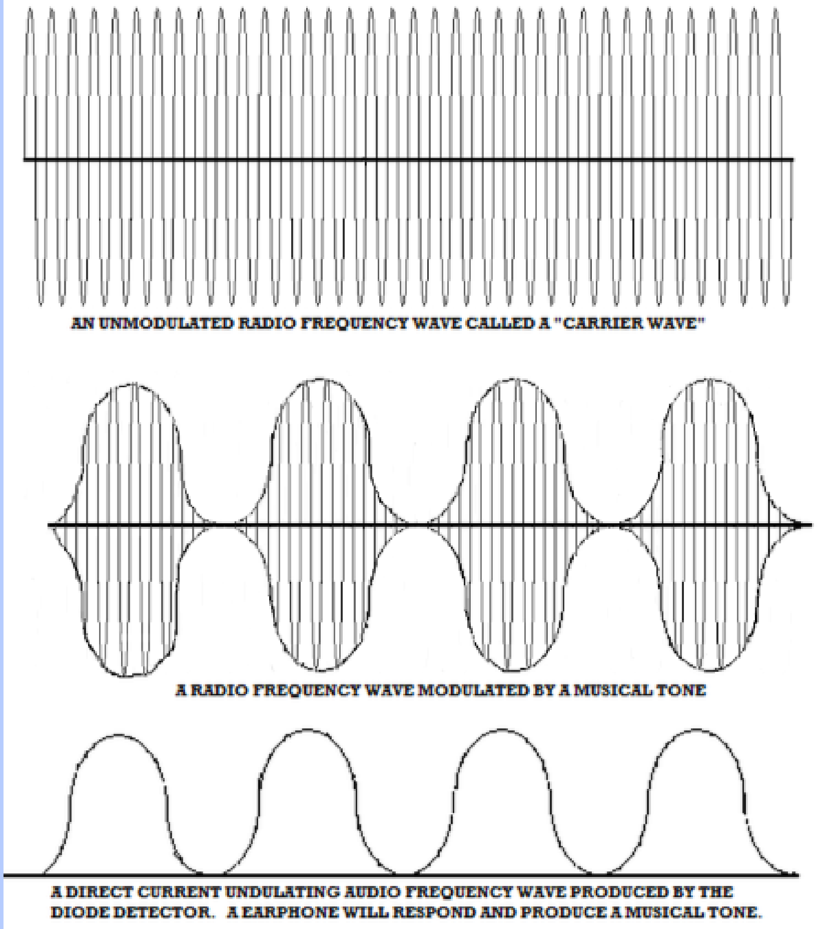

AN UNMODULATED RADIO FREQUENCY WAVE CALLED A "CARRIER WAVE"

A RADIO FREQUENCY WAVE MODULATED BY A MUSICAL TONE

A DIRECT CURRENT UNDULATING AUDIO FREQUENCY WAVE PRODUCED BY THE DIODE DETECTOR. A EARPHONE WILL RESPOND AND PRODUCE A MUSICAL TONE.

Graphs of the various waveforms seen at the tuner stage (top two) and at the detector stage. Unmodulated carrier waves are also called "continuous waves" and can't be heard by a crystal detector except as a kind of a very faint hissing noise. AM radio stations try to avoid sending out unmodulated carrier waves (called "dead air") and so this waveform is rarely seen. More common is the waveform showing lower frequency sound waves superimposed on the carrier wave.

The dark black line is where the voltage is zero as a carrier wave crosses from positive to negative. As shown here, the negative going parts of the wave are stopped by the detector diode so that positive to zero to positive undulations remain. If you reverse the diode, the direct current undulations will be negative to zero to negative, but the same musical tone will be heard in the earphone.

Most crystal radios use only one half of a radio frequency's wave.

The importance of a good antenna and ground system

These radios are not connected to your house's AC power and they use no batteries. A crystal radio has no amplifying circuits and yet they output sound and sound is energy. Where do they get the energy to make sound? Do they somehow use mysterious and occult "free energy from space" or something goofy like that? The energy needed to run these radios doesn't cost you anything directly, but it isn't free or mysterious. The energy is from a tiny fraction left over from the hundreds of watts of radio frequency energy a station broadcasts out to space. The vast, vast majority of the energy a broadcast station radiates out of its antenna system is lost, but a tiny fraction of it may be picked up by an antenna at somebody's house.

Antennas for listening to AM stations can range from tiny "loop" antennas that are extremely inefficient, but good enough for amplified radios, all the way to elaborate "long wire" systems dozens and dozens of feet long and high up in the air. The general rule was discovered by Marconi over a hundred years ago that the longer and higher the antenna, the better it is at capturing some of that energy that is passing by. To produce enough energy so that it can be processed inside the crystal radio circuits and then come out as sound energy strong enough to vibrate the mechanism inside the earphone, the antenna needs to be as long and as high as is practical. Practical is what you have room for and how high you can safely go. Practical antennas can be strung between poles located at the ends of a roof or from the house out to a distant tree or pole. It is true that antennas longer than 60-100 feet and higher than 25 feet do capture more energy, but things soon get very difficult and for very little extra gain at distances and heights beyond what is practical.

In addition to a really good long wire antenna (as long as is practical as mentioned), this radio must have a good ground. Right now, I am using my house's wiring as a ground and it seems to work well. Water pipes, when available to connect to, are supposed to work even better. When using house wiring, it would be a deadly mistake to connect to anything but the AC ground, so be careful.

Sensitive, high impedance headphones or earphones are a must

This radio works reasonably well with good quality high impedance headphones, but my very inexpensive piezoelectric earphone is noticeably louder on weak signals. The tapped tuning coil matches high impedance headphones, but for best selectivity (ability to separate close-by stations), the even higher impedance piezoelectric earphone works best.

A modification so I can use two earphones

After using this little radio set for several days, I had a "brilliant" idea. Since only one side of the radio frequency wave is being "detected" (or rectified) and being turned into undulating direct current waves, what about the other side of the radio frequency wave? Couldn't I detect it too, only in reverse and end up with two direct current waves of opposite polarity so that I could connect up separate earphones? Sure, why not, so that's exactly what I did. I installed a second phone jack and a second 1N34 diode only the new one is in reverse of the other. Then I did some testing and even more testing today. I hate to admit it, but the truth is, I can barely detect a any real difference between using a separate detector and just connecting the second earphone to the same diode. By the way, I noticed some "rattling" on the stronger signals while having two earphones connected, but when I put 24,000 ohm (1/4 watt) loads across the earphones, the tuning didn't seem to be effected, the sound quality improved and the "rattle" was gone.

When I first tried this dual diode system out, there seemed to be a real improvement over using two earphones off the same diode, especially on really weak signals, but on further testing, that difference appears to be minimal at best. At first I was surprised that I had never seen or heard of anybody doing this back in the early days of radio when family members would share a headphone, but now I'm not so sure this is a good idea after all. The theory looks good, but in practice, it's hardly worth doing.

Here's what the circuit now looks like with two earphones connected to the same diode:

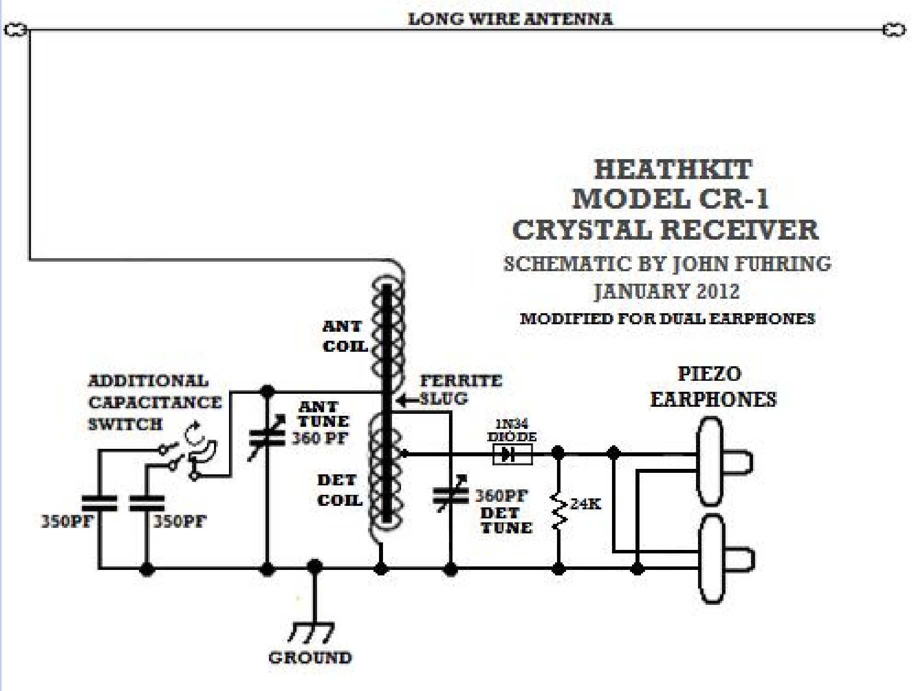

The Heathkit CR-1 modified for two earphones.

Even though using an extra diode didn't really improve the operation of my crystal set as much I thought it would, I liked the idea of having an extra jack for another earphone so I installed it. I can now share my radio's output with somebody else in the bomb shelter in case Atomic War breaks out. Doesn't THAT sound like fun? Seriously, listening with both ears improves my listening experience -- now if there was just something worth listening to.

Tuning this radio for maximum sensitivity and selectivity

The trick to tuning this radio is to first locate a strong local station of known frequency and tune it in with both the antenna tuner and the station tuning controls. If the setting of the antenna tuning control doesn't seem to make much difference, try switching in extra capacitance, but if that doesn't help, consider stringing up a longer antenna. Both controls should be used to peak the signal for maximum loudness. Once you have an idea where on the dial your station is located, move both controls to about where you guess the weaker station you want to listen to might be. Once there, slowly and carefully adjust the station tuning until you hear something and then peak it up with the antenna tuning control. You may have to go back and forth a few times to get maximum volume with minimum interference from nearby and stronger stations. The radio has a switch to add additional capacitance if the variable capacitor isn't enough, but so far I haven't used an antenna that requires extra capacitance.

One of the shortcomings of this and most other crystal radios is that the tuning dial is not calibrated in frequency, but is simply numbered from 100 to 0. It is doubly difficult to try to set the tuning knob with this radio because the larger the number, the lower the frequency. For example, if the station you want to listen to is at 1600 KHz on the AM dial, the Heathkit must be tuned between 10 and 0, if the station is at 800 KHz, set the dial to 50. What is really ironic is that Heathkit put frequency markers on the station tuning control for the Civil Defense frequencies (640 KHz & 1440 KHz), but they left it up to you to guess where the other frequencies should be. Setting the antenna tuning is equally important and again the numbers are backward. For the highest frequencies, the setting is 0 and for the lowest frequencies, the setting is 100.

The best time of the day to have fun with these kinds of radios

During the daytime and during the summer months, about all that can be received with a little set like this, at least around here, is the trash in English that the local broadcasters pander to and a couple of Mexican language and music stations, but in the evening, electrified layers way up high in the outer fringes of the atmosphere allow radio waves to be bent back down to earth so that interesting stations worth listening to, but which are 200 to 300 (sometimes even more) miles away may be heard loudly and clearly. This long distance "refraction" or bending of the radio waves back to earth (a process called "skip") makes it possible to hear these distant stations, but it isn't like listening to the station next door. This "skipping" of radio waves off electrically charged layers in the upper atmosphere makes these signals subject to all kinds of things including interference from even more distant stations, fading in and out, some strange sounding distortions and sometimes there is even an echo effect. These so-called "propagation effects" can be annoying, but if you just wait a few seconds or minute or so, the signal generally returns.

I think that listening for yourself, to hear what Nature is doing to these signals way up in the sky, is part of what makes listening to the Broadcast Band at night so interesting. There is a certain randomness to this skip that is explained to some degree by solar astronomy and physics. Scientists and engineers all over the world closely monitor our local star and measure it for electrical and magnetic activity. Reports and prediction tables are compiled by various scientific agencies and all this stuff is available on the Internet. When you can test these predictions for yourself by listening for "skip," you are participating in a scientific observation in your own little way.

Listening to far distant stations with a crystal radio like this one shows you how ships like the Titanic could communicate with other ships and to shore stations hundreds of miles away with nothing more than crystal radio receivers. Speaking of the Titanic, two thirds of all the people on her died, but without early radios, very much like this crystal set, everybody would have died. If you make one of these radios and actually put it to use, you will see how, with a good antenna and good sharp tuning circuits, individual radio signals can be picked out from the mass of signals out there and your station can be heard from a long way off under the right conditions. For me, this is a very wonderful thing, especially when I consider that these radios do not amplify their signals, but use the tiny amounts of electrical energy that is picked up by the antenna to make sound. Such is the extreme sensitivity of human hearing and the amazing efficiency of the circuits, the detector crystal and the earphone, that microwatts of power, originally coming from a station hundreds of miles away and landing on my antenna wire, can be heard clearly.

How far we have come

When you think about it, it seems that such basic radio technology represented by this little crystal radio is quite old, but in fact, the first Fleming Valve and crystal radios were invented just a little over 100 years ago. When my mother was a girl, people were building crystal sets so that they could listen to the wonderful new radio broadcasts from the big cities. In a relatively short span of time, the time since my grand parents were young in the 1890s to my parent's time, to this very moment, consider just how far we have come with radio, television, GPS, cell phones, satellite broadcasting, wireless Internet, interplanetary missions to space and so much allied technology.

The End

Having arrived this far, obviously you have a superior attention span and reading ability that far exceeds that of the majority of web users. I highly value the opinion of people such as yourself, so I ask you to briefly tell me:

Did you enjoy this article or were you disappointed?

Please visit my guest book and tell me before you leave my website.

If you have any detailed comments, questions, complaints or suggestions, I would be grateful if you would please **E-mail me directly**

If you are a science student or hobbyist, you might be interested in building a radio
that is as simple and cheap to build as a basic crystal radio, but hundreds of times better,
you might like this recommended article:.

Recreating an Alfred P. Morgan crystal set

Analog Dial | Page 1 | Page 2 | Page 3 | Page 4

OK, let's put more holes in the base!

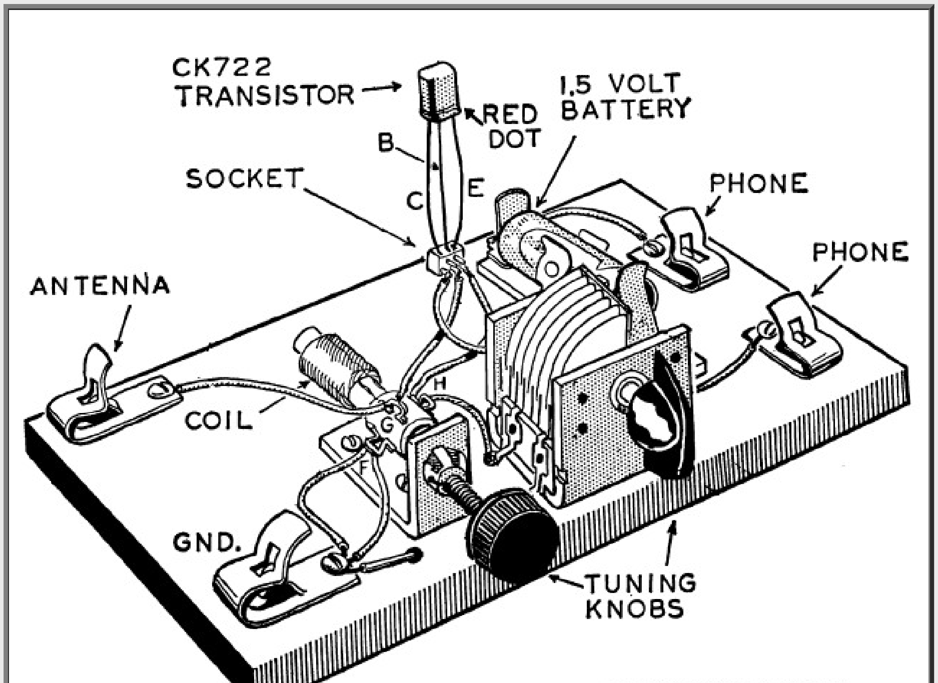

RADIO RECEIVER WITH TRANSISTOR DETECTOR

Further on in "The Boys' Third Book of Radio and Electronics" is this receiver. It was off to Ree Electronics for a coil and a transistor. They didn't have a CK722. The CK722 was designed for hobbyists and Ree was an electronics retail and repair outfit. You wouldn't find a CK722 in any piece of consumer electronics. They sold me a Germanium equivalent.

By the way, the CK722 became the most popular transistor in the world, but inside every one was a REJECT from the hearing aid industry. In the early 1950s Raytheon was producing tens of thousands of transistors a month, but only half of them met specifications. If the rejects worked up to a certain point for gain and noise, they were repackaged as a CK722.

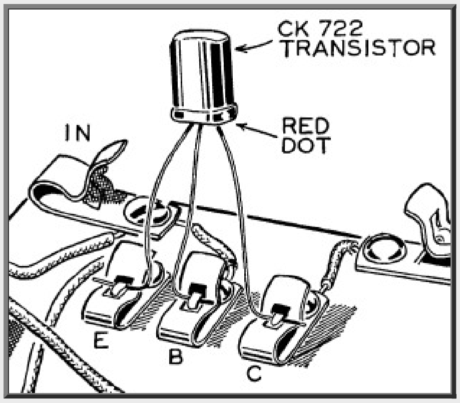

I didn't have a socket, so the transistor would be mounted like this. I realized that by adding three Fahnestock clips for the transistor, I wouldn't have to move any of the others. The diode connections could become the battery connections. Antenna, ground and phone connections could all remain the same.

FOUR SIMPLE, PRACTICAL RADIO RECEIVERS

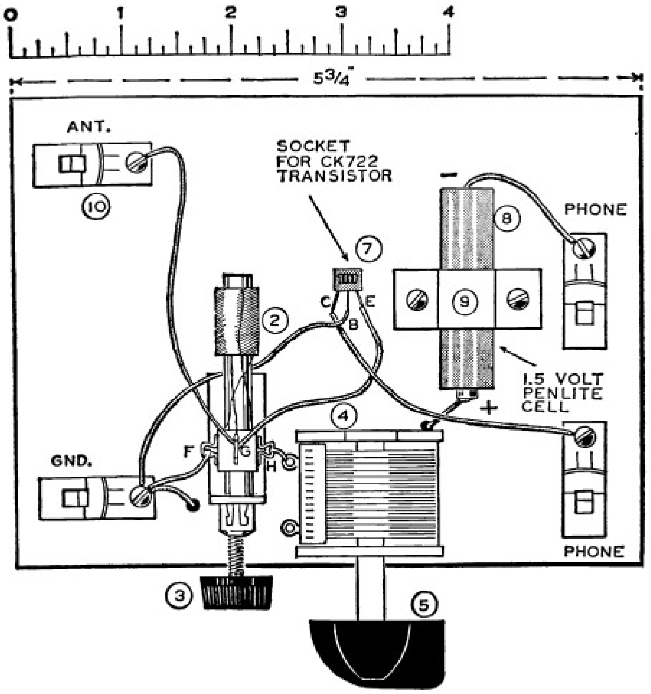
0
1
2
3
4
5 3/4"
ANT.
10
SOCKET FOR CK722 TRANSISTOR
7
C
E
B
8
9
PHONE
2
1.5 VOLT PENLITE CELL
4
GND.
F
G
H
PHONE
3
5

PLAN OF RECEIVER WITH TRANSISTOR DETECTOR

The set was wired according to this drawing. This is important later in the story.

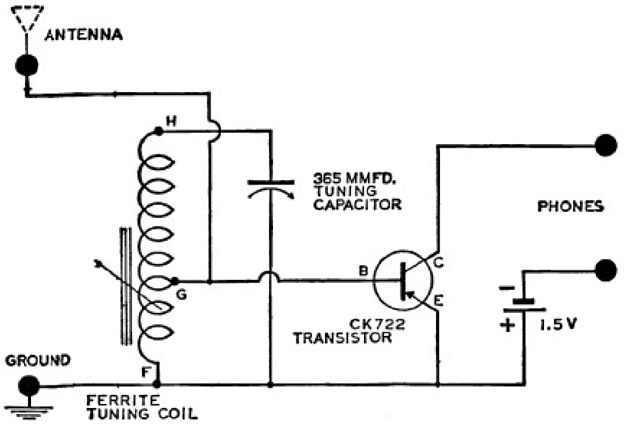
ANTENNA
H
365 MMFD. TUNING CAPACITOR
PHONES
B
C
E
G
CK722 TRANSISTOR
1.5 V
GROUND
F
FERRITE TUNING COIL

CIRCUIT DIAGRAM OF RECEIVER WITH TRANSISTOR DETECTOR

137

The schematic.

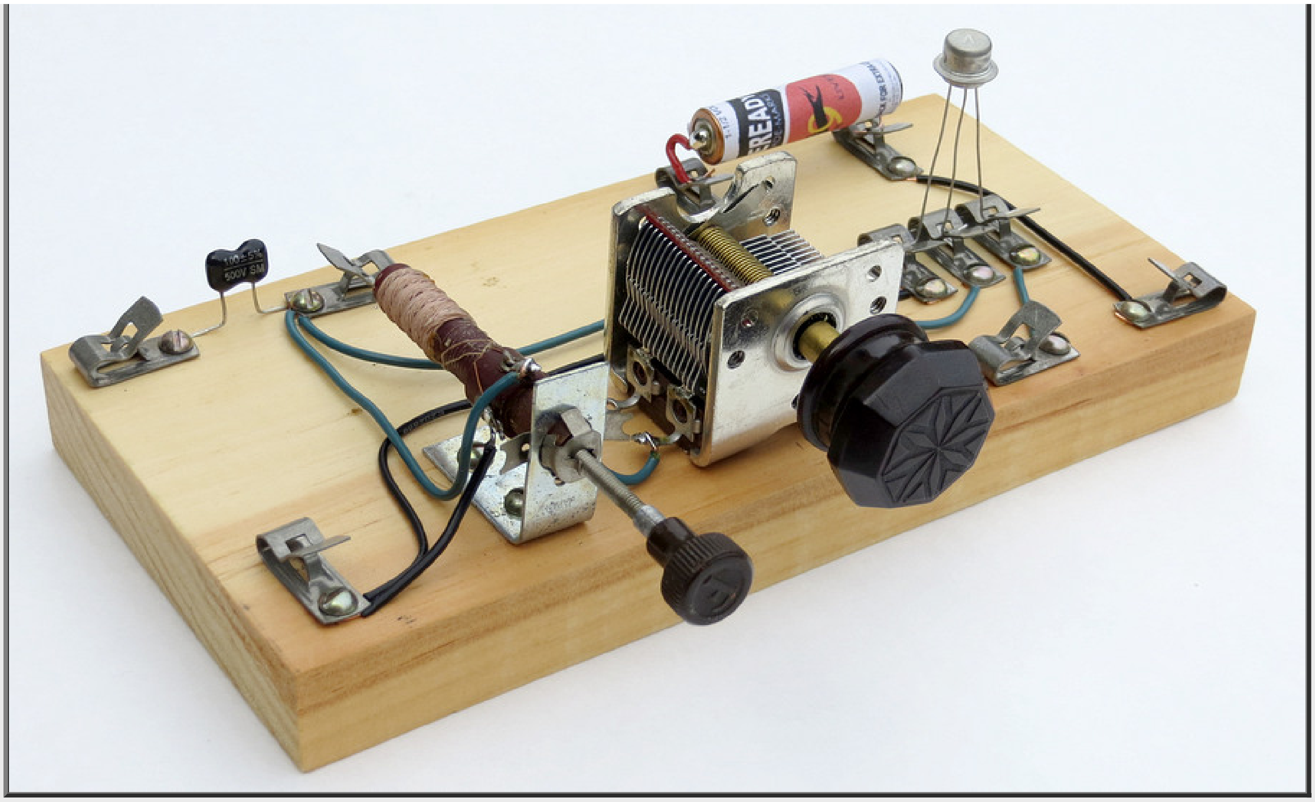

Here is the recreation. The transistor is a CK722 / NTE102 Germanium equivalent. The battery is an AAA skinned with a printed vintage wrapper. I am quite impressed with this. This radio has ONE active component and a battery. The selectivity is very good and our local station, WNPV 1440 AM, came in with plenty of volume.

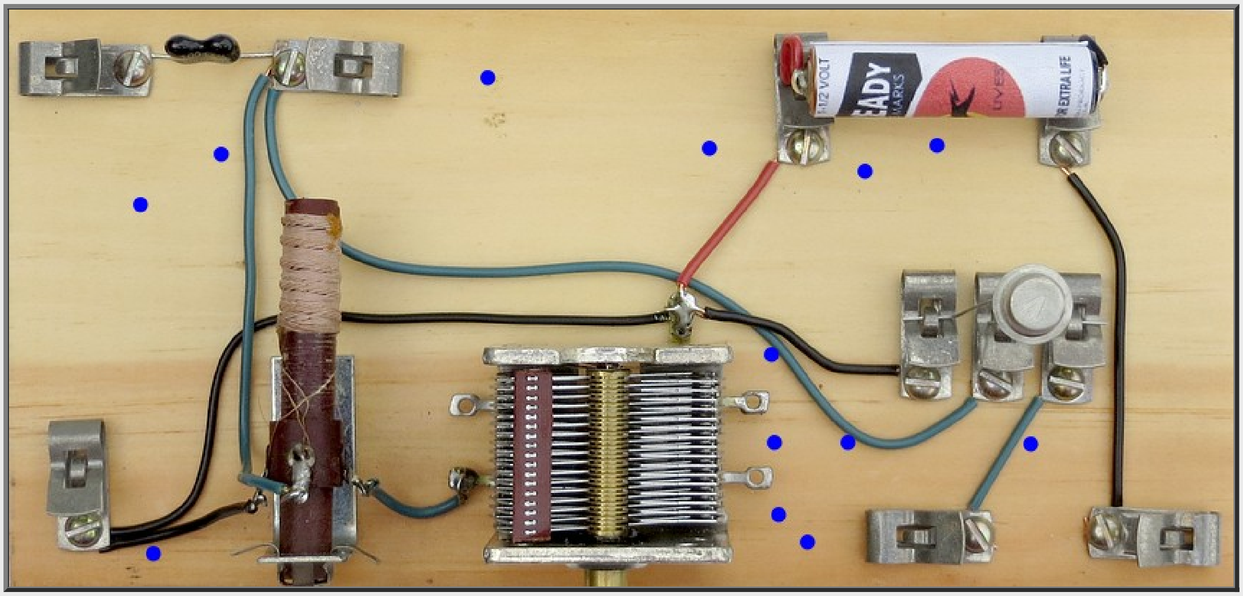

If this radio works so well, why are there so many other holes in the old base? (Click here) before you read on, so that the violin is playing.

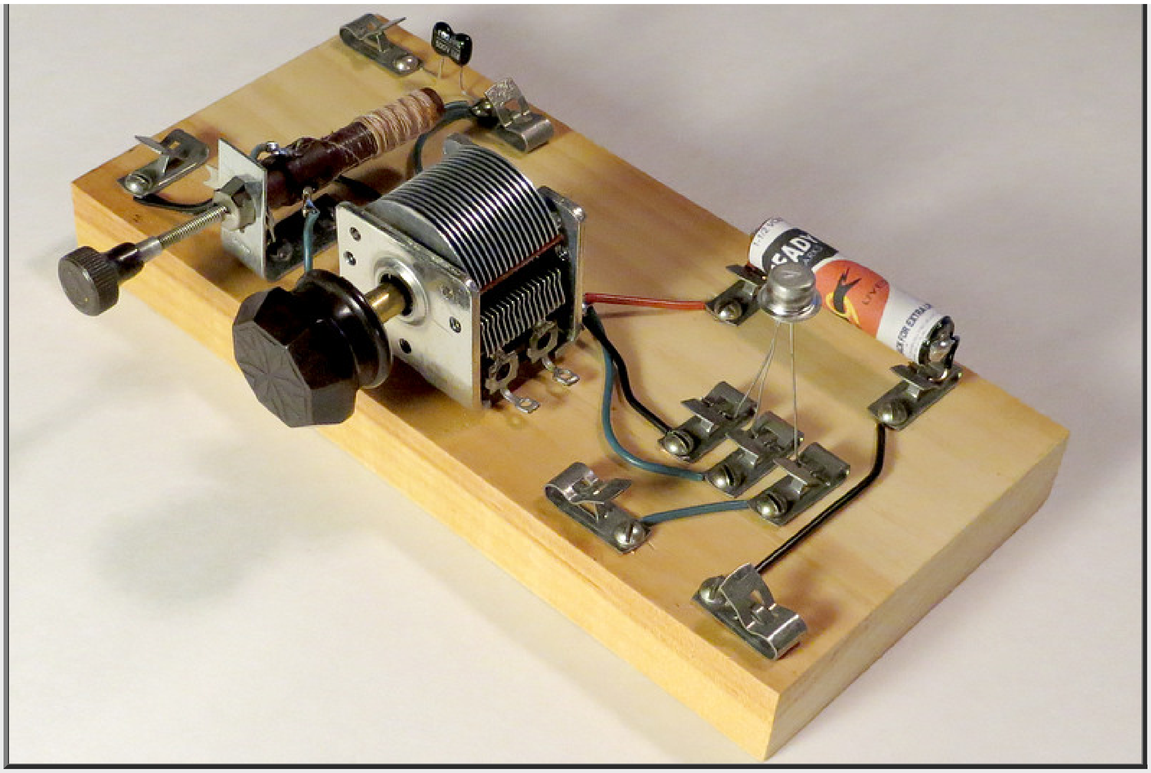

In spite of my efforts in 1966, the radio didn't work. Why didn't it work? I didn't have an actual CK722 transistor, so maybe those guys at Ree Electronics sold me the wrong part. And you probably need that socket. The socket must have something in it that makes the radio work. In other words, I didn't build it EXACTLY like the drawing, so it didn't work.

Also, notice in the drawings that one of the connections going to ground ends in a black dot. I don't know why Morgan drew it this way. I think the wire goes under the base and comes back up behind the variable capacitor. So maybe it didn't work because of that. Maybe the wire had to be under the base for the radio to work.

Actually, there is a MISTAKE in both drawings. The red dot in the top picture is supposed to be on the Collector, not the Emitter. The Base and Emitter connections are swapped in the other drawing. The schematic is correct. I may have fixed the radio by reading the text and looking at the schematic, but I don't remember. It wouldn't have made any difference at the time.

The main reason the radio didn't work was because no current flows through the transistor until you plug in the headphones. The headphones complete the circuit, and are the on/off switch. I didn't have headphones. I had a crystal earplug. The crystal earplug didn't conduct direct current through it, so the set was never "on". It may have worked perfectly but I never knew it.

I could have dropped a 2.2K resistor across the headphone clips, but I didn't know that. My sole source of radio knowledge came from the Alfred P. Morgan books.

So apparently the set was partially dismantled and used for other radio experiments. I still have the transistor I got from Ree Electronics. I plugged it into the set, and it still works. It is so beat up that I didn't use it for the picture.

That music is so sad.

Is that violin music still playing? On the next page are two more crystal sets that didn't work at first.

Next

The Crystal Radio

QST December 1997, pp. 56-57

Copyright © 1997 by the American Radio Relay League, Inc. All rights reserved

By Dave Evison, W7DE

The Crystal Radio

What's old is new again!

Even a casual auditor of the amateur bands will conclude that the hobby is largely populated by wizened old geezers—and I'm one of them. This is a sad situation because most of us entered Amateur Radio as kids. If ham radio is to continue for generations to come, we must get youngsters interested in radio communication while they're still in grammar school and still capable of wonder. We need to reach them while they are still asking questions about everyday matters (that all too often go unanswered) such as, "Why does it get light in here when I push this thing?"

Reaching today's kids won't be easy. With the Internet, cellular telephones and other forms of global communication commonplace in their lives, the true magic of radio lies buried beneath the technology. We must strip away layers upon layers of complicated hardware and reveal radio in its purest form. The perfect tool to achieve this goal is the "lowly" crystal receiver.

Imagine the curiosity you'll inspire when you demonstrate a radio that has no obvious power source whatsoever! The simple crystal receiver described in this article works surprisingly well. It's selective and produces robust headphone volume. I've been able to tune several local AM broadcast stations with ease. Best of all, it's inexpensive, uses readily available parts, and can be built by a youngster (with a little guidance). Of course, adults will love this little receiver, too.

There is a fascinating resurgence of interest in crystal radios. Even hams normally accustomed to owning rigs with four-digit price tags are rediscovering the joys of simple, "powerless" reception. To get a taste of what is going on, pick up a copy of *Crystal Set Projects* from the Xtal Set Society, Box 3026, St Louis, MO 63130; tel 314-725-1172, or visit their Web site at **http://www.midnightscience.com**. (The book was reviewed in the August, 1997 *QST*, page 102.)

About the Circuit

The circuit shown in Figure 1 was developed in the early '30s by Elmer G. Osterhoudt. Selectivity problems plagued most crystal sets of the day due to the proliferation of broadcast stations. Elmer devised a solution. He determined that poor crystal receiver selectivity was caused by the headphones swamping the tuned circuit and lowering its Q. The net result was an inability to separate individual signals. Elmer improved selectivity considerably by connecting the detector near the cold (ground) end of the tuned circuit to minimize the swamping.

Capacitor C2 is used to match the antenna system, and C3 is the main tuning capacitor. Absent from this circuit is the finicky galena-and-catwisker detector. It has been supplanted by a 1N34 germanium diode that is connected near the grounded end of the tuned circuit that includes coil L1. C1 is an RF bypass capacitor.

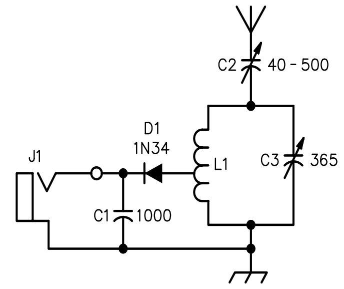

Figure 1—Schematic diagram of the crystal receiver. Parts are available from Radio Shack or Antique Electronic Supply (6221 South Maple Ave, Tempe, AZ 85283; tel 602-820-5411) as noted.

C1—1000 pF mica or disc ceramic trimmer (Radio Shack 272-123)
C2—40 to 500 pF trimmer capacitor (Antique Electronic Supply CV-263)
C3—365 pF variable capacitor (Antique Electronic Supply CV-231)
D1—1N34 germanium diode (Radio Shack 276-1123)
J1—$^{1}/_{4}$-inch, 2-conductor headphone jack (Radio Shack 274-252)
L1—35 feet of 22-gauge enamel wire. See text. (Radio Shack 278-1345)
Fahnestock clips (Antique Electronic Supply SH-11-4034)
2000-Ω headphones (Antique Electronic Supply PA-466)

Headphones

You must use high-impedance headphones (2000 Ω or greater) with crystal receivers. If you attempt to use common 8-Ω headphones, you'll have very disappointing results. You'll find high-impedance headphones at hamfest flea markets, but if you're not up to the hunt, you can buy them from Antique Electronic Supply (see Figure 1).

The Importance of a Good Antenna System

Since the crystal receiver depends completely on the energy from the transmitted signal, you need to gather as much signal out of the air as possible! Remember that this receiver is designed to tune the AM broadcast band. A half-wavelength dipole cut for about 1 MHz would be ideal, but few of us own enough real estate to accommodate a 500-foot long antenna. The practical solution is

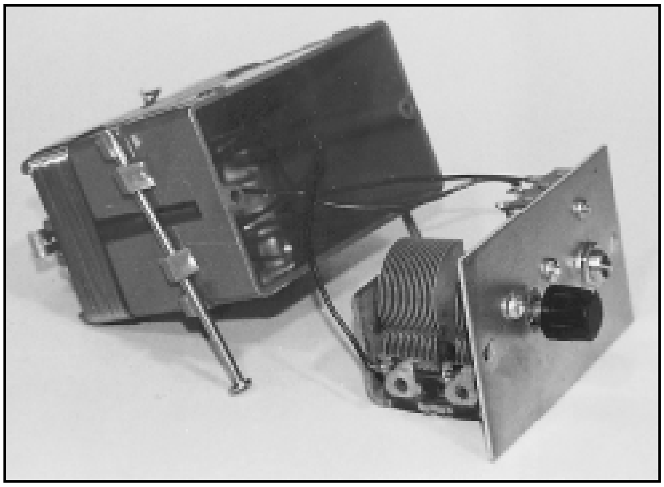

Plastic electrical outlet boxes make terrific enclosures. Note that I've mounted the main tuning capacitor, C3, and the headphone jack on the front using copper-clad PC board.

to simply use an end-fed wire about 100 feet in length. You'll also need to provide a connection to an Earth ground (a cold-water pipe may suffice). With the proper antenna system, you should be able to receive most of the AM stations in your area. If you live in a rural location without an AM broadcaster nearby (within about 50 miles), try using even more antenna wire. Just string it over trees, bushes, or any other supports.

Construction

You can build a crystal receiver into almost anything. I used plastic electrical outlet boxes. They're available at home centers and hardware stores. They are attractive, rugged and inexpensive. They even serve as forms for winding the coil. I made the front panel of my box from a piece of copper-clad PC board. You can cut this material into the proper shape using a hacksaw, and it is easily drilled for mounting parts.

For C3, a miniature variable capacitor salvaged from an old AM-only radio will work, but a single-section air variable is better. If you don't have these capacitors in your junk box, don't despair. You can order one from Antique Electronic Supply.

L1 is wound on the outside rear portion of the enclosure and consists of 32 turns of 22-gauge enamel-covered wire. The coil is tapped nine turns from the grounded end. The tap is made by tightly twisting about 1 inch of the magnet wire and scraping off the enamel insulation. You can secure the windings to the enclosure with transparent plastic tape.

The antenna trimmer capacitor (C1) and the Fahnestock clips for the antenna and ground connections are mounted on the rear of the enclosure. The tuning capacitor, phone jack and detector are mounted on the front panel. Interconnecting wires are fed through holes in the enclosure and soldered to the trimmer capacitor, ground and cold connections.

Operating Adjustments

Connect the antenna, ground and headphones. Tighten the antenna trimmer capacitor C2. Slowly adjust C3 to tune in the stations. Note the number of stations and log their call letters and frequencies. Readjust C1 to provide the best trade-off between volume and selectivity. If you're using your crystal radio near a particularly strong broadcast station, removing the ground connections will often improve selectivity (with some reduction in volume).

Use Your Radio as a Teaching Tool!

You can use your crystal radio to teach a number of important concepts. For example:

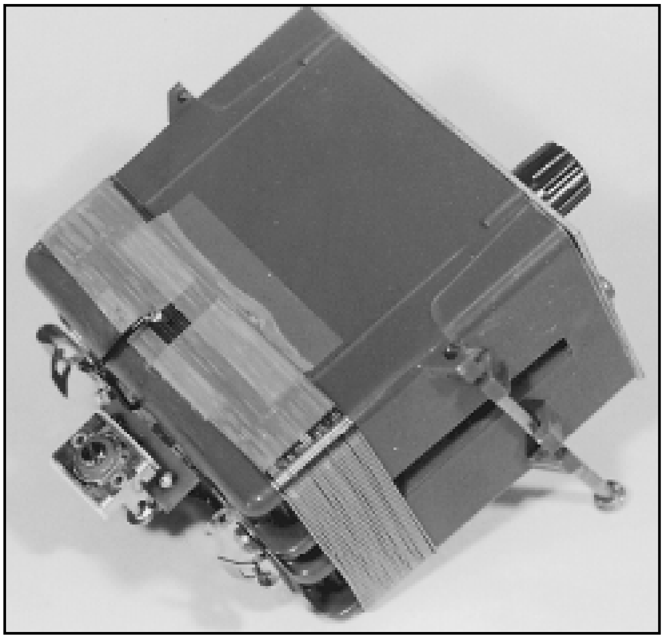

The rear view of the crystal radio enclosure. You can easily see the windings of coil L1 held in place with transparent tape.

Inductance: Applying the "right hand rule," the coil for the tuned circuit can be used to explain how a magnetic field is developed by electron flow through a conductor, and how the magnetic field is increased by forming the conductor into a coil.

Capacitance: Varying capacitance may be illustrated by simply observing how the total area increases as the rotor plates mesh with the stator plates. The compression trimmer capacitor (C1) illustrates the function of a dielectric other than air. It also shows how the capacitance can be varied by changing the separation between the plates.

Resonance, modulation and detection: You can teach these concepts by inviting your student to tune in a station while you display the modulated RF on an oscilloscope! Move the 'scope probe from one side of diode D1 to the other to illustrate "before" and "after."

Magnetism and transduction: High-impedance headphones are ideal instructional aids to demonstrate these ideas. Disassemble a pair of headphones and show your students what's inside. Touch the permanent magnet with a screwdriver so they can see that it is indeed magnetic. Explain how the detected signal flows through the tiny headphone coil, creating a magnetic field that adds or subtracts from the field produced by the permanent magnet. This causes the metal diaphragm to vibrate, creating sound! What began as electrical energy has become mechanical energy—*transduction*.

Summary

Throughout the years, the simple crystal radio has spawned tens of thousands of radio amateurs, engineers and electronic hobbyists. And despite the antiquity of its design, it is still an effective tool for recruitment and instruction. A crystal radio is a delightful device, and the perfect way to introduce the wonder of radio to a young mind—or rekindle it in your own!

---

153 Park Ave
Palo Alto, CA 94306
e-mail w7de@aol.com

No. 96

CRYSTAL SET CONSTRUCTION

SIMPLE INSTRUCTIONS FOR MAKING NINE CRYSTAL SETS

by B. B. BABANI

There is to-day a marked revival of interest in crystal receivers, and this book has been specially prepared to introduce the home constructor to the design and building of the various types. The text is simple and free from technicalities, so that even the young reader may confidently embark on the construction of any of the receivers shown.

CONTENTS

1. AERIALS.
2. EARTHS.
3. CRYSTAL DETECTORS.
4. COILS.
5. THE BEGINNERS' CRYSTAL RECEIVER.
6. FRAME AERIAL RECEIVER.
7. LONG DISTANCE RECEIVER.
8. HIGH GAIN RECEIVER.
9. HIGH SELECTIVITY RECEIVER.
10. BATTERY AIDED CRYSTAL RECEIVER.
11. MEDIUM AND LONG BAND RECEIVER.
12. ULTRA SENSITIVE ALL WAVE RECEIVER.
13. AN ADVANCED CRYSTAL DIODE RECEIVER.

THERE ARE 12 DIAGRAMS

LONDON: BERNARDS (PUBLISHERS) LIMITED

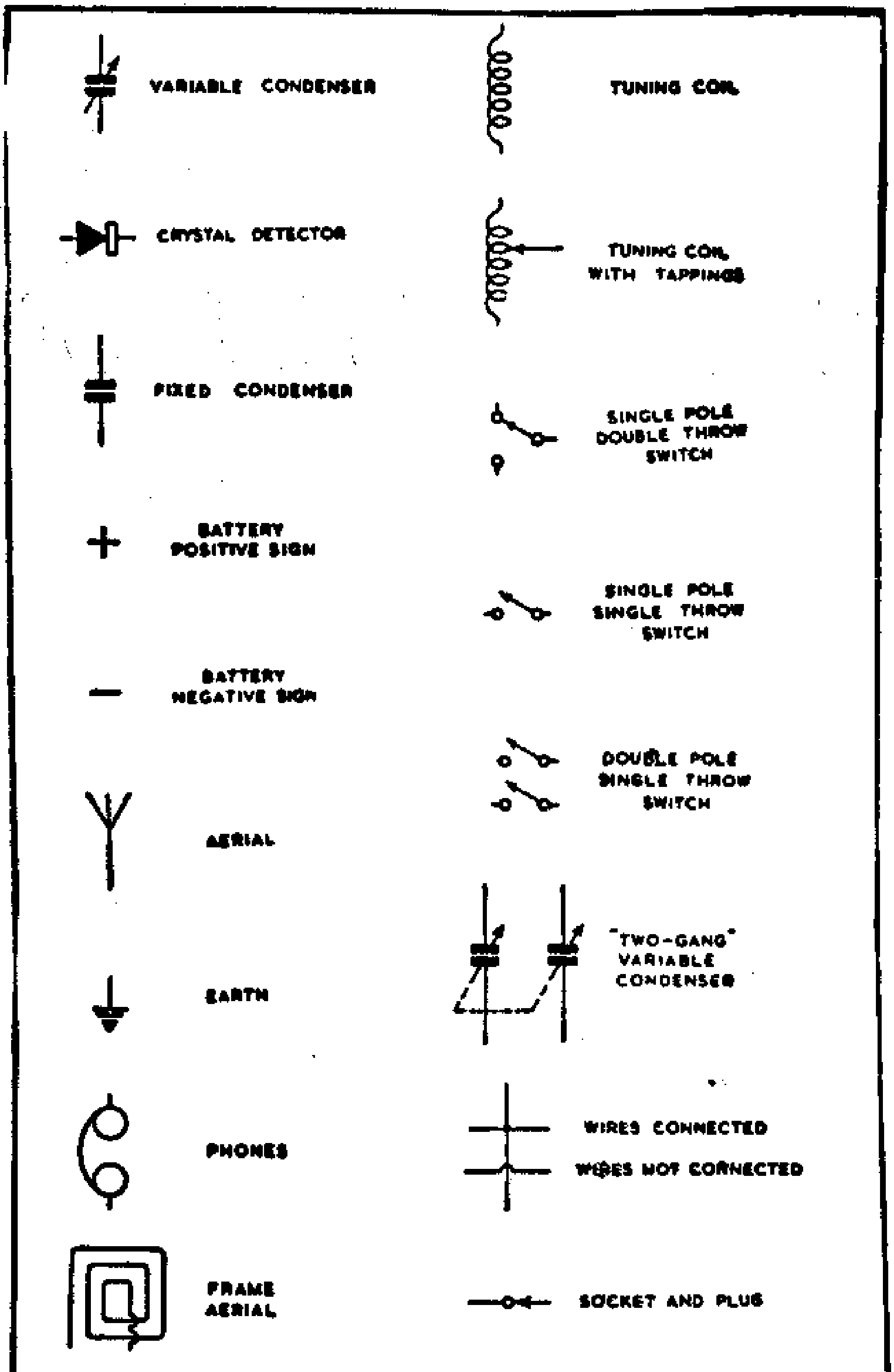
VARIABLE CONDENSER
TUNING COIL
CRYSTAL DETECTOR
TUNING COIL WITH TAPPINGS
FIXED CONDENSER
SINGLE POLE DOUBLE THROW SWITCH
BATTERY POSITIVE SIGN
SINGLE POLE SINGLE THROW SWITCH
BATTERY NEGATIVE SIGN
DOUBLE POLE SINGLE THROW SWITCH
AERIAL
"TWO-GANG" VARIABLE CONDENSER
EARTH
WIRES CONNECTED
WIRES NOT CONNECTED
PHONES
FRAME AERIAL
SOCKET AND PLUG

FIG. I SYMBOLS USED IN THE CIRCUITS DESCRIBED.

1. Aerials.

One of the essentials for satisfactory reception with crystal sets is an efficient aerial. Four different types are suggested and are shown in Fig. 2.

There are several important points to be noted when erecting an aerial for a crystal set.

(1) See that it is suspended as high as possible and is not screened by tall buildings, obstructions, etc.

(2) The aerial itself and the lead-in from the aerial must be adequately insulated. Be sure that the porcelain or plastic insulators do not permit the aerial or lead-in to touch the building, tree, or post from which they are suspended. The lead-in should be brought into the house through an insulated rod.

(3) It is exceedingly important to use insulated copper wire, either single or multi-stranded, for both the aerial and the lead-in.

2. Earths.

An adequate and efficient earth contributes to the good performance of any well-designed crystal set and Fig. 3 shows the best method of obtaining an efficient earth where there is a garden or patch of ground available at a short distance from the receiver.

Obtain a 3-ft. length of ½-in. galvanised iron pipe and at one end drill a hole right through. Insert a bolt with spring washer and locking nut in the hole. Connect the earth lead-in wire to the bolt; the other end of the lead-in to the earth connection on the receiver. Leave both ends of the pipe open. Insert the end opposite the bolt into the ground vertically until only 3-in. are showing above the surface of the surrounding earth.

It is important to see that the surrounding ground is kept damp, if necessary by occasionally pouring a pint or two of water down the open projecting end of the pipe.

Ready-made copper earthing rods with a pointed end for easy hammering into the earth are sometimes obtainable from radio dealers.

Should a garden not be available, an alternative earth is a cold water pipe coming from a rising main. Simply file very lightly the pipe at the back of the tap and wind the copper wire earth lead-in round the filed part and cover the joint with insulating tape.

In no circumstances should a gas pipe be used for an earth.

3. Crystal Detectors.

Most radio dealers can supply very efficient permanent or semi-permanent crystal detectors of proprietary makes. However, there are always available quite a number of ex-Government surplus units which are called crystal valves and, though these were originally designed for a very different use in service equipment, they are easily adaptable and will give extremely efficient results when used in circuits designed for them. To enable the constructor to purchase suitable types for this purpose—and they cost very little—the following list gives the Government reference numbers.

CV 101, CV 102, CV 103, CV 111, CV 112, CV 113, CV 226, CV 241, CV 246, CV 247, CV 253, CV 291, CV 351, CV 361, CV 364, CV 367, CV 727, CV 749, and CV 1785.

One of the satisfactory proprietary makes of this class of crystal detector is the BTH CS7A, which, incidentally, is the commercial equivalent of the CV 253.

For all these Government types, one connection should be made to each end by arranging a holder to fix them into or by carefully soldering the appropriate connecting wires to each end.

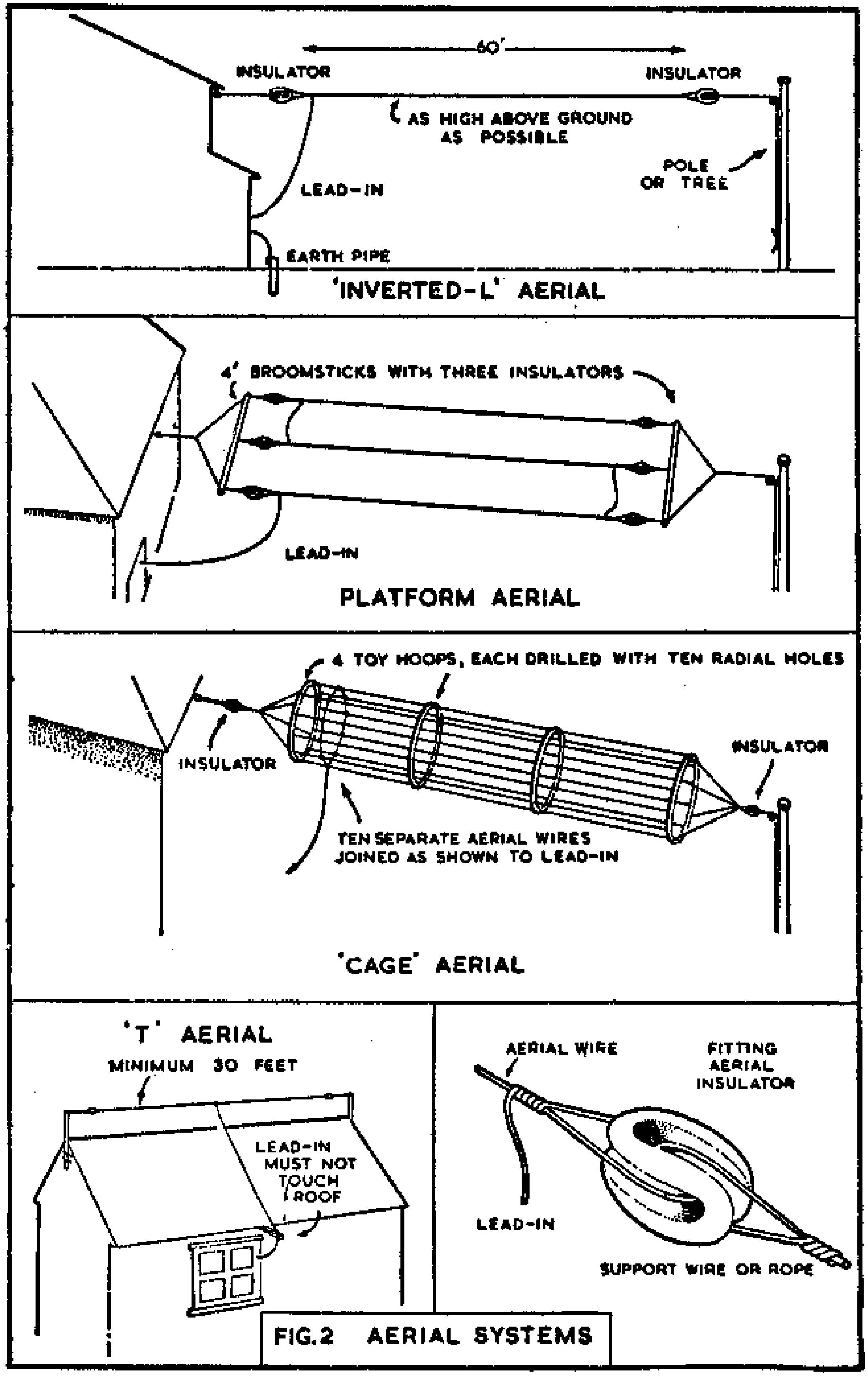
60'
INSULATOR
INSULATOR
AS HIGH ABOVE GROUND AS POSSIBLE
POLE OR TREE
LEAD-IN
EARTH PIPE
'INVERTED-L' AERIAL
4' BROOMSTICKS WITH THREE INSULATORS
LEAD-IN
PLATFORM AERIAL
4 TOY HOOPS, EACH DRILLED WITH TEN RADIAL HOLES
INSULATOR
INSULATOR
TEN SEPARATE AERIAL WIRES JOINED AS SHOWN TO LEAD-IN
'CAGE' AERIAL
'T' AERIAL
MINIMUM 30 FEET
LEAD-IN MUST NOT TOUCH ROOF
AERIAL WIRE
FITTING AERIAL INSULATOR
LEAD-IN
SUPPORT WIRE OR ROPE

FIG.2 AERIAL SYSTEMS

4. Coils.

In Fig. 3 two methods are shown of fixing the coil to the chassis. However, the ingenious constructor may evolve another arrangement.

The important thing to remember is that the coil should be fixed so that it is completely insulated from the metal chassis that may be used.

It is also important to remember that if a metal chassis is used, one must see that all components and wiring, with the exception of earth connections, are adequately insulated from this chassis so as to make sure that no short circuits take place. It is recommended that all crystal sets should be built on a panel, and that both panel and chassis should be of bakelite, paxolin or perspex, as these three materials are efficient insulators in themselves.

5. The Beginners' Crystal Receiver.

The popularity of the crystal receiver among beginners is undoubtedly due to the very modest outlay necessary, and the ease with which such receivers may be built. It is not always appreciated that, to obtain worth while results,

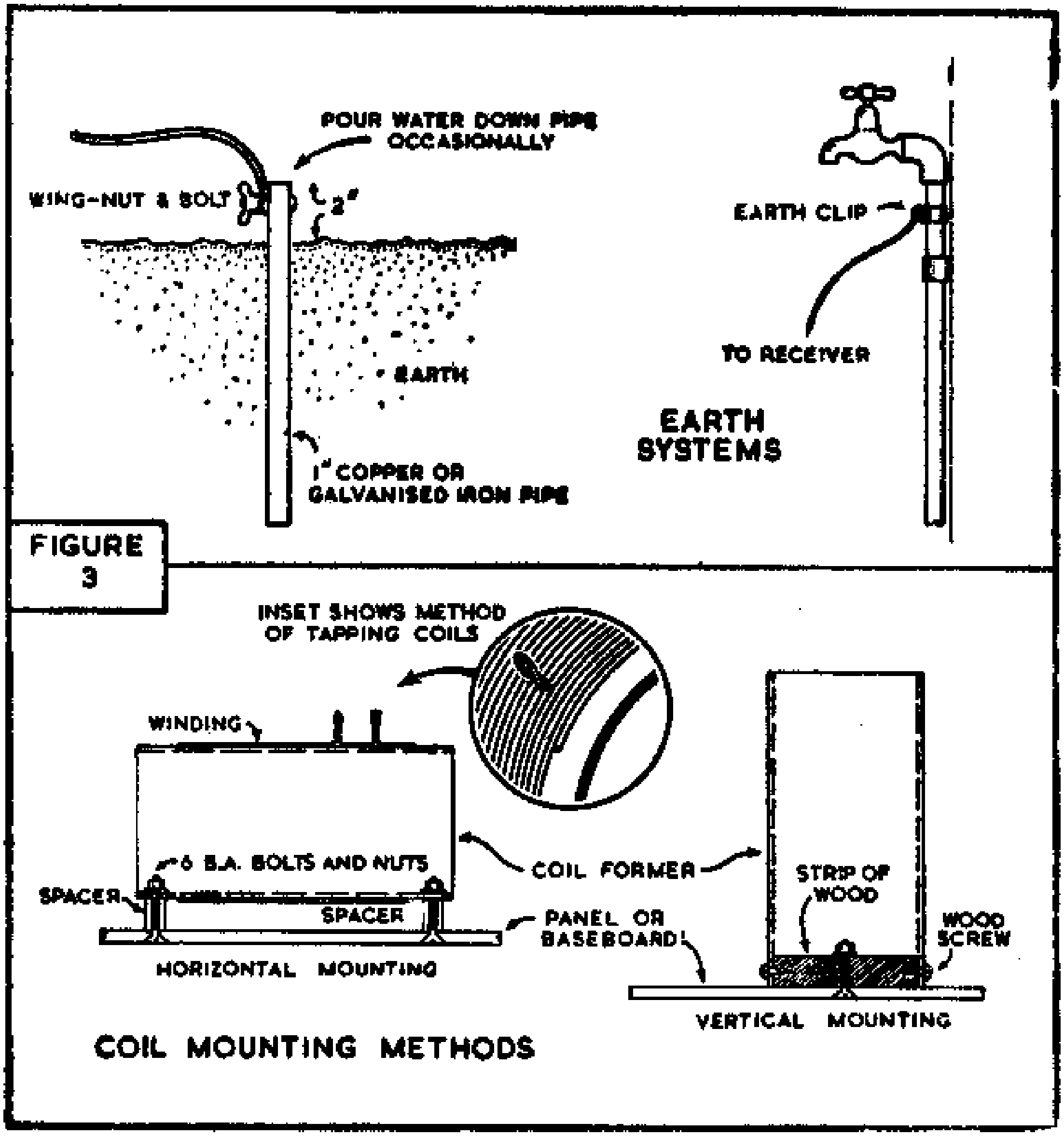
POUR WATER DOWN PIPE OCCASIONALLY
WING-NUT & BOLT
½"
EARTH
1" COPPER OR GALVANISED IRON PIPE
EARTH CLIP
TO RECEIVER
EARTH SYSTEMS
INSET SHOWS METHOD OF TAPPING COILS
WINDING
6 B.A. BOLTS AND NUTS
COIL FORMER
SPACER
SPACER
PANEL OR BASEBOARD
STRIP OF WOOD
WOOD SCREW
HORIZONTAL MOUNTING
VERTICAL MOUNTING
COIL MOUNTING METHODS

FIGURE 3

far more care in avoiding losses is necessary than with a three or four valve receiver. One reason for this is the fact that the crystal receiver can offer no help by way of amplification and it is therefore dependent on an efficient aerial/earth system, low loss design within the receiver itself, and the use of sensitive headphones. The experience gained is well worth the time expended since, when more ambitious receivers are attempted, the constructor will remember the benefits gained by efficient construction and installation with the result that each receiver attempted will be an instrument capable of giving a first class performance.

Losses within the receiver itself may be considerably reduced by efficient coil design. Unfortunately, selectivity is a factor which must not be overlooked; it is this factor which governs the sharpness of the tuning and permits the separation of adjacent powerful stations. The greater the selectivity, the greater the ease with which signals may be separated; but at maximum selectivity the received signals are at minimum strength. In Fig. 4, the coil is tapped at four points, which enables the best point between selectivity and sensitivity to be chosen. The nearer the aerial tap to the earth end of the coil, the greater the selectivity. The pre-set condenser C1 also aids selectivity; at minimum capacity the selectivity is high, and *vice versa.*

Used in conjunction with the coil taps, the constructor should have no difficulty in selecting a condenser setting and tapping point which permit the reception of strong signals and at the same time effect the required separation of powerful stations.

The coil is constructed on a former 3-in. diameter × 2-in., and consists of 50 turns 36 S.W.G. enamelled wire close wound and tapped at every tenth turn. These taps are brought out in the form of loops twisted firmly so that there is no actual break in the winding continuity. After completion the winding may be painted with Durafix to prevent any movement, and the taps and two free ends cleaned with emery cloth to remove the enamel. The layout and wiring is shown in Fig. 4. The headphones should have a resistance of some 4,000 ohms or more and should be of reputable manufacture. Any of the aerials previously described are suitable for use with this receiver.

C2 .0005 variable condenser.
C1 .0002 pre-set condenser.
C3 .001 mica condenser.
1 Crystal detector, semi-permanent pattern.
2 oz. 30 S.W.G. enamelled copper wire.
1 Coil former 3-in. dia. × 2-in.
1 Ebonite or Bakelite panel 6-in. × 5-in. × ⅛-in.
4 terminals, connecting wire, etc.

6. Frame Aerial Receiver for Local Station use on Medium wave bands.

This receiver was specially designed for constructors who are within 5 to 10 miles of a powerful station and who have no facilities for constructing a good outdoor aerial.

The frame itself consists of two ¾-in. × ¾-in. square section wooden rods each 36-in. in length. These are placed together at their centre line to form a cross; they are fastened by a nut and bolt and these in turn pass through an upright piece of wood ½-in. × 1-in. section 2-ft. in length. This upright support is then screwed into a baseboard 8-in. × 8-in. × 1-in. thick. At each of the arms of the 36-in. cross a 4-in. strip of ½-in. square section wood is placed. Reference to Fig. 5B will give the details of the construction of this aerial quite clearly.

It will be noted that approximately 85-ft. of 20 S.W.G. double cotton-covered wire are required for the inductance to cover the medium wave broadcast band. The wire should be wound around the four crossbars, thereby creating

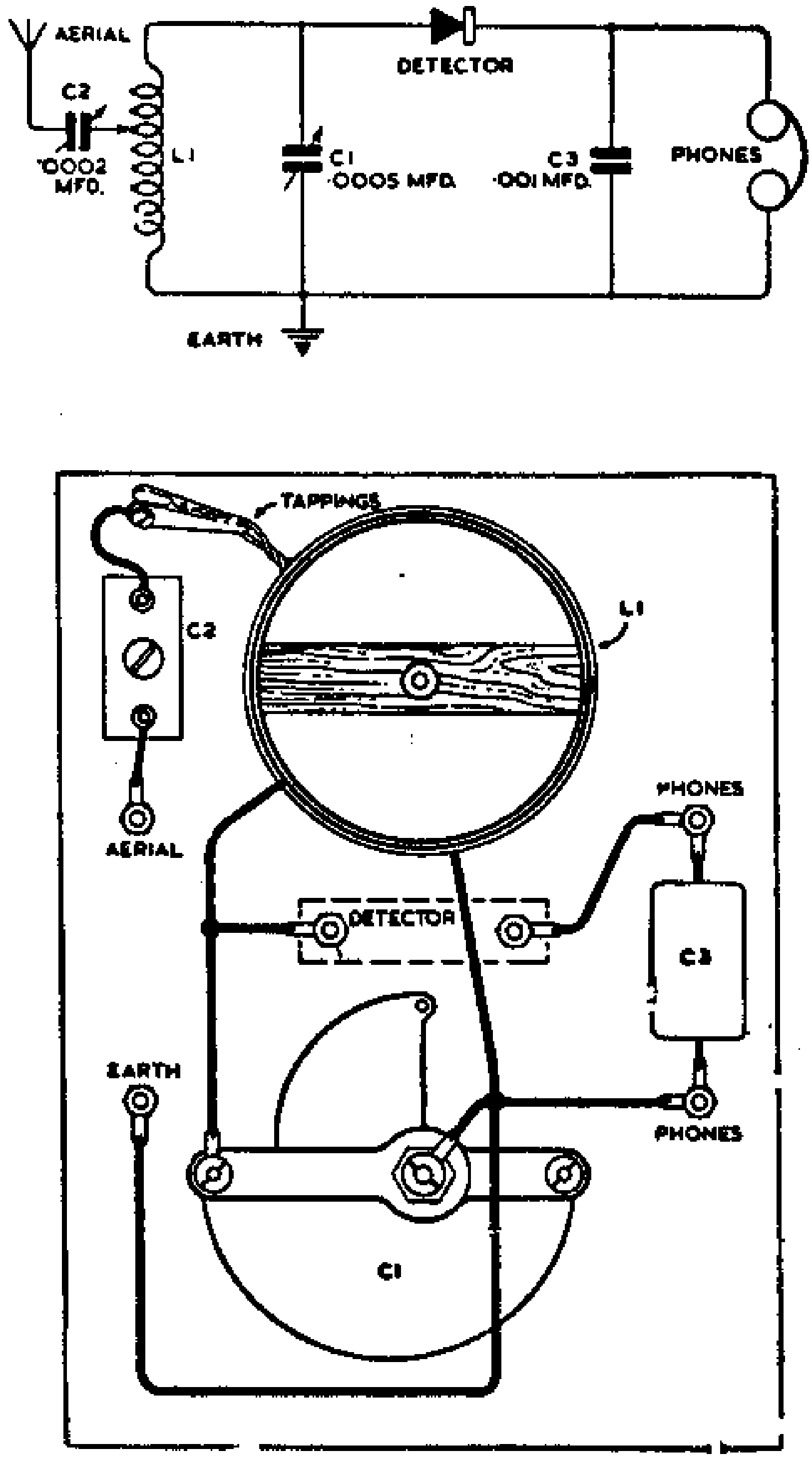
AERIAL
C2
·0002 MFD.
L1
C1 ·0005 MFD.
DETECTOR
C3 ·001 MFD.
PHONES
EARTH
TAPPINGS
C2
L1
PHONES
AERIAL
DETECTOR
C3
EARTH
PHONES
C1

FIG. 4 THE BEGINNER'S CRYSTAL RECEIVER.

a square wire frame, and each turn of the wire should be spaced approximately ⅛-in. The best way to effect this spacing accurately is to file on the 4-in. crossbars a number of small nicks at ⅛-in. intervals. The two ends of the wire should then be brought down on the two terminals which may be screwed into the square baseboard and two leads taken from there to the crystal set itself.

The actual set may be constructed on a small ebonite, paxolin or perspex panel 8-in. × 5-in. which may be screwed to one side of this base supporting piece of wood, thereby making the set virtually self-contained except for the earphones. Fig. 5A shows the circuit of the receiver.

To obtain the best results with this receiver, the aerial should be rotated by hand until the signals become loudest.

1 Semi-permanent crystal detector.
C1 1 .0005 mfd. variable condenser.
C2 1 .0005 mfd. fixed condenser.
1 condenser dial marked 0 to 100 or 0 to 180.
1 pair of high sensitivity earphones.
85-ft. to 90-ft. of 20 S.W.G. double cotton covered wire. (8oz.)
Sufficient connecting wire for wiring purposes.

7. Long Distance Receiver.

This receiver has been specially designed to achieve high sensitivity. With a sufficiently good outdoor aerial effective reception has been obtained up to 150 miles.

The circuit is quite straightforward, as can be seen from Fig. 6 and the only special point is in winding the coil L1.

This coil consists of a 2½-in. diameter former on which are wound a total of 51 turns using 24 S.W.G. double cotton wire. Starting at one end, tap every two turns until there are eight taps; wind a further 15 turns and tap; then a further 10 turns and tap; finally, add 10 turns to end the winding. The wire should be close wound, *i.e.*, with turns touching.

To obtain the best results, the two crocodile clips A and B should be tried on the various coil taps until the loudest reception is obtained. It may be found in use that the loudest reception of any particular station causes another station to be heard at the same time; care should, therefore, be taken to select the appropriate tapping point to separate any two powerful stations that may interfere with one another.

4 oz. 24 S.W.G. DCC copper wire.
Coil former 2½-in. dia. × 3-in.
C1-2 single variable condensers .00035 mfd.
2 Condenser dials marked from 0 to 100 or 0 to 180.
1 Permanent crystal detector.
C3 .001 fixed condenser.
1 pair of high sensitivity earphones (preferably between 4,000 ohms and 8,000 ohms impedance).
2 Crocodile clips.

8. High Gain Receiver.

This receiver has been designed to give good reception up to 50 miles distance from any normal power broadcasting station in the medium band. It is a very selective circuit and will therefore enable the user to separate closely situated stations.

It will be noticed in the circuit (Fig. 7) that the variable condenser C1, has an on/off switch S1 to throw it out of circuit if necessary. It is used to increase the selectivity of the receiver and, in operating, it should be tried with this condenser either in or out of circuit on the various tappings on the coil L1.

L1 consists of 90 turns of 22 S.W.G. double cotton covered wire wound on a 2-in. diameter former and tapped every ten turns. L2 separately wound on a 2-in. former, consists of 90 turns of 22 S.W.G. double cotton covered wire, also tapped as L1. The components required are given on page [illegible]

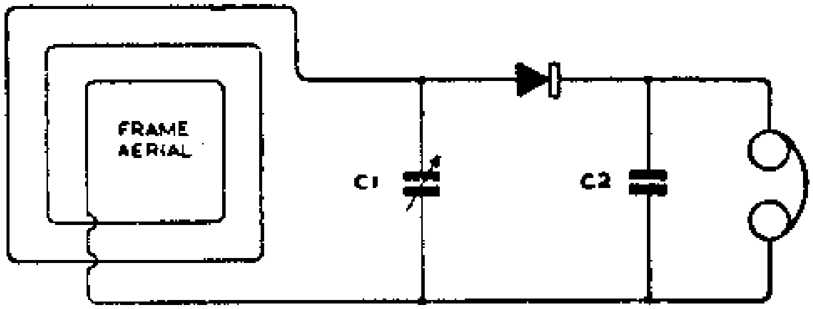
FRAME
AERIAL
C1
C2

FIG. 5A FRAME AERIAL CRYSTAL RECEIVER

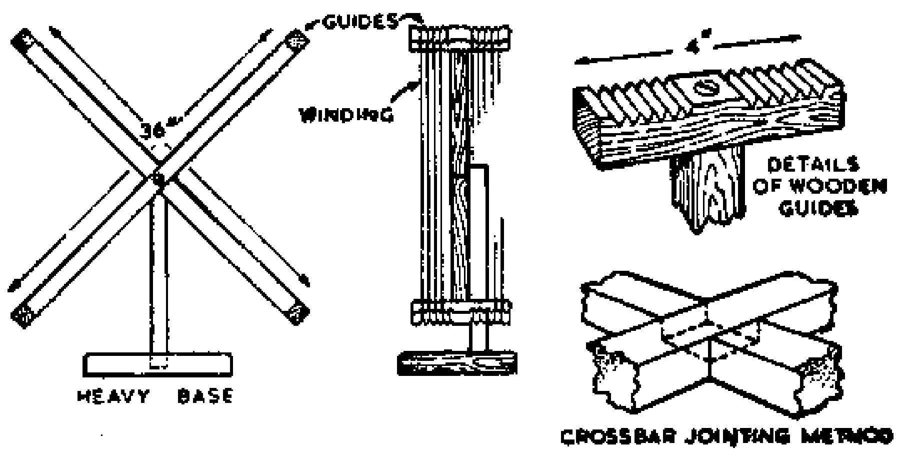
GUIDES
36"
WINDING
HEAVY BASE
4"
DETAILS
OF WOODEN
GUIDES
CROSSBAR JOINTING METHOD

FIG. 5B. DETAILS OF FRAME AERIAL

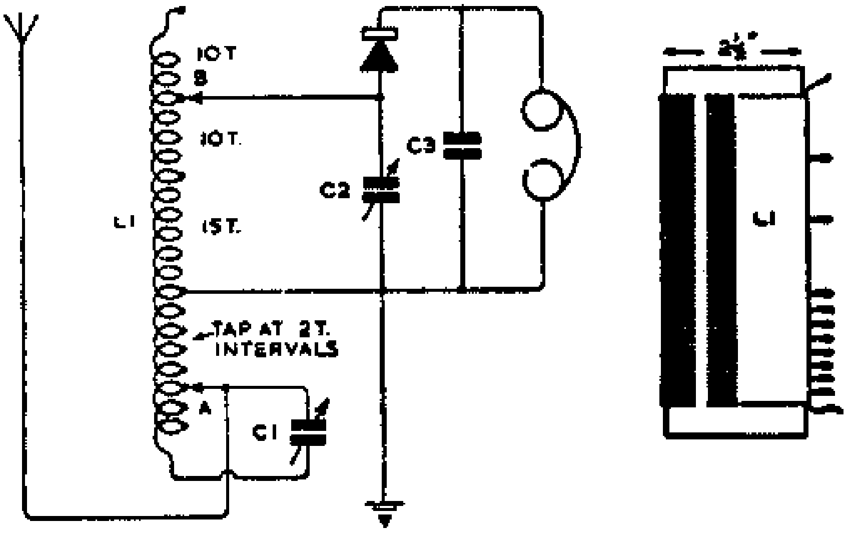
10T
B
10T.
15T.
L1
TAP AT 2T.
INTERVALS
A
C1
C2
C3
2½"
L1

FIG. 6 LONG-DISTANCE CRYSTAL RECEIVER

C1-2 Single variable condensers .00035 mfd.
S1 Single pole single throw switch.
C5 .0001 preset condenser.
1 Permanent crystal detector.
C3-4 .001 mfd. fixed condenser.
1 Pair of high sensitivity earphones.
8 oz. 22 S.W.G. DCC.
2 Dials for variable condenser marked 0 to 100 or 0 to 180.

9. High Selectivity Receiver.

This circuit (see Fig. 8) is for use in high saturation strength areas where it is desired to separate powerful stations. The circuit is not complicated and is extremely effective.

It will be noted that C1 and C2, each .00035 mfd. capacity, are 2 gang unit.

The only special point in this set is the design of coils L1, 2 and 3, which are constructed as follows: L1, 60 turns of 22 S.W.G. double cotton covered wire on a 3 × 4 inches former tapped every five turns; L2, 15 turns of 22 S.W.G. double cotton covered wire on a 3 × 6 former; and L3, 85 turns of 22 S.W.G. double cotton covered wire wound on the same former as L2, the two windings being separated by approximately ¼-in. (see Fig. 8).

Note that the aerial lead is brought down to one of the tappings on wire L1. This should be tried on each of the tappings and the one giving the required selectivity should be used.

C1-2 2-gang variable condenser .00035 mfd.
2 3-in. formers for coils L1-2-3.
8 oz. 22 S.W.G. DCC.
C3 .0005 fixed condenser.
1 Semi-permanent crystal detector.

10. Battery-aided Crystal Receiver.

Fig. 9 shows a design which utilises a 9-volt grid bias battery to improve performance.

The charge built up in the .02 mfd. fixed condenser is changed to a positive current when it passes through the coil L1. This, in turn, is superimposed on the positive charge that is commencing to be rectified by the crystal detector and thereby increases the signal strength.

Coil L1 consists of a 4-in. diameter former wound with 85 turns of 24 S.W.G. double cotton covered wire tapped at 5, 25, 45 and 65 turns from the earth end. The earth end and these four taps are brought out to five small sockets and the connection from the earth is taken to a small socket. The plug should be tried in each of the five positions to find which one gives the greatest signal strength.

The 9-volt battery is connected across the .02 mfd. condenser and special care should be taken to see that the battery is connected correctly, *i.e.*, negative pole to earth.

1 Permanent crystal detector.
1 4-in. dia. × 3-in. former.
8 oz. 20 S.W.G. DCC copper wire.
C1 Variable condenser .0005 mfd.
C2 Fixed condenser .005 mfd.
C3 Fixed condenser .02 mfd.
1 pair of high sensitivity headphones.
1 Clix plug.
5 Clix sockets.
1 9-volt grid bias battery.

11. Medium and Long wave band Receiver.

This set has been designed for reception on two wave bands.

Care must be taken in winding the coils L1, 2, 3 and 4. These are wound on a single 3-in. diameter former. L1 consists of 30 turns of 28 S.W.G. DCC tapped at 8, 16 and 24 turns. L2, 60 turns of 32 S.W.G. DCC. L3, 40 turns of 28 S.W.G. DCC, and L4, 80 turns of 32 S.W.G. DCC. All are close wound and are separated from each other as follows: between L1 and L2, 3/16-in.;

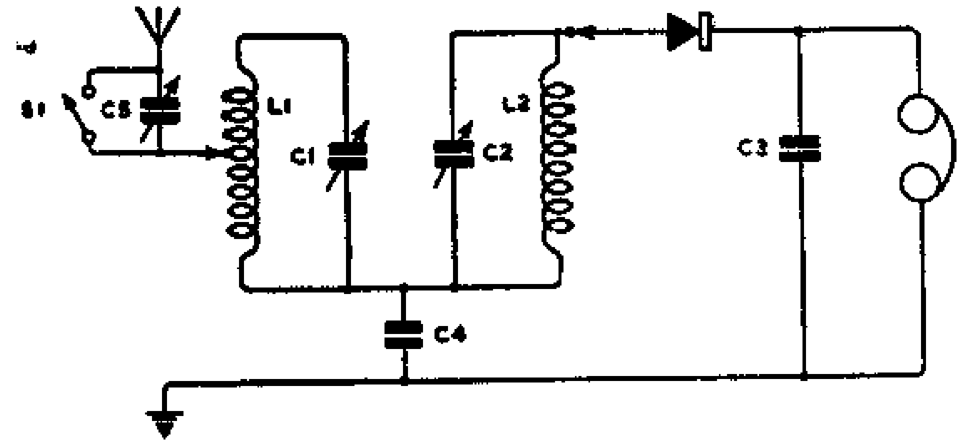
S1
C5
L1
C1
C2
L2
C3
C4

FIG. 7 HIGH-GAIN CRYSTAL RECEIVER

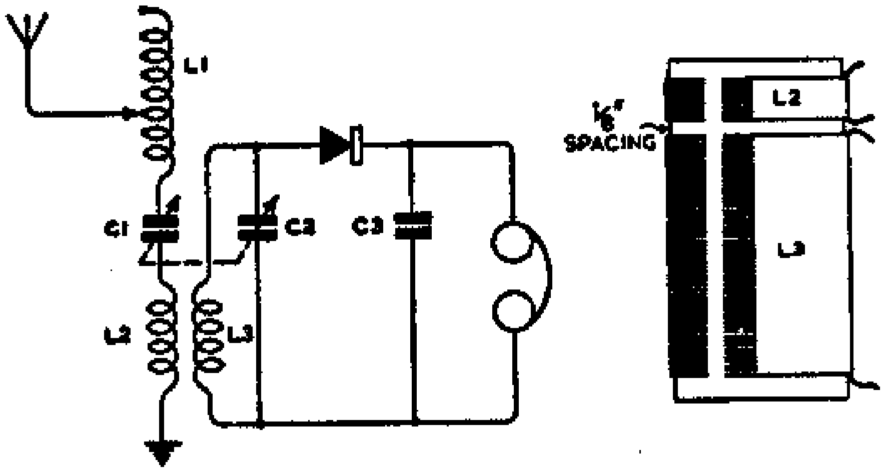
L1
C1
C2
C3
L2
L3
1/8" SPACING
L2
L3

FIG. 8 HIGH SELECTIVITY CRYSTAL RECEIVER

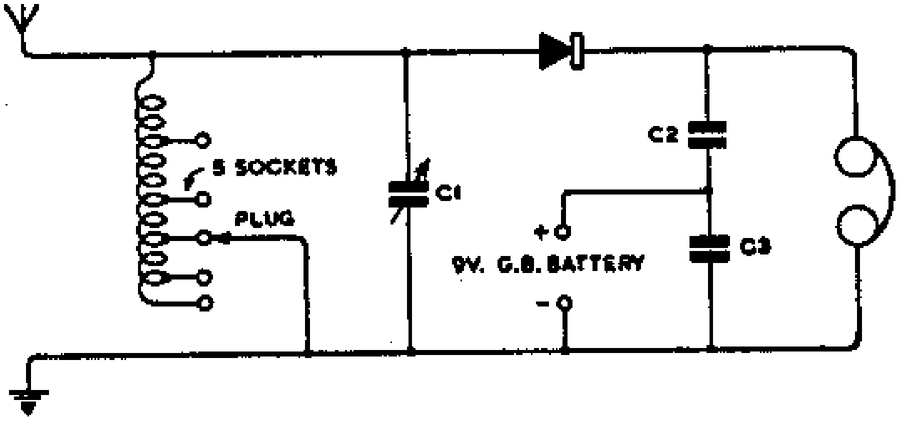
5 SOCKETS
PLUG
C1
C2
C3
+
9V. G.B. BATTERY
−

FIG. 9 BATTERY-AIDED CRYSTAL RECEIVER

between L2 and L3, 1-in.; and between L3 and L4, 3/16-in. See Fig. 10.

C1-2 single variable condensers .0005 mfd.
2 Variable condenser dials marked 0 to 100 or 0 to 180.
1 Clix plug.
3 Clix sockets.
S1 Single pole double throw switch.
C3 .001 mfd. fixed condenser.
1 Pair of high sensitivity earphones.
1 3-in. dia. × 8-in. former.
2oz of both 28 and 32 S.W.G. double cotton covered wire to make these 4 coils.
1 Permanent crystal detector.

In use, the plug should be tried in each of the three sockets to find the point which gives the best results. The switch is used as a wave band change switch.

12. Ultra Sensitive All Wave Receiver.

The receiver shown in Fig. 11 is for long distance reception. It is a novel circuit and, though apparently complicated in construction, it is really very simple to assemble with the minimum of tools and labour.

It will be seen from the circuit diagram that there are two banks of sockets, five black and six red. When using this set the appropriate black and red plugs should be inserted in whichever of the similar coloured sockets give the best results. The coil is quite easy to construct and consists of a primary winding which is inserted at the earth end of the secondary winding. For the medium wave band, the secondary winding L2 consists of 54 turns of 20 S.W.G. enamelled or cotton covered wire. This coil has four taps in addition to the starting point and the end connection, and these taps are taken at six turns, 14 turns, 27 turns and 40 turns from the earth end. The taps are taken out and a connection is made to an appropriate socket, as shown in the diagram.

The coil is wound on a 3-in. diameter former which may be a paper tube, bakelite, paxolin or perspex, and spacing is approximately 17 turns to the inch.

For short wave work down to about 30 metres, L2, the secondary coil, will consist of the same diameter former wound with 15 turns of the same gauge wire, the four taps being taken at 3, 6, 9 and 12 turns from the earth end.

L1, the primary coil, is used for both wave bands and consists of 11 turns of similar gauge wire to that used for L2, wound on a 2-in. × 1-in. former, the winding to be spaced the same as L2. The coil is then inserted at the earth end of whichever secondary coil L2 is being used.

The crystal detector may be one of the semi-permanent types freely available to-day; the old-fashioned cat's whisker pattern; or one of the silicon permanent detectors which are sold by many surplus dealers.

The 'phones should be of a standard type 4,000 ohms impedance although 'phones of other values will probably operate quite satisfactorily with this powerful receiver. It should, however, be stressed that the better the headhones used, the more satisfactory the reception.

Coils for other wave bands may be experimentally designed by the constructor, remembering that four tappings are advisable for the secondary coil L2.

A long aerial is very desirable with this receiver, placed as high as possible. It is recommended that a minimum length of 70-ft. be allowed, to include the aerial and lead-in.

The values of the components are shown in the following components list. The constructor should experience no difficulty in obtaining excellent results from this cleverly designed receiver.

C1 .0005 variable condenser.
C2 .002 fixed condenser.
1 semi-permanent crystal detector.
1 coil former 3-in. dia. × 4½-in.
1 coil former 2-in. dia. × 1-in.
½ lb. 20 S.W.G. enamelled or DCC copper wire.
6 red wander plug sockets.
6 black wander plug sockets.
1 red and 1 black wander plug.

Headphones, terminals, connecting wire

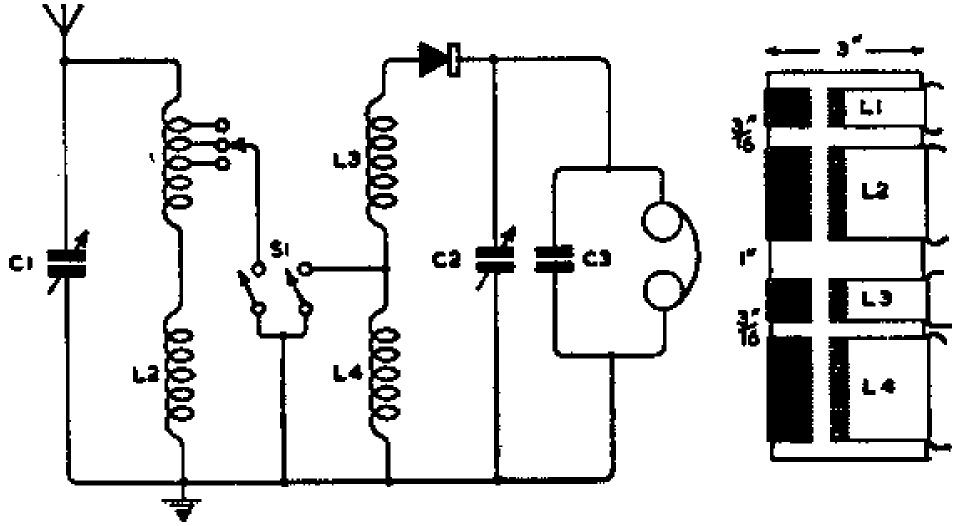
C1
L2
S1
L3
L4
C2
C3
3"
L1
L2
L3
L4
3/16"
1"
3/16"

FIG. 10 MEDIUM AND LONG-WAVE CRYSTAL RECEIVER

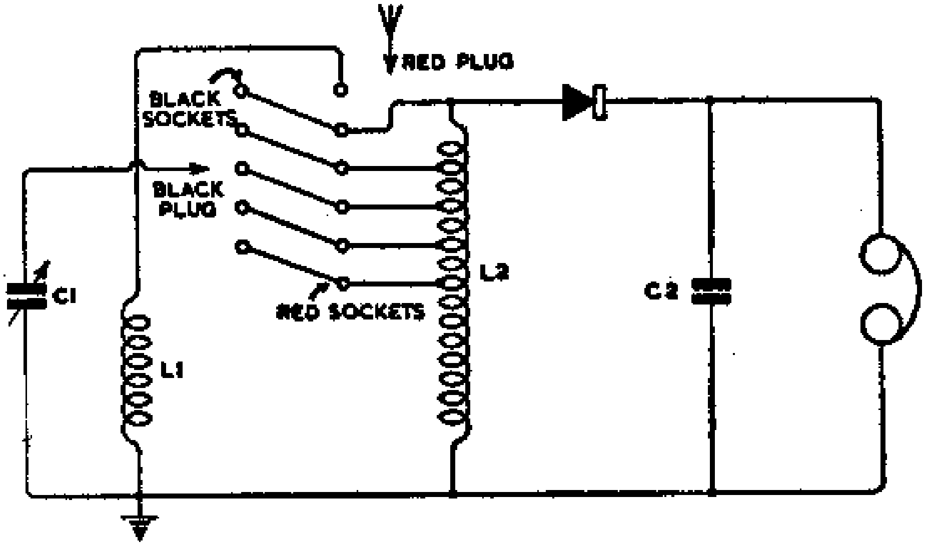
RED PLUG
BLACK SOCKETS
BLACK PLUG
RED SOCKETS
C1
L1
L2
C2

FIG. 11 ULTRA-SENSITIVE ALL-WAVE CRYSTAL RECEIVER

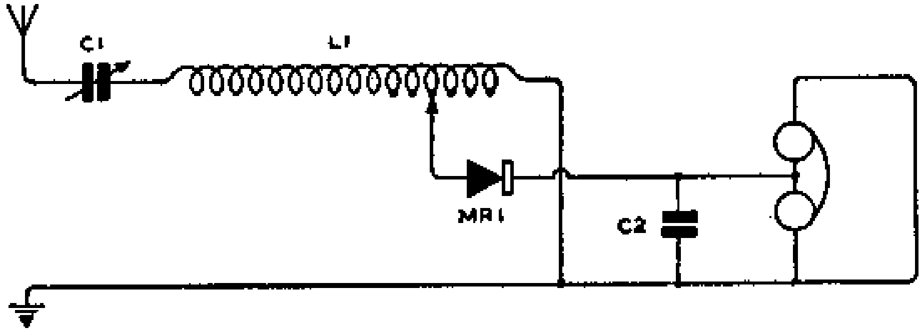

FIG. 12 ADVANCED CRYSTAL DIODE RECEIVER

13. An Advanced Crystal Diode Receiver.

The modern crystal diode of the radar type does not require any adjustment and thus has a distinct advantage over the ordinary crystal detector which needs delicate adjustment—even the semi-permanent type.

Such crystals are now obtainable on both the regular and the surplus markets, and a popular type is the B.T.H. CS7A silicon crystal (also coded as CV 253).

A rather different circuit from the usual crystal set arrangement is needed to suit the characteristics of a crystal diode. The diode must be tapped on to the tuning coil, and it is found that the tuning circuit itself gives best results if a series-tuned acceptor circuit is used. The circuit of a radar crystal receiver is shown in Fig. 12. The type of aerial used with this circuit has a very great bearing on the behaviour of the receiver for, in effect, the tuned circuit, with a series resistance equivalent to the reflected crystal load resistance, is in series with the capacitance of the aerial to earth and the aerial's effective series resistance.

At resonance—when the combination is tuned to any particular signal—the inductance resonates with the capacitances of the tuning condenser and the aerial in series, and the final effect is that the reflected load of the crystal, in series with the coil's R.F. resistance, is paralleled across the aerial's series resistance.

For maximum power transfer, the crystal load resistance must be made equal to the sum of the aerial series resistance and the coil R.F. resistance, and so the method in which the crystal is tapped into the tuning coil, and the exact capacitance required to tune the receiver to any required signal, must depend to a very great extent on the aerial itself.

At the same time the crystal resistance varies with the signal strength, the resistance being high for weak signals and dropping by as much as 50 per cent. and more for strong signals, so that this effect also has a bearing on the correct coil tap. The output impedance of the crystal also varies similarly, affecting the matching of the headphones into the crystal diode, and so for any set of conditions, the receiver requires to be matched up to both the aerial and the signal being received for best results. This would mean a series of coil taps and (theoretically, not practically) a matching transformer between the diode and the headphones; but in practice it will be found that a receiver using standard parts can be built up to give very good results under various conditions.

It has been shown already that the tuning of the receiver depends to a great extent on the characteristics of the aerial, and while a .0005 mfd. variable capacitor is shown in Fig. 12 as the tuner, the constructor must be prepared to experiment with different capacitance values until the required station is tuned in. The range of reception given by the receiver is quite good, if a really long and high aerial, and a good earth connection, are used; but no more than the local station signal can be expected, and the tuning therefore must be adjusted to suit the station frequency.

It must be mentioned that the headphones are shown parallel connected. High resistance headphones of the 4,000 ohm type must be used, and if two of these are connected in parallel rather than in the more usual series method, they will provide a roughly accurate match to the crystal. If more than one pair of headphones are to be connected in, then the pairs of headphones may be left series connected in the usual manner, the sets of headphones being connected in parallel.

The capacitor across the headphone terminals completes the crystal R.F. circuit, and any value between about 0.001 and 0.005 mfd. will serve. The higher capacitance will, of course, by-pass some of the higher audio frequencies, so that if good steady reception is obtained some experiment with this condenser is also worth while.

C1, .0005 mfd. variable condenser.
C2, .005 mfd. fixed condenser.
M.R.1 CV 253 Crystal diode.
Coil former 1½-in. dia. × 4½-in.
4oz. 26 S.W.G. enamelled copper wire.

Coil details are as follows: 150 turns of 26 S.W.G. tapped at 20, 25, 30, 35, 40 and 45 turns from the earth end. At each tapping-point the wire should be twisted up into a loop and the winding then continued without breaking the wire; when the coil is completed and the ends anchored the tapping loops can be bared.

BERNARDS RADIO BOOKS

No.		Price
56.	Radio Aerial Handbook	2/6
57.	Ultra-Shortwave Handbook	2/6
58.	Radio Hints Manual	2/6
64.	Sound Equipment Manual	2/6
68.	Frequency Modulation Receivers' Manual	2/6
73.	Radio Test Equipment Manual	2/6
83.	Radio Instruments and their Construction	2/6
96.	Crystal Set Construction	1/–
99.	One Valve Receivers	1/6
100.	A Comprehensive Radio Valve Guide, Book 1	5/–
103.	"Radiofolder" A. The Master Colour Code Index for Radio and T.V.	1/6
104.	Three Valve Receivers	1/6
107.	Four Valve Circuits	1/6
108.	Five Valve Circuits	2/6
121.	A Comprehensive Radio Valve Guide, Book 2	5/–
123.	"Radiofolder" F. The Beginners' Push-Pull Amplifier	1/6
126.	Boys' Book of Crystal Sets and Simple Circuits	2/6
129.	Universal Gram Motor Speed Indicator	1/–
134.	F.M. Tuner Construction	2/6
135.	All Dry Battery Portable Construction	2/6
138.	How to Make F.M. and T.V. Aerials, Bands 1, 2 and 3	2/6
141.	Radio Servicing for Amateurs	3/6
143.	A Comprehensive Radio Valve Guide, Book 3	5/–
145.	Handbook of AM/FM Circuits and Components	2/–
146.	High Fidelity Loudspeaker Enclosures	5/–
147.	Practical Tape Recording Handbook	5/–
148.	Practical Transistor Receivers, Book 1	5/–
149.	Practical Stereo Handbook	3/6
150.	Practical Radio Inside Out	3/6
151.	Transistor Superhet Receivers	7/6
155.	Portable Transistor Radio and Radiogram	2/6
156.	Transistor Circuits Manual, No. 1	2/6
157.	A Comprehensive Radio Valve Guide, Book 4	5/–
158.	Radio, Television, Industrial Tubes, Semiconductor and Diodes Equivalents Handbook (208 pages)	9/6
159.	Realistic High Fidelity	5/–
160.	Coil Design and Construction Manual	5/–
161.	Radio, Television and Electronics Data Manual	3/6
162.	High Fidelity Stereo Gramophone	5/–
163.	Transistor Circuits Manual, No. 2	2/6
164.	High Fidelity Tape Recorder for the Home Constructor	2/6
165.	Hartley on Hi-Fi, Book 1, Radio Tuners	5/–
166.	Public Address Systems	2/6
167.	Transistor Circuits Manual, No. 3	2/6
168.	Transistor Circuits Manual, No. 4	2/6
169.	Hi-Fi Transistor F.M. Tuner for the Home Constructor	3/6
170.	Transistor Circuits for Radio Controlled Models	7/6
171.	Super Sensitive Transistorised Pocket Set Construction	3/6
172.	International Radio Stations List	2/6
173.	Practical Transistor Audio Amplifiers	3/6
174.	Transistor Subminiature Receivers Handbook	5/–
175.	Transistorised Test Equipment and Servicing Manual	3/6
176.	Transistor Audio Amplifiers Manual	6/–
177.	Modern Transistor Circuits for Beginners	7/6
178.	Comprehensive Radio Valve Guide, Book 5	6/–
179.	Transistor Circuits Manual No. 5	5/–
180.	British Semiconductor Survey	15/–
181.	22 Tested Circuits Using Micro Alloy Transistors	5/–
182.	"At A Glance" Radio Valve & T.V. Tube Equivalents	3/6
183.	How to Receive Foreign T.V. Programmes on your Television Set by simple Modifications	5/–
	Resistor Colour Code Disc Calculator	1/6
	Engineers' Reference Tables	1/6
	International Radio Tube Encyclopaedia—3rd Edition	63/–

N3FRQ's Homebrew Corner

12-volt 1-tube Regenerative Receiver:

This article first appeared in the Delaware Valley Historic Radio Club's *Oscillator*.

12-volt 1-tube Reflex Receiver:

I was so encouraged by the success of the simple regenerative set that I undertook the design of a much more elaborate one-tube set using a "reflex" circuit. In this receiver the signal passes through the tube twice: first, as radio frequencies and then a second time, as audio. Because there is considerable RF gain before the detector, this radio is far more sensitive than a simple regenerative set.

This is not a project for beginners. The design employs a lot of hard-to-get and potentially expensive parts, almost enough to build a rudimentary superheterodyne. You'll probably need a sweep generator to get the tunable band-pass filter working correctly. You should be prepared to reengineer the circuit to suit the parts on hand.

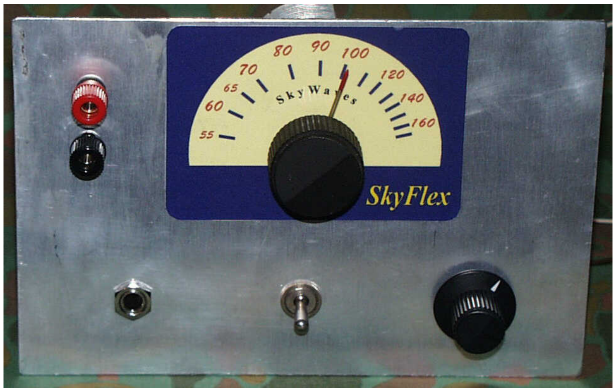

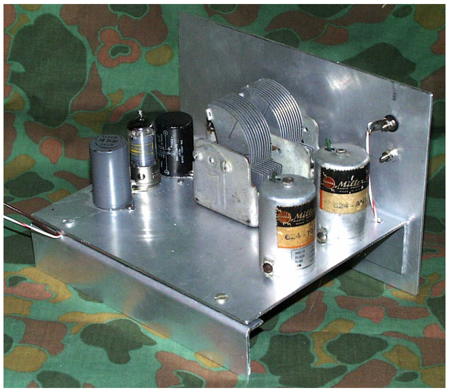

Simple battery tube radio based on the PGXS crystal set.

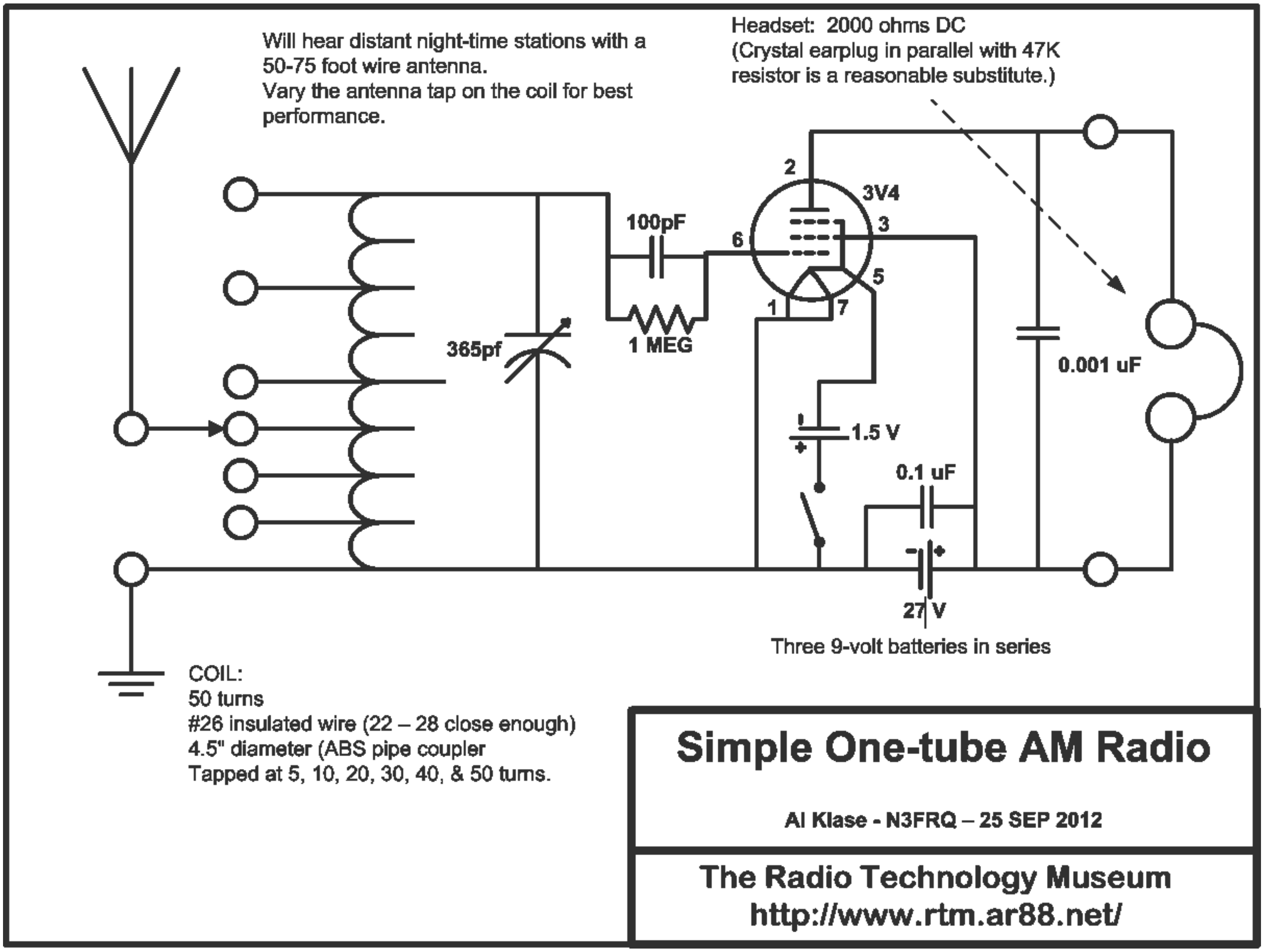

Will hear distant night-time stations with a 50-75 foot wire antenna.
Vary the antenna tap on the coil for best performance.

Headset: 2000 ohms DC
(Crystal earplug in parallel with 47K resistor is a reasonable substitute.)

Three 9-volt batteries in series

COIL:
50 turns
#26 insulated wire (22 – 28 close enough)
4.5" diameter (ABS pipe coupler
Tapped at 5, 10, 20, 30, 40, & 50 turns.

Simple One-tube AM Radio

Al Klase - N3FRQ – 25 SEP 2012

The Radio Technology Museum
http://www.rtm.ar88.net/

1-Tube Regen Based on Above

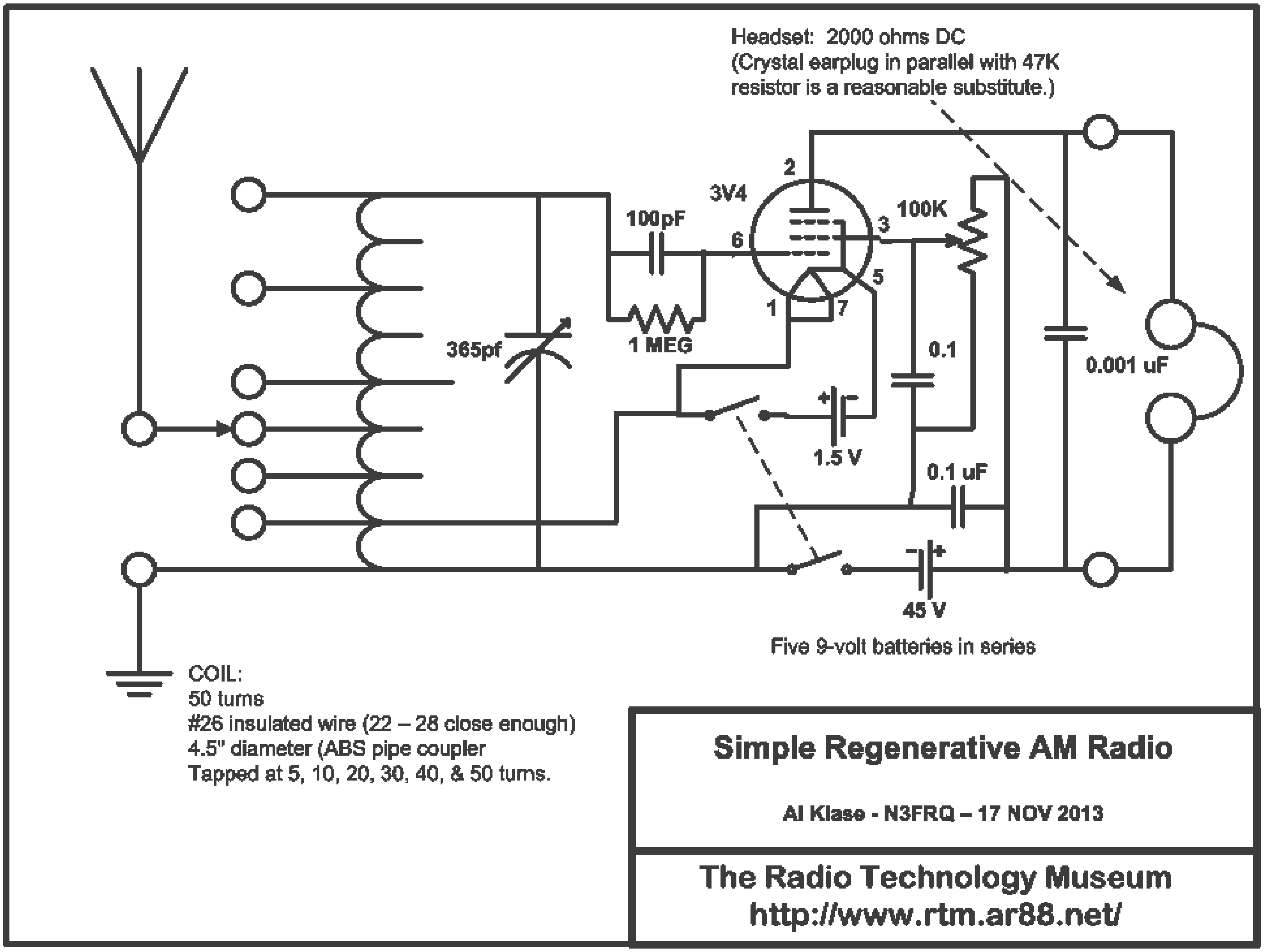
3V4
2
6
3
5
1
7
100pF
1 MEG
365pf
100K
0.1
1.5 V
0.1 uF
45 V
0.001 uF

Headset: 2000 ohms DC
(Crystal earplug in parallel with 47K
resistor is a reasonable substitute.)

Five 9-volt batteries in series

COIL:
50 turns
#26 insulated wire (22 – 28 close enough)
4.5" diameter (ABS pipe coupler
Tapped at 5, 10, 20, 30, 40, & 50 turns.

Simple Regenerative AM Radio

Al Klase - N3FRQ – 17 NOV 2013

The Radio Technology Museum
http://www.rtm.ar88.net/

THE SKYWAVES HIGH-PERFORMANCE CRYSTAL SET

Copyright 1999 - Alan R. Klase - All rights reserved

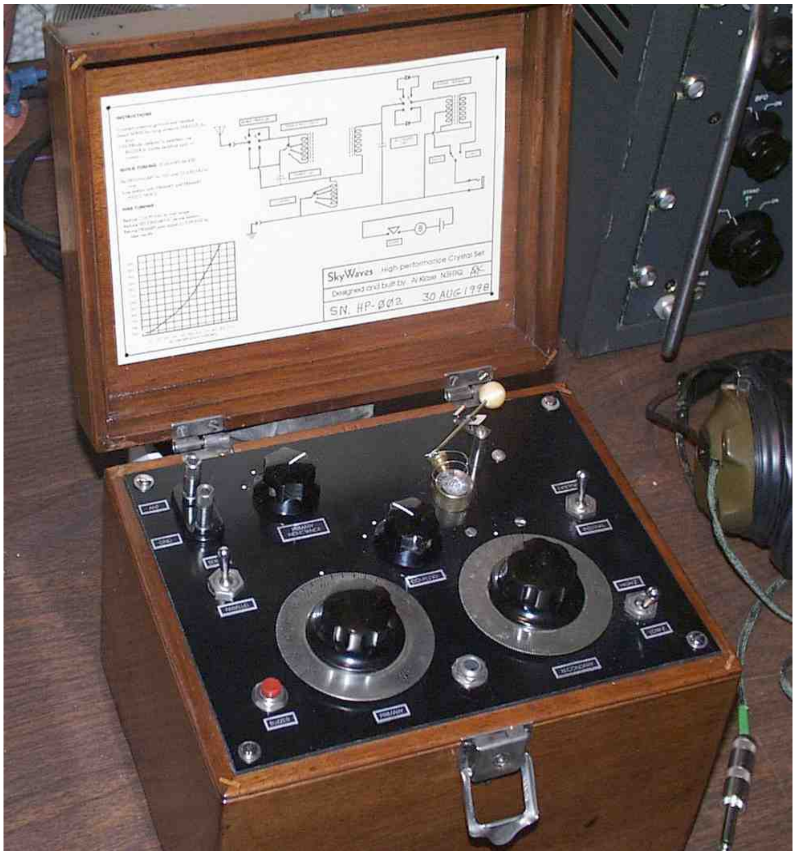

This radio is the result of approximately ten years of sporadic research and experimentation with crystal sets from an engineering perspective. Serious professional development of passive receivers virtually ended with the introduction of reliable vacuum tubes. My conclusions are really pretty straight forward: One needs to pay attention to proper impedance matching at each point in the circuit to assure maximum signal transfer. Multiple tuned circuits with variable coupling allow the operator to trade sensitivity for selectivity. Lastly, a high performance headset can easily be fifty times (17 dB) more sensitive than the instruments usually employed.

THE CIRCUIT

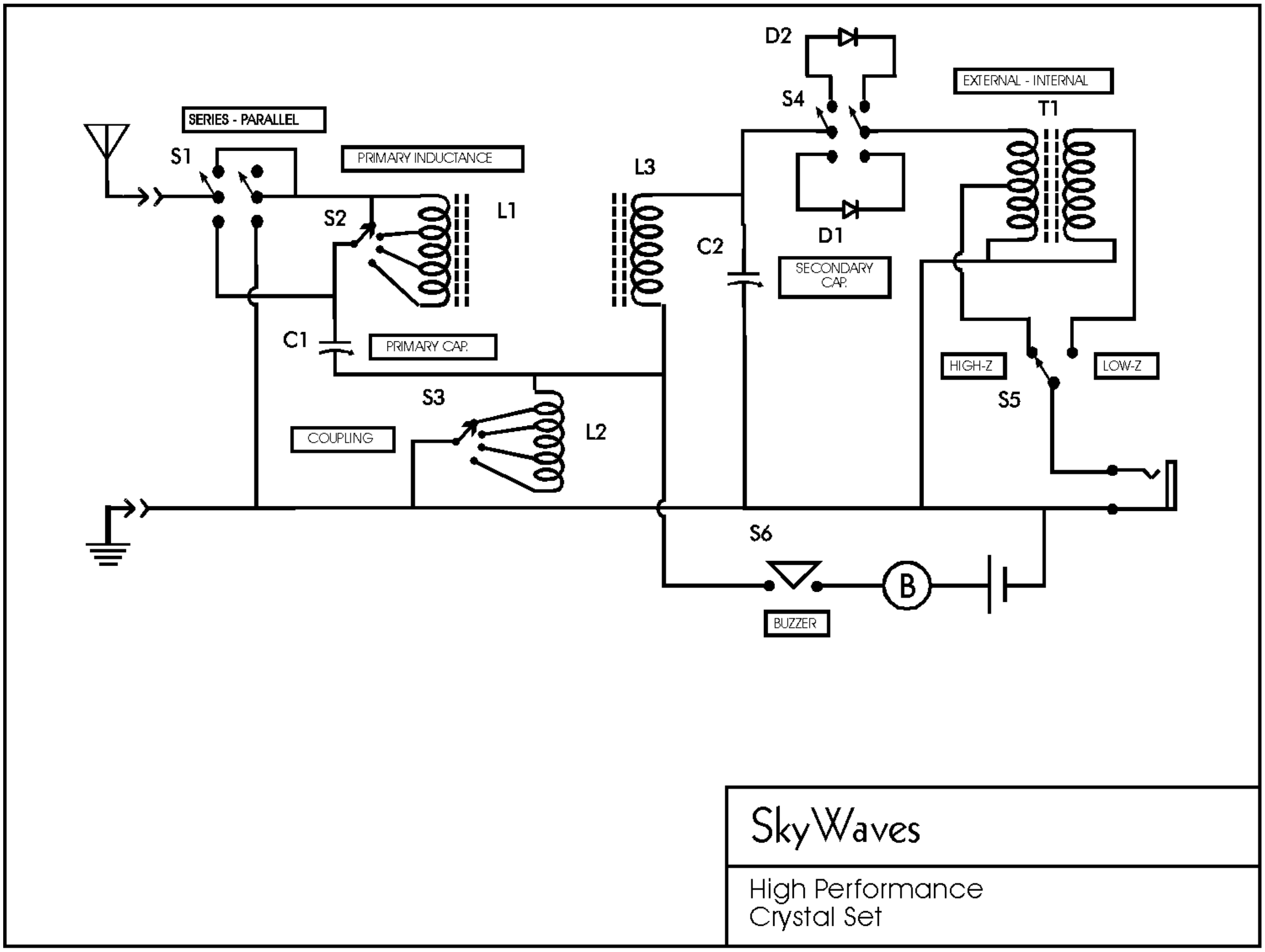

Basically, the “hookup” is a classic two-circuit tuner with variable coupling between the primary (antenna) and secondary (detector) tuned circuits. This is essentially the same architecture used to great effect in the communication receivers of the wireless era. This circuit was almost never used in commercial broadcast crystal sets because of its high parts count resulting in high price.

The primary or antenna circuit consists of inductor L1 and variable capacitor C1. Switch S1 allows the primary circuit to be used either in series or parallel with the antenna and ground connection. The series connection in normally employed, while the parallel arrangement is helpful with shorter antennas. L1 is tapped via switch S2 to provide a range of inductance values.

The secondary or detector circuit comprises inductor L3 and variable capacitor C2. Tapped inductor L2 is in series with the ground returns of both the primary and secondary circuits, and provides a variable mutual inductance which controls the coupling between the two circuits. This feature allows the operator to trade off signal transfer for increased selectivity. The effect is fundamentally the same a achieved with the telescoping coils of the classic “loose coupler”.

The detector is connected to the high-impedance point at the junction of L3 and C2. This applies the maximum available voltage to the detector to help overcome it’s forward junction drop. Although this is the “usual” connection in many sets, it is often counter-productive as the relatively low impedance (10K ohms) of the ubiquitous 2000 ohm headset unduly loads the tuned circuit compromising both signal strength and selectivity. In this set, the detector load is the 100K ohm primary of audio transformer T1. Switch S4 selects either the “external” crystal detector or the “internal” germanium diode.

T1’s low impedance secondary (600 ohms) provides optimal coupling to the extremely sensitive headset included with this radio. Switch S5 may be used to select a medium-high impedance output for use with more conventional headsets.

In a double tuned receiver, it can be difficult to find a sensitive spot on the crystal when proper tuning of the primary and secondary has yet to be achieved. The buzzer circuit provides a broadband RF signal to allow adjustment of the detector in the absence of a tuned station.

THE COMPONENTS

A conscious effort was made to use commercial-quality US-made vintage components wherever possible. The wafer switches are ceramic units with silver plated contacts, the variable capacitors are high-end Hammarlunds, the tuning knobs and binding posts are from James Millen, the toggle switches are Cutler-Hammer, the phone jack is from Switchcraft and the audio transformer is a UTC A-12. Even the buzzer is a Western Electric “KS” spec. part. One notable exception is the coils, these are wound on modern ferrite cores to provide optimum performance in a reasonable space.

The headset was assembled from a vintage headband and sound-powered telephone elements from a WWII TS-10 handset. It is as sensitive as any I’ve encountered.

CONSTRUCTION

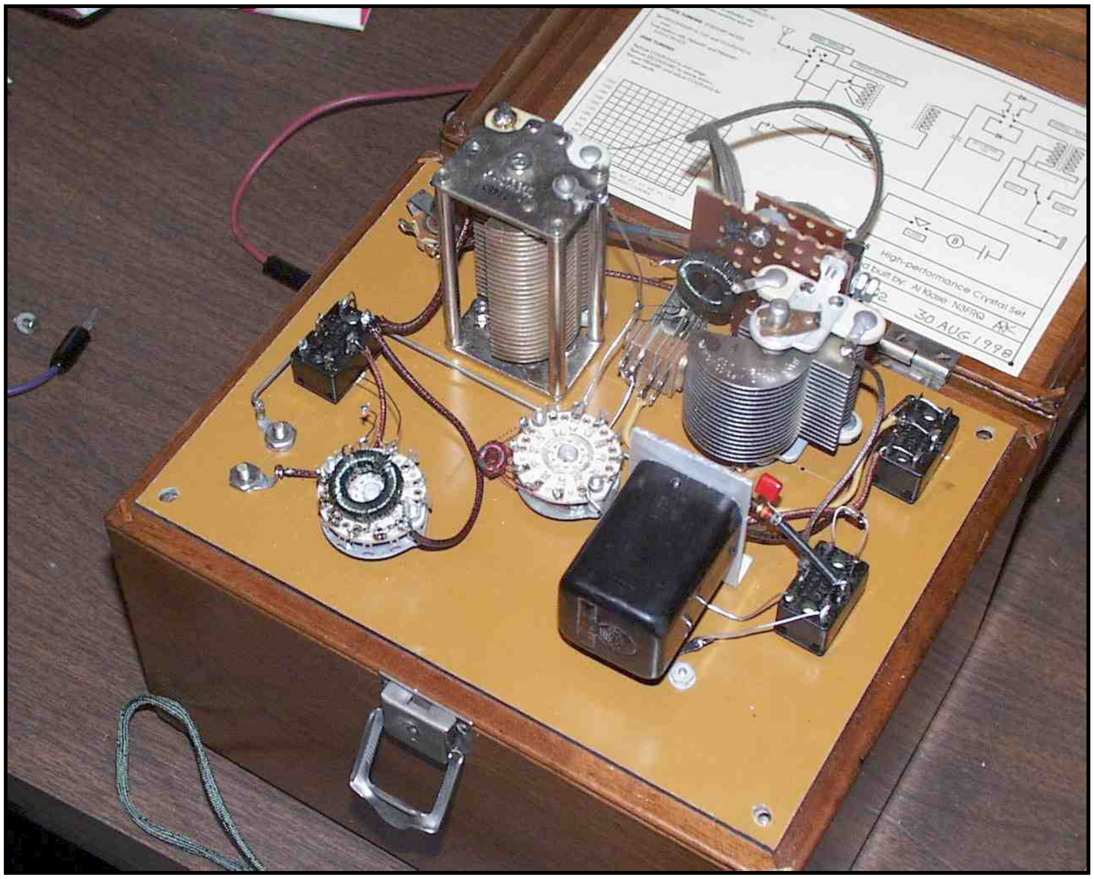

The radio is assembled in an walnut box with a leather carrying handle that once housed a Leeds and Northrup Wheatstone bridge. The bulk of the components are mounted to the Bakelite front panel. Construction technique is best described as 1950’s “model shop”, and is much nicer than an average homebrew set or engineering-lab prototype.

OPERATION

Connect antenna, ground and headset. Select SERIES for long antennas PARALLEL for short.
If the EXTERNAL detector is selected, activate the BUZZER while manipulating the catswisker to locate a sensitive spot on the crystal.

To easily locate strong signals, set the SECONDARY capacitor to 100 and COUPLING to maximum (fully clockwise). Tune the set with the PRIMARY capacitor and PRIMARY INDUCTANCE. With this setup, the inductance L3 acts as an untuned secondary to L1. This was known as the “standby” mode in the classic marine wireless receivers.

To take full advantage of the available sensitivity and selectivity, reduce the COUPLING control to midrange and reduce the SECONDARY capacitor to retune the station. Then, adjust the PRIMARY and COUPLING controls for best results. Tuning a receiver of this sort across the band is definitely a two-hand affair. It’s sometimes helpful to rock one knob back and forth while slowly tuning the other.

The tuning of the secondary circuit is fairly repeatable if the coupling is kept light. The tuning chart inside the lid can be used to set the secondary to a desired frequency. It’s calibration is based on position 2 (from the left) of the coupling switch.

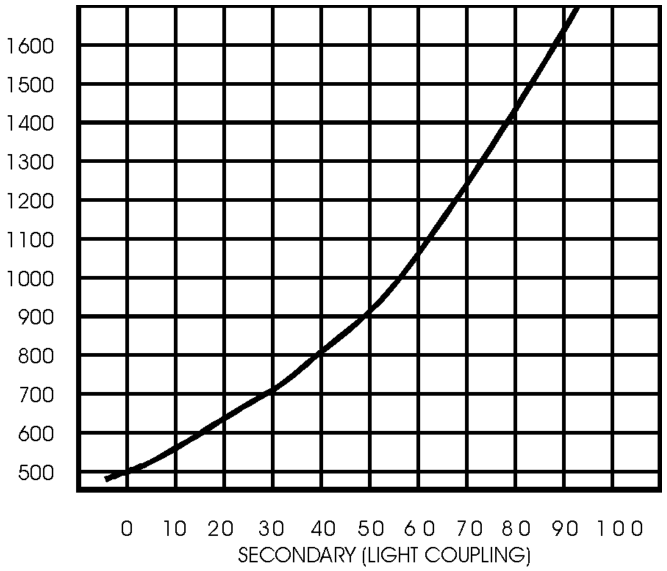

COIL DETAILS

The coils are wound on ferrite toroid cores from Amidon associates. L1 is 75 turns of #26 enameled wire on an FT-82-61 core tapped every 15 turns. Inductance is approximately 400 uH maximum. L2 is 25 turns of #30 on an FT-37-61 core tapped at 3, 10, 15, and 20 turns. Inductance is approximately 35 uH maximum. L3 is 68 turns of #26 on an FT-82-61 core. Inductance is approximately 290 uH.

HOME

Stay Tuned

Crystal Radios & Tube Radios

Crystal Radio Trouble Shooting Page

Here are a few tips to check on. These are common mistakes or problems.

We'll start with the most common problems and work down.

Antenna-

1) Is it long enough? It should be at least 50 feet or so. If it's only 40 feet, don't worry try it anyway. It will probably be OK.

2) Is it high enough? Should be up as high as you can get, but be safe. Don't get near electrical lines etc. Min. would be 8 or 10 feet and still work. I've even seen them thrown out the window and just hanging there and work. It will depend on how close and strong your radio stations are .

3) Make sure your antenna is not "grounded out". That is, no part of the antenna (bare wire) is touching anything like a fence or tree. This will "ground out" the antenna.

Ground-

The other part of the antenna system is the ground. This can be a water pipe or a rod pounded in the ground. If you use a water pipe, make sure your water pipes are not PVC which is a plastic. If you pound a rod in the ground, it will need to be at least 4 or 5 feet deep. Another ground is the screw in the middle of a plug in the wall outlet, but the outlet must be a grounded type to work.

Insulation-

1) Did you strip the insulation off the ends of the wires before connecting them?

2) Did you remove the insulation off the wires at the "Taps". The Taps are where the alligator clip connect on to
the coil?

I know this sounds too simple to be true, but some of us out there know nothing about electrical properties so…….we will check.

Connections-

1) Are all your connections sound? Metal to metal!!!

2) Are they correct? Check them 3 or 4 times at least. This is the most common error and sometimes the hardest to find!

Coil Windings-

1) Did you wind the right amount of winding on the coil?

2) Are they fairly tight to one another. You can't "scatter wind" the coils. One wind neatly along the side of the
next.

Now for the real problems. The rest of the components. That would be the 1) earphone, 2) the diode, 3) and the resistor.

The "Earphone"-

Here the most common mistake is to take an earphone from Radio Shack or a set of headphones or "ear buds" from a regular radio and try to use them. Without getting into a big technical explanation, you can't use them. It must be a "crystal" earphone. Some times called a "ceramic" earphone. They are much more sensitive than stereo earphones or ear buds. You can test your earphones buy removing them from the radio and putting the earphone in your ear. Next take the two wire ends of the cable coming out of the earphone and tap them together a few times. You should hear a "clicking" sound. This is how sensitive they are. DO NOT USE A BATTERY TO TEST THEM. This will mess them up. I have seen lots of earphones brand new that did not work.

Resistor-

The most common mistake here is that you used too small of a resistor here. In other words the values of the ristor is less than 47,000 ohms (47k). The resistor bands should be yellow, purple, orange with a gold or silver band last. Using a resistor of lesser value will "short out" the radio and it will not function. You can also leave out the resistor and it will work, just not as good.

Diode- (most look like they are made from glass, with one black band)

This is almost impossible for a beginner to check without a meter of some kind. If you have a meter that reads "continuity", then put the diode on it and it should read in one direction, but if the diode is turned around it will not read. If you don't have a meter all you can do is try and replace it.

Second most common mistake with the diode is… It has to be a germanium diode to work. A silicon diode used in power configurations will not work (well, not very well).

This radio has been made by hundreds of people and it has worked. Sometimes after a mistake is found, but it will work.

I hate to say this, but maybe you don't have a radio station close buy. If you are in the middle of farm country and no station are near you (more than 100 miles) you might not be able to pick up any thing with this set. 99.999% of the time this is not the case.

If you are still having problems, email me. Maybe I have a part or so that you can have and try……or ask questions at "Tap N Tap" at http://www.midnightscience.com/rapntap/

They have a beginner's forum that is great!!

Return to "Stay Tuned " Home Page

© 1999-2010 Darryl Boyd, All rights reserved

Copyright Notice:

My website is copyrighted. This includes all images, text, drawings.
If you are thinking of downloading my copyrighted items and selling them on ebay (or anywhere) keep this in mind,
I monitor ebay for such violations. I will pursue the issue with ebay and in the courts if necessary.
I do not furnish them for your gain from my hard work.

TOPICCRAZE
VOTE FOR MY SITE!

Search This Site

Google [ ] Google Search
○ WWW ● http://www.crystalradio.net

Send Email To:

darrylboyd@crystalradio.net

Because of antispam techniques, you can not "cut and paste" the above text

---

Return to "Stay Tuned " Home Page

We have made every effort to ensure that the information provided on this website is accurate and up to date.
All information on this website is informational only and no guarantees to the accuracy of any of the projects or circuits on this site or calculators. If you find something you feel is inaccurate, please inform us by e-mail. Be sure to provide a thorough argument to support your case. Confirmed changes will be made as quickly as possible.

The images on this website are protected by copyright. They are the property of the owner of this website and may not be used without the owners permission. Please don't use the images made by us on this website without permission.

Readers' Projects - Sophie's radios

Crystal receiver for FM

A project by **Giacomo Cavuoti**

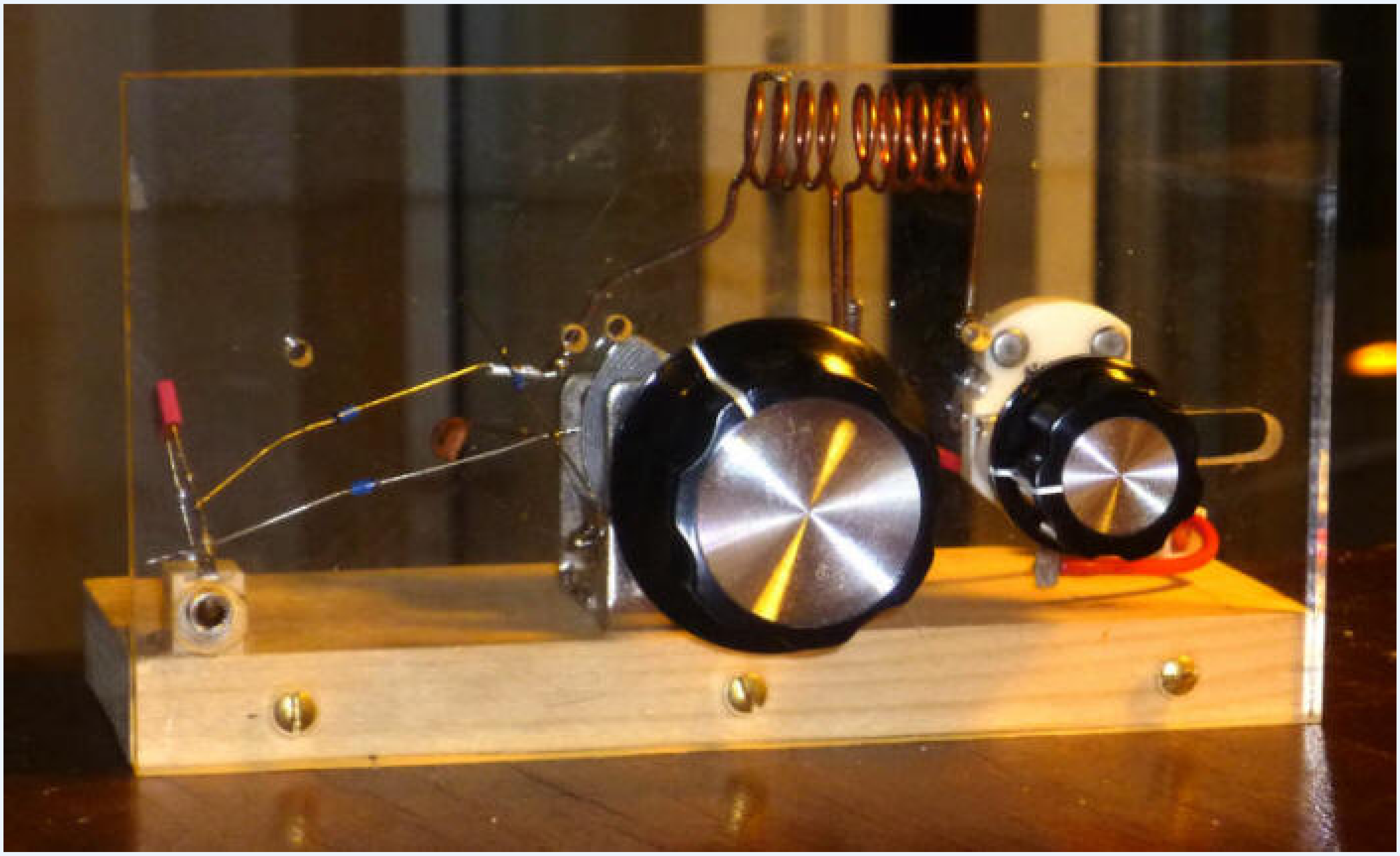

Following Leonardo's invitation I made this simple crystal receiver for FM (click on the diagram below to see it enlarged).

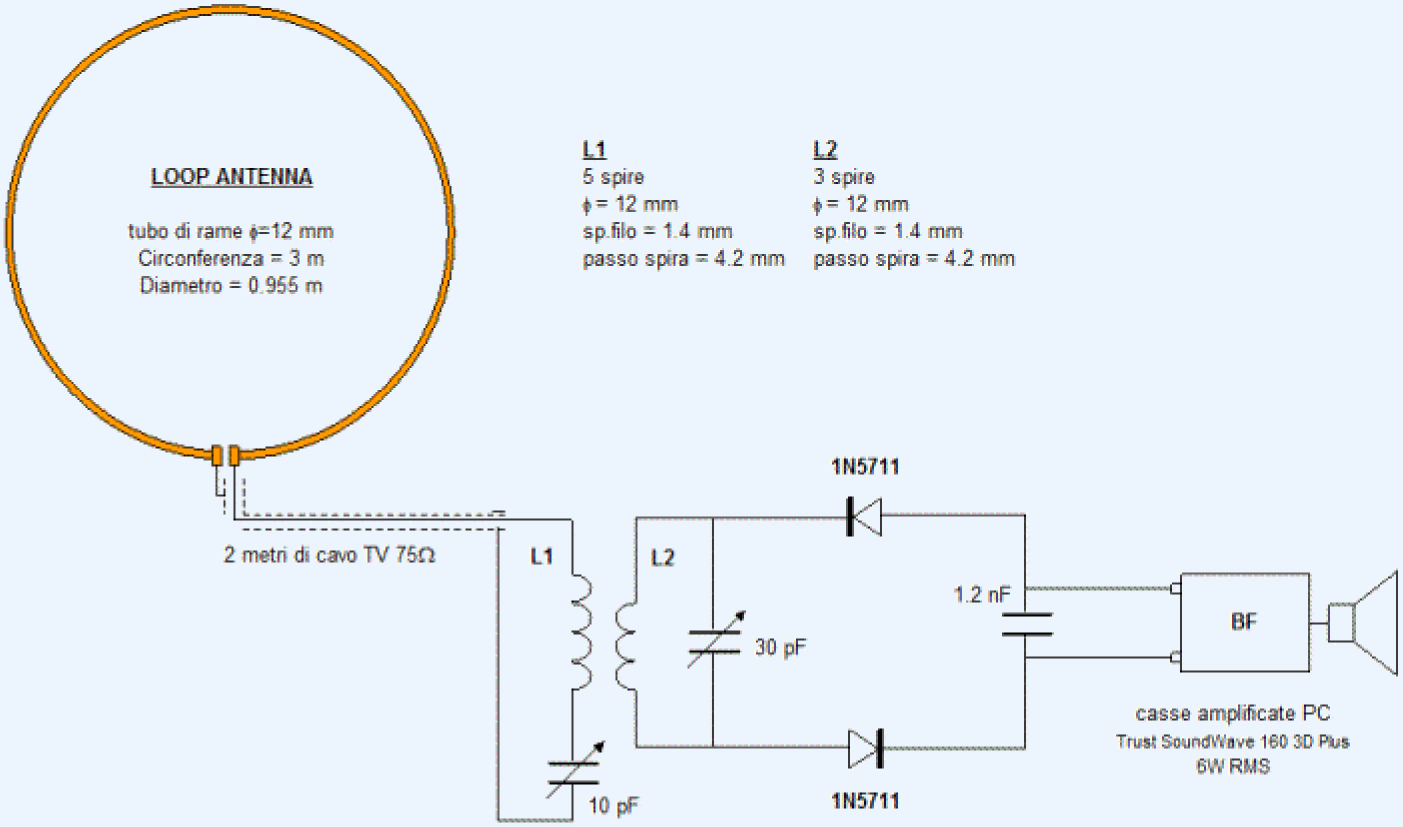
LOOP ANTENNA
tubo di rame ϕ=12 mm
Circonferenza = 3 m
Diametro = 0.955 m
L1
5 spire
ϕ = 12 mm
sp.filo = 1.4 mm
passo spira = 4.2 mm
L2
3 spire
ϕ = 12 mm
sp.filo = 1.4 mm
passo spira = 4.2 mm
1N5711
2 metri di cavo TV 75Ω
L1
L2
1.2 nF
BF
30 pF
casse amplificate PC
Trust SoundWave 160 3D Plus
6W RMS
10 pF
1N5711

The receiver is "powered" by a full wave loop antenna consisting of a single circular coil made of 3 meters of copper pipe for hydraulic systems (measured inductance of about 3 uH). The support is a simple 2-meter long fir-wood strip, 3 cm wide and 1 cm thick, to which the copper rim has been fixed with insulating tape. Another solution to realize the loop antenna in a simple and instantaneous way is to use an aluminum strip 2 mm thick and 2 cm wide. The detectors are two *Schottky* diodes type 1N5711, particularly suitable for VHF thanks to the low capacity (about 2pF). They are currently in production and therefore easy to find. The receiver was tested in a "stiff" situation with the antenna exposed in a recessed balcony (3 x 1.6 meters) on the fourth floor of a 6-storey reinforced concrete building surrounded by other buildings constructed of reinforced concrete walls. As you can see and hear from the two videos I uploaded on Youtube:

http://www.youtube.com/watch?v=YNiuRoIp36E

http://www.youtube.com/watch?v=SJwx8_3mCfI

the reception is good as well as the selectivity, thanks to the appropriate coupling, but very easy to find, between the antenna coil L1 and that of L2 tuning. The audio amplification has been entrusted to two small speakers amplified by PC. The following photos show the first working prototype mounted "in the air" (click on the images to see them better). The photo below the title instead shows an aesthetically more accurate realization.

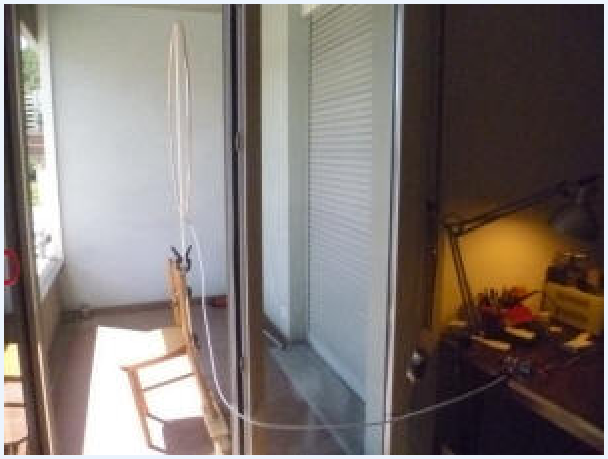

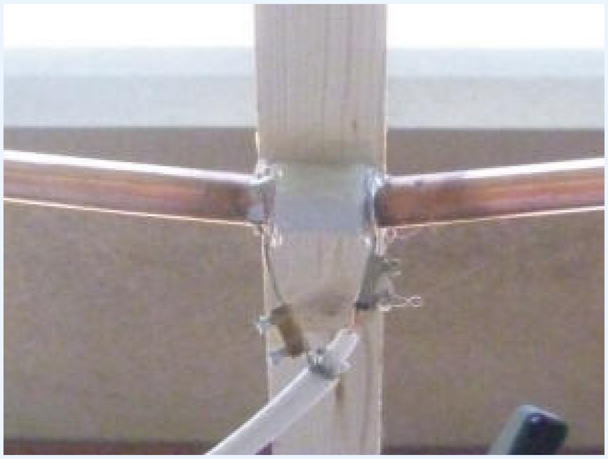

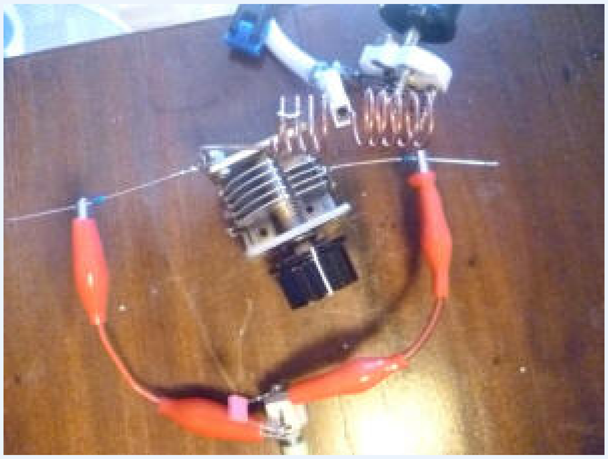

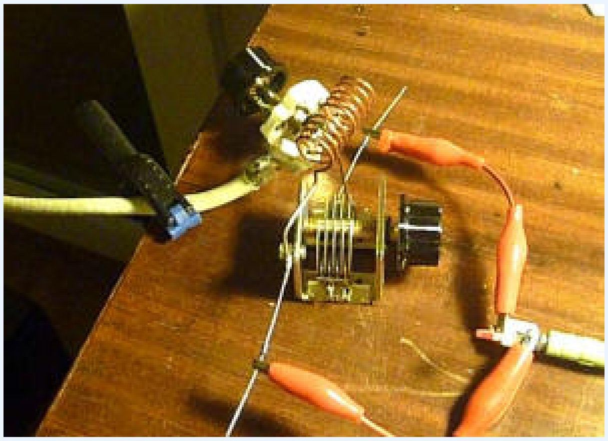

For any clarification, do not hesitate to contact me. Giacomo Cavuoti

Reader projects - Crystal radio
Return to the Main Page

Simple crystal receiver for FM

Carlo Bramanti

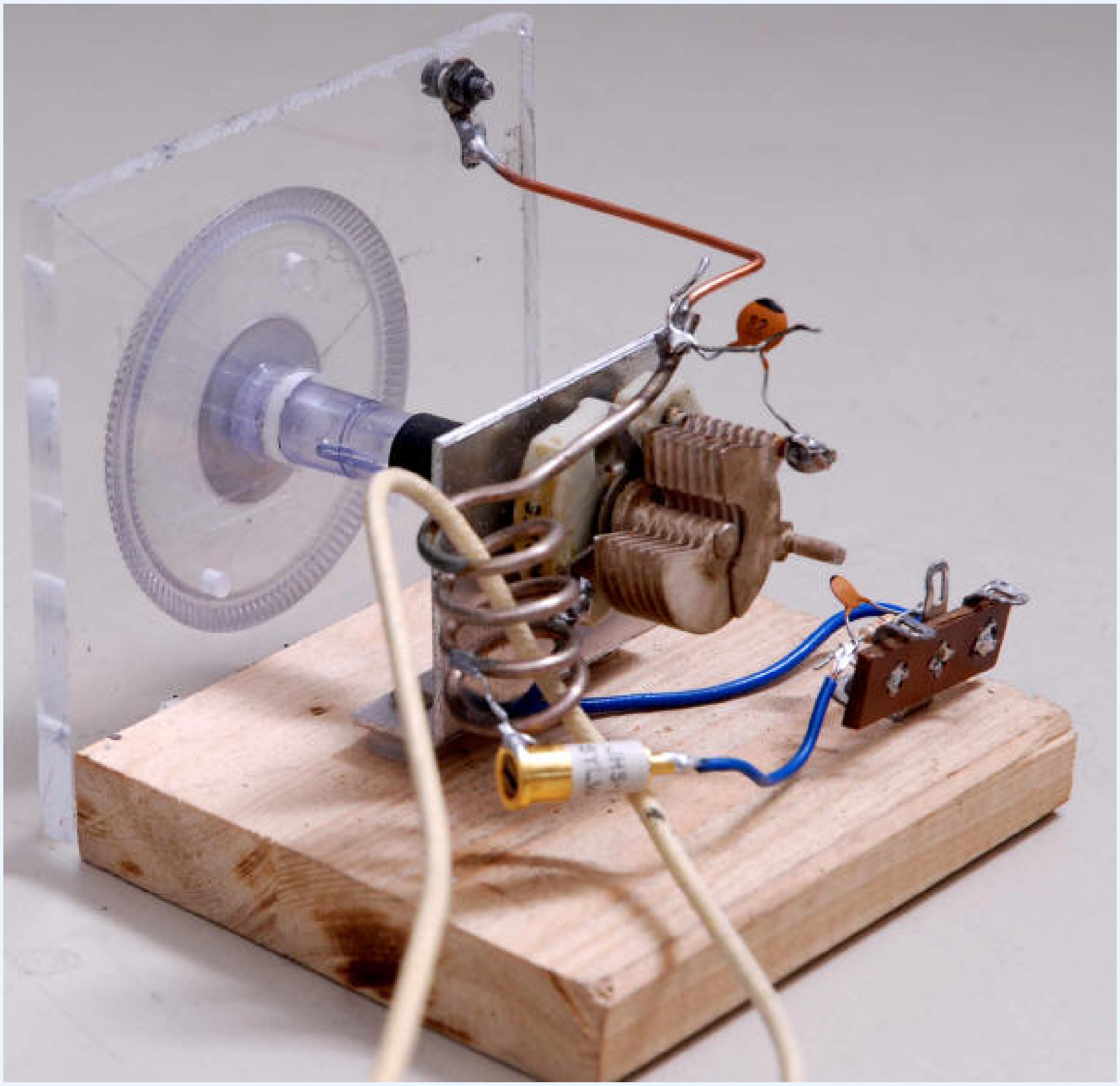

After the success obtained in the realization of the "galena" FM with the resonant cavity, I had promised myself to try also with a simple LC resonant circuit. One of my galena mates favored what he had built without success and I set about making it work.

Operation

With the experience that the audio signal resulting from this detection is very low, I immediately applied the two-transistor amplifier I have already described on these pages. With the fully extended stylus it did not actually work, but by lowering it, you already felt something.

The link

When I connected it to the J antenna it worked fine. The connection that gave me the best results was a simple conductor loop over eighty cm (about ¼ wave), passed through the resonant coil. If we use a cable to move the antenna away, the loop connects directly to the cable. I have no particular experience on these frequencies and I think that those who have more can still improve performance. In any case they are already adequate, even if slightly inferior to that of the cavity, with the advantage that everything is much more comfortable and pleasant.

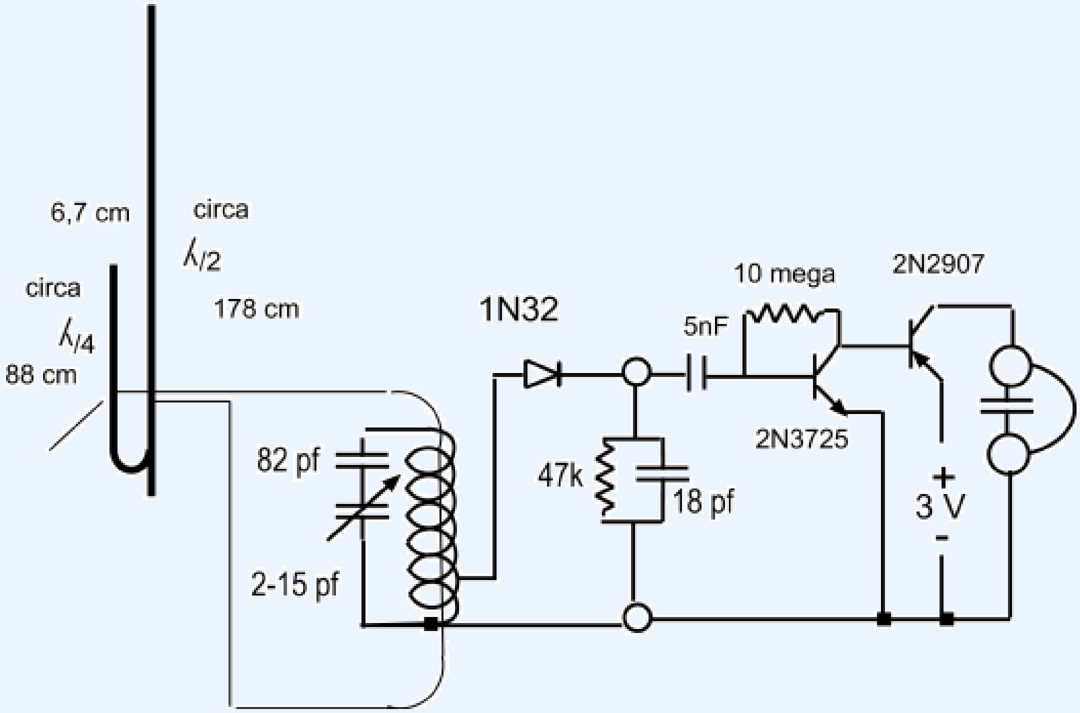

The theory

Normal AM revelation can not reveal weak signals with the consequence that it is useless to amplify after the revelation, since it increases only the volume and not the sensitivity. Instead the detection of the FM also handles very weak signals: therefore, amplifying after the detection, all the weakest stations appear. The amp that I made amplifies a lot (in AM it also acts as a detector, but in FM it must be preceded by a diode); it can be powered by a 3 V lithium drain and it consumes very little: it is even sufficient to disconnect the earphone for almost zero consumption. This makes the use of a switch superfluous.

J antenna

It is an antenna that is easy to realize by bending a copper pipe by fontaniers, with a diameter of 6 o 8 mm. Fixing to a pole or base does not require isolation.

The diode 1N32

I obtained an additional advantage by replacing the germanium diode OA85 with the one with contact tip silicon 1N32, a little old but available and used as a mixer for the K band.

NOTE

In the aforesaid realization the frequencies from 88 to 104 Mc are listened with a rotation of the variable of only 45 degrees, or half of the excursion. Above I hear nothing.

The excursion with trimmers and padders should therefore be extended, while maintaining the optimal L / C ratio.

Realization:

Coil in silver wire from 2 mm.

Average diameter 14 mm

Length 19.5 mm

Number of turns 4

1 ½ coils outlet from the municipality

Ground connection not necessary

Proposal

I would have made the coil with three turns of three millimeters, diameter three cm, length three cm. The Q would be one and a half times higher, which on balance, would improve performance slightly; except try, in this discipline you never know!

See other Bramanti articles on FM Demodulation and Cavity Detector

See also the Bramanti on Galene page

Crystal radio
Return to the Main Page

Crystal Radio for Short Waves

A project by Leonardo Mureddu

For those who enjoy listening to distant and mysterious stations, the bands of the Court Waves undoubtedly offer an endless series of emotions. In fact, the propagation phenomena favor precisely these frequency bands, between 5 and 20MHz (with a wavelength of between 20 and 60 meters), transporting them at intercontinental distances thanks to a series of reflections that occur between the ground and the upper part of the atmosphere (ionosphere). It may therefore happen to receive distinctly in Italy, perhaps only for a few minutes, transmissions from Australia, China or America. Short wave broadcasts were not immediately introduced in the history of radio, indeed for all the '20s the big companies continued to bet on long waves for intercontinental transmissions, and on medium waves for local broadcast broadcasts. It was only in the 1930s, with the development of large planetary-scale communication networks that short waves were chosen as privileged vehicle for transmissions between the motherland and the colonies or the army at war, and for communications between allied nations. The best radio receivers were therefore equipped with numerous wave ranges capable of tuning and receiving broadcasts from all over the world. Nowadays the frequencies of the short waves, although now almost completely forgotten and often absent in the domestic tuners, are still very popular and full of programs in all languages.

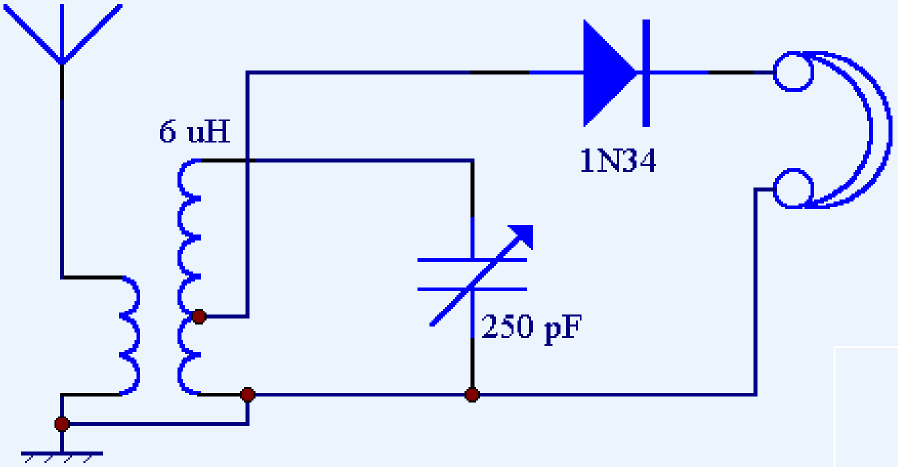

The construction of a crystal receiver for shortwaves is a fairly easy undertaking and can give great satisfaction, especially to those who delight in the so-called "DX", ie listening to broadcasts from great distances.

Scheme of principle

The project I present here is very simple and has no great pretensions, but it can serve as a starting point for experimenting with more complicated and efficient devices. The technique used is somewhat different from that used for medium waves, because higher frequencies require some constructive measures to minimize losses. The scheme is very similar to that of any medium wave receiver. What changes is the inductance of the coil and the construction technique. The receiver consists of an input transformer with a tuned secondary and an intermediate socket for the detector.

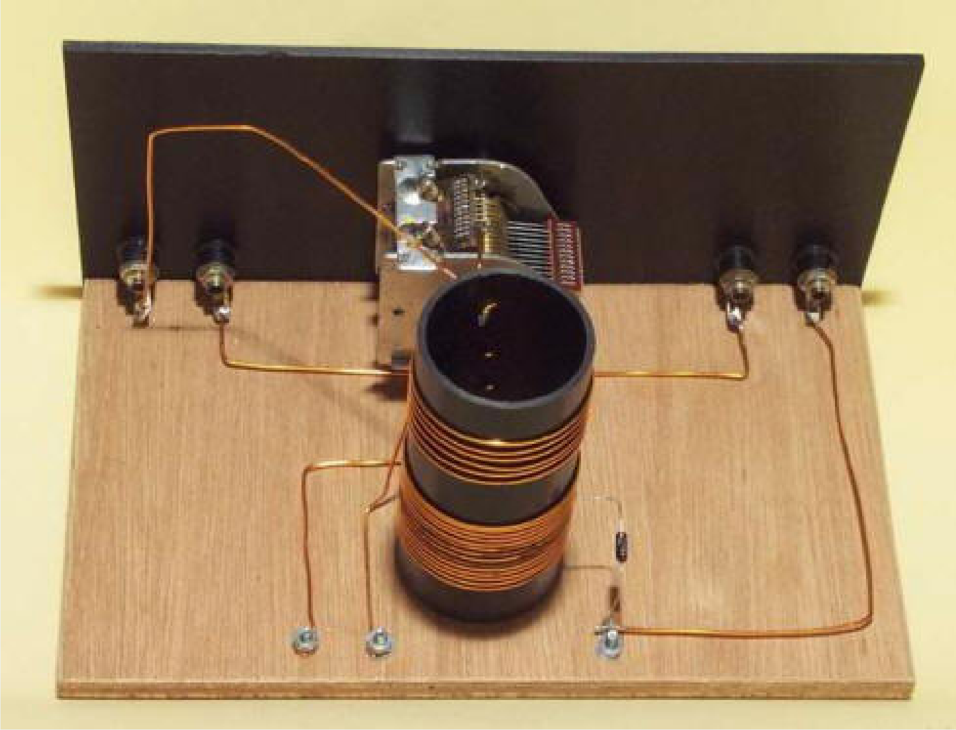

Constructive notes

To obtain satisfactory results, a few rules must be followed, among which the most important is to make a compact assembly using coarse rigid wire for the connections between the various components. In the prototype shown above, 1mm diameter enamelled copper wire was used to make both the coils and the connections. Enamelled copper is preferable to bare or tinned copper, *due to the skin effect*. In fact, the enamel preserves from the oxide the surface of the wire on which the conduction takes place. For the same reason, the use of PVC insulated hard wire is not recommended. It is advisable to mount it on an insulating base, in order to reduce losses due to the proximity effect. The use of a variable air-cooled condenser is strongly recommended, but acceptable results can also be obtained with a miniature capacitor with a polyester dielectric (for transistor radios). The old capacitors with solid dielectric (bakelite or other) must be avoided, as they have too many losses at high frequencies. The tuning coil is made of 14 turns of 1mm wire on a good 4cm diameter support; the turns are spaced so as to have a total length of about 3cm. The detector plug is made at the fifth turn. The antenna coil, wound on the same support, is made up of 6 turns of the same wire, even these slightly spaced. It will be good to equip the variable capacitor with a rather large knob, since in short waves the tuning is always quite critical. The detector plug is made at the fifth turn. The antenna coil, wound on the same support, is made up of 6 turns of the same wire, even these slightly spaced. It will be good to equip the variable capacitor with a rather large knob, since in short waves the tuning is always quite critical. The detector plug is made at the fifth turn. The antenna coil, wound on the same support, is made up of 6 turns of the same wire, even these slightly spaced. It will be good to equip the variable capacitor with a rather large knob, since in short waves the tuning is always quite critical.

Testing

After connecting the antenna (the same used for the medium waves) and the ground, slowly rotate the tuning knob: it should be possible to clearly capture some stations in various languages. Selectivity should be sufficient to separate near-frequency transmissions. You will notice that any received station is subject to a progressive fading, until it disappears altogether and then reappears or replaced by another station. This phenomenon is called "evanescence" (in English "fading"), is typical of short waves and is due to the propagation of ionospheric waves, which are subject to interference phenomena.

I invite the fans to propose changes and improvements, which I will gladly welcome in these pages.

LM

Return to the Main Page

Radio a Cristallo per Onde Corte

Un progetto di Leonardo Mureddu

Per coloro che amano l'ascolto di stazioni lontane e misteriose, le bande delle Onde Corte offrono senza dubbio una serie infinita di emozioni. Infatti i fenomeni della propagazione privilegiano proprio queste bande di frequenza, comprese tra i 5 e i 20MHz (con lunghezza d'onda compresa tra i 20 e i 60 metri), trasportandole a distanze intercontinentali grazie ad una serie di riflessioni che avvengono tra il suolo e la parte alta dell'atmosfera (ionosfera). Può quindi capitare di ricevere distintamente in Italia, magari solo per pochi minuti, trasmissioni provenienti dall'Australia, dalla Cina o dall'America. Le trasmissioni ad onde corte non vennero introdotte subito nella storia della radio, anzi per tutti gli anni '20 le grandi compagnie continuarono a scommettere sulle onde lunghe per le trasmissioni intercontinentali, e sulle onde medie per le trasmissioni a diffusione locale. Fu solo negli anni '30, con lo sviluppo delle grandi reti di comunicazione a scala planetaria che le onde corte furono scelte come veicolo privilegiato per le trasmissioni tra la madrepatria e le colonie o l'esercito in guerra, e per le comunicazioni tra nazioni alleate. I migliori ricevitori radio erano quindi dotati di numerose gamme d'onda in grado di sintonizzare e ricevere trasmissioni da tutto il mondo. Al giorno d'oggi le frequenze delle onde corte, sebbene ormai quasi del tutto dimenticate e spesso assenti nei sintonizzatori domestici, sono ancora molto frequentate e ricche di programmi in tutte le lingue.

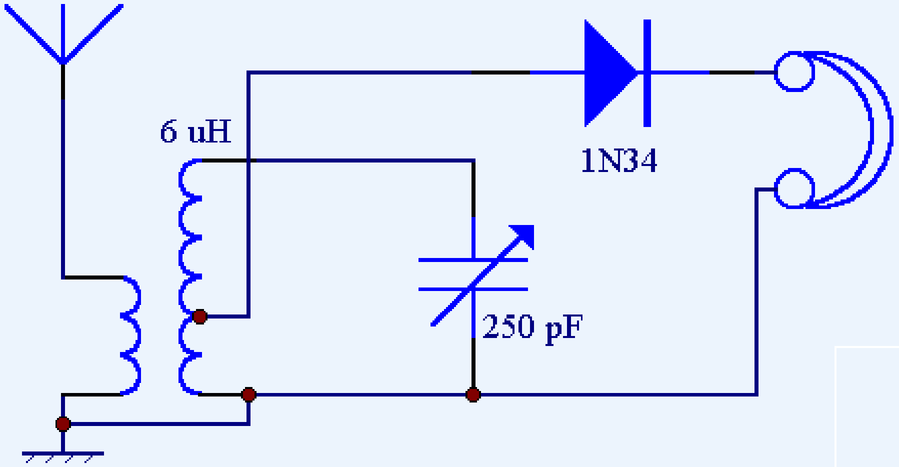

La costruzione di un ricevitore a cristallo per le onde corte è un'impresa tutto sommato abbastanza facile e può dare grandi soddisfazioni, specie a chi si diletta nei cosiddetti "DX", ossia l'ascolto di trasmissioni provenienti da grandi distanze.

Schema di principio

Il progetto che presento qui è molto semplice e non ha grandi pretese, ma può servire come base di partenza per la sperimentazione di apparecchi più complicati ed efficienti. La tecnica usata è un po' diversa da quella usata per le onde medie, perché le frequenze più elevate richiedono alcuni accorgimenti costruttivi per minimizzare le perdite. Lo schema è molto simile a quello di qualunque ricevitore per onde medie. Ciò che cambia è l'induttanza della bobina e la tecnica costruttiva. Il ricevitore è costituito da un trasformatore d'ingresso con secondario accordato e presa intermedia per il rivelatore. Con i valori indicati (6 microhenry per l'induttanza e 250 pF per il condensatore variabile) si copre quasi tutta la gamma delle onde corte, da 5 a 15 MHz.

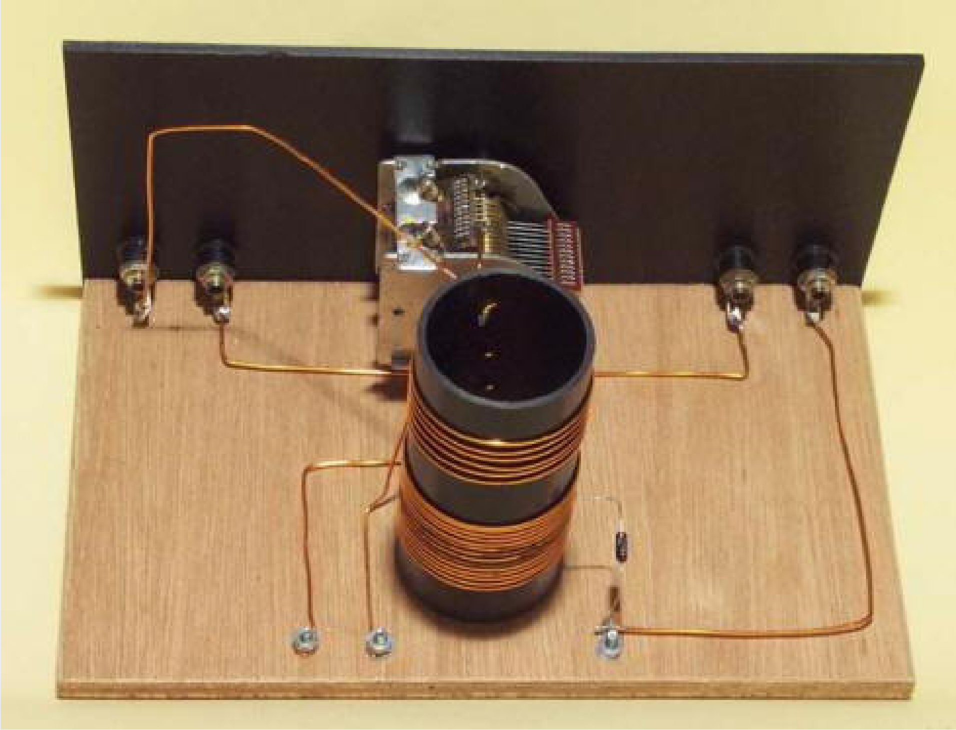

Note costruttive

Per ottenere risultati soddisfacenti occorre attenersi a poche regole, tra cui la più importante è quella di realizzare un montaggio compatto utilizzando filo rigido grosso per i collegamenti tra i vari componenti. Nel prototipo mostrato qua sopra è stato usato del filo di rame smaltato da 1mm di diametro per realizzare sia le bobine che i collegamenti. Il rame smaltato è da preferire al rame nudo o a quello stagnato, per via dell'*effetto pelle*. Infatti lo smalto preserva dall'ossido la superficie del filo su cui avviene la conduzione. Per lo stesso motivo non è consigliabile l'uso di filo rigido isolato in PVC. E' bene realizzare il montaggio su una basetta isolante, in modo da ridurre le perdite per effetto di prossimità. L'uso di un condensatore variabile ad aria è fortemente consigliabile, ma si possono ottenere risultati accettabili anche con un condensatore miniatura con dielettrico in poliestere (per radio a transistor). Sono da evitare assolutamente i vecchi condensatori con dielettrico solido (bakelite o altro), perché presentano troppe perdite alle alte frequenze. La bobina di sintonia è realizzata con 14 spire di filo da 1mm su un buon supporto da 4cm di diametro; le spire sono distanziate in modo da avere una lunghezza totale di circa 3cm. La presa per il rivelatore è realizzata alla quinta spira. La bobina d'antenna, avvolta sullo stesso supporto, è costituita da 6 spire dello stesso filo, anche queste leggermente distanziate. Sarà bene dotare il condensatore variabile di una manopola piuttosto grossa, dal momento che in onde corte la sintonia è sempre alquanto critica.

Collaudo

Collegata l'antenna (la stessa usata per le onde medie) e la terra, ruotate lentamente la manopola di sintonia: dovrebbe essere possibile captare con chiarezza alcune stazioni in varie lingue. La selettività dovrebbe essere sufficiente a separare trasmissioni vicine in frequenza. Noterete che qualunque stazione ricevuta è soggetta ad un affievolimento progressivo, fino a scomparire del tutto per poi riapparire o venire sostituita da un'altra stazione. Questo fenomeno si chiama "evanescenza" (in inglese "fading"), è tipico delle onde corte ed è dovuto alla propagazione delle onde ionosferiche, che sono soggette a fenomeni di interferenza.

Invito gli appassionati a proporre modifiche e miglioramenti, che accoglierò molto volentieri in queste pagine.

L.M.

Torna alla Pagina Principale

High Sensitivity Crystal Set

Build a "crystal radio" that does not require an outside antenna or ground by using a new zero-voltage-threshold MOSFET.

Bob Culter, N7FKI

If you are like me, you may have been bitten by the "radio bug" while building an AM broadcast band crystal set. In my case, I was using a galena "cat's whisker" detector.[1] Such a radio typically required a long outside antenna and ground rod or water pipe ground connection to function, even with the use of a low-threshold (0.3 V) germanium detector diode such as the well-known 1N34A.

To rekindle the spark of your youth or to interest young people in radio, consider building a radio that doesn't need an outside antenna or ground that might be hard to come by with modern high-density housing, antenna restrictions, and the danger of power line proximity or lightning. Fortunately, a new detector has been developed that does just that.

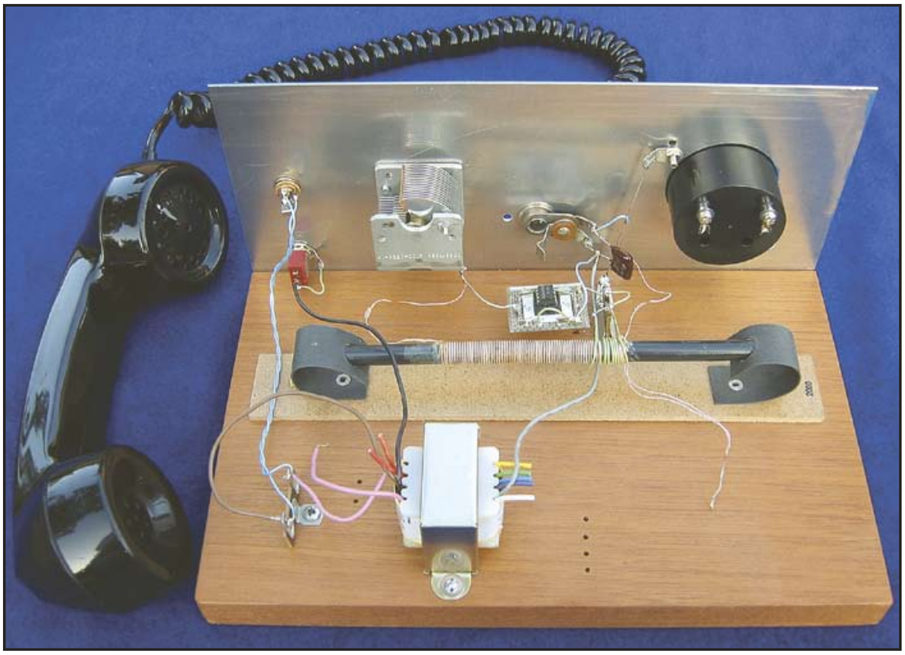

Rear view of the AM band high sensitivity crystal set showing a GTE telephone handset as the headphone. The panel meter is not used for this project.

Anyone who has experimented with simple crystal sets knows that there is a trade-off between sensitivity and selectivity, or the ability to separate stations that are close in frequency. To understand that, refer to Figure 1, the schematic diagram of a typical crystal set. For highest sensitivity, it is desirable to place the detector diode at the top of the LC resonator (which provides the means for selectivity) where the maximum RF voltage appears with the hope of exceeding the 0.3 V threshold of the detector diode. Placing the detector and headphone load at the top of the resonator at point A results in "de-Q-ing" the resonator, largely destroying selectivity.[2] Placing the detector at a tap point on the inductor, typically 10% of the way up from the ground end at point B, results in much less loading of the resonator and preserves selectivity. Unfortunately, the top of the resonator must now see 3 V peak in order for point B to provide 0.3 V to the detector, so sensitivity is severely affected.

Referring to Figure 2, it is possible to have your cake and eat it, too, with a circuit using the newly introduced ALD110900A zero-threshold-voltage MOSFET IC made by Advanced Linear Devices.[3] Figure 2 shows the circuit topology, which was motivated by a similar circuit using a JFET and battery gate bias configuration developed by David W. Cripe, NE4AM (formerly KC3ZQ).[4] The circuit is essentially a synchronous rectifier. It uses the high voltage end of the resonator formed by a ferrite antenna rod, L1, and tuning capacitor, C1, to drive the gate of the MOSFET, resulting in rectifier action from the channel. The MOSFET source and drain replace the anode and cathode of the usual 1N34A diode connected to a separate 8-turn winding on L1, which is equivalent to the 11% tap point on L1.

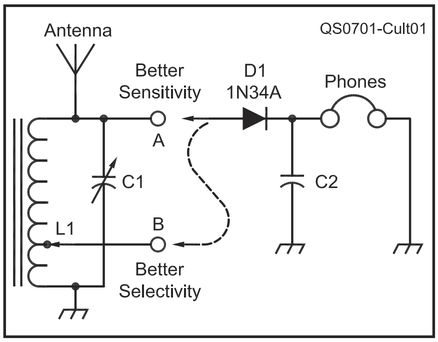

Figure 1 — Schematic diagram of a typical crystal set with diode connected to A for sensitivity and B for selectivity.

In a synchronous rectifier, the MOSFET acts like a switch, turning on during the positive half cycles of the applied RF voltage and off during the negative half cycles. The 2000 pF capacitor, C2, integrates the pulsating dc to create the audio modulation in the headphones. In this circuit, the gate voltages are not high enough to create a perfect switch, and so the MOSFET modulates the resistance of the source-drain channel over some impedance range whose average approximates the impedance of the headphone load for maximum energy transfer.

In the radio shown in Figure 2, both FETs of the dual-FET ALD110900A, U1, have a common-source connection and are connected in parallel to better match the headphone impedance. The resulting detec-

---

[1] Notes appear on page 33.

From January 2007 QST © ARRL

tor has an off resistance (0 V on the gate) of 52 kΩ and an on resistance of 250 Ω with a peak voltage of +5 V on the gate. In this radio, the maximum measured gate voltage is only 0.25 V peak, so the channel modulation creates an average source impedance to the headphones of approximately 25 kΩ. An antique pair of 25 kΩ headphones can be used directly if you have them, but most headphones will require an impedance transformer, in this case the multi-tapped Bogen T-725 public address matching transformer.

Most of the other parts are scarce or no longer in production. You may have to wind your own ferrite rod to emulate the J. W. Miller 2000 antenna rod or adapt the rod from an old AM transistor radio. To wind your own, obtain an Amidon Associates ferrite rod, part number R61-050-750, which is 7.5 inches long and 0.5 inch diameter.[5] Using 15/44 or 15/46 Litz wire, wind 70 turns in the center of the rod spaced to fill 3 inches.[6] The 8-turn secondary can be ordinary magnet wire wound over primary, on the lower end of the main coil. A high-quality 365 pF air variable tuning capacitor can be purchased or obtained from an old radio. A good junk box and resourcefulness are in order.

There is No Magic

Even though the zero threshold of the MOSFET detector greatly improves performance, the available power is low and the acoustic power from the headphones depends on extracting the greatest efficient from every component in the radio.

The use of high-sensitivity earphones is critical to the success of your radio. The best inexpensive stereo earphones have a sensitivity of 108 dB sound pressure level per milliwatt (SPL/mW), where 0 dB SPL is 20 microPascals SPL or the 0.0001 pW/cm^2 threshold of human hearing, and an impedance of 8 to 64 Ω. Although these can be matched to the radio with the Bogen T-725 transformer (which has a separate 8 Ω winding) and offer superior bass response, this is a very low sensitivity compared to what is desired. The receiver of an old dial type or early model tone access pad telephone may be a better choice. I have used an audio signal generator and RadioShack sound level meter to measure the sensitivity of telephone receiver elements such as the GTE D-51030A or Western Electric U-1 or LB-1, which can be obtained from old telephones in thrift shops for several dollars. The sensitivity is about 122 dB SPL/mW or 14 dB better than stereo headphones, although the response is limited to about 3 kHz. If we assume an earphone aperture of 1 cm^2, this is an efficiency of 16% in converting electrical power to sound power. The Knowles Acoustics CM-3152 balanced-armature element

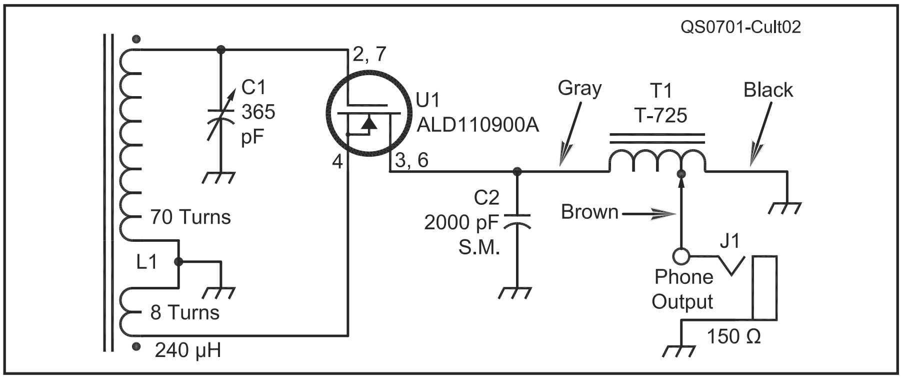

Figure 2 — Schematic and parts list for high sensitivity crystal set.

C1 — 15-365 pF air variable capacitor (www.midnightscience.com or see text).
C2 — 2000 pF silver mica or ceramic capacitor.
J1 — Mono headphone jack.
L1 — 240 μH, 7.5-inch ferrite loop antenna (J. W. Miller Model 2000 or see text). Eight turn secondary wound on lower end, over 70 turn primary winding.
T1 — Bogen T-725 public address matching transformer (www.schmarder.com or www.Grainger.com). Note: Only three of many taps shown to match 25 kΩ output to 150 Ω headphone impedance.
U1 — ALD110900A dual MOSFET (www.mouser.com, part number 585-ALD110900APAL for PDIP-8 or 585-ALD110900ASAL for 8-SOIC).
Headphones — See text.

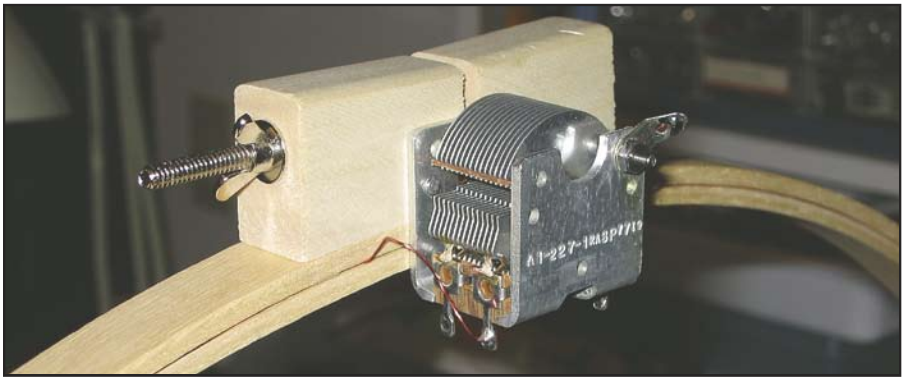

Figure 3 — Variable capacitor mounted on auxiliary quilting-hoop loop antenna.

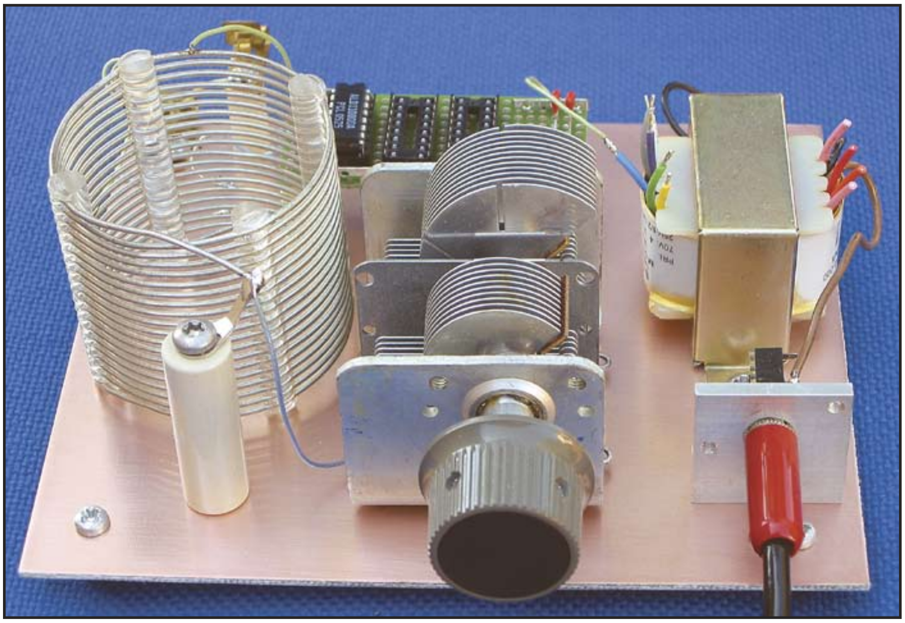

Figure 4 — This photo shows the 90 meter to 40 meter shortwave version of the high sensitivity crystal set.

should offer similar performance.[7] All of these elements have an impedance of 150 Ω at 1 kHz.

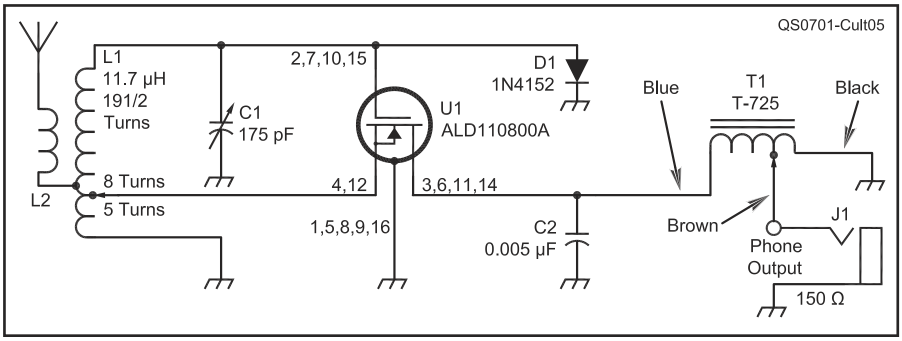

Figure 5 — Schematic and parts list for 90 meter to 40 meter shortwave version of high-sensitivity crystal set.

C1 — 175 pF air variable capacitor.
C2 — 0.005 µF ceramic capacitor (or two 0.0022 µF capacitors in parallel).
D1 — 1N4152 silicon diode (see text).
J1 — Mono headphone jack.
L1 — AirDux coil, 19.5 turns, 2 inch diameter, 10 turns/inch, tapped at 5 and 8 turns.
L2 — 9 turns FT 50-43 toroid (see text).
T1 — Bogen T-725 public address matching transformer. (Note: BLU tap, 5 kΩ, is different from AM version.)
U1 — ALD110800APCL Quad MOSFET (Mouser 585-ALD110800APCL).

Making it Even Better

An additional improvement in received volume can be obtained by coupling a high-Q tunable loop antenna to the ferrite rod antenna in the manner of a "loose coupler." I have used an 11-inch diameter Select-A-Tenna loop to obtain 65 dB SPL A-weighted volume (about normal conversational volume with 3.3 µW electrical input) from a 5 kW station 3 miles away.[8] Replacing the MOSFET with a 1N270 germanium diode resulted in no discernable output. It has been possible to hear seven local stations in a 25 mile radius with no difficulty. You can search for your station locations on the FCC Web site.[9] Signals with 30 dB SPL (whisper level) volume are quite listenable with only 0.1 µW electrical input to the earphone. It only takes 5 to 10 mV peak-to-peak on the source of the MOSFET to provide an adequate volume level.

You can make your own "amplifying" tunable loop antenna with a 23 inch diameter wooden quilting hoop. This will come apart as an inner hoop and a split outer hoop with wooden ends and a tightening screw. Close-wind 12 turns of AWG no. 24 magnet wire in the middle of the outside surface of the inner hoop and tape the ends in place. Next, at about 2 to 3 inch intervals, carefully separate the turns to spread them evenly across the ¾ inch width of the inner hoop and tape in place as you proceed. Mount the outer hoop over the inner hoop and tighten in place using the long metal screw. Mount a 365 pF air variable tuning capacitor on the large wooden blocks as shown in Figure 3. It will be necessary to drill an additional hole in the mounting blocks to pass the shaft of the capacitor. Use a shaft extension with a knob to provide tuning ability. Solder the two ends of the coil to the stator and rotor of the capacitor.

The auxiliary loop that you have created has an inductance of about 240 µH and will tune the entire AM broadcast band with a 365 pF capacitor. Place the loop off the end of the crystal set's ferrite loop. A line across the diameter of the loop should point to the desired station. Tune in a station on the crystal set and peak the volume by tuning the large loop. The received signal strength will be dramatically increased. With this larger loop, the station previously mentioned increased in volume to 72 dB SPL using a GTE telephone receiver.

Shortwave Version

I have constructed a version of this receiver to cover the 90 to 40 meter shortwave spectrum. Figure 4 is a picture of this receiver and Figure 5 shows the schematic diagram and parts list. Shortwave performance is difficult, especially at my US West Coast location where few powerful shortwave stations can be heard and severe fading is evident. With this receiver I was able to reliably receive HSK9, Radio Thailand, broadcasting from the 250 kW International Broadcasting Bureau (IBB) transmitter in Delano, California on 5890 kHz. I was using an external 40 foot wire antenna.

The detector consists of four MOSFETs in parallel using the quad version of the detector IC in order to better match the lower resonator impedance. The source tap point was also changed to the 25% point on the inductor. The resulting loaded Q was 44 at 6 MHz compared to 53 for the AM receiver at 1 MHz. The integrator capacitor was changed to 0.005 µF and the transformer tap was changed to the 5 kΩ tap (BLU). In order to protect the gates of the MOSFET array, the V– and substrate diode connections were grounded and a 1N4152 silicon diode was placed from gate to ground. This diode does not participate in any detector action.

As with any crystal receiver, antenna matching is important to transfer as much power to the receiver as possible. I found that connecting the antenna to the eighth turn of the inductor was optimum. Substantial interference from local FM and TV stations can be greatly reduced by placing a 9 turn FT50-43 toroid in series with the antenna.

Make it Your Own

This article is intended to motivate experimentation in radio at the most fundamental level. If you build either of these sets, you will learn much about high-efficiency radio design and construction, and you will appreciate how the human ear can hear audio signals with energies much less than 1 µW. You will have to adapt the materials you have or can obtain. There is no joy in radio quite like listening to a radio where the only energy is being provided by the station itself, but you will have to cultivate the "art of listening." Hopefully, you can pass on this learning and joy to young people and help them catch the "radio bug." They may well become future hams and extend this wonderful hobby for some time to come.

I would like to thank Wes Hayward, W7ZOI, for valuable discussions during the development of these receivers.

Notes

[1] A. Morgan, *The Boy's First Book of Radio and Electronics*, Charles Scribner's Sons, New York, 1954, pp 126-159.

[2] W. Hayward, R. Campbell, and B. Larkin, *Experimental Methods in RF Design*, ARRL, 2003, pp 3.8-3.9.

[3] **www.aldinc.com/pdf/ALD110800.pdf**.

[4] D. W. Cripe, KC3ZQ, "Nostalgia For The Future," *73 Amateur Radio Today*, Dec 1995, pp 14-16.

[5] **www.amidoncorp.com**.

[6] **www.schmarder.com**.

[7] **www.digikey.com**, part number 423-1036-ND.

[8] **www.selectatenna.com**.

[9] **www.fcc.gov/mb/audio.amq.html**.

Bob has been licensed since 1983 and holds an Advanced class license and a General Radiotelephone Operator License. He has a BA in Physics from Portland State University and works as a design engineer at Phoseon Technology. Bob's interests include QRP, homebrewing, shortwave listening and musical instrument construction. Bob's wife Terry, KA7VAF, is a social worker and is very supportive of his many projects. You can contact Bob at 5860 SW 161st Ave, Beaverton, OR 97007; **n7fki@teleport.com**.

From January 2007 QST © ARRL

BuxComm Corporation

EVINGTON — LYNCHBURG, VIRGINIA 24550
Toll Free: 800 726 2919 or 866 300 1969. FAX 434 525 4919

Shopping Cart - 0 item(s) - $0.00

Log In

View Our Products

Quick Links

Information

Search

Contact Us

AEA, KPC, TNC, KAM & RASCAL Cable Selection Guide | **New Products** | *** Ham Radio Handbooks & Help files** | **ABOUT US**

Home :: Diodes, LED, Signal, Switching, & Zener :: 1N34A 1N34 (D07 case) point-contact 1N34A Germanium diodes use as TV, FM, and Crystal Radio Detectors. Pack of 4

Product 2/13

prev listing next

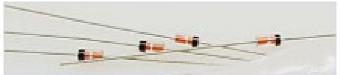

larger image

1N34A
1N34 (D07 case) point-contact 1N34A Germanium diodes use as TV, FM, and Crystal Radio Detectors. Pack of 4

$1.95

cat# 1N34A Crystal Radio Signal Detection. Pack of 4

The original, point-contact 1N34A Germanium diode are the classic 1N34A, These are the glass 1N34A that have a highly sensitive, point-contact to the Germanium chip inside. All have long, full length leads.
The 1N34A Germanium diodes are for use as TV, FM, and Crystal Radio Signal Detection. Pack of 4

***CLICK HERE to read the* Germanium_vs_SiliconDiodeTesting.pdf. This short document will help you to understand the difference between Germanium and Silicon diodes.**

***These are premium quality, and prime; Before our 1N34a Diodes are* packaged, *they are tested, and MUST meet the specifications* shown in the following table:**

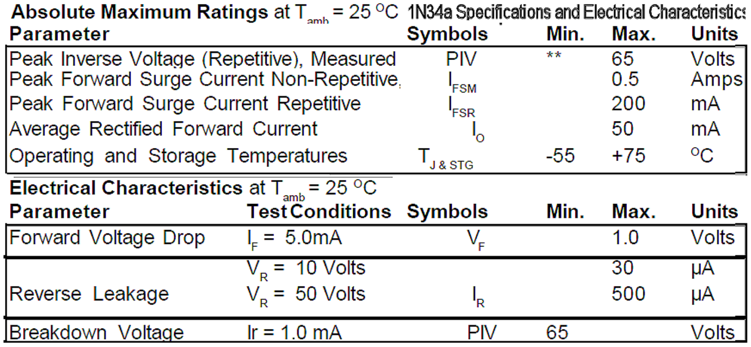

Absolute Maximum Ratings at T_{amb} = 25 °C 1N34a Specifications and Electrical Characteristic

Parameter	Symbols	Min.	Max.	Units
Peak Inverse Voltage (Repetitive), Measured	PIV	**	65	Volts
Peak Forward Surge Current Non-Repetitive,	I_{FSM}		0.5	Amps
Peak Forward Surge Current Repetitive	I_{FSR}		200	mA
Average Rectified Forward Current	I_O		50	mA
Operating and Storage Temperatures	$T_{J \& STG}$	-55	+75	°C

Electrical Characteristics at T_{amb} = 25 °C

Parameter	Test Conditions	Symbols	Min.	Max.	Units
Forward Voltage Drop	I_F = 5.0mA	V_F		1.0	Volts
	V_R = 10 Volts			30	µA
Reverse Leakage	V_R = 50 Volts	I_R		500	µA
Breakdown Voltage	Ir = 1.0 mA	PIV	65		Volts

Used in various music amplifiers as "FUZZ" creators, and Use to build various Crystal Radio circuits. Crystal Radio's do not need any power source, only a wire antenna 10 to 20 feet high, and 20 to 40 feet long.

1N34A, Germanium Diode,TV FM AM Radio **...**
Diagram below is a Germanium Diode AM/FM Crystal Radio.

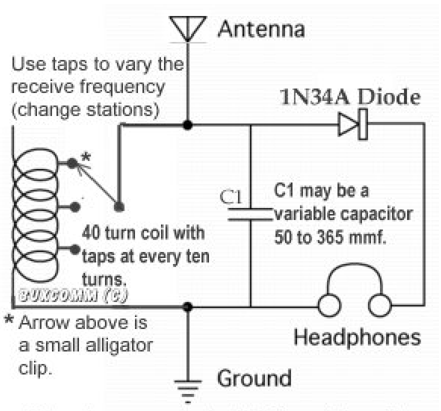
Antenna
Use taps to vary the receive frequency (change stations)
1N34A Diode
*
C1
C1 may be a variable capacitor 50 to 365 mmf.
40 turn coil with taps at every ten turns.
BUXCOMM (C)
* Arrow above is a small alligator clip.
Headphones
Ground

Simple, easy to build Crystal radio.

1N34a,

Shipping Weight: 0.1lbs

Add to Cart:
1
add to cart

Customers who bought this product also purchased...

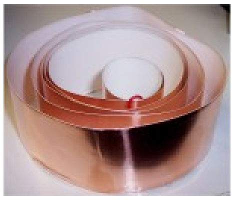

15X36COPPER
36 inches long, by 1-1/2 inches wide copper tape with conductive, adhesive back.

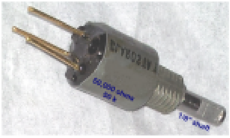

50K-POT
Volume or Squelch control (potentiometer), is an ideal audio control for that sound-card to transceiver interface

UV5R BAOFENG-SPKRMIC
Speaker -Mic for BAOFENG, Kenwoods, POFUNG, HYT, PUXING, LINTON, TYT, RETREVIS, & Wouxon

13k8XFMR
Audio Transformer Transformateur Audio 1300 ohms to 8 Ohm Ohms

7304A
Fits RG 58 RG 58 U RG 58 AU etc. The connector for repeater sites and antenna that require type N fit

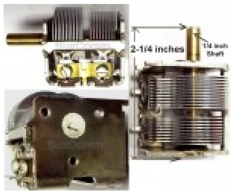
2-1/4 inches
1/4 inch Shaft

AIRVARI365 **Crystal Radio, Dual Gang, air variable, Capacitor. Large section is 365 pf.**

Home | BuxCommCo Shipping & Returns Policy | * Read what our Customers say | New Products | Site Map | HTML Sitemap

Copyright © 2018 www.PacketRadio.com. Powered by Zen Cart
Designed by: TotallyZenned
Your IP Address is: 96.64.239.61

Parse Time: 0.199 - Number of Queries: 235 - Query Time: 0.0406239033203

Crystal Set Mobile

By Al Klase – N3FRQ – December 2006

Back in the 1990's there was a Yahoo Internet "club", devoted to crystal radio development called "The Ferrite Core." One of the perennial topics of conversation was the difference between operating a crystal set near high-power stations as opposed to a more rural setting. Someone, who had read Aesop's Fables likened this to "The City Mouse and the Country Mouse." The country mouse needed a sensitive radio hooked to a large antenna, but could faintly hear distant stations without a lot of interference. The city mouse's crystal set could drive a loudspeaker using only the bed springs for an antenna, but he couldn't hear much beyond the city limits.

I've spent most of my life as a country mouse, but recently I've been frequently finding myself in Jersey City. Indeed, crystal set operation is very different there. Short antennas work pretty well, and even my nearly-deaf rocket radio picks up multiple stations. It's even possible to operate a properly designed set with the antenna terminal connected to ground with the headphone cord and listener serving as the actual antenna.

I decide to apply what I learned down town to a couple of radios I was building as Christmas presents. The circuit is very similar to the NJARC Pretty-Good Crystal Set, but the coil is wound on a ferrite toroid to save space, and there are additional antenna taps on the "high end" of the coil to better accommodate short antennas.

I recently found myself in Maywood, near Hackensack, with some time to kill while Peggy rehearsed before a Sunday-afternoon concert. I took the city mouse for a walk in a nearby park. With 20 feet of wire, 5 feet in the air, and the ground wire clipped to a metal park bench, I could comfortably choose from five or six stations. I could still hear the Jets game on WABC, with just the ground connected to the antenna terminal.

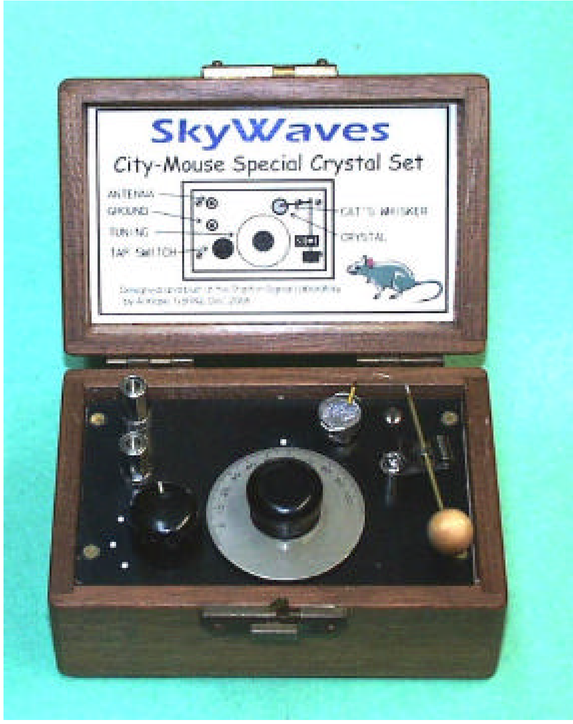

Still having some time to waste, I picked up a four-foot stick as I returned to the car. I grounded the set to the ignition key, and extended the antenna wire out the sunroof with the stick. Sure enough, the Jets were still ahead. I wound the rest of the wire around the stick as a loading coil.

The ugly stick

After an enjoyable concert (The Four Seasons and a Brandenburg Concerto w/harpsichord), I listened to the crystal set down the turnpike through the Meadow Lands, past all the towers, all the way home. I think most were on night-time power by now, as none were obnoxiously loud. I need to try that again, in broad daylight, with a speaker.

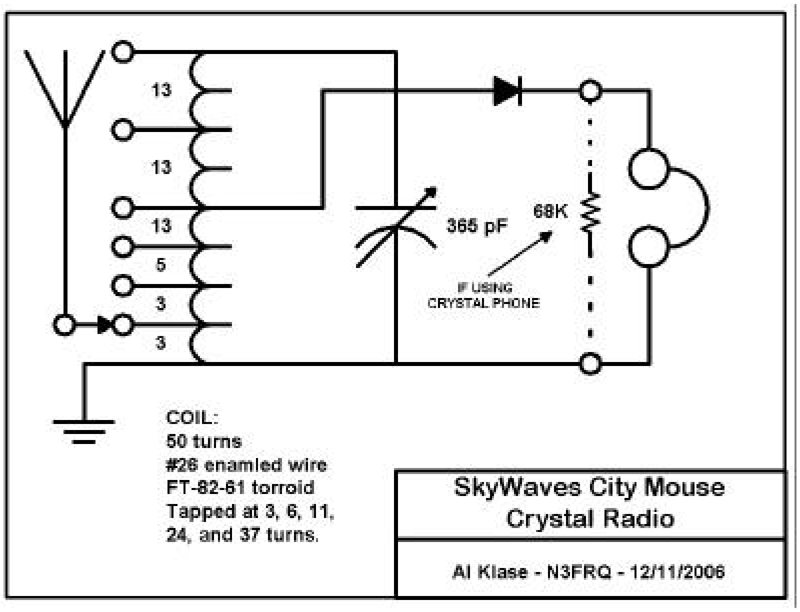
13
13
13
5
3
3
365 pF
68K
IF USING
CRYSTAL PHONE
COIL:
50 turns
#26 enamled wire
FT-82-61 torroid
Tapped at 3, 6, 11,
24, and 37 turns.
SkyWaves City Mouse
Crystal Radio
Al Klase - N3FRQ - 12/11/2006

Crystal Set Mobile

By Al Klase – N3FRQ – December 2006

Back in the 1990's there was a Yahoo Internet "club", devoted to crystal radio development called "The Ferrite Core." One of the perennial topics of conversation was the difference between operating a crystal set near high-power stations as opposed to a more rural setting. Someone, who had read Aesop's Fables likened this to "The City Mouse and the Country Mouse." The country mouse needed a sensitive radio hooked to a large antenna, but could faintly hear distant stations without a lot of interference. The city mouse's crystal set could drive a loudspeaker using only the bed springs for an antenna, but he couldn't hear much beyond the city limits.

I've spent most of my life as a country mouse, but recently I've been frequently finding myself in Jersey City. Indeed, crystal set operation is very different there. Short antennas work pretty well, and even my nearly-deaf rocket radio picks up multiple stations. It's even possible to operate a properly designed set with the antenna terminal connected to ground with the headphone cord and listener serving as the actual antenna.

I decide to apply what I learned down town to a couple of radios I was building as Christmas presents. The circuit is very similar to the NJARC Pretty-Good Crystal Set, but the coil is wound on a ferrite toroid to save space, and there are additional antenna taps on the "high end" of the coil to better accommodate short antennas.

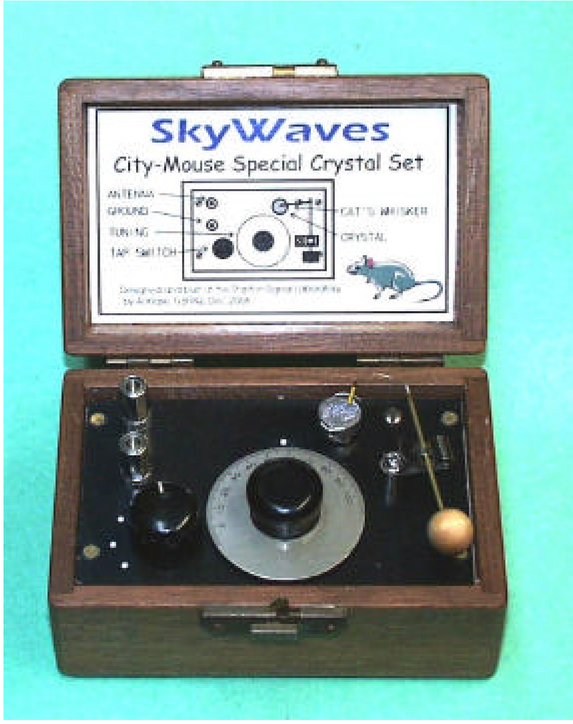

I recently found myself in Maywood, near Hackensack, with some time to kill while Peggy rehearsed before a Sunday-afternoon concert. I took the city mouse for a walk in a nearby park. With 20 feet of wire, 5 feet in the air, and the ground wire clipped to a metal park bench, I could comfortably choose from five or six stations. I could still hear the Jets game on WABC, with just the ground connected to the antenna terminal.

Still having some time to waste, I picked up a four-foot stick as I returned to the car. I grounded the set to the ignition key, and extended the antenna wire out the sunroof with the stick. Sure enough, the Jets were still ahead. I wound the rest of the wire around the stick as a loading coil.

The ugly stick

After an enjoyable concert (The Four Seasons and a Brandenburg Concerto w/harpsichord), I listened to the crystal set down the turnpike through the Meadow Lands, past all the towers, all the way home. I think most were on night-time power by now, as none were obnoxiously loud. I need to try that again, in broad daylight, with a speaker.

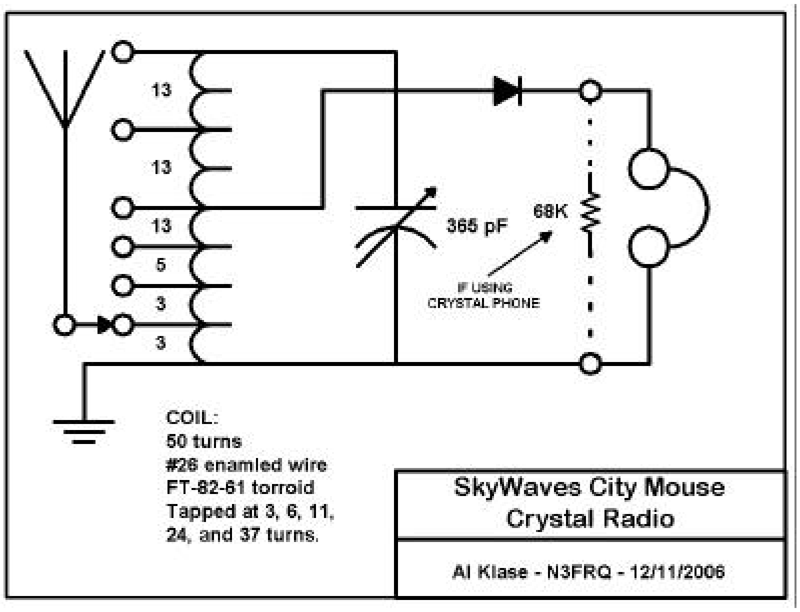
13
13
13
5
3
3
365 pF
68K
IF USING
CRYSTAL PHONE
COIL:
50 turns
#26 enamled wire
FT-82-61 torroid
Tapped at 3, 6, 11,
24, and 37 turns.
SkyWaves City Mouse
Crystal Radio
Al Klase - N3FRQ - 12/11/2006

THE CRYSTAL-SET RADIO

100 more Crystal Set plans

Everyone wants to make a RADIO.

The simplest radio is a **CRYSTAL SET**. (Sometimes called a **Crystal-Set Radio** or **Xtal Radio Set** or **Crystal Diode Radio**.)

However a **Crystal Set** needs a number of components that are very hard to get: (tuning capacitor with knob) and (crystal earpiece for $1.25).
The "air" TUNING CAPACITOR (20p to 415p) is not easily available and the postage is expensive. The germanium diode is a special component and the aerial coil wound on a ferrite slab is difficult to obtain.

But you don't need these components. They can be substituted.

There are hundreds of websites on the internet describing the **CRYSTAL SET** and if you want to build a "normal" set, you can Google these sites or 100 more Crystal Set plans. Many of them sell kits too.

But this article is different.

We are going to have all the fun of making a **CRYSTAL SET** but with modern components and easy-to-make components and with an amplifier stage. The output is loud so you don't need a long antenna. And we are going to make our own TUNING CAPACITOR and a very simple aerial coil (called a FRAME AERIAL) as well as replacements for the germanium diode (use a TRF radio IC or a transistor) and in place of hi-impedance headphones (use a piezo diaphragm) and a crystal earpiece equivalent (a piezo diaphragm).
It's even better to have one of each type of component so you can compare the performance, so no matter how many parts your get, nothing will be wasted.

We are also going to explain the fundaments of how the circuit works as even the simplest circuit has a number of very important features that are used in many other circuits.

But first we are going to learn about the components and how they combine to make the circuit work.
When two or more components are connected together they sometimes produce a completely different result to the capabilities of either item.
This is the case with a capacitor and inductor in parallel. An inductor is simply a coil - turns of wire on a cardboard tube - called a former and the centre of the coil is AIR. It is called an air-cored coil or air-cored inductor.
Each component (the coil and capacitor) is called a PASSIVE DEVICE - in other words it does not amplify, but when the are connected together they create a result very near to amplification. And they also produce a result of picking up a huge number of signals and only allowing one signal to appear across the pair. A truly amazing result.

We start by placing a capacitor across the coil.
There is so much activity in the air, from radio, TV, taxi and mobile phone usage that the air is filled with electromagnetic radiation.
This radiation will cut the turns of the inductor (the coil) and produce a microscopic voltage in the turns. This is enough to start the two components passing energy back and forth at a rate

determined by their values. This is the basis of our first discussion.
These two components are called a TUNED CIRCUIT and make up our first building block called THE FRONT END.
.

THE FRONT END

This consists of a coil and capacitor. These two components are in PARALLEL and the signal (called the RADIO SIGNAL or RADIO WAVE) passes through the centre of the coil and produces a voltage in the turns of the coil. The electromagnetic wave has to pass through the centre of the coil.
This voltage is the result of a mass of signals that are interfering with each other and producing a signal called BACKGROUND NOISE.
The voltage can be increased by an external aerial (called an ANTENNA) and it consists of ALL THE LOCAL radio stations (and everything else).
The voltage is a mass of signals and is absolutely useless as it represents all the stations AT THE SAME TIME.
However, across the coil is a capacitor and these signals charge the capacitor with the very small voltage produced by the energy of the signals. When the capacitor is charged, it delivers its voltage to the coil. The coil accepts the energy and converts it to magnetic flux.
After a very short period of time the capacitor becomes discharged and the magnetic flux collapses and produces a voltage in the coil of the opposite polarity to charge the capacitor again in the opposite direction.
These two components keep oscillating back and forth, using the tiny amount of energy from the stations.
There is a natural frequency for the capacitor and coil to pass energy back and forth and one of the stations will provide energy to assist this natural frequency.
When this happens, the amplitude of the signal increases and the signals from all the other stations cancel themselves out and only one signal (waveform) remains.
This signal is called the NATURAL FREQUENCY OF RESONANCE and it corresponds exactly to one of the radio stations.

The end result is a waveform that is the exact same frequency as one of the radio stations and when voice or music is played, the amplitude of the waveform increases and decreases. This will be the signal you hear in the earpiece or speaker.
It is called an AMPLITUDE MODULATED signal and that is where we get **AM RADIO** from.

Understanding the concept of a CAPACITOR and INDUCTOR in parallel is very important. They form a TUNED CIRCUIT that has a natural RESONANT FREQUENCY.
Here is a very similar analogy. You have a heavy metal ball on a long string attached to the gutter on your house - just like a pendulum. You can push the ball very lightly with a finger and after a number of pushes you will be able to get the heavy ball swinging in a very large arc.
The only way to keep it swinging is to push it very lightly at exactly the right time. If you push it at the wrong time it will eventually stop swinging.
The parallel tuned circuit is exactly like the ball. It wants to oscillate at a particular frequency.
All the radio stations are pushing and pulling the circuit at the wrong times and nothing is happening. But one radio station pushes at exactly the right time and the circuit starts to oscillate.
All the other stations are fighting each other just like one person pushing the ball sideways and another pushing the ball from the opposite side. The results cancel each other and you are the only one assisting the swing.
The TUNED CIRCUIT can also be called a FILTER with a very narrow BAND-PASS frequency but our simple explanation describes the operation much more clearly.

There are a few other terms used to describe the components in the font end:

LOOP STICK ANTENNA -This is an alternate name given to the coil of wire wound on a ferrite rod or slab. It also has the name ROD ANTENNA or FERRITE ROD ANTENNA.
The winding can be enamelled wire or flexible wire called LITZ WIRE. This is very fine strands of enamelled wire twisted together and covered in cotton. The purpose of changing a thick wire to lots of very thin wires is to prevent the radio signals creating loops of signals within the wire and these signals will cancel each other and not produce a signal out the end of the wire.
In our experiments, we have not noticed any difference in a coil made with ordinary enamelled wire and Litz wire.

Here is a set of components to make your own Crystal set from **Scott's Electronic Parts**:
You can see the rod antenna, germanium diode, crystal earpiece, capacitor and resistor. The kit costs about $9.00 plus postage and includes knob, clips and screws but no board to mount the parts.

Here is the circuit for a Crystal Set:

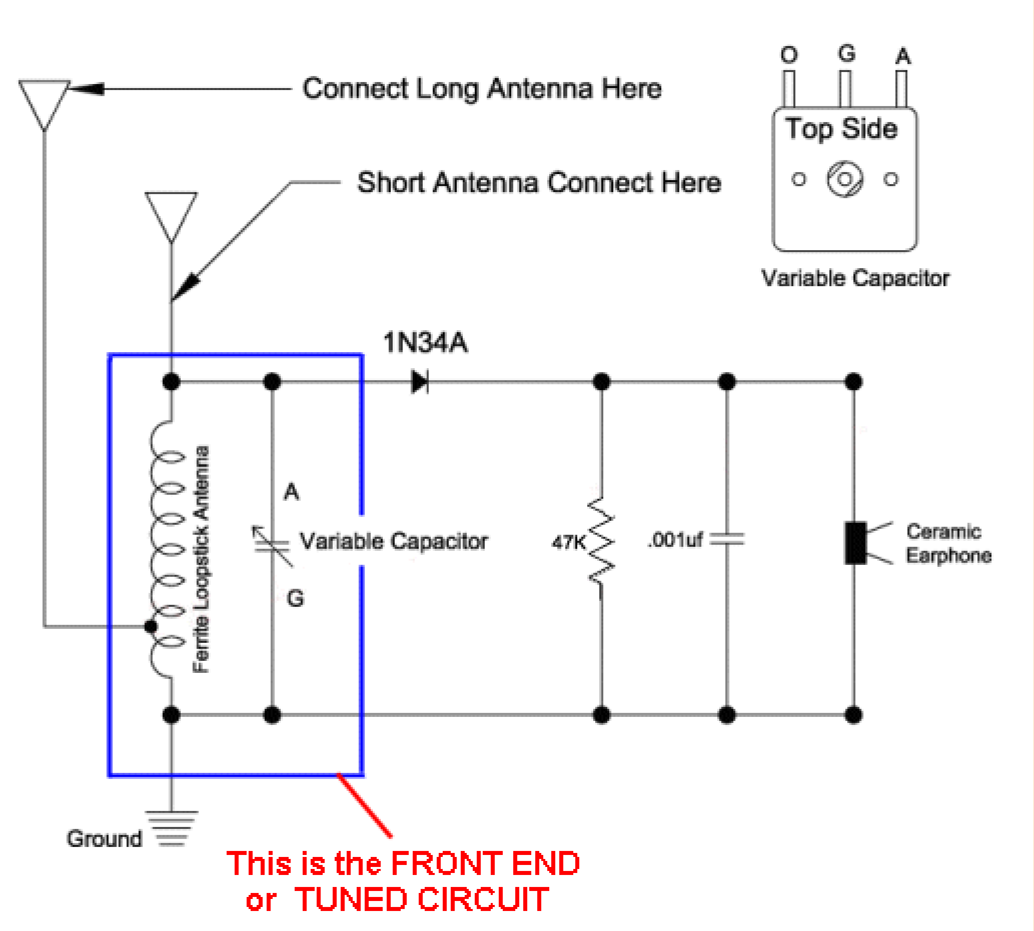
Connect Long Antenna Here
Short Antenna Connect Here
O
G
A
Top Side
Variable Capacitor
1N34A
A
Variable Capacitor
G
Ferrite Loopstick Antenna
47K
.001uf
Ceramic Earphone
Ground
This is the FRONT END or TUNED CIRCUIT

Here are the components mounted on a board called BREADBOARD:

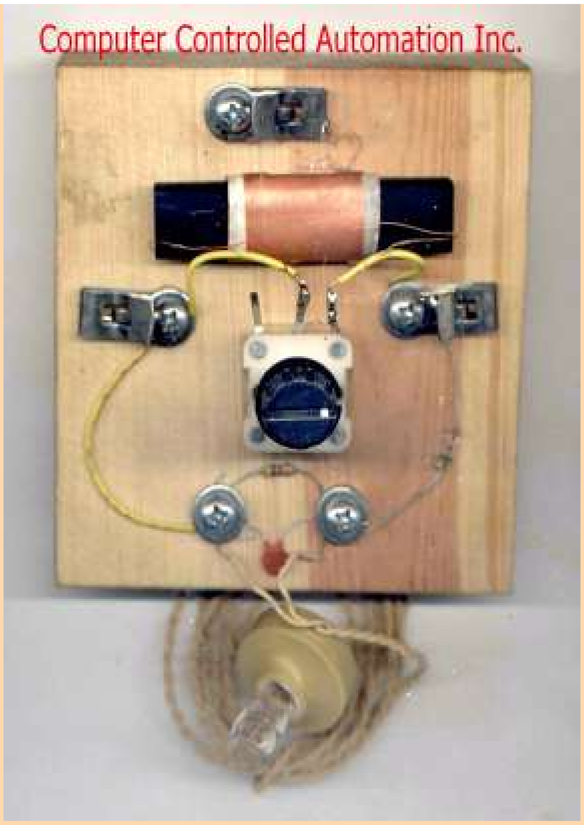

The top clip connects to a long antenna. The left clip connects to ground and the right clip connects to a short antenna.

Let me clear up a point. You do not need a ferrite rod antenna for the coil. You can use an ordinary coil wound on a cardboard tube and it will work just as well if you are using an outside antenna.
Here is the circuit using a home-made coil.

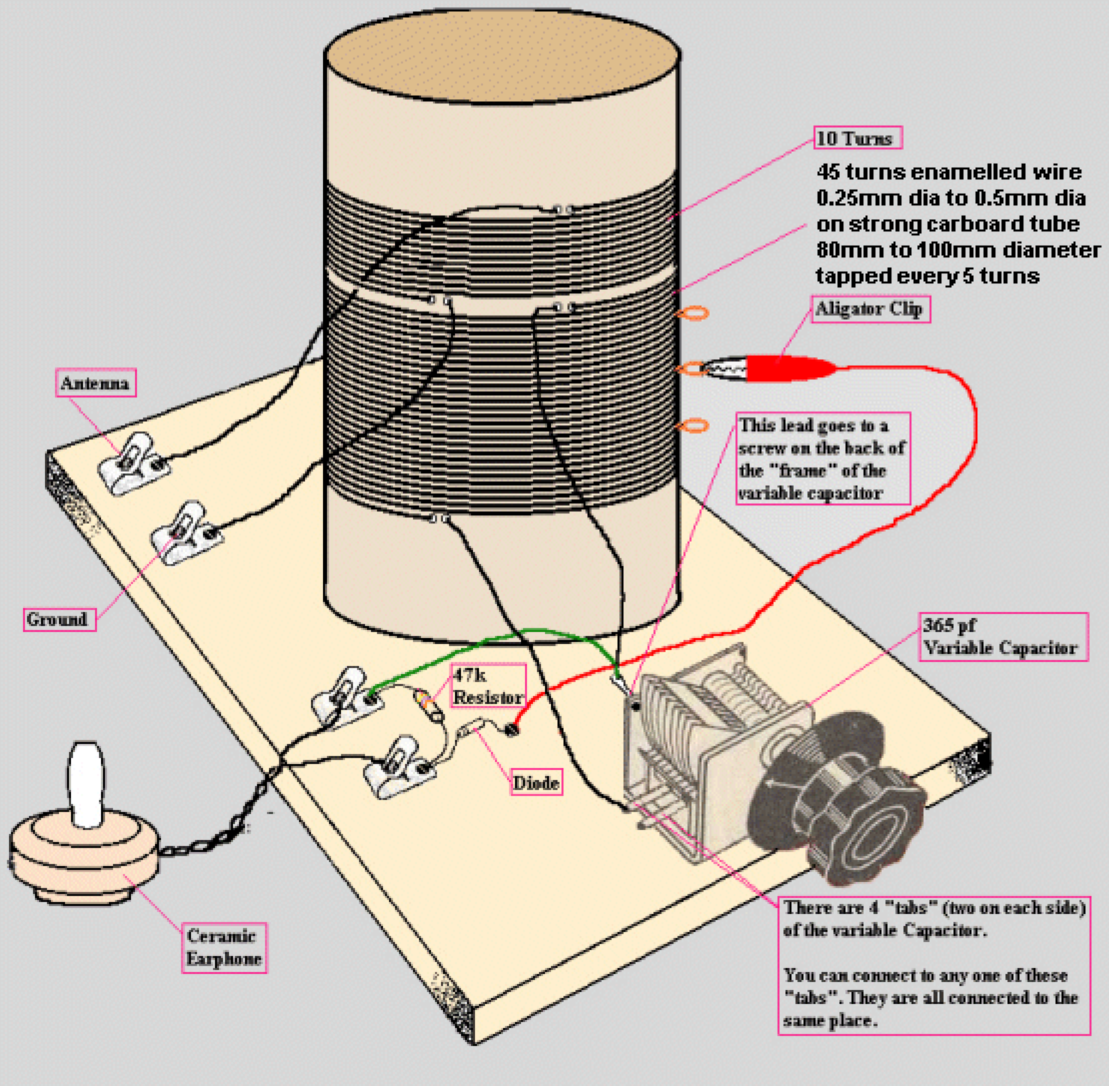
10 Turns
45 turns enamelled wire
0.25mm dia to 0.5mm dia
on strong carboard tube
80mm to 100mm diameter
tapped every 5 turns
Aligator Clip
Antenna
This lead goes to a screw on the back of the "frame" of the variable capacitor
Ground
365 pf
Variable Capacitor
47k
Resistor
Diode
Ceramic
Earphone
There are 4 "tabs" (two on each side) of the variable Capacitor.
You can connect to any one of these "tabs". They are all connected to the same place.

The tappings on the coil allow a wide band of radio stations to be tuned.
Each tapping allows a different portion of the band to be covered.

The next part to understand is this:
The coil and capacitor must not be LOADED. In other words, you cannot connect anything to this combination because the signal it is producing will be "taken away" or "removed" or "considerably reduced" by the item you are connecting to the circuit.
These two components are called a TUNED CIRCUIT and when they are not loaded they pick up all the radio stations, one station at a time, when the natural resonant frequency of the coil and capacitor exactly match the frequency of the radio station. The circuit actually "rejects" all the radio stations except one. Because all the other stations are trying to make the Tuned Circuit oscillate at a different frequency and it does not do this.
The result of the TUNED CIRCUIT oscillating under NO LOAD conditions produces a waveform that is very high and this gives the circuit GOOD SELECTIVITY. The circuit can select one station and reject nearby stations.
It also has good SENSITIVITY as it can pick up weak stations.
If you load the circuit, only the strongest signal will be detected and it will be spread across the full range of the tuning capacitor.
Obviously the theory is more-complex but we are explaining the end-result.
Theory talks about the "Q" value of the coil and this is its ability to produce a very good output when the magnetic flux collapses and the "Q" value increases when the circuit is not loaded. .
Although these voltages are very small (in the order of microvolts or millivolts) the result is very important as the rest of the circuit will be amplifying this waveform a few thousand times.
As we explained above, pushing the weight on a string only needs a push of 1 cm and eventually the weight will swing 1 metre. This is a gain of 100:1 The same thing happens with the tuned circuit. The incoming radio signal is in the order of microvolts, but the coil and capacitor will produce a signal as high as 500 millivolts. This is an improvement or "gain" of more than 1,000 and is referred to as the "Q" of the circuit.
You will also notice the TUNED CIRCUIT is not connected to any supply voltage. It does not have

be connected. It generates its own waveform from the signals in the air.
It should not have any DC current flowing through it via the supply as this would put a load on the circuit and reduce its operation.

However we must "pick-off" the signal so it can be amplified.
This must be done with a very high impedance circuit.

THE CONVERTER (detector) - THE DIODE

The next part of the circuit is the CONVERTER. Commonly called the DETECTOR.
It converts the RADIO FREQUENCY signal to an AUDIO FREQUENCY signal.
This is the job of the DIODE.
The radio frequency signal is a very high frequency signal (say one million cycles per second) and it is sending a tone of one thousand cycles per second through the air-waves.
What is happening is this: The one megahertz signal has a certain amplitude and over a range of the first one-thousand cycles, the amplitude gradually decreases and then increases again. If you look at the tops of this 1,000 cycles you will see a waveform that corresponds to the one kilo-Hertz signal.
The 1MHz signal is picked up by the coil and capacitor in the front end and makes it oscillate.
The radio frequency signal is gradually getting larger over 500 cycles then smaller over the next 500 cycles and this increase and decrease represents the 1,000 cycles per second tone.
This is the waveform (the signal) that passes through the diode. This will be explained further in a moment.
The diode does not pass any signals less than 200mV as the first 200mV is lost in the junction of the diode. This means the signals start to appear on the other end of the diode when they are above 200mV.

This is how the diode works:
Across the crystal earpiece is a capacitor. The capacitor gets charged via the diode.
The diode is present to stop the capacitor getting discharged when the waveform is in the wrong direction. (by this we mean - when the waveform is lower or smaller in amplitude than the voltage on the capacitor).
And the waveform is in the wrong direction about 50% of the time. To charge the capacitor for one-half-cycle requires 500 "little increments" in voltage with each increment adding a microscopic increase in voltage. We don't want this voltage to reduce when the waveform is reversing direction and the diode stops the voltage flowing back to the Tuned Circuit.
During the next half of the cycle when the pulses are getting smaller and smaller, the voltage on the capacitor is "bled off" by the load resistor.
The crystal earpiece detects this voltage. What we mean, is the diode allows the voltage to rise (increase) on the capacitor via lots of little "pulses" and the voltage increases in the form of a sinewave to a maximum amount. This voltage is passed to the crystal earpiece.
Once the voltage rises to a maximum, the little pulses of energy are not quite as strong, and the voltage on the capacitor reduces to form the second portion of the sinewave. This voltage is always being passed to the crystal earpiece and you can hear it as an audio signal.

GERMANIUM OR SILICON DIODE

The preferred type of diode for a Crystal Set is germanium. This is because it drops only about 0.3v.
But a silicon diode can be used, even though it drops about 0.7v, if the radio stations are very loud (close by).
You have to remember, you need a very good aerial (and a water-pipe earth) to get any results with a Crystal Set because you are asking the signal to provide the energy to drive the earpiece.
By simply adding a transistor, you are improving the performance 100 times and the long antenna can be reduced to a FRAME ANTENNA and the earth can be the metal frame of your soldering iron.

THE EARPIECE or EARPHONE
also The Magnetic Earpiece or CRYSTAL EARPIECE

The earphone or earpiece used in a Crystal Set must be a high impedance device because the crystal set does not produce a high current and cannot drive a low-impedance earpiece. That's why a CRYSTAL EARPIECE is ideal.
It has a crystal glued to the back of the earpiece and connected to its top surface is an aluminium diaphragm. When the crystal expands and contracts as a result of a voltage applied via two electrodes, the diaphragm moves and you can hear the signal. It exhibits a very high impedance

because it consists of a crystal and no coil of wire is contained inside the case.
If you do not have a Crystal Earpiece, you can make your own from the shell of an 8 ohm earpiece and a piezo diaphragm. Only the front part of the earpiece is used.

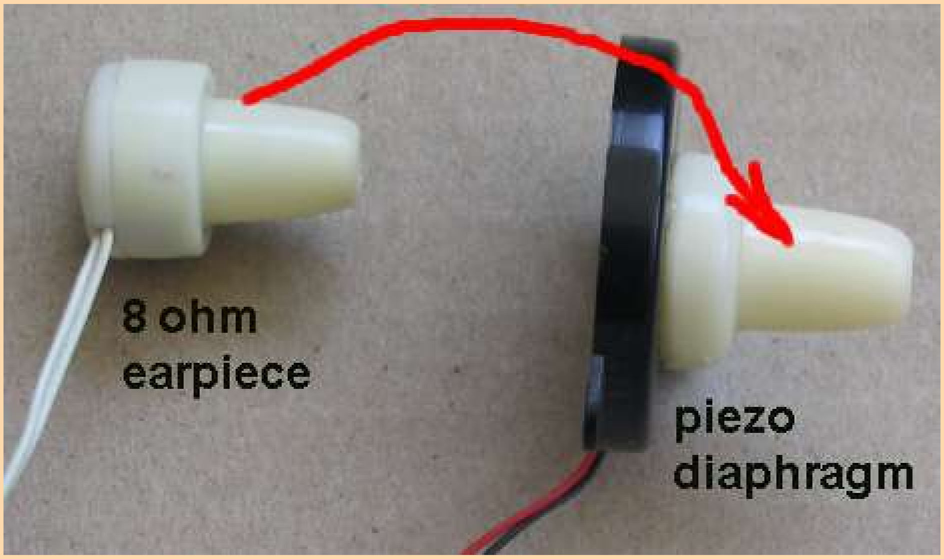

Make your own Crystal Earpiece

Hit the 8 ohm earpiece on the side and the front comes off. Glue the front onto a piezo diaphragm with hot-melt glue. See photo above.
The piezo diaphragm is a ceramic substrate that deflects in the presence of a voltage. It is quite sensitive and you can hear the audio quite clearly.
The waveform emerging from the diode in a Crystal Set is called AUDIO and although it has an amplitude of a few hundred millivolts, it does not have any current associated with it. The crystal earpiece and the piezo diaphragm react to this voltage.

THE 8ohm EARPIECE

The 8 ohm earpiece can be used with our 8ohm Buffer stage shown below.

16ohm 32 ohm and 64 ohm EARPIECE(s)

Earpieces and headsets from mobile phones are 16 ohm or 32 ohm per earpiece and are terminated via a stereo 2.5mm or 3.5mm plug. The earpieces are connected in SERIES to get the best coupling to our radio circuits and you need to find the two pins on a stereo socket to produce series connection. Get a multimeter and switch to "ohms." Try all the pins and you will get a click in the left ear then the right ear. Keep searching until you get a click in both earpieces at the same time. Use these two pins.

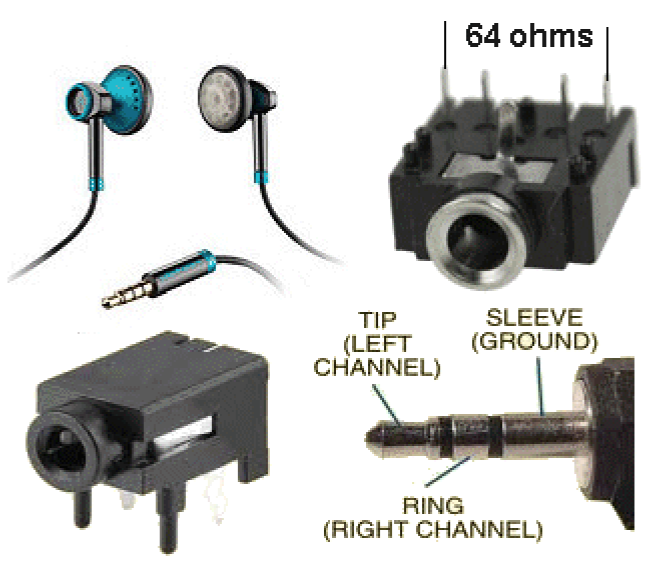

Stereo mobile phone headset - unusually 32R or 64R

PROBLEMS

The biggest problem with a Crystal Set is the need for a long antenna.
The first 200mV to 300mV of a signal is lost in the diode and you need a long antenna to pick up a signal so the output of the TUNED CIRCUIT has enough voltage to drive the high-impedance earpiece.
This requires an outside aerial 5 metres long and 3 metres high.
This is not practical for most hobbyists so we will be adding an amplifying stage to the crystal set so a shorter (smaller) aerial can be used.

THE FRAME AERIAL or FRAME ANTENNA or FRAME COIL

The aerial coil shown in the photo above is a ferrite slab with about 80 turns of Litz wire. You can find one of these in an old broken AM radio or from a parts-shop. In the instructions below we show how to make your own Ferrite Rod Antenna
An equally-good substitute is a frame antenna made by winding insulated wire in a rectangle around wooden sticks.

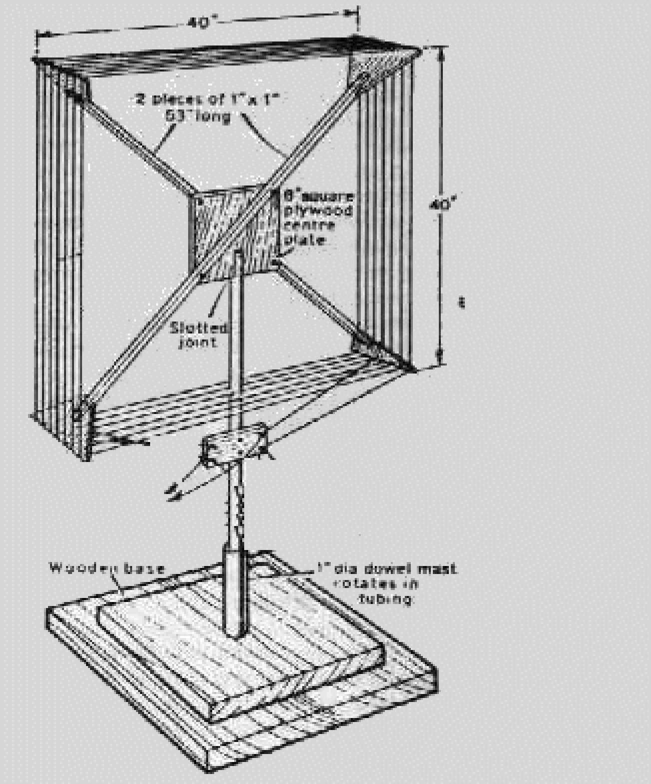
40"
2 pieces of 1" x 1"
53" long
8" square
plywood
centre
plate
40"
Slotted
joint
Wooden base
1" dia dowel mast
rotates in
tubing

FRAME ANTENNA

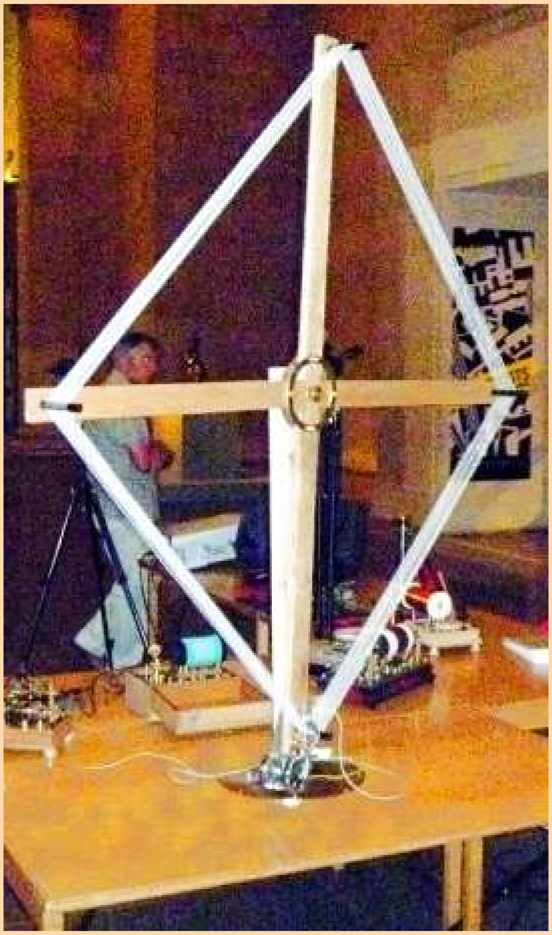

15 turns on a diamond frame

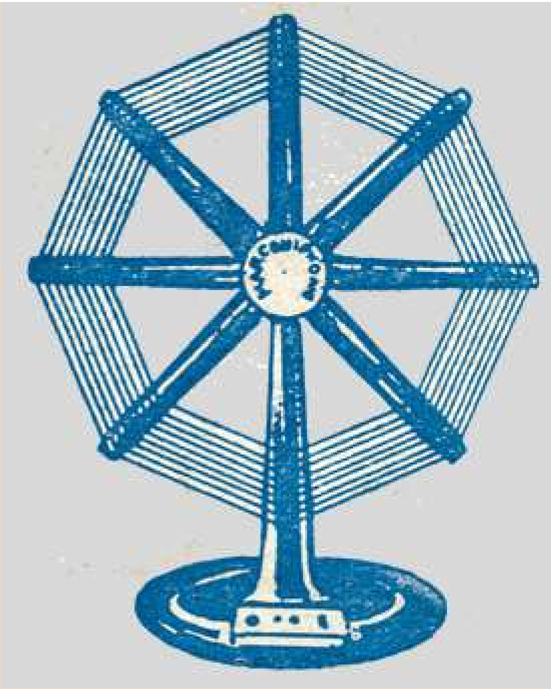

One of the earliest Frame Antennas

The Frame Aerial can be as large as 100cm x 100cm or as small as 10cm x 10cm around a plastic chocolate box.

Here are two FRAME ANTENNAS:

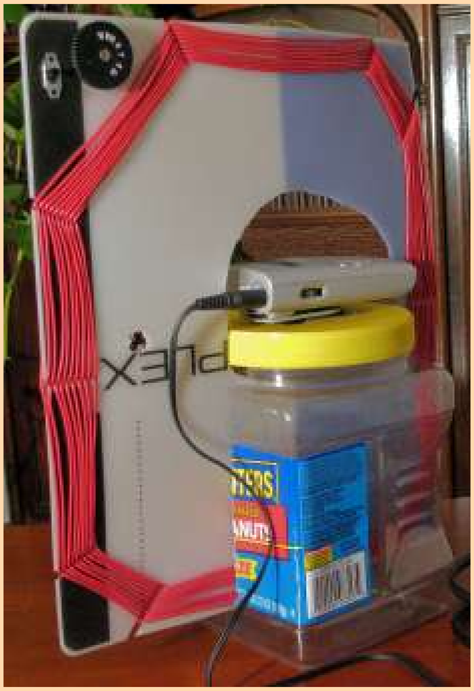

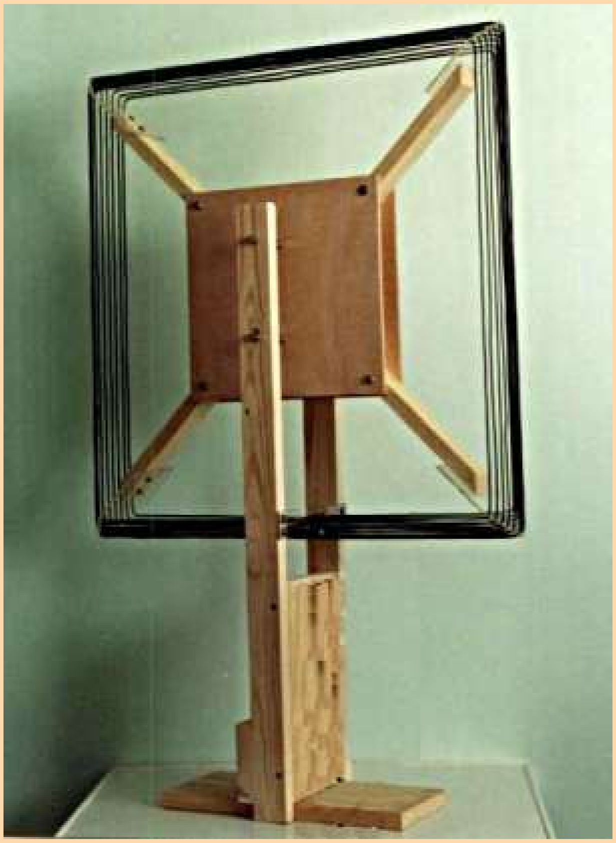

This will work just as good as a ferrite slab antenna. The slab antenna is just 100 times smaller. The slab antenna was invented so a transistor radio could be built in a small case. But if it is not available, you can wind 20 turns around a plastic chocolate box and it will work just as good.

Alternately you can wind 20 turns around a biscuit tin. Put a pencil on the tin and wind the turns over the pencil too. Remove the pencil and it will be easy to remove the turns. Use tape to keep the turns together.

The FRAME AERIAL does two things. It picks up the radio waves and it becomes the coil (called the INDUCTOR) in the TUNED CIRCUIT. It must be placed away from metal objects, such as a refrigerator.

BASKET WEAVE COIL

There is no point making a complex BASKET WEAVE COIL as it will not work any better than simply jumble winding all the turns at the maximum circumference of the coil, because the energy capturing capability of the coil relies entirely on the amount of flux lines passing through the centre of the coil.
By increasing the centre of the coil, the amount of flux is increased for the same coil size.
In fact, the simplest and cheapest is to wind turns around a box, as explained later in this article, or make a frame antenna as shown above. Technically speaking, a round coil has the best performance but only by a few percent.

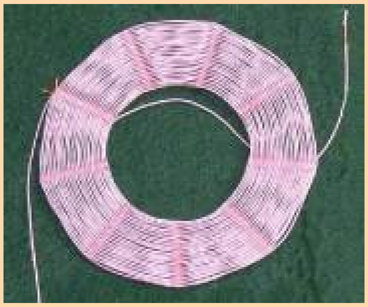

A BASKET WEAVE COIL

THE VARIABLE INDUCTANCE TUNING COIL

Whenever the size or shape of the coil is changed, (or the number of turns), the natural frequency of the Tuned Circuit will change and a different radio station will be picked up.
This means tuning across the band can be done by altering the characteristics of the coil while keeping the value of the capacitor fixed.
Changing the inductance can be done in many different ways.
The coil can have taps every 5 turns and an alligator clips selects the correct tap. But very few radio stations will correspond exactly to each tap.
Another way is to have a slider move up and down the turns as shown in the following image:

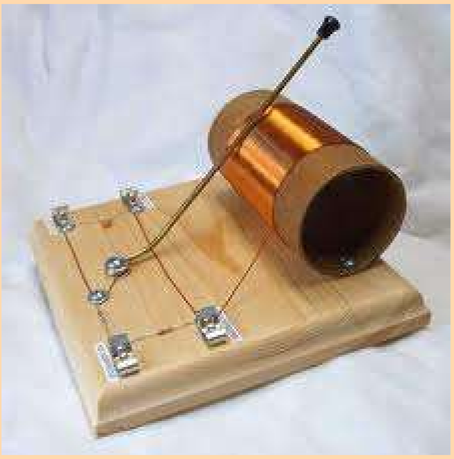

The slider makes contact where the insulation has been removed. But it may touch two turns at the same time and create a "shorted turn" and reduce the "Q" of the coil.

Another way is to move a ferrite bar (rod) in and out of the coil:

THE SLUG TUNED COIL

To tune across the radio band, the natural frequency of oscillation of the TUNED CIRCUIT must be adjusted (changed). This can be done by changing the value of the capacitor or the value of the inductor.

The value of the inductor can be changed by adding or removing turns or changing the amount of magnetic material in the centre of the coil.
A ferrite bar can be screwed in and out of the coil or slid in and out and this component is called a **SLUG TUNED COIL**.
The following diagram shows a **SLUG TUNED CRYSTAL SET**:

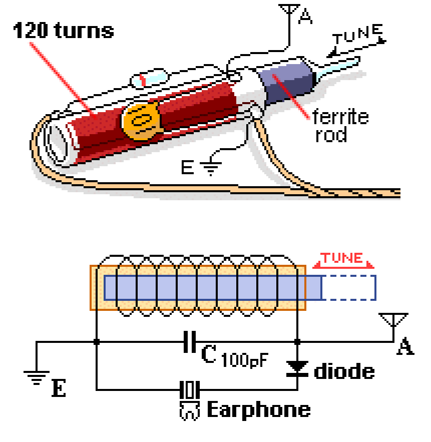
120 turns
A
TUNE
ferrite rod
E
TUNE
A
C 100pF
diode
E
Earphone

By changing the value of the 100p capacitor, different parts of the band can be picked up.

The photo shows a slug tuned coil using 60 turns of insulated wire on a 10mm tube (or any tube that will fit over a 8-10mm ferrite rod) and a circuit containing an AM radio chip plus a buffer driver transistor:

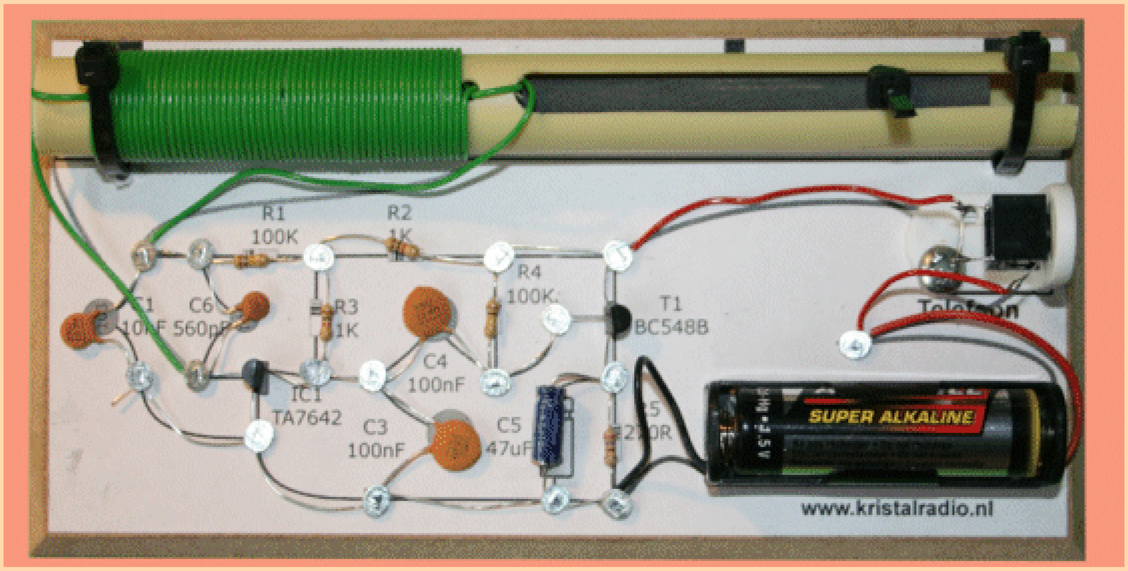
R1 100K
R2 1K
R4 100K
T1 BC548B
Telefoon
C1 10nF
C6 560pF
R3 1K
C4 100nF
IC1 TA7642
C3 100nF
C5 47uF
R5 220R
SUPER ALKALINE
www.kristalradio.nl

A SLUG-TUNED RADIO

The circuit above is has a broad-band amplifier consisting of 10 transistors (IC1) and they are directly coupled (connected) to each other because it is not possible to "manufacture" a capacitor inside the IC. The IC has 3 terminals (pins, legs) and it looks like an ordinary transistor. Experimenting with this type of IC has shown that it is no better than 2 ordinary transistors connected in a direct-coupling arrangement.

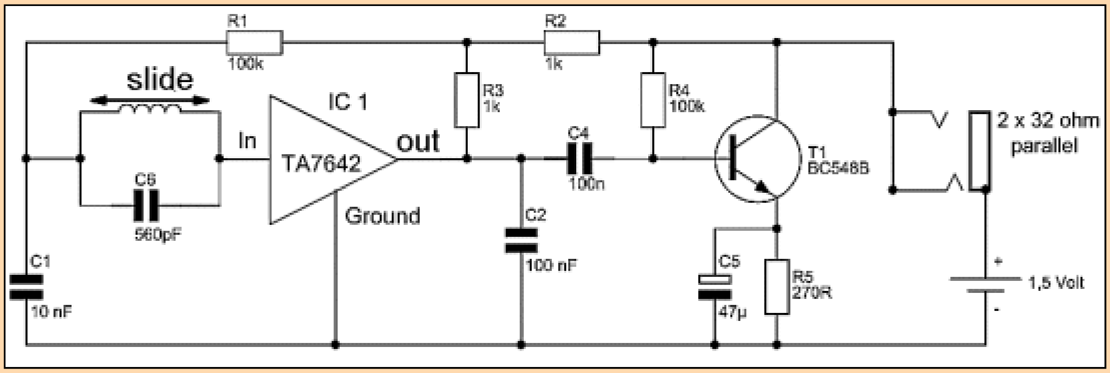
R1
100k
slide
IC 1
In
TA7642
out
Ground
C6
560pF
C1
10 nF
R3
1k
R2
1k
C4
100n
C2
100 nF
R4
100k
T1
BC548B
C5
47µ
R5
270R
2 x 32 ohm
parallel
+
1,5 Volt
-

Here is the address of the site for the slug-tuned radio.
http://www.kristalradio.nl/
Unfortunately the site is in Dutch and the kit is not available. However the photos give a clear picture of the how the parts are connected.

The inductance of the coil can also be altered by winding another coil and placing it near the first coil so that the magnetic field interacts with each other and changes the inductance of the circuit. This is called a VARIABLE INDUCTANCE TUNING COIL.
You can have one coil inside the other, two coils near each other or two flat coils side-by-side. Any two coils will interact with each other.

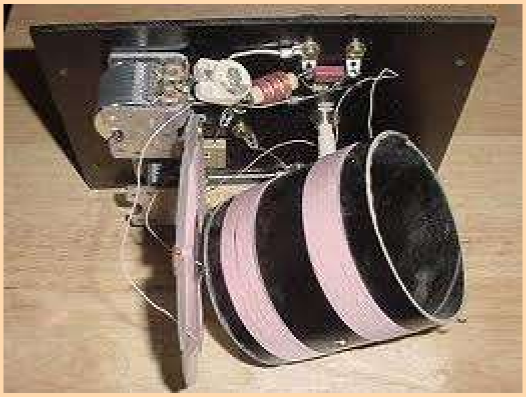

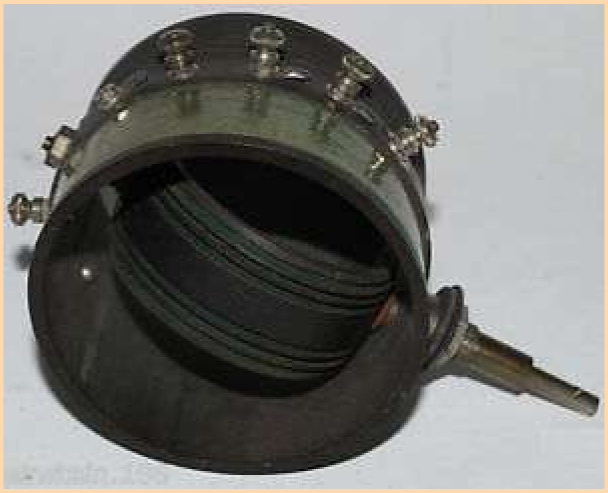

An Inductive TUNING COIL called a VARIOMETER

MAKING YOUR OWN FERRITE ROD ANTENNA

You can make your own FERRITE ROD ANTENNA by winding 60 to 80 turns of 0.25mm enamelled wire onto a 9mm ferrite rod or slab. If you wind it on a paper sleeve, you can move the coil along the rod to get the best performance. When the rod is slid out of the coil, the inductance changes considerably. However the inductance does change very slightly when the coil is moved along the rod.

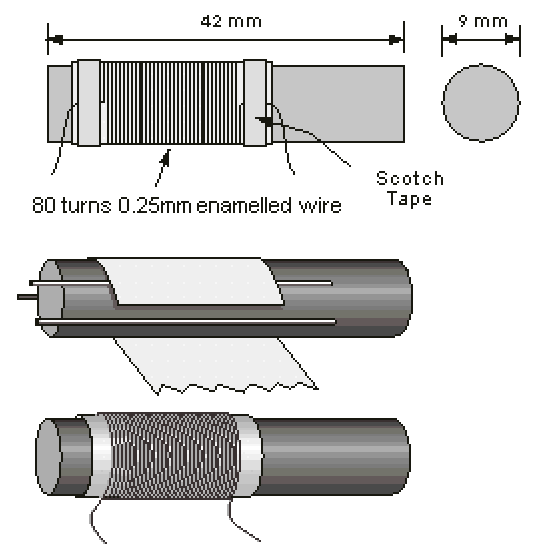

Make your own ferrite antenna

Now we come to the tuning capacitor::

THE TUNING CAPACITOR

The "C" in the "LC" TUNED CIRCUIT can be fixed or variable. When it is variable, it is called a TUNING CAPACITOR. The sheets of aluminium in the air tuning capacitor below are called PLATES and the moving plates are called VANES. The fixed plates make up the STATOR.
The space between the plates is AIR. The photo shows a single capacitor. If two capacitors are connected to the same shaft it is called a GANGED CAPACITOR.
The plates do not come fully out of mesh and that's why the capacitor has a minimum value. The maximum capacitance is when the plates are fully meshed. The odd shape of the plates is designed to produce a fairly constant increase in capacitance as the plates are engaged.

An air tuning capacitor:

Air Tuning Capacitor
(Variable Capacitor)

The capacitor can be made much smaller by using thinner vanes and placing plastic between the vanes. Plastic increases the capacitance about 3 times to 10 times.

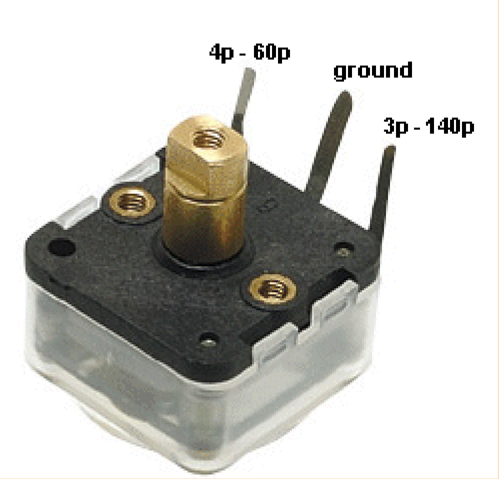

Tuning Capacitor as found in a pocket radio

The tuning capacitor can be replaced with a home-made equivalent that will work just the same.
You need:
4 sheets of aluminium foil (cooking foil) 10cm x 10cm.
4 sheets of thin cardboard 15cm x 20cm (cut A4 sheets in half).

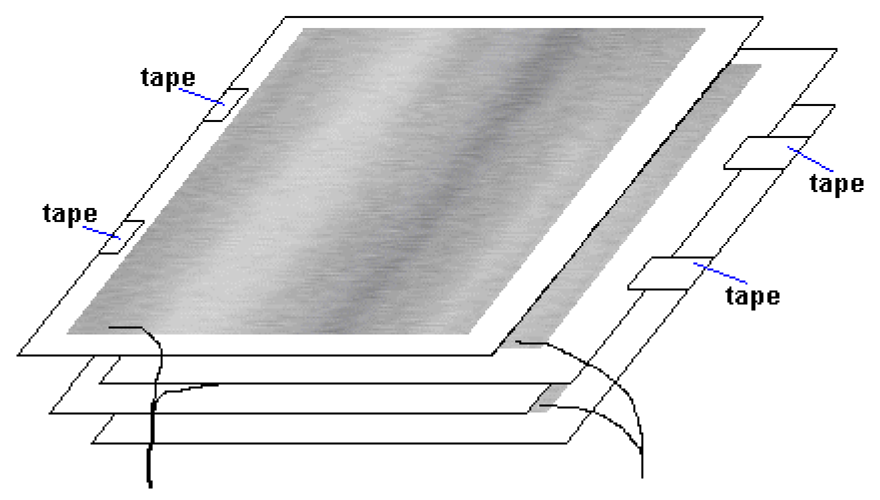

HOME-MADE CAPACITOR

Tape a sheet of aluminium foil to each sheet of cardboard with sticky-tape around all 4 sides. Take one strand of wire from a length of hook-up flex and sticky-tape the end to each sheet of aluminium to make good contact. Place 2 sheets on top of each other and move the top sheet slightly to the left and sticky-tape the edge so they don't move. Do this with the other two sheets but move the top sheet to the right. Now interleave the sets. Connect the wire from the first sheet to the third sheet. Connect the wire from the second sheet to the fourth sheet.
The cardboard (or paper) between the aluminium sheets increases the capacitance three times.
The capacitance decreases when the sheets are moved apart and the capacitance increases when the sheets are moved in. The capacitance also INCREASES when the sheets are squashed together such as when a book is placed on them.
You can also make a smaller capacitor by making each sheet smaller and using 6 sheets. You can then add a 100p or 220p in parallel with the home-made capacitor, to select the lower part of the band.

EACH CIRCUIT

Each circuit we describe in the following set of circuits is an improvement or advancement on the previous. We also offer a number of different types of aerial coils, amplifying stages and earphones. Some of the circuits use easy-to-obtain components and home-made equivalents for hard-to-get items. There will be something in this section for everyone to build.

In all radio circuits you will encounter **TWO MAIN PROBLEMS**:

If the FRONT END (the Coil and Capacitor) is loaded too much by the "pick-off" of the amplifying stages, you will only get one station.
If you get squealing or "motor-boating," try a different circuit and layout as the components you are using, plus the voltage of the battery, will need changing.
You cannot always increase the voltage of the supply and get a louder output. Sometimes the increased voltage will stop the circuit working or it may introduce too much gain that the circuit starts to squeal.
The Radio IC (ZN414) DOES NOT WORK on a voltage above 1.5v and some of the transistor circuits completely stop working with a higher voltage. This has to do with the biasing arrangements and if the circuit is designed for a low voltage, you need to keep to the suggested voltage and experiment with a slight increase in voltage and see what happens.
Building a radio is not easy as the enormous amount of amplification of the combined stages creates a feedback loop via the power rail that sets the circuit into oscillation. This effect gets worse with a higher supply voltage and we will explain this further with each of the circuits.

MAKING A CRYSTAL SET

You can buy a **CRYSTAL SET** kit (see the photo of the kit, above) or the individual components (a kit is the cheapest) or use the replacement for the **FERRITE ANTENNA COIL** (16 turns to 20 turns on a 150mm biscuit tin) and/or the **TUNING CAPACITOR** made from aluminium foil and cardboard sheets.
You will need an outside antenna and an earth (such as a water tap or the frame of your soldering iron) to pick up the radio stations.
If you cannot put up an outside antenna, you will need to add one or more amplifying stages and this will allow you to reduce the length of the antenna and increase the volume of the audio.

ADDING AMPLIFYING STAGES TO A CRYSTAL SET

You can add two different types of amplifying stages to a crystal set.
You can connect amplifying stage(s) to the FRONT END and these will be designed to put less load on the front end so the sensitivity and selectivity increases. These stages work at the frequency of the radio signal and they are called RF STAGES (Radio Frequency Stages).
You can build these stages out of individual components or use a chip called a RADIO CHIP or RADIO IC (integrated circuit) for less than $2.00.
The chip contains 5 stages of amplification and these are RF stages (or RF AMPLIFYING STAGES) and the concept is called TRF. (Tuned Radio Frequency).
It is not easy to get this type of amplifier working because the stages produce a very high overall gain and you get a lot of "motor-boating" and squealing if the gain is not controlled. The gain must be reduced when a strong signal is being passed through the circuit because a strong signal will produce a large output and this will be so large that some of the waveform will find its way to the front of the amplifier via the power rail and start to be amplified again. To prevent the output getting too large, the circuit has a negative feedback line - called the AGC line - Automatic Gain Control.
It would be very difficult to reproduce these 5 stages of amplification with discreet components and that's why it is best to use an IC.

The next stage is a DIODE to convert the RF (Radio Frequency) to AF (Audio Frequency). This can be done with the diode-characteristics of a base-emitter junction in a transistor and we will show the alternatives.

Any stages after the diode are AUDIO STAGES or AUDIO AMPLIFIER STAGES.
The main job of the AUDIO AMPLIFIER is to increase the DRIVE CAPABILITY.
In other words, increase the current capability of the circuit for an 8 ohm speaker or 8ohm earpiece (or 16 or 32 ohm).
This is a very difficult thing to do and requires at least 2 stages.
The LOAD you can put on a Crystal Set must be 10,000 ohms or higher. (if you put a lower resistance (impedance) on the output, you will load the FRONT END and reduce its ability to separate the stations.
That's why a crystal earpiece is normally used with a crystal set. It puts almost NO LOAD on the circuit.
If you put a load on the circuit the result will be only one or two stations across the whole dial and only the most powerful station will be received.

If you don't have a crystal earpiece, you will have to use an 8 ohm earpiece. This will require an IMPEDANCE CONVERTING CIRCUIT of 1,250:1
This is a simple way of saying we want the 8 ohm earpiece to appear as 10,000 ohms to the

crystal set.
To produce an overall gain of 1250, we need two stages of amplification.
If a transistor has a gain of 70, it will it will produce an impedance conversion of 70 times. This is a realistic value. Transistors with a gain of 200 will have a gain of about 70 when fitted to a circuit.
This means the other transistor needs to have a gain of about 20 and that is easy to achieve.

ADDING AMPLIFYING STAGES TO THE FRONT OF A CRYSTAL SET

Adding stages to the front of a crystal set are called RF STAGES (Radio Frequency Stages) because they amplify the RADIO STATION SIGNAL.
It does not matter if you amplify RF signals or AF signals. The result is the same.
The only difference is this: The frequency of RF signals is much higher (1,000 times higher) and the coupling capacitors can be much smaller.
This allows an RF amplifier to be built into an IC - called a Radio Chip.
One of the most popular Radio IC's is ZN414 or YS414. This chip has been copied by other manufacturers as: MK484, TA7642 and LMF501T.
All the chips are the same but the pinout is different.
These chips work on a 1.5v supply and if the voltage is increased above 1.5v, the gain of the stages increases to a point of total distortion.
To prevent strong signals producing distortion on 1.5v supply, the output is passed back to the input via a 150k resistor. This feedback line is called the AGC (Automatic Gain Control).
The chip contains 5 stages of amplification plus a stage that converts the RF signal to AF (Detector Stage). This means the signal diode in a Crystal Set is not needed.

Here is the circuit of the TA7642 Radio Chip. It performs the same as the ZN414 Radio Chip.

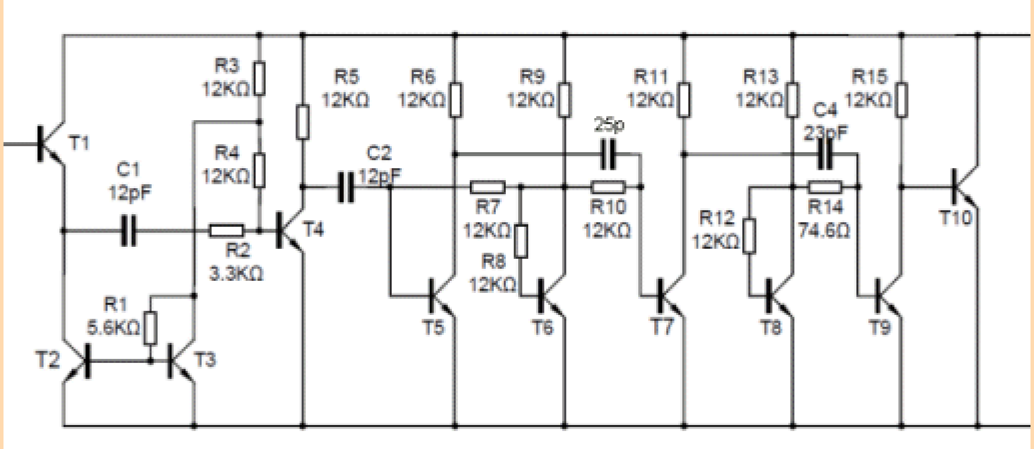
R3
12KΩ
R5
12KΩ
R6
12KΩ
R9
12KΩ
R11
12KΩ
R13
12KΩ
R15
12KΩ
C4
23pF
25p
T1
C1
12pF
R4
12KΩ
C2
12pF
R7
12KΩ
R10
12KΩ
R12
12KΩ
R14
74.6Ω
T10
T4
R2
3.3KΩ
R8
12KΩ
R1
5.6KΩ
T5
T6
T7
T8
T9
T2
T3

Here is the BLOCK DIAGRAM of the ZN414 Radio Chip:

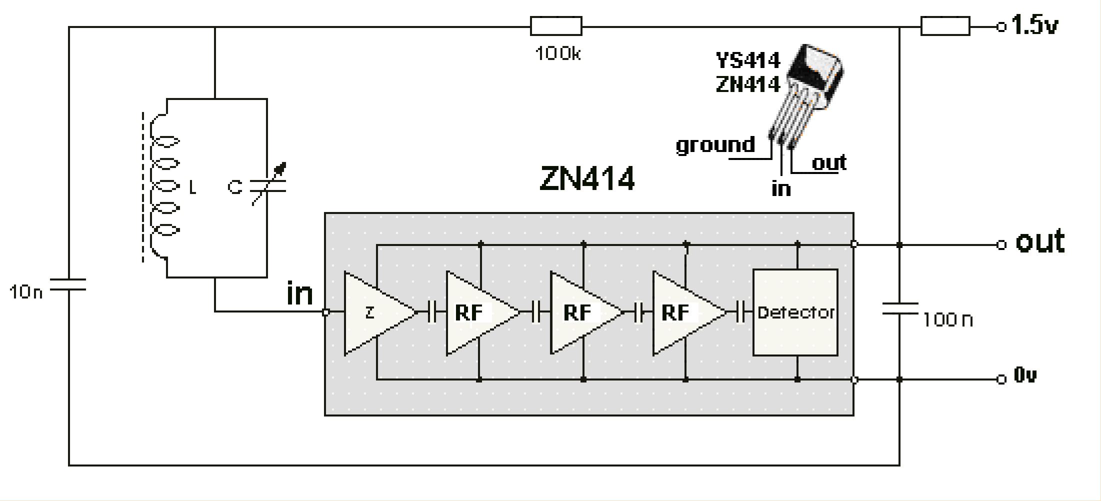
100k
1.5v
YS414
ZN414
ground
out
in
ZN414
L
C
10n
in
Z
RF
RF
RF
Detector
out
100n
0v

The ZN414 chip can be purchased from Talking Electronics for $1.00 plus postage

USING THE ZN414 RADIO IC

By using the ZN414 radio IC (or any if the equivalents) you can create a POCKET RADIO to drive a headphone or speaker.
But it is not easy to use the chip. The main problem is receiving the strong signals without producing distortion and then being able to pick up the weak stations.
A fixed 100k feedback resistor does not provide adequate control and a TRF radio has limited capabilities.
That's why radio manufacturers make SUPERHETRODYNE receivers. Even though they are more complex, the result is far superior.
However a simple TRF set can be made with the Radio IC and a few stages of audio amplification.

The following circuit uses just the Radio IC and a crystal earpiece or the home-made earpiece described above:
You can use a home-made FRAME ANTENNA or a home-made FERRITE ROD ANTENNA and a home-made VARIABLE CAPACITOR.

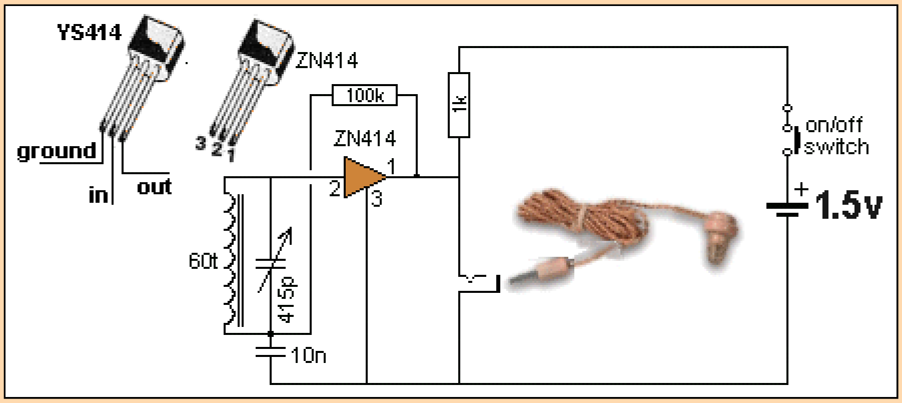

The Simplest ZN414 Radio

Connecting the ground (0v rail) to the frame of your soldering iron or a water tap will increase the output volume. The circuit above shows a Crystal Earpiece. Using a Crystal Earpiece may require adding a 10n across the earpiece to improve the output volume. The substitute Piezo Earpiece is effectively a 20n capacitor and an additional capacitor is not needed.

2 TRANSISTOR RADIO

Here is a simple 2-Transistor radio.
The secret to its performance is the 7 turn "pick-off" from the FRONT END (the TUNED CIRCUIT).
The ratio of 7 turns to 60 turns means a small percentage of the voltage generated in the tuned circuit is passed to the transistor. Thus it puts a small load on the TUNED CIRCUIT.
I don't want to go into any mathematics. The turns ratio is 60:7 = 8 but the effect of the 7 turns "pick-off" has an effect called the IMPEDANCE EFFECT and this is the SQUARE OF THE TURNS RATIO. Thus the IMPEDANCE EFFECT is 8 x 8 = 64. This means the "pick-off" (the LOADING EFFECT) is just a few percent. The front end can produce voltages as high as 500mV because a crystal set can produce a voltage high enough to pass through a diode (350mV) and have sufficient to drive a crystal earpiece.
Even though the front end has a "step-down" ratio, the voltage out the 7 turns will be sufficient to drive the first transistor.
The "transformer" does 2 things: It reduces the loading on the tuned circuit ENORMOUSLY and it produces an output with a higher current than is circuiting in the front end. Even though the transistor is turned ON and biased by the 33k, it is classified as a low-impedance load as far as the front end is concerned and the input signal has to be accompanied by a certain amount of current, otherwise the transistor will not respond to the voltage. The 7-turn "pick-off" is able to provide this current.
Both transistors are biased ON via the 33k base-bias resistors and thus the first transistor responds to the slightest millivolt signal.
This circuit was tested and had the same performance as the **Simplest ZN414 Radio Circuit** above. It can be operated on 1.5v to 6v and the strongest stations tend to overload on 6v. A short antenna is needed.

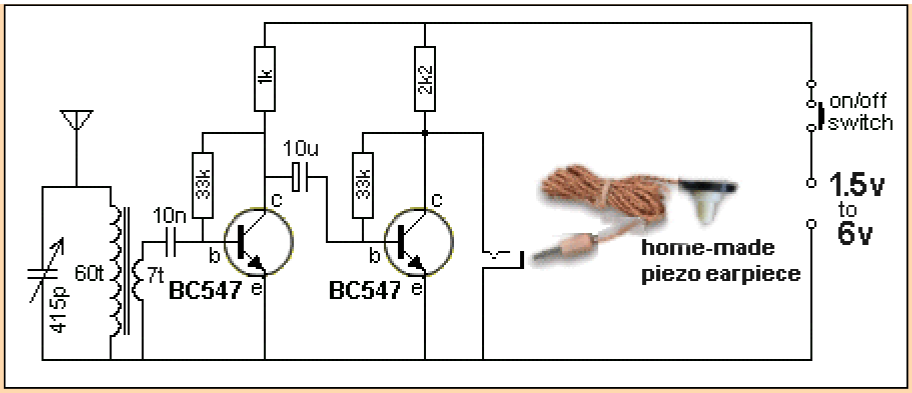
1k
2k2
on/off switch
1.5v to 6v
10u
33k
33k
10n
c
c
b
b
60t
7t
415p
BC547 e
BC547 e
home-made piezo earpiece

SIMPLEST 2-TRANSISTOR RADIO
using a very-high-impedance earpiece

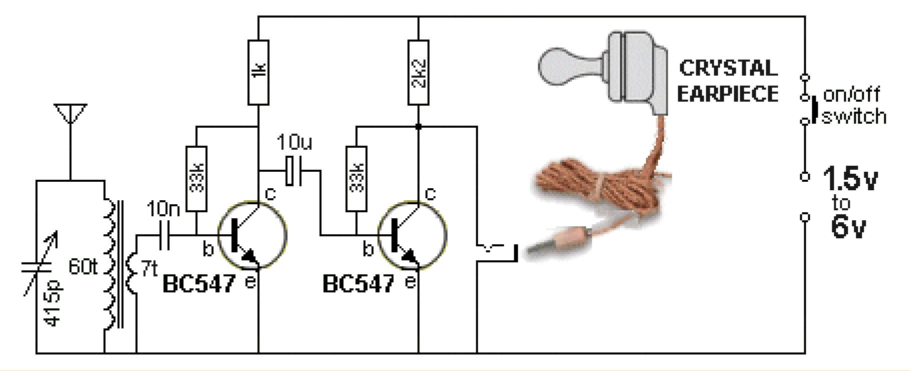
1k
2k2
CRYSTAL EARPIECE
on/off switch
1.5v to 6v
10u
33k
33k
10n
c
c
b
b
60t
7t
415p
BC547 e
BC547 e

SIMPLEST 2-TRANSISTOR RADIO
using a crystal earpiece

ADDING AN IMPEDANCE MATCHING STAGE

You can add an IMPEDANCE MATCHING STAGE to the output of the circuit above so a low-impedance earpiece can be used.
We call it an IMPEDANCE MATCHING STAGE because this is the correct technical term. It is an AMPLIFYING stage but it amplifies the CURRENT because the second transistor cannot drive an 8 ohm LOAD. 8 ohms is a very low resistance and if it is connected directly to the second transistor, the output will be almost zero.
The reason for this is covered in our discussion: The Transistor Amplifier.
This stage will not increase the volume but simply match the 8 ohm load to the circuit above.
It is very difficult to connect a LOAD to this type of circuit because it will take more current from the battery and cause the supply voltage to fluctuate. These fluctuations will be passed to the first stage and cause variations in the signal. This will be amplified by the first and second transistors in the form of a low-frequency buzzing called **MOTOR-BOATING**.
The only way to reduce or remove this noise is to add an electrolytic across the power rails and reduce the supply voltage. The third transistor simply takes the waveform on the output of the second transistor and delivers it to the earphone with a higher current. It is called an IMPEDANCE MATCHING STAGE as it effectively increases the 8 ohm load by a factor of about 100.

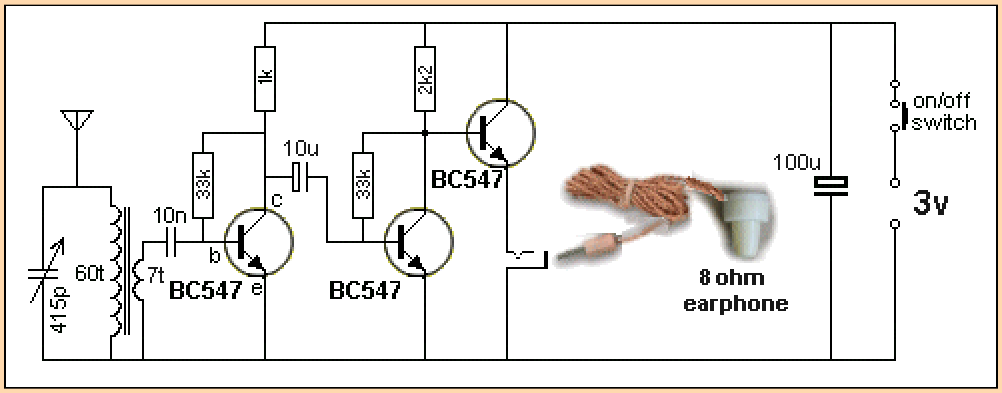
1k
2k2
on/off switch
100u
3v
10u
33k
33k
BC547
10n
c
b
60t
7t
415p
BC547 e
BC547
8 ohm earphone

3-TRANSISTOR RADIO with buffer stage for 8 ohm earpiece

The 3rd transistor converts the 8R to about 800R

You can use 16 ohm, 32 ohm or 64 ohm in place of the 8R earpiece and these will give better performance as they will take less current and improve the stability of the circuit.
Low-impedance earphones create "motor-boating" due to the peaks of current and this can be very hard to fix.

A 2-TRANSISTOR RADIO with REGENERATION

The next stage in our discussion to get better performance is a feature called **REGENERATION**.
Regeneration sends a small output signal back to a previous stage in the form of POSITIVE FEEDBACK to INCREASE the original signal. The signal on the emitter of the first transistor is the same amplitude as the signal entering the base but the FRAME ANTENNA has a turns ratio of 5:15 and this increases the signal on the receiving section of the antenna by up to 3 times. But we want the returning signal to be just above the amplitude of the receiving signal and so a resistive adjustment (attenuator) is provided on the emitter to deliver just the right amplitude.
This has the effect of increasing the amplitude on the TUNED CIRCUIT and is just like reducing the load on the circuit.
As we have mentioned above, when the tuned circuit is lightly loaded, it will pick up a station at the exact frequency of transmission and if the dial is changed slightly, the station will disappear. This quality is called SELECTIVITY.
At the same time, the Tuned Circuit will pick up weak stations and this is called SENSITIVITY.
The quality of a receiver depends on the loading of the TUNED CIRCUIT.

Here is the original circuit from Elektor Magazine with the prototype made on matrix board and fixed to a base-board with a frame antenna made from two sticks of wood. The photo shows a speaker but the output is so low that you really need headphones.

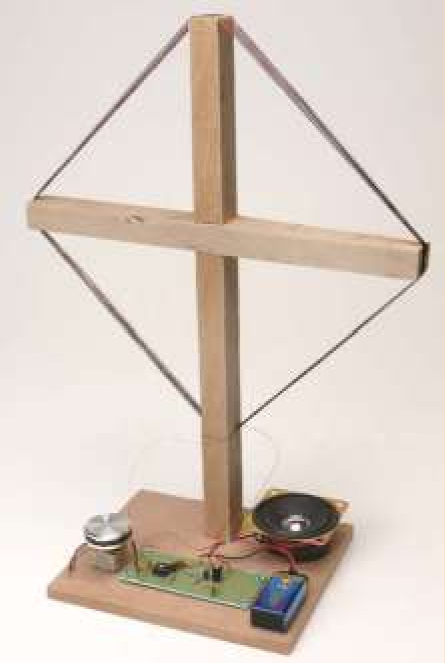

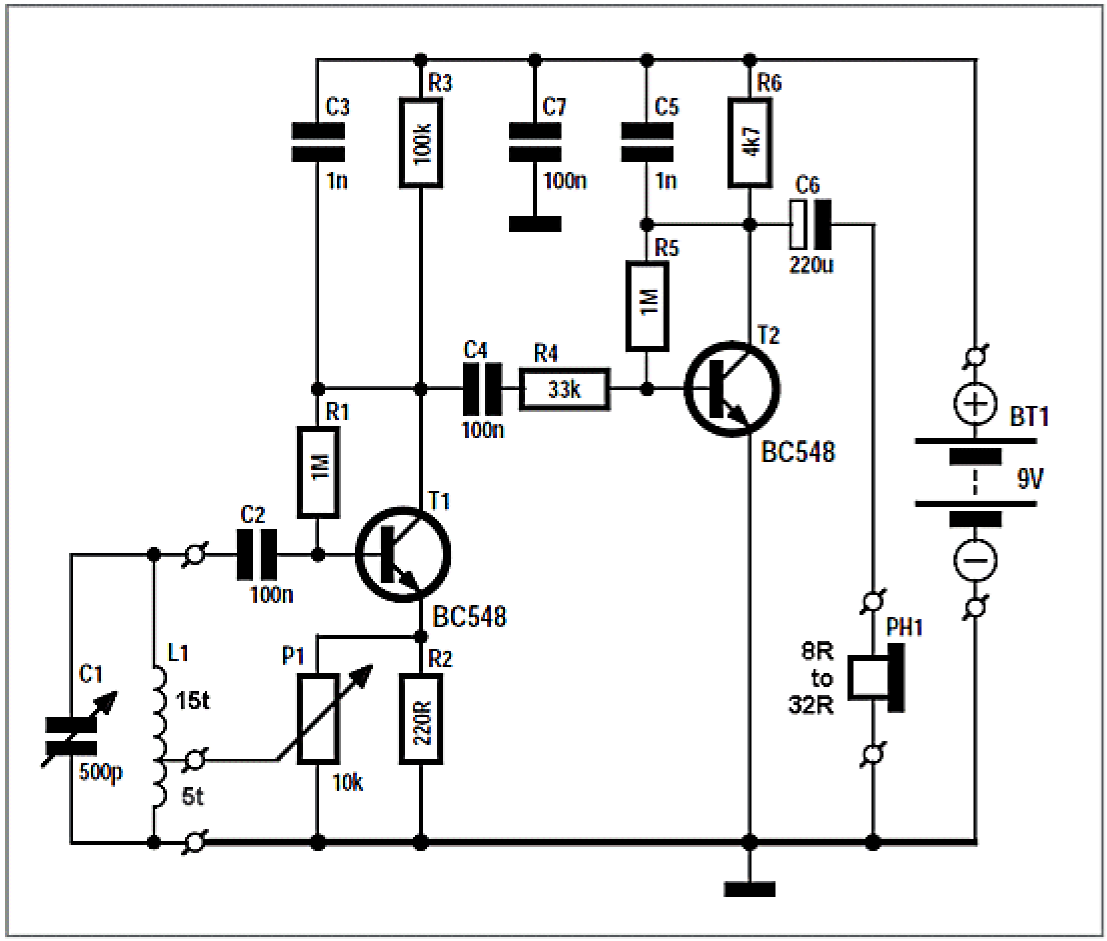
C3
1n
R3
100k
C7
100n
C5
1n
R6
4k7
C6
220u
R5
1M
C4
100n
R4
33k
T2
BC548
BT1
9V
R1
1M
T1
C2
100n
BC548
C1
500p
L1
15t
5t
P1
10k
R2
220R
PH1
8R
to
32R

2 TRANSISTOR REGENERATIVE RADIO from Elektor Magazine

The circuit is very complex and the output will be very low as the circuit cannot drive a low-impedance earphone via a 4k7 load resistor. The 4k7 resistor is actually driving the speaker (the transistor is simply discharging the 220u). The 4k7 only allows 32/4700 x 9 = 61mV to appear across the earphone - a very poor result.
The skill of designing a transistor stage is covered in our comprehensive eBook: The Transistor Amplifier and you wont make a mistake like this !!!
The circuit above can be simplified and we can add the REGENERATIVE feature to our **Simplest 2-Transistor Radio** circuit:

Our circuit uses a 15 turn circular FRAME ANTENNA 15cm diameter and a 5 turn REGENERATION coil.

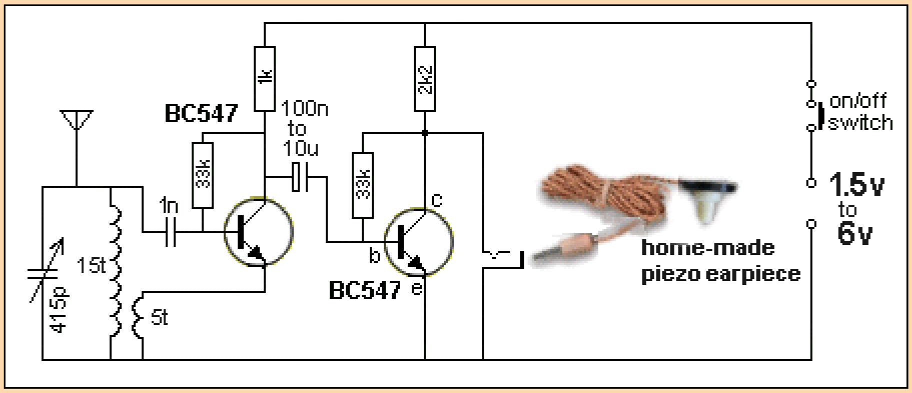
BC547
1k
100n
to
10u
2k2
on/off
switch
33k
33k
1n
c
b
e
BC547
15t
5t
415p
1.5v
to
6v
home-made
piezo earpiece

2 TRANSISTOR RADIO with REGENERATION

The regeneration coil is brought near the main coil and as it gets closer you can hear the audio get louder. If this does not happen, turn the coil around.
Early radios used this technique and the operator had to adjust the coil by hand. No-one minded because radio was a fascination and the simplest radio cost more than a weeks wages. To listen

to a broadcast through headphones was an amazement and listeners would sit all night with headphones listening to music.
This is very fiddly and by adding an extra buffer stage, we can use a Frame Antenna with a very clever "pick-off" that does not load the front end. This gives the circuit very good sensitivity and selectivity without regeneration.

3 TRANSISTOR RADIO

Here is our final design for the simplest self-contained 3-Transistor Radio using our home-made Tuning Capacitor and 250mm x 350mm Frame Antenna. It picks up the local stations and drives a low-impedance earphone or set of earphones (from a mobile phone).

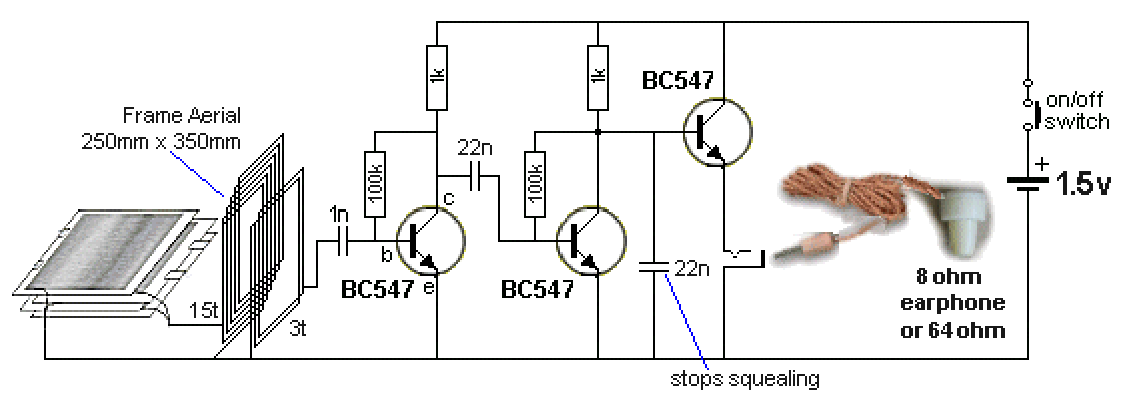

3-TRANSISTOR RADIO

The circuit performs very well and uses readily-available components. The 22n across the output is essential to stop squealing.
The secret to sensitivity and selectivity is the turns-ratio on the Frame Antenna. The 3-turn "pick-off" puts very little load on the front end and this allows the stations to be tuned with our home-made Tuning Capacitor.
The circuit contains all the features we have discussed above and only needs a 1.5v supply.
Build this circuit before you buy any expensive tuning capacitors, IC's or ferrite slab antennas as you will not get any better results.
This is called a TRF circuit and because the stages operate at Radio Frequency or Audio Frequency. Due to the high amount of amplification, the circuit can start to squeal (feedback, motorboat) due to the layout.
You may need to shorten or lengthen the leads or move the parts slightly - it's that critical.
However the result is a portable radio that needs no earth and will pick up the strong stations.
You can try connecting the 0v rail to the metal part of a soldering iron to increase the number of stations.

LOADING

The whole success of picking up a radio station is the RECEIVING CIRCUIT. The receiving circuit is the coil and the signal in the air (from the radio station) must go down the centre of the coil.
It cannot pass over the top or the bottom of the coil. Only the signal that goes down the centre of the coil is received.
As you can see, the centre of the coil is not very big and it is amazing that the signal can pass down the centre. But it does, and that is the only signal that will be amplified.
This signal is passed to the capacitor and we have explained how the signal is gradually increased and increased in amplitude until it is as large as 500mV. The signal from the radio station may be as small as a few millivolts, but as it keeps pushing the "swing" back and forth, the amplitude get larger and larger.
If you put your finger on the "swing" you will prevent it get larger and larger and it only requires the slightest touch of your finger to prevent the swing gaining full amplitude.
In electronic terms, your finger is called LOADING THE CIRCUIT and since we have to pass the signal to further stages of amplification, we need to "tap" or "load" or "pick-off" a signal.
The aim is to load the circuit as least as possible because the actual energy entering the circuit is very small.
In fact, this is all the energy we can remove as that is all the energy entering it.
Because a very small amount of energy is entering the "front-end" we classify it having a very

high impedance. It is very difficult to provide a value of impedance for this circuit because impedance has the term "Z" and the circuit is operating a very high frequency so resistance values are not the same as impedance values.

The actual resistance of the circuit is ONE OHM but the impedance is more like 10,000 ohms to 100,000 ohms.

We can explain its high impedance if we put a 100,000 ohm resistor across the circuit. The waveform will be reduced very slightly. If we put a 10,000 ohm resistor across the circuit, the signal will be reduced a reasonably large amount. If we put 1,000 ohms across the circuit it will stop working.

This means a load of 100,000 ohms will have the least effect.

In a crystal set, the diode creates NO LOAD until a voltage of 350mV is reached. It then passes excess voltage to a crystal earpiece that has a very high impedance. That's why a crystal set will produce a good output. The LOADING is very small.

When a transistor is connected to the TUNED CIRCUIT, it starts to put a load on the circuit after 600mV and this load is VERY HIGH. The "resistance" of the base-emitter junction is about 1k and the signal will find it very difficult to rise above 600mV because the incoming energy is not sufficient to increase the voltage.

Adding a capacitor between the base and the front end allows the transistor to be self-biased and get a turn-on voltage of about 600mV from a base-bias resistor.

The FRONT END is now separated from the transistor and ANY voltage it is producing will be passed to the transistor via the capacitor.

Whereas, with the crystal set, the first 350mV could be produced without any loading, the circuit is now loaded AT ALL TIMES.

This means we have to load the circuit as lightly as possible to be able to pick up individual stations.

The only way we can do this is to use a capacitor of the smallest practical value and this has to be worked out by trying different values. If the value is too small, the transistor will not detect a small signal. If the value is too large, the circuit will stop working.

Values such as 1n, 10n and 100n are suitable.

Values such as 1u, or 10u will be too large.

The CRYSTAL SET loading and a transistor load are completely different.

The transistor loads the front end ALL THE TIME and that's why you need to use a transformer or other ways to reduce the loading. Sometimes a Field Effect Transistor is used as it puts almost no load on the front end.

2-4-2016

Histats.com
Vis. today 10

Easy To Use, All in One, CRM.
1.5 Million Insightly CRM users worldwide.
insightly
TRY FOR FREE

Easy To Use, All in One, CRM.
1.5 Million Insightly CRM users worldwide.
insightly
TRY FOR FREE

Capacitor Kit
The World's Only Professional Kits Low Price! Fast Shipping; In Stock.
analogtechnologies.com

Capacitor Kit
The World's Only Professional Kits Low Price! Fast Shipping; In Stock.
analogtechnologies.com

CRYSTAL RADIOS AK (/)

Crystal Set with Output Meter

I felt that it was not easy to evaluate sensitivity or level of output of a crystal set because human ears are not objective enough. There are several articles on the net about showing the output level as voltage or current. If you connect a voltmeter at the phone jack, the input impedance of the meter will affect the balance between the set's output impedance and the phones input impedance.

I found Mr. Billy Cheung in Hong Kong published an article about a crystal set with an output indicator (http://theradioboard.com/rb/viewtopic.php?f=2&t=6082). (http://theradioboard.com/rb/viewtopic.php?t=6082) It also seemed very nice set to build, so I did. I bought a DC current meter with -50 to +50 uA range at AliExpress. At that time it was the only model available. Since the meter is large, 10cm x 8cm, the case needs to be fairly big, 21cm x 10cm x 7cm.

There are three modifications (additions) to the original. I added a switch to make one of two gang variable capacitor, C1, a part of the tank circuit or an antenna impedance matching capacitor. A switch was added to increase the number of turns for the regeneration in case of MOSFET detection. The last change was to add a switch to increase capacitance of C2 by adding the second section of the air variable capacitor.

Because the case was large, wiring was easy. Therefore, it did not take so long to build the set, except making a few corrections on my stupid wiring errors, as usual.

Performance:

My antenna is about 12 m long wire hanging in my garden, and the ground is a 50 cm copper bar buried in the garden soil. In mode 1, sensitivity is good. We have three local stations, JOPK, JOPB, and JOVR, transmitted at 12 - 16 km distance with 10 kW power. After fine tuning, following values for output in uA were obtained. The values did not change by using different phones or amplifiers.

	1N60	3DQ
JOPK	6.3	8.5
JOPB	3.8	4.5
JOVR	3.8	5.0

Human feeling of loudness for JOPK cannot find any difference between the two detectors. I have no idea if the numbers above are high or low in my environment. No other stations were heard. Because of the capacitor layout in the output circuit, the output sound does not include high audio frequencies and, therefore, is soft, although I have not checked the frequency distribution.

Selectivity of 3DQ is good, while 1N60 gives broad tuning. However, this difference may not be caused by the detectors but by the circuit difference. In case of the diode, primary tank circuit has no tap, and the FET's source gets the signal from the tap.

In mode 2, over-all performance is not satisfactory. Sensitivity is not quite as good as mode 1, and selectivity is even worse. If you set the taps near the top (tap #2-4), you can hear several SW stations in China and Korea as well as domestic ones, but they can be picked up everywhere on the dial. Therefore, there is no way to know the frequency which the set is tuned. This may be caused by the unused part of the coils, lowering the Q, as Mr. Broesel pointed out in the original post by Mr. Cheung (http://theradioboard.com/rb/viewtopic.php?t=6082).

September 30, 2016
October 1, 2017 (last edited)

Crystal Set with Output Meter

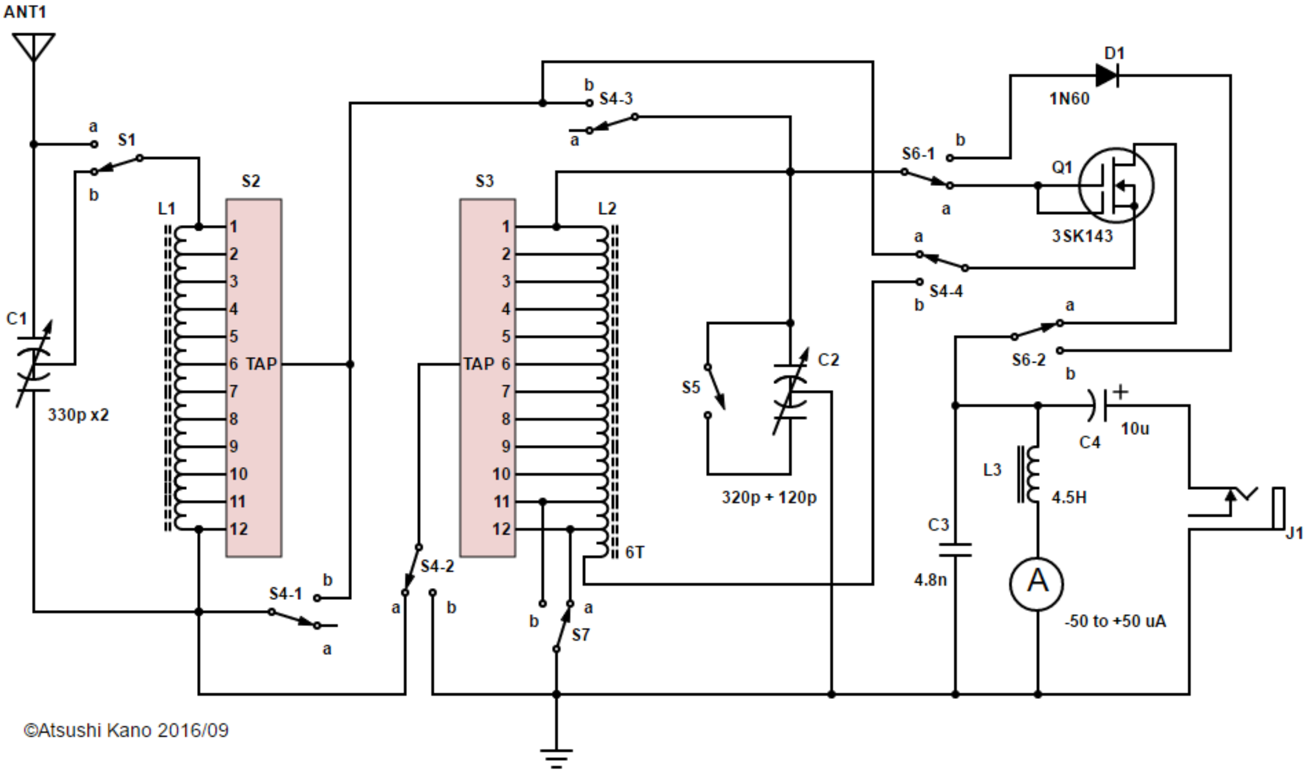
ANT1
S1
a
b
C1
330p x2
L1
S2
1
2
3
4
5
6 TAP
7
8
9
10
11
12
S4-1
a
b
S3
TAP 6
L2
6T
S4-2
S4-3
S7
S5
C2
320p + 120p
S6-1
S4-4
S6-2
D1
1N60
Q1
3SK143
C3
4.8n
L3
4.5H
C4
10u
A
-50 to +50 uA
J1

©Atsushi Kano 2016/09

L1: R40C1 ferrite toroid with 250/46 litz wire. 315 uH. 10 taps at every 6 turn.
L2: same as L1 + 6T for regenaration.
L3: one side of 1:1 transformer. See original article by Cheung.
S1: 1/2 of C1 used (a) as a tank, (b) as an ANT impedance matching.
S2: Tap selecting rotary switch for primary tank.
S3: Tap selecting rotary switch for secondary tank.
S4: DP4T. (a) mode 1, (b) mode 2. See text.
S5: Add 120pF to C2.
S6: Detector selection. (a) 1N60, (b) 3SK143.
S7: (a) original, (b) add 6 turns for regeneration.
Q1: 3SK143 (3DQ). Two gates are connected.
C1: Alps two-section poly-variable capacitor.
C2: Alps two-section air variable capacitor.

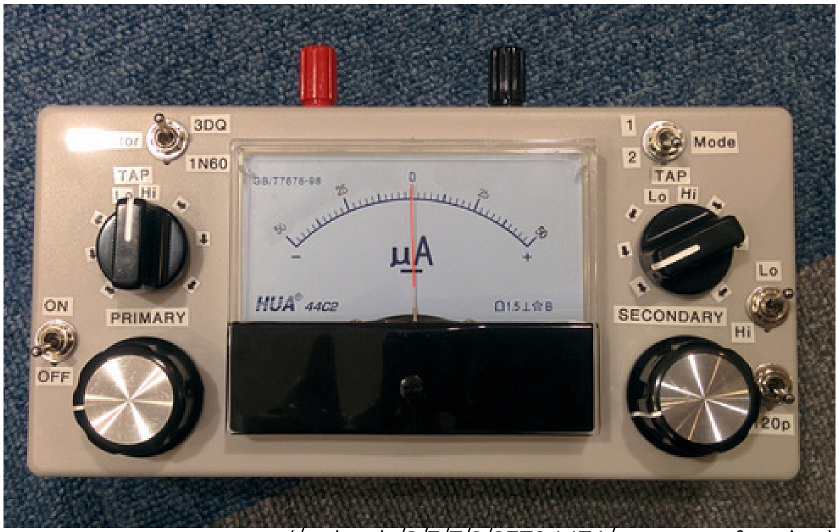

(/uploads/8/3/7/9/83794474/cs_meter_face.jpg)

Face of the set

The case is TAKACHI TW11-8-22G, W110 x D220 x H75.3mm. You can tell how large the meter is.
I put the rotary switches for tap selection upside-down by mistake, so selecting the taps is confusing a little.

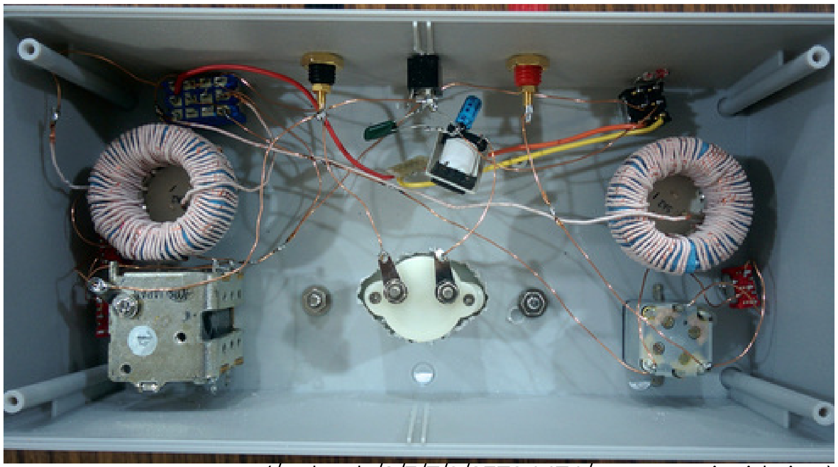

(/uploads/8/3/7/9/83794474/cs_meter_inside.jpg)

Inside

Since few parts were used except switches and variable capacitors, inside the case is almost vacant. The DP4T toggle switch was the most expensive among connectors and switches.
The 3DQ is hidden behind the small universal board, while the 1N60 is connected to the terminals of the switch.

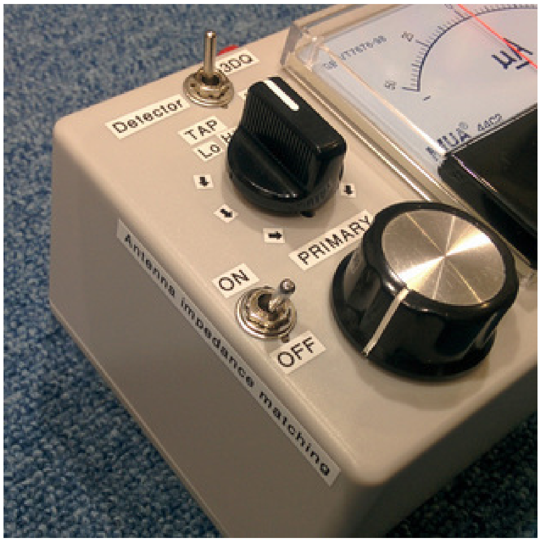
Detector
3DQ
TAP
Lo Hi
PRIMARY
ON
OFF
Antenna impedance matching
μA

(/uploads/8/3/7/9/83794474/cs_meter_left.jpg)

The switch to separate/combine the sections of the variable capacitor, C1 is shown.

ANT
PHONE
GND
3DQ
1N60
μA

(/uploads/8/3/7/9/83794474/cs_meter_top.jpg)

Phone jack and terminals are placed on the top (side) of the case.

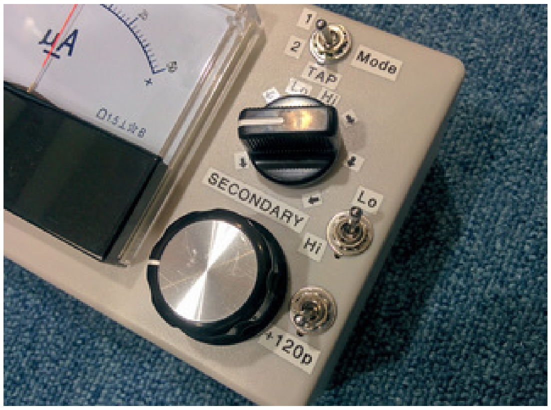
μA
1
2
Mode
TAP
Lo Hi
SECONDARY
Lo
Hi
+120p

(/uploads/8/3/7/9/83794474/cs_meter_right.jpg)

Mode switch and other switches for the secondary tuning are placed on this side.

CRYSTAL RADIOS AK (/)

Crystal Set with Q Multiplier

I found a very interesting YouTube video, showing an application of a Q multiplier to a crystal radio.

The base set is the one published by Mr. (http://www.geojohn.org/Radios/MyRadios/CRKit/CrystalKit.html)John Fuhring (http://www.geojohn.org/Radios/MyRadios/CRKit/CrystalKit.html). It uses two toroidal cores for antenna tuning and station tuning. Linking the two toroids are done by few turns of coupling coils on both of the cores. I used the same cores (T80-15 (http://www.amidoncorp.com/t80-15/)) as in the original article. The linking coils are wound by 24 AWG wire, connected to a 100 ohm variable resistor which attenuate the signal to the secondary coil.

I used 40/60 litz wires for the tuning coils, which was suggested in the Fuhring's article. Another difference between the original and mine is to have an additional tap for the Q multiplier.

The Q multiplier I used was shown in the video by Mr. VK3YE in YouTube (https://www.youtube.com/watch?v=4ix5gNrJy2E). Capacitance between the hot end of the tank circuit and the base of transistor is critical. If it is too large, say 100 pF, it rings all the time. If it is too small, it does not work. At first, I twisted the wire from Q1 to the one from C2, as suggested in the video, the connection was too weak. So, I twisted the two wires as shown in the photo. My guess of the capacitance is about 15 pF.

Performance of the base set is not bad, but not great. There must be many possible reasons that mine may not be as good as the original, although I cannot think of any. The sensitivity is a little less than the Loose Coupler set (LC). The selectivity is as good as the LS, but this set requires no special tuning procedure, which is needed in tuning with the LC. I could listen three local stations very well, but no other stations could be heard even at night with an audio amplifier connected to the output.

With the Q multiplier ON, the sensitivity increased dramatically. Several stations in Tokyo (200 km away) could be heard after tuning with patience. The variable capacitor, especially the secondary for station tuning, needs to be adjusted very little by little, and it also changes its capacitance with my hand. If I assemble the set again, I would 1) use air variable capacitors, instead of poly film ones, 2) add vernier drive mechanism to the secondary variable capacitor, and/or 3) add a variable capacitor with a small capacitance, e. g. 10-20 pF, parallel to C2 for final adjustment.

Fortunately, I have an old SONY ICF-7600D (http://stephan.win31.de/sony76-4.htm) BCL radio, so I used it for making its tuning very easy for specific stations. First, set the receiving frequency of ICF-7600D to the frequency of the station you want to listen and leave it on. Then rotate the tuning dial of the set and find the point which causes the oscillation noise on the 7600D. This means that the set is tuned at the same frequency as the 7600D is receiving, which is digitally displayed on it. Remaining work is not so difficult if the signal is strong enough. Final tuning should make the station audible. I have received 5 stations with 50-100 kW power within the 300 km range in addition to the local ones. FYI: Transmitting power of radio stations in Japan is not as strong as other countries, in general.

Selectivity of the set becomes very sharp with Q multiplier, but since there are no overwhelming stations in my area, I have no trouble separating stations without it.

Playing with this machine is fun, but without Q multiplier, there is not much left to do. With the Q multiplier, it is exciting to listen distant stations, but I feel guilty for using it, because, although it is not directly amplifying the signal like regenerative radio, it is still with a help from a battery.

August 15, 2016

On the night of September 29, 2016, the following stations were caught when Q multiplier was ON. All stations were listened through an audio amplifier.

Frequency, kHz	Call sign	Location
639	JOPB	Shizuoka
711	HLKA	Incheon, Korea
747	JOIB	Sapporo
774	JOUB	Akita w/QRM
810	AFN	Wako w/QRM
828	JOBB	Osaka
846	JOCP	Koriyama
882	JOPK	Shizuoka
954	JOKR	Tokyo
972	HLCA	Danging, Korea
1008	JONR	Osaka
1089	JOHB	Sendai
1134	JOQR	Tokyo
1170	HLSR	Gimje, Korea
1242	JOLF	Tokyo
1314	JOUF	Osaka
1404	JOVR	Shizuoka
1566	KLAZ	Jeju, Korea

All stations are in Japan except those in Korea.

It seems sensitivity of the set is not good when frequency is below 700 kHz. No stations near 1400 kHz were heard because of the strong signal at 1404 kHz.

October 2, 2016

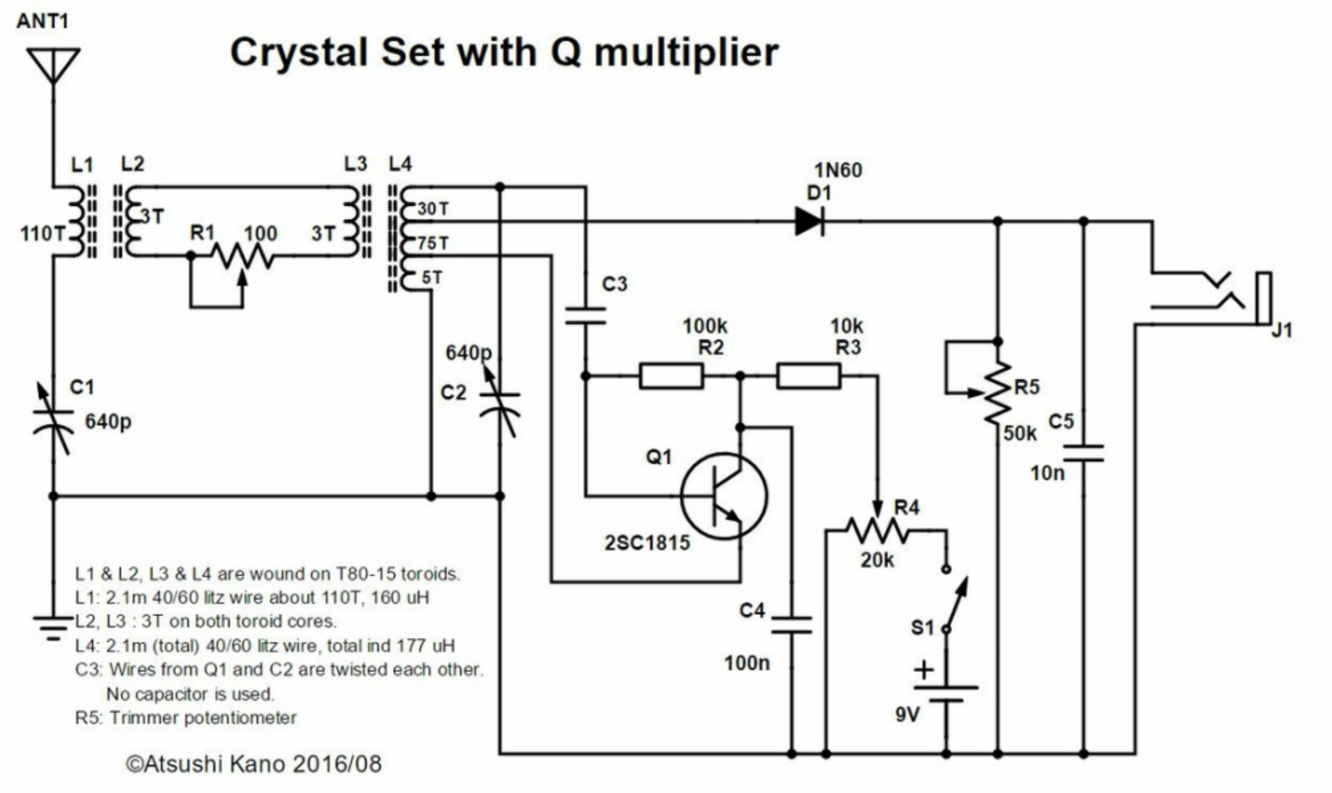
ANT1
Crystal Set with Q multiplier
L1
L2
L3
L4
110T
3T
R1
100
3T
30T
75T
5T
1N60
D1
C3
100k
R2
10k
R3
640p
C2
C1
640p
R5
50k
C5
10n
J1
Q1
2SC1815
R4
20k
S1
C4
100n
+
9V

L1 & L2, L3 & L4 are wound on T80-15 toroids.
L1: 2.1m 40/60 litz wire about 110T, 160 uH
L2, L3 : 3T on both toroid cores.
L4: 2.1m (total) 40/60 litz wire, total ind 177 uH
C3: Wires from Q1 and C2 are twisted each other.
No capacitor is used.
R5: Trimmer potentiometer

©Atsushi Kano 2016/08

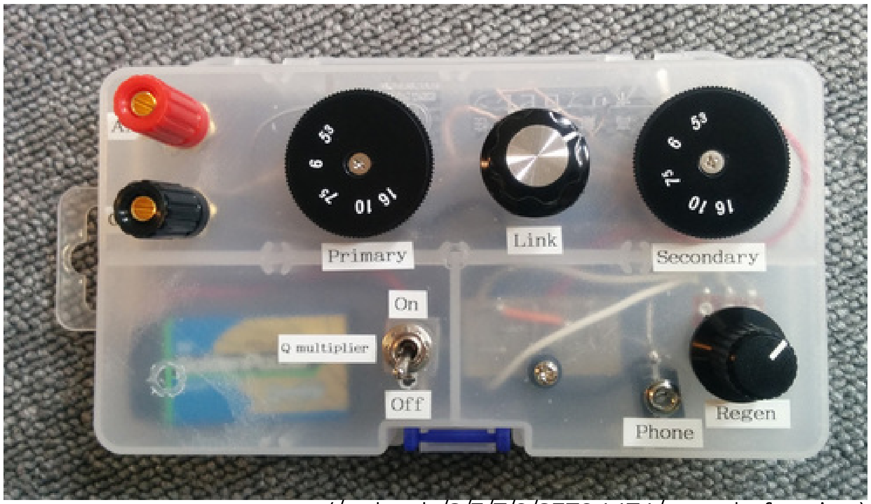

(/uploads/8/3/7/9/83794474/q_mult_face.jpg)

Face of the set

I used a small plastic tackle box sold at 100 yen (about US$1.00).

Top half is the crystal set designed by Mr. John Fuhring and right bottom is the Q multiplier . As I mentioned in the above, it is not easy to tune faint stations because my hand effects on the tuning while it is on the dial.

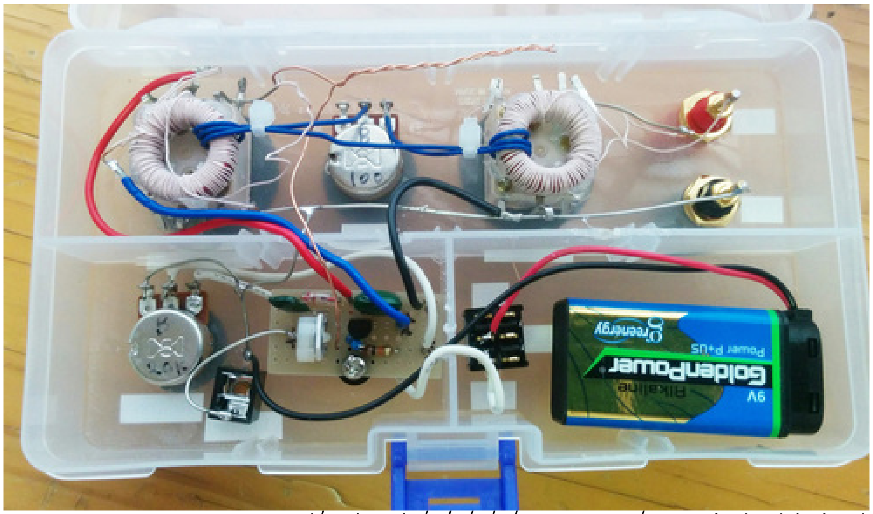

(/uploads/8/3/7/9/83794474/q_mult_inside.jpg)

Inside

I put all the parts for Q multiplier, the detecting diode, and few other parts for the crystal set on a small universal electronic board. The trimmer potentiometer, R5, was set to about 40 k after a few adjustments. Coils are hot-glued to the back of variable capacitors.

You can see the twisted wires, C3, for coupling signal from L4 to the base of Q1. The thinner blue wire forms L2 and L3. Originally it was a litz wire, but the link was too strong, so I changed it to the ordinary AWG 22 wire.

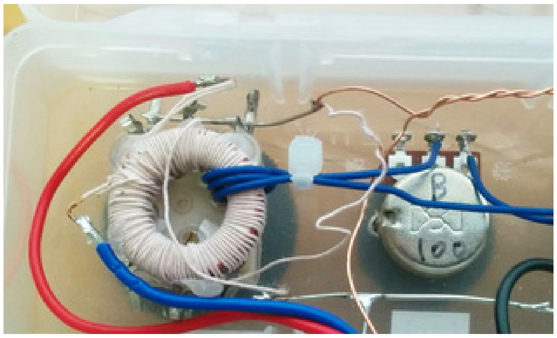

(/uploads/8/3/7/9/83794474/q_mult_l12.jpg)

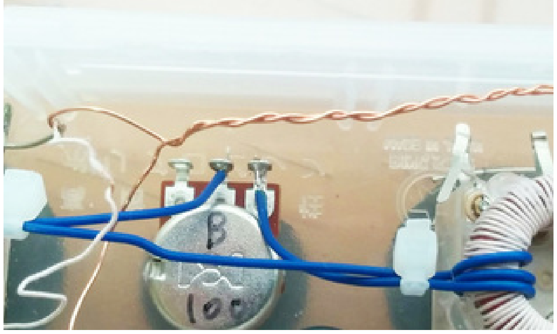

(/uploads/8/3/7/9/83794474/q_mult_c3.jpg)

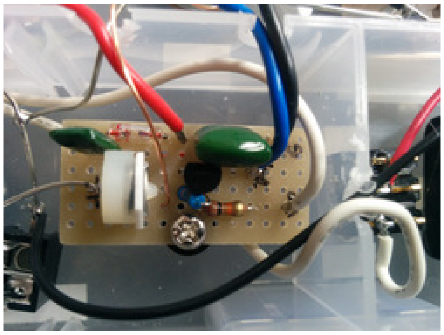

(/uploads/8/3/7/9/83794474/q_mult_pcb.jpg)

CREATE A FREE WEBSITE

POWERED BY

CRYSTAL SETS 1

Introduction

What - Why - How

Home | Contact | Site Map | Radio Stations & Memorabilia | Amateur Radio

Crystal Sets & other pages:

- Build Your Own Crystal Set (Pt 2)
- Spider's Web Crystal Set (Pt 3)
- Crystal Set By Kenneth Rankin (Pt 4)
- Experimental Crystal Set (Pt 5)

- TRF Radios
- Riding On A Wave
- Loop Aerials and ATU's
- Resistor Colour Codes & Capacitor Identification
- Component Suppliers
- The Exciting World of Amateur Radio by MØMTJ

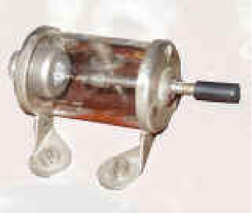

An old fashioned style detector that is housed in a glass tube. The point of the cat's whisker touches the crystal (galena) on the left & is adjusted with the black handle.

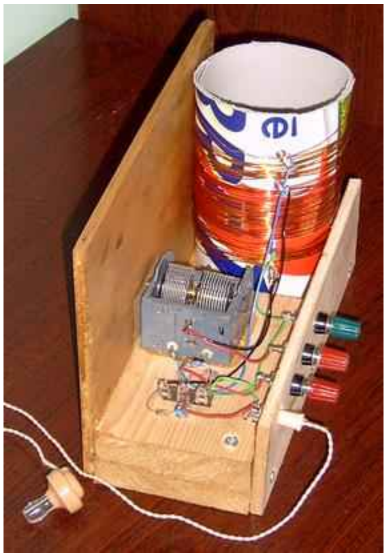

CRYSTAL SETS PART 1 - INTRODUCTION

What They Are, Why They Came Into Being and How YOU Can Build Your Own CRYSTAL SET!

What are Crystal Sets? I am glad you asked! It is a fascinating subject dating back to 1920 when Marconi first started experimenting with radio broadcasting from his station 2MT in Chelmsford.

The earliest listeners to Marconi's experimental radio programmes would have mainly used the most basic of radio sets called a Crystal Set. Companies such as the W.G Pye began producing early radio sets in 1921 which used electronic components called 'valves', the forerunner to today's transistors and silicon chips). These glass valves could be arranged to amplify the weak radio signal to make them louder - enough to operate a loudspeaker. However these radios were expensive and needed electricity or bulky lead acid batteries to operate.

THE MAGIC OF A CRYSTAL SET

The crystal set was popular during the 1920's and 30's because it was cheap to buy, much cheaper if you built one yourself, perhaps only a few shillings. Additionally a Crystal Set did not require electricity or expensive batteries, however it could only provide enough volume for headphone listening, and if it was situated in an area of low signal strength the listener would also require a very quiet room as the volume would be low. However In the 1920's hearing voices and music from a station miles away with no wires (hence the term Wireless) really was a magical experience! Listeners in the early twenties were also tinkerers and experimenters, trying different designs of crystal set in an attempt to get the loudest and clearest reception.

Since crystal sets have no batteries and no mains power, they rely entirely from the electrical energy developed between the aerial an earth connections - producing sounds as if by magic.

The crystal set had to be carefully tuned into the station by making adjustments to a tuning coil and condenser, these early crystal sets had the added complication of using a 'Cats Whisker' as the detector. The detector converted the radio waves received from the radio station into an audio wave that could be heard in the headphones. The Cats Whisker was a fine wire that rested onto a piece of galena, the crystal, and had to be very finely adjusted to obtain the loudest and clearest sound. Once the 'sweet spot' was found it was important not to move it, it would be very frustrating if someone bumped into the table and dislodged it!

To work at all a crystal set requires a very large aerial, but this is nothing more than a long piece of wire, perhaps 20 to 50 yards of insulated copper wire, hung outside around a garden between poles or trees, or even around the loft space. For the aerial to be effective it needs to be balanced with an earth. An earth is simply a copper rod or pipe about one yard long driven into some soft damp earth with a mallet. At the top of the pipe is fitted a jubilee or hose-clip to which is attached a length of insulated copper wire which is then fed into the house, along with the aerial wire, to the room where the crystal set will be operated and connected to the set's terminals.

It is quite possible to build your own crystal set. In its very simplest early form a crystal set consisted of a coil of wire, something called a 'detector' and a pair of very special headphones.

THE CONSTITUENT PARTS

Above: A REAL Crystal Set
You Can Build One Too!
HERE

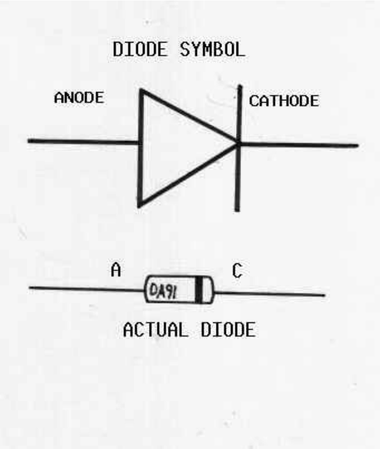
DIODE SYMBOL
ANODE
CATHODE
A
C
OA91
ACTUAL DIODE

An illustration of a modern diode which is often encapsulated in glass and is about 7mm long. Above is the electronic symbol for a diode.

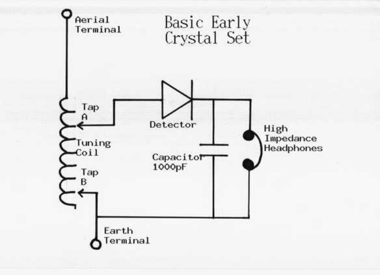
Aerial Terminal
Basic Early Crystal Set
Tap A
Detector
Tuning Coil
Capacitor 1000pF
High Impedance Headphones
Tap B
Earth Terminal

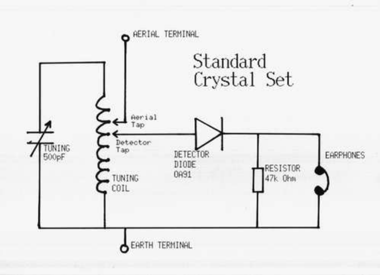
AERIAL TERMINAL
Standard Crystal Set
Aerial Tap
Detector Tap
TUNING 500pF
DETECTOR DIODE OA91
TUNING COIL
RESISTOR 47k Ohm
EARPHONES
EARTH TERMINAL

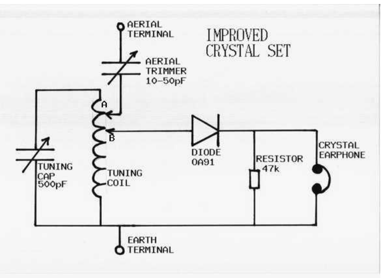
AERIAL TERMINAL
IMPROVED CRYSTAL SET
AERIAL TRIMMER 10-50pF
A
B
DIODE OA91
RESISTOR 47k
CRYSTAL EARPHONE
TUNING CAP 500pF
TUNING COIL
EARTH TERMINAL

The coil of wire is called a Tuning Coil, and the number of turns of wire on the coil determines the wavelength (frequency) that the set is tuned into. Tuning to different stations can be accomplished by varying the number of turms on the coil, or more easily by having many different tapping points on the coil so that adjustments can be made.

To make tuning easier a component called a tuning capacitor can be included in the circuit. In very early sets a tuning capacitor was not always included to keep costs down, or due to their being difficult to obtain.

The Detector converts the radio wave received into an electrical wave that is suitable for the headphones to, in turn, convert into sound waves that can be heard by the human ear. In the very early days of the crystal set the detector consisted of a holder containing a piece of galena crystal that had a very thin and springy wire placed on its surface that had to be very delicately adjusted to find the sweet spot where the radio station could be heard. This was commonly referred to as a "Cat's Whisker". Modern detectors are called diodes and are more efficient than early detectors and cats whiskers. Diodes are still quite readily available and inexpensive. Part numbers for modern diodes include OA90, OA91 and IN34. Diode part numbers that are perhaps now more difficult to obtain include OA47 and OA81.

The headphones have to be of a very special type called high impedance headphones. Because there is no additional power source in a crystal set the current generated in the circuit is tiny - minuscule in fact. Ordinary low impedance headphones, such as Walkman headphones, would present a virtual short circuit to the crystal set allowing the tiny signals to drain away to earth & consequently producing no sound - not very useful!

High impedance headphones, on the other hand, reduce or impede the flow of current down to earth, in effect saving the tiny signals to produce sounds from the headphones that we can then hear.

There is a problem however, these high impedance headphones that were so readily available in the 1920's and 1930's are to so easy to obtain today, but some specialist vintage radio outlets still stock them though the price can be quite high. They can still be seen in museums of course. All is not lost though, today we can obtain a special earphone called, appropriately, a crystal earpiece very easily and far more cheaply than. An electronic component, called a resistor, must be connected accross the crystal earphone to allow a path for DC current to get to earth. The value of the resistor is usually 47,000 Ohms and without it a crystal earphone tends to block DC current and as a consequence the sound will be very quiet and distorted.

Aerial and Earth: For a crystal set work it needs an efficient aerial and earth. Since a crystal radio has no power - no batteries or mains electricity - it relies entirely on the radio wave energy sent out from the radio station's transmitter and collected by the aerial to work. The aerial simply consists of a length of wire, but it necessarily has to be quite long, usually in the order of 10 to 20 meters. For the aerial to be effective the crystal radio set also has to be connected to a good earth point. A good earth often consists of a 3 or 4 foot copper stake driven into the ground, but sometimes a water pipe can be used to reasonable effect. (Safety: Never use the earth pin of a household mains plug)

The circuit diagram (schematic) for a crystal set is shown below. Practical designs are shown on the following pages.

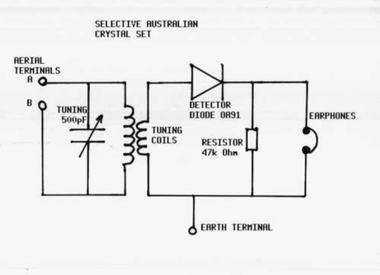
SELECTIVE AUSTRALIAN
CRYSTAL SET
AERIAL
TERMINALS
A
B
TUNING
500pF
TUNING
COILS
DETECTOR
DIODE 0A91
RESISTOR
47k Ohm
EARPHONES
EARTH TERMINAL

No AM radio stations or transmitters in your locality or country?

Has your local medium wave broadcast station closed or been moved to VHF/FM or Digital? Don't worry! You can still build and experiment with crystal sets and TRF radios by also building or buying a simple low power AM transmitter. So, not only can you use your crystal sets but you can also run your own radio station that can be heard in and around your home - playing the music or programmes that you want to hear!

Spitfire & Metzo Complete high quality ready built medium wave AM Transmitters from Vintage Components:
http://www.vcomp.co.uk/index.htm

Spitfire
AM Transmitter

Choice of Spitfire and Metzo transmitters:
Spitfire:
http://www.vcomp.co.uk/spitfire/spitfire.htm

Metzo : AM transmitter including audio compression circuits:
http://www.vcomp.co.uk/metzo/metzo.htm

SSTRAN AMT3000 : Superb high fidelity medium wave AM transmitter kits from SSTRAN. Versions available for 10kHz spacing in the Americas (AMT3000 or AMT3000-SM) and 9kHz spacing in Europe and other areas (AMT3000-9 and AMT3000-9SM). Superb audio quality and a great and well designed little kit to build:
http://www.sstran.com/pages/products.html

POWER
GAIN
MODULATION
COMPRESSION
SSTRAN
AMT3000

http://www.sstran.com/

AM88 LP : AM transmitter kit from North County Radio.
http://www.northcountryradio.com/Kitpages/am88.htm

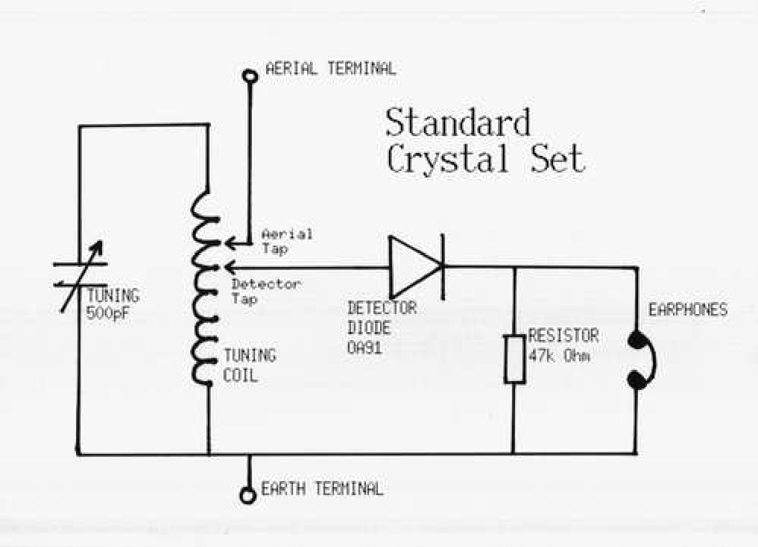
AERIAL TERMINAL
Standard
Crystal Set
Aerial
Tap
Detector
Tap
TUNING
500pF
TUNING
COIL
DETECTOR
DIODE
0A91
RESISTOR
47k Ohm
EARPHONES
EARTH TERMINAL

Above: The circuit diagram for the standard Crystal Radio

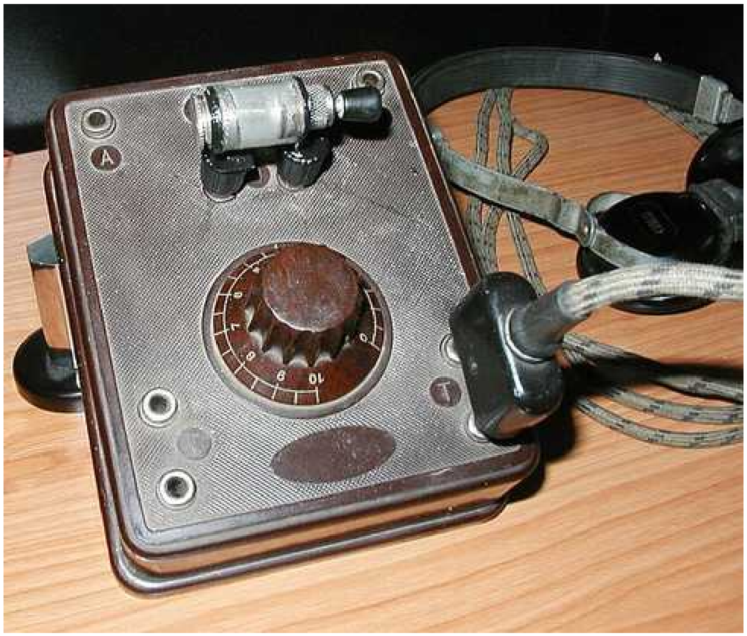

Above: A commercially made crystal set from around 1940

There is no variable tuning capacitor (condenser) in the above design. The capacitor is a fixed component and tuning is achieved by varying the coil inductance with the rotating switch. The detector is a galena crystal, seen at the top of the panel. A pair of high impedance magnetic headphones can just be seen to the right of this shot. (photo sent in by our reader Krysatec in Czechoslovakia - thanks!)

EXCELLENT RESULTS

Almost certainly when building a crystal set today a crystal earphone will need to be used. A crystal earphone is not quite as good as the old high impedance headphones as it is not quite as sensitive and therefore not quite so loud, however today's transmitters are much more powerful than those of the 1920's and so really excellent results are very often obtained - just don't expect to hear everything that you might on a top quality portable radio!

Crystal sets are fascinating because they are so simple to build, and because there is no power required they remain magical to this day. I have built several over the years, with differing designs, and if you'd like to build one yourself, it's easy, and I have described some circuits in part 2 with some more experimental ideas in part 5.

I have a crystal set that receives BBC Radio 4 on long wave and 3 local stations and 3 national stations on medium wave and another that also receives shortwave stations. Not bad for radios with no battery or other power - and such as set could be very useful in an emergency.

More About Diodes - unconventional and new ideas (by Felix Scerri)

While messing around with diodes for crystal sets recently, I tried something as a theory that actually works very well in practice, and with one or two actual advantages.

Ordinarily the ubiquitous 1N914 or IN4148 silicon switching diodes are pretty hopeless as RF detectors unless RF signals are *very* strong! However there is a simple way of making these diodes work very well as RF detectors, by the addition of adjustable voltage bias. I have used adjustable voltage bias with diodes before with very good results, but using bias with IN914 or 1N4148's is a relatively new idea!

Interestingly enough with voltage bias correctly adjusted, 'weak signal' performance and sound quality is actually very good, with that typically 'crisp' silicon diode 'sound'. I use a simple bias network with a 1.5 v AA battery and a 1 Megohm potentiometer applied to the diode directly with the incoming RF fed into the diode through a 1 uf plastic film capacitor (as shown). The use of adjustable bias seems to help most with 'repositioning' the position of the knee in the diodes curve and weak signal performance is excellent, easily as good as any 'proper' germanium diode.

Another 'tweak' that is actually quite beneficial is 'paralleling' several diodes together. This doesn't work with ordinary germanium diodes but definitely works with BAT46 'germanium equivalent' diodes and the effect of three of these diodes together is to make a better 'single' diode. The turn on voltage is noticeably reduced and the 'square law; region of detection is definitely improved and so is the rectified audio quality. These diodes are relatively inexpensive (at least in this country), so it is worth trying. 73 Felix vk4fuq 080214.

ALSO: BUILD A RADIO IN A MATCHBOX!

ENGLAND'S GLORY
MADE IN SWEDEN
QUALITY

Click To Find Out More!

- **Amateur Radio by MØMTJ**

Part 2
Practical Designs To
BUILD YOUR OWN CRYSTAL SET!

Having difficulty in finding components? I have added some ideas for component sources here.
Sources For Older Components

Build Your Own Crystal Set (Part 2) | **Spider's Web Crystal Set (Part 3)**

Crystal Set By Kenneth Rankin (Part 4) | **Experimental Crystal Set (Part 5)**

LINKS

Components:

BOWOOD ELECTRONICS - A friendly, helpful and very speedy source for your electronic components at prices that won't frighten your wallet! - The MK484 IC (the ZN414 replacement) is available from here.

VINTAGE COMPONENTS - A great resource for crystal sets, components, valve radio kits and medium wave AM transmitters!

6V6 - Electronic Nostalgia and Vintage Components

1N34A.com - crystal radio parts, including litz wire is http://1n34a.com

More Links To Electronic Component Suppliers here

Crystal Sets and other interesting radio related topics:

OZ CRYSTAL RADIOS by Austin Hellier

Simple crystal set receivers used by soldiers during the war and by prisoners of war (P.O.W.'s):
THE FOXHOLE and P.O.W RADIOS

Dave Schmarder's Crystal Set Pages
http://www.makearadio.com/crystal/index.php http://www.makearadio.com

"Heart of England Crystal Radio Club" based in Birmingham, West Midlands - http://crystalradioclub.co.uk

Stay Tuned http://www.crystalradio.net/
Designs http://www.crystalradio.net/crystalplans

The Tuggle Two Crystal Set
http://www.crystalradio.net/crystalsets/miketuggle/index.shtml

Jim's Crystal Radio Page http://www.hobbytech.com/crystalradio/crystalradio.htm

VE7SL Crystal Radio DXing http://members.shaw.ca/ve7sl/crystal.html

Gollum's Crystal Receiver World http://www.oldradioworld.de/gollum/

High Fidelity AM/MW Broadcast Band Reception by Felix Scerri
http://sound.westhost.com/articles/am-radio.htm

Crystal Set by G3XBM http://homepage.ntlworld.com/lapthorn/xtal.htm

Experimental Crystal Set Designs by Ken Harthun
http://www.qsl.net/kc4iwt/xtal/SWMystery.htm

Short Wave Crystal Set by Al Klase
http://www.skywaves.ar88.net/SWXS/shortwave.html

Scott's Crystal Radios and links http://oldheadphones.com/crystal/index.html

Crystal Set Analysis by Phil Anderson WØXI (ARRL)
http://www.arrl.org/qex/2008/09/Anderson.pdf

High Sensitivity MOSFET Crystal Set (ARRL)
http://www.arrl.org/qst/2007/01/culter.pdf

MOSFET Crystal Radio by G1EXG http://www.creative-science.org.uk/epadmosfets.html

Peebles Crystal Radios http://www.peeblesoriginals.com/

Crystal Set Kits http://www.peeblesoriginals.com/crystal-kits.html

http://www.peeblesoriginals.com/catalog/45.php

Early Wireless http://www.earlywireless.com/crystal_sets.htm

Stay Tuned Crystal Radio Resources http://www.crystalradio.net/

Birmingham, Alabama Crystal Radio Group http://www.crystalradio.us/index.htm

Borden Radio Company - commercial website (USA)
http://www.xtalman.com/literature.html

Geoff's Crystal Radios - commercial website (UK) http://crystalreceiver.co.uk/

Crystal Radio Supply - commercialwebsite (USA)
http://www.crystalradiosupply.com/index1.html

The Xtal Set Society - commercial websites (USA) http://crystalradio.org/
http://www.midnightscience.com/

Vintage Radio http://www.vintage-radio.com/

PAGES WITH ARTICLES BY RADIO EXPERT FELIX SCERRI:

High Quality AM Receivers:

http://www.crystalradio.net/scerri/index.shtml

http://www.crystalradio.net/

http://www.xtal-sets.com/

Hi Fi AM Receiver Designs:

http://sound.westhost.com/articles/am-radio.htm

TRF Radio Websites:

Gilbert Davey's Radio Sets - A growing resource for all who either remember building radio sets to his designs or would simply like to find out more: http://www.daveysradios.org.uk

Vintage Radio - TRF Radio Designs: http://www.vintageradio.me.uk/radconnav/transtrf/

Vintage Radio - Home Page: http://www.vintageradio.me.uk/

Vintage Radio - Circuit diagrams for crystal sets, TRF and superhet radios, valve and transistor designs: http://www.vintageradio.me.uk/radconnav/radcon.htm

TRF Designs on Birmingham Alabama Crystal Radio Group: http://www.crystalradio.us/1adradios/1ad-2007-2.htm

Regen Radio Tutorial N1TEV : http://www.electronics-tutorials.com/receivers/regen-radio-receiver.htm

Coil Calculations: http://bellsouthpwp2.net/w/u/wuggy/coils.html

E.P.E. Everyday and Practical Electroncs Magazine: http://www.epemag3.com/

The Radio Board Forums - incl Audion Receiver: http://theradioboard.com/rb/viewtopic.php?t=6161&start=30

OTHER ARTICLES

ESP by Rod Elliot - "Mad as Hell" & lead free solder directive:

http://sound.westhost.com/madashell.htm

ESP by Rod Elliot - CFL's (Compact Fluorescent Lamps)

http://sound.westhost.com/articles/incandescent.htm

BOOKS

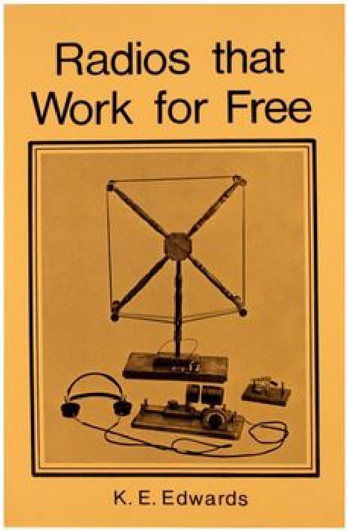

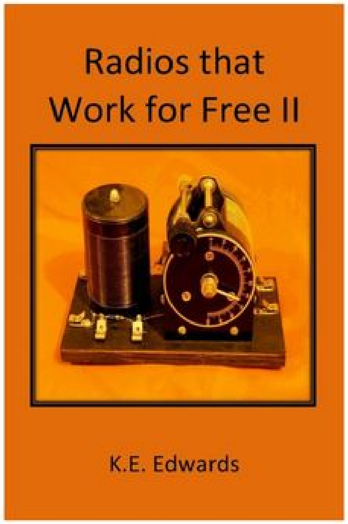

“Radios that Work for Free" and “Radios that Work for Free II” are now available on Amazon Kindle Books! After 38 years in print "Radios that Work for Free” is now an E Book and K. E. Edwards’ new book has just been released as an E Book as well. Crystal set builders worldwide have a new set of crystal set plans and photos to be encouraged and inspired by. These books take the reader from beginner to a competent builder of crystal sets. Take a look inside on Amazon Kindle books.
Hope And Allen Publishing - www.hopeandallen.com

Amateur Radio
MØMTJ Visit my amateur radio pages MØMTJ

Visit my Amateur Radio Pages >

Build Your Own Crystal Set (Part 2) | **Spider's Web Crystal Set (Part 3)**

Crystal Set By Kenneth Rankin (Part 4) | **Experimental Crystal Set (Part 5)**

^Top Of Page

Home | **Contact** | **Site Map** | **Reciprocal Links & Credits** | **Thank You** | **Radio Stations & Memorabilia** | **Amateur Radio**

www.mds975.co.uk © 2005 - 2016

CRYSTAL SETS 3 Spider's Web Coil

With thanks to Chris Dorna

Home | Contact | Site Map | Radio Stations & Memorabilia | Amateur Radio

Crystal Sets Introduction | Resistor & Capacitor Conversion Tables

Crystal Sets (Part1) | **Build Your Own Crystal Set (Part 2)**

Crystal Set by Kenneth Rankin (Part 4) | **Experimental Crystal Sets (Part 5)** | **Crystal Radio Links**

CRYSTAL SETS 3: CHRIS DORNA'S SPIDER'S WEB COIL

Making The Spider's Web

(NB. TRANSLATED FROM DUTCH
so I must apologise for any small errors)

See Chris Dorna's Website HERE!

Generally coils are cylindrical. We have chosen a spiders web coil because these are simple to wind. You must however pay attention that you lay the windings nice and tight when winding the coil.

The material chosen is 3.4 mm MDF as is easy to work with. If you make a former then you could also use an old CD perhaps.

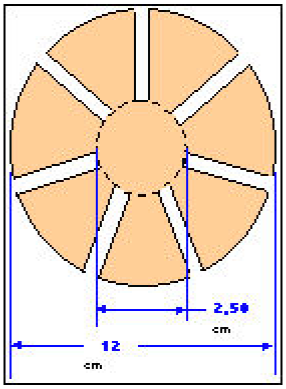

On the former shown on the left 7 segments have been created. For the crystal radio in the article below, the coil has been made with 9 segments at 40° intervals.

The inner diameter is 2.5 cm and the outer diameter is 12 cm.

The former is 4.75 mm thick and there are about 60-70 turns on the winding.

The wire should be reasonably thick as very thin wire can very easily break when winding.

For calculating the coil we used Professor Coyle's Spiderweb Coil Calculator

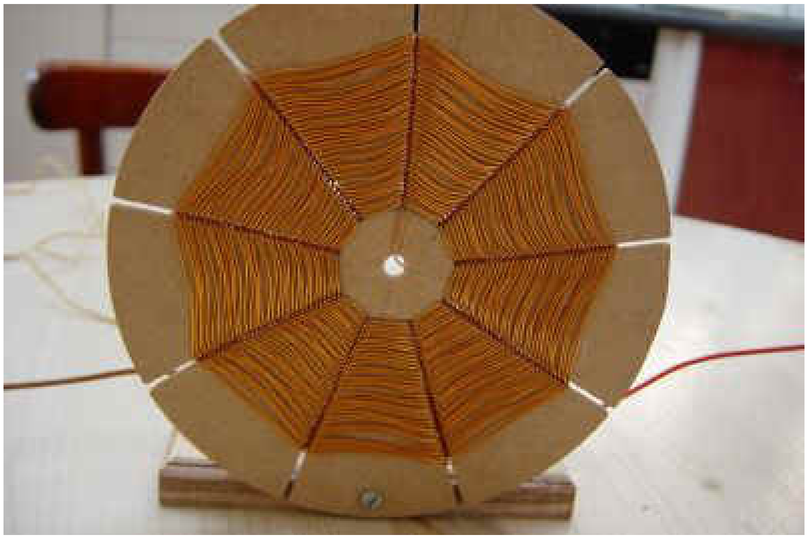

The Completed Coil

The Scouts "Jamboree On The Air"

Below are a couple of photographs of the completed crystal radio that the Vught North Scouts constructed during the their "Jamboree On The Air". The spiders web coil is a very interesting feature of the set and was chosen because it was thought to be easier to wind for the younger scouts.

The scouts were provided with a package that contained all the required components including the former for the coil, suitable wire to wind the coil and a base panel onto which they could build the spiders web crystal set on their Jamboree. They had to ensure, however, that great care was taken when putting up the wire for the large antenna for the wire which was a minimum of 10 meters long. The crystal set also had to be connected to an earth point and a water pipe was used for this.

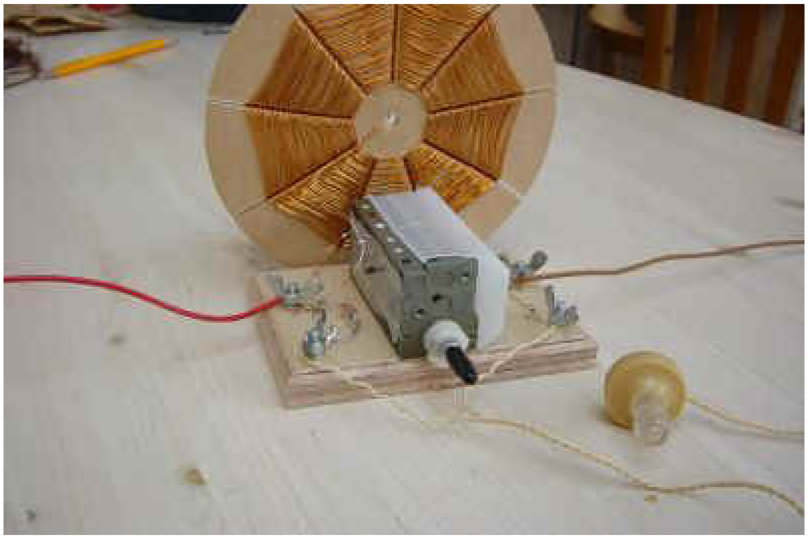

Photo showing the completed crystal set consisting of a germanium diode, a tuning capacitor and the spider web coil. The red wire is the aerial and the brown wire is connected to earth. The crystal earphone is shown in the foreground.

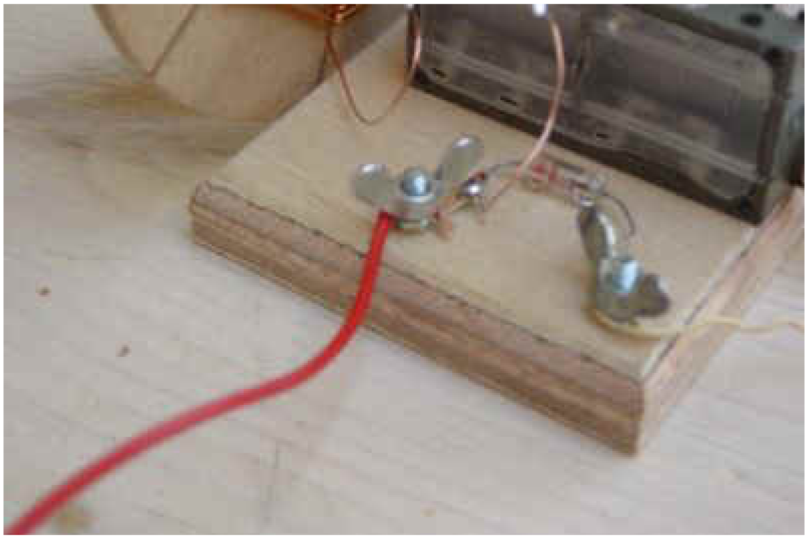

The germanium diode that converts the received radio frequency signal to audio that can be heard in the crystal earphone.

PARTS THAT WERE REQUIRED BY THE SCOUTS WERE PROVIDED IN THE THE PACKAGE:-

1	Base Panel
1	Coil Former
1	Tuning capacitor
1	Reel Of Coil Wire
1	Germanium Diode e.g. OA47, OA90, OA91, IN34 etc
1	Crystal Earphone
All	Screws and fixings required, tools, solder and soldering iron
1	Constructional Description with photograph Click HERE to download instructions. (7MB PDF File)

Despite the clear instructions provided there was soldering involved in the construction of the crystal set and the scouts were guided in this process by experienced members (such as radio amateurs) to ensure safety at all times.

You can get in contact with regard to this project at: kristalradio@dorna.nl
or at pe1dzx@amsat.org

Sincere thanks to Chris Dorna for bringing this exciting project to our attention. Excellent work one and all!

Kenneth Rankin has also sent in the details of the crystal set that he has built and you can see his photographs and notes in CRYSTAL SETS 4

I hope you try building one, it's easy and great fun!

Crystal Sets Part 4 >

No AM radio stations or transmitters in your locality or country?

Spitfire
AM Transmitter

Has your local medium wave broadcast station closed or been moved to VHF/FM or Digital? Don't worry. You can still build and experiment with crystal sets and TRF radios by also buying or even building a simple low power AM transmitter. So, not only can you use your crystal sets but you can also run your own radio station that can be heard in and around your home - playing the music or programmes that you want to hear!

SSTRAN AMT3000 : Superb high fidelity medium wave AM transmitter kits from SSTRAN. Versions available for 10kHz spacing in the Americas (AMT3000 or AMT3000-SM) and 9kHz spacing in Europe and other areas (AMT3000-9 and AMT3000-9SM). Superb audio quality and a great and well designed little kit to build: http://www.sstran.com/pages/products.html

POWER
GAIN
MODULATION
COMPRESSION
SSTRAN
AMT3000

http://www.sstran.com/

Other AM transmitters available:

Spitfire & Metzo Complete, high quality ready built medium wave AM Transmitters from Vintage Components:
http://www.vcomp.co.uk/index.htm Vintage Components offer a choice of the high quality Spitfire and Metzo transmitters:

SPITFIRE AM Medium Wave Transmitter with 100 milliwatt RF output power:

http://www.vcomp.co.uk/spitfire/spitfire.htm

Spitfire
AM Transmitter

METZO AM Medium Wave Transmitter with built in compressor:

Gain
Comp
Mod
Metzo

http://www.vcomp.co.uk/metzo/metzo.htm

AM88 LP A basic AM transmitter kit from North County Radio.
http://www.northcountryradio.com/Kitpages/am88.htm

LINKS:

BOWOOD ELECTRONICS - A friendly, helpful and very speedy source for many of your electronic components at prices that won't frighten your wallet!

THE FOXHOLE and P.O.W RADIOS - Simple crystal set receivers used by soldiers during the war and by prisoners of war (P.O.W.'s).

VINTAGE COMPONENTS - A great resource for crystal sets, components, valve radio kits and medium wave AM transmitters!

6V6 - Electronic Nostalgia and Vintage Components

Crystal Sets (Part1) | Build Your Own Crystal Set (Part 2)

Crystal Set by Kenneth Rankin (Part 4) | Experimental Crystal Sets (Part 5) | Crystal Radio Links

^Top Of Page

Amateur Radio

MØMTJ Visit my amateur radio pages MØMTJ

Visit my Amateur Radio Pages >

Home | Contact | Site Map | Radio Stations & Memorabilia | Amateur Radio

Crystal Sets Introduction | Resistor & Capacitor Conversion Tables

www.mds975.co.uk © 2004 - 2014

CRYSTAL SETS
3
Spider's Web Coil

With thanks to Chris Dorna

Home | Contact | Site Map | Radio Stations & Memorabilia | Amateur Radio

Crystal Sets Introduction | Resistor & Capacitor Conversion Tables

Crystal Sets (Part1) | **Build Your Own Crystal Set (Part 2)**

Crystal Set by Kenneth Rankin (Part 4) | **Experimental Crystal Sets (Part 5)** | **Crystal Radio Links**

CRYSTAL SETS 3: CHRIS DORNA'S SPIDER'S WEB COIL

Making The Spider's Web

(NB. TRANSLATED FROM DUTCH
so I must apologise for any small errors)

See Chris Dorna's Website HERE!

Generally coils are cylindrical. We have chosen a spiders web coil because these are simple to wind. You must however pay attention that you lay the windings nice and tight when winding the coil.

The material chosen is 3.4 mm MDF as is easy to work with. If you make a former then you could also use an old CD perhaps.

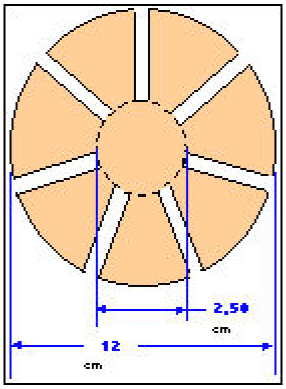

On the former shown on the left 7 segments have been created. For the crystal radio in the article below, the coil has been made with 9 segments at 40° intervals.

The inner diameter is 2.5 cm and the outer diameter is 12 cm.

The former is 4.75 mm thick and there are about 60-70 turns on the winding.

The wire should be reasonably thick as very thin wire can very easily break when winding.

For calculating the coil we used Professor Coyle's Spiderweb Coil Calculator

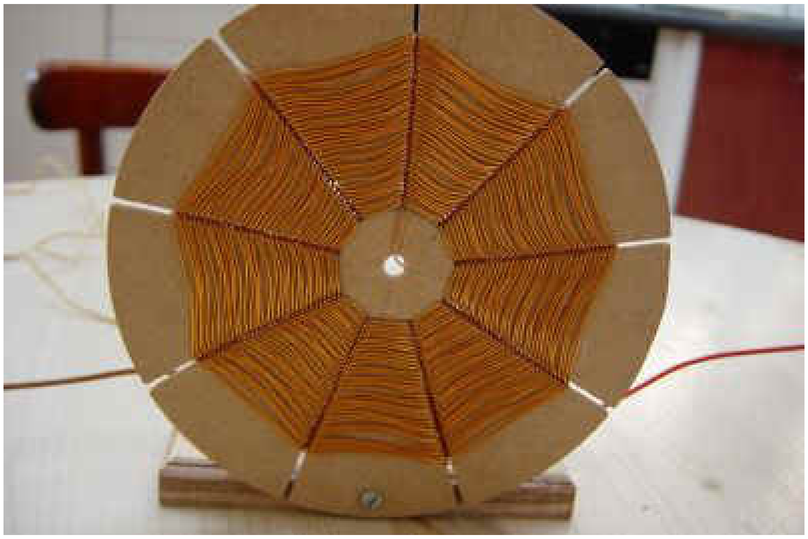

The Completed Coil

The Scouts "Jamboree On The Air"

Below are a couple of photographs of the completed crystal radio that the Vught North Scouts constructed during the their "Jamboree On The Air". The spiders web coil is a very interesting feature of the set and was chosen because it was thought to be easier to wind for the younger scouts.

The scouts were provided with a package that contained all the required components including the former for the coil, suitable wire to wind the coil and a base panel onto which they could build the spiders web crystal set on their Jamboree. They had to ensure, however, that great care was taken when putting up the wire for the large antenna for the wire which was a minimum of 10 meters long. The crystal set also had to be connected to an earth point and a water pipe was used for this.

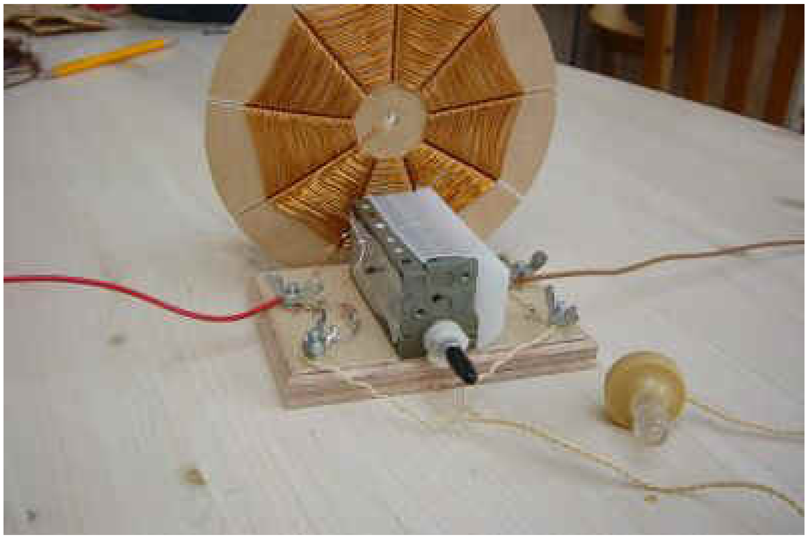

Photo showing the completed crystal set consisting of a germanium diode, a tuning capacitor and the spider web coil. The red wire is the aerial and the brown wire is connected to earth. The crystal earphone is shown in the foreground.

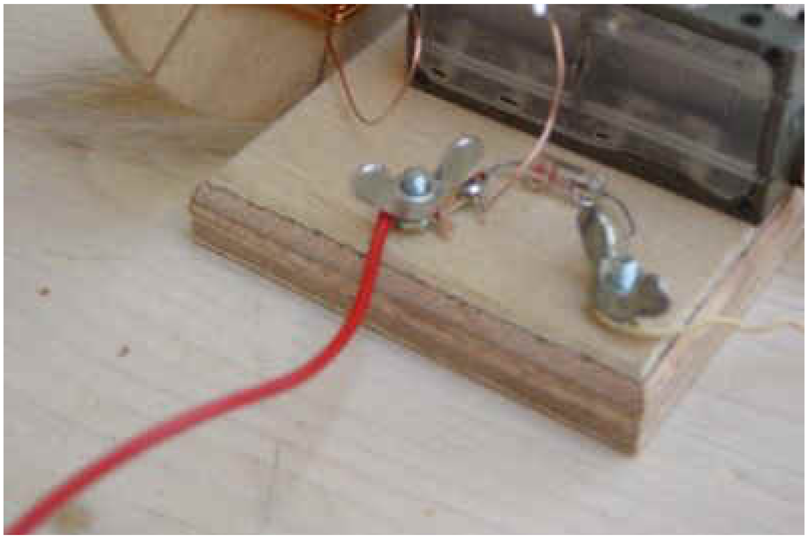

The germanium diode that converts the received radio frequency signal to audio that can be heard in the crystal earphone.

PARTS THAT WERE REQUIRED BY THE SCOUTS WERE PROVIDED IN THE THE PACKAGE:-

1	Base Panel
1	Coil Former
1	Tuning capacitor
1	Reel Of Coil Wire
1	Germanium Diode e.g. OA47, OA90, OA91, IN34 etc
1	Crystal Earphone
All	Screws and fixings required, tools, solder and soldering iron
1	Constructional Description with photograph Click HERE to download instructions. (7MB PDF File)

Despite the clear instructions provided there was soldering involved in the construction of the crystal set and the scouts were guided in this process by experienced members (such as radio amateurs) to ensure safety at all times.

You can get in contact with regard to this project at: kristalradio@dorna.nl
or at pe1dzx@amsat.org

Sincere thanks to Chris Dorna for bringing this exciting project to our attention. Excellent work one and all!

Kenneth Rankin has also sent in the details of the crystal set that he has built and you can see his photographs and notes in CRYSTAL SETS 4

I hope you try building one, it's easy and great fun!

Crystal Sets Part 4 >

No AM radio stations or transmitters in your locality or country?

Spitfire
AM Transmitter

Has your local medium wave broadcast station closed or been moved to VHF/FM or Digital? Don't worry. You can still build and experiment with crystal sets and TRF radios by also buying or even building a simple low power AM transmitter. So, not only can you use your crystal sets but you can also run your own radio station that can be heard in and around your home - playing the music or programmes that you want to hear!

SSTRAN AMT3000 : Superb high fidelity medium wave AM transmitter kits from SSTRAN. Versions available for 10kHz spacing in the Americas (AMT3000 or AMT3000-SM) and 9kHz spacing in Europe and other areas (AMT3000-9 and AMT3000-9SM). Superb audio quality and a great and well designed little kit to build: http://www.sstran.com/pages/products.html

POWER
GAIN
MODULATION
COMPRESSION
SStran
AMT3000

http://www.sstran.com/

Other AM transmitters available:

Spitfire & Metzo Complete, high quality ready built medium wave AM Transmitters from Vintage Components:
http://www.vcomp.co.uk/index.htm Vintage Components offer a choice of the high quality Spitfire and Metzo transmitters:

SPITFIRE AM Medium Wave Transmitter with 100 milliwatt RF output power:

http://www.vcomp.co.uk/spitfire/spitfire.htm

Spitfire
AM Transmitter

METZO AM Medium Wave Transmitter with built in compressor:

Gain
Comp
Mod
Metzo

http://www.vcomp.co.uk/metzo/metzo.htm

AM88 LP A basic AM transmitter kit from North County Radio.
http://www.northcountryradio.com/Kitpages/am88.htm

LINKS:

BOWOOD ELECTRONICS - A friendly, helpful and very speedy source for many of your electronic components at prices that won't frighten your wallet!

THE FOXHOLE and P.O.W RADIOS - Simple crystal set receivers used by soldiers during the war and by prisoners of war (P.O.W.'s).

VINTAGE COMPONENTS - A great resource for crystal sets, components, valve radio kits and medium wave AM transmitters!

6V6 - Electronic Nostalgia and Vintage Components

Crystal Sets (Part1) | **Build Your Own Crystal Set (Part 2)**

Crystal Set by Kenneth Rankin (Part 4) | **Experimental Crystal Sets (Part 5)** | **Crystal Radio Links**

^Top Of Page

Amateur Radio

MØMTJ Visit my amateur radio pages MØMTJ

Visit my Amateur Radio Pages >

Home | Contact | Site Map | Radio Stations & Memorabilia | Amateur Radio

Crystal Sets Introduction | Resistor & Capacitor Conversion Tables

www.mds975.co.uk © 2004 - 2014

CRYSTAL SETS
4
A Crystal Set By Kenneth Rankin

Home | Contact | Site Map | Radio Stations & Memorabilia | Amateur Radio

Crystal Sets Introduction | Resistor & Capacitor Conversion Tables

Crystal Sets (Part1) | Build Your Own Crystal Set (Part 2)

Spider's Web Crystal Set (Part 3) | Experimental Crystal Sets (Part 5) | Crystal Radio Links

CRYSTAL SETS 4: A CRYSTAL SET BUILT BY KENNETH RANKIN

MDS975.co.uk reader Roddie Rankin kindly sent in the details of a crystal set built by his son Kenneth for a school project which won him great praise and merit. The crystal set is so nicely made and the project so well written that I thought it deserved a special page of its own.

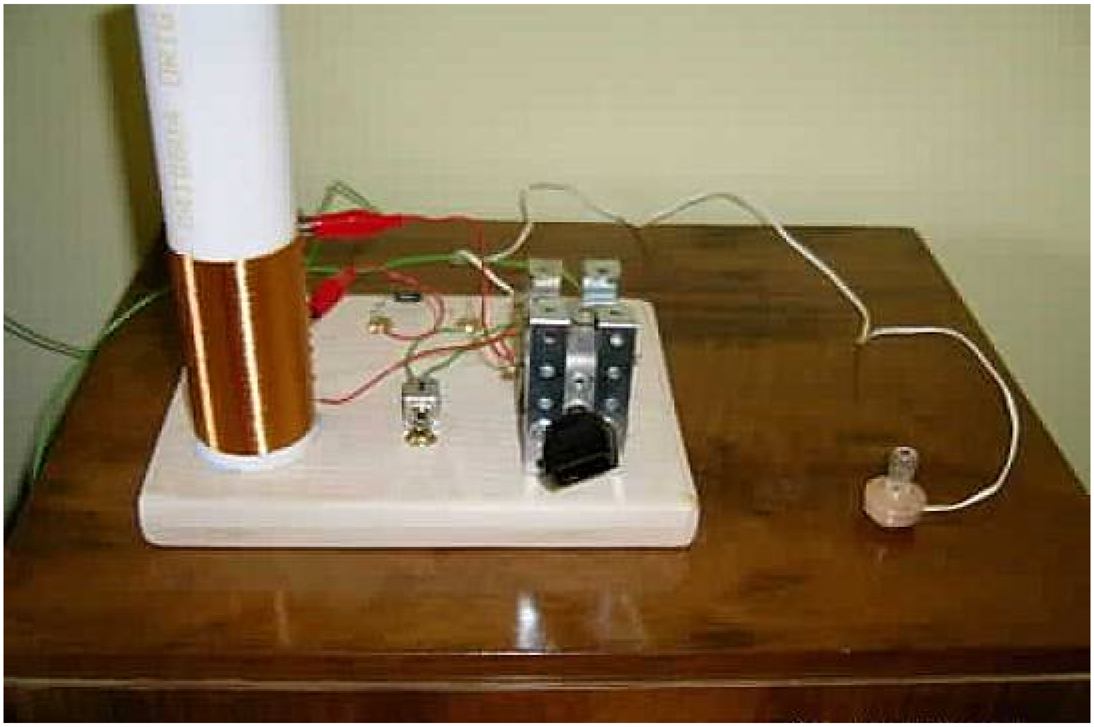

Photograph Of Kenneth Rankin's Rather Super Crystal Set

CONTENTS

1 - What Is A Crystal set?
2 - How A Crystal Set Works
3 - History Of Crystal Sets
4 - How I Built My Crystal Set
5 - The Results I Obtained
6 - Where I Found The Information
7 - Quiz

1 - WHAT IS A CRYSTAL SET?

Perhaps you think, like I did, that a crystal set is just a collection of lovely stones. In fact a crystal set is a very early, simple domestic radio. The photograph above shows the crystal set I built.

The crystal set does not have a battery. It runs completely from the radio waves it picks up in the air. It is absolutely free to run!

Please read on to find out more about this exciting invention!

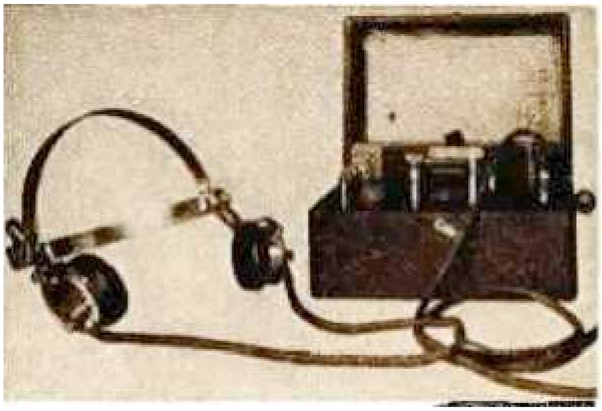

The birth of domestic radio: A deluxe crystal set produced by Marconi for the reception of programmes from the first of the BBC's radio stations, 2LO in London.
[from Pageant Of The Century - Odhams Press]

2 - HOW A CRYSTAL SET WORKS

The diagram below shows the circuit diagram of the crystal set I have built. It consists of a coil (L1), a tuning capacitor (C2), a diode (X1), an aerial, a crystal earpiece and an earth. (Capacitors C1 and C3 are not essential.)

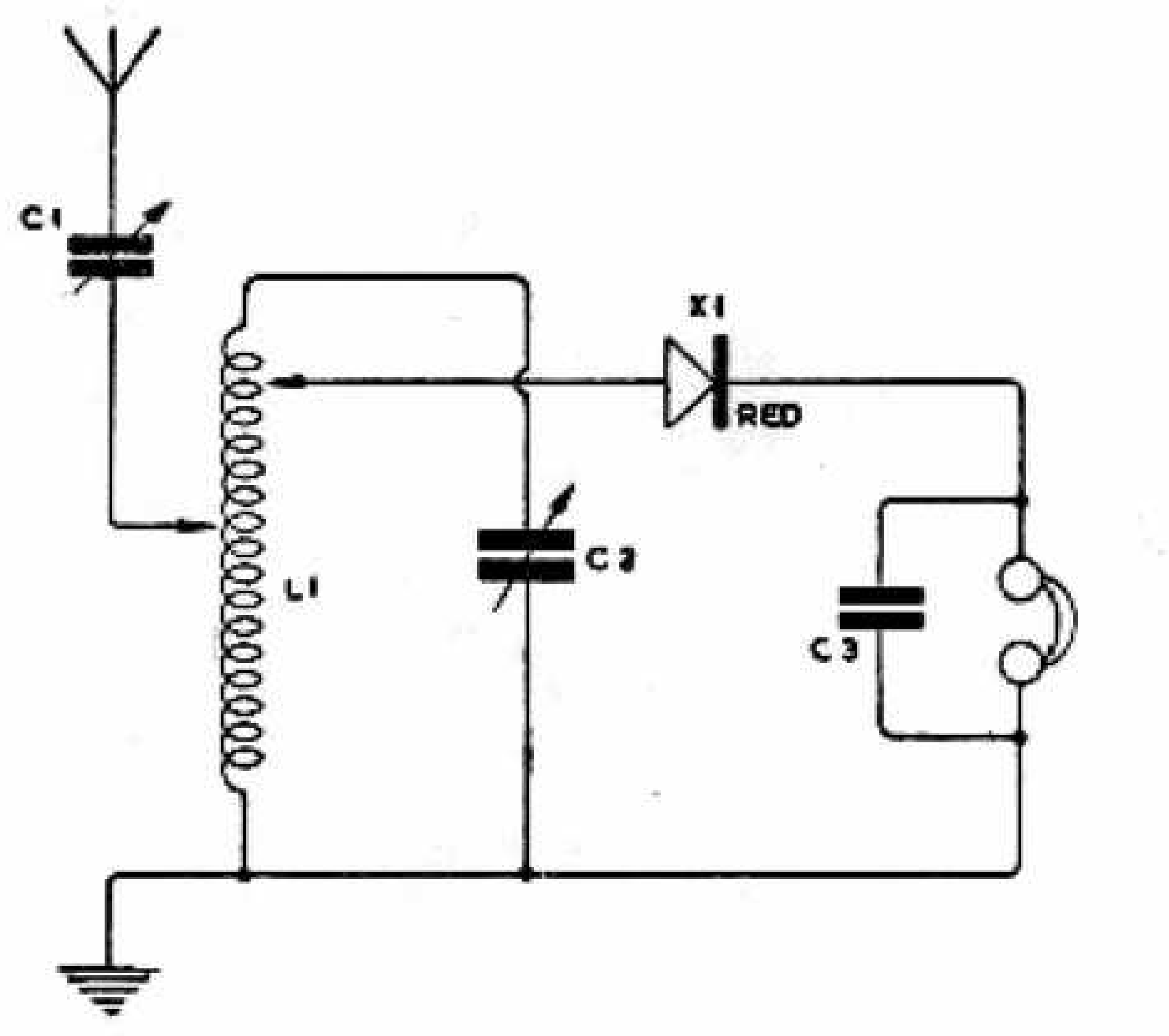

Circuit Diagram Of The Crystal Set

First the aerial picks up radio waves from radio stations all over the world. Then the required station needs to be selected, this is called tuning, and this is achieved with a Tuned Circuit. The Tuned Circuit in the crystal set consists of two items; a component called a variable capacitor and a coil of wire.

By adjusting the value of the capacitor and coil tuning is achieved. The coil taking unwanted signals of low frequency straight through to the earth, which gets rid of them, while the capacitor takes unwanted signals of high frequency straight through to the earth. The remaining signal, in between the high frequency and low frequency signals, is the signal of the required station which is not passed to earth by either the coil or the capacitor. This is the frequency to which the set is now tuned in.

You normally change the frequency to which a set is tuned in to by turning the knob on the tuning capacitor, thus receiving different stations by varying the capacitance of the variable capacitor. The number of turns on the coil can also be changed [called varying the inductance of the coil] and this will give different tuning ranges. A larger number of turns provided longer wavelengths [lower frequencies], while a smaller number of turns of wire provides shorter wavelengths [higher frequencies].

[It is also worth noting that the inductance of a coil can be increased by inserting a component called a Ferrite Rod into the centre of the coil.]

Once tuned into the required station this remaining signal goes to earth by travelling through the diode and earpiece. The diode takes the sound signal out of the radio signal [called demodulation], and passes it to the earpiece. The sound signal travels through the crystal in the earpiece which vibrates the air in the earpiece giving us the sound of the radio station to which you are tuned.

3 - THE HISTORY OF THE CRYSTAL SET

The history of crystal sets begins with the first broadcast stations being opened. This happened in England in 1922. The simplest form of receiver for these broadcasts was the crystal set.

In those early sets they had "cat's whiskers" instead of diodes. A cat's whisker was a fine piece of wire which was adjusted to make a suitable contact with a crystal. The crystal would be made of a substance such as coke or galena. An example from 1923 is shown in the picture below. Notice how well made these instruments were.

There were several limitations to the crystal set: it needed a big aerial, an earth connection, the clumsy cat's whisker, and the weak signal could only be listened to by one person at a time with headphones. Very quickly the crystal set began to be replaced by valve radios with loudspeakers, powered by battery (see the picture from 1923).

In World War II, crystal sets were used by prisoners of war in prison camps to listen to news from home. As they were often badly treated it was hard to get all the right parts to build the set.

By this time crystal sets were no longer used in the home. Instead radios like the attractive 1937 model in the picture were being used.

Nowadays people only use crystal sets as a hobby. There are various different kits available on the market (see websites in the references below).

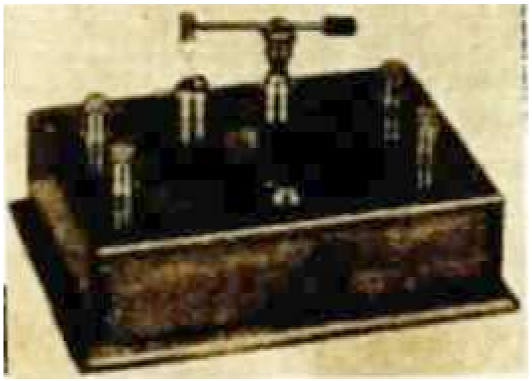

Photograph Of A Crystal Set From 1923

[from Pageant Of The Century - Odhams Press]

Crystal sets were primitive instruments used for 'listening in', and were fitted with the now famous cats whisker. Listening was done though a pair of headphones which meant that only one person could listen at a time.

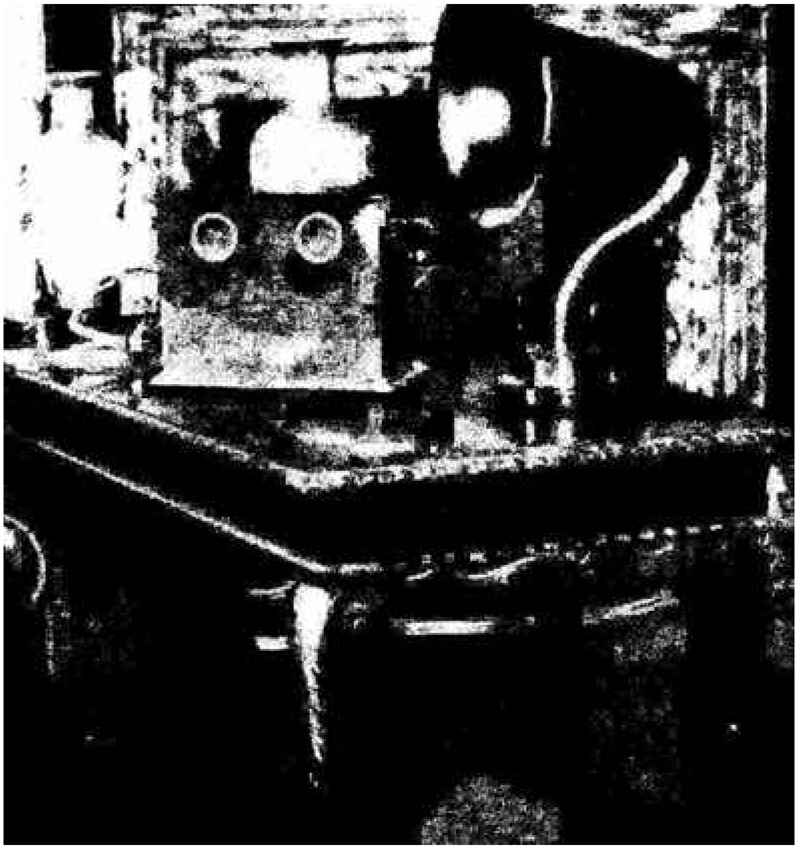

Photograph Of a 1923 Radio

[from Pageant Of The Century - Odhams Press]

Later valve amplification could be added to radios which allowed them to power a loudspeaker so that everyone in a room could listen in.

4 - HOW I BUILT MY CRYSTAL SET

The crystal set is a fun and easy thing to build.

The first thing I had to do was put up an aerial and earth. For the aerial I drilled a hole in the ceiling and fed wire through the hole. Then I went into the attic and took the wire to the other end of the attic, fastening it to rafters as I went. When I got to the other side I fed the wire out of an air vent to the outside of the house. I then went outside tied wire to a nearby tree and joined it up to the wire coming out of the air vent. This gave me an aerial of about 25 metres.

Next the earth. For this I drilled another hole in the ceiling and fed another wire through and out of an air vent at the other side of the house (see previous paragraph). I attached this wire to the metal frame of my swing. This is a good earth because its poles stick deep into the ground.

The coil used enamelled copper wire wound on to a piece of plastic pipe with a diameter of 5 cm. There were 120 turns of wire. Every 10 turns I made a loop on which a crocodile clip could be fixed. See the picture below:

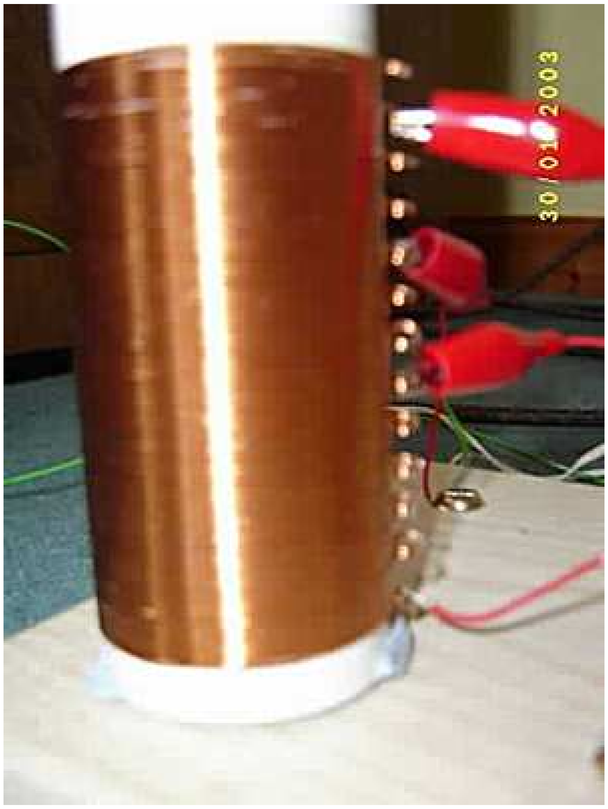

Photograph Of The Coil Showing The Tapping Points & Crocodile Clips

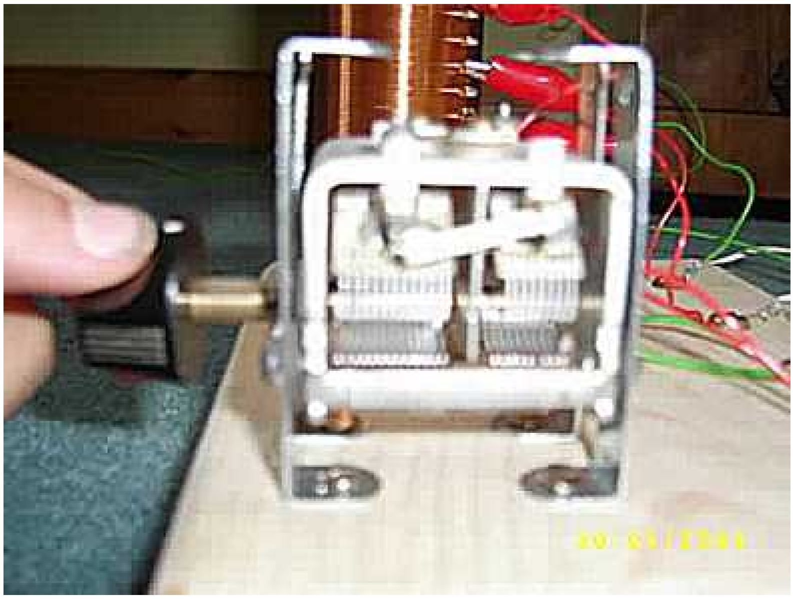

The Tuning Capacitor - Being Carefully Adjusted

For the crystal set itself, I first had to plan what I would be using and where the pieces would be placed (see picture and the plan drawing). I was using the circuit shown above. I built the crystal set on a wooden board about 20cm by 15cm. I drilled 4 holes 4cm apart. I put screws with cup washers in each hole and fastened the wires of the parts of the crystal set and the aerial and earth wires under the cup washers. I fastened the capacitor to the board using Meccano brackets and screws. The coil was fixed to the board using glue. The earpiece plugged into a socket which was fastened to the board using a thunderbolt shaped piece of Meccano.

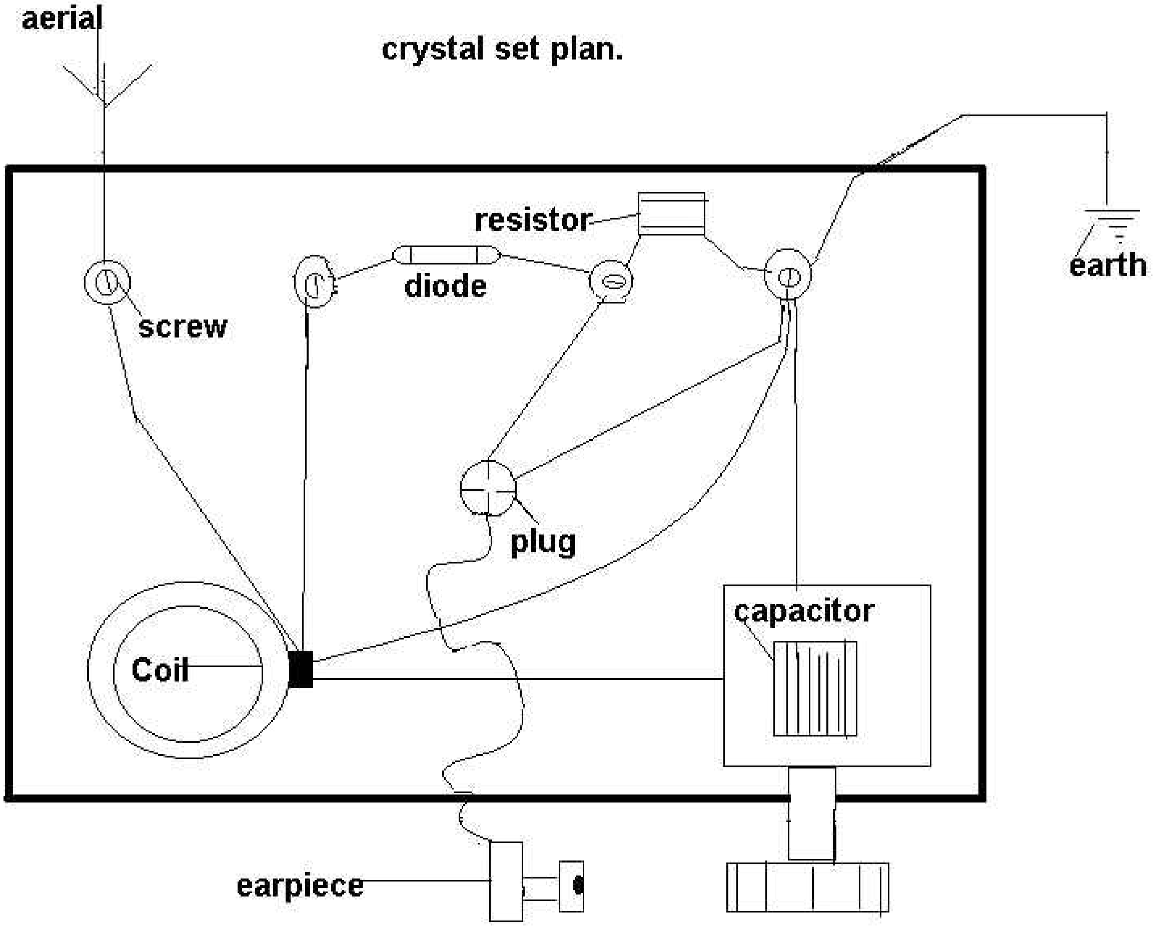

Drawing Showing Layout Of Components

5 - THE RESULTS I OBTAINED

At first all I could hear was a foreign station. But then, I tried late at night and could sometimes hear a station from Florida in English! By tuning the capacitor and moving the crocodile clips up and down the coil I managed to receive a few foreign stations on Short Wave. What I really wanted was to tune in to Medium Wave. I wanted to receive English stations like Radio Scotland and music stations. I tried everything including a longer aerial and using different parts but to no avail. I came to the conclusion that I could not get a strong enough Medium Wave signal in Northerly area of Scotland where we live.

Here is a picture of me operating my crystal set:

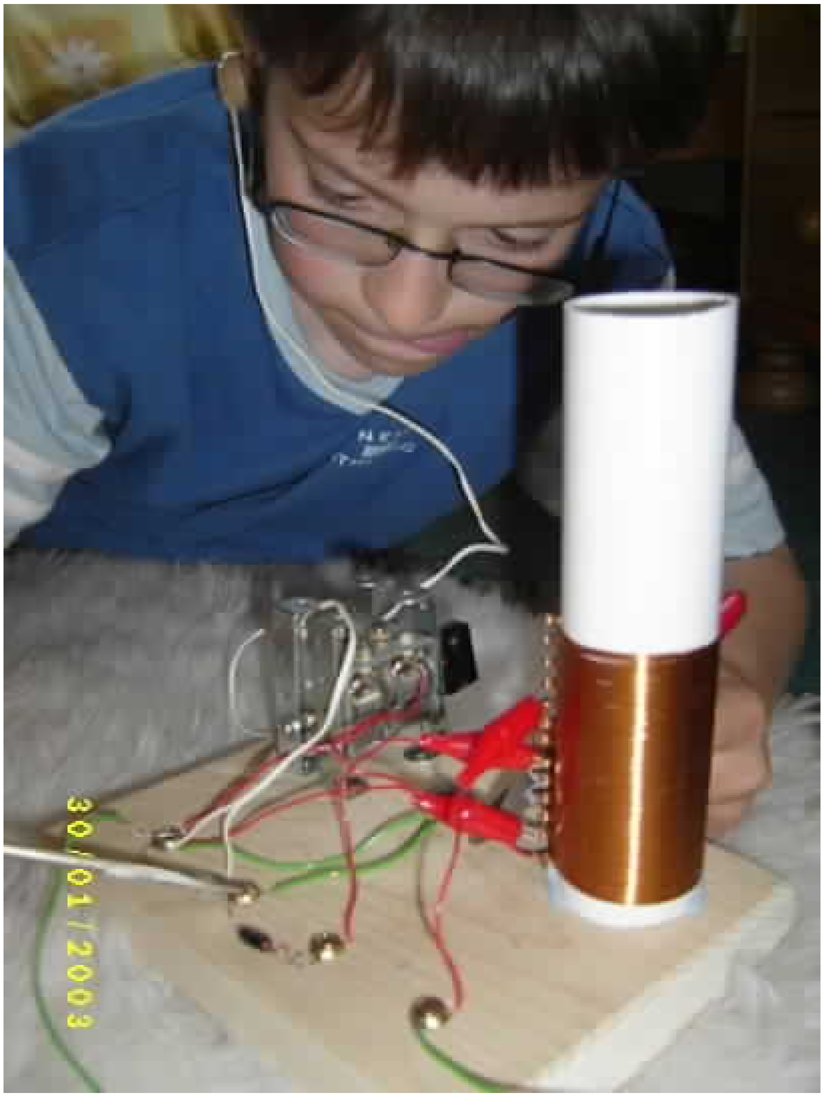

Photograph Showing The Crystal Set In Operation

6 - WHERE I FOUND THE INFORMATION

The History Of Crystal Sets

Book: The Pageant of the Century, Odhams Press, 1933.

www.oldradioworld.de/gollum/dhistor.htm

How I Built My Crystal Set:

Web page describing how to build a crystal set:

http://www.mds975.co.uk/Content/crystalsets2.html (THIS SITE)

Book:

"Making A Transistor Radio" written by Rev. G.C. Dobbs, Ladybird, 1972

Kits Are Available From:

www.midnightscience.com/kits.html

www.oldradioworld.de/gollum/xkits.htm

1N34A.com - crystal radio parts, including litz wire is http://1n34a.com

Other Hobby Websites:

Dave Schmarder's Crystal Set Pages http://www.makearadio.com/crystal/index.php http://www.makearadio.com

"Heart of England Crystal Radio Club" based in Birmingham, West Midlands - http://crystalradioclub.co.uk

More Crystal Set Links here >>

7 - QUIZ QUESTIONS

CRYSTAL SET QUIZ QUESTIONS:

1. What is a crystal set?
2. When were the first crystal sets sold?
3. Name 3 essential parts for a crystal set.
4. What does the earth do?
5. Does the coil pass signals of low frequency or high frequency?
6. How far apart were the screws on my crystal set?
7. What happened when I made the aerial longer?
8. What is this an anagram of: i a c c p r o t a?

(ANSWERS ARE FURTHER DOWN THIS PAGE)

PARTS THAT ARE REQUIRED

1	Wooden Base Panel
1	Coil Former 5cm dia. UPVC Pipe
1	Tuning Capacitor approx 500pF Air-Spaced or Solid Dielectric
1	Reel Of Enamelled Copper Wire for coil
1	Germanium Diode eg OA47 and IN34 which are especially suitable or other alternatives such as OA81, OA90 OA91, 1N94 etc
1	Crystal Earphone
1	47 k Ohm Resistor across earphone
6	½ inch No. 6 Brass Screws and Screw Cups
1	3.5mm mono Jack Socket
	Mounting Hardware (Meccano) for coil, capacitor and jack socket
	PVC covered wire for aerial and earth

A Superb RGD Valve Radio from 1937 - carefully restored in 1994 by Roddie Rankin

With thanks to Kenneth and Roddie Rankin for sending in this excellent article and photographs!

Crystal Sets Part 5 >

QUIZ ANSWERS

1. The first kind of radio
2. 1922.
3. Coil, diode, capacitor, earpiece, aerial, earth.
4. Gets rid of signals you don't want.
5. Low.
6. 4 cm.
7. It did not make much difference.
8. Capacitor.

No AM radio stations or transmitters in your locality or country?

Spitfire
AM Transmitter

Has your local medium wave broadcast station closed or been moved to VHF/FM or Digital? Don't worry. You can still build and experiment with crystal sets and TRF radios by also buying or even building a simple low power AM transmitter. So, not only can you use your crystal sets but you can also run your own radio station that can be heard in and around your home - playing the music or programmes that you want to hear!

SSTRAN AMT3000 : Superb high fidelity medium wave AM transmitter kits from SSTRAN. Versions available for 10kHz spacing in the Americas (AMT3000 or AMT3000-SM) and 9kHz spacing in Europe and other areas (AMT3000-9 and AMT3000-9SM). Superb audio quality and a great and well designed little kit to build: http://www.sstran.com/pages/products.html

POWER
GAIN
MODULATION
COMPRESSION
SSTRAN
AMT3000

http://www.sstran.com/

Other AM transmitters available:

Spitfire & Metzo: Complete, high quality ready built medium wave AM Transmitters from Vintage Components:
http://www.vcomp.co.uk/index.htm Vintage Components offer a choice of the high quality Spitfire and Metzo transmitters:

SPITFIRE AM Medium Wave Transmitter with 100 milliwatt RF output power:

http://www.vcomp.co.uk/spitfire/spitfire.htm

Spitfire
AM Transmitter

METZO AM Medium Wave Transmitter with built in compressor:

Gain
Comp
Mod
Metzo

http://www.vcomp.co.uk/metzo/metzo.htm

AM88 LP A basic AM transmitter kit from North County Radio.
http://www.northcountryradio.com/Kitpages/am88.htm

LINKS:

BOWOOD ELECTRONICS - A friendly, helpful and very speedy source for many of your electronic components at prices that won't frighten your wallet!

THE FOXHOLE and P.O.W RADIOS - Simple crystal set receivers used by soldiers during the war and by prisoners of war (P.O.W.'s).

VINTAGE COMPONENTS - A great resource for crystal sets, components, valve radio kits and medium wave AM transmitters!

6V6 - Electronic Nostalgia and Vintage Components

MORE CRYSTAL RADIO LINKS HERE >>

Crystal Sets (Part1) | Build Your Own Crystal Set (Part 2)

Spider's Web Crystal Set (Part 3) | Experimental Crystal Sets (Part 5) | Crystal Radio Links

^Top Of Page

Amateur Radio
MØMTJ Visit my amateur radio pages MØMTJ

Visit my Amateur Radio Pages >

Home | Contact | Site Map | Radio Stations & Memorabilia | Amateur Radio

Crystal Sets Introduction | Resistor & Capacitor Conversion Tables

www.mds975.co.uk © 2005 - 2014

MDS975 www.mds975.co.uk

CRYSTAL SETS
4
A Crystal Set By Kenneth Rankin

MDS975 www.mds975.co.uk

Home | Contact | Site Map | Radio Stations & Memorabilia | Amateur Radio

Crystal Sets Introduction | Resistor & Capacitor Conversion Tables

Crystal Sets (Part1) | Build Your Own Crystal Set (Part 2)

Spider's Web Crystal Set (Part 3) | Experimental Crystal Sets (Part 5) | Crystal Radio Links

CRYSTAL SETS 4: A CRYSTAL SET BUILT BY KENNETH RANKIN

MDS975.co.uk reader Roddie Rankin kindly sent in the details of a crystal set built by his son Kenneth for a school project which won him great praise and merit. The crystal set is so nicely made and the project so well written that I thought it deserved a special page of its own.

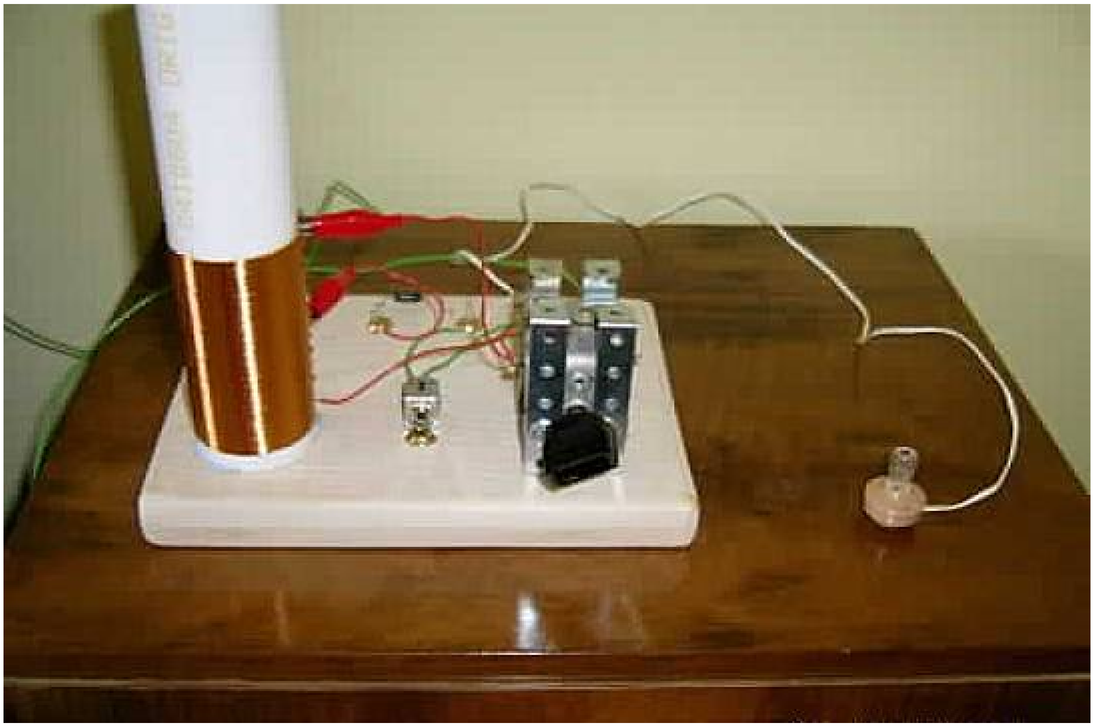

Photograph Of Kenneth Rankin's Rather Super Crystal Set

CONTENTS

1 - What Is A Crystal set?
2 - How A Crystal Set Works
3 - History Of Crystal Sets
4 - How I Built My Crystal Set
5 - The Results I Obtained
6 - Where I Found The Information
7 - Quiz

1 - WHAT IS A CRYSTAL SET?

Perhaps you think, like I did, that a crystal set is just a collection of lovely stones. In fact a crystal set is a very early, simple domestic radio. The photograph above shows the crystal set I built.

The crystal set does not have a battery. It runs completely from the radio waves it picks up in the air. It is absolutely free to run!

Please read on to find out more about this exciting invention!

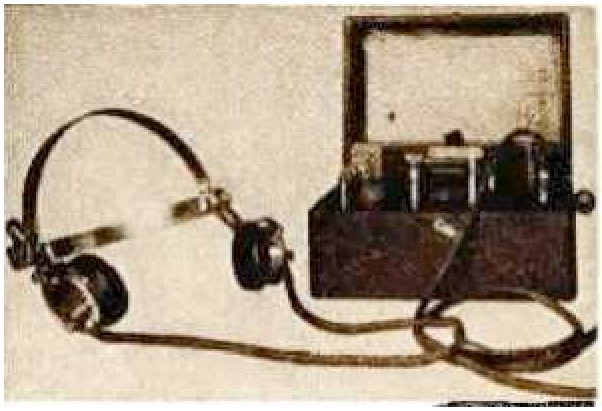

The birth of domestic radio: A deluxe crystal set produced by Marconi for the reception of programmes from the first of the BBC's radio stations, 2LO in London.
[from Pageant Of The Century - Odhams Press]

2 - HOW A CRYSTAL SET WORKS

The diagram below shows the circuit diagram of the crystal set I have built. It consists of a coil (L1), a tuning capacitor (C2), a diode (X1), an aerial, a crystal earpiece and an earth. (Capacitors C1 and C3 are not essential.)

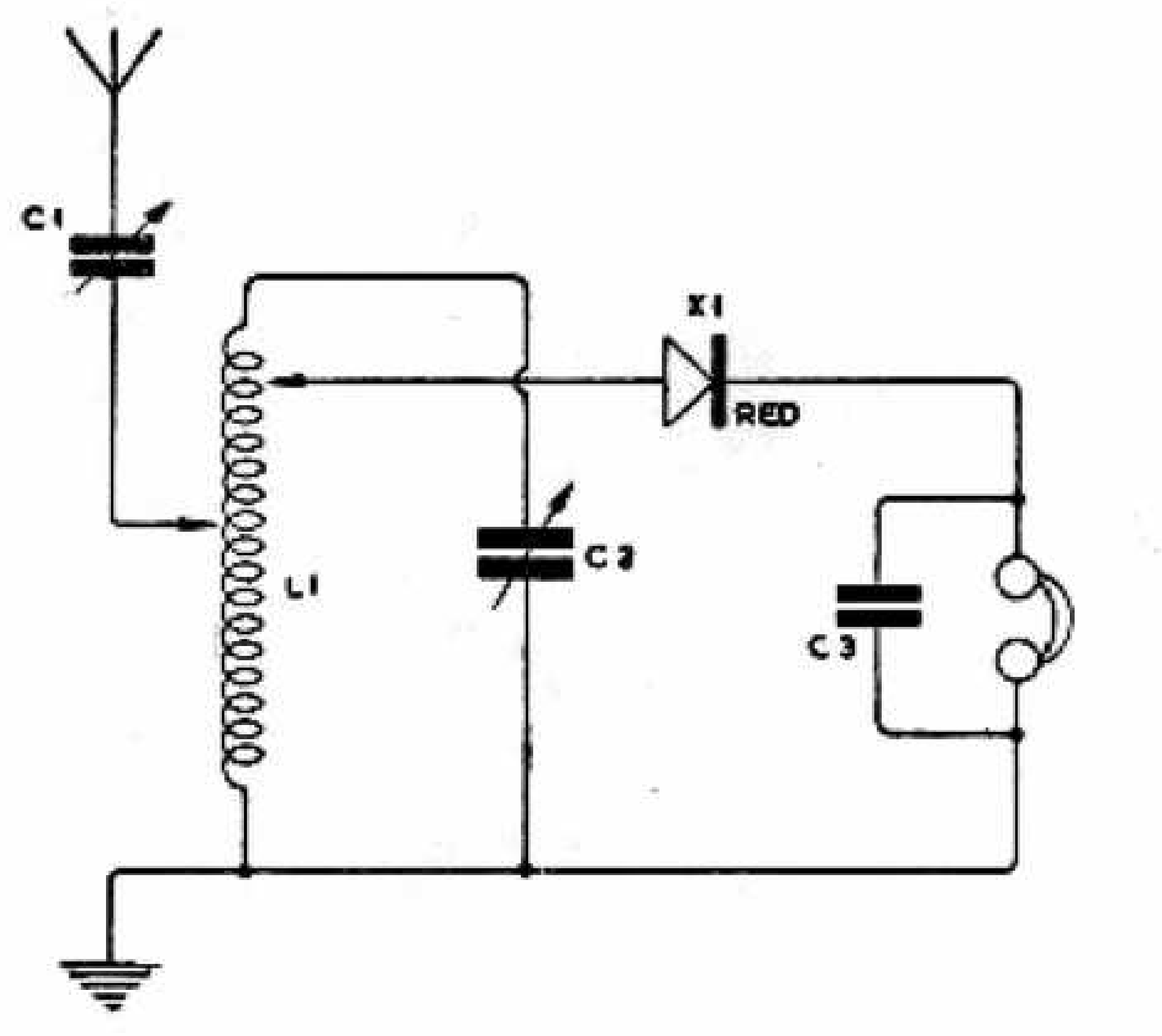

Circuit Diagram Of The Crystal Set

First the aerial picks up radio waves from radio stations all over the world. Then the required station needs to be selected, this is called tuning, and this is achieved with a Tuned Circuit. The Tuned Circuit in the crystal set consists of two items; a component called a variable capacitor and a coil of wire.

By adjusting the value of the capacitor and coil tuning is achieved. The coil taking unwanted signals of low frequency straight through to the earth, which gets rid of them, while the capacitor takes unwanted signals of high frequency straight through to the earth. The remaining signal, in between the high frequency and low frequency signals, is the signal of the required station which is not passed to earth by either the coil or the capacitor. This is the frequency to which the set is now tuned in.

You normally change the frequency to which a set is tuned in to by turning the knob on the tuning capacitor, thus receiving different stations by varying the capacitance of the variable capacitor. The number of turns on the coil can also be changed [called varying the inductance of the coil] and this will give different tuning ranges. A larger number of turns provided longer wavelengths [lower frequencies], while a smaller number of turns of wire provides shorter wavelengths [higher frequencies].

[It is also worth noting that the inductance of a coil can be increased by inserting a component called a Ferrite Rod into the centre of the coil.]

Once tuned into the required station this remaining signal goes to earth by travelling through the diode and earpiece. The diode takes the sound signal out of the radio signal [called demodulation], and passes it to the earpiece. The sound signal travels through the crystal in the earpiece which vibrates the air in the earpiece giving us the sound of the radio station to which you are tuned.

3 - THE HISTORY OF THE CRYSTAL SET

The history of crystal sets begins with the first broadcast stations being opened. This happened in England in 1922. The simplest form of receiver for these broadcasts was the crystal set.

In those early sets they had "cat’s whiskers" instead of diodes. A cat’s whisker was a fine piece of wire which was adjusted to make a suitable contact with a crystal. The crystal would be made of a substance such as coke or galena. An example from 1923 is shown in the picture below. Notice how well made these instruments were.

There were several limitations to the crystal set: it needed a big aerial, an earth connection, the clumsy cat’s whisker, and the weak signal could only be listened to by one person at a time with headphones. Very quickly the crystal set began to be replaced by valve radios with loudspeakers, powered by battery (see the picture from 1923).

In World War II, crystal sets were used by prisoners of war in prison camps to listen to news from home. As they were often badly treated it was hard to get all the right parts to build the set.

By this time crystal sets were no longer used in the home. Instead radios like the attractive 1937 model in the picture were being used.

Nowadays people only use crystal sets as a hobby. There are various different kits available on the market (see websites in the references below).

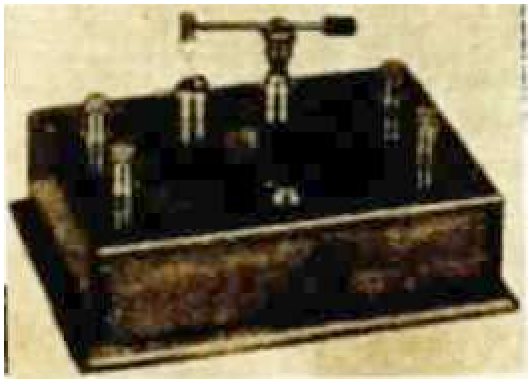

Photograph Of A Crystal Set From 1923

[from Pageant Of The Century - Odhams Press]

Crystal sets were primitive instruments used for 'listening in', and were fitted with the now famous cats whisker. Listening was done though a pair of headphones which meant that only one person could listen at a time.

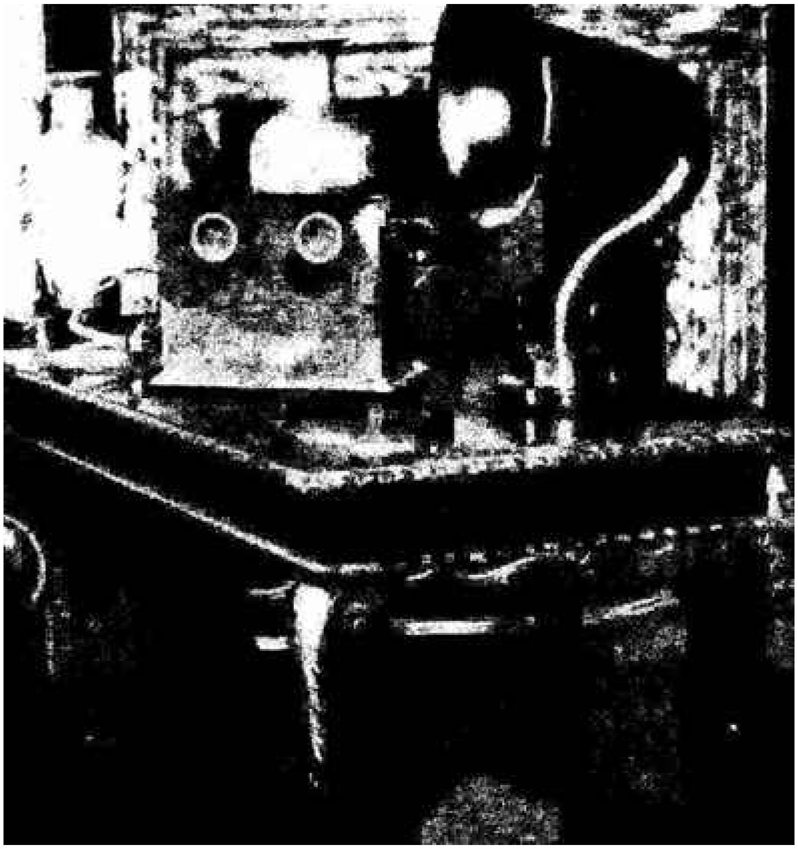

Photograph Of a 1923 Radio

[from Pageant Of The Century - Odhams Press]

Later valve amplification could be added to radios which allowed them to power a loudspeaker so that everyone in a room could listen in.

4 - HOW I BUILT MY CRYSTAL SET

The crystal set is a fun and easy thing to build.

The first thing I had to do was put up an aerial and earth. For the aerial I drilled a hole in the ceiling and fed wire through the hole. Then I went into the attic and took the wire to the other end of the attic, fastening it to rafters as I went. When I got to the other side I fed the wire out of an air vent to the outside of the house. I then went outside tied wire to a nearby tree and joined it up to the wire coming out of the air vent. This gave me an aerial of about 25 metres.

Next the earth. For this I drilled another hole in the ceiling and fed another wire through and out of an air vent at the other side of the house (see previous paragraph). I attached this wire to the metal frame of my swing. This is a good earth because its poles stick deep into the ground.

The coil used enamelled copper wire wound on to a piece of plastic pipe with a diameter of 5 cm. There were 120 turns of wire. Every 10 turns I made a loop on which a crocodile clip could be fixed. See the picture below:

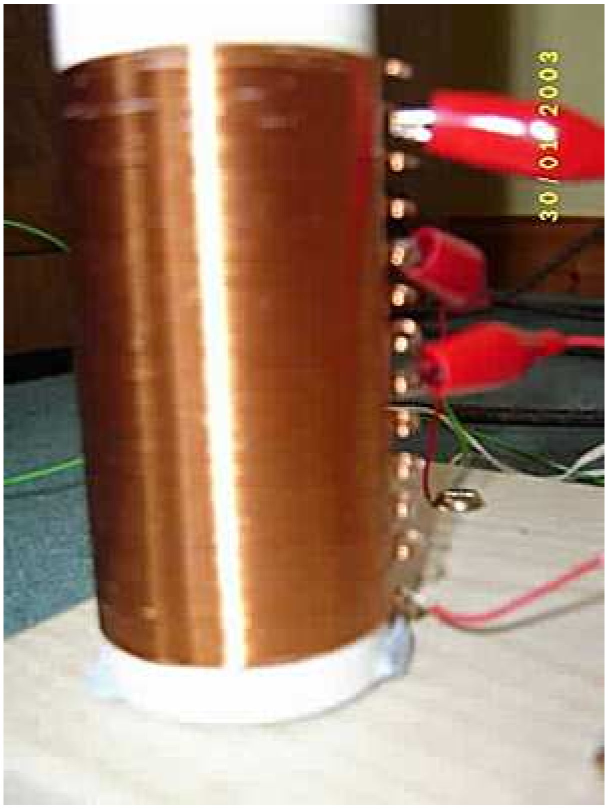

Photograph Of The Coil Showing The Tapping Points & Crocodile Clips

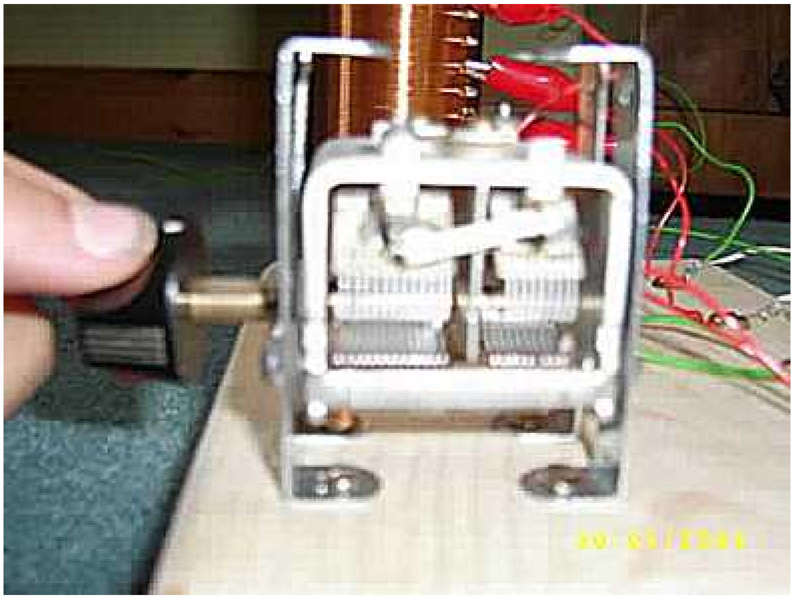

The Tuning Capacitor - Being Carefully Adjusted

For the crystal set itself, I first had to plan what I would be using and where the pieces would be placed (see picture and the plan drawing). I was using the circuit shown above. I built the crystal set on a wooden board about 20cm by 15cm. I drilled 4 holes 4cm apart. I put screws with cup washers in each hole and fastened the wires of the parts of the crystal set and the aerial and earth wires under the cup washers. I fastened the capacitor to the board using Meccano brackets and screws. The coil was fixed to the board using glue. The earpiece plugged into a socket which was fastened to the board using a thunderbolt shaped piece of Meccano.

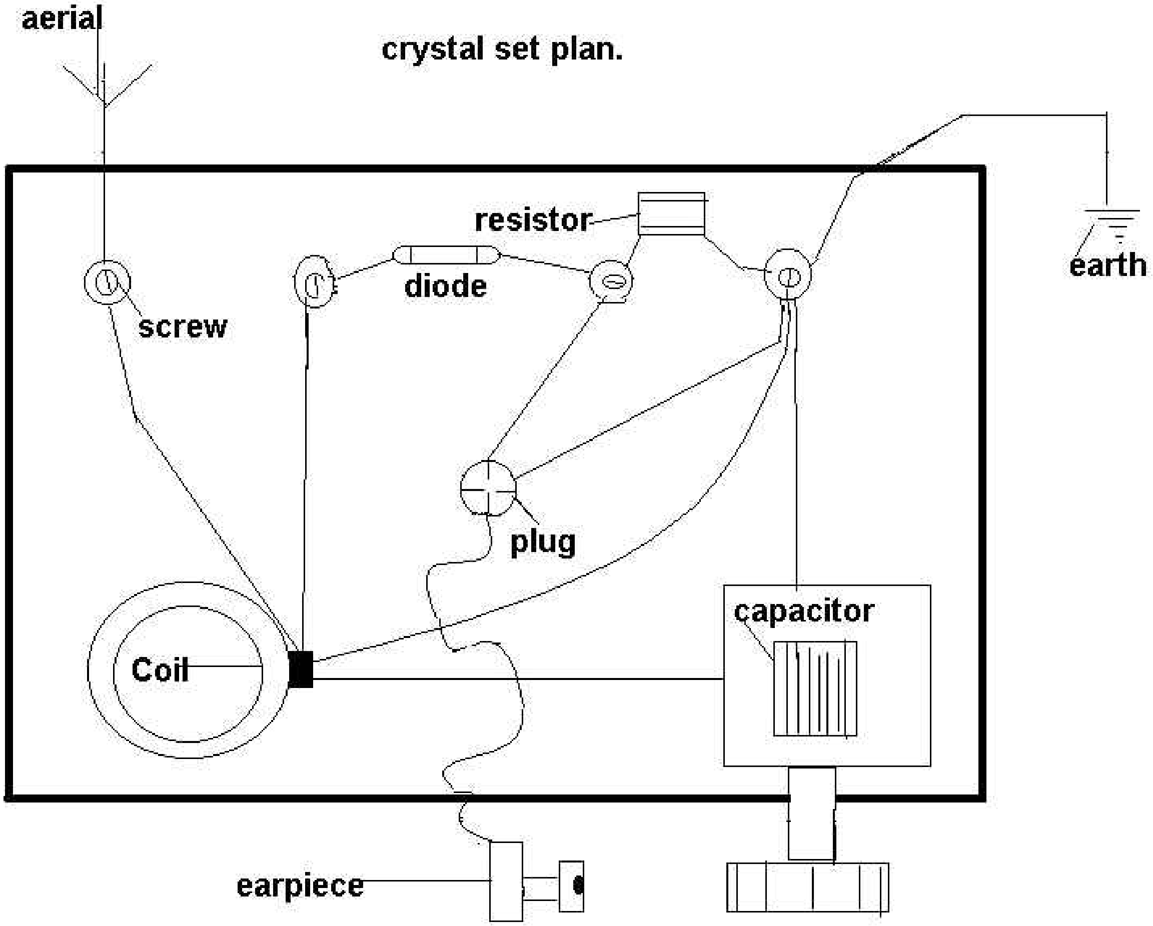

Drawing Showing Layout Of Components

5 - THE RESULTS I OBTAINED

At first all I could hear was a foreign station. But then, I tried late at night and could sometimes hear a station from Florida in English! By tuning the capacitor and moving the crocodile clips up and down the coil I managed to receive a few foreign stations on Short Wave. What I really wanted was to tune in to Medium Wave. I wanted to receive English stations like Radio Scotland and music stations. I tried everything including a longer aerial and using different parts but to no avail. I came to the conclusion that I could not get a strong enough Medium Wave signal in Northerly area of Scotland where we live.

Here is a picture of me operating my crystal set:

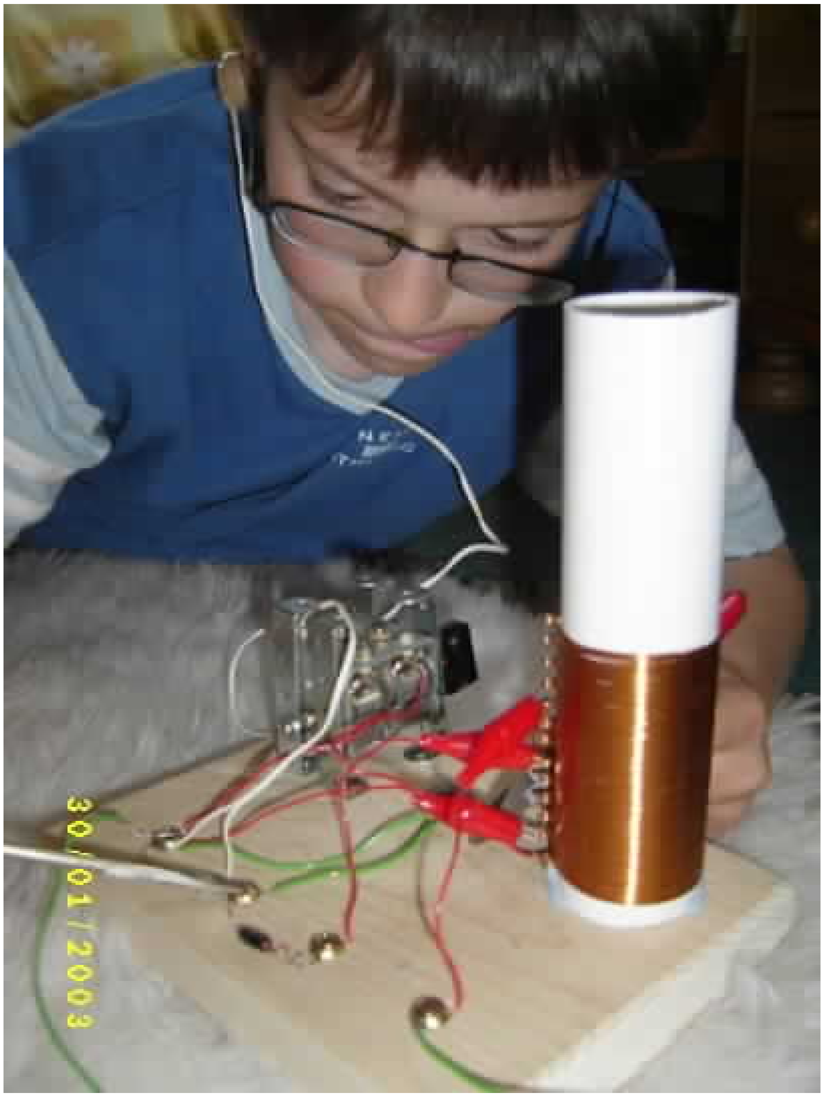

Photograph Showing The Crystal Set In Operation

6 - WHERE I FOUND THE INFORMATION

The History Of Crystal Sets

Book: The Pageant of the Century, Odhams Press, 1933.

www.oldradioworld.de/gollum/dhistor.htm

How I Built My Crystal Set:

Web page describing how to build a crystal set:

http://www.mds975.co.uk/Content/crystalsets2.html (THIS SITE)

Book:

"Making A Transistor Radio" written by Rev. G.C. Dobbs, Ladybird, 1972

Kits Are Available From:

www.midnightscience.com/kits.html

www.oldradioworld.de/gollum/xkits.htm

1N34A.com - crystal radio parts, including litz wire is http://1n34a.com

Other Hobby Websites:

Dave Schmarder's Crystal Set Pages http://www.makearadio.com/crystal/index.php http://www.makearadio.com

"Heart of England Crystal Radio Club" based in Birmingham, West Midlands - http://crystalradioclub.co.uk

More Crystal Set Links here >>

7 - QUIZ QUESTIONS

CRYSTAL SET QUIZ QUESTIONS:

1. What is a crystal set?
2. When were the first crystal sets sold?
3. Name 3 essential parts for a crystal set.
4. What does the earth do?
5. Does the coil pass signals of low frequency or high frequency?
6. How far apart were the screws on my crystal set?
7. What happened when I made the aerial longer?
8. What is this an anagram of: i a c c p r o t a?

(ANSWERS ARE FURTHER DOWN THIS PAGE)

PARTS THAT ARE REQUIRED

1	Wooden Base Panel
1	Coil Former 5cm dia. UPVC Pipe
1	Tuning Capacitor approx 500pF Air-Spaced or Solid Dielectric
1	Reel Of Enamelled Copper Wire for coil
1	Germanium Diode eg OA47 and IN34 which are especially suitable or other alternatives such as OA81, OA90 OA91, 1N94 etc
1	Crystal Earphone
1	47 k Ohm Resistor across earphone
6	½ inch No. 6 Brass Screws and Screw Cups
1	3.5mm mono Jack Socket
	Mounting Hardware (Meccano) for coil, capacitor and jack socket
	PVC covered wire for aerial and earth

A Superb RGD Valve Radio from 1937 - carefully restored in 1994 by Roddie Rankin

With thanks to Kenneth and Roddie Rankin for sending in this excellent article and photographs!

Crystal Sets Part 5 >

QUIZ ANSWERS

1. The first kind of radio
2. 1922.
3. Coil, diode, capacitor, earpiece, aerial, earth.
4. Gets rid of signals you don't want.
5. Low.
6. 4 cm.
7. It did not make much difference.
8. Capacitor.

No AM radio stations or transmitters in your locality or country?

Spitfire
AM Transmitter

Has your local medium wave broadcast station closed or been moved to VHF/FM or Digital? Don't worry. You can still build and experiment with crystal sets and TRF radios by also buying or even building a simple low power AM transmitter. So, not only can you use your crystal sets but you can also run your own radio station that can be heard in and around your home - playing the music or programmes that you want to hear!

SSTRAN AMT3000 : Superb high fidelity medium wave AM transmitter kits from SSTRAN. Versions available for 10kHz spacing in the Americas (AMT3000 or AMT3000-SM) and 9kHz spacing in Europe and other areas (AMT3000-9 and AMT3000-9SM). Superb audio quality and a great and well designed little kit to build: http://www.sstran.com/pages/products.html

POWER
GAIN
MODULATION
COMPRESSION
SSTRAN
AMT3000

http://www.sstran.com/

Other AM transmitters available:

Spitfire & Metzo: Complete, high quality ready built medium wave AM Transmitters from Vintage Components:
http://www.vcomp.co.uk/index.htm Vintage Components offer a choice of the high quality Spitfire and Metzo transmitters:

SPITFIRE AM Medium Wave Transmitter with 100 milliwatt RF output power:

http://www.vcomp.co.uk/spitfire/spitfire.htm

Spitfire
AM Transmitter

METZO AM Medium Wave Transmitter with built in compressor:

Gain
Comp
Mod
Metzo

http://www.vcomp.co.uk/metzo/metzo.htm

AM88 LP A basic AM transmitter kit from North County Radio.
http://www.northcountryradio.com/Kitpages/am88.htm

LINKS:

BOWOOD ELECTRONICS - A friendly, helpful and very speedy source for many of your electronic components at prices that won't frighten your wallet!

THE FOXHOLE and P.O.W RADIOS - Simple crystal set receivers used by soldiers during the war and by prisoners of war (P.O.W.'s).

VINTAGE COMPONENTS - A great resource for crystal sets, components, valve radio kits and medium wave AM transmitters!

6V6 - Electronic Nostalgia and Vintage Components

MORE CRYSTAL RADIO LINKS HERE >>

Crystal Sets (Part1) | Build Your Own Crystal Set (Part 2)

Spider's Web Crystal Set (Part 3) | Experimental Crystal Sets (Part 5) | Crystal Radio Links

^Top Of Page

Amateur Radio

MØMTJ Visit my amateur radio pages MØMTJ

Visit my Amateur Radio Pages >

Home | Contact | Site Map | Radio Stations & Memorabilia | Amateur Radio

Crystal Sets Introduction | Resistor & Capacitor Conversion Tables

www.mds975.co.uk © 2005 - 2014

GALENE 1

commercial gales

The classic commercial galen is of the Vaam type with double basket bottom coil of about fifty turns in parallel to the mica variable from 0-350 pf and is connected in Oudin. With the coil of this capacity we do not get to the short waves because for the resonance we have to calculate also the antenna capacity in parallel.
If we do not want to use a different coil for the high frequencies, it is better to add a socket that puts in series to the antenna a 100pf mica capacitor.

autocostruzione

If we dedicate ourselves to self-construction, the most convenient configuration uses a variable-spooling coil of 80-100 turns on a pipe of eighty millimeters. The wire must be 0.6 mm and the spirals must be spaced about 2 tenths of mm, otherwise we use Litz wire with an external diameter of 0.8-1 mm or the old bell-wire insulated in double cotton , and do not use any variable capacitors.

In the old achievements for the choice of the amount of turns it was done with a sliding slide on the stripped loops. This is not good because you always find at least a couple of coiled turns, which cause losses in the coil.

Another choice was to make, for example, 7 groups of 10 turns, selectable by a switch and a group of 10 turns selectable one by one. This however from too much capacity dispersed through the connections.

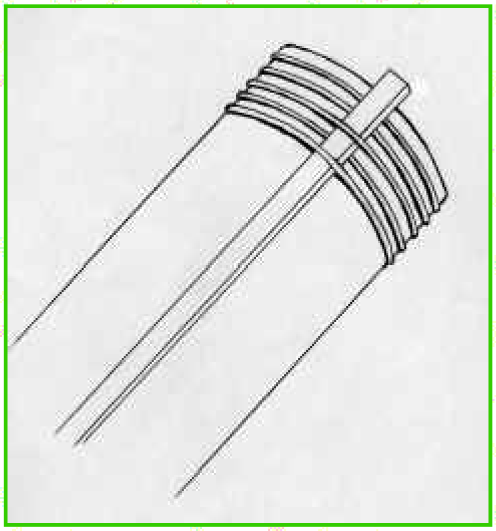

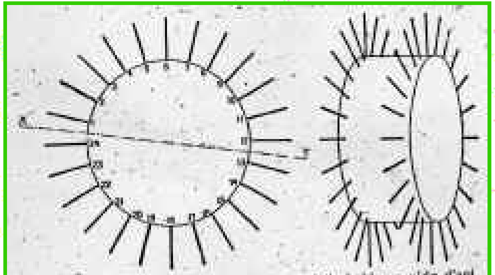

Fig. 124. — Mandrino per l'avvolgimento delle bobine a nido d'api

I chose to make the coil described by wrapping the coils as in the figure and inserting a thin wooden slat under the coils from which I want to make the grip. The pitch of the sockets is variable, dense initially and sparse at the end taking into account interpolation through the use of a small variable of 50 or 100 pf, a value that does little damage, considering the rule that the best yield of the galena is when the tuning is done with the coil only and without capacitors, either in series or in parallel. We will see later the realization of the Tesla circuit but already with Oudin the maximum results are obtained.

some notes

The crystal detection has the property to have a better performance to the strong signals than the weak ones: a weaker signal of 1.4 times in radiofrequency will be 2 times weaker in audio.

This phenomenon leads to a sharpening of the selectivity of the circuit, given that a small dissension is enough to multiply the signal reduction. Except for the case of Milan in the 1920s, where there were two strong stations nearby, the need for a great selectivity is not there. If ever, before the RAI reduced the power of its stations and decimals, a rejector could be needed to eliminate them. I talk about it in my publication IL LIBRO DELLE GALENE by Sandit.

intruders

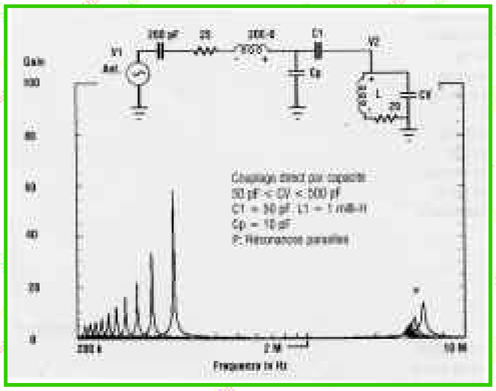

The complaint of finding foreign stations in the evening, which can not be eliminated by tuning, is not due to its fault, but to the intrusion of short-wave stations on our antenna that solicit resonance modes in the antenna-galena system (Virtually every antenna has a resonance on the short waves and the addition of the chord coil fails to eliminate them.

spurious resonances calculated by a French enthusiast.

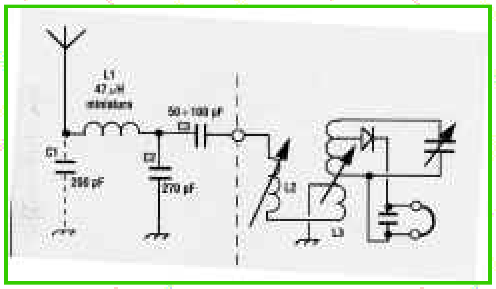
L1
47 µH
miniatura
50 ÷ 100 pF
C3
C1
200 pF
C2
270 pF
L2
L3

2 MHz high-cut filter to eliminate short-wave interference

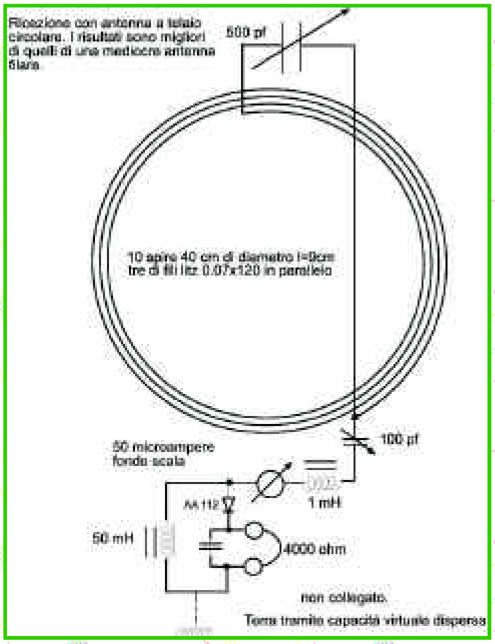
Ricezione con antenna a telaio circolare. I risultati sono migliori di quelli di una mediocre antenna filare.
500 pf
10 spire 40 cm di diametro l=9cm
tre di fili litz 0.07x120 in parallelo
100 pf
50 microampere
fondo scala
AA 112
1 mH
50 mH
4000 ohm
non collegato.
Terra tramite capacità virtuale dispersa

Frame

If we do not have an ideal antenna, but we must adapt to antennas on the balcony of a condominium in reinforced concrete, it may be appropriate to make a frame antenna. Any wired antenna has a much longer effective length than a frame antenna, but it also has a high loss resistance. A frame can have a very low resistance, with a Q even of 1000 and all accounts can transfer more power than a spun like in my case when I receive a double current with the frame. The diode, sensitive to the square of the tensions, works better, at the same power, if the tension is high, compared to a low, so the strong exrtatension of the frame with strong Q is advantageous.
I used a section of a 40 cm diameter wastebasket and skeleton constitution, using the detector I designed. However, a small 1 mH choke is required to eliminate the influence of the FM.
The one illustrated is optimized for the 1368 kHz made with three or better four lengths of conductor, in parallel, placed side by side. The L / C ratio is very delicate and if we want to go down in frequency we have to increase the coils, maybe reducing the wires in parallel, and increasing the capacity to maintain the same ratio that in my realization. For 1MHz we can orient ourselves on 13 turns, below 15-16. The performances are exceptional, except that in an interior of a modern building the reception is reduced to 1/3.

revitalization of an old galena

A simple ice galena in Oudin can become a formidable and sensitive tool without altering it, but simply by connecting the circuit board between the earth terminal of the container and the earth.
This is nothing more than a negative resistance amplifier that has the property of having common input and output and amplify in both directions. Appropriately regulated, it can oscillate in a heterodyne-beat reception circuit or it can stay at the limit of the oscillation, amplifying enormously like a jet device. The two capacitors of the divider may have unsuitable values for certain antennas and may need to be adjusted for stable reception.

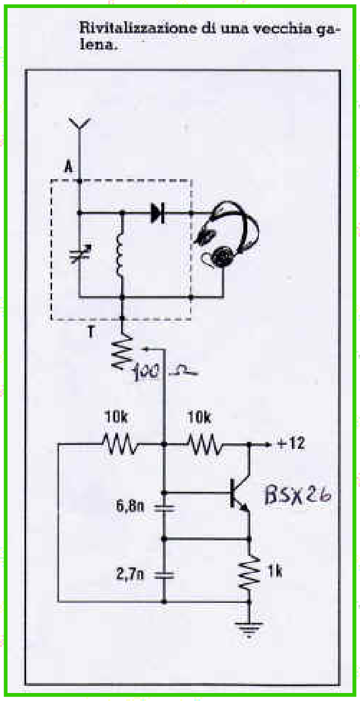
Rivitalizzazione di una vecchia galena.

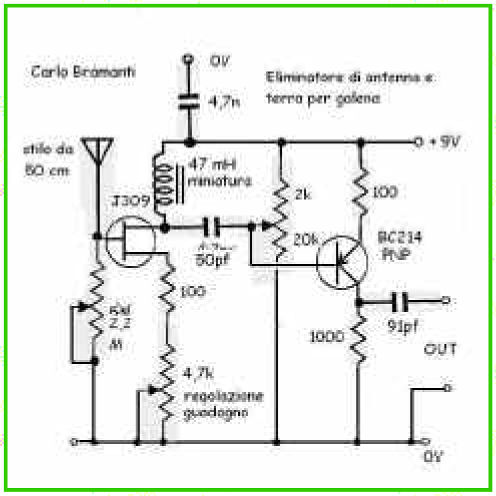

without antenna

If we want to give a demonstration of our galena where there are neither antennas nor earth, something that often happens at the markets, we can use a stylus of about sixty cm, adapting the very high reactive impedance with a high input impedance FET. The earth would become in our body or we can use a short length of hanging wire.

Requires construction experience but it works very well receiving distant stations, since the coil sends back the signal on the antenna with a reaction effect, to be adjusted with one of the potentiometers

help from new technologies

In a 2007 QST article, Bob Cutler introduces a new principle of revelation that is synchronous. In this he is helped by one of those technological miracles which is a MOSFET called EPAD which is satisfied with micropotencies and a supply voltage of 100mV; but something also happens with 10 mV, which is a virtually inaudible signal in galena.
The rectification function is made because the unwanted half-waves polarize the SURCE and inhibit the pitch to the wave itself, while the other half wave passes with good efficiency.

The rectification that we know in the diodes is due to the asymmetry of the curvature of the characteristic and the result is a quadratic trend of the response, ie the strong signal prevails over the weak. In our case, instead, the rectification function will be practically linear with enormous advantage for weak signals.

the original scheme foresees a ferrite rod with 70 turns on which they are wound 8 in the opposite direction, which are those going to the SURCE. I made it with coils in the air and as a result I can listen to the current Florence I, which was transferred far and near the shore at -48 dB, which is impossible with the OA95. With the AAZ17 series you can barely hear .

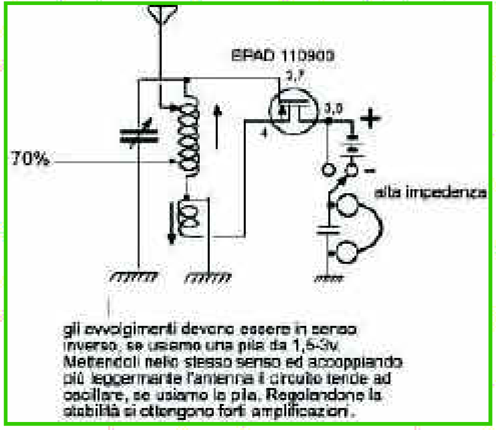

gli avvolgimenti devono essere in senso inverso, se usiamo una pila da 1,5-3v. Mettendoli nello stesso senso ed accoppiando più leggermente l'antenna il circuito tende ad oscillare, se usiamo la pila. Regolandone la stabilità si ottengono forti amplificazioni.

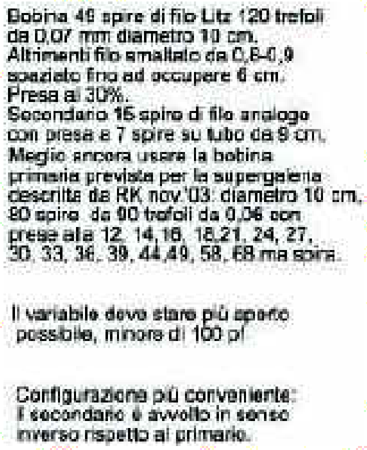

Bobina 49 spire di filo Litz 120 trefoli da 0,07 mm diametro 10 cm.
Altrimenti filo smaltato da 0,8-0,9 spaziato fino ad occupare 6 cm.
Presa al 30%.
Secondario 15 spire di filo analogo con presa a 7 spire su tubo da 9 cm.
Meglio ancora usare la bobina primaria prevista per la supergalena descritta da RK nov.'03: diametro 10 cm. 90 spire da 90 trefoli da 0,06 con prese alla 12, 14,16, 18,21, 24, 27, 30, 33, 36, 39, 44,49, 58, 68 ma spira.

Il variabile deve stare più aperto possibile, minore di 100 pf

Configurazione più conveniente: il secondario è avvolto in senso inverso rispetto al primario.

a p ila

My contribution was the simple addition of a battery in series to the headset: with a 3V alkaline button and an absorption of the microampere, I hear the station in the loudspeaker.
This realization is not a heresy because some diodes, too, see the old carborundum, provided for a feeding, and then, since these components are also available in SMD, you can make a small prism, big as a large diode, with three wires to be connected, without additional components, the ignition occurs when the headset is connected and we have a duration of at least 5 years without intervening.

stolen energy

The EPAD also adapts perfectly to stolen energy solutions, exploiting the rectification of a powerful station or the power supply buzz (see my book).

warning

This circuit can also be implemented with a simple Oudin by connecting to an intermediate socket. In this

in case the turns have the same direction and without power supply there are no problems, but with the battery it becomes unstable, except to insert a resistance in series with the headphone. In this way the gain can be greatly increased, but becomes sensitive to the antenna characteristics. These components should be known to handle, inserted when everything is already on the circuit and do not do so many acrobatics. The EPAD is not easy to find, I took it from MOUSER USA to $ 1.91 each.

a big plus

This circuit, at least in my realization, does not suffer from intrusion by the short waves: in a direct comparison using an AAZ 17, in the dark, Florence I from Coltano becomes unlistenable because of these, while with the synchronous revelation I listen very well.
The battery can be 1.5 V or 3 volts, lithium button. The insertion of the battery allows listening in high impedance speaker.

Experimental realization : the coil has 5 cm of diameter, 80 turns of silver wire of 1mm arranged separated on two combs so to occupy 25 cm of length. Inside there are three 20 cm ferrite rods. It should also be provided that the selected crocodile does not come into contact with other coils . Capacitors are essential.

The performances are formidable!

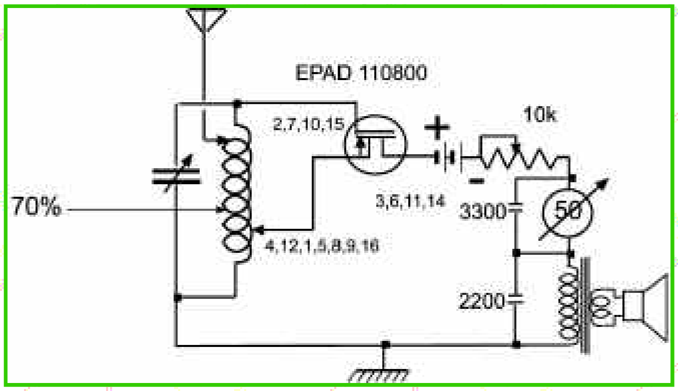

EPAD 110800

This has the same functions as the previous one except that in the casing there are four mosfets that must be connected in parallel: gate = 2,7,10,15 surce = 4,12 drain = 3,6,11,14. Terminals 1,5,8,9,16 can be grounded or surce. The pin 5 is the return of the protection diodes. The absorption passes to about fifteen microampers, absorbed only when the headset is connected, more than double the previous circuit, but the power increases a lot allowing a very intense listening in high impedance speaker or with adapter transformer.

A frame antenna

For the 657 kc I made a frame antenna on the project described: 23 turns wrapped on the basket 40 cm in diameter using three 120 x0.07 litz wires, connected carefully placed side by side wrapping the backhoe uniformly to occupy about twenty cm. The grip of the surce is 1-2 turns if the battery is used, otherwise with 6 turns. The performances are equivalent to my row. Coltano also feels without feeding.

stolen feed

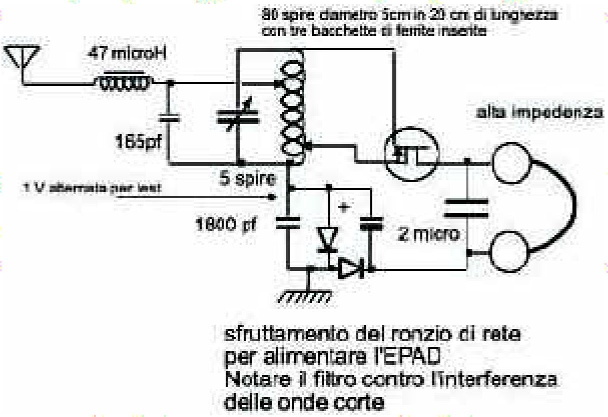
80 spire diametro 5cm in 20 cm di lunghezza
con tre bacchette di ferrite inserite
47 microH
165pf
alta impedenza
5 spire
1 V alternata per test
1800 pf
+
2 micro

sfruttamento del ronzio di rete
per alimentare l'EPAD
Notare il filtro contro l'interferenza
delle onde corte

The low power requirement of the EPAD calls for a stolen power supply. Of course we can not steal from the station we listen to, otherwise we would have invented the perpetual motion, but you can take advantage of the short wave stations holes, in the evening, or the network buzz, always. For the network hum it is sufficient to insert a 5000 pf capacitor between the base of the coil and the ground and rectify. Unfortunately, my 200 pf antenna at 50 hz present 20 megoohm of reactance and does not make it to power the complex, but other cases could be more favorable.

In the figure, a system that gave me positive results.

With the circuit shown my modest antenna results just enough, but in other cases it could give excellent: if you connect the test point to a chassis of some instrument connected to the network the voltage rises rapidly and gives excellent results. Excellent would also be with the light cap. Unfortunately, even if the synchronous detection is much more resistant to the interference of short waves than the traditional diode, in summer the weak Coltano veins me obliterated by those disturbances. So you can see in the diagram the introduction of the basic pass filter that goes well and does not attenuate. Careful to reproduce it faithfully as it is easy to get confused with links.

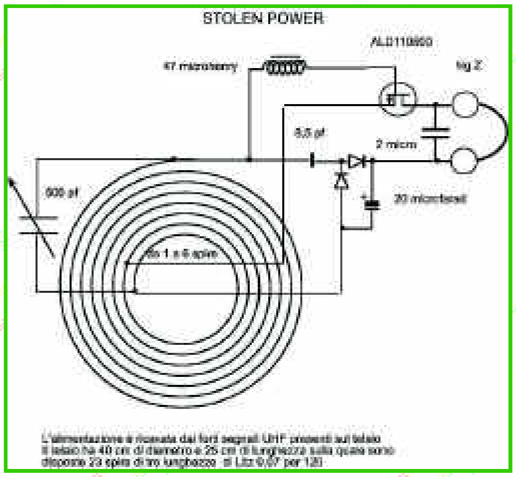
STOLEN POWER
ALD110800
47 microhenry
hg Z
6,5 pf
2 micro
500 pf
20 microfarad
da 1 a 6 spire
L'alimentazione è ricavata dai forti segnali UHF presenti sul telaio
Il telaio ha 40 cm di diametro e 25 cm di lunghezza sulla quale sono
disposte 23 spire di tre lunghezze di Litz 0,07 per 120

Note

I have been suggested to make an ideal rectifier for the detector made with one of the modern integrated circuits and the diode in reaction, as we find in the technical literature. I do not know if it's in our case better than normal diodes.

I tried an expensive zero bias diode from 35 euros but it does half of the AZ17.

go to galena 2

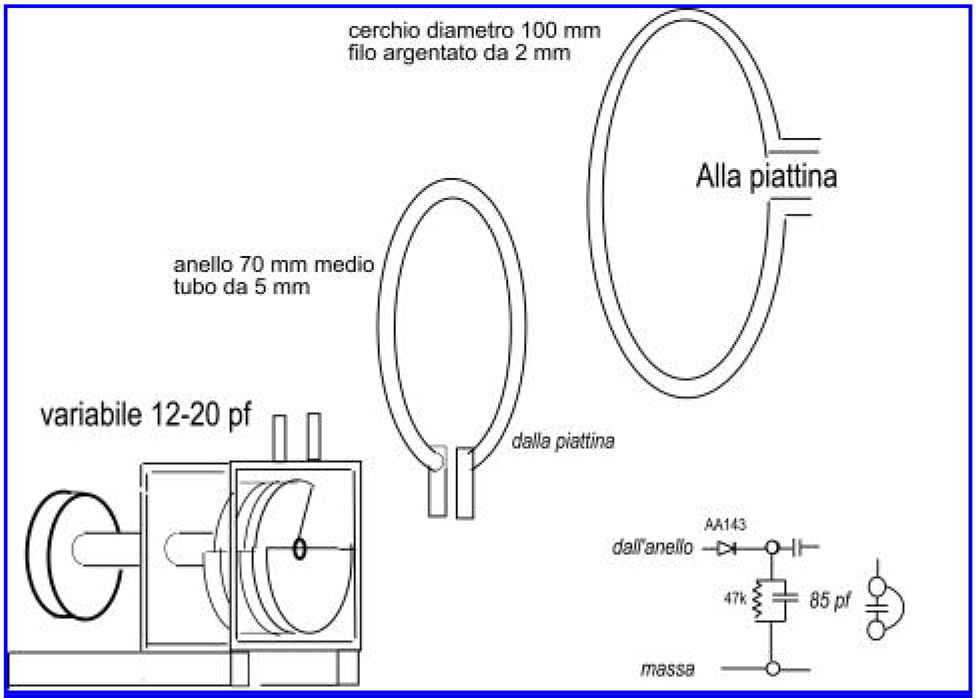
cerchio diametro 100 mm
filo argentato da 2 mm
Alla piattina
anello 70 mm medio
tubo da 5 mm
variabile 12-20 pf
dalla piattina
AA143
dall'anello
47k
85 pf
massa

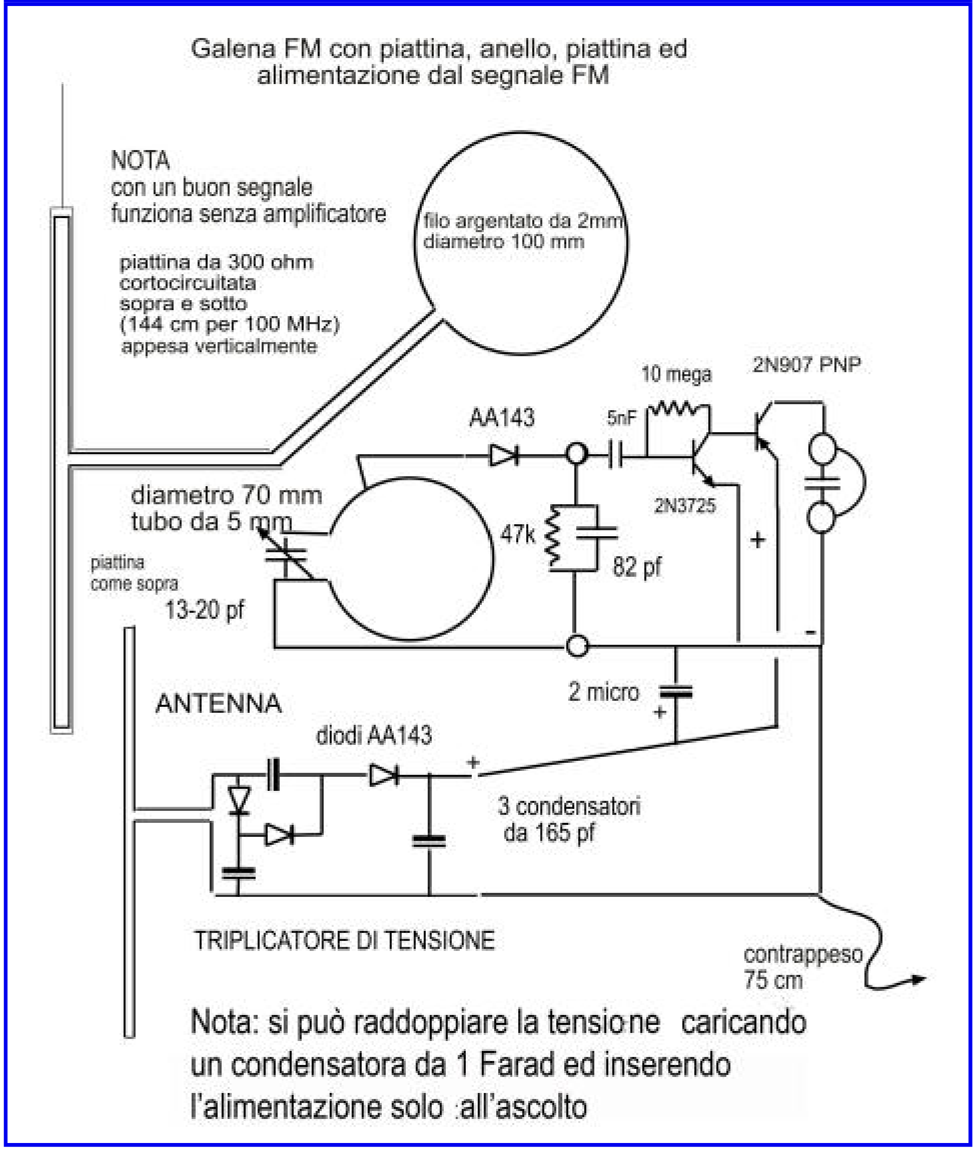
Galena FM con piattina, anello, piattina ed alimentazione dal segnale FM

NOTA
con un buon segnale
funziona senza amplificatore

piattina da 300 ohm
cortocircuitata
sopra e sotto
(144 cm per 100 MHz)
appesa verticalmente

Nota: si può raddoppiare la tensione caricando un condensatora da 1 Farad ed inserendo l'alimentazione solo all'ascolto

filo argentato da 2mm
diametro 100 mm
10 mega
2N907 PNP
AA143
5nF
2N3725
diametro 70 mm
tubo da 5 mm
47k
82 pf
+
piattina
come sopra
13-20 pf
-
2 micro
+
ANTENNA
diodi AA143
+
3 condensatori
da 165 pf
TRIPLICATORE DI TENSIONE
contrappeso
75 cm

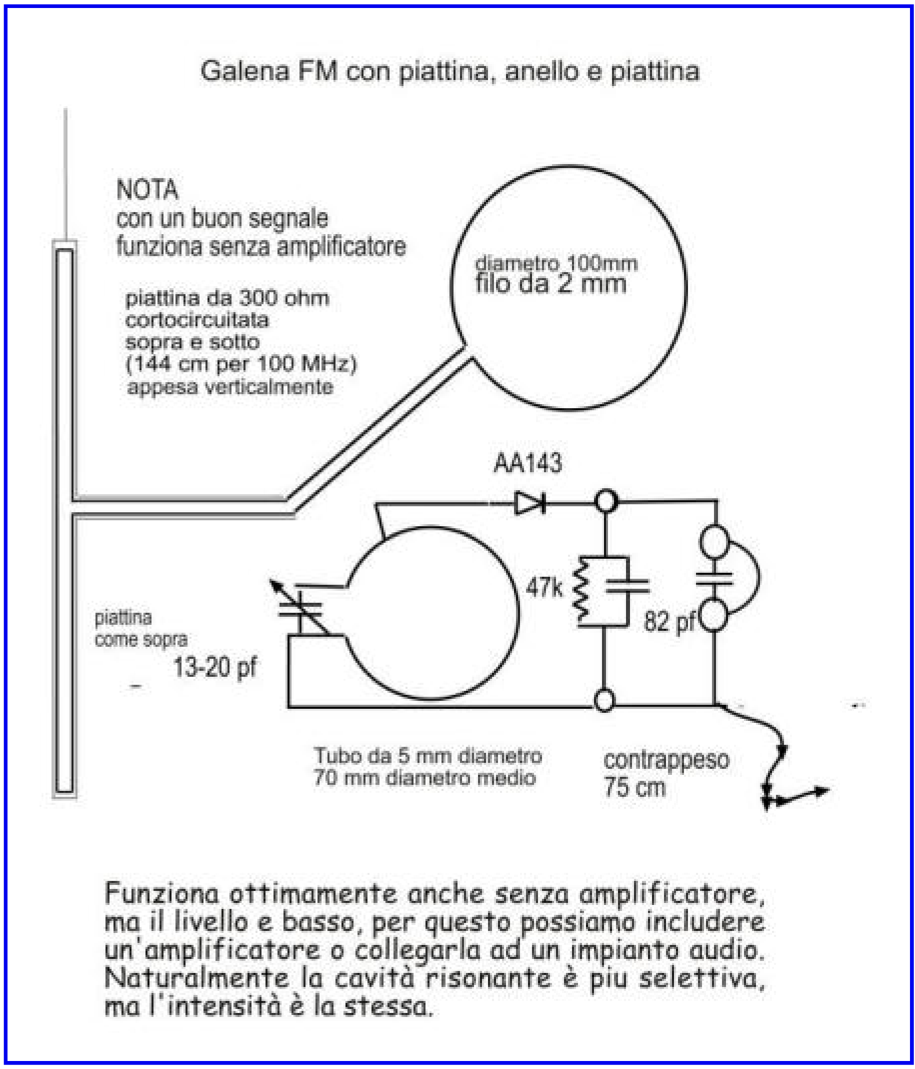

Galena FM con piattina, anello e piattina

NOTA
con un buon segnale
funziona senza amplificatore
piattina da 300 ohm
cortocircuitata
sopra e sotto
(144 cm per 100 MHz)
appesa verticalmente
diametro 100mm
filo da 2 mm
AA143
47k
82 pf
piattina
come sopra
13-20 pf
Tubo da 5 mm diametro
70 mm diametro medio
contrappeso
75 cm

Funziona ottimamente anche senza amplificatore, ma il livello e basso, per questo possiamo includere un'amplificatore o collegarla ad un impianto audio. Naturalmente la cavità risonante è piu selettiva, ma l'intensità è la stessa.

Electronic detector

There are various solutions for achieving revelation in a galena. Several are electronic. If we want to replace the detector with an already created galena, see the VAAM or the RUMA we have the problem that the electronic detector, even with a built-in battery, has three connection wires: antenna, earth and headphones. If we have only two clamps like the traditional detector, a solution must be found to transform the quadripolo into a dipole. The problem that arises is the return to ground of the headset, meaning that of the amplifier that is the negative pole of the power supply.

This is done so that the return occurs through the antenna coils and a radio-frequency stop coil. In this way the device can be inserted in place of the traditional detector.

In the first figure a counterweight connected between the diode and the rheostat can be useful

NOTE

There may be problems with the condenser on the headphone which it is better not to put on. Also the headset used must be low impedance, about 600 ohms. With a 4000 ohm headset, it may not work

in the second figure we see the transistor construction and below that with the integrated 7642

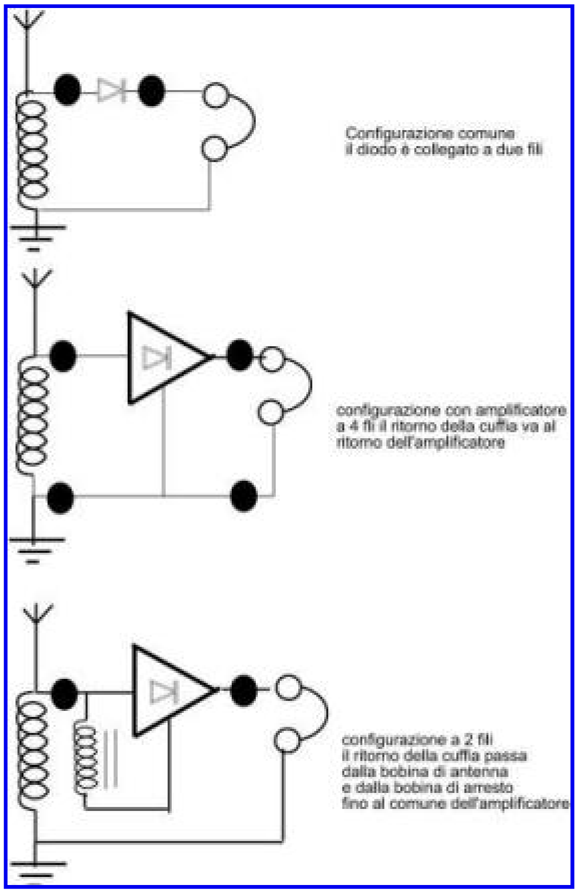
Configurazione comune
il diodo è collegato a due fili
configurazione con amplificatore
a 4 fili il ritorno della cuffia va al
ritorno dell'amplificatore
configurazione a 2 fili
il ritorno della cuffia passa
dalla bobina di antenna
e dalla bobina di arresto
fino al comune dell'amplificatore

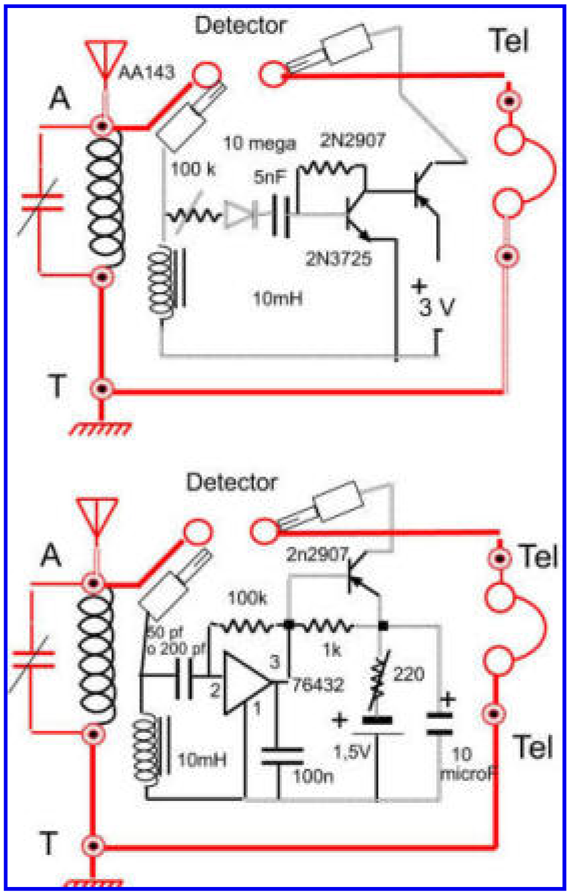
Detector
AA143
A
Tel
10 mega
2N2907
100 k
5nF
2N3725
10mH
+
3 V
T
Detector
A
2n2907
Tel
100k
50 pf
o 200 pf
1k
3
2
1
76432
220
+
1,5V
+
10mH
100n
10
microF
Tel
T

For the solution with integrated 7642 see the supplement for reception am

Reiettore

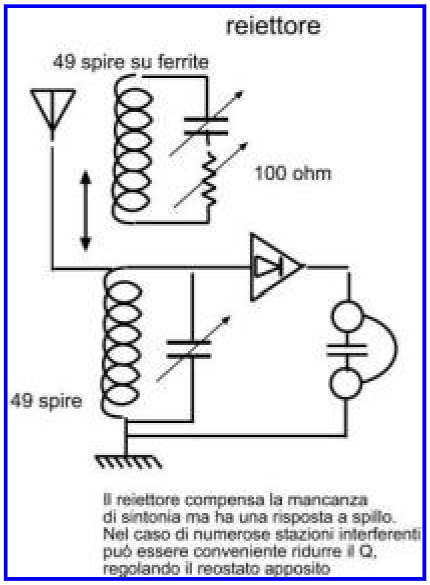
reiettore
49 spire su ferrite
100 ohm
49 spire
Il reiettore compensa la mancanza
di sintonia ma ha una risposta a spillo.
Nel caso di numerose stazioni interferenti
può essere conveniente ridurre il Q,
regolando il reostato apposito

The weakening of the medium wave stations no longer allows us to experiment with the entrance circuit, the corner stone of the galena. We had already lost the thrill of exploring with the cat whisker, using the more stable germanium diode AA143 or AAZ 119, but that had made the nickname "galena" improper. At this point the electronic revelation that brings us back to the experimentation of the input circuits and of the coupling principles that are the basis of the radiotechnics.

The evening disturbances of nearby stations and even those of short waves are so toned. In the evening at the spectrum analyzer I see the local station at -40 dB, a densume of stations between 1400 and 1600 kHz, at the local level or higher, the waves of the 6 MHz range, with many stations always at the level of my local, then something about 12 MHz. After increasingly weak signals. Unfortunately, if a simplified calculation our input resonates on a medium wave station, a deeper calculation appear answers on the short waves that, if they happen on 6 MHz, interfere with our listening, thwarting our efforts to obtain selectivity ..

Already in the '15s in the military receivers did not look so much at the selectivity of the entrance, since the stations were not so many and you could eliminate the disturbers by centering them with the rejector. This can still be done in particular in commercial gales that suffer a lot from interference. : take a ferrite of a radiolina, let's put the classic variable at the ends of the windings and bring it closer to the coil, generally honeycomb, centering with the condenser the area where the disturbances appear. Note that the rejector can also be used for tuning by keeping the main capacitor fixed. It is an interesting experimentation

Something better

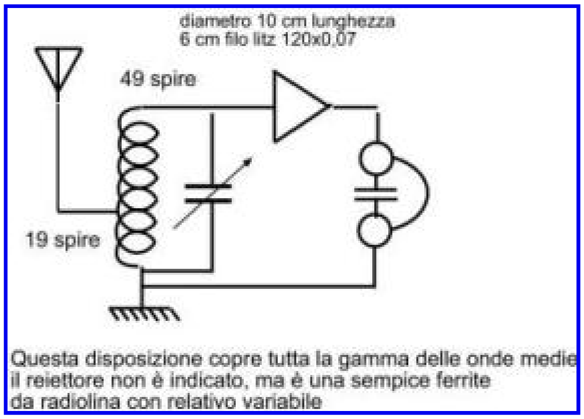
diametro 10 cm lunghezza
6 cm filo litz 120x0,07
49 spire
19 spire
Questa disposizione copre tutta la gamma delle onde medie
il reiettore non è indicato, ma è una sempice ferrite
da radiolina con relativo variabile

What I recommend is a cylindrical bobbin of 10 cm in diameter with 49 turns of Litz thread of 120 strands of 0.07, or of the insulated wire of 0.85-0.90 mm separating if turns with a fishing line. Then two sockets at 1/3 and 2/3 of the winding. Col Litz overcomes the Q of 400 compared to 180 of the normal coils. With that the performance increases a lot 'and the interference decreases. Better if you make the "supergalena" from my publication "Il Libro delle Galene". The condenser on the cuff may be advantageous or not depending on the case.

electronic detector

With the integrated 7642 or MK 484 of the described electronic detectors, reception is very good, but AVC, even if it eliminates interference, makes experimentation more difficult. Then I found several faulty integrals. With the NPN-PNP arrangement, experimentation is possible again and the sensitivity performances are similar to those of the integrated and the loudest volume.

GALENA OMNIBUS

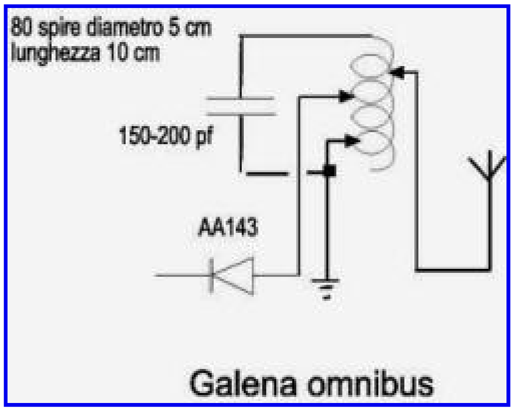

THEat recent reductions in RAI stations in the ether, it did not discourage galene enthusiasts, rather it invited them to improve their antennas and perfect their circuits. During the day, only the local station is received, often with difficulty, so selectivity is not important and a good performance circuit like Oudin is enough. In the evening, however, the entry of many foreign stations confuses reception and requires greater selectivity. The maximum would be obtained with a Tesla arrangement, but it is complex. A circuit that is ideal for selectivity and sensitivity is the Oudin in a three-cursor omnibus configuration. One serves to partialize the coil with reference to ground and fixed condenser, another serves to optimize the socket for the diode, finally the ideal selectivity is obtained with the antenna cursor.

The coil is cylindrical, 5 cm in diameter and 10 in length. The coils are 80 of the largest possible wire, enamelled or, better, with double-cotton insulation that must be peeled corresponding to the sliders.

The principle

The coil of the galena itself can have a Q from 200 to 400, taken alone. If we apply the antenna and the earth to its ends, we transfer the entire antenna loss resistance, that is, about sixty ohms in series to the resistance of the coil loss that is of very few ohms. Thus the Q passes to a fortnight, with consequences on selectivity. If instead we connect the antenna to an intermediate socket, we compensate for the reduction of the signal due to the divider, but we increase the Q by restoring the original overvoltage until reaching an optimal value for both. By reducing the coils on the antenna again at some point we will have the maximum selectivity but less signal.

Another cause of damping of the Q is the load due to the detector. Also for that the grip on the reel should be optimized. This for two cursors. The third is to expand the range of frequencies received and to optimize everything.

Galena FM and flat

My approach to FM was to derive from the strong signal present a little bit of continuous voltage to amplify a galena amplifier to stolen power. The antenna that gave me the best results at the time was the J antenna that I described from various parts.

FM revelation

Since revealing the stations on the side of the selectivity curve it was also possible to listen to FM stations, I used it just for that. Unfortunately, as commonly stated, the intensity of the audio obtained is very poor and it is better to amplify than low. Using resonant cavities I was able to listen comfortably even without amplification, but the problem was to move the antenna away to place it in a place with more signal. Unfortunately I did not manage to couple it with a cable, except for high losses, higher than the natural ones of the cable.

A flat

I then resumed the TV plate dipole that had given me bad results in the beginning: the problem was that, by agreeing to the iconography and what is studied at school, I kept it horizontal. Since, instead, the FM broadcasts are usually vertical polarization, I hung vertically, with the horizontal output plate, and I obtained excellent results such as to allow me to listen, slowly but comfortably without amplification. In the case of the cavity, which has a 50 ohm unbalanced input, the situation improved with a Balun 4: 1.

Stolen feed

Given the excellent results that gives me a circuit to a transistor that I normally use in the AM galena with stolen power, and the high continuous voltage that you can now get (almost 1 volt, on my balcony), I tried to apply it to the galena FM . It already works by connecting on the same signal plate, but the results vary during the day. Since it is not a drama to hang a second band, I used this solution for feeding, with excellent results.

Results

Generally this solution gives me 0.85 Volt and 100 microampers on 400 ohms (the internal resistance of the power supply is about 8000 ohms). It is ineffective the solution of the condenser of 0.1 Farad that I was recharged with detached headphone in the case of 50 Hz in AM (the internal resistance was 15 Megaohm, so the drop given by the absorption was high, now not).

Easier

Simple coil

Applied the ribbon antenna in the old single coil FM galena, the strongest stations felt comfortably without amplification. Better still to stolen power, except in this case the energy can not be made from the signal antenna, but a specific one is needed.

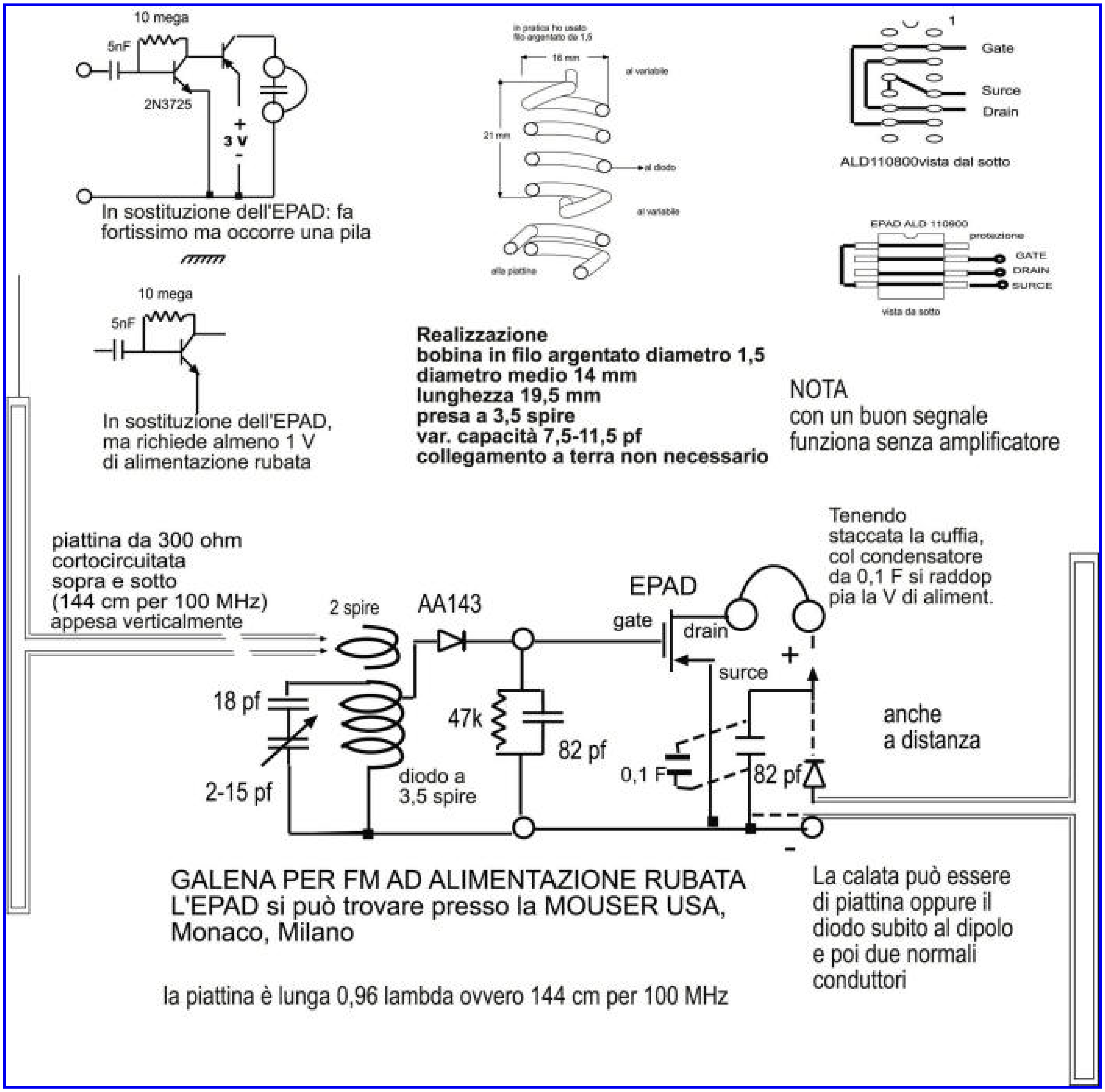
10 mega
5nF
2N3725
+
3 V
-
in pratica ho usato filo argentato da 1,5
16 mm
al variabile
21 mm
al diodo
al variabile
alla piattina
1
Gate
Surce
Drain
ALD110800vista dal sotto
EPAD ALD 110800
protezione
GATE
DRAIN
SURCE
vista da sotto
10 mega
5nF
2 spire
AA143
EPAD
gate
drain
surce
18 pf
2-15 pf
47k
82 pf
0,1 F
82 pf
+
-
diodo a 3,5 spire
anche a distanza

In sostituzione dell'EPAD: fa fortissimo ma occorre una pila

In sostituzione dell'EPAD, ma richiede almeno 1 V di alimentazione rubata

Realizzazione
bobina in filo argentato diametro 1,5
diametro medio 14 mm
lunghezza 19,5 mm
presa a 3,5 spire
var. capacità 7,5-11,5 pf
collegamento a terra non necessario

NOTA
con un buon segnale funziona senza amplificatore

piattina da 300 ohm cortocircuitata sopra e sotto (144 cm per 100 MHz) appesa verticalmente

Tenendo staccata la cuffia, col condensatore da 0,1 F si raddoppia la V di aliment.

GALENA PER FM AD ALIMENTAZIONE RUBATA
L'EPAD si può trovare presso la MOUSER USA, Monaco, Milano

La calata può essere di piattina oppure il diodo subito al dipolo e poi due normali conduttori

la piattina è lunga 0,96 lambda ovvero 144 cm per 100 MHz

Carlo Bramanti December 2012

Galena FM with cavities

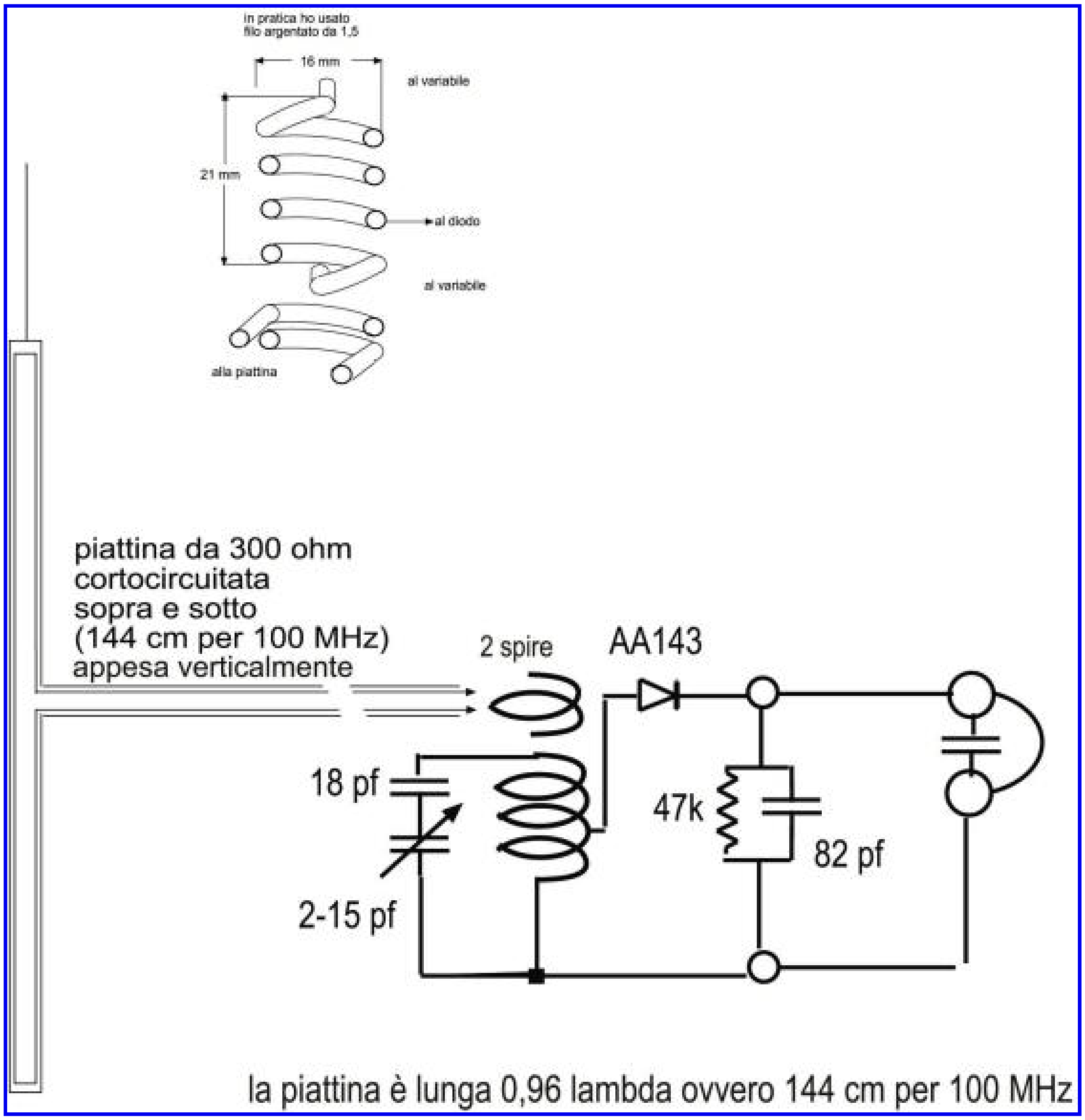
in pratica ho usato
filo argentato da 1,5
16 mm
al variabile
21 mm
al diodo
al variabile
alla piattina
piattina da 300 ohm
cortocircuitata
sopra e sotto
(144 cm per 100 MHz)
appesa verticalmente
2 spire
AA143
18 pf
2-15 pf
47k
82 pf
la piattina è lunga 0,96 lambda ovvero 144 cm per 100 MHz

With a dedicated plate for power it is even better, but it works even without the amplifier.

An amplified frame antenna

If we can build a frame antenna of a very high Q, we can compete with a mediocre line, but the signal is always little and there is nothing to be scandalized if we try to amplify.

In order not to remain on the cattle with a common amplifier, we experiment a solution with a negative resistance amplifier, which also has a considerable didactic implication. These are very particular amplifiers that present the input at the same point of the output and are able to amplify in both directions. For this they were used in telephone lines ...

The criterion is to present itself as a negative resistance, that is to avoid the resistances of the circuit eliminating the losses and even bringing the stage towards the oscillation. In the purely radio field we see it applied to the Q multipliers on the average frequency of the 50s receivers.

We have already used it in the galena with wired antenna, with great success., See the figure below.

Applied to a framed antenna, the matter becomes more critical and delicate, but with a lot of effort we can do it. The rheostat must be adjusted so that the oscillation is not excessive and the station whistle can be intercepted. At that point you adjust the variable and rheostat until the voice appears. If 220 ohm is not enough for the rheostat, try 470.

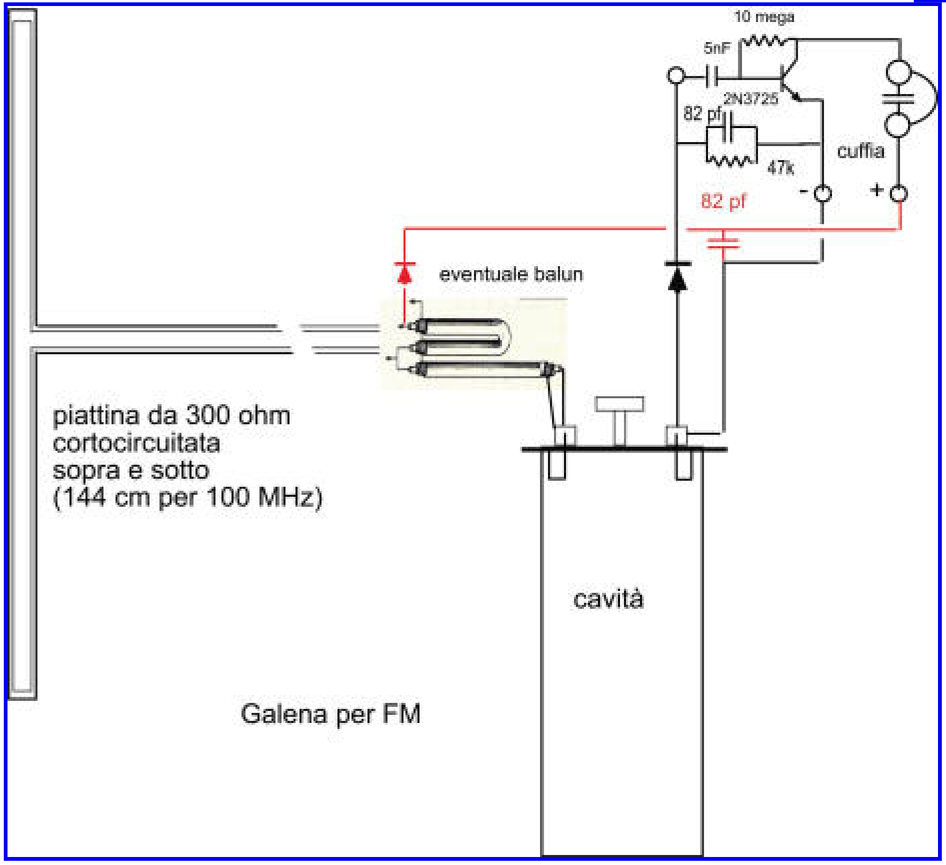

The great thing about this solution is that it does not need to build a frame with a special conductor, so the amplifier thinks about the losses.

In any case, if we want to lessen less, we have excellent results by applying to the chassis the 2 transistor amplifier several times described by me.

You will notice the insertion of 47 microhenry choke coils that serve to eliminate the FM signal that is present in the vertical section of the frame, dangerously saturating.

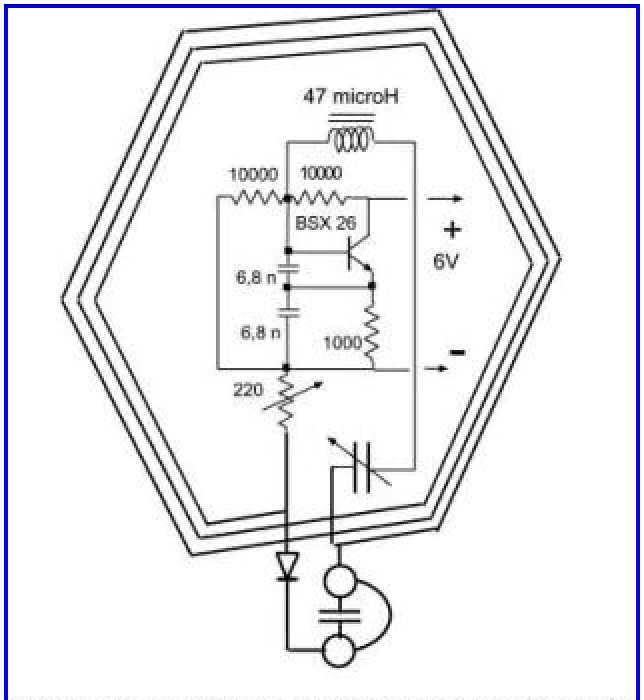
47 microH
10000
10000
BSX 26
+
6V
6,8 n
6,8 n
1000
-
220

Telaio amplificato col sistema ad impedenza negativa o Q multiplier
A me funziona con questi valori, altri tipi di telaio possono ciedere qualche aggiustamento.

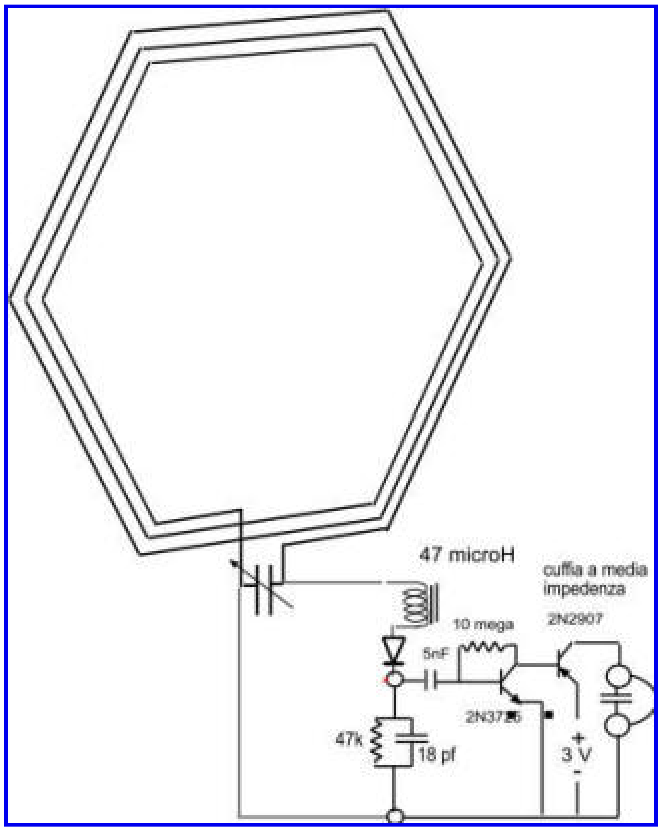
47 microH
cuffia a media impedenza
2N2907
10 mega
5nF
2N37[illegible]6
47k
18 pf
+
3 V
-

Rivitalizzazione di una vecchia galena.

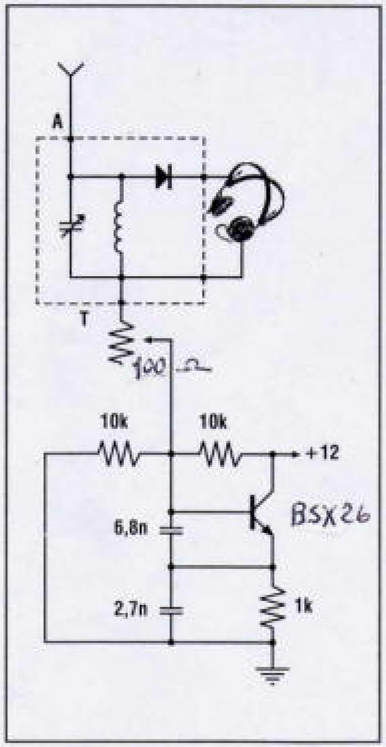
A
T
100 Ω
10k
10k
+12
BSX26
6,8n
2,7n
1k

NEXT

How to Take the Uncertainty Out of Your Trading

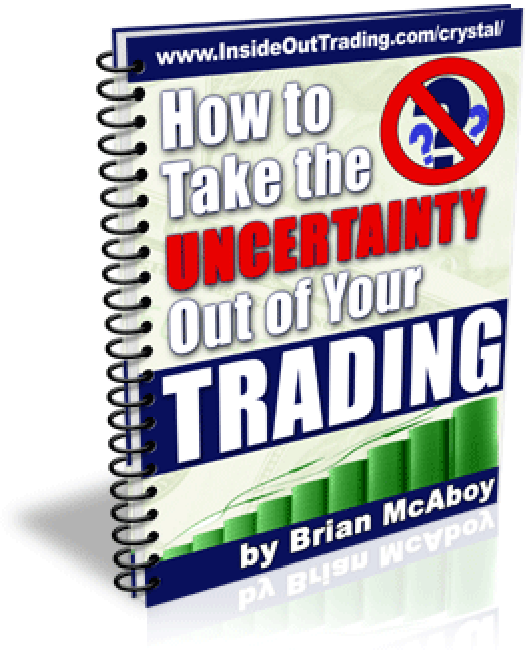
www.InsideOutTrading.com/crystal/
How to Take the UNCERTAINTY Out of Your TRADING
by Brian McAboy

By

Brian McAboy

Inside Out Trading

http://www.insideouttrading.com

Copyright © 2007 by New Ireland Ventures, LLC.

All rights reserved.

You *are NOT* allowed to distribute this report freely.

You may re-sell this report, but ONLY through the www.InsideOutTrading.com website by going to http://www.insideouttrading.com/certainty/signup.php and getting your link to do so there. You may NOT resell this report via any other means.

This publication is designed to provide accurate and authoritative information with regard to the subject matter covered. It is sold with the understanding that the author and the publisher are not engaged in rendering medical, psychological, legal accounting, or any other professional advice. If medical advice or other professional assistance is required, the services of a competent professional should be sought.

Also, the terms “How to Take the Uncertainty Out of Your Trading”, “The Subtle Trap of Trading”, “Trading Performance Analyzer” and “Trader’s Guide to Emotional Management” are pending registration and are fully protected and owned by New Ireland Ventures, LLC. They may not be used without expressed written permission of New Ireland Ventures, LLC or Brian McAboy.

Important Notice and Disclaimer

Trading of securities, including but not limited to, commodities, futures, stocks, options, currencies, CFD’s, etc., can be risky at all times, and involves exposure to substantial risk. It should NOT be assumed that the concepts and techniques suggested in this report will guarantee profitable results or that they will not result in losses. Your trading results are completely and solely your own responsibility.

Purchase of this report constitutes your agreement to the above disclaimer and exempts the author and anyone else involved in the production or distribution of this manual from any liability or litigation.

Introduction

The ultimate goal for all traders is consistent and reliable profits, with a trading operation that allows confidence, security and predictability. Now you may be thinking “Yes it is, but there is so much uncertainty in the markets. How am I ever going to be able to do that?”

While it is true that no one can predict the markets, you can make the uncertainty of the markets a non-issue, thus removing it from your trading. No one has control over the markets, yet there are numerous traders that enjoy trading as described above and they also enjoy consistent and reliable profits year after year as well, trading the same un-controllable and unpredictable markets that you do.

By the end of this report, you too will know exactly how to take the uncertainty out of your trading, and how to raise your confidence and security tremendously on your way to predictable, reliable and profitable trading.

Now, it is important for you to understand something here. This report is being kept extremely brief for two reasons: 1. I respect your time, so I am not filling up pages simply to give it enough “weight” to justify the meager price; 2. I am a firm believer that many solutions don’t need to be 500, 300 or even 100 pages to communicate the core ideas and information. Often this can be accomplished in only a handful of pages, so I try to keep my works as brief and concise as possible. This particular topic does not warrant an entire book, although there are certain matters brought up which do, and at those points I’ll recommend specific resources so that you can explore them further if you so choose.

This report is not intended by any means to tell you everything you need to know about trading. It is intended to raise your awareness to a specific and commonly overlooked problem, give you a simple and effective solution, and then also to give you a different perspective on your trading and how you approach the challenges before you as a trader.

The first part of this report details the problem and the solution. The second part then expands on the approach so that you can make use of it in other areas of your trading. It is my intention with this report to give you a new look at your trading business, and one that serves you well.

Part 1 – A Very Specific Problem

Let me begin by raising your attention to this: if you find this in your trading and choose to not to address it, you are leaving yourself open to a whole slew of emotional challenges in your trading, along with making some of the known best-practices in trading very difficult, if not impossible.

Since releasing the Trading Performance Analyzer, I've had conversation with many traders about backtesting and the importance of system analysis and tracking your key metrics. The reason that many of those conversations was taking place was because the person on the other end of the phone was at the point of struggle and frustration with their trading and they were very excited at prospect of being able to use the Analyzer to help them understand why their past success was eluding them now.

As we got to talking in more detail, they'd tell me about what they'd done in the past and where they were now with their trading. I'd ask about the system they were using and if they'd been able to identify what had changed or what they were trying to get back on track. Usually at this point they'd tell me that they had used this strategy or that and traded in these markets before but now had also moved into other markets, perhaps also making some changes to their system by incorporating different indicators or twists on the methodology.

Now these things are fine and quite commonly done, actually expected. As you develop and grow as a trader, you naturally want to put the things you discover or learn of into use, so your trading system will evolve. That is only natural.

When we got to the point where I asked if they had their trading system written down, the answer was usually half "Yes" and half "Well, sort of" or "Not really" or just plain "No".

The one thing that was common amongst all of them was their answer to the next question: “Is your system well-defined, and are the decision points clear to you?” After some thought, the answer was “No, not really.”

After some further questions, we’d come to the conclusion together that they were trying to trade with a rather discretionary system and / or a mixture of systems, but not a consistent system that was clear and well defined. What they employed on a given day might be subject to a variety of things, and certainly not the same from one week to the next.

Again, back to the goal of consistent profits.

Whether your system is very clear and written down or if it is totally random, you still have a system. The more random it is, the less likely you are to have consistent results, regardless of what the market does. On the same note, the clearer and more consistent your system is, the more consistent your results are likely to be, again, regardless of what the market does.

Now this presents a different challenge: how do you know if you have a clear and well defined system? How clear is clear enough? How well-defined is enough? I bring this up, because even those that had their systems written down, still did not feel that it was as clear as they’d like when we pursued it thoroughly. How do you know?

The Specific Solution to the Problem

At this point, let me share a bit of my background in my previous life outside of trading, because the solution I am going to give you is applicable in many areas, both specific to the business of trading and businesses in other industries.

You see, uncertainty is something that all businesses have to deal with, and there is a specific methodology to addressing it. I personally developed an addition to that methodology that works outside of trading, but is especially effective for the problem at hand, in trading.

I used to be a Quality Manager, which meant that one my primary functions was to ensure that our manufacturing operations ran consistently, efficiently and as cost-effectively as possible.

One tool integral to that effort was the Quality Manual. Now many people have no idea what a Quality Manual is. Simply put, it is the document system which defines the operation and everything that goes on in the operation. Its purpose is to make clear to everyone in detail how the company intends to achieve its goals.

A Quality Manual consists of layers of documents, each answering different questions:

1. Top level: Mission and Vision statements which detail the future of the company and what is to be accomplished
2. Next level: Policies which give an overview of the company rules that are set in stone, and these support the mission and vision statements.
3. Procedures are next and define the general activities of the business, and for each of those activities, answer the questions of:
    a. What is to be done (described generally)?
    b. Why are we doing it (its purpose)?
    c. Who (is responsible for making this happen)?
    d. When (or how often) is this activity to be performed?
    e. In what areas is this applicable (physical or departmental)?
4. Work Instructions support the procedures and answer these questions:
    a. What is to be done (specifically)?
    b. What are the steps to do this task?

Creating these documents can be quite a challenge, where you are defining what people are to be doing on a daily basis. Ideally these documents are clear and complete to the point that those following the procedures and work instructions can do their jobs consistently and according to the established best practices.

When we were first developing the document system, there were many procedures and work instructions written that were NOT clear and wound up needing many revisions. People would have questions in order to do the task at hand, and taking the time to get answers interfered with their efficiency. Either that or things would be done in a random fashion, depending on who was on shift that day or how they were feeling.

In the Quality Department, we had to make sure that all the activities and tasks that mattered (which was most), were documented and documented well. The level of documentation and the clarity and completeness of that documentation were in direct proportion to the consistency of the results, be they product quality, production, general housekeeping, accounting, customer service, etc.

As I trained others to write procedures and work instructions, I found a very effective method for creating these documents in a manner such that they were clear, complete and did not need revision for years.

Here’s the method for the creator of the work instruction:

1. Go through the entire process yourself, writing down every step you take
2. Find someone in the plant that knows their way around (assume that they do know the basics in that area), but hasn’t performed that specific task before and have them go through the work instruction.
3. Write down every question they have, and note anything they miss
4. When you can give the work instruction to someone new to the task and they can do it exactly as you expect and without question or assistance, then you are done. If not, repeat steps 2 and 3.

So how does this apply to the problem at hand, with one’s trading system?

> **When you have your trading system documented clearly and thoroughly to the point that you could hand it to another trader, and they could trade it the same as you without questions or assistance, then you have a sufficiently well-defined system.**

Making the Uncertainty of the Markets a Non-Issue

Now this also solved another problem encountered by the people I spoke of earlier. They were having a tough time trying to backtest their system, because of its lack of clarity. When an activity (like trading system execution) is unclear and variable, it is difficult to repeat. But that is exactly what you are doing when you backtest. You are taking your system back in time and running it through its paces repeatedly to validate it and understand the associated metrics.

Back to our goal of consistency: consistency comes from repeatability in what we do. When your system is clearly and completely defined, especially all the decision points, it becomes very repeatable and thus consistent and reliable.

Through backtesting and determining your system's metrics, you'll know how many winners and losers to expect, what average return-per-trade to expect, what your profit-to-loss ratio is. This is the point where the uncertainty of the markets becomes a none-issue with regards to your trading system.

If you have your trading system thoroughly defined and documented, and then backtest it properly and track your actual trading performance, you'll know what to expect from your system regardless of what the markets do, uncertain as they may be, because you'll have a repeatable and reliable system and you'll know the metrics associated with that system when traded with real market data.

For a very user-friendly and printer-friendly application for backtesting to get your system's metrics, plus a basic guide for backtesting, check out the Trading Performance Analyzer at http://www.insideouttrading.com/tpa/

Part 2 - Taking the Uncertainty Completely Out of Your Trading

Now back to the parallels between trading and all other businesses, and how you can capitalize on what we've discussed above to take the uncertainty completely out of your trading.

Like I mentioned before, all businesses have uncertainty to deal with. A business is not a building, equipment, computers, software, vehicles or machines. A business is people making things happen. It is an operation that involves people and those people are engaged in activities that make it a business.

As a Mechanical Engineer, I learned a long time ago that it doesn’t matter how great things work on paper. You can design the most magnificent machine or equipment layout ever, but when you take it into the real world, you have to include people into its function.

What was truly enjoyable about my time as a Quality Engineer and Manager, is that I got to deal with processes that included both the machines and the people running and maintaining them. The reason that Quality as a business approach and the Quality Manual have taken on increased significance over the years is business leaders have come to understand that profitability is very closely tied to the focus that the business has on quality, both in product and service delivered, but also within the operation itself.

In a process, variability represents risk. The reason that the Quality Manual is of such importance is that it reduces the variability in what the people in the company do. As more and more activities within the company are systemized and clearly defined and documented, the more consistent the company runs. This is the foundation of franchising – building duplicable businesses that can be run with generally available labor – and produce predictable profits, regardless of what town you put one in and who’s going to be working in your franchise business.

Back to trading.

Part of the problem here is the wording. Most traders think of a trading system the way it is defined by Investopedia.com:

> A trading system is simply a group of specific rules, or parameters, that determine entry and exit points for a given equity. These points, known as signals, are often marked on a chart in real time and prompt the immediate execution of a trade.”

By this definition and way of looking at one’s trading system, it is easy to understand why so many traders think that when they sign up for a signal service or buy a particular software package that gives them buy and sell signals, they think that they have a trading system. But they don’t. Not really.

You see, trading is a process, not just a set of rules and signals. That process is a series of actions, it is an activity, and is best defined by procedures.

I bring up the point that your trading is a whole process, not just a trading system to illustrate the importance of reducing variability, thus eliminating uncertainty. Your “trading system” must include everything that impacts your trading results. There can be many factors that introduce variability, and I’m talking about things in what you do, not the markets.

A lesson that we can learn from Quality is to focus on the process before focusing on hitting the target. In manufacturing, process control is imperative to producing in-spec product. If a machine process is both highly variable and off the target specification, there are two approaches that can be taken. You can try to center the process, then reduce the variability or you can tighten the process and then center it. In the automotive industry, the Japanese will generally focus on the process first, then the target, whereas American car-makers often focus on hitting the target first, then tightening the process. Needless to say, the process-focused approach usually reaches the end-goal with considerably much less time and money involved.

Same goes for trading. Understand that your trading system is more than a set of signals and rules. It is a process, an operation. Many people start trading and think that as soon as they have a “trading system”, they have a trading operation. They don’t. That is only one part of the business.

Your trading business is like any other business in certain respects: you have an owner (you), and a worker (you, the trader), probably an accountant to do your books (maybe you, maybe you hire one out), a computer technician for installing software (again probably you), head of Research & Development, etc. etc.

The point is, as the owner of your trading business, you have many hats to wear, many roles to fulfill. As with the trading system, you have a business whether you treat it as one or not. As you define your roles and document the activities for each of those roles more clearly, the more repeatable and predictable your business will become. In the Subtle Trap of Trading, I detail all of the aspects of your trading business, so you might find it helpful to take a peek at it. There are numerous aspects to your business and it is valuable for you to define those roles, document the associated activities and periodically switch hats and fulfill those roles – but with intention, not haphazardly.

Conclusion

One very common problem that traders struggle with is a lack of clarity in their trading system. Perhaps trying to get too technical, or just wanting to find that something that works causes one to introduce variables and inconsistencies into the system, thus adding to the uncertainty of it to be repeatable and reliable.

Without clarity in one's system, you leave yourself open to a host of problems. Most of all deviations and decisions that result in regrettable losses. You'll find it difficult to trade and backtesting will be dicey at best and of little value.

Make sure that you clearly and completely define your system to the point that you could hand it to another trader and have them trade it the same as you. Then you've got a clearly defined system that you can measure, track and improve.

You trading is a business, and the more clearly you define the roles and activities in your business, the more certainty and repeatability you bring to your operation as a whole.

The markets will have their uncertainty and there is nothing you can do to try to control the markets or predict them. You do however have total control over yourself and what YOU do. When you take control of yourself and your operation, you will effectively make the uncertainty of the markets a non-issue and you'll take the uncertainty completely out of your trading.

About the Author

Brian McAboy is a degreed Mechanical Engineer and Certified Quality Engineer, along with being a trading enthusiast.

I hope that you've found this report useful and enjoyable. As you can tell, I take a business approach to trading and bring learning from my roles in other areas of my life here for your benefit.

There are many things to learn and digest as you develop into the trader you wish to be, and there are stages to the development. To assist you and help you accelerate your progress, I've created several resources for you and you can find out more about them at http://www.insideouttrading.com

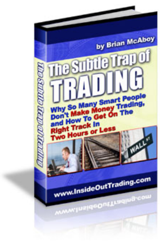
by Brian McAboy
The Subtle Trap of
TRADING
Why So Many Smart People Don't Make Money Trading, and How To Get On The Right Track in Two Hours or Less
WALL
www.InsideOutTrading.com

The Subtle Trap of Trading – the pieces to the puzzle that most new traders miss and only discover they need after they've "hit the wall" – written for new traders, but most often purchased and appreciated by second-wave traders.

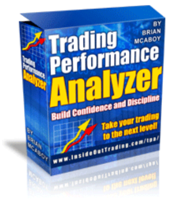
BY BRIAN MCABOY
Trading
Performance
Analyzer
Build Confidence and Discipline
Take your trading to the next level!
www.InsideOutTrading.com/tpa/

The Trading Performance Analyzer – once you have a system (and have it well-defined as discussed above), get your metrics and track them. Build confidence and discipline in your system through backtesting.

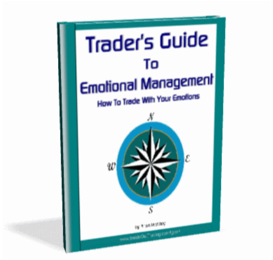
Trader's Guide
To
Emotional Management
How To Trade With Your Emotions
N
E
S
W

The Trader's Guide to Emotional Management – after you have the basics covered, and you have a system well defined, manage your emotions so that you trade your system well and get the most out of your trading.

CRYSTAL RADIO KIT

gti kit™ 8893

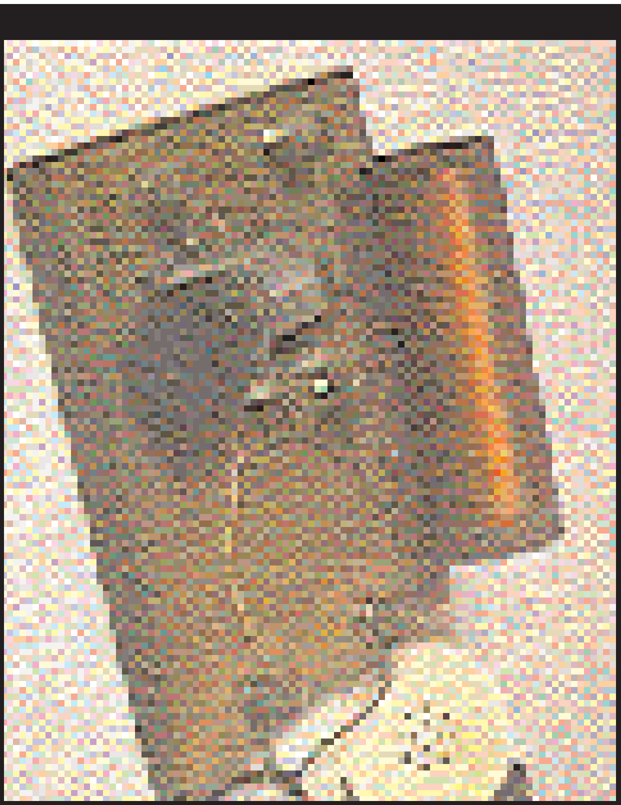

GIBSON TECH ED, INC. 1216 S 1580 W OREM, UT 84058 1-800-422-1100

(A) List of Parts provided in the CRYSTAL RADIO SET

1. 8 Brass furniture thumbtacks
2. 70’ of Copper magnet wire #30
3. 1 Large paper clip
4. 1 - 470 pF disc capacitor
5. 1 - 0.01 µF disc capacitor
6. 1 - Germanium diode
7. 1 - Crystal earphone with wires
8. 2 - Alligator clips with wires
9. 1 - PLMCRS1 instructions

Parts you need to obtain:
1 cardboard tube (toilet papertube)
1 board (wood, 71/2” x 5”)
sticky tape (to hold coil you make)
quick drying varnish, nail polish, or clear radio cement

Tools Needed:
needlenose pliers
a small hammer
wire cutters

(B) Project Description

*The complete **CRYSTAL RADIO SET KIT #8893** includes everything you need except for the wood baseboard, the cardboard tube from a toilet paper roll and varnish, nail polish, or clear radio cement to paint the coil.*

*This little project is a sure interest grabber with students of electronics of any age! It does not require a great amount of skill, but a little application will produce a good basic radio receiver. Construction is fairly simple. **This radio DOES NOT require batteries** to operate, using simple electrical principles to power itself.*

This circuit is for the simplest radio receiver that can be practically constructed - yet it is the basic building block of more highly sophisticated and complex units.

(C) SAFETY FIRST

***DO NOT** attempt to use an electrical wall outlet, or connect to any electrical appliance to ground your radio. Use a cold water tap, or “ground stake”.*

***DO NOT** mount your antenna near to, or over the top of, electrical power lines of any type. If your antenna wire should break it could contact the live power and a severe **ELECTRICAL SHOCK** and possibly **DEATH** could occur.*

***DO NOT** leave your antenna coupled to the radio when it is not used, as it is an attraction to **LIGHTING STRIKES** and/or violent electrical **STATIC DISCHARGE**. Place the free end of the antenna **OUTSIDE** and clear of the building or connect it direct to earth.*

***DO NOT** operate the radio during the approach of, or during an electrical storm .*

GIBSON TECH ED, INC. 1216 S 1580 W OREM, UT 84058 1-800-422-1100

(D) Operating instructions

*In order for the radio set to work you will have to connect it to an **ANTENNA** and a **GROUND.** In some locations an antenna may not be necessary, but a good ground connection is very important. A cold water pipe is usually the best to use, or you can put down an "earth stake" and connect to it.*

The best antenna for this radio can simply be a really long wire, mounted as high as possible. Good results can also be produced by using metal bed spring frames or a fence. Be creative!

Crystal radios are not very selective, nor very sensitive. Several stations may be heard at once, many of them equally as strong as the others. A long antenna (30+ meters) will improve the strength of some stations, helping you tune them in a little more easily.

Operation is that simple.

(E) Theory of Operation

The crystal radio works by creating an oscillator with the coil and C1. The placement of the wiper along the coil varies the amount of inductance in parallel with the capacitor. This tunes our radio by determining the oscillating frequency. The signal is then rectified by the diode and smoothed out by C2. The resulting signal causes a thin ceramic crystal in the earphone to vibrate creating the sound that we hear.

(F) Reading Comprehension

1. The wiper varies the amount of_______________in parallel with the capacitor.

2. The signal is___________by the diode.

3. The crystal _________________ vibrates to create sound.

(G) Complete Parts List

- A (9cm x 13cm) or (7.5in. x 5in.) block of Pine wood, or chipboard.(not included in kit)
- 8 BRASS flathead thumb tacks.
- 2 Alligator Clips, one for antenna, one for ground.
- A discarded cardboard roll from a toilet paper roll.(not included in kit)
- L1 - 70' of 30 Gauge Magnet wire for approximately 145 turns around the cardboard roll.
- A large wire paper clip.
- 300mm length bare copper.
- C1 - A 470 Picofarad ceramic disc capacitor.
- C2 - A 0.01 Microfarad ceramic disc Capacitor.
- D1 - A Germanium Point Contact Diode Type OA90.
- E1 - A CRYSTAL earphone

(H) Schematic Diagram

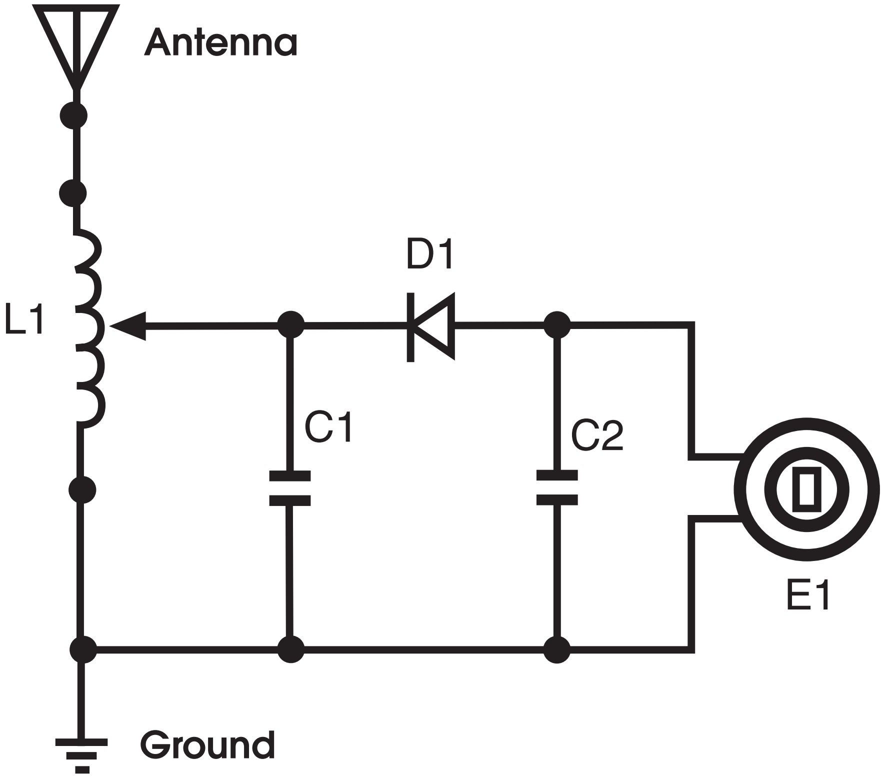

Copyright © 1996 PracTrak, Inc.

(I) Assembly Instructions

Tools needed:

A small hammer, a pair of wire cutters and a pair of needlenose pliers are all the tools required. A roll of household sticky tape is handy for holding wire in place as you wind the coil.

Step 1 (Wood Project Board)
Cut a piece of wood to 5" x 7 1/2" x 1/2"

Step 2 (Layout Pattern)
Cut out the Layout Pattern and glue it to the wood.

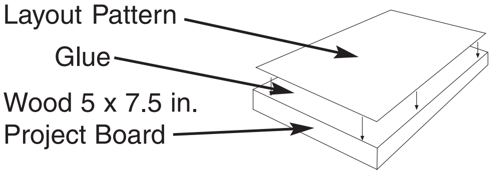

Step 3 (Coil)
Prepare the tube to wind the coil. Measure and cut a 7cm length of the tube. Draw a line all the way around the tube to use as a guide while cutting. Draw two more lines 1cm in from eather end. Cut tabs on both ends exactly across from each other then fold them in, out of the way for cutting off the extra tube length. (See Fig. 1)

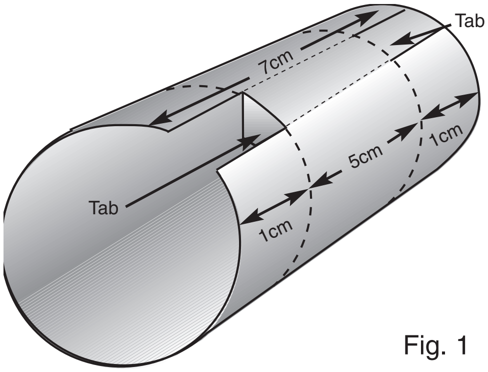

Fig. 1

Step 4 (Coil)
Using scissors, cut the extra length from the tube. Do not cut the tabs off as they will be used to hold the tube to your wood project board.
The easiest way to cut the tube straight is by placing the flat of the scissors against the folded tab and following your line around as shown in Fig. 2.

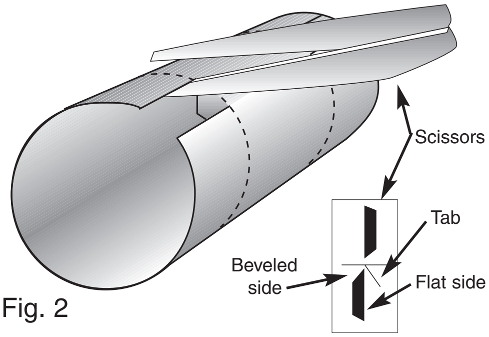

Fig. 2

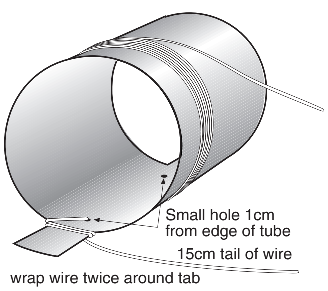

Fig. 3

Step 5 (Coil) (See Fig. 3)
*Make a small hole with a tack, 1cm in from both ends of the tube. Push 15cm of the wire through from the outside of the tube to the inside and wrap the wire tail twice around the tab at one end, this prevents the wire from slipping as you wind. Starting 1cm from the edge of the tube firmly wind on as many turns as will fit between the two holes (approximately 145 turns,) ensuring each turn is straight and tight together **no gaps!***

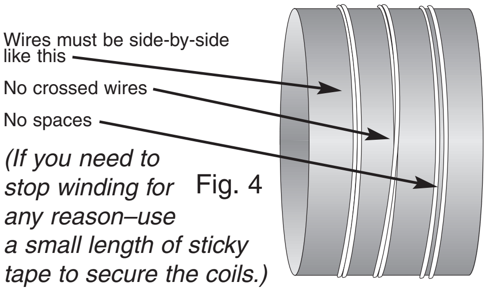

Fig. 4

(If you need to stop winding for any reason–use a small length of sticky tape to secure the coils.)
When you reach tab #2, trim off the wire to 15cm and wrap it twice around the tab. A coat of quick drying varnish or nail polish will secure the coil loops permanently. Coat the cardboard tube at the same time and place the assembly aside to dry.

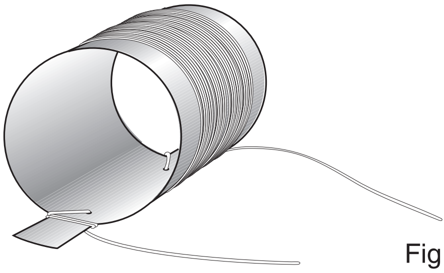

Fig. 6

Step 6

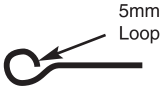

Straighten out a paper clip with pliers and a hammer. In one end make an open loop 5mm in diameter. This paper clip will be used as a wiper on the coil. (See the Pictorial Diagram page 4). All the components and wires are secured with furniture tacks no soldering is required.

For a better connection to the tacks make small loops at the ends of the wire leads on the capacitors and diode and the copper wire which joins the various parts of the circuit together.

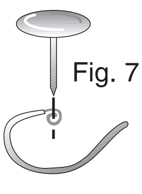

Fig. 7

Step 7 (CoiL)
*Attach the Coil you just made to the board as indicated by the pattern, placing the tacks through the tabs of the tube. (You may need a small hammer to tap the tacks down tightly.) **Do Not** hammer any other tacks down into the board until **all** the other parts have been properly positioned.*

Step 8 (Wiper)
Place the wiper in position on the board and rest it against the coil. From this you can see where the coil will need to be scraped free of insulation. (This area on the coil is shown as a white stripe on the Pictorial Diagram page 4.) The enamel insulation and varnish you coated the coil with can be removed using a fine cut file or fine grit emery paper–taking care not to displace or break any of the coil turns. Make bare an area about 10-15 mm wide, across the face of the coil. The coating also needs to be removed from the leads, then they can be wrapped around the appropriate tacks "A" and "F".

Step 9
*With the straightened paper clip acting as a **WIPER,** the coil will be a **VARIABLE INDUCTOR** (See the Pictorial Diagram page 4). The free end of the wiper is bent at an angle so that it will wipe along the bare wires of the inductor when secured in place on the baseboard.*

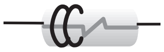

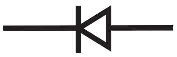

Step 10
Identify the Diode (D1) with the black stripes at the end corresponding to the line shown on the schematic symbol. Place it as shown in the schematic on the board in the correct position between tacks "B" and "C".

Step 11
Identify the capacitors (C1, C2) and place them in position. C1 goes between tacks "B" and "D", and C2 goes between "C" and "E".

Step 12
Take one of the alligator clip leads and wrap the end of the wire around tack "A". Wrap the end of the other alligator clip lead around tack "F". Make sure that you hammer these tacks down securely to prevent the alligator clip wires from pulling away from the tack causing loss of contact.

Step 13
Take bare copper wire and connect tacks "D", "E", and "F".

Step 14
When all of the components are arranged as desired, the tacks can be hammered into place. Now your radio is complete. Happy Listening!

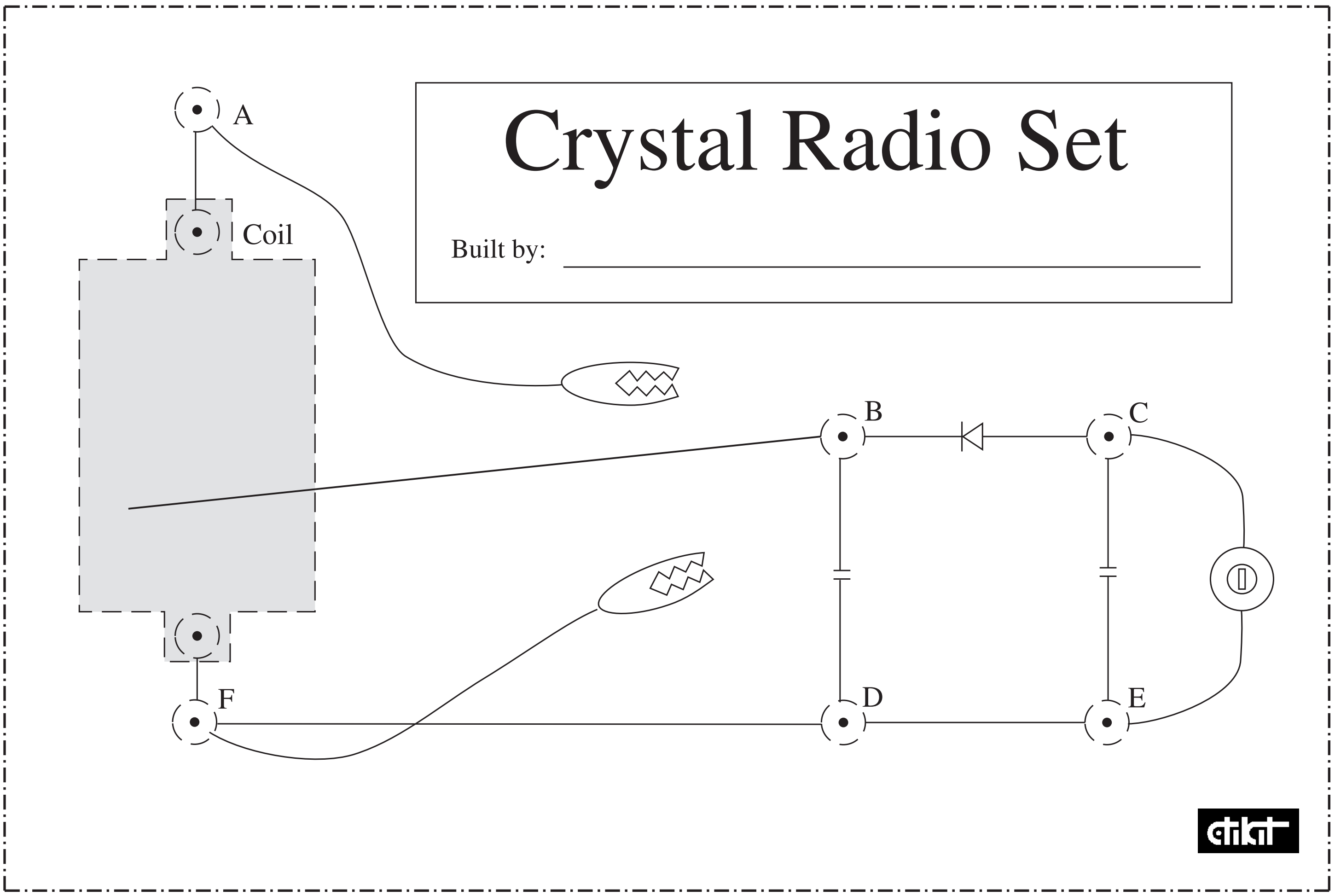
Crystal Radio Set
Built by:
A
Coil
B
C
D
E
F

Cut on dotted line.

(J) Layout Pattern (above, exact size)

Cut out the layout pattern and place it on your board. Build your radio on top of it.

⊙ *= Thumb Tacks*

(K) Pictorial Diagram

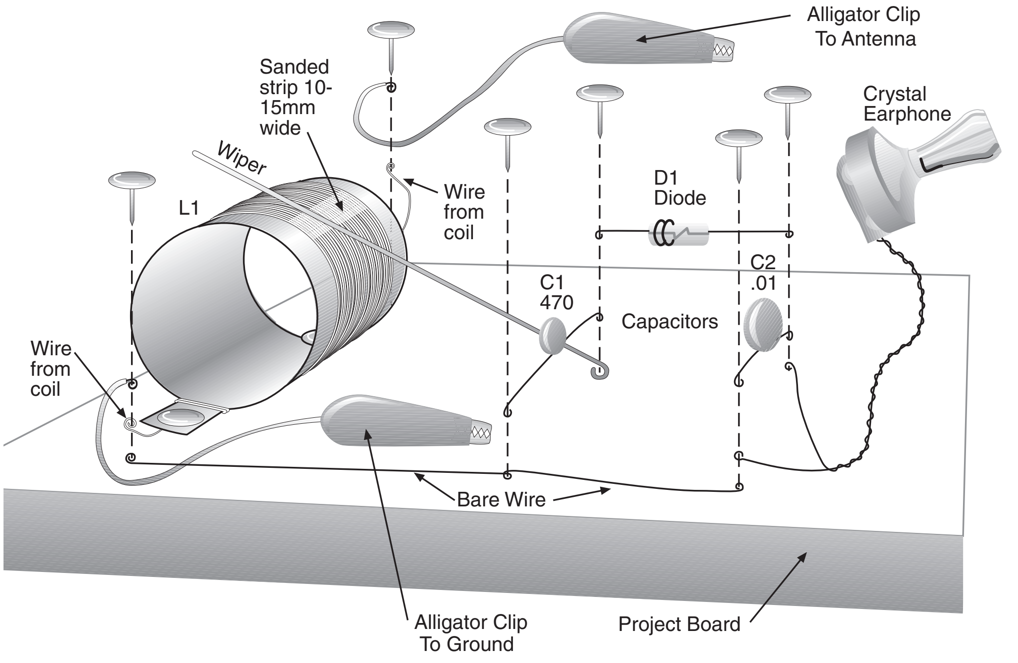
Alligator Clip
To Antenna
Sanded strip 10-15mm wide
Crystal Earphone
Wiper
L1
Wire from coil
D1
Diode
C2
.01
C1
470
Capacitors
Wire from coil
Bare Wire
Alligator Clip
To Ground
Project Board

(L) Web Address

This and other Kits are available on the internet at :
http://www.shopsite.com/kits

Copyright © 1996 PracTrak, Inc.

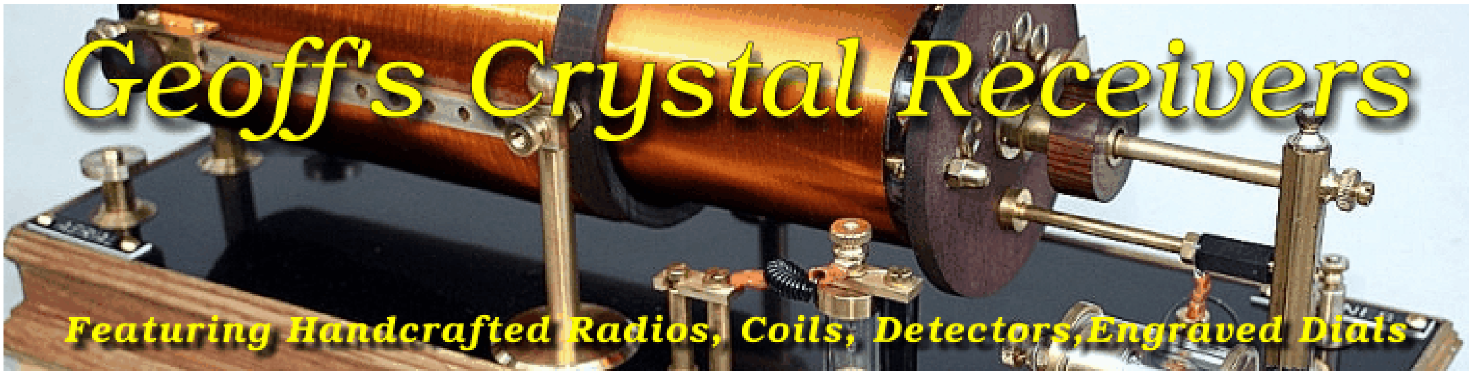
Geoff's Crystal Receivers
Featuring Handcrafted Radios, Coils, Detectors,Engraved Dials

- Home
- Links
- Crystal Sets
- More Crystal Sets
- Loosecoupler

Geoff's Projects Page

Trash Can Radio
with Fools Gold Detector

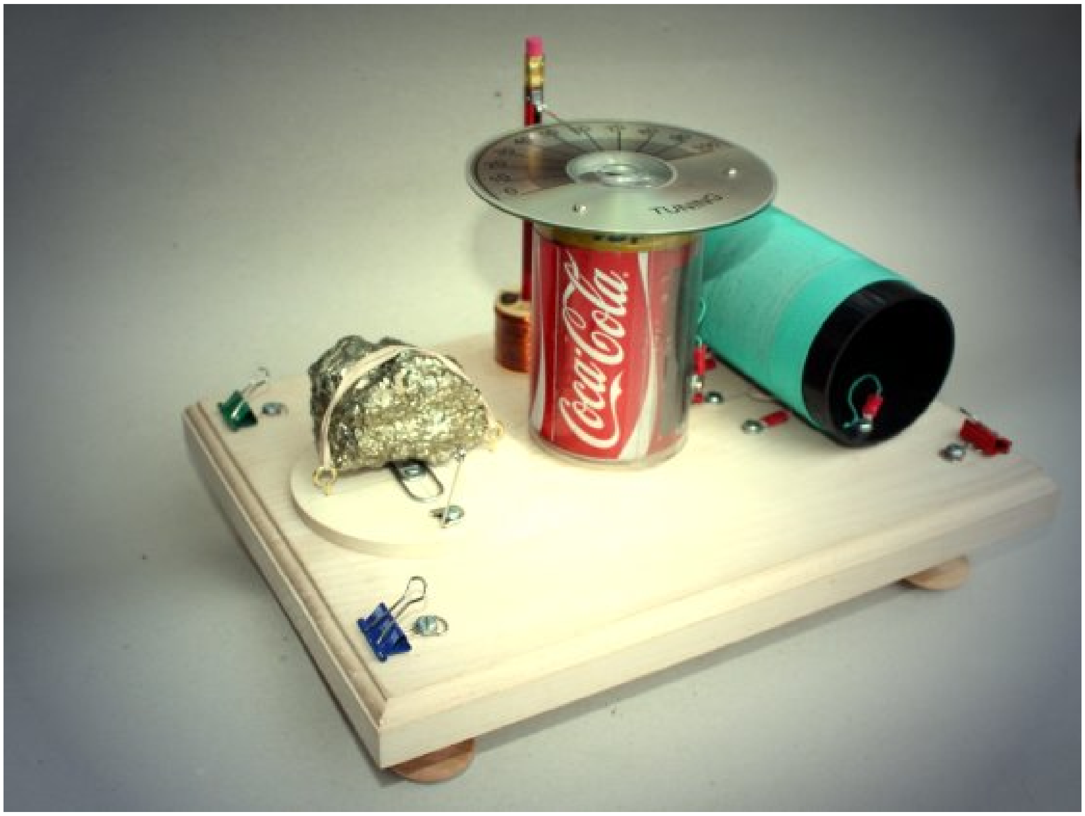

I wanted to make a radio that was very simple to construct so that anybody could make it from readily available materials and parts. I thought it may be useful if there was for example an emergency survival situation or the networks had broken down or you were marooned but really just a nice piece of fun to construct in the holidays. It really works and is quite listenable

This radio was made from what was basically around in any domestic house ie Paper clips pencil Cokacola can cotton reel curtain hooks an elastic band a safety pin and a breadboard. The only thing that was bought by me was some wire but this could have easily been obtained from an old broken down domestic appliance transformer ie from a microwave oven.

The most critical item to make was the main tuning coil this was wound on a 6 inch long by 3 inch diameter piece of drain pipe. I stuck to my old favorite circuit and type of coil. This is the' Mystery Crystal radio' circuit and coil (see link to constructional details below) . It is a very simple and very sensitive circuit and actually works better with a piece of Iron Pyrite (Fools Gold) than a crystal diode like the 1N34 . On the diode there was overlapping of stations but not so on the fools gold.

The main tuning capacitor was made from a cokacola can. I cut away half of the can area. I then stuck another piece of coke can onto a piece of clear plastic Christmas cake wrapper with double sided tape. This was then secured to a CD disc also with double sided tape. The whole thing was made to rotate on the can thus producing a crude but quite effective tuning capacitor The capacitance swing was from 100pf to 400pf. This enabled me to tune most of the medium wave band with ease.

The detector was made from a very large piece of iron pyrite and an elastic band two curtain hooks and a nickel plated safety pin all mounted on a disc of wood . I had not tried such a large piece of pyrite before in a detector but there seemed to be no deterioration of sensitivity due to size or positioning of the safety pin. I found the best part to just touch the pin on was not the flat faces of the crystal but the small crystaline structures or clusters. You do not need a firm pressure but just a light touch contact even just a resting position is the best. It is by no means as fiddly as the Galena type detectors and seems less prone to vibration.

All the set wiring was tucked away under the breadboard , there were only about half a dozen wires anyway. This was a nice little project and I really enjoyed making and listening to it over the Christmas holidays.

Circuit Details

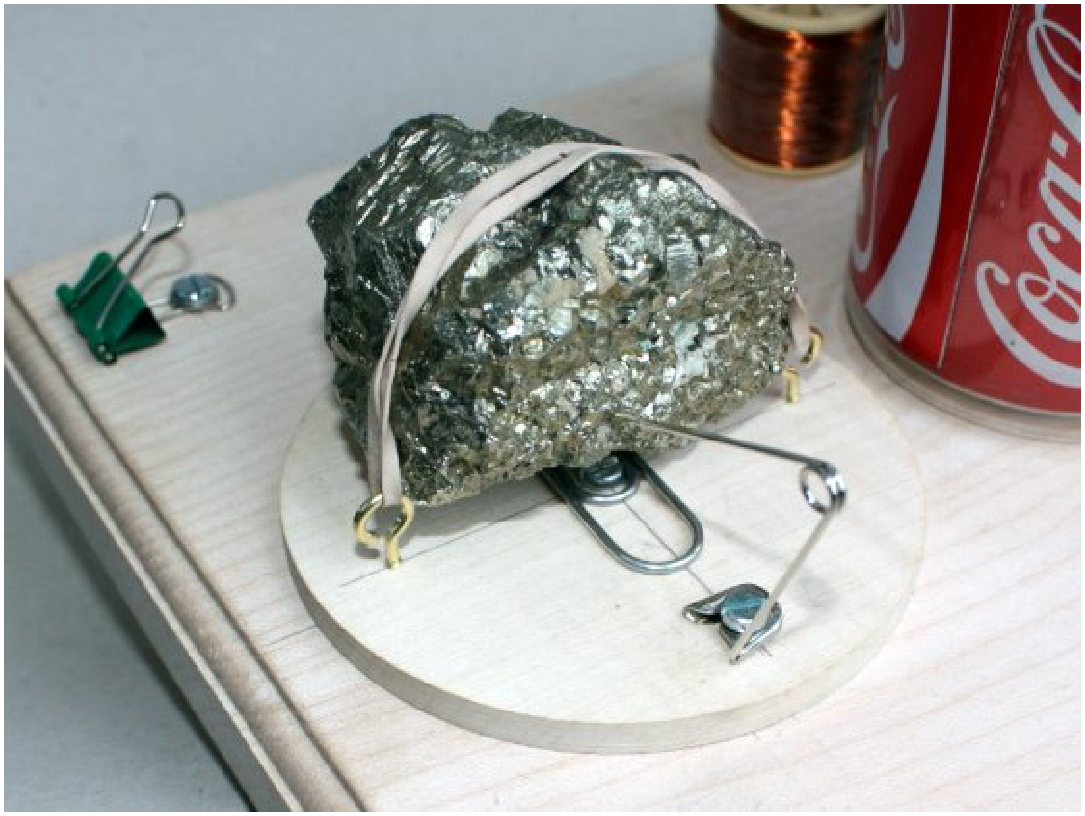

0
10
20
30
40
50
60
70
80
90
100
TUNING

Aerial Parts | Brass Hardware | Capacitors | Coils | Detectors | Diodes
Engraved | Links | Loose Coupler | Misc. Items | Projects |Radios |Radios-2
Radios-3 | Switches | Workshop

© 2009-2015, Geoff Roberts. All Rights Reserved

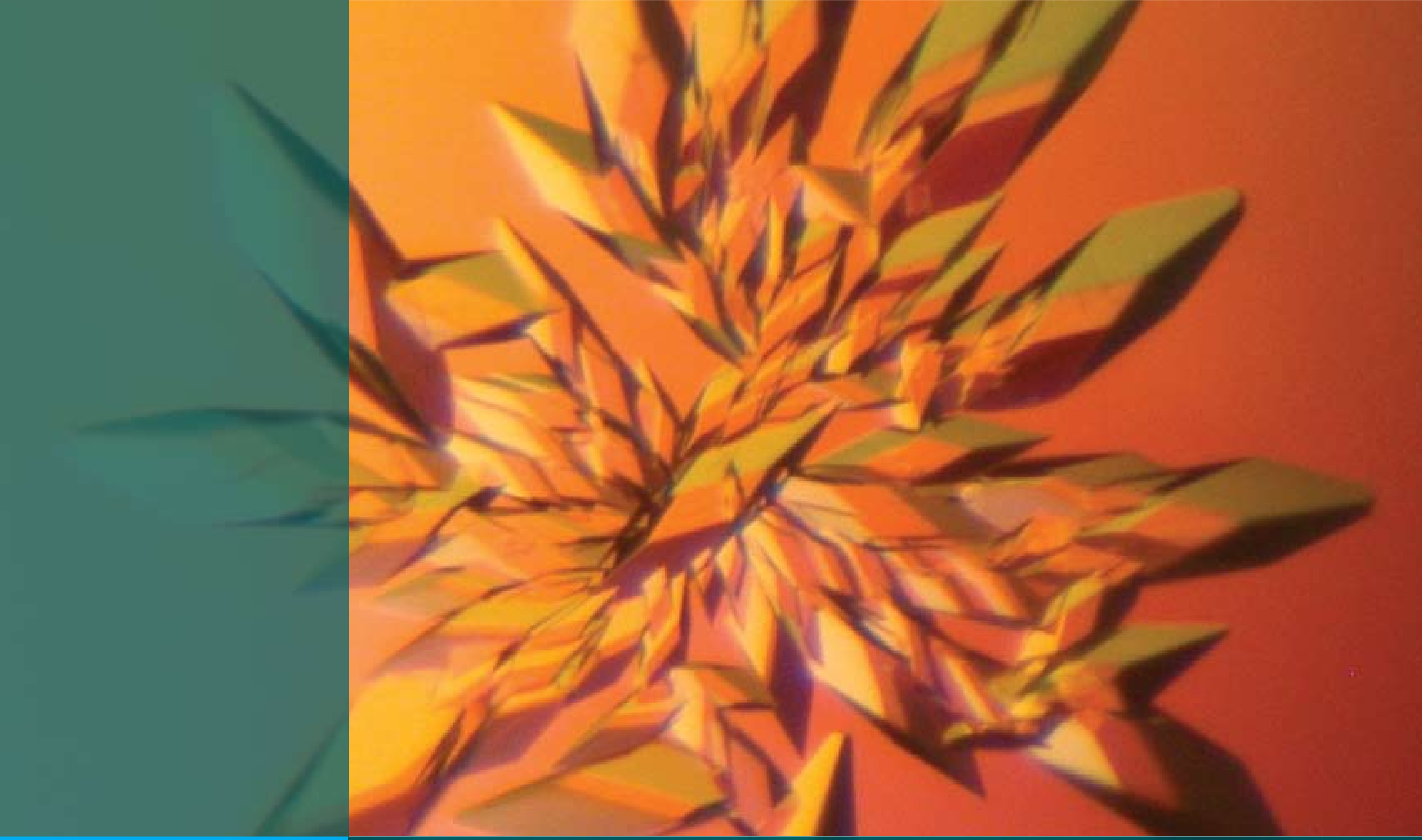

HAMPTON
RESEARCH

CRYSTALLIZATION

VOLUME 18

RESEARCH TOOLS

HAMPTON
RESEARCH
Solutions for Crystal Growth

CONTACT INFORMATION

Business Hours: 7:00 a.m. to 5:00 p.m.
Monday - Friday
(Pacific Standard Time)

Phone: (800) 452-3899 or
(949) 425-1321

Fax: (949) 425-1611

Address: Hampton Research
34 Journey
Aliso Viejo, CA 92656 U.S.A.

CUSTOMER SUPPORT SERVICES

Customer Service:
Phone: (800) 452-3899 or
(949) 425-1321
Fax: (949) 425-1611
Email: info@hrmail.com

Technical Support:
Phone: (949) 425-1321
Fax: (949) 425-1611
Email: tech@hrmail.com

Website: www.hamptonresearch.com

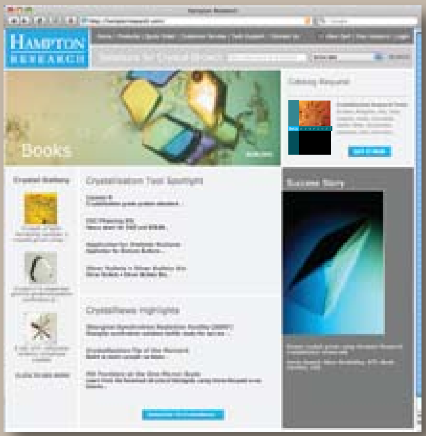

On the cover: Crystals of Big-R protein of Xyllela fastidiosa.

Produced in Brazilian Synchrotron Light Laboratory (LNLS),
Center for Structural Molecular Biology by Rosicler Lázaro Barbosa.

TABLE OF CONTENTS

PAGES

2 - 33 SCREENS

34 - 37 CUSTOM SHOP CRYSTALLIZATION REAGENTS

38 - 59 OPTIMIZE™ CRYSTALLIZATION GRADE REAGENTS

60 - 69 STOCKOPTIONS™ KITS

70 - 105 CRYSTALLIZATION PLATES, HARDWARE & ACCESSORIES

106 -115 TOOLS, SEEDING & RESIN

116 - 135 CRYOCRYSTALLOGRAPHY

136 - 145 CAPILLARY MOUNTS & SUPPLIES

146 - 153 GONIOMETER HEADS & SUPPLIES

154 - 157 XENON DERIVATIZATION

158 - 163 LABELS AND PENS

164 - 173 BOOKS

174 - 179 PROTEIN CRYSTALLIZATION STANDARDS

180 - 267 CRYSTAL GROWTH 101

268 - 275 INDEX

276 - 277 GENERAL INFORMATION

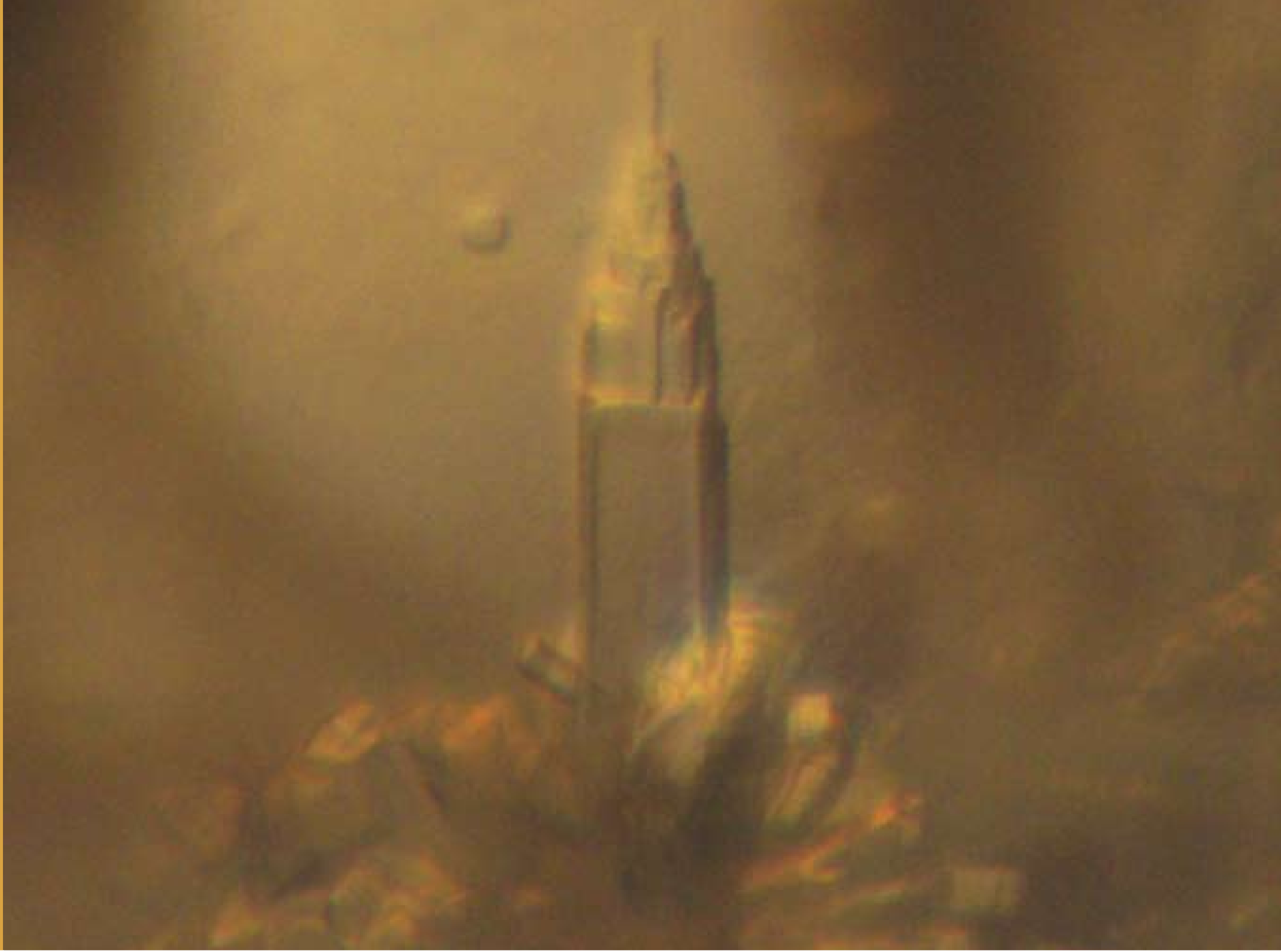

SCREENS

Empire State Building-shaped crystal of neuron guidance receptor protein grown in Hampton Research Crystal Screen reagent 15.

Momchil Kolev, Dorothea Robev and Dimitar Nikolov, Structural Biology Department at Memorial Sloan-Kettering Cancer Center, New York, USA.

TABLE OF CONTENTS

PAGES

5 PCT™ PRE-CRYSTALLIZATION TEST

6 - 7 INDEX™

8 - 9 CRYSTAL SCREEN™

10 - 11 PEGRX™

12 - 13 PEG/ION™

14 - 15 GRID SCREENS™ • QUIK SCREEN™

16 - 17 SALTRX™

18 - 19 MEMBFAC™ • CRYSTAL SCREEN LITE™

20 - 21 NATRIX™

22 - 23 CRYSTAL SCREEN CRYO™

24 NUCLEIC ACID MINI SCREEN™

25 LOW IONIC STRENGTH SCREEN™

26 - 27 SILVER BULLETS™

28 ADDITIVE SCREEN™

29 DETERGENT SCREEN™

30 - 32 HEAVY ATOM KITS™

HAMPTON RESEARCH

Fish shaped protein crystal grown from Hampton Research crystallization reagents.

Paul Morin and Kevin Kish, Bristol Myers-Squibb, Princeton, New Jersey, USA.

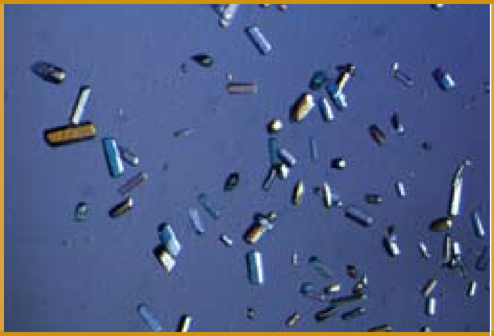

Crystals of human mineralocorticoid receptor complexed with ligand. Grown using Hampton Salt Rx screen.

Kevin P Madauss, GlaxoSmithkline, Research Triangle Park, North Carolina, USA.

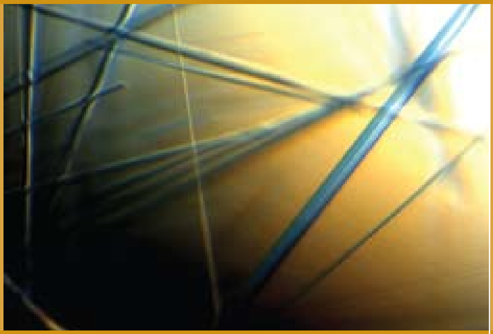

Crystals of kinase complex optimized from a Hampton Research Tacsimate based reagent.

Michelle Quiles, GlaxoSmithKline, RTP, North Carolina, USA.

PCT™ PRE-CRYSTALLIZATION TEST

APPLICATION

- Determine the appropriate sample concentration for crystallization screening

FEATURES

- Conserve sample
- Optimize sample concentration prior to initial screens
- Provide insight to sample homogeneity

DESCRIPTION

The PCT Pre-Crystallization Test is used to determine the appropriate sample concentration for crystallization screening. Sample concentration is a significant crystallization variable. Samples too concentrated can result in amorphous precipitate, while samples too dilute can result in clear drops. Precipitate and clear drops are typical crystallization screen results for reagent conditions which do not promote crystallization and are part of every crystallization screen. However, by optimizing protein concentration for the screen, the number of clear and precipitate results can often be reduced, which in turn results in more efficient sample utilization while at the same time enhancing the chances for crystallization. PCT can minimize or prevent situations where a screen results in an over abundance of precipitate or clear drops.

The PCT kit contains 4 unique, preformulated, sterile filtered reagents used to evaluate protein concentration for crystallization screening. Initially, the sample protein is mixed with two of the reagents to determine if the protein concentration is appropriate for crystallization screening. If the protein is very sensitive to salt and polymer concentration, based on initial PCT results, the protein may be evaluated using a second set of PCT reagents. PCT results will then provide insight to either the appropriate sample concentration or indicate that other diagnostic testing such as native gel electrophoresis or dynamic light scattering should be performed to demonstrate sample homogeneity appropriate for crystallization.

Each PCT kit contains 4 unique reagents, 50 ml each. This kit is useful for researchers who already have crystallization plates and cover slides.

Each PCT kit (with plates) contains 4 unique reagents, 30 ml each, plus 24 well VDX™ Plates with sealant and plain cover slides. This kit is useful for labs without access to crystallization plates and cover slides.

References

1. *Crystallization and x-ray diffraction analysis of human CLEC-2. Aleksandra A. Watson and Christopher A. O'Callaghan. Acta Cryst. (2005). F61, 1094–1096*

Order Information

Each PCT kit contains 4 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 36 for further details.

Cat. No.	Name	Description	Price
HR2-140	**PCT**	50 ml bottles (4 ea), plain cover slides (1 pk)	**$80.00**
HR2-142	**PCT (with Plates)**	30 ml bottles (4 ea), plain cover slides (1 pk), VDX Plates with sealant (5 ea)	**$98.00**
HR2-940-**	**PCT Custom Shop**	185 ml	**$138.00**

** = reagent number A1-B2

PCT Reagent A1/B1 Results	PCT Reagent A2/B2 Results	Action
Heavy amorphous precipitate	Heavy amorphous precipitate	Dilute sample 1:1, repeat steps 1-7
Clear	Clear	Concentrate sample to half the original volume, repeat steps 1-7
Light granular precipitate	Clear	Perform screen
Clear	Light granular precipitate	Perform screen
Light granular precipitate	Light granular precipitate	Perform screen
Heavy amorphous precipitate	Clear	Perform PCT with B1 & B2 / perform diagnostic testing
Clear	Heavy amorphous precipitate	Perform PCT with B1 & B2 / perform diagnostic testing

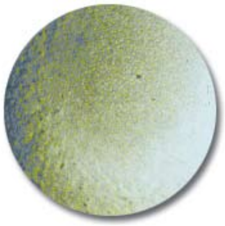

Figure 1. Light Precipitate

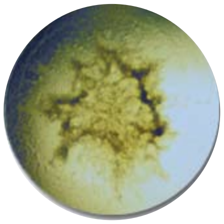

Figure 2. Heavy Amorphous Precipitate

INDEX™ • INDEX HT™

APPLICATION

- Primary, diverse reagent system crystallization screen for proteins, complexes, peptides, nucleic acids, & water soluble small molecules

FEATURES

- Developed at Hampton Research
- A data-driven biased sparse matrix and grid screen
- Screens classic, contemporary, & modern crystallization reagents
- Samples pH 3 to 9
- Compatible with microbatch, vapor diffusion, & liquid diffusion methods
- Specially formulated reagent zones:
  - Traditional salts versus pH
  - Neutralized organic acids
  - High [salt] with low [polymer]
  - High [polymer] with low [salt]
  - Low ionic strength versus pH
  - PEG & Salt versus pH
  - PEG & Salt
- Tube or Deep Well block format

SUCCESS STORY

Crystals of a diabetes related protein.

Courtesy of Allan D'Arcy and Aengus Mac Sweeney.

Morphochem AG

DESCRIPTION

Index is designed as a 96 reagent crystallization screen that combines the strategies of the grid, sparse matrix, and incomplete factorial screening with traditional, contemporary, and new crystallization reagent systems into a highly effective and efficient format.

Index, as the name implies, efficiently samples a series of specially formulated reagent zones to identify which reagent class or classes and pH are effective in producing crystals or limiting sample solubility. Results from Index can be used to design optimization experiments and to identify follow on screens by reagent class. For example, positive results with salt based reagent in Index may be followed up with further screening using SaltRx or Grid Screen Salt. Success with polymer based reagents in Index may be followed up with further screening using PEGRx or PEG/Ion.

Index utilizes a broad, yet refined portfolio of crystallization reagent systems. These include the following: (1) traditional salts such as Ammonium sulfate and Sodium chloride versus pH; (2) neutralized organic acids such as Sodium malonate and Tacsimate; (3) High salt concentration mixed with low polymer concentration as well as high polymer concentration mixed with low salt concentration and; (4) Low ionic strength using polymers such as PEG, MPD, Pentaerythritols versus pH. These reagent systems are formulated across a sparse matrix and incomplete factorial of concentration ranges, sampling a pH range of 3 to 9.

Index contains 96 unique reagents, 10 ml each.

Index HT contains 96 unique reagents in a single Deep Well block format.

Ready-to-use reagents are sterile filtered and formulated with ultra-pure Type 1 water, using the highest purity salts, polymers, organics and buffers.

Measured pH range of kit is 3 to 9 at 25°C
Average measured pH of kit is 6.8 at 25°C
Median measured pH of kit is 6.9 at 25°C

References

1. *The advantages of using a modified microbatch method for rapid screening of protein crystallization conditions. A. D'Arcy, A. Mac Sweeney, M. Stihle and A. Haber. Acta Cryst. (2003). D59, 396-399.*
2. *Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of propionate kinase (TdcD) from Salmonella typhimurium. D. K. Simanshu and M. R. N. Murthy. Acta Cryst. (2005). F61, 52-55.*
3. *Preparation, crystallization and preliminary X-ray analysis of the methionine synthase (MetE) from Streptococcus mutans. T.-M. Fu, X.-Y. Zhang, L.-F. Li, Y.-H. Liang and X.-D. Su. Acta Cryst. (2006). F62, 984-985.*

Order Information

Each Index kit contains 96 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 36 for further details.

Cat. No.	Name	Description	Price
HR2-144	**Index**	10 ml, tube format	**$570.00**
HR2-134	**Index HT**	1 ml, Deep Well block format	**$185.00**
HR2-944-**	**Index Custom Shop**	185 ml	**$138.00**

** = reagent number 1-96

INDEX™ FORMULATION

A1	1.	0.1 M Citric Acid pH 3.5, 2.0 M Ammonium Sulfate
A2	2.	0.1 M Sodium Acetate trihydrate pH 4.5, 2.0 M Ammonium Sulfate
A3	3.	0.1 M Bis-Tris pH 5.5, 2.0 M Ammonium Sulfate
A4	4.	0.1 M Bis-Tris pH 6.5, 2.0 M Ammonium Sulfate
A5	5.	0.1 M HEPES pH 7.5, 2.0 M Ammonium Sulfate
A6	6.	0.1 M Tris pH 8.5, 2.0 M Ammonium Sulfate
A7	7.	0.1 M Citric Acid pH 3.5, 3.0 M Sodium Chloride
A8	8.	0.1 M Sodium Acetate trihydrate pH 4.5, 3.0 M Sodium Chloride
A9	9.	0.1 M Bis-Tris pH 5.5, 3.0 M Sodium Chloride
A10	10.	0.1 M Bis-Tris pH 6.5, 3.0 M Sodium Chloride
A11	11.	0.1 M HEPES pH 7.5, 3.0 M Sodium Chloride
A12	12.	0.1 M Tris pH 8.5, 3.0 M Sodium Chloride
B1	13.	0.1 M Bis-Tris pH 5.5, 0.3 M Magnesium Formate
B2	14.	0.1 M Bis-Tris pH 6.5, 0.5 M Magnesium Formate
B3	15.	0.1 M HEPES pH 7.5, 0.5 M Magnesium Formate
B4	16.	0.1 M Tris pH 8.5, 0.3 M Magnesium Formate
B5	17.	1.4 M Sodium/Potassium Phosphate pH 5.6
B6	18.	1.4 M Sodium/Potassium Phosphate pH 6.9
B7	19.	1.4 M Sodium/Potassium Phosphate pH 8.2
B8	20.	0.1 M HEPES pH 7.5, 1.4 M tri-Sodium Citrate dihydrate
B9	21.	1.8 M tri-Ammonium Citrate pH 7.0
B10	22.	0.8 M Succinic Acid pH 7.0
B11	23.	2.1 M DL-Malic Acid pH 7.0
B12	24.	2.8 M Sodium Acetate trihydrate pH 7.0
C1	25.	3.5 M Sodium Formate pH 7.0
C2	26.	1.1 M di-Ammonium Tartrate pH 7.0
C3	27.	2.4 M Sodium Malonate pH 7.0
C4	28.	35% v/v Tacsimate pH 7.0
C5	29.	60% v/v Tacsimate pH 7.0
C6	30.	0.1 M Sodium Chloride, 0.1 M Bis-Tris pH 6.5, 1.5 M Ammonium Sulfate
C7	31.	0.8 M Potassium Sodium Tartrate tetrahydrate, 0.1 M Tris pH 8.5, 0.5% w/v Polyethylene Glycol Monomethyl ether 5000
C8	32.	1.0 M Ammonium Sulfate, 0.1 M Bis-Tris pH 5.5, 1% w/v Polyethylene Glycol 3350
C9	33.	1.1 M Sodium Malonate pH 7.0, 0.1 M HEPES pH 7.0, 0.5% v/v Jeffamine ED-2001® Reagent pH 7.0
C10	34.	1.0 M Succinic Acid pH 7.0, 0.1 M HEPES pH 7.0, 1% w/v Polyethylene Glycol Monomethyl ether 2000
C11	35.	1.0 M Ammonium Sulfate, 0.1 M HEPES pH 7.0, 0.5% w/v Polyethylene Glycol 8000
C12	36.	15% v/v Tacsimate pH 7.0, 0.1 M HEPES pH 7.0, 2% w/v Polyethylene Glycol 3350
D1	37.	25% w/v Polyethylene Glycol 1500
D2	38.	0.1 M HEPES pH 7.0, 30% v/v Jeffamine M-600® Reagent pH 7.0
D3	39.	0.1 M HEPES pH 7.0, 30% v/v Jeffamine ED-2001® Reagent pH 7.0
D4	40.	0.1 M Citric Acid pH 3.5, 25% w/v Polyethylene Glycol 3350
D5	41.	0.1 M Sodium Acetate trihydrate pH 4.5, 25% w/v Polyethylene Glycol 3350
D6	42.	0.1 M Bis-Tris pH 5.5, 25% w/v Polyethylene Glycol 3350
D7	43.	0.1 M Bis-Tris pH 6.5, 25% w/v Polyethylene Glycol 3350
D8	44.	0.1 M HEPES pH 7.5, 25% w/v Polyethylene Glycol 3350
D9	45.	0.1 M Tris pH 8.5, 25% w/v Polyethylene Glycol 3350
D10	46.	0.1 M Bis-Tris pH 6.5, 20% w/v Polyethylene Glycol Monomethyl ether 5000
D11	47.	0.1 M Bis-Tris pH 6.5, 28% w/v Polyethylene Glycol Monomethyl ether 2000
D12	48.	0.2 M Calcium Chloride dihydrate, 0.1 M Bis-Tris pH 5.5, 45% v/v 2-Methyl-2,4-pentanediol
E1	49.	0.2 M Calcium Chloride dihydrate, 0.1 M Bis-Tris pH 6.5, 45% v/v 2-Methyl-2,4-pentanediol
E2	50.	0.2 M Ammonium Acetate, 0.1 M Bis-Tris pH 5.5, 45% v/v 2-Methyl-2,4-pentanediol
E3	51.	0.2 M Ammonium Acetate, 0.1 M Bis-Tris pH 6.5, 45% v/v 2-Methyl-2,4-pentanediol
E4	52.	0.2 M Ammonium Acetate, 0.1 M HEPES pH 7.5, 45% v/v 2-Methyl-2,4-pentanediol
E5	53.	0.2 M Ammonium Acetate, 0.1 M Tris pH 8.5, 45% v/v 2-Methyl-2,4-pentanediol
E6	54.	0.05 M Calcium Chloride dihydrate, 0.1 M Bis-Tris pH 6.5, 30% v/v Polyethylene Glycol Monomethyl ether 550
E7	55.	0.05 M Magnesium Chloride hexahydrate, 0.1 M HEPES pH 7.5, 30% v/v Polyethylene Glycol Monomethyl ether 550
E8	56.	0.2 M Potassium Chloride, 0.05 M HEPES pH 7.5, 35% v/v Pentaerythritol Propoxylate (5/4 PO/OH)
E9	57.	0.05 M Ammonium Sulfate, 0.05 M Bis-Tris pH 6.5, 30% v/v Pentaerythritol Ethoxylate (15/4 EO/OH)
E10	58.	0.1 M Bis-Tris pH 6.5, 45% v/v Polypropylene Glycol P 400
E11	59.	0.02 M Magnesium Chloride hexahydrate, 0.1 M HEPES pH 7.5, 22% w/v Polyacrylic Acid 5100 Sodium salt
E12	60.	0.01 M Cobalt Chloride hexahydrate, 0.1 M Tris pH 8.5, 20% w/v Polyvinylpyrrolidone K15
F1	61.	0.2 M Proline, 0.1 M HEPES pH 7.5, 10% w/v Polyethylene Glycol 3350
F2	62.	0.2 M Trimethylamine N-oxide dihydrate, 0.1 M Tris pH 8.5, 20% w/v Polyethylene Glycol Monomethyl ether 2000
F3	63.	5% v/v Tacsimate pH 7.0, 0.1 M HEPES pH 7.0, 10% w/v Polyethylene Glycol Monomethyl ether 5000
F4	64.	0.005 M Cobalt Chloride hexahydrate, 0.005 M Nickel (II) Chloride hexahydrate, 0.005 M Cadmium Chloride dihydrate, 0.005 M Magnesium Chloride hexahydrate, 0.1 M HEPES pH 7.5, 12% w/v Polyethylene Glycol 3350
F5	65.	0.1 M Ammonium Acetate, 0.1 M Bis-Tris pH 5.5, 17% w/v Polyethylene Glycol 10,000
F6	66.	0.2 M Ammonium Sulfate, 0.1 M Bis-Tris pH 5.5, 25% w/v Polyethylene Glycol 3350
F7	67.	0.2 M Ammonium Sulfate, 0.1 M Bis-Tris pH 6.5, 25% w/v Polyethylene Glycol 3350
F8	68.	0.2 M Ammonium Sulfate, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene Glycol 3350
F9	69.	0.2 M Ammonium Sulfate, 0.1 M Tris pH 8.5, 25% w/v Polyethylene Glycol 3350
F10	70.	0.2 M Sodium Chloride, 0.1 M Bis-Tris pH 5.5, 25% w/v Polyethylene Glycol 3350
F11	71.	0.2 M Sodium Chloride, 0.1 M Bis-Tris pH 6.5, 25% w/v Polyethylene Glycol 3350
F12	72.	0.2 M Sodium Chloride, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene Glycol 3350
G1	73.	0.2 M Sodium Chloride, 0.1 M Tris pH 8.5, 25% w/v Polyethylene Glycol 3350
G2	74.	0.2 M Lithium Sulfate monohydrate, 0.1 M Bis-Tris pH 5.5, 25% w/v Polyethylene Glycol 3350
G3	75.	0.2 M Lithium Sulfate monohydrate, 0.1 M Bis-Tris pH 6.5, 25% w/v Polyethylene Glycol 3350
G4	76.	0.2 M Lithium Sulfate monohydrate, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene Glycol 3350
G5	77.	0.2 M Lithium Sulfate monohydrate, 0.1 M Tris pH 8.5, 25% w/v Polyethylene Glycol 3350
G6	78.	0.2 M Ammonium Acetate, 0.1 M Bis-Tris pH 5.5, 25% w/v Polyethylene Glycol 3350
G7	79.	0.2 M Ammonium Acetate, 0.1 M Bis-Tris pH 6.5, 25% w/v Polyethylene Glycol 3350
G8	80.	0.2 M Ammonium Acetate, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene Glycol 3350
G9	81.	0.2 M Ammonium Acetate, 0.1 M Tris pH 8.5, 25% w/v Polyethylene Glycol 3350
G10	82.	0.2 M Magnesium Chloride hexahydrate, 0.1 M Bis-Tris pH 5.5, 25% w/v Polyethylene Glycol 3350
G11	83.	0.2 M Magnesium Chloride hexahydrate, 0.1 M Bis-Tris pH 6.5, 25% w/v Polyethylene Glycol 3350
G12	84.	0.2 M Magnesium Chloride hexahydrate, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene Glycol 3350
H1	85.	0.2 M Magnesium Chloride hexahydrate, 0.1 M Tris HCl pH 8.5, 25% w/v Polyethylene Glycol 3350
H2	86.	0.2 M Potassium Sodium Tartrate tetrahydrate, 20% w/v Polyethylene Glycol 3350
H3	87.	0.2 M Sodium Malonate pH 7.0, 20% w/v Polyethylene Glycol 3350
H4	88.	0.2 M tri-Ammonium Citrate pH 7.0, 20% w/v Polyethylene Glycol 3350
H5	89.	0.1 M Succinic Acid pH 7.0, 15% w/v Polyethylene Glycol 3350
H6	90.	0.2 M Sodium Formate, 20% w/v Polyethylene Glycol 3350
H7	91.	0.15 M DL-Malic Acid pH 7.0, 20% w/v Polyethylene Glycol 3350
H8	92.	0.1 M Magnesium Formate, 15% w/v Polyethylene Glycol 3350
H9	93.	0.05 M Zinc Acetate dihydrate, 20% w/v Polyethylene Glycol 3350
H10	94.	0.2 M tri-Sodium Citrate dihydrate, 20% w/v Polyethylene Glycol 3350
H11	95.	0.1 M Potassium Thiocyanate, 30% w/v Polyethylene Glycol Monomethyl ether 2000
H12	96.	0.15 M Potassium Bromide, 30% w/v Polyethylene Glycol Monomethyl ether 2000

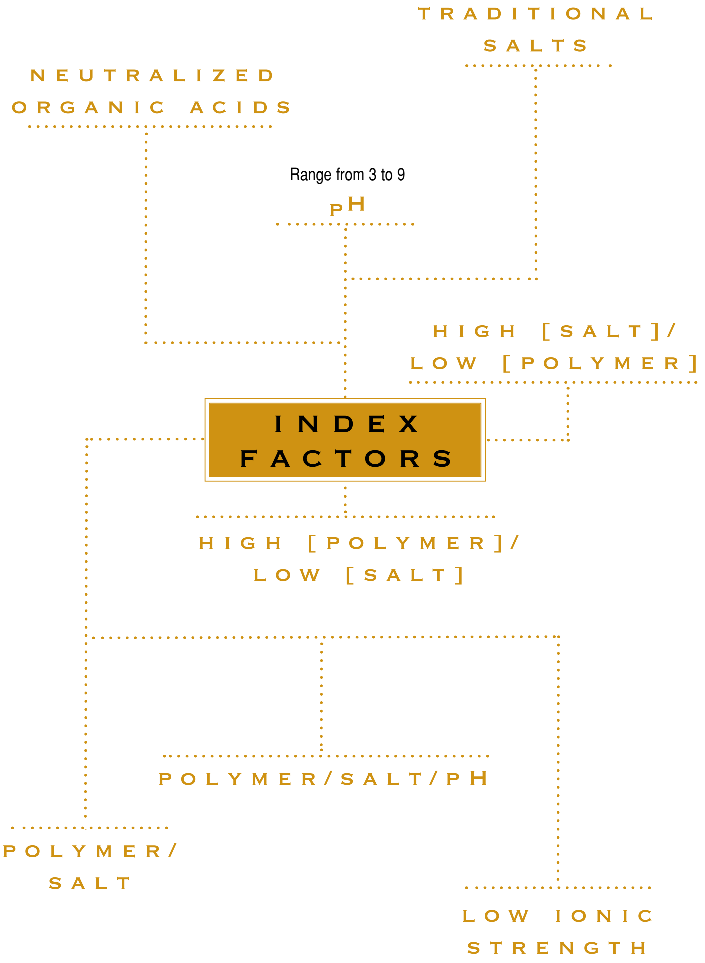

CRYSTAL SCREEN™ • CRYSTAL SCREEN 2™ • CRYSTAL SCREEN HT™

APPLICATION

- Primary screen for proteins, soluble peptides, nucleic acids, & water soluble small molecules
- Sparse matrix additive screen

FEATURES

- The original sparse matrix screen
- Sparse matrix formula efficiently samples salts, polymers, organics, & pH
- Proven effective with more than 1,000 biological macromolecules
- Tube or DeepWell block format

SUCCESS STORY

Crystal of gp120, the exterior envelope glycoprotein of type 1 HIV. Preliminary crystallization conditions obtained using Crystal Screen from Hampton Research.

Courtesy of P.D. Kwong[1], R. Wyatt[2], E. Desjardins[2], J. Robinson[3], F.C. Culp[4], B.D. Hellmig[4], R.W. Sweet[4], J. Sodroski[2], and W.A. Hendrickson[1,5].

[1]Columbia University, [2]Dana-Farber Cancer Institute, [3]Tulane University School of Medicine, [4]GlaxoSmithKline , [5]HHMI-Columbia University

DESCRIPTION

The Crystal Screen and Crystal Screen 2 reagent kits are designed to provide a highly effective and rapid screening method for the crystallization of macromolecules. The screens are simple and practical for finding initial crystallization conditions. The initial crystallization conditions for more than 1,000 proteins, peptides, oligonucleotides, and small molecules have been determined using Crystal Screen.

A highly effective approach to overcome the exhaustive search for suitable crystallization conditions is the use of a sparse matrix method of trial conditions that is biased and selected from known crystallization conditions for macromolecules. The formulation utilized in Crystal Screen and Crystal Screen 2 evaluates 96 unique mixtures of pH, salts, polymers and organics, and their ability to promote crystal growth.

Crystal Screen contains 50 unique reagents, 10 ml each and is based on the sparse matrix formulation first described by Jancarik and Kim in 1991.

Crystal Screen 2, an extension of Crystal Screen, contains 48 unique reagents, 10 ml each and is based on the formulation first described by Cudney et al in 1994.

Crystal Screen HT contains 1 ml each of reagents 1-48 from Crystal Screen and all 48 reagents from Crystal Screen 2 in a single Deep Well block format.

Ready-to-use reagents are sterile filtered and formulated with ultra-pure Type 1 water, using the highest purity salts, polymers, organics and buffers.

References

1. *Jancarik, J. & Kim, S.H. J. Appl. Cryst. 24, 409-411, (1991).*
2. *Expression, purification, crystallization and preliminary X-ray analysis of two arginine-biosynthetic enzymes from Mycobacterium tuberculosis. F. Moradian, C. Garen, L. Cherney, M. Cherney and M. N. G. James. Acta Cryst. (2006). F62, 986-988.*

Order Information

Each Crystal Screen kit contains 50 unique reagents. Each Crystal Screen 2 kit contains 48 unique reagents. To order individual reagents, use Custom Shop™ catalog numbers listed below. Refer to page 36 for further details.

Cat. No.	Name	Description	Price
HR2-110	**Crystal Screen**	10 ml, tube format	**$285.00**
HR2-112	**Crystal Screen 2**	10 ml, tube format	**$285.00**
HR2-130	**Crystal Screen HT**	1 ml, Deep Well block format	**$185.00**
HR2-910-**	**Crystal Screen Custom Shop**	185 ml	**$138.00**
HR2-912-**	**Crystal Screen 2 Custom Shop**	185 ml	**$138.00**

** = reagent number 1-50 (for Crystal Screen)
** = reagent number 1-48 (for Crystal Screen 2)

CRYSTAL SCREEN™ FORMULATION

A1	1.	0.02 M Calcium chloride dihydrate, 0.1 M Sodium acetate trihydrate pH 4.6, 30% v/v (+/-)-2-Methyl-2,4-pentanediol
A2	2.	0.4 M Potassium sodium tartrate tetrahydrate
A3	3.	0.4 M Ammonium phosphate monobasic
A4	4.	0.1 M TRIS hydrochloride pH 8.5, 2.0 M Ammonium sulfate
A5	5.	0.2 M Sodium citrate tribasic dihydrate, 0.1 M HEPES sodium pH 7.5, 30% v/v (+/-)-2-Methyl-2,4-pentanediol
A6	6.	0.2 M Magnesium chloride hexahydrate, 0.1 M TRIS hydrochloride pH 8.5, 30% w/v Polyethylene glycol 4,000
A7	7.	0.1 M Sodium cacodylate trihydrate pH 6.5, 1.4 M Sodium acetate trihydrate
A8	8.	0.2 M Sodium citrate tribasic dihydrate, 0.1 M Sodium cacodylate trihydrate pH 6.5, 30% v/v 2-Propanol
A9	9.	0.2 M Ammonium acetate, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 30% w/v Polyethylene glycol 4,000
A10	10.	0.2 M Ammonium acetate, 0.1 M Sodium acetate trihydrate pH 4.6, 30% w/v Polyethylene glycol 4,000
A11	11.	0.1 M Sodium citrate tribasic dihydrate pH 5.6, 1.0 M Ammonium phosphate monobasic
A12	12.	0.2 M Magnesium chloride hexahydrate, 0.1 M HEPES sodium pH 7.5, 30% v/v 2-Propanol
B1	13.	0.2 M Sodium citrate tribasic dihydrate, 0.1 M TRIS hydrochloride pH 8.5, 30% v/v Polyethylene glycol 400
B2	14.	0.2 M Calcium chloride dihydrate, 0.1 M HEPES sodium pH 7.5, 28% v/v Polyethylene glycol 400
B3	15.	0.2 M Ammonium sulfate, 0.1 M Sodium cacodylate trihydrate pH 6.5, 30% w/v Polyethylene glycol 8,000
B4	16.	0.1 M HEPES sodium pH 7.5, 1.5 M Lithium sulfate monohydrate
B5	17.	0.2 M Lithium sulfate monohydrate, 0.1 M TRIS hydrochloride pH 8.5, 30% w/v Polyethylene glycol 4,000
B6	18.	0.2 M Magnesium acetate tetrahydrate, 0.1 M Sodium cacodylate trihydrate pH 6.5, 20% w/v Polyethylene glycol 8,000
B7	19.	0.2 M Ammonium acetate, 0.1 M TRIS hydrochloride pH 8.5, 30% v/v 2-Propanol
B8	20.	0.2 M Ammonium sulfate, 0.1 M Sodium acetate trihydrate pH 4.6, 25% w/v Polyethylene glycol 4,000
B9	21.	0.2 M Magnesium acetate tetrahydrate, 0.1 M Sodium cacodylate trihydrate pH 6.5, 30% v/v (+/-)-2-Methyl-2,4-pentanediol
B10	22.	0.2 M Sodium acetate trihydrate, 0.1 M TRIS hydrochloride pH 8.5, 30% w/v Polyethylene glycol 4,000
B11	23.	0.2 M Magnesium chloride hexahydrate, 0.1 M HEPES sodium pH 7.5, 30% v/v Polyethylene glycol 400
B12	24.	0.2 M Calcium chloride dihydrate, 0.1 M Sodium acetate trihydrate pH 4.6, 20% v/v 2-Propanol
C1	25.	0.1 M Imidazole pH 6.5, 1.0 M Sodium acetate trihydrate
C2	26.	0.2 M Ammonium acetate, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 30% v/v (+/-)-2-Methyl-2,4-pentanediol
C3	27.	0.2 M Sodium citrate tribasic dihydrate, 0.1 M HEPES sodium pH 7.5, 20% v/v 2-Propanol
C4	28.	0.2 M Sodium acetate trihydrate, 0.1 M Sodium cacodylate trihydrate pH 6.5, 30% w/v Polyethylene glycol 8,000
C5	29.	0.1 M HEPES sodium pH 7.5, 0.8 M Potassium sodium tartrate tetrahydrate
C6	30.	0.2 M Ammonium sulfate, 30% w/v Polyethylene glycol 8,000
C7	31.	0.2 M Ammonium sulfate, 30% w/v Polyethylene glycol 4,000
C8	32.	2.0 M Ammonium sulfate
C9	33.	4.0 M Sodium formate
C10	34.	0.1 M Sodium acetate trihydrate pH 4.6, 2.0 M Sodium formate
C11	35.	0.1 M HEPES sodium pH 7.5, 0.8 M Sodium phosphate monobasic monohydrate, 0.8 M Potassium phosphate monobasic
C12	36.	0.1 M TRIS hydrochloride pH 8.5, 8% w/v Polyethylene glycol 8,000
D1	37.	0.1 M Sodium acetate trihydrate pH 4.6, 8% w/v Polyethylene glycol 4,000
D2	38.	0.1 M HEPES sodium pH 7.5, 1.4 M Sodium citrate tribasic dihydrate
D3	39.	0.1 M HEPES sodium pH 7.5, 2% v/v Polyethylene glycol 400, 2.0 M Ammonium sulfate
D4	40.	0.1 M Sodium citrate tribasic dihydrate pH 5.6, 20% v/v 2-Propanol, 20% w/v Polyethylene glycol 4,000
D5	41.	0.1 M HEPES sodium pH 7.5, 10% v/v 2-Propanol, 20% w/v Polyethylene glycol 4,000
D6	42.	0.05 M Potassium phosphate monobasic, 20% w/v Polyethylene glycol 8,000
D7	43.	30% w/v Polyethylene glycol 1,500
D8	44.	0.2 M Magnesium formate dihydrate
D9	45.	0.2 M Zinc acetate dihydrate, 0.1 M Sodium cacodylate trihydrate pH 6.5, 18% w/v Polyethylene glycol 8,000
D10	46.	0.2 M Calcium acetate hydrate, 0.1 M Sodium cacodylate trihydrate pH 6.5 , 18% w/v Polyethylene glycol 8,000
D11	47.	0.1 M Sodium acetate trihydrate pH 4.6, 2.0 M Ammonium sulfate
D12	48.	0.1 M TRIS hydrochloride pH 8.5, 2.0 M Ammonium phosphate monobasic
	49.	1.0 M Lithium sulfate monohydrate, 2% w/v Polyethylene glycol 8,000
	50.	0.5 M Lithium sulfate monohydrate, 15% w/v Polyethylene glycol 8,000

CRYSTAL SCREEN FACTORS

BUFFER
HEPES sodium
Imidazole
Sodium acetate
Sodium cacodylate
Sodium citrate
TRIS hydrochloride

pH
Range from 4.0 to 9.0

ORGANIC
2-Propanol

NON-VOLATILE ORGANIC
(+/-)-2-Methyl-2,4-pentanediol

POLYMER
Polyethylene glycol 400
Polyethylene glycol 1,500
Polyethylene glycol 4,000
Polyethylene glycol 8,000

SALT
Ammonium acetate
Calcium acetate
Magnesium acetate
Sodium acetate
Zinc acetate
Calcium chloride
Magnesium chloride
Sodium citrate
Magnesium formate
Sodium formate
Ammonium phosphate
Potassium phosphate
Sodium phosphate
Ammonium sulfate
Lithium sulfate
K/Na tartrate

CRYSTAL SCREEN2™ FORMULATION

E1	1.	2.0 M Sodium chloride, 10% w/v Polyethylene glycol 6,000
E2	2.	0.5 M Sodium chloride, 0.01 M Magnesium chloride hexahydrate, 0.01 M Hexadecyltrimethylammonium bromide
E3	3.	25% v/v Ethylene glycol
E4	4.	35% v/v 1,4-Dioxane
E5	5.	2.0 M Ammonium sulfate, 5% v/v 2-Propanol
E6	6.	1.0 M Imidazole pH 7.0
E7	7.	10% w/v Polyethylene glycol 1,000, 10% w/v Polyethylene glycol 8,000
E8	8.	1.5 M Sodium chloride, 10% v/v Ethanol
E9	9.	0.1 M Sodium acetate trihydrate pH 4.6, 2.0 M Sodium chloride
E10	10.	0.2 M Sodium chloride, 0.1 M Sodium acetate trihydrate pH 4.6, 30% v/v (+/-)-2-Methyl-2,4-pentanediol
E11	11.	0.01 M Cobalt(II) chloride hexahydrate, 0.1 M Sodium acetate trihydrate pH 4.6, 1.0 M 1,6-Hexanediol
E12	12.	0.1 M Cadmium chloride hydrate, 0.1 M Sodium acetate trihydrate pH 4.6, 30% v/v Polyethylene glycol 400
F1	13.	0.2 M Ammonium sulfate, 0.1 M Sodium acetate trihydrate pH 4.6, 30% w/v Polyethylene glycol monomethyl ether 2,000
F2	14.	0.2 M Potassium sodium tartrate tetrahydrate, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 2.0 M Ammonium sulfate
F3	15.	0.5 M Ammonium sulfate, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 1.0 M Lithium sulfate monohydrate
F4	16.	0.5 M Sodium chloride, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 2% v/v Ethylene imine polymer
F5	17.	0.1 M Sodium citrate tribasic dihydrate pH 5.6, 35% v/v tert-Butanol
F6	18.	0.01 M Iron(III) chloride hexahydrate, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 10% v/v Jeffamine® M-600®
F7	19.	0.1 M Sodium citrate tribasic dihydrate pH 5.6, 2.5 M 1,6-Hexanediol
F8	20.	0.1 M MES monohydrate pH 6.5, 1.6 M Magnesium sulfate heptahydrate
F9	21.	0.1 M Sodium phosphate monobasic monohydrate, 0.1 M Potassium phosphate monobasic, 0.1 M MES monohydrate pH 6.5, 2.0 M Sodium chloride
F10	22.	0.1 M MES monohydrate pH 6.5, 12% w/v Polyethylene glycol 20,000
F11	23.	1.6 M Ammonium sulfate, 0.1 M MES monohydrate pH 6.5, 10% v/v 1,4-Dioxane
F12	24.	0.05 M Cesium chloride, 0.1 M MES monohydrate pH 6.5, 30% v/v Jeffamine® M-600®
G1	25.	0.01 M Cobalt(II) chloride hexahydrate, 0.1 M MES monohydrate pH 6.5, 1.8 M Ammonium sulfate
G2	26.	0.2 M Ammonium sulfate, 0.1 M MES monohydrate pH 6.5, 30% w/v Polyethylene glycol monomethyl ether 5,000
G3	27.	0.01 M Zinc sulfate heptahydrate, 0.1 M MES monohydrate pH 6.5, 25% v/v Polyethylene glycol monomethyl ether 550
G4	28.	1.6 M Sodium citrate tribasic dihydrate pH 6.5
G5	29.	0.5 M Ammonium sulfate, 0.1 M HEPES pH 7.5, 30% v/v (+/-)-2-Methyl-2,4-pentanediol
G6	30.	0.1 M HEPES pH 7.5, 10% w/v Polyethylene glycol 6,000, 5% v/v (+/-)-2-Methyl-2,4-pentanediol
G7	31.	0.1 M HEPES pH 7.5, 20% v/v Jeffamine® M-600®
G8	32.	0.1 M Sodium chloride, 0.1 M HEPES pH 7.5, 1.6 M Ammonium sulfate
G9	33.	0.1 M HEPES pH 7.5, 2.0 M Ammonium formate
G10	34.	0.05 M Cadmium sulfate hydrate, 0.1 M HEPES pH 7.5, 1.0 M Sodium acetate trihydrate
G11	35.	0.1 M HEPES pH 7.5, 70% v/v (+/-)-2-Methyl-2,4-pentanediol
G12	36.	0.1 M HEPES pH 7.5, 4.3 M Sodium chloride
H1	37.	0.1 M HEPES pH 7.5, 10% w/v Polyethylene glycol 8,000, 8% v/v Ethylene glycol
H2	38.	0.1 M HEPES pH 7.5, 20% w/v Polyethylene glycol 10,000
H3	39.	0.2 M Magnesium chloride hexahydrate, 0.1 M Tris pH 8.5, 3.4 M 1,6-Hexanediol
H4	40.	0.1 M Tris pH 8.5, 25% v/v tert-Butanol
H5	41.	0.01 M Nickel(II) chloride hexahydrate, 0.1 M Tris pH 8.5, 1.0 M Lithium sulfate monohydrate
H6	42.	1.5 M Ammonium sulfate, 0.1 M Tris pH 8.5, 12% v/v Glycerol
H7	43.	0.2 M Ammonium phosphate monobasic, 0.1 M Tris pH 8.5, 50% v/v (+/-)-2-Methyl-2,4-pentanediol
H8	44.	0.1 M Tris pH 8.5, 20% v/v Ethanol
H9	45.	0.01 M Nickel(II) chloride hexahydrate, 0.1 M Tris pH 8.5, 20% w/v Polyethylene glycol monomethyl ether 2,000
H10	46.	0.1 M Sodium chloride, 0.1 M BICINE pH 9.0, 20% v/v Polyethylene glycol monomethyl ether 550
H11	47.	0.1 M BICINE pH 9.0, 2.0 M Magnesium chloride hexahydrate
H12	48.	0.1 M BICINE pH 9.0, 2% v/v 1,4-Dioxane, 10% w/v Polyethylene glycol 20,000

CRYSTAL SCREEN 2 FACTORS

BUFFER
BICINE
HEPES
Imidazole
MES
Sodium acetate
Sodium citrate
TRIS hydrochloride

pH
Range from 4.0 to 9.0

NON-VOLATILE ORGANIC
1,6-Hexanediol
(+/-)-2-Methyl-2,4-pentanediol
Ethylene glycol
Glycerol

ORGANIC
1,4-Dioxane
Ethanol
2-Propanol
tert-Butanol

POLYMER
Ethylene imine polymer
Jeffamine® M-600®
Polyethylene glycol 400
Polyethylene glycol 1,000
Polyethylene glycol 6,000
Polyethylene glycol 8,000
Polyethylene glycol 10,000
Polyethylene glycol 20,000
Polyethylene glycol MME 550
Polyethylene glycol MME 2,000
Polyethylene glycol MME 5,000

SALT
Sodium acetate
Cadmium chloride
Cesium chloride
Cobalt(II) chloride
Iron(III) chloride
Magnesium chloride
Nickel(II) chloride
Sodium chloride
Sodium citrate
Ammonium formate
Ammonium phosphate
Potassium phosphate
Sodium phosphate
Ammonium sulfate
Cadmium sulfate
Lithium sulfate
Magnesium sulfate
Zinc sulfate
K/Na tartrate
Hexadecyltrimethylammonium bromide

APPLICATION

- Primary and secondary, polymer and pH based crystallization screen for biological macromolecules

FEATURES

- Developed at Hampton Research
- PEGRx 1 primary screen variables are polymer type (16 different polymers), polymer molecular weight, pH and low ionic strength.
- PEGRx 2 primary screen variables are polymer type (13 different polymers), polymer molecular weight, pH and secondary reagents which include additives, salts, volatile organics and polyols.
- PEGRx HT combines PEGRx 1 and PEGRx 2 in a single 96 Deep Well block
- pH range 3.5 - 9, using 10 different buffer systems
- Polymer molecular weight range 200 to 20,000
- Tube or Deep Well block format

DESCRIPTION

PEGRx is a primary and secondary, polymer and pH based crystallization screen developed at Hampton Research. It is designed to evaluate polymer based crystallization reagents and pH in low (PEGRx 1) to medium ionic strength (PEGRx 2). It is also designed for use as a secondary or optimization screen to follow the Hampton Research Index screen when polymer based reagents produce hits and interesting solubility leads.

PEGRx 1 is a crystallization reagent kit designed to evaluate an array of polymers of varying molecular weight in a low ionic strength environment versus a wide range of pH. Polymer reagents include Polyethylene glycols, Polyethylene glycol monomethylethers, and Jeffamines. The molecular weight range between 200 and 20,000 is evaluated in a low ionic strength formulation. Ten different buffers are used to span the range of pH between 3.5 and 9. The primary screen variables are polymer type, polymer molecular weight, pH and low ionic strength.

PEGRx 2 is a crystallization reagent kit designed to evaluate an array of polymers of varying molecular weight in a medium ionic strength environment in the presence of additives, salts, volatile organics and polyols versus a wide range of pH. Polymer reagents include Polyethylene glycols and Polyethylene glycol monomethylethers. The polymer molecular weight range between 200 and 20,000 is evaluated in a medium ionic strength formulation. Ten different buffers are used to span the range of pH between 3.5 and 9. The primary screen variables are polymer type, polymer molecular weight, pH and secondary reagents which include additives, salts, volatile organics and polyols.

PEGRx 1 contains 48 unique reagents, 10 ml each.

PEGRx 2 contains 48 unique reagents, 10 ml each.

PEGRx HT contains 1 ml of each reagent from PEGRx 1 and PEGRx 2 in a single Deep Well block format.

Ready-to-use reagents are sterile filtered and formulated with ultra-pure Type 1 water, using the highest purity salts, polymers, organics and buffers.

Order Information

PEGRx 1 and PEGRx 2 kits each contain 48 unique reagents. To order individual reagents, use Custom Shop™ catalog numbers listed below. Refer to page 36 for further details.

Cat. No.	Name	Description	Price
HR2-082	**PEGRx 1**	10 ml, tube format	**$285.00**
HR2-084	**PEGRx 2**	10 ml, tube format	**$285.00**
HR2-086	**PEGRx HT**	1 ml, Deep Well block format	**$185.00**
HR2-982-**	**PEGRx 1 Custom Shop**	185 ml	**$138.00**
HR2-984-**	**PEGRx 2 Custom Shop**	185 ml	**$138.00**

** = reagent number 1-48

PEGRX 1™ FORMULATION

A1	1.	0.1 M Citric acid pH 3.5, 34% v/v Polyethylene glycol 200
A2	2.	0.1 M Sodium citrate tribasic dihydrate pH 5.5, 38% v/v Polyethylene glycol 200
A3	3	0.1 M HEPES pH 7.5, 42% v/v Polyethylene glycol 200
A4	4.	0.1 M Sodium acetate trihydrate pH 4.5, 30% v/v Polyethylene glycol 300
A5	5.	0.1 M BIS-TRIS pH 6.5, 25% v/v Polyethylene glycol 300
A6	6.	0.1 M BICINE pH 8.5, 20% v/v Polyethylene glycol 300
A7	7.	0.1 M Sodium acetate trihydrate pH 4.0, 15% v/v Polyethylene glycol 400
A8	8.	0.1 M MES monohydrate pH 6.0, 22% v/v Polyethylene glycol 400
A9	9.	0.1 M Tris pH 8.0, 30% v/v Polyethylene glycol 400
A10	10.	0.1 M Sodium citrate tribasic dihydrate pH 5.0, 30% v/v Polyethylene glycol monomethyl ether 550
A11	11.	0.1 M Imidazole pH 7.0, 25% v/v Polyethylene glycol monomethyl ether 550
A12	12.	0.1 M BIS-TRIS propane pH 9.0, 20% v/v Polyethylene glycol monomethyl ether 550
B1	13.	0.1 M Sodium acetate trihydrate pH 4.0, 10% v/v Jeffamine® M-600® pH 7.0
B2	14.	0.1 M MES monohydrate pH 6.0, 20% v/v Jeffamine® M-600® pH 7.0
B3	15.	0.1 M Tris pH 8.0, 30% v/v Jeffamine® M-600® pH 7.0
B4	16.	0.1 M Citric acid pH 3.5, 14% w/v Polyethylene glycol 1,000
B5	17.	0.1 M Sodium citrate tribasic dihydrate pH 5.5, 22% w/v Polyethylene glycol 1,000
B6	18.	0.1 M HEPES pH 7.5, 30% w/v Polyethylene glycol 1,000
B7	19.	0.1 M Sodium acetate trihydrate pH 4.5, 30% w/v Polyethylene glycol 1,500
B8	20.	0.1 M BIS-TRIS pH 6.5, 20% w/v Polyethylene glycol 1,500
B9	21.	0.1 M BICINE pH 8.5, 15% w/v Polyethylene glycol 1,500
B10	22.	0.1 M Sodium acetate trihydrate pH 4.0, 10% w/v Polyethylene glycol monomethyl ether 2,000
B11	23.	0.1 M MES monohydrate pH 6.0, 20% w/v Polyethylene glycol monomethyl ether 2,000
B12	24.	0.1 M Tris pH 8.0, 30% w/v Polyethylene glycol monomethyl ether 2,000
C1	25.	0.1 M Sodium citrate tribasic dihydrate pH 5.0, 30% v/v Jeffamine® ED-2001 pH 7.0
C2	26.	0.1 M Imidazole pH 7.0, 20% v/v Jeffamine® ED-2001 pH 7.0
C3	27.	0.1 M BIS-TRIS propane pH 9.0, 10% v/v Jeffamine® ED-2001 pH 7.0
C4	28.	0.1 M Citric acid pH 3.5, 25% w/v Polyethylene glycol 3,350
C5	29.	0.1 M Sodium citrate tribasic dihydrate pH 5.5, 18% w/v Polyethylene glycol 3,350
C6	30.	0.1 M HEPES pH 7.0, 12% w/v Polyethylene glycol 3,350
C7	31.	0.1 M Sodium acetate trihydrate pH 4.0, 10% w/v Polyethylene glycol 4,000
C8	32.	0.1 M MES monohydrate pH 6.0, 14% w/v Polyethylene glycol 4,000
C9	33.	0.1 M Tris pH 8.0, 28% w/v Polyethylene glycol 4,000
C10	34.	0.1 M Sodium acetate trihydrate pH 4.5, 30% w/v Polyethylene glycol monomethyl ether 5,000
C11	35.	0.1 M BIS-TRIS pH 6.5, 20% w/v Polyethylene glycol monomethyl ether 5,000
C12	36.	0.1 M BICINE pH 8.5, 8% w/v Polyethylene glycol monomethyl ether 5,000
D1	37.	0.1 M Sodium citrate tribasic dihydrate pH 5.0, 10% w/v Polyethylene glycol 6,000
D2	38.	0.1 M Imidazole pH 7.0, 20% w/v Polyethylene glycol 6,000
D3	39.	0.1 M BIS-TRIS propane pH 9.0, 30% w/v Polyethylene glycol 6,000
D4	40.	0.1 M Citric acid pH 3.5, 28% w/v Polyethylene glycol 8,000
D5	41.	0.1 M Sodium citrate tribasic dihydrate pH 5.5, 16% w/v Polyethylene glycol 8,000
D6	42.	0.1 M HEPES pH 7.5, 4% w/v Polyethylene glycol 8,000
D7	43.	0.1 M Sodium acetate trihydrate pH 4.5, 10% w/v Polyethylene glycol 10,000
D8	44.	0.1 M BIS-TRIS pH 6.5, 16% w/v Polyethylene glycol 10,000
D9	45.	0.1 M BICINE pH 8.5, 20% w/v Polyethylene glycol 10,000
D10	46.	0.1 M Sodium citrate tribasic dihydrate pH 5.0, 18% w/v Polyethylene glycol 20,000
D11	47.	0.1 M Imidazole pH 7.0, 12% w/v Polyethylene glycol 20,000
D12	48.	0.1 M BIS-TRIS propane pH 9.0, 8% w/v Polyethylene glycol 20,000

Range from 3.5 to 9

pH

PEGRX 1 FACTORS

BUFFERS

BICINE
BIS-TRIS
BIS-TRIS propane
Citric acid
HEPES
Imidazole
MES monohydrate
Sodium acetate
Sodium citrate
Tris

POLYMER

Jeffamine® ED-2001
Jeffamine® M-600®
Polyethylene glycol 200
Polyethylene glycol 300
Polyethylene glycol 400
Polyethylene glycol 1,000
Polyethylene glycol 1,500
Polyethylene glycol 3,350
Polyethylene glycol 4,000
Polyethylene glycol 6,000
Polyethylene glycol 8,000
Polyethylene glycol 10,000
Polyethylene glycol 20,000
Polyethylene glycol MME 550
Polyethylene glycol MME 2,000
Polyethylene glycol MME 5,000

PEGRX 2™ FORMULATION

E1	1.	0.8 M Lithium sulfate monohydrate, 0.1 M Sodium acetate trihydrate pH 4.0, 4% v/v Polyethylene glycol 200
E2	2.	0.2 M Lithium sulfate monohydrate, 0.1 M Sodium citrate tribasic dihydrate pH 5.0, 26% v/v Polyethylene glycol 200
E3	3.	0.05 M Calcium chloride dihydrate, 0.1 M MES monohydrate pH 6.0, 45% v/v Polyethylene glycol 200
E4	4.	28% v/v 2-Propanol, 0.1 M BIS-TRIS pH 6.5, 3% v/v Polyethylene glycol 200
E5	5.	20% v/v Tacsimate pH 7.0, 0.1 M HEPES pH 7.5, 2% v/v Polyethylene glycol 200
E6	6.	10% v/v 2-Propanol, 0.1 M Sodium citrate tribasic dihydrate pH 5.0, 26% v/v Polyethylene glycol 400
E7	7.	0.2 M Ammonium acetate, 0.1 M Sodium citrate tribasic dihydrate pH 5.5, 24% v/v Polyethylene glycol 400
E8	8.	0.2 M Ammonium sulfate, 0.1 M BIS-TRIS pH 6.5, 18% v/v Polyethylene glycol 400
E9	9.	0.19 mM CYMAL®-7, 0.1 M HEPES pH 7.5, 40% v/v Polyethylene glycol 400
E10	10.	6% v/v 2-Propanol, 0.1 M Sodium acetate trihydrate pH 4.5, 26% v/v Polyethylene glycol monomethyl ether 550
E11	11.	1.8 M Ammonium sulfate, 0.1 M BIS-TRIS pH 6.5, 2% v/v Polyethylene glycol monomethyl ether 550
E12	12.	0.15 M DL-Malic acid pH 7.0, 0.1 M Imidazole pH 7.0, 22% v/v Polyethylene glycol monomethyl ether 550
F1	13.	0.1 M Succinic acid pH 7.0, 0.1 M BICINE pH 8.5, 30% v/v Polyethylene glycol monomethyl ether 550
F2	14.	0.1 M Lithium sulfate monohydrate, 0.1 M Sodium citrate tribasic dihydrate pH 5.5, 20% w/v Polyethylene glycol 1,000
F3	15.	0.1 M Sodium malonate pH 8.0, 0.1 M Tris pH 8.0, 30% w/v Polyethylene glycol 1,000
F4	16.	4% v/v (+/-)-2-Methyl-2,4-pentanediol, 0.1 M Citric acid pH 3.5, 20% w/v Polyethylene glycol 1,500
F5	17.	0.2 M L-Proline, 0.1 M HEPES pH 7.5, 24% w/v Polyethylene glycol 1,500
F6	18.	10% v/v 2-Propanol, 0.1 M BICINE pH 8.5, 30% w/v Polyethylene glycol 1,500
F7	19.	0.1 M Sodium chloride, 0.1 M BIS-TRIS propane pH 9.0, 25% w/v Polyethylene glycol 1,500
F8	20.	0.02 M Nickel(II) chloride hexahydrate, 0.02 M Magnesium chloride hexahydrate, 0.02 M Cadmium chloride hydrate, 0.1 M Sodium acetate trihydrate pH 4.5, 24% w/v Polyethylene glycol monomethyl ether 2,000
F9	21.	20% v/v 2-Propanol, 0.1 M MES monohydrate pH 6.0, 20% w/v Polyethylene glycol monomethyl ether 2,000
F10	22.	0.2 M Ammonium citrate tribasic pH 7.0, 0.1 M Imidazole pH 7.0, 20% w/v Polyethylene glycol monomethyl ether 2,000
F11	23.	4.0 M Potassium formate, 0.1 M BIS-TRIS propane pH 9.0, 2% w/v Polyethylene glycol monomethyl ether 2,000
F12	24.	50% v/v Tacsimate pH 4.0, 0.1 M Sodium acetate trihydrate pH 4.5, 1% w/v Polyethylene glycol 3,350
G1	25.	0.10% w/v n-Octyl-β-D-glucoside, 0.1 M Sodium citrate tribasic dihydrate pH 5.5, 22% w/v Polyethylene glycol 3,350
G2	26.	2% v/v Tacsimate pH 7.0, 5% v/v 2-Propanol, 0.1 M Imidazole pH 7.0, 8% w/v Polyethylene glycol 3,350
G3	27.	2% v/v 1,4-Dioxane, 0.1 M Tris pH 8.0, 15% w/v Polyethylene glycol 3,350
G4	28.	18% v/v 2-Propanol, 0.1 M Sodium citrate tribasic dihydrate pH 5.5, 20% w/v Polyethylene glycol 4,000
G5	29.	6% v/v Tacsimate pH 6.0, 0.1 M MES monohydrate pH 6.0, 25% w/v Polyethylene glycol 4,000
G6	30.	0.2 M Magnesium formate dihydrate, 0.1 M Sodium acetate trihydrate pH 4.0, 18% w/v Polyethylene glycol monomethyl ether 5,000
G7	31.	2% v/v Polyethylene glycol 400, 0.1 M Imidazole pH 7.0, 24% w/v Polyethylene glycol monomethyl ether 5,000
G8	32.	0.2 M Sodium formate, 0.1 M BICINE pH 8.5, 20% w/v Polyethylene glycol monomethyl ether 5,000
G9	33.	4% v/v 2-Propanol, 0.1 M BIS-TRIS propane pH 9.0, 20% w/v Polyethylene glycol monomethyl ether 5,000
G10	34.	6% v/v Ethylene glycol, 0.1 M Citric acid pH 3.5, 10% w/v Polyethylene glycol 6,000
G11	35.	0.15 M Lithium sulfate monohydrate, 0.1 M Citric acid pH 3.5, 18% w/v Polyethylene glycol 6,000
G12	36.	10% v/v 2-Propanol, 0.1 M Sodium acetate trihydrate pH 4.0, 22% w/v Polyethylene glycol 6,000
H1	37.	0.2 M Sodium chloride, 0.1 M Sodium acetate trihydrate pH 4.0, 22% w/v Polyethylene glycol 8,000
H2	38.	20% v/v 2-Propanol, 0.1 M Tris pH 8.0, 5% w/v Polyethylene glycol 8,000
H3	39.	10% v/v Polyethylene glycol 200, 0.1 M BIS-TRIS propane pH 9.0, 18% w/v Polyethylene glycol 8,000
H4	40.	15% v/v 2-Propanol, 0.1 M Sodium citrate tribasic dihydrate pH 5.0, 10% w/v Polyethylene glycol 10,000
H5	41.	0.4 M Sodium malonate pH 6.0, 0.1 M MES monohydrate pH 6.0, 0.5% w/v Polyethylene glycol 10,000
H6	42.	0.2 M Potassium sodium tartrate tetrahydrate, 0.1 M BIS-TRIS pH 6.5, 10% w/v Polyethylene glycol 10,000
H7	43.	5% v/v (+/-)-2-Methyl-2,4-pentanediol, 0.1 M HEPES pH 7.5, 10% w/v Polyethylene glycol 10,000
H8	44.	0.2 M Ammonium acetate, 0.1 M Tris pH 8.0, 16% w/v Polyethylene glycol 10,000
H9	45.	5% v/v 2-Propanol, 0.1 M Citric acid pH 3.5, 6% w/v Polyethylene glycol 20,000
H10	46.	1.0 M Sodium malonate pH 5.0, 0.1 M Sodium acetate trihydrate pH 4.5, 2% w/v Polyethylene glycol 20,000
H11	47.	0.2 M Magnesium chloride hexahydrate, 0.1 M Sodium citrate tribasic dihydrate pH 5.0, 10% w/v Polyethylene glycol 20,000
H12	48.	3% w/v Dextran sulfate sodium salt, 0.1 M BICINE pH 8.5, 15% w/v Polyethylene glycol 20,000

Range from 3.5 to 9

pH

BICINE
BIS-TRIS
BIS-TRIS propane
Citric acid
HEPES
Imidazole
MES monohydrate
Sodium acetate
Sodium citrate
Tris

BUFFERS

Additive
Cation
Detergent
Organic Acid
Polyol
Volatile Organic

SECONDARY REAGENTS

PEGRX 2 FACTORS

POLYMERS

Polyethylene glycol 200
Polyethylene glycol 400
Polyethylene glycol 1,000
Polyethylene glycol 1,500
Polyethylene glycol 3,350
Polyethylene glycol 4,000
Polyethylene glycol 6,000
Polyethylene glycol 8,000
Polyethylene glycol 10,000
Polyethylene glycol 20,000
Polyethylene glycol MME 550
Polyethylene glycol MME 2,000
Polyethylene glycol MME 5,000

PEG/ION SCREEN™ • PEG/ION 2 SCREEN™ • PEG/ION HT™

APPLICATION

- Primary or secondary, polymer, salt and pH matrix crystallization screen for biological macromolecules

FEATURES

- Developed at Hampton Research
- PEG/Ion is a sparse matrix profile of anions and cations in the presence of monodisperse Polyethylene glycol 3,350 over pH 4.5 - 9.2
- PEG/Ion 2 screens a complete profile of titrated organic acids at varying pH levels (3.7 - 8.8) in the presence of monodisperse PEG 3,350
- PEG/Ion HT combines PEG/Ion and PEG/Ion 2 in a single 96 Deep Well block
- Tube or Deep Well block format

SUCCESS STORY

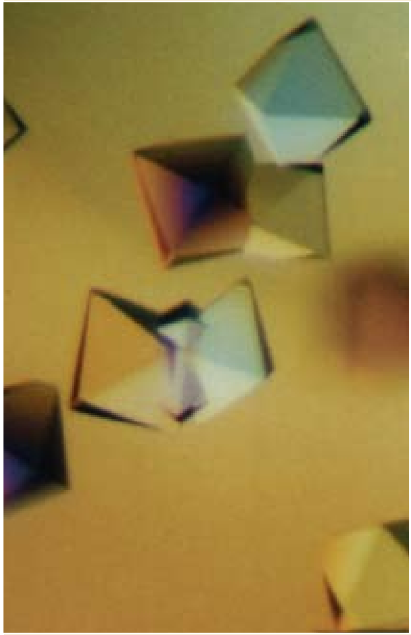

Crystals of Proteinase K. Preliminary crystallization conditions obtained using PEG/Ion Screen from Hampton Research.

Crystals grown at Hampton Research.

DESCRIPTION

PEG/Ion, developed by Hampton Research, is a crystallization screen designed to evaluate monodisperse, high purity Polyethylene glycol 3,350 and 48 unique salts representing a very complete range of anions and cations frequently used in the crystallization of biological macromolecules. The primary screening variables are PEG, ion type, ionic strength, and pH. More than 60% of the published crystallizations utilized PEG as a primary crystallization reagent and in approximately 50% of those reports, the PEG was combined with an ion as a secondary crystallization reagent.

PEG/Ion 2 is an extension to the fundamental crystallization strategy in PEG/Ion. PEG/Ion 2 reagents cover the monodisperse, high purity Polyethylene glycol 3,350 and an array of neutralized and pH adjusted organic acids, multivalent ions, a novel Citrate BIS-TRIS propane buffer system and pH (4 - 8.8). The formulation of PEG/Ion 2 was developed at Hampton Research. Each of the 48 reagents in PEG/Ion 2 contains PEG 3,350 as the polymer (precipitant). The concentration of PEG is varied from 12% w/v to 20% w/v depending upon the type and concentration of buffer/salt paired with the polymer. Thirteen of the forty-eight PEG/Ion 2 reagents contain a separate buffer component. The remaining PEG/Ion 2 reagents are buffered by the titrated organic acid salt. Six of these thirteen conditions feature a novel Citric acid BIS-TRIS propane (CBTP) buffer. The CBTP buffer uses Citric acid and BIS-TRIS propane as the acid base pair to create a two component buffer system effective across pH 2.5 to 9.5. The ratio of Citric acid to BIS-TRIS propane determines the solution pH. Thirty-five of the forty-eight PEG/Ion 2 reagents contain a neutralized or pH adjusted organic acid in the presence of the polymer. Neutralized organic acids are highly effective crystallization salts.[1] Four PEG/Ion 2 reagents feature polyvalent cations. Two of these reagents contain cation mixes, saving sample by screening six different cations with only two reagents. Tryptone, a casein digest combinatorial library of peptides, is included in PEG/Ion 2.

PEG/Ion contains 48 unique reagents, 10 ml each.

PEG/Ion 2 contains 48 unique reagents, 10 ml each.

PEG/Ion HT contains 1 ml of each reagent from PEG/Ion and PEG/Ion 2 in a single Deep Well block format.

Ready-to-use reagents are sterile filtered and formulated with ultra-pure Type 1 water, using the highest purity salts, polymers, organics and buffers.

PEG/Ion 2

Measured pH range of kit is 3.7 to 8.8 at 25°C
Average measured pH of kit is 6.4 at 25°C
Median measured pH of kit is 6.7 at 25°C
Mode measured pH of kit is 6.7 at 25°C

References

1. *A comparison of salts for the crystallization of macromolecules. Alexander McPherson. Protein Science (2001), 10:418-422.*
2. *Searching for silver bullets: An alternative strategy for crystallizing macromolecules. Alexander McPherson and Bob Cudney. Journal of Structural Biology Volume 156, Issue 3 , December 2006, Pages 387-406*

Order Information

PEG/Ion Screen and PEG/Ion 2 Screen kits each contain 48 unique reagents. To order individual reagents, use Custom Shop™ catalog numbers listed below. Refer to page 36 for further details.

Cat. No.	Name	Description	Price
HR2-126	**PEG/Ion Screen**	10 ml, tube format	**$285.00**
HR2-098	**PEG/Ion 2 Screen**	10 ml, tube format	**$285.00**
HR2-139	**PEG/Ion HT**	1 ml, Deep Well block format	**$185.00**
HR2-922-**	**PEG/Ion Screen Custom Shop**	185 ml	**$138.00**
HR2-998-**	**PEG/Ion 2 Screen Custom Shop**	185 ml	**$138.00**

** = reagent number 1-48

PEG/ION SCREEN™ FORMULATION

A1	1.	0.2 M Sodium fluoride, 20% w/v Polyethylene glycol 3,350
A2	2.	0.2 M Potassium fluoride, 20% w/v Polyethylene glycol 3,350
A3	3.	0.2 M Ammonium fluoride, 20% w/v Polyethylene glycol 3,350
A4	4.	0.2 M Lithium chloride, 20% w/v Polyethylene glycol 3,350
A5	5.	0.2 M Magnesium chloride hexahydrate, 20% w/v Polyethylene glycol 3,350
A6	6.	0.2 M Sodium chloride, 20% w/v Polyethylene glycol 3,350
A7	7.	0.2 M Calcium chloride dihydrate, 20% w/v Polyethylene glycol 3,350
A8	8.	0.2 M Potassium chloride, 20% w/v Polyethylene glycol 3,350
A9	9.	0.2 M Ammonium chloride, 20% w/v Polyethylene glycol 3,350
A10	10.	0.2 M Sodium iodide, 20% w/v Polyethylene glycol 3,350
A11	11.	0.2 M Potassium iodide, 20% w/v Polyethylene glycol 3,350
A12	12.	0.2 M Ammonium iodide, 20% w/v Polyethylene glycol 3,350
B1	13.	0.2 M Sodium thiocyanate, 20% w/v Polyethylene glycol 3,350
B2	14.	0.2 M Potassium thiocyanate, 20% w/v Polyethylene glycol 3,350
B3	15.	0.2 M Lithium nitrate, 20% w/v Polyethylene glycol 3,350
B4	16.	0.2 M Magnesium nitrate hexahydrate, 20% w/v Polyethylene glycol 3,350
B5	17.	0.2 M Sodium nitrate, 20% w/v Polyethylene glycol 3,350
B6	18.	0.2 M Potassium nitrate, 20% w/v Polyethylene glycol 3,350
B7	19.	0.2 M Ammonium nitrate, 20% w/v Polyethylene glycol 3,350
B8	20.	0.2 M Magnesium formate dihydrate, 20% w/v Polyethylene glycol 3,350
B9	21.	0.2 M Sodium formate, 20% w/v Polyethylene glycol 3,350
B10	22.	0.2 M Potassium formate, 20% w/v Polyethylene glycol 3,350
B11	23.	0.2 M Ammonium formate, 20% w/v Polyethylene glycol 3,350
B12	24.	0.2 M Lithium acetate dihydrate, 20% w/v Polyethylene glycol 3,350
C1	25.	0.2 M Magnesium acetate tetrahydrate, 20% w/v Polyethylene glycol 3,350
C2	26.	0.2 M Zinc acetate dihydrate, 20% w/v Polyethylene glycol 3,350
C3	27.	0.2 M Sodium acetate trihydrate, 20% w/v Polyethylene glycol 3,350
C4	28.	0.2 M Calcium acetate hydrate, 20% w/v Polyethylene glycol 3,350
C5	29.	0.2 M Potassium acetate, 20% w/v Polyethylene glycol 3,350
C6	30.	0.2 M Ammonium acetate, 20% w/v Polyethylene glycol 3,350
C7	31.	0.2 M Lithium sulfate monohydrate, 20% w/v Polyethylene glycol 3,350
C8	32.	0.2 M Magnesium sulfate heptahydrate, 20% w/v Polyethylene glycol 3,350
C9	33.	0.2 M Sodium sulfate decahydrate, 20% w/v Polyethylene glycol 3,350
C10	34.	0.2 M Potassium sulfate, 20% w/v Polyethylene glycol 3,350
C11	35.	0.2 M Ammonium sulfate, 20% w/v Polyethylene glycol 3,350
C12	36.	0.2 M Sodium tartrate dibasic dihydrate, 20% w/v Polyethylene glycol 3,350
D1	37.	0.2 M Potassium sodium tartrate tetrahydrate, 20% w/v Polyethylene glycol 3,350
D2	38.	0.2 M Ammonium tartrate dibasic, 20% w/v Polyethylene glycol 3,350
D3	39.	0.2 M Sodium phosphate monobasic monohydrate, 20% w/v Polyethylene glycol 3,350
D4	40.	0.2 M Sodium phosphate dibasic dihydrate, 20% w/v Polyethylene glycol 3,350
D5	41.	0.2 M Potassium phosphate monobasic, 20% w/v Polyethylene glycol 3,350
D6	42.	0.2 M Potassium phosphate dibasic, 20% w/v Polyethylene glycol 3,350
D7	43.	0.2 M Ammonium phosphate monobasic, 20% w/v Polyethylene glycol 3,350
D8	44.	0.2 M Ammonium phosphate dibasic, 20% w/v Polyethylene glycol 3,350
D9	45.	0.2 M Lithium citrate tribasic tetrahydrate, 20% w/v Polyethylene glycol 3,350
D10	46.	0.2 M Sodium citrate tribasic dihydrate, 20% w/v Polyethylene glycol 3,350
D11	47.	0.2 M Potassium citrate tribasic monohydrate, 20% w/v Polyethylene glycol 3,350
D12	48.	0.2 M Ammonium citrate dibasic, 20% w/v Polyethylene glycol 3,350

Acetate
Citrate
Chloride
Fluoride
Formate
Iodide
Nitrate
Phosphate
Sulfate
Thiocyanate
Tartrate

ANION

Range from 4.0 to 9.0

pH

PEG/ION FACTORS

POLYMER

Polyethylene glycol 3,350

CATION

Ammonium
Calcium
Lithium
Magnesium
Potassium
Sodium
Zinc

PEG/ION 2 SCREEN™ FORMULATION

E1	1.	0.1 M Sodium malonate pH 4.0, 12% w/v Polyethylene glycol 3,350
E2	2.	0.2 M Sodium malonate pH 4.0, 20% w/v Polyethylene glycol 3,350
E3	3.	0.1 M Sodium malonate pH 5.0, 12% w/v Polyethylene glycol 3,350
E4	4.	0.2 M Sodium malonate pH 5.0, 20% w/v Polyethylene glycol 3,350
E5	5.	0.1 M Sodium malonate pH 6.0, 12% w/v Polyethylene glycol 3,350
E6	6.	0.2 M Sodium malonate pH 6.0, 20% w/v Polyethylene glycol 3,350
E7	7.	0.1 M Sodium malonate pH 7.0, 12% w/v Polyethylene glycol 3,350
E8	8.	0.2 M Sodium malonate pH 7.0, 20% w/v Polyethylene glycol 3,350
E9	9.	4% v/v Tacsimate pH 4.0, 12% w/v Polyethylene glycol 3,350
E10	10.	8% v/v Tacsimate pH 4.0, 20% w/v Polyethylene glycol 3,350
E11	11.	4% v/v Tacsimate pH 5.0, 12% w/v Polyethylene glycol 3,350
E12	12.	8% v/v Tacsimate pH 5.0, 20% w/v Polyethylene glycol 3,350
F1	13.	4% v/v Tacsimate pH 6.0, 12% w/v Polyethylene glycol 3,350
F2	14.	8% v/v Tacsimate pH 6.0, 20% w/v Polyethylene glycol 3,350
F3	15.	4% v/v Tacsimate pH 7.0, 12% w/v Polyethylene glycol 3,350
F4	16.	8% v/v Tacsimate pH 7.0, 20% w/v Polyethylene glycol 3,350
F5	17.	4% v/v Tacsimate pH 8.0, 12% w/v Polyethylene glycol 3,350
F6	18.	8% v/v Tacsimate pH 8.0, 20% w/v Polyethylene glycol 3,350
F7	19.	0.1 M Succinic acid pH 7.0, 12% w/v Polyethylene glycol 3,350
F8	20.	0.2 M Succinic acid pH 7.0, 20% w/v Polyethylene glycol 3,350
F9	21.	0.1 M Ammonium citrate tribasic pH 7.0, 12% w/v Polyethylene glycol 3,350
F10	22.	0.2 M Ammonium citrate tribasic pH 7.0, 20% w/v Polyethylene glycol 3,350
F11	23.	0.1 M DL-Malic acid pH 7.0, 12% w/v Polyethylene glycol 3,350
F12	24.	0.2 M DL-Malic acid pH 7.0, 20% w/v Polyethylene glycol 3,350
G1	25.	0.1 M Sodium acetate trihydrate pH 7.0, 12% w/v Polyethylene glycol 3,350
G2	26.	0.2 M Sodium acetate trihydrate pH 7.0, 20% w/v Polyethylene glycol 3,350
G3	27.	0.1 M Sodium formate pH 7.0, 12% w/v Polyethylene glycol 3,350
G4	28.	0.2 M Sodium formate pH 7.0, 20% w/v Polyethylene glycol 3,350
G5	29.	0.1 M Ammonium tartrate dibasic pH 7.0, 12% w/v Polyethylene glycol 3,350
G6	30.	0.2 M Ammonium tartrate dibasic pH 7.0, 20% w/v Polyethylene glycol 3,350
G7	31.	2% v/v Tacsimate pH 4.0, 0.1 M Sodium acetate trihydrate pH 4.6, 16% w/v Polyethylene glycol 3,350
G8	32.	2% v/v Tacsimate pH 5.0, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 16% w/v Polyethylene glycol 3,350
G9	33.	2% v/v Tacsimate pH 6.0, 0.1 M BIS-TRIS pH 6.5, 20% w/v Polyethylene glycol 3,350
G10	34.	2% v/v Tacsimate pH 7.0, 0.1 M HEPES pH 7.5, 20% w/v Polyethylene glycol 3,350
G11	35.	2% v/v Tacsimate pH 8.0, 0.1 M Tris pH 8.5, 16% w/v Polyethylene glycol 3,350
G12	36.	(0.07 M Citric acid, 0.03 M BIS-TRIS propane / pH 3.4), 16% w/v Polyethylene glycol 3,350
H1	37.	(0.06 M Citric acid, 0.04 M BIS-TRIS propane / pH 4.1), 16% w/v Polyethylene glycol 3,350
H2	38.	(0.05 M Citric acid, 0.05 M BIS-TRIS propane / pH 5.0), 16% w/v Polyethylene glycol 3,350
H3	39.	(0.04 M Citric acid, 0.06 M BIS-TRIS propane / pH 6.4), 20% w/v Polyethylene glycol 3,350
H4	40.	(0.03 M Citric acid, 0.07 M BIS-TRIS propane / pH 7.6), 20% w/v Polyethylene glycol 3,350
H5	41.	(0.02 M Citric acid, 0.08 M BIS-TRIS propane / pH 8.8), 16% w/v Polyethylene glycol 3,350
H6	42.	0.02 M Calcium chloride dihydrate, 0.02 M Cadmium chloride hydrate, 0.02 M Cobalt(II) chloride hexahydrate, 20% w/v Polyethylene glycol 3,350
H7	43.	0.01 M Magnesium chloride hexahydrate, 0.005 M Nickel(II) chloride hexahydrate 0.1 M HEPES sodium pH 7.0, 15% w/v Polyethylene glycol 3,350
H8	44.	0.02 M Zinc chloride, 20% w/v Polyethylene glycol 3,350
H9	45.	0.15 M Cesium chloride, 15% w/v Polyethylene glycol 3,350
H10	46.	0.2 M Sodium bromide, 20% w/v Polyethylene glycol 3,350
H11	47.	1% w/v Tryptone, 0.05 M HEPES sodium pH 7.0, 12% w/v Polyethylene glycol 3,350
H12	48.	1% w/v Tryptone, 0.05 M HEPES sodium pH 7.0, 20% w/v Polyethylene glycol 3,350

Range from 4 to 8

pH

Neutralized Organic Salts
BIS-TRIS
Citric acid BIS-TRIS propane (CBTP)
HEPES
HEPES sodium
Sodium acetate
Sodium citrate
Tris

BUFFERS

Cadmium chloride
Calcium chloride
Cesium chloride
Cobalt(II) chloride
Magnesium chloride
Nickel(II) chloride
Sodium bromide
Zinc chloride

MULTIVALENT IONS

PEG/ION 2 FACTORS

POLYMER

Polyethylene glycol 3,350

ORGANIC SALTS

Ammonium tartrate
Ammonium citrate
Malic acid
Succinic acid
Sodium acetate
Sodium formate
Sodium malonate
Tacsimate™

PEPTIDES/ AMINO ACIDS

Tryptone

GRID SCREENS™ • QUIK SCREEN™

APPLICATION

- Primary or secondary, salt, polymer, organic and pH grid crystallization screen for biological macromolecules

FEATURES

- A systematic grid screen varying salt or polymer or organic versus pH
- Samples pH 4 to 9
- Combine reagents within or between screens to customize your screen
- Combine reagents between rows or columns to create expanded grid screens
- Tube or Deep Well block format

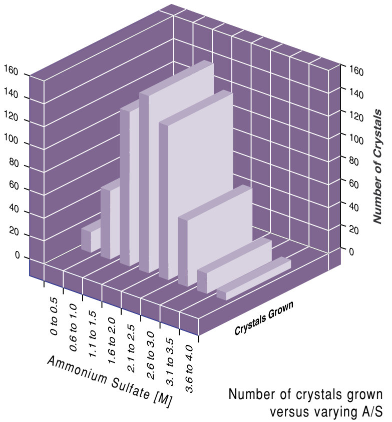

Number of crystals grown versus varying A/S concentrations [M]

DESCRIPTION

Grid Screen
HR2-211

Grid Screen Salt HT™ combines Grid Screen Ammonium Sulfate, Grid Screen Sodium Malonate, Quik Screen and Grid Screen Sodium Chloride into a single 96 Deep Well block format. The block contains 1 ml of each reagent.

Grid Screen Ammonium Sulfate systematically evaluates Ammonium sulfate at four concentrations (0.8, 1.6, 2.4, 3.2 M) versus 6 pH levels (4, 5, 6, 7, 8, 9).

Grid Screen Sodium Malonate evaluates Sodium malonate at six concentrations (1.0, 1.5, 1.9, 2.4, 2.9, 3.4 M) versus four pH levels (4, 5, 6, 7).

Quik Screen evaluates Sodium potassium phosphate at four concentrations (0.8, 1.0, 1.4, 1.8 M) versus six pH levels (5.0, 5.6, 6.3, 6.9, 7.5, 8.2).

Grid Screen Sodium Chloride evaluates Sodium chloride at four concentrations (1.0, 2.0, 3.0, 4.0 M) versus six pH levels (4, 5, 6, 7, 8, 9).

Grid Screen PEG 6000 evaluates Polyethylene glycol at four concentrations (5, 10, 20, 30 %w/v) versus six pH levels (4, 5, 6, 7, 8, 9).

Grid Screen MPD evaluates MPD at four concentrations (10, 20, 40, 65 %v/v) versus six pH levels (4, 5, 6, 7, 8, 9).

Grid Screen PEG/LiCl evaluates Polyethylene glycol at four concentrations (0, 10, 20, 30 %w/v) in the presence of 1.0 M Lithium chloride versus six pH levels (4, 5, 6, 7, 8, 9).

Grid Screens and Quik Screen each contain 10 ml of 24 unique reagents.

Ready-to-use reagents are sterile filtered and formulated with ultra-pure Type 1 water, using the highest purity salts, polymers, organics and buffers.

References

1. *A protein crystallization strategy using automated grid searches on successively finer grids. Patricia C. Weber. Methods: A Companion to Methods in Enzymology Vol. 1, No. 1, August, pp. 31-37, 1990.*
2. *Protein Crystallization; Techniques, Strategies, and Tips. A Laboratory Manual. Edited by Terese M. Bergfors. International University Line, 1999. ISBN 0-9636817-5-3.*

Order Information

Each Grid Screen Salt HT kit contains 96 unique reagents.

Cat. No.	Name	Description	Price
HR2-248	**Grid Screen Salt HT**	1 ml, Deep Well block format	**$185.00**

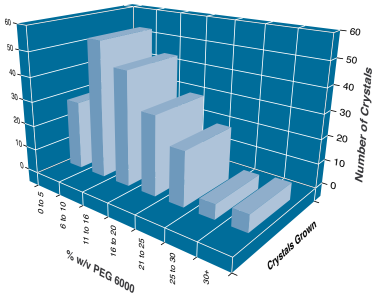

Number of crystals grown versus varying PEG 6000 concentrations (% w/v)

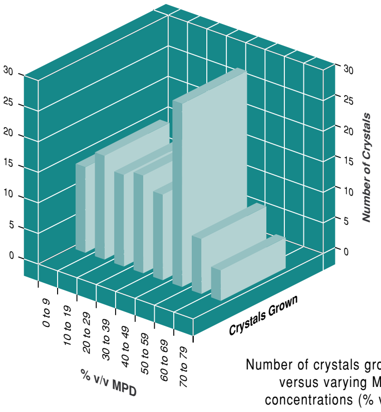

Number of crystals grown versus varying MPD concentrations (% v/v)

Order Information

Each Grid Screen and Quik Screen kit contains 24 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 36 for further details.

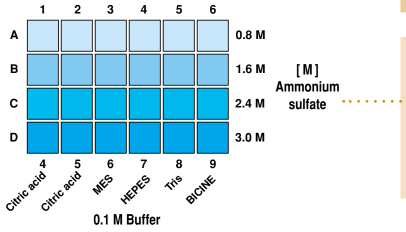
1
2
3
4
5
6
A
B
C
D
0.8 M
1.6 M
2.4 M
3.0 M
[M]
Ammonium
sulfate
4
5
6
7
8
9
Citric acid
Citric acid
MES
HEPES
Tris
BICINE
0.1 M Buffer

Cat. No.	Name	Description	Price
HR2-211	**Grid Screen Ammonium Sulfate**	10 ml, tube format	**$175.00**
HR2-924-**	**Grid Screen Ammonium Sulfate Custom Shop**	185 ml	**$138.00**

** = reagent number A1-D6

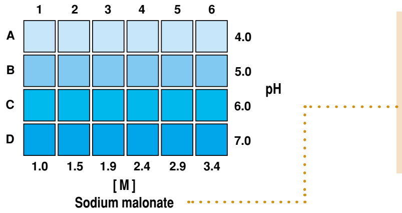
1
2
3
4
5
6
A
B
C
D
4.0
5.0
6.0
7.0
pH
1.0
1.5
1.9
2.4
2.9
3.4
[M]
Sodium malonate

Cat. No.	Name	Description	Price
HR2-247	**Grid Screen Sodium Malonate**	10 ml, tube format	**$175.00**
HR2-947-**	**Grid Screen Sodium Malonate Custom Shop**	185 ml	**$138.00**

** = reagent number A1-D6

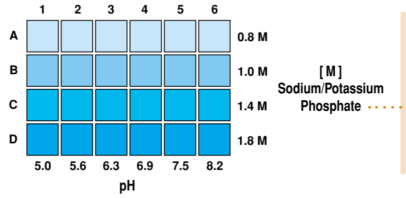
1
2
3
4
5
6
A
B
C
D
0.8 M
1.0 M
1.4 M
1.8 M
[M]
Sodium/Potassium
Phosphate
5.0
5.6
6.3
6.9
7.5
8.2
pH

Cat. No.	Name	Description	Price
HR2-221	**Quik Screen**	10 ml, tube format	**$175.00**
HR2-921-**	**Quik Screen Custom Shop**	185 ml	**$138.00**

** = reagent number A1-D6

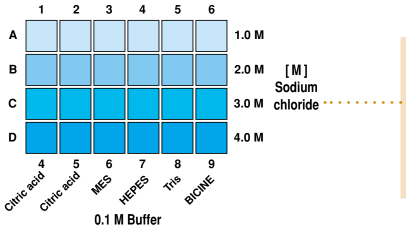
1
2
3
4
5
6
A
B
C
D
1.0 M
2.0 M
3.0 M
4.0 M
[M]
Sodium
chloride
4
5
6
7
8
9
Citric acid
Citric acid
MES
HEPES
Tris
BICINE
0.1 M Buffer

Cat. No.	Name	Description	Price
HR2-219	**Grid Screen Sodium Chloride**	10 ml, tube format	**$175.00**
HR2-932-**	**Grid Screen Sodium Chloride Custom Shop**	185 ml	**$138.00**

** = reagent number A1-D6

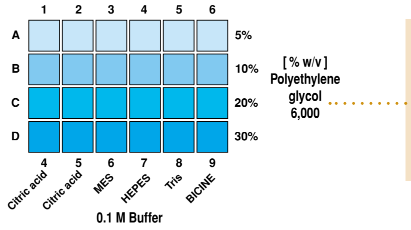
1
2
3
4
5
6
A
B
C
D
5%
10%
20%
30%
[% w/v]
Polyethylene
glycol
6,000
4
5
6
7
8
9
Citric acid
Citric acid
MES
HEPES
Tris
BICINE
0.1 M Buffer

Cat. No.	Name	Description	Price
HR2-213	**Grid Screen PEG 6000**	10 ml, tube format	**$175.00**
HR2-926-**	**Grid Screen PEG 6000 Custom Shop**	185 ml	**$138.00**

** = reagent number A1-D6

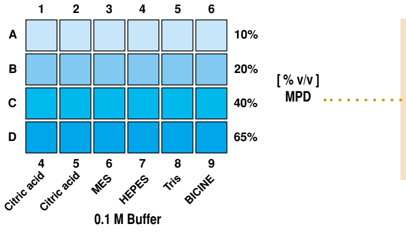
1
2
3
4
5
6
A
B
C
D
10%
20%
40%
65%
[% v/v]
MPD
4
5
6
7
8
9
Citric acid
Citric acid
MES
HEPES
Tris
BICINE
0.1 M Buffer

Cat. No.	Name	Description	Price
HR2-215	**Grid Screen MPD**	10 ml, tube format	**$175.00**
HR2-930-**	**Grid Screen MPD Custom Shop**	185 ml	**$138.00**

** = reagent number A1-D6

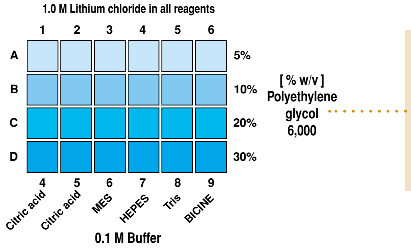
1.0 M Lithium chloride in all reagents
1
2
3
4
5
6
A
B
C
D
5%
10%
20%
30%
[% w/v]
Polyethylene
glycol
6,000
4
5
6
7
8
9
Citric acid
Citric acid
MES
HEPES
Tris
BICINE
0.1 M Buffer

Cat. No.	Name	Description	Price
HR2-217	**Grid Screen PEG/LiCl**	10 ml, tube format	**$175.00**
HR2-928-**	**Grid Screen PEG/LiCl Custom Shop**	185 ml	**$138.00**

** = reagent number A1-D6

SALTRX 1™ • SALTRX 2™ • SALTRX HT™

APPLICATION

- Primary or secondary, salt and pH matrix crystallization screen for biological macromolecules

FEATURES

- Developed at Hampton Research
- Salt versus pH matrix crystallization screen
- Samples pH 4 - 9
- 22 unique salts versus salt concentration and pH
- Compatible with microbatch, vapor diffusion, liquid & gel diffusion methods
- Preformulated, ready to screen
- All salts and buffers in screen readily available as Optimize reagents for reproducing and optimizing crystals
- Tube or Deep Well block format

SUCCESS STORY

Crystals grown using SaltRx, reagent 62. They were grown using the Cyberlab C240 robot and the Neuro Probe plate. These were grown just in time to freeze and take to the synchrotron (no time for optimizations). MAD data were collected on these crystals and the structure has been solved to 2.3Å resolution.

Courtesy of Annie Hassell.

GlaxoSmithKline

DESCRIPTION

SaltRx was developed by Hampton Research as a primary and secondary, salt and pH based crystallization screen for biological macromolecules. Salt is the only primary crystallization reagent (precipitant) utilized. Based on a design of 96 conditions, the screen evaluates a broad portfolio of crystallization salts of varying concentration and pH. The selection, concentration, and pH of the salts were determined by data mining the BMCD[1] and other crystallization databases, crystallization reports in the literature, and internal crystallization trials performed at Hampton Research. Based on this analysis, up to 35% of protein crystallizations involve salt as the primary crystallization reagent.

SaltRx can be used as a primary crystallization screen when salt, ionic strength and pH are desired or suspected as appropriate crystallization variables. It is also useful as a secondary screen when salt based reagents/conditions from screens such as Index, Crystal Screen, and Grid Screen produce crystals and when further screening for additional salt conditions or optimization is desired.

SaltRx is designed to be used as a 96 reagent screen.

SaltRx 1 contains 48 unique reagents, 10 ml each.

SaltRx 2 contains 48 unique reagents, 10 ml each.

SaltRx HT contains 1 ml of each reagent from SaltRx 1 and SaltRx 2 in a single Deep Well block format.

Ready-to-use reagents are sterile filtered and formulated with ultra-pure Type 1 water, using the highest purity salts, polymers, organics and buffers.

References

1. *Gilliland, G.L., Tung, M., Blakeslee, D.M. and Ladner, J. 1994. The Biological Macromolecule Crystallization Database, Version 3.0: New Features, Data, and the NASA Archive for Protein Crystal Growth Data. Acta Crystallogr. D50 408-413.*
2. *Crystallization and preliminary crystallographic analysis of molybdenumcofactor biosynthesis protein C from Thermus thermophilus. S. P. Kanaujia, C. V. Ranjani, J. Jeyakanthan, S. Baba, L. Chen, Z.-J. Liu, B.-C. Wang, M. Nishida, A. Ebihara, A. Shinkai, S. Kuramitsu, Y. Shiro, K. Sekar and S. Yokoyama. Acta Cryst. (2007). F63, 27-29.*

Order Information

SaltRx 1 and SaltRx 2 kits each contain 48 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 36 for further details.

Cat. No.	Name	Description	Price
HR2-107	**SaltRx 1**	10 ml, tube format	**$285.00**
HR2-109	**SaltRx 2**	10 ml, tube format	**$285.00**
HR2-136	**SaltRx HT**	1 ml, Deep Well block format	**$185.00**
HR2-907-**	**SaltRx 1 Custom Shop**	185 ml	**$138.00**
HR2-909-**	**SaltRx 2 Custom Shop**	185 ml	**$138.00**

** = reagent number 1-48

SALTRx 1™ FORMULATION

A1	1.	1.8 M Sodium acetate trihydrate pH 7.0, 0.1 M BIS-TRIS propane pH 7.0
A2	2.	2.8 M Sodium acetate trihydrate pH 7.0, 0.1 M BIS-TRIS propane pH 7.0
A3	3.	1.5 M Ammonium chloride, 0.1 M Sodium acetate trihydrate pH 4.6
A4	4.	1.5 M Ammonium chloride, 0.1 M BIS-TRIS propane pH 7.0
A5	5.	1.5 M Ammonium chloride, 0.1 M Tris pH 8.5
A6	6.	3.5 M Ammonium chloride, 0.1 M Sodium acetate trihydrate pH 4.6
A7	7.	3.5 M Ammonium chloride, 0.1 M BIS-TRIS propane pH 7.0
A8	8.	3.5 M Ammonium chloride, 0.1 M Tris pH 8.5
A9	9.	2.2 M Sodium chloride, 0.1 M Sodium acetate trihydrate pH 4.6
A10	10.	2.2 M Sodium chloride, 0.1 M BIS-TRIS propane pH 7.0
A11	11.	2.2 M Sodium chloride, 0.1 M Tris pH 8.5
A12	12.	3.2 M Sodium chloride, 0.1 M Sodium acetate trihydrate pH 4.6
B1	13.	3.2 M Sodium chloride, 0.1 M BIS-TRIS propane pH 7.0
B2	14.	3.2 M Sodium chloride, 0.1 M Tris pH 8.5
B3	15.	1.0 M Ammonium citrate dibasic, 0.1 M Sodium acetate trihydrate pH 4.6
B4	16.	1.8 M Ammonium citrate dibasic, 0.1 M Sodium acetate trihydrate pH 4.6
B5	17.	1.0 M Ammonium citrate tribasic pH 7.0, 0.1 M BIS-TRIS propane pH 7.0
B6	18.	2.0 M Ammonium citrate tribasic pH 7.0, 0.1 M BIS-TRIS propane pH 7.0
B7	19.	0.7 M Sodium citrate tribasic dihydrate, 0.1 M BIS-TRIS propane pH 7.0
B8	20.	0.7 M Sodium citrate tribasic dihydrate, 0.1 M Tris pH 8.5
B9	21.	1.2 M Sodium citrate tribasic dihydrate, 0.1 M BIS-TRIS propane pH 7.0
B10	22.	1.2 M Sodium citrate tribasic dihydrate, 0.1 M Tris pH 8.5
B11	23.	0.4 M Magnesium formate dihydrate, 0.1 M Sodium acetate trihydrate pH 4.6
B12	24.	0.4 M Magnesium formate dihydrate, 0.1 M BIS-TRIS propane pH 7.0
C1	25.	0.4 M Magnesium formate dihydrate, 0.1 M Tris pH 8.5
C2	26.	0.7 M Magnesium formate dihydrate, 0.1 M BIS-TRIS propane pH 7.0
C3	27.	2.0 M Sodium formate, 0.1 M Sodium acetate trihydrate pH 4.6
C4	28.	2.0 M Sodium formate, 0.1 M BIS-TRIS propane pH 7.0
C5	29.	2.0 M Sodium formate, 0.1 M Tris pH 8.5
C6	30.	3.5 M Sodium formate, 0.1 M Sodium acetate trihydrate pH 4.6
C7	31.	3.5 M Sodium formate, 0.1 M BIS-TRIS propane pH 7.0
C8	32.	3.5 M Sodium formate, 0.1 M Tris pH 8.5
C9	33.	1.2 M DL-Malic acid pH 7.0, 0.1 M BIS-TRIS propane pH 7.0
C10	34.	2.2 M DL-Malic acid pH 7.0, 0.1 M BIS-TRIS propane pH 7.0
C11	35.	1.4 M Sodium malonate pH 7.0, 0.1 M BIS-TRIS propane pH 7.0
C12	36.	2.4 M Sodium malonate pH 7.0, 0.1 M BIS-TRIS propane pH 7.0
D1	37.	2.5 M Ammonium nitrate, 0.1 M Sodium acetate trihydrate pH 4.6
D2	38.	2.5 M Ammonium nitrate, 0.1 M BIS-TRIS propane pH 7.0
D3	39.	2.5 M Ammonium nitrate, 0.1 M Tris pH 8.5
D4	40.	6.0 M Ammonium nitrate, 0.1 M Sodium acetate trihydrate pH 4.6
D5	41.	6.0 M Ammonium nitrate, 0.1 M BIS-TRIS propane pH 7.0
D6	42.	6.0 M Ammonium nitrate, 0.1 M Tris pH 8.5
D7	43.	1.5 M Sodium nitrate, 0.1 M Sodium acetate trihydrate pH 4.6
D8	44.	1.5 M Sodium nitrate, 0.1 M BIS-TRIS propane pH 7.0
D9	45.	1.5 M Sodium nitrate, 0.1 M Tris pH 8.5
D10	46.	4.0 M Sodium nitrate, 0.1 M Sodium acetate trihydrate pH 4.6
D11	47.	4.0 M Sodium nitrate, 0.1 M BIS-TRIS propane pH 7.0
D12	48.	4.0 M Sodium nitrate, 0.1 M Tris pH 8.5

SALTRX 1 FACTORS

Range from 4.6 to 8.5

PH

BIS-TRIS propane
Sodium acetate trihydrate
Tris

BUFFER

ORGANIC SALT

Ammonium citrate dibasic
Ammonium citrate tribasic
Magnesium formate dihydrate
Sodium acetate
Sodium citrate tribasic dihydrate
Sodium formate
Sodium malonate
DL-Malic acid

SALT

Ammonium chloride
Sodium chloride
Ammonium nitrate
Sodium nitrate

SALTRx 2™ FORMULATION

E1	1.	1.0 M Ammonium phosphate monobasic, 0.1 M Sodium acetate trihydrate pH 4.6
E2	2.	1.8 M Ammonium phosphate monobasic, 0.1 M Sodium acetate trihydrate pH 4.6
E3	3.	1.5 M Ammonium phosphate dibasic, 0.1 M Tris pH 8.5
E4	4.	2.4 M Ammonium phosphate dibasic, 0.1 M Tris pH 8.5
E5	5.	1.0 M Sodium phosphate monobasic monohydrate, Potassium phosphate dibasic / pH 5.0
E6	6.	1.0 M Sodium phosphate monobasic monohydrate, Potassium phosphate dibasic / pH 6.9
E7	7.	1.0 M Sodium phosphate monobasic monohydrate, Potassium phosphate dibasic / pH 8.2
E8	8.	1.8 M Sodium phosphate monobasic monohydrate, Potassium phosphate dibasic / pH 5.0
E9	9.	1.8 M Sodium phosphate monobasic monohydrate, Potassium phosphate dibasic / pH 6.9
E10	10.	1.8 M Sodium phosphate monobasic monohydrate, Potassium phosphate dibasic / pH 8.2
E11	11.	0.5 M Succinic acid pH 7.0, 0.1 M BIS-TRIS propane pH 7.0
E12	12.	1.0 M Succinic acid pH 7.0, 0.1 M BIS-TRIS propane pH 7.0
F1	13.	1.5 M Ammonium sulfate, 0.1 M Sodium acetate trihydrate pH 4.6
F2	14.	1.5 M Ammonium sulfate, 0.1 M BIS-TRIS propane pH 7.0
F3	15.	1.5 M Ammonium sulfate, 0.1 M Tris pH 8.5
F4	16.	2.5 M Ammonium sulfate, 0.1 M Sodium acetate trihydrate pH 4.6
F5	17.	2.5 M Ammonium sulfate, 0.1 M BIS-TRIS propane pH 7.0
F6	18.	2.5 M Ammonium sulfate, 0.1 M Tris pH 8.5
F7	19.	0.8 M Lithium sulfate monohydrate, 0.1 M Sodium acetate trihydrate pH 4.6
F8	20.	0.8 M Lithium sulfate monohydrate, 0.1 M BIS-TRIS propane pH 7.0
F9	21.	0.8 M Lithium sulfate monohydrate, 0.1 M Tris pH 8.5e
F10	22.	1.5 M Lithium sulfate monohydrate, 0.1 M Sodium acetate trihydrate pH 4.6
F11	23.	1.5 M Lithium sulfate monohydrate, 0.1 M BIS-TRIS propane pH 7.0
F12	24.	1.5 M Lithium sulfate monohydrate, 0.1 M Tris pH 8.5
G1	25.	1.0 M Magnesium sulfate hydrate, 0.1 M Sodium acetate trihydrate pH 4.6
G2	26.	1.0 M Magnesium sulfate hydrate, 0.1 M BIS-TRIS propane pH 7.0
G3	27.	1.0 M Magnesium sulfate hydrate, 0.1 M Tris pH 8.5
G4	28.	1.8 M Magnesium sulfate hydrate, 0.1 M Sodium acetate trihydrate pH 4.6
G5	29.	1.8 M Magnesium sulfate hydrate, 0.1 M BIS-TRIS propane pH 7.0
G6	30.	1.8 M Magnesium sulfate hydrate, 0.1 M Tris pH 8.5
G7	31.	0.7 M Ammonium tartrate dibasic, 0.1 M Sodium acetate trihydrate pH 4.6
G8	32.	0.7 M Ammonium tartrate dibasic, 0.1 M BIS-TRIS propane pH 7.0
G9	33.	0.7 M Ammonium tartrate dibasic, 0.1 M Tris pH 8.5
G10	34.	1.1 M Ammonium tartrate dibasic, 0.1 M Sodium acetate trihydrate pH 4.6
G11	35.	1.3 M Ammonium tartrate dibasic, 0.1 M BIS-TRIS propane pH 7.0
G12	36.	1.4 M Ammonium tartrate dibasic, 0.1 M Tris pH 8.5
H1	37.	0.6 M Potassium sodium tartrate tetrahydrate 0.1 M BIS-TRIS propane pH 7.0
H2	38.	1.2 M Potassium sodium tartrate tetrahydrate 0.1 M BIS-TRIS propane pH 7.0
H3	39.	0.6 M Potassium sodium tartrate tetrahydrate, 0.1 M Tris pH 8.5
H4	40.	1.2 M Potassium sodium tartrate tetrahydrate, 0.1 M Tris pH 8.5
H5	41.	0.5 M Potassium thiocyanate, 0.1 M Sodium acetate trihydrate pH 4.6
H6	42.	0.5 M Potassium thiocyanate, 0.1 M BIS-TRIS propane pH 7.0
H7	43.	0.5 M Potassium thiocyanate, 0.1 M Tris pH 8.5
H8	44.	4.0 M Ammonium acetate, 0.1 M Sodium acetate trihydrate pH 4.6
H9	45.	0.1 M BIS-TRIS propane pH 7.0, 4.0 M Ammonium Acetate
H10	46.	4.0 M Ammonium acetate, 0.1 M Tris pH 8.5
H11	47.	35% v/v Tacsimate™, 0.1 M BIS-TRIS propane pH 7.0
H12	48.	60% v/v Tacsimate™, 0.1 M BIS-TRIS propane pH 7.0

SALTRX 2 FACTORS

Range from 4.6 to 8.5

PH

BIS-TRIS propane
Sodium acetate trihydrate
Tris

BUFFER

ORGANIC SALT

Ammonium acetate
Ammonium tartrate dibasic
K/Na Tartrate
Succinic acid
Tacsimate™

SALT

Ammonium phosphate mono.
Ammonium phosphate dibasic
Ammonium sulfate
Lithium sulfate monohydrate
Magnesium sulfate hydrate
Na/K Phosphate
Potassium thiocyanate

MEMBFAC™ • CRYSTAL SCREEN LITE™ • MEMBFAC HT™

APPLICATION

- Primary sparse matrix crystallization screen for membrane proteins and samples with limited solubility

FEATURES

- Membrane protein sparse matrix screens[1]
- Sparse matrix formula efficiently samples salts, polymers, organics and pH
- pH range 4.6 - 8.5
- Formulated for use with detergents
- Crystal Screen Lite features a lower ionic strength than the original Crystal Screen
- Tube or Deep Well block format

SUCCESS STORY

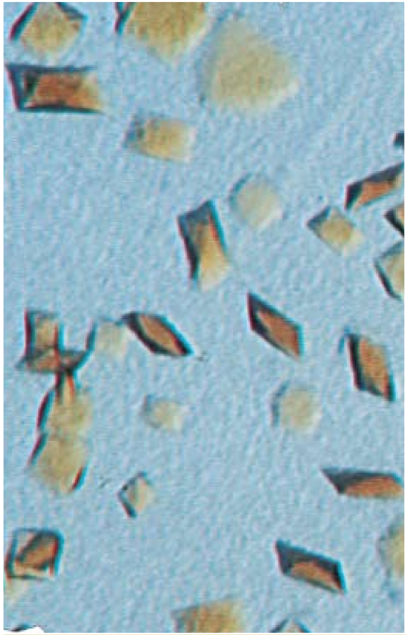

Crystals above are those of the membrane protein succinate:quinone oxireductase. Preliminary crystallization conditions determined using MembFac.

Courtesy of Michael Stowell.

MRC-LMB

DESCRIPTION

MembFac is a highly effective sparse matrix screen specifically designed as a preliminary screen for the crystallization ofmembrane proteins. MembFac is based upon the highly effective biased sparse matrix methodology. The MembFac screen matrix is optimized for the crystallization of membrane proteins.

MembFac contains 48 unique reagent formulations. With this set of 48 conditions, numerous chemicals and chemical combinations are utilized, allowing one to evaluate a large variety of potential crystallization conditions. MembFac covers a broad pH range between 4 and 9.

The Rationale

The basic rationale of MembFac is to perform a sparse matrix screen while changing the detergent dimension because detergents play an important factor in the crystallization of membrane proteins. A MembFac screen is to be completed for each detergent. Detergent screens are available separately.

The Method

The membrane protein of interest is first isolated in the detergent which gives the highest stability and activity. The protein concentration should be approximately 10 to 20 mg/ml and the detergent concentration should be only slightly above the CMC. The sample is then set against MembFac, using all 48 solutions plus the screening detergent(s) of choice. The screening detergent concentration should be approximately one to three times the CMC. Hence, for each detergent screened, one would utilize approximately 1 to 2 mg of sample.

The MembFac protocol was used to crystallize membrane proteins including SQR, fumerate: quinone oxireductase, LHII, and LHCII.

Crystal Screen Lite is a sparse matrix of trial crystallization reagent conditions based upon the Crystal Screen™.[3] The primary screen variables are salt, pH, and precipitant (salts, polymers, volatile organics, and non-volatile organics).[2] Crystal Screen Lite differs from the original Crystal Screen kit in that the primary precipitant reagents are one-half the concentration of that used in the original Crystal Screen formulation. The secondary salts, ions, and buffers remain at the original Crystal Screen concentration. Reducing the primary concentration of the primary precipitant results in a screen which is “more gentle” on the sample and typically produces much less precipitate conditions than the original Crystal Screen.

MembFac contains 48 unique reagents, 10 ml each.

MembFac HT contains 1 ml each of all 48 reagents from MembFac and reagents 1-48 from Crystal Screen Lite in a single Deep Well block format.

Ready-to-use reagents are sterile filtered and formulated with ultra-pure Type 1 water, using the highest purity salts, polymers, organics and buffers.

References

1. *UCLA Crystallization Workshop, June 21, 1993.*
2. *Crystallization of nucleic acids and proteins, Edited by A. Ducruix and R. Giege, The Practical Approach Series, Oxford Univ. Press, 1992.*
3. *McPherson, A., Current approaches to macromolecular crystallization., Eur. J. Biochem. (1990) 189, 1-23.*
4. *Jancarik, J. and Kim, S.H., Sparse Matrix Sampling: a screening method for crystallization of proteins. (1991) J. Appl. Cryst., 24,409-411.*
5. *Protein and Nucleic Acid Crystallization. Methods, A Companion to Methods in Enzymology, Academic Press, Volume 1, Number 1, August 1990.*

Order Information

Each MembFac kit contains 48 unique reagents. Crystal Screen Lite kit contains 50 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 36 for further details.

Cat. No.	Name	Description	Price
HR2-114	**MembFac**	10 ml, tube format	**$285.00**
HR2-128	**Crystal Screen Lite**	10 ml, tube format	**$285.00**
HR2-137	**MembFac HT**	1 ml, Deep Well block format	**$185.00**
HR2-920-**	**MembFac Custom Shop**	185 ml	**$138.00**
HR2-916-**	**Crystal Screen Lite Custom Shop**	185 ml	**$138.00**

** = reagent number 1-48 (for MembFac)

** = reagent number 1-50 (for Crystal Screen Lite)

MEMBFAC™ FORMULATION

A1	1.	0.1 M Sodium chloride, 0.1 M Sodium acetate trihydrate pH 4.6, 12% v/v (+/-)-2-Methyl-2,4-pentanediol
A2	2.	0.1 M Zinc acetate dihydrate, 0.1 M Sodium acetate trihydrate pH 4.6, 12% w/v Polyethylene glycol 4,000
A3	3.	0.2 M Ammonium sulfate, 0.1 M Sodium acetate trihydrate pH 4.6, 10% w/v Polyethylene glycol 4,000
A4	4.	0.1 M Sodium chloride, 0.1 M Sodium acetate trihydrate pH 4.6, 12% v/v 2-Propanol
A5	5.	0.1 M Sodium acetate trihydrate pH 4.6, 12% w/v Polyethylene glycol 4,000
A6	6.	0.1 M Sodium acetate trihydrate pH 4.6, 1.0 M Ammonium sulfate
A7	7.	0.1 M Sodium acetate trihydrate pH 4.6, 1.0 M Magnesium sulfate heptahydrate
A8	8.	0.1 M Magnesium chloride hexahydrate, 0.1 M Sodium acetate trihydrate pH 4.6, 18% v/v Polyethylene glycol 400
A9	9.	0.1 M Lithium sulfate monohydrate, 0.1 M Sodium acetate trihydrate pH 4.6, 1.0 M Ammonium phosphate monobasic
A10	10.	0.1 M Sodium chloride, 0.1 M Sodium acetate trihydrate pH 4.6, 12% w/v Polyethylene glycol 6,000
A11	11.	0.1 M Magnesium chloride hexahydrate, 0.1 M Sodium acetate trihydrate pH 4.6, 12% w/v Polyethylene glycol 6,000
A12	12.	0.1 M Sodium chloride, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 18% v/v Polyethylene glycol 400
B1	13.	0.1 M Lithium sulfate monohydrate, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 12 % w/v Polyethylene glycol 4,000
B2	14.	0.1 M Sodium citrate tribasic dihydrate, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 10% v/v 2-Propanol
B3	15.	0.1 M Sodium chloride, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 12% v/v (+/-)-2-Methyl-2,4-pentanediol
B4	16.	0.1 M Sodium citrate tribasic dihydrate pH 5.6, 1.0 M Magnesium sulfate heptahydrate
B5	17.	0.1 M Sodium chloride, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 12% w/v Polyethylene glycol 4,000
B6	18.	0.1 M Lithium sulfate monohydrate, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 12% w/v Polyethylene glycol 6,000
B7	19.	0.1 M Magnesium chloride hexahydrate, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 4% v/v (+/-)-2-Methyl-2,4-pentanediol
B8	20.	0.1 M Sodium citrate trihydrate dihydrate pH 5.6, 0.1 M Sodium chloride
B9	21.	0.1 M Lithium sulfate monohydrate, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 4% v/v Polyethylene glycol 400
B10	22.	0.1 M ADA pH 6.5, 1.0 M Ammonium sulfate
B11	23.	0.1 M Lithium sulfate monohydrate, 0.1 M ADA pH 6.5, 12% w/v Polyethylene glycol 4,000, 2% v/v 2-Propanol
B12	24.	0.1 M ADA pH 6.5, 1.0 M Ammonium phosphate dibasic
C1	25.	0.1 M Magnesium chloride hexahydrate, 0.1 M ADA pH 6.5, 12% w/v Polyethylene glycol 6,000
C2	26.	0.1 M ADA pH 6.5, 12% v/v (+/-)-2-Methyl-2,4-pentanediol
C3	27.	0.1 M Lithium sulfate monohydrate, 0.1 M ADA pH 6.5, 1.0 M Magnesium sulfate hydrate
C4	28.	0.3 M Lithium sulfate monohydrate, 0.1 M ADA pH 6.5, 4% v/v Polyethylene glycol 400
C5	29.	0.1 M Ammonium sulfate, 0.1 M HEPES sodium pH 7.5, 0.5 M Sodium phosphate dibasic dihydrate, 0.5 M Potassium phosphate dibasic
C6	30.	0.1 M Sodium chloride, 0.1 M HEPES sodium pH 7.5, 10% w/v Polyethylene glycol 4,000
C7	31.	0.1 M Magnesium chloride hexahydrate, 0.1 M HEPES sodium pH 7.5, 18% v/v Polyethylene glycol 400
C8	32.	0.1 M HEPES sodium pH 7.5, 1.0 M Potassium sodium tartrate tetrahydrate
C9	33.	0.1 M Ammonium sulfate, 0.1 M HEPES sodium pH 7.5, 18% v/v Polyethylene glycol 400
C10	34.	0.1 M Ammonium sulfate, 0.1 M HEPES sodium pH 7.5, 10% w/v Polyethylene glycol 4,000
C11	35.	0.1 M Sodium citrate tribasic dihydrate, 0.1 M HEPES sodium pH 7.5, 12% v/v (+/-)-2-Methyl-2,4-pentanediol
C12	36.	0.1 M HEPES sodium pH 7.5, 1.0 M Sodium citrate tribasic dihydrate
D1	37.	0.6 M Magnesium sulfate hydrate, 0.1 M HEPES sodium pH 7.5, 4% v/v Polyethylene glycol 400
D2	38.	0.6 M Magnesium sulfate hydrate, 0.1 M HEPES sodium pH 7.5, 4% v/v (+/-)-2-Methyl-2,4-pentanediol
D3	39.	0.1 M Lithium sulfate monohydrate, 0.1 M HEPES sodium pH 7.5, 0.1 M Potassium sodium tartrate tetrahydrate
D4	40.	0.1 M Lithium sulfate monohydrate, 0.1 M TRIS hydrochloride pH 8.5, 12% v/v (+/-)-2-Methyl-2,4-pentanediol
D5	41.	0.1 M Ammonium phosphate dibasic, 0.1 M TRIS hydrochloride pH 8.5, 0.5 M Sodium phosphate dibasic dihydrate, 0.5 M Potassium phosphate dibasic
D6	42.	0.1 M TRIS hydrochloride pH 8.5, 0.1 M Sodium acetate trihydrate
D7	43.	0.1 M TRIS hydrochloride pH 8.5, 0.1 M Sodium chloride
D8	44.	0.1 M Ammonium phosphate dibasic, 0.1 M TRIS hydrochloride pH 8.5, 12% w/v Polyethylene glycol 6,000
D9	45.	0.1 M Potassium sodium tartrate tetrahydrate, 0.1 M TRIS hydrochloride pH 8.5, 0.4 M Magnesium sulfate hydrate
D10	46.	0.1 M TRIS hydrochloride pH 8.5, 0.2 M Lithium sulfate monohydrate
D11	47.	0.1 M TRIS hydrochloride pH 8.5, 0.5 M Ammonium sulfate
D12	48.	0.1 M Sodium citrate tribasic dihydrate, 0.1 M TRIS hydrochloride pH 8.5, 5% v/v Polyethylene glycol 400

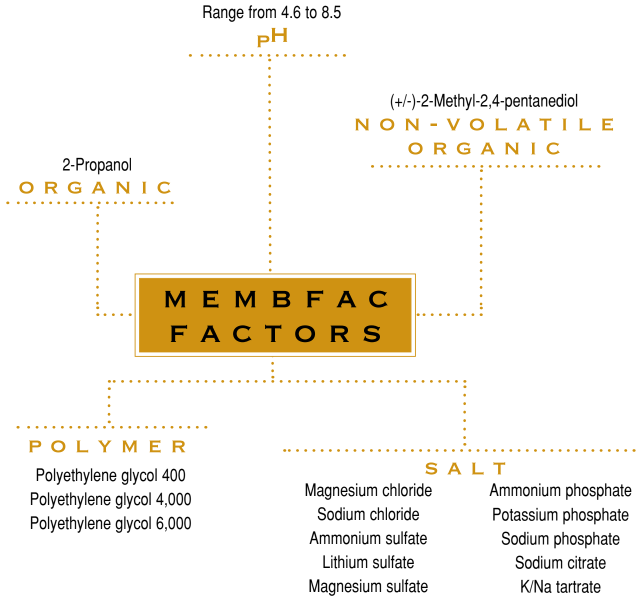
Range from 4.6 to 8.5
pH
(+/-)-2-Methyl-2,4-pentanediol
NON-VOLATILE ORGANIC
2-Propanol
ORGANIC
MEMBFAC FACTORS
POLYMER
Polyethylene glycol 400
Polyethylene glycol 4,000
Polyethylene glycol 6,000
SALT
Magnesium chloride
Sodium chloride
Ammonium sulfate
Lithium sulfate
Magnesium sulfate
Ammonium phosphate
Potassium phosphate
Sodium phosphate
Sodium citrate
K/Na tartrate

CRYSTAL SCREEN LITE™ FORMULATION

E1	1.	0.02 M Calcium chloride dihydrate, 0.1 M Sodium acetate trihydrate pH 4.6, 15% v/v (+/-)-2-Methyl-2,4-pentanediol
E2	2.	0.2 M Potassium sodium tartrate tetrahydrate
E3	3.	0.2 M Ammonium phosphate monobasic
E4	4.	0.1 M TRIS hydrochloride pH 8.5, 1.0 M Ammonium sulfate
E5	5.	0.2 M Sodium citrate tribasic dihydrate, 0.1 M HEPES sodium pH 7.5, 15% v/v (+/-)-2-Methyl-2,4-pentanediol
E6	6.	0.2 M Magnesium chloride hexahydrate, 0.1 M TRIS hydrochloride pH 8.5, 15% w/v Polyethylene glycol 4,000
E7	7.	0.1 M Sodium cacodylate trihydrate pH 6.5, 0.7 M Sodium acetate trihydrate
E8	8.	0.2 M Sodium citrate tribasic dihydrate, 0.1 M Sodium cacodylate trihydrate pH 6.5, 15% v/v 2-Propanol
E9	9.	0.2 M Ammonium acetate, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 15% w/v Polyethylene glycol 4,000
E10	10.	0.2 M Ammonium acetate, 0.1 M Sodium acetate trihydrate pH 4.6, 15% w/v Polyethylene glycol 4,000
E11	11.	0.1 M Sodium citrate tribasic dihydrate pH 5.6, 0.5 M Ammonium phosphate monobasic
E12	12.	0.2 M Magnesium chloride hexahydrate, 0.1 M HEPES sodium pH 7.5, 15% v/v 2-Propanol
F1	13.	0.2 M Sodium citrate tribasic dihydrate, 0.1 M TRIS hydrochloride pH 8.5, 15% v/v Polyethylene glycol 400
F2	14.	0.2 M Calcium chloride dihydrate, 0.1 M HEPES sodium pH 7.5, 14% v/v Polyethylene glycol 400
F3	15.	0.2 M Ammonium sulfate, 0.1 M Sodium cacodylate trihydrate pH 6.5, 15% w/v Polyethylene glycol 8,000
F4	16.	0.1 M HEPES sodium pH 7.5, 0.75 M Lithium sulfate monohydrate
F5	17.	0.2 M Lithium sulfate monohydrate, 0.1 M TRIS hydrochloride pH 8.5, 15% w/v Polyethylene glycol 4,000
F6	18.	0.2 M Magnesium acetate tetrahydrate, 0.1 M Sodium cacodylate trihydrate pH 6.5, 10% w/v Polyethylene glycol 8,000
F7	19.	0.2 M Ammonium acetate, 0.1 M TRIS hydrochloride pH 8.5, 15% v/v 2-Propanol
F8	20.	0.2 M Ammonium sulfate, 0.1 M Sodium acetate trihydrate pH 4.6, 12.5% w/v Polyethylene glycol 4,000
F9	21.	0.2 M Magnesium acetate tetrahydrate, 0.1 M Sodium cacodylate trihydrate pH 6.5, 15% v/v (+/-)-2-Methyl-2,4-pentanediol
F10	22.	0.2 M Sodium acetate trihydrate, 0.1 M TRIS hydrochloride pH 8.5, 15% w/v Polyethylene glycol 4,000
F11	23.	0.2 M Magnesium chloride hexahydrate, 0.1 M HEPES sodium pH 7.5, 15% v/v Polyethylene glycol 400
F12	24.	0.2 M Calcium chloride dihydrate, 0.1 M Sodium acetate trihydrate pH 4.6, 10% v/v 2-Propanol
G1	25.	0.1 M Imidazole pH 6.5, 0.5 M Sodium acetate trihydrate
G2	26.	0.2 M Ammonium acetate, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 15% v/v (+/-)-2-Methyl-2,4-pentanediol
G3	27.	0.2 M Sodium citrate tribasic dihydrate, 0.1 M HEPES sodium pH 7.5, 10% v/v 2-Propanol
G4	28.	0.2 M Sodium acetate trihydrate, 0.1 M Sodium cacodylate trihydrate pH 6.5, 15% w/v Polyethylene glycol 8,000
G5	29.	0.1 M HEPES sodium pH 7.5, 0.4 M Potassium sodium tartrate tetrahydrate
G6	30.	0.2 M Ammonium sulfate, 15% w/v Polyethylene glycol 8,000
G7	31.	0.2 M Ammonium sulfate, 15% w/v Polyethylene glycol 4,000
G8	32.	1.0 M Ammonium sulfate
G9	33.	2.0 M Sodium formate
G10	34.	0.1 M Sodium acetate trihydrate pH 4.6, 1.0 M Sodium formate
G11	35.	0.1 M HEPES sodium pH 7.5, 0.4 M Sodium phosphate monobasic monohydrate, 0.4 M Potassium phosphate monobasic
G12	36.	0.1 M TRIS hydrochloride pH 8.5, 4% w/v Polyethylene glycol 8,000
H1	37.	0.1 M Sodium acetate trihydrate pH 4.6, 4% w/v Polyethylene glycol 4,000
H2	38.	0.1 M HEPES sodium pH 7.5, 0.7 M Sodium citrate tribasic dihydrate
H3	39.	0.1 M HEPES sodium pH 7.5, 2% v/v Polyethylene glycol 400, 1.0 M Ammonium sulfate
H4	40.	0.1 M Sodium citrate tribasic dihydrate pH 5.6, 10% v/v 2-Propanol, 10% w/v Polyethylene glycol 4,000
H5	41.	0.1 M HEPES sodium pH 7.5, 5% v/v 2-Propanol, 10% w/v Polyethylene glycol 4,000
H6	42.	0.05 M Potassium phosphate monobasic, 10% w/v Polyethylene glycol 8,000
H7	43.	15% w/v Polyethylene glycol 1,500
H8	44.	0.1 M Magnesium formate dihydrate
H9	45.	0.2 M Zinc acetate dihydrate, 0.1 M Sodium cacodylate trihydrate pH 6.5, 9% w/v Polyethylene glycol 8,000
H10	46.	0.2 M Calcium acetate hydrate, 0.1 M Sodium cacodylate trihydrate pH 6.5, 9% w/v Polyethylene glycol 8,000
H11	47.	0.1 M Sodium acetate trihydrate pH 4.6, 1.0 M Ammonium sulfate
H12	48.	0.1 M TRIS hydrochloride pH 8.5, 1.0 M Ammonium phosphate monobasic
	49.	0.5 M Lithium sulfate monohydrate, 2% w/v Polyethylene glycol 8,000
	50.	0.5 M Lithium sulfate monohydrate, 7.5% w/v Polyethylene glycol 8,000

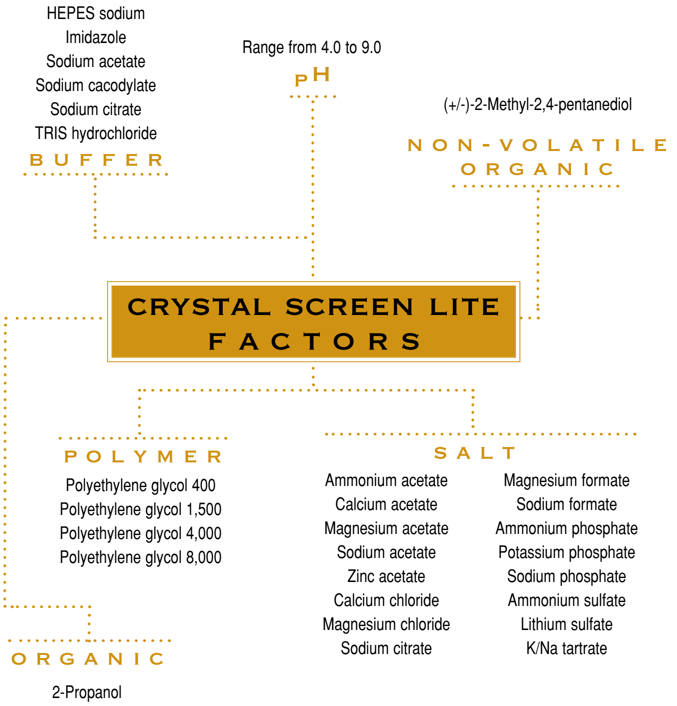
HEPES sodium
Imidazole
Sodium acetate
Sodium cacodylate
Sodium citrate
TRIS hydrochloride
BUFFER
Range from 4.0 to 9.0
pH
(+/-)-2-Methyl-2,4-pentanediol
NON-VOLATILE ORGANIC
CRYSTAL SCREEN LITE FACTORS
POLYMER
Polyethylene glycol 400
Polyethylene glycol 1,500
Polyethylene glycol 4,000
Polyethylene glycol 8,000
SALT
Ammonium acetate
Calcium acetate
Magnesium acetate
Sodium acetate
Zinc acetate
Calcium chloride
Magnesium chloride
Sodium citrate
Magnesium formate
Sodium formate
Ammonium phosphate
Potassium phosphate
Sodium phosphate
Ammonium sulfate
Lithium sulfate
K/Na tartrate
ORGANIC
2-Propanol

NATRIX™

APPLICATION

- Crystallization screen for nucleic acids & protein/nucleic acid complexes

FEATURES

- Nucleic acid sparse matrix screen[1]
- Sparse matrix formula efficiently samples salts, polyols, organics, & pH
- pH range 5.6 - 8.5
- Screen Magnesium, Potassium, & Ammonium anions

SUCCESS STORY

Crystal of MtaN, the N-terminal DNA binding domain of the Multidrug Transporter Activation protein from B. subtilis. Initial crystals grown from Natrix reagent number 29.

Courtesy of Michael Godsey.

Oregon Health Sciences University, Dept. of Biochemistry and Molecular Biology

DESCRIPTION

Natrix is based upon a published protocol which is highly effective for the crystallization of nucleic acids and protein-nucleic acid complexes. A variety of hammerhead ribozymes and other ribozymes, RNAs, DNAs, RNA-drug complexes, and RNA-protein complexes have been crystallized using the Natrix protocol.

By using sparse matrix sampling technology, Natrix allows one to quickly test wide ranges of pH, salts, and precipitants using a very small sample (50 to 100 μl) of nucleic acid.

Natrix is unique in that rather than relying solely on the traditional nucleic acid precipitant MPD, Natrix also utilizes Polyethylene glycols (PEGs) in a variety of molecular weights (200, 400, 4,000, 8,000) as well as 2-Propanol, Polyethylene glycol monomethyl ether (PEG MME), and 1,6-Hexanediol. Many of the polymeric and low molecular weight organic precipitants are combined with various monovalent salts as precipitating agents. This combination of salts and low molecular weight organics and polyalcohols, as well as the utilization of varying chain length PEGs, has proven to be a successful combination for producing nucleic acid and protein-nucleic acid complex crystals.

RNA Crystallization

Typical crystallization trials using Natrix will use the hanging or sitting drop methodology. Sample preparation is straightforward. The highly purified sample is heated to 75°C and annealed in the presence of the appropriate polyamine and buffer. After cooling to room temperature, the sample is mixed with an equal amount of Natrix solution and set against a reservoir containing the same Natrix solution for vapor diffusion equilibration. Typically, crystals are observed within the first week of setting up the trial.

Each Natrix contains 48 unique reagents of varying salt and precipitant composition over six different pH levels. Each preformulated, sterile filtered solution provides sufficient reagent for more than a dozen screens. The method is suitable for hanging drop, sitting drop, sandwich drop, and microbatch crystallization methodologies. Crystallization plates are sold separately. Natrix does not include polyamines. It is recommended that a polyamine such as Spermine be added to the sample prior to screening and that various polyamines be screened AFTER preliminary crystallization conditions have been determined.

Each Natrix kit contains 48 unique reagents, 10 ml each. Ready-to-use reagents are sterile filtered and formulated with ultra-pure Type 1 water, using the highest purity salts, polymers, organics, and buffers.

References

1. *Scott, W. G., Finch, J.T., Grenfell, R., Fogg, J., Smith, T., Gait, M.J., Klug, A., Journal of Molecular Biology (1995) 250: 327-332.*
2. *Purification, crystallization and preliminary X-ray analysis of the BseCI DNA methyltransferase from Bacillus stearothermophilus in complex with it's cognate DNA. E. G. Kapetaniou, D. Kotsifaki, M. Providaki, M. Rina, V. Bouriotis and M. Kokkinidis. Acta Cryst. (2007). F63, 12-14.*
3. *Crystallization and preliminary X-ray diffraction analysis of an Escherichia coli tRNAGly acceptor-stem microhelix. C. Förster, M. Perbandt, A. B. E. Brauer, S. Brode, J. P. Fürste, C. Betzel and V. A. Erdmann. Acta Cryst. (2007). F63, 46-48.*
4. *Combinatorial crystallization of an RNA-protein complex. Danielle Bodrero Hoggan, Jeffrey A. Chao, G. S. Prasad, C. David Stouta and James R. Williamson. Acta Cryst (2003). D59, 466-473*
5. *Cocrystallizing natural RNA with its unnatural mirror image: biochemical and preliminary X-ray diffraction analysis of a 5S rRNA A-helix racemate. Volker A. Erdmanna et al. Acta Cryst. (2007). F63, 839–843.*

Order Information

Each Natrix kit contains 48 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 36 for further details.

Cat. No.	Name	Description	Price
HR2-116	Natrix	10 ml, tube format	$285.00
HR2-918-**	Natrix Custom Shop	185 ml	$138.00

** = reagent number 1-48

NATRIX™ FORMULATION

1. 0.01 M Magnesium chloride hexahydrate, 0.05 M MES monohydrate pH 5.6, 1.8 M Lithium sulfate monohydrate
2. 0.01 M Magnesium acetate tetrahydrate, 0.05 M MES monohydrate pH 5.6, 2.5 M Ammonium sulfate
3. 0.1 M Magnesium acetate tetrahydrate, 0.05 M MES monohydrate pH 5.6, 20% v/v (+/-)-2-Methyl-2,4-pentanediol
4. 0.2 M Potassium chloride, 0.01 M Magnesium sulfate heptahydrate, 0.05 M MES monohydrate pH 5.6, 10% v/v Polyethylene glycol 400
5. 0.2 M Potassium chloride, 0.01 M Magnesium chloride hexahydrate, 0.05 M MES monohydrate pH 5.6, 5% w/v Polyethylene glycol 8,000
6. 0.1 M Ammonium sulfate, 0.01 M Magnesium chloride hexahydrate, 0.05 M MES monohydrate pH 5.6, 20% w/v Polyethylene glycol 8,000
7. 0.02 M Magnesium chloride hexahydrate, 0.05 M MES monohydrate pH 6.0, 15% v/v 2-Propanol
8. 0.1 M Ammonium acetate, 0.005 M Magnesium sulfate heptahydrate, 0.05 M MES monohydrate pH 6.0, 0.6 M Sodium chloride
9. 0.1 M Potassium chloride, 0.01 M Magnesium chloride hexahydrate, 0.05 M MES monohydrate pH 6.0, 10% v/v Polyethylene glycol 400
10. 0.005 M Magnesium sulfate heptahydrate, 0.05 M MES monohydrate pH 6.0, 5% w/v Polyethylene glycol 4,000
11. 0.01 M Magnesium chloride hexahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.0, 1.0 M Lithium sulfate monohydrate
12. 0.01 M Magnesium sulfate heptahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.0, 1.8 M Lithium sulfate monohydrate
13. 0.015 M Magnesium acetate tetrahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.0, 1.7 M Ammonium sulfate
14. 0.1 M Potassium chloride, 0.025 M Magnesium chloride hexahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.0, 15% v/v 2-Propanol
15. 0.04 M Magnesium chloride hexahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.0, 5% v/v (+/-)-2-Methyl-2,4-pentanediol
16. 0.04 M Magnesium acetate tetrahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.0, 30% v/v (+/-)-2-Methyl-2,4-pentanediol
17. 0.2 M Potassium chloride, 0.01 M Calcium chloride dihydrate, 0.05 M Sodium cacodylate trihydrate pH 6.0, 10% w/v Polyethylene glycol 4,000
18. 0.01 M Magnesium acetate tetrahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.5, 1.3 M Lithium sulfate monohydrate
19. 0.01 M Magnesium sulfate heptahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.5, 2.0 M Ammonium sulfate
20. 0.1 M Ammonium acetate, 0.015 M Magnesium acetate tetrahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.5, 10% v/v 2-Propanol
21. 0.2 M Potassium chloride, 0.005 M Magnesium chloride hexahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.5, 0.9 M 1,6-Hexanediol
22. 0.08 M Magnesium acetate tetrahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.5, 15% v/v Polyethylene glycol 400
23. 0.2 M Potassium chloride, 0.01 M Magnesium chloride hexahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.5, 10% w/v Polyethylene glycol 4,000
24. 0.2 M Ammonium acetate, 0.01 M Calcium chloride dihydrate, 0.05 M Sodium cacodylate trihydrate pH 6.5, 10% w/v Polyethylene glycol 4,000
25. 0.08 M Magnesium acetate tetrahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.5, 30% w/v Polyethylene glycol 4,000
26. 0.2 M Potassium chloride, 0.1 M Magnesium acetate tetrahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.5, 10 % w/v Polyethylene glycol 8,000
27. 0.2 M Ammonium acetate, 0.01 M Magnesium acetate tetrahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.5, 30% w/v Polyethylene glycol 8,000
28. 0.05 M Magnesium sulfate hydrate, 0.05 M HEPES sodium pH 7.0, 1.6 M Lithium sulfate monohydrate
29. 0.01 M Magnesium chloride hexahydrate, 0.05 M HEPES sodium pH 7.0, 4.0 M Lithium chloride
30. 0.01 M Magnesium chloride hexahydrate, 0.05 M HEPES sodium pH 7.0, 1.6 M Ammonium sulfate
31. 0.005 M Magnesium chloride hexahydrate, 0.05 M HEPES sodium pH 7.0, 25% v/v Polyethylene glycol monomethyl ether 550
32. 0.2 M Potassium chloride, 0.01 M Magnesium chloride hexahydrate, 0.05 M HEPES sodium pH 7.0, 1.7 M 1,6-Hexanediol
33. 0.2 M Ammonium chloride, 0.01 M Magnesium chloride hexahydrate, 0.05 M HEPES sodium pH 7.0, 2.5 M 1,6-Hexanediol
34. 0.1 M Potassium chloride, 0.005 M Magnesium sulfate hydrate, 0.05 M HEPES sodium pH 7.0, 15% v/v (+/-)-2-Methyl-2,4-pentanediol
35. 0.1 M Potassium chloride, 0.01 M Magnesium chloride hexahydrate, 0.05 M HEPES sodium pH 7.0, 5% v/v Polyethylene glycol 400
36. 0.1 M Potassium chloride, 0.01 M Calcium chloride dihydrate, 0.05 M HEPES sodium pH 7.0, 10% v/v Polyethylene glycol 400
37. 0.2 M Potassium chloride, 0.025 M Magnesium sulfate hydrate, 0.05 M HEPES sodium pH 7.0, 20% v/v Polyethylene glycol 200
38. 0.2 M Ammonium acetate, 0.15 M Magnesium acetate tetrahydrate, 0.05 M HEPES sodium pH 7.0, 5% w/v Polyethylene glycol 4,000
39. 0.1 M Ammonium acetate, 0.02 M Magnesium chloride hexahydrate, 0.05 M HEPES sodium pH 7.0, 5% w/v Polyethylene glycol 8,000
40. 0.01 M Magnesium chloride hexahydrate, 0.05 M TRIS hydrochloride pH 7.5, 1.6 M Ammonium sulfate
41. 0.1 M Potassium chloride, 0.015 M Magnesium chloride hexahydrate, 0.05 M TRIS hydrochloride pH 7.5 10% v/v Polyethylene glycol monomethyl ether 550
42. 0.01 M Magnesium chloride hexahydrate, 0.05 M TRIS hydrochloride pH 7.5, 5% v/v 2-Propanol
43. 0.05 M Ammonium acetate, 0.01 M Magnesium chloride hexahydrate, 0.05 M TRIS hydrochloride pH 7.5, 10% v/v (+/-)-2-Methyl-2,4-pentanediol
44. 0.2 M Potassium chloride, 0.05 M Magnesium chloride hexahydrate, 0.05 M TRIS hydrochloride pH 7.5, 10% w/v Polyethylene glycol 4,000
45. 0.025 M Magnesium sulfate hydrate, 0.05 M TRIS hydrochloride pH 8.5, 1.8 M Ammonium sulfate
46. 0.005 M Magnesium sulfate hydrate, 0.05 M TRIS hydrochloride pH 8.5, 2.9 M 1,6-Hexanediol
47. 0.1 M Potassium chloride, 0.01 M Magnesium chloride hexahydrate, 0.05 M TRIS hydrochloride pH 8.5, 30% v/v Polyethylene glycol 400
48. 0.2 M Ammonium chloride, 0.01 M Calcium chloride dihydrate, 0.05 M TRIS hydrochloride pH 8.5, 30% w/v Polyethylene glycol 4,000

2-Propanol

ORGANIC

Range from 5.6 to 8.5

pH

(+/-)-2-Methyl-2,4-pentanediol
1,6-Hexanediol

NON-VOLATILE ORGANIC

NATRIX FACTORS

POLYMER

Polyethylene glycol 200
Polyethylene glycol 400
Polyethylene glycol 4,000
Polyethylene glycol 8,000
Polyethylene glycol MME 550

SALT

Ammonium acetate	Lithium sulfate
Ammonium chloride	Magnesium chloride
Ammonium sulfate	Magnesium acetate
Calcium chloride	Magnesium sulfate
Lithium chloride	Potassium chloride

CRYSTAL SCREEN CRYO™

APPLICATION

- Crystallization screen for proteins, peptides, & nucleic acids

FEATURES

- Cryo ready reagents for cryocrystallography[1]
- Sparse matrix formula efficiently samples salts, polymers, organics, & pH

METHODOLOGIES

- Screen for cryo and crystallization conditions using only one screen

SUCCESS STORY

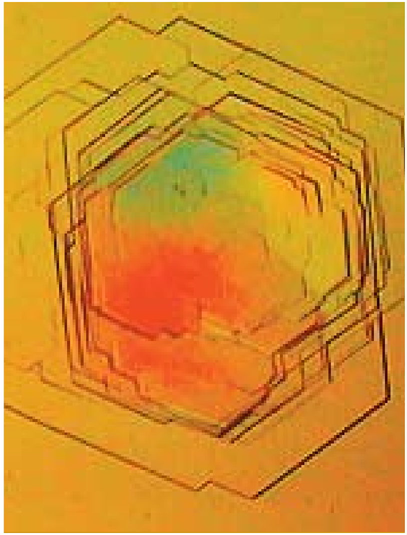

The protein crystal above is a LFA-1 I-domain/Inhibitor complex. Crystallization conditions obtained using Crystal Screen Cryo from Hampton Research.

Courtesy of Joerg Kallen.

Novartis Pharma AG

DESCRIPTION

The proportion of macromolecular x-ray structures solved using cryocrystallographic methods is increasing at an exponential rate. The advantages of macromolecular cryocrystallography are numerous. The most widely used method of flash cooling macromolecular crystals is to support the crystal in a film of cryoprotected mother liquor on or in a small fiber loop which is subsequently cooled in liquid nitrogen at around 100°K. The liquid about the crystal must freeze as an amorphous glass to avoid crystal damage and diffraction from ordered ice. Cryoprotection is achieved by mixing the mother liquor and crystal with a suitable cryoprotectant reagent. The most frequently utilized cryoprotectant is glycerol.

Determining a Cryoprotectant

Determining initial and optimal cryoprotectant concentration is often a process of trial and error. One must find suitable cryoprotectant concentrations which stabilize the crystal while at the same time combine with the crystallization reagent to form an amorphous glass.

Crystal Screen Cryo removes the guesswork from determining the preliminary glycerol concentration required to mix with the crystallization reagent to form an amorphous glass. Crystal Screen Cryo reagents are preformulated with the appropriate amount of glycerol required to form an amorphous glass with each unique reagent composition. Crystals obtained in Crystal Screen Cryo will already have a suitable concentration of glycerol in the reagent. This avoids exhaustive empirical searches for preliminary cryoprotectant conditions. From the preliminary cryoprotectant conditions, cryoprotectant concentration can be optimized to minimize mosaic spread and maximize diffraction resolution.

Crystal Screen Cryo is a complete sparse matrix reagent kit designed to provide a rapid screening method for the crystallization of biological macromolecules in the presence of glycerol, the most frequently utilized cryoprotectant. Crystal Screen Cryo utilizes the original Crystal Screen™ protocol but is optimized to include the appropriate concentration of glycerol required to form an amorphous glass at 100°K. The primary screen variables are salt, pH, precipitant (salts, polymers, volatile organics, and non-volatile organics), and cryoprotectant concentration. The screen is a straightforward, effective, and practical kit for determining preliminary crystallization conditions and provides a good starting point for finding suitable cryoprotectant conditions for macromolecular crystals grown in a wide range of reagents. Crystal Screen Cryo is also effective in determining the solubility of a macromolecule in a wide range of precipitants and pH.

Each Crystal Screen Cryo kit contains 50 unique reagents, 10 ml each. Ready-to-use reagents are sterile filtered and formulated with ultra-pure Type 1 water, using the highest purity salts, polymers, organics, and buffers.

References

1. *Garman, E.F. and Mitchell, E.P., J. Appl. Cryst. (1996) 29, 584- 587.*

Order Information

Each Crystal Screen Cryo kit contains 50 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 36 for further details.

Cat. No.	Name	Description	Price
HR2-122	**Crystal Screen Cryo**	10 ml, tube format	**$285.00**
HR2-914-**	**Crystal Screen Cryo Custom Shop**	185 ml	**$138.00**

** = reagent number 1-50

CRYSTAL SCREEN CRYO™ FORMULATION

1. 0.02 M Calcium chloride dihydrate, 0.1 M Sodium acetate trihydrate pH 4.6, 30% v/v (+/-)-2-Methyl-2,4-pentanediol
2. 0.26 M Potassium sodium tartrate tetrahydrate, 35% v/v Glycerol
3. 0.26 M Ammonium phosphate monobasic, 35% v/v Glycerol
4. 0.075 M Tris hydrochloride pH 8.5, 1.5 M Ammonium sulfate, 25% v/v Glycerol
5. 0.2 M Sodium citrate tribasic dihydrate, 0.1 M HEPES sodium pH 7.5, 30% v/v (+/-)-2-Methyl-2,4-pentanediol
6. 0.16 M Magnesium chloride hexahydrate, 0.08 M Tris hydrochloride pH 8.5, 24% w/v Polyethylene glycol 4,000, 20% v/v Glycerol
7. 0.07 M Sodium cacodylate trihydrate pH 6.5, 0.98 M Sodium acetate trihydrate, 30% v/v Glycerol
8. 0.14 M Sodium citrate tribasic dihydrate, 0.07 M Sodium cacodylate trihydrate pH 6.5, 21% v/v 2-Propanol, 30% v/v Glycerol
9. 0.17 M Ammonium acetate, 0.085 M Sodium citrate tribasic dihydrate pH 5.6, 25.5% w/v Polyethylene glycol 4,000, 15% v/v Glycerol
10. 0.17 M Ammonium acetate, 0.085 M Sodium acetate trihydrate pH 4.6, 25.5% w/v Polyethylene glycol 4,000, 15% v/v Glycerol
11. 0.07 M Sodium citrate tribasic dihydrate pH 5.6, 0.7 M Ammonium phosphate monobasic, 30% v/v Glycerol
12. 0.18 M Magnesium chloride hexahydrate, 0.09 M HEPES sodium pH 7.5, 27% v/v 2-Propanol, 10% v/v Glycerol
13. 0.2 M Sodium citrate tribasic dihydrate, 0.1 M Tris hydrochloride pH 8.5, 30% v/v Polyethylene glycol 400
14. 0.19 M Calcium chloride dihydrate, 0.095 M HEPES sodium pH 7.5, 26.6% v/v Polyethylene glycol 400, 5% v/v Glycerol
15. 0.17 M Ammonium sulfate, 0.085 M Sodium cacodylate trihydrate pH 6.5, 25.5% w/v Polyethylene glycol 8,000, 15% v/v Glycerol
16. 0.075 M HEPES sodium pH 7.5, 1.125 M Lithium sulfate monohydrate, 25% v/v Glycerol
17. 0.17 M Lithium sulfate monohydrate, 0.085 M Tris hydrochloride pH 8.5, 25.5% w/v Polyethylene glycol 4,000, 15% v/v Glycerol
18. 0.16 M Magnesium acetate tetrahydrate, 0.08 M Sodium cacodylate trihydrate pH 6.5, 16% w/v Polyethylene glycol 8,000, 20% v/v Glycerol
19. 0.16 M Ammonium acetate, 0.08 M Tris hydrochloride pH 8.5, 24% v/v 2-Propanol, 20% v/v Glycerol
20. 0.16 M Ammonium sulfate, 0.08 M Sodium acetate trihydrate pH 4.6, 20% w/v Polyethylene glycol 4,000, 20% v/v Glycerol
21. 0.2 M Magnesium acetate tetrahydrate, 0.1 M Sodium cacodylate trihydrate pH 6.5, 30% v/v (+/-)-2-Methyl-2,4-pentanediol
22. 0.17 M Sodium acetate trihydrate, 0.085 M Tris hydrochloride pH 8.5, 25.5% w/v Polyethylene glycol 4,000, 15% v/v Glycerol
23. 0.2 M Magnesium chloride hexahydrate, 0.1 M HEPES sodium pH 7.5, 30% v/v Polyethylene glycol 400
24. 0.14 M Calcium chloride dihydrate, 0.07 M Sodium acetate trihydrate pH 4.6, 14% v/v 2-Propanol, 30% v/v Glycerol
25. 0.07 M Imidazole pH 6.5, 0.7 M Sodium acetate trihydrate, 30% v/v Glycerol
26. 0.2 M Ammonium acetate, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 30% v/v (+/-)-2-Methyl-2,4-pentanediol
27. 0.14 M Sodium citrate tribasic dihydrate, 0.07 M HEPES sodium pH 7.5, 14% v/v 2-Propanol, 30% v/v Glycerol
28. 0.17 M Sodium acetate trihydrate, 0.085 M Sodium cacodylate trihydrate pH 6.5, 25.5% w/v Polyethylene glycol 8,000, 15% v/v Glycerol
29. 0.065 M HEPES sodium pH 7.5, 0.52 M Potassium sodium tartrate tetrahydrate, 35% v/v Glycerol
30. 0.17 M Ammonium sulfate, 25.5% w/v Polyethylene glycol 8,000, 15% v/v Glycerol
31. 0.17 M Ammonium sulfate, 25.5% w/v Polyethylene glycol 4,000, 15% v/v Glycerol
32. 1.5 M Ammonium sulfate, 25% v/v Glycerol
33. 3.6 M Sodium formate, 10% v/v Glycerol
34. 0.07 M Sodium acetate trihydrate pH 4.6, 1.4 M Sodium formate, 30% v/v Glycerol
35. 0.075 M HEPES sodium pH 7.5, 0.6 M Sodium phosphate monobasic monohydrate, 0.6 M Potassium phosphate monobasic, 25% v/v Glycerol
36. 0.065 M Tris hydrochloride pH 8.5, 5.2% w/v Polyethylene glycol 8,000, 35% v/v Glycerol
37. 0.07 M Sodium acetate trihydrate pH 4.6, 5.6% w/v Polyethylene glycol 4,000, 30% v/v Glycerol
38. 0.09 M HEPES sodium pH 7.5, 1.26 M Sodium citrate tribasic dihydrate, 10% v/v Glycerol
39. 0.085 M HEPES sodium pH 7.5, 1.7% v/v Polyethylene glycol 400, 1.7 M Ammonium sulfate, 15% v/v Glycerol
40. 0.095 M Sodium citrate tribasic dihydrate pH 5.6, 19% v/v 2-Propanol, 19% w/v Polyethylene glycol 4,000, 5% v/v Glycerol
41. 0.085 M HEPES sodium pH 7.5, 8.5% v/v 2-Propanol, 17% w/v Polyethylene glycol 4,000, 15% v/v Glycerol
42. 0.04 M Potassium phosphate monobasic, 16% w/v Polyethylene glycol 8,000, 20% v/v Glycerol
43. 24% w/v Polyethylene glycol 1,500, 20% v/v Glycerol
44. 0.1 M Magnesium formate dihydrate, 50% v/v Glycerol
45. 0.16 M Zinc acetate dihydrate, 0.08 M Sodium cacodylate trihydrate pH 6.5, 14.4% w/v Polyethylene glycol 8,000, 20% v/v Glycerol
46. 0.16 M Calcium acetate hydrate, 0.08 M Sodium cacodylate trihydrate pH 6.5, 14.4% w/v Polyethylene glycol 8,000, 20% v/v Glycerol
47. 0.08 M Sodium acetate trihydrate pH 4.6, 1.6 M Ammonium sulfate, 20% v/v Glycerol
48. 0.08 M Tris hydrochloride pH 8.5, 1.6 M Ammonium phosphate monobasic, 20% v/v Glycerol
49. 0.8 M Lithium sulfate monohydrate, 1.6% w/v Polyethylene glycol 8,000, 20% v/v Glycerol
50. 0.4 M Lithium sulfate monohydrate, 12% w/v Polyethylene glycol 8,000, 20% v/v Glycerol

CRYSTAL SCREEN CRYO FACTORS

Glycerol
CRYOPROCTECANT

2-Propanol
ORGANIC

Range from 4.0 to 9.0
pH

(+/-)-2-Methyl-2,4-pentanediol
NON-VOLATILE ORGANIC

POLYMER

Polyethylene glycol 400
Polyethylene glycol 1,500
Polyethylene glycol 4,000
Polyethylene glycol 8,000

SALT

Ammonium acetate
Ammonium phosphate
Ammonium sulfate
Calcium acetate
Calcium chloride
Lithium sulfate
Magnesium acetate
Magnesium chloride
Magnesium formate
Potassium phosphate
K/Na tartrate
Sodium acetate
Sodium citrate
Sodium formate
Sodium phosphate
Zinc acetate

BUFFER

HEPES sodium
Imidazole
Sodium acetate
Sodium cacodylate
Sodium citrate
TRIS hydrochloride

NUCLEIC ACID MINI SCREEN™

APPLICATION

- Crystallization screen for nucleic acid fragments

FEATURES

- Screen a matrix of pH, polyamine, & ions
- Can be used with ribozymes, pseudoknots, RNA, & DNA
- Uses MPD as the primary precipitant

SUCCESS STORY

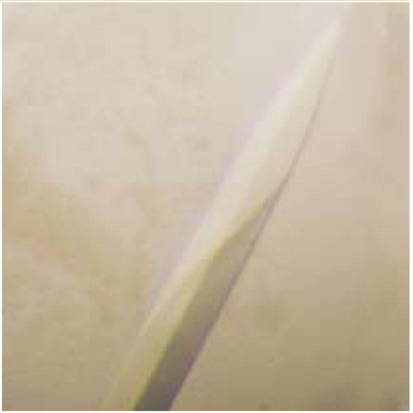

Crystal of RNA tetraloop.

Courtesy of Imre Berger and Li Su. MIT

DESCRIPTION

NUCLEIC ACID MINI SCREEN™
Nucleic Acid Crystallization Kit
HAMPTON RESEARCH

The Nucleic Acid Mini Screen is an efficient screen formulated to assist in the determination of preliminary crystallization conditions of nucleic acid fragments.[1] The formulation is based upon the publication, "A Highly Efficient 24 Condition Matrix for the Crystallization of Nucleic Acid Fragments" where the preliminary crystallization conditions of 35 nucleic acids were determined utilizing this formulation.[1] Samples include DNA-Drug complexes, C-Tetrad and G-Quartet Motifs, RNA oligomers, and others. By using 1 to 4 mM oligonucleotide stock concentration, the screen requires less than 100 µl of sample. The screen is typically performed at 4°C although room temperature incubations can also be performed. To evaluate the effect of equilibration kinetics as well as initial and final sample concentrations, the screens are typically performed using the hanging or sitting drop vapor diffusion method with two drops on a slide (1 µl + 2 µl and 2 µl + 2 µl), side by side. The 24 well format of the mini screen provides for a fast setup and uses small amounts of sample. This makes it possible to cost-effectively screen many sequences as well as variations of a particular sequence. The composition of the Nucleic Acid Mini Screen allows one to apply the formulation to other nucleic acids such as deoxy- and ribozymes, pseudoknots, and tRNAs.

Each Nucleic Acid Mini Screen kit contains 24 unique reagents, 1.0 ml each plus a 250 ml volume of dehydrant (35% v/v MPD). All solutions are sterile filtered and formulated with ultra-pure Type 1 water.

References

1. *Berger, et al., A Highly Effective 24 Condition Matrix for the Crystallization of Nucleic Acid Fragments. Acta Cryst. Section D. (1996) Vol. D52 Part 3, 465-468.*
2. *Adams, A., Nucleic Acids Research (2002) Vol. 30, 719-725.*
3. *Crystallization and preliminary X-ray diffraction analysis of an Escherichia coli tRNAGly acceptor-stem microhelix. C. Förster, M. Perbandt, A. B. E. Brauer, S. Brode, J. P. Fürste, C. Betzel and V. A. Erdmann. Acta Cryst. (2007). F63, 46-48.*

Order Information

Each Nucleic Acid Mini Screen kit contains 24 unique reagents plus dehydrant.

Cat. No.	Name	Description	Price
HR2-118	**Nucleic Acid Mini Screen**	1 ml, tube format + 250 ml bottle	**$195.00**

NUCLEIC ACID MINI SCREEN FORMULATION

Tube #	Precipitant	Buffer	Polyamine	Monovalent Ion	Divalent Ion
1.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 5.5	20 mM Cobalt hexamine	None	20 mM Magnesium chloride
2.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 5.5	20 mM Cobalt hexamine	80 mM Sodium chloride	20 mM Magnesium chloride
3.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 5.5	20 mM Cobalt hexamine	12 mM Sodium chloride, 80 mM Potassium chloride	None
4.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 5.5	20 mM Cobalt hexamine	40 mM Lithium chloride	20 mM Magnesium chloride
5.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 6.0	12 mM Spermine tetrahydrochloride	80 mM Potassium chloride	20 mM Magnesium chloride
6.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 6.0	12 mM Spermine tetrahydrochloride	80 mM Potassium chloride	None
7.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 6.0	12 mM Spermine tetrahydrochloride	80 mM Sodium chloride	20 mM Magnesium chloride
8.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 6.0	12 mM Spermine tetrahydrochloride	80 mM Sodium chloride	None
9.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 6.0	12 mM Spermine tetrahydrochloride	80 mM Sodium chloride, 12 mM Potassium chloride	20 mM Magnesium chloride
10.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 6.0	12 mM Spermine tetrahydrochloride	12 mM Sodium chloride, 80 mM Potassium chloride	None
11.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 6.0	12 mM Spermine tetrahydrochloride	80 mM Sodium chloride	20 mM Barium chloride
12.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 6.0	12 mM Spermine tetrahydrochloride	80 mM Potassium chloride	20 mM Barium chloride
13.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 6.0	12 mM Spermine tetrahydrochloride	None	80 mM Strontium chloride
14.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 7.0	12 mM Spermine tetrahydrochloride	80 mM Potassium chloride	20 mM Magnesium chloride
15.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 7.0	12 mM Spermine tetrahydrochloride	80 mM Potassium chloride	None
16.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 7.0	12 mM Spermine tetrahydrochloride	80 mM Sodium chloride	20 mM Magnesium chloride
17.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 7.0	12 mM Spermine tetrahydrochloride	80 mM Sodium chloride	None
18.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 7.0	12 mM Spermine tetrahydrochloride	80 mM Sodium chloride, 12 mM Potassium chloride	20 mM Magnesium chloride
19.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 7.0	12 mM Spermine tetrahydrochloride	12 mM Sodium chloride, 80 mM Potassium chloride	None
20.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 7.0	12 mM Spermine tetrahydrochloride	80 mM Sodium chloride	20 mM Barium chloride
21.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 7.0	12 mM Spermine tetrahydrochloride	80 mM Potassium chloride	20 mM Barium chloride
22.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 7.0	12 mM Spermine tetrahydrochloride	40 mM Lithium chloride	80 mM Strontium chloride, 20 mM Magnesium chloride
23.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 7.0	12 mM Spermine tetrahydrochloride	40 mM Lithium chloride	80 mM Strontium chloride
24.	10% v/v (+/-)-2-Methyl-2,4-pentanediol	40 mM Sodium cacodylate pH 7.0	12 mM Spermine tetrahydrochloride	None	80 mM Strontium chloride, 20 mM Magnesium chloride

polyamines

Cobalt hexamine

Spermine tetra-HCl

chlorides

+ Lithium
Sodium
Potassium
Magnesium +2
Strontium
Barium

dehydrant

35% v/v MPD

pH

5.5 to 7.0

using

40 mM Sodium cacodylate

LOW IONIC STRENGTH SCREEN™

APPLICATION

- Crystallization screen for intact monoclonal antibodies, monoclonal antibody fragments, & proteins less soluble at low ionic strength

FEATURES

- Screen a complete pH profile (2-12) at low ionic strength
- Enhanced temperature effects due to low ionic strength
- Monodisperse PEG 3350

Low Ionic Strength Screen Components

Buffer

0.05 M Potassium chloride pH 2.0
0.05 M Citric acid pH 3.0
0.05 M Citric acid pH 3.5
0.05 M Citric acid pH 4.0
0.05 M Citric acid pH 4.5
0.05 M Citric acid pH 5.0
0.05 M Citric acid pH 5.5
0.05 M MES monohydrate pH 6.0
0.05 M BIS-TRIS pH 6.5
0.05 M Imidazole pH 7.0
0.05 M HEPES pH 7.5
0.05 M Tris pH 8.0
0.05 M Tris pH 8.5
0.05 M Glycine pH 9.0
0.05 M Glycine pH 9.5
0.05 M Glycine pH 10.0
0.05 M Sodium phosphate dibasic pH 11.0
0.05 M Sodium phosphate dibasic pH 12.0

Precipitant

4% w/v Polyethylene glycol 3,350
8% w/v Polyethylene glycol 3,350
12% w/v Polyethylene glycol 3,350
16% w/v Polyethylene glycol 3,350
20% w/v Polyethylene glycol 3,350
24% w/v Polyethylene glycol 3,350

Dehydrant

24% w/v Polyethylene glycol 3,350

SUCCESS STORY

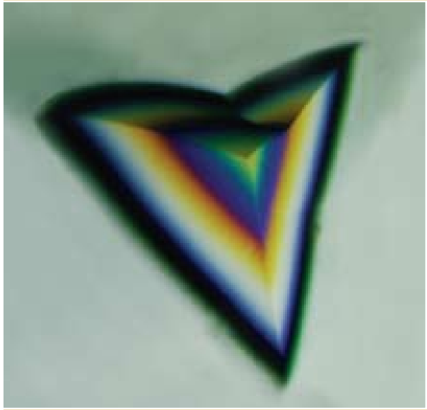

Crystal is that of the intact canine lymphoma antibody. Preliminary crystallization conditions determined using the Low Ionic Strength Screen.

Courtesy of Lisa Harris.

University of California, Irvine

DESCRIPTION

Low Ionic Strength Screen™
Crystallization Kit
HAMPTON RESEARCH

The Low Ionic Strength Screen is effective for determining the preliminary crystallization conditions of intact monoclonal antibodies[1]. However, this screen is not just an intact antibody screen. The screen has effectively determined the preliminary crystallization conditions for numerous monoclonal antibody fragments as well as other soluble proteins. The screen should be utilized as a low ionic strength crystallization screen for proteins where this strategy could be effective in determining preliminary crystallization screens.

In this screen, the concentration of a high purity, monodisperse PEG 3,350 is varied from 4 to 24% w/v (4, 8, 12, 16, 20 and 24% w/v) versus a pH range of 2 to 12 (2, 3, 3.5, 4, 4.5, 5, 5.5, 6.0, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 11, 12). Stock buffer concentrations are 50 mM. Final buffer concentration in the drop is typically 10 mM.

Unique features of this screen: (1) Low ionic strength. Buffer concentration is supplied as 50 mM, resulting in an initial drop concentration of 10 mM (Drop = 4 μl sample, 2 μl buffer, 5 μl precipitant); (2) The Polyethylene glycol (PEG) 3,350 is a special, high-purity, monodisperse preparation with an Mr of 3300-3400. Most PEGs of this molecular weight have an Mr of plus or minus 500 rather than 50; (3) An extremely broad range of pH 2 to 12 is sampled. At low ionic strength, the effects of pH and temperature upon sample solubility are amplified. Hence this screen allows one to critically evaluate the effects of temperature and pH upon sample solubility and crystallization. It is recommended the screen be repeated at several temperatures between 4°C and 37°C to take advantage of the low ionic strength feature.

The format of this screen is different than other screens offered by Hampton Research. The screen is supplied as a set of eighteen 50 mM buffers, six PEG 3,350 concentrations, and a single dehydrant, PEG 3,350. This formulation allows one to screen up to 108 conditions if desired. The formulation is designed to offer options. Coarse sample of pH and/or precipitant concentration can be performed with 24 or even 12 drops. This will eliminate large regions of sampling space and conserve sample. Subsequent screens and perhaps even optimization can encompass an increasingly finer grid matrix of pH, precipitant concentration, as well as other variables significant for sample crystallization. The protocol requires the following pipetting steps for a typical vapor diffusion experiment: (1) Pipet dehydrant to reservoir; (2) Pipet drop; (3) Pipet buffer to drop; (4) Pipet precipitant to drop.

Each kit contains 24 unique reagents of varying pH and PEG 3,350 concentration. Buffers of 50 mM are supplied for pH 2, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 11, and 12 in 1 ml volumes. The precipitant Polyethylene glycol 3,350 is supplied as 4, 8, 12, 16, 20, and 24% w/v in 1 ml volumes. Dehydrant is supplied as 250 ml of 24% PEG 3,350. Combining each pH and precipitant supplied can generate 108 crystallization screening conditions. All solutions are sterile filtered and formulated with ultra-pure Type 1 water.

References

1. Harris, et al., Crystallization of intact monoclonal antibodies, Proteins: Structure, Function, and Genetics (1995) 23:285-289.

Order Information

Each Low Ionic Strength Screen kit contains 24 unique reagents plus dehydrant.

Cat. No.	Name	Description	Price
HR2-120	**Low Ionic Strength Screen**	1 ml, tube format + 250 ml bottle	**$195.00**

SILVER BULLETS™ • SILVER BULLETS BIO™

APPLICATIONS

- Primary, secondary or orthoganol crystallization screen for biological macromolecules
- For use with soluble proteins and membrane proteins
- A methodology to uncover different crystal forms
- Optimization screening in conjunction with preliminary crystallization conditions
- ThermoFluor® assays
- Solubility, folding and stability studies

FEATURES

- Developed at Hampton Research
- Screens a portfolio of small molecules for their ability to establish stabilizing, intermolecular, hydrogen bonding, hydrophobic and electrostatic interactions which could promote lattice formation and crystallization
    - Organic salts & acids
    - Biologically active small molecules
    - Amino acids and peptide
    - Macromolecular digests
- Modular screen design allows Silver Bullets to be screened against virtually any crystallization reagent
- Original & Bio Formulations
- A combinatorial library of more than 1,090 chemicals, of which more than 400 are unique
- Tube or Deep Well block format

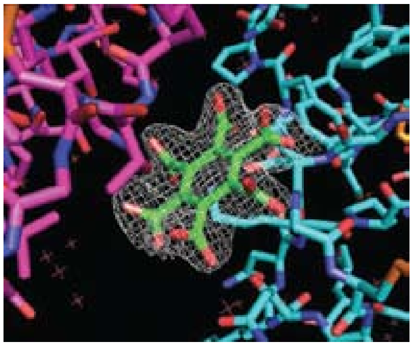

Mellitic acid, a Silver Bullet for Trypsin

DESCRIPTION

An alternative approach to macromolecular crystallization. The Silver Bullets formulation and methodology screens validated small molecule libraries against a few select crystallization reagents as an effective, orthogonal approach to macromolecular crystallization. The formulation offers an alternative and efficient approach that is an excellent complement to current crystallization method-ologies.

Silver Bullets is a library of small molecules that have been shown to promote crystal lattice formation. X-ray diffraction analysis has demonstrated the reagents have the ability to:

- Stabilize the conformation of the protein
- Perturb the interaction of the protein with the solvent
- Participate in forming important lattice contacts
- Build the crystal lattice by forming reversible cross-links between the macromolecules in the crystal

Results with the Silver Bullets approach have been very encouraging, with more than twice as many proteins being crystallized overall as were crystallized from controls free of any small molecules.[1-3] There have been frequent occasions where some exceptional result has been obtained for specific proteins, and numerous examples where new or unusual crystal forms were obtained. X-ray diffraction analysis has revealed the small molecules in the crystal lattice, involved at the centers of hydrogen bonding networks and electrostatic interaction.[3]

Silver Bullets can be used as a primary crystallization screen; as a secondary or orthogonal screen to produce crystals when traditional screens are not successful; as an optimization screen in conjunction with preliminary crystallization conditions; as a methodology to uncover different crystal forms; to promote intramolecular interactions that may stabilize a macromolecules conformation and promote crystallization. Silver Bullets solutions are designed for use with the Silver Bullets PEG/Tacsimate Crystallization Reagents but may also be used with other crystallization reagents.

The Silver Bullets kits are composed of 96 solutions in either screw cap tubes or a single Deep Well block (Greiner 786261) HT format. Each reagent is a mixture of small molecules or macromolecular digest in 0.02 M HEPES sodium pH 6.8 buffer. Each Silver Bullets solution is supplied in a 0.25 ml volume. Each solution contains between 2 and 20 small molecules. Silver Bullets reagents include: Organic salts and acids • Biologically active molecules • Amino acids and peptides • Macromolecular digests. Silver Bullets Bio reagents include: Small organic molecules, organic salts, and organic acids • Biologically active molecules, co-factors, and ligands • Amino acids, peptides, nucleotides, drugs, and carbohydrates • Biochemical pathway intermediates.

References

1. *Searching for silver bullets: An alternative strategy for crystallizing macromolecules. Alexander McPherson and Bob Cudney. Journal of Structural Biology 156 (2006) 387-406.*
2. *A novel strategy for the crystallization of proteins: X-ray diffraction validation. Steven B. Larson, John S. Day, Robert Cudney, and Alexander McPherson. Acta Cryst. (2007) D63, 310-318.*
3. *Development of an alternative approach to protein crystallization. McPherson, Alexander; Nguyen, Chieniang; Larson, Steven B; Day, John S; Cudney, Bob. J Struct Funct Genomics, Volume 8, Number 4, December 2007, 193-198.*

Order Information

Silver Bullets and Silver Bullets Bio kits each contain 96 unique reagents. To order individual reagents, use Custom Shop™ catalog numbers listed below. Refer to page 36 for further details.

Cat. No.	Name	Description	Price
HR2-078	**Silver Bullets**	**0.25 ml, tube format**	**$490.00**
HR2-096	**Silver Bullets HT**	**0.25 ml, Deep Well block format**	**$490.00**
HR2-080	**Silver Bullets Bio**	**0.25 ml, tube format**	**$490.00**
HR2-088	**Silver Bullets Bio HT**	**0.25 ml, Deep Well block format**	**$490.00**
HR2-090	**PEG/Tacsimate pH 5.8**	**1 ml, Deep Well block format**	**$185.00**
HR2-092	**PEG/Tacsimate pH 6.8**	**1 ml, Deep Well block format**	**$185.00**
HR2-094	**PEG/Tacsimate pH 7.8**	**1 ml, Deep Well block format**	**$185.00**
HR2-996-**	**Silver Bullets Custom Shop**	**0.25 ml**	**$65.00**
HR2-988-**	**Silver Bullets Bio Custom Shop**	**0.25 ml**	**$65.00**

**** = reagent number 1-96**

MACROMOLECULAR DIGESTS
AMINO ACIDS
PEPTIDES

COMBINATORIAL
FORMULATION

ORGANIC SALTS
ORGANIC ACIDS

silver • bullets™
FACTORS

COMPATIBLE WITH A DIVERSE
ARRAY OF REAGENTS

BIOLOGICALLY ACTIVE SMALL MOLECULES

COMPOSED OF COMPOUNDS WITH THE ABILITY TO

- ENGAGE IN MULTIPLE, STRONG HYDROGEN BONDS WITHIN THE PROTEIN
- FORM HYDROPHOBIC INTERACTIONS WITH THE PROTEIN
- BECOME INVOLVED IN ELECTROSTATIC INTERACTIONS WITH THE PROTEIN
- ENHANCE, DISRUPT AND ALTER THE PATTERN OF CLUSTERING IN THE CRITICAL NUCLEUS OF THE CRYSTAL
- DEVELOP STABILIZING LATTICE INTERACTIONS WITHIN THE CRYSTAL
- SERVE AS TRANSIENT STABILIZING INTERMEDIARIES

COMPATIBLE WITH A DIVERSE
ARRAY OF REAGENTS

AMINO ACIDS
PEPTIDES
NUCLEOTIDES
CARBOHYDRATES

COMBINATORIAL
FORMULATION

silver • bullets BIO™
FACTORS

SMALL ORGANIC MOLECULES
ORGANIC SALTS
ORGANIC ACIDS

BIOLOGICALLY ACTIVE MOLECULES CO-FACTORS LIGANDS

ADDITIVE SCREEN™ • ADDITIVE SCREEN HT™

APPLICATION

- Manipulate sample-sample & sample-solvent interactions to enhance or alter sample solubility

FEATURES

- 18 classes of reagents
- Highly concentrated (10x) reagent formulation
- 96 unique reagents
- Crystallization or sample solubility optimization
- Tube or Deep Well block format

SUCCESS STORY

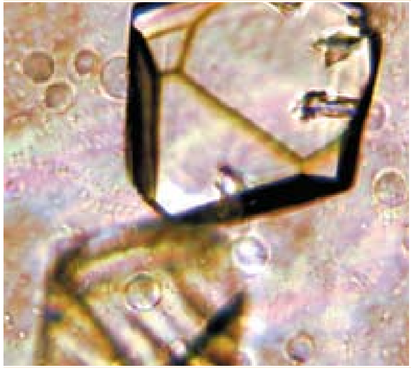

Initial crystals grown using Crystal Screen Reagent Number 9. Crystals were approximately 100 microns in size. With the addition of Dextran Sulfate from the Additive Screen kit, larger crystals were produced as pictured above (approximately 250 microns).

Courtesy of Ranjit K. Deka.

University of Texas Southwestern Medical Center

DESCRIPTION

Additive Screen

Additive Screen is a library of small molecules that can affect the solubility and crystallization of biological macromolecules, including both soluble and membrane proteins. These small molecules can perturb and manipulate sample-sample and sample-solvent interactions, as well as perturb water structure which can alter and improve both the solubility and crystallization of a sample. Additives can stabilize or engender conformity by specific interaction with the macromolecules. There are numerous reports of the use of additives to improve the quality and size of macromolecular crystals.[1-5]

Additive Screen contains 96 unique reagents, 1 ml each, tube format.

Additive Screen HT contains 96 unique reagents, 1 ml each, in a single Deep Well block format.

References

1. *Crystallization of membrane proteins., Edited by Hartmut Michel, CRC Press (1991).*
2. *Crystallization of Nucleic Acids and Proteins: A Practical Approach., Edited by A. Ducruix and R. Giege, Oxford University Press (1992).*
3. *Cudney, R., et al., Screening and optimization strategies for macromolecular crystal growth., Acta Cryst. (1994) D50, 414-423.*
4. *Sousa R., Use of glycerol, polyols and other protein structure stabilizing agents in protein crystallization., Acta Cryst. (1995) D51, 271-277.*
5. *Trakhanov, S. and Quiocho, F.A., Influence of divalent cations on protein crystallization., Protein Science (1995) 4, 9, 1914-1919.*

Order Information

Each Additive Screen kit contains 96 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 36 for further details.

Cat. No.	Name	Description	Price
HR2-428	**Additive Screen**	**1 ml, tube format**	**$490.00**
HR2-138	**Additive Screen HT**	**1 ml, Deep Well block format**	**$490.00**
HR2-428-**	**Additive Screen Custom Shop**	**1 ml**	**$65.00**

** = reagent number 1-96

ADDITIVE SCREEN FORMULATION

Tube #	Additive
1. (A1)	0.1 M Barium chloride dihydrate
2. (A2)	0.1 M Cadmium chloride hydrate
3. (A3)	0.1 M Calcium chloride dihydrate
4. (A4)	0.1 M Cobalt(II) chloride hexahydrate
5. (A5)	0.1 M Copper(II) chloride dihydrate
6. (A6)	0.1 M Magnesium chloride hexahydrate
7. (A7)	0.1 M Manganese(II) chloride tetrahydrate
8. (A8)	0.1 M Strontium chloride hexahydrate
9. (A9)	0.1 M Yttrium(II) chloride hexahydrate
10. (A10)	0.1 M Zinc chloride
11. (A11)	0.1 M Iron(III) chloride hexahydrate
12. (A12)	0.1 M Nickel(II) chloride hexahydrate
13. (B1)	0.1 M Chromium(III) chloride hexahydrate
14. (B2)	0.1 M Praseodymium(III) acetate hydrate
15. (B3)	1.0 M Ammonium sulfate
16. (B4)	1.0 M Potassium chloride
17. (B5)	1.0 M Lithium chloride
18. (B6)	2.0 M Sodium chloride
19. (B7)	0.5 M Sodium fluoride
20. (B8)	1.0 M Sodium iodide
21. (B9)	2.0 M Sodium thiocyanate
22. (B10)	1.0 M Potassium sodium tartrate tetrahydrate
23. (B11)	1.0 M Sodium citrate tribasic dihydrate
24. (B12)	1.0 M Cesium chloride
25. (C1)	1.0 M Sodium malonate pH 7.0
26. (C2)	0.1 M L-Proline
27. (C3)	0.1 M Phenol
28. (C4)	30% v/v Dimethyl sulfoxide
29. (C5)	0.1 M Sodium bromide
30. (C6)	30% w/v 6-Aminohexanoic acid
31. (C7)	30% w/v 1,5-Diaminopentane dihydrochloride
32. (C8)	30% w/v 1,6-Diaminohexane
33. (C9)	30% w/v 1,8-Diaminooctane
34. (C10)	1.0 M Glycine
35. (C11)	0.3 M Glycyl-glycyl-glycine
36. (C12)	0.1 M Taurine
37. (D1)	0.1 M Betaine hydrochloride
38. (D2)	0.1 M Spermidine
39. (D3)	0.1 M Spermine tetrahydrochloride
40. (D4)	0.1 M Hexammine cobalt(III) chloride
41. (D5)	0.1 M Sarcosine
42. (D6)	0.1 M Trimethylamine hydrochloride
43. (D7)	1.0 M Guanidine hydrochloride
44. (D8)	0.1 M Urea
45. (D9)	0.1 M β-Nicotinamide adenine dinucleotide hyd.
46. (D10)	0.1 M Adenosine-5'-triphosphate disodium salt
47. (D11)	0.1 M TCEP hydrochloride
48. (D12)	0.01 M GSH (L-Glutathione reduced), 0.01 M GSSG (L-Glutathione oxidized)
49. (E1)	0.1 M EDTA sodium salt
50. (E2)	5% w/v Polyvinylpyrrolidone K15
51. (E3)	30% w/v Dextran sulfate sodium salt (Mr 5,000)
52. (E4)	40% v/v Pentaerythritol ethoxylate (3/4 EO/OH)
53. (E5)	10% w/v Polyethylene glycol 3,350
54. (E6)	30% w/v D(+)-Glucose monohydrate
55. (E7)	30% w/v Sucrose
56. (E8)	30% w/v Xylitol
57. (E9)	30% w/v D-Sorbitol
58. (E10)	12% w/v myo-Inositol
59. (E11)	30% w/v D-(+)-Trehalose dihydrate
60. (E12)	30% w/v D-(+)-Galactose
61. (F1)	30% v/v Ethylene glycol
62. (F2)	30% v/v Glycerol
63. (F3)	3.0 M NDSB-195
64. (F4)	2.0 M NDSB-201
65. (F5)	2.0 M NDSB-211
66. (F6)	2.0 M NDSB-221
67. (F7)	1.0 M NDSB-256
68. (F8)	0.15 mM CYMAL®-7
69. (F9)	20% w/v Benzamidine hydrochloride hydrate
70. (F10)	5% w/v n-dodecyl-N,N-dimethylamine-N-oxide
71. (F11)	5% w/v n-Octyl-β-D-glucoside
72. (F12)	5% w/v n-Dodecyl-β-D-maltoside
73. (G1)	30% w/v Trimethylamine N-oxide dihydrate
74. (G2)	30% w/v 1,6-Hexanediol
75. (G3)	30% v/v (+/-)-2-Methyl-2,4-pentanediol
76. (G4)	50% v/v Polyethylene glycol 400
77. (G5)	50% v/v Jeffamine M-600® pH 7.0
78. (G6)	40% v/v 2,5-Hexanediol
79. (G7)	40% v/v (±)-1,3-Butanediol
80. (G8)	40% v/v Polypropylene glycol P 400
81. (G9)	30% v/v 1,4-Dioxane
82. (G10)	30% v/v Ethanol
83. (G11)	30% v/v 2-Propanol
84. (G12)	30% v/v Methanol
85. (H1)	40% v/v 1,4-Butanediol
86. (H2)	40% v/v tert-Butanol
87. (H3)	40% v/v 1,3-Propanediol
88. (H4)	40% v/v Acetonitrile
89. (H5)	40% v/v Formamide
90. (H6)	40% v/v 1-Propanol
91. (H7)	5% v/v Ethyl acetate
92. (H8)	40% v/v Acetone
93. (H9)	0.25% v/v Dichloromethane
94. (H10)	7% v/v 1-Butanol
95. (H11)	40% v/v 2,2,2-Trifluoroethanol
96. (H12)	40% v/v 1,1,1,3,3,3-Hexafluoro-2-propanol

Concentrations are actual formulations.

DETERGENT SCREEN™ • DETERGENT SCREEN HT™

APPLICATION

- Prevent and manipulate non-specific aggregation due to hydrophobic interactions

FEATURES

- Kits include popular detergents used in the crystallization of membrane proteins
  - Ionic detergents
  - Non-ionic detergents
  - Zwitterionic detergents
  - Non-detergent Sulfobetaines
  - Synthetic lipids
- Useful with soluble proteins where hydrophobic interactions limit sample solubility
- Detergents are compatible with microbatch under oil crystallization, as well as vapor diffusion, free interface diffusion, & dialysis

SUCCESS STORY

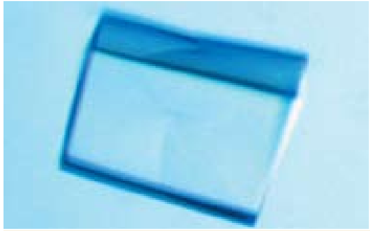

Crystal above is that of ß-Lactamase. Crystal was grown using 15% w/v PEG 6000, 50 mM HEPES pH 7.0, and in the presence of CYMAL-5 and CYMAL-6 detergent from the Hampton Research Detergent Screen kit 1.

Courtesy of James Knox.
University of Connecticut,
Department of Molecular and Cell Biology

DESCRIPTION

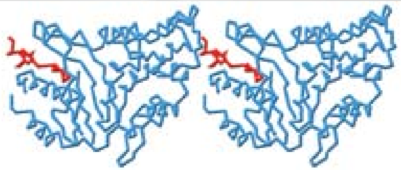

Detergent molecule interacting with beta-lactamase.

Detergent Screen kits are designed to allow the rapid and convenient evaluation of 96 unique detergent reagents for their ability to influence the solubility and crystallization of the sample. These screens are designed to be compatible with most popular crystallization reagents.

The detergents in these kits are capable of manipulating hydrophobic sample-sample interactions which can lead to non-specific aggregation, and prevent or interfere with sample crystallization. The detergents also perturb water structure which may play a role in sample crystallization.

Non-specific aggregation is a common deterrent to the crystallization of soluble macromolecules as well as membrane proteins. There is extensive literature demonstrating the effectiveness of including detergents in the crystallization trial towards preventing non-specific aggregation due to hydrophobic interactions and hence improving crystallization.[1] The kits allow one to screen the most effective detergents used in crystal growth.

The Detergent Screen kits are preformulated so that simple pipetting is all that is required to screen the detergents with the hanging or sitting drop vapor diffusion technique. Screens with these kits are usually performed after preliminary crystallization conditions have been determined, although ab initio screens are also practical. The kits are recommended for both soluble and membrane proteins where non-specific aggregation is a suspected complication or where one simply wishes to screen detergents as an optimization variable. The screens are suitable for hanging drop, sitting drop, microbatch, free interface diffusion and sandwich drop crystallization methodologies.

Each Detergent Screen kit contains 96 unique reagents. 0.25 ml of detergent solution is preformulated at 10 times the reported CMC (unless otherwise noted) in sterile filtered water.

References

1. *Crystallization of membrane proteins. Edited by Hartmut Michel, CRC Press, (1991).*

Order Information

Each Detergent Screen kit contains 96 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 36 for further details.

Cat. No.	Name	Description	Price
HR2-408	**Detergent Screen**	**0.25 ml, tube format**	**$490.00**
HR2-406	**Detergent Screen HT**	**0.25 ml, Deep Well block format**	**$490.00**
HR2-406-**	**Detergent Screen Custom Shop**	**0.5 ml**	**$65.00**

** = reagent number 1-96

DETERGENT SCREEN FORMULATION

Well #	Detergent	CMC (mM)
1. (A1)	BAM	None
2. (A2)	n-Dodecyl-β-iminodipropionic acid, monosodium salt	None
3. (A3)	Dodecyltrimethylammonium chloride	0.046
4. (A4)	CTAB	1.0
5. (A5)	Deoxycholic acid, sodium salt	6.0
6. (A6)	Sodium dodecyl sulfate	8.27
7. (A7)	Sodium cholate	14.0
8. (A8)	Sodium dodecanoyl sarcosine	14.4
9. (A9)	ANAPOE® -X-305	None
10. (A10)	IPTG	None
11. (A11)	n-Hexadecyl-β-D-maltoside	0.0006
12. (A12)	ANAPOE® -58	0.004
13. (B1)	n-Tetradecyl-β-D-maltoside	0.010
14. (B2)	ANAPOE® -80	0.012
15. (B3)	n-Tridecyl-β-D-maltoside	0.033
16. (B4)	ANAPOE® -$C_{12}E_9$	0.05
17. (B5)	ANAPOE® -20	0.059
18. (B6)	Thesit®	0.09
19. (B7)	ANAPOE® -35	0.091
20. (B8)	ANAPOE® -$C_{13}E_8$	0.10
21. (B9)	ANAPOE® -$C_{12}E_8$	0.11
22. (B10)	n-Dodecyl-β-D-maltoside	0.17
23. (B11)	CYMAL® -7	0.19
24. (B12)	ANAPOE® -X-114	0.20
25. (C1)	ANAPOE® -$C_{12}E_{10}$	0.20
26. (C2)	Sucrose monolaurate	0.30
27. (C3)	CYMAL® -6	0.56
28. (C4)	n-Undecyl-β-D-maltoside	0.59
29. (C5)	ANAPOE® -X-405	0.81
30. (C6)	TRITON® X-100	0.90
31. (C7)	ANAPOE® -$C_{10}E_6$	0.90
32. (C8)	n-Decyl-β-D-thiomaltoside	0.90
33. (C9)	Octyl maltoside, fluorinated	1.02
34. (C10)	ANAPOE® -$C_{10}E_9$	1.3
35. (C11)	Big CHAP, deoxy	1.4
36. (C12)	n-Decyl-β-D-maltoside	1.8
37. (D1)	LDAO	2.0
38. (D2)	n-Decanoylsucrose	2.5
39. (D3)	n-Nonyl-β-D-thioglucoside	2.9
40. (D4)	n-Nonyl-β-D-thiomaltoside	3.2
41. (D5)	CYMAL® -5	5.0
42. (D6)	n-Nonyl-β-D-maltoside	6.0
43. (D7)	n-Nonyl-β-D-glucoside	6.5
44. (D8)	HEGA® -10	7.0
45. (D9)	MEGA -10	7.0
46. (D10)	C_8E_5	7.1
47. (D11)	CYMAL® -4	7.6
48. (D12)	C_8E_4	8.0
49. (E1)	n-Octyl-β-D-thiomaltoside	8.5
50. (E2)	n-Octyl-β-D-thioglucoside	9.0
51. (E3)	Hexaethylene glycol monooctyl ether	10.0
52. (E4)	DDAO	10.4
53. (E5)	C-HEGA® -11	11.5
54. (E6)	Pluronic® F-68	17.9
55. (E7)	HECAMEG®	19.5
56. (E8)	n-Octyl-β-D-glucoside	20.0
57. (E9)	n-Octanoylsucrose	24.4
58. (E10)	MEGA-9	25.0
59. (E11)	2,6-Dimethyl-4-heptyl-β-D-malto-pyranoside	27.5
60. (E12)	n-Heptyl-β-D-thioglucopyranoside	29.0
61. (F1)	n-Octyl-β-D-galactopyranoside	29.5
62. (F2)	CYMAL® -3	34.5
63. (F3)	C-HEGA® -10	35.0
64. (F4)	HEGA® -9	39.0
65. (F5)	Dimethyloctylphosphine oxide	40.0
66. (F6)	MEGA-8	79.0
67. (F7)	C-HEGA® -9	108.0
68. (F8)	HEGA® -8	109.0
69. (F9)	CYMAL® -2	120.0
70. (F10)	n-Hexyl-β-D-glucopyranoside	250.0
71. (F11)	C-HEGA® -8	277.0
72. (F12)	CYMAL® -1	340.0
73. (G1)	NDSB-195	N/A
74. (G2)	NDSB-201	N/A
75. (G3)	NDSB-211	N/A
76. (G4)	NDSB-221	N/A
77. (G5)	NDSB-256	N/A
78. (G6)	ZWITTERGENT® 3-14	0.40
79. (G7)	n-Dodecyl-N,N-dimethylglycine	1.5
80. (G8)	FOS-Choline® -12	1.5
81. (G9)	FOS-Choline® -8, fluorinated	2.2
82. (G10)	n-Undecyl-N,N-Dimethlamine-oxide	3.21
83. (G11)	ZWITTERGENT® 3-12	4.0
84. (G12)	DDMAB	4.3
85. (H1)	FOS-MEA® -10	5.25
86. (H2)	CHAPS	8.0
87. (H3)	CHAPSO	8.0
88. (H4)	FOS-Choline® -10	11.0
89. (H5)	n-Decyl-N,N-dimethylglycine	19.0
90. (H6)	FOS-Choline® -9	39.5
91. (H7)	ZWITTERGENT® 3-10	40.0
92. (H8)	CYCLOFOS™ -3	43.0
93. (H9)	FOS-Choline® -8	114.0
94. (H10)	ZWITTERGENT® 3-08	330.0
95. (H11)	LysoFos™ Choline 12	0.7
96. (H12)	LysoFos™ Choline 10	7.0

HEAVY ATOM SCREENS

APPLICATION

- Heavy atoms for multiple isomorphous replacement

FEATURES

- Convenient sets of popular heavy atoms
- Each kit includes a heavy atom tutorial and formulation guide

DESCRIPTION

Heavy Atom Screen Pt
Multiple Isomorphous Replacement

A Heavy Atom Screen kit provides 50 mg of each heavy atom in an o-ring screw cap micro tube. This convenient format provides sufficient material for the preparation of numerous (15 x 100 mM or 1,500 x 1 mM) small volume (100 μl) fresh stock solutions of heavy atoms. The quantity is sufficient for screening and derivatization but avoids the problem of storing large quantities of heavy atoms. Individual heavy atoms from each kit are available in larger quantities on a special order basis for researchers requiring larger amounts of material.

References

1. *Heavy-atom derivatization. Elspeth Garman and James W. Murray, Acta Cryst. (2003). D59, 1903-1913.*
2. *Towards a rational approach for heavy-atom derivative screening in protein crystallography. Johnson Agniswamy, M. Gordon Joyce, Carl H. Hammer, and Peter D. Sun. Acta Cryst. (2008) D64, 354-367.*
3. *Screening for phasing atoms in protein crystallography. Titus J Boggon and Lawrence Shapiro. Structure 2000, Vol 8 No 7, 143-149.*

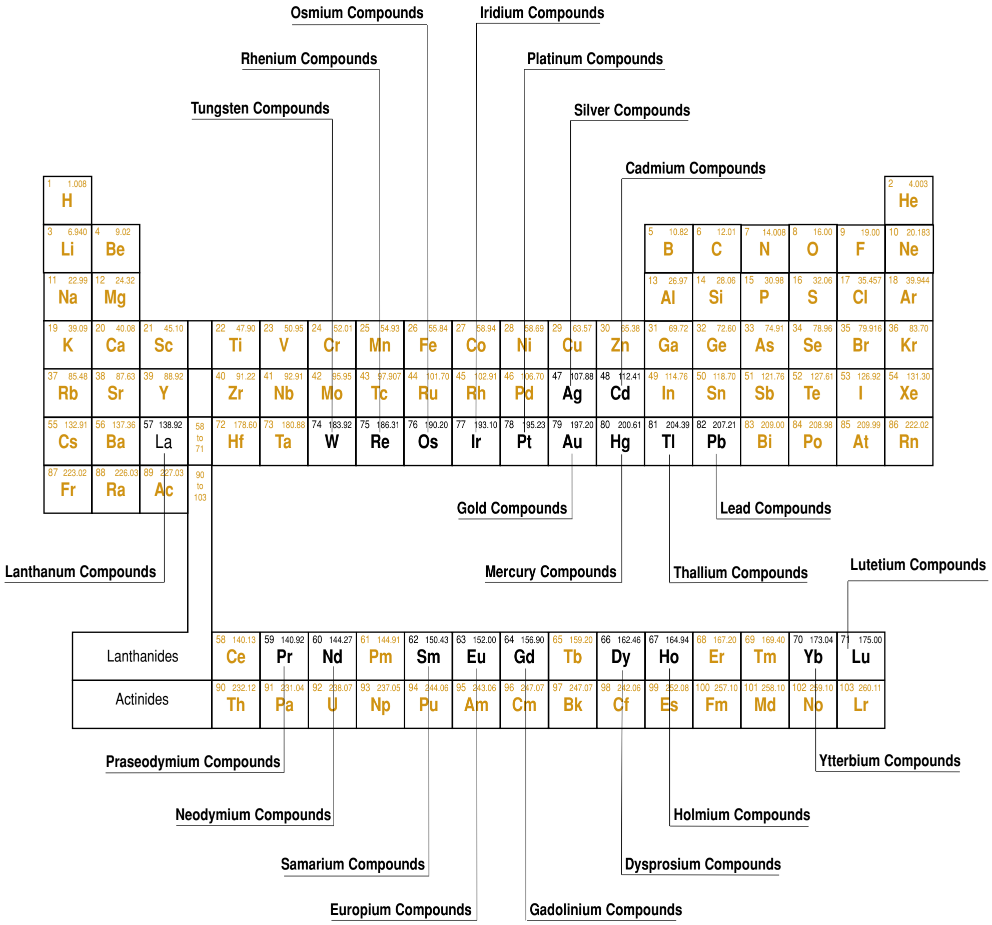
Osmium Compounds
Iridium Compounds
Rhenium Compounds
Platinum Compounds
Tungsten Compounds
Silver Compounds
Cadmium Compounds
Gold Compounds
Lead Compounds
Lanthanum Compounds
Mercury Compounds
Thallium Compounds
Lutetium Compounds
Lanthanides
Actinides
Praseodymium Compounds
Ytterbium Compounds
Neodymium Compounds
Holmium Compounds
Samarium Compounds
Dysprosium Compounds
Europium Compounds
Gadolinium Compounds

SUCCESS STORY

Protein crystal with the heavy atom derivative Potassium osmate.

Kimberly J. Skinner, Structural Biology and Biophysics, Pfizer Global R&D, Sandwich, Kent, United Kingdom.

Heavy Atom Screen Pt (Platinum)

1. Potassium tetrachloroplatinate(II)
2. Ammonium tetrachloroplatinate(II)
3. Potassium hexachloroplatinate(IV)
4. Potassium tetranitroplatinate(II)
5. Potassium tetracyanoplatinate(II) hydrate
6. Dichloro(ethylenediamine)platinum(II)
7. Diammino Platinum Dinitrite
8. Potassium tetrabromoplatinate(II)
9. Potassium hexabromoplatinate(IV)
10. Platinum potassium iodide
11. Platinum potassium thiocyanate
12. Di-μ-iodobis(ethylenediamine)diplatinum(II) nitrate

Order Information

Each Heavy Atom Screen Pt kit contains 12 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 36 for further details.

Cat. No.	Name	Description	Price
HR2-442	**Heavy Atom Screen Pt**	50 mg, tube format	**$418.00**
HR2-442-**	**Heavy Atom Screen Pt Custom Shop**	100 mg	**$80.00**

** = reagent number 1-12

Heavy Atom Screen Au (Gold)

1. Gold(I) potassium cyanide
2. Potassium tetrachloroaurate(III) hydrate
3. Sodium tetrachloroaurate(III) dihydrate
4. Gold(III) chloride
5. Hydrogen tetrachloroaurate(III) trihydrate
6. Potassium tetrabromoaurate(III) dihydrate

Order Information

Each Heavy Atom Screen Au kit contains 6 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 36 for further details.

Cat. No.	Name	Description	Price
HR2-444	**Heavy Atom Screen Au**	50 mg, tube format	**$275.00**
HR2-444-**	**Heavy Atom Screen Au Custom Shop**	100 mg	**$80.00**

** = reagent number 1-6

Heavy Atom Screen Hg (Mercury)

1. Mersalyl acid
2. Ethyl Mercuric Phosphate
3. Mercury(II) chloride
4. Mercury(II) acetate
5. Ethylmercurithiosalicylic acid, sodium salt
6. Phenylmercury acetate
7. Mercury(II) potassium iodide
8. p-Chloromercuribenzoic acid
9. Ethylmercury chloride
10. Mercury(II) bromide
11. Mercury(II) iodide
12. Mercury(II) nitrate monohydrate
13. Mercury(II) cyanide
14. Mercury(II) oxide, yellow
15. Tetrakis(acetoxymercuri)methane

Order Information

Each Heavy Atom Screen Hg kit contains 15 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 36 for further details.

Cat. No.	Name	Description	Price
HR2-446	**Heavy Atom Screen Hg**	50 mg, tube format	**$314.00**
HR2-446-**	**Heavy Atom Screen Hg Custom Shop**	100 mg	**$80.00**

** = reagent number 1-15

Heavy Atom Screen M1

1. Thallium(III) chloride hydrate
2. Thallium(I) chloride
3. Thallium(III) acetate hydrate
4. Lead(II) acetate trihydrate
5. Lead(II) nitrate
6. Lead(II) chloride
7. Silver nitrate
8. Cadmium chloride hydrate
9. Cadmium iodide
10. Potassium hexachloroiridate(IV)
11. Iridium(III) chloride hydrate
12. Sodium hexachloroiridate(III) hydrate
13. Ammonium hexachloroiridate(III) hydrate
14. Potassium hexanitroiridium(III)
15. Potassium osmate(VI) dihydrate
16. Ammonium hexabromoosmate(IV)
17. Potassium hexachloroosmate(IV)
18. Osmium(III) chloride hydrate
19. Acetoxytrimethyllead(IV)

Order Information

Each Heavy Atom Screen M1 kit contains 19 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 36 for further details.

Cat. No.	Name	Description	Price
HR2-448	**Heavy Atom Screen M1**	50 mg, tube format	**$343.00**
HR2-448-**	**Heavy Atom Screen M1 Custom Shop**	100 mg	**$80.00**

** = reagent number 1-19

Heavy Atom Screen M2

1. Sodium tungstate dihydrate
2. Ammonium tetrathiotungstate(VI)
3. Samarium(III) chloride hexahydrate
4. Samarium(III) acetate hydrate
5. Samarium(III) nitrate hexahydrate
6. Lanthanum(III) nitrate hexahydrate
7. Europium(III) nitrate hexahydrate
8. Europium(III) chloride hexahydrate
9. Gadolinium(III) chloride hydrate
10. Lutetium(III) chloride hexahydrate
11. Lutetium(III) acetate hydrate
12. Ytterbium(III) chloride hydrate
13. Dysprosium(III) chloride hexahydrate
14. Praseodymium(III) chloride heptahydrate
15. Neodymium(III) chloride hydrate
16. Holmium(III) chloride hexahydrate
17. Potassium hexachlororhenate(IV)
18. Potassium perrhenate

Order Information

Each Heavy Atom Screen M2 kit contains 18 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 36 for further details.

Cat. No.	Name	Description	Price
HR2-450	**Heavy Atom Screen M2**	50 mg, tube format	**$320.00**
HR2-450-**	**Heavy Atom Screen M2 Custom Shop**	100 mg	**$80.00**

** = reagent number 1-18

I3C PHASING KIT™

APPLICATION

- Produce heavy atom derivative of biological macromolecules for phasing

FEATURES

- I3C kit for soaking or cocrystallization
- Phasing using SAD or SIRAS
- Ready-to-use, no weighing required

DESCRIPTION

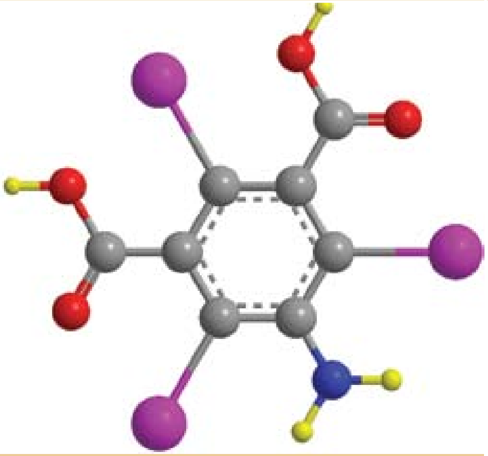

I3C can be used for heavy atom derivatization of biological macromolecules for subsequent single wavelength anomalous dispersion (SAD) or single isomorphous replacement plus anomalous scattering (SIRAS).

The two carboxylic acid groups and one amino group of I3C can interact by way of hydrogen bonds with both the backbone and side chains of proteins. This can result in a relatively highoccupancy of the bound ligand I3C. The three iodine atoms per I3C molecule provide for a strong anomalous signal.

The three iodine atoms in I3C form an equilateral triangle with 6.0 Å side lengths. These triangular structures can readily be identified in the anomalous electron density map.

Kit Contents
- 12 aliquots of I3C (280 mg each) (5-Amino-2,4,6-triiodoisophthalic acid CAS number 35453-19-1)
- 12 aliquots of 2.0 M Lithium hydroxide (650 μl each)
- User Guide

References
1. *A magic triangle for experimental phasing of macromolecules. Beck et al. Acta Cryst. D64, 1179, 2008.*
2. *Structure determination of the cancer-associated Mycoplasma hyorhinis protein Mhp37. Sippel et al. Acta Cryst. D64, 1172, 2008.*
3. *5-Amino-2,4,6 triiodo-isophthalic acid monohydrate. Beck et al. Acta Cryst. E64:1286, 2008.*

Order Information

Cat. No.	Name	Description	Price
HR3-133	**I3C Phasing Kit**	24 tubes	**$155.00**

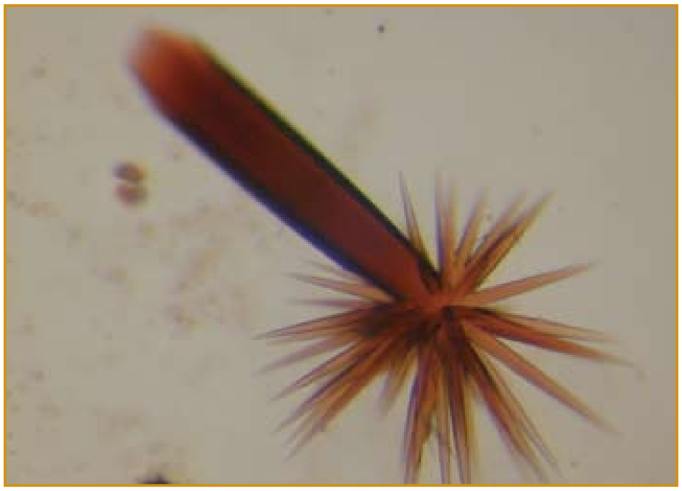

Crystal of cytochrome P450 Cyp121 from Mycobacterium tuberculosis with red color due to the ferric heme iron. Crystal diffracts to 1.4Ang at Proxima1/Soleil, France.

Pascal Belin, iBiTec-s/CEA, Saclay, France.

Crystal Image.

Allan D'Arcy, Switzerland.

Crystal of human phosphatase.

Chris Gray at the Institute of Cancer Research, London, United Kingdom.

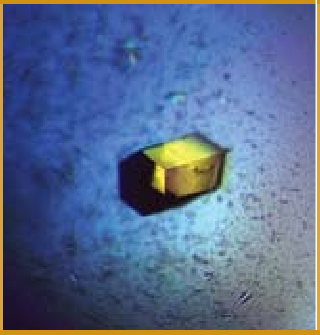

Crystals of apo kinase.
Initial hit from Hampton Research Grid Screen Salt HT.

Michelle Quiles and Annie Hassell, GlaxoSmithKline, RTP, North Carolina, USA.

CUSTOM SHOP CRYSTALLIZATION REAGENTS

Protein crystal grown with additive.

Ranjit Deka, University of Texas Southwest Medical Center, USA.

TABLE OF CONTENTS

PAGES

36 - 37 CUSTOM SHOP™ CRYSTALLIZATION REAGENTS

HAMPTON RESEARCH

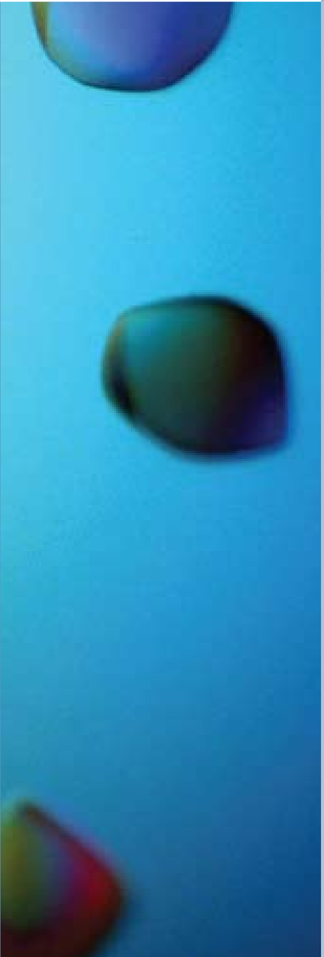

CUSTOM SHOP™ CRYSTALLIZATION REAGENTS

DESCRIPTION

The Hampton Research Custom Shop delivers ready-to-use crystallization reagents for screening, optimizing and producing crystals. The reagents are formulated at the time of order, using the same production protocol, chemicals, and quality control used when making the original Hampton Research crystallization kit.

Formulating the Custom Shop reagents at the time of order ensures consistency as well as saves time and offers convenience. They are shipped within 48 hours. Custom Shop reagents may not be returned for refund or exchange.

To order Custom Shop reagents, use the catalog numbers listed below. If you have any questions, please feel free to contact Customer Service.

When reproducing and optimizing reagent conditions, please be sure to also consider the Hampton Research Optimize™ and StockOptions™ reagents and kits.

Additive Screen - 1.0 ml **$65.00**
Catalog Number: HR2-428-**
** = reagent number 1-96
For example, reagent number A1 = HR2-428-01

Crystal Screen - 185 ml **$138.00**
Catalog Number: HR2-910-**
** = reagent number 1-50
For example, reagent number 1 = HR2-910-01

Crystal Screen 2 - 185 ml **$138.00**
Catalog Number: HR2-912-**
** = reagent number 1-48
For example, reagent number 1 = HR2-912-01

Crystal Screen Cryo - 185 ml **$138.00**
Catalog Number: HR2-914-**
** = reagent number 1-50
For example, reagent number 1 = HR2-914-01

Crystal Screen Lite - 185 ml **$138.00**
Catalog Number: HR2-916-**
** = reagent number 1-50
For example, reagent number 1 = HR2-916-01

Detergent Screen - 0.5 ml **$65.00**
Catalog Number: HR2-406-**
** = reagent number 1-96
For example, reagent number A1 = HR2-406-01

Grid Screen Ammonium Sulfate - 185 ml **$138.00**
Catalog Number: HR2-924-**
** = reagent number A1-D6
For example, reagent number A1 = HR2-924-A1

Grid Screen MPD - 185 ml **$138.00**
Catalog Number: HR2-930-**
** = reagent number A1-D6
For example, reagent number A1 = HR2-930-A1

Grid Screen PEG 6000 - 185 ml **$138.00**
Catalog Number: HR2-926-**
** = reagent number A1-D6
For example, reagent number A1 = HR2-926-A1

Grid Screen PEG/LiCl - 185 ml **$138.00**
Catalog Number: HR2-928-**
** = reagent number A1-D6
For example, reagent number A1 = HR2-928-A1

Grid Screen Sodium Chloride - 185 ml **$138.00**
Catalog Number: HR2-932-**
** = reagent number A1-D6
For example, reagent number A1 = HR2-932-A1

Grid Screen Sodium Malonate - 185 ml **$138.00**
Catalog Number: HR2-947-**
** = reagent number A1-D6
For example, reagent number A1 = HR2-947-A1

Heavy Atom Au (Gold) - 100 mg **$80.00**
Catalog Number: HR2-444-**
** = reagent number 1-6
For example, reagent number 1 = HR2-444-01

Heavy Atom Hg (Mercury) - 100 mg **$80.00**
Catalog Number: HR2-446-**
** = reagent number 1-16
For example, reagent number 1 = HR2-446-01

Heavy Atom M1 - 100 mg **$80.00**
Catalog Number: HR2-448-**
** = reagent number 1-19
For example, reagent number 1 = HR2-448-01

Heavy Atom M2 - 100 mg **$80.00**
Catalog Number: HR2-450-**
** = reagent number 1-18
For example, reagent number 1 = HR2-450-01

Heavy Atom Pt (Platinum) - 100 mg **$80.00**
Catalog Number: HR2-442-**
** = reagent number 1-12
For example, reagent number 1 = HR2-442-01

Index - 185 ml **$138.00**
Catalog Number: HR2-944-**
** = reagent number 1-96
For example, reagent number 1 = HR2-944-01

MembFac - 185 ml **$138.00**
Catalog Number: HR2-920-**
** = reagent number 1-48
For example, reagent number 1 = HR2-920-01

Natrix - 185 ml **$138.00**
Catalog Number: HR2-918-**
** = reagent number 1-48
For example, reagent number 1 = HR2-918-01

PCT - 185 ml **$138.00**
Catalog Number: HR2-940-**
** = reagent number A1-B2
For example, reagent number A1 = HR2-940-A1

PEG/Ion Screen - 185 ml **$138.00**
Catalog Number: HR2-922-**
** = reagent number 1-48
For example, reagent number 1 = HR2-922-01

PEG/Ion Screen 2 - 185 ml **$138.00**
Catalog Number: HR2-998-**
** = reagent number 1-48
For example, reagent number 1 = HR2-998-01

PEGRx 1 - 185 ml **$138.00**
Catalog Number: HR2-982-**
** = reagent number 1-48
For example, reagent number 1 = HR2-982-01

PEGRx 2 - 185 ml **$138.00**
Catalog Number: HR2-984-**
** = reagent number 1-48
For example, reagent number 1 = HR2-984-01

Quik Screen - 185 ml **$138.00**
Catalog Number: HR2-921-**
** = reagent number A1-D6
For example, reagent number A1 = HR2-921-A1

SaltRx 1 - 185 ml **$138.00**
Catalog Number: HR2-907-**
** = reagent number 1-48
For example, reagent number 1 = HR2-907-01

SaltRx 2 - 185 ml **$138.00**
Catalog Number: HR2-909-**
** = reagent number 1-48
For example, reagent number 1 = HR2-909-01

Silver Bullets - 0.25 ml **$65.00**
Catalog Number: HR2-996-**
** = reagent number 1-96
For example, reagent number A1 = HR2-996-01

Silver Bullets Bio - 0.25 ml **$65.00**
Catalog Number: HR2-988-**
** = reagent number 1-96
For example, reagent number A1 = HR2-988-01

StockOptions ADA - 185 ml **$208.00**

Catalog Number: HR2-997-**
** = reagent number 1-13
For example, reagent number 1 = HR2-998-01

StockOptions Bicine - 185 ml **$139.00**

Catalog Number: HR2-999-**
** = reagent number 1-15
For example, reagent number 1 = HR2-999-01

StockOptions Bis-Tris - 185 ml **$151.00**

Catalog Number: HR2-906-**
** = reagent number 1-21
For example, reagent number 1 = HR2-906-01

StockOptions Bis-Tris Propane - 185 ml **$201.00**

Catalog Number: HR2-993-**
** = reagent number 1-33
For example, reagent number 1 = HR2-993-01

StockOptions Citric Acid - 185 ml **$144.00**

Catalog Number: HR2-904-**
** = reagent number 1-44
For example, reagent number 1 = HR2-904-01

StockOptions HEPES - 185 ml **$172.00**

Catalog Number: HR2-902-**
** = reagent number 1-15
For example, reagent number 1 = HR2-902-01

StockOptions Imidazole - 185 ml **$162.00**

Catalog Number: HR2-995-**
** = reagent number 1-17
For example, reagent number 1 = HR2-995-01

StockOptions MES - 185 ml **$138.00**

Catalog Number: HR2-943-**
** = reagent number 1-20
For example, reagent number 1 = HR2-943-01

StockOptions pH - 185 ml

Catalog Number: HR2-941-**

Reagents 1-39	**$139.00**
Reagents 40-45	**$151.00**

** = reagent number 1-45
For example, reagent number 1 = HR2-941-01

StockOptions Sodium Acetate - 185 ml **$150.00**

Catalog Number: HR2-933-**
** = reagent number 1-21
For example, reagent number 1 = HR2-933-01

StockOptions Sodium Cacodylate - 185 ml **$221.00**

Catalog Number: HR2-939-**
** = reagent number 1-24
For example, reagent number 1 = HR2-939-01

StockOptions Sodium Citrate - 185 ml **$173.00**

Catalog Number: HR2-935-**
** = reagent number 1-24
For example, reagent number 1 = HR2-935-01

StockOptions Sodium HEPES - 185 ml **$190.00**

Catalog Number: HR2-931-**
** = reagent number 1-15
For example, reagent number 1 = HR2-931-01

StockOptions Tris - 185 ml **$169.00**

Catalog Number: HR2-900-**
** = reagent number 1-21
For example, reagent number 1 = HR2-900-01

StockOptions Tris Hydrochloride - 185 ml **$139.00**

Catalog Number: HR2-937-**
** = reagent number 1-21
For example, reagent number 1 = HR2-937-01

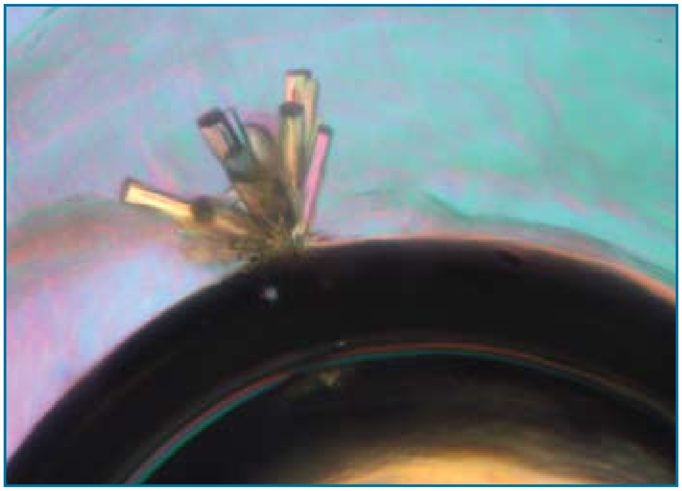

Crystals growing on a bubble.

Michelle Dechene, North Carolina State University, North Carolina, USA.

OPTIMIZE™ CRYSTALLIZATION GRADE REAGENTS

Protein crystal.

Alexey Rak, Max-Planck-Institut fur Molekulare Physiologie, Department of Physical Biochemistry, Dortmund, Germany.

TABLE OF CONTENTS

PAGES

40 - 42 OPTIMIZE™ - POLYMERS
42 OPTIMIZE™ - ORGANICS (VOLATILE)
42 OPTIMIZE™ - ORGANICS (NON-VOLATILE)
43 - 49 OPTIMIZE™ - SALTS
50 - 52 OPTIMIZE™ - BUFFERS
52 OPTIMIZE™ - SOLUBILIZING AGENTS (NDSB)
53 OPTIMIZE™ - REDUCING AGENT
54 OPTIMIZE™ - CRYOPROTECTANTS
55 - 56 OPTIMIZE™ - OILS
57 OPTIMIZE™ - SILICA HYDROGEL KIT
57 OPTIMIZE™ - LM AGAROSE
58 OPTIMIZE™ - IZIT CRYSTAL DYE™

HAMPTON
RESEARCH

APPLICATION

- Crystallization grade reagents for formulating screens or for optimization

FEATURES

- Quantitive formulation
- Lot to lot consistency
- Synergistic with Hampton Research screens & kits
- Quality and convenience
- Sterile filtered

DESCRIPTION

Optimize reagents are quantitatively formulated macromolecular crystallization grade reagents designed specifically for the crystallization of proteins, peptides, and nucleic acids. Each Optimize solution is formulated using high purity salts, polymers, and buffers. Sterile filtered Optimize reagents are formulated at convenient, ready to use concentrations. Optimize reagents remove the guesswork and make the process of reproducing preliminary screening conditions and general optimization faster, easier, and more convenient. When using Optimize reagents, a scientist moves directly from the screen to the optimization with no time wasted searching for and formulating salts, buffers, and viscous polymers. Each Optimize reagent is supplied with a Certificate of Analysis as documentation of reagent quality and performance. Optimize buffer reagents are supplied with calibrated titration curves to make titration easy, fast, and accurate.

POLYMERS

Ethylene imine polymer

HR2-599 50% solution 200 ml $78.00

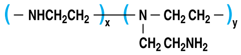
$\left(-NHCH_2CH_2-\right)_x\left(-\underset{CH_2CH_2NH_2}{N}-CH_2CH_2-\right)_y$

Synonyms: PEI or Poly(ethylene imine) or
Poly(ethyleneimine) solution or Ethyleneimine polymer solution
M_r 600,000 - 1,000,000 CAS [9002-98-6] EC No 205-793-9
Density: 1.08 g/mL at 20°C Measured pH Range: 12.0 - 12.1 at 25°C
Refractive Index n20/D: 1.450 Conductivity Range: 244.3 - 257.0 µS at 25°C

Jeffamine® ED-2001 pH 7.0

HR2-597 50% w/v solution pH 7.0 200 ml $98.00

Synonyms: O,O'-Bis(2-aminopropyl) polypropylene glycol-block-polyethylene glycol-block-polypropylene glycol 1,900 $CH_3CH(NH_2)CH_2(OCH_2CH_2)_nOCH_2CH(NH_2)CH_3$
M_r ~2,000 CAS [65605-36-9]
Refractive Index Range: 1.40637 - 1.40666 at 20°C
Conductivity Range: 4.0 - 5.7 mS at 25°C
pH before titration with Hydrochloric acid is > 10
Final measured pH of HR2-597 is 7.0

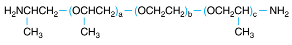
$H_2N\underset{CH_3}{C}HCH_2-(O\underset{CH_3}{C}HCH_2)_a-(OCH_2CH_2)_b-(OCH_2\underset{CH_3}{C}H)_c-NH_2$

$b \approx 38.7, (a+c) \approx 6.0$

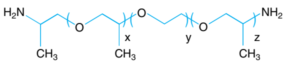

$y \approx 39, (x+z) \approx 6$

Jeffamine® M-600®

HR2-501 50% v/v solution pH 7.0 200 ml $94.00
HR2-503 100% solution (untitrated pH >12) 200 ml $94.00

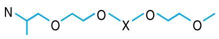

Synonyms: O-(2-Aminopropyl)-O'-(2-methoxyethyl)polypropylene glycol 500 or Polypropylene glycol 500 mono-2-aminoethyl mono-2-methoxyethyl ether $CH_3OCH_2CH_2O[CH(CH_3)CH_2O]_nCH_2CH(NH_2)$
M_r 600 CAS [77110-54-4] Density: 0.983 g/mL at 20°C

HR2-501: Refractive Index Range: 1.40754 - 1.40988 at 20°C
Conductivity Range: 8.8 - 10.3 mS at 25°C

HR2-503: Refractive Index Range: 1.44501 - 1.44546 at 20°C
Conductivity Range: 0.0 - 0.1 µS at 25°C

Pentaerythritol ethoxylate (3/4 EO/OH)

HR2-743 40% v/v solution 200 ml $46.00

Synonyms: Pentaerythritol ethoxylate (15/4 EO/OH)
$C(CH_2(OCH_2CH_2)_nOH)_2$ Average M_n ~270
CAS [30599-15-6] EC No 500-071-2
Density: 1.2 g/mL at 25°C (lit.)
Measured pH Range: 7.5 - 8.7 at 25°C
Refractive Index: 1.39719 at 20°C
Conductivity Range: 37.8 - 40.5 µS at 25°C

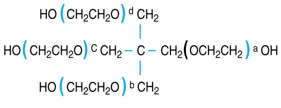

a + b + c + d ~ 3

Pentaerythritol ethoxylate (15/4 EO/OH)

HR2-745 50% v/v solution 200 ml $54.00

Synonyms: Pentaerythritol ethoxylate (3/4 EO/OH)
$C(CH_2(OCH_2CH_2)_nOH)_4$, N ~3.75 Average M_n ~797
CAS [30599-15-6] EC No 500-071-2
Density: 1.2 g/mL at 25°C (lit.)
Measured pH Range: 5.2 - 6.0 at 25°C
Refractive Index Range: 1.41401 - 1.41578 at 20°C
Conductivity Range: 4.2 - 11.4 µS at 25°C

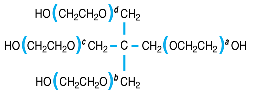

a + b + c + d ~ 15

Pentaerythritol propoxylate (5/4 PO/OH)

HR2-739 50% v/v solution 200 ml $54.00

Synonyms: Pentaerythritol propoxylate
$C[CH_2[OCH_2CH(CH_3)]_nOH]_4$, N ~5 Average M_n ~426
CAS [9051-49-4] EC No 500-030-9
Density of 100% solution: 1.05 g/mL at 25°C
Measured pH Range: 7.7 - 8.3 at 25°C
Refractive Index: 1.40424 - 1.40433 at 20°C
Conductivity Range: 3.8 - 4.7 µS at 25°C

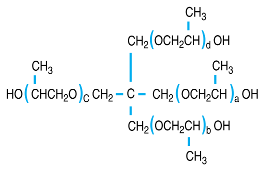

a + b + c + d ~ 5

Pentaerythritol propoxylate (17/8 PO/OH)

HR2-741 50% v/v solution 200 ml $58.00

Synonyms: Pentaerythritol propoxylate
$C(CH_2(OCH_2CH(CH_3))_nOH)_4$, N = 2.1 Average M_n ~629
CAS [9051-49-4] EC No 500-030-9
Density: 1.05 g/mL at 25°C (lit.)
Measured pH Range: 4.3 - 4.6 at 25°C
Refractive Index: 1.40418 at 20°C
Conductivity Range: 6.1 - 6.9 µS at 25°C

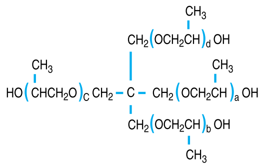

a + b + c + d ~ 5

Poly(acrylic acid sodium salt) 5,100

HR2-773 50% w/v solution 200 ml $95.00

Synonyms: Poly(sodium acrylate) or Sodium polyacrylate
$C_3H_3NaO_2$ M_r 94.04 CAS [9003-04-7]
Measured pH Range: 6.4 - 6.9 at 25°C

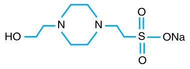

Refractive Index Range: 1.40716 - 1.40821 at 20°C Conductivity Range: 30.5 - 41.1 mS at 25°C

Polyethylene glycol 200

HR2-601 100% solution 200 ml $30.00

Synonyms: PEG 200 $H(OCH_2CH_2)_nOH$ M_r 190 - 210
CAS [25322-68-3] EC No 500-038-2
Density: 1.124 g/mL at 20°C
Measured pH Range: 8.0 - 10.2 at 25°C

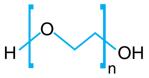

Refractive Index Range: 1.46008 - 1.46083 at 20°C Conductivity Range: 2.5 - 2.8 µS at 25°C

Polyethylene glycol 300

HR2-517	**100% solution**	**200 ml**	**$45.00**

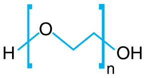

Synonyms: PEG 300 $H(OCH_2CH_2)_nOH$ M_r 285 - 315
CAS [25322-68-3] EC No 500-038-2 Density: 1.125 g/mL at 20°C
Measured pH Range: 8.0 - 8.5 at 25°C
Refractive Index Range: 1.46492 - 1.46535 at 20°C
Conductivity Range: 7.7 - 8.1 µS at 25°C

Peroxides (as H_2O_2)	≤ 0.001%	Co	≤ 0.0005%	Mo	≤ 0.0005%
Al	≤ 0.0005%	Cr	≤ 0.0005%	Na	≤ 0.2%
As	≤ 0.00001%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Bi	≤ 0.0005%	K	≤ 0.2%	SO_4	≤ 0.005%
Ca	≤ 0.001%	Li	≤ 0.0005%	Sr	≤ 0.0005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Zn	≤ 0.0005%
Cl	≤ 0.005%	Mn	≤ 0.0005%		

Polyethylene glycol 400

HR2-603	**100% solution**	**200 ml**	**$53.00**

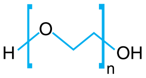

Synonyms: PEG 400 $H(OCH_2CH_2)_nOH$ M_r 380 - 420
CAS [25322-68-3] EC No 500-038-2 Density: 1.126 g/mL at 20°C
Measured pH Range: 9.1 - 9.7 at 25°C
Refractive Index Range: 1.46686 - 1.46761 at 20°C
Conductivity Range: 2.5 - 2.8 µS at 25°C

Polyethylene glycol 1,000

HR2-523	**50% w/v solution**	**200 ml**	**$90.00**

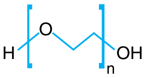

Synonyms: PEG 1,000 $H(OCH_2CH_2)_nOH$ M_r 950 - 1050
CAS [25322-68-3] EC No 500-038-2
Measured pH Range: 6.5 - 7.7 at 25°C
Refractive Index Range: 1.40078 - 1.40104 at 20°C
Conductivity Range: 23.1 - 37.5 µS at 25°C

Peroxides (as H_2O_2)	≤ 0.001%	Co	≤ 0.0005%	Mo	≤ 0.0005%
Al	≤ 0.0005%	Cr	≤ 0.0005%	Na	≤ 0.02%
As	≤ 0.00001%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Bi	≤ 0.0005%	K	≤ 0.02%	SO_4	≤ 0.005%
Ca	≤ 0.001%	Li	≤ 0.0005%	Sr	≤ 0.0005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Zn	≤ 0.0005%
Cl	≤ 0.005%	Mn	≤ 0.0005%		

Polyethylene glycol 1,500

HR2-525	**50% w/v solution**	**200 ml**	**$60.00**

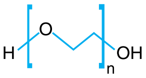

Synonyms: PEG 1,500 $H(OCH_2CH_2)_nOH$ M_r 1,400 - 1,600
CAS [25322-68-3] EC No 500-038-2
Measured pH Range: 6.1 - 8.7 at 25°C
Refractive Index Range: 1.40117 - 1.40127 at 20°C
Conductivity Range: 39.7 - 48.5 µS at 25°C

Polyethylene glycol 3,350 Monodisperse

HR2-591	**flake**	**500 g**	**$118.00**
HR2-527	**50% w/v solution**	**200 ml**	**$80.00**
HR2-837	**25% w/v PEG 3,350, 0.1 M MES pH 5.8**	**100 ml**	**$95.00**
HR2-841	**25% w/v PEG 3,350, 0.1 M HEPES Na pH 6.8**	**100 ml**	**$95.00**
HR2-845	**25% w/v PEG 3,350, 0.1 M BIS-TRIS pH 7.8**	**100 ml**	**$95.00**
HR2-849	**25% w/v PEG 3,350, 0.1 M BIS-TRIS propane pH 7.8**	**100 ml**	**$95.00**

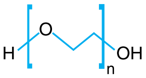

Synonyms: PEG 3,350 $H(OCH_2CH_2)_nOH$ M_r 3,300 - 3,400
CAS [25322-68-3] EC No 500-038-2

HR2-527: Measured pH Range: 7.9 - 9.7 at 25°C
Refractive Index Range: 1.40164 - 1.40208 at 20°C
Conductivity Range: 44.0 - 56.5 µS at 25°C

HR2-837: Measured Conductivity: 826 µS at 25°C

HR2-841: Measured Refractive Index: 1.37193 at 20°C
Measured Conductivity: 4.4 mS at 25°C

HR2-845: Measured Conductivity: 379 µS at 25°C

HR2-849: Measured Conductivity: 3.7 mS at 25°C

Polyethylene glycol 4,000

HR2-605	**flake**	**500 g**	**$80.00**
HR2-529	**50% w/v solution**	**200 ml**	**$80.00**

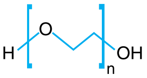

Synonyms: PEG 4,000 $H(OCH_2CH_2)_nOH$ M_r 3,500 - 4,500
CAS [25322-68-3] EC No 500-038-2
HR2-529: Measured pH Range: 6.9 - 9.5 at 25°C
Refractive Index Range: 1.40198 - 1.40279 at 20°C
Conductivity Range: 38.6 - 46.9 µS at 25°C

Polyethylene glycol 6,000

HR2-513	**flake**	**500 g**	**$122.00**
HR2-533	**50% w/v solution**	**200 ml**	**$80.00**

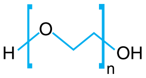

Synonyms: PEG 6,000 $H(OCH_2CH_2)_nOH$ M_r 5,000 - 7,000
CAS [25322-68-3] EC No 500-038-2
HR2-533: Measured pH Range: 6.2 - 9.1 at 25°C
Refractive Index Range: 1.40209 - 1.40249 at 20°C
Conductivity Range: 32.5 - 39.6 µS at 25°C

DNases	none detected	Ca	≤ 0.001%	Mg	≤ 0.0005%
RNases	none detected	Cd	≤ 0.0005%	Mn	≤ 0.0005%
Proteases	none detected	Cl	≤ 0.005%	Mo	≤ 0.0005%
Phosphatases	none detected	Co	≤ 0.0005%	Na	≤ 0.02%
Peroxides (H_2O_2)	≤ 0.001%	Cr	≤ 0.0005%	Ni	≤ 0.0005%
Al	≤ 0.0005%	Cu	≤ 0.0005%	Pb	≤ 0.0005%
As	≤ 0.00001%	Fe	≤ 0.0005%	SO_4	≤ 0.005%
Ba	≤ 0.0005%	K	≤ 0.02%	Sr	≤ 0.0005%
Bi	≤ 0.0005%	Li	≤ 0.0005%	Zn	≤ 0.0005%

Polyethylene glycol 8,000

HR2-515	**flake**	**500 g**	**$122.00**
HR2-535	**50% w/v solution**	**200 ml**	**$80.00**

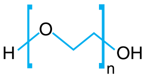

Synonyms: PEG 8,000 $H(OCH_2CH_2)_nOH$ M_r 7,000 - 9,000
CAS [25322-68-3] EC No 500-038-2
HR2-535: Measured pH Range: 6.6 - 9.5 at 25°C
Refractive Index Range: 1.40206 - 1.40261 at 20°C
Conductivity Range: 30.4 - 39.3 µS at 25°C

DNases	none detected	Ca	≤ 0.001%	Mg	≤ 0.0005%
RNases	none detected	Cd	≤ 0.0005%	Mn	≤ 0.0005%
Proteases	none detected	Cl	≤ 0.005%	Mo	≤ 0.0005%
Phosphatases	none detected	Co	≤ 0.0005%	Na	≤ 0.02%
Peroxides (as H_2O_2)	≤ 0.001%	Cr	≤ 0.0005%	Ni	≤ 0.0005%
Al	≤ 0.0005%	Cu	≤ 0.0005%	Pb	≤ 0.0005%
As	≤ 0.00001%	Fe	≤ 0.0005%	SO_4	≤ 0.005%
Ba	≤ 0.0005%	K	≤ 0.02%	Sr	≤ 0.0005%
Bi	≤ 0.0005%	Li	≤ 0.0005%	Zn	≤ 0.0005%

Polyethylene glycol 10,000

HR2-607	**50% w/v solution**	**200 ml**	**$80.00**

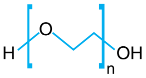

Synonyms: PEG 10,000 $H(OCH_2CH_2)_nOH$ M_r 8,500 - 11,500
CAS [25322-68-3] EC No 500-038-2
Measured pH Range: 6.4 - 8.6 at 25°C
Refractive Index Range: 1.40165 - 1.40254 at 20°C Conductivity Range: 39.4 - 48.3 µS at 25°C

Polyethylene glycol 20,000

HR2-609	**30% w/v solution**	**200 ml**	**$80.00**

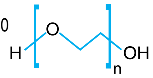

Synonyms: PEG 20,000 $H(OCH_2CH_2)_nOH$ Average M_n ~16,000 - 24,000
CAS [25322-68-3] EC No 500-038-2
Measured pH Range: 4.6 - 10.2 at 25°C
Refractive Index Range: 1.37438 - 1.37476 at 20°C Conductivity Range: 45.3 - 64.5 µS at 25°C

Polyethylene glycol monomethyl ether 550

HR2-611	**100% solution**	**200 ml**	**$85.00**

Synonyms: Methoxypolyethylene glycol or mono-Methyl polyethylene glycol or PEG MME 550
$CH_3O(CH_2CH_2O)_nH$ M_r 470 - 530 CAS [9004-74-4] EC No 215-801-2
Purity ≥ 100.0% Density: 1.089 g/mL at 25°C (lit.) Measured pH Range: 8.5 - 10.5 at 25°C
Refractive Index Range: 1.46144 - 1.46187 at 20°C Conductivity Range: 0.3 - 0.5 µS at 25°C

Polyethylene glycol monomethyl ether 2,000

HR2-613	**50% w/v solution**	**200 ml**	**$85.00**

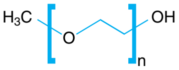

Synonyms: Methoxypolyethylene glycol 2,000 or
mono-Methyl polyethylene glycol 2,000 or PEG MME 2,000
$CH_3O(CH_2CH_2O)_nH$ CAS [9004-74-4] EC No 215-801-2
Purity ≥ 95.0% Density: 1.094 g/mL at 25°C (lit.) Measured pH Range: 7.0 - 8.8 at 25°C
Refractive Index Range: 1.40135 - 1.40161 at 20°C Conductivity Range: 28.0 - 60.2 mS at 25°C

POLYMERS

Polyethylene glycol monomethyl ether 5,000

HR2-615 50% w/v solution 200 ml $85.00

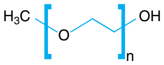

Synonyms: Methoxypolyethylene glycol 5,000 or mono-Methyl polyethylene glycol 5,000 or PEG MME 5,000
$CH_3O(CH_2CH_2O)_nH$ CAS [9004-74-4] EC No 215-801-2
Purity ≥ 95.0% Measured pH Range: 7.7 - 8.9 at 25°C
Refractive Index Range: 1.40199 - 1.40236 at 20°C Conductivity Range: 19.8 - 25.7 mS at 25°C

Polypropylene glycol P 400

HR2-771 100% solution 200 ml $30.00

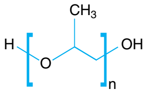

Synonyms: None
$(C_3H_6O)_n$ CAS [25322-69-4]
EC No 500-039-8 Density: 1.01 g/mL at 20°C
Measured pH Range: 3.8 - 8.5 at 25°C
Refractive Index Range: 1.44501 - 1.44827 at 20°C Conductivity Range: 0.0 - 0.1 mS at 25°C

Polyvinylpyrrolidone K 15

HR2-769 50% w/v solution 200 ml $45.00

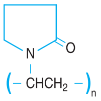

Synonyms: Polyvidone or Povidone or PVP or Poly(1-vinyl-2-pyrrolidinone) homopolymer
$(C_6H_9NO)_n$ Average M_r ~10,000 CAS [9003-39-8] EC No 201-800-4
Measured pH Range: 4.2 - 5.4 at 25°C
Refractive Index Range: 1.42196 - 1.42234 at 20°C Conductivity Range: 476 - 812 µS at 25°C

ORGANICS (VOLATILE)

1,4-Dioxane

HR2-617 100% solution 100 ml $48.00

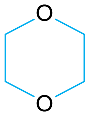

Synonyms: Diethylene oxide or Dioxane
$C_4H_8O_2$ M_r 88.11
CAS [123-91-1]
EC No 204-661-8 Purity ≥ 99.5%

Al ≤ 0.00005%
Ba ≤ 0.00001%
Bi ≤ 0.00001%
Ca ≤ 0.00005%
Cd ≤ 0.000005%
Co ≤ 0.000002%
Cr ≤ 0.000002%
Cu ≤ 0.000002%
Fe ≤ 0.00001%
K ≤ 0.00005%
Li ≤ 0.00001%
Mg ≤ 0.00001%
Mn ≤ 0.000002%
Mo ≤ 0.00001%
Na ≤ 0.00005%
Ni ≤ 0.000002%
Pb ≤ 0.00001%
Sr ≤ 0.00001%
Zn ≤ 0.00001%

2-Propanol

HR2-619 100% solution 200 ml $46.00

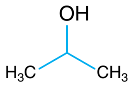

Synonyms: sec-Propyl alcohol or Isopropanol or Isopropyl alcohol
$(CH_3)_2CHOH$ or C_3H_8O M_r 60.10 CAS [67-63-0]
EC No 200-661-7 Purity ≥ 99.5% Density: 0.785 g/mL at 25°C (lit.)
Refractive Index Range: 1.37694 - 1.37748 at 20°C

DNases none detected
RNases none detected
Proteases none detected
Phosphatases none detected
Al ≤ 0.00005%
Ba ≤ 0.00001%
Bi ≤ 0.00001%
Ca ≤ 0.00005%
Cd ≤ 0.000005%
Co ≤ 0.000002%
Cr ≤ 0.000002%
Cu ≤ 0.000002%
Fe ≤ 0.00001%
K ≤ 0.00005%
Li ≤ 0.00001%
Mg ≤ 0.00001%
Mn ≤ 0.000002%
Mo ≤ 0.00001%
Na ≤ 0.00005%
Ni ≤ 0.000002%
Pb ≤ 0.00001%
Sr ≤ 0.00001%
Zn ≤ 0.00001%

ORGANICS (NON-VOLATILE)

Ethylene glycol

HR2-621 100% solution 100 ml $34.00

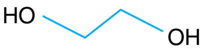

Synonyms: 1,2-Ethanediol
$C_2H_6O_2$ or $HOCH_2CH_2OH$ M_r 62.07
CAS [107-21-1] EC No 203-473-3
Density: 1.113 g/mL at 25°C (lit.) Purity ≥ 99.5%
Measured pH Range: 5.5 - 7.8 at 25°C Refractive Index Range: 1.43174 - 1.43207 at 20°C
Conductivity Range: 0.1 - 0.8 µS at 25°C

A ≤ 0.00005%
Ba ≤ 0.00001%
Bi ≤ 0.00001%
Ca ≤ 0.00005%
Cd ≤ 0.000005%
Cl ≤ 0.001%
Co ≤ 0.000002%
Cr ≤ 0.000002%
Cu ≤ 0.000002%
Fe ≤ 0.00005%
K ≤ 0.00005%
Li ≤ 0.00001%
Mg ≤ 0.00001%
Mn ≤ 0.000005%
Mo ≤ 0.00001%
Na ≤ 0.0001%
Ni ≤ 0.000002%
Pb ≤ 0.00001%
SO_4 ≤ 0.001%
Sr ≤ 0.00001%
Zn ≤ 0.00002%

Glycerol

HR2-623 100% solution 100 ml $61.00

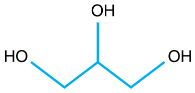

Synonyms: 1,2,3-Propanetriol or Glycerin
$HOCH_2CH(OH)CH_2OH$ or $C_3H_8O_3$ M_r 92.09
CAS [56-81-5] EC No 200-289-5 Density: 1.25 g/mL (lit.)
Purity ≥ 99.5% Measured pH Range: 6.9 - 7.4 at 25°C
Refractive Index Range: 1.47220 - 1.47412 at 20°C Conductivity Range: 0.0 µS at 25°C

DNases None detected
RNases None detected
Proteases None detected
Phosphatases None detected
Ag ≤ 0.0005%
Al ≤ 0.0001%
As ≤ 0.00001%
Ba ≤ 0.0001%
Bi ≤ 0.0001%
Ca ≤ 0.0005%
Cd ≤ 0.0001%
Cl ≤ 0.0001%
Co ≤ 0.0001%
Cr ≤ 0.0001%
Cu ≤ 0.0001%
Fe ≤ 0.0001%
K ≤ 0.002%
Li ≤ 0.0001%
Mg ≤ 0.0001%
Mn ≤ 0.0001%
Mo ≤ 0.0001%
NH_4^+ ≤ 0.0005%
Na ≤ 0.002%
Ni ≤ 0.0001%
Pb ≤ 0.0001%
SO_4 ≤ 0.001%
Sr ≤ 0.0001%
Tl ≤ 0.0005%
Zn ≤ 0.0001%

1,6-Hexanediol

HR2-625 6.0 M (71% w/v) solution 200 ml $52.00

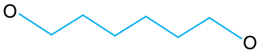

Synonyms: Hexamethylene glycol
$HO(CH_2)_6OH$ or $C_6H_{14}O_2$ M_r 118.18
CAS [629-11-8] EC No 211-074-0 Purity ≥ 97.0%
Measured pH Range: 5.3 - 8.2 at 25°C
Refractive Index Range: 1.42351 - 1.42386 at 20°C Conductivity Range: 0.1 - 1.5 µS at 25°C

(+/-)-2-Methyl-2,4-pentandiol

HR2-627 100% solution 200 ml $62.00

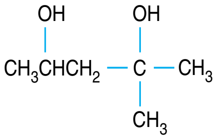

$$CH_3\overset{\overset{\displaystyle OH}{|}}{C}HCH_2-\underset{\underset{\displaystyle CH_3}{|}}{\overset{\overset{\displaystyle OH}{|}}{C}}-CH_3$$

Synonyms: MPD or Hexylene glycol
$CH_3CH(OH)CH_2C(CH_3)_2OH$ or $C_6H_{14}O_2$
M_r 118.18 CAS [107-41-5] EC No 203-489-0
Purity ≥ 99.0% Density: 0.925 g/mL at 25 °C (lit.)
Refractive Index Range: 1.42755 - 1.42787 at 20°C
Conductivity Range: 0.0 - 0.3 µS at 25°C

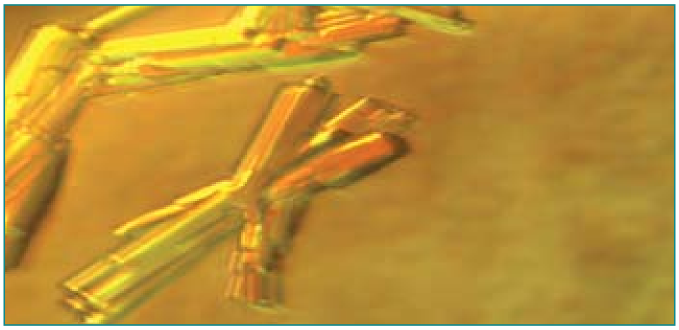

Protein crystals with a "wheat-sheaf" habit.

Jennifer Wingard, University of Maryland, USA.

SALTS

Ammonium acetate

HR2-565 1.0 M solution **100 ml** **\$70.00**
HR2-799 8.0 M solution **200 ml** **\$98.00**

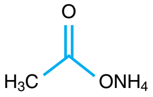

Synonyms: None $C_2H_7NO_2$ or $CH_3CO_2NH_4$
M_r 77.08 CAS [631-61-8] EC No 211-162-9 Purity $\geq$ 99.0%
HR2-565: Measured pH Range: 6.8 - 7.1 at 25°C
Refractive Index Range: 1.34382 - 1.34410 at 20°C
Conductivity Range: 62.0 - 70.5 mS at 25°C
HR2-799: Measured pH Range: 7.6 - 7.9 at 25°C
Refractive Index Range: 1.40811 - 1.40885 at 20°C
Conductivity Range: 68.9 - 80.0 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Na	≤ 0.005%
As	≤ 0.00001%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.0002%	NO_3	≤ 0.001%
Bi	≤ 0.0005%	K	≤ 0.005%	Pb	≤ 0.0005%
Ca	≤ 0.001%	Li	≤ 0.0005%	SO_4	≤ 0.001%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Cl	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%
Co	≤ 0.0005%	Mo	≤ 0.0005%		

Ammonium chloride

HR2-691 5.0 M solution **200 ml** **\$65.00**

Synonyms: Salmiac NH_4Cl M_r 53.49
CAS [12125-02-9] EC No 235-186-4 Purity $\geq$ 99.5%
Measured pH Range: 3.9 - 4.3 at 25°C Refractive Index Range: 1.38063 - 1.38075 at 20°C
Conductivity Range: 515.2 - 588.0 mS at 25°C

Al	≤ 0.0005%	Cu	≤ 0.0002%	Ni	≤ 0.0001%
As	≤ 0.00001%	Fe	≤ 0.0002%	NO_3	≤ 0.0005%
Ba	≤ 0.0005%	K	≤ 0.005%	Pb	≤ 0.0001%
Bi	≤ 0.0005%	Li	≤ 0.0005%	PO_4	≤ 0.0002%
Ca	≤ 0.001%	Mg	≤ 0.0005%	SO_4	≤ 0.002%
Cd	≤ 0.0005%	Mn	≤ 0.0005%	Sr	≤ 0.0005%
Co	≤ 0.0005%	Mo	≤ 0.0005%	Zn	≤ 0.0002%
Cr	≤ 0.0005%	Na	≤ 0.005%		

Ammonium citrate dibasic

HR2-685 2.5 M solution **200 ml** **\$95.00**

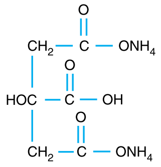

Synonyms: Citric acid ammonium salt or
Diammonium hydrogen citrate or Di-Ammnonium citrate
$C_6H_{14}N_2O_7$ or $(NH_4)_2C_6H_6O_7$ or $HOC(CO_2H)(CH_2CO_2NH_4)_2$
M_r 226.19 CAS [3012-65-5] EC No 221-146-3 Purity $\geq$ 99.0%
Measured pH Range: 4.7 - 4.8 at 25°C
Refractive Index: 1.42003 at 20°C
Conductivity Range: 76.8 - 81.9 µS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Na	≤ 0.005%
As	≤ 0.00001%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Bi	≤ 0.0005%	K	≤ 0.005%	PO_4	≤ 0.0005%
Ca	≤ 0.001%	Li	≤ 0.0005%	SO_4	≤ 0.005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Cl	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%
Co	≤ 0.0005%	Mo	≤ 0.0005%		

Ammonium citrate tribasic pH 7.0

HR2-759 pH 7.0 - 2.5 M solution **200 ml** **\$95.00**

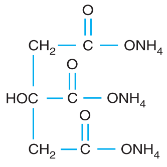

Synonyms: Citric acid triammonium salt
$HOC(CO_2NH_4)(CH_2CO_2NH_4)_2$ or $C_6H_{17}N_3O_7$
M_r 243.22 CAS [3458-72-8] EC No 222-394-5 Purity $\geq$ 97.0%
HR2-759 titrated to pH 7.0 at 25°C using Hydrochloric acid (HR2-581)
Refractive Index Range: 1.43010 - 1.43183 at 20°C
Conductivity Range: 90.4 - 100.8 mS at 25°C

Ammonium fluoride

HR2-689 10.0 M solution **200 ml** **\$71.00**

Synonyms: None NH_4F M_r 37.04 CAS [12125-01-8] EC No 235-185-9
Purity $\geq$ 98.0% Measured pH Range: 7.4 - 8.0 at 25°C Refractive Index: 1.42003 at 20°C
Conductivity Range: 350.8 - 428.8 mS at 25°C

Ag	≤ 0.0005%	Cr	≤ 0.0005%	Na	≤ 0.005%
Al	≤ 0.0005%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Bi	≤ 0.0005%	K	≤ 0.005%	SiF_6	≤ 0.1%
Ca	≤ 0.005%	Li	≤ 0.0005%	SO_4	≤ 0.005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Cl	≤ 0.0005%	Mn	≤ 0.0005%	Ti	≤ 0.0005%
Co	≤ 0.0005%	Mo	≤ 0.0005%	Zn	≤ 0.0005%

Ammonium formate

HR2-659 10.0 M solution **200 ml** **\$134.00**

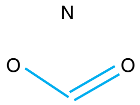

Synonyms: Formic acid ammonium salt
$HCOONH_4$ or CH_5NO_2 M_r 63.06
CAS [540-69-2] EC No 208-753-9 Purity $\geq$ 99.0%
Measured pH Range: 7.0 - 7.3 at 25°C
Refractive Index Range: 1.40348 - 1.40398 at 20°C
Conductivity Range: 579.6 - 706.0 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Na	≤ 0.005%
As	≤ 0.00001%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Bi	≤ 0.0005%	K	≤ 0.005%	SO_4	≤ 0.005%
Ca	≤ 0.001%	Li	≤ 0.0005%	Sr	≤ 0.0005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Zn	≤ 0.0005%
Cl	≤ 0.005%	Mn	≤ 0.0005%		
Co	≤ 0.0005%	Mo	≤ 0.0005%		

Ammonium nitrate

HR2-665 10.0 M solution **200 ml** **\$89.00**

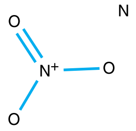

Synonyms: None NH_4NO_3 M_r 80.04
CAS [6484-52-2] EC No 229-347-8 Purity $\geq$ 99.5%
Measured pH Range: 3.8 - 5.6 at 25°C
Refractive Index Range: 1.41941 - 1.41957 at 20°C
Conductivity Range: 942 - 1036 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Na	≤ 0.005%
As	≤ 0.00001%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.0002%	NO_2	≤ 0.0005%
Bi	≤ 0.0005%	K	≤ 0.005%	Pb	≤ 0.0005%
Ca	≤ 0.001%	Li	≤ 0.0005%	PO_4	≤ 0.0005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	SO_4	≤ 0.002%
Cl	≤ 0.0003%	Mn	≤ 0.0005%	Sr	≤ 0.0005%
Co	≤ 0.0005%	Mo	≤ 0.0005%	Zn	≤ 0.0005%

Ammonium phosphate dibasic

HR2-629 3.5 M solution **200 ml** **\$75.00**

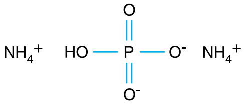

Synonyms: Ammonium hydrogenphosphate or
di-Ammonium hydrogen phosphate (sec) or
Diammonium hydrogen phosphate
$(NH_4)_2HPO_4$ or $H_9N_2O_4P$ M_r 132.06
CAS [7783-28-0] EC No 231-987-8 Purity $\geq$ 99.0%
Measured pH Range: 7.8 - 8.1 at 25°C Refractive Index Range: 1.40135 - 1.40362 at 20°C
Conductivity Range: 87.4 - 105.7 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Na	≤ 0.005%
As	≤ 0.00001%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	NO_3	≤ 0.001%
Bi	≤ 0.0005%	K	≤ 0.005%	Pb	≤ 0.0005%
Ca	≤ 0.001%	Li	≤ 0.0005%	SO_4	≤ 0.002%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Cl	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%
Co	≤ 0.0005%	Mo	≤ 0.0005%		

Ammonium phosphate monobasic

HR2-555 2.5 M solution **200 ml** **\$68.00**

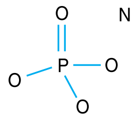

Synonyms: Ammonium dihydrogen phosphate or
Mono-ammonium phosphate or prim-Ammonium phosphate
$NH_4H_2PO_4$ M_r 115.03 CAS [7722-76-1] EC No 231-764-5
Purity $\geq$ 99.5% Measured pH Range: 3.5 - 3.8 at 25°C
Refractive Index Range: 1.36710 - 1.36745 at 20°C
Conductivity Range: 81.9 - 95.5 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Na	≤ 0.005%
As	≤ 0.00005%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.002%	Pb	≤ 0.0005%
Bi	≤ 0.0005%	K	≤ 0.005%	SO_4	≤ 0.005%
Ca	≤ 0.005%	Li	≤ 0.0005%	Sr	≤ 0.0005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Zn	≤ 0.0005%
Cl	≤ 0.005%	Mn	≤ 0.0005%		
Co	≤ 0.0005%	Mo	≤ 0.0005%		

SALTS

Ammonium sulfate

HR2-541 3.5 M solution **200 ml** **$68.00**

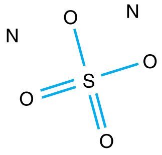

Synonyms: Ammonium sulphate
$(NH_4)_2SO_4$ or $H_8N_2O_4S$ M_r 132.14
CAS [7783-20-2] EC No 231-984-1 Purity ≥ 99.5%
Measured pH Range: 5.0 - 5.2 at 25°C
Refractive Index Range: 1.39131 - 1.39166 at 20°C
Conductivity Range: 313.2 - 424.2 µS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Na	≤ 0.005%
As	≤ 0.00002%	Cu	≤ 0.0002%	Ni	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.0002%	NO_3	≤ 0.001%
Bi	≤ 0.0005%	K	≤ 0.005%	Pb	≤ 0.0002%
Ca	≤ 0.001%	Li	≤ 0.0005%	PO_4	≤ 0.0005%
Cd	≤ 0.0001%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Cl	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.0001%
Co	≤ 0.0005%	Mo	≤ 0.0005%		

Ammonium tartrate dibasic

HR2-679 2.0 M solution **200 ml** **$75.00**
HR2-767 pH 7.0 - 1.6 M solution **200 ml** **$94.00**

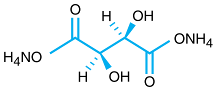

Synonyms: L-(+)-Tartaric acid diammonium salt or
Diammonium tartrate $(NH_4)_2C_4H_4O_6$ or $C_4H_{12}N_2O_6$ M_r 184.15
CAS [3164-29-2] EC No 221-618-9 Purity ≥ 99.5%
HR2-679: Measured pH Range: 6.4 - 6.8 at 25°C
Refractive Index Range: 1.39234 - 1.39254 at 20°C
Conductivity Range: 118.2 - 134.4 mS at 25°C
HR2-767: Titrated to pH 7.0 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.38139 - 1.38147 at 20°C
Conductivity Range: 115.7 - 128.5 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Na	≤ 0.005%
As	≤ 0.00001%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Bi	≤ 0.0005%	K	≤ 0.005%	SO_4	≤ 0.005%
Ca	≤ 0.001%	Li	≤ 0.0005%	Sr	≤ 0.0005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Zn	≤ 0.0005%
Cl	≤ 0.005%	Mn	≤ 0.0005%		
Co	≤ 0.0005%	Mo	≤ 0.0005%		

Cadmium chloride hydrate

HR2-715 1.0 M solution **100 ml** **$98.00**

Synonyms: None $CdCl_2 \cdot xH_2O$ M_r 183.32 (anhyd.) CAS [654054-66-7] EC No 233-296-7
Purity 98.0% Density: 3.327 g/mL at 25°C (lit.) Measured pH Range: 3.3 - 5.0 at 25°C
Refractive Index Range: 1.35455 - 1.35477 at 20°C Conductivity Range: 26.9 - 28.9 mS at 25°C

Cadmium sulfate hydrate

HR2-721 1.0 M solution **100 ml** **$88.00**

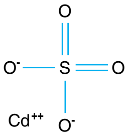

Synonyms: Cadmium sulfate octahydrate $3CdSO_4 \cdot 8H_2O$ M_r 769.52
CAS [7790-84-3] EC No 233-331-6
Purity ≥ 99.0% (calc. based on $CdSO_4 \cdot 8/3\ H_2O$, KT)
Measured pH Range: 3.5 - 4.4 at 25°C
Refractive Index: 1.40310 at 20°C
Conductivity Range: 40.5 - 45.3 mS at 25°C

As	≤ 0.00001%	Cu	≤ 0.0005%	N	≤ 0.0005%
Ca	≤ 0.005%	Fe	≤ 0.0005%	Na	≤ 0.005%
Cl	≤ 0.001%	K	≤ 0.005%	Ni	≤ 0.0005%
Co	≤ 0.0005%	Mg	≤ 0.0005%	Pb	≤ 0.002%
Cr	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.005%

Calcium acetate hydrate

HR2-567 1.0 M solution **100 ml** **$68.00**

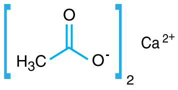

Synonyms: None $(CH_3COO)_2Ca \cdot xH_2O$ $C_4H_6CaO_4 \cdot xH_2O$
M_r 158.17 (anhyd.) CAS [114460-21-8] EC No 200-540-9
Purity ≥ 99.0% Measured pH Range: 7.6 - 7.9 at 25°C
Refractive Index Range: 1.35719 - 1.35722 at 20°C
Conductivity Range: 34.1 - 38.1 mS at 25°C

As	≤ 0.00001%	Cu	≤ 0.0005%	Na	≤ 0.01%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	Ni	≤ 0.0005%
Bi	≤ 0.0005%	K	≤ 0.01%	Pb	≤ 0.0005%
Cd	≤ 0.0005%	Li	≤ 00005%	SO_4	≤ 0.01%
Cl	≤ 0.005%	Mg	≤ 0.1%	Sr	≤ 0.02%
Co	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%
Cr	≤ 0.0005%	Mo	≤ 0.0005%		

Calcium chloride dihydrate

HR2-557 2.0 M solution **100 ml** **$74.00**

Synonyms: None $CaCl_2 \cdot 2H_2O$ M_r 147.01
CAS [10035-04-8] EC No 233-1408 Purity ≥ 99.5%
Measured pH Range: 6.1 - 7.9 at 25°C Refractive Index Range: 1.38108 - 1.38195 at 20°C
Conductivity Range: 170.0 - 196.9 mS at 25°C

As	≤ 0.00001%	Fe	≤ 0.0005%	Na	≤ 0.01%
Ba	≤ 0.0005%	K	≤ 0.005%	Ni	≤ 0.0005%
Bi	≤ 0.0005%	Li	≤ 0.0005%	Pb	≤ 0.0005%
Cd	≤ 0.0005%	Mg	≤ 0.005%	SO_4	≤ 0.005%
Co	≤ 0.0005%	Mn	≤ 0.0005%	Sr	≤ 0.01%
Cr	≤ 0.0005%	Mo	≤ 0.0005%	Zn	≤ 0.0005%
Cu	≤ 0.0005%	N	≤ 0.002%		

Cesium chloride

HR2-719 1.0 M solution **100 ml** **$128.00**

Synonyms: None CsCl M_r 168.36 CAS [7647-17-8] EC No 231-600-2
Purity ≥ 99.0% Measured pH Range: 6.0 - 6.5 at 25°C
Refractive Index: 1.34591 at 20°C
Conductivity Range: 106.5 - 112.9 mS at 25°C

Al	≤ 0.0005%	Cu	≤ 0.0005%	Na	≤ 0.005%
As	≤ 0.00001%	Fe	≤ 0.0005%	Ni	≤ 0.0005%
Ba	≤ 0.0005%	K	≤ 0.005%	Pb	≤ 0.0005%
Bi	≤ 0.0005%	Li	≤ 0.0005%	SO_4	≤ 0.002%
Ca	≤ 0.005%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Cd	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%
Co	≤ 0.0005%	Mo	≤ 0.0005%		
Cr	≤ 0.0005%	N	≤ 0.001%		

Cobalt(II) chloride hexahydrate

HR2-713 1.0 M solution **100 ml** **$64.00**

Synonyms: Cobaltous chloride hexahydrate $Cl_2Co_2 \cdot 6H_2O$ M_r 237.93
CAS [7791-13-1] EC No 231-589-4 Purity ≥ 98.0%
Measured pH Range: 2.1 - 3.6 at 25°C
Conductivity Range: 103.4 - 114.8 mS at 25°C

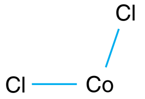

Hexadecyltrimethylammonium bromide

HR2-711 0.05 M solution **200 ml** **$61.00**

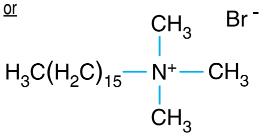

$H_3C(H_2C)_{15}-N^+(CH_3)_3\ Br^-$

Synonyms: CTAB or cetrimide or cetrimonium bromide or
cetyltrimethylammonium bromide or
palmityltrimethylammonium bromide
$C_{19}H_{42}BrN$ or $CH_3(CH_2)_{15}N(Br)(CH_3)_3$ M_r 364.45
CAS [57-09-0] EC No 200-311-3 Purity ≥ 99.0%
Measured pH Range: 5.3 - 5.9 at 25°C
Refractive Index: 1.33573 at 20°C
Conductivity Range: 1109 - 1344 µS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Mo	≤ 0.0005%
As	≤ 0.00005%	Cu	≤ 0.0005%	Na	≤ 0.005%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	Ni	≤ 0.0005%
Bi	≤ 0.0005%	K	≤ 0.005%	Pb	≤ 0.0005%
Ca	≤ 0.001%	Li	≤ 0.0005%	SO_4	≤ 0.005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Co	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%

Iron(III) chloride hexahydrate

HR2-717 1.0 M solution **100 ml** **$58.00**

Synonyms: Ferric chloride hexahydrate
$FeCl_3 \cdot 6H_2O$ or $Cl_3Fe \cdot 6H_2O$ M_r 270.30 CAS [10025-77-1] EC No 231-729-4
Measured pH Range: 0.8 - 1.0 at 25°C Refractive Index: 1.37469 at 20°C
Conductivity Range: 90.0 - 94.8 mS at 25°C

As	≤ 0.0005%	Fe_2^+	≤ 0.002%	NO_3	≤ 0.01%
Ca	≤ 0.005%	K	≤ 0.005%	Pb	≤ 0.002%
Cd	≤ 0.001%	Mg	≤ 0.001%	PO_4	≤ 0.01%
Cl	≤ 0.002%	Mn	≤ 0.1%	SO_4	≤ 0.005%
Co	≤ 0.005%	N	≤ 0.001%	Zn	≤ 0.002%
Cr	≤ 0.01%	Na	≤ 0.01%		
Cu	≤ 0.002%	Ni	≤ 0.005%		

SALTS

Lithium acetate dihydrate

HR2-669 5.0 M solution 200 ml $93.00

Synonyms: Acetic acid lithium salt $CH_3COOLi \cdot 2H_2O$

M_r 102.02 CAS [6108-17-4] EC No 208-914-3

Measured pH Range: 8.3 - 9.2 at 25°C

Refractive Index: 1.38124 at 20°C

Conductivity Range: 24.7 - 26.5 mS at 25°C

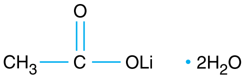

$CH_3-C(=O)-OLi \cdot 2H_2O$

Al	≤ 0.0005%	K	≤ 0.005%	Pb	≤ 0.001%
Ca	≤ 0.005%	Mg	≤ 0.001%	SO_4	≤ 0.05%
Cl	≤ 0.5%	NH_4^+	≤ 0.05%	Zn	≤ 0.0005%
Cu	≤ 0.0005%	Na	≤ 0.005%		
Fe	≤ 0.0005%	P	≤ 0.0005%		

Lithium chloride

HR2-631 10.0 M solution 200 ml $124.00

Synonyms: Lithium chloride anhydrous

LiCl M_r 42.39 CAS [7447-41-8] EC No 231-212-3 Purity ≥ 99.0%

Measured pH Range: 5.7 - 6.6 at 25°C

Refractive Index: 1.41030 at 20°C

Conductivity Range: 107.4 - 132.9 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	N	≤ 0.001%
As	≤ 0.00001%	Cu	≤ 0.0005%	Na	≤ 0.005%
Ba	≤ 0.001%	Fe	≤ 0.0005%	Ni	≤ 0.001%
Bi	≤ 0.0005%	K	≤ 0.005%	Pb	≤ 0.0005%
Ca	≤ 0.001%	Mg	≤ 0.0005%	SO_4	≤ 0.005%
Cd	≤ 0.0005%	Mn	≤ 0.0005%	Sr	≤ 0.0005%
Co	≤ 0.0005%	Mo	≤ 0.0005%	Zn	≤ 0.0005%

Lithium citrate tribasic tetrahydrate

HR2-681 1.5 M solution 200 ml $98.00

Synonyms: Citric acid trilithium salt or Trilithium citrate tetrahydrate

$HOC(COOLi)(CH_2COOLi)_2 \cdot 4H_2O$ or $C_6H_5Li_3O_7 \cdot 4H_2O$

M_r 281.99 CAS [6080-58-6] EC No 213-045-8 Purity ≥ 99.5%

Measured pH Range: 8.4 - 9.3 at 25°C

Refractive Index Range: 1.38910 - 1.38970 at 20°C

Conductivity Range: 30.9 - 32.5 mS at 25°C

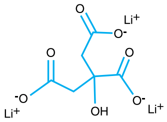

Al	≤ 0.0005%	Co	≤ 0.0005%	Mo	≤ 0.0005%
As	≤ 0.00001%	Cr	≤ 0.0005%	Na	≤ 0.005%
Ba	≤ 0.0005%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Bi	≤ 0.0005%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Ca	≤ 0.001%	K	≤ 0.005%	SO_4	≤ 0.005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Cl	≤ 0.005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%

Lithium nitrate

HR2-697 8.0 M solution 200 ml $85.00

Synonyms: None $LiNO_3$ M_r 68.95

CAS [7790-69-4] EC No 232-218-9 Purity ≥ 99.0%

Measured pH Range: 7.0 - 7.8 at 25°C

Refractive Index: 1.39391 at 20°C

Conductivity Range: 139.2 - 151.5 mS at 25°C

Al	≤ 0.001%	Co	≤ 0.0005%	Mo	≤ 0.0005%
As	≤ 0.00001%	Cr	≤ 0.0005%	Na	≤ 0.02%
Ba	≤ 0.001%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Bi	≤ 0.0005%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Ca	≤ 0.01%	K	≤ 0.02%	SO_4	≤ 0.01%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Cl	≤ 0.005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%

Lithium sulfate monohydrate

HR2-545 2.0 M solution 200 ml $98.00

Synonyms: Lithium sulfate $Li_2SO_4 \cdot H_2O$ or $Li_2O_4S \cdot H_2O$

M_r 127.96 CAS [10102-25-7] EC No 233-820-4 Purity ≥ 99.0%

Measured pH Range: 3.0 - 4.3 at 25°C

Refractive Index Range: 1.36520 - 1.36553 at 20°C

Conductivity Range: 76.4 - 86.5 mS at 25°C

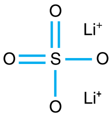

Al	≤ 0.0005%	Co	≤ 0.0005%	Mo	≤ 0.0005%
As	≤ 0.00001%	Cr	≤ 0.0005%	Na	≤ 0.005%
Ba	≤ 0.0005%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Bi	≤ 0.0005%	Fe	≤ 0.0005%	NO_3	≤ 0.001%
Ca	≤ 0.001%	K	≤ 0.005%	Pb	≤ 0.0005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Cl	≤ 0.005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%

Magnesium acetate tetrahydrate

HR2-561 1.0 M solution 100 ml $81.00

Synonyms: Acetic acid magnesium salt

$Mg(CH_3COO)_2Mg \cdot 4H_2O$ or $C_4H_6MgO_4 \cdot 4H_2O$

M_r 214.45 CAS [16674-78-5] EC No 205-554-9

Purity ≥ 99.0% Measured pH Range: 8.1 - 8.4 at 25°C

Refractive Index Range: 1.35767 - 1.35799 at 20°C

Conductivity Range: 32.4 - 36.8 mS at 25°C

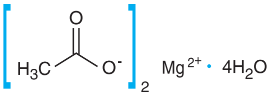

Al	≤ 0.0005%	Co	≤ 0.0005%	Mo	≤ 0.0005%
As	≤ 0.00001%	Cr	≤ 0.0005%	Na	≤ 0.005%
Ba	≤ 0.0005%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Bi	≤ 0.001%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Ca	≤ 0.002%	K	≤ 0.005%	SO_4	≤ 0.02%
Cd	≤ 0.0005%	Li	≤ 0.0005%	Sr	≤ 0.0005%
Cl	≤ 0.005%	Mn	≤ 0.002%	Zn	≤ 0.0005%

Magnesium chloride hexahydrate

HR2-559 2.0 M solution 100 ml $76.00
HR2-803 5.0 M solution 200 ml $98.00

Synonyms: None $MgCl_2 \cdot 6H_2O$ M_r 203.30

CAS [7791-18-6] EC No 232-094-6 Purity ≥ 99.0%

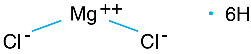

HR2-559: Measured pH Range: 5.1 - 6.0 at 25°C
Refractive Index Range: 1.37544 - 1.37603 at 20°C
Conductivity Range: 136.2 - 160.2 mS at 25°C

HR2-803: Measured pH Range: 3.7 - 5.7 at 25°C Refractive Index: 1.43079 at 20°C
Conductivity Range: 60.0 - 84.5 mS at 25°C

Al	≤ 0.0005%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
As	≤ 0.00001%	Fe	≤ 0.0005%	PO_4	≤ 0.0005%
Ba	≤ 0.0005%	K	≤ 0.2%	Pb	≤ 0.0005%
Bi	≤ 0.0005%	Li	≤ 0.0005%	SO_4	≤ 0.005%
Ca	≤ 0.005%	Mn	≤ 0.0005%	Sr	≤ 0.0005%
Cd	≤ 0.0005%	Mo	≤ 0.0005%	Zn	≤ 0.0005%
Co	≤ 0.0005%	N	≤ 0.0002%		
Cr	≤ 0.0005%	Na	≤ 0.3%		

Magnesium formate dihydrate

HR2-537 1.0 M solution 200 ml $82.00

Synonyms: Diformic acid magnesium salt $C_2H_2O_4Mg \cdot 2H_2O$

M_r 150.38 CAS [557-39-1] EC No 209-173-9 Purity 98.5%

Measured pH Range: 7.6 - 7.8 at 25°C

Refractive Index Range: 1.35214 - 1.35241 at 20°C

Conductivity Range: 45.6 - 50.3 mS at 25°C

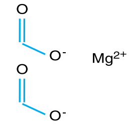

Magnesium nitrate hexahydrate

HR2-657 3.0 M solution 200 ml $83.00

Synonyms: Magnesium nitrate

$Mg(NO_3)_2 \cdot 6H_2O$ $MgN_2O_6 \cdot 6H_2O$ M_r 256.41

CAS [13446-18-9] EC No 233-826-7 Purity ≥ 99.0%

Measured pH Range: 2.7 - 3.8 at 25°C

Refractive Index Range: 1.39348 - 1.39373 at 20°C

Conductivity Range: 112.4 - 136.1 mS at 25°C

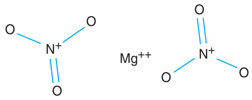

Al	≤ 0.0005%	Co	≤ 0.0005%	Mo	≤ 0.0005%
As	≤ 0.00001%	Cr	≤ 0.0005%	Na	≤ 0.01%
Ba	≤ 0.001%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Bi	≤ 0.001%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Ca	≤ 0.5%	K	≤ 0.01%	SO_4	≤ 0.005%
Cd	≤ 0.0005%	Li	≤ 0.0005%	Sr	≤ 0.001%
Cl	≤ 0.005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%

Magnesium sulfate heptahydrate

HR2-821 2.0 M solution 200 ml $102.00

Synonyms: Epsom salts $MgSO_4 \cdot 7H_2O$ M_r 246.47

CAS [10034-99-8] EC No 231-298-2 Purity ≥ 99.5%

Measured pH: 5.6 at 25°C

Conductivity Range: 51.8 mS at 25°C

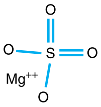

Al	≤ 0.0005%	Co	≤ 0.0005%	Mo	≤ 0.0005%
As	≤ 0.00001%	Cr	≤ 0.0005%	N	≤ 0.002%
Ba	≤ 0.0005%	Cu	≤ 0.0005%	Na	≤ 0.001%
Bi	≤ 0.0005%	Fe	≤ 0.0001%	Ni	≤ 0.0005%
Ca	≤ 0.001%	K	≤ 0.001%	Pb	≤ 0.0005%
Cd	≤ 0.0005%	Li	≤ 0.0005%	Sr	≤ 0.0005%
Cl	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%

SALTS

Magnesium sulfate hydrate

HR2-633 2.5 M solution **200 ml** **$83.00**

Synonyms: Magnesium sulphate hydrate $MgSO_4 \cdot xH_2O$
M_r 120.37 (anhyd.) CAS [22189-08-8] EC No 231-298-2 Purity ≥ 99.0%
Measured pH Range: 6.2 - 8.2 at 25°C
Refractive Index: 1.37187 at 20°C
Conductivity Range: 49.2 - 59.5 mS at 25°C

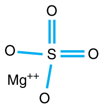

Ca	≤ 0.02%	Cu	≤ 0.005%	Ni	≤ 0.005%
Cd	≤ 0.005%	Fe	≤ 0.005%	Pb	≤ 0.005%
Cl	≤ 0.01%	K	≤ 0.05%	Zn	≤ 0.005%
Co	≤ 0.005%	Na	≤ 0.01%		

DL-Malic acid pH 7.0

HR2-761 pH 7.0 - 3.0 M solution **200 ml** **$62.00**

Synonyms: (±)-2-Hydroxysuccinic acid or DL-Hydroxybutanedioic acid
$HO_2CCH_2CH(OH)CO_2H$ or $C_4H_6O_5$
M_r 134.09 CAS [6915-15-7] EC No 230-022-8 Purity ≥ 99.0%
Titrated to pH 7.0 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.40605 - 1.40638 at 20°C
Conductivity Range: 43.0 - 52.1 mS at 25°C

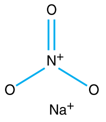

Nickel(II) chloride hexahydrate

HR2-687 4.0 M solution **200 ml** **$69.00**

Synonyms: Nickel chloride $NiCl_2 \cdot 6H_2O$ M_r 237.69
CAS [7791-20-0] EC No 231-743-0 Purity ≥ 98.0%
Measured pH Range: 3.1 - 3.8 at 25°C
Refractive Index Range: 1.43878 - 1.43890 at 20°C
Conductivity Range: 82.2 - 104.6 mS at 25°C

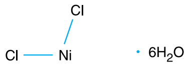

Ca	≤ 0.005%	Cu	≤ 0.005%	Pb	≤ 0.005%
Cd	≤ 0.005%	Fe	≤ 0.005%	Zn	≤ 0.005%
Co	≤ 0.05%	K	≤ 0.01%		

Potassium acetate

HR2-671 5.0 M solution **200 ml** **$85.00**

Synonyms: K(acac) $C_2H_3KO_2$ or CH_3COOK M_r 98.14
CAS [127-08-2] EC No 204-822-2 Purity ≥ 99.0%
Measured pH Range: 7.9 - 8.9 at 25°C
Refractive Index Range: N/A
Conductivity Range: 132.3 - 146.7 mS at 25°C

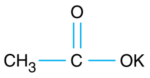
$CH_3-\overset{O}{\overset{\|}{C}}-OK$

DNases	none detected	Proteases	none detected	Pb	≤ 5ppm (parts per million)
RNases	none detected	Phosphatases	none detected		

Potassium bromide

HR2-779 4.0 M solution **100 ml** **$74.00**

Synonyms: None KBr or BrK M_r 119.00
CAS [7758-02-3] EC No 231-830-3 Purity ≥ 99.5%
Measured pH Range: 8.2 - 8.6 at 25°C
Refractive Index Range: 1.38454 - 1.38686 at 20°C
Conductivity Range: 384.6 - 438.8 mS at 25°C

Al	≤ 0.0005%	Co	≤ 0.0005%	Mo	≤ 0.0005%
As	≤ 0.00001%	Cr	≤ 0.0005%	N	≤ 0.001%
Ba	≤ 0.0005%	Cu	≤ 0.0005%	Na	≤ 0.02%
Bi	≤ 0.0005%	Fe	≤ 0.0005%	Ni	≤ 0.0005%
BrO_3	≤ 0.001%	I	≤ 0.001%	Pb	≤ 0.0005%
Ca	≤ 0.001%	Li	≤ 0.0005%	SO_4	≤ 0.005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Cl	≤ 0.1%	Mn	≤ 0.0005%	Zn	≤ 0.0005%

Potassium chloride

HR2-649 4.0 M solution **200 ml** **$80.00**

Synonyms: None KCl M_r 74.55
CAS [7447-40-7] EC No 231-211-8 Purity ≥ 99.5%
Measured pH Range: 5.4 - 8.1 at 25°C Refractive Index Range: 1.36868 - 1.36882 at 20°C
Conductivity Range: 391.0 - 449.6 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	N	≤ 0.001%
As	≤ 0.00001%	Cu	≤ 0.0005%	Na	≤ 0.02%
Ba	≤ 0.001%	Fe	≤ 0.0002%	Ni	≤ 0.0005%
Bi	≤ 0.0005%	I	≤ 0.002%	Pb	≤ 0.0005%
Br	≤ 0.05%	Li	≤ 0.0005%	PO_4	≤ 0.0005%
Ca	≤ 0.005%	Mg	≤ 0.001%	SO_4	≤ 0.003%
Cd	≤ 0.0005%	Mn	≤ 0.0005%	Sr	≤ 0.0005%
Co	≤ 0.0005%	Mo	≤ 0.0005%	Zn	≤ 0.0005%

Potassium citrate tribasic monohydrate

HR2-683 2.5 M solution **200 ml** **$74.00**

Synonyms: Citric acid tripotassium salt or tri-potassium citrate
$HOC(COOK)(CH_2COOK)_2 \cdot H_2O$ or $C_6H_5K_3O_7 \cdot H_2O$
M_r 324.41 CAS [6100-05-6] EC No 231-905-0 Purity ≥ 98.0%
Measured pH Range: 8.2 - 9.9 at 25°C
Refractive Index: 1.42503 at 20°C
Conductivity Range: 80.6 - 91.7 mS at 25°C

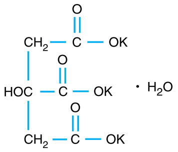

Potassium fluoride

HR2-647 6.0 M solution **100 ml** **$68.00**

Synonyms: None KF or FK M_r 58.10
CAS [7789-23-3] EC No 232-151-5 Purity ≥ 99.5%
Measured pH Range: 8.0 - 10.0 at 25°C
Refractive Index: 1.35344 at 20°C
Conductivity Range: 346.6 - 416.2 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Mn	≤ 0.0005%
As	≤ 0.00001%	Cu	≤ 0.0005%	Mo	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	Na	≤ 0.2%
Bi	≤ 0.0005%	HF	≤ 0.001%	Ni	≤ 0.0005%
Ca	≤ 0.001%	KOH	≤ 0.01%	Pb	≤ 0.0005%
Cd	≤ 0.0005%	K_2SiF_6	≤ 0.003%	SO_4	≤ 0.005%
Cl	≤ 0.005%	Li	≤ 0.0005%	Sr	≤ 0.0005%
Co	≤ 0.0005%	Mg	≤ 0.0005%	Zn	≤ 0.0005%

Potassium formate

HR2-667 14.0 M solution **200 ml** **$118.00**

Synonyms: Formic acid potassium salt HCOOK or $CHKO_2$
M_r 84.12 CAS [590-29-4] EC No 209-677-9 Purity ≥ 99.0%
Measured pH Range: 9.5 - 10.8 at 25°C
Refractive Index: 1.41516 at 20°C
Conductivity Range: 77.9 - 99.4 mS at 25°C

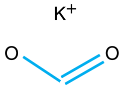

Ca	≤ 0.005%	Cu	≤ 0.005%	Pb	≤ 0.005%
Cd	≤ 0.005%	Fe	≤ 0.005%	SO_4	≤ 0.005%
Cl	≤ 0.005%	Na	≤ 0.5%	Zn	≤ 0.005%
Co	≤ 0.005%	Ni	≤ 0.005%		

Potassium phosphate dibasic

HR2-635 4.0 M solution **200 ml** **$74.00**

Synonyms: Dipotassium hydrogenphosphate or
Dipotassium phosphate or sec.-Potassium phosphate
K_2HPO_4 or HK_2O_4P M_r 174.18 CAS [7758-11-4] EC No 231-834-5
Measured pH Range: 9.6 - 9.8 at 25°C
Refractive Index Range: 1.40781 - 1.40791 at 20°C
Conductivity Range: 126.9 - 148.5 mS at 25°C

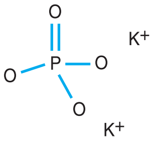

Al	≤ 0.0005%	Co	≤ 0.0005%	Mo	≤ 0.0005%
As	≤ 0.00001%	Cr	≤ 0.0005%	Na	≤ 0.5%
Ba	≤ 0.0005%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Bi	≤ 0.0005%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Ca	≤ 0.001%	Li	≤ 0.0005%	SO_4	≤ 0.005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Cl	≤ 0.005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%

SALTS

Potassium phosphate monobasic

HR2-553 1.5 M solution 200 ml $69.00

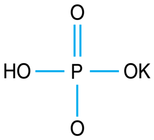

$$\mathrm{HO-\underset{\underset{O}{|}}{\overset{\overset{O}{\|}}{P}}-OK}$$

Synonyms: Monopotassium phosphate or
Potassium dihydrogen phosphate or prim.-Potassium phosphate
H_2KO_4P or KH_2PO_4 M_r 136.09 CAS [7778-77-0]
EC No 231-913-4 Purity ≥ 99.5%
pKa_1 = 2.15 at 25°C
pKa_2 = 6.82 at 25°C
pKa_3 = 12.38 at 25°C
Measured pH Range: 3.8 - 4.3 at 25°C
Refractive Index Range: 1.35379 - 1.35414 at 20°C
Conductivity Range: 67.7 - 78.1 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Na	≤ 0.005%
As	≤ 0.00001%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Bi	≤ 0.0005%	Li	≤ 0.0005%	SO_4	≤ 0.003%
Ca	≤ 0.001%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Cd	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%
Cl	≤ 0.0005%	Mo	≤ 0.0005%		
Co	≤ 0.0005%	N	≤ 0.001%		

Potassium nitrate

HR2-663 3.0 M solution 200 ml $79.00

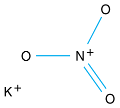

Synonyms: None KNO_3 M_r 101.10
CAS [7757-79-1] EC No 231-818-8 Purity ≥ 99.5%
Measured pH Range: 5.9 - 7.3 at 25°C
Refractive Index: 1.35835 at 20°C
Conductivity Range: 171.4 - 211.1 mS at 25°C

Al	≤ 0.0005%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
As	≤ 0.00001%	Fe	≤ 0.0002%	NO_2	≤ 0.001%
Ba	≤ 0.0005%	IO_3	≤ 0.0005%	Pb	≤ 0.0005%
Bi	≤ 0.0005%	Li	≤ 0.0005%	PO_4	≤ 0.0005%
Ca	≤ 0.005%	Mg	≤ 0.005%	SO_4	≤ 0.002%
Cd	≤ 0.0005%	Mn	≤ 0.0005%	Sr	≤ 0.0005%
Cl	≤ 0.01%	Mo	≤ 0.0005%	Zn	≤ 0.0005%
Co	≤ 0.0005%	NH_4^+	≤ 0.001%		
Cr	≤ 0.0005%	Na	≤ 0.005%		

Potassium sodium tartrate tetrahydrate

HR2-539 1.5 M solution 200 ml $91.00

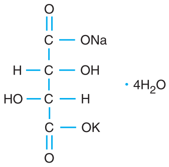

Synonyms: L(+)-Tartaric acid potassium sodium salt or
Rochelle salt or Seignette salt
KOCOCH(OH)CH(OH)COONa · $4H_2O$ or $C_4H_4KNaO_6 \cdot 4H_2O$
M_r 282.22 CAS [6381-59-5] EC No 206-156-8 Purity ≥ 99.5%
Measured pH Range: 8.4 - 9.1 at 25°C
Refractive Index Range: 1.37582 - 1.37599 at 20°C
Conductivity Range: 81.1 - 97.4 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Ni	≤ 0.0005%
As	≤ 0.00001%	Cu	≤ 0.0005%	Pb	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	PO_4	≤ 0.001%
Bi	≤ 0.0005%	Li	≤ 0.0005%	SO_4	≤ 0.005%
Ca	≤ 0.001%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Cd	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%
Cl	≤ 0.0005%	Mo	≤ 0.0005%		
Co	≤ 0.0005%	N	≤ 0.002%		

Potassium sulfate

HR2-675 0.5 M solution 200 ml $65.00

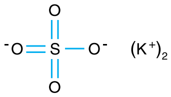

$$\mathrm{^{-}O=\underset{\underset{O}{\|}}{\overset{\overset{O}{\|}}{S}}-O^{-}\quad (K^{+})_2}$$

Synonyms: None K_2SO_4 M_r 174.26
CAS [7778-80-5] EC No 231-915-5 Purity ≥ 99.0%
Measured pH Range: 6.6 - 6.8 at 25°C
Refractive Index Range: N/A
Conductivity Range: 70.8 - 77.8 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Na	≤ 0.005%
As	≤ 0.00001%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Bi	≤ 0.0005%	Li	≤ 0.0005%	PO_4	≤ 0.001%
Ca	≤ 0.005%	Mg	≤ 0.002%	Sr	≤ 0.005%
Cd	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%
Cl	≤ 0.0005%	Mo	≤ 0.0005%		
Co	≤ 0.0005%	N	≤ 0.0005%		

Potassium thiocyanate

HR2-695 8.0 M solution 200 ml $105.00

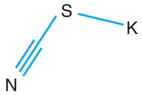

Synonyms: Potassium rhodanide KSCN or CKNS
M_r 97.18 CAS [333-20-0] EC No 206-370-1 Purity ≥ 99.0%
Measured pH Range: 7.2 at 25°C
Refractive Index Range: 1.45988 at 20°C
Conductivity Range: 767.6 mS at 25°C

Al	≤ 0.001%	Fe	≤ 0.0005%	Pb	≤ 0.001%
Ca	≤ 0.001%	Mg	≤ 0.0005%	SO_4	≤ 0.005%
Cl	≤ 0.005%	NH_4^+	≤ 0.003%	Zn	≤ 0.0005%
Cu	≤ 0.0005%	Na	≤ 0.005%		

L-Proline

HR2-775 1.0 M solution 100 ml $60.00

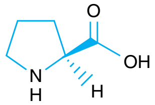

Synonyms: (S)-pyrrolidine-2-carboxylic acid $C_5H_9NO_2$
M_r 115.13 CAS [147-85-3] EC No 205-702-2 Purity ≥ 99.5%
Measured pH Range: 6.0 - 7.0 at 25°C
Refractive Index Range: 1.35260 at 20°C
Conductivity Range: 5.4 - 9.1 µS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Na	≤ 0.005%
As	≤ 0.00001%	Cu	≤ 0.0005%	NH_4^+	≤ .002%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	Ni	≤ 0.0005%
Bi	≤ 0.0005%	K	≤ 0.005%	Pb	≤ 0.0005%
Ca	≤ 0.001%	Li	≤ 0.0005%	SO_4	≤ 0.01%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Cl	≤ 0.01%	Mn	≤ 0.0005%	Zn	≤ 0.0005%
Co	≤ 0.0005%	Mo	≤ 0.0005%		

Sodium acetate trihydrate

HR2-543 3.0 M solution 200 ml $72.00
HR2-763 pH 7.0 - 4.0 M solution 200 ml $65.00

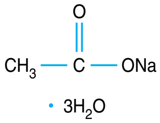

$$\mathrm{CH_3-\overset{\overset{O}{\|}}{C}-ONa\ \cdot 3H_2O}$$

Synonyms: Acetic acid sodium salt or Sodium acetate
$C_2H_3NaO_2 \cdot 3H_2O$ or $CH_3COONa \cdot 3H_2O$ M_r 136.08
CAS [6131-90-4] EC No 204-823-8 Purity ≥ 99.5%
HR2-543: Measured pH Range: 9.0 - 9.7 at 25°C
Refractive Index Range: 1.36432 - 1.36440 at 20°C
Conductivity Range: 73.7 - 79.9 mS at 25°C
HR2-763: Titrated to pH 7.0 at 25°C using Hydrochloric acid (HR2-581)
Refractive Index Range: 1.37343 - 1.37369 at 20°C
Conductivity Range: 74.3 - 77.8 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Mo	≤ 0.0005%
As	≤ 0.00001%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Bi	≤ 0.0005%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Ca	≤ 0.001%	K	≤ 0.005%	PO_4	≤ 0.0005%
Cd	≤ 0.0005%	Li	≤ 0.0005%	SO_4	≤ 0.002%
Cl	≤ 0.0005%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Co	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%

Sodium bromide

HR2-699 5.0 M solution 200 ml $41.00

Synonyms: None NaBr or BrNa M_r 102.89
CAS [7647-15-6] EC No 231-599-9 Purity ≥ 99.0% Measured pH Range: 6.6 - 8.4 at 25°C
Refractive Index Range: 1.39146 - 1.39846 at 20°C
Conductivity Range: 309.0 - 369.4 mS at 25°C

Ca	≤ 0.005%	Cu	≤ 0.005%	Pb	≤ 0.005%
Cd	≤ 0.005%	Fe	≤ 0.005%	SO_4	≤ 0.005%
Cl	≤ 0.1%	K	≤ 0.2%	Zn	≤ 0.005%
Co	≤ 0.005%	Ni	≤ 0.005%		

Sodium chloride

HR2-637 5.0 M solution 200 ml $33.00

Synonyms: None NaCl M_r 58.44 CAS [7647-14-5] EC No 231-598-3 Purity ≥ 99.5%
Measured pH Range: 5.5 - 8.6 at 25°C Refractive Index Range: 1.37712 - 1.37742 at 20°C
Conductivity Range: 281.2 - 354.8 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Mo	≤ 0.0005%
As	≤ 0.00001%	Cu	≤ 0.0005%	N	≤ 0.001%
Ba	≤ 0.0005%	Fe	≤ 0.0001%	Ni	≤ 0.0005%
Bi	≤ 0.0005%	I	≤ 0.001%	Pb	≤ 0.0005%
Br	≤ 0.005%	K	≤ 0.005%	PO_4	≤ 0.0005%
Ca	≤ 0.001%	Li	≤ 0.0005%	SO_4	≤ 0.01%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Co	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%

SALTS

Sodium citrate tribasic dihydrate

HR2-549 1.6 M solution 200 ml $74.00

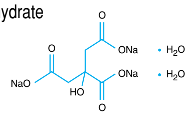

Synonyms: Citric acid trisodium salt dihydrate or Trisodium citrate dihydrate
$C_6H_5Na_3O_7 \cdot 2H_2O$ or $HOC(COONa)(CH_2COONa)_2 \cdot 2H_2O$
M_r 294.10 CAS [6132-04-3] EC No 200-675-3 Purity $\geq$ 99.0 %
Measured pH Range: 8.2 - 8.3 at 25°C
Refractive Index Range: 1.39431 - 1.39467 at 20°C
Conductivity Range: 50.2 - 60.1 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	NH_4^+	≤ 0.001%
As	≤ 0.00001%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Bi	≤ 0.0005%	K	≤ 0.01%	SO_4	≤ 0.005%
Ca	≤ 0.005%	Li	≤ 0.0005%	Zn	≤ 0.0005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%		
Cl	≤ 0.001%	Mn	≤ 0.0005%		
Co	≤ 0.0005%	Mo	≤ 0.0005%		

Sodium fluoride

HR2-645 0.8 M solution 100 ml $63.00

Synonyms: None NaF M_r 41.99 CAS [7681-49-4] EC No 231-667-8 Purity $\geq$ 99.0%
Measured pH Range: 8.7 - 9.9 at 25°C Refractive Index Range: 1.33713 - 1.33716 at 20°C
Conductivity Range: 49.9 - 52.7 mS at 25°C

Al	≤ 0.0005%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
As	≤ 0.00005%	Fe	≤ 0.001%	Pb	≤ 0.0005%
Ba	≤ 0.001%	HF	≤ 0.05%	SiF_6	≤ 0.1%
Bi	≤ 0.0005%	K	≤ 0.02%	SO_4	≤ 0.02%
Ca	≤ 0.001%	Li	≤ 0.0005%	SO_3	≤ 0.005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Cl	≤ 0.005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%
Co	≤ 0.0005%	Mo	≤ 0.0005%		
Cr	≤ 0.0005%	NaOH	≤ 0.04%		

Sodium formate

HR2-547 7.0 M solution 200 ml $68.00
HR2-765 pH 7.0 - 5.0 M solution 200 ml $69.00

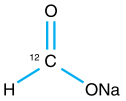

Synonyms: Formic acid sodium salt $CHNaO_2$ or HCOONa
M_r 68.01 CAS [141-53-7] EC No 205-488-0 Purity $\geq$ 99.0%
HR2-547: Measured pH Range: 8.4 - 9.3 at 25°C
Refractive Index Range: 1.38040 - 1.38045 at 20°C
Conductivity Range: 107.5 - 120.5 mS at 25°C
HR2-765: Titrated to pH 7.0 at 25°C using Hydrochloric acid (HR2-581)
Refractive Index Range: 1.36899 - 1.36909 at 20°C
Conductivity Range: 121.8 - 134.7 mS at 25°C

Al	≤ 0.0005%	Co	≤ 0.0005%	Mn	≤ 0.0005%
As	≤ 0.00001%	Cr	≤ 0.0005%	Mo	≤ 0.0005%
Ba	≤ 0.0005%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Bi	≤ 0.0005%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Ca	≤ 0.001%	K	≤ 0.01%	SO_4	≤ 0.005%
Cd	≤ 0.0005%	Li	≤ 0.0005%	Sr	≤ 0.0005%
Cl	≤ 0.005%	Mg	≤ 0.0005%	Zn	≤ 0.0005%

Sodium malonate

HR2-747 pH 4.0 - 3.4 M solution 200 ml $79.00
HR2-749 pH 5.0 - 3.4 M solution 200 ml $79.00
HR2-751 pH 6.0 - 3.4 M solution 200 ml $79.00
HR2-707 pH 7.0 - 3.4 M solution 200 ml $79.00
HR2-807 pH 8.0 - 3.4 M solution 200 ml $79.00
HR2-809 pH 9.0 - 3.4 M solution 200 ml $79.00

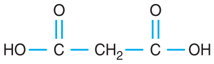

$$HO-\overset{\overset{\displaystyle O}{\|}}{C}-CH_2-\overset{\overset{\displaystyle O}{\|}}{C}-OH$$

Synonyms: Propanedioic acid or Malonic acid
$C_3H_4O_4$ (before titration with NaOH)
M_r 104.06 (before titration with NaOH) CAS [141-82-2] EC No 205-503-0 Purity $\geq$ 99.0%
HR2-747: Titrated to pH 4.0 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.38388 - 1.38458 at 20°C Conductivity Range: 64.2 - 74.1 mS at 25°C
HR2-749: Titrated to pH 5.0 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.39285 - 1.39476 at 20°C Conductivity Range: 63.5 - 77.9 mS at 25°C
HR2-751: Titrated to pH 6.0 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.39940 - 1.39982 at 20°C Conductivity Range: 64.0 - 75.5 mS at 25°C
HR2-707: Titrated to pH 7.0 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.40091 - 1.40112 at 20°C Conductivity Range: 66.8 - 77.1 mS at 25°C
HR2-807: Titrated to pH 8.0 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.40121 - 1.40135 at 20°C Conductivity Range: 72.6 - 73.2 mS at 25°C
HR2-809: Titrated to pH 9.0 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.40119 - 1.40128 at 20°C Conductivity Range: 72.2 - 73.1 mS at 25°C

Sodium nitrate

HR2-661 7.0 M solution 200 ml $75.00

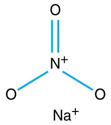

Synonyms: Chile salpeter
$NaNO_3$ M_r 84.99 CAS [7631-99-4] EC No 231-554-3 Purity $\geq$ 99.0%
Measured pH Range: 5.0 - 6.6 at 25°C
Refractive Index Range: 1.38545 - 1.38576 at 20°C
Conductivity Range: 174.4 - 195.3 mS at 25°C

Al	≤ 0.0005%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
As	≤ 0.00001%	Fe	≤ 0.0002%	NO_2	≤ 0.001%
Ba	≤ 0.0005%	IO_3	≤ 0.0005%	Pb	≤ 0.0005%
Bi	≤ 0.0005%	K	≤ 0.005%	PO_4	≤ 0.0002%
Ca	≤ 0.001%	Li	≤ 0.0005%	SO_4	≤ 0.003%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Cl	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%
Co	≤ 0.0005%	Mo	≤ 0.0005%		
Cr	≤ 0.0005%	NH_4^+	≤ 0.002%		

Sodium phosphate dibasic dihydrate

HR2-639 1.0 M solution 200 ml $94.00

Synonyms: sec-Sodium phosphate or di-sodium hydrogen phosphate dihydrate or
Disodium hydrogen phosphate dihydrate or Disodium phosphate
$HNa_2O_4P \cdot 2H_2O$ or $Na_2HPO_4 \cdot 2H_2O$ M_r 177.99
CAS[10028-24-7] EC No 231-448-7 Purity $\geq$ 99.0%
Measured pH Range: 8.7 - 8.9 at 25°C
Refractive Index Range: 1.35711 - 1.35717 at 20°C
Conductivity Range: 62.3 - 69.0 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	N	≤ 0.001%
As	≤ 0.00001%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Bi	≤ 0.0005%	K	≤ 0.005%	SO_4	≤ 0.005%
Ca	≤ 0.001%	Li	≤ 0.0005%	Sr	≤ 0.0005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Zn	≤ 0.0005%
Cl	≤ 0.001%	Mn	≤ 0.0005%		
Co	≤ 0.0005%	Mo	≤ 0.0005%		

Sodium phosphate monobasic monohydrate

HR2-551 4.0 M solution 200 ml $94.00

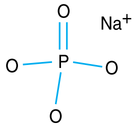

Synonyms: Monosodium phosphate or
Sodium dihydrogen phosphate monohydrate
$H_2NaPO_4 \cdot H_2O$ or $NaH_2PO_4 \cdot H_2O$ M_r 137.99 CAS [10049-21-5]
EC No 231-449-2 Purity $\geq$ 99.0%
Measured pH Range: 3.4 at 25°C
Refractive Index Range: 1.38216 - 1.38229 at 20°C
Conductivity Range: 58.8 - 60.5 mS at 25°C

As	≤ 0.00005%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Ca	≤ 0.001%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Cd	≤ 0.0005%	K	≤ 0.005%	SO_4	≤ 0.003%
Cl	≤ 0.0005%	Mg	≤ 0.0005%	Zn	≤ 0.0005%
Co	≤ 0.0005%	Mn	≤ 0.0005%		
Cr	≤ 0.0005%	N	≤ 0.001%		

Sodium potassium phosphate (Quik Optimize)

HR2-223 Quik Optimize 100 ml, 2 bottles $175.00

Quik Optimize kit includes:
4.0 M Sodium phosphate monobasic monohydrate, 100 ml (HR2-551)
M_r 137.99 $NaH_2PO_4 \cdot H_2O$ CAS [10049-21-5] EC No 231-449-2
4.0 M Potassium phosphate dibasic, 100 ml (HR2-635)
M_r 174.18 K_2HPO_4 CAS [7758-11-4] EC No 231-834-5

Each kit contains a dilution table to create any pH (between 5.0 and 8.2) and any concentration (between 0.2 and 4.0 M) in 0.2 increments. This kit can be used to reproduce Quik Screen (HR2-221) reagent conditions as well as formulate optimization conditions.

Sodium sulfate decahydrate

HR2-673 1.0 M solution 200 ml $75.00

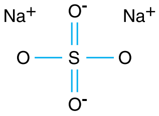

Synonyms: Glauber's salt
$Na_2SO_4 \cdot 10H_2O$ M_r 322.20 CAS [7727-73-3]
EC No 231-820-9 Purity $\geq$ 99.0%
Measured pH Range: 6.0 - 7.8 at 25°C
Refractive Index Range: 1.35184 - 1.35190 at 20°C
Conductivity Range: 89.0 - 96.1 mS at 25°C

Ca	≤ 0.002%	Fe	≤ 0.0005%	N	≤ 0.0005%
Cd	≤ 0.0005%	K	≤ 0.002%	Ni	≤ 0.0005%
Cl	≤ 0.0005%	Li	≤ 0.0005%	Pb	≤ 0.0005%
Co	≤ 0.0005%	Mg	≤ 0.0005%	PO_4	≤ 0.001%
Cr	≤ 0.0005%	Mn	≤ 0.0005%	Sr	≤ 0.0005%
Cu	≤ 0.0005%	Mo	≤ 0.0005%	Zn	≤ 0.0005%

SALTS

Sodium tartrate dibasic dihydrate

HR2-677 1.5 M solution 200 ml $85.00

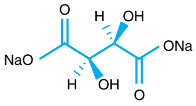

Synonyms: L-(+)-Tartaric acid disodium salt or Disodium tartrate dihydrate or Sodium tartrate dihydrate
$C_4H_4Na_2O_6 \cdot 2H_2O$ M_r 230.08 CAS [6106-24-7]
EC No 212-773-3 Purity $\geq$ 99.0%
Measured pH Range: 7.5 - 8.8 at 25°C
Refractive Index Range: 1.37600 - 1.37625 at 20°C
Conductivity Range: 67.9 - 71.0 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Ni	≤ 0.0005%
As	≤ 0.00001%	Cu	≤ 0.0005%	Pb	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	PO_4	≤ 0.0005%
Bi	≤ 0.0005%	K	≤ 0.005%	SO_4	≤ 0.005%
Ca	≤ 0.001%	Li	≤ 0.0005%	Sr	≤ 0.0005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Zn	≤ 0.0005%
Cl	≤ 0.0005%	Mn	≤ 0.0005%		
Co	≤ 0.0005%	Mo	≤ 0.0005%		

Sodium thiocyanate

HR2-693 8.0 M solution 200 ml $75.00

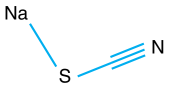

Synonyms: Sodium isothiocyanate or Sodium rhodanate or Sodium rhodanide or Sodium sulfocyanate
NaSCN or CNNaS M_r 81.07 CAS [540-72-7]
EC No 208-754-4 Purity $\geq$ 98.0%
Measured pH Range: 5.4 - 6.9 at 25°C
Refractive Index Range: 1.45963 - 1.46046 at 20°C
Conductivity Range: 137.1 - 161.5 mS at 25°C

Ca	≤ 0.02%	Cu	≤ 0.005%	Pb	≤ 0.005%
Cd	≤ 0.005%	Fe	≤ 0.005%	SO_4	≤ 0.05%
Cl	≤ 0.05%	K	≤ 0.2%	Zn	≤ 0.005%
Co	≤ 0.005%	Ni	≤ 0.005%		

Succinic acid pH 7.0

HR2-709 pH 7.0 - 1.2 M solution 200 ml $94.00

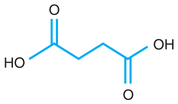

Synonyms: Dicarboxylic acid C4. or Butanedioic acid
$C_4H_6O_4$ or $HOOCCH_2CH_2COOH$ M_r 118.09 CAS [110-15-6]
EC No 203-740-4 Purity $\geq$ 99.5%
pKa_1 = 4.2 at 25°C pKa_2 = 5.6 at 25°C
Titrated to pH 7.0 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.36373 - 1.36407 at 20°C
Conductivity Range: 70.1 - 79.7 mS at 25°C

Ca	≤ 0.005%	Fe	≤ 0.0005%	Ni	≤ 0.0005%
Cd	≤ 0.0005%	K	≤ 0.005%	Pb	≤ 0.0005%
Cl	≤ 0.001%	Mg	≤ 0.001%	PO_4	≤ 0.001%
Co	≤ 0.0005%	Mn	≤ 0.0005%	SO_4	≤ 0.003%
Cr	≤ 0.0005%	N	≤ 0.001%	Zn	≤ 0.0005%
Cu	≤ 0.0005%	Na	≤ 0.005%		

Tacsimate pH 4, 5, 6, 7, 8, & 9

HR2-823	**100% solution pH 4.0**	**200 ml**	**$88.00**
HR2-825	**100% solution pH 5.0**	**200 ml**	**$88.00**
HR2-827	**100% solution pH 6.0**	**200 ml**	**$88.00**
HR2-755	**100% solution pH 7.0**	**200 ml**	**$88.00**
HR2-829	**100% solution pH 8.0**	**200 ml**	**$88.00**
HR2-813	**100% solution pH 9.0**	**200 ml**	**$88.00**
HR2-839	**55% Tacsimate pH 6.0, 0.1 M MES pH 5.8**	**100 ml**	**$95.00**
HR2-843	**55% Tacsimate pH 7.0, 0.1 M HEPES pH 6.8**	**100 ml**	**$95.00**
HR2-847	**55% Tacsimate pH 8.0, 0.1 M BIS-TRIS pH 7.8**	**100 ml**	**$95.00**
HR2-851	**55% Tacsimate pH 8.0, 0.1 M BTP pH 7.8**	**100 ml**	**$95.00**

Tacsimate is a pH titrated mixture of organic acids: 1.8305 M Malonic acid, 0.25 M Ammonium citrate tribasic, 0.12 M Succinic acid, 0.3 M DL-Malic acid, 0.4 M Sodium acetate trihydrate, 0.5 M Sodium formate, 0.16 M Ammonium tartrate dibasic

HR2-823: Titrated to pH 4.0 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.38590 - 1.38655 at 20°C Conductivity Range: 73.1 - 79.8 mS at 25°C
HR2-825: Titrated to pH 5.0 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.39432 - 1.39476 at 20°C Conductivity Range: 76.2 - 83.1 mS at 25°C
HR2-827: Titrated to pH 6.0 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.39981 - 1.40017 at 20°C Conductivity Range: 77.4 - 79.9 mS at 25°C
HR2-755: Titrated to pH 7.0 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.40102 - 1.40171 at 20°C Conductivity Range: 75.7 - 90.7 mS at 25°C

Continued next column....

continued....Tacsimate pH 4, 5, 6, 7, 8, & 9

HR2-829: Titrated to pH 8.0 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.40129 - 1.40166 at 20°C Conductivity Range: 74.9 - 79.8 mS at 25°C
HR2-813: Titrated to pH 9.0 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.40142 - 1.40151 at 20°C Conductivity Range: 75.6 - 75.9 mS at 25°C
HR2-839: Titrated to pH 5.8 at 25°C using Hydrochloric acid (HR2-581)
Refractive Index Range: 1.37622 at 20°C Conductivity Range: 94.2 mS at 25°C
HR2-843: Titrated to pH 6.8 at 25°C using Hydrochloric acid (HR2-581)
Refractive Index Range: 1.37822 at 20°C Conductivity Range: 92.6 - 96.0 mS at 25°C
HR2-847: Titrated to pH 7.8 at 25°C using Hydrochloric acid (HR2-581)
Refractive Index Range: 1.37808 at 20°C Conductivity Range: 93.8 mS at 25°C
HR2-851: Titrated to pH 7.8 at 25°C using Hydrochloric acid (HR2-581)
Refractive Index Range: 1.37949 at 20°C Conductivity Range: 93.7 mS at 25°C

Trimethylamine N-oxide dihydrate

HR2-777 1.0 M solution 100 ml $96.00

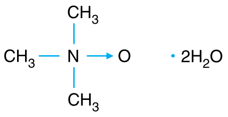

Synonyms: TMANO or N,N-Dimethylmethanamine oxide
$(CH_3)_3NO \cdot 2H_2O$ M_r 111.14 CAS [62637-93-8]
EC No 214-675-6 Purity $\geq$ 98.0%
Measured pH Range: 8.0 - 9.3 at 25°C
Refractive Index: 1.34327 at 20°C
Conductivity Range: 31.0 - 67.1 µS at 25°C

Tryptone

HR2-835 10% w/v Tryptone 100 ml $69.00

Measured pH Range: 7.0 - 7.1 at 25°C
Refractive Index: 1.35052 at 20°C
Conductivity Range: 7.7 - 8.5 mS at 25°C

Zinc acetate dihydrate

HR2-563 1.0 M solution 100 ml $74.00

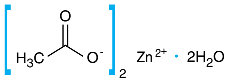

Synonyms: Zinc diacetate
$Zn(CH_3COO)_2 \cdot 2H_2O$ or $C_4H_6O_4Zn \cdot 2H_2O$ M_r 219.52
CAS [5970-45-6] EC No 209-170-2 Purity $\geq$ 98.0%
Measured pH Range: 5.5 - 5.8 at 25°C
Refractive Index Range: 1.35739 - 1.35749 at 20°C
Conductivity Range: 16.0 - 17.6 mS at 25°C

Ca	≤0.005%	Cu	≤0.005%	Ni	≤0.005%
Cd	≤0.005%	Fe	≤0.005%	Pb	≤0.005%
Cl	≤0.005%	K	≤0.01%	SO_4	≤0.005%
Co	≤0.005%	Na	≤0.01%		

Zinc chloride

HR2-811 2.0 M Zinc chloride 100 ml $56.00

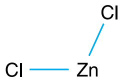

Synonyms: Zinc chloride anhydrous $ZnCl_2$ M_r 136.30
CAS [7646-85-7] EC No 231-592-0 Purity $\geq$ 98.0%
Measured pH Range: 3.1 - 4.1 at 25°C
Refractive Index Range: 1.37596 - 1.37597 at 20°C
Conductivity Range: 90.2 - 104.1 mS at 25°C

Ca	≤ 0.001%	Fe	≤ 0.0005%	Na	≤ 0.001%
Cd	≤ 0.0005%	K	≤ 0.001%	Ni	≤ 0.0005%
Co	≤ 0.0005%	Mg	≤ 0.0005%	NO_3	≤0.003%
Cr	≤ 0.0005%	Mn	≤ 0.0005%	Pb	≤ 0.001%
Cu	≤ 0.0005%	N	≤ 0.0005%	SO_4	≤0.002%

Zinc sulfate heptahydrate

HR2-641 2.0 M solution 100 ml $67.00

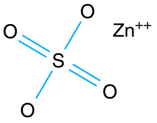

Synonyms: None $ZnSO_4 \cdot 7H_2O$ M_r 287.56
CAS [7446-20-0] EC No 231-793-3 Purity $\geq$ 99.5%
Measured pH Range: 3.4 - 4.0 at 25°C
Refractive Index Range: 1.38259 - 1.38264 at 20°C
Conductivity Range: 54.1 - 57.4 mS at 25°C

As	≤ 0.00001%	Cr	≤ 0.0005%	Mn	≤ 0.0002%
Ca	≤ 0.001%	Cu	≤ 0.0005%	N	≤ 0.0005%
Cd	≤ 0.0005%	Fe	≤ 0.0005%	Na	≤ 0.001%
Cl	≤ 0.0005%	K	≤ 0.001%	Ni	≤ 0.0005%
Co	≤ 0.0005%	Mg	≤ 0.0005%	Pb	≤ 0.001%

BUFFERS

ADA Buffer (Useful pH Range: 5.6 - 7.5)

HR2-507	**0.5 M solution**	**100 ml**	**$88.00**
HR2-817	**pH 6.5 - 1.0 M solution**	**100 ml**	**$98.00**

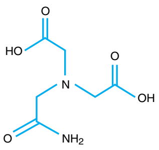

Synonyms: N-(2-Acetamido)iminodiacetic acid or
N-(Carbamoylmethyl)iminodiacetic acid
$C_6H_{10}N_2O_5$ or $H_2NCOCH_2N(CH_2CO_2H)_2$ M_r 190.16
CAS [26239-55-4] EC Number 247-530-0 Purity ≥ 99.0%
pKa = 6.6 at 25°C
HR2-507: Titrate to useful pH range using Sodium hydroxide (HR2-583)
Refractive Index: 1.35012 at 20°C
Conductivity Range: 23.9 - 28.6 mS at 25°C
HR2-817: Titrated to pH 6.5 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.36776 - 1.36803 at 20°C
Conductivity Range: 42.7 - 43.8 mS at 25°C

Al ≤ 0.0005%
As ≤ 0.00001%
Ba ≤ 0.0005%
Bi ≤ 0.0005%
Ca ≤ 0.001%
Cd ≤ 0.0005%
Cl ≤ 0.1%
Co ≤ 0.0005%
Cr ≤ 0.0005%
Cu ≤ 0.0005%
Fe ≤ 0.0005%
K ≤ 0.005%
Mg ≤ 0.0005%
Mn ≤ 0.0005%
Mo ≤ 0.005%
Na ≤ 0.05%
Ni ≤ 0.0005%
Pb ≤ 0.0005%
SO_4 ≤ 0.005%
Sr ≤ 0.0005%
Zn ≤ 0.0005%

BICINE Buffer (Useful pH Range: 7.4 - 9.3)

HR2-509	**1.0 M solution**	**100 ml**	**$76.00**
HR2-723	**pH 9.0 - 1.0 M solution**	**100 ml**	**$76.00**

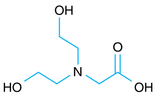

Synonyms: N,N-Bis(2-hydroxyethyl)glycine
$C_6H_{13}NO_4$ M_r 163.17 CAS [150-25-4] EC No 205-755-1
Purity ≥ 99.0% pKa = 8.3 at 25°C
HR2-509: Titrate to useful pH range using Sodium hydroxide (HR2-583)
Measured pH Range: 5.1 - 5.3 at 25°C
Refractive Index Range: 1.35805 - 1.35821 at 20°C
Conductivity Range: 114.3 - 114.7 µS at 25°C
HR2-723: Titrated to pH 9.0 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.36078 - 1.36104 at 20°C
Conductivity Range: 24.3 - 24.7 mS at 25°C

BIS-TRIS Buffer

HR2-781	**pH 5.5 - 1.0 M solution**	**100 ml**	**$124.00**
HR2-783	**pH 6.5 - 1.0 M solution**	**100 ml**	**$124.00**
HR2-906-24	**pH 7.8 - 1.0 M solution**	**185 ml**	**$151.00**

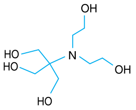

Synonyms: 2,2-Bis(hydroxymethyl)-2,2',2"-nitrilotriethanol or
2-Bis(2-hydroxyethyl)amino-2-(hydroxymethyl)-1,3-propanediol or
Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane
$C_8H_{19}NO_5$ M_r 209.24 CAS [6976-37-0] EC No 230-237-7
Purity ≥ 99.0% pKa = 6.4 at 20°C pKa = 6.5 at 25°C
HR2-781: Titrated to pH 5.5 at 25°C using Hydrochloric acid (HR2-581)
Refractive Index Range: 1.37130 - 1.37150 at 20°C
Conductivity Range: 38.4 - 43.1 mS at 25°C
HR2-783: Titrated to pH 6.5 at 25°C using Hydrochloric acid (HR2-581)
Refractive Index Range: 1.36860 - 1.36926 at 20°C
Conductivity Range: 22.9 - 28.7 mS at 25°C
HR2-906-24: BIS-TRIS pH 7.8 - 1.0 M solution is a Custom Shop reagent and is used to reproduce and optimize Crystallization Reagents for use with the Silver Bullets kits.

Ca ≤ 0.001%
Cd ≤ 0.0005%
Cl ≤ 0.005%
Co ≤ 0.0005%
Cr ≤ 0.0005%
Cu ≤ 0.0005%
Fe ≤ 0.0005%
K ≤ 0.005%
Mg ≤ 0.0005%
Mn ≤ 0.0005%
Na ≤ 0.005%
Ni ≤ 0.0005%
Pb ≤ 0.0005%
SO_4 ≤ 0.005%
Zn ≤ 0.0005%

BIS-TRIS propane pH 7.0 Buffer

HR2-795	**pH 7.0 - 1.0 M solution**	**100 ml**	**$119.00**

Synonyms: 1,3-Bis[tris(hydroxymethyl)methylamino]propane
$CH_2[CH_2NHC(CH_2OH)_3]_2$ or $C_{11}H_{26}N_2O_6$
M_r 282.34 CAS [64431-96-5] EC No 264-899-3
Purity ≥ 99.0% pKa_1 = 6.8 at 25°C pKa_2 = 9.0 at 25°C
HR2-795: Titrated to pH 7.0 at 25°C using Hydrochloric acid (HR2-581)
Refractive Index Range: 1.39139 - 1.39143 at 20°C
Conductivity Range: 41.7 - 42.5 mS at 25°C

Citric acid pH 3.5 Buffer (Useful pH Range: 2.2 - 6.5)

HR2-757	**pH 3.5 - 1.0 M solution**	**100 ml**	**$32.00**

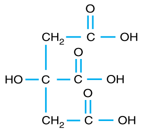

Synonyms: Citric acid anhydrous
$C_6H_8O_7$ or $HOC(COOH)(CH_2COOH)_2$
M_r 192.12 CAS [77-92-9] EC No 201-069-1 Purity ≥ 99.5%
pKa_1 = 3.13 at 25°C pKa_2 = 4.76 at 25°C pKa_3 = 6.4 at 25°C
HR2-757: Titrated to pH 3.5 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: N/A
Conductivity Range: 28.7 - 33.3 mS at 25°C

Al ≤ 0.0005%
As ≤ 0.00001%
Ba ≤ 0.0005%
Bi ≤ 0.0005%
Ca ≤ 0.001%
Cd ≤ 0.0005%
Cl ≤ 0.0005%
Co ≤ 0.0005%
Cr ≤ 0.0005%
Cu ≤ 0.0005%
Fe ≤ 0.0005%
K ≤ 0.005%
Li ≤ 0.0005%
Mg ≤ 0.0005%
Mn ≤ 0.0005%
Mo ≤ 0.0005%
Na ≤ 0.005%
Ni ≤ 0.0005%
Oxalate (C_2O_4) ≤ 0.05%
Pb ≤ 0.0005%
PO_4 ≤ 0.0005%
SO_4 ≤ 0.002%
Sr ≤ 0.0005%
Tartrate ($C_4H_4O_6$) ≤ 0.2%
Zn ≤ 0.0005%

Citric acid BIS-TRIS propane Buffer (Useful pH Range: 2.5 - 9.5)

HR2-831	**1.0 M Citric acid**	**100 ml**	**$33.00**
HR2-833	**1.0 M BIS-TRIS propane**	**100 ml**	**$119.00**

Citrate BIS-TRIS propane (CBTP) is a buffer system useful across pH 2.5 to 9.5. Using CBTP one can use a single buffer system to screen pH 2.5 to 9.5 by varying the ratio of Citric acid to BIS-TRIS propane. Although the two buffer reagents are made available separately for convenient, individual replacement, they are designed to be used together, with the supplied CBTP buffer titration table to create a crystallization buffer system covering pH 2.5 to 9.5.
HR2-831: Measured pH Range: 1.2 - 1.5 at 25°C
Refractive Index: 1.35682 at 20°C Conductivity Range: 7.8 - 8.7 mS at 25°C
HR2-833: Measured pH Range: 11.2 - 11.3 at 25°C
Refractive Index: 1.38076 at 20°C Conductivity Range: 309.0 - 364.0 µS at 25°C

HEPES Buffer (Useful pH Range: 6.8 - 8.2)

HR2-585	**1.0 M solution**	**100 ml**	**$84.00**
HR2-785	**pH 7.0 - 1.0 M solution**	**100 ml**	**$84.00**
HR2-729	**pH 7.5 - 1.0 M solution**	**100 ml**	**$84.00**

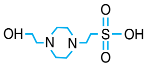

Synonyms: 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid or
N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) or
2-[4-(2-Hydroxyethyl)-1-piperazinyl]ethanesulfonic acid
$C_8H_{18}N_2O_4S$ M_r 238.31 CAS [7365-45-9] EC No 230-907-9
Purity ≥ 99.0% pKa = 7.5 at 25°C
HR2-585: Titrate to useful pH range using Sodium hydroxide (HR2-583)
Measured pH Range: 5.5 - 5.5 at 25°C
Refractive Index Range: 1.37121 - 1.37186 at 20°C
Conductivity Range: 161.2 - 178.4 µS at 25°C
HR2-785: Titrated to pH 7.0 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.37161 at 20°C Conductivity Range: 7.1 mS at 25°C
HR2-729: Titrated to pH 7.5 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.37184 at 20°C Conductivity Range: 12.1 mS at 25°C

HEPES sodium Buffer (Useful pH Range: 6.6 - 8.5)

HR2-577	**1.0 M solution**	**100 ml**	**$98.00**
HR2-733	**pH 7.5 - 1.0 M solution**	**100 ml**	**$98.00**
HR2-931-01	**pH 6.8 - 1.0 M solution**	**185 ml**	**$206.00**

Synonyms: HEPES sodium salt or
4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid sodium salt or
N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) sodium salt
$C_8H_{17}N_2NaO_4S$ M_r 260.29 CAS [75277-39-3] EC No 278-169-7
Purity ≥ 99.5% pKa = 7.5 at 25°C

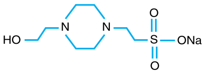

HR2-577: Titrate to useful pH range using Hydrochloric acid (HR2-581)
Measured pH Range: 10.5 - 10.9 at 25°C
Refractive Index Range: N/A
Conductivity Range: 25.9 - 28.2 mS at 25°C
HR2-733: Titrated to pH 7.5 at 25°C using Hydrochloric acid (HR2-581)
Refractive Index Range: 1.37713 - 1.37743 at 20°C
Conductivity Range: 40.6 - 42.1 mS at 25°C
HR2-931-01: HEPES sodium pH 6.8 - 1.0 M solution is a Custom Shop reagent and is used to reproduce and optimize Crystallization Reagents for use with the Silver Bullets kits.

BUFFERS

Imidazole Buffer (Useful pH Range: 6.2 - 7.8)

HR2-573 1.0 M solution 100 ml $68.00

HR2-819 pH 7.0 - 1.0 M solution 100 ml $82.00

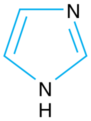

Synonyms: 1,3-Diaza-2,4-cyclopentadiene or Glyoxaline
$C_3H_4N_2$ M_r 68.08 CAS [288-32-4] EC No 206-019-2
Purity ≥ 99.5% pKa = 6.95 at 25°C

HR2-573: Titrate to useful pH range using Hydrochloric acid (HR2-581)
Measured pH Range: 10.2 - 10.6 at 25°C
Refractive Index: 1.34423 at 20°C
Conductivity Range: 72.2 - 85.2 µS at 25°C

HR2-819: Titrated to pH 7.0 at 25°C using Hydrochloric acid (HR2-581)
Refractive Index Range: 1.34847 - 1.34872 at 20°C
Conductivity Range: 46.1 - 49.2 mS at 25°C

Ca	≤ 0.001%	Cu	≤ 0.0005%	Na	≤ 0.005%
Cd	≤ 0.0005%	Fe	≤ 0.0005%	Ni	≤ 0.0005%
Cl	≤ 0.005%	K	≤ 0.005%	Pb	≤ 0.0005%
Co	≤ 0.0005%	Mg	≤ 0.0005%	SO_4	≤ 0.005%
Cr	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%

MES monohydrate Buffer (Useful pH Range: 5.2 - 7.1)

HR2-587 0.5 M solution 100 ml $98.00

HR2-787 pH 6.5 - 1.0 M solution 100 ml $98.00

HR2-943-07 pH 5.8 - 1.0 M solution 185 ml $138.00

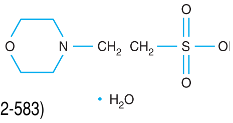

Synonyms: MES or 2-(N-morpholino)ethanesulfonic acid or
4-morpholineethanesulfonic acid monohydrate
$C_6H_{13}NO_4S \cdot H_2O$ M_r 213.25 CAS [145224-94-8]
EC No 224-632-3 Purity ≥ 99.0% pKa = 6.1 at 25°C

HR2-587: Titrate to useful pH range using Sodium hydroxide (HR2-583)
Measured pH Range: 3.2 - 3.4 at 25°C
Refractive Index Range: 1.34690 - 1.34731 at 20°C
Conductivity Range: 166.4 - 259.0 µS at 25°C

HR2-787: Titrated to pH 6.5 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index Range: 1.36197 - 1.36229 at 20°C
Conductivity Range: 19.2 - 25.6 mS at 25°C

HR2-943-07: MES monohydrate pH 5.8 - 1.0 M solution is a Custom Shop reagent and is used to reproduce and optimize Crystallization Reagents for use with the Silver Bullets kits.

Ca	≤ 0.005%	Cu	≤ 0.0005%	Na	≤ 0.005%
Cd	≤ 0.0005%	Fe	≤ 0.0005%	Ni	≤ 0.0005%
Cl	≤ 0.005%	K	≤ 0.005%	Pb	≤ 0.0005%
Co	≤ 0.0005%	Mg	≤ 0.0005%	SO_4	≤ 0.005%
Cr	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%

Sodium acetate trihydrate Buffer (Useful pH Range: 3.6 - 5.6)

HR2-569 1.0 M solution 100 ml $29.00

HR2-789 pH 4.5 - 1.0 M solution 100 ml $29.00

HR2-731 pH 4.6 - 1.0 M solution 100 ml $29.00

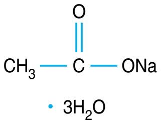

$$CH_3-\overset{\overset{\displaystyle O}{\|}}{C}-ONa \cdot 3H_2O$$

Synonyms: Acetic acid sodium salt $C_2H_3NaO_2 \cdot 3H_2O$ or
$CH_3COONa_2 \cdot 3H_2O$ M_r 136.08 CAS [6131-90-4]
EC No 204-823-8 Purity ≥ 99.5% pKa = 4.8 at 25°C

HR2-569: Titrate to useful pH range using Hydrochloric acid (HR2-581)
Measured pH Range: 8.7 - 9.2 at 25°C
Refractive Index Range: 1.34405 - 1.34408 at 20°C
Conductivity Range: 48.0 - 51.7 mS at 25°C

HR2-789: Titrated to pH 4.5 at 25°C using Hydrochloric acid (HR2-581)
Refractive Index Range: 1.34558 - 1.34564 at 20°C
Conductivity Range: 60.8 - 67.7 mS at 25°C

HR2-731: Titrated to pH 4.6 at 25°C using Hydrochloric acid (HR2-581)
Refractive Index Range: 1.34553 - 1.34556 at 20°C
Conductivity Range: 59.3 - 63.4 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Mo	≤ 0.0005%
As	≤ 0.00001%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Bi	≤ 0.0005%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Ca	≤ 0.001%	K	≤ 0.005%	PO_4	≤ 0.0005%
Cd	≤ 0.0005%	Li	≤ 0.0005%	SO_4	≤ 0.002%
Cl	≤ 0.0005%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Co	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%

Sodium cacodylate trihydrate Buffer (Useful pH Range: 5.0 - 7.4)

HR2-575 1.0 M solution 100 ml $178.00

HR2-737 pH 6.5 - 1.0 M solution 100 ml $178.00

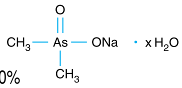

Synonyms: Cacodylic acid sodium salt trihydrate or
Dimethylarsinic acid sodium salt or Dimethylarsonic acid sodium salt
$C_2H_6AsNaO_2 \cdot 3H_2O$ or $(CH_3)_2AsO_2Na \cdot 3H_2O$ M_r 214.03
CAS [6131-99-3] EC No 204-708-2 pKa = 6.2 at 25°C Purity ≥ 98.0%

HR2-575: Titrate to useful pH range using Hydrochloric acid (HR2-581)
Measured pH Range: 8.3 - 8.5 at 25°C
Refractive Index: 1.35596 at 20°C
Conductivity Range: 33.3 - 37.7 mS at 25°C

HR2-737: Titrated to pH 6.5 at 25°C using Hydrochloric acid (HR2-581)
Refractive Index Range: 1.35651 - 1.35707 at 20°C
Conductivity Range: 43.5 - 45.4 mS at 25°C

Sodium citrate tribasic dihydrate Buffer (Useful pH Range: 3.0 - 6.2)

HR2-571 1.0 M solution 100 ml $41.00

HR2-735 pH 5.6 - 1.0 M solution 100 ml $41.00

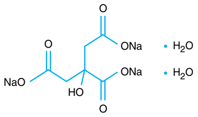

Synonyms: Citric acid trisodium salt dihydrate or
Trisodium citrate dihydrate $C_6H_5Na_3O_7 \cdot 2H_2O$ or
$HOC(COONa)(CH_2COONa)_2 \cdot 2H_2O$ M_r 294.10
CAS [6132-04-3] EC No 200-675-3 Purity ≥ 99.0%
pKa_1 = 3.1 at 25°C pKa_2 = 4.8 at 25°C pKa_3 = 5.4 at 25°C

HR2-571: Titrate to useful pH range using Hydrochloric acid (HR2-581)
Measured pH Range: 8.1 - 8.3 at 25°C Refractive Index: 1.37370 at 20°C
Conductivity Range: 60.0 - 67.5 mS at 25°C

HR2-735: Titrated to pH 5.6 at 25°C using Hydrochloric acid (HR2-581)
Refractive Index Range: 1.37423 - 1.37502 at 20°C
Conductivity Range: 67.4 - 72.4 mS at 25°C

Al	≤ 0.0005%	Co	≤ 0.0005%	Mn	≤ 0.0005%
As	≤ 0.00001%	Cr	≤ 0.0005%	Mo	≤ 0.0005%
Ba	≤ 0.0005%	Cu	≤ 0.0005%	NH_4^+	≤ 0.001%
Bi	≤ 0.0005%	Fe	≤ 0.0005%	Ni	≤ 0.0005%
Ca	≤ 0.005%	K	≤ 0.01%	Pb	≤ 0.0005%
Cd	≤ 0.0005%	Li	≤ 0.0005%	SO_4	≤ 0.005%
Cl	≤ 0.001%	Mg	≤ 0.0005%	Zn	≤ 0.0005%

Tris Buffer (Useful pH Range: 7.0 - 9.0)

HR2-589 1.0 M solution 100 ml $44.00

HR2-725 pH 8.5 - 1.0 M solution 100 ml $44.00

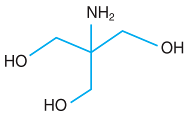

Synonyms: Trizma® base or THAM or Tris base or
2-Amino-2-(hydroxymethyl)-1,3-propanediol or
Tris(hydroxymethyl)aminomethane or Trometamol
$NH_2C(CH_2OH)_3$ or $C_4H_{11}NO_3$ M_r 121.14 CAS [77-86-1]
EC No 201-064-4 Purity ≥ 99.8% pKa = 8.1 at 25°C

HR2-589: Titrate to useful pH range using Hydrochloric acid (HR2-581)
Measured pH Range: 10.9 - 11.0 at 25°C
Refractive Index Range: 1.35044 - 1.35048 at 20°C
Conductivity Range: 197.0 - 223.0 µS at 25°C

HR2-725: Titrated to pH 8.5 at 25°C using Hydrochloric acid (HR2-581)
Refractive Index Range: 1.35364 - 1.35369 at 20°C
Conductivity Range: 18.1 - 21.5 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Na	≤ 0.005%
As	≤ 0.00001%	Cu	≤ 0.0005%	Ni	≤ 0.0005%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	Pb	≤ 0.0005%
Bi	≤ 0.0005%	K	≤ 0.005%	SO_4	≤ 0.0005%
Ca	≤ 0.001%	Li	≤ 0.0005%	Sr	≤ 0.0005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Zn	≤ 0.0005%
Cl	≤ 0.0005%	Mn	≤ 0.0005%		
Co	≤ 0.0005%	Mo	≤ 0.0005%		

BUFFERS

TRIS hydrochloride Buffer (Useful pH Range: 7.0 - 9.0)

HR2-579 1.0 M solution 100 ml $45.00
HR2-727 pH 8.5 - 1.0 M solution 100 ml $45.00

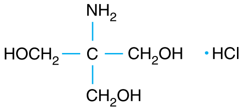

$HOCH_2-C(NH_2)(CH_2OH)-CH_2OH \cdot HCl$

Synonyms: TRIS HCl or Trizma® hydrochloride or Tris(hydroxymethyl)aminomethane hydrochloride
$NH_2C(CH_2OH)_3 \cdot HCl$ or $C_4H_{11}NO_3 \cdot HCl$
M_r 157.60 CAS [1185-53-1] EC No 214-684-5
Purity ≥ 99.0% pKa = 8.1 at 25°C

HR2-579: Titrate to useful pH range using Sodium hydroxide (HR2-583)
Measured pH Range: 3.9 - 4.1 at 25°C Refractive Index: 1.35882 at 20°C
Conductivity Range: 49.1 - 53.3 mS at 25°C

HR2-727: Titrated to pH 8.5 at 25°C using Sodium hydroxide (HR2-583)
Refractive Index: 1.35961 at 20°C Conductivity: 60.3 mS at 25°C

Al	≤ 0.0005%	Cr	≤ 0.0005%	Mo	≤ 0.0005%
As	≤ 0.0001%	Cu	≤ 0.0005%	Na	≤ 0.005%
Ba	≤ 0.0005%	Fe	≤ 0.0005%	Ni	≤ 0.0005%
Bi	≤ 0.0005%	K	≤ 0.005%	Pb	≤ 0.0005%
Ca	≤ 0.001%	Li	≤ 0.0005%	SO_4	≤ 0.005%
Cd	≤ 0.0005%	Mg	≤ 0.0005%	Sr	≤ 0.0005%
Co	≤ 0.0005%	Mn	≤ 0.0005%	Zn	≤ 0.0005%

Hydrochloric acid

HR2-581 1.0 M solution 100 ml $27.00

Synonyms: None HCl M_r 36.46 CAS [7647-01-0] EC No 231-595-7
Conductivity Range: 320.4 - 378.2 mS at 25°C

Ag	≤ 0.000001%	Cu	≤ 0.000002%	Ni	≤ 0.000002%
Al	≤ 0.000005%	Fe	≤ 0.00002%	PO_4	≤ 0.0005%
As	≤ 0.0000005%	Hg	≤ 0.000001%	Pb	≤ 0.000002%
Ba	≤ 0.000002%	K	≤ 0.00001%	SO_4	≤ 0.0001%
Bi	≤ 0.00001%	Li	≤ 0.000001%	SO_3	≤ 0.0001%
Br	≤ 0.005%	Mg	≤ 0.00001%	Sr	≤ 0.000001%
Ca	≤ 0.00005%	Mn	≤ 0.000001%	Tl	≤ 0.000005%
Cd	≤ 0.000001%	Mo	≤ 0.000002%	Zn	≤ 0.000005%
Co	≤ 0.000001%	Na	≤ 0.00005%		
Cr	≤ 0.000002%	NH_4^+	≤ 0.0001%		

Sodium hydroxide

HR2-583 1.0 M solution 100 ml $27.00

Synonyms: Caustic soda NaOH or HNaO M_r 40.00 CAS [1310-73-2]
EC No 215-185-5 Purity ≥ 98.0%

Al	≤ 0.0005%	Cu	≤ 0.0005%	P	≤ 0.0005%
Ca	≤ 0.0005%	Fe	≤ 0.0005%	Pb	≤ 0.001%
Cd	≤ 0.0005%	K	≤ 0.02%	SO_4	≤ 0.05%
Cl	≤ 0.005%	Mg	≤ 0.0005%	Zn	≤ 0.0005%

SOLUBILIZING AGENTS

NDSB-195

HR2-703 powder 5 g $98.00

Synonyms: Dimethylethylammonium propane sulfonate or Non-detergent sulfobetaine $C_7H_{17}NO_3S$ M_r 195.3
Purity > 95.0% by TLC

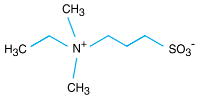

NDSB-201

HR2-701 powder 5 g $27.00

Synonyms: 3-(1-Pyridino)-1-propane sulfonate
$C_8H_{11}NO_3S$ M_r 201.2
CAS Number 15471-17-7 EC No 2394913
Purity > 99.0%

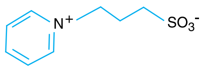

NDSB-211

HR2-793 powder 5 g $181.00

Synonyms: Dimethyl-2(-Hydroxyethyl)-(3-Sulfopropyl)-ammonium, Inner Salt or Non detergent sulfobetaine 211
$C_7H_{17}NO_4S$ M_r 211.28 Purity > 99.0%

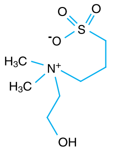

NDSB-221

HR2-791 powder 5 g $133.00

Synonyms: Non-detergent sulfobetaine 221 or 3-(1-Methylpiperidinium)-1-propane Sulfonate
$C_9H_{19}NO_3S$ M_r 221.3 Purity > 99.0%

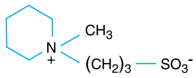

NDSB-256

HR2-705 powder 5 g $98.00

Synonyms: Dimethylbenzylammonium propane sulfonate
$C_{12}H_{19}NO_3S$ M_r 257.4
Purity > 99.0% by TLC

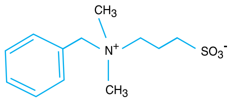

WHAT ARE NDSBS?

DESCRIPTION

Non Detergent Sulfobetaines (NDSB) are a new family of non denaturing protein solubilizing agents with a wide range of applications. NDSBs have been used for protein extraction, solubilization, and crystallization. NDSBs are zwitterionic, they can be removed by dialysis since they do not form micelles, and they do not alter the biophysical properties of biological buffers.

This particular class of reagent was developed by Laurent Vuillard (University of Cambridge, Department of Pathology, UK) in collaboration with T. Rabilloud (CENG, Grenoble, France).

References:
1. *Vuillard, L., T. Rabilloud, M.E. Goldberg, A New Additive for Protein Crystallization., FEBS Letters., 1994., 353(3):, 294-296.*
2. *Vuillard, L., et al., Protein crystallography with non detergent sulfobetaines., J. Cryst. Growth. (1996) 168, 150-154.*
3. *Vuillard, L., Rabilloud, T., Goldberg, M.E., Eur. J. Biochem., (1998) 256, 1, 128-135.*
4. *http://www.path.cam.ac.uk/~lv213/*

Crystal picture of RPA in an earring motif.

Courtesy of Gloria Borgstahl.

University of Toledo
Department of Chemistry

TCEP HYDROCHLORIDE

APPLICATION

- Stable reducing agent for crystallization trials

FEATURES

- Highly soluble
- Odorless and non-volatile
- Effective at acidic and alkaline pH

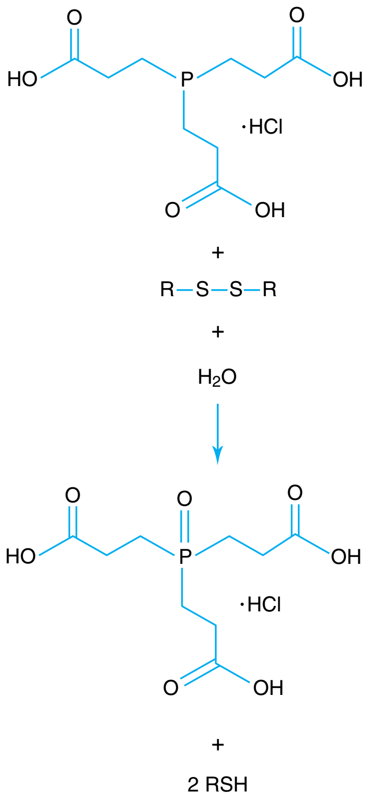

DESCRIPTION

Try using TCEP hydrochloride instead of dithiothreitol (DTT) as a reducing agent in your crystallization setups. TCEP hydrochloride is soluble in water to 310 g per L. TCEP hydrochloride is Tris (2-carboxyethyl)phosphine hydrochloride. It is an odorless (non-volatile) reducing agent that is more stable and effective than DTT or 2-Mercaptoethanol. Unlike DTT, TCEP hydrochloride retains its reducing power at acid pHs (pH 5) and at pHs above 7.5. It is unreactive toward other functional groups found in proteins. Unlike DTT (dithiothreitol), TCEP hydrochloride does not contain a free thiol, and therefore does not require removal before reaction with a thiol-reactive reagents. It reduces disulfides but apparently not mercury–thiol bonds. TCEP hydrochloride is compatible with many heavy atoms and may be used during heavy atom derivatization. It is more stable at a higher pH and at higher temperatures than is DTT and for a longer period of time in buffers without metal chelators such as EGTA.

References

1. *Tris(2-carboxyethyl)phosphine stabilization of RNA: comparison with dithiothreitol for use with nucleic acid and thiophosphoryl chemistry. Rhee SS, Burke DH. Anal Biochem 325, 137-43 (2004) PN51541.*
2. *A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. Getz EB, Xiao M, Chakrabarty T, Cooke R, Selvin PR. Anal Biochem 273, 73-80 (1999) PN34481*
3. *Burns, J.A., et al. (1991). Selective reduction of disulfides by tris-(2-carboxyethyl)-phosphine. J. Org. Chem. 56, 2648-2650*
4. *Oda , Y., et al. (2001). Nature Biotech19, 379-382*

Order Information

Cat. No.	Name	Description	Price
HR2-651	**TCEP hydrochloride**	1 g	**$46.00**
HR2-801	**TCEP hydrochloride**	10 g	**$290.00**

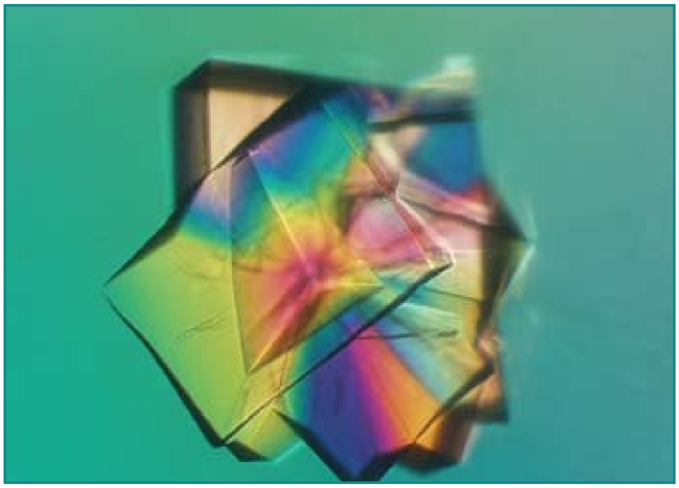

Hanging drop vapor diffusion crystal.

Ertugrul Cansizoglu, University of Texas, Southwestern, USA.

PERFLUOROPOLYETHER PFO-X175/08

APPLICATION

- Cryoprotectant

FEATURES

- Perfluoropolyether oil
- Low viscosity
- Low surface tension

DESCRIPTION

Used during cryocrystallography to displace and reduce the amount of water (mother liquor, reagent) on the crystal after the crystal is mounted in a cryoloop. Coating the crystal with oil can minimize evaporation from the crystal and reduce exposure and slow diffusion of air (oxygen) to the crystal.

References

1. *Structure of the 'open' form of Aspergillus nidulans 3-dehydroquinate synthase at 1.7 Å resolution. C.E. Nichols, A.R. Hawkins and D.K. Stammersa. Acta Crystallographic Section D, Volume 60, Part 5, Pages 971-973, May 2004.*
2. *H. Hope, Annu. Rev. Biophys. Chem. 1990 19:107-126*

Order Information

Cat. No.	Name	Description	Price
HR2-814	**Perfluoropolyether PFO-X175/08**	1 ml	**$25.00**

PARATONE-N

APPLICATION

- Cryoprotectant

DESCRIPTION

Paratone-N is a viscous cryoprotectant for both small and large molecule crystallography.[1] Mildly air-unstable, small molecule compounds can be coated with Paratone-N under an inert atmosphere. The Paratone-N protected crystal sample can be cryocooled in a chilled nitrogen gas stream. Paratone-N has also been used successfully as a cryoprotectant for biological macromolecule crystals. It is used during cryocrystallography to displace and reduce the amount of water (mother liquor, reagent) on the crystal after the crystal is mounted in a cryoloop. Coating the crystal with Paratone-N can minimize evaporation from the crystal and reduce exposure and slow diffusion of air (oxygen) to the crystal.

Paratone-N is also known as Parabar 10312, Paratone 8277, and Infineum V8512.

References

1. *H. Hope, Cryocrystallography of biological macromolecules: a generally applicable method. Acta Cryst. (1988) B44, 22-26.*
2. *Structure of the ligand-binding domain (LBD) of human androgen receptor in complex with a selective modulator LGD2226. F. Wang, X.- Liu, H. Li, K.- Liang, J. N. Miner, M. Hong, E. A. Kallel, A. van Oeveren, L. Zhi and T. Jiang. Acta Cryst. (2006). F62, 1067-1071*
3. *H. Hope, Annu. Rev. Biophys. Chem. 1990 19:107-126*
4. *S. Parkin and H. Hope, J. Appl. Cryst. (1998) pages 945-953*

Order Information

Cat. No.	Name	Description	Price
HR2-643	**Paratone-N**	100 ml	**$55.00**

Trypsin jungle.

Allan D'Arcy, Switzerland.

FLUORINERT™ FC-70 FLUID

APPLICATION

- Drop floating crystallization

FEATURES

- Viscosity similar to water with approximately 75% greater density
- Stable
- Completely fluorinated

figure 1
Floating drop vapor diffusion technique

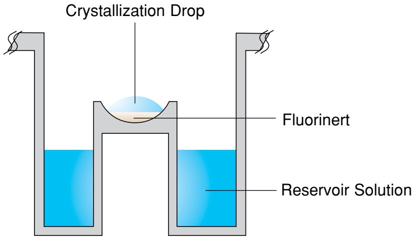

figure 2
Floating and stirring technique

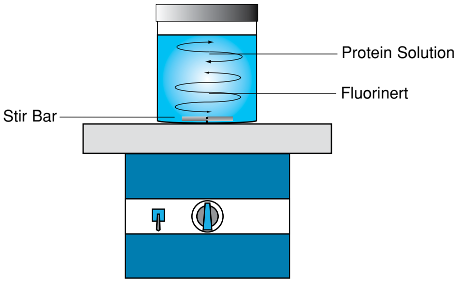

DESCRIPTION

Floating Drop Vapor Diffusion

The floating drop vapor diffusion technique[1] has been described using Fluorinert FC-70. The transparent and high density FC-70 liquid is not miscible with most crystallization reagents, samples, oil, or water. When Fluorinert liquid and the sample/reagent are mixed, the sample/reagent drop immediately separates from the Fluorinert and floats on top of the Fluorinert (see figure 1). Crystals grown using the floating drop technique do not stick to the crystallization plate. As the sample/reagent/crystal is not miscible with the Fluorinert liquid, complete separation of the crystal from the Fluorinert is straight forward. Crystals grown using the floating drop technique can be removed from the drop without mechanical damage. Drop volumes can vary widely using the floating drop method. Handling of the crystal is improved.

Floating and Stirring Technique (FAST)

The floating and stirring technique (FAST)[2] has been described using Fluorinert FC-70 liquid. The method, which involves a sample and reagent mixture applied over a non-miscible dense liquid (Fluorinert FC-70) without contact and slow stirring can also be combined with slow cooling or warming (See figure 2). The floating and stirring method has been reported to accelerate growth of the crystal as well as prevent subsequent spontaneous nucleation. The method has been used with seed crystals to promote the growth of a larger crystal without the appearance of subsequent crystals.

Fluorinert: FC-70 Fluid
Mr: 820
CAS number: [86508-42-1]
Pour point: -25°C
Boiling point: 215°C
Nonflammable
Vapor pressure: 15 pascals
Liquid density: 1,940 kg/m3
Kinematic viscosity: 12 centistokes
Absolute viscosity: 24 centipoise
Coeffecient of expansion: 0.0010°C-1
Surface tension: 18 dynes/cm
Refractive index: 1.303
Water solubility: 8 ppmw
Ozone depletion potential: 0
Dielectric strength: 40 kV, 0.1” gap
Dielectric constant: 1.98
Electrical resistivity: 2.3 x 1015 ohm cm
Appearance: Clear, colorless

References

1. *Application of a two-liquid system to sitting-drop vapour-diffusion protein crystallization. Adachi, H. et al, Acta Cryst. (2003) D59, 194-196*
2. *Promotion of large protein crystal growth with stirring solution. Adachi, H. et al. Jpn. J. Appl. Phys. Vol. 41 (2002) pp.1025-1027*
3. *Two-liquid hanging-drop vapour-diffusion technique of protein crystallization. Hiroaki Adachi et al. Japanese Journal of Applied Physics. Vol. 43, No. 1A/B, 2004, pp.L79-L81.*

Order Information

Cat. No.	Name	Description	Price
HR2-797	**100% Fluorinert FC-70 Fluid**	100 ml	**$99.00**

MICROBATCH CRYSTALLIZATION OILS

APPLICATION

- Microbatch crystallization

FEATURES

- Screen different temperatures without condensation
- Protect the sample from oxidation
- Under oil crystallization

DESCRIPTION

Oils used for microbatch and modified microbatch crystallization under oil.

Al's Oil is a 50:50 (volume:volume) mixture of Paraffin Oil and Silicon Oil.

Paraffin oil
Synonym: Mineral oil
CAS Number [8042-47-5]
EC Number 2324558
RTECS PY8047000
Appearance Colorless, clear, viscous liquid
Refractive Index n20/D 1.467 (literature) n20/D 1.468
IR Spectroscopy Grade
Density 0.85 g/ml at 20°C
Viscosity 40-42 CST at 25°C

References

1. *The advantages of using a modified microbatch method for rapid screening of protein crystallization conditions. Allan D'Arcy et al. Acta Cryst. (2003). D59, 396-399*

Order Information

Cat. No.	Name	Description	Price
HR3-411	**Paraffin Oil**	100% - 250 ml	**$40.00**
HR3-421	**Paraffin Oil**	100% - 1 L	**$100.00**
HR3-415	**Silicon Oil**	100% - 250 ml	**$150.00**
HR3-423	**Silicon Oil**	100% - 1 L	**$350.00**
HR3-413	**Al's Oil**	(50:50 Paraffin:Silicon) - 250 ml	**$90.00**
HR3-417	**Combo Oil Pack**	(Paraffin, Al's, & Silicon Oil) - 250 ml of each	**$225.00**

CONTAINERLESS CRYSTALLIZATION

APPLICATION

- Contact free crystallization by microbatch under oil

FEATURES

- Utilizes two immiscible oils to create a simulated containerless system

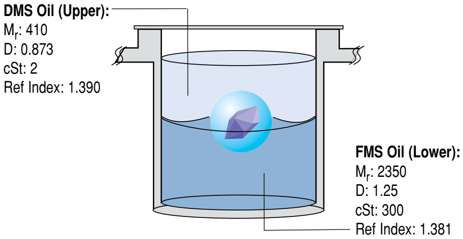

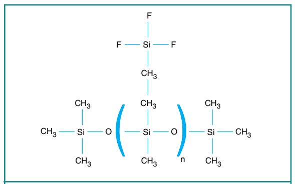

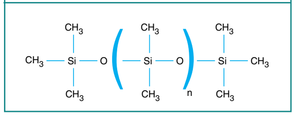

DESCRIPTION

Containerless crystallization is a micromethod of batch crystallization of proteins under conditions where the sample and precipitant have no contact with the surface of the crystallization plate. Drops (2 to 20 μl) of sample combined with crystallization reagent are pipetted at the interface between two layers of inert and immiscible silicon oils contained in a 24, 48, or 96 well plate and sealed. Plates and plain cover slides are available separately. It is easier to position drops into the center of the oil using plates with larger reservoirs (i.e. 24 and 48 well plates) but this requires more oil. When using the 96 well plates, drops of higher density (polymer based reagents such as PEG) tend to migrate to the side walls of the plate.

The method relies upon the use of two immiscible oils (FMS and DMS). Poly-3, 3, 3-Trifluoropropylmethylsiloxane (FMS), a branched silicone compound, has a higher density than Polydimethylsiloxane (DMS). Therefore, one can create a bilayer with a distinct interface by placing the two oils in a reservoir. Neither DMS nor FMS is miscible with water. The oils are compatible with most crystallization reagents including but not limited to buffers, salts, polymers, and MPD. Additional protein or crystallization reagents can be added to the drop in the oil. Crystal mounting as well as seeding can be performed while the drop is in the oil. During crystal mounting, the drop will maintain a spherical shape and decrease in size as sample is withdrawn. During seeding, crystal seeds will sediment to the bottom of the drop. It has been observed that large, perfect, single crystals can be grown using the containerless crystallization method. Crystals typically nucleate at the oil/water interface and grow into the drop.

Order Information

Cat. No.	Name	Description	Price
HR2-593	**DMS Oil**	100 ml	**$73.00**
HR2-595	**FMS Oil**	100 ml	**$171.00**

OPTIMIZE™ - SILICA HYDROGEL KIT

APPLICATION

- Quick & easy kit format for crystallization in gels

FEATURES

- Proprietary Silica Hydrogel formulation with neutral pH
- Gel matrix can reduce nucleation & sedimentation

DESCRIPTION

Gels are a very efficient media for growing macromolecular crystals.[1-5] Silica gels in particular have the advantage in that they are stable, usable over a wide range of temperatures (0-60°C), and are compatible with a wide variety of precipitants and additives used for crystal growth.

Gels can reduce nucleation and sedimentation, provide added stability, and allow crystals to grow larger. The porous network minimizes natural convection. Crystals are suspended in the gel network so that they do not form sediments and can grow free from strain exerted by the container or other crystals. Heterogeneous and secondary nucleation are reduced in the presence of a Silica Hydrogel.[3] The Silica Hydrogel can be used for liquid-gel, liquid-gel-liquid, vapor diffusion, as well as dialysis crystallization methodologies. The gel is compatible with a wide range of salts, polymers, organic solvents, and buffers used for macromolecular crystallization in a pH range from 3 to 10.

Each Silica Hydrogel kit contains 12 tubes of Sodium silicate solution and 12 tubes of Acetic acid solution, 500 μl each. All solutions are sterile filtered and formulated using ultra-pure water. Crystallization accessories are sold separately.

References

1. *Robert, M.C. & Lefaucheux, F., J. Crystal Growth (1988) 90, 358.*
2. *Provost, K. & Robert, M.C., J. Crystal Growth (1991) 110, 258.*
3. *M.C. Robert, K. Provost, & F. Lefaucheux, Crystallization of Nucleic Acids and Protein, A Practical Approach, Oxford Univ Press (1992) 127-143.*
4. *McPherson A., Methods in Enzymology (1985) 114, 112.*
5. *Cudney, B., Patel, S., McPherson, A., Acta Cryst. (1994) D50, 479-483.*

Order Information

Cat. No.	Name	Description	Price
HR2-310	**Silica Hydrogel**	500 µl, 24 tubes	**$95.00**

OPTIMIZE™ - LM AGAROSE

APPLICATION

- Crystallization in agarose gel

FEATURES

- Gel matrix can reduce nucleation and sedimentation
- Crystallization grade
- Low melting agarose

DESCRIPTION

Low melting (LM) agaroses are the result of a derivatization process by organic synthesis. The main properties of these agaroses are their low melting and gelling temperatures when compared with standard agaroses. LM agaroses have higher clarity (gel transparency) than gels of standard agaroses. The gelling temperature of LM agaroses is 24 to 28°C.

The structure of the polysaccharide is that of a galactan, formed by linking agarobioses by links 1-3, 1-4. This chemical structure gives agaroses the capacity to form strong gels even at low temperatures. The gels have a macroreticular structure with a very open mesh which can be adjusted simply by varying the concentration of the agarose. The macroreticule structure of the agarose gel is formed by hydrogen bonds, which makes the gel reversible, transforming the gel into a solution by heating. The absence of ionic groups makes the gel a neutral structure. With no interaction, macromolecules can migrate through the gel mesh, making the gel an efficient sieve for biological macromolecules.

Order Information

Cat. No.	Name	Description	Price
HR8-092	**LM Agarose**	10 g	**$58.00**

OPTIMIZE™ - IZIT CRYSTAL DYE™

APPLICATION

- Differentiate protein crystals from salt crystals

FEATURES

- 0.22 micron sterile filtered solution
- Small molecule dye penetrates solvent channels of macromolecular crystals, coloring the crystals blue.
- Salt crystals cannot absorb Izit and remain colorless

DESCRIPTION

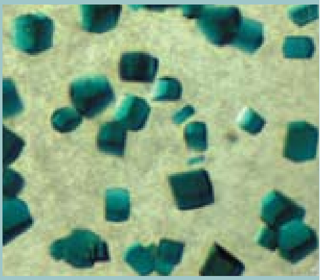

A crystal! Is it protein or is it salt? Have you ever asked yourself this question? Sure you could mount the crystal in question and take a quick look at the diffraction pattern but perhaps you're too lazy or you don't have an x-ray facility. Well, there's always the crush test. One simply takes a MicroTool™ or Crystal Probe™ and crushes the suspect crystal. A click or solid crunch is indicative of salt while a powder or silent destruction might indicate one just destroyed a perfectly good protein crystal. Izit is here to help. Simply place one μl of Izit in the sample drop and wait for an hour or so. Izit is a small molecule dye which will fill the solvent channels in protein crystals, coloring the crystals blue. With the appropriate dilution, Izit will in fact leave a clear drop with blue crystals, as illustrated in the picture to the right. Salt crystals do not possess these large solvent channels. Therefore, Izit cannot enter the crystal, leaving one with a clear crystal and a blue drop. Izit is especially nice for small microcrystals or questionable precipitate. The 0.5 ml vial of Izit is sufficient for more than thousands of crystallization drops.

Each Izit vial contains 0.5 ml of Izit dye. All solutions are formulated using ultra-pure water and are sterile filtered. Crystallization accessories are sold separately.

Order Information

Cat. No.	Name	Description	Price
HR4-710	**Izit Crystal Dye**	0.5 ml vial	**$15.00**

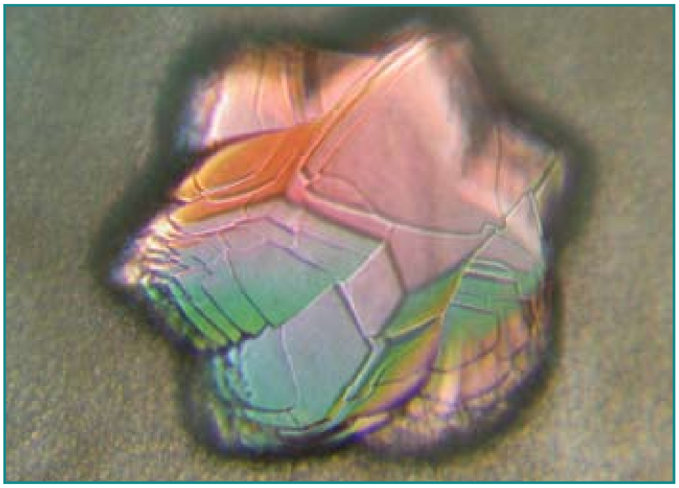

Picturesque crystal of multidrug efflux pump AcrB.

Markus Seeger, University Zürich, Switzerland.

Nuclear receptor crystal.

Gilbert Bey, ALIX, Illkirch, France.

Crystals of Thermus thermophilus enolase.

From the group of Paola Fucini,
Max Planck Institute for Molecular Genetics, Berlin, Germany.

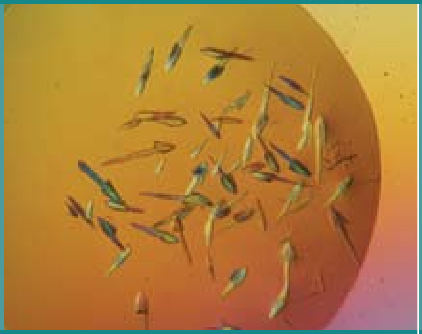

Protein crystals.

Kasumi Kobayashi, Taiji Nakae and Hiroyuki Akama.
The Kitasato Institute, Kanagawa, Japan.

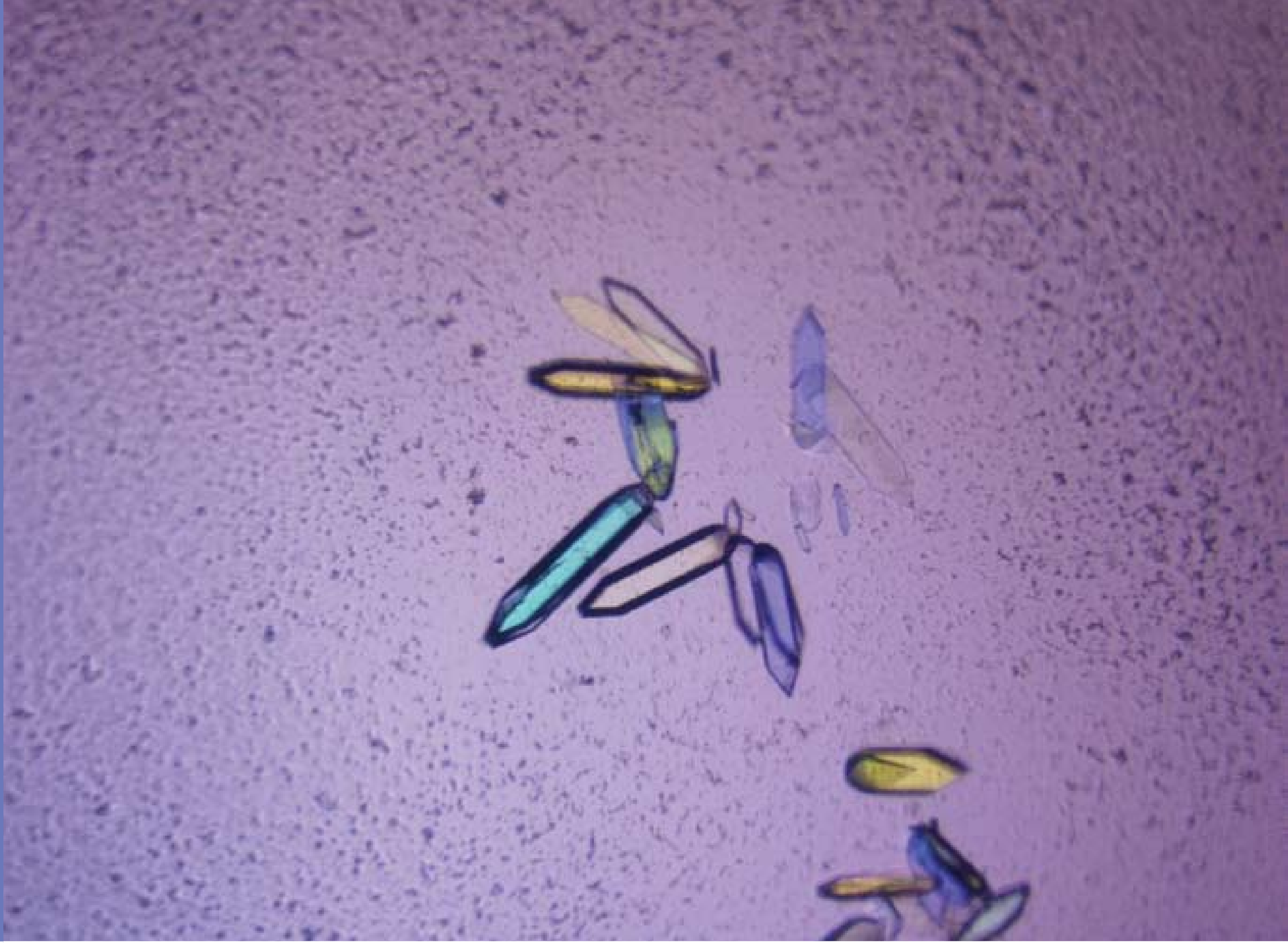

STOCKOPTIONS™ KITS

Protein crystal.

Alexey Rak, Max-Planck-Institut fysiologie, Department of Physical Biochemistry, Dortmund, Germany.

TABLE OF CONTENTS

PAGES

62 STOCKOPTIONS™ SALT

63 STOCKOPTIONS™ PH

64 - 67 STOCKOPTIONS™ BUFFER

68 STOCKOPTIONS™-CRYOPRO™

HAMPTON RESEARCH

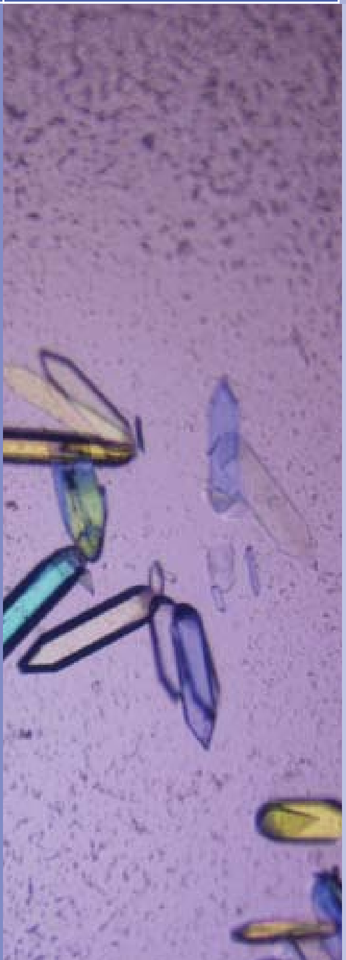

STOCKOPTIONS™ SALT

APPLICATION

- Crystallization grade salt reagent stock solutions for screen formulation and optimization

FEATURES

- Preformulated and sterile filtered
- Easy transition from screening to optimization
- Synergistic with Hampton Research screens, kits, & reagents
- 49 unique salts, including organic acids
- Highly concentrated, ready to dilute

STOCKOPTIONS SALT KIT FORMULATION

Reagent Number	Reagent
1	1.0 M Ammonium acetate
2	5.0 M Ammonium chloride
3	2.5 M Ammonium phosphate monobasic
4	10.0 M Ammonium fluoride
5	10.0 M Ammonium formate
6	2.5 M Ammonium citrate dibasic
7	3.5 M Ammonium phosphate dibasic
8	10.0 M Ammonium nitrate
9	3.5 M Ammonium sulfate
10	2.0 M Ammonium tartrate dibasic
11	1.0 M Calcium acetate hydrate
12	2.0 M Calcium chloride dihydrate
13	5.0 M Lithium acetate dihydrate
14	10.0 M Lithium chloride
15	1.5 M Lithium citrate tribasic tetrahydrate
16	8.0 M Lithium nitrate
17	2.0 M Lithium sulfate monohydrate
18	1.0 M Magnesium acetate tetrahydrate
19	2.0 M Magnesium chloride hexahydrate
20	1.0 M Magnesium formate dihydrate
21	3.0 M Magnesium nitrate hexahydrate
22	2.5 M Magnesium sulfate hydrate
23	4.0 M Nickel(II) chloride hexahydrate
24	5.0 M Potassium acetate
25	4.0 M Potassium chloride
26	2.5 M Potassium citrate tribasic monohydrate
27	1.5 M Potassium phosphate monobasic
28	6.0 M Potassium fluoride
29	14.0 M Potassium formate
30	3.0 M Potassium phosphate dibasic
31	3.0 M Potassium nitrate
32	1.5 M Potassium sodium tartrate tetrahydrate
33	0.5 M Potassium sulfate
34	8.0 M Potassium thiocyanate
35	3.0 M Sodium acetate trihydrate
36	5.0 M Sodium chloride
37	1.6 M Sodium citrate tribasic dihydrate
38	5.0 M Sodium phosphate monobasic monohydrate
39	0.8 M Sodium fluoride
40	7.0 M Sodium formate
41	1.0 M Sodium phosphate dibasic dihydrate
42	3.4 M Sodium malonate pH 7.0
43	7.0 M Sodium nitrate
44	1.0 M Sodium sulfate decahydrate
45	1.5 M Sodium tartrate dibasic dihydrate
46	8.0 M Sodium thiocyanate
47	1.2 M Succinic acid pH 7.0
48	1.0 M Zinc acetate dihydrate
49	2.0 M Zinc sulfate heptahydrate

DESCRIPTION

StockOptions reagent kits are reagent tool boxes for the macromolecular crystal grower. They offer precisely formulated, high quality crystallization grade reagent stocks in convenient, cost-effective kits. The chemicals utilized in these kits are the same crystallization grade, ultra-high purity chemicals utilized in the Hampton Research kits such as Crystal Screen™ and Crystal Screen 2™. StockOptions reagents are carefully formulated under strict quality standards to ensure reliable performance and lot-to-lot consistency.

Each reagent in a StockOptions kit is available in convenient concentration, making crystal setups quick and easy. Gone is the tedious task of finding and sourcing reagents, as well as costly and time-consuming reagent formulation.

Preformulated reagents also reduce the activation energy between the discovery of preliminary screen conditions and the task of setting experiments for optimization. The generation of custom screens or optimized conditions now simply involves pipetting StockOptions reagents from convenient kits.

StockOptions Salt contains 49 unique salts, preformulated at convenient stock concentrations, each in 10 ml volumes, all in a single kit with a 5" x 9" footprint that saves precious lab space. It is designed to help researchers improve the speed, accuracy, precision, and quality of the formulation of crystallization optimization solutions. Researchers can use the individual StockOptions reagents to formulate custom screen solutions or accurately reproduce standard screen solutions from Hampton Research crystallization kits. All one needs to do is select the reagent and pipet.

The convenience also reduces the chance of errors. Preformulated stocks remove calculation, measurement, and formulation errors. No more second guessing how the reagent was formulated, what specific chemical was used, when it was made, or how to precisely reproduce that reagent when it is gone and more reagent is required for additional setups.

StockOptions are cost-effective, time saving reagents. When a StockOptions kit is purchased, one is using reagents as preformulated stocks in reasonable volumes. You buy only the reagents you need, not a large container of raw material that may sit out on shelves for years to come. Waste is further reduced since there is no chance for formulation or measuring errors. When you need larger volumes, Hampton Research offers individual, preformulated, sterile filtered Optimize crystallization reagents which include salts, polymers, organics solvents, and buffers. Each of the 49 salts offered in StockOptions Salt is available individually as an Optimize™ crystallization grade reagent. Optimize, Custom Shop, StockOptions, and all Hampton Research kits are synergistic research tools.

StockOptions kits also lower the costs associated with making crystallization reagents since there is no need to purchase sterile filters, filtration devices, or pre-sterilized storage containers. Cost savings are also extended to labor since time can now be better utilized for sample production and purification or setting crystallization experiments.

References

1. *Purification, crystallization and preliminary crystallographic characterization of the caspase-recruitment domain of human Nod1. T. Srimathi, S. L. Robbins, R. L. Dubas, J.-H. Seo and Y. C. Park. Acta Cryst. (2007). F63, 21-23.*

Order Information

Each StockOptions Salt kit contains 49 unique reagents.

Cat. No.	Name	Description	Price
HR2-245	**StockOptions Salt**	10 ml, tube format	**$285.00**

APPLICATION

- Crystallization grade buffer stocks for crystal screening, optimization, & production

FEATURES

- Titrated, preformulated, & ready-to-use buffers
- pH 2.2 - 11.0
- Easy transition from screening to optimization and production
- Synergistic with Hampton Research screens, kits, & reagents
- Concentrated stocks
- Sterile filtered

STOCKOPTIONS PH KIT FORMULATION

Reagent Number	pH	Reagent	Titrant
1	2.2	1.0 M Citric acid	Sodium hydroxide
2	2.4	1.0 M Citric acid	Sodium hydroxide
3	2.6	1.0 M Citric acid	Sodium hydroxide
4	2.8	1.0 M Citric acid	Sodium hydroxide
5	3.0	1.0 M Citric acid	Sodium hydroxide
6	3.2	1.0 M Citric acid	Sodium hydroxide
7	3.4	1.0 M Citric acid	Sodium hydroxide
8	3.6	1.0 M Citric acid	Sodium hydroxide
9	3.8	1.0 M Citric acid	Sodium hydroxide
10	4.0	1.0 M Sodium acetate trihydrate	Hydrochloric acid
11	4.2	1.0 M Sodium acetate trihydrate	Hydrochloric acid
12	4.4	1.0 M Sodium acetate trihydrate	Hydrochloric acid
13	4.6	1.0 M Sodium acetate trihydrate	Hydrochloric acid
14	4.8	1.0 M Sodium acetate trihydrate	Hydrochloric acid
15	5.0	1.0 M Sodium citrate tribasic dihydrate	Hydrochloric acid
16	5.2	1.0 M Sodium citrate tribasic dihydrate	Hydrochloric acid
17	5.4	1.0 M Sodium citrate tribasic dihydrate	Hydrochloric acid
18	5.6	1.0 M Sodium citrate tribasic dihydrate	Hydrochloric acid
19	5.8	1.0 M Sodium citrate tribasic dihydrate	Hydrochloric acid
20	6.0	1.0 M Sodium cacodylate trihydrate	Hydrochloric acid
21	6.2	1.0 M Sodium cacodylate trihydrate	Hydrochloric acid
22	6.4	1.0 M Sodium cacodylate trihydrate	Hydrochloric acid
23	6.6	1.0 M Sodium cacodylate trihydrate	Hydrochloric acid
24	6.8	1.0 M Sodium cacodylate trihydrate	Hydrochloric acid
25	7.0	1.0 M HEPES sodium	Hydrochloric acid
26	7.2	1.0 M HEPES sodium	Hydrochloric acid
27	7.4	1.0 M HEPES sodium	Hydrochloric acid
28	7.6	1.0 M HEPES sodium	Hydrochloric acid
29	7.8	1.0 M HEPES sodium	Hydrochloric acid
30	8.0	1.0 M TRIS hydrochloride	Sodium hydroxide
31	8.2	1.0 M TRIS hydrochloride	Sodium hydroxide
32	8.4	1.0 M TRIS hydrochloride	Sodium hydroxide
33	8.6	1.0 M TRIS hydrochloride	Sodium hydroxide
34	8.8	1.0 M TRIS hydrochloride	Sodium hydroxide
35	9.0	0.5 M CAPSO	Sodium hydroxide
36	9.2	0.5 M CAPSO	Sodium hydroxide
37	9.4	0.5 M CAPSO	Sodium hydroxide
38	9.6	0.5 M CAPSO	Sodium hydroxide
39	9.8	0.5 M CAPSO	Sodium hydroxide
40	10.0	0.5 M CAPS	Sodium hydroxide
41	10.2	0.5 M CAPS	Sodium hydroxide
42	10.4	0.5 M CAPS	Sodium hydroxide
43	10.6	0.5 M CAPS	Sodium hydroxide
44	10.8	0.5 M CAPS	Sodium hydroxide
45	11.0	0.5 M CAPS	Sodium hydroxide

DESCRIPTION

StockOptions reagent kits are reagent tool boxes for the macromolecular crystal grower. They offer precisely formulated, high quality crystallization grade reagent stocks in convenient, cost-effective kits. The chemicals utilized in these kits are the same crystallization grade, ultra-high purity chemicals utilized in the Hampton Research kits such as Crystal Screen™ and Crystal Screen 2™. StockOptions reagents are carefully formulated under strict quality standards to ensure reliable performance and lot-to-lot consistency.

StockOptions pH is a convenient and complete kit with buffers pHed from 2.2 to 11.0 in increments of 0.2 pH units. Each reagent is available in convenient concentration, making crystal setups quick and easy. Gone is the tedious task of finding and sourcing reagents, as well as costly and time-consuming reagent formulation.

The generation of custom screens or optimized conditions now simply involves pipetting StockOptions pH reagents from one convenient kit. This portfolio of buffers between pH 2.2 and 11, along with wide arrays of salts, polymers, and organic solvents also stimulates creativity since all of the tools are readily and conveniently available. All one needs to do is select the reagent and pipet.

The convenience also reduces the chance of errors. Preformulated stocks remove calculation, measurement, and formulation errors. No more second guessing how the reagent was formulated, what specific chemical was used, when it was made, or how to precisely reproduce that reagent when it is gone and more reagent is required for additional setups.

StockOptions pH is cost-effective. When a StockOptions kit is purchased, one is using reagents as preformulated stocks in reasonable volumes. You buy only the reagents you need, not a large container of raw material that may sit out on shelves for years to come. When you need larger volumes, Hampton Research offers individual, preformulated, sterile filtered Optimize™ crystallization reagents which include salts, polymers, organics solvents, and buffers. Optimize, Custom Shop, StockOptions, and all Hampton Research kits are synergistic research tools.

References

1. *Structure of the ribosomal interacting GTPase YjeQ from the enterobacterial species Salmonella typhimurium. D. K. Stammers et al. Acta Cryst. (2007). F63, 922–928*
2. *Increasing the size of microcrystals by fine sampling of pH limits. A. McPherson. J. Appl. Cryst. (1995). 28, 362-365.*

Order Information

Each StockOptions pH kit contains 45 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 37 for further details.

Cat. No.	Name	Description	Price
HR2-241	**StockOptions pH**	10 ml, tube format	**$275.00**
HR2-941-**	**StockOptions pH Custom Shop**	185 ml	
	Reagents 1-39		**$139.00**
	Reagents 40-45		**$151.00**

** = reagent number 1-45

STOCKOPTIONS™ BUFFER

APPLICATION

- Focused, titrated buffer stocks for crystal screening, optimization & production

FEATURES

- pH range 2.2 - 6.5 (44 unique reagents)
- 0.1 pH unit increments for fine tuning pH
- Compatible with Hampton Research screens, kits & reagents
- 1.0 M concentrated stocks

DESCRIPTION

StockOptions Buffer kits are reagent toolboxes for the macromolecular crystal grower. StockOptions reagent kits offer precisely formulated, high quality crystallization grade reagent stocks in convenient, cost-effective kits. The chemicals utilized in StockOptions kits are the same crystallization grade, ultra-high purity chemicals utilized in the Hampton Research kits. StockOptions reagents are carefully formulated under strict quality standards to ensure reliable performance and lot-to-lot consistency.

Preformulated stocks remove calculation, measurement, and formulation errors. No more second guessing how the reagent was formulated, what specific chemical was used, when it was made, or how to precisely reproduce that reagent when it is gone and more reagent is required for additional setups.

STOCKOPTIONS™ CITRIC ACID

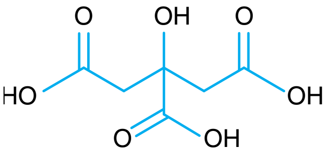

pH 2.2 - 6.5

StockOptions Citric Acid buffer kit is a preformulated, sterile filtered set of titrated buffer stocks. The reagents are supplied as 1.0 M stock solutions in 10 ml volumes. Each StockOptions Citric Acid buffer reagent is carefully titrated using Sodium hydroxide. StockOptions Citric Acid is comprised of 44 unique reagents covering the pH range of 2.2 to 6.5 in 0.1 pH unit increments.

References

1. *Increasing the size of microcrystals by fine sampling of pH limits. A. McPherson. J. Appl. Cryst. (1995). 28, 362-365.*

Order Information

Each StockOptions Citric Acid kit contains 44 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 37 for further details.

Cat. No.	Name	Description	Price
HR2-104	**StockOptions Citric Acid**	10 ml, tube format	**$195.00**
HR2-904-**	**StockOptions Citric Acid Custom Shop**	185 ml	**$144.00**

** = reagent number 1-44

STOCKOPTIONS™ SODIUM ACETATE

StockOptions Sodium Acetate buffer kit is a preformulated, sterile filtered set of titrated buffer stocks. The reagents are supplied as 1.0 M stock solutions in 10 ml volumes. Each StockOptions Sodium Acetate buffer reagent is carefully titrated using Hydrochloric acid. The kit is comprised of 21 unique reagents covering the pH range of 3.6 to 5.6 in 0.1 pH unit increments.

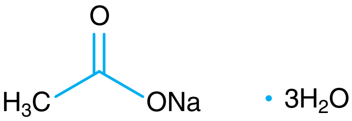

pH 3.6 - 5.6

References

1. *Crystal structures of reduced, oxidized, and mutated human thioredoxins: evidence for a regulatory homodimer. Weichsel, A; Gasdaska, JR; Powis, G; Montfort, WR. Structure 4, 735- 751, 1996.*
2. *Increasing the size of microcrystals by fine sampling of pH limits. A. McPherson. J. Appl. Cryst. (1995). 28, 362-365.*

Order Information

Each StockOptions Sodium Acetate kit contains 21 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 37 for further details.

Cat. No.	Name	Description	Price
HR2-233	**StockOptions Sodium Acetate**	10 ml, tube format	**$195.00**
HR2-933-**	**StockOptions Sodium Acetate Custom Shop**	185 ml	**$150.00**

** = reagent number 1-21

STOCKOPTIONS™ SODIUM CITRATE BUFFER

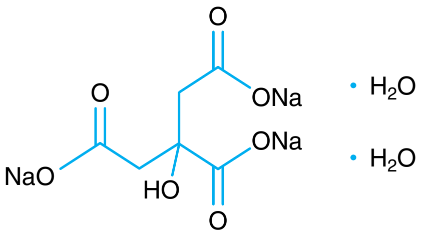

pH 4.2 - 6.5

StockOptions Sodium Citrate buffer kit is a preformulated, sterile filtered set of titrated buffer stocks. The reagents are supplied as 1.0 M stock solutions in 10 ml volumes. Each StockOptions Sodium Citrate buffer reagent is carefully titrated using Hydrochloric acid. The kit is comprised of 24 unique reagents covering the pH range of 4.2 - 6.5 in 0.1 pH unit increments.

Formulated using Sodium citrate tribasic dihydrate and HCl.
pKa1 = 3.1 pKa2 = 4.8 pKa3 = 5.4

References

1. *Increasing the size of microcrystals by fine sampling of pH limits. A. McPherson. J. Appl. Cryst. (1995). 28, 362-365.*

Order Information

Each StockOptions Sodium Citrate kit contains 24 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 37 for further details.

Cat. No.	Name	Description	Price
HR2-235	**StockOptions Sodium Citrate**	10 ml, tube format	**$195.00**
HR2-935-**	**StockOptions Sodium Citrate Custom Shop**	185 ml	**$173.00**

** = reagent number 1-24

STOCKOPTIONS™ SODIUM CACODYLATE BUFFER

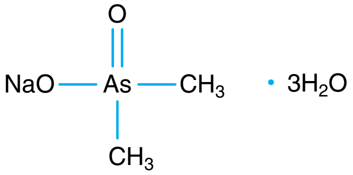

$(CH_3)_2As(O)ONa \cdot 3H_2O$

pH 5.1 - 7.4

StockOptions Sodium Cacodylate buffer kit is a preformulated, sterile filtered set of titrated buffer stocks. The reagents are supplied as 1.0 M stock solutions in 10 ml volumes. Each StockOptions Sodium Cacodylate buffer reagent is carefully titrated using Hydrochloric acid. The kit is comprised of 24 unique reagents covering the pH range of 5.1 to 7.4 in 0.1 pH unit increments.

References

1. *Increasing the size of microcrystals by fine sampling of pH limits. A. McPherson. J. Appl. Cryst. (1995). 28, 362-365.*

Order Information

Each StockOptions Sodium Cacodylate kit contains 24 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 37 for further details.

Cat. No.	Name	Description	Price
HR2-239	**StockOptions Sodium Cacodylate**	10 ml, tube format	**$250.00**
HR2-939-**	**StockOptions Sodium Cacodylate Custom Shop**	185 ml	**$221.00**

** = reagent number 1-24

STOCKOPTIONS™ MES BUFFER

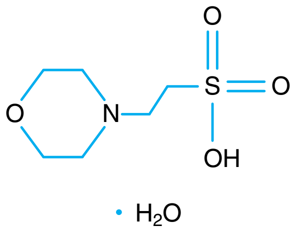

pH 5.2 - 7.1

StockOptions MES buffer kit is a preformulated, sterile filtered set of titrated buffer stocks. The reagents are supplied as 1.0 M stock solutions in 10 ml volumes. Each StockOptions MES buffer reagent is carefully titrated using Sodium hydroxide. The kit is comprised of 20 unique reagents covering the pH range of 5.2 to 7.1 in 0.1 pH unit increments.

References

1. *Characterization, crystallization and preliminary crystallographic analysis of human recombinant cyclooxygenase-2. Marco, SD; Priestle, JP; Grutter, MG; Wennogle, LP; Boyar, W. Acta Crystallogr D 53, 224- 226, 1997.*
2. *Increasing the size of microcrystals by fine sampling of pH limits. A. McPherson. J. Appl. Cryst. (1995). 28, 362-365*

Order Information

Each StockOptions MES kit contains 20 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 37 for further details.

Cat. No.	Name	Description	Price
HR2-243	**StockOptions MES**	10 ml, tube format	**$195.00**
HR2-943-**	**StockOptions MES Custom Shop**	185 ml	**$138.00**

** = reagent number 1-20

STOCKOPTIONS™ BIS-TRIS

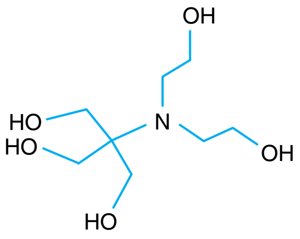

pH 5.5 - 7.5

StockOptions Bis-Tris buffer kit is a preformulated, sterile filtered set of titrated buffer stocks. The reagents are supplied as 1.0 M stock solutions in 10 ml volumes. Each StockOptions Bis-Tris buffer reagent is carefully titrated using Hydrochloric acid. The kid is comprised of 21 unique reagents covering the pH range of 5.5 to 7.5 in 0.1 pH unit increments.

References

1. *Increasing the size of microcrystals by fine sampling of pH limits. A. McPherson. J. Appl. Cryst. (1995). 28, 362-365.*

Order Information

Each StockOptions Bis-Tris kit contains 21 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 37 for further details.

Cat. No.	Name	Description	Price
HR2-106	**StockOptions Bis-Tris**	10 ml, tube format	**$195.00**
HR2-906-**	**StockOptions Bis-Tris Custom Shop**	185 ml	**$151.00**

** = reagent number 1-21

STOCKOPTIONS™ HEPES

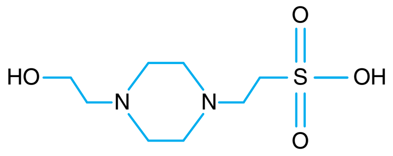

pH 6.8 - 8.2

Each kit contains 15 unique 10 ml volumes of 1.0 M HEPES buffer solution titrated to the indicated pH using NaOH.

References

1. *The expression, purification and crystallization of the epsilon subunit of the F1 portion of the ATPase of Escherichia coli. Codd, R; Cox, GB; Guss, JM; Solomon, RG; Webb, D. J Mol Biol 228, 306- 309, 1992.*
2. *Increasing the size of microcrystals by fine sampling of pH limits. A. McPherson. J. Appl. Cryst. (1995). 28, 362-365.*

Order Information

Each StockOptions HEPES kit contains 15 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 37 for further details.

Cat. No.	Name	Description	Price
HR2-102	**StockOptions HEPES**	10 ml, tube format	**$195.00**
HR2-902-**	**StockOptions HEPES Custom Shop**	185 ml	**$172.00**

** = reagent number 1-15

STOCKOPTIONS™ SODIUM HEPES

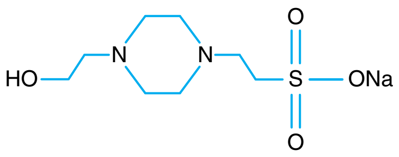

pH 6.8 - 8.2

StockOptions Sodium HEPES buffer kit is a preformulated, sterile filtered set of titrated buffer stocks. The reagents are supplied as 1.0 M stock solutions in 10 ml volumes. Each StockOptions Sodium HEPES buffer reagent is carefully titrated using Hydrochloric acid. The kit is comprised of 15 unique reagents covering the pH range of 6.8 to 8.2 in 0.1 pH unit increments.

References

1. *Increasing the size of microcrystals by fine sampling of pH limits. A. McPherson. J. Appl. Cryst. (1995). 28, 362-365.*

Order Information

Each StockOptions Sodium HEPES kit contains 15 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 37 for further details.

Cat. No.	Name	Description	Price
HR2-231	**StockOptions Sodium HEPES**	10 ml, tube format	**$195.00**
HR2-931-**	**StockOptions Sodium HEPES Custom Shop**	185 ml	**$190.00**

** = reagent number 1-15

STOCKOPTIONS™ TRIS

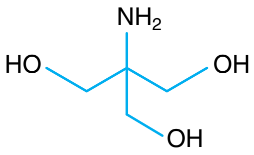

pH 7.0 - 9.0

StockOptions Tris buffer kit is a preformulated, sterile filtered set of titrated buffer stocks. The reagents are supplied as 1.0 M stock solutions in 10 ml volumes. Each StockOptions Tris buffer reagent is carefully titrated using Hydrochloric acid. The kit is comprised of 21 unique reagents covering the pH range of 7.0 to 9.0 in 0.1 pH unit increments.

References

1. *Crystallization and Preliminary X-ray Diffraction Analysis of the ArsC Protein from the Escherichia coli Arsenical Resistance Plasmid, R773. deMel, VSJ; Doyle, MA; Gladysheva, TB; Oden, KL; Martin, PD; Rosen, BP; Edwards, BFP. J Mol Biol 242, 701- 702, 1994.*
2. *Increasing the size of microcrystals by fine sampling of pH limits. A. McPherson. J. Appl. Cryst. (1995). 28, 362-365.*

Order Information

Each StockOptions Tris kit contains 21 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 37 for further details.

Cat. No.	Name	Description	Price
HR2-100	**StockOptions Tris**	10 ml, tube format	**$195.00**
HR2-900-**	**StockOptions Tris Custom Shop**	185 ml	**$169.00**

** = reagent number 1-21

STOCKOPTIONS™ TRIS HYDROCHLORIDE

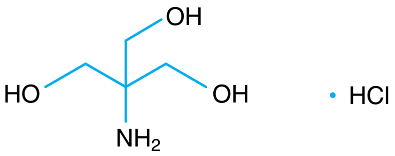

pH 7.0 - 9.0

StockOptions Tris Hydrochloride buffer kit is a preformulated, sterile filtered set of titrated buffer stocks. The reagents are supplied as 1.0 M stock solutions in 10 ml volumes. Each StockOptions Tris Hydrochloride buffer reagent is carefully titrated using Sodium hydroxide. The kit is comprised of 21 unique reagents covering the pH range of 7.0 to 9.0 in 0.1 pH unit increments.

References

1. *Increasing the size of microcrystals by fine sampling of pH limits. A. McPherson. J. Appl. Cryst. (1995). 28, 362-365.*

Order Information

Each StockOptions Tris Hydrochloride kit contains 21 unique reagents. To order individual reagents, use Custom Shop™ catalog number listed below. Refer to page 37 for further details.

Cat. No.	Name	Description	Price
HR2-237	**StockOptions Tris Hydrochloride**	10 ml, tube format	**$195.00**
HR2-937-**	**StockOptions Tris Hydrochloride Custom Shop**	185 ml	**$139.00**

** = reagent number 1-21

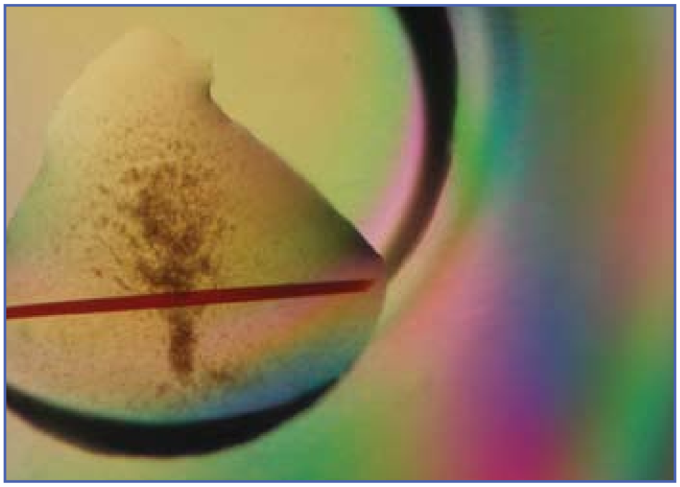

"The Lord of the Ring" - Crystal of bacterial di-haem Cyt c4 grown using the Hampton Research Additive Screen.

Ivana Tomcova, Institute of Physical Biology, Academy of Science of the Czech Republic.

CRYOPRO™

APPLICATION

- Cryoprotectant reagent kit

FEATURES

- 36 unique cryoprotectants
- Includes cryoprotectant tutorial
- Preformulated, ready-to-use
- Polyols, organics, oils, polymers, sugars, & salts

FORMULATIONS

Reagent Number	Cryoprotectant
1	100% Glycerol
2	100% Ethylene glycol
3	100% Polyethylene glycol 200
4	100% Polyethylene glycol 400
5	80% v/v Polyethylene glycol 600
6	60% w/v Polyethylene glycol 4,000
7	50% w/v Polyvinylpyrrolidone K 15
8	100% (+/-)-2-Methyl-2,4-pentanediol
9	6.0 M 1,6-Hexanediol
10	100% 1,2-Propanediol
11	100% Paratone-N
12	100% Paraffin Oil
13	100% NVH Oil
14	100% Dimethyl sulfoxide (DMSO)
15	100% 2-Propanol
16	100% Ethanol
17	100% Methanol
18	70% w/v D-(+)-Sucrose
19	35% w/v meso-Erythritol
20	70% w/v Xylitol
21	15% w/v myo-Inositol
22	20% w/v D-(+)-Raffinose pentahydrate
23	50% w/v D-(+)-Trehalose dihydrate
24	70% w/v D-(+)-Glucose monohydrate
25	100% 2,3-Butanediol
26	100% L-(+)-2,3-Butanediol
27	5.0 M Lithium acetate dihydrate
28	10.0 M Lithium chloride
29	4.0 M Lithium formate monohydrate
30	8.0 M Lithium nitrate
31	2.0 M Lithium sulfate monohydrate
32	3.4 M Sodium malonate pH 7.0
33	3.5 M Magnesium acetate tetrahydrate
34	5.0 M Sodium chloride
35	7.0 M Sodium formate
36	7.0 M Sodium nitrate

DESCRIPTION

As in selecting reagents for crystallization, the selection of a suitable cryoprotectant involves some trial and error as well as a screening.[1-15] A suitable cryoprotectant, when mixed with the crystal and crystallization reagent, will cool to cryogenic temperature without ice formation and damage to the crystal. To assay for the proper concentration of cryoprotectant in the reagent used to grow the crystal, one can mix the cryoprotectant with the crystallization reagent and employ the desired cooling method. For example, place the solution in a CryoLoop™ and place the CryoLoop in a cryostream. Observe for ice formation either visually or with x-ray diffraction. Upon cooling, a transparent drop and x-ray diffraction pattern, free of powder diffraction rings or "ice rings" indicates success. The appearance of a cloudy drop or "ice rings", indicates an inappropriate cryoprotectant concentration or cryoprotectant. Incrementally increase the concentration and/or composition of the cryoprotectant serially 5 to 10% and repeat until the cooled drop remains clear while in the cryostream. Once a clear drop is achieved in the cryostream, this is typically a good starting point for cryopreservation of the crystal.

Some crystals can simply be dipped or washed quickly in a simple cryoprotectant such as 30% glycerol for successful cryopreservation. However, when all else fails, a rational assay of each cryoprotectant with incremental increases in cryoprotectant concentration, as well as a test of mixtures (i.e. a mixture of sugars or a sugar mixed with Ethylene glycol) may be required to determine the best cryoprotectant for a crystal.

Each CryoPro - Cryoprotectant kit contains 36 unique cryoprotectants including organics, oils, polyols, salts, sugars, and polymers. 35 of the 36 reagents are supplied in 10 ml volumes. A single reagent L-(+)-2,3-Butanol is supplied in a 0.2 ml aliquot. CryoPro is convenient and cost-effective.

Ready-to-use cryoprotectants are sterile filtered and formulated with ultra-pure water, using high purity reagents. Many of the individual cryoprotectants are available as a 100 ml or 200 ml Optimize™ reagent. Please refer to the CryoPro Formulations pdf file to find the catalog number for the equivalent Optimize.

References

1. *Boutron, P. (1987). Non-equilibrium formation of ice in aqueous solutions: efficiency of polyalcohol solutions for vitrification. In: Pegg, D.E. & Karow, A.M. Jr. (eds). The biophysics of organ prese*
2. *Garman, E.F., & Mitchell, E.P. (1996). Glycerol concentrations required for cryoprotection of 50 typical protein crystallization conditions. J. Appl. Cryst. 29, 584-587.*
3. *Garman, E.F., & Schneider, T.R. (1997). Macromolecular Cryocrystallography. J. Appl. Cryst. 30, 211-237.*
4. *Hope, H. (1988). Cryocrystallography of biological macromolecules: a generally applicable method. Acta Cryst. B 44, 22-26.*
5. *Kottke, T., & Stalke, D. (1993). Crystal handling at low temperatures. J. Appl. Cryst. 26, 615-619.*
6. *Kwong, P.D. Liu, Y. (1999). Use of cryoprotectants in combination with immiscible oils for flash cooling macromolecular crystals. J. Appl. Cryst. 32, 102-105.*
7. *Mehl, P. (1989). Experimental dissection of devitrification in aqueous solutions in 1,3-butanediol. Cryobiology. 26, 567-568.*
8. *Parkin, S., & Hope, H. (1998). Macromolecular cryocrystallography: Cooling, mounting, storage and transportation of crystals. J. Appl. Cryst. 31, 945-953.*
9. *Petcock, J.M., Wang, Y.-F., DuBois, G.C., Harrison, R.W., & Weber, I.T. (2001). Effects of different post-crystallization soaking conditions on the diffraction of Mtcp1 crystals. Acta Cryst. D57, 763-*
10. *Petsko, G.A. (1975). J. Mol. Biol. 96, 381-392.*
11. *Rodgers, D.W. (1994). Cryocrystallography. Structure. 2, 1135-1140. 12. Schneider, T.R. (1997). Cryocrystallography of biological macromolecules. Acta Physica Polonica A. 91, 739-744.*
12. *Teng, T.-Y. (1990). J. Appl. Cryst. 23, 387-391.*
13. *Walker, L.J., Moreno, P.O., Hope, H. (1998). Cryocrystallography: effect of cooling medium on sample cooling rate. J. Appl. Cryst. 31, 954-965.*
14. *Watenpaugh, K.D. (1991). Curr. Op. Struct. Biol. 1, 1012.*
15. *Schneider, T.R. (1997), Cryocrystallography of biological macromolecules. Acta Physica Polonica A. 91, 739-744.*

Order Information

Each CryoPro kit contains 36 unique reagents.

Cat. No.	Name	Description	Price
HR2-132	CryoPro	10 ml, tube format	$285.00

Crystals shaped like a dragonfly wing.

Yi-Wei Change, Institute of Molecular Biology,
Academia Sinica, Taiwan.

"Butterfly"-crystal of hydroxynitrile lyase from Prunus amygdalus grown with the Hampton Research Grid Screen PEG/LiCl.

Andreas Giessauf at the Institute for Chemie,
Karl-Franzens-Universität Graz, Austria.

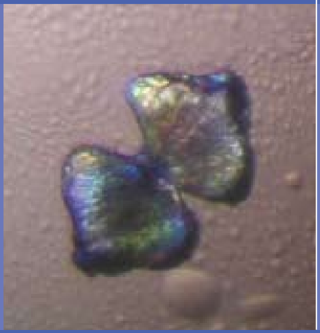

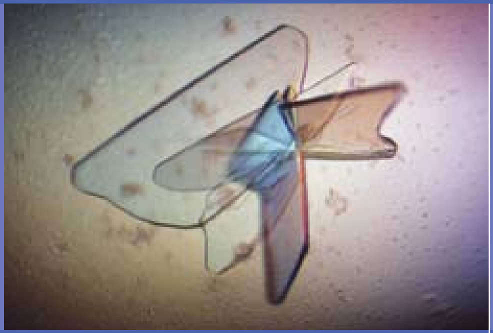

Crystals of the FERM domain of Focal Adhesion Kinase. The final precipitant was derived from an initial hit obtained with condition 17 from Hampton Research Crystal Screen.

Derek F.J. Ceccarelli, Samuel Lunenfeld Research Institute,
Mount Sinai Hospital, Toronto, Ontario, Canada.

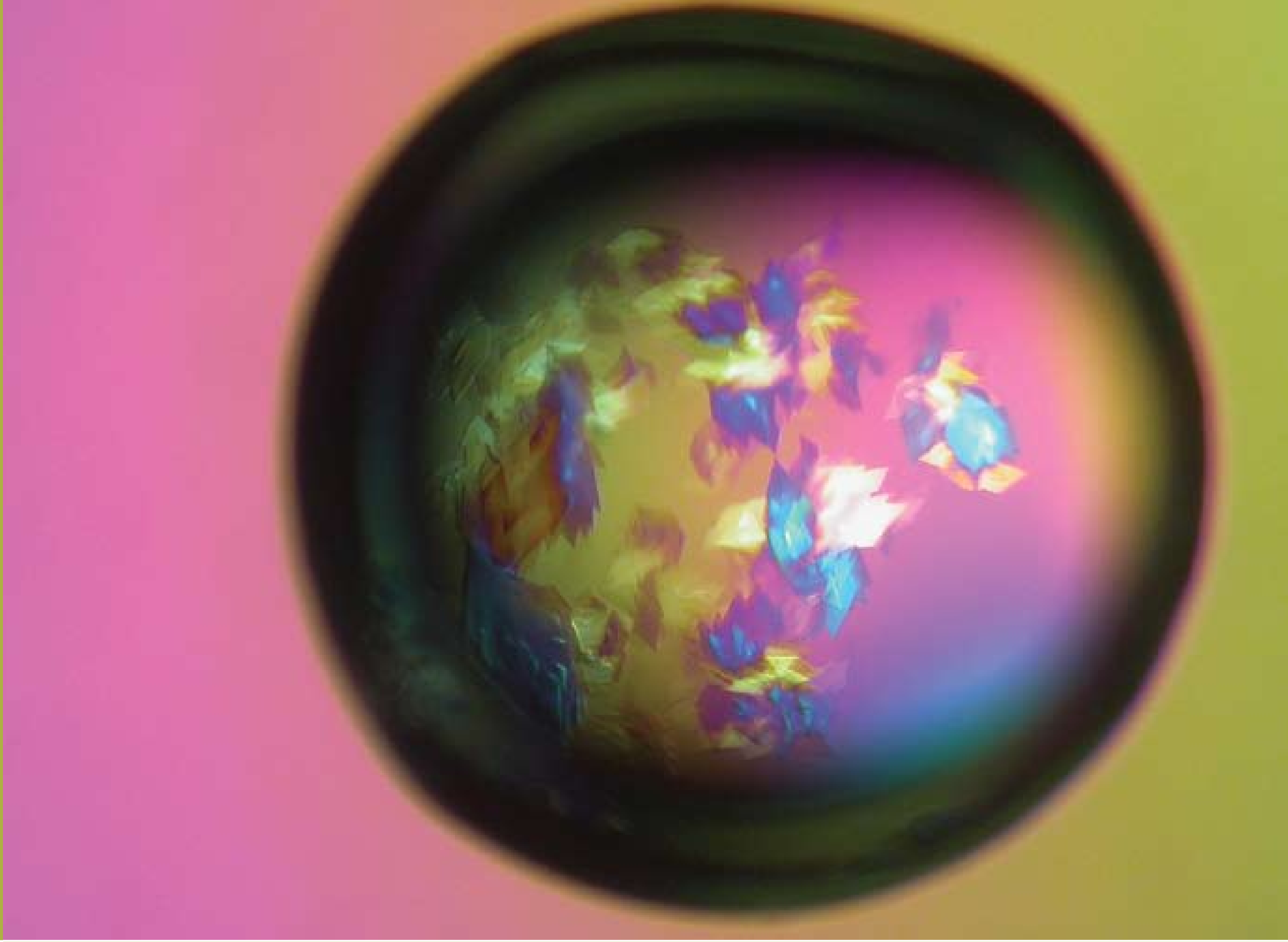

CRYSTALLIZATION PLATES, HARDWARE & ACCESSORIES

Crystals of a frameshift promoting RNA pseudoknot from the coronavirus IBV, grown using the Hampton Research Natrix screen.

Simon Pennell, MRC National Institute for Medical Research, United Kingdom.

TABLE OF CONTENTS

HAMPTON
RESEARCH

PAGES

72 - 77 24 WELL CRYSTALLIZATION PLATES
78 - 79 48 WELL CRYSTALLIZATION PLATES
79 72 WELL CRYSTALLIZATION PLATES
80 - 90 96 WELL CRYSTALLIZATION PLATES
90 384 WELL CRYSTALLIZATION PLATES
91 - 95 COVER SLIDES & RELATED TOOLS
95 GLASS SITTING DROP RODS™
96 MICRO-BRIDGES®
97 9 WELL GLASS PLATE
97 - 98 SEALANTS, SEALING GREASE & OILS
99 - 101 SEALING FILMS, TAPES, MATS & COVERS
102 - 103 DIALYSIS BUTTONS™, MEMBRANES & APPLICATORS
104 GRANADA CRYSTALLIZATION BOX® & CAPILLARIES

24 WELL CRYSTALLIZATION PLATES

		SEALS							
Plate	# Wells	12 mm Cover Slide	18 mm Cover Slide	22 mm Cover Slide	Crystal Clear Sealing Film	ClearSeal Film	1.88" Tape	3" Tape	AlumaSeal II Film
Cryschem™	24						X		
Cryschem™ 24-1 SBS	24				X	X		X	
VDX™	24			X					
VDXm™	24		X						
Greiner ComboPlate™	24		X						
Linbro®	24			X					
Intelli-Plate™ 24 Well	24				X	X		X	
Intelli-Plate™ 48 Well	48				X	X		X	
VDX48™	48	X							
Douglas Instruments CrystalClear Strips™	96				X	X		X	
Corning®	96				X	X		X	
Greiner CrystalQuick™	96				X	X		X	
Intelli-Plate™	96				X	X		X	
MRC/Swissci	96				X	X		X	
MASTERBLOCK®	96								X

CRYSCHEM™ PLATE

APPLICATIONS

- Sitting drop crystallization
- Heavy atom soaks
- Seeding

FEATURES

- Optically clear, concave sitting drop post
- Smooth, flat top for easy seal

MATERIAL

- Optically clear polystyrene

Top

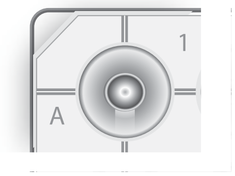

Side

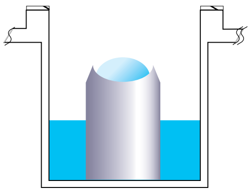

DESCRIPTION

The Cryschem Plate is a 24 well sitting drop plate that is sealed with clear sealing tape. The plate is supplied with a raised cover so the plate can also be sealed with vacuum grease and plain glass or plastic cover slides (22 mm diameter). The stackable, 24 well plates have distortion-free, flat bottom wells and a concave sitting drop post in the center of the well for concentric equilibration of the drop with the reagent. Each plate is individually sealed. The top of the plate and each of the 24 wells is a single, flat plane for smooth and easy seal application. Well spacing provides for a large seal area between wells to prevent cross contamination and evaporation. Approximate size: 15.1 cm x 10.6 cm x 2.2 cm. Approximate well size: 1.6 cm x 1.5 cm. Typical fill volume: 500 to 1,000 μl. Well capacity: 1.5 ml. Maximum drop volume on post: 40 μl. Recommended seal: HR3-511 Crystal Clear Sealing Tape (1.88 inch x 43.7 yard roll, with cutter) or HR4-511 Crystal Clear Sealing Tape (1.88 inch x 60 yard roll, without cutter).

The Cryschem 24-1 SBS Plate is a smaller, SBS footprint 24 well sitting drop microplate that is sealed with tape or film. The plate is supplied with a raised cover so the plate can also be sealed with vacuum grease and plain glass or plastic cover slides (18 mm diameter). The stackable, 24 well plates have distortion-free, flat bottom wells and a concave sitting drop post in the center of the well for concentric equilibration of the drop with the reagent. Each plate is individually sealed. The top of the plate and each of the 24 wells is a single, flat plane for smooth and easy seal application. Well spacing provides for a large seal area between wells to prevent cross contamination and evaporation. Approximate size: 127.8 mm x 85.5 mm x 14.4 mm. Typical fill volume: 500 to 700 μl. Maximum drop volume on post: 12 μl. Recommended seal: HR3-609 Crystal Clear Sealing Film (100 pack), HR4-521 ClearSeal Film (100 pack) or HR4-506 Crystal Clear Sealing Tape (3 inch x 55 yard roll).

References

1. *Purification, crystallization and preliminary X-ray crystallographic analysis of ST1022, a putative member of the Lrp/AsnC family of transcriptional regulators isolated from Sulfolobus tokodaii strain 7. Shigeyuki Yokoyama et al. Acta Cryst. (2007). F63, 964–966*

Order Information

Cat. No.	Name	Description	Price
HR3-158	**Cryschem Plate**	24 plate case	**$95.00**
HR3-160	**Cryschem Plate**	100 plate case	**$350.00**
HR1-002	**Cryschem 24-1 SBS Plate**	50 plate case	**$150.00**

- Hanging Drop Crystallization

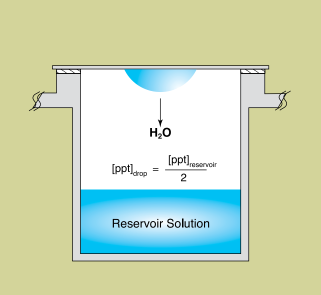
H$_2$O
[ppt]$_{drop}$ = [ppt]$_{reservoir}$ / 2
Reservoir Solution

- Sitting Drop Crystallization with Micro-Bridges or Glass Rods

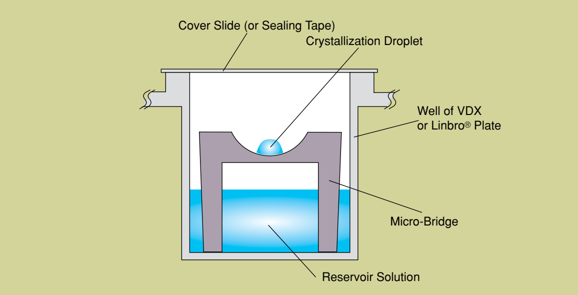
Cover Slide (or Sealing Tape)
Crystallization Droplet
Well of VDX or Linbro® Plate
Micro-Bridge
Reservoir Solution

- Dialysis Crystallization with Dialysis Buttons

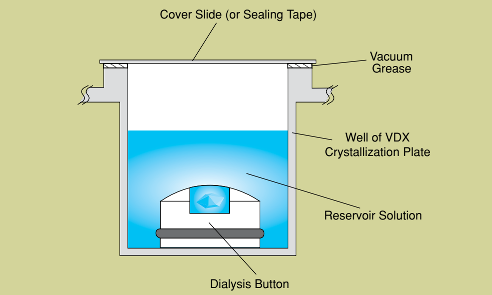
Cover Slide (or Sealing Tape)
Vacuum Grease
Well of VDX Crystallization Plate
Reservoir Solution
Dialysis Button

INTELLI-PLATE™ 24-4 WELL

APPLICATION

- Sitting drop crystallization

FEATURES

- 4 drop wells per reservoir
- Optically clear polystyrene
- Alphanumeric labeled drop wells visible under microscope

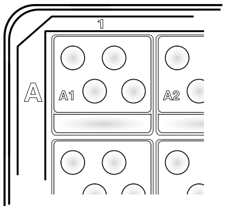

Intelli-Plate™ 24-4 well

DESCRIPTION

The Intelli-Plate 24-4 is a 24 well sitting drop plate for crystallization screening and optimization. It features 4 reagent wells along the y-axis (A-D) and 6 reagent wells along the x-axis (1-6). The reagent reservoir is typically filled with 250 μl and is capable of holding between 200 and 650 μl. The 4 drop wells are arranged at the top of each well and can hold up to 8 μl (5 μl for hydrophobic reagents such as MPD).

The Hampton Research ClearSeal Film™ and Sealing Film Applicator are used to seal the Intelli-Plate when a sitting drop experiment is performed. The advanced film is a transparent, polyolefin film with clear, pressure-sensitive, silicone adhesive. The film is pre-cut to perfectly fit 96 well crystallization plates and is complete with 1 cm wide, perforated end tabs which makes handling and positioning of the film easy, without fear of fingerprints getting into the view field. The Sealing Film Applicator is a small, hand-held applicator which should be used to create a consistent film seal across the entire crystallization plate. The Intelli-Plate can also be sealed using one of the clear sealing tapes supplied in a roll format.

The height of all Intelli-Plate 24-4 is 0.560 inches. The Art Robbins Instruments catalog number for this plate is 102-0004-00.

Order Information

Cat. No.	Name	Description	Price
HR3-114	**Intelli-Plate 24-4 well**	40 plate case	**$200.00**

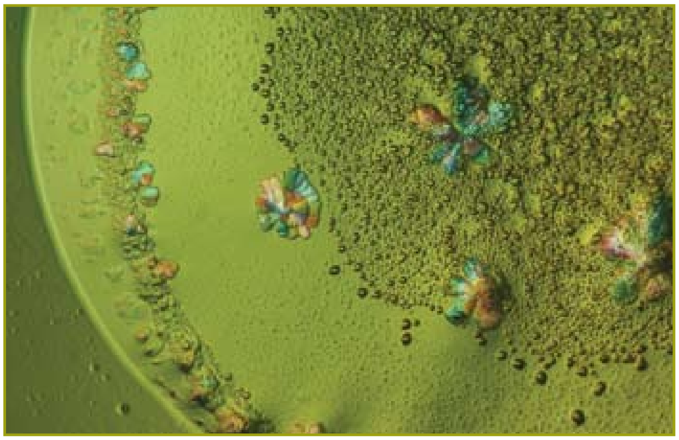

Proteins crystals.

Kasumi Kobayashi, Taiji Nakae and Hiroyuki Akama.
The Kitasato Institute, Kanagawa, Japan.

24 WELL CRYSTALLIZATION PLATES

VDX™ PLATE

APPLICATIONS

- Hanging drop crystallization
- Sitting drop crystallization with Micro-Bridges® or Glass Sitting Drop Rods™
- Dialysis crystallization with Dialysis Buttons™

FEATURES

- Compatible with 22 mm diameter square and circle cover slides
- 24 well plate with raised cover
- Extremely versatile and cost-effective crystallization platform
- Optically clear plastic

DESCRIPTION

24 well crystallization plate for hanging drop or sitting drop vapor diffusion crystallization (when used with Micro-Bridges or Glass Rods), or Dialysis crystallization (when used with Dialysis Buttons). Stackable, optically clear plastic 24 well plates with raised covers (to allow room for cover slides) and flat bottoms for exceptional optics. Raised, wide rings around each reservoir (well) minimize cross contamination and allow each well to be individually sealed with 22 mm diameter circle or square cover slides. Plates are individually wrapped and supplied without sealant. 15.0 cm x 10.8 cm plate footprint is convenient for manual pipetting and well access while at the same time compatible with some automated liquid handling systems. Approximate dimensions: 15.0 cm x 10.6 cm x 2.2 cm. Approximate well size: 1.7 cm x 1.6 cm. Typical well volume: 500 to 1,000 μl. Well capacity: 3.5 ml.

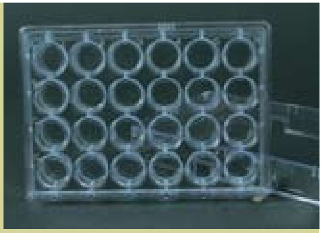

Order Information

Cat. No.	Name	Description	Price
HR3-142	**VDX Plate**	24 plate case	**$65.00**
HR3-140	**VDX Plate**	100 plate case	**$250.00**

VDX™ PLATE WITH SEALANT

APPLICATIONS

- Hanging drop crystallization
- Sitting drop crystallization with Micro-Bridges® or Glass Sitting Drop Rods™
- Dialysis crystallization with Dialysis Buttons

FEATURES

- Sealant applied to 24 wells, ready to seal with cover slides
- Compatible with 22 mm diameter circle and square cover slides
- 24 well plate with raised cover
- Optically clear plastic
- Extremely versatile and cost-effective crystallization platform

DESCRIPTION

24 well crystallization plate for hanging drop or sitting drop vapor diffusion crystallization (when used with Micro-Bridges or Glass Rods), or dialysis crystallization (when used with Dialysis Buttons). Stackable, optically clear plastic 24 well plates with raised covers (to allow room for cover slides) and flat bottoms for exceptional optics. Raised, wide rings around each reservoir (well) minimize cross contamination and allow each well to be individually sealed with 22 mm diameter circle or square cover slides. Plates are individually wrapped and supplied with applied sealant. 15.0 cm x 10.8 cm plate footprint is convenient for manual pipetting and well access while at the same time compatible with some automated liquid handling systems. Approximate dimensions: 15.0 cm x 10.6 cm x 2.2 cm. Approximate well size: 1.7 cm x 1.6 cm. Typical well volume: 500 to 1,000 μl. Well capacity: 3.5 ml.

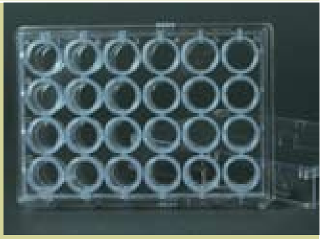

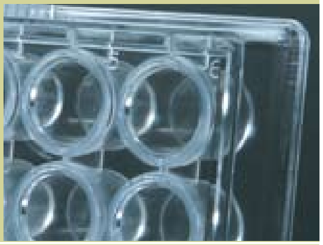

Order Information

Cat. No.	Name	Description	Price
HR3-172	**VDX Plate with sealant**	24 plate case	**$110.00**
HR3-170	**VDX Plate with sealant**	100 plate case	**$400.00**

VDXm™ PLATE

APPLICATION

- Hanging drop crystallization

FEATURES

- 24 well plate with raised cover
- Compatible with 18 mm diameter circle cover slides
- Optically clear plastic
- 128 mm x 85 mm microplate footprint

DESCRIPTION

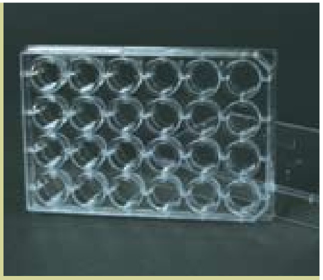

24 well plate for hanging drop vapor diffusion crystallization with a microplate footprint. Stackable, optically clear plastic, 24 well plates with raised covers (to allow room for cover slides) and flat bottoms for exceptional optics. Raised, wide rings around each reservoir (well) minimize cross contamination and allow each well to be individually sealed with 18 mm diameter circle cover slides. Plates are individually wrapped and supplied without sealant. The microplate footprint is space-saving and compatible with automated liquid handling systems. Approximate dimensions: 128 mm x 85 mm. Approximate well ID: 14.4 mm. Typical well volume: 100 to 500 μl.

Order Information

Cat. No.	Name	Description	Price
HR3-108	**VDXm Plate without sealant**	50 plate case	**$130.00**

VDXm™ PLATE WITH SEALANT

APPLICATION

- Hanging drop crystallization

FEATURES

- 24 well plate with raised cover
- Sealant applied to 24 wells, ready to seal with cover slides
- Compatible with 18 mm diameter circle cover slides
- Optically clear plastic
- 128 mm x 85 mm microplate footprint

DESCRIPTION

24 well plate with sealant for hanging drop vapor diffusion crystallization with a 128 mm x 85 mm microplate footprint. Stackable, optically clear plastic, 24 well plates with raised covers (to allow room for cover slides) and flat bottoms for exceptional optics. Raised, wide rings around each reservoir (well) minimize cross contamination and allow each well to be individually sealed with 18 mm diameter circle cover slides. Plates are individually wrapped and supplied with applied sealant. The microplate footprint is space-saving and compatible with automated liquid handling systems. Approximate dimensions: 128 mm x 85 mm. Approximate well ID: 14.4 mm. Typical well volume: 100 to 500 μl.

Hanging Drop Crystallization

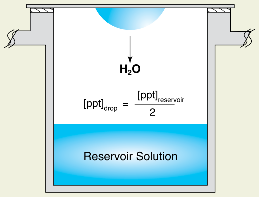

Order Information

Cat. No.	Name	Description	Price
HR3-306	**VDXm Plate with sealant**	50 plate case	**$210.00**

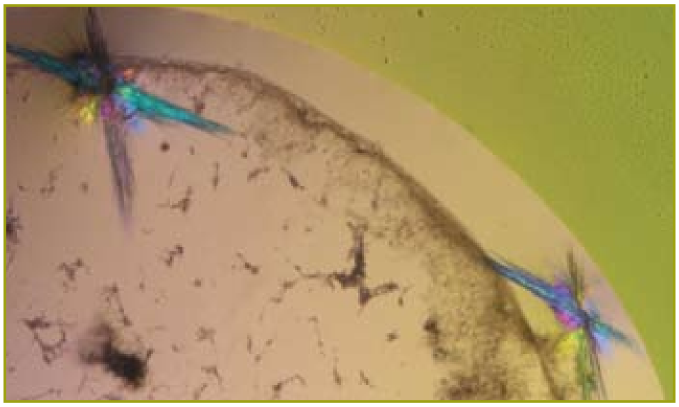

Crystals of a nucleoporin complex.

James Partridge, Massachusetts Institute of Technology, USA.

GREINER COMBOPLATE™ AND CRYSTALBRIDGE™

APPLICATION

- Hanging and sitting drop crystallization

FEATURES

- 24 reagent wells
- Sitting drop when used with CrystalBridge
- ComboPlate has flat, raised rings about each well to minimize spills and cross contamination
- ComboPlate seals with 18 mm glass or plastic cover slides
- Slightly raised protection lid

MATERIALS

- Polystyrene

DESCRIPTION

The Greiner ComboPlate is a 24 well plate with a raised cover in an SBS microplate footprint. The plate, supplied without sealant, can be sealed with an 18 mm diameter cover slide and sealant for hanging drop crystallization. When used with the Greiner CrystalBridge, the ComboPlate can be used for sitting drop crystallization and sealed with sealing film/tape or 18 mm diameter cover slide and sealant. Dimensions: 127.76 mm x 85.48 mm x 18.8 mm. Well depth: 16.28 mm. 3.3 ml well capacity. Typical fill volume: 0.5 - 1.0 ml. Supplied with cover. The Greiner CrystalBridge is a polystyrene sitting drop pedestal with a concave drop well (45 μl capacity) and is designed specifically for the ComboPlate.

Order Information

Cat. No.	Name	Description	Price
HR3-200	**24 Well ComboPlate with cover, Greiner 662150**	24 plate case	**$61.00**
HR3-150	**CrystalBridge for ComboPlate, Greiner 662145**	100 pack	**$78.00**
HR3-154	**CrystalBridge for ComboPlate, Greiner 662145**	400 pack	**$276.00**

LINBRO® PLATE

APPLICATIONS

- Hanging drop crystallization
- Sitting drop crystallization with Micro-Bridges® or Glass Sitting Drop Rods™
- Dialysis crystallization with Dialysis Buttons™

FEATURES

- 24 well plate with cover
- Compatible with 22 mm diameter circle and square cover slides
- Optically clear plastic

DESCRIPTION

The Linbro 24 well plate with cover is made of clear, rigid, polystyrene and used primarily for hanging drop vapor diffusion crystallization. Sitting drop vapor diffusion may be performed when used with Micro-Bridges or Glass Sitting Drop Rods. Plates supplied without sealant. Compatible with 22 mm diameter circle and square cover slides. Each plate is individually wrapped and sterile. Wide, raised rings around each reservoir minimizes cross contamination and allows each individual well to be sealed using vacuum grease, DC 7 Release Compound or immersion oil and a siliconized or plain cover slide. Cover is not raised but can be raised by placing wax or mounting clay in each of the four corners of the cover. The stackable plates have distortion-free, flat-bottom wells of fine clarity. Wells are identified by lettered rows A through D and numbered columns 1 through 6. Approximate dimensions: 15.0 cm x 10.8 cm x 2.2 cm. Approximate well size: 1.7 cm x 1.6 cm. Typical well volume: 700 to 1,000 μl. Individual well capacity: 3.5 ml

References

1. *Expression, purification and crystallization of 2-oxo-hept-4-ene-1,7-dioate hydratase (HpcG) from Escherichia coli C. T. Adachi, A. Izumi, D. Rea, S.-Y. Park, J. R. H. Tame and D. I. Roper. Acta Cryst. (2006). F62, 1010-1012.*

Order Information

Cat. No.	Name	Description	Price
HR3-110	**Linbro Plate**	50 plate case	**$302.00**
HR3-112	**Linbro Plate**	100 plate case	**$592.00**

VDX™ PLATE (MODULAR)

APPLICATIONS

- Sitting, hanging drop, and dialysis crystallization
- Modular wells for temperature studies[1]

FEATURES

- Removable wells
- 24 well plate with raised cover
- Wells available with or without sealant
- Use with 22 mm diameter circle and square cover slides
- Optically clear plastic

DESCRIPTION

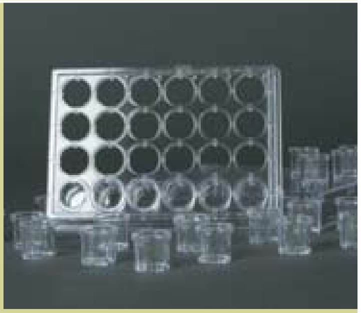

The Modular VDX Plate is a standard 24 well VDX Plate™ with individually removable wells. The modular format, consisting of a VDX frame and removable wells creates a very flexible plate format for screening and optimization of crystallization experiments.

Preliminary crystallization screens can be set in the Modular VDX Plate. Once the results are scored, the wells can be removed and organized into separate frames for optimization and observation. Preliminary crystallization screens typically evaluate a diverse array of reagents, reagent concentrations, pH values, and other variables that can influence sample solubility and crystallization. These crystallization variables, combined in a single well can have a unique effect on the sample solubility and are useful for screening preliminary crystallization conditions. The crystallization screen typically delivers a wide range of solubility results, presented in some wells as clear drops, others as precipitate, phase separation, or crystals. Clear drops can indicate the relative supersaturation of the sample in the drop as too low for crystallization, while precipitate can indicate the relative supersaturation as too high. If the sample has a temperature dependent solubility, the temperature of the experiment can be raised or lowered, and the effect of this change can be evaluated based upon the appearance of the drop. For example, a clear drop at room temperature moved to 10°C might result in a change in sample solubility, hence precipitate, phase separation, or even a crystal. A precipitate at room temperature, moved to 30°C may again result in a change in sample solubility, hence precipitate, phase separation, or a crystal. When a standard crystallization plate with fixed wells is used for screening, the decision to move the entire plate to a different temperature can be complicated since all the wells in the plate must be moved together. This may be less than ideal when crystals appears in one of those wells, or there is a mix of clear, precipitated, and phase sep drops where each well requires a different course of action. Each solubility result or class (clear, precipitate, phase sep, etc.) might require a unique temperature evaluation. With the Modular VDX Plate, individual wells can be removed from the frame, organized by crystallization score or solubility into a separate frame or frames, and the effect of temperature evaluated on a more discrete basis. For example, clear drops can be removed from the screen frame, complete with their respective equilibrated reservoir solution, placed into a separate frame and that newly created module can be incubated at a lower temperature. Drops with precipitate can be removed from the screen frame, organized into a new module and be incubated at an elevated temperature to see if an increase in temperature improves sample solubility and promotes crystallization.

Beyond screening, the Modular VDX Plate also offers enhanced plate flexibility and customization during optimization of preliminary crystallization experiments. Optimization experiments can be performed in Modular VDX frames, utilizing removable wells to reorganize the plate as the optimization proceeds. More promising conditions can be grouped into a single frame, reducing the actual utilized active screen space required. Less promising experiments can be archived and stored into other Modular VDX frames.

Identification of the Modular VDX Plates is simple. Frames and well sides can be labeled using a variety of color coded Tough-Tag labels.

The Modular VDX Plates are also handy for the organization of drops and wells containing crystals. Wells containing crystals suitable for diffraction analysis can be combined, stored, and carefully transported in modules, avoiding the extra baggage of wells containing clear drops and precipitate.

Just like the original VDX Plate, the Modular VDX Plate is compatible with Micro-Bridges®, Glass Sitting Drop Rods™, as well as 22 mm circle and square cover slides.

References

1. *A modular plate for the optimization of crystallization experiments. J. Appl. Cryst. (2002) 35, 509-510.*

Order Information

Cat. No.	Name	Description	Price
HR3-198	**Modular VDX Frames**	24 frames with covers	**$57.00**
HR3-196	**Modular VDX Wells**	no sealant, 576 wells	**$70.00**
HR3-204	**Modular VDX Well**	with sealant, 96 wells	**$25.00**

48 WELL CRYSTALLIZATION PLATES

INTELLI-PLATE 48-2 AND 48-3

APPLICATION

- Sitting Drop Crystallization

FEATURES

- SBS Footprint
- Intelli-Plate 48-2 Drop Volume:
  20 µl & 4 µl
  Reservoir: 500 µl
- Intelli-Plate 48-3 Drop Volume: 4 µl
  Reservoir: 500 µl

DESCRIPTION

HR8-152

HR8-156

The Intelli-Plate 48-2 is a 48 well sitting drop plate for optimization. The Intelli-Plate 48-2 features 6 reagent wells along the y-axis (A-F) and 8 reagent wells along the x-axis (1-8). The reagent reservoir is typically filled with 200 µl and is capable of holding between 100 and 500 µl. The optimization drop wells are arranged at the top of each well. The left sample drop well holds between 4 µl or less while the elongated sample well holds 20 µl or less.

The IntelliPlate 48-3 is a 48 well sitting drop plate for screening. The Intelli-Plate 48-3 features 6 reagent wells along the y-axis (A-F) and 8 reagent wells along the x-axis (1-8). The reagent reservoir is typically filled with 200 µl and is capable of holding between 100 and 500 µl. The screening drop wells are arranged at the top of each well. The three sample drop wells each hold between 4 µl or less.

The Hampton Research ClearSeal Film™ and Sealing Film Applicator are used to seal the Intelli-Plate when a sitting drop experiment is performed. The advanced film is a transparent, polyolefin film with clear, pressure-sensitive, silicone adhesive. The film is pre-cut to perfectly fit 96 well crystallization plates and is complete with 1 cm wide, perforated end tabs which makes handling and positioning of the film easy, without fear of fingerprints getting into the view field. The Sealing Film Applicator is a small, hand-held applicator which should be used to create a consistent film seal across the entire crystallization plate. The Intelli-Plate can also be sealed using one of the clear sealing tapes supplied in a roll format.

The height of all Intelli-Plates is 0.560 inches.

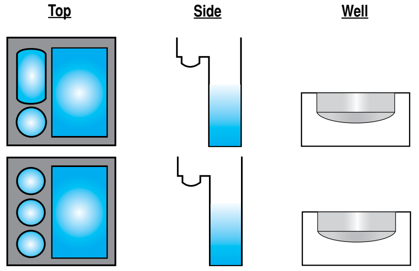

Order Information

Cat. No.	Name	Description	Price
HR8-150	Intelli-Plate 48-2 well	10 plate case	$85.00
HR8-152	Intelli-Plate 48-2 well	40 plate case	$300.00
HR8-154	Intelli-Plate 48-3 well	10 plate case	$85.00
HR8-156	Intelli-Plate 48-3 well	40 plate case	$300.00

Individual Side Well View

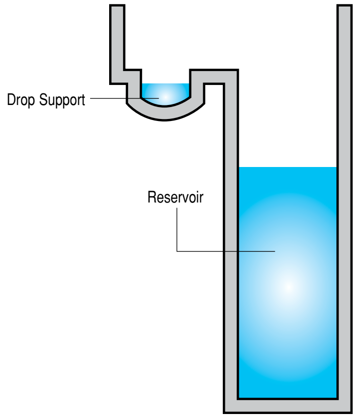

Individual Top Well View Intelli-Plate & Intelli-Plate 48-2

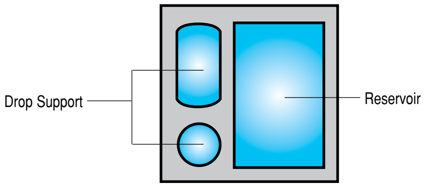

Individual Top Well View Intelli-Plate 48-3

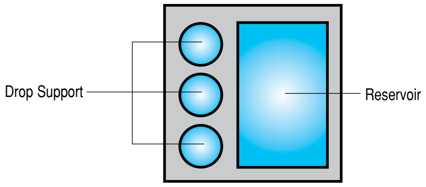

VDX48™ PLATE WITH SEALANT

APPLICATION

- Hanging drop crystallization

FEATURES

- Sealant applied to 48 wells, ready to seal with cover slide
- Compatible with 12 mm diameter circle cover slides
- Optically clear plastic
- 128 mm x 85 mm microplate footprint

Hanging Drop Crystallization

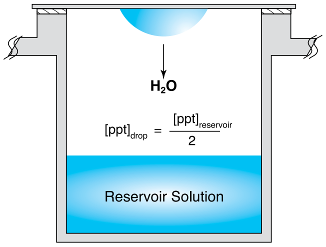

DESCRIPTION

48 well plate with sealant for hanging drop vapor diffusion crystallization with a microplate footprint. Stackable, optically clear plastic, 48 well plates with raised covers (to allow room for cover slides)and flat bottoms for exceptional optics. Raised, wide rings around each reservoir (well) minimize cross contamination and allow each well to be individually sealed with 12 mm diameter circle cover slides. Plates are individually wrapped and supplied with applied sealant. The microplate footprint is space-saving and compatible with automated liquid handling systems. Approximate dimensions: 128 mm x 85 mm. Approximate well ID: 9 mm. Typical well volume: 100 to 300 μl.

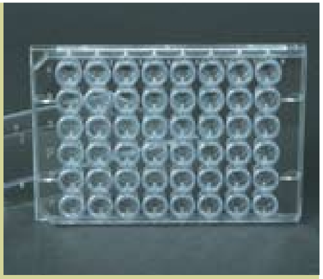

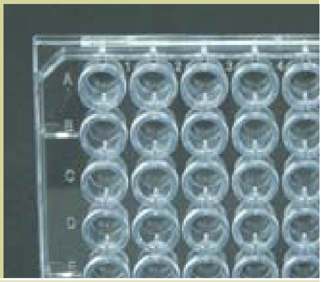

Order Information

Cat. No.	Name	Description	Price
HR3-275	**VDX48 Plate with sealant**	50 plate case	**$335.00**

72 WELL CRYSTALLIZATION PLATES

MICROBATCH 72 WELL PLATE (GREINER)

APPLICATION

- Microbatch crystallization

FEATURES

- Conical, flat bottom wells
- Hydrophobic or hydrophilic versions
- Compact size
- Supplied with cover

MATERIAL

- Polystyrene

DESCRIPTION

The 72 Well Microbatch Plate consists of a 6 by 12 well layout. Paraffin, Silicon, or Al's Oil is poured into the trough area of the plate, submerging all 72 wells under oil. Sample and reagent are then pipetted into each of the wells. The plate is available as a plasma-treated, hydrophilic version for screening or an untreated, hydrophobic version for optimization. The hydrophilic version gives better liquid handling with small volumes, while the hydrophobic version reduces crystal nucleation and helps to prevent the crystals from sticking to the plastic. Plate dimensions: 83.3 mm x 58.0 mm x 10.0 mm. Conical well with 11 μl drop volume. Flat bottom well diameter: 1.3 mm. Supplied with cover.

Order Information

Cat. No.	Name	Description	Price
HR3-120	**72 Well Microbatch Plate, Greiner 654180**	treated, hydrophylic - 10 plate case	**$10.00**
HR3-122	**72 Well Microbatch Plate, Greiner 654180**	treated, hydrophylic - 100 plate case	**$106.00**
HR3-086	**72 Well Microbatch Plate, Greiner 654102**	untreated, hydrophobic - 10 plate case	**$10.00**
HR3-087	**72 Well Microbatch Plate, Greiner 654102**	untreated, hydrophobic - 100 plate case	**$106.00**

CRYSTALCLEAR STRIPS™ (DOUGLAS INSTRUMENTS)

APPLICATION

- Sitting drop vapor diffusion crystallization

FEATURES

- 96 well sitting drop plate
- Concave or flat drop area
- Modular frame and well design
- One or two sample drop wells per reagent well

MATERIAL

- Polystyrene

DESCRIPTION

HR3-128

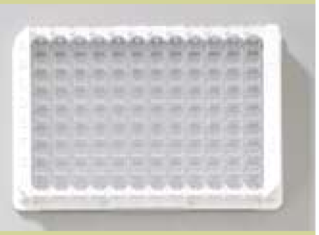
HR3-162

CrystalClear Strips 96 well sitting drop plates are suitable for both automatic and manual crystallization. They are particularly useful for screening, since they feature a space-saving microplate footprint. Sample and reagent drops (50 nl to 4 μl) are dispensed on a shelf on one side of each well. 50 to 100 μl of crystallization reagent is placed in the well. The strip is sealed with clear sealing tape. Losses by evaporation though the body of the strip and the tape are around 0.25% of the reservoir per day.

CrystalClear D (Douglas Instruments CCLEAR-D/1-10)
The CrystalClear D Strips, with indent (HR3-128) have a single depression for the sample and reagent drop on the shelf above each reservoir. This prevents hydrophobic reagents from running off the shelf into the reservoir. This version should be used for reagents and screens that contain low molecular weight alcohols and detergents. The spacing between strips is 9 mm (the regular SBS spacing). However, the spacing between sample chambers along a D strip is only 8.45 mm. Therefore, if you want to use a multichannel pipette to fill either the reservoirs or the sample wells, use a 6 or 12-channel pipette.

CrystalClear P (Douglas Instruments CCLEAR-P/1-10)
The CrystalClear P Strips, without indent (HR3-162) have a small, raised, circular platform on the shelf where samples are placed. This provides improved viewing but is not suitable for solutions containing hydrophobic reagents such as low molecular weight alcohols and detergents. However, drops below 0.2 μl (200 nl) containing alcohol (and other hydrophobic reagents) will generally work with all plates. The spacing between strips is 9 mm (the regular SBS spacing). However, the spacing between sample chambers along a P strip is only 8.45 mm. Therefore, if you want to use a multichannel pipette to fill either the reservoirs or the sample wells, use a 6 or 12-channel pipette.

CrystalClear Duo (Douglas Instruments CCLEAR-DUO/1-10)
The CrystalClear Duo Strips is a 96 well sitting drop vapor diffusion plate. The plate features SBS 9 mm spacing in both directions, which makes the plate suitable for use with all dispensing robots and in all observation systems. The plate offers low volume reagent wells for vapor diffusion with two depressions on the shelf for samples. A plate consists of a single frame and twelve strips, snapped into position, secured in the frame. Each strip contains eight wells. Each of the eight wells contains a single reagent well and two sample drop wells. Sample drop wells are flat. Frame is color coded green.

CrystalClear Strips are packaged as twelve removable 8 well strips per frame (for a total of 96 wells per frame) with ten frames (plates) per case. CrystalClear Strips are sealed using clear sealing tape or film. Approximate plate size: 12.8 cm x 8.6 cm.

Order Information

Cat. No.	Name	Description	Price
HR3-128	CrystalClear D Strips, with indent	10 plate case	$169.00
HR3-162	CrystalClear P Strips, without indent	10 plate case	$169.00
DI-043	CrystalClear Duo Strips	10 plate case	$145.00

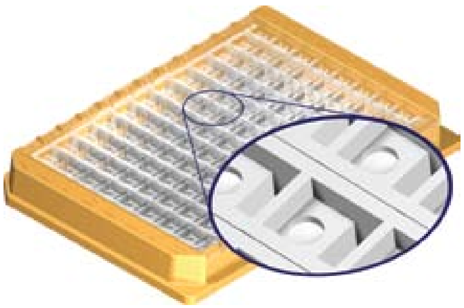
HR3-128 CrystalClear D Strips, with indent

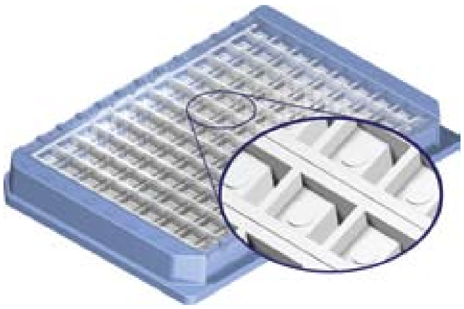
HR3-162 CrystalClear P Strips, without indent

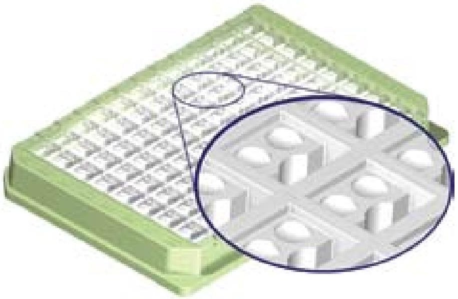
DI-043 CrystalClear Duo Strips

CRYSTALEX™ (CORNING)

APPLICATION

- Sitting drop crystallization

FEATURES

- 96 well sitting drop plate
- Protein well shape is either conical bottom (3773) or flat bottom (3785)
- Compatible with 8 & 12 channel pipetters

MATERIAL

- Advanced optically clear polymer

DESCRIPTION

Corning® CrystalEX 96 Well Crystallization plates are designed for sitting drop vapor diffusion, high throughput protein crystallization. The plates feature 96 protein wells corresponding to 96 reagent wells. The maximum reagent well volume is 210 μl with a typical working volume of 25 to 150 μl. The maximum sample well volume is 10 μl for the conical bottom and 7 μl for the flat bottom. The flat bottom (1.5 mm diameter) sample wells provide better crystal viewing for automated imaging systems. The well surfaces are treated to be hydrophilic for improved drop mixing. Both versions are compatible with 8 and 12 channel manual pipetters, automated liquid handling work stations, and imaging systems. Both versions are manufactured from an advanced polymer with high resistance to commonly used protein crystallization reagents. The advanced polymer material provides for low-background polarization and high optical clarity so protein crystals may be viewed under polarized light with minimal background interference. Low water absorption of the polymer prevents loss of protein drop volume. The plates can be sealed using clear sealing tape or film.

	Corning 3773	Corning 3785
Protein well shape	Conical bottom	Flat bottom
Protein well volume	10 μl	7 μl
Protein well dimensions (top/bottom)	3 mm	3 mm/1.5 mm
Protein well depth	3.1 mm	3.1 mm
Number of reagent/protein wells	96/96	96/96
Reagent well volume	210 μl	210 μl
Recommended reagent working volume	25-150 μl	25-150 μl

Order Information

Cat. No.	Name	Description	Price
HR3-273	**CrystalEX 96 Well, Conical Bottom, Corning 3773**	10 plate case	**$110.00**
HR3-271	**CrystalEX 96 Well, Conical Bottom, Corning 3773**	50 plate case	**$500.00**
HR3-113	**CrystalEX 96 Well, Flat Bottom, Corning 3785**	10 plate case	**$110.00**
HR3-115	**CrystalEX 96 Well, Flat Bottom, Corning 3785**	50 plate case	**$500.00**

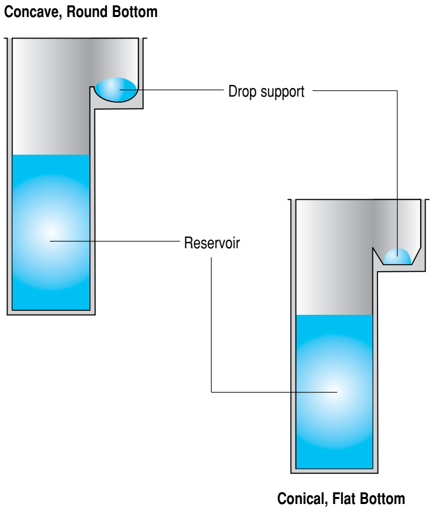

Individual Top Well View

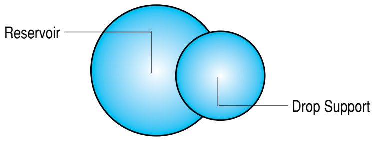

CRYSTALEX™ SECOND GENERATION (CORNING)

APPLICATION

- Sitting drop crystallization

FEATURES

- 96 well sitting drop plate
- Round, flat, and conical flat bottom drop well shapes
- 1, 2 or 4 microliter drop wells
- 1, 3 or 5 drop wells per reservoir
- Hydrophilicity treated or untreated
- PZero or COC plate material

DESCRIPTION

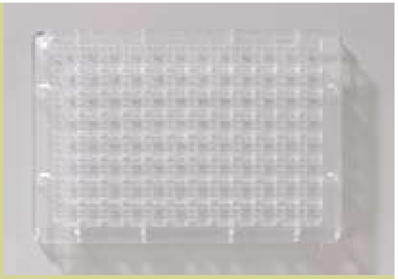

The second generation of Corning® 96 Well, sitting drop format plates are built to SBS specifications, making them well suited for high throughput crystallization and are fully compatible with robotic equipment. The plates are available in several different configurations with varying drop well shapes, plate materials, and number of drop wells per reagent well. The basic plate design is one reagent well flanked by one or three drop wells, with SBS standard spacing between the centers of adjacent well clusters. One may choose a plate with small (1 µl), medium (2 µl), or large (4 µl) drop well volumes. The choice of round, flat, or conical flat well shapes are available. The PZero polymer provides for zero background polarization and is non-birefringent. PZero plates are not treated. The COC polymer offers high chemical resistance. Both types of plastic feature improved transparency. The reservoir numbers are embossed on each individual well for easy identification. Drop well locations conform to SBS standards for robotic handling. The low-volume reagent well saves on reagent costs. The plates can be sealed using clear sealing tape or film.

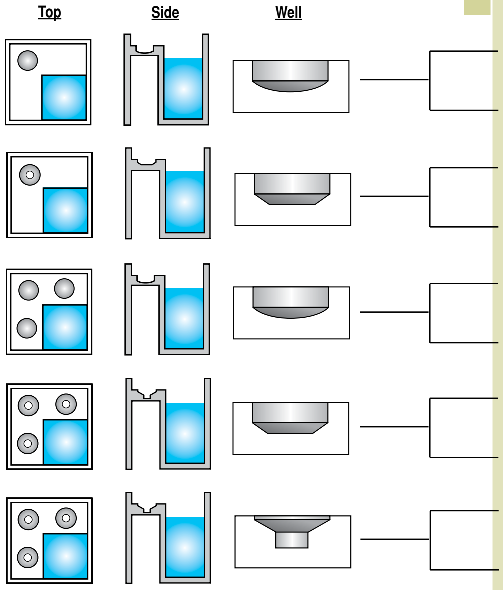

Order Information

Cat. No.	Name	Description	Price
HR8-135	Corning 3556	4 µl round drop well, 1 drop well, COC, untreated - 10 plate case	$105.00
HR8-134	Corning 3556	4 µl round drop well, 1 drop well, COC, untreated - 50 plate case	$490.00
HR8-137	Corning 3551	4 µl conical flat drop well, 1 drop well, COC, treated - 10 plate case	$105.00
HR8-136	Corning 3551	4 µl conical flat drop well, 1 drop well, COC, treated - 50 plate case	$490.00
HR8-139	Corning 3552	2 µl round drop well, 3 drop well, PZero - 10 plate case	$105.00
HR8-138	Corning 3552	2 µl round drop well, 3 drop well, PZero - 50 plate case	$490.00
HR8-141	Corning 3553	2 µl conical flat drop well, 3 drop well, PZero - 10 plate case	$105.00
HR8-140	Corning 3553	2 µl conical flat drop well, 3 drop well, PZero - 50 plate case	$490.00
HR8-147	Corning 3550	1 µl conical flat drop well, 3 drop well, PZero - 10 plate case	$105.00
HR8-146	Corning 3550	1 µl conical flat drop well, 3 drop well, PZero - 50 plate case	$490.00

CRYSTALQUICK™ 96 WELL SITTING DROP PLATE (GREINER)

APPLICATION

- Sitting drop vapor diffusion crystallization
- Hanging drop vapor diffusion crystallization with CrystalDrop™ Lid

FEATURES

- 96 well sitting drop plate
- Round or flat bottom drop wells
- Hanging drop with CrystalDrop™ Lid

MATERIAL

- Plastic

DESCRIPTION

The Greiner CrystalQuick 96 well sitting drop format plates are built to SBS specifications, making them well suited for high throughput crystallization and are fully compatible with robotic equipment. The low volume reagent well saves on reagent costs. The plates are available in round or square drop well shapes, polystyrene, LBR and hydrophobic plate materials, and a varying number of drop wells per reagent well. The CrystalQuick is available in a cyclic polyolefin low birefringence (LBR) material for polarized crystal imaging, exceptional transparency, high chemical resistance, and low water absorption. The low profile CrystalQuick version reduces the amount of space required for storage. In combination with the CrystalDrop™ Lid, the CrystalQuick plates enable simultaneous experiments using both sitting drop and hanging drop methodologies. The plates can be sealed using clear sealing tape or film.

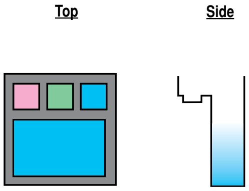

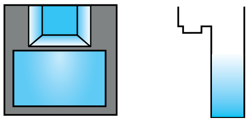

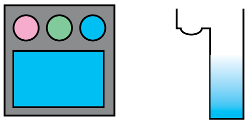

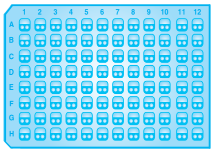

Order Information

Cat. No.	Name	Description	Price
HR3-192	CrystalQuick 96 Well, Greiner 609101	4 µl square drop well, 3 drop well - 10 plate case	$130.00
HR3-190	CrystalQuick 96 Well, Greiner 609101	4 µl square drop well, 3 drop well - 40 plate case	$490.00
HR3-094G	CrystalQuick Plus 96 Well, Greiner 609830	4 µl square drop well, 3 drop well, LBR, hydrophobic - 10 plate case	$139.00
HR3-095G	CrystalQuick Plus 96 Well, Greiner 609830	4 µl square drop well, 3 drop well, LBR, hydrophobic - 40 plate case	$528.00
HR8-148	CrystalQuick Plus 96 Well, Greiner 609130	4 µl square drop well, 3 drop well, hydrophobic - 10 plate case	$130.00
HR8-149	CrystalQuick Plus 96 Well, Greiner 609130	4 µl square drop well, 3 drop well, hydrophobic - 40 plate case	$490.00
HR3-088	CrystalQuick 96 Well, Greiner 609801	4 µl square drop well, 3 drop well, LBR, hydrophobic - 10 plate case	$139.00
HR3-089	CrystalQuick 96 Well, Greiner 609801	4 µl square drop well, 3 drop well, LBR, hydrophobic - 40 plate case	$528.00
HR3-302	CrystalQuick 96 Well, Greiner 609171	4 µl square drop well, 1 drop well, low profile - 20 plate case	$230.00
HR3-304	CrystalQuick 96 Well, Greiner 609171	4 µl square drop well, 1 drop well, low profile - 80 plate case	$830.00
HR3-092G	CrystalQuick Plus 96 Well, Greiner 609180	4 µl square drop well, 1 drop well, hydrophobic, low profile - 20 plate case	$266.00
HR3-093G	CrystalQuick Plus 96 Well, Greiner 609180	4 µl square drop well, 1 drop well, hydrophobic, low profile - 80 plate case	$1,011.00
HR3-284	CrystalQuick 96 Well, Greiner 609871	4 µl square drop well, 1 drop well, LBR, low profile - 20 plate case	$255.00
HR3-285	CrystalQuick 96 Well, Greiner 609871	4 µl square drop well, 1 drop well, LBR, low profile - 80 plate case	$880.00
HR3-283	CrystalQuick 96 Well, Greiner 609120	2 µl round drop well, 3 drop well - 10 plate case	$130.00
HR3-281	CrystalQuick 96 Well, Greiner 609120	2 µl round drop well, 3 drop well - 40 plate case	$490.00
HR3-090	CrystalQuick 96 Well, Greiner 609820	2 µl round drop well, 3 drop well, LBR - 10 plate case	$139.00
HR3-091	CrystalQuick 96 Well, Greiner 609820	2 µl round drop well, 3 drop well, LBR - 40 plate case	$528.00
HR3-096	CrystalDrop Lid, Greiner 609150	2 flat wells, ungreased - 10 lid case	$54.00
HR3-097	CrystalDrop Lid, Greiner 609150	2 flat wells, ungreased - 40 lid case	$201.00
HR3-137	Greased CrystalDrop Lid, Greiner 609050	2 flat wells, greased - 10 lid case	$233.00
HR3-138	Greased CrystalDrop Lid, Greiner 609050	2 flat wells, greased - 40 lid case	$850.00

INTELLI-PLATE™ (ART ROBBINS INSTRUMENTS)

APPLICATION

- Sitting drop crystallization

FEATURES

- 96 reservoir wells with 2 or 3 corresponding sample wells
- 96 well Intelli-Plate is compatible with 8 and 12 channel pipetters for manual set up
- Footprint and well spacing meet microplate industry standards for automation
- Intelli-Plate 96-2 well plate features a maximum fill volume of 140 µl for the reservoir well, 10 µl for the upper sample well, & 4 µl for the lower sample well
- Compatible with robotic equipment for automation
- Intelli-Plate Flat Shelf 96 well plate features a maximum fill volume of 300 µl for the reservoir well and a flat shelf for the sample drop
- Intelli-Plate 96-3 plate features a maximum fill volume of 140 µl for the reservoir well, 1 µl for the sample well
- Custom bar coding available - please inquire

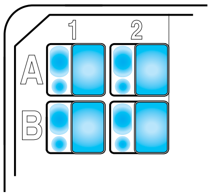

Intelli-Plate™ 96-2 well

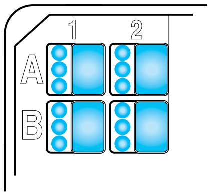

Intelli-Plate™ 96-3 well

DESCRIPTION

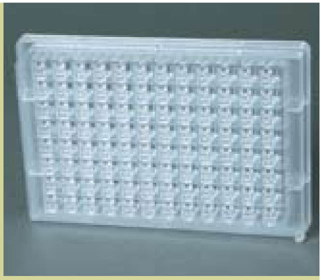

The Intelli-Plate 96-2 Original crystallization plate is an optically clear plate for sitting drop vapor diffusion crystallization. It has been built to SBS specifications with 8 vertical wells versus 12 horizontal wells. This plate features two locations for sample per reservoir. The sample drop locations are located along the left side of the reagent reservoir, along the Y-axis of the plate. The sample wells are concave depressions on the ledge above the adjacent, flat bottom reagent reservoir. One sample well is located above the second sample well. The top sample well can hold 10 *μl* or less of sample. The lower sample well can hold 4 *μl* or less of sample. The reagent reservoir is typically filled with 70 *μl* of reagent and is capable of holding up to 140 *μl* of reagent. The sidewalls separating adjacent wells or reservoirs are 0.9 and 1.0 mm thick in order to offer a larger area for sealing the plate and separation of the reservoirs. The vertical wells are labeled A-H along the left side of the plate and the horizontal wells are labeled 1-12 along the top of the plate. The Intelli-Plate 96-2 is quite rigid, with virtually no torsional flex and is designed for either manual or automated pipetting. All wells have standard 9 mm spacing to conform to SBS standards. The Art Robbins Instruments catalog number for this plate is 102-0001-00.

The Intelli-Plate 96-3 well crystallization plate is an optically clear plate for sitting drop vapor diffusion crystallization. It has been built to SBS specifications with 8 vertical wells versus 12 horizontal wells. This plate features three identical locations for sample per reservoir. The sample drop locations are located along the left side of the reagent reservoir, along the Y-axis of the plate. The sample wells are concave depressions on the ledge above the adjacent, flat bottom reagent reservoir. Each identical well features a round bottom for easy crystal harvesting. Each drop well can hold up to 1 *μl* of sample. The reagent reservoir is typically filled with 70 *μl* of reagent and is capable of holding up to 140 *μl* of reagent. The sidewalls separating adjacent wells or reservoirs are 0.9 and 1.0 mm thick in order to offer a larger area for sealing the plate and separation of the reservoirs. The vertical wells are labeled A-H along the left side of the plate and the horizontal wells are labeled 1-12 along the top of the plate. All wells have standard 9 mm spacing to conform to SBS standards.

The Intelli-Plate 96 Flat Shelf crystallization plate is an optically clear plate for sitting drop vapor diffusion crystallization. It has been built to SBS specifications with 8 vertical wells versus 12 horizontal wells. This plate features a single flat shelf for sample drops. The sample drop shelves are located along the left side of the reagent reservoir, along the Y-axis of the plate. The reagent reservoir is typically filled with 100 *μl* of reagent and is capable of holding up to 300 *μl* of reagent. Maximum drop volume is 10 *μl* for the upper sample well, & 4 *μl* for the lower sample well. The sidewalls separating adjacent wells or reservoirs are 0.9 and 1.0 mm thick in order to offer a larger area for sealing the plate and separation of the reservoirs. The vertical wells are labeled A-H along the left side of the plate and the horizontal wells are labeled 1-12 along the top of the plate. The Intelli-Plate 96 Flat Shelf is quite rigid, with virtually no torsional flex and is designed for either manual or automated pipetting. All wells have standard 9 mm spacing to conform to SBS standards.

The Intelli-Plate 96-2 LVR is a Low Volume Reservoir version of the original Intelli-Plate. The Art Robbins Instruments catalog number for this plate is 102-0001-00. The plates can be sealed using clear sealing tape or film. The height of all Intelli-Plates is 0.560 inches.

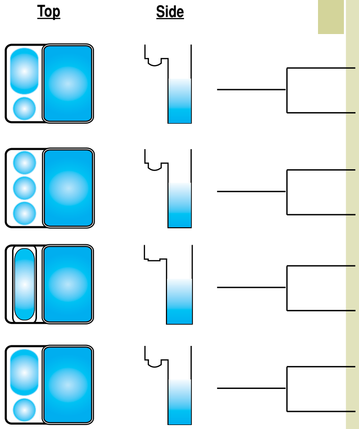

Order Information

Cat. No.	Name	Description	Price
HR3-297	**Intelli-Plate 96-2 Original**	10 plate case	**$90.00**
HR3-299	**Intelli-Plate 96-2 Original**	40 plate case	**$330.00**
HR3-183	**Intelli-Plate 96-3 well**	10 plate case	**$90.00**
HR3-185	**Intelli-Plate 96-3 well**	40 plate case	**$330.00**
HR8-171	**Intelli-Plate 96 Flat Shelf**	10 plate case	**$90.00**
HR8-172	**Intelli-Plate 96 Flat Shelf**	40 plate case	**$330.00**
HR3-143	**Intelli-Plate 96-2 LVR**	10 plate case	**$90.00**
HR3-145	**Intelli-Plate 96-2 LVR**	40 plate case	**$330.00**

MRC 2 WELL CRYSTALLIZATION PLATE (SWISSCI)

APPLICATION

- Sitting drop crystallization

FEATURES

- Lens effect drop wells for enhanced optics
- Drop volume: 100 nl to 5 µl
- Micro-numbering alongside drop volumes
- Reservoir volume: 50 to 100 µl
- SBS format
- Rigid plate structure
- Drop well allows easier crystal harvesting
- Wide partition walls between wells improve sealing
- Developed in conjunction with the MRC Laboratory of Molecular Biology in Cambridge, United Kingdom
- Two drop wells per reservoir

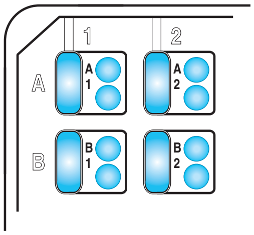

MRC 2 Well Crystallization Plate

DESCRIPTION

The MRC 2 Well Crystallization Plate manufacted by Swissci is a 2 drop chamber, 96 well crystallization plate for sitting drop vapor diffusion. This plate is sometimes referred to as the Innovaplate™ SD-2 crystallography plate.

The plate was developed at the MRC Laboratory of Molecular Biology (Cambridge, United Kingdom) in collaboration with Dr. Jan Löwe. Designed to meet the stringent requirements specified by crystallographers, and made of proprietary new materials that facilitate enhanced imaging, the MRC Crystallization Plate provides easier access and removal of crystals, improved mixing of reagent and sample, and reduced sample and reagent volumes.

Easy Crystal Retrieval
Raised wide wells make the crystal mounting especially easy.

Easy Viewing
The wells are a wide conical shape and have a lens effect for perfect illumination. The micro-numbering ensures you will never get lost again when using the microscope. The optically superior polymer is UV transmissible and may be used to differentiate between salt and protein crystals.

Better Sealing
Wide partition walls between the wells give plenty of area for sealing with tape or film. No central bending occurs in this very robust structure.

Wide Range of Volumes
Typical volumes are 50 to 100 μl of reservoir and 100 nl to 5 μl drop size. The 192 optical wells offer twice the number of experiments of experimental constructions.

SBS Standard
The plate is designed to the 96 well SBS standard for all common holders and external alphanumeric identification. The MRC Crystallization Plate is suitable for centrifugation.

The plates can be sealed using clear sealing tape or film.

Innovaplate is a trademark of Innovadyne Technologies, Inc.

Order Information

Cat. No.	Name	Description	Price
HR3-082	MRC 2 Well Crystallization Plate	10 plate case	$130.00
HR3-083	MRC 2 Well Crystallization Plate	40 plate case	$490.00

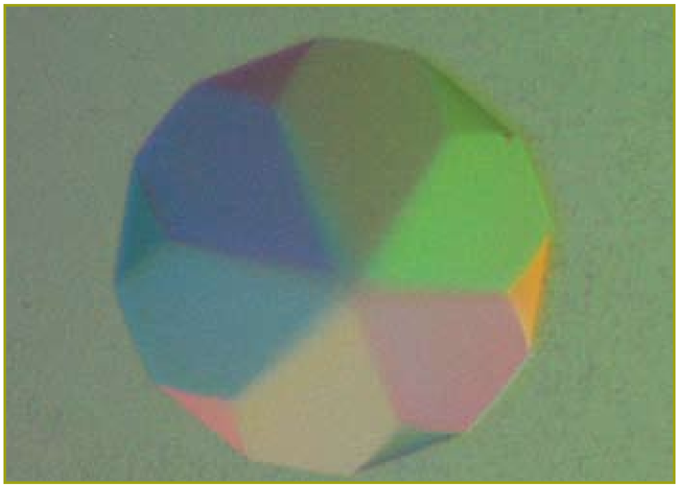

Protein crystal.

Bebiana Moura,
Institute for Molecular and Cell Biology (IBMC) at Oporto, Portugal.

MRC 3 WELL CRYSTALLIZATION PLATE (SWISSCI)

APPLICATION

- Sitting drop crystallization

FEATURES

- Lens effect drop wells for enhanced optics
- Drop volume: 100 nL to 5 µl
- Micro-numbering alongside drop volumes
- Reservoir volume: 50 µl
- SBS format
- Drop well allows easier crystal harvesting
- Three drop wells per reservoir
- Low profile for easier storage

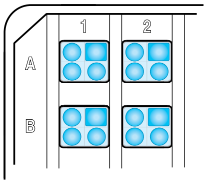

MRC 3 Well Crystallization Plate

3 circular drop wells and 1 square reservoir

DESCRIPTION

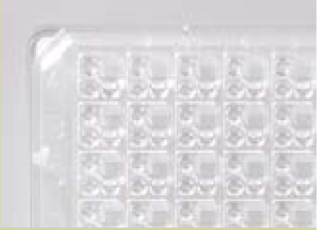

The MRC 3 Well Crystallization Plate is presented in a 96-well plate format that offers unique properties which make it ideal for both nanoliter crystallization screening and microliter optimization alike. Made from optically superior polymer (UVP) and with a new design of the wells, the plate allows easy crystal viewing and retrieval.

Easy Crystal Retrieval
Raised wide wells make the crystal mounting especially easy.

Easy Viewing
The wells are a wide conical shape and have a lens effect for perfect illumination. The micro-numbering ensures you will never get lost again (visible by microscope). The optically superior polymer is UV transmissible and may be used to differentiate between salt and protein crystals. Grown crystals are easy to identify and to remove from well due to a low-binding polymer. Plate with 3 wells for each sample, better growing security with triplicates or the ability to use well two and well three as mixing stations. Wells fill without micro-droplets jumping out due to static effects. The profile allows for easier storage. Low volume buffer well enables savings on reagents.

Better Sealing
Wide partition walls between the wells give plenty of area for good sealing with tape. No central bending occurs in this very robust structure.

Wide Range of Volumes
Typical volumes are 50 μl of reservior and 100 nl to 5 μl drop size. The 288 optical wells offer three times the number of experimental constructions.

SBS Standard
The plates are designed to the 96-well SBS standard for all common holders and external numbering (A-H, 1-12) with corner location make the plate easy to use in a robotic sampler. The 3 well plate is suitable for centrifugation.

The plates can be sealed using clear sealing tape or film.

Order Information

Cat. No.	Name	Description	Price
HR3-123	**MRC 3 Well Crystallization Plate**	10 plate case	**$125.00**
HR3-125	**MRC 3 Well Crystallization Plate**	40 plate case	**$470.00**

MASTERBLOCK® 96 DEEP WELL PLATE (GREINER)

APPLICATION

- 96 Deep Well plate for crystallization screen reagents

FEATURES

- Sterile, individually sealed plates
- Polypropylene
- Seal with AlumaSeal II, Crystal Clear Sealing Film, Cap Mats, or Robolid
- SBS format

DESCRIPTION

The Greiner 780261 MASTERBLOCK is a 96 Deep Well polypropylene block (plate). Each well has a fill volume of 1.2 ml. The MASTERBLOCK is useful for holding crystallization screen reagents when using liquid handling automation or multi channel pipets. Each plate is sterile and packaged individually. Round bottom well for complete reagent recovery. ClearChoice™ virgin polypropylene resin (autoclavable to 121°C). SBS recommended dimensions. Alphanumerically coded wells with chimney-style, round well top.

Seal with AlumaSeal II, Crystal Clear Sealing Film, Cap Mats, or Robolid. Plate Lid is useful to cover the opened MASTERBLOCK when not in immediate use to minimize evaporation. MASTERBLOCK is also great for thermal sealing due to the raised chimney about each well.

Order Information

Cat. No.	Name	Description	Price
HR3-105	**MASTERBLOCK 96 Deep Well polypropylene plate**	50 plate case	**$210.00**

MRC UNDER OIL 96 WELL CRYSTALLIZATION PLATE (SWISSCI)

APPLICATION

- Microbatch crystallization

FEATURES

- Easy crystal retrieval
- Easy viewing
- Drop Volume 100 to 500 nl
- Oil Volume 20 µl
- SBS Standard
- Ultra low binding polymer, no static

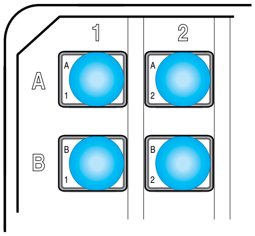

MRC Under Oil Crystallization Plate

DESCRIPTION

The MRC Under Oil 96 Well Crystallization Plate is designed for microbatch crystallization.

The plate was developed at the MRC Laboratory of Molecular Biology (Cambridge, United Kingdom) in collaboration with Dr. Jan Löwe and colleagues.

It is a result of many years of experience in successful robotic high-throughput crystallization and combines many of the features not earlier available to the crystallographer.

The new MRC Under Oil 96 Well Crystallization Plate is designed for microbatch crystallization using paraffin, silicon oil, or a mixture of the two. Following the initial experiment, evaporation of the drop through the oil allows for second crystallization stage, enabling further crystal growth as a consequence of concentration. This is different from other experiments in that the conditions are then extreme in nature and permit new conditions to arise.

The new MRC Under Oil 96 Well Crystallization Plate offers unique properties that make it ideal for both nanoliter crystallization screening and microliter optimization alike. Made from an optically superior polymer and with a new design of the wells, the plate allows easy crystal viewing and retrieval. The advantages of the new MRC Under Oil 96 Well Crystallization Plate:

Easy Crystal Retrieval
Raised wide wells make the crystal mounting especially easy

Easy Viewing
The wells are wide conical and have a polished surface on both sides important for perfect illumination. The micro numbering ensures that you will never get lost again (visible by microscope). The optically superior polymer is even UV transmissible and may be used to differentiate between salt and crystals.

Better Sealing
Wide partition walls between the wells give plenty of area for good sealing with tape for the initial experiments of microbatch. No central bending occurs in this very robust structure.

Recommended Volumes
Typical volumes validated for these plates are 20 μl of oil with a shot through sample delivery of 100 to 500 nl. The 20 μl volume of the individual wells gives the user a wide range of macromolecular crystallization possibilities.

SBS Standard
The plates are designed to the 96 well SBS standard for all common holders ad external numbering (A - H, 1 - 12) with corner location that make the plate easy to use in a robotic sampler. The plate can also be centrifuged for better results. The unique MRC Under Oil 96 well Crystallization plate offers a new way of microbatch crystallography. The 96 wells are optically perfect, designed to observe crystals under a microscope.

Unique Polymer
The proprietary polymer is optically perfect - ultra low binding and guaranteed to have central drop location in the well. There are no static effects and thus micro-droplet jumping is avoided.

Order Information

Cat. No.	Name	Description	Price
HR3-102	MRC Under Oil Crystallization Plate	10 plate case	$120.00
HR3-104	MRC Under Oil Crystallization Plate	40 plate case	$432.00

IMP@CT™ PLATE WITH RESERVOIR (GREINER)

APPLICATION

- Microbatch crystallization

FEATURES

- Conical, flat bottom wells
- Clear and black μClear®
- Double rim reservoir

MATERIAL

- Polystyrene

DESCRIPTION

The Greiner 96 Well Imp@ct Plate for microbatch crystallization has conical, flat bottom wells that promote central location of the sample drop and ease crystal harvesting. The plate has a flat, optically clear bottom. A raised perimeter wall surrounds the plate to contain Paraffin Oil, Al's Oil, or Silicon Oil. The entire inner area of the plate or each individual sample well may be filled with oil. The plate features a double rim reservoir to hold water or reagent to control and manipulate vapor diffusion from the microbatch drops. Plate dimensions: 127.76 mm x 85.48 mm x 14.4 mm. Conical flat bottom well with 8 μl drop volume. Flat bottom well diameter: 1.33 mm.

Greiner 673170 plate is clear. Greiner 673096 plate is black with clear drop area.

Plates supplied without cover. Plate Lids HR3-084 or HR3-085 (Greiner 656190) are available separately.

Order Information

Cat. No.	Name	Description	Price
HR3-098	**96 Well Imp@ct Plate, Greiner 673170**	with reservoir, no cover - 10 plate case	**$118.00**
HR3-099	**96 Well Imp@ct Plate, Greiner 673170**	with reservoir, no cover - 40 plate case	**$444.00**
HR3-100	**96 Well Imp@ct Plate, Greiner 673096**	with reservoir, no cover, black μClear - 10 plate case	**$135.00**
HR3-101	**96 Well Imp@ct Plate, Greiner 673096**	with reservoir, no cover, black μClear - 40 plate case	**$482.00**
HR3-084	**Plate Lid, Greiner 656190**	10 lid case	**$10.00**
HR3-085	**Plate Lid, Greiner 656190**	40 lid case	**$35.00**

IMP@CT™ PLATE WITHOUT RESERVOIR (GREINER)

APPLICATION

- Microbatch crystallization

FEATURES

- Conical, flat bottom wells
- Supplied with cover
- Microplate footprint

MATERIAL

- Polystyrene

DESCRIPTION

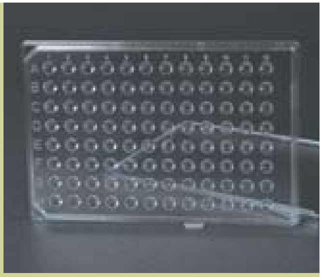

The Greiner 673101 96 Well Imp@ct Plate has a low profile format with individual conical shaped sidewalls and a smooth, flat bottom for enhanced microscopic investigation of crystals. Conical well shape avoids spreading of the sample and reagent droplet away from the center of the well. Typical well capacity for Paraffin, Silicon, or Al's Oil is 10 to 20 μl per well. Vertical wells are labeled A-H. Horizontal wells are labeled 1-12. Stackable with supplied cover. Plate Dimensions: 127.76 mm x 85.48 mm x 8.0 mm. Conical well with 19 μl drop volume. Flat bottom well diameter: 2 mm.

Order Information

Cat. No.	Name	Description	Price
HR3-295	**96 Well Imp@ct Plate, Greiner 673101**	with cover, without reservoir - 10 plate case	**$98.00**
HR3-293	**96 Well Imp@ct Plate, Greiner 673101**	with cover, without reservoir - 40 plate case	**$350.00**

MICROBATCH 96 WELL PLATE

APPLICATION

- Microbatch crystallization

FEATURES

- 96 concave bottom wells
- Microplate footprint
- Supplied with cover

MATERIAL

- Polystyrene

DESCRIPTION

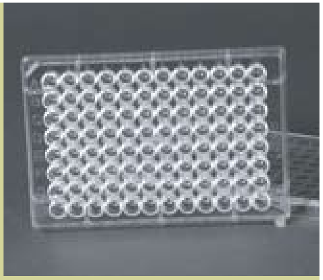

The 96 Well Microbatch Plate is an 8 by 12 well microplate with cover. Paraffin, Silicon, or Al's Oil is dispensed into each individual well. Sample and reagent are then pipetted into the wells below the oil. Individual well bottoms are "U" shaped to center the drop under oil. Oil can be rapidly (between 3 to 20 seconds) and neatly dispensed into the plate using V&P Scientific VP195B Multi-Spense or V&P Scientific VP195B1-96. After sample and reagent are dispensed under the oil using an 8 channel and single channel pipet, the plates are then centrifuged at very low speed for 5 to 10 minutes to mix and center the drops. The covered plates can be stored between 0-37°C and viewed free of condensation as the covers can be removed for viewing without drying the drops when using the microbatch method.

Order Information

Cat. No.	Name	Description	Price
HR3-265	**96 Well Microbatch Plate, with cover**	10 plate case	**$50.00**
HR3-267	**96 Well Microbatch Plate, with cover**	50 plate case	**$235.00**

VAPOR BATCH 96 WELL PLATE (DOUGLAS INSTRUMENTS)

APPLICATIONS

- Sitting drop and microbatch crystallization
- Anaerobic crystallization

FEATURES

- 96 conical, flat bottom wells
- Supplied with cover
- Microbatch and vapor diffusion from the same plate
- Compact size
- Hydrophobic or hydrophilic versions

MATERIAL

- Polystyrene

DESCRIPTION

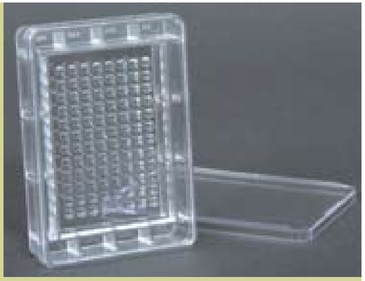

The Douglas Instruments Vapor Batch Plate is designed for both microbatch and sitting drop vapor diffusion crystallization. It has 96 wells in the central region of the plate and several reservoirs around the outside of the plate. Oil is placed in the central region followed by sample and reagent for a microbatch experiment. When the outside reservoirs are filled with reagent, they can be used to preserve microbatch crystals by preventing the drops from drying out. The reservoirs can also be used for common dehydrant, sitting drop vapor diffusion experiments where up to 96 wells are equilibrated against a single reagent or dehydrant. Preliminary experiments suggest this method may find more hits in screening experiments than the conventional method of using the same solution in the reservoir and the drop. Wells are individually labeled. The Vapor Batch Plate is available as a plasma-treated hydrophilic version for screening or an untreated, hydrophobic version for optimization. The hydrophilic version gives better liquid handling with small volumes, while the hydrophobic version reduces crystal nucleation and helps to prevent the crystals from sticking to the plastic. Vapor Batch Plate Holders which provide a Linbro®, VDX™, or SBS footprint are available on a special order basis. Approximate plate dimensions: 81 mm x 55 mm x 20 mm. Drop well depth: 1.8 mm. Typical drop volume: 1 to 9 μl. Maximum drop volume: 20 μl. Recommended oil fill: 2.5 ml. Recommended reagent/dehydrant volume: 8 ml.

Order Information

Cat. No.	Name	Description	Price
DI-038	**Vapor Batch Plate, treated, hydrophilic**	10 plate case	**$66.00**
DI-040	**Vapor Batch Plate, treated, hydrophilic**	80 plate case	**$475.00**
DI-039	**Vapor Batch Plate, untreated, hydrophobic**	10 plate case	**$66.00**
DI-041	**Vapor Batch Plate, untreated, hydrophobic**	80 plate case	**$475.00**

96 WELL CRYSTALLIZATION PLATES

MULTICHANNEL PIPETTER BASIN

APPLICATIONS

- Pipet oil reservoir for microbatch crystallization
- Reagent reservoir for crystallization reagents when setting Silver Bullets screens

FEATURES

- Basin with cover
- Compatible with 8 & 12-channel pipets

MATERIAL

- Polypropylene

DESCRIPTION

The Multichannel Pipetter Basin is a polypropylene trough that allows for rapid reproduction of similar reagents or oils throughout an entire 96 Well crystallization plate. Basin comes with cover.

Order Information

Cat. No.	Name	Description	Price
HR3-269	**Multichannel Pipetter Basin**	each	**$7.00**

384 WELL CRYSTALLIZATION PLATES

CRYSTALEX™ 384 WELL FLAT BOTTOM PLATE (CORNING)

APPLICATION

- Sitting drop crystallization

FEATURES

- 192 sample wells & 192 reagent wells
- Flat bottom sample well
- Low background polarization
- SBS footprint

MATERIAL

- Advanced polymer

METHODOLOGIES

- Use the CrystalEX 384 plate to screen four different reservoir solutions against the same crystallization solution to expand screening space (Newman 2005).

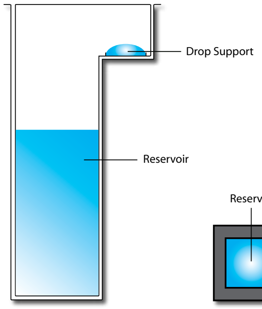

Individual side well view

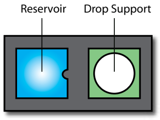

Individual top well view

DESCRIPTION

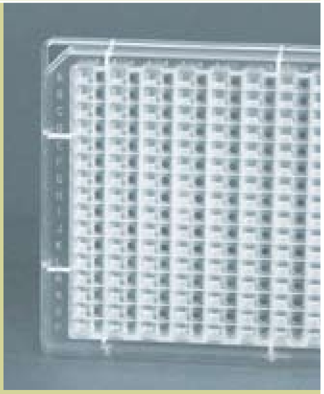

The Corning® CrystalEX is designed for full automation in crystal screening and built to meet industry standards for 384 well microplate footprint and well locations. Features 192 reservoir wells with a 105 μl volume and 192 corresponding protein wells with a maximum volume of 4 μl. Typical reagent well volumes are 50 μl and sample well volumes are 1 μl. The reagent wells and sample wells are positioned to be compatible with multihead dispensing equipment (up to 96 well heads). The plate is manufactured from an advanced polymer with high resistance to commonly used solvents, including acetone, acetic acid, butanone, ethanol, iso-propanol, methanol, DMSO, nitric acid (65%), sulfuric acid (40%), hydrochloric acid (36%), and ammonia solution (33%). The advanced polymer features low background polarization and high optical clarity which allows crystals to be viewed under polarized light with minimal background interference. The low water absorption of the polymer prevents loss of protein drop volume. The plates can be sealed using Crystal Clear Sealing Tape or ClearSeal Film™. The plate is not treated.

References

1. *Expanding screening space through the use of alternative reservoirs in vapor-diffusion experiments. Janet Newman. Acta Cryst. (2005). D61, 490-493.*

Order Information

Cat. No.	Name	Description	Price
HR8-056	**CrystalEX 384 Well Flat Bottom Plate**	10 plate case	**$210.00**
HR8-058	**CrystalEX 384 Well Flat Bottom Plate**	50 plate case	**$875.00**

SILICONIZED GLASS COVER SLIDES

APPLICATIONS

- Hanging, sitting or sandwich drop crystallization

FEATURES

- Siliconized, hydrophobic glass surface
- 12 mm, 18 mm, & 22 mm glass diameter
- 0.22 mm & 0.96 mm glass thickness
- Circle & square

DESCRIPTION

Siliconized glass cover slides allow a droplet to be suspended in a position which provides near optimal conditions for vapor diffusion with the surrounding reservoir solution. These siliconized glass cover slides provide a consistent, high quality finish for crystallization experiments. The hydrophobic surface produces a drop which "stands well" and does not flatten on the glass. The siliconized surface prevents the adhesion of crystals and precipitate onto the glass surface. Use vacuum grease, DC Release Compound, or immersion oil to seal cover slide to plate. Available in 12, 18 or 22 mm diameter circles and 22 mm diameter squares. Available in 0.22 mm or 0.96 mm glass thickness. The 0.96 mm thick slides are virtually unbreakable during crystallization handling.

Order Information

Size	Thickness	Fits	Cat. No.	Name	Description	Price
12	│	VDX48	HR3-278T	Siliconized circle cover slides for Tecan robot	0.5 oz pack (~240 slides)	$40.00
			HR3-280T	Siliconized circle cover slides for Tecan robot	5 oz case (~2,400 slides)	$380.00
			HR3-277	12 x 0.22 mm Siliconized circle cover slides	0.5 oz pack (~240 slides)	$40.00
			HR3-279	12 x 0.22 mm Siliconized circle cover slides	5 oz case (~2,400 slides)	$380.00
12	┃	VDX48	HR8-088	12 x 0.96 mm Thick Siliconized circle cover slides	1 oz pack (~106 slides)	$25.00
			HR8-090	12 x 0.96 mm Thick Siliconized circle cover slides	10 oz case (~1,060 slides)	$238.00
18	│	VDXm	HR3-239	18 x 0.22 mm Siliconized circle cover slides	0.5 oz pack (~100 slides)	$22.00
			HR3-241	18 x 0.22 mm Siliconized circle cover slides	5 oz case (~1,000 slides)	$210.00
18	┃	VDXm	HR3-515	18 x 0.96 mm Thick Siliconized circle cover slides	1 oz pack (~45 slides)	$20.00
			HR3-517	18 x 0.96 mm Thick Siliconized circle cover slides	10 oz case (~450 slides)	$190.00
22	│	VDX	HR3-231	22 x 0.22 mm Siliconized circle cover slides	1 oz pack (~120 slides)	$25.00
			HR3-233	22 x 0.22 mm Siliconized circle cover slides	10 oz case (~1,200 slides)	$238.00
22	┃	VDX	HR3-247	22 x 0.96 mm Thick Siliconized circle cover slides	3 oz pack (~75 slides)	$27.00
			HR3-249	22 x 0.96 mm Thick Siliconized circle cover slides	30 oz case (~750 slides)	$256.00
22	│	VDX	HR3-215	22 x 0.22 mm Siliconized square cover slides	1 oz pack (~100 slides)	$22.00
			HR3-217	22 x 0.22 mm Siliconized square cover slides	10 oz case (~1,000 slides)	$210.00
22	┃	VDX	HR3-223	22 x 0.96 mm Thick Siliconized square cover slides	4 oz pack (~75 slides)	$25.00
			HR3-225	22 x 0.96 mm Thick Siliconized square cover slides	40 oz case (~750 slides)	$238.00

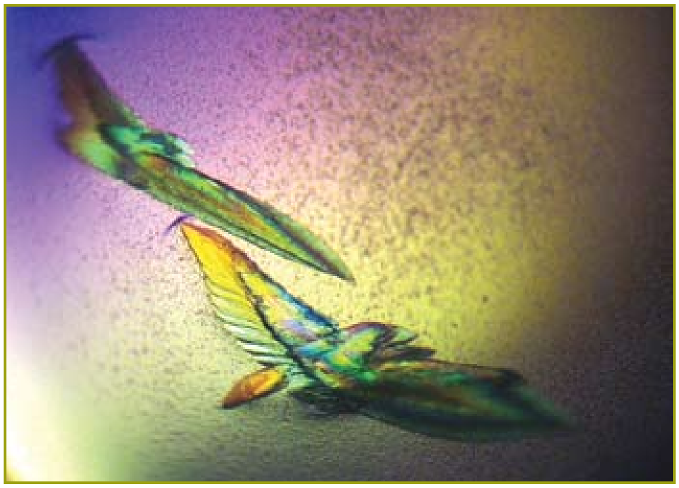

Flying eagles crystals. A mutant of AppA photoreceptor protein.

Vladimira Dragnea, Department of Biology, Indiana University, Bloomington, Indiana, USA.

GLASS COVER SLIDE GIZMO DISPENSER

APPLICATION

- Dispense circle or square cover slides for hanging drop crystallization

FEATURES

- Dispense siliconized or plain cover slides
- Platform displays one to six cover slides at a time for 18 and 22 mm slides
- Platform displays one to twelve cover slides at a time for 12 mm slides
- Available for 12, 18 or 22 mm diameter cover slides for either original (0.22 mm) or thick (0.96 mm) glass thickness
- Black platform for slides allows one to see if drops are clear or precipitated

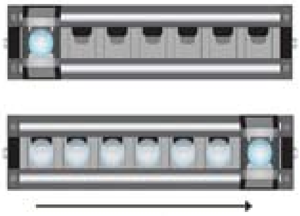

Circle Cover Slide Gizmo

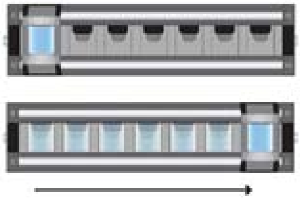

Square Cover Slide Gizmo

DESCRIPTION

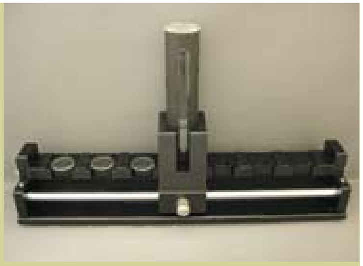

The Cover Slide Gizmo dispenses siliconized or plain slides for convenient drop building during hanging drop vapor diffusion crystallization. Load the cover slides into the dispenser. A clear plastic cover or sealed dispenser keeps slides clean. Slide the loaded dispenser along the platform and six cover slides are dispensed onto the platform. A black background on the platform makes for easy drop visualization as one builds the drop(s). Use finger or forceps to place the completed cover slides onto the crystallization plate.

Each Cover Slide Gizmo includes a single cover so you only need to order HR8-166 if you lose the cover for the HR8-162 or HR8-164. Manual operation, no power required.

Order Information

Cat. No.	Name	Description	Price
HR8-167	**Cover Slide Gizmo**	Use for 12 x 0.22 mm circle cover slides	**$125.00**
HR8-168	**Cover Slide Gizmo**	Use for 12 x 0.96 mm circle cover slides	**$125.00**
HR8-169	**Cover Slide Gizmo**	Use for 18 x 0.22 mm circle cover slides	**$125.00**
HR8-170	**Cover Slide Gizmo**	Use for 18 x 0.96 mm circle cover slides	**$125.00**
HR8-162	**Cover Slide Gizmo**	Use for 22 x 0.22 mm circle cover slides	**$125.00**
HR8-163	**Cover Slide Gizmo**	Use for 22 x 0.96 mm circle cover slides	**$125.00**
HR8-164	**Cover Slide Gizmo**	Use for 22 x 0.22 mm square cover slides	**$125.00**
HR8-165	**Cover Slide Gizmo**	Use for 22 x 0.96 mm square cover slides	**$125.00**
HR8-166	**Replacement Cover**	Fits HR8-162 or HR8-164	**$15.00**

Protein crystal grown using the Hampton Research Crystal Screen.

Enrique Lemus Fuentes, Jefe de carrera de ing. en alimentos, Universidad Tecnológica de la Mixteca.

PLAIN GLASS COVER SLIDES

APPLICATIONS

- Hanging, sitting or sandwich drop crystallization

FEATURES

- Hydrophilic glass surface
- 12 mm, 18 mm, & 22 mm glass diameter
- 0.22 mm & 0.96 mm glass thickness
- Circle & square

DESCRIPTION

Plain, non-siliconized, hydrophilic glass cover slides. Useful for hanging drop, sitting drop, and sandwich drop vapor diffusion crystallization methods. The hydrophilic surface creates a drop which is more flat than a drop on a siliconized surface. This offers enhanced imaging for very small drops. Use vacuum grease, DC Release Compound, or immersion oil to seal cover slide to plate.

Order Information

Cat. No.	Name	Description	Price
HR3-207T	**12 mm Plain circle cover slides for Tecan robot**	0.5 oz pack (~240 slides)	**$38.00**
HR3-209T	**12 mm Plain circle cover slides for Tecan robot**	5 oz case (~2,400 slides)	**$361.00**
HR3-235	**18 x 0.22 mm Plain circle cover slides**	0.5 oz pack (~100 slides)	**$20.00**
HR3-237	**18 x 0.22 mm Plain circle cover slides**	5 oz case (~1,000 slides)	**$190.00**
HR3-227	**22 x 0.22 mm Plain circle cover slides**	1 oz pack (~120 slides)	**$23.00**
HR3-229	**22 x 0.22 mm Plain circle cover slides**	10 oz case (~1,200 slides)	**$219.00**
HR3-243	**22 x 0.96 mm Thick Plain circle cover slides**	3 oz pack (~75 slides)	**$25.00**
HR3-245	**22 x 0.96 mm Thick Plain circle cover slides**	30 oz case (~750 slides)	**$238.00**
HR3-211	**22 x 0.22 mm Plain square cover slides**	1 oz pack (~100 slides)	**$20.00**
HR3-213	**22 x 0.22 mm Plain square cover slides**	10 oz case (~1,000 slides)	**$190.00**
HR3-219	**22 x 0.96 mm Thick Plain square cover slides**	4 oz pack (~75 slides)	**$23.00**
HR3-221	**22 x 0.96 mm Thick Plain square cover slides**	40 oz case (~750 slides)	**$219.00**

OPTICLEAR™ PLASTIC COVER SLIDES

APPLICATION

- Hanging, sitting and sandwich drop crystallization

FEATURES

- 0.22 mm thickness
- RNase-free; handle with forceps to prevent contamination
- Compatible with most crystallization reagents
- Siliconization is not required
- 18 mm circle, 22 mm circle, or 22 mm square

DESCRIPTION

RNase-free, hydrophobic covers are lighter weight than glass. Plastic cover slides are ready to use without washing or siliconization. Hydrophobic working surfaces are protected with clean release liners to prevent RNase contamination. The plastic cover slides remain flat and will not curl, even at high temperatures. Plastic cover slides do not chip, crack, or break and are dust-free. Square cover slides are supplied protected in a sandwich polyethylene film. The circle cover slides are packed between fiber-free paper slips. Use vacuum grease, DC Release Compound, or immersion oil to seal cover slide to plate.

Order Information

Cat. No.	Name	Description	Price
HR8-082	**22 mm OptiClear Plastic circle cover slides**	100 pack	**$42.00**
HR8-084	**22 mm OptiClear Plastic circle cover slides**	1,000 pack	**$373.00**
HR8-074	**22 mm OptiClear Plastic square cover slides**	100 pack	**$28.00**
HR8-076	**22 mm OptiClear Plastic square cover slides**	1,000 pack	**$248.00**
HR8-078	**18 mm OptiClear Plastic circle cover slides**	100 pack	**$42.00**
HR8-080	**18 mm OptiClear Plastic circle cover slides**	1,000 pack	**$373.00**

PEN-VAC®

APPLICATION

- Holding and manipulating small, flat surfaces such as glass cover slides

FEATURES

- Self-contained vacuum
- No power needed
- Compact

DESCRIPTION

PEN-VAC is a great tool for handling small, flat surfaced objects such as plain and siliconized cover slides, as well as plastic slides. Simply depress the side button, touch the tip to the object to be lifted and release the button. PEN-VAC creates a totally self-contained vacuum that lifts up to 50 grams for up to one minute. Less than an ounce, the lightweight aluminum body fits in a pocket. No power supply needed. The kit includes one PEN-VAC with 5 3/4" blue body, (3) angled and (3) straight probes with (2) each 1/8", 1/4", and 3/8" diameter blue silicone vacuum cups. It is packaged in a clear plastic storage case.

Order Information

Cat. No.	Name	Description	Price
HR3-251	**PEN-VAC Kit**	each	**$95.00**

COVER SLIDE VACUUM GADGET™

APPLICATION

- Holding and manipulating small, flat surfaces such as glass cover slides

FEATURES

- Self-contained vacuum
- No power needed
- Compact

DESCRIPTION

The Cover Slide Vacuum Gadget™ allows one to pick up a single cover slide, invert the slide and gadget, rest the gadget on the bench so the slide rests as if on a pedestal, pipet the drop onto the slide, then use the gadget to place and release the slide onto the crystallization plate. The Gadget is a hand held vacuum bulb with a single Buna-N vacuum cup. Squeeze the bulb, press the vacuum cup to the slide, ease the squeeze on the bulb to create a vacuum and the slide is held in place at the end of the Gadget. When ready to position and place the slide give the Gadget a gentle squeeze to release the vacuum. Single bulb with 3/8" (9.53 mm) diameter cup.

Order Information

Cat. No.	Name	Description	Price
HR8-098	**Cover Slide Vacuum Gadget**	each	**$27.00**

PLATE STAND

APPLICATION

- Convenient plate holder

DESCRIPTION

A convenient holder for a crystallization plate, tilted at a 30° angle to comfortably view samples and pipet without strain. Equipped with a neoprene pad and rubber feet to prevent slippage. Works well with smaller, 96 well footprint plates.

Order Information

Cat. No.	Name	Description	Price
HR4-468	**Plate Stand**	each	**$41.00**

DUSTER - CANNED AIR

APPLICATION

- Dust removal

DESCRIPTION

Remove dust from cover slides, plates, and other laboratory supplies.

Order Information

Cat. No.	Name	Description	Price
HR4-411	**Duster - Canned Air**	10 oz can	**$15.00**

GLASS SITTING DROP RODS™

APPLICATION

- Sitting drop crystallization

FEATURES

- Perform sitting drop experiments in 24 well hanging drop plates
- Made of glass, may be siliconized
- Thermal mass stabilizes drop temperature

MATERIAL

- Optically clear glass

DESCRIPTION

Solid Glass Sitting Drop Rods with a concave depression in the top can hold up to 50 μl and fit into the well of a 24 well VDX™ Plate or Linbro® Plate. The rods can stand freely in the well, or a small amount of vacuum grease can be applied to the bottom of the rod to secure it to the plate. They also offer enhanced temperature stability due to the mass of the rod. The rods are 14 mm tall, 10 mm in diameter, and the depth of the concave depression is 2 mm. The rods are supplied without siliconization.

Order Information

Cat. No.	Name	Description	Price
HR3-146	**Glass Sitting Drop Rods**	24 pack	**$100.00**
HR3-148	**Glass Sitting Drop Rods**	144 pack	**$540.00**

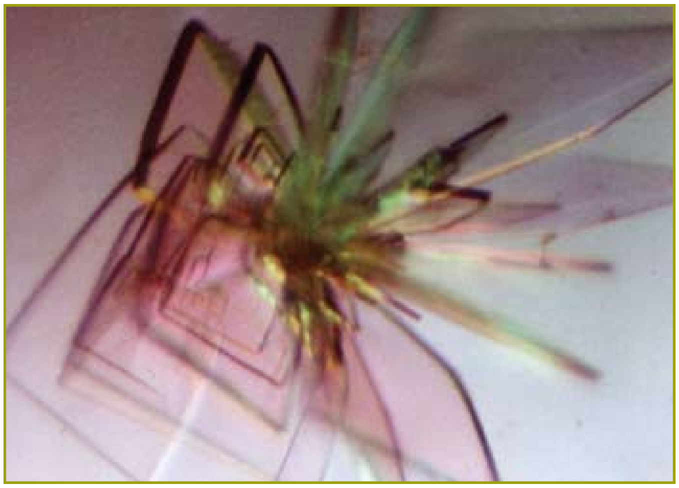

ATPase-domain complex of HlyB.

Jelena Zaitseva and Lutz Schmitt, Institute of Biochemistry, Heinrich Heine University, Duesseldorf, Germany.

MICRO-BRIDGES®

MICRO-BRIDGES®

APPLICATIONS

- Sitting drop crystallization
- Heavy atom soaks
- Seeding

FEATURES

- Perform sitting drop experiments in 24 well hanging drop plates
- Removable

MATERIAL

- Optically clear polystyrene

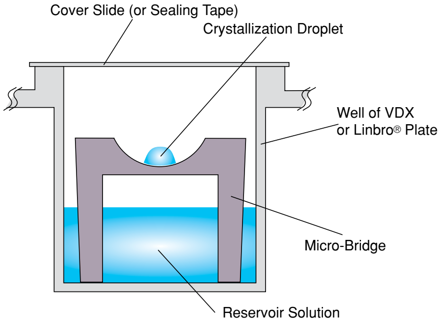

DESCRIPTION

Micro-Bridges are small devices in the shape of a bridge that are designed to carry out sitting drop vapor diffusion crystallization when placed in a VDX™ Plate or Linbro® Plate. A single Micro-Bridge fits neatly into the reservoir of a standard 24 well VDX Plate for a sitting drop crystallization experiment. Once placed inside the wells, Micro-Bridges are stable and there is no need to stick them to the wells with grease or adhesive. It is therefore possible to transfer them to other wells during or after a crystallization experiment. Why sitting drop? Placing the droplet in the indentation greatly reduces the risk of losing the protein by accident. Crystallization can be carried out in the presence of detergents and organic solvents which are compatible with polystyrene (such as MPD, iso-propanol, and ethanol). The protein drop is less affected by condensation problems. Soaking and seeding experiments can be carried out easily. Crystals can be transported more securely. Larger drop volumes can be used. Micro-Bridges have a concave indentation in the top surface of the bridge which holds the sample droplet during a crystallization experiment and prevents the droplet from spreading over a large area. Made from polystyrene, these parts are highly transparent and suitable for most crystallizations. The surface of the indentation is highly polished to facilitate the visual inspection of the drops under a microscope. The maximum drop volume for the Micro-Bridge is 35 μl. Reservoirs can be sealed with plain cover slides and vacuum grease or clear sealing tape.

Order Information

Cat. No.	Name	Description	Price
HR3-310	**Micro-Bridges**	100 pack	**$65.00**
HR3-312	**Micro-Bridges**	400 pack	**$361.00**

MICRO-BRIDGES® POLYPROPYLENE

APPLICATIONS

- Sitting drop crystallization
- Small molecule crystallization
- Heavy atom soaks

FEATURES

- Perform sitting drop experiments in 24 well hanging drop plates
- Removable

MATERIAL

- Clarified polypropylene

DESCRIPTION

Same great Micro-Bridges, but manufactured from clarified polypropylene. These Micro-Bridges are resistant to most organic solvents, and are especially useful with crystallization experiments that involve detergents and other hydrophobic reagents. They resist drop-spreading typically observed when using reagents such as detergents, volatile organics such as iso-propanol and ethanol, and non-volatile organics such as MPD. These Micro-Bridges also resist acetone, dioxane, acetonitrile, 2,2,2 trifluoroethanol, and other aggressive organic solvents.

Order Information

Cat. No.	Name	Description	Price
HR3-340	**Micro-Bridges Polypropylene**	100 pack	**$75.00**
HR3-342	**Micro-Bridges Polypropylene**	400 pack	**$255.00**

9 WELL GLASS PLATE

9 WELL GLASS PLATE & SANDWICH BOX SETUP

APPLICATIONS

- Sitting drop crystallization
- Seeding
- Heavy atom soaks
- Cryo transfers

FEATURES

- Siliconized, 9 well glass plates
- Concave wells
- Polystyrene box and support

MATERIALS

- Siliconized glass and polystyrene

DESCRIPTION

When setting up a Sandwich Box, one pipets 25 ml of crystallization reagent or dehydrant into the bottom of the Sandwich Box, then places a support into the box. The 9 well depression plate is placed upon the support. Drops of sample and reagent are pipetted into the depressions and the lid of the Sandwich Box is sealed closed with vacuum grease. Why Sandwich Boxes? They allow one to use very, very large drops. They are optically superior to plastic plates and glass slides, and offer different equilibration kinetics than other crystallization plates. Each of the Sandwich Box components can be washed and reused so there is little waste with the Sandwich Box Setup. These plates can be used during screening and optimization, but are best suited for final optimization and production of crystals. Since the glass plates can be removed from the boxes, crystal mounting is convenient. The siliconized 9 well depression plates are also useful for heavy atom soaks, cryo solution dilution and transfers, and seeding experiments where serial dilutions are involved. Sandwich Box Setups are available as a complete set or as individual components so one can customize the system to meet their needs. Square cover slides can be used to seal individual reservoirs but are not typically used in a crystallization setup. The Sandwich Box Setup consists of a square plastic box (4 5/16” x 4 5/16” x 1 1/8”), a polystyrene plate support, and a siliconized, glass plate (4” L x 3 3/8” W (100 mm x 85 mm)) with nine concave depressions (7/8” O.D. x 1/4” D (22 mm x 7 mm)).

References

1. *In situ X-ray crystallography. A. McPherson. J. Appl. Cryst. (2000). 33, 397-400.*

Order Information

Cat. No.	Name	Description	Price
HR3-136	**Sandwich Box Setup**	contains 6 siliconized, 9 well glass plates, 6 plastic supports, and 6 sandwich boxes with covers	**$299.00**
HR3-134	**Siliconized 9 Well Glass Plate**	6 plate pack	**$273.00**
HR3-132	**Sandwich Box with cover**	40 box case	**$93.00**

SEALANTS, SEALING GREASE & OILS

IMMERSION OIL (TYPE A, B AND NVH)

APPLICATIONS

- Useful for sealing cover slides to crystallization plates
- Cryoprotection

DESCRIPTION

Useful for sealing cover slides to crystallization plates. Type A oil has low viscosity (150 centistokes), Type B has medium viscosity (1250 centistokes), Type 300 has medium/low viscosity (300 centistokes), and Type NVH has a very high viscosity (21,000 centistokes). This works well for incubations above 25°C.

Order Information

Cat. No.	Name	Description	Price
HR3-611	**Sample Pack**	Contains the following: Type A - 14.8 ml, Type B - 7.4 ml, Type 300 - 7.4 ml, Type NVH - 7.4 ml	**$16.00**
HR3-613	**Type A**	4 oz bottle	**$16.00**
HR3-615	**Type B**	4 oz bottle	**$16.00**
HR3-617	**Type NVH**	4 oz bottle	**$18.00**

DOW CORNING® VACUUM GREASE

APPLICATION

- Standard sealant for hanging, sitting, and sandwich drop vapor diffusion setups

DESCRIPTION

Dow Corning® Vacuum Grease. The standard sealant for hanging, sitting, and sandwich drop vapor diffusion setups. Stiff, non-melting, non-drying silicone sealant maintains its consistency from -40° to 260°C. Chemical-resistant and non-carbonaceous. Squeeze the grease into a 10 cc syringe fitted with a 200 μl pipet tip trimmed to desired diameter and you are ready to grease, or try the Grease Kit™.

DOW CORNING
high vacuum grease

Order Information

Cat. No.	Name	Description	Price
HR3-510	**Dow Corning Vacuum Grease**	150 g tube	**$34.00**

DOW CORNING® 7 RELEASE COMPOUND GREASE

DESCRIPTION

The Dow Corning® 7 Release Compound Grease sealant is similar to vacuum grease but is less viscous. Being less viscous, it is easier to push through a syringe when applying the sealant manually. Squeeze the grease into a 10 cc syringe fitted with a 200 μl pipet tip trimmed to desired diameter and you are ready to grease.

Order Information

Cat. No.	Name	Description	Price
HR3-508	**Dow Corning 7 Release Compound Grease**	150 g tube	**$22.00**

REPLACEMENT GREASE CARTRIDGES

APPLICATION

- For the Grease Machine L-100

DESCRIPTION

Replacement grease cartridges for the Grease Machine L-100

Order Information

Cat. No.	Name	Description	Price
HR3-202	**Grease Cartridge**	5 pack	**$56.00**

CRYSTAL CLEAR SEALING FILM

APPLICATION

- Sealing film used to seal sitting drop crystallization experiments

FEATURES

- Fits SBS format microplates
- No special applicator required
- Optically clear

DESCRIPTION

Crystal Clear Sealing Film is an optically transparent sealing film for SBS format 24, 48, 96 and 384 well plates. The 2 mil (0.05 mm) thick, optically clear, non-fluorescing polyester film with 1 mil (0.025 mm) thick acrylic custom adhesive is designed for the convenient and secure sealing of crystallization plates. A split, optically opaque polyester backing with two end tabs assures a clean, uniform, adhesive layer and allows for easy and accurate positioning on the plate. Before use, peel off the opaque center protective polyester backing to reveal the optically clear sealing film. Film length: 5.625" Film width: 3.125"; End tabs: 0.375"; Perforations: 0.415" from edge of film; Corner radius: 3/32". DNase-, RNase-, and nucleic-acid-free, non-sterile. Recommended temperatures from -40 to +120°C.

Order Information

Cat. No.	Name	Description	Price
HR3-609	**Crystal Clear Sealing Film**	100 pack	**$108.00**

CLEARSEAL FILM™ AND APPLICATOR

APPLICATION

- Sealing film used to seal sitting drop crystallization experiments

FEATURES

- Optically clear
- Easy to handle
- Pressure sensitive adhesive
- For best results use specially designed applicator

DESCRIPTION

ClearSeal Film fits all SBS format 96 well plates. Minimal evaporation, minimal cross-well contamination, easy to handle, and optically clear. Single coated, 2 mil (0.05 mm), clear polyolefin film with pressure-sensitive silicone-based adhesive. Supplied as a 141.3 mm x 79.4 mm film with polyester release liner (backing). Seals well to polypropylene, polystyrene, and polycarbonate plates. Compatible with aqueous and organic solvents. Suitable for use between -70 and 110°C. For best seal, the plate must be sealed with the Sealing Film Applicator. Limited shelf life. Must be used before expiration date.

The Sealing Film Applicator is a specially designed tool for the proper application of the ClearSeal Film. The design and rigid plastic allows for the application of even and consistent pressure which releases the pressure sensitive adhesive on the film to properly seal the film to the plate.

Order information

Cat. No.	Name	Description	Price
HR4-523	**ClearSeal Film**	25 pack	**$56.00**
HR4-521	**ClearSeal Film**	100 pack	**$198.00**
HR4-525	**Sealing Film Applicator**	each	**$5.00**

UVP HANGING DROP MRC PLATE SEAL

APPLICATION

- Hanging drop sealing film for MRC plates

FEATURES

- Designed for use with the 2 or 3 drop MRC plates
- UV compatible
- X-ray diffraction capability

DESCRIPTION

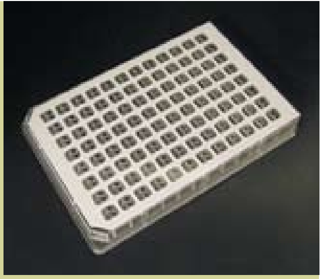

The new revolutionary Swissci AG UVP Hanging Drop Crystallization Plate Seal comes ready-to-use with dust free protective coating and a specialty polymer base. The seals accommodate up to 3 separate drops of protein and fit the standard 2 or 3 drop MRC plates. The product has a 100 micron thin layer of UVP specialty polymer developed for compact drop creation and ability to shoot x-ray without any noticeable diffraction.

The seals can be run with drops in the 96, 192 or 288 drop positions. The seal covers sit directly over the microplate crystallization reagent wells and are closed with a long-term resistant closing adhesive which is validated not to ingress into sample, nor lead to seal corruption. To remove an individual crystal, the seal may be peeled back or simply cut out with a scalpel. The material is very thin, although still remaining as a perfect evaporation barrier and can be sliced with a knife very easily.

The plate seals are produced in a class 10,000 clean room environment and guaranteed to be dust and scratch free. Optical quality is of the highest level available. The seals are able to be used in the UV range thus enabling the user to recognize salt from protein crystals under the ultra violet light source microscope.

Order Information

Cat. No.	Name	Description	Price
HR3-607	UVP Hanging Drop MRC Plate Seal	50 pack	$470.00

CRYSTAL CLEAR SEALING TAPE

APPLICATION

- Sealing tape used to seal sitting drop crystallization experiments

FEATURES

- Optically clear
- Compatible with a wide range of crystallization reagents

Plate	1.88 inch	3 inch
Corning®		X
Cryschem™	X	
Cryschem™ 24-1 SBS		X
Douglas Instruments CrystalClear Strips™		X
Greiner		X
Intelli-Plate™		X
MRC/Swissci		X

DESCRIPTION

These optically clear tapes are compatible with protein crystallization reagents. The Crystal Clear Sealing Tapes use solvent based adhesives which are compatible with aqueous crystallization reagents.

Catalog number HR3-511 is a 1.88 inch (48 mm) wide, 3 mil tape on a 43.7 yard (40 M) roll with a 1.5 inch core and is supplied with a green dispenser/cutter.

Catalog number HR4-511 is a 1.88 inch wide, 3 mil tape on a 60 yard roll with a 3 inch core and no dispenser/cutter.

Two strips of the HR3-511 or HR4-511 will seal a Corning®, Cryschem™, CrystalClear Strip™, Greiner, Intelli-Plate™, Linbro®, MRC, VDX™ and VDXm™ plate.

Catalog number HR4-506 is a 3 inch wide, 3 mil tape on a 54.86 yard roll with a 3 inch core and no dispenser/cutter.

One strip of the HR4-506 will seal a Corning, Cryschem 24-1 SBS, CrystalClear Strip, Greiner, Intelli-Plate, MRC and VDXm plate.

See chart for tape and plate compatibility.

Order Information

Cat. No.	Name	Description	Price
HR3-511	1.88 inch wide Crystal Clear Sealing Tape	1.88 inch roll, x 43.7 yard, with cutter	$7.00
HR4-511	1.88 inch wide Crystal Clear Sealing Tape	1.88 inch x 60 yard roll	$9.00
HR4-506	3 inch wide Crystal Clear Sealing Tape	3 inch x 55 yard roll	$10.00

ALUMASEAL™ II SEALING FILM AND APPLICATOR

APPLICATION

- Sealing film used to reseal HT format screen kits

FEATURES

- Excellent seal
- Film conforms to raised chimney wells
- Easily pierceable with single or multichannel pipetters and robotic probes
- Less evaporation than clear films
- Heat & cold resistant, recommended for temperatures from -80 °C to +120 °C
- Certified DNase-, RNase-, and nucleic-acid-free
- Excellent barrier properties, virtually no reagent evaporation or drying

DESCRIPTION

A 38 μm soft non-permeable aluminum foil sealing film with strong medical-grade adhesive, AlumaSeal II sealing films eliminate the need for heat-sealing devices or mats during the resealing of reagents in Deep Well blocks. Each sealing film measures 82.6 x 142.9 mm and offers sufficient sealing area for all 96 Deep Well blocks. Length between the perforations with end tabs removed is 125.4 mm. Compared to other aluminum foils, AlumaSeal II has less tendency to roll back on itself when removing the backing paper and conforms well to the plate during application.

AlumaSeal II is a soft, pierceable adhesive film designed for the convenient and rapid sealing of reagent blocks. A multiple split backing with two end tabs allows for easy, accurate positioning and secure sealing. The use of an adhesive sealing film minimizes evaporation and helps to prevent well-to-well cross contamination in reagent blocks. AlumaSeal II films are easily pierced by pipettte tips or robotic probes or piercing tools for direct reagent recovery without significant gumming by adhesive.

Order Information

Cat. No.	Name	Description	Price
HR8-069	**AlumaSeal II Sealing Film**	100 pack	**$65.00**
HR4-525	**Sealing Film Applicator**	each	**$5.00**

CAP MAT FOR MASTERBLOCK®

APPLICATION

- Snap seal MASTERBLOCK 96 Deep Well plate.

FEATURE

- Polypropylene

DESCRIPTION

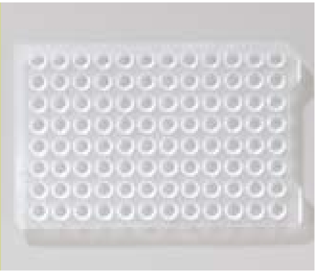

Polypropylene mat snap seals MASTERBLOCK 96 Deep Well plate.

Order information

Cat. No.	Name	Description	Price
HR3-103	**Cap Mat for MASTERBLOCK**	50 pack	**$160.00**

ROBOLID™

APPLICATION

- Deep Well block cover for HT screens

FEATURES

- Lid dimensions are suitable for a wide range of standard microplates and liquid handling work stations
- No well-to-well contamination to ensure multiple application and removal of lid
- Silicone sealing plugs are suitable with organic solvents and low extractables
- Designed for low evaporation under standard storage conditions from -20 to 37°C

DESCRIPTION

The Robolid combines the sealing abilities of a silicone rubber mat and the automation friendliness of a polystyrene lid. Essentially, it is a silicone mat with tapered well plugs specially bonded to a polystyrene plate cover. It is compatible with a wide range of liquid handling automation. The Robolid can be used to cover a 96 Deep Well block for low evaporation, short-term storage, or to minimize evaporation or contamination during pipetting. The Robolid is compatible with the 96 Deep Well blocks used for Hampton Research HT kits such as Crystal Screen HT™, Index HT™, and SaltRx HT™.

Order Information

Cat. No.	Name	Description	Price
HR3-111	**Robolid**	25 pack	**$223.00**

How To: Using Dialysis Buttons

Description

Crystallization by dialysis is an easy variation to the typical vapor diffusion method used to grow crystals. In the dialysis method, the sample in question is separated from the "precipitant" by a semi-permeable membrane which allows small molecules such as ions, additives, buffers, and, salts to pass but prevents biological macromolecules from crossing the membrane. Equilibration kinetics depend upon the molecular weight cutoff of the dialysis membrane, the precipitant, the ratio of the volume, the concentration of the components inside and outside of the dialysis cell, and the geometry of the cell.

Step 1 - Place sample into dialysis cell

Pipet sample into the dialysis chamber. Amount of sample can vary depending on the size of the Dialysis Button.

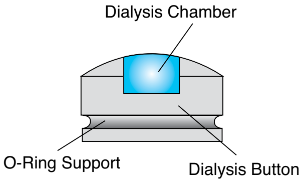

Step 2 - Placing the membrane

Choose a dialysis membrane that best suits your molecular weight cutoff of choice. Place the membrane across the top of the Dialysis Button. Using a golf tee or applicator, press down on the membrane firmly, making sure no air is trapped between the membrane and dialysis chamber. Roll the O-ring over the golf tee and onto the Dialysis Button. Make sure the O-ring is placed securely in the O-ring support, holding down the membrane.

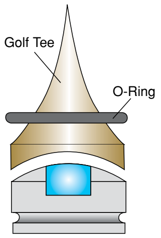

Step 3 - Begin dialysis

Place the sealed Dialysis Button into a VDX™ Plate or Linbro® Plate. Fill the reservoir with crystallization reagent and let the dialysis begin.

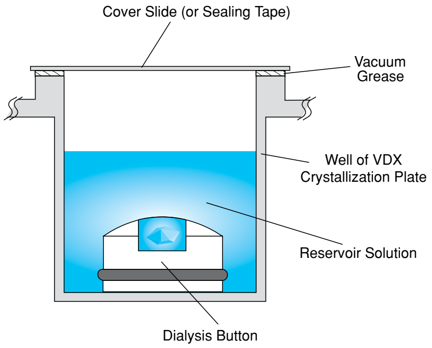
Cover Slide (or Sealing Tape)
Vacuum Grease
Well of VDX Crystallization Plate
Reservoir Solution
Dialysis Button

DIALYSIS BUTTONS™

APPLICATION

- Crystallization by dialysis, protein folding, & small volume sample dialysis

FEATURES

- Low volume dialysis
- Fits in 24 well plates
- Use over and over again

DESCRIPTION

In the dialysis method, the sample in question is separated from the "precipitant" by a semi-permeable membrane which allows small reagent molecules to pass but prevents biological macromolecules from crossing the membrane.[1-3] Dialysis Buttons are either machined from transparent perspex or injection molded from polystyrene, and are the size of a small button. The sample is placed in this chamber so as to create a slight dome of liquid at the top of the button. A dialysis membrane is placed over the top of the button/sample and is held in place with an O-ring. The O-ring is held in place by a groove in the Dialysis Button. Dialysis Buttons are supplied with O-rings. The golf tee is supplied with the 5 μl to 100 μl buttons.

References

1. *McPherson, A., Preparation and Analysis of Protein Crystals, Krieger Publishing, 88-91 (1992).*
2. *Durcruix, A, and Giege, R., Crystallization of Nucleic Acids and Proteins, A Practical Approach, Oxford University Press, 78-82 (1992).*
3. *Zeppenzauer, M., et al., Acta Chem Scan (1967) 21, 1009.*

Order Information

Cat. No.	Name	Description	Price
HR3-336	**Dialysis Buttons Sampler**	5 of each size	**$88.00**
HR3-314	**5 µl Dialysis Button**	50 pack	**$88.00**
HR3-316	**10 µl Dialysis Button**	50 pack	**$88.00**
HR3-318	**15 µl Dialysis Button**	50 pack	**$88.00**
HR3-320	**20 µl Dialysis Button**	50 pack	**$88.00**
HR3-322	**25 µl Dialysis Button**	50 pack	**$88.00**
HR3-326	**50 µl Dialysis Button**	50 pack	**$88.00**
HR3-328	**100 µl Dialysis Button**	50 pack	**$88.00**
HR3-330	**200 µl Dialysis Button**	50 pack	**$88.00**
HR3-332	**350 µl Dialysis Button**	50 pack	**$88.00**

APPLICATOR FOR DIALYSIS BUTTONS™

APPLICATION

- Application of dialysis membrane to Dialysis Buttons

FEATURES

- Manufactured of glass
- Two sizes available
  - Small Applicator (5 - 100 µl Dialysis Buttons)
  - Large Applicator (200 & 350 µl Dialysis Buttons)

DESCRIPTION

The Applicator makes easier work of applying the dialysis membrane and O-ring to Dialysis Buttons. It positions and holds the dialysis membrane on the buttons and also allows the easy application and position of the O-ring to secure the dialysis membrane onto the Dialysis Button. The Applicator for Dialysis Buttons is manufactured of glass and is available in two sizes. The Small Applicator (HR4-348) is designed for use with 5 to 100 μl Dialysis Buttons. The Large Applicator (HR4-350) is designed for use with 200 and 350 μl Dialysis Buttons.

Order Information

Cat. No:	Name:	Description	Price
HR4-348	**Small Applicator**	for use with 5 to 100 µl Dialysis Buttons	**$47.00**
HR4-350	**Large Applicator**	for use with 200 and 350 µl Dialysis Buttons	**$47.00**

DIALYSIS MEMBRANE DISCS™ FOR BUTTONS

APPLICATION

- Membranes for dialysis

DESCRIPTION

Spectra/Por® regenerated cellulose dialysis membrane discs are 33 mm, circular, pre-cut membranes available with the following molecular weight cutoffs:

- 3,500 MW cutoff
- 6,000 - 8,000 MW cutoff
- 12,000 - 14,000 MW cutoff

SPECTRUM
Spectra/Por
molecularporous membrane
3

Order Information

Cat. No.	Name	Description	Price
HR3-338	**Dialysis Membrane Discs**	cutoff 3,500 - 50 pack	**$195.00**
HR3-344	**Dialysis Membrane Discs**	cutoff 6,000 to 8,000 - 50 pack	**$195.00**
HR3-346	**Dialysis Membrane Discs**	cutoff 12,000 to 14,000 - 50 pack	**$195.00**
HR4-348	**Small Applicator**	for use with 5 to 100 µl Dialysis Buttons	**$47.00**
HR4-350	**Large Applicator**	for use with 200 and 350 µl Dialysis Buttons	**$47.00**

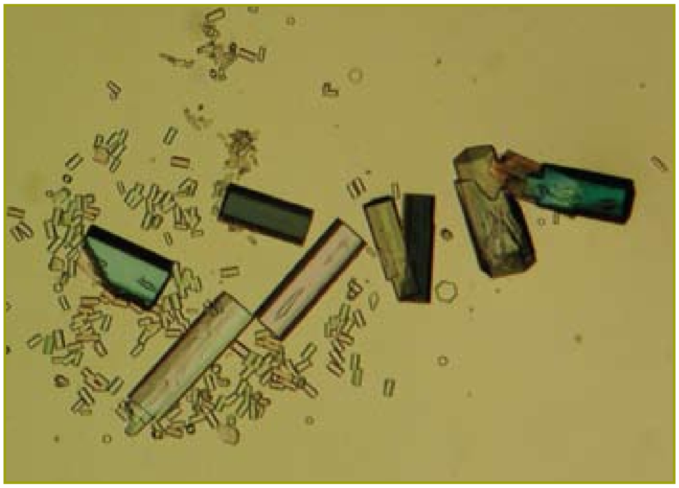

Protein crystals.

Kasumi Kobayashi, Taiji Nakae and Hiroyuki Akama.
The Kitasato Institute, Kanagawa, Japan.

GRANADA CRYSTALLIZATION BOX® & CAPILLARIES

APPLICATION

- Crystallization using gels in capillaries

FEATURES

- Crystallization inside glass or quartz capillaries
- Can be used with or without gel
- Counter-diffusion or batch method

MATERIAL

- Optically clear polystyrene

figure 1

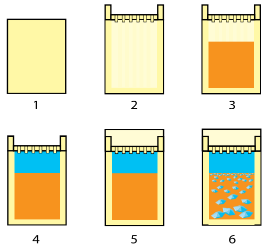

figure 2

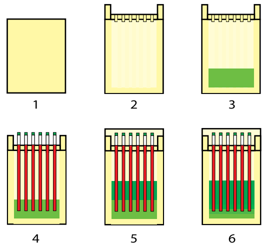

figure 3

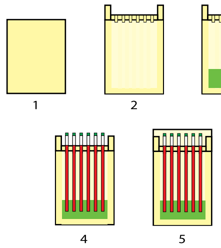

figure 4

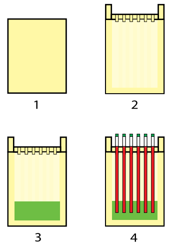

DESCRIPTION

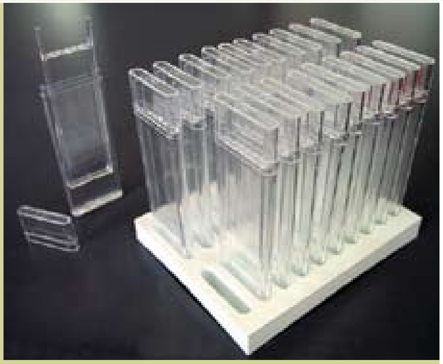

The Granada Crystallization Box (GCB) consists of four elements made of polystyrene:

1. A reservoir to introduce the gel
2. A guide to hold capillaries
3. A cover
4. A holder to maintain the boxes

The GCB has been designed to be used in four different ways:

1. To grow crystals inside gels under diffusion-controlled mass transport (figure 1).
2. To grow crystals inside capillaries with un-gelled precipitating agent by the counter-diffusion technique (figure 2).
3. To grow crystals inside capillaries with gelled precipitating agent by the counter-diffusion technique (figure 3).
4. To grow crystals inside capillaries by the batch method (figure 4).

Capillaries and gel are not included with the Granada Crystallization Box.

The capillaries described on this page are designed for use with the Granada Crystallization Box. The capillaries are extremely thin walled. The internal diameter for these capillaries is of 0.1, 0.2 and 0.3 mm. The respective external diameters are 0.17, 0.33 and 0.40 mm. The borosilicate glass capillaries are available in four different lengths of 30, 40, 50 or 100 mm. Although the capillaries are designed for use with the Granada Crystallization Box they can also be used for general free-interface diffusion crystallization.

References

1. *Granada Crystallisation Box: a new device for protein crystallisation by counter-diffusion techniques. Garcia-Ruiz JM, Gonzalez-Ramirez LA, Gavira JA, Otalora F. Acta Crystallogr D Biol Crystallogr. 2002 Oct;58 (Pt 10 Pt 1):1638-42.*
2. *Counterdiffusion methods for macromolecular crystallization. Garcia-Ruiz JM. Methods Enzymol. 2003;368:130-54.*
3. *A simplified counter diffusion method combined with a 1D simulation program for optimizing crystallization conditions. H. Tanaka, K. Inaka, S. Sugiyama, S. Takahashi, S. Sano, M. Sato and S. Yoshitomi. J. Synchrotron Rad. (2004). 11, 45-48.*

Order Information

Cat. No.	Name	Description	Price
HR3-194	**Granada Crystallization Box**	20 plates and 1 holder	**$261.00**
HR4-677	**Round Capillary**	100 mm x 0.1 mm - 25 pack	**$30.00**
HR4-678	**Round Capillary**	100 mm x 0.2 mm - 25 pack	**$30.00**
HR4-679	**Round Capillary**	100 mm x 0.3 mm - 25 pack	**$30.00**
HR4-680	**Round Capillary**	50 mm x 0.1 mm - 25 pack	**$26.00**
HR4-681	**Round Capillary**	50 mm x 0.2 mm - 25 pack	**$26.00**
HR4-682	**Round Capillary**	50 mm x 0.3 mm - 25 pack	**$26.00**
HR4-683	**Round Capillary**	40 mm x 0.1 mm - 25 pack	**$26.00**
HR4-684	**Round Capillary**	40 mm x 0.2 mm - 25 pack	**$26.00**
HR4-685	**Round Capillary**	40 mm x 0.3 mm - 25 pack	**$26.00**
HR4-686	**Round Capillary**	30 mm x 0.1 mm - 25 pack	**$26.00**
HR4-687	**Round Capillary**	30 mm x 0.2 mm - 25 pack	**$26.00**
HR4-688	**Round Capillary**	30 mm x 0.3 mm - 25 pack	**$26.00**
HR8-092	**LM Agarose**	10 g bottle	**$58.00**

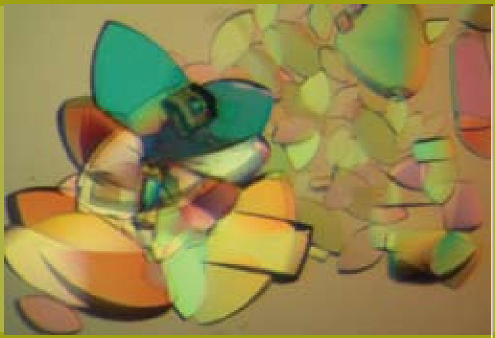

Crystals of the Deinococcus radiodurans 50S ribosome subunit.

From the group of Paola Fucini, Max Planck Institute for Molecular Genetics, Berlin, Germany.

Protein crystals.

Kasumi Kobayashi, Taiji Nakae and Hiroyuki Akama. The Kitasato Institute, Kanagawa, Japan.

Deoxyribonuclease crystals grown in dAMP, GDP, naladixic acid, 15% w/v Polyethylene glycol 3,350.

John Day, University of California Irvine, USA.

TOOLS, SEEDING & RESIN

Crystal image "UFO".

Chongping Chen, Dr. Jaffe Lab, Fox Chase Cancer Research Center.

TABLE OF CONTENTS

PAGES

108 MICRO-TOOLS™

109 MICRO-TOOLS II™

110-113 FORCEPS

113 CRYSTAL PROBE™

113 CRYSTAL PENCIL™

114 SEED BEAD™

115 SEEDING TOOL

115 CHELATING RESIN

HAMPTON RESEARCH

MICRO-TOOLS™

APPLICATION

- Precision instruments for crystal manipulation

FEATURES

- Made from chemical resistant, hardened tool steel
- Interchangeable tools tips

SUCCESS STORY

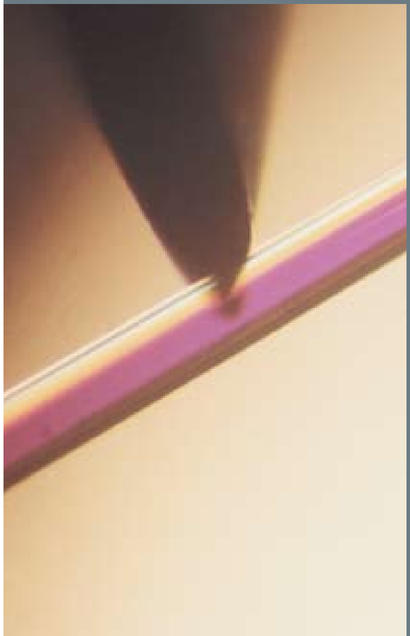

Crystal shown being cut using the Hampton Research Micro-Knife.

Courtesy of Lisa Edberg.

University of Alabama, Birmingham
Center for Macromolecular Crystallography

DESCRIPTION

Micro-Tools are the smallest available precision instruments for laboratory use. These tools are realistically proportioned for macromolecular and small molecule crystallization work. The high quality, precision Micro-Tools ease microscopic manipulations. The eight specifically engineered Micro-Tools are useful in a variety of applications in the crystallization lab. Eight tools, a handle and wood storage box.

Micro-Tools are designed to be versatile as well as useful for specific crystallization manipulations. For example, the Micro-Scale is very useful for measuring crystal dimensions as well as serving as a way to record the dimensions by photography for reference or publication. The Micro-Chisel and Micro-Spade are helpful for separating and splitting blade and whisker clusters during seeding or mounting. The Micro-Prober and Micro-Needle come in handy when probing precipitate or when manipulating small or large crystals for seeding as well as during mounting. The Micro-Spatula can be used to move crystals as well as split single crystals or clusters. The Micro-Scraper is helpful when working with very small crystals.

The Micro-Tools Set is indispensable during crystal observation, manipulation, seeding, and mounting. The set includes a single anodized aluminum handle which holds each of the eight unique tips, all supplied in an attractive alderwood instrument case. The case provides protective storage of the Micro-Tools as well as a convenient carrying case.

The tips are constructed from hardened tungsten steel and are resistant to most common reagents used for crystal growth. The threaded tool tips are easily interchanged with the threaded aluminum handle. The diameter of the tool shaft is 0.250 mm (0.010 inches). The actual dimension at the very tip of the tool will vary with the particular tool design.

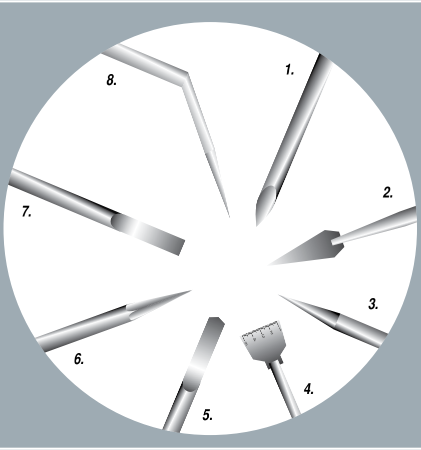

Micro-Tool Set:
1. Micro-Scraper
2. Micro-Spatula
3. Micro-Needle
4. Micro-Scale
5. Micro-Spade
6. Micro-Knife 20°
7. Micro-Chisel
8. Micro-Prober 45°

Order Information

Cat. No.	Name	Description	Price
HR4-811	**Micro-Tools Set**	each	**$298.00**
HR4-813	**Micro-Scale**	each	**$55.00**
HR4-815	**Micro-Spade**	each	**$48.00**
HR4-817	**Micro-Prober 45°**	each	**$48.00**
HR4-819	**Micro-Scraper**	each	**$54.00**
HR4-821	**Micro-Chisel**	each	**$54.00**
HR4-823	**Micro-Knife 20°**	each	**$54.00**
HR4-825	**Micro-Spatula**	each	**$52.00**
HR4-827	**Micro-Needle**	each	**$46.00**
HR4-829	**Micro-Tool Handle**	each	**$33.00**
HR4-831	**Micro-Tools Carrying Case**	each	**$56.00**

MICRO-TOOLS II™

APPLICATION

- Precision instruments for crystal manipulation

FEATURES

- Made from chemical resistant, tungsten steel and hardened tool steel
- Interchangeable tools tips
- 8 unique tools designed to extend the original Micro-Tools set

DESCRIPTION

Micro-Tools II Set is an extension to the smallest available precision instruments for laboratory use. These tools are realistically proportioned for macromolecular and small molecule crystallization work. The high quality, precision Micro-Tools II ease microscopic manipulations. The eight specifically engineered Micro-Tools II are useful in a variety of applications in the crystallization lab. Eight tools, a handle and wood storage box.

The Micro-Tools II are designed to be versatile as well as useful for specific crystallization manipulations. For example, the Micro-Manipulator is handy for manipulating long, thin blades or for separating long, thin blade clusters. Awkward positions encountered during crystal manipulation can sometimes be avoided using tools with an unusual and flexible approach angle such as the Micro-Prober 90° and the Micro-Hook 90°. Seeding experiments can be performed using the Micro-Brush. The Ultra Micro-Needles are extremely delicate tools crafted from 5 micron radius tungsten that are very handy for seeding as well as manipulations requiring a soft, fine, and controlled touch.

The Micro-Tools II Set is indispensable during crystal observation, manipulation, seeding, and mounting. The set includes a single anodized aluminum handle which holds each of the eight unique tips, all supplied in an attractive alderwood instrument case. The case provides protective storage of the Micro-Tools as well as a convenient carrying case.

The tips are constructed from hardened tungsten steel and are resistant to most common reagents used for crystal growth. The threaded tool tips are easily interchanged with the threaded aluminum handle. The diameter of the tool shaft is 0.250 mm (0.010 inches). The actual dimension at the very tip of the tool will vary with the particular tool design.

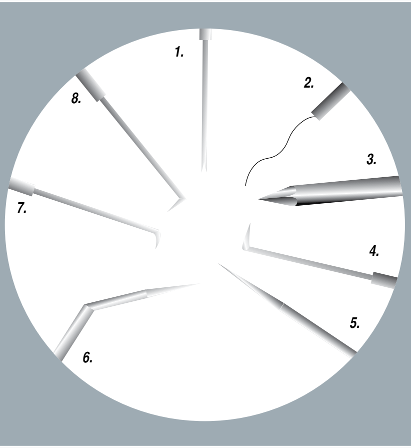

Micro-Tool II Set :
1. Manipulator
2. Micro-Brush
3. Micro-Knife 45°
4. Micro-Prober 90°
5. Ultra Micro-Needle, straight, tungsten, 5 micron radius
6. Ultra Micro-Needle, bent 30°, tungsten, 5 micron radius
7. Micro-Hook 90°
8. Ultra Micro-Needle, bent 90°, tungsten, 5 micron radius

Order Information

Cat. No.	Name	Description	Price
HR4-837	**Micro-Tools II Set**	each	**$298.00**
HR4-839	**Micro-Manipulator**	each	**$49.00**
HR4-841	**Micro-Knife 45°**	each	**$49.00**
HR4-843	**Micro-Prober 90°**	each	**$48.00**
HR4-845	**Micro-Brush**	each	**$44.00**
HR4-847	**Micro-Hook 90°**	each	**$49.00**
HR4-849	**Ultra Micro-Needle, Straight**	each	**$41.00**
HR4-851	**Ultra Micro-Needle 30°**	each	**$49.00**
HR4-853	**Ultra Micro-Needle 90°**	each	**$49.00**
HR4-829	**Micro-Tool Handle**	each	**$33.00**
HR4-831	**Micro-Tools Carrying Case**	each	**$56.00**

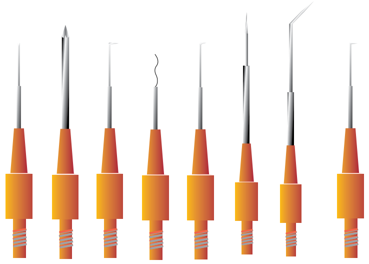

FORCEPS

ULTRA-FINE FORCEP

FEATURES

- Length - 115 mm
- Tip - 9.5 mm to point

DESCRIPTION

High quality, precision straight forceps with an extra fine point. Smooth, non-serrated tip and handle. Great for mounting CryoLoops.

Order Information

Cat. No.	Name	Description	Price
HR4-855	Ultra-Fine Forcep	each	$10.00

SLIDE-TENSION FORCEP

FEATURES

- Length - 150 mm
- Tip - 10.3 mm to point

DESCRIPTION

Straight forcep with adjustable "locking" tension bar. Increase the width of the tip opening by setting the bar to the desired position. Slide bar can be positioned to "lock" forceps and hold cover slides, CryoLoops, MicroTubes, and other small parts.

Order Information

Cat. No.	Name	Description	Price
HR4-857	Slide-Tension Forcep	each	$8.00

45° ANGLED FORCEP

FEATURES

- Length - 137 mm
- Tip - 2.5 mm to point

DESCRIPTION

All purpose angled forcep with blunt end. Tip has serration for firm grip. Good for retrieving. 45° angle makes for easy manipulation. Especially nice for flipping cover slides. Has serrated non-slip handle.

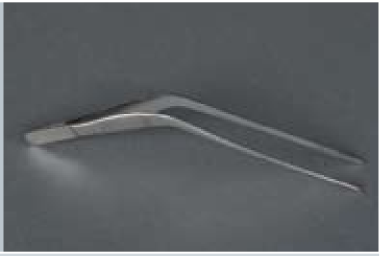

Order Information

Cat. No.	Name	Description	Price
HR4-859	45° Angled Forcep	each	$10.00

30° TIP ANGLED FORCEP

FEATURES

- Length - 119 mm
- Head Length - 11.5 mm
- Head Width - 6.1 mm

DESCRIPTION

Smooth, wide, angled, flat tip that will leave delicate items undamaged, unlike tweezers with sharp points. Tip is angled down 30° from the handle. Has serrated, non-slip handle. Good for handling cover slides and filter paper.

Order Information

Cat. No.	Name	Description	Price
HR4-861	30° Tip Angled Forcep	each	$8.00

SMALL FORCEP

FEATURES

- Length - 140 mm
- Tip Width - 2.5 mm

DESCRIPTION

All purpose straight forcep with blunt end. Tip has serrated surface for firm grip. Good for retrieving. Has serrated non-slip handle.

Order Information

Cat. No.	Name	Description	Price
HR4-863	**Small Forcep**	each	**$9.00**

MEDIUM FORCEP

FEATURES

- Length - 200 mm
- Tip Width - 2.5 mm

DESCRIPTION

All purpose straight forcep with blunt end. Tip has serrated surface for firm grip. Good for retrieving. Has serrated non-slip handle.

Order Information

Cat. No.	Name	Description	Price
HR4-865	**Medium Forcep**	each	**$13.00**

EXTRA LARGE FORCEP

FEATURES

- Length - 300 mm
- Tip Width - 2.5 mm

DESCRIPTION

All purpose straight forcep with blunt end. Tip has serrated surface for firm grip. Good for retrieving. Has serrated non-slip handle.

Order Information

Cat. No.	Name	Description	Price
HR4-869	**Extra Large Forcep**	each	**$10.00**

FORCEPS

STRAIGHT MICROFORCEP

FEATURES

- Length - 113 mm
- Tip Length - 31 mm
- Tip Width - 3.1 mm to point

DESCRIPTION

Ultra-thin, smooth, pointed tip for very precise work. Great for retrieving or holding small parts. Smooth, non-serrated handle and tip.

Order Information

Cat. No.	Name	Description	Price
HR4-871	**Straight Microforcep**	each	**$13.00**

45° TIP ANGLED FORCEP

FEATURES

- Length - 113 mm
- Tip Length - 10.9 mm
- Tip Width - 3.5 mm to point

DESCRIPTION

Straight forceps with angled tip. Easy to use forceps when space is limited. Tip is slanted at a 45° angle. Smooth, non-serrated handle and tip.

Order Information

Cat. No.	Name	Description	Price
HR4-875	**45° Tip Angled Forcep**	each	**$8.00**

CURVED TIP MICROFORCEP

FEATURES

- Length - 113 mm
- Tip Length - Curved to point

DESCRIPTION

Ultra-thin, smooth, pointed tip for very precise work. Great for retrieving or holding small parts. Smooth, non-serrated handle and tip.

Order Information

Cat. No.	Name	Description	Price
HR4-879	**Curved Tip Microforcep**	each	**$9.00**

ANGLED TIP LOCKING FORCEP

FEATURES

- Length - 156 mm
- Tip Length - 18 mm
- Tip Width - 1.7 mm

DESCRIPTION

All purpose angled tip forcep. Tip has serrated surface for firm grip. Good for retrieving. When closed and squeezed, forcep locks to clamp and hold item in place. Lock is released by moving "lock" toward tip with thumb. Serrated, non-slip handle. Angled for easy manipulation.

Order Information

Cat. No.	Name	Description	Price
HR4-881	**Angled Tip Locking Forcep**	each	**$13.00**

LOCKING FORCEP

FEATURES

- Length - 156 mm
- Tip Width- 1.7 mm

DESCRIPTION

All purpose straight forcep. Tip has serrated surface for firm grip. Good for retrieving. When closed and squeezed, forcep locks to clamp and hold item in place. Lock is released by moving “lock” toward tip with thumb. Serrated, non-slip handle.

Order Information

Cat. No.	Name	Description	Price
HR4-883	**Locking Forcep**	each	**$13.00**

CRYSTAL PROBE™

APPLICATIONS

- Crystal manipulation
- Seeding

FEATURES

- Stainless steel probe with plastic handle

DESCRIPTION

Disposable Crystal Probe manipulators made from a 0.12 mm x 30 mm stainless steel, pointed-end probe attached to a 20 mm plastic handle. Useful for breaking apart blade clusters and general crystal manipulation. Small size and disposable format makes them convenient for trips to the synchrotron.

Order Information

Cat. No.	Name	Description	Price
HR4-217	**Crystal Probe**	12 pack	**$15.00**

CRYSTAL PENCIL™

APPLICATIONS

- Crystal manipulation
- Seeding
- Crystal mounting

DESCRIPTION

Useful for manipulating crystals for seeding and cryocrystallography. The 0.7 mm tip accepts the Hampton Research Mounted CryoLoops™ or MicroTubes™ with your own cryoloop which creates an easy to hold, comfortable, and convenient tool for manipulating crystals. CryoLoops, MicroTubes, and Mounted CryoLoops sold separately.

Order Information

Cat. No.	Name	Description	Price
HR4-835	**Mechanical Crystal Pencil**	each	**$3.00**

SEED BEAD™

APPLICATION

- Generate seeds of protein crystals

FEATURES

- Easily generate consistent seed stocks
- Use serial dilution to control the number of seeds introduced into the drop

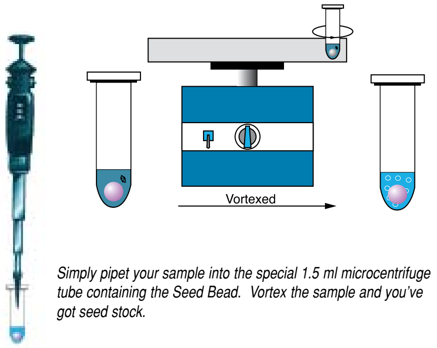

Simply pipet your sample into the special 1.5 ml microcentrifuge tube containing the Seed Bead. Vortex the sample and you've got seed stock.

DESCRIPTION

The Seed Bead kit is used to create a seed stock for performing subsequent seeding experiments. The Seed Bead kit contains 24 Seed Beads manufactured from PTFE, individually contained in a special 1.5 ml microcentrifuge tube.

Seeding allows one to grow crystals in the metastable zone. Crystallization in this zone provides control, reproducibility and an improved likelihood of a successful crystallization experiment. Also, crystals can grow from seeds but cannot spontaneously nucleate. By placing a seed or solution of seeds in a drop which is saturated to the metastable zone, one can use the seeds to grow larger single crystals. By controlling the number of seeds introduced into the drop, one can control the number of crystals grown. It is not practically possible to measure and know the number of seeds introduced to a drop, but by performing serial dilutions from a concentrated seed stock, one can control the number of crystals grown in the drop.

Using the Seed Bead kit, one can create crystal seed stock for subsequent seeding experiments. Crystals are placed in the microcentrifuge tube with the PTFE Seed Bead and either vortexed or sonicated to generate a seed stock. Then by performing serial dilutions, one can control the number and size of crystals in the experiment.

The Seed Bead kit is useful for the preparation of seed stocks for automated and semi-automated microseeding (D'Arcy 2007, Harlos 2008).

Each kit contains 24 special microcentrifuge tubes with Seed Beads. Crystallization accessories are sold separately.

References

1. *Stura, E.A., Wilson, I.A., Methods: A Companion to Methods in Enzymology (1990) 1, 38-49.*
2. *Stura, E.A., Wilson, I.A., "Seeding Techniques" in Crystallization of Nucleic Acids and Proteins: A Practical Approach. Oxford University Press (1992) 99-126.*
3. *Luft, J.R., DeTitta, G.T., Poster: Using Ultrasound in the preparation of Micro-Seed Stock for the Crystallization of Macromolecules, 1997 ACA Meeting, St Louis, MO. Publication in preparation.*
4. *J.R. Luft and G.T. DeTitta, Methods in Enzymology (1997) 276, 110-131.*
5. *Structure of an orthorhombic form of xylanase II from Trichoderma reesei and analysis of thermal displacement. Watanabe et al. Acta Cryst. (2006). D62, 784-792.*
6. *Crystallization and preliminary crystallographic analysis of p40phox, a regulatory subunit of NADPH oxidase. K. Honbou, S. Yuzawa, N. N. Suzuki, Y. Fujioka, H. Sumimoto and F. Inagaki. Acta Cryst. (2006). F62, 1021-1023 (Used seed bead to optimize)*
7. *Purification, crystallization and preliminary X-ray diffraction study of human ribosomal protein L10 core domain. Yuji Kobayashi et al. Acta Cryst. (2007). F63, 950–952*
8. *Semi-automated microseeding of nanolitre. Thomas S. Walter, Erika J. Mancini, Jan Kadlec, Stephen C. Graham, Rene´ Assenberg, Jingshan Ren, Sarah Sainsbury, Raymond J. Owens, David I. Stuart, Jonathan M. Grimes and Karl Harlos Acta Cryst. (2008). F64, 14–18*
9. *An automated microseed matrix-screening method. Allan D'Arcy,a* Frederic Villard and May Marsh. Acta Cryst. (2007). D63, 550–554*

Order Information

Cat. No.	**Name**	**Description**	**Price**
HR2-320	**Seed Bead kit**	24 tubes with Seed Beads	**$56.00**

SEEDING TOOL

APPLICATION

- Streak Seeding

FEATURES

- Natural fiber
- Built-in handle/cover

DESCRIPTION

During streak seeding, one touches the Seeding Tool to crystalline material to dislodge, remove and transfer small crystals (seeds) to a drop that will support the growth of potentially larger and more perfect crystals. A seed can provide a template on which additional macromolecules can assemble and under the proper conditions, grow to form a large single crystal. Using seeding can avoid problems associated with growing crystals from spontaneous nucleation. Seeds can grow into larger crystals in the metastable region of the solubility curve, which is a region of lower, relative supersaturation. One can also streak seed from phase separation or amorphous material as a diagnostic to confirm whether the material is crystalline. The Seeding Tool is a 1 cm natural fiber attached to a stainless steel pin and plastic handle. The Seeding Tool is supplied with a cover to protect the fiber and when the cover is placed on the back side of the Seeding Tool, it makes an excellent handle.

References

1. *Seeds to crystals. Terese Bergfors. Journal of Structrual Biology 142 (2003) 66-76.*
2. *Applications of the streak seeding technique in protein crystallization. Stura and Wilson. Journal of CrystaL Growth, 110 (1991) 270-282.*

Order Information

Cat. No.	Name	Description	Price
HR8-133	**Seeding Tool**	5 pack	**$30.00**

CHELATING RESIN™

APPLICATION

- Remove trace metals from water, reagents, & samples

FEATURES

- Scavenge multivalent metal ion contaminants
- Remove nickel from sample following purification over a nickel column

DESCRIPTION

Chelating Resin can be used to remove trace metals from water, reagents, and macromolecular samples[1-3].

Chelating Resin has a high preference for mercury, copper, nickel, lead, zinc, cobalt, cadmium, iron, manganese, barium, calcium, strontium, and magnesium. Its selectivity for divalent ions over monovalent ions is high (5,000:1) and it has a very high attraction for transition metals, even in concentrated salt solutions. The resin will scavenge multivalent metal ion contaminants without altering the concentration of nonmetallic ions. In most cases the resin does not have any effect on protein concentration or activity.

Chelating Resin is a high purity, 100-200 mesh resin. This resin is a styrene divinylbenzene copolymer containing paired iminodiacetate ions which act as chelating groups in binding polyvalent metal ions. The nominal capacity is 2.0 meq/dry gram or 0.4 meq/ml resin bed. The density is 0.65 g/ml. The resin can be classed with weakly acidic cation exchange resins due to the presence of carboxylic acid groups but differs from ordinary exchange resins due to its high selectivity for metal ions and high bond strength. The resin is stable over the entire pH range and functionally active from pH 2 to 14. The resin is stable for at least two years when stored in the original sealed container at 23°C.

References

1. *Removal of Nickel from protein solutions following purification over a Nickel column. Hassell, A., Recent Advances in Macromolecular Crystallization, Bischenberg, FRANCE (1997).*
2. *Dunn, M.F., Pattison, S.E., Storm, M.C., and Quiel, E., Removal of metals from enzyme solutions., Biochem. (1980) 19, 718.*
3. *Ray, W.J., Burgner, J.W., and Post, C.B., Purification of NMR reagents., Biochem. (1990) 29, 2770.*

Order Information

Cat. No.	Name	Description	Price
HR2-312	**Chelating Resin**	10 g bottle	**$38.00**

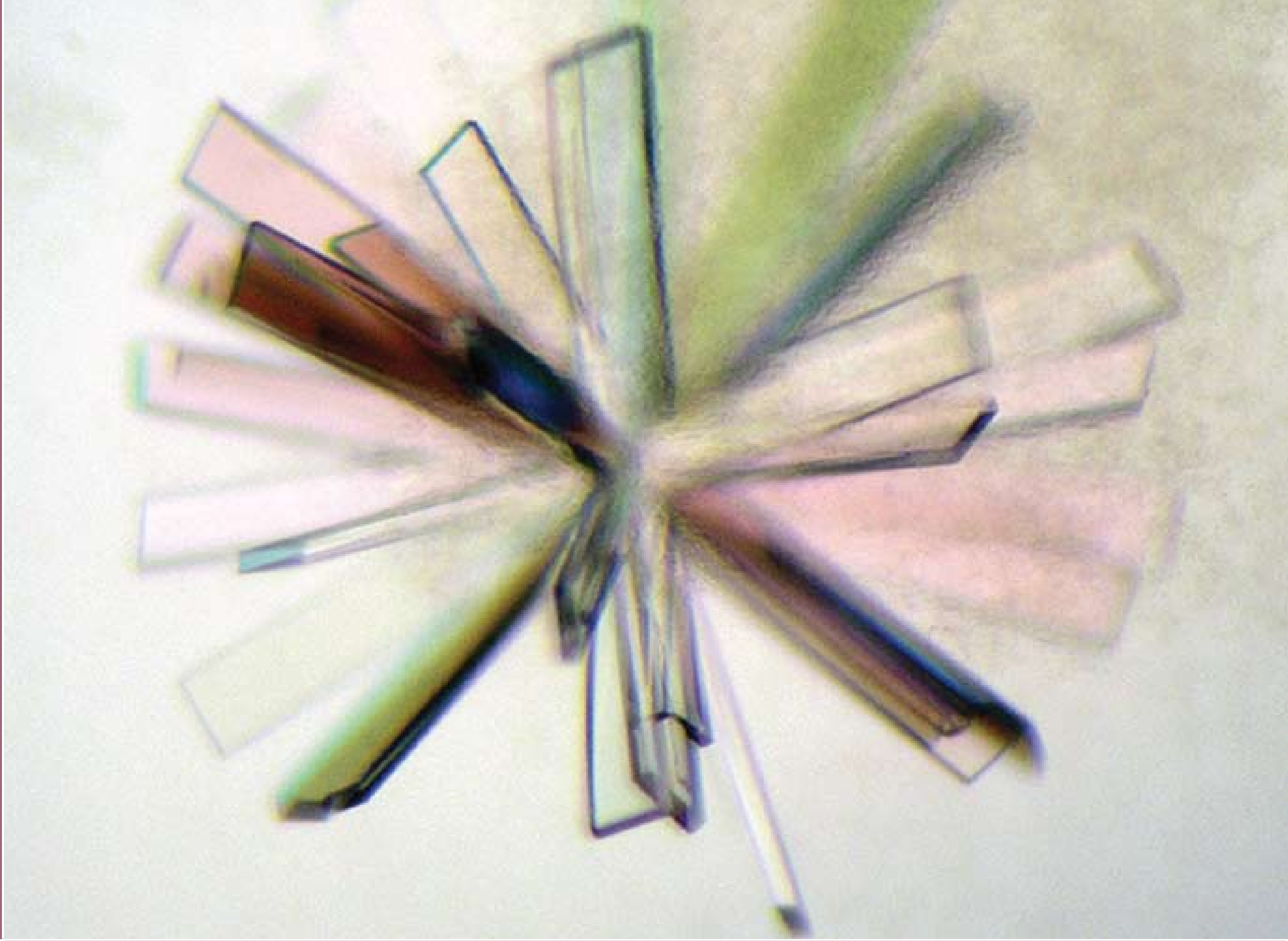

CRYOCRYSTALLOGRAPHY

A star of H. influenzae carbonic anhydrase crystals.

Roger S. Rowlett, Department of Chemistry, Colgate University, Hamilton, New York, USA.

TABLE OF CONTENTS

PAGES

118 - 120 CRYSTALCAP™ HT SYSTEMS

121 - 122 CRYSTALCAP™ MAGNETIC SYSTEMS

122- 123 CRYSTALCAP™ SYSTEMS

124 - 127 CRYOLOOPS & MICROTUBES

128 - 132 CRYO TOOLS

133 CANES, SLEEVES & CODERS

134 - 135 DEWARS

HAMPTON
RESEARCH

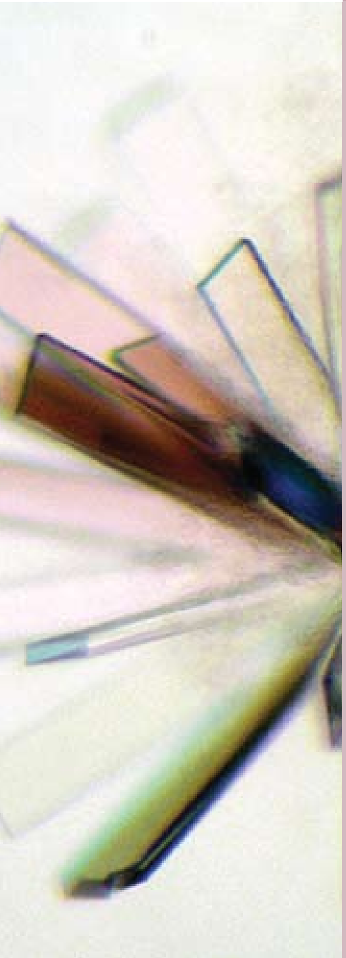

CRYSTALCAP™ COPPER MAGNETIC HT

APPLICATION

- Cryocrystallography

FEATURES

- Specially designed vial with magnetic ring
- Cap magnetically attaches to vial
- Copper jacket to reduce icing on MicroTube
- Vented vial sold separately
- No threads
- Bar coded, color coded, alphanumeric, magnetic cap
- Assembled with Mounted CryoLoops, ready-to-use; vials sold separately
- Cap features a flat ledge used by grippers such as SSRL Stanford Automated Mounting System (SAM)

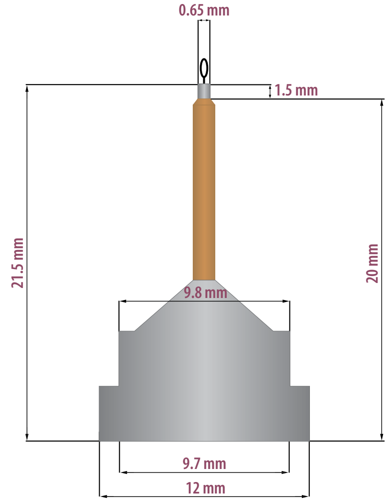

CrystalCap Copper Magnetic HT

DESCRIPTION

CrystalCap Copper Magnetic HT is designed to prevent ice formation along the copper pin/MicroTube™ in a cryostream where the flow is positioned perpendicular, at an angle, or non-collinear to the MicroTube. The CrystalCap Copper Magnetic HT is a non-threaded, 1.8 ml (approximate) cryo vial storage container with a specially engineered cap/pin for cryocrystallography. Each cap is manufactured from an alloy base which allows the cap to be magnetically secured to a magnetic base in the goniometer head. Into the alloy cap is threaded and bonded a solid, 3 mm diameter copper pin. The end of the copper pin is machined to an aerodynamic 60° and is assembled with a MicroTube and CryoLoop. A ring magnet is molded into the top end of the vial so that when the cap is positioned in the vial, the ring magnet holds the cap on the vial during cryogenic storage. The vial has two holes for venting. The 18 mm designates the distance between the top of the copper pin and the bottom of the CrystalCap Copper Magnetic where contact is made with the magnetic base.

For crystal transfer under cryo temperature use the 18 mm CryoTong (catalog number HR4-637 for the 110 mm overall tool length or catalog number HR5-112 for the 180 mm overall tool length).

The cap features a two dimensional bar coded and alphanumeric 16 x 16 data matrix. Each cap is also color coded.

HR8-181 CrystalCap Copper Magnetic HT is a two dimensional bar coded and alphanumerically labeled cap with copper pin; no Mounted CryoLoop and no vial.

Note: Caps with Mounted CryoLoops are sold without vials. Vials sold separately.

Color Coded Cap	CryoLoop Size
Red	0.025 - 0.05 mm
Green	0.05 - 0.1 mm
Yellow	0.1 - 0.2 mm
Blue	0.2 - 0.3 mm
Blue/Red	0.3 - 0.4 mm
Green/Red	0.4 - 0.5 mm
Yellow/Red	0.5 - 0.7 mm
Yellow/Green	0.7 - 1.0 mm

The CrystalCap Copper Magnetic HT is not compatible with the EMBL/ESRF SC3 sample changer. For EMBL/ESRF SC3 and other SPINE sample changers please use the CrystalCap HT.

Order Information

Cat. No.	Name	Description	Price
HR8-173	**CrystalCap Copper Magnetic HT**	0.025 - 0.05 mm CryoLoop - 30 pack	**$195.00**
HR8-174	**CrystalCap Copper Magnetic HT**	0.05 - 0.1 mm CryoLoop - 30 pack	**$195.00**
HR8-175	**CrystalCap Copper Magnetic HT**	0.1 - 0.2 mm CryoLoop - 30 pack	**$195.00**
HR8-176	**CrystalCap Copper Magnetic HT**	0.2 - 0.3 mm CryoLoop - 30 pack	**$195.00**
HR8-177	**CrystalCap Copper Magnetic HT**	0.3 - 0.4 mm CryoLoop - 30 pack	**$195.00**
HR8-178	**CrystalCap Copper Magnetic HT**	0.4 - 0.5 mm CryoLoop - 30 pack	**$195.00**
HR8-179	**CrystalCap Copper Magnetic HT**	0.5 - 0.7 mm CryoLoop - 30 pack	**$195.00**
HR8-180	**CrystalCap Copper Magnetic HT**	0.7 - 1.0 mm CryoLoop - 30 pack	**$195.00**
HR4-904	**CrystalCap Magnetic Vial**	Vial ONLY - 30 pack	**$84.00**
HR8-181	**CrystalCap Copper Magnetic HT**	Cap ONLY - 30 pack	**$130.00**

CRYSTALCAP™ COPPER MAGNETIC ALS HT

APPLICATION

- Cryocrystallography

FEATURES

- Specially designed vial with magnetic ring
- Cap magnetically attaches to vial
- Copper jacket to reduce icing on MicroTube
- Vented vial sold separately
- No threads
- Bar coded, color coded, alphanumeric, magnetic cap
- Assembled with Mounted CryoLoops, ready-to-use; vials sold separately
- Compatible with ALS style sample handlers, mounters
- Cap has ledge free, conical shape used by grippers in ALS - style automounters. For use at LBNL Berkeley Center for Structural Biology, with the PXRR automounter at Brookhaven NSLS, and at CHESS

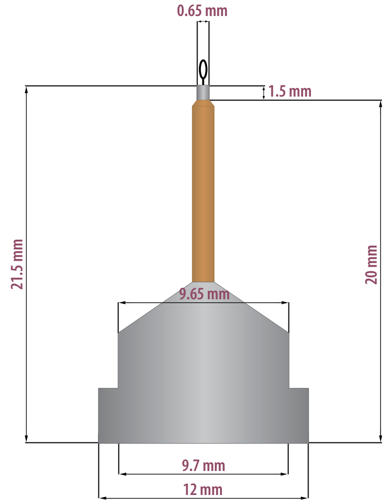

CrystalCap Copper Magnetic ALS HT

DESCRIPTION

CrystalCap Copper Magnetic ALS HT is designed to prevent ice formation along the copper pin/MicroTube™ in a cryostream where the flow is positioned perpendicular, at an angle, or non-collinear to the MicroTube. The CrystalCap Copper Magnetic ALS HT is a non-threaded, 1.8 ml (approximate) cryo vial storage container with a specially engineered ALS compatible cap/pin for cryocrystallography. Each ALS style cap is manufactured from an alloy base which allows the cap to be magnetically secured to a magnetic base in the goniometer head. Into the alloy cap is threaded and bonded a solid, 3 mm diameter copper pin. The end of the copper pin is machined to an aerodynamic 60° and is assembled with a MicroTube and CryoLoop or is available without a MicroTube and CryoLoop (HR8-182). A ring magnet is molded into the top end of the vial so that when the cap is positioned in the vial, the ring magnet holds the cap on the vial during cryogenic storage. The vial has two holes for venting. The 18 mm designates the distance between the top of the copper pin and the bottom of the CrystalCap Copper Magnetic where contact is made with the magnetic base.

For crystal transfer under cryo temperature use the 18 mm CryoTong (catalog number HR4-637 for the 110 mm overall tool length or catalog number HR5-112 for the 180 mm overall tool length).

The cap features a two dimensional bar coded and alphanumeric 16 x 16 data matrix. Each cap is also color coded.

HR8-182 is a CrystalCap Copper Magnetic ALS HT cap only. No MicroTube, no CryoLoop, no vial.

Note: Caps with Mounted CryoLoops are sold without vials. Vials available separately.

Color Coded Cap	CryoLoop Size
Red	0.025 - 0.05 mm
Green	0.05 - 0.1 mm
Yellow	0.1 - 0.2 mm
Blue	0.2 - 0.3 mm
Blue/Red	0.3 - 0.4 mm
Green/Red	0.4 - 0.5 mm
Yellow/Red	0.5 - 0.7 mm
Yellow/Green	0.7 - 1.0 mm

The ALS style automated sample changer accepts the Hampton Research CrystalCap Copper Magnetic ALS HT.

The CrystalCap Copper Magnetic ALS HT is not compatible with the EMBL/ESRF SC3 sample changer. For EMBL/ESRF SC3 and other SPINE sample changers please use the CrystalCap HT.

Order Information

Cat. No.	Name	Description	Price
HR8-184	**CrystalCap Copper Magnetic ALS HT**	0.025 - 0.05 mm CryoLoop - 30 pack	**$195.00**
HR8-186	**CrystalCap Copper Magnetic ALS HT**	0.05 - 0.1 mm CryoLoop - 30 pack	**$195.00**
HR8-188	**CrystalCap Copper Magnetic ALS HT**	0.1 - 0.2 mm CryoLoop - 30 pack	**$195.00**
HR8-190	**CrystalCap Copper Magnetic ALS HT**	0.2 - 0.3 mm CryoLoop - 30 pack	**$195.00**
HR8-192	**CrystalCap Copper Magnetic ALS HT**	0.3 - 0.4 mm CryoLoop - 30 pack	**$195.00**
HR8-194	**CrystalCap Copper Magnetic ALS HT**	0.4 - 0.5 mm CryoLoop - 30 pack	**$195.00**
HR8-196	**CrystalCap Copper Magnetic ALS HT**	0.5 - 0.7 mm CryoLoop - 30 pack	**$195.00**
HR8-198	**CrystalCap Copper Magnetic ALS HT**	0.7 - 1.0 mm CryoLoop - 30 pack	**$195.00**
HR4-904	**CrystalCap Magnetic Vial**	Vial ONLY - 30 pack	**$84.00**
HR8-182	**CrystalCap Copper Magnetic ALS HT**	Cap ONLY - 30 pack	**$130.00**

CRYSTALCAP™ HT SYSTEMS

CRYSTALCAP™ HT (SPINE)

APPLICATION

- Cryocrystallography

FEATURES

- Specially designed vial with magnetic ring
- Cap magnetically attaches to vial
- Vented vial sold separately
- No threads
- Bar coded, color coded, alphanumeric, magnetic cap
- Assembled with Mounted CryoLoops, ready-to-use; vials sold separately
- Compatible with automated sample handlers
  - SPINE sample changer
  - ALS style sample mounters
  - Brookhaven National Lab PXRR automounter
  - ESRF automated sample changer

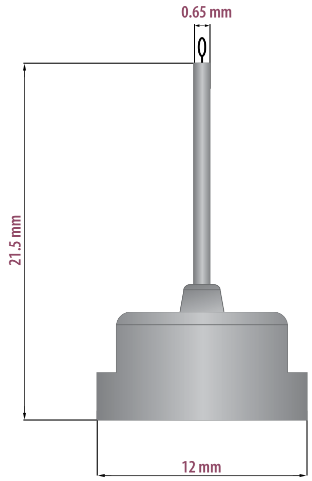

CrystalCap HT

DESCRIPTION

The CrystalCap HT is a complete crystal mount for manual and automated cryocrystallography. The cap supports the MicroTube and CryoLoop and is made from a corrosion resistant magnetic alloy. Chamfered edges on the cap avoid potential blocking during transfers. The cap design minimizes material to reduce the cooling/melting/drying temperature cycle when the sample holder is transferred. The cap features a two dimensional, bar coded and alphanumeric 16 x 16 data matrix. The alloy cap, alloy MicroTube, and synthetic CryoLoop feature an overall sample holder length of 22 mm (measured from the base of the cap to beam position). The magnetic vial, which is sold separately, is vented, features chamfered edges for enhanced cap positioning and a magnetic alloy bottom for stability. The CrystalCap HT is compatible with numerous commercial and academic automated sample handling systems. It is available as individual components (cap only, vial only, or cap and vial) or as an assembled crystal mount system (cap, MicroTube, and CryoLoop). The CryoLoop is a 20 micron diameter, synthetic material.

For crystal transfer under cryo temperature use the 18 mm CryoTong (catalog number HR4-637 for the 110 mm overall tool length or catalog number HR5-112 for the 180 mm overall tool length).

HR8-094 is the Crystal Cap HT cap only, no color code, no two dimensional bar code, no white background on bottom of cap, no alphanumeric side labeling, and is the cap only, without Mounted CryoLoop.

Note: Caps with Mounted CryoLoops are sold without vials. Vials available separately.

Color Coded Cap	CryoLoop Size
Red	0.025 - 0.05 mm
Green	0.05 - 0.1 mm
Yellow	0.1 - 0.2 mm
Blue	0.2 - 0.3 mm
Blue/Red	0.3 - 0.4 mm
Green/Red	0.4 - 0.5 mm
Yellow/Red	0.5 - 0.7 mm
Yellow/Green	0.7 - 1.0 mm

The CrystalCap HT is compatible with the SPINE sample changer, ALS style sample mounters, the Brookhaven National Lab PXRR automounter, and the ESRF automated sample changer.

Order Information

Cat. No.	Name	Description	Price
HR8-118	**CrystalCap HT**	0.05 mm CryoLoop - 30 pack	**$195.00**
HR8-120	**CrystalCap HT**	0.05 - 0.1 mm CryoLoop - 30 pack	**$195.00**
HR8-122	**CrystalCap HT**	0.1 - 0.2 mm CryoLoop - 30 pack	**$195.00**
HR8-124	**CrystalCap HT**	0.2 - 0.3 mm CryoLoop - 30 pack	**$195.00**
HR8-126	**CrystalCap HT**	0.3 - 0.4 mm CryoLoop - 30 pack	**$195.00**
HR8-128	**CrystalCap HT**	0.4 - 0.5 mm CryoLoop - 30 pack	**$195.00**
HR8-130	**CrystalCap HT**	0.5 - 0.7mm CryoLoop - 30 pack	**$195.00**
HR8-132	**CrystalCap HT**	0.7 - 1.0 mm CryoLoop - 30 pack	**$195.00**
HR4-637	**CryoTong for CrystalCap HT**	(same as CryoTong for 18 mm) - 1 each	**$55.00**
HR5-112	**CryoTong for CrystalCap HT**	(same as Long CryoTong for 18 mm) - 1 each	**$61.00**
HR8-112	**CrystalCap HT**	Cap ONLY - 60 pack	**$81.00**
HR8-114	**CrystalCap HT**	Vial ONLY - 30 pack	**$84.00**
HR8-116	**CrystalCap HT**	Cap/Vial ONLY - 30 pack	**$155.00**
HR8-094	**CrystalCap HT**	Cap ONLY, without bar coding - 30 pack	**$77.00**

CRYSTALCAP™ MAGNETIC

APPLICATION

- Cryocrystallography

FEATURES

- Magnetic cap
- Specially designed vial with magnetic ring
- Cap magnetically attaches to vial
- Vented vial
- No threads
- Vial without magnetic base

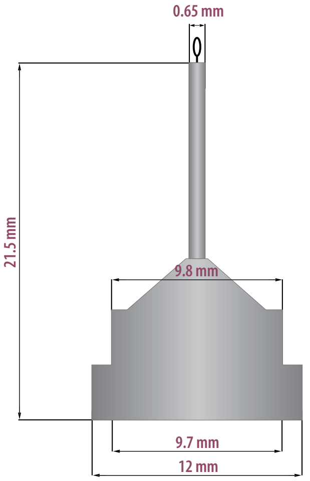

CrystalCap Magnetic

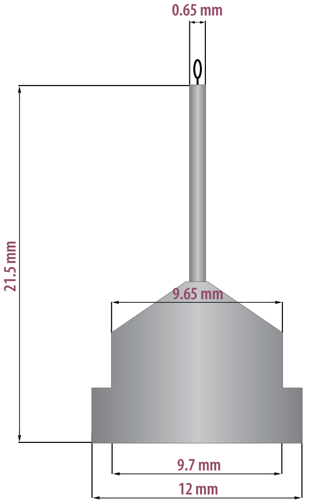

CrystalCap Magnetic ALS

DESCRIPTION

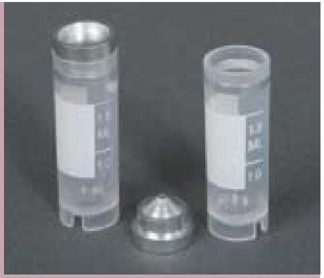

The CrystalCap Magnetic is a non-threaded, 1.8 ml (approximate) cryo vial storage container with a specially engineered cap for cryocrystallography. Each cap is manufactured from an alloy base which allows the cap to be magnetically secured to a magnetic base in the goniometer head. A ring magnet is molded into the top end of the vial so that when the cap is positioned in the vial, the ring magnet holds the cap on the vial during cryogenic storage. The tip of the metal cap is ready to accept a MicroTube™ and CryoLoop™ or Mounted CryoLoop™. CrystalCap Magnetic is available with or without a vial. The vial contains two small holes for venting.

CrystalCap Magnetic ALS (HR4-779) is compatible with ALS style sample changers.

In the comparison image above, the CrystalCap Magnetic (notice ledge at bottom of taper) appears on the left while the CrystalCap Magnetic ALS (no ledge at bottom of taper) appears on the right. Both cap styles can accept and use the same CrystalCap Magnetic Vial.

Brookhaven National Lab PXRR automounter is compatible with the CrystalCap Magnetic ALS.

The automounter at the Berkeley Center for Structural Biology at Lawrence Berkeley National Labs is compatible with the CrystalCap Magnetic ALS.

ALS style sample mounters accept the CrystalCap Magnetic ALS.

The ESRF automated sample changer accepts the Hampton Research CrystalCap Magnetic with an 18 mm mounted cryoloop.

Order Information

Cat. No.	Name	Description	Price
HR4-731	**CrystalCap Magnetic**	with Vial - 10 pack	**$60.00**
HR4-733	**CrystalCap Magnetic**	with Vial - 60 pack	**$300.00**
HR4-902	**CrystalCap Magnetic**	without Vial - 60 pack	**$145.00**
HR4-779	**CrystalCap Magnetic ALS**	without Vial - 60 pack	**$150.00**
HR4-904	**CrystalCap Magnetic Vial**	Vial ONLY - 30 pack	**$84.00**

Fab crystals growing on a thread.

Allen B. Edmundson,
Oklahoma Medical Research Foundation, USA.

CRYSTALCAP™ MAGNETIC SYSTEMS

CRYSTALCAP™ COPPER MAGNETIC

APPLICATION

- Cryocrystallography

FEATURES

- Magnetic cap
- Specially designed vial with magnetic ring
- Cap magnetically attaches to vial
- Copper jacket to reduce icing on MicroTube
- Vented vial
- No threads
- Vial without magnetic base

DESCRIPTION

CrystalCap Copper Magnetic is designed to prevent ice formation along the copper pin/MicroTube™ in a cryostream where the flow is positioned perpendicular, at an angle, or non-collinear to the MicroTube. The CrystalCap Copper Magnetic is a non-threaded, 1.8 ml (approximate) cryo vial storage container with a specially engineered cap/pin for cryocrystallography. Each cap is manufactured from an alloy base which allows the cap to be magnetically secured to a magnet base in the goniometer head. Into the alloy cap is threaded and bonded a solid, 3 mm diameter copper pin. The end of the copper pin is machined to an aerodynamic 60° and has a 0.65 mm opening designed to accept a 10 mm MicroTube and CryoLoop™ or Mounted CryoLoop™ snapped at the 10 mm mark to properly position the CryoLoop in the beam. A ring magnet is molded into the top end of the vial so that when the cap is positioned in the vial, the ring magnet holds the cap on the vial during cryogenic storage. The tip of the metal cap is ready to accept a MicroTube and CryoLoop or Mounted CryoLoop. The vial has two holes in the vial for venting. The sampler pack contains two of each copper pin length (10, 12, 14, 16, 18, 21, 24 mm). The length is the distance between the top of the copper pin and the bottom of the CrystalCap Copper Magnetic where contact is made with the magnetic base.

The ESRF automated sample changer accepts the Hampton Research CrystalCap Copper Magnetic 18 mm.

Order Information

Cat. No.	Name	Description	Price
HR4-737	**CrystalCap Copper Magnetic - 10 mm**	with Vial - 30 pack	**$190.00**
HR4-739	**CrystalCap Copper Magnetic - 12 mm**	with Vial - 30 pack	**$190.00**
HR4-741	**CrystalCap Copper Magnetic - 14 mm**	with Vial - 30 pack	**$190.00**
HR4-743	**CrystalCap Copper Magnetic - 16 mm**	with Vial - 30 pack	**$190.00**
HR4-745	**CrystalCap Copper Magnetic - 18 mm**	with Vial - 30 pack	**$190.00**
HR4-747	**CrystalCap Copper Magnetic - 21 mm**	with Vial - 30 pack	**$190.00**
HR4-749	**CrystalCap Copper Magnetic - 24 mm**	with Vial - 30 pack	**$190.00**
HR4-900	**CrystalCap Copper Magnetic - 18 mm**	without Vial - 30 pack	**$117.00**
HR4-904	**CrystalCap Magnetic Vial**	Vial ONLY - 30 pack	**$84.00**

CRYSTALCAP™ SYSTEMS

CRYSTALCAP™

APPLICATION

- Cryocrystallography

FEATURES

- Cap threads to vial
- Vented cap
- Magnetic base plate in cap for attachment to a magnetic base
- Cap accepts 0.65 mm diameter MicroTube or Mounted CryoLoop

DESCRIPTION

The CrystalCap is a threaded, 1.8 ml cryo vial storage container with a specially engineered cap for cryocrystallography. Each cap contains a magnetic base plate which allows the CrystalCap to be magnetically secured to a magnetic base in the goniometer head. The opposite end of the CrystalCap is ready to accept a MicroTube and CryoLoop or Mounted CryoLoop. The CrystalCap lid is vented to allow cryogen to pass in and out of the container. The CrystalCap is aerodynamically efficient to minimize ice buildup.

Order Information

Cat. No.	Name	Description	Price
HR4-911	**CrystalCap**	10 pack	**$55.00**
HR4-913	**CrystalCap**	60 pack	**$285.00**
HR4-914	**CrystalCap Ex, 0.5 - 0.7 mm**	30 pack	**$225.00**

CRYSTALCAP™ - COLORED

APPLICATION

- Cryocrystallography

FEATURES

- Cap threads to vial
- Caps are colored to make identification & organization easier
- Vented cap
- Magnetic base plate in cap for attachment to a magnetic base
- Cap accepts 0.65 mm diameter MicroTube and Mounted CryoLoop

DESCRIPTION

The CrystalCap is now available is six different colors: Red, Orange, Yellow, Green, Blue, and Pink. The vial container remains clear. Only the cap lid is colored. Same CrystalCap features are color coded to make identification, storage, and recovery faster, easier, and less confusing.

The CrystalCap is a threaded, 1.8 ml cryo vial storage container with a specially engineered cap for cryocrystallography. Each cap contains a magnetic base plate which allows the CrystalCap to be magnetically secured to a magnetic base in the goniometer head. The CrystalCap lid is vented to allow cryogen to pass in and out of the vial container.

Order Information

Cat. No.	Name	Description	Price
HR8-014	**CrystalCap Colored - Sampler**	60 pack (10 of each color)	**$300.00**
HR8-002	**CrystalCap Colored - Red**	60 pack	**$300.00**
HR8-004	**CrystalCap Colored - Orange**	60 pack	**$300.00**
HR8-006	**CrystalCap Colored - Yellow**	60 pack	**$300.00**
HR8-008	**CrystalCap Colored - Green**	60 pack	**$300.00**
HR8-010	**CrystalCap Colored - Blue**	60 pack	**$300.00**
HR8-012	**CrystalCap Colored - Pink**	60 pack	**$300.00**

CRYSTALCAP™ COPPER

APPLICATION

- Cryocrystallography

FEATURES

- Cap threads to vial
- Copper jacket to reduce icing on MicroTube
- Vented cap
- Zinc plated steel base plate for attachment to a magnetic base

DESCRIPTION

CrystalCap Copper is designed to prevent ice formation when using the MicroTube™ in a cryostream where the flow is positioned perpendicular, at an angle, or non-collinear to the MicroTube. The CrystalCap Copper uses a 1.8 ml cryo vial storage container and a cap which contains a sealed, zinc-plated metal core which allows the CrystalCap Copper to be secured to a magnetic platform on the goniometer head. Threaded into the metal base is a solid, 3 mm diameter copper pin. The end of the copper pin is machined to an aerodynamic 60° and has a 0.65 mm opening to accept ONLY a 10 mm MicroTube and CryoLoop™ or Mounted CryoLoop™ snapped at the 10 mm mark. The CrystalCap Copper lid is vented to allow cryogen to pass in and out of the container. The length (10, 12, 14, 16, 18, 21, or 24 mm) is the distance between the top of the copper pin and the bottom of the CrystalCap Copper where contact is made with the magnetic base.

Order Information

Cat. No.	Name	Description	Price
HR4-969	**CrystalCap Copper - 10 mm**	30 pack	**$180.00**
HR4-971	**CrystalCap Copper - 12 mm**	30 pack	**$180.00**
HR4-973	**CrystalCap Copper - 14 mm**	30 pack	**$180.00**
HR4-665	**CrystalCap Copper - 16 mm**	30 pack	**$180.00**
HR4-975	**CrystalCap Copper - 18 mm**	30 pack	**$180.00**
HR4-977	**CrystalCap Copper - 21 mm**	30 pack	**$180.00**
HR4-979	**CrystalCap Copper - 24 mm**	30 pack	**$180.00**

MOUNTED CRYOLOOP™ - 20 MICRON

APPLICATION

- Cryocrystallography

FEATURES

- Synthetic CryoLoop on stainless steel MicroTube
- MicroTube specially engineered with EasySnap notches
- Complete range of loop diameters from 0.025 - 1.0 mm

DESCRIPTION

Mounted CryoLoops with 20 micron diameter nylon. These nylon loops are pre-staked to hollow, stainless steel MicroTubes™ that are used to mount, freeze, and secure the crystal during cryocrystallographic procedures and x-ray data collection. The MicroTube is 24 mm in length and is specially engineered with EasySnap™ notches at the 10, 12, 14, 18, and 21 mm measures. To obtain the desired length of MicroTube, simply snap the MicroTube at the desired length and stake it to the CrystalCap Magnetic™ or CrystalCap Copper Magnetic™. These nylon loops show minimal diffraction, are thin for fast freezing, strong, and aerodynamic. Most data collection geometries prefer or require a 22 mm length between the base of the cap and the beam. For a 22 mm overall length using the CrystalCap, CrystalCap Magnetic and CrystalCap HT, snap the MicroTube at the second notch from the bottom. Snap the MicroTube at the fifth notch from the bottom (notch closest to the CryoLoop) and attach to an 18 mm CrystalCap Copper or CrystalCap Copper Magnetic to create a 22 mm length between the base of the cap and the beam.

Order Information

Cat. No.	Name	Description	Price
HR4-953	**Mounted CryoLoop - 20 micron**	Sampler - 30 pack (5 of each diameter size 0.1 - 1.0 mm)	**$78.00**
HR4-313	**Mounted CryoLoop - 20 micron**	0.025 - 0.05 mm - 25 pack	**$62.00**
HR4-625	**Mounted CryoLoop - 20 micron**	0.05 - 0.1 mm - 25 pack	**$62.00**
HR4-955	**Mounted CryoLoop - 20 micron**	0.1 - 0.2 mm - 25 pack	**$62.00**
HR4-957	**Mounted CryoLoop - 20 micron**	0.2 - 0.3 mm - 25 pack	**$62.00**
HR4-959	**Mounted CryoLoop - 20 micron**	0.3 - 0.4 mm - 25 pack	**$62.00**
HR4-961	**Mounted CryoLoop - 20 micron**	0.4 - 0.5 mm - 25 pack	**$62.00**
HR4-963	**Mounted CryoLoop - 20 micron**	0.5 - 0.7 mm - 25 pack	**$62.00**
HR4-965	**Mounted CryoLoop - 20 micron**	0.7 - 1.0 mm - 25 pack	**$62.00**

MOUNTED CRYOLOOP™ - 10 MICRON

APPLICATION

- Cryocrystallography

FEATURES

- Synthetic CryoLoop on stainless steel MicroTube
- MicroTube specially engineered with EasySnap notches
- Complete range of loop diameters from 0.05 - 0.5 mm

DESCRIPTION

Mounted CryoLoops with 10 micron diameter nylon. These nylon loops are pre-staked to hollow, stainless steel MicroTubes™ that are used to mount, freeze, and secure the crystal during cryocrystallographic procedures and x-ray data collection. The MicroTube is 24 mm in length and is specially engineered with EasySnap™ notches at the 10, 12, 14, 18, and 21 mm measures. To obtain the desired length of MicroTube, simply snap the MicroTube at the desired length and stake it to the CrystalCap Magnetic™ or CrystalCap Copper Magnetic™. These nylon loops show minimal diffraction, are thin for fast freezing, strong, and aerodynamic. The 10 micron Mounted CryoLoop shows less background diffraction than the 20 micron Mounted CryoLoop. Please note that the 10 micron Mounted CryoLoops are very flexible due to their small diameter. Most data collection geometries prefer or require a 22 mm length between the base of the cap and the beam. For a 22 mm overall length using the CrystalCap, CrystalCap Magnetic and CrystalCap HT, snap the MicroTube at the second notch from the bottom. Snap the MicroTube at the fifth notch from the bottom (notch closest to the CryoLoop) and attach to an 18 mm CrystalCap Copper or CrystalCap Copper Magnetic to create a 22mm length between the base of the cap and the beam.

Order Information

Cat. No.	Name	Description	Price
HR4-993	**Mounted CryoLoop - 10 micron**	Sampler - 25 pack (5 of each size)	**$67.00**
HR4-995	**Mounted CryoLoop - 10 micron**	0.05 - 0.1 mm - 25 pack	**$67.00**
HR4-997	**Mounted CryoLoop - 10 micron**	0.1 - 0.2 mm - 25 pack	**$67.00**
HR4-999	**Mounted CryoLoop - 10 micron**	0.2 - 0.3 mm - 25 pack	**$67.00**
HR4-615	**Mounted CryoLoop - 10 micron**	0.3 - 0.4 mm - 25 pack	**$67.00**
HR4-617	**Mounted CryoLoop - 10 micron**	0.4 - 0.5 mm - 25 pack	**$67.00**

ADJUSTABLE MOUNTED CRYOLOOP™

APPLICATION

- Cryocrystallography

FEATURES

- Adjustable loop orientation
- CryoLoops are pre-mounted
- Synthetic CryoLoop on adjustable stainless steel MicroTube
- MicroTube specially engineered with EasySnap notches
- Available in 20 micron diameter nylon
- Customize your CryoLoop orientation

DESCRIPTION

Similar to the Mounted CryoLoop™ in design and function, the Adjustable Mounted CryoLoop allows one to bend and orient the CryoLoop to multiple, unique positions previously unattainable with typical cryo setups. The Adjustable Mounted CryoLoop is a 20 micron diameter synthetic loop mounted inside a malleable stainless steel sleeve, which in turn, is mounted inside the EasySnap MicroTube™.

Using a specially designed L-shaped tool with a positioning notch, one can manipulate the angle and orientation of the CryoLoop by bending and adjusting the angle of the malleable stainless steel insert which holds the CryoLoop. The Adjustable Mounted CryoLoop may be positioned several times. Note: Repeated manipulation will fatigue the sleeve which can lead to failure of the sleeve. The Adjustable Mounted CryoLoop is designed to work with the CrystalCap System components.

The MicroTube is 24 mm in length and is specially engineered with EasySnap™ notches at the 10, 12, 14, 18, and 21 mm measures. To obtain the desired length of MicroTube, simply snap the MicroTube at the desired length and stake it to the CrystalCap Magnetic™ or CrystalCap Copper Magnetic™. These nylon loops show minimal diffraction, are thin for fast freezing, strong, and aerodynamic. Most data collection geometries prefer or require a 22 mm length between the base of the cap and the beam. For a 22 mm overall length using the CrystalCap, CrystalCap Magnetic and CrystalCap HT, snap the MicroTube at the second notch from the bottom. Snap the MicroTube at the fifth notch from the bottom (notch closest to the CryoLoop) and attach to an 18 mm CrystalCap Copper or CrystalCap Copper Magnetic to create a 22 mm length between the base of the cap and the beam.

Order Information

Cat. No.	Name	Description	Price
HR5-900	**Adjustable Mounted CryoLoop**	Sampler - 30 pack (5 of each diameter size 0.1 - 1.0 mm)	**$98.00**
HR4-336	**Adjustable Mounted CryoLoop**	0.025 - 0.05 mm - 25 pack	**$82.00**
HR4-338	**Adjustable Mounted CryoLoop**	0.05 - 0.1 mm - 25 pack	**$82.00**
HR8-102	**Adjustable Mounted CryoLoop**	0.1 - 0.2 mm - 25 pack	**$82.00**
HR8-072	**Adjustable Mounted CryoLoop**	0.2 - 0.3 mm - 25 pack	**$82.00**
HR8-104	**Adjustable Mounted CryoLoop**	0.3 - 0.4 mm - 25 pack	**$82.00**
HR8-106	**Adjustable Mounted CryoLoop**	0.4 - 0.5 mm - 25 pack	**$82.00**
HR8-108	**Adjustable Mounted CryoLoop**	0.5 - 0.7 mm - 25 pack	**$82.00**
HR8-110	**Adjustable Mounted CryoLoop**	0.7 - 1.0 mm - 25 pack	**$82.00**
HR5-902	**Tool for Adjustable Mounted CryoLoop**	each	**$35.00**

Flower shaped crystal of a hypothetical protein from S.Aureas.

Diana Benetteraj, Nickolay Chirgadze Lab, Clinical Genomics Centre, University Health Network Max Bell Research Centre, Toronto, Ontario, Canada.

MICROTUBE™

APPLICATION

- CryoLoop support for CrystalCap

FEATURES

- Laser-cut steel pin
- Use to mount CryoLoop to CrystalCap

DESCRIPTION

MicroTube
HR4-928

The MicroTube is the mount and support for the CryoLoop™ for crystallographers who prefer to mount their own CryoLoop. It fits into the pinhole opening on the CrystalCap™ or CrystalCap Copper™ (use 10 mm only) and is secured with epoxy or Super Glue. A CryoLoop is then secured into the opposite end of the MicroTube. The MicroTube is a laser-cut, 0.65 mm diameter stainless steel hollow tube. They are available in a variety of lengths to accommodate the wide range of geometries presented by various x-ray data collection systems. The 18 mm length MicroTube is the standard length used at most data collection facilities. Try a MicroTube FitKit pack to determine the best length for your application.

Determining MicroTube Length

Follow the instructions to determine which size works best for you. The following considerations have been taken in determining the length which is best suited for you:

Most XYZ goniometer heads have 5 mm of Z adjustment. The length of exposed loop stem on a Mounted CryoLoop is approximately 0.5 - 0.7 mm. When mounted properly, the length of exposed MicroTube set in a CrystalCap Copper is approximately 1.0 mm.

Recommendations:

Use 18 mm MicroTube with CrystalCap
Use 10 mm MicroTube with CrystalCap Copper 18 mm

Order Information

Cat. No.	Name	Description	Price
HR4-928	**MicroTube FitKit**	6 pack (1 of each size)	**$25.00**
HR4-915	**MicroTube - 10 mm**	60 pack	**$59.00**
HR4-917	**MicroTube - 12 mm**	60 pack	**$59.00**
HR4-919	**MicroTube - 14 mm**	60 pack	**$59.00**
HR4-921	**MicroTube - 18 mm**	60 pack	**$59.00**
HR4-923	**MicroTube - 21 mm**	60 pack	**$59.00**
HR4-925	**MicroTube - 24 mm**	60 pack	**$59.00**
HR4-318	**Epoxy Dual Syringe**	25 ml tube	**$10.00**
HR4-346	**Epoxy & Hardener**	35 g tube (2 pack)	**$9.00**
HR4-316	**Super Glue**	2 g tube	**$5.00**

CRYOLOOPS™ - 20 MICRON

APPLICATION

- Cryocrystallography

FEATURES

- Synthetic CryoLoop
- Complete range of loop diameters from 0.05 - 1.0 mm
- 20 micron diameter nylon

DESCRIPTION

CryoLoops in 20 micron diameter nylon. They are used to mount, freeze, and secure the crystal during cryocrystallographic procedures and x-ray data collection. These nylon CryoLoops show minimal diffraction, are thin for fast freezing, strong, and aerodynamic. Supplied as 10 loops per strip, 6 strips per package. Stake CryoLoops into MicroTubes™ using epoxy or Super Glue.

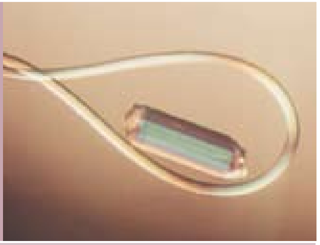

Crystal shown being collected using a Hampton Research CryoLoop.

Picture courtesy of Lisa Edberg. Center for Macromolecular Crystallography University of Alabama, Birmingham.

Order Information

Cat. No.	Name	Description	Price
HR4-941	**CryoLoop - 20 micron**	Sampler - 70 pack (10 of each size)	**$79.00**
HR4-623	**CryoLoop - 20 micron**	0.05 - 0.1 mm - 60 pack	**$69.00**
HR4-929	**CryoLoop - 20 micron**	0.1 - 0.2 mm - 60 pack	**$69.00**
HR4-931	**CryoLoop - 20 micron**	0.2 - 0.3 mm - 60 pack	**$69.00**
HR4-933	**CryoLoop - 20 micron**	0.3 - 0.4 mm - 60 pack	**$69.00**
HR4-935	**CryoLoop - 20 micron**	0.4 - 0.5 mm - 60 pack	**$69.00**
HR4-937	**CryoLoop - 20 micron**	0.5 - 0.7 mm - 60 pack	**$69.00**
HR4-939	**CryoLoop - 20 micron**	0.7 - 1.0 mm - 60 pack	**$69.00**

CRYOLOOPS™ - 10 MICRON

APPLICATION

- Cryocrystallography

FEATURES

- Synthetic CryoLoop
- Complete range of loop diameters from 0.05 - 0.5 mm
- 10 micron diameter nylon

DESCRIPTION

CryoLoops are available in 10 and 20 micron diameters. They are used to mount, freeze, and secure the crystal during cryocrystallographic procedures and x-ray data collection. These synthetic CryoLoops show minimal diffraction, are thin for fast freezing, strong, and aerodynamic. Supplied as 10 loops per strip, 6 strips per package. Stake CryoLoops into MicroTubes™ using Super Glue (HR4-316). Please note that the 10 micron CryoLoops are very flexible due to their small diameter.

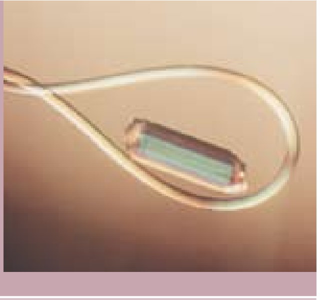

Order Information

Cat. No.	Name	Description	Price
HR4-981	**CryoLoop - 10 micron**	Sampler - 50 pack (10 of each size)	**$69.00**
HR4-983	**CryoLoop - 10 micron**	0.05 - 0.1 mm - 60 pack	**$79.00**
HR4-985	**CryoLoop - 10 micron**	0.1 - 0.2 mm - 60 pack	**$79.00**
HR4-987	**CryoLoop - 10 micron**	0.2 - 0.3 mm - 60 pack	**$79.00**
HR4-989	**CryoLoop - 10 micron**	0.3 - 0.4 mm - 60 pack	**$79.00**
HR4-991	**CryoLoop - 10 micron**	0.4 - 0.5 mm - 60 pack	**$79.00**

Crystal shown being collected using a Hampton Research CryoLoop.

Picture courtesy of Lisa Edberg. Center for Macromolecular Crystallography University of Alabama, Birmingham.

EPOXY

APPLICATIONS

- Sealing capillary tubes
- Securing MicroTubes into CrystalCaps
- Securing CryoLoops into MicroTubes

DESCRIPTION

5 minute, fast drying epoxy for sealing capillary tubes or securing MicroTubes into CrystalCaps or securing CryoLoops into MicroTubes. Sets in 5 minutes, can be handled in 15 minutes. Full bond strength in 1 hour. Requires no heat.

DEVCON
5 Minute Epoxy

Order Information

Cat. No.	Name	Description	Price
HR4-318	**Epoxy Dual Syringe**	25 ml tube	**$10.00**
HR4-346	**Epoxy & Hardener**	35 g tube (2 pack)	**$9.00**

SUPER GLUE

APPLICATION

- Staking CryoLoops to MicroTubes

DESCRIPTION

Super Glue for securing CryoLoops into MicroTubes.

DEVCON
Super Glue II

Order Information

Cat. No.	Name	Description	Price
HR4-316	**Super Glue**	2 g tube	**$5.00**

CRYOTONG™ - STANDARD

APPLICATION

- Crystal transfer under cryo temperature

FEATURES

- Artery clamp style closure
- 30 seconds of cryo temperature
- 7 different sizes
- 110 mm length

DESCRIPTION

The CryoTong is a tool used to manually transfer a crystal mounted on a CrystalCap™ from liquid nitrogen to a magnetic base in a goniometer head positioned in a cryogenic stream, and then back to liquid nitrogen. The one-piece, compact CryoTong is available in two lengths. The standard CryoTong is approximately 110 mm in length from the edge of the tool head to the end of the handle. The artery clamp style maintains the CryoTong in the closed position until the clamp is squeezed, which opens the opposing heads. The heads are non-magnetic stainless steel. The handle is magnetic stainless steel. The inside of the head is machined to closely surround the CrystalCap with loop and crystal in place. A small retaining lip is machined into the lower portion of the head to prevent the CrystalCap from slipping out when the tool is in the closed position. The CryoTong can maintain the temperature of the crystal at -160°C for up to 30 seconds during room temperature crystal transfers. The CryoTong is available in seven different sizes to fit 10 to 24 mm pin heights. Each size is designed to fit any of the the CrystalCap systems. Choose the proper CryoTong size based upon the pin length required by the configuration of the x-ray data collection hardware used. The 18 mm CryoTong (HR4-637) is to be used with the CrystalCap HT systems and any CrystalCap configured to 18 mm.

Order Information

Cat. No.	Name	Description	Price
HR4-631	**CryoTong - 10 mm**	each	**$55.00**
HR4-633	**CryoTong - 12 mm**	each	**$55.00**
HR4-635	**CryoTong - 14 mm**	each	**$55.00**
HR4-667	**CryoTong - 16 mm***	each	**$55.00**
HR4-637	**CryoTong - 18 mm**	each	**$55.00**
HR4-639	**CryoTong - 21 mm**	each	**$55.00**
HR4-641	**CryoTong - 24 mm**	each	**$55.00**

***The 16 mm CryoTong only works with the CrystalCap Copper 16 mm and the CrystalCap Copper Magnetic 16 mm.**

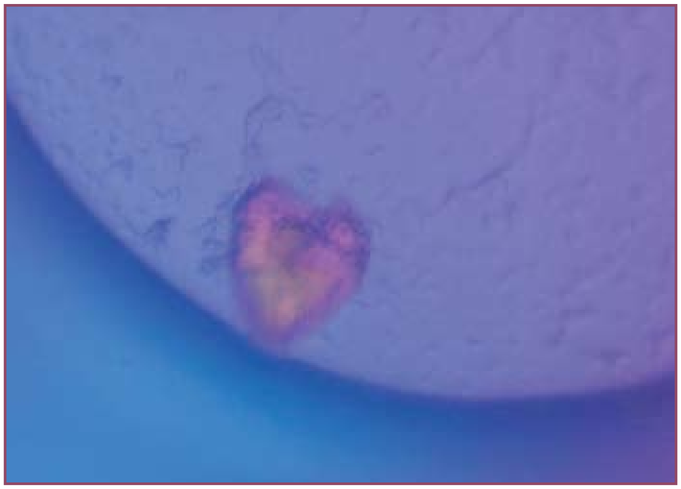

Valentine's Day crystal.

Igor Nederlof, Key Drug Prototyping BV, The Netherlands.

CRYOTONG™ - LONG

APPLICATION

- Crystal transfer under cryo temperature

FEATURES

- Artery clamp style closure
- 30 seconds of cryo temperature
- 7 different sizes
- 180 mm length

DESCRIPTION

The CryoTong is a tool used to manually transfer a crystal mounted on a CrystalCap™ from liquid nitrogen to a magnetic base in a goniometer head positioned in a cryogenic stream, and then back to liquid nitrogen. The one-piece, compact CryoTong is available in two lengths. The long CryoTong is approximately 180 mm in overall length. The artery clamp style maintains the CryoTong in the closed position until the clamp is squeezed, which opens the opposing heads. The heads are non-magnetic stainless steel. The handle is magnetic stainless steel. The inside of the head is machined to closely surround the CrystalCap with loop and crystal in place. A small retaining lip is machined into the lower portion of the head to prevent the CrystalCap from slipping out when the tool is in the closed position. The CryoTong can maintain the temperature of the crystal at -160°C for up to 30 seconds during room temperature crystal transfers. The CryoTong is available in seven different sizes to fit 10 to 24 mm pin heights. Each size is designed to fit all CrystalCap systems. Choose the proper CryoTong size based upon the pin length required by the configuration of the x-ray data collection hardware used. The 18 mm CryoTong (HR5-112) is to be used with the CrystalCap HT and any other CrystalCap configured for 18 mm.

HR5-114 Long CryoTong 18 mm, 180° features the head in alignment with the handle of the CryoTong.

HR5-113 Long CryoTong 18 mm, 45° features the head at a 45° angle to the handle of the CryoTong.

Order Information

Cat. No.	Name	Description	Price
HR5-104	**Long CryoTong - 10 mm**	each	**$61.00**
HR5-106	**Long CryoTong - 12 mm**	each	**$61.00**
HR5-108	**Long CryoTong - 14 mm**	each	**$61.00**
HR5-110	**Long CryoTong - 16 mm***	each	**$61.00**
HR5-112	**Long CryoTong - 18 mm**	each	**$61.00**
HR5-113	**Long CryoTong - 18 mm, 45°**	each	**$61.00**
HR5-114	**Long CryoTong - 18 mm, 180°**	each	**$61.00**
HR5-118	**Long CryoTong - 21 mm**	each	**$61.00**
HR5-120	**Long CryoTong - 24 mm**	each	**$61.00**

***The 16 mm Long CryoTong only works with the CrystalCap Copper 16 mm and the CrystalCap Copper Magnetic 16 mm.**

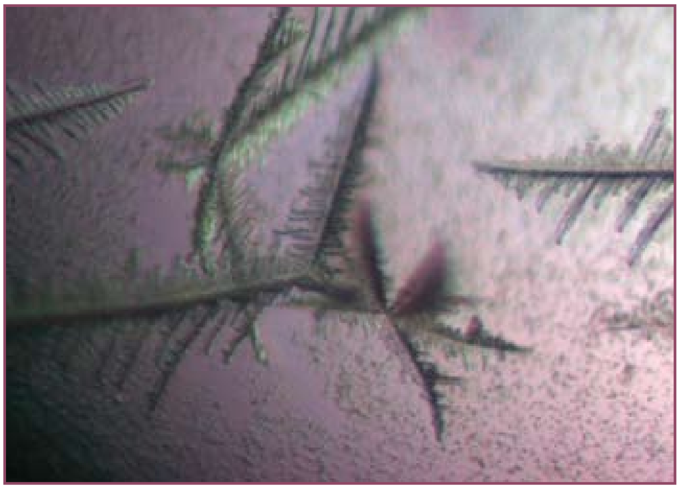

Christmas tree crystals.

Heidi Roth, Department of Structural Biology, University of Wuerzburg, Germany.

CRYSTALWAND™ - STANDARD

APPLICATION

- CrystalCap & CrystalCap Copper handling tool

DESCRIPTION

The CrystalWand is a chrome-plated steel wand, 9.6 mm in diameter with a magnet molded inside one end. The magnet is designed to fit the CrystalCap and CrystalCap Copper. The Crystal Wand makes it easy to manipulate, seed, mount, store, and handle crystals with the CrystalCap system. The CrystalWand is available in two styles (no tab and with tab) and two lengths (standard 130 mm and long 205 mm overall length). The long CrystalWand has a plastic handle.

Note: CrystalWand is not to be used with the CrystalCap Magnetic or CrystalCap Copper Magnetic or CrystalCap HT. Use the CrystalWand Magnetic for these products.

CrystalWand With Tab

CrystalWand No Tab

Order Information

Cat. No.	Name	Description	Price
HR4-951	**CrystalWand no Tab, 130 mm**	each	**$36.00**
HR4-619	**CrystalWand with Tab, 130 mm**	each	**$36.00**

CRYSTALWAND™ - LONG

APPLICATION

- CrystalCap and CrystalCap Copper handling tool

DESCRIPTION

The Long CrystalWand is a chrome-plated steel wand, 9.6 mm in diameter with a magnet molded inside one end. The magnet is designed to fit the CrystalCap and CrystalCap Copper. The Long Crystal Wand makes it easy to manipulate, seed, mount, store, and handle crystals with the CrystalCap System. It has a plastic handle and is available in two styles (no tab and with tab) and two lengths (standard 130 mm and long 205 mm overall length).

Note: Long CrystalWand is not to be used with the CrystalCap Magnetic or CrystalCap Copper Magnetic or CrystalCap HT. Use the CrystalWand Magnetic for these products.

Order Information

Cat. No.	Name	Description	Price
HR4-600	**Long CrystalWand no Tab, 205 mm**	each	**$46.00**
HR4-602	**Long CrystalWand with Tab, 205 mm**	each	**$46.00**

CRYSTALWAND™ MAGNETIC

APPLICATION

- CrystalCap Magnetic & CrystalCap Copper Magnetic handling tool

DESCRIPTION

The CrystalWand Magnetic is designed to be used exclusively with the CrystalCap Magnetic, CrystalCap Copper Magnetic, and CrystalCap HT systems during transfer of the caps from the vial to the goniometer head and from the goniometer to the vial. The 6 1/2” (165 mm) long chrome plated steel wand features a plastic housing enclosing a spring tensioned plunger that when depressed, moves a non-magnetic steel platform away from the magnet housed in the end of the wand. This causes the steel CrystalCap to detach from the wand and be placed readily into the vial or the CryoTong™. When the platform is retracted, the wand securely holds the steel CrystalCap Magnetic. The CrystalWand Magnetic 45° (McMiken Tool) offers the same functionality but is bent at a 45° angle at approximately 2.5 cm from the magnet end.

Order Information

Cat. No.	Name	Description	Price
HR4-729	**CrystalWand Magnetic, Straight**	each	**$61.00**
HR4-315	**CrystalWand Magnetic, 45° (McMiken Tool)**	each	**$56.00**

VIAL CLAMP™ - STRAIGHT

APPLICATION

- Vial support and manipulation

FEATURES

- Straight end tip
- Hemostat style closure

DESCRIPTION

The Vial Clamp - Straight is a chrome plated, hemostat style tool. It has a tip shaped to hold the storage vial of all the CrystalCap™ systems straight when the clamp is closed. The clamp can be locked using the hemostat style closure. This clamp makes it easy to dip the storage vial into liquid nitrogen for crystal storage. The complete length of the clamp is straight, with an overall length of 195 mm. When the vial is placed in the clamp, the length of it is positioned perpendicular, 90° to the length of the clamp.

Order Information

Cat. No.	Name	Description	Price
HR4-670	Vial Clamp - Straight	each	$49.00

VIAL CLAMP™ - CURVED

APPLICATION

- Vial support and manipulation

FEATURES

- Curved end tip
- Hemostat style closure

DESCRIPTION

The Vial Clamp - Curved is a chrome plated, hemostat style tool. It has a tip shaped to hold the storage vial of all the CrystalCap systems at an angle when the clamp is closed. The clamp can be locked using the hemostat style closure. The end of the clamp, where the vial is held, is curved at either a 45°/135° or 110°/ 70° angle. When the vial is placed in the clamp, the length of it is positioned to the clamp at either of those same angles. The overall length of the clamp is 195 mm.

Order Information

Cat. No.	Name	Description	Price
HR4-671	Vial Clamp - Curved 45°/135°	each	$49.00
HR4-672	Vial Clamp - Curved 110°/70°	each	$75.00

TUBE CLAMP™

APPLICATION

- 15 ml vial support & manipulation

FEATURES

- Curved end tip
- Hemostat style closure

DESCRIPTION

The 7 1/2” (190 mm) long, 15 ml Tube Clamp is useful for holding and manipulating 15 ml screw top centrifuge tubes during cryocrystallographic procedures involving liquid propane and liquid nitrogen. The jaws of the clamp are designed to fit the outside diameter (16 mm) of most 15 ml screw top centrifuge tubes. When the vial is placed in the clamp, the length of it is positioned to the clamp at either a 45° or 135° angle, depending upon how one holds the clamp.

Order Information

Cat. No.	Name	Description	Price
HR4-727	Tube Clamp	each	$61.00

CRYSTALCAP™ HOLDER

APPLICATION

- CrystalCap support inside stainless steel dewars

FEATURES

- Adjustable height
- Supports up to 8 CrystalCaps
- 15 ml vial position for liquid propane
- 500 and 1,000 ml versions

DESCRIPTION

CrystalCap Holders are convenient stands for holding all CrystalCap systems in liquid nitrogen dewars during crystal cryogenic procedures. Available in two sizes. Both holders feature spring-clips to secure the CrystalCap Vials in the holder. Both stands have 3" of adjustable Z (height) which allows one to raise or lower the CrystalCaps to the appropriate height, depending upon the liquid nitrogen level in the dewar. Each holder is supplied with a socket driver for height adjustment.

500 ml version holds five CrystalCaps, two 15 ml vials (for liquid propane procedures) and fits into the 500 ml dewar. 1,000 ml version holds eight CrystalCaps, two 15 ml vials (for liquid propane procedures) and fits into the 1,000 ml dewar.

The HR4-705 is approximately 98 mm wide (measured at corners of hexagon)/92 mm wide (measured at flats of hexagon) by approximately 120 mm tall.

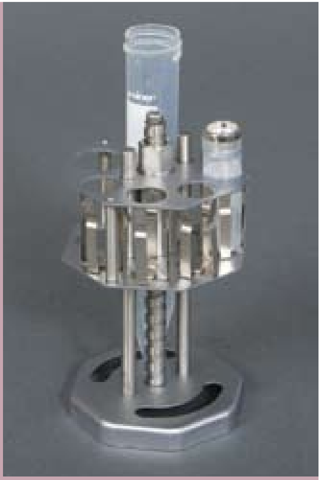

Order Information

Cat. No.	Name	Description	Price
HR4-707	CrystalCap Holder - 500 ml	each	$109.00
HR4-705	CrystalCap Holder - 1,000 ml	each	$109.00
HR4-706	Replacement Height Adjustment Tool	each	$15.00

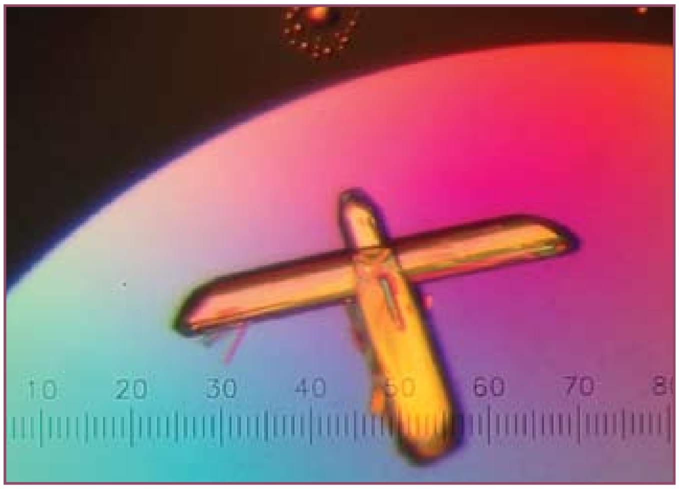

Protein crystals.

Ivana Tomcova, University of South Bohemia & Academy of Science of the Czech Republic, Nove Hrady, Czech Republic.

CRYOCANE

APPLICATION

- Organized storage of CrystalCaps

FEATURES

- 2 styles: 5 or 6 vial holding canes
- Made of aluminum
- Compatible with the CrystalCap systems

DESCRIPTION

Aluminum CryoCanes firmly hold CrystalCap™, CrystalCap Copper, CrystalCap Magnetic, and CrystalCap Copper Magnetic vials for storage in dewar-type, liquid nitrogen freezers. The 5 vial CryoCane offers base tabs for proper CrystalCap seating whereas the 6 vial CryoCane offers free alignment. Length of 5 vial version: 11 5/16" (28.7 cm). Length of 6 vial version: 11 13/16" (30 cm).

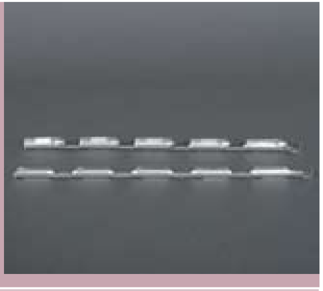

Order Information

Cat. No.	Name	Description	Price
HR4-709	**CryoCane 5 Vial Holder**	12 pack	**$24.00**
HR4-711	**CryoCane 6 Vial Holder**	12 pack	**$26.00**

CRYOCANE COLOR CODERS

APPLICATION

- Identify & organize CryoCanes

FEATURES

- Compatible with the CrystalCap systems
- Made of aluminum
- 5 colors to choose from

DESCRIPTION

Colored aluminum tabs slip into the ends of CryoCane holders to identify a set of vials, a researcher's work, or sample. Flat writing surface allows for additional identification. Available in five different colors.

Order Information

Cat. No.	Name	Description	Price
HR4-722	**CryoCane Color Coder - Sampler**	60 pack (12 of each color)	**$50.00**
HR4-713	**CryoCane Color Coder - White**	12 pack	**$10.00**
HR4-715	**CryoCane Color Coder - Yellow**	12 pack	**$10.00**
HR4-717	**CryoCane Color Coder - Blue**	12 pack	**$10.00**
HR4-719	**CryoCane Color Coder - Green**	12 pack	**$10.00**
HR4-721	**CryoCane Color Coder - Red**	12 pack	**$10.00**

CRYOSLEEVE

APPLICATION

- Use with CryoCanes

FEATURES

- Will not become brittle
- Made of polyvinyl
- Compatible with the CrystalCap systems

DESCRIPTION

Clear polyvinyl sleeve encloses CrystalCap on a CryoCane Holder for extra security during handling, transport, and storage. Will not become brittle during freezing or thawing. Unlike cardboard, will not shred, tear, or fall to pieces after repeated use. 10 13/16" length (27.5 cm).

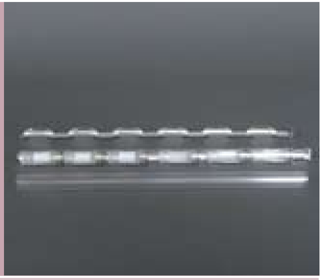

Order Information

Cat. No.	Name	Description	Price
HR4-708	**CryoSleeve**	12 pack	**$18.00**

DEWARS

STAINLESS STEEL DEWAR

APPLICATION

- Cryogenic procedures

FEATURES

- Made of 304 stainless steel
- Available in 500 ml and 1,000 ml sizes
- Compatible with CrystalCap Holders

DESCRIPTION

Unbreakable, 304 stainless steel construction dewars for cryocrystallography. Convenient for the flash freezing of crystals, these dewars store and transport low temperature liquid gases such as liquid nitrogen and can withstand temperatures from -196°C to 204°C. The dewars are impervious to vibrations. In comparison tests, these stainless steel dewars outperform common metal dewars by 20% and perform comparable to glass modules. Only the 1,000 ml dewar comes with a stainless steel carrying handle. Each dewar is supplied with a stainless steel cover. Covers are also available separately.

500 ml stainless steel dewar flask (2.5"ID, 3.4"OD, 7.0" depth, 8" height) - right in the picture. 1,000 ml stainless steel dewar flask (3.9"ID, 4.8" OD, 6.2" depth, 6.9" height) - left in the picture.

Order Information

Cat. No.	Name	Description	Price
HR4-695	**Dewar Flask - 500 ml with cover**	each	**$90.00**
HR4-697	**Dewar Cover - 500 ml**	each	**$20.00**
HR4-699	**Dewar Flask - 1,000 ml with cover**	each	**$120.00**
HR4-701	**Dewar Cover -1,000 ml**	each	**$20.00**

GLASS DEWAR

APPLICATION

- Cryogenic procedures

FEATURES

- Low form, shallow 600 ml base
- Low sidewall for easier access
- Manufactured for service of liquid nitrogen

DESCRIPTION

Low form, glass, shallow dewar for use with materials ranging from dry ice/acetone to liquid nitrogen. This 600 ml full base hemispherical bottom dewar is a low form with full aluminum protective coating. Low sidewall allows easier access to cryogen.

Dimensions: ID 140 mm, inside depth 65 mm, OD 165 mm, total height 95 mm. Accepts 500 ml flask. Manufactured by Pope Scientific.

For optimum results, temper the Dewar per instructions included with each dewar. Refer to The Care and Use of Pope Dewar Flasks.

Order Information

Cat. No.	Name	Description	Price
HR5-102	**Low Form Dewar**	each	**$268.00**

FOAM DEWAR

APPLICATION

- Cryocrystallography

FEATURES

- Each dewar is supplied with a lid
- Proprietary foam construction
- Standard (800 ml) and Tall (2 liter)

DESCRIPTION

Spearlab Cryogenic Foam Dewar (800 ml) - Standard: The standard foam dewar shape is circular, with a protruding handle, so that it resembles a large teardrop. Each dewar is supplied with a matching foam lid to insulate the contents from ambient air. Dimensions of the cylindrical cavity in this vessel are 5.8" in diameter by 2.8" deep, so that it easily holds 800 ml of liquid nitrogen.

Spearlab Cryogenic Foam Dewar (2 liter) - Tall: The tall foam dewar shape is a tapered octagon on the outside with a cylindrical interior. The tapered octagon features a wide, stable base. Each dewar is supplied with a matching foam lid to insulate the contents from ambient air. Dimensions of the cylindrical cavity in this vessel are 3.5" in diameter by 12.5" deep, so that it holds approximately 1,970 ml of liquid nitrogen. The outside dimensions of the tall dewar are 15.25" tall with a 6.25" wide top and a 7.75" wide base (measured on the flats of the octagon). The cover is 6.5" in diameter and 0.5" thick.

The patent pending design of the Spearlab Cryogenic Foam Dewar makes it easy to handle and safer to use than a traditional low profile glass dewar. Also, because of its lower thermal mass, a foam vessel will cause less boil off when it is filled with liquid nitrogen. Additionally, the dewar will accumulate less frost during regular use. The end result is that less liquid nitrogen is consumed.

HR4-673

HR4-675

Order Information

Cat. No.	Name	Description	Price
HR4-673	**Cryogenic Foam Dewar (800 ml) - Standard**	each	**$100.00**
HR4-675	**Cryogenic Foam Dewar (2 liter) - Tall**	each	**$160.00**

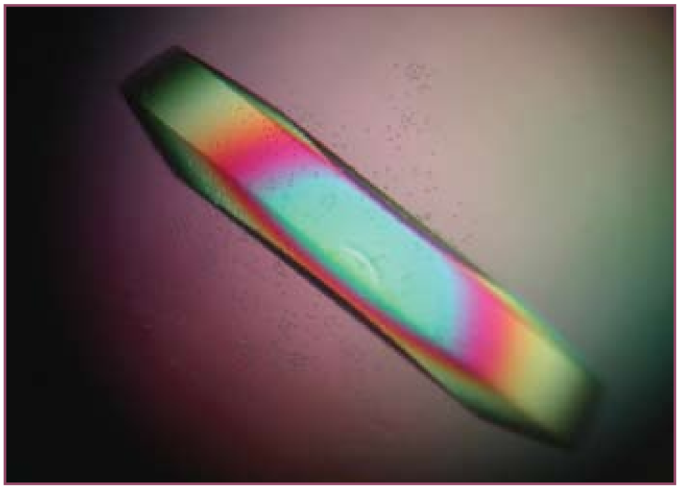

Crystal of a protein-DNA complex that grew over wells containing 5% PEG 8000.

Dr. Christopher Davies, Medical University of South Carolina, USA.

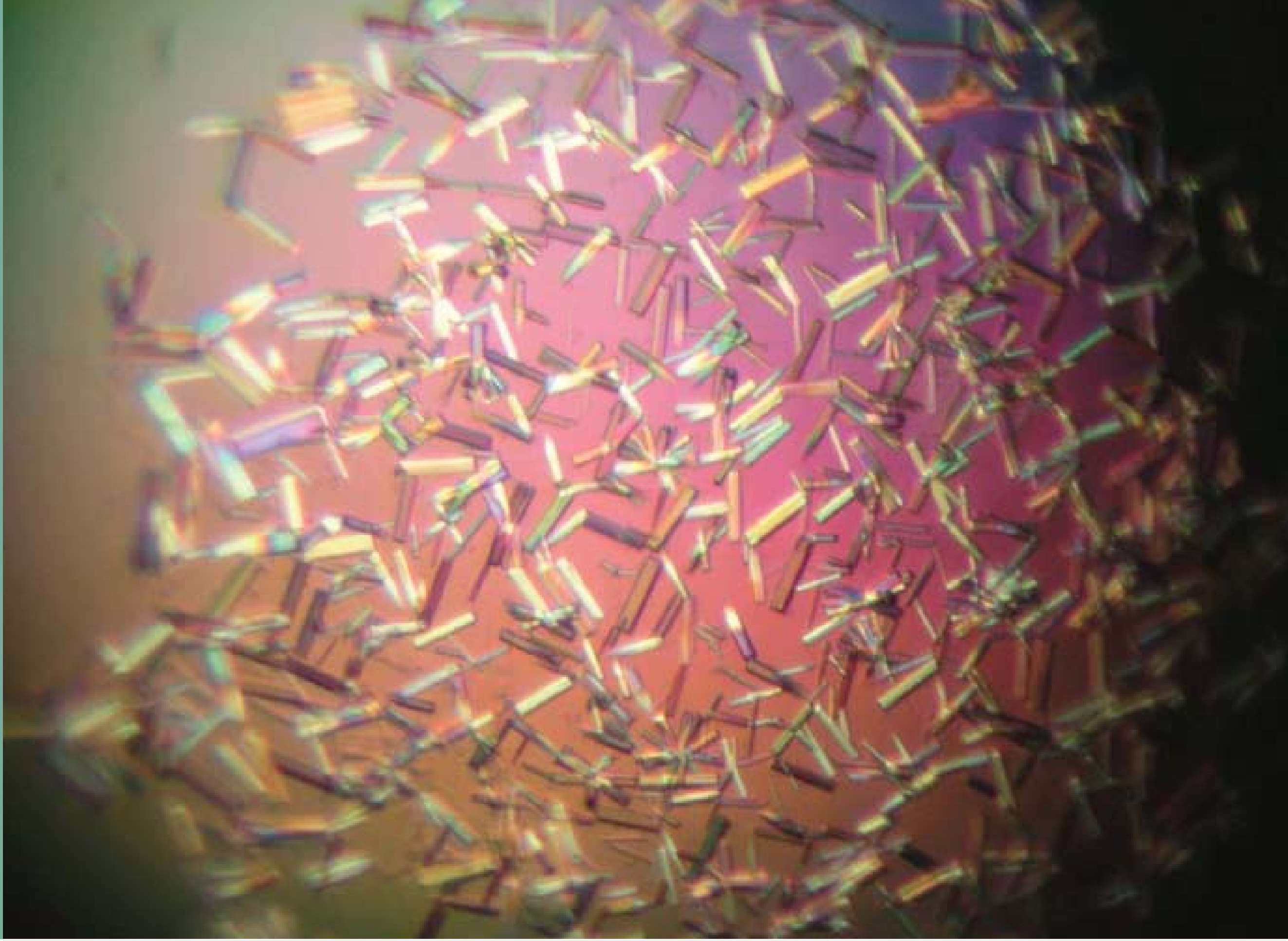

CAPILLARY MOUNTS & SUPPLIES

Crystal flakes - Crystals from a RNA duplex.

Vanessa Delfosse and Dr Claudine Mayer,
Laboratoire de Recherche Moléculaire sur les Antibiotiques, Paris, France.

TABLE OF CONTENTS

PAGES	
138 - 140	CAPILLARIES
140	GLASS FIBERS
141	WICKS
142	CAPILLARY SEALANTS
143 - 144	WAXES & CLAY
145	MAGNETIC MOUNTS & SUPPORTS

HAMPTON
RESEARCH

APPLICATIONS

- X-ray data collection
- Liquid-liquid diffusion crystallization
- Gel acupuncture crystallization

FEATURES

- 3 glass types
- Thin walled

DESCRIPTION

Crystal Clear Glass and Quartz capillaries that are extremely thin walled (approximately 10 microns wall thickness). The length of the capillary has a well defined diameter, with one end having a funnel shape and the other end closed. Capillaries have a wall thickness of 0.01 mm and an overall length of 80 mm. Capillaries are available in a wide range of outside diameters from 0.1 to 2.0 mm (5.0 mm for quartz only). The capillaries are designed to mount, hold, and store small molecule and biological macromolecular crystals for x-ray data collection. Capillaries can also be used for crystal density measurements and crystal growth experiments. The capillaries can be sealed tightly against moisture and gases using wax, epoxy, or other sealing materials.

In determining what glass or quartz capillary is right for you, please refer to the "Linear Absorption Coefficient μ cm^{-1}" table. This table indicates the amount of radiation that is absorbed by the capillary during x-ray data collection.

Linear Absorption Coefficient μ cm^{-1}

Glass Type	μ CuK L Radiation	μ MoK L Radiation	Softening Temp.
10	126.0	14.7	708°C
50	71.0	7.35	815°C
Quartz	75.8	8.2	1,730°C

HOW TO: MOUNTING CRYSTALS USING CAPILLARIES

Step 1 - Choosing a crystal

Locate the crystal of interest to be mounted in the capillary. Measure the size of crystal. This will determine what diameter capillary to use.

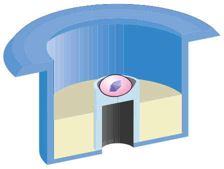

Step 2 - Cutting the capillary

Select the size of capillary. Score the capillary near the sealed end using the Capillary Cutting Stone. Snap the capillary at the score mark. A clean cut every time!

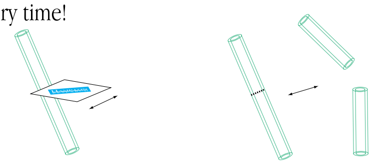

Step 3 - Make a slug

On the funnel end of the capillary attach a flexible adaptor to use a syringe or pipet. Suck up a small amount of mother liquor. This will be your mother liquor slug to prevent the crystal from drying out during data collection.

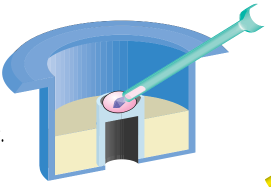

Step 4 - Obtaining the crystal

Remove the capillary from the drop and place an air gap into the capillary. This will separate the slug from the crystal.

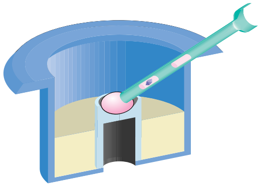

Place the capillary back into the drop and suck up the crystal. Try not to obtain too much mother liquor while picking up the crystal.

Step 5 - Removing excess mother liquor

Using either the Paper Wicks or the MicroWick, remove any excess mother liquor from around the crystal. There should only be a minimal amount of mother liquor surrounding the crystal. This will help reduce background scattering and reduce crystal slippage.

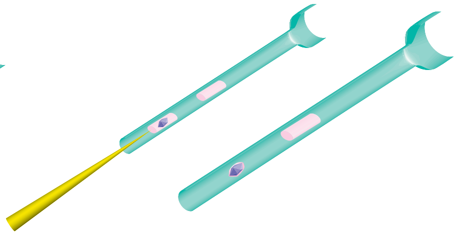

Step 6 - Sealing one end

Using either Beeswax or Capillary Wax, seal off the open end of the capillary. Do not overheat the wax for this will move or even damage the crystal.

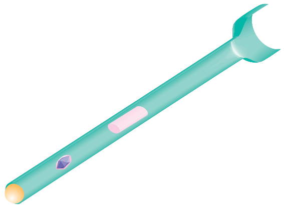

Step 7 - Sealing the other end

Trim off the funnel part of the capillary using the Capillary Cutting Stone as described in Step 2.

Finally, seal off the open end of the capillary as described in Step 6.

Using Mounting Clay or an Adjustable Crystal Mount, and a Brass Pin, place the capillary containing your crystal onto the goniometer head. You are now ready to collect data.

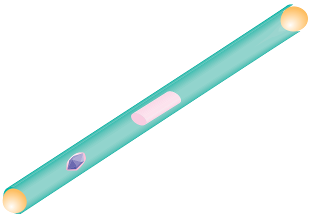

SPECIAL GLASS 10 (SODA LIME GLASS)

Order Information

Cat. No.	Name	Description	Price
HR6-152	Special Glass 10 Capillary	0.1 mm - 25 pack	$65.00
HR6-154	Special Glass 10 Capillary	0.2 mm - 25 pack	$65.00
HR6-156	Special Glass 10 Capillary	0.3 mm - 25 pack	$65.00
HR6-158	Special Glass 10 Capillary	0.4 mm - 25 pack	$65.00
HR6-160	Special Glass 10 Capillary	0.5 mm - 25 pack	$65.00
HR6-162	Special Glass 10 Capillary	0.6 mm - 25 pack	$65.00
HR6-164	Special Glass 10 Capillary	0.7 mm - 25 pack	$65.00
HR6-166	Special Glass 10 Capillary	0.8 mm - 25 pack	$65.00
HR6-168	Special Glass 10 Capillary	0.9 mm - 25 pack	$65.00
HR6-170	Special Glass 10 Capillary	1.0 mm - 25 pack	$65.00
HR6-172	Special Glass 10 Capillary	1.5 mm - 25 pack	$65.00
HR6-174	Special Glass 10 Capillary	2.0 mm - 25 pack	$85.00

0.8 mm
ial Glass 10 Capi

GLASS NUMBER 50 (BOROSILICATE)

Order Information

Cat. No.	Name	Description	Price
HR6-104	Glass Number 50 Capillary	0.1 mm - 25 pack	$78.00
HR6-106	Glass Number 50 Capillary	0.2 mm - 25 pack	$78.00
HR6-108	Glass Number 50 Capillary	0.3 mm - 25 pack	$78.00
HR6-110	Glass Number 50 Capillary	0.4 mm - 25 pack	$78.00
HR6-112	Glass Number 50 Capillary	0.5 mm - 25 pack	$78.00
HR6-114	Glass Number 50 Capillary	0.6 mm - 25 pack	$78.00
HR6-116	Glass Number 50 Capillary	0.7 mm - 25 pack	$78.00
HR6-118	Glass Number 50 Capillary	0.8 mm - 25 pack	$78.00
HR6-120	Glass Number 50 Capillary	0.9 mm - 25 pack	$78.00
HR6-122	Glass Number 50 Capillary	1.0 mm - 25 pack	$78.00
HR6-124	Glass Number 50 Capillary	1.5 mm - 25 pack	$88.00
HR6-126	Glass Number 50 Capillary	2.0 mm - 25 pack	$98.00

1.0 mm
Number 50 Capi

QUARTZ

Order Information

Cat. No.	Name	Description	Price
HR6-128	Quartz Capillary	0.1 mm - 25 pack	$118.00
HR6-130	Quartz Capillary	0.2 mm - 25 pack	$118.00
HR6-132	Quartz Capillary	0.3 mm - 25 pack	$118.00
HR6-134	Quartz Capillary	0.4 mm - 25 pack	$118.00
HR6-136	Quartz Capillary	0.5 mm - 25 pack	$118.00
HR6-138	Quartz Capillary	0.6 mm - 25 pack	$118.00
HR6-140	Quartz Capillary	0.7 mm - 25 pack	$118.00
HR6-142	Quartz Capillary	0.8 mm - 25 pack	$118.00
HR6-144	Quartz Capillary	0.9 mm - 25 pack	$118.00
HR6-146	Quartz Capillary	1.0 mm - 25 pack	$118.00
HR6-148	Quartz Capillary	1.5 mm - 25 pack	$118.00
HR6-150	Quartz Capillary	2.0 mm - 25 pack	$128.00
HR6-151	Quartz Capillary	2.5 mm - 15 pack	$128.00
HR6-175	Quartz Capillary	3.0 mm - 15 pack	$240.00
HR6-177	Quartz Capillary	4.0 mm - 5 pack	$240.00
HR6-179	Quartz Capillary	5.0 mm - 5 pack	$285.00

0.9 mm
Quartz Capillary

CAPILLARIES

CAPILLARY CUTTING STONE™

APPLICATION

- The easiest way to cut capillaries

FEATURES

- Score and snap

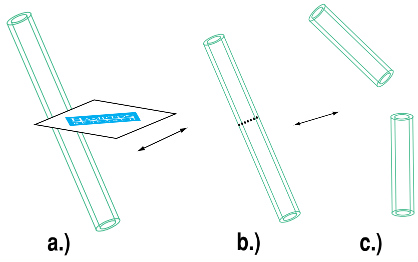

DESCRIPTION

These small (25 mm x 25 mm x 0.65 mm) cutting stones are used to make clean cuts of glass and quartz capillaries. Use the edge of the stone to etch the capillary then gently snap the capillary in two. This is useful for cutting the tips off closed capillaries used for mounting single crystals for x-ray diffraction analysis or for cutting capillaries to the desired length.

HAMPTON
RESEARCH

Order Information

Cat. No.	Name	Description	Price
HR4-334	**Capillary Cutting Stone**	each	**$15.00**

GLASS FIBERS

APPLICATION

- Crystal mounts for small molecules

FEATURES

- Two sizes: 0.1 - 0.3 mm & 0.3 - 0.5 mm
- Length: 40 mm

DESCRIPTION

40 mm length glass fibers for small molecule crystal mounts. Available in 0.1 - 0.3 mm and 0.3 - 0.5 mm diameter glass. Simple to use. Glue the glass fiber to a brass specimen pin. Adhesives for attaching glass fibers to the brass pins can be found in Capillary Essentials. Add a very small amount of the adhesive of choice to the tip of the glass fiber. Carefully bring the fiber into solid contact with the crystal face and leave in this orientation long enough for the glue to set firmly. Position onto goniometer head and collect data.

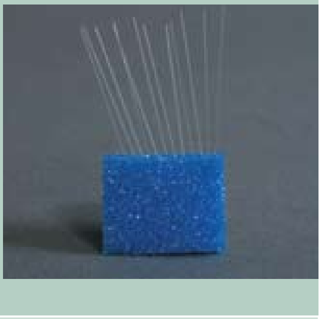

Order Information

Cat. No.	Name	Description	Price
HR8-030	**Glass Fiber**	0.1 - 0.3 mm diameter - 25 pack	**$15.00**
HR8-032	**Glass Fiber**	0.3 - 0.5 mm diameter - 25 pack	**$15.00**

PAPER WICKS™

APPLICATIONS

- Remove excess mother liquor from inside capillaries and CryoLoops
- Remove stray drops

DESCRIPTION

Sterile Paper Wicks are available in a variety of sizes. Smaller diameter wicks (in 25 mm & 55 mm lengths) are useful for wicking away mother liquor from within capillary tubes and CryoLoops while mounting crystals. Larger diameter wicks (in 25 mm & 55 mm lengths) are useful for removing stray/excess solvent from crystallization setups.

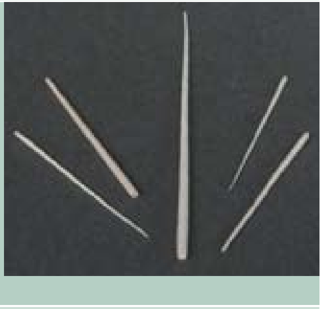

Order Information

Cat. No.	Name	Description	Price
HR4-110	**Assorted Paper Wicks - 25 mm**	XXX Fine - X Fine - 200 wicks	**$15.00**
HR4-122	**Paper Wicks - 25 mm**	XXX Fine - 200 wicks	**$15.00**
HR4-112	**Paper Wicks - 25 mm**	XX Fine - 200 wicks	**$15.00**
HR4-114	**Paper Wicks - 25 mm**	X Fine - 200 wicks	**$15.00**
HR4-116	**Paper Wicks - 25 mm**	Fine - 200 wicks	**$15.00**
HR4-211	**Paper Wicks - 55 mm**	X Fine Long - 100 wicks	**$15.00**
HR4-213	**Paper Wicks - 55 mm**	Medium Long - 100 wicks	**$15.00**

MICROWICK™

APPLICATION

- Remove excess mother liquor from inside capillaries

DESCRIPTION

MicroWick is useful for removing mother liquor from inside capillaries during crystal mounting. MicroWick is a non-metallic syringe needle used for filling and removing liquid from capillaries and micropipettes. MicroWick features a 70 mm tip, 0.164 mm OD, 0.1 mm ID (34 gauge) tip which will fit into most capillaries and micropipettes with an internal diameter of 0.2 mm or larger. The MicroWick needle is constructed from a combination of plastic and fused silica - no metal components are used. The MicroWick tip elasticity is sturdy and flexible though not unbreakable (much more flexible than glass!). Moderate bending will not block or damage the needle. The MicroWick luer fitting allows easy coupling to syringes and syringe filters. Each MicroWick contains one needle and one 1 ml syringe. Crystallization and capillary accessories sold separately.

Order Information

Cat. No.	Name	Description	Price
HR4-330	**MicroWick**	each	**$31.00**

CAPILLARY SEALANTS

EPOXY

APPLICATIONS

- Sealing capillaries
- Staking MicroTubes into CrystalCaps
- Mount crystals to glass fibers

DESCRIPTION

5 minute, fast drying epoxy for sealing capillary tubes or staking MicroTubes™ into CrystalCaps™. Sets in 5 minutes, can be handled in 15 minutes. Full bond strength in 1 hour. Requires no heat!

Order Information

Cat. No.	Name	Description	Price
HR4-318	**Epoxy Dual Syringe**	25 ml tube	**$10.00**
HR4-346	**Epoxy & Hardener**	35 g tube (2 pack)	**$9.00**

DUCO® CEMENT

APPLICATIONS

- Sealing capillaries
- Mount crystals to glass fibers

DESCRIPTION

Old school cement for sealing glass capillaries when mounting crystals. Requires no heat!

Order Information

Cat. No.	Name	Description	Price
HR4-320	**Duco Cement**	29 ml tube	**$7.00**

SUPER GLUE

APPLICATIONS

- Staking glass fibers to brass specimen pins
- Staking CryoLoops to MicroTubes

DESCRIPTION

Super glue for staking glass fibers to brass specimen pins, crystal mounting, and sticking fingers together.

Order Information

Cat. No.	Name	Description	Price
HR4-316	**Super Glue**	2 g tube	**$5.00**

BEESWAX™

APPLICATIONS

- Sealing capillaries
- Keep bees busy

DESCRIPTION

Pure, natural, and soft Beeswax™ - golden in color. Beeswax has a melting point of 63°C/146°F. Beeswax comes in seven easy to peel wrapped sticks making it very convenient to handle. Very useful for sealing the ends of capillary tubes when mounting crystals.

Order Information

Cat. No.	Name	Description	Price
HR4-312	Beeswax Stick	7 sticks	$30.00

CAPILLARY WAX™

APPLICATION

- Sealing capillaries

DESCRIPTION

A hard, brittle wax orange in color useful for sealing crystals in glass capillaries. Non-malleable. Melt with soldering iron or other heat device and apply to glass or quartz capillaries. When applied properly, Capillary Wax will provide an airtight seal. Also referred to as “sticky wax”.

Order Information

Cat. No.	Name	Description	Price
HR4-328	Capillary Wax	40 g pack (approximate)	$25.00

WAX PEN™

APPLICATION

- Sealing capillaries

DESCRIPTION

The Wax Pen™ is a hot pen, powered by two AA batteries (supplied). Pressing a switch heats the tips of the pen to a temperature that melts Beeswax and Capillary Wax which are used for sealing capillaries.

Replacement tips are available in a pack of three different shaped tips (straight loop, straight flat, and curved loop).

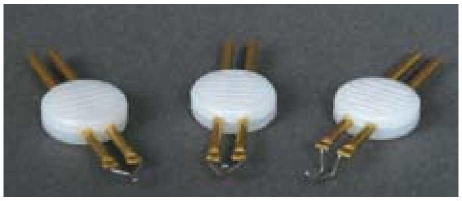

Order Information

Cat. No.	Name	Description	Price
HR4-342	Wax Pen	each	$50.00
HR4-344	Wax Pen Replacement Tips	3 tip pack	$21.00

RED STICKY WAX

APPLICATIONS

- Stabilizing the capillary on the goniometer head or brass specimen pin
- Prop open lids of Linbro® plates
- Make bendable buddies

DESCRIPTION

Soft, malleable, cylindrical wax ropes with similar properties as mounting clay. Uses include propping open the lids of Linbro® crystallization plates, supporting small items, and stabilizing capillaries.

Order Information

Cat. No.	Name	Description	Price
HR4-310	**Red Sticky Wax**	44 wax ropes	**$20.00**

FOUR COLOR MOUNTING CLAY

APPLICATION

- Stabilizing the capillary on the goniometer head or brass specimen pin

FEATURES

- Color code capillary mounts

DESCRIPTION

Adhesive, malleable, non-hardening clays have many uses in the crystallization lab. Useful for holding capillaries during crystal mounting as well as, stabilizing the capillary in the goniometer head. Useful for propping open the lids of crystallization plates. Great for color coding experiments, capillaries, and Linbro® crystallization plates. Each 4 oz brick of clay is individually wrapped.

Order Information

Cat. No.	Name	Description	Price
HR4-326	**Four Color Mounting Clay**	16 oz pack (4 oz each blue/yellow/red/green)	**$12.00**

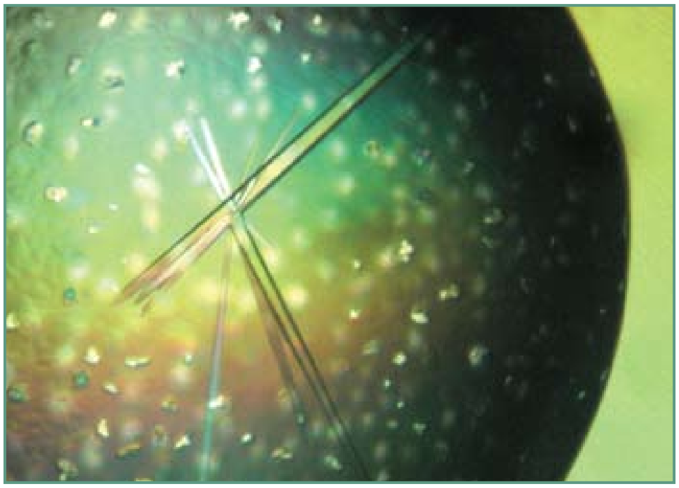

Vvn protein with DNA "diamond clock".

Lucy Doudeva, Hanna Yuan's lab, Academia Sinica, Taiwan.

ADJUSTABLE CRYSTAL MOUNT™

APPLICATION

- Adjustable, magnetic crystal mounting with capillaries or fibers

FEATURES

- Adjustable Z
- Magnetic mount
- Fits XYZ Goniometer Heads with Z Platform or magnetic base
- Fast, easy convenient crystal positioning and retrieval

DESCRIPTION

The Adjustable Crystal Mount is a magnetic piece (approximately 12 mm in height, 9.8 mm in diameter) designed to position a brass specimen pin (below) and either a capillary or glass fiber onto a goniometer head for x-ray data collection. The Adjustable Crystal Mount is engineered to fit onto the Z Platform or magnetic base within a goniometer head. The recessed bottom properly positions the Crystal Mount over the center of the goniometer head and prevents the mount from sliding across the goniometer head during data collection. The bottom of the Crystal Mount is flat and closed to prevent the brass specimen pin from protruding through the bottom of the Mount. The magnetic design allows for quick, convenient, and rapid placement and retrieval of the Crystal Mount. Finally, the Crystal Mount is designed with a single allen screw which allows the Z-translation of the brass specimen pin to be adjusted (up and down) and then secured in position by tightening the allen screw using the Goniometer Head Key.

Order Information

Cat. No.	Name	Description	Price
HR8-028	**Adjustable Crystal Mount**	5 pack	**$58.00**

BRASS SPECIMEN PIN™

APPLICATIONS

- Capillary and glass fiber mounts
- Small molecules and macromolecules

FEATURES

- Available with or without a platform

DESCRIPTION

3 mm brass specimen pins with platform for capillary mounting using Four Color Mounting Clay™ or Red Sticky Wax™. The platform has a diameter of 9 mm which adds for extra stability in maintaining a straight capillary for x-ray data collection. Conforms to IUCr standards.

Order Information

Cat. No.	Name	Description	Price
HR4-661	**Brass Specimen Pin**	25 pack	**$32.00**
HR4-663	**Brass Specimen Pin with Platform**	25 pack	**$46.00**

GONIOMETER HEADS & SUPPLIES

Crystal of nuclear receptor.

Virginie Chavant, Structural Biology and Genomics Platform, CEBGS / IGBMC, Illkirch France.

TABLE OF CONTENTS

PAGES

148 XYZ HEATED GONIOMETER HEADS
149 XYZ HEATED GONIOMETER HEAD SUPPLIES
150 XYZ STANDARD GONIOMETER HEADS
151 - 152 XYZ GONIOMETER HEAD SUPPLIES
153 MAGNETIC BASE SUPPORT™

HAMPTON RESEARCH

XYZ HEATED GONIOMETER HEADS

XYZ HEATED GONIOMETER HEAD™

FEATURES

- Heated goniometer head
- Prevent icing on goniometer head & pin
- Head height: Cryo: 34.5 mm
  Capillary: 38.0 mm
- Adjustments: X & Y: ± 2.5 mm
  Z: ± 5 mm

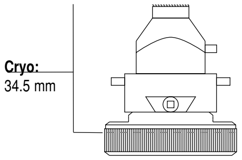

Goniometer Head with Cryo Mount

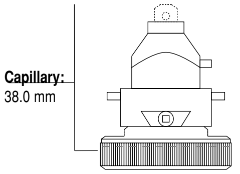

Goniometer Head with Capillary Mount

DESCRIPTION

34.5 mm height with Z Platform™ magnetic base for cryocrystallography. 38.0 mm height with Z Capillary™ for capillary mounts (measurements from the top of a lowered Z Platform/Z Capillary to the point where the base mount of the goniometer head attaches to the goniometer). ± 2.5 mm X and Y adjustments, 5 mm Z adjustment. Ceramic heating element in base of goniometer head maintains the head and mounting pin at a temperature that prevents icing. Temperature controller requires 120/230 volt AC. Head is equipped with an ACA/IUCr standard thread and support.

Each XYZ Heated Goniometer Head includes a heated goniometer head, Z capillary mount, Z Platform 9,000 gauss magnetic base, adjustment key, and storage case. Temperature controller, connection cable, and transformer (120v USA or 230v International) are required and sold separately. Crystallization and goniometer head accessories available separately.

Base thread is M30, 1 mm/thread.

Order Information

Cat. No.	Name	Description	Price
HR4-767	**XYZ Heated Goniometer Head**	each	**$685.00**

XYZ SHORT HEATED GONIOMETER HEAD™

FEATURES

- Heated goniometer head
- Prevent icing on goniometer head & pin
- Head height: Cryo: 27.5 mm
  Capillary: 31.0 mm
- Adjustments: X & Y: ± 2.5 mm
  Z: ± 5 mm

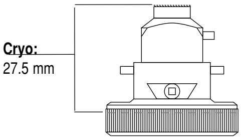

XYZ Short Heated Goniometer Head with Cryo Mount

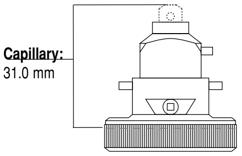

XYZ Short Heated Goniometer Head with Capillary Mount

DESCRIPTION

27.5 mm height with Z Platform™ magnetic base for cryocrystallography. 31.0 mm height with Z Capillary™ for capillary mounts (measurements from the top of a lowered Z Platform/Z Capillary to the point where the base mount of the goniometer head attaches to the goniometer). ± 2.5 mm X and Y adjustments, 5 mm Z adjustment. Ceramic heating element in base of goniometer head maintains the head and mounting pin at a temperature that prevents icing. Temperature controller requires 120/230 volt AC. Head is equipped with an ACA/IUCr standard thread and support.

Each XYZ Short Heated Goniometer Head includes a short heated goniometer head, Z capillary mount, Z Platform 9,000 gauss magnetic base, adjustment key, and storage case. Temperature controller, connection cable, and transformer (120v USA or 230v International) are required and sold separately. Crystallization and goniometer head accessories available separately.

Base thread is M30, 1 mm/thread.

Order Information

Cat. No.	Name	Description	Price
HR4-765	**XYZ Short Heated Goniometer Head**	each	**$685.00**

TEMPERATURE CONTROLLER

FEATURES

- Temperature controller for all XYZ heated heads

DESCRIPTION

Temperature Controller for the XYZ Heated™ and XYZ Short Heated™ goniometer heads. Includes connection cable.

Order Information

Cat. No.	Name	Description	Price
HR4-775	Temperature Controller	each	$98.00

TRANSFORMER

FEATURES

- Power supply for all XYZ heated heads

DESCRIPTION

Transformer (power supply) for the XYZ Heated and XYZ Short Heated goniometer heads. Available for both 120v and 230v applications.

Order Information

Cat. No.	Name	Description	Price
HR4-669	24V Transformer - 120V	each	$41.00
HR4-755	24V Transformer - 230V	each	$45.00

CONNECTION CABLE

FEATURES

- Connection cable for all XYZ heated heads

DESCRIPTION

Replacement connection cable for the XYZ Heated and XYZ Short Heated goniometer heads. Connects goniometer head to temperature controller.

Order Information

Cat. No.	Name	Description	Price
HR4-657	Connection Cable	each	$11.00

TEMPERATURE CONTROLLER FUSE

FEATURES

- Fuse for temperature controllers

DESCRIPTION

Replacement fuse for the temperature controller.

Order Information

Cat. No.	Name	Description	Price
HR4-777	Temperature Controller Fuse	3 pack	$8.00

XYZ STANDARD GONIOMETER HEADS

XYZ STANDARD GONIOMETER HEAD™

FEATURES

- Head height: Cryo: 34.5 mm
  Capillary: 38.0 mm
- Adjustments: X & Y: ± 2.5 mm
  Z: ± 5 mm

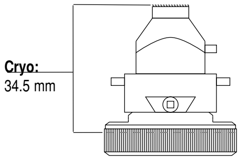

Goniometer Head with Cryo Mount

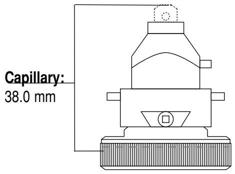

Goniometer Head with Capillary Mount

DESCRIPTION

34.5 mm height with Z Platform™ magnetic base for cryocrystallography. 38.0 mm height with Z Capillary™ for capillary mounts (measurements from the top of a lowered Z Platform/Z Capillary to the point where the base mount of the goniometer head attaches to the goniometer). ± 2.5 mm X and Y adjustments, ±5 mm Z adjustment. Includes XYZ Goniometer Head, Z Capillary Mount, Z Platform 9,000 gauss magnetic base, adjustment key, and storage case. Head is equipped with an ACA/IUCr standard thread and support.

Base thread is M30, 1 mm/thread.

Order Information

Cat. No.	Name	Description	Price
HR4-647	XYZ Standard Goniometer Head	each	$590.00

XYZ SHORT STANDARD GONIOMETER HEAD™

FEATURES

- Head height: Cryo: 27.5 mm
  Capillary: 31.0 mm
- Adjustments: X & Y: ± 2.5 mm
  Z: ± 5 mm

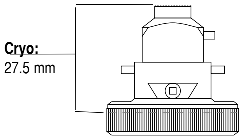

XYZ Short Goniometer Head with Cryo Mount

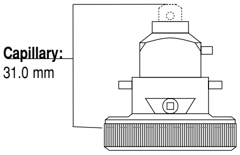

XYZ Short Goniometer Head with Capillary Mount

DESCRIPTION

27.5 mm height with Z Platform™ magnetic base for cryocrystallography. 31.0 mm height with Z Capillary™ for capillary mounts (measurements from the top of a lowered Z Platform/Z Capillary to the point where the base mount of the goniometer head attaches to the goniometer) . ± 2.5 mm X and Y adjustments, ±5 mm Z adjustment. Includes XYZ Short Goniometer Head, Z Capillary mount, Z Platform 9,000 gauss magnetic base, medium (6,000 gauss) magnetic base, adjustment key, and storage case. Head is equipped with an ACA/IUCr standard thread and support.

Base thread is M30, 1 mm/thread.

Order Information

Cat. No.	Name	Description	Price
HR4-643	XYZ Short Standard Goniometer Head	each	$590.00

STORAGE CASE FOR GONIOMETER HEAD

FEATURES

- Plastic storage case for IUCr standard thread goniometer heads

DESCRIPTION

Black plastic base threads securely into clear plastic cover and is threaded to hold IUCr standard thread goniometer heads.

Base thread is M30, 1 mm/thread.

Order Information

Cat. No.	Name	Description	Price
HR4-753	**Storage Case for Goniometer Head**	each	$20.00

GONIOMETER HEAD KEY

FEATURES

- Works with all XYZ heads

DESCRIPTION

Goniometer head key for use with all XYZ goniometer heads. This key is included with the purchase of any goniometer head.

Order Information

Cat. No.	Name	Description	Price
HR4-659	**Goniometer Head Key**	each	$11.00

SET SCREWS

FEATURES

- Replacement set screws for all XYZ heads

DESCRIPTION

Replacement set screws are available for the X and Y translation (2 mm x 4 mm) and for Z translation collar (2 mm x 1 mm).

Order Information

Cat. No.	Name	Description	Price
HR6-100	**X & Y Translation Set Screw**	3 pack	$9.00
HR6-102	**Z Translation Collar Set Screw**	3 pack	$9.00

Z CAPILLARY™ SUPPORT

FEATURES

- Capillary pin support for all XYZ heads
- Supports brass specimen pins

DESCRIPTION

Replacement Z Capillary Support for all XYZ goniometer heads.

Order Information

Cat. No.	Name	Description	Price
HR4-769	**Z Capillary**	For all XYZ short heads - each	**$37.00**
HR4-771	**Z Capillary**	For all XYZ heads - each	**$37.00**

ADJUSTABLE CRYSTAL MOUNT

APPLICATION

- Adjustable, magnetic crystal mounting with capillaries or fibers

FEATURES

- Adjustable Z
- Magnetic mount
- Fits XYZ Goniometer Heads with Z Platform or magnetic base
- Fast, easy, and convenient crystal positioning and retrieval

DESCRIPTION

The Adjustable Crystal Mount is a magnetic piece (approximately 12 mm in height, 9.8 mm in diameter) designed to position a brass specimen pin (below) and either a capillary or glass fiber onto a goniometer head for x-ray data collection. The Adjustable Crystal Mount is engineered to fit onto the Z Platform or magnetic base within a goniometer head. The recessed bottom properly positions the crystal mount over the center of the goniometer head and prevents the mount from sliding across the goniometer head during data collection. The bottom of the crystal mount is flat and closed to prevent the brass specimen pin from protruding through the bottom of the mount. The magnetic design allows for quick, convenient, and rapid placement and retrieval of the crystal mount. Finally, the crystal mount is designed with a single allen screw which allows the Z-translation of the brass specimen pin to be adjusted (up and down) and then secured in position by tightening the allen screw using the Goniometer Head Key.

Order Information

Cat. No.	Name	Description	Price
HR8-028	**Adjustable Crystal Mount**	5 pack	**$58.00**

BRASS SPECIMEN PIN™

APPLICATIONS

- Capillary and glass fiber mounts
- Small molecules and macromolecules
- Available with or without platform

DESCRIPTION

3 mm brass specimen pins with platform for capillary mounting using Four Color Mounting Clay™ or Red Sticky Wax™. The platform has a diameter of 9 mm which adds extra stability in maintaining a straight capillary for x-ray data collection. Conforms to IUCr standards.

Order Information

Cat. No.	Name	Description	Price
HR4-661	**Brass Specimen Pin**	25 pack	**$32.00**
HR4-663	**Brass Specimen Pin with Platform**	25 pack	**$46.00**

MAGNETIC BASE

FEATURES

- Available in 3 strengths
- Magnetic support for CrystalCap

DESCRIPTION

The CrystalCap Magnetic Base (9.3 mm diameter support) is a machined brass piece fitted with a magnet. The base fits into the goniometer head, providing a secure base for the CrystalCap Magnetic System™. Fits most goniometer heads. The CrystalCap Magnetic Base is available in three strengths: strong, medium, and light. The stem of the magnet is 3 mm in diameter. The length of the stem is 5 mm. The thickness of the platform is 3.6 mm.

Order Information

Cat. No.	Name	Description	Price
HR4-943	**Magnetic Base - Strong**	9,000 gauss - each	**$35.00**
HR4-627	**Magnetic Base - Medium**	6,000 gauss - each	**$35.00**
HR4-629	**Magnetic Base - Light**	4,000 gauss - each	**$35.00**

MAGNETIC BASE SUPPORT Z PLATFORM™ - LIGHT

FEATURES

- Light strength (6,000 gauss) magnetic support for all XYZ heads
- Magnetic support for CrystalCap

DESCRIPTION

Z Platform™ with 6,000 gauss magnet for all XYZ goniometer heads. Less pull and reduced "snapping" than Z Platform with 9,000 gauss magnet.

Order Information

Cat. No.	Name	Description	Price
HR4-651	**Z Platform - Light**	For all XYZ short heads - each	**$37.00**
HR4-653	**Z Platform - Light**	For all XYZ heads - each	**$37.00**

MAGNETIC BASE SUPPORT Z PLATFORM™ - STRONG

FEATURES

- Strong strength (9,000 gauss) magnetic support for all XYZ heads
- Magnetic support for CrystalCap

DESCRIPTION

Z Platform™ with 9,000 gauss magnet for all XYZ goniometer heads. Same Z Platform that is included with the purchase of any goniometer head.

Order Information

Cat. No.	Name	Description	Price
HR4-761	**Z Platform - Strong**	For all XYZ short heads - each	**$35.00**
HR4-763	**Z Platform - Strong**	For all XYZ heads - each	**$35.00**

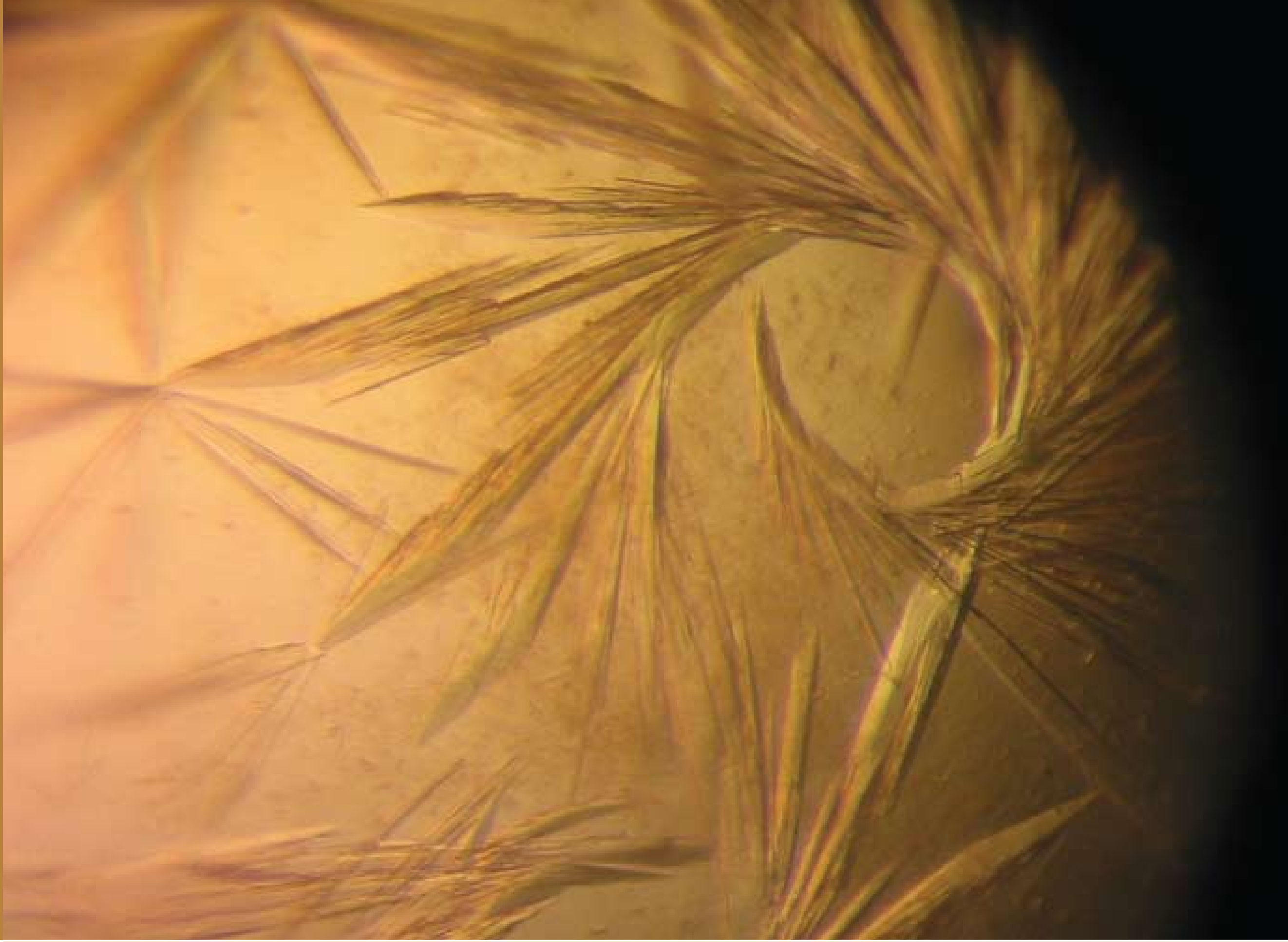

XENON DERIVATIZATION

Protein Crystals.

Anja Lehweß-Litzmann, Institut for Biochemie & Biotechnologie, Martin-Luther-Universität Halle-Wittenberg.

TABLE OF CONTENTS

PAGES

156 XENON CHAMBER™

157 XENON RECOVERY SYSTEM™

HAMPTON
RESEARCH

XENON CHAMBER™

APPLICATION

- Produce xenon derivatives of biological macromolecular crystals

FEATURES

- Operation pressure range of 0 to 600 psi
- Small pressure chamber for minimum xenon gas consumption
- Pressure regulated safety valve
- Safety lock prevents opening of pressurized chamber
- Safety shield between pressure chamber and user
- Quick release connectors between xenon gas supply and pressure chamber
- High quality xenon pressure regulator with needle valve for fine control
- Unique track design allows rapid crystal transfer from xenon to dewar for freezing
- Mini-Vial/Wick system prevents crystal dehydration
- CrystalCap compatible

DESCRIPTION

The Xenon Chamber is a pressure chamber designed to produce xenon derivatives of biological macromolecular crystals.

Aside from the preparation of a pure sample and subsequent crystallization, an often rate-limiting step in the determination of biological macromolecular structures by x-ray diffraction analysis is the formation of isomorphous heavy atom derivatives. Typically, isomorphous derivatives are formed by diffusing heavy atoms such as lead or mercury based compounds into the crystal with the anticipation that the heavy atom molecules will bind in an ordered fashion to each biological macromolecule. In order to obtain two or more successful isomorphous derivatives, a researcher may need to evaluate close to 100 different heavy atom compounds by manually transferring a number of crystals to different mother liquors each containing a different heavy atom compound.

Enter Xenon

Xenon is a noble gas which binds to specific sites in a biological macromolecule. There are numerous examples demonstrating that xenon-macromolecule complexes can serve as heavy atom derivatives.[1-8] The xenon-macromolecule complex is obtained by placing the macromolecular crystal under a pressurized xenon atmosphere. A clear advantage in using xenon is that one can simply screen xenon by pressurizing a loop-mounted crystal rather than having to set up plates and having to transfer crystals from numerous drops.

Enter the Xenon Chamber

The Xenon Chamber is a simple, yet effective device designed to pressurize loop-mounted biological macromolecular crystals in the presence of xenon gas at room temperature. Crystals mounted in loops such as the CrystalCap system are placed into the Xenon Chamber. Once sealed, the chamber is pressurized with xenon gas so that the crystal and macromolecules are equilibrated in a vapor saturated xenon atmosphere. Following depressurization of the chamber, the loop-mounted crystal is simply lifted and slid along the Xenon Chamber track and quickly lowered into a dewar for freezing in liquid nitrogen or propane. The transfer from pressurized xenon to the freeze is accomplished in seconds.

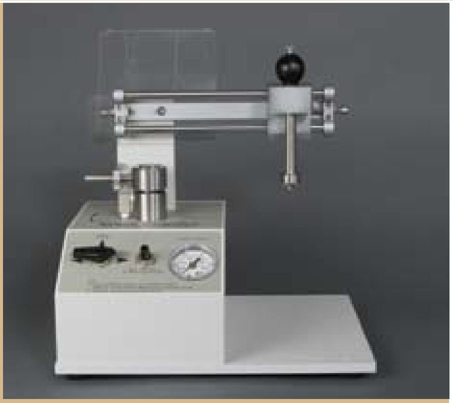

Specifications

The Xenon Chamber measures 14 1/4" tall, 13 1/2" wide, and 10 1/4" deep. The pressure chamber measures 9.5 mm (diameter) by 29 mm (depth).

The Xenon Chamber has an operation pressure range of 0 to 600 psi (0 to 4.1 Mpa, 0 to 41 Bar).

The pressure regulator for the Xenon Chamber is designed to be used with inert gases such as xenon. The CGA number for the pressure regulator is 580. The pressure regulator has a hose length of 21" with a quick connector on the end to attach it to the Xenon Chamber.

Each Xenon Chamber comes complete with a CrystalCap Dewar Stand that fits the 1,000 ml Dewar Flask (HR4-699) and a pack of five Mini-Vials with Wicks. The Xenon Chamber Pressure Regulator is sold separately. Additional Mini-Vials with Wick and Vial Stands may also be obtained separately.

References

1. *PNAS, Oct. 12 1999, Vol.96, No. 21, 11717-11722.*
2. *Soltis, S.M., Stowell, M.H.B., Wiener, M.C., Phillips, G.N. Jr., & Rees, D.C., J. Appl. Cryst. (1997) 30, 190-194.*
3. *Stowell, M.H.B., Soltis, S.M., Kisker, C., Peters, J.W., Schindelin, H., Rees, D.C., Cascio, D., Beamer, L., Hart, P.J., Wiener, M.C., & Whitby, F.G., J. Appl. Cryst. (1996) 29, 608-613.*
4. *Schiltz, M., Fourme, R., Broutin, I., & Prange, T., Structure (1995) 3, 309-316.*
5. *Schitz, M., Prange, T., & Fourme, R., J. Appl. Cryst. (1994) 27, 950-960.*
6. *Vitali, J., Robbins, A.H., Almo, S.C., & Tilton, R.F., J. Appl. Cryst. (1991) 24, 931-935.*
7. *Schiltz, M., Shepard, W., Fourme, R., Prange, T., De La Fortelle, R., & Bricogne, G., Acta Cryst. (1997) D53, 78-92.*
8. *Otwinowski, Z. & Minor, W., Methods Enzymol. Vol. 276, edited by C.W. Carter, Jr. and R.M. Sweet. New York: Academic Press (1997).*

Order Information

Cat. No.	Name	Description	Price
HR4-791	**Xenon Chamber**	each	**$2,400.00**
HR4-793	**Xenon Chamber Pressure Regulator**	each	**$560.00**
HR4-781	**Xenon Chamber Quick Female Connector**	each	**$50.00**
HR4-795	**Xenon Chamber Mini-Vial with Wick**	5 pack	**$75.00**
HR4-799	**Xenon Chamber Vial Stand**	each	**$55.00**

CASE STUDY: XENON DERIVATIVE & CRYSTAL STRUCTURE

A Xenon Derivative of Tt UvrB Crystal Prepared using the Hampton Research Xenon Chamber

DESCRIPTION

Thermus thermophilus UvrB (Tt UvrB), a 76 kDa protein is a component of the nucleotide excision repair system. The Tt UvrB crystal structures was determined by single isomorphous replacement / anomalous scattering from a xenon derivative. The UvrB crystals were pressurized with 500 psi of xenon gas for 15 minutes at room temperature and flash cooled for data collection. Xenon derivatized crystals diffracted to 1.9 angstrom.

Reference:

Machius, M., Henry, L., Palnitkar, M., and Deisenhofer, J. (1999) Proc. Natl. Acad. Sci USA, 96, 11717-11722.

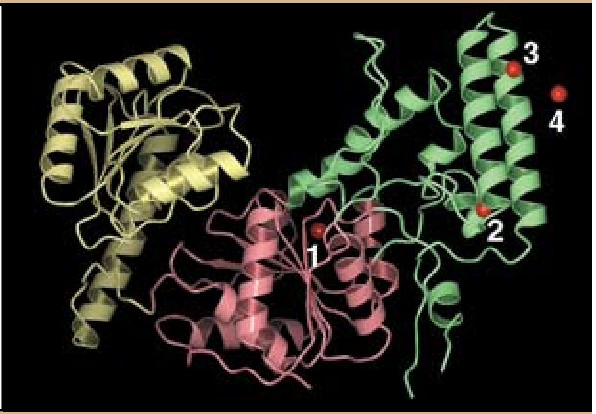

Overall structure of Tt UvrB with the Xenon sites (red spheres).

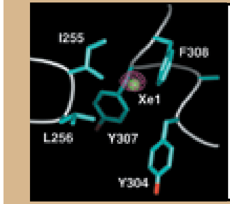

Environment of Xenon Site 1 with Fo-Fc density for the Xenon atom contoured at 10 σ.

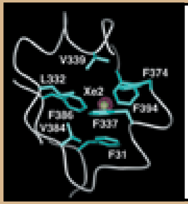

Environment of Xenon Site 2 with Fo-Fc density for the Xenon atom contoured at 10 σ.

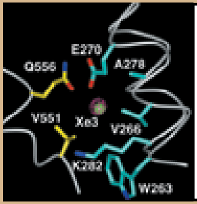

Environment of Xenon Site 3 with Fo-Fc density for the Xenon atom contoured at 10 σ.

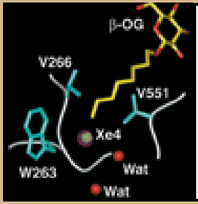

Environment of Xenon Site 4 with Fo-Fc density for the Xenon atom contoured at 10 σ.

XENON RECOVERY SYSTEM™

APPLICATION

- Scavenge and recover xenon gas from low pressure xenon gas cylinders

FEATURES

- Utilize up to 95% of xenon gas in cylinders

DESCRIPTION

Xenon Gas Pump

When using the Xenon Chamber to obtain derivatized crystals, the gas pressure is typically between 200 and 500 psi. A typical 50 liter compressed xenon gas cylinder has an initial pressure of 600 psi. Without some type of recovery system, if the cylinder pressure is less than desired, the cylinder cannot be used and the expensive xenon gas is wasted. With the Xenon Recovery System connected to a compressed xenon gas cylinder, pressures can be increased by a factor of 9. This means a typical compressed xenon gas cylinder is utilized at 95% efficiency versus 50% efficiency without the recovery system. The system easily connects to the Xenon Chamber and Pressure Regulator. Contact Technical Support for complete specifications.

Order Information

Cat. No.	Name	Description	Price
HR4-797	Xenon Recovery System	each	$1,200.00

LABELS & PENS

Proteins crystals.

Kasumi Kobayashi, Taiji Nakae and Hiroyuki Akama. The Kitasato Institute, Kanagawa, Japan.

TABLE OF CONTENTS

PAGES
160 PENS
160 - 161 TOUGH-TAGS® ON A ROLL
161 CRYO-BABIES® AND CRYO-TAGS® ON A ROLL
162 TOUGH-SPOTS® ON A ROLL
163 LASER CRYO-BABIES® AND CRYO-TAGS®
163 LASER TOUGH-SPOTS®

HAMPTON
RESEARCH

PENS

TEENY WRITER

APPLICATION

- Marking crystallization plates, labels, sample tubes, and vials

FEATURES

- Fine point writing tip
- Black ink

DESCRIPTION

The Teeny Writer has an extra fine tip. Recommended for writing on all our labels for samples that will be stored at room temperature and not exposed to chemicals and/or caustic agents. Ink dries in a matter of seconds. The ink is NOT resistant to detergents and reagents used in histology procedures. Color: Black.

Order Information

Cat. No.	Name	Description	Price
HR4-462	Teeny Writer	each	$6.00

SOLVENT RESISTANT PEN

APPLICATION

- Marking of labels and samples

FEATURES

- Fast drying, permanent ink
- Red ink or black ink

DESCRIPTION

The Solvent Resistant Pen is recommended for writing on Tough-Tags. The ultra-permanent ink dries in a matter of seconds and resists solvents such as xylene, alcohol, acetone, and formalin. Testing is advised before use.

Order Information

Cat. No.	Name	Description	Price
HR4-464	Solvent Resistant Pen - Black	each	$6.00
HR4-466	Solvent Resistant Pen - Red	each	$6.00

TOUGH-TAGS® ON A ROLL

TOUGH-TAGS® ON A ROLL - TEENY

APPLICATION

- Labels for 0.2 ml microtubes

FEATURES

- Pre-cut, peel off labels
- Withstands solvents, incubation in high humidity, and freezing to -40°C
- Works best with extra fine point markers (Teeny Writer)
- Convenient, easy to dispense roll

DESCRIPTION

Designed for use with 0.2 ml microtubes, Tough-Tags On a Roll - Teeny will withstand thermal-cycling, autoclaving, boiling water baths and freezing to -40°C.

Label Size: 0.81" x 0.28" (21 mm x 7 mm)
Color: White; Quantity: 1,500 labels per roll

Order Information

Cat. No.	Name	Description	Price
HR4-458	Tough-Tags® On a Roll - Teeny	1 roll	$35.00

TOUGH-TAGS® ON A ROLL

TOUGH-TAGS® ON A ROLL – SIDEWALL

APPLICATION

- Labels for crystallization plates

FEATURES

- Withstands solvents, incubation in high humidity, and freezing to -20°C
- Easy to peel off and affix
- Pre-cut labels fit perfectly on the side of crystallization plates

DESCRIPTION

Tough-Tags On a Roll - Sidewall have been carefully engineered to strongly adhere to all plastics and other materials. Sidewall Tough-Tags withstand organic solvents, caustic agents, humid incubators, boiling water baths, and freezer temperatures without peeling. These pre-cut labels fit perfectly onto the sides of crystallization plates allowing for visual identification of the individual plates in a stack.

Label Size: 1.50" x 0.25" (38 mm x 6 mm)
Color: White; Quantity: 1,000 labels per roll

Order Information

Cat. No.	Name	Description	Price
HR4-460	**Tough-Tags® On a Roll - Sidewall**	1 roll	**$35.00**

CRYO-BABIES® AND CRYO-TAGS® ON A ROLL

APPLICATION

- Labels for cryo labware cryogenic storage

FEATURES

- Freezable in liquid nitrogen
- Chemically resistant
- Made of a special smear-resistant material that will accept any marking instrument
- Pre-cut labels on a roll in dispenser box

DESCRIPTION

CRYO-TAGS
CRYO-BABIES
CRYO-BABIES

Labels for cryogenic storage. Withstand conventional and cryogenic storage (vapor and liquid phase nitrogen storage) to -196°C. The labels will also withstand boiling water baths to 100°C and dry heat up to 150°C. Polyolefin labels adhere to all plastics, glass, and metals. They resist detergents, oils, solvents, caustic agents, and other challenges without peeling or falling off. Great for labeling plastics and other materials for cryogenic storage. Cryo-Babies are ideal for1.5-2.0 ml or larger tubes. Apply label to tube or container BEFORE freezing.

Catalog #HR4-404 Label Size: 1.28" x 0.50" (33 mm x 13 mm)
Catalog #HR4-406 Label Size: 1.50" x 0.75" (38 mm x 19 mm)
Color: White; Quantity: 1,000 labels per roll

Order Information

Cat. No.	Name	Description	Price
HR4-404	**Cryo-Babies® On a Roll**	1 roll	**$36.00**
HR4-406	**Cryo-Tags® On a Roll**	1 roll	**$36.00**

TOUGH-SPOTS® ON A ROLL

TOUGH-SPOTS® ON A ROLL - SMALL

APPLICATION

- Circular labels for 0.5 - 2.0 ml tube tops

FEATURES

- For microtube cap or lid identification
- Freezable in liquid nitrogen
- Autoclavable and boilable
- Chemically resistant
- Won't dry out, fall off, or tear

DESCRIPTION

Tough-Spots On a Roll - Small are chemically inert, polyvinyl labels that adhere to all plastics. They are 3/8" diameter circular labels that are useful for tagging and easily identifying a wide range of small containers in the crystallization lab. Available in five different colors.

Label Size: 3/8" diameter (9.5 mm diameter)
Quantity: 1,000 labels per roll

Order Information

Cat. No.	Name	Description	Price
HR4-436	**Tough-Spots® On a Roll - Small**	White - 1 roll	**$31.00**
HR4-438	**Tough-Spots® On a Roll - Small**	Red - 1 roll	**$31.00**
HR4-440	**Tough-Spots® On a Roll - Small**	Yellow - 1 roll	**$31.00**
HR4-442	**Tough-Spots® On a Roll - Small**	Green - 1 roll	**$31.00**
HR4-444	**Tough-Spots® On a Roll - Small**	Blue - 1 roll	**$31.00**

TOUGH-SPOTS® ON A ROLL - LARGE

APPLICATION

- Labels for 1.5 - 2.0 ml tube tops

FEATURES

- For microtube cap or lid identification
- Freezable in liquid nitrogen
- Autoclavable and boilable
- Chemically resistant
- Won't dry out, fall off, or tear

DESCRIPTION

Tough-Spots On a Roll - Large are chemically inert, polyvinyl labels that adhere to all plastics. They are 1/2" diameter circular labels that are useful for tagging and easily identifying a wide range of small containers in the crystallization lab.

Label Size: 1/2" diameter (13 mm diameter)
Color: White; Quantity: 1,000 labels per roll

Order Information

Cat. No.	Name	Description	Price
HR4-446	**Tough-Spots® On a Roll - Large**	White - 1 roll	**$31.00**

LASER CRYO-BABIES® AND CRYO-TAGS®

LASER CRYO-BABIES®

APPLICATION

- Laser format labels for cryogenic storage

FEATURES

- Pre-cut, peel off labels sized to fit microcentrifuge tubes and other containers
- Temperature resistant from -196 to 100°C

DESCRIPTION

Labels for cryogenic storage. Withstand conventional and cryogenic storage (vapor and liquid phase nitrogen storage) to -196°C. The labels will also withstand boiling water baths to 100°C and dry heat up to 150°C. Polyolefin labels adhere to all plastics, glass, and metals. They resist detergents, oils, solvents, caustic agents, and other challenges without peeling or falling off. Ideal for 1.5 to 2.0 ml or larger tubes.

Label Size: 1.28 x 0.50" (33 mm x 13 mm)
Color: White; Quantity: 85 labels per sheet (1,700/pack)

Order Information

Cat. No.	Name	Description	Price
HR4-402	Laser Cryo-Babies®	1 pack	$70.00

LASER CRYO-TAGS®

APPLICATION

- Laser format labels for cryogenic storage

FEATURES

- Pre-cut, peel off labels sized to fit microcentrifuge tubes and other containers
- Temperature resistant from -196 to 100°C

DESCRIPTION

Labels for cryogenic storage. Withstand conventional and cryogenic storage (vapor and liquid phase nitrogen storage) to -196°C. The labels will also withstand boiling water baths to 100°C and dry heat up to 150°C. Polyolefin labels adhere to all plastics, glass, and metals. They resist detergents, oils, solvents, caustic agents, and other challenges without peeling or falling off. Ideal for labeling plastics and other materials for cryogenic storage.

Label Size: 1.50 x 0.75" (38 mm x 19 mm)
Color: White; Quantity: 60 labels per sheet (1,200/pack)

Order Information

Cat. No.	Name	Description	Price
HR4-400	Laser Cryo-Tags®	1 pack	$70.00

LASER TOUGH-SPOTS®

APPLICATION

- Laser format labels for 0.5 ml - 2.0 ml tube tops

FEATURES

- Sized to fit a variety of tubes and containers
- Pre-cut, peel off spots for easy identification
- Temperature resistant - can withstand 100°C water baths and dry heat up to 150°C
- Can withstand liquid and vapor phase nitrogen to -196°C

DESCRIPTION

Tough-Spots Small are chemically inert, polyvinyl labels that adhere to all plastics. They are 3/8" diameter circular labels that are useful for tagging and easily identifying a wide range of small containers in the crystallization lab.

Label Size: 3/8" diameter (9.5 mm diameter)
Color: White; Quantity: 192 labels per sheet (3,840/pack)

Order Information

Cat. No.	Name	Description	Price
HR4-456	Laser Tough-Spots®	1 pack	$70.00

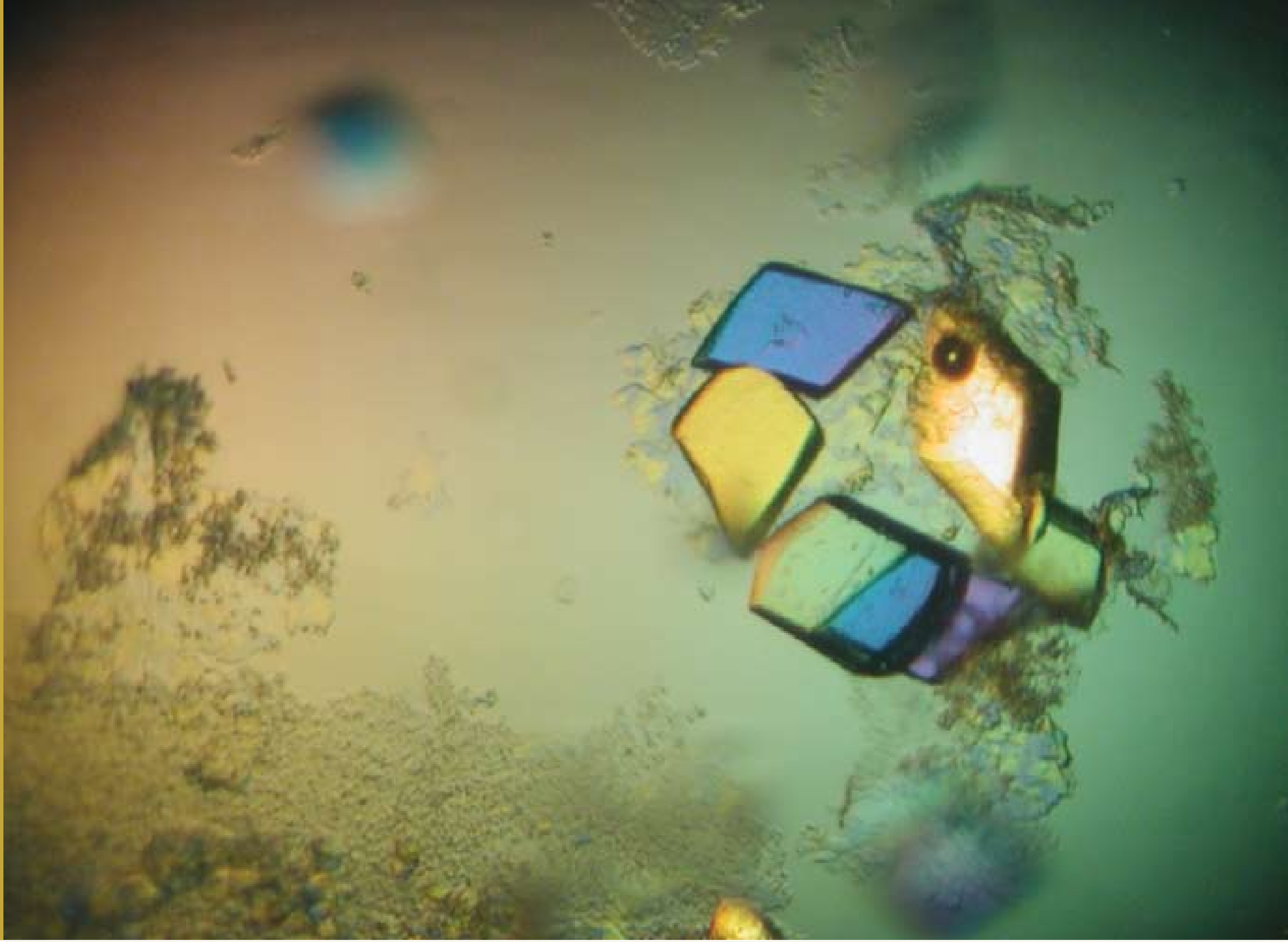

BOOKS

Proteins crystals.

Kasumi Kobayashi, Taiji Nakae and Hiroyuki Akama. The Kitasato Institute, Kanagawa, Japan.

TABLE OF CONTENTS

PAGES

166 MACROMOLECULAR CRYSTALLOGRAPHY
166 MACROMOLECULAR CRYSTALLOGRAPHY PROTOCOLS, VOLUME 1
167 PROTEIN CRYSTALLIZATION STRATEGIES FOR STRUCTURAL GENOMICS
167 PRACTICAL PROTEIN CRYSTALLOGRAPHY
167 PREPARATION AND ANALYSIS OF PROTEIN CRYSTALS
168 PROTEIN METHODS
168 METHODS IN ENZYMOLOGY - PART C
168 METHODS IN ENZYMOLOGY - PART D
169 CRYSTALLOGRAPHY MADE CRYSTAL CLEAR
169 OUTLINE OF CRYSTALLOGRAPHY FOR BIOLOGISTS
169 INTRODUCTION TO MACROMOLECULAR CRYSTALLOGRAPHY
170 PRINCIPLES OF PROTEIN X-RAY CRYSTALLOGRAPHY
170 MEMBRANE PROTEIN PURIFICATION & CRYSTALLIZATION, A PRACTICAL GUIDE
170 METHODS AND RESULTS IN CRYSTALLIZATION OF MEMBRANE PROTEINS
171 MACROMOLECULAR CRYSTALLIZATION METHODS, VOLUME 34, NUMBER 3
171 JOURNAL OF STRUCTURAL BIOLOGY
172 PRESENT AT THE FLOOD: HOW STRUCTURAL MOLECULAR BIOLOGY CAME ABOUT
172 CRYSTALS AND LIFE: A PERSONAL JOURNEY
172 PROTEIN CRYSTALLOGRAPHY: A CONCISE GUIDE
173 PROTEIN CRYSTALLIZATION
173 VIEWERS

HAMPTON
RESEARCH

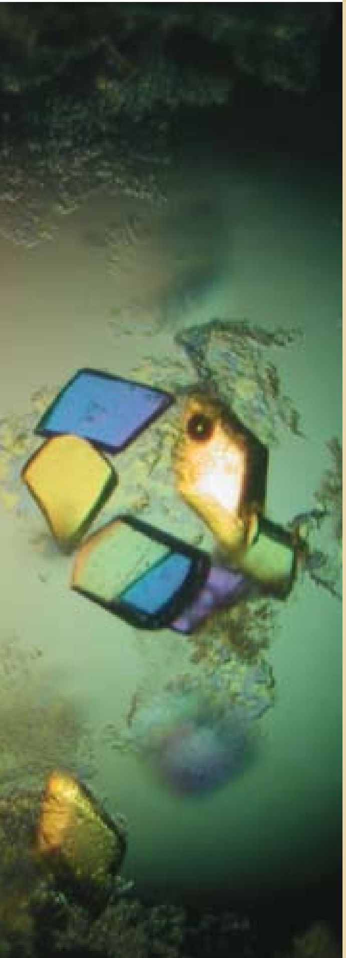

MACROMOLECULAR CRYSTALLOGRAPHY

- Authors: Mark Sanderson and Jane Skelly
- Publisher: Oxford University Press, 2007
- ISBN-10: 0198520972
- ISBN-13: 978-0198520979
- Hardcover: 400 pages

DESCRIPTION

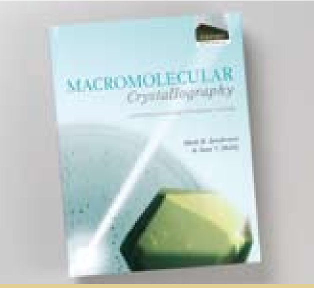

Macromolecular Crystallography - conventional and high-throughput methods

Macromolecular crystallography is the study of macromolecules (proteins and nucleic acids) using x-ray crystallographic techniques in order to determine their molecular structure. The knowledge of accurate molecular structures is a pre-requisite for rational drug design, and for structure-based function studies to aid the development of effective therapeutic agents and drugs. The successful determination of the complete genome (genetic sequence) of several species (including humans) has recently directed scientific attention towards identifying the structure and function of the complete complement of proteins that make up that species; a new and rapidly growing field of study called 'structural genomics'. There are now several important and well-funded global initiatives in operation to identify all of the proteins of key model species. One of the main requirements for these initiatives is a high-throughput crystallization facility to speed-up the protein identification process. The extent to which these technologies have advanced, calls for an updated review of current crystallographic theory and practice. This practical reference book features the latest conventional and high-throughput methods, and includes contributions from a team of internationally recognized leaders and experts. It will be of relevance and use to graduate students, research scientists and professionals currently working in the field of conventional and high-throughput macromolecular crystallography.

Author Information

Mark Sanderson is an experienced protein and nucleic-acid crystallographer and co-editor of Methods in Molecular Biology, who has solved several key ab initio structures, notably nucleic-acid proteins. Jane Skelly has widespread experience in the expression and crystallisation of macromolecules. Both authors lecture undergraduates in structural biology.

Order Information

Cat. No.	Name	Description	Price
HR5-224	**Macromolecular Crystallography**	each	**$156.00**

MACROMOLECULAR CRYSTALLOGRAPHY PROTOCOLS, VOLUME 1

- Author: Sylvie Doublie
- Publisher: Humana Press, 2007
- ISBN-10: 1588292924
- ISBN-13: 978-1588292926
- Hardcover: 312 pages

DESCRIPTION

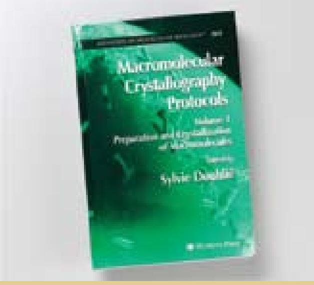

Macromolecular Crystallography Protocols, now in two volumes, examines major developments that have occurred since publication of the acclaimed first edition nearly a decade ago. Volume 1, Preparation and Crystallization of Macromolecules and Volume 2, Structure Determination, explore recent advances that have accelerated the pace of structural determination and made crystallography accessible to a broader range of investigators. Volume 1 is composed of detailed protocols for the preparation and optimization of crystals, including tips from the experts on the best methods for inducing proteins to adopt their crystalline form. Volume 2 complements the first volume by addressing laboratory techniques for crystal handling and structural characterization, as well as computational techniques for data collection, phasing, and refinement. The volume concludes with a detailed and insightful survey of available crystallographic software. These volumes will be an indispensable reference for obtaining macromolecular crystals and determining their three-dimensional structure.

Order Information

Cat. No.	Name	Description	Price
HR5-215	**Macromolecular Crystallography Protocols, Vol. 1**	each	**$40.00**

PROTEIN CRYSTALLIZATION STRATEGIES FOR STRUCTURAL GENOMICS

- Editor: Naomi E. Chayen
- Publisher: International University Line, 2007
- ISBN-10: 097207743X
- ISBN-13: 978-0972077439
- Hardcover: 290 pages

DESCRIPTION

Research advances in recent years have opened up the scope for the development of new methods and tools to overcome the bottleneck of protein crystallization. A variety of parameters that could previously not be explored are now accessible thanks to sophisticated apparatus and the development of new science-based techniques to monitor and control the process of crystallization. However, in order to become useful to the structural genomics effort, it is vital to miniaturize and automate these techniques and adapt them to cope with the vast numbers of "leads" resulting from the high-throughput screening procedures. Such efforts are those of the immediate future and the focus of this book.

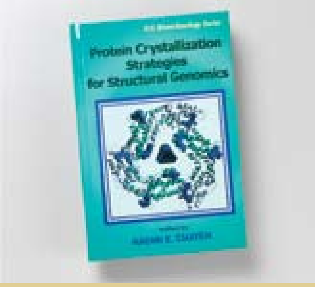

Order Information

Cat. No.	Name	Description	Price
HR5-230	**Protein Crystallization Strategies for Structural Genomics**	each	**$109.00**

PRACTICAL PROTEIN CRYSTALLOGRAPHY

- Author: Duncan E. McRee
- Publisher: Academic Press; 2nd edition, 1999
- ISBN-10: 0124860524
- ISBN-13: 978-0124860520
- Hardcover: 504 pages

DESCRIPTION

Covers all aspects of protein crystallography. It is a practical handbook intended to be used by anyone who wants to solve a structure by protein crystallography. This book is divided into four chapters: (1) Laboratory Techniques; (2) Data Collection Techniques; (3) Computational Techniques; and (4) Protein Crystallography Cookbook.

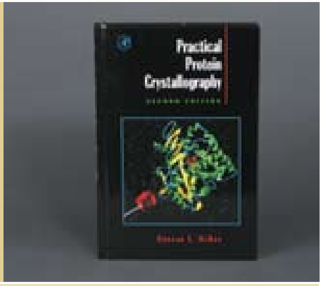

Order Information

Cat. No.	Name	Description	Price
HR5-211	**Practical Protein Crystallography**	each	**$142.00**

PREPARATION AND ANALYSIS OF PROTEIN CRYSTALS

- Author: Alexander McPherson
- Publisher: Krieger Publishing, 1989
- ISBN-10: 089464355X
- ISBN-13: 978-0894643552
- Hardcover: 384 pages

DESCRIPTION

This hard-to-find book is a classic. The first text written specifically for macromolecular crystallization. This twelve chapter book provides an interface between the techniques and practices common to most biochemists and the procedures familiar to x-ray diffractionists.

PREPARATION AND ANALYSIS OF PROTEIN CRYSTALS

Order Information

Cat. No.	Name	Description	Price
HR5-213	**Preparation and Analysis of Protein Crystals**	each	**$75.00**

PROTEIN METHODS

- Authors: Daniel M. Bollag, Michael D. Rozycki, & Stuart J. Edelstein
- Publisher: Wiley-Liss Publishing; 2nd edition, 1996
- ISBN-10: 0471118370
- ISBN-13: 978-0471118374
- Paperback: 415 pages

DESCRIPTION

This text is a handy and practical guide of protein methods which includes 36 pages dedicated to a straightforward, practical introduction to macromolecular crystallization. Chapter titles include: Preparation for Protein Isolation, Protein Extraction and Solubilization, Protein Concentration Determination, Concentrating Protein Solutions, Gel Electrophoresis, IEF, Immunoblotting, Ion Exchange, Affinity and Gel Filtration Chromatography, and Hanging Drop Crystallization.

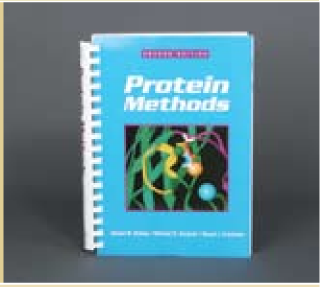

Order Information

Cat. No.	Name	Description	Price
HR5-223	Protein Methods	each	$98.00

METHODS IN ENZYMOLOGY - PART C

- Editor: Charles W. Carter, Jr.
- Publisher: Elsevier, 2003
- ISBN-10: 0121822710
- ISBN-13: 978-0121822712
- Hardcover: 368 pages

DESCRIPTION

Macromolecular Crystallography, Part C (Methods in Enzymology, Volume 368). Topics covered include: methodological methods in crystals and data analysis.

Methods in ENZYMOLOGY

Order Information

Cat. No.	Name	Description	Price
HR5-210	Methods in Enzymology - Vol 368, Part C	each	$150.00

METHODS IN ENZYMOLOGY - PART D

- Editor: Charles W. Carter, Jr.
- Publisher: Elsevier, 2003
- ISBN-10: 0121827771
- ISBN-13: 978-0121827779
- Hardcover: 679 pages

DESCRIPTION

Macromolecular Crystallography, Part D (Methods in Enzymology, Volume 374). Topics covered include: Phases, Map Interpretation and Refinement, Analysis and Software.

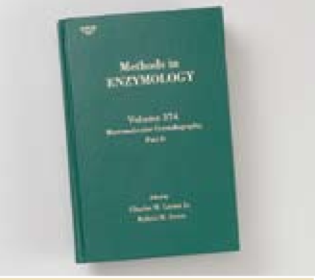

Order Information

Cat. No.	Name	Description	Price
HR5-212	Methods in Enzymology Volume 374, Part D	each	$150.00

CRYSTALLOGRAPHY MADE CRYSTAL CLEAR

- Author: Gale Rhodes
- Publisher: Academic Press; 3rd edition, 2006
- ISBN-10: 0125870736
- ISBN-13: 978-0125870733
- Paperback: 352 pages

DESCRIPTION

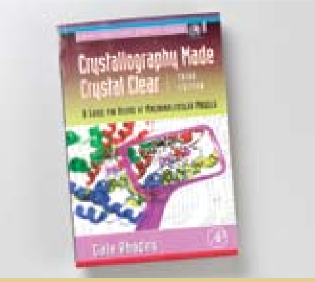

This book is an exciting resource for those who are vitally interested in macromolecular models derived from x-ray crystallography but are unfamiliar with, or daunted by, crystallography itself. This book is for readers who want an intellectually satisfying understanding of how three-dimensional models of proteins and nucleic acids are derived from x-ray diffraction analysis of crystals. This understanding is essential for wise use of crystallographic models in studying enzyme catalysis, protein-mediated transport and recognition, protein/nucleic acid interaction, and protein folding. It is also vital in interpreting data from chemical, kinetic, thermodynamic, and spectroscopic studies of macromolecules.

Order Information

Cat. No.	Name	Description	Price
HR5-233	Crystallography Made Crystal Clear	each	$63.00

OUTLINE OF CRYSTALLOGRAPHY FOR BIOLOGISTS

- Author: David Blow
- Publisher: Oxford University Press, 2002
- ISBN-10: 0198510519
- ISBN-13: 978-0198510512
- Paperback: 248 pages

DESCRIPTION

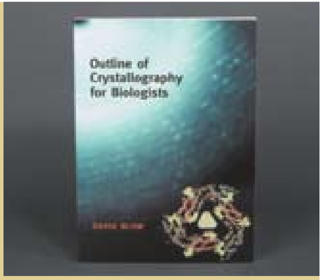

Outline of Crystallography for Biologists is intended for researchers and students in the biological sciences who require an insight into the methods of x-ray crystallography without needing to learn all the relevant theory. The main text is purely descriptive and is readable by those with minimal mathematical knowledge. Some mathematical detail is given throughout in boxes, but these can be ignored. Theory is limited to the essentials required to comprehend issues of quality. There is an extensive reference section and suggestions for further reading for those who wish to delve deeper. The first part, 'Fundamentals', presents the underlying ideas which allow x-ray structure analysis to be carried out and provides an appropriate background to courses in structural determination. The second part, 'Practice', gives more information about the procedures employed in the course of crystal structure determination. The emphasis is on the quality measures of x-ray diffraction analysis to give the reader a critical insight into the quality and accuracy of a structure determination and to enable the reader to appreciate which parts of a structure determination may have caused special difficulty. There is no pretence of completeness and many matters discussed in standard crystallography texts are deliberately omitted. However, issues not brought out in the standard texts are discussed, making it a useful resource for non-practicing crystallographers as well.

Order Information

Cat. No.	Name	Description	Price
HR8-086	Outline of Crystallography for Biologists	each	$82.00

INTRODUCTION TO MACROMOLECULAR CRYSTALLOGRAPHY

- Author: Alexander McPherson
- Publisher: Wiley-Blackwell; 2nd edition, 2009
- ISBN-10: 0470185902
- ISBN-13: 978-0470185902
- Paperback: 267 pages

DESCRIPTION

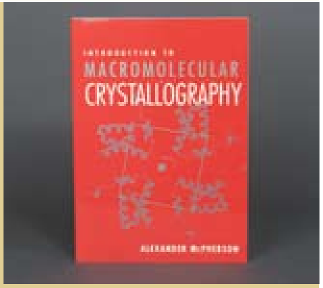

This book provides a comprehensive, approachable summary of the field of crystallography, from the fundamental theory of diffraction and properties of crystals to applications in determining macromolecular structure.

Dedicated to providing a complete introduction to the subject that does not assume a background in physics or math, the book's contents flow logically from basic principles to key methods, such as those for solving phase problems, interpretation of Patterson maps, and the difference Fourier method. The Introduction includes: Practical Instruction on the Interpretation of Data and Methods for Determining Phases.

Order Information

Cat. No.	Name	Description	Price
HR8-062	Introduction to Macromolecular Crystallography	each	$80.00

PRINCIPLES OF PROTEIN X-RAY CRYSTALLOGRAPHY

- Author: Jan Drenth
- Publisher: Springer; 3rd edition, 2006
- ISBN-10: 0387333347
- ISBN-13: 978-0387333342
- Hardcover: 332 pages

DESCRIPTION

Strong on crystallography, light on crystal growing (17 pages). Fifteen chapters including Crystallizing a Protein, X-Ray Sources and Detectors, Crystals, Theory of X-Ray Diffraction By a Crystal, Average Reflection Intensity and Distribution of Structure Factor Data, Special Forms of the Structure Factor, Phase Improvement, The Solution of the Phase Problem by the Isomorphous Replacement Method, Anamolous Scattering, Molecular Replacement, Direct Methods, Refinement, Combination of Phase Information, Phase Information from Partial Structure Data, Solvent Flattening, and Molecular Averaging.

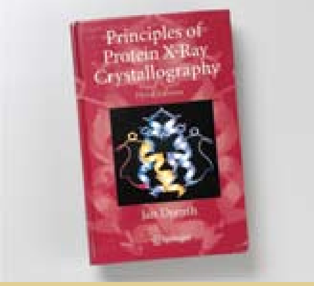

Order Information

Cat. No.	Name	Description	Price
HR5-219	**Principles of Protein X-Ray Crystallography**	each	**$90.00**

MEMBRANE PROTEIN PURIFICATION & CRYSTALLIZATION, A PRACTICAL GUIDE

- Editors: Carola Hunte, Gebhard Von Jagow, Hermann Schagger
- Publisher: Academic Press; 2nd edition, 2002
- ISBN-10: 0123617766
- ISBN-13: 978-0123617767
- Spiral-bound: 316 pages

DESCRIPTION

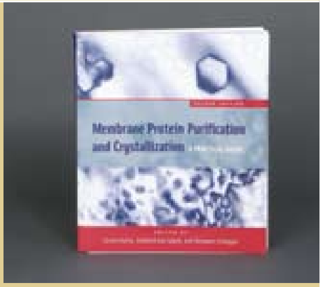

This guide presents isolation and crystallization techniques in a concise form, emphasizing the critical aspects unique to membrane proteins. It explains the principles of the methods and provides protocols of general use, permitting researchers and students new to this area to adapt these techniques to their particular needs. Key features: Provides general guidelines and strategies for isolation and crystallization of membrane proteins. Gives detailed protocols that have wide application, and low specialized equipment needs. Emphasizes recent progress in production and purification of recombinant membrane proteins, especially of histidine-tagged and other affinity-epitope-tagged proteins. Summarizes recent developments of Blue-Native PAGE, a high resolution separation technique, which is independent of the use of recombinant techniques, and is especially suited for proteomic analyses of membrane protein complexes. Gives detailed protocols for membrane protein crystallization, and describes the production and use of antibody fragments for high resolution crystallization. Presents a comprehensive guide to 2D-crystallization of membrane proteins.

Order Information

Cat. No.	Name	Description	Price
HR5-241	**Membrane Protein Purification and Crystallization**	each	**$98.00**

METHODS AND RESULTS IN CRYSTALLIZATION OF MEMBRANE PROTEINS

- Editor: So Iwata
- Publisher: International University Line, 2002
- ISBN-10: 0963681796
- ISBN-13: 978-0963681799
- Hardcover: 380 pages

DESCRIPTION

This text consists of four parts. Parts 1 and 2 include an introduction as well as general principles and techniques in membrane protein crystallization. This is followed by chapters covering the use of detergents, crystallization in lipidic cubic phases, the use and generation of antibody fragments, porin as a model, and crystallization of membrane proteins in oils. Part 3 focuses on specific examples of membrane proteins whose structures have been solved. Part 4 covers the design of a kit for membrane protein crystallization.

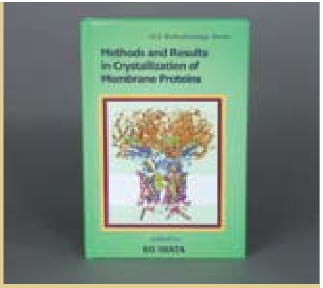

Order Information

Cat. No.	Name	Description	Price
HR5-237	**Methods and Results in Crystallization of Membrane Proteins**	each	**$109.00**

MACROMOLECULAR CRYSTALLIZATION METHODS, VOLUME 34, NUMBER 3

- Edited by: Alexander McPherson
- Publisher: Elsevier; Methods, Volume 34, Number 3, November 2004
- ISSN: 1046-2023
- Paperback: 172 pages

DESCRIPTION

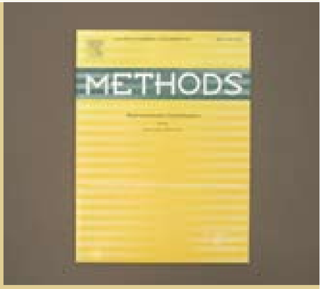

This is a single issue of Methods, a companion journal to Methods in Enzymology, focusing on rapidly developing techniques. This single issue is dedicated to Macromolecular Crystallization and begins with an Introduction to Protein Crystallization by Alexander McPherson. The issue features the following additional thirteen articles. Protein crystallization and phase diagrams, Neer Asherie. Growth and disorder of macromolecular crystals: insights from atomic force microscopy and x-ray diffraction studies, Alexander J. Malkin and Robert E. Thorne. Ions from the Hofmeister series and osmolytes: effects on proteins in solution and in the crystallization process, Kim D. Collins. Effects of naturally occurring osmolytes on protein stability and solubility; issues important in protein crystallization, D.W. Bolen. Practical aspects of using the microbatch method in screening conditions for protein crystallization, Allan D'Arcy, Aengus Mac Sweeney, and Alexander Haber. Automated systems for protein crystallization, Joel Bard, Kimberly Ercolani, Kristine Svenson, Andrea Olland, and Will Somers. Lipidic cubic phases as matrices for membrane protein crystallization, Peter Nollert. The use of recombinant and molecular engineering in protein crystallization, Zygmunt Derewenda. A pedestrian guide to membrane protein crystallization, Michael Wiener. Crystallization data mining in structural genomics: using positive and negative results to optimize protein crystallization screens, Rebecca Page and Raymond C. Stevens. Predictive models for protein crystallization, Bernhard Rupp and Junwen Wang. Crystallization of RNA and RNA-protein complexes, Ailong Ke and Jennifer A. Doudna. Macromolecular cryocrystallography – methods for cooling and mounting protein crystals are cryogenic temperatures, J.W. Pflugrath.

Order Information

Cat. No.	Name	Description	Price
HR8-061	**Macromolecular Crystallization**	each	**$40.00**

JOURNAL OF STRUCTURAL BIOLOGY

- Editor: Alexander McPherson
- Publisher: Academic Press; Volume 142, Number 1, April 2003
- ISSN: 1047-8477
- Paperback: 231 pages

DESCRIPTION

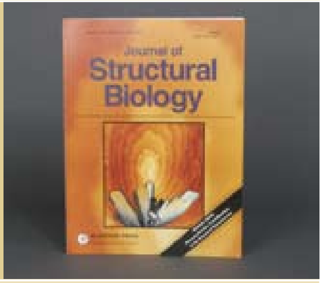

This special issue of the Journal of Structural Biology focuses on Macromolecular Crystallization in the Structural Genomics Era. Following an Introduction by Alexander McPherson, there are 19 regular articles focused on the crystallization of biological macromolecules. If crystallization is your thing, you need this. Thanks to everyone contributing to this special issue. Well done everyone!

Order Information

Cat. No.	Name	Description	Price
HR8-100	**Journal of Structural Biology**	each	**$60.00**

Crystal image.

Sylvia Luckner, Lehrstuhl für Biotechnik, University Erlangen-Nürnberg Germany.

PRESENT AT THE FLOOD: HOW STRUCTURAL MOLECULAR BIOLOGY CAME ABOUT

- Author: Richard Dickerson
- Publisher: Sinauer Associates, 2005
- ISBN-10: 0878931686
- ISBN-13: 978-0878931682
- Paperback: 307 pages

DESCRIPTION

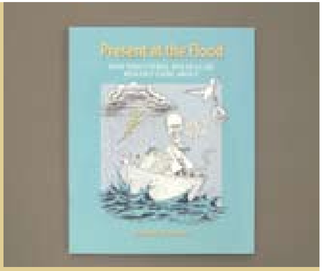

This book chronicles a revolution in molecular biology—the crucial 30 years (roughly between 1933 and 1963) during which our ideas about proteins and nucleic acids changed from those of formless, functionless organic chemicals into precisely structured molecular machines with specific biological purpose. Proteins evolved from being colloidal micelles or globules with no specific structure (or even sequence) into quite precisely structured molecular catalysts, carrier proteins, and information-sensing agents. Indeed, the very idea that the amino acids of a protein were linked in a specific order in long linear chains was not accepted initially. During this same time period, DNA changed from being a sterile repeating polymer of no particular function (the tetranucleotide hypothesis) into a double helix that serves as the archive of genetic information.

Order Information

Cat. No.	Name	Description	Price
HR5-214	**Present at the Flood**	each	**$35.00**

CRYSTALS AND LIFE: A PERSONAL JOURNEY

- Author: Cele Abad-Zapatero
- Publisher: International University Line, 2002
- ISBN-10: 0972077405
- ISBN-13: 978-0972077408
- Paperback: 300 pages

DESCRIPTION

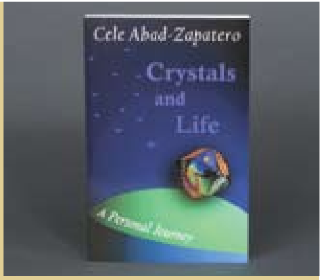

This book introduces crystallography, the field, and applications of structural biology using metaphors from the arts, music, poetry, and architecture. The reader will find a tapestry filled by personal experiences, historical anecdotes, biographical snapshots of scientific heroes of the field, and crystallographic concepts illustrated with artistic analogies. All these elements create a concoction that makes crystallography accessible, comprehensible, intriguing, inspiring, and beautiful. The book is divided in sections covering the following: the basic elements of crystallography, novel technologies, practical applications and future perspectives. Since crystallography is quintessentially a visual science, the illustrations play a very important role in providing an excellent set of landmarks to guide the reader on this personal journey of discovery.

Order Information

Cat. No.	Name	Description	Price
HR5-239	**Crystals and Life: A Personal Journey**	each	**$15.00**

PROTEIN CRYSTALLOGRAPHY: A CONCISE GUIDE

- Authors: Eaton E. Lattman and Patrick J. Loll
- Publisher: The Johns Hopkins University Press, 2008
- ISBN-10: 0801888085
- ISBN-13: 978-0801888085
- Paperback: 152 pages

DESCRIPTION

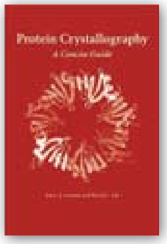

The proteome remains a mysterious realm. Researchers have determined the structures of only a small fraction of the proteins encoded by the human genome. Crystallography continues to be the primary method used to determine the structures of the remaining unknown proteins. This imaging technique uses the diffraction of x-rays to determine a protein's three-dimensional molecular structure. Drawing on years of research and teaching experience, Eaton E. Lattman and Patrick J. Loll use clear examples and abundant illustrations to provide a concise and accessible primer on protein crystallography. Discussing the basics of diffraction, the behavior of two- and three-dimensional crystals, phase determination (including MIR and MAD phasing and molecular replacement), the Patterson function, and refinement, Lattman and Loll provide a complete overview of this important technique, illuminated by physical insights. The crisp writing style and simple illustrations will provide beginner crystallographers with a guide to the process of unraveling protein structure.

Order Information

Cat. No.	Name	Description	Price
HR5-221	**Protein Crystallography: A Concise Guide**	each	**$37.00**

PROTEIN CRYSTALLIZATION

- Editor: Terese Bergfors
- Publisher: International University Line; 2nd edition, 2009
- ISBN-10: 0972077448
- ISBN-13: 978-0972077446
- Hardcover: 474 pages

DESCRIPTION

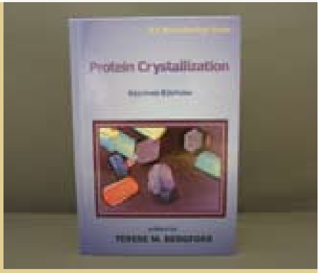

Completely revised and updated, Protein Crystallization, 2nd Edition is a greatly expanded follow-up to the best-selling 1st edition. Completely new chapters on high-throughput methods, mass spectrometry, microcalorimetry, counterdiffusion, heavy-atom derivatization, selenomethionine-labeling, rational strategies for crystallization, and protein modification to improve crystallization. Updated chapters on formulation of the protein before crystallization, characterization of the protein by dynamic light scattering, classic methods and the phase diagram, seeding, and cryoprotection of the crystals. Thirty full-color plates for evaluating crystallization drops. Separate section of laboratory exercises, ideal for crystallization courses. A-Z glossary.

Order Information

Cat. No.	Name	Description	Price
HR5-231	**Protein Crystallization**	each	**$99.00**

VIEWERS

STEREOPTICON STEREO VIEWER

FEATURES

- Folding 3-D viewer

DESCRIPTION

The Stereopticon is punched from black plastic and folds completely flat to fit inside a book. The Stereopticon uses plastic lenses.

Order Information

Cat. No.	Name	Description	Price
HR4-513	**Stereopticon**	each	**$9.00**
HR4-515	**Stereopticon**	10 pack	**$81.00**

STEREO VIEWER

FEATURES

- 3-D stereo glasses with folding wire legs for compact storage

DESCRIPTION

The Stereo Viewer features an adjustable nose bridge. It is made of high impact, crystal-clear plastic with a 2.2 magnification, 4.7 focal length and adjustable inter-pupillary distance.

Order Information

Cat. No.	Name	Description	Price
HR4-517	**Stereo Viewer**	each	**$12.00**

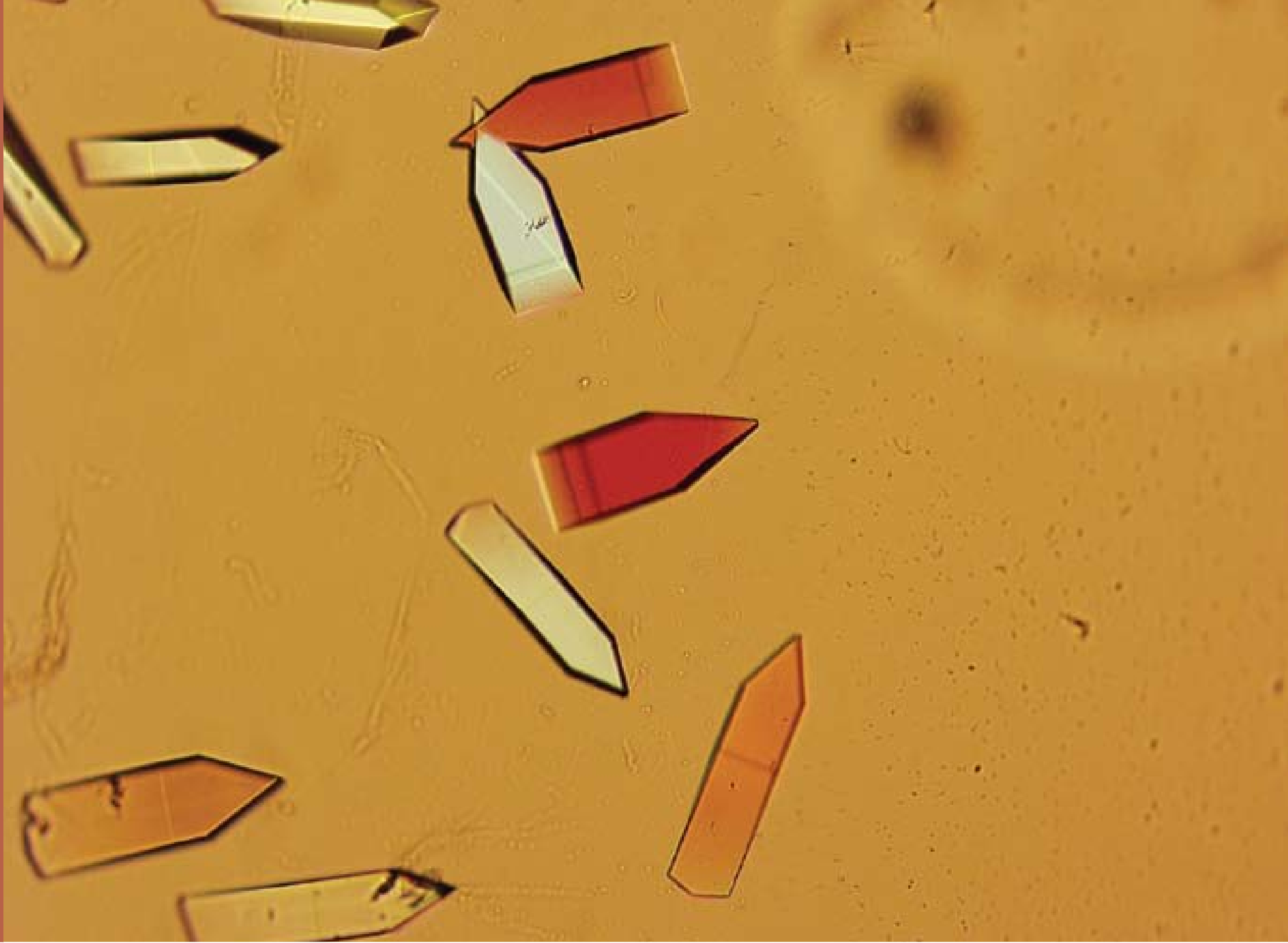

PROTEIN CRYSTALLIZATION STANDARDS

Protein crystal.

Alexey Rak, Max-Planck-Institut fur Molekulare Physiologie, Department of Physical Biochemistry, Dortmund, Germany.

TABLE OF CONTENTS

PAGES

176 GLUCOSE ISOMERASE

177 LYSOZYME

178 XYLANASE

179 LIPASE B

HAMPTON
RESEARCH

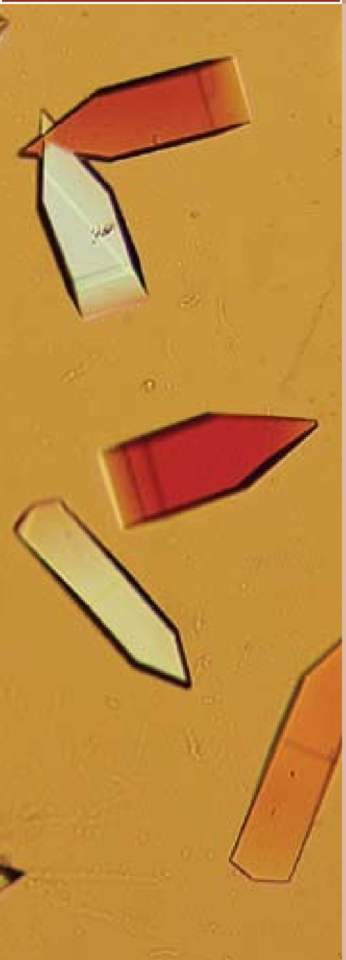

GLUCOSE ISOMERASE

APPLICATION

- Crystallization grade protein standard

FEATURES

- Protein standard for crystallization studies and applications
- Crystallization training and student demos

DESCRIPTION

Glucose Isomerase can be crystallized by sitting, hanging or sandwich drop vapor diffusion, batch, liquid/liquid diffusion, pH shift, temperature jump, as well as by dialysis. It can be crystallized rapidly across a broad range of screen conditions which makes the enzyme an excellent candidate for crystallization studies and demonstrations. Glucose Isomerase is supplied as a crystal suspension in the following medium: Approximately 33 mg/ml in 6 mM TRIS hydrochloride pH 7.0, 0.91 M Ammonium sulfate, 1 mM Magnesium sulfate. Glucose Isomerase is prepared by Macrocrystal Oy, exclusively for Hampton Research.

References

1. *Carrell, H. L., et al., Proc. Natl. Acad. Sci. USA (1989) Vol 86, 4440-4444.*
2. *Glucose Isomerase from Streptomyces rubiginosus - potential molecular weight standard for small-angle x-ray scattering. Maciej Kozak. J. Appl. Cryst. (2005). 38, 555-558.*

Order Information

Cat. No.	Name	Description	Price
HR7-100	**Glucose Isomerase**	1 g	**$284.00**
HR7-102	**Glucose Isomerase**	100 mg	**$60.00**

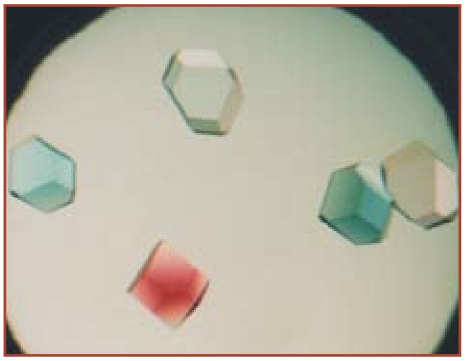

Glucose Isomerase Crystals

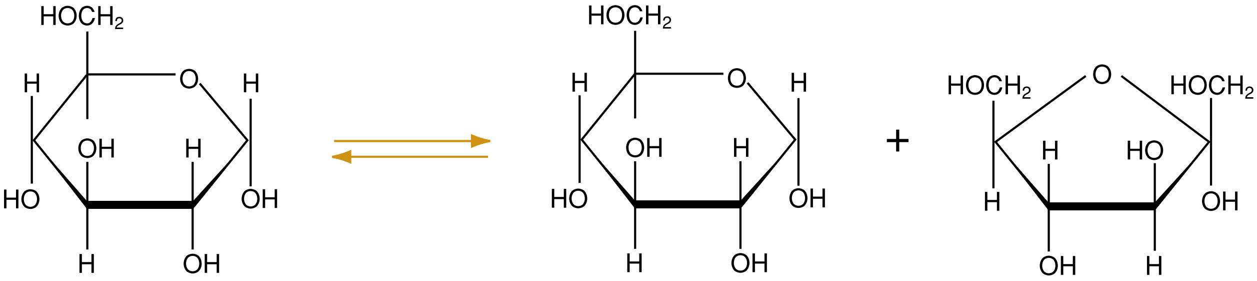

Glucose **Glucose** **Fructose**

Glucose Isomerase catalyzes the isomerization reaction shown. The reaction is reversible and goes to equilibrium, approximately equal concentrations of fructose and glucose.

LYSOZYME

APPLICATION

- Crystallization grade protein standard

FEATURES

- Protein standard for crystallization studies and applications
- Crystallization training and student demos
- Hen egg white source

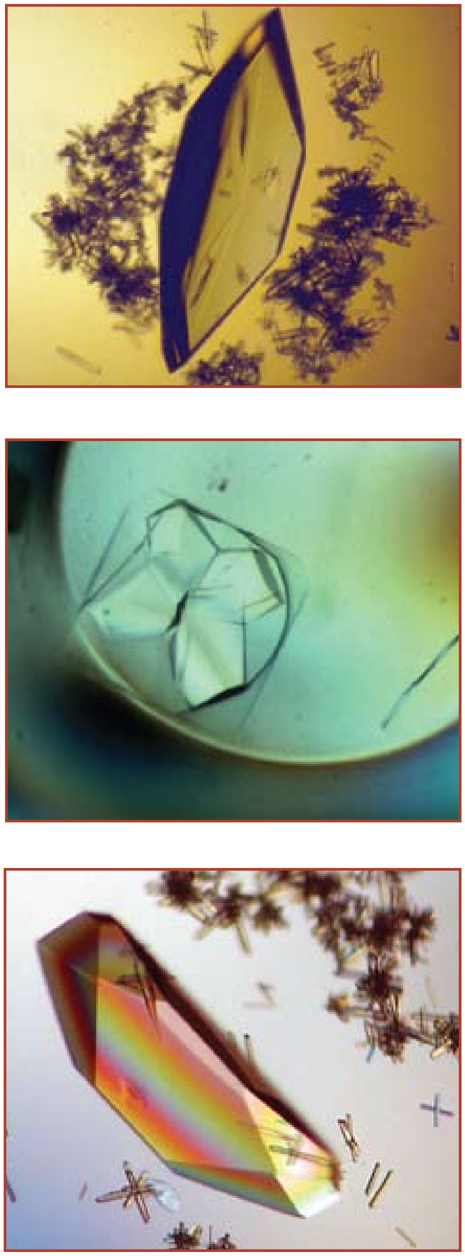

Lysozyme crystals grown using the Hampton Research Silver Bullets screen

DESCRIPTION

A very cost-effective and convenient crystallization grade lysozyme (hen egg white). The Lysozyme Kit is a convenient buffer and lysozyme set for teaching, demonstrations or instrumentation setup and validation procedures. The kit features 12 vials with 20 mg of lysozyme and 12 vials of 0.02 M Sodium acetate trihydrate pH 4.6 buffer. To prepare the lysozyme for crystallization, simply add the supplied buffer to the protein. No weighing, no messing with fluffy crystalline powder, no mess, and no variability. Each tube can make up to 1 ml of a 20 mg/ml stock or 200 μl of a 100 mg/ml stock. Twelve sets of tubes and a stable shelf life make this a handy kit to keep in the fridge. Lysozyme is available separately in an 8 g vial of lysozyme powder. To dissolve the powder, use the Lysozyme Crystallization Buffer. For teaching and demonstrations requiring crystals in a short period of time, one should consider the 15-Minute Lysozyme Crystallization Reagent (30% w/v Polyethylene glycol monomethyl ether 5,000, 1.0 M Sodium chloride, 0.05 M Sodium acetate trihydrate pH 4.6). For flexibility, one can also use the Lysozyme Crystallization Buffer stock (1.0 M Sodium acetate trihydrate pH 4.6) or the StockOptions Sodium Acetate buffer kit and 5.0 M Sodium chloride to create custom grid screens and practicals for demonstrating and teaching the fundamentals of optimization of sample concentration and reagent concentration. The Grid Screen Sodium Chloride kit is a convenient kit for demonstrating the effect of varying pH and reagent concentration on the solubility and crystallization of a protein.

Order Information

Cat. No.	Name	Description	Price
HR7-108	**Lysozyme Kit**	12 x 20 mg plus 12 x 1 ml solubilization buffer	**$60.00**
HR7-110	**Lysozyme**	8 g	**$60.00**
HR2-731	**Lysozyme Crystallization Buffer**	1.0 M Sodium acetate trihydrate pH 4.6, 100 ml	**$29.00**
HR2-805	**15 Minute Lysozyme Crystallization Reagent**	30% w/v PEG MME 5,000, 1.0 M Sodium chloride, 0.05 M Sodium acetate trihydrate pH 4.6, 100 ml	**$39.00**
HR2-637	**Sodium Chloride for Lysozyme Crystallization**	5.0 M Sodium chloride, 200 ml	**$33.00**
HR2-233	**StockOptions Sodium Acetate kit**	1.0 M Sodium acetate trihydrate pH 3.6 to 5.6, 21 reagents	**$195.00**
HR2-219	**Grid Screen Sodium Chloride kit**	Sodium chloride versus pH crystallization screen kit	**$175.00**

XYLANASE

APPLICATION

- Crystallization grade protein standard

FEATURES

- Protein standard for crystallization studies and applications
- Crystallization training and student demos

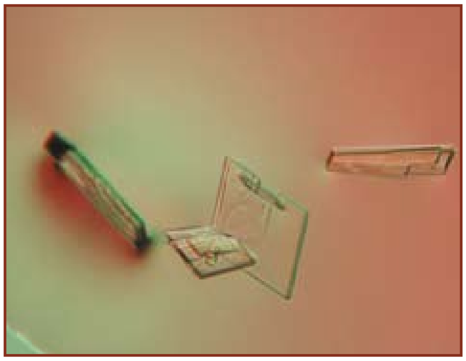

Xylanase Crystals

DESCRIPTION

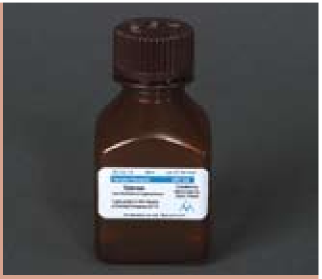

Xylanase can be crystallized by sitting, hanging or sandwich drop vapor diffusion, batch, liquid/liquid diffusion, pH shift, temperature jump, as well as by dialysis. It can be crystallized rapidly across a broad range of screen conditions which makes the enzyme an excellent candidate for crystallization studies and demonstrations. It is supplied as a solution, produced by dissolving pure xylanase crystals with phosphate buffer and glycerol. The solution is filtered with a 0.2 micron filter and has the final formulation:

Xylanase Concentration: 36 mg/ml
Absorbance at 280 nm: 97
Activity: 648,000 nkat/ml
Glycerol Concentration: 43% w/v
Na/K phosphate: 0.18 M
pH: 7
Specific Gravity: 1.124 g/ml

Xylanase is prepared by Macrocrystal Oy, exclusively for Hampton Research.

References

1. *Torronen, A., et al., The EMBO Journal (1994) Vol 13, No 11, 2493-2501.*
2. *Torronen, A. ,et al., J. Mo. Biol. (1993) Vol 233, 313-316.*
3. *Crystals of family 11 xylanase II from Trichoderma longibrachiatum that diffract to atomic resolution. Natalia Moiseeva and Marc Allaire. Acta Cryst. (2004). D60, 1275-1277.*
4. *Structures of an orthorhombic form of xylanase II from Trichoderma reesei and analysis of thermal displacement. Watanabe et al. Acta Cryst. (2006). D62, 784-792.*
5. *Cryogenic (<20 K) helium cooling mitigates radiation damage to protein crystals. B. Leif Hanson et al. Acta Cryst (2007) D63, 486-492.*

Order Information

Cat. No.	Name	Description	Price
HR7-104	**Xylanase**	1 g	**$463.00**
HR7-106	**Xylanase**	100 mg	**$69.00**

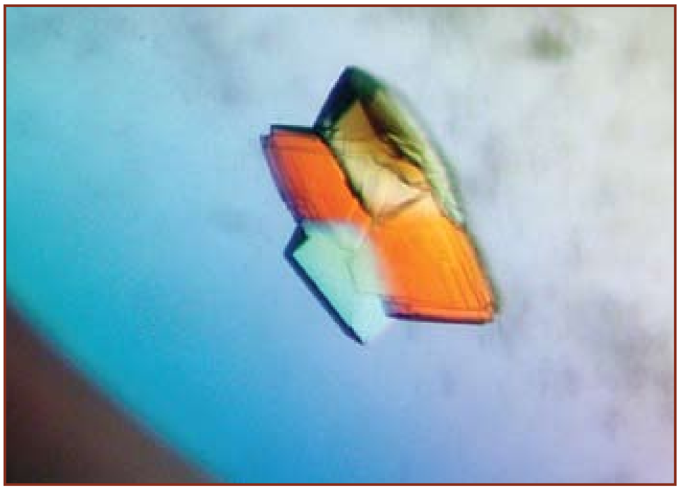

Protein crystals grown using the Silver Bullets screen from Hampton Research.

Bob Cudney & Peter Nguyen, Hampton Research; Alex McPherson, University of California, Irvine.

APPLICATION

- Protein standard for crystallization studies and applications
- Crystallization training and student demos

Lipase B Crystals

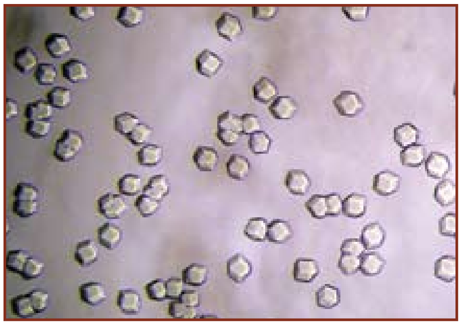

Lipase B Crystals Grown by Macrocrystal Oy

DESCRIPTION

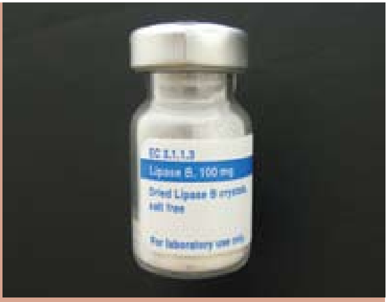

Lipase B is a pure, dried, crystalline form of Lipase B from Candida antarctica produced by submerged fermentation of a genetically modified Aspergillus oryzae microorganism. It is a white crystal powder where Lipase B is present as crystals and there are no other ingredients or buffering salts.

Lipases (EC 3.1.1.3) make up a diverse group of water soluble enzymes that catalyze the hydrolysis of ester bonds in water–insoluble, lipid substrates. Lipases are characterized by their drastically increased esterase activity when adsorbed to a lipid surface (interfacial activation). Lipases are a form of esterases. They perform essential roles in the digestion, transport, and processing of dietary lipids such as triglycerides, fats, oils in most, if not all, living organisms.

Chemical name, synonym Triacylglycerol hydrolase EC 3.1.1.3
M_r 35 kDa
pH 5.0 - 7.0
Isoelectric point 6.0
Appearance: White crystal powder

References

1. *Crystallographic and molecular-modeling studies of lipase B from Candida antarctica reveal a stereospecificity pocket for secondary alcohols. Uppenberg, J., Ohrner, N., Norin, M., Hult, K., Kleywegt, G.J., Patkar, S., Waagen, V., Anthonsen, T., Jones, T.A. Biochemistry v34 pp. 16838-51, 1995.*
2. *The sequence, crystal structure determination and refinement of two crystal forms of lipase B from Candida antarctica. Uppenberg, J., Hansen, M.T., Patkar, S., Jones, T.A. Structure v2 pp. 293-308, 1994.*
3. *Crystallization and preliminary X-ray studies of lipase B from Candida antarctica. Uppenberg J., Patkar S.; Bergfors T., Jones, T. A.. Journal of Molecular Biology, 1994, vol. 235, no2, pp. 790-792.*

Order Information

Cat. No.	Name	Description	Price
HR7-099	**Lipase B**	100 mg	**$184.00**

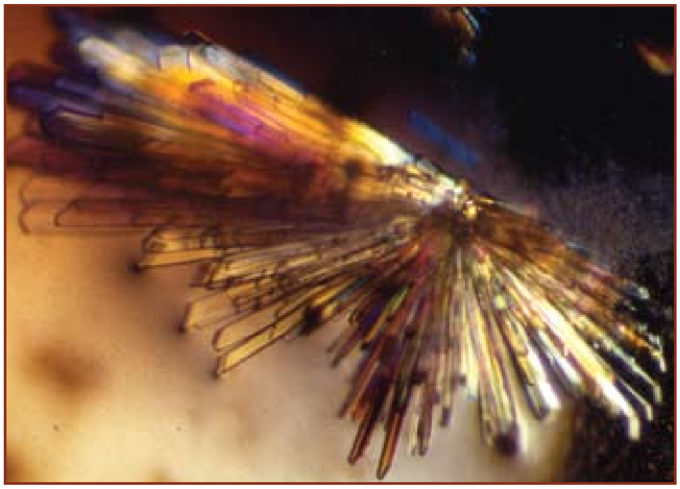

Protein crystals grown using the Silver Bullets screen from Hampton Research.

Bob Cudney & Peter Nguyen, Hampton Research; Alex McPherson, University of California, Irvine.

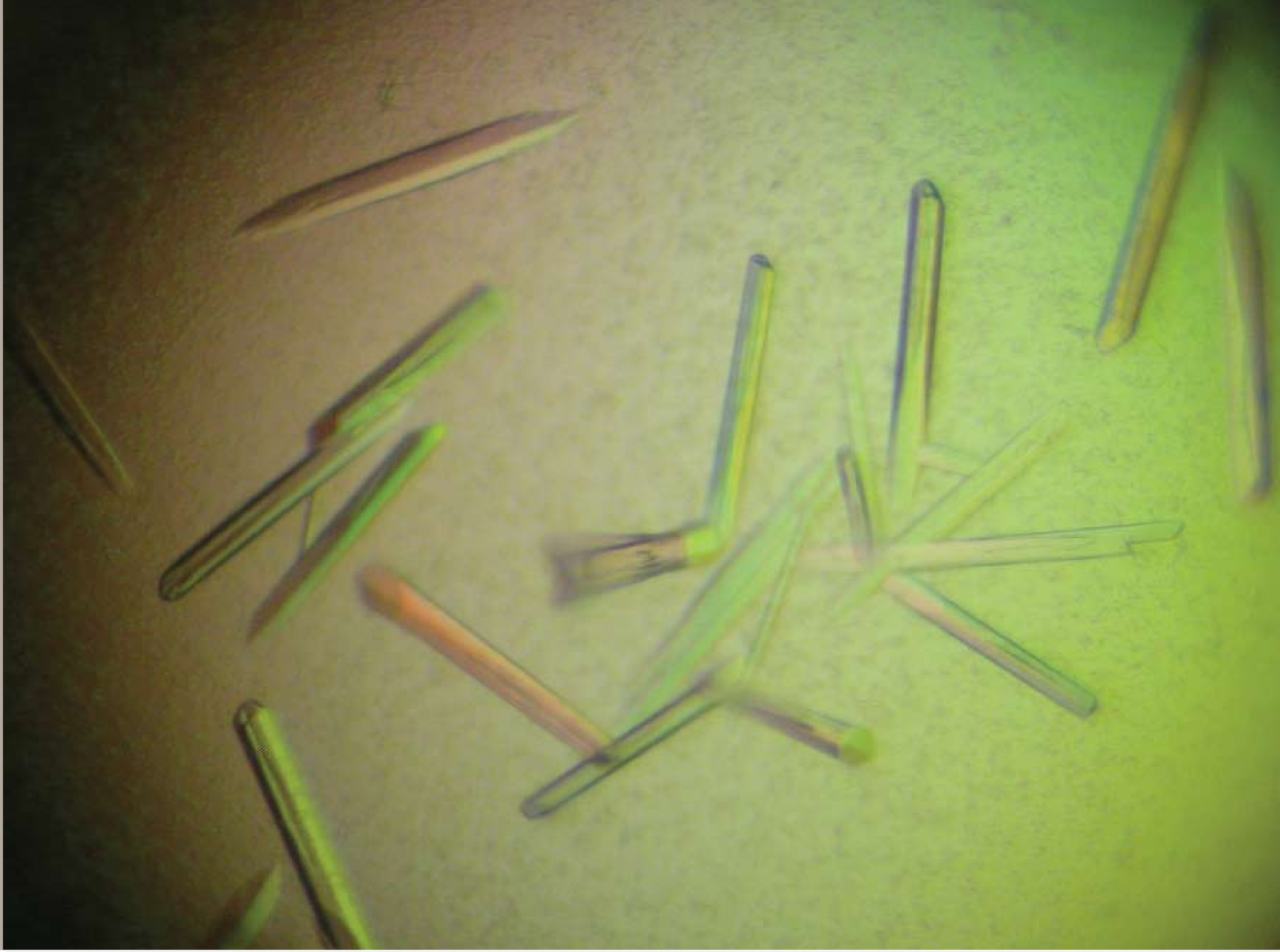

CRYSTAL GROWTH 101

Protein crystals.

Kimberly J. Skinner, Structural Biology and Biophysics, Pfizer Global R&D, Sandwich, Kent, United Kingdom.

TABLE OF CONTENTS

PAGES

182 - 183 PRELIMINARY SAMPLE PREPARATION
184 - 185 CRYSTAL GROWTH TECHNIQUES
186 HANGING DROP VAPOR DIFFUSION CRYSTALLIZATION
187 - 189 SITTING DROP VAPOR DIFFUSION CRYSTALLIZATION
190 CRYSTALLIZATION UNDER OIL
191 - 192 DIALYSIS CRYSTALLIZATION
193 VIEWING CRYSTALLIZATION EXPERIMENTS
194 BIREFRINGENCE
195 - 196 SALT OR BIOLOGICAL CRYSTALS?
197 pH VERSUS NUMBER OF CRYSTALS
198 pH POINTS TO PONDER
198 CRYSTAL GROWTH VIA pH RELAXATION
199 PROTEIN CONCENTRATION VERSUS NUMBER OF CRYSTALS
200 - 201 CRYSTALLIZATION POLLING BOOTH
202 CRYO QUICKIES
203 HOW TO: CRYOCRYSTALLOGRAPHY AND THE CRYSTALCAP SYSTEM
204 USING HALIDES FOR PHASING MOLECULAR STRUCTURES
205 MICROSEEDING
206 DETERGENTS AT WORK
207 - 208 PROTEIN CRYSTALLIZATION RECIPES
209 - 210 TEMPERATURE AS A CRYSTALLIZATION VARIABLE
212 - 219 TIPS FROM RAMC
220 - 234 CRYSTALLIZATION TIPS A-Z
235 - 263 CRYSTALLIZATION TIPS
264 - 265 CRYSTALLIZATION BUFFERS TABLE
266 - 267 SOLUBILITY TABLE

Lyophilization

Avoid lyophilization. Even though there are many examples of proteins which crystallize after lyophilization (lysozyme, thaumatin, hemoglobin, etc.), lyophilization is to be avoided when possible. If the protein is lyophilized, it needs to be dialyzed before crystallization. Dialyze the protein against deionized water or a stabilization buffer before crystallization. Dialysis will remove non-volatile buffers and other chemicals that may have been present before lyophilization.

Ammonium Sulfate Precipitation

Avoid using Ammonium sulfate precipitation as a final purification and/or concentration step. It is often very difficult to completely remove all the Ammonium sulfate by a desalting column of dialysis. The remaining trace amounts can interfere with crystallization screening results and create reproducibility problems. It is not uncommon for trace amounts of Ammonium sulfate in the sample to cause precipitation or excessive nucleation in screen conditions containing polyethylene glycol and salt.

Batches

Avoid combining different purification batches for crystallization trials. Purification conditions and procedures are never identical so each batch should be screened separately.

Profile the Protein

Ideally, you will purify your own protein, but this is not always reality. So, it is always a good idea to characterize your protein before beginning crystallization experiments. Profiling your protein before crystallization can often provide valuable clues during screening and optimization of crystallization conditions. Assay to seriously consider:

- SDS-PAGE
- Native PAGE or Dynamic Light Scattering
- IEF (Isoelectric Focusing) Gel
- Mass Spectroscopy

The results of these assays can:

- Determine the purity of the sample
- Determine the homogeneity of the sample
- Identify batch to batch variations
- Identify stability problems with the sample

How Pure?

How pure should the protein sample be for crystallization trials? As pure as possible. That's some answer, is it not? Integrating common sense into the question, we might arrive at the following answer. For initial screening, the sample should be at least 90 to 95% pure on a Coomassie stained SDS-PAGE. Finally, it does no harm to screen an "impure sample" as one can always perform further purification. Remember, crystallization used to be considered a very powerful purification tool (and still is!).

If the initial screen does not produce crystals, any promising results, or it becomes next to impossible to improve crystal quality during optimization, one should consider further purification of the sample.

Storing the Sample

Most proteins can be stored successfully at 4°C or -70°C.

Check with the person preparing the protein or compare your protein to a similar protein in the literature for best storage temperature.

Ideally, one should assay the activity and stability of the protein before storage and then later on at various points in time to determine the sample storage stability.

Repeated freezing and thawing of the sample should be avoided. Aliquot the sample into multiple small microcentrifuge tubes. Make the aliquots small enough so that the entire aliquot can be consumed in the experiment after thawing.

Sometimes people like to add Glycerol (10 to 50% v/v) to help proteins better tolerate freezing. Avoid this if possible since it is often difficult to remove Glycerol by dialysis or filtration. The presence of Glycerol is a crystallization variable. It can behave as a precipitant, an additive, or cryoprotectant and therefore can influence the outcome of a crystallization experiment.

In general, it is better to store proteins more concentrated than diluted. When too dilute, adsorption of the protein onto the storage container can lead to significant losses. However, precipitation can sometimes be a problem when the protein is stored too concentrated.

Sample Handling

Be nice to your protein. Remember that proteins make an excellent food source for microbes. Protect your sample from microbes by storing the protein at less than 4°C and not leaving it for extended periods of time at room temperature.

When thawing a sample or mixing a lyophilized sample into solution, do not shake or vortex the protein. Avoid foaming the sample. Foam can be a sign of denaturation.

Allow the sample to equilibrate to the temperature where the crystallization experiments will be set up and/or incubated before setting the experiment.

The field of crystal growth is full of opinions and controversy. There are several opinions on what should be done with the sample just prior to setting the crystallization experiment. Let's have a look at those opinions.

Some like to filter the sample through a 0.2 micron (or smaller, but be sure to compare the MW of your sample to the pore size of your filter so as not to stick your sample on the filter) pore size filter into a sterile container. Filtration can remove microbial contamination (but not the proteases) as well as sample aggregation. Turbid sample solutions with lots of precipitate should be solubilized or centrifuged before filtration to avoid the ugly experience of sticking the sample to a filter membrane. Use filters with the smallest possible dead volume to minimize sample loss. Some of the centrifugal microfiltration devices are certainly worth consideration. Before filtering the sample, wash or flush the filter with a small amount of the sample buffer/storage solution. This will test the filter for compatibility with your sample buffer and remove any trace glycerol which can sometimes be present from the manufacturer. If possible, test filter a small aliquot of the sample, and measure the activity/OD before filtering the entire sample. Do this to test the adherence of the sample to the filter media. Read and follow the instructions supplied with the filter before introducing the sample to the filter.

Some like to centrifuge the sample. Centrifugation removes large sample aggregates and amorphous debris. Post centrifugation views can provide a visual clue of aggregation/precipitate for seemingly clear solutions. Following centrifugation, use only the supernatant for crystallization trials.

Others prefer to avoid filtration or centrifugation before setting crystallization experiments. One view is that the presence of amorphous material or aggregates can enhance the changes for crystallization by acting as nucleants.

To Azide or Not

Sodium azide (NaN^3) is an anti-microbial preservative that is sometimes used to protect samples and crystallization reagents from microbial contamination. Sodium azide is toxic and should be handled with care. Typical Sodium azide concentrations are 1 mM or if you prefer % measurements, between 0.02% and 0.1% w/v.

If you choose to use Sodium azide remember that:

- It is toxic to humans as well as microbes.
- It is an inhibitor for some proteins and may become an unintentional ligand for your sample.
- It can interfere with heavy atom derivatization.
- Some metal azides are explosive.
- There are reports where eliminating Sodium azide from the experiment improved crystallization.

Alternatives to Sodium Azide include Thymol & Thimerosal.

A final alternative to the use of antimicrobials is the use of proper sterile technique and materials. Sterile filter all samples and reagents into sterile containers. Store samples and reagents at 4°C or lower. Use sterile pipet tips. Keep your work area clean. Develop a sterile technique with your crystallization setups. With common sense, sterile reagents and sample, good technique, and sterile pipet tips, one can successfully avoid the use of chemical antimicrobials in the crystallization lab.

Label & Organize Samples

Label samples clearly with the sample identification, batch identification, and date of storage. Small cryo labels can be very useful here. Color coding samples can be a nice organization tool. For the sake of easy organization and identification, it is sometimes more convenient to nest samples. For example, store batches of small microcentrifuge tubes in 10 ml or 50 ml centrifuge tubes and organize them by batch or sample.

It is prudent to write down and hold onto detailed notes concerning the purification, storage, and handling of the sample. It is obvious that one should also maintain records of crystallization trials which should include:

- Sample information
- Name of sample
- Sample identification (batch, storage location, storage temperature, etc.)
- Sample buffer composition, additives, ligands, etc.
- Sample concentration
- Crystallization experiment information
- Method
- Drop size and composition
- Reagent composition
- Temperature
- Date
- Name of person performing experiment

Questions to Ponder about the Sample

- Does a similar sample exist and has it been crystallized?
- Does the sample contain free cysteines?
- Does the sample contain additives such as Sodium azide, ligands, inhibitors, or substrates?
- Is the protein glycosylated?
- Is the protein phosphorylated?
- Is the protein N-terminal methylated?
- At what temperature is the protein stable?
- How does sample solubility and stability change temperature?
- How does sample solubility and stability change with pH?
- Does the sample bind metals?
- Is the protein sensitive to proteolysis?
- What class of protein am I working with (antibody, virus, enzyme, membrane protein)?
- What have been the most successful approaches with my class of protein?
- What is the source of the sample?
- How was the sample purified and stored before it arrived into my hands?
- What is in the sample container besides the sample (buffer, additives, etc.)?
- Is the sample pooled purification aliquots or a single batch?
- How much sample do I have and how much more is available?
- How pure is the sample?
- How homogeneous is the sample?
- Does anyone possess any solubility information on this sample?
- What is unique about this protein?
- What is necessary chemically and physically to maintain a stable, active sample?

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

There are several techniques for setting up crystallization experiments (often termed "trials"), including sitting drop vapor diffusion, hanging drop vapor diffusion, sandwich drop, batch, microbatch under oil, dialysis, and free interface diffusion. Here we offer an overview of these crystallization techniques.

Sitting & Hanging Drop Crystallization

Sitting and hanging drop methodologies are very popular because they are easy to perform, require a small amount of sample, and allow a large amount of flexibility during screening and optimization.

Using the sitting drop technique (figure 1), one places a small (0.1 to 40 µl) droplet of the sample mixed with crystallization reagent on a platform in vapor equilibration with the reagent. The initial reagent concentration in the droplet is less than that in the reservoir. Over time the reservoir will pull water from the droplet in a vapor phase such that an equilibrium will exist between the drop and the reservoir. During this equilibration process, the sample is also concentrated, increasing the relative supersaturation of the sample in the drop.

figure 1

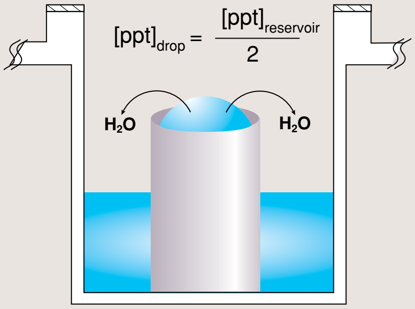

The advantages of the sitting drop technique include speed and simplicity. The disadvantages are that crystals can sometimes adhere to the sitting drop surface making removal difficult. This disadvantage can turn into an advantage where occasionally the surface of the sitting drop can assist in nucleation. The sitting drop is an excellent method for screening and optimization. During production, if sticking is a problem, sitting drops can be performed in the Sandwich Box Setup.

Sitting drop crystallization may be performed using Micro-Bridges® or Glass Sitting Drop Rods™ with VDX™ or Linbro® plates. Both plates can be sealed with clear sealing tape or plain cover slides for easy viewing and access. Sitting drop crystallization may also be preformed using the Cryschem™ Plate. The Cryschem Plate is a specially designed plate with a post already in the center of the reservoir. Finally, sitting drop crystallization can also be performed using one of the 48 or 96 well plates made by Art Robbins Instruments (Intelli-Plate™), Corning (CrystalEX™), Douglas Instruments (CrystalClear Strips™), Greiner (CrystalQuick™), or Swissci (MRC).

Using the hanging drop technique (figure 2), one places a small (0.1 to 20 µl) droplet of the sample mixed with crystallization reagent on a siliconized glass cover slide inverted over the reservoir in vapor equilibration with the reagent. The initial reagent concentration in the droplet is less than that in the reservoir. Over time the reservoir will pull water from the droplet in a vapor phase such that an equilibrium will exist between the drop and the reservoir. During this equilibration process, the sample is also concentrated, increasing the relative supersaturation of the sample in the drop.

figure 2

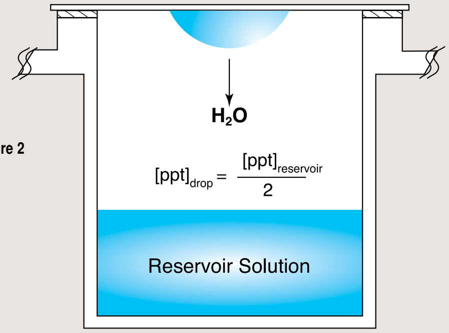

The advantages of the hanging drop technique include the ability to view the drop through glass without the optical interference from plastic, flexibility, reduced chance of crystals sticking to the hardware, and easy access to the drop. The disadvantage is that a little extra time is required for setups.

Sandwich Drop Crystallization

The sandwich drop crystallization method is illustrated in figure 3. The sample solution that is mixed with the precipitant is placed in the middle of a lower siliconized glass cover slide. A siliconized glass cover slide is then set in position along an upper edge. This allows for a small amount of space between the cover slides but is close enough so the drop is sandwiched between them. This technique offers an alternate equilibration method. However, the set up can be tedious and the plates designed for this method typically have a larger footprint than most.

figure 3

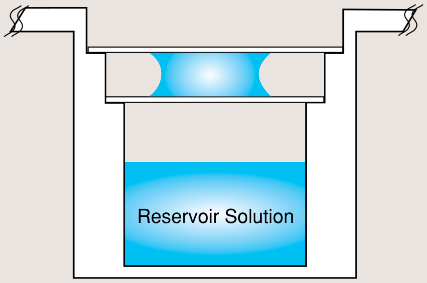

Free Interface Diffusion

Free interface diffusion crystallization is used less frequently than sitting or hanging drop vapor diffusion but it is one of the methods used by NASA in microgravity crystallization experiments and at least one company has automated and miniaturized the method. Using this method, one actually places the sample in liquid contact with the precipitant. When doing so, one attempts to create a clearly defined interface between the sample and the precipitant. Over time the sample and precipitant diffuse into one another and crystallization may occur at the interface, or on the side of high sample/low precipitant or low sample/high precipitant. The technique allows one to screen a gradient of sample precipitant concentration combinations. The technique can readily be performed in small

figure 4

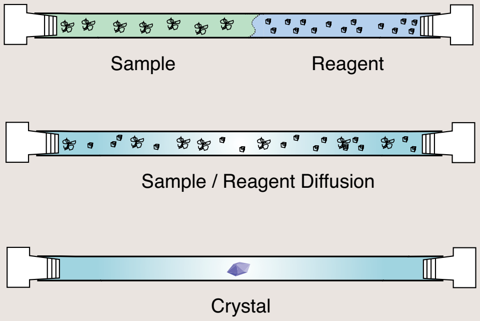

capillaries (figure 4).

Batch

Batch crystallization is a method where the sample is mixed with the precipitant and appropriate additives, creating a homogeneous crystallization medium requiring no equilibration with a reservoir. The technique is popular with small molecule crystallographers. The advantages to the technique are speed and simplicity but the disadvantage is that only a narrow space of precipitant/sample concentration can be sampled in a single experiment. A batch experiment can be readily performed in a capillary, small container, or plate with a small reservoir. One must be very close to the conditions which promote crystal growth in order for this technique to be successful.

Microbatch Under Oil

In this technique, a small drop of the sample combined with the crystallization reagent is pipetted under a layer of oil. For a true microbatch, the drop is placed under paraffin oil (figure 5) which allows little to no evaporation, nor concentration in the drop. A modified microbatch can be performed when the drop is placed under a mixture of paraffin oil and silicon oil, or straight silicon oil (figure 6). Such oils permit water vapor to permeate from the drop and allow sample and reagent to concentrate. Unless the drop is equilibrated with a reservoir, water will leave the drop until only solids remain.

The benefits of crystallization under oil include the use of very small sample and reagent volumes with less concern for unwanted evaporation, the minimization of surface interaction with the sample, the ability to precisely control sample and reagent concentrations during the experiment, and the minimization of condensation during temperature fluctuations.

figure 5

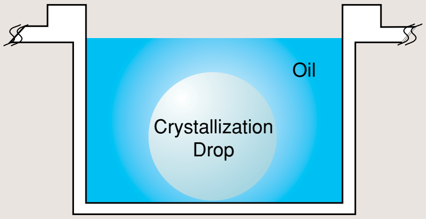

figure 6

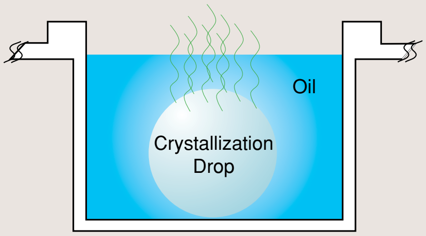

Dialysis Crystallization

Dialysis crystallization involves placing the sample in a Dialysis Button™ which is sealed with a dialysis membrane. Water and some precipitants are then allowed to exchange while retaining the sample in the cell. The Dialysis Button is placed into a suitable container holding the precipitant or crystallization media (figure 7). For example, one might dialyze a sample requiring a high ionic strength for solubility against a solution of low ionic strength. The technique allows for salting in and salting out, as well as pH crystallization techniques.

figure 7

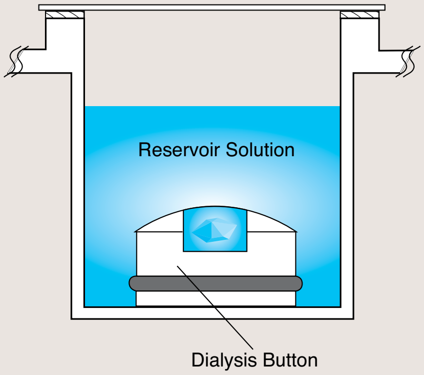

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

HANGING DROP VAPOR DIFFUSION CRYSTALLIZATION

Solutions for Crystal Growth

The hanging drop vapor diffusion technique is a very popular method for the crystallization of macromolecules. The principle of vapor diffusion is straightforward. A drop composed of a mixture of sample and reagent is placed in vapor equilibration with a liquid reservoir of reagent. Typically the drop contains a lower reagent concentration than the reservoir. To achieve equilibrium, water vapor leaves the drop and eventually ends up in the reservoir. As water leaves the drop, the sample undergoes an increase in relative supersaturation. Both the sample and reagent increase in concentration as water leaves the drop for the reservoir. Equilibration is reached when the reagent concentration in the drop is approximately the same as that in the reservoir.

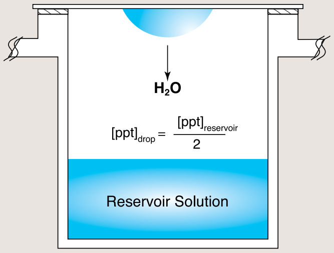

figure 1
Process of Vapor Diffusion

Benefits of Hanging Drop Crystallization

- Can be cost-effective.
- Sample and reagents in contact with a siliconized glass surface.
- Easy access to crystals.
- Can perform multiple drops (experiments) with a single reservoir.
- Good optics for viewing experiments.

Using the VDX™ Plate

The VDX Plate is a 24 well plate manufactured from clear polystyrene. It is typically sealed with vacuum grease (HR3-510) and siliconized 22 mm circle or square glass cover slides. The VDX Plate is also available with sealant which is a convenient time saver. Rows of the plate are labeled A-D and columns are labeled 1-6.

1. Apply a bead of vacuum grease along the top edge of the raised reservoir A1 of the VDX Plate. It is recommended that one apply the vacuum grease prior to pipetting the reagent. Create a circular bead on the upper edge of the reservoir. Do not complete the circle. Leave a 2 mm opening between the start and finish of the circular bead. Apply the cover slide, press to relieve the air pressure and twist to close the gap. Or simply use the VDX Plate with sealant. These plates come pregreased.

figure 2

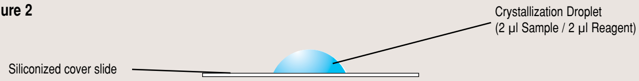

2. Pipet 1.0 ml of crystallization reagent into reservoir A1 of the VDX Plate. *(Note: Recommended reservoir volume is 0.5 to 1.0 ml)*
3. Clean a siliconized 22 mm circle or square cover slide by wiping it with lens paper and blowing it with clean, dry compressed air. Pipet 1 µl of sample into the center of a siliconized 22 mm circle or square cover slide. *(Note: Recommended total drop volume is 0.1 to 20 µl)*
4. Pipet 1 µl of reagent from reservoir A1 into the drop on the cover slide containing the sample. *(Note: Some prefer to mix the drop while others do not. Proponents of mixing leave the pipet tip in the drop while gently aspirating and dispensing the drop with the pipet. Mixing ensures a homogeneous drop and consistency drop to drop. Proponents of not mixing the drop simply pipet the reagent into the sample with no further mixing.)*
5. Holding the cover slide with forceps, the PEN-VAC®, Cover Slide Vacuum Gadget™, or on the edge between your thumb and forefinger, carefully yet without delay, invert the cover slide so the drop is hanging from the cover slide.
6. Position the cover slide onto the bead of grease on reservoir A1. Gently press the slide down onto the grease and twist the slide 45° to ensure a complete seal.
7. Repeat for reservoirs 2 through 24.

VDX Plate Tips

- Note the VDX Plate has a raised cover to protect the cover slides during transport and storage.
- To access a drop and/or reservoir, simply grasp the edge of the cover slide with forceps or fingertips, twist and pull gently.
- VDX Plates can be stacked for convenient storage.
- One can pipet multiple drops onto the cover slide. This technique is often useful when screening additives since one can use the same reservoir with multiple drops with each drop containing a different additive. This technique can also be used to screen different drop sizes and ratios versus the same reservoir. Use care to avoid mixing the drops during pipetting, plate transport, and plate viewing.
- The 0.96 mm Thick glass cover slides or the plastic cover slides are very durable and most tolerant to rough handling.

Plates for Hanging Drop Crystallization

- VDX™
- VDX with sealant
- VDXm™
- VDXm with sealant
- VDX48™ with sealant
- Modular VDX™
- Linbro®
- Greiner ComboPlate™

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

SITTING DROP VAPOR DIFFUSION CRYSTALLIZATION

The sitting drop vapor diffusion technique is a popular method for the crystallization of macromolecules. The principle of vapor diffusion is straightforward. A drop composed of a mixture of sample and reagent is placed in vapor equilibration with a liquid reservoir of reagent. Typically the drop contains a lower reagent concentration than the reservoir. To achieve equilibrium, water vapor leaves the drop and eventually ends up in the reservoir. As water leaves the drop, the sample undergoes an increase in relative supersaturation. Both the sample and reagent increase in concentration as water leaves the drop for the reservoir. Equilibration is reached when the reagent concentration in the drop is approximately the same as that in the reservoir.

figure 1
Process of Vapor Diffusion

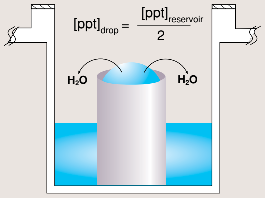

Benefits of Sitting Drop Crystallization

- Can be cost-effective.
- Can be time-efficient.
- Often easier when using detergents, organics and hydrophobic reagents.
- Drops can be positioned in a stable sitting position.
- Compatible with gels.

Using the Cryschem™ Plate

The Cryschem Plate is a 24 well plate manufactured from clear polystyrene. Each well contains a post in the center which is elevated above the bottom of the reservoir. The smooth, concave depression in the post can hold up to 40 µl drops and the reservoir can hold up to 1.2 ml of reagent. The Cryschem Plate is sealed with either clear sealing tape or film, or plain 22 mm circle or square glass cover slides. Rows are labeled A-D and columns are labeled 1-6.

1. Pipet 0.7 ml of crystallization reagent into reservoir A1 of the Cryschem Plate. *(Note: Recommended reservoir volume is 0.5 to 1.0 ml)*
2. Pipet 1 µl of sample into the post of reservoir A1. *(Note: Recommended total drop volume is 0.1 to 40 µl)*
3. Pipet 1 µl of reagent from reservoir A1 into the drop in post A1. *(Note: Some prefer to mix the drop while others do not. Proponents of mixing leave the pipet tip in the drop while gently aspirating and dispensing the drop with the pipet. Mixing ensures a homogenous drop and consistency drop to drop. Proponents of not mixing the drop simply pipet the reagent into the sample with no further mixing.)*
4. Repeat steps 1 through 3 for the remaining 23 reservoirs.
5. Seal the Cryschem Plate with clear sealing tape or film.

figure 2

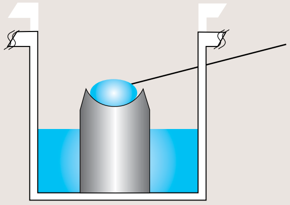

Crystallization Droplet
(2 µl Sample / 2 µl Reagent)

Cryschem Plate Tips

- Use Crystal Clear Sealing Tape to seal two rows at a time.
- To access a drop and/or reservoir of a Cryschem Plate sealed with tape, simply make a circular incision in the tape using the inside of the reservoir as a guide. Use a sharp blade to cut the tape and hold the incised piece of tape with forceps. The opening can be sealed with another strip of tape or a plain 22 mm circle or square glass cover slide and vacuum grease.

Using Micro-Bridges®

The Micro-Bridge is a small bridge (inverted U) manufactured from clear polystyrene or clarified polypropylene which contains a smooth, concave depression in the center of the top region of the bridge (figure 3). The Micro-Bridge can hold up to 40 µl drops. It is inserted into the reservoirs of VDX™ or Linbro® plates to perform a sitting drop vapor diffusion experiment. The design is such that the bridge is quite stable in the reservoir and does not require the Micro-Bridge to be bonded to the plate. The Micro-Bridge can be removed from the plate for crystal manipulation and observation if desired.

1. Pipet 1.0 ml of crystallization reagent into reservoir A1 of a VDX plate. *(Note: Recommended reservoir volume is 0.5 to 1.0 ml)*
2. Place a clean (blow the Micro-Bridge with clean, dry compressed air before use) Micro-Bridge into the bottom of reservoir A1 such that the concave depression in the Micro-Bridge is facing up.
3. Pipet 1 µl of sample into the Micro-Bridge in reservoir A1. *(Note: Recommended total drop volume is 0.1 to 40 µl)*
4. Pipet 1 µl of reagent from reservoir A1 into the drop in the Micro-Bridge A1. *(Note: Some prefer to mix the drop while others do not. Proponents of mixing leave the pipet tip in the drop while gently aspirating and dispensing the drop with the pipet. Mixing ensures a homogenous drop and consistency drop to drop. Proponents of not mixing the drop simply pipet the reagent into the sample with no further mixing.)*
5. Repeat steps 1 through 3 for the remaining 23 reservoirs.
6. Seal the plate with clear sealing tape or sealant and plain glass cover slides.

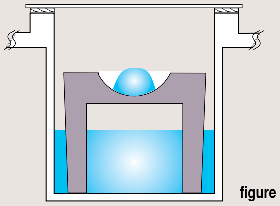

figure 3

Micro-Bridge Tips

- To access a drop and/or reservoir sealed with tape, simply make a circular incision in the tape using the inside of the reservoir as a guide. Use a sharp blade to cut the tape and hold the incised piece of tape with forceps. The opening can be sealed with another strip of tape or a plain 22 mm circle or square glass cover slide and vacuum grease.
- Micro-Bridges can be removed for crystal seeding, mounting, manipulation, and observation.
- Micro-Bridges are designed as disposable devices. It is not recommended to wash and re-use Micro-Bridges.
- Micro-Bridges cannot be siliconized or autoclaved.

SITTING DROP VAPOR DIFFUSION CRYSTALLIZATION

Using Glass Sitting Drop Rods™

The Glass Sitting Drop Rod is a small, solid rod manufactured from clear glass that has a smooth, concave depression in the center of the top region (figure 4). The opposite end of the glass rod is flat and it can hold up to a 100 µl drop. It is inserted into the reservoirs of VDX™ plates to perform a sitting drop vapor diffusion experiment. The Glass Sitting Drop Rod can be secured to the bottom of the plate using vacuum grease or can be left unattached to the bottom of the plate. It can be removed from the plate for crystal manipulation and observation if desired.

1. Pipet 1.0 ml of crystallization reagent into reservoir A1 of a VDX plate. *(Note: Recommended reservoir volume is 0.5 to 1.0 ml; Additional Note: If vacuum grease will be used to secure the Glass Sitting Drop Rod to the plate, apply the grease to the rod and insert the rod prior to pipetting reagent into the reservoir.)*
2. Place a clean (blow the Glass Sitting Drop Rod with clean, dry compressed air before use), siliconized glass rod into the bottom of reservoir A1 such that the concave depression in the Glass Sitting Drop Rod is facing up.
3. Pipet 1 µl of sample into the Glass Sitting Drop Rod in reservoir A1.
4. Pipet 1 µl of reagent from reservoir A1 into the drop in the glass rod A1. *(Note: Some prefer to mix the drop while others do not. Proponents of mixing leave the pipet tip in the drop while gently aspirating and dispensing the drop with the pipet. Mixing ensures a homogenous drop and consistency drop to drop. Proponents of not mixing the drop simply pipet the reagent into the sample with no further mixing.)*
5. Repeat steps 1 through 3 for the remaining 23 reservoirs.
6. Seal the plate with two strips of clear sealing tape or a plain glass cover slide and sealant.

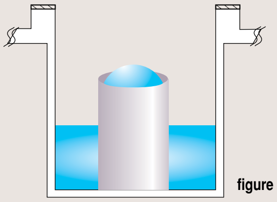

figure 4

Glass Sitting Drop Rods Tips

- To access a drop and/or reservoir of a plate sealed with tape, simply make a circular incision in the tape using the inside of the reservoir as a guide. Use a sharp blade to cut the tape and hold the incised piece of tape with forceps. The opening can be sealed with another strip of tape or a plain 22 mm circle or square glass cover slide and vacuum grease.
- Glass Sitting Drop Rods can be removed for crystal seeding, mounting, manipulation, and observation.
- Siliconize Glass Sitting Drop Rods before use.
- Glass Sitting Drop Rods may be washed and used over and over again. However, if vacuum grease is used to secure the Glass Sitting Drop Rod to the bottom of the reservoir, we wish you good luck in completely removing the grease from the rod.
- Glass Sitting Drop Rods may be autoclaved.

CrystalClear Strips™

CrystalClear Strips consist of a plastic frame and 12 polystyrene strips with a total experiment capacity of 96 per plate. Each strip contains 8 reservoirs and platforms. Typical reservoir volumes are 100 µl. The smooth, concave depression on the platform above the reservoir can hold up to a 10 µl drop. The CrystalClear Strip is sealed with clear sealing tape or film.

Pipet 100 µl of crystallization reagent in each of the 96 reservoirs. Pipet 1 µl of sample into the depression on the ledge of the first reservoir A1. Pipet 1 µl of reagent from reservoir A1 into the drop on the ledge above reservoir A1. Repeat steps 2 and 3 for the remaining 95 reservoirs and wells. Seal the CrystalClear Strips with clear sealing tape or film.

CrystalClear Strips Tips

- The strips are available with and without a concave depression for drop placement.
- While pipetting reagents into the reservoirs, place a clean pipet tip into the first empty reservoir. Move the clean tip ahead one empty reservoir with each reagent addition. This will help one keep track of their pipetting position in the plate.

Sandwich Box™

The Sandwich Box consists of a square polystyrene box, a plastic support, and a siliconized 9 well glass plate. The Sandwich Box is used when a common dehydrant system is desired as well as very large drops. Enormous drops can be pipetted into the siliconized glass wells. The siliconized glass plates offer excellent optics and can be removed from the plastic box to inspect the drop for birefringence without optical interference from plastic. Sandwich Boxes offer unique vapor equilibration kinetics and are very easy to access for crystal seeding, manipulation, and mounting. The plates are often used for heavy atom screening and derivatization and are useful for long-term crystal storage when each well is sealed with a glass slide and vacuum grease.

Open the Sandwich Box and place a plastic support, bottom side facing up into the box. Apply a bead of vacuum grease to the outer top edge of the box or the outer lower edge of the lid. Pour 25 ml of crystallization reagent or common dehydrant into the Sandwich Box. Place the siliconized 9 well glass plate on top of the inverted plastic support. Pipet the sample into one of the nine wells. Add the appropriate crystallization reagent to each drop. Place the cover on the Sandwich Box.

Sandwich Box Tips

- Apply a thin bead of vacuum grease around a single depression of the glass plate and seal the depression with a plain glass cover slide for long term crystal storage.
- Use a siliconized glass depression plate to test a small amount of sample for solubility with various crystallization reagents.

96 Well Plates

The 96 well sitting drop plates offer a variety of drop well configurations and flexibility in a standard microplate footprint. The 8 x 12 reagent wells in 9 mm offset can be filled with automated liquid handling systems or manual, single, and multichannel pipets with a typical reagent fill volume of up to 100 µl. The diversity of the various sample drop wells allow for automated and manual pipetting into a variety of well shapes, volumes and geometries. Materials range from optically clear polystyrene to low birefringent polymers. The plates can be manually or automatically sealed with optically clear sealing tape or film.

Plates for Sitting Drop Crystallization

- Cryschem™
- VDX™ with Micro-Bridges® or Glass Sitting Drop Rods™
- Linbro® with Micro-Bridges or Glass Sitting Drop Rods
- Greiner ComboPlate™ with CrystalBridge™
- Douglas Instruments CrystalClear Strips™
- Sandwich Box™
- Intelli-Plate™
- Corning Crystal*EX* ™
- Greiner CrystalQuick™
- Douglas Instruments Vapor Batch
- Swissci MRC

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

CRYSTALLIZATION UNDER OIL

Method

Crystallization under oil is a method where a small drop of sample combined with the crystallization reagent of choice is pipetted under a layer of oil. This is also known as microbatch crystallization.

Oils can also be used as a barrier between the reservoir and the drop in traditional hanging or sitting drop crystallization experiments. This is known as vapor diffusion rate control.

Description of Microbatch

The crystallization of proteins under a thin layer of paraffin oil was originally described by Chayen et al. (Appl. Cryst. 23 (1990) 297). In this technique, a small drop of sample combined with the reagent of choice is pipetted under a small layer of paraffin oil (figure 1). The oil generally used is a mineral oil of branched paraffins in the C20+ range and allows for little to no diffusion of water through the oil. Essentially all of the reagents in a batch or microbatch experiment are present at a specific concentration. Furthermore, there is no significant concentration of either the protein or the reagent within the crystallization drop.

figure 1

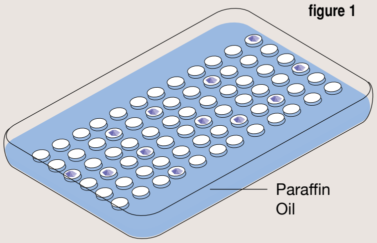

Description of Modified Microbatch

D'Arcy et al. (A novel approach to crystallizing proteins under oil, Journal of Crystal Growth 168 (1996) 175-180) modified the microbatch under oil technique by using silicone fluids which are polymeric compounds composed of repeating dimethylsiloxane units -(Si(CH$_3$)$_2$-O-)$_n$-. Using a mixture of 1:1 silicon oil and paraffin oil, also known as Al's Oil, one can perform a microbatch experiment under oil and have diffusion of water from the drop through the oil, hence a microbatch experiment that does allow for concentration of the sample and the reagents in the drop (figure 2).

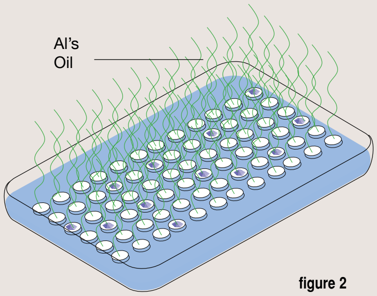

figure 2

Performing Microbatch / Modified Microbatch

Pipet 6 ml of 100% paraffin oil or 6 ml of 1:1 paraffin/silicon oil (Al's Oil) into a 72 Well Microbatch Plate, as shown in figure 1 or 2. *Note: One can also utilize other ratios of paraffin oil and silicon oil to vary the rate of diffusion from the drop (higher % of silicon oil = more rapid diffusion and evaporation).*

Pipet the sample into the appropriate cone-shaped depression in the microbatch plate followed by addition of the reagent. Typical drop ratios and final drop sizes are 1:1 and 1 to 2 µl. Drops up to 10 µl can be achieved under oil using the microbatch plate. Place plate cover over microbatch plate to prevent dust and debris from entering experiment. The microbatch method can also be performed in round bottom, clear 96 well plates. The oil, reagent, and sample can be mixed quickly and efficiently for fast throughput by using an 8 or 12 channel pipetter.

Description of Vapor Diffusion Rate Control

Chayen (A novel technique to control the rate of vapour diffusion, giving larger protein crystals J. Appl. Cryst 30 (1997) 198-202) has described a technique where oils can be used to vary the rate of vapor diffusion. Using mixtures of paraffin and silicon oil, Chayen reported fewer, larger crystals in the drop.

Using a standard hanging or sitting drop vapor diffusion setup, the drop is first mixed with reservoir solution, thus preventing oil from entering the drop. Then, 200 µl of oil is applied over the reservoir solution (figure 3). The oil acts as a barrier to vapor diffusion between the reservoir and the drop. Using 100% paraffin oil allows limited amount of vapor diffusion to occur, which in turn causes the drop to behave like a batch experiment. The drop will then eventually dry up due to evaporation through the polystyrene plate. Using 100% silicon oil will give results similar to that when no oil is used. When using a mixture of the two oils, the rate of vapor diffusion between the drop and the reservoir may be controlled. The rate of vapor diffusion is also a function of thickness of the oil layer over the reservoir. Chayen evaluated oil volumes between 100 and 700 µl. Oil volumes of 50 to 100 µl resulted in crystals similar to the control without oil. Oil volumes greater than 700 µl have a significant delay in the onset of crystallization, with improved crystal size. Results using hanging drop were more pronounced than sitting drop which may be due to either surface effects or the drop geometry in relation to the reservoir which could influence vapor diffusion kinetics.

figure 3

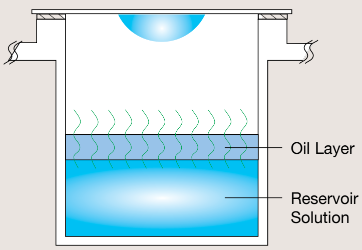

Performing Vapor Diffusion Rate Control

Prepare a VDX or Cryschem Plate for a sitting or hanging drop vapor diffusion experiment. After the reservoir has been added to the drop, pipet between 200 and 700 µl of a mixture of paraffin/silicon oil onto the reservoir (figure 3). Seal the plate.

Plates for Microbatch Crystallization

- Douglas Instruments Vapor Batch
- Greiner 72 Well Microbatch
- Greiner Imp@ct™
- Swissci MRC Under Oil 96 Well Crystallization Plate

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

Method

Crystallization by dialysis is an easy variation to the typical vapor diffusion method used to grow crystals. In the dialysis method, the sample in question is separated from the "precipitant" by a semi-permeable membrane which allows small molecules such as ions, additives, buffers, and salts to pass but prevents biological macromolecules from crossing the membrane. Equilibration kinetics depend upon the molecular weight cut-off of the dialysis membrane, the precipitant, the ratio of the volume, the concentration of the components inside and outside of the dialysis cell, and the geometry of the cell.

Description

The Dialysis Buttons™ offered by Hampton Research are machined from transparent perspex. The button has a chamber which varies from 5 to 350 µl depending upon which size button one chooses to use. The sample is placed in this chamber so as to create a slight dome of liquid at the top edge of the button. A dialysis membrane (having the appropriate molecular weight cut-off) is placed over the top of the button/sample and is held in place with an O-ring. The O-ring is held in place by a groove in the button. Dialysis Buttons are notoriously tricky to set up since beginners often trap air bubbles between the sample solution and the membrane which impedes dialysis. With a little practice using a golf tee or applicator, one can master the technique. Dialysis Buttons are supplied with O-rings and a golf tee. Dialysis membrane discs and applicators are available separately. The applicator is used to apply the membrane and the O-ring to the buttons.

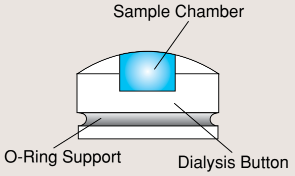
figure 1

Using the Dialysis Button

A typical dialysis experiment is used to take the sample from the presence of a high ionic strength solution to a lower ionic strength solution. However, the technique can just as easily be used to proceed from low ionic strength to a higher ionic strength. This is accomplished by placing the sample in high ionic strength into the Dialysis Button, sealing the button with a dialysis membrane, and placing the sealed button in a solution of ionic strength lower than that inside the button. Salts, ligands, and compounds smaller than the pore size of the dialysis membrane will leave the button as long as their concentration is lower on the opposite side of the membrane. Once the concentration of the diffusible species is the same on both sides of the membrane, the system is in equilibrium.

Cleaning

The buttons can be cleaned with soap and deionized water. Do not clean them with organic solvents as this may turn the perspex opaque.

Practical Example

The following two practicals offer examples of how to set up a dialysis experiment.

Practical 1 - Carboxypeptidase A

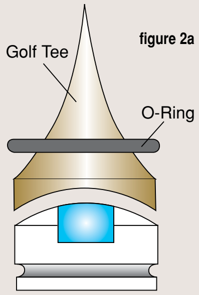
figure 2a

1. Using Carboxypeptidase A (Sigma-Aldrich CO386 or CO261), make an 8 to 20 mg/ml solution of the Carboxypeptidase A in 20 mM Tris HCl pH 7.5, 1.5 M LiCl.
2. Place 100 ul of 10 mg/ml carboxypeptidase in 20 mM Tris HCl, 1.5 M LiCl, pH 7.5 in a 100 ul Dialysis Button. The droplet should have a slight dome shape following the hemispheric edge of the top of the button.

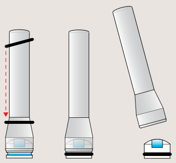
figure 2b

3. Seal the button with dialysis membrane. Using a one inch (2.5 cm) square of dialysis membrane which has equilibrated in water, place the membrane over the top of the button. Place an inverted golf tee on top of the membrane and button. Roll the O-ring down the applicator until the O-ring rolls off onto the edge of the button. Roll the O-ring into the machined groove on the edge of the button. Remove the applicator. There should be no bubbles between the membrane and the sample inside the button. Bubbles will prevent dialysis. Some researchers prefer to use the rubber tipped plunger of a syringe with a modified syringe body to apply the O-ring. Others prefer to use an applicator. Try each and see which method works best.
4. Place 0.9 ml of 20 mM Tris HCl, pH 7.5 in the reservoir of a VDX™ Plate, a Linbro® Plate, or small chamber which can be sealed.
5. Place the Dialysis Button in the well, membrane side up. Be sure the reservoir solution covers the top of the membrane/button. Seal the VDX plate using grease and a cover slide.
6. Observe under a microscope. Crystals will appear within 2 to 3 days. Final concentration of LiCl will be 0.15 M.

Reference for the above protocol: Dr. Jim Pflugrath and Dr. Gary Gilliland, Cold Spring Harbor Laboratory Protein Crystallography Workshop.

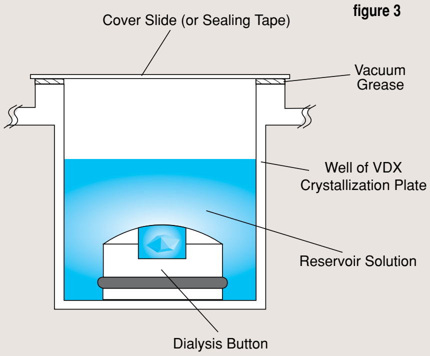
figure 3
Cover Slide (or Sealing Tape)
Vacuum Grease
Well of VDX Crystallization Plate
Reservoir Solution
Dialysis Button

Practical 2 - Lysozyme

1. Prepare 10 mg/ml lysozyme in 50 mM Sodium acetate trihydrate pH 4.5. Filter the solution using a 0.2 micron filter.
2. Fill a 100 µl Dialysis Button with 100 µl of the lysozyme solution as described for the carboxypeptidase practical.
3. Pipet 1 ml of 50 mM Sodium acetate trihydrate buffer into a small (5 ml) beaker.
4. Place the filled button, membrane side up in the beaker.
5. Pipet a small amount of concentrated Sodium chloride into the beaker such that the final concentration of Sodium chloride in the beaker is 0.2 M. Seal the beaker with parafilm and store at room temperature.
6. Increase the concentration of Sodium chloride each day by 0.2 M. Repeat until crystals are observed in the button.

Reference for the above protocol: Crystallization of nucleic acids and proteins, a practical approach. Edited by A. Ducruix and R. Giege, Oxford University Press, 1992.

Considerations

Just as in a vapor diffusion experiment, the path is often as important as the endpoint in a dialysis experiment. The path is the equilibration course which the solution inside and outside the button take toward achieving equilibrium. This course can be changed by manipulating the following:

- Ratio of button volume/reservoir volume
- Button and reservoir components & concentration
- Molecular weight cutoff of dialysis membrane
- Viscosity of solutions
- Plus the usual assortment of crystallization variables including pH, sample concentration, temperature, etc.

Variations of Dialysis

Macrodialysis

The sample is loaded into dialysis tubing of the appropriate molecular weight cutoff and is dialyzed against the appropriate reservoir solution. This method typically requires at least 100 µl of sample and can be performed with liters of sample in large dialysis tubing.

Zeppenzauer Cells

Capillary tubes are closed with dialysis tubing or gel plugs. See Zeppenzauer, M. 1971, Methods In Enzymology, 22, 253.

Microcap Dialysis

The sample is placed in a glass capillary with one end sealed with wax, the other with dialysis membrane. The tube is placed in a microcap/small centrifuge tube filled with the appropriate reagent. See Crystallization of nucleic acids and proteins, a practical approach, Edited by A. Ducruix and R. Giege, Oxford University Press, 1992.

Double Dialysis

This method reduces the rate of equilibration and can provide enhanced control over the crystallization of the sample. Simply put, a Dialysis Button is prepared and placed inside a reservoir sealed with a dialysis membrane, which is in turn placed inside another reservoir. Confused? See Thomas, D.H., et al., 1989, Protein Engineering, 2, 489.

References and Readings

1. *Crystallization of nucleic acids and proteins, Edited by A. Ducruix and R. Giege, The Practical Approach Series, Oxford Univ. Press, 1992.*
2. *Preparation and analysis of protein crystals. McPherson, A. Eur. J. Biochem. 189, 1-23, 1990.*
3. *Zeppenzauer, M. et al., Crystal. of horse liver alcohol dehydrogenase complexes from alcohol solutions. Acta Chem Scan, 21, 1099, 1967.*
4. *Christopher Bunick, A.C.T. North, and Gerald Stubbs., Evaporative microdialysis: an effective improvement in an established method of protein crystallization. Acta Cryst. D56, 1430-1431, 2000.*

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

HAMPTON RESEARCH

Solutions for Crystal Growth

VIEWING CRYSTALLIZATION EXPERIMENTS

figure 1
Typical observations in a crystallization experiment.

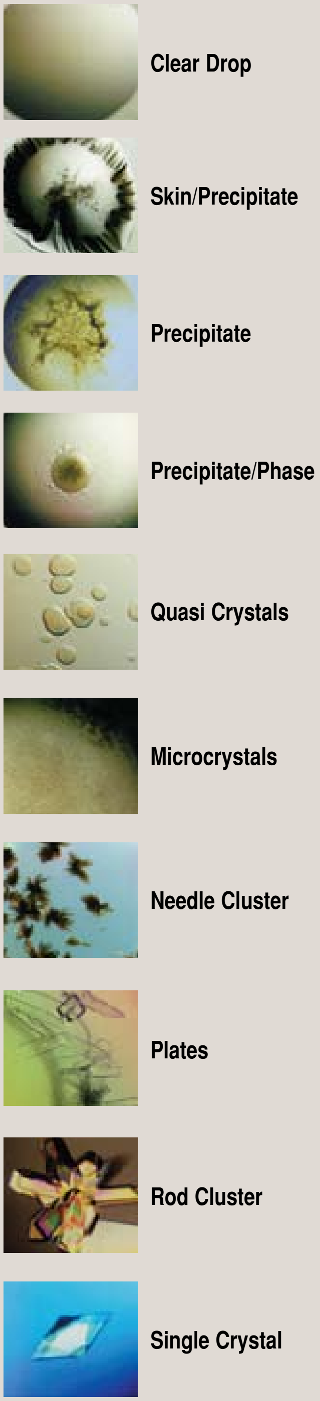

Clear Drop

Skin/Precipitate

Precipitate

Precipitate/Phase

Quasi Crystals

Microcrystals

Needle Cluster

Plates

Rod Cluster

Single Crystal

Observing the Experiment

Gently set the plate onto the observation platform. If the platform is smooth and free of protrusions, one may simply slide the plate in the X and Y directions on top of the viewing platform to view each of the drops. Use low magnification to view and center the drop in the field of view. Scan the drops at 20 to 40x magnification. When something suspicious appears, increase the magnification to 80 or 100x for a better view. Scan the entire depth of the drop using the fine focus control of the microscope. Sometimes crystals will form at different depths of the drop because different areas of the drop can equilibrate at different rates. Also, crystals sometimes form at the top of a drop and as the crystal gains mass, it will fall to a lower portion of the drop. Scrutinize everything until you are familiar with the differences between crystals, microcrystals, precipitate, and sweater fuzz. True crystals will feature edges. Precipitate does not have edges. Crystals can appear as needles, blades, walnuts, spherulites, plates, and various geometric shapes. Crystals vary in size anywhere from a barely observable 20 microns to 1 or more mm but most seem to fall in the range of 0.05 to 0.5 mm. Figure 1 shows typical examples of what one might observe in a crystallization experiment.

Diffractable Crystal

Crystals useful for x-ray diffraction analysis are typically single, 0.05 mm or larger, and free of cracks and defects.

Differentiating Between Microcrystals & Precipitate

Microcrystals (less than 0.02 mm) can be difficult to differentiate from precipitate, especially under low power or with a low to medium quality microscope. Differentiate microcrystals from amorphous precipitate by looking for birefringence (light colored shiny spots under a polarizer in dark field mode = crystals). Other tests to differentiate crystals from precipitate include streak seeding, or the use of a small amount (1 µl) of colored, low molecular weight, water soluble dye (crystal violet and methylene blue will often penetrate the solvent channels of macromolecular crystals and color them, where as precipitate will not be colored).

Precipitate can appear as clumps, fine wispy clouds, or anything in between and can range in color from white to yellow, brown, or rust. During screening and very preliminary optimization, one may wish to observe the drops immediately after setup, one day later, and each day thereafter for the first week. Observations may be performed once a week thereafter until the drops turn into a crust of deceased sample and reagent. Never throw plates away until the drop is dead. Why? Most crystallization plates are made from polystyrene which allows for some evaporation over time. Evaporation leads to increased relative supersaturation and over time, may also lead to an increase in crystals. Time also can lead to changes in the protein (denaturation of less stable forms, proteolytic cleavage, and other changes) which might promote crystallization. Take careful notes during observations and be especially conscious of changes that occur between observations. Most crystallization observations are done at room temperature since this is where one will find most microscopes and it is most comfortable.

Cold Experiments

4°C experiments may be observed by moving the microscope into a cold room. Allow time for the microscope to equilibrate to 4°C to prevent fogging of the optics as well as unnecessary temperature transfer from the warm microscope to the cold experiment. Wear a warm jacket with gloves to stay as comfortable as possible in the cold room. Excessive moisture in a cold room can be very destructive to a microscope so check with your maintenance group to keep the cold room as dry as possible. If a cold room is unavailable, one is forced simply to work fast, moving plates from an incubator to the microscope carefully and making rapid, yet thorough observations. Move only one plate at a time and gently close the incubator door between transfers since slamming the door will cause vibrations which can influence crystallization. Cold experiments tend to fog up rapidly, especially if the light source is hot (if no infrared filters or light pipes are used). This is difficult to avoid and is one reason researchers prefer working in cold rooms.

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

BIREFRINGENCE

The technical mumbo jumbo first. The physical properties of isotropic materials, such as glasses, liquids, and amorphous materials do not depend on direction. However, most properties of a wide variety of crystals (including liquid crystals) do show such variation. This anisotropy of physical properties originates in the anisotropic build-up of the materials (crystal structure). Anisotropy in the optical properties of uniaxial crystals is referred to as either birefringence or dichroism, depending on whether the index of refraction or the absorption coefficient is concerned. Birefringence means that there are two distinct speeds with which light can propagate, depending on the direction of propagation. When a light ray splits into two beams as it passes through a material, the effect is called birefringence (or double refraction) and the material is birefringent. If you look at something through a birefringent material, you'll see double. The word birefringence comes from the Latin bi- (twice) plus refringere (to break up). Thus, the light rays are “broken in two” by a birefringent material. One well-known example of a birefringent medium is crystalline calcite (calcium carbonate). If you look at the world through a clear crystal of calcite (calcium carbonate), you will see double. Place such a crystal on a drawing, and you'll see two overlapping copies of the drawing. The molecular structure of calcite causes double refraction, in which each light ray is split into two rays that emerge from the crystal at slightly different angles. Calcite shows this more clearly than most crystals, but quartz and many other crystalline minerals also split light ray.

Now, the practical interpretation for crystal growers. You might hear the word birefringence used quite often by crystal growers when viewing crystals under a microscope. Here, crystal growers are stretching the definition of the term birefringence to describe the colorful display produced by biological macromolecular crystals when polarized light is passed through the crystal. See figure 1.

figure 1

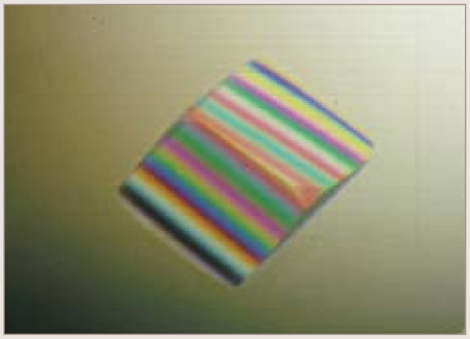

A light microscope with polarizing optics is required to observe birefringence. The following path is a typical setup where light passes from the light source through the first polarizing lens, next the specimen (crystal), then the second polarizing optic, finally the magnifying optics and into your eye. On many typical polarization setups, the second polarizing filter can be rotated while the specimen is stationary. Rotating the polarizing optic without something to rotate the plane of polarized light in the path (i.e. a crystal) will result in one seeing light, dark, light, dark, as the filter is rotated. But if a crystal with birefringent properties (i.e. a biological macromolecular crystal) is positioned in between the two polarizing filters, one will observe changing colors as the polarizing filter is rotated. Specifically, when the polarizing filters are aligned such that the field is dark, a birefringent object (crystal) will glow with color.

Birefringence is one way we can differentiate amorphous precipitate from microcrystals in a drop when viewed under a microscope. Precipitate does not have birefringent properties while most biological macromolecular crystals do.

One drawback with using birefringence in today's crystal growth world is that most of the crystallization devices utilized are made from plastic such as polystyrene and polypropylene. These plastics are optically active and can be birefringent. In fact, often times the colors we see displayed in crystals are contributions from the birefringent plastic. See figure 2. However, it is still possible to observe microcrystalline birefringence in the plastic trays, but there is usually a contributory effect from the plates used to grow the crystals. One way to avoid this is to grow crystals in a glass device or at least observe the crystals in a path that is free of birefringent plastics. Some crystallization plates are available in a low birefringent material.

figure 2

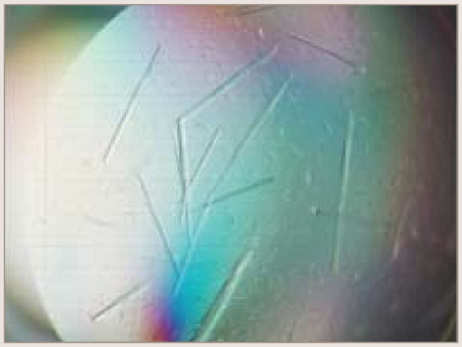

Birefringent precipitates will glow, sparkle, or glisten. See figure 3.

figure 3

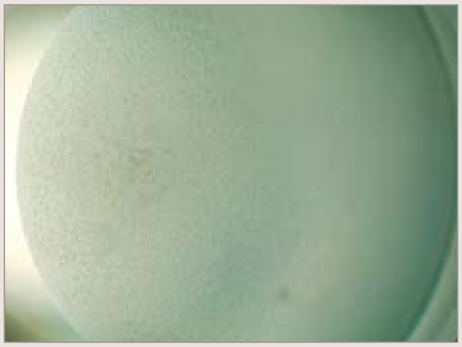

To test for birefringence, position the polarizers so the field of view is dark WITHOUT the crystallization plate or setup. Place the tray into position on the microscope. If a crystal is birefringent, some of the light passing through the crystal will be rotated and pass through the second (analyzing) polarizing filter. The intensity of the transmitted light will increase and decrease as the crystal or the polarizer is rotated. Remember, birefringence is NOT ALWAYS clearly visible when polystyrene or polypropylene is in the light path (i.e. when you use plastic slides or crystallization plates). However, a birefringent crystal viewed in most plastic trays or plastic cover slides will have a different color than the background (i.e. plastic plate) or foreground (plastic slide). Finally, birefringence is a property of crystals, both biological (proteins, peptides, and nucleic acids) and inorganic crystals (salts). Birefringence is MORE pronounced in inorganic crystals (salt).

A quick comment on what to do with birefringent precipitates. Streak seeding is a common and often successful method of taking advantage of microcrystalline precipitate to grow large single crystals.

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

SALT OR BIOLOGICAL CRYSTALS?

First of all, before deciding on how to distinguish whether the crystal is inorganic (salt) or biological (protein), take a picture of the crystal before you destroy it or make it disappear.

The most definitive test is to obtain an x-ray diffraction pattern of the crystal. A diffraction pattern of a protein crystal may look like that in figure 1. Want to take a less direct approach? Read on.

The following are ways to differentiate a biological crystal from an inorganic crystal.

1. Dehydration

A biological crystal typically has very significant solvent content and will dehydrate when removed from the drop or when the drop is allowed to evaporate from around the crystal (figure 2). Inorganic crystals typically do not possess large solvent channels and have very little solvent content. Removing an inorganic crystal from the drop or allowing the drop to evaporate from around the crystal will typically not destroy or change the appearance of the inorganic crystal.

2. Physical Manipulation

A biological crystal behaves more like an ordered gel than a hard crystal and will powder, crumble, or break easily when touched with a probe such as a Micro-Tool or needle (figure 3). Inorganic crystals can also break apart, but they require more force and typically make a click or crunching sound when breaking apart under the force of a probe. The inorganic crystals are typically more dense than protein crystals and once broken, the pieces fall quickly and stay put on the bottom of the drop.

3. Birefringence

A biological crystal, unless it is cubic, will be weakly birefringent under cross polarizers (figure 4). Inorganic crystals are typically strongly birefringent under cross polarizers. Some plastic plates and materials are also birefringent so this test is more easily performed and interpreted in an all-glass environment or in a plate made from a low birefringent plastic.

4. Dye absorption

A biological crystal typically has large solvent channels which will accommodate a small molecule dye. Small molecule dyes can travel into these solvent channels and color the crystal (figure 5). Inorganic crystals do not possess such solvent channels and will not absorb the small molecule dye. Dyes are chemicals and have solubility limits. So it is possible that a crystal grown in a reagent of high relative supersaturation may have a reagent concentration that will precipitate or even crystallize the dye. Most dyes under such conditions will crystallize into needles or whiskers and of course be colored. So before adding dye, take a picture or memorize the location of crystals in the drop in case the crystal itself forms crystals. Finally, diluting the drop with dye can sometimes decrease the relative supersaturation in the drop to the point where the biological crystal will dissolve. Once the drop equilibrates again with the reservoir, the crystal may reappear and it may appear in a location different from the original crystal. For dyes, consider Izit (HR4-710) from Hampton Research.

5. Control Experiment

Using the same sample buffer, same reagent, same volumes and same hardware, set an experiment identical to the one that produced the crystals, except leave the sample out of the experiment. Simply replace the sample with the sample buffer. Don't get lazy and use water instead of sample buffer because the sample buffer may be a variable in the formation of the crystals. If a crystal forms that appears visually similar to the original crystal you likely have an inorganic crystal in the original setup.

6. Run a Gel

Collect, wash, and dissolve the crystals. Run the sample on SDS-PAGE. If a band indicates the presence of your sample, there is a high probability that your original crystals are biological.

Hampton Research would like to thank Joe Luft at the Hauptman Woodward Medical Research Institute for his help in putting together the information about inorganic and biological crystals.

figure 1

X-ray diffraction pattern from a protein crystal.

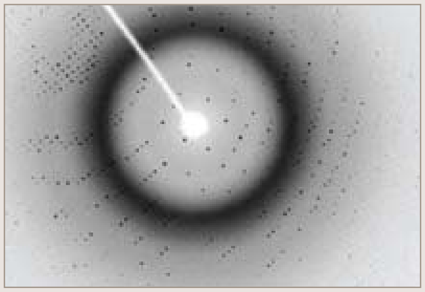

figure 2

Below is a happily hydrated protein crystal. On the next page is an unhappy, dehydrated protein crystal, removed from the drop.

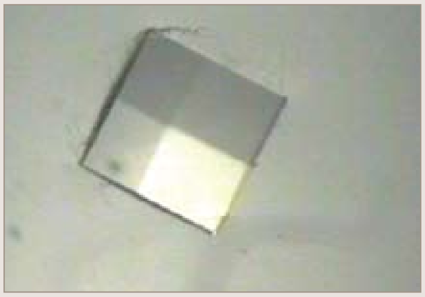

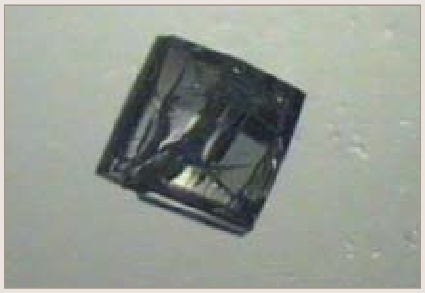

Dehydrated
Protein Crystal

figure 3

A protein crystal before and after being bullied by a probe.

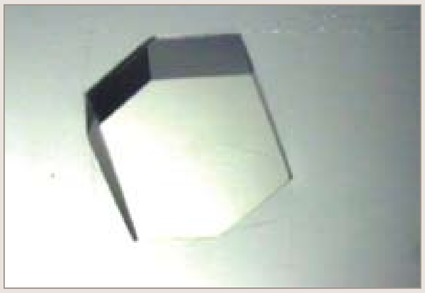

Before: Pre-Crush

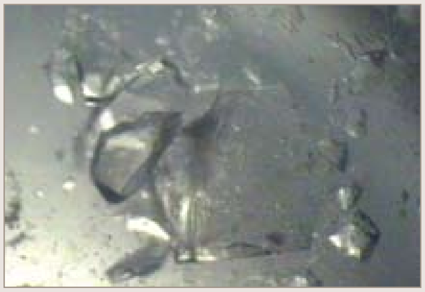

After: Crush

figure 4

A weakly birefringent biological crystal.

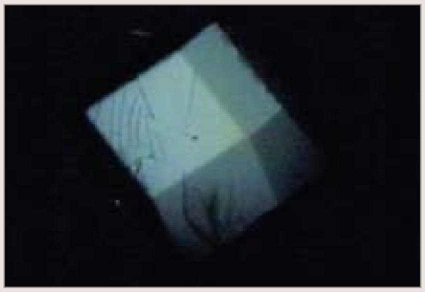

figure 5

Protein crystal stained with a dye.

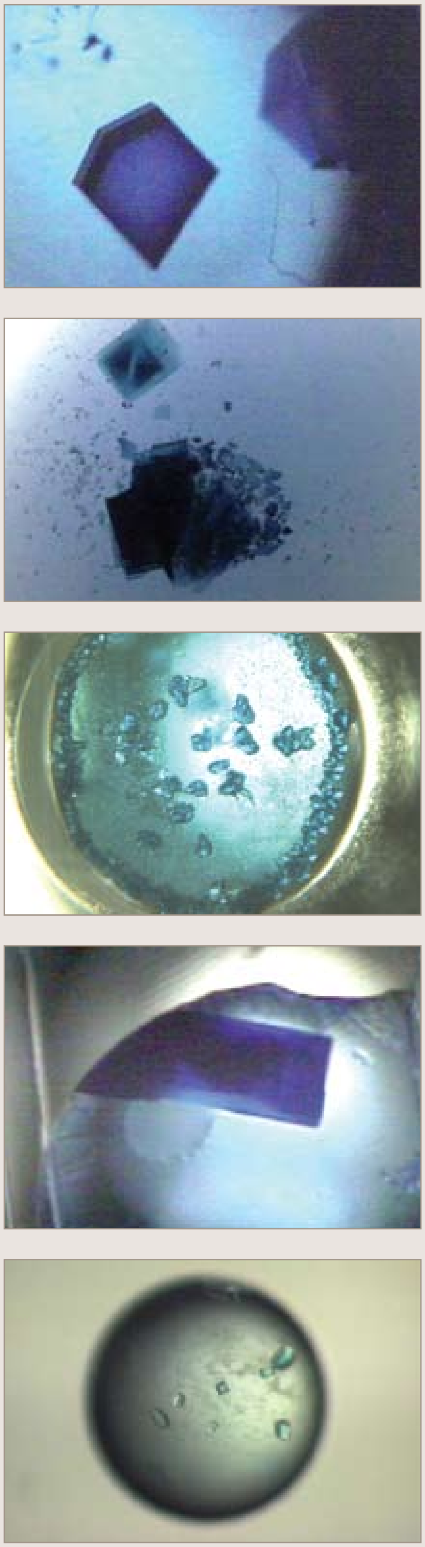

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

pH VERSUS NUMBER OF CRYSTALS

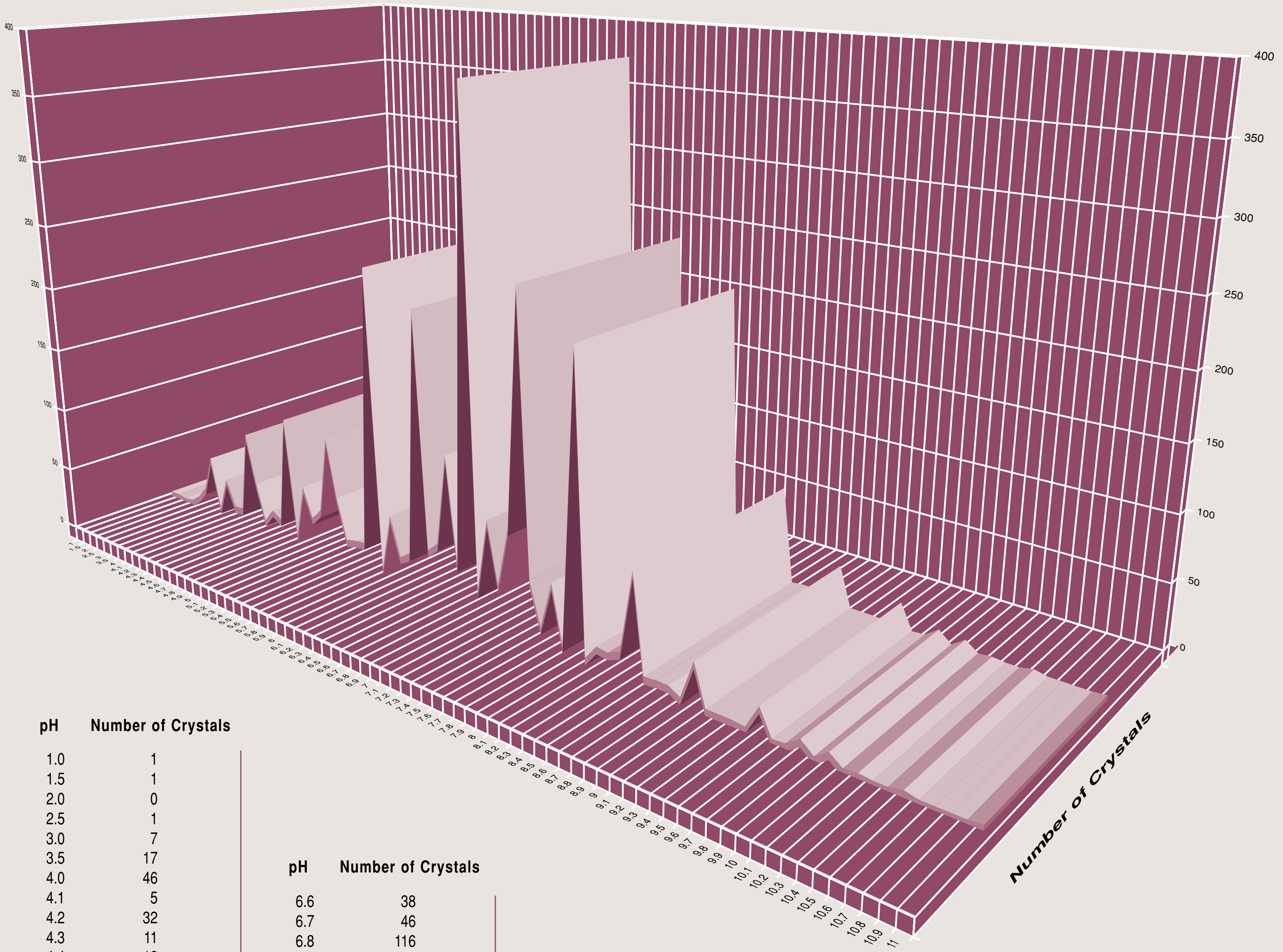

pH	Number of Crystals
1.0	1
1.5	1
2.0	0
2.5	1
3.0	7
3.5	17
4.0	46
4.1	5
4.2	32
4.3	11
4.4	12
4.5	78
4.6	39
4.7	10
4.8	21
4.9	16
5.0	102
5.1	8
5.2	51
5.3	28
5.4	37
5.5	98
5.6	53
5.7	21
5.8	23
5.9	24
6.0	239
6.1	9
6.2	54
6.3	24
6.4	29
6.5	217

pH	Number of Crystals
6.6	38
6.7	46
6.8	116
6.9	35
7.0	379
7.1	21
7.2	78
7.3	32
7.4	77
7.5	251
7.6	45
7.7	13
7.8	49
7.9	6
8.0	220
8.1	3
8.2	15
8.3	14
8.4	18
8.5	79
8.6	7
8.7	8
8.8	8
8.9	3
9.0	32

pH	Number of Crystals
9.1	2
9.2	2
9.3	2
9.4	1
9.5	19
9.6	0
9.7	1
9.8	8
9.9	0
10.0	7
10.1	0
10.2	1
10.3	0
10.4	0
10.5	2
10.6	0
10.7	0
10.8	0
10.9	0
11.0	1

pH versus number of crystals grown for 2,953 biological macromolecules reported from the BMCD.

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

pH POINTS TO PONDER

- Never use a pH probe containing AgCl when measuring the pH and/or titrating the pH of Tris and Tris HCl buffers since the AgCl will complex with the Tris which can lead to erroneous pH readings. Use a Tris-compatible pH probe when adjusting the pH of Tris buffers.
- Follow the manufacturer's recommended use and maintenance instructions to keep the pH probe accurate and precise.
- pH measurements are only as good as the probe.
- Glass electrode probes are more precise and accurate than gel-filled plastic probes.
- Calibrate pH probes each and every day. Refresh calibration standard solutions frequently to avoid contamination and evaporation.
- A two-point calibration with the "pH to be measured" sandwiched between the calibration points is better than a single or triple point calibration.
- Bacterial contamination of pH standards, storage solution, and water can cause pH fluctuations in reagents.

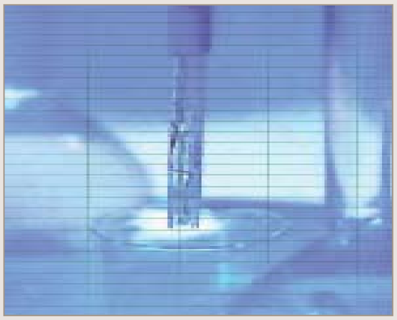

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

CRYSTAL GROWTH VIA pH RELAXATION

Try the following if sample solubility is pH dependent. Determine a pH and salt concentration where the protein is insoluble. A simple way to do this, if sample supply is limited, is to review old drops and look for those which contain precipitate. Review the data and select the reagent, reagent concentration, and pH most often associated with precipitate. Or, if sample is abundant, using a microscope, a microscope slide, or cover slide, empirically perform solubility tests by adding 1 µl of protein to 1 µl of buffer (and/or precipitant) and watch the drop for the formation of precipitate within a minute or so (not too long or the drop will evaporate and you will always observe precipitate!). Once reagent and pH points have been determined to promote a sample solubility minima, set hanging or sitting drop vapor diffusion experiments with a smattering of the solubility minima conditions. Once the sample has precipitated, add acetic acid (acidic) or ammonium hydroxide (basic) to the drop (try 1 µl of 0.1 M stock and adjust the amount and/or concentration of the stock accordingly) until the precipitate is redissolved. Seal the reservoir. The addition of the acid or the base will, in most experiments, solubilize the precipitate. Since acetic acid and ammonium hydroxide are volatile, the acidic/basic reagent will leave the drop. As the acidic/basic reagent leaves the drop, the pH of the drop will relax to approximately the original pH prior to the addition of the acidic/basic reagent. As the relaxation approaches the original pH, the protein will approach a point of solubility minima and here is when there is the chance for crystallization to occur. Variables to consider when refining the pH relaxation technique include initial pH, buffer concentration, buffer type, precipitant type, precipitant concentration, concentration and volume of acetic acid/ammonium hydroxide added to the drop, reservoir volume, and drop volume. One can also incorporate additives (organic solvents, chaotropes, detergents, ions, salts, etc.) into the procedure to evaluate the effect of the additive on sample-sample and sample-solvent interactions. Sound tedious? Well, it's not as bad as it sounds. Approach the technique as both a crystallization experiment and a crude solubility measurement. If no crystals are grown using the method, all is not lost. One has at least obtained a great deal of pH and reagent dependent solubility data which can be used to design additional crystallization experiments. What should one do with the old experiments full of precipitate and no crystals? Move the plates to warmer and then cooler temperatures. Note changes in sample solubility. See a difference? Consider temperature relaxation as another crystallization method.

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

HAMPTON
RESEARCH

Solutions for Crystal Growth

PROTEIN CONCENTRATION VERSUS NUMBER OF CRYSTALS

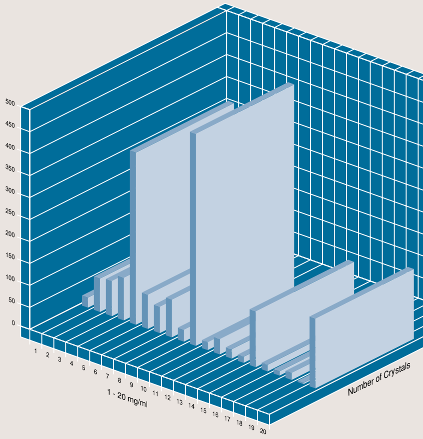

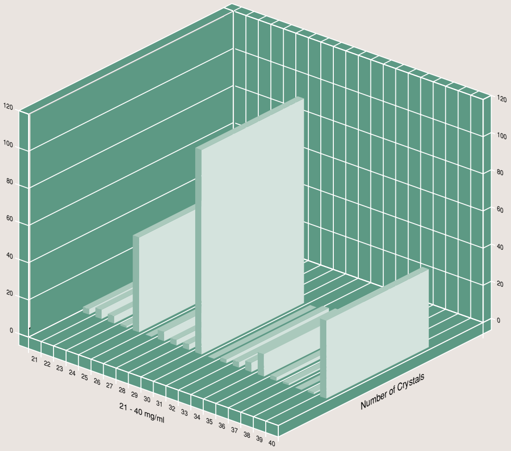

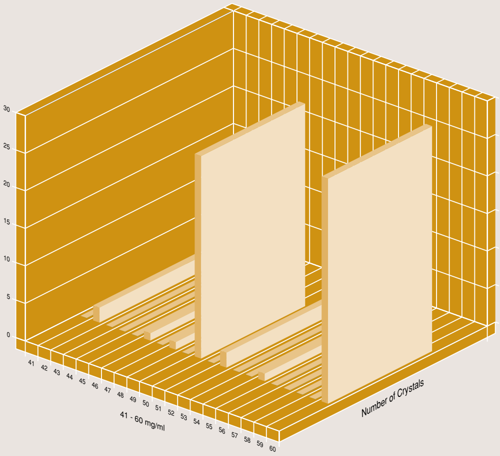

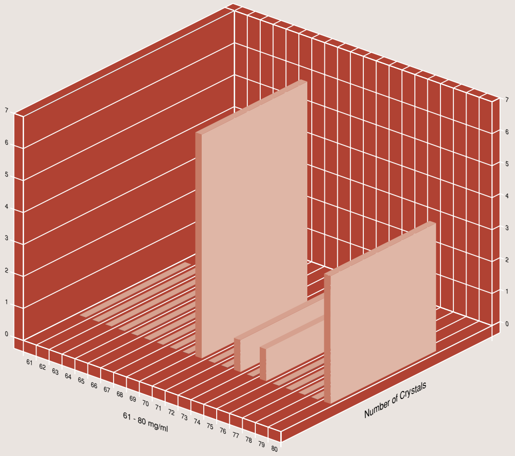

[Protein]	Number of Crystals
1	23
2	76
3	80
4	97
5	392
6	78
7	60
8	86
9	27
10	483
11	18
12	35
13	21
14	14
15	127
16	15
17	10
18	16
19	3
20	159
21	3
22	5
23	4
24	1
25	51
26	0
27	5
28	3
29	3
30	110
31	0
32	2

[Protein]	Number of Crystals
33	3
34	5
35	12
36	2
37	1
38	0
39	1
40	42
41	0
42	2
43	0
44	0
45	0
46	1
47	0
48	1
49	0
50	27
51	0
52	2
53	0
54	0
55	1
56	0
57	0

[Protein]	Number of Crystals
58	0
59	0
60	30
61	0
62	0
63	0
64	0
65	0
66	0
67	0
68	0
69	0
70	7
71	0
72	0
73	1
74	0
75	1
76	0
77	0
78	0
79	0
80	4

Protein concentration (mg/ml) versus number of crystals grown for 2,953 biological macromolecules reported from the BMCD.

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

CRYSTALLIZATION POLLING BOOTH

What temperatures do you routinely use for crystallization?

Answer	Percentage
4°C only	3%
25°C only	43%
4°C and 25°C only	33%
4°C and 25°C and other temperatures	21%

How important is it to keep your chosen crystallization temperature constant during the entire crystallization experiment (e.g. during crystal growth and crystal inspection)?

Answer	Percentage
Very important	45%
Important	35%
Neither important nor unimportant	7%
Slightly important	8%
Not important at all	4%

What crystallization method do you use most often for crystal optimization?

Answer	Percentage
Hanging Drop Vapor Diffusion	66%
Sitting Drop Vapor Diffusion	31%
Microbatch	3%
Free Interface Diffusion	0%
Dialysis	0%

What crystallization method do you use most often for screening?

Answer	Percentage
Hanging Drop Vapor Diffusion	36%
Sitting Drop Vapor Diffusion	58%
Microbatch	5%
Free Interface Diffusion	0.5%
Dialysis	0.5%

Based on your own experience, if your protein has a His tag, do you cleave it or leave it when you screen for crystallization conditions?

Answer	Percentage
Cleave it (remove His tag) and screen	26%
Leave tag intact and screen	15%
Leave it & screen. No xtals, cleave & screen.	59%

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

Cryobuffer

We've had good success using the well solution directly as the foundation of a cryobuffer in several situations where crystals cannot be grown directly in the presence of cryoprotectant, and where crystals don't tolerate transfer to artificial mother liquors. The basic protocol is as follows:

1. Remove 100 µl of the well solution after crystals have grown.
2. Split this sample into two 50 µl aliquots.
3. Add 7.5 mg of Dextrose (glucose) to the first aliquot and 15 mg of Dextrose to the second. Dissolve by gently pipetting with a wide-bore tip. This will give two sequential well solutions that now contain 15% and 30% w/v Dextrose. If all the dextrose won't go into the second aliquot, spin hard and remove the supernatant.
4. Transfer the crystal to aliquot 1, equilibrate for 3 minutes, then to aliquot number 2, then freeze.

We've had a few crystals that routinely crack or blow up when transferred to artificial mother liquor that behave well when transferred to well solution plus glucose. We assume that there is some aspect of the crystal drop (pH, ionic tension, precipitant concentration) that is more effectively reproduced within the well than by separately prepared mother liquors.

The nice thing about the protocol above is that you don't get much of a volume increase when dry dextrose is dissolved in the well solution, so the components in the solution are not diluted.

Finally, if you don't get a really good freeze, you can try to add about 5% v/v Glycerol to aliquot 2 in addition to the 30% w/v Dextrose.

Thank you Barry Stoddard for sharing this tip.

Cryo Tip

When adding glycerol to crystallization experiments for cryogenic data collection, remember that some proteins are more soluble in glycerol and one might have to increase the concentration of the primary precipitant (Example: Increase Ammonium sulfate from 2.0 M to 2.1 M or 2.2 M in the presence of 20% v/v Glycerol).

Cryo-Soaking Method

Eom et al. have reported improved diffraction using a novel cryo-soaking method. Crystals were grown using the following condition: 1.8 M Ammonium sulfate, 0.1 M Sodium citrate pH 5.1 (hanging drop, 0.5 ml reservoir). To improve diffraction quality, the crystals were sequentially transferred to cryosolvents containing 5 to 25% v/v Glycerol in five steps (5, 10, 15, 20, and 25%). Equilibration at each step was 30 seconds. Note that the cryoprotectant (glycerol) also contains the mother liquor used to grow the crystals. Equilibration at each step was 30 seconds. After the final transfer, the crystals were equilibrated for 30 minutes. The cryosolvent containing the crystal(s) was then air dried until Ammonium sulfate crystals began to appear. The authors report this took about 20 minutes at room temperature. Diffraction resolution using the novel cryo-soaking method improved from 3 angstrom to 2.2 angstrom. (Editors Note: Yet another case of crystal dehydration?)

Reference:

Acta Cryst (1999) D55 Pages 1601-1603 "Crystallization and preliminary X-ray crystallographic studies of the D2 region of the skeletal ryanodine receptor". Seha Kim, Dong Wook Shin, Do Han Kim and Soo Hyun Eom. E-mail: eom@kjist.ac.ky

Crystal Annealing

Flash-cooling crystals can sometimes increase the mosaicity of biological macromolecular crystals. In some cases, macromolecular crystal annealing can reduce the mosaicity of flash-cooled crystals without affecting molecular structure. Crystal annealing involves cycling a flash-cooled crystal to room temperature and then back to cryogenic temperature. The procedure can also be applied to sometimes restore diffraction from flash-cooled crystals that were not properly handled to and from cryogenic storage. Crystal annealing does not seem to improve a poorly diffracting crystal suffering from molecular disorder. The essential features of the crystal annealing procedure are the following: The crystal is first flash-cooled. The crystal is removed from the cryostream and quickly transferred to a drop of cryoprotectant at crystal growth temperature and allowed to remain in the drop for at least three minutes. The drop should be covered to prevent evaporation. Finally, the crystal is remounted on a loop and flash-cooled.

Reference:

Macromolecular Crystal Annealing: Overcoming Increase Mosaicity Associated with Cryocrystallography, J.M. Harp, D.E. Timm, and G.J. Bunick, Acta Cryst. (1998) D54 622-628.

Halides for Phasing Proteins

In the publication, "Novel approach to phasing proteins: derivatization by short cryo-soaking with halides", Dauter et al. described the procedure with four different proteins including lysozyme, RNAase A, subtilisin, and xylanase. Hey! Congratulations for taking the time to test beyond lysozyme! We wish more folks would take the time to test novel crystallization "discoveries" with more than just lysozyme! Anyway, Sodium iodide or Sodium bromide was added to the crystallization reagent and cryoprotectant mixture for the derivatization. It was noted that 1.0 M Sodium bromide did not seem to affect the quality and diffracting power of the tested crystals. 1.0 M Potassium iodide did mess with the crystallization of lysozyme and xylanase so the Potassium iodide concentration was reduced to 0.35 and 0.5 M for these two proteins (CrystalNews Editor Note: Both lysozyme and xylanase were crystallized in high salt where the other two proteins were crystallized in MPD and PEG. So it seems one might need to adjust the NaBr or KI concentration for high salt crystallizations and perhaps not too much for polymer or non-volatile organic crystallizations. Who knows? More results needed for a consensus). Crystals were transferred to the halide laced mother liquor for 15 to 45 seconds. Crystals were then flash-cooled (you are welcome Sean) for data collection. A fifth protein, in fact, a new protein to be described in a future publication was also successfully tested with this novel protocol.

For more details about the identification of anamolous sites, the phasing, and application to an unknown structure, see the following reference about quick cryo-soaking with halides as a possible alternative method for phasing protein crystal structures.

Reference:

Acta Cryst. 2000, D56, 232-237.

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

HAMPTON RESEARCH

Solutions for Crystal Growth

HOW TO: CRYOCRYSTALLOGRAPHY AND THE CRYSTALCAP SYSTEM

FLOWCHART FOR CRYO COOLING OF CRYSTALS

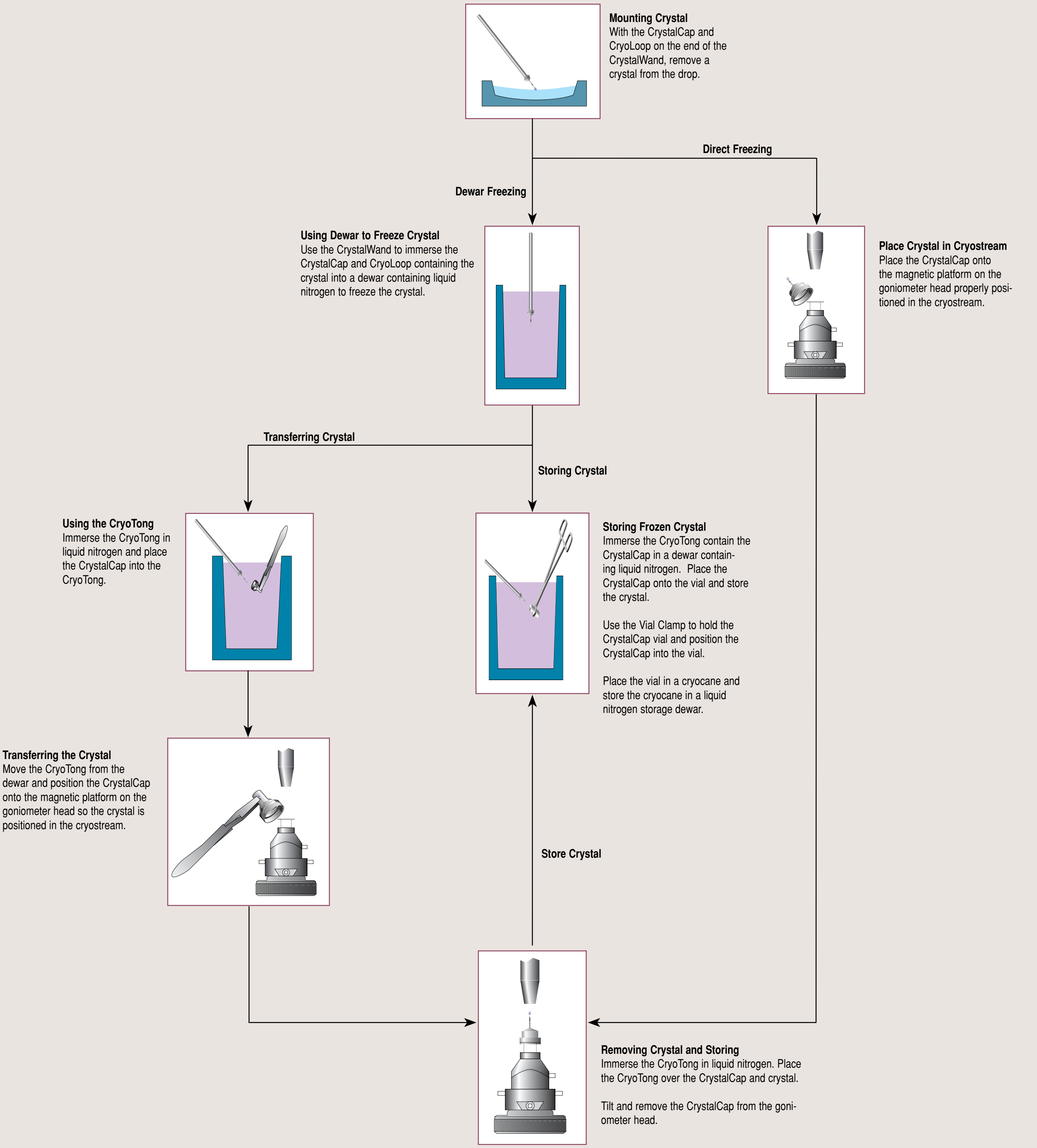

Mounting Crystal
With the CrystalCap and CryoLoop on the end of the CrystalWand, remove a crystal from the drop.

Using Dewar to Freeze Crystal
Use the CrystalWand to immerse the CrystalCap and CryoLoop containing the crystal into a dewar containing liquid nitrogen to freeze the crystal.

Place Crystal in Cryostream
Place the CrystalCap onto the magnetic platform on the goniometer head properly positioned in the cryostream.

Using the CryoTong
Immerse the CryoTong in liquid nitrogen and place the CrystalCap into the CryoTong.

Storing Frozen Crystal
Immerse the CryoTong contain the CrystalCap in a dewar containing liquid nitrogen. Place the CrystalCap onto the vial and store the crystal.

Use the Vial Clamp to hold the CrystalCap vial and position the CrystalCap into the vial.

Place the vial in a cryocane and store the cryocane in a liquid nitrogen storage dewar.

Transferring the Crystal
Move the CryoTong from the dewar and position the CrystalCap onto the magnetic platform on the goniometer head so the crystal is positioned in the cryostream.

Removing Crystal and Storing
Immerse the CryoTong in liquid nitrogen. Place the CryoTong over the CrystalCap and crystal.

Tilt and remove the CrystalCap from the goniometer head.

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

USING HALIDES FOR PHASING MOLECULAR STRUCTURES

Preliminary research shows that bromide (as well as iodide) can diffuse into protein crystals when soaked with the appropriate solution and can successfully be used for phasing.[1] Halide soaked crystals can be used for MAD, as well as multiple or single isomorphous replacement with anomalous scattering or for single anomalous diffraction. The procedure has been termed "Halide Cryosoaking" by Dauter and Dauter.[1] In simplest terms, the procedure involves dipping the crystal for a short period of time into a cryoprotectant solution that contains a significant concentration of halide salt. Although no single recipe will suffice for all proteins since each crystal has a unique crystallization recipe and will require different cryoprotectant cocktails, there are some general suggestions to follow at this time. First, there are currently more successful examples using bromide than iodide. The soak time is approximately 10 to 20 seconds. Longer soaks have not lead to more incorporation of halide ions for the examples to date. The concentration range of sodium bromide for soaking is approximately 0.25 to 1 M. Higher concentrations of halide ions may lead to more sites with higher occupancies and increased phasing power. Factors influencing the success of the procedure include the resolution and quality of the x-ray diffraction data, crystal symmetry, packing density, and pseudo-symmetric arrangements of molecules.

Tips for successful Halide cryosoak include:

1. Initially, preserve the formulation of the crystallization reagent used to grow the crystal as well as the formulation for a successful cryosoak and then add the halide salt. In other words, leave everything constant and add the halide salt.
2. If the crystallization reagent contains salt, try substituting the halide salt, especially if the salt is Sodium chloride.
3. High concentrations of the halide salt can serve as a cryoprotectant without the addition of other traditional cryoprotectants (Glycerol, MPD, sucrose).
4. Experiment with soak conditions. Vary the concentration of the original reagents, the concentration of the halide salt, and the soak time.

Examples of successful crystallization reagents optimized for halide cryosoaking:

Original Condition (Black) Halide Cryosoak Condition (Blue)

1.0 M Ammonium sulfate, 5 mM guanidine, 10% Glycerol, 0.1 M Sodium citrate pH 3.3[3]

1.0 M Ammonium sulfate, 5 mM guanidine, 18% Glycerol, 0.1 M Sodium citrate pH 3.3, 1 M Sodium bromide[3]

1.4 M Lithium sulfate, 0.1 M Tris pH 7.5[3]

1.2 M Lithium sulfate, 0.1 M Tris pH 7.5, 1 M Sodium bromide, 14% Glycerol[3]

1.0 M Sodium chloride, 0.1 M Sodium acetate pH 4.7[2]

0.1 M Sodium acetate pH 4.7, 1.0 M Sodium bromide, 30% Glycerol[2]

50% MPD, 0.1 M Sodium acetate pH 5.4[2]

50% MPD, 0.1 M Sodium acetate pH 5.4, 1.0 M Sodium bromide[2]

12% PEG 4,000, 0.1 M Citrate, 1.0 M Sodium chloride, 10 mM Calcium chloride, pH 6.0[2]

12% PEG 4,000, 10 mM Citrate, 10 mM Calcium chloride, 25% Glycerol, 1.0 M Sodium bromide[2]

10% Ammonium sulfate, 0.1 M TRIS HCl pH 7.4[2]

10% Ammonium sulfate, 0.1 M TRIS HCl pH 7.4, 25% Glycerol, 1.0 M Sodium bromide[2]

References

1. Entering a new phase: Using solvent halide ions in protein structure determination. Dauter, Z. & Dauter, M., Structure, Vol 9, R21-26, Feb 2001.
2. Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. Dauter, Z., Dauter, M., & Rajashankar, K.R. Acta Cryst. (2000) D56, 232-237.
3. Practical experience with the use of halides for phasing macromolecular structures: a powerful tool for structural genomics. Dauter, Z., Li, M. and Wlodawer, A.

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

MICROSEEDING

Background

A crystallization experiment typically begins with the sample in a stabilizing solution of water and a buffer, salt, reducing agent, ligand, or other reagent. Prior to mixing the sample with crystallization reagent, this sample solution is undersaturated with respect to the macromolecule in question (sample). In an undersaturated sample solution, no crystals can nucleate, nor can crystals grow from seeds. Upon addition of a crystallization reagent, the relative supersaturation of the sample is increased. Assuming the crystallization reagent decreases the solubility of the sample to increase the relative supersaturation, three events can take place. In the first stage of supersaturation, the metastable zone, spontaneous homogenous nucleation cannot occur, but crystal growth from seeds can occur. Moving further into supersaturation, the labile zone, spontaneous homogeneous nucleation and crystal growth can occur. Further into supersaturation, the precipitation zone, precipitation of the sample from solution occurs. See figure 1 below.

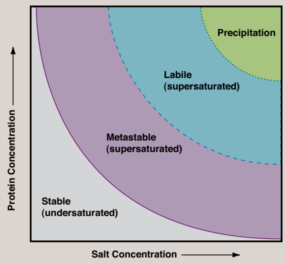

figure 1

The diagram is divided into 4 zones:

1. **Stable:** undersaturated where crystallization is not possible.
2. **Metastable:** supersaturated where nuclei cannot form but crystals from seeds can grow.
3. **Labile:** supersaturated where crystallization can occur.
4. **Precipitation**

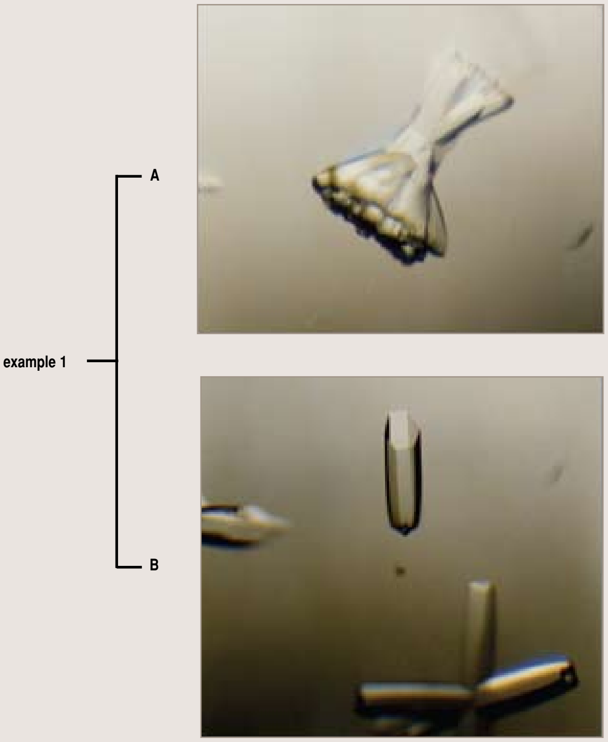

Seeding

Seeding allows one to grow crystals in the metastable zone. Why would one want to do this? For control, reproducibility, and to improve the likelihood of a successful crystallization experiment. In the metastable zone, crystals can grow from seeds but cannot spontaneously nucleate. By placing a seed or solution of seeds in a drop which is saturated to this zone, one can use the seeds to grow larger single crystals. By controlling the number of seeds introduced into the zone drop, one can control the number of crystals grown. It is not practically possible to measure and know the number of seeds introduced to a drop, but by performing serial dilutions from a concentrated seed stock, one can control the number of crystals grown in the drop.

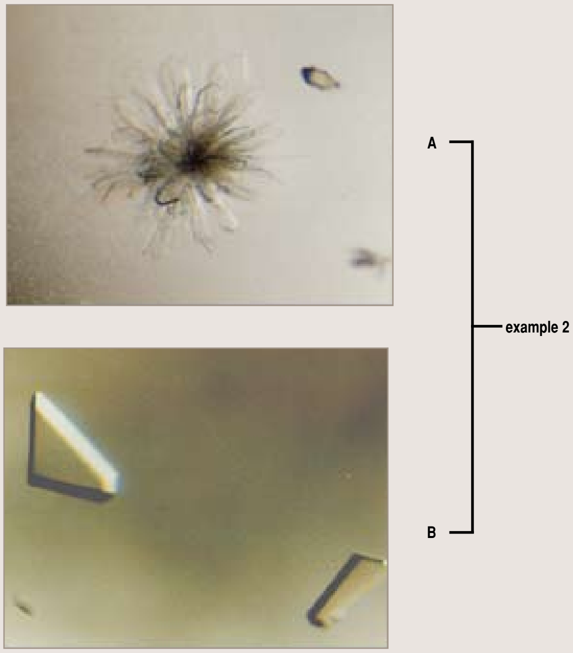

Seeding Examples

Example 1: (A) Initial crystals of the catalytic domain of CheB methyltransferase were grown from Crystal Screen 2™, reagent 42. (B) Microseeds were taken from (A) to grow diffraction quality crystals.

Example 2: (A) Initial crystals of CheR methyltransferase were grown producing platelet clustered crystals. (B) Microseeds were taken from (A) to grow quality 3-dimensional crystals.

Thank you Ann West and Snezana Djordjevic for your examples of microseeding.

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

DETERGENTS AT WORK

The image below is a view of the crystal packing of ß-lactamase. The image demonstrates the detergent molecule linking SHV-1 molecules in the crystal lattice. The detergent binding to the ß-lactamase is CYMAL®-6. It appears as though the well-ordered detergent molecules are linking protein molecules in the crystal lattice. The hydrophobic tail of the detergent is buried in the protein and the hydrophilic head is hydrogen bonded with a neighboring molecule.

ß-lactamase resisted producing any decent crystals until the detergents CYMAL®-5 and CYMAL®-6 were included in the crystallization drop. The crystals were grown in a relatively low ionic strength environment (15% w/v PEG 6,000, 50 mM HEPES pH 7.0, and detergent) via sitting drop vapor diffusion.

The crystals and the structure demonstrate the efficacy of detergents as additives to improve crystallization of soluble proteins and bring to light the molecular linking capabilities that small molecules can have in biological macromolecular crystals.

CYMAL®-5 and CYMAL®-6 are part of the Hampton Research Detergent Screen.

Images and discussion courtesy of James Knox, University of Connecticut, Department of Molecular and Cell Biology 1999.

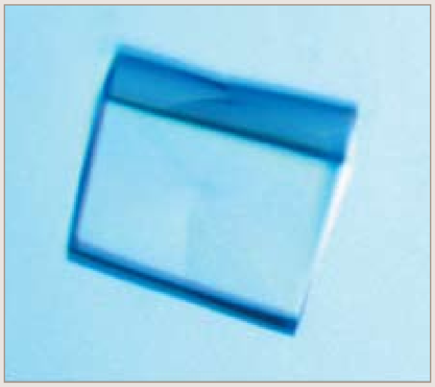

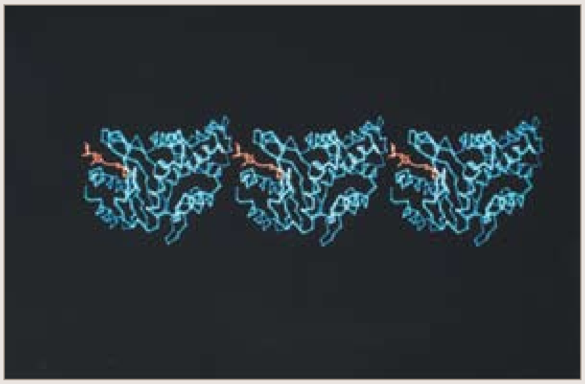

CRYSTALLIZATION EXPERIMENTS ON THE PHOTOSYSTEM II REACTION CENTER FROM PISUM SATIVUM

Submitted by Ivana Kuta Smatanova, University of South Bohemia, Photosynthesis Research Center and Laboratory of Biomembranes, Branisovska 31, CZ-370 05 Ceske Budejovice, Czech Republic

Crystallization trials have been performed in Cryschem™ Plates for sitting drops and in capillary tubes at 277K or 289K. 15-mg/ml (1.3 mg/ml chlorophyll a) protein has been mixed with different type of detergents and additives used in the crystallization experiments. We have either used the Hampton Research MembFac kit for membrane protein screening or solutions prepared in our own lab. 50 commercial solutions and 50 different kind of detergents have been available for experiments. According to the published information about a crystallization of photosynthetic reaction center core complexes we have used solutions containing PEG as precipitant. Concerning detergents we have proceeded in accordance with literature data and up to this day we have tried 15 different detergents, detergents with different alkyl chain length and mixtures of detergent as well. Processes inside the drops have been observed during the period of 2-4 weeks at 289K and 4-6 weeks at 277K. The major part of drops contained precipitates, separated phases or segregated carotenoids in various forms. We have already found for crystallization experiments acceptable pH rate around 7.00 (±0.50), the precipitant PEG 4000 or PEG 6000 (in the case of using a commercial solutions without PEGs, we have observed the clear drops without precipitate or other formations). For detergents, in the presence of 1-heptyl-ß-D-glucoside, 1-octyl-ß-D-glucoside, 1-nonyl-ß-D-glucoside and their thio forms such as heptyl-ß-D-thioglucoside and 1-octyl-ß-D-thioglucoside, different type of carotenoid formations are separated from the protein samples. We have also performed crystallization experiments either in the absence or in the presence of additional amphiphiles (heptanetriol) that has been used to modify detergent characteristics and to facilitate membrane protein crystallization.[1] Nowadays crystallization experiments are still in the progress; we have tried to use new kind of detergents and their mixtures and new crystallization conditions as well.

Reference

1. *Marone P.A., Thiyagarajan P., Wagner A.M. and Tiede, D.M. "Effect of detergent alkyl chain length on crystallization of a detergent-solubilized membrane protein: correlation of protein-detergent particle size and particle-particle interaction with crystallization of the photosynthesis reaction center from Rh. sphaeroides" J. Cryst. Growth 207 (1999) 214-225.*

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

PROTEIN CRYSTALLIZATION RECIPES

The following crystallization recipes can be useful for teaching and practicing protein crystallization or for fooling your PI into thinking you have grown huge crystals of that impossible-to-crystallize protein.

Lysozyme

Protein source: Hampton Research HR7-108 or HR7-110
Worthington Biochemical Corporation
Sigma-Aldrich
Recommended concentration: 20 to 100 mg/ml
Sample Buffer: 0.1 M Sodium acetate pH 4.6
Crystallization reagent: 8% w/v Sodium chloride, 0.1 M Sodium acetate pH 4.6
10% w/v Sodium chloride, 0.1 M Sodium acetate pH 4.6, 10% v/v Glycerol
Additional Reagents: Hampton Research Index screen reagents 8, 22, 28, 31, 34, 40, 58, 59, 69, 86, 88 and Hampton Research Grid Screen Sodium Chloride reagents
Method: Hanging or sitting drop vapor diffusion, microbatch, free interface diffusion
Procedure: Mix equal amounts of lysozyme with reagent, incubate at 4°C or room temperature

Lysozyme crystals in 15 minutes

Recommended concentration: 100 mg/ml
Sample Buffer: 0.1 M Sodium acetate pH 4.6
Crystallization reagent: 30% w/v PEG MME 5,000, 1.0 M Sodium chloride, 0.05 M Sodium acetate trihydrate pH 4.6 (Hampton Research HR2-805)
Method: Hanging or sitting drop vapor diffusion, microbatch, free interface diffusion
Procedure: Mix equal amounts of lysozyme with reagent, incubate at 4°C or room temperature

Thaumatin

Protein source: Worthington Biochemical Corporation or Sigma-Aldrich
Recommended concentration: 20 to 40 mg/ml
Sample buffer: DI water, 25 mM BIS-TRIS propane or Imidazole pH 6.5
Crystallization reagent: 1.0 M Potassium sodium tartrate
Additional reagents: Hampton Research Index screen reagents 26, 31, 78, 86
Method: Hanging or sitting drop vapor diffusion, microbatch, free interface diffusion
Procedure: Mix equal amounts of protein with reagent, incubate at 4°C or room temperature

Glucose isomerase

Protein Source: Hampton Research HR7-100 or HR7-102
Recommended concentration: 10 to 40 mg/ml
Sample buffer: 10 mM HEPES, 1 mM Magnesium chloride hexahydrate
Crystallization reagent:

- 1.5 - 2.5 M Ammonium sulfate pH 6.0 - 9.0
- 10 - 15 % w/v PEG 4,000 - 8,000, 0.2 M salt *, 6.0 - 9.0
- 0.6 - 1.0 M Sodium citrate tribasic dihydrate, pH 6.0 - 8.0
- 0.1 - 0.3 M Magnesium formate dihydrate
- 1.0 M Sodium formate, pH 5.0 - 7.0
- 10 - 20 % v/v MPD, 0.2 M salt *, pH 6.0 - 9.0
- 10 - 20 % v/v PEG 400, 0.2 M salt *, pH 6.0 - 9.0
- 10 - 20 % 2-Propanol, 0.2 M salt *, pH 6.0 - 9.0

* Ammonium sulfate, Magnesium acetate, Magnesium chloride, Calcium acetate and other salts
Additional reagents: Hampton Research Index screen reagents including but not limited to 4, 5, 14, 48, 49, 67, 92
Method: Hanging or sitting drop vapor diffusion, microbatch, free interface diffusion
Procedure: Mix equal amounts of protein with reagent, incubate at 4°C

Glucose isomerase for cryo

Special thanks to Laurie Betts for the the cryo work and data!
Protein Source: Hampton Research HR7-100 or HR7-102
Recommended concentration 20 to 30 mg/ml
Sample buffer 10 mM HEPES, 1 mM Magnesium chloride hexahydrate
Crystallization reagent: 30% v/v 2-Methyl-2,4-pentanediol, 0.1 M Sodium Cacodylate pH 6.5, 0.2 M Magnesium Acetate tetrahydrate (Crystal Screen Cryo reagent 21) or 30% v/v Polyethylene Glycol 400, 0.1 M HEPES - Na pH 7.5, 0.2 M Magnesium Chloride hexahydrate (Crystal Screen Cryo reagent 23)
Method: Hanging or sitting drop vapor diffusion, microbatch, free interface diffusion
Method: Mix equal amounts of Glucose isomerase and reagent. Mount crystal in CryoLoop. Swish mounted crystal through reservoir to increase MPD or PEG 400 concentration in crystal to decrease mosaicity. No swish for more mosaicity. Crystal can diffract to 1.6 Angstrom, indexes in I-centered orthorhombic with a=92.7 b=97.4 c=102.7. Mosaicity is about 0.5 to 0.6.

Alpha Lactalbumin

Protein source: Sigma-Aldrich
Recommended concentration 10 to 20 mg/ml
Sample buffer: 10 mM TRIS hydrochloride pH 8.5
Crystallization reagent: 30% w/v Polyethylene glycol 3,350, 0.1 M TRIS hydrochloride pH 8.5, 0.2 M Lithium sulfate
Additional reagents: Hampton Research Index screen reagent including but not limited to
4, 5, 6, 74
Method: Hanging or sitting drop vapor diffusion, microbatch, free interface diffusion
Procedure: Mix equal amounts of protein with reagent, incubate at 4°C or room temperature

Catalase

Protein source: Worthington Biochemical Corporation or Sigma-Aldrich
Recommended concentration: 10 to 20 mg/ml
Sample buffer: 10 mM HEPES pH 7.0
Crystallization reagent: 20% w/v Polyethylene glycol 3,350, 0.1 M HEPES pH 7.0
Additional reagents: Hampton Research Index screen reagents including but not limited to 39, 55, 63, 64, 88, 89
Method: Hanging or sitting drop vapor diffusion, microbatch, free interface diffusion
Procedure: Mix equal amounts of protein with reagent, incubate at 4°C or room temperature

Ribonuclease S
Protein source: Sigma-Aldrich
Recommended concentration: 40 mg/m
Sample buffer: 10 mM HEPES pH 7.0
Crystallization reagent: 30% Polyethylene glycol 4,000, 0.1 M Sodium citrate tribasic dihydrate pH 5.6, 0.2 M Ammonium acetate
Additional reagents: Hampton Research Crystal Screen reagent including but not limited to 9, 20
Method: Hanging or sitting drop vapor diffusion, microbatch, free interface diffusion
Procedure: Mix equal amounts of protein with reagent, incubate at 4°C or room temperature

Proteinase K
Protein source: Sigma-Aldrich
Recommended concentration: 20 mg/ml
Sample buffer: 10 mM HEPES pH 7.0
Crystallization reagent: 0.4 M Potassium sodium tartrate tetrahydrate
Additional reagents: Hampton Research Crystal Screen reagent including but not limited to 2, 3, 4, 6, 7, 9, 10, 11, 15, 16, 17, 18, 20, 22, 25, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39, 40, 41, 42, 44, 45, 46, 47, 48, 49, 50
Method: Hanging or sitting drop vapor diffusion, microbatch, free interface diffusion
Procedure: Mix equal amounts of protein with reagent, incubate at 4°C or room temperature

Trypsin (bovine pancreas)
Protein source: Worthington Biochemical Corporation or Sigma-Aldrich
Recommended concentration: 15 to 60 mg/ml
Sample buffer: 10 mM Calcium chloride, 10 mg/ml Benzamidine hydrochloride, 20 mM HEPES pH 7.0
Crystallization reagent: 4% w/v Polyethylene glycol 4,000, 0.2 M Lithium sulfate monohydrate, 0.1 M HEPES pH 7.0, 15% v/v Ethylene glycol
Additional reagents: Hampton Research Crystal Screen reagents including but not limited to 4, 15, 16, 20, 28, 30, 31
Method: Hanging or sitting drop vapor diffusion, microbatch, free interface diffusion
Procedure: Mix equal amounts of protein with reagent, incubate at 4°C or room temperature

Ferritin
Protein source: Sigma-Aldrich
Recommended concentration: 20 mg/ml
Sample buffer: 10 mM HEPES pH 7.0
Crystallization reagent: 0.8 Ms Ammonium sulfate, 0.1 M HEPES pH 7.5, 60 mM Cadmium sulfate
Additional reagents: Crystal Screen 2 reagents 12, 34 and Index screen reagent 64 and PEG/Ion 2 reagent 42 and PEGRx 2 reagent 20
Method: Hanging or sitting drop vapor diffusion, microbatch, free interface diffusion
Procedure: Mix equal amounts of protein with reagent, incubate at 4°C or room temperature

Xylanase
Protein source: Hampton Research HR7-104, HR7-106
Recommended concentration: 3 to 20 mg/ml
Sample buffer: 0.09 M Sodium potassium phosphate pH 7, 21% v/v Glycerol
Crystallization reagent: 1.0 M Sodium potassium phosphate pH 7
Additional reagents: Quik Screen reagents A5, B5, C5, D5, A6, B6, C6, D6
Method: Hanging or sitting drop vapor diffusion, microbatch, free interface diffusion
Procedure: Mix equal amounts of protein with reagent, incubate at 4°C or room temperature

Lipase B
Protein source: Hampton Research HR7-099
Recommended concentration: 10 to 100 mg/ml
Sample buffer: Deionized water or 10 mM Sodium acetate pH 4.6
Crystallization reagent: 1.0 M Sodium potassium phosphate pH 7
Additional reagents: Quik Screen reagents A5, B5, C5, D5, A6, B6, C6, D6
Method: Hanging or sitting drop vapor diffusion, microbatch, free interface diffusion
Procedure: Mix equal amounts of protein with reagent, incubate at 4°C or room temperature

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

TEMPERATURE AS A CRYSTALLIZATION VARIABLE

Temperature and Crystallization

Temperature can be a significant variable in biological macromolecule and small molecule crystallization.(1-5) Temperature often influences nucleation and crystal growth by manipulating the solubility and supersaturation of the sample. Temperature has also been shown to be an important variable with phase separation in detergent solutions during membrane protein crystallization.[7]

Control and manipulation of temperature during the screening, optimization and production of crystals is a prerequisite for successful and reproducible crystal growth of proteins with temperature dependent solubility. Christopher et al.,[5] testing 30 randomly chosen proteins, found 86% demonstrated a temperature dependent solubility and suggested that temperature induced crystallization could be a generally useful technique. Temperature was shown to affect quantity, size, and quality of the crystals, as well as sample solubility and preliminary crystallization data.

One advantage of temperature is that it provides precise, quick, and reversible control of relative supersaturation. Using temperature, in addition to standard crystallization variables such as pH, reagent composition and concentration, and sample concentration, can increase the probability of producing crystals as well as uncover new crystallization conditions for a sample. Additional crystallization conditions may uncover reagent formulations more amicable to heavy atom derivatization, cryoprotection, and optimization or at least offer options. Temperature is amenable to control and can be used to carefully manipulate crystal nucleation and growth. This control can also be used to etch or partially dissolve the crystal and then grow it back in an attempt to improve size, morphology, and quality or assist with seeding. Temperature control is noninvasive and can manipulate sample solubility and crystallization with altering reagent formulation.

Typically, crystallization screens and experiments are performed at room temperature and possibly 4°C. A reasonable range of temperature to screen and optimize for protein crystallization is 4 to 37°C and some proteins have been crystallized at 60°C (glucagon and choriomammotropin). Temperature incubations above room temperature should be monitored closely for evaporation from the drop and reservoir. A 2 µl hanging drop vapor diffusion experiment at 37 °C can evaporate in as little as 48 hours depending upon the plate and the quality of seal. Microbatch under Paraffin Oil can minimize evaporation problems. In the case of room temperature incubations, temperature control and stability are often minimal since the experiments may be left in the open room. In an open room, temperature fluctuations may be significant, especially over a 24 hour period and on weekends when thermostatic control of the room environment can fluctuate 10 degrees or more. Incubation at 4°C and other temperatures are often more stable since the incubation is performed in some type of incubator. Another source of temperature fluctuation occurs while viewing experiments. The light microscope is a heat source and extended viewing can significantly alter the temperature of small drops. Quick, efficient viewing can minimize temperature changes. Also, remember to turn off the light source when leaving plates on the stage in one position for more than a few seconds.

While controlled temperature can be important for consistent results, temperature fluctuation can be useful in obtaining high quality crystals by screening a larger range of crystallization conditions.[8] For a sample with temperature dependent solubility, changes in temperature can equate to changes in a crystallization reagent condition. Hence, a sparse matrix screen takes on a new dimension when screened at multiple temperatures, or ramped over several different temperatures over a period of time.

How does one test for the effect of temperature and temperature dependent solubility without consuming a lot of sample? One solution is to set a single crystallization screen at one temperature, allow the experiment to incubate for a week, record the results, and then move the plate to another temperature. Allow the experiment to incubate for a few days to a week at the new temperature and record the results. If one notices changes is solubility (i.e. clear drop turning to precipitate, or precipitate turning to clear drops) between the two temperatures, then the sample has temperature dependent solubility and temperature should be explored as a crystallization variable.

Temperature gradients can be used for screening and optimization of proteins with temperature dependent solubility. For screening, set the experiment at one temperature, allow it to equilibrate, and then slowly change the temperature to a second temperature. In general, ramp the temperature so that the sample is exposed to an increase in relative supersaturation as the temperature changes over time. In other words, ramp from high to low temperature if the sample is more soluble at high temperatures. This can be accomplished using a programmable temperature incubator. A temperature gradient or ramp, allows one to slowly approach temperatures where a sample may have a decrease in solubility with a corresponding increase in relative supersaturation. Published examples of temperature gradient or temperature ramp crystallization include elastase9 (25 to 20°C gradient), alpha-amylase10 (25 to 12°C gradient), and insulin11 (50 to 25°C gradient).

To demonstrate how screening temperature could affect and enhance the results obtained from a preliminary crystallization screen, a temperature incubator was used to screen four different temperatures. Using Glucose isomerase and Crystal Screen™, sitting drop vapor diffusion experiments were set using Cryschem™ plates at 4, 15, 25, and 37°C. Drops were observed daily and the results were quite interesting. Glucose Isomerase crystallized in 19 conditions at 25°C, 23 conditions at 15°C, 28 conditions at 4°C, and 12 conditions at 37°C. A similar approach with Trypsin, yielded crystals in 8 conditions at 15°C, 4 conditions at 25°C, and 7 conditions at 32°C. In the case of Trypsin, a single set of Cryschem plates were set and then simply moved from one temperature to another over a period of one week, scoring results before each temperature change.

Temperature Tips

- For proteins with "normal" solubility, in high salt they will be more soluble at cold than at warm temperatures.

- For proteins with "normal" solubility, in low salt they will be more soluble at warm than at cold temperatures.

continued next page....

- For proteins with "normal" solubility, they will precipitate or crystallize from lower concentration of PEG, MPD, or organic solvent more slowly at low than at high temperatures.

- Temperature effects can be more pronounced in low ionic strength reagent conditions.

- Do not use the appearance or non-appearance of crystals at various temperatures to gauge the effectiveness of temperature as a crystallization variable. Rather, use the difference in the solubility at different temperatures to gauge the effect temperature has on sample solubility. If an effect is observed, explore temperature as a crystallization variable.

- Temperature can affect different crystal forms and growth mechanisms.[12]

- When incubating experiments below and above room temperature and viewing experiments at room temperature, condensation can be a problem. To minimize and avoid condensation with vapor diffusion experiments, stack a dummy plate (with the reservoir filled with water and sealed) at the bottom and top of the stack of plates. This will slow the temperature change in the sandwiched plates and minimize condensation.

- The microbatch method works well for temperature exploration. In a traditional microbatch experiment, the relative supersaturation of the system does not change since in theory, there is no vapor diffusion. However, if the sample exhibits temperature dependent solubility, temperature can be used to manipulate sample solubility in a microbatch experiment. Another plus of using microbatch is the lack of condensation while viewing the experiment. Covers of microbatch plates can be removed for a clear view.

- Condensation with a hanging drop can lead to an alteration when the condensation mixes with the drop. In a sitting drop, condensation does not mix with the drop unless the condensation falls from above.

- To dry up condensation, add a small amount of concentrated salt solution to the reservoir. Keep in mind this might also further dehydrate your drop.

- Ideally, one should set the experiment at the eventual incubation temperature and all reagents, samples, and plates should be equilibrated to the incubation temperature. This is a reality for room temperature and 4°C setups for those of us with cold rooms. For the rest of us, we can set the experiment at room temp and then toss it into the incubator. Or, for 4°C setups, one can cheat. Simply incubate the reagents, sample, plates and slides in the refrigerator before set up. During the set up, place materials in a tray full of ice. Maintain the plates on ice during the set up. Seal and move smartly to the 4°C incubator.

- Nucleic acid temperature stability allows one to examine temperatures between 4 and 35°C.

- Increasing temperature increases the disorder of reagent molecules. Varying the temperature of a crystallization experiment can manipulate sample-sample as well as sample-reagent and reagent-reagent interactions. Such manipulations may have an impact on interactions which control nucleation and crystal growth. In addition, such interactions may have an impact on crystal packing as well as the termination of crystal growth.

- Temperature can be a habit modifier and change the crystal lattice. For example, at temperatures below 25°C and in the presence of sodium chloride and acidic pH, the tetragonal form of lysozyme is favored. Under similar reagent conditions above 25°C, the orthorhombic form is favored.[13]

- The preparation of heavy atom isomorphous derivatives can depend upon the temperature of the experiment. In most cases, it seems the soak temperature is the same as the crystallization temperature.

References

1. *Giege, R., and Mikol, V., Trends in Biotechnology (1989) 7, 277.*
2. *McPherson, A., European J. Biochemistry (1990) 189, 1.*
3. *A. Ducruix and R. Giege, Editors, Crystallization of Nucleic Acids and Proteins: A Practical Approach, IRL Press at Oxford University Press, 1991.*
4. *Lorber, B., and Giege, R., Journal of Crystal Growth (1992) 122, 168-175.*
5. *Christopher, G.K., Phipps, A.G., and Gray, R.J., Journal of Crystal Growth (1998) 191, 820-826.*
6. *Haser, R., et al., Journal of Crystal Growth (1992) 123, 109-120.*
7. *Garavito, R.M., and Picot, D., Journal of Crystal Growth (1991) 110, 89.*
8. *Drenth, J., Crystal Growth (1988) 90, 368.*
9. *Shotton, D.M., Hartley, B.S., Camerman, H., Hofmann, T., Nyborg, S.C., and Rao, L., Journal of Molecular Biology (1968) 32, 155-156.*
10. *McPherson, A., and Rich, A., Biochem. Biophys. Acta (1972) 285, 493-497.*
11. *T.L. Blundell, and L.N. Johnson, Protein Crystallography, Academic Press (New York) 1976, 59-82 (method by Guy Dodson).*
12. *A. McPherson, Crystallization of Biological Macromolecules, Cold Spring Harbor Laboratory Press, 1999.*
13. *Ataka, M., and Tanaka, S., Biopolymers (1986) 25, 337.*

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

Crystal image.

Anastasia Tziridis, Institute for Biotechnologie
Martin-Luther-Universität Halle-Wittenberg, Germany

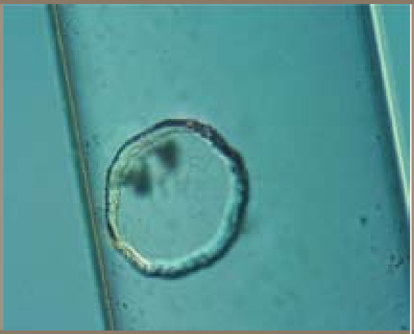

Protein crystal grown by counter diffusion technique in 0.2mm capillary in the presence of 0.1 % agarose.

Pavlina Rezacova,
UT Southwestern Medical Center at Dallas, Texas, USA.

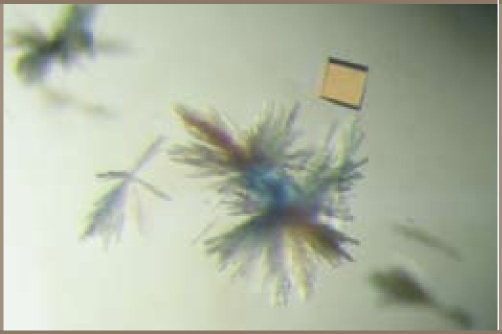

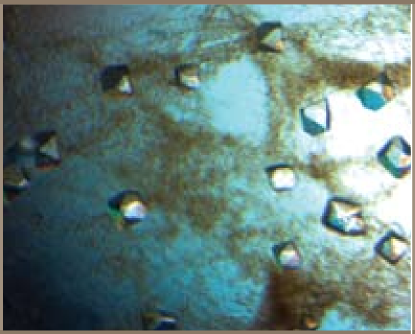

Crystals of human mineralocorticoid receptor complexed with ligand. Grown using Hampton Salt Rx screen.

Kevin P Madauss, GlaxoSmithklinem Research Triangle Park,
North Carolina, USA.

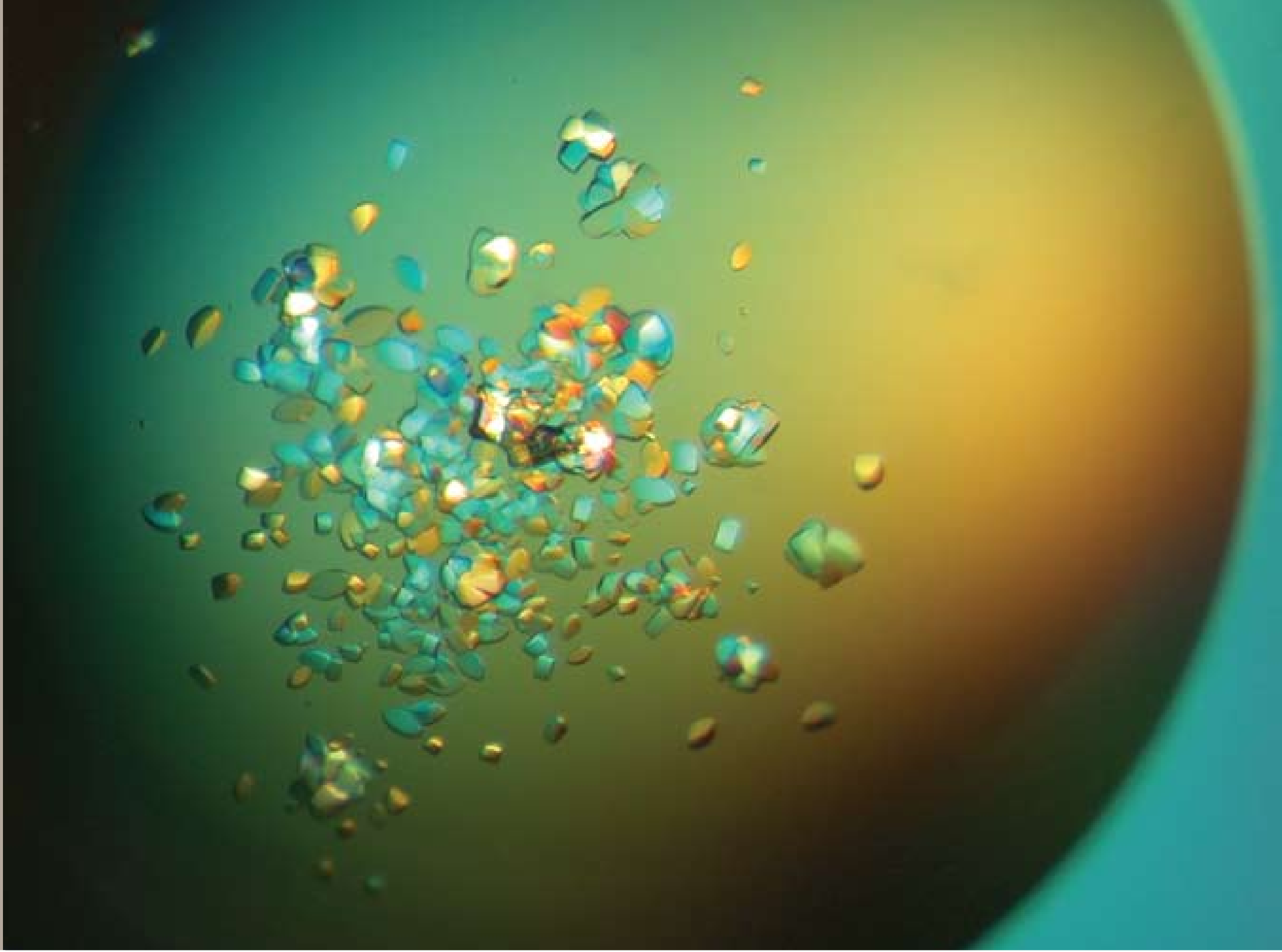

HAMPTON RESEARCH

TIPS FROM RAMC

Crystals of the Deinococcus radiodurans 50S ribosome subunit.

From the group of Paola Fucini, Max Planck Institute for Molecular Genetics, Berlin, Germany.

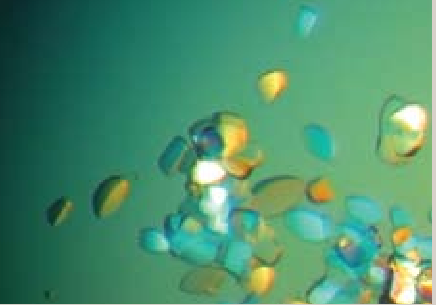

Filtering the Protein

Naomi Chayen

To reduce the number of crystals and increase their size, try filtering the protein solution prior to setting up the experiment. Try the following filter sizes: 0.22, 0.1 micron and 300 kD molecular weight cut-off. Try the Millipore centrifugal filters.

Pseudo-Microseeding

Mike Sintchak

When working with crystals that grow fairly rapidly (one day) try the following. Pipette multiple protein drops (2 to 4 works best) onto the cover slide. Using a single pipette tip, get the reservoir solution to mix with the first drop. Now, go back into the reservoir with the same tip to get the reservoir solution for the second drop. Continue for the remaining drops with the same tip. In certain cases, seeding starts very quickly, so by using the same tip one can introduce minute seeds to successive drops. Use the same cover slide with multiple drops to minimize evaporation.

Low Molecular Weight PEGs

Lesley Haire

When screening with low molecular weight PEGs try microbatch. Crystals appear rapidly with PEG 400-2000. To convert to vapor diffusion use 0.2 M buffer in the well and a 1:1 drop ratio. Try using a positive displacement pipette such as the Anachem Microman 1 - 10 µl. These are much more accurate.

Purest is Not the Best

Michal Harel

A protein which was purified and showed some faint bands of contaminants on a native gel was crystallized and solved successfully. The same protein, purified by HPLC and resulting in a single band native gel did not crystallize.

Cryoprotectant

Elspeth Garman

When making up cryoprotectant solutions containing glycerol, put a test tube of glycerol in a beaker of warm water. The viscosity falls and it is easier to pipette accurately.

Mass spectrometry

David Leys

We found mass spectrometry like ESMS and MALDI highly efficient in determining impurities and/or microheterogeneities in our protein sample/batch. In most cases it is a simple, straight forward method which requires a minimum amount of sample. In some cases it has shown to detect impurities/microheterogeneities when other techniques did not.

Shape of Drop

Ursula Kamlott, Hoffmann-La Roche: ursula.kammlott@roche.com

One of my proteins produced only zillions of tiny useless crystals when I mixed the drops the conventional way-mixing well, overlaying, etc. the protein with precipitant solution. Large, gorgeous crystals were produced when I crossed the drop, creating a gradient within the drop. This worked best, setting up sitting drops with vapor diffusion.

Don't Flip

Dennise Dombroski

When removing crystals from a hanging drop, I sometimes find that the biggest crystals fall against the coverslip and are impossible to resuspend without damage. I had our glass shop make a stand to transfer the coverslip that enabled me to manipulate the crystals more easily.

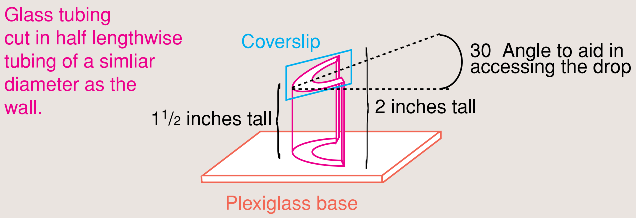
Glass tubing cut in half lengthwise tubing of a simliar diameter as the wall.
Coverslip
30 Angle to aid in accessing the drop
1 1/2 inches tall
2 inches tall
Plexiglass base

Extreme Soak

Jirundon Yuvaniyama

For soaking crystals with compounds with limited solubility I have tried two "extreme ways" (although not much- and more experiments should be tried):

- Leave some solid compound in the soaking solution.
- Dissolve some compound in n-octanol and layer the octanol solution on top of the soaking experiment.

These two provide (hopefully) more or less constant concentration of the compound in the soaking solution. The octanol layer may help reduce air oxidation by preventing direct contact of soaking solution to air.

I found that 200 µl of soaking solution in a well of the 24 well Linbro® plate is a good volume to work with:

—Not too little that possibly causes concentration change due to evaporation (Sealed well) and not too much that the crystals get lost in the solution.

The Glycerol Effect

Anil Mistry

Glycerol has many benefits but also some drawbacks. We found it to be beneficial with one protein we were working with; this protein is a transpeptidase called Mur A. The protein is quite soluble and could concentrate to 20 mg/ml but it lost activity over time when stored at 80°C. We therefore dialysed it into 50% v/v Glycerol to see if the activity could be retained for longer, this would allow us to make a large batch rather than regular smaller batches, for crystallisation. On dialysis we found a significant reduction in the volume of protein, so much so that the protein had concentrated from 5-10 mg/ml to almost 50 mg/ml. Activity was also found to be retained with no significant loss after 6 months at 80°C. This gave us a method for storing large batches of Mur A at 80°C, without losing activity and also resulted in a sample pre-concentrated for crystallisation and containing a cryo-protectant. Other sugars gave similar effects; sucrose, sorbitol, etc., but none were as effective as glycerol in achieving the 50 mg/ml final concentration.

TIPS FROM RAMC

PAGE 2-7

Iodoacetic Acid
Bernie Santarsiero

Add a small amount of (~ 1%) iodoacetic acid to buffer solutions. This helps prevent aggregation by carboxymethylation of cys. Also, iodoacetic acid seems to help form salt bridges and aid in crystallization.

Soaking Crystals to Improve Resolution
Irene Weber

Try soaking poorly diffracting crystals in higher concentration of precipitate, ammonium sulfate, or PEG. It may take several weeks so test after 1 or 2 months.

Rapid Preliminary Screening of Protein for Aggregation Using Protein Quantities
Tom Zarembinski

When only small amounts of protein are available, it is not feasible to screen many compounds which promote monodispersim using a dynamic light scattering machine. To detect aggregation, we use a pseudo-native gel approach: 1 l of protein is mixed with 1 l of additive from Hampton Additive or Detergent Screens. These samples are then incubated at room temperature for 20-30 minutes and then placed in 2x sample buffer containing no DTT, no SDS and are not boiled. These samples are run on a standard SDS-PAGE gel. We have screened many additives using this approach and it has given us leads for subsequent optimization of protein buffers.

HPLC Profile
Glenn Dale

Keep an HPLC (Reverse phase) profile of your protein before crystallisation and after crystal formation. It can be used as a quality control and tells you if any modifications have occurred.

Don’t Throw Away Without Looking Close
Kalevi Visuri

Look closely at your old test tubes when cleaning the place. Proteins do crystallize on the walls of the tube when stored in a cold room. I picked up my tubes from the wastebasket and an x-ray was made from an old supersaturated protein tube.

Concentration of Protein without Aggregation
Paul Reichert

Use centipreps (millipore) for concentrating protein >O.5 ml. Protein concentrating away from membrane. (No micro high concentration, ppt on membrane), works very nicely for a number of proteins.

Preserve Hampton Solutions
Cheryl Janson

To preserve Hampton solution when you or a co-worker gets a “hit” and suspect the Hampton solution may be “magic” or you cannot reproduce crystals with lab-made solution, make the reservoir solution from lab ingredients and use your homemade solution for reservoirs. Use the Hampton “magic” solution only for the drops thus using a few µl per experiment rather than O.5 ml. Saves having a 48 well screening solution set with one tube empty and the rest at 8 ml and still allows the superstitions or in Alex’s case, the contaminated solution to reproduce crystals.

Recycle Your Precipitate
Neali Armstrong

If your protein refolding reaction has a low yield and produces lots of precipitate try collecting the precipitate, resolubilize in GuHCl, and refold again. This material is sometimes more pure than the washed inclusion bodies.

90% Solutions
Anna Stevens

When optimizing or making solutions for a random scan, omit the buffer (@ final O.1 M) so that your stock is 90%. Prior to putting your stock in the well, add 100 µl of 1M buffer of your choice. Next, add 900 µl of 90% stock and mix. This reduces the number of tubes for crystallization (precipitant) stocks and allows flexibility in buffer identity and pH range. Be sure to make plenty of (~20 mL) of precipitant, you’ll need 900 µl per buffer.

Check Both Liquid Nitrogen + Stream Flash Cooling
Hans Parge

If your crystal does not freeze well in the cold stream try liquid nitrogen or vice versa.

To Determine the Optimal Concentration of Your Protein for Screen I and Screen II
Jaru Jancarik

Try (setup) drop no. 6 first (of Screen I). It should produce light precipitation if ppt appears too heavy, reduce the protein concentration by 1/2 and try again. If there is no precipitation in drop 6 try drop 4. If there is no precipitation in either of the drops concentrate the protein 2 fold and try again.

Cryoprotectant Additive
Laura Pelletier

I tried adding 1-10 mg/ml BSA in the cryoprotectant when soaking crystals that would crack. The one time it was used, the crystal did not crack and froze nicely. I don’t know if the BSA was the reason for successful freezing. I was wondering if anyone else has tried this.

Slow Cryosoaks for Improved Crystal Stability
Bryan Prince & Melissa Harris

In a sitting drop well, cryoprotectant (20% Glycerol in crystallization buffer) should be dribbled down the side of the well.

Crosslinking of Crystals
Clare Stevenson

When crystals are fragile or you want to transfer them to a different mother liquor e.g. for heavy atom soaking, why not try crosslinking your crystals with

0.1% glutaraldehyde in your mother liquor. This can be done by adding the glutaraldehyde directly to the drop or placing it next to the drop and allowing for vapour diffusion. This crosslinking enabled us to solve Mod A at 1.2 A resolution.

Know When Enough is Enough!

Brent Segelke

As a graduate student, I spent countless hours and quantities of protein trying to get crystals of a single construct. We never got crystals. Another group got the structure of a homologous protein that was auto-digestive. Had we stopped after 400 trials and altered our construct, perhaps we would have faired better, we couldn't have faired any worse. There is a reasonable statistical argument to demonstrate that 400 trials are a good limit.

Buffer Screening

Anil Mistry

To find the best buffer system which will keep your protein happy, stable, and soluble for concentration prior to crystallisation, setup your protein (at 1-2 mg/ml) in hanging drops over 1 ml well solutions containing a series of buffers at various pHs, with various additives/stabilizers, etc. (but no precipitant!). Checking for drops which are clear will give an idea of which solutions keep the protein happy. In addition, as the system attains equilibrium some in situ concentration of protein will be induced due to the bulk difference between the drop and well volumes, hence an idea of how the protein behaves upon concentration in this solution will also be observed. Modifying a clear or slightly clear drop by adding a higher concentration of buffer to the well may even produce crystals. However, the main piece of information this method can produce is an idea of which buffer system and additives to put your protein into prior to concentration a crystallisation screening.

Soft Crystals

Allan D'Arcy

If you have crystals that are very sensitive to being touched (they break) or stick to the glass or dish, use a pointed strip of parafilm to move them. Otherwise, grow the crystals on parafilm and punch "wells" around the crystals to move it.

Crystal Annealing

Clare Stevenson

As I said in my talk, give it a go. You might be surprised!!

"No More 4°C"

Marie Anderson

Prepare trays for crystallisation, leave at 4°C, and fill polystyrene box or flat container with ice. Imbed a metal plate in the ice and set out cover slips. When you're ready to set out crystallisations place the trays in the bed of the ice and prepare drops, when finished transfer to 4°C. Simple but it works.

Cross Seeding to Generate Crystals of a Related Protein or Protein/Inhibitor Complex

Margaret O'Gara

If your protein or protein complex fails to crystallize try seeding from crystals of the same protein (if it's a protein you want to crystallize) or a related protein for apo protein crystals. Serial micro seeding works best, make sure you look at the drops after seeding to identify any visible crystal seeds in there (i.e. not new crystals).

Advice for the Follically - Challenged Crystallographer

Jonathon Hadden

It has been long accepted that crystallizers with an abundance of facial hair have highly successful careers (Leeds University personal observation). This was thought to stem from the fact that matter found its way from the hair into the trials and acted as a nucleation centre-BUT SERIOUSLY; Addition of small grains of sand to a crystallization drop that is "close to producing" crystals can aid in nucleation and crystal growth.

Raise the DMSO?

Melissa Harris

If higher DMSO does not damage your protein, try higher concentrations of DMSO (10-20%) for crystallization. It can help when dealing with insoluble compounds and is an excellent cryoprotectant. Freeze directly from drop!

TCEP as a Reducing Agent

Barbara Brandhuber

Use TCEP instead of DTT as a reducing agent in your protein solution. It isn't oxidized as quickly as DTT. Be sure to watch the pH of your solutions because it is very acidic.

Temperature Variation

Irene Weber

To grow crystals at different temperatures around room temperature search the lab for spots that are consistently at higher or lower temperatures. A difference of several degrees can be found. Temperature shifts can be easily made by moving crystals to a different place. (Check office shelves too!) [Discovered by Charles Reed in my lab]

Drop Drying Technique

Anil Mistry & Ron Rubin

When setting up drops with a protein which has low solubility a low starting concentration has to be used. Setup larger drops (5-10 µl) and leave them to stand "dry" for 3-5 minutes at room temperature /4°C prior to inverting over a well of a Linbro® plate, this should allow some pre-concentration.

TIPS FROM RAMC

Using DLS to Test for Irreversible Aggregation

Anil Mistry

Generally, people concentrate a protein to 10 mg/ml or higher, then dilute 2-fold when setting up their hanging drops by doing a 1:1 minx. What if in concentrating to 10 mg/ml aggregation has been induced which is irreversible, such when pipetting your 1:1 mix hanging drop, it already contains aggregates, a bad starting point.

(For monodisperse protein samples)

Using DLS test the limit of concentrated protein, i.e., the maximum concentration that can be achieved before a polydisperse signal is obtained. Then test samples, up to this limit, by concentrating up to this limit and test for irreversible aggregation by diluting a concentrated sample to a number of levels and test for monodispersity. With the sensitivity of current DLS equipment even samples at 10 mg/ml should be measurable. In this way and within the limits of your DLS machine it should be possible to find out whether you will have aggregates when you dilute your concentrated protein 2-fold when setting up a hanging drop.

Pickled Crystals Unusual Additives! (A True Story)

Michael Hickey

Pickle juice was added as an additive to a mutant form of a crystallized native protein. The mutant could not be crystallized in near similar conditions of the native. By chance, the components of pickle juice were read and found to contain compounds used in crystallization (i.e. glycerol, PEG 400, citric acid, acetic acid, alum, and a few vitamins). This juice (Sweet & Snappy Vlassic brand) was filtered (0.45 µl and adjusted to neutrality. It was then added to various PEG's (that crystallized the native form) and set up with the mutant. Crystals formed after 1 week! Trying to "add back" single components of the pickle juice to determine which component was responsible gave no crystals, the pickle juice (~1%) was necessary. Hint/ Tip: Commercially available food/ detergent solutions ought not be discounted as additives for crystallization!

Improve Your Crystals in Size, Shape, and Quantity

Nham Nguyen

After you have crystals, open the coverslip, remove the mother liquid in the droplet, dissolve the crystal with 3 µl of H20 or buffer and add 3 µl of well solution. Close the coverslip. The crystal will appear again. My crystals diffract ~ 4 A, sometimes I get twins that diffract ~2.5 A.

No Hits! What Next?

Anil Mistry, Pfizer

No hits from a screen, what next? You can always set up another screen! Or, you can make a list of all drops that have precipitate and use the precipitate as a potential seed stock. Streak seed from the precipitate into new drops that have been set up at 50% and 75% of the original screening solution. There may be nucleation sites or microscopic seeds in the precipitates that may grow at lower precipitate saturation. Better still, streak seed from the precipitate to a clear zone into the same drop to recycle the drop.

Micro-Seeding With A Cryoloop

Anna Aagaard, AstraZeneca

For reproducible micro-seeding by hand use a cryoloop to fish out your seed from the seed stock and transfer them to the drop. Use a 0.3-0.4 mm cryoloop.

Reduce To Enlarge

Zhanna Druzina, SGX Pharma

For bigger crystals try to add 0.5-1 microliter of 14 M beta-mercaptoethanol to the reservoir after the protein drop was set up.

Complex Screening

Annie Hassell, GlaxoSmithKline

Problem: Can grow crystal but no protein-ligand crystals. Tip: Take the conditions from the apo crystals and develop a focused optimization screen (24 well maximum). Screen complexes using cross seeding and the focused screen and three drop ratios (1:1, 2:1 and 2:3).

Change the Method

Annie Hassell, GlaxoSmithKline

Problem: Poor crystal quality. Tip: Change tray type or crystallization method. For example, initial screens done in sitting drop tray and crystal quality improved in 96 well hanging drop tray.

Watch Out For Ruts

Brandon Collins, Boehringer-Ingelheim

Every project, every protein, every construct is unique. Be careful of knowing too much. Just because things did or did not work in the past does not mean things will work that way for the next project.

Heavy Atoms To The Rescue

Problem: Poor diffraction. Tip: Heavy atom soaks to stabilize floppy regions of the protein.

Don't But All Your Eggs In One Basket

Joseph Luft, Hauptman-Woodward Institute

Don't focus all of your optimization efforts on a single crystallization condition. If you have several different crystallization conditions identified for a sample go after them. Crystals of the same protein produced from different chemical conditions and/or temperatures will have unique physical properties. These properties will determine how easy the crystal can be looped (physical stability), cryoprotected and ultimately how well the crystal diffracts X-rays. Avoid single points of failure, go after several hits.

© 1991-2008 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

Insoluble Ligands

Doug Marcotte, Biogen IDEC

If ligands can't be soaked into crystal or co-crystallization is ligand specific try seeding into drops that contain ligand of interest. When soaking crystals with insoluble ligands try adding the cryoprotectant to the soaking solution. This can help solubilize the ligand and also cryoprotect (so less handling). Doesn't work so well when salt is the cryoprotectant, but may well when it's glycerol, DMSO or ethylene glycol.

Getting Away From Clear Drops

Annie Hassell, GlaxoSmithKline

Problem: Drops are all/mostly clear. Tip: Remove stabilizing agents (salt, glycerol, etc) from the protein buffer. Then do crystallization screens.

Avoiding Excessive Precipitation

Vaheh Oganesyan, MedImmune

At suboptimal protein concentrations the interface between protein solution and crystallization screen solution may exhibit excessive precipitation. To avoid this, before adding screen, solution add 1 microliter of water. Downside of this is equilibration will take slightly longer. Upside of this is decreased osmotic shock for protein and less precipitation.

Heavy Precipitate Control

Janet Newman, CSIRO

Put one conditions of 40% TCA into your standard screen. This should precipitate out all of your protein, so that you have an idea of what heavy precipitate should look like.

Look At The Big Picture

Edward Snell, Hauptman-Woodward Medical Research Institute

Don't consider a crystallization result in isolation. Look for neighbors in chemical space and use those results to provide chemical directions for optimization. If a cryocooled crystal does not diffract well. You cannot tell if it is the crystal, cryoprotectant or cooling that is causing the problem. Look at room temperature data before moving on.

Poor Nulceation

Mei Xu, Novartis

Consider the case of poor nucleation and seeding did not work. Tip: Mix protein and well solution then use pipette tip to cross the drop into branched shape. Crystal may grow in the branches of the drop.

Seeding Suggestion

Paris Ward, GlaxoSmithKline

To increase you choices of producing more optimal crystal condition or conditions using seeds, try the following. Program a small volume liquid handler to dispense your protein and seed solution directly into 96 well commercial screens and/or an additive screen.

Delete To Succeed

Heidi Schubert, University of Utah

Recent success with loop deletions. Sequence alignments reveal either charged loops and/or loop insertions relative to homologues. I have removed 3 to 34 amino acids and retained high expression soluble protein and novel crystals.

Seeking Stability

Annie Hassell, GlaxoSmithKline

Problem: Unstable protein. Tip: Add 1-5% low molecular weight (200 to 1,000) PEG directly to the protein and then screen.

Doh!

Jim Pflugrath, Rigaku

Start with big crystals! Add at least 5% glycerol to everything.

Stabilize For Cryo

Laura Pelletier, SGX Pharmaceuticals

Stabilize crystals in cryo by adding protein buffer components into the cryo. For example, most commonly I add 100 to 150 mM NaCl plus reservoir components plus cryoprotectant(s).

Ratios

Ayse Sinem Ozyurt, SGX Pharmaceuticals

Play with protein-mother liquor ratio, especially with low solubility proteins. Try different concentration of protein coupled with streak seeding.

Eliminate One Offs

Nancy Bump, Millennium Pharmaceuticals

After looking at the results of initial screening or of additive screen, pick several of the best "hits" and screen in 96 well format, 6 or 8 identical drops of each favorite before scaling up. Helps to eliminate "one offs" and save time.

Complex It!

Paul Reichert, Schering Plough Research Institute

Apo protein is monidisperse but won't crystallize. Complex it! Complex it! Complex it!

Matrix Seeding Tweak

Armando Villasenor, Roche Palo Alto

If your crystal seeds withstand large serial dilutions, try matrix seeding via the reservoir by doing the following. 1) Create crystal seeds as described by Allan D'Arcy. 2) Dispense seed into reservoirs containing reservoir. 3) Aispirate/dispense to mix seed in reservoir. 4) Dispense mother liquor droplet containing seed onto protein drop.

A Little Salt Please

Neil Grodsky, Pfizer Global Research & Development, La Jolla Labs

If you do not see any crystal growth in several days after set-up (more than 1 to 2 weeks) and the drop are not all clear, add salt (such as ammonium sulfate) to 0.5 M to the drops. Even though ammonium sulfate salt crystals might form, you might actually get protein crystals. This worked for me recently.

Urea To Solubilize

Gloria Borgstahl, Eppley Institute UNMC

Non denaturing (less than 3 M) concentrations of urea can be helpful to solubilize your protein.

Multivalents

Simon Low, Pfizer, Inc.

Try stoichiometric levels of multivalents, cations as additives. They may be necessary for crystallization but sometimes the levels found in commercial screens are too high and toxic.

Cooking for Crystals

Liping Wang, GlaxoSmithKline

Heat treatment of protein complex to obtain diffraction quality crystals. Original complex has no initial crystal hits. Heat treat protein complex (25 to 80 degrees Celsius) for various times (5 to 30 minutes). Centrifuge to get rid of aggregated protein and screen again.

Toothpick Tip

Paul Reinfelds, Illinois Institute of Technology

When using pre greased trays, take a toothpick and remove a bit of grease from each well. Now as you push down on your cover slip, you turn it a few degrees. This will allow the air to escape and the turn will form an airtight seal over each well.

Thermofluor To The Rescue

Jackie Day, Pfizer St Louis

Thermostability. Monitor the effect of additives, buffers, ligands, etc. on melt temp of your protein. We have seen in multiple cases that the most thermostable construct, buffers, additive yields the best or only crystals. How? We use Bio-Rad's iQ5 iCycler (a PCR instrument) as it has 5 sets of filters for excitational emission and hydrophobic dyes that fluoresce upon binding (protein unfolding).

Getting Out Of A Sticky Situation

Michael Wiener, University of Virgina

Problem: Crystals adhering to plastic of sitting drop plate, and mechanical dislodging (by cryoloop, tool, etc) does not work. Solution: Stan a fine gauge syringe needle into the plastic, near but not into the crystal. This often distorts/disrupts the plastic near the crystal and breaks the seal.

Liquid Bridge

Margarete Neu, GSK Stevenage, UK

Getting bigger crystal by low tech / low cost counter diffusion. If you are faced with either no nucleation or showers of crystals and the usual tricks including seeding do not work, try this: On a cover slide, set the protein and precipitant drop (example 1 microliter plus 1 microliter) separate, but very close to each other. Then, with a whisker or pipette tip streak through the drops to form a connecting bridge between the protein and precipitant solution. Invert cover slide and place over well. Crystals will form along the gradient and "self screen" for best conditions.

Low Concentration Complexing

Elizabeth Fry, Abbott Laboratories

When working with compounds for co-crystallization, if the compounds are highly insoluble in protein buffer (50 micromolar or less) we often employ low concentration complexing. We dilute the protein and then add in diluted compound, so tha the compound is added close, or at least closer to a concentration where it is soluble and the content of DMSO in the protein sample remains less than 2%. The protein-compound complex is then concentrated for crystallization trials. This has helped us with several projects with highly insoluble compounds.

Ionic Liquids

Christopher Bonagura, Exelixis, Inc.

In experiments using model proteins we found that Ionic Liquids (IL's) specifically 1-Butyl-2-methyl imidazolium chloride gave increased numbers of crystallization outcomes compared to the IL controls. A large number of the crystals obtained had precipitated outcomes in the IL controls. In many other cases the IL and crystal had an improved morphology (needles to plates, plates to 3D crystals) over the IL controls. Tip: Using an IL such as 1-Butyl-2-methyl imidazolium chloride as an additive to improve chances of getting a crystal from conditions which otherwise would give precipitate. Marc Pusey, MI Research, Inc. Our favorite cryoprotectant. 1x UCP (Ultimate Cryo Protectant) 8% Glycerol, 8% Ethylene glycol, 9% Sucrose, 2% Glucose. We make a 2x solution. Generally add this 1:1 with reservoir. The ratio can be modified, for example 1.2 microliter 2x UCP : 0.8 microliter reservoir, or 1.4 microliter 2x UCP : 0.6 microliter reservoir, or 0.6 microliter 2x UCP : 1.4 microliter reservoir, etc. I believe this has been successful in cryoprotecting some 70% of all of our systems, resulting in more than 50 solutions of these targets regardless of previous cryogenic treatments. Author in unknown to me, but the credit is published in a singled Hencrickson paper. I was tipped off 6 years ago.

Protect The Bridge

Shirley Robert, York Structural Biology Laboratory

If you can't get your Se-met protein to crystallize try leaving out the DTT/TCEP during purification and crystallization. Sometimes there are disulfide bridges near crystallization contacts that need preserving for crystallization to take place. Collected the Se edge is still possible.

DMSO For Cryo

Rich Romero, SGX Pharmaceuticals

Try 20 to 30% DMSO as cryo. This has worked well in a number of cases for me and I've added this to a very short list of cryos that I personally use. Note: If your mother solution contains a high salt concentration the DMSO will cause it to precipitate out of solution. So beware!

No Fog

Beat Blattmann, University of Zurich

4 degrees Celsius crystallization plates prepared at room temperature always have a condensation problem on the plate seal. To avoid condensation cover the finished plates with two lids and plate it in the cold room for 20 minutes or on top of a cold metal block, The will reduce the temperature of the reservoir solution while the 2 lids delay the temperature change from the top long enough to avoid condensation.

Anti Slip Tip

Barbra Pagarigan, Celgene

Ever been manually sealing a plate only have it slip out from under you? The result is usually death for hanging drop plates, and with sitting drop plates your best bet is hoping the drop is not splashed onto the seal above. To ensure the plate stays put when applying pressure, we use a "grip pad" in our lab. Simply place the grip pad onto the bench, set plate on top and seal as usual. The grip pad prevents the plate from slipping out from under the compression tool, usually a brayer, used for ensuring the please seal is applied correctly. The grip pad can be cut from the material commercially available for lining tool shop drawers. In addition, we created a fixed plate holder that encloses the entire plate to guarantee the brayer does not slip off the plate when sealing, a common occurrence when manually sealing many plates. Our plate holder is custom cut from a hard rubber to fit both thee 24 well and 96 well plates. The plates sit slightly above the platform to ensure both ends are sealed and for easy removal. The sides of the platform are rounded to ensure the brayer has a smooth path of travel.

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

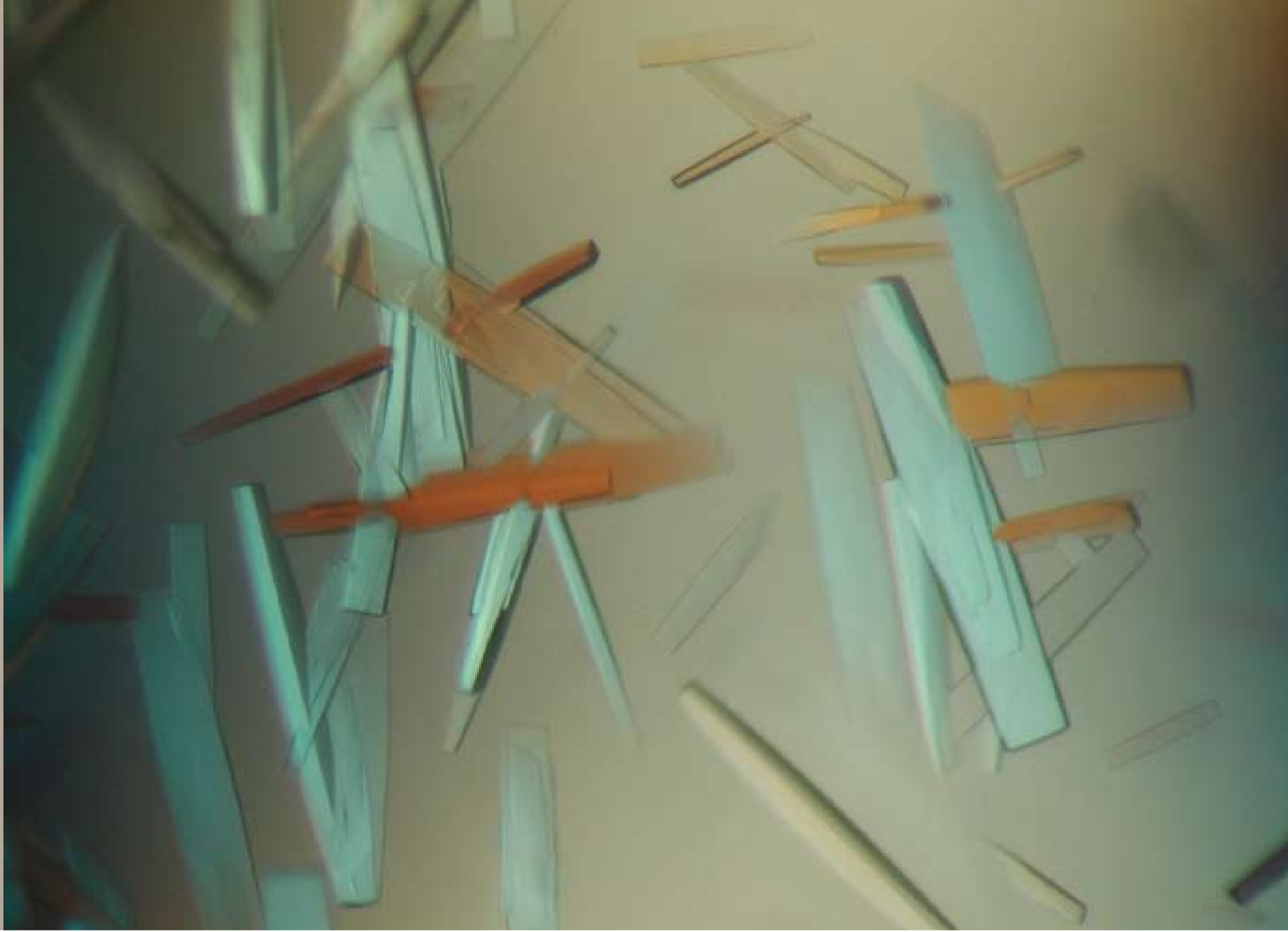

HAMPTON
RESEARCH

CRYSTALLIZATION TIPS FROM A TO Z

Crystals of apotransferrin (bovine milk) grown in fumaric acid, pimelic acid, sebacic acid, hexadecanedioic acid, maleic acid, glutaric acid, suberic acid, dodecanedioic acid, oxamic acid, 15% w/v polyethylene glycol 3,350.

Peter Nguyen, Hampton Research, USA.

A is for Aggregation

Aggregation can be a deterrent to the crystallization of biological macromolecules including proteins, peptides, and nucleic acids. The presence of sample aggregation can be detected by either dynamic light scattering or native gel electrophoresis. Aggregation might be caused by hydrophobic patches on the surface of the sample, differently charged isoforms, differently phosphorylated isoforms, mixtures of methylated and non-methylated samples, glycosylation, as well as electrostatic interactions. Aggregation can be due to autologous aggregation where the protein is aggregating with itself or heterologous contamination where the sample is aggregating with other proteins. In the case of heterologous contamination, further purification of the sample should be seriously considered. In the case of autologous aggregation that precludes crystallization, one might consider:

Using molecular biology to manipulate intra and inter molecule interactions by modifying the sample sequence (alter, add, or delete residues).

Using chemical additives, like those listed below, to manipulate sample-sample and sample solvent interactions:

- Detergents
- Chaotropes (urea, guanidine hydrochloride, hydrochloric acid, etc.)
- Electrostatic agents
- Alcohols (isopropanol, methanol, ethanol, etc.)
- Salts (sodium chloride, potassium chloride, sodium fluoride, etc.)
- Polyols (glycerol, PEG 400, etc.)
- Ligands, inhibitors, co-factors, and metals
- Use temperature to prevent aggregation (0°C and 60°C)
- Consider a fusion protein
- Remove C-terminus or N-terminus
- Truncate domains
- Remove His-tag

In some cases aggregates can be removed by centrifugation or filtration.

In some cases the aggregates can be removed by mixing the sample with the crystallization reagent, allowing the sample to incubate for 15 minutes, centrifuging the sample/reagent mixture, removing the precipitate, and setting the drop with the supernatant.

B is for Birefringence

The technical mumbo jumbo first. The physical properties of isotropic materials, such as glasses, liquids, and amorphous materials, do not depend on direction. However, most properties of a wide variety of crystals (including liquid crystals) do show such variation. This anisotropy of physical properties originates in the anisotropic build-up of the materials (crystal structure). Anisotropy in the optical properties of uniaxial crystals is referred to as either birefringence or dichroism, depending on whether the index of refraction or the absorption coefficient is concerned. Birefringence means that there are two distinct speeds with which light can propagate, depending on the direction of propagation. When a light ray splits into two beams as it passes through a material, the effect is called birefringence (or double refraction) and the material is birefringent. If you look at something through a birefringent material, you'll see double. The word birefringence comes from the Latin bi- (twice) plus refringere (to break up). Thus, the light rays are "broken in two" by a birefringent material. One well-known example of a birefringent medium is crystalline calcite (calcium carbonate). If you look at the world through a clear crystal of calcite (calcium carbonate), you will see double. Place such a crystal on a drawing, and you'll see two overlapping copies of the drawing. The molecular structure of calcite causes double refraction, in which each light ray is split into two rays that emerge from the crystal at slightly different angles. Calcite shows this more clearly than most crystals, but quartz and many other crystalline minerals also split light ray.

Now, the practical interpretation for crystal growers. You might hear the word birefringence used quite often by crystal growers when viewing crystals under a microscope. Here, crystal growers are stretching the definition of the term birefringence to describe the colorful display produced by biological macromolecular crystals when polarized light is passed through the crystal.

A light microscope with polarizing optics is required to observe birefringence. The following path is a typical setup. Light passes from the light source through the first polarizing lens, then the specimen (crystal) then the second polarizing optic, the magnifying optics and then into your eye. On many typical polarization setups, the second polarizing filter can be rotated while the specimen is stationary. Rotating the polarizing optic without something to rotate the plane of polarized light in the path (i.e. a crystal) will result in one seeing light, dark, light, dark, as the filter is rotated. But if a crystal with birefringent properties (i.e. a biological macromolecular crystal) is positioned in between the two polarizing filters, one will observe changing colors as the polarizing filter is rotated. Specifically, when the polarizing filters are aligned such that the field is dark, a birefringent object (crystal) will glow with color.

Birefringence is one way we can differentiate amorphous precipitate from microcrystals in a drop when viewed under a microscope. Precipitate does not have birefringent properties while most biological macromolecular crystals do.

One drawback with using birefringence in today's crystal growth world is that most of the crystallization devices utilized are made from plastic such as polystyrene and polypropylene. These plastics are optically active and can be birefringent. In fact, often times the colors we see displayed in crystals are contributions from the plastic birefringence. However, it is still possible to observe microcrystalline birefringence in the plastic trays, but there is usually a contributory effect from the plates used to grow the crystals. One way to avoid this is to grow crystals in a glass device or at least observe the crystals in a path that is free of birefringent plastics.

Birefringent precipitates will glow, sparkle, or glisten.

To test for birefringence, position the polarizers so the field of view is dark WITHOUT they crystallization plate or setup. Place the tray into position on the microscope. If a crystal is birefringent, some of the light passing through the crystal will be rotated and pass through the second (analyzing) polarizing filter. The intensity of the transmitted light will increase and decrease as the crystal or the polarizer is rotated. Remember, birefringence is not ALWAYS clearly visible when plastic is in the light path (i.e. when you use plastic slides or crystallization plates). However, a birefringent crystal viewed in a plastic tray or plastic cover slide will have a different color than the background (i.e. plastic plate) or foreground (plastic slide). Finally, birefringence is a property of crystals, both biological (proteins, peptides, and nucleic acids) and inorganic crystals (salts). Birefringence is MORE pronounced in inorganic crystals (salt).

A quick comment on what to do with birefringent precipitates. Streak seeding is a common and often successful method of taking advantage of microcrystalline precipitate to grow large single crystals. But that starts with an S so we cannot talk about that this time!

C is for Condensation

Condensation on cover slides or tape may be a problem when incubating plates at cool or warm temperatures.

To minimize condensation try the following: Place a dummy plate above and below the plates being used for crystallization. For example, on a stack of 5 or 10 plates, place a dummy plate on the bottom of the stack as well as the top of the stack. The reservoirs (wells) of the dummy plates should contain water as the reservoir solution. The reservoirs should be sealed either with plain glass cover slides and vacuum grease or clear sealing tape to prevent the reservoir solution from evaporating.

Pipet the reservoirs and drops at the temperature where the plates will be incubated. This means enjoying a cold room during setups. One can cheat and avoid a cold room by incubating reagents, pipet tips, plates, and slides at the cold temperature. When ready to set the experiment, set the plates in an ice bath and quickly set the plate while the plate (as well as reagents and protein) are on ice. Conceivably one could also do this for warm temperature incubations where an elevated temperature incubator is available. The idea is to avoid a RAPID change in temperature with the sealed crystallization experiment (plate). By minimizing the temperature difference between the setup and incubation temperatures, condensation can be minimized.

Try sitting drop vapor diffusion. Condensation is especially bad for hanging drop vapor diffusion experiments since water condensing on the slide can run or merge into the crystallization drop. To avoid such a merger, try sitting drop vapor diffusion. Condensation on the tape or cover glass above a sitting drop experiment may still make viewing the drop difficult, but at least the drop will not become diluted with condensation unless the condensation is so bad that it rains in your experiment!

Finally, avoid vapor diffusion altogether and go microbatch.

D is for Drafts

Temperature is often a significant and overlooked crystallization variable. Temperature can sometimes be an unwanted crystallization variable in labs subject to temperature changes. Temperature fluctuations are especially troublesome because unless they are intentional and accurately produced and recorded, the temperature changes are difficult or impossible to reproduce. Keep the following in mind when selecting a location to store crystallization experiments:

How does the temperature of the room fluctuate over the course of a day, over the weekend, during the month, over the year?

Place a thermometer in a container full of water (allowed to equilibrate to the location temperature) and observe the temperature fluctuation over time to determine the temperature stability of the location.

Is the area subject to drafts such as opened and closed doors, under a heating or air conditioning duct, or near a heat producing device?

Avoid storing experiments near windows or walls that are in contact with the outside of the building.

Is the heating or air conditioning turned off in the evening or on weekends in your facility? If yes, substantial temperature changes can occur, especially during very cold or hot days/nights.

In an effort to maintain a more constant “room temperature”, store the plates in a quality, commercial incubator set at the desired “room temperature”. Since most of us cannot swallow spending money for a room temperature incubator, consider storing the plates in an unplugged or non-working refrigerator, styrofoam box, ice box, or cooler.

Use a quality incubator to store crystallization experiments.

E is for Ethylene Glycol

Ethylene glycol. Also known as 1,2-Ethanediol as well as glycol. Molecular mass of 62.07. HOCH$_2$CH$_2$OH. A nice little polyol that can be useful as a cryoprotectant. Typically used as a cryoprotectant in the concentration range of 10 to 30% v/v. Also useful as a non volatile organic additive when used in the concentration range of 1 to 5% v/v and in some cases, much higher (15 to 25% v/v). As an additive, ethylene glycol has been reported to sometimes increase the size and quality of the crystal. Finally, it is useful as a crystallization reagent in the concentration range of 25 to 30% v/v. It is even thought that ethylene glycol can have a structure-stabilizing effect and may be useful in the crystallization of flexible proteins.

F is for Fab Fragments

Fab and Fv fragments can be used as co-crystallization agents.

Fab fragments have decent solubility properties and bind specifically to antigens with reasonable equilibrium constants (105 to 108 M-1). Fab fragments can, in some instances, effectively transform aggregated protein into soluble, monodisperse protein, suitable for crystallization trials. Since antibodies are frequently available from related biochemical studies, they can often be applied more readily than molecular biology (point mutants, truncations, molecular engineering). However, it seems MOBO gets faster and easier with each passing year. It is often useful to consider screening several different Fab fragments that recognize different epitopes for crystallization trials. Sometimes Fab fragments can immobilize a region of a protein sample, reducing sample flexibility, enhancing conformational homogeneity of the sample which in turn can enhance chances for crystallization.

Using an Fv fragment in lieu of a Fab fragment might have some advantage since there is no flexible elbow to inhibit crystallization.

When preparing antigen-antibody complexes for crystallization, one should carefully select the antibody, prepare a homogeneous Fab species, and prepare the Fab-sample complex with proper stoichiometry. Ideally the Fab should not interfere with the native sample conformation or activity.

The presence of a Fab fragment in the crystal can assist in the crystallographic structure determination by allowing one to utilize molecular replacement.

Fab and Fv fragments should be considered useful tools in the crystallization toolbox for manipulating sample solubility, conformational flexibility, and crystal lattice contacts, as well as being a possible tool for molecular replacement.

Reference:

The use of antibody fragments for crystallization and structure determinations. LC Kovari, C. Momany and MG Rossman. Structure 15 December 1995, 3: 1291-1293.

G is for Gap

When greasing crystallization plates, it is often useful to leave a small gap in the bead of grease applied to the plate. For example, begin greasing at 12:00 and apply the bead clockwise until 10:00. A gap with no grease is left between 10:00 and 12:00. Place the cover slide onto the bead of grease. Depress the slide onto the grease. As the slide is pressed onto the bead of grease, pressure is allowed to escape through the gap. Before the slide bottoms out on the plate, twist the slide approximately 15 minutes (90°) to smear the grease across the gap and properly seal the cover slide to the plate.

H is for 1,6 Hexanediol

1,6 Hexanediol ($C_6H_{14}O_2$ M_r 118.18) is a non-volatile alcohol which most likely precipitates biological macromolecules by lowering the chemical activity of water. Non-volatile organics can be effective crystallization reagents for proteins and are particularly appropriate for nucleic acids. This non-volatile alcohol steals water molecules from the biological macromolecule through hydrogen bonding and reduces the electrostatic screening effectiveness of the water. Typical stock solutions of 6.0 M (about 71% w/v) are typical for this crystallization reagent. Typical reservoir concentrations for this reagent are 1 to 4 M across a broad pH range. 1,6 Hexanediol can be utilized in the presence of numerous anions and cations, as well as a broad range of salts. It can be a substitute crystallization reagent for MPD, PEG 400, and PEG MME 550. It can also be utilized as an additive in the concentration range of 0.1 to 0.2 M. 1,6 Hexanediol is a waxy solid or flake. When formulated with water, the reaction is endothermic and one might need to warm the solution to effect complete dissolution at high levels of supersaturation.

1,6 Hexanediol is utilized as a precipitant in Crystal Screen 2 and is available as a preformulated, sterile filtered reagent from Hampton Research (Catalog Number HR2-625; 6.0 M 1,6 Hexanediol, 200 ml).

I is for Interface Diffusion

Crystallization by interface diffusion was popularized by Ray Salemme (F.R. Salemme 1972, Arch. Biochem. Biophys. 151, 533). The method is also called free interface diffusion and liquid-liquid diffusion. The methodology is straightforward. Interface diffusion is classically performed in a cylinder with a small diameter, such as a capillary. Using a capillary with both ends open, the sample is pipetted into the capillary. Next, the crystallization reagent is added to the capillary. For true interface diffusion to occur, the sample and the reagent should touch one another inside the capillary. This can be accomplished by making sure the sample is positioned at the inside end of the capillary before adding the crystallization reagent. When adding the crystallization reagent, use a pipet tip with a small diameter opening such as a gel loading tip. Carefully and slowly pipet the reagent from the pipet tip and allow a bead of the reagent to form at the tip. Hold the capillary parallel to the ground and gently touch the tip of the capillary to the reagent bead. With a combination of gently pipetting, capillary action and proper capillary tilt, fill the capillary with the appropriate amount of crystallization reagent without air gaps. Done correctly, the inside of the capillary should contain a single cylinder of liquid. Filling the capillary should be done gently and slowly to minimize mixing the sample and the reagent. Actually, one should add the more dense solution into the capillary first, followed by the less dense solution to minimize mixing. If done properly, one can see an interface between the sample and the reagent. It might be best to start with equal amounts of sample and reagent. Later, one can vary the ratio of sample to reagent for optimization. Once filled with reagent and sample, the capillary can be sealed with a non-drying clay or wax (sealant). Be careful not to add too much sealant at a time to the end of the capillary since the sealant will displace the liquid in the capillary and too much sealant will result in liquid being lost from the capillary. Storage of the capillary can be done in several ways. Some like to simply stick the capillaries vertically into clay or wax. But this means one must remove and contort the capillary for viewing under a microscope. Instead, one can use a petri dish (square ones work really well) and line the petri dish with two thin lines of wax or clay about the size of a toothpick. The lines of wax or clay should be parallel to one another and positioned apart from one another just under the length of the capillary. The capillaries can be set upon the lines of clay in neat rows. A dozen or so capillaries can be stored in a small square petri dish. Storage and viewing is quick and simple. The petri dish also facilitates easy labeling.

A nice feature of interface diffusion crystallization is the potential of a concentration gradient of sample and reagent along the length of the capillary. Although pipetting, gravity, and movement make this interface and gradient less than perfect, gradients do form and one can often visualize the concentration gradient as crystals of different size and number form at different points along the gradient (length of capillary), presumably due to the change in relative supersaturation along the length of the capillary. For example, numerous, small crystals might form at high levels of supersaturation, fewer and larger crystals might form at ideal levels of supersaturation, and no crystals may form at very low levels of supersaturation.

Permutations of interface diffusion include the following. Doing it in microgravity. Freezing one layer before adding another to ease pipetting (good for larger volumes). Freezing the entire experiment then slowly melting the slug to introduce temperature as a crystallization variable. Placing a gel plug between the sample and the reagent. Using high throughput plates with tall, thin cylinder-like wells for “vertical” interface diffusion.

J is for Jeffamine® Reagent

The Jeffamine polyoxyalkyleneamines contain primary amino groups attached to the terminus of a polyether backbone. They are thus “polyether amines.” The polyether backbone is based either on propylene oxide (PO), ethylene oxide (EO), or mixed EO/PO. Jeffamines are synthesized as either monoamines (M-series), diamines (D series), or triamines, and are made in a variety of molecular weights, ranging up to 5,000. The ED-series are aliphatic diamines structurally derived from the propylene oxide capped polyethylene glycol.

The wide range of molecular weights, amine functionality, oxide type, and distribution provides flexibility in synthetic design of compounds made from Jeffamine Reagents. For the most part, Jeffamine Reagent products undergo typical amine reactions and are low viscosity liquids, exhibiting low vapor pressure.

Jeffamine® Reagents originated chemically at the Texaco Chemical Company as lubricants and fuel additives and are now most frequently used in manufacturing adhesives, coatings, epoxies, and curing agents. Scary what we crystal growers will use as crystallization reagents, is it not?

Jeffamine Reagents worked their way into the protein crystallization community in the late 70's when Dulio Cascio (then in the lab of Alexander McPherson at the University of California Riverside now University of California Los Angeles) evaluated them as polymeric precipitants "similar" to the polyethylene glycols.

Alexander McPherson published the use of Jeffamine Reagents in a crystallization publication "Two approaches to the rapid screening of crystallization conditions" J Crystal Growth 1992, 122: 161-167.

The first protein structure solved using Jeffamine Reagents was that of Xylose Isomerase in the publication by Lloyd et al, "Crystallization and preliminary X-ray diffraction studies of xylose isomerase from Thermoanaerobacterium thermo sulfurigenes strain 4B" J Mol Biol 1994, 240: 504-506.

Jeffamine Reagents are discussed by Lesley Lloyd Haire and others in Terese Bergfor's wonderful book "Protein Crystallization - Techniques, Strategies, and Tips, A Laboratory Manual (IUL Biotechnology Series 1999 ISBN 98-075232).

The following Jeffamine Reagents have been tried in protein crystallization experiments:

- Jeffamine D-230 Reagents
- Jeffamine D-400 Reagents
- Jeffamine ED-600 Reagents
- Jeffamine ED-900 Reagents
- Jeffamine ED-2001 Reagents
- Jeffamine M-600 Reagents

Hampton Research has experienced the most success with Jeffamine ED-2001 Reagent and Jeffamine M-600 Reagent.

Jeffamine Reagents can be used "like polyethylene glycols" in the crystallization of proteins, peptides, and nucleic acids. In fact, some crystal growers substitute Jeffamine Reagents for PEGs during optimization to see if using them can improve the crystals or offer new conditions for further optimization.

Jeffamine Reagents can be formulated with most of the salts, buffers, organic solvents, and additives used in biological macromolecular crystallization.

There is very little literature describing the use of Jeffamine Reagents as crystallization reagents and not a single report of them being used as a crystallization reagent in the Biological Macromolecular Crystallization Database at the time of this writing.

(http://wwwbmcd.nist.gov:8080/bmcd/bmcd.html)

Obtaining crystals in Jeffamine Reagents is a mixed bag. It is a good thing since there are crystals. It is a bad thing since they are a pain to formulate.

The formulation of Jeffamine Reagents as crystallization reagents is tricky and tedious. They are very alkaline chemicals and must be titrated to neutrality or the desired experimental pH before use. Titration is typically performed using hydrochloric acid. A significant amount of hydrochloric acid must be used to titrate the Jeffamine Reagents to pH values between 2 and 11. pH changes are slow until pH 9, and change very rapidly as one approaches neutrality. The addition of hydrochloric acid results in a significant temperature increase in the reagent. Accurate pH recordings require repeated pH, cool, pH, titrate; repeat cycles. Many salts, including anions and cations, can be used in conjunction with Jeffamine Reagents. Typical salt/Jeffamine Reagent ratios and concentrations are similar to those used with "related" polyethylene glycols. Phase separation can be a bigger problem with Jeffamine Reagent / salt mixtures than for PEGs, but can be minimized by titrating the Jeffamine Reagent to neutrality or close to the final working pH before adding the salt and/or buffer. Having as much water present in the formulation also helps prevent phase separation.

Jeffamine M-600 Reagent is utilized as a crystallization reagent in Crystal Screen 2 (Catalog Number HR2-112).

Hampton Research offers the following preformulated, ready-to-use Jeffamine Reagents:

Jeffamine ED-2001 Reagent 50% w/v solution pH 7.0, 200ml sterile filtered (HR2-597)

Jeffamine M-600 Reagent 50% w/v solution pH 7.0, 200ml sterile filtered (HR2-501)

Hampton Research also offers the following Jeffamine Reagent which has not been titrated:

Jeffamine M-600 Reagent 100% solution pH >12, 200 ml (HR2-503)

Jeffamine is a registered trademark of the Huntsman Petrochemical Corporation.

K is for Krypton

Named from the Greek kryptos or "hidden", krypton is neither green, nor a solid material that can defeat Superman. Rather it is another noble gas discovered in 1898 by Ramsay and Travers. It ranks sixth in abundance in the atmosphere. Krypton gas is used in various kinds of lights, from small, bright, flashlight bulbs to special strobe lights for airport runways. As with the other noble gases, krypton is isolated from the air by liquefaction.

Krypton and crystals. The use of xenon and krypton at high pressure is becoming a popular method for the phasing of proteins. There are several homemade as well as commercially available pressure cells for creating xenon and krypton derivatives.

In essence, the crystal is exposed to the krypton or xenon gas at high pressures for a short period of time in order to allow the krypton or xenon to bind to the protein within the crystal. Following incubation and depressurization, the crystal is flash cooled in liquid nitrogen and mounted onto a goniometer in a cryostream for X-ray diffraction analysis. Freezing of the crystal is necessary to prevent the diffusion, release and loss of gaseous krypton or xenon. As a side note, methods of generating and analyzing xenon derivatives in glass or quartz capillaries have been described (see references below).

Krypton and xenon binding sites are generally different from those for other heavy atom sites so the screening of krypton and xenon can be used as a follow up when the initial soaks in heavy atoms are not successful.

When the specialized hardware is available for screening krypton and xenon under high pressure, the method is a convenient and rapid way for screening successful derivatives for phasing.

References:

- *High-pressure krypton gas and statistical heavy-atom refinement: a successful combination of tools for macromolecular structure determination. Schiltz, M., Shepard, W., Fourme, R., Prange, T., de La Fortelle, E. and Bricogne, G. Acta Cryst. D53: 78-92, 1997*
- *Exploring hydrophobic cavities in proteins using xenon and krypton noble gas. Prange,T., Schiltz, M., Pernot, L., Colloc'h, N., Longhi, S., Bourguet, W. and Fourme, R. Protein Struct. Funct. and Genetics 30(1): 61-73, 1998.*
- *Solubility of krypton and xenon in blood, protein solutions, and tissue homogenates. Yeh, S.Y., Peterson, R.E. J Appl Physiol 5: 1041-1047, 1965.*
- *Flash freezing isomorphous xenon or krypton derivatives of protein crystals. Sauer, O., Dutzler, R and Kratky, C. ECM 17, Seventeenth European Crystallographic Meeting (Lisboa, Portugal 24/28 aug.1997). Book of Abstracts p 18 (ref. MS1.6-4)*
- *Xenon and Krypton at LURE http://www.lure.u-psud.fr/sections/Xenon/XENON.HTM*
- *Tutorial for Krypton-Elastase SIRAS refinement http://utica.med.jhmi.edu/sharp/tutorials/KrEl.html*

L is for Lyophilization

Avoid lyophilization of your protein for storage and concentration if you plan to use it for crystallization experiments. Lyophilization is a not-so-gentle procedure and can prevent crystallization. Yes, we know that lyophilized lysozyme will crystallize quite readily as well as other "Sigma proteins" but we have seen, as well as others, many a protein that will not crystallize following lyophilization. And we have yet to see a protein that would only crystallize after lyophilization. So don't go there.

If you have a lyophilized protein, there is hope. First of all, one should solubilize the protein in water or a stabilization buffer and dialyze the protein exhaustively against water or the stabilization buffer. Dialysis is an important step which will help to remove residual, non-volatile buffers and reagents, as well as low molecular contaminants.

L is for Lipidic Cubic Phases

Membrane proteins can be crystallized in a lipid phase where the crystallization detergent diffuses into the lipid phase and crystal growth proceeds through lateral diffusion of the protein molecules.

References:

- *Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. Proc Natl Acad Sci 1996, 93: 14532-14535*

Related Reference:

- *X-ray crystallography of membrane proteins: Concepts and applications of lipidic mesophases to three-dimensional membrane protein crystallization G PROTEIN-COUPLED RECEPTORS. 2000. p.365-388*

L is for Light Scattering

Dynamic light scattering, also known as DLS, also known as photon correlation spectroscopy, also known as quasi-elastic light scattering, can be utilized to screen crystallization candidates for monodispersity.

DLS measures the translational diffusion coefficient of a macromolecular undergoing Brownian motion in a solution. What? In essence, DLS measures the intensity of light scattered by molecules in solution. In turn, this measurement can tell you the size distribution of the protein molecules in solution.

Crystal growers should use DLS to differentiate monodisperse solutions from polydisperse solutions. Homogeneous, non-aggregated monodisperse proteins have a high probability of producing crystals. Proteins with non-specific aggregates and heterogeneous samples are less likely to crystallize. DLS allows one to screen protein samples quickly, with small volumes of sample, at different temperatures. The procedure is non-invasive so the protein can be recovered. If DLS is not an option, one can obtain similar results using native polyacrylamide gel electrophoresis or size exclusion chromatography. But DLS is faster and more sensitive.

For a great story on DLS, as well as practical information on methods for DLS, read Terese Bergor's chapter on Dynamic Light Scattering in Protein Crystallization - Techniques, Strategies, and Tips - A Laboratory Manual.

M is for Mass Spectrometry

Mass spectrometers (MS) use the difference in mass-to-charge ratio (m/e) of ionized atoms or molecules to separate them from each other. MS is therefore useful for quantitation of atoms or molecules and also for determining chemical and structural information about molecules. Molecules have distinctive fragmentation patterns that provide structural information to identify structural components.

MS can be a nice tool along with Dynamic Light Scattering (DLS), sodium dodceyl sulfate polyacrylamide gel electrophoresis (SDS PAGE), iso-electric focusing (IEF), and size exclusion chromatography (SEC) to evaluate the purity, homogeneity, and monodispersity of the sample prior to crystallization. Likewise, if crystallization problems exist, MS can be a nice tool to identify possible sources of the problem. MS can also be used to detect and possibly identify impurities in crystallization reagents.

In the case of recombinant protein samples, an accurate measurement of the molecular weight will inform one about the presence of post-translational modifications. Dependent on the physical properties of the compound and the MS technique used, the molecular weight may be determined within an accuracy of one Dalton. The use of MS measurements will give sequence information where classical methods fail or may lead to ambiguous results as in the case of N-terminal blocked peptides, glycosylation or identification of the C-terminus. Next to this, MS is the ideal analytical method to support sample clean-up procedures and as a routine check for sample impurities of all kinds. As for sample requirements for MS, the samples should be free of non-volatile salts and the procedure typically requires between 1 ug and several nmoles, depending on the kind of analysis performed.

N is for Nucleus

For a crystal to grow, the system must be supersaturated, in non-equilibrium. But first, a stable nucleus must form. A stable nucleus is an aggregate of the macromolecule. A stable nucleus is an aggregate of such size, shape, and properties that it will enlist new molecules into it's growing surface faster than others are lost into solution.

N is for NIST/CARB BMCD

The NIST/CARB BMCD is a Biological Macromolecule Crystallization Database. The internet address for this freely accessible database is:

http://wwwbmcd.nist.gov:8080/bmcd/bmcd.html

The BMCD contains crystal data and the crystallization conditions, which have been compiled from literature. The current version of the BMCD contains 3547 crystal entries from 2526 biological macromolecules for which diffraction quality crystals have been obtained. These include proteins, protein:protein complexes, nucleic acids, nucleic acid: nucleic acid complexes, protein: nucleic acid complexes, and viruses.

Use the BMCD to see if your sample has been crystallized or to see if a related sample has been crystallized. Use this information to design crystallization experiments for your sample.

N is for Nuclease Inhibitors

Nucleases can be a real pain during the crystallization of nucleic acids. Here are some nasty things nuclease can do to your sample: Modification of sample size, charge, or hydrophobicity, partial or total loss of activity, or utter desctruction of the sample. Traces of nuclease can be difficult to detect even when overloading electrophoresis gels and can obviously cause sample damage during purification, concentration, and storage of samples. What is one to do? Well, one can include a nuclease inhibitor in the prep or sample to protect the sample from ribonuclease and deoxyribonucleases. Inhibitors of nucleases include RNasin (from Promega), ribonucleaoside-vanadyl complexes, and DEPC. Inhibitors of deoxyribonucleases include DEPC and chelators such as EDTA or EGTA.

O is for Osmium

Osmium is a good reactant for ribose moieties and the 3' terminus of RNA for heavy atom derivatives.

O is for Organic Solvents

The most popular volatile organic solvents used in biological macromolecular crystallization have been ethanol, acetone, isopropanol, tert-butanol, 1,3-propanediol, acetonitrile, DMSO, methanol, and 1,3-butyrolactone. Organic solvents can be utilized as a primary precipitant (buffered or unbuffered), as a secondary precipitant in the presence of salt or polymer (primary precipitant), or as additives. The most popular nonvolatile organics have been MPD and 1,6-hexanediol. Organic solvents act as precipitants by lowering the chemical activity of water. This means they steal water molecules from biological macromolecules in solution, through a process of hydrogen bonding. This in turn reduces the dieletric constant of the solution. Current popular thought is that organic solvents should be used at low temperatures (4°C or lower) and at the lowest possible ionic strength, keeping in mind to include whatever is necessary to stabilize the sample (buffer, divalent cations, etc.).

P is for Protease Inhibitors

Proteolytic modification of proteins can be a tool for making small active fragments of proteins, which might have enhanced solubility characteristics compared to the native protein, which in turn might make the protein more amenable to crystallization.

However, proteases can also be trouble makers when they are uninvited guests to a protein sample. When protease contamination can be a problem, one can consider adding a protease inhibitor to prevent cleavage of the sample by these nasty proteases. Protease contamination can occur when these nasty beasts are carried over from isolation and purification of the sample from a natural source or expressed proteins. Also, if your lab is situated in an area (on the same floor, building, etc.) where microbial research, plant research, or other work involving the possible generation of fungal and bacterial organisms is possible, then one might consider adding protease inhibitors to the sample for protection. Why? The presence of fungal or microbial contamination in your sample, reagents, or crystallization related plates, capillaries, etc. can lead to growth of such organisms with subsequent release of proteases from these organisms so they can use your innocent little sample for food. Some crystal growers like to include sodium azide or thymol in all their reagents and sample as a deterrent to microbial growth, which prevents the possibility of microbial agents growing and secreting proteases into the sample solution. But sodium azide and thymol can sometimes bind the sample, are toxic, and in some cases do not live well with heavy atoms so some feel it is not wise to simply include these agents as a multi-vitamin in the crystallization cocktail. Rather, clean workspace, sterile filtered samples and reagents, sterile pipet tips, and good technique go a long way to preventing microbial contamination. Well, back to the P for Protease Inhibitors. Here are a few protease inhibitors and their targets.

Protease:	Metalloproteases
Inhibitor:	Chelators like EDTA and EGTA, bestatin, amastatin, thiol derivatives, hydroxamic acid, phosphoramidon.
Protease:	Aspartic Acid Proteases
Inhibitor:	Pepstatins and statin derived inhibitors.
Protease:	Cysteine Proteases
Inhibitor:	Thiol binding reagents, peptidyldiazomethanes, epoxysuccinyl peptides, cystatins, peptidyl chloromethanes.
Protease:	Most
Inhibitor:	DEPC
Protease:	Serine Proteases
Inhibitor:	Trypsin inhibitors, leupeptin, boronic acids, cyclic peptides, DIFP, PMSF, Pefabloc, aminobenzamidine, 3,4-dichloro isocoumarin, chymostatin.

Q is for Quintillion

In the USA and France, quintillion is the number 1 followed by 18 zeroes. In the UK and Germany, it is the number 1 followed by 30 zeroes. Use this term to impress your lab mates, as in "Hey, have you tried my new screen with a quintillion different conditions?" Just think of it, a screen with a quintillion conditions. Gotta get a hit with that one. Problem is, most of us are lacking a quintillion mg of protein...

Q is for QELS (Quasi Elastic Light Scattering)

Another name for DLS (see D).

Q is for Quiet

Vibration has been reported to lead to excessive nucleation and crystals of questionable quality. It is often suggested that crystallization experiments be incubated in a location with minimal vibration. Avoid cabinets that are frequently opened and closed, or countertops where equipment such as centrifuges or vortexes live. Incubators can be another sore spot for vibration, especially poorly insulated incubators with compressors that are not well isolated.

R is for Reducing Agents

Reducing agents are substances that cause other chemical species to be reduced or gain electrons. In order for reducing agents to cause the gaining of electrons on some other chemical species, they must undergo oxidation. Therefore reducing agents undergo oxidation when they do their job.

Dithiothreitol (DTT), beta-mercaptoethanol (beta-me), and Tris(2-Carboxyethyl)-Phosphine Hydrochloride (TCEP HCl) are sulfhydryl protective reducing agents. Reducing agents are typically used to prevent the oxidation of free sulfhydryl residues (cysteines) in the protein. Such oxidation can lead to non-specific aggregation of the sample, sample heterogeneity, inactivity, or denaturation of the sample. In a typical crystallization experiment, reducing agents are used in the concentration range of 1 to 10 mM in the crystallization drop.

Beta-me has one sulfhydryl group and is the weakest of the three reducing agents discussed here, lasting perhaps two to three days. The supply of beta-me should be replenished every two to three days in the crystallization experiment to maintain the effectiveness of the reducing agent. Since beta-me is volatile, it can be added to the reservoir of vapor diffusion experiments for diffusion into the crystallization drop.

DTT has two sulfhydryl groups and lasts about three to seven days in a typical crystallization experiment. DTT is not a strong volatile like beta-me (although it does possess a strong odor) and should be added directly to the crystallization drop when possible. Another consideration when working with DTT is to use the microdialysis method for crystallization since DTT can be added to the dialysis solution to replenish the supply of reducing agent.

TCEP HCl is stronger than both beta-me and DTT, lasting about 2 to 3 weeks in a typical crystallization experiment. TCEP hydrochloride can acidify the crystallization solution. Like DTT, TCEP HCl needs to be added directly to the crystallization drop, hence, microdialysis is a consideration.

Beta-me, DTT, and TCEP HCl can behave differently. Like all additives, if you find a class of compounds has an effect on sample stability or crystallization, then screen a variety of compounds in that class to see which one is best for your application. What we are trying to suggest is that like all additives, one might consider evaluating all three reducing agents to see which one works best for your sample.

There have been reports where reducing agents have been used as successful crystallization additives where there were no free sulfhydryls in the sample.

Reducing agents can bind metals and trace metal compounds, inactivating the reducing agent and the metal. This can make heavy atom derivatization in the presence of reducing agents, difficult and frustrating. EDTA can be added to the crystallization experiment to avoid inactivation of the reducing agent by metals. Keep in mind that if your sample needs metals for activity or stability that EDTA will keep metals from your sample.

When working at an alkaline pH, beta-me and TCEP HCl are more stable than DTT. TCEP HCl is more stable at acid pH than DTT.

DTT reduces nickel ions and can cause problems when purifying His-tagged proteins. To avoid this complication, try beta-me or TCEP HCl as a reducing agent instead of DTT when purifying His-tagged proteins.

L-cysteine is also a reducing agent. It's usefulness in crystallization is limited since it likes to form small hexagonal plate shaped crystals. L-cysteine can be a useful crystallization additive, but keep those pesky hexagonal plates in the back of your mind when working with L-cysteine.

If oxidation of the sample is expected or anticipated, reducing agents should be present during the preparation of the protein (when possible) and should be included, added, or replenished during the final preparation of the sample for crystallization. In vapor diffusion experiments, the reducing agent can be included or added to the crystallization reagent in the reservoir to minimize or prevent dilution of the reducing agent in the sample drop.

When deciding on the appropriate concentration of reducing agent for the sample, consider the number of free sulfhydryls in the sample as well as sample concentration. More free sulfhydryls and higher sample concentration mean that one should consider using higher concentrations of reducing agent.

Arsenical compounds are not compatible with reducing agents. Cacodylic acid, or sodium cacodylate is an arsenical compound and popular crystallization buffer and is not compatible with reducing agents. Other compounds not compatible with reducing agents include ammonium nitrate, hydroperoxide, potassium perchlorate, sodium nitrate. Check the Material Safety Data Sheet (MSDS) for these reagents for specifics about the incompatibilities of these chemicals with reducing agents.

S is for Sample Preparation

Preparing the Protein for Crystallization

Lyophilization

Avoid lyophilization. Even though there are many examples of proteins which crystallize after lyophilization (lysozyme, thaumatin, hemoglobin), lyophilization is to be avoided when possible. If the protein is lyophilized, it needs to be dialyzed before crystallization. Dialyze the protein against deionized water or a stabilization buffer before crystallization. Dialysis will remove non-volatile buffers and other chemicals which may have been present before lyophilization.

Ammonium Sulfate Precipitation

Avoid using ammonium sulfate precipitation as a final purification and/or concentration step. It is often very difficult to completely remove all the ammonium sulfate by a desalting column of dialysis. The remaining trace amounts of ammonium sulfate can interfere with crystallization screening results and create reproducibility problems. It is not uncommon for trace amounts of ammonium sulfate in the sample to cause precipitation or excessive nucleation in screen conditions containing polyethylene glycol and salt.

Batches

Avoid combining different purification batches for crystallization trials. Purification conditions and procedures are never identical so each batch should be screened separately.

Profile the Protein

Ideally, you will purify your own protein, but this is not always reality. So, it is always a good idea to characterize your protein before beginning crystallization experiments. Profiling your protein before crystallization can often provide valuable clues during screening and optimization of crystallization conditions. Assays to seriously consider:

- SDS-PAGE
- Native PAGE or
- Dynamic Light Scattering
- IEF (Isoelectric Focusing) Gel
- Mass Spectroscopy

The results of these assays can:

- Determine the purity of the sample
- Determine the homogeneity of the sample
- Identify batch-to-batch variations
- Identify stability problems with the sample

How Pure?

How pure should the protein sample be for crystallization trials? As pure as possible. That's some answer, is it not? Integrating common sense into the question, we might arrive at the following answer. For initial screening, the sample should be at least 90 to 95% pure on a Coomassie stained SDS-PAGE. Finally, it does no harm to screen an "impure sample" as one can always perform further purification. Remember, crystallization used to be considered a very powerful purification tool (and still is!).

If the initial screen does not produce crystals, any promising results, or it becomes next to impossible to improve crystal quality during optimization, one should consider further purification of the sample.

Storing the Sample

Most proteins can be stored successfully at 4°C or -70°C.

Check with the person preparing the protein or compare your protein to a similar protein in the literature for best storage temperature.

Ideally, one should assay the activity and stability of the protein before storage and then later on at various points in time to determine the sample storage stability.

Repeated freezing and thawing of the sample should be avoided. Aliquot the sample into multiple small microcentrifuge tubes. Make the aliquots small enough so that the entire aliquot can be consumed in the experiment after thawing.

Sometimes people like to add glycerol (10 to 50% v/v) to help proteins better tolerate freezing. Avoid this if possible since it is often difficult to remove glycerol by dialysis or filtration. The presence of glycerol is a crystallization variable. Glycerol can behave as a precipitant, an additive, or cryoprotectant and therefore can influence the outcome of a crystallization experiment.

In general, it is better to store proteins more concentrated than diluted. When too dilute, adsorption of the protein onto the storage container can lead to significant losses. However, precipitation can sometimes be a problem when the protein is stored too concentrated.

Label samples clearly with the sample identification, batch identification, and date of storage. Small cryo labels can be very useful here. Color coding samples can be a nice organization tool. For the sake of easy organization and identification, it is sometimes more convenient to nest samples. For example, store batches of small microcentrifuge tubes in 10 ml or 50 ml centrifuge tubes and organize them by batch or sample.

Sample Handling

Be nice to your protein. Remember that proteins make an excellent food source for microbes. Protect your sample from microbes by storing the protein at less than 4° C and not leaving the sample for extended periods of time at room temperature.

When thawing a sample or mixing a lyophilized sample into solution do not shake or vortex the protein. Avoid foaming the sample. Foam can be a sign of denaturation.

Allow the sample to equilibrate to the temperature where the crystallization experiments will be set up and/or incubated before setting the experiment.

The field of crystal growth is full of opinions and controversy. There are several opinions on what should be done with the sample just prior to setting the crystallization experiment. Let's have a look at those opinions.

Some like to filter the sample through a 0.2 micron (or smaller, but be sure to compare the MW of your sample to the pore size of your filter so as not to stick your sample on the filter) pore size filter into a sterile container. Filtration can remove microbial contamination (but not the proteases) as well as sample aggregation. Turbid sample solutions with lots of precipitate should be solubilized or centrifuged before filtration to avoid the ugly experience of sticking the sample to a filter membrane. Use filters with the smallest possible dead volume to minimize sample loss. Some of the centrifugal microfilration devices are certainly worth consideration. Before filtering the sample, wash or flush the filter with a small amount of the sample buffer / storage solution before filtering the sample. This will test the filter for compatibility with your sample buffer and remove any trace glycerol which can sometime be present from the manufacturer. If possible, test filter a small aliquot of the sample, and measure the activity / OD before filtering the entire sample. Do this to test the adherence of the sample to the filter media. Read and follow the instructions supplied before introducing the sample to the filter.

Some like to centrifuge the sample. Centrifugation removes large sample aggregates and amorphous debris. Post centrifugation views can provide a visual clue of aggregation/precipitate for seemingly clear solutions. Following centrifugation, use only the supernatant for crystallization trials.

Others prefer to avoid filtration or centrifugation before setting crystallization experiments. One view is that the presence of amorphous material or aggregates can enhance the changes for crystallization be acting as nucleants.

To Azide or Not

Sodium azide (NaN3) is an anti-microbial preservative that is sometimes used to protect samples and crystallization reagents from microbial contamination. Sodium azide is toxic and should be handled with care. Typical sodium azide concentrations are 1 mM or if you prefer % measurements, between 0.02% and 0.1% w/v.

If you choose to use sodium azide remember that:

- It is toxic to humans as well as microbes.
- It is an inhibitor for some proteins and may become an unintentional ligand for your sample.
- It can interfere with heavy atom derivatization.
- Some metal azides are explosive.
- There are reports where eliminating sodium azide from the experiment improved crystallization.

Alternatives to Sodium Azide Include Thymol and Thimerosal.

A final alternative to the use of antimicrobials is the use of proper sterile technique and materials. Sterile filter all samples and reagents into sterile containers. Store samples and reagents at 4°C or lower. Use sterile pipet tips. Keep your work area clean. Develop a sterile technique with your crystallization setups. With common sense, sterile reagents and sample, good technique, and sterile pipet tips, one can successfully avoid the use of chemical antimicrobials in the crystallization lab.

Record Keeping and Organization

It is prudent to write down and hold onto detailed notes concerning the purification, storage, and handling of the sample. It is obvious that one should also maintain records of crystallization trials which should include:

- Sample information
- Name of sample
- Sample identification (batch, storage location, storage temperature, etc)
- Sample buffer composition, additives, ligands, etc.
- Sample concentration
- Crystallization experiment information
- Method
- Drop size and composition
- Reagent composition
- Temperature
- Date
- Person performing experiment

Although it may seem trivial, a little AR, and excessive, it is reasonable to write down anything that could become a crystallization variable.

Maintaining accurate and complete records of experimental observations is obviously important and will be covered in more detail elsewhere. In general, try to use both descriptive and quantitative comments when developing crystallization records.

Maintain a clean and well organized workplace for your crystallization setups. Cleanliness will help to prevent chemical, microbial, and miscellaneous contamination. Organization will prevent errors and save time.

Questions to Ponder About the Sample

- Does a similar sample exist and has it been crystallized?
- Does the sample contain free cysteines?
- Does the sample contain additives such as sodium azide, ligands, inhibitors, or substrates?
- Is the protein glycosylated?
- Is the protein phosphorylated?
- Is the protein N-terminal methylated?
- At what temperature is the protein stable?
- How does sample solubility and stability change with temperature?
- How does sample solubility and stability change with pH?
- Does the sample bind metals?
- Is the protein sensitive to proteolysis?
- What class of protein am I working with (antibody, virus, enzyme, membrane protein)?
- What have been the most successful approaches with my class of protein?
- What is the source of the sample?
- How was the sample purified and stored before it arrived into my hands?
- What is in the sample container besides the sample (buffer, additives, etc.)?
- Is the sample pooled purification aliquots or a single batch?
- How much sample do I have and how much more is available?
- How pure is the sample?

- How homogeneous is the sample?
- Does anyone possess any solubility information on this sample?
- What is unique about this protein?
- What is necessary chemically and physically to maintain a stable, active sample?

T is for Temperature

Temperature can be a significant variable in biological macromolecule and small molecule crystallization[1-5]. Temperature often influences nucleation and crystal growth by manipulating the solubility and supersaturation of the sample. Temperature has also been shown to be an important variable with phase separation in detergent solutions during membrane protein crystallization[7].

Control and manipulation of temperature during the screening, optimization, and production of crystals is a prerequisite for successful and reproducible crystal growth of proteins with temperature dependent solubility. Christopher et al.[5], testing 30 randomly chosen proteins, found 86% demonstrated a temperature dependent solubility and suggested that temperature induced crystallization could be a generally useful technique. Temperature was shown to affect quantity, size, and quality of the crystals, as well as sample solubility and preliminary crystallization data.

One advantage of temperature is that it provides precise, quick, and reversible control of relative supersaturation. Using temperature in addition to standard crystallization variables such as sample concentration, reagent composition and concentration, as well as pH, can increase the probability of producing crystals as well as uncover new crystallization conditions for a sample. Additional crystallization conditions may uncover reagent formulations more amicable to heavy atom derivatization, cryoprotection, and optimization or at least offer options. Temperature is amenable to control and can be used to carefully manipulate crystal nucleation and growth. This control can also be used to etch or partially dissolve, then grow back the crystal in an attempt to improve crystal size, morphology, and quality or assist with seeding. Temperature control is noninvasive and can manipulate sample solubility and crystallization without altering reagent formulation.

Typically, crystallization screens and experiments are performed at room temperature and possibly 4°C. A reasonable range of temperature to screen and optimize for protein crystallization is 4 to 45°C and some proteins have been crystallized at 60°C (glucagon and choriomammotropin). Temperature incubations above room temperature should be monitored closely for evaporation from the drop and reservoir. A 2 µl hanging drop vapor diffusion experiment at 37°C can evaporate in as little as 48 hours depending upon the plate and quality of seal. Microbatch under paraffin oil can minimize evaporation problems. In the case of room temperature incubations, temperature control and stability are often minimal since the experiments may be left in the open room. In an open room, temperature fluctuations may be significant, especially over a 24-hour period and on weekends when thermostatic control of the room environment can fluctuate 10° or more. Incubation at 4°C and other temperatures are often more stable since the incubation is performed in some type of incubator. Another source of temperature fluctuation occurs while viewing experiments. The light microscope is a heat source and extended viewing can significantly alter the temperature of small drops. Quick, efficient viewing can minimize temperature changes. Also, remember to turn off the light source when leaving plates on the stage in one position for more than a few seconds.

While controlled temperature can be important for consistent results, temperature fluctuation can be useful in obtaining high quality crystals by screening a larger range of crystallization conditions[8] since for a sample with temperature dependent solubility, changes in temperature can equate to changes in a crystallization reagent condition. Hence, a sparse matrix screen takes on a new dimension when screened at multiple temperatures, or ramped over several different temperatures over a period of time.

How does one test for the effect of temperature and temperature dependent solubility without consuming a lot of sample? One solution is to set a single crystallization screen at one temperature, allow the experiment to incubate for a week, record the results, and then move the plate to another temperature. Allow the experiment to incubate for a few days to a week at the new temperature and record the results. If one notices change in solubility (i.e. clear drop turning to precipitate, or precipitate turning to clear drops) between the two temperatures, then the sample has temperature dependent solubility and temperature should be explored as a crystallization variable.

Temperature gradients can be used for screening and optimization of proteins with temperature dependent solubility. For screening, set the experiment at one temperature, allow the experiment to equilibrate and then slowly change the temperature to a second temperature. In general, ramp the temperature so that the sample is exposed to an increase in relative supersaturation as the temperature changes over time. In other words, ramp from high to low temperature if the sample is more soluble at high than low temperatures. A temperature gradient or ramp allows one to slowly approach temperatures where a sample may have a decrease in solubility with a corresponding increase in relative supersaturation. Published examples of temperature gradient or temperature ramp crystallization include elastase[9] (25 to 20°C gradient), alpha-amylase[10] (25 to 12°C gradient), and insulin[11] (50 to 25°C gradient).

To demonstrate how screening temperature could affect and enhance the results obtained from a preliminary crystallization screen, the Hampton Research M6 Mini Temperature Incubator was used to screen 4 different temperatures. Using glucose isomerase and Crystal Screen, sitting drop vapor diffusion experiments were set using Cryschem™ plates at 4, 15, 25, and 37°C. Drops were observed daily and the results were quite interesting. Glucose isomerase crystallized in 19 conditions at 25°C, 23 conditions at 15°C, 28 conditions at 4°C, and 12 conditions at 37°C. A similar approach with Trypsin, yielded crystals in 8 conditions at 15°C, 4 conditions at 25°C, and 7 conditions at 32°C. In the case of Trypsin, a single set of Cryschem plates were set and the plates simply moved from one temperature to another over a period of a week, scoring results before each temperature change.

Temperature Tips

- In high salt, proteins with normal solubility will be more soluble at cold than at warm temperatures.
- In low salt, proteins with normal solubility will be more soluble at warm than at cold temperatures.
- Proteins with normal solubility will precipitate or crystallize from a lower concentration of PEG, MPD, or organic solvent more slowly at low than at high temperatures.

- Diffusion rates are less and equilibration occurs more slowly at low than at high temperatures. Crystallization may occur more slowly at low than at high temperatures.

- Temperature effects can be more pronounced at low ionic strength reagent conditions.

- Do not use the appearance or non-appearance of crystals at various temperatures to gauge the effectiveness of temperature as a crystallization variable. Rather, use the difference in the solubility at different temperatures to gauge the effect temperature has on sample solubility. If an effect is observed, explore temperature as a crystallization variable.

- Temperature can affect different crystal forms and growth mechanisms[12].

- When incubating experiments below and above room temperature and viewing experiments at room temperature, condensation can be a problem. To minimize and avoid condensation with vapor diffusion experiments, stack a dummy plate with reservoir filled with water and sealed, at the bottom and top of the stack of plates. This will slow the temperature change in the sandwiched plates and minimize condensation.

- The microbatch method works well for temperature exploration. In a traditional microbatch experiment, the relative supersaturation of the system does not change since, in theory there is no vapor diffusion. However, if the sample exhibits temperature dependent solubility, temperature can be used to manipulate sample solubility in a microbatch experiment. Another plus of using microbatch is the lack of condensation while viewing the experiment. Covers of microbatch plates can be removed for a clear view.

- Condensation with a hanging drop can mean alteration of your drop when the condensation mixes with the drop. Condensation with a sitting drop can mean there will be no mixing of the condensation with your drop, unless the condensation falls into the drop. Moral, sitting drop has less chance of mixing with condensation.

- To dry up condensation, add a small amount of concentrated salt solution to the reservoir. Keep in mind this might also dry your drop a bit.

- Ideally, one should set the experiment at the eventual incubation temperature and all reagents, samples, and plates should be equilibrated to the incubation temperature. This is a reality for room temperature and 4°C setups for those of us with cold rooms. For the rest of us, we can set the experiment at room temp and then toss it into the incubator. Or, for 4°C setups, one can cheat. Simply incubate the reagents, sample, plates, and slides in the refrigerator before setup. During the setup, place materials in a tray full of ice. Maintain the plates on ice during the setup. Seal and move smartly to the 4°C incubator.

- Nucleic acid temperature stability allows one to examine temperatures between 4 and 35°C.

- Increasing temperature increases the disorder of reagent molecules. Varying the temperature of a crystallization experiment can manipulate sample-sample as well as sample-reagent and reagent-reagent interactions. Such manipulations may have an impact on interactions that control nucleation and crystal growth. In addition, such interactions may have an impact on crystal packing as well as the termination of crystal growth. Hence, temperature can impact nucleation, growth, packing, and termination.

- Temperature can be a habit modifier and change the crystal lattice. For example, at temperatures below 25°C and in the presence of sodium chloride and acidic pH, the tetragonal form of lysozyme is favored. Under similar reagent conditions above 25°C, the orthorhombic form is favored[13].

- The preparation of heavy atom isomorphous derivatives can depend upon the temperature of the experiment. In most cases, it seems the soak temperature is the same as the crystallization temperature.

References

1. *Giege, R., and Mikol, V., Trends in Biotechnology (1989) 7, 277.*
2. *McPherson, A., European J. Biochemistry (1990) 189, 1.*
3. *A. Ducruix and R. Giege, Editors, Crystallization of Nucleic Acids and Proteins: A Practical Approach, IRL Press at Oxford University Press, 1991.*
4. *Lorber, B., and Giege, R., Journal of Crystal Growth (1992) 122, 168-175.*
5. *Christopher, G.K., Phipps, A.G., and Gray, R.J., Journal of Crystal Growth (1998) 191, 820-826.*
6. *Haser, R., et al., Journal of Crystal Growth (1992) 123, 109-120.*
7. *Garavito, R.M., and Picot, D., Journal of Crystal Growth (1991) 110, 89.*
8. *Drenth, J., Crystal Growth (1988) 90, 368.*
9. *Shotton, D.M., Hartley, B.S., Camerman, H., Hofmann, T., Nyborg, S.C., and Rao, L., Journal of Molecular Biology (1968) 32, 155-156.*
10. *McPherson, A., and Rich, A., Biochem. Biophys. Acta (1972) 285, 493-497.*
11. *T.L. Blundell, and L.N. Johnson, Protein Crystallography, Academic Press (New York) 1976, 59-82 (method by Guy Dodson).*
12. *A. McPherson, Crystallization of Biological Macromolecules, Cold Spring Harbor Laboratory Press, 1999.*
13. *Ataka, M., and Tanaka, S., Biopolymers (1986) 25, 337.*

U is for Units

Molarity is the number of moles of solute per liter and in represented as M, such as 3.0 M ammonium sulfate. Molarity is the ratio between the moles of dissolved solute (solid stuff) and the volume of solution (liquid stuff) in liters. The accepted volume of the solution is 1 L, so a 1M (molar) solution would be 1M = 1 mole of solute/1 L solution. Molarity is a way of determining the concentration of a solution. Dilute solutions are typically expressed in terms of millimolarity (mM) where 1 mM = 0.001 M. Typically, in crystallization we are asked to make something like a 3.5 M solution of ammonium sulfate. To do this we need to know the molecular weight of ammonium sulfate (132.14 g/mole), the volume of solution to make (let's make 500 ml or 0.5 L), and the desired concentration (3.5 M doh!). Then we calculate:

$$\#\ \text{grams required} = (3.5\ \text{mole/liter})(0.5\ \text{liter})(132.14\ \text{g/mole}) = 231.25\ \text{g}$$

So, we then weigh 231.25 g of ammonium sulfate and add deionized water to dissolve the ammonium sulfate and then adjust the final volume to 0.5 liter (500 ml). Do NOT add 500 ml of water to 231.25 g of ammonium sulfate. Ideally one should use the most precise measuring instrument possible such as a volumetric flask. The less desirable instrument would be a graduated cylinder and the least desirable would be a beaker.

Molarity is typically used as a concentration unit for salts 1,6-hexanediol, detergents, and some additives.

% w/v (percent weight/volume) is often used when formulating high molecular weight polyethylene glycols which begin life as solids as well as some additives. % w/v is the weight of a solute in a given volume. % w/v = gram per 100 ml. For example, 50% w/v PEG 3350 is 50 g of PEG 3350 in a final volume of 100 ml and

NOT 50 g of PEG 3350 plus 50 g of water. This means weigh 50 g of PEG into a graduated cylinder or volumetric flask and bring the final volume to 100 ml with water. To clarify, in order to make a 500 ml of a 30% w/v solution of PEG 8000, simply weigh 150 g of PEG 8000 and bring it to a final volume of 500 ml in deionized water. Often the mistake of adding 150 g of PEG 800 to 350 ml of water is thought of as a 30% w/v solution. This is a huge no-no and makes one look like a lab rookie. Get it right and look cool.

%v/v (percent volume/volume) is often used when formulating liquid, low molecular weight polymers (PEG 400), organics (MPD), and organic solvents (iso-propanol) into stock solutions. % v/v is the volume of a solute in a given volume and NOT a volume plus a volume. For example, a 50% v/v iso-propanol is 50 ml of iso-propanol brought to a final volume of 100 ml with water and is NOT 50 ml of iso-propanol plus 50 ml of water. Once again, the latter is a big no-no and makes you look like a Homer.

Additional comment regarding % w/v and % v/v. Do not ask someone "What is your definition of % w/v?". There is only one correct definition. Therefore all other definitions are wrong. Keep life simple and use only one definition of % w/v and % v/v. The correct definition. There is only one correct way to make these solutions. There is one occasion where one can ask the question, "What is your definition of % w/v?" and this is when someone new enters the lab and you want to check them out and see if they know anything about growing crystals or you want to steal your buddy's crystallization reagents but you want to be sure they make them properly before you risk getting into trouble for stealing junk solutions.

% saturation is the concentration of material in solution as a percent of the maximum concentration possible at the given temperature. A saturated solution is one where there is equilibrium between undissolved solute and dissolved solute. To make a saturated solution, a salt is added to water and often warmed to enhance solubilization. Complete dissolution is desired. Upon cooling, some of the solute (salt) will crystallize out and leave behind a saturated solution. The actual concentration of a saturated stock depends upon the temperature of the solution. For example, at 0°C, 127.5 g of potassium iodide can be dissolved into 100 ml of water, but at 20°C, 144 g of potassium iodide can be dissolved into 100 ml of water. Therefore, depending upon whether the solution is kept at room temperature or in the cold, the concentration can be very different. % saturation is an old school way to make salt solutions for crystallization. However, since we often perform crystallization at different temperatures, the actual concentration in the bottle, reservoir, or drop can be very different. Plus, exact reproduction of a stock solution not only depends upon careful mass and volume measurement, but also temperature. Keep life simple, avoid reproducibility hassles and stick with M, % w/v and % v/v.

Mg/ml (mg/ml) is typically used to express or determine protein concentration. Oftentimes we like to whip up a batch of lysozyme to test a new or stupid crystallization idea or simply prove to ourselves that we can crystallize a protein (especially those of us working with kinases, membrane proteins, and other chewing gum macromolecules). To make a 20 mg/ml lysozyme solution we might weigh 20 mg of lysozyme and simply add 1 ml of buffer (20 mg plus 1 ml). However, others might weigh 20 mg of lysozyme and add 980 µl (0.98 ml) (20 mg in 1 ml). Same rough definition, but different concentrations of lysozyme. Might not make a difference unless you are trying to reproduce someone else's work and you want to be sure you are comparing apples to apples and not apples to oranges.

A final note on units. Knowing how to properly calculate, formulate, and document crystallization reagents is an excellent foundation for becoming a crystal Jedi master. Just the same, knowing how to properly calibrate and use an analytical and top loading balance and select the appropriate glassware for the job are often overlooked, yet important crystallization variables. Quality and consistency are big time crystallization variables so it is important to learn and master good lab technique. While nobody ever won a Nobel Prize for cooking, most Nobel Laureates can cook up a storm.

V is for Vibration

Vibration is often associated (in a crystallographers mind) with excessive nucleation. It is popular belief that the best crystals grow when there is NOT a significant source of vibration. It is believed this is particularly important during the optimization and production stages of crystallization. Hence, it is suggested that during optimization and production, one set crystallization experiments and avoid the burning curiosity to look at the plates every single day. Try leaving the setups alone during optimization and production.

Since vibration can lead to excessive nucleation, it seems one might want to move the plates some during screening. One typically does move plates during screening as it is typically recommended to view experiments each day for the first week and once a week thereafter until the drop dries out. Maybe we need to build vibrators into our screening incubators. Along the same line, it is sometimes observed that plates which have been sitting for some time, then viewed, often have crystals forming in the days following the latent observation. So perhaps one should consider giving the plates a Monty Python nudge-nudge or move them about to stimulate nucleation.

To avoid vibration, place plates in cupboards, cabinets, or incubators that are infrequently opened and closed. Installing hydraulic door dampers to prevent slamming doors or cupboards is one way to reduce vibration. Some incubators vibrate more than other units and peltier based heating and cooling incubators may be desirable from this perspective. However, some of the circulation fans in these units vibrate quite a bit, so be sure to check out the fan and even consider replacing it with a higher quality unit with less oscillation and vibration.

W is for Water

The role of water in protein crystallization is significant as protein crystals are highly hydrated. The amount of water in a drop or a crystal, the purity, the chemical composition, and how the water interacts with itself, the sample, and reagents are all critical variables in a protein crystallization experiment. Consistency is typically a good thing to keep in mind when trying to reproduce crystallization results. Being consistent with one's treatment, handling and use of water can be a important crystallization. Consider the source of water. One should use pure and fresh water for crystallization experiments. Type II analytical grade water with a resistivity of less than 18.2 MΩ-cm, with a total organic carbon (TOC) of less than 30 ppb and low bacteria count is a good source of water for protein crystallization. Purchasing and using a quality water purification system is one part of the water source equation. Different water systems and different methods of water purification produce different water. For example, glass distilled water typically has a pH of 5.5 while reverse osmosis and deionization systems typically produce water with a pH of 7.

Another part of the water equation is the proper use and maintenance of the water purification system. Carefully follow the manufacturer's use, maintenance, and service guidelines. Establish regular quality control practices to ensure the consistent quality of the water. Using a water purification cartridge beyond its expiration date or failing to clean the system on a regular basis can lead to water with higher resistivity as well as microbial contamination. Microbial contamination can change the pH of the water as well as expose your sample to proteloytic modification. Use fresh water. Purified water is deionized and can leach zinc, lead, copper, iron, aluminum and other substance from glass or plastic storage containers. Purified water stored in plastic containers for several weeks has nearly the same level of TOC as tap water.

Use fresh, pure water for your crystallization experiments. When having trouble reproducing experiments, consider water, the difference in purity and the difference in source as potential crystallization variables.

X is for Xenon

Xenon is a noble gas that binds to specific sites in a macromolecule. Xenon-protein complexes can often serve as heavy atom complexes for MIR structure determination. Heavy atom derivatives of protein crystals can be produced by pressurizing native crystals with xenon gas. In the vast majority of cases, the modification of the mother liquor to determine soaking conditions is avoided since the crystal and mother liquor are simply placed in a chamber pressurized with xenon gas. Another advantage of working with xenon is that it interacts weakly with a protein so isomorphism of the derivative with the native crystal is high. Also, the number of binding sites as well as the binding occupancies can often be changed by altering the pressure of the xenon gas. Xenon binding sites often differ from heavy metal binding sites so it can be useful to try xenon when traditional heavy atom soaks fail. Xenon binding is often reversible so if one has very few crystals, the same crystal could be used for heavy atom soak.

The interaction between xenon and the protein is typically confined to weak dispersion forces and does not involve electrostatic interactions. Therefore, xenon should bind to different locations of the protein compared to most other heavy atom compounds. The weakness of the interaction between xenon and the protein makes it unlikely that xenon will interfere with crystal contacts. Xenon derivatives usually show good isomorphicity with the native crystal, resulting in high phasing power.

Xenon seems to have little or no effect on the pH or ionic strength of the mother liquor. The weakness of the interactions between xenon and proteins requires a significant xenon concentration in the crystal's mother liquor to enforce sufficient occupation of a potential binding site so most uses of xenon reported in the literature make use of high pressure equipment for xenon derivatization. At this time is appears cryocooling does not alter the protein's xenon binding properties.

Xenon derivatization can be performed at room temperature and data collection with xenon derivatives can be performed at room temperature using specially made pressure cells and capillaries or at cryogenic temperatures using CryoLoops.

Another application of xenon in protein crystals is the crystallographic imaging of disordered areas of lipids or detergents in crystals of membrane proteins.

References for using xenon as a derivative:

1. *Binding of Xenon to Sperm Whale Myoglobin. Schoenborn B.P.; Watson, H.C.; Kendrew, J.C. (1965). Nature, 207, 28-30.*
2. *Cavities in proteins: structure of a metmyoglobin-xenon complex solved to 1.9&Angring. Tilton, R.F.; Kuntz, L.D.; Pesko, G.A. (1984) Biochemistry 23. 2849-2857.*
3. *Using Xenon as a Heavy Atom for Determining Phases in Sperm Whale Metmyoglobin. Vitali, J.; Robbins, A.H.; Almo, S.C.; Tilton, R.F. (1991). Journal of Applied Crystallography, 24, 931-935.*
4. *On the Preparation and X-ray Data Collection of Isomorphous Xenon Derivatives. Schiltz, M.; Prange, T.; Fourme, R. (1994). Journal of Applied Crystallography, 27, 950-960.*
5. *Successful flash-cooling of xenon-derivatised myoglobin crystals. Soltis, S.M.; Stowell, M.H.B.; Wiener, M.C.; Philips, G.N.; Rees, D.C. (1997).Journal of Applied Crystallography, 30, 190-194.*
6. *Freeze-Trapping Isomorphous Xenon Derivatives of Protein Crystals. Sauer, O.; Schmidt, A.; Kratky, C. (1997). Journal of Applied Crystallography, 30, 476-486.*
7. *Protein Crystallography at Ultra-Short Wavelengths: Feasibility Study of Anomalous Dispersion Experiments at the Xenon K-Edge. Schiltz, M., Kvick, A., Svensson, O., Shepard, W., De LaFortelle, E., Prange, T., Kahn, R. & Fourme, R. (1997). Journal of Synchrotron Radiation , 4, 287-297.*
8. *A method to stabilize reduced and/or gas treated protein crystals by flash cooling under a controlled atmospher*
9. *Xavier Vermede et.al. J. App. Cryst, (1999) 32(3) 505-509*
10. *Better structures from better data through better methods: a review of developments in de novo macromolecular phasing techniques and associated instrumentation at LURE. Fourme, R. et al (1999) J. Synch. Rad. 6(4) 834-844*
11. *Timmins, P., Pebay-Peroula, E. & Welte, W. (1994). Biophys. Chem. 53, 27-36.*

Y is for Yummy

Protein samples are yummy treats for microbes such as bacteria, yeast and fungus which secrete proteases that like to chop your sample into tasty bites, ruining your crystallization setup. These same microbial menaces also like to dine on polymer based crystallization reagents such a polyethylene glycols and even buffers. So even if the bugs and their secreted proteases do not chop your sample into bits, they can degrade reagents, alter the pH of the solution, or generate chemical species which can influence crystallization and even make reproducing conditions a pain in the tush. To prevent or at least minimize your sample and reagents from becoming Sunday Crystallization Brunch for microbes, consider the following suggestions:

Sterile filter the sample into a sterile tube using a 0.2 micron filter before setup to remove microbes. Do not leave the sample on the bench at room temperature for extended periods. Store unused sample appropriately (4, -20, or -80°C; only you know best). Those of you working with engineers, physicists, computer scientists, or folks from outside the life science field, take a moment to realize that quarks, gluons, hard drives and other physical things these folks are used to handling do not support microbial growth. Take a little time to help these crystallization rookies understand what happens to samples left on the bench for a day or three and the significance of keeping things clean to prevent microbial (as well as chemical) contamination. Case in point – Caller: "My lysozyme stock that was growing crystals no longer grows crystals." Tech Support: "Where are you storing the lysozyme when it is not being used?" Caller: "Same place as the reagents." Tech: "Which is?" Caller: "On the bench (i.e. room temperature)." Tech Support: "For how long?" Caller: "Oh, not long, maybe a few days, a couple weeks at the most". Hey, nothing wrong with these types of questions. We all gotta learn some time. Just be sure and help that engineer, computer scientist and new post doc from a physics lab who is helping you with your new AFM machine or testing the high throughput robot, to understand proteins are food for microbes. Who knows, maybe they will help you with the next Windows update installation.

Sterile filter water into a sterile container before formulating crystallization reagents that cannot or should not be filtered (detergents, gels, and some high molecular weight polymeric agents too big for filtering).

Sterile filter (0.2 micron pore size) buffers, salts, polymers, and diluted organics into sterile containers. Sterile containers such as polypropylene or PETG plastic are readily available today and are cost-effective, time-saving alternatives to glass and autoclaving.

When using a stock that has been sitting about for more than a couple of weeks, give the bottle a swirl and look for signs of microbial growth such as a settle, faint white, off white or yellow to brown precipitate. Sometimes slightly precipitated salts will resemble microbial growth to the untrained eye. To help differentiate precipitate from microbial growth, try warming the solution in your hands for 5 to 10 minutes. If the material disappears it might well be precipitate. Microbial growth will not dissolve. Although we are not endorsing the sniffing of reagents, especially since some crystallization reagents can be hazardous, precipitates will not smell, while microbial growth will often stink.

Use sterile pipet tips when pipetting reagents, sample, and water into plates. Keep your filthy paws away from your pipet tips and do not touch pipet tips to the counter, your lab partner or other non-sterile items. If you set your pipet down it had better not have a tip on it or the tip will likely touch something non-sterile.

Keep crystallization plates in their sealed wrappers until just before use. This will prevent airborne microbes from setting up home in your crystallization plates. Most crystallization plates are not offered with a claim of sterility but most makers of these plates operate the injection molds in clean areas and package the plates straight out of the mold so there is little risk of contamination. Use them as is and put the worry into keeping them clean once the plate is out of the package.

If you do crystallization in a lab that also grows microbe bugs, especially fungal microbes, be especially careful as these airborne spore forming nasties can set up camp in crystallization plates quite easily.

Z is for Zeppenzauer

Microvolume crystallization by dialysis was first reported by Zeppenzauer in 1971 (Methods in Enzymology, ibid ref. 1. Vol. 22, page 253). The Zeppenzauer cells are homemade from capillaries with the ends of the tubes either covered with dialysis membrane or plugged with acrylamide, silica, or agarose gel. If covering the tube with a dialysis membrane, one can use small o-rings to secure the membrane over the tube. The capillary is then placed in an appropriate chamber with a volume of reagent sufficient to cover the ends of the tube.

Z is for Zinc

1 to 5 mM zinc (most frequently as zinc chloride or zinc sulfate) can be a useful crystallization additive and typically reduces the solubility of the sample. Zinc is not very soluble and can readily precipitate or crystallize out of reagents as higher concentrations.

Z is for Zero

Zero or “0” is often used to represent a clear drop when scoring crystallization experiments. While on the subject of zero or nothing, consider using -1 or some other more interesting and exciting score to represent a non-experiment. A non-experiment is a drop that has fallen off or a drop that was never pipetted (human or robot error). Better to use a non-experiment score such as -1 than a “0” since someone reading your scores might mistake a “0” as a clear drop.

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

Additives

Glycerol

Looking for yet another additive to try during optimization? Try Glycerol. Typical concentrations in the drop can range from 3 to 15%. Glycerol may limit non-specific aggregation and could be useful if you plan to use cryo techniques during data collection.

What Additive to Use

When trying to decide which additive might be useful during crystallization, try the following. If one has crystals and wants to try using additives to improve the crystal size or quality, go back to the initial crystallization screening plates. Review the plates, looking for conditions where neither precipitate nor crystals were/are observed. Now, review the results, looking for a common reagent ingredient present where no precipitate or crystals are found. For example, one might find that all drops with Isopropanol remained clear. One could then try adding Isopropanol to the current crystallization condition to see if it could improve the crystal size and/or quality. If one observes a difference in the crystal in the presence of isopropanol, then one might consider evaluating other additives in the class of alcohols such as Ethanol, Methanol, tert-Butanol and others. If one has no crystals, but plenty of precipitate, phase separation and clear drops, follow the above analysis and try adding the common reagent ingredient found in the clear drops to those drops which contained precipitate or phase separation. It is possible this agent could improve or at least manipulate sample solubility. The above tip was submitted from Jarmila Jancarik from the laboratory of Sung Ho Kim at the University of California, Berkeley. Thank you Jaru! When using this approach it might be reasonable to discern between concentration independent and dependent precipitation when trying to decide which agent to pursue as an additive. Try evaluating the concentration independent agents first and then look at the other agents if sample quantity permits. For example, if one observes precipitate in 15 to 40% PEG but not in 5 to 10% PEG, it might simply be a concentration. However, if one observes no precipitate in 45 to 60% v/v MPD, one could guess that MPD is a reasonable agent to evaluate as an additive.

Crystal Screen as an Additive Screen

Although Hampton Research does offer a specifically formulated Additive Screen™ (HR2-428), here is a tip submitted by the crystallography group at GSK (North Carolina) when one already has a sparse matrix screen or two laying about the lab. When screening additives, try adding 50 µl of each Crystal Screen™ reagent to 950 µl of the "best" crystallization conditions thus far, in order to see if any of the reagents in Crystal Screen might serve as good additives. Crystal Screen 2™ is an especially good kit for this technique since it contains numerous divalent cations such as Jeffamine® Reagent and a few other "novel" agents. Jeffamine® is a registered trademark of the Huntsman Petrochemical Corporation.

Benzalkonium Chloride

Try Benzalkonium chloride (Fluka 12060). This cationic surface active agent has been reported to be useful as a crystallization additive with membrane proteins and may be useful for soluble proteins. We've been using it in the drop between 1 and 3% w/v in water. Try a 10% w/v stock solution in water and dilute into the drop to 1-3%.

Sodium Thiosulfate to Prevent Intermolecular Disulfide Bridges

The presence of Thiosulfate in the protein solution was essential to promote crystal growth and to avoid the formation of unstable and weakly diffracting crystals.[1] This is likely to be a consequence of the intrinsic capability of the reduced thiol group of the active-site cysteine to form disulfide bridges, leading to the destabilization of the protein native structure. Sulfane sulfur-donor compounds such as $Na_2S_2O_3$ are likely to either keep the protein in the persulfurated form or to prevent intermolecular disulfide bridges leading to unfolding and aggregation.[2]

References

1. *Crystallization and preliminary crystallographic characterization of LmACR2, an arsenate/antimonate reductase from Leishmania major. D. Bisacchi, Y. Zhou, B. P. Rosen, R. Mukhopadhyay and D. Bordo. Acta Cryst. (2006). F62, 976-979.*
2. *Bordo, D., Forlani, F., Spallarossa, A., Colnaghi, R., Carpen, A., Bolognesi, M. & Pagani, S. (2001). Biol. Chem. 382, 1245–1252.*

Dissolving Hydrophobic Additives into Oil

Try dissolving the small molecule additive into paraffin or silicon oil, and use this mixture to cover the sample drop. This can be used with sitting drop, vapor diffusion, or with microbatch under oil methods. The oil acts as a reservoir that may contain excess small molecules that (you hope) will be fed into the crystals.

Aggregation

Sample Buffer

Non-specific aggregation problems? Watch your sample buffer. The selection of an appropriate sample buffer/pH can often ward off aggregation problems.

Cryo

Cryo Buffer

We've had good success using the well solution directly as the foundation of a cryobuffer in several situations where crystals cannot be grown directly in the presence of cryoprotectant, and where crystals don't tolerate transfer to artificial mother liquors. The basic protocol is as follows: (1) Remove 100 µl of the well solution after crystals have grown; (2) Split this sample into two 50 µl aliquots; (3) Add 7.5 mg of Dextrose (Glucose) to the first aliquot and 15 mg to the second aliquot. Dissolve by gentle pipeting with a wide-bore tip. This will give two sequential well solutions that now contain 15% and 30% w/v Dextrose. If all the Dextrose won't go into the second aliquot, spin hard and remove the supernatant; (4) Transfer the crystal to aliquot number 1, equilibrate for 3 minutes, then to aliquot number 2, then freeze. We've had a few crystals that routinely crack or blow up when transfered to artificial mother liquor that behave well when transfered to well solution plus Glucose. We assume that there is some aspect of the crystal drop (pH, ionic tension, precipitant concentration) that is more effectively reproduced within the well than by separately prepared mother liquors. The nice thing about the protocol above is that you don't get much of a volume increase when dry Dextrose is dissolved in the well solution, so the components in the solution are not diluted. Finally, if you don't get a really good freeze, you can try to add about 5% v/v Glycerol to aliquot number 2, in addition to the 30% w/v Dextrose.

PEGs for Cryo

High molecular PEGs are also good cryoprotectants. If crystals are obtained from relatively high concentration of PEGs (e.g., 30% of PEG 3350), you can cryoprotect them simply by raising the concentration of the PEG a little bit (e.g., 40% of PEG 3350).

J. Appl. Cryst. (2006). 39, 244-251
Effects of cryoprotectant concentration and cooling rate on vitrification of aqueous solutions. V. Berejnov, N. S. Husseini, O. A. Alsaied and R. E. Thorne
Synopsis: Critical concentrations required for vitrification of aqueous solutions are determined for fourteen common cryoprotectants, for sample volumes ranging over four orders of magnitude and covering the range of interest in protein crystallography.

Mounting Thin Crystals

To mount very thin crystals onto CryoLoops™, first dip the nylon loop into 0.5% Formvar® solution (Fluka # 09819) to form a thin film. The film provides extra support for fragile crystals, and can result in much sharper reflections with just slightly higher background. To clean the loop, dip it in alcohol to dissolve the support. Two notes: (1) the technique works only for crystals grown without organic solvents; and (2) take precautions not to breathe vapor from the Formvar solution--the solvent is 1,2-dichloroethane. It is a standard support for electron microscopy grids.

The Potential Benefits of Cryogenic Data Collection

Reduced radiation damage.

Decreased thermal motion and disorder.

Potential for improved resolution.

Increased crystal lifetime.

Crystals can be stored and shipped.

Cryo continued...

Cannot Get Cryo to Work?

Try x-ray data collection at room temperature.

Evaluate other cryoprotectants. Try CryoPro™ from Hampton Research (HR2-132).

Try the idential procedure again with another crystal.

Vary the time and temperature of the crystal handling steps.

Try annealing.

Match the osmotic pressure of your cryoprotectant to the osmotic pressure of the reagent producing the crystal. Crystallization reagents with lower salt concentrations require a higher percentage of cryo reagent for cryo protection than crystallization reagents with higher salt solutions (Garman 1999). Osmolality tables (Weast 1988-1989) can be used to estimate the osmolality of reagents.

References
1. *Cool data: quantity and quality. Elspeth Garman. Acta Cryst. (1999). D55, 1641-1653.*
2. *Weast, R. C. (1988-1989). Editor. Handbook of Chemistry & Physics, 69th ed. Boca Raton, Florida: CRC Press.*

Cryo and Detergents

Be aware of the potential for detergent concentration mismatch between your mother liquor and the cryosolution. This particularly happens with vapor diffusion setups: there is a delicate balance of "free" detergent in the mother liquor versus the proportion of the detergent which is bound to the protein. Dropping a xtal into the cryosolution shocks the crystal with a bolus of extra free detergent. Hence, and counterintuitively, you may need to reduce the detergent concentration in the cryosolution to keep everything in balance. Try titrating down from 1% to even as low as 0.4% in the cryosolution. Under the conditions you are using, the CMC of bOG is suppressed below the usual 0.67% (w/v).

Also, the behavior of many of the alkyl glycoside detergents is very temperature sensitive. So be careful about the temperature of all the solutions you use.

R. Michael Garavito, Ph.D.
Submitted to CCP4 bulletin board February 2007
Edited by Hampton Research Corp.

Cleaning a CryoLoop™

Soak the CryoLoop in a 0.5-2.0% detergent solution for 15 minutes or more. To improve the action of the cleaning solution, try increasing the temperature of the solution. The use of ultrasonic or other agitation will further improve the cleaning action. Finally, lengthen the dwell time or increase the concentration of the solution.

Paper Wicks™ can be used for delicate scrubbing of the CryoLoop. Dip the end of the wick into a detergent solution or 2-Propanol, and gently clean the nylon CryoLoop.

Compressed air can be used to remove debris from a CryoLoop. Hold the tip of the air sprayer 3 inches from the CryoLoop and use gentle pressure (less than 25 psi).

Paper Wicks - 55 mm X Fine Long - 100 wicks (HR4-211)

Paper Wicks - 55 mm Medium Long - 100 wicks (HR4-213)

Duster - Canned Air - 10 oz can (HR4-411)

Micro-90® Concentrated Alkaline Cleaning Solution (International Products Corporation, www.ipcol.com, 609-386-8770)

Detergents

Detergent to Protein Ratio and Pea Type as Crystallization Variables

Ben-Shem et al (Acta Cryst. (2003). D59, 1824-1827) in the crystallization of higher plant photosystem I found the detergent to chlorophyll ratio had to be carefully optimized in all purification steps in order to produce ordered crystals. Crystals were produced in 22.5 mM MES-bis-tris pH 6.6, 0.5% v/v PEG 400, 8.1 mM Ammonium citrate, 6% w/v PEG 6,000. Crystals appeared in 2 to 3 days and matured in size within two weeks time, yet the loss of sharp edges and a degradation of diffraction quality was observed after an additional two weeks time. The initial diffraction resolution of 20 Angstrom was improved to 6 Angstrom through a seemingly tedious refinement of isolation, crystallization and cryo conditions. Again, the detergent to chlorophyll ratio as well as the pea type and growing conditions with adjustment of the preparation to seasonal changes were essential to improvement of crystal quality.

Amphophiles

When screening detergents as additives, be sure to evaluate small amphophiles such as 1,2,3-Heptanetriol, Benzamidine, Ethanol, Dioxane, 1,6-Hexanediol, Ethylene glycol, and Butyl ether for their ability to "manipulate" micelles.

Detergents, Oils and Microbatch

Detergents can be used successfuly as crystallization reagents in oil based microbatch experiments. Measurements by Barenda et al and Loll et al indicated no significant loss of detergent will occur by migration of the detergent into the oil.

References

1. *Thomas R.M. Barenda and Bauke W. Diskstra. Oils used in microbatch crystallization do not remove a detergent from the drops they cover. Acta Cryst. (2003). D59, 2345-2347.*
2. *Loll et al. Compatibility of detergents with the microbatch-under-oil crystallization method. Acta Cryst. (2003). D59, 114-116.*

Handling Crystals

Separating Twins

The following procedure for separating twins is reported in "Structural Basis for Double-Stranded RNA Processing by Dicer", Ian J. MacRae, Kaihong Zhou, Fei Li, Adrian Repic, Angela N. Brooks, W. Zacheus Cande, Paul D. Adams, Jennifer A. Doudna. Science 13 January 2006: Vol. 311. no. 5758, pp. 195 - 198.

Crystals of the protein Dicer grown at 18°C by vapor diffusion in hanging drops composed of equal volumes of protein solution (9 mg/ml) and reservoir solution (28% PEG-400, 0.1 M $MgCl_2$, 0.1 M NaCl, 5 mM TCEP, 1 mM DTT and 0.1 M MES, pH 6.5). Resulting crystals were pseudo-merhodrally twinned, producing an apparent spacegroup of P4222. Substituting $MgCl_2$ with $MnCl_2$ resulted in crystals that were macroscopically twinned. Individual crystals, which belong to the spacegroup P21212, were extracted by soaking the largest twinned crystals in reservoir solution containing 15% PEG-400 to weaken the twinning lattice contacts, followed by gentle mechanical prodding with a cat whisker. The resulting crystal shards were transferred back into full strength reservoir solution, cooled to 4°C for 1 hour and then cryo-cooled by plunging into liquid N2.

Handling Crystals continued...

14 Things to Say About Twins

1. Try additives.
2. Try DNA shuffling to introduce random mutations.
3. Try agarose gel crystallisation.
4. Try a new construct ("having tried everything else for 3 years before...").
5. Try to work on very small crystals and/or using a very small beam of a microfocus beamline at a synchrotron to isolate a single domain and get less twinning.
6. Destabilise the crystal to separate the two halves (when possible) as in http://www.sciencemag.org/cgi/data/311/5758/195/DC1/1
7. If the crystals are obtained with Mg exchange it to Mn.
8. Change in crystallisation pH.
9. Change in crystallisation temperature.
10. You can try switching proteins. Lysozyme usually does the trick.
11. Co-crystallisation with partner proteins/domains or with ligands.
12. "Hmmm. Who knows, eh?"
13. Change crystallisation setup (e.g., from microbatch to hanging drops).
14. Test crystals at room temperature as well: "(EBV protease) which became twins only upon freezing, at room temperature the crystals were untwinned (but resolution was much worse as well)".

Source: CCP4 Newsgroup, Stefano Benini Ph.D.

Crystal Annealing

Macromolecular crystal annealing (MCA) can help overcome increase mosaicity associated with cryocrystallography (Harp et al 1998, 1999). One might consider annealing if diffraction is uncharacteristically poor after flash cooling. The process cycles a flash cooled crystal to ambient temperature and then to cryogenic temperature and requires no special equipment. The annealing process does not improve a poorly diffracting crystal suffering from molecular disorder. The annealing protocol assumes that adequate cryo protection is available or that the crystal may be flash cooled using an oil and that the crystal diffracts well. The crystal is flash cooled a cryo stream. In the MCA procedure, the crystal is removed from the cryo stream and placed in a large (300 microliter) volume of the optimal reagent/cryo/oil solution from which the crystal was originally grown/mounted. Cover this drop to prevent evaporation. Incubation time at room temperature is typically 3 minutes. Extended incubation times are okay, but shortening the incubation time can produce inconsistent results. MCA has been successfully reported for both small and large crystals as well as for crystals with low (30%) or high solvent (70%) content.

Other forms of crystal annealing, one termed flash annealing (Yeh and Hol 1998), the other termed annealing on the loop (Harp 1999), have been used successfully, especially on crystals with low (30%) solvent content. For flash annealing the crystals remain in the loop and on the mount. The cryo stream is diverted for 1.5-2.0 seconds, then reflash cooled for 6 seconds before repeating the process for a total of 3 rounds of rewarming and flash cooling. For annealing on the loop, a variable length of time for warming is used without multiple rounds of warming and reflash cooling. Warming time for cooling on the loop is determined by observing the crystal in the loop while the stream is diverted and waiting until the drop in the loop is clear before reflash cooling the crystal. Warming time is typically proportional to the size of the crystal. For the flash annealing and annealing on the loop methods one might carefully reduce the amount of liquid in the loop using a paper wick or micro wick as a variable to improve annealing results. But be careful not to let the drop dry too much. Finally, one might be prepared for the formation of ice on the crystal and loop when the stream is blocked. Ice can be removed with the delicate use of a paper wick or fiber.

In closing, one may attempt the quick annealing methods first, although the MCA seems to be more general.

References

1. *Macromolecular crystal annealing: evaluation of techniques and variables. Harp et al. Acta Cryst. (1999). D55, 1329-1334.*
2. *Macromolecular crystal annealing: overcoming increased mosaicity associated with cryocrystallography. Harp et al. Acta Cryst. (1998). D54, 622-628.*
3. *A flash-annealing technique to improve diffraction limits and lower mosaicity in crystals of glycerol kinase. Yeh and Hol. Acta Cryst. (1998). D54, 479-480.*

Handling Crystals continued...

Minimizing Radiation Damage

Small molecules which behave as electron and free radical scavengers can be used for co crystallization and soaking to reduce radiation damage to the macromolecular crystal. Recommended scavengers include oxidized glutathione, nicotinic acid, 5,5'-Dithiobis(2-nitrobenzoic acid) (DTNB), and ascorbic acid. Typical scavenger concentration is 0.2 M in the original crystallization reagent. A quick (less than 20 second) soak can be used to introduce the scavenger to the crystal. Prepare fresh solutions (add solid material to crystallization reagent in plate well) just before the soak.

References

1. *How to avoid premature decay of your macromolecular crystal: a quick soak for long life. Brice Kauffmann, Manfreid Weiss, Victor Lamzin, and Andrea Schmidt. Structure 14, 1099-1105 July 2006.*
2. *Investigation of possible free radical scavengers and metrics for radiation damage in protein cryo-crystallography. Murray and Garman. J. Synchrotron Radiation 9, 347-354, 2002.*
3. *Blundell, T.L. and John, L.N. (1976) Protein Crystallography (New York, Academic Press).*

Microbatch Crystal Mounting Tips

Here is the compiled 'tricks of the trade' for microbatch crystal mounting from the CCP4B community. Rebecca Page, Brown University, September 2006.

Original Post to CCP4 community:
I find mounting crystals grown in microbatch drops difficult compared to mounting crystals grown out of sitting drops, primarily because of the oil layer. Do most people simply use the oil layer as a cryoprotectant? If not, do you typically try to remove the oil prior to transferring the crystal to a cryoprotectant? I'd like to compile 'tricks of the trade' for mounting crystals out of microbatch plates. Any advice would be helpful and I'll compile the list of responses and repost for the community.

Summary of responses:
Sorry if this sounds like a crazy suggestion, but sometimes the simplest things work. Did you try to mount the crystal directly on a loop and see if it diffracts? The oil can be cryoprotectant.
The setup would be:

1. Fish the crystal with a loop
2. Do not care if it comes across the oil layer and it retains the oil
3. Mount in the cryostream
4. Shoot x-rays and see if you have diffraction

If everything fails and you have few more drops of crystals and do not know how to freeze mount, here is another way you can try :- If you have few crystals in the drop under oil and the drop size (excluding oil!!) is few µl: add 10 µl of mother liquor (you can get the mother liquor conc. based on few trials; it should be few % more than the final conc. of the one you had used while setting up the microbatch, a decent start will be 5 % more), allow it to stand for few minutes. Then use the classical cappillary mounting method to suck the crystals slowly out and onto a cover slip. If possible try to remove as much as possible (you may not be able to remove everything) the halo of oil surrounding the drop on cover slip (using the same capillary watching under microscope to make sure you are not sucking out the crystals). The use of mother liquor (the 10 ul) is to basically to reduce the oil that come along when you suck the crystals, so if necessary you can use more µl's. Now quickly scoop the crystal with a loop and dip in the cryoprotectant and freeze it, in the usual way. This has worked for me, though it may need some more standardization depending on your case.

In one case I had crystals grown under paraffin oil in microbatch, which required 25% glycerol for cryoprotection. However, when I transferred them directly to the base condition + 25% glycerol, they would invariably break up. So I ended up picking up a crystal with a loop, pulling it through the oil, and putting it on a coverslip. Then I removed the surrounding liquid and quickly replaced it with 4 µl of base solution + 5% glycerol, observing the process under the microscope. The crystal seemed to hold up fine, and I then progressively replaced the solution with base condition + 10%, 15%, 20% and finally 25% glycerol. The resulting crystal froze nicely, and provided a 2A dataset. Any oil that was still around the crystal from the first transfer was probably completely removed by the cryo/washing steps, as other people have already described.

Handling Crystals continued...

Microbatch Crystal Mounting Tips...continued

We do the same, I prepare 2-3 drops of cryoprotectant on a microscope slide, draw the crystal through the oil and "wash" in successive cryo drops until I see no more oil coming off. Paraffin oil itself is mot a good cryo protectant, if you want to use paratone oil you can use the same procedure, or first get rid of the paraffin oil in artificial motherliquor drops, then tranfer to paratone.

I have very tough time mounting crystals from microbatch. Essentially, I was adding the cryoprotectant to the drop and then transferring crystals to another drop of cryoprotectant. Yet, I lost many crystals before successful mounting. At times, I tried to remove the oil layer and then add cryoprotectant, but it was always difficult.

I harvested exclusively from microbatch under oil for a little over half a year due to a crystal form that grew in 20%+ volatile alcohol (hanging drop tended to swirl about something fierce when unsealed). I found that the main problem is getting the crystal up through the oil, as paraffin oil tended to knock crystals out of loops when pulling out of the aqueous layer. Al's oil is a little easier, but still not a cake walk. However, residual oil on the surface of the drop/loop didn't significantly affect diffraction. It was a little harder to see the crystals in the loop with blobs of oil getting in the way, but not impossible

----

A few schemes that worked well for me:

1. Step-wise supplement the drop with small volumes of mother liquor + over-concentrated cryo. For example, if starting with a 2ul drop, add 0.5ul at a time of mother liquor + 30-40% glycerol. Once you've done this 4-6 times, you'll have (relatively) gently moved into 20% glycerol while staying under oil. Different crystal forms may demand different steps/concentrations of the cryo stock.

or...

2. Use a pipete to remove all but a thin layer of oil from the top of the drop and then increase the drop volume to 4-6ul with mother liquor. This leaves a thin film of oil to slow down evaporation, but isn't quite so hard to pull crystals through. If you aren't in a volatile condition, you could just remove all of the oil by flooding the drop this way . Then transfer to a separate well with cryo (also covered with a slick of oil, if the condition is volatile).

3. As for crystal/loop handling, the hanging drop method of setting the loop parallel to the surface of the drop and then pulling up is awful for microbatch, as this causes the oil to shove the crystal out of the loop almost every time. I found that 45 degrees to nearly perpendicular to the surface of the drop tended to work best.

A nice way to use cryo is to add it under the oil into the drop let it sit for the time you need and just fish it with the oil then the crystals will be already protected. It worked for me and others in the lab and it also enlarge the volume of the drop so you can easily fish.

I don't typically use the oil as a cryoprotectant.

First I remove *most* of the oil from the plate, and then remove the bulk of the oil from the well where I'll be working. Next, I use a piece of clay to hold the plate at an angle, which makes it easier to get at the well. Then, I either

(a) fish the crystal directly from the well (through the remaining oil) and transfer it to a small drop of cryoprotectant on a cover slip; or

(b) I use a pipet to transfer the drop to a cover slip and then work from there (in the latter case, you'll bring some oil along, which will coat the drop after you deposit it on the cover slip, which helps prevent evaporation and allows you to perform the subsequent steps at a leisurely rate).

Handling Crystals continued...

Microbatch Crystal Mounting Tips...continued

We pour off the oil and then add a harvest solution (~10-20 ul to each terasaki plate well) with ~2% more PEG than that in the droplet (remember that the drop concentrates with time - sometimes fine tuning of the harvest solution PEG concentration is required but generally 2% works.). This keeps the crystals stable and displaces any oil remaining at the top of the droplet. Sometimes there is a skin on top of the droplet. This is removed by a capillary or acupuncture needle. Then we remove the crystals as per sitting drop. We have tried to cryo-protect the crystals by just pulling them up through the oil but this has not worked for our crystals. It may be possible to add a cryoprotectant to the harvest solution but we haven't tried it.

Comments

1. Most people simply leave the oil, and loop the crystal out through the oil.

2. The crystal must fit the loop well, or the surface tension will drag it off the loop.

3. People who are used to working with oil claim it is easier than sitting or hanging drop because the oil prevents evaporation. Therefore you don't have to work fast.

4. A loop with a bent handle can be helpful, because you have to harvest crystals from above.

5. Tip from Jeroen Mesters: if you want to use the oil as a cryoprotectant, you must DRY the oil first. Use ordinary paraffin oil, and put 1 ml aliquots into Epi-tubes. Dry tubes e.g. in a SpeedVac overnight. Freeze tubes, and use for a day, then discard. To use as a cryoprotectant, place a drop of dry oil on a glass slide. Drag the loop with the crystal through the oil along the surface of the glass. Small drops of mother liquor will stick to the slide. Remove as much mother liquor from around the crystal as possible, and freeze in the normal way. (Have I got that right Jeroen?)

I've mounted many crystals from microbatch and sitting drop plates and haven't found it difficult to mount xtals from either! (well, of course a bit difficult when I first started mounting xtals!!)

Typically I try to remove the oil by dragging the xtals gently in the cryoprotectant drop until I see the oil layer is removed (at much as I can judge visually!). It's a tricky job removing the oil completely, depending on the cryo (eg. paratone!!!!) so great attention is needed in doing this part! I transfer the xtals to a fresh drop of cryo before mounting.

----

1. Try to change to hanging drop vapor diffusion, by halving the precipitant.
2. Flood the microbatch drop with the first stabilizing/cryosolution, then remove the oil layer.

We usually dip the loop straight through the oil drop and try to 'fish' out the crystal. We then transfer it to a fresh drop of cryosolution and scoop it back out before mounting it in into the beam. This however isn't as easy as it sounds and takes a little practice. We do sometimes remove the oil drop (partially) to make it a little easier. Furthermore, if we use oil as a cryoprotectant, we would transfer it to a paratone (not paraffin oil) drop before mounting the crystal.
I have never heard anything about using the paraffin oil on top of the drop as a cryoprotectant. However, I found out that using paratone as a cryoprotectant is also very difficult, and as for all the described techniques, practice makes perfect.

Crystal and Age

How Does the Age of a Crystal Affect X-Ray Diffraction Data?

Reading and References:
The crystal structure of yeast phenylalanine tRNA at 2.0 Å resolution: cleavage by Mg2+ in 15-year old crystals. Luca Jovinea,, Snezana Djordjevica and Daniela Rhodes. Journal of Molecular Biology, Volume 301, Issue 2, 11 August 2000, Pages 401-414.

Abstract: We have re-determined the crystal structure of yeast tRNAPhe to 2.0 Å resolution using 15 year old crystals. The accuracy of the new structure, due both to higher resolution data and formerly unavailable refinement methods, consolidates the previous structural information, but also reveals novel details. In particular, the water structure around the tightly bound Mg2+ is now clearly resolved, and hence provides more accurate information on the geometry of the magnesium-binding sites and the role of water molecules in coordinating the metal ions to the tRNA. We have assigned a total of ten magnesium ions and identified a partly conserved geometry for high-affinity Mg2+ binding. In the electron density map there is also clear density for a spermine molecule binding in the major groove of the T?C arm and also contacting a symmetry-related tRNA molecule. Interestingly, we have also found that two specific regions of the tRNA in the crystals are partially cleaved. The sites of hydrolysis are within the D and anticodon loops in the vicinity of Mg2+.

Atomic resolution structure of a succinimide intermediate in E.coli CheY. Simonovic M, Volz K. J Mol Biol. 2002 Sep 27;322(4):663-7.

Abstract: Isomerization of aspartate to isoaspartate occurs spontaneously in proteins, causes changes in protein structures, and correlates positively with the aging processes of many organisms, including Alzheimer disease in humans. Aspartate isomerization proceeds through an unstable cyclic succinimide intermediate. There are few protein structure determinations that have characterized the intermediates and products of this isomerization reaction. Here we report the discovery of an unusually stabilized succinimide ring in the 1.1A structure of the Escherichia coli CheY protein, as determined from a crystal eight years old. The ring is formed by the side-chain of aspartate 75 and the backbone nitrogen of glycine 76 in an exposed loop of the molecule. Stabilization of the succinimide is through interaction of a sulfate ion oxygen atom with the imide nitrogen atom. Formation of the ring caused conformational changes in the loop, but did not alter the overall structure of the protein.

Using rational screening and electron microscopy to optimize the crystallization of succinate:ubiquinone oxidoreductase from Escherichia coli. R. Horsefield, V. Yankovskaya, S. Törnroth, C. Luna-Chavez, E. Stambouli, J. Barber, B. Byrne, G. Cecchini and S. Iwata. Acta Crystallographica Section D, Biological Crystallography Volume 59, Part 3 (March 2003).

Synopsis: Crystals of SQR that diffract to 2.6 Å were obtained by rational screening and sample quality analysis using electron microscopy.
From the publication: "It proved critical to freeze the crystals within 72 h of crystallization set-up. Crystals that were frozen after this time limit showed no diffraction. This alteration in properties was apparent by the change in colour from deep orange to pale yellow that was observed in crystals more than four weeks old (Fig. 1c). The deep orange colour of the crystals is attributed to the presence of haem b within the protein. The loss or breakdown of haem, demonstrated by the change of colour in the crystals, could lead to structural instability and consequently loss of diffraction. "

CryoLoops for X-Ray Data Collection at Room Temperature

We use a much easier test for mounting crystals at room temperature. Just coat the crystal with Paratone-N (HR2-643) and mount your crystal in a standard CryoLoop. The Paratone-N will slow down evaporation enough - no special tools required. You don't even need to remove all the liquid as you would do for flash-cooling the crystal in Paratone-N. A major advantage is that you can use the same crystal to collect under cryo conditions and directly compare the impact of cooling the crystal.

Reference: CCP4 Bulletin Board January 16, 2009 from Filip Van Petegem, The University of British Columbia.

Methods, Procedures, Technical

Using Stock Solutions from the Refrigerator

When using stock solutions from the refrigerator or any one that has been standing for some time, be sure to thoroughly mix it before use. Storing stock solutions in the refrigerator can result in condensation forming on the inside upper portion of the bottle. If this condensation falls into the stock solution, it will create a heterogeneous stock concentration which in turn could make reproducing the experiment difficult. Condensation can occur with any seal when there is a temperature change in the environment. So it is possible for this phenomena to occur with room temperature stored stocks if the temperature in the room changes significantly. To avoid such problems, it is generally good technique to mix crystallization reagents prior to pipetting.

Want Control and Accuracy?

Use a gel-loading pipet tip to dispense small drops with better control and accuracy of drop placement.

Mixing Drops

If you routinely mix the sample with the precipitant by aspirating and dispensing the drop several times, try setting up the drop without mixing. Simply pipet the precipitant into the sample and let the two diffuse together without mixing. On the other hand, if you usually set up your drop without mixing, try mixing. During the 2005 RAMC meeting, a show of hands survey among 100 participants showed most people do not mix their drops.

Make Connections

Communicate. Call (no e-mail allowed) a colleague and ask them their favorite crystallization trick or tip. Offer up one or two of yours.

PEG 400 and MPD

If you are using PEG 400 as a precipitant with some success, try MPD. If you are using MPD, try PEG 400.

Read the Classics

Read the classics. Try Laboratory Experiments in Biological Chemistry by James B. Sumner and G. Fred Somers, Academic Press 1949.

Trying Different Precipitants

Escape the local minima of using the same precipitants. If something worked once, great. Try it again, but remember to try something new.

Check Out Temperature

Try different temperatures. Room temperature and 4°C are the norm and work quite fine. Anything between 0°C and 37°C is fair game. If you do not have an incubator, simply look for a cool or warm space in the building to incubate the plates.

Free Interface Diffusion

Try free interface diffusion as a crystallization technique. This is easy to do and requires only a small amount of sample. Using a 20 µl microcapillary pipet, pipet 2 to 5 µl of sample into one end of the capillary. Next, carefully pipet 2 to 5 µl of precipitant SLOWLY into the capillary without introducing an air bubble. Seal the ends with wax or clay. It helps to pipet the less dense material onto the material of higher density (i.e. pipet sample onto 30% PEG 8,000, pipet 15% Isopropanol onto the sample). Use gel-loading pipet tips to load the sample and precipitant.

Removing Mother Liquor

Tired of cutting filter paper for wicks to remove mother liquor from capillaries during crystal mounting? Try our Paper Wicks™. The 55 mm X Fine Long (HR4-211) work great in 0.75 mm capillaries.

Methods, Procedures, Technical continued...

PEG Sensitivity

Keep stocks and working solutions of PEGs in the dark. Use amber or solid colored bottles or wrap clear bottles in foil. If you want to be really careful, purge with Argon and freeze them.

About Growth Rate

As a general rule, larger and finer crystals are obtained if the growth rate is minimized. If you are growing microcrystals, reducing their growth rate by more gradual equilibration with the reservoir may result in their attaining larger size.

Reducing Agents

Consider using a reducing agent as a variable in the crystallization. Include a small amount (0.001 to 0.01 M) of a mild reductant such as TCEP hydrochloride, Cysteine, B-Mercaptoethanol, Glutathione, or Dithiothreitol in the mother liquor. Besides protecting sensitive Cysteine residues, there is some indication that antioxidants play some other undefined role in crystallization. Try levels of reducing agent much higher than that generally thought necessary to maintain sulfhydryl function.

Do Not Disturb

Once crystallization conditions are relatively well established, set up a crystallization trial and do not examine or disturb it in any way for several weeks. Premature handling of crystallization trials can convert a few promising growth centers into a massive shower of microcrystals.

Using Sodium Azide

Keep crystallization samples and hardware free from microbial presence by filtering all solutions through a 0.22 micron filter. Store stocks and working solutions in sterile containers. In most cases this negates the need for antimicrobials such as Sodium azide.

Bioseparation

Consider crystallization as a tool for bioseparation.

Temperature Effects

Proteins in high salt are usually more soluble at lower temperatures than higher temperatures. The reverse is generally true for PEG and MPD. Keep this in mind during initial screening and optimization. Also keep in mind that this is not always true.

pH pHun

pH adjustments in the drop can be made by adding a small amount of concentrated Ammonium hydroxide to the drop or reservoir to increase the pH, or Acetic acid to lower the pH.

Cold Organics

If you are using organic solvents as precipitating agents, be sure to include 4°C incubations with your trials.

How to Dilute

A 1:10 dilution can be defined as 1 part of stock plus 9 parts of diluent. For example, to make 1,000 µl of 0.1 M buffer from 1.0 M buffer, simply pipet 100 µl of 1.0 M buffer into a tube and add 900 µl of diluent (water).

Methods, Procedures, Technical continued...

What is Your Definition of % W/V?

Percent Weight/Volume (% w/v) = the weight in grams of a solute per 100 milliliter of a solution. For example, a 50% w/v solution of PEG 6,000 is made by adding exactly 50 g of PEG 6,000 solid to a total, final volume of 100 ml. NOT by adding 50 g of PEG 6,000 to 100 ml of water.

Removing Urea and Acrylamide

Urea and Acrylamide contamination of RNA samples from denaturing gels cannot be removed by dialysis. Acrylamide and Urea can be removed using a strong cation exchange resin. Joe Ng, UAB.

Path and Endpoint

The path is as important as the endpoint. Vary the method, drop ratios, and reservoir volume to alter the path to the same endpoint as a variable during optimization.

Inorganic False Positives

Beware of inorganic salt crystals when using Phosphate, Carbonate, or Borate buffers!

Condensation

Tired of condensation problems in the plates at 4°C? Fill two dummy plates with water, seal and place one at the top and the other at the bottom of your stack of plates. Voila! No condensation problem!

Taking Care of PEGs

Polyethylene glycols are inherently unstable and are prone to produce aldehydes and peroxides. The presence of these oxidation byproducts increase with the age of the PEG and is promoted in the presence of UV light. One can slow the production of aldehydes and peroxides by storing the PEG at low temperatures and in the absence of light and oxygen. Flooding the storage bottles with Argon will help limit the oxidative effects of oxygen on the PEG. An increase in the oxidative byproducts will inherently change the pH, conductivity, and ionic strength of the PEG.

Sodium Azide and Heavy Metals

Never use Sodium azide in crystallization reagents that contain heavy metals since the combination poses the danger of forming explosive metal azide salts (See Problems associated with the use of azide as an inhibitor of microbial activity in soil, M. Rozycki and R. Bartha, 1981, Appl. Env. Microbiol. 41: 833-836).

Equilibration Kinetics

Recently we chatted about ways to alter the equilibration kinetics in a hanging or sitting drop vapor diffusion experiment. Another way to fool with the kinetics is to perform the experiment at a cool temperature. Mikol et al have described how both the protein and precipitant are concentrated three to five time faster at 20°C than at 3°C. For details see V. Mikol, J.L. Rodeau, R. Giege, 1990, Anal. Biochem. 186, 332-339, Experimental determination of water equilibration rates in the hanging drop method of protein crystallization.

Physical Chemical Aspects

Always take into consideration the effect that salts, polymers, organic solvents, and additives might have on the pH of the buffer selected for the crystallization experiment. To be safe, check the pH of an experimental solution after all of the reagent components have been added, mixed, and the solution has equilibrated to the temperature where the buffer was titrated and the experiment is to be performed.

Methods, Procedures, Technical continued...

Using Different Reservoir Volumes

Although one typically uses 1,000 µl of reservoir solution in a VDX™ or Linbro® Plate, one can actually get away with 500 µl of reservoir solution before having to begin worrying about the excessive effects of evaporation which occurs when using polystyrene crystallization plates. The utilization of less volume in the reservoir can alter the kinetics of equilibration between the reservoir and the drop. Less volume equates to slower (longer) equilibration times. This is not necessarily a bad thing when screening for preliminary crystallization conditions. For a good read on the kinetics of macromolecular crystallization see "Kinetic Aspects of Macromolecular Crystallization" by Joe Luft and George DeTitta, page 110-130, Methods in Enzymology, Vol. 276, Part A, 1997.

Crystal Annealing

Flash-cooling crystals can sometimes increase the mosaicity of biological macromolecular crystals. In some case, macromolecular crystal annealing can reduce the mosaicity of flash-cooled crystals without affecting molecular structure. Crystal annealing involves cycling a flash-cooled crystal to room temperature and then back to cryogenic temperature. The procedure can also be applied to sometimes restore diffraction from flash-cooled crystals that were not properly handled to and from cryogenic storage. Crystal annealing does not seem to improve a poorly diffracting crystal suffering from molecular disorder. The essential features of the crystal annealing procedure are the following. The crystal is first flash-cooled. The crystal is removed from the cryostream and quickly transferred to a drop of cryoprotectant at crystal growth temperature and allowed to remain in the drop for at least three minutes. The drop should be covered to prevent evaporation. Finally, the crystal is remounted on a loop and flash-cooled. Reference: Macromolecular Crystal Annealing: Overcoming Increase Mosaicity Associated with Cryocrystallography, J.M. Harp, D.E. Timm, and G.J. Bunick, Acta Cryst. (1998) D54 622-628.

Isoelectric Homogeneity

Protein preparations which exhibit microheterogeneity are sometimes not amenable to crystallization or yield poorly diffracting crystals. Isoelectric heterogeneity is one such biochemical crystallization variable. Protein isoforms are often the result of modifications such a phosphorylation, methionine oxidation, or glycosylation. Such modifications can cause very small changes in the isoelectric point (pI) of proteins. These isoforms can sometimes be resolved by chromatographic techniques such as ion exchange (Crystallization of Isoelectrically Homogeneous Cholera Toxin, Westbrook & Spangler, Biochemistry 1989, 28 1333-1340). Conventional preparative isoelectric focusing techniques utilizing carrier ampholytes can often resolve isoforms, but suffer from pH gradient instabilities and residual contamination of the proteins by ampholytes. An alternative technique that can avoid these complications, PrIME (Hoefer Pharmacia Biotech) has been used successfully to further purify proteins which have subsequently been crystallized (Fab fragment of anti HIV gp-41 monoclonal antibody and isoforms of epidermal growth factor - Righettia et al).

Stabilizing Crystals with Glutaraldehyde

Soak crystals overnight in a solution comprised of the mother liquor utilized to grow the crystals plus 0.5 to 2% v/v Glutaraldehyde. Glutaraldehyde cross-links proteins through the epsilon amino group of lysines. Cross-linking can be used to stabilize fragile crystals for seeding or data collection.
Reference: F.A Quiocho and F.M. Richards. Proc. Natl. Acad. Sci. U.S.A. 52, 833 (1964).

Building Blocks

During the growth of a high quality crystal, identical components are incorporated into a periodic lattice. Thus, sample homogeneity, the presence of uniform biological macromolecules, is typically essential for the growth of high quality crystals.

Homogeneity

If members of a homogeneous population alter their form, then the population is no longer homogeneous. Stability of a population is important in order to maintain homogeneity. One should take measures to preserve the stability of the sample and make it resistant to change, in order to prevent denaturation and aggregation.

Read the Tea Leaves

We set screens in anticipation of growing crystals. In reality, we are looking for ways to alter the solubility of the sample in order to grow a crystal. Therefore, review screens looking for more than crystals. Look for variables that alter the sample's solubility. The big hitters are sample concentration, reagent type and concentration, pH, and temperature.

Methods, Procedures, Technical continued...

Reverse Vapor Diffusion

Reverse vapor diffusion can sometimes be used to crystallize a protein in low ionic strength. This is an alternative method to dialysis that is less tedious to set up but also allows the protein concentration to become more dilute as the experiment proceeds, whereas dialysis can maintain a constant protein concentration. Begin with the concentrated protein in salt. The concentration of salt should be sufficient to maintain the solubility of the sample, perhaps 0.1 to 0.5 M salt. Set the sample for crystallization using the hanging or sitting drop vapor diffusion method. Pipet deionized water into the reagent reservoir. Pipet the high ionic strength protein sample onto the cover slide (hanging drop) or post (sitting drop). Do not add water to the sample. Using only deionized water in the reagent reservoir will allow water to diffuse from the reagent reservoir into the drop. As the drop size increases by vapor diffusion, the ionic strength in the drop will decrease. The protein concentration will also decrease, so it is a good idea to begin with a concentrated protein sample. It is also helpful to avoid having too much salt in the samples since this may cause the sample drop to absorb too much water and dilute the protein concentration. For cryoprotection, one can add Glycerol to the reagent reservoir and mix this with the sample drop.

Gentle Crystal Cross-Linking with Gutaraldehyde

To cross-link fragile crystals for a chance at improved handling, solution changes or cryo, try the following: If you have hanging drops, then transfer the cover slide to a Cryschem™ 24 well sitting drop plate or place a Micro-Bridge™ into the 24 well hanging drop plate. Pipette 2 to 5 µl of 25% Glutaraldehyde into the sitting drop post and pipette your crystallization reagent into the surrounding reservoir. Seal the experiment. Allow the Glutaraldehyde to vapor diffuse into the sample drop containing the crystal. Thirty minutes to six hours is generally sufficient. Glutaraldehyde relies primarily on lysine residues so the number of lysines as well as temperature can be two significant variables influencing cross-linking time. Note that amines will interfere with Glutaraldehyde so Ammonium sulfate or Tris buffer will have to be dialyzed or exchanged out and substituted before cross-linking is attempted.

Note: Glutaraldehyde is toxic and precaution was taken against exposure to the reagent by conducting the cross-linking reactions in a chemical fume hood.

Reference:

1. J. Appl. Cryst. (1999). 32, 106-112 . A gentle vapor-diffusion technique for cross-linking of protein crystals for cryocrystallography. Carol J. Lusty

Cryschem Plate - Circle Your Hits

Circle your drops with crystals or promising results. With a Cryschem plate sealed using tape, simply use a Sharpie® to circle the drops that look promising and those that have crystals. This makes it easy to spot these drops later during follow up. We suppose you could get fancy and use different colors to code the results, but a simple black Sharpie circle around a well is easy to spot, even through a stack of plates.

Seeding from Wild Type Crystals to Grow Mutant Crystals

If you are able to obtain crystals of the native or wild type protein and cannot get crystals of variants or mutants of the native protein, try streak seeding. Streak seed from the wild type crystals into drops containing the variant/mutant.

Glover et al. streak seeded from wild type crystals, either immediately or after a day of equilibration. The seeding stock solution was prepared by diluting a drop with small crystals 100-fold using the reservoir solution. Streak seeding was performed by dipping a hair into the seeding stock solution and rinsed twice in the reservoir. Crystals grew to 100-400 micron in 1 to 2 weeks.

Reference:

1. Principles of protein-DNA recognition revealed in the structural analysis of Ndt80-MSE DNA complexes. Jason S. Lamoureux and J.N. Mark Glover. Structure 14, 555-565, March 2006.

Using Evaporation to Control Crystal Nucleation

Jovine (2000) used an evaporation method to control excessive nucleation and produce larger crystals.

A vapor diffusion experiment is set with the sample and reagent concentration where no crystals form when the drop and reservoir are in equilibration with one another. Nucleation is effected by opening each vapor diffusion experiment (remove cover slide, film or tape) for a precise amount of time. The amount of time used by Jovine was 100-120 seconds. The amount of time is empirically determined and will depend upon drop volume, reservoir volume, temperature, and reagent formulation. The opening increases the evaporation of water from the drop and creates a small and temporary increase in the relative supersaturation, hopefully enough to promote nucleation before the system returns to equilibrium when the experiment is sealed. If the drop remains clear, the procedure can be repeated until crystals start to appear. If too many crystals appear, reduce the amount of evaporation time.

Methods, Procedures, Technical continued...

Using Evaporation to Control Crystal Nucleation continued...

This method can also be applied to experiments which remain clear, as a way to effect a change in the drop's relative supersaturation, which could result in a crystal, precipitate or phase separation.

Reference:

1. *A simple technique to control macromolecular crystal nucleation efficiently using a standard vapour-diffusion set up. Luca Jovine. J. Appl. Cryst. Volume 33, Part 3, Number 2, 988-989 (2000).*

Alternative Reservoirs in Vapor Diffusion Experiments

Setting up vapor diffusion crystallization experiments against different reservoir solutions may have a profound effect on the outcome of crystallization experiments. It has been demonstrated that one way to increase crystallization space is to set the same conditions over different reservoirs (Newman 2005). So, rather than screen more reagents, one can screen the same reagent set over two or three different reservoirs (such as Sodium chloride, Ammonium sulfate and Polyethylene glycol 3,350), in addition to the crystallization reagent. Using the screen reagent in four sets of drops and then using the reagent in addition to three alternative reservoirs expands the crystallization space as the path of equilibration and endpoint is unique for each reservoir solution.

Recommended reservoir solutions
1.25 - 2.0 M Sodium chloride
1.0 - 1.5 M Ammonium sulfate
30-50% PEG 3,350
20% PEG 3,350, 0.2 M Sodium chloride

Replacement reservoir solutions (approximate)
0.5 M Lithium chloride = 35% PEG 4,000
1.5-2.0 M Lithium chloride = 2.5 M Ammonium sulfate

Sodium chloride should be expected to act as a simple desiccant. Ammonium sulfate can desiccate and affect the pH of the drop through the gaseous NH_3 which is in equilibrium with the NH_4^+ in solution. PEG 3,350 should desiccate, although with different kinetics compared to the salts (Luft & DeTitta 1995).

Alternative reservoir solutions have also been referred to as common dehydrants, a generic reservoir, and desiccants. The use of a single, simple, concentrated solution for the reservoir in vapor diffusion crystallization experiments has been happening for decades (Dunlop & Hazes, 2005; Hempel, 1968; Luft et al., 1994; McPherson, 1992).

To speed delivery of the reservoir for 48 and 96 well crystallization plates that are set up manually, the reservoir solution can be transferred to a Multichannel Pipetter Basin (HR3-269) and delivered to the plate using a multichannel pipette. For automated experiments, the reservoir solution can be transferred to a MASTERBLOCK® 96 Deep Well Plate (HR3-105) or a Multichannel Pipetter Basin (HR3-269).

References :

1. *Expanding screening space through the use of alternative reservoirs in vapor-diffusion experiments. Janet Newman. Acta Cryst. (2005). D61, 490-493.*
2. *Single Crystals of Transfer RNA from Formylmethionine and Phenyalanine Transfer RNA's.Arnold Hempel et al. Science, New Series, Vol. 162, No. 3860 (Dec. 20, 1968), 1384-1387.*
3. *A modified vapor-diffusion crystallization protocol that uses a common dehydrating agent. Dunlop & Hazes. Acta Cryst. (2005). D61, 1041-1048.*
4. *Luft et al. (1994). J. Appl. Cryst. 27, 443-452.*
5. *Luft & DeTitta. (1995). Acta Cryst. D51. 780-785.*
6. *McPherson. (1992). J. Cryst. Growth, 122, 161-167.*

Methods, Procedures, Technical continued...

Skin on Drop

Are you seeing skin form on the surface of your drops containing sample and reagent and wondering what to do?

Skins on the drop are considered bad. These skins are believed to be a layer of denatured protein. Skins can slow down the rate of vapor diffusion.

Pick away the skin before mounting your crystal. If the skin wraps around the crystal, it can affect the limit of diffraction.

To avoid skins, one can try crystallize under oil.[1] Either put a few µl of oil on top of the sitting drop, or use the microbatch under oil method.

Keep in mind that covering the sitting drop with oil will affect the rate of vapor diffusion equilibration. The following oils are ranked in order of the rate of equilibration allowed by the oil: Paraffin Oil (slowest rate), Als' Oil (medium rate), Silicon Oil (highest rate).

Reference:

1. Pearl, B.O'Hara, R.Drew, S.Wilson. Crystal Structure of AmiC: the controller of transcription antitermination in the amidase operon of Pseudomonas earuginosa. EMBO Journal 13. (1994), pp 5810-5817.

Vibration

Should one be concerned about vibration during crystallization?

1. "Many years ago at Imperial College we did a fairly systematic experiment on vibration using lysozyme (hanging drops in Linbro trays). We placed one tray in the basement on a very solid concrete base, one tray was placed in an incubator, frequently used and known to vibrate a little (with frequent door opening and closing), and the third tray was placed on a rather strongly vibrating metal plate hanging from a ceiling on a piece of wire. The vibrations were produced by a small motor with an acentric piece of metal attached to its rotor. The trays were examined at regular intervals. Crystals grew in all of them. The basement tray gave fewer but bigger crystals, but this may have been a small temperature effect - temperature was not controlled very well. Crystals from all trays were nice, they diffracted to the same resolution, with virtually identical Wilson distribution. After this experiment we stopped worrying about vibration." Tadeusz J. Skarzynski, 2006 CCP4 Bulletin Board

2. "...This is not to say, however, that absolute silence or stillness must prevail in the presence of growing crystals. At the Massachusetts Institute of Technology all the tRNA and protein crystals were grow in a cold room that had a giant compressor attached and contained, as well, several ancient centrifuges and shaker baths. One could observe standing waves in the reservoirs of all the vapor diffusion chambers and frequently could scarcely converse above the sound. Whether this had a positive or negative effect, we could not be sure. It is more likely that dramatic changes in the environment, as those caused by handling, are more disruptive than ambient conditions themselves." Alexander McPherson (1976) "The Growth and Preliminary Investigation..." in Methods of Biochemical Analysis, Vol. 23, page 284.

3. CCD Video Observation of Microgravity Crystallization of Lysozyme and Correlation with Accelerometer Data. E. H. Snell, T. J. Boggon, J. R. Helliwell, M. E. Moskowitz and A. NadarajahActa Cryst. (1997). D53, 747-755.
Synopsis: Stepped growth rates and crystal movements have been observed by CCD video in microgravity lysozyme protein crystallization experiments which correlate with `g-jitter' accelerations, especially of low frequencies. The spurts and lulls of crystal growth may, therefore, explain the residual mosaic block structure seen in protein crystal mosaicity and topography measurements.
Additional Comment by James Holton: They saw a very clear conneciton between vibrations and crystal growth rates and indeed crystal quality. Things like "astronaut exercising" coincided with increased crystal growth rate a short time later. So, it would appear that even in the most controlled environments vibration control can be a challenge, but at least on the Shuttle, where everything gets logged, you can (could) look for relationships.
And a comment from Eddie Snell: Basically, for vibration there are two factors, frequency and magnitude. Any high magnitude vibration is bad, i.e. the sudden short duration, hopefully low frequency one when you drop your tray :) For the same magnitude, low frequencies (a few Hz) are worse than higher. frequencies (mains frequency). High frequencies are quickly damped in a liquid environment. Measuring vibrations is fairly easy, translating that measurement into something useful is not. In the case James mentioned, we were looking at the reduced acceleration environment. Certain crystallization processes lasted over a longer time period than on the ground - in this case the presence of a halo of depleted solution around the crystal caused by the diffusion limited transport of the protein to the crystal face. Every time an astronaut exercised, this halo broke down and a growth spurt occurred (except for one astronaut who didn't exercise very hard at all, however he/she shall remain nameless). On the ground, this depletion zone is only there for a short period of time if at all. Transport and therefore growth is dominated by convection - the density differences cause flow of sample across the growing crystal face. Any effect from vibration at high frequency in a liquid medium would only be meaningful if the magnitude of that vibration was large. Diffusion processes and surface tension effects in the drop and Brownian motion and surface dynamics on the crystal face probably dominate the crystallization process at the macromolecule scale. Although I don't think vibration will be too big of an issue unless the magnitude is large I personally like small bench top Peltier controlled

continued next page...

Methods, Procedures, Technical continued...

Vibration continued...

forced air incubators. I can rule vibration out as a variable. Vibration isolation may make a difference with large magnitude vibrations, however vibration reduction would be cheaper. The best thing to do is try a simple experiment with a few of your favorite proteins. Set up a tray in one of the rooms if a colleague has one available and duplicate trays in other incubators set at the same temperature. If there is any big difference you have your answer (assuming it is not light, temperature control or a million other differences). Frequently I have often found that nice physical explanations in crystal growth are often poor approximations and later prove to be incorrect - a little shaking might be good after all. Perhaps that is why there are so many crystallography groups in California :)

His Tag and Ni Help Phasing

Structure of Mycobacterium tuberculosis mtFabD, a malonyl-CoA:acyl carrier protein transacylase (MCAT). Hemza Ghadbane, Alistair K.Brown, Laurent Kremer, Gurdyal S. Besraa and Klaus Futterera. Acta Cryst. (2007). F63, 831–835. The crystal structure of M. tuberculosis mtFabD, the mycobacterial MCAT, has been determined to 3.0 A ° resolution by multi-wavelength anomalous dispersion. Phasing was facilitated by Ni2+ ions bound to the 20-residue N-terminal affinity tag, which packed between the two independent copies of mtFabD.

Crystallization of (insoluble) Ligand Protein Complexes

Crystallizing ligands with limited or no water solubility can be tricky. Here are some suggestions.

A good review on the crystallization of protein ligand complexes can be found in Crystallization of protein–ligand complexes, Annie Hassel et al, Acta Cryst. (2007). D63, 72–79.

Solubilize the ligand in DMSO so it is maximally concentrated (100mM works fine). Add enough ligand to achieve two to three fold excess ligand to protein. Keep the DMSO concentration to no more than 3% to avoid damaging the protein. Set crystallization experiment using this sample. If you cannot achieve a high enough stock concentration of DMSO to be below the 3% threshold, dilute you protein in the storage buffer to approximately 1mg/ml. Add compound to two to three fold excess, incubate and co-concentrate to the desired concentration. This may help to avoid the DMSO shock. Alternatively one can incubate the concentrated protein with the compound solubilized in water for 24 to 48 hours and solubility of the ligand will be sufficient to complex with the protein. (Carsten Schubert ccp4bb December 2007)

One method that worked for me was to dissolve my ligand in 100% DMSO, as suggested in the previous response, then add a 3 molar excess of ligand to protein so that the final concentration of DMSO in the protein-ligand solution was no greater than 10% - of course the maximum concentration of DMSO that your protein can suffer will be protein specific but you could investigate this by incrementally adding DMSO to just a solution of your protein at your working crystallization concentration then measuring scattering at 600nm in a spectrophotometer to determine the critical DMSO concentration that causes your protein to precipitate (if you have sufficient protein to 'waste'!). If your ligand binds tightly to your protein at an equimolar concentration you can then remove excess ligand and DMSO by passing your sample through a G25 Sephadex column. (Rob Hussey ccp4bb December 2007)

If you grow crystals in polyethylene glycols or similar reagent you might try to solubilize the compound in a small amount of this reagent. This is helpful if you want just a 1:4 protein:ligand ratio. Sometimes solubility is low even in 5% DMSO (or diluted solutions of glycerol, alcohols and similar molecules). In these cases setting up drops in the presence a saturated solution and some precipitate of the compound may also lead to good co-crystals. As one molecule passes from the saturated solution to the “bound” state, a new molecule is solubilized from the precipitate, which gradually dissolves and passes from the solution to its binding site in the protein. Indeed, this sounds like a “soaking” experiment and it works well if the compound is colored, so that you can see if the crystals actually become of the same color. Just remember to wash them thoroughly before measurements, in order to remove traces of the ligand precipitate that would result in poor diffraction. (Marco Mazzorana ccp4bb December 2007)

We routinely obtain structures from protein solutions with a big pellet of ligand in the bottom of the tube. For co-crystallizations we add 1mM compound to a 0.3mM solution of the protein and incubate overnight. Many of the compounds are only soluble to 50 micromolar, so we get a lot of precipitate. The next day, we spin the tube at high speed, and use the supernatant for crystallization trials. We have started from 100 mM stocks in 100% DMSO or ethanol. This has worked for compounds ranging for picomolar to micromolar affinity, which surprised us, but it worked. (Kendall Nettles ccp4bb December 2007)

Methods, Procedures, Technical continued...

Glutaraldehyde to Differentiate Salt from Protein Crystals

Place the crystals in a low ionic strength buffered solution containing 0.2-.2.0% v/v Glutaraldehyde. Protein crystals will quickly be fixed (faster than they dissolve) into a light golden, gelatinous lump. Sometimes the crystals retain a crystal-like shape, other times the crystals leave just a rubbery blob. In contrast to protein crystals, salt crystals should dissolve over time and should not be colored.

One can add a small amount of 2% v/v Glutaraldehyde to the protein drop. Protein crystals in the presence of Glutaraldehyde will then turn a light golden color. Crystals fixed like this can then be put into a low ionic strength solution where salt crystals should dissolve. You can easily transfer Glutaraldehyde into a protein drop by vapor diffusion by adding it to the reservoir to make it 2-3% v/v.

A caveat to using this method is that there should be no free amines around other than on the protein (i.e., no Ethanolamine or Tris buffer, no Ammonium ions, etc.).

Note: Glutaraldehyde is toxic, causing severe eye, nose, throat and lung irritation, along with headaches, drowsiness and dizziness. It is quite volatile with a strong, obnoxious odor.

A suggestion from Michael Garavito, Michigan State University.

Making a Bigger Crystal or Changing Crystal Form

1. Screen additives. See Hampton Research Additive Screen™ and Detergent Screen™.
2. Use a crystallization screen as an additive screen. Make the reservoir 90% original reagent plus 10% crystallization screen reagent (Crystal Screen™, Index™, etc). Repeat using the entire screen.
3. Change the crystallization method: Sitting drop, hanging drop, microbatch, dialysis, free interface diffusion, etc.
4. Try limited proteolysis. Reference: In situ proteolysis for protein crystallization and structure determination. Nature Methods - 4, 1019 - 1021 (2007).
5. Add fresh protein to the drop once crystal growth has ceased.
6. Add water to the drop once crystal growth has ceased. Crystals may dissolve and grow larger or a new crystal may form.
7. Try the Silver Bullets™ screen from Hampton Research. Reference: Searching for silver bullets: An alternative strategy for crystallizing macromolecules. Alexander McPherson and Bob Cudney. Journal of Structural Biology 156 (2006) 387-406.

Optimization

Sialidase to the Rescue

Sialidase I from Clostridium perfringens (Glyko, Inc. Rosedale, NY, USA) was used by Ogawa et al (Acta Cryst. (2003). D59, 1831-1833) to remove terminal sialic acid residues and minimize the heterogeneity of the glycosyl structure on the protein and enhance crystallization. The protocol utilized 1 mU of Sialidase I in 50 microliters of 50 mM Tris-HCl buffer pH 6.8 at 310 K for 12 hours.

Substituting Sodium Malonate for Ammonium Sulfate

Xing and Xu (Acta Cryst. (2003) D59, 1816, 1818) recently describe the crystallization of the PX domain of cytokine independent survival kinase in Ammonium sulfate. Optimization of the crystallization seemingly involved the subtle manipulation of several crystallization variables, including using drop ratios (1:2 drop ratio), the removal of DTT to allow crystal growth after nucleation, as well as the substitution of Sodium malonate for Ammonium sulfate. The authors reported the Sodium malonate dramatically increased the reproducibility of the crystals and also acted as a good cryoprotectant. Crystals grown from 2.0 M Sodium malonate could be cryoprotected directly from the drop. The crystals grown in Sodium malonate were more resistant to physical shock compared to those grown in Ammonium sulfate.

Sample Purification as an Optimization Variable

If one is able to obtain crystals of a protein but the crystals do not diffract well, try further purification of the sample and repeat the crystallization conditions. Oftentimes crystals do not diffract well because of impurities or aggregates present in the sample. Impurities in the sample can cause dislocations and defects in the crystal, which can lead to poor diffraction. Further purification can sometimes remove these impurities and aggregates, resulting in crystals with improved diffraction. Oftentimes the reason crystals do not diffract can be traced to the molecules in the crystal not being well ordered in the lattice. Therefore, there is conformational flexibility or the molecules do not take a desirable or fixed conformation because the lattice forces are too weak. Or sometimes there are severe defects in the crystal lattice, which lead to disorientation.

Drop Size

Try increasing the drop size to grow larger crystals.

Remove Aggregates

Centrifuge the sample at 2,000 g for 10 to 15 minutes immediately prior to setup to remove aggregates and amorphous material.

Using Additives

During optimization of crystallization conditions, examine any additive that might for some reason tend to stabilize or engender conformity by specific interaction with the macromolecule.

Saving Kits

To reproduce crystallization conditions from a screen without taking additional solution from the screening kit (and making all your lab partners want to do evil things to you), simply set an additional drop or drops on the cover slide next to the original drop. One can set multiple drops on the same cover slide over the same reservoir. Pipet the sample onto the cover slide, then pipet the precipitant from the reservoir into the sample drop. Lastly, seal the cover slide over the reservoir.

Optimization continued...

Varying Ionic Strength

It has been suggested that lattice contacts within a crystal can be manipulated using ionic strength, which in turn can be used to influence crystal morphology. Rather than the type of salt used (Sodium chloride, Sodium acetate, Sodium citrate, etc.), one may manipulate the concentration of the salt or change crystal morphology. Typical salt concentrations to be screened for such manipulation are 0 to 500 mM.

Explore All Angles

Explore as many optimization opportunities for crystallization as possible. Optimize every hit from a screen. When you reach an optimization dead end for a particular hit, review the screens again and optimize crystal leads from other hits. Optimize different hits simultaneously if sample availability permits. This can save time if one hit leads to a dead end (poorly diffracting crystal, non-isomorphous derivative, etc.).

Beauty is Only Skin Deep

One's tendency is to optimize the largest, best looking crystals. Common sense. However, there are plenty of reports where small, miserable, ugly crystals have eventually produced the best data, so do not ignore the ugly ducklings.

Control

Control is a good thing. Having a handle on the variables which affect the crystallization and the ability to control these variables is key to the reproducible growth of high quality crystals.

Fresh Ligands

When working with ADP, ATP, NAD, NADH or other chemically unstable ligands as crystallization additives, be sure to use freshly prepared working solutions or properly stored stocks. These agents are sensitive to degradation over time which can influence the outcome of a crystallization experiment. Also, consider the pH of these ligands in solution since the pH of the additive agent can influence the final pH of the drop. It might be necessary to titrate the ligand solution before addition to the drop.

Increasing Crystal Size

How might one increase the size of a crystal which is grown from dialyzing the protein into pure deionized water? The following replies were posted on the cc04bb (http://www.ccp4.ac.uk) on October 21, 2005.

George Kontopidis: If I was you the first thing I will try is to increase protein concentration. Also after crystal formed you can take the solution to room temperature that also may help.With the existing crystals just open the container or the vial where the crystals formed to allow further evaporation (do not let it dry out) which will result to increase of protein concentration and hopefully increase crystal size.

A.ErtugrulCansizoglu: You might try vapor diffusion against water. As such you keep you protein sample in a mild-salty condition- where it doesn't actually crystallize, then set up your drops using pure water. This worked for a case in our lab. Slowing down the crystallization made the crystals bigger.

Sameer Velankar: I once crystallized a protein in similar conditions. For that matter many Thymidylate synthase crystallize by decreasing the salt concentration. You can add very little salt (15-25mM) in the drop and keep water or 2-5mM salt in the well. This controls the growth of the crystals and if you are lucky you will get bigger crystals.

Christopher F. Snook:I would suggest increasing the drop volume, i.e. going from say 4 to 8 microliters. This has a two-fold effect: an increase in the protein available for crystallisation and a reduction in the equilibration of the drop with the well solution. Another possiblity would be to increase the concentration which may increase the nucletaion at the expense of crystal growth.

deacona@mail.rockefeller.edu: I have never encontered such a situation, but maybe some microdialysis from your storage buffer against pure water may help somewhat? I believe this is a common strategy for antibody crystallization. As usual, micro/macro-seeding, varying protein concentration etc may also be of use. Just some thoughts...P.S. These may be of some interest J Biol Chem. 1970 May 25;245(10):2763-4 A crystallographic investigation of a human IgG immunoglobulin.Edmundson AB, Wood MK, Schiffer M, Hardman KD, Ainsworth CF, Ely KR.

Optimization continued...

Increasing Crystal Size...continued

J Mol Biol. 1997 Dec 19;274(5):748-56. Use of organic cosmotropic solutes to crystallize flexible proteins: application to T7 RNA polymerase and its complex with the inhibitor T7 lysozyme. Jeruzalmi D, Steitz TA.

Oleg Tsodikov: I would still spin down the crystals (or dissolve them in the presence of small amount of salt or glycerol in and some buffer) and then screen in a variety of different conditions.

David Aragao: We had, some years ago, a very similar case (salting inn?). Our crystals started to appear in the purification process everytime the ionic strenght decreased. The final crystals were big and we got 1.2 Å resolution from them. I know there are nice comercial dialysis buttons nowadays but we made our ones at the time. Maybe the description of our optimization process can help you (see materials and methods). Acta Crystallogr D Biol Crystallogr. 2003 Apr;59(Pt 4):644-53.

jjwarren@duke.edu: I would recommend that you check out Jeruzalmi, D. and Steitz, T.A. (1997) Use of organic cosmotropic solutes to crystallize flexible proteins: application to T7 RNA polymerase and its complex with the inhibitor T7 lysozyme. J. Mol. Biol., 274, 748-56. Their methods are generally applicable, but they focus heavily on proteins that crystallize at low ionic strength...

Christopher Colbert: Seriously, you might want to consider screening for additives.

Ewa Skrzypczak-Jankun: Try salt - increasing the ionic strength often makes crystals more bulky. Also adding salt will change saturation point and can slow crystallization If nothing else NaCl is CHEAP. Do you have any information about affinity to ions? what your protein likes and dislikes? digging in the literature can provide valuable clues.

Stephen Prince: You could try slowing the rate of cooling - there was an old recipe for insulin crystallization that involved slow cooling. One used a thermos flask containing water at above room temp, the protein mix was floated in this in a small test tube, the thermos was sealed and placed in a polystyrene box and left undisturbed for a week or so, crystals grew in the tubes as the solution cooled. You could try the same trick in a cold room ...

Jonathan Caruthers: I've seen the same effect on at least two occasions - actually I believe hemoglobin was first crystallized by dialyzing blood in water if memory serves. In any event, we managed to grow diffractable crystals simply by removing them from the fridge, partially melting them, and putting them back in the fridge to re-grow. This works pretty well.

William Scott: The main concern is stability. If you can find a more conventional crystallization, it might ultimately make your life easier.
The good thing is you know it will crystallize readily. (I assume this is a good thing to know, unless you are trying to do NMR)

Raji Edayathumangalam: If you want to pursue these very crystals - try oil, which retards nucleation and you get bigger (and often better) crystals. We've had a case before where crystals falling out of the protein solution weren't the best bet. They were inherently disordered - not the optimal crystals. Crystal screens yielded crystals in other conditions that diffracted to higher than 2Ang.

Jon Caruthers : actually, to follow up here, I grew crystals of a MBP-protein fusion construct using diet-coke as a buffer. It turned out that these crystals diffracted better than anything grown with any other buffer of similar pH (2Å vs. ~3.5Å with other buffers), so maybe there's something to this idea of using CO2 as a buffer.

Tassos Papageorgiou: Try dissolving the crystals in salt and dialyse them afterwards against water might help. Have you try to see whethre crystals grow in low ionic strength?

David Waterman: I had a similar case with a protein which crystallised on a drop in ionic strength. The protein was soluble in ~100 mM K Phos, and started forming crystals when this concentration was reduced by diluting with water (or glycerol, which had the bonus of being a cryoprotectant). The biggest crystals (blocks with sides up to 0.3-0.4 mm) were obtained after setting up a hanging drop of concentrated protein over a reservoir of water alone (no mixing!) and allowing this drop to grow slowly by vapour diffusion. These crystals were difficult to cryo-protect though, so a compromise including glycerol in the reservoir and mixing with the protein drop was found. It was certainly much easier than fiddling with dialysis buttons, though I might have been lucky.

Peter Moody: I crystallised a DNA repair protein by dialysis to remove the salt that kept it soluble. Crystal quality was much improved with 300mM sugar (Maltose seemed the best).

Optimization continued...

No Diffraction - What to Try?

a. The protein may have a high degree of structural flexibility and therefore structural heterogeneity. Try screening ligands, additives, co-factors and other small and large molecules that might help the protein assume a more rigid conformation.

b. Structural variations in the protein due to enzymatic modification, partial deamination, partial oxidation, other post translational modifications can lead to sample heterogeneity. Analyze your sample and conditions from expression to solubilization to purification to crystallization to identify what might lead to sample heterogeneity. Remove it or change it. If working with an enzyme, bind an inhibitor or ligand to the enzyme in an attempt to lock into a more rigid and homogeneous conformation. Reduce the number of steps and the amount of time between expression and crystallization.

c. If the crystal is growing overnight, the high rate of crystal growth may indicate a high level in the number of defects in the crystal lattice. Slow the rate of crystal growth. Reduce reagent concentration or protein concentration. Layer silicon oil over the reagent reservoir to slow vapor diffusion. Use less reagent in the reservoir. Evaluate drop dilution and drop ratios. For example, dilute the protein 1:1 with your sample buffer. Set vapor diffusion experiments with diluted sample:reagent drop ratios of 1:1, 2:1 and 3:1. The final protein concentration will remain similar between the original and diluted samples but the path of equilibration will be different and the rate will be lower than for the original drop.

d. Mount the crystal, pre cryo, in a capillary and test for diffraction to confirm that cryo is not causing the poor diffraction. If room temperature data is better than cryo, your cryo method and cryo procedure might benefit from further refinement.

e. Screen additives.

f. Screen crystallization kits as additives. Mix 70 to 85 µl of the optimized crystallization reagent with 30 to 15 µl of solutions from your favorite Hampton Research screen (Crystal Screen™, Crystal Screen 2™, Index™, SaltRx™, etc.) to create a 100 µl solution. Mix thoroughly. Pipet 100 µl of this solution into a 96 well sitting drop plate. Create a drop mixing equal amounts of reagent and sample, seal and allow to equilibrate. Varying the ratio of optimized reagent: screen reagent and drop ratio can also be tried. Watch out for false positive salt crystals as these random mixes may create insoluble salts. Look in the reservoir for salt precipitates or crystals. If one prefers to use 24 well plates, simply change the ratio of optimized crystallization reagent:screen reagent. For example, in a Cryschem™ Plate, try 350-450:150-50. In a VDX™ Plate, try 750-900:250:100.

References:

1. *Personal communication with Annie Hassell at GlaxoSmithKline, circa 1996.*
2. *Acta Cryst. (2005). D61, 646–650. Crystallization of foot-and-mouth disease virus 3C protease: surface mutagenesis and a novel crystal-optimization strategy. James R. Birtley and Stephen Curry.*

Desiccating Crystals to Improve Diffraction

Desiccating, or drying a protein crystal can be a way to salvage or improve the diffraction quality. The procedure consists of removing a non-diffracting crystal from the X-ray beam, plunging it into a soaking solution made of the original crystallization reagent supplemented with a suitable cryoprotectant (glycerol, ethylene glycol, MPD, PEG 400, etc.) and then allowing the drop to dry in the evaporating sitting or hanging drop for 15 minutes to several hours. Try 1 ul cryo reagent plus 9 ul of original crystallization reagent for the drop. Remount the treated crystals and examine for diffraction. The procedure is originally described by Chantal Abergel. Reference: Spectacular improvement of X-ray diffraction through fast desiccation of protein crystals. Chantal Abergel. Acta Cryst. (2004). D60, 1413-1416.

Dissolve and Restore Protocol for Growing Larger Crystals

Impurities adsorbing onto the crystalline surface can prevent further growth of a crystal. Repair and growth of an impurity poisoned crystal is possible by using a dissolve and restore protocol described by Plomp et al (2003). By temporarily applying undersaturation conditions, an impurity adsorbed layer can be removed, which followed by saturation conditions can revive crystal growth with the potential for crystal growwth without the incorporation of impurities. Regular cycles of short periods of undersaturation followed by longer periods of supersaturation can be applied to dissolve and restore, with the potential of a larger crystal with fewer impurities and defects. Temperature manipulation (cycling) for samples with temeprature dependent solubility is a convenient approach. Altering and cycling sample and reagent concentration is another approach (Heinreichs et al (1992), Koeppe et al (1975), Przybylska (1989)).

References:

1. *Plomp et al (2003). Repair of impurity-poisoned protein crystal surfaces. Proteins: Structure, Function, and Genetics 50:486-495.*
2. *Heinreichs et al (1992). Growth of single protein crystalsin a periodically solubility gradient: description of the method and first results. J. Cryst. Growth 122:186-193.*
3. *Koeppe et al (1975). A pulsed diffusion technique for the growth of protein crystals for X-ray diffraction. J. Mol. Biol. 98: 155-160.*
4. *Przybylska (1989). A double cell for controlling nucleation and growth of protein crystals. J. Appl. Cryst. 1989:22:115-118.*

Optimization continued...

Heavy Atoms and Salt Reagents

Heavy atom derivatives can sometimes be difficult to obtain in the presence of sulfate salt crystallization reagents such as ammonium sulfate or lithium sulfate.

However, difficult does not mean impossible, so be sure to screen heavy atoms in the salt based condition before trying to cure a problem that might not exist.

When one is unable to obtain succcessful derivatives in the presence of sulfate salt based crystallization reagents, one might try transferring the crystals to high concentrations of a similar acetate salt based crystallization. For example, Ammonium acetate instead of Ammonium sulfate, or Lithium acetate instead of Lithium sulfate. This will maintain the high ionic strength of the crystallization reagent yet increase the solubility of some heavy atoms. Acetate, in higher concentrations can also act as a cryo salt.

Try replacing the original salt with Sodium malonate or Tacsimate. Malonate: a versatile cryoprotectant and stabilizing solution for salt-grown macromolecular crystals. T. Holyoak, T. D. Fenn, M. A. Wilson, A. G. Moulin, D. Ringe and G. A. Petsko. Acta Cryst. (2003). D59, 2356-2358.

Try halide soaking using KI or KBr. Novel approach to phasing proteins: derivatization by short cryo-soaking with halides Z. Dauter, M. Dauter and K. R. Rajashankar Acta Cryst. (2000). D56, 232-237.

One might also consider trying 100 mM Cobalt hexammine or Iridium hexammine as derivatives in the presence of acetate salts.

Try cross-linking. A gentle vapor-diffusion technique for cross-linking of protein crystals for cryocrystallography. Carol J. Lusty. J. Appl. Cryst. (1999). 32, 106-112.

Sulfate Ions in the Active Site

Problem: Enzyme crystals grown in Ammonium sulfate having trouble binding ligand.

Solution: This is not an unusual situation with sulfate, and, yes, sulfate often occupies phosphate binding sites, particularly when you're work with 1-3 molar concentration of Ammonium sulfate. The simplest way to remove sulfate from the crystal is to transfer to high concentrations of citrate. From my own graduate work on GPDH in the 1970's, we just transferred the crystals from 3.0 M AS to 1.44 M Sodium citrate. While the crystals weren't stable in citrate for long periods of time (days), that is not an issue today with cryocrystallography and flash-freezing. You might even try formate or malonate, which can have a modest cyroprotectant behavior. Another, and perhaps more relevant, method is the one devised by Bill Ray in the late 1980's for phosphoglucomutase (PGM). Apo-PGM was crystallized in high AS, which caused the same problems as you are experiencing. Ever the perfectionist, Bill devised a systematic method to transfer PGM crystals from AS to a PEG solution that allowed the formation of enzyme complexes (Biochemistry 30, 6866, 1991). Bill also looked at glycine as a replacement as well. So there are a number of ways to "desalt" crystals without resorting to cross-linking, while preserving good diffraction characteristics. Michael Garavito, Michigan State University.

Vary the Method of Crystallization

Using the crystallization reagent that produced the initial hit, or reagent formulations which have been partially optimized, set additional crystallization experiments using other methods and techniques. If sitting drop vapor diffusion was used to produce the initial hit, try hanging drop, vapor diffusion, microbatch, free interface diffusion, or dialysis. Try different plates and different apparatus using the same method to discover whether the difference in equilibration will produce a better crystal, a crystal in a shorter period of time, or a different crystal form. The volume of the chamber or well, the space between the drop and the reagent, the shape of the chamber and the well, and also the method and time required to set the experiment may affect the outcome of the crystallization experiment.

The path is as important as the starting and ending points in determining ideal nucleation and growth conditions.

Optimization continued...

Omit the Irrelevant

If the crystallization reagent producing a hit is a complex mixture containing, for example, salt, detergent, additive, metal ions, ligands, reducing agent, EDTA, cryogen, or other chemicals, prepare optimization reagents based on the hit, but systematically omit each of the components, one at a time. Which chemicals appear to make a difference in the crystal size, number and quality? Omit from future optimization experiments any chemical which appears to be irrelevant. Identify and remove irrelevant chemicals before adding the evaluating new chemicals. As with the sample buffer, keep the crystallization reagent as simple as possible. Include only chemicals which promote and enhance sample stability, homogeneity and crystallization.

Optimization is an Art

The skill of crystal optimization is an art form that improves with effort, thought, discussion, reflection, reading, learning, trial and error, patience, tinkering and time.

The path of optimization is rarely linear and the variables affecting crystallization are not independent. This makes it impossible to sample all crystallization variables and evaluate them at fine intervals.

Optimization may follow a main trail, but there are typically many branches in the trail to consider, explore and evaluate. And typically, time and sample availability do not allow one to explore each and every branch.

So one is left with the question of what they should try or what they should do next to increase the size or improve the quality of of the crystal. And there are usually plenty of recommendations to consider. But which recommendations should be considered and tried and in what order or priority, and which ones should be ignored for now and tried later, or perhaps discarded altogether?

It is the art of an experienced crystal grower to consider the results of systematic experiments of many different samples over much time and use this knowledge base to synthesize the best plan for growing perfect crystals.

Optimization can be tedious and require numerous iterations before one uncovers which combination chemical and physical variables produce the perfect crystal.

Optimization is an art and a technical skill. As such, it is dynamic and improves in direct proportion to the number of challenges, the degree of the challenges, and experience.

A good crystal grower never is, they simply get better and better and better...

Build a Tool Box

A skilled craftsman knows they cannot do their job without the proper tools. Maybe some jobs can be accomplished with one or two common and simple tools, but a skilled craftsman, over time, will build not only experience, but also a tool box. A tool box of both general and specialized tools that will allow them to handle all tasks in an efficient manner. And over time, a skilled craftsman gains appreciation of the importance that quality, reliability and support bring when deciding which tools to acquire.

A crystal grower also needs to build a tool box. This tool box will consist of apparatus and devices, plates and seals, kits and reagents; some general, some specialized. Initially the crystal grower will want to acquire tools that have general utility. Later, more specialized tools can be added so that when the need arises, these tools will be available to use and evaluate. Access to a useful and broad portfolio of tools ensures the crystal grower will have the ability to evaluate those variables which could lead to the perfect crystal.

You can only use and evaluate what you have and know how to use.

Sample Preparation

Imidazole and Crystallization

Should one remove 1.0 M Imidazole from the sample after elution from a nickel column? In many cases it is good to have (an) additional purification step(s) following the elution of the sample from a Ni column. Size exclusion chromatography (SEC) is a convenient and effective way to remove the Imidazole. If one finds that Imidazole is an important variable in maintaining sample homogeneity, one can also simply reduce the concentration from 1.0 M to a more reasonable level (0.02 – 0.2 M) for inclusion of crystallization reagents. In some cases, a protein might not crystallize in the presence of Imidazole. To remove the Imidazole, one can try precipitating the protein in Ammonium sulfate and resuspending the sample in low salt buffer and purify by ion exchange or SEC. One can also remove Imidazole by dialysis. In one instance (personal communication from Artem G. Evdokimov) decent crystals could only be obtained using 0.6 M Imidazole/acetate buffer and two Imidazole molecules were seen occupying hydrophobic spots on the protein surface. Sometimes the reason Ni column purified proteins precipitate without Imidazole is that the Ni ions leak from the Ni column. The presence of Imidazole may have been chelating this excess and trace Ni and once removed, the remaining Ni may cause the protein with a His tag to aggregate in solution. One may be able to remove this excess Ni using a Ni chelating resin (Hampton Research catalog number HR2-312). Or one can leave the Imidazole buffer with the sample in a more reasonable concentration (0.02 – 0.1 M) or try Citrate buffer (chelator) or EDTA. If you are working with a metalloprotein and need metals in the sample, using the chelating resin rather than leaving a chelating reagent in the sample is obviously a better choice.

Heat Treatment

If a protein does not crystallize or gives poorly diffracting crystals, one might try heat treatment of the sample. This is a quick experiment with minimal equipment required. The procedure can reduce and/or eliminate protein that is not folded properly from the sample. Place the sample at 4°C for one hour. Incubate the sample at 37°C for one hour. Centrifuge the sample or filter using a 0.22 micron filter. Screen the heat treated sample for crystallization. The time course and temperature may require optimization for best results. One can monitor the heat treatment using dynamic light scattering or an activity assay. Tip from Annie Hassell, GSK – IUCr 2005.

Sample Too Soluble To Crystallize

The "solution" we hit upon for a problem protein with a high solubility profile was to cocrystallize it with an Fab fragment. Not the easiest route, but it worked. Catherine L. Lawson, Rutgers University. Reference: Li H, Lawson CL. (1995) Crystallization and preliminary X-ray analysis of Borrelia burgdorferi outer surface protein A (OspA) complexed with a murine monoclonal antibody Fab fragment. J Struct Biol. (115)335-337.

Dealing with Glycosylation

When having trouble removing glycosylation from a protein before crystallization, try a sequential exo-glycosidase(s) then endo-glycosidase treatment in order give the endoglycosidase better access. Or, try to prevent glycosylation in the first place and try glycosilation inhibitors (tunicamycine for example inhibits N-glycosilation). Another option is to try the expression of the protein in GnTI(i) HEK293S cells which are unable to synthesize complex N-glycans (Proc Natl Acad Sci U S A. 2002 Oct 15;99(21):13419-24. Reeves PJ, Callewaert N, Contreras R, Khorana HG. Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line). One more alternative suggestion is to use your original expression system, but with Asn-Asp mutations of the glycosylation sites; assuming of course the mutant proteins will still be soluble which is straightforward to test. If the protein is resistant to treatment with Endo H or PNGase F under native conditions, try the Endo H treatment under only slightly denaturating conditions. Try screening a gradient of denaturating conditions to see if you can find a concentration that allows deglycosylation while not unfolding your protein.

There are reports where a fully glycosylated protein is able to be crystallized. So it may be worth a try to crystallize the glycosylated form. If the sample will still not crystallize, try deglycosylation, preventing glycosylation, or another cell line.

Sample Preparation continued...

Deglycosylation Summary

Deglyosylation Summary
(built on tips from the CCP4 community)

Glycoproteins are proteins to which a carbohydrate chain is covalently attached. Proteins modified in this way in many cases present a considerable challenge to X-ray structure analysis, because they do not easily form crystals. This is because the attached carbohydrate chains are often heterogeneous and flexible and interfere with the formation of crystal contacts. A well-known strategy to tackle this problem is to deglycosylate the glycoproteins prior to crystallization (see below for practical tips). This is usually done with endoglycosidases such as PNGase F (resulting in virtually complete removal of all N-linked glycans) or Endo H, Endo F1, F2, F3 (for partial deglycosylation). Also exoglycosidases can be used, alone or in combination with endoglycosidases. Alternatively, the glycosylation sites can be mutated (substitution of S/T to Ala or N to Gln or Asp in the conserved NXS/T motif). Expression in lower organisms like E. coli or Pichia pastoris is also an option, or - if a less radical procedure is required - genetically modified expression systems can be used that do not contain certain glycosyltransferases (e.g. CHO Lec cells).

Practical tips for enzymatic deglycosylation:

Several companies like Sigma-Aldrich and ProZyme have excellent theoretical and practical information about deglycosylation on their web pages, which are well worth studying. Both N- and O-glycosylation occur naturally in proteins, with N-glycosylation being the more common modification. N-glycans can generally be removed rather effectively by a single enzyme, PNGase F, whereas this is not the case for O-glycosylation, where several enzymes have to act in concert to exert the same effect. Deglycosylation can be tried both under native and denaturing conditions. While native conditions are in general preferable for crystallization purposes, deglycosylation under denaturing conditions is more effective and should for this reason always been done in parallel as a positive control. To achieve higher efficiency even under native conditions, one may experiment with using higher temperatures (up to 37°C, instead of 0-4°C) and/or longer reaction times (1-5 days; preferably adding some Sodium azide to the mixture to prevent bacterial growth).

In some cases, native deglycosylation does not work well. In these cases, one may try to deglycosylate the protein under denaturing conditions and then refold the protein. Alternatively, one may use only slightly denaturing conditions, by applying various mild detergents (like β-octyl-glucoside, Chaps, Triton X-100, SDS, etc.). One can also try to incubate the reaction mixture in a sonicating water bath. Yet another option is to deglycosylate the protein only partially using exoglycosidases (neuraminidase, galactosidase, etc.) instead of endoglycosidases.

In practice, one often starts out with a protein solution (concentrated to 0.5 to 5 mg/ml in water or a suitable buffer) and adds the glycosidase of choice (e.g. PNGase F, Endo H, Endo F1, F2, F3, neuraminidase or enzyme kits) in ratios varying from ca. 1:15 to 1:2000 (w:w). The deglycosylation reaction is then monitored regularly by taking samples and analyzing them on SDS-PAGE or IEF gels. A final check is preferably done by mass spectrometry. With some of the glycosidases like neuraminidase, one should be very careful to fully remove the enzyme (e.g. by gelfiltration), since it crystallizes easily even in minute concentrations. Another way to ensure complete removal is to use glycosidases as fusion proteins (coupled e.g. to glutathione-S-transferase; see Grueninger-Leitch, 1996) and passing the mixture through a GST affinity column (glutathione Sepharose) after the reaction is completed.

Of course, it is always well worth a try to also crystallize the protein in its fully glycosylated form. Some proteins even crystallize better in their glycosylated form, due to the involvement of the glycan chains in crystal contacts. Other proteins behave best when partially deglycosylated.

Glycobiology tools:

If one does get crystals from the glycosylated protein, one has to deal with handling carbohydrates, so here are a few practical tips: Modeling the ligand into the binding site, with or without supporting electron density, of course requires a 3D structural model of the carbohydrate ligand. This can be obtained either from scratch (e.g. using the modeling tool 'Sweet' from the www.glycosciences.de website, which converts carbohydrate sequences into 3D-models) or from a structure database (via the Uppsala HIC-Up server at http://xray.bmc.uu.se/hicup/ or directly from the PDB (http://www.pdb.org/) or other appropriate databases such as the Cambridge Structural Database (CSD; http://www.ccdc.cam.ac.uk/). The model of the carbohydrate ligand should then be refined and checked as carefully as the protein structure itself. Currently, however, tools for analyzing carbohydrate structures are not as widely known and applied as those for protein or DNA structures (see Kleywegt, 2003). A good tip is to check out the website at http://www.glycosciences.de. A couple of tools, such as pdb-care (to check carbohydrate residues in pdb files for errors) or carp (which generates Ramachandran-like plots for carbohydrates) or even GlyProt (to identify glycosylation sites in proteins and automatically attach them in silico) make life of structural glycobiologists significantly easier. Another database, currently under development, is EuroCarbDB (http://www.eurocarbdb.org/databases). Also, questions concerning carbohydrate nomenclature may be resolved by consulting the web site http://www.chem.qmul.ac.uk/iupac/2carb/.

by Ute Krengel, University of Oslo, September 2006.

Sample Preparation continued...

Compound Solubility with DMSO

Compounds produced in medicinal chemistry efforts often have low aqueous solubility when compared to biological ligands. When poor compound solubility is suspected, one may try Dimethylsulfoxide (DMSO) to solubilize the compound. Dissolve the compound in 100% DMSO and THEN dilute to a lower DMSO concentration. Dissolving a compound in dilute DMSO can be a very slow kinetic process.

Using L-Arg and L-Glu to Improve Solubility

Add 50 mM L-Arginine together with 50 mM L-Glutamic acid to the protein sample to improve protein solubility and long-term stability. Golovanov et al. reported a 3 to 8 fold improvement in solubilization with six different test proteins.

Reference:

1. A simple method for improving protein solubility and long-term stability. Alexander P. Golovanov, Guillaume Haubergue, Stuart A. Wilson, and Lu-Yun Lian. J. Am. Chem. SOc. 2004, 126, 8933-8939.

What Sample Concentration Should I Use for My First Crystallization Screen?

For most soluble proteins, 10 to 15 mg/ml in a sample buffer that promotes the sample stability, homogeneity and monodispersity will work fine for an initial crystallization screen.

A pre crystallization test such as the Hampton Research PCT™ (HR2-140) can be used to better determine the appropriate protein concentration for crystallization screening.

What to Use and What to Avoid in the Sample Buffer for Crystallization

Use the minimal number of chemicals and the minimal concentration of these chemicals necessary to promote sample stability, homogenity and monodispersity.

Use a buffer concentration between 10 -25 mM. This will allow crystallization reagents with a buffer concentration of 50 - 100 mM to manipulate the pH of the sample during the experiment.

Current trends seems to favor the use of 200 mM or less NaCl or KCl in the sample. Using no NaCl or KCl is just fine as well. It simply seems most overexpressed proteins are purified and prepared into a final sample buffer contain NaCl.

If the sample is known or expected to have free sulfhydyl residues on the surface, including a reducing agent such as TCEP hydrochloride can help to prevent oxidation of these residues. Oxidation of sulfhydryl residues can lead to sample aggregation. Avoid reducing agents if the sample has disulfide bonds in order to prevent cleaving.

If your screens contain di or poly valent cations such as Calcium, Zinc, Iron and others, avoid Phosphate, Borate and Carbonate buffers. Such mixtures can results in false positive salt crystals.

Phosphorylated Protein Samples

The crystallization of proteins with phosphorylation sites can often be difficult. It is imagined that such samples are susceptible to heterogeneity due to the presence of multiple species of protein in various states of phosphorylation. Careful sample purification and preparation is perhaps more of a significant crystallization variable than any particular crystallization trick or specialized screen. One can use a strong anion exchange chromatography with a shallow elution gradient to prepare a homogenous sample for crystallization. Isolation of a single phosphorylated form can produce crystals where as the presence of multiple phosphorylated forms can lead to precipitate, microcrystals or crystals less suitable for diffraction compared to those from a homogeneous sample. One may use mass spec to characterize the different elution samples following chromatography to assist in the selection of the best sample for crystallization.

Batch to Batch Sample Variation

Do not mix different batches of protein. Each batch of protein should be considered and treated separately. The reason being that the expression and purification conditions as well as the chemical and physical conditions are not identical for each batch. Even subtle differences between each batch can affect the outcome of a crystallization experiment. Each batch should be characterized and documented. Run at least an SDS-PAGE. If possible, consider also a native PAGE, IEF-PAGE, analytical gel filtration, mass spec and dynamic light scattering to characterize each batch. Save some protein from each batch for later experimentation, comparison, documentation, and reproducibility.

Screening

Co-Crystallize
Try crystallizing your sample with a substrate, coenzyme, inhibitor, or ligand. Sometimes these agents serve to fix the macromolecule in a more compact and stable form. This may impose a greater degree of structural homogeneity and increase the likelihood of crystallization.

Lose the Carboxy
If you cannot get diffraction quality crystals of your protein, try both Carboypeptidase-A and Carboxypeptidase-B to trim the carboxy terminus of the protein in order to generate smaller protein fragments which might be more amicable to crystallization.

High Quality Choppers
Use sequence grade Trypsin, Chymotrypsin, SV8 Protease and subtilisin along with inhibitors to generate small, active fragments for crystal screening. (Aled Edwards, McMaster University, CANADA & Brian McKeever, Merck, USA)

Seeing Double
Perform screens in duplicate at both 4°C and room temperature. If you notice a difference in solubility between the two temperatures, be sure to include temperature as a variable during optimization.

Screening Temperature with Limited Sample
Not enough sample to perform screens in duplicate at two different temperatures? Set up screens at room temperature. After four weeks move the plates to 4°C and watch for changes in solubility. Notice a difference? Use temperature as a variable during optimization.

Screening with Volatiles
Remember when using volatile organics as the precipitating agent that no reservoir needs to be added to the drop since distillation and equilibration proceed from the reservoir to the drop.

No Crystals?
No crystals, only precipitate in initial screens? Bring precipitant/protein concentration to just below supersaturation and vary pH and temperature to manipulate sample solubility.

Low Ionic Strength
Sample soluble at low ionic strength but forms a precipitate when dialyzed against water? This "salting in" effect can be used as a method to grow crystals. Begin with the sample in low ionic strength and dialyze the sample against a solution of lower ionic strength. Or, try changing the pH or temperature with the sample at a constant ionic strength. Charge distribution, surface features, or conformation may change as a function of these variables which in turn may influence sample solubility and the ability to crystallize the sample.

Sequence Homology Search
Uncertain about the ligand requirements for your sample? If you have the amino acid sequence for your protein, consider a sequence homology search to see if the sample is structurally related to other proteins with known ligand binding. Then try screening crystallization conditions including the ligands identified through the homology search.

Screening Continued...

Going from Nano Drops to Micro Drops

If you are having trouble reproducing crystallization initially found using nano drops (100 nl plus 100 nl for example) and are having trouble growing the crystals in the same conditions at µl volumes (1 ul plus 1 ul for example), one might consider the following tips.

Heather Ringrose (Pfizer UK) has suggested that if you initially screen with 100 nl (reservoir) + 200 nl (protein) you will pick up conditions that scale up to 1 + 1 µl.

Patrick Shaw-Stewart (Douglas Instruments, UK) reported that data-mining of published high-throughput results suggests that increasing the salt can help when scaling up from nanodrops. Presumably this speeds up equilibration. Patrick suggests that if you have a hit where the main precipitant is salt, try increasing the salt concentration in the reservoir solution by 1.0 M or more. Also, If you have a hit where the main precipitant is PEG or another organic, try increasing reservoir salt concentration by, say, 200 mM.

Using Old Screens and Reagents

Should one use an old crystallization screen kit that is past its "Best If Used By" date, or make screen and optimization reagents with old stock solutions?

The main issue would be the ability to reproduce any crystallization results achieved with the old reagents. With time, even at 4 and -20°C, the pH, ionic strength, concentration and stability of the reagents will change. Reagents containing Polyethylene glycol undergo oxidation which produces increased levels of aldehydes and peroxides which in turn lowers the pH and increases the ionic strength and produces small molecule species which can affect crystallization. Evaporation from the reagents, even when frozen, increases the concentration of solutes and decreases the concentrations of solvents. The result is that one might experience different results with an old reagent compared to a new reagent. An old reagent might not produce crystals as effectively as a new reagent. Or, the old reagent might produce crystals not able to be grown with the new reagent. In such cases it can be difficult or impossible to reproduce or optimize a crystal produced with an old reagent, using freshly prepared reagents. However, in these situations, some have reported success using the old reagent as an additive into freshly prepared reagents. For example, mix 1 part old reagent with 9 parts new reagent into the reagent well, mix and then add this reagent to the sample drop. This procedure will incorporate some level of pH, ionic strength and small molecule change from the old reagent into the new reagent.

The crystallization problems associated with old reagents versus new reagents seem sample-dependent. For some samples, a change in pH, ionic strength or the presence of a small molecule contaminant produced with age may have no affect on the crystallization. For some samples, the change might be advantageous and for still others, it might be other detrimental.

Biasing Crystallization Screen Reagents Towards Success

The following is a crystallization screening and optimization technique, which can be used in an attempt to improve the quality and size of a crystal or produce a different crystal form. First, one must begin with an initial crystallization reagent (ICR) which has produced some form of a crystal, microcrystalline precipitate or similar promising result. The ICR will be used to bias a subsequent crystallization screening experiment by adding the ICR to the crystallization drop, along with the sample and the screen reagent. Pipette the crystallization screen into the crystallization plate reservoir. One may choose to either repeat the original screen or choose a different set of screen reagents. When creating the drop, mix the protein, screen reagent and ICR in a 3:2:1 ratio. 3 parts sample : 2 parts screen reagent : 1 part ICR. The inclusion of the ICR in the drop modifies the crystallization screen, biasing the screen reagents with a promising reagent mix which may increase the likelihood for crystallization, improved crystals, or a different crystal form.

Possible modifications to this technique include: (1) varying the ratio of sample : screen reagent : ICR; and (2) modifying the bias such that one uses the ICR as the screen reagent and uses the original or new screen as the bias (3 parts protein : 2 parts ICR : 1 part screen reagent). This is essentially using the crystallization screen as an additive screen.

References

The role of bias in crystallization conditions in automated microseeding. F. J. St John, B. Feng and E. Pozharski. Acta Cryst. (2008). D64, 1222-1227.

Enhancing protein crystallization through precipitant synergy. Majeed S, Ofek G, Belachew A, Huang CC, Zhou T, Kwong PD. Structure. 2003 Sep;11(9):1061-70.

© 1991-2009 Hampton Research Corp. All rights reserved. Printed in the United States of America.
This guide or parts thereof may not be reproduced in any form without the written permission of the publishers.

CRYSTALLIZATION BUFFERS TABLE

IN ORDER OF PH RANGES

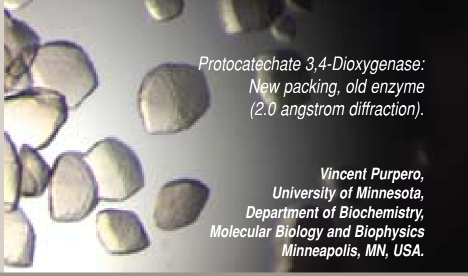

Protocatechate 3,4-Dioxygenase:
New packing, old enzyme
(2.0 angstrom diffraction).

Vincent Purpero,
University of Minnesota,
Department of Biochemistry,
Molecular Biology and Biophysics
Minneapolis, MN, USA.

DESCRIPTION

Tabled below are buffers typically used in creating buffered crystallization reagents. Buffers should be used in the indicated ranges.

BUFFER	pH	
Potassium chloride	1.0	2.2
Citric acid pKa(25°C) (1) 3.13, (2) 4.76, (3) 6.4	2.2	6.5
Sodium citrate tribasic dihydrate pKa(25°C) (1) 3.1, (2) 4.8, (3) 5.4 Citric acid trisodium salt dihydrate	3.0	6.2
Sodium acetate trihydrate pKa(25°C) 4.8 Acetic acid sodium salt	3.6	5.6
Sodium cacodylate trihydrate pKa(25°C) 6.2 Cacodylic acid sodium salt trihydrate	5.0	7.4
MES monohydrate pKa(25°C) 6.15 2-(N-Morpholino)ethanesulfonic acid	5.2	7.1
ADA pKa(25°C) 6.6 N-(2-Acetamido)iminodiacetic acid	5.6	7.5
BIS-TRIS pKa(25°C) 6.5 2,2-Bis(hydroxymethyl)-2,2',2''-nitrilotriethanol	5.8	7.2
ACES pKa(25°C) 6.8 N-(2-Acetamido)-2-aminoethanesulfonic acid	6.1	7.5
PIPES pKa(25°C) 6.8 1,4-Piperazinediethanesulfonic acid	6.1	7.5
Imidazole pKa(25°C) 6.95 1,3-Diaza-2,4-cyclopentadiene, Glyoxaline	6.2	7.8
MOPSO pKa(25°C) 6.9 β-Hydroxy-4-morpholinepropanesulfonic acid	6.2	7.6
BIS-TRIS propane pKa(25°C) (1) 6.8, (2) 9.0 1,3-Bis[tris(Hydroxymethyl)methylamino]propane	6.3	9.5
BES pKa(25°C) 7.1 N,N-Bis(2-hydroxyethyl)-2-aminoethanesulfonic acid	6.4	7.8
MOPS pKa(25°C) 7.2 3-(N-Morpholino)propanesulfonic acid	6.5	7.9
HEPES sodium pKa(25°C) 7.5 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid sodium salt	6.6	8.5
HEPES pKa(25°C) 7.5 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid	6.8	8.2

© 1991-2009 Hampton Research Corp. All rights reserved.

BUFFER	pH	
TES pKa(25°C) 7.5 N-[Tris(hydroxymethyl)methyl]-2-aminoethanesulfonic acid	6.8	8.2
MOBS pKa(25°C) 7.6 4-(N-Morpholino)butanesulfonic acid	6.9	8.3
DIPSO pKa(25°C) 7.6 3-(N,N-Bis[2-hydroxyethyl]amino)-2-hydroxypropanesulfonic acid	7.0	8.2
Tris pKa(25°C) 8.1 Tris(hydroxymethyl)aminomethane	7.0	9.0
TRIS hydrochloride pKa(25°C) 8.1 Tris(hydroxymethyl)aminomethane hydrochloride	7.0	9.0
TAPSO pKa(25°C) 7.6 2-Hydroxy-3-[tris(hydroxymethyl)methylamino]-1-propanesulfonic acid	7.0	8.2
HEPPSO pKa(25°C) 7.5 N-(Hydroxyethyl)piperazine-N'-2-hydroxypropanesulfonic acid	7.1	8.5
POPSO pKa(25°C) 7.8 Piperazine-N,N'-bis(2-hydroxypropanesulfonic acid)	7.2	8.5
EPPS pKa(25°C) 8.0 4-(2-Hydroxyethyl)-1-piperazinepropanesulfonic acid; HEPPS	7.3	8.7
TEA pKa(25°C) 7.8 Triethanolamine	7.3	8.3
BICINE pKa(25°C) 8.3 N,N-Bis(2-hydroxyethyl)glycine	7.4	9.3
Tricine pKa(25°C) 8.1 N-[Tris(hydroxymethyl)methyl]glycine	7.4	8.8
HEPBS pKa(25°C) 8.3 N-(2-Hydroxyethyl)piperazine-N'-(4-butanesulfonic acid)	7.6	9.0
TAPS pKa(25°C) 8.4 N-[Tris(hydroxymethyl)methyl]-3-aminopropanesulfonic acid	7.7	9.1
TABS pKa(25°C) 8.9 N-tris(Hydroxymethyl)methyl-4-aminobutanesulfonic acid	8.2	9.6
AMPSO sodium salt pKa(25°C) 9.0 3-([1,1-Dimethyl-2-hydroxyethyl]amino)-2-hydroxypropanesulfonic acid	8.3	9.7
CHES pKa(25°C) 9.3 2-(Cyclohexylamino)ethanesulfonic acid	8.6	10.0
Glycine pKa(25°C) 2.35 Aminoacetic acid, Aminoethanoic acid, Glycocoll	8.6	10.6
CAPSO pKa(25°C) 9.6 3-(Cyclohexylamino)-2-hydroxy-1-propanesulfonic acid	8.9	10.3
CAPS pKa(25°C) 10.4 3-(Cyclohexylamino)-1-propanesulfonic acid	9.7	11.1
CABS pKa(25°C) 10.7 4-(Cyclohexylamino)-1-butanesulfonic acid	10.0	11.4
Potassium chloride	12.0	13.0

© 1991-2009 Hampton Research Corp. All rights reserved.

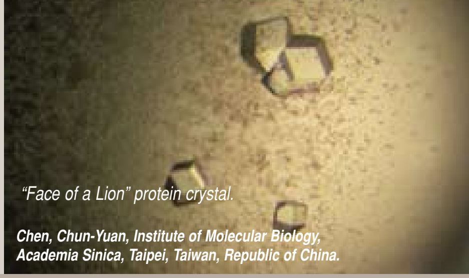

“Face of a Lion” protein crystal.

Chen, Chun-Yuan, Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan, Republic of China.

SOLUBILITY TABLE

CRYSTALLIZATION REAGENT SOLUBILITY

DESCRIPTION

The following table provides data for formulating a saturated solution of the reagents listed at the temperature designated.

Reagent	Formula	MW	Temp (°C)	g/100 ml	[M]
Ammonium bromide	NH_4Br	97.95	15	53.8	5.5
Ammonium chloride	NH_4Cl	53.49	20	26.7	5.0
Ammonium citrate dibasic	$(NH_4)_2C_6H_6O_7$	226.19	25	56.54	2.5
Ammonium fluoride	NH_4F	37.04	20	37.0	10.0
Ammonium formate	$HCOONH_4$	63.06	20	63.0	10.0
Ammonium iodide	NH_4I	144.94	25	94.2	6.5
Ammonium nitrate	NH_4NO_3	80.04	25	90.2	11.2
Ammonium phosphate dibasic	$(NH_4)_2HPO_4$	132.06	25	46.2	3.5
Ammonium phosphate monobasic	$NH_4H_2PO_4$	115.03	25	28.7	2.5
Ammonium sulfate	$(NH_4)_2SO_4$	132.14	20	46.2	3.5
Ammonium tartrate dibasic	$(NH_4)_2C_4H_4O_6$	184.15	20	36.8	2.0
Barium nitrate	$Ba(NO_3)_2$	261.34	25	10.2	0.3
Cadmium bromide tetrahydrate	$CdBr_2 \cdot 4H_2O$	344.28	25	94.0	2.7
Cadmium chloride hemipentahydrate	$CdCl_2 \cdot 2.5H_2O$	228.35	25	97.2	4.2
Cadmium iodide	CdI_2	366.22	20	73.0	1.9
Cadmium sulfate hydrate	$3CdSO_4 \cdot 8H_2O$	769.52	25	70.3	0.9
Calcium chloride hexahydrate	$CaCl_2 \cdot 6H_2O$	219.08	25	67.8	3.0
Calcium sulfate dihydrate	$CaSO_4 \cdot 2H_2O$	172.17	25	0.208	0.01
Cesium bromide	$CsBr$	212.81	22	89.8	4.2
Cesium chloride	$CsCl$	168.36	25	126.3	7.5
Cesium iodide	CsI	259.81	23	74.1	2.8
Cesium nitrate	$CsNO_3$	194.91	25	26.1	1.3
Cesium sulfate	Cs_2SO_4	361.87	25	129.8	3.5
Citric acid monohydrate	$HOC(COOH)(CH_2COOH)_2 \cdot H_2O$	210.14	25	88.6	4.2
Copper(II) bromide	$CuBr_2$	223.35	25	102.5	4.5
Copper(II) chloride dihydrate	$CuCl_2 \cdot 2H_2O$	170.48	25	80.0	4.6
Copper(II) sulfate pentahydrate	$CuSO_4 \cdot 5H_2O$	249.68	25	22.3	0.8
Iron(III) sulfate heptahydrate	$FeSO_4 \cdot 7H_2O$	278.01	25	52.8	1.8
Lithium acetate dihydrate	$CH_3COOLi \cdot 2H_2O$	102.02	20	51.0	5.0

© 1991-2009 Hampton Research Corp. All rights reserved.

Reagent	Formula	MW	Temp (°C)	g/100 ml	[M]
Lithium chloride	$LiCl$	42.39	20	42.4	10.0
Lithium citrate tribasic tetrahydrate	$HOC(COOLi)(CH_2COOLi)_2 \cdot 4H_2O$	281.99	20	42.3	1.5
Lithium fluoride	LiF	25.94	18	0.27	0.1
Lithium nitrate	$LiNO_3$	68.95	20	55.2	8.0
Lithium sulfate monohydrate	$Li_2SO_4 \cdot H_2O$	127.96	20	25.6	2.0
Magnesium bromide hexahydrate	$MgBr_2 \cdot 6H_2O$	292.20	18	83.1	2.8
Magnesium chloride hexahydrate	$MgCl_2 \cdot 6H_2O$	203.30	20	40.7	5.0
Magnesium formate dihydrate	$C_2H_2O_4MG \cdot 2H_2O$	150.38	25	150.04	1.0
Magnesium nitrate hexahydrate	$Mg(NO_3)_2 \cdot 6H_2O$	256.41	20	76.9	3.0
Nickel(II) chloride hexahydrate	$NiCl_2 \cdot 6H_2O$	237.69	20	95.1	4.0
Potassium acetate	CH_3COOK	98.14	20	49.1	5.0
Potassium bromide	KBr	119.00	25	56.0	4.7
Potassium chloride	KCl	74.55	25	22.4	3.0
Potassium citrate tribasic monohydrate	$HOC(COOK)(CH_2COOK)_2 \cdot H_2O$	324.42	20	81.2	2.5
Potassium fluoride	KF	58.10	20	34.9	6.0
Potassium formate	$HCOOK$	84.12	20	117.7	14.0
Potassium iodide	KI	166.00	25	103.2	6.2
Potassium nitrate	KNO_3	101.10	25	30.3	3.0
Potassium phosphate dibasic	K_2HPO_4	174.18	22	52.3	4.0
Potassium phosphate monobasic	KH_2PO_4	136.09	22	20.4	1.5
Potassium sodium tartrate tetrahydrate	$KOCOCH(OH)CH(OH)COONa \cdot 4H_2O$	282.22	20	42.3	1.5
Potassium sulfate	K_2SO_4	174.27	22	8.7	0.5
Potassium thiocyanate	$KSCN$	97.18	22	77.7	8.0
Sodium acetate trihydrate	$CH_3COONa \cdot 3H_2O$	136.08	22	40.8	3.0
Sodium chloride	$NaCl$	58.44	22	29.2	5.0
Sodium citrate tribasic dihydrate	$HOC(COONa)(CH_2COONa)_2 \cdot 2H_2O$	294.10	22	47.1	1.6
Sodium fluoride	NaF	41.99	22	3.4	0.8
Sodium formate	$HCOONa$	68.01	22	47.6	7.0
Sodium iodide	NaI	149.89	25	124.3	8.2
Sodium nitrate	$NaNO_3$	84.99	22	59.5	7.0
Sodium phosphate dibasic dihydrate	$Na_2HPO_4 \cdot 2H_2O$	177.99	22	8.9	0.5
Sodium phosphate monobasic monohydrate	$NaH_2PO_4 \cdot H_2O$	137.99	22	69.0	5.0
Sodium sulfate decahydrate	$Na_2SO_4 \cdot 10H_2O$	322.20	22	32.2	1.0
Sodium tartrate dibasic dihydrate	$C_4H_4Na_2O_6 \cdot 2H_2O$	230.08	22	34.5	1.5
Sodium thiocyanate	$NaSCN$	81.07	22	64.8	8.0
Zinc acetate dihydrate	$Zn(CH_3COO)_2 \cdot 2H_2O$	219.50	22	22.0	1.0
Zinc sulfate heptahydrate	$ZnSO_4 \cdot 7H_2O$	287.56	22	57.5	2.0

© 1991-2009 Hampton Research Corp. All rights reserved.

INDEX

Jack straws.

Ulrike Demmer, Max-Planck-Institute for Biophysics, Molecular Membrane Biology, Frankfurt, Germany

INDEX

A

- ADA....50
- Additive Screen....28
- Additive Screen HT....28
- Adjustable Crystal Mount....145, 152
- Adjustable Mounted CryoLoops....125
- Agarose....57, 104
- Al's Oil....56
- ALS Style CrystalCap....121
- AlumaSeal II Sealing Film....101
- Ammonium acetate....43
- Ammonium chloride....43
- Ammonium citrate dibasic....43
- Ammonium fluoride....43
- Ammonium formate....43
- Ammonium nitrate....43
- Ammonium phosphate dibasic....43
- Ammonium phosphate monobasic....43
- Ammonium sulfate....44
- Ammonium tartrate dibasic....44
- Applicators....103

B

- Batch Crystallization....79, 87-89
- Beeswax....143
- BICINE....50
- BIS-TRIS....50
- BIS-TRIS propane....50
- Books....166-173
- Brass Pin....145, 152
- Brass Pin with Platform....145, 152
- Buffer Kits....63-67
- Buffer Reagents....50-52

C

- Cadmium chloride hydrate....44
- Cadmium sulfate hydrate....44
- Calcium acetate hydrate....44
- Calcium chloride dihydrate....44
- Canned Air - Duster....95
- Cap Mat for MASTERBLOCK Plate....101
- Capillaries....138-140
- Capillary Cutting Stone....140
- Capillary Supplies....138-145
- Capillary Support....138-145
- Capillary Wax & Clay....143-144
- Cesium chloride....44
- Chelating Resin....115
- Circle Cover Slides....91-93
- Citric acid....50
- Citric acid BIS-TRIS propane....50
- Clamp....131
- Clay....144
- ClearSeal Film....99
- ClearSeal Film Applicator....99
- Cobalt(II) chloride hexahydrate....44
- Containerless Crystallization....56
- Corning Crystal*EX* 96 Well Plates....81-82
- Corning Crystal*EX* 384 Well Plates....90
- Cover Slides....91-95
- Cover Slide Dispensers....92
- Cover Slide Holder....92, 94
- Cover Slide Vacuum Gadget....94
- Cryo Labels....161,163
- CryoCane Color Coders....133
- CryoCanes....133
- Cryocrystallography....118-135
- CryoLoops....124-127
- CryoPro Kit....68
- Cryoprotectants....54, 68
- CryoSleeves....133
- CryoTongs....128-129
- Cryschem 24-1 SBS Plate....72
- Cryschem Plate....72
- Crystal Clear Sealing Film....99
- Crystal Clear Sealing Tape....100
- Crystal Dye - Izit....58
- Crystal Pencil....113
- Crystal Probe....113
- Crystal Screen 2 Kit....8-9
- Crystal Screen Cryo Kit....22-23
- Crystal Screen HT....8-9
- Crystal Screen Kit....8-9
- Crystal Screen Lite Kit....18-19
- CrystalBridge....76
- CrystalCap Colored....123
- CrystalCap Copper....123
- CrystalCap Copper Magnetic ALS HT....119
- CrystalCap Copper Magnetic HT....118
- CrystalCap Ex....122
- CrystalCap Holder....132
- CrystalCap HT Systems....118-120
- CrystalCap Magnetic....121
- CrystalCap Magnetic Systems....121-122
- CrystalCap Systems....122-123
- CrystalClear Strips....80
- CrystalDrop Lids....83
- Crystal*EX* 96 Well Plates....81-82
- Crystal*EX* 384 Well Plates....90
- Crystallization Plates & Accessories....72-104
- CrystalQuick Plates....83
- CrystalWands....130
- CrystalWand Magnetic....130
- CTAB....44
- Custom Shop....36-37
- Cutting Stone....140

D

- Deep Well MASTERBLOCK....86
- Detergent Screen....29
- Detergent Screen HT....29
- Dewar Flasks....134-135
- Dialysis Button Applicators....103
- Dialysis Buttons....102
- Dialysis Membrane Discs....103
- 1,4-Dioxane....42
- DL-Malic acid....46
- DMS Oil....56
- Douglas Instruments Plates....80, 89
- Dow Corning 7 Release Compound....98
- Dow Corning Vacuum Grease....98
- Duco Cement....142
- Dye....58

E

- Epoxy....126-127, 142
- Ethylene glycol....42
- Ethylene imine polymer....40

F

- Fluorinert....55
- FMS Oil....56
- Foam Dewars....135
- Forceps....110-113

G

- Gels....57
- Gizmo - Cover Slide Dispenser....92
- Glass Capillaries....138-139
- Glass Cover Slides....91-93
- Glass Fibers....140
- Glass Number 50 Capillaries....138-139
- Glass Sitting Drop Rods....95
- Glucose Isomerase....176
- Glue....126-127, 142
- Glycerol....42
- Goniometer Heads & Supplies....148-153
- Granada Crystallization Box....104
- Grease....98
- Grease Cartridges....98
- Greiner 24 Well ComboPlate....76
- Greiner 96 Well Imp@ct Plates....88
- Greiner CrystalBridge....76
- Greiner CrystalDrop Lid....83
- Greiner CrystalQuick Plates....83
- Greiner Microbatch....79
- Greiner Plate Lids....88
- Grid Screens....14-15
  - Grid Screen Ammonium Sulfate
  - Grid Screen MPD
  - Grid Screen PEG 6000
  - Grid Screen PEG/LiCl
  - Grid Screen Salt HT
  - Grid Screen Sodium Chloride
  - Grid Screen Sodium Malonate

H

- Heavy Atom Screens....30-32
- HEPES....50
- HEPES sodium....50
- 1,6-Hexanediol....42
- Hydrochloric acid....52
- Hydrogel....57

I

- I3C Phasing Kit....32
- Imidazole....51
- Immersion Oil....97
- Imp@ct Plates....88
- Index HT....6-7
- Index Kit....6-7
- Individual Screen Reagents....36-37
- Intelli-Plates....73, 78, 84
  - Intelli-Plate 24-4 well....73
  - Intelli-Plate 48-2 well....78
  - Intelli-Plate 48-3 well....78
  - Intelli-Plate 96 Flat Shelf....84
  - Intelli-Plate 96-2 LVR....84
  - Intelli-Plate 96-2 Original....84
  - Intelli-Plate 96-3 well....84
- Iron(III) chloride hexahydrate....44
- Izit - Crystal Dye....58

J

- Jeffamine....40

L

- L-Proline....47
- Labels & Pens....160-163
- Linbro Plate....76
- Lipase B....179
- Lithium acetate dihydrate....45
- Lithium chloride....45
- Lithium citrate tribasic tetrahydrate....45
- Lithium nitrate....45
- Lithium sulfate monohydrate....45
- LM Agarose....57, 104
- Long CryoTongs....129
- Long CrystalWands....130
- Low Form Dewar....134
- Low Ionic Strength Screen Kit....25
- Lysozyme....177
- Lysozyme Crystallization Reagent....177

M

- Magnesium acetate tetrahydrate....45
- Magnesium chloride hexahydrate....45
- Magnesium formate dihydrate....45
- Magnesium nitrate hexahydrate....45
- Magnesium sulfate heptahydrate....45
- Magnesium sulfate hydrate....46
- Magnetic Bases....153
- Malic acid....46
- MASTERBLOCK Deep Well Plate....86
- MembFac HT....18-19
- MembFac Kit....18-19
- MES monohydrate....51
- Micro-Bridges....96
- Micro-Bridges Polypropylene....96
- Micro-Tools....108-109
- Microbatch Plates....79, 87-89
- Microdialysis....102-103
- Microdialysis Buttons....102
- Microdialysis Membrane Discs....103
- MicroTubes....126
- MicroWick....141
- Modular VDX Plates....77
- Modular VDX Plates with sealant....77
- Mounted CryoLoops....124-125
- Mounting Clay - Four Color....144
- MPD....42
- MRC Plates....85-87

MRC Plate Seal 100
Multichannel Pipetter Basin 90

N

Natrix Kit 20-21
NDSB 52
Nickel(II) chloride hexahydrate 46
Nucleic Acid Mini Screen Kit 24

O

Oils 55-56
OptiClear Cover Slides 93
Optimize Reagents 40-58
Organic (Non-Volatile) Reagents 42
Organic (Volatile) Reagents 42

P

Paper Wicks 141
Paraffin Oil. 56
Paratone-N. 54
PCT Kit 5
PEG 40-42
PEG/Ion 2 Screen Kit 12-13
PEG/Ion Screen Kit 12-13
PEG/Ion HT 12-13
PEGRx 1 Kit 10-11
PEGRx 2 Kit 10-11
PEGRx HT 10-11
PEN-VAC Kit 94
Pencil 113
Pens 160
Pentaerythritol 40
Perfluoropolyether 54
pH Kits 63-67
Plain Cover Slides 93
Plastic Cover Slides 93
Plate Lids 83, 88
Plate Stand 94
Plates & Accessories 72-104
  9 Well Plate 97
  24 Well Plates 72-77
  48 Well Plates 78-79
  72 Well Plates 79
  96 Well Plates 80-90
  384 Well Plate 90
Poly(acrylic acid sodium salt) 5,100 40
Polyethylene glycol 40-42
Polymer Reagents. 40-42
Polypropylene glycol P 400 42
Polyvinylpyrrolidone K 15 42
Potassium acetate 46
Potassium bromide 46
Potassium chloride 46
Potassium citrate tribasic monohydrate 46
Potassium fluoride 46
Potassium formate 46
Potassium phosphate dibasic 46, 48
Potassium phosphate monobasic 47
Potassium nitrate 47
Potassium sodium tartrate tetrahydrate 47
Potassium sulfate 47
Potassium thiocyanate 47
Pre-Crystallization Test 5
Proline 47
2-Propanol 42
Protein Standards 176-179

Q

Quartz Capillaries 138-139
Quik Optimize Kit 48
Quik Screen Kit 14-15

R

Reagents 40-58
Red Sticky Wax 144
Reducing Agent 53
Robolid 101
Round Capillaries 104

S

Salt Kit 62
Salt Reagents 43-49
SaltRx 1 Kit 16-17
SaltRx 2 Kit 16-17
SaltRx HT 16-17
Sandwich Box Setup 97
Screens 5-32
Sealants 97-98
Sealing Film Applicator 99
Sealing Films 99-101
Sealing Tapes 100
Seed Bead Kit 114
Seeding 114-115
Seeding Tool 115
Set Screws 151
Silica Hydrogel 57
Silicon Oil 56
Siliconized Cover Slides 91
Silver Bullets Bio HT 26-27
Silver Bullets Bio Kit 26-27
Silver Bullets HT 26-27
Silver Bullets Kit 26-27
Sitting Drop Rods 95
Slide Dispensers 92
Slides 91-95
Sodium acetate trihydrate 47, 51
Sodium bromide 47
Sodium cacodylate trihydrate 51
Sodium chloride 47
Sodium citrate tribasic dihydrate 48, 51
Sodium fluoride 48
Sodium formate 48
Sodium hydroxide 52
Sodium malonate 48
Sodium nitrate 48
Sodium phosphate dibasic dihydrate 48
Sodium phosphate monobasic monohyd. 48
Sodium potassium phosphate 48
Sodium sulfate decahydrate 48
Sodium tartrate dibasic dihydrate 49
Sodium thiocyanate 49
Solubilizing Agents 52
Solvent Resistant Pens 160
Special Glass 10 Capillaries 138-139
Square Cover Slides 91-93
Stereo Viewer 173
Stereopticon 173
StockOptions Buffer Kits 64-67
  StockOptions Bis-Tris Kit 66
  StockOptions Citric Acid 64
  StockOptions HEPES Kit 66
  StockOptions MES Kit 65
  StockOptions Sodium Acetate Kit 64
  StockOptions Sodium Cacodylate Kit 65
  StockOptions Sodium Citrate Kit 65
  StockOptions Sodium HEPES Kit 66
  StockOptions Tris Kit 67
  StockOptions Tris Hydrochloride Kit 67
StockOptions CryoPro Kit 68
StockOptions Kits 62-68
StockOptions pH Kit 63
StockOptions Salt Kit 62
Succinic acid 49
Super Glue 126-127, 142

T

Tacsimate 49
Tapes 100
TCEP hydrochloride 53
Trimethylamine N-oxide dihydrate 49
Tris 51
TRIS hydrochloride 52
Tryptone 49
Tube Clamp 131

U

UVP Hanging Drop MRC Plate Seal 100

V

Vacuum Gadget 94
Vacuum Grease 98
Vapor Batch Plates 89
VDX48 Well Plate with sealant 79
VDX Plate 74
VDX Plate with sealant 74
VDX Plate - Modular 77
VDXm Plate 75
VDXm Plate with sealant 75
Vial Clamps 131
Viewers 173

W

Wax Pen 143
Waxes 143-144
Wicks 141

X

Xenon Chamber & Accessories 156
Xenon Derivatization 156-157
Xenon Recovery System 157
Xylanase 178
XYZ Heated Goniometer Heads 148-149
XYZ Standard Goniometer Heads 150-152

Z

Z Capillaries 152
Z Platforms 153
Zinc acetate dihydrate 49
Zinc chloride 49
Zinc sulfate heptahydrate 49

INDEX

HR1-

HR1-002 72

HR2-

HR2-078 26-27
HR2-080 26-27
HR2-082 10-11
HR2-084 10-11
HR2-086 10-11
HR2-088 26-27
HR2-090 26-27
HR2-092 26-27
HR2-094 26-27
HR2-096 26-27
HR2-098 12-13
HR2-100 67
HR2-102 66
HR2-104 64
HR2-106 66
HR2-107 16-17
HR2-109 16-17
HR2-110 8-9
HR2-112 8-9
HR2-114 18-19
HR2-116 20-21
HR2-118 24
HR2-120 25
HR2-122 22-23
HR2-126 12-13
HR2-128 18-19
HR2-130 8-9
HR2-132 68
HR2-134 6-7
HR2-136 16-17
HR2-137 18-19
HR2-138 28
HR2-139 12-13
HR2-140 5
HR2-142 5
HR2-144 6-7
HR2-211 14-15
HR2-213 14-15
HR2-215 14-15
HR2-217 14-15
HR2-219 14-15, 177
HR2-221 14-15
HR2-223 48
HR2-231 66
HR2-233 64, 177
HR2-235 65
HR2-237 67
HR2-239 65
HR2-241 63
HR2-243 65
HR2-245 62
HR2-247 14-15
HR2-248 14-15
HR2-310 57
HR2-312 115
HR2-320 114
HR2-406 29, 36
HR2-408 29
HR2-409 39
HR2-428 28, 36
HR2-442 30-31, 36
HR2-444 30-31, 36
HR2-446 30-31, 36
HR2-448 30-31, 36
HR2-450 30-31, 36
HR2-501 40
HR2-503 40
HR2-507 50
HR2-509 50
HR2-513 41
HR2-515 41
HR2-517 41
HR2-523 41
HR2-525 41
HR2-527 41
HR2-529 41
HR2-533 41
HR2-535 41
HR2-537 45
HR2-539 47
HR2-541 44
HR2-543 47
HR2-545 45
HR2-547 48
HR2-549 48
HR2-551 48
HR2-553 47
HR2-555 43
HR2-557 44
HR2-559 45
HR2-561 45
HR2-563 49
HR2-565 43
HR2-567 44
HR2-569 51
HR2-571 51
HR2-573 51
HR2-575 51
HR2-577 50
HR2-579 52
HR2-581 52
HR2-583 52
HR2-585 50
HR2-587 51
HR2-589 51
HR2-591 41
HR2-593 56
HR2-595 56
HR2-597 40
HR2-599 40
HR2-601 40
HR2-603 41
HR2-605 41
HR2-607 41
HR2-609 41
HR2-611 41
HR2-613 41
HR2-615 42
HR2-617 42
HR2-619 42
HR2-621 42
HR2-623 42
HR2-625 42
HR2-627 42
HR2-629 43
HR2-631 45
HR2-633 46
HR2-635 46, 48
HR2-637 47, 177
HR2-639 48
HR2-641 49
HR2-643 54
HR2-645 48
HR2-647 46
HR2-649 46
HR2-651 53
HR2-657 45
HR2-659 43
HR2-661 48
HR2-663 47
HR2-665 43
HR2-667 46
HR2-669 45
HR2-671 46
HR2-673 48
HR2-675 47
HR2-677 49
HR2-679 44
HR2-681 45
HR2-683 46
HR2-685 43
HR2-687 46
HR2-689 43
HR2-691 43
HR2-693 49
HR2-695 47
HR2-697 45
HR2-699 47
HR2-701 52
HR2-703 52
HR2-705 52
HR2-707 48
HR2-709 49
HR2-711 44
HR2-713 44
HR2-715 44
HR2-717 44
HR2-719 44
HR2-721 44
HR2-723 50
HR2-725 51
HR2-727 52
HR2-729 50
HR2-731 51, 177
HR2-733 50
HR2-735 51
HR2-737 51
HR2-739 40
HR2-741 40
HR2-743 40
HR2-745 40
HR2-747 48
HR2-749 48
HR2-751 48
HR2-755. 49
HR2-757 50
HR2-759 43
HR2-761 46
HR2-763 47
HR2-765 48
HR2-767 44
HR2-769 42
HR2-771 42
HR2-773 40
HR2-775 47
HR2-777 49
HR2-779 46
HR2-781 50
HR2-783 50
HR2-785 50
HR2-787 51
HR2-789 51
HR2-791 52
HR2-793 52
HR2-795 50
HR2-797 55
HR2-799 43
HR2-801 53
HR2-803 45
HR2-805 177

HR2-807 ... 48
HR2-809 ... 48
HR2-811 ... 49
HR2-813 ... 49
HR2-814 ... 54
HR2-817 ... 50
HR2-819 ... 51
HR2-821 ... 45
HR2-823 ... 49
HR2-825 ... 49
HR2-827 ... 49
HR2-829 ... 49
HR2-831 ... 50
HR2-833 ... 50
HR2-835 ... 49
HR2-837 ... 41
HR2-839 ... 49
HR2-841 ... 41
HR2-843 ... 49
HR2-845 ... 41
HR2-847 ... 49
HR2-849 ... 41
HR2-851 ... 49
HR2-900-** ... 37, 67
HR2-902-** ... 37, 66
HR2-904-** ... 37, 64
HR2-906-** ... 37, 66
HR2-906-24 ... 50
HR2-907-** ... 16-17, 36
HR2-909-** ... 16-17, 36
HR2-910-** ... 8-9, 36
HR2-912-** ... 8-9, 36
HR2-914-** ... 22-23, 36
HR2-916-** ... 18-19, 36
HR2-918-** ... 20-21, 36
HR2-920-** ... 18-19, 36
HR2-921-** ... 14-15, 36
HR2-922-** ... 12-13, 36
HR2-924-** ... 14-15, 36
HR2-926-** ... 14-15, 36
HR2-928-** ... 14-15, 36
HR2-930-** ... 14-15, 36
HR2-931-** ... 37, 66
HR2-931-01 ... 50
HR2-932-** ... 14-15, 36
HR2-933-** ... 37, 64
HR2-935-** ... 37, 65
HR2-937-** ... 37, 67
HR2-939-** ... 37, 65
HR2-940-** ... 5, 36
HR2-941-** ... 37, 63
HR2-943-** ... 37, 65
HR2-943-07 ... 51
HR2-944-** ... 6-7, 36
HR2-947-** ... 14-15, 36
HR2-982-** ... 10-11, 36
HR2-984-** ... 10-11, 36
HR2-988-** ... 26-27, 36
HR2-993-** ... 37
HR2-995-** ... 37
HR2-996-** ... 26-27, 36
HR2-997-** ... 37
HR2-998-** ... 12-13, 36
HR2-999-** ... 37

HR3-

HR3-082 ... 85
HR3-083 ... 85
HR3-084 ... 88
HR3-085 ... 88
HR3-086 ... 79
HR3-087 ... 79
HR3-088 ... 83
HR3-089 ... 83
HR3-090 ... 83
HR3-091 ... 83
HR3-092G ... 83
HR3-093G ... 83
HR3-094G ... 83
HR3-095G ... 83
HR3-096 ... 83
HR3-097 ... 83
HR3-098 ... 88
HR3-099 ... 88
HR3-100 ... 88
HR3-101 ... 88
HR3-102 ... 87
HR3-103 ... 87
HR3-104 ... 104
HR3-105 ... 86
HR3-108 ... 75
HR3-110 ... 76
HR3-111 ... 101
HR3-112 ... 76
HR3-113 ... 81
HR3-114 ... 73
HR3-115 ... 81
HR3-120 ... 79
HR3-122 ... 79
HR3-123 ... 86
HR3-125 ... 86
HR3-128 ... 80
HR3-132 ... 97
HR3-133 ... 32
HR3-134 ... 97
HR3-136 ... 97
HR3-137 ... 83
HR3-138 ... 83
HR3-140 ... 74
HR3-142 ... 74
HR3-143 ... 84
HR3-145 ... 84
HR3-146 ... 95
HR3-148 ... 95
HR3-150 ... 76
HR3-154 ... 76
HR3-158 ... 72
HR3-160 ... 72
HR3-162 ... 80
HR3-170 ... 74
HR3-172 ... 74
HR3-183 ... 84
HR3-185 ... 84
HR3-190 ... 83
HR3-192 ... 83
HR3-194 ... 104
HR3-196 ... 77
HR3-198 ... 77
HR3-200 ... 76
HR3-202 ... 98
HR3-204 ... 77
HR3-207T ... 93
HR3-209T ... 93
HR3-211 ... 93
HR3-213 ... 93
HR3-215 ... 91
HR3-217 ... 91
HR3-219 ... 93
HR3-221 ... 93
HR3-223 ... 91
HR3-225 ... 91
HR3-227 ... 93
HR3-229 ... 93
HR3-231 ... 91
HR3-233 ... 91
HR3-235 ... 93
HR3-237 ... 93
HR3-239 ... 91
HR3-241 ... 91
HR3-243 ... 93
HR3-245 ... 93
HR3-247 ... 91
HR3-249 ... 91
HR3-251 ... 94
HR3-265 ... 89
HR3-267 ... 89
HR3-269 ... 90
HR3-271 ... 81
HR3-273 ... 81
HR3-275 ... 79
HR3-277 ... 91
HR3-278T ... 91
HR3-279 ... 91
HR3-280T ... 91
HR3-281 ... 83
HR3-283 ... 83
HR3-284 ... 83
HR3-285 ... 83
HR3-293 ... 88
HR3-295 ... 88
HR3-297 ... 84
HR3-299 ... 84
HR3-302 ... 83
HR3-304 ... 83
HR3-306 ... 75
HR3-310 ... 96
HR3-312 ... 96
HR3-314 ... 102
HR3-316 ... 102
HR3-318. ... 102
HR3-320 ... 102
HR3-322 ... 102
HR3-326 ... 102
HR3-328 ... 102
HR3-330 ... 102
HR3-332 ... 102
HR3-336 ... 102
HR3-338 ... 103
HR3-340 ... 96
HR3-342 ... 96
HR3-344 ... 103
HR3-346 ... 103
HR3-411 ... 56
HR3-413 ... 56
HR3-415 ... 56
HR3-417 ... 56
HR3-421 ... 56
HR3-423 ... 56
HR3-508 ... 98
HR3-510 ... 98
HR3-511 ... 100
HR3-515 ... 91
HR3-517 ... 91
HR3-607 ... 100
HR3-609 ... 99
HR3-611 ... 97
HR3-613 ... 97
HR3-615 ... 97
HR3-617 ... 97

HR4-

HR4-110 ... 141
HR4-112 ... 141
HR4-114 ... 141

HR4-CONTINUED

HR4-116 141
HR4-122 141
HR4-211 141
HR4-213 141
HR4-217 113
HR4-310 144
HR4-312 143
HR4-313 124
HR4-315 130
HR4-316 126-127, 142
HR4-318 126-127, 142
HR4-320 142
HR4-326 144
HR4-328 143
HR4-330 141
HR4-334 140
HR4-336 125
HR4-338 125
HR4-342 143
HR4-344 143
HR4-346 126-127, 142
HR4-348 103
HR4-350 103
HR4-400 163
HR4-402 163
HR4-404 161
HR4-406 161
HR4-411 95
HR4-436 162
HR4-438 162
HR4-440 162
HR4-442 162
HR4-444 162
HR4-446 162
HR4-456 163
HR4-458 160
HR4-460 161
HR4-462 160
HR4-464 160
HR4-466 160
HR4-468 94
HR4-506 100
HR4-511 100
HR4-513 173
HR4-515 173
HR4-517 173
HR4-521 99
HR4-523 99
HR4-525 99, 101
HR4-600 130
HR4-602 130
HR4-615 124
HR4-617 124
HR4-619 130
HR4-623 126
HR4-625 124
HR4-627 153
HR4-629 153
HR4-631 128
HR4-633 128
HR4-635 128
HR4-637 120, 128
HR4-639 128
HR4-641 128
HR4-643 150
HR4-647 150
HR4-651 153
HR4-653 153
HR4-657 149
HR4-659 151
HR4-661 145, 152
HR4-663 145, 152
HR4-665 123
HR4-667 128
HR4-669 149
HR4-670 131
HR4-671 131
HR4-672 131
HR4-673 135
HR4-675 135
HR4-677 104
HR4-678 104
HR4-679 104
HR4-680 104
HR4-681 104
HR4-682 104
HR4-683 104
HR4-684 104
HR4-685 104
HR4-686 104
HR4-687 104
HR4-688 104
HR4-695 134
HR4-697 134
HR4-699 134
HR4-701 134
HR4-705 132
HR4-706 132
HR4-707 132
HR4-708 133
HR4-709 133
HR4-710 58
HR4-711 133
HR4-713 133
HR4-715 133
HR4-717 133
HR4-719 133
HR4-721 133
HR4-722 133
HR4-727 131
HR4-729 130
HR4-731 121
HR4-733 121
HR4-737 122
HR4-739 122
HR4-741 122
HR4-743 122
HR4-745 122
HR4-747 122
HR4-749 122
HR4-753 151
HR4-755 149
HR4-761 153
HR4-763 153
HR4-765 148
HR4-767 148
HR4-769 152
HR4-771 152
HR4-775 149
HR4-777 149
HR4-779 121
HR4-781 156
HR4-791 156
HR4-793 156
HR4-795 156
HR4-797 157
HR4-799 156
HR4-811 108
HR4-813 108
HR4-815 108
HR4-817 108
HR4-819 108
HR4-821 108
HR4-823 108
HR4-825 108
HR4-827 108
HR4-829 108, 109
HR4-831 108, 109
HR4-835 113
HR4-837 109
HR4-839 109
HR4-841 109
HR4-843 109
HR4-845 109
HR4-847 109
HR4-849 109
HR4-851 109
HR4-853 109
HR4-855 110
HR4-857 110
HR4-859 110
HR4-861 110
HR4-863 111
HR4-865 111
HR4-869 111
HR4-871 112
HR4-875 112
HR4-879 112
HR4-881 112
HR4-883 113
HR4-900 122
HR4-902 121
HR4-904 118-119, 121-122
HR4-911 122
HR4-913 122
HR4-914 122
HR4-915 126
HR4-917 126
HR4-919 126
HR4-921 126
HR4-923 126
HR4-925 126
HR4-928 126
HR4-929 126
HR4-931 126
HR4-933 126
HR4-935 126
HR4-937 126
HR4-939 126
HR4-941 126
HR4-943 153
HR4-951 130
HR4-953 124
HR4-955 124
HR4-957 124
HR4-959 124
HR4-961 124
HR4-963 124
HR4-965 124
HR4-969 123
HR4-971 123
HR4-973 123
HR4-975 123
HR4-977 123
HR4-979 123
HR4-981 127
HR4-983 127
HR4-985 127
HR4-987 127
HR4-989 127
HR4-991 127
HR4-993 124
HR4-995 124
HR4-997 124
HR4-999 124

HR5-

HR5-102 ... 134
HR5-104 ... 129
HR5-106 ... 129
HR5-108 ... 129
HR5-110 ... 129
HR5-112 ... 120, 129
HR5-113 ... 129
HR5-114 ... 129
HR5-118 ... 129
HR5-120 ... 129
HR5-210 ... 168
HR5-211 ... 167
HR5-212 ... 168
HR5-213 ... 167
HR5-214 ... 172
HR5-215 ... 166
HR5-219 ... 170
HR5-221 ... 172
HR5-223 ... 168
HR5-224 ... 166
HR5-230 ... 167
HR5-231 ... 173
HR5-233 ... 169
HR5-237 ... 170
HR5-239 ... 172
HR5-241 ... 170
HR5-900 ... 125
HR5-902 ... 125

HR6-

HR6-100 ... 151
HR6-102 ... 151
HR6-104 ... 139
HR6-106 ... 139
HR6-108 ... 139
HR6-110 ... 139
HR6-112 ... 139
HR6-114 ... 139
HR6-116 ... 139
HR6-118 ... 139
HR6-120 ... 139
HR6-122 ... 139
HR6-124 ... 139
HR6-126 ... 139
HR6-128 ... 139
HR6-130 ... 139
HR6-132 ... 139
HR6-134 ... 139
HR6-136 ... 139
HR6-138 ... 139
HR6-140 ... 139
HR6-142 ... 139
HR6-144 ... 139
HR6-146 ... 139
HR6-148 ... 139
HR6-150 ... 139
HR6-151 ... 139
HR6-152 ... 139
HR6-154 ... 139
HR6-156 ... 139
HR6-158 ... 139
HR6-160 ... 139
HR6-162 ... 139
HR6-164 ... 139
HR6-166 ... 139
HR6-168 ... 139
HR6-170 ... 139
HR6-172 ... 139
HR6-174 ... 139
HR6-175 ... 139
HR6-177 ... 139
HR6-179 ... 139

HR7-

HR7-099 ... 179
HR7-100 ... 176
HR7-102 ... 176
HR7-104 ... 178
HR7-106 ... 178
HR7-108 ... 177
HR7-110 ... 177

HR8-

HR8-002 ... 123
HR8-004 ... 123
HR8-006 ... 123
HR8-008 ... 123
HR8-010 ... 123
HR8-012 ... 123
HR8-014 ... 123
HR8-028 ... 145, 152
HR8-030 ... 140
HR8-032 ... 140
HR8-056 ... 90
HR8-058 ... 90
HR8-061 ... 171
HR8-062 ... 169
HR8-069 ... 101
HR8-072 ... 125
HR8-074 ... 93
HR8-076 ... 93
HR8-078 ... 93
HR8-080 ... 93
HR8-082 ... 93
HR8-084 ... 93
HR8-086 ... 169
HR8-088 ... 91
HR8-090 ... 91
HR8-092 ... 57, 104
HR8-094 ... 120
HR8-098 ... 94
HR8-100 ... 171
HR8-102 ... 125
HR8-104 ... 125
HR8-106 ... 125
HR8-108 ... 125
HR8-110 ... 125
HR8-112 ... 120
HR8-114 ... 120
HR8-116 ... 120
HR8-118 ... 120
HR8-120 ... 120
HR8-122 ... 120
HR8-124 ... 120
HR8-126 ... 120
HR8-128 ... 120
HR8-130 ... 120
HR8-132 ... 120
HR8-133 ... 115
HR8-134 ... 82
HR8-135 ... 82
HR8-136 ... 82
HR8-137 ... 82
HR8-138 ... 82
HR8-139 ... 82
HR8-140 ... 82
HR8-141 ... 82
HR8-146 ... 82
HR8-147 ... 82
HR8-148 ... 83
HR8-149 ... 83
HR8-150 ... 78
HR8-152 ... 78
HR8-154 ... 78
HR8-156 ... 78
HR8-162 ... 92
HR8-163 ... 92
HR8-164 ... 92
HR8-165 ... 92
HR8-166 ... 92
HR8-167 ... 92
HR8-168 ... 92
HR8-169 ... 92
HR8-170 ... 92
HR8-171 ... 84
HR8-172 ... 84
HR8-173 ... 118
HR8-174 ... 118
HR8-175 ... 118
HR8-176 ... 118
HR8-177 ... 118
HR8-178 ... 118
HR8-179 ... 118
HR8-180 ... 118
HR8-181 ... 118
HR8-182 ... 119
HR8-184 ... 119
HR8-186 ... 119
HR8-188 ... 119
HR8-190 ... 119
HR8-192 ... 119
HR8-194 ... 119
HR8-196 ... 119
HR8-198 ... 119

DI-

DI-038 ... 89
DI-039 ... 89
DI-040 ... 89
DI-041 ... 89
DI-043 ... 80

GENERAL INFORMATION

ORDERS

By phone:

Telephone	(949) 425-1321
Toll Free in the USA	(800) 452-3899
Toll Free in Canada	(800) 452-3899

By Fax:

Fax	(949) 425-1611

By Mail:

Hampton Research Corp.
34 Journey
Aliso Viejo, CA 92656-3317 USA

By E-mail: info@hrmail.com

On the Internet: www.hamptonresearch.com

Telephone Orders

Hampton Research is available by telephone between 7:00 am and 5:00 pm Pacific Standard Time, Monday through Friday.

Fax Orders

Fax ordering is available 24 hours a day, 7 days a week. Fax orders are confirmed within 24 hours of receipt.

When placing an order please provide the following:

- Purchase order number
- Account name
- Shipping address
- Billing address
- Name of person to whose attention the order should be shipped
- Name and telephone number of contact person
- Product catalog number, description, size, and quantity

Order by Credit Card

Hampton Research accepts American Express, Visa and MasterCard orders via telephone, fax, and a secure internet shopping cart in USA dollar amounts only. Please provide credit card number, expiration date, card name, and billing address.

Standing Orders

Hampton Research can establish periodic shipments or large orders as standing orders. Please contact Customer Service.

Use of Products

Hampton Research products are for laboratory research use only and are not to be used for any other purpose including in vitro diagnostics, foods, drugs, medical devices, medical procedures, cosmetics, for/with/on humans, animals or plants, or for commercial purpose.

Pricing

Pricing is in USA dollars. Pricing is subject to change. Please contact Hampton Research or reference our website for current and correct pricing if you require this information before placing an order. Prices are confirmed at time of order.

Damaged Products

If you receive damaged products, notify Customer Service within 24 hours of receipt by telephone, fax, or e-mail to obtain instructions regarding claims.

Returns

Products may not be returned without prior authorization by Customer Service. A Return Authorization form must be completed and included with each return for proper handling. Please inspect packages upon receipt to be sure you ordered and received the items you need and expect. Notify Hampton Research of any discrepancies within 24 hours of receipt.

Shipping

All shipments are FOB Aliso Viejo, California USA. All orders received by 3:00 pm Pacific Standard Time will be shipped the same day unless the Shipping Department is out on a mountain bike ride. Orders requiring manufacture will be processed and shipping will be confirmed at the time of order. Shipping and handling charges are prepaid and added to the invoice.

Holidays

Hampton Research observes various USA holidays and random days off as well. Should your order arrive on a holiday, it will be processed and shipped the following work day.

Resale

Hampton Research products are not for resale except for distributors contracted by us.

Payment

Payment terms are net 30 days from shipment date. Payment is due in USA dollars. All bank drafts must be paid in USA dollars and drawn from a USA bank.

Suggestions and Comments

Hampton Research welcomes all suggestions, comments, and complaints pertaining to our products, service, and just about anything related to crystallization. Hampton Research is serious about delivering the best crystallization research tools and service to the crystallization research community. We appreciate you telling us what needs to be changed, improved, created or scrapped. Please direct ideas, comments and rantings to Bob Cudney. Telephone: (949) 425-1321 Extension 200. E-mail: bob@hrmail.com.

MSDS

Material Safety Data Sheets are available for products which are hazardous. Each Hampton Research product is intended for use by a qualified professional, trained in good laboratory practices and for those familiar with the potential hazards associated with laboratory research products.

Warranty and Liability

Hampton Research warrants that products conform to descriptions contained in the product specific literature and technical data sheets. Hampton Research is not liable for any claims or applications not listed in the Hampton Research literature. Hampton Research is not liable for the misuse of products, nor should information disclosed in the product literature be construed as a recommendation to use Hampton Research products in violation of any patents or applicable laws. Hampton Research's liability shall not exceed the purchase price of the product. Hampton Research shall have no liability for direct, indirect, consequential, or incidental damages arising from the use, results of use, or inability to use its products.

Technical Service

Technical Service is here to help with the most current and accurate information related to Hampton Research crystallization research tools and just about any topic related to protein crystallization. We welcome inquiries regarding biological macromolecular crystallization from product inquiries, screening and optimization strategies, or general brainstorming. We're here to help. There are numerous internet-based, technical resources at www.hamptonresearch.com or you can call, e-mail, or fax us. Telephone: (949) 425-1321 option 2. E-mail: tech@hrmail.com Fax: (949) 425-1611

CrystalNews E-mail Newsletter

Sign up for CrystalNews, an e-mail newsletter containing up-to-date information about biological macromolecular crystallization and crystallography. Tips, crystallographer related reading, meetings, community job opportunities, product announcements, and more. Send your e-mail to crystalnews@hrmail.com or visit www.hamptonresearch.com.

Thank You

Thank you for using Hampton Research crystallization research tools. Thank you to our customers, friends, families, fellow employees, and vendors who allow us to make a living doing what we enjoy (growing crystals). Best of success with your crystals!

TRADEMARKS

The following trademarks or registered trademarks are the property of the companies listed:

Anatrace - Anapoe, C-HEGA, CYMAL, FOS-Choline, HECAMEG, MEGA
Art Robbins Instruments - Intelli-Plate
BASF Corporation - Pluronic
Destin Arzneimittel GmBH - Thesit
Devcon - Duco Cement
Diversified Biotech - Tough-Spots, Tough-Tags
Douglas Instruments - CrystalClear Strips
Dow Chemical Company -TRITON
Dow Corning Corporation - 7 Release Compound, High Vacuum Grease
EMD Chemicals - ZWITTERGENT
EXCEL Scientific - AlumaSeal II
Greiner Bio-One - ComboPlate, CrystalBridge, CrystalDrop, CrystalQuick, MASTERBLOCK
Huntsman Petrochemical Corporation - Jeffamine
Johnson & Johnson - ThermoFluor
MP Biomedicals - Linbro
Nalgene - CryoCane, CryoSleeve
Sigma-Aldrich - Trizma
Spectrum - Spectra/Por
Texaco - M-600
Triana Science & Technology - Granada Crystallization Box
Virtual Industries - PEN-VAC

The following trademarks or registered trademarks are the property of Hampton Research Corporation:

Additive Screen, Additive Screen HT, Beeswax Stick, Brass Specimen Pin, Capillary Cutting Stone, Capillary Wax, Chelating Resin, ClearSeal Film, Cover Slide Vacuum Gadget, CryoLoops, CryoPro, CryoTong, Cryschem, Crystal Probe, Crystal Screen, Crystal Screen 2, Crystal Screen Cryo, Crystal Screen HT, Crystal Screen Lite, CrystalCap, CrystalCap HT, CrystalWand, Custom Shop, Detergent Screen, Detergent Screen HT, Dialysis Buttons, Dialysis Membrane Discs, Glass Sitting Drop Rods, Goniometer Head, Grid Screen, Grid Screen Salt HT, Heavy Atom, Index, Index HT, Izit Crystal Dye, Low Ionic Strength Screen, Magnetic Base, Mechanical Crystal Pencil, MembFac, Micro-Bridges, Micro-Tools, Micro-Tools II, MicroTube, MicroWick, Modular VDX, Mounted CryoLoops, Natrix, Nucleic Acid Mini Screen, Optimize, Paper Wicks, PCT, PEG/Ion 2 Screen, PEG/Ion Screen, PEG/Ion Screen HT, PEGRx, Quik Optimize, Quik Screen, SaltRx, SaltRx HT, Seed Bead, Silica Hydrogel, Silver Bullets, Silver Bullets Bio, Silver Bullets Bio HT, Silver Bullets HT, StockOptions, Tacsimate, Tube Clamp, VDX, VDX48, VDXm, Vial Clamp, Wax Pen, Xenon Chamber, Xenon Recovery System, Z Capillary, Z Platform

Crystal Radio

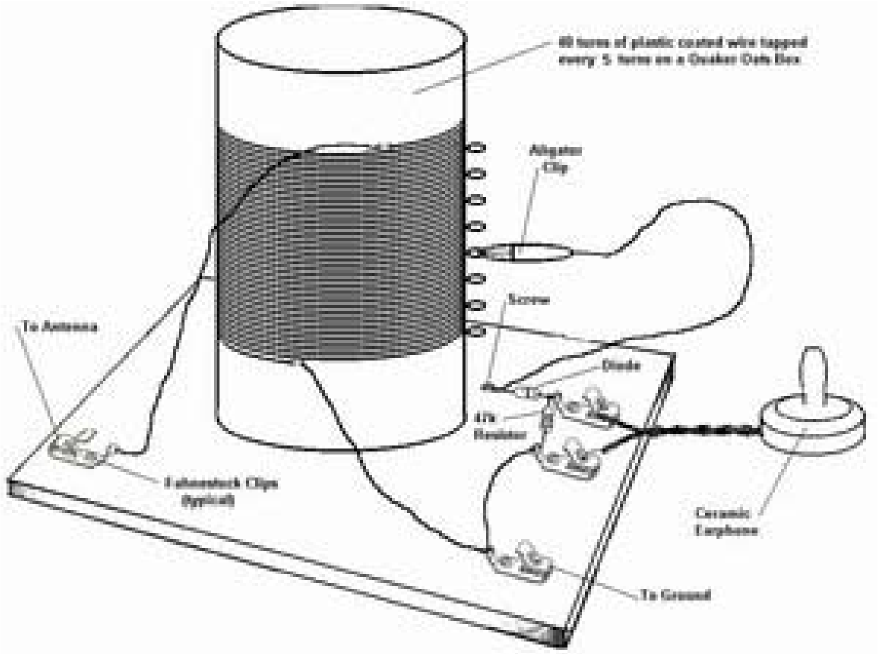
40 turns of plastic coated wire tapped every 5 turns on a Quaker Oats Box
Alligator Clip
Screw
To Antenna
Diode
47k Resistor
Fahnestock Clips (typical)
Ceramic Earphone
To Ground

Crystal Radio

Abstract

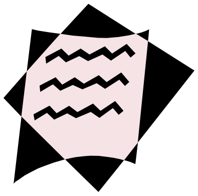

Students learn the principles of radio waves by building a crystal radio.

Equipment

1. Antenna
2. Copper Wire
3. Tank Circuit
4. Plastic Insulation
5. Detector
6. Germanium Semiconductor Crystal
7. Earphone
8. Piezoelectric Device

Grade Level

This activity is suitable for Elementary and Middle School Students.

State Standards Met

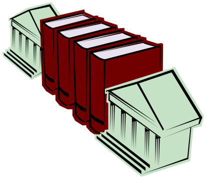

Standard 1 – Analysis, Inquiry, and Design
Standard 4 – Physical Setting and Living Environment

Build a Crystal Radio

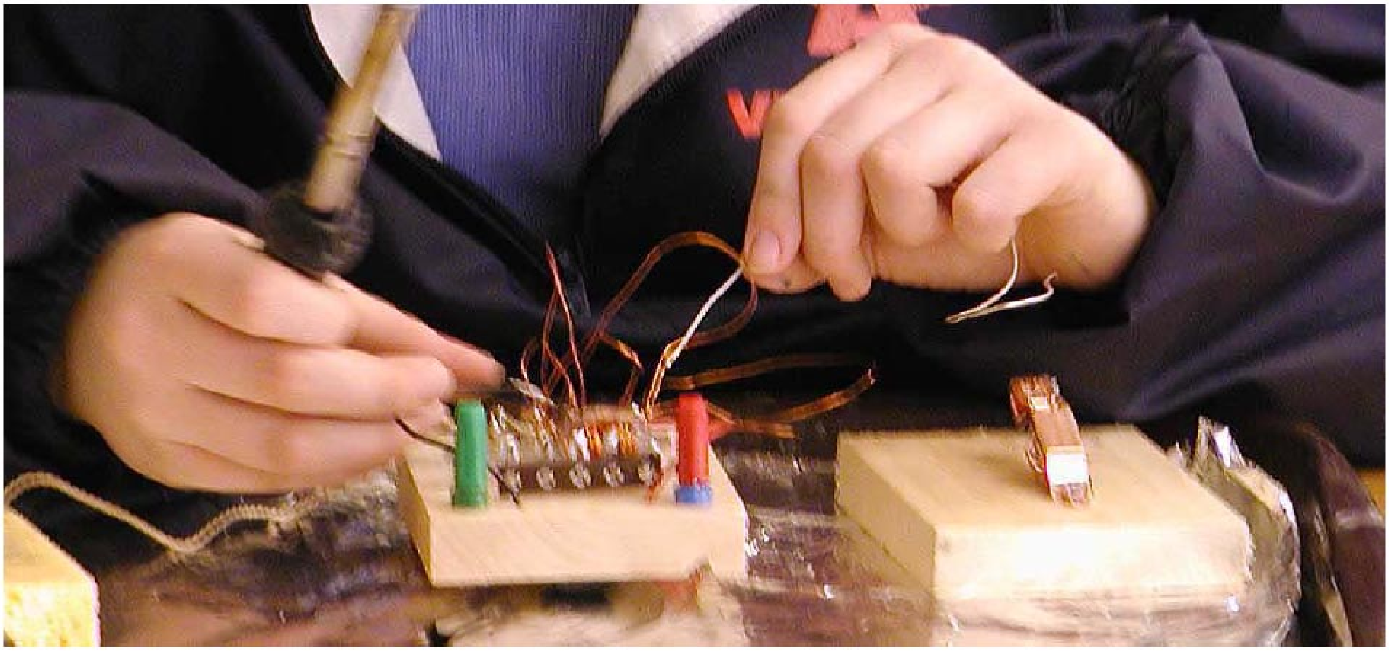

This program is supported by:
the Cornell Center for Materials Research,
and
the National Science Foundation

Support for the Cornell Center for Materials Research is provided through the National Science Foundation as part of the Materials Research Science and Engineering Centers (MRSEC) program. Additional support is provided by Cornell University.

Make Your Own Crystal Radio!

Here at the Cornell University Center for Materials Research Radio Factory you'll make your own radio wave powered crystal radio set. You'll learn a lot about how your radio works and the special materials that you'll use to make it go. You'll also gain some valuable electronics construction skills such as the ability to recognize components and join them together by soldering.

How Your Radio Will Work

Here is a chart that shows how signals flow through the four different sections of the radio and are changed along the way.
In this chart we say a bit about what each section does, give its official radio builders name, and introduce a special material that is needed for the section to work

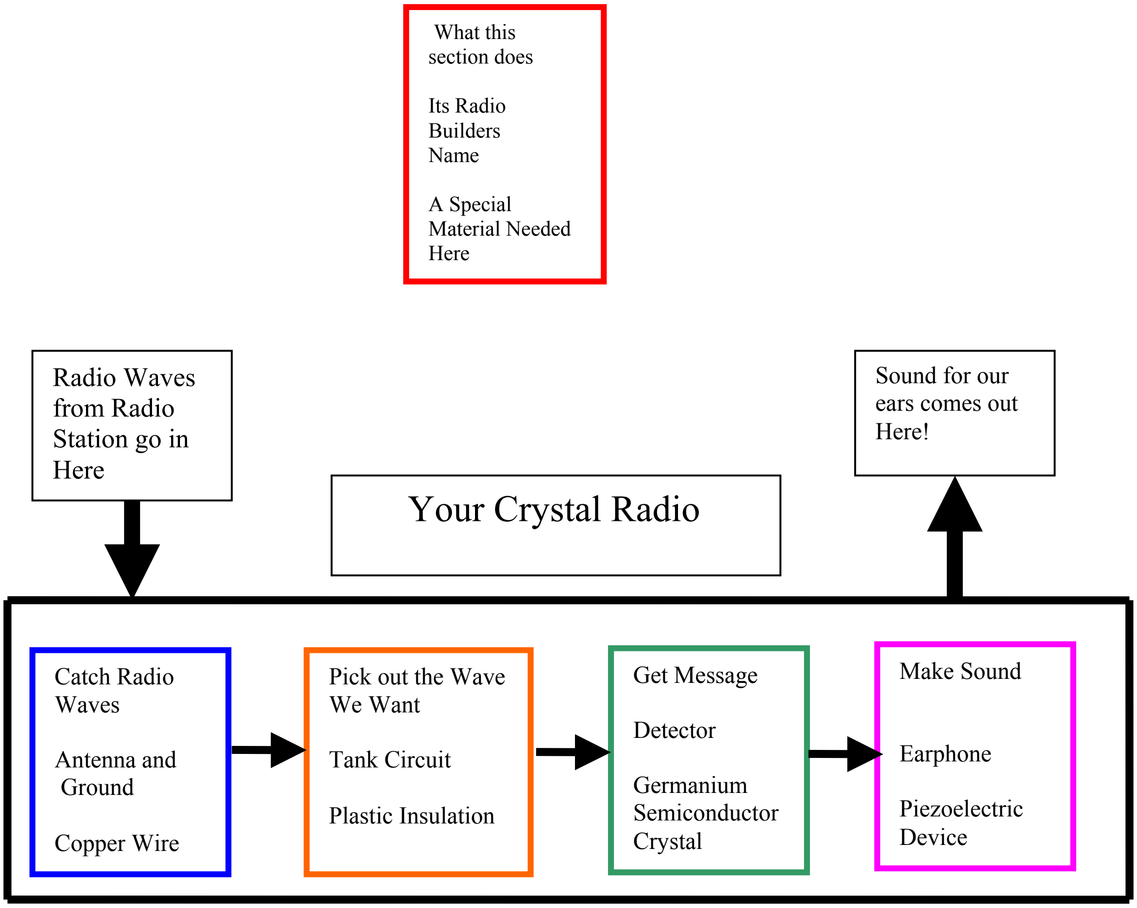
What this section does
Its Radio Builders Name
A Special Material Needed Here
Radio Waves from Radio Station go in Here
Your Crystal Radio
Sound for our ears comes out Here!
Catch Radio Waves
Antenna and Ground
Copper Wire
Pick out the Wave We Want
Tank Circuit
Plastic Insulation
Get Message
Detector
Germanium Semiconductor Crystal
Make Sound
Earphone
Piezoelectric Device

Support for Cornell Center for Materials Research is provided through NSF Grant DMR-0079992
Copyright 2003 CCMR Educational Programs. All rights reserved.

Here’s what the different sections in the radio do and the role the special materials play.

1) Catch Radio Waves:

The radio stations launch electric and magnetic forces through space that come and go a million times in a second. These electromagnetic vibrations are not always the same. In fact, the radio station changes their strength much more slowly, thousands of times per second to carry the messages in the form of words or music that they wish us to hear. You can send your voice a whole lot further by getting it to ride a radio wave than by simply yelling. For our radio wave catcher we can often use large pieces of metal that readily pass electricity. Things that pass electricity well are called conductors. We use electric forces in the radio waves to set electric current moving back and forth in time to the wave in the copper metal wires (more good conductors) that connect our radio at the antenna and ground connections. We need the material copper metal to bring these currents to the radio. Radio builders call the antenna the conductor that catches the wave in space. It can be the wire that comes with your set alone or a piece of metal such as an air duct. The ground is a connection to earth that allows the desired electric currents to pass from the antenna and through your radio. When running a home crystal set one can frequently use a cold water pipe, so long as it’s made out of a metal (often copper). The ``ground’’ doesn’t have to actually be connected to the earth. When we look for a good antenna and ground combination, what we are really seeking are two large conductors that are appropriately shaped and oriented (that means trying to see what works!) and are not already connected to each other. We can then hook up your radio between them (one side to the antenna connection, the other side to the ground connection) and catch the radio wave currents that pass back and forth between them. Please pay attention to avoid connecting your radio to power lines. I can’t warn you enough about how dangerous that is. There are lots more ideas about making good antenna and ground connections later in these notes.

To keep from losing our radio signal, we prevent wires we don’t want connected from touching with the colored plastic around them. Things that won’t allow electricity to pass through them, like the plastic, are called insulators. You’ll need many forms of insulators in your radio. See if you and your instructor can find some.

2) Pick out the Wave We Want:

Just as a swing swings best when we pump it at the right rate, the tank circuit composed of the coil and a capacitor will set an electric current sloshing back and forth if they are tuned for the right station, Around Ithaca, WTKO vibrates waves at 1,470,000 times a second and WHCU vibrates waves at 870,000 times a second. By adjusting the tank circuit to the right rate of electric vibration we not only select the station we want, but it is made stronger, just the reason we pump a swing at the right rate to make it swing well. To tune our tank circuit we change the size of the variable capacitor by adjusting the black tuning knob, just like in a regular radio. In the tank circuit the radio energy is swapped back and forth between the coil where it is stored as a magnetic field and the capacitor where it’s in the form on an electric field. The vibrations are managed in exactly this energy exchange process. The corresponding

Support for Cornell Center for Materials Research is provided through NSF Grant DMR-0079992

Copyright 2003 CCMR Educational Programs. All rights reserved.

process in a swing is to exchange energy of motion (greatest when you are swing through the low point) and energy in a gravity field (greatest when you are at the high points). When we want to store energy in a magnetic field, we use turns of metal wire to flow electricity in a loop. In the capacitor, we also use metal, but as plates that are close to each other, but not touching. We keep them from touching by using plastic insulation. That's a special material needed for the tank circuit. See if you can spot the metal/plastic sandwiches in the capacitor. Can you see what happened inside the capacitor when you turn the tuning knob?

3) Get the Message: To pull the message of music or speech off the vibrating current from the tank circuit is the neatest trick of all. This happens in the detector stage. This accomplished by our most special material: the germanium crystal diode. This is why our radio is called a crystal radio. Germanium, like so many materials that are at the heart of today's radio, TV's, electronics toys and computers, is a material that conducts electricity much more than an insulator, but not as well as a metal. Our crystal diode has a special property that is essential for pulling the message off the vibrating current: it passed current better one way than in the opposite direction. This allows the slow changes in the vibration that contains the electrical signals containing music or speech messages to be detected.

We can explain this in detail as follows: following the tuner stage, the radio uses a germanium diode to extract the audio signal from the much higher frequency radio wave that has carried it to this point. Since this signal is coded in the form of strength variations of the radio frequency wave, the detector acts in an analogous manner to the DC power supplies that have as their input the AC line voltage in a house. The principle it uses to extract the audio wave is to act as a one-way valve for electric current. Rather than confront the earphone of the last stage with an AC voltage that swings too fast by equal amounts toward positive and negative voltages (besides which, our ears would not detect such rapid variations anyway) the diode passes current in only one direction. Once smoothed by the earphone circuitry, this wave appears as immediately recognizable sounds. To repeat, the diode pulls the audio message off the radio wave that carried it great distances by passing its rapidly oscillating current in only one direction. The resultant signal is then smoothed in the last stage into a slowly varying wave. If we had admitted both directions of current flow, successive swings of the radio wave would have cancelled each other out during smoothing.

Can you think of other things at home that need to do tricky electrical transformations to do their job?

4) Make Sound: The last step is to turn the electric currents representing speech or music back into sound that we can here. Our earphone does this. It uses another special material that has the property of being piezoelectric: when we apply electricity to it, it changes its size a little bit. But that's enough to make the air around it move to form a sound wave that we can hear.

I'll bet you know other devices that we use all the time to make electrical currents into sound. What about the other way around?

Support for Cornell Center for Materials Research is provided through NSF Grant DMR-0079992

Copyright 2003 CCMR Educational Programs. All rights reserved.

One of the neatest features of your radio is that it gets all its power from the radio wave itself. It does all these jobs without being plugged into the wall or using a battery.

Here are a couple of views of one way to build your radio. The different sections are outlined in the second view to indicate by color the four main parts of the radio.

Support for Cornell Center for Materials Research is provided through NSF Grant DMR-0079992
Copyright 2003 CCMR Educational Programs. All rights reserved.

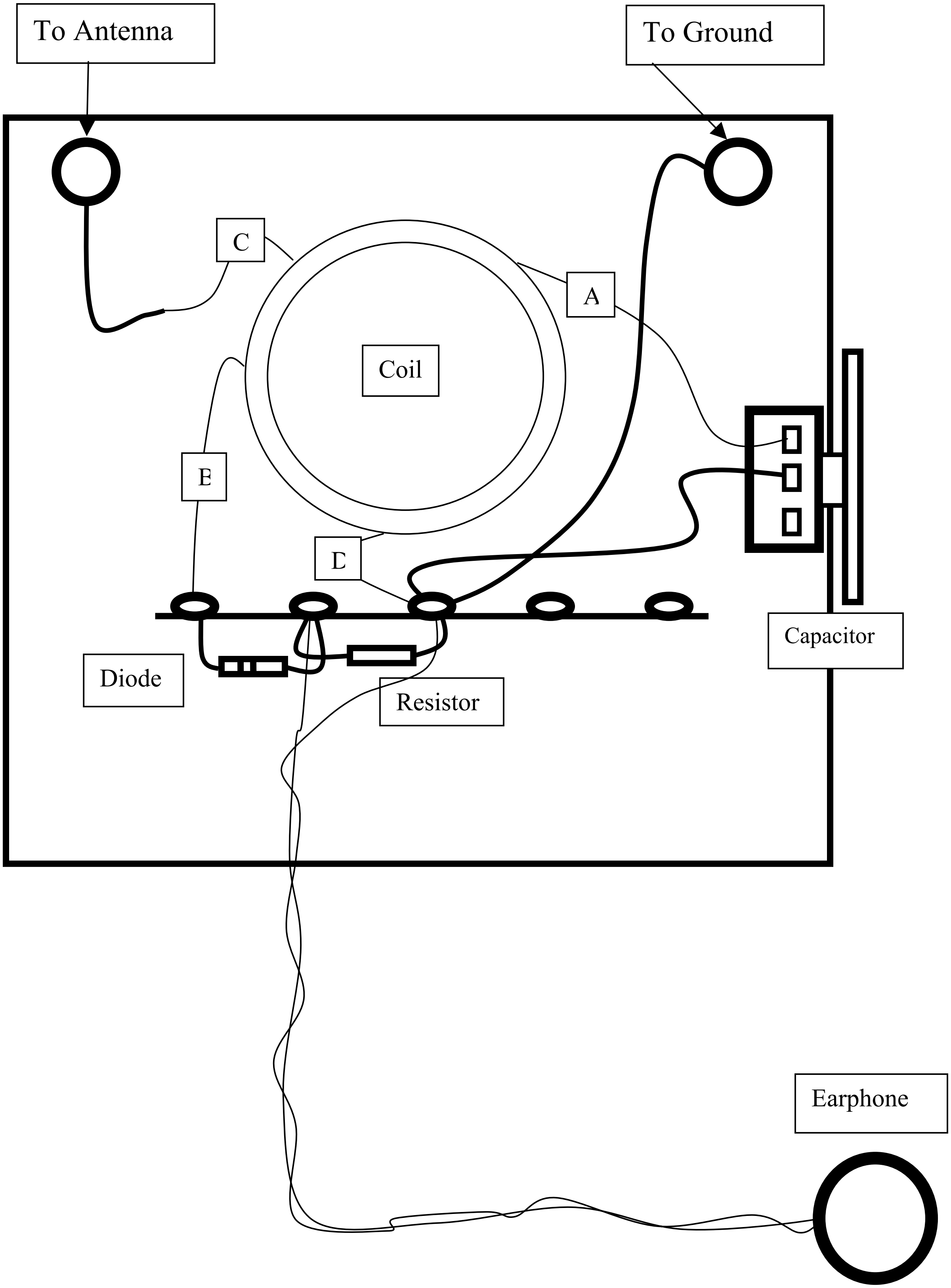
To Antenna
To Ground
C
A
Coil
E
D
Capacitor
Diode
Resistor
Earphone

Support for Cornell Center for Materials Research is provided through NSF Grant DMR-0079992
Copyright 2003 CCMR Educational Programs. All rights reserved.

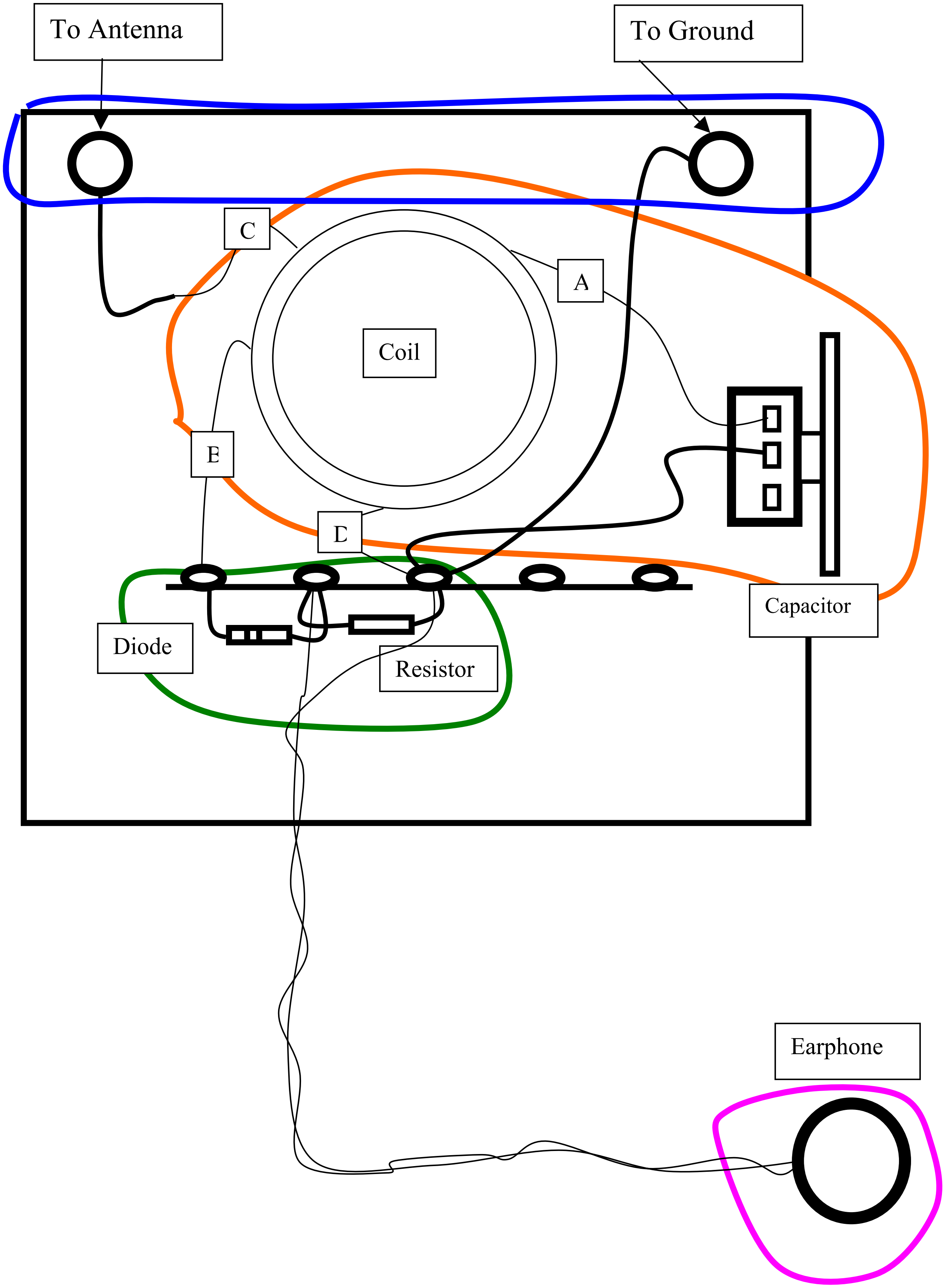
To Antenna
To Ground
C
A
Coil
E
D
Capacitor
Diode
Resistor
Earphone

Support for Cornell Center for Materials Research is provided through NSF Grant DMR-0079992
Copyright 2003 CCMR Educational Programs. All rights reserved.

Using your radio:

Important Notes for Radio Builders and Parents

Antennas and Grounds and Warnings!!!

You'll need to connect your antenna and ground to the appropriate terminals through the clips on wires that you made up. Please be very careful not connect to live electric wires (conduit casing is fine for a ground however). Most importantly don't throw antenna wires over any utility cables: it is really really really dangerous if you get on an electric power wire. Also, don't ground to gas lines. Outdoor antennas are fine as long as they are temporary and are not left connected to your radio or house in times of possible lightning strikes. If you use a permanent outdoor antenna, you will need a lightning arrestor. A licensed electrician knows how to set one up, but that could be costly.

A swell ground is provided by a metal (not plastic, that won't conduct electricity!) cold water pipe. We have also had good luck with the metal nut on the outside of the cable connector for the local cable television service (be careful not to connect to the center conductor). You can make an antenna by stringing up your antenna wire as far as it will stretch. Outdoors it helps to get it off the ground, at least by a few feet. An antenna may also be provided by large pieces of metal such as a metal doorframe or rain gutter (again, think about the possibility of lightning and don't leave your set connected when not in use). One of the best antennas we have found is an air duct. For an outdoor birthday party, we found that a garden hose faucet made a swell ground. For the antenna, we found that a fine gauge magnet wire (like the #32 we use to wind our coils) sent up about 50' into a treetop by big Mylar helium balloon worked really well! The trick is to be patient and willing to experiment (for example try aiming the antenna in different directions). You should try to bite through paint with your clips to make contact with bare metal. It is possible however to connect by simply wrapping wire without making bare metal. Generally, avoid overlapping the antenna and ground wires. It will be fun to discover what works for you!

Tuning Your Radio:

Once you have set up an antenna and ground, just try turning the capacitor's knob. Good luck and have fun!

Trouble Shooting Defective Earphones

We have found that sometimes a good strong signal will fade away for no apparent reason while you are listening. If this happens to you, try rapping on the earphone (don't smash it so that it breaks! Striking with you finger with the force of a finger snap is the trick). We have found that doing this can quickly bring this back to life (you are likely to have to retune in case you moved the slug while trying to restore the signal). If this

Support for Cornell Center for Materials Research is provided through NSF Grant DMR-0079992
Copyright 2003 CCMR Educational Programs. All rights reserved.

happens to you, please let us know and we'll give you a replacement earphone in exchange for the defective one. We are pretesting our earphones, but you may still get a poor performer.

What to Listen for and When to Listen

You may be surprised to discover more stations at night than you can hear during the day. This is because the high reaches of the atmosphere can sometimes reflect radio waves from far way. If you're persistent, you may become an expert on this effect. Also, you may even pull in stations that are higher in frequency than the usual broadcast band.

Support for Cornell Center for Materials Research is provided through NSF Grant DMR-0079992
Copyright 2003 CCMR Educational Programs. All rights reserved.

More Information and Ideas for Home Experimentation

The World Wide Web is an amazing resource for more information on crystal radios and the crystal detector materials themselves. You can simply search under www.google.com for ``crystal radio.'' Here are some of the excellent sites you will find:

freeweb.pdq.net/headstrong/crystal.htm
www.midnightscience.com
http://www.scitoys.com/scitoys/scitoys/radio/radio
www.thebest.net/wuggy/build.htm

The wuggy website gives a great series of crystal radios under the link: The Project Crystal Radio. Our design is based on the second circuit.
The scitoys website of Simon Field is also not to be missed: his projects are exceptional in their fun, value and ease of building. As you will see, there are many marvelous variations and extensions to our current design. First, there are a great variety of antenna and ground possibilities. Within the radio itself, large tank coils can bring better performance both in signal strength and in ability to pick out one station at a time. You might wish to add a battery-powered one-transistor amplifier for portability (you will still need to have an antenna wire, but no ground is necessary). The regenerative receiver is an amazing few-transistor extension to provide extremely high sensitivity. With this design, you can readily detect waves with higher vibration rates (higher frequencies) that provide worldwide communication.

In order to modify your coil, you can pry it out of the glue that holds it to the base. You might prefer however to preserve your working radio and make a new one to try out a new design.

To learn more about the materials that make your radio work, you may enjoy books from a public library. Any edition of the ARRL's Handbook for Radio Amateurs will give you very useful information about what's inside your radio from the insulator inside your capacitor to the solder you used to make your set...

We look forward very much to news of your crystal radio experiences at home as well as any questions or stories to tell that you might have (grandparents know some of the most interesting radio tales). Our e-mail addresses are cpf1@cornell.edu and outreach@ccmr.cornell.edu.

Sincerely, Carl Franck, Juliane Bauer-Hutchinson, Nev Singhota, Kevin Dilley for the Cornell Center for Materials Research

Support for Cornell Center for Materials Research is provided through NSF Grant DMR-0079992
Copyright 2003 CCMR Educational Programs. All rights reserved.

Radio-Builder's Diagram (Schematic)

Here's another way to understand how the various parts of your radio connect to each other. Radio builders call this a schematic. Each part has a special symbol. Wires are simply lines. We'll bet you can quickly find your fancy coil. Radio builders understand how the signals flow through your radio a lot more easily by looking at this diagram than the physical layout on the early pages. If you get a new idea to try out, you can explain to another radio builder easily by just drawing the new schematic. You might like to get some practice reading schematics by figuring out where our four sections of the radio are on this schematic. Good luck!

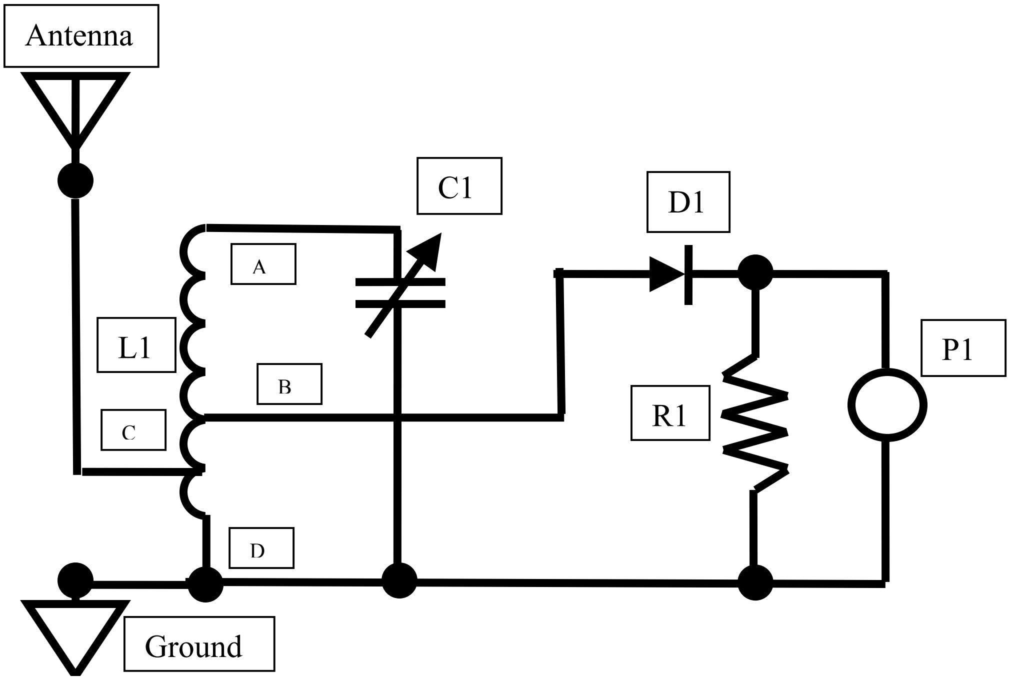

Antenna wire is 15' long medium sized clip on the end
Ground wire is 10' long with large clip on end

L1: Coil is 140 non-overlapping turns of #28 wire on 1.5" (inner diameter) thick walled PVC pipe (3" long) tapped at 20 and 40 turns from the ground side.

C1: 141(max) pf variable capacitor

D1: 1N34A diode

R1: 10K resistor

P1: piezoelectric earphone

Support for Cornell Center for Materials Research is provided through NSF Grant DMR-0079992
Copyright 2003 CCMR Educational Programs. All rights reserved.

Coil Winding Instructions

To wind the coil, tack the wire down with hot glue initially, wind five turns at a time side by side, taking up slack afterwards while compressing the coil after every set of five turns.

Starting from the bottom (wire connection D), wind 20 turns, bring out tap C, wind another 20 turns, bring out tap B, complete with 100 turns to end with connection A.

This version of the Make-A-Radio notes was made 7/28/03. C.F.

Support for Cornell Center for Materials Research is provided through NSF Grant DMR-0079992
Copyright 2003 CCMR Educational Programs. All rights reserved.

CRYSTAL RADIO INITIATIVE

by Eric P. Dollard

THE COLLECTED POSTS FROM ENERGETIC FORUM
OPEN SOURCE BUILDERS
THANK YOU TO DR_GREEN, JAKE AND OTHERS

CRYSTAL RADIO INITIATIVE

by Eric P. Dollard

NOTE: THIS BOOK WILL BE UNDER DEVELOPMENT IN THE FUTURE AND IS CURRENTLY A COMPILATION OF THE POSTS FROM VARIOUS ERIC DOLLARD THREADS AT THE OFFICIAL ERIC DOLLARD FORUM AT ENERGETIC FORUM

http://www.energeticforum.com/eric-dollard-official-forum/

TAKE NOTE THAT THESE POSTS ARE REORDERED ACCORDING TO ERIC DOLLARD'S PREFERENCE SO WILL BE DIFFERENT THAN HOW THEY APPEAR IN THE FORUM.

BETWEEN THESE POSTS AND THE CRYSTAL RADIO INITIATIVE VIDEO, THERE IS ENOUGH INFORMATION TO REPLICATE THE TESLA STYLE CRYSTAL RADIO.

DR_GREEN ALREADY POSTED RESULTS IN THE FORUM SHOWING THE GROUND TRANSMISSION IS INDEED FASTER THAN THROUGH THE AIR.

Cover Layout: Aaron Murakami
A&P Electronic Media
Spokane, Washington

Front Cover Image: Jerry Clement, http://commons.wikimedia.org/wiki/File:Crystal_radio_receiver.jpg

Interior Images by: Eric Dollard and the contributing members of Energetic Forum

Copyright © 2015 A & P Electronic Media

NOTE: ALTHOUGH THE CONTENT OF THE FORUM POSTS ARE IN ENERGETIC FORUM, THIS DOCUMENT HAS A PUBLISHER'S COPYRIGHT FOR THIS PARTICULAR COMPILATION AND MAY NOT BE FREELY DISTRIBUTED. IF YOU WANT TO SHARE THIS WITH OTHERS, REFER THEM TO HTTP://CRYSTALRADIOINITIATIVE.COM TO GET THEIR OWN COPY.

All Rights Reserved, Worldwide. No part of this publication may be reproduced, stored in an electronic retrieval system, or transmitted in any form, or by any means, without the prior, written permission of the copyright holders or the publisher. Unauthorized copying of this digital file is prohibited by International Law.

Digital Edition Published by: A&P Electronic Media

PO Box 10029

Spokane, WA 99209

http://www.emediapress.com/

First Edition – Version 1.01

Released: September, 2015
Digital Format: PDF File
50,000 authorized Downloads

Book Version

Current Version 1.01 – September, 2015

On page 20 of the original version, a full page image was covered up at the end of the page. This has been corrected so that on page 21 it is fully visible.

Original Version 1.00 - April, 2015

These are the forum posts that go together with Eric Dollard's Crystal Radio Initiative video. A concise manual will be created at a later time.

LIABILITY DISCLAIMER

By making use of any information in this book and video package, you agree to the following:

NO WARRANTIES: All of the information provided in this book and video package is provided "AS-IS" and with NO WARRANTIES. No express or implied warranties of any type are made with respect to the information, or any use of the information, on this site. A & P Electronic Media ("A & P") makes no representations and extends no warranties of any type as to the accuracy or completeness of any information or content on this website.

DISCLAIMER OF LIABILITY: A & P specifically DISCLAIMS LIABILITY FOR INCIDENTAL OR CONSEQUENTIAL DAMAGES and assumes no responsibility or liability for any loss or damage suffered by any person as a result of the use or misuse of any of the information or content in this book & video package. A & P or undertakes NO LIABILITY for any loss or damage suffered as a result of the use, misuse or reliance on the information and content of this book and video package.

USE AT YOUR OWN RISK: This book and video package is for informational purposes only. Consult a professional electrical engineer before performing any electrical experiments, including the experiments described in this book and video package. Any electrical experiment may results in injury or death. By voluntarily conducting any experiments described in this book and video package, you assume the risk of any resulting injury.

FOLLOW THE LAW: You must comply with FCC laws as they apply in your area and you accept any and all risks or liabilities if you should break these laws.

SUPPORT ERIC DOLLARD

Before doing anything else, make sure to show Eric Dollard your gratitude by donating through PayPal or through the mail. If there is a current fundraising campaign, please donate to show your support and tell others.

Information on how to donate to the work of Eric Dollard and EPD Laboratories, Inc. 501(c)(3) is on http://ericpdollard.com

TABLE OF CONTENTS

Introduction 8

Crystal Radio Initiative 9

Crystal Radio Magic Thread by Jake 137

Eric's "Crystal Radio Initiative" Challenge 274

INTRODUCTION

As previously mentioned above, this book is currently the reordered compilation of posts in Energetic Forum, which were posted in response to Eric Dollard's Crystal Radio Initiative.

Who will be the first to light a 100 watt bulb from AM ground transmission and prove Tesla right and Einstein wrong? Even if you can only power a 7 watt bulb, you are definitely in the club as someone who has certainly made a point.

Please watch the video first that accompanies this Crystal Radio Initiative package, especially if you're a beginner. Write down a few of the strongest stations that you receive with this and pick one of those frequencies to design your Tesla style crystal radio for.

For the Tesla style crystal radio, you'll have to calculate the coil dimensions so it will be specifically tuned for one of the strongest stations that you receive with the regular crystal radio set.

Once you do this for a small scale Tesla style crystal radio, modify the circuit and try to light one of the very small midget flange bulbs as described in the video.

When you can accomplish that, you are ready to scale it up to try to light even bigger bulbs. As Eric Dollard mentioned in the video, if you can just get to the 7 watt size, then you have definitely proven the point.

Furthermore, you can compare the speed between the above ground reception and the telluric reception to see which is fastest. DR_GREEN has posted this recently in the Eric Dollard thread here: http://www.energeticforum.com/eric-dollard-official-forum/11855-eric-dollard.html - The below ground reception is faster than the light speed above ground transmission by the way. Not only will you be able to prove Tesla right, you'll be proving Einstein wrong.

CRYSTAL RADIO INITIATIVE

01-02-2012, 06:38 PM

t-rex
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Crystal Sets Gone Wild

Crystal Sets Gone Wild

For the diagram shown the coil dimensions are missing, number of turns, etc. A good ground is essential for these kinds of devices. 16 Ground rods in a 10 to 20 foot radius circle, connected to a single ground rod at the center(17th rod), this connection being 10 gauge wire. Dry sand or rock will not ground, so this requires 80, each 14 gauge wires in a 30 foot diameter circle in a star radial configuration, to a center terminal. Without these groundings a Tesla Transformer cannot properly operate, but some "HI-Z" sets may.

The objective here is to scale the "Crystal Set", a step at a time, into a Tesla Transformer for the reception of medium wave band, 300 - 3000 kilocycle A.M. broadcasts. No license is required for this and the broadcast station provides the power.

And this objective cooperates with the primary objective. That is; Who will be the first ham to disprove Einstein's theory? An International contest, but who will sponsor it, Iran maybe?

We have the good fortune in the "Crystal Set Initiative" that, in theory at least, a quarter wave A.M. broadcast tower, and its 120 quarter wave ground radials, must emit a pair of waves as shown by Tesla in his basic diagrams.

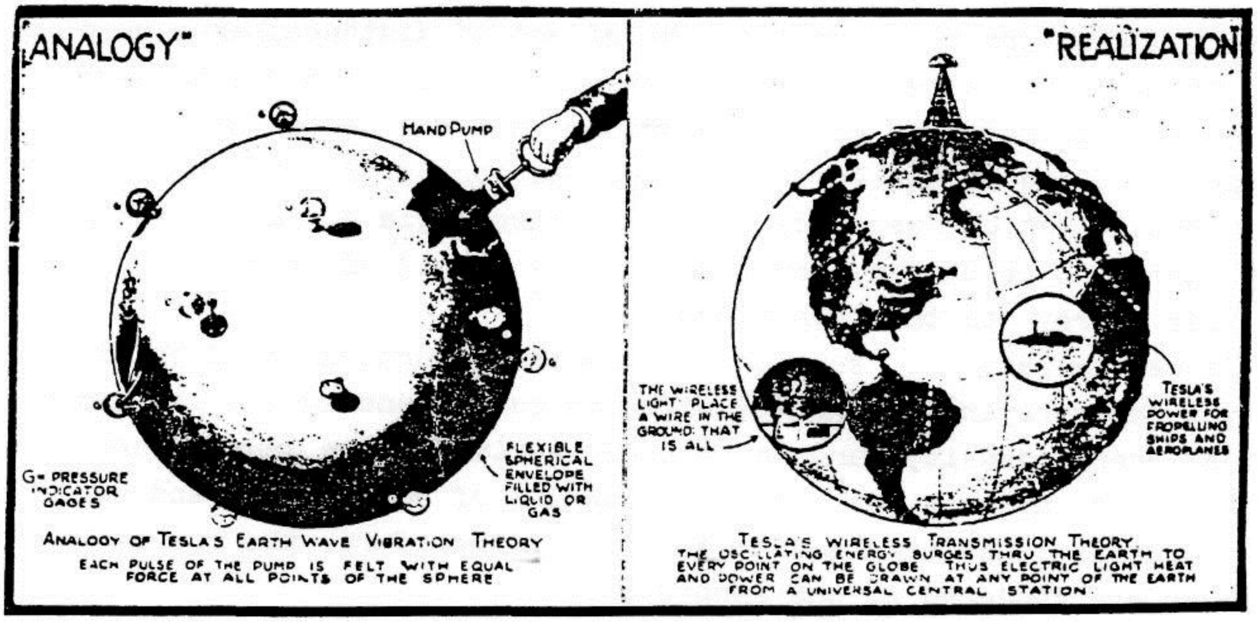
"ANALOGY"
HAND PUMP
G= PRESSURE INDICATOR GAGES
FLEXIBLE SPHERICAL ENVELOPE FILLED WITH LIQUID OR GAS
ANALOGY OF TESLA'S EARTH WAVE VIBRATION THEORY
EACH PULSE OF THE PUMP IS FELT WITH EQUAL FORCE AT ALL POINTS OF THE SPHERE
"REALIZATION"
THE WIRELESS LIGHT: PLACE A WIRE IN THE GROUND: THAT IS ALL
TESLA'S WIRELESS POWER FOR PROPELLING SHIPS AND AEROPLANES
TESLA'S WIRELESS TRANSMISSION THEORY
THE OSCILLATING ENERGY SURGES THRU THE EARTH TO EVERY POINT ON THE GLOBE. THUS ELECTRIC LIGHT HEAT AND POWER CAN BE DRAWN AT ANY POINT OF THE EARTH FROM A UNIVERSAL CENTRAL STATION.

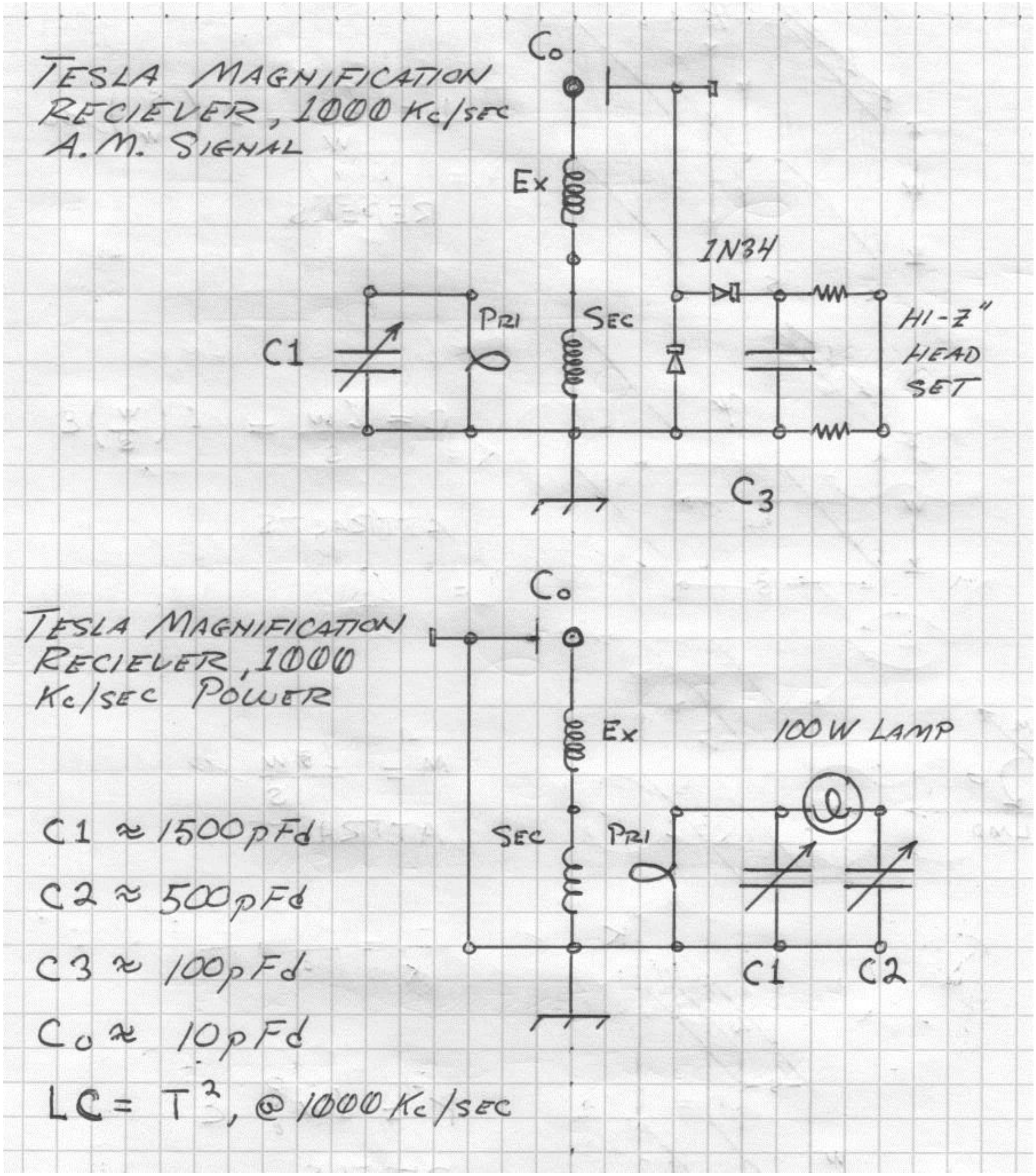
TESLA MAGNIFICATION RECIEVER, 1000 Kc/sec A.M. SIGNAL
Co
Ex
1N34
C1
PRI
SEC
HI-Z" HEAD SET
C3
TESLA MAGNIFICATION RECIEVER, 1000 Kc/sec POWER
Co
Ex
100 W LAMP
SEC
PRI
C1
C2
C1 ≈ 1500 pFd
C2 ≈ 500 pFd
C3 ≈ 100 pFd
Co ≈ 10 pFd
LC = T², @ 1000 Kc/sec

Hence it can be seen that a pair of waves are engendered by this transmission system. (Tower and Star Radials). One wave, Hertzian, is the over ground wave, the other wave, Telluric, is the under ground wave. These two waves arrive at the point of reception in their own distinct time frames, giving rise to a difference in phase. Hence, multiple rings of interference patterns are produced. Since the Hertzian portion, over ground, time frame is based upon the velocity of light, then the Telluric portion, under ground, time frame gives the Telluric velocity. Two crystal sets, one over ground, one under ground, and a basic oscilloscope , that simple. I have done this at Landers.

Concluding, a Tesla Magnification Transformer, properly proportioned can, in theory, actually draw power from a local 50 kW station. Several hundred watts of power reception is likely. This would prove Tesla once and for all. No antenna, just a good ground, and a nice and bright 100 watt light bulb.

This would overturn physics more than any billion dollar C.E.R.N. project. A ham radio operator overturns Einstein for 100 bucks. What a concept.
Read,
Tesla, "The True Wireless"
Tesla, "System of Concatenated Tuned Circuits"
Dollard, "System for the Transmission and Reception of Telluric Electric Waves"
A.R.R.L. "Radio Amatuers Handbook". Chapter "H.F. Transmitters, & Tank Circuits"
73 DE N6 KPH

01-03-2012, 03:58 PM

t-rex ●
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Tesla's Diagram of overground vs underground waves Hetzian/Telluric

Tesla's Diagram of overground vs underground waves Hetzian/Telluric

Note Telluric velocity, a function of angle around the earth, it is a cosecant, infinity at the poles, luminal at the equator. Pi over two is the integral "velocity"

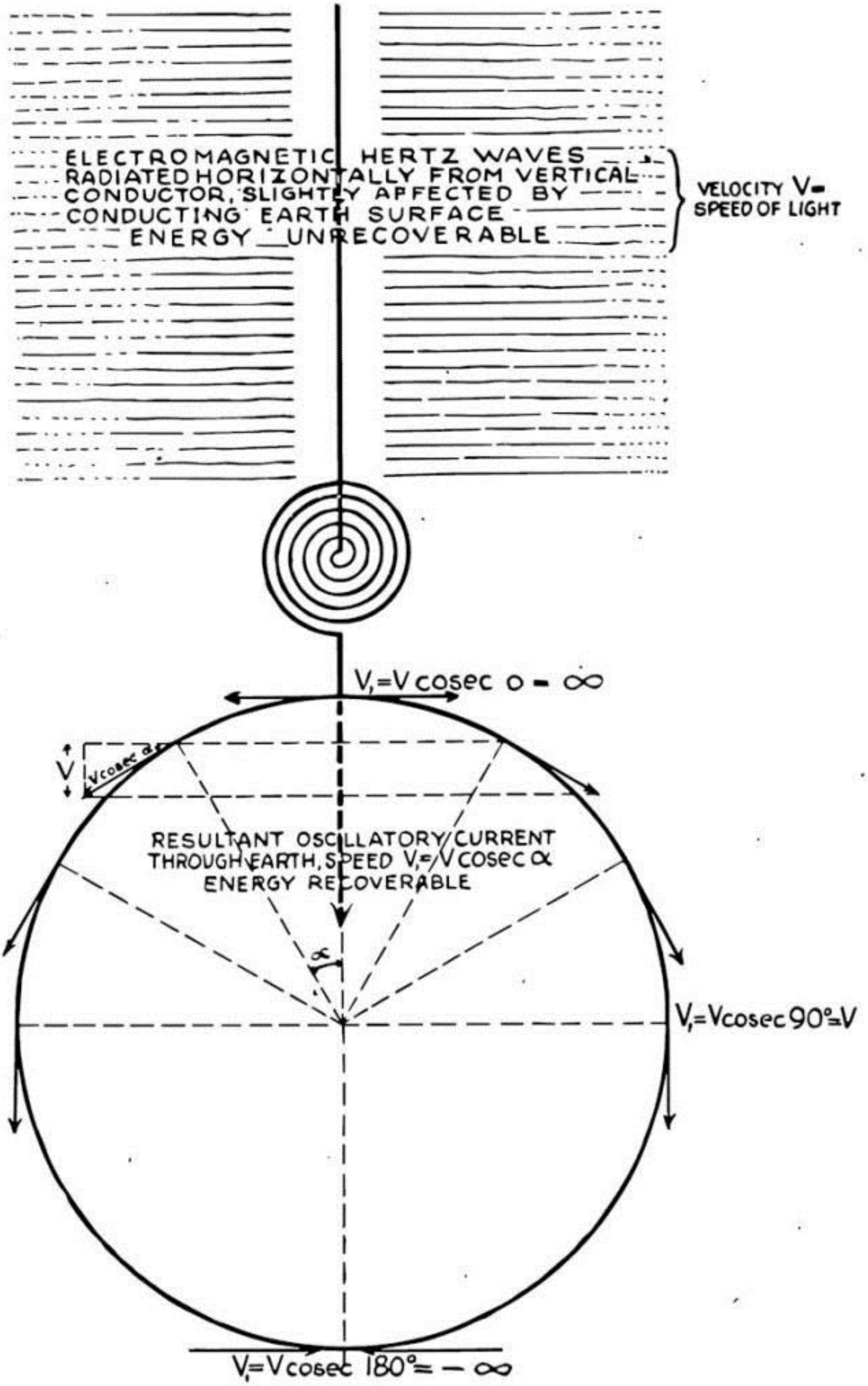
ELECTROMAGNETIC HERTZ WAVES RADIATED HORIZONTALLY FROM VERTICAL CONDUCTOR, SLIGHTLY AFFECTED BY CONDUCTING EARTH SURFACE
ENERGY UNRECOVERABLE
VELOCITY V= SPEED OF LIGHT
V₁=V cosec 0 = ∞
V
V cosec α
RESULTANT OSCILLATORY CURRENT THROUGH EARTH, SPEED V₁=V cosec α
ENERGY RECOVERABLE
α
V₁=V cosec 90°=V
V₁=V cosec 180°= − ∞

02-03-2012, 03:35 PM

t-rex ●
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Notes on the Crystal Set

Notes on the Crystal Set

When measuring the velocity difference between underground and our ground, the two seperate receivers can be brought into phase unison by the use of a test oscillator. Both receivers must have a primary tank circuit, here you can place your measurement equipment.

The detected AM output, audio frequency, signal is immune to receiver phase shift. This is a property of A.M. detection. Use the audio to measure propagation delay.

Wind all coils in the same direction, this is a must! Brass is a very good primary material, good depth of penetration at broadcast frequencies. Wires and clip lead interconnection are the "Road to Absolute Defeat." Coils and Condensers must be merged into each other in broad surfaces. Stacked metal plates on insulators makes good condensers. Transmitter air condensers are good, vacuum condensers are the best. A number of small condensers are better than one big one, less stray inductance. Stray inductance is fatal to success! For smaller units silver mica condensers are the only way, use no others. Paralell them on a low inductance bus.

Do not get sidetracked, experimental and auxilliary coils as well as patent diagrams will get you stuck in the sand forever. The finished Colorado unit dimensions are all in the notebook. It is all there, look at it, not the diversions, and remember everything you thought was Tesla is the exact opposite. It is there; one or two turn primary; a 20% height/width secondary; and a 100% height/width extra coil, all there, wire size, turns, all there, can you find it? It then can be scaled to any frequency and any KVAR capacity within wavelength conditions.

As your unit gets perfected the signal gets stronger and stronger, then diodes start burning out. We then are reaching the engineering objective, that is to light a number 327 pilot lamp. Not much power but too bright for Einstein.

02-06-2012, 02:44 AM

t-rex ●
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Tesla Type Crystal Radio

The design of the Tesla Transformer starts with the secondary resonant coil. In its resonant mode it is a quadrapolar resonator. Two constants exist in the construction of this solenoidal resonator, one is that the height to length ratio must be 20%, the other is that the side by side spacing of the cylindrical conductor turns is 62% the diameter of the cylindrical conductor. The optimum number of turns is 20 on the secondary, with 2 on the primary. This gives the physical size of the coil for a given frequency. Increasing the number of turns reduces the size of the secondary coil in proportion to the increase in turns. Accordingly the conductor diameter also decreases as the copper is spread out over more windings. The ratio of secondary turns to primary turns must equal a constraint of 10 to 1. For a given number of turns the operating frequency becomes the sole independent variable in coil design.

The primary coil is the same diameter as the secondary coil. The ratio of conductor width to coil diameter is 18%. The sheet conductor is closely spiralled into two turns, for a secondary of twenty turns. For a conductor thickness as given by the maximum thickness vs frequency the volume of the primary metal must equal the volume of the secondary metal, based upon equal weights for primary and secondary metals. Brass can be 1.7 times thicker than copper, and bronze can be 2.8 times thicker than copper, for a given maximum thickness and frequency.

The primary condenser should be made of the same amount of metal as the primary coil. Their weights should be about the same. The connecting leads must be short and the same width as the primary conductor. Large, multiplate air condensers are good, so are metal sheets and glass.

The secondary capacitance is best a copper tubing ring around the outside of the H end of the coil. This ring must be open in one spot along its loop so as not to be a shorted turn. The end of the secondary wire connects to this capacity ring. Beer cans make the best elevated capacitors.

The finished transformer must operate on a directly connected ground plane. A large plywood table with aluminum foil glued on to it, this covered with a plastic or glass sheet is the minimum required ground plane. Obviously the aluminum plane must be solidly earthed with special attention to minimizing ground lead inductance. In order to light a lamp a more broadcast station style of ground plane is required, but for testing purposes the metallized plywood "ten by" sheet is just fine. Good luck and good crystal set DX ing.

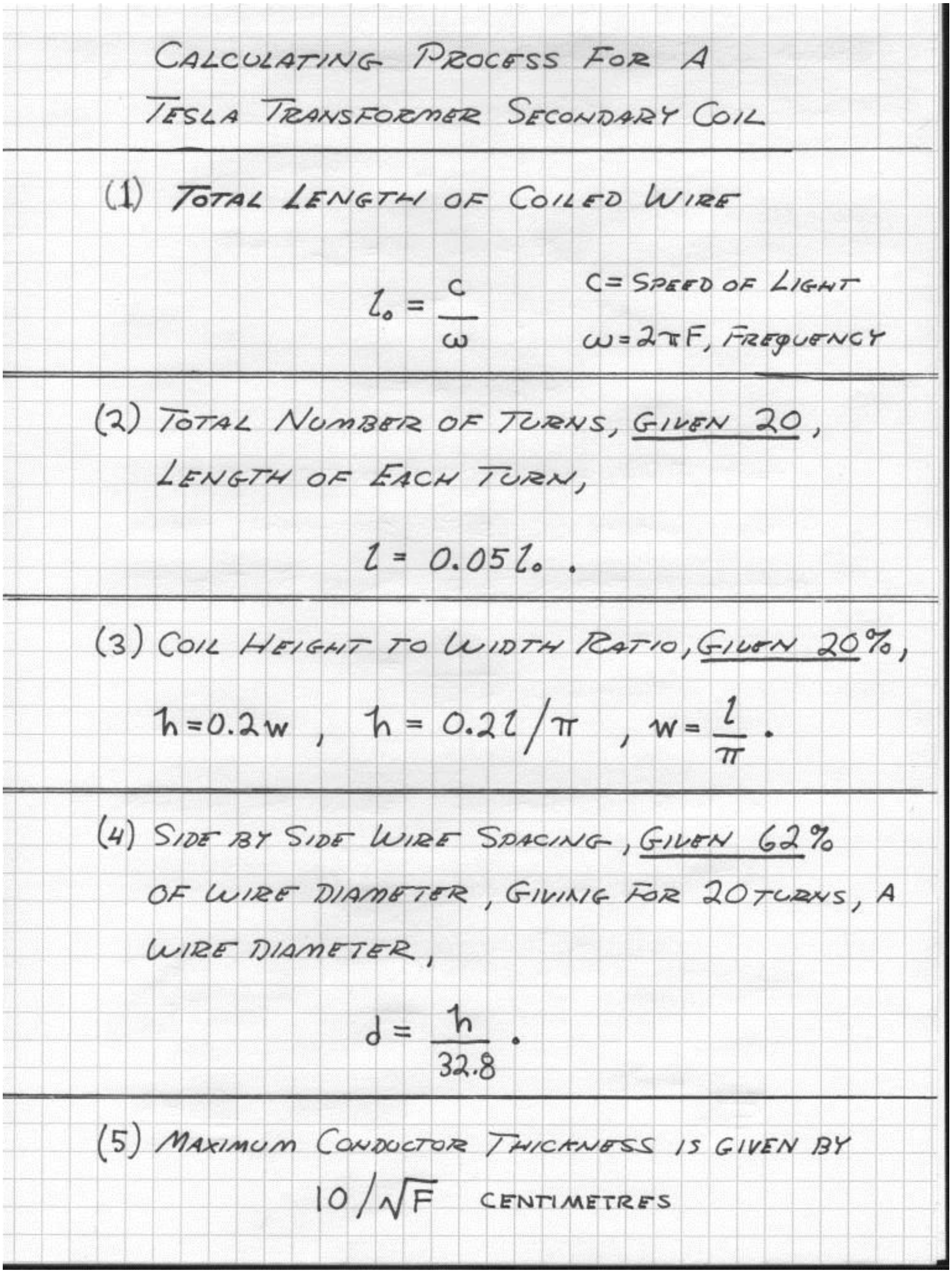

CALCULATING PROCESS FOR A TESLA TRANSFORMER SECONDARY COIL

(1) TOTAL LENGTH OF COILED WIRE

$$l_0 = \frac{c}{\omega}$$

C = SPEED OF LIGHT

$\omega = 2\pi F$, FREQUENCY

(2) TOTAL NUMBER OF TURNS, GIVEN 20,
LENGTH OF EACH TURN,

$$l = 0.05\,l_0 .$$

(3) COIL HEIGHT TO WIDTH RATIO, GIVEN 20%,

$$h = 0.2w \quad , \quad h = 0.2l/\pi \quad , \quad w = \frac{l}{\pi} .$$

(4) SIDE BY SIDE WIRE SPACING, GIVEN 62%
OF WIRE DIAMETER, GIVING FOR 20 TURNS, A
WIRE DIAMETER,

$$d = \frac{h}{32.8} .$$

(5) MAXIMUM CONDUCTOR THICKNESS IS GIVEN BY

$10/\sqrt{F}$ CENTIMETRES

02-08-2012, 06:47 PM

t-rex
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Tesla Resonant Transformer

TESLA TRANSFORMER, 1000 Kc/sec

NOTE PREVIOUS ERRORS: , CORRECTED ARE GIVEN,
QUADRAPOLAR RESONANCE, END; $\alpha_m \alpha_d = 1$ *
TESLA TRANSFORMER, END; $10/\sqrt{F}$, INCHES *

DESIGN FORMULAE:

$$l_0 = \frac{c}{\omega}, \quad c = \text{Velocity of Light}, \quad \omega = 2\pi F, \text{ Frequency.} \quad \Big\} \text{ Wire Length}$$

$$l = \frac{l_0}{N}, \quad N = \text{Number of Turns.} \quad \Big\} \text{ Turn Length}$$

$$w = \frac{l}{\pi}, \quad \text{Coil Diameter}$$

$$h = 0.2w, \quad \text{Coil Height}$$

$$a = \left[N + (N-1)0.62 \right], \quad \text{Spacing Factor}$$

$$d = \frac{h}{a}, \quad \text{Wire Diameter}$$

$$\delta = \frac{10}{\sqrt{F}} = \text{Current Depth in Inches}$$

(*, NOT $\alpha_m / \alpha_d = 1$, NOT CENTIMETRES *)

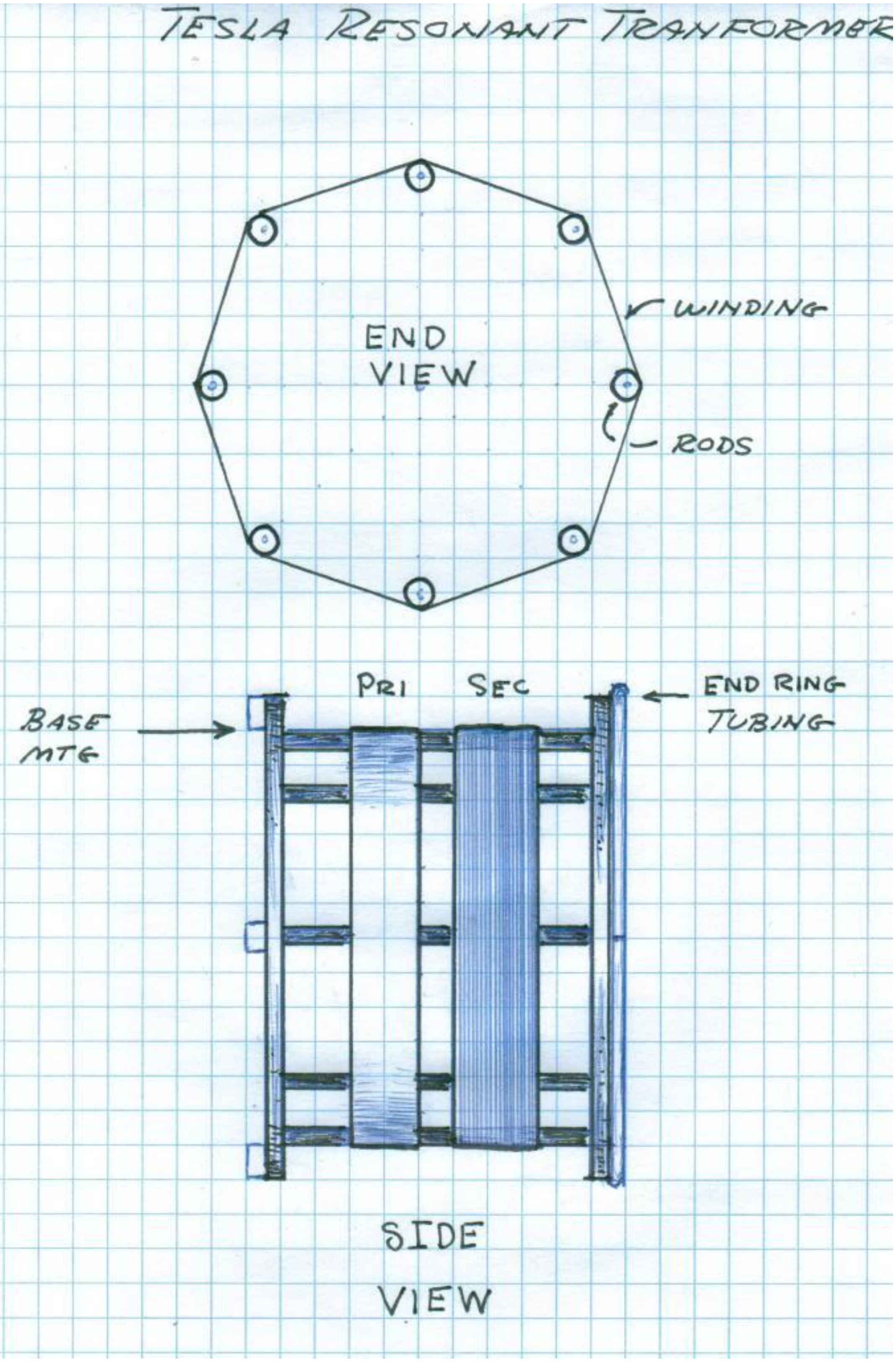
TESLA RESONANT TRANFORMER
WINDING
END
VIEW
RODS
PRI
SEC
END RING
TUBING
BASE
MTG
SIDE
VIEW

GIVEN ;

NUMBER OF TURNS = 30

FREQUENCY OF OPERATION = 1000 Kc/SEC

$l_o = 3 \times 10^{10} / 6.28 \times 10^{6}$ CENTIMETRES

$l_o = 4.78 \times 10^{3}$, WIRE LENGTH , CENTIMETRES

$l = 4.78 \times 10^{3} / 30$ CENTIMETRES

$l = 1.6 \times 10^{2}$, TURN LENGTH , CENTIMETRES

$w = 1.6 \times 10^{2} / \pi$ CENTIMETRES

$w = 5.1 \times 10^{1}$, COIL DIAMETER , CENTIMETRES

$h = 0.2 \times 5.1 \times 10^{1}$ CENTIMETRES

$h = 10.1$, COIL HEIGHT , CENTIMETRES

$a = 30 + 18 = 48$, FACTOR NUMERIC

$d = 10.1 / 48$ CENTIMETRES

$d = 0.21$, WIRE DIA , CENTIMETRES

$\delta = 2.54 / \sqrt{10^{6}}$ CENTIMETRES

$\delta = 0.03$, SKIN DEPTH , CENTIMETRES

(PRIMARY DIAMETER , 51.0 , CENTIMETRES)

TESLA TRANSFORMER, 1000 Kc/sec

NOTE PREVIOUS ERRORS:, CORRECTED ARE GIVEN,

QUADRAPOLAR RESONANCE, END; $\alpha_m \alpha_d = 1$ *

TESLA TRANSFORMER, END; $10/\sqrt{F}$, INCHES *

DESIGN FORMULAE:

$$\left.\begin{array}{l} l_0 = \dfrac{c}{\omega}, \quad \text{C = VELOCITY OF LIGHT,} \\ \omega = 2\pi F, \quad \text{FREQUENCY.} \end{array}\right\} \text{WIRE LENGTH}$$

$$\left. l = \frac{l_0}{N}, \quad \text{N = NUMBER OF TURNS.} \right\} \text{TURN LENGTH}$$

$$w = \frac{l}{\pi}, \quad \text{COIL DIAMETER}$$

$$h = 0.2w, \quad \text{COIL HEIGHT}$$

$$a = \left[N + (N-1)0.62\right], \quad \text{SPACING FACTOR}$$

$$d = \frac{h}{a}, \quad \text{WIRE DIAMETER}$$

$$\delta = \frac{10}{\sqrt{F}} = \text{CURRENT DEPTH IN INCHES}$$

*, NOT $\alpha_m/\alpha_d = 1$, NOT CENTIMETRES *)

02-12-2012, 01:46 PM

[t-rex](#)
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Calculating Properties for Secondary

20 TURN SECONDARY COIL

GENERAL FORMULAE IN CENTIMETRES

CONDUCTOR LENGTH, $l_s = 4.8 \times 10^9 \div F$ cm (1)

COIL DIAMETER, $w_s = 7.6 \times 10^7 \div F$ cm (2)

COIL HEIGHT, $h_s = 1.5 \times 10^7 \div F$ cm (3)

MAXIMUM HOLLOW CONDUCTOR DIAMETER,

$d_o = 4.8 \times 10^5 \div F$ cm (4)

MAXIMUM HOLLOW CONDUCTOR WALL THICKNESS,

$d_1 = 2.5 \times 10^1 \div \sqrt{F}$ cm (5)

MAXIMUM SOLID CONDUCTOR DIAMETER,

$d_s = 5.1 \times 10^1 \div \sqrt{F}$ cm (6)

FREQUENCY OF OPERATION

F, CYCLES PER SECOND PER SEC

02-13-2012, 02:50 PM

t-rex ●
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

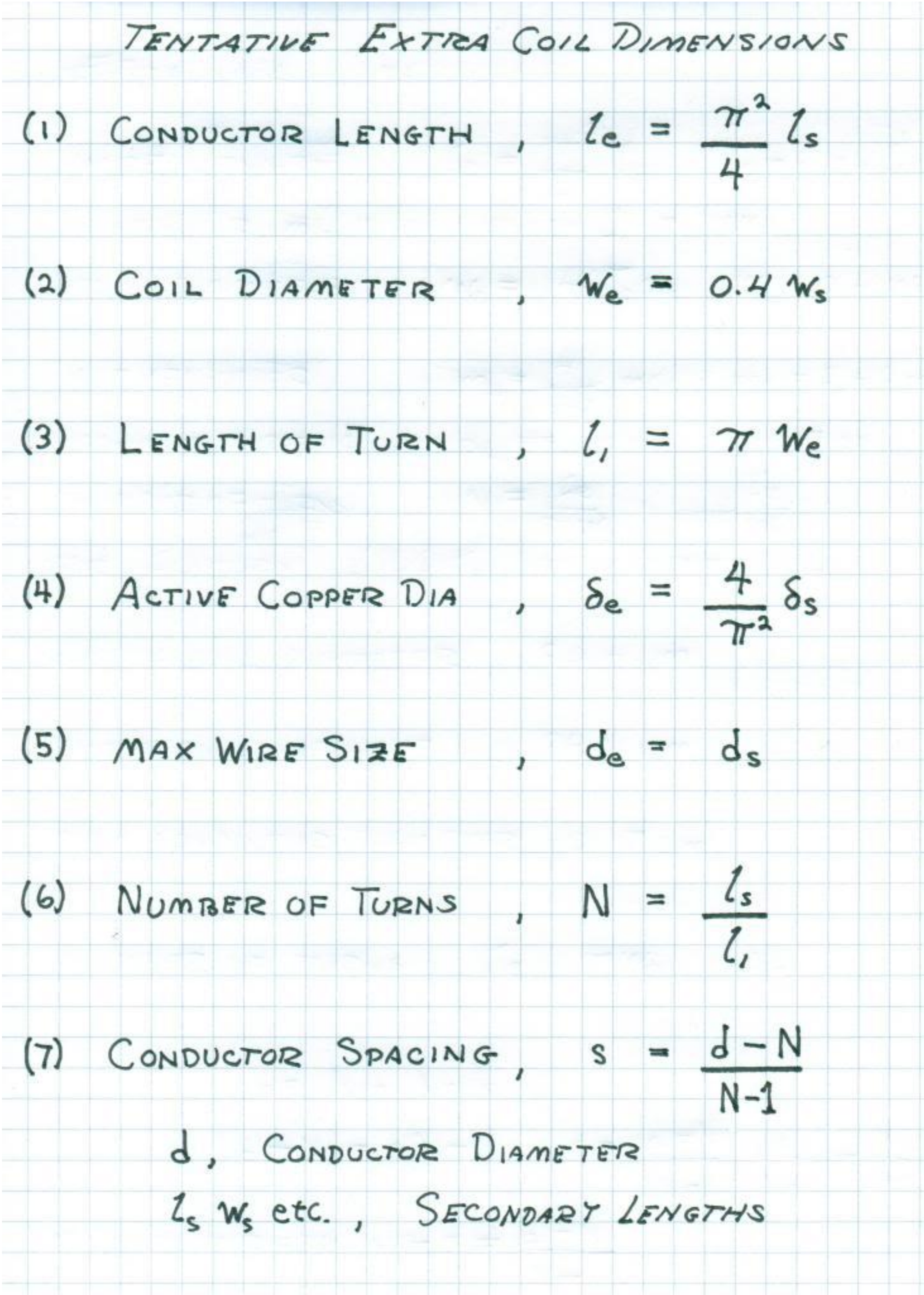

TENTATIVE EXTRA COIL DIMENSIONS

(1) CONDUCTOR LENGTH , $l_e = \frac{\pi^2}{4} l_s$

(2) COIL DIAMETER , $W_e = 0.4\, W_s$

(3) LENGTH OF TURN , $l_1 = \pi W_e$

(4) ACTIVE COPPER DIA , $\delta_e = \frac{4}{\pi^2} \delta_s$

(5) MAX WIRE SIZE , $d_e = d_s$

(6) NUMBER OF TURNS , $N = \frac{l_s}{l_1}$

(7) CONDUCTOR SPACING , $s = \frac{d - N}{N-1}$

d , CONDUCTOR DIAMETER

l_s W_s etc. , SECONDARY LENGTHS

02-14-2012, 06:22 PM

t-rex ●
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

CONCATENATED RESONANT COILS, 1000 Kc/SEC

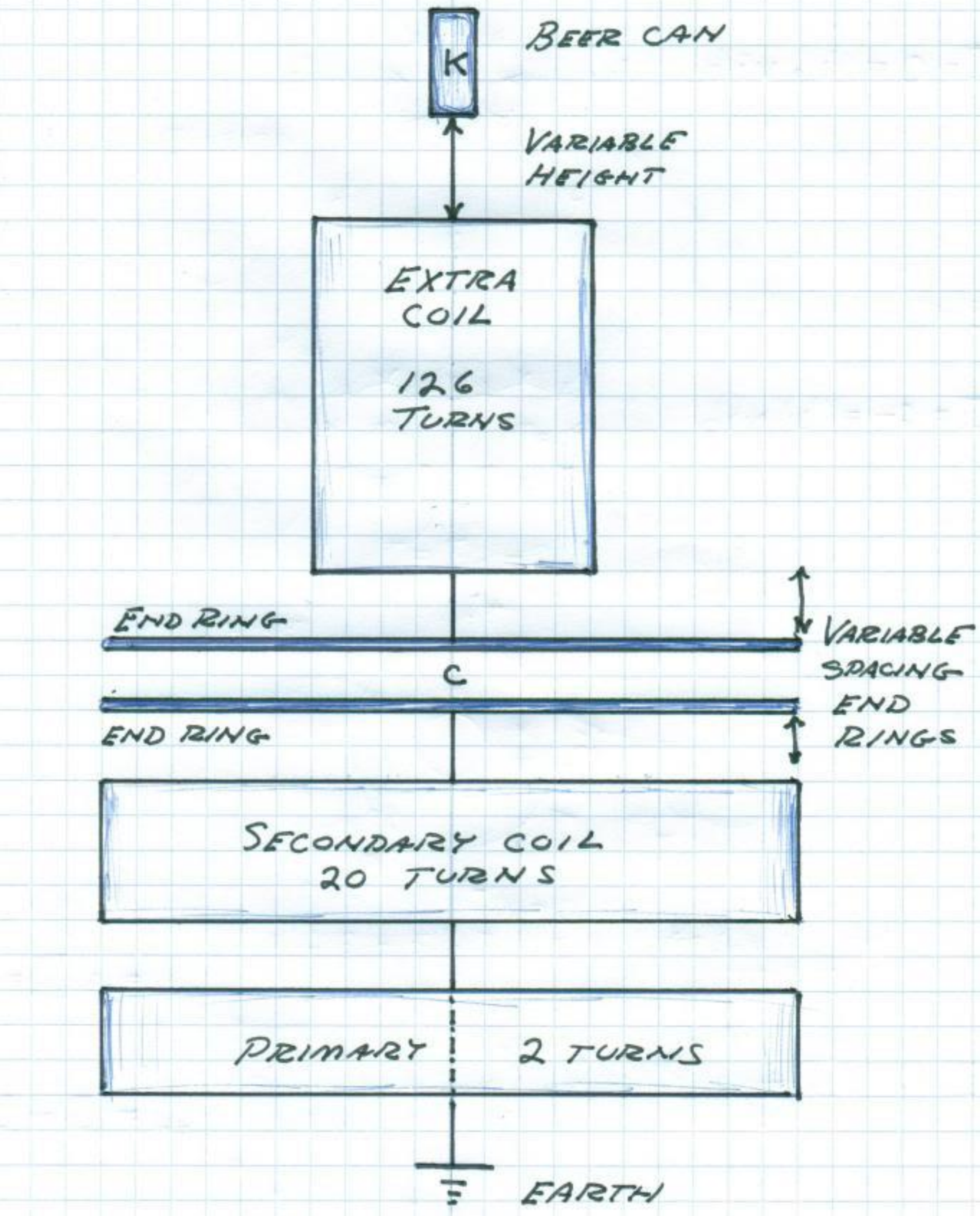
K
BEER CAN
VARIABLE HEIGHT
EXTRA COIL
126 TURNS
END RING
C
END RING
VARIABLE SPACING END RINGS
SECONDARY COIL
20 TURNS
PRIMARY
2 TURNS
EARTH

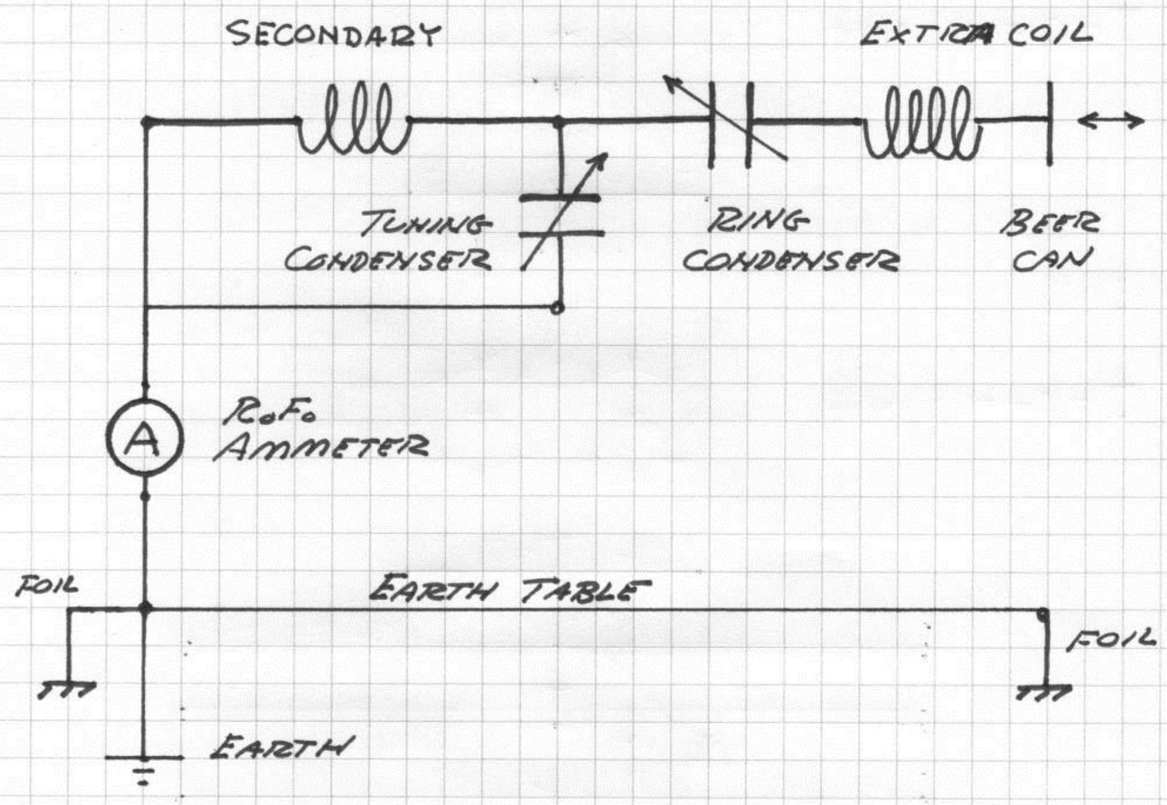
SECONDARY
EXTRA COIL
TUNING CONDENSER
RING CONDENSER
BEER CAN
R.F. AMMETER
A
FOIL
EARTH TABLE
FOIL
EARTH

DIAGRAM OF INTER CONNECTIONS AND TUNING CONDENSERS

04-09-2012, 08:55 AM

Web000x
Silver Member

Join Date: Apr 2009
Posts: 539

@All Crystal Radio Builders,

Eric would like to get some feedback on his given TMT design formulas. The following is a test procedure used to determine the resonance of the coils.

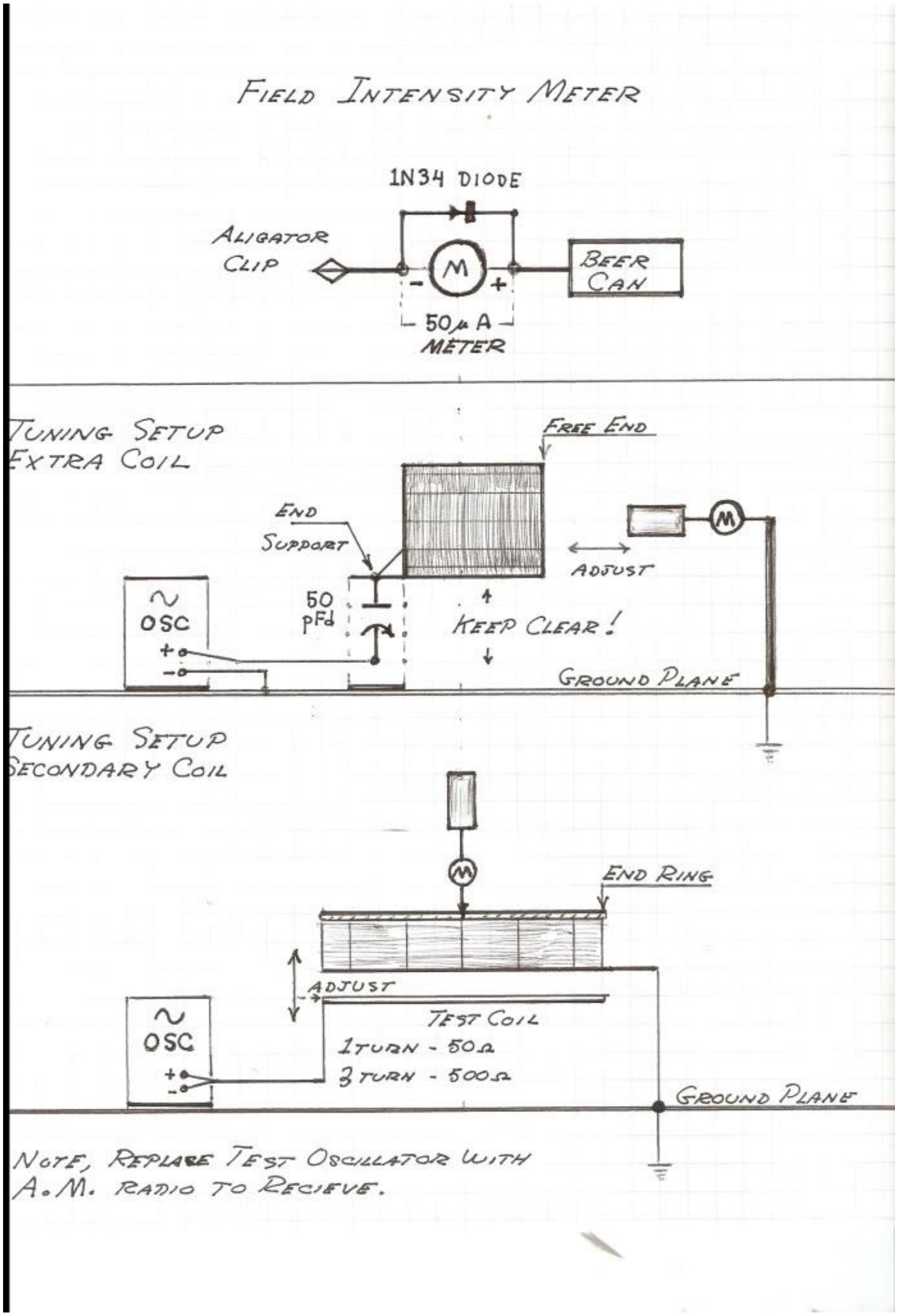
FIELD INTENSITY METER
1N34 DIODE
ALIGATOR CLIP
M
-
+
BEER CAN
50µA METER
TUNING SETUP EXTRA COIL
FREE END
END SUPPORT
M
ADJUST
OSC
+
-
50 pFd
KEEP CLEAR!
GROUND PLANE
TUNING SETUP SECONDARY COIL
M
END RING
ADJUST
OSC
+
-
TEST COIL
1 TURN - 50Ω
3 TURN - 500Ω
GROUND PLANE
NOTE, REPLACE TEST OSCILLATOR WITH A.M. RADIO TO RECIEVE.

The best way to get him the data will be to mail the results to the Lone Pine general delivery address.

Dave

05-14-2012, 12:26 AM

dR-Green
Gold Member

Join Date: Apr 2008
Location: Merlin's Fort
Posts: 1,220

Using these:

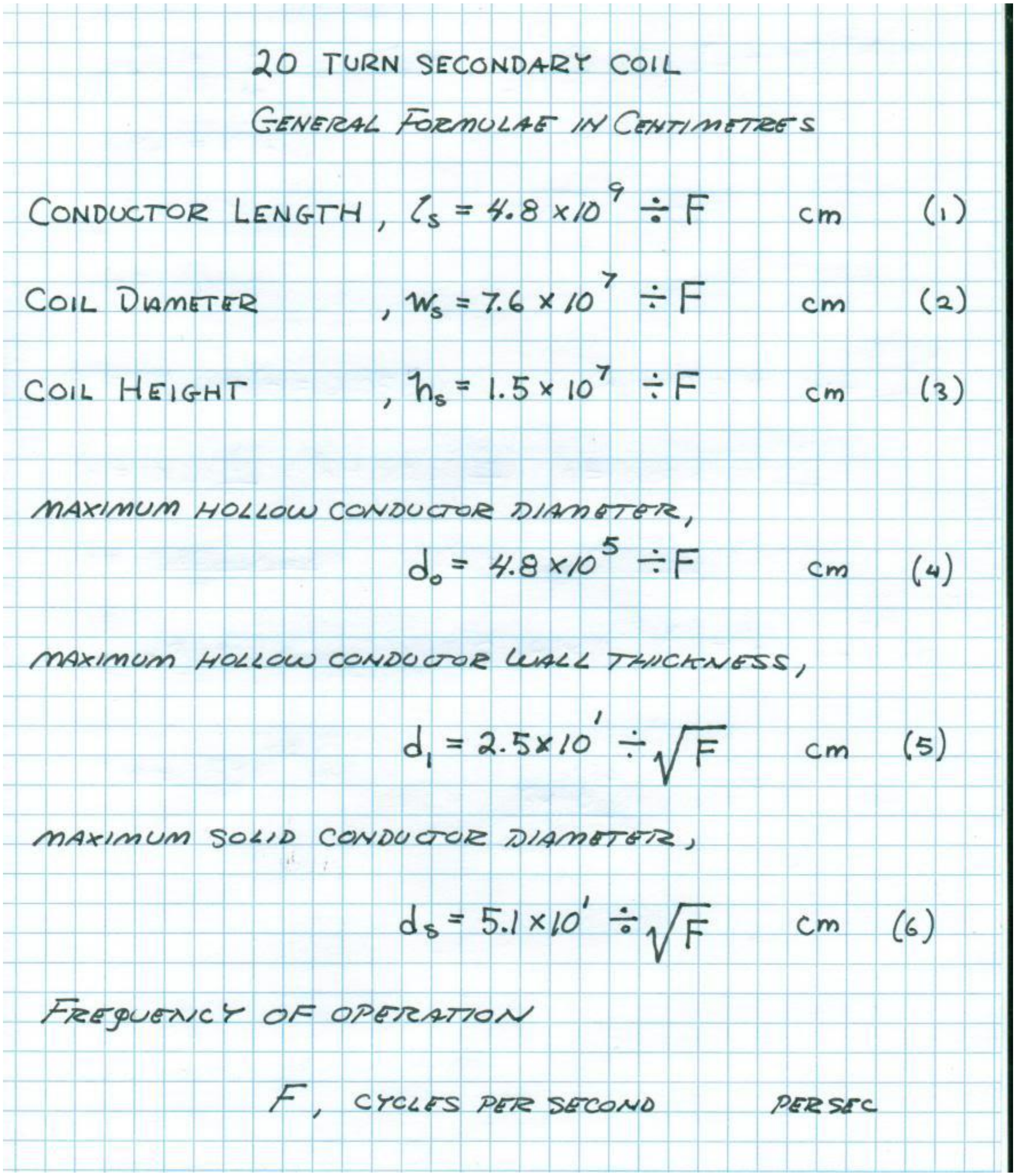

20 TURN SECONDARY COIL

GENERAL FORMULAE IN CENTIMETRES

CONDUCTOR LENGTH, $l_s = 4.8 \times 10^9 \div F$ cm (1)

COIL DIAMETER, $w_s = 7.6 \times 10^7 \div F$ cm (2)

COIL HEIGHT, $h_s = 1.5 \times 10^7 \div F$ cm (3)

MAXIMUM HOLLOW CONDUCTOR DIAMETER,

$d_0 = 4.8 \times 10^5 \div F$ cm (4)

MAXIMUM HOLLOW CONDUCTOR WALL THICKNESS,

$d_1 = 2.5 \times 10^1 \div \sqrt{F}$ cm (5)

MAXIMUM SOLID CONDUCTOR DIAMETER,

$d_s = 5.1 \times 10^1 \div \sqrt{F}$ cm (6)

FREQUENCY OF OPERATION

F, CYCLES PER SECOND PER SEC

TENTATIVE EXTRA COIL DIMENSIONS

(1) CONDUCTOR LENGTH , $l_e = \frac{\pi^2}{4} l_s$

(2) COIL DIAMETER , $W_e = 0.4 W_s$

(3) LENGTH OF TURN , $l_1 = \pi W_e$

(4) ACTIVE COPPER DIA , $\delta_e = \frac{4}{\pi^2} \delta_s$

(5) MAX WIRE SIZE , $d_e = d_s$

(6) NUMBER OF TURNS , $N = \frac{l_s}{l_1}$

(7) CONDUCTOR SPACING , $s = \frac{d - N}{N-1}$

d , CONDUCTOR DIAMETER

l_s W_s etc. , SECONDARY LENGTHS

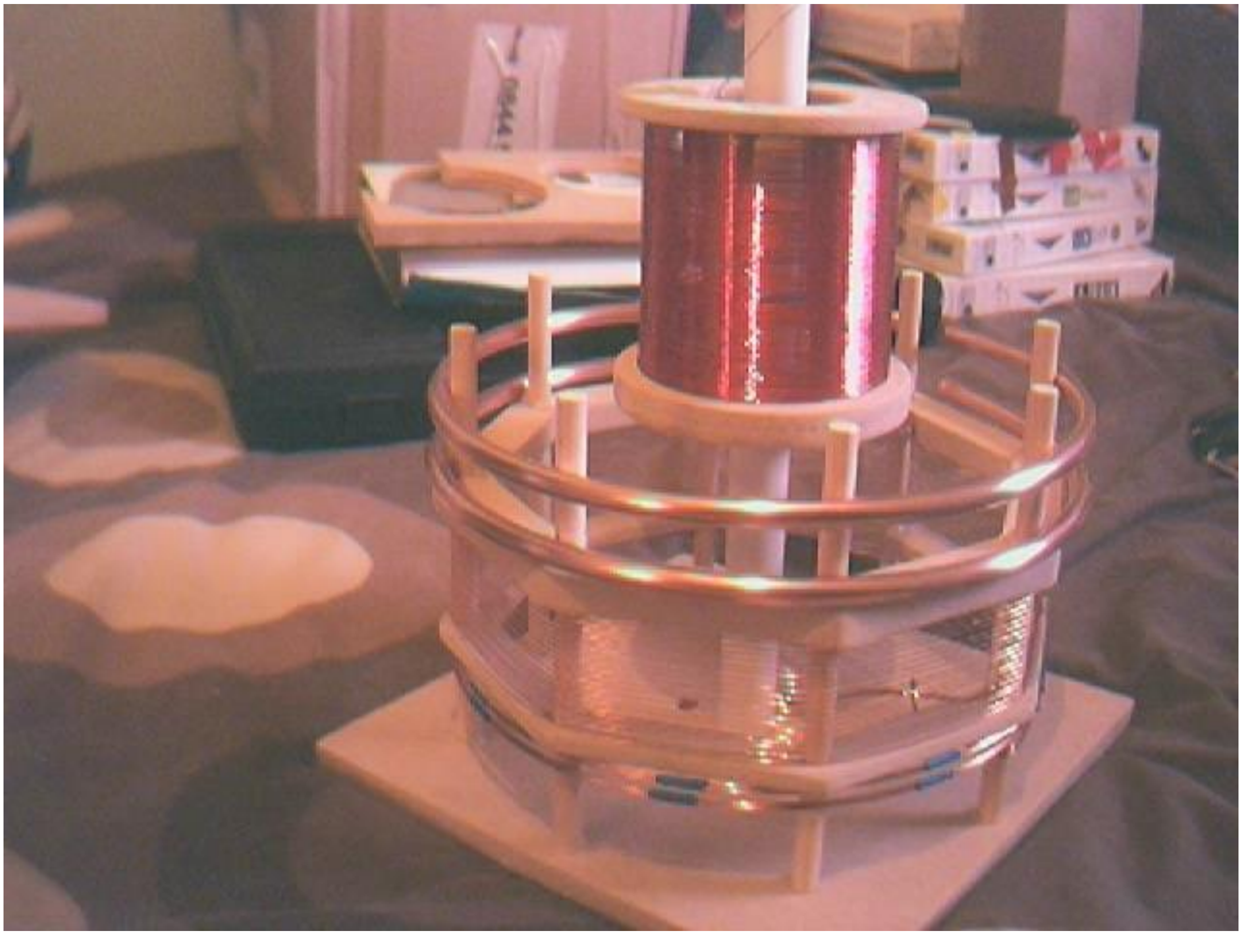

F = 3670000 cycles/sec

Secondary
Height = 4.08cm
Diameter = 20.7cm
Number of turns = 20
Wire = 13 +/- metres 26 SWG (thin wire issues resulting in slight octagon shape therefore reduced length, estimated actual length = approx 12.92 metres)
Approx 18.9 grams of copper in coil

Extra coil
Height = 8.28cm
Diameter = 8.28cm
Number of turns = 126
Wire = 32.27 metres 26 SWG (advisable not to use thicker wire otherwise no room on coil frame, impossible to space 126 turns evenly as is)

Margin of error between calculated and measured frame dimensions after construction = <0.5mm

Primary
2 turns, not flat spiral
Wire = 2 x 20 SWG + 24 SWG parallel wound

Tests results = Forthcoming

05-15-2012, 10:54 AM

t-rex
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

C.R.I. Questions and Answers

Post numbers here refer to the Crystal Radio Magic Thread

Post 73, Jake.

Standing wave on coil indicates proper operation. A voltage node must exist exactly at the neutral terminal. Tesla could move this node about in the earth. This is important and in an optimum state the voltage node is right at the point of earthing. A projected wave velocity exist on the earth surface, infinite at the earth connection, and luminal 90 degrees around the planet. It is a tangent function with Pi over two as the "Effective velocity". Hence velocity here is only a projection on the earth surface, not a true velocity, like your shadow at sundown it goes to infinity.

Post 136, Madhatter.

The B field is a bad way to look at this. It is a physics dingleberry. B is a dimensional relation with a per centimetre square. Here it is best to use versor magnetic induction k Phi, see last part of Theory of Wireless Power. Mutual inductance is longitudinal so that coupling remains for concatenated longitudinal resonators. Wave direction now has a role in coupling.

Post 143, Madhatter

Wave direction coupling makes balance to unbalance transformation impossible, this gives the "dead phase" extra coil. If it is a balanced system, every part of that system must be push-pull. Unbalanced unit has neutral on one side of primary, a balanced unit has neutral on primary center tap. All concatenated coils must be wound in same direction as wave propogation spiraling around them.

Post 156, Jake

Do not replace 100 pF condenser with LED, it will distort and rob power from the audio. Use a completely separate setup for LED. Try Ne 2 neon lamps. Also a 10-15 watt fluorescent lamp is a good standing wave "wand", this for tuning.

I would recommend using a 6SN7 Triode section with cathode drive from function generator and the primary as the plate tank circuit. See the figure. This will greatly improve resonance for tuning. 50 Ohm across primary NFG, 6SN7>6000 Ohm general.

Multiple tuning peaks indicate overly tight coupling between coils. Slowly separate them until only one resonant peak. Multiple modes of stray reception and strong local stations will make grounding and tuning very confusing. An isolation amplifier 6SN7 triode isolates amplifier from function generator interaction and gives more power to light neon lamps. Both sections of 6SN7 in parallel give more power.

73 DE N6KPH

C.R.I. TRIODE DRIVER

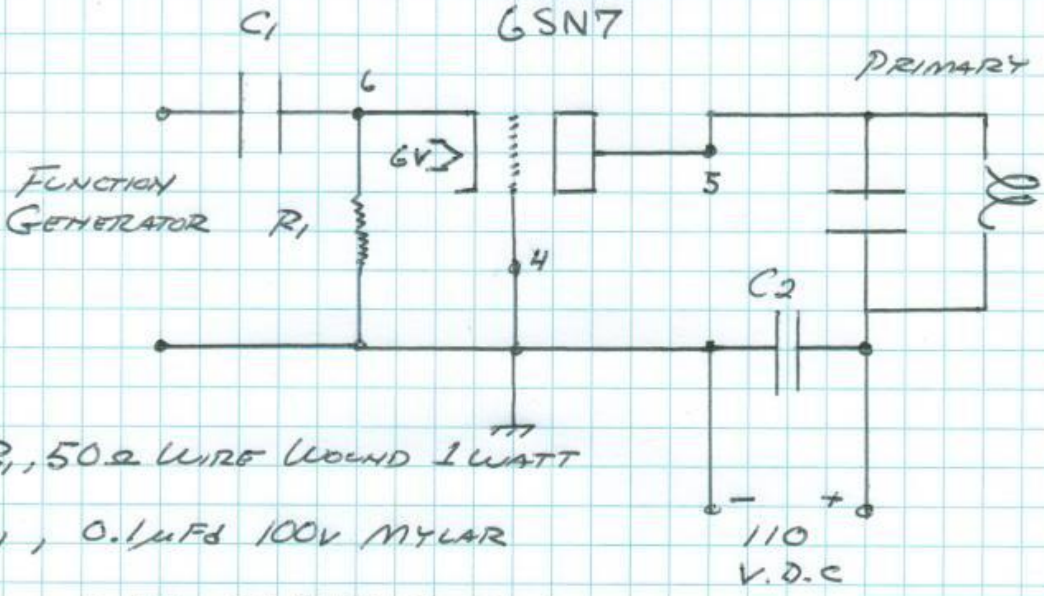
C1
6SN7
6
PRIMARY
6V
5
FUNCTION GENERATOR
R1
4
C2
−
+
110 V.D.C

R1, 50Ω WIRE WOUND 1 WATT

C1, 0.1μFd 100V MYLAR

C2, 0.05μFd 400V CERAMIC

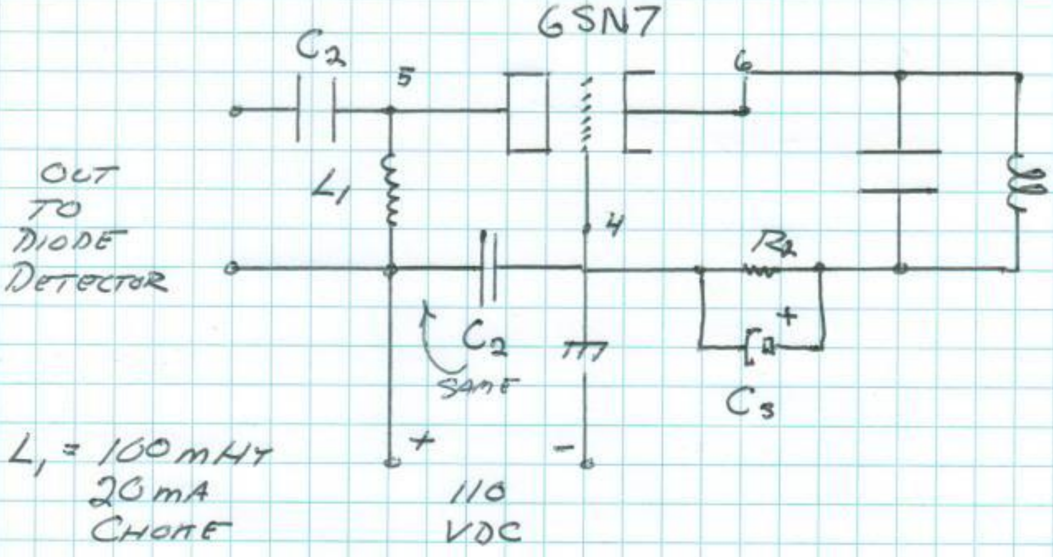
6SN7
C2
5
6
OUT TO DIODE DETECTOR
L1
4
R2
+
C2
SAME
C3
+
−
110 VDC

L1 = 100mHY
20mA
CHOKE

TRIODE BUFFER

C3 = 10μF
15V
TANTALUM

R2 = 100 Ω CARBON
1 WATT

THEORY OF WIRELESS POWER

by Eric Dollard
"Wireless Engineer"

© 1986

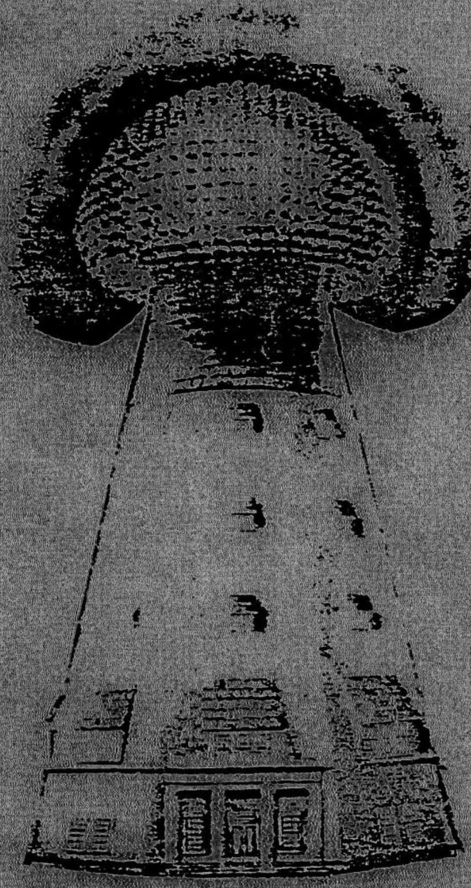

WHAT ABOUT TODAY'S SCIENTISTS?

"The scientists from Franklin to Morse were clear thinkers and did not produce erroneous theories. The scientists of today think deeply instead of clearly. One must be sane to think clearly, but one can think deeply and be quite insane.

"Today's scientists have substituted mathematics for experiments and they wander off through equation after equation and eventually build a structure which has no relation to reality."

- Nikola Tesla -

The Oscillating Current Transformer

The oscillating current transformer functions quite differently than a conventional transformer in that the law of dielectric induction is utilized as well as the familiar law of magnetic induction. The propagation of waves along the coil axis does not resemble the propagation of waves along a conventional transmission line, but is complicated by inter-turn capacitance & mutual magnetic inductance. In this respect the O.C. transformer does not behave like a resonant transmission line, nor a R.C.L. circuit, but more like a special type of wave guide. Perhaps the most important feature of the O.C. transformer is that in the course of propagation along the coil axis the electric energy is dematerialized, that is, rendered mass free energy resembling Dr. Wilhelm Reich's Orgone Energy in its behavior. It is this feature that renders the O.C. transformer usefull for wireless power transmission and reception, and gives the O.C. transformer singular importance in the study of Dr. Tesla's research.

FUNDAMENTALS OF COIL INDUCTION

Consider the elemental slice of a coil shown in fig. 1. Between the turns 1,2 & 3 of the coiled conductor exists a complex electric wave consisting of two basic components. In one component (fig. 2), the lines of magnetic and dielectric flux cross at right angles, hereby producing a photon flux perpendicular to these crossings, hereby propagating energy along the gap, parallel to the conductors and around the coil. This is the transverse electro-magnetic wave. In the other component, shown in fig. 3, the lines of magnetic flux do not cross but unite along the same axis, perpendicular to the coil conductors, hereby energy is conveyed along the coil axis. This is the Longitudinal Magneto-Dielectric Wave.

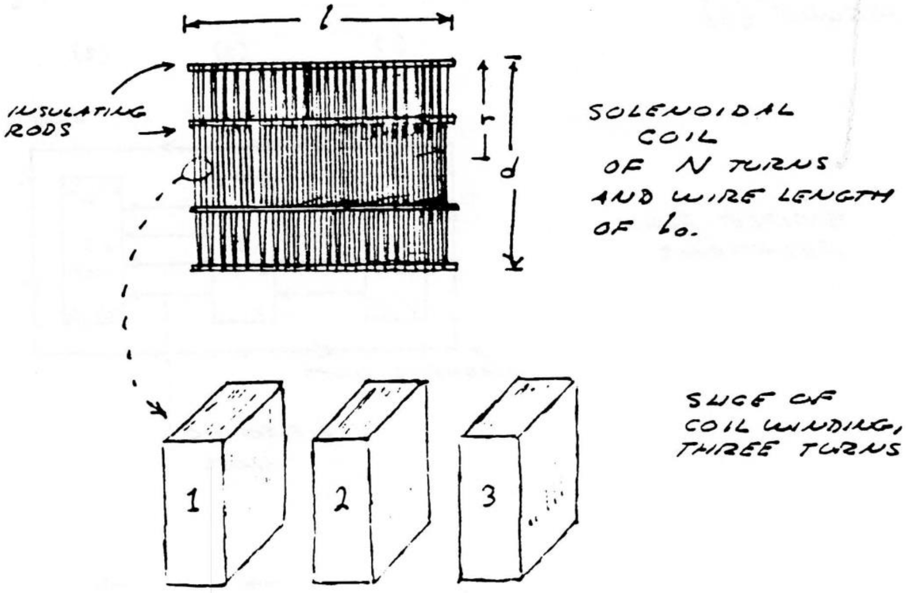
l
INSULATING RODS
r
d
SOLENOIDAL COIL OF N TURNS AND WIRE LENGTH OF l₀.
1
2
3
SLICE OF COIL WINDING, THREE TURNS

FIGURE (1)

Hence, two distinct forms of energy flow are present in the coiled conductor, propagating at right angles with respect to each other, as shown in fig. 4. Hereby a resultant wave is produced which propagates around the coil in a helical fashion, leading the transverse wave between the conductors. Thus the oscillating coil posses a complex wavelength which is shorter than the wavelength of the coiled conductor.

COIL CALCULATION

If the assumptions are made that an alternating current is applied to one end of the coil, the other end of the coil is open circuited. Additionaly external inductance and capacitance must be taken into account, then simple formulae may be derived for a single layer solenoid.

The well known formula for the total inductance of a single layer solenoid is

$$L = \frac{r^2 N^2}{(9r+10l)} \quad x\ 10^{-6}\ \text{Henry (inches)} \qquad (1)$$

Where r is coil radius
l is coil length
N is number of turns

FIGURE (2)

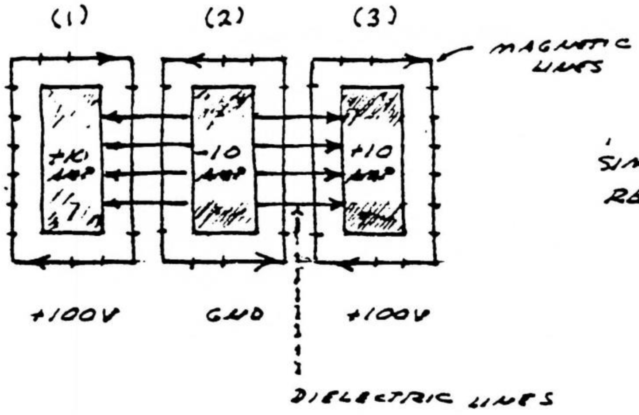
(1)
(2)
(3)
MAGNETIC LINES
+10 AMP
-10 AMP
+10 AMP
+100V
GND
+100V
DIELECTRIC LINES

SIMPLE FLUX RELATIONS

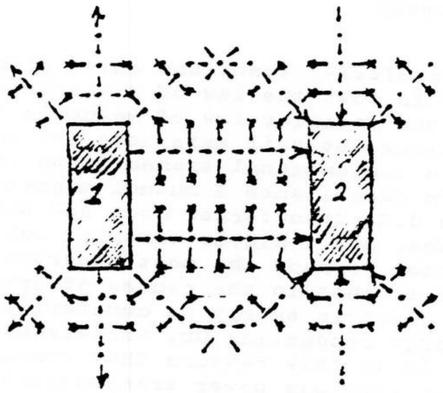

COMPOSITE FLUX PATTERN OF TWO TURNS

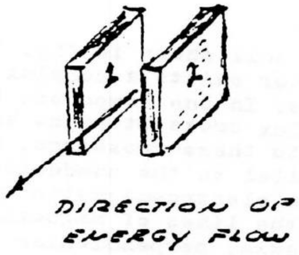
1
2
DIRECTION OF ENERGY FLOW

TRANSVERSE WAVE

FIGURE (3)

SIMPLE FLUX RELATIONS

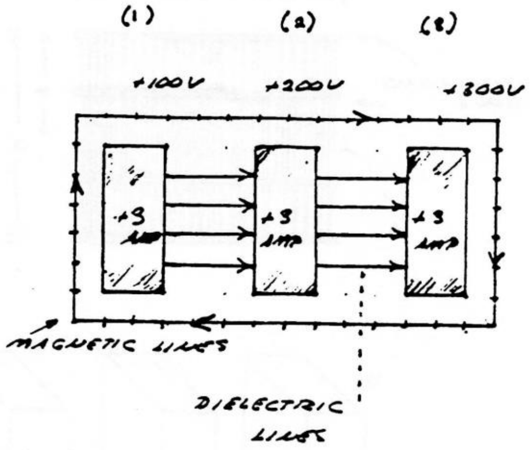
(1)
(2)
(3)
+100V
+200V
+300V
+3 AMP
+3 AMP
+3 AMP
MAGNETIC LINES
DIELECTRIC LINES

COMPOSITE FLUX PATTERN OF TWO TURNS

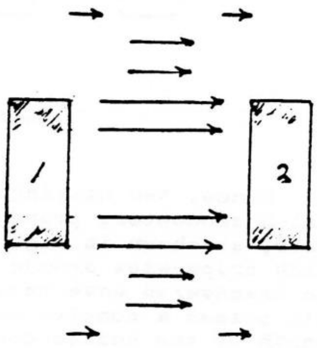

LONGITUDINAL WAVE

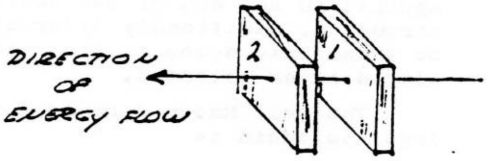
DIRECTION OF ENERGY FLOW
2
1

28

May-June 1986 JBR, Page 18

The capacitance of a single layer solenoid is given by the formula

$$C = pr \quad 2.54 \times 10^{-12} \text{ Farads (inches)} \qquad (2)$$

where the factor p is a function of the length to diameter ratio, tabulated in table (1). The dimensions of the coil are shown in figure (1). The capacitance is minimum when length to diameter ratio is equal to one.

Because the coil is assumed to be in oscillation with a standing wave, the current distribution along the coil is not uniform, but varies sinusoidially with respect to distance along the coil. This alters the results obtained by equation (1), thus for resonance

$$L_o = \tfrac{1}{2}L \qquad \text{Henrys} \qquad (3)$$

likewise, for capacitance

$$C_o = \frac{8}{\pi} C \qquad \text{Farads} \qquad (4)$$

Hereby the velocity of propagation is given by

$$V_o = 1/\sqrt{L_o C_o} = \eta V_c \qquad \text{Units/sec} \qquad (5)$$

Where

$$V_c = 1/\sqrt{\mu \varepsilon} \qquad \text{Inch/sec} \qquad (6)$$

That is, the velocity of light, and

$$V_o = V_c \eta = \left[\frac{1.77}{p} + \frac{3.94}{p} n\right]^{\frac{1}{2}} 2\pi 10^9 \qquad \text{Inch/sec} \qquad (7)$$

Where n = the ratio of coil length to coil diameter. The values of propagation factor η are tabulated in table (2).

Thus, the frequency of oscillation or resonance of the coil is given by the relation

$$F_o = V_o/(l_o \cdot 4) \qquad \text{Cycles/sec} \qquad (8)$$

Where l_o = total length of the coiled conductor in inches.

FIGURE (4)

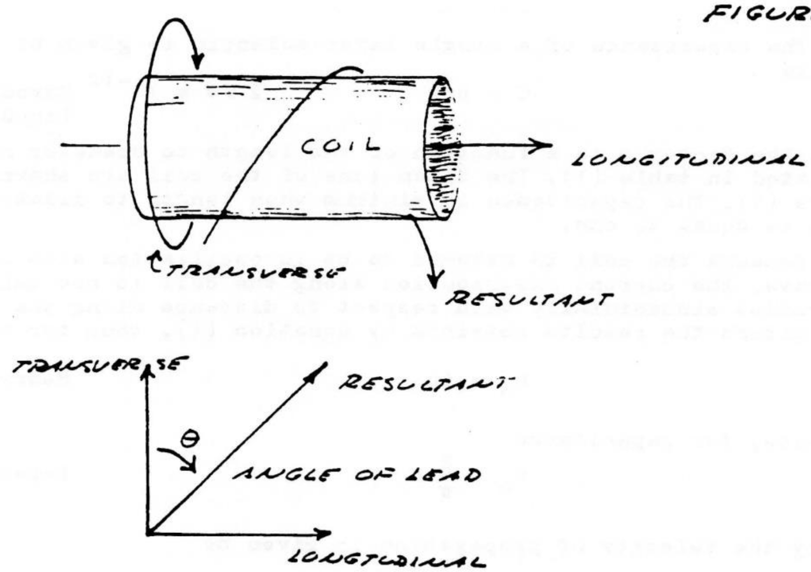

The characteristic impedance of the resonant coil is given by

$$Z_c = \sqrt{\frac{L_o}{C_o}} \qquad \text{Ohms} \qquad (9)$$

Hence,

$$Z_o = NZ_s \qquad \text{Ohms} \qquad (10)$$

Where

$$Z_s = \left[(182.9 + 406.4n)p\right]^{\frac{1}{2}} \qquad \frac{\pi}{2}\,10^3 \qquad \text{Ohms (inches)} \qquad (11)$$

and N = number of turns. The values of sheet impedance, Z_s, are tabulated in table (3).

The time constant of the coil, that is, the rate of energy dissipation due to coil resistance is given by the approximate formula

$$u = R_o/2L_o = \left(\frac{2.72}{r} + \frac{2.13}{l}\right)\pi\sqrt{F_o} \qquad \text{Nepers/sec (inches)} \qquad (12)$$

Where r = coil radius
l = coil length

In general, the dissipation of the coil's oscillating energy by conductor resistance:

1) Decreases with increase of coil diameter, d;
2) Decreases with increase of coil length, l, rapidly when the ratio, n, of length to diameter is small with little decrease beyond n equal to unity;
3) Is minimum when the ratio of wire diameter to coil pitch is 60%.

By examination of the attached tables, (1), (2) & (3), it is seen that the long coils of popular designs do not result in optimum performance. In general, coils should be short and wide, and not longer than n=1. The frequency is usually given as $F_o = V_c/\lambda_o$ which by equation (7) is incorrect. Winding on solid or continous formers rather than spaced slender rods, as shown in figure (1), greatly retards wave propagation as indicated in equation (6), thereby seriously distorting the wave. The dielectric constant of the coil, ε , should be as close to unity as is physically possible to insure high efficiency of transformation.

The equations for the voltampere relations of the oscillating coil are

$$\dot{E}_1 = j\,(Z_c\,Y_o + \delta)\,\dot{E}_o \qquad \text{Complex Input Voltage} \qquad (13)$$

$$\dot{I}_1 = j\,(Y_c\,Z_o + \delta)\,\dot{I}_o \qquad \text{Complex Input Current} \qquad (14)$$

$$Z_1 = \frac{Z_c\,Y_o + \delta}{Y_c\,Z_o + \delta}\,Z_o \qquad \text{Input Impedance, Ohms} \qquad (15)$$

Where

$\dot{E}_o$ = Voltage on elevated terminal

$\dot{I}_o$ = Current into elevated terminal

$Y_c = Z_c^{-1}$

Z_o = Terminal impedance

Y_o = Terminal admittance

$\delta = u/2F_o$ = Decrement

j = root of $\sqrt{-1}$

For negligible losses and absolute values

$$E_1 = (Z_c 2\pi F_o C_o)E_o \qquad \text{Volts} \qquad (16)$$

$$I_1 = (Y_c/2\pi F_o C_o)I_o \qquad \text{Amperes} \qquad (17)$$

Where

C_o = **Terminal capacitance**

By the law of conservation of energy

$$E_1 I_1 = E_o I_o \qquad \text{Volt-Amperes} \qquad (18)$$

If the terminal capacitance is small then the approximate input/output relations of the Tesla coil are given by

$$E_o = Z_c I_1 \qquad \text{Output Volts} \qquad (19)$$

$$I_1 = E_o Y_c \qquad \text{Input Amperes} \qquad (20)$$

$$I_o = Y_c E_1 \qquad \text{Output Amperes} \qquad (21)$$

$$E_1 = I_o Z_c \qquad \text{Input Volts} \qquad (22)$$

*** *** *** ***

TABLE (1)
Coil Capacitance Factor

Length/Width = n	Factor p	Length/Width = n	Factor p
0.10	0.96	0.80	0.46
0.15	0.79	0.90	0.46
0.20	0.70	1.00	0.46
0.25	0.64	1.5	0.47
0.30	0.60	2.0	0.50
0.35	0.57	2.5	0.56
0.40	0.54	3.0	0.61
0.45	0.52	3.5	0.67
0.50	0.50	4.0	0.72
0.60	0.48	4.5	0.77
0.70	0.47	5.0	0.81

TABLE (2)

TABLE (3)

Length/Width = n	V_o Inches/Sec	Percent Luminal Velocity = η
0.10	9.42×10^{9}	79.8%
0.15	10.9	92.2
0.20	12.0	102
0.25	13.0	110
0.30	13.9	118
0.35	14.8	125
0.40	15.6	132
0.45	16.4	139
0.50	17.2	146
0.60	18.4	156
0.70	19.5	165
0.80	20.5	176
0.90	21.4	181
1.00	22.1	187
1.5	25.4	215
2.0	27.6	234
2.5	28.7	243
3.0	29.7	251
3.5	30.3	257
4.0	30.9	262
4.5	31.6	268
5.0	32.4	274
6.0	33.0	279
7.0	33.9	287

**** **** **** **** **** **** ****

Books by Eric Dollard

CONDENSED INTRO TO TESLA TRANSFORMERS. This book is an abstract of theory and construction techniques of Tesla transformers. It is the result of experimental investigations and theoretical considerations. Includes relevant Tesla patents and an article on capacity by Fritz Lowenstein, Tesla's assistant. (BSRA #TE-1)....................$5.50

INTRODUCTION TO DIELECTRIC & MAGNETIC DISCHARGES IN ELECTRICAL WINDINGS. Theory of abrupt electrical oscillations such as those used by Tesla for experimental researches. Contains ELECTRICAL OSCILLATIONS IN ANTENNAE AND INDUCTION COILS by John Miller, 1919. This is one of the few articles containing equations useful to the design of Tesla coils. (BSRA #TE-2)....................................$5.50

05-27-2012, 01:41 PM

t-rex
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Comment and Correction

The following is the corrected version of the equations given in post #404.

Equations (1) and (2) still need to be substituted into equations (6), (7), and (9). Graphs should be made up also. Can someone do this and also reprint in better form the following equations. Next, taking the data from the COlorado notes, these equations should be applied to the final Colorado design, here for all to see, u, Z, etc.

Also the Guillemen line is IMPORTANT, the rest of the MIT pages on it would be helpful here. The Guillemen line is a "single wire" longitudinal transmission structure. Hence it is an analog of a longitudinal wave, or a section of a one wire transmission system. This is known as a "two terminal" network, in distinction to the standard "four terminal" network. Guillemen's rearrangement of the condensers from shunt to series greatly reduced the dielectric stress on these condenser. This allowed for a great reduction of size and weight, important for aircraft. The smallest thermionic gas thyratron suitable for impulse work is the 2050. It is an argon type so not very fast. The smallest hydrogen thyratron is the 3C45. It likes about 1500 volts DC.

As for gas diode tubes of the cold cathodes DC type is the,

OA3, 75 volt neon
OB3, 90 volt argon
OD3, 150 volt helium

and the cold cathode detector triode is the OA4G (neon).

These tube types are usually plentiful and cheap. The OA4G may be a bit hard to find however. There are several miniature types of cold cathode triodes in existence but I do not remember the numbers. I would point out here that blowing them up is not very useful, they do not make them anymore.

Finally, rather than an endless string of experiments doing "strange things" more might be gained by developing a solid theoretical concept of what we are getting into. And for the "free energy fuse box" crowd, forget it. No one is going to use wireless power transmission so she does not have to see the power line to her new "dream home" on top of the hill. Transmission of radio frequency energy for any purpose is subject to government approval, no way around it. Remember that the excuse Franklin Roosevelt, USN used to

outlaw radio was the mess of jamming and interference created by electrical experimenters leading up to 1917, the day that wireless died. This can happen again in the 911 police reality, watch out.

73 DE N6KPH

RE-WORKED AND CORRECTED FORMULAE FOR COIL CALCULATION

(1) FORMULA FOR SCALAR INDUCTANCE OF A SINGLE LAYER SOLENOID.

$$L_0 = \frac{r^2 N^2}{9r + 10l} \times 10^{-6}$$

WHERE L_0 IS THE TOTAL INDUCTANCE IN HENRY

N IS THE NUMBER OF TURNS

r IS THE COIL RADIUS IN INCHES

l IS THE COIL LENGTH IN INCHES.

(2) FORMULA FOR THE SELF CAPACITANCE OF A SINGLE LAYER SOLENOID.

$$C_0 = \rho r \qquad 2.54 \times 10^{-12}$$

WHERE, C_0 IS IN FARAD, r, IS THE RADIUS IN INCHES

ρ IS A PROPORTION FROM TABLE (1)

THE PROPORTION FACTOR IS A FUNCTION OF THE COIL ASPECT RATIO. SEE "SELF CAPACITANCE OF COILS", RADIOTRON DESIGNER HANDBOOK.

HENCE DERIVED IS THE "STATIC" INDUCTANCE AND CAPACITANCE IN THE DIMENSION OF INCHES.

COIL INDUCTANCE, L_0, HENRY (INCH)

COIL CAPACITANCE, C_0, FARAD (INCH)

BECAUSE THE COIL IS IN A PI OVER TWO OSCILLATION THE EFFECTIVE VALUES OF L_0 AND C_0 ARE ALTERED BY A STANDING WAVE.

THE FOLLOWING EFFECTIVE VALUES ARE GIVEN BY:

"OSCILLATIONS, IN ANTENNAE AND INDUCTION COILS", J. MILLER, 1919.

THE EFFECTIVE INDUCTANCE IS GIVEN BY

(3) $$L = \frac{1}{2} L_0$$ HENRY, AND

(4) $$C = \frac{8}{\pi^2} C_0$$ FARAD.

IS THE EFFECTIVE CAPACITY OF THE COIL.

THE EFFECTIVE VELOCITY IS GIVEN BY THE RELATION

(5) $$u^2 = \frac{1}{LC}$$

UNITS PER SECOND

$$u = \frac{1}{\sqrt{LC}}$$

THIS IS NOT AN ACTUAL VELOCITY IN THE NORMAL SENSE. HERE u IS A PROPORTION BETWEEN WINDING LENGTH AND FREQUENCY.

THE LUMINAL VELOCITY IN THE SPACE OCCUPIED BY THE SOLENOIDAL COILED WINDING IS GIVEN BY

(6) $$v^2 = \frac{1}{\mu \epsilon}$$

$$v = \frac{1}{\sqrt{\mu \epsilon}}$$

INCHES PER SECOND

THE PI OVER TWO RESONANT FREQUENCY OF THIS SOLENOIDAL COILED WINDING IS GIVEN BY

(7) $$F_o = \frac{u}{4 \ell_o}$$

CYCLES PER SECOND

WHERE ℓ_o IS THE TOTAL LENGTH OF COILED WIRE IN INCHES.

THE RATIO OF THE EFFECTIVE VELOCITY, u, AND THE LUMINAL VELOCITY, v, GIVES THE PERCENTAGE INCREASE OVER THE VELOCITY OF LIGHT AXIALLY ALONG THE WIRE AXIS.

(8) $$\frac{u}{v} = \eta$$ PERCENT

OR $$u = \eta v$$ INCH PER SECOND.

WHERE

$u < v$, SLOWER THAN LIGHT

$u > v$, FASTER THAN LIGHT

THE NATURAL, OR CHARACTERISTIC, IMPEDANCE OF THE COILED WINDING IN PI OVER TWO OSCILLATION IS GIVEN BY

(9) $$Z_0^2 = \frac{L}{C}$$

$$Z_0 = \sqrt{\frac{L}{C}}$$ OHM

AND RELATING TO THE NUMBER OF TURNS

$$Z_0 = N Z_s$$ OHM

WHERE N IS THE NUMBER OF TURNS

Z IS THE "SHEET IMPEDANCE".

05-27-2012, 03:17 PM

t-rex
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

From Radiotron Designers Handbook

At frequencies between 2 and 6 Mc/s (which are used in receivers providing continuous coverage between the normal short-wave and broadcast bands, or for oscillators giving second harmonic mixing for the short wave band) the progressive universal winding is particularly useful because the winding length of solenoids may be too great with the usual range of former diameters, and the number of turns is insufficient to allow them to be split into many sections.

A set of worm driving gears which spreads the appropriate number of turns over a length approximately equal to the coil diameter is required. This usually necessitates a faster worm drive than is used for broadcast coils. Litz wire with a large number of strands is useful in obtaining maximum Q.

For the 6 to 18 Mc/s band the solenoid is used, almost always with solid wire, and this type of coil remains useful at least up to 100 Mc/s. Increased wire diameters and spacings are used at higher frequencies to obtain form factors similar to those used at low frequencies.

SECTION 2: SELF-CAPACITANCE OF COILS

(i) Effects of self-capacitance (ii) Calculation of self-capacitance of single-layer solenoids (iii) Measurement of self-capacitance.

(i) Effects of self-capacitance

The self-capacitance of a coil is due to the electrostatic coupling between individual turns and between the turns and earth. When the self-capacitance is between uninsulated turns in air its Q may be high, but the greater the amount of dielectric in the field of the coil the greater will be the losses.

Short wave coils of enamelled wire on a solid former do not have serious dielectric losses if good quality materials are used, but at broadcast and intermediate frequencies the universal windings used have comparatively high-loss dielectrics and unless self-capacitances are kept to a low value the reduction in Q may be appreciable.

In the case of coils which are to be tuned over a range of frequencies, comparatively small values of shunt capacitance can have a large effect on the possible tuning range.

In all cases the self capacitance of a coil has an apparent effect on its resistance, inductance and Q, and at frequencies considerably below the self resonant frequency of the coil—

$$\text{apparent inductance} = L\left(1 + \frac{C_0}{C}\right)$$

$$\text{apparent resistance} = R\left(1 + \frac{C_0}{C}\right)^2$$

$$\text{and apparent } Q = Q/\left(1 + \frac{C_0}{C}\right)$$

where C_0 = self-capacitance of coil
C = external capacitance required to tune L to resonance
L = true inductance of coil
R = true resistance of coil
and Q = true Q of coil.

(ii) Calculation of self-capacitance of single-layer solenoids

Until recently the work of Palermo (Ref. C3) had been taken as the standard on the self-capacitance of single-layer solenoids. However Medhurst (Ref. C2) has disputed the theoretical grounds on which Palermo's work is based. As the result of a careful analysis and a large number of measurements he states that the self-capacitance, C_0, of single-layer solenoids with one end earthed, and without leads, is

$$C_0 = HD \ \mu\mu\text{F} \qquad (1)$$

where D = diameter of coil in cm.

and H depends only on the length/diameter ratio of the coil. A table of values of H is given below and when used in conjunction with a capacitance correction for the "live" lead, Medhurst states that the accuracy should be within 5%.

TABLE 1

(iii) Measurement of self-capacitance

A graphical method of self-capacitance determination due to Howe is shown in Fig. 11.2. Values of external capacitance in $\mu\mu$F required to tune a coil to resonance are plotted against $1/f^2$ where f is the resonant frequency in Mc/s. The self-capacitance is the negative intercept of the straight line with the capacitance axis.

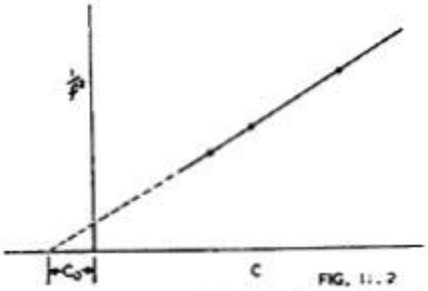

Fig. 11.2. *Construction for determination of self-capacitance of coil.*

An alternative method is to determine the external capacitances necessary to tune a coil to frequencies of f and $2f$. If the two capacitances are C_1 and C_2 respectively,

$$\text{self-capacitance} = \frac{C_1 - 4C_2}{3}.$$

SECTION 3 : INTERMEDIATE-FREQUENCY WINDINGS

(i) Air-cored coils (ii) Iron-cored coils (iii) Expanding selectivity i-f transformers (iv) Calculation of gear ratios for universal coils (v) Miscellaneous considerations.

(i) Air-cored coils

Most commercial receivers have two i-f transformers with four circuits tuned to about 455 Kc/s. To obtain adequate selectivity in such a case the required Q for each winding (mounted on the chassis but not connected in the circuit) is 100 or more.

This Q can readily be obtained with air cored coils of about 1 mH inductance provided that litz wire and comparatively large coil cans are used. Without litz it is difficult to exceed a Q of 50, but even this may be sufficient when more i-f stages than usual are used.

When a single pie is used, Q is to some extent dependent on coil shape but if five times the winding depth plus three times the winding width equals the external diameter, Q will be close to the maximum for the wire and type of winding.

There are three methods of increasing the Q, and when all are used to practical commercial limits, production Q figures of 150 can be maintained. The first requirement is litz wire, probably nine strands for a Q of the order of 100, and twenty or more strands for a Q of 150. Large coil cans are necessary, up to two inches in diameter, with formers of such a size as to make the outside coil diameter little more than half the can diameter. Lastly, self-capacitance must be reduced to a minimum, because its Q is always low and in the case of a single pie winding it may amount to say 25 $\mu\mu$F, a large percentage of the total capacitance. Splitting the winding into pies reduces self-capacitance, at the same time improving the form factor, and a suitable compromise between Q and economy is obtained by winding with three pies. Self-capacitance is also reduced by winding narrow pies, although a limit is set by the difficulty of winding litz wires with a cam of less than 0.1 inch, or perhaps even $\frac{1}{8}$ inch. In addition, as the cam is reduced the height of the coil increases and larger losses from damping of the coil by the can may more than offset the reduced losses in a smaller self-capacitance.

When the windings have more than one pie, the inductance is dependent on pie spacing. For this reason, and to increase the speed of winding, multi-section coils are wound with the former located by a gate, a different slot in the gate being used for each individual pie. Spacing between primary and secondary windings is kept constant by means of a double winding finger which winds primary and secondary at the same time.

06-11-2012, 12:09 AM

dR-Green
Gold Member

Join Date: Apr 2008
Location: Merlin's Fort
Posts: 1,220

Secondary tuning test results. Metallic connection to meter test first.

Using 20 Turn Secondary calculations

http://i168.photobucket.com/albums/u...aryCoil001.jpg

F = 3670000 cycles/sec

Height = 4.08cm
Diameter = 20.7cm
Number of turns = 20
Calculated wire length = 13.07 metres
Wire used after octagon shaping of thin wire around frame = Approx 12.92 metres 26 SWG
Approx 18.9 grams of copper in coil

No variable condenser used with secondary for any of the following tests.

Test Coil = 1 turn 1.5mm wire

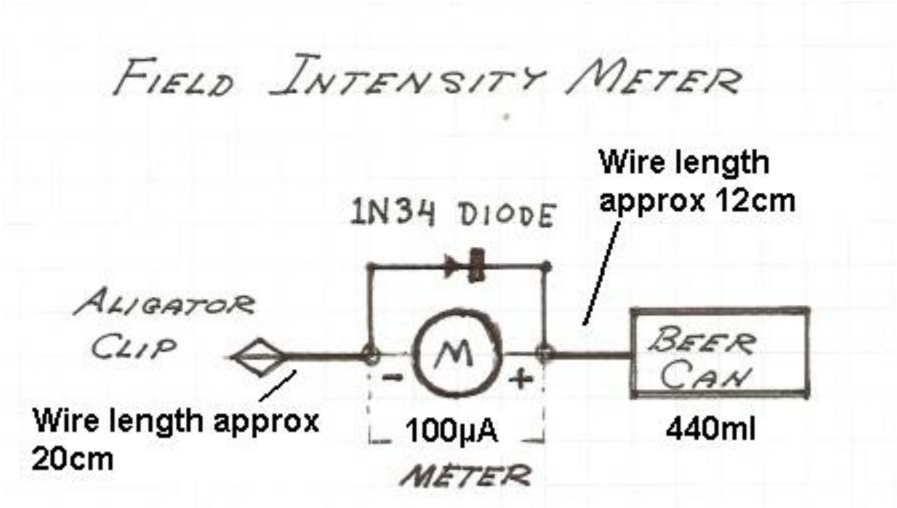
FIELD INTENSITY METER
Wire length approx 12cm
1N34 DIODE
ALIGATOR CLIP
M
-
+
BEER CAN
Wire length approx 20cm
100µA METER
440ml

Setup

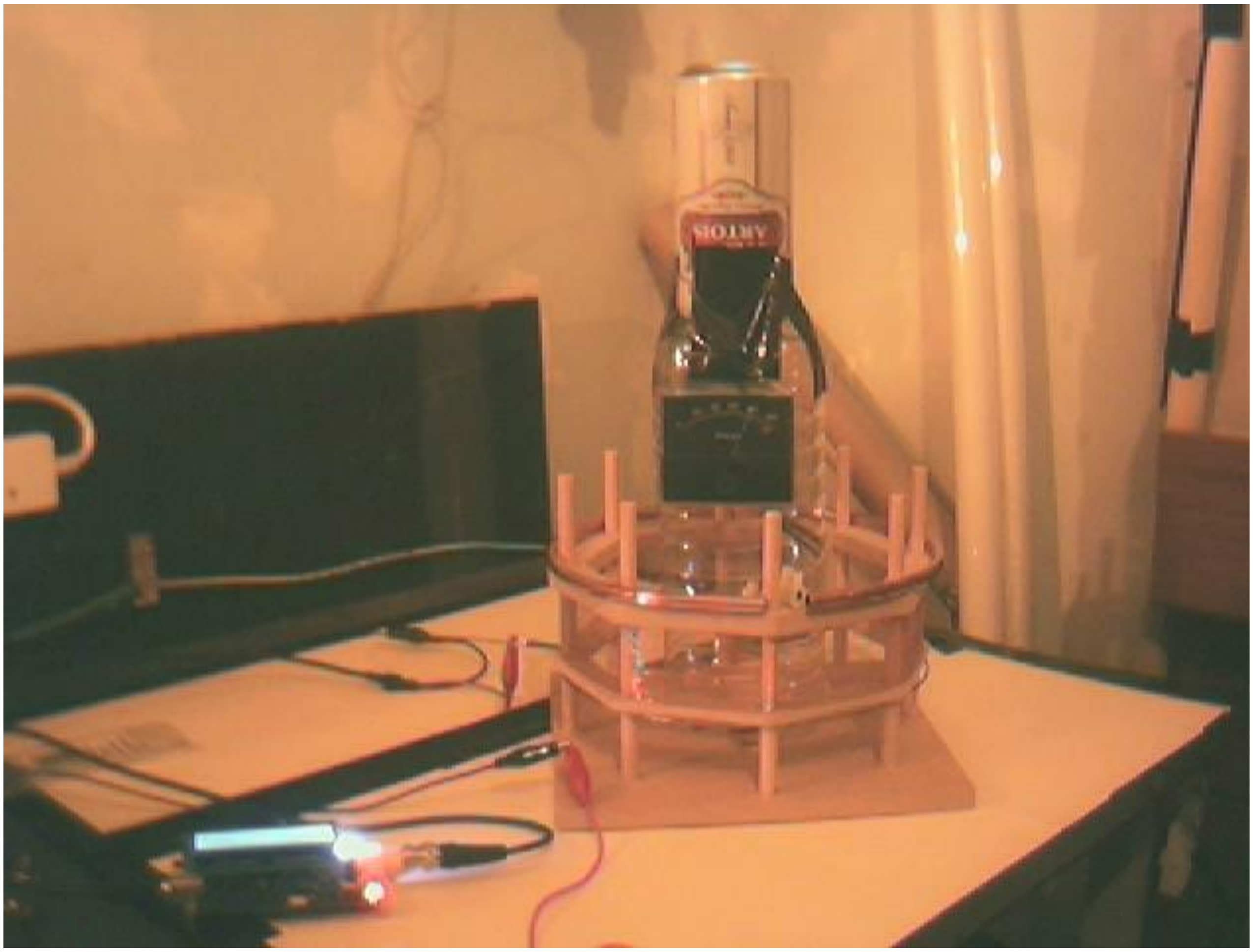

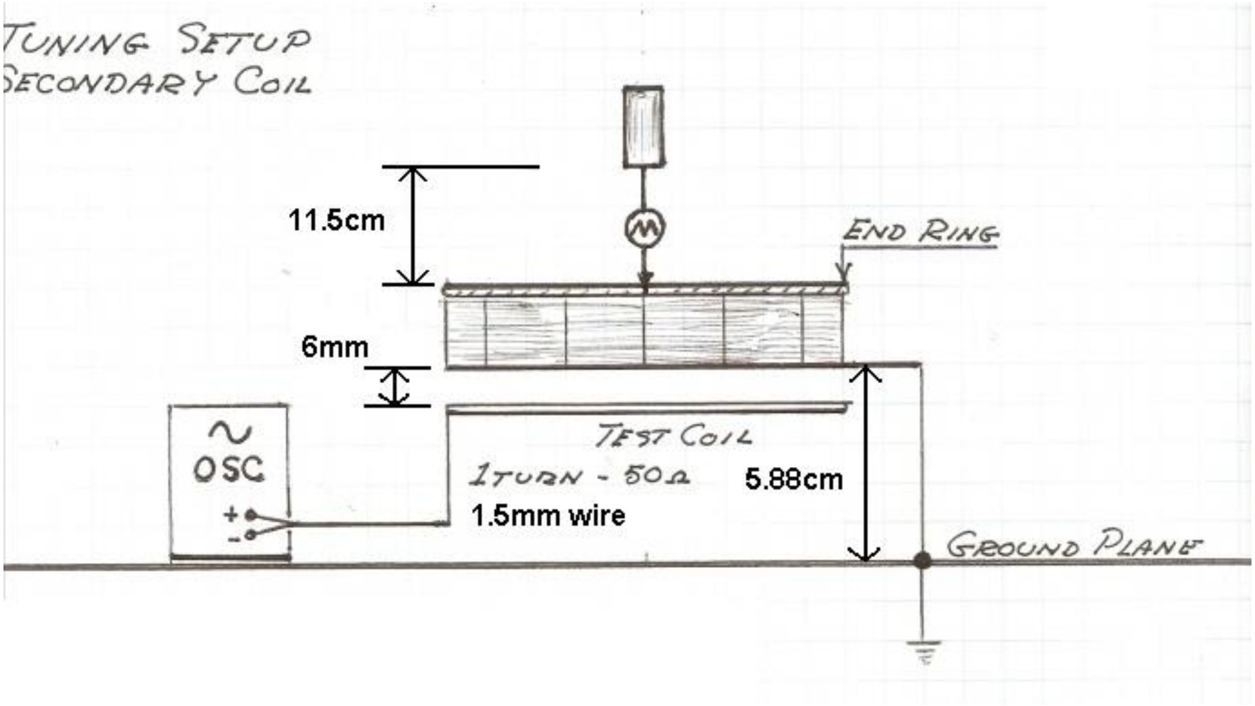
TUNING SETUP
SECONDARY COIL
11.5cm
M
END RING
6mm
OSC
+
-
TEST COIL
1TURN - 50Ω
1.5mm wire
5.88cm
GROUND PLANE

Frequency Sweep

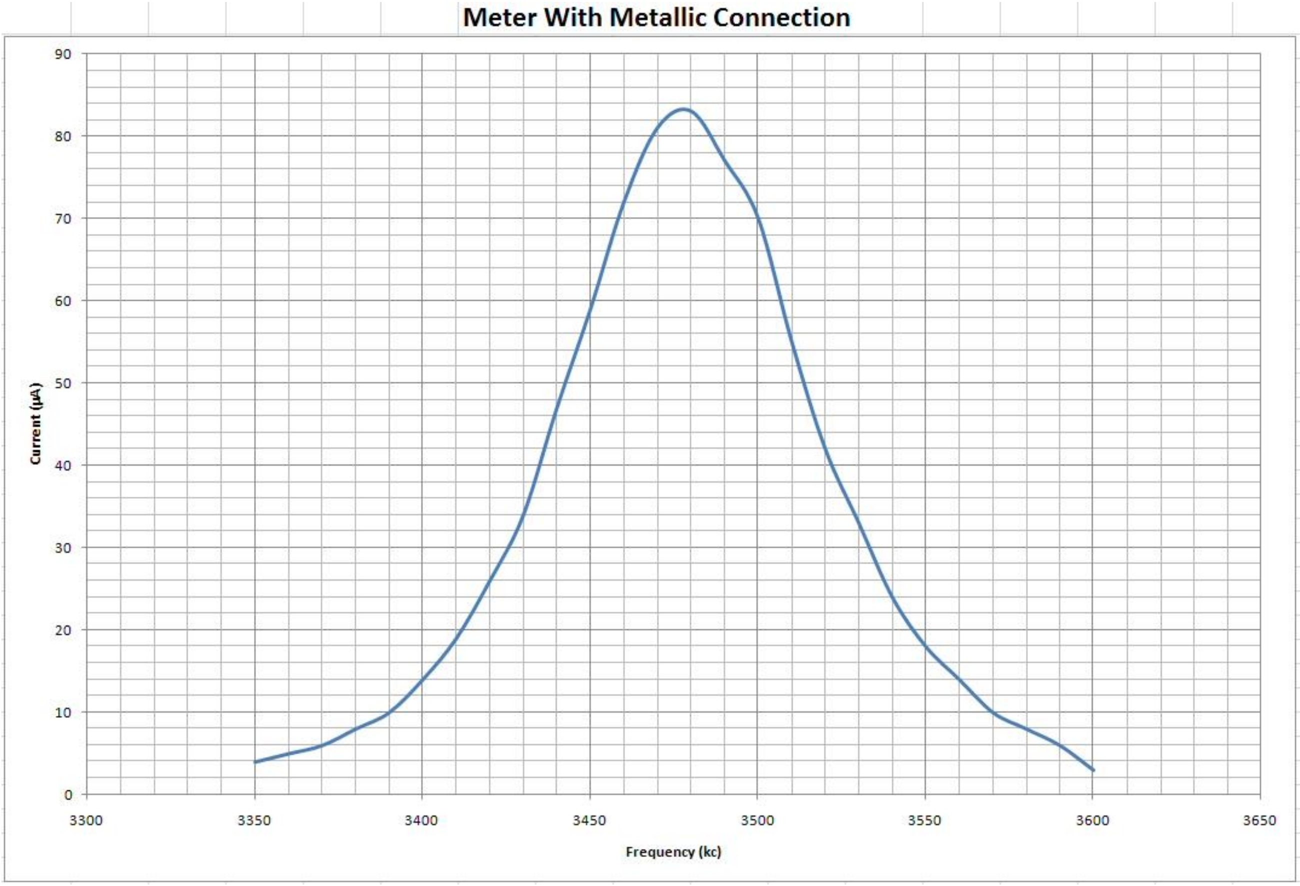

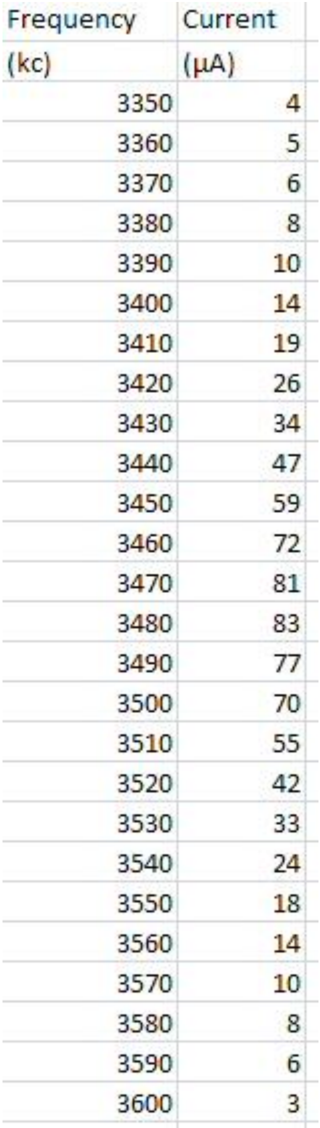

Frequency (kc)	Current (µA)
3350	4
3360	5
3370	6
3380	8
3390	10
3400	14
3410	19
3420	26
3430	34
3440	47
3450	59
3460	72
3470	81
3480	83
3490	77
3500	70
3510	55
3520	42
3530	33
3540	24
3550	18
3560	14
3570	10
3580	8
3590	6
3600	3

Notes
Metallic connection
Wire length from top ring to meter approx 20cm
Wire length from meter to beer can approx 12cm
Beer can = 440ml
Longer signal generator USB (power supply) cable decreases meter reading from 84µA to 80µA

Signal Generator Output
Vrms = 441mV
Vpeak/peak = 1.32V

Coil Peak Output
84µA @ 3478.5 kc

06-11-2012, 12:23 AM

dR-Green

Gold Member

Join Date: Apr 2008

Location: Merlin's Fort

Posts: 1,220

Setup

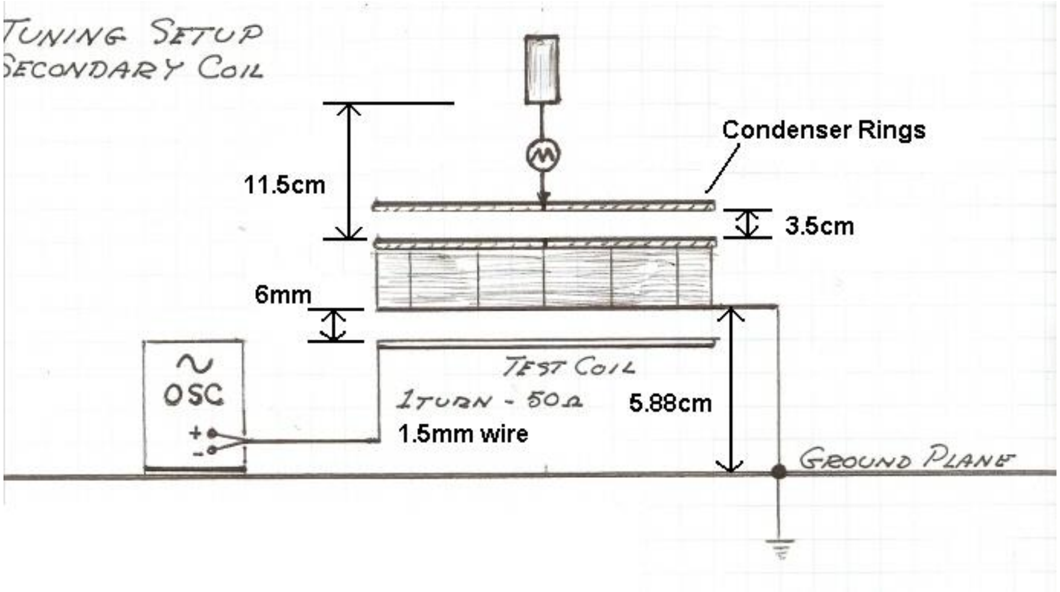
TUNING SETUP
SECONDARY COIL
11.5cm
Condenser Rings
3.5cm
6mm
OSC
+
-
TEST COIL
1 TURN - 50Ω
1.5mm wire
5.88cm
GROUND PLANE

Frequency Sweep

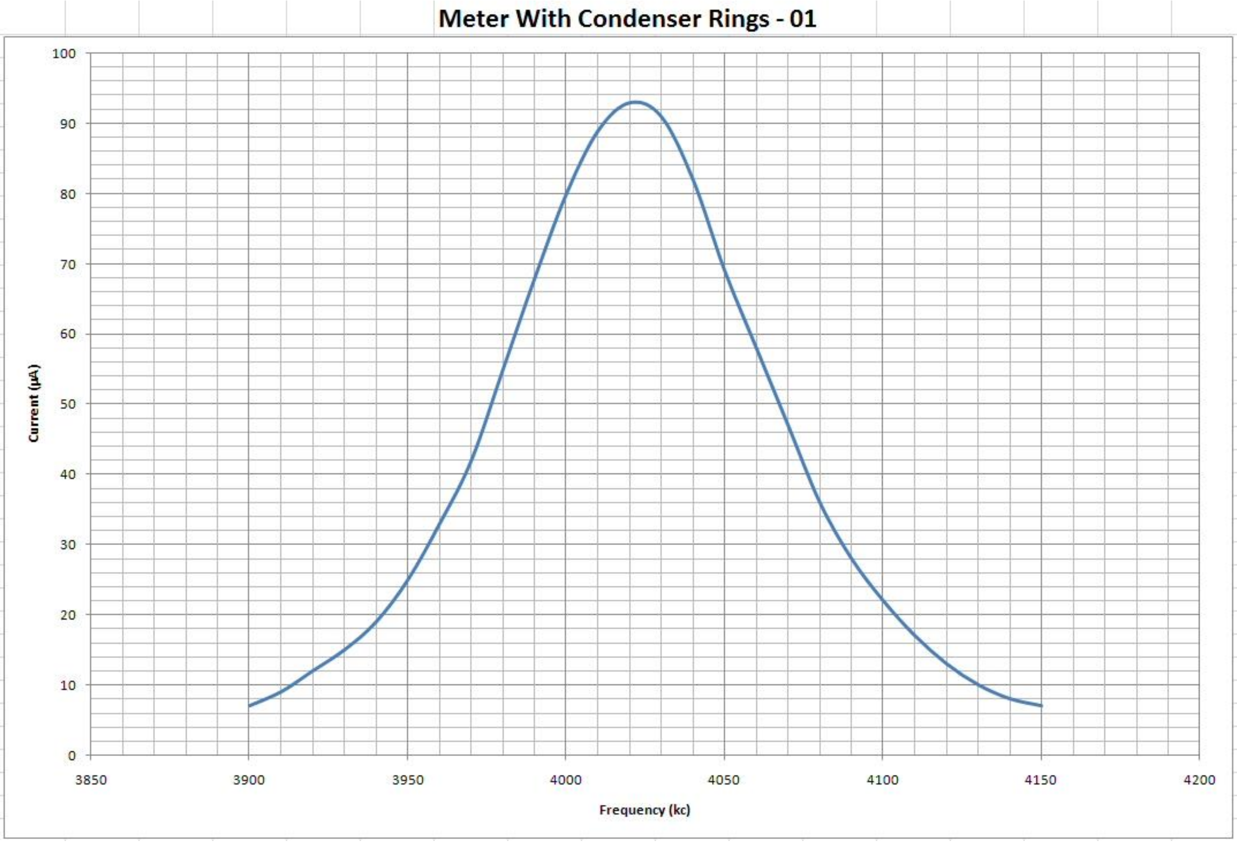
Meter With Condenser Rings - 01
Current (µA)
100
90
80
70
60
50
40
30
20
10
0
3850
3900
3950
4000
4050
4100
4150
4200
Frequency (kc)

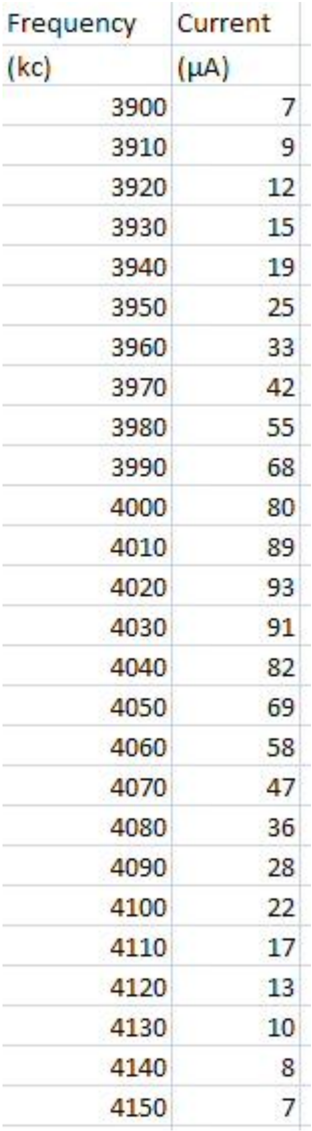

Frequency (kc)	Current (µA)
3900	7
3910	9
3920	12
3930	15
3940	19
3950	25
3960	33
3970	42
3980	55
3990	68
4000	80
4010	89
4020	93
4030	91
4040	82
4050	69
4060	58
4070	47
4080	36
4090	28
4100	22
4110	17
4120	13
4130	10
4140	8
4150	7

Notes

Condenser rings connection
Distance between rings = 3.5cm
Test coil to secondary distance = 6mm
Ground plane to bottom of secondary distance = 5.88cm

Signal Generator Output
Vrms = 1.203V
Vpeak/peak = 3.56V
Coil Peak Output
94µA @ 4020.8 kc

06-11-2012, 12:32 AM

dR-Green
Gold Member

Join Date: Apr 2008
Location: Merlin's Fort
Posts: 1,220

Setup

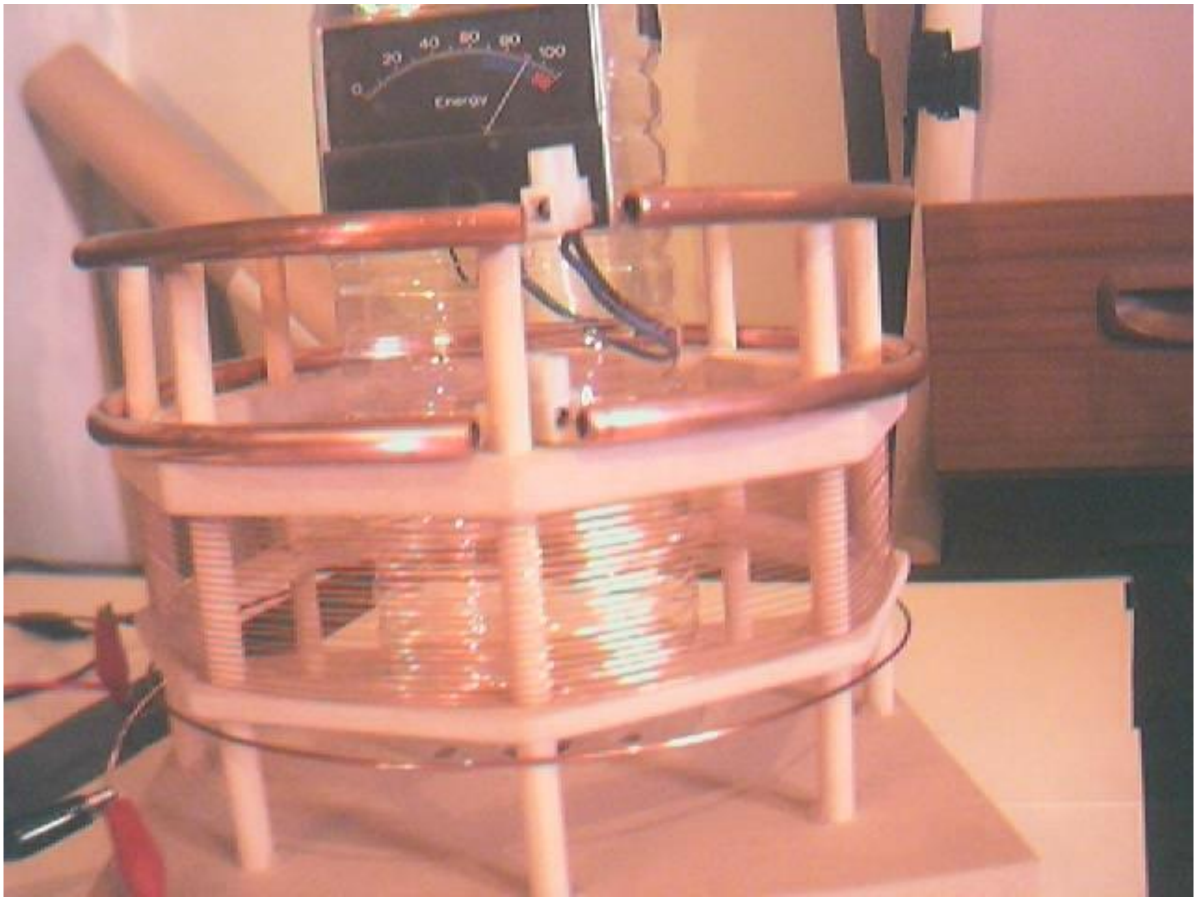

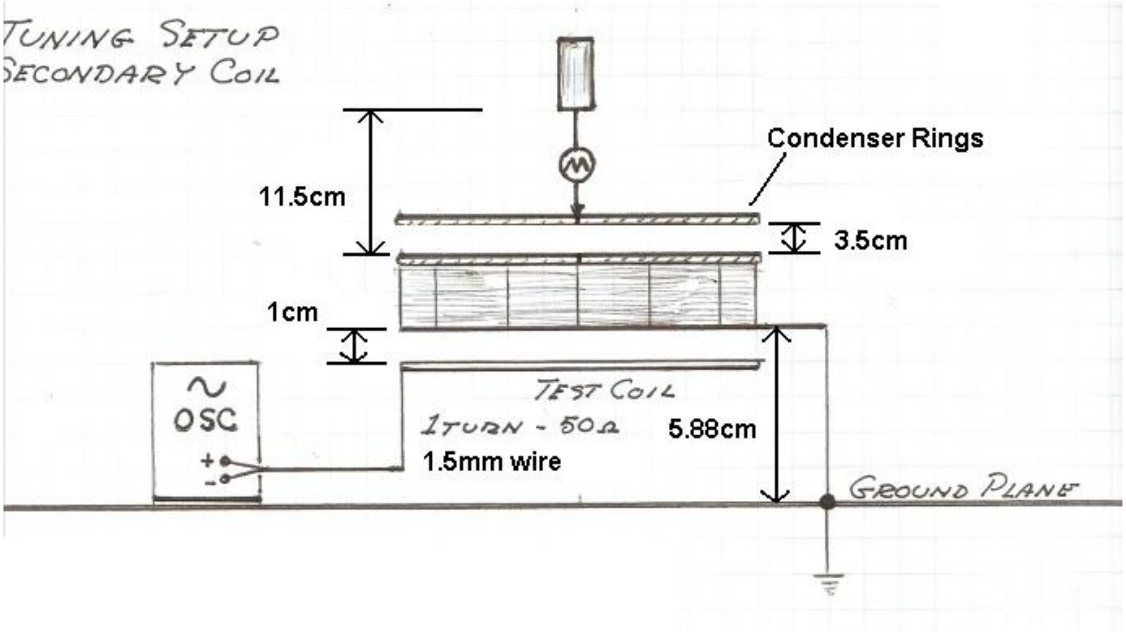
TUNING SETUP
SECONDARY COIL
Condenser Rings
11.5cm
3.5cm
1cm
OSC
+
-
TEST COIL
1TURN - 50Ω
1.5mm wire
5.88cm
GROUND PLANE

Frequency Sweep

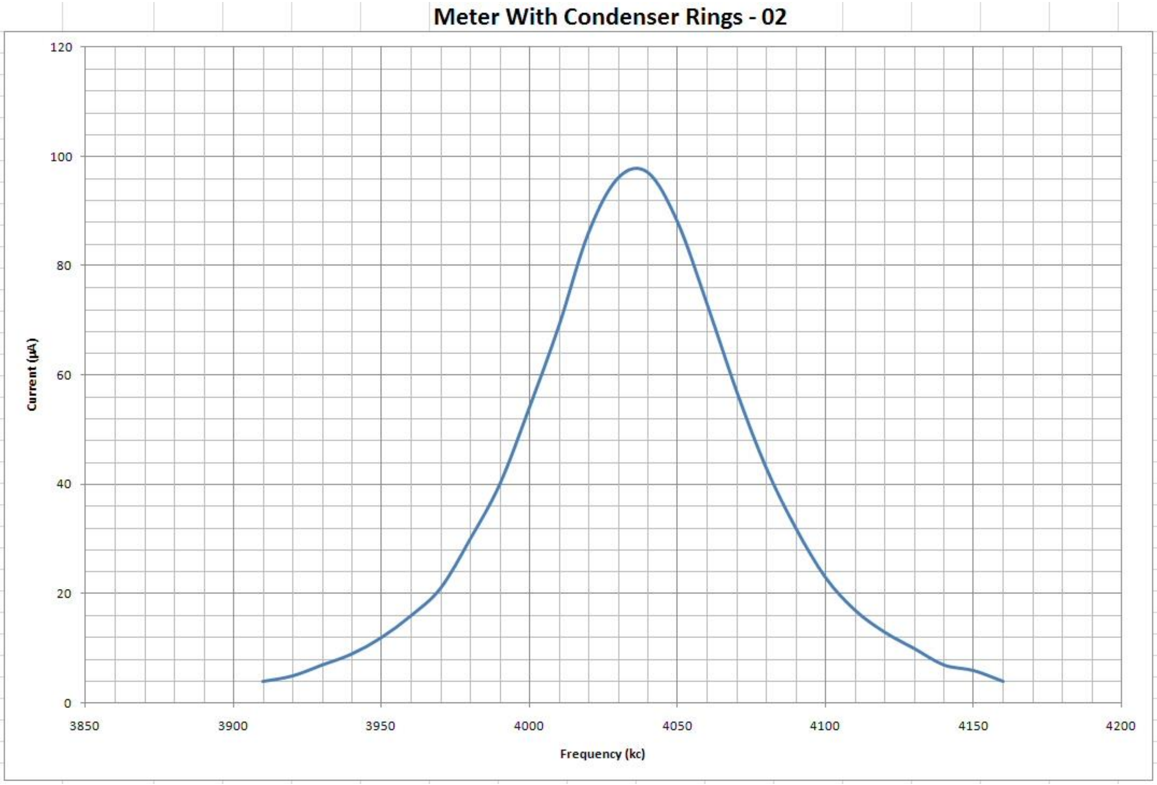

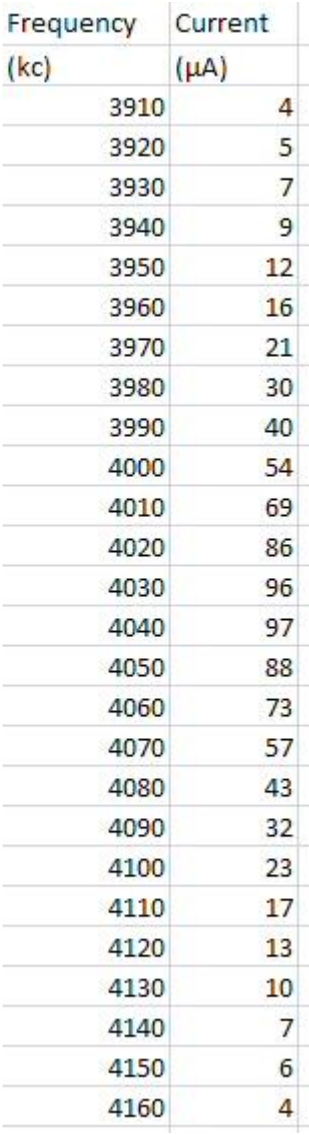

Frequency (kc)	Current (µA)
3910	4
3920	5
3930	7
3940	9
3950	12
3960	16
3970	21
3980	30
3990	40
4000	54
4010	69
4020	86
4030	96
4040	97
4050	88
4060	73
4070	57
4080	43
4090	32
4100	23
4110	17
4120	13
4130	10
4140	7
4150	6
4160	4

Notes

Condenser rings connection
Test coil to secondary distance = 1cm

Signal Generator Output
Vrms = 1.203V
Vpeak/peak = 3.56V

Coil Peak Output
98µA @ 4034.5 kc

Next test will be the extra coil.

06-15-2012, 12:14 AM

dR-Green
Gold Member

Join Date: Apr 2008
Location: Merlin's Fort
Posts: 1,220

After 8 hours or so of work, I've made some changes and did the extra coil tests again. As mentioned above 7.4cm of lead-out wire was removed from the top end of the coil. Beer can distance from extra coil is increased to 28cm. Extra coil raised to 52cm above ground plane. Additional 2.5pF capacitance measured.

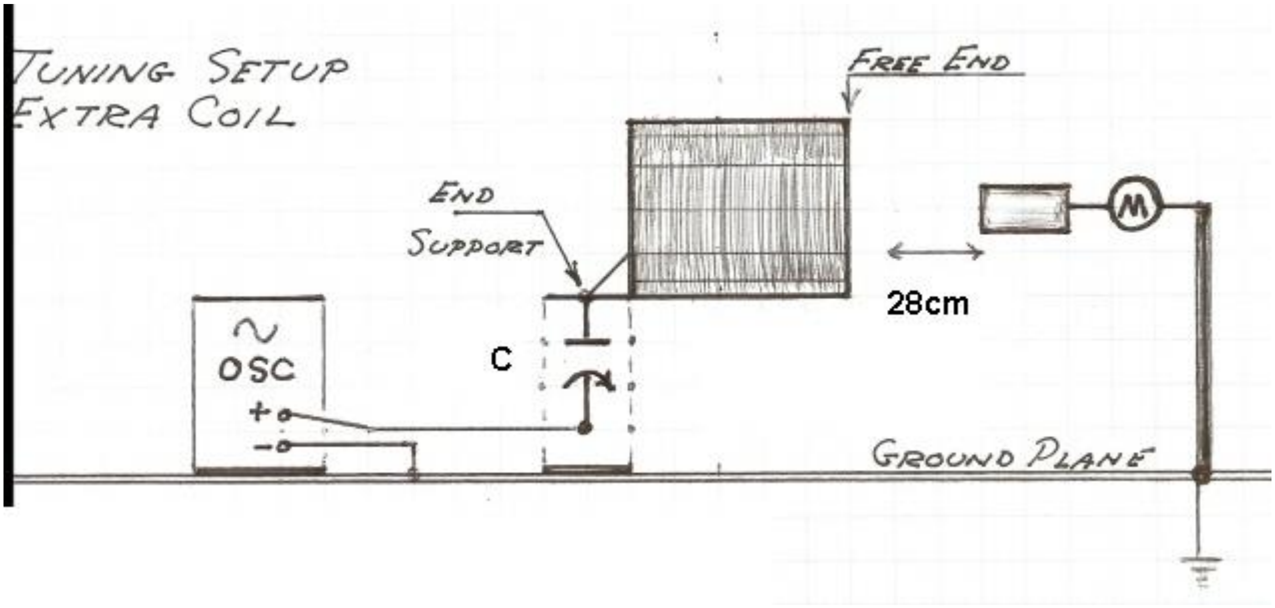

New frequency response with metallic connection

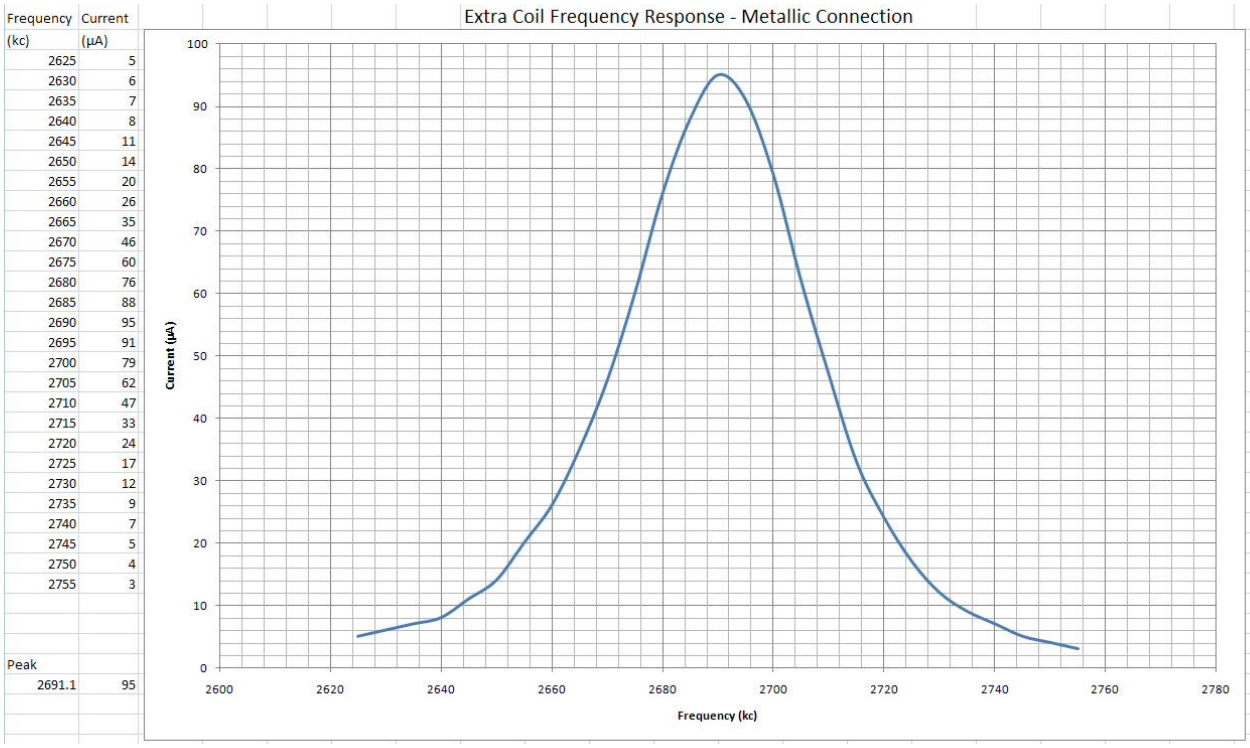

Frequency (kc)	Current (µA)
2625	5
2630	6
2635	7
2640	8
2645	11
2650	14
2655	20
2660	26
2665	35
2670	46
2675	60
2680	76
2685	88
2690	95
2695	91
2700	79
2705	62
2710	47
2715	33
2720	24
2725	17
2730	12
2735	9
2740	7
2745	5
2750	4
2755	3
Peak	
2691.1	95

Extra Coil Frequency Response - Metallic Connection
Current (µA)
0
10
20
30
40
50
60
70
80
90
100
2600
2620
2640
2660
2680
2700
2720
2740
2760
2780
Frequency (kc)

Frequency response with different capacitors (charting not possible with 2.5pF and 28cm beer can distance - meter range too high):

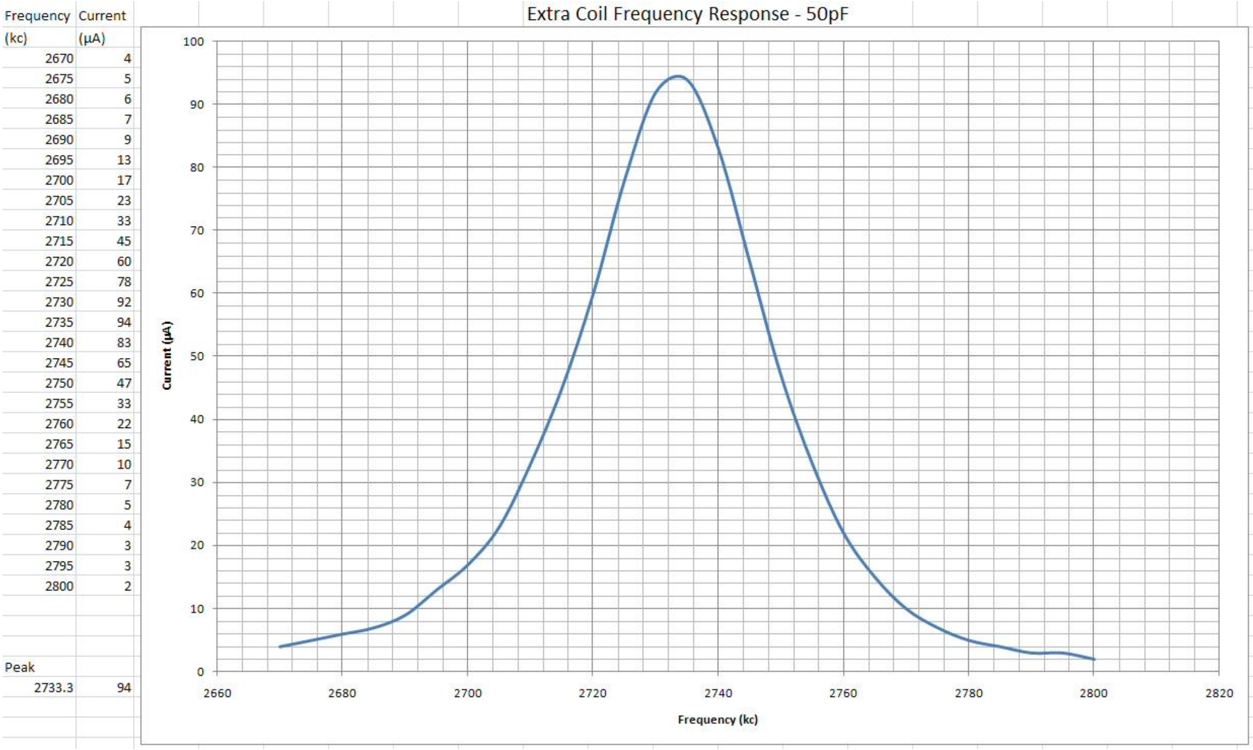

Frequency (kc)	Current (µA)
2670	4
2675	5
2680	6
2685	7
2690	9
2695	13
2700	17
2705	23
2710	33
2715	45
2720	60
2725	78
2730	92
2735	94
2740	83
2745	65
2750	47
2755	33
2760	22
2765	15
2770	10
2775	7
2780	5
2785	4
2790	3
2795	3
2800	2
Peak	
2733.3	94

Extra Coil Frequency Response - 50pF
Current (µA)
0
10
20
30
40
50
60
70
80
90
100
2660
2680
2700
2720
2740
2760
2780
2800
2820
Frequency (kc)

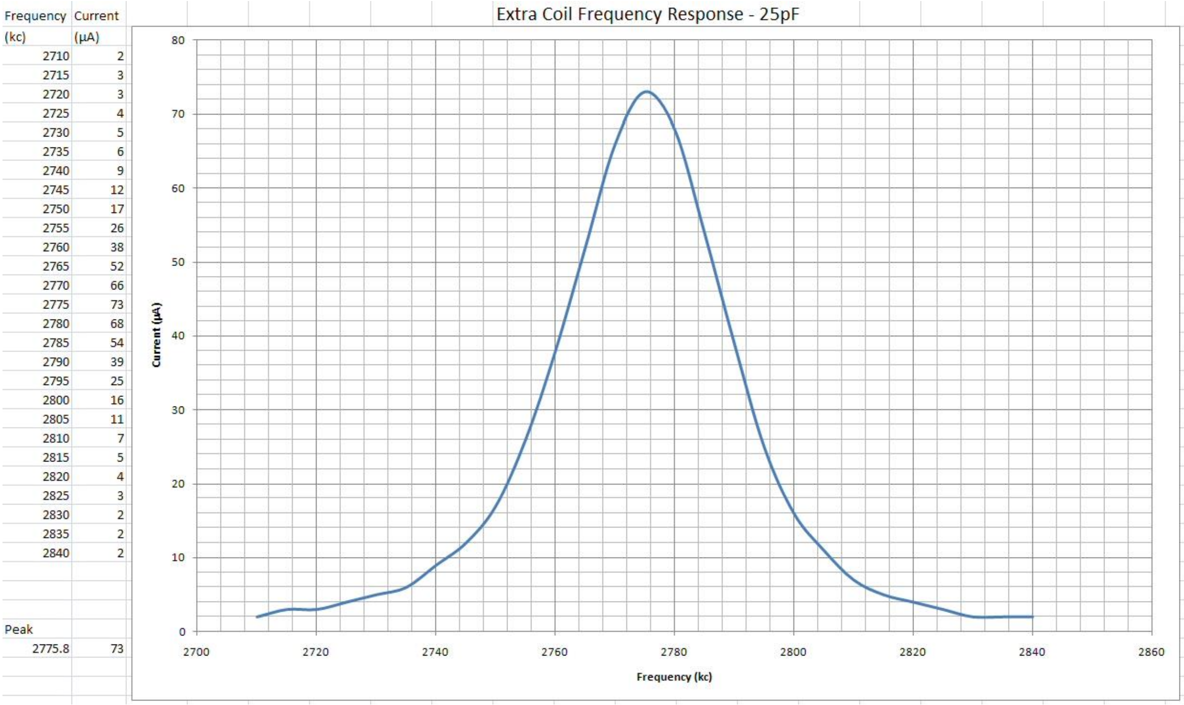

Frequency (kc)	Current (µA)
2710	2
2715	3
2720	3
2725	4
2730	5
2735	6
2740	9
2745	12
2750	17
2755	26
2760	38
2765	52
2770	66
2775	73
2780	68
2785	54
2790	39
2795	25
2800	16
2805	11
2810	7
2815	5
2820	4
2825	3
2830	2
2835	2
2840	2
Peak	
2775.8	73

Extra Coil Frequency Response - 25pF

Current (µA)
0
10
20
30
40
50
60
70
80
Frequency (kc)
2700
2720
2740
2760
2780
2800
2820
2840
2860

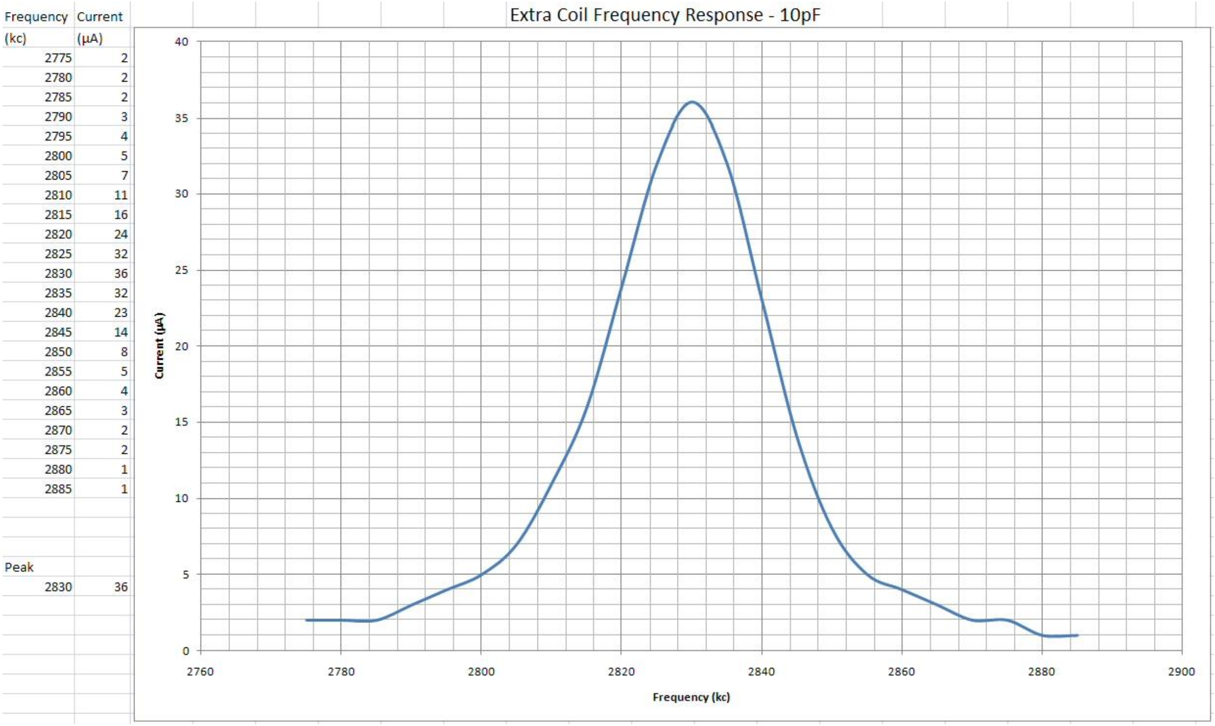

Frequency (kc)	Current (µA)
2775	2
2780	2
2785	2
2790	3
2795	4
2800	5
2805	7
2810	11
2815	16
2820	24
2825	32
2830	36
2835	32
2840	23
2845	14
2850	8
2855	5
2860	4
2865	3
2870	2
2875	2
2880	1
2885	1
Peak	
2830	36

Extra Coil Frequency Response - 10pF

Current (µA)
0
5
10
15
20
25
30
35
40
Frequency (kc)
2760
2780
2800
2820
2840
2860
2880
2900

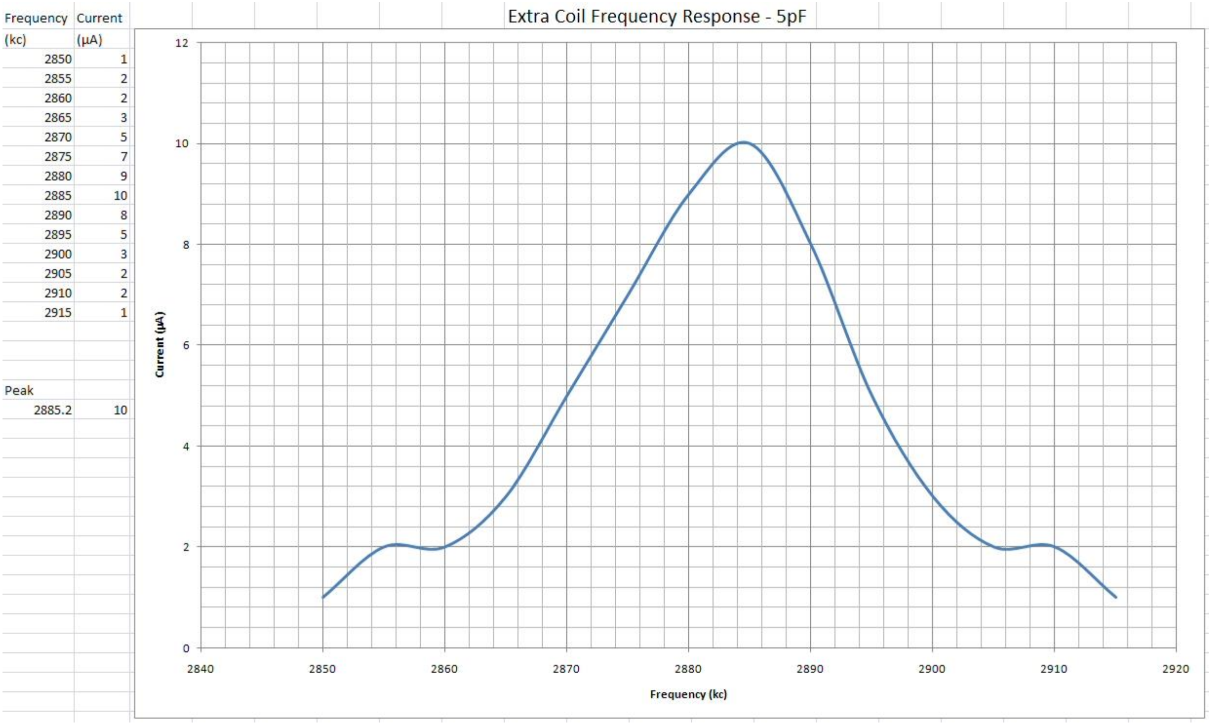

Frequency (kc)	Current (µA)
2850	1
2855	2
2860	2
2865	3
2870	5
2875	7
2880	9
2885	10
2890	8
2895	5
2900	3
2905	2
2910	2
2915	1
Peak	
2885.2	10

New frequency vs capacitance graph

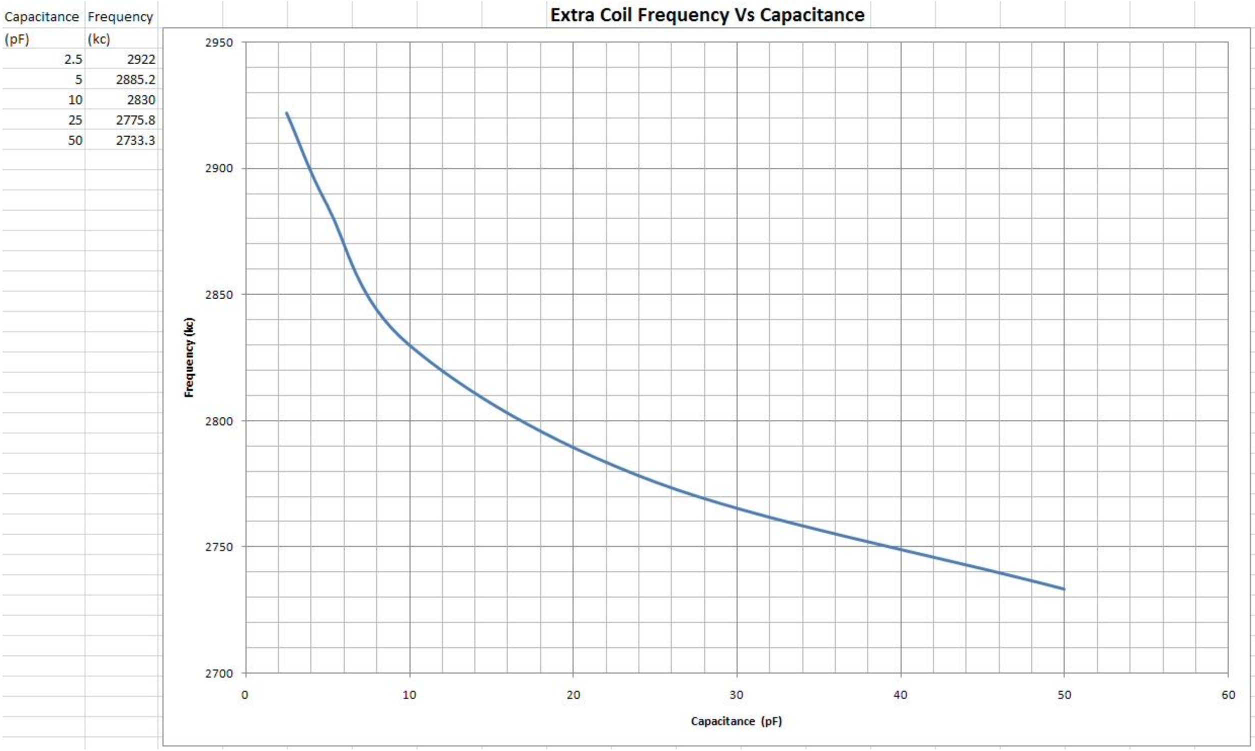

Capacitance (pF)	Frequency (kc)
2.5	2922
5	2885.2
10	2830
25	2775.8
50	2733.3

Given

Quote:

> Originally Posted by **T-rex**
> *1) Extra Coil Luminal Velocity*
> *2.31 Mc/sec*

2) Extra Coil Maximum Velocity Frequency
2.922 Mc/sec

3) Percent Operating Velocity to Luminal Velocity
126 Percent

06-15-2012, 12:00 PM

t-rex
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Dr. Green #396 (I)

Consideration of Coil Self Capacitance
1) Extra Coil Diameter
8.25 cm

2) Capacitance factor from Radiotron table for a coil aspect ratio of 100%
.46 Numeric

3) Multiply this by the coil diameter
3.8 picofarads

4) This is the "free space" self capacitance of this extra coil. This means a true air wound coil.

Now to determine the actual working capacitance of this extra coil resulting from the coil form and stray lead parasitic capacitance.

1) The theoretical velocity factor from the Wireless Power Table for a coil with a 100% aspect ratio:
187% numeric

2) The actual measured velocity factor
124% numeric

3) Taking the ratio of the free space to actual measured velocity gives
1.51 numeric

Note, close to 1.54, pi/2

4) Velocity varies as the inverse of the square root of the capacitance
2.28 numeric

5) Multiplication of the free space capacitance by this squared factor
8.7 picofarads

6) This is the burdened coil capacitance resulting from the coil form and connecting leads.

Observe that the 45 degree knee or asymptote point is this derived value of capacitance. This effective coil capacitance is the square root of its self capacitance, C and its mutual elastance, K.

Dr Green #396 (II)

Considerations of the magnification factor as derived from the bandwidth of the resonance curve, this the half power, or 71% voltage points on the curve.

1) Determine the 71% points on the extra coil curve (for 10 pFd).

Lower Sideband Point
2821 Kilocycles/sec

Upper Sideband Point
2831 Kilocycles/sec

Hence the half power bandpass is 17 Kilocycles or 6%

2) Two factors are here derived. Factor d is the ratio of the bandwidth to the center frequency, 2830 Kc/sec.
0.006

Factor n is the inverse of factor d:
167

3) Here derived is the magnification factor of this extra coil,
n = 167 numeric

General Conclusions

This extra coil is now quantified as to its general electrical behavior:

v = 124% luminal velocity

C = 8.7 pico
Farad
n = x 167 magnification factor

73 DE N6KPH

06-15-2012, 04:39 PM

t-rex ●
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Dr Green Extra Coil

The inductance, characteristic impedance and resonant frequency of this extra coil is now developed.

1) The magnetic inductance of the extra coil by Wheeler's formula (Theory of Wireless Power) is 846 microHenry

2) The effective inductance for a cosine quarter wavelength current distribution is

by Steinmetz, two over Pi the total inductance,
539 microHenry

By Miller, one half the total inductance,
423 microHenry

3) The equivalent capacitance of the extra coil is given as
8.7 picoFarad

4) The effective capacitance for a sine quarter wave distribution is

by Steinmetz, two over Pi times C,
5.5 picoFarads

by Miller, eight over Pi squared times C,
7.1 picoFarads

4)The Characteristic coil impedance is defined as the square root of the ratio of inductance to capacitance

By Steinmetz

9.9 Kilo Ohm

by Miller
7.7 Kilo Ohm

5) The angular frequency is defined as the inverse of the square root of the product of inductance and capacitance

18 Kilo-radians per second
or
2860 Kilocycles/sec

This concludes the calculation of the important characteristics of the extra coil under study

73 DE N6KPH

06-16-2012, 01:56 PM

t-rex ●
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Nhopa #704 (I)

Basic Secondary Coil Characteristics:

1) From Table 2 of the Theory of Wireless Power, for an aspect ratio of 10 percent, the free space coil velocity effective is
78%

2) The actual frequency is given as
1188 Kilocycles/sec

The luminal frequency is derived as
1856 Kilocycles/sec

And hence the actual velocity effective is given
64% luminal velocity

Note, 63.7% is one over the value of Pi over two

3) From table 1 of the Theory of Wireless Power for a 62 cm diameter coil is given as
60 picoFarads

This is the free space capacitance of the coil.

The actual capacitance is derived from the square of the ratio of free space to actual velocities of this secondary coil.

Ratio of Velocities
126%

Squared giving
1.58 numeric

Capacitance, actual
95 picoFarads

4) Tabulating this data

Luminal velocity
100%

Free space velocity
78%

Actual Velocity
64%

Free Space Capacitance
60 picoFarads

Actual Capacitance 95 picoFarads

Conclusion, coil form and stray lead capacitance add 35 picoFarad of capacitance burden.

Nhopa #704 (II)

The natural, or characteristic impedance, of this secondary coil as a transmission network is derived.

1) By Wheelers formula in the Theory of Wireless Power the total magnetic inductance is given as
444 microHenries

The actual capacitance been given as
95 picoFarads

2) Determining the effective values of inductance and capacitance resulting from the Cosine and Sine quarter wave space distributions the following are derived

By Steinmetz method it is, the effective inductance
283 microHenry

The effective capacitance
60.5 picoFarad

And the effective values from the Miller method it is, the effective inductance
222 microHenry

The effective capacitance
77.1 picoFarad

3) The natural impedance is defined as the square root of the ratio of the inductance to the capacitance, giving

Steinmetz, 2200 Ohm

Miller, 1700 Ohm

4) Hence the primary transmission factors are determined for the secondary coil

v, propagation constant
64%

Z, the natural impedance
1700, or 2200 Ohm
1950 Ohm averaged

Nhopa #704 (III)

It seems that your test coil is too close to secondary, back it away instead of using a ballast resistor. Then look for distance to give the sharpest curve. The magnification factor is the inverse of the bandwidth between the 70.7% down points on each side of the curve.

ALSO NOTE EVERYONE!

I cannot see attachments in coffee shop, I cannot! Post full size pictures then I can respond. Remember also that this provides visual entertainment to the person who is working his/her computer. Remember the coffee shop customer is often a cowboy, or a foreign hiker from France or Italy, they think a Volt is a car by Chevy.

Finally from the Geometric Algebra, Dr Green, and Nhopa material the secondary and extra coils are quantified. It now is a simple matter for any competent person to design a coil for any desired frequency. It is now engineer-able.

73 DE N6KPH

06-17-2012, 02:44 AM

dR-Green
Gold Member

Join Date: Apr 2008
Location: Merlin's Fort
Posts: 1,220

Quote:

> Originally Posted by **T-rex**
> *This concludes the calculation of the important characteristics of the extra coil under study*

Thank you Eric. I've just completed all the necessary secondary tests, the best I'll be able to do without rewinding it with thicker wire, and/or making a new frame. The preferred plan in that case is to make a bigger coil.

I'm now using a different measuring method so I can see the readings and change the frequency on the computer without getting close and affecting things:

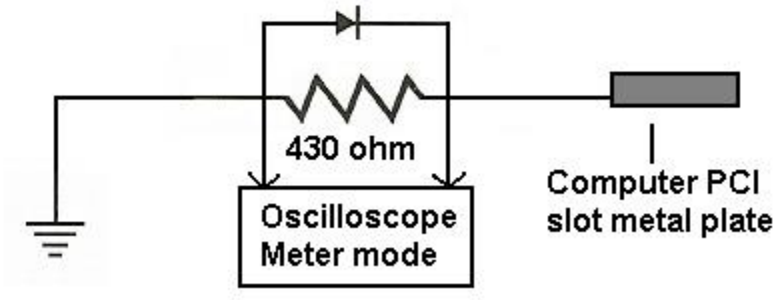

As arranged

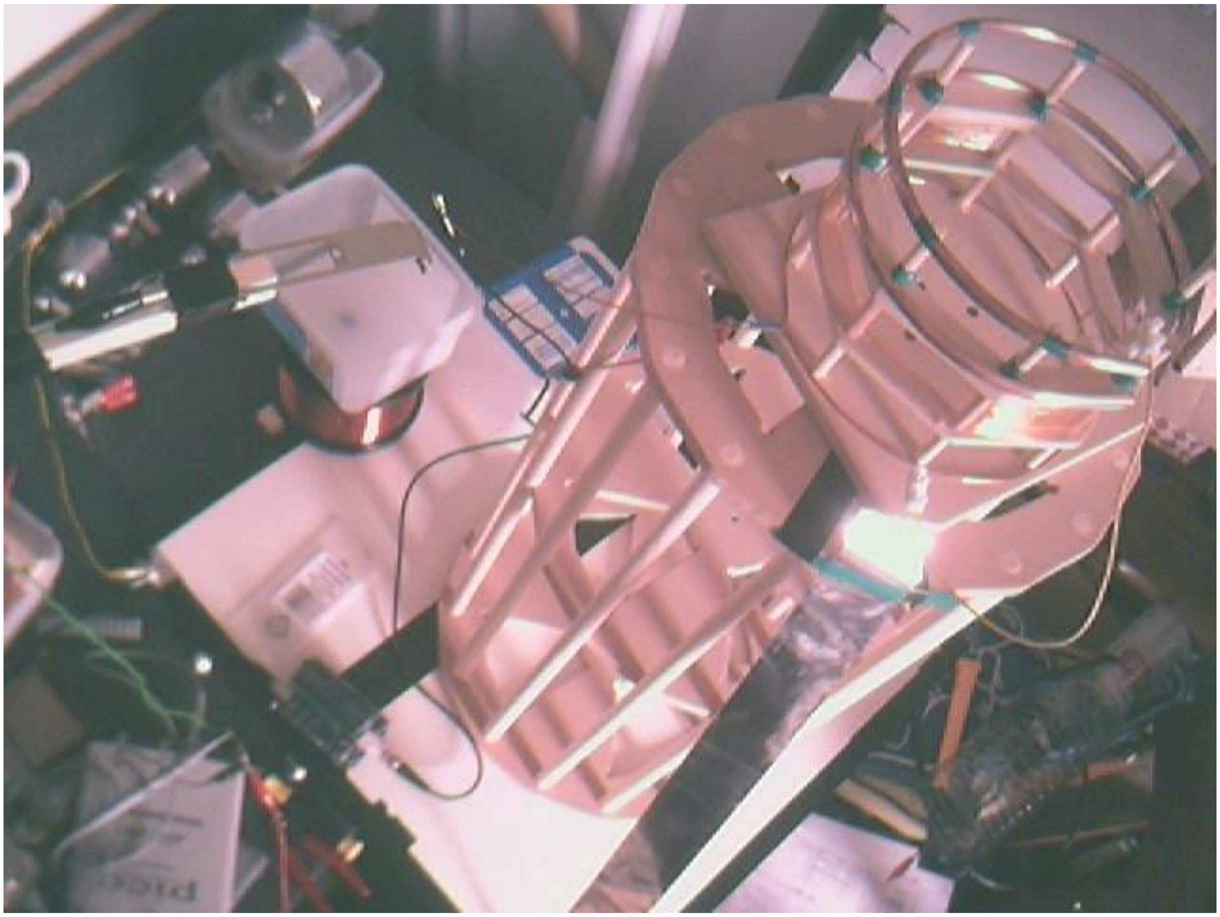

The frequency response graphs should now be within 0.3mV accuracy at each sample point and the written peak frequency values within 300 cycles/sec accuracy.

Used the condenser rings as suggested - upper ring connected to secondary neutral. The frame is only just big enough to set the rings at the correct distance. I think ideally the frame should be a little bigger.

The "final" results are as follows:

Secondary without condenser rings peak: 4126 kc
Secondary with condenser rings peak: 3670 kc
Extra Coil with 10pF peak: 2833.1 kc
Extra Coil with metallic connection peak: 2694.8 kc

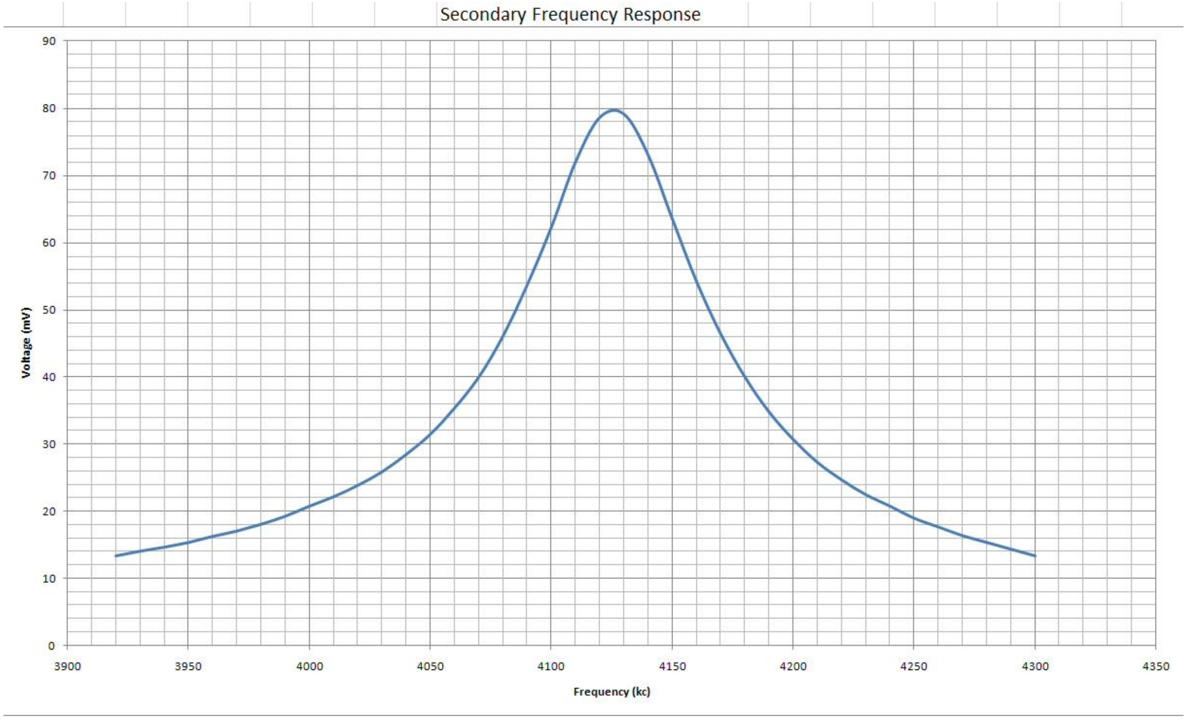
Secondary Frequency Response
Voltage (mV)
0
10
20
30
40
50
60
70
80
90
3900
3950
4000
4050
4100
4150
4200
4250
4300
4350
Frequency (kc)

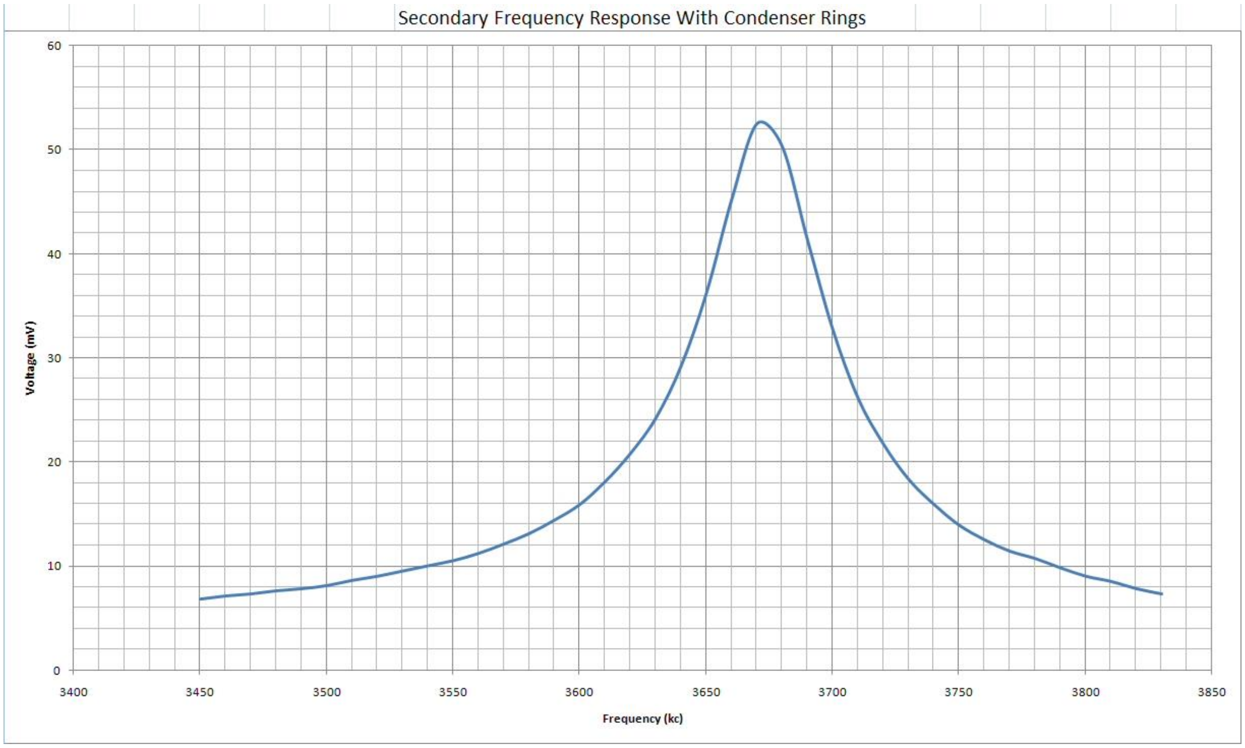
Secondary Frequency Response With Condenser Rings
Voltage (mV)
0
10
20
30
40
50
60
3400
3450
3500
3550
3600
3650
3700
3750
3800
3850
Frequency (kc)

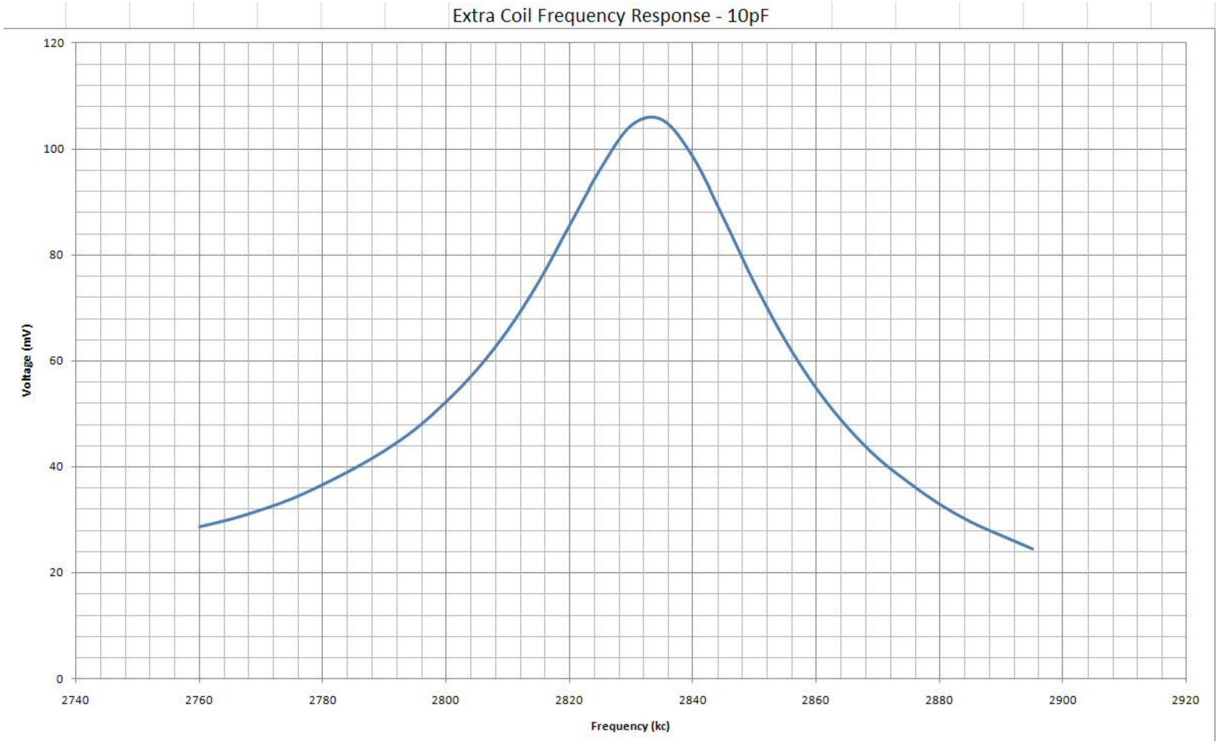
Extra Coil Frequency Response - 10pF
Voltage (mV)
0
20
40
60
80
100
120
Frequency (kc)
2740
2760
2780
2800
2820
2840
2860
2880
2900
2920

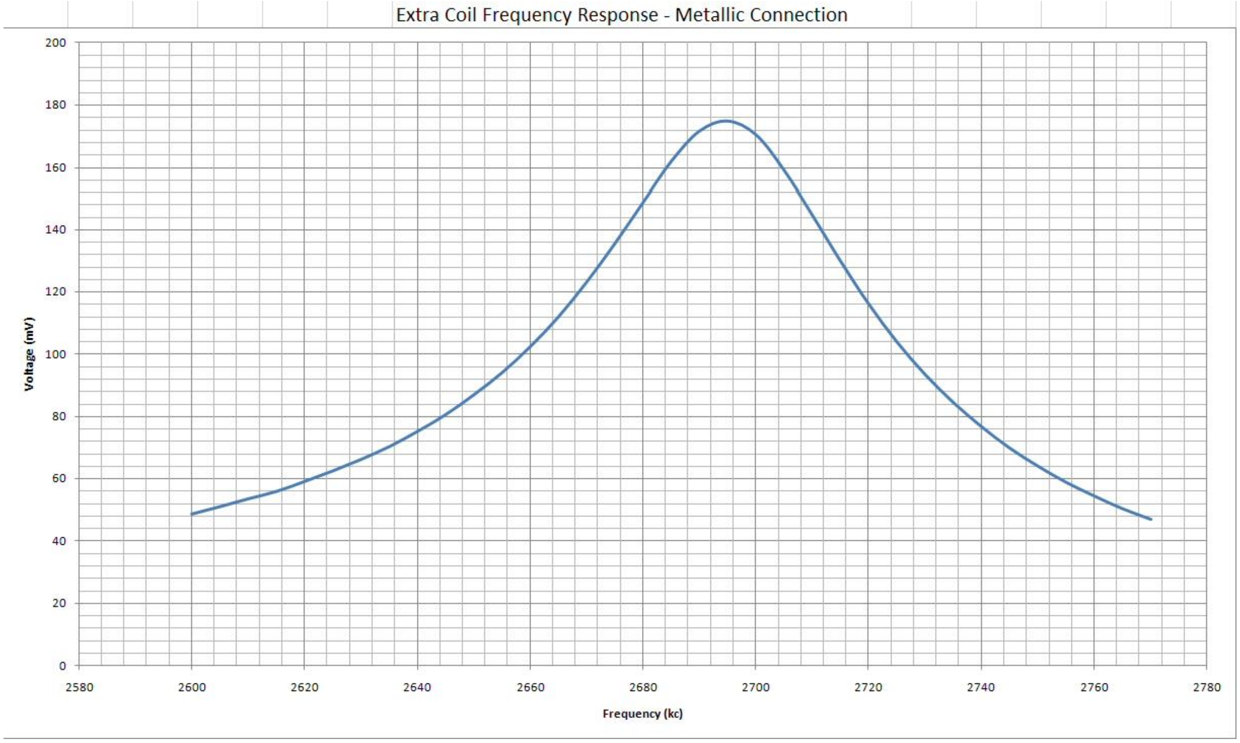
Extra Coil Frequency Response - Metallic Connection
Voltage (mV)
0
20
40
60
80
100
120
140
160
180
200
Frequency (kc)
2580
2600
2620
2640
2660
2680
2700
2720
2740
2760
2780

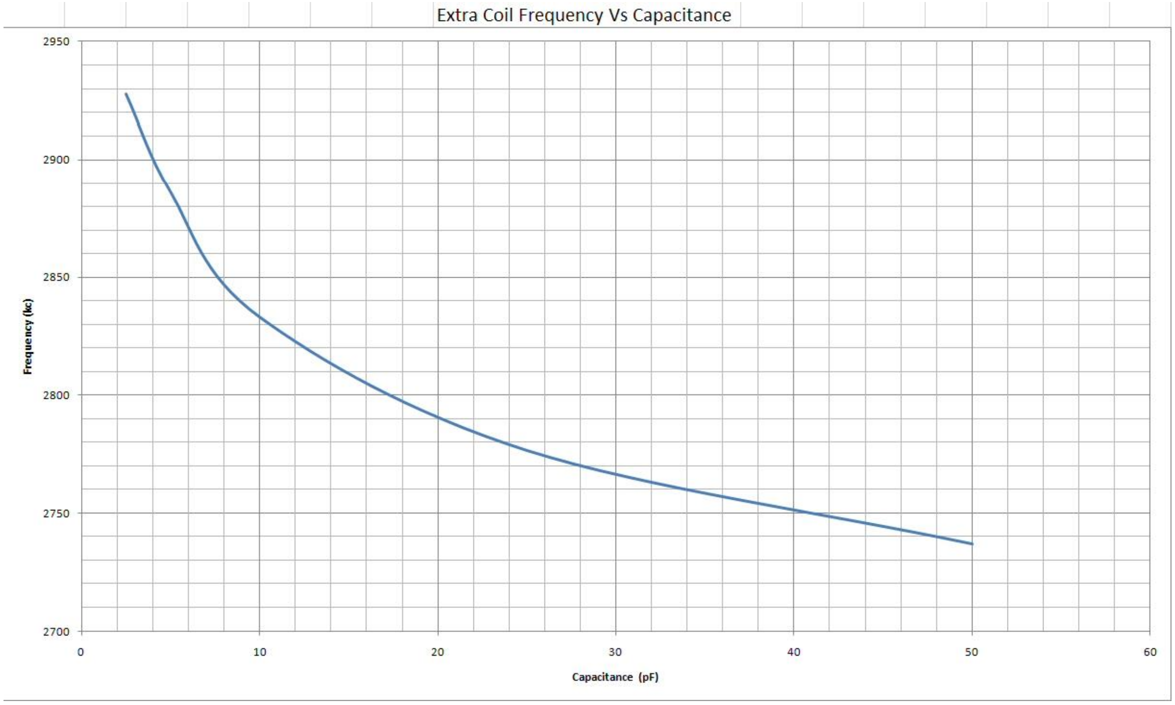

Frequency Vs Capacitance sample points:

2.5pF 2927.6 kc
5pF 2886.1 kc
10pF 2833.1 kc
25pF 2776.6 kc
50pF 2737 kc

06-19-2012, 01:33 PM

t-rex
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Extra Coil Calculations and Experiments

1) Tesla, Colorado Springs Notes. Final Extra Coil is found on page 318;
Diameter, 8.4 feet
Length, 8.0 feet
Number of turns, 100 numeric

Now, convert this to meters;
Diameter, 2.55 meter
Length, 2.42 meter

To find the total wire length, mean length of turn multiplied by number of turns;
Coil Circumference, 8.0 meter
Total Wire Length, 800 meter

This establishes the luminal quarter wavelength. The whole wave is hence:
Lambda equals 3200 meters

Dividing this into the velocity of light in meters per second gives
94 Kilocycles per second

This result is the luminal resonant frequency of an equivalent TEM line. We will call this frequency 100 percent luminal velocity.

2) Tesla, Colorado SPrings Notes, page 357, Extra Coil Grounded at neutral end, lamp as capacity terminal. Frequency of resonance is given as:
93.7 Kilocycles/sec
Near 100 percent luminal

3) Tesla, Colorado Springs Notes, page 363, extra coil free with a 38cm end capacitance. The resonant frequency is given as
116.3 Kilocycles/sec
near 124% Luminal

4) Here given by Tesla for the least coupling to the extra coil is 124% luminal velocity. For a "square" coil the free space velocity is 187% the velocity of light. The measured reduction, or retardation, is a result of the dielectric material which supports the coiled windings. The value 124% is seen to agree here in the experiments given by:
1) Geometric Algebra
2) Dr. Green
3) Nhopa
4) Nikola Tesla

All have been burdened to the same extent by dielectric influence and stray lead capacitance.

The challenge here, as was faced by Tesla, is how to make this "self-capacitance" even smaller. Every connecting lead and nearby insulator burdens the coil. The coil form itself is the principle culprit. Is it possible to free the coil to 157%? I did it once with a "PIDUX"air wound coil held together with 6 plastic rods, no more. In conclusion we have now an experimentally verified propagation constant of 124%, the rest is engineering.

73 DE N6KPH

06-24-2012, 07:27 PM

dR-Green
Gold Member

Join Date: Apr 2008
Location: Merlin's Fort
Posts: 1,220

This was the setup in the previous test, condenser rings capacitance measured at 11.8pF to tune the secondary to 3670 kc.

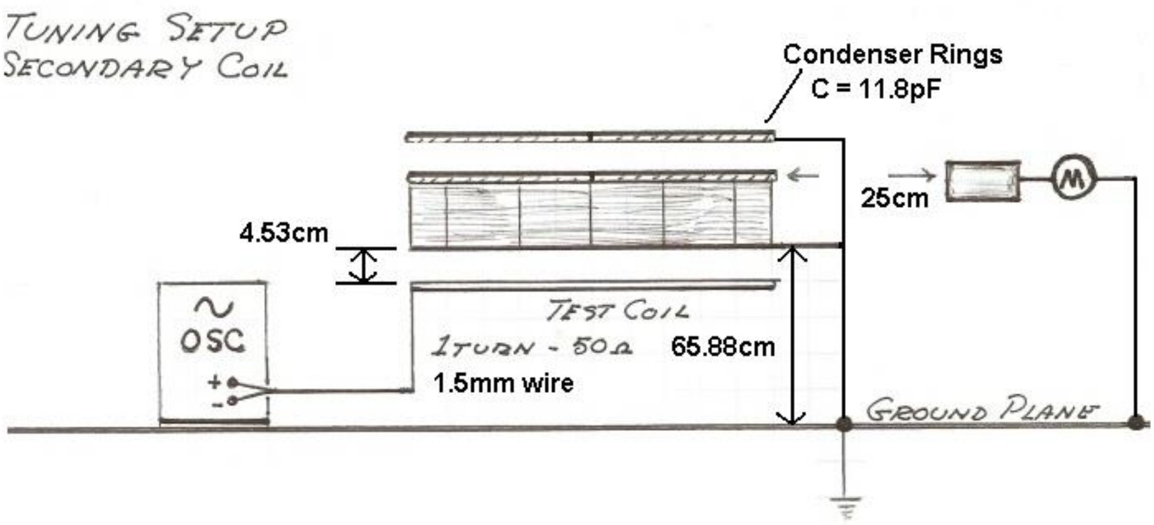
TUNING SETUP
SECONDARY COIL
Condenser Rings
C = 11.8pF
25cm
M
4.53cm
OSC
+
-
TEST COIL
1TURN - 50Ω
65.88cm
1.5mm wire
GROUND PLANE

With the extra coil, the required capacitance is measured at 12.59pF. Set up as so:

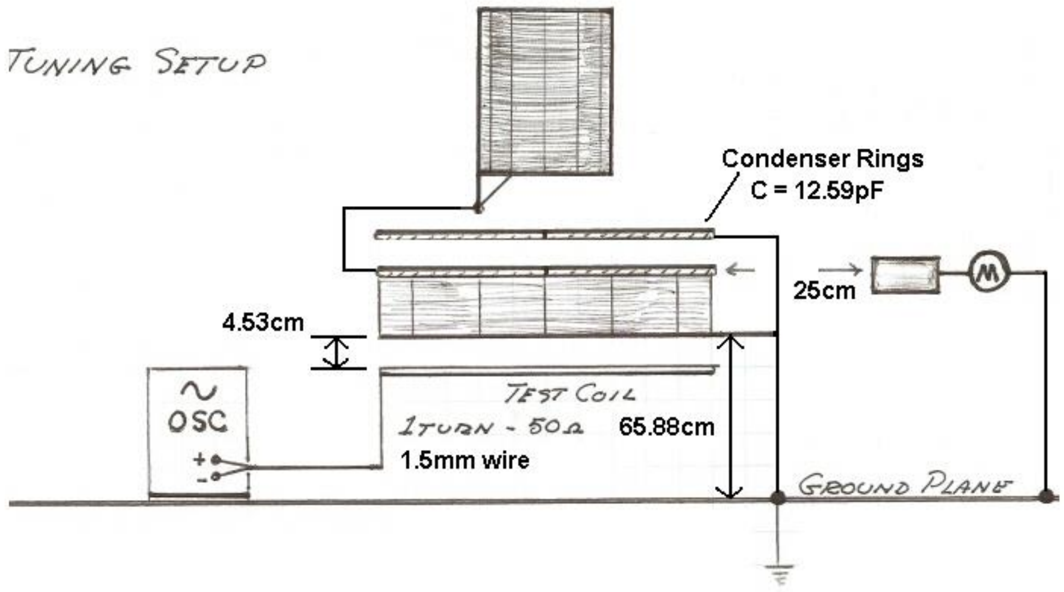
TUNING SETUP
Condenser Rings
C = 12.59pF
25cm
4.53cm
OSC
+
-
TEST COIL
1TURN - 50Ω
65.88cm
1.5mm wire
M
GROUND PLANE

Two resonant peaks found with the rings set at 12.59pF, 3670 kc reading 22.7mV, and 2344.2 kc reading 41.3mV.

(The previously mentioned "unreliable" test with the AV plug seemed to show the opposite in terms of light output vs measured voltage).

Assuming I calculated it right, taking 70.7% from post #706:

Magnification Factor @ 3670 kc = 90.20524
Magnification Factor @ 2344.2 kc = 110.44

3670 / 2344.2 = 1.565566

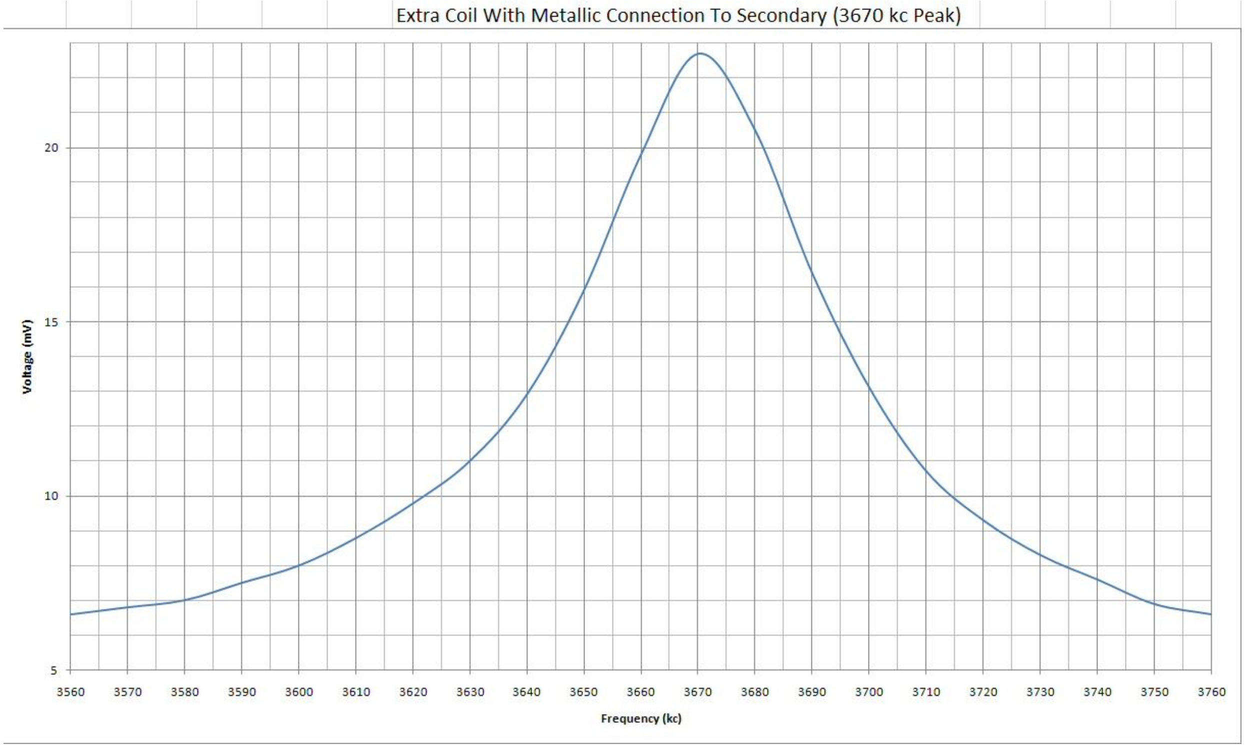
Extra Coil With Metallic Connection To Secondary (3670 kc Peak)
Voltage (mV)
5
10
15
20
Frequency (kc)
3560
3570
3580
3590
3600
3610
3620
3630
3640
3650
3660
3670
3680
3690
3700
3710
3720
3730
3740
3750
3760

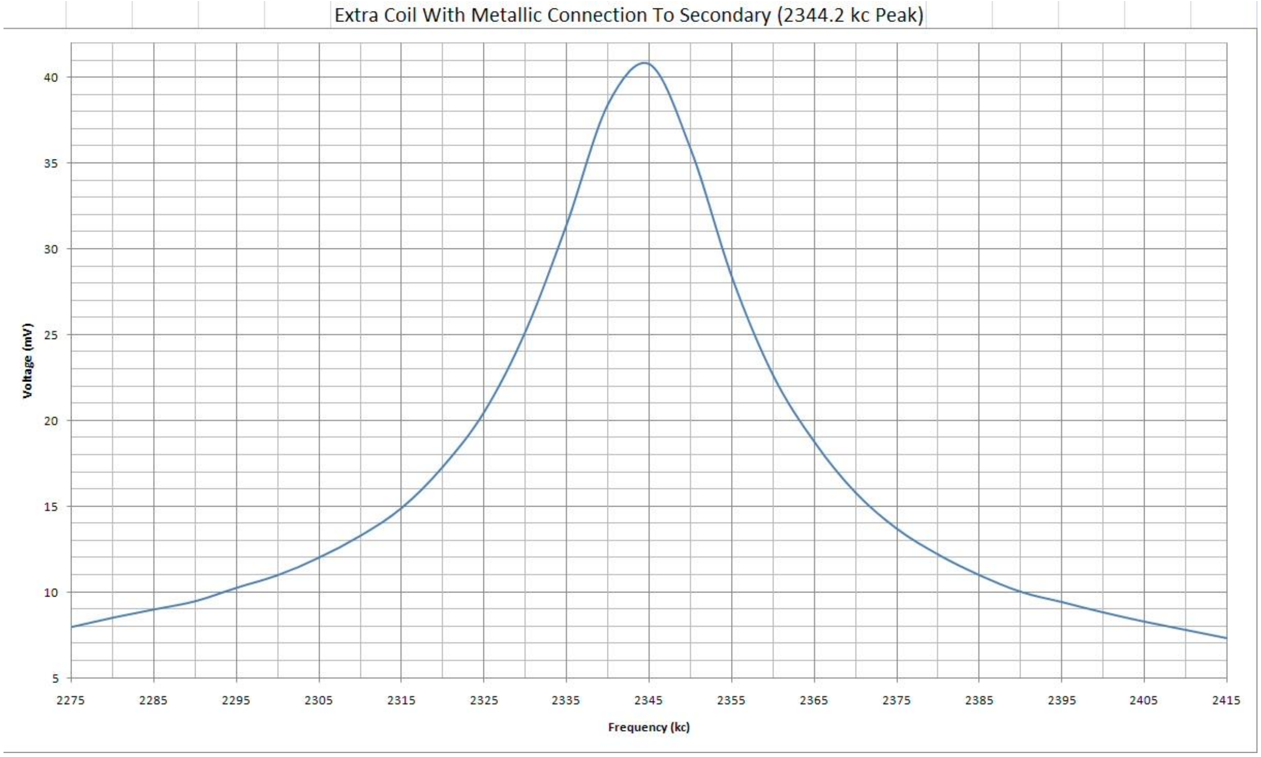
Extra Coil With Metallic Connection To Secondary (2344.2 kc Peak)
Voltage (mV)
5
10
15
20
25
30
35
40
Frequency (kc)
2275
2285
2295
2305
2315
2325
2335
2345
2355
2365
2375
2385
2395
2405
2415

06-25-2012, 11:48 PM

dR-Green
Gold Member

Join Date: Apr 2008
Location: Merlin's Fort
Posts: 1,220

Quote:

> Originally Posted by **T-rex**
> *Next test is now to measure the potential of the high end of the extra coil and graph the peaks.*

I wasn't happy with the consistency after leaving things overnight and moving the "probes" around and getting rulers in there to measure things etc, so I reset the whole thing and did the secondary test again and extra coil simultaneously. This is now the setup:

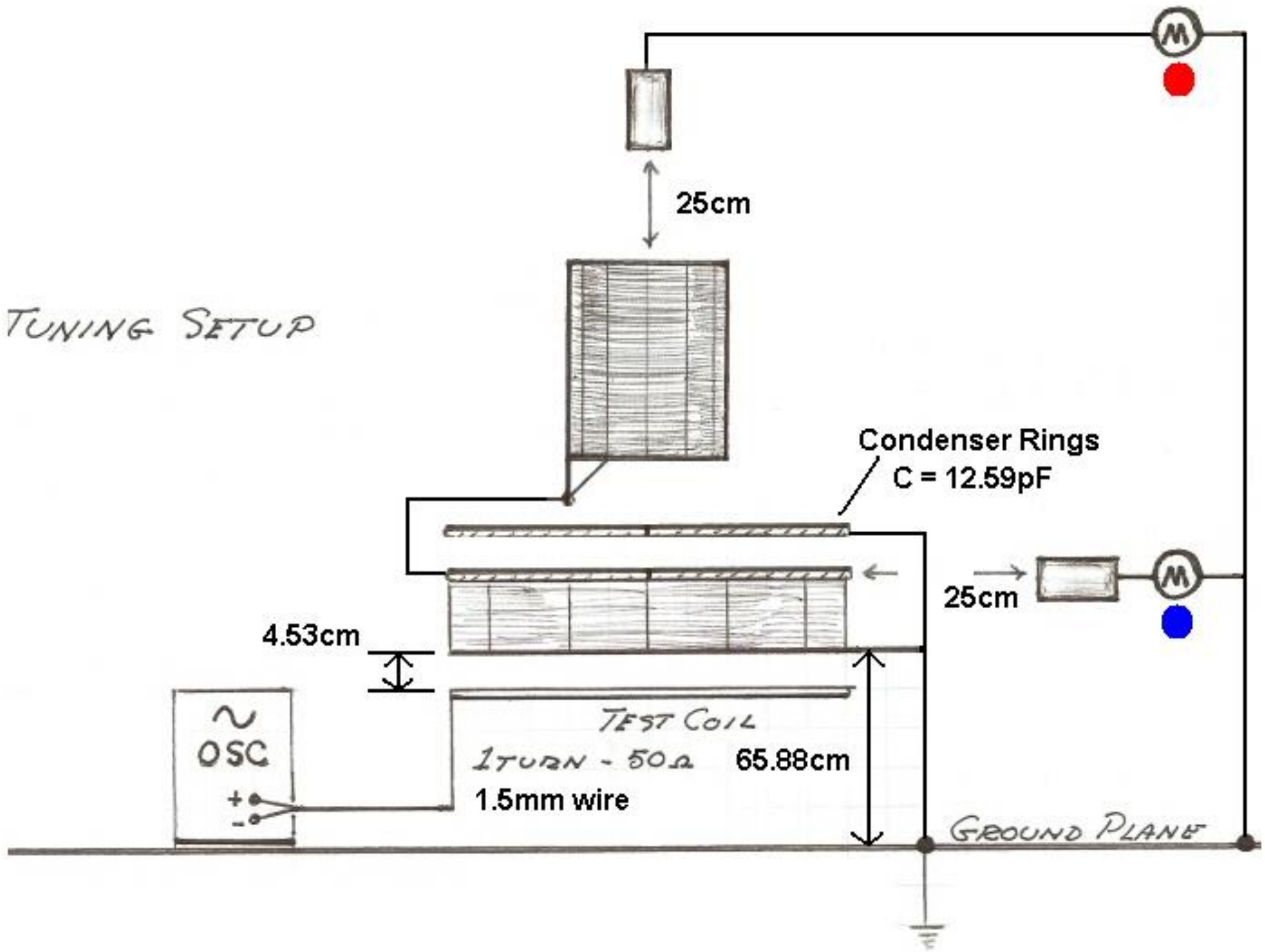

TUNING SETUP
M
25cm
Condenser Rings
C = 12.59pF
M
25cm
4.53cm
OSC
+
-
TEST COIL
1TURN - 50Ω
1.5mm wire
65.88cm
GROUND PLANE

Data based on most recent results:

Extra Coil:
126 turns
Diameter = 8.28cm
Height = 8.28cm
Conductor length = 32.271 metres

Luminal frequency = 2322.4 kc
Direct = 2694.8 kc - Magnification Factor = 57.07992
10pF = 2833.1 kc - Magnification Factor = 81.5891

Secondary Coil:
20 turns
Diameter = 20.7cm
Height = 4.08cm
Conductor length = 13.079 metres calculated (approx 12.92 metres used)

Luminal frequency = 5730.4 kc calculated (5800.9 kc for 12.92m)
Free maximum frequency = 4126 kc - Magnification Factor = 63.70233
Rings = 3670 kc - Magnification Factor = 83.71159

Concatenated Coils:
First Peak Resonance: 3670 kc
Second Peak Resonance: 2340 kc

Ratio = 1.568376

Even closer to Pi over two.

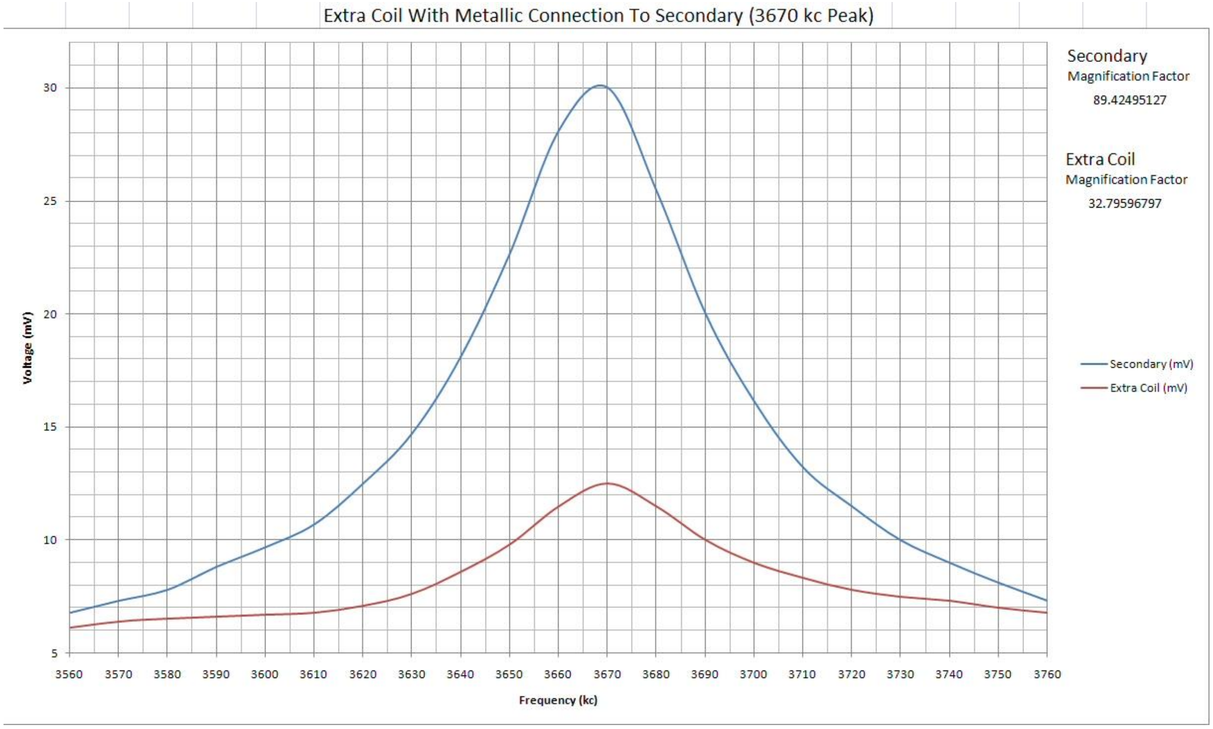
Extra Coil With Metallic Connection To Secondary (3670 kc Peak)
Voltage (mV)
30
25
20
15
10
5
3560
3570
3580
3590
3600
3610
3620
3630
3640
3650
3660
3670
3680
3690
3700
3710
3720
3730
3740
3750
3760
Frequency (kc)
Secondary
Magnification Factor
89.42495127
Extra Coil
Magnification Factor
32.79596797
Secondary (mV)
Extra Coil (mV)

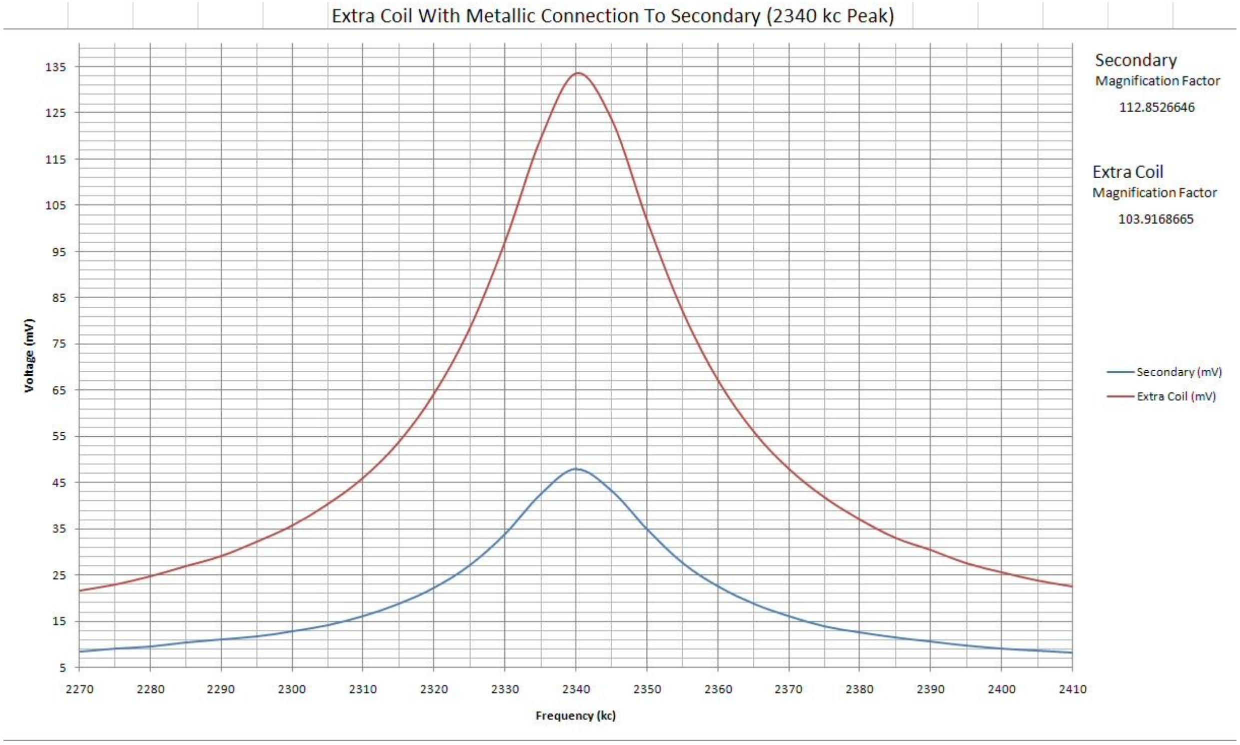
Extra Coil With Metallic Connection To Secondary (2340 kc Peak)
Voltage (mV)
135
125
115
105
95
85
75
65
55
45
35
25
15
5
2270
2280
2290
2300
2310
2320
2330
2340
2350
2360
2370
2380
2390
2400
2410
Frequency (kc)
Secondary
Magnification Factor
112.8526646
Extra Coil
Magnification Factor
103.9168665
Secondary (mV)
Extra Coil (mV)

06-26-2012, 08:59 PM

dR-Green
Gold Member

Join Date: Apr 2008
Location: Merlin's Fort
Posts: 1,220

So far I've borrowed the inductor (1 metre of wire) with ferrite rod from a basic crystal radio kit. I'm unable to get anything to happen with an incandescent or the fluorescents I have. Incandescents tested were a 3V torch bulb and a Mini Maglite AA bulb both measuring 0.8-1 ohm, fluorescent was a dismantled CFL, 4W and 6W tubes. I might need to make a simple (transistor) amplifier out of parts I have around to get a bit more power than the signal generator can provide.

06-26-2012, 10:21 PM

t-rex ●
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Dr. Green Coils

The secondary winding is the prime mover of this transformer. Here the system operating frequency is established , and then locked in place by a primary resonator. This in itself establishes a telluric transmission network, a basic resonant transformer with a mono-polar connection at the secondary neutral. Another transformer can be constructed, identical in every way except wound in an opposite direction to the first coil. Interlinking the neutrals of these two counter-wound transformers enables a one wire transmission between the two units. This is the basic telluric system. No more is needed than this.

The extra coil is employed in order to magnify the potential to much higher values than possible with only a secondary winding*. This however involves ratios of refractions and reflection that are difficult to calculate or even understand. Hence its complications are best avoided in basic telluric resting.

*This will increase receiving sensitivity, or "gain"

Two modes are possible for extra coil in relation with the secondary coil. Both involve quarter wave resonant rise, this the fundamental of resonant transformation. Its also known as constant potential to constant current transformation. A constant potential is a zero impedance (short circuit) a constant current is a zero admittance (open circuit). Departure from these zero values alters the coil distribution to something other than a quarter wave.

This quarter wave can exist in a distinct pair of manifestations. The first mode is when the quarter wave is distributed over the length of both extra and secondary windings as a whole, a pair of eighth waves let us say. This is the TANDEM mode. A multiplication in potential is derived hereby since the extra coil exhibits a higher transmission impedance thereby giving rise to a greater EMF between turns and thus a higher termination potential. All photos of my Bolinas and Integratron setups operated in this mode. It is the easy one to achieve.

The second mode of the extra coil and secondary coil connection involves two quarter wave distributions, one on each coil. This is not to be considered a half wave however. This mode is the <b>CONCATENATED</> connection.It compounds the quarter wave resonant rise of the secondary coil with another quarter wave rise in the extra coil, hence a concatenated resonant rise. This is the holy grail of resonant transformer design and unheard of potentials may be gained in this manner. To derive this analytically is extremely difficult, it is an advance transmission line problem. It might not even be possible to calculate or even achieve this mode of resonance, but we are going to give it a try. Tesla dreamed of this mode but electrical knowledge was in its infancy in Tesla's Colorado years. Also note that D. Dawson has become a zealot in drawing us back into the tandem mode of operation, which he claims to have mastered. It has come to pass that he is now becoming a detractor, his denouncements abound. I hereby ignore his posts and caution others that a confusion may develop here for the beginner in this un-charted territory.

In reviewing the Dr. Green material it can be seen that this extra coil is too large, or the wire is too long. Its resonant frequency, even with the lightest coupling, is far below that of the standardizing secondary frequency.

Extra Coil Direct: 2695 Kc/sec
Secondary, rings: 3670 Kc/sec
Frequency ratio: 72% lower
Wavelength ratio: 136% longer

Here seen is that a smaller coil is needed. However to start with, turns can be removed from the existing extra coil. But this shortens the coil, counteracting the turn removal by lowering the velocity and hence lengthening the wavelength of the coil. As a side note, both the secondary and existing extra coil should be tested independently over a wide frequency range to determine if other modes exist. All odd harmonics are resonant modes. Using the information gained from the design and construction of the existing extra coil it is possible to design and construct one for 3670 Kc/sec.

Constructing a pair of counter wound magnification transformers and connecting their neutrals can help minimize the burden of transmission to earth, a significant distorting influence. When possible this should be done. It must be remembered that this effort is really for the experienced transmission engineer, not the HAM operator, and surely not for the beginner. But Tesla was a beginner also, and through systematic experimentation made great progress with what little knowledge of standing and travelling waves in compound circuits that existed in his day. And of course he got no assistance from the pedants like Pupin. Just for fun take this problem of resonance to your local university E.E. professor, but make sure he does not see your license plate number. Is not this the way of modern society, the ones who are supposed to help are useless at best and hostile at worst? Then we have the shills?

73 DE N6KPH

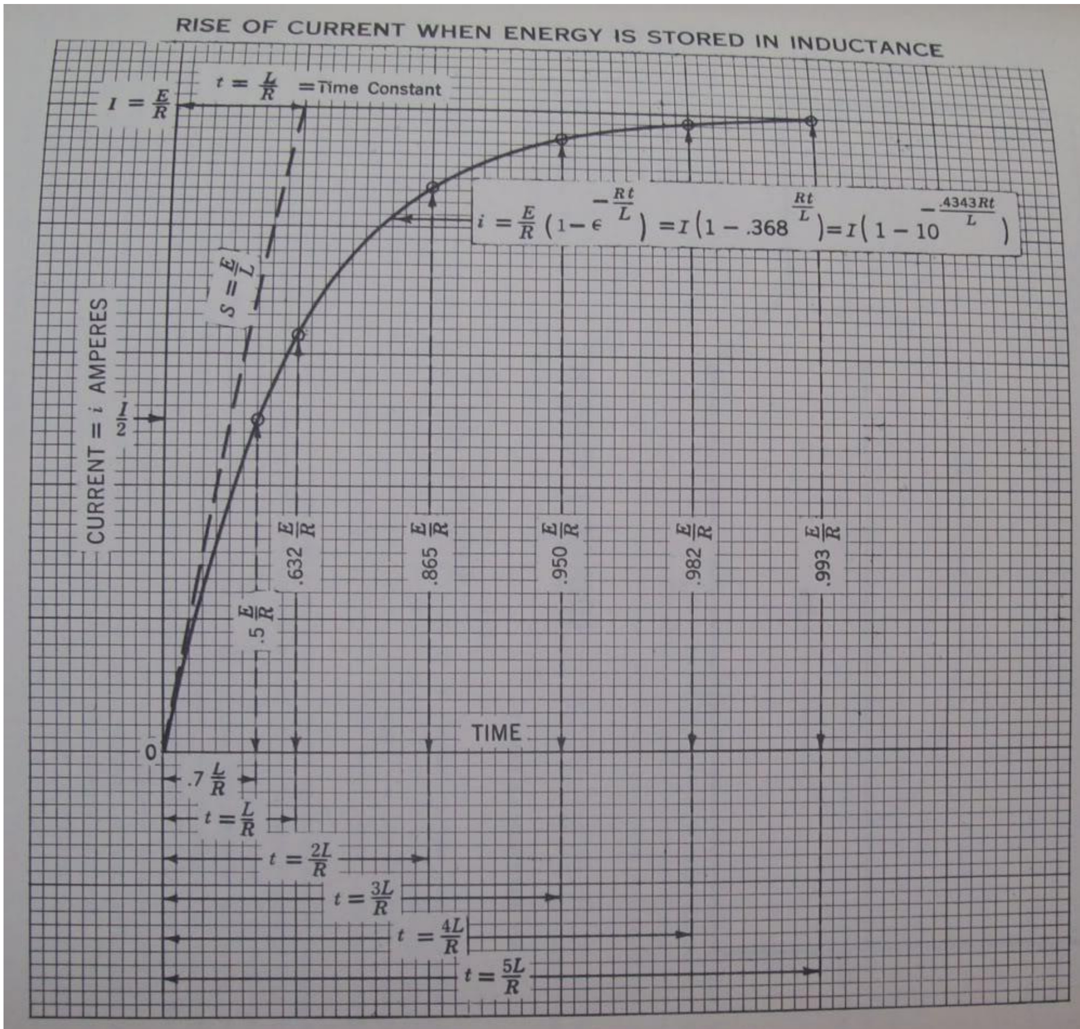

FIGURE 63 CHART 1

See also Figure 64, Chart 1

07-03-2012, 02:19 AM

dR-Green
Gold Member

Join Date: Apr 2008
Location: Merlin's Fort
Posts: 1,220

Quote:

> Originally Posted by **T-rex**
> *It compounds the quarter wave resonant rise of the secondary coil with another quarter wave rise in the extra coil, hence a concatenated resonant rise. This is the holy grail of resonant transformer design and unheard of potentials may be gained in this manner. To derive this*

> *analytically is extremely difficult, it is an advance transmission line problem. It might not even be possible to calculate or even achieve this mode of resonance, but we are going to give it a try.*

I don't know, but there are certainly two modes of operation that seem to be very different. The easiest way to get the 3670 kc resonance is to set the condenser rings to maximum capacitance, with some additional extra coil terminal capacitance. The size of the required terminal increases rapidly as the condenser rings capacity is decreased. In my opinion there is a balance to be found somewhere between the proper resonance of the secondary with the condenser rings and the extra coil with its terminal capacitance.

New extra coil:

Diameter = 8.28cm
Height = 8.28cm
64.25 Turns
Wire length = 16.47 metres 20 SWG (0.9mm)
Luminal frequency = 4550.8 kc
Free space frequency = 8509.996 kc

Terminal capacitance = connected via approx 36cm 14 SWG earth wire, 15cm diameter ball + apple pie tin resting on top + 2 tinfoil "leaves" for fine tuning

Secondary Coil:
20 turns
Diameter = 20.7cm
Height = 4.08cm
Conductor length = 13.079 metres calculated (approx 12.92 metres used)

Luminal frequency = 5730.4 kc calculated (5800.9 kc for 12.92m)
Free maximum frequency = 4126 kc - Magnification Factor = 63.70233
Rings (11.8pF) = 3670 kc - Magnification Factor = 83.71159

Concatenated Coils:
First Peak Resonance: 3670 kc
Second Peak Resonance: 1838 kc

Ratio = 1.996735582

First Peak Resonance Magnification Factor:
Extra coil = 108.2117057
Secondary = Unmeasurable

Second Peak Resonance Magnification Factor:
Extra coil = 88.59539188
Secondary = 102.0317531

Rings capacitance = 29.13pF

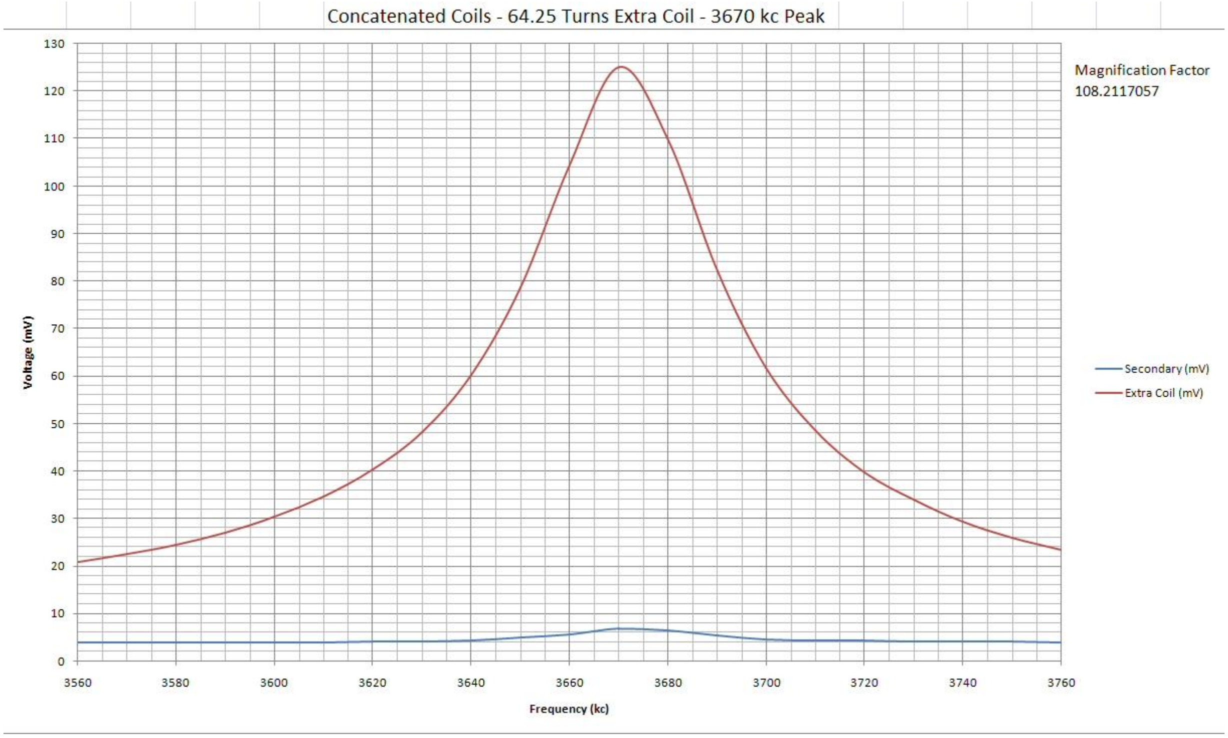

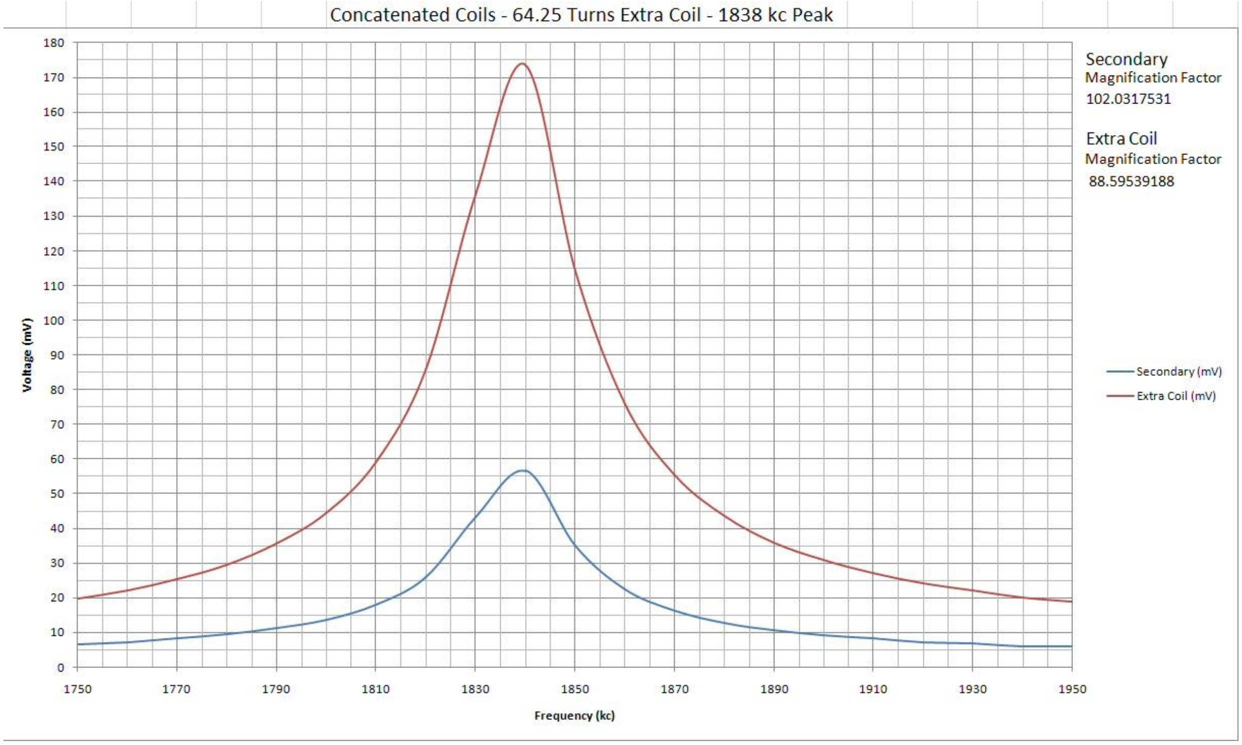

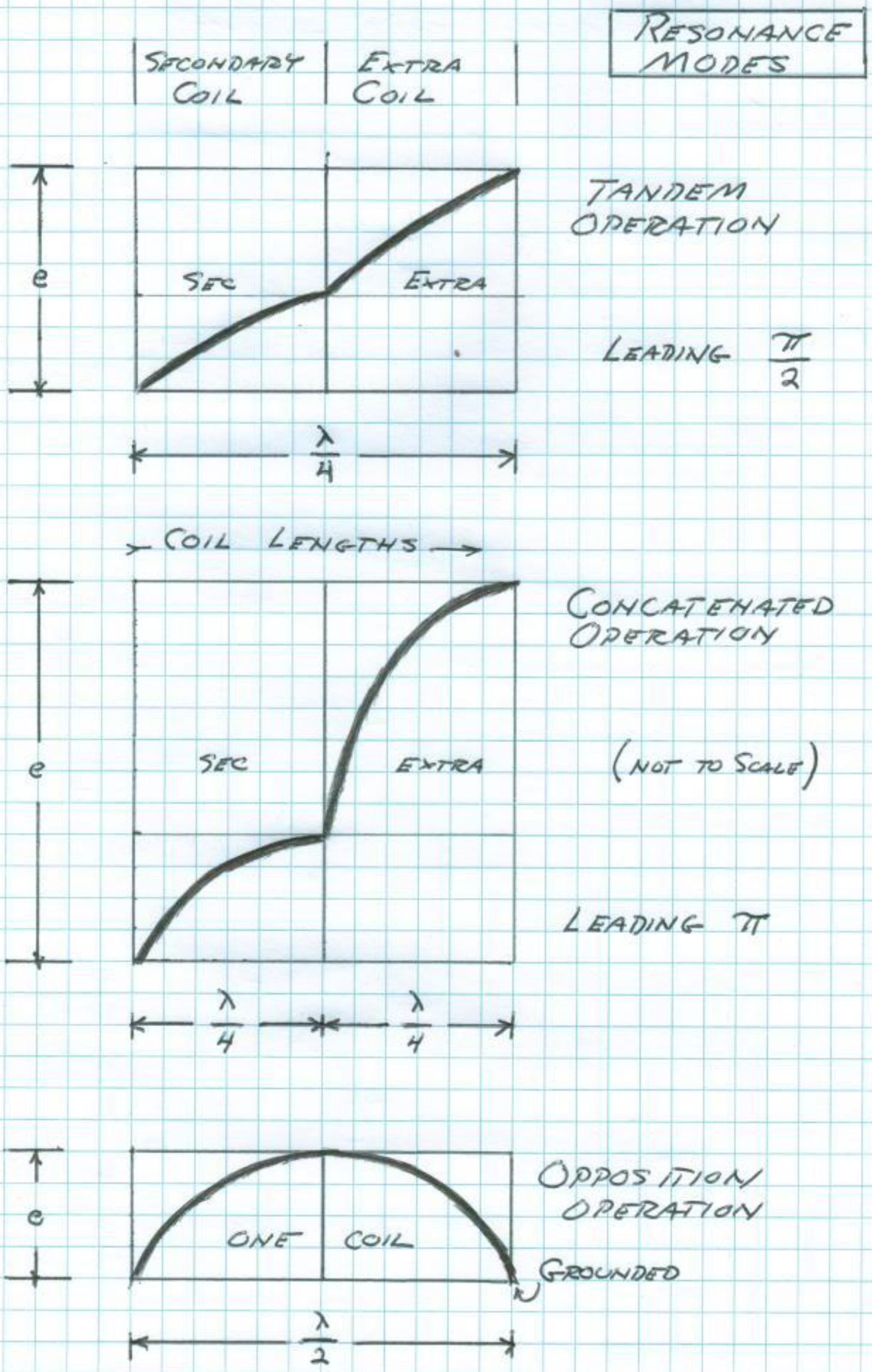
RESONANCE MODES
SECONDARY COIL
EXTRA COIL
e
SEC
EXTRA
TANDEM OPERATION
LEADING $\frac{\pi}{2}$
$\frac{\lambda}{4}$
COIL LENGTHS
e
SEC
EXTRA
CONCATENATED OPERATION
(NOT TO SCALE)
LEADING π
$\frac{\lambda}{4}$
$\frac{\lambda}{4}$
e
ONE COIL
OPPOSITION OPERATION
GROUNDED
$\frac{\lambda}{2}$

02-04-2012, 02:37 PM

t-rex
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Quadrapolar Resonance

QUADRAPOLAR RESONANT CIRCUIT

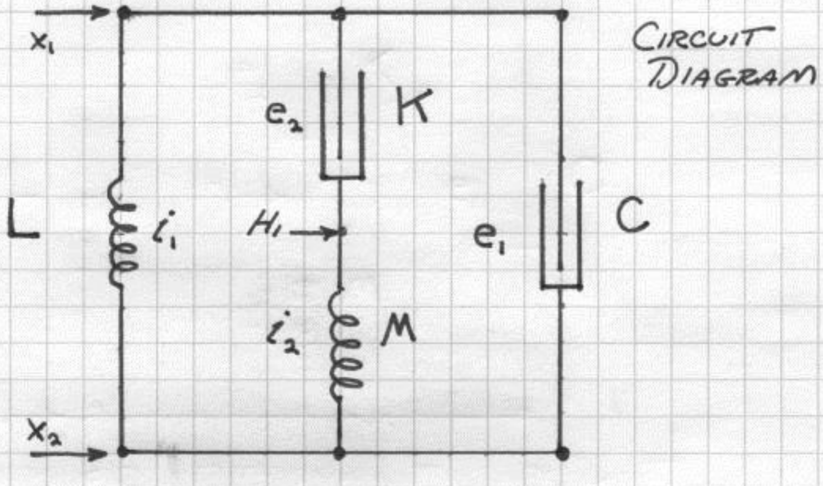
x_1
x_2
e_2
K
L
i_1
H_1
e_1
C
i_2
M
CIRCUIT DIAGRAM

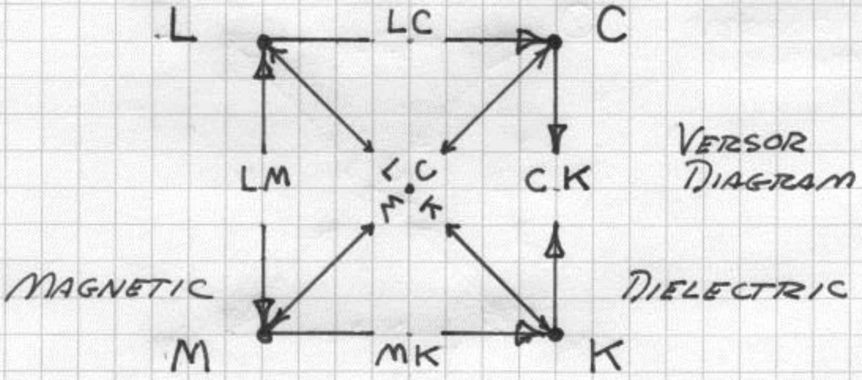
L
LC
C
LM
L C M K
C K
VERSOR DIAGRAM
MAGNETIC
DIELECTRIC
M
MK
K

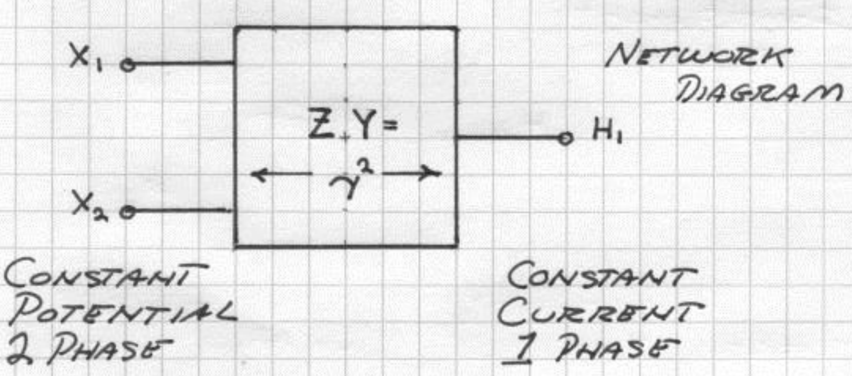
X_1
X_2
$Z\,Y = \gamma^2$
H_1
NETWORK DIAGRAM
CONSTANT POTENTIAL 2 PHASE
CONSTANT CURRENT 1 PHASE

$$i_1^2 L = i_2^2 M^{-1} = e_1^2 C = e_2^2 K^{-1}$$

$$\omega_1^2 = \omega_2^2 \,, \quad P_1 = P_2$$

THE FUNDAMENTAL RELATIONS FOR QUADRAPOLAR RESONANCE

SHUNT ADMITTANCE

$$Y_1 = \frac{i_1}{e_1} \; : \; Y_1 = \sqrt{\frac{C}{L}}$$

SIEMENS

SERIES IMPEDANCE

$$Z_1 = \frac{e_2}{i_2} \; : \; Z_1 = \sqrt{\frac{K}{M}}$$

OHM

TRANSFER ADMITTANCE

$$Y_0 = \frac{i_2}{e_1} \; : \; Y_0 = \sqrt{MC}$$

SIEMENS

TRANSFER IMPEDANCE

$$Z_0 = \frac{e_2}{i_1} \; : \; Z_0 = \sqrt{LK}$$

OHM

SHUNT FREQUENCY

$$\omega_1^{-1} = \sqrt{LC}$$

SECOND PER RADIAN

SERIES FREQUENCY

$$\omega_2^{+1} = \sqrt{MK}$$

RADIAN PER SECOND

TRANSFER FREQUENCY

$$\nu_0^{-1} = \sqrt{\frac{C}{M}}$$

SECOND PER RADIAN

TRANSFER FREQUENCY

$$\nu_0^{+1} = \sqrt{\frac{K}{L}}$$

RADIAN PER SECOND

MAGNETIC AND DIELECTRIC RELATIONS

$$\omega_s^{-2} = L_s C_s$$

MAGNETIC PROPORTION

$$\alpha_m = \sqrt{LM}$$

NUMERIC

DIELECTRIC PROPORTION

$$\alpha_d = \sqrt{CK}$$

NUMERIC

NATURAL INDUCTANCE

$$L_s = \sqrt{\frac{L}{M}}$$

HENRY

NATURAL CAPACITANCE

$$C_s = \sqrt{\frac{C}{K}}$$

FARAD

$$LCMK = \left[\begin{array}{l|l} \omega_1^2 = \omega_2^2 & (LC)^{-1} = (MK) \\ \\ P_1 = P_2 & (MC)^{-1} = (LK) \\ \\ \frac{\alpha_m}{\alpha_d} = 1 & (CK)^{-1} = (LM) \end{array} \right] = 1$$

THE CONDITION OF QUADRAPOLAR RESONANCE

06-01-2012, 01:12 PM

t-rex ●
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Tentative Experimental Circuits, Use at your own risk!

Regenerative receiver is a 2C22 triode, and the pulse generator is a 2050 thyratron, and a 5R4GYB dual diode rectifier.

05-31-2012, 01:18 PM

t-rex
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Pulse Generators, etc

(1) The Guillemen line, or network, can be regarded as an elementary electro-static condenser. It has two terminals, and stores electro-static, or dielectric, energy. this in a manner like the basic condenser. There is one important difference however, this "Guillemen Condenser" is a "loaded condenser".

The common condenser discharge is of a hyperbolic nature, this in Nepers per second. This is not a pulse, it is an impulse. The variation of a pulse is indefinite in time or a discontinuous function, not a hyperbolic one. The hyperbolic function is a result of a rate of change of a quantity which at each instant is proportional to the magnitude of the quantity itself. This gives rise to Epsilon, or base e natural logarithms. It is a law of energy and thermodynamics. The pulse, in distinction to the impulse, represents a constant flow of electric energy, hence constant power.

(2) Previously discussed was the loading on antenna and long distance telephone lines. For the most part the networks described were all four terminal structures, with the exception of reactance arms. The output terminals were the distant telephone, or the Aether itself. The Guillemen network is a two terminal structure and cannot be considered a transmission network as normally understood. Now it is a transmission network in a counter spatial form. So as with a loaded electro-magnetic telephone pair, a four terminal network, the Guillemen line is a loaded dielectric network, a two terminal network. In his book "communication networks", volume 2, Guillemen is close to a wave equation for counterspace. The "Type A" line can be considered a longitudinal magneto-dielectric transmission line. This is important.

(3) For pulse experimentation I recommend the 2050 Argon Thyratron, this instead of a 5C22. Likewise, a 5R4G dual half-wave rectifier can replace the pair of 705 H.V. diodes. This is where to start.

The OA4G has application as a radiant matter detector. The main circuit consists of a Western Electric Telephone Bell in series with the plate supply. The bell condenser should be connected in shunt to the bell windings. A contact should be fitted to the bell armature to break the cathode connection to the OA4G in order to reset the tube. See the Marconi Coherer detector.

(4) Th cosmic anode lead connects to the trigger or grid lead of the OA4G. There must be NO DC leakage to ground. For cosmic anodes see Vassilatos "Vril Compendium", and the experiments of Luigi Galvani.

(5) As for detectors into the outer limits the common regenerative detector is the circuit of choice. It is a development of E.H. Armstrong, see "Empire of the Air". The Regenerative

Detector is also developed in the Russian parametric paper, an important study. Radionic receivers here are possible, depending on the metallic-dielectric geometry used. Connection with a magnification transformer is possible.

Finally, the bulbs. The neon sign shops receive their gas in large glass globes. It is at a bit less than atmospheric pressure and when exhausted it of course is at the same pressure as the neon tube under construction. These bulbs then are junk to the neon guy. Great glom potential! These bulbs can be pumped further with a diffusion pump and a small ion gun can be placed in its neck, or a Tesla electrode fitted inside. Many possibilities.

(6) So sell your Bedini junk off on e-bay to the Chinese. Dispense with the Grey tube secrets and the rest of the useless lot. These are only dummy loads to tax your resources and keep your mind chasing its own tail. It is established here that the old "Electrical Experimenter" can come back to life, the possibilities are infinite.

73 DE N6KPH

TESLA TRANSFORMER PULSE GENERATOR

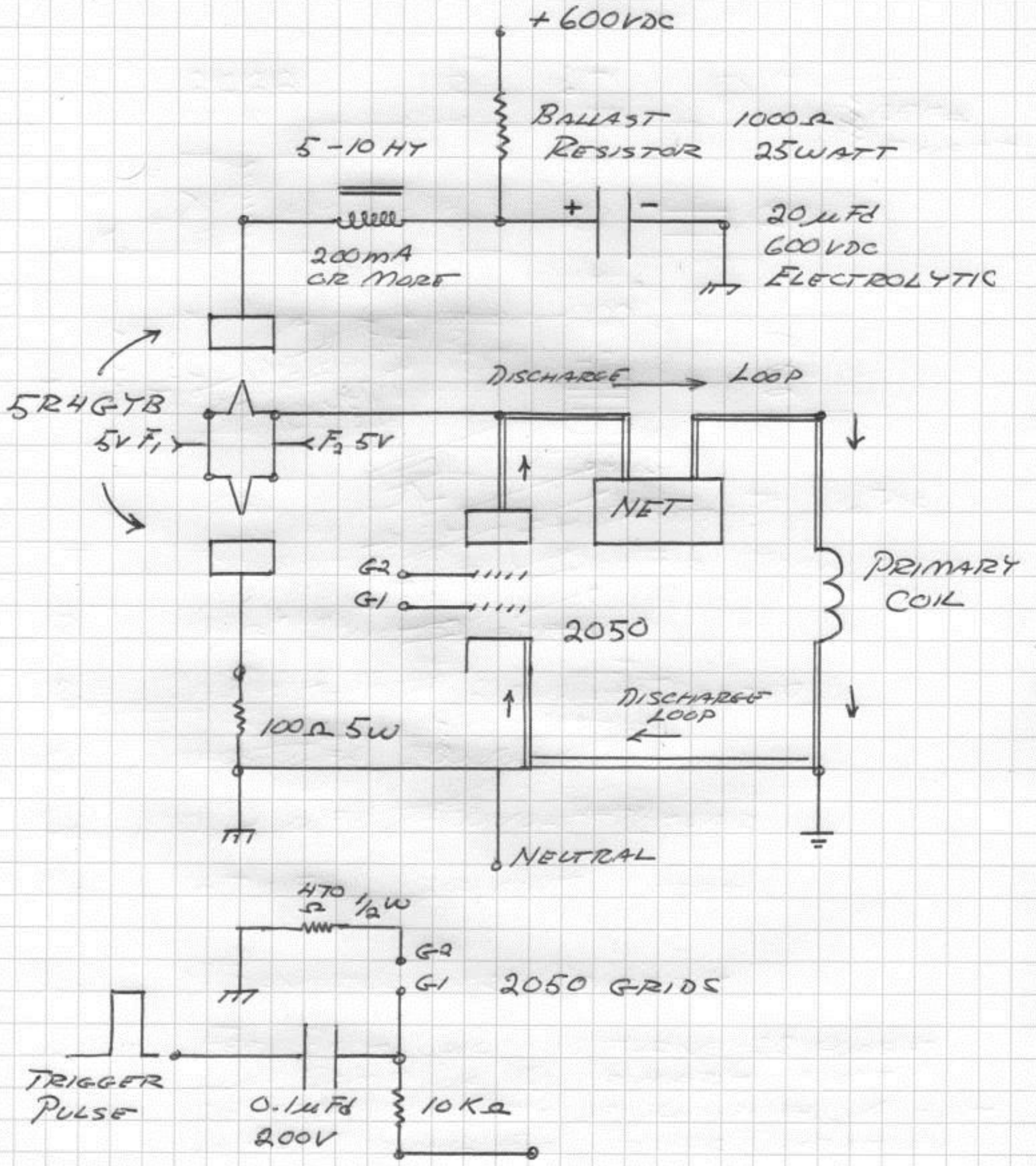
+600VDC
BALLAST RESISTOR
1000Ω 25WATT
5-10 HY
200mA OR MORE
+
-
20μFd 600VDC ELECTROLYTIC
5R4GYB
5V F1
F2 5V
DISCHARGE → LOOP
NET
G2
G1
2050
PRIMARY COIL
DISCHARGE LOOP
100Ω 5W
NEUTRAL
470 Ω ½W
G2
G1
2050 GRIDS
TRIGGER PULSE
0.1μFd 200V
10KΩ

REGENERATIVE MAGNIFYING RECIEVER

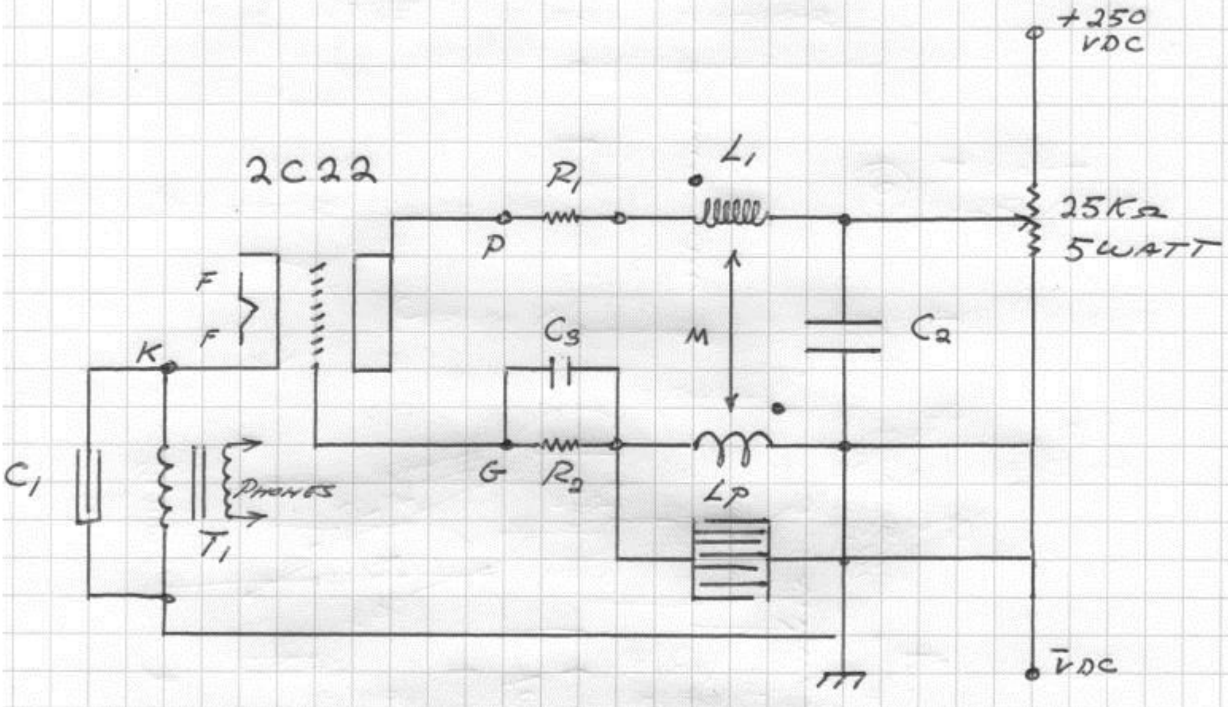

T_1 ,	TRANSISTOR RADIO SPEAKER TRANSFORMER
R_1 ,	100Ω 1WATT CARBON
R_2 ,	1.2 MEG-OHM 1/2 WATT CARBON
C_1 ,	0.05 µFd 200 VOLT CERAMIC
C_2 ,	0.10 µFd , 400 VOLT MYLAR
C_3 ,	47 pFd , 200 VOLT CERAMIC
L_p ,	TRANSFORMER PRIMARY
L_1 ,	FEED BACK COIL, 10 TURNS 22 GUAGE LOOP SAME DIA AS PRIMARY NOTE! OBSERVE POLARITY OF WINDINGS. ADJUST L_p - L_1 SPACING FOR PROPER OPERATION

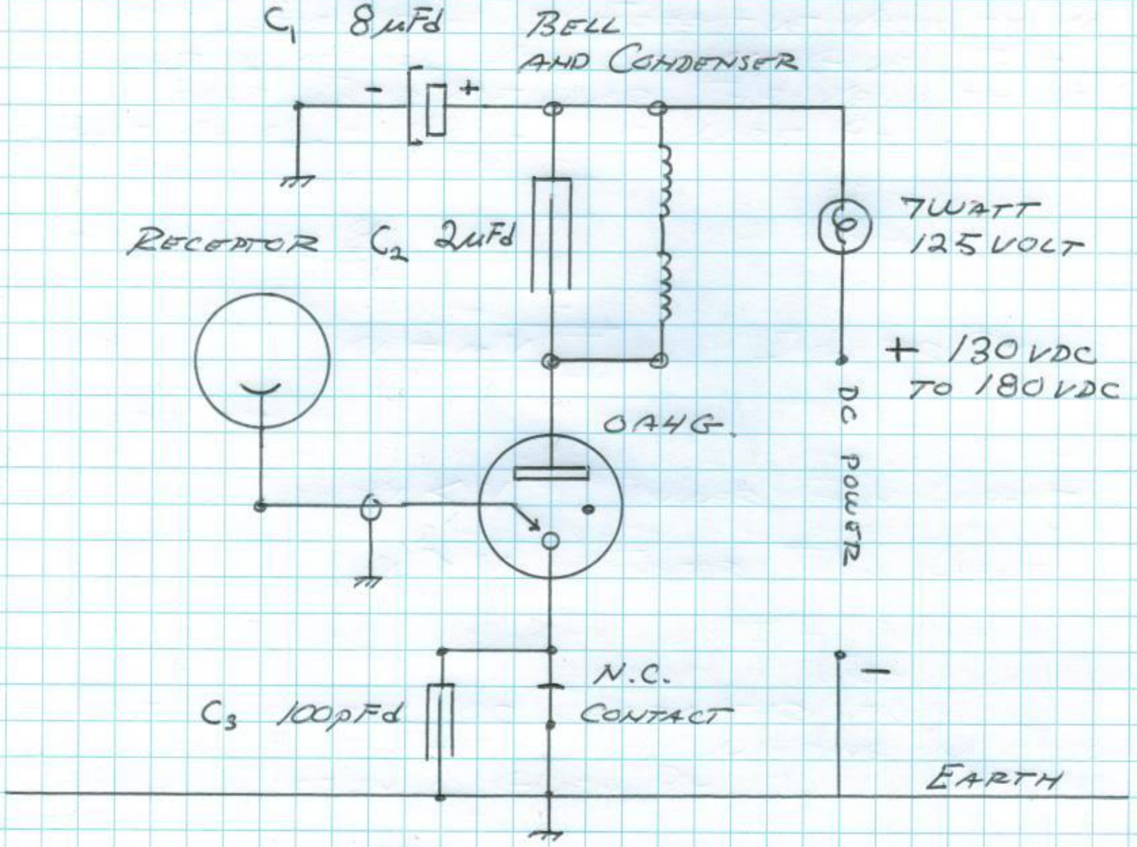
C1 8μFd
BELL
AND CONDENSER
- +
7WATT
125VOLT
RECEPTOR
C2 2μFd
+ 130VDC
TO 180VDC
DC POWER
OA4G
N.C.
CONTACT
-
C3 100pFd
EARTH

COSMIC RAY DETECTOR

SEE R.C.A. APPLICATION
NOTES ON OA4G RELAY SERVICE

C1 PLATE STORAGE CONDENSER 8μFd, 200V ELECTROLYTIC
C2 TUNING CONDENSER 2μFd, 400V OIL / PAPER
C3 ARC SUPPRESSOR 100pFd, 400V CERAMIC

coupling shown in Fig. 5-5. The same order of inductance is required as with the pentode grid-leak detector described previously.

The plate detector is more sensitive than the diode since there is some amplifying action in the tube, but less so than the grid-leak detector. It will handle considerably larger signals than the grid-leak detector, but is not quite so tolerant in this respect as the diode. Linearity, with the self-biased circuits shown, is good. Up to the overload point the detector takes no power from the tuned circuit, and so does not affect its Q and selectivity.

Infinite-Impedance Detector

The circuit of Fig. 5-6 combines the high signal-handling capabilities of the diode detector with low distortion (good linearity), and, like the plate detector, does not load the tuned circuit it connects to. The circuit resembles that of the plate detector, except that the load resistance, R_1, is connected between cathode and ground and thus is common to both grid and plate circuits, giving negative feed-back for the audio frequencies. The cathode resistor is by-passed for r.f. (C_2) but not for audio, while the plate circuit is by-passed to ground for both audio and radio frequencies. R_2 forms, with C_3, an RC filter to isolate the plate from the "B" supply at a.f. An r.f. filter, consisting of a series r.f. choke and a shunt condenser, can be connected between the cathode and C_4 to eliminate any r.f. that might otherwise appear in the output.

The plate current is very low at no signal, increasing with signal as in the case of the plate detector. The voltage drop across R_1 similarly increases with signal, because of the increased plate current. Because of this and the fact that the initial drop across R_1 is large, the grid cannot be driven positive with respect to the cathode by the signal, hence no grid current can be drawn.

REGENERATIVE DETECTORS

By providing controllable r.f. feed-back or regeneration in a triode or pentode detector circuit, the incoming signal can be amplified many times, thereby greatly increasing the sensitivity of the detector. Regeneration also increases the effective Q of the circuit and increases the selectivity because the maximum regenerative amplification takes place only at the frequency to which the circuit is tuned. The grid-leak type of detector is most suitable for the purpose. Except for the regenerative connection, the circuit values are identical with those previously described for this type of detector, and the same considerations apply. The amount of regeneration must be controllable, because maximum regenerative amplification is secured at the critical point where the circuit is just about to oscillate, and the critical point in turn depends upon circuit conditions, which may vary with the frequency to which the detector is tuned. In the oscillating condition, a regenerative detector can be detuned slightly from an incoming c.w. signal to give *autodyne* reception.

Fig. 5-7 shows the circuits of regenerative detectors of various types. The circuit of A is for a triode tube, with a variable by-pass condenser, C_3, in the plate circuit to control regeneration. When the capacity is small the tube does not regenerate, but as it increases toward maximum its reactance becomes smaller until a critical value is reached where there is sufficient feed-back to cause oscillation. If L_2 and L_3 are wound end-to-end in the same direction, the plate connection is to the outside of the plate or "tickler" coil, L_3, when the grid connection is to the outside of L_2.

The circuit of 5-7B is for a pentode tube, regeneration being controlled by adjustment of the screen-grid voltage. The tickler, L_3, is in the plate circuit. The portion of the control resistor between the rotating contact and ground is by-passed by a large condenser (0.5 μfd. or more) to filter out scratching noise when the arm is rotated. The feed-back is adjusted by varying the number of turns on L_3 or the coupling between L_2 and L_3, until the tube just goes into oscillation at a screen potential of approximately 30 volts.

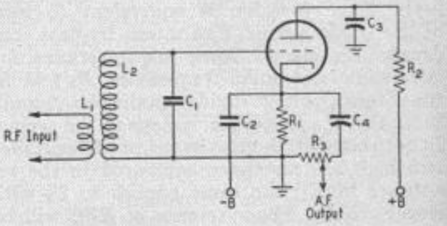

Fig. 5-6 — **The infinite-impedance detector. The input circuit, L_2C_1, is tuned to the signal frequency. Typical values for the other components are:**

C_2 — 250 μμfd. R_1 — 0.15 megohm.
C_3 — 0.5 μfd. R_2 — 25,000 ohms.
C_4 — 0.1 μfd. R_3 — 0.25-megohm volume control.

A tube having a medium amplification factor (about 20) should be used. Plate voltage should be 250 volts.

Circuit C is identical with B in principle of operation, except that the oscillating circuit is of the Hartley type. Since the screen and plate are in parallel for r.f. in this circuit, only a small amount of "tickler" — that is, relatively few turns between the cathode tap and ground — is required for oscillation.

Smooth Regeneration Control

The ideal regeneration control would permit the detector to go into and out of oscillation smoothly, would have no effect on the frequency of oscillation, and would give the same value of regeneration regardless of frequency and the loading on the circuit. In practice, the effects of loading, particularly the loading that occurs when the detector circuit is coupled to an antenna, are difficult to overcome. Likewise, the regeneration is usually affected by the frequency to which the grid circuit is tuned.

In all circuits it is best to wind the tickler at the ground or cathode end of the grid coil, and to use as few turns on the tickler as will allow the detector to oscillate easily over the whole tuning range at the plate (and screen, if a pentode) voltage that gives maximum sensitivity. Should the tube break into oscillation suddenly as the regeneration control is advanced, making a click, the operation often can be made smoother by changing the grid-leak resistance to a higher or lower value. The wrong grid leak plus too-high plate and screen voltage are the most frequent causes of lack of smoothness in going into oscillation.

Antenna Coupling

If the detector is coupled to an antenna, slight changes in the antenna constants (as when the wire swings in a breeze) affect the frequency of the oscillations generated, and thereby the beat frequency when c.w. signals are being received. The tighter the antenna coupling is made, the greater will be the feed-back required or the higher will be the voltage necessary to make the detector oscillate. The antenna coupling should be the maximum that will allow the detector to go into oscillation smoothly with the correct voltages on the tube. If capacity coupling to the grid end of the coil is used, only a very small amount of capacity will be needed to couple to the antenna. Increasing the capacity increases the coupling.

At frequencies where the antenna system is resonant the absorption of energy from the oscillating detector circuit will be greater, with the consequence that more regeneration is needed. In extreme cases it may not be possible to make the detector oscillate with normal voltages, causing so-called "dead spots." The remedy for this is to loosen the antenna coupling to the point that permits normal oscillation and smooth regeneration control.

Body Capacity

A regenerative detector occasionally shows a tendency to change frequency slightly as the hand is moved near the dial. This condition (**body capacity**) can be caused by poor design of the receiver, or by the antenna if the detector is coupled directly to it. If body capacity is present when the antenna is disconnected, it can be eliminated by better shielding, and sometimes by r.f. filtering of the 'phone leads. Body capacity that is present only when the antenna is connected is caused by resonance effects in the antenna, which tend to raise the whole detector circuit above ground potential. A good, short ground connection should be made to the receiver and the length of the antenna varied electrically (by adding a small coil or variable condenser in the antenna lead) until the effect is minimized. Loosening the coupling to the antenna circuit also will help.

Hum

Hum at the power-supply frequency may be present in a regenerative detector, especially when it is used in an oscillating condition for c.w. reception, even though the plate supply itself is free from ripple. The hum may result from the use of a.c. on the tube heater, but effects of this type normally are troublesome

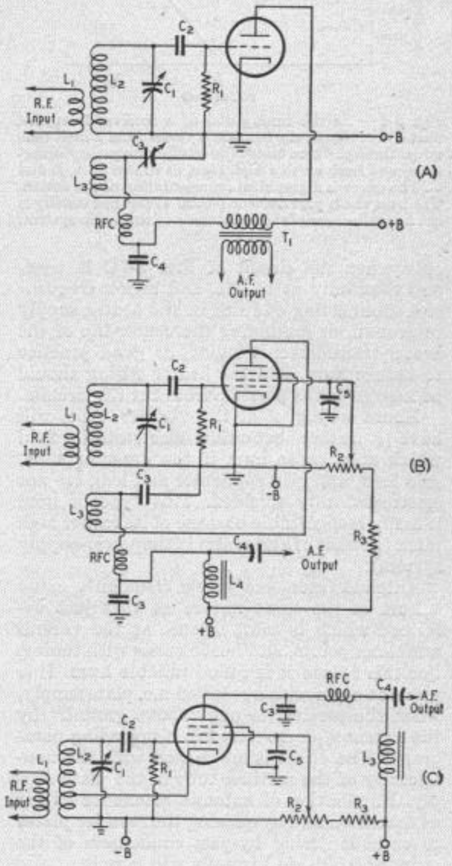

Fig. 5-7 — Triode and pentode regenerative detector circuits. The input circuit, L_2C_1, is tuned to the signal frequency. The grid condenser, C_2, should have a value of about 100 μμfd. in all circuits; the grid leak, R_1, may range in value from 1 to 5 megohms. The tickler coil, L_3, ordinarily will have from 10 to 25 per cent of the number of turns on L_2; in C, the cathode tap is about 10 per cent of the number of turns on L_2 above ground. Regeneration-control condenser C_3 in A should have a maximum capacity of 100 μμfd. or more; by-pass condensers C_3 in B and C are likewise 100 μμfd. C_5 is ordinarily 1 μfd. or more; R_2, a 50,000-ohm potentiometer; R_3, 50,000 to 100,000 ohms. L_4 in B (L_3 in C) is a 500-henry inductance. C_4 is 0.1 μfd. in both circuits. T_1 in A is a conventional audio transformer for coupling from the plate of a tube to a following grid. *RFC* is 2.5 mh. In A, the plate voltage should be about 50 volts for best sensitivity. Pentode circuits require about 30 volts on the screen; plate potential may be 100 to 250 volts.

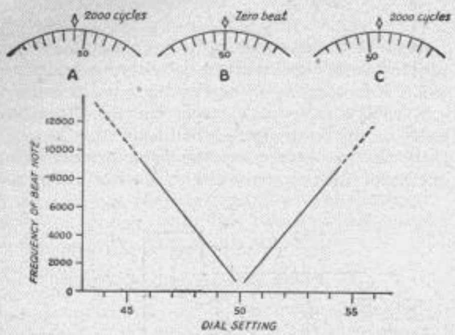

Fig. 5-8 — As the tuning dial of a receiver is turned past a c.w. signal, the beat-note varies from a high tone down through "zero beat" (no audible frequency difference) and back up to a high tone, as shown at A, B and C. The curve is a graphical representation of the action. The beat exists past 8000 or 10,000 cycles but usually is not heard because of the limitations of the audio system.

only when the circuit of Fig. 5-7C is used, and then only at 14 Mc. and higher frequencies. Connecting one side of the heater supply to ground, or grounding the center-tap of the heater-transformer winding, is good practice to reduce hum, and the heater wiring should be kept as far as possible from the r.f. circuits.

House wiring, if of the "open" type, will have a rather extensive electrostatic field which may cause hum if the detector tube, grid lead, and grid condenser and leak are not electrostatically shielded. This type of hum is easily recognizable because of its rather high pitch (a result of harmonics in the power-supply system).

Antenna resonance effects frequently cause a hum of the same nature as that just described which is most intense at the various resonance points, and hence varies with tuning. For this reason it is called **tunable hum.** It is prone to occur with a rectified-a.c. plate supply, when the receiver is put "above ground" by the antenna, as described in a preceding paragraph. The effect is associated with the non-linearity of the rectifier tube in the plate supply. Elimination of antenna resonance effects as described and by-passing the rectifier plates to cathode (using by-pass condensers of the order of 0.001 μfd.) usually will cure it.

Tuning

For c.w. reception, the regeneration control is advanced until the detector breaks into a "hiss," which indicates that the detector is oscillating. Further advancing the regeneration control after the detector starts oscillating will result in a slight decrease in the strength of the hiss, indicating that the sensitivity of the detector is decreasing.

The proper adjustment of the regeneration control for best reception of c.w. signals is where the detector just starts to oscillate, when it will be found that c.w. signals can be tuned in and will give a tone with each signal depending on the setting of the tuning control. As the receiver is tuned through a signal the tone first will be heard as a very high pitch, then will go down through "zero beat" (the region where the frequencies of the incoming signal and the oscillating detector are so nearly alike that the difference or beat is less than the lowest audible tone) and rise again on the other side, finally disappearing at a very high pitch. This behavior is shown in Fig. 5-8. It will be found that a low-pitched beat-note cannot be obtained from a strong signal because the detector **"pulls in"** or **"blocks"**; that is, the signal tends to control the detector in such a way that the latter oscillates at the signal frequency, despite the fact that the circuit may not be tuned exactly to resonance. This phenomenon, commonly observed when an oscillator is coupled to a source of a.c. voltage of approximately the frequency at which the oscillator is operating, is called **"locking-in"**; the more stable of the two frequencies assumes control over the other. "Blocking" usually can be corrected by advancing the regeneration control until the beat-note occurs again. If the regenerative detector is preceded by an r.f. amplifier stage, the blocking can be eliminated by reducing the gain of the r.f. stage. If the detector is coupled to an antenna, the blocking condition can be satisfactorily eliminated by advancing the regeneration control or loosening the antenna coupling.

The point just after the detector starts oscillating is the most sensitive condition for c.w. reception. Further advancing the regeneration control makes the receiver less prone to blocking by strong signals, but also less capable of receiving weak signals.

If the detector is in the oscillating condition and a 'phone signal is tuned in, a steady audible beat-note will result. While it is possible to listen to 'phone if the receiver can be tuned to exact zero beat, it is more satisfactory to reduce the regeneration to the point just before the receiver goes into oscillation. This is also the most sensitive operating point.

Audio-Frequency Amplifiers

The ordinary detector does not produce very much audio-frequency power output — usually not enough to give satisfactory sound volume, even in headphone reception. Consequently, audio-frequency amplifiers are used after the detector to increase the power level. One amplifier usually is sufficient for headphones, but two stages generally are used where the receiver is to operate a loudspeaker. A few milliwatts of a.f. power are sufficient for headphones, but a loudspeaker requires a watt or more for good room volume.

The Superregenerative Receiver

The simplest type of v.h.f. receiver is the superregenerator, for many years the most popular receiver for v.h.f. work. It affords fair sensitivity with few tubes and elementary circuits, and though it has largely been replaced by the more effective superheterodyne for home-station use, it still has many v.h.f. applications. Its disadvantages are lack of selectivity, poor signal-to-noise ratio on weak signals, and its tendency to radiate a strong signal which causes severe interference.

Its selectivity may be improved somewhat and its interference capabilities reduced by the addition of an r.f. stage, a refinement which should be considered a necessity if the receiver is to be used in a locality where there are other stations operating on the same band. If no r.f. stage is used, as in portable applications where economy of space and battery drain are primary considerations, the detector should be operated with the lowest plate voltage that will permit superregeneration, in order to reduce its interference range.

From a practical aspect, superregenerative receivers may be divided into two general types. In the first the quenching voltage is developed by the detector itself, called a "self-quenched" detector. In the second, a separate low-frequency oscillator is used to generate the quench voltage. Self-quenched detectors have found wide favor, particularly for portable work; but it is possible to achieve better performance with the separately-quenched type, particularly as the frequency approaches the upper limit of the tube's capabilities.

Superregeneration Principles

The limit to which ordinary regenerative amplification can be carried is the point at which oscillation commences, since at that point further amplification ceases. The superregenerative detector overcomes this limitation by introducing into the detector circuit an alternating voltage of a frequency somewhat above the audible range, the value being between 20 and 200 kc. depending on the signal frequency. Because the oscillations are constantly being interrupted by this quenching voltage the regeneration can be greatly increased, and the amplified signal will build up to tremendous proportions. A one-tube superregenerative receiver is capable of an inherent sensitivity approaching the thermal-agitation noise level of the tuned circuit, and may have an antenna input sensitivity of two microvolts or better.

Because of its inherent characteristics, the superregenerative circuit is suitable only for the reception of modulated signals, and operates best on the very-high frequencies. Typical superregenerative circuits for separately-quenched and self-quenched detectors are shown in Fig. 12-21, but the basic circuit may be any of the various arrangements used for straight regenerative detectors.

In the self-quenched detector the frequency of the quench oscillation depends upon the feed-back and upon the time constant of the grid leak and condenser, the oscillation being a "blocking" or "squegging" in which the grid accumulates a strong negative charge which does not leak off rapidly enough through the grid leak to prevent a relatively slow variation of the operating point.

The greater the difference between the quenching and signal frequencies the greater the amplification, because the signal then has a longer period in which to build up during the nonquenching half-cycle when the resistance of the circuit is negative. This ratio should not exceed a certain limit, however, for during the quenched or nonregenerative intervals the input selectivity is merely that of the Q of the tuned circuit alone.

Because of the greater amplification, the hiss noise when a super-

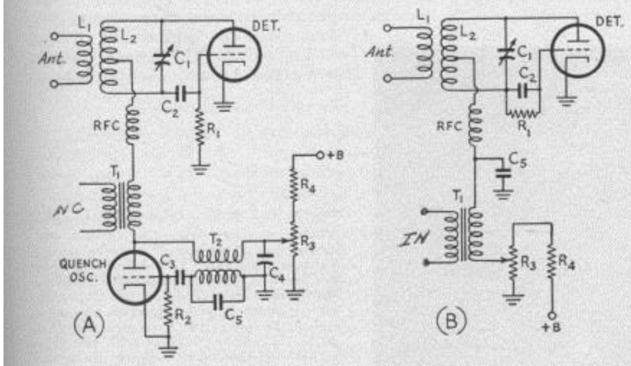

Fig. 12-21 — (A) Superregenerative detector circuit using a separate quench oscillator. (B) Self-quenched superregenerative detector circuit. L_2C_1 is tuned to the signal frequency. Typical values for other components are:

C_2 — 47 μμfd.
C_3 — 470 μμfd.
C_4 — 0.1 μfd.
C_5 — 0.001–0.005 μfd.
R_1 — 2 — 10 megohms.
R_2 — 47,000 ohms.
R_3 — 50,000-ohm potentiometer.
R_4 — 47,000 ohms.
RFC — R.f. choke, value depending upon frequency. Small low-capacitance chokes are required for v.h.f. operation.
T_1 — Audio transformer, plate-to-grid type.
T_2 — Quench-oscillator transformer.

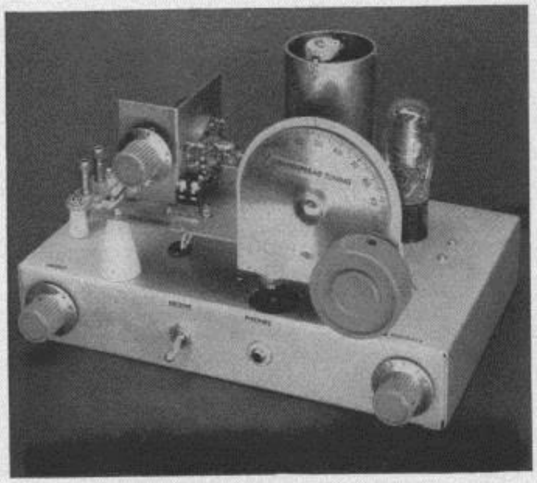

Fig. 12-22 — Front view of the coaxial-line receiver. The r.f. amplifier tuning control is at the left and the main control, for the concentric-line detector circuit, is at the right side of the unit. The audio gain control, send-receive switch, 'phone jack and regeneration control can be seen in that order, from left to right, across the front wall of the chassis.

regenerative detector goes into oscillation is much stronger than with the ordinary regenerative detector. The most sensitive condition is at the point where the hiss first becomes marked. When a signal is tuned in, the hiss will disappear to a degree that depends upon the signal strength.

Lack of hiss indicates insufficient feed-back at the signal frequency, or inadequate quench voltage. Antenna-loading effects will cause dead spots that are similar to those in regenerative detectors and can be overcome by the same methods. The self-quenching detector may require critical adjustment of the grid-leak and grid-condenser values for smooth operation, since these determine the frequency and amplitude of the quench voltage.

● A COAXIAL-LINE SUPERREGENERATIVE RECEIVER FOR 220 MC.

The performance of a superregenerative receiver, both as to selectivity and smoothness of operation, can be improved by the use of a coaxial line tank in the grid circuit of the detector, in place of the customary coil and condenser. Addition of an r.f. amplifier stage will improve sensitivity, reduce radiation, and make antenna coupling less critical. A superregenerative receiver for 220 to 240 Mc. incorporating these features is shown in Figs. 12-22–12-25.

The r.f. tube is a 954 acorn with a conventional tuned circuit in its grid. The plate circuit is a self-resonant loop, which is coupled to the concentric line grid circuit of the 6AK5 detector. The detector output is fed through a quench filter to a

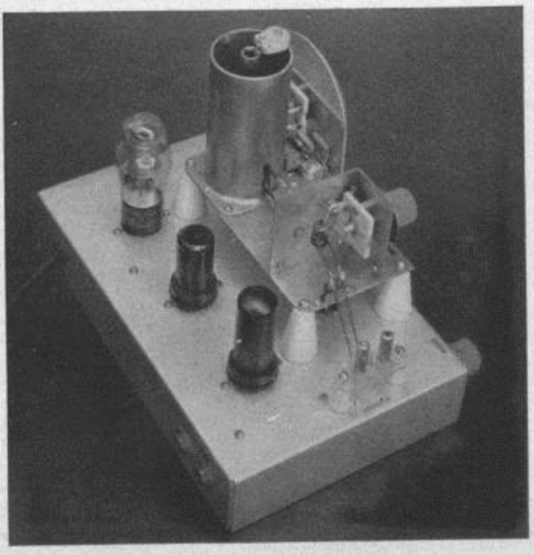

Fig. 12-23 — Rear view of the superregenerative receiver. The r.f. circuits are mounted on a copper shelf to the left of the antenna terminals. The detector tuning condenser is mounted on a small panel to the front of the coaxial line, and the bandset condenser is soldered across the open end of the line. The r.f. stage is mounted on an "L"-shaped bracket with the tube socket and plate-circuit components on the left side and the grid circuit on the right side. Audio tubes and voltage regulator are in line across the rear of the chassis.

05-13-2012, 03:17 PM

t-rex
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Notes on Tesla Transceiving 2

You are all playing with my first fully mathematical design, no books, no experiment, just a pad of paper in the seat of my Corolla sitting in between giant rocks, one 14,500 feet high! And you all have shown it to work. It must be noted here the given extra coil is a compromise, Colorado Unit will not completely scale down to broadcast band without excessive turns of fine wire on a tiny extra coil.

Note that a primary condenser is a must and it must match the KVar of the primary coil. Equal surfaces or equal weight for coil and condenser, a good "Rule of Thumb". There must be no breaks, gaps, etc between the coil and condenser, NONE! Tesla says this in his refined designs. (Fig 1)

The final setup at Colorado Springs, the FINAL ONE, should go on forum;

1) Coil heights, widths,
2) Number of turns,
3) Wire Gauge, and number of wires in primary bundle,
4) Length of wire in secondary and extra coils,
5) Operating frequencies,
6) Primary condenser connections and auxilliary devices in with the primary system

This is of extreme importance, do not get hung up on test coils, un-related experiments, or early transformer construction, we want ONLY his finished product known to be his final perfection for earth resonance. Can anyone do this and not put out distractive parts of Tesla's notebook? Let us see. Also a compendium of everyone's unit is important here.

In conclusion I am very happy to see how all this grew out of a few pages of equations scribbled in the front seat of my Corolla to an International Experimental Effort. From my first efforts in 1980 at RCA to a final design equation in 2012. It was long path to follow.

Soon my name in this will be forgotten by everyone and some will claim it their own idea, then I can join Tesla in oblivion. As Peter Lindemann expressed when asked about the widespread use of my material without my knowledge or consent, that Eric is just an A-hole so no one wants to involve him in his own discoveries. My multi-wave oscillator antenna is in wide use through Lindemann, Borderland, etc and I get nothing. Well it is just a small bit of the over 50 Kilobux Borderland embezzled from me. Tom Brown told me that I owed it to (Her), that, (She), needed it more than I did. Well (She) is dead now and the money is spent. The notes, record, equipment of my work is long gone, good buddy Tom Brown. It serves Al-Gora well that all the records of my RCA solar-terrestrial work vanished, so who is helping who?

When I undertook providing this forum with the means of duplicating a part of Tesla's work, this in the form of the C.R.I., it was tacitly assumed that it would be used responsibly, no weapon development, no interference to other services, only legal and responsible use. This requirement has been violated. Moreover flouted attitudes of punk personalities serve to negate my own interest in making my efforts known to this forum, or anyone at all. It in no way serves my own interest to provide any info here, and I am in fact serving as a prostitute on this forum, spreading my legs for a few sheckels to keep the car running. You can rest assured once a coyote is not pinned down it will just vanish into invisibility.

73-DE-N6KPH

Note Jake,
1) This is the starting point of the "World System" (on 400 Kc!), a "Ham radio" net.
2) A.M. stations are pilot signals for the experimenter, this makes it easy and legal to work with Tesla Systems, a receive only development.

COLORADO SPRINGS PRIMARY ENERGY TRANSFER CIRCUIT, MAIN TELLURIC TRANSFORMER.

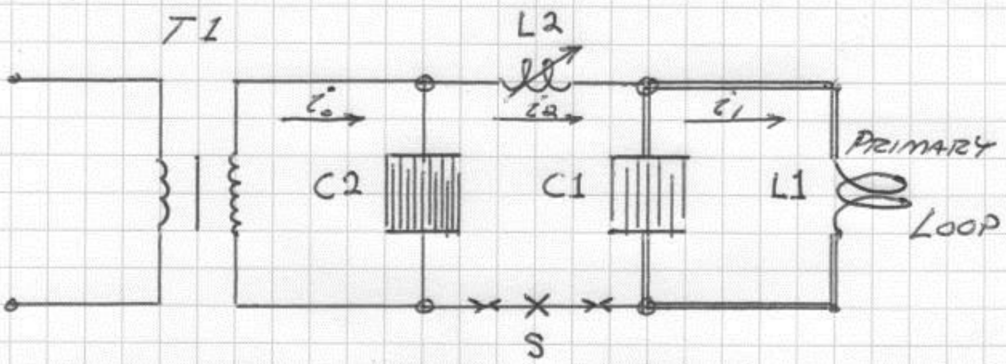

T_1 , 50KVA , 50 KV CONSTANT CURRENT TRANSFORMER.

L_1 , 2 TURN PRIMARY LOOP, CABLED 10GUAGE WIRES.

C_1 , CONDENSER, WATER BOTTLE LEYDEN JARS.

L_2 , INDUCTANCE COIL, ADJUSTABLE.

C_2 , CONDENSER, WATER BOTTLE LEYDEN JARS.

S , CIRCUIT CLOSER, MERCURY TURBINE SWITCH.

RELATIONS

$C2 \geq 3 \cdot C1$, $L2 \leq \frac{1}{3} \cdot L1$, APPROXIMATE

$L2 \cdot C2 = L1 \cdot C1 = \frac{1}{\omega^2}$, EXACTLY EQUAL

$$i_1 > i_2 > i_0$$

06-20-2012, 02:50 PM

t-rex
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Engineering analysis of the Colorado Springs Tesla Magnifying Transformer

Preface: the following material and calculations are taken from the Tesla Colorado Springs Notebook. An index of important pages is given at the end.

1) The Colorado Transformer consists in part of a basic quarter wave resonant transformer, referred to in the notes as the primary-secondary structure. The primary circuit is of the balanced, open loop break, configuration. This break in the primary loop gives rise to steep transient waves. This in turn gives rise to a multitude of harmonic travelling waves upon the coiled windings which constitute the Tesla Transformer. This is discussed in “Condensed Intro to Tesla Transformers.” Like did Tesla, we will focus on the sinusoidal, single frequency, distributions.

2) Next is the extra coil. This coil here operates with a propagation constant less than an eighth of a wavelength. Therefore the coil is operating as a simple inductance coil, not as a distributed network. The distributed network capacity of this mode can be expressed as a definite terminal capacitance. Hence the inductance of this coil is adequately represented by its static inductance.

3) Finally is the capacity mast. The effective electrostatic capacity of this mast is given. To convert from C.G.S. “centimeters” first divide by the speed of light squared, in centimeters per second, then multiply by ten to the plus ninth. This gives Farads. It works out to 1.1 times C.G.S. gives picoFarads. How easy.

The capacity of the mast consists of two components, one is the self capacity to earth in per Farads, the other is the mutual elastivity K to the ionosphere in per Farads. The measured mast capacity is a resultant, the square root of the ratio of self capacity, C, to the mutual elastivity, K. The actual values of C and K are unknown. The square root of the product of C and K is the propagation constant. The smaller the value of this constant, the greater is the electro-static coupling to the ionosphere.

4) It should be noted that the earth connection at this location was very poor. It was not adequate for the system neutrals. This gives rise to stray mutual inductance between neutral connections. This also led to travelling waves on the two wire 1000 volt power line that powered the Tesla system. Standing waves in the distant generator windings shorted out these coils burning out the town’s generator station.

In conclusion, the Colorado Tesla Transformer is most basic. An extremely high electromotive force is established through the employment of a large lumped series resonant circuit. This series circuit consists of a static inductance coil, the extra coil, and a static capacitance, the capacity mast. This is a basic LC circuit.

This LC circuit is fed by a constant current resonant transformer as a source of low frequency alternating current energy. The operating frequency is near 45 Kilocycles.

5) The propagation constants and the transmission impedances can be derived from the basic physical dimensions of this system. It is fortunate that we have the “RadioTron Designers Handbook”, Tesla did not. He had no frequency counter, no scope, no W.W.V. time standard, Nothing! Think about this.

6) **SECONDARY COIL DIMENSIONS AND CONSTANTS**

Diameter: 15 meters
Height: 1 meter
Number of Turns: 17 numeric
Mean Length of Turn: 47 meters
Total Length of Turns: 800 meters
Luminal Wavelength: 3200 meters
Self Capacitance: 1500 picoFarad
Self Inductance: 10 milliHenry
Luminal Frequency: 94 Kc/sec
Free Space Frequency: 64 Kc/sec
Actual Frequency: 43 Kc/sec
Free Space Propagation: 68%
Actual Propagation: 46%
Transmission Impedance: 2500 Ohm
Dielectric Burden: 330 picoFarads

7) **EXTRA COIL DIMENSIONS AND CONSTANTS**

Diameter: 8.4 feet
Height: 8.0 feet
Number of Turns: 100 numeric
Mean Length of Turn: 8 meters
Total Length of Turns: 800 meters
Luminal Wavelength: 3200 meters
Self Capacitance: 112 picoFarad
Self Inductance: 25 milliHenry
Luminal Frequency: 94 Kc/sec
Free Space Frequency: 176 Kc/sec
Actual Frequency: 116 Kc/sec
Free Space Propagation: 187%
Actual Propagation: 123%
Transmission Impedance: 15 Kilo-Ohm
Dielectric Burden: 26 picoFarads

8) **MAST CAPACITANCE EFFECTIVE VALUE:**

Measured Capacitance is given approximately at 320 picoFarads

The self capacitance of the extra coil is given as 112 picoFarad.

The total capacitance is hence given,
Total End Capacity, 432 picoFarad

This end capacity is series resonant with a 25 milliHenry inductance coil. This gives rise to a propagation constant, (the resonant frequency) and to a transfer impedance.

9) The propagation constant, or frequency, is defined as the inverse of the quantity consisting of the square root of the inductance times the capacitance, this root then multiplied by two Pi. Hence cycles per second.

The product,
25 milliHenry times 432 picoFarad
Results in a frequency of
47 Kilo-cycles

10) The transfer impedance is given as the square root of the ratio of the inductance to the capacitance.

25 milliHenry divided by 432 picoFarad
Results in an impedance of
7600 Ohm

This represents the output impedance of the Tesla Magnifying Transformer.

11) The basic circuit is shown in figure 1:

FIG 1

ACTUAL
BASIC LAYOUT

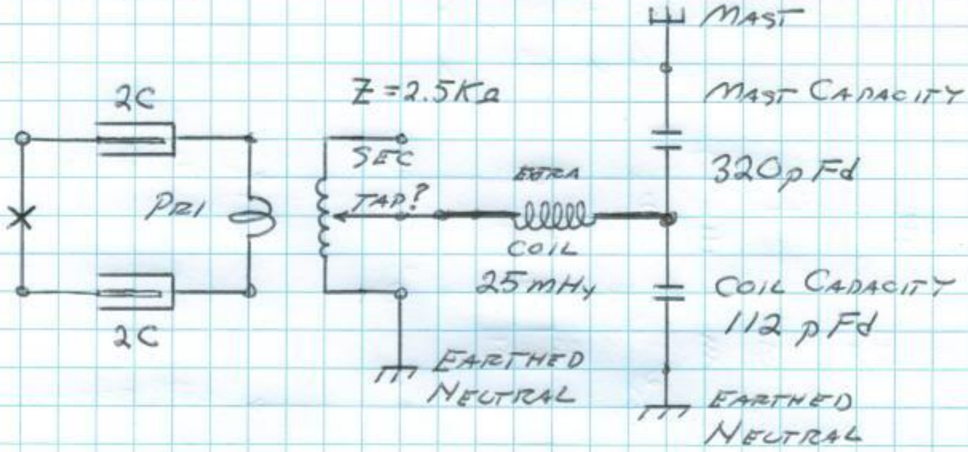
MAST
2C
Z = 2.5 KΩ
MAST CAPACITY
SEC
320pFd
PRI
TAP?
EXTRA
COIL
25mHy
COIL CAPACITY
112 pFd
2C
EARTHED
NEUTRAL
EARTHED
NEUTRAL

EQUIVILENT
BASIC CIRCUIT

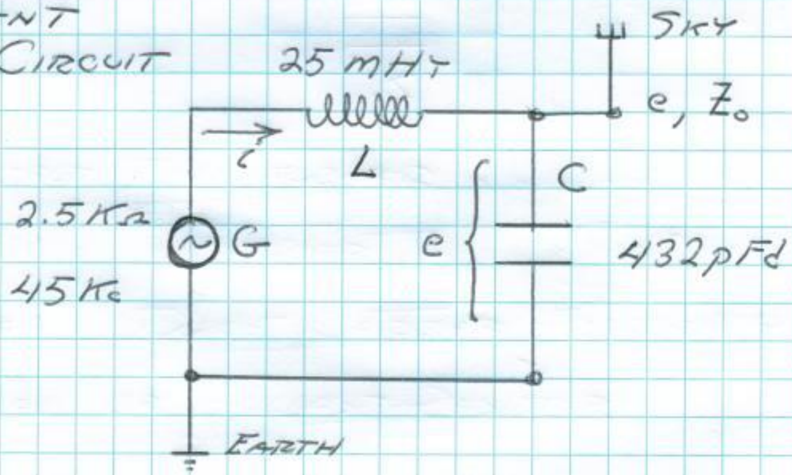
SKY
25 mHY
e, Z₀
i
L
C
2.5KΩ
G
e
432pFd
45Kc
EARTH

$e = i Z_0$ = MAST POTENTIAL

$e/i = Z_0$ = TRANSFER IMPEDANCE

$$Z_0 = \sqrt{\frac{L}{C}} \qquad \omega_0 = \frac{1}{\sqrt{LC}}$$

$$F_0 = \frac{\omega_0}{2\pi}$$

In basic terms, a 45 Kc/sec 2500 Ohm source of alternating current energy energizes a large series resonant circuit consisting of an inductance coil and elevated capacity structure. At 45 Kilo-cycles per second, every ampere supplied to the extra coil gives rise to 7.6 Kilovolts at the elevated capacity terminal. The effects of higher harmonics and distributed constants can raise this potential by no more than 50% over the base 7.5 KV. Hereby, for a capacity potential of 1000 KV, the input current to the extra coil must be 132 amperes at 45 Kilocycles per second.

12) This derives the total KiloVolt-Ampere reactive activity of the extra coil in 1000 KV operation, hence, the total activity;

132 MegaVolt-Ampere

This is a substantial activity of 174 thousand horsepower, remembering that this is an average value, the peak value is unknown.

13) In conclusion, the Colorado Springs Tesla Magnification Transformer is a very simple system, no more than a giant series resonant circuit with an approximately sinusoidal waveform rendered asymmetrical by a moderate dampening constant. The extra coil is not operating as a transmission network of distributed constants, but only as a basic inductance coil. Hence the length of wire on the coil is somewhat **immaterial.** The mast capacitance swamps out the extra coil self capacitance, this suppressing transmission modes of higher harmonics.

This transformer is seen to be somewhat disappointing to the theorist. No fancy travelling wave interaction nor any stupendous harmonic conjunctions, only a **Big Series Resonant Circuit.** It can be surmised that the Colorado Springs Transformer was incapable of full mono-polar operation, due to the suppression of distributed constants. What is fortunate is that with the material hereby provided the Colorado Springs Tesla Transformer can be scaled down to any convenient size for experimentation. This setup is now quantified, and the mystics can remain silent.

73 DE N6KPH

Reference Index (From Colorado Springs Notes)
Page 43 & 58 Primary Circuit
Page 67 Primary Length
Page 203 Primary/Secondary Dimensions
Page 206 Secondary Frequency
Page 211 Secondary Inductance
Page 260 to 267 Capacity Mast
Page 318 Extra Coil Dimensions
Page 357 Extra Coil Ground Frequency
Page 359 Extra Coil Mast Frequency
Page 363 Extra Coil Free Frequency

An excellent addendum to this is the John Miller Paper "Electrical Oscillations in Antennae and Induction Coils" provided with Eric Dollard's paper "Introduction to Dielectric and Magnetic Discharges in Electrical Windings". It can be seen here: http://www.tuks.nl/pdf/Eric_Dollard_...0remake%29.pdf

06-28-2012, 01:06 AM

t-rex ●
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Resonant Magnification

Today we have the good fortune of the permission to experiment in the transformer yard of a department of water and power generating station. The station station is down for maintenance and one bank of line transformers is available for our experiments. This bank has a wye connected secondary of 60 Kilo-Volt line to neutral voltage. The primary windings are in a delta connected 2.4 KiloVolt line to line voltage, these to the generators.

Now one transformer is removed from the bank, its 2.4 Kilo-Volt primary will remain unused. The secondary coil is asymmetrical for grounded neutral connection. Hence the secondary coil has a high voltage bushing for line connection, and a distribution voltage bushing for neutral connection. See Figure (1). This is an "odd order series" winding, or a "quarter wave" type configuration.

This single phase power transformer has the following specifications:
Full Load Capacity: 3000 KiloVolt - Amperes
Secondary Voltage: 60 KiloVolts
Basic Impulse Level: 300 KiloVolts

Measurements on the secondary windings give the following results:
D.C. Winding Resistance: 24 Ohms
60 cycle Winding Reactance: 60 KiloOhm

Hence the Inductance is derived as: 160 Henry

And the ratio of reactance to resistance of this winding is derived as:
2500 Numeric

A 60 cycle condenser bank is now constructed to give an equal and opposite reactance to that of the 60 KV secondary coil. At 60 KV the reactive power in this winding is:
60 KiloVars

Thus the condenser must also develop a counteracting power,
60 KiloVars

And this results from a condenser reactance of:
60 Kilo-Ohm
The same as that in the transformer winding. Finally, this condenser bank must withstand the basic impulse level developed by the secondary winding, a maximum of
300 KiloVolts

Hence established is a series connected condenser bank, of ten 30 KV units, each with a capacity of
.4 microFarads

Or a total of .04 microFarads.

Here now in the transformer yard is our experimental network, about 50 tons in weight.

The test voltage in the experiments will be 120 volts AC, 60 cycles per second, from a lighting transformer.

The first experiment is shown in Figure (2). The 120 volts is connected to the secondary winding of the transformer. The current is the voltage divided by the reactance,
2 mA

Now in place of the transformer winding, the condenser is connected the 120 volt supply. The current is also
2 mA

The Back E.M.F. of both the coil and the condenser are equal since the reactance of the coil and condenser are equal. Here however it is equal and opposite.

The next experimental setup is shown in figure (3). Here now the condenser is connected in series with the winding of the transformer. It is found that a current of
5 Amperes

is now being delivered by the 120 volt supply. The transformer is emitting a deep, powerful, growl and the condenser bank is crackling with static sparks. The voltage on the line bushing is measured to be
300 KiloVolts.
The maximum voltage of the transformer.

HOLY TESLA MR WIZARD!! WHAT HAPPENED?!!

The electrical network utilized is shown in figure (4). It is a basic series RLC circuit;
Resistance, 24 Ohm
Inductance, 160 Henry
Capacitance, 0.04 microFarad

Shown in figure (4) is the remarkable condition that the full line voltage of 120 volts appears across the resistor, just as if the coil and condenser **DID NOT EVEN EXIST**. Also interesting is that the high voltage of the coil counteracts the high voltage of the condenser giving a net voltage of zero.

The Electromotive force developed by the coil is the current times the reactance,
5 Ampere, 60 Kilo-Ohm
300 KiloVolt

Likewise the electrostatic potential contained by the condenser is the current times the reactance,
5 Ampere, 60 Kilo-Ohm
300 KiloVolt

The real power of energy consumption is the product of the E.M.F. and current dissipated by the resistance
600 watt

The image power of energy storage is the product of the voltage developed times the current developing the voltage, E.M.F. or potential
1500 KiloVolt-Amperes

Taking the ratios gives the voltage magnification,
300 KV per 120V,
or 2500 numeric

And the power magnification,
1500 KVar per 600 Watt, or
2500 numeric.

To summarize, today in the D.W.P. transformer yard we simulated a 60 cycle Tesla Transformer in lumped parameter form. This "Tesla Coil" developed a maximum potential of 300 KiloVolts with an activity of 1500 KiloVars. Its magnification factor is 2500. This system is not able to transmit however because the sum of all the currents in this setup are equal to zero in accord with Kirchoff's Law. This is a simulation only to demonstrate the basic principles of resonance as developed by Tesla.

73 DE N6KPH

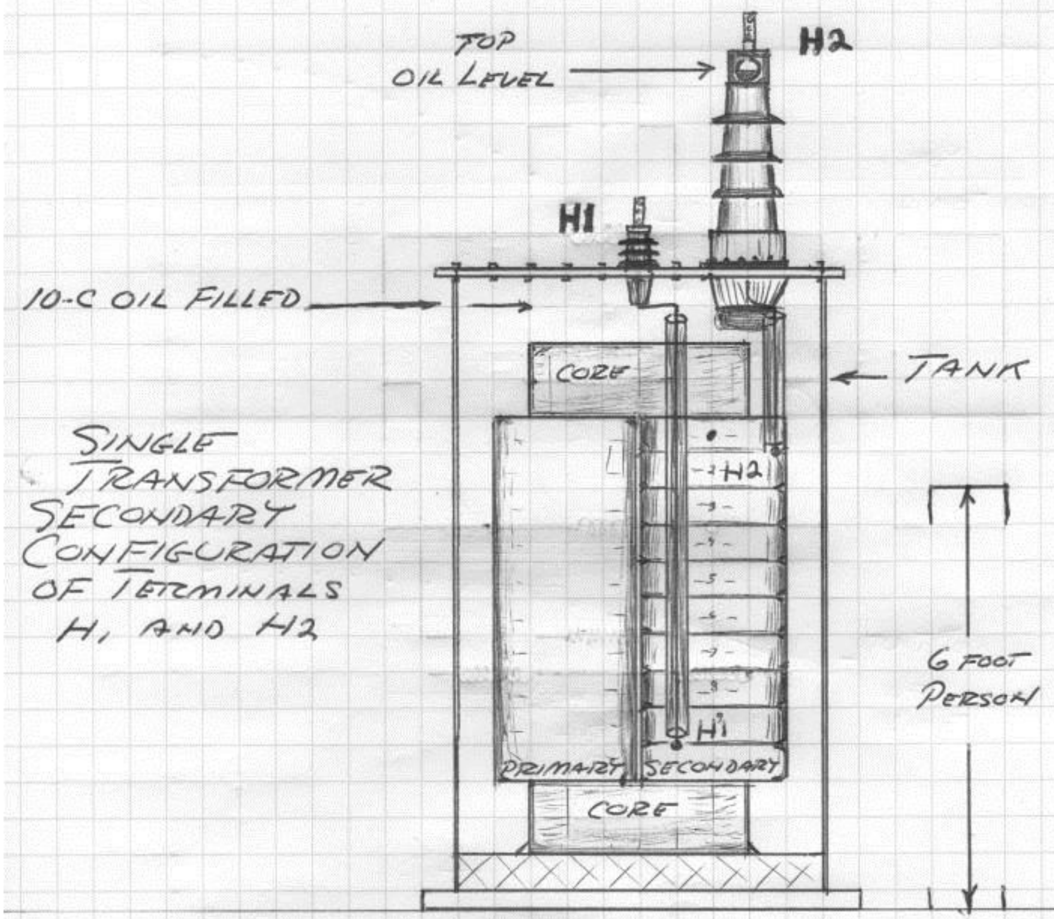
TOP OIL LEVEL
H2
H1
10-C OIL FILLED
TANK
CORE
SINGLE TRANSFORMER SECONDARY CONFIGURATION OF TERMINALS H1 AND H2
H2
3
4
5
6
7
8
H1
6 FOOT PERSON
PRIMARY
SECONDARY
CORE

SINGLE PHASE TRANSFORMER

3000 KVA

60,000 VOLT

2400 VOLT

RATIO 25:1

FIGURE (1)

FIGURE - 2

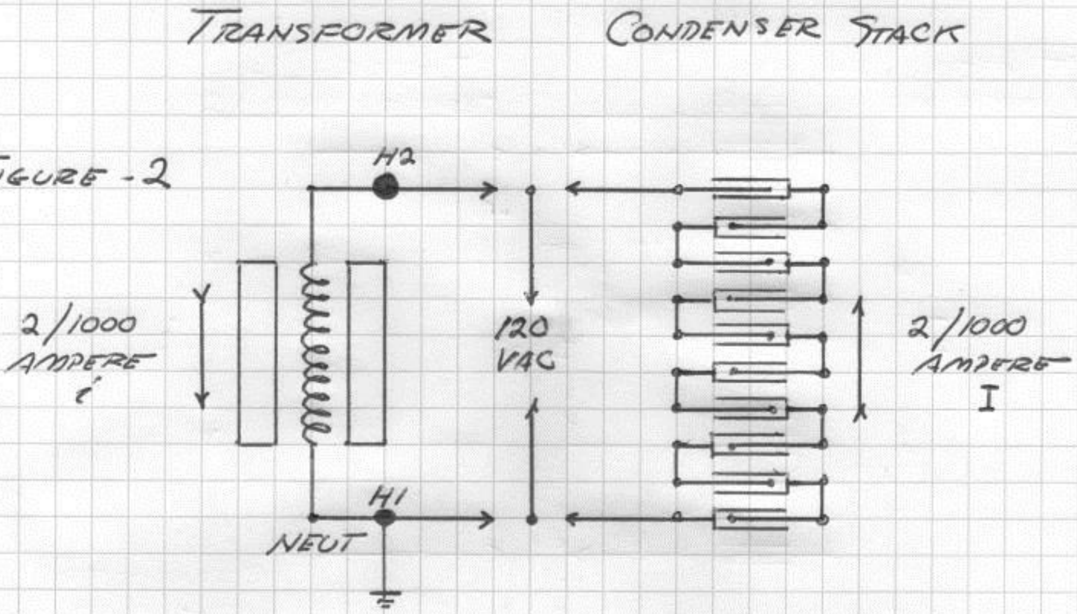
TRANSFORMER
CONDENSER STACK
H2
2/1000 AMPERE i
120 VAC
2/1000 AMPERE I
H1
NEUT

FIGURE - 3

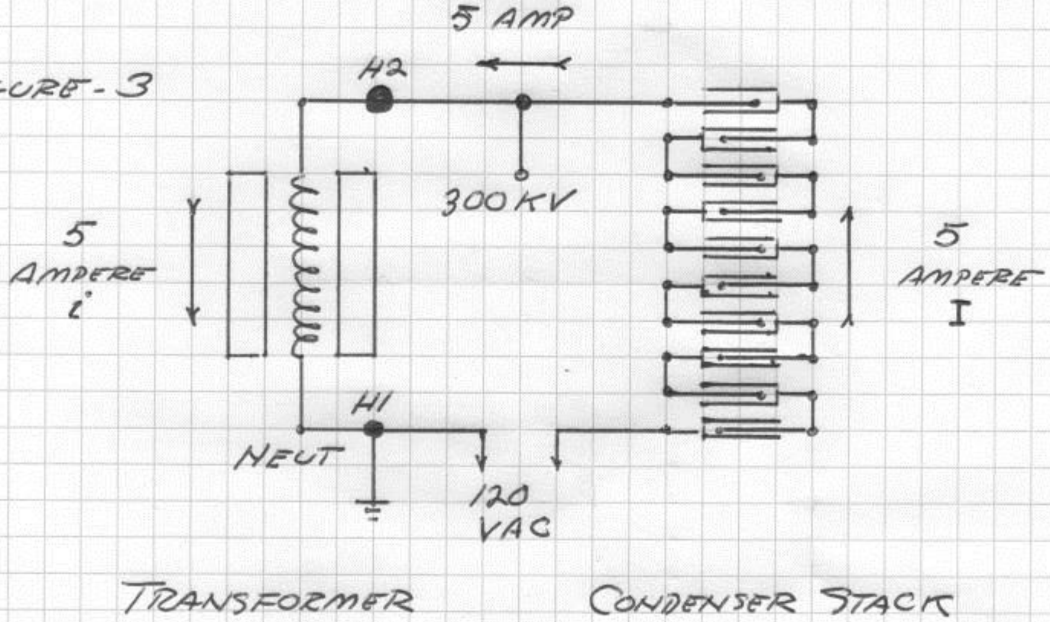
5 AMP
H2
300 KV
5 AMPERE i
5 AMPERE I
H1
NEUT
120 VAC
TRANSFORMER
CONDENSER STACK

CIRCUIT CONNECTIONS

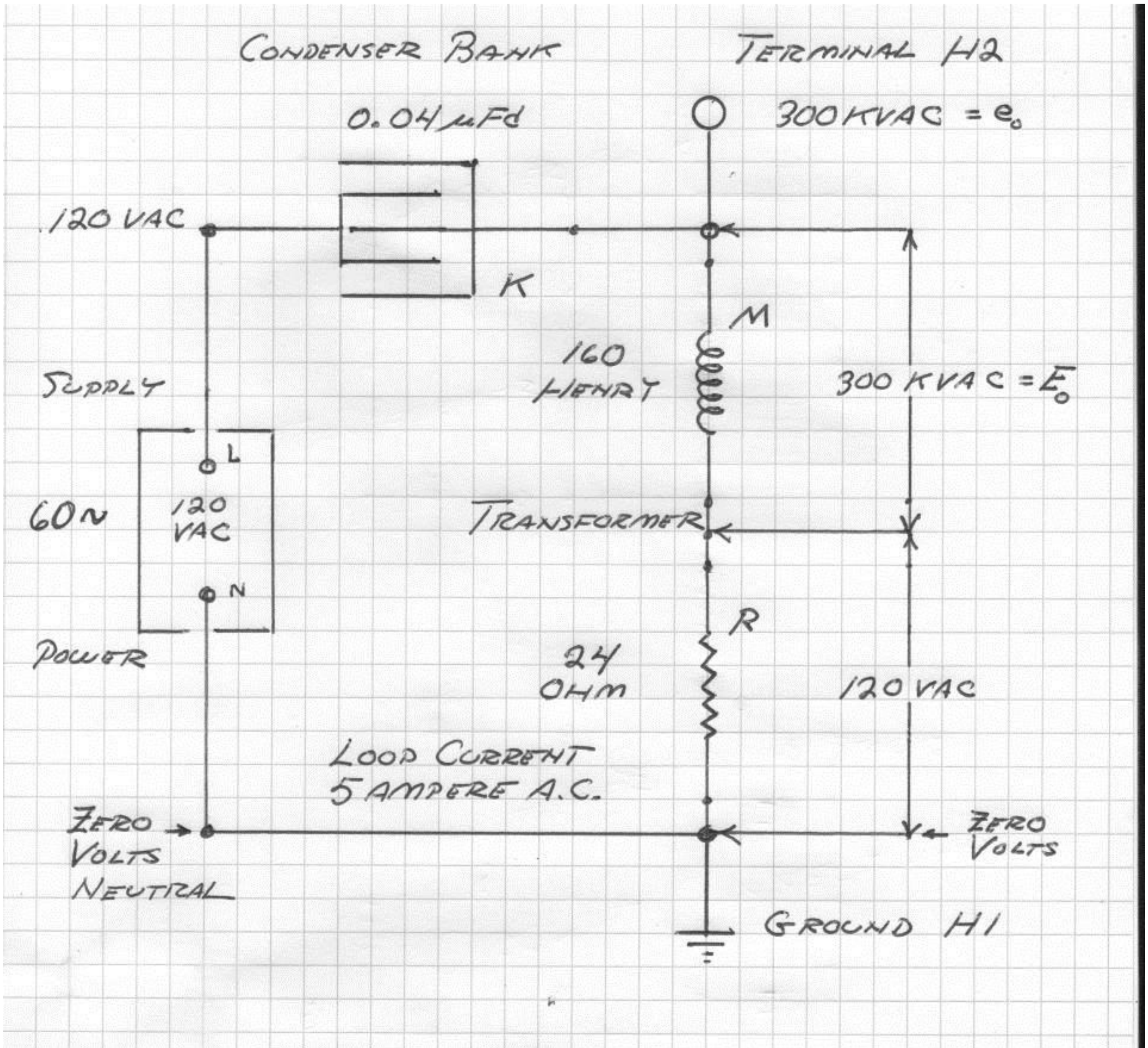
CONDENSER BANK
TERMINAL H2
0.04 µFd
300 KVAC = e_0
120 VAC
K
M
160 HENRY
300 KVAC = E_0
SUPPLY
L
60~
120 VAC
N
TRANSFORMER
POWER
R
24 OHM
120 VAC
LOOP CURRENT
5 AMPERE A.C.
ZERO VOLTS NEUTRAL
ZERO VOLTS
GROUND H1

MK SERIES RESONANT CIRCUIT

FIGURE - 4

06-29-2012, 03:09 AM

t-rex
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

60 KV Transformer

(1) Experiments continue today with the 3000 KVA 60 KV generating station line transformer. No information exists on the exact distributed constants within this transformer therefore the constants will be developed by the use of effective values as externally measured.

A source of variable frequency alternating current is applied to the 2.4 KV primary winding of the line transformer. The neutral, H1, of the secondary winding is grounded. An electro-static voltmeter is connected from neutral, H1, to line, H2, of the 60 KV secondary winding. It is found that the fundamental quarter wave frequency is 716 cycles per second. This is the natural frequency of the transformer.

(2) The scalar inductance was determined by 60 cycle measurement to be 160 Henry. Since the current in the winding in resonance (self) is of a sine distribution along the winding length the effective value of inductance is two divided by Pi times the scalar inductance. Hence the effective inductance is derived:
100 Henry

(3) For a frequency of 716 cycles per second the angular frequency in radians per second is two Pi times 716 giving:
5000 radians per second

The product of the inductance and the capacitance of an oscillating circuit gives the inverse of the square of the angular frequency. By algebraic operation it is hence the capacitance in Farads is given by the inverse of the product of the inductance and the square of the angular frequency. Here derived is the effective capacitance,
500 picoFarad

And since the electro-static potential also is a sine function along coil length multiplication by Pi divided by two gives the total capacitance,
785 picoFarads

These values are for 716 cycles per second only, the Pi over two, or quarter wave resonance, of the winding.

Hereby the basic energy storage constants of the transformer secondary are established for a condition of self resonance. Now rather than a 60 cycle power transformer it is a 716 cycle resonant, or Tesla, transformer.

(4) In this self resonant mode this transformer not only has a frequency in radians per second, but also has a wavelength in radians per centimeter. At 60 cycles the winding is space scalar, that is, the current is the same in each and every element of length along the winding conductor. At 716 cycles the current not only varies along the length in intensity, but also varies in phase. In terms of effective values the current at the neutral, H1, is not equal to, or in phase with, the current at the line, H2. Also at quarter wave the current is zero at the line terminal since this terminal is not connected but left free. The circuital laws of Kirchoff no longer apply and the currents no longer

sum to zero. Hereby a mono-polar current appears to ground at the neutral, H1, of the secondary winding.

(5) In order to analyse this transformer winding as a transmission network two factors must be known:

I) The propagation constant as derived from the product of the effective inductance and the effective capacitance,

II) The characteristic impedance as derived from the ratio of the effective inductance and the effective capacitance.

The propagation constant is defined by the frequency, or per second. The characteristic impedance is derived from the square root of the ratio of the effective capacitance. While the frequency is invariant if the Miller, or Steinmetz methods of determining the effective values, the impedance is not and will differ between methods of deriving effective values. For simplicity the Steinmetz method is used here. Taking the effective inductance as,
100 Henry

And the effective capacitance as,
500 picoFarad

Gives the characteristic impedance as,
450 Kilo-Ohm

This can be called the natural impedance of the secondary winding at its natural frequency of 716 cycles per second.

These are the fundamental externally established constants for the 60 KV secondary winding.

(6) Hereby the input vs output transformation of this secondary winding can be determined. The winding can now be considered a "Black Box" two terminal reactance arm, this akin to the Guillemen line shown previously. In this network of coiled windings the output potential is the input current multiplied by the natural impedance. Hence for a current applied to neutral (H1) terminal, one ampere multiplying this by 450 Kilo-Ohms gives a potential at line terminal (H2) as
450 KiloVolts

This causing the bushing spark gap to flash over. Since the B.I.L. of this transformer is 300 KV then this potential divided by the natural impedance derives the maximum permissible neutral current as,
.667 Ampere

At a frequency of 716 cycles per second. In this manner, when excited through the neutral to develop a high potential at the line, the 60 KV secondary winding is operating as an extra coil. The primary has no function in operation and therefore this is no longer a resonant transformer.

Thus this 60 KV transformer can serve two distinct functions:

I) A resonant transformer with a low voltage primary and a high voltage resonant secondary,

II) an extra coil as a single resonant high voltage winding series fed through its neutral.

73 DE N6KPH

06-30-2012, 12:39 PM

t-rex

Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

60 KV Transformer, Theory:

(1) In the previous experiments two distinct forms of resonance have been demonstrated.

I) Lumped parameter resonance, circuital

II) Distributed parameter resonance, wavelength

In condition one the dielectric energy storage coefficient, capacitance, is external to the winding structure. The current IN at H1 is the current OUT at H2. There is no difference in magnitude, nor any displacement in time between the currents at H1 and those at H2, the same current.

In condition two the dielectric energy storage is within the winding structure and that in the space between the winding and the core and tank. This can no longer be expressed as a capacitance in the normal sense of the word. Dielectric energy storage is now a complex quantity in the dimension of space. For the most part is is inscrutable and all that can be measured outside of the tank is the effective capacitance at a given operating frequency. A graphic illustration is given by Tesla in his "True Wireless" extract. The effective capacitance of a coiled winding is the square root of the ratio of the capacitance, C, from winding to core and tank, and the elastance, K (in per Farad), existing within the interturn counterspace. Hence, Farad per, per Farad, the square root of which gives the effective capacity. In static or D.C. operation the energy storage coefficient, K, vanishes, leaving only the self capacitance, C. This measure as the static capacitance of the transformer winding. Hence the effective capacitance is a function fo the time rate of energy exchange. It is important to recognize here that a Pure Dilectric Wave exists. The space distribution of dielectric induction is given by the propagation constant. This constant is a Time Scalar, that is, the formation of this dielectric distribution is instantaneous from end to end on the coiled winding. No magnetic induction exists here.

(2) The lumped parameter condition lends itself to circuital representation. Series coefficients are

given as inductance, L, and elastance, K, Henry and per Farad. Shunt coefficients are given as Enductance, M, and Capacitance, C, per Henry and Farad.

The components that comprise the circuit are lumped, or Space Scalar. The shape, size, geometry, and etc. play no definite role in the energy storage coefficients. In the magnetic "components" the dielectric field is forced out of existence, and likewise in the dielectric "components" the magnetic field is forced out of existence. The magnetic wants an expansive space, the dielectric wants a contractive space. See "Disruptive Discharges" in my Four Quadrant writing. Review This.

(3) The distributed parameter condition CANNOT be given as a circuital representation. Inductance and Capacitance lose their common meaning and can only be expressed as effective values, such as derived from space integrals. The magnetic and dielectric energy storagecoefficients now exist in a different set of dimensions, as space derivatives of the energy storage coefficients. Dimensionally it is;

Magnetic
I) Distributed Inductance: Henry per centimeter
II) Distributed Enductance: per Henry-centimeter

and Dielectric
III) Distributed capacitance: Farad per centimeter
IV) Distributed Elastance: per Farad-centimeter

The designations, C, K, L, M, are to be taken on this per centimeter basis when expressing energy storage in distributed form.

It is of interest to note here that, L, and, C, now are dimensionally equivalent to the magnetic permeability, and the dielectric permittivity,

Mu, or Henry per centimeter
Epsilon, or Farad per centimeter

Hence these coefficients, L, and, C, are akin to the one over C square relation of electro-magnetic wave propagation. Conversely M, and, K, are of a unique dimensional constitution and K, is more akin to a sound wave condition. The product of, M, and K, do not yield an electro-magnetic condition, there is no per velocity square as given by the product of, L, and, C. The dimensional relation of, M, and, K, yield a magneto-dielectric condition as a per counter-velocity square. Hence if velocity is given as

C, in miles per hour

It is that counter-velocity is given as

u, in per mile per hour.

These propagations exist at right angles to each other on the coiled winding in a Four "dimensional" space as the physic would like to say. This complex configuration defies any existing analysis.

(3) Here in the winding with distributed constants exists a transmission line condition, propagation is involved from one end of the winding, H1, to the other end of the winding, H2.

It is however, that this propagation not only exists in space, such as a two wire phone line, but now also exists in counter-space. The dimensional relation of velocity no longer has meaning in defining this propagation. These considerations are the primary stumbling blocks to resonant transformer calculations. The physicist declares the situation as "incapable of being treated by mathematics" and thus declares it impossible to exist. Then comes ridicule, and finally totalitarian suppresion, AKA Al-Gorda.

4) Now it is left to the experimenter. Tesla knew very little of what has just been given here on the resonance of windings. He had one foot in lumped constants, and the other foot in distributed constants, and no measuring instruments that are taken for granted today. It can be plainly seen that here is a transmission line problem, just taken to the next level. Basic two wire line theory is almost beyond human conception as plainly given by Heaviside, Steinmetz, and Guillemen, in thousands of pages. Now it goes octaves deeper.

73 DE N6KPH

06-30-2012, 03:52 PM

t-rex ●
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

60 KV Transformer, Final

(1) Given in the previous transmissions is that there exists three distinct modes in which this transformer can operate.

I) As a power transformer for machine frequencies. The inductance is space scalar and the dielectric energy content is negligible.

II) As a resonant transformer for VLF frequencies. The inductance is a sine function of space and a considerable quantity of dielectric energy content exists, equal to the magnetic energy content. The capacitance is now a cosine function of distance. This resonant energy exchange is powered by a loose coupled source of Alternating Current through the primary winding.

III) As an extra coil for VLF frequencies. Again equal energies in a resonant exchange but now series fed through the neutral to earth connection. The primary coil now is inactive in this operation.

2) Some further experiments are in order here.

The DC, or static, resistance of the 60 KV secondary winding is measured to be:
24 Ohm

Handy is a car battery:
12 Volt

Now connecting the line, H2, terminal to its line vacuum breaker, Figure 1, and connecting the line side of the breaker to the positive terminal of the car battery, a basic current loop is formed.

The negative terminal of the car battery is connected to an ammeter and then to the transformer neutral. The neutral, H1, is grounded to station ground.

3) The vacuum breaker is closed completing the current loop, this loop a basic, L,R, circuit of 24 Ohm and 160 Henry. Upon closure a stopwatch counts time from the instant of this closure, T, equals zero. Observing the ammeter it is seen to rise very slowly. At the point when the ammeter reads
316 milliAmpere

The stopwatch is checked. The span of time to reach 63.2 percent of the maximum value is called the magnetic time constant, u. The stopwatch reads
6.7 seconds

The ammeter continues to climb, in ever smaller increments closer and closer to 500 milliAmp. It is noticed it takes over 33 seconds for the inductance to reach full charge. See figure 63.

At full charge the winding current is one half Ampere. The static winding inductance is 160 Henry. Hence the magnetic energy storage is derived as
20 Watt-second

4) Now the vacuum breaker is tripped, this instantaneously breaking the loop. A streamer several feet long erupts from the line, H2, bushing insto space. Since no escape of energy exists the magnetic energy is abruptly converted into dielectric energy at a fundamental frequency and its odd order harmonics. This was given as
716 cycles per second

The effective values of inductance and capacitance were determined to be 100 Henry
and
500 picoFarad

By the law of conservation of energy the magnetic energy equals the dielectric energy. By algebraic operation the ratio of EMF to current is the square root of the ratio of 100 Henry and

500 picoFarad,
450 Kilo_ohm

And thus from a 12 Volt car battery the EMF developed at the line terminal, H2, is derived as
225 KiloVolts peak

This assuming a sinusoidal waveform. Here the harmonics were burned off in the streamer issuing from H2.

(5) The increase in frequency from 60 cycle to 716 cycle gives an increase of the effective resistance to 72 Ohm, at 716 cycles per second. The effective inductance is 100 Henry. The reactance of this inductance at the oscillating frequency, Omega,
4500 radians per second

is derived to be Omega L or
450 Kilo-Ohm

And the time constant of the 716 cycle oscillation is derived from the effective inductance in proportion to the effective resistance, L, over, R. It is however that the oscillatory energy is magnetic only half the time, the other half is dielectric. Hence the time constant is given as one half the ration of inductance to resistance, one half of
1.39

Gives the oscillatory time constant as,
.7 seconds

At this time the magnitude of the oscillation has diminished to one over Epsilon, or
36.8 percent

The oscillation becomes inaudible in about five times this, or,
3.5 seconds

See figure 60.

(6) The magnification factor of this oscillating winding is determined by the ratio of the effective winding reactance to the effective winding resistance,
450 Kilo-Ohm
to
72 Ohm
Giving
6250 numeric

The initial, or peak, current is one half an ampere this the effective value, or per square root of two, is given,
356 milliAmpere RMS

The initial, or peak, EMF is given as 225 Kilo-Volt, hence the effective value is,
159 Kilo-Volt, RMS

The product of these values gives the total oscillatory activity in the transformer winding,
56.6 Kilo-Volt-Ampere

(7) The losses in the windings are given by the square of the effective current times the effective resistance, I squared R, or the Back EMF developed by the losses is the current times the resistance, I, times, R, or

25.6 Volt

Here from are derived the basic transformation parameters in the operation of the 60 KV secondary winding as a resonant extra coil, this for an output potential of 160 KV at 716 cycles per second.

Input (H1) Power Requirement
9.1 Watt

Input Impedance, E, over, I, or,
72 Ohm

This a resistive load as seen by the source of alternating electric energy. Hence, a 10 watt audio amplifier, on the 50 Ohm output tap, will give rise to a resonant oscillation, building up over a time span of 3.5 seconds, this with an activity of 56.6 KiloVars and an electrostatic potential of 160 KiloVolts. Here established is the concept, -- "Magnification Transformer"

73 DE N6KPH

FIGURE 1

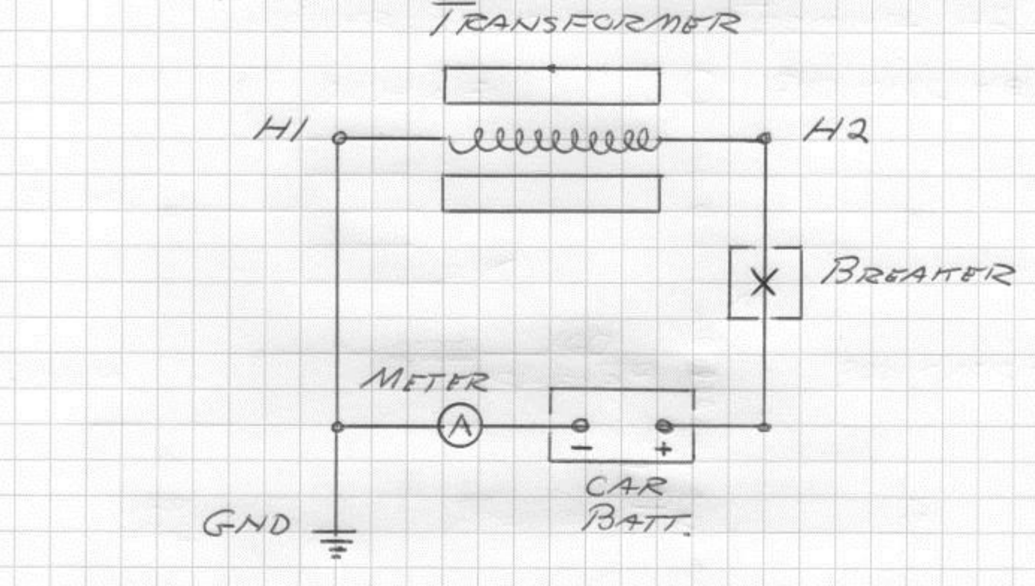
TRANSFORMER
H1
H2
BREAKER
METER
A
−
+
CAR BATT.
GND

IMPULSE DISCHARGE

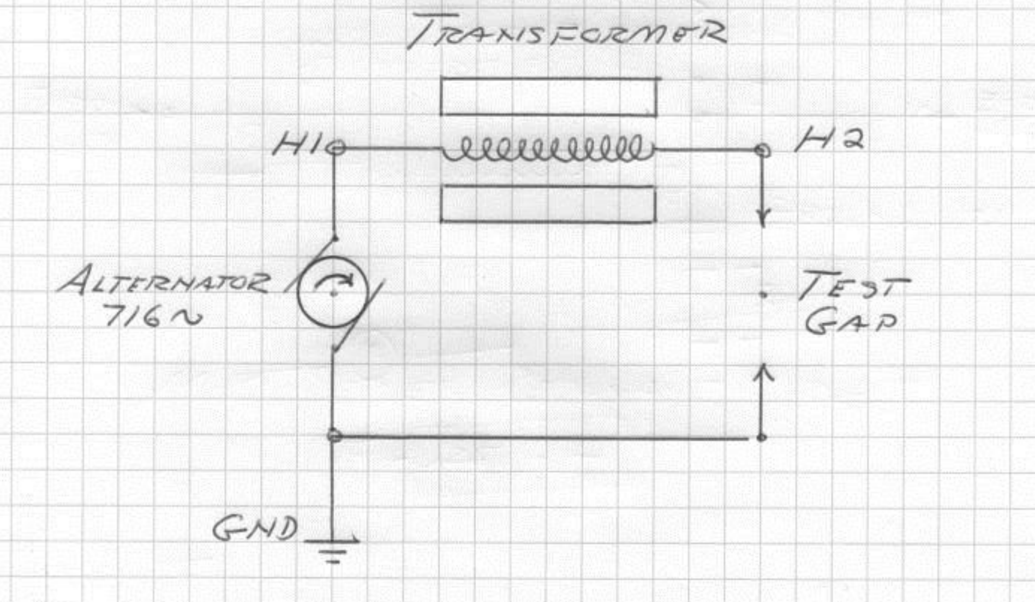
TRANSFORMER
H1
H2
ALTERNATOR 716 ∼
TEST GAP
GND

EXTRA COIL RESONANCE

From Perrine's book "Physics and Mathematics in Electrical Communications":

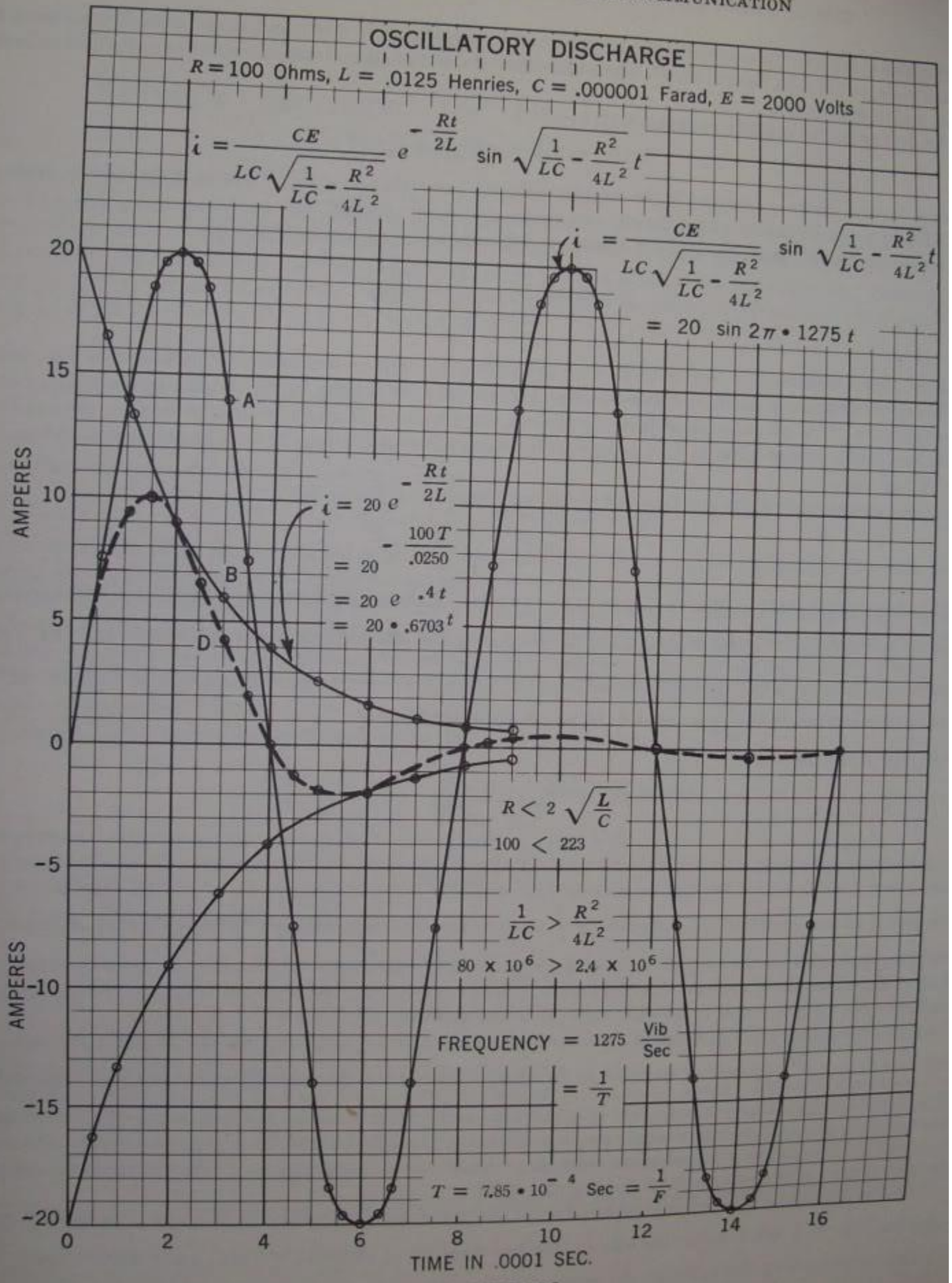
OSCILLATORY DISCHARGE
$R = 100$ Ohms, $L = .0125$ Henries, $C = .000001$ Farad, $E = 2000$ Volts
$i = \frac{CE}{LC\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}} e^{-\frac{Rt}{2L}} \sin \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}\, t$
$i = \frac{CE}{LC\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}} \sin \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}\, t$
$= 20 \sin 2\pi \cdot 1275\, t$
A
B
D
$i = 20\, e^{-\frac{Rt}{2L}}$
$= 20^{-\frac{100T}{.0250}}$
$= 20\, e^{.4t}$
$= 20 \cdot .6703^{t}$
$R < 2\sqrt{\frac{L}{C}}$
$100 < 223$
$\frac{1}{LC} > \frac{R^2}{4L^2}$
$80 \times 10^6 > 2.4 \times 10^6$
FREQUENCY = 1275 $\frac{\text{Vib}}{\text{Sec}}$
$= \frac{1}{T}$
$T = 7.85 \cdot 10^{-4}$ Sec $= \frac{1}{F}$
AMPERES
AMPERES
20
15
10
5
0
-5
-10
-15
-20
0
2
4
6
8
10
12
14
16
TIME IN .0001 SEC.

FIGURE 60

POSTS FROM THE CRYSTAL RADIO MAGIC THREAD STARTED BY JAKE

http://www.energeticforum.com/renewable-energy/10731-crysta-radio-magic.html

03-03-2012, 08:54 AM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Crysta Radio Magic

Well it's not magic(sorry 4 the deception) but it is a radio, my first. It works and I can hear the one station I was hoping for; 1620kHz, 10,000 watts, 6 miles away.

I started this because of
Quote:

> Originally Posted by **T-rex**
> *Here is how the contest works. We learn how to make a "Crystal Set", just as everybody should. It is a "Rite of Passage". A crystal set is an A.M. radio that uses no battery, the magnification factor of its tank circuit powers the radio. Hence it can be seen that the A.M. broadcast station's transmitted energy is powering the crystal set, JUST AS TESLA ENVISIONED. Wow Mr. Wizard that is fantastic. Let's start today.*
> *Break more to follow*

Problem: It was easy, very easy. No matter what I do I can hear this station. I was expecting to have to "tune in" this station. So instead I minimized the components. An inductor(300uH, 60 turns, 20gauge), 1N34 germanium diode, crystal ear bud, and antenna(just about everything works(10 feet of 20 gauge, my body, cookie sheet, and even a beer can)) Some work better than others.
No capacitor??? Holding the coil is the best capacitor I found.
No ground. If I connect it to ground I either get a 60Hz hum when I use a water pipe or lower volume if I stick 2 feet of 6 gauge in the ground. I guess that means I need a better ground. Off to home depot for 10' water pipes.

Two questions for the group: (the goal is to maximize the reception of 1620kHz with only a crystal ear piece)

-There are a number of online calculators that allow you to calculate capatance for a given inductance or inductance for a given capatance, both for a given frequency? What "ratio" works best? (is it better to have more capatance or more inductance).

-Ground-What is good enough? Does it keep getting better as your ground system grows and gets deeper. Is it better to go deeper or wider(radial star like Dollard describes) I have lots of clay around here and I am going to try to water drill with the copper pipe as it goes down. How much better would 20' be over 10'?

Ok more than 2 questions but I'm learning.

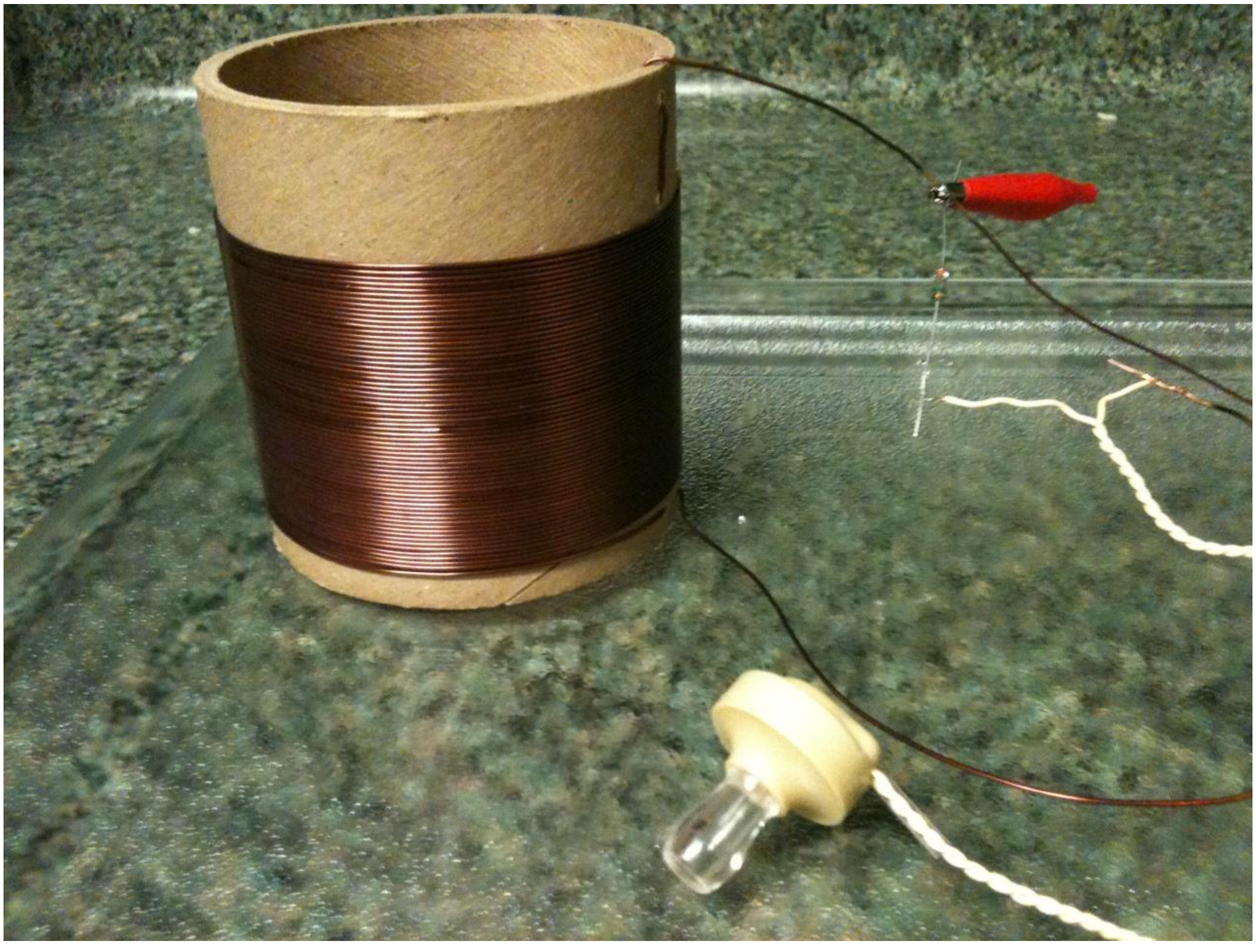

03-03-2012, 12:46 PM

gyula
Silver Member

Join Date: Apr 2007
Posts: 753

Hi Jake,

Your biggest "problem" (may sound strange) is your closeness to the 10kW AM station... I mean the station makes too much field strength this is why almost anything behaves like an antenna, touching the coil makes big changes, etc.

First I would suggest using a coupling coil of 10-15 turns, placed to the bottom of your big air core coil, this coupling coil upper wire end would be connected to a 1 or 2m long wire (this latter serves as an antenna) and the lower end would go to Earth. Do not connect the big coil lower end to the coupling coil's earth point, ok?
Then you should find the capacitor needed for resonating your big coil to be a parallel LC circuit at 1620kHz. If you have now 300uH, then you would need 32.2pF to tune it to 1620kHz, the best would be to use a 40pF or 60pF trimmer capacitor. See link later for online LC calculator.

Then see this schematic here and I suggest making another 20-22 turn coil (not critical) onto your big air core coil, also to its bottom part, maybe over to the coupling coil mentioned above, this newer coil would serve as the L4 coil does in this schematic:
http://hibp.ecse.rpi.edu/~john/hi_power_xtal_set.gif
and you would need the diodes D1, D2, C3 and C4 of course, your crystal earpiece would connect to points A and B directly. (Do not connect this second coil to the earth point either.)

See this link where the schematics is referred to, it may be useful reading for you: High-Power Crystal Radio

I think your water tap earth is good enough but connect it ONLY to the antenna coupling coil I mentioned making first above, then you should not have 60Hz hum.

Another possibility is to use an AM ferrite antenna, like seen here:
Scotts Electronics Ferrite Loopstick Antenna Coil Crystal Radio Coils AM Coils but if you google you can find elsewhere too. Of course you do not have to make any antenna coupling coil for it, the Litze wire coil on the rod completes the antenna and if you choose the 680uh Litz wire coil with a tap at 470uh about 20 turns from end, then this 20 turn can serve as L4 for the diodes (you can separate the two wires which led out as a twisted wire to make the tap). This 680uH ferrite antenna needs 14pF capacitor across it.
I used this online calculator:
LC Resonance Calculator
(Notice the L value needs to be entered in nH, not in uH and f is in MHz.)

Normally it is better here to use higher L and lower C values but because of you want to receive a station at the very edge of the AM band the usual coils used for AM reception seems too high in value this is why so small pF value the tuning capacitor has (14pF or 32pF). You may wish to remove say 10-15 turns from your 60 turn air core coil to reduce its inductance and then increase the tuning cap higher or when using ferrite antenna the coil on the rod can be slided a bit off towards and / or beyond the rod's edge. The small cap values needed for 1620kHz resonance with such 300-680uH coils explain why you found good reception when you were holding the coil: your hand and body capacitance, your close presence to the coil supplied the 32pF tuning...

Hope this helps,

Gyula
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#3

03-03-2012, 03:03 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Extra thanks

That is exactly the kind of advice I was looking for. I appreciate it very much. I suffered information overload when I tried searching for the "ideal" capatance/inductance ratio. There are just so many ways to do it there does not seem to be an ideal way.

Im off to snip, tap, wind, and take off some turns.

Thanks,
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

#4
03-06-2012, 09:07 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Whats the ground for?

I mean: I know what the ground is for but if you can receive a radio signal fine without a ground connection what additional benefit would you get if you attache to a good ground??
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

#5
03-07-2012, 02:54 PM

gyula
Silver Member

Join Date: Apr 2007
Posts: 753

IF you can receive a station fine without ground connection, then I do not think you need ground for THAT station. I think it all depends on the sufficient field strength for a radio station at a particular location, a good ground can contribute to a much better reception (in a given receiving

antenna case) when you wish to listen to a distant station which does not have a strong field 'presence' at your location.
What additional benefit? In your 6 miles away station case probably nothing (I have never had such chance lol), though you could probably use an even shorter piece of wire when you have a good ground. Remember the station transmitter antenna for the AM bands usually works against a big ground system and an isolated metal tower of a given vertical length is the radiator. So in a distant station case you can benefit from grounding because radiated station energy is taken also from the ground, not only from the 'air'.

Gyula
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#6
03-07-2012, 04:35 PM

dR-Green
Gold Member

Join Date: Apr 2008
Location: Merlin's Fort
Posts: 1,220

Quote:

> Originally Posted by **jake**
> *I mean: I know what the ground is for but if you can receive a radio signal fine without a ground connection what additional benefit would you get if you attache to a good ground??*

A Tesla receiver won't work without the ground because it's designed to receive the underground signal. Herein lies the fallacy of so-called "wireless energy transmission"
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

"Knowledge is cosmic. It does not evolve or unfold in man. Man unfolds to an awareness of it. He gradually discovers it." - Walter Russell

"Once men died for Truth, but now Truth dies at the hands of men." - Manly P. Hall

03-25-2012, 07:18 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

next generation crystal radio

So once I felt confortable with a typical crystal radio and could tune most of the stations on the high end of the AM band. I started to put together a Tesla Resonant Transformer. This is part of the crystal initative on the Dollard Thread.

Below are the results of three generations of coil formers. The former you see is stout and will allow you to pull the wire tight while winding. I am still missing the primary, topload, capacitors. BUT

It blows my crystal radio with tuning capacitor away. Attached are some pics. The DSO pic is my night time reception it came down from 1.3v earlier this afternoon with +4v peaks. This got me excited for a while but then I realized that my scope maxes out at 1MHz. Your looking at 1.62MHz rectified
audio. 100us/div and 50mv/div, the yellow line is 500mv.

Do these numbers mean anything or is working out of the range of my scope produce garbage numbrs???

And finally I was able to dimly light an LED in line with the germanium diode.

Any one know how to hook up a joule theif without loading down the circuit??

It looks like there is enought there to run one but doing it without loading down the circuit is eluding me.

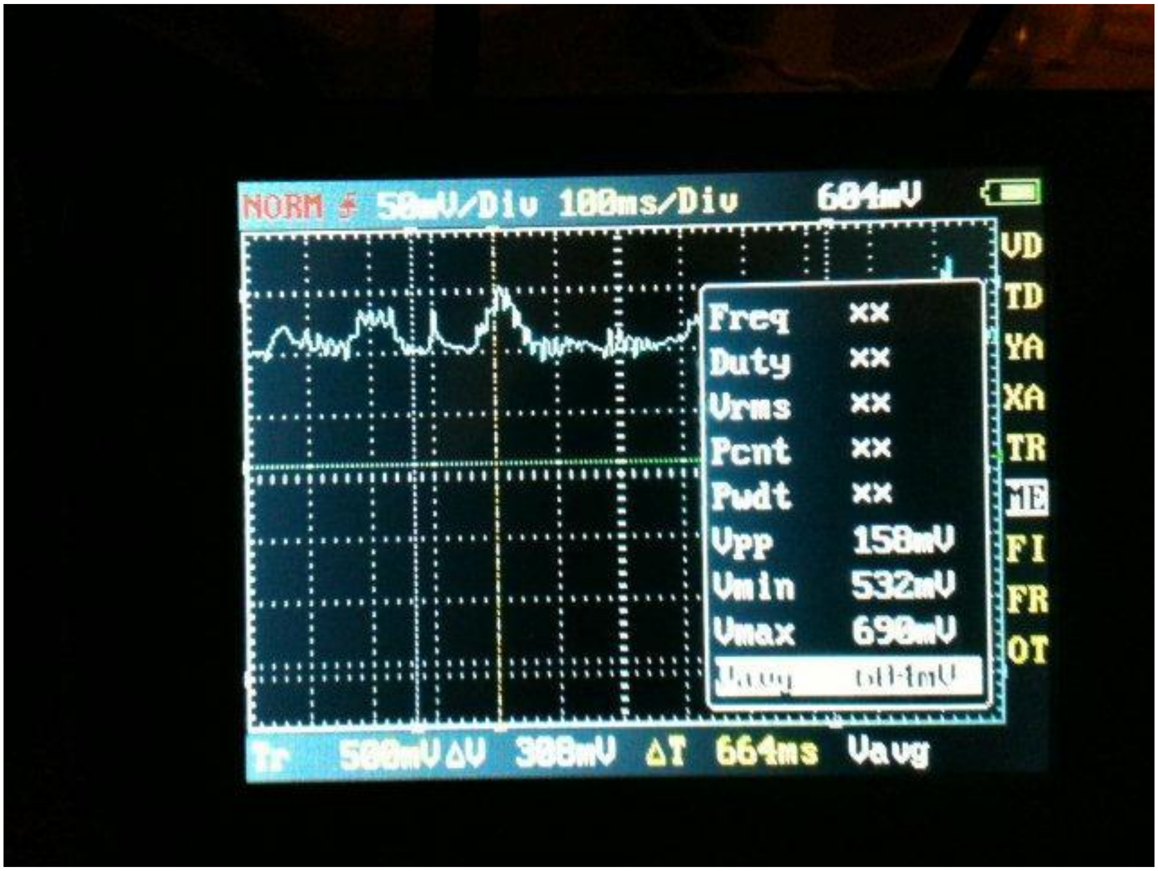
NORM
50mV/Div 100ms/Div
604mV
VD
TD
YA
XA
TR
ME
FI
FR
OT
Tr 500mV ΔV 300mV ΔT 664ms Vavg

Freq xx
Duty xx
Vrms xx
Pcnt xx
Pwdt xx
Vpp 158mV
Vmin 532mV
Vmax 690mV
Vavg [illegible]

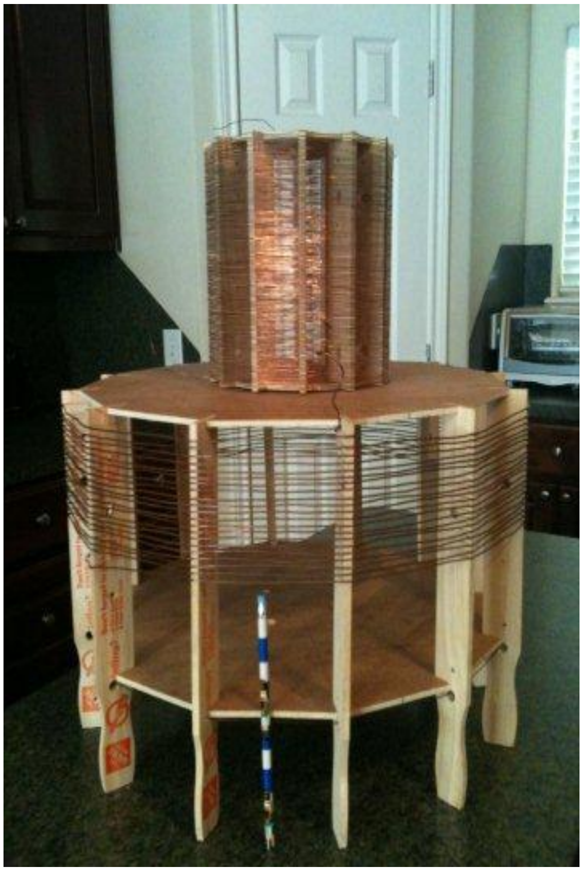

03-26-2012, 10:24 AM

[madhatter](#)
Senior Member

Join Date: Nov 2010
Posts: 457

the primary coil is key to the ground receptive part of the set. the primary and secondary resonate and based on the capacitance and geometry the radiation resistance will be eliminated in the extra coil giving rise to FTL reception. On my set the primary is not large enough as I based it purely on surface area however after reading the reports of others and noting their relations and the effects between my set, I tentatively state that the surface area of winding capacitance is what needs to me matched, not the bulk surface area. I'm working on the equations to see what this turns up. I'm also looking at the skin effect and mass as well, no point in leaving any stone unturned.

Nice build btw, looks good. the extra coils is such a PITA to wind too. I wouldn't worry about 'skip' distance from being too close to the transmitter in fact I'd say that's a prime example of ground current reception, proof that this works.
[Share on facebook](#) [Share on twitter](#) [Share on google_plusone_share](#) [Share on reddit](#) [Share on pinterest_share](#) [Share on linkedin](#)

#9

03-26-2012, 08:01 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **madhatter**
> *For example, the crystal set is a prime example of a information rich effects. I personally have the set built but have not gotten the results I want, however it's not been without results as it as given me answers to questions on capacitance by noting the differences in my results to others. however lets say we have one build that is catalog'd and every parameter of the physical construction known. from that it can be altered and the changes noted and then using that it can direct the mathematics to derive the needed equations to refine and move forward.*
>
> *right now the one thing I've seen is the surface area and capacitance equations, capacitance is frustrating itself as it's always an assumed infinite plane yet we know it's not. high frequency skin effect and resistance of the mass of the 'director' and the geometry all play a role, as seen it gets very complex. I've spent a couple hundred hrs already on just calculations and have now just started to notice a few things I'd like to explore, with some more research into hyperbolic trig I may have some equations that can be tested in the field. I'll share this freely with anyone willing to test on their crystal sets. ultimately I'd like to see if we can all build matching sets and see if we can network them together as Tesla intended.*

I am thinking the same thing. Matching sets would get noticed. Im thinking somewhere in the 2-3MHz range. That would make everything smaller and the brass shim stock would fit the bill. You could put that together for under 50 bucks. But we'd have to find a transmitter for the testing.

but before we get too far ahead of ourselves. We need to light an indicator bulb.

As far as my build. I am still looking for a long enough flat piece of copper.. I sent out a few emails looking for overruns, but the standard size seems to be 100" and I need 118". I found a thin walled copper tube at the scrap yard and bought enough.. Ive been contemplating slicing it open and flattening it out.

and now back to what I think was your point. The math needed to go along with the cyrstal initative. I want it. Been doing lots of thinking.

Lets start simple.. and I am really look forward to hearing from all those interested..

How do you calculate the C and K of a TRT secondary?
How do you calculate the L and M of a TRT secondary?
i.e. 20 turns(T), wire diameter(d) in mm, turn gap(g) in mm,

Can we even use the standard inductance and capacitance equations or are they contaminated?
bring on the versors

Ill post my answer tomorrow? after I watch the SFTS videos If i can think of one.

jake
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#10

03-26-2012, 09:56 PM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

Quote:

> Originally Posted by **jake**
> *I am thinking the same thing. Matching sets would get noticed. Im thinking somewhere in the 2-3MHz range. That would make everything smaller and the brass shim stock would fit the bill. You could put that together for under 50 bucks. But we'd have to find a transmitter for the testing.*
>
> *but before we get too far ahead of ourselves. We need to light an indicator bulb.*
>
> *As far as my build. I am still looking for a long enough flat piece of copper.. I sent out a few emails looking for overruns, but the standard size seems to be 100" and I need 118". I found a thin walled copper tube at the scrap yard and bought enough.. Ive been contemplating slicing it open and flattening it out.*
>
> *and now back to what I think was your point. The math needed to go along with the cyrstal*

> *initative. I want it. Been doing lots of thinking.*
>
> *Lets start simple.. and I am really look forward to hearing from all those interested..*
>
> *How do you calculate the C and K of a TRT secondary?*
> *How do you calculate the L and M of a TRT secondary?*
> *i.e. 20 turns(T), wire diameter(d) in mm, turn gap(g) in mm,*
>
> *Can we even use the standard inductance and capacitance equations or are they contaminated?* ***bring on the versors***
>
> *Ill post my answer tomorrow? after I watch the SFTS videos If i can think of one.*
>
> *jake*

Here is where hyperbolic algebra come in, Alexander Macfarlane has some very good books on hyperbolic and elliptical algebra. useful eq. for adapting to the coils can be found in Kennelys work as he does utilize Steinmetz and Heaviside's operators making it easier to cross reference his work.

the complex quantities notes of Steinmetz are also very helpful in explaining the j operator and power of the square root of -1.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#11
03-27-2012, 10:16 AM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **madhatter**
> *Here is where hyperbolic algebra come in, Alexander Macfarlane has some very good books on hyperbolic and elliptical algebra. useful eq. for adapting to the coils can be found in Kennelys work as he does utilize Steinmetz and Heaviside's operators making it easier to cross*

> *reference his work.*
>
> *the complex quantities notes of Steinmetz are also very helpful in explaining the j operator and power of the square root of -1.*

Thats where I am having trouble. I need to study.

So if all the adapting and math is down would it be able to calculate "values" for C,K,L,M and thier ratios at any point during the cycle of a TRT??? <---Please answer pretty please.. any one. What are the roadblocks?

That would help enormously in tuning the setup. It's almost impossible to tune three coils with 4 parameters that are changing without having either a solid math foundation or a natural ability to feel what you are doing. I'm working with the former at the moment. I'm not thinking about numbers; just C,K,M,L and what they are doing at every spot in the setup at every point in the cycle and and how I can add, reduce, and change their "rate".

Am I talking jibberish? or just a few months behind.
currently rewatching video 3 of the SFTS.

Also regarding the surface area aspect... Don't give up on mass just yet. Tesla's use of the word "stout" appears in lots of primaries.

My primary was calculated to 3m by 8.4cm by 0.1mm for a grand total of 26.4 cm^3.. Does that sound stout to you?
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Sponsored Links

EPD LABORATORIES
TESLA ADVANCED SEISMIC
WARNING SYSTEM #2
YOU CAN HELP SAVE A MILLION LIVES

SUPPORT EPD LABORATORIES, INC.

#12

03-27-2012, 02:31 PM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

Quote:

> Originally Posted by **jake**
> *Thats where I am having trouble. I need to study.*
>
> *So if all the adapting and math is down would it be able to calculate "values" for C,K,L,M and thier ratios at any point during the cycle of a TRT??? <---Please answer pretty please.. any one. What are the roadblocks?*

parametric changes are nearly impossible to calculate as they would occur on the fly and you'd need to run a simulator to get close. there is no program in existence that can do that for these parameters. you can get a rough layout that would need 'tuning' with the math and knowledge available at the moment.

Hyperbolic geometry is closer to nature, the cantenary slope is a prime example. a rope hung between two points exhibits this form. what also needs to be added into the equation is counter-space, due to the time invariant 'dimension' of the dielectric.

Quote:

> Originally Posted by **jake**
> *That would help enormously in tuning the setup. It's almost impossible to tune three coils with 4 parameters that are changing without having either a solid math foundation or a natural ability to feel what you are doing. I'm working with the former at the moment. I'm not thinking about numbers; just C,K,M,L and what they are doing at every spot in the setup at every point in the cycle and and how I can add, reduce, and change their "rate".*
>
> *Am I talking jibberish? or just a few months behind.*
> *currently rewatching video 3 of the SFTS.*
>
> *Also regarding the surface area aspect... Don't give up on mass just yet. Tesla's use of the word "stout" appears in lots of primaries.*
>
> *My primary was calculated to 3m by 8.4cm by 0.1mm for a grand total of 26.4 cm^3.. Does that sound stout to you?*

I'm still researching every day, it's an undeveloped field. the majority of mathematics of manifolds and abstract algebra all conform to relativity, that makes it that much harder to develop a working set of equations as one needs to filter out the conforming terms and relations. it's mighty pervasive, one of the reasons why the bulk of reference material is pre-1927.

around the turn of the century there was a divide on quaternions as well, Alexander Macfarlane also postulated with hyperbolic quaternions and that can be seen in his lectures on the mathematics of space.

so an abstract non relativistic conforming algebraic hyperbolic counter-space geometry needs to be developed it would seem. a simple thing really

I don't abide by the reductionist view of quantum physics or a static state. we exist in a dynamic interconnected manifold, one change effects all. any equations that try to handle this will have a type of 'Heisenberg' uncertainty to use it loosely, it's ever changing and a static 'snapshot' will not explain the dynamic relations. personally I think true advancement will occur when apparatus become 'living' not sentient but more organic. loose terms there, but think of a plant vs a solar panel, the solar panel is rigidly fixed in it's parameters where the plant can adapt and change as needed constantly.

I've diverged enough there. To make any real gains a framework needs to be developed to model the manifold of the dielectric and then develop the basics of each phenomen, i.e. the e,E,I,i k K,c etc...
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#13

03-31-2012, 10:11 AM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Does anyone have tips for soldering 14ga solid to 1/2" copper pipe?

Do you wrap the wire around the pipe and solder all the way around or just solder ~1/2" however???

Still trying to calculate MKLC. Watching SFTS videos again this time taking notes.

@ mad

What do you mean by manifold?

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Last edited by jake; 03-31-2012 at 10:13 AM. Reason: extra questions

#14

03-31-2012, 10:40 AM

Parav
Member

Join Date: Nov 2008
Posts: 49

Soldering Copper

Quote:

> Originally Posted by **jake**
> *Does anyone have tips for soldering 14ga solid to 1/2" copper pipe?*
>
> *Do you wrap the wire around the pipe and solder all the way around or just solder ~1/2" however???*
>
> *Still trying to calculate MKLC. Watching SFTS videos again this time taking notes.*
>
> *@ mad*
>
> *What do you mean by manifold?*

Hi Jake,

To solder together thick pieces of solder , an ordinary soldering iron just does not provide enough heat as the heat gets dissipated into the thicker material too easily.
Use a propane torch -like what plumbers use to solder copper pipe fittings.
Also , make sure that the copper pipe is lightly sanded at the area you want to soldered -to get rid of the oxidization---or use that flux paste plumbers use to help clean off the oxidization while it's being soldered . Good luck--[I] hope this helps.

Paul
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

#15

03-31-2012, 10:42 AM

Parav
Member

Join Date: Nov 2008
Posts: 49

Quote:

> Originally Posted by **Parav**
> *Hi Jake,*
>
> *To solder together thick pieces of solder , an ordinary soldering iron just does not provide enough heat as the heat gets dissipated into the thicker material too easily.*
> *Use a propane torch -like what plumbers use to solder copper pipe fittings.*
> *Also , make sure that the copper pipe is lightly sanded at the area you want to soldered -to get rid of the oxidization---or use that flux paste plumbers use to help clean off the oxidization while it's being soldered . Good luck--[I] hope this helps.*
>
> *Paul*

oops typo-- I meant when soldering two pieces of thicker copper together
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#16

03-31-2012, 10:28 PM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

Here's a decent explanation on manifolds.
Manifold - Wikipedia, the free encyclopedia
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#17

04-09-2012, 09:00 AM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

I this typical RF

On a TRT secondary and extra coil:

{extra coil}------{secondary}-----Ground

If I tap between the extra coil and secondary run a germanium diode, my body, and LED in series to ground I can dimly light the led.

{top of secondary}--{+germanium diode-}---{right hand}---{body}----{left hand}---{+LED-}----Ground

The LED will show light and its getting brighter the more I learn.

Is this a normal RF???

I can see 2.2v avg across the LED when I leave my body out of the circuit. 2.6max and 2.5pp. These are all estimated because the values fluctuate with the broadcast signal. They stabalize a bit when I put a 330pF cap across the LED.
Also adding capacitance to the lead that goes from secondary to diode in the form of me touching the insulated lead has a large effect. I was able to recreate this effect using a the piece of aluminium(i.e. the ground terminals of a dule gang air variable cap). Using the variable cap the way it should be used did not work. Made things completely stop working and the voltage dissapear. I found this on accident.

So is this something you would see on a typical crystal radio?
Is it RF or somethine else?
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#18

04-09-2012, 04:03 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *So is this something you would see on a typical crystal radio?*

Not unless the antenna was near a strong RF source. Do you have any radio or TV transmit towers near you? If you watch MrFlathunter (Seth) videos you will see he is able to light things with his passive resonator but only if near the transmitting exciter.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#19
04-09-2012, 06:53 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **xee2**
> *Not unless the antenna was near a strong RF source. Do you have any radio or TV transmit towers near you? If you watch MrFlathunter (Seth) videos you will see he is able to light things with his passive resonator but only if near the transmitting exciter.*

I have a 10kw AM station about 6 miles away. I looked up mrflathunter and found two more that show similar results of dimly lighting LEDs with nothing but a coil and diode. Do you have any advice on how to run a joule theif off these? I listed the voltages I can measure above.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#20
04-09-2012, 06:56 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *Does anyone have tips for soldering 14ga solid to 1/2" copper pipe?*

> *Do you wrap the wire around the pipe and solder all the way around or just solder ~1/2" however???*
>
> *Still trying to calculate MKLC. Watching SFTS videos again this time taking notes.*
>
> *@ mad*
>
> *What do you mean by manifold?*

Solder wire to standoff. Attach standoff to pipe with screw.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#21

04-09-2012, 07:07 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *Do you have any advice on how to run a joule theif off these?*

I have posted a number of low power Joule thief circuits. This may be a good choice ... no toroid Joule thief 3 - YouTube

Rectify radio output with diode and capacitor. Run Joule thief from capacitor. You will need at least 0.5 to 0.7 volts to run Joule thief. You do not need 10,000 uF capacitor, use what ever you have. Reduce resistor value to make LED brighter.

Same circuit with toroid >> 2 uA Joule thief - YouTube
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Last edited by xee2; 04-09-2012 at 07:36 PM.

Sponsored Links

EPD LABORATORIES
TESLA ADVANCED SEISMIC
WARNING SYSTEM #2
YOU CAN HELP SAVE A MILLION LIVES

SUPPORT EPD LABORATORIES, INC.

#22

04-11-2012, 07:58 PM

WARNING
BURN HAZARD

jake

Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **xee2**
> *I have posted a number of low power Joule thief circuits. This may be a good choice ... no toroid Joule thief 3 - YouTube*
>
> *Rectify radio output with diode and capacitor. Run Joule thief from capacitor. You will need at least 0.5 to 0.7 volts to run Joule thief. You do not need 10,000 uF capacitor, use what ever you have. Reduce resistor value to make LED brighter.*
>
> *Same circuit with toroid >> 2 uA Joule thief - YouTube*

Wish I had seen that before I untwisted 500 feet of 20 gauge bell wire. That would have been fun to try with 500'.

Thanks,

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

#23

04-11-2012, 08:06 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **Kokomoj0**
> *jake, use a full wave germanium bridge and filter caps.*

Are you saying ditch the joule thief and just add a bridge?

Can't be that simple can it?

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#24
04-11-2012, 08:13 PM

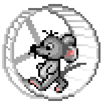

Kokomoj0
Senior Member

Join Date: Jul 2011
Posts: 386

I said this in the other thread, make a ful wave bridge out of your germ diodes and use filter caps. I would be shocked if you couls pull in any amount of power like that because of the low current density. The whole idea of pulling "significant" power off of these radio stations seems absurd to me.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#25
04-11-2012, 08:14 PM

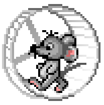

Kokomoj0
Senior Member

Join Date: Jul 2011
Posts: 386

Quote:

> Originally Posted by **jake**
> *Are you saying ditch the joule thief and just add a bridge?*
>
> *Can't be that simple can it?*

It will give you filtered dc, then you can do with it what you want.

you know thinking about it you may not be able to do a fw bridge, maybe only 1/2 wave, I forget the ckt config.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Last edited by Kokomoj0; 04-11-2012 at 08:20 PM.

Quote

#26

04-11-2012, 09:41 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **Kokomoj0**
> *I said this in the other thread, make a ful wave bridge out of your germ diodes and use filter caps. I would be shocked if you couls pull in any amount of power like that because of the low current density. The whole idea of pulling "significant" power off of these radio stations seems absurd to me.*

That's what I'm trying to figure out.

File:Replica-of-first-transistor.jpg - Wikipedia, the free encyclopedia

I like this picture. It says alot about what we can do once we learn how.

I've learned alot about working with radio frequency just tuning the basic crystal radio.

So far my TRT is a Just a large crystal radio with two coils. But it will light an LED continuously all day when the station is @ 10kW. It's dim but it's on. Not blinking(filter cap), one germanium diode and led across the secondary.

It may all be just standard RF reception. I don't know yet. I didn't think that would be possible. But it's not tuned yet. Unless I got it right the first rime.

The full wave should help as long as the extra voltage drop does not cause problems.

Thanks,

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#27

04-12-2012, 04:38 PM

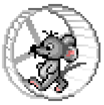

Kokomoj0

Senior Member

Join Date: Jul 2011
Posts: 386

http://i168.photobucket.com/albums/u...taldiagram.jpg

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Last edited by Kokomoj0; 04-12-2012 at 04:41 PM.

#28

04-13-2012, 07:57 PM

WARNING
BURN HAZARD

jake

Senior Member

Join Date: Nov 2007
Posts: 169

Why?

As the distance between the primary and secondary is increased the resonant frequency of the coil increases ~5-10kHz per inch and the amperage measured between the can and the top of the secondary also goes up.

Any ideas?

I'm playing with the secondary and primary like in

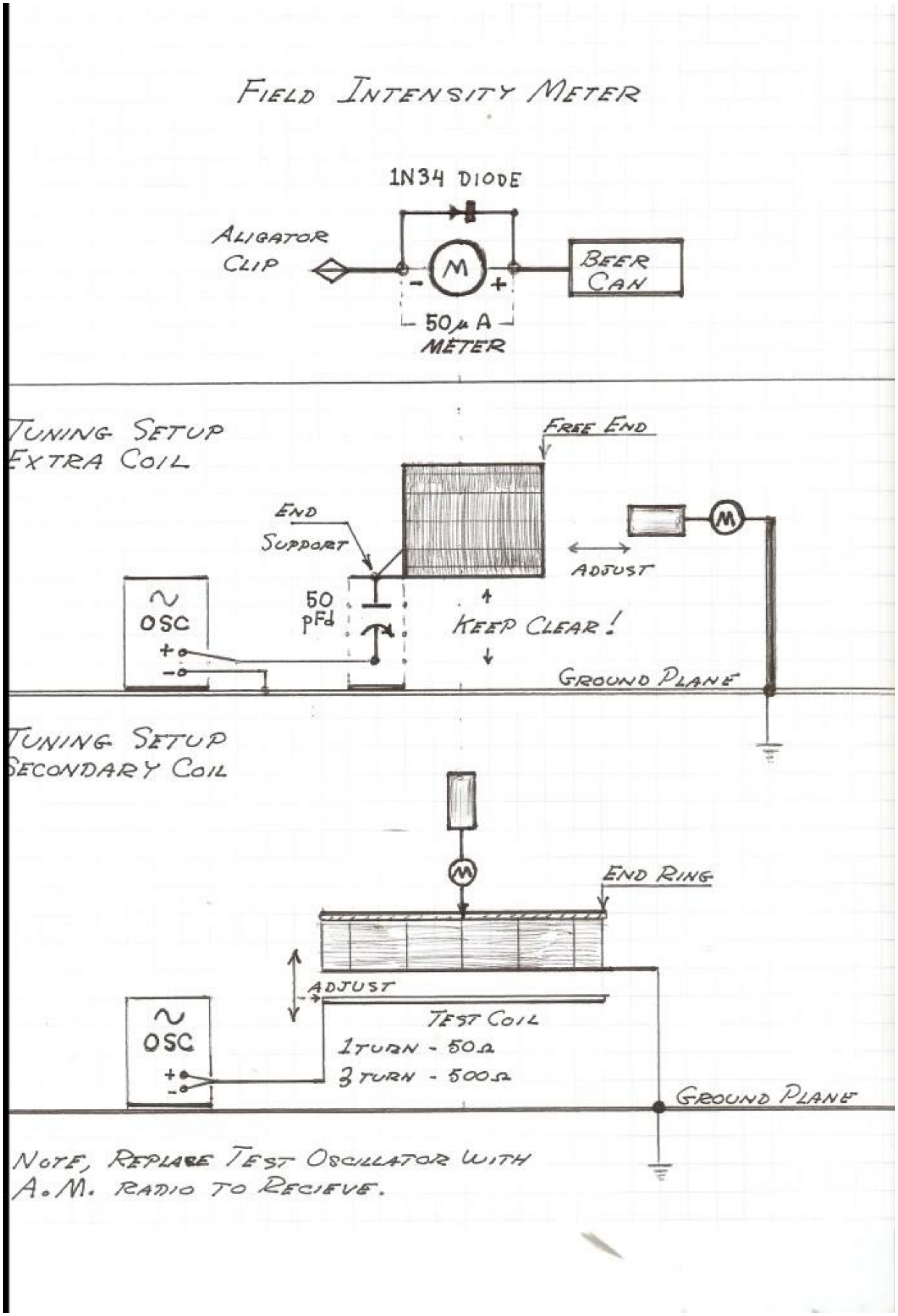
FIELD INTENSITY METER
1N34 DIODE
ALIGATOR CLIP
M
-
+
BEER CAN
50µA METER
TUNING SETUP EXTRA COIL
FREE END
END SUPPORT
M
ADJUST
50 pFd
OSC
+
-
KEEP CLEAR!
GROUND PLANE
TUNING SETUP SECONDARY COIL
M
END RING
ADJUST
OSC
+
-
TEST COIL
1 TURN - 50Ω
3 TURN - 500Ω
GROUND PLANE
NOTE, REPLACE TEST OSCILLATOR WITH A.M. RADIO TO RECIEVE.

I did not post this in the Dollard thread because I am missing the split ring on top. I am also using more of a current indicator with an adjustable range for readings so there are no actual values I cam use.

Currently the best frequency is 2300kHz. But this will change (+) 500kHz) depending on the coil

#29

04-16-2012, 08:34 AM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Getting Better

I'm getting lonely. Is anyone else getting ANY results out of the crystal radio initiative setup?
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#30

04-16-2012, 06:16 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *Any ideas?*

Nice drawings.

When the two coils are close they "load" each other. This increases inductance of each coil, decreases Q of resonators, and provides more energy transfer between coils. In crystal radios increasing coupling between coils will provide more energy transfer but also decrease resonator Q. Usually minimum coupling is best since this produces highest resonator Q. If both coils are resonant at the same frequency then the Q of one coil will increase the effective Q of the other coil. This is how coupled resonator filters work. But the resonant frequency will change as the coils are brought closer together so they must be re-tuned after they are moved closer together.

I suggest you try coupling two parallel resonant circuits and tune to same frequency. This should give improved reception. Best if you can tune to frequency of near by radio station.

Or just make one parallel resonant circuit with large coil and weakly couple to it with a single loop or a small capacitor.

I would think your power detector would work better if the diode was in series with the amp meter. The circuit you have would work with a volt meter.

Please keep posting your results. I hope you will be able to light an LED from the RF signal.

04-16-2012, 08:05 PM

Web000x
Silver Member

Join Date: Apr 2009
Posts: 539

Jake,

Eric said that it is not easy to get three coils to resonate together. It took him about a year at Landers to properly tune his receiver. As xee2 was saying, your magnification factor (Q) will be much higher if you can get the coils into a symbiotic resonance. This heightened magnification factor is apparently what allowed Eric to light the light bulb.

He also said that if you are having trouble getting these coils to resonate then don't use the extra coil and it would still work. However, I don't know if this will light a light bulb without the heightened magnification factor. This might just be for a telluric receiver.

Have you tried using an aerial (dipole) and the TMT style antenna at the same time for a comparison on an oscilloscope? This is what Eric is talking about in order to see the difference in phase angle of the two signals to see that the telluric transmission is propagated faster than the above ground transmission. This would be the evidence of a faster than light propagation.

Dave
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

Sponsored Links

EPD LABORATORIES
TESLA ADVANCED SEISMIC
WARNING SYSTEM #2
YOU CAN HELP SAVE A MILLION LIVES

SUPPORT EPD LABORATORIES, INC.

Download SOLAR SECRETS by Peter Lindemann
***Free - Get it now*: Solar Secrets**

#32

04-16-2012, 08:14 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

The credit for the drawings goes to Eric Dollard. Those are his pictures from the Dollard thread.

And the LED is lit. I am able to take the large 20 turn coil and attach it to either my fireplace or stove vent hood or large metal object(i.e. antenna) and light an LED to a decent glow. It is not full brightness but it was constant and even my wife freaked out a little when I said there was no battery. The other end of the coil has a Dr Pep can.

My first attemts to get a joule thief to brighten up the light were a failure but I will revisit the JT idea a little later as attaching the coil to an antenna now allowes me to light the LED to a level that does not require the JT.

Without attaching to an antenna I can only sometimes get a very faint glow.

ALSO this coil is not resonant at 1620kHz. This is the station that is supplying the power. It is 10kw more than 5 miles away. And at night it goes down to 1kw. Even at night I can see very very faint blips of the LED.

The coil attached to the fireplace surround and DP can is resonant at ~2200kHz depending on where and how I excite it. I KNOW that LED will go balistic if I can get it tuned to 1620. Using a function generator at 1620 will not light the led. Turning it to 2200 will light the LED beyond a glow and one might say bright enough.

PROBLEM!! - Tuning for brightness does not seem to like external capacitors and REALLY does not like alligator clip leads. To get it to where it is now required respacing the turns of the coil and tapping the coil in the right spot. but even that does not let it drop down to 1620

Eventually when I figure out how to tune the coil without needing capacitors I will move on to the extra coil. Using the extra coil now will either help or hinder depending on what I have going on. So I took it off and will revist it when I get the secondary(large 20 turn coil) to 1620.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

#33

04-16-2012, 08:51 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **Web000x**
> *Jake,*
>
> *Eric said that it is not easy to get three coils to resonate together. It took him about a year at Landers to properly tune his receiver. As xee2 was saying, your magnification factor (Q) will be much higher if you can get the coils into a symbiotic resonance. This heightened magnification factor is apparently what allowed Eric to light the light bulb.*
>
> *He also said that if you are having trouble getting these coils to resonate then don't use the extra coil and it would still work. However, I don't know if this will light a light bulb without the heightened magnification factor. This might just be for a telluric receiver.*
>
> *Have you tried using an aerial (dipole) and the TMT style antenna at the same time for a comparison on an oscilloscope? This is what Eric is talking about in order to see the difference in phase angle of the two signals to see that the telluric transmission is propagated faster than the above ground transmission. This would be the evidence of a faster than light propagation.*
>
> *Dave*

I saw this post after I posted my last post.

It is quite a troubleshooting mission to tune three coils. I am having extream difficulty tuning one and I don't even know if I am doing that right. but I think I am getting a feel for it. Its hard to tell how much of the tuning should be done with distributed elements alone. As I noted in my last post adding a capacitor no matter how tiny seems to reduce performance. Putting my hand ~1" from the soda can peaks performance. Touching the can kills it.

I have assembled all that is necessary for version 2. It will be the same dimensions as the first unless I can figure out how to lower the resonant frequency about 400kHz. Instead of using 14ga I will be using 2.21mm OD x 0.91mm ID copper tubing. This will more than double the active copper compared to what I have now. I don't know how important that is but it feels right. The primary will have slightly more and can be narrowed down to match. Same for the copper plate primary cap. Now thats a lot of copper. Trust me I have been buying it by the pound at the scrap yard as it comes in.

I do not have a scope capable of the comparison. But I am sure if I can light a 100w light bulb my wife will let me buy one.

Thanks for the input.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#34
04-16-2012, 09:22 PM

dR-Green
Gold Member

Join Date: Apr 2008
Location: Merlin's Fort
Posts: 1,220

Quote:

> Originally Posted by **jake**
> *I am having extream difficulty tuning one and I don't even know if I am doing that right. but I think I am getting a feel for it. Its hard to tell how much of the tuning should be done with distributed elements alone. As I noted in my last post adding a capacitor no matter how tiny seems to reduce performance. Putting my hand ~1" from the soda can peaks performance. Touching the can kills it.*

Use a metal plate connected to earth/bottom of the secondary with adjustable distance from the can, and position it for peak performance. It won't like a variable cap very much across the secondary (unless maybe it actually goes down to zero pF), the type with rotating plates. I think vacuum variable caps might be usable here but they are expensive, so a metal plate near the can should do the trick. You can also lower the frequency by using a bigger can/pieces of metal. I think it's the surface area that applies in this case. Try connecting progressively larger pieces of metal and see what happens.

Also the bulb that Eric has recommended since is 28v 1.12 watt. I think that will be quite a task. The most I've seen on the scope with my flat spiral coil is 210mV so there's no possibility of lighting the bulb with that. I expect I need much better grounding which will be done at some point, but there's no way of adjusting the windings either so that's about the most I'll be able to do with that coil for the moment. [edit] Although I'm not entirely sure if it will even be possible with a flat spiral at all
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on

pinterest_share Share on linkedin

"Knowledge is cosmic. It does not evolve or unfold in man. Man unfolds to an awareness of it. He gradually discovers it." - Walter Russell

"Once men died for Truth, but now Truth dies at the hands of men." - Manly P. Hall

---

Last edited by dR-Green; 04-16-2012 at 09:32 PM.

#35

04-16-2012, 09:27 PM

Web000x
Silver Member

Join Date: Apr 2009
Posts: 539

Quote:

> Originally Posted by **jake**
> *I saw this post after I posted my last post.*
>
> *It is quite a troubleshooting mission to tune three coils. I am having extream difficulty tuning one and I don't even know if I am doing that right. but I think I am getting a feel for it. Its hard to tell how much of the tuning should be done with distributed elements alone. As I noted in my last post adding a capacitor no matter how tiny seems to reduce performance. Putting my hand ~1" from the soda can peaks performance. Touching the can kills it.*
>
> *I have assembled all that is necessary for version 2. It will be the same dimensions as the first unless I can figure out how to lower the resonant frequency about 400kHz. Instead of using 14ga I will be using 2.21mm OD x 0.91mm ID copper tubing. This will more than double the active copper compared to what I have now. I don't know how important that is but it feels right. The primary will have slightly more and can be narrowed down to match. Same for the copper plate primary cap. Now thats a lot of copper. Trust me I have been buying it by the pound at the scrap yard as it comes in.*
>
> *I do not have a scope capable of the comparison. But I am sure if I can light a 100w light bulb my wife will let me buy one.*
>
> *Thanks for the input.*

I assume that you are using the resonance detection method that Eric had me post recently?

As far dropping the resonant frequency of your coil, Eric had nothing to go off of except his memory for developing those formulas for design parameters. This is all in the experimental stage of getting these formulas correct. It is possible that you need to readjust the constants given by Eric by setting the coil dimensions equal to the measured resonance/constant relations and solve for a new constant. Then you could more precisely target the desired frequency in your next build. That is what I would do anyway if I were trying to adjust the frequency.

Are you trying to parallel a cap with the secondary? I think it needs to resonate at the desired frequency without a cap.

Also, are you finding that there are two resonant frequencies for the secondary that exist in a pi/2 proportionate relation?

Dave
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#36

04-16-2012, 09:57 PM

Web000x
Silver Member

Join Date: Apr 2009
Posts: 539

Quote:

> Originally Posted by **dR-Green**
> *Also the bulb that Eric has recommended since is 28v 1.12 watt. I think that will be quite a task. The most I've seen on the scope with my flat spiral coil is 210mV so there's no possibility of lighting the bulb with that. I expect I need much better grounding which will be done at some point, but there's no way of adjusting the windings either so that's about the most I'll be able to do with that coil for the moment. [edit] Although I'm not entirely sure if it will even be possible with a flat spiral at all*

It is hard to catch every detail that Eric throws me when I speak to him, but I am almost certain that there needs to be an extra coil in order to get the magnification factor to skyrocket which will in turn light the light bulb. I don't think I've seen an extra coil on your setup, or have I?

Dave
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#37

04-16-2012, 10:29 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Yes there is another point at ~7200kHz. It's small by comparison to the other.

I was using the method you posted for the secondary to the best of my knoledge. But that's not much. Here is what I did:

For the test coil I used a pretty stout flat copper strip. I could only get it around two times. Placing the function generator across 1 turn of the test coil I set it to work. (always wanted to say that) I was not sure what the 50ohms was for but I figured it was accounted for by the 50 in the function generator.

Come to think of it I was not grounded. Or not grounded well. I will try again tomorrow outside where I have a semi decent ground. Am I correct about the 50ohms? Does the size of wire on the test coil matter?

And yes I have/had an extra coil. I took it off and things started to get better. That's when I realized something was wrong. Then you posted the tuning diagrams and I left it off. But the wire loosened up and now it has shorts all over the place. I'm using uninsulated wire for everything. I think I will move to insulated on my next extra coil.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#**38**

04-17-2012, 07:52 AM

Web000x
Silver Member

Join Date: Apr 2009
Posts: 539

That is not the frequency that I'm looking for. See if your meter shows *any* response around 1400 KHz or 3455 KHz.

I could be wrong but I think that there is not supposed to be a resistor in the excitation loop since there is no symbolic representation of one in the drawing. My interpretation of it is that if you have a function generator with an impedance of 50 ohm then you would use a 1 turn loop with no resistor and for 500 ohm impedance a 3 turn loop with no resistor. This is because you are interested in delivering the magnetizing force to the loop rather than trying to get a maximum power transfer between the oscillator and the excitation loop (where the majority of power would be delivered to the resistance). Try it with and without the resistor just to see if there is a difference.

Also, nobody will light a lightbulb without a GOOD ground. When I was quizzing Eric about how good a 'good ground' might be he said that the ground is a horrible conductor of electricity and that what you are really doing is creating a capacitive coupling to the Earth with the wires so make sure you have PLENTY of wires laying down against or buried in the Earth.

Dave

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Last edited by Web000x; 05-08-2012 at 07:25 AM.

Quote

#39

04-17-2012, 07:54 AM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *Eventually when I figure out how to tune the coil without needing capacitors*

To lower frequency just increase inductance (increase number of turns keeping everything else the same).

Suggestion: make coil with a lot of extra turns and then cut off turns until the frequency is where you want it.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Last edited by xee2; 04-17-2012 at 09:02 AM.

Quote

#40

04-17-2012, 09:45 AM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Awesome

@ Web000x:
Good Grounding is on the horizon. What do you think of Eric's recommendation of the 80 ground radials? Is this needed in addition to the 16 ground rods or can this be used in place of the 16 rods. It would be easier to lay out compared to the rods. but the center rod will be 10-20' deep in good soil.

@ xee2:
I wish it was that easy. Problem is the coil I am using now was calculated to have a specific diameter, hight and length. I can't change one without screwing up the others.

BUT I have a slightly smaller diameter cardboard tube. I think it would be a good idea to build a setup based on you way of thinking then I could compare the two. Smaller diameter will allow more turns greater induction for same length of wire.
Can't wait to get out of work.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

____

Quote

#**41**
04-17-2012, 10:19 AM

Web000x
Silver Member

Join Date: Apr 2009
Posts: 539

It sounds like he is saying that the 16 ground rods should be used if you are in a place that stays relatively moist. If the ground is dry sand or rock, you need to use the 80 ground rods of 14 gauge wire. I don't see why you could not use the 80 rod configuration in normal soil. (But I am also not Eric. I'll ask when he calls me next)

Be careful about using solid tubing to wrap coils. The dielectric constant needs to be a minimum. Anything above relative dielectric constant = 1 will retard the wave. This is why Eric builds all of his transformers with wooden dowels, to expose the wires to the dielectric constant of air.

Dave
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Sponsored Links

EPD LABORATORIES
TESLA ADVANCED SEISMIC
WARNING SYSTEM #2
YOU CAN HELP SAVE A MILLION LIVES

[SUPPORT EPD LABORATORIES, INC.]

#[42]
04-17-2012, 02:42 PM

[xee2]
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *BUT I have a slightly smaller diameter cardboard tube. I think it would be a good idea to build a setup based on you way of thinking then I could compare the two. Smaller diameter will allow more turns greater induction for same length of wire.*
> *Can't wait to get out of work.*

Not exactly true. Here are some links that may help you:

Inductance Calculator

Web Based Resonant Frequency Calculator
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#[43]
04-20-2012, 09:33 AM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

What is it?

So now I can light the LED to a continious glow at night when the statio is at 1000w. The larger primary made that happen. The correct capatance across the primary eliminates all blinking and the little hand is no longer needed.

There is very little improvement when the power goes up in the morning.

What it is:

Secondary - 20 turn coil of 14 gauge wrapped according to EPD. With attention to getting the coil to the proper EPD dimensions. So spacing was what it was.

Primary - Almost 2 turns of Stout Flat copper.
84.5mm wide
2857.5mm long(10mm short) Best I Could Do. Now I need to shorten the Secondary
1.35mm thick (STOUT)

I made this from a 1.25"dia copper pipe. Split it lenght wise on a table saw. Hammered an ax down the lenght to open it up. Used pliers to open one end wide enough to get it in a sheet metal roller and rolled it flat in 12 passes.

FWBR - Of 1N34 diodes across the secondary or from almost any point on the secondary to bottom of secondary.

LED - across the DC of the FWBR.

Ground or antenna? - Bottom of secondary is connected to Fireplace shroud/metal chimney. It is grounded via propane copper tubing but it also acts like an antenna since its big metal and sticks out the roof.

coments and questions are welcome.

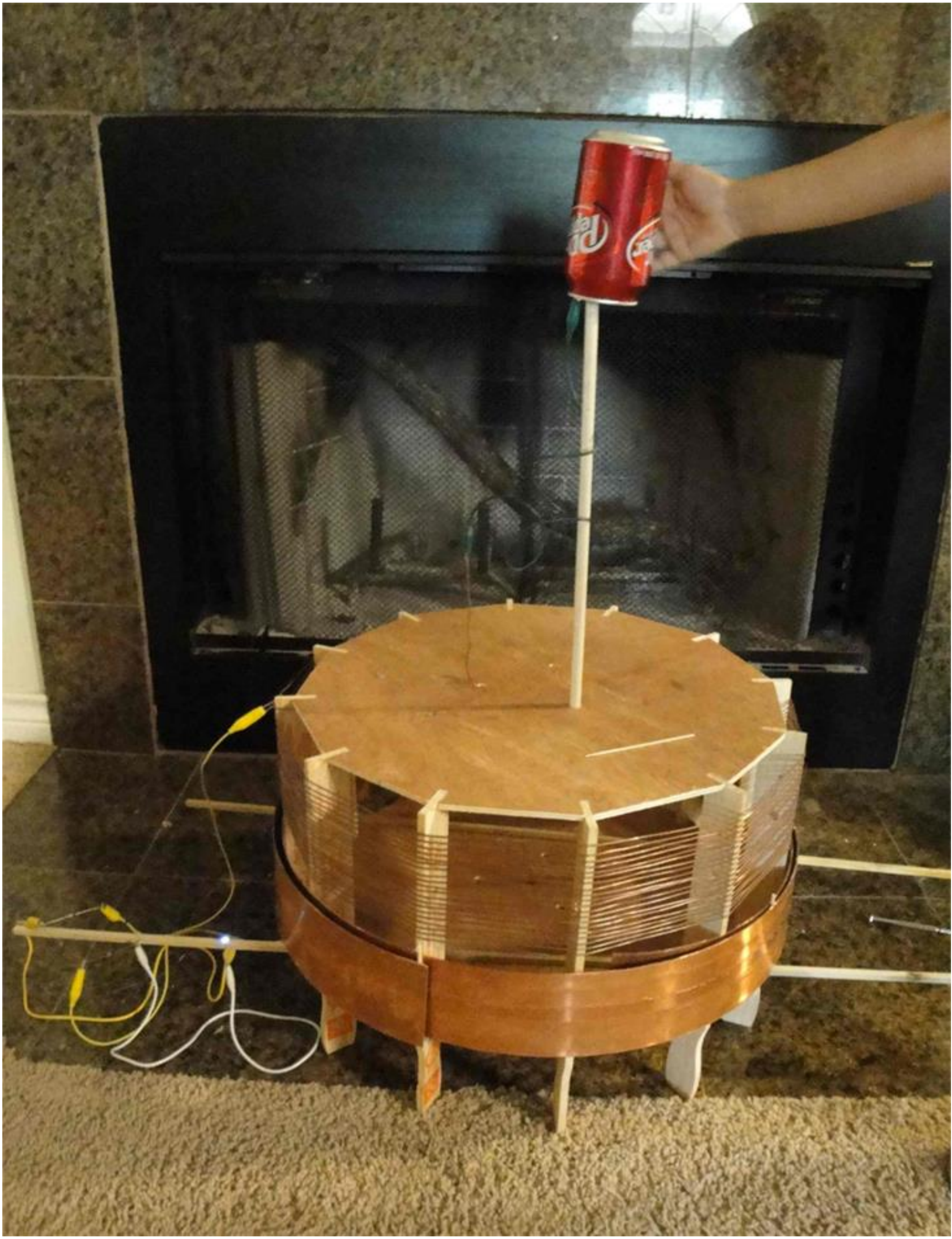

04-20-2012, 12:19 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *So now I can light the LED to a continious glow at night when the statio is at 1000w.*

Congratulations. Thanks for the report and photo. Is your only antenna the short wire on top of the primary coil (or are you connecting to an external antenna wire in addition to the fireplace ground)? Please post a link to the page where you got the EPD info.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#45

04-20-2012, 01:33 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

@ jake

Replacing LED with a capacitor and then using capacitor as a battery for Joule thief circuit should produce brighter LED. You must be getting at least 2 volts out of FWR in order to light the LED so you should have no problem driving a low current Joule thief circuit.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#46

04-20-2012, 06:48 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **xee2**
>
> *Congratulations. Thanks for the report and photo. Is your only antenna the short wire on top of the primary coil (or are you connecting to an external antenna wire in addition to the fireplace ground)? Please post a link to the page where you got the EPD info.*

http://www.energeticforum.com/179223-post956.html

First four equations are the most important as far as I can tell. (wire length, turn length, coil diameter, coil hight) Primary width is 18% of coil diameter but I cant find the post. It's somewhere on the "Peter what happened to Eric" thread, search the T-Rex posts. If you can figure out how to incorporate the other items into your build without loosing the first 4 the the better according to EPD. Not having the correct size tube, I used 14 gauge as recommended by Eric. Coax is also recommended by Eric but I felt this was a compromise to another extream, so I went with the 14 gauge solid.

There are a few excell calculators on the yahoo group. I wrote the TRTcalc3.xls. These will do the math for you. If you are not a member PM me and I will send it to you. We cant post excel files here. Now that I think about it I should update that spreadsheet again.

As you know from you exciter experience there are variables in designing a coil. This way, adds a few more variables in the mix. This coil is based on a few ideas EPD described in his writings and old school HAM techniques, brought back to life by EPD. Spaced windings, active copper volume match
and Erics formule were a few of my considerations.

The extra coil is still a bit of a mystery. I am not using one. I made one but I have not figured out what it is supposed to do, and how it does it. Erics dimensions were tenetive(as he described) at best and not all the equations make sense to me at least. I am almost ready to revisit that.

ALSO I got an LED AV plug to glow just as well as the FWBR. One wire to the top of the secondary and another lead off the negative of the LED laying on the tile.

The fireplace is metal and sticks out of the roof so its like an antenna. but its grounded so I dont know what to call it.

EDIT:!!!
OK fireplace not needed!!!

I connected to the mains ground and adjusted the CAN, now the LED is on, bottom of the secondary is mains ground and soda can antenna.

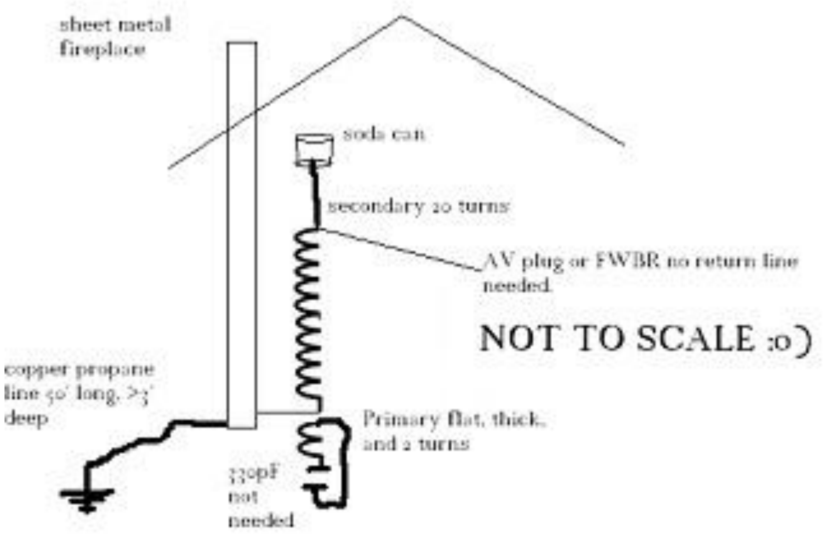
sheet metal fireplace
soda can
secondary 20 turns
AV plug or FWBR no return line needed.
NOT TO SCALE :o)
copper propane line 50' long, >3' deep
Primary flat, thick, and 2 turns
330pF not needed

04-20-2012, 09:10 PM

Leviathan
Member

Join Date: Jan 2012
Posts: 57

I don't think this is off topic but this link is to a instruct-able on how to make a Tesla lightning detector. It looks really interesting this guy has videos of his device detecting light, and sound waves as well as EM waves. I've been wanting to build one for a long time but never got around to it ;p

Make the Tesla Spiral Antennas
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#48

04-21-2012, 12:43 AM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

@ jake

Thanks for the info. I find lighting an LED with just a few feet of wire for an antenna to be impressive. What kink of results did EPD get with his?

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#**49**

04-21-2012, 03:10 PM

WARNING
BURN HAZARD

jake

Senior Member

Join Date: Nov 2007
Posts: 169

This is the theory for the crystal set initative AND an appropriate experiment for validation. To my knowledge it has not been accomplished. It is based on the following post.

http://www.energeticforum.com/renewa...tml#post173955

Quote:

> Originally Posted by **T-rex**
>
> *Crystal Sets Gone Wild*
>
> *For the diagram shown the coil dimensions are missing, number of turns, etc. A good ground is essential for these kinds of devices. 16 Ground rods in a 10 to 20 foot radius circle, connected to a single ground rod at the center(17th rod), this connection being 10 gauge wire. Dry sand or rock will not ground, so this requires 80, each 14 gauge wires in a 30 foot diameter circle in a star radial configuration, to a center terminal. Without these groundings a Tesla Transformer cannot properly operate, but some "HI-Z" sets may.*
>
> *The objective here is to scale the "Crystal Set", a step at a time, into a Tesla Transformer for the reception of medium wave band, 300 - 3000 kilocycle A.M. broadcasts. No license is required for this and the broadcast station provides the power.*
>
> *And this objective cooperates with the primary objective. That is; Who will be the first ham to disprove Einstein's theory? An International contest, but who will sponsor it, Iran maybe?*
>
> *We have the good fortune in the "Crystal Set Initiative" that, in theory at least, a quarter wave A.M. broadcast tower, and its 120 quarter wave ground radials, must emit a pair of waves as shown by Tesla in his basic diagrams.*
>
> *Hence it can be seen that a pair of waves are engendered by this transmission system. (Tower*

> *and Star Radials). One wave, Hertzian, is the over ground wave, the other wave, Telluric, is the under ground wave. These two waves arrive at the point of reception in their own distinct time frames, giving rise to a difference in phase. Hence, multiple rings of interference patterns are produced. Since the Hertzian portion, over ground, time frame is based upon the velocity of light, then the Telluric portion, under ground, time frame gives the Telluric velocity. Two crystal sets, one over ground, one under ground, and a basic oscilloscope , that simple. I have done this at Landers.*
>
> *Concluding, a Tesla Magnification Transformer, properly proportioned can, in theory, actually draw power from a local 50 kW station. Several hundred watts of power reception is likely. This would prove Tesla once and for all. No antenna, just a good ground, and a nice and bright 100 watt light bulb.*
>
> *This would overturn physics more than any billion dollar C.E.R.N. project. A ham radio operator overturns Einstein for 100 bucks. What a concept.*
> *Read,*
> *Tesla, "The True Wireless"*
> *Tesla, "System of Concatenated Tuned Circuits"*
> *Dollard, "System for the Transmission and Reception of Telluric Electric Waves"*
> *A.R.R.L. "Radio Amatuers Handbook". Chapter "H.F. Transmitters, & Tank Circuits"*
> *73 DE N6 KPH*
>
> ___________________
>
> *If you consider these tranmissions helpful, please donate to Eric Dollard via PayPal at DollardDonations@gmail.com*

Some more build info form the "Peter where's Eric" thread that was locked.

I wonder if Aaron could sticky it to the begining of the EPD thread.

Quote:

> Originally Posted by **T-rex**
> *Tesla Type Crystal Radio*
>
> --
>
> *The design of the Tesla Transformer starts with the secondary resonant coil. In its resonant mode it is a quadrapolar resonator. Two constants exist in the construction of this solenoidal resonator, one is that the height to length ratio must be 20%, the other is that the side by side spacing of the cylindrical conductor turns is 62% the diameter of the cylindrical conductor. The optimum number of turns is 20 on the secondary, with 2 on the primary. This gives the physical size of the coil for a given frequency. Increasing the number of turns reduces the size of the secondary coil in proportion to the increase in turns. Accordingly the conductor diameter also decreases as the copper is spread out over more windings. The ratio of secondary turns to primary turns must equal a constraint of 10 to 1. For a given number of*

> *turns the operating frequency becomes the sole independent variable in coil design.*
>
> *The primary coil is the same diameter as the secondary coil. The ratio of conductor width to coil diameter is 18%. The sheet conductor is closely spiralled into two turns, for a secondary of twenty turns. For a conductor thickness as given by the maximum thickness vs frequency the volume of the primary metal must equal the volume of the secondary metal, based upon equal weights for primary and secondary metals. Brass can be 1.7 times thicker than copper, and bronze can be 2.8 times thicker than copper, for a given maximum thickness and frequency.*
>
> *The primary condenser should be made of the same amount of metal as the primary coil. Their weights should be about the same. The connecting leads must be short and the same width as the primary conductor. Large, multiplate air condensers are good, so are metal sheets and glass.*
>
> *The secondary capacitance is best a copper tubing ring around the outside of the H end of the coil. This ring must be open in one spot along its loop so as not to be a shorted turn. The end of the secondary wire connects to this capacity ring. Beer cans make the best elevated capacitors.*
>
> *The finished transformer must operate on a directly connected ground plane. A large plywood table with aluminum foil glued on to it, this covered with a plastic or glass sheet is the minimum required ground plane. Obviously the aluminum plane must be solidly earthed with special attention to minimizing ground lead inductance. In order to light a lamp a more broadcast station style of ground plane is required, but for testing purposes the metallized plywood "ten by" sheet is just fine. Good luck and good crystal set DX ing.*

There is no step by step guide to tuning this. Following basic DX, crystal set rules do not necessary apply IMHO.

All the theory is in here: [http://www.energeticforum.com/renewa...tml#post15858]

I do not want to talk about theory of operation here. Just observations and diffferences between this design and a normal crystal set design. For an unbiased Theory pull out the T-rex posts from the above thread and read. Not that there is not other great information in that thread but it can make you mind wander. Read the other posts after you determine if you understand Eric or not. I found page 26 a great place to start. Lots of good T-rex posts before that but they are a bit advanced and its hard to apply to those lessons before you understand the basics and that is what this is about.

Learning the basics.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

#50

04-21-2012, 03:17 PM

WARNING
BURN HAZARD

jake

Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **xee2**
> *@ jake*
>
> *Thanks for the info. I find lighting an LED with just a few feet of wire for an antenna to be impressive. What kink of results did EPD get with his?*

What powers the wireless AV plug?

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#51

04-21-2012, 04:46 PM

madhatter

Senior Member

Join Date: Nov 2010
Posts: 457

RF Probe

For those who may not have RF test gear, I found a simple RF circuit you can make to read RF voltages on a DMM.

N5ESE's Classic RF Probe

I made one this afternoon took about 30mins, used an aluminum cigar tube for the RF shielded body.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Sponsored Links

EPD LABORATORIES
TESLA ADVANCED SEISMIC
WARNING SYSTEM #2
YOU CAN HELP SAVE A MILLION LIVES

SUPPORT EPD LABORATORIES, INC.

#52

04-21-2012, 06:33 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *What powers the wireless AV plug?*

The AV plug appears to not have a return, but it actually does. The return path is through capacitive coupling. There is always some amount of capacitive coupling between any two conductors. Touching one side of the LED will increase the capacitance since that allows the high frequency current to flow over the surface of your body and increases the area of the capacitive plate. Normally this requires higher voltage than you have in order to work wirelessly so I am surprised that you can get it to work without a return wire with the voltage you have. The circuit is just a diode rectifier feeding an LED. The diodes prevent the reverse voltage across the LED from getting high enough to burn it out (LEDs can only have about 10 volts of reverse voltage before they will be destroyed).

see >>> one wire circuit - YouTube

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#53

04-21-2012, 06:43 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Awesome

Quote:

> Originally Posted by **madhatter**
> *For those who may not have RF test gear, I found a simple RF circuit you can make to read RF voltages on a DMM.*
>
> *N5ESE's Classic RF Probe*
>
> *I made one this afternoon took about 30mins, used an aluminum cigar tube for the RF shielded body.*

That is perfect! But wait lets make it better.

I don't care about accuracy. I want maximum sensitivity and stability so I can tell where the peak is. Becides ditching the the resistor should I consider another sized cap? 0.01uF seems a little large for use on a crystal set

@ all,

When using 0.5v PP on the function generator(lowest setting) across 1 turn of the primary I can use an AV plug to trace the secondary and "see" the high and low points along the entire secondary. Middle 5 turns have close to nothing. All the umph is at the top and bottom of the secondary.

Now to get that umph into the extra coil through the cap without loosing anything.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#54
04-21-2012, 07:37 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

@ jake

I calculated the field strength voltage 5 miles away from a 1,000 watt transmitter to be 0.03 volts per meter. This is a rough calculation but gives approximate value to be expected. You are getting much more out. I am impressed with your design. I worked several years as an antenna engineer developing very advanced antenna designs, so it takes a bit to impress me.

For a 10,000 watt transmitter this increases to about 0.10 volts per meter.

After I did the calculations the hard way, I found an online calculator that gives about the same answers:

Volt per meter and power density calculator

EDIT: A tuned antenna will suck up the RF energy much like an electronic vacuum cleaner, so I would expect your antenna to produce more than the ambient volts. It is actually possible to measure the lack of signal behind a tuned antenna which shows it has absorbed energy from an area much larger than itself.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Last edited by xee2; 04-21-2012 at 08:03 PM.

#55

04-21-2012, 08:19 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

xee2 you rock

The funny thing is: The coil responds best when excited at ~2200kHz. So I may be getting signal from another station but I cant hear it because 1620 is so loud.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

#56

04-21-2012, 08:31 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *I consider another sized cap? 0.01uF seems a little large for use on a crystal set*

What cap? There is no reason to go over 100 pf for coupling at 1 MHz. The smaller the coupling cap is the higher the resonator Q will be.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#**57**
04-21-2012, 08:34 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *xee2 you rock*
>
> *The funny thing is: The coil responds best when excited at ~2200kHz. So I may be getting signal from another station but I cant hear it because 1620 is so loud.*

No. If you are hearing 1620 then that is where the power is coming from. You will hear the station delivering the most power.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#**58**
04-21-2012, 09:09 PM

Loadstone
Member

Join Date: Jan 2010
Posts: 66

Hi Jake....nice work.

Thought i'd post my setup here for you.
I'm getting 25+ volts from my circuit and i can run a wall clock from it very well.
I made my own 2 gang vari cap from some bits of aluminium glued to a wooden dowel.
I'm using about 30 feet of data cable as antenna but anything works...even just an iron wire....
My antenna is about 12 to 15 feet in the air.
My coil is tuned to 774 khz but i also get 1550 khz as well.
The 774 station is about 20 kilometers away from me, and i think it is a 50,000 w transmitter....i think?
Here is a link showing my setup running a clock.
[clock radio - YouTube]

I wind the primary with taps, then i just wind the secondary straight over the primary.
I'd like to see someone else try this setup and see how it performs.
Cheers.
Scotty.

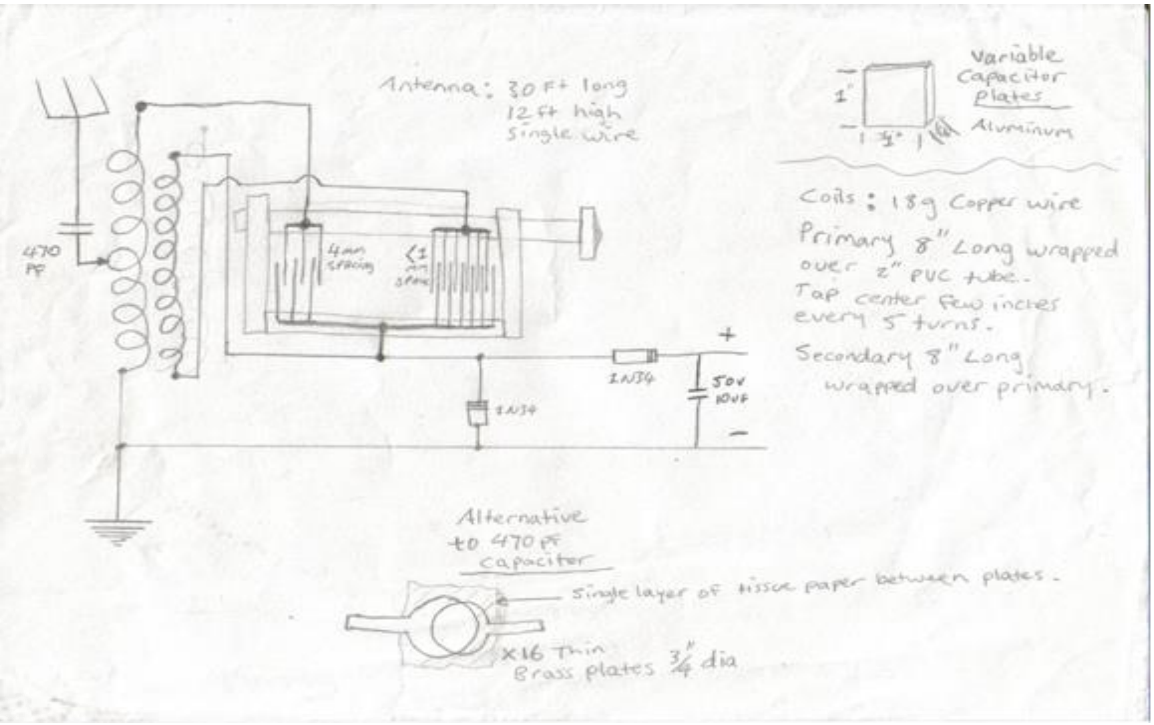
Antenna: 30 Ft long
12 ft high
Single wire
470 PF
4mm spacing
<1 mm space
1N34
1N34
50v 10uF
+
-
Variable Capacitor plates
1"
1 ½"
Aluminum
Coils: 18g Copper wire
Primary 8" Long wrapped over 2" PVC tube.
Tap center few inches every 5 turns.
Secondary 8" Long wrapped over primary.
Alternative to 470 pF capacitor
Single layer of tissue paper between plates.
X16 Thin Brass plates ¾" dia

04-21-2012, 09:59 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Standing/Stationary "wave"

@ all: Does this make sense?? Is this normal antenna theory stuff?

I am using bare 14gauge wire. So I can tap it at any point, and I do. There is a standing/stationary wave along the entire coil when excited at ~2200kHz. Max voltage is on the soda can, and again max at the bottom of the secondary by the grounding at ~2200kHz. With an area in the center of tht gives no readings. In this case I am feeding it with a function generator.

When attached to my magic "fireplace gouund " I get hightest signal close to the top of the secondary but not on the can and the voltage does not go up at the bottom.

This tells me I need more interwinding capacitance or a capacitor across or the split ring the secondary think
I think it's time for me to learn YouTube. I'm a video posting virgin. But the standing wave visual is cool.

And finally, my detector is an AV plug with germanium diodes. Crystal ear phone is attached across the LED and I can hear the station and the light glows when I tap points. This is the on the unpowered setup. Better yet I can hear the station on the "AV LED-head piece" 1" away from the secondary without touching, and it gets louder the closer you get.

Which means I am radiating. Which is bad.

So is this normal stuff?
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

____

Quote

#60
04-21-2012, 10:37 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

@ loadstone

NICE!

Wish I had seen that earlier. Tried earlier on to get a joule thief to work but failed. Then I got good light from the coil directly and dropped that idea. I'm borrowing a function generator so I'm spending my time working with that. I'll eventually get a thief to steal radiowaves.

Did you ever work with 1N34 germanium diodes? How do thay compare to the ones in the video?

04-21-2012, 11:58 PM

Loadstone
Member

Join Date: Jan 2010
Posts: 66

@ Jake....i have used 1N34's but in my clip they are a different type...i think 1n60.....
I started off getting low voltages and refined it until i can get up to 28 volts out @ 500 ua.....not much power but i can run various small devices or charge batteries now.
It's all fun....i also tinker with the capacity versus inductance ect....teaches you alot!
Scotty.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

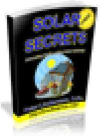

Download SOLAR SECRETS by Peter Lindemann
***Free - Get it now*: Solar Secrets**

#62
04-22-2012, 12:11 PM

jake
Senior Member

Join Date: Nov 2007
Posts: 169

FM anyone?

@loadstone,

Have you ever tried power recovery with an FM station??
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

#63
04-22-2012, 01:40 PM

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **xee2**
> *@ jake*
>
> *I calculated the field strength voltage 5 miles away from a 1,000 watt transmitter to be 0.03 volts per meter. This is a rough calculation but gives approximate value to be expected. You are getting much more out. I am impressed with your design. I worked several years as an antenna engineer developing very advanced antenna designs, so it takes a bit to impress me.*
>
> *For a 10,000 watt transmitter this increases to about 0.10 volts per meter.*
>
> *After I did the calculations the hard way, I found an online calculator that gives about the same answers:*
>
> *Volt per meter and power density calculator*
>
> *EDIT: A tuned antenna will suck up the RF energy much like an electronic vacuum cleaner, so I would expect your antenna to produce more than the ambient volts. It is actually possible to measure the lack of signal behind a tuned antenna which shows it has absorbed energy from an area much larger than itself.*

Just noticed this edit:

Right now it does nothing to a radio when I plug hook up the secondary but I can still light the led.

One previous setup it was making the radio louder when I hooked it up. I think it was when I grounded it to the house ground but the led would just fliker every once in a while.

This is where the extra coil and ring capacitor come in to play. I dont want to change the wire length, coil hight or coil diameter. So all I have is capacitance and I have tried: a 3 turn 505pf variable air cap. A 112pf HV air cap. and a couple of other crystal radio/pocket radio caps and they just dont do much of anything across the secondary.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#64
04-22-2012, 04:54 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *One previous setup it was making the radio louder when I hooked it up.*

Your house radio resonator and the crystal set resonator are mutually coupled. See ... AM Radio with wireless Loop Antenna - YouTube

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#65

04-22-2012, 06:27 PM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

The crystal radio project is to recover Teluric currents, not hertzian waves. This is why ground is THE most important part of the project.

recovering hertzian waves and magnifying them to power LED's etc.. is nothing new, novel but not what Eric was going for.

Diodes, 1N34A have a very low FW voltage, and a high 50V breakdown. 1N60 is nearly twice as high in forward voltage with a lower 40v breakdown, although the 1N60 is supposedly mfg to higher tolerances.

a ground impedance of 1ohm or less is going to be the biggest and for most an impossible requirement.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#66

04-22-2012, 06:47 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **madhatter**
> *The crystal radio project is to recover Teluric currents, not hertzian waves. This is why ground is THE most important part of the project.*
>
> *recovering hertzian waves and magnifying them to power LED's etc.. is nothing new, novel but not what Eric was going for.*
>
> *Diodes, 1N34A have a very low FW voltage, and a high 50V breakdown. 1N60 is nearly twice as high in forward voltage with a lower 40v breakdown, although the 1N60 is supposedly mfg to higher tolerances.*
>
> *a ground impedance of 1ohm or less is going to be the biggest and for most an impossible requirement.*

Yes, you are correct. However one can use the Hertzian waves to tune their secondary, just like we are doing with the function generator. LEDs make pretty good HF indicator lights.

So whats you take on my standing wave question? last page
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#67

04-22-2012, 07:06 PM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

Quote:

> Originally Posted by **jake**
> *Yes, you are correct. However one can use the Hertzian waves to tune their secondary, just like we are doing with the function generator. LEDs make pretty good HF indicator lights.*
>
> *So whats you take on my standing wave question? last page*

Not sure I understand what you're after, could you explain a bit more?

Also this is going to be frustrating, but all leads will impact the L&C of the circuit, I found that by using the coil itself it's a better compramise then adding in wire length. the leads and LED also effect the circuit too, it's a big balance game.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#68

04-22-2012, 07:53 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

[QUOTE=madhatter;189674]Not sure I understand what you're after, could you explain a bit more?[Quote]

As far as the Standing wave. Should there be one along the lenght of the wire?

Quote:

> Originally Posted by **madhatter**
> *Also this is going to be frustrating, but all leads will impact the L&C of the circuit, I found that by using the coil itself it's a better comprAmise then adding in wire length. the leads and LED also effect the circuit too, it's a big balance game.*

Yes, Yes, and Yes I noticed that.

A wise man once said "we were going to need a new set of tools to work with these new concepts" or something along those lines. Once I realized I could easily light an led I thought that it could make a handy dandy indicator. THEN..

I get my hands on a 10MHz function generator. I start testing and messing but lacking a fuctional ground I can only take my results as mere observations of a crystal radio and NOT a TRT. What did I learn....that leads are the enemy. Thats when I made the leap form FWBR to AV plug and that made all the difference. Thanks to all those who helped with that. No leads and 1 point of contact. That makes a nice little indicator. With that indicator I can touch the coil at any point. And it appears to make very little change to the resonant frequency. The change I do see is usually caused by me holding the AV plug. BUT I cant get the AV plut to light on a stick.!

jake, jastermake @ yahoogroup

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#69

04-22-2012, 09:11 PM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

[QUOTE=jake;189683][QUOTE=madhatter;189674]Not sure I understand what you're after, could you explain a bit more?
Quote:

> As far as the Standing wave. Should there be one along the lenght of the wire?
>
> jake, jastermake @ yahoogroup

I'm going to be a bit lazy here and point you to this, Standing waves and resonance : TRANSMISSION LINES it may help answer some questions, if you've got more after, let me know.

the length, are you referring to the physical wire length or the length of the coil? there is also the issue of electrical length. gets more involved with RF and transmission, esp here since C is no longer a fixed value.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#70

04-22-2012, 10:57 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *As far as the Standing wave. Should there be one along the lenght of the wire?*

For parallel resonator or self resonant coil:
When the resonator is at resonance there will be a standing wave on the inductor wire with a peak voltage measured across the ends. Note that this is an RF wave not DC. Best way to measure this is with scope using very small capacitor as probe. There will be multiple peaks if the resonator is resonating at a harmonic of fundamental resonator frequency.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Last edited by xee2; 04-22-2012 at 11:39 PM.

#71

04-23-2012, 11:15 AM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Oooops!!!

So I went outside soldered into a ground wire and set to work. With 1 turn, 1/2" flat copper stock primary. Tried different arials, different hights, and so on and so on. Got it down to 1620(heard it on the radio 15' away) with a big aluminum salad bowl. , I went back to no terminal and put on my big fat 2 turn primary. Started testing with this when

A small white exlporer sticker on side said "Texas Communicat........." Could not read the whole thing. Pulled into my driveway waved nicely at me then drove off nice and slow....

There is a company called "Texas Communications" that runs communication services and it may have been one of their trucks. but still It did not feel right.

Did I break a rule?? FCC or other??

Function Generator signal level never went past 1/3. and was usually all the way down..

BK Precision 4017A, 10mHz, 10v(max) p-p into 50ohm, minium .25v p-p across my scope.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#72

04-23-2012, 11:52 AM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
>
> *So I went outside soldered into a ground wire and set to work. With 1 turn, 1/2" flat copper stock primary. Tried different arials, different hights, and so on and so on. Got it down to 1620(heard it on the radio 15' away) with a big aluminum salad bowl. , I went back to no terminal and put on my big fat 2 turn primary. Started testing with this when*
>
> *A small white exlporer sticker on side said "Texas Communicat........." Could not read the whole thing. Pulled into my driveway waved nicely at me then drove off nice and slow....*
>
> *There is a company called "Texas Communications" that runs communication services and it may have been one of their trucks. but still It did not feel right.*
>
> *Did I break a rule?? FCC or other??*
>
> *Function Generator signal level never went past 1/3. and was usually all the way down..*
>
> *BK Precision 4017A, 10mHz, 10v(max) p-p into 50ohm, minium .25v p-p across my scope.*

You broke no rules. He probably pulled into your driveway by mistake and then realized he was not where he should be.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#73

04-23-2012, 12:12 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Just remembered there is also a strong resonant point at 7MHz that I played with. Does this change anything?
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Last edited by jake; 04-23-2012 at 12:14 PM.

#74
04-23-2012, 01:32 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *Just remembered there is also a strong resonant point at 7MHz that I played with. Does this change anything?*

I doubt that your signal generator will put out enough power to interfere with anything unless you have neighbors within a 100 feet and then they would need to have a radio tuned to the frequency you are using (unlikely).
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Last edited by xee2; 04-23-2012 at 01:35 PM.

#75
04-23-2012, 02:25 PM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

The FCC has no sense of humor, be careful.
wiki link on FCC broadcasting regs
Title 47 CFR Part 15 - Wikipedia, the free encyclopedia

20dBm = 100mW is the maximum allowed, also there is limits on antennae length. ground and above ground total to 3meters.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#76

04-23-2012, 10:58 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Im Back!

So before I was freaked out earlier. I was tring to figure out the standing issue I have.
@ mad, I like it when you lazy.

So what I really wanted to ask was why is there only a partial wave along the coil at resonance? With a 1 turn primary I get 1/4 of a wave along the entire coil. Meaning the the can has highest "potential and the potential along the coductor goes down to nothing as you approach ground.

However a few days ago I had 1/2 wave at resonance when I was using my primary instead of a 1 turn copper strip. Meaning highest potential on terminal, the 5 turns on in the middle of the coil had nothing and voltage went back up(actually -) as you approached the ground.

I had just got my primary back on when. I pooped my pants.. Spent the rest of the day at FCC.. and wiki.

What did I learn there: 2mHz-3.5mHz is wierd and is allocated to maritime, ship to shore, and other , but the coast guard uses it as well and I can't forget 2.1735MHz mobile distress and calling frequency.. I was all over that one today. With the right termination/salad bowl. I learned I can keep it under 2MHz and still play as long as I keep it turned down all the way..

I used a FWBR made from SF35s and SF37s diodes. Across the DC of the FWBR I used a 100pF cap. Powering the function generator at 1.8MHz the max I could get was 90mA straight across a DMM out from the rectifier. How do I go from that to "milliwatts of DC input power to the final RF stage"?

time to do the pots.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#77

04-24-2012, 08:22 AM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *and voltage went back up(actually -) as you approached the ground.*

You must have had a bad ground connection. Ground should never have a voltage on it (it is the reference for all other voltages).

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#78

04-24-2012, 08:30 AM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *I used a FWBR made from SF35s and SF37s diodes. Across the DC of the FWBR I used a 100pF cap. Powering the function generator at 1.8MHz the max I could get was 90mA straight across a DMM out from the rectifier. How do I go from that to "milliwatts of DC input power to the final RF stage"?*

You need to measure imput impedance of antenna to determine how much power is going into it. But it will look like like a small capacitor. To get power into the antenna you need to make a matching network to convert generator impedance to complex conjugate of antenna impedance. As is, 90% of power appllied to your wire antenna is being reflected back to generator (less than 10% is going into antenna). So the amount of power your antenna is readiating is very small.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#79

04-24-2012, 09:11 AM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **xee2**
> *You must have had a bad ground connection. Ground should never have a voltage on it (it is the reference for all other voltages).*

Awesome.. Thats why I went outside and soldered directly to decent "ground". Did not stay out long enough to prove it to myself.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

____

#80

04-24-2012, 09:24 AM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **xee2**
> *You need to measure imput impedance of antenna to determine how much power is going into it. But it will look like like a small capacitor. To get power into the antenna you need to make a matching network to convert generator impedance to complex conjugate of antenna impedance. As is, 90% of power appllied to your wire antenna is being reflected back to generator (less than 10% is going into antenna). So the amount of power your antenna is readiating is very small.*

I was hoping for more answers like this...

BUT what the #?*! did you just say..

I never wanted to transmit.. never planned on it.. still dont want to. I'm more of a receiver.

Is "very small" enough to make a radio 15' away scream like a banchee. By scream like a banchee I mean, completely drown out the local station that was coming in loud and clear? And boost the volume?

I have absolutely no reference of radiated RF and its capabilities. So I appreciate the replies.

jake

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#81

04-24-2012, 09:42 AM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *I was hoping for more answers like this...*
>
> *BUT what the #?*! did you just say..*
>
> *I never wanted to transmit.. never planned on it.. still dont want to. I'm more of a receiver.*
>
> *Is "very small" enough to make a radio 15' away scream like a banchee. By scream like a banchee I mean, completely drown out the local station that was coming in loud and clear? And boost the volume?*
>
> *I have absolutely no reference of radiated RF and its capabilities. So I appreciate the replies.*
>
> *jake*

At 15 feet you are coupling to the receiver directly not picking up the RF radiation. A wavelength at 1 MHz is 1000 feet so 15 feet is only a few percent of a wavelength away. This is called near field coupling and is very strong.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#82

04-25-2012, 05:17 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Really 90%

Quote:

> Originally Posted by **xee2**
> *You need to measure imput impedance of antenna to determine how much power is going into it. But it will look like like a small capacitor. To get power into the antenna you need to make a matching network to convert generator impedance to complex conjugate of antenna impedance. As is, 90% of power appllied to your wire antenna is being reflected back to generator (less than 10% is going into antenna). So the amount of power your antenna is readiating is very small.*

FIELD INTENSITY METER

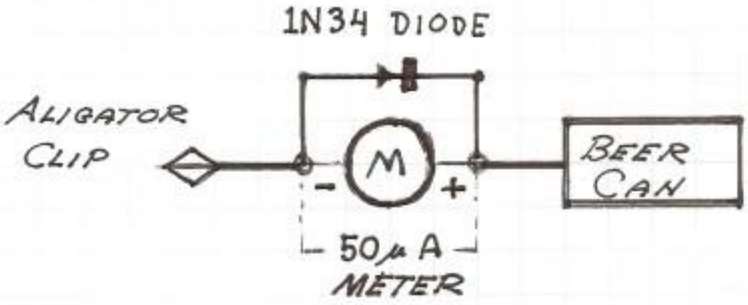
1N34 DIODE
ALIGATOR CLIP
M
-
+
BEER CAN
50µA METER

TUNING SETUP EXTRA COIL

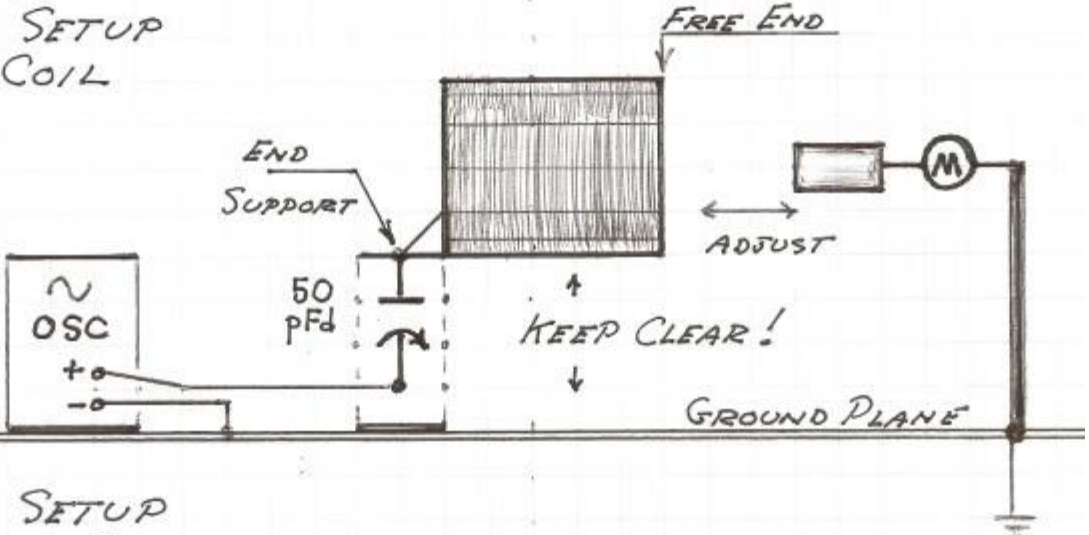
FREE END
END SUPPORT
50 pFd
OSC
+
-
M
ADJUST
KEEP CLEAR!
GROUND PLANE

TUNING SETUP SECONDARY COIL

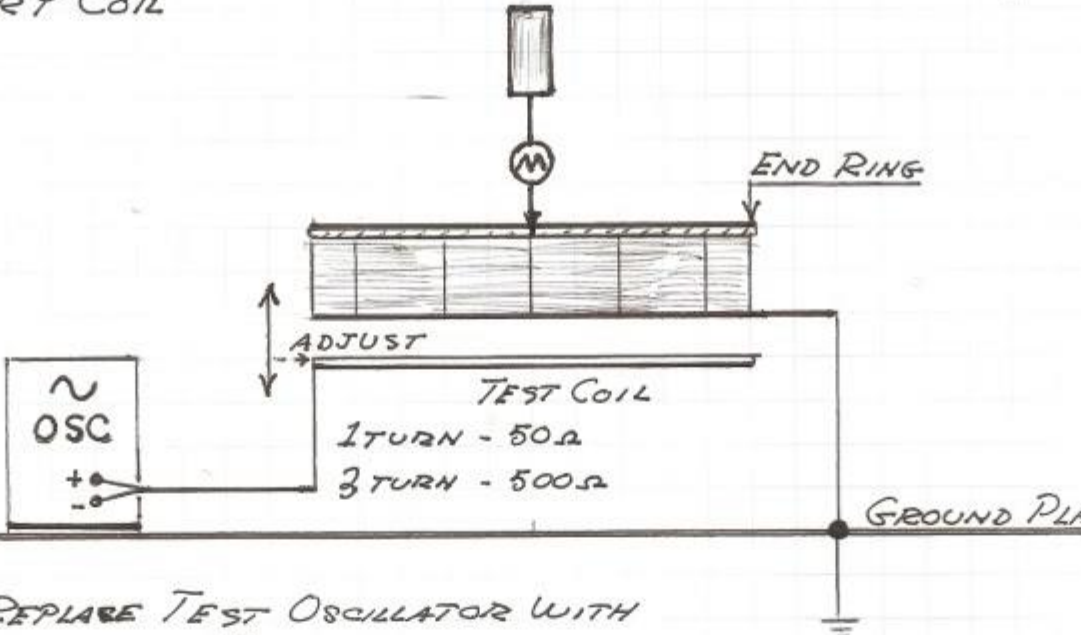
M
END RING
ADJUST
TEST COIL
1 TURN - 50Ω
3 TURN - 500Ω
OSC
+
-
GROUND PL

NOTE, REPLACE TEST OSCILLATOR WITH A.M. RADIO TO RECIEVE.

#83

04-25-2012, 06:01 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *In the bottom diagram 90% is being reflected to the generator???*

It is hard to tell from diagram. I was assuming you were just connecting a wire to signal generator. In this case there is a whole circuit attached. But I suspect that most of the energy is being reflected back into the generator. Getting an electrically small antenna (length only a small fraction of a wavelength) to radiate all of the power being fed to it requires much work. When the length gets to be 1/4 wavelength the antenna impedance will be close enough to 50 ohms to radiate most of power. But less than than 1/4 wavelength will reflect most of the power fed to it from a 50 ohm source.

If you are tuning with the set up in diagram, be aware that the frequency is going to be effected by the loop attached to the generator. I think it would be better to use a short wire well away from the receiver resonator to transmit with when tuning the resonator.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#84

04-25-2012, 06:26 PM

dR-Green
Gold Member

Join Date: Apr 2008
Location: Merlin's Fort
Posts: 1,220

Quote:

> Originally Posted by **Web000x**
> *I don't think I've seen an extra coil on your setup, or have I?*

No you haven't, not yet. I suppose I could make an extra coil first and test it on the spiral
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

"Knowledge is cosmic. It does not evolve or unfold in man. Man unfolds to an awareness of it.

He gradually discovers it." - Walter Russell

"Once men died for Truth, but now Truth dies at the hands of men." - Manly P. Hall

#85

04-25-2012, 06:34 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **xee2**
> *I think it would be better to use a short wire well away from the receiver resonator to transmit with when tuning the resonator.*

Being close to the resonator will also effect tuning. I recommend sweeping the frequency to see where it resonates, adjust, move away, sweep frequency to see where it is resonant, ... etc.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#86

04-25-2012, 06:47 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **xee2**
> *It is hard to tell from diagram. I was assuming you were just connecting a wire to signal generator. In this case there is a whole circuit attached. But I suspect that most of the energy is being reflected back into the generator. Getting an electrically small antenna (length only a small fraction of a wavelength) to radiate all of the power being fed to it requires much work. When the length gets to be 1/4 wavelength the antenna impedance will be close enough to 50 ohms to radiate most of power. But less than than 1/4 wavelength will reflect most of the power fed to it from a 50 ohm source.*
>
> *If you are tuning with the set up in diagram, be aware that the frequency is going to be effected by the loop attached to the generator. I think it would be better to use a short wire*

> *well away from the receiver resonator to transmit with when tuning the resonator.*

-1 turn primary flat 1/2". in series across the Function generator
-20 turn secondary 29.5-30.0 meters. about 2" above the secondary, same diameter
-Coil was most active at 2000kcycles. It lit an av plug the brightest at the greatest distance.

Considering these three observations, and considering the 20 turn secondary as the antenna;

Can I conclude the coil is acting like a (1/4 lamda) 150m antenna? (i.e. I have electrically lengthened the 30m wire I used in the secondary)
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#87
04-25-2012, 07:45 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *Can I conclude the coil is acting like a (1/4 lamda) 150m antenna? (i.e. I have electrically lengthened the 30m wire I used in the secondary)*

It is possible but there is no way for me to tell if that is happening. But the resonant frequency is going to change when you disconnect the generator if the generator is tightly coupled to the resonator. I think your objective is to get the receiver tuned to the correct frequency, not to make a transmitter tuned to the correct frequency.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#88
04-25-2012, 09:00 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **xee2**
> *I think your objective is to get the receiver tuned to the correct frequency, not to make a transmitter tuned to the correct frequency.*

You are correct

and on the way there I'm disecting each component.

I do not know what frequency the secondary should resonate as it is shown in the diagram. What part the secondary plays in the TRT I do not know. I think Eric may be the only one. But this is not a TRT, just yet.

So I am looking at the coil from all angles. How it was a radiator in this case study, not exactly what I was expecting but Ill take what I can get.

When excited @ 2000kcycles it radiates quite readily but at 1620 not so well . Same goes for receiving I thought...

'CAN' or Beer can/ dr pepper can, is the standard end termination for the Crystal Radio Initative. Dollard recommended it and I think those building are also using it in testing.

I need a lot of 'CAN' to bring it down to 1620. But how much 'CAN' will be provided by the extra coil, I do not know. The two split rings in the final design of the TRT will allow for some adjustment. Of 'what' is still to be determined. So much I do not know.

As far as the 150m question I was just trying to figure out why it ended up that way since I was only using 30m of wire. But I think I got that part figured out.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#89

04-25-2012, 09:48 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Trt

The final product.

by EPD

CONCATENATED RESONANT COILS, 1000K

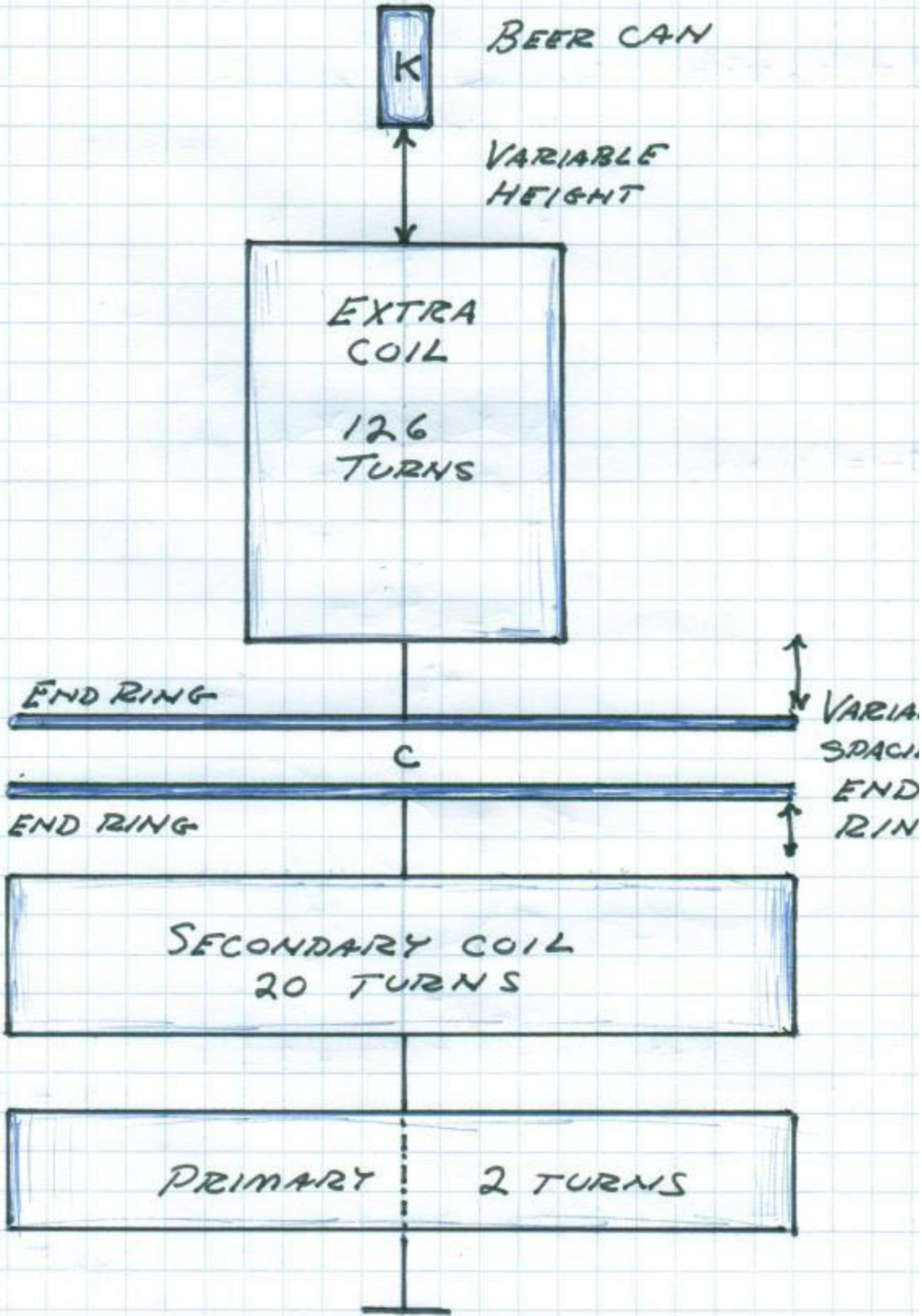
K
BEER CAN
VARIABLE
HEIGHT
EXTRA
COIL
126
TURNS
END RING
VARIA
C
SPACI
END
END RING
RIN
SECONDARY COIL
20 TURNS
PRIMARY
2 TURNS

#90

04-25-2012, 10:31 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

@ jake

Wow, that looks complicated.

To lower your frequency you need to add parallel capacitance to the coil. Probably about 5 to 10 pF. You can make an adjustable capacitor with two sheets of copper overlapped with insulator in between and held together with a clothes pin.

04-26-2012, 11:07 AM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

I've seen videos of people lighting av plugs wirelessly from an SEC. What kind of voltages(primary and secondary)/frequencies are needed to get this effect?
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Download SOLAR SECRETS by Peter Lindemann
Free - Get it now: Solar Secrets

#92

04-26-2012, 03:03 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

I have several videos posted on youtube at xee2vids. 1.5 volts will light fluorescent tube a foot away wirelessly.
Slayer exciter circuit - 1 AA battery - YouTube
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Last edited by xee2; 04-26-2012 at 03:19 PM.

#93

04-26-2012, 04:51 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

xee2,
What kind of voltages were your getting on the secondary coil?
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#94

04-26-2012, 05:24 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *xee2,*
> *What kind of voltages were your getting on the secondary coil?*

Unknown. but I think over 20,000 volts.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#95

04-28-2012, 01:13 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

20kv nice

I played around with HV a lot last year. 50uF spark discharges of up to 5000V. Neon sign transformers and spark gaps are fun. I could light neons 15' away when stuck into the ground. The whole plastic as conductor also worked with NSTs.

I also did 3 Henry discharges/oscillations. Only 24v but scary with the amps I knew were there. I melted 4 neons in parallel this way. It only took 12 volts to melt 1 neon.

Any following Tesla's advice I went with smaller and smaller caps and faster and faster frequencies. That is when Eric came out with the crystal initiative and I thought that would be a good way to learn the in and outs of HF.

As I proceed I am trying to figure out what is HF related and what is HV related. I am so used to working with and reading about HF/HV at the same time I forgot that they are each their own entity.

I am starting to like this HF stuff.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#96

04-28-2012, 09:03 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

@madhatter,

What causes the two small peaks on each side of the large resonant point? I don't have a scope but they "show" up as audio on the piezo. It lets me know I'm getting close.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#97

04-28-2012, 10:09 PM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

it may have to do with the phase shift, before resonance the current leads the voltage at resonance it is in phase and after the current lags the emf. The shift may be what you're picking up.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#98
04-29-2012, 08:10 AM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *@madhatter,*
>
> *What causes the two small peaks on each side of the large resonant point? I don't have a scope but they "show" up as audio on the piezo. It lets me know I'm getting close.*

I am not sure what your setup is so this is hard to answer. But it is probably the modulation on the carrier. This shows up as sidebands in the frequency domain. You can sometimes even see this from signal generators due to 60 Hz leakage.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#99
04-29-2012, 07:09 PM

David G Dawson
Senior Member

Join Date: Mar 2008
Posts: 380

Smokey - Introduction

Hello All,
Here to introduce myself and have an Extra Coil discovery that you all need to know about which I will post shortly.
Have been over at 'n6kph' and working with, David, Jake, Scott and Gary Peterson of 'TFCBooks' and all have been most helpful.
This introduction is a copy of one I made at n6kph in response to a Post:

Gary,
These tests of Erics are designed to gain some distance tuning knowledge with the Extra Coil and with the Secondary test, to tune in the local AM Station via the Hertzian component and not the Telluric.

I have four Oscilloscopes, 1 x 100 Mhz (HP 1741A), 2 x Tektronix Valve and a Telequipment solid state which is UK Tektronix.
Have a fully set up Lab that has basically taken me 10 years to acquire or build the required tools.
Only thing I don't have is a Spectrum Analyser and working on that now.
Am not a Ham but fully qualified to be so and have a considerable collection of Vacuum Tubes that go along with my research, Moray/Ion Valve/Smith/Kapanadze type research.
I am 70 years of age and been at this full time since retiring 10 years ago.
I am ex Radio Station, RANavy, NASA Tidbinbilla Space Track and finally electronic designing for the Auto Industry.
Repairing a Yaesu FT-901 Transceiver due to its AM driver capabilities ready for the next step but these initial problems need to be overcome with respect to design.
Yes, I have seen your repeated information but can only do one step at a time and will get to do your Oscilloscope tests at the right time which is not right now as the working data is suspect as is my Extra Coil.
I am also into Aetheric Weather Engineering after Wilhelm Reich and TJ Constable etc which is another story again but helps with the Aetheric ubderstanding associated with all of this.
I have 13 devices here that can make it rain and all of which are not recognised by 'science' just as is Tesla's TMT and why I am also following this path.
Letting you know where I come from.

Smokey
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#100
04-29-2012, 07:18 PM

David G Dawson
Senior Member

Join Date: Mar 2008
Posts: 380

Extra Coil Math Discovery

This Post is repeated at 'n6kph'.

A Discovery - Extra Coil:
I find repeatedly that simple Maths is a significant part of discovery and has allowed me to build many devices that others have failed at or not even bothered with, all because it appeared too hard.
I now have two Extra Coils, one considered wrong Math and the newer considered the correct Math.
Question to me was that the old worked and the new doesn't but why?
You may remember me saying that my first attempt at this coil worked as I was picking up my local 2 Kw station with the Headphone capacitor antenna some 8 to 10" away from the Extra Coil and it was proven to be a Telluric signal and not Hertzian.

The Math:
Old Coil data - works:
CalcF = 1.557 Mhz (1.557 x pi/2 = 2.4457) MULTIPLY
53 turns of 14 awg (same as Secondary wire size) with a diameter of 195.2mm
L = 30.829 meters (9.7244 Mhz)
Actual 1/4 wave Fres = 2.4655 Mhz AFTER F 1.557 Mhz
Error long by 198 Hz which means I need more turns over the 53 I did have but probably only one required and this should increase signal strength as well.

New Coil data - doesn't work:
Calc F = 1.557 Mhz (1.557 / pi/2 = 991.2 Khz) DIVIDE
125 turns of 21 awg with diameter of 195.2mm.
L = 75.61 meters (3.965 Mhz)
Actual 1/4 wave Fres = 991.4 Khz BEFORE F 1.557 Mhz

What this is saying is that you need to find the 1/4 Wave point from your number of turns and that needs to match your signal frequency MULTIPLIED by pi/2. (1.557 x (pi/2) = 2.4457 Mhz)

When I had completed the first coil I was happy with the Math and why I went ahead with the build.
It is only now, on analysis that the Math can be presented on paper and you can see what is happening here.
What I first built was, I feel, correct as the pi component is at the correct point AFTER the centre point going on to the full 1/4 wave and not before the 1/8th wave point.
This AFTER point is where Tesla's 'standing wave' is permitted/allowed to occur.
The second coil has its pi point BEFORE the 1/8 wave point and cannot resonate correctly and will be difficult to tune as is noted in the documentation many times.
What is of significance here is exactly where the pi point is (where your reference signal is) and is not a harmonic or octave or any of the regular resonance points but WITHIN THE FULL QUARTER WAVE at the 1/4 wave/pi/2 point and this is where your reference signal is located. (2.4457/(pi/2) = 1.557 Mhz)

If your Extra Coil does not work or is difficult to tune then the above will explain why.
If you then build a new Coil to the guide above, you will meet with guaranteed success as proven above.

I now have a difficult challenge to figure out if Tesla was on the wrong side with his pi point and working on that now.
I will be building a new coil on a new more open frame but adding one more turn to home in on the exact point as I still have the original wire ready to rewind.
Will advise as I proceed.
Hope you can understand what is documented above.
You MUST know your simple Maths if you are ever going to succeed in this business as I relate from empirical experience.
Good luck with your builds.
Will introduce myself over at Energetic and present this document as well.
Thanks.

Smokey

PS:
If anybody can see an obvious piece of data that I have overlooked, please respond.
David,
Would like to hear Eric's comments on this 'discovery'
Thanks.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#101
04-29-2012, 08:04 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

WELCOME

Good to see you here.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#**102**
04-29-2012, 08:23 PM

Web000x
Silver Member

Join Date: Apr 2009
Posts: 539

Quote:

> Originally Posted by **David G Dawson**
> *This Post is repeated at 'n6kph'.*
>
> *A Discovery - Extra Coil:*
> *I find repeatedly that simple Maths is a significant part of discovery and has allowed me to build many devices that others have failed at or not even bothered with, all because it appeared too hard.*
> *I now have two Extra Coils, one considered wrong Math and the newer considered the correct Math.*
> *Question to me was that the old worked and the new doesn't but why?*
> *You may remember me saying that my first attempt at this coil worked as I was picking up my local 2 Kw station with the Headphone capacitor antenna some 8 to 10" away from the Extra Coil and it was proven to be a Telluric signal and not Hertzian.*
>
> *The Math:*
> *Old Coil data - works:*
> *CalcF = 1.557 Mhz (1.557 x pi/2 = 2.4457) MULTIPLY*
> *53 turns of 14 awg (same as Secondary wire size) with a diameter of 195.2mm*
> *L = 30.829 meters (9.7244 Mhz)*
> *Actual 1/4 wave Fres = 2.4655 Mhz AFTER F 1.557 Mhz*
> *Error long by 198 Hz which means I need more turns over the 53 I did have but probably only one required and this should increase signal strength as well.*
>
> *New Coil data - doesn't work:*
> *Calc F = 1.557 Mhz (1.557 / pi/2 = 991.2 Khz) DIVIDE*
> *125 turns of 21 awg with diameter of 195.2mm.*
> *L = 75.61 meters (3.965 Mhz)*
> *Actual 1/4 wave Fres = 991.4 Khz BEFORE F 1.557 Mhz*
>
> *What this is saying is that you need to find the 1/4 Wave point from your number of turns and that needs to match your signal frequency MULTIPLIED by pi/2. (1.557 x (pi/2) = 2.4457 Mhz)*
>
> *When I had completed the first coil I was happy with the Math and why I went ahead with the build.*

> *It is only now, on analysis that the Math can be presented on paper and you can see what is happening here.*
> *What I first built was, I feel, correct as the pi component is at the correct point AFTER the centre point going on to the full 1/4 wave and not before the 1/8th wave point.*
> *This AFTER point is where Tesla's 'standing wave' is permitted/allowed to occur.*
> *The second coil has its pi point BEFORE the 1/8 wave point and cannot resonate correctly and will be difficult to tune as is noted in the documentation many times.*
> *What is of significance here is exactly where the pi point is (where your reference signal is) and is not a harmonic or octave or any of the regular resonance points but WITHIN THE FULL QUARTER WAVE at the 1/4 wave/pi/2 point and this is where your reference signal is located. (2.4457/(pi/2) = 1.557 Mhz)*
>
> *If your Extra Coil does not work or is difficult to tune then the above will explain why.*
> *If you then build a new Coil to the guide above, you will meet with guaranteed success as proven above.*
>
> *I now have a difficult challenge to figure out if Tesla was on the wrong side with his pi point and working on that now.*
> *I will be building a new coil on a new more open frame but adding one more turn to home in on the exact point as I still have the original wire ready to rewind.*
> *Will advise as I proceed.*
> *Hope you can understand what is documented above.*
> *You MUST know your simple Maths if you are ever going to succeed in this business as I relate from empirical experience.*
> *Good luck with your builds.*
> *Will introduce myself over at Energetic and present this document as well.*
> *Thanks.*
>
> *Smokey*
>
> *PS:*
> *If anybody can see an obvious piece of data that I have overlooked, please respond.*
> *David,*
> *Would like to hear Eric's comments on this 'discovery'*
> *Thanks.*

Smokey,

I like the Energetic Forum 'style' so I will ask questions here. Also, Eric is more apt to see info posted on the forum and *not* the Yahoo group since he doesn't know how to navigate it. Discussion here in the open is going to preserve these conversations *on one thread* for Eric to see since there is no way I can remember all details that he might like to know. How did you arrive at your calculations? I only briefly looked but couldn't find how you got a calculated frequency of 1.557 MHz.

That being said, if you can present your findings in a manner that builds them up from Eric's initial design equations to where your experiments have taken you, I will refer Eric to this thread. He told me the other day that he wanted to check the forum again to see if there has been any activity so he is looking for somebody with a computer to use.

Geometric_Algebra has been the only one to be able to functionally correspond with Eric thru the postal service. Eric received a letter from Geometric_Algebra that included coil dimensions, measurements, errors, etc. and sent back a response that is on the Eric P. Dollard thread. If you send Eric something that is a worthwhile question, he will respond. However, since you seem to be on the other side of the world, you should document your results here on this forum either in this thread or Eric's thread so that you are using a medium that Eric is used to.

I would be more than willing to use my own printer and paper to get questions to Eric for everybody that doesn't have convenient access the the United States Postal Service. However, I have no money and my printer is out of ink. If somebody wants to help with these problems, I will print and mail all day long. Otherwise, hope that Eric can have access to the forum long enough to fully digest your post or you'll have to mail him from your location.

Sorry about the inconveniences,

Dave
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#103
04-29-2012, 08:50 PM

David G Dawson
Senior Member

Join Date: Mar 2008
Posts: 380

Extra Coil Confirmation

Confirmation:
Here is the Tesla data based on the Colorado Spring's Notes:
Extra Diameter = 2.52m
One Turn 2pir = 7.9168m
Turns = 100 Total Length = 791.7m
Full Wave = 378.67 Khz
1/4 Wave = 94.6675 Khz
Operating Signal Point 94.6675 MULTIPLY pi/2 = 148.703 Khz
Where Tesla should have been working and NOT here:
94.6675 DIVIDE pi/2 = 60.267 Khz where he was working and supported in the CS Notes 'about 59 Khz'.

Think this is sufficient evidence to point to the base reason why the Extra Coil is so hard to tune and is very much not so at the AFTER and not BEFORE 1/8 Wave point.
It may have been difficult for Tesla to get to the higher frequency where he should and probably knew he should be working at and this is his compromise.
He may have also considered that what he was attempting to achieve would work from both sides and chose the DIVIDE path when he should have gone the MULTIPLY way.
Hope this makes it clear as to what is required with your Extra Coil to make it work.

Am satisfied now that this is the case and will rebuild my Extra Coil to the new dimensions which will be:
F = 1.557 Mhz MULTIPLY (pi/2) = 2.4457 Mhz = 1/4 Wave.
Full wave = 9.78292 Mhz = 30.644 meters
Diameter = 195.2mm (r = 97.6 mm)
1 Turn = 2pir = 613.23 mm
Number of Turns = 30644/613.23 = 49.97
Wire size, same as Secondary and 14 awg in my case.
Number of turns very close to my original figure of 53 but would suggest that you add about 5 extra turns and then trim or better to probe back on turns that you have cleared of insulation when you can get a Sig Gen set on the exact frequency of your station and then tune for maximum gain in the Headphones or on the Intensity meter.
Hope this explanation is now clear to everyone.

Will be honest and say that using 21 awg for the new Extra did not appear correct as I could not see being able to pump this frail device with any watts to achieve a goal.
This IS different!
Will now go build my new Extra Coil and make this fact for good as it has already been proven by empirical means.
I have Tesla nodding approval.
Thanks.

Smokey
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#104
04-29-2012, 10:35 PM

David G Dawson
Senior Member

Join Date: Mar 2008
Posts: 380

Radio Station Frequency

David,

The 1.557 Mhz is my local Radio Station that I am tuning into.
If you follow my last Post it is fully exampled there and shows what I feel Tesla would like to have done but couldn't due to technical shortages of his equipment or he may have just gone the wrong way.

Wish I could help more from afar but not easy - I have 5 Computers here, all the old ones I have fixed their problems and sit here unused or just kept as data monitors.

Think this IS a significant discovery and just needs builders to first understand what I have proposed and then to build following my examples.
The important issue here is where your AM Radio Station sits in the 1/4 Wave and needs to be at the 0.637 point of the 1/4 Wave of your Coil.
From my previous example - 1.557/2.4457 = 0.637 or on the HIGH side of your local AM Radio Station signal
I will now go away and work on what that point is, as it must have some other significance besides pi.
The Tesla way was 1.557/991.2 = 1.5708 or better as a comparison 991.2/1.557 = 0.637 the same but on the LOWER side of your Radio signal.

I knew 0.637 was something to do with AC and here is the connection:

Measuring the Sine Wave

The Average value of a Sine Wave.
Note now the association here with pi and the average value of a Sine or AC wave and this in itself speaks volumes.
This is where the 'magnification' takes place.
Now that we have a 'something' to work with we may be able to make some considerable steps in evolving this Free Enegy that we seek in all of our Coil designs.
Hope you have all been able to follow this unravelling so to speak as it just simply now makes all that much sense.
I have a Coil to build.

As a side note, please note that I have been at Energetic since March 2008 sitting on the sidelines and waitimg for my moment.
Hopefully it has arrived.

Smokey
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#105

04-30-2012, 12:47 AM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

some fun with math.
300,000,000 x Pi/2 = 300,000,000 x 1.5707 = 471,238,898.04
300,000,000/471,238,898.04 = 0.636619772
1/(Pi/2) = 0.636619772
2/Pi = 0.636619772
0.636619772 x (2 x Pi) "revolution" = 4 "radians"
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#106
04-30-2012, 06:56 AM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **madhatter**
> *some fun with math.*
> *300,000,000 x Pi/2 = 300,000,000 x 1.5707 = 471,238,898.04*
> *300,000,000/471,238,898.04 = 0.636619772*
> *1/(Pi/2) = 0.636619772*
> *2/Pi = 0.636619772*
> *0.636619772 x (2 x Pi) "revolution" = 4 "radians"*

Not fun.

So it comes down to "A/(A*Pi/2)=0.63661977"
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#107

04-30-2012, 07:11 AM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

@ David G Dawson

Have you built an antenna using the same EPD design that Jake is using? If so how well does it work?
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#**108**
04-30-2012, 07:56 AM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

Quote:

> Originally Posted by **jake**
> *Not fun.*
>
> *So it comes down to "A/(A*Pi/2)=0.63661977"*

why not fun?

I think you mean, 'A/(A(Pi/2)=0.636619772'

What I was getting at was that 0.6366.... is a function of Trig. As Smokey mentioned the A/C useage it's the rms multiplier value for 1/2 of the sine wave of a rectified wave for DC. and .707 is the full wave. a lot of these 'numbers' will factor back out and can seem 'magical', they do help in recognizing patterns and relations though.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#**109**
04-30-2012, 02:46 PM

ewizard
Gold Member

Join Date: Jan 2010
Location: Outside the Matrix
Posts: 1,015

Quote:

> Originally Posted by **David G Dawson**
> *David,*
> *The 1.557 Mhz is my local Radio Station that I am tuning into.*
> *If you follow my last Post it is fully exampled there and shows what I feel Tesla would like to have done but couldn't due to technical shortages of his equipment or he may have just gone the wrong way.*
> *<snip<*
> *Smokey*

That's why I thought you chose 1.557 Mhz. I can get lucky sometimes Good to see you here Smokey! I've been on Bruce Perreault's list for something like 15 years and have seen you there as well as the OU forum. Always enjoy your posts. While I've built crystal radio's long ago I'm not really up to speed on this thread. Would you say your concept is anything like a scalar energy receiver?
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

There is no important work, there are only a series of moments to demonstrate your mastery and impeccability. Quote from Almine

Quote

#110
04-30-2012, 05:23 PM

David G Dawson
Senior Member

Join Date: Mar 2008
Posts: 380

Extra Coil detail

Quote:

> Originally Posted by **xee2**
> *@ David G Dawson*
>
> *Have you built an antenna using the same EPD design that Jake is using? If so how well does it work?*

Yes, and it doesn't work.
My old design as per my instructions worked first off without any problems and was picking up the signal from the 1N34 probe head into high impedance Headphones from 8 to 10 " away from

the Extra and this was on a piece of metal used as a neutralising capacitor in an old RF transmitter without any specific direction of alignment.
It was obvious that the coil was radiating as well as magnifying the signal and was of a Telluric nature.
It is very clear in Tesla's Colorado Springs Notes that the length of the Secondary be equal to the Extra and this is NOT the case in Eric/Jake's example but is in mine.
I will put up some pics of both so you can see what I am talking about.

Just getting used to the procedures here at Energetic and always thought it was a neat means of display and communication.

Busy day here building the 3rd Coil but will come back later and put my formula into a step by step instruction so that it can be easily followed.
I am/was most familiar with spreadsheets but will leave that to the keen ones like Jake if he would like to build a new sheet for this working proposal at some later date after I have the 3rd coil working.

The final Colorado Spring's organisation only had 17 turns on the Secondary but I am using 20 as instructed but have a tap at 17 for testing purposes.
I, like Eric, spent many hours digesting the CSNotes and have come away with a different theory than Eric which works.
Eric leaves me for dead with the engineering data side but I have intuition on my side which has worked very well in the past.
Thinking about my working theory and I need to progress on that from an engineering viewpoint.

I have spelt it all out in the Posts but may be difficult to follow.
No, Phi is not involved here so far as no connection has been found.

On the new coil I will allow for fine tuning for a peak as the last one was just as built and could probably have been improved considerably and will fit this stage into the final Math.
Currently reading Eric's 'The Tesla Transformer' and can see some new holes there that need filling.
Later on this.

Smokey
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#111
04-30-2012, 07:38 PM

xee2 Join Date: Jan 2010

Silver Member Posts: 580

Quote:

> Originally Posted by **David G Dawson**
> *Just getting used to the procedures here at Energetic and always thought it was a neat means of display and communication.*

Thanks for the reply. For your information, you can post images here but the storage space is very small so it will fill up quick. This is probably best to start with. But you will need to use ImageShack® - Online Photo and Video Hosting and paste the link into your post here when your space is filled up. They now require setting up an account. Setting image size to 640x480 or 800x600 will fit on screen best. There is a preview button that allows you to check how your post will look before actually submitting the post.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

____

#112
04-30-2012, 08:10 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Whats going to happen next?

I got my hands on a piece of copper flashing 59" x 7". That is about as good as it gets for me. Right size, Right volume, or close enough to start. This will be my primary capacitor and connecting lengths. My thought is a simple two plate capacitor.

cut it like this.. and the thinner strips solder to the primary.

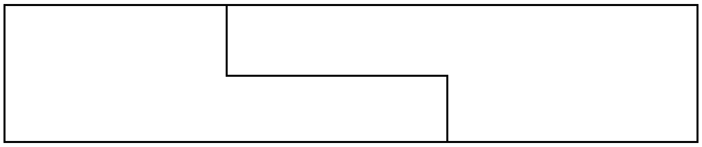

By jastermake at 2012-04-30

BUT, before I cut...

What will the 2 turn primary and capacitor do to the resonant frequency of the secondary when I excite them both? I'll be using the same 1 turn strip and function generator I used to test the secondary?

What else should I consider when designing this capacitor? Please keep in mind this capacitor may play more than one role in this setup. I believe Eric called this capacitor the 'energy transformer' in one video.

So.... what will it do?

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Last edited by jake; 04-30-2012 at 08:42 PM. Reason: imageshack its easy!!

Quote

#**113**

04-30-2012, 08:29 PM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

Quote:

> Originally Posted by **David G Dawson**
> *Yes, and it doesn't work.*
> *My old design as per my instructions worked first off without any problems and was picking up the signal from the 1N34 probe head into high impedance Headphones from 8 to 10 " away from the Extra and this was on a piece of metal used as a neutralising capacitor in an old RF transmitter without any specific direction of alignment.*
> *It was obvious that the coil was radiating as well as magnifying the signal and was of a Telluric nature.*
> *It is very clear in Tesla's Colorado Springs Notes that the length of the Secondary be equal to the Extra and this is NOT the case in Eric/Jake's example but is in mine.*
> *I will put up some pics of both so you can see what I am talking about.*
>
> *Just getting used to the procedures here at Energetic and always thought it was a neat means of display and communication.*
>
> *Busy day here building the 3rd Coil but will come back later and put my formula into a step by*

> *step instruction so that it can be easily followed.*
> *I am/was most familiar with spreadsheets but will leave that to the keen ones like Jake if he would like to build a new sheet for this working proposal at some later date after I have the 3rd coil working.*
>
> *The final Colorado Spring's organisation only had 17 turns on the Secondary but I am using 20 as instructed but have a tap at 17 for testing purposes.*
> *I, like Eric, spent many hours digesting the CSNotes and have come away with a different theory than Eric which works.*
> *Eric leaves me for dead with the engineering data side but I have intuition on my side which has worked very well in the past.*
> *Thinking about my working theory and I need to progress on that from an engineering viewpoint.*
>
> *I have spelt it all out in the Posts but may be difficult to follow.*
> *No, Phi is not involved here so far as no connection has been found.*
>
> *On the new coil I will allow for fine tuning for a peak as the last one was just as built and could probably have been improved considerably and will fit this stage into the final Math.*
> *Currently reading Eric's 'The Tesla Transformer' and can see some new holes there that need filling.*
> *Later on this.*
>
> *Smokey*

Jan 2, 1900:

New Extra coil, 98 turns last modification.

25' 11" turn length = 98.99" Dia coil.

Still a 1:1 coil and there is notation of various 'tap' points on the primary and secondary to bring them into resonance.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#114

04-30-2012, 08:34 PM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

a note also in the CSN of Tesla on Picture V. Extra coil connected to the secondary at 3/4 turns from ground connection. any idea on what 3/4 turns means? 3/4 of the turn or 75% of the total number of turns?
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#115
04-30-2012, 09:12 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *What will the 2 turn primary and capacitor do to the resonant frequency of the secondary when I excite them both?*

When close together the inductance of each coil will be increased by the other coil. The coil with the highest Q will primarily effect self resonant frequency. You should set SRF of secondary then gradually move primary closer re-tuning the SRF of each coil as you move them together.

This is how I measure self resonant frequency.

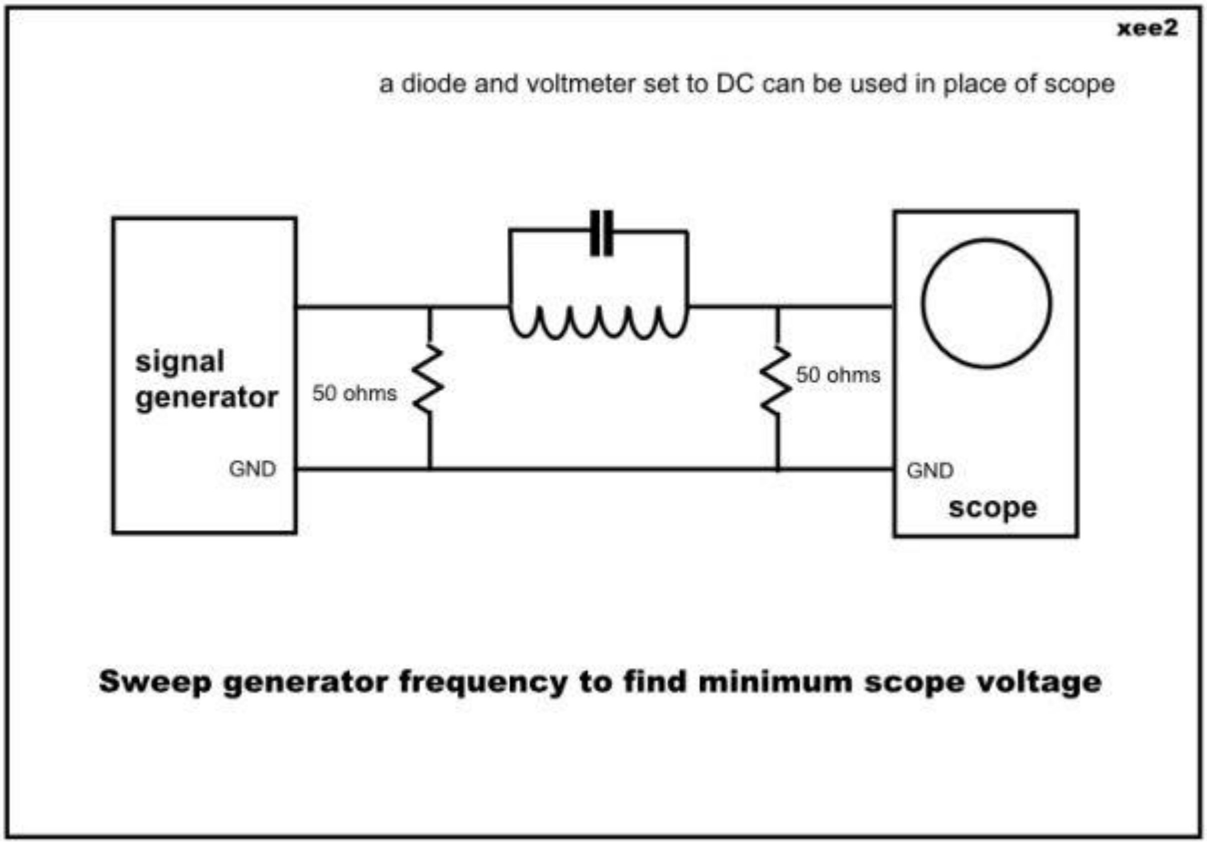
xee2
a diode and voltmeter set to DC can be used in place of scope
signal generator
GND
50 ohms
50 ohms
GND
scope
Sweep generator frequency to find minimum scope voltage

resonant frequency test

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#116

05-01-2012, 01:54 AM

David G Dawson
Senior Member

Join Date: Mar 2008
Posts: 380

Extra Coil Math

Goint to be away for a week so will make an itemised listing of the Math now.

Crystal Set Iniative Extra Coil:
In brackets are the figures for my situation.

1. Write down the frequency of the local AM Radio Station you are building the Crystal Set to

receive from. (1.557 Mhz)
2. Multiply this number by pi/2 - this is the 1/4 Wave frequency (2.4457 Mhz)
3. Multiply this number by 4 (9.7829 Mhz) - this is the Full Wave frequency
4. Use your Frequency & Wavelength calculator to find the total length of wire required - this is what I use:

FREQUENCY & WAVELENGTH CALCULATOR

My length of wire = 30.645 meters or 30645mm
I use the same size awg (14) as for the Secondary which means the Copper masses are equal for all of Primary/Secondary and Extra Coils.
Length of Secondary = length of Extra.

5. Calculate Coil diameter with We = 0.4 Ws (195.2mm)
6. Calculate 1 turn using 2pi r = (613.24mm)
7. Divide this number into your total length (30645/613.24 = 49.97) = number of Turns
Suggest at this point that you wind a few extra for tuning purposes and then with Headphones on, tune for a maximum by probing back along the coil.
8. Use another tin can about 6 to 12" away from your Extra Coil to hang your 1N34 probe onto, doesn't need to be oriented in any way.
Those of you with excessive local transmitters may be blown off the map so be warned.
Mine is only a 2 Kw.

If at step 2 you were to DIVIDE that number you would arrive at the point where both Eric/Jake's Spreadsheet & Tesla would be working at.
This is in my view the wrong point as it brings you back into a 1/8th Wave and not just before the 1/4 Wave where the 0.637 'average AC Wave' becomes a significant factor.
This is the exact point where your local AM Radio Station sits at.

That's it.
Will work on the theory while I am away as I need to know the electrics behind all of this.
David,
Would be appreciated if you could get this Post to Eric as he may be able to see what is happening here.
This one he can simply copy down the instructions on paper from the phone.
First time I have ever seen a pi and average AC wave association.
Bottom line is that this arrangement works.
Goint to enjoy this holiday, first in about 5 years.
May call in while I am away.
Thanks.

Smokey
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#117
05-01-2012, 03:27 AM

David G Dawson
Senior Member

Join Date: Mar 2008
Posts: 380

Why Hard to Tune?

Have now drawn both the Multiply and Divide situations out on paper and you can see why the Eric/Tesla situation will make it difficult to tune.
In the MULTIPLY way which is my way, the 0.637 'average AC wave' point is on the UPCURVE of the frequency waveshape slope.

In the DIVIDE way that same point is on the DOWNCURVE slope.
So what we are seeing is an opportunity to tune in the first instance and only a brief and parting opportunity in the second instance.
That is sufficient evidence for me as mine required NO tuning where the other is said 'to be difficult to tune' and there is your answer as to why above.

OK have attached the drawing and hope it comes out viewable.
Math also shows the difference in length involved and how it is calculated.
Have always had this notion that all of this should not be difficult if we are to work at the correct operating points.
Thanks.

Smokey
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Last edited by David G Dawson; 09-18-2012 at 01:51 AM.

#118
05-01-2012, 07:18 AM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **T-rex**
> *5) PI over two; this factor appears twice in the Crystal Radio Initiative, C.R.I. The (20%) secondary coil "effective" propogation velocity is the velocity of light in the space in which it occupies. Here the Pi over two factoris an offsetof a unit value to shorten the winding to compensate for the heavy external loading that the secondary is loaded with, that is, the external inductance and capacitance from the earth and extra coil connections. The aspect ratio of 1 to 1 on the extra coil gives an "effective" propogation of 187% the velocity of light, hence the winding length must be made greater in order to get a quarter wave resonance at a higher speed. This is brought down to Pi over 2 percent, 157%, the velocity of light to compensate for the burden upon the extra coil, this mostly arising from the dielectric upon which this coil is wound. The secondary is Pi over two shorter, the extra coil is Pi over two longer. Golly Mr. Wizard that's Pi squared over four! Do not hunt for magic where it is not, this is all experimental. So go experiment.*

This is all I know about the extra coil according to Eric.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#119

05-01-2012, 07:44 AM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

Quote:

> Originally Posted by **David G Dawson**
> *Have now drawn both the Multiply and Divide situations out on paper and you can see why the Eric/Tesla situation will make it difficult to tune.*
> *In the MULTIPLY way which is my way, the 0.637 'average AC wave' point is on the UPCURVE of the frequency waveshape slope.*
>
> *In the DIVIDE way that same point is on the DOWNCURVE slope.*
> *So what we are seeing is an opportunity to tune in the first instance and only a brief and parting opportunity in the second instance.*
> *That is sufficient evidence for me as mine required NO tuning where the other is said 'to be difficult to tune' and there is your answer as to why above.*
>
> *OK have attached the drawing and hope it comes out viewable.*
> *Math also shows the difference in length involved and how it is calculated.*
> *Have always had this notion that all of this should not be difficult if we are to work at the*

> *correct operating points.*
> *Thanks.*
>
> *Smokey*

the 0.637 is not the AC average but the rms of 1/2 of the sine wave of a DC rectified current. AC average or rms of a full wave is 0.707 or sine45*, cos45*.

If the extra coil is designed around C and harmonics at velocities of C it will only continue to receive hertzian waves. The extra coil dimensions that Eric supplied should be useless in reception of hertz transmission in air.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

#120
05-01-2012, 11:04 AM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

1/4 and 3/4 according to EPD

Quote:

> Originally Posted by **T-rex**
> *@All struggling with 1/4 and 3/4*
>
> *There is no inductance, capacitance, or wavelength. The wavlenght I gave of c/omega is not the standard wavelength. The standard way is c/f, then take 1/4th of that. That's the electromagnetic wavelength. The coils are not electromagnetic devices. Wire length is calculated by c/omega. A factor of 2pi is involved, so it's c/2*pi*f. There is no quarter. This gives a wirelength that is now 2 over pi shorter. There is no inductance and capacity to compensate or think about, the coil is both. They are not separated. Read the article in 7 parts just put in, paying special attention to the section labeled disruptive discharges. In the Colorado Springs all you want to look for is the final dimensions of the Magnifying Transmitter. Other than some circuit diagrams and final dimensions there's not much there of any use for coil design. Tesla was figuring stuff out as he was going along, he was struggling to figure out what was going on, he had nothing to work with. So read the Colorado Springs notes and share in his confusion.*

I miss the old thread and all the goodies hidden in it.

05-01-2012, 05:11 PM

David G Dawson
Senior Member

Join Date: Mar 2008
Posts: 380

Wish to Correct

Quote:

> Originally Posted by **jake**
> *I miss the old thread and all the goodies hidden in it.*

MadHatter,
You are incorrect and please read the URL I posted as it very clearly states that the 0.637 point is AVERAGE VALUE (Vav) and 0.707 is RMS VALUE (Vrms).
I have been in this game a long time and know what is what.
Please check your facts before responding to anything I have presented.
I don't mind criticism but this is BS!

Jake,
All very well but look at this:
1.557 x 2pi = 9.7829 Mhz and then go look at my drawing as it is the very same 9.7829 Mhz.
1.557 x (pi/2) x4 = 9.7829 Mhz.
All the waveshape drawing that I presented represents is the point at which we are working at with our AM Radio Station signal with respect to a Hertzian 1/4 Wave system.

If you people are not going to listen to an oldtimer with a great deal of experience in OU research, then I will just disappear into the bush just like Eric has done.

One thing I should make note of is the direction of winding.
My Primary and Secondary are both CCW but for some unknown reason I stood on the wrong side whilst winding my first WORKING Extra and wound it CW - opposite.
The new Extra is also wound CCW but does not work or is difficult to get to work and my drawing shows why.

Eric states clearly to wind ALL Coils in the same direction but I may have accidentally stumbled across another discovery which I will prove out on my return.

Not going to sit here working out all the Math problems and for all of you to not understand and discredit.
If you are so blind that you cannot see what I am presenting then I will simply fade into the distance, again like Eric.

What you may not realise is that what is being presented before your very eyes is the entire secret to winding coils like Don Smith and Kapanadze use but nobody knows why they work and all their efforts were/are of an empirical nature.

What Kapanadze is currently using is a split Copper main core and this is his Primary as he is attempting to build as Eric has shown with all the Math that I am fully supporting as shown above and replicates a Tesla Magnifying Transformer (TMT).

What the Kapanadze device now needs is the Secondary and Extra Coils to be wound using the data that I have supplied and we should have our OU device.
The only outstanding question is the direction to wind the Extra Coil and I will clear that up when I return.

I am sorry to be working with people that cannot see the light when it has all been presented before you and very clearly at that with full reference to Eric's theory and in agreement.

Didn't you hear that my Extra Coil WORKED first time using Eric's Math but on the MULTIPLY side and NOT the DIVIDE side - what do you need to see the LIGHT!?

If I am beginning to sound like Eric then you may wake up and realise why, some day.

The information passed on from this 'discovery' even shows clearly that Tesla may have made a mistake with his operating fequency and he chose the wrong side of his 96 Khz base frequency and used the 60 Khz side instead of the 149 Khz side.
This may be why he had tuning problems as his work was of an empirical nature without a solid working Math base which we now have thanks to Eric and some accidental discoveries.
Tesla may also have had technical equipment problems at the higher frequency and why he chose the lower.

WAKE UP!

Smokey
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

Download SOLAR SECRETS by Peter Lindemann
Free - Get it now: **Solar Secrets**

#122
05-01-2012, 05:47 PM

ewizard
Gold Member

Join Date: Jan 2010
Location: Outside the Matrix
Posts: 1,015

Thanks Smokey! That's a biggy and I believe I get it now. Please don't take any offense at those who may not yet get it. There are lots of different levels people are at here. BTW you now have your own Folder on my hard drive

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

There is no important work, there are only a series of moments to demonstrate your mastery and impeccability. Quote from Almine

Quote

#123

05-01-2012, 05:55 PM

ewizard
Gold Member

Join Date: Jan 2010
Location: Outside the Matrix
Posts: 1,015

Thanks also for the info on the split copper. Zilano mentioned that also recently and about 2 years ago I came across a video that demonstrated something quite impossible looking with a split copper tube device. I was trying to get more info at the time but some said it was nothing special. Looks like maybe it is after all. I really appreciate your showing the calcs and the coil info too. I should know the calcs being a Ham but I've been out of touch so long with that it's good to see it here again as it should be easy to extrapolate any needed calcs for other frequencies or sizes from what you have shown.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

There is no important work, there are only a series of moments to demonstrate your mastery and impeccability. Quote from Almine

Quote

#124

05-01-2012, 05:56 PM

Farmhand
Platinum Member

Join Date: Jun 2010
Location: Australia
Posts: 3,364

In the Colorado Springs notes Tesla does mention winding half or part of the secondary the opposite way or another coil or something, I'll try to find the reference . ahah page 69 figure 2. A few of the top turns of the secondary

Cheers

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

____

First Imagine it. Then Build it.

#125

05-01-2012, 06:29 PM

WARNING
BURN HAZARD

jake

Senior Member

Join Date: Nov 2007
Posts: 169

Outch!!

Quote:

> Originally Posted by **David G Dawson**
> *MadHatter,*
> *You are incorrect and please read the URL I posted as it very clearly states that the 0.637 point is AVERAGE VALUE (Vav) and 0.707 is RMS VALUE (Vrms).*
> *I have been in this game a long time and know what is what.*
> *Please check your facts before responding to anything I have presented.*
> *I don't mind criticism but this is BS!*
>
> *Jake,*
> *All very well but look at this:*
> *1.557 x 2pi = 9.7829 Mhz and then go look at my drawing as it is the very same 9.7829 Mhz.*
> *1.557 x (pi/2) x4 = 9.7829 Mhz.*
> *All the waveshape drawing that I presented represents is the point at which we are working at with our AM Radio Station signal with respect to a Hertzian 1/4 Wave system.*
>
> *If you people are not going to listen to an oldtimer with a great deal of experience in OU research, then I will just disappear into the bush just like Eric has done.*
>
> *One thing I should make note of is the direction of winding.*
> *My Primary and Secondary are both CCW but for some unknown reason I stood on the wrong side whilst winding my first WORKING Extra and wound it CW - opposite.*
> *The new Extra is also wound CCW but does not work or is difficult to get to work and my drawing shows why.*
>
> *Eric states clearly to wind ALL Coils in the same direction but I may have accidentally*

> *stumbled across another discovery which I will prove out on my return.*
>
> *Not going to sit here working out all the Math problems and for all of you to not understand and discredit.*
> *If you are so blind that you cannot see what I am presenting then I will simply fade into the distance, again like Eric.*
>
> *What you may not realise is that what is being presented before your very eyes is the entire secret to winding coils like Don Smith and Kapanadze use but nobody knows why they work and all their efforts were/are of an empirical nature.*
>
> *What Kapanadze is currently using is a split Copper main core and this is his Primary as he is attempting to build as Eric has shown with all the Math that I am fully supporting as shown above and replicates a Tesla Magnifying Transformer (TMT).*
>
> *What the Kapanadze device now needs is the Secondary and Extra Coils to be wound using the data that I have supplied and we should have our OU device.*
> *The only outstanding question is the direction to wind the Extra Coil and I will clear that up when I return.*
>
> *I am sorry to be working with people that cannot see the light when it has all been presented before you and very clearly at that with full reference to Eric's theory and in agreement.*
>
> *Didn't you hear that my Extra Coil WORKED first time using Eric's Math but on the MULTIPLY side and NOT the DIVIDE side - what do you need to see the LIGHT!?*
>
> *If I am beginning to sound like Eric then you may wake up and realise why, some day.*
>
> *The information passed on from this 'discovery' even shows clearly that Tesla may have made a mistake with his operating fequency and he chose the wrong side of his 96 Khz base frequency and used the 60 Khz side instead of the 149 Khz side.*
> *This may be why he had tuning problems as his work was of an empirical nature without a solid working Math base which we now have thanks to Eric and some accidental discoveries.*
> *Tesla may also have had technical equipment problems at the higher frequency and why he chose the lower.*
>
> *WAKE UP!*
>
> *Smokey*

Smokey,

Math does not mean much to me. Sure it's a tool that simplifies life but to me.. It's just numbers, there all relative.. I need to understand... Once I do that the math is a piece of cake...

I see the Dollard quote as being directly related you your situtation. You did it... Now don't give up on us. Because we are trying to catch up.. It takes us more then two days to build something new. We are trying. I don't have 20 years coil winding experience. And we are in the middle putting together the "whats going on" or at least I am

Now buck up, get back to work, and tell us what happens with your next extra coil.... I'm taking notes, trust me...

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#126
05-01-2012, 07:28 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

@ David G Dawson

It is great to have someone who understands what EPD was doing. Can you explain the purpose of the "primary" copper sheet coil. As far as I can tell it is just adding capacitance to the secondary coil by way of mutual coupling.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#127
05-01-2012, 07:49 PM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

Quote:

> Originally Posted by **David G Dawson**
> *MadHatter,*
> *You are incorrect and please read the URL I posted as it very clearly states that the 0.637 point is AVERAGE VALUE (Vav) and 0.707 is RMS VALUE (Vrms).*
> *I have been in this game a long time and know what is what.*
> *Please check your facts before responding to anything I have presented.*

> *I don't mind criticism but this is BS!*

Yes, it is the average value, my mistake in calling it rms. It is the average of half the sine wave and is normally applied to a rectified waveform.

Look forward to your results, and further development of the equations your using.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#**128**

05-01-2012, 08:02 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **xee2**
> *When close together the inductance of each coil will be increased by the other coil. The coil with the highest Q will primarily effect self resonant frequency. You should set SRF of secondary then gradually move primary closer re-tuning the SRF of each coil as you move them together.*
>
> *This is how I measure self resonant frequency.*

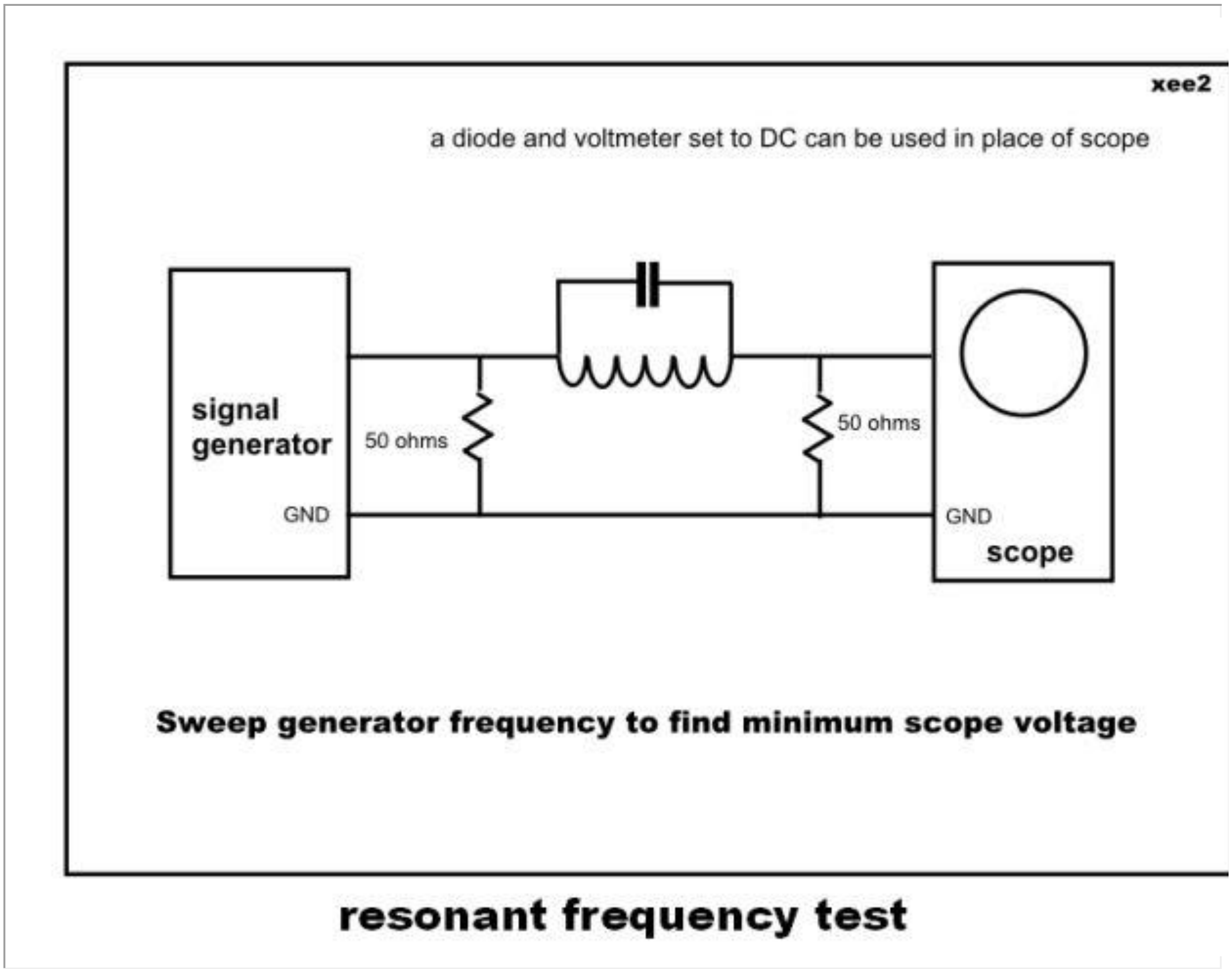
xee2
a diode and voltmeter set to DC can be used in place of scope
signal generator
GND
50 ohms
50 ohms
GND
scope
Sweep generator frequency to find minimum scope voltage
resonant frequency test

6 - 330pF across the primary and the system is happiest at 1640. It is now louder with only a ground(house ground) than it was with the fireplace antenna and I can listen to the piezo earpiece well without even touching the diode to the secondary. It will also light an led.

So now I need that 2 plate capacitor I was talking about earlier. That way the capacitor metal will match the primary metal.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#129

05-01-2012, 08:06 PM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

David, you're not going to like this and I mean no insult to you at all. Your calcs are still based on the velocity of C, Eric is clear and adamant that this is wrong! any and all variations of the harmonics based on the propagation speed of light will result in the wrong direction being taken.

If I have somehow missed where you've taken this into account please point it out to me

I agree with Eric on the increased velocity propagation to what extent is where it needs fine tuning. the rotation of the B field is where the velocity start to increase and is a cos value, how far that rotation is, well... that's a good question.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

#130
05-01-2012, 08:19 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> :
> *6 - 330pF across the primary and the system is happiest at 1640. It is now louder with only a ground(house ground) than it was with the fireplace antenna and I can listen to the piezo earpiece well without even touching the diode to the secondary. It will also light an led.*

Fantastic.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

#131
05-01-2012, 08:57 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **T-rex**
>
> *5) PI over two; this factor appears twice in the Crystal Radio Initiative, C.R.I. The (20%) secondary coil "effective" propogation velocity is the velocity of light in the space in which it occupies. Here the Pi over two factoris an offsetof a unit value to shorten the winding to compensate for the heavy external loading that the secondary is loaded with, that is, the external inductance and capacitance from the earth and extra coil connections. The aspect ratio of 1 to 1 on the extra coil gives an "effective" propogation of 187% the velocity of light, hence the winding length must be made greater in order to get a quarter wave resonance at a higher speed. This is brought down to Pi over 2 percent, 157%, the velocity of light to compensate for the burden upon the extra coil, this mostly arising from the dielectric upon which this coil is wound.* ***The secondary is Pi over two shorter, the extra coil is Pi over two longer.*** *Golly Mr. Wizard that's Pi squared over four! Do not hunt for magic where it is not, this is all experimental. So go experiment.*
>
> .

If the secondary is Pi over two shorter, the extra coil is Pi over two longer what is the actual length in rad.

It been a long time since I worked in radians. Is this correct:

pi/2 shorter = ((3*pi)/2)*'?'
pi/2 longer = ((5*pi)/2)*'?'

?= cycle or wave or...

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#132

05-01-2012, 10:03 PM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

Quote:

> Originally Posted by **jake**
> *If the secondary is Pi over two shorter, the extra coil is Pi over two longer what is the actual length in rad.*
>
> *It been a long time since I worked in radians. Is this correct:*
>
> *pi/2 shorter = ((3*pi)/2)*'?'*
> *pi/2 longer = ((5*pi)/2)*'?'*
>
> *?= cycle or wave or...*

Pi/2 = 1.5707
1.57... rads is 90*

Pi = 180*
1 rad = 180*/pi = 57.29577*

90* is the multiplication of -j rotation. reference to Steinmetz and the operator, also it's the rotation of the B field and hence the reason for the increased velocity.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#133
05-02-2012, 09:55 AM

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **madhatter**
> *Pi/2 = 1.5707*
> *1.57... rads is 90**

> *Pi = 180**
> *1 rad = 180*/pi = 57.29577**
>
> *90* is the multiplication of -j rotation. reference to Steinmetz and the operator, also it's the rotation of the B field and hence the reason for the increased velocity.*

The secondary length is given by C/omega or C/(2piF). As stated by Eric this is pi/2 shorter.. Shorter than what(w)? [pi/2 = 90* = 1/4 cycle]

So pi/2 less than one cycle is 3pi/2 or 3/4 of a cycle.

So what is the standard that we are taking away from? In terms of pi.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#134
05-02-2012, 11:21 AM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

Quote:

> Originally Posted by **jake**
> *The secondary length is given by C/omega or C/(2piF). As stated by Eric this is pi/2 shorter.. Shorter than what(w)? [pi/2 = 90* = 1/4 cycle]*
>
> *So pi/2 less than one cycle is 3pi/2 or 3/4 of a cycle.*
>
> *So what is the standard that we are taking away from? In terms of pi.*

The secondary is in radians not degrees, so the wave length is equal to it's arc length.

Erics comment of the pi over two longer or shorter is by my understanding the reference to -j operator which is a pi/2 multiplication rotation of the B field.

I'm working on some other relations for the extra coil at the moment using hyperbolic trig and will post as soon as I have something.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on

pinterest_share Share on linkedin

Quote

#135

05-02-2012, 11:34 AM

xee2

Silver Member

Join Date: Jan 2010

Posts: 580

@ jake

Is this what your circuit looks like?

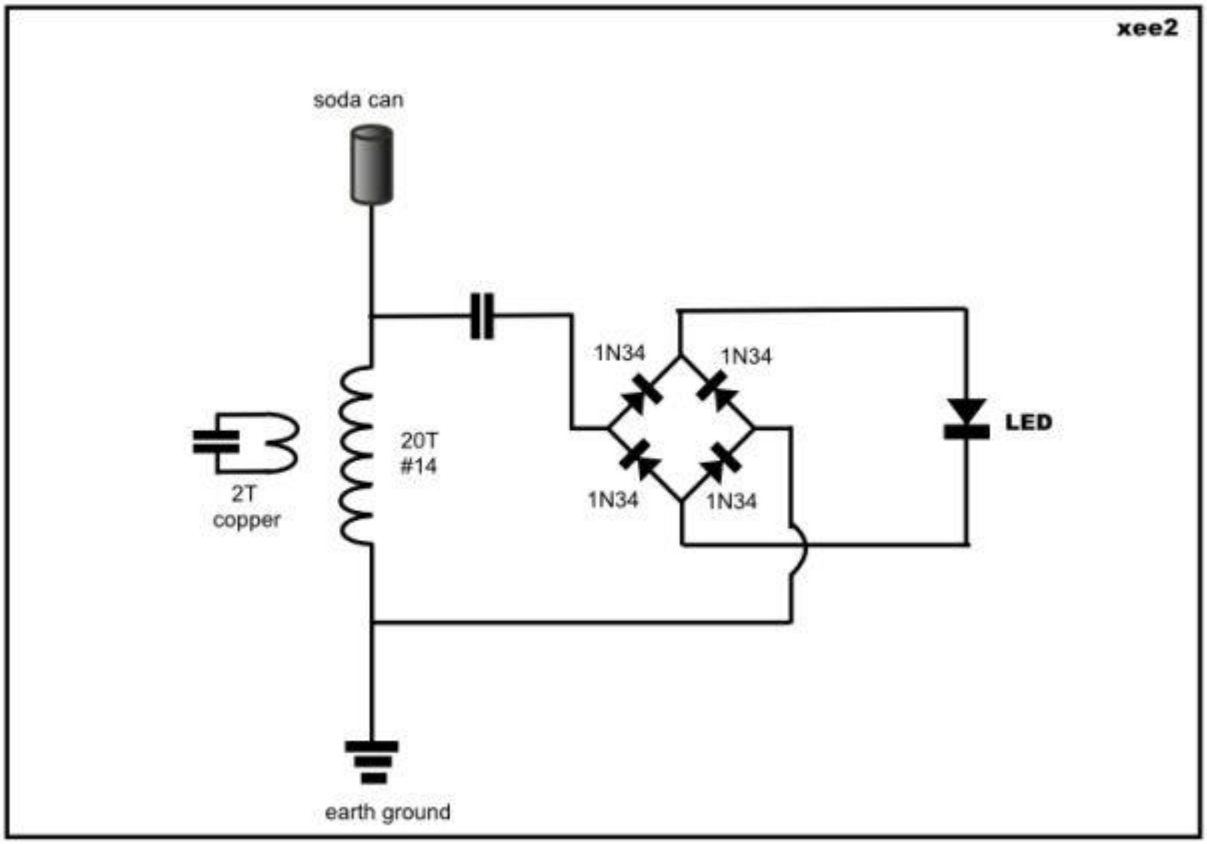
xee2
soda can
1N34
1N34
LED
20T
#14
2T
copper
1N34
1N34
earth ground

Jake's crystal radio

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#136

05-02-2012, 11:46 AM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

I wanted to post a comment.
I realize that we can twist and fiddle fart all we want with the mathematics to get results we want. the math in reality should be the theory and the experimental lab work is testing that theory, when it doesn't match or the results are different its back to the drawing board to re-do or re-calculate the math again.

Math is a VERY powerful tool in setting up a theory and testing it. If there is a pattern where experimental results fit the mathematics used then expanding on the mathematical theory gets a bit easier however one still has to be careful that we don't start to effect the results by manipulating the results thru the mathematics used.

Thankfully a lot of the base math is known and results are predictable, we are though when getting to the extra coil moving into the unknown. coil winding and manipulation of fields thru various winding designs is nothing new, this is not what Eric or Tesla was doing though. instead of trying to force or manipulate the fields Tesla was using the base parameters of the fields to manipulate them. Steinmetz covers this with the J operator in hyperbolic form, Clifford algebra is another useful tool though it's basis was quaternions - another story for another day!

The goal is to get to L&C to match in harmony, nondestructive resonance. once achieved there should be a rotation of the B field to where the conjugate pair is no more, this will give rise to velocity of the wave faster than C. simple in theory but highly complicated in practice due to the sheer number of parameters involved and that the mathematics for engineering this are not finished.

Now here's the 64 trillion dollar question (inflation), when the B field is longitudinal to the E field, is the extra coil still inductively coupled to the secondary? or is it now in electrostatic phase coupling? it needs both to achieve resonance, but once there what happens? is it the elusive sink and radiator Steinmetz theorized?

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#137
05-02-2012, 11:52 AM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

Quote:

> Originally Posted by **xee2**
> *@ jake*
>
> *Is this what your circuit looks like?*

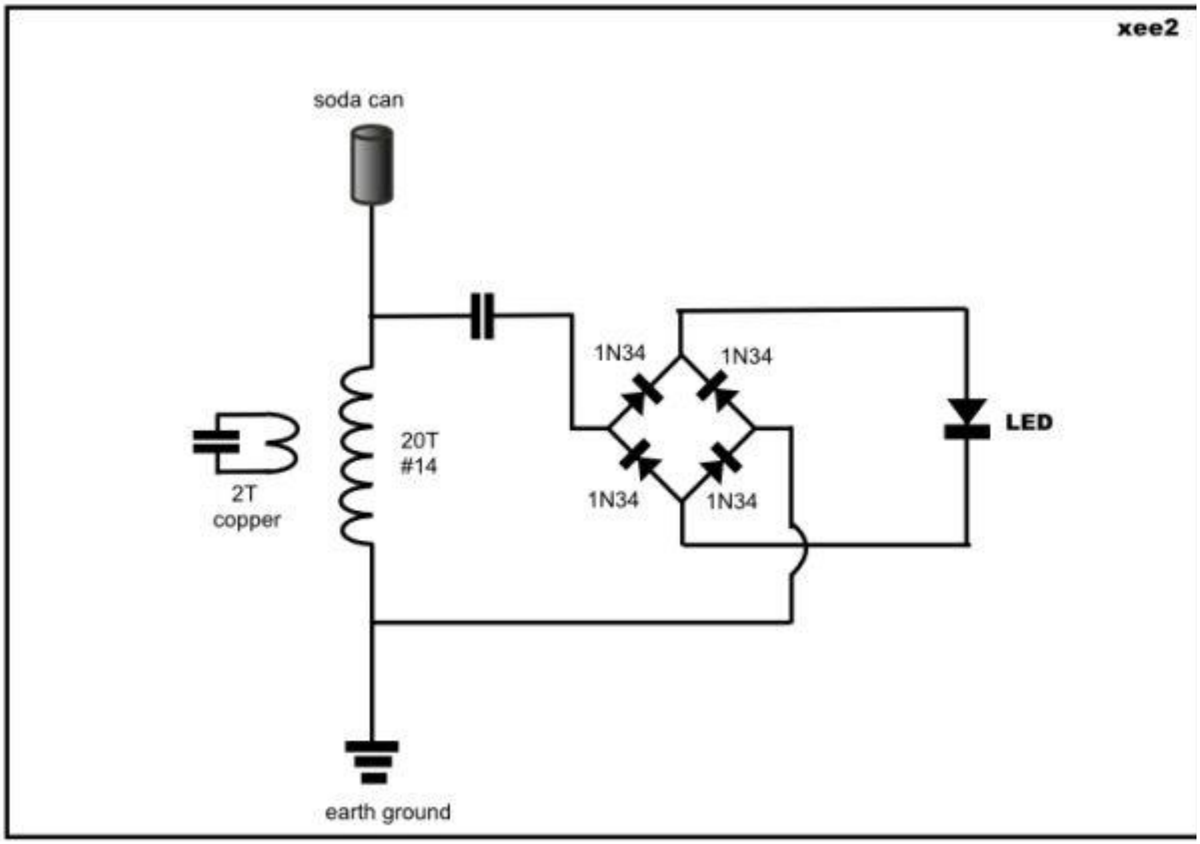

Jake's crystal radio

you don't need a full bridge rectifier or capacitor for the LED, or at least I don't. single diode and led with leads shunting top and ground will light the LED nicely.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on

pinterest_share Share on linkedin

#138
05-02-2012, 12:27 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

@ madhatter

Thanks for the info. How far are you away from the transmitter? I would try this but the nearest transmitter to me is about 50 miles away.

I know nothing about this antenna design, but as an ex-ham the top coil looks like an inductive matching coil that is used on ham whip antennas that are shorter than 1/4 wavelength. When the whip is less than 1/4 wavelength it looks like a large capacitive reactance. Putting an inductor in series with the capacitance cancels out the reactive components and makes the whip look like a small real impedance instead of a large reactive impedance. This makes it easier to get power into and out of the whip.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#139
05-02-2012, 01:06 PM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

Quote:

> Originally Posted by **xee2**
> *@ madhatter*
>
> *Thanks for the info. How far are you away from the transmitter? I would try this but the nearest transmitter to me is about 50 miles away.*
>
> *I know nothing about this antenna design, but as an ex-ham the top coil looks like an inductive matching coil that is used on ham whip antennas that are shorter than 1/4 wavelength. When*

> *the whip is less than 1/4 wavelength it looks like a large capacitive reactance. Putting an inductor in series with the capacitance cancels out the reactive components and makes the whip look like a small real impedance instead of a large reactive impedance. This makes it easier to get power into and out of the whip.*

Ahh, well the radio setup was too problomatic and that transmitter is roughly 36miles away.

I turned my attention to smaller scale and built two sets CW and CCW coil sets at 4Mhz.

This is a bit different from std radio reception, the secondary is in radian length. but the extra is very different. It's basic design is based on the propagation velocity increased by 157.08% over C. which increases the length to fit the increased wave length.

the secondary is 1.570 radians in length or 90*, the extra is also the same 90* but at the increased velocity of C by 157%. the spacing and wire ga are designed to create a self resonance within a 1:1 H/D ratio. As you can imagine nearbody capacitance becomes an issue as it's highly sensitive. The extra coil calcs seem to confuse many as well, there is a fair amount of information by Eric Dollard on the reasons and research behind the equations.

It starts from the point that the speed of light is due to the conjugate pair of E & B fields, now by rotating the B field by theta degrees it is possible to increase the velocity by cos(theta).
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#140
05-02-2012, 02:12 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **xee2**
> *@ jake*
>
> *Is this what your circuit looks like?*

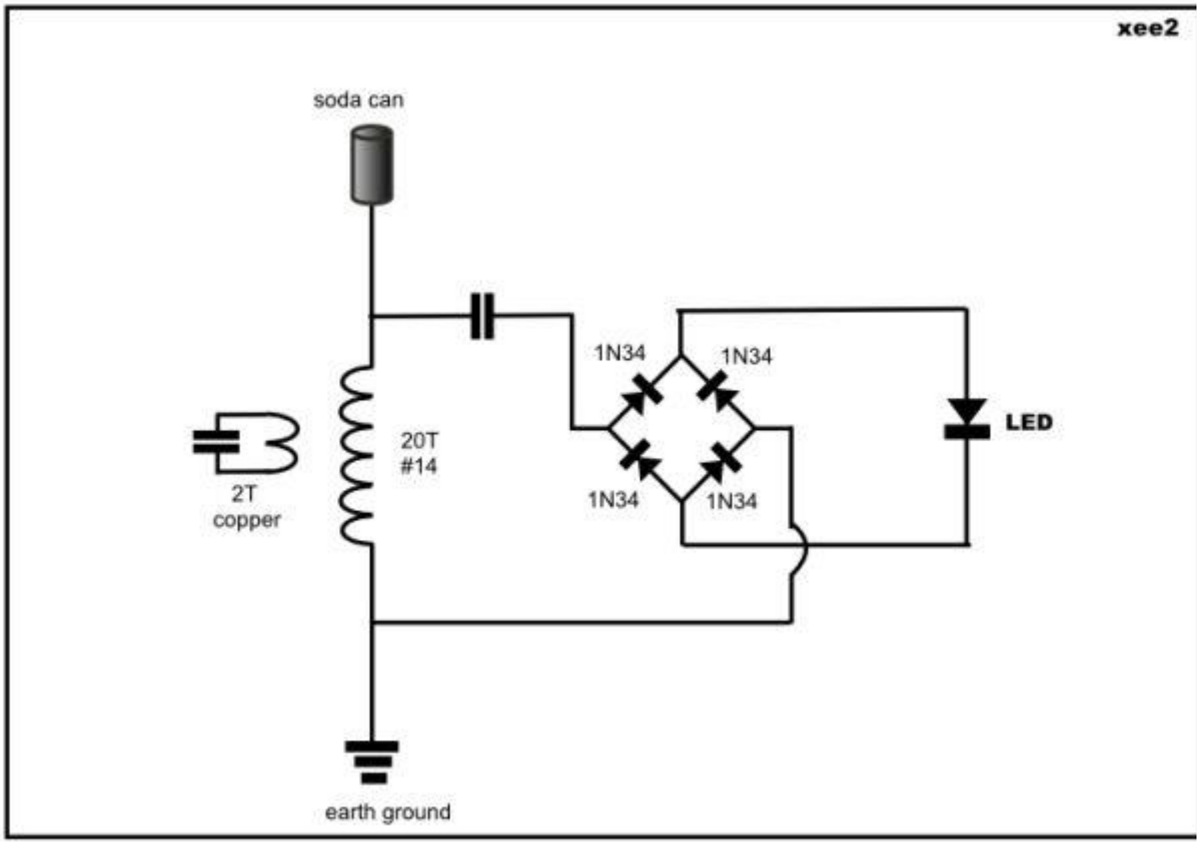

Jake's crystal radio

I started with something very similar to that schematic. But then the ground line from the FWBR came off and it still worked fine with little to no difference. It actually worked better if I held the ground line instead of attaching it to the bottom of the secondary. That reminded me of an AV plug so I tried one and it worked, so I stopped using the FWBR.

The secondary has nothing attached to it except the soda can terminal and the ground. The soda can terminal is optional and the size can be changed to for more or less capacitance. I don't know exactly Eric recommends the can.

The primary has almost 2nF of capacitance attached to it. (6 silver mica in parallel) This will be replaced with a 2 plate capacitor. The LED and Piezo earpiece are both attached to the same AV plug. The piezo is in parallel with the LED on the AV plug. This can be touched to almost any part of the secondary or primary for a listen. It will start to pick up audio about 1" away from the coil.

If there is not enough capacitance on the secondary (not across), hand and body capacitance plague the system. You need to be near it to for it to work. Once I had enough capacitance I could approach without much issue but if I get with 1" it shuts down.

I went with the soda can on the secondary which gave me an SRF of ~2000kHz for the secondary alone. Adding the primary brought that down. Adding capacitance to the primary keeps bringing it lower. But there is a cost in efficiency.

Adding capacitors across the secondary was difficult because I needed so little capacitance. One lead from the top and one from the bottom laid close together was about as good as I could get. Or I could just hold my hand about 1 inch from the coil.

What did you use to draw the schematic?? I need something like that.

The question remains is this all related to TEM?
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#141
05-02-2012, 02:25 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

@ jake

Thanks for the update.

Quote:

> Originally Posted by **jake**
> *What did you use to draw the schematic?? I need something like that.*

Download Serif DrawPlus

This is a free version.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#142
05-02-2012, 07:09 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Uploaded with ImageShack.us

I'll get better at the schematics.

Thats what it looks like today. Replacing the can with 1 turn 1/2" copper tube made very little change. All of which could be offset by tube hight or capacitance on the primary.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#143

05-03-2012, 08:22 PM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

I finally finished the extra coil for the other set. So now I have two sets, one CW and one CCW.

Setting them up in series so that the two primaries are back to back and axial the entire series will resonate at 4Mhz, that took some fiddling.

disconnecting the extra coils and leaving them inductively coupled, and the two primaries inductive as well. setting about to tune the extras has been strange, the CW extra tuned easily enough, the CCW extra, nope. I've removed 3 turns got it as far as 3.5Mhz adding capacitance lowers it and I don't feel comfortable removing more turns. something odd with this setup.

For an experiment I fed the two primaries the same signal from the generator, so that signal goes to outer turn of CW and inner turn of CCW, opposite for ground lead. the secondaries are in series with a small capacitor and the free end has 7sq/in of aluminum foil as capacitance on it.

the extras are sitting horizontal so that the axis is vertical, they are sitting 45* off plane of the primary secondary axis at roughly 1.5" from the secondaries edge. top of each extra has 2sq/in of foil. the top of the CW extra is an LED with two diodes. The led lights brightly at 4.1Mhz, placing my hand anywhere near the secondary kills the light, near the extra has no effect. taking the LED circuit and placing it on a wooden dowel and moving about the entire arrangement the

LED lights only when touched to the extra or secondary foil end. If I take the circuit and hold it via a plastic clip and move it about it will light when near the secondary free end or extra coils, my body capacitance seems to enable it to light wireless.

I'm only feeding it a 0dbm signal at .5vdc. CW extra coil measures ~3vdc wirelessly.

strange indeed. Oh and the magnetic field near the extra coils is lower then background field! very very weird.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#144

05-04-2012, 09:32 AM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **madhatter**
> *I finally finished the extra coil for the other set. So now I have two sets, one CW and one CCW.*
>
> *Setting them up in series so that the two primaries are back to back and axial the entire series will resonate at 4Mhz, that took some fiddling.*
>
> *disconnecting the extra coils and leaving them inductively coupled, and the two primaries inductive as well. setting about to tune the extras has been strange, the CW extra tuned easily enough, the CCW extra, nope. I've removed 3 turns got it as far as 3.5Mhz adding capacitance lowers it and I don't feel comfortable removing more turns. something odd with this setup.*
>
> *Is there anything different about the coil form?*
>
> *Is there any change in resonant frequency when you change the orientation?*
>
> *For an experiment I fed the two primaries the same signal from the generator, so that signal*

> *goes to outer turn of CW and inner turn of CCW, opposite for ground lead. the secondaries are in series with a small capacitor and the free end has 7sq/in of aluminum foil as capacitance on it.*
>
> *the extras are sitting horizontal so that the axis is vertical, they are sitting 45* off plane of the primary secondary axis at roughly 1.5" from the secondaries edge. top of each extra has 2sq/in of foil. the top of the CW extra is an LED with two diodes. The led lights brightly at 4.1Mhz, placing my hand anywhere near the secondary kills the light, near the extra has no effect. taking the LED circuit and placing it on a wooden dowel and moving about the entire arrangement the LED lights only when touched to the extra or secondary foil end. If I take the circuit and hold it via a plastic clip and move it about it will light when near the secondary free end or extra coils, my body capacitance seems to enable it to light wireless.*
>
> *I see this happen as well. I am working on in.*
>
> *I'm only feeding it a 0dbm signal at .5vdc. CW extra coil measures ~3vdc wirelessly.*
>
> *How are you feeding it?*
>
> *strange indeed. Oh and the magnetic field near the extra coils is lower then background field! very very weird.*

Thats why you have to type words here without meaning

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#145

05-07-2012, 04:04 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

@ jake

The following test would provide a good measure of how well your crystal radio is performing. Find highest output voltage you can get with a specified antenna length.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on

pinterest_share Share on linkedin

Last edited by xee2; 05-07-2012 at 04:07 PM.

#**146**

05-08-2012, 12:28 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **xee2**
> *@ jake*
>
> *The following test would provide a good measure of how well your crystal radio is performing. Find highest output voltage you can get with a specified antenna length.*

It will show 2-3v avg. or 5-6v peak to peak if I set it up like that. No antenna. The digital multimeter and DSO nano average values are close.

BUT it needs to be grounded to something in my house or attached to my house, or "virtual grounded" to a copper plate on the counter. The entire electrical system including the mains ground and water pipes all are contaminated with the signal. i.e. I can faintly hear the station with just a diode and no coils..

Outside with a wire stuck into the ground away from the house I just get a signal of ~50mv. Without the wire in the ground I get nada.. (I need to work on a good ground)

05-08-2012, 12:32 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Primary capacitor plate and lead

Its going to be dificult to solder this to the primary. I have two of these. The other is the mirror image.

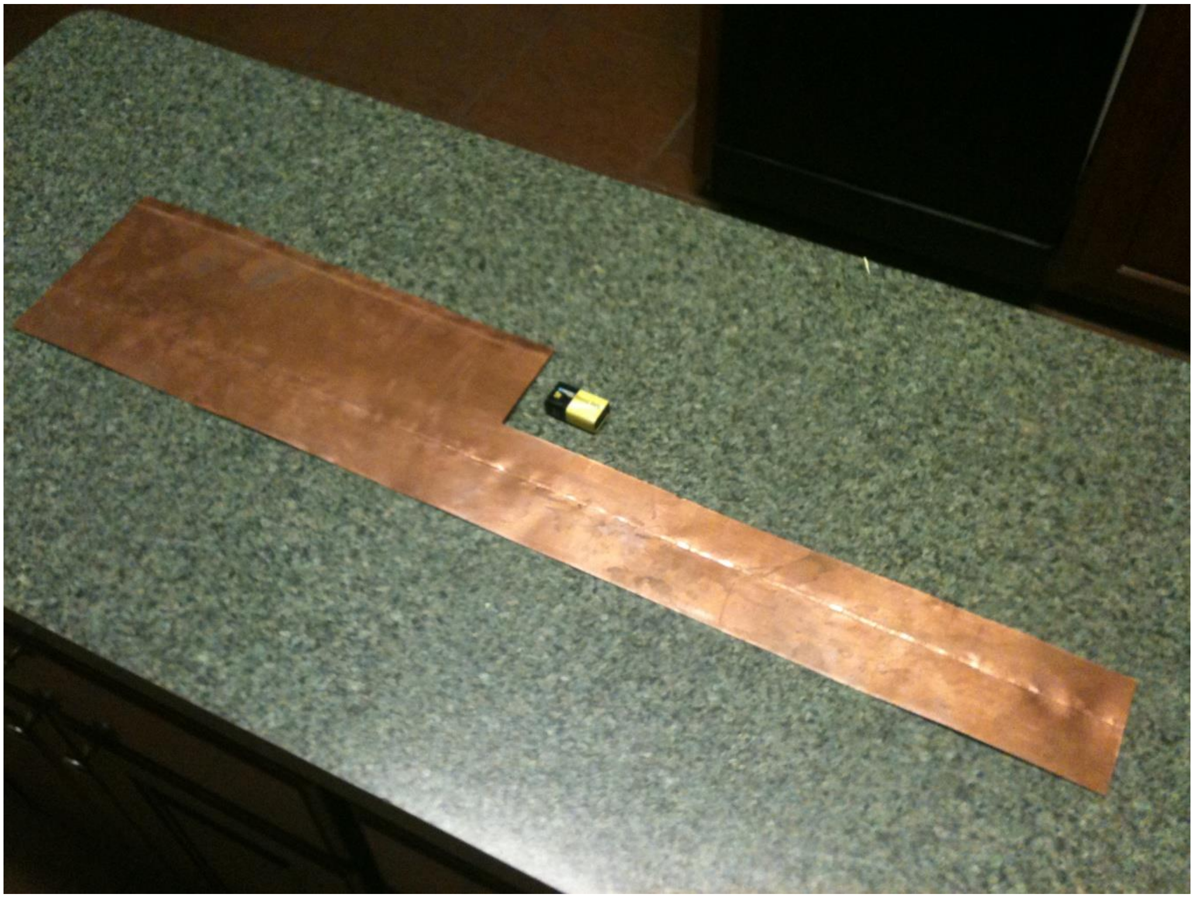

Uploaded with ImageShack.us
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#148
05-08-2012, 01:25 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *Its going to be dificult to solder this to the primary. I have two of these. The other is the*

> *mirror image.*

Soldering is best, but just fastening with screws should work.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#149
05-08-2012, 01:28 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *Outside with a wire stuck into the ground away from the house I just get a signal of ~50mv. Without the wire in the ground I get nada.. (I need to work on a good ground)*

Is this 50 mV with the resonator?
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#150
05-08-2012, 01:51 PM

dR-Green
Gold Member

Join Date: Apr 2008
Location: Merlin's Fort
Posts: 1,220

I think the secondary should be mostly left alone. It can be tuned for best signal by using adjustable condenser plates across it, and the primary also tuned using variable capacitors. One affects the other. This can all be seen on an oscilloscope across the primary. The plate connecting to the diodes "sucks" energy out of it and the voltage in the primary is lowered, also the plate itself affects the tuning so the condenser plates need to be adjusted to compensate. Using a simplified "power" setup with no load makes all this easily visible.

I also get a higher voltage in the primary at 1.4-2.something MHz when the station I'm tuning to is 882kHz, there's only noise at the higher frequencies/higher voltage, so I don't think just measuring the voltage is an accurate way of tuning it or any reliable way of determining what's

going on. The signal should be in the primary otherwise there'll be no possibility of powering anything. The radio setup isn't arranged across the primary as a regular Tesla type receiver, rather in this case the primary is just having a resonant effect to maximise the "radiated" signal at the top of the secondary for reception through the plate antenna. In my opinion 🙂So I'd say the place to be looking is mostly at the primary.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

____

"Knowledge is cosmic. It does not evolve or unfold in man. Man unfolds to an awareness of it. He gradually discovers it." - Walter Russell

"Once men died for Truth, but now Truth dies at the hands of men." - Manly P. Hall

05-08-2012, 07:11 PM

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **xee2**
> *Is this 50 mV with the resonator?*

Yes, just like the last sechmatic you posted. With ~2nF on the primary and no antenna.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

____

Quote

Download SOLAR SECRETS by Peter Lindemann
Free - Get it now: Solar Secrets

#152

05-08-2012, 07:59 PM

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **dR-Green**
> *Using a simplified "power" setup with no load makes all this easily visible.*

What simplification? Please elaborate.

Quote:

> Originally Posted by **dR-Green**
> *The signal should be in the primary otherwise there'll be no possibility of powering anything. .....So I'd say the place to be looking is mostly at the primary.*

With only the primary setup, (coil and capacitor) and the secondary (no capacitor), No terminal.. and grounded to the house mains or metal fireplace. There are two points of interest. One when the point of highest "activity" is at the top of the secondary, the other is a bit off that when "activity" is highest at one terminal of the primary condenser (only one the other terminal shows nothing).
At both of these points it will wirelessly light an av plug(led and 2 diodes in circle) about 0.25" away.

I'm not sure what makes an av plug light an led wirelessly. I thought it was high voltage but there is none of that here.
It is a loop and it is in the near field so it might be induction??

Anyone know what making the led light up?

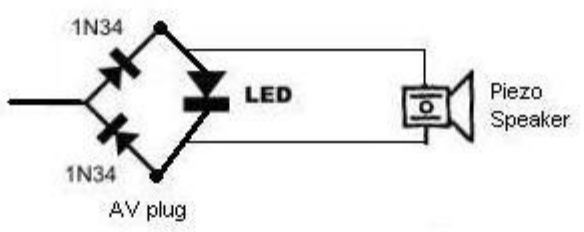

Uploaded with ImageShack.us
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

#153
05-08-2012, 08:50 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *the point of highest "activity" is at the top of the secondary,*

This is the normal high voltage point at resonance.

I suspect your radio is working with the house ground because there is a good ground at the circuit breaker box for the power lines. I also suspect that the circuit is not working outside because you do not have it well grounded. Your soil probably has too little moisture to provide a good ground. Grounding to dry soil is like not having a ground. For the ground to function it must have good conductivity to a large area. Even if there is not a good ground for the power lines, the wiring in your house covers a large area with good conductivity which is probably large enough to be a good ground.

However, I am puzzled by your being able to get a signal without the coil. This may indicate that the power line ground is not very good and the house wiring is acting like an antenna.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Last edited by xee2; 05-08-2012 at 09:06 PM.

#154
05-08-2012, 09:23 PM

dR-Green
Gold Member

Join Date: Apr 2008
Location: Merlin's Fort
Posts: 1,220

Quote:

> Originally Posted by **jake**
> *What simplification? Please elaborate.[/URL]*

If you remove all the diodes from the radio setup and the load from the power setup then you are left with only a capacitor across the primary so a very simple arrangement. As far as I can make out the bulb affects the capacitance/tuning so the 2nd cap is possibly in series with the bulb to reduce the extra capacitance. Either way if you just use the single cap and no load you can see on the scope how the reception (or "activity" in the primary) changes by adjusting the primary capacitance and the secondary condenser plates accordingly.

With my spiral for example by setting it up like that and tuning for best signal in the primary with no loads at all except a scope, to hear the radio all I need to do is place the C0 plate from Eric's radio diagram 25+cm away from the secondary terminal. That reduces the voltage in the primary but not too much because it's so far away relatively. This is probably where the flat spiral and TMT start to differ more significantly, but employing variable condenser plates and a separate receiver/C0 plate is a big improvement over using the C0 plate as both the receiver and tuning condenser, before needing to be within 5cm distance of the terminal to bring the frequency down enough and still hear the radio. Now it's already tuned without the C0 plate as can be seen in the primary.

Does the LED still light if you remove the speaker? It could be acting like a microphone although I don't see why it would work only in certain places in that case. It would be interesting if it still lights without the speaker
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

"Knowledge is cosmic. It does not evolve or unfold in man. Man unfolds to an awareness of it. He gradually discovers it." - Walter Russell

"Once men died for Truth, but now Truth dies at the hands of men." - Manly P. Hall

Quote

#155

05-08-2012, 09:33 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **xee2**
> *This is the normal high voltage point at resonance.*
>
> *I suspect your radio is working with the house ground because there is a good ground at the circuit breaker box for the power lines. I also suspect that the circuit is not working outside because you do not have it well grounded. Your soil probably has too little moisture to provide a good ground. Grounding to dry soil is like not having a ground. For the ground to function it must have good conductivity to a large area. Even if there is not a good ground for the power lines, the wiring in your house covers a large area with good conductivity which is probably*

> *large enough to be a good ground.*

The ground component is the next item on my agenda as soon as the large plate primary condenser is working at the same level as the current set up.
Thanks to recent Dollard transmissions on his thread I can finish that up. without excuse.

But the ground is still a mystery to me. Important as it is, how it is supposed to fuction is not clear in my head. I want to lay the ground system, one piece(radial) at at time and get a feel for what changes it makes to this setup. I also have not decided if I want to use the 17 rod or 80 wire radial Eric recommended.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#156
05-08-2012, 10:05 PM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **dR-Green**
> *If you remove all the diodes from the radio setup and the load from the power setup then you are left with only a capacitor across the primary so a very simple arrangement. As far as I can make out the bulb affects the capacitance/tuning so the 2nd cap is possibly in series with the bulb to reduce the extra capacitance. Either way if you just use the single cap and no load you can see on the scope how the reception (or "activity" in the primary) changes by adjusting the primary capacitance and the secondary condenser plates accordingly.*

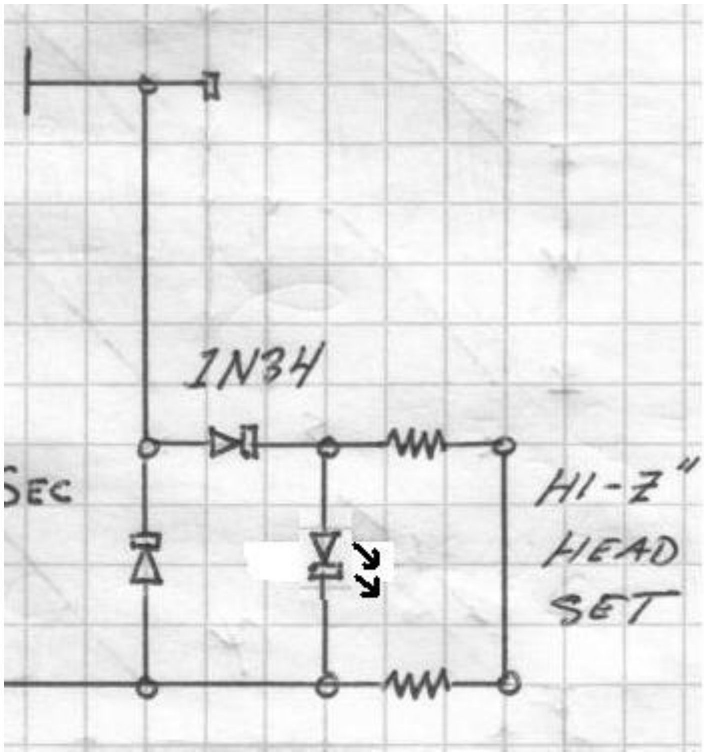

Nice, you just made me notice that the radio setup doide/capacitor/high z arrangement looks like the "AV plug / speaker" I made. Just replace the capacitor with the LED.

Quote:

> Originally Posted by **dR-Green**
> *With my spiral for example by setting it up like that and tuning for best signal in the primary with no loads at all except a scope, to hear the radio all I need to do is place the C0 plate from Eric's radio diagram 25+cm away from the secondary terminal. That reduces the voltage in the primary but not too much because it's so far away relatively. This is probably where the flat spiral and TMT start to differ more significantly, but employing variable condenser plates and a separate receiver/C0 plate is a big improvement over using the C0 plate as both the receiver and tuning condenser, before needing to be within 5cm distance of the terminal to bring the frequency down enough and still hear the radio. Now it's already tuned without the C0 plate as can be seen in the primary.*
>
> *Does the LED still light if you remove the speaker? It could be acting like a microphone although I don't see why it would work only in certain places in that case. It would be interesting if it still lights without the speaker*

Yes it will light without the speaker. I have three different plugs I have been playing with. They all light the led.
One is just like the one in the picture. Another is just like that but also has a ~50pf cap across it. The last does not have the speaker or capacitor.

Capacitance is needed on either side of the LED between the diodes and LED. I can hold it or sometimes just attaching an lead will work. If I feed it a signal using an extra 1 turn tuning coil it will light without the capacitance on the LED.

ALSO and a bit off topic but... IF YOU THUMP THE CRYSTAl SPEAKER THE LED BLINKS SUPER BRIGHT... It cought me off guard. And took me about 10 seconds to realize what happened.

I knew the potential for that to happen was there but I never thought it would do that to an LED with a little tiny thump.

now back to work.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#**157**
05-10-2012, 04:53 PM

jpolakow
Senior Member

Join Date: Aug 2011
Posts: 152

There's more material on the Eric Dollard thread, the recent writing on the Bolinas Antenna that applies to this. There's more information on the subject there. Eric asked me to make sure you guys saw that information. Also a lot of your questions have been answered there. Eric also said he is very impressed with your guys' work so far.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Please help support my indiegogo campaign: Cosmic Induction Generator

#**158**
05-12-2012, 08:03 AM

WARNING
BURN HAZARD

jake
Senior Member

Join Date: Nov 2007
Posts: 169

Quote:

> Originally Posted by **jpolakow**
> *There's more material on the Eric Dollard thread, the recent writing on the Bolinas Antenna that applies to this. There's more information on the subject there. Eric asked me to make sure you guys saw that information. Also a lot of your questions have been answered there. Eric also said he is very impressed with your guys' work so far.*

Thanks, I always try keep up to date on the Dollard thread. Great info. I hope Eric knows how much we appreciate his experience, direction and help.

@ all,

When I took my primary off to try and solder on the capacitor I realized that I don't know which way it should be wound.

Do you wrap in the same direction as the secondary, then overlap the first turn? (that's how I have it now) OR after one turn go under the first turn?

Does it matter?

Also how close are the turns on the primary? Eric said closely spaced. Does that mean as close as possible. I have about 1/8" spacing now but I was thinking I could get closer and use paper as the spacer. It becomes important to know this before you add you capacitor. Changing the spacing after I have attached the capacitor will cause the plates to be misaligned.

What effect does spacing have on the primary and how does that carry over to the secondary? I tried to experiment with this but I noticed that if I make the spacing larger I end up shorter on the number of turns. That is changing two variables at once and I can't draw any conclusions.

My thoughts are with 2 turns fixed: more spacing = more inductance(increased area), but also less capacitance between turns. ???

Thanks,
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#159
05-12-2012, 12:43 PM

xee2
Silver Member

Join Date: Jan 2010
Posts: 580

Quote:

> Originally Posted by **jake**
> *Does it matter?*

I would say that it does not matter for capacitance. However, the current flow should be in the same direction as the secondary turns. The spacing between turns changes the self capacitance of the loop. Less spacing equals more capacitance. Moving turns closer together is like adding parallel capacitance so it will effect the tuning.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

#160
05-13-2012, 11:17 AM

Aether
Member

Join Date: Aug 2010
Posts: 36

Editeditedit
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Last edited by Aether; 05-13-2012 at 11:18 AM. Reason: wrong thread

#161
05-15-2012, 10:58 AM

t-rex
Senior Member

Join Date: Aug 2011
Location: In the bushes
Posts: 394

Post numbers here refer to the Crystal Radio Magic Thread

Post 73, Jake.

Standing wave on coil indicates proper operation. A voltage node must exist exactly at the neutral terminal. Tesla could move this node about in the earth. This is important and in an optimum state the voltage node is right at the point of earthing. A projected wave velocity exist on the earth surface, infinite at the earth connection, and luminal 90 degrees around the planet. It is a tangent function with Pi over two as the "Effective velocity". Hence velocity here is only a projection on the earth surface, not a true velocity, like your shadow at sundown it goes to infinity.

Post 136, Madhatter.

The B field is a bad way to look at this. It is a physics dingleberry. B is a dimensional relation with a per centimetre square. Here it is best to use versor magnetic induction k Phi, see last part of Theory of Wireless Power. Mutual inductance is longitudinal so that coupling remains for concatenated longitudinal resonators. Wave direction now has a role in coupling.

Post 143, Madhatter

Wave direction coupling makes balance to unbalance transformation impossible, this gives the "dead phase" extra coil. If it is a balanced system, every part of that system must be push-pull. Unbalanced unit has neutral on one side of primary, a balanced unit has neutral on primary center tap. All concatenated coils must be wound in same direction as wave propogation spiraling around them.

Post 156, Jake

Do not replace 100 pF condenser with LED, it will distort and rob power from the audio. Use a completely separate setup for LED. Try Ne 2 neon lamps. Also a 10-15 watt fluorescent lamp is a good standing wave "wand", this for tuning.

I would recommend using a GSN7 Triode section with cathode drive from function generator and the primary as the plate tank circuit. See the figure. This will greatly improve resonance for tuning. 50 Ohm across primary NFG, GSN7>6000 Ohm general.

Multiple tuning peaks indicate overly tight coupling between coils. Slowly separate them until only one resonant peak. Multiple modes of stray reception and strong local stations will make grounding and tuning very confusing. An isolation amplifier GSN7 triode isolates amplifier from function generator interaction and gives more power to light neon lamps. Both sections of GSN7 in parallel give more power.

73 DE N6KPH

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

SUPPORT ERIC DOLLARD'S WORK AT EPD

LABORATORIES, INC.

Purchase Eric Dollard's Books & Videos - 70% of the sale goes to Eric and EPD Laboratories:
Eric Dollard Books & Videos
Donate by Paypal: Donate to EPD Laboratories

Edit Quote

#162
05-15-2012, 12:40 PM

madhatter
Senior Member

Join Date: Nov 2010
Posts: 457

Quote:

> Originally Posted by **T-rex**
> *Post numbers here refer to the Crystal Radio Magic Thread*
>
> *Post 136, Madhatter.*
>
> *The B field is a bad way to look at this. It is a physics dingleberry. B is a dimensional relation with a per centimetre square. Here it is best to use versor magnetic induction k Phi, see last part of Theory of Wireless Power. Mutual inductance is longitudinal so that coupling remains for concatenated longitudinal resonators. Wave direction now has a role in coupling.*

I'll re-review that, also I was pointed towards the 2002 Oerstad Medal Lecture that covers the need to use geometric algebra -Clifford algebra- to reduce the fragmentation physics has undergone with the avoidance of vectors.

Quote:

> Originally Posted by **T-rex**
>
> *Post 143, Madhatter*
>
> *Wave direction coupling makes balance to unbalance transformation impossible, this gives the "dead phase" extra coil. If it is a balanced system, every part of that system must be push-pull. Unbalanced unit has neutral on one side of primary, a balanced unit has neutral on primary center tap. All concatenated coils must be wound in same direction as wave propogation spiraling around them.*

I figured that may be the issue when driving the CW set and then placing a sheet of plexi between them then the CCW was able to be tuned. more experimentation to follow as time permits.

Quote:

> Originally Posted by **T-rex**
>
> *Post 156, Jake*
>
> *Do not replace 100 pF condenser with LED, it will distort and rob power from the audio. Use a completely separate setup for LED. Try Ne 2 neon lamps. Also a 10-15 watt fluorescent lamp is a good standing wave "wand", this for tuning.*
>
> *I would recommend using a GSN7 Triode section with cathode drive from function generator and the primary as the plate tank circuit. See the figure. This will greatly improve resonance for tuning. 50 Ohm across primary NFG, GSN7>6000 Ohm general.*
>
> *Multiple tuning peaks indicate overly tight coupling between coils. Slowly separate them until only one resonant peak. Multiple modes of stray reception and strong local stations will make grounding and tuning very confusing. An isolation amplifier GSN7 triode isolates amplifier from function generator interaction and gives more power to light neon lamps. Both sections of GSN7 in parallel give more power.*
>
> *73 DE N6KPH*

Thanks for the input and the time to respond to our efforts here.

Have you read any of Julius Adam Strattons' notes? I was going over his notes on electromagnetic theory from 1941 as he has what I've found to be the most coverage of electrostatic energy and fields.
excerpt:
Conductors play an especially prominent role in electrostatics. For the purposes of a purely macroscopic theory it is sufficient to consider a conductor as a closed domain within which charge moves freely. If the conductor is a metal or electrolyte, the flow of charge is directly proportional to the intensity E of the electric field: J = aE. Charge is free to move on the surface of a conductor but can leave it only under the influence of very intense external fields or at high temperatures (thermal emission). If the conductor is in electrostatic equilibrium all flow of charge has ceased, whence it is evident that at every interior point of a conductor in an electrostatic field the resultant field intensity E is zero, and at every point on its surface the tangential component of E is zero. Furthermore the electrostatic potential within a conductor is constant and the surface of every conductor is an equipotential. Let us suppose that an

uncharged conductor is introduced into a fixed external field E . In the .first instant there occurs a transient current.

POSTS FROM THE Eric's "Crystal Radio Initiative" Challenge THREAD STARTED BY SEBACID

http://www.energeticforum.com/crystal-radio-initiative/14371-erics-crystal-radio-initiative-challenge.html

08-21-2013, 03:30 AM

Sebacid
Junior Member

Join Date: Jul 2012
Posts: 17

⚠Eric's "Crystal Radio Initiative" Challenge

I thought this challenge needs some promoting to get us all involved in it.

I want people who are taking on this challenge to post their progress, results, questions, photos etc. in this thread.

Basically anything that involves this challenge belongs to this thread. We can assist each other in problems, setbacks, windings etc. and generally work together in completing Eric's challenge.

The objective would be to see how many of us can successfully complete this challenge. Hopefully it also encourages more people to experiment in this field.

Remember, the idea of the challenge is to get us *thinking* and *doing* things on our own, so please try it yourself first, then if you hit a wall that you can't breach, post here and ask help.

Don't give blueprints or schematics for a working system! If you manage to complete this challenge, post the date of completion here and we can compare results later in another thread specifically made for complete, tested and working setups.

In short, get off your monitor and start experimenting! 😉

-Sebastian Bowles -Lost in Northern Territory
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

____

Quote

Sponsored Links

EPD LABORATORIES
TESLA ADVANCED SEISMIC
WARNING SYSTEM #2
YOU CAN HELP SAVE A MILLION LIVES

SUPPORT EPD LABORATORIES, INC.

Download SOLAR SECRETS by Peter Lindemann
Free - Get it now: Solar Secrets

#2

08-22-2013, 05:45 AM

Sebacid
Junior Member

Join Date: Jul 2012
Posts: 17

Some excellent advice and info by Dr. Green

Tesla Magnifying Transmitter/Eric Dollard Type Coils Compendium

Some very helpfull scientific data and info on coil building etc. here. Use it to your advantage and learn from it.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#3

08-22-2013, 08:31 AM

LutherG
Member

Join Date: Apr 2007
Posts: 91

Quote:

> Originally Posted by **Sebacid**
> *I thought this challenge needs some promoting to get us all involved in it.*
>
> *I want people who are taking on this challenge to post their progress, results, questions, photos etc. in this thread.*
>
> *Basically anything that involves this challenge belongs to this thread. We can assist each other in problems, setbacks, windings etc. and generally work together in completing Eric's challenge.*
>
> *The objective would be to see how many of us can successfully complete this challenge. Hopefully it also encourages more people to experiment in this field.*

> *Remember, the idea of the challenge is to get us thinking and doing things on our own, so please try it yourself first, then if you hit a wall that you can't breach, post here and ask help.*
>
> *Don't give blueprints or schematics for a working system! If you manage to complete this challenge, post the date of completion here and we can compare results later in another thread specifically made for complete, tested and working setups.*
>
> *In short, get off your monitor and start experimenting!*
>
> *-Sebastian Bowles -Lost in Northern Territory*

Hi Sebastian,

What specifically is this challenge and where can I read up on it?

Best regards,

Luther
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Electrostatic charges manipulating magneto-gravitic streams...

#4

08-23-2013, 12:52 AM

Sebacid
Junior Member

Join Date: Jul 2012
Posts: 17

Quote:

> Originally Posted by **LutherG**
> *Hi Sebastian,*
>
> *What specifically is this challenge and where can I read up on it?*
>
> *Best regards,*
>
> *Luther*

Hi Luther,

The challenge is located amidst the jungle in this forum in Eric's posts as T-Rex.

Here is a short copy describing it:

"The objective here is to scale the “Crystal Set”, a step at a time, into a Tesla Transformer for the reception of medium wave band, 300 – 3000 kilocycle A.M. broadcasts. No license is required for this and the broadcast station provides the power.And this objective cooperates with the primary objective. That is; Who will be the first ham to disprove Einstein’s theory? An International contest, but who will sponsor it, Iran maybe?We have the good fortune in the “Crystal Set Initiative” that, in theory at least, a quarter wave A.M. broadcast tower, and its 120 quarter wave ground radials, must emit a pair of waves as shown by Tesla in his basic diagrams.

Hence it can be seen that a pair of waves are engendered by this transmission system. (Tower and Star Radials). One wave, Hertzian, is the over ground wave, the other wave, Telluric, is the under ground wave. These two waves arrive at the point of reception in their own distinct time frames, giving rise to a difference in phase. Hence, multiple rings of interference patterns are produced. Since the Hertzian portion, over ground, time frame is based upon the velocity of light, then the Telluric portion, under ground, time frame gives the Telluric velocity. Two crystal sets, one over ground, one under ground, and a basic oscilloscope , that simple. I have done this at Landers.Concluding, a Tesla Magnification Transformer, properly proportioned can, in theory, actually draw power from a local 50 kW station. Several hundred watts of power reception is likely. This would prove Tesla once and for all. No antenna, just a good ground, and a nice and bright 100 watt light bulb.This would overturn physics more than any billion dollar C.E.R.N. project. A ham radio operator overturns Einstein for 100 bucks. What a concept."

Read,
Tesla, “The True Wireless”
Tesla, “System of Concatenated Tuned Circuits”
Dollard, “System for the Transmission and Reception of Telluric Electric Waves”
A.R.R.L. “Radio Amatuers Handbook”. Chapter “H.F. Transmitters, & Tank Circuits”

Here is another link that has most of Eric's posts shortlisted:

Eric Dollard | Gestalt Reality

Hope this helps,

-Seb
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#5

08-24-2013, 04:42 PM

wayne.ct
Senior Member

Join Date: Jan 2011
Posts: 356

What would this look like?

I have looked at this and done a bit of reading at the context and I don't really have a good feeling for where to start. I am thinking perhaps I should apply some of the functions and formulae that EPD has gathered and perhaps compute the size and other parameters of the necessary coils and capacitors? If someone would suggest some specific numbers and create an example of the calculations, that would seem like a helpful starting point. First, pick a typical AM frequency, such as that of your local 50 KVA AM station, and use that as an example. It seems we need to solve a simultaneous equation. There are two values for inductors and also two? values for capacitors. Then there is the actual circuit diagram. Again, I have a picture in my mind, but I'm not sure if it is right. If you want to post you sample data, my eyes are open. Thx.

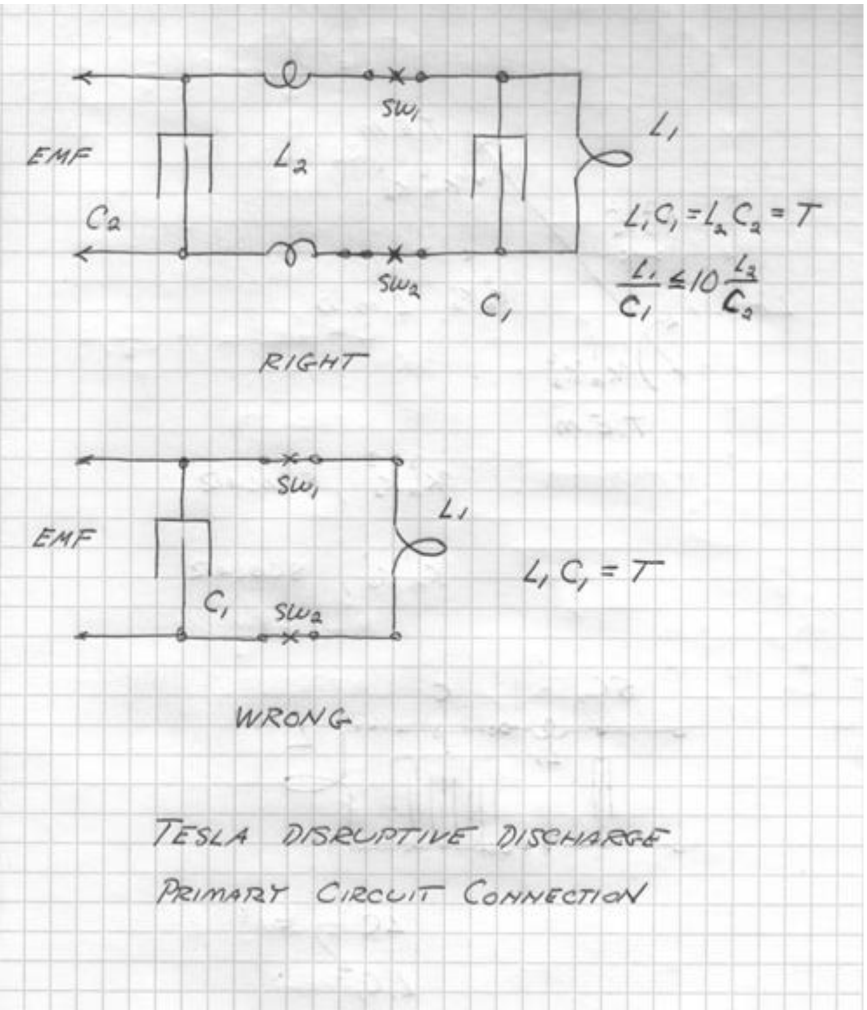
SW₁
EMF
L₂
L₁
C₂
$L_1C_1 = L_2C_2 = T$
SW₂
$\frac{L_1}{C_1} \leq 10\frac{L_2}{C_2}$
C₁
RIGHT
SW₁
L₁
EMF
$L_1C_1 = T$
C₁
SW₂
WRONG
TESLA DISRUPTIVE DISCHARGE
PRIMARY CIRCUIT CONNECTION

08-24-2013, 10:14 PM

Sebacid
Junior Member

Join Date: Jul 2012
Posts: 17

Quote:

> Originally Posted by **wayne.ct**
> *I have looked at this and done a bit of reading at the context and I don't really have a good feeling for where to start. I am thinking perhaps I should apply some of the functions and formulae that EPD has gathered and perhaps compute the size and other parameters of the necessary coils and capacitors? If someone would suggest some specific numbers and create an example of the calculations, that would seem like a helpful starting point. First, pick a typical AM frequency, such as that of your local 50 KVA AM station, and use that as an example. It seems we need to solve a simultaneous equation. There are two values for inductors and also two? values for capacitors. Then there is the actual circuit diagram. Again, I have a picture in my mind, but I'm not sure if it is right. If you want to post you sample data, my eyes are open. Thx.*
> *Attachment 13624*

Hi Wayne,

All the information necessary to do this can be found in this forum under Eric's posts as T-Rex. See links in my previous reply. You just have to get your hands dirty and look for the information. That is a part of the challenge, to get people *doing*.

But here's some advice to get you started:

"The design of the Tesla Transformer starts with the secondary resonant coil. In its resonant mode it is a quadrapolar resonator. Two constants exist in the construction of this solenoidal resonator, one is that the height to length ratio must be 20%, the other is that the side by side spacing of the cylindrical conductor turns is 62% the diameter of the cylindrical conductor. The optimum number of turns is 20 on the secondary, with 2 on the primary. This gives the physical size of the coil for a given frequency. Increasing the number of turns reduces the size of the secondary coil in proportion to the increase in turns. Accordingly the conductor diameter also decreases as the copper is spread out over more windings. The ratio of secondary turns to primary turns must equal a constraint of 10 to 1. For a given number of turns the operating frequency becomes the sole independent variable in coil design.The primary coil is the same diameter as the secondary coil. The ratio of conductor width to coil diameter is 18%.

The sheet conductor is closely spiralled into two turns, for a secondary of twenty turns. For a conductor thickness as given by the maximum thickness vs frequency the volume of the primary metal must equal the volume of the secondary metal, based upon equal weights for primary and secondary metals. Brass can be 1.7 times thicker than copper, and bronze can be 2.8 times thicker than copper, for a given maximum thickness and frequency.The primary condenser should be made of the same amount of metal as the primary coil. Their weights should be about the same. The connecting leads must be short and the same width as the primary conductor.

Large, multiplate air condensers are good, so are metal sheets and glass.The secondary capacitance is best a copper tubing ring around the outside of the H end of the coil. This ring must be open in one spot along its loop so as not to be a shorted turn. The end of the secondary wire connects to this capacity ring. Beer cans make the best elevated capacitors.The finished transformer must operate on a directly connected ground plane. A large plywood table with aluminum foil glued on to it, this covered with a plastic or glass sheet is the minimum required ground plane. Obviously the aluminum plane must be solidly earthed with special attention to minimizing ground lead inductance. In order to light a lamp a more broadcast station style of ground plane is required, but for testing purposes the metallized plywood “ten by” sheet is just fine. Good luck and good crystal set DX ing.

TESLA RESONANT TRANFORM

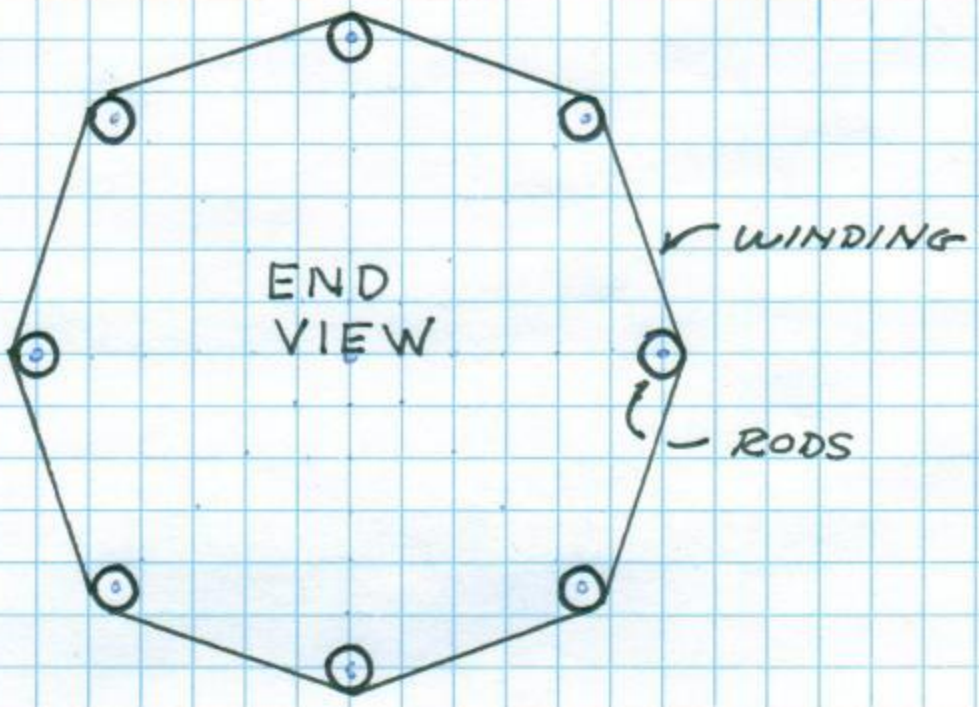
WINDING
END VIEW
RODS

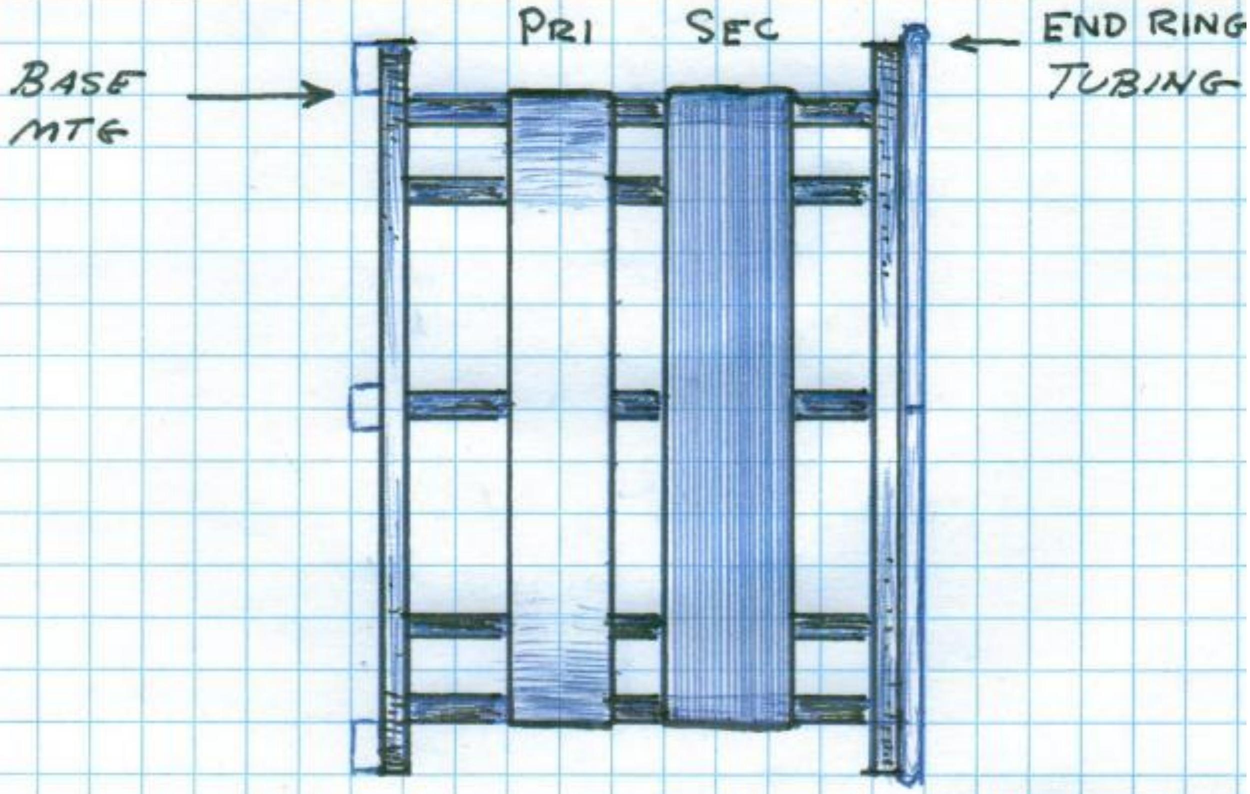
PRI
SEC
END RING TUBING
BASE MTG

SIDE

#7

08-25-2013, 07:28 PM

wayne.ct
Senior Member

Join Date: Jan 2011
Posts: 356

Sample calculations

OK, based on your suggestions and the calculations you have posted, I can visualize the transformer. It will be small enough to hold in one hand.

Here is what I have done so far:

I will be using 750 KHz as my center frequency as an example. This frequency is used by the famous WSB.

WSB is a 50,000 watt clear channel broadcasting station in Atlanta, GA.

The antenna is located in Tucker, GA according to Wikipedia.

Calculation #1

L = Total length of coiled wire

L = f/w = 299,792,458 (meter / second) / 2 * PI * 750,000 (/second)

= 299,792,458 / 2 * 3.14159 * 750,000 (meters) = 63.62 meters = 63.62 meters * .3048 ft / meter

= 19.39 feet

Calculation #2

L = Length of each turn = 19.39 feet / 20 = 0.9695 feet = 11.63 inches

Calculation #3

Circumference C = L; Diameter = Circumference / PI = 3.7 inches

H = Coil height = 0.2 * Diameter = 0.2 * (11.63 inches / PI)

= 0.74 inches

Calculation #4

Max diameter of wire. Space between strands of wire is 62 percent of wire diameter.

20 turns take 0.74 inches ===> 1 turn takes 0.74 inches / 20 = 0.037 inches

0.037 inches = 162 percent of wire diameter ===> 100 percent of wire diameter = 0.037 / 1.62

Wire diameter = 0.02286 inches (max.)

AWG 23 has a diameter of 0.0226 inches / 0.573 mm.

AWG 23 has resistance of 66.79 ohms / Km. which is 20.36 mOhms / ft.

For 19.39 ft, the coil will have 395 mOhms or 0.395 Ohms resistance.

I thought it would be larger than this. So much for my intuition.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

There is a reason why science has been successful and technology is widespread. Don't be afraid to do the math and apply the laws of physics.

Quote

#8

08-26-2013, 02:47 AM

Sebacid
Junior Member

Join Date: Jul 2012
Posts: 17

Quote:

> Originally Posted by **wayne.ct**
> *OK, based on your suggestions and the calculations you have posted, I can visualize the transformer. It will be small enough to hold in one hand.*
>
> *Here is what I have done so far:*
>
> *I will be using 750 KHz as my center frequency as an example. This frequency is used by the famous WSB.*
>
> *WSB is a 50,000 watt clear channel broadcasting station in Atlanta, GA.*
>
> *The antenna is located in Tucker, GA according to Wikipedia.*
>
> *Calculation #1*
>
> *L = Total length of coiled wire*

> *L = f/w = 299,792,458 (meter / second) / 2 * PI * 750,000 (/second)*
>
> *= 299,792,458 / 2 * 3.14159 * 750,000 (meters) = 63.62 meters = 63.62 meters * .3048 ft / meter*
>
> *= 19.39 feet*
>
> *Calculation #2*
>
> *L = Length of each turn = 19.39 feet / 20 = 0.9695 feet = 11.63 inches*
>
> *Calculation #3*
>
> *Circumference C = L; Diameter = Circumference / PI = 3.7 inches*
>
> *H = Coil height = 0.2 * Diameter = 0.2 * (11.63 inches / PI)*
>
> *= 0.74 inches*
>
> *Calculation #4*
>
> *Max diameter of wire. Space between strands of wire is 62 percent of wire diameter.*
>
> *20 turns take 0.74 inches ===> 1 turn takes 0.74 inches / 20 = 0.037 inches*
>
> *0.037 inches = 162 percent of wire diameter ===> 100 percent of wire diameter = 0.037 / 1.62*
>
> *Wire diameter = 0.02286 inches (max.)*
>
> *AWG 23 has a diameter of 0.0226 inches / 0.573 mm.*
>
> *AWG 23 has resistance of 66.79 ohms / Km. which is 20.36 mOhms / ft.*
>
> *For 19.39 ft, the coil will have 395 mOhms or 0.395 Ohms resistance.*
>
> *I thought it would be larger than this. So much for my intuition.*

There is a small mistake in your calculations...

L = Total length of coiled wire

L = *f/w* THIS IS WRONG, it should be c/w! (but you have calculated it correctly) = 299,792,458 (meter / second) / 2 * PI * 750,000 (/second)

= 299,792,458 / 2 * 3.14159 * 750,000 (meters) = 63.62 meters ***= 63.62 meters * .3048 ft / meter***

= *19.39 feet* THIS IS WRONG!

If converted then 63,62 meters gives roughly about 209 feet (63,62x3,28=208,67).

That shouldn't fit in your hand anymore... ☺

Remember you need two different setups, one wave, Hertzian, is the over ground wave, the other wave, Telluric, is the under ground wave. These two waves arrive at the point of reception in their own distinct time frames, giving rise to a difference in phase. Hence, multiple rings of interference patterns are produced. Since the Hertzian portion, over ground, time frame is based upon the velocity of light, then the Telluric portion, under ground, time frame gives the Telluric velocity. Two crystal sets, one over ground, one under ground, and a basic oscilloscope , that simple.

A great link for most information needed is provided in my first answer in this thread. Happy experimenting!

-Seb
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Last edited by Sebacid; 08-26-2013 at 12:37 PM.

#9

08-28-2013, 10:03 AM

Sebacid
Junior Member

Join Date: Jul 2012
Posts: 17

Just finished my calculations for my secondary coil and it will be a whopping 76cm in diameter! I am going with 1000Kc/sec Frequency.

I finished the work on the wooden bases and supports, all that is left is to bake and laquer them before assembling.

My extra coil's diameter/height is going to be 30cm.

Some calculations with other stuff is still going on, but I'm going to build two identical coils of the exact same size, one for overground Hertzian wave reception and the other for underground Tellurical wave reception.

The coils will be identical but the setup will be different as the goal is to light a 100W lightbulb with the other coil.

We'll see how things go once I have built them.

-Seb
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#10
08-30-2013, 12:48 AM

Sebacid
Junior Member

Join Date: Jul 2012
Posts: 17

Quote:

> Originally Posted by **Sebacid**
> *Just finished my calculations for my secondary coil and it will be a whopping 76cm in diameter! I am going with 1000Kc/sec Frequency.*
>
> *I finished the work on the wooden bases and supports, all that is left is to bake and laquer them before assembling.*
>
> *My extra coil's diameter/height is going to be 30cm.*
>
> *Some calculations with other stuff is still going on, but I'm going to build two identical coils of the exact same size, one for overground Hertzian wave reception and the other for underground Tellurical wave reception.*
>
> *The coils will be identical but the setup will be different as the goal is to light a 100W lightbulb with the other coil.*
>
> *We'll see how things go once I have built them.*
>
> *-Seb*

Thinking of replacing solid wooden supports, with hollow pvc tubing (thin) to minimize stray inductance? Any thoughts on this?
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#11

08-30-2013, 03:43 AM

dR-Green
Gold Member

Join Date: Apr 2008
Location: Merlin's Fort
Posts: 1,220

Hi Sebastian. Yes PVC tubing will be better than wood. Also ideally you should design the coil for a specific frequency/radio station, 1000 kc is a general example. When tuned to different frequencies the magnification factor will alter, the goal is to have highest magnification factor at the intended frequency so different tuning will work against this. Although you will need to experiment to find the highest magnification factor with a given design anyway.

For the overground wave reception you should use a normal crystal radio or a normal radio receiver. You need to receive the signal as normal and then measure the difference.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

"Knowledge is cosmic. It does not evolve or unfold in man. Man unfolds to an awareness of it. He gradually discovers it." - Walter Russell

"Once men died for Truth, but now Truth dies at the hands of men." - Manly P. Hall

Quote

Sponsored Links

EPD LABORATORIES
TESLA ADVANCED SEISMIC
WARNING SYSTEM #2
YOU CAN HELP SAVE A MILLION LIVES

SUPPORT EPD LABORATORIES, INC.

#12

08-30-2013, 05:04 AM

Sebacid
Junior Member

Join Date: Jul 2012
Posts: 17

Quote:

> Originally Posted by **dR-Green**

> *Hi Sebastian. Yes PVC tubing will be better than wood. Also ideally you should design the coil for a specific frequency/radio station, 1000 kc is a general example. When tuned to different frequencies the magnification factor will alter, the goal is to have highest magnification factor at the intended frequency so different tuning will work against this. Although you will need to experiment to find the highest magnification factor with a given design anyway.*
>
> *For the overground wave reception you should use a normal crystal radio or a normal radio receiver. You need to receive the signal as normal and then measure the difference.*

Hi Dr. Green,

Thank you for your insight, I will have to do some searching on a AM radio station near me (Finland). The problem here is that almost all of the broadcasting stations have ceased using the AM frequencys. I think the nearest one is in Holland?! Not sure on this, must check.

In the calculations for the extra coil, I get a totally different total lenght of coil if I use the formulae by Eric to calculate it?

Tesla's Colorado extra coil had the same amount of wire as the secondary (800m). Should I go with this or trust the formulae?

Thanks again,

-Seb

Post edit:

Found a radio station from Sweden, Stockholm radio that broadcasts at 1710Khz AM frequency. I think it is still operational, so I will try working with this frequency. The last AM radiostation in Finland ceased to operate in 2007. Any other known operational frequencys are welcome! 🙂 Time to do the calculations again! 🙂
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Last edited by Sebacid; 08-30-2013 at 05:55 AM. Reason: Adding information

Quote

#13

08-30-2013, 06:32 AM

dR-Green
Gold Member

Join Date: Apr 2008
Location: Merlin's Fort
Posts: 1,220

I forgot to mention, use white or clear etc plastic rods. Apparently to make black and grey they put carbon in it, so that won't be good.

Eric's design is different to Tesla's. Eric's theory involves concatenated resonance with the extra coil as a distributed network, whereas Tesla's is lumped. So it's completely experimental. I've tried 3 different extra coil variations, and Eric's latest comments on it was that the secondary might need to be of a lower inductance, so my latest design is a bit different. I have 17 turns secondary with 15% height to diameter ratio, and the extra coil, when calculating wavelength from frequency and speed of light, has 1.19 times longer wire length because of the "faster than light" propagation, wire length is extended to bring the frequency back down.

[edit] Probably the easiest way to find whether a radio station is operational and suitable or not is to use a radio. When you find one then you can find info on it on the internet.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

"Knowledge is cosmic. It does not evolve or unfold in man. Man unfolds to an awareness of it. He gradually discovers it." - Walter Russell

"Once men died for Truth, but now Truth dies at the hands of men." - Manly P. Hall

Last edited by dR-Green; 08-30-2013 at 06:57 AM.

Quote

#**14**
08-30-2013, 10:44 AM

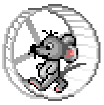

Kokomoj0
Senior Member

Join Date: Jul 2011
Posts: 386

for accurate time comparison how do you compensate and adjust for delay differences that result from the inherent design differences between the telluric and atmospheric versions?

and one more....

how do we calculate the maximum theoretical (ideal conditions) power that can be drawn from a transmitter with a particular receiver design/situation
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Last edited by Kokomoj0; 08-30-2013 at 10:48 AM.

#**15**

08-30-2013, 11:07 AM

dR-Green
Gold Member

Join Date: Apr 2008
Location: Merlin's Fort
Posts: 1,220

Quote:

> Originally Posted by **Kokomoj0**
> *for accurate time comparison how do you compensate and adjust for delay differences that result from the inherent design differences between the telluric and atmospheric versions?*

Use an oscillator and tune both receivers to be in phase.

Quote:

> Originally Posted by **Kokomoj0**
> *how do we calculate the maximum theoretical (ideal conditions) power that can be drawn from a transmitter with a particular receiver design/situation*

Beats me. How's your geological skills?
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

"Knowledge is cosmic. It does not evolve or unfold in man. Man unfolds to an awareness of it. He gradually discovers it." - Walter Russell

"Once men died for Truth, but now Truth dies at the hands of men." - Manly P. Hall

#**16**

09-01-2013, 10:14 AM

wayne.ct
Senior Member

Join Date: Jan 2011
Posts: 356

This is out of my league

Correcting my error and still using 750 KHz as my center frequency, this is what I get...

Calculation #1

L = Total length of coiled wire

L = c/w = 299,792,458 (meter / second) / 2 * PI * 750,000 (/second)

= 299,792,458 / 2 * 3.14159 * 750,000 (meters) = 63.62 meters = 63.62 meters / .3048 (meter / ft)

= 208.7 feet

Calculation #2

L = Length of each turn = 208.7 feet / 20 = 10.44 feet = 10 feet 5.23 inches

Calculation #3

Circumference C = L; Diameter = Circumference / PI = 3.32 feet = 3 feet 3.86 inches

H = Coil height = 0.2 * Diameter = 0.2 * (10.44 feet / PI)

= 7.97 inches

Calculation #4

Max diameter of wire. Space between strands of wire is 62 percent of wire diameter.

20 turns take 7.97 inches ===> 1 turn takes 7.97 inches / 20 = 0.4 inches

0.4 inches = 162 percent of wire diameter ===> 100 percent of wire diameter = 0.4 / 1.62

Wire diameter = 0.246 inches (max.)

AWG 3 has a diameter of 0.2294 inches / 5.83 mm.

AWG 3 has resistance of 0.646 ohms / Km. which is 0.197 mOhms / ft.

For 208.7 ft, the coil will have 395 mOhms or 0.4111 Ohms resistance.

Now we have a manly sized inductor! 'Sorry to say I will not be building that any time soon! Thanks for the ride!

At the same time, my intuition was telling me this was going to be a large coil. So, I feel a bit better knowing my gut was in the ballpark.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on

pinterest_share Share on linkedin

There is a reason why science has been successful and technology is widespread. Don't be afraid to do the math and apply the laws of physics.

#17

09-02-2013, 02:14 AM

Sebacid
Junior Member

Join Date: Jul 2012
Posts: 17

Quote:

> Originally Posted by **wayne.ct**
> _Correcting my error and still using 750 KHz as my center frequency, this is what I get..._
>
> _Calculation #1_
>
> _L = Total length of coiled wire_
>
> _L = c/w = 299,792,458 (meter / second) / 2 * PI * 750,000 (/second)_
>
> _= 299,792,458 / 2 * 3.14159 * 750,000 (meters) = 63.62 meters = 63.62 meters / .3048 (meter / ft)_
>
> _= 208.7 feet_
>
> _Calculation #2_
>
> _L = Length of each turn = 208.7 feet / 20 = 10.44 feet = 10 feet 5.23 inches_
>
> _Calculation #3_
>
> _Circumference C = L; Diameter = Circumference / PI = 3.32 feet = 3 feet 3.86 inches_
>
> _H = Coil height = 0.2 * Diameter = 0.2 * (10.44 feet / PI)_
>
> _= 7.97 inches_
>
> _Calculation #4_
>
> _Max diameter of wire. Space between strands of wire is 62 percent of wire diameter._
>
> _20 turns take 7.97 inches ===> 1 turn takes 7.97 inches / 20 = 0.4 inches_

> *0.4 inches = 162 percent of wire diameter ===> 100 percent of wire diameter = 0.4 / 1.62*
>
> *Wire diameter = 0.246 inches (max.)*
>
> *AWG 3 has a diameter of 0.2294 inches / 5.83 mm.*
>
> *AWG 3 has resistance of 0.646 ohms / Km. which is 0.197 mOhms / ft.*
>
> *For 208.7 ft, the coil will have 395 mOhms or 0.4111 Ohms resistance.*
>
> *Now we have a manly sized inductor! 'Sorry to say I will not be building that any time soon! Thanks for the ride!*
>
> *At the same time, my intuition was telling me this was going to be a large coil. So, I feel a bit better knowing my gut was in the ballpark.*

You can always use another frequency (higher), then the coil, wire etc. size will be smaller. Try for something in between 2000-3000Kc/sec? As Dr. Green said earlier, the idea is to magnify the chosen frequency to the max!

Don't give up, build a smaller one first, experiment on it and if succesfull, upscale it on the next one!

Cheers,

-Seb
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

#18

09-13-2013, 01:22 AM

Sebacid
Junior Member

Join Date: Jul 2012
Posts: 17

Wire?

Does it matter if I use insulated wire for the winding? Or is it better if it was copper?
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

Quote

#19

09-23-2013, 07:02 AM

orgonaut314
Senior Member

Join Date: Jul 2013
Posts: 138

Quote:

> Originally Posted by **Sebacid**
> *Does it matter if I use insulated wire for the winding? Or is it better if it was copper?*

Every insulation will increase the self capacitance. Now I think it was better to keep that one low as possible.

Apart from that be careful with the material you use. If there is Orgon than organic materials like wood or materials like coper are preferable. In another type of research we found out that Orgon does become sickening if you use the wrong materials that your body does not like when you would eat or drink them. I would be careful with pvc that contains chlore. Even the thin insulation on the coper wire might be wrong. But this is only important if you want to make a coil that does not make people get an head-age.
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

#20

08-01-2014, 12:02 PM

Jeff Pearson
Member

Join Date: Jan 2012
Location: Taos NM
Posts: 85

I was experimenting with crystal sets last year
Was listening to clear channel stations 2 states away at night with crappy antenna. Powered piezo headset i made. no power source for radio except radio waves. I learned a lot doing this. was gonna try ground antenna next Harris Rogers style.

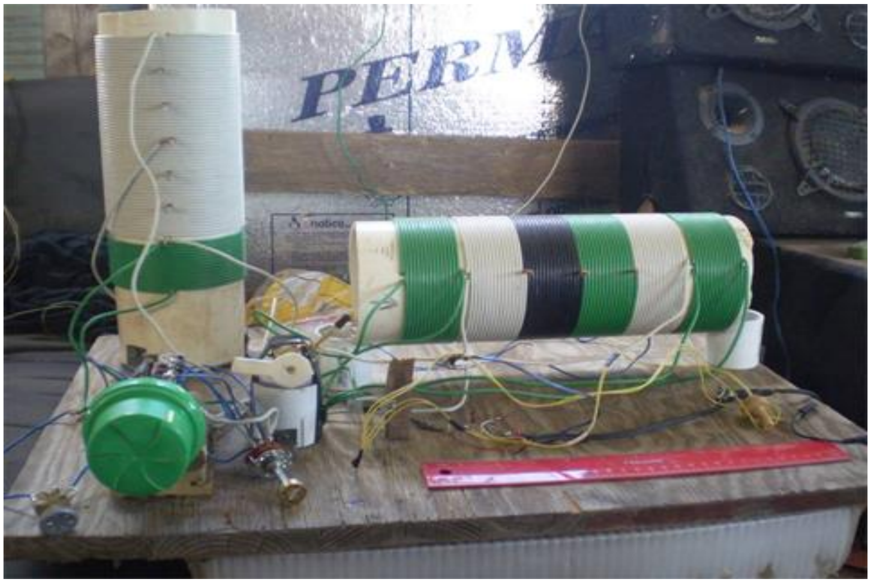

08-07-2014, 07:01 AM

parkham
Junior Member

Join Date: Jul 2014
Location: US/IN/IN
Posts: 12

I hope this isn't a completely question - When he says "Total number of turns", Given 20 - what does he mean by 20? Would that be 20 turns per inch?

The same for the next formula, "Given 20%" What exactly does that mean as well?

Please pardon my ignorance. I've got all the components together to start on a crystal radio - I want to fully understand before proceeding though.

Thanks,
Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

Quote

Sponsored Links

EPD LABORATORIES
TESLA ADVANCED SEISMIC
WARNING SYSTEM #2
YOU CAN HELP SAVE A MILLION LIVES

SUPPORT EPD LABORATORIES, INC.

#22

08-07-2014, 01:35 PM

dR-Green

Gold Member

Join Date: Apr 2008
Location: Merlin's Fort
Posts: 1,220

"Total number of turns = 20" means exactly what it says. 20 turns in total on the coil.

The 20% is in reference to the height/length to diameter ratio of the coil. The coil is 20% high as it is wide. If diameter = 100cm then height = 20cm.

Share on facebook Share on twitter Share on google_plusone_share Share on reddit Share on pinterest_share Share on linkedin

---

"Knowledge is cosmic. It does not evolve or unfold in man. Man unfolds to an awareness of it. He gradually discovers it." - Walter Russell

"Once men died for Truth, but now Truth dies at the hands of men." - Manly P. Hall

A & P Electronic Media

Leading Digital Publishers of Advanced Energy Science

http://emediapress.com

Here are a couple of Eric Dollard's book & video packages...

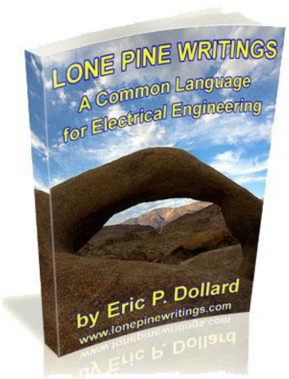
LONE PINE WRITINGS
A Common Language for Electrical Engineering
by Eric P. Dollard
www.lonepinewritings.com

LONE PINE WRITINGS

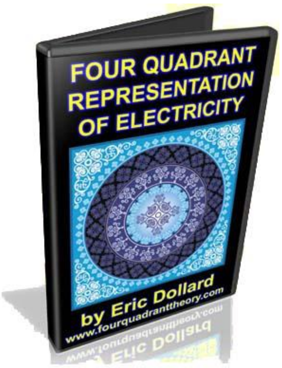
FOUR QUADRANT REPRESENTATION OF ELECTRICITY
by Eric Dollard
www.fourquadranttheory.com

FOUR QUADRANT REPRESENTATION OF ELECTRICITY

VISIT OUR FULL CATALOG OF NEARLY 40 GAME-CHANGING BOOKS & VIDEOS!

http://emediapress.com

JOIN OUR AFFILIATE PROGRAM AND EARN 60% ON ALL OUR DIGITAL PRODUCTS

TESLA
CHARGERS

Customer Service: info@teslachargers.com Phone: (509) 474-0538 Hours: 0900-1700 Pacific Time

Tesla Chargers

World's Most Efficient, Effective & Advanced Battery Chargers and Solar Charge Controllers

Built to Order

http://teslachargers.com

ENERGETIC FORUM

Visit Eric Dollard's Official Forum

Join for FREE and meet others who are building technologies based on Eric Dollard's work!

http://www.energeticforum.com/eric-dollard-official-forum/

Energy Science & Technology Conference

Another * EXCLUSIVE * One-Of-A-Kind Conference
Produced by A & P Electronic Media

Held every year in July in the Coeur d'Alene, Idaho area – next to Spokane, Washington

MEET ERIC DOLLARD IN PERSON!

Join the Free Conference Update Mailing List

http://energyscienceconference.com

SUPPORT ERIC DOLLARD

Please donate to Eric Dollard and EPD Laboratories, Inc so he can continue advancing the Electrical Sciences!

Donate to EPD Laboratories, Inc.
501(c)(3) – Your donations are tax deductible

Make Checks or Money Orders
Payable to: EPD Laboratories, Inc.

EPD Laboratories, Inc.
PO Box 31
Davenport, CA 95017

or

DONATE TO ERIC DOLLARD BY PAYPAL

USE THIS PAYPAL EMAIL: info@epdlabs.org

http://ericpdollard.com/donate-to-eric-dollard/

Crystal Radio Kit 4

Mike's Electronic Parts, LLC

Mike's Electronic Parts

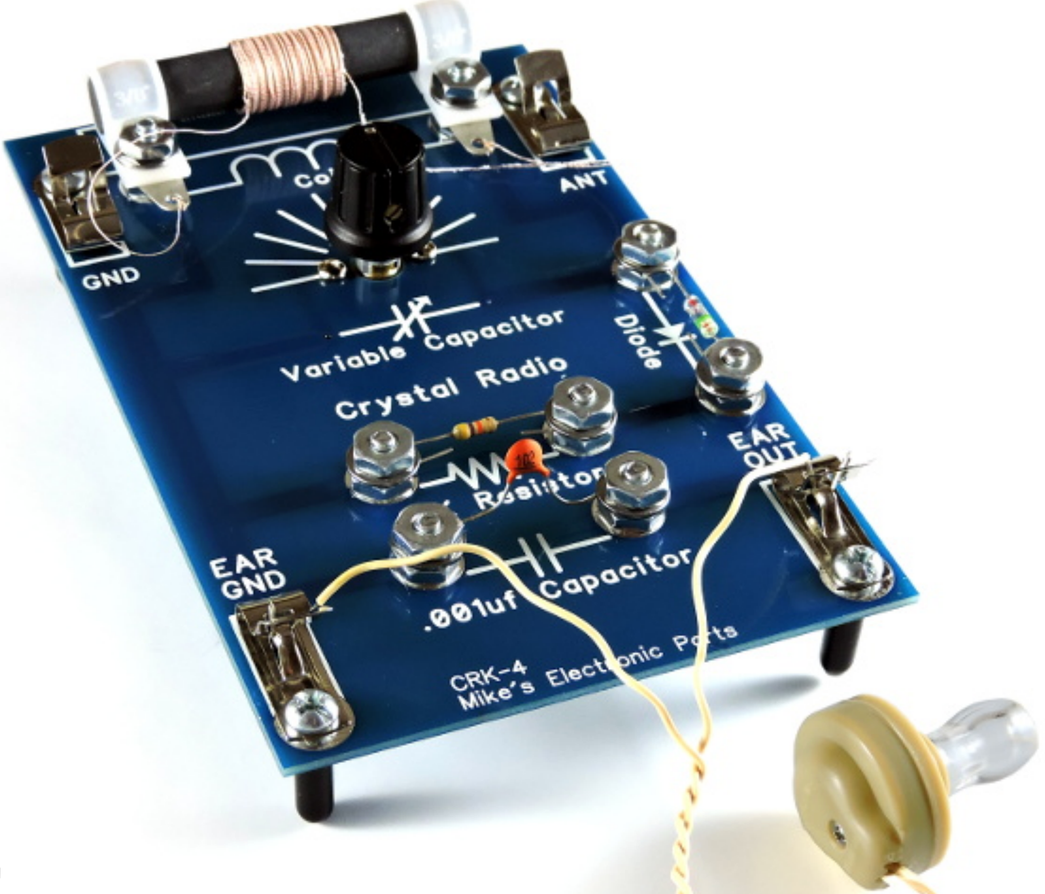
Coil
ANT
GND
Variable Capacitor
Diode
Crystal Radio
Resistor
EAR OUT
EAR GND
.001uf Capacitor
CRK-4
Mike's Electronic Parts

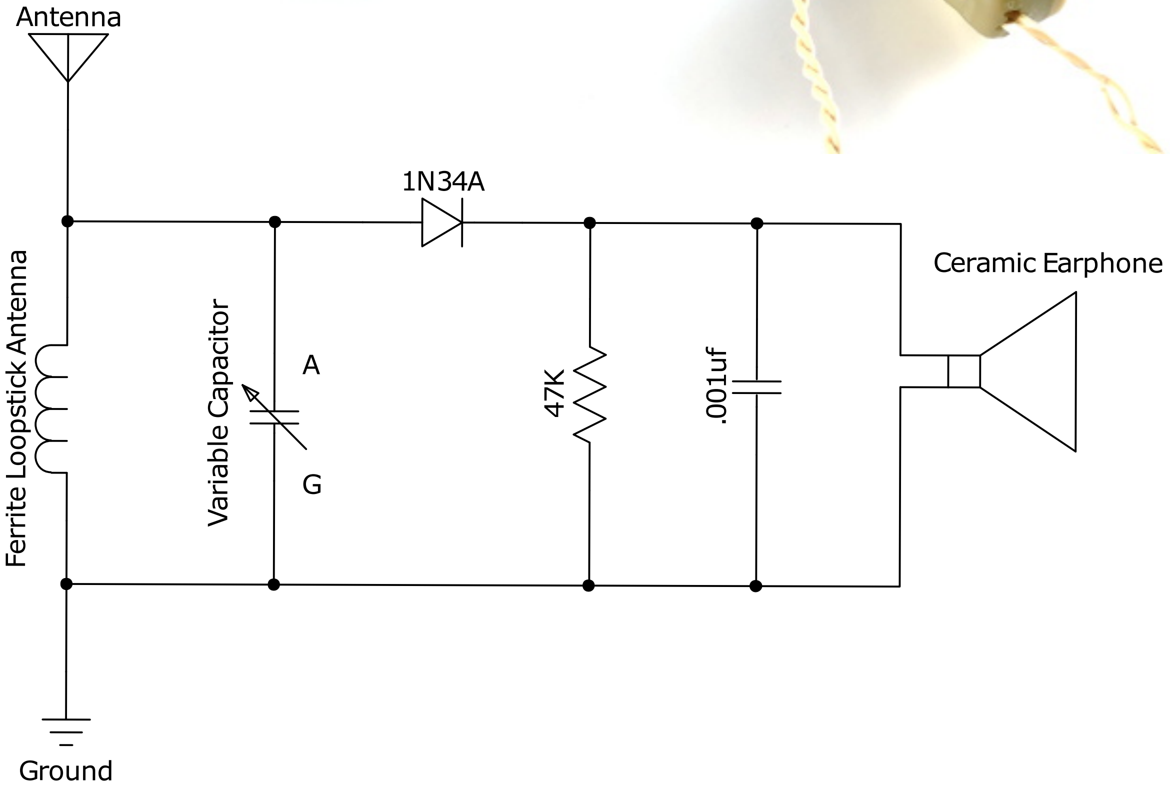
Antenna
1N34A
Ceramic Earphone
Ferrite Loopstick Antenna
Variable Capacitor
A
G
47K
.001uf
Ground

Crystal Radio Kit 4
Mike's Electronic Parts, LLC

Parts List
- 1 - Circuit Board
  - 1 - Variable Capacitor Pre-soldered To Circuit Board
- 1 - Ferrite Loopstick Antenna
  - 2 - Solder Lugs Pre-soldered To Ferrite Loopstick Antenna
- 1 - 1N34A Germanium Diode
- 1 - .001uf Capacitor
- 1 - 47K Resistor
- 1 - Ceramic Earphone
- 4 - Fahnestock Clips
- 1 - Knob with shaft and screw
- 2 - Coil Clamps
- 6 - 3/8" #6 Screws
- 2- 1/2" #6 Screws
- 4 - 3/4" #6 Screws
- 20 - #6 Nuts
- 8 - #6 Washers
- 4 - Rubber Feet
- 1 - ~20 feet antenna wire (color may vary)
- 1 - ~ 5 feet ground wire (color may vary)

Tools Required
Philips Screwdriver
Small (Jewelers) Screwdriver
5/16" Wrench

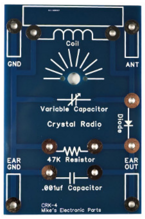
Coil
GND
ANT
Variable Capacitor
Crystal Radio
Diode
47K Resistor
EAR GND
EAR OUT
.001uf Capacitor
CRK-4
Mike's Electronic Parts
Circuit Board

Variable Capacitor

Knob with shaft and screw

Ceramic Earphone

Antenna Wire

Ground Wire

Ferrite Loopstick Antenna

1N34A Germanium Diode

47K Resistor

.001uf Capacitor

Fahnestock Clip

Coil Clamp

Rubber Foot

3/8" #6 Screw

1/2" #6 Screw

3/4" #6 Screw

#6 Washer

#6 Nut

Crystal Radio Kit 4
Mike's Electronic Parts, LLC

Crystal Radio Kit 4 is a great way to get started in electronics and radio building. This is a fun and simple project for both beginners and expert builders alike. Everything needed to build a fully functional radio, including an antenna and ground wire, is included in the kit. Easy to assemble with no soldering required. Crystal Radio Kit 4 is a great project to build with your child or grandchild, boy scout groups and schools.

The kit contains small parts that may be a choking hazard. Adult supervision is always advised while building the radio kit.

Assembly

1. Locate and familiarize yourself with all the parts in the parts list.
2. Assemble the legs.
   Items used: 4 - 3/4" #6 Screws, 4 #6 Nuts, 4 - Fahnestock Clips, 4 - Rubber Feet.
   Insert a 3/4" #6 screw through a Fahnestock Clip. Insert the screw with fahnestock clip through the circuit board. Attach a #6 nut from the bottom side of the circuit board. Snug the nut, do not over tighten. Push a rubber foot over the bottom of the screw. Lightly turn the rubber foot clockwise to ensure it is fully pushed into the screw.
   Repeat for all four legs.

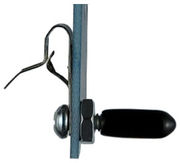

3. Assemble the coil screws.
   Items used: 2 - 1/2" #6 Screws, 2 #6 Nuts.
   Locate the Coil location on the circuit board. Insert a 1/2" screw through the circuit board from the bottom in the hole to the left of the coil schematic symbol. Use a #6 nut from the top to fasten the screw to the circuit board.
   Repeat for the hole to the right side of the coil schematic symbol.

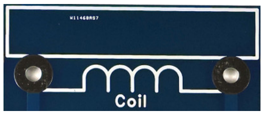

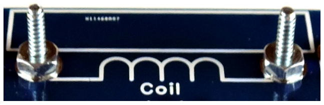

4. Assemble the posts.
   Items used: 6 - 3/8" #6 Screws, 6 #6 Nuts.
   Push a 3/8" #6 screw through the circuit board from the bottom. Use a #6 nut from the top to fasten the screw to the circuit board.
   Repeat for the remaining holes in the circuit board.

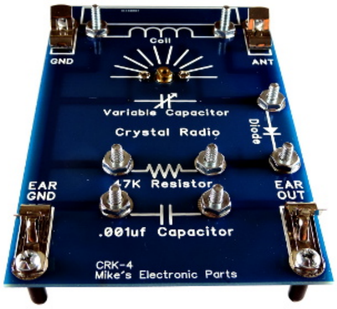

Assembly Continued

5. Assemble the coil.
   Items used: 1 - Ferrite Loopstick Antenna, 2 - Coil Clamps, 2 - #6 Washers, 2 - #6 Nuts.
   Locate the Coil location on the circuit board. Place the solder lugs over the screws at the end of the Coil symbol. Slide a coil clamp over each end of the ferrite rod. Place the coil clamp over the solder lug and secure with a #6 washer and #6 nut at each end.
   Double check that the coil is between the coil clamps and that the wires are not being pinched by the coil clamp.

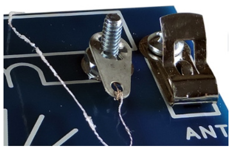

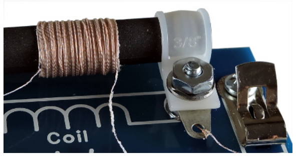

6. Assemble the 47K resistor
   Items used: 1 - 47K Resistor, 2 - #6 Washers, 2 - #6 Nuts.
   Locate the 47K Resistor location on the circuit board. Place the resistor across the nuts. Bend each end of the wire around the screw. Place a #6 washer and #6 nut on top of the wire. Tighten the nut to secure the resistor.

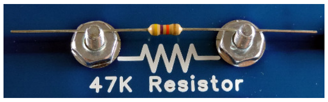

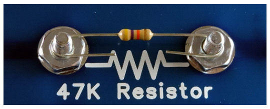

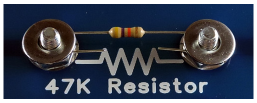

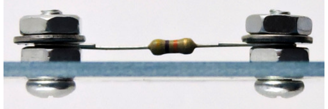

7. Assemble the .001uf Capacitor and 1N34A as described in Step 6 above.

8. Assemble the Knob.
   Attach the Knob shaft using the knob shaft screw to the variable capacitor shaft on the top side of the circuit board. Tighten the shaft screw. Turn the knob shaft to the full left position. Loosen the set screw in the side of the knob. Slide the knob over the knob shaft. Align the dot on the top of the knob with the far left dial markings on the circuit board. Tighten the set screw. The dot on the knob should point to the circuit board markings from full left to full right of the knob rotation.

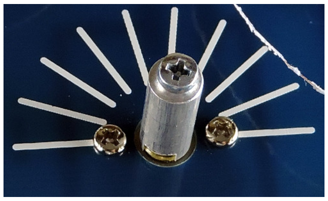

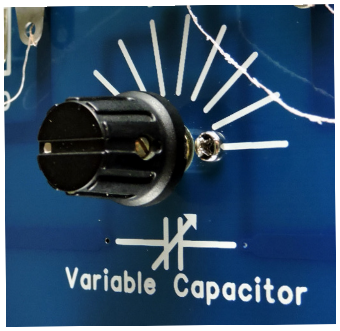

Assembly Continued

9. Attach the earphone.
   Items used: 1 - Ceramic Earphone.
   Locate the Fahnestock clips marked EAR. Push down on the fahnestock clip. Slide the end of the earphone wire through the opening. Release the fahnestock clip. Note: Do not clip the wire by the plactic coating. The bare metal portion of the wire must be clamped by the fahnestock clip.

Antenna

The bare end of the wire connects to the fahnestock clip marked ANT. The antenna wire should be placed as high as feasible. Run the antenna horizontal to the ground in as straight a line as possible. If there is not enough space to run the antenna in a straight line, the wire can be placed in an arc or loop shape as well. Experiment with the antenna placement and shape to achieve the best results.

Do not use the antenna outdoors during bad weather.

Ground

The use of a ground is optional. There are many options for connecting a ground. You can just lay the ground wire along the floor in the opposite direction from the antenna. You can connect the end of the ground wire to a metal object, like a metal desk. You can also connect the end of the ground wire to a physical earth ground if you have access to one. Some metal plumbing in a home may be physically earth grounded for example.

Congratulations! Your Crystal Radio is ready to use.

Use the tuning knob to locate local radio stations. Keep in mind that the crystal radio will only receive strong stations. Unlike a traditional battery powered radio, the power to drive the sound you hear is from the signal the antenna is able to pick up. Experiment moving the antenna to different locations to increase stations and/or volume.

If you have any questions at this stage, please refer to the photo on page 1 and double check any directions and photos applicable.

Jeffrey La Favre June 20, 2015

Crystal Radio Project

Radio receivers capture radio waves broadcast by radio stations and convert those waves into sound that you can hear. For this project you will construct a crystal radio, a simple type of radio receiver that does not require a battery or other power supply. A crystal radio operates only on power that it captures from the radio waves transmitted by the radio station.

The original crystal radios of the early twentieth century utilized special crystalline minerals as detectors, converting the radio waves into the audio signals that can be heard when connected to a headset. These naturally occurring special minerals act as diodes in the crystal radio circuit. Galena is one example of these special minerals (https://en.wikipedia.org/?title=Galena). A device called a "cat's whisker," a spring-loaded wire, is used with the crystal to make the crystal operate as a diode. It can be a tedious exercise to adjust the contact point between the crystal and the cat's whisker so that they operate together as a diode. And without careful handling of the radio, the adjustment can be lost, necessitating a repeat of the frustrating exercise of adjusting the cat's whisker.

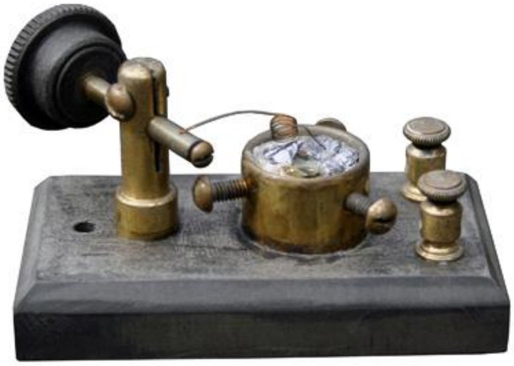

Figure 1 crystal and cat's whisker

With the availability of germanium and silicon-based diodes, it is no longer necessary to use crystals, but this type of radio is still called a crystal radio even when a natural crystalline mineral is not used as the detector. The term crystal radio is now used generally for any radio that operates only on power captured from the radio wave.

There has been a renewed interest in crystal radios over the past 15 years. Some Ham Radio operators have become interested in designing very efficient crystal radios, capable of receiving radio broadcasts from distances well exceeding 1,000 miles. There are contests for DX (distance) crystal radio operators. One contest winner, located in Hawaii, was able to hear a radio station in Cuba with his fancy crystal radio! Using the same radio you will build, I have been able to hear radio stations as far away as Georgia and Iowa, although I "cheated" by connecting an amplifier to my crystal radio (you have to give me a break, I don't hear very well anymore).

Radio waves are of the same class of radiation as light. Both are **electromagnetic radiation**. The distinguishing feature between radio waves and light is the **wavelength** of the radiation. Visible light, the kind humans can detect with their eyes, ranges in wavelength from 400 to 700 nanometers. A nanometer is one-billionth of a meter. Therefore, wavelengths of light are very short. Compared to visible light, the wavelengths of radio waves are very long. The shortest wavelengths of electromagnetic radiation that are considered radio waves are 1 millimeter in length and the longest are 100 kilometers in length (100 kilometers is equal to 62 miles). That is quite a large range in wavelengths.

History of Radio

James Clerk Maxwell, a Scottish physicist, established the theoretical basis for electromagnetic radiation in 1864 (https://en.wikipedia.org/wiki/James_Clerk_Maxwell). That year is a good starting point for this brief history of radio. Twenty-four years later, another physicist, Heinrich Rudolf Hertz, conducted experiments in transmitting radio waves through the air. He is generally credited with providing the experimental evidence for Maxwell's theory. The unit **Hertz** is used for the frequency of radio waves to honor the accomplishments of Heinrich Hertz (https://en.wikipedia.org/wiki/Heinrich_Hertz).

As the news of Hertz's experiments spread, a number of scientists and inventors began to experiment with radio waves. The most well-known was probably the Italian inventor, Guglielmo Marconi, who is credited with building the first successful wireless telegraphy system utilizing radio transmission through the air (https://en.wikipedia.org/wiki/Guglielmo_Marconi). His initial development work with radio waves occurred in the 1890's. By the middle of the decade, Marconi was able to transmit a radio signal up to a distance of 2 miles and over hills.

The signals transmitted by Marconi's equipment were in the form of Morse code, a series of short and long sounds heard as clicks or tones. In order to understand messages sent by this method, the listener needed to learn Morse code, not a trivial exercise. It was desired that the human voice could be transmitted by radio waves so that this technology could be of use to more people. This was accomplished in 1900 by a Brazilian priest named Roberto Landell de Moura.

During the period of 1900 to about 1920, radio grew from a curiosity to a powerful means of communication. At the end of this period, commercial radio stations began broadcasting programming to thousands of listeners. Station 8MK (now known as WWJ) claims to have broadcast the first radio news program on August 31, 1920, from Detroit, Michigan. During this time in radio history many people used crystal radios to listen to these early broadcasts.

The decades of 1920, 1930 and 1940 are often referred to as the **Golden Age of Radio** (https://en.wikipedia.org/?title=Old-time_radio). During the 1950's television replaced radio as the dominant form of news and entertainment delivered electronically. It is important to understand that television broadcasts also utilize radio waves for transmission. Television is a more complex system than radio and was therefore developed after radio.

Crystal radios are very simple in design and construction and were fairly cheap to buy during the 1920's. At the time most people could not afford the expensive radios that utilized vacuum tubes to amplify the sound. In fact, crystal radios are so simple in construction that some people built their own, saving even more money. In this project you will see that building a crystal radio is fairly easy. In addition, you will discover the disadvantages of a crystal radio, which soon caused them to fall from favor with the public when the cost of better radios became affordable (https://en.wikipedia.org/wiki/History_of_radio).

Wavelength and Frequency

There are various ways to characterize radio waves. I have already mentioned wavelength. Perhaps the easiest way to think about radio waves is to imagine that they look like water waves. Of course they are not water waves, this is just a convenient way to think about something we can't see with our eyes. If you drop a stone in a pool of water, a series of waves move across the surface of the water away from the location where the stone entered the water. In a similar manner we can imagine radio waves radiating away from the radio station transmitting antenna. For the water waves we can see that there are places we can name the tops of the waves and there are places we can name the bottoms of the waves. If we measure the distance between the top of one wave and the next, then that is the wavelength. We can also measure the distance from the bottom of one wave to the bottom of the next wave. That will also be the wavelength and will be the same length measured from the top of one wave to the next. In fact, we can measure the wavelength from any position of the wave to the next exact same position.

In the case of radio waves, we don't have a way to actually see them, but we can use instruments to measure the wavelength. In fact, it may be more convenient to use another measure of a radio wave, **frequency**. Frequency is the number of waves that pass a specific point in space within a specified time (usually one second). If we count just one wave passing a measurement point in one second, the frequency is one cycle per second or one H**ertz (Hz).**

Radio waves travel at the same speed as light through air, because they are both electromagnetic radiation. From the work of the famous physicist, Albert Einstein, we know that the speed of light is the ultimate speed limit in the Universe, nearly 300,000 kilometers per second (to be more accurate, 299,792.458 kilometers per second in a vacuum*, i.e.,* in space – in the Earth's atmosphere the speed is slightly slower, about 299,700 kilometers per second).

Suppose we set up an experiment where we could make light circle the Earth at its equator. The circumference of the Earth at the equator is 40,074 km. How many times would the light circle the Earth in one second? The light would travel 299,700 km, or more than 7 times around the Earth in one second (299,700 km/40,074 km = 7.479)!

If we know the frequency of a radio wave, then we can calculate the wavelength. This is possible because we know the velocity of electromagnetic radiation in air (299,700 km/sec or 299,700,000 m/sec). Suppose we measure a radio wave with an instrument and determine the frequency to be 1,100 kHz (1,100,000 Hz). That just happens to be the frequency broadcast by a strong AM radio station in Cleveland, which you will be able to hear with your crystal radio.

What is the wavelength of a 1,100 kHz radio wave? In one second we know the radio wave travels 299,700,000 meters through air. And during that time 1,100,000 waves will pass an observation point. If we could count the number of waves along a distance of 299,700,000 meters, we would count 1,100,000 waves. Therefore, the length of one wave would be 299,700,000 m / 1,100,000 waves = 272 meters per wave (or wavelength). The wavelength of a radio wave of 1,100 kHz frequency is 272 meters.

Amateur radio operators, also known as Ham Radio operators, have permission to transmit radio within a number of radio bands, or small ranges of frequencies. These bands are often referred to by the wavelength in meters. For example, there are amateur bands at 40, 80 and 160 meters wavelength, as well as other bands.

If you want to design a radio to capture a radio broadcast, then it is important to know both the frequency and the wavelength of the radio waves. This will help you design the tuning circuit and antenna required for good reception. For example, a good antenna design is achieved with a long wire, divided in the middle, with a total length of one-half the wavelength of the radio wave. This kind of antenna is named a dipole. There are also other types of antennas that perform well.

In designing the tuning circuit of a radio, it is important to know the frequencies of the radio stations you wish to tune. For example, in the United States, AM radio stations broadcast on frequencies ranging from 540 to 1,610 kHz. If you are listening to a radio station and hear the announcer say: “This is WTAM 1100” then you can understand that the station named WTAM is broadcasting at a frequency of 1,100 kHz.

You will begin your construction of the crystal radio by winding two coils of wire around pieces of plastic pipe. These become the **inductors** of your radio. An inductor is a new category of electronic component that you have not worked with before, but is a very important type of component.

You have already learned about capacitors and have even made a capacitor from aluminum foil and polyethylene sheet for insulation. Recall that capacitors can store electrical energy. This is done by collecting positive and negative charge on the opposing plates of the capacitor, which creates an electric field. Energy is contained in the capacitor by the created electric field.

In some ways an inductor is similar to a capacitor. Inductors can also store electrical energy. In the case of an inductor, the energy is stored in a **magnetic** field. Capacitors store energy in an electric field and inductors store energy in a magnetic field.

When a capacitor is paired with an inductor, a very useful circuit can be designed to tune a specific frequency of radio waves. This circuit is sometimes called a **LC circuit** or **tank circuit**. In the case of the word “tank,” we can think of it as a container for the specific frequency of radio waves. The letters **L** and **C** represent inductance and capacitance. The letter C is obvious but not the letter L. The letter L represents inductance in electrical equations and the unit used is the **Henry**. The inductance of a coil of wire can be measured in Henries. You have already learned that the unit of capacitance is the Farad, and hopefully remember that a one Farad capacitor is a very large capacitor. The same can be said for an inductor. A one Henry inductor is a very large coil of wire. In the case of our crystal radio, it is more convenient to measure the inductance in units of microhenries (µH). There are one million microhenries to one Henry.

The Henry is defined as follows.

> *In an inductor, if the current is changing at a constant rate of one ampere per second, a one Henry inductor will generate a one volt potential difference across the inductor.*

That may be more difficult to understand than the definition of a Farad for capacitance. All you really need to understand is that inductors can store electrical energy in a magnetic field and we can measure the storage capacity in terms of the unit called a Henry.

Capacitors and inductors behave differently when voltage is applied across them. It is this difference that makes them a good pair in a tuning circuit. When voltage is applied across a capacitor, at first, high current flows into the capacitor. As the current charges the capacitor, voltage across the plates rises, which causes a back voltage, reducing the current flow. The exact opposite happens with an inductor. When voltage is first applied to an inductor, there is a nearly instantaneous build-up of back voltage, resulting in a very low initial current flow. As time goes on, the back voltage decreases and the current flow increases, just the opposite of a capacitor. These two behaviors, that complement each other, provide a good tuning circuit. With an inductor of a specific value paired to a capacitor of a specific value, the circuit containing them will allow the **resonance** of only a small range of radio frequencies. This is one method for tuning radios to a specific frequency. By adjusting the value of either the inductor or capacitor (or both), it is possible to change the frequency of resonance in the circuit. In radios it is more common to use an adjustable capacitor to change the tuning. By adjusting the capacitor, different radio stations can be tuned.

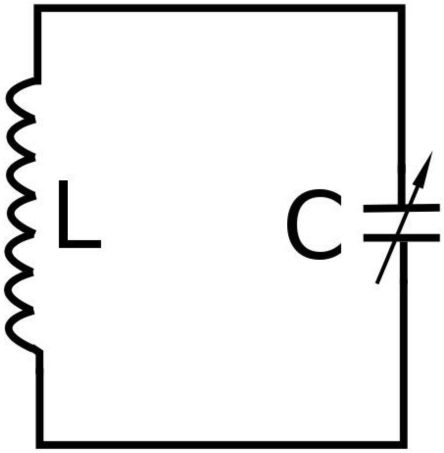

Figure 2 Schematic of an LC circuit

Figure 2 shows the schematic of a LC circuit of the same type you will construct for your crystal radio project. Antenna and ground connections can be added to the circuit (for example connect antenna to top wire between coil and variable capacitor and ground wire to bottom wire between coil and variable capacitor). A connection for a "detector" can be added into the circuit as well. The detector extracts the sound signal from the radio waves. Another way to connect antenna and ground is by the "loose coupling" of another coil. That is the method you will use. The radio signal is transferred from the antenna coil to the coil of the LC circuit without any wire connection. This is made possible by the sharing of a magnetic field between the two coils.

All of these additions can make the circuitry more difficult to understand, so I have just provided you with the tank circuit in Figure 2. Radio signals delivered to the tank circuit from the antenna are in the form of **alternating current**, similar to the current found in your house electrical wiring. The difference is one of frequency. House current has a frequency of 60 Hz, while radio frequencies are much higher (for example, our radio station AM 1100, which has a frequency of 1,100,000 Hz). The purpose of the tank circuit is to provide a resonating circuit for the alternating current of a specific frequency (a specific radio station). At the tuned frequency, the current will make one cycle of traveling first one direction and then the other, matched to the frequency. It is the job of the LC circuit to capture radio waves of a specific frequency and rejecting radio waves of other frequencies. Otherwise, we would have to listen to many stations at one time, talking over each other.

Imagine that we had a circle of wire, cut in one place by inserting a magic mirror. The job of the magic mirror would be to reflect the current flowing in the circuit so that it bounces back and forth, like alternating current. Now suppose the length of the wire is 136 meters (half of 272 meters, the wavelength of a radio wave at a frequency of 1100 kHz). The current in the loop would be in resonance (tuned) to a frequency of 1100 kHz.

Well, we don't have a magic mirror to bounce the current back and forth. Even if we did, having to make a circle of wire 136 meters long would not be very practical. But what if we put a capacitor and an inductor in our circular circuit? The capacitor cuts the circuit, just like our magic mirror. Then we can make a compact circuit that will resonate at the desired frequency. It takes time for the capacitor to charge and discharge, allowing current to flow for a relatively long period of time even in a circuit of short length. The inductor reduces the current flow in the circuit to insure that the current will flow for the proper amount of time to make the circuit resonant. By adjusting the capacitance and inductance of the circuit, current will cycle back and forth at a rate to match a specific frequency. By adjusting the value of a variable capacitor, it is possible to change the frequency of resonance within a certain range. The current cycles back and forth at the same rate as the frequency of the tuned radio signal.

Understanding resonance in an electrical circuit can be difficult. Think about a child on a swing who is being pushed by a parent. Every time the swing comes back, just at the right moment, when the movement backwards stops, the parent gives a push forward to the child. This keeps the child swinging. This is like a resonating circuit. Now suppose the parent gave a push to the child while the swing was still moving backward. That would have a braking effect on the swinging, and with each cycle the swinging would be less and less. This would be the same as a circuit that was not resonant.

Keep in mind that the antenna of a radio intercepts radio waves at many different frequencies. It is the job of the tuning circuit to select a narrow range of frequencies (hopefully just one radio station) and reject all other frequencies. Only the frequencies that resonate well in the tank circuit will be heard. The energy of non-resonant frequencies will be absorbed by the tank circuit and not passed on through the radio.

We would like our crystal radio to tune all of the AM broadcast band, so that we can listen to various AM radio stations. The FCC, an organization that governs radio in the United States, has set the range of frequencies for the AM band (540 to 1610 kHz). You can search for licensed radio stations at their web site where you will also find technical information such as location of the transmitting antenna and the amount of electrical power used in transmitting the signal https://www.fcc.gov/encyclopedia/am-query-broadcast-station-search

This information can be helpful to you while using your crystal radio. Remember, your radio operates only on the power transmitted by the radio station. A station that is close by or one that uses a high amount of power will be easier to hear.

Calculating the Resonant Frequency of a LC circuit

Suppose for a minute that you are assigned to design your own crystal radio. You have the basic knowledge regarding how a LC circuit is constructed (see Figure 2). However, you don't know how many turns your coil should have or what the capacitance range should be for the variable capacitor. There are two formulas that you can use to help with the design of the LC circuit. It is possible to make your own variable capacitor, but most people elect to buy a manufactured one. The capacitor you will use has an adjustable capacitance range of about 25 to 400 pF (0.000025 to 0.000400 µF) which is commonly used for tuning radio frequencies in the AM band. Now that we have established the capacitance, we need to calculate the amount of inductance needed in the coil. The lowest frequency we wish to tune is 540 kHz. Just to make sure our radio will tune this frequency, we can design for a slightly lower frequency, say 538 kHz. We can use a formula to calculate the required inductance. Our formula requires the frequency to be in MHz, not kHz, so we divide 538 kHz by 1,000 to convert to 0.538 MHz. Now we can use the formula below to calculate the required inductance.

Formula for calculating inductance

$$L = \frac{1}{C(2\pi f)^2}$$

where L is the inductance of the coil in microhenries (µH)
f is the frequency in MHz
C is the capacitance of the capacitor in microfarads (µF)

The lowest frequency tuned will be when the capacitor is adjusted to the highest capacitance.

Step one: multiply the frequency by 2π: 6.283 X 0.538 MHz = 3.380

Step two: square the result of step one (i.e., multiply it by itself): 3.380 X 3.380 = 11.42

Step three: multiply the result of step two by the capacitance: 11.42 X 0.000400 µF = 0.004568

Step four: divide 1 by the result of step three: 1/0.004568 = 219 µH

Our coil should have an inductance of 219 microhenries. We can just round that off to 220 microhenries.

If we use a 220 µH coil, then what will the highest frequency tuned be? To make that calculation we can rearrange the formula to this:

$$f = \frac{1}{2\pi\sqrt{LC}}$$

The highest frequency tuned will be when the capacitor is adjusted to the lowest capacitance.

Step one: multiply the inductance by the capacitance: 0.000025 µF X 220 µH = 0.0055

Step two: take the square root of the product of step one: sqrt 0.0055 = 0.07416

Step three: multiply the result of step two by π: 3.142 X 0.07416= 0.2330

Step four: multiply the result of step three by 2: 2 X 0.2330 = 0.466

Step five: divide 1 by the result of step five: 1/0.466 = 2.15 MHz or **2,150 kHz**

Therefore, we can expect our radio to tune from 538 kHz to 2,150 kHz if we use a 220 µH coil, more than enough to cover the AM broadcast band.

Now we need another formula to help us construct a 220 µH coil:

$$L = \frac{(d^2 n^2)}{18d + 40I}$$

where L = the coil inductance in microhenries
d = the diameter of the coil in inches
n = the number of turns of the coil
I = the length of the coil in inches

You will wind 22 gauge magnet wire around a piece of PVC pipe to make the coil. The outside diameter of the pipe is 3.5 inches. Technically, we should add the diameter of the wire to the diameter of the pipe, but we will not do this because the wire diameter is very small compared to the pipe diameter (the value d is the diameter of the coil, from the center of the wire thickness on one side of the coil to the center of the wire thickness on the opposite side of the coil).

There are three independent variables in the formula above (d, n and I) that we must know in order to calculate the inductance. We already know the diameter (d), which leaves two independent variables. It would be helpful if we could establish the value of one more independent variable. What we really want to know is the number of turns required for the coil. Therefore, we need to know the value for I, the length of the coil. We are using 22 gauge wire, which has a diameter of 0.0253 inches. But that does not account for the thin coating of enamel used for insulation or the small amount of space between windings that usually occurs when winding the coil. When I wound the wire around the pipe, I found that each turn of the coil had a width of 0.0278 inches. We can use that number to calculate the length of the coil (I). We will use an iteration process to find the number of turns of wire required for a 220 µH coil. Suppose we start by finding the inductance of a coil with 40 turns:

$$L = \frac{3.5^2 x\, 40^2}{(18\, x\, 3.5) + (40\, x\, 1.11)}$$

If we have a coil of 40 turns, the length of the coil will be 0.0278 inches X 40 = 1.11 inches, the value of l.

By solving the above equation we find that a coil of 40 turns will have an inductance of 182 µH. We need a coil with 220 µH inductance, so it will need more than 40 turns. Let us try 45 turns:

$$L = \frac{3.5^2 x\, 45^2}{(18\, x\, 3.5) + (40\, x\, 1.25)}$$

A coil with 45 turns will be 45 X 0.0278 inches = 1.25 inches long. When we solve the above equation we find that a coil of 45 turns will have an inductance of 220 µH. Perfect! Our coil must have 45 turns to create an inductor with 220 µH. Now you are ready to make your coil!

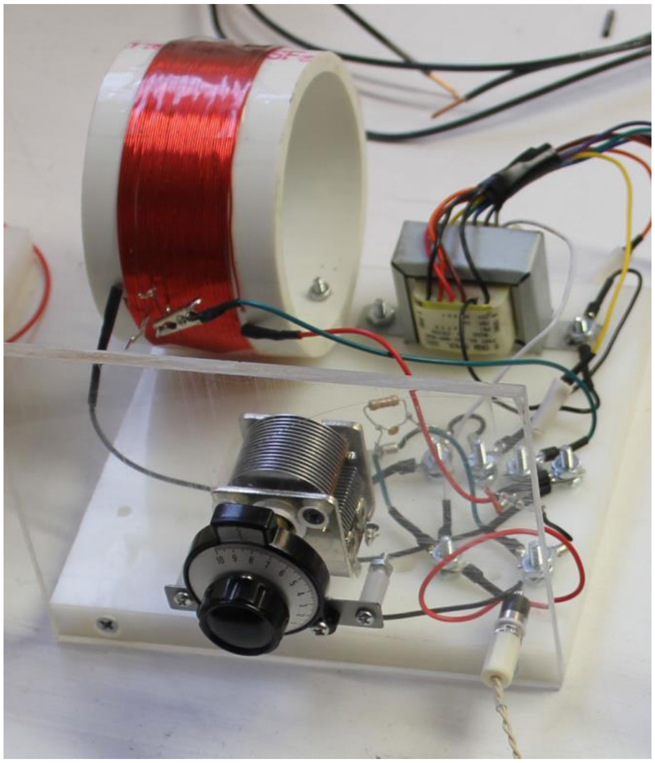

Figure 3 Crystal Radio

Crystal Radio Project

Figure 3 is a photo of the main part of the crystal radio that you will construct. The coil constructed with red wire is located at the back of the radio on the left side. The variable capacitor is located at the front of the radio and has a knob attached for adjusting the capacitance. You should notice that there are some more parts of the radio. You will learn about these in subsequent lessons.

Winding the tuning coil

In order to facilitate your project, I have already cut the wire for the coil for you. I made some calculations and measurements first. Your coil will contain two sections, one of 28 turns and one of 17 turns (45 turns total). You will construct a "tap" where the two coil sections meet. In this case the detector of your radio will be connected to this tap (I will explain why later).

You will use 22 gauge magnet wire for the coil. This wire has a very thin coating of enamel (paint) for insulation. The insulation is needed to prevent shorting of the coils since the turns of wire will touch each other when you tightly wind the coil. The diameter of 22 gauge wire is 0.0253 inches, which will be a factor in calculating the amount of wire needed to make one turn on the pipe. To get a good estimate of the wire needed for one turn, we can add the diameter of the wire to the diameter of the pipe: 3.5" + 0.0253" = 3.5253". Then we can multiply the diameter by pi to find the circumference, which is the same as the length of wire needed: π X 3.5253" = 11.075". I tried another method to determine the amount of wire needed for one turn. I actually wrapped a piece of wire around the pipe and cut it carefully so that the amount of wire was just enough to make one turn. I measured this amount of wire and found it was about 11 3/32" long, which is the same as 11.094" long. My calculated value was 11.075" and the actual measurement was 11.094". There is not much difference between these two, but we should use the larger number so that we will have enough wire. You will need to have some extra wire at the ends of the coil to make connections so we will add 2 extra inches.

28 turn coil: 28 turns X 11.094 inches = 310.63 inches + 2 inches = about **312.5 inches**

17 turn coil: 17 turns x 11.094 inches = 188.60 inches + 2 inches = about **190.5 inches**

Before you start winding wire on the pipe, you need to remove the enamel insulation from the four ends of the two pieces of wire. You can do this with a piece of sandpaper. The insulation is red colored, so you should be able to tell when it has been removed. Make sure you remove ALL insulation for a length of about one inch from the end of the wire. If you leave some insulation on the wire, you may not be able to solder connections later (solder won't stick to the enamel insulation).

I have already drilled holes in the coil form (pipe) for mounting the coil on the radio base and for inserting the ends of the wire. At each end of the 45 turns of wire, the wire is threaded into a hole in the pipe and then back out through another hole near the first hole (make sure you use the holes that are paired, not the single holes that will be used to mount the coil). After you have threaded one end of the wire through the holes, so that about ½ inch of the wire end is protruding from the pipe, start wrapping the wire around the coil core (Figure 4). To do this, hold the pipe with threaded wire end to the right (Figure 4). The wire that you are wrapping around the pipe should sit on the top side of the pipe and you should rotate the pipe toward you (make the top side of the pipe rotate toward you). While you are wrapping the wire, place your thumb on top of the wire and apply some pressure. This will keep the wire windings tight. After you have a few turns done, make sure that the turns of wire are all touching (squeeze them together if there is space between the wire coils). The turns of wire should be tight in each wrap and also right next to each other.

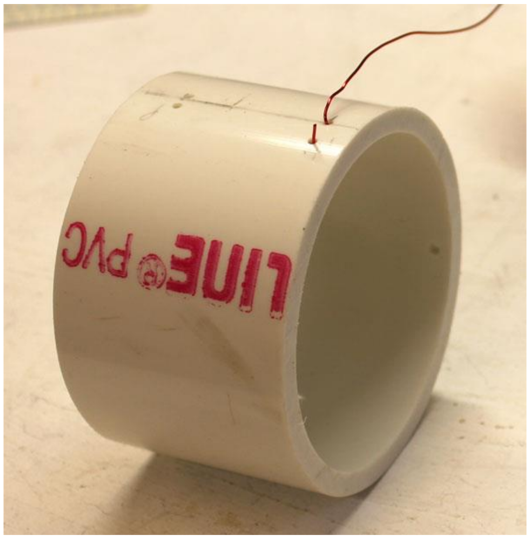

Figure 4 preparing the coil form for winding

After you have wrapped the 312.5 inches of wire around the pipe 28 turns, you should have about one inch of wire remaining. Using some packing tape, tape down the coil to prevent it from unwinding.

Take one end of the 190.5 inch piece of wire and twist it with the end of the wire you just coiled on the pipe. The length of the twisted part should be about ¼ inch. This will form the tap on your coil. If a soldering iron is available, solder the twisted wires together. The twisted wires should point out from the coil so that you can attach a wire later.

Continue wrapping the coil in the same direction. The second piece of wire should be enough to make an additional 17 turns. After you have made 17 turns past the tap, there should be very little wire left. If there is not enough wire to reach the pair of holes on the pipe, then unwind the coil by one turn. If you unwind one turn there will be plenty of extra wire. It will not make much difference if the coil has 44 or 45 turns, just a slight difference in the tuning range. Thread the wire end through the two holes on the pipe as provided, just as you did for the start of the coil. Make sure the wire of the last coil is tight after you thread the wire through the holes so that the coil does not loosen up. You can apply a couple of pieces of clear packing tape over the wire coils to keep them from spreading across the pipe. It is best if you can't see the white color of the pipe between the turns of wire. Or at least keep the open space between turns to a minimum. Congratulations! You have finished wrapping your tuning coil. Construction of the remainder of your radio will be covered in other lessons soon to follow.

"Amateur Wireless" Handbooks

CRYSTAL RECEIVING SETS

And How to Make Them

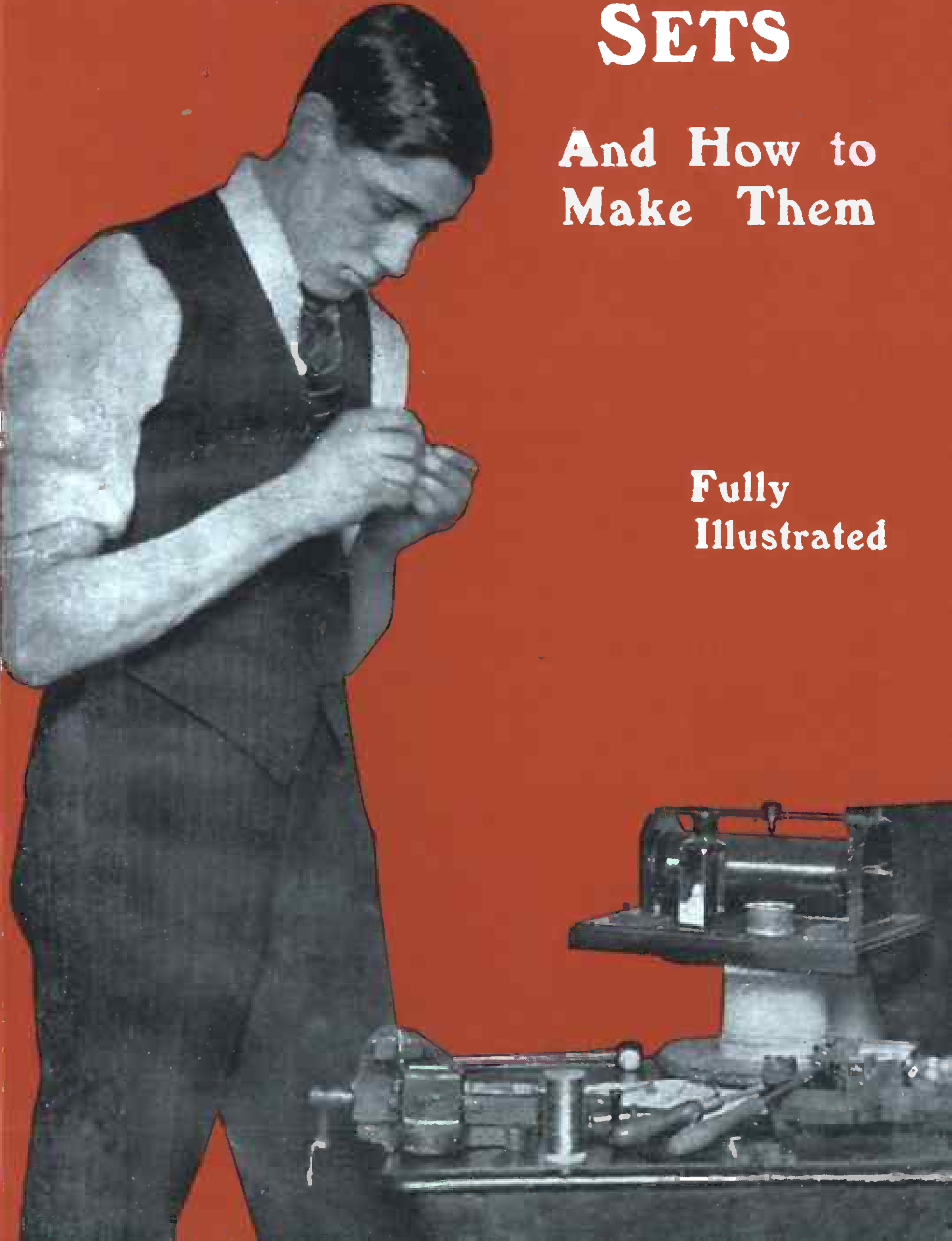

Fully Illustrated

The Best Paper for Wireless Amateurs

Amateur Wireless
And Electrics

The Leading Weekly

Devoted exclusively to the interests of Wireless Amateurs; explains the Underlying Principles; Makes the Way Plain to Beginners; Shows them How to Start; What to Do; What to Buy; What to Make, and How to Make It. Gives Expert Replies to Readers' Questions; no matter what the difficulty is, the Reader is helped out of it by Highly Skilled Experts; Wireless Club and Society Fixtures; and the Latest News of Wireless Developments. Fully illustrated by Helpful Diagrams and Photographs.

Every Thursday
3D.

ON SALE EVERYWHERE

Cassells, Publishers, London.

SIMPLE CRYSTAL RECEIVING SETS

AND HOW TO MAKE THEM

EDITED BY
BERNARD E. JONES
Editor of "Amateur Wireless," etc. etc.

WITH 114 ILLUSTRATIONS

CASSELL AND COMPANY, LTD
London, New York, Toronto and Melbourne

PREFACE

THIS Handbook deals, in a simple, straightforward manner, with the making of a number of different kinds of crystal sets representative of those in present-day use. It is uniform in style and price with other members of the "Amateur Wireless" Handbook Series, notably with two of them, namely "Wireless Component Parts and How to Make Them" and "Simple Valve Receiving Sets and How to Make Them," the former containing detailed instructions on the making of the various components forming part of many kinds of wireless receiving apparatus, and the latter treating of valve sets in exactly the same way as crystal sets are dealt with in this present Handbook.

This book has been compiled from the pages of "Amateur Wireless," the well-known weekly journal for wireless amateurs, and the name of the author responsible for each of the various chapters is given in many cases immediately below the chapter heading.

It is believed that the information is presented in such simple fashion as not to cause the reader any trouble when putting it into practice, but should he be in difficulty with regard to any matter dealt with in this book, or to any other subject within the scope of the weekly journal mentioned, he should send a query to "Amateur Wireless," La Belle Sauvage, London, E.C.4, *but should take particular care to accompany his query with a stamped and addressed envelope and a coupon from the current issue of that journal.*

THE EDITOR.

CONTENTS

CHAPTER PAGE
1. A Simple and Efficient Receiving Set . . . 1
2. A Single-slider Set 16
3. Set with Semicircular Tuner 25
4. Crystal Set with Tapped Single Coil . . 33
5. A Loose-coupled Set 45
6. Set with Plug-in Basket Coils 58
7. Combined Crystal and Valve Receiver . . 68
8. Some Miniature Receiving Sets 73
9. Crystal Circuits 85
10. How Crystals Work 91
11. Making a Buzzer 99
12. Receiving C.W. Signals on a Crystal Set . 103
13. Making a Wavemeter from a Crystal Set . 108
14. Converting Low-resistance Phones . . . 111
15. The Morse Code 121
Index 124

SIMPLE CRYSTAL RECEIVING SETS AND HOW TO MAKE THEM

CHAPTER I

A Simple and Efficient Receiving Set

As the result of recent experiments conducted in the United States of America to produce an efficient crystal receiving set that at the same time would be reasonably simple and inexpensive to make, a circular was issued by the American Bureau of Standards giving constructional details of a set evolved in accordance with the result of those experiments. The set, it is

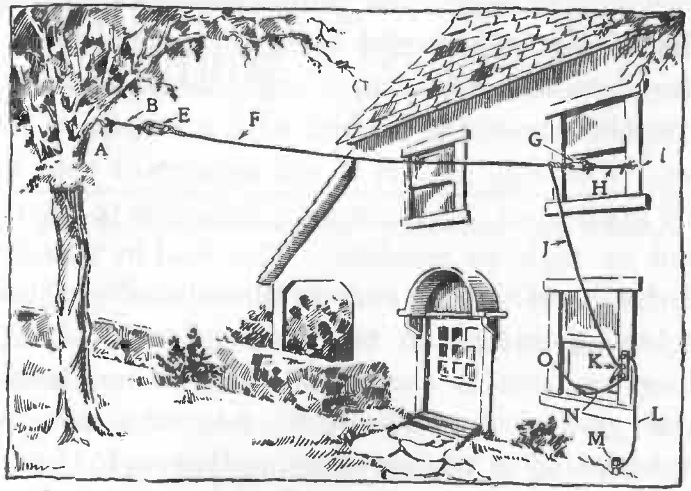

Fig. 1.—General View of the Aerial and its Supports and Connections.

thought, forms a fitting introduction to this book. It has been made by thousands of people, including hundreds of readers of "AMATEUR WIRELESS," and its efficiency is well proved by ample practical experience.

Elements of the Set.—In this receiving set there are five essential parts: the aerial, lightning switch, ground or earth connection, tuner and detector, and telephone receivers, and the construction of each (excluding the last-named, which, of course, must as a rule be purchased) is described under its proper heading.

The details regarding the aerial and earth hold true for any set, but it must be remembered that the aerial, both as regards type and system of erection, is largely dependent upon individual circumstances.

The Aerial.—The aerial is simply a wire suspended between two elevated points, and should not be less than 30 feet above the ground with a length of about 75 feet. (See Fig. 1.) It is not important that it be strictly horizontal. It is, in fact, desirable to have the far end as high as possible. The lead-in wire from the aerial itself should run as directly as possible to the lightning switch. If the position of the adjoining buildings or trees is such that the distance between them is greater than about 85 feet, the aerial can still be held to a 75-foot distance between the insulators by increasing the length of the piece of rope (B) to which the far end is attached. The rope (H) securing the aerial insulator to the house should not be lengthened to overcome this difficulty, because by so doing the lead-in wire (J) would be lengthened.

The parts will be referred to by reference to the letters appearing in Figs. 1 and 2.

A and I are screw eyes sufficiently strong to anchor the aerial at the ends. B and H are pieces of rope $\frac{3}{8}$ in. or $\frac{1}{2}$ in. in diameter, just long enough to allow the aerial to swing clear of the two supports. B is a piece of $\frac{3}{8}$-in. or $\frac{1}{2}$-in. rope sufficiently long to make the distance between E and G about 75 feet. A single-block pulley may be used, but is not really essential.

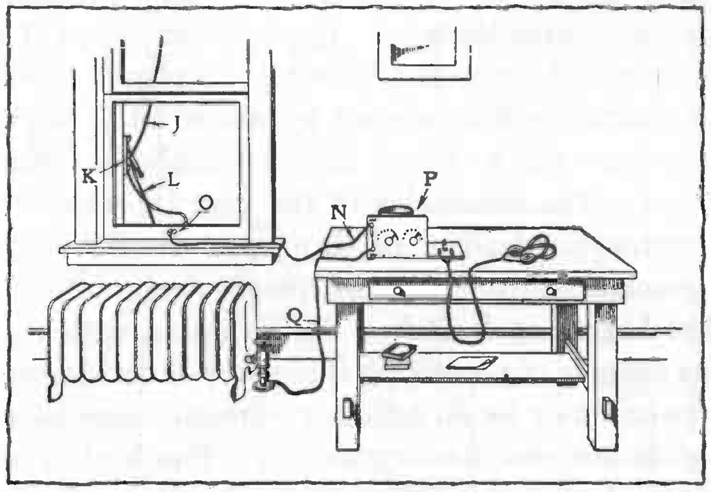

Fig. 2.—The Receiver and its External Connections.

E and G are two insulators which may be constructed of any dry hardwood of sufficient strength to withstand the strain of the aerial; blocks about $1\frac{1}{2}$ in. × 2 in. × 10 in. will serve. The holes should be drilled, as shown in Fig. 1, sufficiently far from the ends to give proper strength. If wood is used the insulators should be boiled in paraffin wax for about an hour. If porcelain wiring cleats are available they may be substituted instead of the wood insulators. If any unglazed porcelain is used as insulators, it should be boiled in paraffin wax in the same way as the wood. The wire for the aerial may be No. 14 or 16 copper wire either bare or insulated. The end farthest from the receiving set may be secured to the insulator (E) by any satisfactory method, care being taken not to kink the wire. Draw the other end of the wire through the other insulator (G) to a point where the two insulators are separated by about 75 ft.; twist the *insulator* (G) so as to form an anchor as shown in Fig. 1. The remainder of the wire (J) which now constitutes the lead-in or drop-wire should be just long enough to reach the lightning switch.

The Lightning Switch.—The lightning switch K is for the purpose of protecting the system from lightning. This switch may be an ordinary porcelain-base, single-pole double-throw battery switch. The lead-in wire (J) is attached to this switch at the middle point. The switch blade should always be thrown to the upper clip when the receiving set is not actually being used and to the lower clip when it is desired to receive signals.

The Earth Lead.—L is the earth wire for the lightning switch; it may be a piece of the same wire as used for the aerial, of sufficient length to connect the upper clip of the lightning switch (K) to the clamp on the earth rod (M), which is a piece of iron pipe or rod driven 3 ft. to 6 ft. into the ground, preferably where the ground is moist, and extending a sufficient

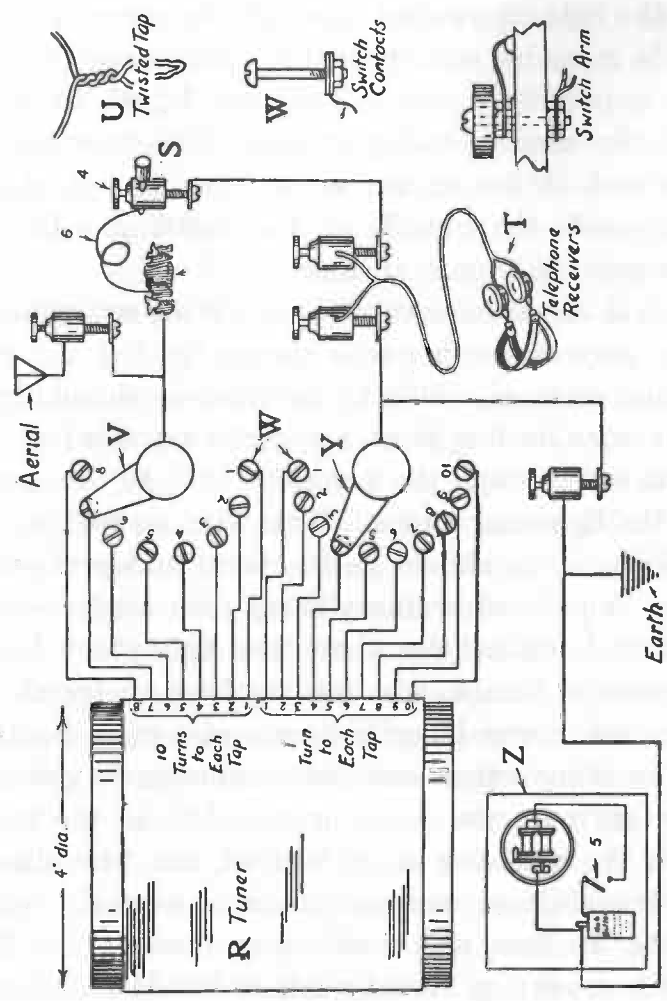

Fig. 3.—The Receiving Set in Complete Detail.

distance above the ground in order that the ground clamp may be fastened to it.

The Lead-in.—N is a wire leading from the lower clip of the lightning switch through the porcelain tube (O) to the receiving set terminal marked aerial (Fig. 3).

O is a porcelain tube of sufficient length to reach through the window casing or wall. This tube should be mounted in the casing or wall so that it slopes down towards the outside of the building. This is done to prevent ingress of rain.

General Arrangement.—The receiving set installed in some part of the house is shown by Fig. 2. P is the actual receiver, which is described in detail later. N is the wire leading from the aerial terminal of the receiving set through the porcelain tube to the lower clip of the lightning switch. This wire, as well as the wire shown by Q, should be insulated and preferably flexible. A piece of ordinary hemp cord might be unbraided and utilised for these two important leads. Q is a piece of flexible wire leading from the receiving set terminal marked earth to a water-pipe, heating system or some other metallic conductor to ground. If there are no water-pipes or radiators in the room in which the receiving set is located, the wire should be run out of doors and connected to a special earth below the window, and which may consist of a few yards of wire netting buried a few inches deep in damp soil, or an iron rod as mentioned.

Crystal Detector.—The construction of a crystal detector (S, Fig. 3) may be very simple although quite satisfactory. The crystal, as it is ordinarily

purchased, may be unmounted or mounted in a little block of metal. For mechanical reasons the mounted type may be more satisfactory, but that is of no great consequence. It is very important, however, that a good crystal be used. It is probable also that a galena crystal will be the most satisfactory.

The crystal detector may be made up of a tested crystal, three wood-screws, a short piece of copper wire, a nail, setscrew type of terminal, and a wooden knob or cork. The crystal is held in position on the wooden base by three brass wood-screws, as shown in Fig. 4. A bare copper wire may be wrapped tightly round the three brass screws for contact. The assembling of the rest of the crystal detector is quite clearly shown in Fig. 4. If, however, the reader wishes to adopt any other type of detector, he will find almost every type illustrated and explained in "Wireless Component Parts, and How to Make Them," uniform in style and price with the present volume.

Telephones.—It is desirable to use a pair of telephone receivers connected by a head band, usually called a double telephone head-set (T, Fig. 3). The telephone receiver may be of any of the standard commercial makes having a resistance of between 2,000 and 3,000 ohms. The double telephone receivers will cost more than all the other parts of the station combined, but it is desirable to get them, especially if one intends to improve the receiving set later. A single telephone receiver with a head band may be used, though it gives results somewhat less satisfactory.

Accessories.—Under the heading of accessory equipment may be listed terminals, switch-arms, switch-contacts, test-buzzer, dry battery, and boards on which to mount the complete apparatus. The terminals, switch-arms and switch-contacts may all be readily purchased. There is nothing peculiar about the pieces of wood on which the equipment is mounted, except that they should be soaked in melted paraffin wax for an hour or two.

Winding the Tuner.—Having obtained a piece of cardboard tubing 4 in. in diameter and about $\frac{1}{2}$ lb. of No. 24 (or No. 26) d.c.c. copper wire, start the winding of the tuner (see R, Fig. 3). Punch two holes in the tube about $\frac{1}{2}$ in. from one end, not shown in the diagram. Weave the wire through these holes in such a way that the end of the wire will be quite firmly anchored, leaving about 12 in. of the wire free for connections. Start with the remainder of the wire to wrap the several turns in a single layer about the tube, tightly and closely together. After ten complete turns have been wound on the tube, hold those turns snugly while a tap is taken off. This tap is made by forming a 6-in. loop of the wire and twisting it together at such a place that it will be slightly staggered from the first tap. This method of taking off taps is shown quite clearly at U (Fig. 3). Proceed in this manner until seven twisted taps have been taken off at every ten turns. After these first seventy turns have been wound on the tube, then take off a 6-in. twisted tap for every succeeding single turn until ten additional turns have been wound on the tube.

After winding the last turn of wire, anchor the end by weaving it through two holes punched in the tube, much as was done at the start, leaving about 12 in. of wire free for connecting. It is to be understood that each of the eighteen taps is slightly staggered from the one just above, so that the several taps will not be bunched along one line on the cardboard tube.

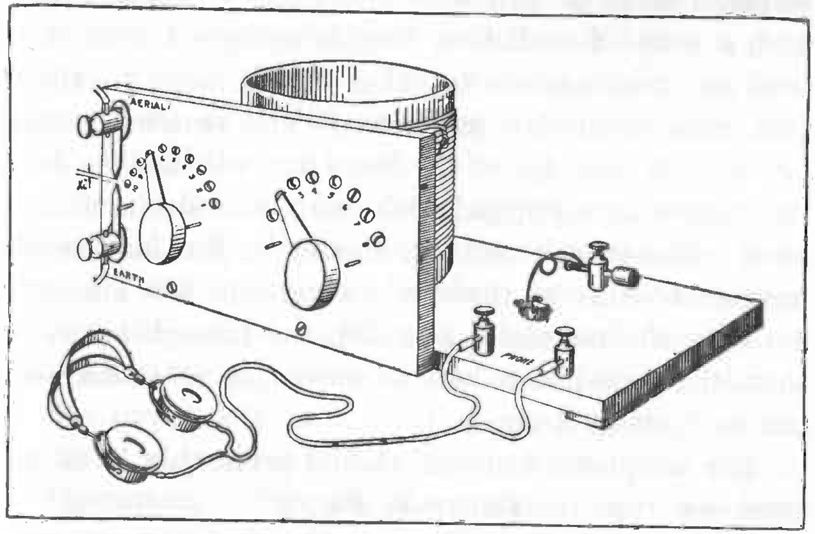

Fig. 4.—View of the Receiver.

It will be advisable, after winding the tuner, to dip it in hot paraffin wax.

Upright Panel and Base.—Having completed the tuner to this point, set it aside and construct the upright panel shown in Fig. 4. This panel may be a piece of wood approximately $\frac{1}{2}$ in. thick. The positions of the several holes for the terminals, switch-arms and switch-contacts may first be laid out and drilled. The aerial and earth terminals may be ordinary $\frac{1}{8}$-in.

brass screws of sufficient length and supplied with three nuts and two washers. The first nut binds the bolt to the panel, the second nut holds one of the short pieces of stiff wire, while the third nut holds the aerial or earth wire, as the case may be. The switch-arm with knob (shown in Fig. 3) may be purchased in the assembled form, or it may be constructed from a thin slice cut from a broom-handle and a bolt of sufficient length equipped with four nuts and two washers together with a narrow strip of thin brass somewhat as shown. The switch contacts (w, Fig. 3) may be of the regular type, or they may be brass bolts equipped with one nut and one washer each. The switch-contacts should be just sufficiently near each other so that the switch-arm will not drop between the contacts, but also far enough apart so that the switch-arm can be set so as to touch only one contact at a time.

The telephone terminal should preferably be of the set-screw type, as shown in Fig. 3.

Wiring.—Having constructed the several parts just mentioned and mounted them on the wood base, the next matter is to connect the several taps to the switch-contacts and attach the other necessary wires. Scrape the cotton insulation from the loop ends of the sixteen twisted taps as well as from the ends of the two single-wire taps coming from the first and last turns. Fasten the bare ends of these wires to the proper switch contacts as shown by the corresponding numbers in Fig. 3. Be careful not to cut or break any of the looped taps. It would be preferable to fasten the connecting wires to the switch-contacts by binding them between the washer and the nut, as shown at w (Fig. 3). A wire is run from the back of the terminal marked earth (Fig. 3) to the back of the left-hand switch-arm bolt (Y), thence to underneath the left-hand binding post marked phones in Fig. 4. A wire is then run from underneath the right-hand terminal marked phones to underneath the binding post w (Fig. 3), which forms a part of the crystal detector. Take a piece of No. 24 bare copper wire about 2½ in. long and twist one end tightly round the nail passing through binding post s; its other end will rest gently by its own weight on the crystal. The bare copper wire which has been wrapped tightly round the three brass wood-screws holding the crystal in place is led to and fastened at the rear of the right-hand switch-arm bolt (v), thence to the upper left-hand binding post marked aerial. As much as possible of this wiring is shown in Fig. 3.

Operation.—After all the parts of the set have been constructed and assembled the first essential operation is to adjust the little piece of wire, which rests lightly on the crystal, to a sensitive point. This may be accomplished in several different ways; the use of a miniature buzzer transmitter is very satisfactory. Assuming that the most sensitive point on the crystal has been found by the method described later, the rest of the operation is to get the set in resonance or in tune with the station from which one wishes to hear messages. The tuning is attained by adjusting the inductance of the tuner. That is, one or both of the switch-arms are rotated until the proper number of turns of wire of the tuner are made a part of the metallic circuit between the aerial and earth, so that, together with the capacity of the aerial, the receiving circuit is in resonance with the particular transmitting station. It will be remembered that there are ten turns of wire between each of the first eight switch contacts and only one turn of wire between each two of the other contacts. The tuning of the receiving set is best accomplished by setting the right-hand arm (v., Fig. 3) on contact (1) and rotating the left-hand switch-arm over all its contacts. If the desired signals are not heard, move the right-hand switch-arm to contact (2) and again rotate the left-hand switch-arm throughout its range. Proceed in this manner until the desired signals are heard.

The Test Buzzer.—As mentioned previously, it is easy to find the most sensitive spot on the crystal by using a test buzzer (z, Fig. 3). The test buzzer is used as a miniature local transmitting set. When connected to the receiving set, as shown at z, Fig. 3, the current produced by the buzzer will be converted into sound by the telephone receivers and the crystal, the loudness of the sound depending on what part of the crystal is in contact with the fine wire.

In order to find the most sensitive spot connect the test buzzer to the receiving set as directed, close the switch 5, Fig. 3, set the right-hand switch-arm on contact point No. 8, fasten the telephone receivers to the binding posts marked phones, loosen the set-screw of the binding post slightly and change the

position of the fine wire (6, Fig. 3) to several positions of contact with the crystal until the loudest sound is heard in the 'phones, then tighten the binding post set-screw (4) slightly.

The illustrations to this chapter were published by " Popular Radio " and prepared by them from the drawings issued by the U.S.A. Government, and the Editor of the present Handbook wishes to express his indebtedness for their use to the American publication named.

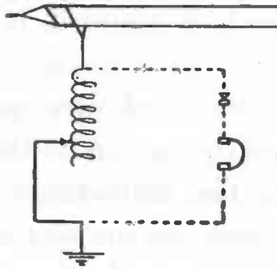

Fig. 5.—Diagram of Oscillatory Circuit.

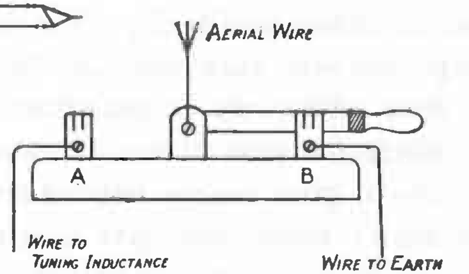

Fig. 6.—Details of Lightning Switch and Connections.

The Range of the Set.—It is impossible to state offhand what the range of a set will be, for there are different factors that govern the operation. Briefly these are as stated below :

(1) The distance of the transmitting station and its direction.

(2) The power used by the station ; this, of course, varies enormously.

(3) The height and length of the receiving aerial.

(4) The position of the receiving station. If it is on a hill or a level plain, and is not sheltered by trees

or neighbouring houses, these conditions confer great advantages.

(5) The efficiency of the receiving set. This depends on (*a*) good insulation, (*b*) proper adjustment of detector, and (*c*) good telephones.

Hints on Operation.—In order that signals may be received, it must be remembered that the receiving set must be in " tune " with the vibrations in the ether set up by the transmitting station. To tune-in stations of different wave-lengths either the inductance or capacity, or both, of the oscillatory circuit (see Fig. 5) must be altered. The method adopted in a simple set is to vary the inductance.

Now inductance necessitates a length of wire, preferably in the form of turns ; two turns, provided they are close, give nearly four times the inductance of one turn ; three turns give nine times the inductance, and so on. Also the inductance increases in proportion to the square of the diameter of the turns in the same way. By having, say, 100 turns of wire in a former (each turn being close to its neighbour but well insulated from it), and having a slider, we can vary the inductance over a wide range, increasing by a single turn at a time. The slider has certain disadvantages. It wears the wire by constant rubbing and is liable to short-circuit two or more turns, which will waste some of the energy. In order to overcome these objections, and also to make an inductance suitable for panel mounting, the turns are brought out to two switches, one called the " tens switch " and the other the " units." Each stud on the tens switch

has ten complete turns connected to it, so that by moving the knob of this switch the inductance is varied by comparatively large amounts at a time. To tune more closely the units switch is used, ten studs being connected to the last ten turns on the coil.

By using both switches together it is possible to include any number of turns varying from one to the greatest number on the coil.

The lightning switch is shown enlarged by Fig. 6. A single-pole change-over switch is used. The aerial wire is connected to the arm. The lead to the tuning inductance is connected to A and the earth wire to B. When the set is in use the switch should be over to the left (A), and when out of use it should be over to the right (B), thus earthing the aerial.

CHAPTER II

A Single-slider Set

By PAUL D. TYERS.

Remembering the statement in the previous chapter that the longer the wave-length the greater is the amount of inductance required to tune it in, it will be obvious that it is impossible to wind up one coil to receive all wave-lengths. The range of the set here described is from about 350 metres to 600 metres.

Construction of the Tuner.—The coil will be wound on a cylinder 6 in. long and 3 in. in diameter. Cardboard postage tubes are very suitable for the purpose if they are previously prepared. A 6-in. length is cut off with a saw and the surface and ends are carefully smoothed with very fine glass-paper. Three small holes are drilled in each end of the tube in the position shown in Fig. 7. The tube is then heated in an oven to drive out any absorbed moisture, and while it is still hot it is given a coat of thick shellac varnish. This is prepared by dissolving shellac in methylated spirit. The tube should be heated and varnished at least three times.

Wooden ends are fixed to the cylinder, which serve as a support for the slider, as follows: Two wooden discs about $\frac{3}{8}$ in. thick are cut out with a fretsaw of such size as tightly to fit into the ends of the tube. This will be understood by reference to Fig. 8, and

it will be seen that they are afterwards secured by small brass screws passed through the sides of the tube. The end pieces are made from any hardwood, and should be about $4\frac{1}{2}$ in. square and $\frac{1}{2}$ in. thick. These are fixed to the discs by brass screws after the coil has been wound.

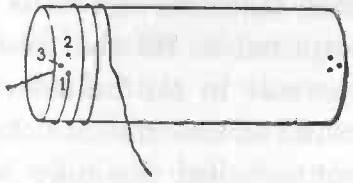

Fig. 7.—Method of Fixing Winding.

Winding the Coil.—When winding a very large coil it is usual to arrange some form of winding apparatus, but it is unnecessary in this case. Referring to Fig. 7, about 6 in. of free end of the wire is passed down hole No. 1, up hole No. 2, and down hole No. 3, the end being put through the loop which has been formed inside the tube between holes No. 1 and No. 2. It will be found that this will fix the end of the wire quite firmly. The winding is then started, making each turn close against the other. It is essential to keep the wire fairly tight, but not so tight as to strain it. When the other end of the coil is reached the wire is fastened off in a similar manner to the first

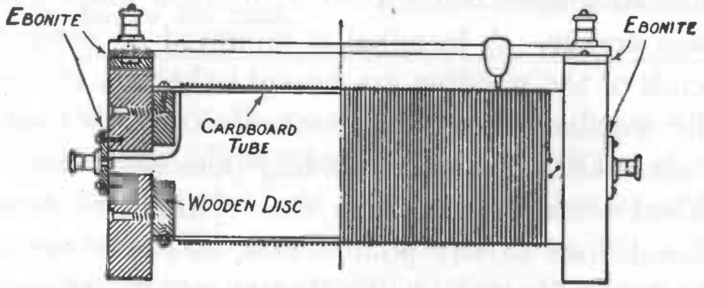

Fig. 8.—Half Section of Tuning Coil.

turn. The best wire is double silk-covered, but this, of course, is rather expensive, and cotton- or enamel-covered may be substituted. It should be remembered that enamelled wire is thinner than silk-covered, and therefore if this is employed more turns will be required to fill the winding space, with a consequent increase in the inductance of the coil. It is not very usual to use silk- or cotton-covered wire with a sliding contact, but this may be done if the winding is given a thin coat of shellac.

Mounting the Coil.—Although a connection is only required to one end of the winding, it is useful to bring both ends to terminals for use in other circuits. A hole about $\frac{3}{4}$ in. in diameter is made in the centre of each wooden disc and end piece. Two squares of ebonite about $1\frac{1}{2}$ in. long are placed over the holes in the end pieces, being fixed with small brass screws at each corner. A terminal is mounted on each, and the ends of the winding are brought through the holes in the wooden disc, being connected to the back of the terminals. Thus the winding does not come into electrical contact with either the cylinder or wooden ends and discs at any point; this, of course, ensures the best possible insulation. Having reached this stage the end pieces can be screwed to the wooden discs, thus firmly fixing the coil.

The Sliding Contact.—The slider may be made if desired, but as it is a standard detail it can be purchased cheaply. It consists of a piece of moulded ebonite containing a spring and brass plunger which makes contact on the winding of the coil. Usually

the slider is made to fit a $\frac{1}{4}$-in.-square brass rod, as shown in the diagram. A piece of square brass rod is cut about 7 in. long (that is, the length of the tube and end pieces), and a hole is drilled in each end to take a small terminal screw. Two pieces of ebonite are next prepared. These should be made about 1 in. long and $\frac{1}{8}$ in. thick. The width is that of the end piece, the ebonite being fixed to this as shown in Fig. 9. It will be noticed that a hole is drilled in the middle of the ebonite, a small part of the end piece being cut away to accommodate the head of the terminal screw. The screw is put through the ebonite, passed through the hole in the square brass, and the two parts of the terminal are screwed up tightly. The same process is repeated at the other end of the square brass, and the pieces of ebonite are then screwed to the ends.

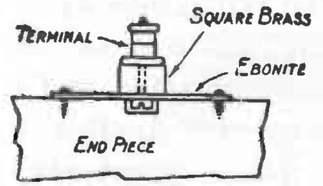

Fig. 9.—Method of Fixing Slider Rod to End Piece.

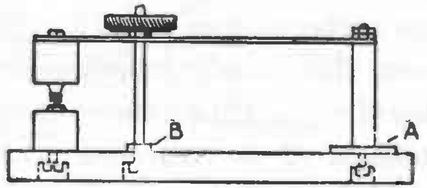

Fig. 10.—Simple Detector.

Two terminals are connected to the slider rod, whereas only one is needed. This makes the appearance of the coil symmetrical, and the additional terminal may sometimes be of use in different circuits.

The Contact Line.—The coil is now complete with the exception of the coil contact line; where the slider touches the winding the insulation must be removed. This is best accomplished by rubbing it with a piece of very fine glass-paper, using the edge of a straight piece of wood as a guide, so that a neat line of bare copper wire appears. Only just sufficient insulation should be removed to ensure the slider making an efficient contact.

The Crystal Detector.—There are a number of different forms of detectors (see Chapter II of the "Amateur Wireless" Handbook "Wireless Component Parts"), the most stable type probably being the carborundum and steel-plate combination. However, this usually requires a potentiometer, and it is therefore a little more complicated. Another type consists of a very fine wire (called a cat-whisker) which bears lightly on the surface of the crystal. This combination is perhaps a little more sensitive than others, but at the same time it is rather erratic in behaviour. The set here described has a perikon detector.

The detector will be understood by reference to Fig. 10. The base may be of ebonite, 4 in. by 3 in. and $\frac{1}{2}$ in. thick. The crystal cups may be drilled or turned from brass rod $\frac{1}{2}$ in. in diameter, or they may be bought for a few pence. The success of the detector depends upon its rigidity; a weak or springy detector will never keep sensitive for more than a few minutes, and it will be found a source of everlasting annoyance. One crystal cup is fixed near the end of the base, as shown in Fig. 10, and the other is screwed to the end of a piece of brass $2\frac{1}{2}$ in. long, $\frac{1}{2}$ in. wide, and a little less than $\frac{1}{16}$ in. thick.

A thick brass screw about 2 in. long is passed through a hole drilled at the other end of the base. A strong washer (A, Fig. 10) is put on the screw, together with a length of stout brass tube about $\frac{3}{8}$ in. external diameter. This is followed by the brass strip, which is finally secured by two nuts. The length of the brass tube will depend, of course, upon the height of the crystal cups. A similar but longer screw is fixed to the base with a nut B, and the end is allowed to project through a slit in the brass arm. A nut to fit the screw is soldered to a small piece of thin brass, which is then screwed to an ebonite knob. It will be understood that by screwing the knob downwards the two cups will be brought nearer together, thus varying the pressure between the crystals. The crystals are set in the cups with either molten solder or Wood's metal. It is usual to employ a pointed piece of bornite or copper pyrites resting on a flat face of a piece of zincite. Two terminals mounted on the base complete the detector; connection is made, of

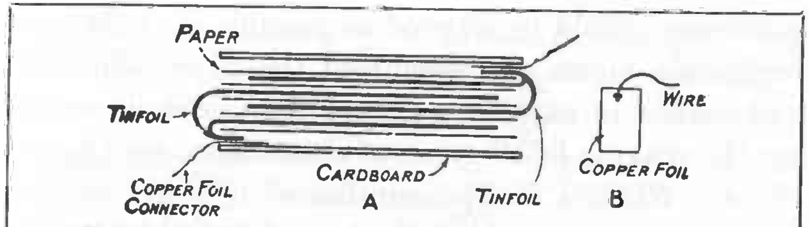

Fig. 11.—Diagram showing Construction of Telephone Condenser.

course, to the fixed cup and the screw holding the upright.

The Telephone Condenser.—This can be made very simply from tinfoil and paper. The actual value of the condenser is not of very great importance, so that the following details need not be adhered to very strictly. Eight pieces of tinfoil 3 in. by $1\frac{1}{2}$ in. and nine pieces of thick writing paper $2\frac{1}{2}$ in. by $2\frac{1}{2}$ in. will be required. Fig. 11 illustrates the method of assembling the condenser, only four plates being shown for clearness. It is essential that the insulation of the condenser should be as good as possible, and therefore before the plates are assembled the paper should be well soaked in paraffin wax. (For full details regarding the making of all types of condensers, see Chapter IV of " Wireless Component Parts.") When the tinfoil plates are assembled, the four plates which project on one side are bent on to the top piece of paper, and the other four are bent on to the bottom piece of paper. The length of the projecting pieces should be a little over half an inch.

Two pieces of waxed cardboard are cut to the same size as the paper to act as a kind of cover. Connection has yet to be made to the plates, and this is best done as follows: Two short lengths of copper wire are soldered to two small pieces of copper foil. These are placed one against each of the bent pieces of tinfoil, being slipped between the top and bottom pieces of paper and the cardboard. This should be quite clear from the diagram. To ensure an efficient contact the condenser is pressed well together, and it is then firmly bound round with some tough brown paper which is secured with a little gum. To serve as a protection against damp, the finished condenser is painted with hot wax, great care being taken to cover the parts where the copper foil emerges from the paper. If desired, the condenser may be placed in a box, when the wires would be brought to two terminals mounted on a piece of ebonite on the top.

Connecting the Apparatus.—Fig. 12 is divided into two parts. One part shows how the apparatus appears when it is connected up, and the other part shows the standard way of indicating the various components. It will be noticed that a switch has been included between the telephones and the detector. It is so arranged that either of two detectors may be used at will. This is a very convenient method of working, as, should one detector suddenly become insensitive in the middle of an important signal, it is only necessary to put the switch in the other position, thereby connecting up the second detector. As soon as the signal

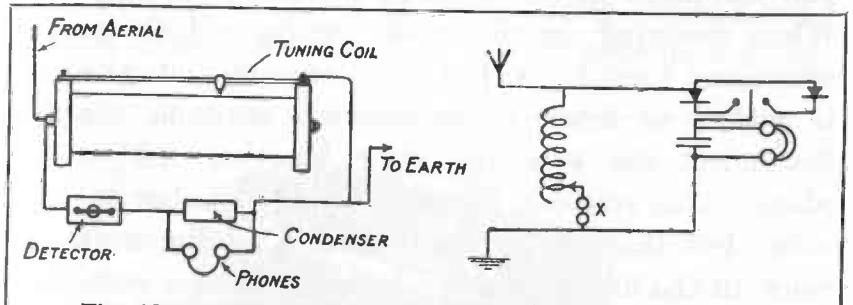

Fig. 12.—Details of Connections showing Arrangement of Apparatus and Wiring Diagram.

has finished, it is possible immediately to readjust the faulty detector. The switch is also very useful for comparing two different types of detectors, since it is possible to change from one to the other without disconnecting any of the apparatus.

Using Loading Coils.—The tuning coil, as previously stated, is designed to receive wave-lengths up to about 600 metres. However, if it is desired to receive longer waves, all that is necessary is to insert two terminals in the position marked x in Fig. 12. When receiving on the small tuning coil these are connected together with a piece of wire, but when it is wished to listen to longer-wave stations, simply disconnect the wire and insert another coil in its place. This coil can be made exactly similar to the other, but there is no need to fix a sliding contact, since all the fine tuning can be accomplished with the original coil. It is only necessary to tap the coil at varying intervals, bringing the tappings to a multipoint switch.

A coil of this description is usually called a loading coil, since it loads the aerial to the required wave-length.

CHAPTER III

Set With Semicircular Tuner

By O. J. RANKIN.

THE tuning inductance of the set described in this chapter (see Fig. 13) has been specially designed for the efficient reception of broadcasting, "dead-end" effects being obliterated as far as possible by only winding on enough wire to just cover the broadcasting wave-lengths.

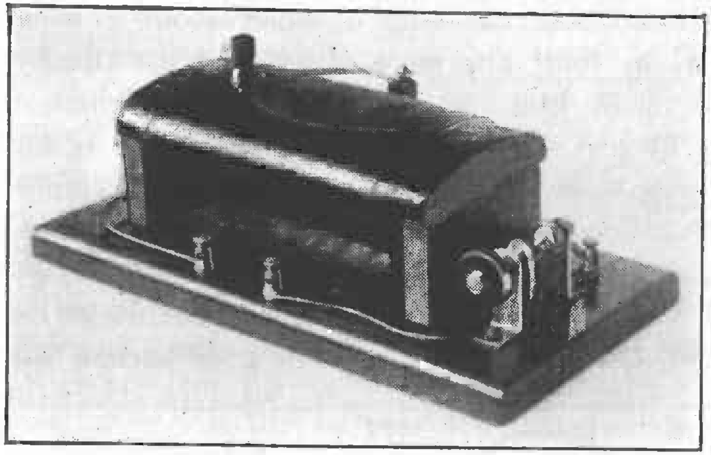

Fig. 13.—Photograph of Set with Semicircular Tuner.

Materials.—The materials and parts required to build this receiver comprise one cardboard cylinder or "former," 6 in. long by 4 in. diameter, ½ lb. bobbin of No. 24 (or nearest) enamelled copper wire, one baseboard $\frac{1}{2}$ in. thick, 14 in. long and 8 in. wide, one piece of board $\frac{3}{4}$ in. thick, $3\frac{7}{8}$ in. wide and $7\frac{1}{2}$ in. long, two pieces of board each about $\frac{1}{2}$ in. thick, $3\frac{7}{8}$ in. wide and 3 in. long, one strip of wood about $\frac{1}{2}$ in. square by $3\frac{1}{2}$ in. long, one piece of spring brass (hard-rolled brass) 4 in. long by $\frac{3}{8}$ in. wide, also one piece $1\frac{1}{2}$ in. long by $\frac{1}{2}$ in. wide, two strips of $\frac{1}{8}$-in. brass $1\frac{3}{4}$ in. long by $\frac{1}{2}$ in. wide, one crystal cup and crystal (preferably hertzite or permanite), one small terminal with nut, four medium-size wood-screw terminals, length of 2 B.A. screwed rod and six nuts, two ebonite or wooden knobs, one spring washer, some good shellac varnish, about one yard of connecting wire and systoflex, and a quantity of small nails, screws, and washers and other sundries.

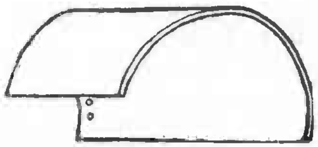

Fig. 14.—Former for Coil.

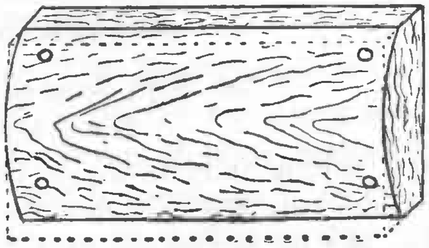

Fig. 15.—Wooden Support for Former.

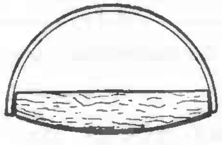

Fig. 16.—End View of Former.

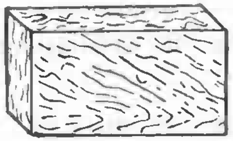

Fig. 17.—End Support.

Inductance Coil.—Cut the cardboard cylinder lengthwise, as shown by Fig. 14, and put aside the other half, as this will not be required. Take the thick piece of board, $3\frac{7}{8}$ in. wide by $7\frac{1}{2}$ in. long, and shape it as shown in Fig. 15. Fit this to the inside of the semi-cylinder, as shown in Fig. 16, and secure it with a few small brass nails driven in so that the heads are flush with the cardboard. See that the straight edges of the cardboard are flush with the curved edge of board, and that an equal portion of board is left projecting at each end. Give this a coat of shellac varnish inside and out, and when perfectly dry apply a second coat. When this is quite dry screw on the two end supports (Fig. 17), which are the two pieces of board $3\frac{7}{8}$ in. wide by 3 in. long, as shown in Figs. 18 and 19. Attach the end of the wire from the bobbin to a small screw T (Fig. 19), apply another coat of shellac, and while this is still "tacky," wind on the wire, taking care that there are no spaces between the layers, particularly where it covers the top curved portion. Having finished it, leave about six or eight inches of wire at the end, and pass it through the holes shown in the cardboard in Fig. 14 to prevent it

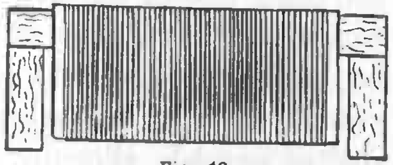

Fig. 18

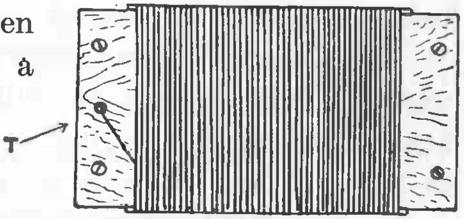

Fig. 19

Figs. 18 and 19.—Elevation and Plan of Coil.

unwinding. Apply shellac freely over the winding and allow the coil to dry.

Next take the piece of spring brass 4 in. long by $\frac{3}{8}$ in. wide, round off the corners with a smooth file, and drill a hole in each end (see Figs. 20 and 21), one of which must take a length of the 2 B.A. rod and the other a round-headed wood-screw. Cut off about 1 in. of the screwed B.A. rod and fit one end into a suitable wood or ebonite knob, as shown in Fig. 22. Bend the other end of the strip slightly, so that it is parallel to the projecting part of the rod. The latter should be made quite smooth and rounded off to a blunt point. Give the baseboard at least three coats of shellac varnish, allowing each coat to dry thoroughly before applying the next.

The Detector.—The construction of the crystal detector should present no difficulties. The two $\frac{1}{8}$-in. brass strips should be cut out and drilled as shown in Figs. 23 and 24, and bent at right angles at the dotted lines. The shank of the terminal fitted to one crystal cup (Fig. 25) should slide freely in the slot (Fig. 24). This small terminal should, of course, be fitted before mounting the crystal.

Take the piece of spring brass $1\frac{1}{2}$ in. long by $\frac{1}{2}$ in. wide as shown by Fig. 26, solder the fine wire feeler or "cat's whisker" U (Fig. 27) to one end and bend to the shape shown. Solder the other end to the base of the piece of brass (Fig. 23). The hole at the top of the pillar should be tapped 2 B.A. An alternative method is to drill the hole slightly larger than the diameter of the detector screw and solder on a nut

previously filled with clay to prevent the solder adhering to the thread. The detector screw (Fig. 28) is self-explanatory. A suitable wood or ebonite knob is clamped tightly to a short length of the 2 B.A. rod with two nuts, as shown, and the end of the rod rounded off to a point. The general arrangement of the detector is clearly shown in Fig. 29.

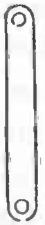
Fig. 20

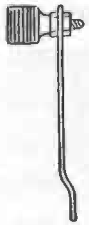
Fig. 21

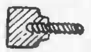
Fig. 22

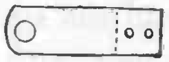
Fig. 23

Figs. 20 to 22.—Details of Contact.

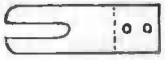
Fig. 24

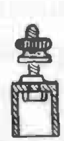
Fig. 25

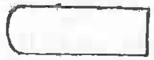
Fig. 26

Figs. 23 to 28.—Details of Detector.

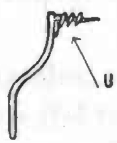

Fig. 27

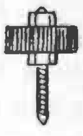
Fig. 28

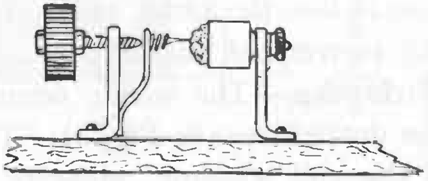
Fig. 29. Detector Assembled.

Assembling.—Place the mounted coil on the centre of the baseboard, move it along to the left about $1\frac{1}{2}$ in., see that everything is quite square, and with a pencil mark all round the bases of supports. Drill four holes large enough to take the screws, and screw the two parts firmly together as shown in Fig. 30. Take the strip of $\frac{1}{2}$-in.-square wood, place it at the back and on a line with the centre of coil, and attach it to the base in a similar way. The correct position of this, which is the support for the inductance switch arm, is shown in Figs. 30 and 31. Now screw down the switch-arm with a medium-size round-headed wood screw and the spring washer. The point must make good contact with every turn of wire on the coil, and to get it perfect it may be found necessary to bend the arm slightly one way or the other. Rotate the arm across the coil several times so as to make a distinct mark across it, as shown in Fig. 31, then with a piece of emery-paper wrapped round any pointed instrument carefully remove the enamel along that line, leaving the bare copper to make contact with the point on the switch arm. Screw the detector to the base and place the aerial, earth and telephone terminals in any convenient position.

Wiring-up.—The wiring connections are indicated by the dotted lines in Fig. 31. Take the end left over from the coil winding, measure the distance to the aerial terminal marked A, remove the enamel, clean thoroughly with emery-paper, and wind it once round the shank of the terminal. Insulating sleeving should be slipped over all connecting wires. Connect another short length of wire from the aerial terminal to the detector pillar which holds the crystal cup, placing the end under one of the screws or soldering it direct to any convenient part of the pillar. This latter

method is always preferable. From the detector pillar carrying the detector screw connect a wire to the nearest phone terminal, from the other phone terminal to the earth terminal marked E, and from E to the

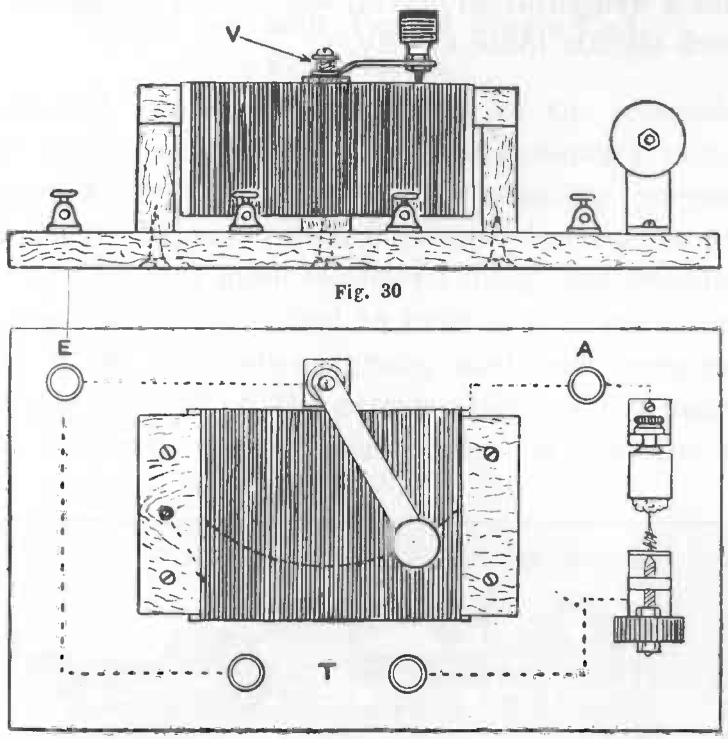

Fig. 30

Fig. 31

Figs. 30 and 31.—Elevation and Plan of Complete Receiver.

spring washer V (Fig. 30), which, of course, should be soldered. It may be necessary to add that the other end of the coil winding terminates at E.

It will be easily understood that the point on the switch arm represents the end of the winding in the sense that only the turns of wire between the connected end of the coil and the switch-arm contact are actually being used. By rotating the switch-arm the inductance is made variable.

This receiver is extremely simple to operate, and with a well-adjusted crystal will receive telephony and spark signals quite clearly.

CHAPTER IV

Crystal Set with Tapped Single Coil

By A. J. COOPER.

Single-coil Tuner.—A photograph of the completed tuner is shown by Fig. 32. The secondary circuit detailed later is not essential for receiving purposes and can be dispensed with if desired, but its use permits of finer and more selective tuning, and therefore the constructor is advised to include it in his instrument. The bar-switch, studs, and two terminals shown in Fig. 32 on the extreme right of the instrument illustrate that portion of the arrangement re-

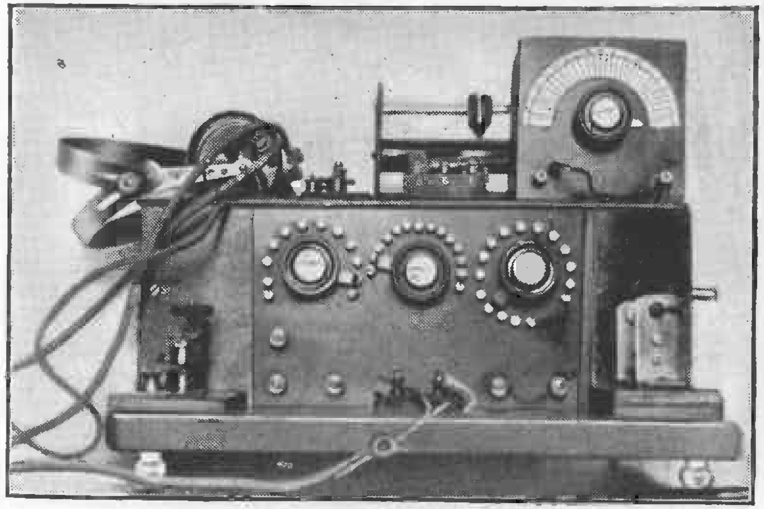

Fig. 32.—Photograph of Complete Tuner.

lating to this secondary circuit, and will be dealt with later on.

Should the constructor decide to embody the secondary circuit in his set he should prepare the panel for its reception according to the sketches, etc., as to do so later will necessitate dismantling the set.

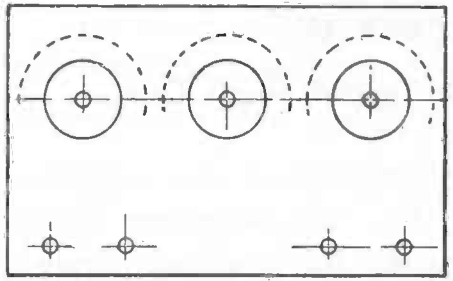

Fig. 33.—Lay-out of Ebonite Panel.

The material detailed does not include the material for the closed circuit, which necessitates an additional 12 studs, 2 terminals and bar-switch.

Panel.—The panel which carries the tuning switches

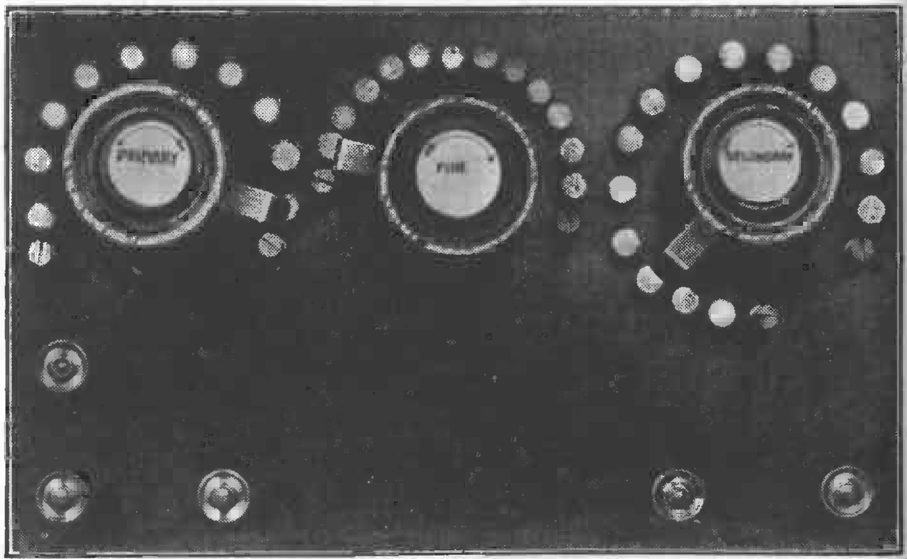

Fig. 34.—Photograph of Panel and Tuning Switches.

will first be dealt with, and a sketch showing the lay-out is given in Fig. 33.

First take the ebonite and smooth off the rough edges to the dimensions specified, that is, 8¾ in. by 5¼ in. Prepare a sheet of paper marked with the required dimensions and glue it to the ebonite. The holes required for two sets of 10 contact studs, etc., can now be drilled. The sizes of the holes and distances apart are not given, as these must be bored to the dimensions of the material obtained. The positions of the contact studs will vary according to the area of their bearing surface, but they should be placed so that the bar-switch cannot fall between them. When boring tapping holes for screw-threads a drill must be put through of the dimension of the inside of the thread for which the hole is intended. Care must be taken when working on ebonite not to exert too much pressure, as the material is very easily broken

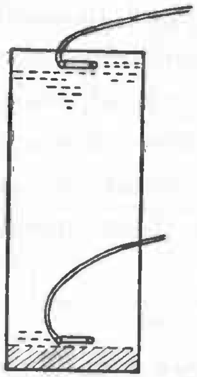

Fig. 35.—Method of Securing Wire.

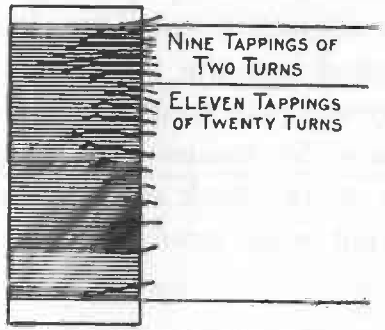

Fig. 36.—Details of Tappings.

and also the ebonite might be torn away on the under side, leaving an ugly, uneven edge. When all the holes are drilled, those which are intended for screw-threads should be tapped according to the size for which they are intended.

A photograph of the completed panel is shown in Fig. 34, and this incorporates the bar-switch and studs and terminals for the secondary coil. After the insertion of the contact studs and terminals (the threads of which should previously be dipped in shellac) in their respective positions, the bar-switches should be mounted and the tip of the switches adjusted so as to give equal contact on the centre of each stud. Ensure, by means of pinning or soldering, that the nuts on the back of the panel on these bar-switches will not work loose.

Materials for Instrument and Aerial

¼ lb. No. 28 s.w.g. d.c.c. wire.

A cardboard cylindrical former 3½ in. in diameter and 5 in. long. (Should a former of these dimensions not be procurable a Horlick's Malted Milk dummy will meet requirements.)

2 bar-switches with two nuts and spring washer.

Small quantity of shellac varnish.

Crystal cups, crystals, and sundry brass-strip and screws.

2 terminals with large bearing surfaces between the nuts.

2 binding-post terminals (for telephones).

1 pair of high-resistance phones (total resistance of 4,000 ohms).

1 fixed telephone condenser .003 microfarads.

1 sheet of ebonite 8¾ in. by 5¼ in.

20 contact studs.

Wood for case, quantity specified in diagram, Fig. 37, ⅝ in. thick (this wood should be very well seasoned).

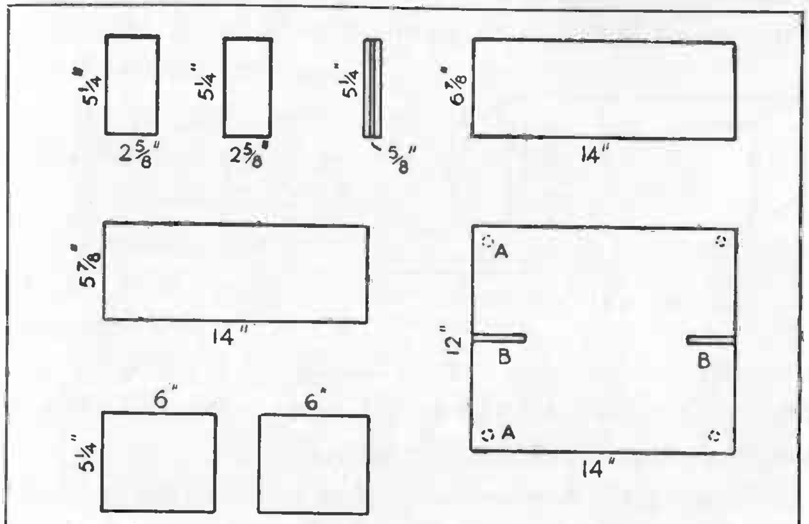

Fig. 37.—Details of Material for Case.

4 small china insulators (feet for instrument).

100 ft. of 7/22 bare or insulated copper wire.

8 2-in. china " reel " insulators with ½-in. diameter hole in centre.

A pole or poles not shorter than 22 ft. if possible.

Tuning Coil.—Should the coil former be of cardboard it should be put into a moderately hot oven for five minutes and then shellac-varnished. If the former is of ebonite, paxolin or any other specially-prepared material, this process need not be resorted to.

Shellac varnish may be made by dissolving about 4 oz. of shellac in half a pint of methylated spirit.

Now take the former and cut out a disc of wood $\frac{5}{8}$ in. in thickness to fit exactly in one end and glue it in position. Two holes of about $\frac{1}{16}$ in. should be made about $\frac{1}{2}$ in. apart on this end of the former and the end of the No. 28-gauge wire inserted and made fast in these holes, being wrapped round as shown in the sketch Fig. 35, leaving a free end protruding from one hole about 12 in. long. Proceed to wrap the wire round the former, taking off tappings every twenty turns for the first 200 turns, and then every two turns for another 18 turns.

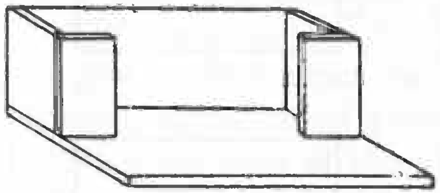

Fig. 38.—Case Assembled.

Tapping the Coil.—The method of making the tapping is to take the wire and turn a length of about 12 in. back on itself, twisting it two or three times between the finger and thumb and leaving a loop of 6 in. When the winding of the coil is completed, thoroughly clean the cotton covering from the wire where it is twisted, apply some powdered resin and solder it. When the soldering is completed, one side of the wire loop can be cut, and this will result in a lead of about 12 in. long being left, which is subsequently soldered to the contact stud.

An alternative method of making the tapping leads is to leave the wire twisted and to bare the end of the loop and solder this to the contact stud, but this method is not so neat or efficient. A loop of 9 in. should be taken if this method is used in order to allow sufficient wire to reach the contact stud.

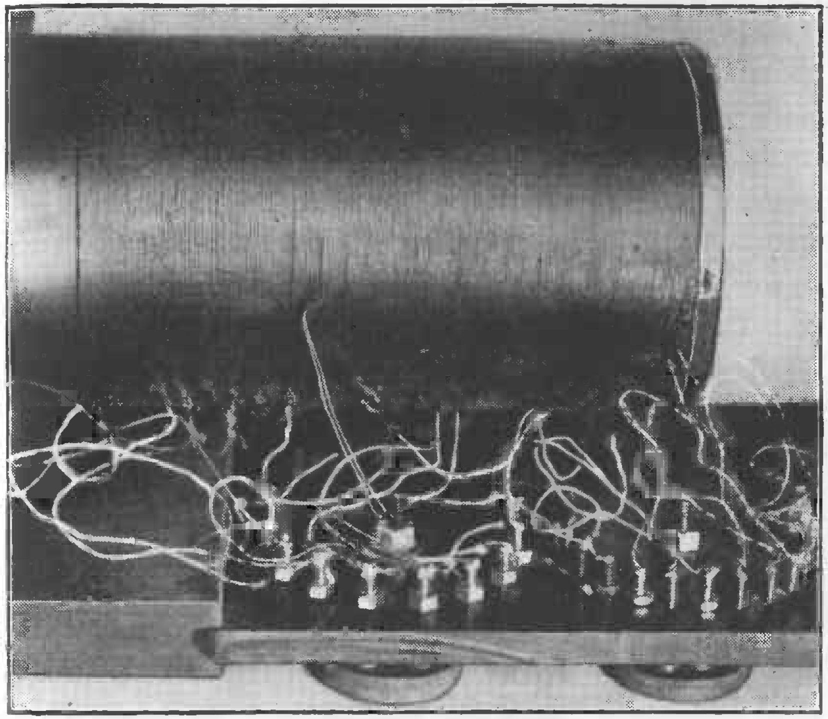

Fig. 39.—Photograph of Coil Connected to Panel.

Finish off the winding in the same manner as it was commenced, leaving a free end of 1 ft. On completion of the winding and the soldering the coil should be given two or three good coats of shellac. The tappings are illustrated in Fig. 36.

Case.—The next operation is to prepare the case for the reception of the coil and the panel. This should be of $\frac{5}{8}$-in. hardwood. Mahogany, teak or oak is to be preferred, but ordinary white wood or pine can be used if it is well seasoned. The case may be french-polished or stained and varnished, according to taste. Prepare the case according to the dimensions and lay-out given in Fig. 37, and use only screws in its construction. The right-hand side and top and bottom should be easily removable, otherwise the constructor can carry out the work according to his skill in joinery, providing the dimensions given are adhered to. The case when finished should appear as in Fig. 38.

A photograph of the coil with wires soldered is given by Fig. 39.

In order to attach the wires of the coil to the panel, as shown in the photograph (Fig. 39), lay the panel on the bench with the bar-switches assembled in position and its top towards the constructor, and place the coil with the end with the wooden end piece to the left and solder the free end to the first contact stud and the loops in rotation to the remaining studs. Keep the leads as short as possible consistent with leaving room to fix the coil and panel in the case and for other wiring. The tappings of twenty turns should now be on the left-hand switch and the two-turn tappings on the right-hand switch, the second free end being soldered to the last right-hand contact stud when completed.

It will now be seen that when the bar-switches are in a position where they point to each other the

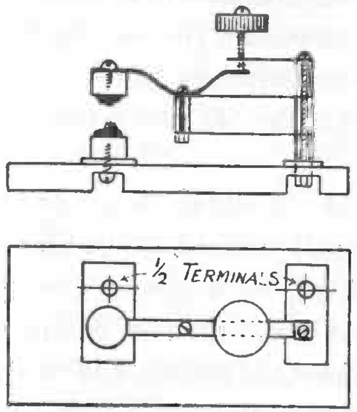

Fig. 40.—Elevation and Plan of Detector.

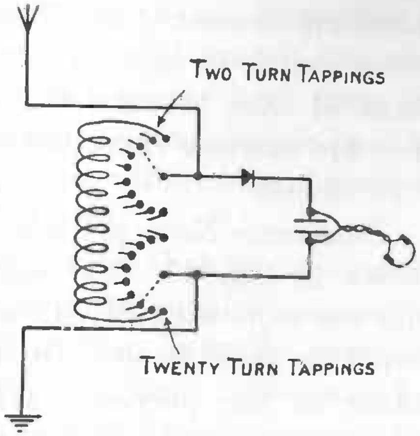

Fig. 42.—Circuit Diagram of Receiver.

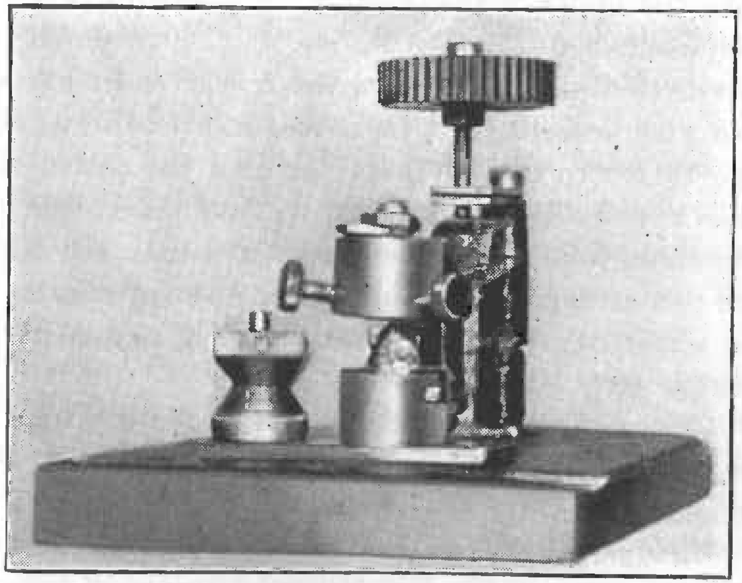

Fig. 41.—Photograph of Detector.

minimum amount of coil is in circuit, and that rotating either switch to right or left increases the number of turns of wire between the bar-switches until, when they are extreme right and left, the whole of the coil is in circuit.

Detector.—Now prepare the crystal detector as shown in Fig. 40. The crystals should be soldered into the cups with Wood's metal, with the screws inserted in their bases. On no account use ordinary solder for this purpose. When the screws are withdrawn they will leave a path for their re-entry into the cups when affixing them in position on the detector. Zincite and copper-pyrites or zincite and bornite are recommended, the zincite in each case being the upper of the crystals. A photograph of the completed detector, which should be mounted on ebonite or dry, well-seasoned wood, is given by Fig. 41. Two such detectors can be made, and it is an advantage to incorporate both of them in the instrument with a suitable two-way switch, as one might be sometimes found to be more sensitive than the other, and this arrangement also facilitates the comparing of the sensitivity of one crystal or set of crystals with another set.

The constructor may now complete the wiring of the panel with 18 S.W.G. tinned copper wire (see Fig. 42).

Wiring.—Particular care should be taken that all connections are made fast either by means of nuts or soldering. When applying the soldering-iron to terminals or studs in the panel do not apply it too hot,

as this is likely to spoil the insulating properties of the panel. A wet rag may be held against the stud being soldered, and this will take the heat from the panel. Ordinary 1/18 rubber-covered electric lighting cable can be used for the wiring with the braid stripped off (not the rubber), or bell wire.

It will be noted by the reader that the number of contact studs on the front of the panel illustrated in the photograph differs from that actually provided for, and also there is an extra terminal. The reason for this is that the writer has used the instrument for certain experimental purposes, such experiments necessitating an additional number of studs, etc., being used, but there is no need for the present construction to embody them.

Assembling.—The crystal detector and phones might be mounted in the position shown in the photograph (Fig. 32), and this has been found to be very convenient, the wiring being under the base and the instrument being kept insulated by the small china insulators. The constructor can, however, use his own discretion as to the position of these accessories. Two detectors are shown in the photograph, one being the perikon and the other a " cat's whisker." A potentiometer, carborundum detector and condenser are also shown on the top of the instrument ; these will be described in the next chapter. The wire between the back of the aerial and earth terminals and the bar-switches should be electric-bell flexible wire, as this will allow for rotation of the switch.

Fix the coil by screwing it to the inside of the left-hand side of the case in a convenient position about midway across the side and about ½ in. from the base of the instrument.

Testing the Set.—To operate the present instrument the crystal detector should be adjusted, and the contact between the two crystals should be light. A buzzer might be sounded near the coil or earth lead, and when the note of this buzzer is heard in the telephones the crystal is in a sensitive condition. Rotate the left-hand switch until signals are heard, and finish by fine tuning on the right-hand switch.

CHAPTER V

A Loose-coupled Set

By A. J. COOPER.

THE apparatus described in this chapter is of a more advanced type than that described in the previous chapter, and includes a tuner with a secondary closed circuit. The general constructional details, however, excluding the loose-coupler, are the same as in the set described in that chapter.

Materials.—The materials required are as follows: A cylindrical former 4½ in. long by 2¾ in. in diameter, 1 oz. 36 d.c.c. wire, a length of ⅜-in. dowelling, 12 contact studs, 1 bar-switch, 2 terminals, 15 ft. of electric-bell flexible wire.

The Former.—First take the former and prepare two wood or ebonite end pieces exactly to fit in the end, with a hole ⅜ in. in diameter bored through the centre of both, as illustrated at A in Fig. 43. If wood is used it should be perfectly dry and of a thickness of about ½ in. Should ebonite be used, it need be only ⅜ in. thick. In one of the end pieces drill ten or twelve small holes, as illustrated at B, Fig. 43.

The former, if of cardboard material, should be prepared with shellac varnish as specified for the primary former. Now draw a straight line down one side, mark off ½ in. from each end, and beginning just inside these marks bore twelve holes right through the former at equal distances from each other. The positions of these holes should be marked on the former before the commencement of operations. A second hole should now be drilled near the first and last holes for making the end of the wire fast. A diagram of the former prepared to these instructions is given by Fig. 44.

Winding the Former.—Now take the 36-gauge wire and commence winding it on to the former in precisely the same manner as adopted for the primary coil, with the exception that the end of the wire has first of all a length of the flexible electric-bell wire soldered to it and is then pushed through the hole into the former. A slip of paper numbered 1 should now be attached to this flexible lead, which should be about a foot long. Proceed with the winding, and meeting the first hole strip the cotton covering off the twisted wire and solder a flexible lead to this; insert it through the hole and number it 2. Proceed with the winding, and ensure that the next turn of wire bears closely against the one just soldered, and so carry on the soldering and numbering of the leads as the holes are reached until the coil is completed, when twelve flexible leads should be obtained from the end of the former farthest from the starting end. The leads will be numbered 1 to 12. Now insert the end pieces into the former, securing them in position by small brass screws, the one with the small holes being utilised at the end of the coil from which the leads emerge, the leads being passed through the small holes. The numbered tickets should be taken off the leads one at a time whilst they are being pushed through, and then re-attached after the lead is through.

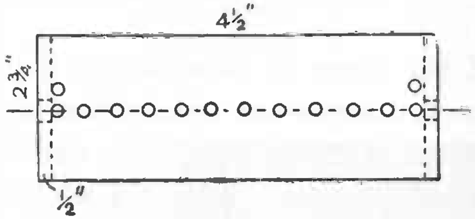

Fig. 44.—Diagram of Former.

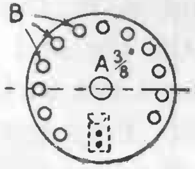

Fig. 43.—End Piece.

The primary-coil end piece should now have a hole bored through it ⅜ in. in diameter, and it is preferable that this should be done with the end piece detached from the coil.

The Sliding Arrangement.—Next take the dowelling and cut off a length exactly to fit inside and across the length of the case—that is, 12½ in. long if the case has been made to the specification—and prepare another piece 8 in. long.

Prepare a small brass right-angle bracket 1 in. each way from the apex of the angle, and bore two holes

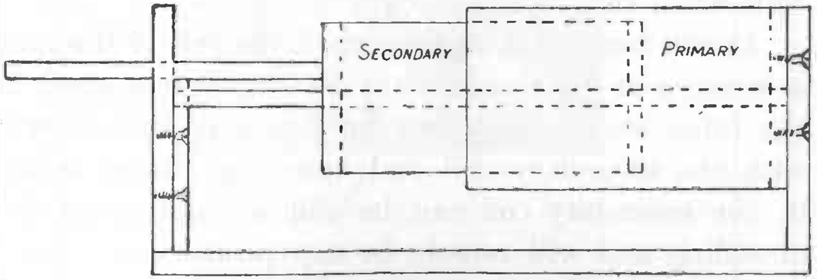

Fig. 45.—Diagram of Loose-coupler.

in each side to take small wood screws $\frac{1}{4}$ in. long and about $\frac{1}{16}$ in. thick.

On the end of the piece of dowelling 8 in. long make a flat surface and screw one side of the bracket to it. Prepare a piece of wood $\frac{1}{2}$ in. thick and $2\frac{3}{4}$ in. by $2\frac{3}{4}$ in. and take off the end of the case which does not hold the primary coil. This side would be the right-hand side when the case is facing the constructor. Now pass the piece of dowelling which is to fit exactly into the case through the secondary coil, and see that it slides easily and freely; then place the end in the hole just made in the end piece of the primary. Position the coils and dowelling in the case against the left-hand side about 1 in. clear of the base, and place the piece of wood $2\frac{3}{4}$ in. by $2\frac{3}{4}$ in. against the other end of the dowelling and resting on the base of the case. Adjust the coils and dowelling to a horizontal position, and mark on the piece of wood the centre position of the dowelling and bore a hole $\frac{3}{8}$ in. in diameter at this position. Now screw the wood to the end piece of the case which has been removed, seeing that it is in a line with the centre line of the coils when they are positioned in the case.

It will now be clear that when the end of the case is screwed in position and the primary coil secured to the other end of the case by means of two screws, with the secondary coil and dowelling placed inside it, the secondary coil can be slid up and down the dowelling and will remain in any position.

Now take the piece of dowelling with the right-angle bracket attached to it and secure the other side

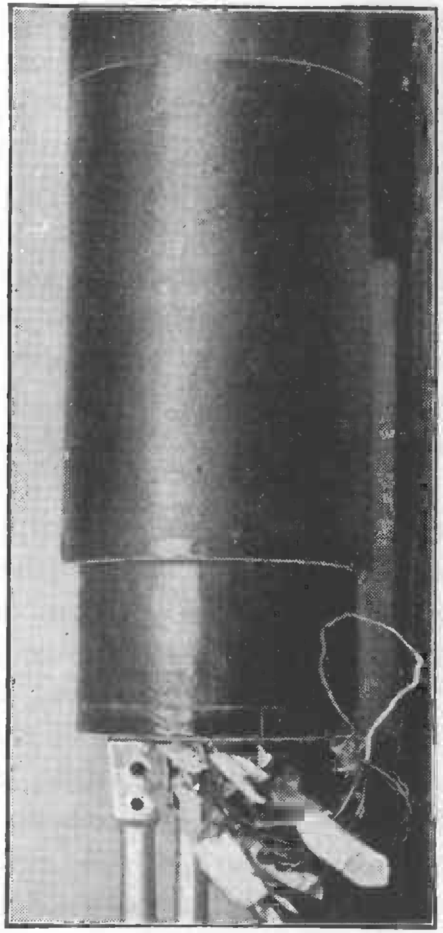

Fig. 46.—Photograph of Loose-coupler.

of the right-angle bracket to the secondary-coil end piece through which the flexible leads emerge. A hole must now be bored through the end of the case which has been removed to allow this piece of dowelling to project, and this forms the rod by which the coupling of the coils is varied. The coil when viewed from the back will now appear as in Figs. 45 and 46.

Solder the ends of the flexible leads and the contact studs of the secondary bar-switch, fix the coils in position in the case, and wire up the remainder of the circuit, as in the diagram (Fig. 47). From this diagram it will be seen that the secondary coil, crystal detector, phones and variable condenser are in an entirely self-contained or closed circuit, and the primary coil is in a separate circuit in a position between aerial, and earth.

The Closed Circuit.—There is no apparent connection between the two circuits, but the connection is made by induction between the two coils. To put it simply, the oscillating current passing through the primary coil on its way to earth sets up magnetic lines of force about the primary, and the secondary coil being close enough to intersect these lines of force, a corresponding current is induced in this closed secondary circuit. These high- or radio-frequency oscillations are rectified or filtered out by the crystal " detector," which allows a high percentage of low-frequency current (sometimes termed audio-frequency) to pass and produce an audible note in the telephones.

The Variable Condenser.—Although not essential, it is customary to employ a variable condenser in

conjunction with the close or secondary circuit, as by its means very fine tuning is possible.

Such a condenser can take several forms for practical work (see the companion handbook " Wireless Component Parts," which also explains capacity calculations), the cheapest and possibly easiest to construct being of the sliding type and made from old photographic negatives and sheet zinc. The glass (with the film cleaned off) is interposed and supported between sheets of zinc of a slightly smaller area, and the movable or sliding sheets of zinc are pushed in between the glass sheets and the fixed zinc plates above and below. In this case the glass is the dielectric.

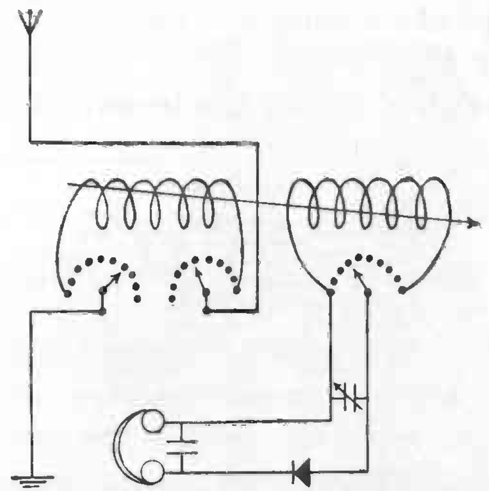
Fig. 47.—Circuit Diagram for Loose-coupled Crystal Set.

A diagram showing the construction of a variable condenser of this nature is given in Fig. 48.

The capacity of a condenser is measured in microfarads, and the size of condenser suitable for the present purpose is .0003 microfarads.

Other forms of condenser are in existence, such as the tubular and rotary-vane type. It is this latter type which is considered by the majority of experimenters as being the most suitable for wireless work generally, as it is compact and smooth in operation. A photograph of a condenser of the capacity prescribed is shown by Fig. 49.

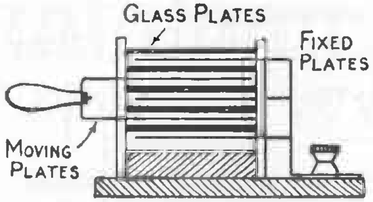

Fig. 48.—Sliding-type Condenser.

Mounting the Variable Condenser.—When assembling the variable condenser, which should have an ebonite panel and base, one of the fixed plates should be placed on the ebonite panel and used as a template for marking off the position of the ties which hold these fixed plates in position. A rotary plate should be placed so that its straight edge is square with the straight edge of the fixed plate but not quite touching it, and the small projecting piece should just fit into the recess cut into the fixed plate for its reception. When the centre of the square hole in the rotating plate has been marked on the ebonite all the holes may be bored with drills of an appropriate size. The

fixed-vane tie rods should be screwed into the ebonite, and for this purpose a tap will be required in order to put a thread in the hole in the panel. The rotary-vane spindle must of course be a close fit to the shoulder on the spindle. When the top panel has been prepared it can be used as a template for the base in which corresponding holes will be required.

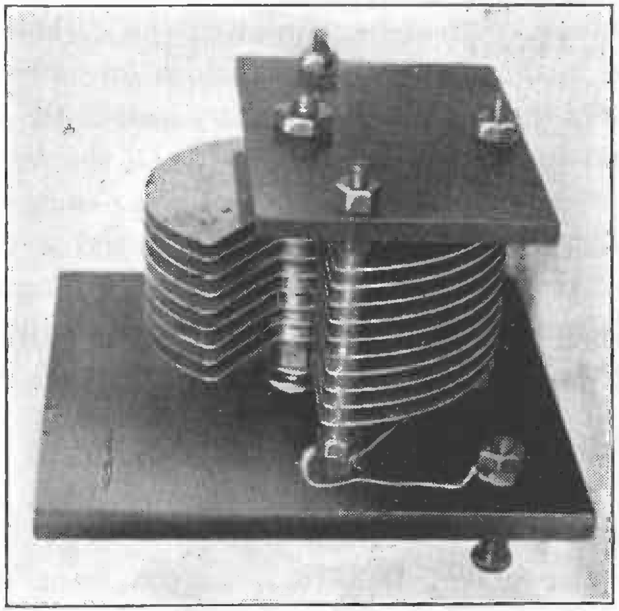

Fig. 49.—Photograph of ·0003 mfd. Variable Condenser.

Additional Experimental Instruments.—It is probably known to many readers that a carborundum detector requires a potentiometer and battery for use in conjunction with it, and that these are required in order to create a difference in potential between the carborundum crystal and steel plate. The reader may like to construct a simple type of carborundum detector, and for this reason dimensions and details are given in Figs. 50 and 51.

Making the Potentiometer.—The potentiometer consists of a piece of wood (broom handle will do) 1 in. in diameter by 4 in. long, wound for 3⅛ in. of its length with No. 36 d.s.c. Eureka resistance wire and mounted between suitable supports. A slider makes contact on the top surface and runs on a piece of brass rod ¼ in. square. When incorporating the carborundum detector in the circuit the wire which formerly led to one side of the perikon detector is joined to the square brass rod by means of a terminal, and the two ends of the resistance wire are connected to each side of a 4-volt dry battery. There is also looped or twisted with one of these battery wires a wire leading to the steel plate on the crystal detector, and another wire runs from the crystal to the telephones. A switch should be incorporated in the potentiometer to cut off the battery current when not in use. The buzzer is used as explained in the previous chapter, and is obtainable for a very small sum.

The Catwhisker Detector.—In the front of the instrument should be placed a simple catwhisker detector, as used in conjunction with certain crystals —for example, galena, rectarite, hertzite, etc. This is simple to construct, and Fig. 52 shows the general arrangement. A photograph is shown by Fig. 53.

The bracket is a piece of brass 3 in. by ¾ in. bent to an angle of 90 degrees. A hole is drilled ½ in. from the top of the long side (2 in. long) to take the screw

which secures the crystal cup. A hole is drilled on the other side (1 in. long) to take a terminal.

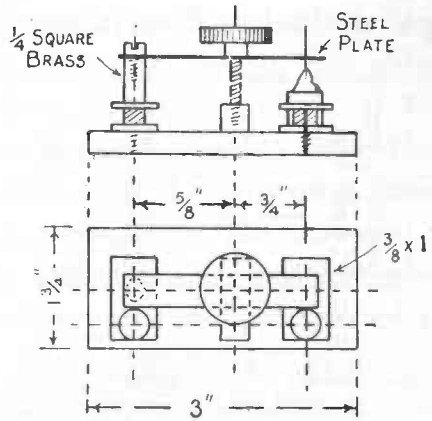

Fig. 50.—Details of Carborundum Detector.

The other bracket is a similar piece of brass bent in the same manner and to exactly the same dimensions.

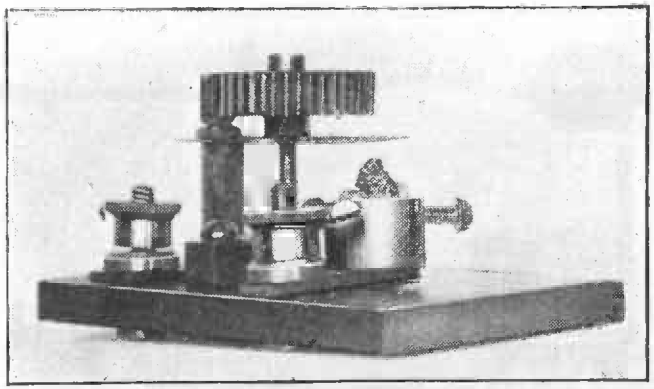

Fig. 51.—Photograph of Carborundum Detector.

A hole is drilled ½ in. from the top of the 2-in. piece, but of a larger diameter—in fact, almost large enough to allow a brass ball to pass through it, but not quite. This brass ball might be ½ in. in diameter, but not larger; the one used on the instrument in the photograph was taken from the top of an old bed rail. A piece of springy brass is prepared 2 in. by ¾ in., and has a similar hole drilled in it. The bracket and the spring are then placed together and a hole drilled through the component parts about ½ in. from the foot of the bracket to take a clamping screw, which should be provided with a nut. The brass ball has a hole drilled right through it to take a piece of stout brass wire 3½ in. long, which is then soldered in place, allowing one end on one side of the brass ball to be 1½ in. long.

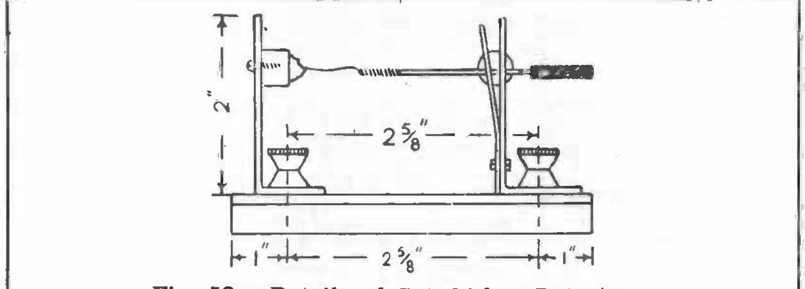

Fig. 52.—Details of Catwhisker Detector

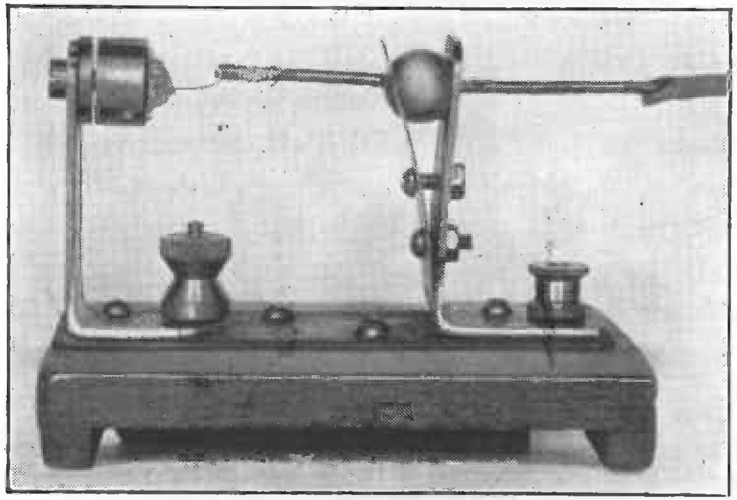

Fig. 53.—Photograph of Catwhisker Detector.

Place the ball in position over the hole in the bracket and fit over it the spring brass, and clamp in position by means of the nut and screw shown. Solder to the long side of the stout brass wire a spiral of wire 36 S.W.G., and on the other end fit a piece of ebonite rod or rubber to act as an insulator when adjusting the position of the fine wire on the crystal. Fit the crystal cup with a piece of crystal and mount the whole on a suitable base of ebonite or hard. dry wood.

CHAPTER VI

Set with Plug-in Basket Coils

By H. J. TURPIN.

An unusual type of crystal receiver is described in this chapter, one in which is embodied extreme selectivity. Instead of the orthodox slider and cylindrical inductance coil or the variometer methods, the writer has adopted a system of combined adjustment of the aerial and earth connections, each of which can be independently connected to any particular part of an inductance coil of the " basket " type. It is claimed that by this method (which is explained more fully later) an amount of wire equal to $\frac{1}{2}$ in. can be added to or subtracted from the total quantity in the inductance circuit depending upon the relative position of the aerial and earth tuning plugs. The set is shown photographically in Figs. 53A and B.

In order to ensure that the instrument is as compact as possible it is made in a simple box form, one end of the case being used to carry the crystal detector which is entirely covered in by a suitably designed lid, while the main part contains the ebonite panel and the basket inductance coil.

Case.—The complete case is shown pictorially in Fig. 54, and is made throughout of mahogany $\frac{1}{4}$ in. thick, the overall dimensions being $5\frac{5}{8}$ in. by 4 in. by

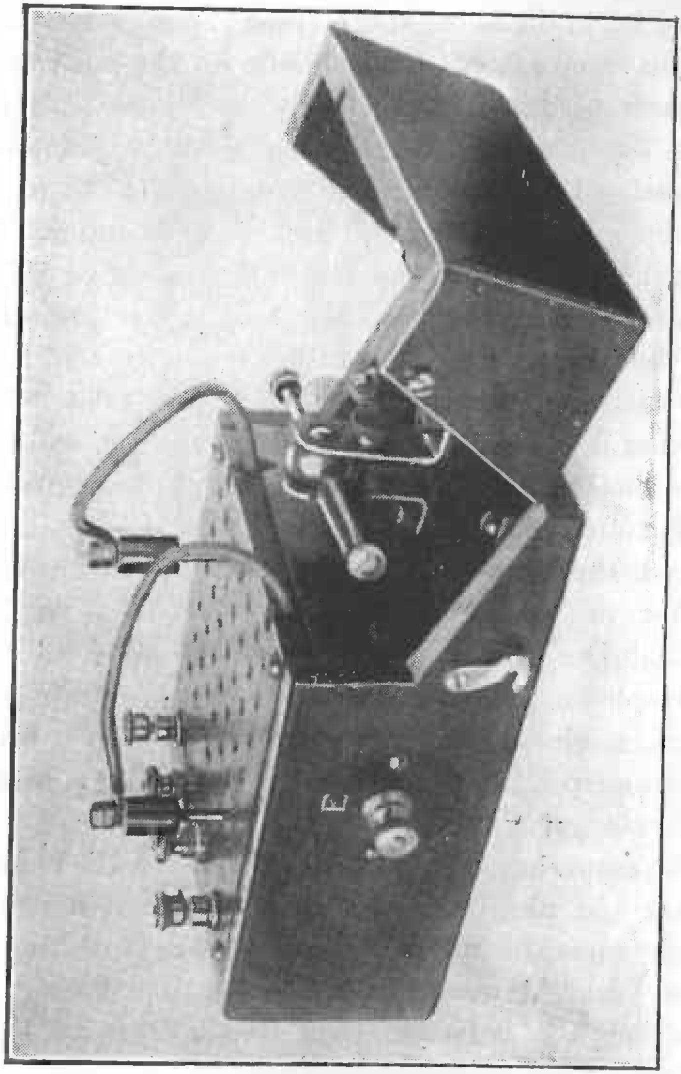

[Fig. 53a.—Photograph of Receiver shown Open.

2 in. deep, and the inside dimensions of the chamber containing the "basket" coil are $3\frac{1}{2}$ in. by $3\frac{1}{2}$ in. Holes $\frac{7}{8}$ in. in diameter and recessed $1\frac{1}{4}$ in. in diameter by $\frac{1}{16}$ in. deep are cut on each side for the main aerial and earth terminals respectively, as shown at A and E, Fig. 54. Beyond this the figure is self-explanatory.

Panel.—The panel is made of ebonite $3\frac{3}{4}$ in. by $3\frac{1}{2}$ in. by $\frac{1}{4}$ in. thick arranged as in Figs. 55 and 56. The tuning holes are made by screwing a piece of 2 B.A. rod into the ebonite until the end is just projecting from the other side; the rod is then sawn off almost flush with the panel and roughly filed flat. This operation is repeated for the forty-two plugs and then all are finally filed flush with the panel face, finishing off with medium emery cloth and oil.

Mark the centres of the plugs with a centre punch and then drill each one $\frac{1}{8}$ in. in diameter by $\frac{7}{32}$ in. deep and slightly round off the edges of the holes.

Telephone terminals are fitted in the positions indicated, as also are another pair connected by a bridge of brass strip so that on removing the bridge a loading coil for longer wave-lengths can be fitted.

The slots shown in Fig. 56 are filed in the ebonite, so that the plug leads for aerial and earth can be brought outwards when the instrument is in use, and can be folded downwards and tucked inside the cover containing the detector when the instrument is not required.

Detector.—This is built up entirely of old pieces, the ball being a brass one (from an old bedstead) and plugged and drilled for the wire holder. A simple

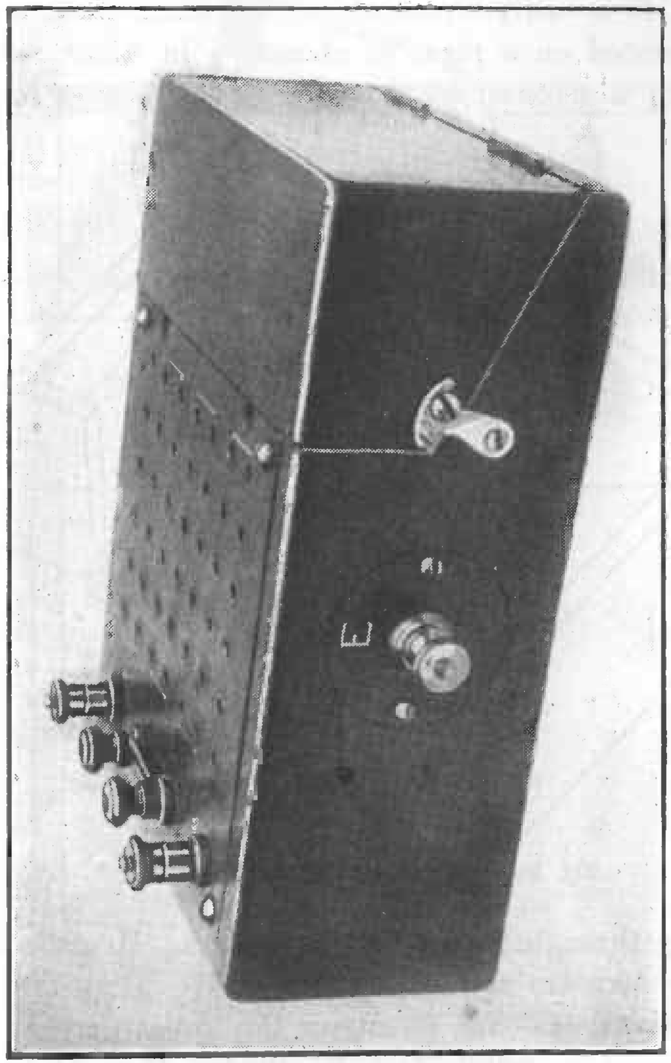

Fig. 53b.—Photograph of Receiver shown Closed.

method of holding the wire is to screw the end of the rod and then clamp the " whisker " between two nuts. It is easily replaceable. The complete detector is mounted on a piece of ebonite $\frac{1}{8}$ in. thick, which in turn is screwed to the base by four brass screws passing through four celluloid washers. Hertzite and copper wire are used as a detector.

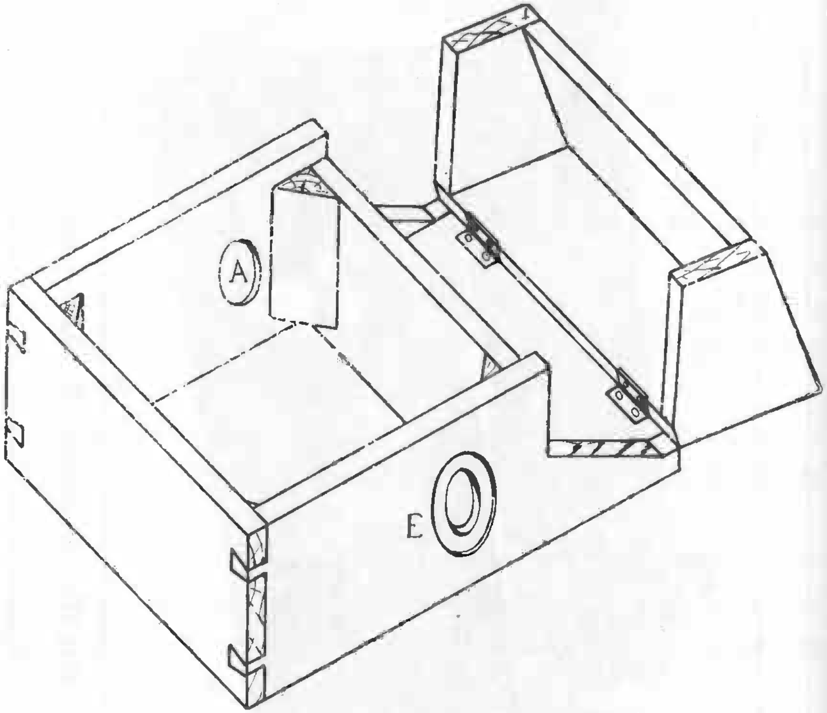

Fig. 54.—Details of Containing Case.

Inductance.—Fig. 60 shows the appearance of the former cut in celluloid $\frac{1}{16}$ in. thick. It is $3\frac{1}{2}$ in. in diameter and contains 9 vanes, the bottom of which is on the circumference of a circle $\frac{3}{4}$ in. in diameter. Seventy-two turns of No. 26 d.s.c. wire are wound on

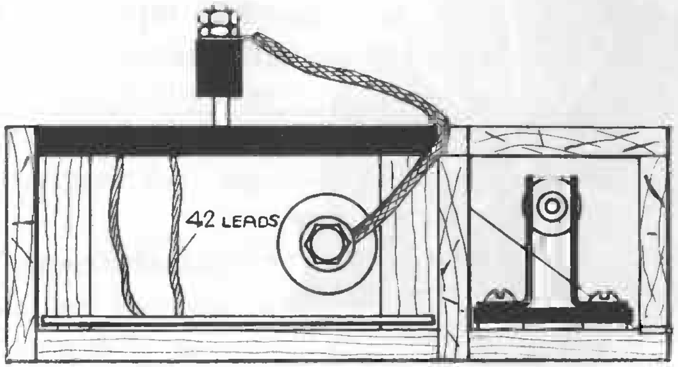

Fig. 55

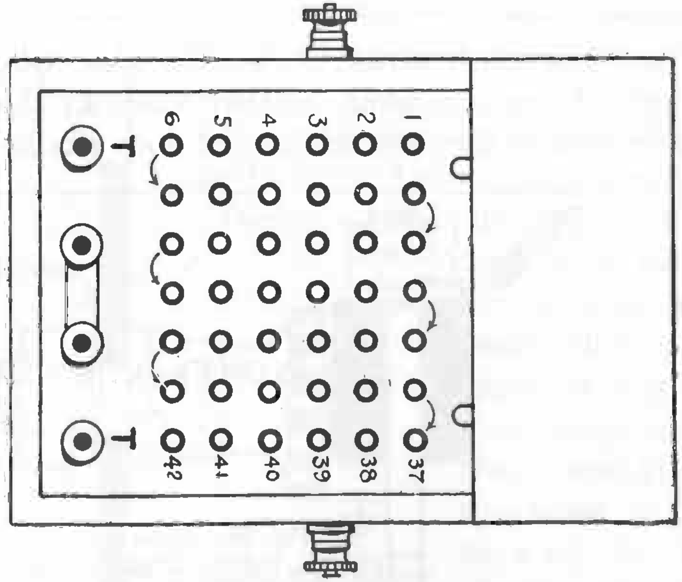

Fig. 56

Figs. 55 and 56.—Section and Plan of Receiver with Plug-in Basket Coils.

and tappings taken off in the following manner:—Starting from the inside, the first tapping is taken after the first turn, the second is taken at the next

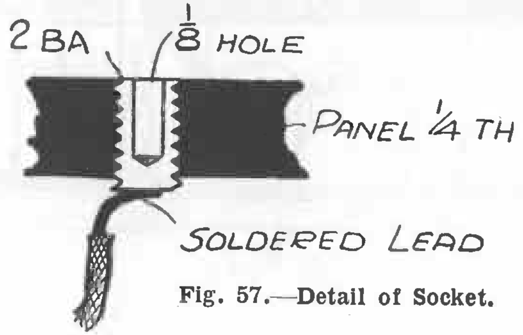

Fig. 57.—Detail of Socket.

one and one-ninth turns, i.e. on the vane following the end of one revolution, so that when all the tappings are made they are not heaped upon one another

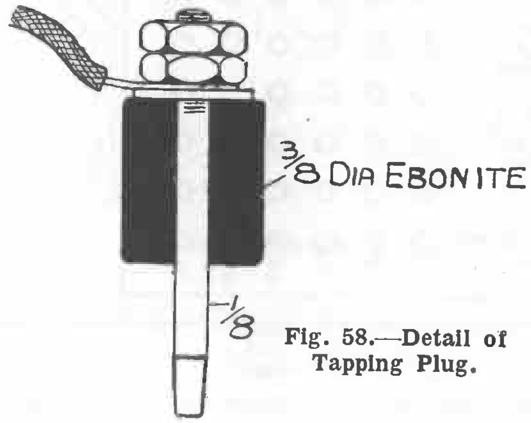

Fig. 58.—Detail of Tapping Plug.

on the same vane. The third tapping is taken after the next one and one-ninth turns, and so on. When completed, 33 tappings should be taken as just

described and 9 tappings taken at every four and one-ninth turns. The latter are for coarse tuning, and the former for fine tuning.

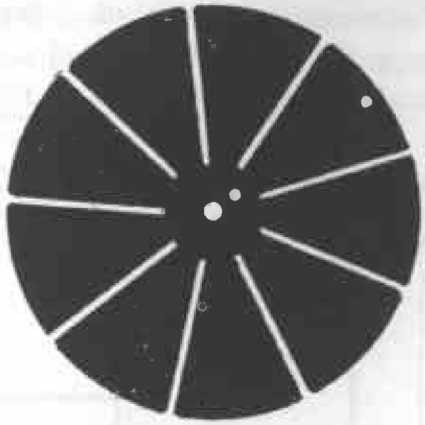
Fig. 60.—Tuning-coil Former.

Connections.—For inductance coil use a five-stranded flexible wire (called "baby" flex) and make each 5 in. long. Solder all connections, and on completion soak the whole in paraffin wax.

The plug numbers in Fig. 56 indicate the order of attaching connections.

Tuning.—As an illustration of the method of tuning, for example, it is found that 2LO comes in very well when plugs are in Nos. 2 and 41, and also when in Nos. 29 and 42, but there is a very slight difference in the result, as the difference between the length of wire

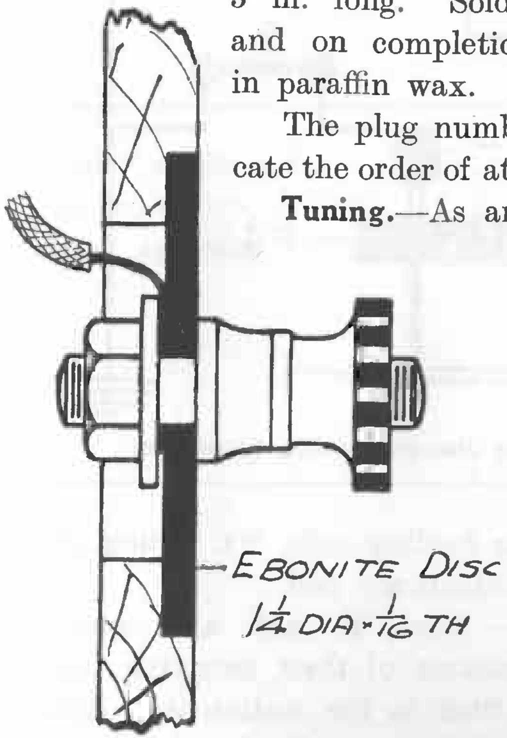

Fig. 59.—Method of Fixing Terminals.

between each pair of holes is very small. In this case the final tuning is obtained by moving the aerial and earth plugs together into either pair of holes.

Explanatory details are shown in Figs. 54 to 60 and a wiring diagram in Fig. 61, in which figure B represents bridge for loading coils, P L tuning plugs, P A panel, and I inductance coil.

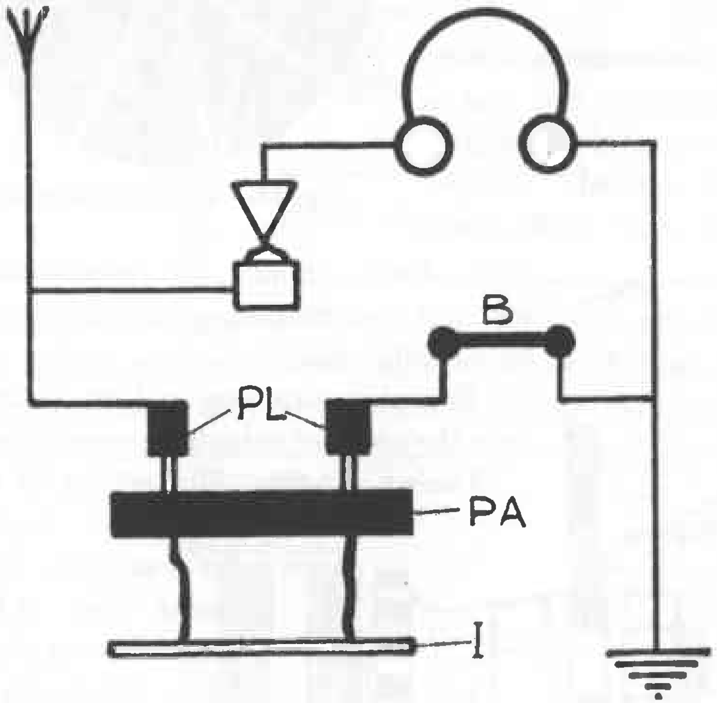

Fig. 61.—Circuit Diagram showing Connections.

Tuning Cards.—Most, if not all, amateurs memorise the calibrations of their receiving sets or have to tune each time to the station it is desired to hear. It may not be generally known that the use of a tuning card would be of great assistance. These are used in just the same way as a business man uses a file index. An example of such a card is shown above, though, of course, this may be modified according to individual requirements.

Stn.	γ	A.T.I.	A.T.C.	C.C.C.	Coil Nos.			Coupling	H.T. (volts)	Strength	Remarks
					P	S	R				

CHAPTER VII

Combined Crystal and Valve Receiver

By P. T. BEARD

THE receiver dealt with in this chapter is not new, but its many advantages are usually unknown to the newcomer in the wireless field. It is as easy to make as a simple one-valve receiver. Briefly the receiver consists of a valve used as an amplifier of radio-frequency oscillations, in conjunction with a crystal used as a detector or rectifier.

The Tuner.—An ebonite or prepared cardboard cylinder 4 in. long by 3 in. in diameter is wound with forty turns of No. 20 d.c.c. copper wire, tapped off to a seven-point switch from the following turns: 6, 12, 18, 24, 30, 36 and 40. Between turns 12 and 13 a space of $\frac{1}{4}$ in. is left as shown in Fig. 62.

Reaction Coil.—This is a basket-wound coil of No. 26 s.c.c. copper wire wound on a cork 1 in. in diameter with thirteen pieces of cycle-wheel spoke or blanket pins spaced equally round its circumference. The wire is simply wound in and out round the spokes (Fig. 63) until the coil is just of sufficient diameter to rotate inside the inductance coil without touching. When the reaction coil is wound, leave about 6 in. of loose wire at each end for connections, and soak the coil in a bath of melted paraffin wax. When the wax is just about to set, remove the coil and allow it to

harden. Afterwards the spokes can be withdrawn and the cork taken away.

A piece of brass rod 5 in. long and $\frac{1}{8}$ in. in diameter is now passed through the inductance coil in the space left between turns 12 and 13, and the coil lashed securely to the brass rod by means of adhesive tape. A hardwood or ebonite knob is fastened to the end of the brass rod so that the reaction coil can be rotated inside the inductance coil (Fig. 64).

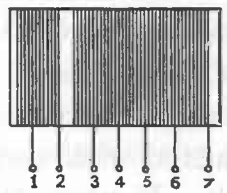

Fig. 62.—Inductance Coil.

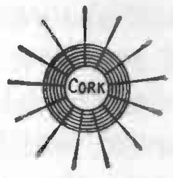

Fig. 63.—Reaction Coil.

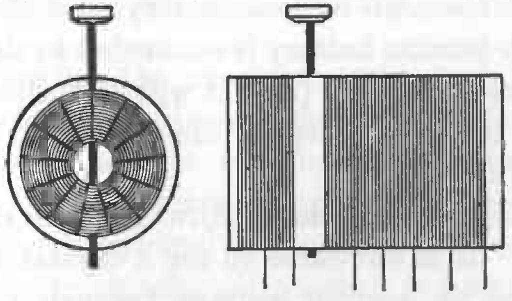

Fig. 64.—End and Side Views of Mounted Reaction Coil Fitted in Inductance Coil.

A variable condenser of about ·0005 microfarad capacity is connected between the aerial end of the inductance (A) and the handle of the seven-point switch, and the tuner is complete (Fig. 65).

The complete receiver is shown in the diagram (Fig. 65). It will be seen that the aerial is joined to the inductance coil at the end nearest to No. 1 tapping. The earth is connected to the arm of the seven-point switch, and the variable condenser is placed between the aerial and earth. The grid of the valve (G) is connected to the aerial terminal through a grid leak and condenser (C), which may be omitted with very little detriment to the signal strength. If used, the grid leak should be of two megohms resistance and the grid condenser of ·0003 microfarad capacity.

One connection of the reaction coil (which is, of course, inside the inductance coil) is connected to the plate of the valve (P), and the other connection is taken to the telephone terminal which is nearest to the positive terminal of the high-tension battery (M). Take careful note of this part of the circuit. A connection is taken from this telephone terminal to the positive of the high-tension battery, and the negative of the high-tension battery is connected to the negative of the filament battery (H). It will be noticed that the negatives of both batteries are joined to the earth terminal (E).

The Crystal Detector.—Now for the crystal detector (K). It is advisable to use a crystal which does not require an applied voltage through a potentiometer. A connection is taken from the plate of the valve (P) to one terminal of the crystal detector (K). The other terminal of the crystal is joined to the telephone terminal opposite to the one which is connected to the reaction coil. A fixed condenser (N) of

about ·01 microfarad capacity placed across the phones and high-tension battery as shown will improve results. The valve filament is supplied from a 6-volt accumulator through a variable resistance (J) of 5 ohms.

Operating the Set.—Phones of from 4,000 to 8,000 ohms total resistance should be used, or 120-ohm phones will do as well if a telephone transformer is inserted between the phones and phone terminals.

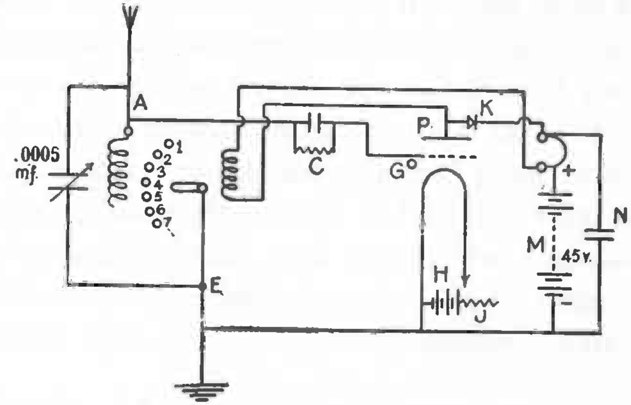

Fig. 65.—Diagram of Connections for Complete Receiver.

When testing the receiver, make sure that the crystal is correctly adjusted, then rotate the reaction coil; at the same time tap the aerial terminal with the finger. If clicks are heard in the phones each time the aerial terminal is touched, the receiver is ready to receive signals. If no click is heard, reverse the connections to the reaction coil; this will give the desired result.

With careful adjustment of the crystal the strength of signals will be from 200 per cent. to 300 per cent. louder than those obtained when using the valve alone.

Another advantage of this circuit is that by simply switching off the filament current signals can still be received on the crystal. This is an asset when the filament battery runs down in the middle of an interesting transmission.

CHAPTER VIII

Some Miniature Receiving Sets

Whilst a set carefully built on orthodox lines is a necessity for those living some distance from a broadcasting station, the simplest of sets, even those which may be classed as freaks, are capable of giving excellent results within a range of a few miles of a broadcasting centre. Some have yielded good results over a range of more than 100 miles, but this must be taken as exceptional, and not to be expected as a general rule. The sets described in this chapter are miniature sets of the simplest possible construction.

A Basket-coil Receiver

This set, as previously intimated, is limited as regards range ; that is, it can be described as a single-purpose instrument suitable for broadcast reception only.

An elevation and plan of the instrument is shown by Figs. 66 and 67.

The set consists of an ebonite block supported on four legs, between which are situated inductance coils of the " basket " type. On the top of this block is secured the crystal detector and terminals.

Constructing the Coils.—The construction of the coils is the first operation, as really the size of the base is dependent upon the outside diameter of these, but if the latter are made in accordance with Fig. 68 sufficient wire can be wound on to suit the majority of " broadcast " aerials. Disks $2\frac{3}{4}$ in. in diameter are cut from waxed cardboard, celluloid, or ebonite, and nine slots, each $\frac{3}{32}$ in. wide, cut therein, so that they reach down to the circumference of a circle 1 in. in diameter. A hole $\frac{1}{8}$ in. in diameter is drilled in the centre, and two pairs of holes about $\frac{1}{16}$ in. in diameter are made for the purpose of securing the ends of the wire.

Winding.— The wire is put on as follows: Wind on uniformly 44 turns of No. 26 d.s.c. wire on each disk and secure the ends through the holes just mentioned. It will be necessary to test these coils, the aim being to obtain perfect tuning when they are in the assembled position shown in Fig. 66.

Coupling these coils together and connecting to a crystal detector in the manner shown in the diagram (Fig. 69), it will be possible to determine the best relative position of the coils. (It must be remembered that it is important that the windings should both be in the same direction.) If, for instance, good reception is obtained with the coils 12 in. apart, then some wire must be removed from each coil, taking off one turn at a time, until with the coils about 1 in. apart good results are obtained. In the writer's case the coils each contain 38 turns, in others 36 and 44 turns have been necessary, depending on the natural wave-length of the aerial. This completed, leave 6 in. of wire projecting at each end for connecting purposes, give the

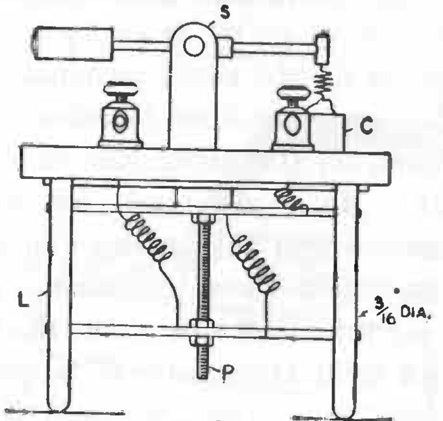

Fig. 66

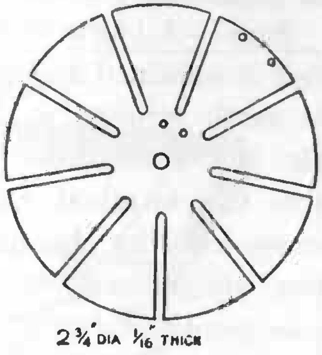

Fig. 68.—Former for Winding Coils.

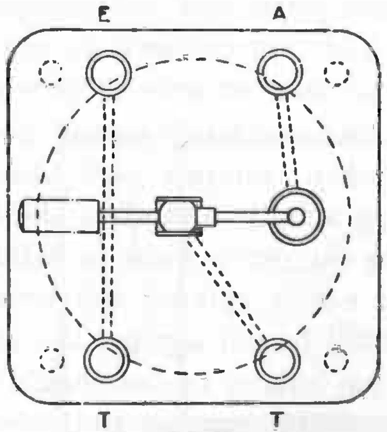

Fig. 67

Figs. 66 and 67.—Elevation and Plan of Miniature Basket-coil Receiver.

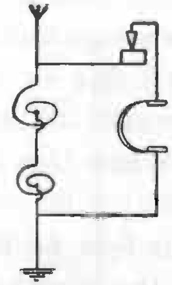

Fig. 69.—Circuit.

whole a coat of shellac varnish, and set aside until the remaining portion of the instrument is completed.

Base.—A block of ebonite $3\frac{1}{4}$ in. by $3\frac{1}{4}$ in. by $\frac{3}{8}$ in. thick is obtained and drilled for the aerial terminal A, the earth terminal E, the two telephone terminals T (Fig. 67), and counterbored on the under side to receive the terminal nuts. Holes are made for the screws securing the stirrup S and the crystal cup C, after which channels are cut underneath connecting these points as shown. On the under side of the block five holes are drilled (not right through) and tapped 4 B.A. to receive the four legs L (Fig. 66), and the centre spindle P, which is a piece of rod 2 in. long screwed 4 B.A.

Assembling.—Assemble the four terminals, crystal cup and stirrup and make permanent connections in the channels provided. The best connections are obtained by taking three or four strands of thin wire and twisting them together, as single-strand coarse wire is liable to break when clamped hard under a nut; screw in the centre spindle and then the four legs. When this is done cut out a piece of celluloid (Fig. 66) to fit over the centre spindle and between the four legs, so that it will lie flat against the under side of the ebonite base (no nuts or screws should project). Before finally assembling, connect up the inductance coils to the respective terminals, replace the celluloid and place on the first coil, taking care to insert a rubber washer $\frac{1}{2}$ in. in diameter by $\frac{3}{16}$ in. thick between. Clamp up by means of a 4 B.A. nut. Solder the wire connection between the two coils, and

twist the intervening length into a neat coil and assemble the second disk at the predetermined distance by

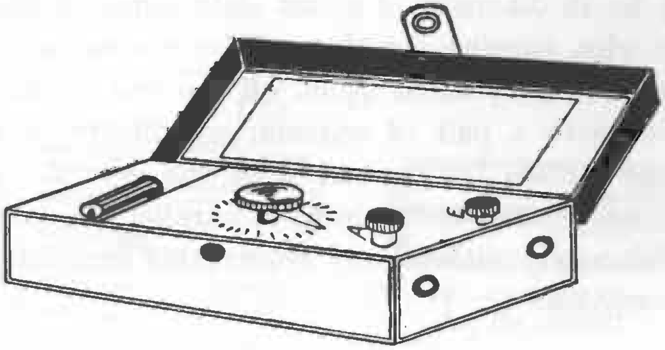

Fig. 70.—General View of Vest-pocket Receiver.

a 4 B.A. nut on each side, at the same time observing the correct direction of the coil winding.

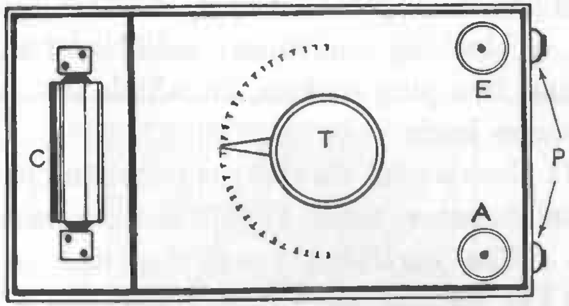

Fig. 71.—Plan of Vest-pocket Receiver.

Check over the connections carefully in accordance with the diagram Fig. 69.

The Detector.—Hertzite crystal and copper wire is probably the most satisfactory combination. The copper wire (catwhisker) is preferably made into a coil $\frac{1}{4}$ in. in diameter of about eight turns of No. 36-gauge wire, turning out the end for contact with the crystal. Cutting about $\frac{1}{32}$ in. off the end of the catwhisker with a pair of scissors periodically provides a clean surface for contact with the crystal. Four headphones (each 4,000 ohms) in parallel can be used simultaneously without any appreciable diminution of the reception.

Building a Vest-pocket Receiver

The sketch (Fig. 70) gives a general idea of the appearance of a vest-pocket receiver contained in a safety-razor case. The instrument has only one adjustment, the small tuning knob in the centre of the panel.

Elements of the Set.—Briefly, the receiver consists of a variable tuning inductance, a cartridge crystal detector, a blocking condenser, aerial and earth terminals, and two plug sockets, to which are connected the telephone leads.

Fig. 71 gives a plan view of the panel and position of the crystal detector, while Fig. 72 is a diagram of connections. The cartridge containing the crystal is shown in Fig. 73, and the tuning inductance in Fig. 74.

The Case and Panel.—The first item requiring consideration is the containing case. The ordinary safety-razor case is ideal for this purpose. An ebonite panel is now cut to fit tightly into the case, leaving a small partition in which the crystal is housed. The ebonite panel must be securely supported, either by means of a small strip fixed in the inside of the case or by means of small countersunk screws passing through the sides.

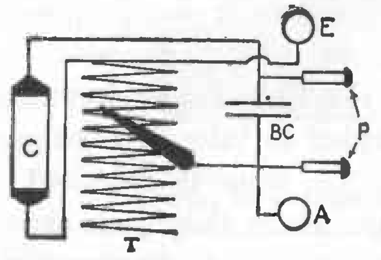

Fig. 72.—Diagram of Connections of Vest-pocket Receiver.

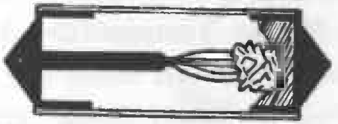

Fig. 73.—Cartridge Containing the Crystal.

The Tuning Coil.—To the under side of the ebonite panel is fitted the tuning inductance, which consists of No. 32 enamelled wire wound on an ebonite former $\frac{3}{16}$ in. thick, 2 in. deep, and as wide as can be got into the case, that is, a shade smaller than the ebonite panel itself.

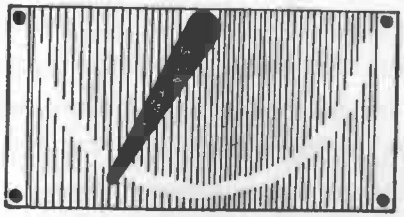

Fig. 74.—Tuning Inductance.

The Contact Finger.—A small contact finger is now made up, soldered to a short spindle, and arranged to make contact on the bared portion of the inductance winding, as shown in Fig. 74. This contact finger is moved by means of the tuning knob T on the top of the panel. A pointer may be fitted to enable comparative tests to be made, a scale being engraved on the panel, as indicated in Fig. 71. Two small terminals are fitted at one end of the panel for the purpose of connecting the aerial and earth wires to the set.

Phone Terminals.—At the end of the containing case two small holes are drilled to take two ordinary valve sockets; these must be securely mounted in position. The holes through which they pass should preferably be bushed with ebonite to increase the insulation efficiency of the set.

A pair of valve legs, corresponding to the sockets, should be soldered to the phone, which, for compactness' sake, may consist of a single 4,000-ohm receiver without a head-band.

The Detector.—The cartridge crystal detector is very simple in construction: a short length of ebonite or fibre tube is obtained and two small metal caps made to fit tightly into each end. One cap should be filled with Wood's alloy, and a good sample of a galena crystal mounted therein. The other cap is fitted with a metal rod $\frac{1}{8}$ in. in diameter, to the end of which are soldered several wires of about No. 26-gauge copper wire. Hard-drawn wire should be used on account of its springiness.

With regard to the metal end cap, these are best turned out of solid brass rod, the ends being pointed for the spring-clip holder provided. This holder may consist of two right-angled brass strips screwed to the

bottom of the containing case. It is, of course, understood that the cartridge is assembled so that the bunch of copper wires makes contact with the crystal in several different places.

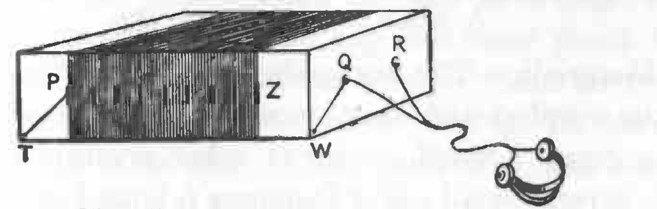

Fig. 75.—Match-box Receiver.

The Condenser.—The blocking condenser consists of eleven sheets of tinfoil, 2 in. by $1\frac{1}{2}$ in., interleaved with wax paper and mounted, underneath the panel on the bottom of the case.

Connections.—The diagram of connections is given in Fig. 72. The telephones are connected to the plug sockets P, the aerial and earth to A and E, B C

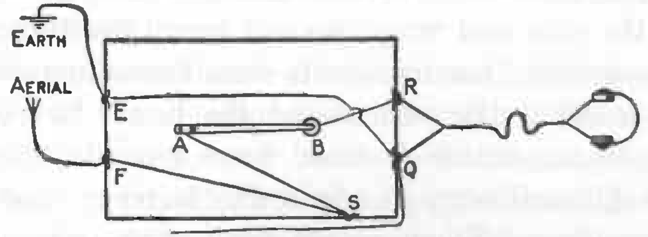

Fig. 76.—Diagram of Connections of Match-box Receiver.

being the blocking condenser, T the tuner, and C the crystal.

Range of Set.—Using an aerial of the regulation length and height the broadcast concerts should come through very clearly, while a certain amount of Morse and amateurs' transmissions may be picked up under good conditions.

Quite sharp tuning can be obtained with the wiring as given in Fig. 72.

A Match-box Receiver

Materials.—The materials required for this set are of the simplest and merely comprise : 1½ oz. of No. 26 d.c.c. wire, 1 small galena or other crystal, 1 match-box, several small paper fasteners (about ¼ in.), 4 large paper fasteners (about ½ in.), a small strip of thin brass, and the screw cap of a metal-polish tin.

Winding the Tuner.—The amount of wire, etc., depends on the size of the match-box used ; the best is one of the largest. At each end of the box make two holes about 1 in. apart, and through these holes put the large fasteners. About a ¼ in. from the end on one side of the box put a small fastener P (Fig. 75), with the head inwards and through the cover only. Take the wire and wrap the end round the fastener P on the outside, leaving about 6 in. for wiring. Then begin to wind the wire round the box. Every five turns put in another fastener, head inwards, bare the wire a bit and wrap it round the fastener, and continue winding. When within ¼ in. of the end put in another fastener, cut the wire, bare the end and wrap it round the last fastener.

Through the inner box part put a fastener S, as shown in Fig. 76, so that its head will touch the head of the fastener Z when the box is closed.

The Detector.—The detector is easily made with

the piece of galena, the strip of brass, and the screw-cap. Take the screw-cap and cut four slits in the side to about half-way down as in Fig. 77. Bore a hole in the bottom of the cap and fasten it down to the inside part of the box with a small fastener as B (Fig. 76). Then put the crystal in the cap and bend down the edges, thus holding the crystal firmly. Take the strip of brass and cut it into the shape of Fig. 78, boring a hole in the wide end. Then bend it to the shape shown in Fig. 79, and fasten it with a fastener at the proper distance from the cup.

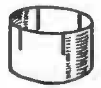

Fig. 77.—Crystal Cup.

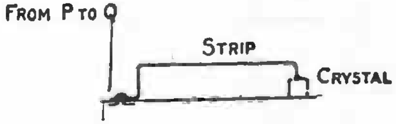

Fig. 79.—Detector.

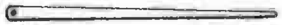

Fig. 78.—Detail of Detector.

Wiring Connections.—Having completed the tuning coil and the detector the next thing is the wiring. Through the outer cover in the bottom corner bore a small hole T (Fig. 75), and through it thread the wire from P. Run wire along the inside of the outer cover so that it comes out at W at the other end, then join it to Q. Next attach a wire to the cup B and connect it to R on the inside. Then attach a wire to the strip A and connect that wire to the fastener S ; connect the fastener S to the fastener F, and attach a wire to Q on the inside and join it to E.

If desired, a small knob can be glued on to the box between E and F for the purpose of tuning.

The head-phones are attached to the terminals R and Q, the aerial is attached to F and the earth to E.

Operation.—In use the point of the strip A should be moved about until signals are heard. Tuning is then effected by pulling the inner part of the box out.

CHAPTER IX

Crystal Circuits

THE illustrations (Figs. 80 to 85) in this chapter show diagrammatically (as distinct from conventionally) the common circuits used in crystal sets.

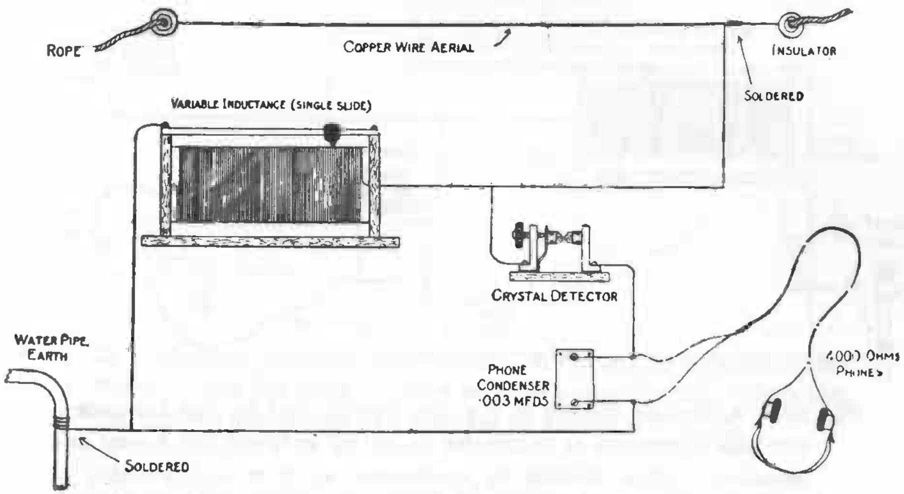
Rope
Copper Wire Aerial
Insulator
Soldered
Variable Inductance (Single Slide)
Crystal Detector
Water Pipe. Earth
Phone Condenser ·003 MFDS
4000 Ohms Phones
Soldered

Fig. 80.—**Diagram showing the Method of Connecting up a Single-slider Tuning Inductance to a Crystal Detector. This manner of Coupling is termed "Direct-coupling."**

There are various types of tuning devices in general use, the more common, perhaps, being the single-layer solenoid type. These may be divided up into two classes—the "tapped" and the sliding contact tuners. Dealing with the single coil direct-coupled tuners, the most effective as regards fine tuning is doubtless the single slider tuning coil as illustrated in Fig. 80. Here it will be seen that tuning is effected by simply sliding a spring contact along the coil. It is possible to obtain fairly accurate tuning to within one complete turn of wire on the coil. For finer tuning a variable condenser is often inserted in the aerial system, in series with the tuning coil as in Fig. 81. This condenser certainly facilitates short-wave tuning, but is often undesirable owing to its effect in causing unbalanced tuning. Another type of sliding contact tuner is the two-slider. The original object of a two-slider coil was to form an auto-transformer coupling which would give a reasonable transformer step-up in signal strength without the usual losses attendant to the use of a loose-coupled air-core transformer. The wiring of this two slider is shown by Fig. 82. Here it will be seen that the one coil governs two circuits—the aerial, one portion of the coil and earth being one circuit, and the crystal, phones, one section of the coil and earth, the other circuit. The aerial to earth circuit is the primary, and the crystal, coil, and telephone circuit forms the secondary. Fairly correct tuning is obtained from the primary circuit, but when the secondary circuit is adjusted with a greater number of turns of wire to obtain the necessary transformer step-up, the secondary circuit becomes out of "tune." To obviate this, a variable condenser is inserted in parallel across the secondary section of the coil as in Fig. 83. By means of this condenser, accurate

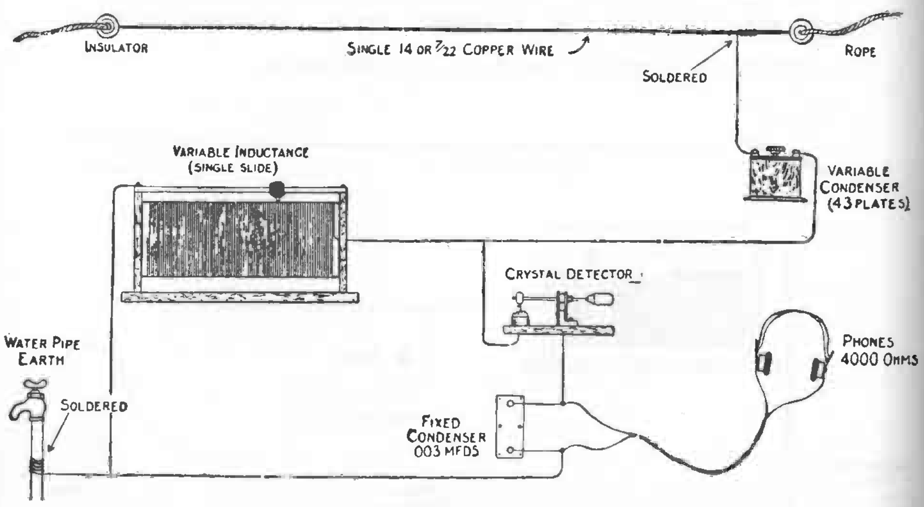
Insulator
Single 14 or 7/22 Copper Wire
Soldered
Rope
Variable Inductance (Single Slide)
Variable Condenser (43 Plates)
Crystal Detector
Water Pipe Earth
Soldered
Phones 4000 Ohms
Fixed Condenser 003 MFDS

Fig. 81.—A Similar Circuit to Fig. 80, except that in this instance a Variable Condenser is Connected in Series between the Aerial and Receiver. This Method is Inefficient as it is not possible to obtain Correct Tuning.

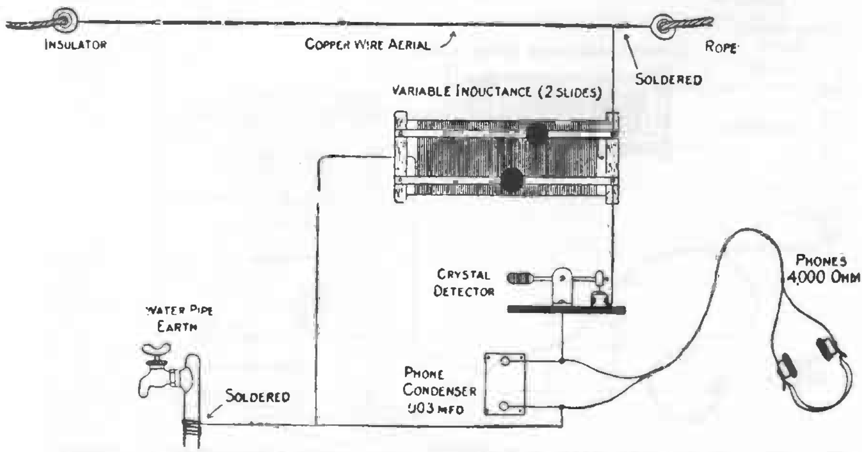
Insulator
Copper Wire Aerial
Rope
Soldered
Variable Inductance (2 Slides)
Crystal Detector
Phones 4000 Ohms
Water Pipe Earth
Soldered
Phone Condenser 003 MFD

Fig. 82—Diagram showing One Way of Connecting up a Double-slider Tuner. This Coupling is known as "Auto-transformer" Coupling. The Circuit above may be Improved as shown in Fig. 83.

secondary tuning is possible, allowing scope at the same time for any reasonable transformation ratio of turns between the primary and secondary sections of the tuning coil. Provided accurate tuning is employed, this type of tuner is probably the most desirable from the point of view of the crystal user.

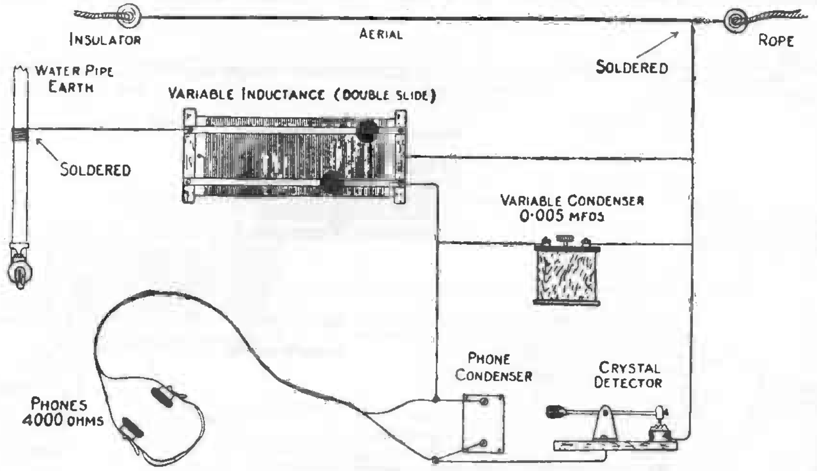
INSULATOR
AERIAL
ROPE
SOLDERED
WATER PIPE EARTH
VARIABLE INDUCTANCE (DOUBLE SLIDE)
SOLDERED
VARIABLE CONDENSER 0·005 MFDS
PHONE CONDENSER
CRYSTAL DETECTOR
PHONES 4000 OHMS

Fig. 83.—A Circuit with a Variable Condenser Connected in Parallel across the "Secondary." A "Transformer Step-up" Effect between Primary and Secondary is Produced which gives Selective Tuning and very Efficient Results.

Another method of tuning is that known as "tapping" the coil. Fig. 84 illustrates the single switch-arm type. Correct tuning, however, is not possible unless the tappings are accurately taken at calibrated points for certain wavelengths. A tuning condenser may, however, be connected between the aerial and earth terminals of the coil to facilitate accurate tuning. A more efficient type of tapped coil is that shown by Fig. 85 where two switch-arms are employed. One

switch-arm is designed to make contact over a portion of the coil which is tapped in units of turns, the other switch-arm making contact over the remainder of the coil which is tapped in sections each equalling the total number of turns covered by the unit switch-arm. Suitably designed, and for short wavelengths, it is possible with this type of tuner to obtain correct

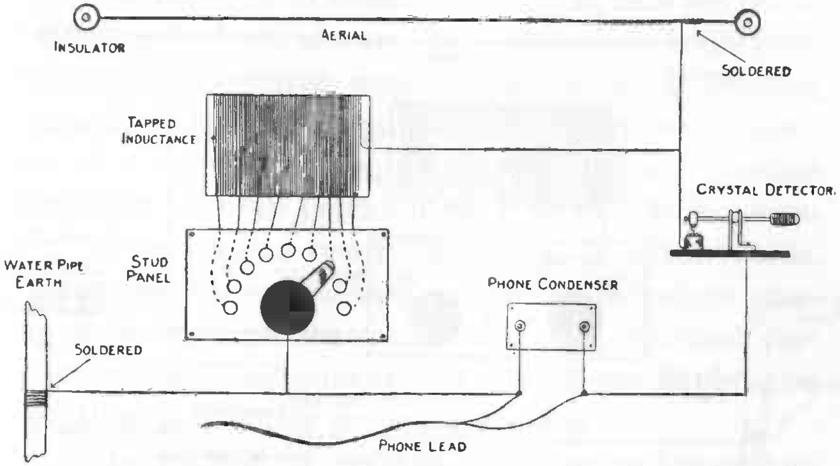
INSULATOR
AERIAL
SOLDERED
TAPPED INDUCTANCE
CRYSTAL DETECTOR.
WATER PIPE EARTH
STUD PANEL
PHONE CONDENSER
SOLDERED
PHONE LEAD

Fig. 84.—Diagram showing the Connections for a Tapped Inductance Employing only One Switch-arm. This Circuit does not allow of very Fine Tuning.

tuning to within one turn of wire on the coil. This type of tuner has one great advantage over the slider method, inasmuch as the wire on the coil is not worn away or cut through by constant use of a slider. These fine particles of copper, worn from the coil by the slider, tend to bridge neighbouring turns of wire on the slider coil. In the tapped coil this trouble is eliminated, but the advantage is lost if the tapping studs are poorly insulated. On each diagram will be seen the necessary crystal detector for detecting and rectifying the signals, and also the telephone condenser and the telephones. The telephone condenser is not absolutely an essential instrument, as the self-capacity of the telephone windings and the telephone leads is often adequate. However, it will generally be found that a fixed condenser of capacity between ·001 and ·005 microfarads will considerably improve the note of the signal heard in the telephones. All bare wire connections should be thoroughly clean and soldered, all terminal connections should be clean and tightly bolted. Do not grind or break crystal surfaces unnecessarily and periodically wash the surface of the crystal with methylated spirits by means of a small brush.

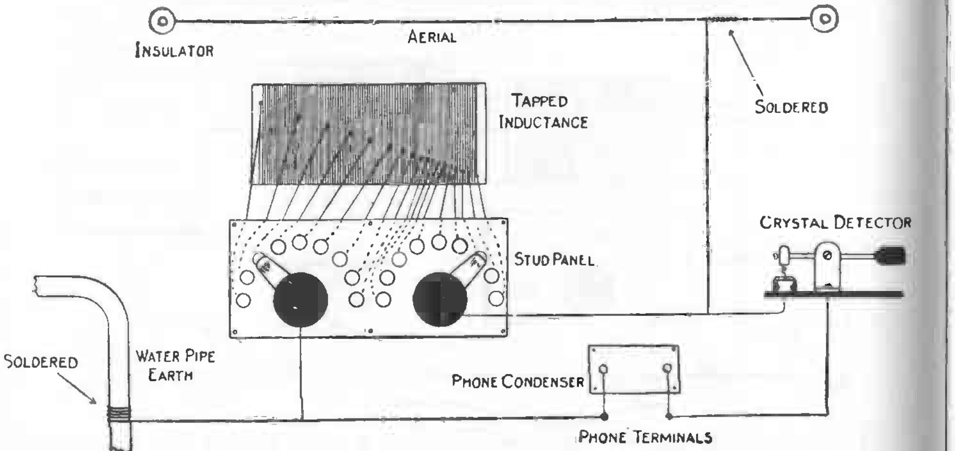

Fig. 85.—An Effective Means of Wiring-up a Tapped Coil to Two Switch-arms which Facilitates Fine Tuning.

CHAPTER X

How Crystals Work

What the Crystal Does.—Wireless signals are transmitted in the form of waves created in the ether, which, striking the receiver aerial, cause small alternating currents to travel up and down it. As these currents are of a very high frequency, the ordinary wireless telephone receiver cannot detect them. It is necessary, therefore, to rectify them, or, in other words, change them into unidirectional currents, before passing them through the telephones. The rectified currents produce a movement of the telephone diaphragms and thus the signals are made audible.

Certain kinds of crystals, such as carborundum, zincite, bornite, silicon, galena, molybdenite and iron pyrites (see Figs. 86 to 93, which show crystals enlarged to about three times normal size), have the peculiar property of being able to rectify small alternating currents. An explanation is given to show how they do this.

How a Crystal Rectifies.—A crystal rectifies because it allows current to flow through it in one direction and does not, to any appreciable extent, allow current to flow through it in the opposite direction. It is practically a uni-directional conductor. The ratio of currents flowing in opposite directions through

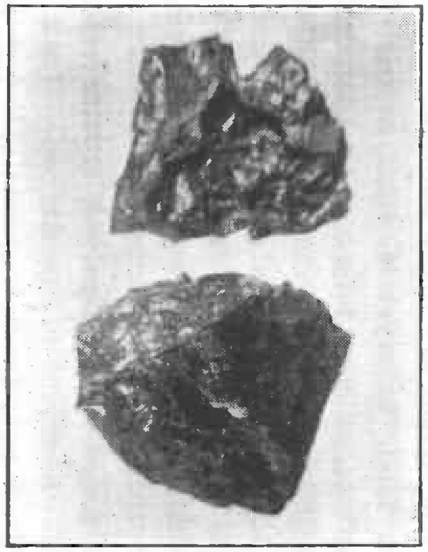
Fig. 86.—Silicon.

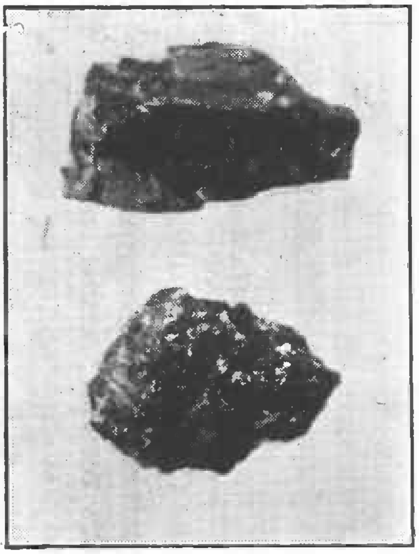
Fig. 87.—Galena and Special Galena.

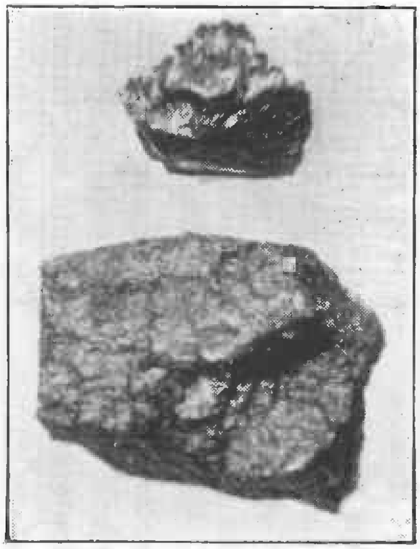
Fig. 88.—Bornite.

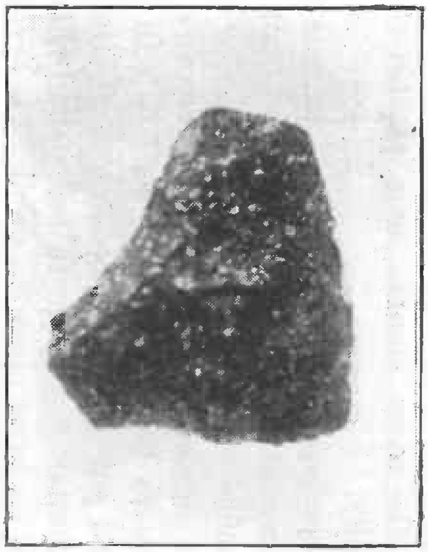
Fig. 89.—Radiocite.

a good crystal should be about 40 : 1. It will be understood that the crystal acts electrically, very much in the same manner as a ratchet-wheel acts mechanically.

Carborundum Crystals.—Carborundum is probably the best crystal for all-round use. It is not affected by mechanical vibration nor by fairly heavy electrical discharges across it, and it is very sensitive. This crystal has to be used in conjunction with a 4-volt battery and potentiometer, because it requires a small initial potential or voltage to be applied across it to bring it to its most sensitive point. Carborundum is used with a flat steel surface, the rectification taking place at the point of contact. This crystal is manufactured by fusing carbon and silicon together in an electric furnace. In practice three kinds of carborundum crystal are met with : (*a*) a hard dark variety, having great metallic lustre which has poor rectifying properties ; (*b*) a soft crystal of a pale-green colour, a colour due to the presence of copper and iron salts (this type of crystal is a very poor rectifier) ; (*c*) a dark grey crystal which will be found to give the best results as regards sensitivity and stability of action.

The Perikon Detector.—The most popular crystal detector is without doubt the perikon detector. This is a combination of either zincite and bornite or else zincite and tellurium. The latter combination is usually found to give the better results. Zincite, a zinc-oxide ore, is dark red in colour ; bornite is a compound of copper and iron sulphides, being bluish grey

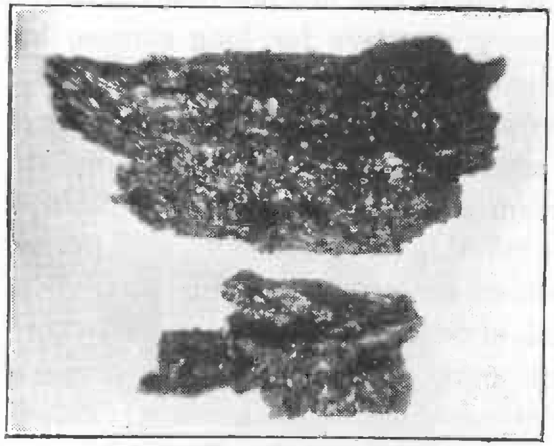

Fig. 90.—Carborundum.

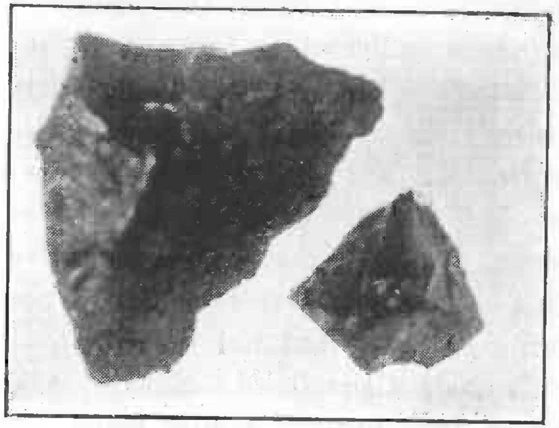

Fig. 91.—Copper Pyrites.

in appearance and fairly heavy. Tellurium is a light-coloured metal. The perikon detector does not require a potentiometer because the voltage of the incoming signal is sufficient to bring it to its sensitive point. It has the advantage of being fairly easy to adjust. Also when a crystal becomes slightly insensitive it can be scratched with a knife and a new surface used. To obtain the best results it is usually found that the crystals must make a light contact with each other. This detector is more sensitive in action than the carborundum type, but not so robust.

Other Crystals Considered.—Silicon, galena, molybdenite and iron pyrites, whilst being serviceable for experimental work, do not give such satisfactory general results as either carborundum or the perikon detector. All these crystals are minerals. Silicon is light grey in colour and has not many sensitive points. Galena is very sensitive for long ranges, but strong signals destroy its sensitivity. It is bluish black in colour, has a metallic lustre, and is heavy. When using this crystal it is usual to have a graphite contact in conjunction with it. Molybdenite is dark in appearance and is suitable for the reception of strong signals; it is convenient for use in portable sets. Iron pyrites is a cubical crystal, being very metallic and heavy. It possesses many of the properties of carborundum but is not so sensitive.

Finding the Best Crystal.—The only sure way to obtain a good crystal is to search for sensitive points amongst several specimens. It is wise to remember that sensitivity is not everything; a good crystal

Fig. 92.—Permanite.

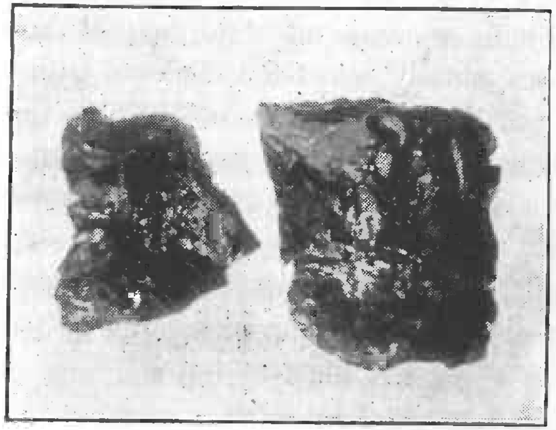

Fig. 93.— Zincite.

should have the qualities of being constant in action and of not being affected by atmospheric disturbances. Such crystals as " radiocite " and " permanite " are usually sensitised forms of galena. They are very sensitive but not constant ; a fine copper wire contact is used in conjunction with them.

There is plenty of instruction and amusement to be had with a good crystal set. The range of a crystal set is about 300 miles for ordinary ship stations and 20 or 30 miles for high-power telephony.

CHAPTER XI

Making a Buzzer

By A. W. HULBERT

THIS chapter deals with a new type of buzzer, one which utilises a circular armature or diaphragm instead of the usual flat strip of metal. Due to the fact that this diaphragm is supported all round, it is possible to have the magnet almost touching the diaphragm, resulting in a more sensitive instrument and one capable of producing an extremely high-pitched note.

The overall dimensions of the instrument are also considerably less than the standard instrument, and having a minimum of working parts it is very easy to make.

Constructional Details.—Fig. 94 is a section of the instrument. A circular ebonite case B, with a snap-on lid, is fitted with an electro-magnet M, which is wound on a strip of soft iron P which serves the double purpose of a core for the magnet and also a support. The soft iron strip P is bent round and mounted on the bottom of the ebonite case with the aid of two screws, the opposite arm to that carrying the coil M being in contact with the diaphragm D, which rests on a ledge turned in the ebonite case. The distance between the diaphragm and the pole of the magnet M is approximately $\frac{1}{32}$ in.

The lid of the ebonite case B is undercut to enable the diaphragm to vibrate. Mounted on the top is a contact screw S and a terminal F. The other terminal (not shown in Fig. 94) is connected to one end of the magnet winding, the other end being earthed to the polepiece P. The passage of the current through the instrument is therefore from the battery or accumulator to the terminal F, and the metal plate to the contact screw S, across the contacts and the diaphragm D to the pole-piece P, and through the magnet winding to the terminal T (Fig. 95), and thence back to the battery.

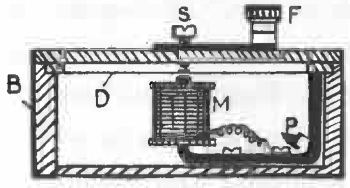

Fig. 94

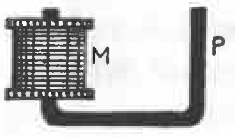

Fig. 96

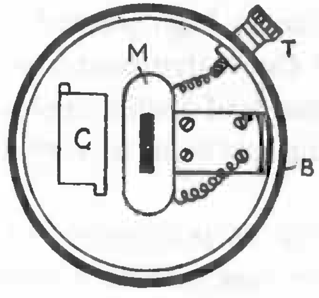

Fig. 95

Fig. 94.—Section of Complete Buzzer.

Fig. 95.—Details of Interior showing Connections.

Fig. 96.—Pole-piece and Coil.

Due to the fact that the snap-on lid clamps the diaphragm securely all round, it is possible, as mentioned previously, to have the diaphragm very close to the magnet, thereby producing a high-pitched note.

The Magnet.—With regard to the construction of the magnet M, this is wound on a metal former made up by soldering two metal cheeks to one arm of the metal strip P. This strip should measure $\frac{5}{8}$ in. wide by $1\frac{1}{2}$ in. long, and may be cut from sheet-iron $\frac{1}{2}$ in. thick. This is bent up as shown in Fig. 96; that portion forming the core of the magnet should be insulated by means of waxed silk, which should also be gummed on to the inside of the cheeks. The magnet is wound with No. 36 d.s.c. wire, the last layer but one being covered with a strip of varnished silk. The final layer of wire should be wound on very carefully and secured by looping the last turn under itself, the complete coil then being soaked in melted paraffin wax.

The Pole-pieces.—The complete coil and iron core P should now be placed on a surface plate or other flat base and the heights of the two pole-pieces carefully measured. The arm or pole-piece on which the coil is wound must be $\frac{1}{32}$ in. shorter than the other. The height of the other pole-piece must be such that it just reaches a shade higher than the ledge in the ebonite case B, so that when the metal diaphragm D is in position direct contact is made with the pole-piece P.

The Diaphragm.—The diaphragm D may be one taken from an old telephone receiver, the only addition being a contact stud soldered in the centre. Both this and the contact on the screw S should be of german-silver or other suitable alloy. The actual sparking may be considerably reduced by fitting a small condenser C (Fig. 95) in the case and connecting this across the contacts.

The Case.—The ebonite case B can easily be turned up by anyone possessing a lathe, or one may be purchased from an electrical dealer. The battery voltage may be anything from 2 to 4 volts, and when the buzzer is used exclusively for testing faults in the circuits the terminals may be dispensed with and a length of twin flexible lead soldered on permanently.

CHAPTER XII

Receiving C.W. Signals on a Crystal Set

By E. H. ROBINSON

It is frequently stated that a simple crystal receiving set will enable the listener only to pick up spark stations and telephony, and that valves will have to be installed before continuous-wave (C.W.) signals can be satisfactorily received. This is true to a certain extent, but this chapter shows how some of the louder C.W. stations can be received by means of a very simple addition.

C.W. Signals.—The reason why C.W. signals cannot be heard on ordinary crystal circuits is that when the key at the transmitting station is pressed down a continuous rectified current passes through the head telephones of the receiver, not an interrupted one as in the case of spark and telephony, with the result that no sound is heard except perhaps a faint click when the key is depressed or raised. If, however, an interrupting device, capable of making and breaking the circuit at a high rate, say 500 times a second, is inserted between the crystal and the phones, then any continuous current in the phones will be broken up into a musical note which is clearly audible.

The "Tikker."—Such an arrangement is shown diagrammatically in Fig. 97. In this figure T is a rapidly rotating cog-wheel against which a small metal tongue rests lightly. The use of such an interrupter is the invention of the Danish scientist Pedersen, who used it for reception of signals from the famous arc-transmitting system which bears Poulsen's name. Pedersen called the interrupter a "tikker," and this is the name by which the device is still known.

The cog-wheel T may conveniently be the fastest-moving wheel in the mechanism of a clock. It will be necessary to remove the escapement so that the wheels rotate freely. If the spring is now wound up and the mechanism allowed to run it will go at a tremendous speed and be run down in a few seconds; a governing device must, therefore, be added, and this can easily be done by soldering, or otherwise fixing, a small vane cut out of thin sheet brass to the most rapidly revolving spindle as shown in Fig. 98. Owing to the friction of the air with this vane the mechanism will not be able to "race," and will run nicely for a considerable time. The actual size of the vane for best working will depend on the nature of the clockwork and will have to be determined in individual cases by trial; the larger the vane the slower the mechanism will run. In the writer's own tikker the vane is only ¾ in. across and ¼ in. wide, but it is sufficient to ensure a steady run of about twenty minutes. The only other thing that remains to be done is to fix a light tongue which has to make contact with the cogs of the fastest wheel in the train.

The method of mounting is shown in Fig. 99. A is an ebonite strip, sufficiently thick to be rigid, through which a hole has been made at each end. By means of a bolt and fly-nut one end of the ebonite strip is fixed to a point C in the frame, and that at the other end will come near the wheel T, a hole being made at C to accommodate the bolt. The tongue B is cut to the shape shown in Fig. 100 out of springy sheet brass about the thickness of drawing-paper. An ordinary pair of scissors will cut the brass quite easily, and the

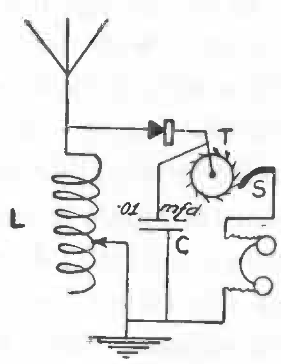

Fig. 97.—Circuit Diagram of Interrupter.

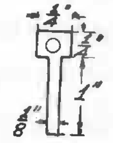

Fig. 100. —Metal Tongue.

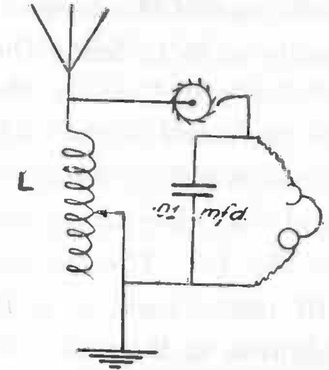

Fig. 101.—Pedersen System of C.W. Reception.

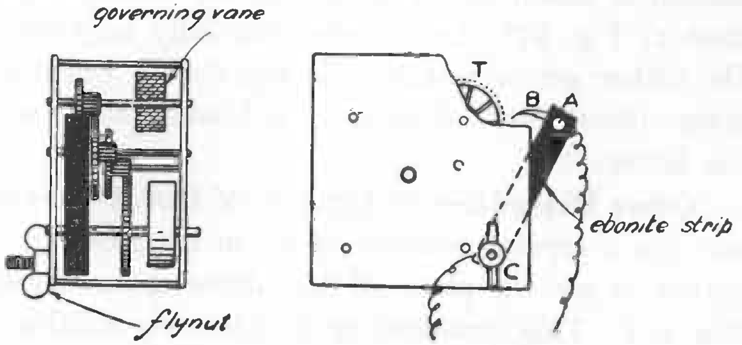

Fig. 98

Fig. 99

Figs. 98 and 99.—Front and Side Elevations of Interrupter.

hole shown in the wide end may be pushed through by any sharp tool. The tongue is fastened to the upper end of the ebonite strip A by means of a nut and bolt through the hole which has been made there, and the thin part of the tongue is given a twist through a right angle so as to bring the free end square with the teeth of the wheel T. A piece of flexible wire is connected to the clock-frame and another to the tongue ; one of these wires is connected to one of the phone leads and the other to one of the detector terminals as shown in Fig. 97. The condenser C, having a capacity of about .01 microfarad, is shunted across the tikker and telephones, as it greatly improves the signals. Some little adjustment of the speed of the tikker may be necessary before good results are obtained. When a C.W. station is tuned in (by means of the variable inductance L, Fig. 97), the detector carefully adjusted and the tikker set in motion, the signals are heard as an intermittent musical note or a hissing note, usually the latter.

Other Methods.—As a matter of fact, Pedersen did not use a crystal detector at all in his receiver. The tikker is put in place of the detector as shown in Fig. 101. This arrangement is extremely reliable, but, in the writer's experience, is not quite so sensitive as when a crystal detector is included. However, as crystal detectors are very liable to go out of adjustment the writer would recommend the method shown by Fig. 101.

Where a crystal detector is to be used, either a good piece of zincite touched by a sharp piece of

bornite or copper pyrites, or a piece of galena touched by a small spiral of 40-gauge copper wire, will be the most suitable. Crystals, like carborundum, which require a battery and potentiometer, cannot be used satisfactorily in conjunction with a tikker. Much depends on getting good specimens of crystal.

A gramophone can be used as a tikker by connecting one wire to the works and allowing the other to rub lightly against the edge of the revolving turntable

CHAPTER XIII

Making a Wavemeter from a Crystal Set

The components of a crystal set—or some of them—can be assembled to produce a wavemeter that gives approximately truthful results.

With a coil of known inductance, a variable condenser, crystal and phones, a useful wavemeter can be rigged up. The circuit is as shown in Fig. 102.

First calculate from the formula:

$$1{,}885 \sqrt{L \times C}$$

(which simply means 1,885 multiplied by the square root of the product of the inductance and capacity) the wave-length range the coil and condenser will give. For example, condenser ·0001 to ·001, and coil, say, 2,000 mhy. = 850 L C to 2,550 L C approx. Next prepare a chart on squared paper (Fig. 103) as follows: Mark off the condenser degrees equidistant on the vertical edge. Start at, say, 900 metres and square the units as follows: 9^2, 10^2, 11^2, 12^2, etc., = 81, 100, 121, 124. 100 − 81 = 19; 121 − 100 = 21. There will be an increase of two in every case after the first number of squares is found; the others need not be worked out. The wave-length scale starts at 900 metres, therefore count 19 squares and mark 1,000 m. from this mark. If 19 squares is too large and half is taken then the increase will be halved.

Calibrating.—When the scale is prepared start the calibrating as follows: Set the receiver to a known wave-length adjustment within the scale or get a station on a known wave that can be relied upon, say G.F.A. on calibration waves. Then listen in on the wavemeter circuit, coupling it closely at first to the C.W. receiver and varying the condenser until a rustling noise is heard, which will be very sharp tuning on the shorter waves. This point will be more quickly found if the aerial terminal of the receiver is tapped with a moistened finger. The clicks heard when a set is oscillating will be heard in the wavemeter when it is

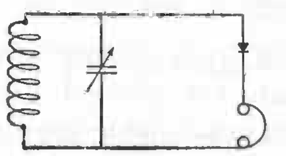

Fig. 102.—Circuit Diagram of Wavemeter.

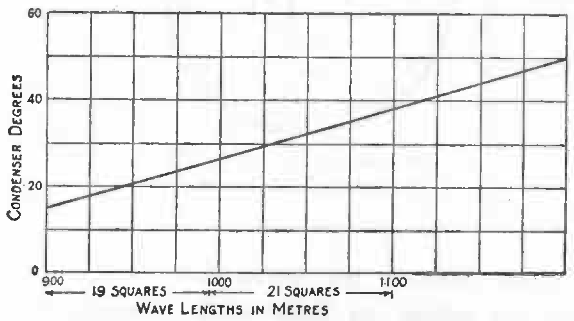

Fig. 103.—Chart for Purpose of Calibration.

in tune and sufficiently coupled. The tuning must be completed without the finger or capacity effect will give a false tune, though the difference is usually small.

The wave-length is marked on the chart at the point corresponding with degrees—wave-length. Two or more such points are plotted and joined and produced, when any wave-length between the limits can be found. To use the wavemeter it is set to the wave-length, and while listening in the receiver is tuned until the hiss is heard ; then transfer on to the receiver and listen for signals. Different values of coils may be used and scales plotted.

If a calibrated condenser is available the chart can be plotted without the aid of the receiver. A buzzer can be used, when the crystal and phones will not be necessary.

CHAPTER XIV

Converting Low-resistance Phones

Meaning of Resistance.—With a few exceptions, it is necessary that the telephone headgear for wireless work be wound to a high resistance, anything from 60 to 4,000 ohms per receiver being general. The statement " high resistance " is perhaps apt to mislead the reader ; what is actually meant, of course, is

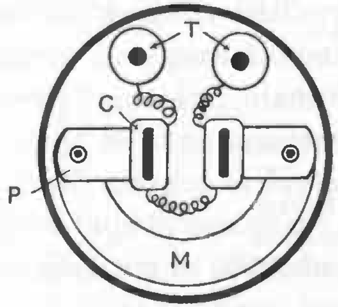

Fig. 104.—Diagram of Interior of Receiver.

that the magnet coils must be wound with a great number of turns of fine wire in order to produce a maximum magnetic effect with very little current. The resultant resistance is, therefore, merely a necessary evil, unavoidable in all magnetic devices which are more or less voltage operated.

Another point about wireless phones is that the diaphragms are much thinner than those used on ordinary commercial telephones. This makes them more sensitive to small variations in the magnetic force from the coils. At the same time, the diaphragms are more easily damaged, an excessive current causing them to buckle.

Construction of Receivers.—The general arrangement of a receiver of the common watch-pattern type is shown in Fig. 104, the ebonite earpiece and diaphragm having been removed. A permanent magnet M of semicircular shape is screwed to the outer containing case, and mounted on the poles are two pole-pieces P, which are bent at right angles and carry the magnet windings C. The ends of the windings are connected direct to the terminals T, which pass through insulated bushes in the metal case and serve to connect the coils C with the main receiving instruments.

The coils C when removed from the ordinary receiver will be found to contain fairly thick wire. The task is to strip them and rewind with much finer wire to a resistance of 2,000 ohms each receiver. Fig. 105 shows the general appearance of the coil when wound, the figures in the sketch corresponding with the reference letters in Figs. 104 and Figs. 106 to 108.

It will be seen that with such a small coil and with only a bent strip of metal P to hold, winding the empty bobbin by hand would be an extremely awkward job ; in fact, it would be almost an impossibility to wind the wire with sufficient evenness to get anything like the required amount on the coils.

Again, the wire being so fine it is necessary to use

great care to avoid breakage, which would mean starting all over again, and also the coil being so small, it would be almost impossible to hold it long enough to get all the wire on at once without the hand becoming cramped.

There are several ways of getting over this difficulty, such as with a winding machine.

Device for Winding.—Perhaps the best method is to use a small electric motor. The arrangement is clearly shown in Fig. 108, E being a small electric 4- or 6-volt motor, using current from an accumulator and having a variable resistance in the circuit to regulate the speed. Clamped to the end of the armature shaft is a circular block of ebonite or fibre B on which the telephone coil C is mounted, being secured in position by means of the clamping screw C S, which passes through the holes in the pole-piece. It will be

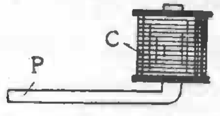

Fig. 105.—Wound Coil on Magnet.

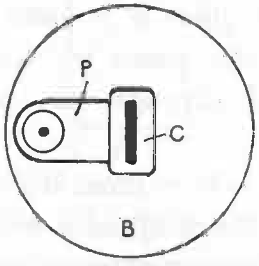

Fig. 106

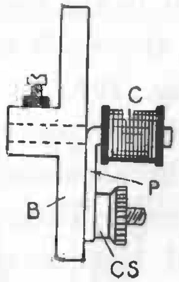

Fig. 107

Figs. 106 and 107.—Magnet and Spool Mounted for Winding.

noticed that the coil C is arranged to rotate as near as possible in the line of the armature shaft, so that, in spite of the coil itself being rectangular in shape, it will revolve with as little wobble as possible. This is very necessary, as any tendency of the coil to revolve eccentrically will tend to break the very fine wire. The ebonite block B is turned up in the lathe and a hole bored through whilst still in the chuck to take the armature spindle; this can either be a good tight fit in the first instance or the block may be clamped on the shaft by means of a small set-screw, shown in Fig. 107. A second clamping screw C S is fitted, to hold the coil C in position for winding. This screw must be accurately positioned to ensure the coil C running true. The best plan is to hold the coil in position and then mark the place for the screw by means of a scriber passed through the hole in the pole-piece P.

Revolving the Magnet.—It is most important that the hole to take the armature spindle be accurately bored, and also that the armature spindle runs true, otherwise the coil C will wobble and difficulty will be experienced, not only in the wire continually breaking, but also in getting it to fill in close to the cheeks of the coil. If the motor E is of the type in which the armature shaft extends at both ends, the opposite end to which the block B is mounted may be used to carry a small knurled knob fitted with a short piece of brass wire about $\frac{1}{8}$ in. in diameter to form a handle. This will prove particularly useful for controlling the motor or for winding portions of the coil by hand.

Winding.—The motor should be connected up to the accumulator and regulating-resistance and tested to see that it is in running order. The block B is then permanently fixed to the armature shaft and one of the coils to be wound clamped in position and trued up. Some fairly thin shellac varnish should be near at hand, together with a strip of silk ribbon, this latter cut to the width of the inside of the coil.

A single layer of silk ribbon is stuck on the core of the coil with shellac varnish, one layer of wire is then wound on and covered with a second layer of ribbon soaked in shellac. The free end of the wire is then carefully brought outside the coil and secured to the clamping screw C S for the time being. It is taken for granted, of course, that the sides of the coil C are already insulated with varnished ribbon or paper; at the same time it is a good plan to stick a small piece of varnished ribbon over the free end of the wire, to insulate it from the subsequent layers of wire

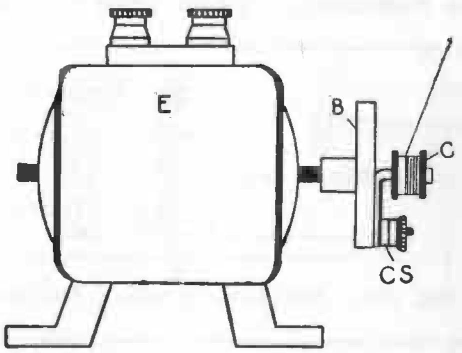

Fig. 108.—Motor with Magnet Mounted.

when these are wound on. The bobbin on which the wire is purchased should be mounted on a small brass rod so that it is free to rotate, and with the wire resting lightly across the second finger of the right hand, the motor should be started on the first speed, gradually increasing the speed of the motor.

If one has access to a gramophone another method of winding is to obtain an empty cotton reel which has a hole through its centre just large enough to fit over the projecting portion of the spindle of the gramophone turntable. Then take the bobbin which

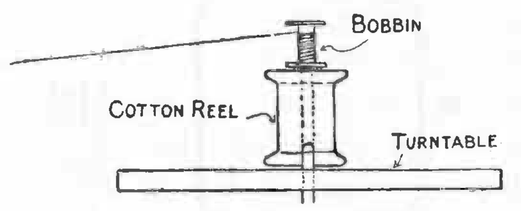

Fig. 109.—Telephone Winding Bobbins.

is to be wound and fix it to the top of the reel by means of a spot of glue or sealing-wax. The wire can then be attached to the bobbin and the turntable allowed to revolve as in Fig. 109. The speed at which it revolves can, of course, be regulated by the controller which is fitted to most gramophones.

When the coil is almost full a single layer of silk ribbon should be put on, the last layer of the wire being wound over this, leaving about six turns from the end empty. The top layer is then held in position with the thumb and finger, while an assistant soft-solders a short length of double silk-covered No. 28

S.W.G. wire to the winding. The last six turns will consist of thick wire, when the coil should be covered with two layers of ribbon and the whole thoroughly

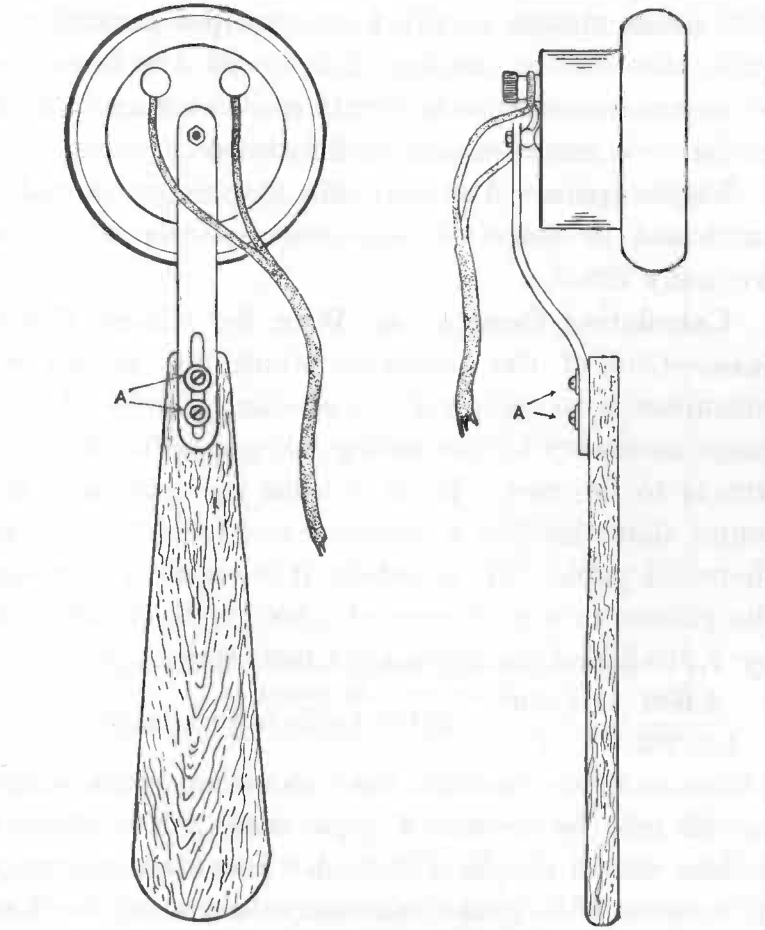

Figs. 110 and 111.—Telephones Fitted with Handles.

soaked in shellac varnish and allowed to dry before touching again. The second coil should be treated in the same way.

Assembling.—When both coils are dry they may be remounted on the magnet poles in the receiver case, and the winding connected to the terminals.

Having connected up to the terminals T the whole of the case should be filled with melted paraffin wax up to the level of the top of the coils, the loose connecting wires will thus be firmly embedded and a break in the wire rendered almost impossible.

Diaphragms.—A special thin diaphragm should be purchased in place of the comparatively thick one originally fitted.

Calculating Quantity of Wire for Given Resistance.—One of the problems which the reader will encounter is to calculate the amount of wire of given gauge necessary for rewinding. Suppose No. 42 gauge wire is to be used. From a table of resistances it is found that this has a resistance of 1,910·5 ohms per thousand yards. If, therefore, it is required to rewind the phones to a resistance of 4,000 ohms, divide 4,000 by 1,910·5 and multiply by 1,000, thus :

$$\frac{4{,}000}{1{,}910\cdot5} \times \frac{1{,}000}{1} = 2{,}093 \text{ yards approximately.}$$

This is a better method than calculating the weight, as will now be shown. Copper wire of No. 42 S.W.G. weighs ·14533 lb. per 1,000 yd. To obtain the *weight* of wire of this gauge necessary to rewind to 4,000 ohms the calculation becomes

$$\frac{4{,}000}{1{,}910\cdot5} \times \frac{1{,}000}{1} \times \frac{\cdot14533}{1{,}000} = \cdot304175 \text{ lb.}$$

The difficulty of gauging this amount of wire is at once apparent, apart from the fact that the weight

given does not and cannot take the weight of the covering into consideration.

Testing Phones.—The test commonly used for phones was by means of the handiest and weakest of batteries—a penny and a two-shilling piece, between which a bit of moist paper was placed. The end of one wire from the phones was connected to the " copper," and then the silver coin touched lightly (and repeatedly) with the end of the other wire.

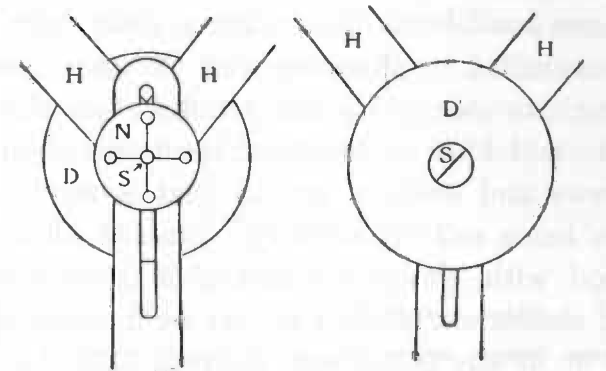

Fig. 112 Fig. 113

Figs. 112 and 113.—Outside and Inside View of Improved Head-band.

Nowadays a similar test is made with a single piece of sheet aluminium, which should be damp. Simply connect one wire of the phone to one end of the aluminium plate, and then touch the plate with the other wire, when, if the phone is properly adjusted each contact will cause a click in the phones.

Another method of testing phones is less difficult and perhaps more efficient than testing by the " weak battery " or sixpence and penny method.

The two leads from the earphones are connected to the two terminals of an ordinary electric bell. No battery whatever is put in circuit. The clapper is pressed forward to the bell and then quickly released. It will vibrate several times before coming to rest. A similar vibration will take place and will be heard plainly in the phones. The explanation of the action is that there is a small amount of magnetism in the magnet of the bell which usually works it.

Handles for Phones.—Ladies often complain of the phone head-band disarranging their hair. This can be remedied by disconnecting the head-band and attaching to short wooden handles, as shown in Figs. 110 and 111. A length of hard dry wood and a few screws and washers are all that is required, the sketches being self-explanatory. Smooth all edges off the wood with glasspaper and give it two or three coats of shellac varnish, or it may be finished in black enamel or in any other way desired. Fig. 110 shows a back view of the phone and link connected to the wooden handle by the screws and washers A, and the side view and general arrangement is clearly shown in Fig. 111. Another method of eliminating the difficulty is to cut out of pasteboard (or other suitable material) two discs 2 in. in diameter. Remove the screws at each side of the head-bands. Pass a screw through the centre of each disc and replace on the head-bands. Fig. 112 shows the outside view, and in this figure D is the disc, H the head-bands, N the nut and S the screw. Fig. 113 (inside view) shows how the attachment shields all moving parts.

CHAPTER XV

The Morse Code

Most owners of crystal receiving sets will be interested in learning the Morse code, for it is interesting to be able to pick up and translate the messages sent between the various ship and shore stations. One excellent method of memorising it is to learn the code in four different groups, the easiest signs being in the first group, which should be learned thoroughly.

First Group.

E · is opposite T —
I · · ,, ,, M — —
S · · · ,, ,, O — — —
H · · · ·

Second Group.

A · — is opposite N — ·
U · · — ,, ,, D — · ·
V · · · — ,, ,, B — · · ·

Third Group.

W · — — is opposite G — — ·
R · — · ,, ,, K — · —
P · — — · ,, ,, X — · · —
L · — · · ,, ,, Y — · — —
F · · — · ,, ,, Q — — · —

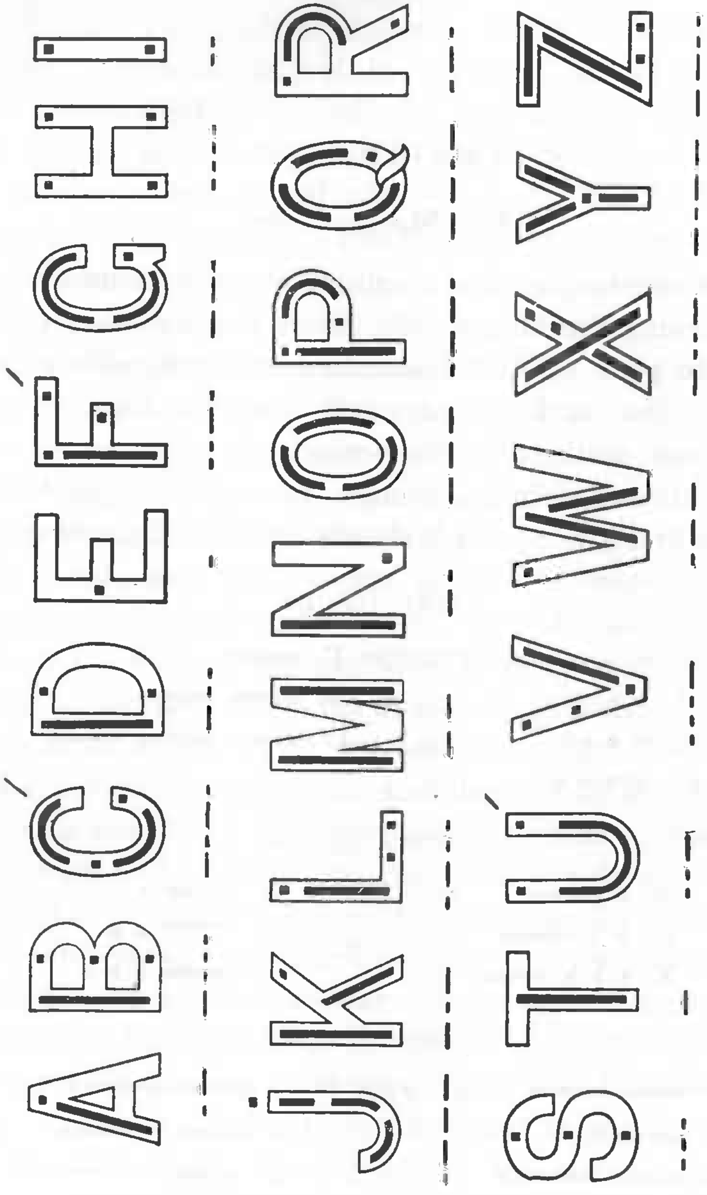

Fig. 114.—Graphical Method of Memorising Morse.

Fourth Group.

C — · — ·
J · — — —
Z — — · ·
} No opposites

Fig. 114 shows another method of memorising the Morse code. The code is learnt rapidly simply by visualising the code letters as superimposed upon the alphabetical letters. On thinking of any letter the mind conjures up a mental picture of it and of the dots and dashes of the Morse code lying on it and actually forming part of its shape. The twenty-six letters can be learnt in as many minutes, but only practice, practice, and yet more practice at sending and receiving can make perfect.

INDEX

AERIAL erection, 2
Auto-transformer coupling, 87

BASKET-COIL receiver, miniature, 73
——, tuner, plug-in, 58
Basket coils, winding, 62, 69, 74
Bornite crystal, 93
Buzzer, making, 99
—— test, applying, 12

CARBORUNDUM crystal, 95
Case for set, making, 40, 58
Circuits for crystal sets, 85
Closed circuit explained, 50
Coil, mounting, 18
——, tapping, 38
Condenser, phone, making, 22, 81
Contact, sliding, making, 18
Continuous-wave signals, receiving, 103
Copper pyrites, 95
Coupling, auto-transformer, 87
Crystal and valve receiver, 68
—— detector, making, 6, 20, 28, 42, 54, 60, 70, 78, 80, 82
——, finding best, 96
Crystals: how they rectify, 91
——, types of, 91

DIRECT-COUPLED circuit, 85

EARTH lead, 4

GALENA crystal, 92

HEAD-BANDS for phones, 119

INDUCTANCE coil (*see* Tuning coil)
Interrupter, 103

LEAD-IN, connecting to set, 6
Lightning switch, purpose of, 4
Loading coils, use of, 24
Loose-coupled set, 45
—— set, sliding arrangement for, 47

MATCH-BOX receiver, 82
Miniature crystal sets, 73
Morse code, memorising, 121

OFFICIAL receiving set, 1
Operation of crystal set, 11, 14, 71, 84

PEDERSEN, 104
Permanite crystal, 97
Phone condenser, making, 22, 81
Phones, calculating quantity of wire for, 118
——, constructional details of, 112
——, converting low-resistance, 111
——, handles for, 120
——, improved head-bands for, 119
——, selecting, 7
——, testing, 119
——, winding device for, 113
Plug-in basket-coil tuner, 58
Pocket receiver, 78
Potentiometer, making, 54
Poulsen, 104

RADIOCITE crystal, 93
Reaction coil, 68
Rectification by crystal explained, 91

SETTING of set, recording, 66
Silicon crystal, 92
Single-coil set, tapped, 33
Single-slider set, 16
—— with semicircular tuner, 25
Sliding arrangement for loose-coupler, 47
—— contact, making, 18

TAPPED single-coil set, 33
Tapping coils, 38
Testing receiving set, 44
"Tikker," making, 103
Transformer step-up effect, 88
Tuning cards, use of, 66
—— coil, basket, winding, 62, 69, 74
—— ——, cylindrical, tapping, 38
—— —— ——, winding, 8, 16, 27, 37, 46, 79, 82

VALVE and crystal receiver, 68
Variable condenser, 50, 52
Vest-pocket receiver, 78

WAVEMETER, calibrating, 109
——, making from crystal set, 108
Winding device for phones, 113

Zincite crystal, 97

PRINTED BY CASSELL & COMPANY, LIMITED, LA BELLE SAUVAGE, LONDON, E.C.4.
F75.623

There's Money in Knowing How!

Knowing how to do "odd jobs" and repairs, and how to make things for the home, means the saving of expense in many directions. Most men could be "handymen" but lack the necessary knowledge.

"WORK" is a long-established weekly which supplies this knowledge in a simple, straightforward manner. Its articles are full of working details and explanatory diagrams, and are written in simple, non-technical language. Every issue contains several pages devoted to solving, in comprehensive answers, the difficulties of readers.

As a practical guide for the amateur it is unrivalled and unique.

As a mine of information it is inexhaustible.

3D
Weekly

BUY A COPY TO-DAY

WORK
A WEEKLY JOURNAL FOR AMATEUR MECHANICS

CASSELLS
La Belle Sauvage, E.C.4

ECONOMIC ELECTRIC LTD.

TELEPHONE: MUSEUM 1055.

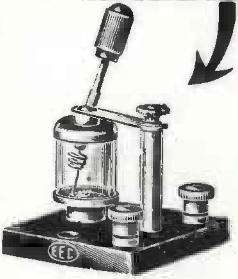

THAT OUR "MAXSIG" CRYSTAL DETECTOR

IT'S A FACT

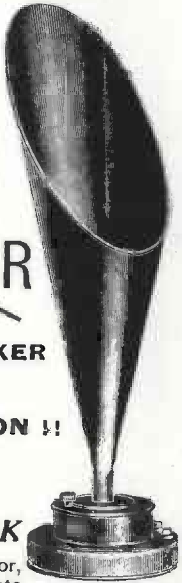

ECONOMIC
Radio
ELECTRIC
EEC

WILL OPERATE A LOUD SPEAKER TEN MILES FROM A BROADCASTING STATION !!

APPLY NOW for the

"EEC" RADIO BOOK

which describes this wonderful Detector, and gives the keenest prices for complete CRYSTAL & VALVE SETS, Component Parts, Material, etc. Undoubtedly the Radio Constructors' *VADE MECUM.* Comprises 40 pages, over 250 illustrations and circuit diagrams for Crystal and Valve Receivers. Send us 4d. in stamps and secure this wonderful publication.

Head Office: 10, FITZROY SQUARE, LONDON, W.1.	*Showrooms:* 303, EUSTON RD., N

Cave of the Crystals

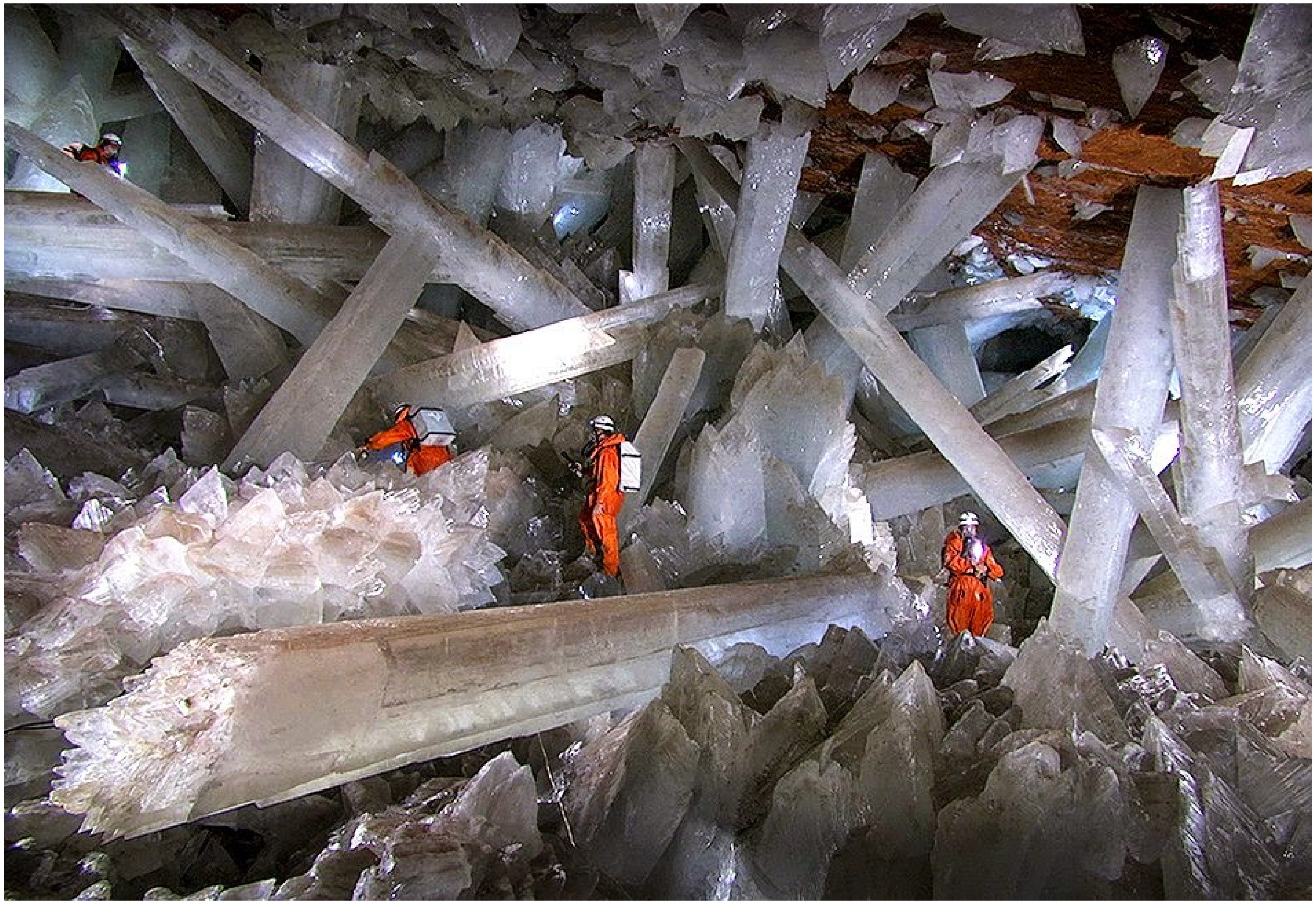

Cave of the Crystals

The Cave of the Crystals is a natural marvel in Chihuahua, Mexico. Its main chamber houses some of the largest natural crystals ever discovered, hence the name. The magma within makes the cave unbearably hot and humid, and thus it remains largely unexplored.

Nevertheless, this subterranean spectacle is unquestionably one of the world’s most remarkable displays of natural mineral formations.

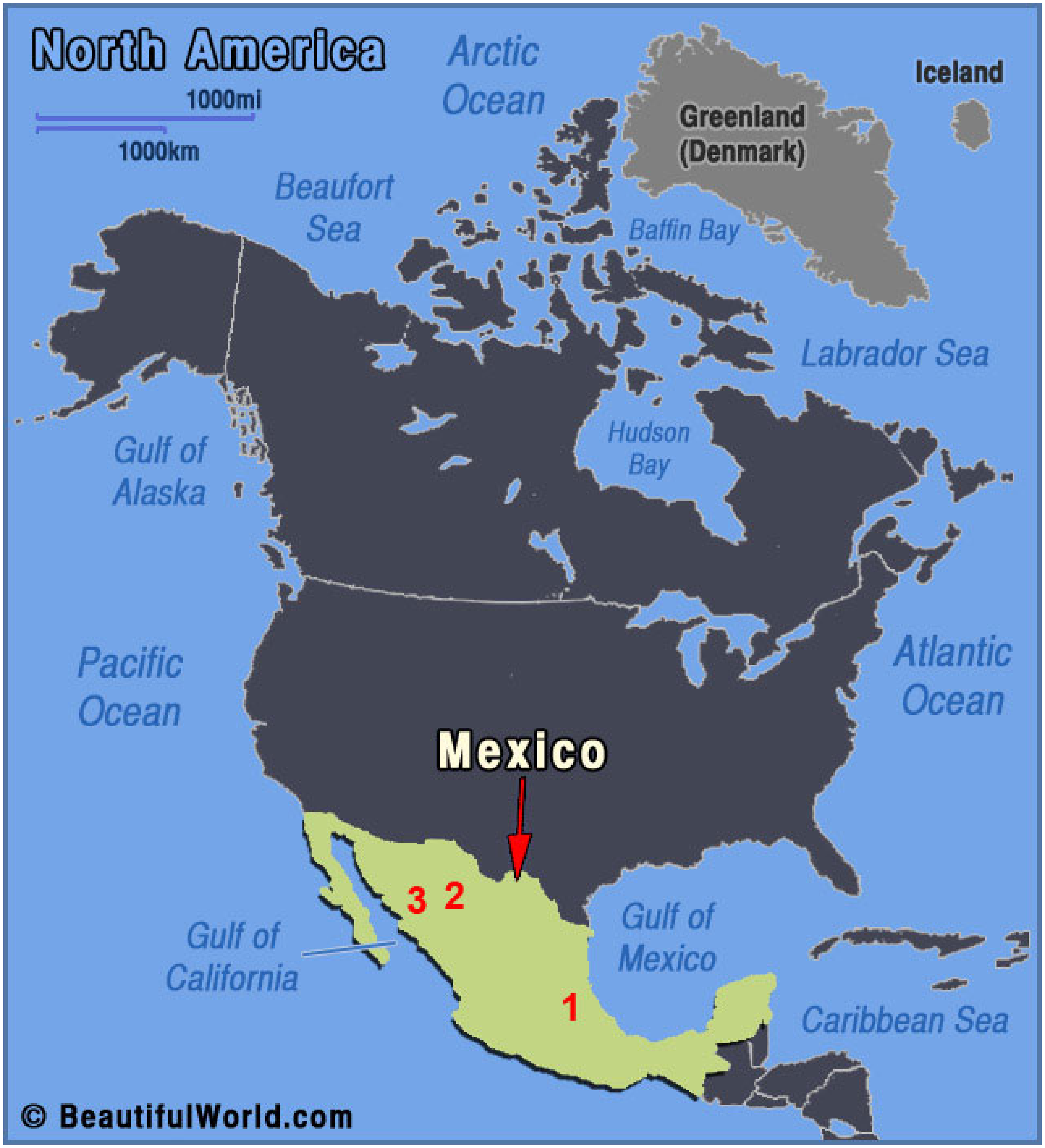
North America
1000mi
1000km
Arctic Ocean
Beaufort Sea
Greenland (Denmark)
Iceland
Baffin Bay
Labrador Sea
Hudson Bay
Gulf of Alaska
Pacific Ocean
Atlantic Ocean
Mexico
3 2
1
Gulf of California
Gulf of Mexico
Caribbean Sea
© BeautifulWorld.com

1 Las Grutas de Tolantongo 2 Cave of Crystals

3 Copper Canyon

Cave of the Crystals Map

Cave of the Crystals Facts

The Cave of the Crystals is connected to the Naica Mine, 300m (980 ft.) below the ground in Naica, Chihuahua, Mexico. Despite the Naica Mine having been worked for centuries, the Cave of the Crystals remained undiscovered until the year 2000, when two miners, pursuing traces of silver, broke into the extensive subterranean chamber. Just days later, an iron door was erected in order to protect the entrance.

Whilst the cave is referred to singularly, as the ‘Cave of the Crystals’ (Cueva de los Cristales), it is in fact made up of three co-joined caves, with one far larger than the others.

This gigantic cavern is home to towering stacks of selenite; rising so tall that humans appear miniscule in comparison. In fact, the tallest crystals have been measured at 11m (36 ft.) high, weighing approximately 55 tonnes.

These giant crystals have experienced steady slow growth over the past half a million years. Approximately 2-3 miles (3-5km) under the surface is a giant pool of magma. The caverns became flooded over time as a result of groundwater seeping in. This water was rich in gypsum minerals, and the heat generated by the magma worked like a cooker, maintaining a steady heat for 500,000 years. The gypsum in the water then transformed into the translucent selenite crystals that we see today. Whilst up until now the crystals have grown steadily and slowly, they will grow no further unless the cave floods once again.

Beneath the cave lies a pool of magma, rising up from the Earth’s core. This heat from this magma makes the cave unbearable for any substantial amount of time. In fact, no human can spend more than a mere few minutes inside the cave. Temperatures of 150°F (65.6°C) have been recorded in the main chamber, whilst the smaller caverns are slightly cooler at 100°F (37.8°C). To add an extra layer of danger, conditions are close to 100% relative humidity.

The story of a miner who managed to bypass the protective iron door illustrates the aforementioned risks perfectly. Despite knowing the risks, he had entered the chamber with plastic bags filled with air. This, he hoped, would buy himself enough time to enter and steal some of the precious crystals within. Alas, the man was found dead the next day; cooked, as though he were food in an oven.

The Naica mine is still active, and should you wish to visit, it is possible to do so with prior approval. However, bear in mind that only the upper caverns are available for access (under supervision), with the lower caverns currently considered too dangerous due to extreme temperatures. Only scientists and employees are able to access these prohibited areas.

BEAUTIFULWORLD.COM

BEAUTIFULWORLD.COM

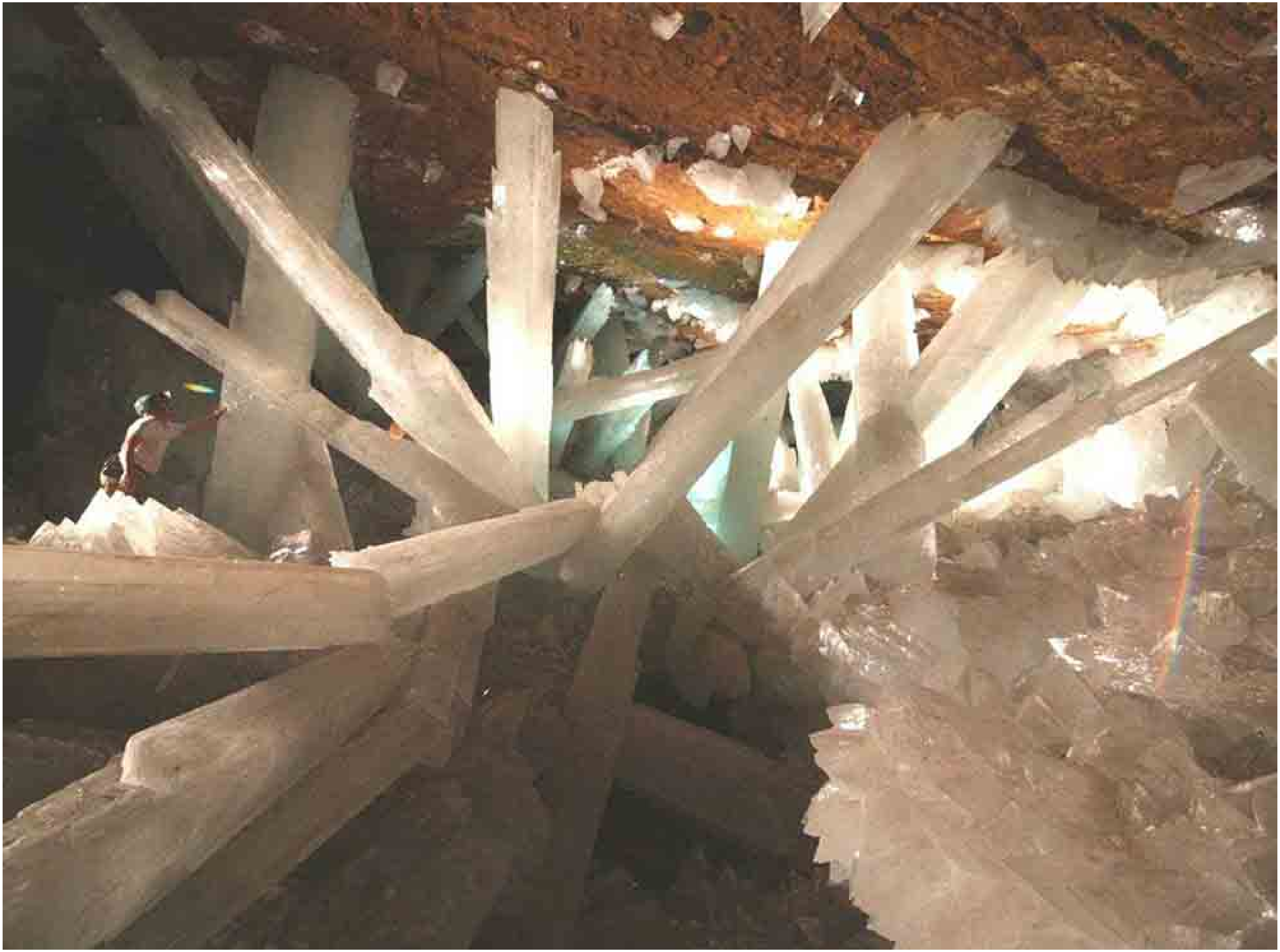

The Naica Mine of Chihuahua, Mexico, is a working mine that is known for its extraordinary crystals. Naica is a lead, zinc and silver mine in which large voids have been found, containing crystals of selenite (gypsum) as large as 4 feet in diameter and 50 feet long. The chamber holding these crystals is known as the Crystal Cave of Giants, and is approximately 1000 feet down in the limestone host rock of the mine.

The crystals were formed by hydrothermal fluids emanating from the magma chambers below. The cavern was discovered while the miners were drilling through the Naica fault, which they were worried would flood the mine. The Cave of Swords is another chamber in the Naica Mine, containing similar large crystals.

The Naica mine was first discovered by early prospectors in 1794 south of Chihuahua City. They struck a vein of silver at the base of a range of hills called Naica by the Tarahumara Indians. The origin in the Tarahumara language seems to mean "a shady place". Perhaps here in the small canyon there was a grove of trees tucked away by a small canyon spring.

From that discovery, until around 1900, the primary interest was silver and gold. Around 1900 large-scale mining began as zinc and lead became more valuable.

During the Mexican Revolution the mine was producing a great deal of wealth. Revolutionary troops entered the town and demanded money from the owners. One of them was assassinated when he refused to pay, causing the mine to shut down from 1911 to 1922.

Just before the mine was closed, the famous Cave of Swords was discovered at a depth of 400 feet. Due to the incredible crystals, it was decided to try to preserve this cave. While many of the crystals have been collected, this is still a fascinating cave to visit. In one part there are so many crystals on one of the walls, they appear to be like an underwater reef moving in a gentle undulating motion in an ocean current.

In April 2000, brothers Juan and Pedro Sanchez were drilling a new tunnel when they made a truly spectacular discovery. While Naica miners are accustomed to finding crystals, Juan and Pedro were absolutely amazed by the cavern that they found. The brothers immediately informed the engineer in charge, Roberto Gonzalez. Ing. Gonzalez realized that they had discovered a natural treasure and quickly rerouted the tunnel. During this phase some damage was done as several miners tried to remove pieces of the mega-crystals, so the mining company soon installed an iron door to protect the find. Later, one of the workers, with the intention of stealing crystals, managed to get in through a narrow hole. He tried to take some plastic bags filled with fresh air inside, but the strategy didn't work. He lost consciousness and later was found thoroughly baked.

When entering the cave our group is issued helmets, lanterns, rubber boots, and gloves. One must then be driven by truck into the main mining tunnel called Rampa Sn. Francisco. While the vertical drop is approximately 1000 feet, the drive is almost a half mile long. The heat steadily increases and women have been observed to begin "glowing". The truck stops in front of a concrete wall with a steel door. The intense heat can prevent brain functioning.

At the end of the tunnel there are three or four steps into the aperture of the cavern itself. It is in this short tunnel. In this short distance the temperature and humidity goes from being uncomfortably warm to literally a blast furnace.

Momentarily, the penetrating heat is forgotten as the crystals pop into view on the other side of the "Eye of the Queen". The entire panorama is now lighted and the cavern has a depth and impressive cathedral-like appearance that was not visible on earlier trips with just our headlamps.

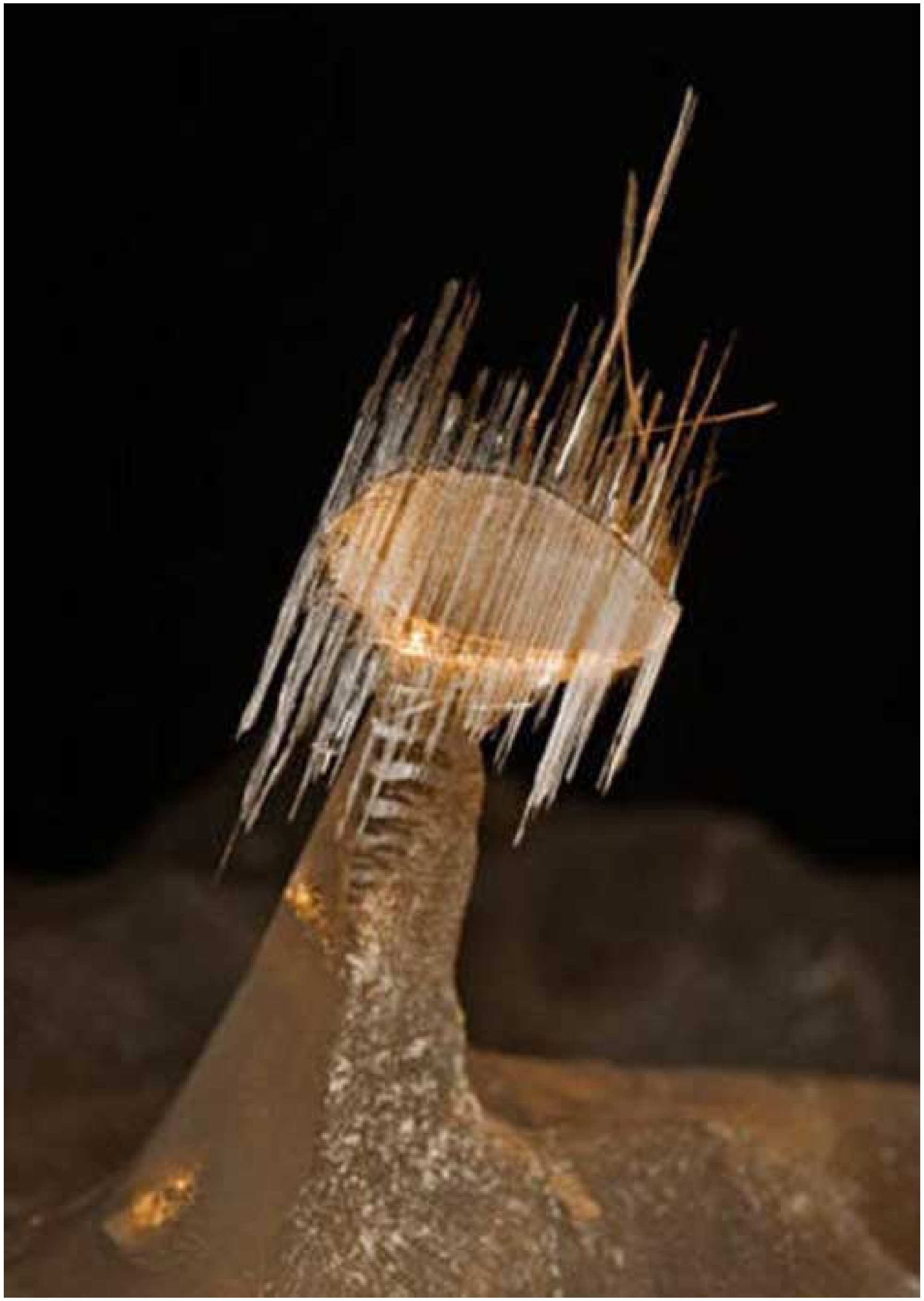

When inside the great cathedral of crystals, the pressure of intense heat create a gamut of emotions and perhaps hallucinations. One can only remain for a short period of time.

Geologists report that these natural crystal formations are incredibly complex, yet so simple. They have a magical or metaphysical personality independent of their chemical structures. There is a magma chamber two to three miles below the mountain and that heat from this compressed lava travels through the faults up into the area of the mine. Super heated fluids carry the minerals the miners are seeking as well as form the crystals. The mine is ventilated; otherwise, it could not be worked. Some parts, however, are not air-conditioned, such as the Cave of the Crystals, and there you feel the heat from the magma deep below. The fluids travel along the Naica fault, enter voids in the bedrock, and then form entirely natural structures that are not easily explained scientifically.

In April 2000, the mining company became confident that the water table on the other side of the fault had been lowered sufficiently to drill.

When they did this, it is almost as if a magical veil of reality was breached and an entirely new world was discovered. Two caverns filled with the Earth's largest crystals were immediately revealed. More discoveries are expected to be made in this magical kingdom of intense natural beauty.

Selenite, the gypsum crystal, named after the Greek goddess of the moon, Selene, due to its soft white light, is said to have many metaphysical and healing benefits. Selenite powder has been used cosmetically for thousands of years to enhance one's natural beauty. It is believed that this crystal assists with mental focus, growth, luck, immunity, and soothes the emotions.

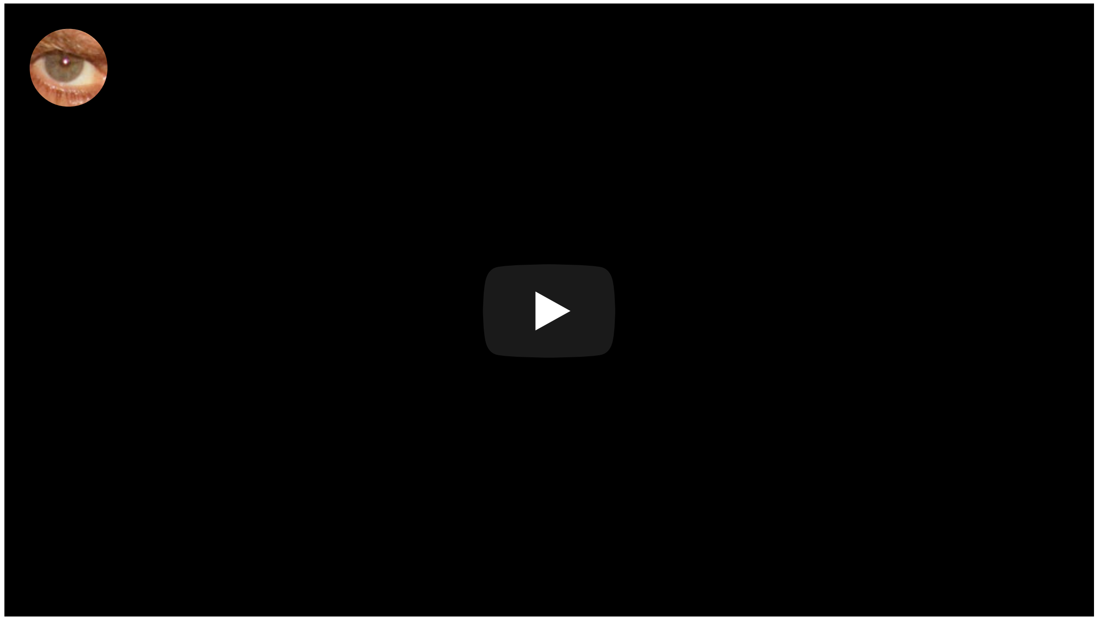

In the News ...

Naica's crystal caves hold long-dormant life BBC - February 20, 2017
Scientists have extracted long-dormant microbes from inside the famous giant crystals of the Naica mountain caves in Mexico - and revived them. The organisms were likely to have been encased in the striking shafts of gypsum at least 10,000 years ago, and possibly up to 50,000 years ago. It is another demonstration of the ability of life to adapt and cope in the most hostile of environments. "Other people have made longer-term claims for the antiquity of organisms that were still alive, but in this case these organisms are all very extraordinary - they are not very closely related to anything in the known genetic databases," said Dr Penelope Boston. The new director of Nasa's Astrobiology Institute in Moffett Field, California, described her findings here at the annual meeting of the American Association for the Advancement of Science (AAAS).

Crystals Break Slow Growth Record Seeker - September 13, 2011
Buried a thousand feet below Mexico's Naica mountain, the "Cave of Crystals" are the slowest growing ever measured. Researchers have measured the slowest crystal growth ever, which created the massive, underground gypsum crystals found in Mexico's Naica mine. Some crystals may have taken as much as a million years to form, growing by the width of a hair each century. The largest crystals in the mine are at 36 feet long and more than three feet across.

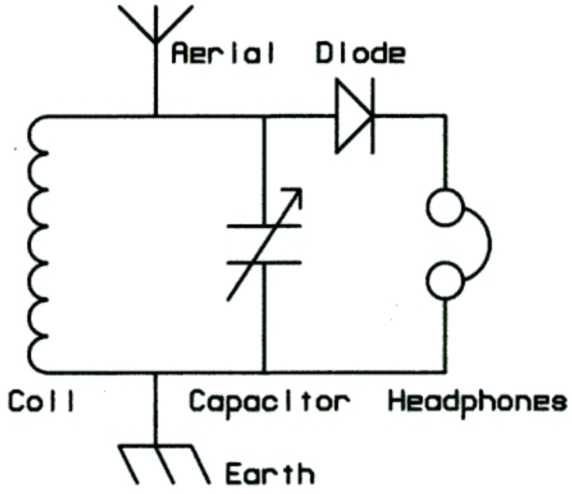
Aerial
Diode
Coil
Capacitor
Headphones
Earth

Basic Circuit

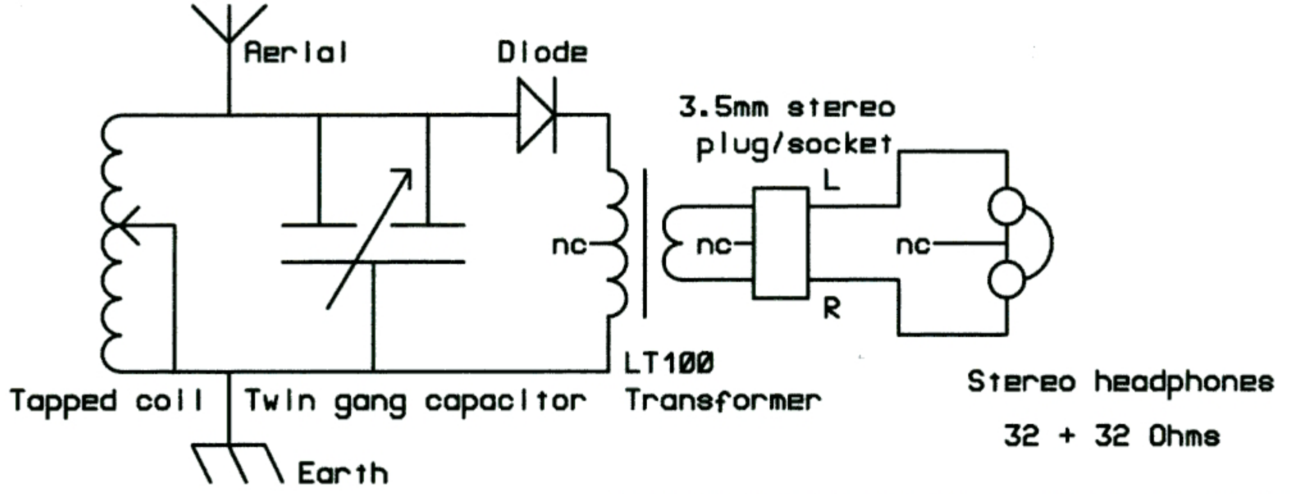
Aerial
Diode
3.5mm stereo plug/socket
L
R
nc
nc
nc
Tapped coil
Twin gang capacitor
LT100 Transformer
Stereo headphones
32 + 32 Ohms
Earth

Practical Circuit

(common headphone connection not used)

nc = no connection

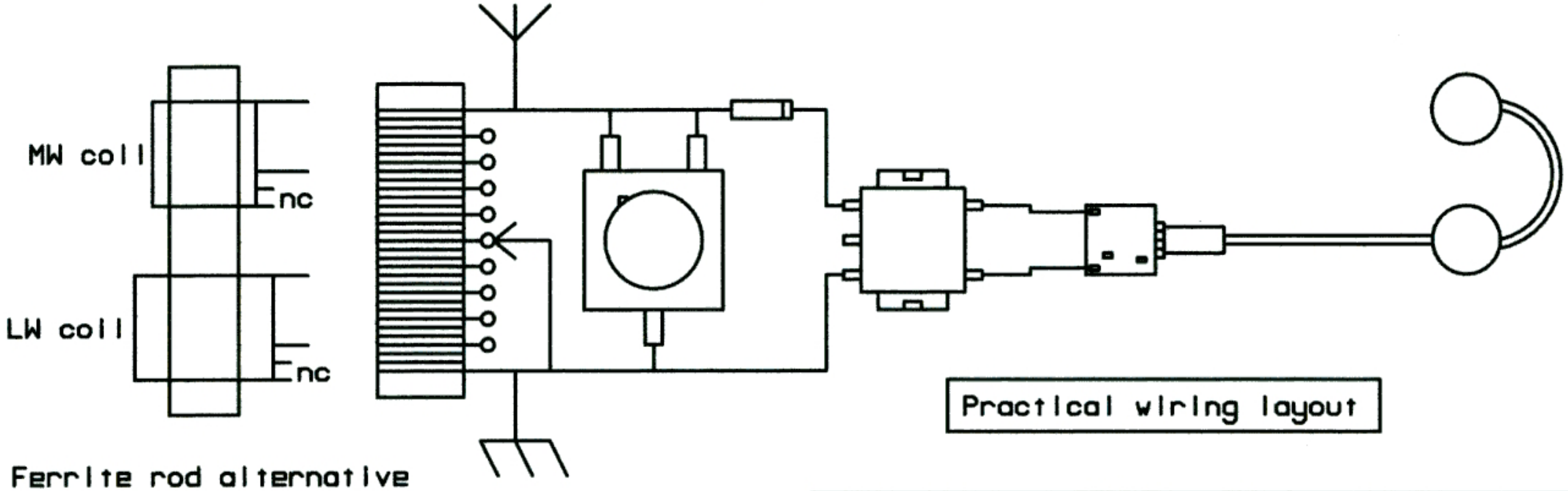
MW coil
nc
LW coil
nc
Ferrite rod alternative

Practical wiring layout

DRN:- FGR
DATE:- 11/05/2007
JOB No:- 2625
ISSUE:- 1

TITLE

OU BBC Crystal Set

THIS DRAWING IS THE PROPERTY OF THE OPEN UNIVERSITY AND MAY NOT BE REPRODUCED OR DISCLOSED TO A THIRD PARTY WITHOUT WRITTEN PERMISSION. © 98 OPEN UNIVERSITY

Yet another crystal set am broadcast band radio

ralph klimek 2001

Why do I or the world need yet another crystal-set design ?

This little project was motivated by me wanting to show the kids some cool technology. A radio that required no batteries or no source of external power, save that so generously provided by the advertisers of commercial AM broadcast stations. I revived for me rather nice memories of when I was very young trying to coax the last bit of performance from my earliest radios. It was allways a tradeoff between loudness and selectivity. In Melbourne Australia this really only meant being able to discriminate between the government stations at 620 and 770 khz which ran a 50KW transmitter and the more numerous commercial stations between 800 and 1600Khz any only running only 5KW. The meant in practice, being only able to resolve the big government radio and the commercials were reduced to a cacophony all trampling on each other. In those days nearly all the commercials carried horse racing or football on the weekends and this made for amusing listening. Often they would all call the same horse race and you could make out 5 stations at once blabbling out the same race call.

The classic crystal set design that is still taught to kids requires a coil of high inductance, chosen to resonate with what was the standard broadcast gang of 415pF max.

This is actually a rather poor choice. It means a poor Q at the low end and allmost NO Q at the high end, which is why here in Melbourne the commercials could not be discriminated. That meant the "cool" stuff, the pop music and shock jocks could not be heard by me. So I acquired a taste for govenment news and solid classical music. I remember falling asleep while listening to the Week in Business or the haunting sounds of "In Choirs and Places" (radio 3AR) just before nightly close down with one of those horrible crystal earpieces (I called them "infecto-phones") making my ear very sore. But it was worth it because one of those shows in the sixties was the inspiring ABC Weekly Science Report which would lead me on to better things.

So this silly little project caused me to revisit the concept of Q and bandwidth and resonance. The classical crystal set performance was ruined by poor understanding of Q. The classical crystal set inductor has high inductance, presumeably to resonate with a cheap low maximum value variable capacitor. High inductance is not a bad thing if that coil has low resistance, but at RF that is hard to do.

This radio's resonator has very small inductance and very high capacitance. Performance is not limited by the inductors resistance but only by the loading. The load is nothing more that the losses in the antenna and the rectifier. The diode is ,off course, germanium. It feeds a step down transformer, not of audio design, but a mains frequency plug pack transformer which drives high impedance (600ohm) voice coil headphones. Any cheap ones will do. The coil should be wound with the heaviest gauge wire to hand, the ferrite core is optional, it serves merely to shorten the length of wire required and to make the coil size practical. Allthough not done here, the coil should be tapped half way to provide for 2 octaves of tuning. The large tuning capacitance effectively absorbs the capacitance of the antenna.

Now in practice, with an antenna that consists of nothing more than a 10 meter long heavy gauge wire running along the back fence, less than 3 meters above the ground and a good earth, this radio can resolve all of Melbournes' stations with minimal interference. The really suprising thing, it is possible to drive a large diameter loud speaker with this radio. I wouldnt call it loud, but rather, call it noticeable. At late night it could even be described as objectionable!

...and someone else is paying for the power!

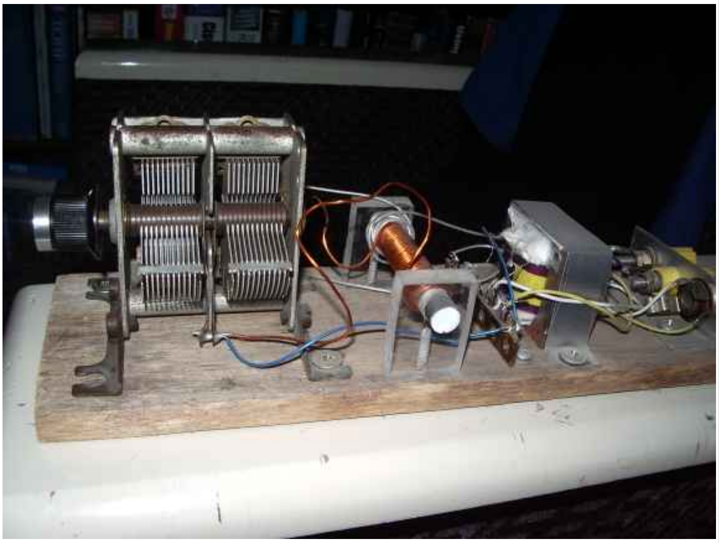	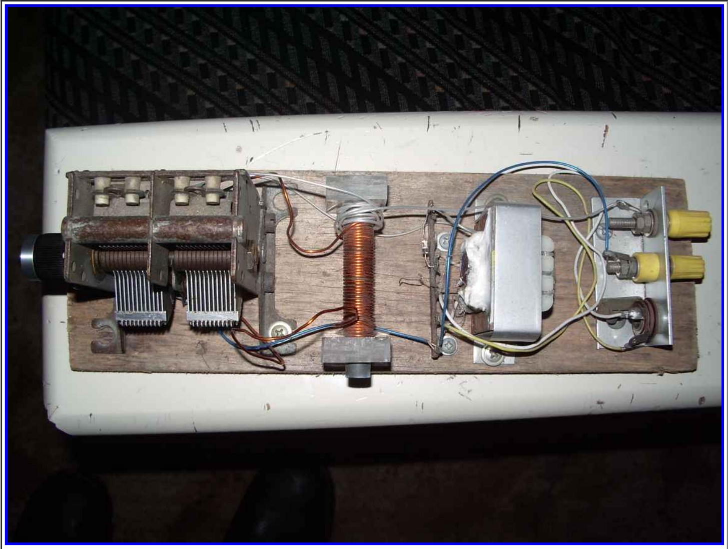
old broadcast gang parrelled to give 900pF Max	Inductor has about 30 to 40 turns of 16 gauge wire

homepage

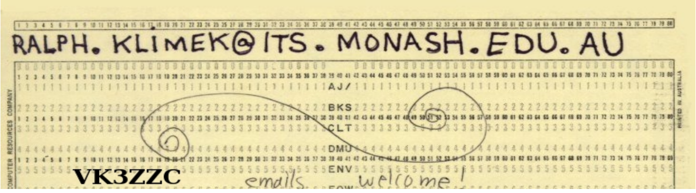

RALPH.KLIMEK@ITS.MONASH.EDU.AU

VK3ZZC

emails welcome!

AJ/
BKS
CLT
DMU
ENV
COMPUTER RESOURCES COMPANY
PRINTED IN AUSTRALIA

THE BOY'S BOOK
of
CRYSTAL SETS

BY

W. J. MAY

⁂

BERNARDS (PUBLISHERS) LTD.
THE GRAMPIANS
WESTERN GATE
LONDON W.6

General Editor
WALTER J. MAY

First Published—March, 1954
Second Impression—June, 1954
Third Impression—April, 1955
Fourth Impression—January, 1956
Fifth Impression—February, 1957
Sixth Impression—August, 1957
Seventh Impression—March, 1958
Eighth Impression—October, 1958
Ninth Impression—April, 1959
Tenth Impression—December, 1959
Eleventh Impression—April, 1960
Twelfth Impression—October, 1960

⁂

Printed by V. Cooper & Partners, Ltd., Flitcroft St., W.C.2
for Bernards (Publishers) Ltd., The Grampians, Western Gate, London, W.

PREFACE

The Publishers make no apology for reviving interest in the crystal set. For too long the technical press has immersed itself in an orgy of complex superheterodyne receivers and television equipment, all of which is necessarily expensive. In the 1920's, crystal set construction was a fascinating and inexpensive pastime indulged in by the majority of boys of that day. Today this pleasurable hobby can be even more attractive, the introduction of the germanium crystal diode and high performance coils, has opened up new paths which will capture the interest of boys of all ages. Perhaps most important of all, not only can a receiver capable of impeccable reproduction be produced without any technical knowledge; but the cost is no greater than that of a few visits to the cinema

❧

CONTENTS

Page

Preface - - - - - 4

Introduction - - - - 5

Receiver No. 1 - - - - 12

Receiver No. 2 - - - - 13

Receiver No. 3 - - - - 14

Receiver No. 4 - - - - 16

Receiver No. 5 - - - - 19

Receiver No. 6 - - - - 20

Receiver No. 7 - - - - 23

Receiver No. 8 - - - - 24

Receiver No. 9 - - - - 26

Receiver No. 10 - - - - 27

Receiver No. 11 - - - - 30

Receiver No. 12 - - - - 31

INTRODUCTION

Before any attempt is made to construct a receiver, it is necessary to examine the problems which surround the crystal set, so that the best can be obtained from any of the designs attempted.

Firstly, it must be understood that the crystal set as it is to-day does **not** provide any amplification. It relies entirely on what is fed into it via the aerial and earth system and gives a very faithful replica of the original transmission.

From this it will be obvious that the aerial and earth system must be as efficient as possible if the final results are to be in any way outstanding. This because these are the only means by which the signals are fed to the receiver.

Secondly it is necessary to understand the nature of the transmitted signal, then it will be easy to understand the working of the set and to appreciate the function of each of the components.

When crystal sets first became popular, very little information was generally available, at least, not in a form that school-boys could understand.

We all built sets of all shapes and sizes, with coil designs that had to be seen to be believed, but very few of us had much idea of how they worked.

I well remember the case of a cousin of mine who, in those days acquired a magnificent variable capacitor or condenser as it was then called, this instrument had a most impressively engraved dial of polished ebonite, brass vanes and nickel plated end plates. Having observed similar dials on several highly priced commercial receivers in the town, he at once ripped out the somewhat tattered coil from the family receiver and replaced it with this device. The profound silence which ensued caused considerable amazement and dismay until a better informed adult explained the mysteries of L and C to him. Readers of this manual, however, will be better informed and there is no risk of failure with any of the designs described provided the instructions are carefully followed.

The Signal

To commence, we will assume that an orchestra is playing in a broadcasting studio. Since the principle of radio transmission is electrical it is necessary to change the sound produced by the orchestra into an electrical equivalent. This is carried out by the microphone, which picks up the sound and changes it into minute electric currents.

As they appear at the output of the microphone they are too small to be of use and accordingly are passed through a high power amplifier. These amplified currents could now be transmitted, but unfortunately, as we shall learn, owing to the inherent nature of the signal in this state, transmission over any useful distance would be impractical.

When the music from the orchestra is transformed into electrical currents they are in the form of alternating currents, usually called A.C., that is they rise to a maximum in one direction, fall to a minimum, rise to a maximum in the opposite direction and then fall to minimum again. This process is repeated over and over again. One complete rise and fall in each direction is called a cycle and is drawn in Fig. 1. Every time a note is struck on a piano, vibrations are sent out which reach the ear enabling you to hear it. These vibrations are also spoken of as cycles, they rise and fall in intensity the same way as an alternating current. The number of cycles radiated by any given note over a period of one second are referred to as its frequency. Middie C on the piano sends out 261 cycles every second and is known as having a frequency of 261. The microphone also " hears " the note and in the case of middie C produces minute A.C. at 261 cycles. This can be drawn as in Fig. 2, the only difference between Fig. 1 and 2 is, that the time factor is given so that the frequency can be identified. The higher the pitch of a note the higher the frequency and the lower the pitch the lower the frequency. On a piano the frequency of the top note is 3515 cycles and that of the lower 27 cycles. Those of you who have listened to an organ in a concert hall will have noticed that when a very deep note was played, it sounded like a growl to the ear, but the vibrating fre-

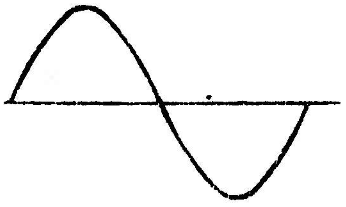

FIG 1

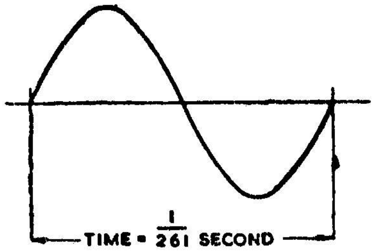

FIG 2

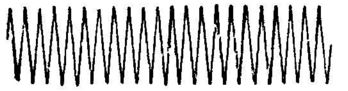

FIG 3

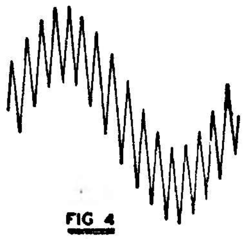

FIG 4

quency could be distinctly felt through the seat. Higher notes have too high a frequency to be observed in this manner.

The range of sounds which can be detected by the human ear are known as audio or low frequencies. From this you will understand that a low frequency amplifier is one which amplifies sound.

So that the transmitter will carry the programme over a useful distance it is necessary to radiate high frequencies. Now, as explained, the programme to be transmitted consists of low frequencies, and to overcome the difficulty, the transmitter generates A.C. of high frequency and combines it with the low frequencies. It will now be understood that the transmitted signal consists essentially of two different parts, a high and a low frequency content.

Fig. 3 gives a representation of the high frequency signal generated by the transmitter. In the case of the London Home Service, the frequency is 908000 cycles.

When referring to a high frequency signal on the medium or long wave-band it is usual to express the frequency in thousands of cycles, thus 908000 cycles becomes 908 kilo-cycles, which in turn may be abbreviated to 908 k/cs. It might at first be thought that by adding the low frequency or L.F. signal to the high frequency or H.F. carrier a form such as in Fig. 4 would result, such a combination is useless, and, so that the original L.F. content can be satisfactorily extracted by the receiver, the L.F. signal must vary the amplitude or output power of the H.F. signal as in Fig. 5.

It is in this form that the signal arrives at the receiving aerial. The aerial in itself is incapable of discriminating between one signal and another, and countless signals will be collected by the aerial at any one time. Many of these are too weak to be of use but the stronger ones must be sorted out since there is no point in receiving several programmes at once.

Fig. 6 shows the basic circuit of the input to a crystal. The coil L possesses a quality known as inductance, and the capacitor C, that of capacitance. If the coil had no capacitance whatever across it, all signals arriving at the aerial would be effectively short circuited to earth. As a matter of interest, it is impossible to obtain this state of affairs since even without any additional capacitor any coil **must** contain a certain amount of self capacitance.

By combining a coil and capacitor as in Fig. 6 a peculiar effect is observed, at one particular frequency, the signals are not short circuited to earth, but are developed across the coil. In other words the combined effect of L and C no longer provides a short circuit, **but only at one particular frequency.** If the value of C is altered the effect will be observed at a different frequency; likewise by altering L the frequency at which the effect will take place can be changed. There is a name for this phenomena, the frequency at which it occurs with any given L and C combination is known as the **resonant frequency.**

The values of the coils and capacitors shown in this manual have been carefully chosen so that resonance will be obtained at all frequencies where stations are broadcasting. Broadcasting stations work in bands of frequencies, those of major interest to crystal set constructors are the medium wave-band 1200 k/cs—600 k/cs and the long wave-band 300 kc/s—150 k/cs.

Usually a variable capacitor is used with a fixed inductance to cover one band and an additional coil switched in to increase the inductance to cover the other. In this way the L and C combination can be adjusted to provide resonance at the desired frequency of any given station. In other words you can **select the** station you want by varying C, that is turning the dial of the variable capacitor. This procedure is referred to as tuning. Having selected or tuned the required station it still remains necessary to change the form of the signal back to that of the original transmission. This process is called detection or de-modulation.

Examination of Fig. 5 will show that the signal has been duplicated, in other words, as it rises in one direction it also rises equally in the other. In this form the signal is useless since each half of the signal cancels the other, and if this signal is applied to a pair of headphones silence will result. Obviously some provision must be made to get rid of the unwanted half of the signal, and it is here that the crystal detector must be considered. This device will pass current in one direction only, ignoring any signal in the opposite direction, so that if the signal of Fig. 5 is passed through such a crystal, that of Fig. 7 will result. There is still the H.F. content to be reckoned

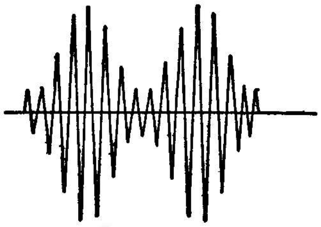

FIG 5

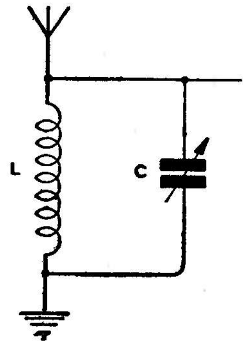

FIG 6

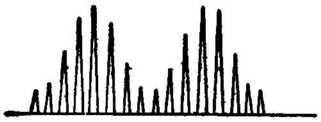

FIG 7

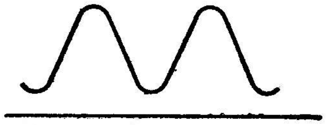

FIG 8

with, fortunately this is easily dealt with, a capacitor connected across the headphone terminals effectively disposes of this, leaving only the audio or L.F. content, as shown in Fig. 8. This audio content, which is a faithful replica of the original transmission is fed to the headphones. These in turn reverse the process of the microphone and transform the electrical currents into sound waves acceptable to the human ear.

Briefly then, your requirements are as follows:—

(1) A good aerial and earth installation, to make the most of the available signals.
(2) A receiver containing:—
  (a) Some form of coil and capacitor (L & C) combination to select or tune in the wanted station.
  (b) A crystal to get rid of the unwanted half of the signal (detection).
  (c) A fixed capacitor across the headphones to get rid of any remaining carrier.
(3) A pair of sensitive high resistance headphones.

Aerials

By this time the intending constructor will be able to appreciate the necessity of a good aerial. It is a point which cannot be over-emphasised. Assuming that you are in the fortunate position of being able to erect an outdoor aerial there are two main considerations, height and length.

One of the best that can be used is the inverted L shown diagramatically in Fig. 9.

It should be erected as high as is practical, every foot counts. The horizontal wire, that is the aerial proper, should have a minimum length of 60′ to which of course the length of the down lead is added. Where it is impossible to erect an aerial with an ideal horizontal length, a compromise must be effected.

Fig. 10 shows a three wire spreader aerial which gives quite a good effective length.

Suitable wire for a receiving aerial will not set any problems, stranded copper about 7/22 gauge is the best. 7/22 means that it consists of 7 strands of 22-gauge wire. This wire may be obtained covered, and for the present purpose is better than the plain or enamelled kind.

Note that insulators are used between the actual aerial wire and its anchoring supports, it is important that these are used, otherwise leakage will occur which will of course spoil its efficiency. Fig. 11 and 12 show how the wire may best be attached to two of the most common types of insulator available. Fig. 11 is of porcelain and is usually referred to as an egg insulator, whereas Fig. 12 shows a more modern (and more expensive) type in glass. If you are using the glass pattern, one is usually sufficient at each end of the aerial, but with the egg type two should be used. No doubt many readers will not be in a position to erect an out-door aerial, and must necessarily be content with an indoor installation.

The next best thing to a good out-door aerial is a replica constructed in a loft. If this form of construction is used, care must be taken when feeding the down lead to avoid close contact with the wall of the house. At the point where the lead feeds under the eaves, a length of rubber tubing can be used to cover the wire. The lead is fed through one of the small spaces left for ventilation purposes. A general idea is given by Fig. 13.

A less elaborate but quite effective aerial can be obtained by using a bed-spring. A length of 7/22 copper insulated wire is connected between the receiver and the spring. The spring should first be cleaned with emery cloth. Remove about 2″ of insulation from the wire and bind it tightly round the prepared spring. The joint may be covered with insulating tape which is obtainable from all electrical stores for a few pence.

Such aerials are quite popular since so many crystal sets are built for bedroom use.

When an aerial is required in the living room, the picture rail can be conveniently used. Insulated screw-eyes are fixed at intervals of about 3 feet along the rail, the wire is firmly anchored to one of them, and stretched right round the room until you arrive back at the starting point. By use of one of the insulated screw-eyes, the wire is secured and the down lead fed to the receiver. The idea is illustrated in Fig. 14.

Soldering

Before leaving the subject of aerials, a few words on soldering will not come amiss. Down leads on outdoor aerials should be soldered, and the same applies to loft types. Apart from aerial leads; earth leads and all the connections in the receiver will require soldering.

For the type of soldering necessary in radio construction an electric iron is the best solution. One of the small types marketed by Adcola or Henleys will be found admirable for the job. These are excellent for actual set construction but are hardly large enough for soldering the down lead to an outdoor type of aerial. Here the heat is dissipated much more quickly and a larger iron is required in order to get the solder to flow.

Ordinary irons which may be heated by a gas flame can be obtained very cheaply from most ironmongers' stores. Assuming the use of such an iron, first heat the iron until the copper bit is giving off a green coloured flame, the iron is now at the correct operating temperature. The bit is now discoloured or oxidised and its tip

✓

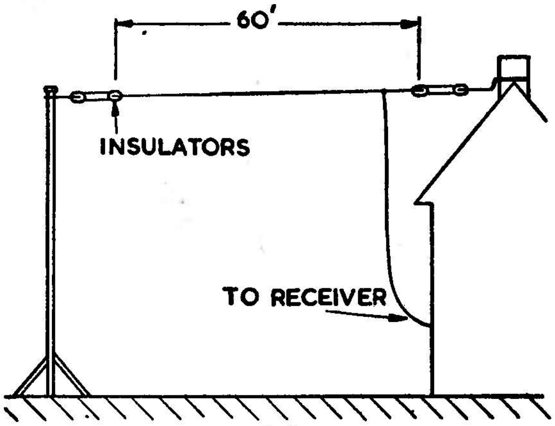
60'
INSULATORS
TO RECEIVER

INVERTED 'L' AERIAL

FIG. 9

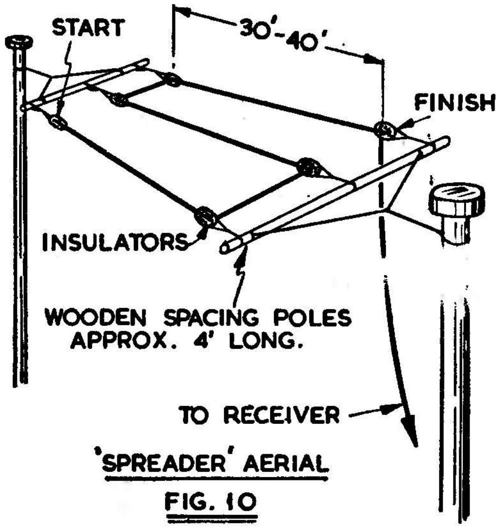
START
30'-40'
FINISH
INSULATORS
WOODEN SPACING POLES
APPROX. 4' LONG.
TO RECEIVER

'SPREADER' AERIAL

FIG. 10

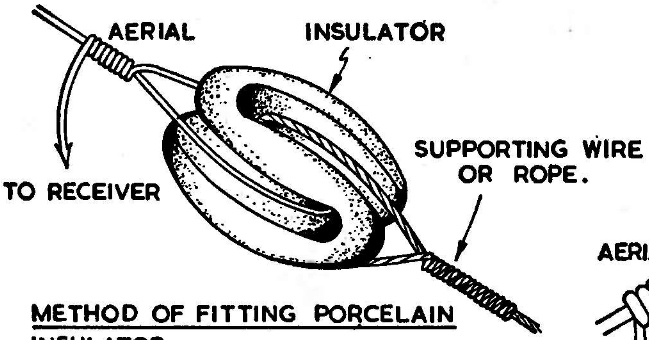
AERIAL
INSULATOR
TO RECEIVER
SUPPORTING WIRE
OR ROPE.

METHOD OF FITTING PORCELAIN INSULATOR

FIG. 11

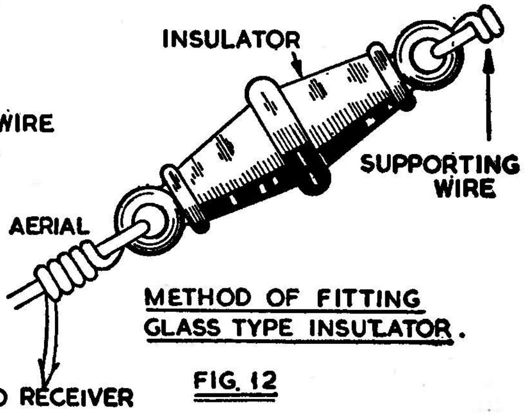
INSULATOR
SUPPORTING
WIRE
AERIAL
TO RECEIVER

METHOD OF FITTING GLASS TYPE INSULATOR.

FIG. 12

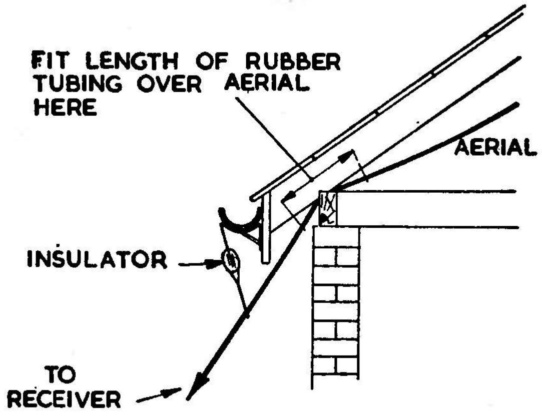
FIT LENGTH OF RUBBER
TUBING OVER AERIAL
HERE
AERIAL
INSULATOR
TO
RECEIVER

LOFT AERIAL.

FIG. 13

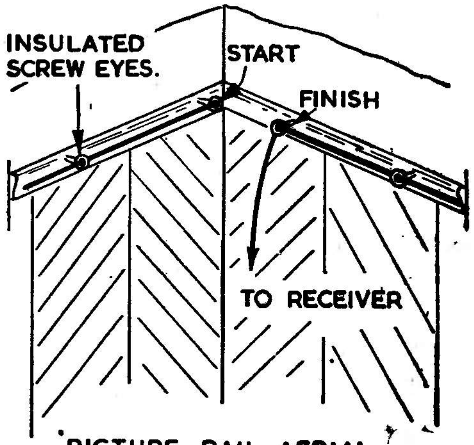
INSULATED
SCREW EYES.
START
FINISH
TO RECEIVER

'PICTURE RAIL AERIAL.'

FIG. 14

should be quickly cleaned with an **old** file which should be kept specially for the purpose. A better idea is to obtain a small block of sal-ammoniac and rub the tip of the hot iron on it. This will clean the tip of the bit perfectly. Next take a length of cored solder such as Ersin Multicore, and melt a little on to the prepared tip faces, now smooth it evenly over the surface with a piece of old rag (be very careful not to burn your fingers) the iron is now "tinned" and ready for use. When heating the iron, be very careful not to let the bit overheat or get red hot, otherwise the tinned surface will be destroyed and the whole process will have to be gone through again.

An electric iron will not overheat, and the tinning will last much longer than with ordinary types, and since the heat does not deteriorate there is less likelihood of making faulty joints.

Having obtained a tinned iron, the process of soldering joints is remarkably simple and anyone with a little patience can acquire the art in a very short time.

To solder two copper wires together: clean the wire with emery cloth, apply a prepared iron and some cored solder to it, the solder will flow evenly over the wire thus tinning it. Repeat the process with the remaining wire and then twist the two together. Now apply the iron to the joint from the underside and the cored solder to the joint on top. Solder will flow evenly over the joint. Remove both iron and solder and allow to set. The solder will harden or set in a few seconds but during this period the joint must not be touched or moved, as otherwise the joint will be "dry" and quite useless mechanically or electrically. On radio components, tags for soldering are already tinned though if they are old, or discoloured it is best to re-tin them. Normally, however, it is only necessary to twist the connecting wire to it and apply the iron and solder as explained.

Remember: **never** apply the solder to the iron and then the iron to the joint, always apply the iron **and** the solder to the joint. It is however a good thing to apply a little solder to the iron tip even when it is perfectly tinned just before making a new joint.

One final "DON'T," you will have noticed that the solder referred to is "cored," that is, it contains resin and other substances through its centre, plain solder as used by electricians and plumbers will not do, as it is the resin or flux

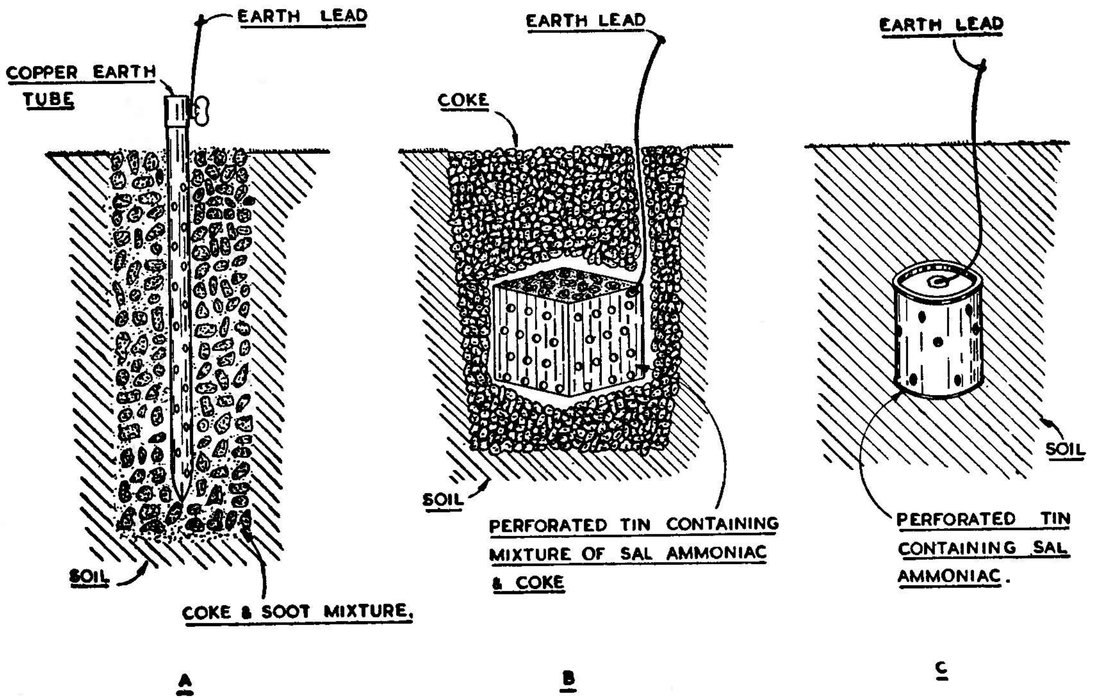
EARTH LEAD
COPPER EARTH TUBE
SOIL
COKE & SOOT MIXTURE.
A
EARTH LEAD
COKE
SOIL
PERFORATED TIN CONTAINING MIXTURE OF SAL AMMONIAC & COKE
B
EARTH LEAD
SOIL
PERFORATED TIN CONTAINING SAL AMMONIAC.
C

FIG. 15.

as it is called which makes the solder flow evenly and permits a good electrical joint.

Equally important do not attempt to use a separate flux or soldering fluid with plain solder or to "help" the cored solder; since these are almost certain to cause eventual corrosion and will destroy any components that have been contaminated.

Earths

The provision of a good earth is just as important as the rest of the installation.

Much disappointment would be avoided if this fact were not lost sight of, to avoid any slip up in this direction I am proposing to outline several well tried and efficient earth systems.

If it can be obtained, an earth rod, specially designed for the purpose provides the basis of a good earth connection. These were very popular some years ago when the majority of receivers were either crystal or battery operated, but in these days of modern mains driven sets they are not often used and consequently not always readily available.

Fig. 15 shows how it is used. First excavate a hole to a depth of three feet and fill with a mixture of soot and coke. Drive in the earth rod, which is a hollow copper tube, perforated. The lead to the receiver is connected at the top, and should be of covered 7/22 gauge, as used for the aerial system. Though not obvious it is important that this lead-in should be in covered wire, otherwise a number of indifferent earth contacts are likely to be made at various points along its length until it reaches the set. This is very undesirable and will spoil the efficiency of the system. It is essential to keep the soil surrounding the rod moist, which is one reason why the tube is hollow, and care must be taken to pour water into the tube at intervals.

A very efficient earth is the percolative type also popular at one time. Due to the chemicals used, it will extract moisture from its surroundings, thus maintaining a permanently moist earth.

It should be installed as in Fig. 15b.

The container is of copper or zinc, anything else will quickly rust away, again a good quantity of coke is used. Fill the container with a mixture of sal-ammoniac and coke, and then bury in coke as illustrated. This earth will not require further attention. It is possible to use powdered calcium chloride instead of sal-ammoniac but unless a chemist can be persuaded to make some up it is better to stick to sal-ammoniac.

If a zinc container is not available and you do not use or cannot easily get coke, Fig. 15c should be used. This makes a better earth than many so-called "earths" that I have come across. Obtain as large a tin as possible, make a number of holes as shown. Solder the lead in the bottom and fill it with sal-ammoniac. Replace the lid and bury in the ground. The tin will eventually rust away, but the replacement cost is negligible.

If it is quite impossible to make direct contact with the ground, a water pipe must be pressed into service. This should be a main pipe feeding straight to ground and not a hot water pipe or one fed from a tank. Scrape the pipe clean and twist the lead in tightly around it, a copper clip is even better. Do not attempt to solder on to the water pipe; since cold water is flowing through the pipe, it is extremely unlikely that your soldering iron will heat up the water supply sufficiently to allow a sound electrical joint, though you may spring a leak.

No attempt should be made to utilise gas pipes; the possibility of causing a fire is certainly very remote but, they make incredibly bad earth connections due to a number of joints made before true ground is reached. These joints are at best only semi-conductors, at least from an electrical standpoint.

Headphones

Since the late war there have been a large number of head-phone sets available on the surplus market. These may be roughly divided into two types, high-impedance and low-impedance. For the crystal sets detailed in this book high-impedance 'phones are required and the low-impedance pattern will not be suitable unless a matching transformer is used. As this is likely to cost more than the rest of the installation including the set, it will be as well to avoid them.

High-impedance types have an impedance of 2000Ω to 4000Ω whereas the low-impedance types are usually 600Ω.

The remaining consideration is weight; often cheap headphones are very heavy and uncomfortable to wear, every endeavour should be made to obtain 'phones as light in weight as possible.

Crystals

All the sets shown in this book have been designed to work with modern germanium crystals rather than the older galena crystal.

These germanium crystals require no adjustment, which, in itself, removes the main objection to this class of receiver. Suitable crystals are available from the following manufacturers and on the surplus market: Mullard, G.E.C., Brimar, Westinghouse and B.T.H. They are of robust construction, some in glass, others in ceramic or plastic, but must not be subjected to heavy knocks, otherwise the contact point may become dislodged.

From the information given, you should now be able to install an excellent aerial and earth system, and have some idea of how the crystal set works, so it is time to pass on to the actual receiver construction.

CONSTRUCTION 1

If you examine the following circuits you will find that each one is different. In most cases the difference lies in the coil design and/or the method by which the crystal and aerial is tapped into it. Each of these circuits has its own particular advantage to suit different conditions and the ideal circuit in some localities is not necessarily the best in others. It is not just a matter of a given circuit giving louder results than another, if it were there would be no point in showing more than one.

The main problem is to obtain adequate selectivity without reducing the volume level.

A receiver is said to be selective when it tunes sharply, a set with poor selectivity allows the stations to spread over the dial and when used near a transmitter will receive the local stations mixed together, which of course is useless.

Consider Fig. 16a, this is a very simple receiver, with no special attempt to provide any great amount of selectivity. In areas where signal

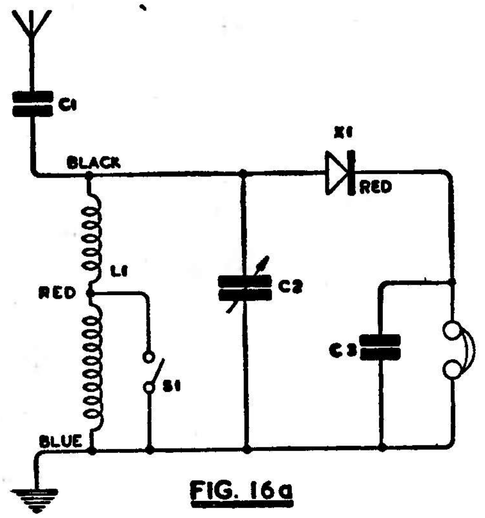

FIG. 16a

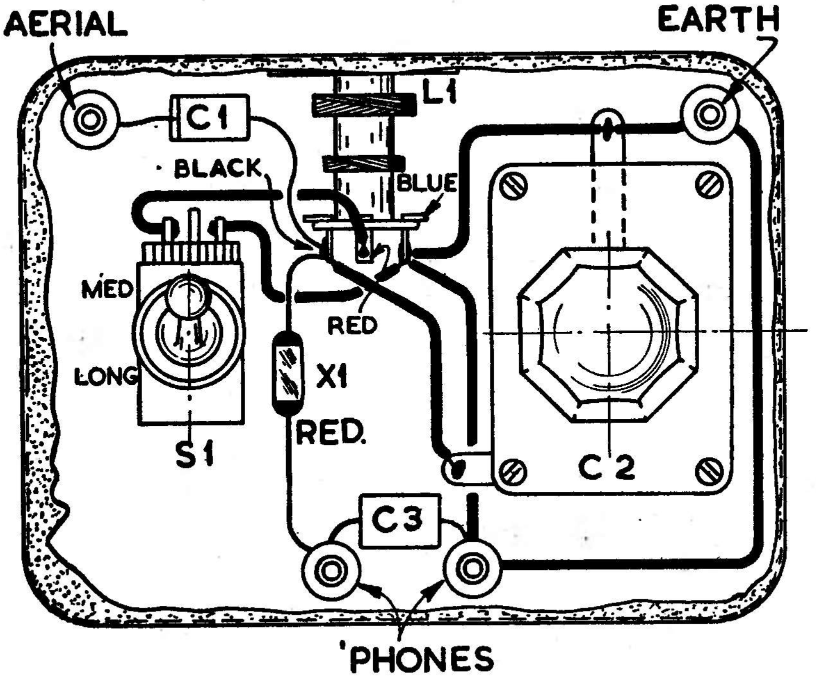

LID SHOWN BROKEN AWAY TO REVEAL POSITION OF COMPONENTS.

FIG 16 b

strength is not high, or a short aerial is used, it will probably be ideal.

It would have been quite easy to increase the selectivity by providing a tap on the coil for the aerial as in Fig. 17a, but unfortunately, as the selectivity is increased overall volume is likely to decrease, so that unless you live close enough to the transmitter to have a large signal available and therefore need the selectivity, the circuit of Fig. 16a will be quite satisfactory.

Capacitor C1 is to prevent the aerial damping the circuit too heavily because this would flatten the tuning unnecessarily, however with small aerials it may be better to take the aerial direct to the black tag on L1. The signals are selected or tuned by L1 and C2, X1 is the crystal and C3 the capacitor across the 'phones to prevent unwanted carrier or R.F. reaching the 'phones.

A practical diagram (Fig. 16b) is provided showing the layout and all the wiring. You will require nuts and bolts to fix the coil, about ½", 4BA size will do, the other parts have locking nuts provided. An old 2-oz. tobacco tin makes a very good container and keeps the size down. Note that the metal box is connected to earth. With the exception of AC/DC receivers the metal work on any receiver or amplifier is connected to earth.

When S1 is open as in the diagram, the set will tune in long-wave stations, but when closed the medium wave-band will be received.

Components List, Fig. 16a

C1	100pF mica capacitor.
C2	500pF tuning capacitor, solid dielectric.
C3	1000pF mica capacitor.
X1	Germanium crystal.
L1	Crystal Set Coil. R.E.P.
S1	Single Pole toggle switch.
4	Insulated wander-plug sockets and plugs.
1	2-oz. tobacco tin (or similar container).

Make sure that the wander-plug sockets are of the **insulated** type, otherwise the metal case will join all the sockets together electrically.

Try to follow the theoretical diagram when wiring, a little practice will soon enable you to wire up a set without a practical diagram, which is a great advantage because often only the theoretical diagram is given when circuits are detailed in the technical press.

Fig 28 on page 34 gives a list of symbols used on the theoretical diagrams, so that you can readily identify the components.

2

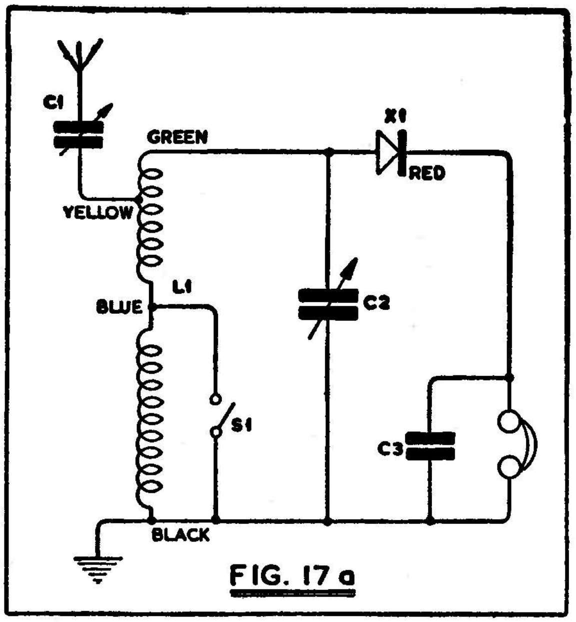

FIG. 17a

Fig. 17a circuit is similar in many respects to that of Fig. 16a. The difference is purely one of selectivity, tuning will certainly be sharper, and even with comparatively inefficient aerials the design will put up a very good performance. There is no reason why the construction should not follow the same lines as the previous receiver but, by enlarging the set a little and using an air spaced tuning capacitor, the efficiency is improved. Observe that the colour coding on the coil, an R.E.P. Dual Range (Blue Box) is different from that of the Crystal set coil used on the previous design, and make sure it is correctly wired in.

C1 is a mica compression capacitor, often referred to as a "trimmer" and because it is adjustable, permits the set to be matched to aerials of varying lengths. Construction is carried out on a square panel of bakelite, perspex or wood, as shown in Fig. 17b. Perspex being clear like glass presents a most attractive finish, provided the set is neatly wired. It must be drilled slowly, however, otherwise the generated heat of the drill will make the hole wander, and spoil the panel.

After the set has been wired and tested a small wooden container can be made to house the completed set.

When carrying out reception tests, a little experiment is well worth while, to get the best out of the set. Try removing the germanium crystal from the green tag on L1 and connecting it to yellow, at the same time removing the lead from C1 to yellow and connecting it to green. Once the best arrangement has been found the wiring can be left permanently in that position. Note that two tags on the coil are unused, this is intentional because on this circuit the extra winding connected to these tags is not required. With the switch S1 closed, the coil will cover the medium wave-band, and when open, long wave stations can be received. It is customary to abbreviate the expression medium wave-band to M.W. and long wave-band to L.W.

Components List, Fig. 17a

C1	100pF mica trimmer capacitor.
C2	500pF variable capacitor (air spaced).
C3	1000pF mica capacitor.
L1	Dual Range Coil R.E.P. (Blue Box).
S1	Single Pole toggle switch.
X1	Germanium Crystal.
4	Terminals or wander-plugs and sockets.
	Perspex, bakelite or wood for mounting panel.

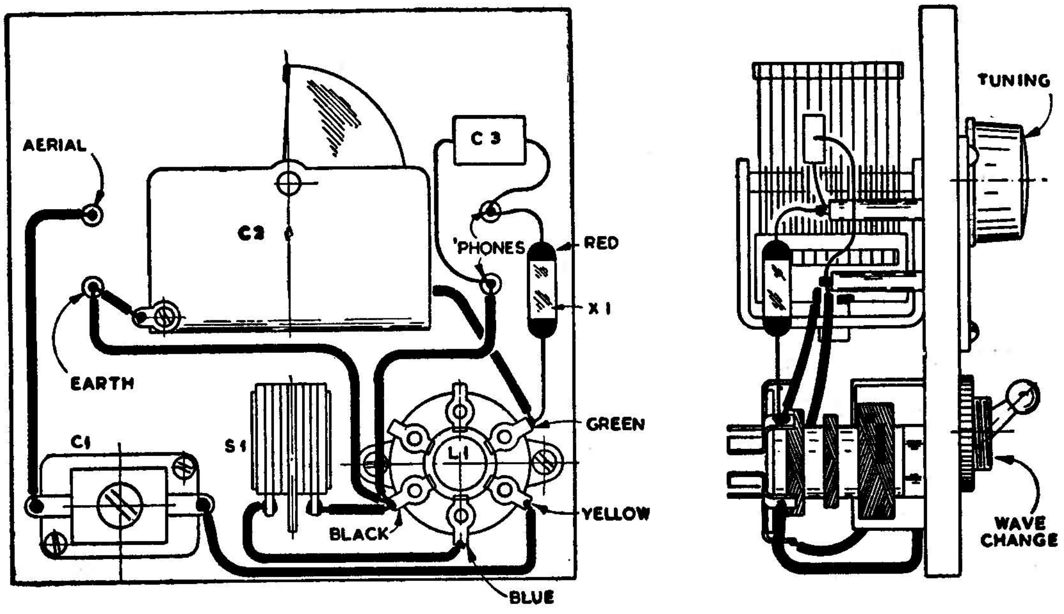

FIG. 17b

3

The design shown in Fig. 18a and 18b is more flexible than either of the preceding sets. That is, it can be varied to suit local conditions to a greater extent than the first two sets. This is made possible by the large number of taps provided on the coil.

You will notice that in this instance the coil is home-constructed on a cardboard former and is much larger than commercially produced coils. Coils wound in this manner are called **solenoids**, the turns are wound on side by side; in this case to a depth of some three inches. The diameter of the coil is made purposely large so as to obtain high efficiency. Modern commercial coils are invariably wave-wound and quite often are litzendraht which is the German for litz wire. This wire is made up of a number of strands of fine copper wire, each strand is enamelled to insulate it from the others, the whole is then silk covered. Litzendraht is more efficient than solid copper wire, and by using this and wave-winding, manufacturers can produce an efficient coil which is also small. Unfortunately wave-wound coils cannot be produced without a complex winding machine and home-constructed coils must take the solenoid form. It would be possible to use litzendraht but it is not easy to obtain by the reel, it is most expensive, and is difficult to handle. This last point is because at termination points, each strand must be cleaned of its enamel before a joint is made, and if one strand is broken, its advantage over plain copper wire is lost. How-

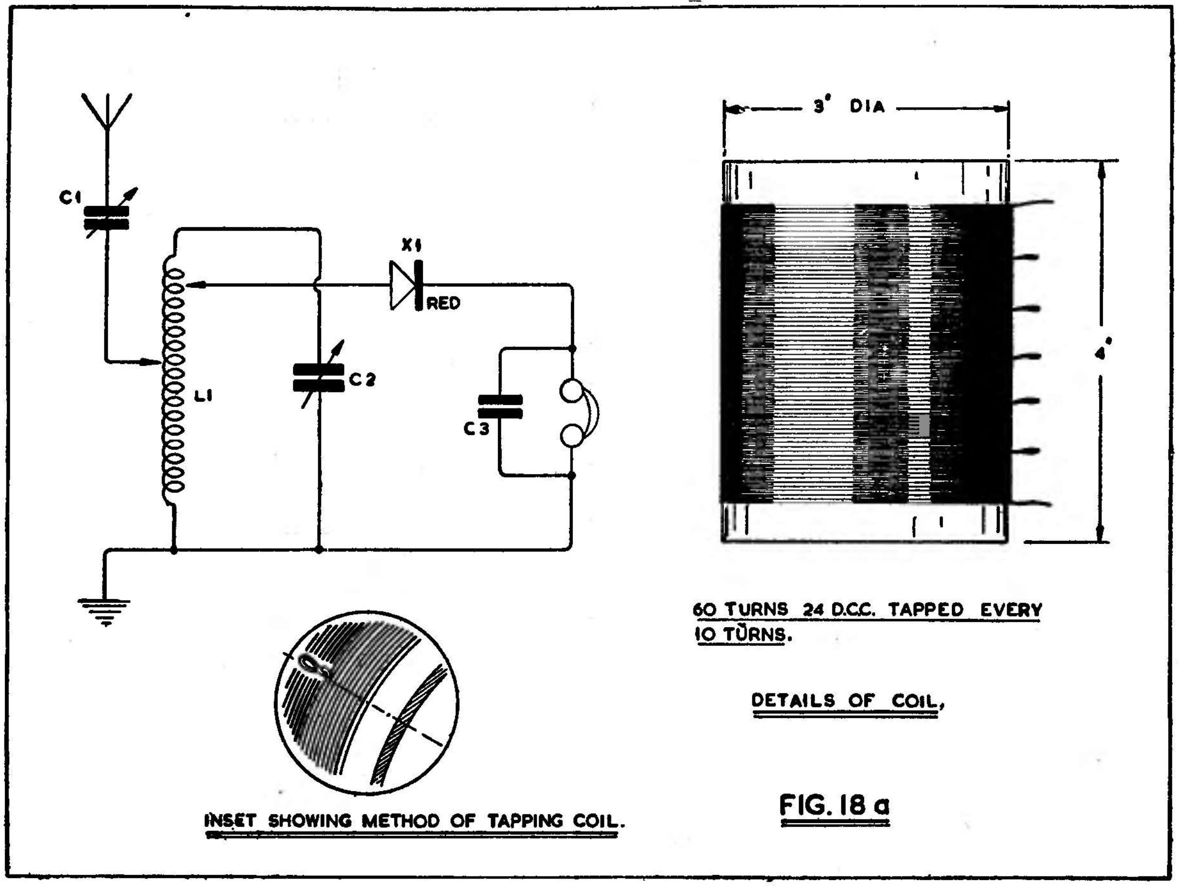
C1
X1
RED
C2
L1
C3
3″ DIA
4″
60 TURNS 24 D.C.C. TAPPED EVERY 10 TURNS.
DETAILS OF COIL.
INSET SHOWING METHOD OF TAPPING COIL.

FIG. 18a

ever, by using a former of reasonably large dimensions an efficient coil can be produced cheaply using ordinary copper wire.

The wire must be firmly secured at the start and finish of the windings. Pierce three small holes about ¼″ apart ½″ from the end of the former. Pass the wire through the first from the outside, return it through the second and pass it back again through the third. Leave some 6″ of wire at the end to make off the connection. It will now be possible to wind the turns on tightly without wire slipping.

Count on ten turns and make a loop 1″ long. The method of preparing loops is shown in Fig. 18a, loops or taps should always be made in this manner, never by baring the wire and soldering a further length of wire to it. Carry on with the winding, making off the taps every ten turns as directed until the coil is complete.

Construction is carried out on a wooden base-board and front panel. Secure the coil to the base-board as shown on Fig. 18b, mount the remaining components and proceed with the wiring.

When testing out, it will be found that the further the aerial is tapped down the coil towards the earth end, the greater the selectivity. A position should be found which permits separation of local stations without excessive loss of sensitivity.

The crystal tap is adjusted for best results and different settings of C1 tried out, when choosing the best position for the aerial tap.

Components List, Fig. 18a

C1	450pF padder.
C2	500pF variable capacitor.
C3	1000pF mica capacitor.
L1	See text and Fig. 18a.
X1	Germanium Crystal.
2	Crocodile Clips (to connect leads to coil taps).
4	Wander-plugs and sockets.

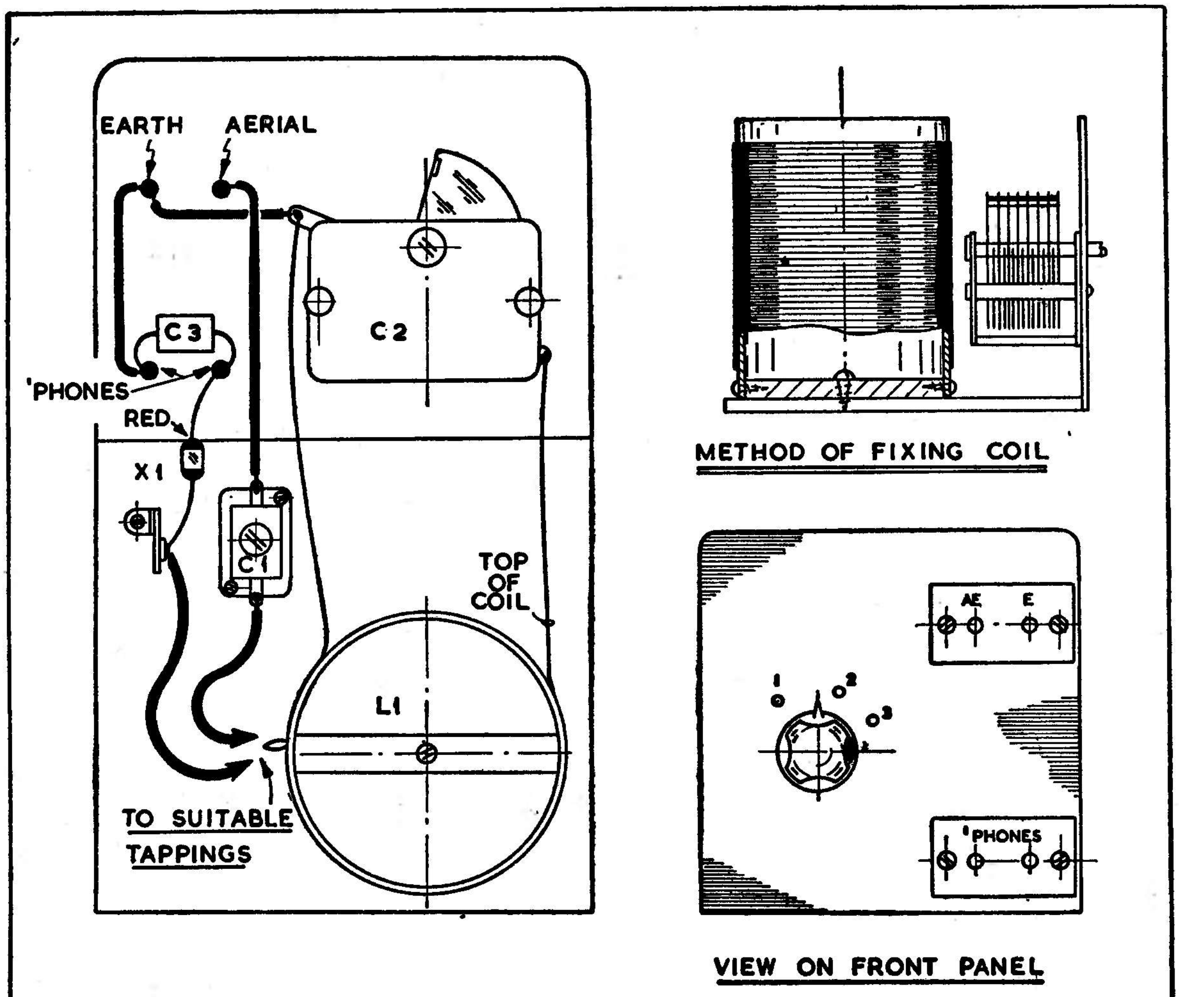

METHOD OF FIXING COIL

VIEW ON FRONT PANEL

FIG. 18 b

4

A most unusual design is reproduced in Fig. 19a.

It first appeared in America about three years ago, and it certainly does offer some advantage over more conventional sets. The tuning circuits are duplicated and two crystals are used, so the circuit may be described as a full-wave receiver. Signals received will certainly be louder than with simple sets, but care must be taken with the coil winding, as with the other receivers using home-made coils, otherwise results will be disappointing. Be very careful to ensure that all three windings are in the same direction, this is very important.

The distance between each winding should be ¼". After the coil has been wound it is a good plan to warm it before a fire and paint the windings with "Durafix." Heating the coils makes the "Durafix" run freely. This substance sets quite hard and there will be no risk of the windings loosening. A small quantity of enamelled wire will cover the requirements of this coil, a 2 oz. reel will provide more than sufficient.

Before wiring in the germanium crystals, examine them carefully, note that one end is coloured red or in some cases marked with a positive sign thus +. You will notice, the sign is the same as the addition symbol used in arithmetic.

It is essential that both the red or positive ends are connected together, note that this is clearly marked on Fig. 19b. The receiver cannot work if

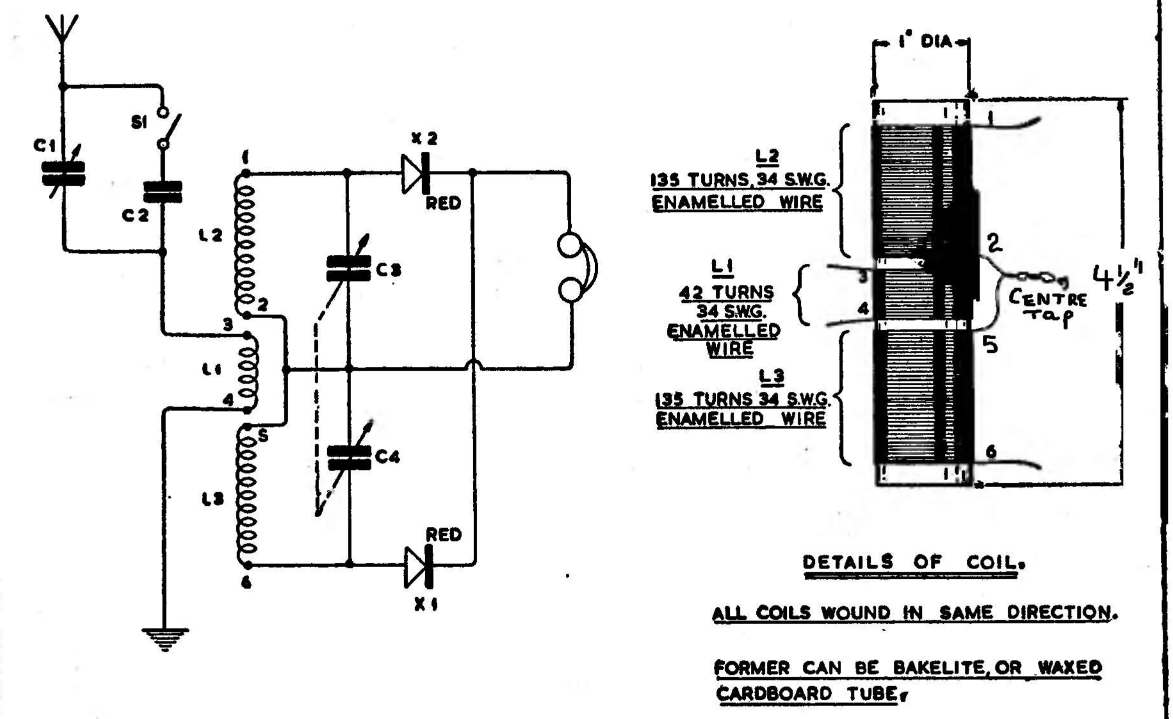
C1
S1
C2
L2
1
2
3
L1
4
5
L3
6
X2
RED
C3
C4
RED
X1
1" DIA
L2
135 TURNS, 34 S.W.G. ENAMELLED WIRE
L1
42 TURNS 34 S.W.G. ENAMELLED WIRE
L3
135 TURNS 34 S.W.G. ENAMELLED WIRE
1
2
3
4
5
6
CENTRE TAP
4½"

DETAILS OF COIL.

ALL COILS WOUND IN SAME DIRECTION.

FORMER CAN BE BAKELITE, OR WAXED CARDBOARD TUBE.

FIG. 19a

one of the crystals is connected the reverse way round.

A baseboard and panel form of assembly is used for this set which looks quite attractive if housed in a small polished cabinet. Dimensions are not at all critical but Fig. 19b gives a general idea of the layout that should be used. To avoid any unnecessary losses the coil should be mounted on perspex.

Obtain two strips of perspex 5½" long by ½" wide, drill fixing holes at both ends of each strip. It is best to clamp the two together when drilling these holes. Place one strip on the board in position, then put the coil and remaining strip over the first one and screw down. The coil will be clamped neatly and rigidly into position.

Twin socket bakelite strips are used to carry the aerial/earth and 'phone connections, if desired terminals mounted on pieces of bakelite could be used. These socket strips can be purchased from most shops selling components. If feet are not provided, drill two holes and screw into the edge of the base-board. Tag-strips are used to anchor some of the wires from the coil, and the germanium crystals. These also can be purchased for a few pence. At least two tags are necessary on one, and three on the other, though if the strips have more tags than required it is of no consequence. C3-C4, the ganged tuning capacitors have trimmers fitted, this should be stipulated when purchasing, and when first operating the set these small trimming capacitances are set about half-way. The idea of these trimmers is, that they compensate for any difference between the self-capacity of the tuning coils. With the switch S.1 open, that is in the OFF position, turn

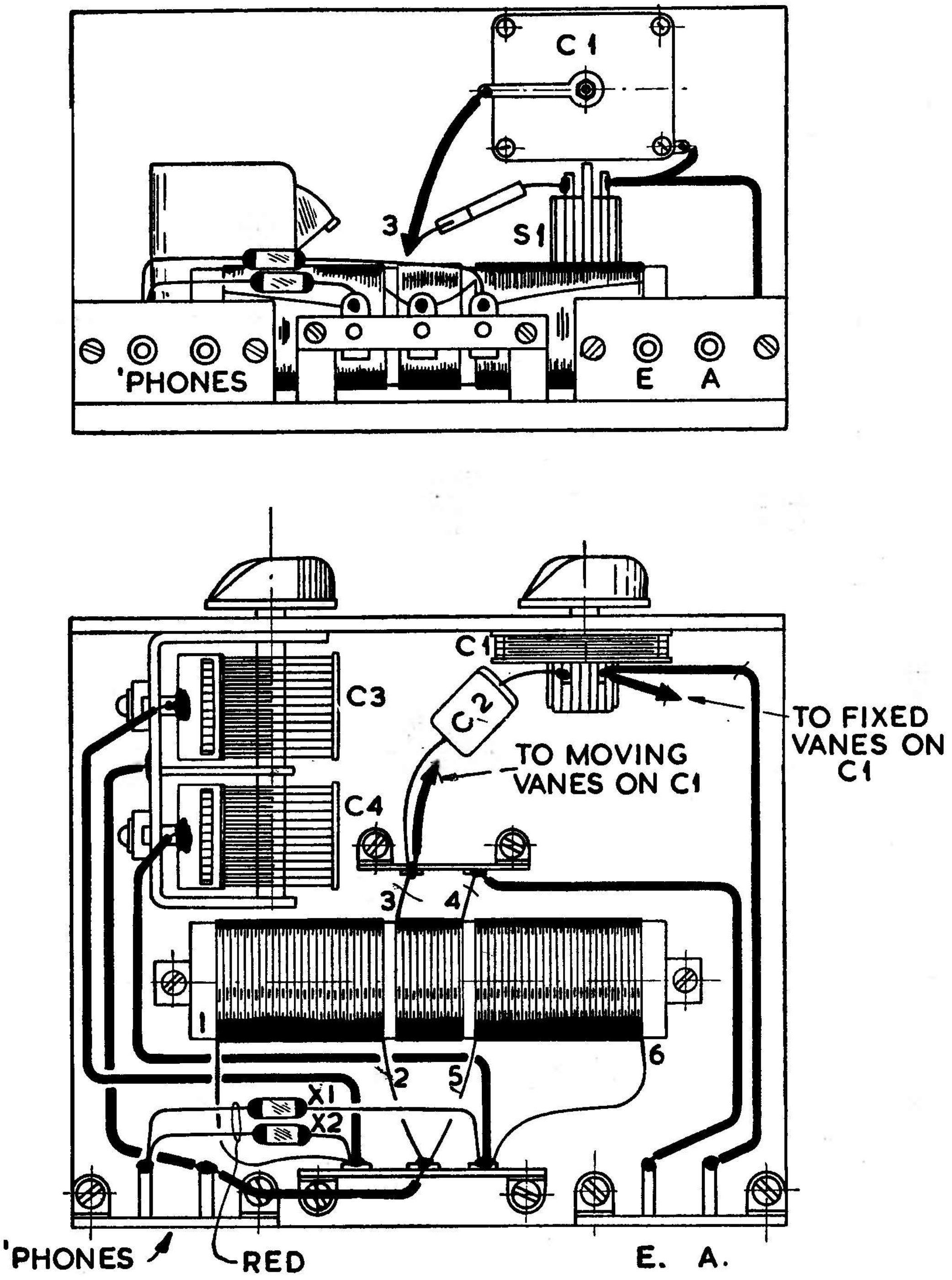
C1
3
S1
'PHONES
E
A
C1
C3
C2
C4
TO FIXED VANES ON C1
TO MOVING VANES ON C1
3
4
1
2
5
6
X1
X2
1
'PHONES
RED
E.
A.

FIG. 19 b

in a station at the high-frequency end of the band, that is with the vanes of C3-C4 towards the disengaged position. Adjust C1 for maximum volume, without interference from other stations. Now adjust the trimmers for loudest headphone strength. It may be necessary to run over the adjustments several times for best results. Once they are correctly set, the trimmers require no further adjustment. The tuning coil covers the M.W. band, 1500 kc/s. to 600 kc/s., as the capacitor C3-C4 is advanced and the vanes start to mesh, the frequency to which the receiver tunes **decreases** so that with the vanes disengaged the set is tuned to 1500 kc/s. approximately and at full mesh, 600 kc/s. Switch S1 is used as follows, for stations **lower** in frequency than 850 kc/s it should be in the ON position with the contact closed, but for stations **higher** in frequency, it is in the OFF position. As an example both the London Home Service and London Light Programme transmitters on the M.W. band operate at a higher frequency than 850 kc/s. If any doubt exists as to the operating frequency of a given station, reference should be made to the Radio Times which quotes both wave-lengths and frequency. If you know the wave-length in metres, it is simple to find the frequency. Divide 300,000 by the wave-length in metres, the dividend equals the frequency in kilocycles, i.e., 300,000 ÷ 300 metres=1,000 kc/s. In the same way, dividing 300,000 by the frequency in kilocycles will produce the wave-length in metres.

To use the receiver, tune in the signal by C3-C4, adjust C1 for maximum volume without allowing stations to overlap.

Components List, Fig. 19a

C1	300pF solid dielectric variable capacitor.
C2	1000pF mica capacitor.
C3-4	500pF twin gang variable capacitor.
L1-2-3	See text and Fig. 19a.
X1-2	Germanium Crystal.
2	Tag-strips.
2	Twin socket strips.
S1	Single Pole toggle switch.

5

Whenever crystal sets are discussed, constructors are apt to think in terms of medium-wave reception. When you consider that at least 98% of published circuits are designed for this band (sometimes with the long-wave band thrown in as an afterthought) it is understandable. However, if you can provide a good outdoor aerial (and an equally good earth) there is a lot of fun to be had listening to the short-wave bands. This receiver is designed specially for short-wave reception.

A metal chassis is used for construction, these can be obtained ready made in aluminium from most good supply houses. It can be quite small and on the original model a 6″ x 4″ was used. Low loss components are used so as to obtain greatest efficiency on the short wave bands. The coil is a commercial product which plugs into a 4-pin base. Best results were obtained on the 3mc/s to 7mc/s band though coils covering other bands are available.

Note that the tuning capacitor C1 is smaller than normally used and has a maximum capacitance of 140pF.

Coil type 706/R covers the 3mc/s to 7mc/s band, but the set is equally satisfactory on the M.W. band, and to cover this coil type 706P should be used. This coil has an iron dustcore which can be adjusted to make the coil cover the required band. The effect of the core is as if turns were being added or removed from an ordinary coil. When dealing with the short waves it is customary to refer to the frequencies in terms of mega-cycles (mc/s) rather than kilo-cycles. One mega-cycle is equivalent to 1000 kilo-cycles, that is 1,000,000 cycles.

Layout and wiring can be clearly followed from Fig. 20b and a theoretical diagram is shown in Fig. 20a. Care should be taken when arranging a mounting for the aerial terminal. The best plan is to cut a 1″ hole in the chassis, mount the aerial terminal on a piece of perspex and screw into the chassis. This avoids any unnecessary losses between aerial and chassis. The earth terminal is screwed directly into the chassis.

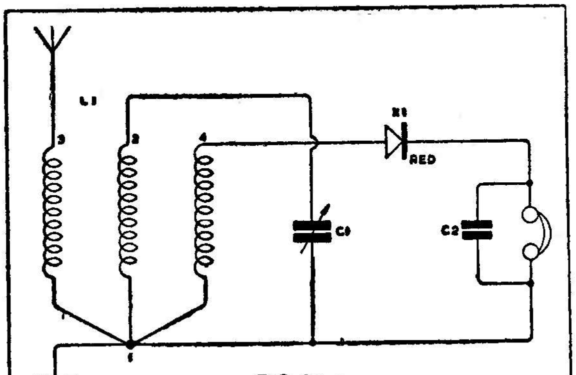

FIG. 20 a

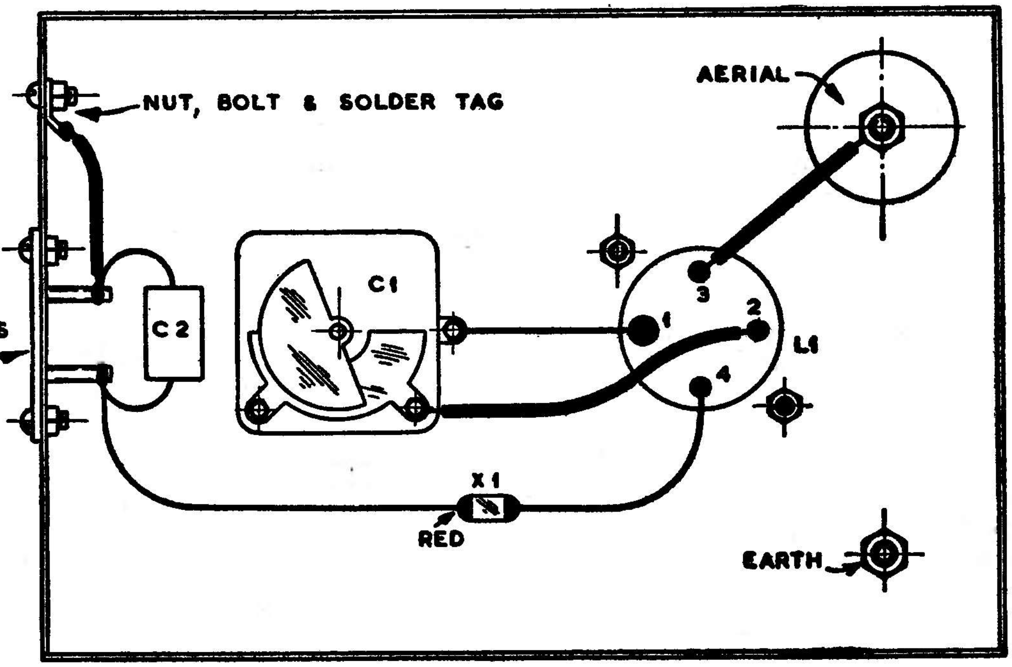

FIG. 20 b

Wiring as shown in Fig. 20a is likely to give the best results, but other arrangements are possible and to get the best out of the set they should be tried.

Refer to Fig. 20a. The aerial is connected to pin 3 on L1, C1 to pin 2 and X1 the crystal to pin 4. If you have a long aerial try this combination—Aerial to pin 4, C1 to pin 2, and the crystal to pin 3. Connections to pin 1 are not altered. In a few cases the following arrangement may prove best. Crystal and C1 to pin 2, aerial to pin 3, pin 4 left free and again pin 1 is unaltered.

Tuning on the short-wave bands is more critical than on the medium waves so tune very slowly over the band and remember signal strength is likely to vary from day to day.

Components List, Fig. 20a

C1	140pF variable capacitor Eddystone 586.
C2	1000pF mica capacitor.
1	Coil holder Eddystone 707.
L1	3m/c—7m/c coil Eddystone 706/R.
1	Engraved tuning dial.
1	6"x4"x2½" chassis.
2	Terminals (Aerial-Earth).
1	Twin socket strip ('phones).
X1	Germanium Crystal.

6

Fig. 21a-b is yet another design from America and appeared in Radio Craft some years ago, designed by Mr. W. J. Spain. The original used a silicon crystal which preceded the development of the germanium. Results are certainly very good though of course on the model built here, a germanium crystal was used. Selectivity can be adjusted to suit all conditions and even at short distances from local stations. Good volume can be obtained without the programmes overlapping. Home made coils are used and for best results a 75′ outdoor aerial is desirable.

Baseboard and panel construction is used, which can conveniently be housed in a small wooden cabinet.

Operation is a little more complicated than some of the other designs, but the results certainly merit the extra trouble taken. As with the other receivers much of the success obtainable is due to the coils and these must be carefully made otherwise results will be disappointing.

Layout and wiring can be followed quite easily from Fig. 21b and requires no special comment.

First make the coils. L1 is a tapped coil, the taps should be made in the same way as shown in Fig. 18b. It consists of 90 turns of 22 D.C.C. (double cotton covered) copper wire tapped at the following number of turns, 5, 10, 15, 25, 30, 40, 50, 60, 70 and 80. The former used has a diameter of 2". L2 is also wound on a 2" diameter former, this coil however is not tapped, it consists of 110 turns of 38 D.C.C. or enamelled copper wire. Great care must be taken when handling this wire, it is quite fine and in the hands of the inexperienced will easily break.

After wiring has been carried out the set can be put into operation.

Selectivity is controlled by S2, in position B tuning is broad but in position A it can be quite sharp.

When searching for a station switch S2 to B. Tune the station and switch S2 to A. The frequency range is controlled by the position of the tap on L1, a good plan is to start with the 5 turn tap on L1.

Components List, Fig. 21a

C1	500pF variable capacitor.
C2	500pF variable capacitor.
C3	1000pF Mica Capacitor.
X1	Germanium Crystal.
L1 & L2	See text.
S1	Single Pole toggle switch.
S2	Single Pole 2-way toggle switch.
2	Twin socket strips (Aerial/Earth and 'Phones).

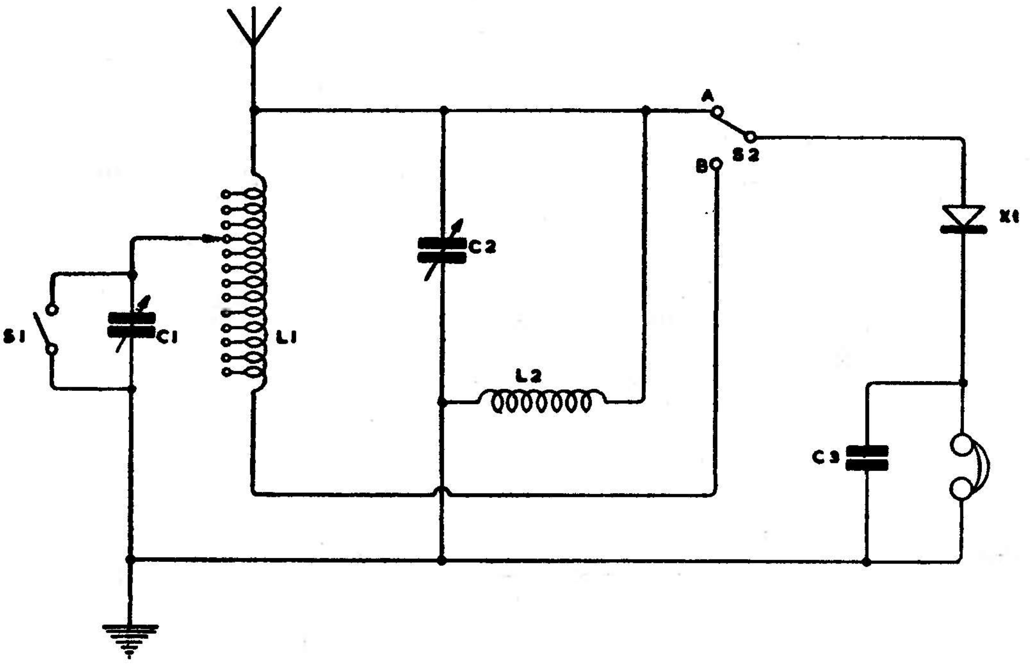

FIG. 21 a

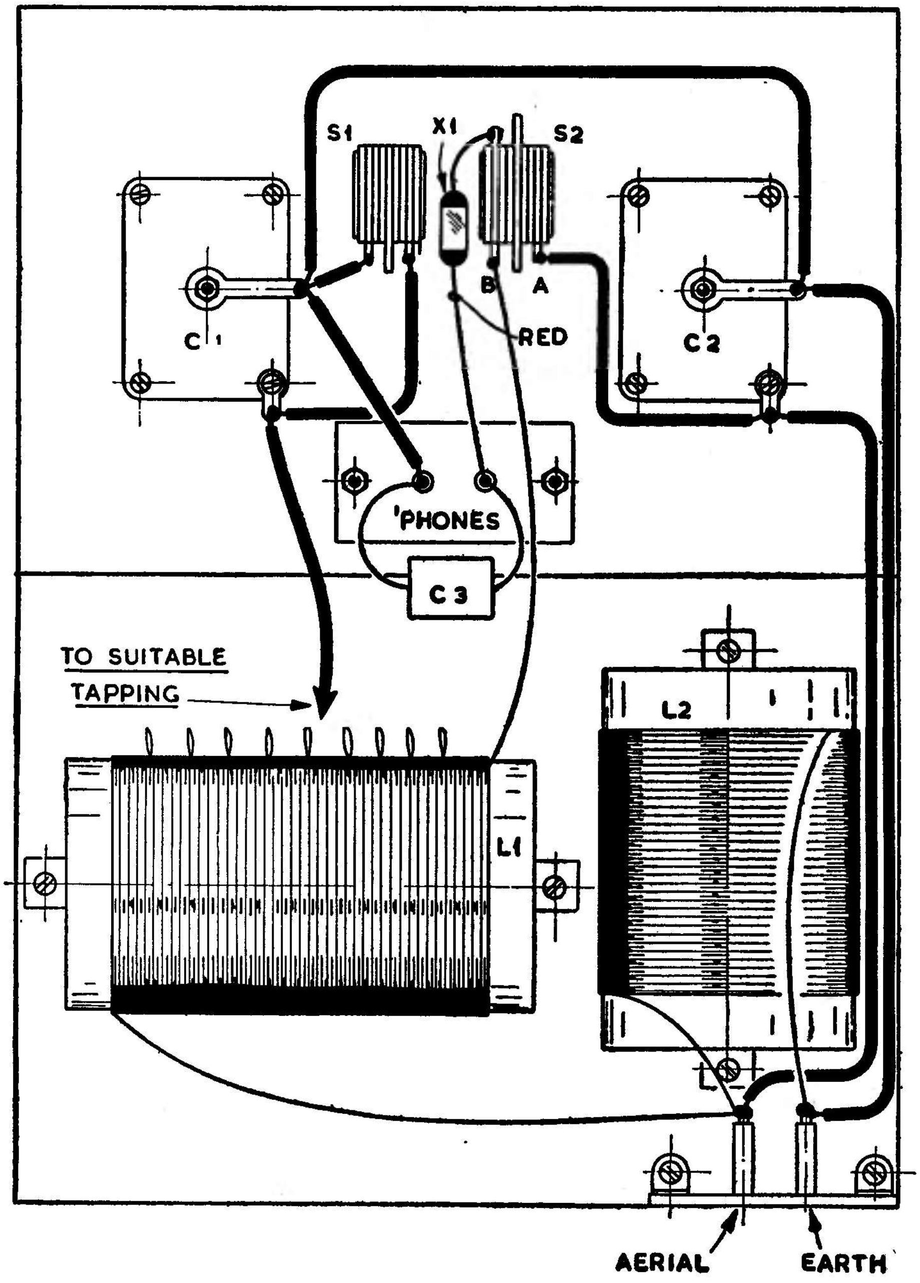
S1
X1
S2
C 1
B
A
RED
C 2
'PHONES
C 3
TO SUITABLE TAPPING
L2
L1
AERIAL
EARTH

FIG. 21 b

Keep S1 open and C1 for tuning, adjusting the tap on L1 for best results. Very powerful signals are best brought in with S1 closed. When using C1 and adjusting the taps on L1, switch S2 is kept in the A position.

It will be found that C2 acts as a fine control on C1. One division of the tuning dial on C1 is roughly equal to a movement of 20 divisions on C2.

Note that the coils are mounted at right angles to one another and that in this instance wood has been used as far as possible, an aluminium or other metal chassis is not suitable.

7

A very useful receiver for both medium and long-wave reception is possible with the circuit of Fig. 22a-b. A home-made coil is used. On the original a 1¼" diameter former was used but in fact this is not very critical and a 1½" former could be used. An interesting point is the construction of the long-wave coil section. Solenoid coils as used for the medium wave-band are not efficient if the length is too great relative to the diameter. Now as the long-wave section has 300 turns, if wound as a solenoid the winding length would be very great and a lot of efficiency would be lost. To overcome the difficulty the coil is pile wound. If the whole coil were to be wound in one pile it would still be inefficient because of the capacity formed in the winding, but by dividing it into five sections a coil of reasonable efficiency results. The same remarks concerning winding apply to this coil as to all the others in

Components List, Fig. 22a

C1	300pF Mica Capacitor.
C2	500pF Variable Capacitor.
C3	1000 Mica Capacitor.
X1	Germanium Crystal.
L1-2	See Text.
S1	Single Pole Toggle Switch.

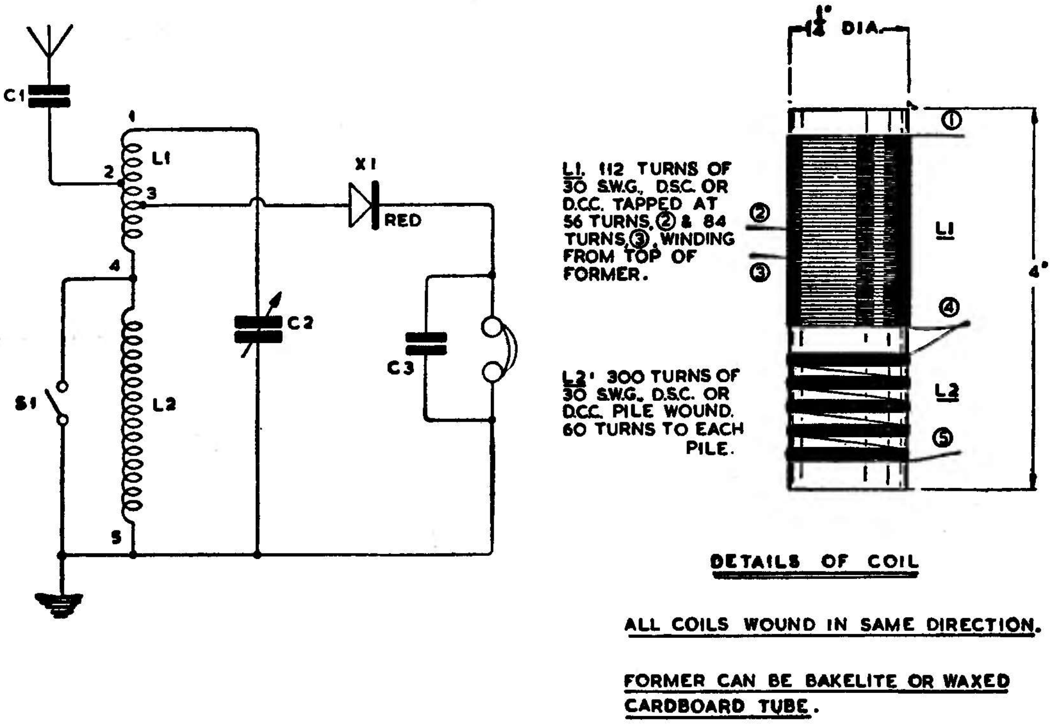
C1
1
2
3
L1
4
S1
L2
5
C2
X1
RED
C3
1¼" DIA.
①
②
③
④
⑤
L1
L2
4"
L1. 112 TURNS OF 30 S.W.G., D.S.C. OR D.C.C. TAPPED AT 56 TURNS, ② & 84 TURNS, ③, WINDING FROM TOP OF FORMER.
L2' 300 TURNS OF 30 S.W.G., D.S.C. OR D.C.C. PILE WOUND. 60 TURNS TO EACH PILE.
DETAILS OF COIL
ALL COILS WOUND IN SAME DIRECTION.
FORMER CAN BE BAKELITE OR WAXED CARDBOARD TUBE.

FIG. 22 a

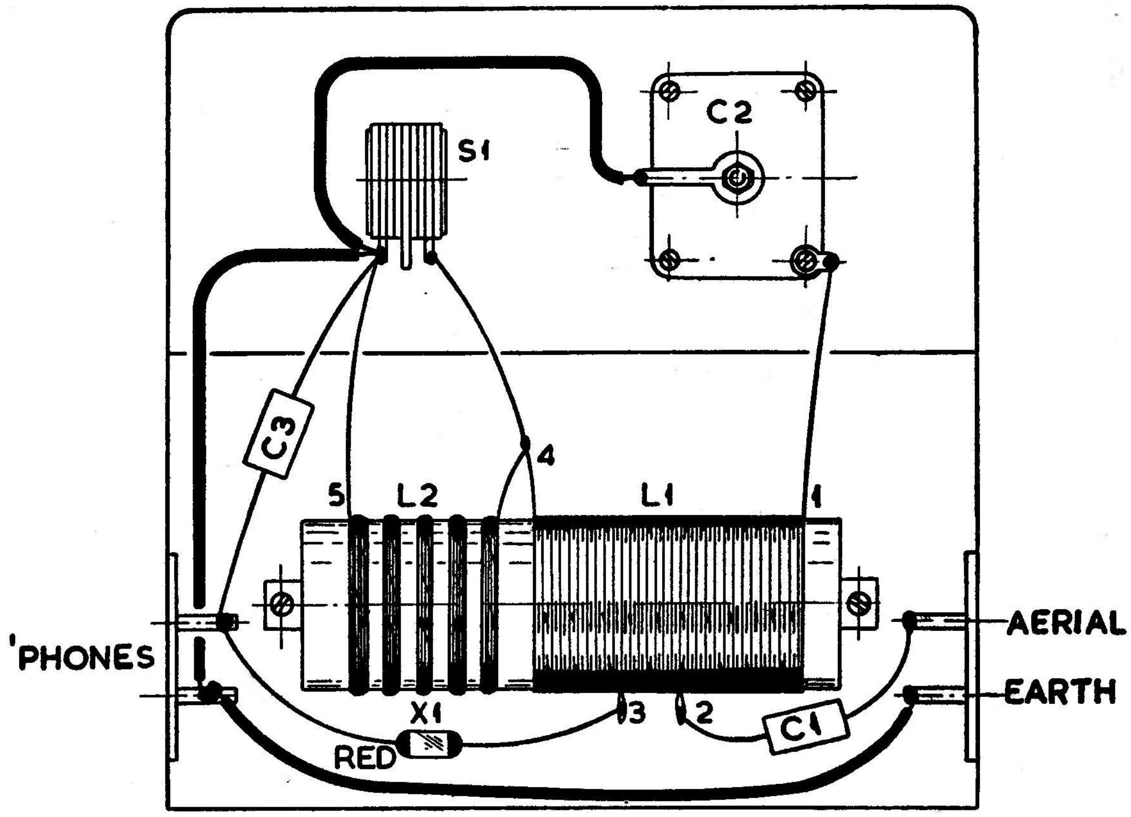

FIG. 22 b

this book, and readers should not, by this time, encounter any difficulties.

The design provides good selectivity with good sensitivity, and provided the set is not required to operate within a few miles of a powerful transmitter, no trouble will be encountered from interference between adjacent stations.

A wooden base-board and front panel is best for this design, and the wiring and layout can be clearly followed from Fig. 22b. Switch S1 is open for long-wave reception and closed for the medium-wave band.

Coil mounting can be on the same principle as for the full-wave design Fig. 19a-b.

8

If you look at all the other circuits shown in this book you will find that in each case the tuning capacitor and tuning coil are in parallel, this is by no means essential and the circuit of Fig. 23a shows a series arrangement that is the tuning capacitor is between the coil and earth instead of being wired up across it. This design also appeared in **Radio Craft** a few years ago.

An interesting point is that the coil is a conversion adapted from another type of component. To make this coil you must obtain an old I.F. transformer as used in superheterodyne receivers.

You will require one designed to work anywhere between 450 and 470 kc/s. Some of the very early ones worked at 110 kcs/ but these are unsuitable.

The I.F. transformer must also be of the type tuned by an iron dust core at each end.

Dismantle the transformer, remove the parallel capacitors and mounting wires so that you are left with two coil bobbins and their former. Be very careful when un-soldering the wires from the supports because the coils are usually wound in litzendraht and must remain intact. Remove 25% (approximately) of the turns from one coil which we will call the secondary, and completely remove the iron dust slug from this coil, now cut the former in half, and mount as in Fig. 23a. Warm the coils with heat from your soldering iron or from a fire and gently slide the coils into position at the ends of the former as in Fig. 23a. The wax will set again holding the coils in their new position.

Note that the wires from the coils are connected to tags on the mountings. These are pieces of bakelite, and usually the original end pieces used on the I.F. transformer can be used.

A word about soldering litzendraht. Do not attempt to use emery paper. Prepare the end of wire by removing the outer silk covering and dip the end in methylated spirits. Ignite with a match and after about five seconds wipe with a cloth held between finger and thumb. This will quench the flame and will remove all traces of enamel from the wire. It is a good plan to experiment with the odd length of wire removed from the secondary coil. If the coils prove to be constructed of ordinary copper wire, it can of course be prepared with fine emery cloth in the normal manner.

Components List, Fig. 23a

C1	500pF variable capacitor.
X1	Germanium Crystal.
L1	See Text.
4	Plugs and sockets or terminals.

Wiring is straight forward as can be seen from Fig. 23b and requires no explanation.

Operation is as follows, selectivity is controlled by varying the distance between coils with the wing nut, and to an extent by the position of the primary iron dust core. Remember that the aerial and earth installations are as important with this set as with any other crystal receiver. No parallel capacitance appears necessary across the 'phones, the set works quite happily relying on the self-capacitance of the head-phones.

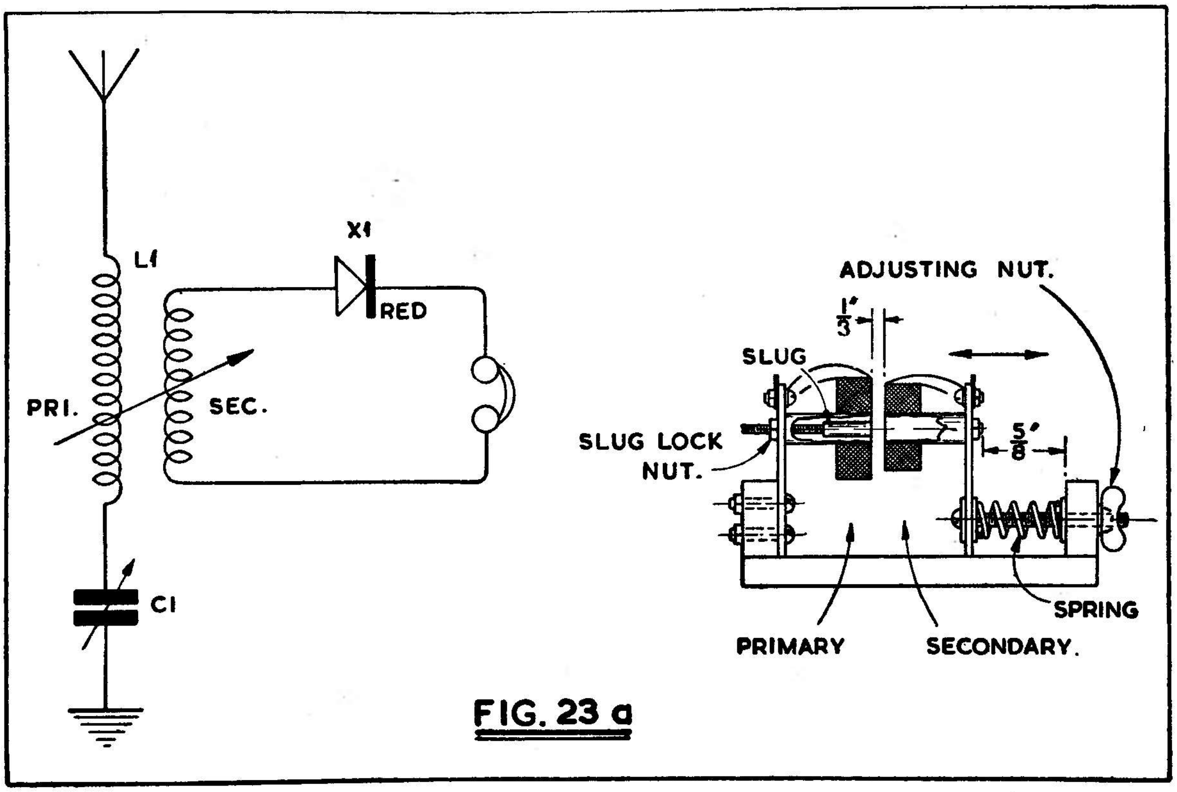

FIG. 23 a

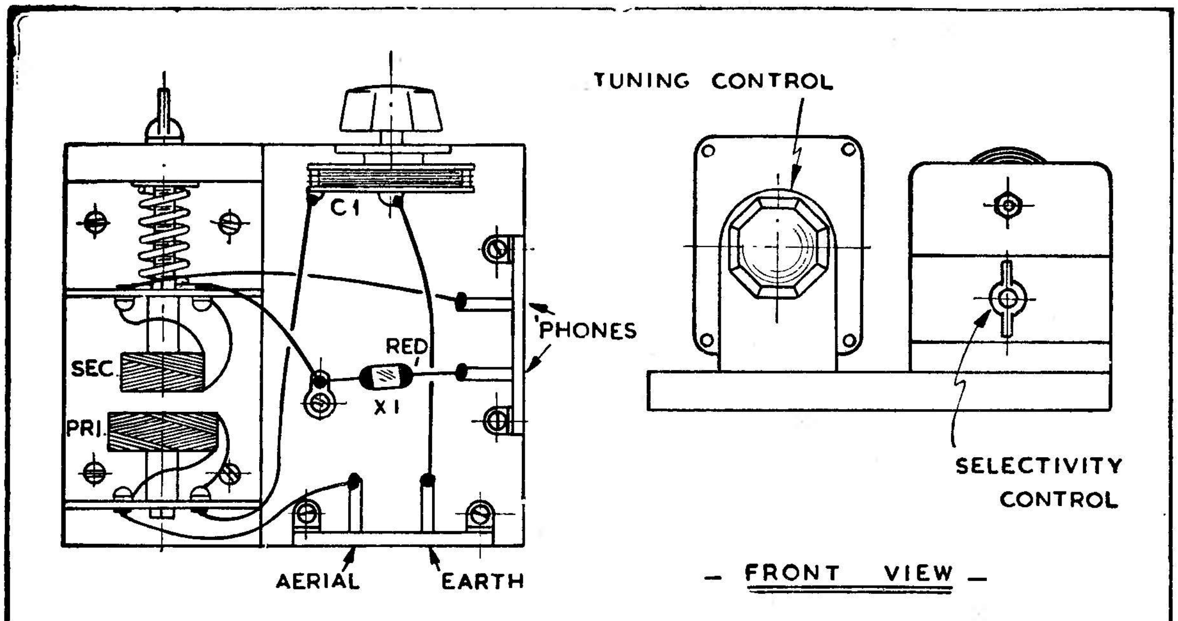
TUNING CONTROL
C1
SEC.
PRI.
RED
X1
PHONES
AERIAL
EARTH
SELECTIVITY CONTROL
— FRONT VIEW —

FIG. 23 b

9

So as to give a truly varied selection of circuits, some of the sets have been designed around commercial coils whilst others use coils which must be home-constructed.

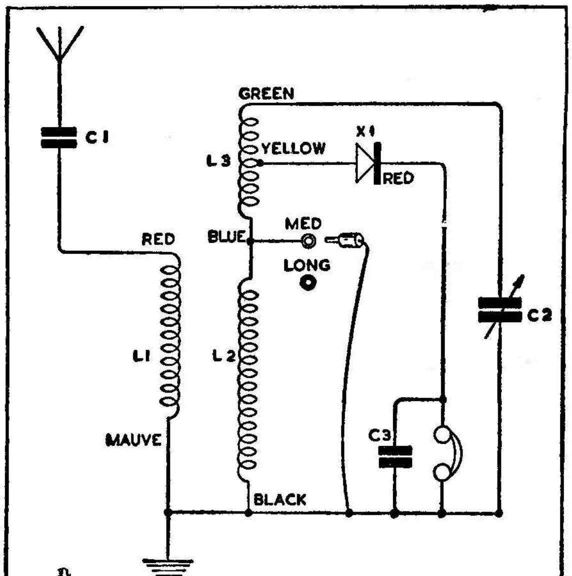
C1
GREEN
L3
YELLOW
X1
RED
RED
BLUE
MED
LONG
C2
L1
L2
MAUVE
C3
BLACK

FIG. 24 a

Most enthusiasts like to build their own coils but when it is a question of something really small combined with efficiency a commercial coil is the obvious choice.

Fig. 24a-b uses a tobacco tin or a similar container just as the design shown in Fig. 16a.

This receiver however is a little more selective and is more suitable than the earlier design, if you are near a transmitter or have a very long aerial. Take great care to connect the coil according to the colours shown on the diagrams. Sockets are used for wave-changing, this cheapens the cost a little and they are just as efficient.

Components List, Fig. 24a

C1	100pF Mica capacitor.
C2	500pF variable capacitor solid dielectric.
C3	1000pF Mica capacitor.
X1	Germanium crystal.
L1-2	R.E.P. dual range coil. Blue box.
6	Insulated wander-plug sockets.
5	Wander Plugs.
2oz.	Tobacco tin or similar container.

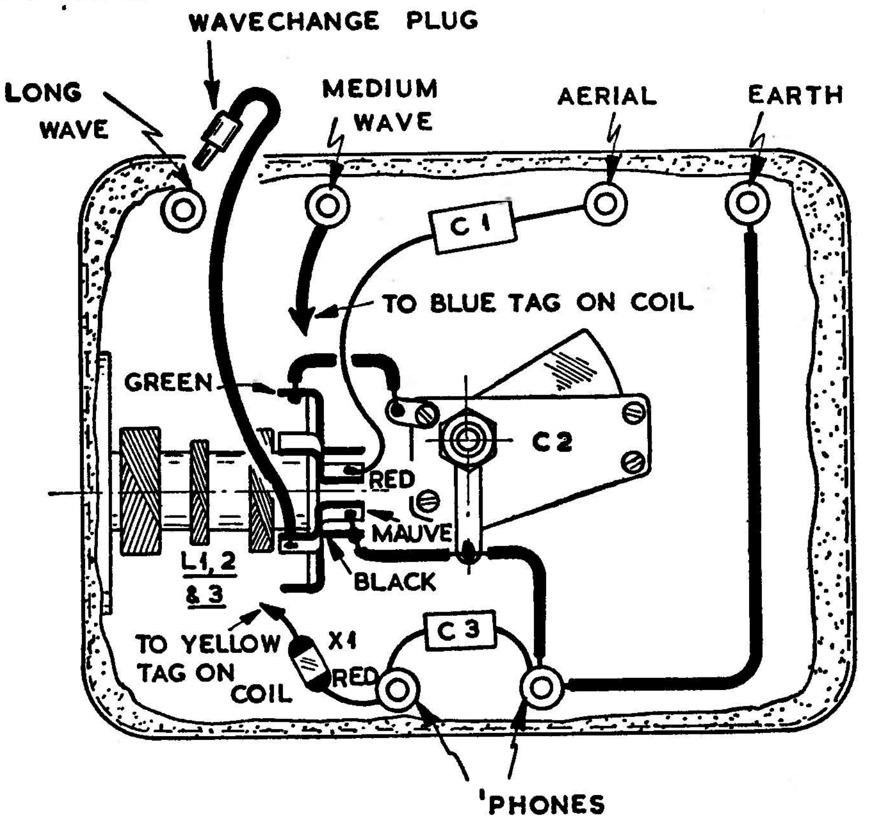

LID SHOWN BROKEN AWAY TO REVEAL POSITION OF COMPONENTS.

FIG. 24 b

10

In certain cases, signals from the local transmitter are too powerful to permit a crystal set with a normal coil to be of much use. I myself live in Hertfordshire a few miles from the London transmitters, and find a more elaborate circuit is necessary. Now in radio circles it is a well-known fact that the greater the number of tuned circuits, the greater the overall selectivity. Look at Fig. 25a, you will see that the usual L and C tuning arrangements have been duplicated. This type of circuit is known as a Band-Pass Filter. There are many kinds of such filters, this particular type is an inductively-coupled filter, note that energy from the first half (L2) is conveyed to L3 by way of two small inductors L1 and L4.

Tuning is quite sharp and the "feel" of the set is quite different from that of single coil designs. A small chassis is ideal for construction. 6"x4"x2½" was used on the original, though it could have been smaller.

Fig. 25b gives the layout and wiring.

Take particular notice that L1-2-3 is mounted horizontally and at right angles to L4-5-6, this is to prevent the coupling from being too "tight" which would ruin the idea of the filter. A good aerial and earth is of course necessary. It may be found on test that the selectivity is too great

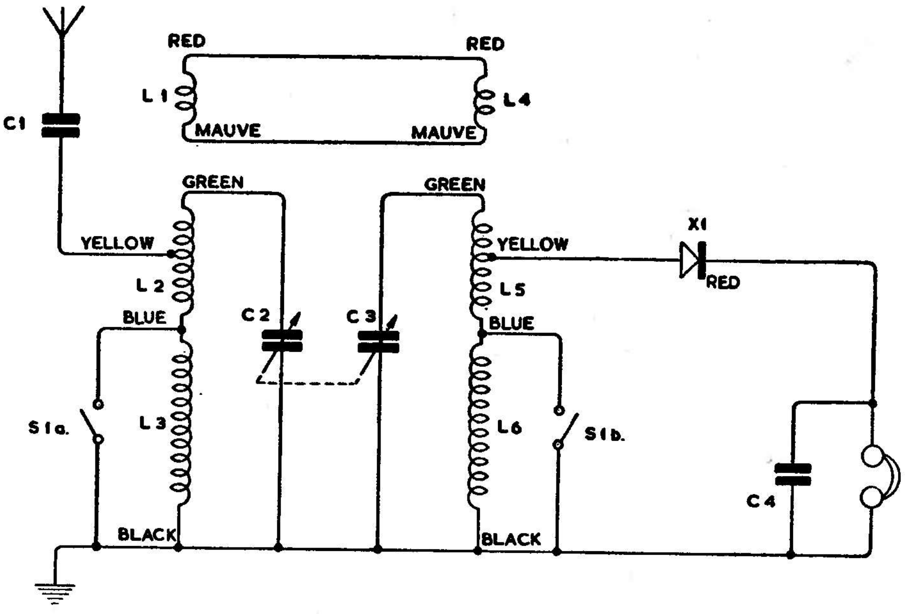

FIG. 25 a.

and that volume has suffered. This can be ingeniously overcome by connecting a very small capacitance between the two Green tags on the coils. The simplest way is to connect it across the fixed vane connections on C2-C3. Values for this extra capacitance are a matter for experiment usually a value between 10pF and 47pF is sufficient. A small 50pF trimmer could be used and adjusted to suit. The ideal setting is to enable powerful adjacent stations to be received just short of overlap so that as much volume as possible is obtained.

There is one point to be observed with the ganged capacitor. It should be of the type fitted with trimmers though of course, there is no reason why you should not fit them yourself. These are to allow for differences between coils to be balanced out, and consist of a small variable capacitance across each main section.

When setting up the receiver, first set each trimmer at half way and tune in a station near the high-frequency end of the band (vanes nearly out) adjust the trimmers for maximum volume. If extra coupling capacity is added as described, try adjusting them further.

Once properly set up, no further adjustment is necessary.

Components List, Fig. 25a

C1 100pF Mica capacitor.
C2-3 2 x 500pF variable capacitor (see text).
C4 1000pF Mica capacitor.
L1-2-3 Dual Range Coil R.E.P. (Blue Box).
L4-5-6 Dual Range Coil R.E.P. (Blue Box).
X1 Germanium Crystal.
S1a-b 2 pole single throw toggle switch.
4 Terminals or plugs and sockets.

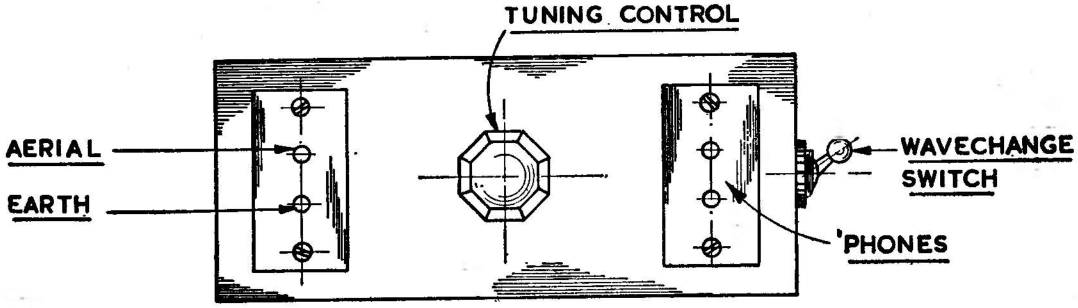
TUNING CONTROL
AERIAL
EARTH
WAVECHANGE SWITCH
'PHONES

FRONT OF COMPLETED RECEIVER.

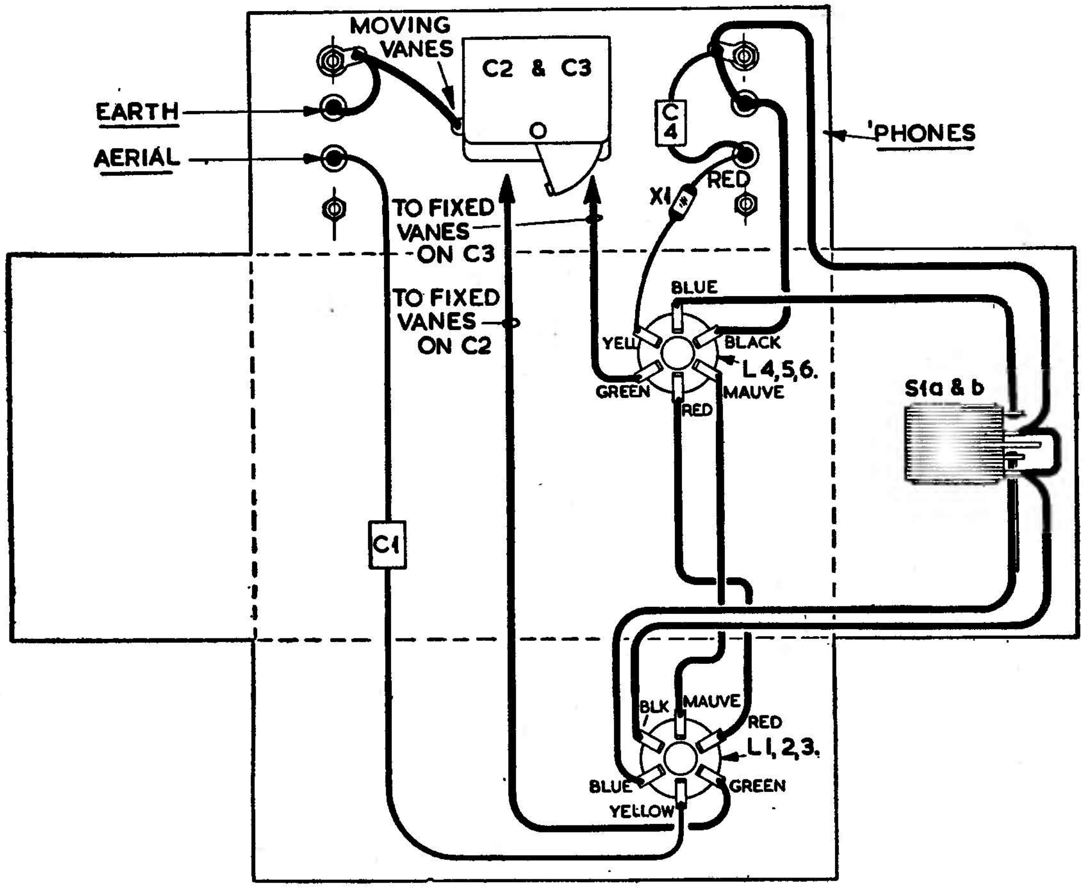
MOVING VANES
C2 & C3
EARTH
AERIAL
C4
'PHONES
X1
RED
TO FIXED VANES ON C3
TO FIXED VANES ON C2
BLUE
YELL
BLACK
L 4,5,6.
GREEN
RED
MAUVE
S1a & b
C1
BLK
MAUVE
RED
L1,2,3.
BLUE
YELLOW
GREEN

EXPLODED UNDER-CHASSIS VIEW

FIG. 25 b

11

Another band-pass circuit is shown in Fig. 26a. This time home-made coils are used. This filter is not inductively coupled but entirely capacitive, relying on C3. As in the previous circuit it must be adjusted to give just sufficient " broadness " to the tuning to provide adequate volume without station overlap. To avoid unwanted coupling a screen is mounted between the coils, this is necessary because due to the size of the coils they would be bound to couple quite tightly without it. Such precautions were not necessary with the circuit of Fig. 25a, as the commercial coils are much smaller physically and the risk of unwanted coupling reduced.

The coils for this receiver are for medium wave reception only, this keeps the size down. Coil winding procedure has already been fully explained, and covers all the requirements of these coils. Base-board and panel construction is used, and for more accurate tuning since the coils are home-made, separate tuning capacitors are recommended. Fig. 26b gives the layout and practical wiring.

Components List, Fig. 26a

C1	100pF Mica capacitor.
C2	500pF variable capacitor.
C3	50pF Mica trimmer.
C4	500pF variable capacitor.
C5	1000pF mica capacitor.
L1-2	See text.
X1	Germanium Crystal.
4	Terminals or plugs and sockets.

For those who do not wish to bother with these a ganged capacitor can be used, but tuning may not be as good as with separate ones. The screen mounted between the coils is of aluminium or copper and should be about 1″ higher than the coils and the same length as the baseboard depth.

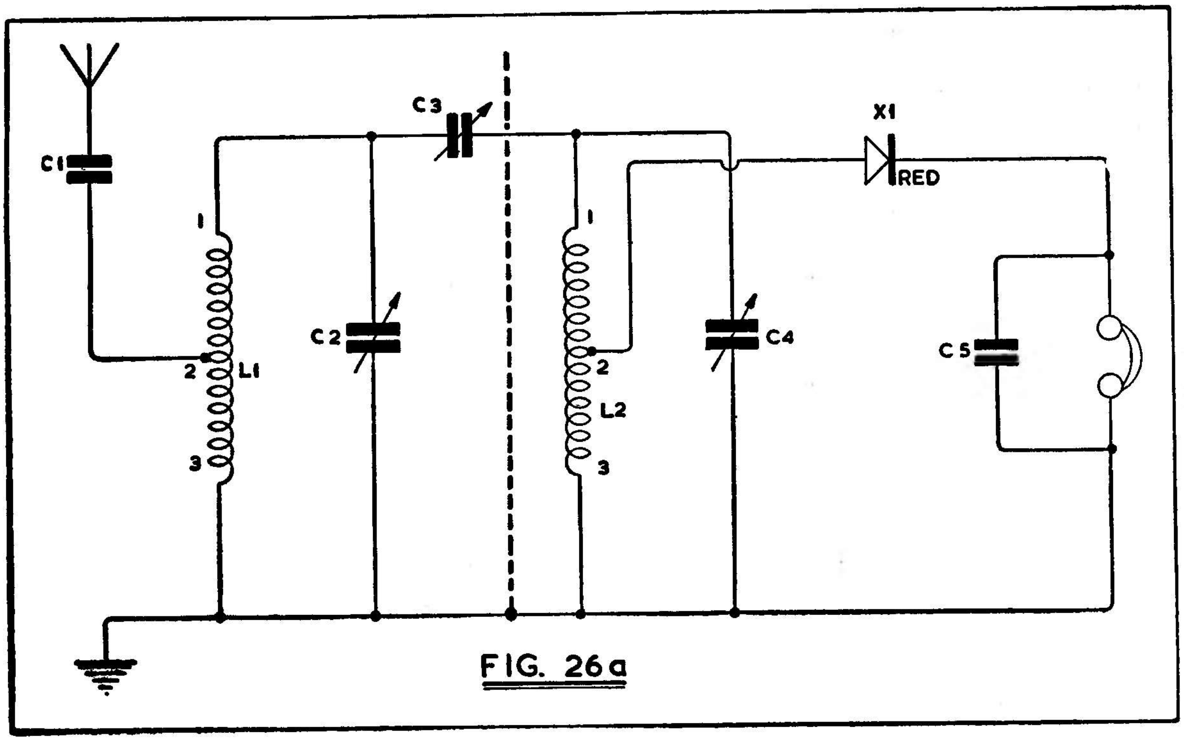

FIG. 26a

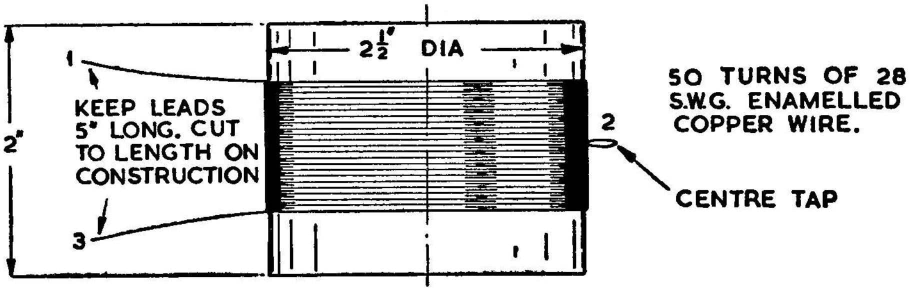
2½" DIA
1
KEEP LEADS 5" LONG. CUT TO LENGTH ON CONSTRUCTION
2"
3
2
50 TURNS OF 28 S.W.G. ENAMELLED COPPER WIRE.
CENTRE TAP

COIL DETAILS.

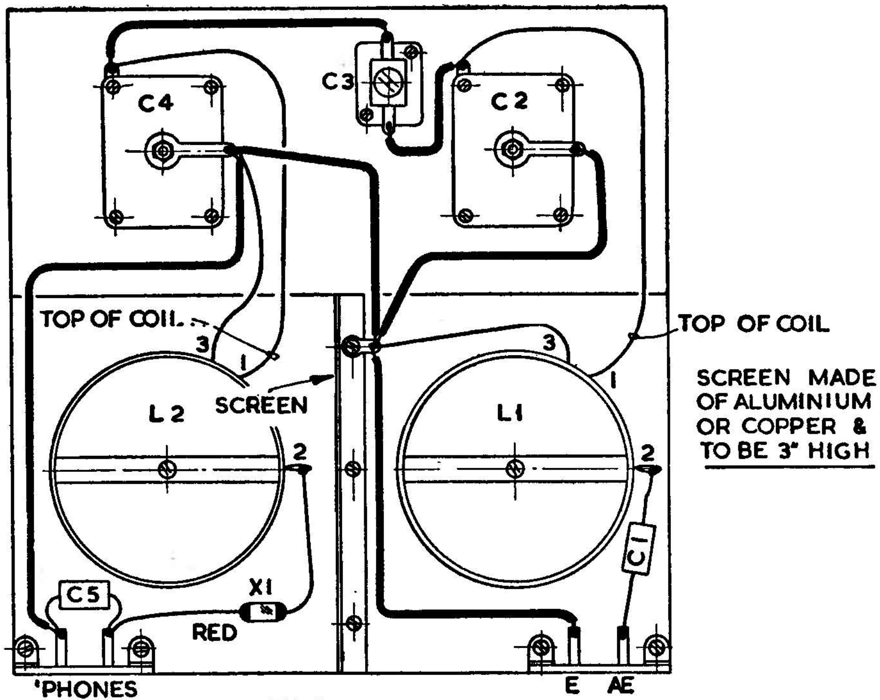
C4
C3
C2
TOP OF COIL
3
1
L2
SCREEN
2
TOP OF COIL
3
1
L1
2
SCREEN MADE OF ALUMINIUM OR COPPER & TO BE 3" HIGH
C1
C5
X1
RED
'PHONES
E
AE

FIG. 26 b

12

Fig. 27a-b uses a band-pass filter for tuning, and the coils are home-constructed.

With this design a ganged capacitor is quite suitable as the coils tune quite accurately. It is essential that the coupling condenser C3 is **non-inductive,** there will be no difficulty in obtaining a new component of this pattern but it is as well to avoid old components which may not employ this form of construction.

Receivers using this type of filter are used to the best advantage when situated a short distance from a powerful transmitter and if selectivity is not important this particular design is not the most suitable.

Both medium-waves and long-waves are catered for: consider L1-2-3, this is the first half of the filter.

L1 is the aerial coupling coil, with S1 closed, the medium-waves are received. This coil is inductively coupled to L2 and L3. Again with S2 open L2 and L3 combined cover the long-waves and when S2 is closed the medium waves are covered.

Tuning for this half of the filter is by means of the capacitor C1, which is one half of the ganged capacitor. A screen is erected between the two sets of coils comprising the filter, and coupling is effected by C3, an earth return for the coils is provided by R1.

L4 and L5 operated in a similar manner to L2 and L3, tuning is by means of C2, the remaining section of the ganged capacitor. You will realise that the coils must be accurately wound otherwise tuning will not remain constant between the two sections over the band.

The ganged capacitor should be fitted with trimmer capacitors when purchased so that the two circuits can be balanced.

You will have noticed that this circuit uses three sets of switch contacts. A three-pole two-way wafer switch is the best choice otherwise either two or even three separate switches are necessary. All the necessary coil details are given in the diagrams, the tap on L4 is made in the same way as shown for other coils in the manual. One other component used only on this design, is the resistor R1, its size is not important because the current flowing is so small it can be ignored.

A ¼ watt rating is adequate, its precise value is not critical so that a 20% tolerance component is quite suitable.

Resistors are coded by colours and the 1000Ω type you require will bear the following sequence. At one end will be painted three coloured rings.

Components List, Fig. 27a

C1-2	500pF 2 gang, variable capacitor.
C3	0.05mfd. non-inductive paper capacitor.
C4	1000pF mica capacitor.
R1	1000 ohm resistor.
X1	Germanium Crystal.
S1-2-3	3-pole 2-way wafer switch.
L1-2-3 & L4-5	See text.
4	Terminals (Aerial, Earth and 'Phones).
	Sheet copper or aluminium for screen.

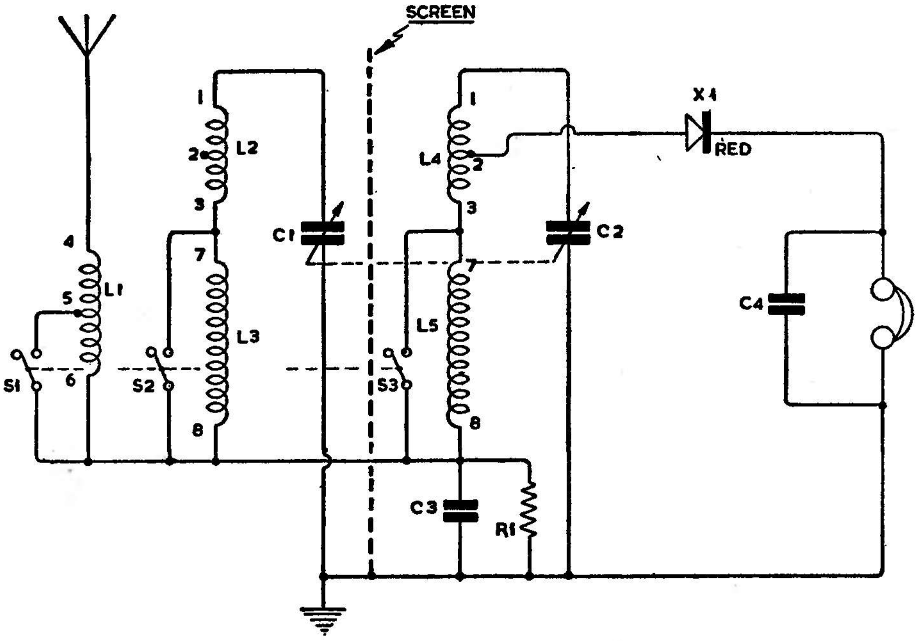

FIG. 27a

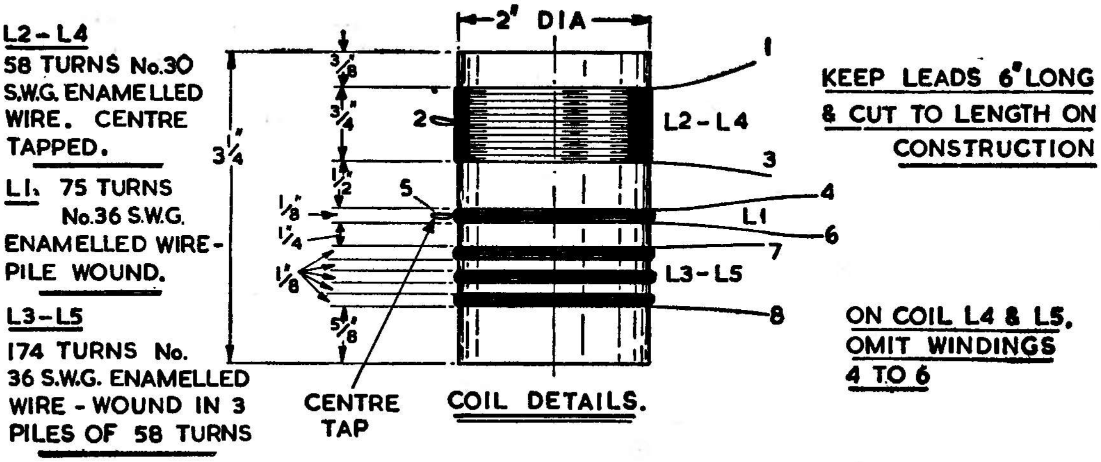
L2 - L4
58 TURNS No.30 S.W.G. ENAMELLED WIRE. CENTRE TAPPED.
L1. 75 TURNS No.36 S.W.G. ENAMELLED WIRE - PILE WOUND.
L3 - L5
174 TURNS No. 36 S.W.G. ENAMELLED WIRE - WOUND IN 3 PILES OF 58 TURNS
3¼"
⅜"
¾"
½"
⅛"
¼"
⅛"
⅝"
2' DIA
1
2
3
4
5
6
7
8
L2 - L4
L1
L3 - L5
CENTRE TAP
KEEP LEADS 6" LONG & CUT TO LENGTH ON CONSTRUCTION
ON COIL L4 & L5, OMIT WINDINGS 4 TO 6

COIL DETAILS.

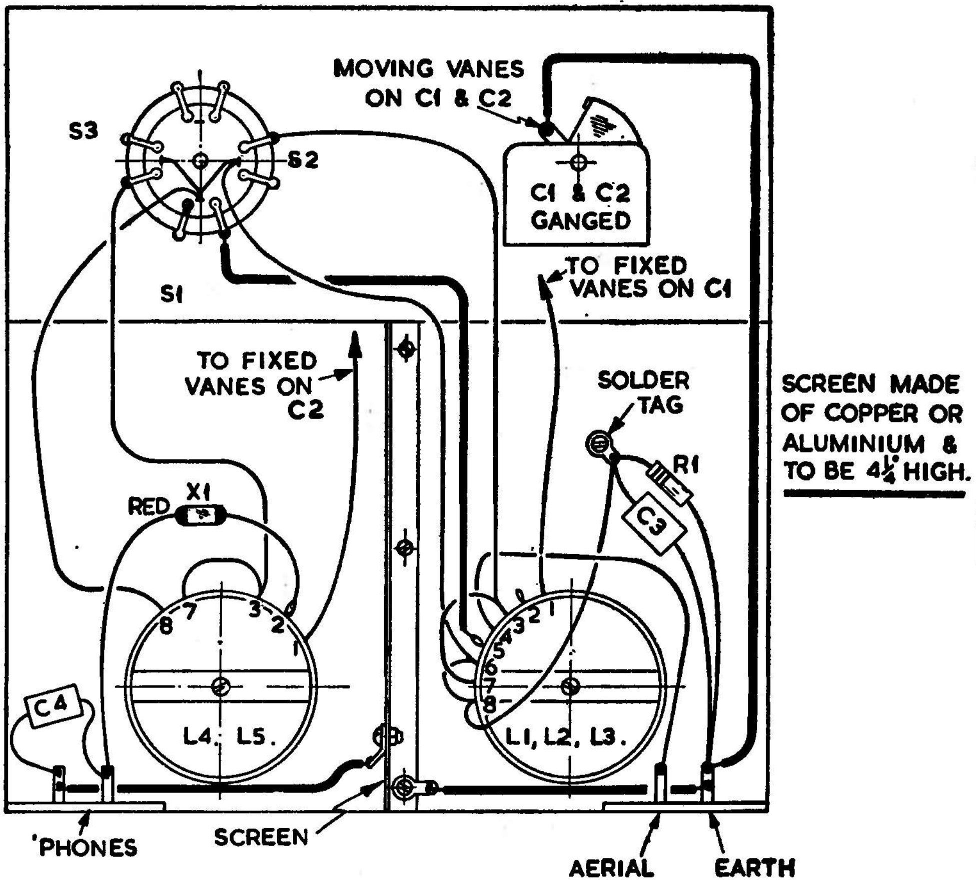
MOVING VANES ON C1 & C2
S3
S2
S1
C1 & C2 GANGED
TO FIXED VANES ON C1
TO FIXED VANES ON C2
SOLDER TAG
R1
C3
SCREEN MADE OF COPPER OR ALUMINIUM & TO BE 4¼" HIGH.
RED
X1
8
7
3
2
1
L4, L5.
4
3
2
1
5
6
7
8
L1, L2, L3.
C4
'PHONES
SCREEN
AERIAL
EARTH

FIG. 27 b

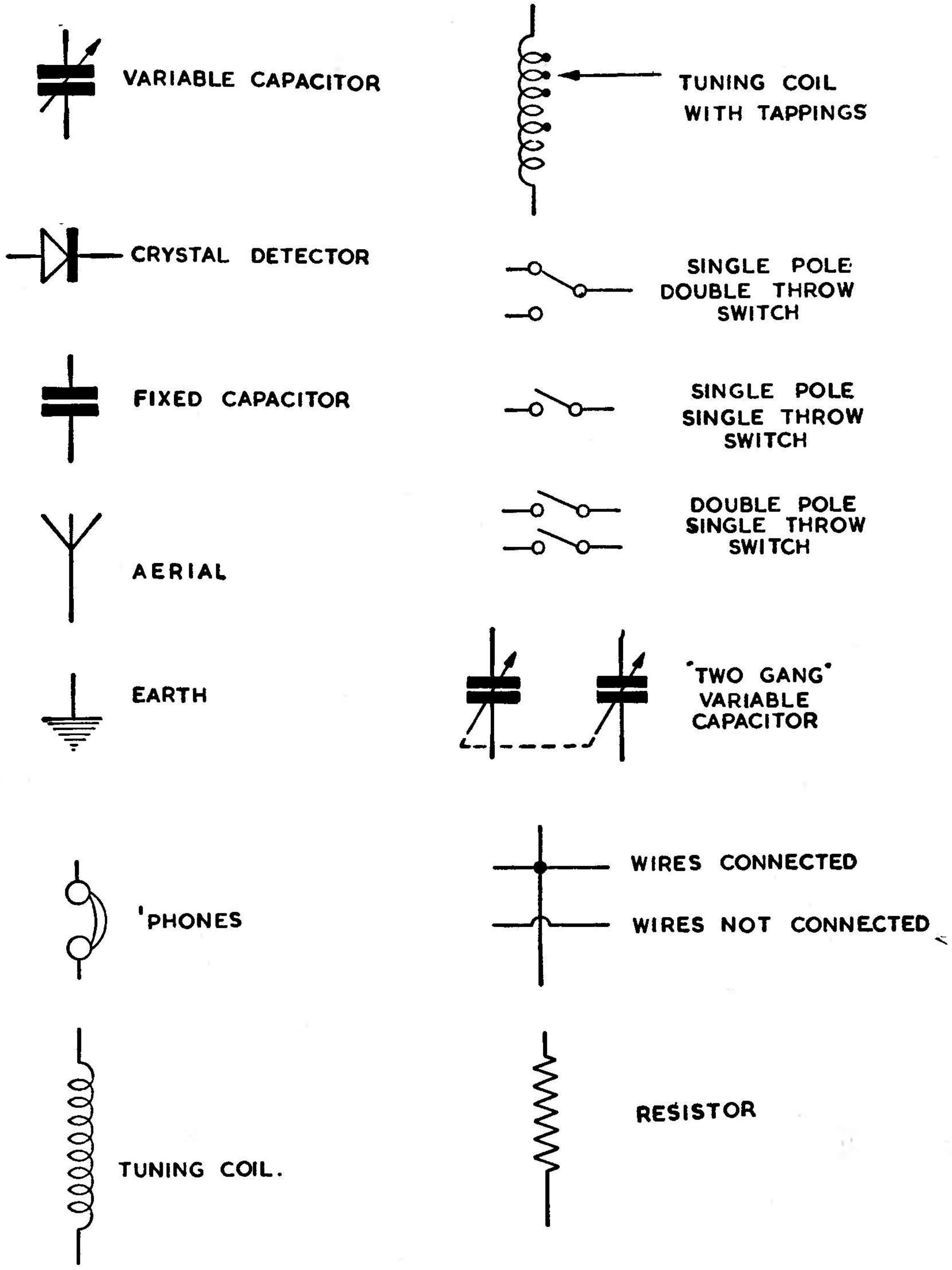
VARIABLE CAPACITOR
TUNING COIL WITH TAPPINGS
CRYSTAL DETECTOR
SINGLE POLE DOUBLE THROW SWITCH
FIXED CAPACITOR
SINGLE POLE SINGLE THROW SWITCH
DOUBLE POLE SINGLE THROW SWITCH
AERIAL
EARTH
'TWO GANG' VARIABLE CAPACITOR
WIRES CONNECTED
WIRES NOT CONNECTED
'PHONES
RESISTOR
TUNING COIL.

SYMBOLS USED IN THE CIRCUITS DESCRIBED.

FIG. 28

The first one is BROWN which represents 1, the first figure of the value, the second BLACK, this indicates the second numeral is nought, and the third colour is RED which tells us that two further ciphers or noughts are added to give the total value. From this you can see, we have 1 plus a nought equalling 10, plus two further noughts which total 1000. Some earlier types of resistor were painted differently, the whole body colour represented the first figure, one tip was coloured to represent the second figure, and a painted dot in the centre gave the number of ciphers to be added.

There are no other special points regarding construction except perhaps that if a long aerial is used, a small 100pF capacitor might be tried in the aerial lead to prevent damping the first tuned circuit.

When setting up the receiver, tune to a station at the high-frequency end of the M.W. band and set the trimmers for loudest volume, no other adjustments are necessary.

CONCLUSION

The twelve receivers described are representative of the best crystal set designs available to-day. Their construction will provide many hours of useful enjoyment, and the results will give lasting pleasure.

A final word about components, if no actual maker is specified, any good class component can be used. Switches, fixed and variable capacitors and crystals, are available from numerous manufacturers all of which are invariably of excellent quality. Where a particular manufacturer's product is called for, the specification should be adhered to. The commercial coils specified are freely available from most supply houses specialising in components for constructors, but in case of difficulty write to Bernards (Publishers) Ltd., The Grampians, Western Gate, London, W.6, who will, on receipt of a S.A.E. be pleased to supply the address of your nearest stockist. Coil formers are not always easy to obtain and in this case advice should be sought from Post Radio Supplies, 395, Queensbridge Road, London, E.8, who will be pleased to help in any way possible.

* *

US005188738A

United States Patent [19]

Kaali et al.

[11] **Patent Number: 5,188,738**

[45] **Date of Patent: * Feb. 23, 1993**

[54] ALTERNATING CURRENT SUPPLIED ELECTRICALLY CONDUCTIVE METHOD AND SYSTEM FOR TREATMENT OF BLOOD AND/OR OTHER BODY FLUIDS AND/OR SYNTHETIC FLUIDS WITH ELECTRIC FORCES

[76] Inventors: **Steven Kaali**, 88 Ashford Ave., Dobbs Ferry, N.Y. 10522; **Peter M. Schwolsky**, 4101 Cathedral Ave., NW., Washington, D.C. 20016

[*] Notice: The portion of the term of this patent subsequent to Aug. 18, 2009 has been disclaimed.

[21] Appl. No.: **615,437**

[22] Filed: **Nov. 16, 1990**

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 562,721, Aug. 6, 1990, abandoned.

[51] **Int. Cl.**^{5} **B01D 35/06**; A61K 41/00

[52] **U.S. Cl.** .... **210/748**; 128/419 R; 128/421; 128/783; 128/784; 204/131; 204/164; 204/186; 204/302; 210/243; 422/22; 422/44; 604/4

[58] **Field of Search** .... 210/243, 748, 764; 128/419 R, 421, 783, 784; 604/4; 422/22, 44; 204/131, 164, 186, 242, 275, 302, 305

[56] References Cited

U.S. PATENT DOCUMENTS

592,735	10/1897	Jones	204/242
672,231	4/1901	Lacomme	204/275
2,490,730	12/1949	Dubilier	204/305
3,692,648	9/1972	Matloff et al.	204/129
3,753,886	8/1973	Myers	204/186
3,878,564	4/1975	Yao et al.	210/648
3,965,008	6/1976	Dawson	422/22
3,994,799	11/1976	Yao et al.	210/321.64
4,473,449	9/1984	Michaels et al.	204/101
4,616,640	10/1986	Kaali et al.	128/130
4,770,167	9/1988	Kaali et al.	128/788
4,932,421	6/1990	Kaali et al.	128/831
5,049,252	9/1991	Murrell	210/243
5,058,065	10/1991	Slovak	128/783
5,133,932	7/1992	Gunn et al.	210/748

FOREIGN PATENT DOCUMENTS

995848	7/1983	U.S.S.R.	210/243

OTHER PUBLICATIONS

Proceedings of the Society for Experimental Biology & Medicine, vol. 1, (1979), pp. 204–209, "Inactivation of Herpes Simples Virus with Methylene Blue, Light and Electricity"—Mitchell R. Swartz et al.

Journal of the Clinical Investigation published by the American Society for Clinical Investigations, Inc., vol. 65, Feb. 1980, pp. 432–438—"Mechanisms of Photodynamic Inactivation of Herpes Simplex Viruses" —Lowell E. Schnipper et al.

Journal of Clinical Microbiology, vol. 17, No. 2, Feb. 1983, pp. 374–376, "Photodynamic Inactivation of Pseudorabier Virus with Methylene Blue Dye, Light and Electricity"—Janine A. Badyisk et al.

Primary Examiner—Robert A. Dawson
Assistant Examiner—Sun Uk Kim
Attorney, Agent, or Firm—Charles W. Helzer

[57] ABSTRACT

A new alternating current process and system for treatment of blood and/or other body fluids and/or synthetic fluids from a donor to a recipient or storage receptacle or in a recycling system using novel electrically conductive treatment vessels for treating blood and/or other body fluids and/or synthetic fluids with electric field forces of appropriate electric field strength to provide electric current flow through the blood or other body fluids at a magnitude that is biologically compatible but is sufficient to render the bacteria, virus, parasites and/or fungus ineffective to infect or affect normally healthy cells while maintaining the biological usefulness of the blood or other fluids. For this purpose low voltage alternating current electric potentials are applied to the treatment vessel which are of the order of from about 0.2 to 12 volts and produce current flow densities in the blood or other fluids of from one microampere per square millimeter of electrode area exposed to the fluid being treated to about two milliamperes per square millimeter.

31 Claims, 6 Drawing Sheets

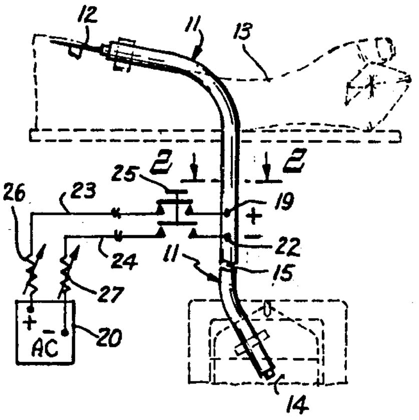

Fig. 1

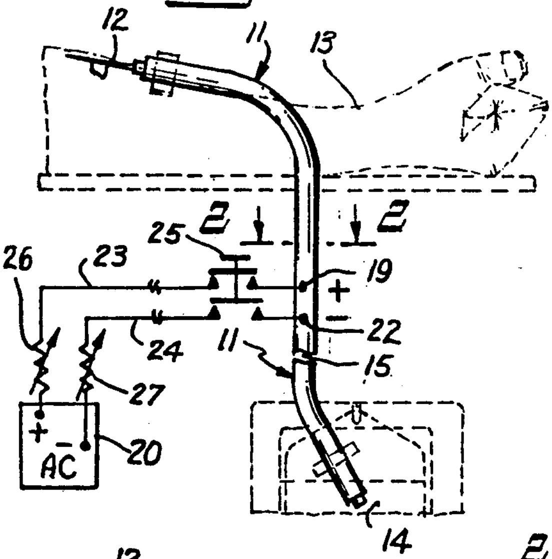
12
11
13
2
2
25
26
23
19
+
22
11
15
24
27
+
AC
20
14

Fig. 2

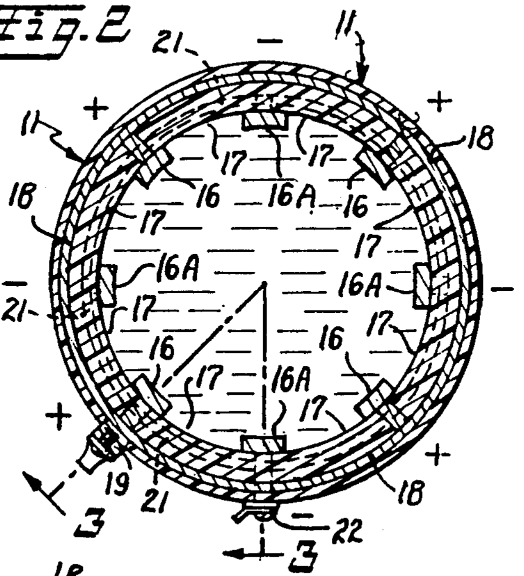
21
11
+
+
11
18
16
16A
16
17
17
18
17
17
16A
16A
21
17
16
16
17
17
16A
+
19
21
18
+
22
3
3

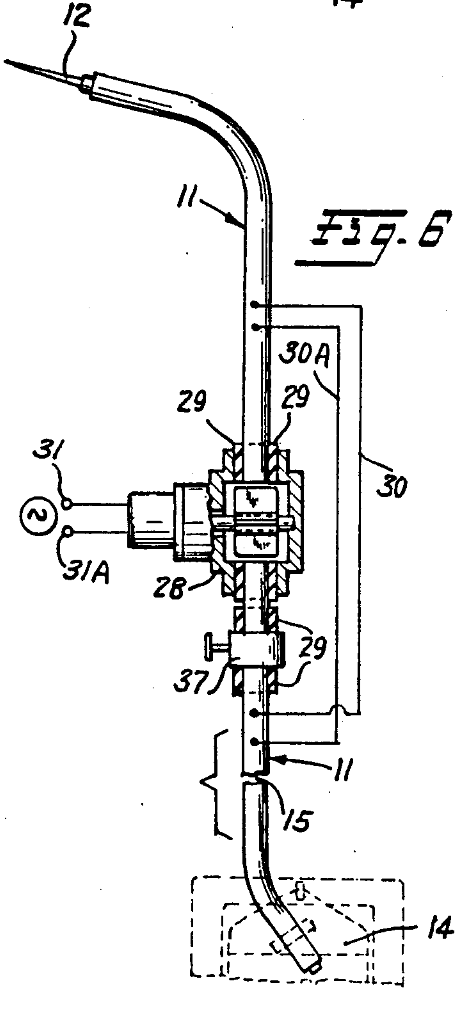
12
11
30A
29
29
31
30
31A
28
29
37
29
11
15
14

Fig. 6

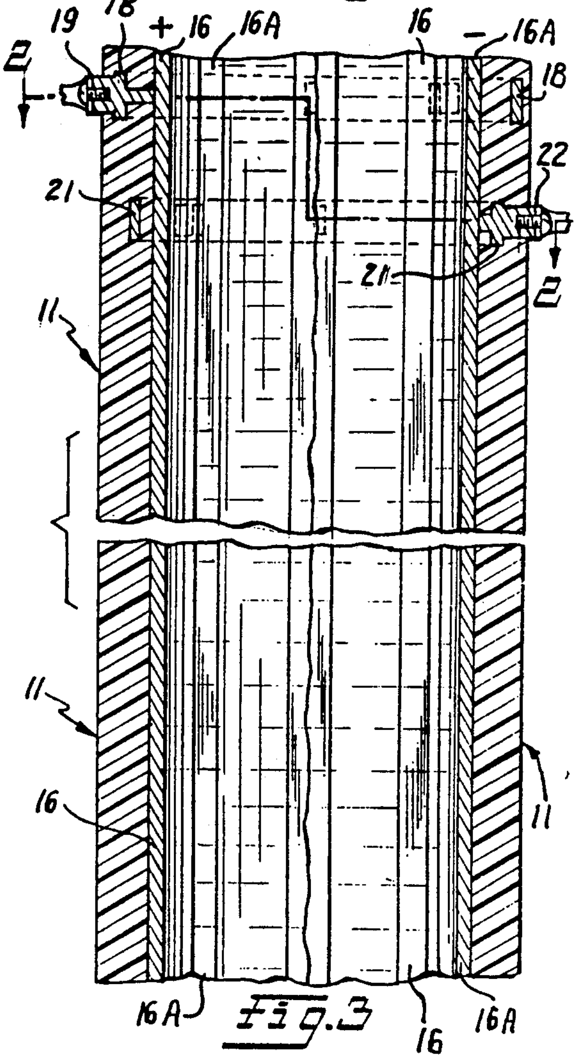
19
18
+
16
16A
16
16A
18
2
21
22
21
2
11
11
16
11
16A
16
16A

Fig. 3

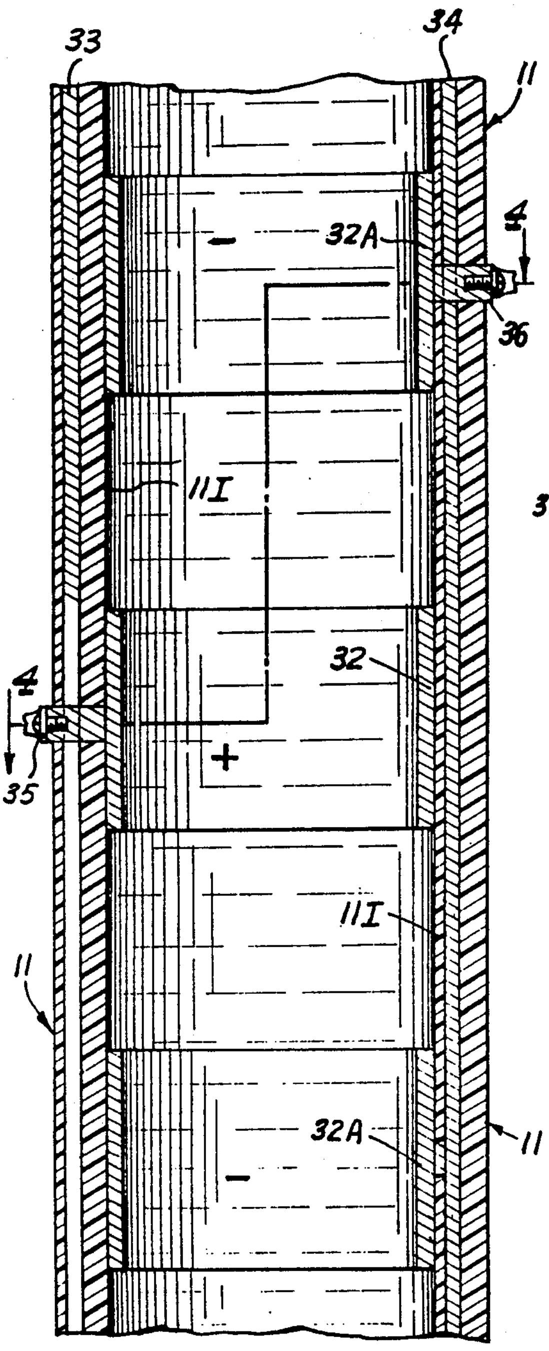
33
34
11
32A
−
4
36
11I
32
+
4
35
11I
11
32A
−
11

Fig. 5

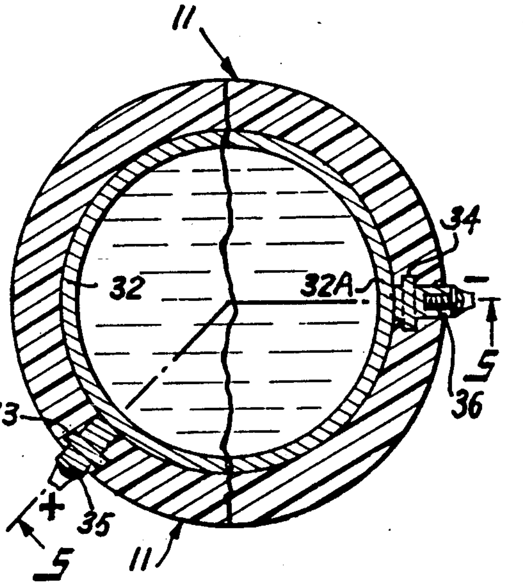
11
32
32A
34
−
5
36
33
+
35
5
11

Fig. 4

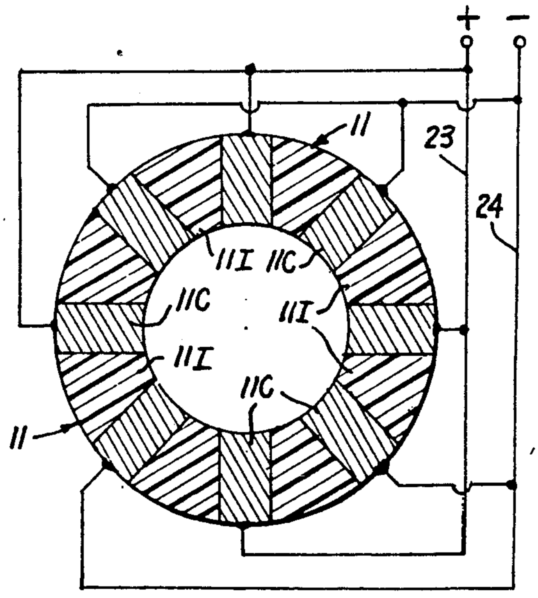
+
−
11
23
24
11I
11C
11C
11I
11I
11C
11

Fig. 7

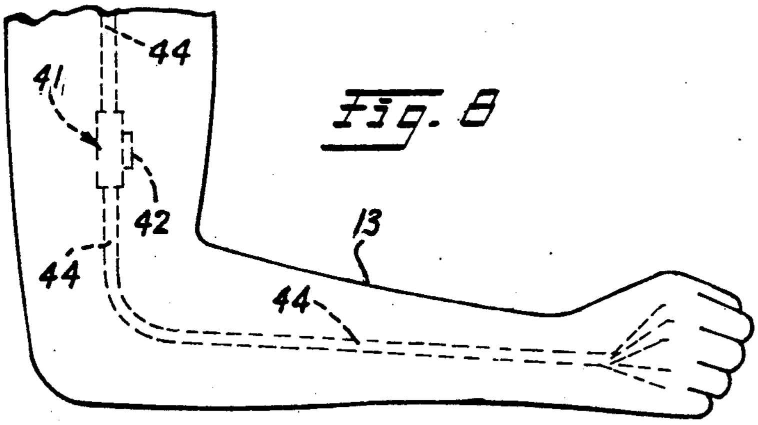
44
41
42
44
13
44

Fig. 8

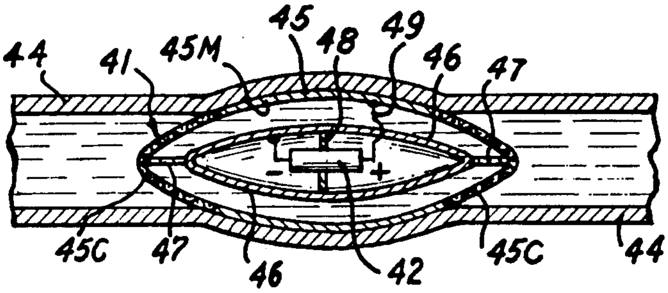
44
41
45M
45
48
49
46
47
45C
47
46
42
45C
44
-
+

Fig. 9

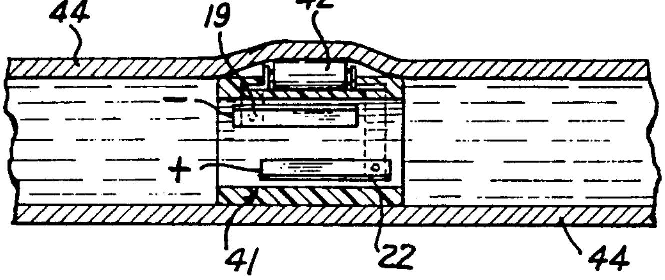
44
19
42
-
+
41
22
44

Fig. 10

Fig. 11

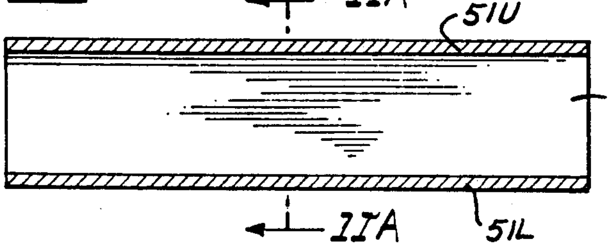
11A
51U
52R
51L
11A

Fig. 11A

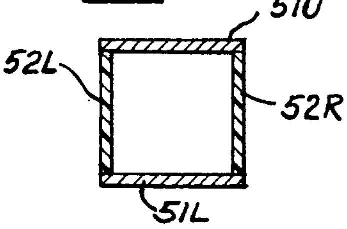
51U
52L
52R
51L

Fig. 12A

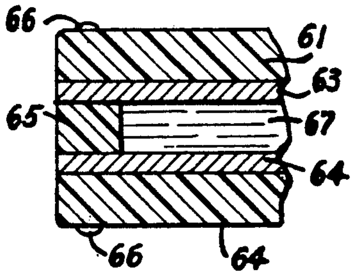
66
61
63
65
67
64
66
64

Fig. 12

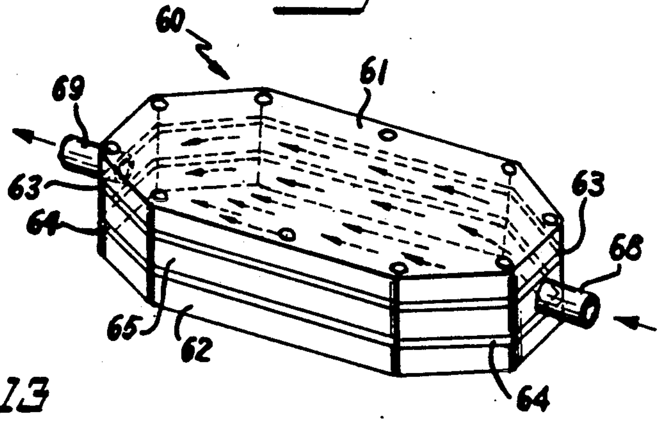
60
61
69
63
64
63
68
65
62
64

Fig. 13

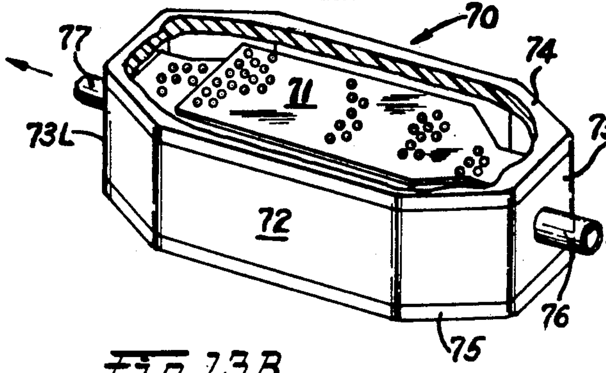
70
77
74
71
73L
73R
72
76
75

Fig. 13A

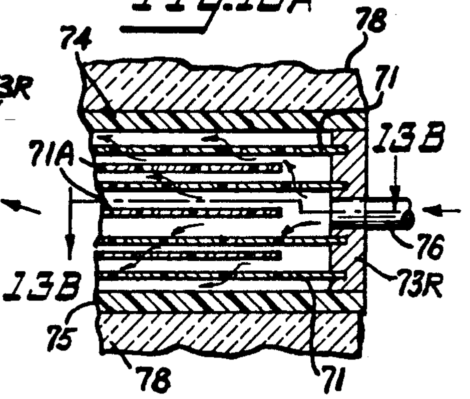
78
74
71
71A
13B
76
13B
73R
75
78
71

Fig. 13B

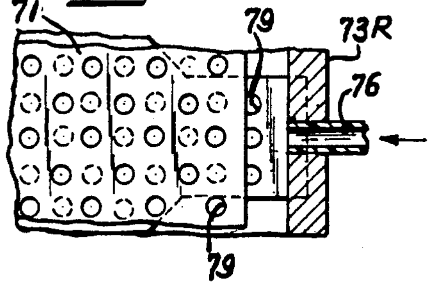
71
79
73R
76
79

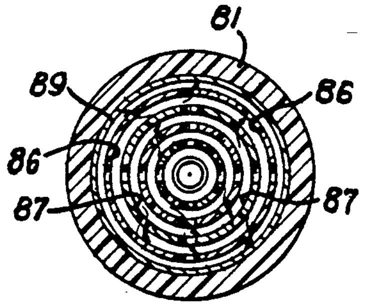
81
89
86
86
87
87

Fig. 14A

Fig. 14

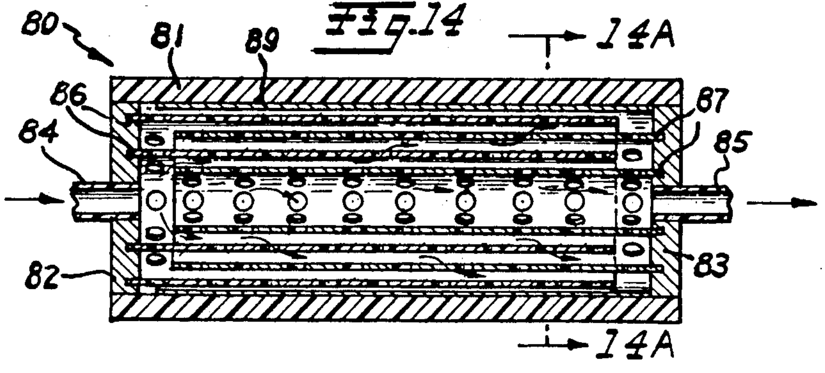
80
81
89
14A
86
87
84
85
82
83
14A

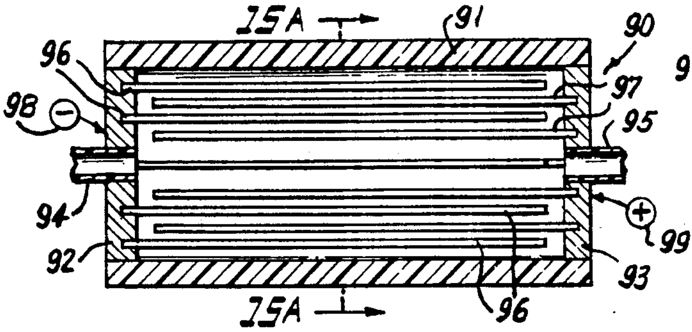
15A
91
90
96
97
98
95
94
99
92
93
96
15A

Fig. 15

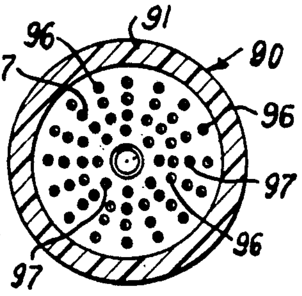
91
96
90
97
96
97
96
97

Fig. 15A

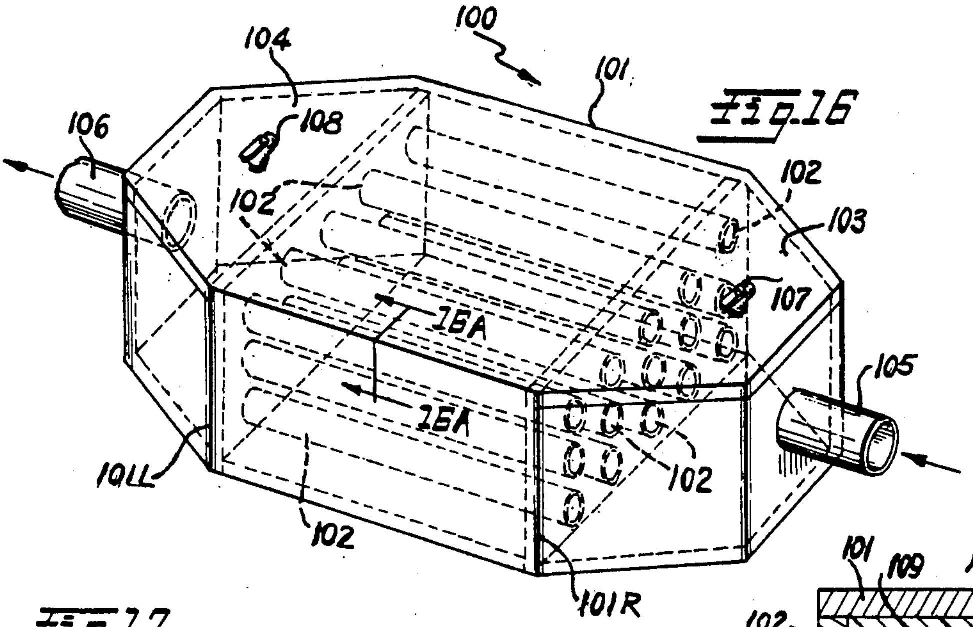
100
104
101
108
106
102
102
103
107
16A
16A
105
101L
102
102
101R

Fig. 16

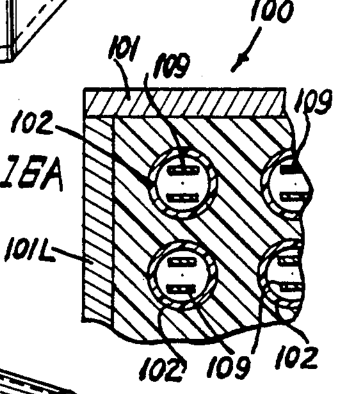
100
101
109
109
102
101L
102
109
102

Fig. 16A

Fig. 17

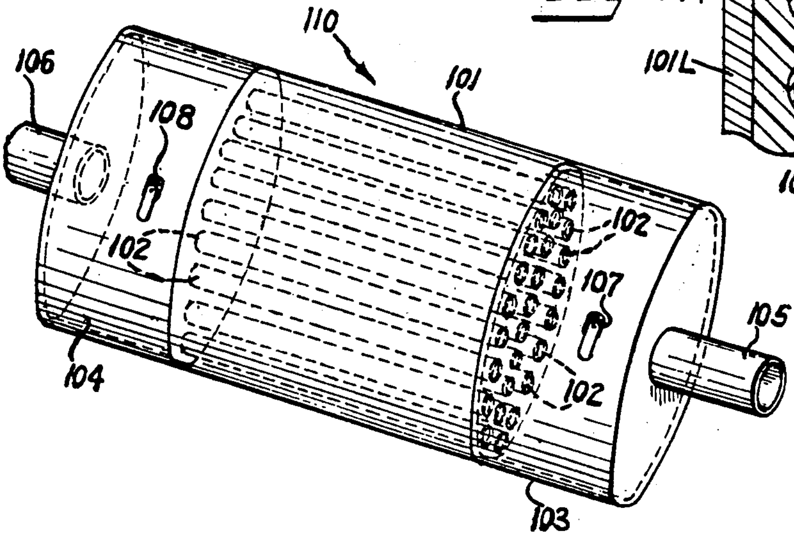
110
106
101
108
102
102
107
105
104
102
103

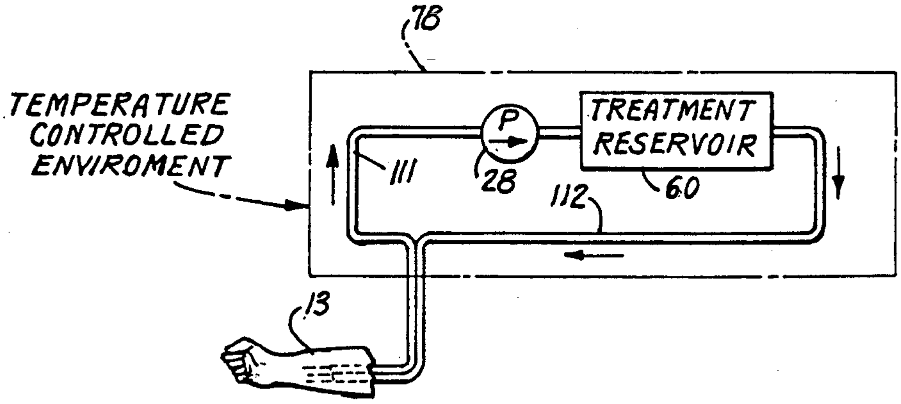
78
TEMPERATURE
CONTROLLED
ENVIROMENT
P
TREATMENT
RESERVOIR
111
28
112
60
13

Fig.18

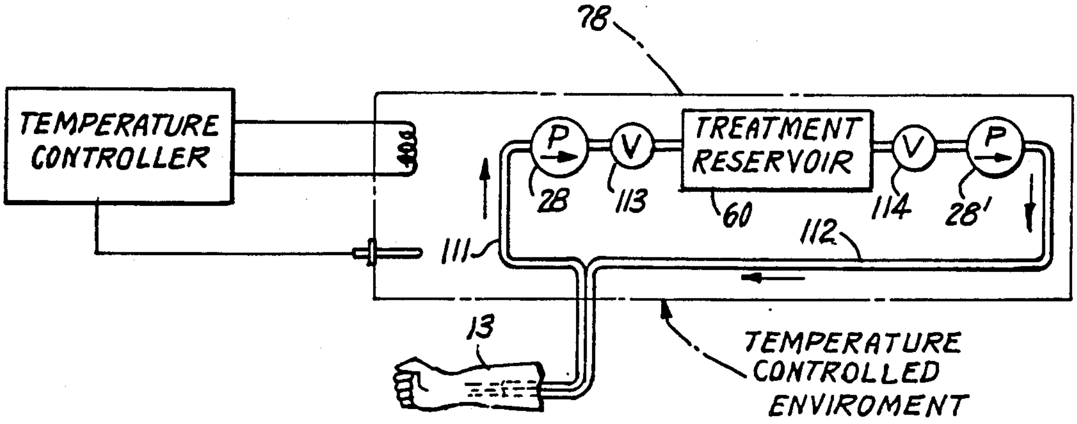
78
TEMPERATURE
CONTROLLER
P
V
TREATMENT
RESERVOIR
V
P
28
113
60
114
28'
111
112
13
TEMPERATURE
CONTROLLED
ENVIROMENT

FIG. 19

ALTERNATING CURRENT SUPPLIED ELECTRICALLY CONDUCTIVE METHOD AND SYSTEM FOR TREATMENT OF BLOOD AND/OR OTHER BODY FLUIDS AND/OR SYNTHETIC FLUIDS WITH ELECTRIC FORCES

FIELD OF INVENTION

This is a continuation-in-part application of prior U.S. patent application Ser. No. 07/562,721 filed Aug. 6, 1990, now abandoned

This invention relates to novel electrically conductive methods and systems employing electrically conductive vessels provided with electrically conductive surfaces for use in subjecting blood and/or other body fluids and/or synthetic fluids such as tissue culture medium to direct treatment by alternating current electric forces.

BACKGROUND PROBLEM

It is now well known in the medical profession and the general public that blood collected in a blood bank from a large number of donors may be contaminated by contaminants such as bacteria, virus, parasites and/or fungus obtained from even a single donor. While screening of donors has done much to alleviate this problem, the screening of donors can and does miss occasional donors whose blood is unfit for use. When this occurs and the unfit blood is mixed with otherwise usable blood, the entire batch must be discarded for transfusion purposes. Because of this problem, the present invention has been devised to attenuate any bacteria, virus (including the AIDS HIV virus) parasites and/or fungus contained in blood contributed by a donor to the point that any such contaminant is rendered ineffective for infecting a normally healthy human cell, but does not make the blood biologically unfit for use in humans. Similar problems exist with respect to treatment of other body fluids, such as amniotic fluids. The treatment method and system is also applicable to mammals other than humans.

In addition to the above, there is a need for methods and systems for the treatment of blood and other body fluids both in in-situ processing wherein the treated blood and/or other body fluids are withdrawn from the body, treated and then returned to the body in a closed loop, recirculating treatment process that is located near but outside the patient's body, or the treatment can be effected through implanted treatment system components.

In co-pending United States application serial No. 07/615,800 entitled "Electrically Conductive Methods and Systems for Treatment of Blood and Other Body Fluids with Electric Forces"-Steven Kaali and Peter M. Schwolsky, inventors, filed concurrently and co-pending with this application, a similar treatment method and system employing direct current excitation potentials is described and claimed. The disclosure of co-pending application Ser. No. 07/615,800 hereby is incorporated into this application in its entirety.

SUMMARY OF INVENTION

The present invention provides new electrically conductive methods and systems using alternating electric current excitation potentials for treating blood and/or other body fluids, such as amniotic fluids, and/or synthetic fluids such as tissue culture medium from a donor to a transfusion recipient or to a storage receptacle, or for recirculating a single donor's or patient's blood or other body fluids. The treatment can be accomplished in a treatment system external of the body or by implant devices for purging contaminants using a novel electrically conductive vessel for direct electric treatment of blood or other body fluids, such as amniotic fluids, with alternating current electric field forces of appropriate electric field strength to attenuate such contaminants to the extent that bacteria, virus, fungus, and/or parasites contained in the blood or other body fluids are rendered ineffective to infect and/or affect normally healthy human cells. The treatment, however, does not render the blood or other body fluids biologically unfit for use in humans or other mammals after the treatment. The new methods and systems according to the invention achieve these ends without requiring time consuming and expensive processing procedures and equipment in addition to those normally required in the handling of blood or other body fluids or synthetic fluids. The invention can be used to achieve the electric field force treatment during the normally occurring transfer processing from a donor to a recipient or to a collection receptacle, or recirculation of a single donor's or patient's blood or other body fluids, such as amniotic fluids.

BRIEF DESCRIPTION OF DRAWINGS

The above and many other objects, features and attendant advantages of this invention will be appreciated more readily as the invention becomes better understood from a reading of the following detailed description, when considered in connection with the accompanying drawings, wherein like parts in each of the several figures are identified by the same reference characters, and wherein:

FIG. **1** is a diagrammatic, fragmentary, elevational view of a new blood transfer system using a novel alternating current electrically conductive treatment vessel in the form of conductive tubing to directly treat blood being transferred to a storage receptacle with electric field forces according to the invention;

FIG. **2** is an enlarged, horizontal cross sectional view of the novel electrically conductive tubing treatment vessel taken across lines **2—2** of FIG. **1**;

FIG. **3** is a longitudinal, vertical sectional view of the novel electrically conductive tubing treatment vessel taken along the staggered section lines **3—3** of FIG. **2**;

FIG. **4** is a view similar to FIG. **2** showing a different construction of the novel electrically conductive tubing treatment vessel;

FIG. **5** is a view similar to FIG. **3**, taken along the staggered section lines **5—5** of FIG. **4**;

FIG. **6** is a diagrammatic, fragmentary, elevational view showing a different modification of a novel blood transfer system using a novel electrically conductive tubing treatment vessel, and which employs a blood pump and a blood flow regulator;

FIG. **7** is an enlarged cross sectional view, similar to FIG. **2** that shows an electrically conductive tubing treatment vessel fabricated from longitudinally extending, integrally molded strips of alternate polarity, conductive polymer interconnected by integrally molded, insulating, longitudinally extending strips made of polymer or other insulating material;

FIG. **8** is a diagrammatic, fragmentary elevational view showing a different form of a blood transfer system according to the invention wherein a small electrically conductive vessel in the form of a short piece of tubing and a miniaturized battery power source are implanted in the arm of a human being to provide a novel electrically conductive blood and other body fluid treatment system which operates in a closed loop, recirculating manner;

FIG. **9** is a partial, diagrammatic sectional view of the upper arm portion of a human being and shows in greater detail the construction of a specially designed miniaturized, electrically conductive treatment vessel with associated miniaturized battery electric power source and DC to AC power converter for use in the implant treatment system shown in FIG. **8**;

FIG. **10** illustrates the details of construction of a somewhat different form of miniaturized electrified treatment tubing for use in an implanted treatment system of the type shown in FIG. **8** and built according to the invention;

FIGS. **11** and **11A** illustrate still a different construction for the electrified treatment tubing for use in practicing the invention wherein the tubing has a square or rectangular cross section with upper and lower conductive sides and intervening right and left sides separating the two conductive sides made from plastic or other suitable electrical insulating material;

FIG. **12** is a perspective top and side view of a novel electrified, closed, octagonally-shaped, flat, box-like treatment vessel having an enlarged cross sectional area relative to the cross sectional diameter of the inlet and outlet tubes supplying the interior of the treatment vessel;

FIG. **12A** is a partial, cross sectional view of the enlarged treatment vessel shown in FIG. **12**;

FIG. **13** is a perspective view of a second form of enlarged cross sectional area treatment vessel having an exterior shape similar to that of FIG. **12**, but wherein the electrically conductive electrodes of the treatment vessel comprise interleaved conductive plates with one set of alternate ones of the plates being electrically insulated from the remaining set, and wherein different polarity electric potentials are applied to the respective sets. If desired, the electrode plates may be formed from an electrically conductive porous material;

FIG. **13A** is a partial, cross sectional view taken through the electrically conductive treatment vessel shown in FIG. **13**;

FIG. **13B** is a sectional view taken through staggered line **13B**—**13B** of FIG. **13A**;

FIG. **14** is a longitudinal sectional view of still a different form of enlarged diameter electrified treatment vessel wherein the vessel is in the form of an elongated cylinder, and the sets of conductive electrodes mounted therein are concentrically arrayed within the interior of the treatment vessel and maintained at different electric potentials;

FIG. **14A** is a cross sectional view of FIG. **14** taken through plane A—A;

FIG. **15** is an enlarged longitudinal sectional view of still another form of an enlarged cross sectional area treatment vessel according to the invention wherein the electrically conductive electrodes of the treatment vessel are comprised by longitudinally extending needle-like electrodes with alternate ones of the needle-like electrodes being provided with opposite polarity electric potentials;

FIG. **15A** is a cross sectional view of the treatment vessel shown in FIG. **15** taken through plane A—A of FIG. **15**;

FIG. **16A** is a partial cross-sectional view taken through **16A**—**16A** of FIG. **16**;

FIG. **16** is a perspective view of still another form of enlarged cross sectional area treatment vessel according to the invention wherein the treatment vessel comprises a relatively large block of insulating material having parallel, longitudinally extending, open ended tubes formed through its length. The tubes are provided with electrically separated, opposed, parallel extending conductive plate electrodes which have opposite polarity electric potentials applied thereto. The ends of the tubes open into and are supplied from, or supply, respective reservoirs formed on the respective ends of the central block of insulating material containing the tubes, with inlet and outlet conduits for body fluids to be treated connected to the free ends of the respective reservoirs;

FIG. **17** is a perspective view of an enlarged cross sectional area treatment vessel similar to FIG. **16** wherein the body of the treatment vessel is cylindrical in nature;

FIG. **18** is a diagrammatic, fragmentary elevational view of a human blood or other body fluid treatment system according to the invention employing one of the larger cross sectional dimension fluid treatment vessels shown in any one of FIGS. **12–16** of the drawings, and which is suitable for use in a continuous flow through recirculating body fluid treatment system; and

FIG. **19** is a diagrammatic, fragmentary elevational view of still another human blood or other body fluid, closed loop, recirculating treatment system according to the invention designed for use with the enlarged diameter fluid treatment vessels illustrated in FIGS. **12–16**, and which employs both inlet and outlet fluid pumps on each side of the treatment vessel. With this arrangement the system can be operated in an intermittent manner to allow batch treatment of the body fluids to fully take place before passage of the body fluids being treated back to the patient.

BEST MODE OF PRACTICING INVENTION

FIG. **1** is a schematic illustration of one form of a novel blood and other body fluid treatment system according to the invention. FIG. **1** shows an electrically conductive blood and/or other body fluid treatment vessel constructed according to the invention which is in the form of intravenous tubing **11** interconnected between a hypodermic needle **12** and a blood storage receptacle **14**. The needle **12** is inserted in an artery or vein of the arm **13** of a blood donor and the tubing **11** leads from the arm **13** to the receptacle **14**. Alternatively, the system could be set up to transfer blood from the storage receptacle **14** to the arm of a recipient or could be designed to recirculate the blood through electrified tubing **11** back to the donor. The electrically conductive tubing **11** may be of any desired length as indicated by the break at **15** so that it can be appropriately set up to lead from a comfortable position for the donor from whose arm **13** the blood is being taken to a proper storage location for the receptacle **14**. The greater the length of the electrified portion of tubing **11**, then the more extended is the exposure of the blood (or other body fluid) to the electric field force effects and low level, biologically compatible current flow through the body fluid being treated thereby assuring adequate electrification treatment of the fluid without impairing the biological usefulness of the blood or other body fluid being treated.

FIG. **2** is a cross sectional view of the electrically conductive tubing **11** taken through plane 2—2 of FIG. **1**. The tubing **11** may be from 1 to about 20 millimeters in inside diameter, although it may be larger or smaller in diameter depending upon the intended application. For example, if the blood transfer system is for the purpose shown in FIG. **6**, then the tubing may have a cross sectional dimension of about 5 millimeters. However, if the intended use is in an implanted blood treatment system, such as shown in FIG. **8**, then the tubing diameter must be designed to result in a flow-through rate corresponding to the natural circulatory blood flow rate of the patient in which the system is implanted, and must be long enough to assure effective electrification treatment at the flow rate selected. The tubing **11** is formed from plastic, rubber, medical grade polymer, or other suitable material which is compatible with human fluids and/or tissue. A plurality of physically separated, electrically conductive surface segments form opposed, parallel electrodes shown at **16** and **16A** on the inside of tubing **11** from electrically conductive materials such as platinum, platinum alloys, silver, silver or platinum covered alloys, or other similar conductive materials such as conductive polymers, or silver or platinum covered polymers which are compatible with human fluids and tissue. The spacing between opposed electrodes **16** and **16A** is of the order of 1 to 19 millimeters and perhaps may be more or less dependent upon the application and the conductivity of the body fluids being treated.

FIG. **3** is a longitudinally extending sectional view along the axis of tubing **11** taken through staggered section lines 3—3 of FIG. **2**. From FIG. **3** of the drawings it will be seen that the electrically conductive surface segments **16** and **16A** all comprise longitudinally extending, zebra-like stripe or strip electrodes which extend longitudinally in parallel with the longitudinal axis of the tubing **11**. In between each longitudinally extending conductive stripe electrode **16** or **16A** is a longitudinally extending electric insulating area **17** which electrically isolates the alternate electrically conductive, zebra-like stripe electrodes **16** and **16A** one from the other.

As best shown in FIG. **3**, a first set of alternate electrically conductive surface stripes **16** are electrically connected in common to a first annular terminal buss **18** which circumferentially surrounds the tubing **11** and is embedded within the sidewalls of the tubing **11** at a suitable point along its length. The design is such that the first annular terminal buss **18** is electrically isolated from the remaining second set of alternate, electrically conductive surface stripe electrodes **16A** and is electrically connected through a conductor terminal **19** to an alternating current source of electric excitation potential. AC source **20** may comprise the output from an AC to AC voltage converter for converting 110 volt AC potential to the desired 0.2 volts to 12 volts for use in the invention. For those treatment systems which are to be implanted as described hereafter, the AC source may comprise a miniaturized DC to AC converter for converting the DC voltage from a miniaturized battery to low voltage (0.2 to 12 volts) AC. As best depicted in FIG. **2**, all of the first set of positive electrically conductive stripes **16** are physically and electrically connected in common to the first annular terminal buss **18** so that all of the conductive stripes **16** are maintained at a constant, alternating current electric excitation potential.

A second annular terminal buss **21**, which circumferentially surrounds the tubing **11**, is embedded within the tubing **11** at a point along its length displaced from the position of the first annular terminal buss **18** and is spaced inwardly towards the inside diameter of the tubing relative to the first annular buss **18**. By this arrangement it is possible to electrically connect the remaining second set of alternate electrically conductive surface stripes **16A** in common to the second annular terminal buss **21** in a manner such that the second annular terminal buss is electrically isolated from the first annular terminal buss **18** as well as the first set of alternate electrically conductive surface stripes **16**. As shown in FIG. **3**, the second annular terminal buss **21** is provided with an outside terminal conductor connection **22** for connecting the annular buss **21** and annular buss **18** across AC source **20** as shown in the system drawing of FIG. **1**. The second set of alternate electrically conductive surface stripes **16A** are all provided with internal connector studs which physically and electrically connect all of the **16A** stripes in common to the second annular terminal buss **21** so that all of these conductive stripes will be maintained at a potential opposite to that from the potential applied to the first set of electrically conductive stripes **16** by annular buss **18**.

As described earlier, the AC source of electric potential **20** may constitute an AC to AC converter for converting 110 volt AC to 0.2 to 12 volt AC or a DC to AC converter for converting 12 volt DC to 0.2 to 12 volt AC. The AC source **20** is connected to the conductor terminals **19** and **22** through electric supply conductors **23** and **24** preferably by a double pole, double throw, on-off control switch **25**. In preferred embodiments of the invention, voltage controlling variable resistors **26** and **27** also are included in the electric supply conductors **23** and **24** in order to control the value of the excitation voltage developed between the alternate sets of conductive surface stripes **16, 16A**.

In operation, the donor whose blood is to be taken, or the recipient who is to be given blood, or is to have his or her blood recycled, is made comfortable on a cot with his or her arm **13** extended and the interconnecting electrically conductive tubing **11** having the hypodermic needle **12** for withdrawal, or supplying, or recycling of blood set up as shown in FIG. **1**. When both the donor/recipient and the system is in readiness, the control switch **25** is closed so that an electric field is built up across the oppositely disposed electrically conductive zebra-like stripes **16, 16A**, etc. Voltages of the order of from 0.2 to 12 volts are applied to the conductive surfaces **16, 16A** For this purpose it is important to note that the hypodermic needle should be electrically isolated via conventional electrically insulating IV tubing from any of the zebra stripe electrodes **16, 16A** so that the donor/recipient does not receive a shock. By this precaution, he or she will not even be aware of the existence of the electric field within the electrically conductive tubing **11**. With the treatment system thus conditioned, the hypodermic needle is inserted into a vein in the donor's/recipient's arm and blood is withdrawn, given, or recycled through the tubing **11**.

As the blood passes through the electric fields produced within the electric conductive tubing **11** it will be subjected to and treated by biologically compatible electric current flow through the blood or other body fluid with a current density of from one microampere per square millimeter (1 $\mu A/mm^2$) of electrode cross sectional area exposed to the fluid to about two milliamperes per square millimeter (2 mA/mm^2) dependent upon field strength of the electric field gradient existing between electrodes **16** and **16A**, the space between the electrodes **16**, **16A** and the conductivity (resistivity) of the body fluid being treated. Recent experiments have proven that exposure to electric fields induced by supply voltages in the range produces electric current flow through blood of the order of 1 to 100 microamperes. Effectiveness is dependent upon length of time of treatment in conjunction with the magnitude of the biologically compatible current flow. For example, treatment of virus in media at 100 microamperes for 3 minutes has been observed to substantially attenuate (render ineffective) the AIDS virus. Similar treatment at other field strength values and lengths of time will have a similar attenuating effect on bacteria, virus, parasites and/or fungus which are present in blood or other body fluids being treated. By controlling the length of time and field strength values that blood is subjected to the electric field forces, undesirable contaminants such as virus, bacteria, fungus and/or parasites will be adequately attenuated to the point that they are rendered ineffective by the sustained action of the electric current flow as the blood travels from the hypodermic needle **12** to the storage bag **14**, or vice versa, or in a recycling mode. The length of travel of the blood through the sustained electric field induced current flow also can be adjusted so that the blood is subjected to the electric field force for time periods of the order of from one to six minutes at least. At the current values noted above this is believed adequate to attenuate (render ineffective) bacteria, virus (including the AIDS virus), parasites and/or fungus entrained in blood or other body fluids, but does not render the fluids unfit for human use or impair their biological usefulness.

The species of the invention shown in FIGS. **2** and **3** is advantageous since it is possible to fabricate the treatment tubing by preforming the conductive segments **16** and **16A** on the tubing walls while it is in a flat planar condition, and then rolling the walls into tubular form using a suitable mandrel. The adjoining longitudinal edges of the planar member after rolling are thereafter heat sealed along a longitudinally extending seam located within one of the electrically insulating sections **17**. Particular attention must be paid to the juncture of the ends of the annular terminal busses **18** and **21** during the rolling and heat sealing steps to assure that good electrical interconnection and continuity at these junctures of the annular terminal busses is provided in the completed treatment tubing. The conductive electrode segments **16**, **16A** may be electro-deposited, chemically formed, separately formed conductive polymer surfaces, or conductive foil or wires adhesively secured to the side walls of the tubing **11** in advance of the rolling and sealing using techniques well known in the printed circuit and integrated circuit manufacturing technologies.

FIG. **6** is a diagrammatic, fragmentary, elevational view of a modified blood treatment system using the novel electrically conductive treatment tubing in accordance with the invention. In the FIG. **6** embodiment of the invention, a blood pump **28** of conventional, commercially available construction is inserted in the tubing **11** at some point along its length. The blood pump **28** is electrically isolated from the zebra striped conductive surfaces **16**, **16A** by suitable insulators **29** formed on the blood input-output connections of pump **28**. Provision for electrically bypassing the blood pump **28** (if need be) is made through the shunt conductors **30**, **30A** which maintain electrical continuity of the alternating current excitation potential applied to the conductive stripes **16**, **16A** on each side of pump **28**. For convenience, the alternating current excitation source **20** and its connection to the electrically conductive tubing **11** has not been shown in FIG. **6** but would have to be provided. A separate source of excitation current for running the blood pump **28** is provided from a conventional 110 volt alternating current source through the input terminals **31**, **31A**.

In systems employing a blood pump, it may be desirable in some applications to provide a blood flow regulating valve **37** inserted in the system at the output of blood pump **28** and within the by-pass loop **30**, **30A** for the conductive stripes **16**, **16A**. By thus controlling blood flow, the electrified transfer system safely can be employed in a closed loop recycling system for withdrawing blood from a patient,. electrically treating the blood as described above and then returning the electrically treated blood to the patient. This procedure is referred to herein as recycling. The system of FIG. **6** also can be used in those situations where the blood flow of a donor's blood is not sufficient to assure supply of an adequate amount of blood to or from the collection receptacle **14** or other recipient. It may also be desirable to have a blood flow regulating valve such as **37** in non-pump systems.

FIGS. **4** and **5** of the drawings show another embodiment of the invention wherein the electrically conductive treatment tubing **11** includes electrically conductive electrode segments **32** and **32A** which are in the form of zebra stripes that extend radially around the inside diameter of tubing **11** in spaced-apart, alternating polarity, conductive annular bands **32** and **32A** separated by insulating surface bands **11I** which serve to electrically isolate the respective first set of conductive zebra stripes **32** from the second set of conductive zebra stripes **32A**. The first set of alternate ones of the electrically conductive annular stripes **32** are electrically connected in common to a first longitudinally extending terminal buss bar **33** that is embedded within tubing **11** in parallel with the longitudinal axis of the tubing and electrically isolated from the remaining second set of alternate electrically conductive annular stripes **32A**. The first longitudinally extending terminal buss bar **33** is designed for connection to one output terminal of a source, such as **20**, of alternating current electric excitation potential through a supply conductor connection **35** on the exterior surface of the tubing **11**.

A second longitudinally extending terminal buss bar **34** is embedded within the body of tubing **11** and is electrically connected to the remaining second set of alternate electrically conductive annular stripes **32A**. The second longitudinally extending terminal buss bar **34** is electrically isolated from the first longitudinally extending terminal buss **33** and the first set of altefnate electrically annular stripes **32**. Terminal buss bar **33** is designed for connection to a second output terminal for the alternating current source of electric excitation potential. For this purpose an input supply conductor connection **36** is directly connected through the exterior surface of tubing **11** and to the second longitudinally treatment extending terminal buss bar **34**.

In operation, the embodiment of the invention shown in FIGS. **4** and **5** is physically arranged in a blood treatment system in the manner illustrated in FIG. **1** of the drawings with the positive polarity and negative polarity zebra annular stripes being connected to the respective output terminals of AC source **20** via control switch **25**. If required, a blood pump such as **28** and blood flow regulating valve **37** shown in FIG. **6** can be included in the blood transfer system employing electrified tubing as shown in FIGS. **4** and **5**.

Similar to the system shown in FIG. **1**, a blood transfer system employing the embodiment of the invention shown in FIGS. **4** and **5** would be electrically excited in advance of injection of the hypodermic needle **12** into the arm of a blood donor so that all blood passing through the tubing **11** will be subjected to electric forces produced between the alternate polarity annularly formed conductive bands **32** and **32A**. Experience with the invention will establish what length is required for the electrification field. However, for initial installations the length of the electrified field as related to the flow of blood through electrified tubing **11** should correspond to at least the 1-6 minute treatment time mentioned earlier. This is achieved by using an extended array of the alternate annular zebra bands **32** and **32A** of adequate length to assure thorough subjection of blood to electric current flow produced between the alternating polarity zebra stripes **32** and **32A**. The electric field force intensity applied to the blood by means of the electrified tubing is anticipated to be of the order of from 0.2 to 12 volts similar to the embodiment of the invention shown in FIGS. **1-3**.

In place of supplying continuous alternating current excitation to the conductive stripes **16, 16A** of FIGS. **2** and **3** or **32, 32A** of FIGS. **4** and **5**, it also is possible to excite these electrically conductive segments of tubing **11** with pulsed waveform direct current excitation potentials. For use in this manner, the pulse rate of the pulsed waveform excitation potentials must be sufficiently high to maintain continuous current flow through blood being treated. In addition, it may be desirable to couple a bank of storage capacitors in parallel across respective pairs of opposite polarity electrically conductive segments **16, 16A** and **32, 32A** where operation in a pulsed DC mode is desired.

FIG. **7** of the drawings is a cross sectional view of another embodiment of the invention which is substantially different from those previously described. In FIG. **7**, the material used for fabrication of the tubing **11** is one of the new space-age polymer materials which can be either highly electrically conductive, insulating, or semiconducting and may have values of conductivity ranging from essentially fully conductive to insulating. In the embodiment of the invention of FIG. **7**, the conductive surface areas on the inside diameter of the tubing **11** are actually formed into segments, such as **11C**, of the cross sectional area of the tubing **11** fabricated from the highly conductive polymer material. The intervening segments of the tubing **11I** which separate the conductive segments **11C** are integrally formed from the highly insulating polymer material. Suitable positive polarity and negative polarity potentials are applied to the exterior surface areas of alternate ones of the sets of conductive polymer segments **11C** from a source of electric potential via the conductors **23** and **24** as illustrated schematically in FIG. **7**.

It will be appreciated that the embodiment of the invention shown in FIG. **7** is much simpler and hence less expensive to make in that it requires fewer processing steps than the embodiments of the invention shown in FIGS. **1-6**. In other respects, the embodiment of the invention shown in FIG. **7** would be used in a blood transfer system similar to that shown in FIG. **1** or **6** with or without a blood pump **28** and blood flow regulating valve **37** to effect transfer of blood from a donor to a receptacle or recipient in the event of a transfusion or recycling. During the blood transfer process, again it would be necessary to provide alternating current excitation potentials across the spaced-apart, alternate sets of electrically conductive polymer segments **11C** prior to passing blood through the tubing **11**. This will assure that all of the blood being transferred is subjected to the electric field forces produced between the alternate conductive surfaces **11C**. As a variation of the FIG. **7** embodiment, which visualizes that the segments **11C** and **11I** all extend longitudinally and parallel to the longitudinal axis of tubing **11**, it would be possible, but more elaborate to design, to employ alternate radially surrounding annular conductive segments **11C** and interlacing insulating segments **11I** similar to FIG. **5**, but such fabrication would require somewhat more complex terminal buss bar electric supply connections **23** and **24** than those shown in FIG. **7**.

FIG. **8** is a fragmentary, diagrammatic, elevational view showing a form of blood treatment system according to the invention wherein a small electrically conductive vessel **41** in the form of a short piece of electrified tubing and a combined miniaturized DC to AC converter and battery power source **42** are implanted in the arm of a human being. The electrified tubing **41** may be in the form of any of the prior disclosed electrified tubing structures described with relation to FIGS. **1-7**, but which are fabricated in miniaturized form so that the tubing **41** and power package **42** can be inserted in a section of or surrounding a vein **44** of the arm **13** of a patient whose blood is being treated. The implantation is such that the blood through the patient's vein **44** naturally is pumped through the short piece of electrified tubing **41** while circulating blood to the hand of the patient to thereby form a closed loop, recirculating, implanted treatment system that comprises an integral part of the circulatory system of the patient being treated. Because the parameters of such an implanted system are necessarily small, a single passage through the implanted electrified tube **14** may accomplish relatively little attenuation of contaminants in the blood. Therefore, it is the repeated passage of small portions of the patient's blood continuously twenty-four hours a day and for as many days as are needed which will gradually attenuate the contaminants to the point where they are rendered ineffective as described earlier.

FIG. **9** is a partial, fragmentary, sectional view of the upper arm portion **13** of a vein or artery of a patient in which a treatment system according to the invention has been implanted, and shows in greater detail the construction of a specialized, miniaturized, electrically conductive treatment vessel with associated miniaturized battery electric power source and DC to AC converter for use in an implanted treatment system as shown in FIG. **8**. In FIG. **9**, the electrified vessel **41** is in the form of an outer housing **45** that is in the shape of a football which is implanted within the interior walls **44** of an artery or a vein. The outer housing **45** is comprised by a central, cylindrically-shaped portion **45M** of solid conductor such as platinum which is biocompatible with human blood and tissue and has integrally formed, conically-shaped porous ends **45C** which are attached to and form an electrically conductive screen grid (at the same potential) as the mid portion **45M**. The conical end portions **45C** both are perforated and may be in the nature of a screen or mesh wire and of the same material composition as the mid portion **45M**. Disposed within the outer housing **45** is a inner housing **46** which is tear-drop shaped and secured within the central portion **45M** of the outer housing by suitable insulating support spider legs **47**. The inner housing **46** likewise is formed from platinum or other suitable biocompatible conductive material and has supported within its interior a miniaturized AC source comprising a miniaturized battery and AC to DC converter **42** secured to the conductive walls of inner housing **46** by conductive support legs **48**. The support legs **48** serve as terminal connectors from one terminal of AC power converter **42** to the inner housing **46** so that it is maintained at one polarity excitation potential. The remaining opposite polarity terminal of miniaturized AC source **42** is connected through an insulated conductor **49** to the central portion **45M** of outer housing **45** whereby the entire outer housing including the meshed conical end portions **45C** are maintained at an opposite polarity potential from the inner housing **46**.

Prior to implantation in a patient, the electrified vessel shown in FIG. **9** is activated by connection to AC source **42** so that an electric field gradient is produced across the space between the inner and outer housings **45** and **46**. Following implantation of the activated, electrified treatment vessel **41**, its presence in a vein or artery will cause all blood flowing through the vein or artery to pass between the side walls of the inner and outer housings **45** and **46** so as to be subjected to the electric field force gradient existing in these spaces. The presence of the electric field forces will induce a current flow through the blood passing between the interior and outer housings as explained above which will result in attenuating bacteria, virus, parasites and/or fungus which are present in the blood as contaminants. Here again, because of the relatively small portion of the total blood flowing in a patient that will be treated by the device within a given time period, it is the repeated, recycling process treatment of the blood over a prolonged period of time that will result in attenuation of the contaminants in the blood to the point where such contaminants are rendered ineffective as described earlier.

In order to further assure adequate treatment of the blood of a patient receiving the implant device, it is recommended that the blood be treated in an external treatment processing facility such as described earlier in FIGS. **1** and **6** or to be described hereinafter with relation to FIGS. **18** and **19** in which the total capacity of the treatment system is greater whereby substantial attenuation effect can be achieved in a comparatively shorter time period yet to be determined, and then the in vitro implant treatment system such as shown in FIGS. **8**, **9** and **10** can be used to maintain the attenuated condition and to prevent any subsequent build up of contaminants after the initial treatment, if determined to be desirable.

FIG. **10** is a fragmentary, diagrammatic view of a partial vein or artery **44** showing in greater detail the cylindrical or tubular electrified treatment vessel **41** originally described with relation to FIG. **8**. This implant treatment vessel **41** is miniaturized so that it is in effect an open-ended cylinder in shape and has a diameter comparable to that of a large vein or artery and so that it can be grafted or implanted into the vein or artery as illustrated in FIG. **10**. The tubular treatment vessel **41** may be designed pursuant to FIGS. **2** and **3** of the drawings, for example. For this application, the battery source of power and interconnected DC to AC converter **42** are annular in shape and are slipped over the tubular treatment vessel **41** in the manner shown. In FIG. **10** a longitudinal sectional view of the hollow annular-shaped treatment vessel **41** and AC power source **42** is illustrated. At the point where the battery driven AC power source **42** fits over the tubular treatment vessel **41**, the respective terminals of the AC power source **42** are exposed to engage the corresponding positive and negative supply terminals **19** and **22** of the tube **41** so that the resulting structure has a minimum exterior profile to facilitate implantation. From a comparison of FIG. **10** to FIG. **9** of the drawings, it will be appreciated that the FIG. **9** treatment vessel introduces some flow restriction in the vein or artery in which it is implanted and for this reason the construction shown in FIG. **10** is preferred.

FIGS. **11** and **11A** of the drawings illustrate a construction for the electrified treatment vessel **51** wherein the treatment vessel is in the form of square or rectangular cross sectionally-shaped open-ended tubing. The treatment tubing **51** provided with a square or rectangular shape so that provision of opposed, parallel conductive electrode surfaces **51U** and **51L** is greatly simplified as best seen in FIG. **11A** of the drawings, which is a cross sectional view taken through plane **11A—11A** of FIG. **11**. By fabricating the upper and lower surfaces of the tubing **11** from electrically conductive material such as platinum, etc., and separating the upper and lower surfaces **51U** and **51L** by electrically insulating side walls **52R** and **52L**, provision of the electrically isolated, opposed, parallel electrode surfaces is simplified and the resulting treatment vessel introduces minimum restriction to flow of blood. By connecting the upper surface **51U** to one terminal of the AC power source **42** and connecting the lower surface **51L** to the opposite terminal, AC electrification of the interior area of the tubing wherein the fluids to be treated flow is readily achieved with a greatly simplified electrode structure. Variations of this structural feature wherein the side insulating surfaces **52R** and **52L** are curved with their concave surfaces facing each other and the cross sectional area of the upper and lower conductive surfaces **51U** and **51L** tailored to provide a desired current density, tubular treatment vessels such as shown in FIGS. **11** and **11A** could be readily provided for use in implantation devices such as that illustrated in FIG. **8**.

FIG. **12** is a perspective view of a novel, electrified, closed, octagonally-shaped, flat, box-like treatment vessel **60** according to the invention which provides an enlarged cross-sectional area relative to the cross sectional diameter of the inlet and outlet tubing supplying the interior of the treatment vessel whereby increased through-put of a fluid being treated can be achieved in a given time period. The treatment vessel **60** shown in FIG. **12** is comprised essentially of upper and lower, octagonally-shaped, flat insulating plates **61** and **62**, respectively, of an insulating material which is compatible with human blood and/or other body fluids. Disposed immediately below and above the upper and lower plates **61** and **62** are octagonally-shaped, conductive electrode members **63** and **64**, respectively, which are separated and electrically isolated one from the other by a surrounding electric insulating gasket member **65**. The entire structure is sandwiched together and held in assembled relation by threaded thru-pins **66** as best seen in FIG. **12A** of the drawings. The insulating gasket **65** which may be of teflon defines an open space **67** between the two conductive electrode members **63** and **64** into which the blood or other body fluid to be treated is introduced via inlet and outlet conduits **68** and **69**. Alternating current electric potentials are applied across the respective conductive plates **63** and **64** to produce an electric field force across the intermediate space **67** through which the fluids being treated flow between electrode plates **63** and **64**. By thus structuring the treatment vessel, increased treatment surface area is provided to the blood or other body fluid flowing through the space **67** whereby in a given time period an increased quantity of fluids can be treated.

FIG. **13** is a perspective view of another form of enlarged cross sectional area treatment vessel **70** having an exterior shape similar to that of the treatment vessel shown in FIG. **12**. The electrified treatment vessel shown in FIG. **13** differs from that in FIG. **12**, however, in the construction of its electrically conductive electrodes which comprise a plurality of interleaved, conductive, flat, electrode plates **71** and **71A**. The electrode plates **71** are secured in and project inwardly from a right hand (RH) conductive end plate **73R** as shown in FIG. **13A**. The alternate set of flat electrode plates **71A** are secured to and project inwardly from a corresponding conductive end plate **73L** on the left hand end of the treatment vessel **70**. The conductive end plates **73R** and **73L** and coacting insulating side plates **72** which insulate the conducting end plates from one another, form an octagonally-shaped box frame which is closed by upper and lower insulating top and bottom insulating plates **74** and **75**. The conductive end plates **73R** and **73L** have a central opening formed therein into which inlet and outlet tubes **76** and **77** are secured as best seen in FIG. **13** for providing inlet and outlet flow through connection to the treatment vessel **70**.

The alternate sets of flat electrode plates **71** and **71A** extend parallel to one another and are provided with alternating current electric potentials supplied across the respective sets of interleaved electrode plates via the respective conductive end members **73R** and **73L**. If desired, the respective flat conductive electrode plates **71** and **71A** may be fabricated from a perforated material as shown in FIG. **13B** of the drawings. Also, it may be desirable that some form of thermal insulation, or a thermally controlled chamber be provided around the exterior of the treatment vessel **70** as indicated by the thermal insulation **78** shown in FIG. **13A**.

In operation, electrified treatment vessel **70** shown in FIGS. **13, 13A** and **13B** functions in essentially the same manner as was described earlier with respect to FIGS. **1-7** to effect attenuation of contaminants such as bacteria, virus and fungus contained in blood and/or other body fluids being treated in the flow through treatment vessel of FIG. **13**.

FIG. **14** is a longitudinal sectional view of still another form of enlarged cross sectional area, electrified treatment vessel **80**. The treatment vessel **80** shown in FIG. **14** is in the form of an open-ended, elongated cylinder **81** whose cylindrical walls are fabricated from an insulating material which is biocompatible with human blood and/or other body fluids and whose open ends are closed by circular-shaped conductive end pieces **82** and **83**. Inlet and outlet tubular openings **84** and **85** are provided to the interior of cylindrical housing **81** through centrally formed apertures in the circular end plates **82** and **83**. Within the interior of the cylindrical, insulating housing **81** at least two, separate, concentric, perforated, cylindrically-shaped electrode members **86** and **87** are provided which extend longitudinally through the interior of the outer cylindrical housing **81**. The first set of concentric, perforated, electrically conductive electrodes **86** is embedded in and supported by the conductive end plate **82** which serves as an electrical terminal for applying electric potentials to all of the concentric electrode member **86**. Similarly, the concentric, perforated, conductive electrode member **87** is physically supported by and electrically connected to the conductive end plate **83** for the supply of alternating current potentials thereacross. Additionally, if desired, one or more additional perforated concentric electrode members similar to **86** may be spaced apart from the inner concentric electrode member **86** outwardly along the diameter of the circular end member **82** with additional perforated concentric electrode members **87** being sandwiched between the two electrode members **86** and spaced apart therefrom so as to provide an electric field force between all the spaced apart, separated electrically conductive electrode members **86** and **87**. Additionally, if desired, a conductive surface **89** may be formed around the interior walls of the outer, insulating cylindrical housing member **81** and electrically connected to the conductive end plate **82** or **83**. This will assure that the entire interior of the treatment **80** vessel cross sectional area is crossed by the electric field force and all blood or other body fluid passing the cylindrical housing member **81** is subjected to biologically compatible low electric current flow as a consequence of the alternating current electric fields produced between the different concentric electrode members including the coated surface **89** within the interior insulating housing member **81**.

In operation, the embodiment of the invention shown in FIG. **14** and **14A** operates in substantially the same manner as described with relation to earlier embodiments of the invention to assure production of biologically compatible electric current flow through the blood or other body fluid being treated in the treatment vessel **80**.

FIG. **15** is a longitudinal sectional view of still another embodiment of an enlarged cross-sectional area treatment vessel **90**. The treatment vessel **90** again comprises an outer, hollow, open-ended cylindrically-shaped, insulating body member **91** whose open ends are closed by electrically conductive, circular end plates **92** and **93**, respectively. Inlet and outlet tubular openings **94** and **95** are provided through the central axial opening in the conductive end plates **92** and **93** for passage of blood and/or other body fluids being treated into the interior of the treatment vessel **90**. The conductive end plates **92** and **93** have respective sets of opposite polarity potential needle-like electrodes **96** and **97**, respectively, projecting therefrom inwardly into the interior of the treatment vessel **90**. Alternating current electric potentials are applied to the respective conductive end plates **92** and **93** through respective AC supply terminals indicated at **98** and **99**. If desired, and in order to assure complete saturation of the entire volumetric area within treatment vessel **90** with electric fields, a conductive coating similar to that shown at **89** in FIG. **14** can be provided to the inner surface of the hollow, cylindrically-shaped outer body member **91** of treatment vessel **90**.

FIG. **15A** is a cross sectional view taken through plane A–A of FIG. **15** and shows how the array of needle-like electrodes appear within the interior of the treatment vessel **90**. In operation, the treatment vessel **90** will function in substantially the same manner as has been described previously with relation to earlier described embodiments of the invention.

FIG. **16** is a perspective view of still another form of enlarged cross sectional area treatment vessel **100** according to the invention and FIG. **16A** is a partial cross sectional view taken through plane **16A—16A** of FIG. **16**. The treatment vessel **100** comprises a relatively large rectangular-shaped block **101** of electrical insulating material which is biocompatible with blood and/or other human body fluids. The insulating block **101** has a plurality of parallel, longitudinally extending, open-ended, tubular-shaped openings **102** formed therein through the entire length of the block. The tubes **102** are provided with electrically isolated, opposed, parallel extending conductive plate electrodes **109** as best shown in FIG. **16A**, which have alternating current electric potentials applied thereacross. One set of these electrodes, formed for example by the lower electrode **109** in each tube, extend out to and engage a conductive surface coating formed on one end of the insulating block, for example **101R**, and the remaining upper electrodes **109** form a second set which extend out of the left hand end of the tubes and contact a conductive coating formed on the remaining end **101L** of block **101**. Alternating current electric potentials are connected across the respective conductive surfaces **101R** and **101L** so that a potential difference exists between the sets of electrodes **109** within each longitudinally extending tube in block **101**. The ends of the tubes **102** open into and are supplied from, or supply, respective header reservoirs **103** and **104** formed on the respective opposite ends of the block of insulating material **101**. Each of the reservoirs **103** and **104** has a centrally formed opening for receiving either an inlet tube **105** applied to header **103** or an outlet tube **106** secured to header **104** for supply of blood or other body fluids to be treated to and from the treatment vessel **100**. If desired, a blood pump or other fluid pump can be inserted between the supply tube **105** and header **103**, or between outlet tube **106** and the or outlet from the header reservoir **104**, or both. Alternatively, both inlet and outlet pumps can be used. In operation, the electrified treatment vessel **100** shown in FIG. **16** functions in the same manner as those species of treatment vessels described previously.

For some treatment applications, it may be desirable to provide exhaust vents such as shown at **107** and **108** in FIG. **16** to the inlet reservoir **103** and/or the outlet reservoir **104** with the vents that can be selectively operated by valves that can be automatically or manually controlled for venting off gases that might be trapped in the tops of reservoirs and which otherwise might interfere with the proper operation of the electrified treatment vessel. In a similar manner, suitable venting apparatus may be provided to other of the large cross sectional area electrified treatment vessels described previously.

FIG. **17** is a perspective view of still another enlarged cross-sectional area treatment vessel **110** which is similar in all respects to the treatment vessel shown in FIG. **16** with the exception that the body or block of insulating material **101** through which the elongate tubular openings are made, is cylindrically shaped as illustrated in FIG. **17**. In other respects, the embodiment of the invention shown in FIG. **17** would be identical to FIG. **16** in the fabrication and operation of its component parts including the reservoir headers **103** and **104** and would operate in a similar manner.

FIG. **18** is a diagrammatic, sketch of a human blood or other body fluid treatment system employing one of the larger cross-sectional dimension fluid treatment vessels **60**, such as any one of those shown in FIGS. **12–17** of the drawings. The particular fluid treatment system shown in FIG. **18** is for a continuous flow-through recirculating body fluid treatment wherein blood is withdrawn from the arm **13** of a patient and supplied through IV tubing **111** to a commercially available blood pump **28** and thence to an electrified treatment vessel **60**. The treatment vessel **60** may be like any of the treatment vessels described with relation to FIGS. **12–17** of the drawings wherein the blood or other body fluid being treated is exposed to a low voltage, low current electric current flow for attenuating to the point of rendering them ineffective, any contaminants entrained in the blood, such as bacteria, virus and fungus. The treated blood appearing at the output of the treatment vessel **60** then is recirculated back through IV tubing **112** to the arm **13** of the patient whose blood or other body fluid is being treated. If desired, IV tubing **111** and **112** could also be treatment tubing such as described in FIGS. **1–7** and **11**. This could provide double treatment for the fluid if that were desirable. In the event that the entire treatment does not take place in an air conditioned, temperature controlled room, then it may be desirable to provide a temperature controlled enclosure indicated by dotted lines **78** around at least the pump **28**, electrified treatment vessel **60** and the interconnecting IV tubing sections **111** and **112** in order to assure maintaining a substantially constant viscosity of the blood or body fluid being treated.

Normally, the system of FIG. **18** would be used in a continuous flow-through recirculating treatment system wherein blood from the patient's arm **13** is supplied through pump **28** to the treatment vessel **60** where it is treated and then discharged back through tubing section **112** to the arm of the patient. The flow rate of the blood thus processed would be adjusted to correspond substantially to the natural flow rate of blood circulated through the patient's body to the extent possible.

In addition to operation in the above manner, it would also be possible to operate the system of FIG. **18** in a stopped-flow, batch treatment manner wherein the blood pump is intermittently stopped to allow for more extended electrical treatment of the blood or other body fluid contained in the treatment vessel **60** during the period of time (referred to as the dwell time) that the blood pump is stopped thereby assuring fuller electrification treatment and the greater attenuation of the bacteria, virus, parasites and/or fungus entrained in the blood.

FIG. **19** is a diagrammatic sketch of a form of closed loop, flow-through recirculating treatment system according to the invention that is somewhat similar to the system shown in FIG. **18**. FIG. **19** differs from FIG. **18** in that an inlet pump **28** and an outlet pump **28'** are connected to, respectively, the intake to and outlet from the electrified treatment vessel **60**. If desired, an inlet control valve **113** and an outlet control valve **114** also can be interconnected between the inlet pump **28** and the intake to the treatment vessel **60** and between the output from the treatment vessel **60** and the intake to the outlet blood pump **28'**. These inlet and outlet control valves indicated at **113** and **114** preferably are automatically operated in a time sequence which allows the system of FIG. **19** to be operated as a two pump, start-stop flow through system. When operated in this manner, the first pump **28** is allowed to operate and discharge blood from the arm **13** of the patient to be pumped into the treatment vessel **60** and thereafter is closed off with both the inlet and outlet valves **113** and **114** in their closed condition. At this point electrification treatment of the blood or other body fluid takes place for a predetermined, scheduled time period to assure adequate attenuation to the point of rendering ineffective the contaminant bacteria, virus, parasites or fungus. Upon completion of the pre-scheduled treatment period, the outlet valve **114** is opened and outlet pump **28'** actuated to return the treated blood to the arm of the patient **13**. Operation in this semi-continuous, start-stop, batch fashion will assure that adequate electrified treatment of the blood has been accomplished while achieving this end in a somewhat continuous manner suitable for use in a closed loop, recycling blood treatment process.

PRACTICAL USES OF INVENTION

While the disclosure herein presented has been directed to principally the electrical treatment of blood, it is believed obvious to those skilled in the art that the invention can be applied with corresponding effect to other body fluids which are electrically conductive for the treatment of contaminants such as bacteria, virus, parasites and/or fungus contained therein. Further, while voltages of the order of from about 0.2 volts to 12 volts AC have been indicated as preferable, it is possible that certain virus may be attenuated (or attenuated at a faster rate) if they are subjected to greater electric current magnitudes of the order of 500 microamperes for shorter time periods. Acceptable current magnitudes normally would require an excitation voltage of from 0.2 to 12 volts. However, in certain cases where faster or more complete attenuation of the contaminants in body fluids may be desired under certain circumstances and conditions, the excitation voltage supplied to the conductive tubing may in fact exceed the 0.2 to 12 volt range indicated for most treatments.

Although it is uncertain what is specifically causing the attenuation of the contaminants (virus, bacteria, parasites and/or fungus), some possible explanations have been put forward. One is that the attenuation is caused simply by the direct affect of the electric current and voltage. Another entails the following. When a voltage is applied to the electrodes, a small current will flow through the electrically conductive medium. The applied voltage and ensuing current will induce changes in the complex biologically active fluid. Current can flow through the media if positive and/or negative charges are transported through said media. The transport might induce changes in the charge distribution of the biologically active molecules thus changing their biological activity. Furthermore, the voltage and current can induce the production or elimination of different ions, radicals, gases and/or PH levels which may affect, alone or in combination, the biologically active molecules and/or cells. The above products of the electrical processes may either be very short lived and stay in the close proximity of the electrodes or can diffuse or mix in the bulk of the media and react with the biologically active molecules or cells to result in their attenuation.

Having described several embodiments of new and improved electrically conductive treatment methods and vessels for use in practicing the novel method for the treatment of blood and/or other body fluids with electric field forces and treatment systems employing the same, it is believed obvious that other modifications and variations of the invention will be suggested to those skilled in the art in the light of the above teachings. It is therefore to be understood that changes may be made in the particular embodiments of the invention described which are within the full intended scope of the invention as defined by the appended claims.

What is claimed is:

1. An electrically conductive vessel for direct electric treatment of bacteria, and/or virus, and/or parasites and/or fungus entrained in blood and/or other body fluids and/or synthetic fluids contained within or flowing through the vessel in the presence of electric field forces, said electrically conductive vessel being fabricated with only biologically compatible material contacting the fluid being treated and with an array of at least two or more spaced-apart, opposed electrically conductive electrode segments formed of biologically compatible conductive material on or in the interior surface of the vessel and exposed to blood or other fluids contained in or flowing through the vessel, said electrically conductive electrode segments being electrically isolated from each other and extending over or through a portion of the length of the vessel, and means for applying low voltage alternating current non-biologically damaging electric potentials to the electrically conductive electrode segments whereby electrical field forces are produced between the electrically conductive electrode segments that induce biologically compatible current flow through the blood and/or other fluids contained in or flowing through the vessel so as to attenuate bacteria, virus, parasites and/or fungus contained in the blood and/or other fluids by the action of the electric current flow therethrough to thereby render the bacteria, virus, parasites and/or fungus ineffective while not impairing and maintaining the biological usefulness of the fluids.

2. An electrically conductive vessel according to claim **1** wherein the low voltage alternating current electric potentials are in the range from about 0.2 volts to 12 volts and induce electric current flow densities in the blood or other fluids of from one microampere per square millimeter ($1\ \mu A/mm^2$) to about two milliamperes per square millimeter ($2\ mA/mm^2$).

3. An electrically conductive vessel according to claim **2** wherein the vessel is in the form of tubing and is inserted in a flow-thru blood treatment system between a hypodermic needle employed to withdraw and/or supply blood from a donor and/or to a recipient and/or a blood storage receptacle or to a patient in a blood recycling system.

4. An electrically conductive vessel according to claim **2** wherein the vessel is part of a system and is in the form of tubing and a blood pump is inserted in the tubing between a donor and a recipient or a receptacle, and the system further includes means for electrically isolating the blood pump from the electrically conductive vessel, means for regulating blood flow rate from the blood pump output and means for maintaining electrical continuity throughout a desired length of the conductive vessel.

5. An electrically conductive vessel according to claim **2** wherein the vessel is in the form of tubing and the electrically conductive electrode segments are in the form zebra stripes which extend longitudinally parallel with the longitudinal axis of the tubing with the alternate electrically conductive electrode stripes being separated by alternate electrically insulating stripes for electrically isolating the alternate electrically conductive electrode stripes one from the other, a first set of alternate ones of the electrically conductive electrode stripes being electrically connected in common to a first annular terminal buss formed on and circumferentially surrounding the tubing and electrically isolated from the remaining second set of alternate electrically conductive electrode stripes, said first annular terminal buss being designed for connection to one supply terminal of a source of alternating current electric excitation potential, and a second annular terminal buss circumferentially surrounding the tubing and electrically connected to the remaining second set of alternate electrically conductive electrode stripes, said second annular terminal buss being electrically isolated from the first annular terminal buss and the first set of alternate electrically conductive electrode stripes and being designed for connection to a second supply terminal of a source of alternating current electric excitation potential.

6. Electrically conductive tubing according to claim **5** wherein the tubing is inserted in a flow-thru blood treatment system between a hypodermic needle employed to withdraw and/or supply blood from a donor and/or to a recipient and/or a blood storage receptacle or to a patient in a blood recycling system.

7. Electrically conductive tubing according to claim **5** wherein a blood pump is inserted in the tubing between a donor and a recipient and/or a receptacle, and the tubing is a part of a system which further includes means for electrically isolating the blood pump from the electrically conductive tubing, means for regulating blood flow rate from the blood pump output, and means for electrically interconnecting the input and output sides of the tubing around the blood pump and the blood flow regulating means whereby electrical continuity is maintained throughout a desired length of the tubing.

8. An electrically conductive tubing according to claim **2** wherein the vessel is in the form of tubing and the electrically conductive electrode segments are in the form of zebra stripes which extend radially around the inside diameter of the tubing in alternating conductive and insulating annular bands whereby alternate conductive bands are electrically isolated one from the other by respective insulating bands, a first set of alternate ones of the electrically conductive annular electrode stripes being electrically connected in common to a first longitudinally extending terminal buss that is formed on the tubing in parallel with the longitudinal axis thereof and electrically isolated from the remaining second set of alternate electrically conductive annular electrode stripes, said first longitudinally ext ending terminal buss being designed for connection to a first supply terminal of a source of alternating current electric excitation potential, and a second longitudinally extending terminal buss electrically connected to the remaining second set of alternate electrically conductive annular electrode stripes, said second longitudinally extending terminal buss being electrically isolated from the first longitudinally extending terminal buss and the first set of alternate electrically conductive annular electrode stripes and being designed for connection to a second supply terminal of a source of alternating current electric excitation potential.

9. Electrically conductive tubing according to claim **8** wherein the tubing is inserted in a flow-thru blood treatment system between a hypodermic needle employed to withdraw and/or supply blood from a donor and/or to a recipient and/or a blood storage receptacle or to a patient in a blood recycling system.

10. Electrically conductive tubing according to claim **9** wherein a blood pump is inserted in the tubing between a donor and a recipient and/or a receptacle, and the tubing is part of a system that further includes means for electrically isolating the blood pump from the electrically conductive tubing, means for regulating blood flow from the output of the blood pump, and means for electrically interconnecting the input and output sides of the tubing around the blood pump and blood flow regulating means whereby electrical continuity is maintained through a desired length of the tubing.

11. An electrically conductive vessel according to claim **2** wherein the walls of the vessel itself are formed from electrically conductive polymer material that is compatible with human tissue and blood and/or other body fluids with the electrically conductive portions being formed into desired patterns of spaced apart electrically conductive electrode segments physically interconnected by integrally formed electrically insulating tubing walls portions which electrically isolate a first array of electrode segments from a second array of electrode segments.

12. An electrically conductive vessel according to claim **11** wherein the vessel is in the form of tubing and the electrically conductive electrode segments are in the form of zebra stripes which extend longitudinally parallel with the longitudinal axis of the tubing with the alternate electrically conductive electrode stripes being separated by alternate electrically insulating stripes for electrically isolating the alternate electrically conductive electrode stripes one from the other, a first set of alternate ones of the electrically conductive electrode stripes being electrically connected in common to a first annular terminal buss formed on and circumferentially surrounding the tubing and electrically isolated from the remaining second set of alternate electrically conductive electrode stripes, said first annular terminal buss being designed for connection to one supply terminal of a source of alternating current electric excitation potential, and a second annular terminal buss circumferentially surrounding the tubing and electrically connected to the remaining second set of alternate electrically conductive electrode stripes, said second annular terminal buss being electrically isolated from the first annular terminal buss and the first set of alternate electrically conductive electrode stripes and being designed for connection to a second supply terminal of a source of alternating current electric excitation potential.

13. Electrically conductive tubing according to claim **12** wherein the tubing is inserted in a flow-thru blood treatment system between a hypodermic needle employed to withdraw and/or supply blood from a donor and/or to a recipient and/or a blood storage receptacle or to a patient in a blood recycling system.

14. Electrically conductive tubing according to claim **13** wherein a blood pump is inserted in the tubing between a donor and a recipient and/or a receptacle, and the tubing is part of a system which further includes means for electrically isolating the blood pump from the electrically conductive tubing, means for regulating blood flow from the output of the blood pump, and means for electrically interconnecting the input and output sides of the tubing around the blood pump and blood flow regulating means whereby electrical continuity is maintained throughout a desired length of the tubing.

15. An electrically conductive vessel according to claim **11** wherein the vessel is in the form of tubing and the electrically conductive electrode segments are in the form of zebra stripes which extend radially around the inside diameter of the tubing in alternating conductive and insulating annular bands whereby alternate conductive bands are electrically isolated one from the other by respective insulating bands, a first set of alternate ones of the electrically conductive annular electrode stripes being electrically connected in common to a first longitudinally extending terminal buss that is formed on the tubing in parallel with the longitudinal axis thereof and electrically isolated from the remaining second set of alternate electrically conductive annular electrode stripes, said first longitudinally extending terminal buss being designed for connection to a first supply terminal of a source of alternating current electric excitation potential, and a second longitudinally extending terminal buss electrically connected to the remaining second set of alternate electrically conductive annular electrode stripes, said second longitudinally extending terminal buss being electrically isolated from the first longitudinally extending terminal buss and the first set of alternate electrically conductive annular electrode stripes and being designed for connection to a second supply terminal of a source of alternating current electric excitation potential.

16. Electrically conductive tubing according to claim **15** wherein the tubing is inserted in a flow-thru blood treatment system between a hypodermic needle employed to withdraw and/or supply blood from a donor and/or to a recipient and/or a blood storage receptacle or a patient in a blood recycling system.

17. Electrically conductive tubing according to claim **16** wherein a blood pump is inserted in the tubing between a donor and a recipient and/or a receptacle, and the tubing is part of a system that further includes means for electrically isolating the blood pump from the electrically conductive tubing, means for regulating blood flow from the output of the blood pump, and means for electrically interconnecting the input and output sides of the tubing around the blood pump and the blood flow regulating means whereby electrical continuity is maintained throughout a desired length of the tubing.

18. A fluid treatment process for attentuating bacteria, and/or virus, and/or parasites, and/or fungus, existing in blood and/or other body fluids and/or synthetic fluids within a treatment vessel having only biologically compatible internal and conductive electrode surfaces therein contacting fluid being treated thereby maintaining the biological usefulness of the blood or other fluids being treated comprising subjecting the fluid within the treatment vessel to low voltage, low alternating current electric field forces within non-biologically damaging electric field forces for producing a biologically compatible current flow through the blood or other fluids for a predetermined period of time sufficient to attenuate bacteria and/or virus, and/or parasites and/or fungus contained in the blood or other fluids to thereby render them ineffective while maintaining the biological usefulness of the fluids being treated.

19. The product of the process according to claim **18**.

20. A fluid treatment process according to claim **18** wherein the low voltage alternating current electric potentials are in the range from about 0.2 to 12 volts and induce electric current flow densities in the blood or other fluids of from one microampere per square millimeter (1 $\mu A/mm^2$) to about two milliamperes per square millimeter (2 mA/mm^2).

21. The product of the process according to claim **20**.

22. A fluid treatment system for attentuating bacteria, and/or virus, and/or parasites, and/or fungus existing in blood and/or other body fluids and/or synthetic fluids being treated without biological damage to the blood or other fluids comprising an electrically conductive vessel formed at least in part of biologically compatible conductive material for contacting blood or other fluids to be treated, means for subjecting the blood or other fluids within the conductive vessel to low voltage, low alternating current electric field forces for producing biologically compatible current flow through the blood or other fluids for a predetermined period of time sufficient to attenuate bacteria and/or virus, and/or parasites, and/or fungus contained in the blood or other fluids to thereby render such contaminants ineffective while maintaining the biological usefulness of the blood or other fluids.

23. A fluid treatment system according to claim **22** wherein the low voltage alternating current electric potentials are in the range from about 0.2 to 12 volts and produce electric current flow densities in the blood or other body fluids of from one microampere per square millimeter (1 $\mu A/mm^2$) to about two milliamperes per square millimeter (2 A/mm^2).

24. A fluid treatment system according to claim **22** wherein the system comprises a plurality of components including an electric power source all of which the miniaturized and implanted in the body of a patient being treated to form a closed loop, continuous recirculating body fluid treatment system.

25. A fluid treatment system according to claim **22** wherein the conductive vessel is in the form of an open ended tube to allow flow-thru treatment of blood and other fluids and is miniaturized along with an electric power source for supply of alternating current electric potentials thereto whereby the system may be implanted in human beings and other mammals to operate as a continuous recirculating fluid treatment process.

26. A fluid treatment system according to claim **22** wherein the conductive vessel in the vicinity of the spaced-apart opposed electrically conductive electrode segments is provided with an enlarged cross sectional area wherein enlarged electrically conductive electrode segment surface areas are provided to act on the blood or other fluids flowing through the vessel thereby increasing the through-put and/or effectiveness of the treatment accomplished within the vessel for a given dwell time.

27. A body fluid treatment system according to claim **26** wherein the electrically conductive vessel comprises an enlarged rectangular-shaped body of electrical insulating material having a plurality of parallel, longitudinally extending tubular openings formed all the way through the insulating material from one end to the other and having spaced-apart electrically conductive metal strips secured to respective opposite sides of all of the tubes in opposed, parallel relationship, one set of corresponding conductive strips of all of the tubes extending out of the ends of each tube on one side or end of the body of electrical insulating material and contacting a conductive surface forming a terminal buss for all conductive strips of the set, and the remaining set of conductive strips projecting out of the opposite ends of the respective tubes on the opposite end of the insulating block to engage a conductive terminal surface, and header reservoirs formed on each of the ends of the body of electrical insulating material into which the ends of the tubular openings are connected, each header having a respective inlet or outlet opening for supply of blood and/or other fluids for treatment thereto.

28. A fluid treatment system according to claim **27** wherein the enlarged insulating clock member is cylindrically shaped and the header reservoirs at each end of the block member are correspondingly cylindrically shaped.

29. A fluid treatment system according to claim **27** wherein selectively operated gas vents are provided in the top of the respective header reservoirs of the electrically conductive vessel.

30. A fluid treatment system according to claim **26** wherein the electrically conductive vessel is in the form of an enlarged cross sectional area treatment vessel of substantially greater cross sectional area than the inlet and outlet conduits supplying body fluids to be treated to the vessel and wherein the enlarged cross sectional area vessel is included in a blood transfer system between a hypodermic needle employed to withdraw and/or supply blood from a donor and/or to a recipient and/or a blood storage receptacle or to a patient in a continuous flow-thru blood recycling system.

31. A fluid treatment system according to claim **30** wherein a blood pump is inserted in the flow path of the blood or other fluid either to or from the enlarged cross sectional area vessel, or both, and are located in a tubing system between the donor and recipient or receptacle, and the system further includes means for regulating blood flow rate from or to the enlarged cross sectional area treatment vessel via the inlet or outlet pumps or both.

* * * * *

rexresearch.com

Danielle ZUROVCIK

Negative-Pressure Wound Healing Device

Accelerated wound-healing with vacuum-assisted skin dressing

---

http://www.technologyreview.com/news/418098/a-cheap-portable-wound-healing-device/?a=f
March 19, 2010

A Cheap, Portable Wound-Healing Device

After the Haiti earthquake, physicians tested a vacuum pump meant to speed healing.

by

Emily Singer

In mid-February, about a month after a massive earthquake leveled much of Port-au-Prince, Haiti, a wound-care team from Brigham and Women's hospital in Boston traveled to the devastated capital. The team's task was to help care for scores of patients suffering from the large open wounds that accompany amputations, crushed limbs, and other injuries. Among the team was MIT graduate student Danielle Zurovcik, who arrived ready to test a device she had developed as part of her thesis research–a cheap and portable version of the negative-pressure devices currently used to speed wound healing in hospitals.

Zurovcik and her collaborators hope the device, which costs about $3, will provide a way to improve care for patients after the emergency phase of relief efforts, including life- and limb-saving surgeries, has ended. Even after many of the emergency medical teams leave the disaster zone, the dangers of chronic wounds remain high.

“My experience in Haiti and other major earthquakes is that after the acute medical response, such as amputating limbs and setting fractures, the major disease burden is wounds,” says Robert Riviello, a trauma surgeon at Brigham and Women’s, and Zurovcik’s collaborator. Negative-pressure therapy decreases the need to change wound dressings from one to three times per day to once every few days, a major benefit when medical staff is in short supply.

Negative-pressure devices, which act like a vacuum over the bandaged wound, have become a central part of wound therapy in the United States over the last decade. They speed healing up to threefold, depending on the type of wound, and in some cases eliminate the need for plastic surgery or skin grafts. A number of commercial versions are available in the U.S. and are used to treat burns and chronic wounds such as bed sores or diabetic foot ulcers. While scientists don’t exactly know why this treatment accelerates the healing process, it likely helps by removing some of the fluid and bacteria that accumulates at the injury site and by increasing blood flow to the wound. The pressure itself may also help healing by bringing together the edges of the wound and delivering mechanical pressure, which has been shown to spur cell growth, says Dennis Orgill, a surgeon at Brigham and Women’s who was not involved in the project.

Existing devices are often heavy, about five to 10 pounds, and require an energy source to create the vacuum, making them difficult to apply in disaster settings. Texas-based KCI, the leading maker of negative-pressure machines, has a portable version that’s battery powered, but it costs approximately $100 per day to rent. A number of companies are working on even more portable versions, say Orgill.

But Zurovcik, inspired by a burn surgeon’s plea, went a step further, designing a human-powered device that applies pressure via a simple bellows pump weighing less than half a pound. By improving the seal around the wound dressing to reduce air leaks, Zurovcik cut the pump’s power requirements from about 14 watts to 80 microwatts, which comes from a hand pump.

“To basically take a toilet plunger and produce negative pressure over a prolonged period of time, that is really great,” says Kristian Olson, a physician at Massachusetts General Hospital, in Boston, who was not involved in the project. “Not only do I see it answering this need in developing countries, I think it could really enhance home therapy for chronic wounds in the U.S.”

Zurovcik and Riviello had been planning a trial of the device in Rwanda – Riviello spends about half his time working in Africa–when the earthquake hit Haiti. Colleagues treating the first waves of injury victims told the duo that their device might be of help, so they joined a wound-care team headed for University Hospital, a few blocks from the leveled palace in Port-au-Prince. (Commercial negative-pressure devices, known as VACS, were employed in various relief efforts in Haiti, including $2 million worth of equipment donated by KCI.)

Working in stifling tents filled with patients, the team tended to those whose doctors had left and were in need of follow-up care. Of the hundreds of patients assessed, the researchers chose eight people suffering from a variety of injuries– amputations, open tissue wounds, open fractures, crushing injuries (where the skin had to be opened to give the muscle room to expand), infected surgical wounds, and bedsores from being paralyzed–appropriate for negative-pressure therapy. “Because this was a disaster setting, we didn’t feel it was an appropriate place for rolling out a randomized controlled trial,” says Riviello. (They cared for other patients with typical dressings.)

The surgeon would first apply a sponge over the cleaned wound and then cover it with a plastic seal. A tube fed through a small hole in the plastic connected to the pump, which was manually compressed to create negative pressure. The team trained patients’ families, who often took on typical nursing duties, to charge, or pump, the device. “We learned that family members are interested in being trained and motivated to keep the device charged because they saw the benefits for their loved ones," says Riviello. "They were tremendously reliable. We saw patients twice a day, but it became clear that we could come back days later and the device would still be charged."

Because the researchers were in Haiti for just 10 days, they weren't able to determine if the device helped patients heal faster. But it did seem to keep the wounds cleaner, says Riviello, and reduced the need to change bandages, which is painful for the patient. In fact, one patient requested the treatment after observing how a neighbor in the next bed was subjected to fewer painful dressing changes, says Zurovcik. Now back in Cambridge, she is tinkering with the prototype, trying to further improve the pressure seal and the amount of negative pressure the device can deliver.

The team plans a larger test in Rwanda, where it will likely put the device to broader use. People in poor countries are much less likely to survive severe burns, for example, which can be helped with negative-pressure therapy. And the rate of complications from diabetes, such as foot ulcers, is skyrocketing in these countries as well, says Olson.

Modifiable Occlusive Skin Dressing
US2014031735

[PDF]

Occlusive tissue dressings and methods including an elastomeric drape and a liquid component, at least partially cross-linked at least after one of drying and curing, suitable for application at a dressing-to-skin interface in order to create a substantially air-tight seal. The same or a different liquid component may be applied by a user at a tube-to-dressing interface in order to create a similar air-tight seal around the tube, if not occlusively sealed during its manufacture.

CROSS-REFERENCE TO RELATED APPLICATION

[001] This application claims priority to U.S. Patent Application No. 13/745,690 filed January 18, 2013 and U.S. Provisional Application No. 61/588,121 filed 18 January 2012.

FIELD OF THE INVENTION

[002] This invention relates to dressings intended to provide a fluid-impervious barrier over skin, and more particularly to dressings suitable for negative pressure wound therapy.

BACKGROUND OF THE INVENTION

[003] Negative pressure wound therapy ("NPWT") is an effective technology for treating open wounds. NPWT devices were originally accepted by the US Food and Drug Administration ("FDA") in 1995, when the FDA approved a 510(K) for the Kinetic Concepts Inc. ("KCI")'s V.A.C.(R) device. The definition of NPWT devices by the FDA has changed over the years; in general terms, its definition is: a system that is used to apply negative pressure for wound management purposes, including the removal of fluids (i.e., wound exudates, irrigation fluids, and infectious materials). The negative pressure is applied through a porous dressing positioned into or over the wound cavity, depending on wound type and depth, or over a flap or graft; the dressing distributes the pressure while removing fluids from the wound. NWPT systems typically include:

<-> Non-adhesive wound dressing used to fill the wound cavity (e.g., a sterilized medical sponge or gauze; a.k.a., non-adhesive packing materials);

<-> Drainage tube placed adjacent to or into the dressing;

<-> Occlusive transparent film placed over the dressing (and potentially the drainage tube) and adhered to

the skin to maintain a seal;

<-> Collection container for drained fluids from the wound; and

<-> Low pressure vacuum source.

[004] NPWT has been approved by the FDA to treat many wound types: chronic, acute, traumatic, sub-acute and dehisced wounds, partial-thickness burns, ulcers (such as diabetic, venous or pressure), surgically closed incisions (a.k.a., closed surgical incisions), flaps and grafts. The prescribed therapy time depends on wound type, wound dimensions, and patient conditions; it typically lasts from four weeks to four months. Disposable dressing components are changed approximately every three days.

[005] Extensive clinical trials have demonstrated the success of negative pressure in healing the approved wound types by applying a controlled negative pressure typically between 20mmHg and 200mmHg. Most studies applied a constant vacuum pressure, with 125mmHg being the most common, although cyclic and intermittent studies are currently underway. Evidence supporting the use of NPWT in the treatment of chronic, non-healing wounds exists primarily in the form of nonrandomized, controlled trials; prospective and retrospective large and small case series; single-center studies; and single case studies, with few randomized, controlled clinical trials. Studies also exist that demonstrate NPWT benefits in healing acute wounds. Additionally, since 2006, benefits of managing surgical incisions post-operatively have been shown with improved clinical outcomes; at least ten studies have been published to date. From these studies, proven medical benefits of NPWT treatment include:

<-> Promotes blood flow (perfusion) at the wound;

<-> Removes interstitial fluid (a.k.a., wound exudates), reduces edema;

<-> Decreases counts of bacteria and infectious materials;

<-> Increases rate of granulation tissue formation, reducing scar tissue formation, increases growth factors and fibroblasts;

<-> Uniformly draws the wound edges together;

<-> Provides a protected healing environment; and

<-> Provides a moist environment.

[006] Although significant clinical evidence exists to support the benefit of NPWT as a safe therapy in healing chronic wounds, it is possible during NPWT to rupture a vein or artery. Usually, a machine safety alarm will signify a fluid leak rate that exceeds the rate that the machine was designed for. This alarmed leak rate typically includes the combination of both air and liquid, and typically has an upper safety limit of the minimum blood flow rate possible out of a wound cavity with an actively bleeding vein or artery. If a vein or artery accidently ruptures, the system must shut down. Therefore, it is very important to have a safety feature that stops blood flow if this occurs, in order not to exsanguinate the patient.

[007] Lina et al. describe in U.S. Patent No. 7,611,500 and WO1996/005873 an initial apparatus used for NPWT. In practice, the device proved to be effective; however, one major limitation was detected: the high electrical grid power source needed to operate the device limited the mobility of a patient. Therefore, future refinements, such as that described by Hunt et al. in U.S. Patent No. 6,142,982, incorporated rechargeable batteries for the power source. Batteries increased patient mobility, but time was limited by the life of the batteries between charges. Additionally, battery management became an issue, especially for facilities with a high number of NPWT patients, and electrical grid power was still needed to recharge the batteries.

[008] Eliminating the need for electrical power, via the grid or batteries, would create a more widely applicable, clinically viable therapy. The power requirement variability of a system is dependent on the desired vacuum pressure, rate of wound exudate removal from the wound cavity, and the leak rate of air into the system. As the air leak rate increases, more power is needed to supply a continuous negative pressure at

a predetermined value or threshold range at the wound bed. Air leakage into the NPWT system requires the most power of any other component. Air leaks are the obstacle to creating a vacuum system that does not require a continuous external power source or frequent recharging of its internal power storage. Therefore, the feasibility of a mechanical NPWT system is heavily reliant on the seal quality of every interface in the system. The dressing system has been identified as the main source of air leaks in current NPWT systems, particularly at the interfaces between 1) the dressing and the skin and 2) the tube and the dressing. The amount of air leaks into these interfaces determines the time frequency that the pump needs to be recharged and the magnitude of vacuum pressure applied to the wound cavity at a specific time. These two latter characteristics are dependent system parameters.

[009] Few mechanical NPWT systems are currently available, as described by the present inventor in "Development of a simplified Negative Pressure Wound Device" submitted in 2007 for her Master of Science in Mechanical Engineering at the Massachusetts Institute of Technology. Certain lower-pressure, mechanical devices were disclosed later by Hu et al. in U.S. Patent Application No. 2010/0228205. Current mechanical systems typically use sophisticated-material, planar dressings, such as hydrocolloid dressings, to try to solve the air leak problem. However, the inherent geometry mismatch of a planar dressing and the contoured skin surface often leads to air leaks. The mechanical devices therefore are only applicable for select, relatively flat surfaces on the body and, even then, it is difficult to eliminate air leaks entirely. [010] Non-electrical pumps are at the low end of the spectrum of medical pumps, typically utilizing bladder pumps and capillary action materials. Bladder pumps are used for both extracting and inserting fluids. By their physical characteristics, they are governed by non-linear spring like properties. Currently, bladder pumps are used in wound treatments for drainage purposes, particularly for internal, body cavity drainage. C. R. Bard, Inc. manufacturers many of these non-electrical pumps; one bladder model frequently used to drain internal cavities is commonly referred to as a Jackson Pratt Drain.

[011] There are various limitations to applying NPWT with existing mechanical, bladder pumps. There are no pressure gauges on the pumps and, therefore, the user does not know the initial magnitude of the negative pressure pulled, and cannot monitor the pressure during therapy. Additionally, there are no air leak detection systems for the current pumps, except to visually watch for the expansion of the bladder at a rate higher than expected. If the pump is clear, one can also visually monitor if the expansion rate is due to air leaks or due to drainage fluid.

[012] Capillary action materials are also currently used to treat wounds by providing very low negative pressure treatment, too low to be considered NPWT. This form of treatment is usually found in dressings such as small topical bandages to provide NPWT- like benefits to very small, self-healing wounds, such as blisters and brush burns. Treating a wound with this technology enhances the healing environment. Capillary action materials are filled with small capillaries between the wound and outside environment. A negative pressure is applied by capillary action of fluid flowing from the wound to the outside environment, thereby, removing interstitial fluid. One example of a capillary action material is Johnson & Johnson's First Aid Advanced Care Advanced Healing Adhesive Pads.

[013] Dressing technologies have tried to address the issue of air leaks into NPWT systems. This is important to both electrical and mechanical systems to reduce their necessary power requirements. In mechanical systems, it is necessary for clinically relevant device functionality, such that power input and pump recharge time is reasonable for a caregiver and/or patient to perform. For electrical systems, air leak reduction reduces the number of, if not completely eliminates, false-positive, alarmed emergency system shutdowns. Air leak reduction allows battery designs to last longer on one battery charge and use lower power capacity batteries altogether. Air leak elimination potentially eliminates the need for a continuous power supply, as the vacuum pressure can be maintained in the occlusive environment within a specified threshold, for which the timeframe depends on the pump parameters and exudate removal rate (typically less than 100 mL/day) from the wound.

[014] Currently, most NPWT dressings (the drape component) are thin, planar, tape-like adhesive dressings that must be applied to a contoured area of skin. A backing on the dressing must be removed to expose the adhesive, and then the dressing is applied to the skin. The pre-application handling of the dressings alone introduces a probability for air leaks, as the dressing typically folds onto itself or creases very easily due to its low bending stiffness; many dressings are thinner than a piece of standard paper, and the bending stiffness of a material is proportional to the inverse of its thickness cubed. As a dressing is applied, it must

often fold onto itself in order to accommodate for a geometrical mismatch between the planar dressing and the contours of the body surrounding the wound to be treated. This creates creases, also referred to herein as wrinkles, in the dressing that have a high potential for causing air leaks into the NPWT system.

[015] Adding to the geometrical mismatch, the dressings often become less adhesive due to the introduction of foreign materials onto the adhesive before dressing application. This is most common and almost unavoidable at the edges of the dressing due to handling by the caregiver. At times, the caregiver's hands introduce enough foreign particles onto the adhesive to forbid further adhesion of that area of the dressing. In the U.S., this often happens when a caregiver uses powdered gloves. This is a critical issue as the edges of the dressing are an area where leak propagation from the edge of the dressing to the wound cavity is potentially very high, based on the theory of interface fracture mechanics.

[016] For the electrical NPWT systems, a thin plastic, adhesive backed dressing is typically used. Electrical NPWT dressing systems have not readily addressed the air leak issues listed above that form at the dressing-to-skin interface. Instead, dressing iterations have focused on air leaks at the tube-to-dressing interface. When NPWT was first introduced into the market, the drainage tube was inserted into the wound cavity through the edge of the dressing. This introduced a high potential for air leaks, which often alarmed the shut-off system. Caregivers began to solve this problem by raising the tube from the skin surface at the dressing edge, and pinching the dressing under the tube before the dressing contacts the skin. This caused the dressing to adhere to itself in an upside-down "T" pattern onto the skin. [017] Eventually, some of the NPWT dressing, commercial designs incorporated their own solutions to the high air leak rate at the tubing interface. Out of these solutions, the T.R.A.C. Pad by KCI was highly effective, which is driving the current design trends. The T.R.A.C. Pad prefabricates the drainage tube to a semi-rigid, tubing connector, which is then attached to a small, circular, planar adhesive dressing (a.k.a., drape). All of these connections are made air-tight during its manufacture. The tubing does not travel beyond the plane of the adhesive dressing, and therefore its opening remains at the skin surface. When the T.R.A.C. Pad is used, the standard dressing is initially applied to the wound, without a tubing connection. Then, a small incision is made in the dressing, over the wound cavity; this hole may also be prefabricated into the drape component of the dressing during its manufacture. The film backing of the circular adhesive component is removed from the Pad, and the tube opening is centered over the incision. Since the adhesive surface of the Pad is small, it is easier to handle than the procedure of tunneling the tube into the initial dressing. Although the Pad does not guarantee elimination of air leaks at the tube-to-dressing interface, it highly reduces the probable amount of air leaks into the dressing, based on its ergonomic design and small profile. A minimal amount of air leaks is almost unavoidable for all applications with planar adhesive components, due to the geometrical mismatch and user handling that still remain.

[018] Many efforts have been made in order to overcome the identified barriers of low end, mechanical pumps for application in NPWT. Most of the focus has been on reducing air leaks and creating more predictable vacuum sources. New materials used in NPWT dressings have been the main driver in reducing the air leak rate into the system at the dressing-skin interface. These materials are often not new to wound dressings; however, they are new to NPWT. Pump design has been the focus of creating more predictable vacuum sources; mechanical components, such as linear or constant force springs, are often introduced into the system and maintain a more predictable behavior throughout therapy.

[019] Only one mechanical NPWT system is on the market today, but is not widely used: SNaP(R) Wound Care System by Spiracur (Sunnyvale, CA). The SNaP(R) Wound Care System uses a hydrocolloid dressing with specific mechanical connectors from the tube to the dressing, in order to accommodate for air leaks; the provided hydrocolloid dressing is relatively small in size. Hydrocolloids are used in many wound-dressing systems, and are a common trend in the NPWT market. They are stiffer and thicker than most common, adhesive, planar, NPWT specific dressings. This causes the dressing to fold onto itself less during its handling and application. However, it cannot accommodate for geometrical mismatch without creases, especially as the dressing surface area increases. Since the dressing is stiffer and thicker, these creases are difficult to seal in an air-tight manner, due to its increased bending stiffness. Therefore, hydrocolloids are often only applicable to smaller wounds. Much effort is currently being taken to make them thinner, in order to increase their applicable surface area and accommodate more for contours, such as the Replicare Thin Hydrocolloid Dressing by Smith and Nephew. Hydrocolloids rely on their extremely sticky adhesive properties to account for increased skin adhesion and reduced air leaks. If they come in contact with wound exudate, the polymers in the hydrocolloid swell with water until saturation, forming a gel, which is held

solid in its adhesive matrix structure.

[020] In the SNaP(R) Wound Care System, the hydrocolloid dressings are connected to the tubing with a mechanical connector component, similar to the T.R.A.C. Pad, KCI. The SNaP(R) Wound Care System eliminates any potential air leaks from this mechanical connector by prefabricating it to the center of the entire dressing during manufacture. The prefabrication eliminates any potential air leaks at the tube-to-dressing interface due to user interface and geometrical mismatch, but it is not capable of being moved on the dressing surface. Therefore, it may need to be placed on an inconvenient area of the wound, such as a location that is uncomfortable for the patient. Additionally, the tube runs parallel to the plane of the drape; the direction of the tube along the plane of the drape is fixed. Since the dressings are not typically round, the tube path may be required to travel in an undesirable path, in order to cover the wound area with the preset shape of the drape.

[021] For its vacuum source, the SNaP(R) Wound Care System uses a complex system, driven by constant force springs. Therefore, as the pump expands, mainly due to air leaks and potentially exudate removal, the pressure remains relatively constant for the length of the pressure application. This system is expensive and highly technical when compared to the non-electrical pumps at the low end of the medical pump spectrum (e.g., bladder pumps); however, it is the first commercial mechanical NPWT pump, which has been proven to be a potential NPWT pump design. Since air leaks into the dressing system remain highly probable, depending on wound location and caregiver experience, the successful application of the SNaP(R) Wound Care System is limited in practice.

SUMMARY

[022] Occlusive skin dressings according to the present invention preferably provide one or more of the following advantages:

<-> a conformable dressing system that can be altered if desired and applied to substantially all areas of the skin surface;

<-> a dressing system that is ergonomic;

<-> dressings that are easy to obtain and re-obtain by the user, through conveniences in storage;

<-> dressings, pumps, systems, and methods to administer NPWT without the need for electrical power;

<-> minimizing the amount of air leaks into the NPWT system;

<-> detecting air leaks into the NPWT system;

<-> compatible with light-weight, easily transportable and low cost pumps; and

<-> mechanical methods to minimize the possibility of exsanguinating the patient.

[023] Occlusive dressings according to the present invention overcome the aforementioned drawbacks by being truly air-tight. One principal application of this technology is to facilitate administration of mechanical NPWT. A liquid component is applied at the dressing-to-skin interface in order to create a substantially air-tight seal preferably for at least 48 hours, more preferably for at least 72 hours. Preferably the same or different liquid component is applied at the tube-to-dressing interface in order to create a similar air-tight seal. In some embodiments, the liquid components may be made of rubber polymers applied by touch, by squeezing a dispenser, or by spraying the polymers with an atomization process.

[024] This invention features a kit suitable for occlusively sealing a wound penetrating the skin of a patient, including a drape formed as a thin sheet of an organic, preferably elastomeric material, substantially impervious to fluid transfer of air and bodily fluids, having first and second surfaces. A biocompatible adhesive is at least one of (1) disposed on at least the first surface of the drape and (2) capable of contacting at least a portion of at least the first surface of the drape. When the kit includes the biocompatible adhesive disposed on at least a portion of the first surface of the drape, the kit further includes at least a first

removable liner sheet covering the first surface of the drape. In some embodiments, a second removable liner sheet covers the second surface of the drape, especially when adhesive is also disposed on the second surface of the drape. The kit further includes at least one container of at least one sealant component that is capable of being delivered as a sealant in a liquid state at pre-selected ambient conditions, the sealant as delivered being at least partially cross-linked at least after one of drying and curing, and which is capable of at least one of drying and curing within thirty minutes, preferably within twenty minutes and, more preferably, within ten minutes, after application of the sealant as a layer to the edges of the drape after the drape is applied to the skin surrounding the wound.

[025] In some embodiments, the drape and the sealant after one of drying and curing are elastomeric. In a number of embodiments, the drape and the sealant are derived from substantially the same material, such as a type of a latex compound or a type of silicone compound. In certain embodiments, the adhesive is a silicone-based adhesive and is disposed on at least a majority of each of the first and second surfaces of the drape as a solid coating or in a pattern such as a grid or concentric circles. At least one container of a sealant component enables manual application of the sealant in some embodiments, such as by squeezing the container and, in other embodiments, at least one container is a removable vial or cartridge insertable into a dispensing apparatus or other applicator. In a number of embodiments, the kit further includes a flexible tube having a first end and a second end connectable to a source of negative pressure such as a bellows, especially a novel bellows which unrolls, or other mechanical vacuum source. Preferably, the kit further includes a flange having at least one of (1) a central passage through which the first end of the tube is insertable and (2) a central passage communicating with a connector capable of mating with the first end of the tube. In one embodiment, the first end of the tube includes a feature such as a spiral cut to resist blockage of the tube. In some embodiments, the kit includes at least one non-stick handling component. In a number of embodiments, the kit further includes at least one wound packing material.

[026] This invention may also be expressed as a method of constructing an occlusive dressing over a wound penetrating the skin of a patient by selecting a drape formed as a thin sheet of an elastomeric material, substantially impervious to fluid transfer, and having first and second surfaces. A biocompatible adhesive is selected that is at least one of (1) disposed on at least the first surface of the drape, preferably with a first removable liner sheet covering the first surface of the drape and (2) applied to at least one of (i) the skin of the patient surrounding the wound and (ii) at least a portion of at least the first surface of the drape. Optionally, a second removable liner sheet covers the second surface of the drape. The method includes removing the first removable liner, if present, and placing the drape onto the skin surrounding the wound, removing the second removable liner if present, and applying a sealant that is in a liquid state as applied, the sealant being at least partially cross-linked at least after one of drying and curing, on at least the edges of the drape and on the skin adjacent to the drape in one or more layers. The method further includes at least one of drying and curing the sealant within thirty minutes, preferably within twenty minutes, after application of the sealant to the edges of the drape in at least one layer.

[027] In certain embodiments, the adhesive is disposed on at least a majority of each of the first and second surfaces of the drape, and/or the method includes pressing on the second surface of the drape in the vicinity of any wrinkles in the drape, preferably before sealant is applied in that vicinity. In some embodiments, a flexible tube is selected having a first end and a second end connectable to a source of negative pressure such as a bellows or other mechanical vacuum source. Preferably, the first end of the tube (1) is inserted through a central passage of a flange secured to the drape or (2) is mated with a connector on a flange having a central passage communicating with the connector. In one embodiment, the first end of the tube includes a feature such as a spiral cut to resist blockage of the tube. In some embodiments, the wound is packed with gauze or other fluid-pervious material prior to placing the drape on the skin.

[028] This invention may be further expressed as a method of constructing an occlusive dressing over a wound, penetrating the skin of a patient, by at least one of (1) packing the wound with a fluid-pervious material and (2) covering at least a portion of the wound with a protective material. The method further includes applying, such as by spraying, an elastomeric material that is in a liquid state, and is at least partially cross-linked at least after one of drying and curing, over the packed material and onto skin surrounding the wound to create an occlusive drape as a thin sheet substantially impervious to fluid transfer, having a first, inner surface and a second, outer surface. The method includes at least one of drying and curing the elastomeric material within thirty minutes after application of the elastomeric material as a layer.

BRIEF DESCRIPTION OF THE DRAWINGS

[029] In what follows, preferred embodiments of the invention are explained in more detail with reference to the drawings, in which:

[030] FIG. 1 is a schematic expanded perspective view of a drape, flange and tube with first and second liners prior to application of a liquid sealant according to the present invention;

[031] FIGS. 2 and 3 illustrate a novel first end of the tube of FIG. 1 being inserted through the novel, preferably symmetrical flange;

[032] FIG. 4 is a schematic perspective view of an alternative novel first end of a tube;

[033] FIGS. 5A and 5B illustrate repositioning of the upright tube into a desired side orientation;

[034] FIGS. 6 and 7 show a drape being covered by an upper liner to manufacture a dressing according to the present invention;

[035] FIG. 8 shows a hole punched in the dressing of FIG. 7;

[036] FIGS. 9 and 10 shows a tube assembly being inserted onto the dressing of FIG. 8 with the edge of the flange being sealed to the drape;

[037] FIG. 11 shows a protective liner being added to the dressing of FIG. 10;

[038] FIG. 12 illustrates how a user can cut the dressing of FIG. 11 to conform to a wound;

[039] FIG. 13 shows a handling tab being added to the dressing of FIG. 12;

[040] FIGS. 14 and 15 illustrate debriding and packing an open wound;

[041] FIG. 16 is a perspective view of the underside of the dressing of FIG. 1 1 with the bottom, inner protective layer being removed;

[042] FIG. 17 is a schematic top plan view of a dressing according to the present invention attached to skin surrounding the wound;

[043] FIG. 18 is a schematic perspective view of the dressing of FIG. 17 with the upper protective liner being removed;

[044] FIG. 19 shows liquid sealant being applied to the edges of the drape of FIG. 18;

[045] FIGS. 19A and 19B illustrate modifying the coverage of a dressing according to the present invention;

[046] FIG. 20 is a schematic expanded view of a vial of sealant with a non-stick finger protector optionally positionable within the vial for storage and transportation;

[047] FIGS. 21 A and 2 IB show a dispensing apparatus with removable cartridge of liquid sealant;

[048] FIG. 22 is an enlarged perspective view of the cartridge of FIGS. 21 A and 21B;

[049] FIG. 23A is schematic perspective view of a hand-powered squeeze applicator for liquid sealant;

[050] FIGS. 23B and 23C are enlarged views of the outlet with and without a removable strip covering the dispensing openings;

[051] FIGS. 24 and 25 are schematic top plan views illustrating non-stick gloves and finger covers, respectively, integrated into a liner;

[052] FIGS. 26-28 are schematic top plan views liners having different shapes;

[053] FIG. 29 is a schematic side view of a dressing according to the present invention being applied to the heel of a foot;

[054] FIGS. 30 and 30A are enlarged schematic views of the dressing of FIG. 29 with a fold being created and then pressed flat to enhance conformance to the heel;

[055] FIG. 31 is a cross-sectional view of a known bellows pump;

[056] FIG. 32 is a perspective view of a novel rolling bellows pump;

[057] FIG. 33 is a flow chart of a sample occlusive dressing method;

[058] FIGS. 34 A and B are diagrams comparing active versus passive flow NPWT systems;

[059] FIG. 35 is a diagram of a method to prevent undermining of the dressing due to exudate build-up in the passive NPWT system;

[060] FIG. 36 is a diagram of some of the specific components of an occlusive wound dressing embodiment;

[061] FIG. 37 is a diagram of a tube connection method with a spiral end tube;

[062] FIG. 38 is a schematic top plan view of the wound shown in FIG. 14 with the additional step of applying a protective covering over the wound;

[063] FIG. 39 is a view of FIG. 38 with a hole cut in the protective covering;

[064] FIG. 40 is a view of FIG. 39 with a tube assembly placed over the hole;

[065] FIG. 41 is a view of FIG. 40 with liquid drape material applied over the protective covering and onto surrounding skin to construct a dressing according to the present invention;

[066] FIG. 42 is a schematic perspective view of the dressing of FIG. 41; and

[067] FIG. 43 is a schematic perspective view of a novel flange according to the present invention with integral connector.

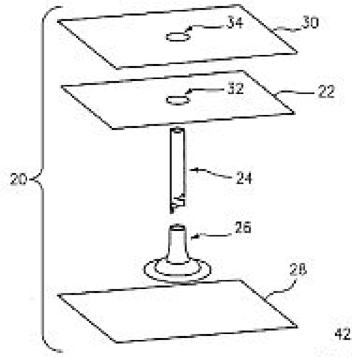
34
30
32
22
20
24
26
28

FIG. 1

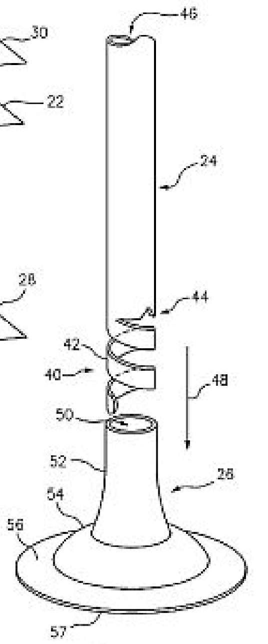
46
24
44
42
40
48
50
52
26
54
56
57

FIG. 2

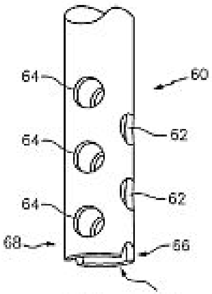
64
60
62
64
64
62
68
66
70

FIG. 4

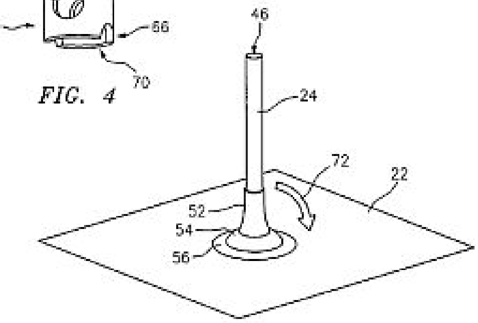
46
24
72
22
52
54
56

FIG. 5A

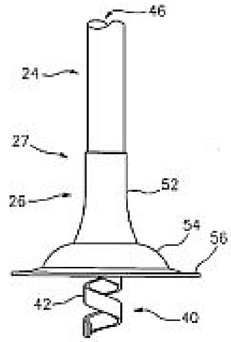
46
24
27
52
26
54
56
42
40

FIG. 3

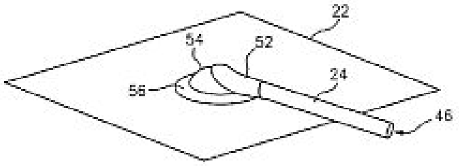
22
54
52
56
24
46

FIG. 5B

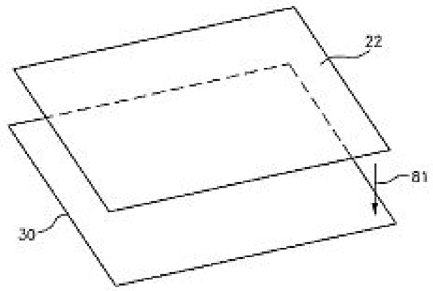

FIG. 6

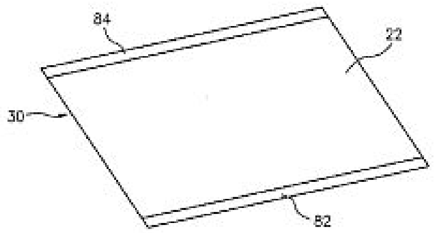

FIG. 7

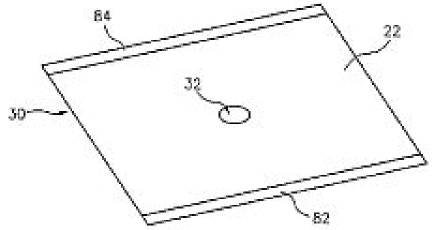
84
22
32
30
82

FIG. 8

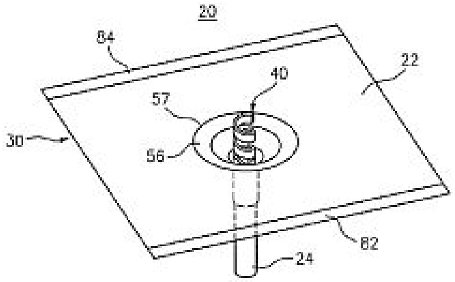
20
84
40
22
57
30
56
24
82

FIG. 10

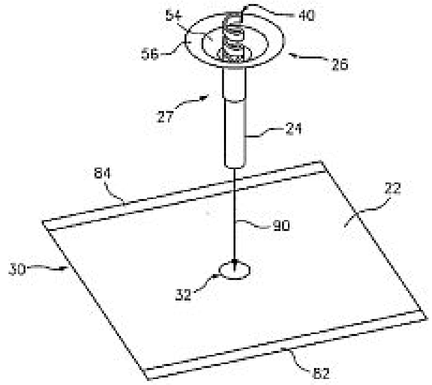
54
40
56
26
27
24
84
22
90
30
32
82

FIG. 9

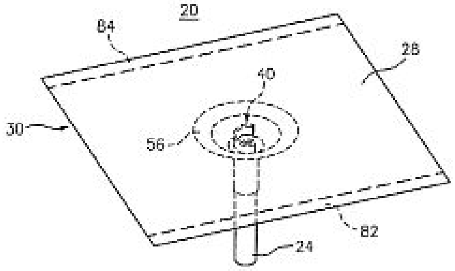
20
84
28
40
30
56
82
24

FIG. 11

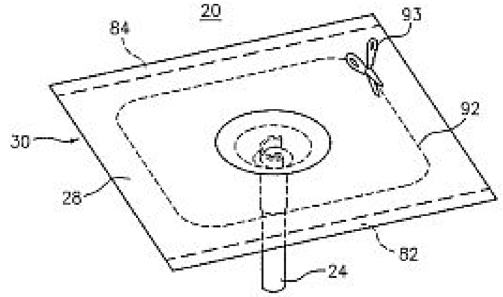
20
84
93
92
30
28
82
24

FIG. 12

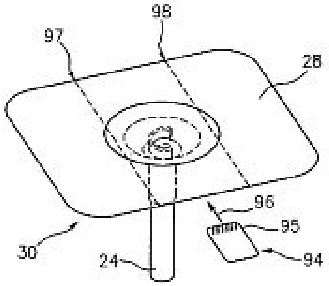
97
98
28
96
95
30
24
94

FIG. 13

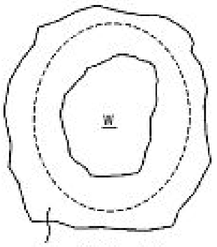

FIG. 14

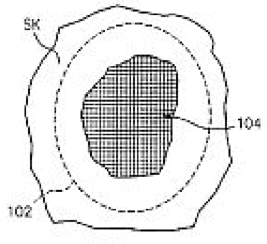

FIG. 15

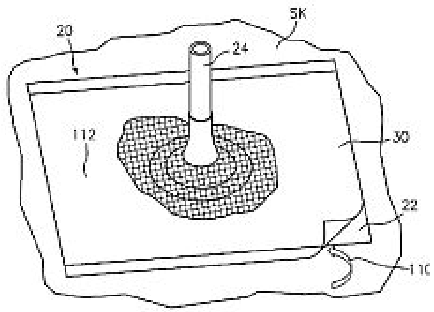
20
SK
24
112
30
22
110

FIG. 18

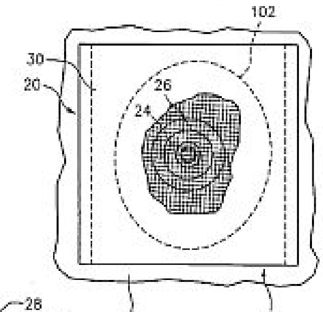
102
30
20
26
24
SK
22

FIG. 17

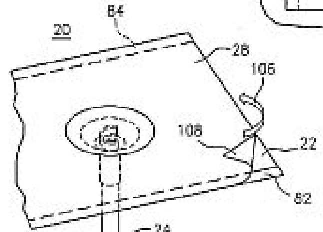
84
20
28
106
108
22
82
24

FIG. 16

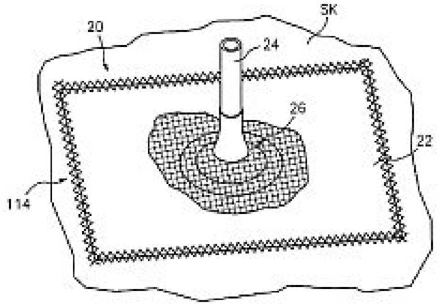
20
SK
24
26
22
114

FIG. 19

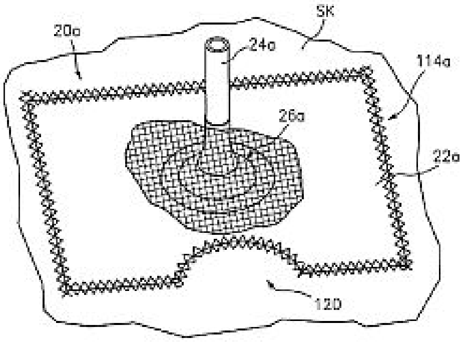
20a
SK
24a
114a
26a
22a
120

FIG. 19A

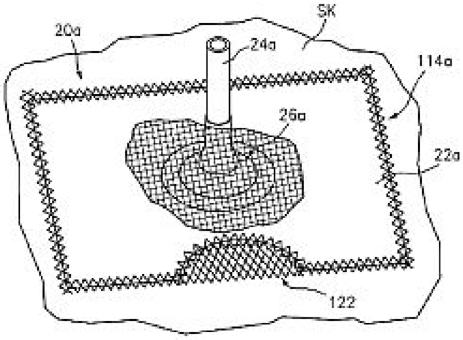
20a
SK
24a
114a
26a
22a
122

FIG. 19B

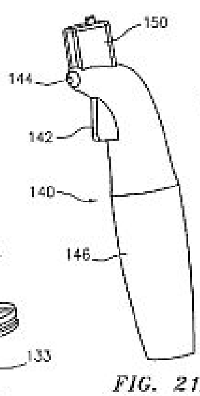
150
144
142
140
146

FIG. 21A

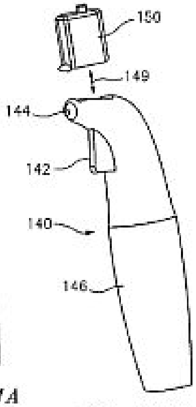
150
149
144
142
140
146

FIG. 21B

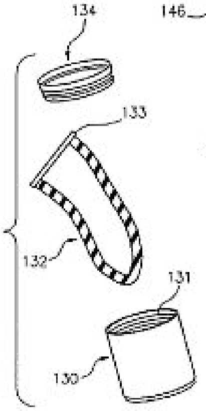
134
133
132
131
130

FIG. 20

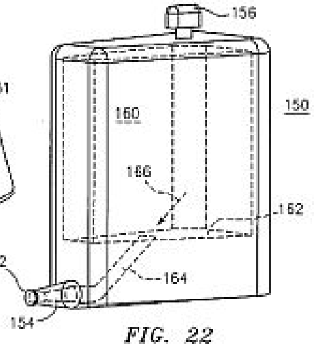
156
160
150
166
162
152
164
154

FIG. 22

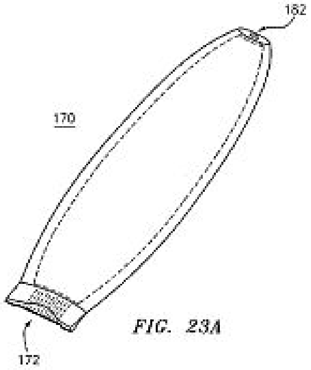

FIG. 23A

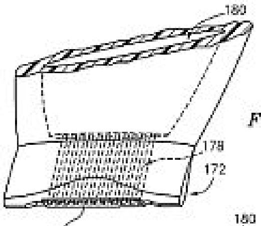
180
178
172
174

FIG. 23B

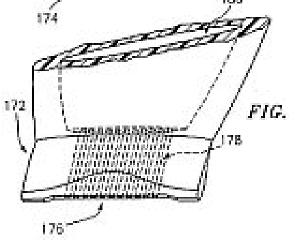
180
172
178
176

FIG. 23C

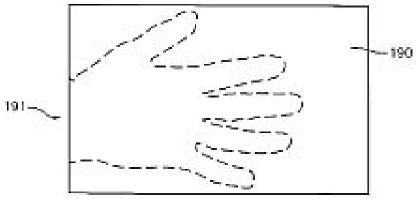

FIG. 24

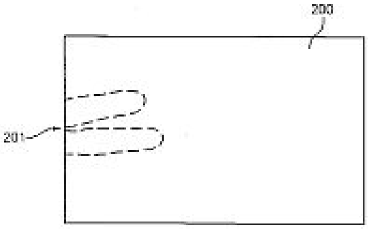

FIG. 25

13/19

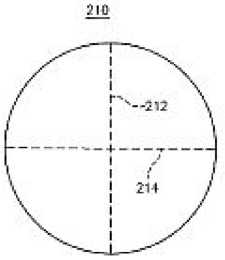

FIG. 26

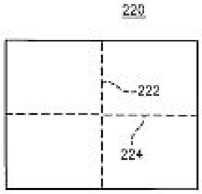

FIG. 27

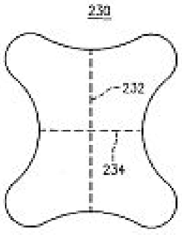

FIG. 28

14/19

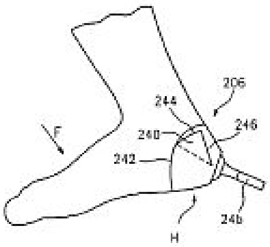
206
244
246
240
242
F
H
24b

FIG. 29

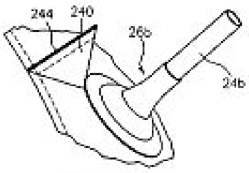
244
240
26b
24b

FIG. 30

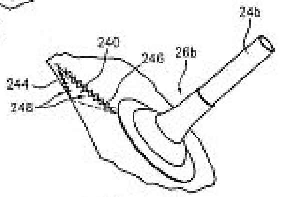
24b
240
246
26b
244
248

FIG. 30A

15/19

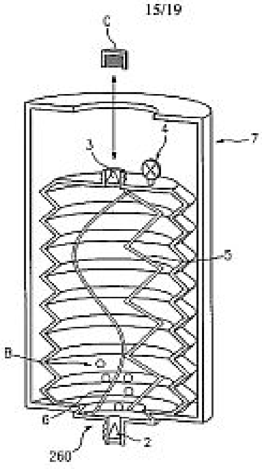
C
3
4
7
5
B
6
260
2

FIG. 31

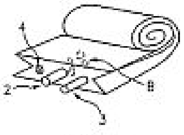

FIG. 32

16/19

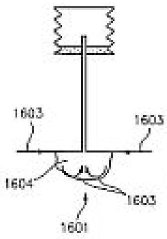
1603
1603
1604
1603
1601

FIG. 34A

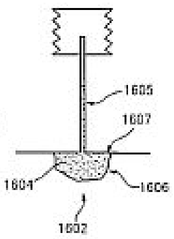
1605
1607
1604
1606
1602

FIG. 34B

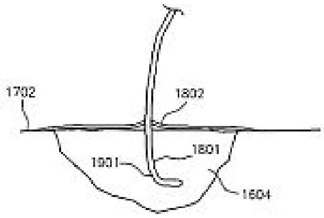
1702
1802
1801
1901
1604

FIG. 36

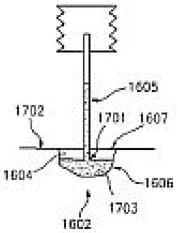
1605
1702
1701
1607
1604
1606
1703
1602

FIG. 35

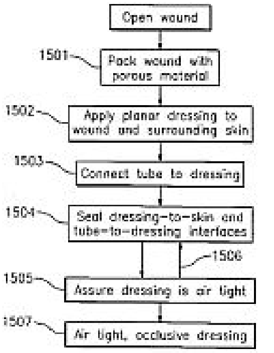
Open wound
1501
Pack wound with porous material
1502
Apply planar dressing to wound and surrounding skin
1503
Connect tube to dressing
1504
Seal dressing-to-skin and tube-to-dressing interfaces
1506
1505
Assure dressing is air tight
1507
Air tight, occlusive dressing

FIG. 33

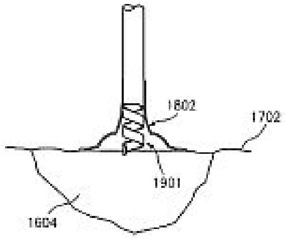
1802
1702
1901
1604

FIG. 37

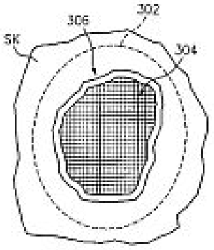
306
302
SK
304

FIG. 38

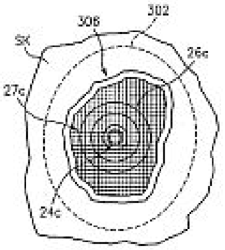
308
302
SK
26c
27c
24c

FIG. 40

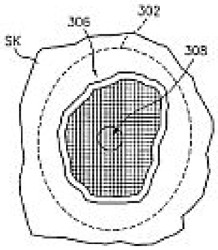
306
302
SK
308

FIG. 39

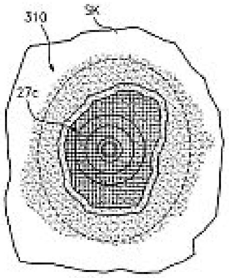

FIG. 41

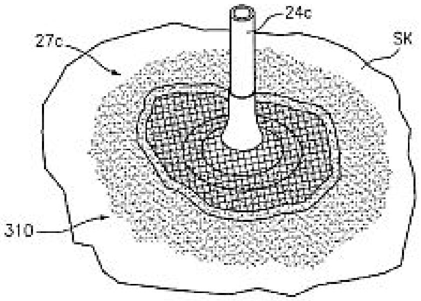

FIG. 42

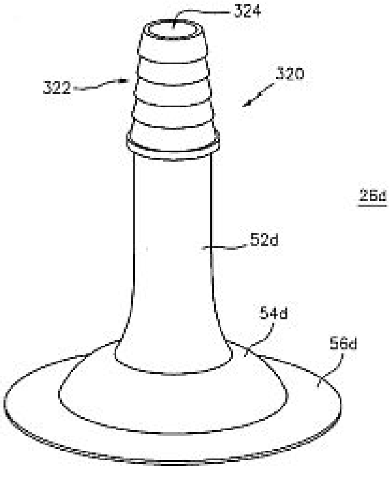

FIG. 43

DETAILED DESCRIPTION

[068] This invention may be accomplished by a kit, dressing system or method utilizing a drape formed as a thin sheet of an organic, preferably elastomeric material, substantially impervious to fluid transfer of air and bodily fluids for preferably at least 48 hours, more preferably at least 72 hours, having first and second surfaces. Preferably, a biocompatible adhesive is disposed on, applied to or contacted with, at least the first surface of the drape. In a number of constructions, a first removable liner sheet covers the first surface of the drape and, optionally, a second removable liner sheet covers the second surface of the drape. The invention further utilizes a container of at least one sealant component that is capable of being delivered as a sealant in a liquid state at pre-selected ambient conditions, the sealant as delivered being at least partially cross-linked at least after one of drying and curing, and which is capable of at least one of drying and curing within thirty minutes, preferably within twenty minutes and, more preferably, within ten minutes after application of the sealant as a layer to the edges of the drape after the drape is applied to the skin surrounding the wound.

[069] The occlusive dressings presently disclosed address the power/mobility and air leak issues by eliminating the need for an electrical power source and by maintaining reliably air-tight interfaces, particularly at 1) the dressing and the skin and 2) the tube and the dressing. The disclosed dressing systems and their connection methods allow for reliable, mechanical NPWT systems. Not only does this eliminate patient mobility and battery management issues, but it also allows NPWT to be administered in austere environments, where electricity is often scarce and harsh environments require robust products. Multiple disclosed embodiments support an inexpensive, robust therapy method for global application. Additionally, dressings according to the present invention are MRI- compatible.

[070] In order to obtain an air-tight skin dressing, the present occlusive dressings use a liquid sealant. This liquid sealant may dry and cure fast, even immediately or effectively immediately, upon application to the skin or other dressing components, into a continuous, occlusive film or sheet of material. The drying and curing processes may occur simultaneously, may be driven by evaporation, may require a curing agent and/or accelerator, and/or may be enhanced or controlled with a curing agent and/or accelerator. Any extra

additives (e.g., curing agents and accelerators) may be added just before, during, and/or after the sealant application process, depending on its chemical reaction with the sealant and its rate.

[071] The liquid sealant bonds to the component(s) that it is meant to seal. The ability of Van der Waals forces to provide the bond strength without an added adhesive is based on the material and its thickness. Theoretically, the debond toughness (strength of the bond) must be greater than the debonding energy, and the debonding energy is proportional to: the thickness of the material, the strain in the material squared, and the elastic modulus of the material. Specifically (on a first order basis; as its basis is a small strain analysis), the bond strength of a thin film must abide by Equation 1, where [Gamma] is the debond toughness, [sigma] is the debonding energy, [Omega] is a dimensionless prefactor, h is the thickness of the film, [epsilon][tau] is the strain in tension, and Ef is the elastic modulus of the film, in order to maintain adhesion to the skin in tension:

[Gamma] > [sigma] = [Omega]1[iota][epsilon][tau][3/4](1)

Therefore, a highly elastic, thin film presents the ideal material properties for reduced, required adhesion strength, increasing the functional applicability of the Van der Waals forces.

[072] An additional adhesive, such as a silicone-based, latex-based, or acrylic- based glue, having one or more components, might be employed to produce the desired bond strength (for example, Liqui-Tape Silicone Adhesive, Waterproof by Walker Tape Co., West Jordan, UT). This adhesive can be applied under the liquid sealant or chemically mixed with the liquid sealant prior to its application, depending on its chemical make-up and final mixing properties. When applied under the sealant, the adhesive may need to become tacky (a.k.a., applied set time) prior to sealant overlay. A fast-setting, two-part sealant that is mixed prior to use may be useful in some circumstances, such as Skin Tite(R) silicone available from Smooth On, Easton, Pennsylvania, which is ACMI Certified Safe and may be used by itself or mixed with a thickener, such as Thi-vex(R) thickener, also available from Smooth On. A polymer sealant, or other material with the ability to bond into a continuous occlusive sheet, with adhesive-like properties due to high Van der Waals forces may be desirable, where no additional adhesive is needed.

[073] Rubber polymers, such as latex, synthetic rubber, and hypoallergenic latex, are examples of polymers with desired properties for both the dressing-to-skin and tube-to- dressing interfaces. For example, Deviant Liquid Latex from Deviant, a subsidiary of Envision Design, San Jose, CA and Liquid Latex Fashions Body Paint from Liquid Latex Fashions, Warrington, PA were both demonstrated to seal the dressing at both dressing interfaces. The drying and curing time for the latex was significantly reduced by applying the liquid to the skin with an atomization process, which is further disclosed in the sections below, by adding alcohol, which helps to absorb the water that evaporates from the latex, and/or by flowing a gas across the sealant for convection drying. For most applications, the curing/drying time was lowered to immediately (at most 1 minute) from the 5-10 minutes previously stated by Deviant at http://www.liquidlatex.net/.

[074] Examples of suitable latex materials include Vytex Natural Rubber Latex (NRL), a brand of natural rubber latex produced and marketed by Vystar Corporation, Duluth, GA. Vytex is manufactured by Revertex Malaysia and distributed by Centrotrade Minerals and Metals, Inc. Protein test results show that Vytex NRL typically has 90% fewer antigenic proteins than Hevea natural rubber latex. Therefore, Vytex causes less exposure and developed latex sensitivities. The Vytex has two products with different levels of ammonia; ammonia is a stabilizer and preservative, and both functionally are feasible for the NPWT liquid sealant and drape components, although alternative stabilizers to ammonia may irritate the skin less. Liquid latex for body painting typically contains ammonia, which is what has been applied to patients during field studies with no irritations. Vytex NRL, low ammonia compound, has provided functional, occlusive drape and sealant components on clean, unwounded skin in a lab setting.

[075] Yulex Corporation, Phoenix, AZ creates hypoallergenic latex from guayule (Parthenium argentatum). Yulex's guayule biorubber emulsions and solids have none of the sensitizing antigenic proteins found in traditional Hevea latex and is considered a safe alternative for people with Type I allergies. Yulex's biorubber emulsions are registered with the Personal Care Product Council and its INCI name is Parthenium argentatum Bark Extract. This is a presently preferred material for the NPWT dressing and sealant, in order to provide a non-allergenic material option. Yulex presently has ammonia and ammonia-free options.

[076] Synthetic materials such as nitrile rubber and neoprene are alternatives to natural rubber that do not have allergy-provoking proteins, but can also generally have poor elasticity with higher risk of break rates and viral penetration rates. Therefore, they are less ideal for many of the dressing applications according to the present invention, but may be suitable in some circumstances, particularly for the drape for which curing on the skin and drying time are not issues. Other multi-part materials, such as Room Temperature Vulcanizing silicones and certain polyurethanes which are two-part materials with base and curative components, may be acceptable in some applications.

[077] Extremely low stiffness, which is achievable with many rubber-type materials, increases its bonding ability through Van der Waals forces alone. The high elasticity capable of being achieved using rubber polymers accommodates for the high levels of tensional strain reached at the skin surface during large deformation body movements. Additionally, the material properties of rubber polymers may also accommodate for the tendency to buckle when compressive strains are applied, depending on any initial interface crack sizes and adhesion strength. A desirable sealant accommodates for the large variability over time and surface area of the skin surface strains experienced during large deformation human motions; in the literature, the maximum large deformation strain is indicated to be approximately 0.45 in tension and 0.3 in compression. As rubber mechanical properties are sufficient to achieve structural integrity, the Van der Waals adhesive properties determine the applicable occlusive sealants, and depending on the polymer, an additional adhesive may be necessary.

[078] The liquid sealant should have viscosity and curing properties, preferably including minimal shrinkage, that enable it to conform to all contact surfaces during the application and curing processes, such that no air leak channels at the interface are present after its application. At the dressing-to-skin interface, the sealant should conform to the folds and creases in the skin that are often bridged when applying a standard, planar wound dressing. These types of bridged cracks at all component interfaces are often a significant source of air leaks into the system without a liquid sealant. Once a crack exists, crack propagation occurs in tension and compression with reduced, applied strains, so air leak channels can form overtime with reduced strain magnitudes. Therefore, eliminating any initial cracks at all of the interfaces is desirable. At the dressing-to-skin interface, structures, such as hair, often create opportunities for crack propagation and air leaks into a wound dressing, and therefore, hair is often shaved before dressing applications. The need to shave the hair from an infectious standpoint is not desirable, as the shaving process creates trauma at the hair follicles and increases the risk of infection. With a liquid sealant, these structures can be completely enclosed in the air-tight sealant, and therefore, are not a source of crack propagation under the sealant and do not typically require removal prior to the sealant application, as cracks at the dressing edges are most critical to seal, in order to resist crack propagation due to tension. In some constructions, adhesive on the first surface of the drape is sufficiently thick and/or flowable to seal around hairs and skin crevices and to minimize crack propagation.

[079] The sealant thickness, number of components, wound location, and sealant viscosity determines the optimal sealant application method(s). The liquid sealant may have a very high to low viscosity, as long as it can completely wet the contact surfaces. If mechanically applied (e.g., brush or "painting" application, roller application, sponge/dabbing application, squeegee or other squeeze-type application, application by-hand (i.e., finger) with or without a non-stick cover, etc.), a viscosity that avoids run-off due to gravity is preferable in order for the sealant to be ergonomically applicable to any wound location. This leads to higher viscosities and is limited at the low viscosity range. Painting is not the preferred application method; when painting the sealant, it is difficult to achieve a constant thickness. If the thickness varies significantly over its surface area, the mechanical properties and debonding energy will also vary significantly, which may cause occlusive dressing failure. Painting also has other drawbacks, as it is easy to trap air bubbles in the sealant, which are a source of cracks for crack propagation. Also, it is difficult to produce and maintain a very thin coat, which significantly increases the necessary Van der Waals bonding strength; it increases the stiffness of the final dressing and decreases its ability to conform to large tissue strains.

[080] Spraying is a preferred method of applying the sealant. Two types of spraying procedures are possible: 1) an aerosol process which propels the liquid sealant with a pressurized liquid or gas propellant that forces the liquid sealant through an atomizing nozzle, and 2) a shearing process which shears the liquid sealant with a pressurized gas or liquid causing atomization. When atomized, the layer of sealant material can be made thin enough that run-off is less of an issue, and therefore a range of lower viscosities can be used for their desired wetting characteristics. Additionally and in combination, the small atomized particles

fill in the structures on the skin for wetting purposes. The spraying technique is limited at the high viscosity range, as too high of a viscosity sealant will not be capable of being sprayed with reasonable pressures and velocities for application in the clinical setting onto skin. However, this is not seen as a negative aspect since liquids with very high viscosities often do not properly wet the complex contours of the skin surface.

[081] The shearing process may be preferred over the aerosol process. One reason for this preference is that nozzles clog easily with long polymer chains, unless the liquid can be further thinned. Using the shearing process, the shearing fluid and sealant fluid may be kept separate until they are both external to the nozzle head. Therefore, internal clogging of the nozzle does not occur when properly designed, including a fluid filter (if necessary) and the proper nozzle orifice size. Gas is the preferred shearing fluid, as it does not add additional liquid to the system for drying purposes, it is easy to propel since it can be compressed to high pressure levels, and it helps to dry the sealant when spraying it onto the skin. Higher viscosities and materials with long polymer chains are capable of being sprayed by the shearing method rather than the propellant method, although the viscosities and chemical chains that can be accommodated with the propellant method can be increased with complex nozzle design.

[082] Additives such as curing agents, accelerators, convection drying agents, and adhesives may be applied via separate application methods, if they are not mixed with the sealant prior to application. Their application method may be via painting or spraying. The application of these additive components and the sealant may occur in a multi-step process. They may be stored and applied from separate containers with the same or different application methods in series or in parallel. However, they may also be applied in parallel or series from the same containing body. One example is a parallel spraying process, for which three ports exist: the sealant port, the shearing fluid port, and an accelerator port; these three components can combine during the atomization process in the spray nozzle where the three ports may interact. Another example is a spray apparatus that allows the amount of sealant (and potential accelerator) to be controlled, such that it may be shut-off; the shearing gas then becomes a convective drying gas.

[083] Various polymers with rubber-like properties were determined to have the desired sealant properties. Additionally, a preferred sealant cures immediately or within a few seconds after surface contact. With these characteristics, the polymer tends to have long and heavy polymer chains, and therefore, requires the higher atomizing forces capable with the shearing process. As gas is used to atomize the polymer, there is a desirable range of gas pressure, velocity, and volume flow combinations that are required for the desired, continuous-film outcome. Filtered air, pure oxygen, and carbon dioxide are examples of applicable shearing gases that can be readily used, and are often available in the clinical setting at the desired pressures and volume flow rates. They are also readily available outside the clinical setting. Using these gases, the necessary, shearing atomization process is capable of being designed into a miniaturized handheld device. This process and design is similar to the consumer use of the aerosol embodiment commonly found in consumer products and is further disclosed in the Dressing

Application Methods section.

[084] The thickness of a desired seal embodiment can be built-up in a successive layered, lamination process. A material that has a strong affinity for itself with either strong Van der Waals forces or chemical bonds that form between its layers, such that the final material behaves as a continuous one-layer sealant is desirable. The desired thickness is the minimal thickness needed for strength and to achieve the desired occlusive properties, which is material dependent. This thickness is often thinner than the thickness that can be reliably and uniformly achieved through a painting process, and therefore a spraying process is often preferred. The atomization process provides a method to achieve the thinnest functional sealant thickness.

[085] Occlusive dressings are beneficial beyond NPWT and in combination with advanced NPWT features. Some proven benefits of occlusive properties are highlighted here. The occlusive characteristic may enhance the penetration and absorption of topically applied medications, such as ointments, powders and creams, which can be beneficial in combination with standard wound dressings and with therapies, such as NPWT. The V.A.C. Instill Therapy Unit (KCI) was meant to combine instillation therapy with NPWT. Instillation, as defined by the V.A.C. Instill documentation, includes both: 1) the introduction and removal of topical solutions in liquid form and 2) the ability to flush out and clean a wound through a rigorous irrigation technique. The main caregiver complaint about this and other instillation-purposed dressings is that they often leak liquid during the instillation process, especially during a rigorous irrigation procedure,

which further induces air leaks during continued therapy. The occlusive seal and dressings disclosed in this disclosure would solve any leak issues that arise. Often the irrigation process introduces leaks by propagating cracks in the dressing; by eliminating these cracks, the sealant and dressing techniques in this disclosure significantly reduce the potential for leaks and leak formation during instillation. The port(s) needed for instillation fluid insertion and removal can be directly connected to the disclosed occlusive dressing embodiments with the same tube-to-dressing connection methods that are disclosed in the Tube-to-Dressing Interface section in this disclosure.

[086] Although the presently disclosed occlusive dressings were developed with NPWT system in mind, they can be used for any application for which an occlusive (a.k.a., air tight and water tight), air tight, or water tight seal to the skin is desirable. Therefore, they are applicable in multiple fields beyond NPWT, and more generally in the field of skin sealants and their methods. Truly occlusive dressings create a control volume over the area of tissue that they are applied, which is a desirable feature for multiple applications, many which are disclosed in this application document.

[087] The occlusive dressings discussed in this disclosure are the first skin dressings to provide a control volume, as no other dressing to-date is proven to be (reliably) truly occlusive. This would benefit the enhancement of advanced healing therapies that are sensitive to any variation in the environment, such as stem cell based therapies, for which complete control of the environment is necessary to achieve deterministic results. If a specific air leak is desirable, its rate can be precisely controlled into the control volume through precision valves. Currently, there is no accurate predetermination for the air leak rate into any wound dressing, especially since most dressing air leaks have variability over time and with body movement. Furthermore, truly occlusive dressings may be used in in vivo acute toxicity tests of dermal irritation and sensitization. The test animal is shaved and the test material is applied to the skin and wrapped in an occlusive material. The skin is then exposed after 23 hours and an assessment for redness and edema is made; this assessment is repeated 48 hours later. [088] FIG. 1 is a schematic expanded perspective view of a dressing assembly 20 including a drape 22, a novel flange 26 and a tube 24 with first and second protective liners 28 and 30 prior to application of a liquid sealant according to the present invention. Drape 22 and second liner 30 define holes 32 and 34, respectively, through which tube 24 is insertable.

[089] FIGS. 2 and 3 illustrate a novel first end 40 of the tube 24 of FIG. 1 being inserted through the novel symmetrical flange 26 to form a tube assembly 27, FIG. 3. First end 40, also referred to as distal end 40, includes a spiral extension 42 which, in one construction, is formed by making a helical cut into the distal end of tube 24. In other constructions, a separate component having a helical shape or other geometric shape is attached to serve as a deterrent to clogging of distal end 40. Spiral extension 42 minimizes possible blockage of lumen 46 through tube 24. Another anti-blockage construction is illustrated in FIG. 4 with a tube distal end 60 defining perforations 62 and 64, notches 66 and 68, and a blunt tip 70. Perforations 62 and 64 can be formed as diamond-shaped cutouts, circular holes, or other geometries. A notch 44, FIG. 2, is also created in this construction to further minimize the possibility for lumen 46 through tube 24 to become obstructed, as described in more detail below.

[090] Arrow 48, FIG. 2, represents distal end 40 being inserted through passage 50 in flange 26 defined by a sleeve region 52, a rotation region 54, and an adhesion region 56 having an outer edge 57. Sleeve region 52 is adhered to tube 24, as described in more detail below, at a final location such as shown in FIG. 3. In another construction, sleeve region 52 is attached, via adhesion, welding or other air-tight connection process, to a short piece of tube with a connector that connects to a longer piece of tubing. In yet another construction, the flange includes an integral connector capable of mating with flexible tubing such as shown in FIG. 43. Rotation region 54 serves as a flexible ball joint in this construction. As depicted in FIGS. 5A and 5B, tube 24 can be manipulated in the direction of arrow 72, FIG. 5A, to a desired side orientation as shown in FIG. 5B. Lumen 46 through tube 24 remains open in some constructions because distal end 40 extends beyond adhesion region 56, with notch 44 preferably below sleeve region 52 but above adhesion region 56, so that rotation region 54 does not collapse onto itself. In other constructions, the flange is sufficiently short and wide to minimize the possibility of pinching closed. Sleeve region 52 is sealed in an airtight, fluid-impervious manner in some constructions by first applying a silicone adhesive, such as Liqui-Tape adhesive from Walker Tape Company as mentioned above, to the portion of the outer surface of tube 24 which will be brought in contact with sleeve region 52 during assembly. A liquid sealant can be applied to the junction of tube 24 and flange sleeve region 52 to further occlude possible fluid escape at that junction.

[091] In some constructions, flange 26 is manufactured directly onto tube 24, via a dipping, molding or spraying process. In constructions where flange 26 is constructed entirely from, or coated with, a material that has an affinity for itself, sleeve region 52 may self-adhere to rotation region 54 and adhesion region 56, to the extent that region 56 is exposed, when folded against itself as shown in FIG. 5B. Where the material forming the exterior of flange 26 has an affinity for the material of drape 22, especially for materials containing latex compounds, the exterior of sleeve region 52 will also adhere to drape 22 at least to some extent; latex-type material applied to the surface of tube 24 will further enhance this adhesion. Fixing the tube 24 into a fixed orientation such as shown in FIG. 5B may be especially beneficial for bed-ridden or less mobile patients so that the tube can be positioned to avoid the patient lying on the tubing for long periods of time or to avoid compromised areas around the wound. In other circumstances where the tube remains movable, it can be easily repositioned because rotation region 54 remains flexible and the tube can be monitored and moved frequently to assure that tissue is not degraded from lying on the tube in one position for an extended period. Especially for active patients, the tube 24 can be periodically re-positioned by the patient or by a healthcare professional.

[092] FIGS. 6 and 7 show a drape 22 being covered by an upper liner 30 to manufacture a dressing according to the present invention. Preferably, drape 22 has a thickness ranging from 2 microns to 0.4mm, especially in portions which will be applied to skin; a greater thickness in the center portion to be located over a wound is less critical for occlusivity. In some constructions, adhesive is pre-applied on the upward-facing surface shown in FIGS 6 and 7, which will be placed in contact with skin during use; in other constructions, adhesive is also placed on the opposite side of drape 22, to be covered by liner 30, as indicated by arrow 81 in FIG. 6, for storage and handling. The adhesive is applied as a uniform coating in some constructions and, in other constructions, as concentric circles or other non-uniform pattern. Preferably, liner 30 has extensions 82 and 84 which extend beyond the drape 22 to facilitate handling of the dressing without touching any adhesive, and to enable easy removal of the liner 30 from the drape 22 after placement on a patient.

[093] FIG. 8 shows a hole 32 punched in both layers of the dressing of FIG. 7.

Hole 34, FIG. 1 , is not visible in FIG. 8.

[094] FIGS. 9 and 10 shows a tube assembly 27 being inserted, arrow 90, onto the dressing of FIG. 8 with the adhesive region 56 to edge 57 of the flange 26 being sealed to the drape 22 utilizing the pre-applied adhesive. Additional adhesive or sealant can be added around edge 57 or pre-applied to region 56 as desired.

[095] FIG. 11 shows a protective liner 28 being added to the dressing 20 of FIG.10. Protective liner 28 protects the skin-side adhesive, when pre-applied, until liner 28 is removed as illustrated in FIG. 16 below. Distal end 40 with anti-clogging feature 42 is sufficiently relaxed and short in length to be contained under liner 28. Liner 28 preferably extends beyond drape 22 over regions 82 and 84 of liner 30.

[096] FIG. 12 illustrates how a user can cut the dressing 20 of FIG. 11, along dashed line 92 using scissors 93 for example, to conform to a wound.

[097] FIG. 13 shows a handling tab 94 being added, arrow 96, to the dressing 20 of FIG. 12, which is especially useful if liner extensions 82 and 84 are cut away. In this construction, tab 94 is attached by adhesive 95 to one of liners 28 and 30 to assist removal of the selected liner. Additionally or as an alternative, perforations 97, 98 create locations for easy removal of the liners. If perforations are utilized, it is preferred that the top and bottom liners have perforations that are aligned along different angles. The preferred angle difference is substantially perpendicular, that is, at about ninety degrees offset. Perforations are preferred when extensions 82, 84 are not provided. The dressings are very difficult to handle with medical gloves on, which are required for sterility. Therefore, handling tabs eliminate the need for the clinician to touch the adhesive. This is also desirable since powdered gloves tend to cause the adhesive to adhere to the powder and loose its adhesion properties. Another solution would be to provide non-stick finger or hand covers, such as shown in FIGS. 24 and 25, similar to the finger sealant applicator shown in FIG. 20. This is especially important when dressings are re-shaped, potentially cutting off handling features, and when the user is removing the top protective liner and folding down the top folds. Ideally, additional handling components are built into the packaging components or protective liners. Such as, the package that

the drape comes in, turns inside out to form a sterile, handling glove, or the bottom liner is used to maneuver the higher-level of adhesive interactions when dealing with the top liner. The bottom liner may have a cut-out (pre-perforated) glove, FIG. 24, with the non-stick side, e.g. silicone-coated, initially facing the adhesive; preferably, a non-stick coating is provided on both sides for both right- or left- handed application.

[098] FIGS. 14 and 15 illustrate debriding an open wound W and cleaning the wound cavity and surrounding skin SK, preferably at least 3 cm in width as indicated by dashed line 102, with standard cleaning methods such as with alcohol and gauze wipes. Typically, the next step is to pack the open wound W with fluid pervious material 104 such as gauze, open-cell foam or a sponge.

[099] FIG. 16 is a perspective view of the underside of the dressing 20 of FIG. 11 with the liner 28 being removed as indicated by arrow 106, such as by pulling on corner 108, to expose drape 22 with pre-applied adhesive.

[0100] FIG. 17 is a schematic top plan view of a dressing 20 according to the present invention attached via adhesive drape 22 to skin SK surrounding the wound. When negative pressure therapy is desired, a source of negative pressure is connected to tube 24 such that its lumen is in communication with the wound cavity.

[0101] FIG. 18 is a schematic perspective view of the dressing 20 of FIG. 17 with the upper protective liner 30 being removed, as indicated by arrow 110. Dashed line 112 represents a perforation or pre-cut line to assist removal of liner 30 without sliding it over tube 24.

[0102] FIG. 19 shows liquid sealant 114 being applied to the edges of the drape 22 of FIG. 18. The preferred sealant embodiment has as width of 2-3cm and is centered over the edge of the drape 22.

[0103] If the dressing is applied to contoured surfaces on the body, such as described below in relation to FIGS. 29-3 OA, folds in the planar dressing may be necessary for adhering to the surface of the skin, in order to match the surface contour. These folds typically travel from the outer edge towards the tube 24. In this situation, the preferred application method is to minimize the number of folds by creating a few large folds. Preferably, there are no more than four folds, divided substantially equally around the periphery of tube 24. These folds are created when adhering the drape to the surface of the skin, forming a "T". Then, when the top protective liner is removed, the folds are adhered to the surface of the drape with the adhesive on the top of the drape. Preferably, the folds form individual triangles on the top surface of the drape. The folds are then pressed to lie flat and be completely adhered to the surface of the drape. Sealant is then applied to the edge of each fold to seal off the area between the fold and the top of the drape from the surrounding environment. This additional sealant preferably connects with the sealant placed around the outer edge 114, FIG. 19, for example. Preferably, the additional sealant is applied at substantially the same time as the original sealant with the same sealant material. Any drape material that would extend onto the skin, beyond the original edge of the drape, when folded preferably is cut off before pressing the folded drape material against the skin to lie within the original edge of the drape. Patterns may be provided on the upper protective liner to direct the user where to put the folds, when needed.

[0104] FIGS. 19A and 19B illustrate modifying the coverage of a dressing 20a according to the present invention, with drape 22a, tube 24a and flange 26a. If the dressing 20a is too small to cover the desired skin area around the wound, preferably by 3 -5cm, for instance if the user cut away too much of the dressing, as shown by cut-out 120, while reshaping the drape for easier application or in order to avoid a complex contour near the wound cavity, the user can use sealant to reconstruct the dressing as shown by additional sealant 122, FIG. 19B. A modified occlusive dressing 20a is thereby achieved. However, it is preferable for the drape 22a itself to cover the entire wound edge, in order to protect the wound cavity from the sealant material.

[0105] FIG. 20 is a schematic expanded view of a vial 130 of sealant, with closure threads 131 to receive a cap 134, with a non-stick finger protector 132, shown in cross- sectional view, preferably with a rim 133, optionally positionable within the vial 130 for storage and transportation.

[0106] FIGS. 21A and 21B show a dispensing apparatus 140 with removable cartridge 150 of liquid sealant. Dispensing apparatus 140 has a finger trigger 142 and a nozzle 144 in this construction and can be powered by a cylinder of compressed gas, such as a C02 cartridge, contained within the housing 146. Preferably, the apparatus is gravity fed. Because there need not be a needle valve, such as found in typical air guns to stop the flow of fluid, an adhesive tab 152, FIG. 22, is initially removed from tip 154, and the cartridge 150 is inserted into the apparatus 140 as represented by arrow 149, FIG. 21B. A plug 156, FIG. 22, is then removed, such as by twisting, to expose an air hole at the top of cartridge 150 and activate apparatus 140 and allow sealant to flow or be sprayed out of nozzle 144. The apparatus 140 can be set aside temporarily, with nozzle 144 directed upwards, between sealant layer applications.

[0107] FIG. 22 is an enlarged perspective view of the cartridge of FIGS. 21 A and 2 IB with internal chamber 160, raised floor 162, and slope 164 in this construction to assist gravity feed of sealant liquid to tip 154, as indicated by arrow 166. In other constructions, a multi-component sealant is delivered utilizing a separate chamber for each component. The components are mixed during delivery in a down-stream mixing chamber or in a mixing nozzle such as the 3M(TM) Scotch- Weld(TM) EPX(TM) Mixing Nozzle currently available from 3M Company, St. Paul, Minnesota. Other multi-component delivery systems can be utilized such as those commercialized by Henkel Loctite Corporation, Rocky Hill, Connecticut. One or more of the sealant components can be a powder or other state as long as the final sealant is delivered in a liquid state, including as liquid droplets via shearing or a propellant.

[0108] FIG. 23A is schematic perspective view of a hand-powered squeeze applicator 170 for liquid sealant.

[0109] FIGS. 23B and 23C are enlarged views of the squeegee-type outlet 172 with and without a removable strip 174 covering the dispensing openings 176 of passages 178 communicating with inner chamber 180. In this construction, a removable tab 182, FIG. 23 A, allows air to enter chamber 180 during delivery of the sealant.

[0110] FIGS. 24 and 25 are schematic top plan views illustrating non-stick gloves 191 and finger covers 201 formed in liners 190 and 200, respectively. These applicators are non-stick, such as by a non-stick silicone coating, on only one side in some constructions and, in other constructions, are non-stick on two sides.

[0111] FIGS. 26-28 are schematic top plan views liners 210, 220 and 230 with indicator lines 212 and 214, 222 and 224, and 232 and 234, respectively, having different shapes for selected locations and contours of a patient. Shorter lines 222 and 234 have priority if folds are needed; a symmetrical shape such as a square or the circular shape of liner 210, FIG. 26, has fold lines of equal priority.

[0112] FIG. 29 is a schematic side view of a dressing 20b according to the present invention being applied to the heel H of a foot F. Flange 26b is positioned with tube 24b communicating with a wound in heel H. One large fold 240 is shown, with edges 244 and 246.

[0113] FIGS. 30 and 30A are enlarged schematic views of the dressing of FIG. 29 with a fold 240 being created and then pressed flat to enhance conformance to the heel. All edges 242, 244 and 246 should be sealed with sealant 248 according to the present invention.

[0114] There are multiple different methods of using the sealant described in this disclosure at the dressing-to-skin interface. The first method is to use the sealant in conjunction with current, commercial skin dressings (or dressings with similar embodiments), in order to achieve occlusive properties. In order to do this, the dressing is first applied to the skin, step 1502, FIG. 33, after the wound is packed, step 1501 ; typically, the dressing (a.k.a., drape component) is a planar adhesive tape form. The drainage tube may enter into the dressing at the dressing-to-skin interface, or it may have its own connector that requires an incision into the dressing above the wound cavity, step 1503. The dressing system is applied with its recommended procedure. Then, all dressing-to-skin interfaces are sealed with the sealant and potentially additional adhesive, step 1504.

[0115] At the dressing-to-skin interface, the sealant contact with the skin should be biocompatible. The sealant should conform to and seal off the folds and creases in the skin, which are often bridged when applying a standard, planar wound dressing. These cracks are a significant source of air leaks into the system without a liquid sealant with the proper wetting properties. The proper wetting properties are achieved by applying the liquid sealant directly to the skin and dressing in its liquid form through a painting process or through spraying the liquid with an atomization process that eliminates liquid run-off and that may achieve a more uniform, thin film.

[0116] Once a crack in the planar dressing exists, crack propagation may occur in tension and compression with reduced, applied strains. Therefore, sealing any initial cracks in the dressing-to-skin interface is desirable. Also, properly sealing the dressing-to-skin interface at the edge of the dressing deters any air leaks from future crack propagations, as the sealant hinders the propagation from reaching the outside environment. If an additional adhesive is used between the sealant and dressing-to-skin interface, then the adhesive should adhere to the skin, dressing, and sealant to form the necessary bond strength. The adhesive or its applied components should also conform to the folds and creases in the skin and/or dressing. The adhesive should be compatible with the skin, dressing, and sealant when applying the adhesive under the sealant, or when mixing the sealant with an adhesive component prior to application.

[0117] Use of the liquid sealant can permit elimination of the current commercial dressings (or similar dressing embodiments; a.k.a., the drape component). The liquid sealant can be applied directly over the wound cavity and wound packing material. In some embodiments, the packing material may require an additional liquid tight barrier if the liquid sealant can be absorbed into the packing material. Additionally, a liquid tight barrier may need to exist at the interface between the packing material and the wound edge, as the sealant could potentially leak into this barrier, depending on the application technique of the packing material, which may not be desirable. A gap at the interface between the packing material and the wound edge may be disruptive to the sealant in creating a continuously occlusive film, or the potential of the sealant contacting the inside tissue of the wound cavity may need to be eliminated. These barriers may be of an occlusive nature; in this case, the sealant should be applied at any of their non-occlusive edges; however, the sealant may also cover the entire surface area, which may help to maintain the adhesion of the barriers. The barriers can be made of multiple materials from adhesive and non-adhesive polymer films to clays and pastes, for example. Barriers mentioned in this description are different from the standard wound dressings, as the standard wound dressings' adhesion to the skin forms structural and adhesion integrities of the dressing-to-skin interface, and the barriers currently discussed are used to protect the wound from the sealant component and are not necessarily intended to provide any structural support beyond that purpose.

[0118] Maceration of the skin under a truly occlusive dressing may be of concern to the caregiver. This can be solved with a material selection solution, as a one-way, directional occlusive sealant material can be used that allows the skin to breathe and its moisture to evaporate without letting air into the system. Similar material properties are commonly found today in materials used for sports apparel. Additionally, this can be solved from a design perspective. The sealant application area can be made narrow enough that the moisture of the tissue under the dressing can diffuse around the seal. If a larger surface area of seal adhesion is necessary, a web of sealant can be applied to allow diffusion around the webbing. Additionally, the sealant can vary in thickness via the atomization process, where a thick enough dressing for occlusive properties is sprayed around the wound edge or dressing-to-skin interface. This application can maintain a narrow width, and then the rest of the dressing can be made into a thinner layer that is breathable based on a different number of lamination layers or by using different spraying variables and techniques. This thinner part of the dressing can maintain a continuous film embodiment with the occlusive barrier, as the debonding energy of the thinner part is significantly decreased due to the reduction in thickness, increasing the effective bond strength. Additionally, this breathable component can be webbed over the surface, instead of encompassing a continuous film embodiment.

[0119] The tube-to-dressing interface should be sealed if the connection is not prefabricated to be occlusive during its manufacturing process, as it is in the Spiracur dressing. The sealant should bond to both materials found at the tube-to-dressing interface and form an occlusive seal spanning the interface, step 1504, FIG. 33. Three methods can be used for this sealant interface and its components: 1) dressing components that were not originally prefabricated to be occlusive can be pre-assembled and sealed prior to dressing application (most desirable from an occlusive results reliability perspective); 2) dressing components can be preassembled prior to dressing application, but the seal is applied after dressing application; or 3) the tube connection method is fabricated and sealed to the dressing during or post dressing application.

[0120] The first method provides the user with a method to prefabricate a custom dressing that has an occlusive tube-to-dressing interface. This eliminates many potential air leaks, and for the first time, allows custom, prefabricated, occlusive dressings to be made in the clinical setting. Method two is convenient if the liquid sealant is the same for all dressing interfaces; therefore, all the interfaces (tube-to-dressing and dressing-to-skin) can be sealed in one step after the dressing application. However, this method requires that the pre- assembly configuration is stable during its application, before any sealant is applied. For method three, less prep-work needs to be performed by the caregiver. If this sealant method is ergonomic and repeatable without any prefabrication, then this method can significantly cut-down on dressing time, which is a significant personnel and cost savings for the care center. The ergonomic and repeatable characteristics depend on the tubing connector designs.

[0121] Multiple tubing connector designs can be manufactured for sealing purposes to be used for all three methods. Three basic design concepts can span many embodiments. These three design concepts are:

[0122] 1) Puncture the dressing with the drainage tube, such that the drape fits snuggly against the tube. Then, apply the sealant at the tube-to-dressing interface. With this method, the tube can recess into the wound cavity at a custom length as indicated by extended distal end 1801, FIG. 36. If the adhesion force of the sealant needs to be increased, an additional adhesive can be added under the sealant or mixed with the sealant, or the tubing and/or dressing can be pre-coated with a material that the sealant has an affinity for. In practice, rubberized polymers typically have a strong affinity towards themselves, even if the under layer is previously cured. Multiple drainage holes, as illustrated in FIG. 4, or a spiraled cut pattern, FIGS. 2 and 3, in the portion of the tube extending into the dressing is preferred, in order to prevent the tube from occluding against saturated packing material or with particles in the wound exudate.

[0123] 2) The same concept as in concept 1, except with a different tube entry into the dressing. This concept is for the case where an initial planar dressing is used. Two pieces of the planar dressing cover the wound from two different sides, and they meet above the wound cavity in a "T" joint. The tube is placed through this "T" joint into the wound cavity before the "T" joint is sealed. Then, all of the interfaces are sealed with the liquid sealant.

[0124] 3) The same concept as in concept 1, except at the tube-to-dressing interface, a prefabricated foot 1802, FIG. 36, also referred to herein as a flange, is attached (preferably air-tight) to the tube in order to provide a planar surface to seal to the dressing. In one functional embodiment, the foot 1802 is made of a flexible material that the sealant has a strong affinity for and no additional adhesive is necessary. The material of the foot may be tapered in thickness, such that it thins to meet at its edge(s) with the dressing, which may be more desirable for reliable, occlusive sealant application. The tube can connect to this foot 1802 in many orientations; however, it is often preferable to minimize the dressing profile. However, often when minimizing the dressing profile, the tube is in an orientation that cannot be readjusted after dressing application. Therefore, the tube may connect perpendicular to the skin surface, and by using non-kink tubing and/or the flexibility of the foot 1802 material allows the tube to be oriented in any orientation post dressing application, such that the tube will not kink and occlude itself. For this concept, the tube may not puncture the drape, but instead, the hole (a.k.a., incision) may be pre-cut; the foot should extend beyond the hole (a.k.a., incision).

[0125] 4) The fourth concept is the similar to concept 3, except the tubing does not extend into the wound cavity, FIG. 37. Therefore, an incision is made into the dressing, and the tube opening is positioned over the center of the incision, as in the T.R.A.C. Pad. The foot should extend beyond the incision and is sealed to the dressing with the liquid sealant or occlusively pre-sealed during its manufacture. In this embodiment, the end of the tube should be designed to stop potential occlusion onto the foot, onto packing materials, or with wound exudate substances. Therefore, if the foot is connected to the tube above the skin surface, the end of the tube may have a spiral cut along its length, up to its interface (intersection of the upper portion of foot 1802 and tube distal end 1901 in FIG. 37) with the foot. Additionally, an anti-occluding material may be placed at the end of the tube between the foot and the dressing. This anti-occlusive material may be a large pore, open cell sponge.

[0126] In the tube-to-dressing connection, as with all sealed interfaces, an additional adhesive may be added if the bonding strength needs to be increased. The foot may also be initially adhered with a tape or adhesive to the dressing prior to sealant application. The tube connectors can exist in many similar embodiments to those listed above; however, a limited number of examples are given here in order to illustrate the basic connections and the occlusive dressings. The tube-to-dressing interface may be occlusively pre-sealed during its manufacture. Additionally, the component attached at the interface may only consist of a tube connector (which may or may not contain a segment of tubing) that is additionally connected to a longer piece of tubing that then attaches to the pump. Examples of occlusive tube connectors are barbed connectors that connect directly with a tube, specific connectors that interlock with each other and are required on each end of the connected components, and a compression fit seal such as a cylindrical hole in rubber that the tube can be occlusively pressed into.

[0127] As previously stated, handling a dressing with a planar tape embodiment may cause the adhesive to weaken prior to dressing application. Therefore, specific handling devices for the caregiver can be included with this dressing component. These devices may include non-stick gloves, such as PTFE gloves, FIG. 24, or non-stick fingertips, FIGS. 20 and 25. Handing tabs that extend from the dressing may also be incorporated into the dressing design. These tabs may be a part of the dressing (a.k.a., drape) that are torn-off after applying the dressing, or they may be extensions of a removable backing material that is attached to the dressing as shown in FIG. 13. [0128] For application of the sealant, many application embodiments and methods are possible. For mechanical applications, including painted applications, the applicator embodiment can be a brush, roller, sponge, spatula, or other similar embodiment to apply paint in a "spreading" fashion. These spreading devices can be attached to a container (preferably refillable) of liquid sealant for a continuous feed of sealant to the applicator; this may be gravity fed (passive or user controlled), or the applicator may be prepped with sealant by dipping the applicator into a container of sealant. Although painting is not the preferred application method for the liquid dressing, it may be preferred if a high viscous sealant material is used to span large gaps, such as that between the packing material and the wound edge, the potentially high ridges of a hydrocolloid at its skin interface, or the large creases, gaps, and folds in a hydrocolloid dressing, due to its high stiffness and thickness and geometrical mismatch.

[0129] For sprayed applications, the device to atomize the sealant with a shearing process can be a refillable spray gun or airbrush, with an external pressurized gas supply, or this functionality can be incorporated into a miniature, handheld spray can, which can be rechargeable and refillable. Each embodiment has a design specific envelope of pressure, velocity and volume flow of gas that is required to shear the sealant, such that it forms a thin film, continuous layer on the skin. If the operation is outside the envelope, the droplets of the spray may be too large and will not spray as a continuous layer, but will sputter onto the skin, or the gas may not shear the fluid out of the fluid opening. In a functional embodiment, the liquid sealant is gravity fed into a center opening in a nozzle, and pressurized gas shears the sealant through a circumferential ring around the sealant nozzle opening. Multiple nozzles may exist for one or both fluids. Particularly, the spray pattern may be controlled through the shearing of the sealant from multiple gas ports, aimed in different shearing directions across the liquid sealant nozzle. In a handheld device, the pressurized gas may be generated from a miniature gas cylinder, such as a high pressure, liquid carbon dioxide cartridge. The spraying device may be charged by the caregiver when he or she activates the charged canister of gas.

[0130] Once the dressing-to-skin and tube-to-dressing interfaces are sealed (either during dressing application or during its manufacture), the caregiver should monitor the pump to assure that air is not leaking into the system above a predetermined threshold, typically zero, step 1505, FIG. 33. This can be done visually, for example, by monitoring the expansion of the pump (a.k.a., mechanical pump) or with an air leak test that is further disclosed in the pump descriptions, or it can be sensed using pressure sensors to detect the vacuum pressure over time (particularly, with a mechanical pump, if the pressure changes continuously with internal pump volume). If too high of an air leak exists, the dressing-to- skin and/or tube-to-dressing interfaces can be resealed with the liquid sealant by removing the previously applied sealant material, or overlaying the new sealant over the previously applied seal material, step 1506, FIG. 33. This is an iterative process until the desired air leak threshold is achieved, step 1507.

[0131] When a truly occlusive wound dressing is used for NPWT, the behavior of the system changes from an active flow system, FIG. 34A, to a passive flow system 1602, FIG. 34B. When air leaks, arrows 1603, FIG. 34A into the active system, the system has an active flow of fluid (both air and wound exudate) that both removes the exudate from the wound cavity 1604 and tends to dry out the wound cavity. With an air-tight system 1602, FIG. 34B, the flow of the exudate toward the pump is no longer an active flow, but tends to build up, 1605, even into the tube over time, maintaining a passive flow to the vacuum source. A pressure differential still exists at the surface of the wound bed 1606 and, thus, negative pressure is still being applied to the wound bed; however, the wound cavity volume 1604 fills with exudate fluid 1605 over time. This characteristic may have increased healing benefits compared to standard NPWT, as it maintains a moist, healing environment at the wound site, while also maintaining NPWT vacuum pressure benefits.

[0132] With this build-up of fluid 1605, FIGS. 34B and 35, the dressing-to-skin interface adhesion 1607 may be compromised over time by the exudate, and the exudate may eventually undermine the dressing and leak out of the dressing-to-skin interface. The rate of exudate removal, size of the wound cavity, and time between dressing changes determine the build-up characteristics. If there is a chance that the dressing may be compromised, it can be prevented with multiple methods, including:

[0133] 1) A sealant or additional adhesive that can withstand the exudate build-up may be applied. For this case, the sealant and/or additional adhesive should be applied as close to the wound edge as possible. This is difficult if a standard dressing was used. Planar dressings typically leak over the three-day dressing period if fluid build-up occurs. This is because the exudate often degrades the adhesive by undermining the dressing at the wound edge at the locations of initial creases in the dressing. Therefore, a dressing without initial cracks at the wound edge is preferred; however, the dressing application described in the previous section only seals the outer edge of the dressing. To solve this problem, a flexible adhesive, with flexibility and adhesive properties such as those of a 30+ day silicon wig glue, may be initially applied at the wound edge under the adhesive planar dressing. This can fill in any initial cracks at the wound edge and prevent exudate-caused degradation.

[0134] 2) A barrier can be applied at the wound edge, after the wound packing material is inserted. This barrier may be made of highly absorbent material, in order to reduce the chance of overspill of exudate due to factors, such as gravitational effects.

[0135] 3) The tube end can be recessed into the wound cavity below the plane of the surface of the skin 1702, as indicated by arrow 1701, FIG. 35. Therefore, the drainage line of fluid 1703, and hence the build-up of exudate will not build-up to the wound edge 1607, and degrade the adhesive. This technique may not be possible if the wound is superficial.

[0136] 4) A purge valve to let a controlled, temporary air leak into the dressing system to clear the fluid can be incorporated into the dressing system. This valve can be incorporated using the same connection methods as described in the Tube-to-Dressing Interface section in this disclosure. This would cause the fluid to actively flow into the fluid collection canister during the initial pressure drop in the system. The pump can be reset, if necessary.

[0137] 5) The wound packing material can be made from materials with a low resistance to the flow of exudate and a low absorption, which would encourage the fluid to passively move through the system at a faster rate in a path more direct to the drainage tube. Depending on the rate of exudate removal, this may not fix the problem if it is a very low rate. In this case, the packing material should be designed to direct flow to the drainage tube and specifically away from the wound edge.

[0138] 6) If the dressing is fabricated completely out of the liquid sealant (potentially with an additional adhesive) with no planar dressing component, then no cracks will exist at the wound edge when it is properly applied, and therefore, no cracks will initially exist for the exudate to undermine.

[0139] Although any mechanical or electrical vacuum source may be applied to the occlusive dressings in this disclosure, a mechanical system may be preferred due to the significant benefits over electrical pumps. Mechanical vacuum pumps and methods are provided for medical application in negative pressure wound therapy (PWT) that would be compatible with the disclosed dressings. A number of known pumps are described by the present inventor in "Development of a simplified Negative Pressure Wound Device" submitted in 2007 for her Master of Science in Mechanical Engineering at the Massachusetts Institute of Technology. The pump is initially set and then governed by a linear or nonlinear spring force. The pump enclosure may act as a collection chamber; however, a separate collection chamber may exist in series with the pump.

[0140] In one embodiment, the pump is a plastic bellows, shown in FIG. 31, where the enclosure and spring can be the same component. The pump is compressed manually and then attached to the tube of the wound dressing. A negative pressure is applied through expansion of the bellows due to the spring characteristics of its material and design. The pressure gradient of the device continuously decreases over the expansion of the standard bellows due to its linear spring-like properties. Referring to the above description, one skilled in the art would realize that other embodiments exist: the device could be constructed of a different material bellows, and/or the device could contain an additional spring 5 in parallel with the bellows in order to vary the spring constant without changing the material properties and design of the bellows itself. If there are no air leaks into the system, then the bellows would remain at a constant expansion length, and therefore, at a constant pressure. The bellows can be collapsed to any desired therapy pressure from maximum compression to zero.

[0141] In addition to the standard bellows, another embodiment of bellows can resemble a constant force spring, in order to decrease the pressure gradient. In one embodiment of this design, the bellows resembles a long tube that, when fully compressed, is rolled onto itself, similar to a tape measure, as shown in FIG. 32. As it unrolls and expands from its flattened to open cross-section, it creates negative pressure in the tube to which it is connected. For the tube to unroll following the expansion of the bellows, the spring constant of the bellows must be higher than the spring constant of the constant force spring unrolling. The unrolling can also be mechanically dampened, for example by adhesion, or forced to unroll after expansion by structural limiters. In this embodiment, a long, cylindrical tube can replace the bellows, as it presents similar characteristics.

[0142] In all of the pumps described above, orientation of the device is independent of the magnitude of negative pressure pulled and the proper operation of the device. Therefore, the device is highly transportable. Referring to the above descriptions, one skilled in the art would realize that other embodiments exist; however, only selected embodiments are described in detail. To change pressures in a pump design, separate pumps can be made with different material properties and/or dimensions, or components can be swapped for different pressure results.

[0143] The negative pressure generated is governed by the material and mechanical properties of the container and/or balloon and the spring constant. Using a non-constant force spring (such as a common linear spring 5, FIG. 31, the pump may be used for negative pressure wound therapy that does not require a specific, constant pressure (in the case that the internal volume of the pump is expected to expand), although the variance in pressure can be reduced through material property selection and design. Using a constant force spring with a constant area, a constant vacuum pressure can be pulled throughout treatment, even if there is a change in the internal volume of the pump. This is the basis for design of the rolling bellows (FIG. 32), and the syringe concept discussed in the next section. Additionally, a more constant pressure with pump expansion can be achieved with a non- constant force spring by designing the force/area ratio to be constant, such as the balloon design with a small (constant) diameter to length ratio and a bellows with a varying cross- sectional area. Additionally, constant pressure can be achieved over time if no air leaks into the system, causing geometrical changes in the pump configuration. In this case, the pumps should be made out of materials that do not degrade when applying negative pressure overtime due to properties such as stress relaxation.

[0144] The pump is initially set and then governed by gravity. It includes an expansion container that expands due to an applied force such as a weight. The pump enclosure may act as a collection chamber; however, a separate collection chamber may exist in series with the pump. In one embodiment, the pump includes a rolling diaphragm syringe (similar to a friction free diaphragm air cylinder). Negative pressure amplitude is governed by the diameter of the syringe and the magnitude of the attached weight. One skilled in the trade would realize that a similar device could also be constructed of any sealed piston syringe. Referring to the above descriptions, the device could also include a linear spring in parallel with the syringe or a constant force spring in series with the syringe for expansion, eliminating the need for weight. This embodiment would then fall under the spring governed pumps described in the previous section. A rolling diaphragm can also be achieved using a rubber ball design. One hemisphere of the rubber ball is held rigid in its inflated position, such as by bonding it to the inside of a rigid hemisphere, and the other hemisphere is compressed into it. The embodiment of the pump resembles a bowl. Then, the bowl is oriented so that its hollow side is facing down. A weight is hung from the ball (i.e., a rubber ball) on the hollow inside of the hemisphere, and the wound drainage tube is connected to the internal volume of the pump (preferably through the top of the rigid hemisphere). The weight pulls a negative pressure as the ball returns back to the shape of a sphere.

[0145] Another embodiment for a gravity governed pump is created by a siphon. The pump enclosure may act as a collection chamber; however, a separate collection chamber may exist in series with the pump. The pressure pulled is equal to:

rho*g*h (2)

wherein rho is the density of the fluid in the column, g being the gravitational constant, and h being the height of the column). The fluid should be compatible with the wound (such as saline), unless a check valve is used to assure separation of the pump fluid from the wound cavity. The pump can be configured in two ways, depending on the patient situation and the desired pressure:

[0146] 1. The pump can include a column of fluid that exists in a tube directly connected to the wound. The lower (preferably closed-expandable) container of fluid can rest at the desired height on a separate mechanism (such as a hanging hook or floor), or could be attached to a lower extremity of the patient, again at the desired height. The diameter of the tube would determine the pressure gradient: the larger the diameter, the lower the pressure gradient as fluid is collected.

[0147] 2. The pump can include two bodies of fluid with a tube from the higher body of fluid to the wound. The mobility of the patient would be determined by the tube length and the mechanism used to carry the pump (for instance, a rolling stand could be used). The diameter of the higher container would determine the pressure gradient: the larger the diameter, the lower the pressure gradient as fluid is collected.

[0148] Integrating the spring governed pumps with the gravity governed concept allows for further performance. Then, the magnitude of negative pressure a spring governed pump can obtain is not completely limited by the material properties of the container, the design, and the spring constant combination. Additional weights can be attached to one end of the pump in series with the spring, in order to pull a higher negative pressure. (For the bladder concept, portions of the bladder may need structural support, so mat the bladder does not collapse on itself as the weight acts on it.) The weights should be attached between the pump and ground. Even though in this form the orientation of the pump should be maintained, varying the additional weight is a simple solution to achieving multiple pressures beyond that of the original pump properties.

[0149] A container evacuation pump is not continuously governed by a force exerted on the container. Instead, the pump is simply an evacuated rigid chamber that is continuously monitored through a pressure gauge, such as gauge 4 in FIGS. 31 and 32. Alternatively, a mechanical check gauge would be used, with an optimal pressure range. When the vacuum pressure decreases to a certain, predetermined level, a notification mechanism is activated and the rigid chamber is recharged. Recharging can be by a pump or by human suction. In this embodiment, the rigid chamber can act as the collection container, or a separate non-structural, expandable container can be inserted into the rigid chamber that is directly connected to the wound drainage tube. An expandable collection chamber can be integrated into any of the mechanical pump concepts disclosed in this disclosure, in order to collect the fluid inside the pump body, acting as a collection liner instead of a completely separate collection canister.

[0150] To administer NPWT, the pump is connected to the wound drainage tube, and the container is then evacuated. Air leaks and wound drainage rate determines the pressure gradient, and the pressure range is determined by the maximum pressure pumped and the recharge notification pressure. The maximum pressure pumped can be limited by a pressure activated inlet valve.

[0151] As generally applies to all of the above-mentioned pumps, a sequence of steps should be followed. First, the tube connected directly to the dressing should be clamped shut between the dressing and the collection chamber, preferably at the collection chamber end. Then, the pump and collection chamber should be disconnected. If necessary, the collection chamber should be emptied, and/or the proper sterilization procedures should be performed; component 8, FIG. 31, represents a rubber plug, not shown because integral floor 260 is utilized instead, for increased access to the interior of the collection chamber in an alternative construction. The pump should then be reset, and the pump and collection chamber reconnected to the tube. Remove the clamp to begin NPWT again. If a dressing change is also desired, there is no need to use a clamp to keep the dressing sealed. Also, if the collection chamber does not need to be emptied and/or sterilized, then the tube should be clamped between the dressing and the pump, preferably after the collection chamber if it is separate from the pump at the pump end of the tube.

[0152] An air leak test can be incorporated into the mechanical pumps, except for the first (1) siphon concept. In the second (2) siphon concept, the higher container is turned upside-down for the initial air leak test. Most air leaks originate at the dressing interfaces. In a purely mechanical pump, air leaks fill the limited volume, causing the maximum time between pump resets to decrease. To eliminate these air leaks and create a reliable, repeatable therapy, devices according to the present disclosure may include an air leak test. By using the air leak test, the purely mechanical pumps have been proven to be capable of lasting throughout the recommended timeframe between dressing changes (3 days). However, this test is not necessary for the occlusive seals and dressings disclosed in this disclosure, but can provide a visual reassurance to the caregiver and patient that the dressing was applied properly and no significant leaks exist in the system.

[0153] The air leak test is in the collection chamber. The tube from the wound that enters into the collection chamber enters into a wound compatible solution (such as saline). When applying the NPWT, one should confirm that the end of the tube is submerged in the solution and should look at the solution for air bubbles B, FIGS. 31 and 32, (any air that initially exists in the tube may create air bubbles; therefore, one should wait about 1-2 seconds for additional air bubbles). If air bubbles are detected, the dressing should be sealed until no air bubbles are detected. This resealing may be to completely redress the wound, to smooth out the air leaks in the current dressing, or to reinforce the current dressing with additional dressing components. Once no air bubbles are detected, the pump may need to be reset depending on how much air entered into the pump.

[0154] A safety feature of the collection chamber is to limit the amount of liquid capable of being collected. If the collected liquid were blood due to destruction of a vein or artery, there exists a possibility that the patient may die due to fatal bleeding. The collection chamber should be limited to less than 300cc of liquid to keep the patient at a safe range from possible exsanguination. Therefore, if the pump design can pull more than 300cc of fluid, a safety feature should be implemented. If the pump acts as the collection chamber, the safety feature should limit its expansion volume. This can be done in various ways through the introduction of limiting, internal (FIG. 31, component 6) and/or external (FIG. 31, component 7) structural components. If an external collection chamber exists, then a safety feature should stop the negative pressure after 300cc is collected. This can be done by "plugging" the system with a mechanism, such as a float-stop valve.

[0155] Prior to the existence of a truly occlusive dressing, a benefit in the external collection chamber was that the pump can be larger than 300cc, and therefore, account for more air leaks into the system. However, with a truly occlusive dressing, the benefits include that the external collection chamber and its fluids can be easily removed for lab testing purposes, and the pump requires a less rigorous cleaning procedure between dressing changes. However, these benefits are more readily solved with a volume specific container with no rigidity, and containing no initial volume of fluid that may contaminate a exudate sample, if desired, that can be inserted into any of the mechanical pump concepts disclosed in this disclosure, in order to collect the fluid inside the pump body, acting as a collection liner instead of a completely separate collection canister. The 300cc limitation is recommended for the average adult; however, the limitation volume may vary based on the patient. This volume variation can be designed into multiple pump or collection chamber sizes, or into a single, limit adjustable pump or collection chamber.

[0156] Another pump safety feature is a one-way valve incorporated in the tube between the wound and the collection chamber, such as component 2, FIGS. 31 and 32. This mechanism assures that fluid from the pump and collection chamber does not flow back into the dressing. It also can be used as the tube clamping mechanism for resetting the pump or emptying the collection chamber, depending on placement in the tube. This mechanism can also be incorporated into the tube connector on the collection chamber.

[0157] Another mechanism that may be included is used to evacuate the initial air found in the system after no air leaks are detected. The current method is to clamp the tube near the pump and to reset the pump until the initial air is evacuated from the system. This can also be accomplished by including a one-way- valve incorporated into the tube connector on the pump, such as component 2, FIGS. 31 and 32, and another one-way-valve incorporated between the interior cavity of the pump and atmosphere, component 3, shown with a cap C in FIG. 31. With this design, one can continue to compress (reset) the pump until the desired vacuum is maintained; the system does not need to be disconnected. The one-way-valve open to atmosphere can be capped after therapy begins.

[0158] This mechanism cannot be easily integrated to eliminate the need for resetting the pump in the design that includes a rubber balloon that is inserted into an orifice of an air-tight container and the two siphon pumps. For the balloon design, a connection to the container can be made to incorporate the attachment of a separate pump with the oneway-valve and check valve design. This pump can be attached for initial balloon inflation and container evacuation, and then detached between dressing changes. In the two siphon concepts, a pump can be attached to evacuate the space above the column of fluid, raising the fluid level to the desired height. The pump can be detached for extended therapy, between dressing changes.

[0159] An individual sealant component may be packaged by itself to make any skin dressing occlusive. Alternatively, the sealant can be packaged as part of a mechanical NPWT kit, including a mechanical pump and its pre-attached components, tubing with flexible foot and pre-attached tubing connector and optional one way valve, dressing adhesive film to cover the packing material (if necessary), the sealant material in a handheld spray container, a wound packing material, and skin prep (if necessary). Additionally, if there is an adhesive dressing tape-like film that should be handled by the caregiver, then non-stick fingertip covers maybe included for better adhesion outcomes. Non-powdered gloves may also be included, so that the Van der Waals forces for sealant attachment are not altered due to powder on the skin surface. One skilled in the art would realize that kit components may be swapped for their different functional embodiments, discussed above. Also, additional components may be added or put into additional kits that are used in typical dressing changes, such as wound debridement tools, or additional wound therapies, such as medications with their corresponding introduction and (potentially) removal ports through the dressing, into the wound cavity.

[0160] As many dressing systems are identified in this disclosure, one skilled in the art would realize that the liquid sealing method can be used in combination with any tissue (a.k.a., skin) dressing in order to create an air-tight seal. As many pumps are identified in this disclosure, one skilled in the art would realize that any pump combined with the occlusive dressing systems would have similar performance characteristics.

[0161] One technique according to the present invention for constructing an occlusive dressing over a wound includes at least one of (1) packing the wound with a fluid- pervious material and (2) covering at least a portion of the wound with a protective material. The method further includes applying, such as by spraying, an organic material, preferably elastomeric, that is in a liquid state, and is at least partially cross-linked at least after one of drying and curing, over the packed material and onto skin surrounding the wound to create an occlusive drape as a thin sheet substantially impervious to fluid transfer, having a first, inner surface and a second, outer surface. As utilized herein, the term "organic material" includes matter in various forms that include carbon atoms, including silicone rubbers. The method includes at least one of drying and curing the elastomeric material within thirty minutes after application of the elastomeric material as a layer.

[0162] FIG. 38 is a schematic top plan view of the wound shown in FIG. 14, here with a cleaned skin zone indicated by dashed line 302 and with packing material 304, such as gauze or a sponge. Additionally, a protective covering material 306 is applied over the wound when intended liquid drape material has a sufficiently lower viscosity and longer set time to be absorbed into the packing material 304. Protective material 306 prevents liquid drape material from flowing into the wound cavity. In some constructions, protective material 306 is a solid impermeable or semi-permeable polymeric sheet, which may be utilized with or without adhesive. In other constructions, protective material 306 is a claylike substance that can be molded and packed over packing material 304 and around a tube inserted into the packing material 304.

[0163] FIG. 39 is a view of FIG. 38 with a hole 308 cut in the protective covering 306, if an opening in the protective covering 306 has not already been formed or maintained.

[0164] FIG. 40 is a view of FIG. 39 with a tube assembly 27c, with tube 24c and flange 26c, having sleeve 52d, rotation region 54d and adhesion region 56d, placed over the hole 308. Tube assembly 27c is maintained in position with adhesive in some techniques and, in other techniques, is manually held in place. Alternatively, tube 24c can directly puncture protective covering 306 such that covering 306 maintains a tight seal around tube 24c, potentially with an additional seal barrier such as clay or adhesive. In another construction, flange 26c is sufficiently large in diameter to completely cover the wound.

[0165] FIG. 41 is a view of FIG. 40 with liquid drape material 310 applied over the protective covering 306 and onto surrounding skin SK to cover skin zone 302 to thereby construct a dressing according to the present invention. The liquid drape material 310 is applied by spraying or application technique such that material 310 firmly attaches to skin SK surrounding the wound and covers any protective cover layer 306, if utilized, as well as creating an air-tight seal around flange 26d. FIG. 42 is a schematic perspective view of the dressing of FIG. 41 In other constructions, the drape is constructed directly onto a tube without utilizing a separate flange.

[0166] FIG. 43 is a schematic perspective view of a novel flange 26d according to the present invention with integral connector 320 having a barb-type engagement feature 322 defining passage 324. Engagement feature 322 is insertable into end of a tube.

[0167] Although specific features of the present invention are shown in some drawings and not in others, this is for convenience only, as each feature may be combined with any or all of the other features in accordance with the invention. While there have been shown, described, and pointed out fundamental novel features of the invention as applied to one or more preferred embodiments thereof, it will be understood that various omissions, substitutions, and changes in the form and details of the devices illustrated, and in their operation, may be made by those skilled in the art without departing from the spirit and scope of the invention. For example, it is expressly intended that all combinations of those elements and/or steps that perform substantially the same function, in substantially the same way, to achieve the same results be within the scope of the invention. Substitutions of elements from one described embodiment to another are also fully intended and contemplated. It is also to be understood that the drawings are not necessarily drawn to scale, but that they are merely conceptual in nature. It is the intention, therefore, to be limited only as indicated by the scope of the claims appended hereto. Other embodiments will occur to those skilled in the art and are within the following claims.

---

http://en.wikipedia.org/wiki/Negative-pressure_wound_therapy

Negative-pressure wound therapy

Application of a vacuum pump using a foam dressing to a wound

Negative-pressure wound therapy (NPWT) is a therapeutic technique using a vacuum dressing to promote healing in acute or chronic wounds and enhance healing of first and second degree burns. The therapy involves the controlled application of sub-atmospheric pressure to the local wound environment,[1] using a sealed wound dressing connected to a vacuum pump.[2][3] The use of this technique in wound management increased dramatically over the 1990s and 2000s[4] and a large number of studies have been published examining NPWT.[5] NPWT appears to be useful for diabetic ulcers[6] and management of the open abdomen (laparotomy) [7] but further research is required for other wound types.[8]

Overview

NPWT promotes wound healing by applying a vacuum through a special sealed dressing. The continued vacuum draws out fluid from the wound and increases blood flow to the area.[2] The vacuum may be applied continuously or intermittently, depending on the type of wound being treated and the clinical objectives. Typically, the dressing is changed two to three times per week.[3] The dressings used for the technique include open-cell foam dressings and gauze, sealed with an occlusive dressing intended to contain the vacuum at the wound site.[1] Where NPWT devices allow delivery of fluids, such as saline or antibiotics to irrigate the wound,[9][10] intermittent removal of used fluid supports the cleaning and drainage of the wound bed.[11]

In 1995, Kinetic Concepts was the first company to have a NPWT product cleared by the US Food and Drug Administration.[12] Following increased use of the technique by hospitals in the US, the procedure was approved for reimbursement by the Centers for Medicare and Medicaid Services in 2001.[1]

Technique

General technique for NPWT is as follows: a dressing or filler material is fitted to the contours of a wound (which is covered with a non-adherent dressing film) and the overlying foam is then sealed with a transparent film. A drainage tube is connected to the dressing through an opening of the transparent film. A vacuum tube is connected through an opening in the film drape to a canister on the side of a vacuum pump.[13] or vacuum source, turning an open wound into a controlled, closed wound[2] while removing excess fluid from the wound bed to enhance circulation and remove wound fluids. This creates a moist healing environment and reduces edema.[5][13] The technique is usually used with chronic wounds or wounds that are expected to present difficulties while healing (such as those associated with diabetes).[3]

Three types of filler material are used over the wound surface: open-cell foam, gauze and transparent film, or honeycombed textiles with a dimpled wound contact surface.

Foam dressings are used to fill open cavity wounds and can be cut to size to fit wounds. The foam dressing is applied, filling the wound and then a film drape is applied over the top to create a seal around the dressing.

Open weave cotton gauze can be covered with a transparent film, and a flat drain is sandwiched in gauze and placed onto the wound. The film drape covers the wound and create a complete seal, and then the drain is connected to the pump via the tubing.[14]

Layers of non-woven polyester, joined by a silicone elastomer, has a non-adherent wound contact surface made up of numerous small semi-rigid dome structures,[15] though this solution has not completed a clinical trial as of December 2012.[16]

With all three techniques, once the dressing is sealed the vacuum pump can be set to deliver continuous or intermittent pressures, with levels of pressure depending on the device used,[13][14][17] varying between −125 and −75 mmHg depending on the material used and patient tolerance.[15][18] Pressure can be applied constantly or intermittently.[13]

The dressing type used depends on the type of wound, clinical objectives and patient. For pain sensitive patients with shallow or irregular wounds, wounds with undermining or explored tracts or tunnels, gauze may be used, while foam may be cut easily to fit a patient’s wound that has a regular contour and perform better when aggressive granulation formation and wound contraction is the desired goal.[19]
Effectiveness

A 2007 Cochrane Review stated that the evidence comparing NPWT to alternative care was flawed and required more study, but the evidence did support improved healing and called for more, better quality research to be conducted.[8] A 2010 systematic review found "consistent evidence of the benefit of NPWT" in the treatment of diabetic ulcers of the feet. Results for bedsores was "conflicting" and research on "mixed wounds" was of poor quality, but promising. The review did not find evidence of increased significant complications. The review concluded "There is now sufficient evidence to show that NPWT is safe, and will accelerate healing, to justify its use in the treatment of diabetes-associated chronic leg wounds. There is also evidence, though of poor quality, to suggest that healing of other wounds may also be accelerated."[6]

References

a b c Lillis, Karin (2003). "Effective wound care requires look at total patient picture". Healthcare Purchasing News 27 (1): 32. ISSN 0279-4799

a b c Moody, Yasmeen (19 July 2001). "Advances in healing chronic wounds". The Ithaca Journal (Ithaca, NY). p. 10A

a b c Fogg, Erich (27 August 2009). "Best treatment of nonhealing and problematic wounds". Journal of the American Academy of Physician Assistants 22 (8): 46, 48. PMID 19725415.

Driscoll, P (24 October 2009). "Negative Pressure Wound Therapy (Gauze and Foam)". Advanced Medical Technologies

a b Gupta, Subhas; Bates-Jensen, Barbara; Gabriel, Allen; Holloway, Allen; Niezgoda, Jeffrey; Weir, Dot (2007). "Differentiating Negative Pressure Wound Therapy Devices: An Illustrative Case Series". Wounds 19 (1 (Supplement)): 1–9

a b Xie, X.; McGregor, M.; Dendukuri, N. (November 2010). "The clinical effectiveness of negative pressure wound therapy: a systematic review". Journal of Wound Care 19 (11): 490–5. PMID 21135797.

Fitzgerald JEF, Gupta S, Masterson S, Sigurdsson HH. Laparotomy Management using the ABThera™ Open Abdomen Negative Pressure Therapy System in a Grade IV Open Abdomen Secondary to Acute Pancreatitis. International Wound Journal 2012. PMID 22487377

a b Ubbink, Dirk T; Westerbos, Stijn Joël; Evans, Debra; Land, Lucy; Vermeulen, Hester (2008). "Topical negative pressure for treating chronic wounds". In Ubbink, Dirk T. Cochrane Database of Systematic Reviews 16 (3): CD001898. doi:10.1002/14651858.CD001898.pub2. PMID 18646080.

Gerry R, Kwei S, Bayer L, Breuing KH (July 2007). "Silver-impregnated vacuum-assisted closure in the treatment of recalcitrant venous stasis ulcers". Ann Plast Surg. 59 (1): 58–62. doi:10.1097/01.sap.0000263420.70303.cc. PMID 17589262

Wendling, Patrice (April 2008). "Vacuum-assisted wound therapy uses expanded". Skin & Allergy News. Retrieved 11 January 2011

Moch D, Fleischmann W, Westhauser A (1998). "Instillationsvakuumversiegelung: ein erster Erfahrungsbericht" [Instillation vacuum sealing—report of initial experiences]. Langenbecks Archiv für Chirurgie (in German) 115: 1197–9. PMID 9931834.

"Vacuum Assisted Closure Wound Therapy Cleared for Partial Thickness Burns". Reuters Health Medical News. January 27, 2003.

a b c d Baxter, Helena; Ballard, Kate (2001). "Vacuum-Assisted Closure". Nursing Times 97 (35): 51–2. PMID 11957602. Retrieved 11 January 2011.

a b Miller, Michael S.; Brown, Rhonda; McDaniel, Cheryl (1 September 2005). "Negative pressure wound therapy options promote patient care". Biomechanics. p. 49.

a b Glat, Paul (8 July 2010). "The use of negative pressure wound therapy with Bio-Dome™ dressing technology". Today's Wound Clinic. Retrieved 20 January 2011.[unreliable medical source?]

http://clinicaltrials.gov/show/NCT01107821

Michael S., Miller (February 2009). "Multiple approaches offer negative pressure options". Biomechanics.

Morykwas, Michael J.; Argenta, Louis C.; Shelton-Brown, Erica I.; McGuirt, Wyman (1997). "Vacuum-assisted closure: a new method for wound control and treatment". Annals of Plastic Surgery 38 (6): 553–62. doi:10.1097/00000637-199706000-00001. PMID 9188970.

Long, Mary Arnold; Blevins, Anne (2009). "Options in negative pressure wound therapy". Journal of Wound, Ostomy and Continence Nursing 36 (2): 202–11. doi:10.1097/01.WON.0000347664.10217.2e.

---

https://www.ecri.org/Documents/Press%20Releases/Negative_Pressure_Wound_Therapy_Devices.pdf

Agency for Healthcare Research and Quality
Technology assessment: negative-pressure wound therapy devices

---

http://www.fda.gov/MedicalDevices/Safety/AlertsandNotices/TipsandArticlesonDeviceSafety/ucm225038.htm

(Article reprinted from *Nursing2010*, September issue, p.64-66.)

Negative pressure wound therapy: Use with care

By Nasrin Mirsaidi, MSN, RN, CNOR

NEGATIVE Pressure wound therapy (NPWT) is used to treat acute and chronic wounds. A vacuum source creates continuous or intermittent negative pressure inside the wound to remove fluid, exudates, and infectious materials to prepare the wound for healing and closure.[1] NPWT systems consist of a vacuum pump, drainage tubing, a foam or gauze wound dressing, and an adhesive film dressing that covers and seals the wound.

Since 1997, when the first device indicated for NPWT was cleared by the FDA for marketing in the United States, the system has evolved considerably. New device and dressing features have been added that make the system safer, more efficient, and more user-friendly in care settings other than hospitals, such as in long-term care facilities, rehab centers, and even patients' homes.

Although many patients have benefited from NPWT, adverse events – including deaths and serious injuries – have been reported to the FDA. The following is a case example:

An older adult underwent bilateral femoral endarterectomy for severe peripheral arterial disease. About 1 week after the procedure, the patient required surgery for wound exploration and evacuation of blood clots and then was diagnosed with an infection of the right groin incision. An NPWT system was applied to the wound and the patient was discharged home on antibiotic and coagulation therapy.

Ten days later, the patient experienced a massive hemorrhage from the right groin wound. When paramedics arrived at the home, they provided basic life and advanced cardiovascular life support. After transport to the ED, the patient was pronounced dead. The cause of death was exsanguination.

What went wrong?

The most serious adverse events associated with NPWT, blood loss and infection, can be prevented if healthcare professionals pay close attention to:

Device labeling, including warnings, instructions for use, indications, and contraindications.

Criteria for patient selection, wound type, and appropriate care setting based on each patient's needs.

What precautions can you take?

These strategies are recommended to prevent complications:

Patient selection and wound type.

Patients should be thoroughly evaluated for bleeding risk. The patient in the case example had several risk factors for bleeding, including history of postoperative blood clots in the wound, anticoagulation therapy, and wound infection. Besides being at high risk for infection, groin wounds are also prone to delayed wound closure due to hip joint movement.

High-risk patients and selection of care setting.

NPWT should be used in an appropriate healthcare setting based on the patient's need for monitoring and wound care. At high risk for bleeding and infection, this patient should have been in a healthcare facility where healthcare professionals are available around the clock.

Wound care-specific considerations.

Many reported adverse events are the result of inappropriate methods of dressing change. For instance, bleeding can occur from improper removal of dressings that have adhered to or are imbedded in the tissues. Bleeding can also occur if the dressing is applied directly to exposed or superficial vessels in or around the wound that haven't been completely covered and protected as recommended by the device manufacturer before initiating NPWT.

Wound infections may develop if pieces of dressing are retained in the wound. To prevent this complication, device instructions recommend counting the total number of dressing pieces used in the wound and documenting that number on the outer adhesive film dressing and in the patient's medical record. Teach the caregiver to count the number of dressing pieces removed to make sure that all pieces are accounted for.

Training of healthcare providers, patients, and caregivers.

Healthcare providers using NPWT devices should undergo appropriate training on device use, including its indications and contraindications, and recognition and management of potential complications.

NPWT training for patients and their caregivers who will be using the device at home should include how to:

Safely operate the device; provide a copy of printed instructions for patient use from the specific device manufacturer

Respond to audio and visual alarms

Perform dressing changes

Recognize signs and symptoms of complications, such as redness, warmth, and pain associated with possible infection

Contact appropriate healthcare providers, especially in emergency situations

Respond to emergency situations; for instance, if bright red blood is seen in the tubing or canister, to immediately stop NPWT, apply direct manual pressure to the dressing, and activate emergency medical services.

Use return demonstrations to evaluate patients' and caregivers' skills and comprehension of training materials.

On November 13, 2009, the FDA published a preliminary public health notification on the use of NPWT systems. It will be updated according to findings of an ongoing investigation. Please review the links below for further details. You play an important role in keeping your patients safe when using NPWT.

http://www.fda.gov/MedicalDevices/Safety/AlertsandNotices/PublicHealthNotifications/ucm190658.htm

http://www.fda.gov/MedicalDevices/Safety/AlertsandNotices/PatientAlerts/ucm190476.htm

http://www.fda.gov/ForConsumers/ConsumerUpdates/ucm193277.htm

REFERENCE

1. Agency for Healthcare Research and Quality. Technology assessment: negative-pressure wound therapy devices. 2009.

https://www.ecri.org/Documents/Press%20Releases/Negative_Pressure_Wound_Therapy_Devices.pdf

---

http://www.journalofwoundcare.com/cgi-bin/go.pl/library/article.html?uid=79697;article=JWC_19_11_490
Journal of Wound Care, Vol. 19, Iss. 11, 10 Nov 2010, pp 490 - 495

The clinical effectiveness of negative pressure wound therapy: a systematic review

X. Xie, M. McGregor, N. Dendukuri

Objective: To estimate the efficacy of negative pressure wound therapy (NPWT), on the basis of a systematic review of reported randomised controlled trials (RCTs).

Method: A systematic literature search for relevant RCTs was carried out. The credibility of the outcome of each study was evaluated using a specially constructed instrument.

Results: We identified 17 RCTs, of which five had not been included in previous reviews or health technology assessments. For diabetic foot ulcers (seven RCTs), there was consistent evidence of the benefit of NPWT compared with control treatments. For pressure ulcers (three RCTs), results were conflicting. In trials involving mixed wounds (five RCTs), evidence was encouraging but of inadequate quality. Significant complications were not increased.

Conclusion: There is now sufficient evidence to show that NPWT is safe, and will accelerate healing, to justify its use in the treatment of diabetes-associated chronic leg wounds. There is also evidence, though of poor quality, to suggest that healing of other wounds may also be accelerated.

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

Roof Mount Antennas - LTE Multiband/GPS
Pulse/Larsen Panther Series Antennas, for transmitting mission critical data.
tessco.com/Public_Safety/VehicleAntennas

Dave's Homemade Radios

The sweetest music comes from a radio you made!

Two Radios in one — My #24 Crystal Radio

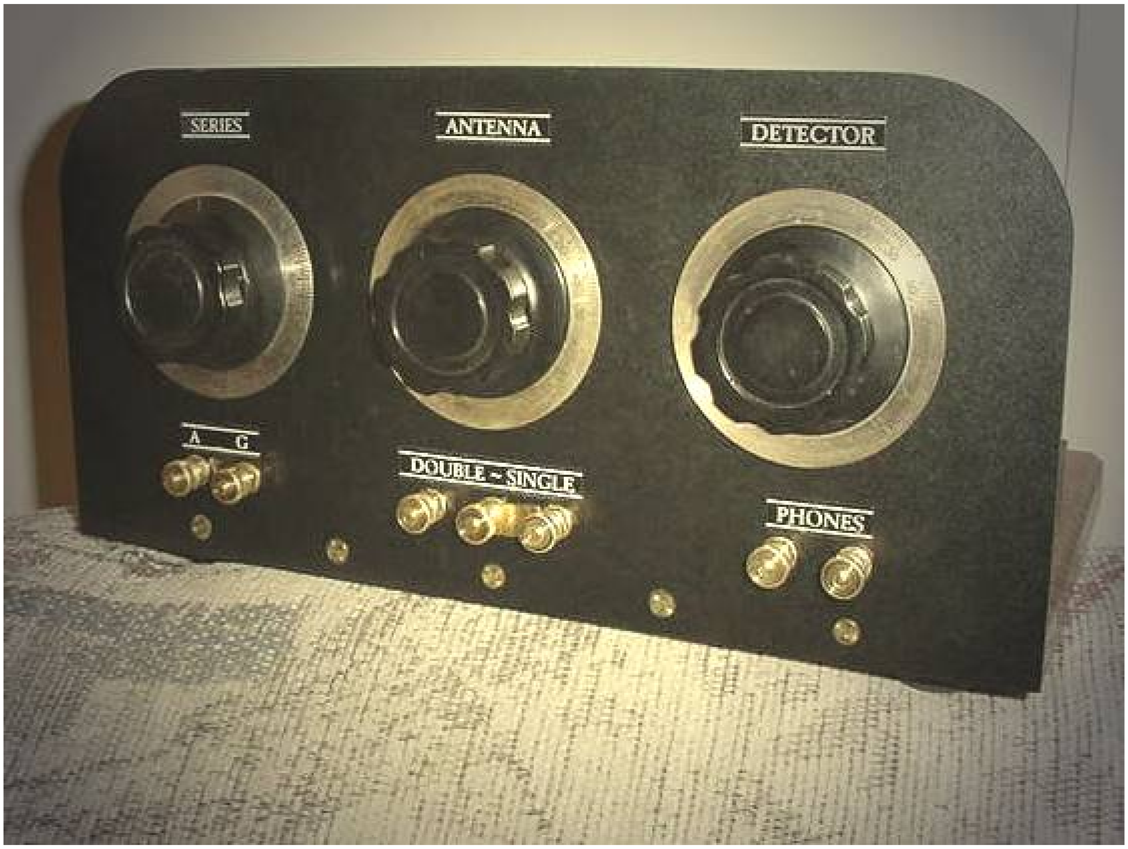

A- (javascript:changeFontSize(-1)) A+ (javascript:changeFontSize(1)) (javascript:changeFontSize(1))

Hi there! This is another of my recently revised radios. Not that I built them wrong to begin with, but I come across new and better materials to make my sets. I don't mind going back and revisiting a set now and again.

Building on the success of my last (23.php) radio, I proudly offer this crystal set for consideration for your next construction project. This is another one of my dual circuit crystal radios. The idea for this set comes from Mike Tuggle's Lyonodyne 17 (http://www.crystalradio.net/crystalsets/lyonodyne/index.shtml) and experience that I have gotten from my past sets. I found that with a very simple switching system, I can make this radio's circuit act like two different ones. I have wanted to build a Lyonodyne type set for a while but couldn't resist adding the extra switch.

This radio features the use of three variable capacitors. While at first, this is quite a chore to tune, after a while you would get used to where the dial is to be set for certain stations. You also have a choice of two radio circuits in one radio. One is the single coil that works very well and is featured in my #12 (12.php) and other sets. A hidden feature of this radio circuit is that the unused coil / capacitor becomes a wave trap! This wave trap can be used to reduce the signal of a loud interfering blowtorch station, allowing you to hear other, weaker stations that are close in frequency.

The two coil receiver is similar to the Lyonodyne circuit. The series tuning capacitor has been transposed so it is in the antenna line, rather than the ground line. I also decided to use two capacitors, rather than one capacitor. This increases the tuning complexity of this radio. The circuit was explained in some forums, but I just couldn't wrap my arms around using one ganged capacitor, so I used two. Since this is my radio, I built it my own way. It is kind of a "German" in me.

The antenna coil is the original one I used in this set. I could have used 660/46 litz on both coils, but this works dandy the way it is. The best inductance in the antenna coil will depend on your antenna. It is cut and try but you have leeway. My antenna is a single wire about 75 feet long.

The detector coil is made from 660 strand 46 gauge litz wire. This is the first coil I wound with the big litz. It came up short from the 240 µH that I wanted for my other project but will work pretty well in my single / double coil arrangement. The worst case is that the low end of the band can't be tuned without an extra capacitor.

The coils are mounted using dowel rods and 1/2 inch square wood blocks. I have used this scheme in my #23 (23.php) and #16 (16.php) sets with good results. I made it so the antenna coil could slide to adjust the amount of coupling used. Most of the time I use them fully apart for the best selectivity. If you experience real weak reception, you can move the coils closer together.

My revision included changing the detector circuit to match the extra high performance of the detector coil. This included using the 100k to 1.5k ohm audio transformer. You can pick out the transformer that meets your headphones requirements, or build a matching box (../misc-stuff/transformer.php) . The coil and the transformer turn this set into a high performance radio. If you can, try to find good quality variable capacitors. The ones commonly available really aren't that good in high performance radios. The capacitors I selected have a ceramic insulation on the stator.

The circuit details is where it starts to get interesting. Look at the circuit below while I rattle on about this set. When the switch is set to double, you see an antenna input, series capacitor connecting to a coil / capacitor parallel circuit. If you set the switch to single, you see an antenna input, series capacitor also connecting to a coil / capacitor parallel circuit. In the case of the double switch position, there are two tuned circuits, compared to a single tuned circuit in the single position. I just found that highly interesting and also helped me in how I should tune this set.

Here is how I tune this radio. There are three variable capacitors, the **series** (in series with the antenna), the **antenna** tuning (across the antenna coil) and the **detector** tuning (across the tapped coil). Tuning to maximum means that the capacitor plates are fully meshed and minimum means the plates are open. Ok? Here we go.

Single position tuning instructions: Tuning starts at the high end of the band and works downwards. Set the detector and series capacitors fully open. Fully mesh the antenna tune capacitor (if you are near the low end of the band, this capacitor should be fully open as it is a wave trap). Rotate the series capacitor as you listen for stations. If you hear a strong station, and want to increase the selectivity (at the cost of some sensitivity), mesh the plates of the detector capacitor a little and then peak the series capacitor to maximize the signal. After you make your first pass, set the detector capacitor at a third of rotation and start tuning again. Do this until you have reached the bottom of the band. The antenna capacitor can be tuned to tune out a loud station so you can hear others. Try it.

Double position tuning instructions: Tuning is a lot like in the single position. Set the antenna capacitor at minimum. Rotate the detector capacitor a little at a time while sweeping the series capacitor. You will be tuning starting at the high end of the band. When you hear a signal, then start peaking each capacitor starting with the antenna, then detector, then back to series. Keep repeating that until the signal is at full strength. Having numbered knobs and keeping a log of the settings helps finding the station again.

As a little "extra", I have included a little circuit to use if your sets don't tune the full range. It is a padder arrangement. Instead of hooking two gangs of a capacitor together for full range tuning, adding a padder capacitor will allow you to tune only to the bottom of the band. That way, you can spread out your tuning range to fit the full rotation of the capacitor. It does require the use of a 2 gang capacitor.

I haven't talked too much about construction details on this page. I have a new construction hints (../misc-stuff/cp.php) page on my site. Also, take a look back through my previous sets for little nuggets of hints. If you have a question or two, or would like to chat about crystal sets, I am only an e-mail away.

Best wishes and good DX. Dave - N2DS

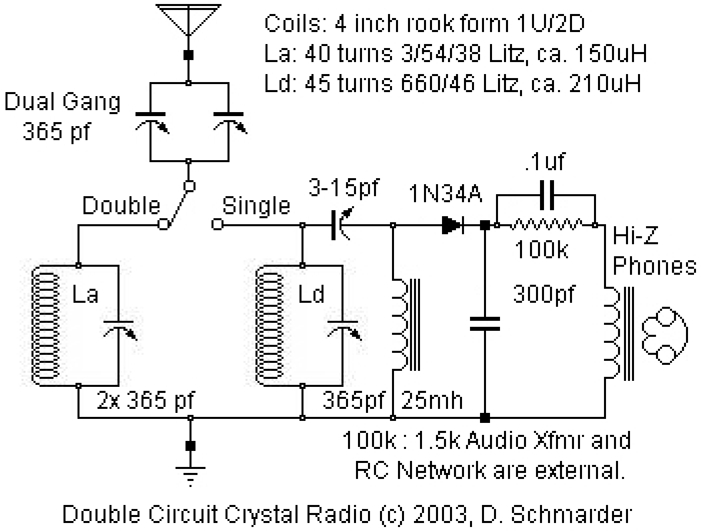
Coils: 4 inch rook form 1U/2D
La: 40 turns 3/54/38 Litz, ca. 150uH
Ld: 45 turns 660/46 Litz, ca. 210uH
Dual Gang
365 pf
.1uf
3-15pf
1N34A
Double
Single
100k
Hi-Z
Phones
La
Ld
300pf
2x 365 pf
365pf
25mh
100k : 1.5k Audio Xfmr and
RC Network are external.
Double Circuit Crystal Radio (c) 2003, D. Schmarder

Schematic

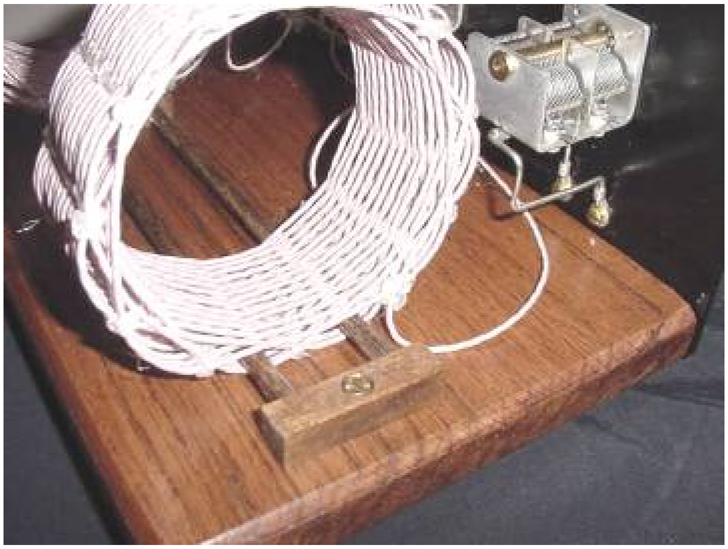

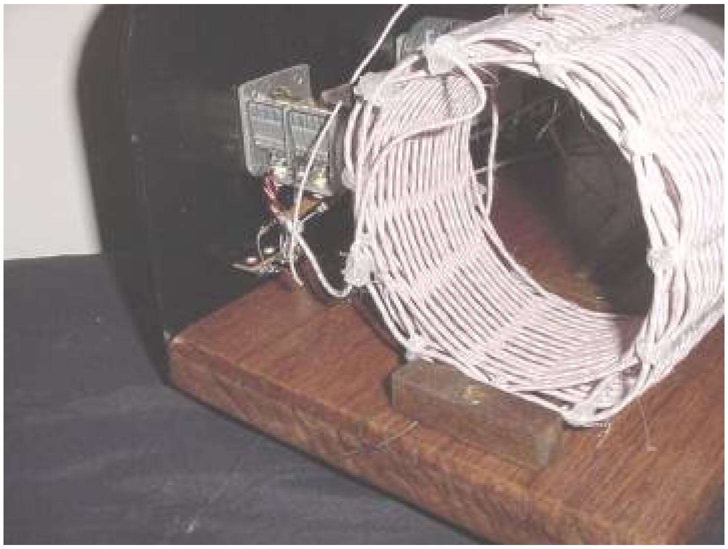

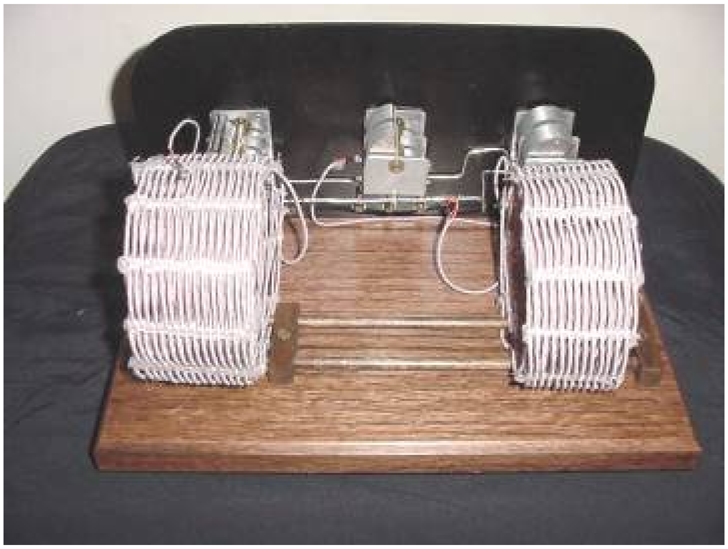

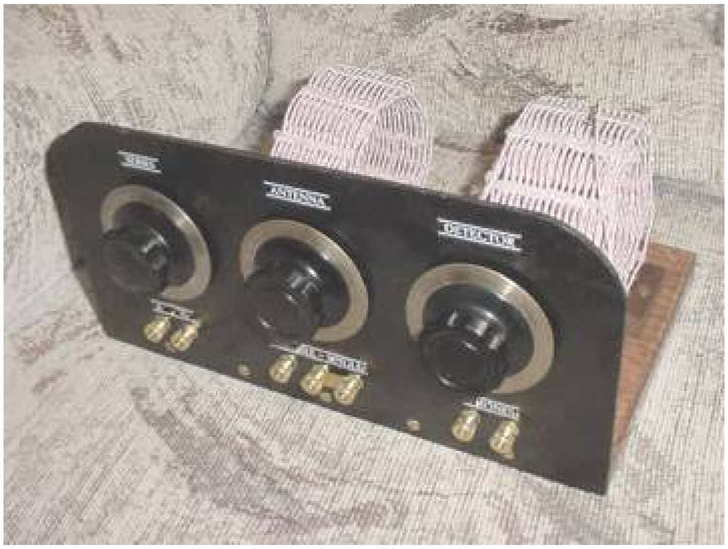

Padder option: Use the trimmer padder
to set the lowest frequency when the
dual 365 is set to maximum.

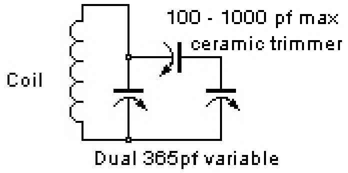
100 - 1000 pf max
ceramic trimmer
Coil
Dual 365pf variable

[(https://plus.google.com/u/0/110073333267145506266/posts)](https://plus.google.com/u/0/110073333267145506266/posts)

[Closeout Radio Parts Available (http://1n34a.com)](http://1n34a.com)

Crystal Radio Society Kit

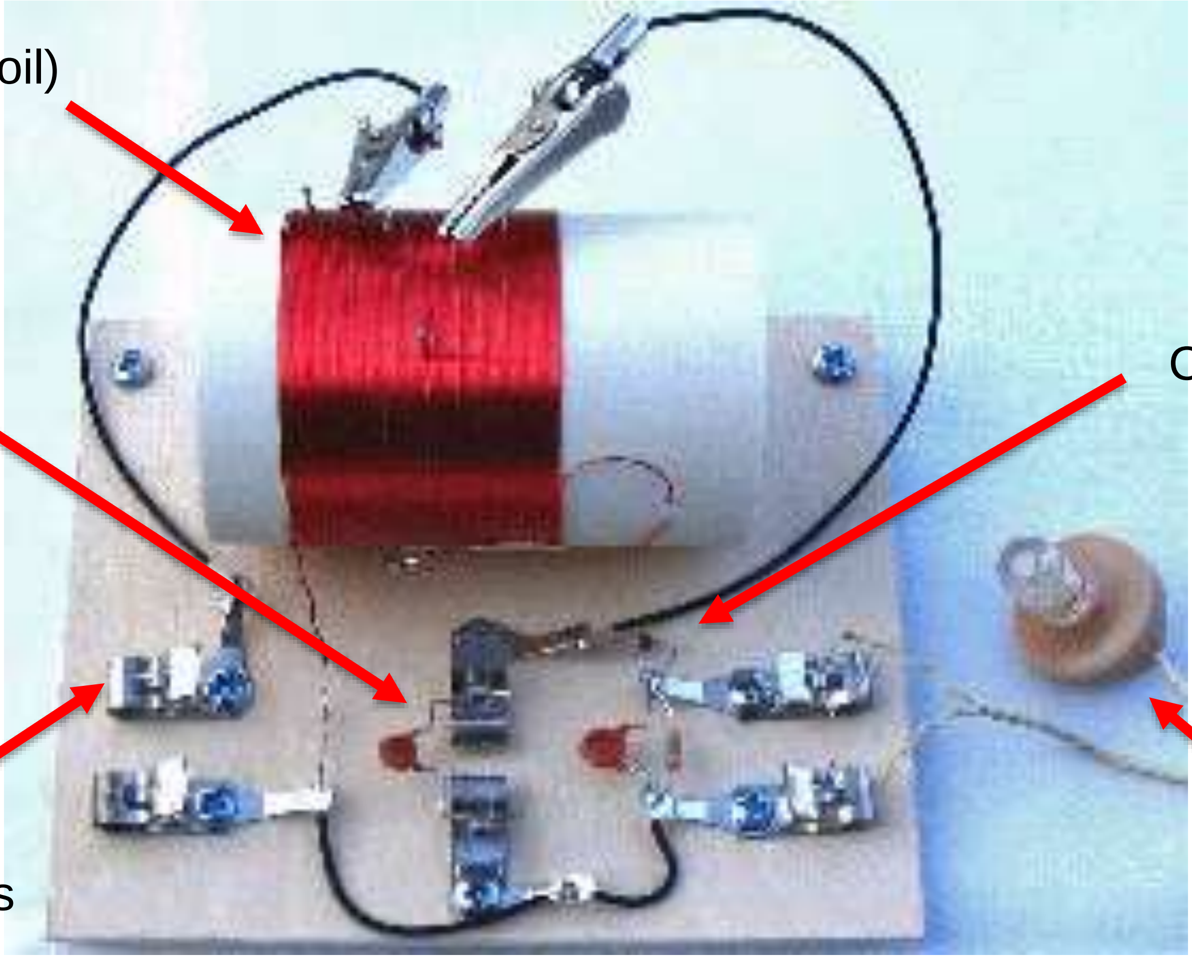

Inductor (Coil)

Capacitor

Crystal Diode

Fahnstock Clips

Earphone

Cat’s Whisker Detector

Step back in time with a galena detector kit and experience the soft detection that comes from minerals.

Delight in the first broadcast reception using the galena as a replacement for those 1N34 diodes

Wood base, mounted galena, phosphor-bronze cat whisker, 3-inch brass rod, compression spring, and brass parts.

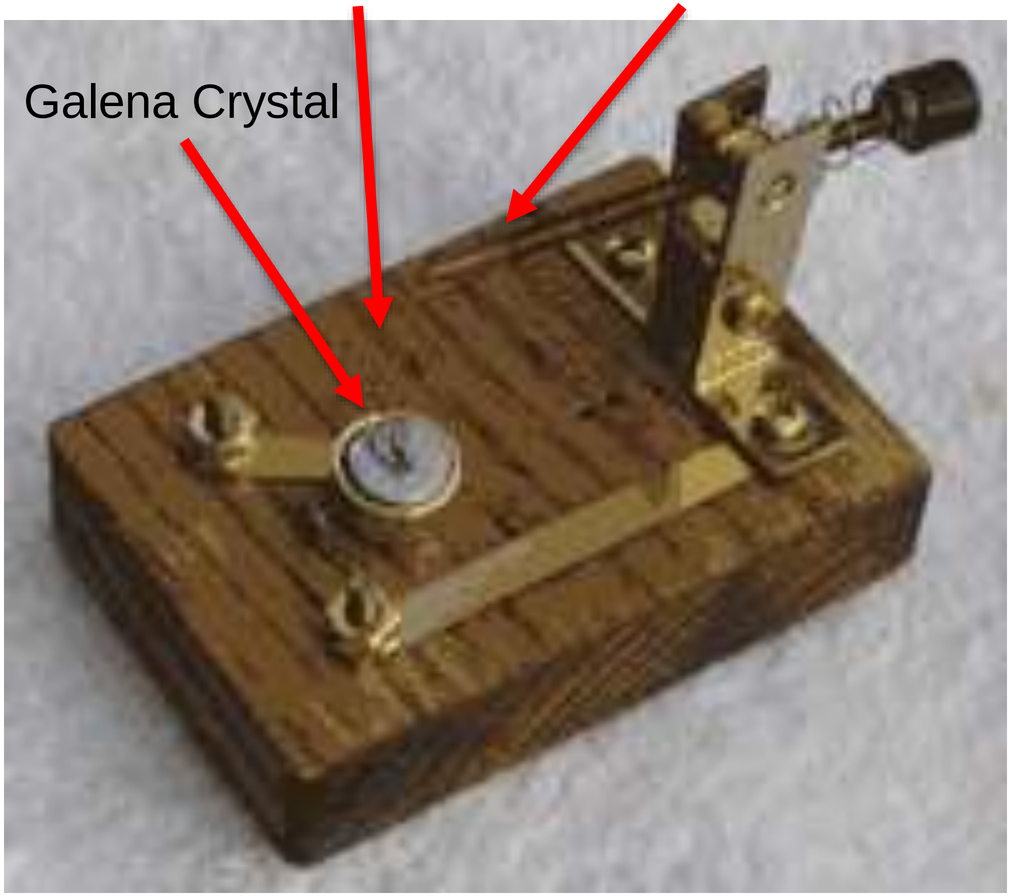
Galena Crystal
Positioning Arm
Galena Crystal

Foxhole Radio Pictorial Schematic

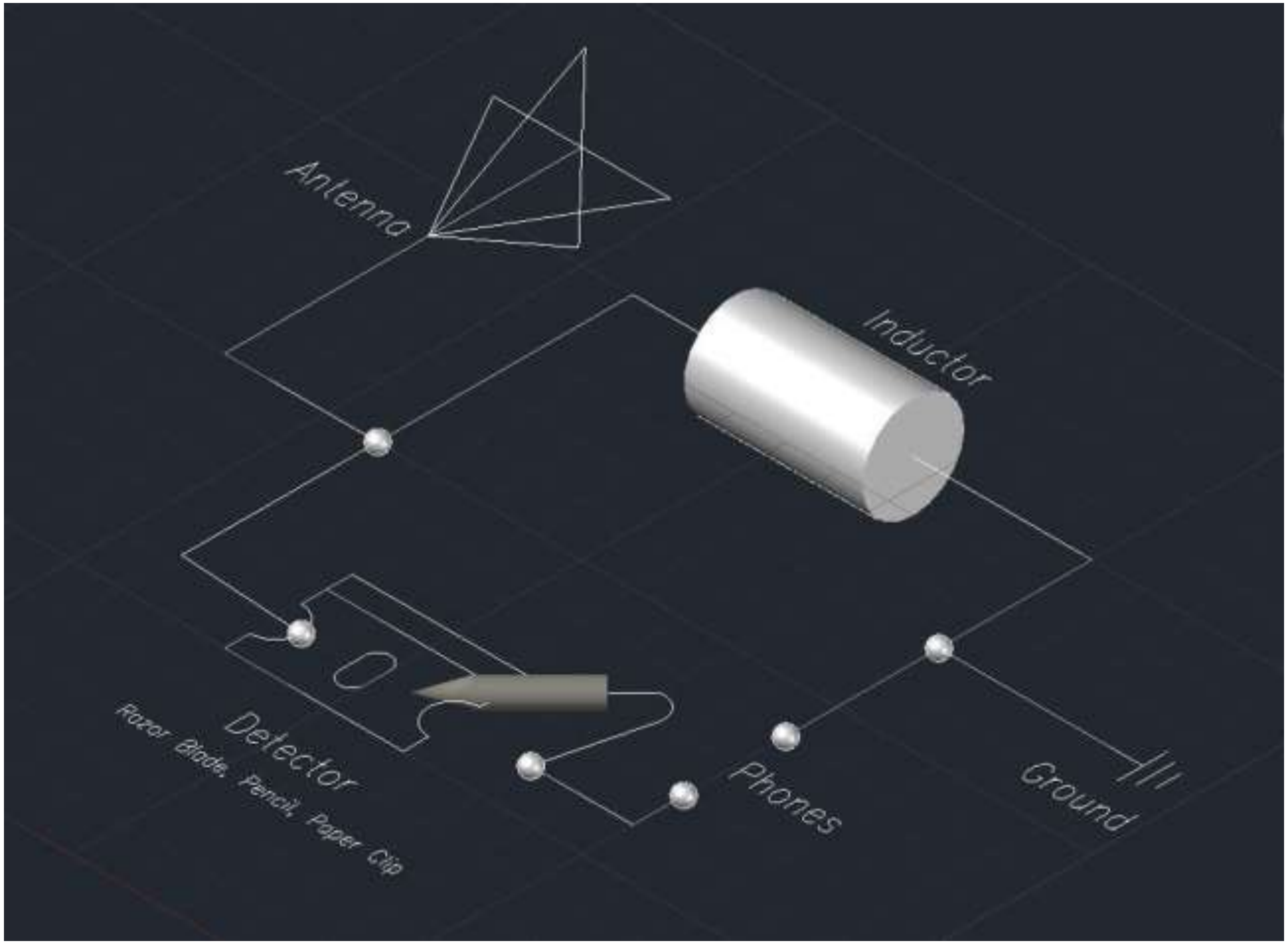
Antenna
Inductor
Detector
Razor Blade, Pencil, Paper Clip
Phones
Ground

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) **International Patent Classification** ^{6} :
H02K 19/18, 21/20, 21/36

A1

(11) **International Publication Number:** **WO 95/08210**

(43) **International Publication Date:** 23 March 1995 (23.03.95)

(21) **International Application Number:** PCT/IB94/00296

(22) **International Filing Date:** 12 September 1994 (12.09.94)

(30) **Priority Data:**
248647 13 September 1993 (13.09.93) NZ

(71) **Applicant** *(for all designated States except US)*: HIGHVIEW DEVELOPMENTS LIMITED

(72) **Inventor; and**
(75) **Inventor/Applicant** *(for US only)*: DEPALMA, Bruce [US/NZ];

(74) **Agents:**

(81) **Designated States:** AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LT, LU, LV, MD, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) **Title:** A HOMOPOLAR GENERATOR

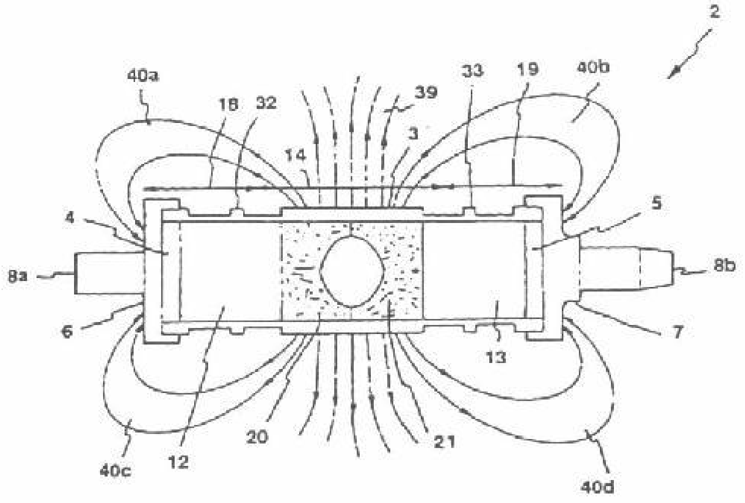

(57) **Abstract**

A single piece homopolar generator (1) for use alone or in combination with a like or known generators comprising: current collectors co-operating with a rotor body (2) wherein the rotor body (2) comprises: at least one electrically conductive member (3), a central zone (14) between the zones (18, 19), at least two actually aligned magnets (12, 13), arranged so that the poles (20, 21) of the magnets (12, 13) oppose to produce flux lines which pass through and exit the central zone (14) of said rotor body (2) in a direction radial to the axis

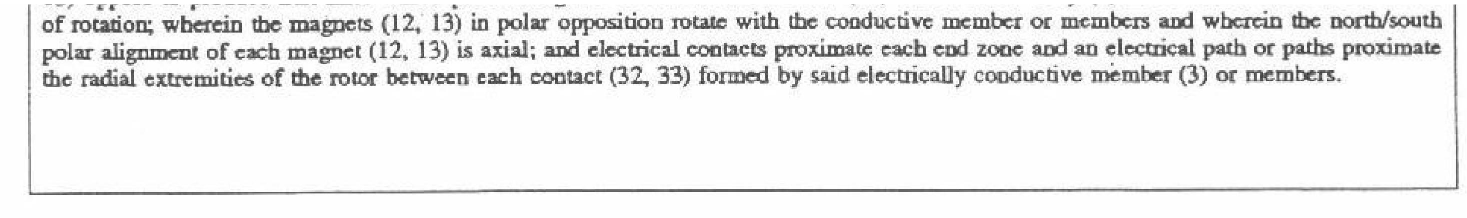

of rotation; wherein the magnets (12, 13) in polar opposition rotate with the conductive member or members and wherein the north/south polar alignment of each magnet (12, 13) is axial; and electrical contacts proximate each end zone and an electrical path or paths proximate the radial extremities of the rotor between each contact (32, 33) formed by said electrically conductive member (3) or members.

A HOMOPOLAR GENERATOR

BACKGROUND

[5] The present invention relates to an electric generator. The generator described herein is referred to as a quadrapole generator because of the four distinct magnetic poles involved in the machine.

PRIOR ART

[10] In 1831 Michael Faraday performed the initial experiments which resulted in the discovery of the dynamo. In one of his experiments a copper disc was secured to a cylindrical magnet with paper intervening the two. The poles of the magnet were aligned along the axis of the copper disc. Wires of a galvanometer brushed the centre and circumference of the copper
[15] disc respectively. It was discovered that upon rotation of the copper disc and magnet, an electrical potential was created between the terminals of the galvanometer. This simple construction is known as a homopolar generator. Importantly, this experiment revealed that a potential difference was created across the copper disc when it was rotated through a magnetic field, irrespective of whether the magnet was rotated with the copper disc or[20] remained stationary.

Another early generator was the two piece design by Faraday where a conducting disc is revolved adjacent to the poles of fixed magnets.

[25] Homopolar generators produce low voltages at high currents. In the later 1800's these unipolar generators were used in metal reduction and plating applications where high currents are required. In the early 1900's however, the development of commutated DC and AC generators which could develop higher voltages at lower operating speeds led to the decline in use of homopolar generators, except for specialised applications.

[30]Another prior art generator involves the combination of two one piece homopolar generators similar to that designed by Michael Faraday in 183 1 and mounted in common on a central supporting conducting shaft. This generator was constructed with magnet poles aligned in opposition so that they were voltage additive between two current collector rings encircling
[5] the centers of the tandem rotating magnets. The current generated by this generator flows radially inward in a conducting disc located centrally within and co-axially disposed within one magnet through the connecting axle and then radially outward in a disc co-axially disposed within the second magnet. The mechanism of voltage generation in this generator was similar to that in the previously described one piece Faraday homopolar generator[10] wherein the magnetic flux lines within the magnets are perpendicular to the conducting disc co-rotating with and centrally disposed within each permanent magnet assembly.

One disadvantage of this generator is that the current output is limited by the diameter of the supporting axle. If the axle is larger, it is necessary to have larger holes in the magnets [15] through which reverse flux may pass. The necessity for the hole through the magnets and the reverse flux problem reduces magnet strength and voltage.

The copper discs of this generator were subdivided into two spirals to produce a self magnetising effect with current withdrawal which counteracted partially the high internal [20] resistance ofthe long current path through the two copper spirals and the axle. The 50mm diameter shaft limited current output to four kiloamperes. Above this current level excessive heating would occur.

Another disadvantage of this prior art generator is that the dumb-bell shaped rotor lacks[25] rigidity compared with the rotor of the present invention to be described below. This affects ease of balancing the rotor.

DESCRIPTION OF INVENTION

According to the present invention there is provided a single piece homopolar generator which has one moving part, the rotor and in which the desired electrical potential is produced

[5] without the mutual interaction of a second member (stator). This generator includes an electrically conductive member such as a cylindrical tube having two magnets therein which) when the generator is in operation, rotate with the tube. The cylindrical version of this generator, known as a Quadrapole, is not an immediately apparent development of the original Faraday 'one-piece' axially rotated magnet experiment since the vector directions of

[10] the (radial) magnetic flux lines and axially flowing electrical current are interchanged in their respective directions in comparison to the previously described Faraday Disc experiment.

The one-piece, rotor only version of what is presently known as a cylindrical homopolar generator has not hitherto previously been known.

[15]

Throughout the specification the term 'homopolar' can be taken to mean the repulsion of like magnetic fields, i.e. N-N or S-S which can alter the direction of magnetic flux lines and produce a radial pattern in the central zone of the rotor.

[20] Within the last ten years certain materials such as rare earth, Neodymium-Iron-Boron $Nd_2Fe_{14}B$), and Samarium-Cobalt (SaCo) permanent magnets, and Niobium-Tin or Niobium-Titanium superconductive magnet wire have become available. With these materials it becomes practical to fabricate magnetic structures impossible to realise with iron and copper wire. The configuration of the present invention exploits the advantages thatmodern magnetic materials provide[25].

It is an object of the present invention to provide an improved generator or to at least provide the public with a useflil choice.

[30] In one broad form of the invention there is provided an electric generator comprising:

a single piece homopolar generator for use alone or in combination with like or known generators comprising:

> current collectors co-operating with a rotor body wherein the rotor body comprises; at least one electrically conductive member,
>
> [5] a central zone between end zones,
>
> at least~two axially aligned magnets, arranged so that like poles of the magnets oppose to produce flux lines which pass through and exit the central zone of said rotor body in a direction radial to the axis of rotation;
>
> wherein the magnets in polar opposition rotate with the conductive member or members and[10] wherein the north/south polar alignment of each magnet is axial; and electrical contacts proximate each end zone and an electrical path or paths proximate the radial extremities of the rotor and between each contact formed by said electrically conductive member or members.

[15] Preferably there is one electrically conductive member comprising a cylindrical tube and the magnets are permanent magnets which are permanently fixed with respect to the tube. The poles of the magnets are preferably orientated co-axially with the axis of the tube and the tube is preferably rotated at high speed.

[20] Alternatively, the pefformance characteristics of the generator may be achieved by use of alternative structural arrangements which receive and retain the magnets and other rotor components. For example, it would be possible to use an array of radially disposed conductors such as rods providing electrical paths connecting electrical contacts on the rotor Alternatively, the generator may comprise concentric cylinders or a nest of cylindrical tubes [25] whose axes are parallel.

According to another embodiment there is provided an electnc generator as hereinbefore described including an electrically conductive compensation tube provided about said central zone and spaced apart therefrom, an end of said compensation tube being electrically connected to the contact adjacent thereto, the other end of said compensation tube being electrically connected to a generator output terminal.

The advantages of the homopolar generator according to the present invention include the [5] following:

> a solid magnet across the frill internal diameter of the tube providing higher and uniform magnetic field and the elimination of current flow through the magnet and an increased current carrying capacity now only limited by the size of the current collectors.

[10] In another broad form the present invention comprises;

> a rotor for use with a generator as hereinbefore described, the rotor comprising; a rotor body comprising;
>
> at least one electrically conductive member, a central zone between end zones,
>
> [15] at least two axially aligned magnets arranged so that like poles of the magnets oppose to produce flux lines which when the rotor is in use pass through and exit the central zone of said rotor body in a direction radial to the axis of rotation;
>
> wherein the magnets in polar opposition rotate with the conductive member or members and wherein the north/south polar in alignment of each magnet is axial; and
>
> [20] electrical contacts proximate each end zone and an electrical path or paths proximate the radial extremities of the rotor between each contact formed by said electrically conductive member or members.

In an alternative form, the invention comprises;

> [25] a current collector for use with a generator as hereinbefore described, the current collector comprising;

a two part body one of which pans is detachably attached to the other part, wherein when the two parts are attached, a central bore is formed which receives a rotor,

means located at least partially within said bore for creating a seal between the rotor and the [30] outside of the bore, wherein said means forms a circumferential recess within the bore in which an electrically conductive material is located and which is in electrical contact with electrical contacts on a rotor in the generator, wherein, the space between the base of the recess and the electrical contacts of rotor is filled with liquid metal or eutectics providing an electrical path between the rotor and the current collector.

[5] Preferably the electrical contacts are machined into the rotor and comprise a circumferential nng on each end zone, providing an electrical connection between the surface of the cylindrical tube and the conductive liquid metal or eutectics.

[10] In a flinher form according to the system aspect the present invention comprises:

a system for generating electricity using a single piece homopolar generator; the system comprising;

the single piece generator, having one moving part, the rotor, and in which the desired electrical potential is produced without mutual interaction of a stator,

[15] a power source to drive the generator,

a field of energy influence within which the generator is situated and with which the generator interacts,

wherein the interaction between the generator and the field influences the output of the generator by supplementing energy input to the generator from said power source.

[20] A rotor may be constructed of multiple concentric conducting cylinders. In the zone of zero or low magnetic flux pertaining to the region encircling the centers of rare-earth magnets or super-conducting solenoids, muitiple sliding liquid metal contacts may be established thus enabling a series connection of the portions of the concentric conducting cylinders in the [25] voltage generating region between the opposing poles of the rotating magnets contained within the nested cylinders. Voltage addition by connecting a series of concentric conducting cylinders in a one-piece cylindrical homopolar generator has not previously been known.

Because of the existence of a region of zero radial magnetic field in a zone encircling the [30] center of a cylindrical permanent magnet, i.e. the neutral zone, current extraction from the rotating member is taken at this point. Current extraction by means of a liquid metal sliding contact in this zone eliminates any electro-magnetic forces which might act to disturb the liquid metal contact during current extraction. A zone of zero magnetic flux also eliminates electrical currents circulating transversely through the conductive body of a current collector [5] because of inhomogeneities in voltage across the width of the liquid metal sliding contact.

With the Quadrapole, the magnets are arranged NSSN or SNNS and the fact that the magnetic flux lines emerge radially from the center of the conducting cylinder is because of the mutual repulsion of opposing directions of like (homopolar) force.

[10] In the conventional two-piece cylindrical homopolar machine, magnetic flux lines are caused to emerge radially from the central voltage generation segment of the cylindrical rotating member by fixed iron pole pieces which encircle the rotating cylinder and form part of a stator structure which closes the magnetic flux paths in fixed external loops back to each axle [15] of the machine. The two piece closed path construction makes no use of the mutually repulsive effect of homopolar magnetic fields because in the closed path construction the magnetic field internal to the cylinder is directed to flow radially outward by low magnetic reluctance external pole pieces.

[20] Without the provision of external pole pieces and a closed magnetic flux path, the attainable magnetic field strength within such a machine would be so low as to render the machine not suitable for commercial application. Rare earth high strength permanent magnets make it possible to obtain high strength and usefiil radially directed magnetic flux lines without closed magnetic flux paths. The radially directed flux arises from mutual repulsion of homopolar [25] fluxfields.

The key requirements of the cylindrical one-piece homopolar generator as herein described are that all parts of the rotor including the magnets must rotate together and there is no closure of the magnetic flux paths by fixed ferromagnetic yokes, - stators.

[30] If the permanent rare-earth magnets are replaced with super-conducting electrical solenoidal coils, the coils must rotate with the cylinder. The magnetic fields produced when they are cooled and energised must be poled NSSN or SNNS and the spacing of the coils adjusted to produce radial flux lines perpendicular to the central voltage producing segment, (of the [5] rotating conductive cylinder enclosing and supporting the magnet solenoids). The mutual repulsion of homopolar flux fields is employed to create radially diverging flux lines in the central zone

The present invention in all its forms will now be described in more detail according to a [10] preferred but non-limiting embodiment and with reference to the accompanying illustrations wherein:

Figure 1: shows a long sectional view through a generator rotor according to a preferred embodiment;

Figure 2: [15] shows an exploded view of the generator incorporating the rotor of figure 1 according to a preferred embodiment of the invention;

Figure 3: shows an isometric exploded view of one current collector for use with the generator;

Figure 4: shows an assembled view of the current collector of figure 3,

Figure 5: [20] shows the rotor of figure 1 seated in part of current collectors and showing the relationship of the rotor to the electrical contacts and seals.

Figure 6: shows a long section through the generator rotor of figure ~ with lines of magnetic flux indicated,

Figure 7: shows an embodiment of the rotor according to a preferred embodiment of [25] the present invention including magnetic compensation,

Figure 8. shows an isometric view of the completed generator with output terminals according to a preferred embodiment of the invention, and

Figure 9: shows the generator of figure 8 from a rear view driven via a drive belt by a drive motor.

Referring now to figure 1 there is shown a sectional view ofa rotor 2 for use with a generator I (see figure 2) according to a preferred embodiment of the invention. Rotor 2 comprises an electrically conductive cylindrical tube 3 which may include end plates 4 and 5. End plates 4 and 5 each preferably comprise an aluminium disc. Fixed to the ends of [5] cylindrical tube 3 and covering plates 4 and 5 respectively are non magnetic stainless steel end caps 6 and 7. End caps 6 and 7 terminate in shaft ends Ba and Sb respectively. End caps 6 and 7 are preferably affixed to the cylindrical tube 3 by means of screw threads 6a and 7a located on end caps 6 and 7 respectively. Alternatively, end caps 6 and 7 may be fixed via an internal thread (not shown) on inner surface 3a of cylindrical housing 3 or fixed with a glue [10] or friction fitted. The electrically conductive tube 3 of rotor 2 may comprise as an alternative hollow members such as but not limited to a sphere or cube. Shaft ends Ba and Sb may be integral with or are detachably connected to end caps 6 and 7 and are co-axial with cylindrical tube 3. Rotor shaft ends Ba and Sb are, when in situ, surrounded by bearing assemblies 9a and 9b (see figure 2) respectively allowing free rotation of the cylindrical tube [15] 3 upon rotation of rotor shaft B. Once bearings 9a and 9b (see figure 2) are fitted to the rotor shaft ends Ba and Sb, the bearings are contained within stationary supports 10 and 11 (see figure 2). Cylindrical tube 3 rotates freely about its axis when driven via shaft ends Sa or Sb. Fixed to cylindrical tube 3 are permanent magnets ~2 and 13 which rotate with the cylindrical tube 3 when the generator operates.

[20] The magnets are onented so that their like poles oppose (in this case the north poles) resulting in magnetic flux lines being directed radially outwardly from central zone 14 of cylindrical tube 3. Throughout the specification the term 'central zone' can be taken to mean that region in the centre of the rotor wherein the output voltage is generated when the rotor [25] is rotated. A cavity 15 is formed between two shaped cast iron pole pieces 20 and 21 between magnets 12 and ~3.

As well as havuig a central zone 14, cylindrical tube 3 includes end zones 18 and 19 wherein the central zone 14 is disposed between the end zones.

[30] Referring to figure 2 there is shown an exploded view of the generator of figure I including the rotor 2 of figure 1, current collectors 22 and 23 and their interrelationship with the cylindrical tube 3.

[5] Figure 2 also shows connected to electrically conductive cylindrical tube 3 end caps 6 and 7 terminating in rotor shaft ends Sa and Sb respectively.

Current collectors 22 and 23 both of which are identical are located at contacts 32 and 33 in end zones 18 and 19 respectively of cylindrical tube 3. Each collector is located along [10] cylindrical tube 3 in a neutral region of each end zone in a flux field where the concentration of flux is low. As current collectors 22 and 23 are identical, only current collector 22 will be described in detail and with reference to figure 3 below to avoid duplication.

As the rotor of the generator I is rotated, a voltage potential develops between contacts 32 [15] and 33. Power output is drawn from the generator via the two current collectors 22 and 23. The mercury provides the electrical contact between the cylindrical tube 3 and current collectors 22 and 23 which are in electrical contact with output terminals 24 and 25 (see figure 8). It has been found that using the pole configuration shown in figure 1 that 3 to 4 times the voltage output of a standard homopolar generator may be obtained using magnets [20] having the same field strength.

Referring now to figure 3 there is shown an exploded isometric view of a typical current collector. Figure 3 shows an enlargement of the current collector 22 of figure 2 comprising two parts 27 and 28 which preferably are symnietrical about their plane of separation and [25] which together form a contact body housing 26 (see figure 4). Body parts 27 and 28 are preferably manufactured from a high conductivity material eg. copper. To form current collector 22 each of parts 27 and 28 which include semi circular bores 29a and 29b respectively receive substantially semi circular and preferably plastic sealing elements 30a, 30b, 30c and 30d which when in situ and mated together form circular recess 31 (see figure 4). [30] In use, recess 31 receives liquid mercury which provides the electrical contact with contact 32 on cylindrical tube 3. Contact 32 rotates in circular recess 31. Body parts 27 and 28 are mated together by means of bolts or locking screws 34 and 35.

Figure 4 shows the current collector 22 of figure 3 assembled. When body parts 27 and 28 [5] are mated together a seal is created by plastic seal 30 formed by sealing elements 30a, 30b, 30c and 30d thereby preventing the escape of liquid mercury during operation of the generator. It will be appreciated that plastic seal 30 can be an integral member as an alternative to formation by separate elements. There is a small clearance between sealing elements 30a, 30b, 30c, 30d and the rotor 2. Screw threads are machined on the sealing [10] lands 58, 59, 60, 61 (see figure 1) of the cylindrical tube 3, so that any leakage of mercury is returned when the rotor is rotating to recess 3 1 - the electrical contact zone.

As an alternative to use of liquid metal contacts, electrical brushes which are widely used in electrical machinery may be used. However, the generator, according to the present [15] invention produces low voltage at very high currents which is generaijy unsuitable for solid sliding contacts. The preferred contacts are conducting liquid metals such as mercury or eutectics such as sodium-potassium or gallium-indium. The use of liquid metal electrical contacts gives the advantage of lower electrical resistance, lower mechanical friction and low wear.

[20] Electrical contacts 32 and 33 are machined into the cylindrical tube 3. Preferably a number of annular ridges 32a and 33a may be formed on contacts 32 and 33 respectively. Contacts 32 and 33 are when surrounded by current collectors 22 and 23 separated by a very small clearance between the conductive surface of recess 3 1 (in the case of contact housing body 26)[25]. There is a corresponding arrangement in contact assembly 23. Preferably that clearance for each contact is 0.5mm or less.

Each of current collectors 22 and 23 include capillary lines. As the capi]lary line arrangements for current collectors 22 and 23 are the same, the following description will [30] relate to the capillary line for current collector 22 shown assembled in figure 4. Referring to figure 3 it can be seen that body part 28 of current collector 22 includes mercury reservoir 44 which feeds into capillary line 42 with flow of liquid metal into recess 31 being controlled by means of valve 46. In use, rotor 2 is rotated and then liquid metal is introduced from reservoir 44 via the capillary line 42 to the space between the circumferential contact 32 [5](see figure 1) and recess 31 of current collector 22 (see figure 3).

Centrifugal forces and viscous drag cause liquid metal to be taken up on the contacts 32 and 33 of cylindrical tube 3 to form a circumferential ring of liquid metal bead encircling those contacts. Thus, for current collector 22 mercury is in contact with the surfaces within recess [10] 31 of contact body housing 26. Similarly for contact assembly 23. Because the liquid metal bead is held in place by a combination of centrifigal and viscous forces the clearance between contact 32 and recess 31 can be quite large (for instance; 2mm). Once the cylindrical tube is rotating and the liquid metal has been introduced, the apparatus will operate equally well either horizontally or vertically. For satisfactory operation the liquid [15] metal should wet the inner surfaces of recess 3 1. In the case of mercury, to a achieve proper amalgamation it is preferred that a process is employed to remove oxide from the surface of the body parts 27 and 28 prior to introduction of the mercury.

Referring to figure 5 there is shown the rotor 2 of figure 1 seated in part of current collectors [20] 22 and 23. It can be seen that contacts 32 and 33 locate in recesses 31 and 37 respectively. Recess 31 is formed by plastic seal elements 30a, 30b and recess 37 is formed by seal elements 38a and 38b. Seal elements 30a, 30b, 38a and 38b engage respectively sealing lands SB, 59, 60 and 61 which have helical threads which urge any mercury that escapes recesses 31 and 37 back into those recesses when the rotor rotates.

[25] Referring now to figure 6 there is shown a long section view of the rotor 2 of figure 1 showing the disposition of the flux lines relative to the cylindrical tube 3. Cylindrical tube 3 is shown including permanent magnets 12 and 13. Magnets 12 and 13 are preferably permanent magnets and may either be conventional magnets, rare earth metal magnets or [30] super conducting magnets. The magnets 12 and 13 may each be formed from a plurality of magnetic elements or other magnetic material. As cylindrical tube 3 is preferably formed of a high strength, high electrical conductivity copper alloy the cylindrical tube can be rotated at very high speeds. The polar opposititíon configuration of magnets 12 and] 3 produce flux lines which pass through and exit the central zone of cylindrical tube 3 in a direction that is perpendicular to the axis of the cylindrical tube 3 in central zone 14. From figure 6 it can be seen that there is a concentration of radial flux in central region 14 as depicted by flux lines 39. Flux lines 40a, b, c and d are concentrated in end zones 18 and 19 as shown.

Rotation of the conductive cylindrical tube 3 with the magnetic flux generates a potential difference between contacts 32 and 33 (refer figure 1). The electric potential between terminals 32 and 33 is given by the relationship

$$E = 10^{-8} .v. Bn\ dl$$

where:

E is the potential difference between the contacts 32 and 33 in volts,

Bn is the flux density (in Gauss) normal to the surface of cylindrical tube 3, i.e. acting radially to the axis of rotation of the rotor.

l is the distance in cm between contacts 32 and 33; and

v is the tangential velocity of the surface of cylindrical tube 3 in cm/second.

Accordingly, I will effectively be the length of cylindncal tube 3 adjacent poles 16 and 17.

As the potential difference is proportional to the tangential velocity of the cylinder it is preferable that the cylindrical tube 3 be built as large as possible to achieve optimal output voltage, and be rotated as fast as possible, for example, up to 100,000 rpm or beyond if physical limits permit.

Due to the high currents generated by the generator, super conducting materials are particularly suitable to be incorporated in or used with cylindrical tube 3.

A generator of the type hereinbefore described can produce very high output currents (multiples of kiloamperes) at low voltages. The withdrawal of high electrical currents from the generator results in a magnetic field consisting of circular flux tines enclosing the central zone 14 of the cylindrical tube 3. A method of cancelling or at least minimising these effects will now be described with reference to figure 7 of the drawings.

Referring to figure 7 there is shown a schematic representation of a generator similar to that depicted in figure 1 including magnetic compensation means. In the embodiment shown a conductive compensation tube 48 is positioned about and spaced apart from, the central zone 14 of cylindrical tube 3. One end 49 of compensation tube 48 is electrically connected to contact 50 by conductor 5 1. The opposite end *52* of compensation tube 48 is connected to outward terminal 53 of the generator by conductive connection 54. Compensation tube 48 remains stationary while cylindrical body 3 rotates relative thereto. Compensation tube 48 produces compensatory circular magnetic flux which cancels the field generated by withdrawal of current. This works in the same manner as a coaxial cable; whereby equal currents flow in opposite directions thus the magnetic fields thereby produced cancel each other. The cancellation of the magnetic fields due to the high currents in the generator is important, since the field distortion (armature reaction) produced by these currents when uncompensated can limit the power output of the machine by altering the perpendicularity of the flux lines to the rotating cylindrical tube 3.

Referring now to figure B there is shown an isometric view of a filly assembled generator including output terminals 24 and 25 in communication with current collectors 22 and 23 with spacing block 56 therebetween. Outside current collectors 22 and 23 are bearing supports 10 and 11 which receive bearings 9a and 9b (see figure 2). Outside bearing support 11 is drive wheel 57.

Figure 9 shows the assembled generator of figure 8 with an electric motor 62 connected thereto via drive belt 63 which engages drive wheel 57.

It is thus seen that the present invention provides a generator having improved performance over known homopolar generators. Particularly, in this invention the arrangement of the magnets in polar opposition provides increased flux concentrations, increasing the output voltage of the generator for the same strength of magnet employed. The magnetic compensation method of the invention allows the generator to operate at high power outputs without substantial effect on the perpendicularity of the flux lines. Finally, contacts using liquid metal reduce the electrical and frictional losses of the generator.

Where in the foregoing description reference has been made to integers or components having known equivalents then such equivalents are herein incorporated as if individually set forth.

For example, it is to be appreciated that cylindrical tube 3 need not by cylindrical, although a cylindrical tube is preferred.

The magnets which are the source of flux must rotate integrally with the cylindrical tube 3 in polar opposition with flux lines parallel to the axis of rotation even though the output voltage is generated by a 900 curvature of these lines intersecting the rotating conductive cylindrical tube 3 No fixed external pole pieces or magnets can be used for the purpose of magnetic field enhancements.

As an example of the performance of the generator using known magnets, it could be expected that a voltage output of 1.5 V.D.C. could be achieved at a rotational speed of 12,000 rpm. A power output of 10KW maybe obtained from the machine by the withdrawal of 6,670 amperes of electrical current. At this current under 400 watts will be dissipated in the rotor as heat. A realistic determination of allowable current flow based on rotor heating would be 12,000 amperes, with 6,000 amperes taken from each side of the machine. Generator drive can be achieved by use of any applicable electrical, mechanical, internal combustion, water or wind power.

Although this invention has been described by *way* of example it is to be appreciated that improvements and/or modifications may be made thereto without departing from the scope or spirit of the invention, such as but not limited to:

use of low friction bearings, for example air bearings;

operating the device in a vacuum sealed environment to reduce windage drag;

modifications to the magnetic field to enhance the performance, utility and regulation of the generator.

The claims defining the invention are as follows:

1. A single piece homopolar generator for use alone or in combination with a like or known generators comprising;

current collectors co-operating with a rotor body wherein the rotor body comprises;
at least one electrically conductive member,

a central zone between end zones,

at least two axially aligned magnets, arranged so that like poles of the magnets oppose to produce flux lines which pass through and exit the central zone of said rotor body in a direction radial to the axis of rotation;

wherein the magnets in polar opposition rotate with the conductive member or members and wherein the north/south polar alignment of each magnet is axial; and

electrical contacts proximate each end zone and an electrical path or paths proximate the radial extremities of the rotor between each contact formed by said electrically conductive member or members.

2. A generator according to claim 1 wherein said electrically conductive member or members comprise(s) a cylindrical tube(s) and wherein the magnets in each said tube(s) are spaced apart and are disposed in axial alignment with the axis(es) of the tube(s).

3. A generator according to claim 2 wherein the rotor comprises one cylindrical tube and the generator includes current collectors which connect electrically with the cylindrical tube by sliding brush or liquid metal contacts or both, wherein the current collectors are each located along the tube in a neutral region of each end zone in a flux field where the concentration of flux lines is low.

4. A generator according to claim 3 wherein each current collector includes an element which provides a seal to prevent egress of liquid metal from a current collecting zone between the tube and each current collector.

5. A generator according to claim 4 wherein the niagnets are either superconductive solenoid repelling magnets or rare earth magnets such as Nd_2 Fe_{14} B or SaCo.

6. A generator according to claim 5 wherein the speed of the generator is within a range up to 100,000 rpm.

7. A generator according to claim 6 wherein the tube is supported by shafts extending from each end and which bear on fixed bearing supports.

8. A generator according to claim 7 wherein the cylindrical tube is formed from superconducting materials.

9. A generator according to claim 7 wherein the cylindrical tube is manufactured from Beryllium-Copper alloy.

10. A generator according to claim B or 9 wherein the brushes are carbon or copper graphite.

11. A generator according to claim B or 9 wherein the liquid metal is either mercury, sodium-potassium eutectic or gallium~indium eutectic as the conductive material.

12. A generator according to any of the foregoing claims wherein multiple generators are connected in senes.

13. A generator according to claim I wherein the generator is adapted with cooling means whereby a cooling liquid or gas is passed through the rotor and/or current collectors of the machine during operation.

14. A rotor for use with a generator as hereinbefore described, the rotor comprising;

a rotor body comprising;
at least one electrically conductive member,

a central zone between end zones,

at least two axially aligned magnets, arranged so that like poles of the magnets oppose to produce flux lines which when the rotor is in use pass through and exit the central zone of said rotor body in a direction radial to the axis of rotation,

wherein the magnets in polar opposition rotate with the conductive member or members and wherein the north/south polar alignment of each magnet is axial;

and electrical contacts proximate each end zone and an electrical path or paths proximate the radial extremities of the rotor between each contact formed by said electrically conductive member or members.

15. A rotor according to claim 14 wherein said electrically conductive member or members comprise(s) cylindrical tube(s) and wherein the magnets in each said tube(s) are spaced apart and are disposed in axial alignment with the axis(es) of the tube(s).

16. A rotor according to claim 15 wherein the rotor comprises one cylindrical tube.

17. A rotor according to claim 16 wherein the contacts are circumferential about the cylindrical tube.

18. A rotor according to claim 17 wherein the electrical contacts include annular ridges.

19. A rotor according to claim 18 wherein the central zone includes two shaped cast iron pole pieces forming a cavity therebetween.

20. A current collector for use with a generator as hereinbefore descnbed, the current collector comprising,

a two part body one of which pans is detachably attached to the other part, wherein when the two parts are attached a central bore is formed which receives a rotor, means for fitting at least partially within said bore to create a seal between the rotor and the outside of the bore wherein said means allows the formation of a circumferential recess within the bore in which an electrically conductive material is located and which is in electrical contact with electrical contacts on the rotor.

21. A current collector according to claim 20 wherein the two parts are symmetrical about their line of separation.

22. A current collector according to claim 21 wherein said sealing means comprises a plastic insert having four elements two of which engage one part of the current collector and two of which engage the other part.

23. A current collector according to claim 22 wherein the electrically conductive material is mercury or liquid metal eutectics.

24. A current collector according to claim 23 wherein one part of the current collector includes a passage in communication with the recess in the bore and which receives an outlet of a mercury or liquid metal eutectic reservoir, the reservoir including a valve which regulates the flow of said mercury or liquid metal eutectic from said reservoir via said passage into said recess.

25. A compensation tube disposed concentrically about the cylindrical tube of the rotor as hereinbefore described and which produces compensatory circular magnetic flux.

26. A compensation tube according to claim *25* wherein the compensation tube is connected to a first contact on the rotor and the other is connected to an output terminal of a generator as hereinbefore described.

27. A system for generating electricity using a single piece homopolar generator; the system comprising;

the single piece generator, having one moving part, the rotor, and in which the desired electrical potential is produced without mutual interaction of a stator, a power source to drive the generator, a field of energy influence within which the generator is situated and with which the generator interacts, wherein the interaction between the generator and the field influences the output of the generator by supplementing energy input to the generator from said power source.

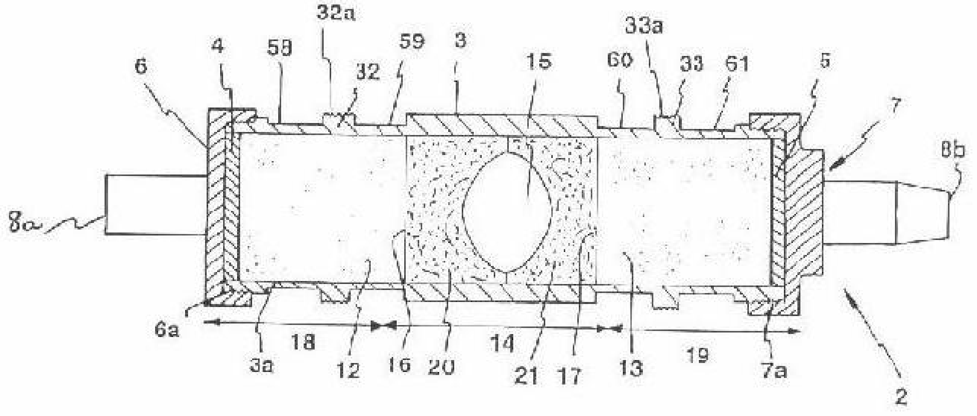
32a
33a
4
58
59
3
15
60
33
61
5
6
32
7
8b
8a
6a
18
14
19
3a
12
16
20
21
17
13
7a
2

FIGURE 1

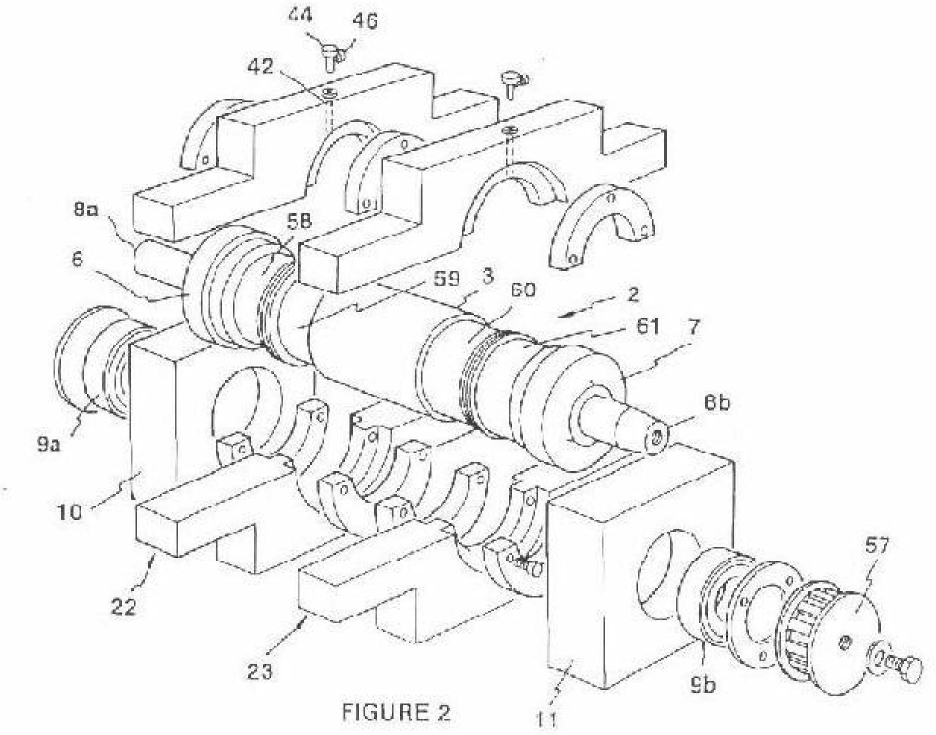
44
46
42
8a
58
6
59
3
60
2
61
7
8b
9a
10
22
23
57
9b
11

FIGURE 2

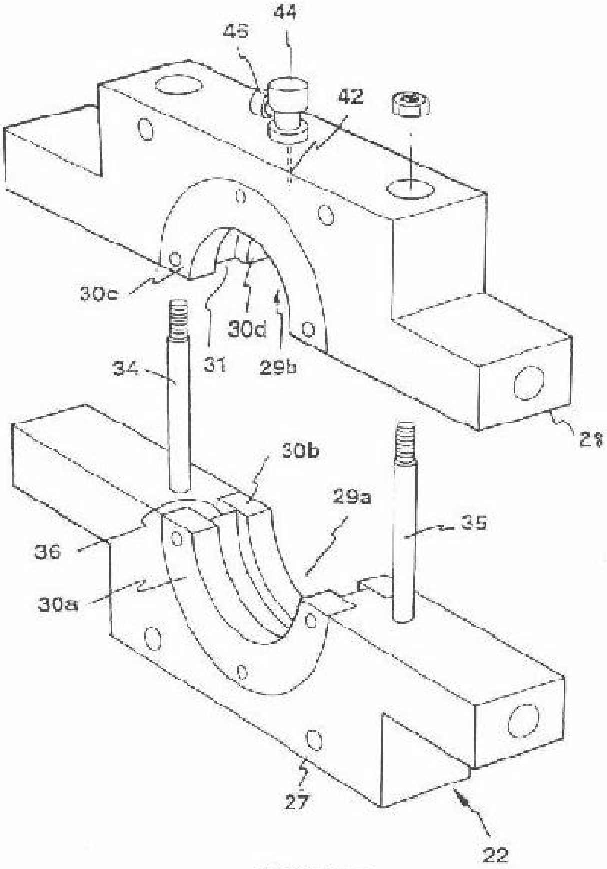
44
46
42
30c
30d
31
29b
34
28
30b
29a
35
36
30a
27
22

FIGURE 3

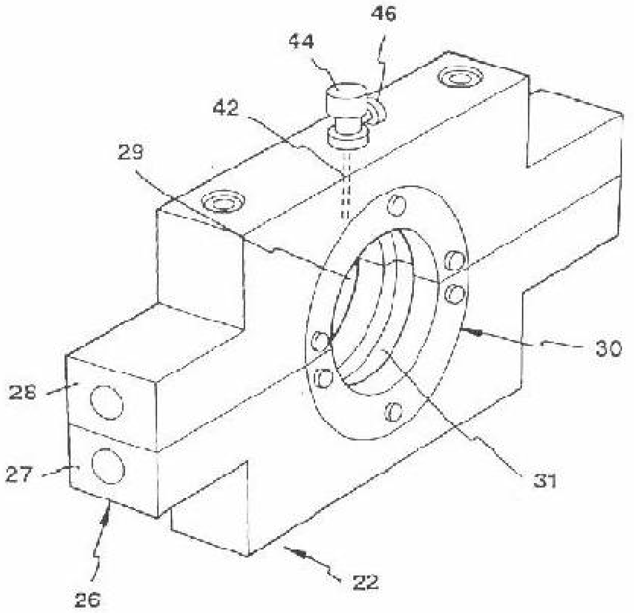
46
44
42
29
30
28
27
31
26
22

FIGURE 4

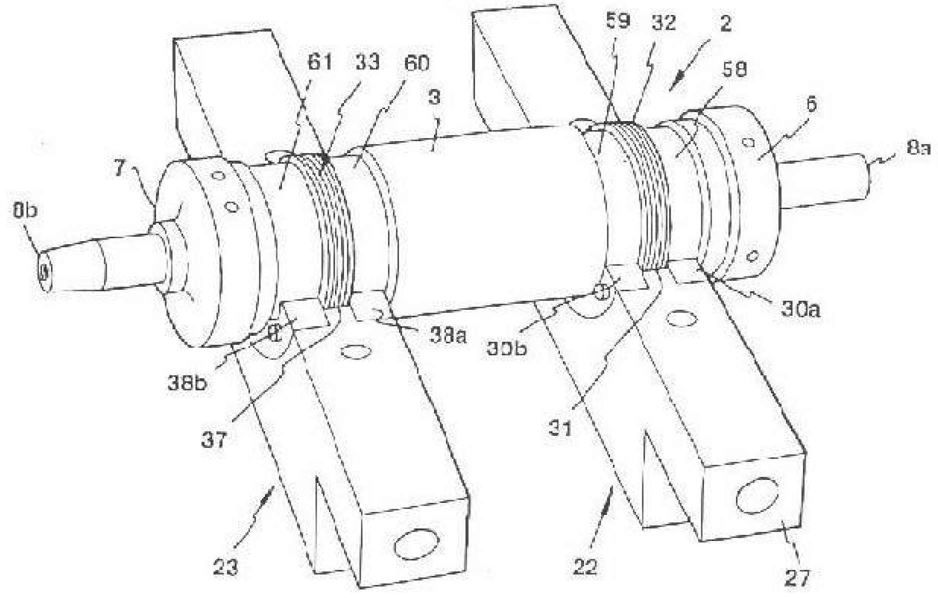
59
32
2
61
33
60
58
3
6
7
8a
8b
30a
38a
30b
38b
37
31
23
22
27

FIGURE 5

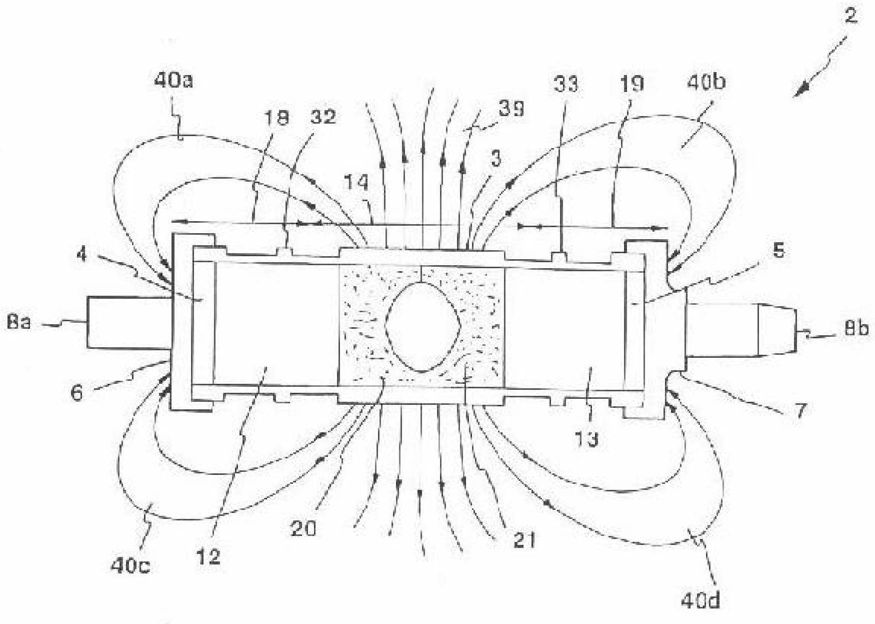
2
40a
18
32
14
39
3
33
19
40b
4
5
8a
8b
6
7
13
12
20
21
40c
40d

FIGURE 6

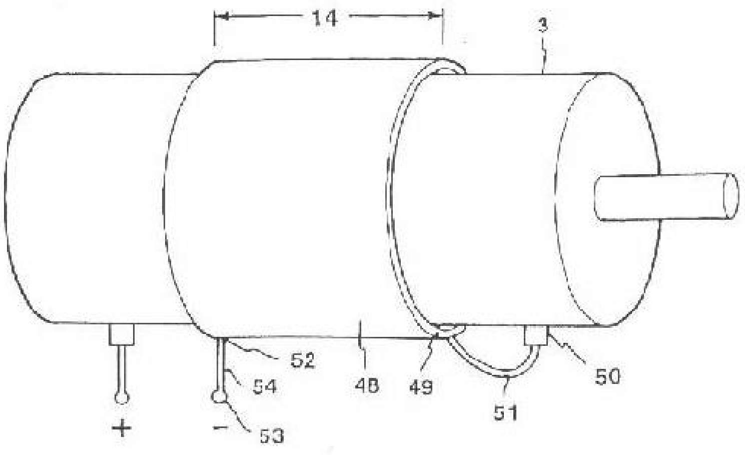
14
3
52
54
48
49
50
51
+
−
53

FIGURE 7

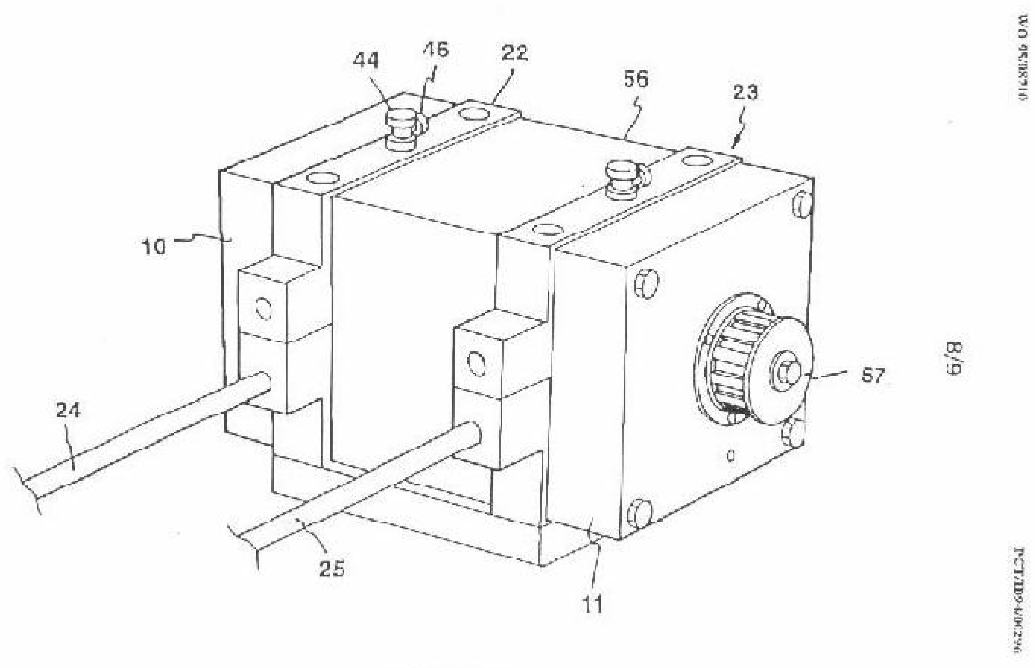
44
46
22
56
23
10
24
25
11
57
WO 95/18716
8/9
PCT/IB94/00296

FIGURE 8

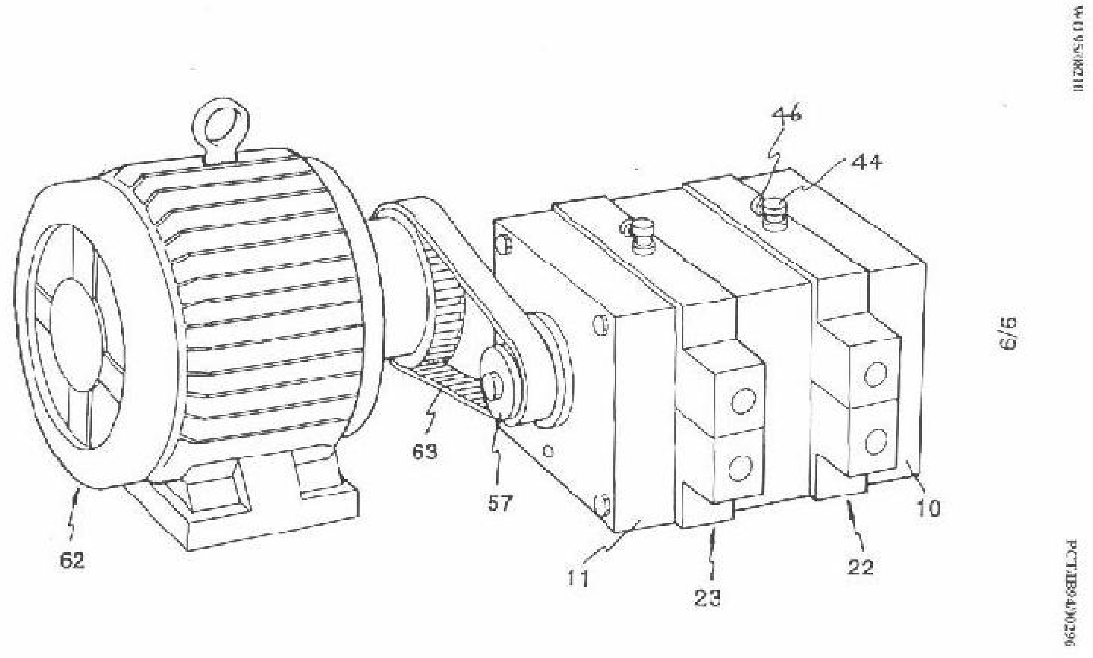
46
44
63
57
62
11
23
22
10
WO 95/18716
9/9
PCT/IB94/00296

FIGURE 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB 94/00296

A.	**CLASSIFICATION OF SUBJECT MATTER**
Int. Cl.^{6} HO2K 19/18, 21/20, 21/36	
According to International Patent Classification (IPC) or to both national classification and IPC	
B.	**FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)
IPC : HO2K 19/18, 19/20, 21/20, 21/36, 1/22, 1/24, 1/27

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
AU : as above

Electronic data base consulted during the international search (name of data base, and where practicable, search terms used)
DERWENT
JAPIO

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	US,A, 4399381 (CHABRERIE) 16 August 1983 (16.08.83)	
A	US,A, 3668447 (HAYASAKA) 6 June 1972 (06.06.72)	
A	EP,A2, 0343457 (SIEMENS AKTIENGESELLSCHAFT) 29 November 1989 (29.11.89)	
A	FR,A1, 2603433 (HINDRE) 4 March 1988 (04.03.88)	

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

*	Special categories of cited documents :
"A"	document defining the general state of the art which is not considered to be of particular relevance
"E"	earlier document but published on or after the international filing date
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O"	document referring to an oral disclosure, use, exhibition or other means
"P"	document published prior to the international filing date but later than the priority date claimed
"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&"	document member of the same patent family

Date of the actual completion of the international search 9 January 1995 (09.01.95)	Date of mailing of the international search report 17 Jan 1995 (17.01.95)
Name and mailing address of the ISA/AU AUSTRALIAN INDUSTRIAL PROPERTY ORGANISATION PO BOX 200 WODEN ACT 2606 AUSTRALIA Facsimile No. 06 2853929	Authorized officer **E. PERRIS** Telephone No. (06) 2832167

Form PCT/ISA/210 (continuation of first sheet (2)) (July 1992) copmak

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB 94/00296

Category*	Citation of document, with indication, where appropriate of the relevant passages	Relevant to Claim No.
A	FR,A, 1114142 (SOCIETE D'APPLICATIONS ET DE FABRICATIONS INDUSTRIELLES) 9 April 1956 (09.04.56)	
A	FR,A, 1114167 (SOCIETE D'APPLICATIONS ET DE FABRICATIONS INDUSTRIELES) 9 April 1956 (09.04.56)	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB 94/00296

Box I **Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)**

This international search report has not established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II **Observations where unity of invention is lacking (Continuation of item 2 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

(a) Claims 1-19

Claims 1-19 define a single piece homopolar generator having a rotor with axially aligned magnets so that the magnetic poles oppose to produce flux lines.

(b) Claims 20-24

Claims 20-24 are directed to a current collector with a two part body defining a central zone that receives a rotor.

(c) Claims 25-26

Claims 25-26 define a compensatory tube which produces a compensatory circular flux.

(d) Claim 27

Claim 27 defines a system (any system) which uses a single piece homopolar generator.

Since the abovementioned groups of claims do not share any of the technical features identified, a "technical relationship" between the inventions, as defined in PCT rule 13.2 does not exist. Accordingly the international application does not relate to one invention or to a single inventive concept.

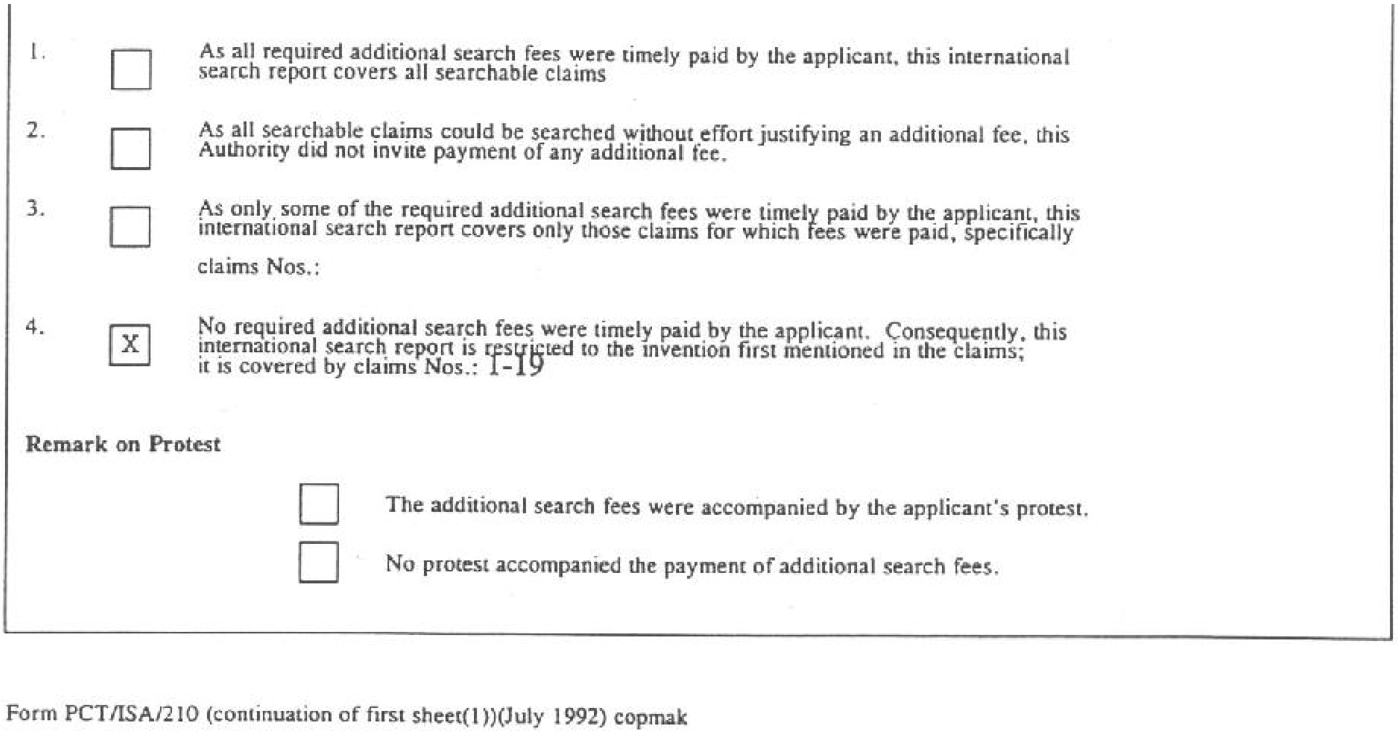

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-19

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet(1))(July 1992) copmak

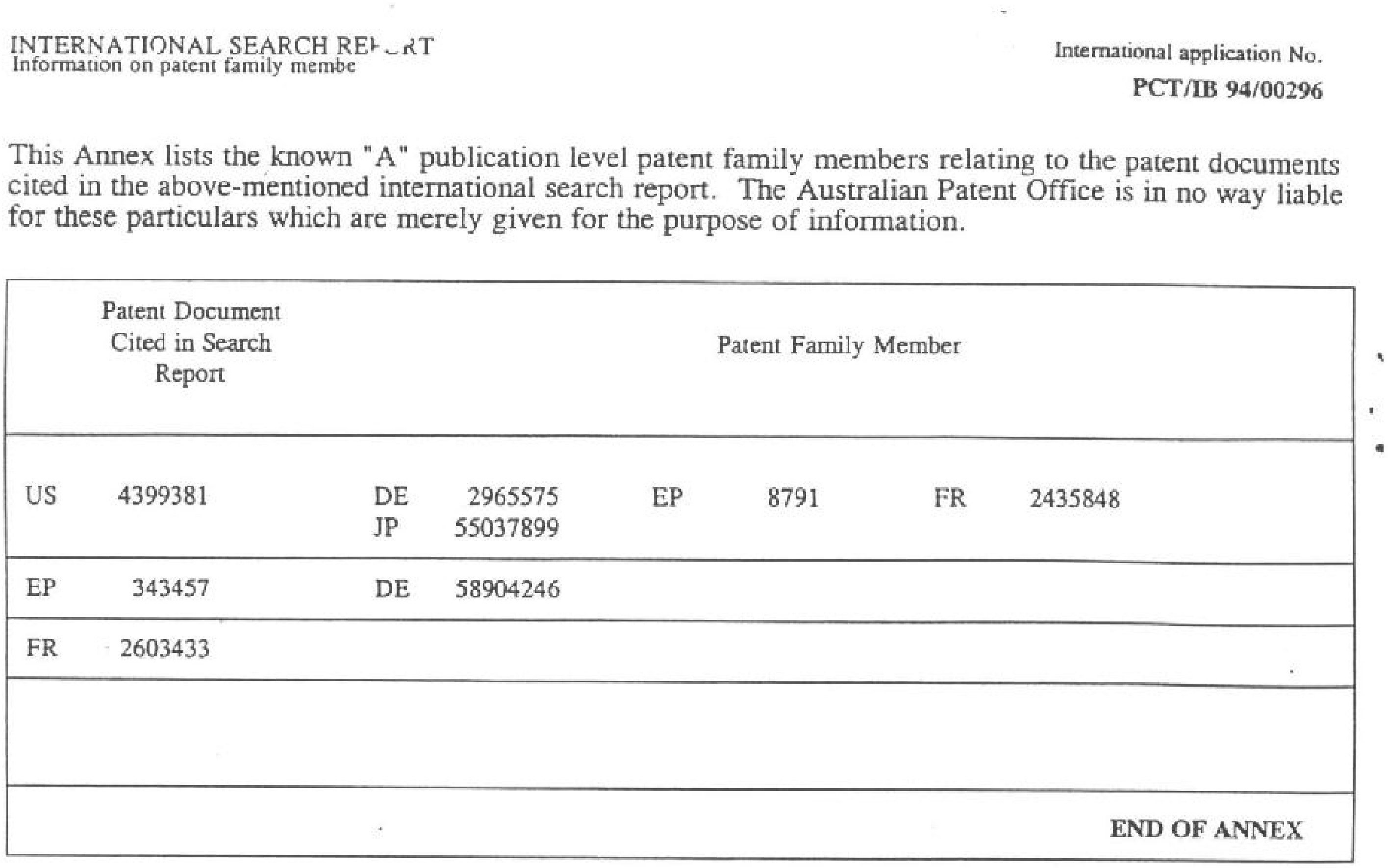

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/IB 94/00296

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report	Patent Family Member		
US 4399381	DE 2965575 JP 55037899	EP 8791	FR 2435848
EP 343457	DE 58904246		
FR 2603433			
			END OF ANNEX

Back to Homepage

US 20150102675A1

(19) **United States**

(12) **Patent Application Publication**

Dinwiddie et al.

(10) **Pub. No.: US 2015/0102675 A1**

(43) **Pub. Date: Apr. 16, 2015**

(54) POWER RECEIVER FOR EXTRACTING POWER FROM ELECTRIC FIELD ENERGY IN THE EARTH

(71) Applicant: **Earth Energies, Inc.**, Johns Creek, GA (US)

(72) Inventors: **John Dinwiddie**, Cary, NC (US); **Terry L. Wright**, Suwanee, GA (US); **David R. Ames**, Johns Creek, GA (US)

(21) Appl. No.: **14/509,772**

(22) Filed: **Oct. 8, 2014**

Related U.S. Application Data

(60) Provisional application No. 61/889,894, filed on Oct. 11, 2013.

Publication Classification

(51) **Int. Cl.**

H02J 4/00 (2006.01)

(52) **U.S. Cl.**

CPC *H02J 4/00* (2013.01)

(57) ABSTRACT

A resonant transformer connected between a ground terminal and elevated terminal draws current from the earth's electric field through a primary winding of the transformer. An impulse generator applies a high voltage impulse to the primary winding of the resonant transformer to cause current to flow from the ground terminal through the primary winding. The flow of current through the primary winding of the resonant transformer induces a current in the secondary winding, which may be converted and filtered to a usable form, e.g. 60 Hz AC or DC.

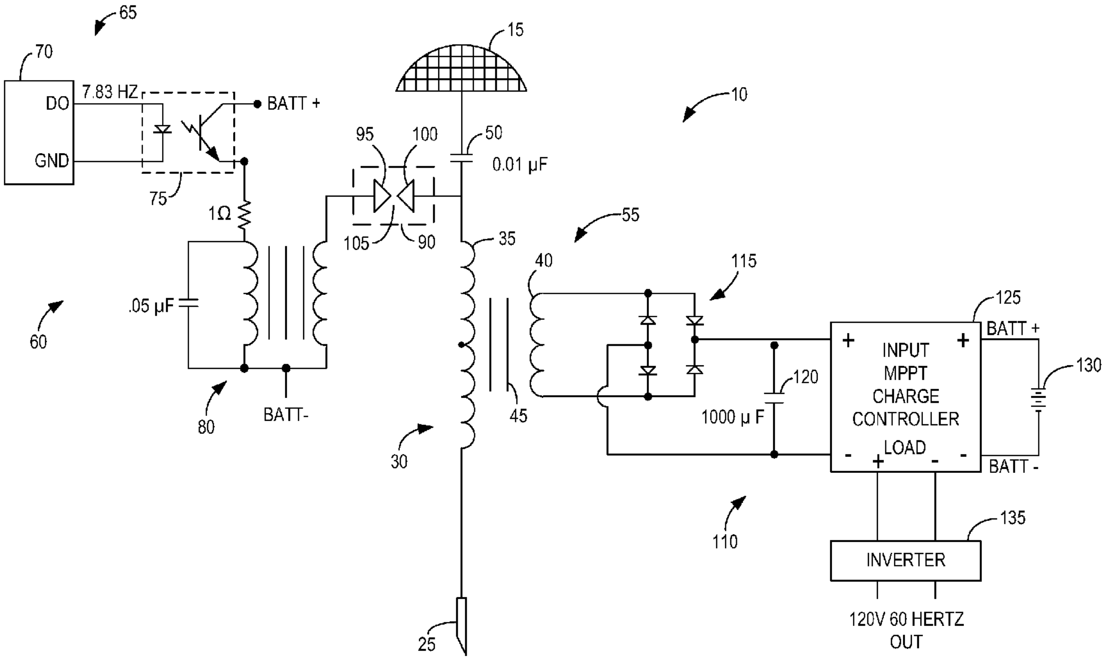
65
70
DO
GND
7.83 HZ
BATT +
75
1Ω
.05 µF
60
80
BATT-
15
95
100
50
0.01 µF
105
90
35
30
25
10
55
40
45
115
120
1000 µ F
125
INPUT MPPT CHARGE CONTROLLER
LOAD
BATT +
BATT -
130
110
135
INVERTER
120V 60 HERTZ OUT

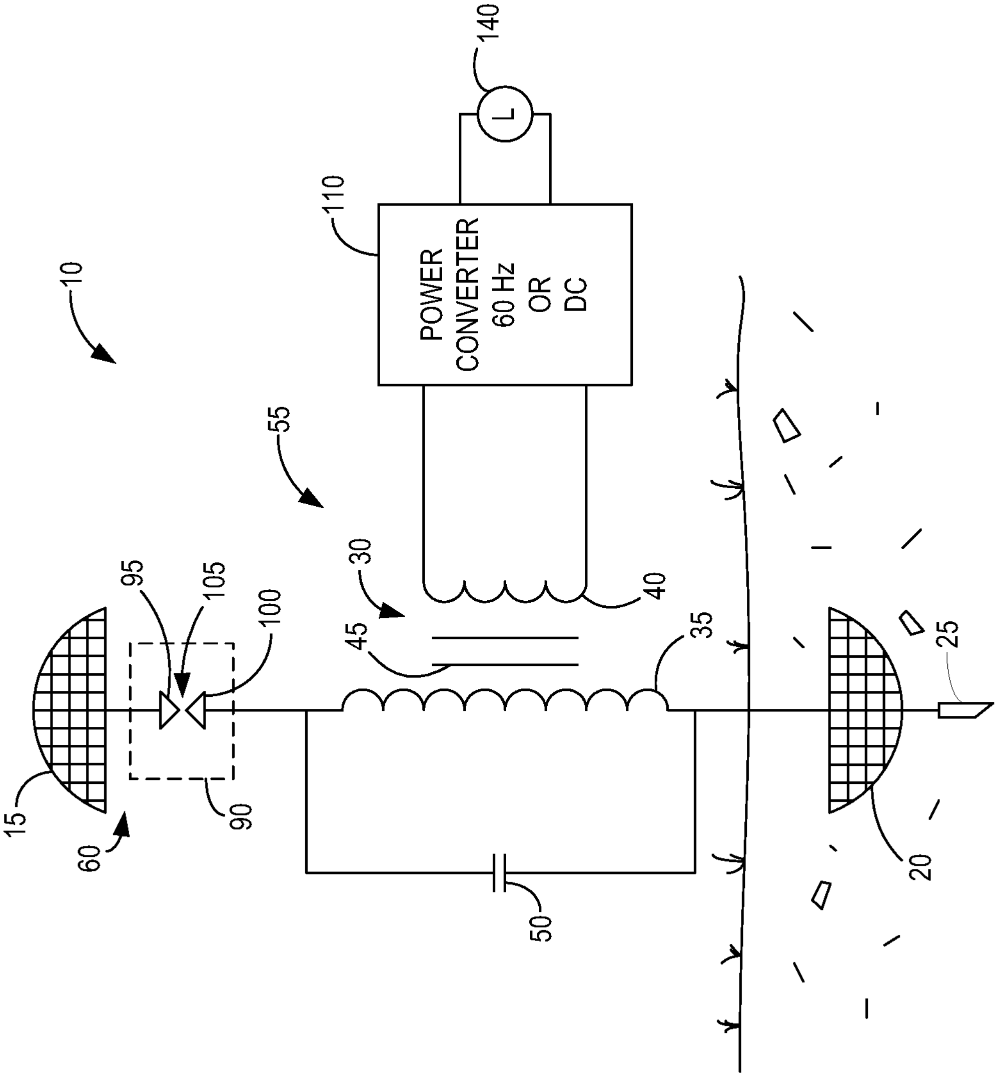
POWER
CONVERTER
60 Hz
OR
DC
L

FIG 1

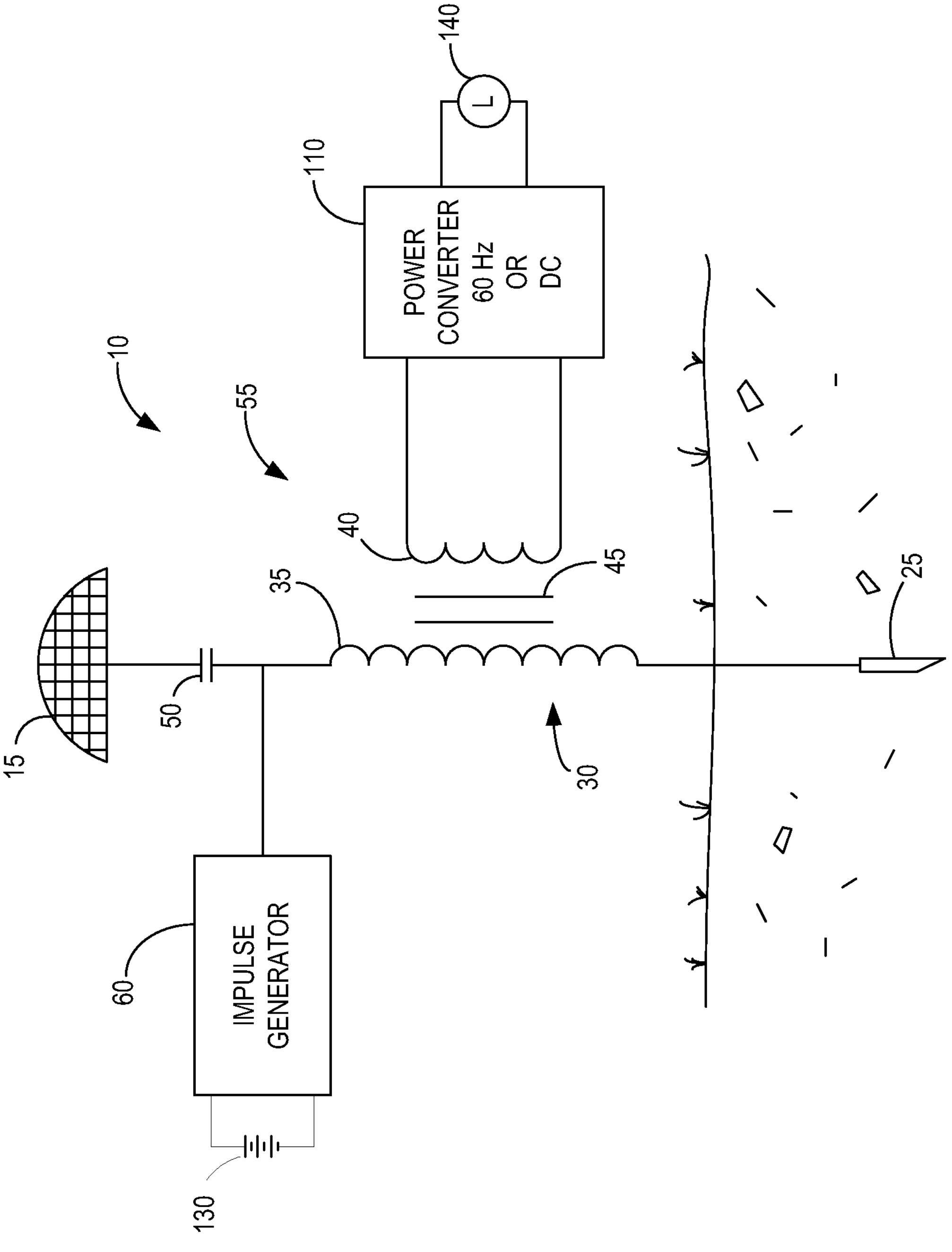
L
POWER CONVERTER 60 Hz OR DC
IMPULSE GENERATOR

FIG 2

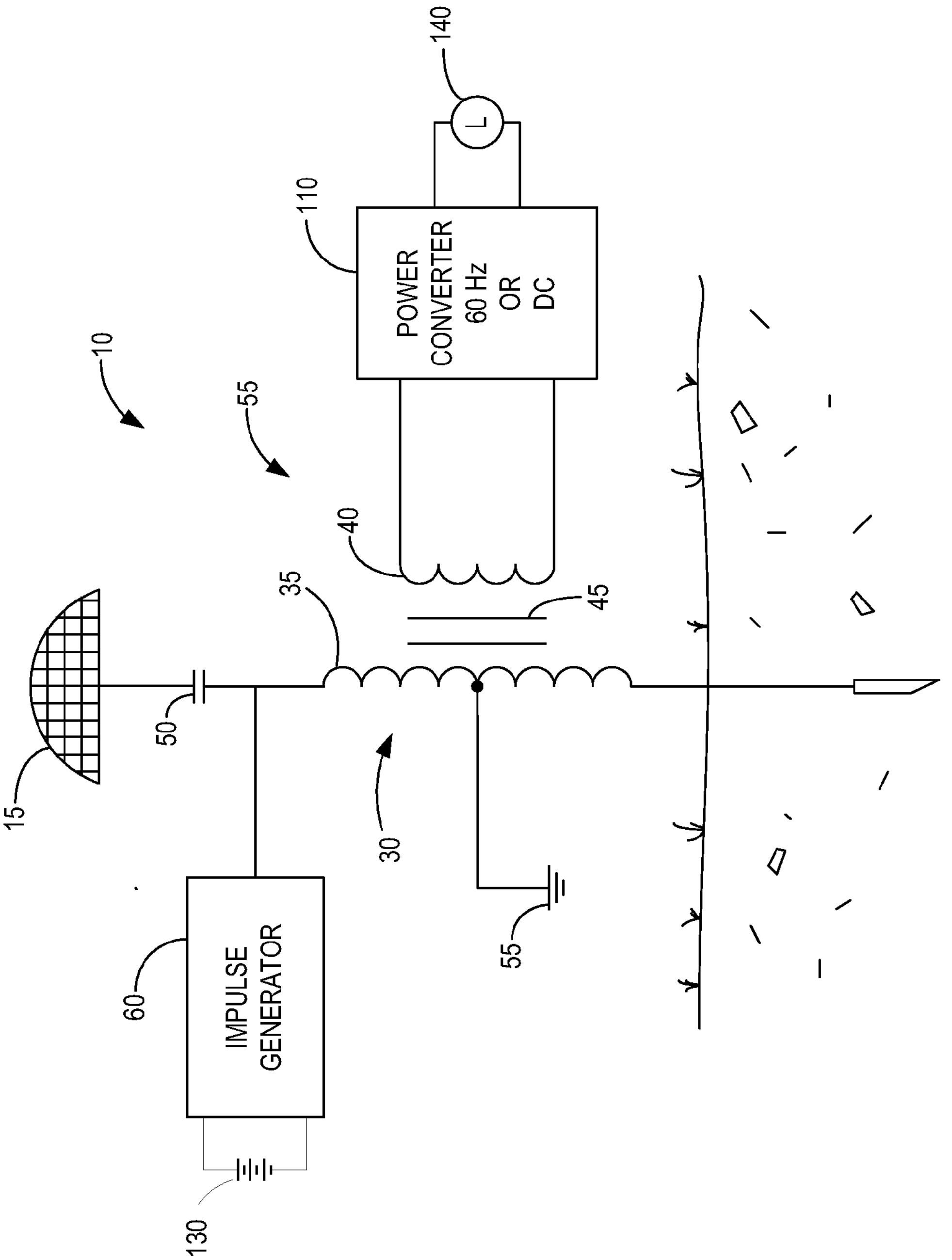
L
POWER CONVERTER 60 Hz OR DC
IMPULSE GENERATOR

FIG 3

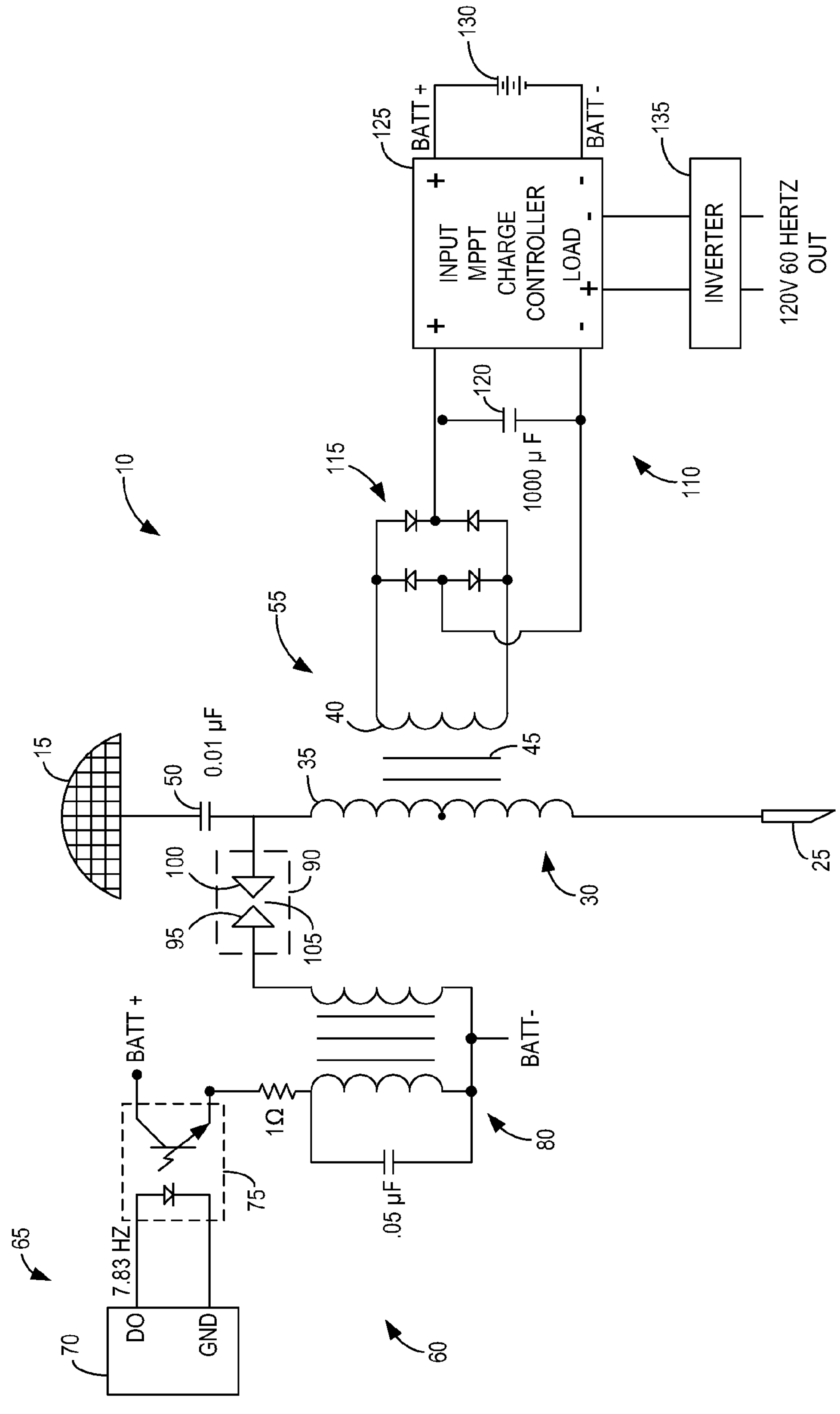
10
15
0.01 µF
50
35
45
40
55
25
30
90
100
95
105
115
120
1000 µ F
110
125
BATT +
130
BATT -
135
INPUT
MPPT
CHARGE
CONTROLLER
LOAD
+
+
-
-
-
+
INVERTER
120V 60 HERTZ
OUT
BATT +
BATT-
1Ω
80
.05 µF
75
7.83 HZ
65
70
DO
GND
60

FIG 4

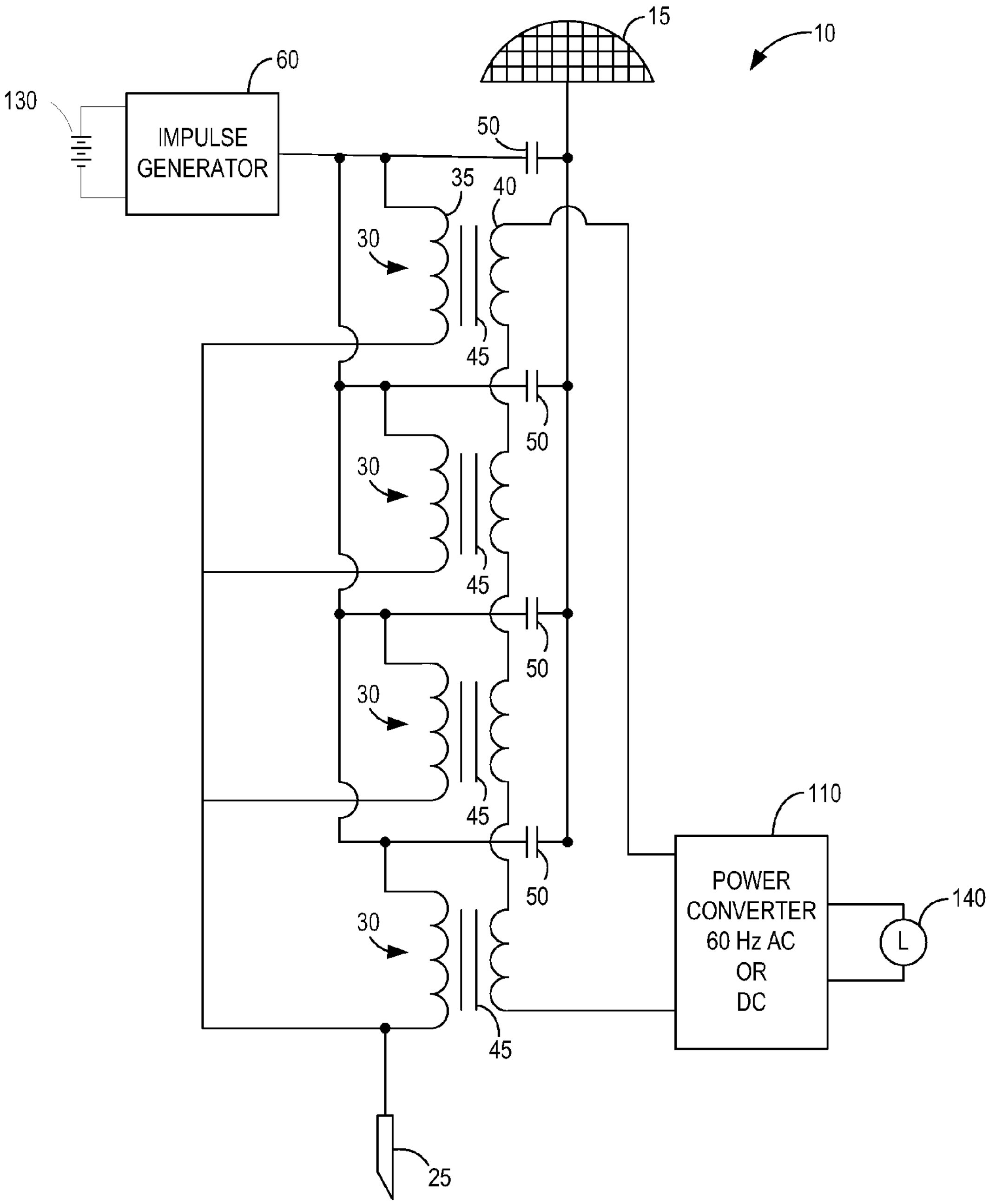
IMPULSE GENERATOR
POWER CONVERTER 60 Hz AC OR DC
L

FIG 5

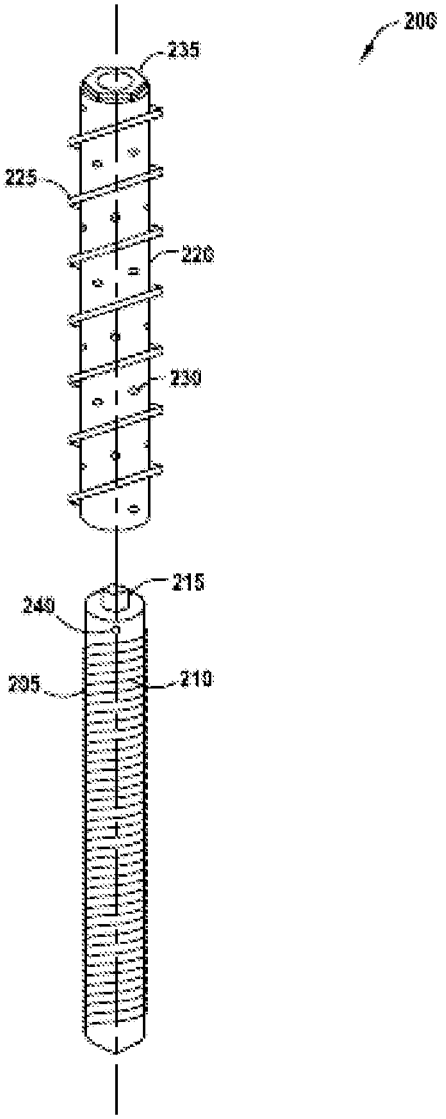
200
235
225
220
230
215
240
205
210

FIG 6A

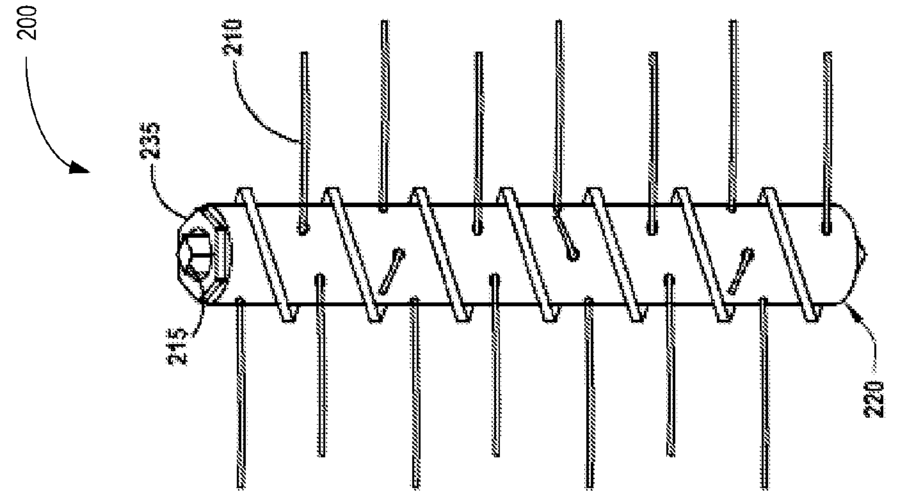
200
235
210
215
220

FIG 6C

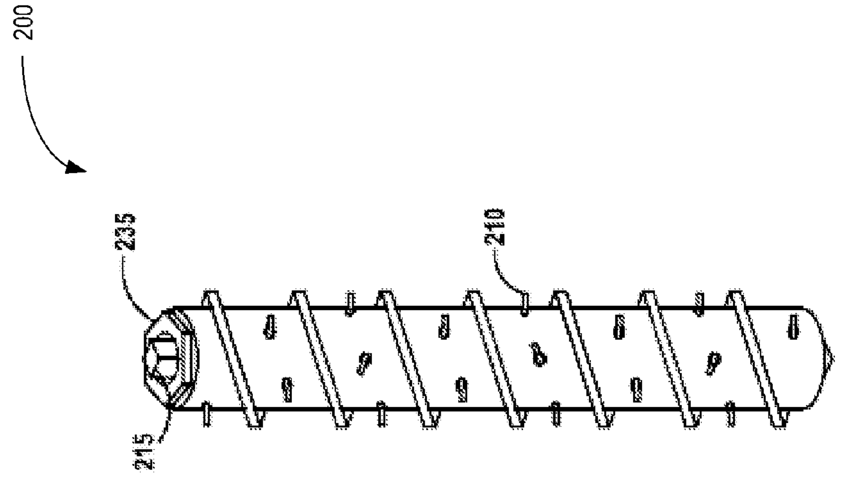
200
235
210
215

FIG 6B

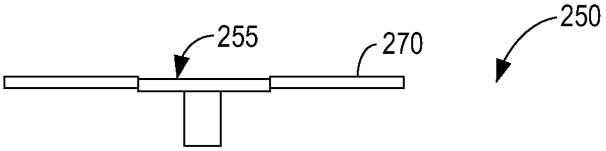

FIG 7A

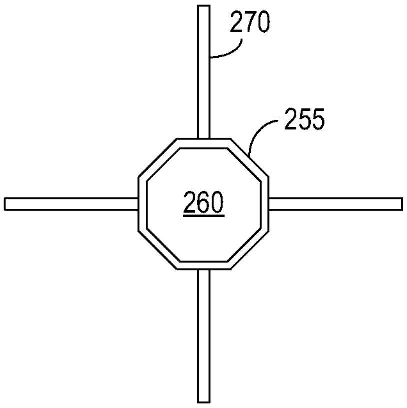

FIG 7B

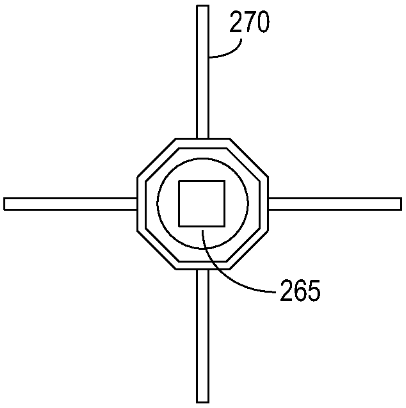

FIG 7C

POWER RECEIVER FOR EXTRACTING POWER FROM ELECTRIC FIELD ENERGY IN THE EARTH

[0001] This application claims the benefit of U.S. Provisional Application No. 61/889,894, filed 11 Oct. 2013, the disclosure of which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

[0002] The present invention relates generally to renewable energy, and more particularly to methods and apparatus for extracting energy from subsurface electrical fields beneath the earth's surface.

BACKGROUND

[0003] The earth and the ionosphere cavity may be viewed as a global electric circuit. Electrical currents are constantly flowing within the earth and its atmosphere. Within the earth, the majority of the earth's energy is carried by extremely low frequency (ELF) and ultralow frequency (ULF) waves in the 0-200 Hz frequency range. The earth's rotating magnetic field and positive lightning are two energy sources that sustain the ELF/ULF waves within the earth and the atmosphere.

[0004] A great deal of research has been devoted to studying the electric field present in the earth's ionosphere cavity. Joseph M. Crawley, the "Fair Weather Atmosphere as a Power Source", Proceedings ESA Annual Meeting on Electrostatics 2011; O. Jefimenko, "Operation of Electric Motors from Atmospheric Electric Field," American Journal of Physics, Vol. 39, Pgs. 776-779, 1971; M. L. Breuer, "Usability of Tapping Atmospheric Charge as a Power Source," Renewable Energy, Vol. 28, Pgs. 1121-1127, 2003. Numerous attempts have been made in the past to extract electrical energy from the earth's atmosphere. For example, U.S. Pat. No. 1,540,998 to Plauson describes a system for converting atmospheric electrical energy into usable power. These past attempts have been successful in producing only small amounts of power from the electrical field in the earth's ionosphere cavity. The modest success of these experiments compared to results from other renewable energy sources, such as solar and wind, has tempered further research and prevented widespread use of the electric field in the ionosphere cavity as an energy source.

SUMMARY

[0005] The present invention relates to a power receiver for extracting power from electric fields beneath the earth's surface. In embodiments of the present disclosure, a resonant transformer connected to a ground terminal draws current from the earth's electric field through the primary winding of the transformer. Current flow through the resonant transformer is induced by applying a high voltage impulse to the primary winding. The flow of current through the primary winding of the resonant transformer induces a current in the secondary winding, which may be converted and filtered to a usable form, e.g. 60 Hz AC or DC.

[0006] In some embodiments of the power receiver, the resonant frequency of the resonant transformers is below 200 Hz.

[0007] In some embodiments of the power receiver, the resonant transformer comprises a ferro-resonant transformer.

[0008] In some embodiments, the power receiver further comprises an elevated terminal.

[0009] In some embodiments of the power receiver, the primary winding of the resonant transformer is connected between the ground terminal and elevated terminal.

[0010] In some embodiments of the power receiver, the elevated terminal comprises an upper capacitive plate coupled to the earth's ionosphere cavity.

[0011] In some embodiments of the power receiver, the impulse generator comprises the upper capacitive plate and a spark gap connected between the upper capacitive plate and the primary winding of the resonant transformer. The spark gap comprises a pair of electrodes separated by a gap and configured to generate a spark when a voltage difference between the electrodes reaches a predetermined level.

[0012] In some embodiments of the power receiver, the impulse generator comprises a pulse generator for generating low voltage pulses, a step-up transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses, and a spark gap connected between the step-up transformer and the primary winding of the resonant transformer to generate a spark responsive to the high voltage impulses from the step-up transformer.

[0013] In some embodiments of the power receiver, the impulse generator comprises a pulse generator for generating low voltage pulses, and a step-up transformer connected to the primary winding of the resonant transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses.

[0014] In some embodiments of the power receiver, the impulse generator comprises a solid state spark generator.

[0015] In some embodiments of the power receiver, the resonant transformer includes a capacitor connected in parallel with the primary winding.

[0016] In some embodiments of the power receiver, the resonant transformer includes a capacitor connected in series with the primary winding between the impulse generator and the elevated terminal.

[0017] In some embodiments, the power receiver comprises multiple resonant transformers having primary windings connected in parallel between the ground terminal and the elevated terminal.

[0018] In some embodiments of the power receiver, the resonant transformers have different resonant frequencies.

[0019] In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are all below 200 Hz.

[0020] In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

[0021] Another embodiment of the power receiver comprises a resonant circuit connected to a ground terminal disposed below the surface of the earth, an impulse generator for generating and applying a high voltage electrical impulse to the resonant circuit to induce current flow from the ground terminal through the resonant circuit, and a power conversion circuit connected to the resonant circuit to convert electrical current flowing through the resonant circuit to a desired form. The resonant circuit has a resonant frequency below 200 Hertz.

[0022] In some embodiments of the power receiver, the resonant circuit comprises a resonant transformer having a primary winding, a secondary winding, and resonant capacitor connected in series with the primary winding.

[0023] In some embodiments of the power receiver, the resonant circuit comprises multiple resonant transformers having primary windings connected in parallel to the ground terminal.

[0024] In some embodiments of the power receiver, the resonant transformers have different resonant frequencies.

[0025] In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are all below 200 Hz.

[0026] In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

[0027] Other embodiments of the disclosure comprise a ground terminal for a power receiver. In one embodiment, the ground terminal comprises a ground shaft configured for insertion beneath the surface of the earth, a hollow cylinder surrounding the ground shaft and having a plurality of openings, and a plurality of ground wires connected at one end to the ground shaft. The ground wires are wound around the ground shaft and have free ends protruding through respective openings in the hollow shaft so that rotation of the ground shaft relative to the hollow cylinder causes the ground wires to extend radially into the earth.

[0028] Other embodiments of the disclosure comprise methods of extracting power from the earth. In one embodiment, the method comprises applying a high voltage impulse to resonant circuit coupled to a ground terminal disposed beneath the surface of the earth to initiate resonance in the resonant circuit and induce the flow of current from the ground terminal to the resonant circuit, and converting the current flowing from the ground terminal into the resonant circuit into a useful form.

[0029] In some embodiments of the method, the resonant circuit comprises a resonant transformer including a primary winding coupled to the ground terminal and a second winding coupled to a power converter, and applying a high voltage impulse to resonant circuit comprises applying a high voltage impulse to the primary winding of the resonant transformer.

[0030] In some embodiments of the method, applying a high voltage impulse to the primary winding of the resonant transformer comprises applying an impulse in the range to 10,000 to 40,000 volts to primary winding of the transformer.

BRIEF DESCRIPTION OF THE DRAWINGS

[0031] FIG. **1** illustrates a first exemplary embodiment of a power receiver.

[0032] FIG. **2** illustrates a second exemplary embodiment of a power receiver.

[0033] FIG. **3** illustrates a third exemplary embodiment of a power receiver.

[0034] FIG. **4** illustrates a fourth exemplary embodiment of a power receiver.

[0035] FIG. **5** illustrates a fifth exemplary embodiment of a power receiver.

[0036] FIG. **6**A is an exploded perspective view of an exemplary ground antenna array for the power receiver.

[0037] FIG. **6**B is a perspective view of an assembled ground antenna array before being deployed.

[0038] FIG. **6**C is a perspective view of an assembled ground antenna array after being deployed.

[0039] FIG. **7**A is a side view of an insertion tool for installing the ground antenna array.

[0040] FIG. **7**B is a top view of the insertion tool for installing the ground antenna array.

[0041] FIG. **7**C is a bottom view of the insertion tool for installing the ground antenna array.

DETAILED DESCRIPTION

[0042] Referring now to the drawings, a power receiver for extracting energy from the earth's electric field are illustrated and indicated generally by the numeral **10**. Various embodiments of the power receiver **10** are described and similar reference numbers are used throughout the description to indicate similar components.

[0043] The power receiver **10** converts energy in the ELF/ULF waves to useful form, e.g. 60 Hz AC or DC. The power receiver **10** is essentially a resonance circuit that resonates at the natural resonance frequencies in the earth's electric field. These resonance frequencies, known as Schumann resonance frequencies, occur at 7.83 Hz, 14.3 Hz, 20.8 Hz, 27.3 Hz, and 33.8 Hz. A high voltage impulse initiates resonance within the power receiver **10**. In the resonant mode, the impedance of the power receiver **10** is reduced to near zero thus inducing ground currents to flow into the power receiver **10** where the ground currents are converted to useful form.

[0044] FIG. **1** illustrates a first embodiment of the power receiver **10**. The power receiver **10** comprises a resonant transformer **30** connected between an elevated terminal **15** and ground terminal **25**. In this embodiment, the elevated terminal **15** is capacitively coupled to electric fields within the earth's ionosphere cavity and functions as an upper capacitive plate. A lower capacitive plate **20** is connected to the ground terminal **25** beneath the surface of the earth.

[0045] The resonant transformer **30** comprises a primary winding **35**, secondary winding **40**, ferromagnetic core **45**, and capacitor **50**. One end of the primary winding **35** is connected to the lower capacitive plate **20** and ground terminal **25**. The opposite end of the primary winding **35** is connected via a spark gap **90** to the elevated terminal **15**. The capacitor **50** is connected in parallel with the primary winding **35** of the resonant transformer **30** to form an LC circuit **55** with a resonance frequency range of between about 0.1 and 200 Hz. In a preferred embodiment, the resonant transformer has a Q of about 10 or greater and resonance frequency in the range of about 0.1-200 Hertz. For example, the resonant transformer **30** may have a resonance frequency of about 7.83 Hz, the fundamental Schumann resonance frequency. The secondary winding **40** of the resonant transformer **30** is connected to a power converter **110** as will be hereinafter described in greater detail. The power converter **110** converts the energy extracted from the earth's electric field by the power receiver **10** into a usable form for driving a load **140**.

[0046] The elevated terminal/upper capacitive plate **15** comprises an insulated, dish-shaped plate with a large radius of curvature. The capacitance and resistance of the elevated terminal is chosen for receiving broadband electric field frequencies in the 0-200 Hz range. The upper capacitive plate **15** is sized to maximize to the extent practical coupling with the electric field in the earth's ionosphere cavity.

[0047] The lower capacitive plate **20** is also a dish-shaped plate with a large radius of curvature. One function of the lower capacitive plate **20** is to collect charge from the earth's ground currents and provide an instantaneous source of current as hereinafter described. The capacitance and resistance of the lower capacitive plate **20** is selected to promote the flow of current from the ground with minimal losses.

[0048] The spark gap **90** connected between the elevated terminal **15** and resonant transformer **30** comprises a pair of electrodes **95**, **100** separated by an evacuated air gap **105**. Electrode **95** is connected to the upper capacitive plate **15**. Electrode **100** is connected to the resonant transformer **30**. The spark gap **105** prevents electrical discharge from the upper capacitive plate **15** to the earth's ionosphere cavity. The spark gap **90** in combination with the elevated terminal **15** function as an impulse generator that applies a high voltage impulse of about 10,000-40,000 volts to the primary winding **35** to initiate resonance in the transformer **30**.

[0049] In operation, the capacitive coupling of the upper capacitive plate **15** induces a high voltage operating current in the upper capacitive plate **15**. The upper capacitive plate is connected to a first electrode **95** to the spark gap **90**. When the voltage difference between the electrodes **95** and **100** reaches a threshold, a spark forms across the electrodes **95**,**100** and a high voltage impulse is applied to the primary winding **35** of the resonant transformer **30**. This high voltage impulse initiates resonance within the transformer **30**.

[0050] In resonant mode, the impedance of the resonance transformer is reduced to nearly zero allowing current to flow from the capacitive plate **20** and ground terminal **25** through the primary winding **35** of the transformer **30**, which in turn induces current in the secondary winding **40**. Power converter **110** converts the current flowing through the secondary winding **40** into a usable form for driving a load **140**. The transformer **30** will continue to resonate for a short period of time. By providing high voltage impulses to the primary winding **35** of the resonant transformer **30** at periodic intervals, it is possible to maintain a continuous flow of current from the earth into the resonant transformer **30**, thus producing a continuous supply of power.

[0051] FIG. **2** discloses a second embodiment of the primary receiver **10**. This embodiment includes a resonant transformer **30** connected between an elevated terminal **15** and ground terminal **25**. The resonant transformer **30** comprises a primary winding **35**, secondary winding **40**, ferromagnetic core **45** and a high voltage capacitor **50**. One end of the primary winding **35** is connected to the ground terminal **25**. The opposite end of the primary winding **35** is connected to the elevated terminal **15**. The capacitor **50** has a capacitance of about 0.01 micro-farads. In contrast to the previous embodiment, capacitor **50** is connected in series with the primary winding **35** and elevated terminal **15** and forms a LC circuit **55** with a Q of about 10 or greater and a resonance frequency in the range of about 0.1 to 200 Hz. In a preferred embodiment, the resonance frequency of the transformer **30** is 7.83 Hz, the fundamental Schumann resonance frequency. An impulse generator **60** is connected between the primary winding **35** of the resonant transformer **30** and the series capacitor **50** and applies a high voltage impulse to the primary winding **35** of the resonant transformer **30**. A battery **130** or other external power source supplies power to the impulse generator **60**. As previously described, the high voltage impulse applied by the impulse generator **60** initiates resonance within the resonant transformer **30** inducing current flow from the ground terminal **25** into the primary winding **35** of the resonant transformer **30**. The flow of current from the ground terminal **25** through the primary winding **35** induces current in the secondary winding **40**. Power converter **110** converts the electrical energy in the current flowing through the primary winding **40** into a usable form.

[0052] In contrast to the first embodiment, it is not required to capacitively couple the elevated terminal **15** in the second embodiment to the earth's ionosphere cavity. Rather, the elevated terminal **15** in this embodiment provides lightning protection and dissipates some of the energy flowing into the power receiver **10** to the earth's ionosphere cavity. Also, in contrast to the first embodiment, the capacitor **50** is connected in series between the primary winding **35** of the transformer **30** and the elevated terminal **15**. Those skilled in the art will appreciate that the capacitor **50** could also be connected in parallel rather than series with the primary winding **35** as shown in FIG. **1**. Another difference is that the impulse generator **60** has an external power source. The amount of energy generated by the power receiver **10**, however, is far greater than the energy needed to generate high voltage impulses. The first embodiment does not require an external power source to generate high voltage impulses.

[0053] FIG. **3** illustrates a third embodiment of the power receiver **10**. This embodiment is essentially the same as the embodiment shown in FIG. **2**. The main difference is that a center tap of the primary winding **35** in the resonant transformer **30** is connected to an electrical ground **55**. It should be appreciated that the electrical ground **55** may be different than the earth ground. When the center tap of the resonant transformer **30** is grounded at a distance away from the ground terminal **25** (e.g. 50 ft to 100 ft), the power receiver **10** becomes a transmitter via the ground loop formed.

[0054] FIG. **4** illustrates the power receiver **10** of FIG. **2** in greater detail. The power receiver includes a resonant transformer **30** connected between a ground terminal **25** and an elevated terminal **15**. The resonant transformer **30** includes a primary winding **35**, secondary winding **40**, ferromagnetic core **45** and series capacitor **50** configured as previously described. The resonant transformer **30** may have a Q of about 10 and a resonance frequency in the range of about 0.1 to 200 Hz. The resonant transformer **30** may be made using an Allison transformer (part #1530BP120R) connected in series with a 0.01 micro-farad capacitor, such as the Condensor Products high voltage capacitor (part #TC 103-17-125). An impulse generator **60** is connected between the primary winding **35** of the resonant transformer **30** and the series capacitor **50** and applies a high voltage impulse in the range of about 10,000 to 40,000 volts to the primary winding of the transformer **30**. A battery **130** or other external power source supplies power to the impulse generator **60**. The power converter **110** connects to the secondary winding **40** of the resonant transformer **30** for converting current in the secondary winding of the transformer to a useful form.

[0055] The impulse generator **60** comprises a pulse generator **65** for generating low voltage pulses, a step-up transformer **80** for converting the low voltage pulses from the pulse generator **65** to high voltage pulses, and a spark gap **90** for generating sparks responsive to the high voltage pulses from the step-up transformer **80**.

[0056] The pulse generator **65** comprises a square wave generator **70** and solid state relay **75**. The square wave generator **70** generates a digital pulse stream. In one embodiment, the digital pulse stream generates a square waveform with a frequency of about 7.83 Hz. The frequency of the digital pulse stream is selected to match the resonance frequency of the transformer **30**, though such is not necessarily required. The pulse stream output from the square wave generator **70** is applied to the solid state relay **75**. The solid state relay **75** is connected between a battery or other power source and a first winding of the step-up transformer **80**. The solid state relay **75** functions as a switch that is activated responsive to the waveform from the square wave generator **70** to provide a continuous stream of low voltage pulses from the battery to the first winding of the step-up transformer **80**. A 1 ohm resistor is connected between the solid state relay **75** and step-up transformer **80**.

[0057] The step-up transformer **80** converts the low voltage pulses from the pulse generator **65** to high voltage pulses that are applied to the spark gap **90**. The step-up transformer has a 0.5 micro-farad capacitor connected in parallel with the primary winding of the step-up transformer **80**. The step-up transformer produces pulses at the output of about 30,000 to 40,000 volts.

[0058] The spark gap **90** comprises a pair of electrodes **95**, **100** separated by an air gap **105**. As previously described, when the voltage potential between the electrodes **95**, **100** reaches a threshold, a spark forms between the electrodes **95**, **100** and supplies a nearly instantaneous, high voltage impulse to the primary winding **35** of the resonant transformer **30**. This high voltage impulse initiates resonance in the resonant transformer **30** inducing current flow from the ground terminal **25** through the primary winding **35** of the resonant transformer **30**.

[0059] The power converter **110** comprises a bridge rectifier **115**, filter capacitor **120**, charge controller **125**, and inverter **135**. The bridge rectifier **115** converts the AC current flowing through the secondary winding **40** of the resonant transformer to a DC current. a Filter capacitor **120** removes unwanted frequencies from the DC current. The filter capacitor **120** has a capacitance of about 1000 micro-farads. The DC current is input to the charge controller **125**. The charge controller **125** may, for example, comprise a maximum power point tracking (MPPT) charge controller, which is commonly used in solar power generating systems. The charge controller **125** applies a small amount of energy to a battery **30** to charge the battery **130**. As previously noted, the battery **130** serves as a power source for the impulse generator **60**. The remaining current is supplied to an inverter **135**, which converts the DC current to an AC current with a desired voltage and frequency, e.g., 120 volts/60 Hz AC. The power converter **110** as shown in FIG. **4** may be utilized in the embodiment shown in FIGS. **1**, **2** and **3**.

[0060] FIG. **5** illustrates a power converter **10** according to another embodiment. The power converter **10** comprises a plurality of resonant transformers **30** connected between a ground terminal **25** and elevated terminal **15**. Each of the resonant transformers **30** comprises a primary winding **35**, secondary winding **40**, ferromagnetic core **45** and series capacitor **50**. The primary windings **35** of the resonant transformers **30** are connected in parallel. The secondary windings **40** are connected in series with the power converter **110**. An impulse generator **60** applies a high voltage impulse to the primary windings **35** of the resonant transformers **30**. A battery **130** or other external power source supplies power to the impulse generator **60**. The power converter **110** converts the current in the power converter circuit to a usable form for driving a load **140**.

[0061] In one embodiment, each of the resonant transformers **30** shown in FIG. **5** is configured to have a different resonant frequency. In one embodiment, the resonant transformers **30** are configured to resonate at frequencies of 7.83 Hz, 14.8 Hz, 20.3 Hz and 26.8 Hz respectively. Additional resonant transformers **30** could be added to operate at other resonance frequencies.

[0062] FIGS. **6**A-**6**C illustrate a high quality ground antenna array **200** which may be used as a ground terminal **25**. The ground antenna array **200** comprises a generally cylindrical ground shaft **205** disposed with a hollow cylinder **220** and a plurality of reinforced, heavy gauge ground wires **210** attached at one end to the ground shaft **205**. The ground shaft **205** and ground wires **210** should be highly conductive and have low resistance to supply current from the ground to the power receiver **10**. In one embodiment, the ground wires **210** may be copper or other highly-conductive metal. The end of the ground shaft may be pointed to facilitate insertion into the earth. A connection port on the ground shaft **220** is provided to electrically connect the ground antenna array **220** to the resonant transformer **30**.

[0063] The hollow cylinder **220** has external threads **25** to facilitate insertion into the ground. A rotator nut **235** is fixedly secured to the top end of the hollow shaft **220**. A square shaft **215** protrudes from the top end of the ground shaft **205** into the opening in the rotator nut **235**. FIG. **6**B. A tool **250**, shown in FIG. **7**, engages with the rotator nut **235** and square shaft **215** during insertion of the ground antennas array **200** into the ground as will be hereinafter described.

[0064] The insertion tool **250** is shown in FIG. **7**. The insertion tool **250** includes a tool body **255** having a first socket **260** on one side to fit the rotator nut **235** on the hollow cylinder **220** and a second socket **265** on the other side to fit the square shaft **215** on the ground shaft **205**. Arms **270** extend from the outer periphery of the tool body **255** for manually or mechanically turning the insertion tool **250**.

[0065] Before the antenna array **200** is deployed, the ground wires **210** are wound around the ground shaft **205** with the free ends protruding slightly from respective openings **230** in the hollow cylinder **220** to a distance not to exceed one half (½) the depth of the external threads **225** on the hollow cylinder **220**. FIG. **6**B illustrates the ground antenna array **200** before deployment. FIG. **6**C illustrates the ground antenna array in a deployed configuration.

[0066] Installation of the ground antenna array **200** is performed in two stages. In the first stage, a hole slightly smaller in diameter than the threads **235** of the hollow cylinder **220** is drilled into the Earth to a depth matching the length of the hollow cylinder **220** or slightly longer. The hole is filled with water and the water is allowed to soak into the soil. After the ground is softened, the hollow cylinder **220** is rotated using the insertion tool **250** to insert the ground antenna array **200** into the ground. The first socket **260** of the insertion tool **250** is engaged with the rotator nut **230** and the insertion tool **250** is turned by hand or a mechanized rotating shaft fitted and attached to the tool arms **270** to thread the ground assembly into the hole. During the initial insertion of the ground antenna array **200**, the ground shaft **205** is fixed to the hollow shaft **220** and rotates with the hollow shaft. The hollow cylinder **220** is rotated until it reaches the full depth of the hole.

[0067] Once the ground antenna array **200** has been fully inserted into the earth, the insertion tool **250** is flipped over and the second socket **265** of the insertion tool **250** is engaged with the square shaft **215**. The insertion tool **250** is turned by hand or a mechanized rotating shaft fitted and attached to the tool arms **270** to rotate the ground shaft **205**. During the second phase, the ground shaft **205** rotates freely inside the hollow cylinder **220**. Rotation of the ground shaft **205** causes the reinforced ground wires **210** to extend radially into the earth. The ground shaft **220** is rotated until the ground wires are fully extended. The ends of the ground wires may be sharpened to aid in the extension of the ground wires during the second phase.

[0068] After the ground antenna array **200** is deployed, a connection cable **280** is attached to a connection port **240** on the ground shaft **220** to electrically connect the ground antenna array **220** to the resonant transformer **30** in the power receiver **10**.

What is claimed is:

1. A power receiver for extracting electrical energy from the earth's electric field, said power receiver comprising:

a resonant transformer connected to a ground terminal disposed below the surface of the earth;

an impulse generator for generating and applying a high voltage electrical impulse to a primary winding of the resonant transformer to induce current flow from the ground terminal through the primary winding of the transformer; and

a power conversion circuit connected to a secondary winding of the resonant transformer to convert electrical current flowing through the secondary winding to a desired form.

2. The power receiver of claim **1** wherein a resonant frequency of the resonant transformers is below 200 Hz.

3. The power receiver of claim **1** wherein the resonant transformer comprises a ferro-resonant transformer.

4. The power receiver of claim **1** further comprising an elevated terminal, and wherein the primary winding of the resonant transformer is connected between the ground terminal and elevated terminal.

5. The power receiver of claim **4** wherein the elevated terminal comprises an upper capacitive plate coupled to the earth's ionosphere cavity.

6. The power receiver of claim **5** wherein the impulse generator comprises:

the upper capacitive plate;

a pair of electrodes separated by a spark gap, said electrodes connected between the upper capacitive plate and the primary winding of the resonant transformer and configured to generate a spark when a voltage difference between the electrodes reaches a predetermined level.

7. The power receiver of claim **1** wherein the impulse generator comprises:

a pulse generator for generating low voltage pulses;

a step-up transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses;

a spark gap connected between the step-up transformer and the primary winding of the resonant transformer to generate a spark responsive to the high voltage impulses from the step-up transformer.

8. The power receiver of claim **1** wherein the impulse generator comprises:

a pulse generator for generating low voltage pulses;

a step-up transformer connected to the primary winding of the resonant transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses.

9. The power receiver of claim **1** wherein the impulse generator comprises a solid state spark generator.

10. The power receiver of claim **1** wherein said resonant transformer includes a capacitor connected in parallel with the primary winding.

11. The power receiver of claim **1** wherein said resonant transformer includes a capacitor connected in series with the primary winding between the impulse generator and the elevated terminal.

12. The power receiver of claim **1** wherein comprise multiple resonant transformers having primary windings connected in parallel between the ground terminal and the elevated terminal.

13. The power receiver of claim **12** wherein the resonant transformers have different resonant frequencies.

14. The power receiver of claim **13** wherein the resonant frequencies of the resonant transformers are all below 200 Hz.

15. The power receiver of claim **12** wherein the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

16. A power receiver for extracting electrical energy from the earth's electric field, said power receiver comprising:

a resonant circuit connected to a ground terminal disposed below the surface of the earth, said resonant circuit having a resonant frequency below 200 Hertz

an impulse generator for generating and applying a high voltage electrical impulse to the resonant circuit to induce current flow from the ground terminal through the resonant circuit; and

a power conversion circuit connected to the resonant circuit to convert electrical current flowing through the resonant circuit to a desired form.

17. The power receiver of claim **16** wherein the impulse generator comprises:

a pulse generator for generating low voltage pulses;

a step-up transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses;

a spark gap connected between the step-up transformer and resonant circuit to generate a spark responsive to the high voltage impulses from the step-up transformer.

18. The power receiver of claim **16** wherein the impulse generator comprises:

a pulse generator for generating low voltage pulses;

a step-up transformer connected to the resonant circuit for converting the low voltage pulses provided by the pulse generator to high voltage impulses.

19. The power receiver of claim **16** wherein the resonant circuit comprises a resonant transformer having a primary winding, a secondary winding, and resonant capacitor connected in parallel with the primary winding.

20. The power receiver of claim **16** wherein the resonant circuit comprises a resonant transformer having a primary winding, a secondary winding, and resonant capacitor connected in series with the primary winding.

21. The power receiver of claim **16** wherein the resonant circuit comprises multiple resonant transformers having primary windings connected in parallel to the ground terminal.

22. The power receiver of claim **21** wherein the resonant transformers have different resonant frequencies.

23. The power receiver of claim **22** wherein the resonant frequencies of the resonant transformers are all below 200 Hz.

24. The power receiver of claim **23** wherein the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

25. A ground terminal for a power receiver, said ground terminal comprising:

a ground shaft configured for insertion beneath the surface of the earth;

a hollow cylinder surrounding the ground shaft and having a plurality of openings; and a plurality of ground wires connected at one end to the ground shaft said ground wires being wound around the ground shaft and having free ends protruding through respective openings in the hollow shaft r so that rotation of the ground shaft relative to the hollow cylinder causes the ground wires to extend radially into the earth.

26. A method of extracting power from the earth, said method comprising:

applying a high voltage impulse to resonant circuit coupled to a ground terminal disposed beneath the surface of the earth to initiate resonance in the resonant circuit and induce the flow of current from the ground terminal to the resonant circuit; and

converting the current flowing from the ground terminal into the resonant circuit into a useful form.

27. The method of claim **26** wherein resonant circuit comprises a resonant transformer including a primary winding coupled to the ground terminal and a second winding couple to a power converter, and wherein applying a high voltage impulse to resonant circuit comprises applying a high voltage impulse to the primary winding of the resonant transformer.

28. The method of claim **27** wherein applying a high voltage impulse to the primary winding of the resonant transformer comprises applying an impulse in the range to 10,000 to 40,000 volts to primary winding of the transformer.

* * * * *

US 20190348842A1

(19) **United States**
(12) **Patent Application Publication**
Dinwiddie et al.
(10) **Pub. No.: US 2019/0348842 A1**
(43) **Pub. Date: Nov. 14, 2019**

(54) POWER RECEIVER FOR EXTRACTING POWER FROM ELECTRIC FIELD ENERGY IN THE EARTH

(71) Applicant: **Earth Energies, Inc.**, Alpharetta, GA (US)

(72) Inventors: **John Dinwiddie**, Pittsboro, NC (US); **Terry L. Wright**, Suwanee, GA (US); **David R. Ames**, Johns Creek, GA (US)

(21) Appl. No.: **16/521,764**

(22) Filed: **Jul. 25, 2019**

Related U.S. Application Data

(63) Continuation of application No. 15/397,281, filed on Jan. 3, 2017, now Pat. No. 10,389,138, which is a continuation-in-part of application No. 14/509,772, filed on Oct. 8, 2014, now Pat. No. 9,564,268.

(60) Provisional application No. 61/889,894, filed on Oct. 11, 2013.

Publication Classification

(51) **Int. Cl.**
H02J 4/00 (2006.01)

(52) **U.S. Cl.**
CPC . ***H02J 4/00*** (2013.01); *H05F 7/00* (2013.01)

(57) ABSTRACT

A resonant transformer connected between a ground terminal and elevated terminal draws current from the earth's electric field through a primary winding of the transformer. An impulse generator applies a high voltage impulse to the primary winding of the resonant transformer to cause current to flow from the ground terminal through the primary winding. The flow of current through the primary winding of the resonant transformer induces a current in the secondary winding, which may be converted and filtered to a usable form, e.g. 60 Hz AC or DC.

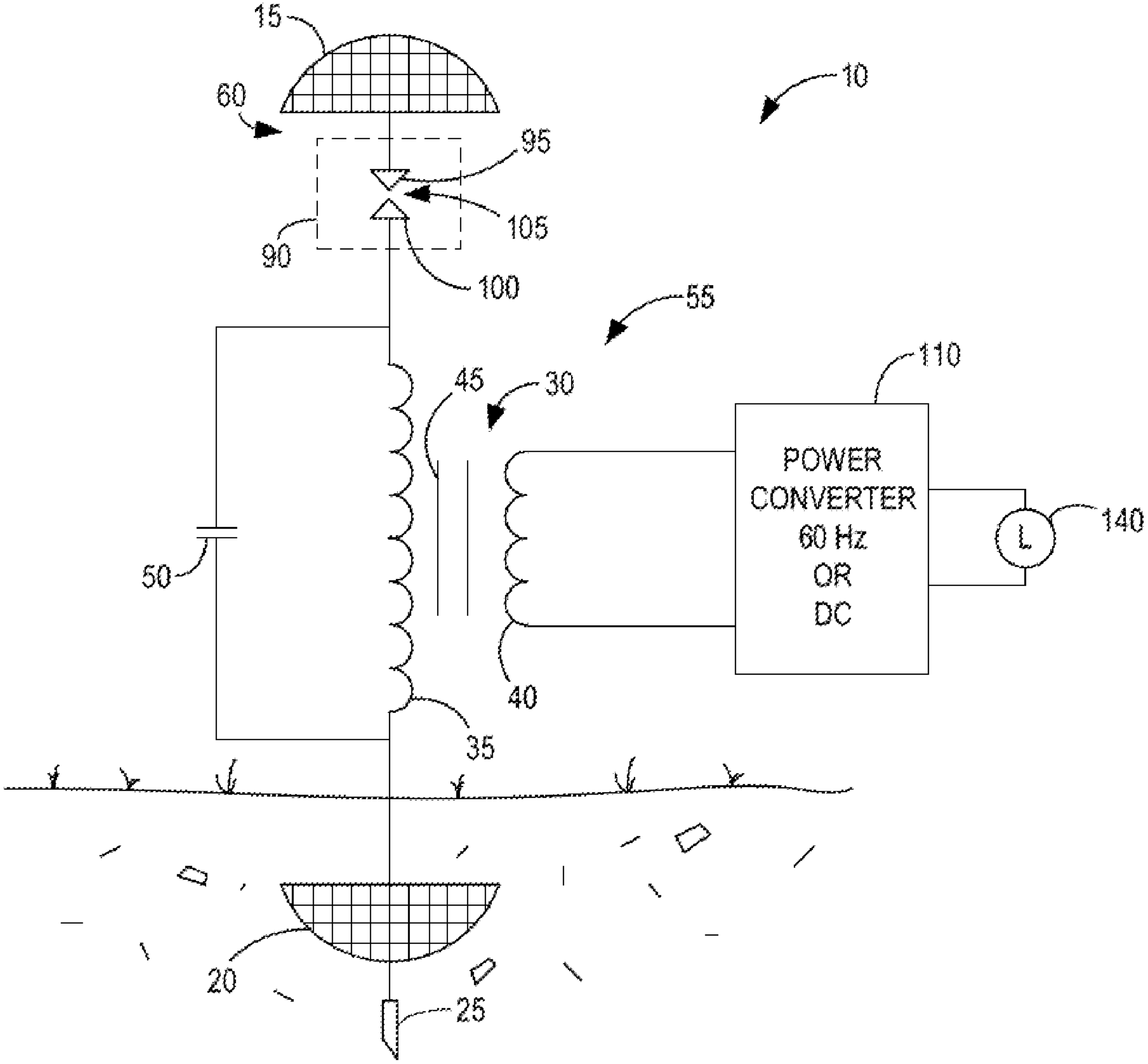

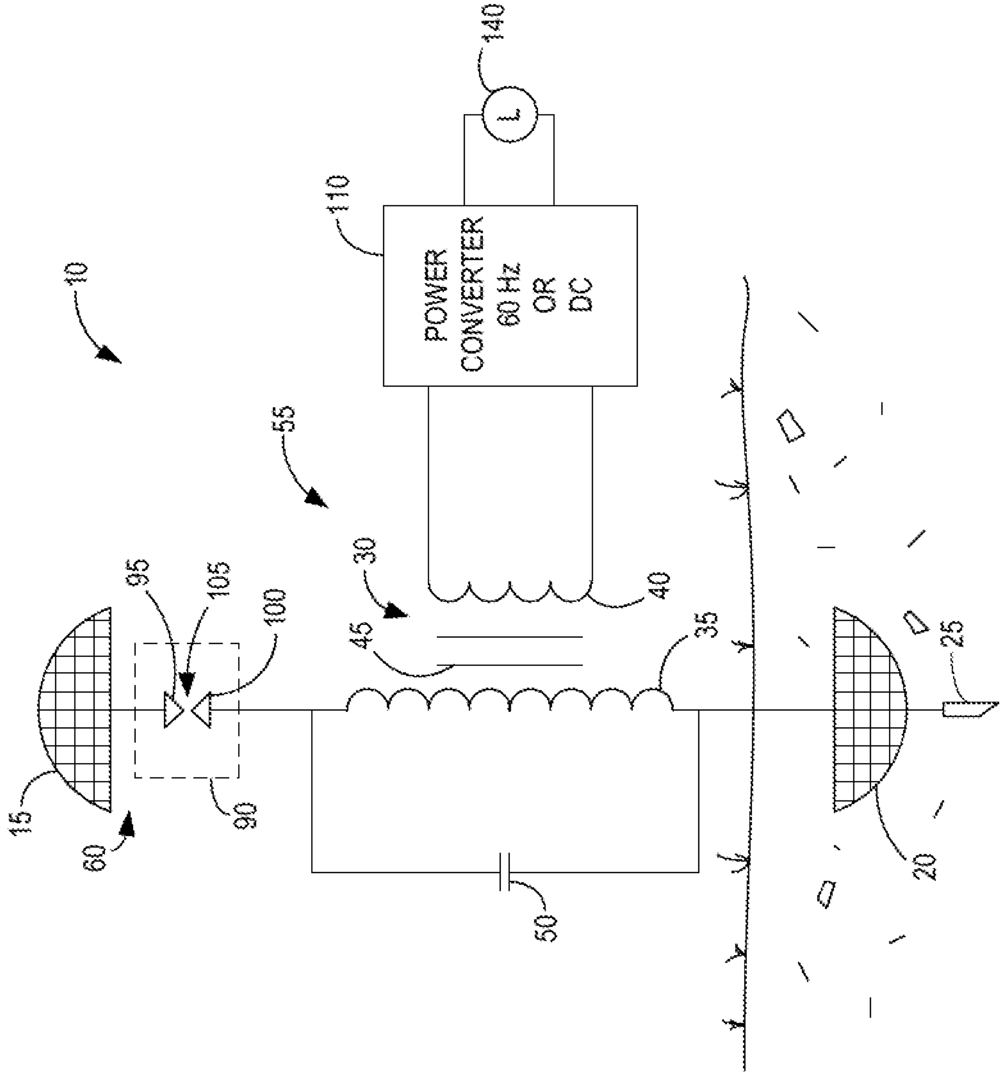
L
POWER CONVERTER 60 Hz OR DC

FIG 1

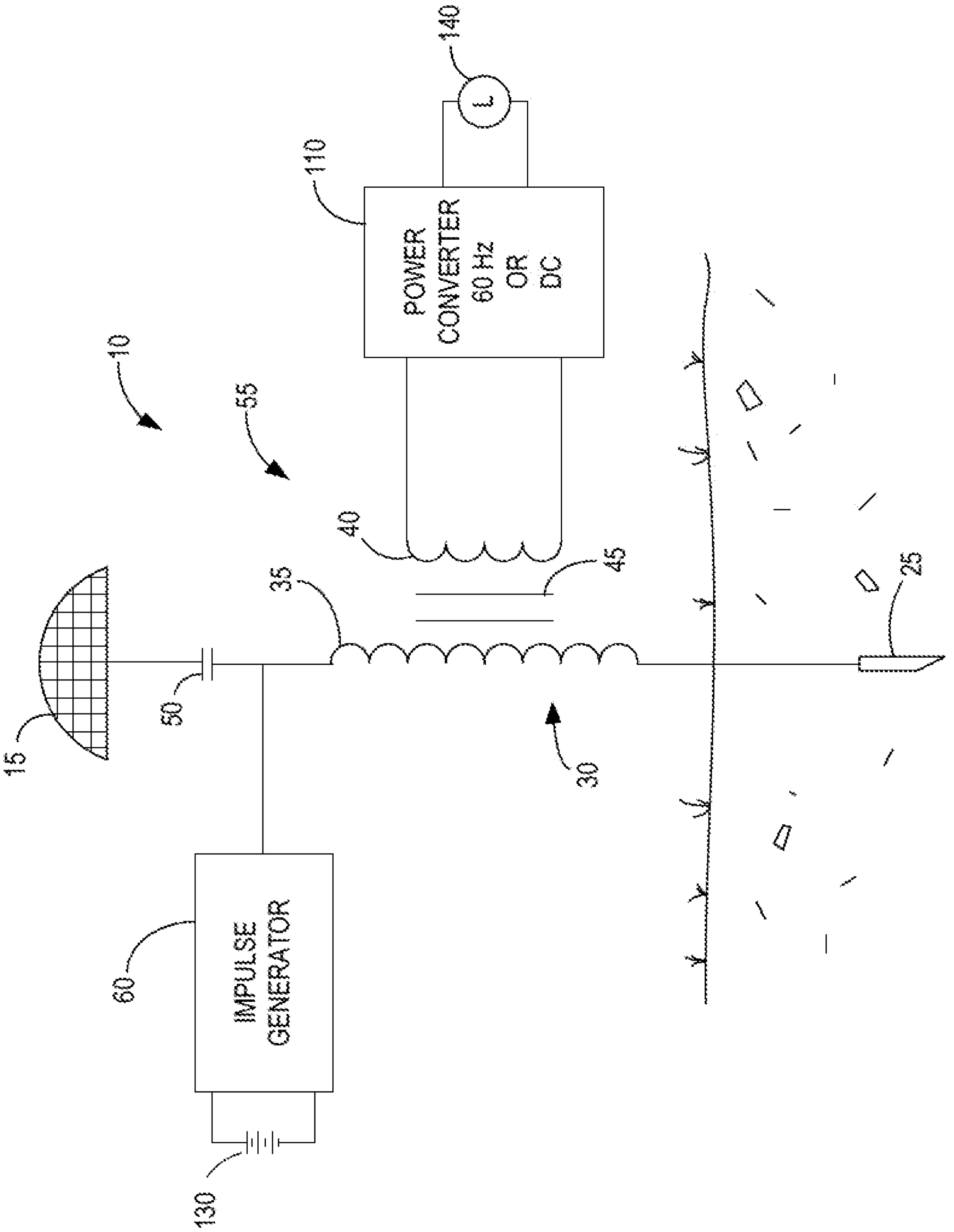
L
POWER CONVERTER 60 Hz OR DC
IMPULSE GENERATOR

FIG 2

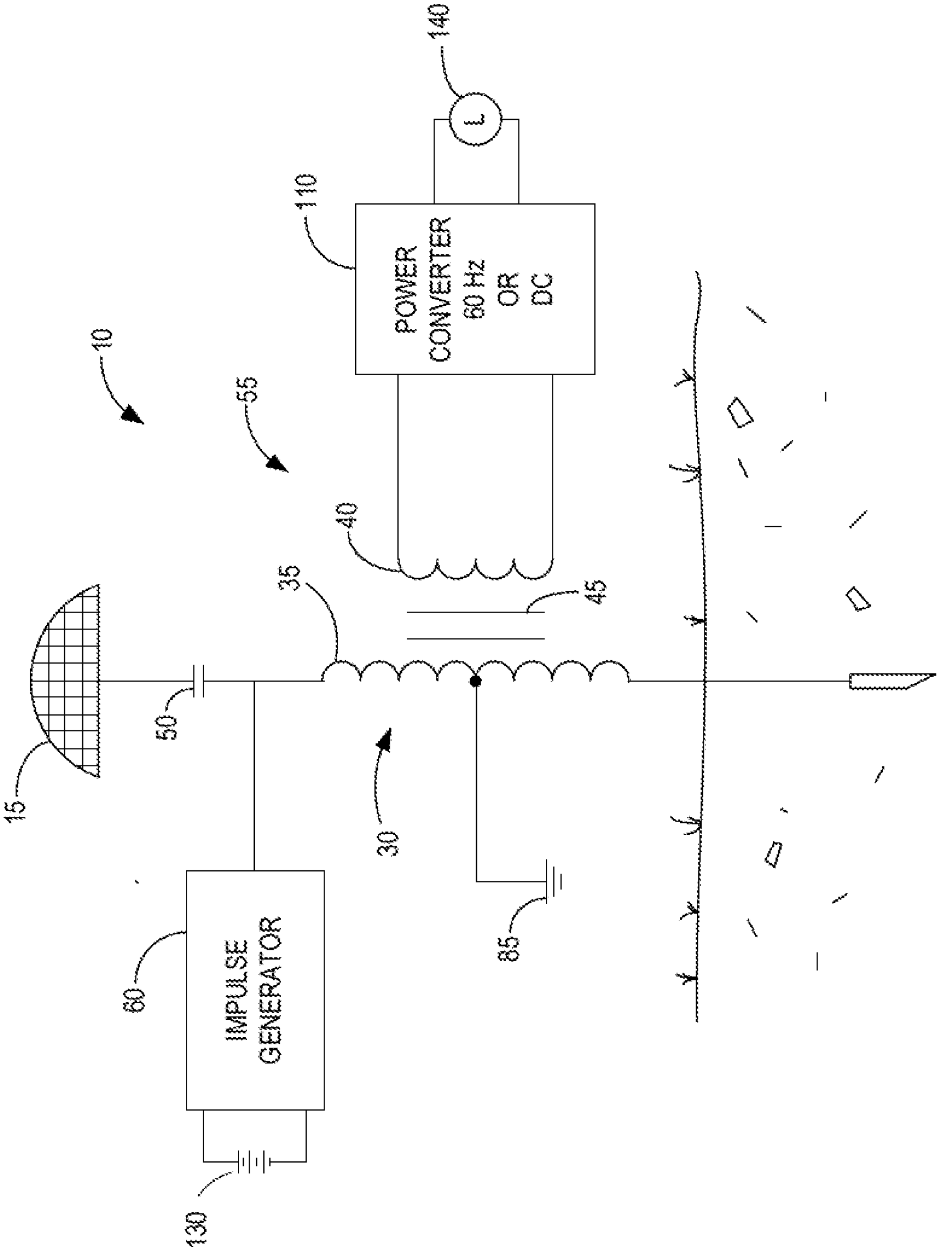
L
POWER CONVERTER 60 Hz OR DC
IMPULSE GENERATOR

FIG 3

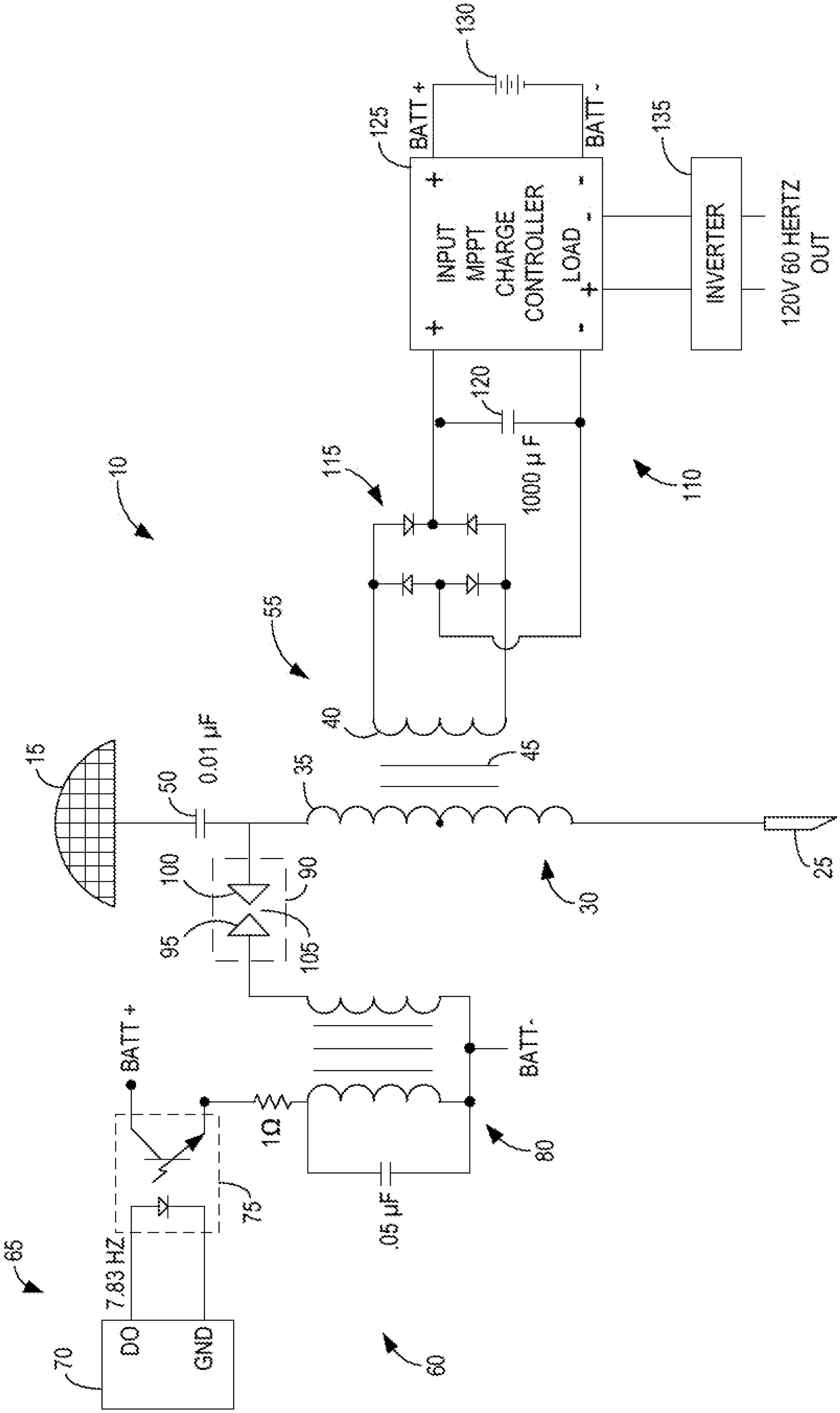
10
15
0.01 µF
50
35
45
40
55
115
120
1000 µ F
110
130
BATT +
BATT -
125
INPUT
MPPT
CHARGE
CONTROLLER
LOAD
+
-
135
INVERTER
120V 60 HERTZ
OUT
100
90
95
105
30
25
65
70
DO
GND
7.83 HZ
75
BATT +
1Ω
.05 µF
80
BATT-
60

FIG 4

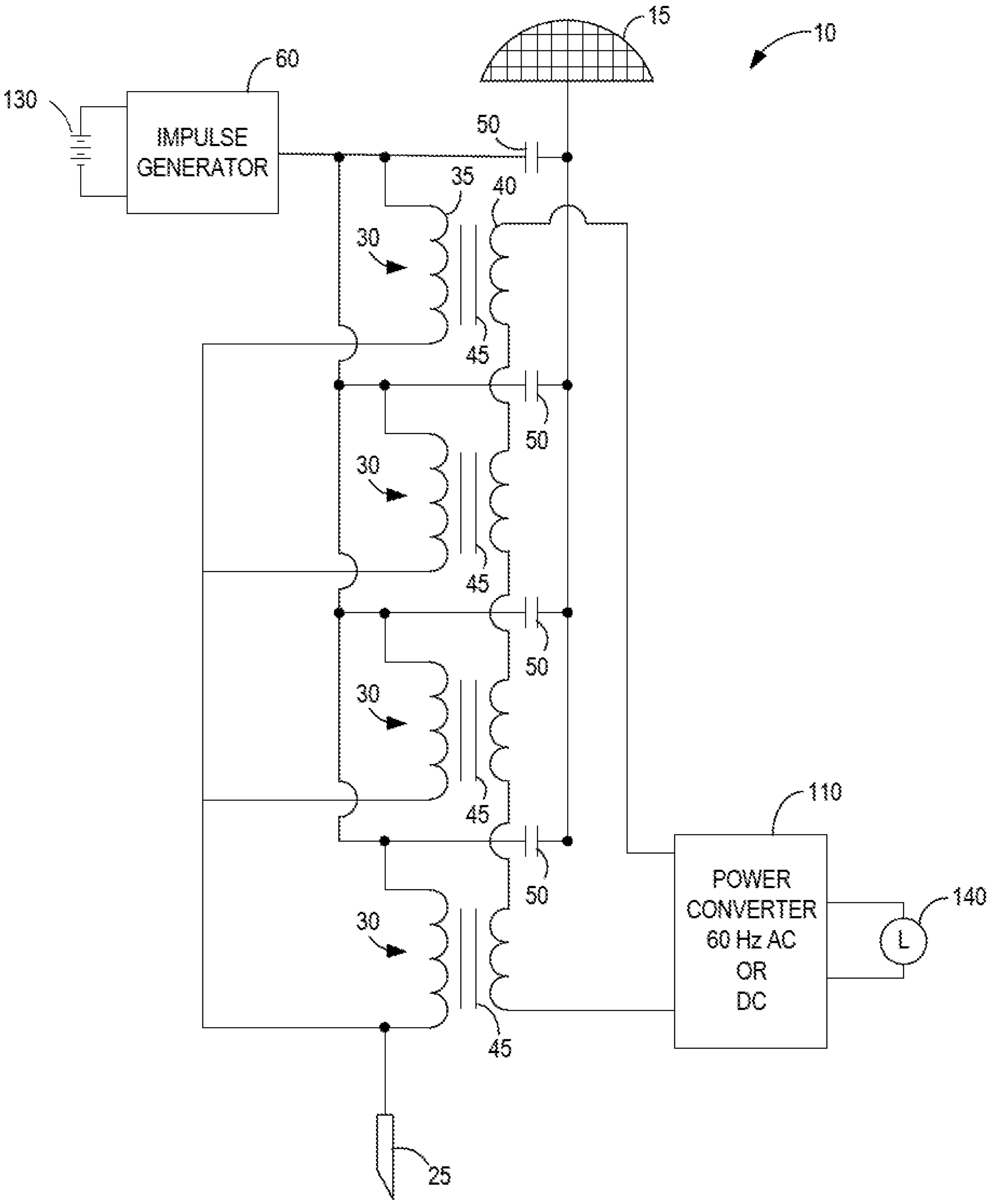
IMPULSE GENERATOR
POWER CONVERTER 60 Hz AC OR DC
L

FIG 5

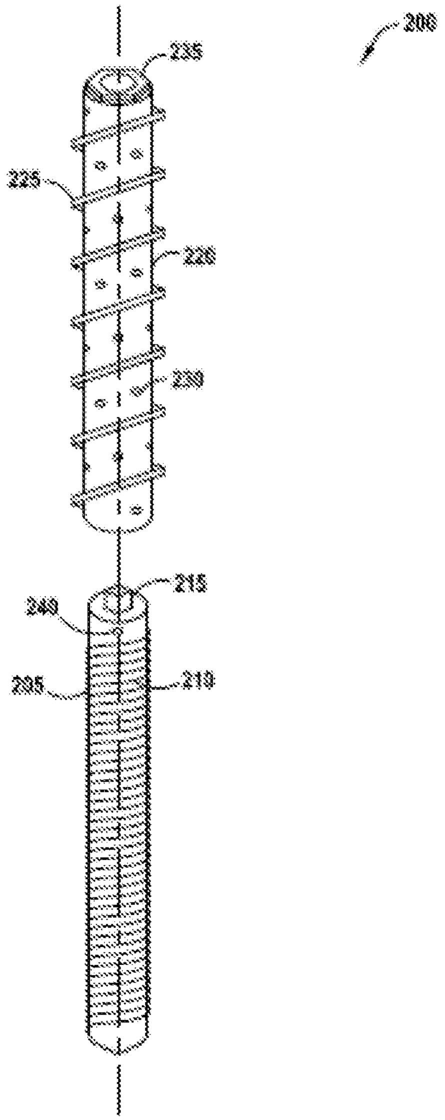
200
235
225
220
230
215
240
235
210

FIG 6A

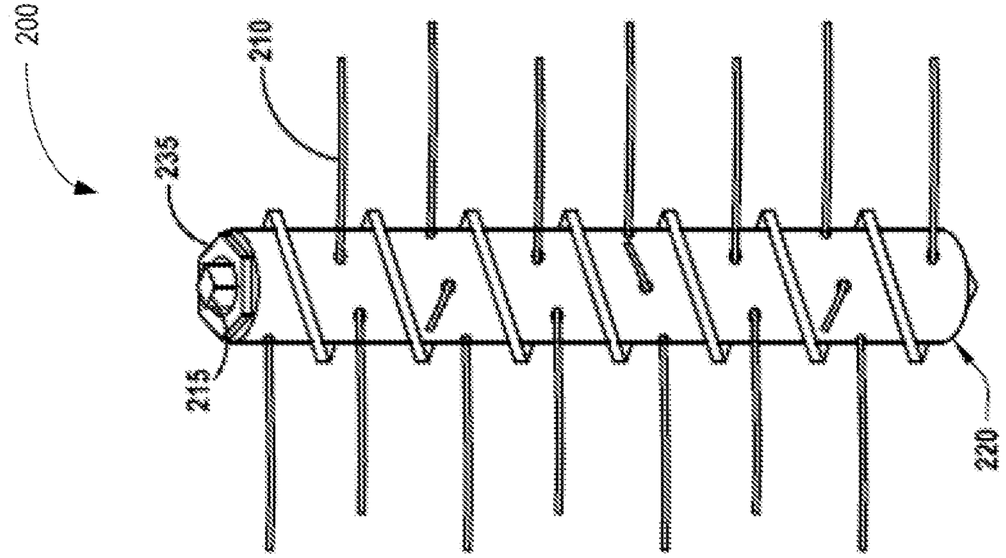
200
235
210
215
220

FIG 6C

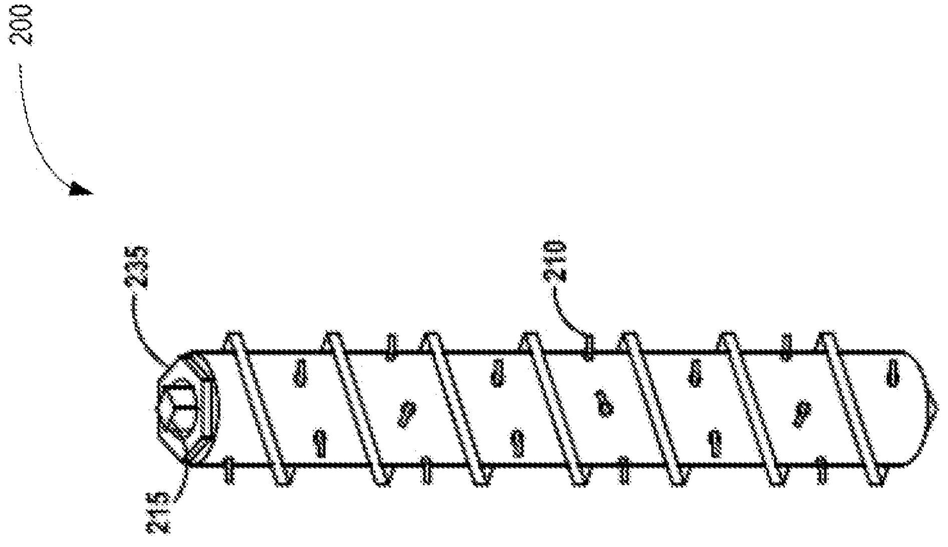
200
235
210
215

FIG 6B

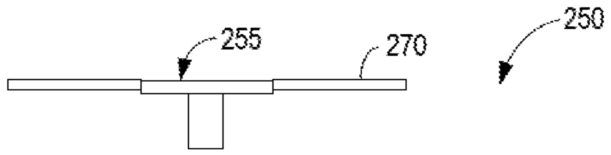

FIG 7A

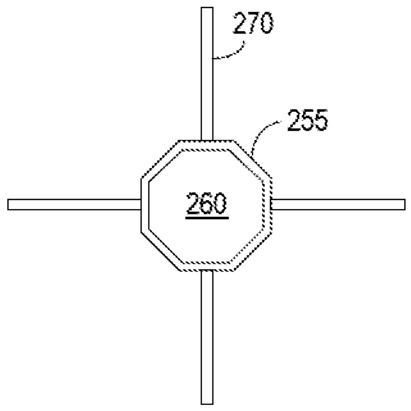

FIG 7B

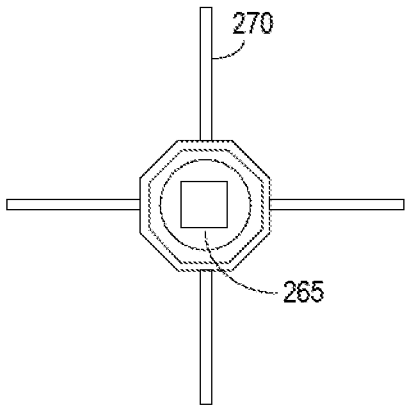

FIG 7C

POWER RECEIVER FOR EXTRACTING POWER FROM ELECTRIC FIELD ENERGY IN THE EARTH

[0001] This application is a continuation of prior U.S. application Ser. No. 15/397,281 filed Jan. 3, 2017, which is a continuation-in-part of prior U.S. application Ser. No. 14/509,772 filed Oct. 8, 2014, U.S. Pat. No. 9,564,268, which claims the benefit of U.S. Provisional Application No. 61/889,894 filed Oct. 11, 2013, the disclosures of all of which are incorporated by reference herein in their entirety.

TECHNICAL FIELD

[0002] The present invention relates generally to renewable energy, and more particularly to methods and apparatus for extracting energy from subsurface electrical fields beneath the earth's surface.

BACKGROUND

[0003] The earth and the ionosphere cavity may be viewed as a global electric circuit. Electrical currents are constantly flowing within the earth and its atmosphere. Within the earth, the majority of the earth's energy is carried by extremely low frequency (ELF) and ultralow frequency (ULF) waves in the 0-200 Hz frequency range. The earth's rotating magnetic field and positive lightning are two energy sources that sustain the ELF/ULF waves within the earth and the atmosphere.

[0004] A great deal of research has been devoted to studying the electric field present in the earth's ionosphere cavity. Joseph M. Crawley, the "Fair Weather Atmosphere as a Power Source", Proceedings ESA Annual Meeting on Electrostatics 2011; O. Jefimenko, "Operation of Electric Motors from Atmospheric Electric Field," American Journal of Physics, Vol. 39, Pgs. 776-779, 1971; M. L. Breuer, "Usability of Tapping Atmospheric Charge as a Power Source," Renewable Energy, Vol. 28, Pgs. 1121-1127, 2003. Numerous attempts have been made in the past to extract electrical energy from the earth's atmosphere. For example, U.S. Pat. No. 1,540,998 to Plauson describes a system for converting atmospheric electrical energy into usable power. These past attempts have been successful in producing only small amounts of power from the electrical field in the earth's ionosphere cavity. The modest success of these experiments compared to results from other renewable energy sources, such as solar and wind, has tempered further research and prevented widespread use of the electric field in the ionosphere cavity as an energy source.

SUMMARY

[0005] The present invention relates to a power receiver for extracting power from electric fields beneath the earth's surface. In embodiments of the present disclosure, a resonant transformer connected to a ground terminal draws current from the earth's electric field through the primary winding of the transformer. Current flow through the resonant transformer is induced by applying a high voltage impulse to the primary winding. The flow of current through the primary winding of the resonant transformer induces a current in the secondary winding, which may be converted and filtered to a usable form, e.g. 60 Hz AC or DC.

[0006] In some embodiments of the power receiver, the resonant frequency of the resonant transformers is below 200 Hz.

[0007] In some embodiments of the power receiver, the resonant transformer comprises a ferro-resonant transformer.

[0008] In some embodiments, the power receiver further comprises an elevated terminal.

[0009] In some embodiments of the power receiver, the primary winding of the resonant transformer is connected between the ground terminal and elevated terminal.

[0010] In some embodiments of the power receiver, the elevated terminal comprises an upper capacitive plate coupled to the earth's ionosphere cavity.

[0011] In some embodiments of the power receiver, the impulse generator comprises the upper capacitive plate and a spark gap connected between the upper capacitive plate and the primary winding of the resonant transformer. The spark gap comprises a pair of electrodes separated by a gap and configured to generate a spark when a voltage difference between the electrodes reaches a predetermined level.

[0012] In some embodiments of the power receiver, the impulse generator comprises a pulse generator for generating low voltage pulses, a step-up transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses, and a spark gap connected between the step-up transformer and the primary winding of the resonant transformer to generate a spark responsive to the high voltage impulses from the step-up transformer.

[0013] In some embodiments of the power receiver, the impulse generator comprises a pulse generator for generating low voltage pulses, and a step-up transformer connected to the primary winding of the resonant transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses.

[0014] In some embodiments of the power receiver, the impulse generator comprises a solid state spark generator.

[0015] In some embodiments of the power receiver, the resonant transformer includes a capacitor connected in parallel with the primary winding.

[0016] In some embodiments of the power receiver, the resonant transformer includes a capacitor connected in series with the primary winding between the impulse generator and the elevated terminal.

[0017] In some embodiments, the power receiver comprises multiple resonant transformers having primary windings connected in parallel between the ground terminal and the elevated terminal.

[0018] In some embodiments of the power receiver, the resonant transformers have different resonant frequencies.

[0019] In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are all below 200 Hz.

[0020] In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

[0021] Another embodiment of the power receiver comprises a resonant circuit connected to a ground terminal disposed below the surface of the earth, an impulse generator for generating and applying a high voltage electrical impulse to the resonant circuit to induce current flow from the ground terminal through the resonant circuit, and a power conversion circuit connected to the resonant circuit to convert electrical current flowing through the resonant circuit to a desired form. The resonant circuit has a resonant frequency below 200 Hertz.

[0022] In some embodiments of the power receiver, the resonant circuit comprises a resonant transformer having a primary winding, a secondary winding, and resonant capacitor connected in series with the primary winding.

[0023] In some embodiments of the power receiver, the resonant circuit comprises multiple resonant transformers having primary windings connected in parallel to the ground terminal.

[0024] In some embodiments of the power receiver, the resonant transformers have different resonant frequencies.

[0025] In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are all below 200 Hz.

[0026] In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

[0027] Other embodiments of the disclosure comprise a ground terminal for a power receiver. In one embodiment, the ground terminal comprises a ground shaft configured for insertion beneath the surface of the earth, a hollow cylinder surrounding the ground shaft and having a plurality of openings, and a plurality of ground wires connected at one end to the ground shaft. The ground wires are wound around the ground shaft and have free ends protruding through respective openings in the hollow shaft so that rotation of the ground shaft relative to the hollow cylinder causes the ground wires to extend radially into the earth.

[0028] Other embodiments of the disclosure comprise methods of extracting power from the earth. In one embodiment, the method comprises applying a high voltage impulse to resonant circuit coupled to a ground terminal disposed beneath the surface of the earth to initiate resonance in the resonant circuit and induce the flow of current from the ground terminal to the resonant circuit, and converting the current flowing from the ground terminal into the resonant circuit into a useful form.

[0029] In some embodiments of the method, the resonant circuit comprises a resonant transformer including a primary winding coupled to the ground terminal and a second winding coupled to a power converter, and applying a high voltage impulse to resonant circuit comprises applying a high voltage impulse to the primary winding of the resonant transformer.

[0030] In some embodiments of the method, applying a high voltage impulse to the primary winding of the resonant transformer comprises applying an impulse in the range to 10,000 to 40,000 volts to primary winding of the transformer.

BRIEF DESCRIPTION OF THE DRAWINGS

[0031] FIG. **1** illustrates a first exemplary embodiment of a power receiver.

[0032] FIG. **2** illustrates a second exemplary embodiment of a power receiver.

[0033] FIG. **3** illustrates a third exemplary embodiment of a power receiver.

[0034] FIG. **4** illustrates a fourth exemplary embodiment of a power receiver.

[0035] FIG. **5** illustrates a fifth exemplary embodiment of a power receiver.

[0036] FIG. **6**A is an exploded perspective view of an exemplary ground antenna array for the power receiver.

[0037] FIG. **6**B is a perspective view of an assembled ground antenna array before being deployed.

[0038] FIG. **6**C is a perspective view of an assembled ground antenna array after being deployed.

[0039] FIG. **7**A is a side view of an insertion tool for installing the ground antenna array.

[0040] FIG. **7**B is a top view of the insertion tool for installing the ground antenna array.

[0041] FIG. **7**C is a bottom view of the insertion tool for installing the ground antenna array.

DETAILED DESCRIPTION

[0042] Referring now to the drawings, a power receiver for extracting energy from the earth's electric field are illustrated and indicated generally by the numeral 10. Various embodiments of the power receiver **10** are described and similar reference numbers are used throughout the description to indicate similar components.

[0043] The power receiver **10** converts energy in the ELF/ULF waves to useful form, e.g. 60 Hz AC or DC. The power receiver **10** is essentially a resonance circuit that resonates at the natural resonance frequencies in the earth's electric field. These resonance frequencies, known as Schumann resonance frequencies, occur at 7.83 Hz, 14.3 Hz, 20.8 Hz, 27.3 Hz, and 33.8 Hz. A high voltage impulse initiates resonance within the power receiver **10**. In the resonant mode, the impedance of the power receiver **10** is reduced to near zero thus inducing ground currents to flow into the power receiver **10** where the ground currents are converted to useful form.

[0044] FIG. **1** illustrates a first embodiment of the power receiver **10**. The power receiver **10** comprises a resonant transformer **30** connected between an elevated terminal **15** and ground terminal **25**. In this embodiment, the elevated terminal **15** is capacitively coupled to electric fields within the earth's ionosphere cavity and functions as an upper capacitive plate. A lower capacitive plate **20** is connected to the ground terminal **25** beneath the surface of the earth.

[0045] The resonant transformer **30** comprises a primary winding **35**, secondary winding **40**, ferromagnetic core **45**, and capacitor **50**. One end of the primary winding **35** is connected to the lower capacitive plate **20** and ground terminal **25**. The opposite end of the primary winding **35** is connected via a spark gap **90** to the elevated terminal **15**. The capacitor **50** is connected in parallel with the primary winding **35** of the resonant transformer **30** to form an LC circuit **55** with a resonance frequency range of between about 0.1 and 200 Hz. In a preferred embodiment, the resonant transformer has a Q of about 10 or greater and resonance frequency in the range of about 0.1 -200 Hertz. For example, the resonant transformer **30** may have a resonance frequency of about 7.83 Hz, the fundamental Schumann resonance frequency. The secondary winding **40** of the resonant transformer **30** is connected to a power converter **110** as will be hereinafter described in greater detail. The power converter **110** converts the energy extracted from the earth's electric field by the power receiver **10** into a usable form for driving a load **140**.

[0046] The elevated terminal/upper capacitive plate **15** comprises an insulated, dish-shaped plate with a large radius of curvature. The capacitance and resistance of the elevated terminal is chosen for receiving broadband electric field frequencies in the 0-200 Hz range. The upper capacitive plate **15** is sized to maximize to the extent practical coupling with the electric field in the earth's ionosphere cavity.

[0047] The lower capacitive plate **20** is also a dish-shaped plate with a large radius of curvature. One function of the lower capacitive plate **20** is to collect charge from the earth's ground currents and provide an instantaneous source of current as hereinafter described. The capacitance and resistance of the lower capacitive plate **20** is selected to promote the flow of current from the ground with minimal losses.

[0048] The spark gap **90** connected between the elevated terminal **15** and resonant transformer **30** comprises a pair of electrodes **95**, **100** separated by an evacuated air gap **105**. Electrode **95** is connected to the upper capacitive plate **15**. Electrode **100** is connected to the resonant transformer **30**. The spark gap **105** prevents electrical discharge from the upper capacitive plate **15** to the earth's ionosphere cavity. The spark gap **90** in combination with the elevated terminal **15** function as an impulse generator that applies a high voltage impulse of about 10,000-40,000 volts to the primary winding **35** to initiate resonance in the transformer **30**.

[0049] In operation, the capacitive coupling of the upper capacitive plate **15** induces a high voltage operating current in the upper capacitive plate **15**. The upper capacitive plate is connected to a first electrode **95** to the spark gap **90**. When the voltage difference between the electrodes **95** and **100** reaches a threshold, a spark forms across the electrodes **95**,**100** and a high voltage impulse is applied to the primary winding **35** of the resonant transformer **30**. This high voltage impulse initiates resonance within the transformer **30**.

[0050] In resonant mode, the impedance of the resonance transformer is reduced to nearly zero allowing current to flow from the capacitive plate **20** and ground terminal **25** through the primary winding **35** of the transformer **30**, which in turn induces current in the secondary winding **40**. Power converter **110** converts the current flowing through the secondary winding **40** into a usable form for driving a load **140**. The transformer **30** will continue to resonate for a short period of time. By providing high voltage impulses to the primary winding **35** of the resonant transformer **30** at periodic intervals, it is possible to maintain a continuous flow of current from the earth into the resonant transformer **30**, thus producing a continuous supply of power.

[0051] FIG. **2** discloses a second embodiment of the primary receiver **10**. This embodiment includes a resonant transformer **30** connected between an elevated terminal **15** and ground terminal **25**. The resonant transformer **30** comprises a primary winding **35**, secondary winding **40**, ferromagnetic core **45** and a high voltage capacitor **50**. One end of the primary winding **35** is connected to the ground terminal **25**. The opposite end of the primary winding **35** is connected to the elevated terminal **15**. The capacitor **50** has a capacitance of about 0.01 micro-farads. In contrast to the previous embodiment, capacitor **50** is connected in series with the primary winding **35** and elevated terminal **15** and forms a LC circuit **55** with a Q of about 10 or greater and a resonance frequency in the range of about 0.1 to 200 Hz. In a preferred embodiment, the resonance frequency of the transformer **30** is 7.83 Hz, the fundamental Schumann resonance frequency. An impulse generator **60** is connected between the primary winding **35** of the resonant transformer **30** and the series capacitor **50** and applies a high voltage impulse to the primary winding **35** of the resonant transformer **30**. A battery **130** or other external power source supplies power to the impulse generator **60**. As previously described, the high voltage impulse applied by the impulse generator **60** initiates resonance within the resonant transformer **30** inducing current flow from the ground terminal **25** into the primary winding **35** of the resonant transformer **30**. The flow of current from the ground terminal **25** through the primary winding **35** induces current in the secondary winding **40**. Power converter **110** converts the electrical energy in the current flowing through the primary winding **40** into a usable form.

[0052] In contrast to the first embodiment, it is not required to capacitively couple the elevated terminal **15** in the second embodiment to the earth's ionosphere cavity. Rather, the elevated terminal **15** in this embodiment provides lightning protection and dissipates some of the energy flowing into the power receiver **10** to the earth's ionosphere cavity. Also, in contrast to the first embodiment, the capacitor **50** is connected in series between the primary winding **35** of the transformer **30** and the elevated terminal **15**. Those skilled in the art will appreciate that the capacitor **50** could also be connected in parallel rather than series with the primary winding **35** as shown in FIG. **1**. Another difference is that the impulse generator **60** has an external power source. The amount of energy generated by the power receiver **10**, however, is far greater than the energy needed to generate high voltage impulses. The first embodiment does not require an external power source to generate high voltage impulses.

[0053] FIG. **3** illustrates a third embodiment of the power receiver **10**. This embodiment is essentially the same as the embodiment shown in FIG. **2**. The main difference is that a center tap of the primary winding **35** in the resonant transformer **30** is connected to an electrical ground **85**. It should be appreciated that the electrical ground **85** may be different than the earth ground. When the center tap of the resonant transformer **30** is grounded at a distance away from the ground terminal **25** (e.g. 50 ft to 100 ft), the power receiver **10** becomes a transmitter via the ground loop formed.

[0054] FIG. **4** illustrates the power receiver **10** of FIG. **2** in greater detail. The power receiver includes a resonant transformer **30** connected between a ground terminal **25** and an elevated terminal **15**. The ground terminal **25** may comprise a ⅝-inch×8-foot copper ground rod, such as the ERICO 615880UPC. The elevated terminal **15** may comprise a 90% copper mesh formed into a hemisphere with a radius of about 9 inches. The elevated terminal **15** may be elevated at a height of approximately 6 feet above the ground.

[0055] The resonant transformer **30** includes a primary winding **35**, secondary winding **40**, ferromagnetic core **45** and series capacitor **50** configured as previously described. The resonant transformer **30** may have a Q of about 10 and a resonance frequency in the range of about 0.1 to 200 Hz. The resonant transformer **30** may be made using an Allanson transformer (part #1530BP120R) connected in series with a 0.01 micro-farad capacitor, such as the Condensor Products high voltage capacitor (part #TC 103-17-125). The resonant transformer **30** is used in a step-down configuration. The center tap of the resonant transformer **30** may optionally be connected to a ground.

[0056] An impulse generator **60** is connected between the primary winding **35** of the resonant transformer **30** and the series capacitor **50** and applies a high voltage impulse in the range of about 10,000 to 40,000 volts to the primary winding of the transformer **30**. A battery **130** or other external power source supplies power to the impulse generator **60**. The power converter **110** connects to the secondary winding **40** of the resonant transformer **30** for converting current in the secondary winding of the transformer to a useful form.

[0057] The impulse generator **60** comprises a pulse generator **65** for generating low voltage pulses, a step-up transformer **80** for converting the low voltage pulses from the pulse generator **65** to high voltage pulses, and a spark gap **90** for generating sparks responsive to the high voltage pulses from the step-up transformer **80**.

[0058] The pulse generator **65** comprises a square wave generator **70**, such as a Sinometer VC2002 function signal generator, and solid state relay **75**. The square wave generator **70** generates a digital pulse stream. In one embodiment, the digital pulse stream generates a square waveform with a frequency of about 7.83 Hz. The frequency of the digital pulse stream is selected to match the resonance frequency of the transformer **30**, though such is not necessarily required. The pulse stream output from the square wave generator **70** is applied to the solid state relay **75**. The solid state relay **75** is connected between a battery or other power source and a first winding of the step-up transformer **80**. The battery may comprise a 12 V, 7.0 A/H sealed lead acid battery, such as the ELB 1270A by Lithonia Lighting. The solid state relay **75** functions as a switch that is activated responsive to the waveform from the square wave generator **70** to provide a continuous stream of low voltage pulses from the battery to the first winding of the step-up transformer **80**. A 1 ohm resistor is connected between the solid state relay **75** and step-up transformer **80**.

[0059] The step-up transformer **80** may comprise a Transco 15 kV, 30 mA neon sign transformer (part # S15612). The step-up transformer **80** converts the low voltage pulses from the pulse generator **65** to high voltage pulses that are applied to the spark gap **90**. The step-up transformer has a 0.5 micro-farad capacitor connected in parallel with the primary winding of the step-up transformer **80**. The step-up transformer produces pulses at the output of about 30,000 to 40,000 volts.

[0060] The spark gap **90** comprises a pair of electrodes **95**, **100** separated by an air gap **105**. A suitable spark gap electrode pair is the Information Unlimited SPARK05 ¼-inch×1-inch tungsten electrodes. As previously described, when the voltage potential between the electrodes **95**, **100** reaches a threshold, a spark forms between the electrodes **95**, **100** and supplies a nearly instantaneous, high voltage impulse to the primary winding **35** of the resonant transformer **30**. This high voltage impulse initiates resonance in the resonant transformer **30** inducing current flow from the ground terminal **25** through the primary winding **35** of the resonant transformer **30**.

[0061] The power converter **110** comprises a bridge rectifier **115**, filter capacitor **120**, charge controller **125**, and inverter **135**. A suitable rectifier is the Micro Commercial Components 10 amp, 1000 volt bridge rectifier (Part #GBJL 1010). The bridge rectifier **115** converts the AC current flowing through the secondary winding **40** of the resonant transformer to a DC current. A filter capacitor **120** removes unwanted frequencies from the DC current. A suitable capacitor **120** is Cornell Dubilier 1000 uF 450 VDC capacitor (part #383LX102M450N082). The filter capacitor **120** has a capacitance of about 1000 micro-farads. The DC current is input to the charge controller **125**. The charge controller **125** may, for example, comprise a maximum power point tracking (MPPT) charge controller, such as a Tracer 4215 BN MPPT Solar Charge Controller, which is commonly used in solar power generating systems. The charge controller **125** applies a small amount of energy to a battery **30** to charge the battery **130**. As previously noted, the battery **130** serves as a power source for the impulse generator **60**. The remaining current is supplied to an inverter **135**, which converts the DC current to an AC current with a desired voltage and frequency, e.g., 120 volts/60 Hz AC. A suitable inverter **135** is the 1500 W Pure Sine power inverter (AIMS) (part # PWRI1500125). The power converter **110** as shown in FIG. **4** may be utilized in the embodiment shown in FIGS. **1**, **2** and **3**.

[0062] FIG. **5** illustrates a power receiver **10** according to another embodiment. The power receiver **10** comprises a plurality of resonant transformers **30** connected between a ground terminal **25** and elevated terminal **15**. Each of the resonant transformers **30** comprises a primary winding **35**, secondary winding **40**, ferromagnetic core **45** and series capacitor **50**. The primary windings **35** of the resonant transformers **30** are connected in parallel. The secondary windings **40** are connected in series with the power converter **110**. An impulse generator **60** applies a high voltage impulse to the primary windings **35** of the resonant transformers **30**. A battery **130** or other external power source supplies power to the impulse generator **60**. The power converter **110** converts the current in the power converter circuit to a usable form for driving a load **140**.

[0063] In one embodiment, each of the resonant transformers **30** shown in FIG. **5** is configured to have a different resonant frequency. In one embodiment, the resonant transformers **30** are configured to resonate at frequencies of 7.83 Hz, 14.8 Hz, 20.3 Hz and 26.8 Hz respectively. Additional resonant transformers **30** could be added to operate at other resonance frequencies.

[0064] FIGS. **6A-6C** illustrate a high quality ground antenna array **200** which may be used as a ground terminal **25**. The ground antenna array **200** comprises a generally cylindrical ground shaft **205** disposed with a hollow cylinder **220** and a plurality of reinforced, heavy gauge ground wires **210** attached at one end to the ground shaft **205**. The ground shaft **205** and ground wires **210** should be highly conductive and have low resistance to supply current from the ground to the power receiver **10**. In one embodiment, the ground wires **210** may be copper or other highly-conductive metal. The end of the ground shaft may be pointed to facilitate insertion into the earth. A connection port on the ground shaft **220** is provided to electrically connect the ground antenna array **220** to the resonant transformer **30**.

[0065] The hollow cylinder **220** has external threads **25** to facilitate insertion into the ground. A rotator nut **235** is fixedly secured to the top end of the hollow shaft **220**. A square shaft **215** protrudes from the top end of the ground shaft **205** into the opening in the rotator nut **235**. FIG. **6B**. A tool **250**, shown in FIG. **7**, engages with the rotator nut **235** and square shaft **215** during insertion of the ground antennas array **200** into the ground as will be hereinafter described.

[0066] The insertion tool **250** is shown in FIG. **7**. The insertion tool **250** includes a tool body **255** having a first socket **260** on one side to fit the rotator nut **235** on the hollow cylinder **220** and a second socket **265** on the other side to fit the square shaft **215** on the ground shaft **205**. Arms **270**

extend from the outer periphery of the tool body **255** for manually or mechanically turning the insertion tool **250**.

[0067] Before the antenna array **200** is deployed, the ground wires **210** are wound around the ground shaft **205** with the free ends protruding slightly from respective openings **230** in the hollow cylinder **220** to a distance not to exceed one half (½) the depth of the external threads **225** on the hollow cylinder **220**. FIG. **6**B illustrates the ground antenna array **200** before deployment. FIG. **6**C illustrates the ground antenna array in a deployed configuration.

[0068] Installation of the ground antenna array **200** is performed in two stages. In the first stage, a hole slightly smaller in diameter than the threads **235** of the hollow cylinder **220** is drilled into the Earth to a depth matching the length of the hollow cylinder **220** or slightly longer. The hole is filled with water and the water is allowed to soak into the soil. After the ground is softened, the hollow cylinder **220** is rotated using the insertion tool **250** to insert the ground antenna array **200** into the ground. The first socket **260** of the insertion tool **250** is engaged with the rotator nut **230** and the insertion tool **250** is turned by hand or a mechanized rotating shaft fitted and attached to the tool arms **270** to thread the ground assembly into the hole. During the initial insertion of the ground antenna array **200**, the ground shaft **205** is fixed to the hollow shaft **220** and rotates with the hollow shaft. The hollow cylinder **220** is rotated until it reaches the full depth of the hole.

[0069] Once the ground antenna array **200** has been fully inserted into the earth, the insertion tool **250** is flipped over and the second socket **265** of the insertion tool **250** is engaged with the square shaft **215**. The insertion tool **250** is turned by hand or a mechanized rotating shaft fitted and attached to the tool arms **270** to rotate the ground shaft **205**. During the second phase, the ground shaft **205** rotates freely inside the hollow cylinder **220**. Rotation of the ground shaft **205** causes the reinforced ground wires **210** to extend radially into the earth. The ground shaft **220** is rotated until the ground wires are fully extended. The ends of the ground wires may be sharpened to aid in the extension of the ground wires during the second phase.

[0070] After the ground antenna array **200** is deployed, a connection cable **280** is attached to a connection port **240** on the ground shaft **220** to electrically connect the ground antenna array **220** to the resonant transformer **30** in the power receiver **10**.

What is claimed is:

1. A power receiver for extracting electrical energy from the earth's electric field, said power receiver comprising:

a resonant transformer connected to a ground terminal disposed below the surface of the earth;

an impulse generator for generating and applying a high voltage electrical impulse to a primary winding of the resonant transformer to induce current flow from the ground terminal through the primary winding of the transformer; and

a power converter connected to a secondary winding of the resonant transformer to convert an alternating current flowing through the secondary winding to a direct current.

2. The power receiver of claim **1** wherein a resonant frequency of the resonant transformers is below 200 Hz.

3. The power receiver of claim **1** wherein the resonant transformer comprises a ferro-resonant transformer.

4. The power receiver of claim **1** wherein the impulse generator comprises:

a pulse generator for generating low voltage pulses;

a step-up transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses;

a spark gap connected between the step-up transformer and the primary winding of the resonant transformer to generate a spark responsive to the high voltage impulses from the step-up transformer.

5. The power receiver of claim **1** wherein the impulse generator comprises a solid state spark generator.

6. The power receiver of claim **1** wherein said resonant transformer includes a capacitor connected in parallel with the primary winding.

7. The power receiver of claim **1** wherein said resonant transformer includes a capacitor connected in series with the primary winding.

8. A method of extracting electrical energy from the earth's electric field, the method comprising:

applying a high voltage electrical impulse to a primary winding of a resonant transformer connected to a ground terminal disposed below the surface of the earth to induce current flow from the ground terminal through the primary winding of the transformer; and

converting an alternating current flowing through a secondary winding of the resonant transformer to a direct current.

9. The method of claim **8** wherein a resonant frequency of the resonant transformers is below 200 Hz.

10. The method of claim **8** wherein the resonant transformer comprises a ferro-resonant transformer.

11. The method of claim **8** wherein the impulse generator comprises:

a pulse generator for generating low voltage pulses;

a step-up transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses;

a spark gap connected between the step-up transformer and the primary winding of the resonant transformer to generate a spark responsive to the high voltage impulses from the step-up transformer.

12. The method of claim **8** wherein the impulse generator comprises a solid state spark generator.

13. The method of claim **8** wherein said resonant transformer includes a capacitor connected in parallel with the primary winding.

14. The method of claim **8** wherein said resonant transformer includes a capacitor connected in series with the primary winding.

* * * * *

Diodes

Back to the index

Which diode can I use best in my crystal receiver?

Maybe you think a diode with a voltage drop as low as possible, then also small signals at the detector circuit are detected.
But diodes with a low voltage drop, also have a high reverse current (leaking current), this will load the detector circuit heavier, the Q of the detector circuit reduces, and with that also the voltage across the LC circuit.
At a lower input voltage the diode will give much more losses, and it can happen that despite the lower voltage drop of the diode, you have less voltage at the load resistor.
Besides that, reduction of circuit Q will also gives a less selective receiver.

For every 20 mV less voltage drop, the reverse current will approximately double.

Germanium, silicon, en schottky diodes.

Depending on the material they are made from, we can distinguish germanium diodes, silicon diodes and schottky diodes.
There are some more types, which are not discussed here.

Silicon diodes have the highest voltage drop (about 0.5 Volt) and are for this reason not very useable for crystal receivers.
Unless we use a small DC bias current, which brings the diode already a little bit in conduction.

Germanium diodes have a low voltage drop (about 0.1 - 0.2 Volt) and are often used in crystal receivers.
The properties like voltage drop and reverse current can vary a lot between two germanium diodes of the same type.
In practice we can best test several germanium diodes in our receiver and then choose the best.
The diode resistance RD of germanium diodes is most times rather low, and only useable in crystal receivers with a low Q (low sensitivity and low selectivity).
For high performance receivers, we can better use a suitable schottky diode.

Schottky diodes have a voltage drop of about 0.25 Volt.
The differences in properties between two diodes of the same type are often small.
Schottky diodes with the correct resistance RD are very useable in high quality crystal receivers.

The given voltage drop is normally measured at a forward current of about 1 mA.
Also if we measure the voltage drop of a diode with a multimeter, the test current shall be about 1 mA.
But also below this voltage drop the diode can conduct current, and can rectify a RF (radio frequency) signal.
Only the current through the diode is then much smaller.
When receiving very weak stations, the current through the diode can be e.g. only 10 nA.
At such a low current, the voltage drop of the diode is also much lower then at 1 mA.

Detected voltage as function of the input voltage

If we rectify a RF signal with a diode we can distinguish two situations.

Situation 1: Rectifying in the linear region
If the input voltage is high enough (well above the voltage drop of the diode at 1 mA), the output voltage of the diode will be about proportional to the input voltage.
So double input voltage, gives about double output voltage.

The output voltage is almost equal to the peak value of the input voltage.
The power losses in the diode are in this region very low compared to the rectified power.

Situation 2: Recifying in the square law region
If the input voltage is low, lower then the voltage drop of the diode (at 1 mA) then the situation is completely different.
The input of the diode behaves for the RF signal like a resistor with value RD.
The output of the diode behaves like a DC voltage source in series with a resistor, the value of this resistor is also equal to RD.
The value of the DC voltage source is square law related to the amplitude of the RF input signal.
So double input voltage, gives 4 times as much detected DC voltage at the output
In the square law region the output voltage of the diode will be much lower then the input voltage, the diode gives much power loss between input and output.
The lower the input voltage, the higher the losses.
The higher the input voltage, the lower the diode losses.
When further increasing the diode input voltage, we gradually come into the linear detection region.
When receiving weak stations, detection takes place in the square law region.

Between the linear and square law region, there is a region not linear, and not square law but somewhere in between.
This region is not discussed here.

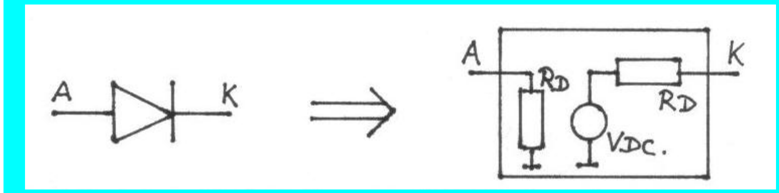

Equivalent circuit of a diode at low input voltages.

Via this link you find a measurement on several schottky diodes, which shows detection in the square law region takes place at input voltages below 200 mVpp.

Diode resistance RD.

At zero voltage, diodes have a certain resistance.
This resistance at zero Volt we call RD.
The lower the reverse leaking current of the diode, the higher resistance RD.
When detecting small signals (in the square law region) the input of the diode also behaves like a resistor with value RD.

But how do we know the RD of a diode?
We can calculate it with the formula:
formula 1: **RD = 0.000086171 x n x TK / Is**

RD = diode resistance at zero Volt (unit: Ohm)
n = *ideality factor*, the lower this factor the better, between 1.0 and 1.1 is a very good value.
TK = temperature in Kelvin (= temperature in ºC + 273)
Is = *saturation current* (unit: A)
x = multiply

The values of n and Is can (sometimes) be found in the diode datasheet.

In the following table some types of schottky diodes, with the values for n, Is and Rd, the maximum reverse voltage and the diode capacitance at zero voltage.

type diode	n	Is at 25 ºC	RD at 25 ºC	maximum reverse voltage	capacitance	
5082-2835	1.08	22 nA	1260 kΩ	8 Volt	1 pF	datasheet
BAT85	?	?	± 200 kΩ ??	30 Volt	10 pF	datasheet
HSMS 2820	1.08	22 nA	1260 kΩ	15 Volt	1 pF	datasheet
HSMS 2850	1.06	3000 nA	9.07 kΩ	2 Volt	0.3 pF	datasheet
HSMS 2860	1.10	38 nA	743 kΩ	4 Volt	0.3 pF	datasheet

To decrease the value RD, we can connect more diodes in parallel, with two the same diodes parallel the value of RD shall halve.
With 3 diodes parallel, the value of RD shall be divided by 3 etc..

Diode resistance when using bias current.

We can decrease the value of RD by sending a small DC bias current (e.g. 0.1 uA) in forward direction through the diode.
The higher the bias current, the lower RD will be.
With the following formula we can calculate the diode resistance RD, when we make use of a DC bias current.

Formula 2: **RD = 0.000086171 x n x TK /(Ib + Is)**

RD= diode resistance at certain DC bias current Ib (unit: Ohm)
n= *ideality factor* of the diode
TK = Temperature in Kelvin (= temperature in ºC + 273)
Ib= DC bias current through the diode in A
Is =*saturation current* of the diode in A

A diode with a certain RD value at a certain bias current, gives the same receiving performance as a diode without bias current with the same RD.

Influence of temperature on "saturation current: Is"

The saturation current (Is-value) is strongly depending on temperature.
A temperature increase of 1 ºC will increase the Is value by about 7 %.
In datasheets, the Is value is most times given at 25 ºC.
If the diode temperature is not 25 ºC, but another value "T", then we must multiply Is with a factor 1.07^(T-25).
T = diode temperature
^ = raise to the power of

Ideality factor n

The ideality factor n of a diode indicates how good the diode performs with regard to an ideal diode.
A (not existing) ideal diode has a value of n=1.
At low input signals, the maximum available detected output power is proportional to 1/n.
So doubling the n will halve the output power (this only applies at weak signals).

Diode capacitance

Between the two connections of the diode there will be a certain capacitance (capacitor value), when this capacitance is fairly high (e.g. 10 pF) the tuning range at high frequencies is limited.

At increasing reverse voltage across the diode the capacitance will reduce, also the detected voltage in a crystal receiver is such a reverse voltage.
Through this, the frequency of the circuit can shift upwards when receiving strong signals.
On the next page: experiments with a detector unit you find in table 3 a measurement about the frequency shift.

Measuring the Is value of a diode.

We can measure the Is value of a diode as follows:
Send a small current through the diode, the value of this current (ID) must be about 1 µA.
Measure the voltage across the diode (VD).

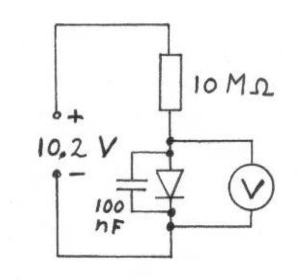

Circuit diagram for measuring the Is value of a diode.
The voltage across the diode is about 0.2 Volt.
The voltage across the resistor is about 10 Volts, so the current is about 1 µA.

The voltmeter must have a resistance of at least 10 MΩ.
The 100 nF capacitor reduces the influence of radio signals and hum on the measurement.

Calculate Is with the formula:

formula 3: **Is = ID / (e^ (VD /(0.0257xn))-1)**

Is = *saturation current* of the diode in nA
ID = current through the diode in nA, (1 µA = 1000 nA)
e = base of the natural logarithms, this is about 2.718
^ = raise to the power of
VD = voltage across the diode in Volt
n = *ideality factor* of the diode, if you don't know this value, take for instance: n= 1.08

More information about measuring the Is you can find on the website of Ben Tongue , in his articles number 4 and 16.

I measured the Is value of several diodes, and calculated the diode resistance RD.
Also some European germanium types are measured.
Several diodes are measured of the type OA95 and AA119

Diode	VD (Volt) at 1 µA	Is (nA)	RD (kΩ).
HSMS282K	0.1341	7.9	3428
HSMS282K 2 parallel	0.118	14.5	1867
HSMS286K (1 diode)	0.1116	18.3	1479
5082-2800	0.1871	1.14	23756
5082-2835	0.1464	5.04	5373
5082-2835 2 parallel	0.1289	9.5	2850
BAT 82	0.136	7.3	3710
BAT 85	0.0686	90.8	298
OA81 (germanium)	0.0225	800	34

OA95 #1 (germanium)	0.0272	600	45
OA95 #2	0.0221	821	33
OA95 #3	0.0271	604	45
OA95 #4	0.0304	502	54
AA116 (germanium)	0.0441	256	106
AA119 #1 (germanium)	0.0320	461	59
AA119 #2	0.0363	370	73
AA119 #3	0.0428	272	100

The HSMS282K is the same as the HSMS2820, only the HSMS282K has 2 equal diodes in one package.

The Is value of the HSMS282K, the HSMS286K and the 5082-2835 is lower than the value in the datasheet, this has also been noticed by other people.
Ben Tongue wrote me, that the Is value of the 5082-2835 has been reduced over the years by the manufacturer.
Also temperature has big influence, I measured at 18 ºC, in datasheets the Is value is given at 25 °C.
A increase from 18 °C to 25 °C will increase the Is by 60 %.

Back to the index

United States Patent [19]
Ensley

[11] **4,456,912**
[45] **Jun. 26, 1984**

[54] HIGH ENERGY MICROWAVE DEFENSE SYSTEM

[75] Inventor: **Donald L. Ensley**, Danville, Calif.

[73] Assignee: **Harvest Queen Mill & Elevator Company**, Plainview, Tex.

[21] Appl. No.: **240,420**

[22] Filed: **Apr. 3, 1972**

[51] **Int. Cl.**^{3} **G01S 13/86**
[52] **U.S. Cl.** .... **343/6 R**; 330/4
[58] **Field of Search** .... 330/4; 343/6 R, 5 R

[56] **References Cited**

U.S. PATENT DOCUMENTS

2,962,585	11/1960	Bolef et al.	330/4 X
3,258,767	6/1966	Jacobsen	330/4 X
3,427,611	2/1969	Enenstein	343/6 R

FOREIGN PATENT DOCUMENTS

1914250	10/1970	Fed. Rep. of Germany	244/3.16

OTHER PUBLICATIONS

"Lasers: Devices and Systems-Part III," by S. Y. Vogel, Electronics, Nov. 1961, pp. 81-85.

Primary Examiner—T. H. Tubbesing
Attorney, Agent, or Firm—Richards, Harris & Medlock

[57] ABSTRACT

A defense system wherein a radar intercepts and follows a target and includes an antenna array and a low level source of microwave energy for exciting the elements of the array to produce radar returns and define the path. A microwave storage reservoir is supplied upon demand with high energy microwave energy and is coupled to microwave feed channels leading to the elements of the antenna array. Microwave dump switches at the juncture of the channels and the reservoir control flow of energy from the reservoir. Control means responsive at least in part to the presence of returns periodically actuate the dump switches for flow of microwave energy from the reservoir to the target via the antenna array.

9 Claims, 5 Drawing Figures

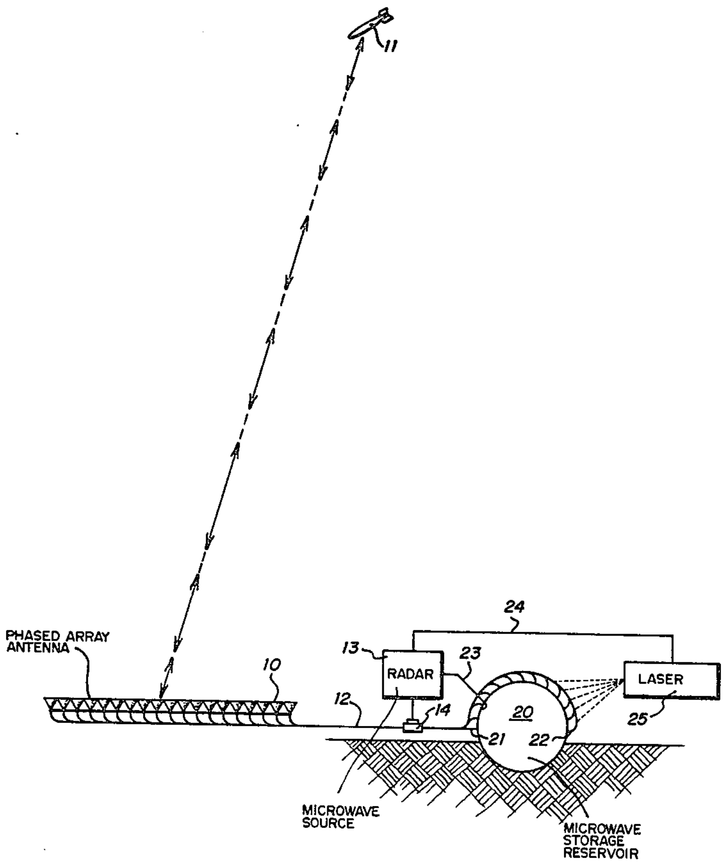

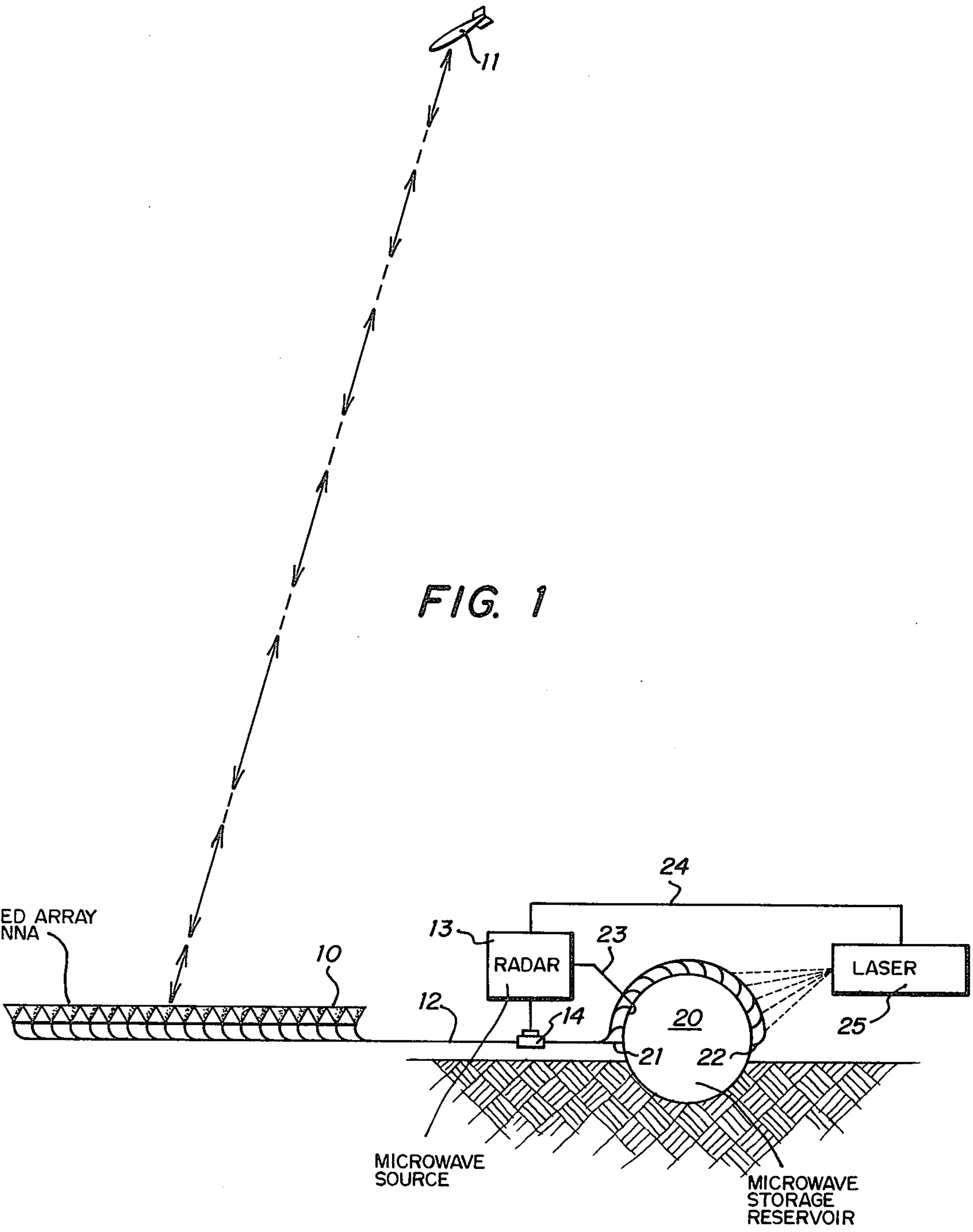
PHASED ARRAY ANTENNA
RADAR
LASER
MICROWAVE SOURCE
MICROWAVE STORAGE RESERVOIR

FIG. 1

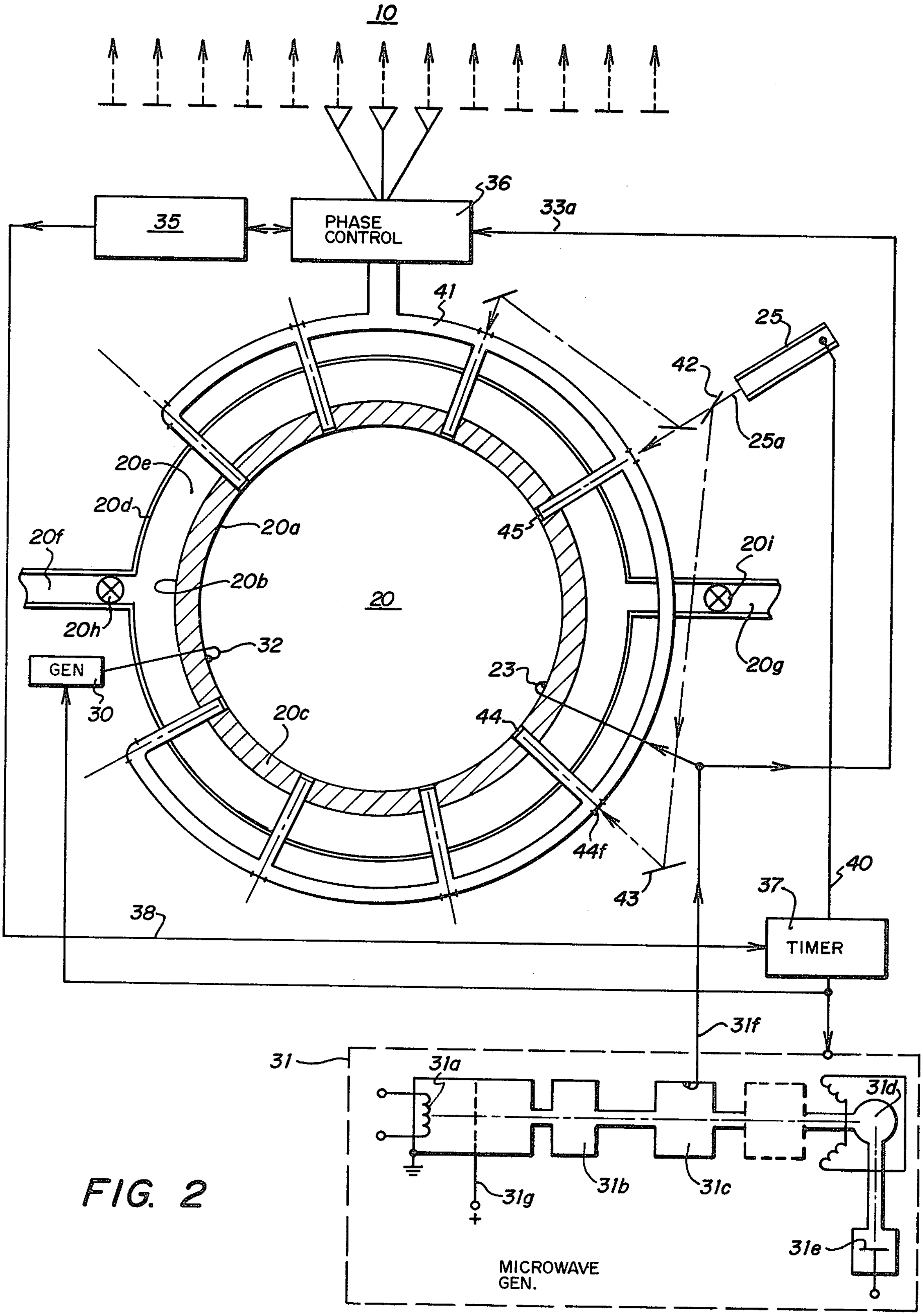
PHASE CONTROL
GEN
TIMER
+
MICROWAVE GEN.

FIG. 2

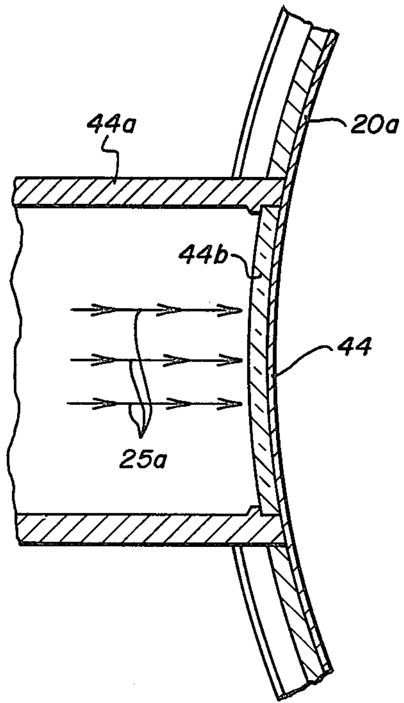
44a
20a
44b
44
25a

FIG. 3

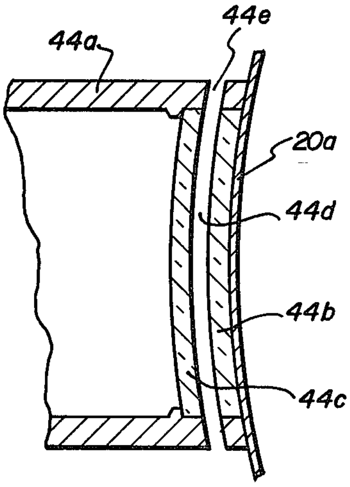
44e
44a
20a
44d
44b
44c

FIG. 4

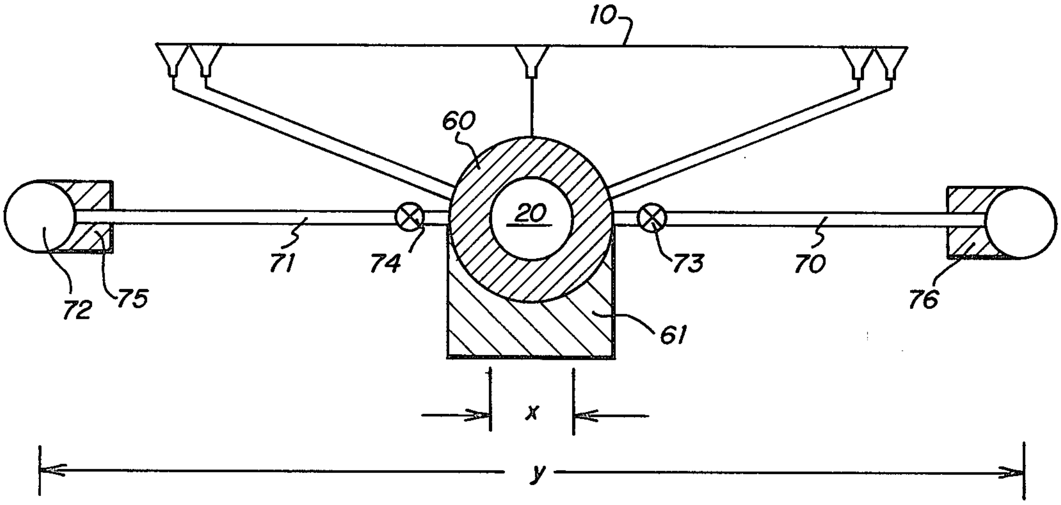
10
60
20
72
75
71
74
73
70
76
61
x
y

FIG. 5

HIGH ENERGY MICROWAVE DEFENSE SYSTEM

This invention relates to the utilization of high energy electromagnetic waves, and more particularly to a method and system for accumulating and releasing high energy electromagnetic waves in a medium such as the earth's atmosphere.

In national defense, expenditures of time and money are being made in the area of ballistic missiles defenses. One purpose is to provide a defense system specifically directed to destruction of enemy missiles during flight long before a target is reached. The high cost of such systems is occasioned by the necessity of propelling and guiding a physical mass along a path at such speeds as to enable it to collide with or otherwise destroy an enemy missile in flight.

The present invention avoids much of the difficulty encountered in such systems by generating and focusing electromagnetic energy of high intensity to a target to impact the target with electromagnetic energy traveling at the speed of light. Radar systems heretofore have been developed for locating bodies traveling at high speeds through the earth's atmosphere and tracking them. Radars are highly developed and provide reliable information which in an early warning sense heralds the approach of missiles, for example, and provides data to define the path and identify the target point.

The present invention is provided to operate in conjunction with a radar system to generate and store a high quantity of electromagnetic energy and then to dump such energy from storage, feeding the same to an antenna for transmission along the path controlled by the radar. The electromagnetic energy will be focused onto the airborne body with such magnitude as to impart destructive action thereto.

More particularly, in accordance with the present invention, a system is provided for operation of a phased array radar to intercept and follow the path of a body spaced from the radar and includes an antenna array and a source of microwave energy for exciting the array to produce target dependent radar returns.

Further, one or more evacuated microwave storage reservoirs are provided with microwave feed channels leading from the reservoir to elements of the antenna array. Microwave dump switches located at the juncture of the channels and the reservoir permit extraction of energy from the reservoir. Means responsive at least in part to the presence of radar returns actuate the dump switches for flow of microwave energy from the reservoir to the body via the antenna array.

In a further aspect, the invention employs the method of focusing a phased array radar onto a body traveling along a path spaced from the radar, repeatedly storing over substantial intervals microwave energy in a microwave storage reservoir, periodically dumping the energy stored from the radar via channels leading to the array for transmission of high energy microwaves to said body. Microwave energy is fed over a substantial period of time into a reservoir having superconductive walls with windows therein normally coated with a superconductive material which responsive to local heating momentarily become transmissive to the microwave energy stored in the reservoir.

For a more complete understanding of the present invention and for further objects and advantages thereof, reference may now be had to the following description taken in conjunction with the accompanying drawings in which:

FIG. 1 is a schematic representation illustrating operation of the present invention;

FIG. 2 is a schematic sectional view of the reservoir of FIG. 1;

FIG. 3 is a view of a portion of the wall of the reservoir of FIG. 1 illustrating a window therein;

FIG. 4 illustrates a modified form in detail of one of the windows;

FIG. 5 illustrates an embodiment of the invention;

Referring now to FIG. 1, a schematic representation of the invention has been shown wherein a phased array radar antenna **10** is provided on the earth's surface for tracking aircraft or projectiles such as the projectile **11** during flight through the earth's atmosphere at a range of the order of a hundred miles. The antenna **10** is fed by way of channels **12** from a radar control unit **13** which is coupled by way of a tee **14** to the channels **12**. The radar **13** in conjunction with antenna **10** operates in a conventional manner to transmit energy into space and to detect returns reflected from projectile **11**.

The channels **12** also lead to a microwave reservoir which is a resonant sphere **20**. The sphere **20** is connected to a plurality of channels, such as channels **21** and **22**, to the antenna **10** all by way of tees represented by the tee **14**.

The radar unit **13** is also coupled by way of channel **23** to sphere **20**. The radar unit **13** is connected by way of a control channel **24** to laser units **25**. As will later be explained, the channels **21** and **22** leading to the sphere **20** are coupled to the sphere **20** by way of dump switches located at the juncture of sphere **20** and the channels **21** and **22**. The dump switches are actuated by energy transmitted by way of laser beams from the laser **25**. Sphere **20** has superconducting walls maintained at a supercooled condition as by liquid helium circulating in the walls. Laser beams from laser **25** are focused onto windows locally to heat them to cause the microwave energy stored within sphere **20** to be dumped into the channels **12** and thence transmitted by way of radar antenna **10** to impinge the projectile **11**. The microwave energy from the radar antenna **10** travels at the speed of light making it possible to repeatedly impact the projectile **11** during its course of flight. The repetition rate depends upon the rate at which build-up of energy in sphere **20** reaches the necessary level to transmit bursts of electromagnetic energy from antenna **10** which would be destructive of the projectile **11** or parts thereof. Thus, high energy bursts of microwave radiation are directed at target **11** which could be, for instance, an I.C.B.M. warhead, a satellite, a submarine launched ballistic missile or an aircraft. The system also acts, at a much lower power level, as its own acquisition phased array radar system.

The high power system is composed of a multiplicity of megawatt power tubes in radar unit **13** coupled to the large superconducting lined resonant cavity **20**. By providing a high "Q" cavity, only minimal power is required to offset high frequency skin losses. Typically cavity **20** may have a radius of 10 meters with a field level of 1×10^6 volt/cm therein. Stored energy of 1×10^8 joules can be maintained during operative periods. Pure niobium forming the walls of resonator **20** has heretofore similarly been employed in systems in which such field levels were present. Other alloys and materials can operate at even higher levels in which case the system dimensions could be reduced. By applying very rapid rise time ($\approx 1\times10^{-8}$ sec.) heating to local areas of thin film superconducting dump switches the stored energy is coupled via a multiplicity of high pressure inert gas horns into the atmosphere without breakdown. In about 1×10^{-6} seconds the energy is coupled out of resonator **20**. Thus, a pulse of duration 1×10^{-6} to 1×10^{-7} seconds with power of 1×10^{14} to 1×10^{15} watts is generated. Such horns form the integral phased array antenna **10**. The antenna thus excited illuminates with a central energy lobe 2×10^{-4} radians wide when using 3 cm radiation from a 500 foot base. At a 100 mile range, the central lobe has a width of 100 feet. An object with a linear dimension of 5 feet would intercept approximately $(5/100)^2$ or ¼% of the burst of radiated energy. This is $\approx 3\times10^5$ joules which would cause significant damage such as blowing off several grams of vaporized material upon absorption. At the same time, shocks would be generated in the object of as much as 100 kilobars.

A recycle time of one second, and hence a pulse rate of 1/second may be maintained by a power source of 1×10^8 watts. A target moving at 5 miles a second would receive about 200 bursts in 100 miles with shocks of varying strengths. By increasing the amount of power available, a pulse rate of 10 per second is feasible and 200 bursts could impact either on one, or several targets within range.

The effects on a body of high energy bursts are twofold: (1) the shock kill mechanism and (2) the material ablation could reach proportions sufficient to deflect and randomize the impact point sufficiently to make the system inaccurate and useless. If a satellite has a nonmetal absorptive skin in the X-band, minimizing the radar echo cross section, such a skin would ideally absorb most of the burst energy, ablating from 100 gms/burst at a 100 mile range to 10 Kg/burst at a 10 mile range. This may reduce the shock strength into the 10 Kilobar range, but would cause sufficient ablative damage and orbital error to cause a large random miss-distance which could be from 2 to 10 miles for a one ton object.

Normally power is required only to operate low energy radar search and acquisition and to run the vacuum pumps to maintain the main resonator liquid helium bath temperature and void gas dump dewar temperature. During an emergency such as an I.C.B.M. attack, about 2×10^8 watts per site is required. This requirement may only last, at a given site, for a few minutes. The power is used to fill the resonator from beam power tubes. The resonator with niobium walls at liquid helium temperatures has "Q's" of the order of 1×10^{11}. Replacement of energy is required only once every Q/ω seconds (1 to 2 seconds). Helium is not cooled against the skin losses directly, but rather exposed to a low pressure, large volume, precooled gas dump chamber which is pumped down to a low temperature over a longer period of time. The gas dump chamber operates for a few minutes with no external power required. The dump chamber preferably is a large racetrack shaped ring buried around the resonator, with a minor radius of the order of 30 feet and a major diameter of 1000 feet. Such system may operate for about two minutes with no rise in the bath temperature of the resonator. In such period of operation about 1×10^6 liters of liquid helium bath would boil off.

In an emergency, resources of a major power grid would be diverted to sites at which this invention is installed. Some sites with magneto hydrodynamic generators capable of independent operation might also be used, but the diversion or use as the case may be would be of short duration.

Referring now to FIG. 2, the sphere **20** has been shown in greater detail and comprises an inner liner or wall **20***a* and an intermediate liner **20***b*. The zone **20***c* between liners **20***a* and **20***b* is a metallic honeycomb filled zone through which liquid helium is caused to flow to maintain the inner liner **20***a* at about the temperature of liquid helium, thus making it superconductive. Preferably, the inner liner **20***a* is a suitable metallic shell having a niobium or niobium alloy film thereon which may be made superconductive.

The outer shell **20***d* encompasses the intermediate shell **20***b* with a zone **20***e* therebetween which similarly has liquid hydrogen therein. Pipes **20***f* and **20***g* lead through the outer liner **20***d* to a gas dump, later to be described. More particularly, a plurality of pipes **20***f* and **20***g* extend radially from sphere **20** and are provided with valves **20***h* and **20***i*.

The sphere **20** is loaded on command with microwave energy from a plurality of generator sets such as the sets **30** and **31**. The generator set **30** is coupled to the sphere by way of an input channel **32**. The generator set **31** is connected into the sphere **20** by way of a channel **33**. The microwave generators **30** and **31** may comprise multiple units such as klystron tubes which are maintained at cryogenic temperatures. As illustrated in FIG. 2, several tubes are to be included in unit **31**, only one being shown. The tubes are identical and each includes a cathode **31***a* in a first cavity which leads to a buncher cavity **31***b* which in turn is connected to an output cavity **31***c*. The beam then passes through a magnetic deflecting unit **31***d* to a dump anode **31***e*. The output cavity **31***c* is connected by way of channel **31***f* to the input channel **33** leading to the sphere **20**. A grid is maintained at a very high potential, of the order of 200,000 volts, for operation at high power levels as will be described wherein a plurality of like tubes in each unit **30** and **31** are operated in parallel.

Generators **30** and **31** operate under the control of the radar control unit **35**. The control unit **35** controls the phase of microwaves in the various channels in unit **36** so that the beam from the radiating elements in antenna **10** can be focused or directed to any given direction above the plane defined by the phased array antenna elements. Targets are identified by the control unit **35** operating at low power supplied via channel **33***a*. In response to detecting a target a timer **37** is actuated by way of channel **38** to turn generators **30** and **31** on to high power thereby initiating high level flow of microwave energy into sphere **20**. Timer **37** also applies periodically a trigger pulse by way of channel **40** to each of a plurality of lasers. Only one laser **25** has been shown producing a laser beam which momentarily heats windows in the walls of the sphere **20** in order to dump the microwave energy stored therein into the wave guides **41** which lead to the phase controlled unit **36**. As indicated in FIG. 2, the laser beam **25***a* passes through while being partially reflected by a mirror **42**. The beam thus reflected is again reflected by a mirror **43** to impinge a window **44** in the wall of sphere **20**. The portion of the beam passing through mirror **42** illuminates and heats the skin on window **45**. The other windows similarly are simultaneously heated. The laser beams together with the structure of the window **44**, **45** etc. operate as dump switches. As indicated in FIG. 2, there are many dump switches located around the wall of the sphere **20**, only eight such window dump switches being shown in FIG. 2. The dump switches lead to the wave guide **41** so that the energy may be transmitted through the phase control unit to the various elements in the antenna **10** to radiate a high energy burst into the atmosphere and to the target to which the array **10** is directed.

In FIG. **3** the structure of the window **44** has been shown in section. A wave guide **44***a* passes through the wall of sphere **20** to the layer **20***a* which is a superconducting skin lining sphere **20**. A quartz plate **44***b* is mounted in the wave guide **44***a* immediately behind the layer **20***a*. The quartz layer is thus transmissive as to the laser beam **25***a* which is directed axially along the wave guide **44***a*. The quartz is a high dielectric material but capable of supporting the film **20***a* and capable of passing the microwaves stored in sphere **20** when film **20***a* on the face of the quartz plate **44***b* is heated by the laser beam **25***a* to a temperature above the superconducting temperature.

A modified form of the window is shown in FIG. **4**. In this modification, the film **20***a* is backed by a first quartz plate **44***b*. A second quartz plate **44***c* is positioned behind plate **44***b* and is spaced therefrom so that a flow channel **44***d* is created between the confronting surfaces of the quartz plates **44***b* and **44***c*. The wave guide **44***a* has perforations **44***e* therein registering with the zone between plates **44***b* and **44***c*. By this means, liquid helium may flow between the two quartz plates from the zone **20***c* lying between the liners **20***a* and **20***b* of FIG. **2**.

In FIG. **5** the system has been illustrated in a generalized form wherein the phased array antenna **10** is supported above the sphere **20**. The sphere **20** is provided with an outer cover or shell **60** within which all of the components of the radar may be mounted except that a base portion **61** will be provided to house the microwave generators and associated electronics. The entire volume within housing **60** and within the unit **61** will be maintained preferably at cryogenic temperature.

A reservoir of liquified helium is maintained available for heat transfer from the zone encompassed by shell **60**. More particularly, a plurality of radially extending tubes **70** and **71** lead to a chamber **72** which may be the racetrack shaped chamber above referred to. Valves **73** and **74** may be opened so that heat may be transmitted from components within the shell **60** to the reservoir of cooled helium in the chamber **72**. The liquid helium in chamber **72** would be maintained cool by conventional liquifying apparatus contained in the units **75** and **76**. Normally when the system is operating in the search and acquisition mode only, valves **73** and **74** will be closed. The system in units **75** and **76** would be maintained on an operative basis to keep the helium in chamber **72** liquified. When high power bursts are to be generated, the valves **73** and **74** are opened to make available for cooling the vast reservoir of helium in chamber **72**.

By way of example, the dimension (x) of the sphere **20** may be of the order of 60 feet. The maximum diameter (y) of the chamber **20** may be of the order of 1000 feet. The system would accommodate about two million liters of helium for cooling purposes.

Before a radar burst is triggered, the field level inside the resonator **20** is brought up to full power, requiring about 2×10^8 watts using beam power tubes as microwave source for excitation. This requires approximately one second.

The laser trigger pulse is applied via the wave guides through a quartz window **44***f*, FIG. **2**, located on each guide into the intended focal plane of each dump window. The focal plane in FIG. **3** is within the quartz window **44***b* in resonator **20** just behind the thin niobium film **20***a* so that very fast local heating is applied to the film via conduction in the quartz. Low temperature quartz conducts heat away from the thin niobium film caused by the passage of a high power microwave pulse during the dump operation. The thermal diffusivity of quartz below 5° K. is less than or about the same as a metal. Quartz makes an ideal dielectric backing and heat sink for low temperature applications. When the superconductor film **20***a* is operating near 2° K., a 200 A° thickness is sufficient to completely reflect the field and establish the resonator modes. When the temperature is raised locally about 5° to 10° K., the local skin depth in the film **20***a* increases and the material switches into the normal conducting state. In the normal state, the film represents only about 1% of a skin depth. The radiation escapes through the ports very rapidly, being coupled and impedance matched to the traveling wave guide mode at the same frequency. During the approximately one microsecond when the power is flowing through the film, the current density in the metal in the normal state is,

$$\bar{J}=\sigma E_w \alpha Q^2 E_w \tag{1}$$

where:

E_w=traveling wave amplitude;

σ=conductivity; and

Q=the Q of the liner **20***a*.

For a traveling mode to transmit the power from the cavity in time Δt requires a field E given by

$$\Delta t \frac{E^2}{8\pi} C = \frac{1\times10^8}{NS} \tag{2}$$

where N is the number of windows in area S each, C the speed of light. This means for NS approximately equal to one half the sperical resonator surface, and $\Delta t\approx1$ μsec. that $E_{peak}=5\times10^4$ volt/cm. After transmission to the high pressure horns in antenna **10**, this field level E_{peak} is finally reduced by increasing the radiating area to about ten times the resonator area to a value of about 1×10^6 volt/meter. Dealing with low temperature quartz behavior, the film **20***a* is thermally coupled to a diffusion depth approximately equal to 0.05 cm for 1 μsec. in the quartz. This provides a mass of about ten grams for a one microsecond pulse, and area approximately equal to 100 cm² per window. For a temperature rise not to exceed T=5° K. in the quartz to this depth, this requires no more than $\Delta E=M_Q C_Q \Delta T$ (where M_Q is the mass of the quartz, and C_Q is the specific heat of the quartz) joules be deposited in the film by the pulse. The pulse will carry 1×10^5 joules of electromagnetic energy per window. From the above relations, only about 200 joules can be accommodated and still use the low temperature quartz coupling. But this is the amount expected for the energy deposit, as the following shows: For the normal metal condition the cavity "Q" is known to drop to about 1×10^5 at 10° K. This means all the cavity energy would be absorbed in a time given by $Q/\omega=1\times10^{-5}$ sec. For 1% of a mean free path in the film (1% of a skin depth) only about 1×10^4 watts/cm² is deposited or about a few joules per microsecond per window. This is well within the capacity of the available heat sink. Note that if this energy were deposited only within the metal skin of area 100 cm², the temperature rise would be at least a few thousand degrees K., and would probably ablate the film.

In a phased array system, the microwave beam can be made very directional and approach a theoretical paraboloid reflector surface. The approximate angular width of the ½ energy points can be estimated from the relation for a paraboloid:

$$\theta=(7\times10^{4})/fD \qquad (3)$$

where

θ=beam angle in degrees;

f=frequency in megahertz; and

D=diameter of phased array (paraboloid) feet.

Hence, for a "baseline" or equivalent paraboloid diameter of 500 feet and f=10,000 Hz (3 cm. radiation), $\theta=2\times10^{-4}$ radian. At a range of 100 miles (500,000 feet) the main energy lobe covers about 100 feet. If D is measured in cm., f in c.p.s., and the angle θ in radians, $\theta=(3\times10^{10})/fD$.

A target having linear dimension, L, such that $L>>\lambda$, has a radar intercept cross section approximately equal to its geometrical projected area. Hence, if R=range in cm and δ=fraction of total pulse intercepted, $\delta=(\pi/2)(L/R)^2(D/\lambda)^2$. For example, if $L\approx5$ feet, R=500,000 feet, $(L/R)^2=1\times10^{-10}$, $(D/\lambda)=0.25\times10^{8}$, $\delta=\pi\times10^{-10}0.25\times10^{8}=0.005$ or ½%. At 33 miles, this is 4% and at 10 miles≈40%. A target moving at 5 mi./second would receive from 30-40 pulses in 100 miles at a rate of 2/second. The average energy is that intercepted at 33 miles in this case, being the average over range.

$$\langle\delta\rangle_R=\frac{\pi}{2}(D/\lambda)^2(L)^2\frac{1}{(R_{max}-R_{min})}\int_{R_{min}}^{R_{max}}\frac{dR}{R^2} \qquad (4)$$

Assuming $R_{max}\approx100$ miles, $R_{min}\approx10$ miles, a target either has (1) generated within it≈30, 200 kilobar shocks or (2) about 30, 20 kilobar shocks, plus an ablation of about 50 Kg of material causing a consequent change in momentum of about 1-10% and so a randomization of from one to ten miles in circular error probable. This depends upon whether the target (1) reflects part of the radiation with metallic like skin depths or (2) is very absorptive at X-band. In either case, components inside the target may be deformed and rendered inoperative. Tracking the target is done via phase shifting electronically in known manner as generally shown in U.S. Pat. Nos. 3,460,149 to Logan and 3,454,945 to Hyltin.

The target will not be able to manuever to avoid the burst since it travels at the speed of light. A target traveling at 5 miles a second could move only ten feet during a maximum range 100 mile transit time for the burst. It will be located near the center of the central lobe. It would move only 0.03 feet even for a one microsecond pulse duration.

For a 1×10^{8} joule system, and a one microsecond pulse, 1×10^{14} watts is radiated. At a field level of 1.5×10^{6} volt/meter, about 6×10^{9} watts/M^2 can be radiated. This requires a total radiating area of about 2×10^{4} M^2. Hence, a 100 meter square array will prevent atmospheric breakdown.

The system including cavity **20** then is a large honeycomb structure lined with a superconductor which is spherical and excited to a field level at the wall of about 1×10^{6} volt/cm. to avoid magnetic breakdown. This requires a radius of about ten meters to store 1×10^{8} joules, thus a 60 foot diameter sphere. The sphere is cooled by superfluid liquid helium in the honeycomb and surrounding bath. With little power this helium bath can be maintained against heat leak to the atmosphere via a simple dewar wall design.

During operation, a helium gas dump chamber, surrounding the cold liquid helium reservoir, is used to pump the helium vapor during an emergency; the void dewar is recycled and maintained at about 1.7° K. by ordinary vacuum pumps when not in emergency use, although over a much longer time. Coupling from the sphere **20** to the radiation impedance is accomplished by first vacuum wave guide transmission to each of vacuum side horn windows. From there the horns are high pressure inert gas (e.g. argon) coupled between the high vacuum side and the atmospheric window side. About 1000 horns with an exit area≈10 M^2 are employed in antenna **10**. They are part of the phased array radiation system and are preferably fed via about 100 dump switch areas on sphere **20**.

In conventional klystron tubes the pulse length obtainable is determined primarily by the electron beam dump anode. For example, tube **31** of FIG. **2** may be operated conventionally to generate a 10 megawatt pulse for a few microseconds by a 100 amp, 200,000 volt beam at about 50% efficiency at $\lambda\approx3$ cm (X-band).

The beam is generated at cathode **31***a*, accelerated by the 200,000 volt potential on grid **31***g* and sent through the buncher cavity **31***b*. The bunched beam is then drifted to maximum bunching time and then either put back through the same cavity (reflex klystron) or through an output cavity **31***c*. After this stage, high energy microwaves are extracted. The beam is still at 100 amperes at about 100,000 volts. This is a 10 megawatt beam. This beam is absorbed onto an anode. The anode can only absorb this energy for a short pulse time without over heating and destruction. Hence pulses of only a few microseconds have been possible.

In accordance with this invention means are provided to eliminate this drawback. Microwaves at 10 megawatts are generated, continuous wave. Some small skin losses in the cavity walls must be taken into account, but this is a relatively minor problem.

In this invention large cryogenic capacity is utilized. This is accomplished by employing two features:

(1) use of superconducting output cavities to eliminate wall skin losses, and

(2) allowing the electron beam to pass out from the output cavity into a vacuum region pipe through a port.

The beam is then passed through another resonant output cavity system (klystron) **31***f* and in like manner used over again several times until most of the beam energy is used up. It then is diverted through a superconducting magnet beam diverter and dumped onto a large absorbing and if necessary ablative anode. Alternately, the beams are decelerated through a D.C. potential to deliver residual power back to the power supply system.

With known techniques (100 amp, 200,000 volts), a continuously operating electron beam system of klystrons operates continuous wave.

In order to power up the main resonator with energy 1×10^{8} joules/sec=100 microwave, about 10 klystron power tubes are required. The unit may run continuously. It can be multiple stage and then diverted to the decelerating electrode **31***e*. Sphere **20** is fed from a continuous electron beam power tube system which may employ 2 to 5 tubes per beam and 2 to 5 beams.

Having described the invention in connection with certain specific embodiments thereof, it is to be understood that further modifications may now suggest themselves to those skilled in the art and it is intended to cover such modifications as fall within the scope of the appended claims.

What is claimed is:

1. A high energy microwave system for directing energy bursts of microwave radiation at a target comprising:
   - (a) a microwave cavity having superconducting walls for storing microwave energy,
   - (b) generating means for feeding waves of microwave energy into said cavity,
   - (c) transmitting means coupled to said cavity for directing microwave energy toward said target, and
   - (d) tracking means connected to said transmitting means for initiating flow of microwave energy to said cavity upon locating said target and triggering said transmitting means a predetermined time after initiating said flow to said cavity for directing a burst of electromagnetic energy to said target.

2. A defense system comprising:
   - (a) a phased array radar adapted to intercept and follow a target path including an antenna array and a low level channel of microwave energy for exciting the elements of said array to produce radar returns and define said path,
   - (b) a microwave storage reservoir,
   - (c) means to supply upon interception of said target a high level flow of microwave energy to said reservoir,
   - (d) microwave feed channels leading from said reservior to the elements of said antenna array,
   - (e) microwave dump switches at the juncture of said channels with said reservoir, and
   - (f) means responsive at least in part to said returns for actuating said dump switches for flow of bursts of microwave energy from said reservoir to said target via said antenna array.

3. The combination set forth in claim 2 wherein said reservoir is maintained at a temperature near 0° K. to prevent losses in the walls thereof.

4. The combination set forth in claim 2 wherein said dump switches comprises:
   - (a) wave guides extending to the wall of said reservoir,
   - (b) a superconductive layer lining said reservoir and covering each of said wave guides,
   - (c) a heat conductive dielectric member in each said wave guide backing said layer, and
   - (d) means to heat said member locally to change the conductivity of said layer at the mouth of each wave guide.

5. The combination set forth in claim 4 wherein said layer is of a niobium material and said dielectric member is quartz.

6. The combination set forth in claim 4 wherein said means to heat said dielectric member is a laser producing a beam directed onto said member.

7. A method of employing high level microwave energy bursts to do work on a moving target at a distance comprising:
   - (a) locating and tracking said target with low energy microwave radar transmissions,
   - (b) upon locating said target initiating flow of high level microwaves to cryogenically controlled storage, and
   - (c) after reaching a predetermined level of energy in storage abruptly dumping said energy via the same path as said radar transmissions to impact said target by locally heating exit locations in the walls of said storage to trigger dumping of said energy.

8. A defense method comprising:
   - (a) operating a phased array radar to intercept and follow a target path by focusing an antenna array in a low level channel of microwave energy to produce radar returns defining said path,
   - (b) supplying upon interception of a target a high level flow of microwave energy to a cryogenic microwave storage reservoir, and
   - (c) responsive at least in part to said returns, actuating localized zones in the walls of said reservoir to render them transmissive for dumping microwave energy from said reservoir onto said target through said antenna array.

9. A high energy microwave system for directing energy bursts of microwave radiation at a target comprising:
   - (a) a microwave reservoir of the order of 10 meters radius lined with a niobium layer,
   - (b) a liquid helium system to maintain said layer superconducting during storage of microwave energy in said reservoir,
   - (c) generating means for feeding microwave energy into said reservoir,
   - (d) wave guides leading from said reservoir,
   - (e) a phased array radar means coupled to said wave guides for directing microwave energy from said reservoir toward said target,
   - (f) a quartz barrier closure for each of said wave guides contacting said barrier in heat transfer relation, and
   - (g) laser means connected to said phased array radar means for heating said quartz to cause flow of microwave energy from said reservoir in predetermined relation to flow of energy to said reservoir.

* * * * *

US007629918B2

(12) **United States Patent**
Brown et al.

(10) **Patent No.: US 7,629,918 B2**
(45) **Date of Patent: Dec. 8, 2009**

(54) **MULTIFUNCTIONAL RADIO FREQUENCY DIRECTED ENERGY SYSTEM**

(75) Inventors: **Kenneth W. Brown**, Yucaipa, CA (US); **David J. Canich**, Upland, CA (US); **Russell F. Berg**, Upland, CA (US)

(73) Assignee: **Raytheon Company**, Waltham, MA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 300 days.

(21) Appl. No.: **11/300,876**

(22) Filed: **Dec. 15, 2005**

(65) **Prior Publication Data**

US 2007/0139247 A1 Jun. 21, 2007

(51) **Int. Cl.**
G01S 13/66 (2006.01)
G01S 7/38 (2006.01)

(52) **U.S. Cl.** **342/13**; 342/67; 342/74; 342/90; 342/97; 89/1.11

(58) **Field of Classification Search** 244/54–61, 244/3.11–3.3; 342/13, 67, 74, 54, 90, 43, 342/729, 61, 14
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2,703,399	A	*	3/1955	Williams et. al.	342/59
2,709,773	A	*	5/1955	Getting et. al.	318/16
2,745,095	A	*	5/1956	Stoddard	342/58
2,840,808	A	*	6/1958	Woodward	342/163
2,908,870	A	*	10/1959	Hardin et al.	331/87
3,417,700	A	*	12/1968	Furlani	102/209
3,688,313	A	*	8/1972	Kern	342/97
3,714,898	A	*	2/1973	Ziemba	102/215
3,946,233	A	*	3/1976	Erben et al.	250/347
4,004,487	A	*	1/1977	Eichweber	89/1.815
4,035,799	A	*	7/1977	Hsiao	342/162
4,109,883	A	*	8/1978	Korr et al.	244/3.1
4,119,967	A	*	10/1978	Johnson et al.	342/365
4,130,821	A	*	12/1978	Goldie	342/198
4,214,534	A	*	7/1980	Richter et al.	102/215
4,319,242	A	*	3/1982	Lewis	342/67
4,360,812	A	*	11/1982	Peperone	342/128
4,456,912	A	*	6/1984	Ensley	342/13
4,661,819	A	*	4/1987	Lewis	342/201

(Continued)

OTHER PUBLICATIONS

International Search Report for corresponding Application No. PCT/US2007/001831 dated Oct. 29, 2007.

Primary Examiner—Thomas H Tarcza
Assistant Examiner—Matthew M Barker
(74) *Attorney, Agent, or Firm*—Renner, Otto, Boisselle & Sklar, LLP

(57) **ABSTRACT**

An RFDE system includes an RFDE transmitter and at least one RFDE antenna. The RFDE transmitter and antenna direct high power electromagnetic energy towards a target sufficient to cause high energy damage or disruption of the target. The RFDE system further includes a targeting system for locating the target. The targeting system includes a radar transmitter and at least one radar antenna for transmitting and receiving electromagnetic energy to locate the target. The RFDE system also includes an antenna pointing system for aiming the at least one RFDE antenna at the target based on the location of the target as ascertained by the targeting system. Moreover, at least a portion of the radar transmitter or the at least one radar antenna is integrated within at least a portion of the RFDE transmitter or the at least one RFDE antenna.

22 Claims, 8 Drawing Sheets

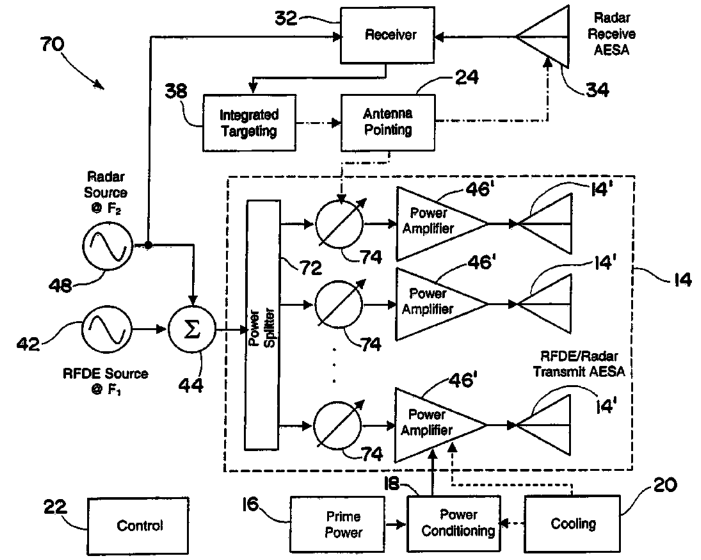

U.S. PATENT DOCUMENTS

4,740,790	A *	4/1988	Hess et al.	342/361
4,890,109	A *	12/1989	Gagliardi	342/14
5,013,151	A *	5/1991	Hughes	356/139.08
5,189,244	A *	2/1993	Weldon	89/8
5,191,343	A *	3/1993	Danzer et al.	342/21
5,387,917	A *	2/1995	Hager et al.	342/68
5,389,931	A *	2/1995	Anderson et al.	342/70
5,432,546	A *	7/1995	Cargill	348/144
5,606,313	A *	2/1997	Allen et al.	340/10.31
5,613,039	A *	3/1997	Wang et al.	706/24
5,726,657	A *	3/1998	Pergande et al.	342/202
5,747,720	A *	5/1998	Schnurr et al.	89/1.11
5,777,572	A *	7/1998	Janusas	342/13
5,780,839	A *	7/1998	Livingston	250/203.2
5,838,262	A *	11/1998	Kershner et al.	340/945
5,977,910	A *	11/1999	Matthews	342/368
6,014,922	A *	1/2000	Livingston	89/1.11
6,265,704	B1 *	7/2001	Livingston	250/203.2
6,347,001	B1	2/2002	Arnold et al.	
6,396,577	B1 *	5/2002	Ramstack	356/141.1
6,424,286	B1 *	7/2002	Smith et al.	342/14
6,545,630	B1 *	4/2003	Apa et al.	342/14
6,731,234	B1 *	5/2004	Hager et al.	342/94
6,832,740	B1 *	12/2004	Ransom	244/3.13
6,859,163	B2 *	2/2005	Alford et al.	342/26 D
6,872,960	B2 *	3/2005	Pepper et al.	250/495.1
6,906,659	B1 *	6/2005	Ramstack	342/54
6,967,614	B1 *	11/2005	Wardell et al.	342/60
7,046,187	B2 *	5/2006	Fullerton et al.	342/54
7,205,934	B2 *	4/2007	Hall	342/175
7,212,148	B1 *	5/2007	Torres	342/13
7,312,744	B1 *	12/2007	Ramstack	342/54
2003/0016421	A1 *	1/2003	Small	359/172
2004/0075884	A1	4/2004	Byren et al.	
2005/0012655	A1 *	1/2005	Lalezari et al.	342/62
2005/0115385	A1	6/2005	Rodriguez	
2005/0115387	A1 *	6/2005	Smith et al.	89/1.11
2006/0022115	A1 *	2/2006	Byren	250/201.9
2006/0082488	A1	4/2006	Keller, III	
2006/0238199	A1 *	10/2006	Larsen	324/329
2007/0040725	A1 *	2/2007	Lowell et al.	342/13
2007/0051233	A1 *	3/2007	Duge	89/1.13

* cited by examiner

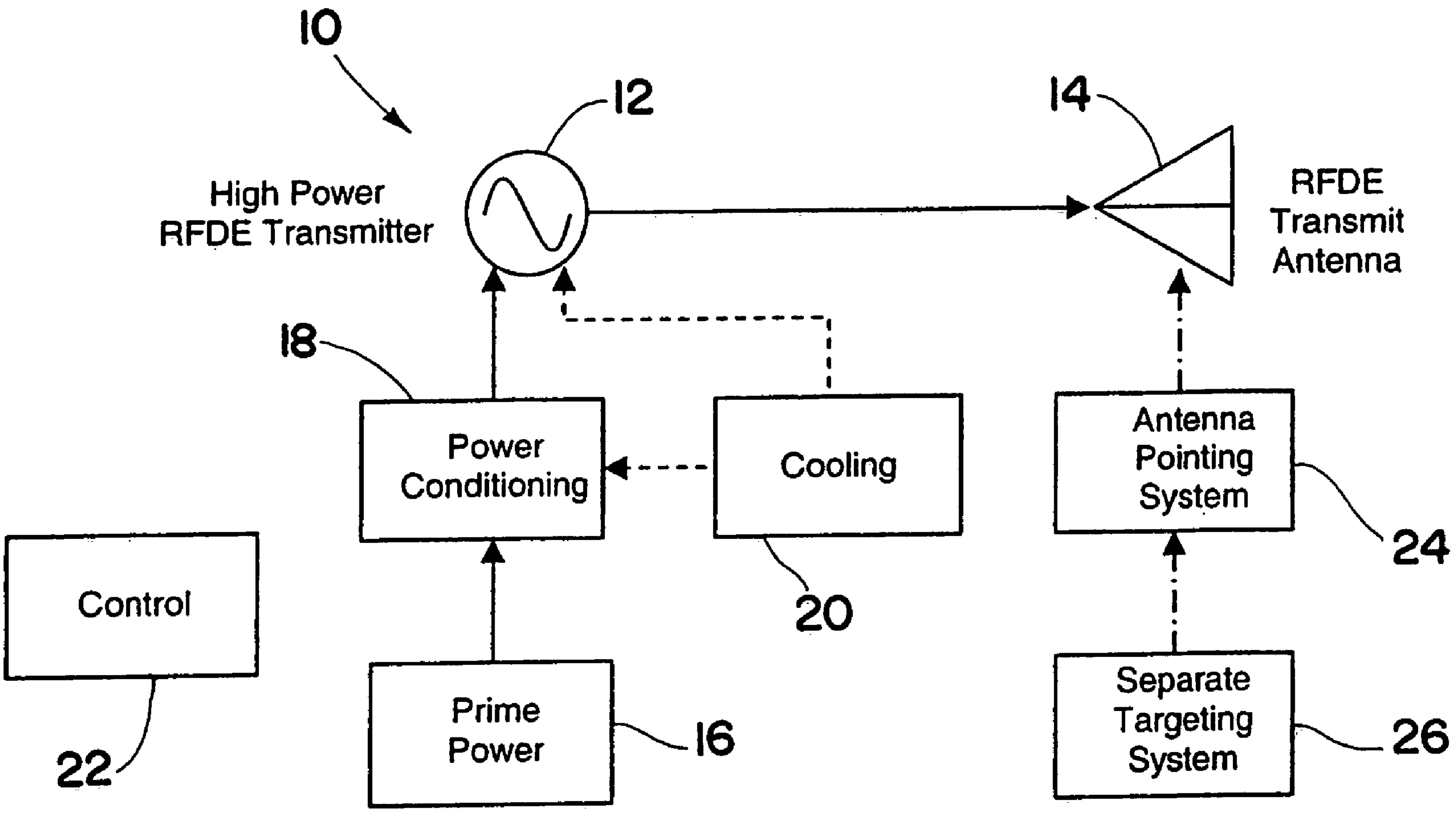
10
12
High Power
RFDE Transmitter
14
RFDE
Transmit
Antenna
18
Power
Conditioning
Cooling
Antenna
Pointing
System
24
Control
20
22
Prime
Power
16
Separate
Targeting
System
26

FIG. 1
Conventional Art

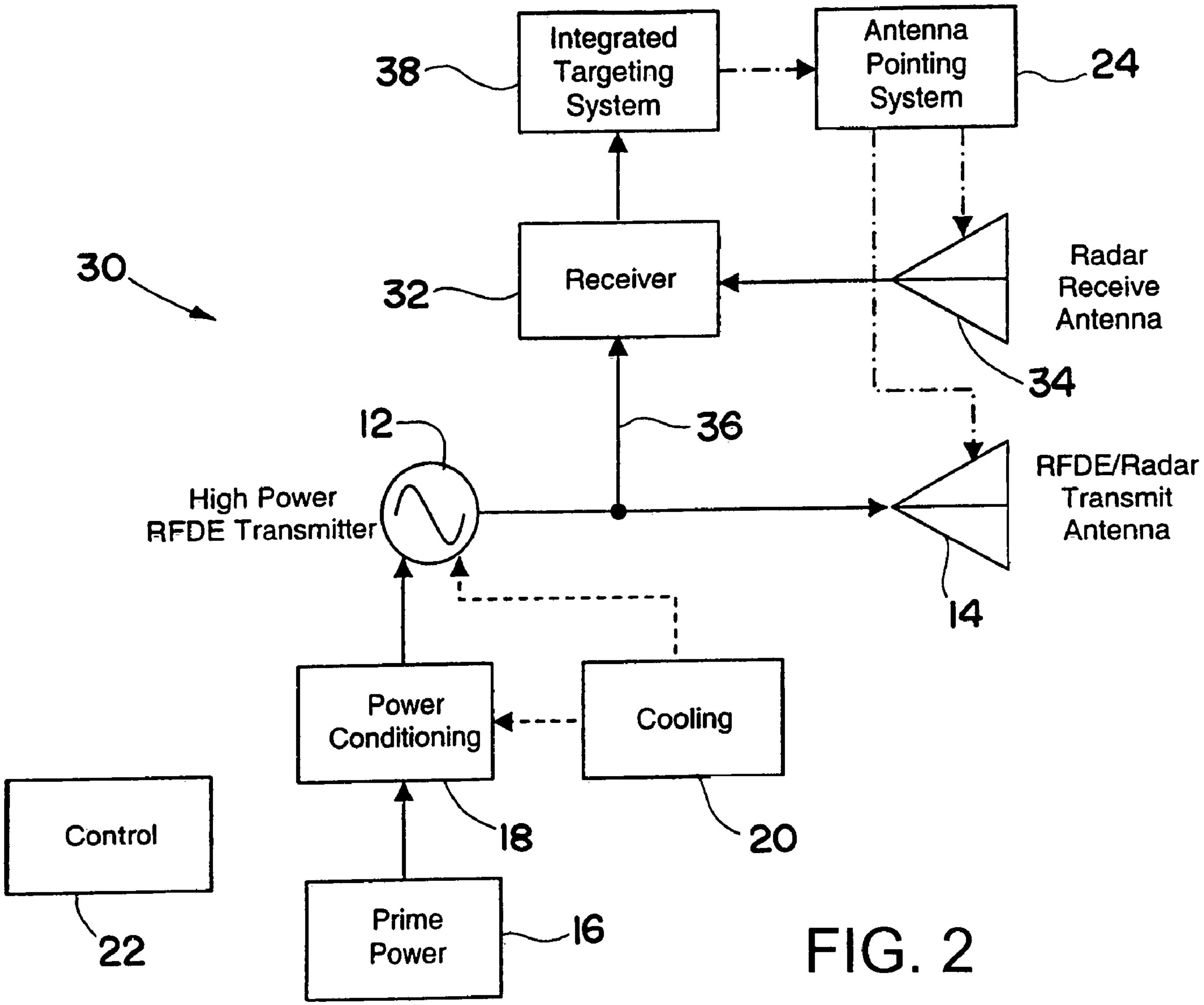
Integrated Targeting System
38
Antenna Pointing System
24
Receiver
32
Radar Receive Antenna
34
30
12
36
High Power RFDE Transmitter
RFDE/Radar Transmit Antenna
14
Power Conditioning
Cooling
Control
18
20
22
Prime Power
16

FIG. 2

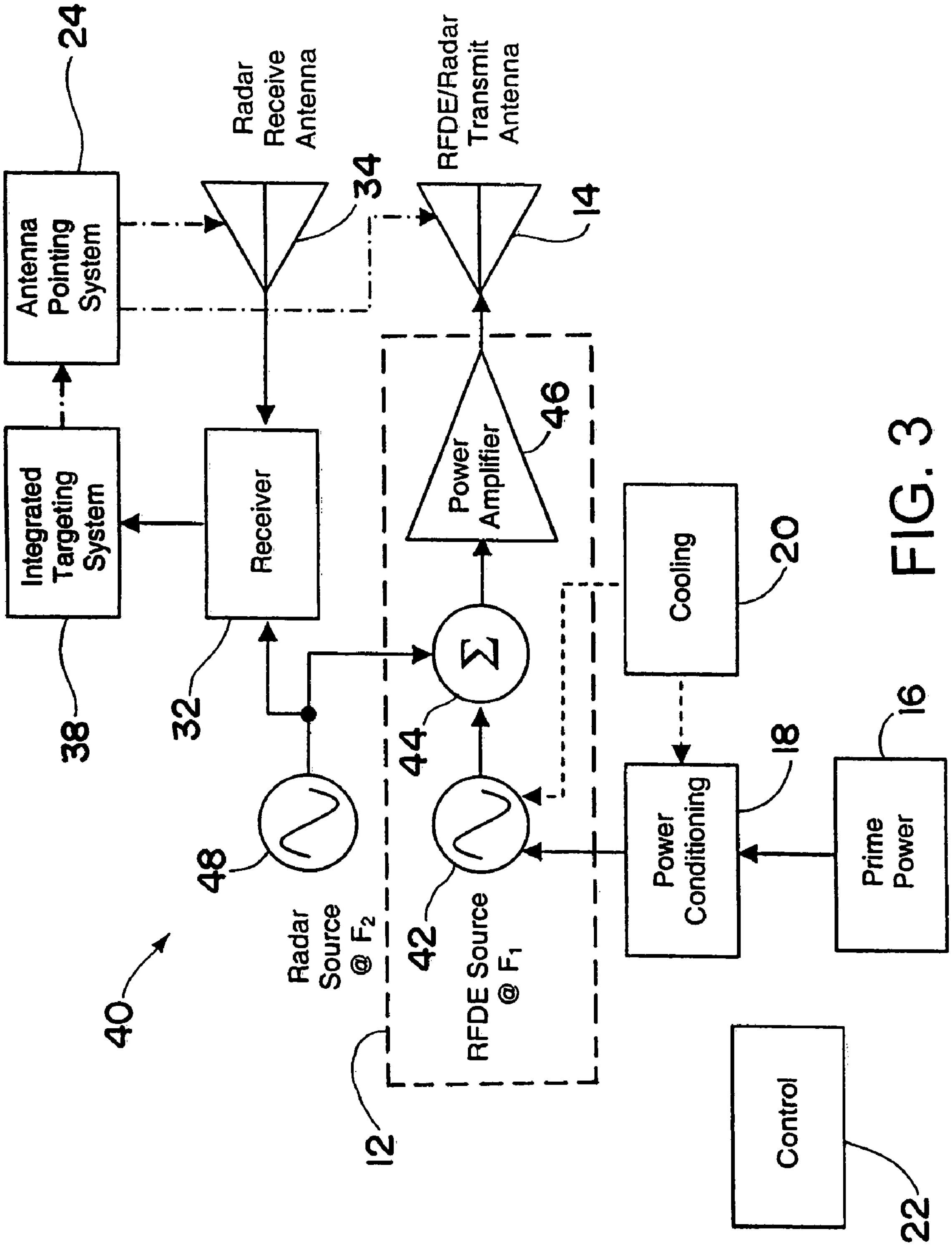
Antenna Pointing System
24
Integrated Targeting System
38
Radar Receive Antenna
34
RFDE/Radar Transmit Antenna
14
Receiver
32
Power Amplifier
46
Σ
44
Radar Source @ F_2
48
RFDE Source @ F_1
42
40
12
Cooling
20
Power Conditioning
18
Prime Power
16
Control
22

FIG. 3

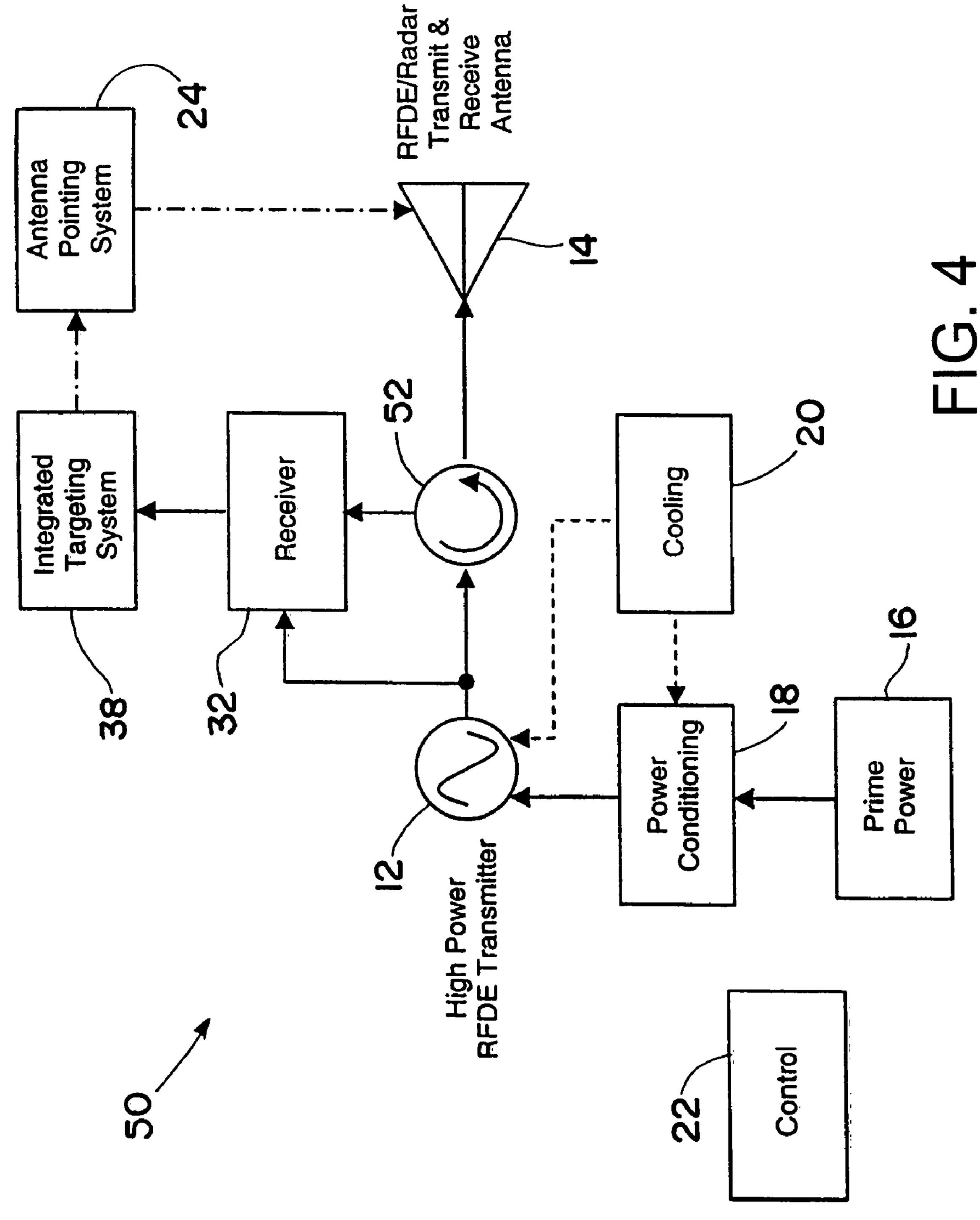
Antenna Pointing System
24
RFDE/Radar Transmit & Receive Antenna
14
Integrated Targeting System
38
Receiver
32
52
Cooling
20
Power Conditioning
18
Prime Power
16
High Power RFDE Transmitter
12
50
Control
22

FIG. 4

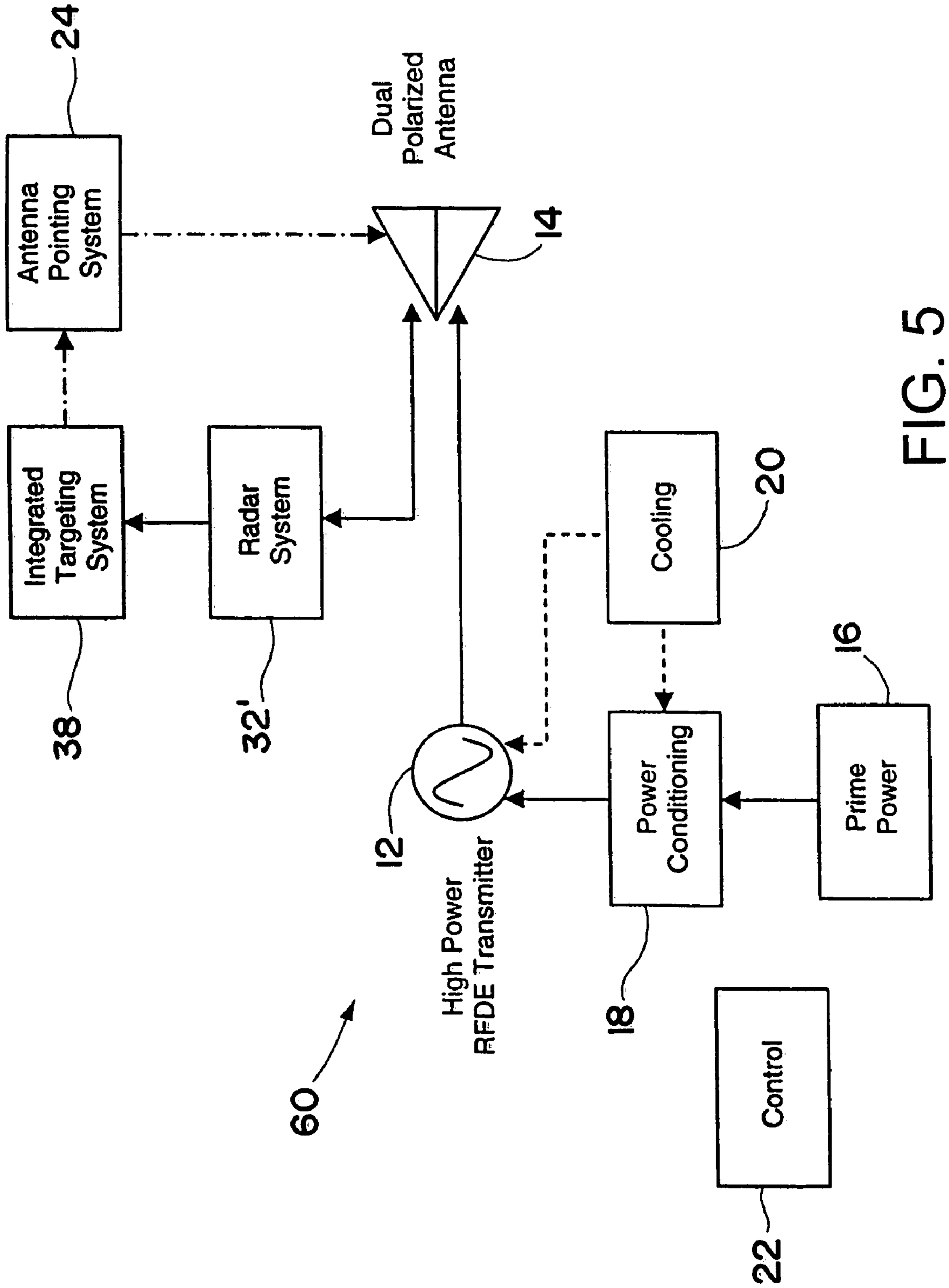
Antenna Pointing System
24
Dual Polarized Antenna
14
Integrated Targeting System
38
Radar System
32'
Cooling
20
High Power RFDE Transmitter
12
Power Conditioning
18
Prime Power
16
Control
22
60

FIG. 5

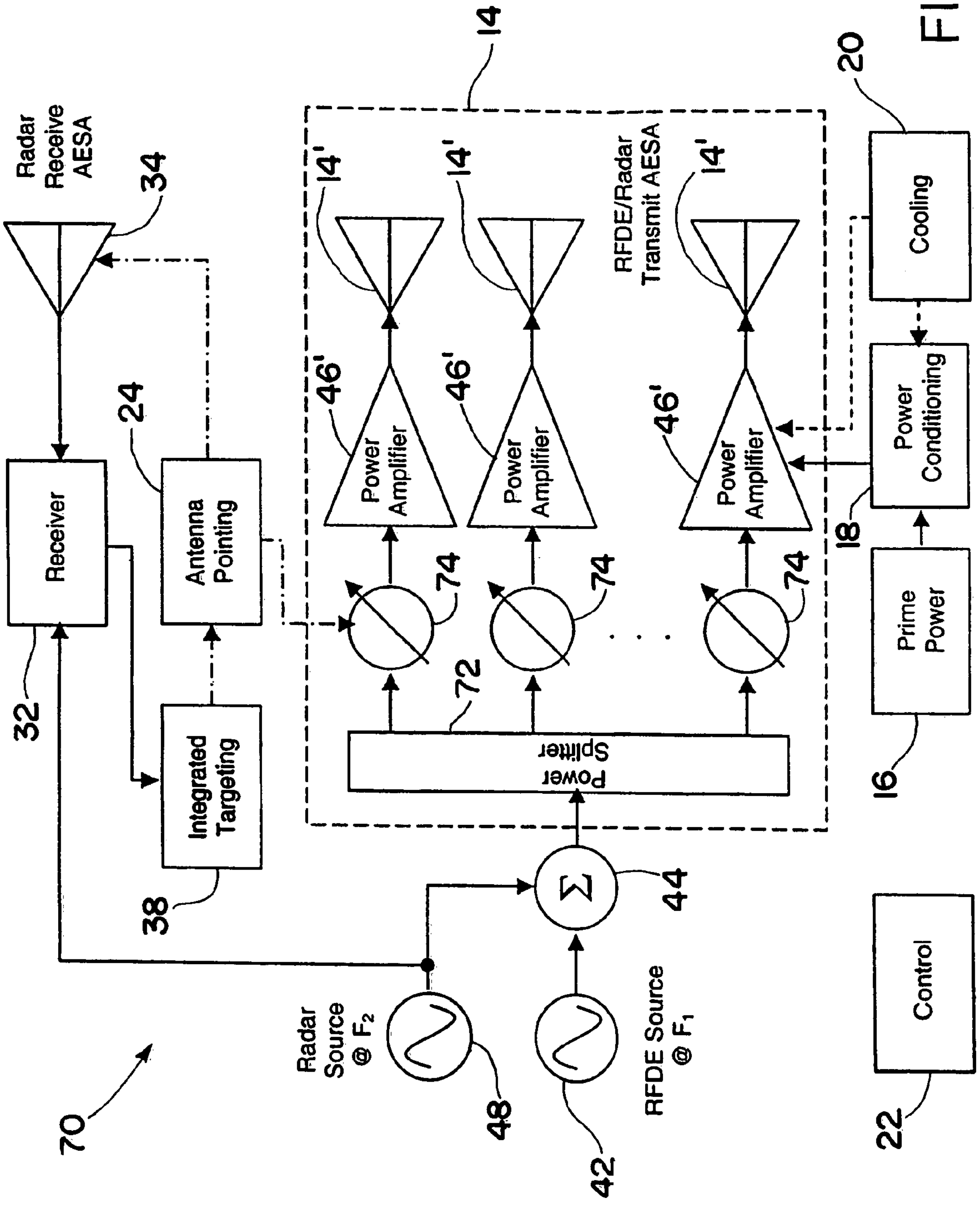
70
Radar Source @ F_2
48
RFDE Source @ F_1
42
Σ
44
32
Receiver
Radar Receive AESA
34
24
Antenna Pointing
Integrated Targeting
38
14
14'
14'
14'
46'
46'
46'
Power Amplifier
Power Amplifier
Power Amplifier
RFDE/Radar Transmit AESA
74
74
74
72
Power Splitter
20
Cooling
Power Conditioning
18
Prime Power
16
Control
22

FIG. 6

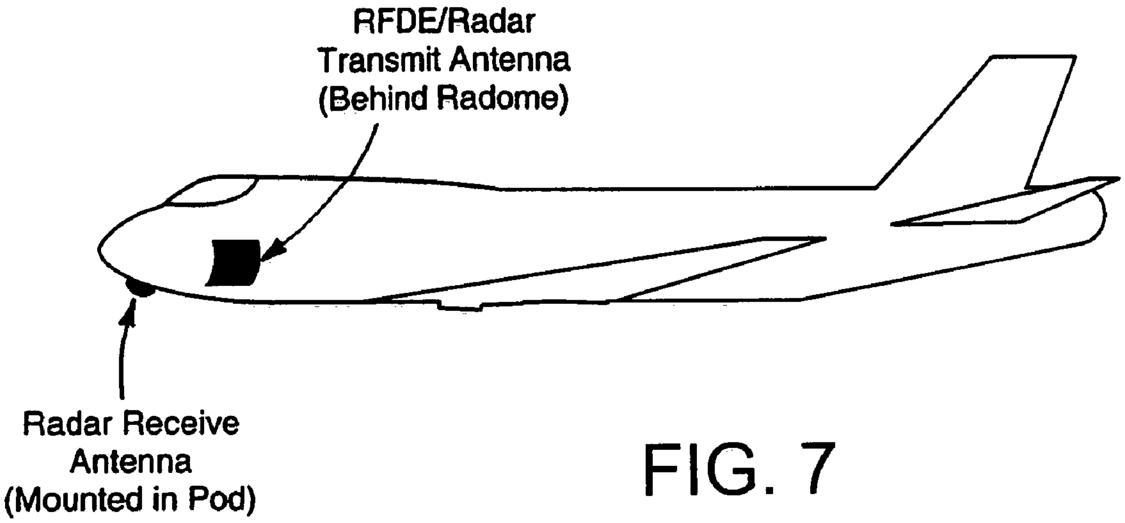
RFDE/Radar
Transmit Antenna
(Behind Radome)
Radar Receive
Antenna
(Mounted in Pod)

FIG. 7

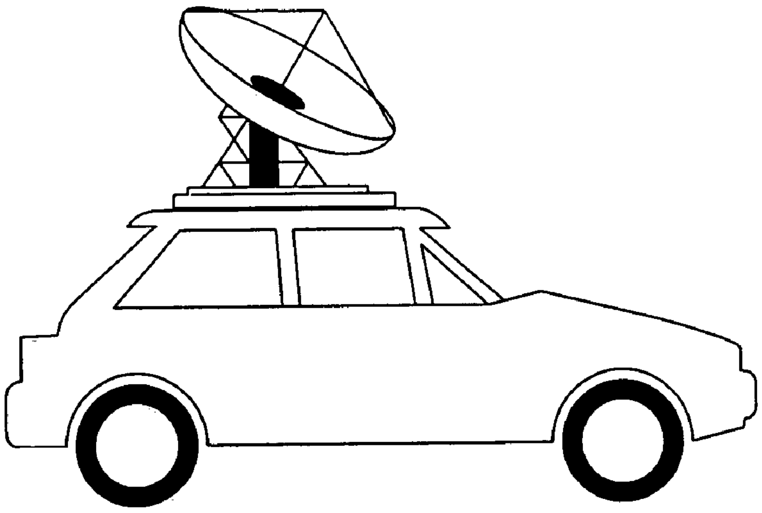

FIG. 8

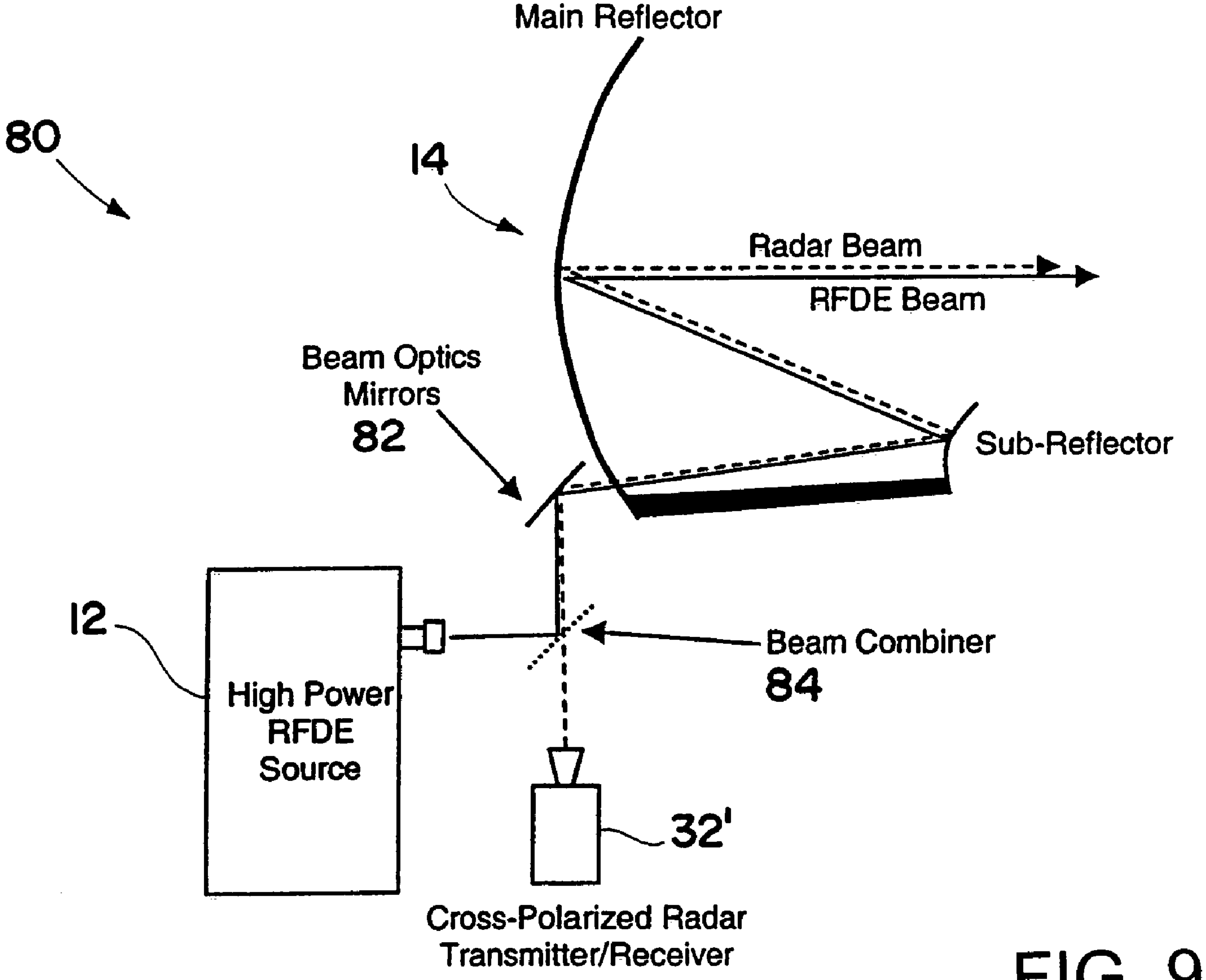
Main Reflector
80
14
Radar Beam
RFDE Beam
Beam Optics
Mirrors
82
Sub-Reflector
12
High Power
RFDE
Source
Beam Combiner
84
32'
Cross-Polarized Radar
Transmitter/Receiver

FIG. 9

MULTIFUNCTIONAL RADIO FREQUENCY DIRECTED ENERGY SYSTEM

TECHNICAL FIELD

The present invention relates generally to radio frequency directed energy (RFDE) systems, and more particularly to multifunctional type RFDE systems.

BACKGROUND OF THE INVENTION

Radio frequency directed energy (RFDE) systems are known in the art for directing high power RF, microwave and/or millimeter wave electromagnetic energy to destroy or disrupt a target. Although RFDE systems typically serve as military weapons, RFDE systems need not be limited to weapon systems. For example, RFDE systems of the present invention may be used for non-military purposes such as destroying or disrupting foreign objects, contaminants, undesirable atmospheric conditions, or other types of targets.

As for weapon systems, it is important to distinguish between an RFDE weapon system and an electronic warfare system. A primary difference between an RFDE weapon and an electronic warfare system is power and kill mode. An electronic warfare system makes use of a priori knowledge of a target it is designed to jam or disrupt. An electronic warfare system uses such a priori knowledge of a target's characteristics (e.g., frequency of operation, method of operation, etc.) to disrupt or confuse the target with "finesse", or a relatively low amount of power.

On the other hand, an RFDE weapon system can go after a broad range of targets (electronics, biological, ordinance, structures, etc.) due to its relatively large radiated power. A priori knowledge of the intended target characteristics is typically not required because the RFDE weapon either burns-out or overwhelms its target by the shear amount of power it radiates.

An ongoing problem with RFDE systems is targeting—accurately pointing the RF directed energy beam at the intended target and establishing an accurate range from the system to the target. To date, the RFDE system targeting problem has been addressed by using what may be referred to as auxiliary add-on systems. These add-on systems could include a stand-alone radar system, a stand-alone laser range finder, stand-alone optical or infrared imaging system, etc. However, these add-on systems add significant cost to the RFDE system. In addition, these add-on systems add significant complexity by requiring calibration of the alignment between the RFDE system and the stand-alone targeting system.

FIG. **1** is a block diagram of a typical RFDE system **10**. In its simplest form, the system **10** includes a high power transmitter **12** transmitting thru a high power antenna **14**. The transmitter **12** operates at RF, microwave or millimeter wave frequencies. The system **10** operates based on a prime power source **16**, such as an AC mains, generator, high capacity battery system, etc. A power conditioning block **18** conditions power delivered from the power source **16** so as to be suitable for powering the transmitter **12**. A cooling system **20** provides appropriate cooling to the power conditioning block **18** and the high power transmitter **12** as needed. A control block **22** provides appropriate control among the various sub-systems.

The RFDE weapon system **10** further includes an antenna pointing system **24** for aiming the high power antenna **14**, and thus the high power electromagnetic energy beam transmitted therefrom, at the target. The pointing system **24** typically is driven by coordinate data identifying the direction and range of the intended target. Such coordinate data is provided by a stand-alone targeting system **26**. As is noted above, the targeting system **26** is an add-on often in the form of a stand-alone radar system, a stand-alone laser range finder, stand-alone optical or infrared imaging system, etc. As is also noted above, however, these add-on systems add significantly to the cost and complexity of the RFDE system.

In view of the aforementioned shortcomings associated with conventional RFDE systems, there is a strong need in the art for an RFDE system which is not subject to the cost and complexity associated with conventional targeting systems.

SUMMARY OF THE INVENTION

The RFDE system of the present invention eliminates the need for a separate, stand-alone targeting system by integrating a targeting system within the RFDE system itself. The RFDE system is multi-functional in that all or part of the RFDE system hardware that functions to direct high power electromagnetic energy also functions to obtain and provide targeting information to aim the high power electromagnetic energy beam. For example, the RFDE transmitter is not only used as the source of the directed electromagnetic energy, but is also used as a radar transmitter for targeting an object. A relatively simple radar receiver may then be added to the RFDE system. The cost of the overall system is substantially reduced since an expensive radar transmitter is not required.

Moreover, the complexity of the system is reduced as calibration of the alignment between the RFDE system and a stand-alone targeting system becomes unnecessary.

According to one aspect of the invention, a multi-functional RFDE system is provided. The RFDE system includes an RFDE transmitter and at least one RFDE antenna. The RFDE transmitter and antenna direct high power electromagnetic energy towards a target sufficient to cause high energy damage or disruption of the target. The RFDE system further includes a targeting system for locating the target. The targeting system includes a radar transmitter and at least one radar antenna for transmitting and receiving electromagnetic energy to locate the target. The RFDE system also includes an antenna pointing system for aiming the at least one RFDE antenna at the target based on the location of the target as ascertained by the targeting system. Moreover, at least a portion of the radar transmitter or the at least one radar antenna is integrated within at least a portion of the RFDE transmitter or the at least one RFDE antenna.

According to another aspect of the invention, a method is provided operating an RFDE system. The method includes the steps of utilizing an RFDE transmitter and at least one RFDE antenna to direct high power electromagnetic energy towards a target sufficient to cause high energy damage or disruption of the target; utilizing a targeting system to locate the target, the targeting system including a radar transmitter and at least one radar antenna for transmitting and receiving electromagnetic energy to locate the target; aiming the at least one RFDE antenna at the target based on the location of the target as ascertained by the targeting system; and integrating at least at portion of the radar transmitter or the at least one radar antenna within, at least a portion of the RFDE transmitter or the at least one RFDE antenna.

To the accomplishment of the foregoing and related; ends, the invention, then, comprises the features hereinafter fully described and particularly pointed out in the claims. The following description and the annexed drawings set forth in detail certain illustrative embodiments of the invention. These embodiments are indicative, however, of but a few of the various ways in which the principles of the invention may be employed. Other objects, advantages and novel features of the invention will become apparent from the following detailed description of the invention when considered in conjunction with the drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a block diagram of a conventional RFDE system;

FIG. **2** is a block diagram of a multi-functional RFDE system in accordance with a first embodiment of the present invention;

FIG. **3** is a block diagram of a multi-functional RFDE system in accordance with a second embodiment of the present invention;

FIG. **4** is a block diagram of a multi-functional RFDE system in accordance with a third embodiment of the present invention;

FIG. **5** is a block diagram of a multi-functional RFDE system in accordance with a fourth embodiment of the present invention;

FIG. **6** is a block diagram of a multi-functional RFDE system in accordance with a fifth embodiment of the present invention;

FIG. **7** illustrates a multi-functional RFDE system mounted within an aircraft in accordance with an embodiment of the present invention;

FIG. **8** illustrates a multi-functional RFDE system mounted within a wheeled vehicle in accordance with an embodiment of the present invention; and

FIG. **9** is a schematic diagram of a multi-functional RFDE system incorporating a reflector-type antenna in accordance with an embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The present invention will now be described with reference to the drawings, in which like reference numerals are provided to refer to like elements throughout.

The RFDE system of the present invention integrates a targeting system, such as a radar targeting, system, into an otherwise conventional RFDE system. There are several ways that the targeting system can be integrated into the RFDE system as explained herein. The particular embodiments described below are meant to be merely exemplary. The present invention contemplates not only the particular embodiments described herein, but any system in which a targeting system is integrated in part or in whole within the RFDE system.

Referring to FIG. **2**, an RFDE system **30** is shown in accordance with an embodiment of the present invention. Since many of the elements of the RFDE system **30** in FIG. **3** are similar to those in the conventional system **10** discussed above with respect to FIG. **1**, only the relevant differences with be discussed herein for sake of brevity. In this particular embodiment, a portion of an otherwise conventional RFDE system (notably the high power transmitter **12** and/or the transmit antenna **14**) forms part of the targeting system. Specifically, during a targeting mode, the high power RFDE transmitter **12** is controlled by the control block **22** to transmit standard radar tracking signals thru the antenna **14**. The radar tracking signals may be any type of conventional radar signal such a pulse or continuous wave radar. The power level of the tracking signals may be a high powered signal, such as the RFDE signal itself, or a relatively low powered signal as more typical in radar tracking applications.

In the embodiment of FIG. **2**, a separate radar receiver **32** and radar receive antenna **34** are used in conjunction with the integrated RFDE/radar transmitter **12**. In one embodiment, the RFDE high power output beam intended to destroy or disturb a target is transmitted using the RFDE/radar transmitter **12** and antenna **14**. The radar receive antenna **34** receives portions of the RFDE output beam reflected by the target back towards the system **30**. The radar receiver **32** processes the reflected return signals using conventional techniques in order to identify the location of the target. For example, the radar receiver **32** can be coherently linked to the transmitted RFDE output waveform (represented by line **36**) so that Doppler processing can be achieved and the direction and range of the target identified.

The radar receiver **32** provides the target location information to an integrated targeting system block **38** which feeds the location information to the antenna pointing system **24**. Such operation allows the antenna **14** to be directed in both search and track radar functions.

As will be appreciated, the RFDE system **30** in FIG. **2**, as with the various other embodiments of the invention described herein, can operate in both an RFDE mode and a tracking mode. Both modes may be carried out simultaneously as described above, for example, where the high power electromagnetic energy output waveform of the RFDE system also serves as the radar tracking system transmit signal. Alternatively, the RFDE system **30** may switch between the RFDE mode and the tracking mode using a separate RFDE high power beam and lower power radar transmit signals, respectively. So long as the system **30** switches between the two modes rapidly enough so as not to lose track of the target, operation between RFDE mode and tracking mode may be time-division multiplexed.

Typically the radar transmitter is one of the most expensive portions of a radar tracking system. Therefore, by using the RFDE transmitter **12** and antenna **14** to function as the radar transmitter and antenna for targeting, the cost of the targeting system can be drastically reduced. Also, using the radar return of the RFDE high power beam itself to determine the target location can substantially improve the beam pointing accuracy of the RFDE system **30**. By using the RFDE high power beam to determine the location of the target, the power density on target will be maximized when standard radar tracking techniques are employed (e.g., monopulse, continuous scan, etc.).

Those of ordinary skill will appreciate that the RFDE transmitter **12** can be any transmitter suitable for transmitting an RFDE high power beam. For example, the RFDE transmitter **12** may be a single or multiple tube source, or solid state source. Moreover, it will be appreciated that the antenna **14** can be any type of suitable high power antenna which can be mechanically and/or electronically pointed and scanned via the antenna pointing system **24**. The transmitter/antenna can also be comprised of an active electronically steered array (AESA), for example, where an array of high power amplifiers/antennas is utilized. The radar receive antennas **34** can be any type of suitable antenna for receiving the radar return signals. As with the antenna **14**, the radar receive antenna **34** is mechanically and/or electronically pointed and scanned via the antenna pointing system **24**. The antenna pointing system **24** can be a mechanical gimbal or a beam steering computer controlling phase shifters in an electronically steerable array.

FIG. **3** illustrates another embodiment representing how a tracking system can be integrated with an RFDE system. The embodiment of FIG. **3**, as with the other embodiments described herein, shares many of the same elements as FIG. **2**, and thus again only the relevant differences between the embodiments will be discussed for sake of brevity.

Specifically, FIG. **3** illustrates an RFDE system **40** in which the power amplifier of the tracking system is integrated within the power amplifier of the RFDE system. More particularly, the RFDE transmitter **12** includes a low power RFDE signal source **42** operating at a first frequency and input to an adder **44**. The output of the adder **44** is input to a power amplifier **46** which amplifies the output before being radiated by the transmit antenna **14**. A low power radar signal source **48** at a second frequency is also input to the adder **44**. In addition, the low power radar signal source **48** is input to the radar receiver **32** to provide for coherent processing. The adder **44** thus outputs the combined RFDE signal source and radar signal source to the power amplifier **46**. The power amplifier **46** can be any suitable type of amplifier including, for example, an injection locked magnetron, a klystron, a solid-state amplifier, etc., or an array of any of these types of amplifiers in an AESA embodiment.

In the embodiment of FIG. **3**, a separate low power radar signal from the signal source **48** is used. This signal is combined with the RFDE signal from the RFDE signal source **42** prior to the combined signal being amplified by the power amplifier **46**. The frequencies of the RFDE signal and the radar signal do not have to be at the same frequency. In fact, they can be completely independent of each other within the bandwidth constraints of the power amplifier **46** and the RFDE transmit antenna **14**. It will be appreciated that significant isolation can be achieved between the RFDE and radar signals by filtering out the RFDE signal at the receive antenna **34** and/or radar receiver **32**.

Again, it will be appreciated that the RFDE system **40** of FIG. **3** may operate in an RFDE mode and a targeting mode. As in the other embodiments described herein, such modes may be carried out simultaneously or in time-multiplexed fashion. In the case where the RFDE signal source **42** and radar signal source **48** are different, one may consider such operation as frequency-multiplexed as will be appreciated.

FIG. **4** illustrates yet another embodiment of the present invention. In this embodiment, the RFDE system **50** integrates the radar receive antenna into the same antenna **14** serving as the RFDE and radar transmit antenna. This is accomplished by means of a high power circulator **52** which routes the RFDE/radar transmit signals from the shared transmitter **12** through to the antenna **14**. Reflected signals received by the antenna are routed by the circulator **52** to the receiver **32** for processing. In this example, as in the others discussed herein, the RFDE transmit signal may also be the radar transmit signal. The return signal is received by the same antenna **14** and coupled through the circulator **52** to the receiver **32**. The radar system thus can provide target information feedback to the antenna pointing system **24**.

The high power circulator **52** can be problematic in that it can be difficult to design a circulator that can handle the typical total power radiated by an RFDE system. Nevertheless, with improvements in materials and technologies such a circulator may someday be commercially feasible. Furthermore, the embodiment of FIG. **4** certainly is very suitable for an AESA system where the output power is broken-up among an array of transmit elements as discussed in more detail below.

In some integrated RFDE/targeting applications it may be desirable only to share the antenna between the RFDE and targeting systems. The RFDE system and the targeting system otherwise operate independently. An example of such an embodiment is shown in FIG. **5**.

Specifically, the embodiment of FIG. **5** illustrates a radar system **32'** which functions essentially independently of the RFDE system **60** with the exception of sharing a dual polarized antenna **14**. The RFDE system transmits the RFDE high power beam from the transmitter **12** via the dual polarized antenna **14** using one polarization (e.g., vertical, right hand circular, etc.). The radar system **32'**, with its own transmitter/receiver, transmits the radar transmit signal via the antenna **14** using the orthogonal polarization (e.g., horizontal, left hand circular, etc.). The reflected radar return signal received by the antenna **14** is processed by the radar system **32'** to provide target location information, again using conventional techniques. As in the other embodiments, the location information is provided to the integrated targeting system **38** which provides the information to the antenna pointing system **24**. The embodiment of FIG. **5** can be used to provide tracking information and/or is especially suitable for providing range information for the RFDE system.

FIG. **6** illustrates an embodiment in which AESAs or phased array antennas may be incorporated within the present invention. The embodiment of FIG. **6** is fundamentally the same as the embodiment of FIG. **3**, with the exception that the power amplifier **46** is represented by an array of power amplifiers **46'** included with the RFDE/radar transmit AESA antenna **14**. The radar receive antenna **34** may similarly comprise an AESA antenna.

As is shown in FIG. **6**, the combined signal from the adder **44** is input to a power splitter **72** within the AESA antenna **14**. The power splitter **72** separates the signal and provides the split signal to respective phase shifters **74** and power amplifiers **46'** corresponding to respective radiator elements **14'** in the antenna **14**. The antenna pointing system **24** may steer the antenna **14** by adjusting the phase of the phase shifters **74**, as will be appreciated.

Continuing to refer to FIG. **6**, it will further be appreciated that the radar source **48** and RFDE source **42** can be at different frequencies and still radiate in the same direction. As long as the phase shifters **74** are selected to provide a true time delay (which is fairly common in the art), the antenna **14** may be steered concurrently for both frequencies.

A multifunctional RFDE system of the present invention can be employed on a variety of platforms. For example, FIG. **7** illustrates an embodiment in which the system is employed on an aircraft. The combined RFDE/radar transmitter **12** and antenna **14** (not shown), for example, are mounted to radiate out the side of the aircraft. A pod is then mounted beneath the aircraft, containing the radar receiver **32** and receive antenna **34** (also not shown).

FIG. **8** shows an embodiment in which the RFDE system of the present invention can be employed on a wheeled vehicle. For example, the RFDE and radar systems are mounted in the back of the vehicle and share a common antenna. One possible such system **80** is shown in FIG. **9**. A high power RFDE source **12** radiates into a beam transport system comprised of mirrors (e.g., **82**) suitable for the frequency of operation. A cross-polarized radar transmit signal from the radar system **32'** is then injected into the RFDE beam path by means of a beam combiner/splitter **84**. The RFDE and radar transmit signal are then simultaneously transmitted from the common antenna **14**, in this embodiment a reflector-type antenna. The radar return signal is received by the antenna **14** and directed back to the cross-polarized radar system **32'** via the mirrors **82** and beam combiner/splitter **84**.

Although the invention has been shown and described with respect to certain preferred embodiments, it is obvious that equivalents and modifications will occur to others skilled in the art upon the reading and understanding of the specification. For example, although the present invention has been described primarily in the context of a conventional radar based targeting system, other types of radar-like targeting systems which rely on the transmission and return of electromagnetic energy (e.g., laser-based, infrared, etc.) can also be used without departing from the scope of the invention.

The present invention includes all such equivalents and modifications, and is limited only by the scope of the following claims.

What is claimed is:

1. A multi-functional radio frequency directed energy (RFDE) system, comprising:

an RFDE transmitter and at least one RFDE antenna for directing high power electromagnetic energy towards a target sufficient to cause high energy damage or disruption of the target;

a targeting system for locating the target, the targeting system including a radar transmitter and at least one radar antenna for transmitting and receiving electromagnetic energy to locate the target; and

an antenna pointing system for aiming the at least one RFDE antenna at the target based on the location of the target as ascertained by the targeting system,

wherein at least a portion of the radar transmitter or the at least one radar antenna is integrated within at least a portion of the RFDE transmitter or the at least one RFDE antenna to provide simultaneous transmission of the high power electromagnetic energy and the electromagnetic energy to locate the target.

2. The multi-functional RFDE system of claim **1**, wherein the at least one radar antenna is embodied at least partially within the at least one RFDE antenna.

3. The multi-functional RFDE system of claim **1**, wherein the radar transmitter is embodied at least partially within the RFDE transmitter.

4. The multi-functional RFDE system of claim **3**, wherein the radar transmitter and the RFDE transmitter comprise a common RF power amplifier.

5. The multi-functional RFDE system of claim **4**, wherein the electromagnetic energy for locating the target is at a first frequency, and the high power electromagnetic energy is at a second frequency different from the first.

6. The multi-functional RFDE system of claim **1**, wherein the at least one radar antenna functions to transmit the electromagnetic energy for locating the target, and the at least one radar antenna is embodied at least partially in the at least one RFDE antenna.

7. The multi-functional RFDE system of claim **6**, wherein the at least one radar antenna includes a first radar antenna that functions to transmit the electromagnetic energy for locating the target and to transmit the high power electromagnetic energy, and a second radar antenna that functions to receive the electromagnetic energy reflected from the target in order to locate the target.

8. The multi-functional RFDE system of claim **7**, wherein the first radar antenna comprises a multi-element phased array.

9. The multi-functional RFDE system of claim **6**, wherein the at least one radar antenna that functions to transmit the electromagnetic energy for locating the target also functions to receive the electromagnetic energy reflected from the target in order to locate the target.

10. The multi-functional RFDE system of claim **9**, wherein the at least one radar antenna comprises a dual-polarized antenna.

11. The multi-functional RFDE system of claim **1**, wherein the system comprises a beam combiner for combining the high power electromagnetic energy with the electromagnetic energy for locating the target in a path between the RFDE transmitter and the RFDE antenna.

12. The multi-functional RFDE system of claim **1**, wherein the system is configured for operation in a mobile vehicle.

13. The multi-functional RFDE system of claim **12**, wherein the mobile vehicle is a wheeled-vehicle.

14. The multi-functional RFDE system of claim **12**, wherein the mobile vehicle is an aircraft.

15. A method of operating a multi-functional radio frequency directed energy (RFDE) system, comprising the steps of:

utilizing an RFDE transmitter and at least one RFDE antenna to direct high power electromagnetic energy towards a target sufficient to cause high energy damage or disruption of the target;

utilizing a targeting system to locate the target, the targeting system including a radar transmitter and at least one radar antenna for transmitting and receiving electromagnetic energy to locate the target;

aiming the at least one RFDE antenna at the target based on the location of the target as ascertained by the targeting system; and

integrating at least a portion of the radar transmitter or the at least one radar antenna within at least a portion of the RFDE transmitter or the at least one RFDE antenna to provide simultaneous transmission of the high power electromagnetic energy and the electromagnetic energy to locate the target.

16. The method of claim **15**, wherein the at least one radar antenna is embodied at least partially within the at least one RFDE antenna.

17. The method of claim **15**, wherein the radar transmitter is embodied at least partially within the RFDE transmitter.

18. The method of claim **17**, wherein the radar transmitter and the RFDE transmitter comprise a common RF power amplifier.

19. The method of claim **18**, wherein the electromagnetic energy for locating the target is at a first frequency, and the high power electromagnetic energy is at a second frequency different from the first.

20. The method of claim **15**, wherein the at least one radar antenna functions to transmit the electromagnetic energy for locating the target, and the at least one radar antenna is embodied at least partially in the at least one RFDE antenna.

21. A multi-functional radio frequency directed energy (RFDE) system, comprising:

an RFDE transmitter and at least one RFDE antenna for directing high power electromagnetic energy towards a target sufficient to cause high energy damage or disruption of the target;

a targeting system for locating the target, the targeting system including a radar transmitter and at least one radar antenna for transmitting and receiving electromagnetic energy to locate the target; and

an antenna pointing system for aiming the at least one RFDE antenna at the target based on the location of the target as ascertained by the targeting system,

wherein at least a portion of the radar transmitter or the at least one radar antenna is integrated within at least a portion of the RFDE transmitter or the at least one RFDE antenna and the high power electromagnetic energy is used as the electromagnetic energy to locate the target.

22. A method of operating a multi-functional radio frequency directed energy (RFDE) system, comprising the steps of:

utilizing an RFDE transmitter and at least one RFDE antenna to direct high power electromagnetic energy towards a target sufficient to cause high energy damage or disruption of the target;

utilizing a targeting system to locate the target, the targeting system including a radar transmitter and at least one radar antenna for transmitting and receiving electromagnetic energy to locate the target;

aiming the at least one RFDE antenna at the target based on the location of the target as ascertained by the targeting system; and

integrating at least a portion of the radar transmitter or the at least one radar antenna within at least a portion of the RFDE transmitter or the at least one RFDE antenna,

wherein the high power electromagnetic energy is used as the electromagnetic energy to locate the target.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 7,629,918 B2
APPLICATION NO. : 11/300876
DATED : December 8, 2009
INVENTOR(S) : Brown et al.

Page 1 of 1

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

On the Title Page:

The first or sole Notice should read --

Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 230 days.

Signed and Sealed this

Second Day of November, 2010

David J. Kappos

David J. Kappos
Director of the United States Patent and Trademark Office

22nd March 2013

DIY - Battery-Free Tiny Crystal Radio

Build a Tiny Crystal Radio that runs without battery.

[http://www.blogger.com/blogger.g?blogID=2311644141825422235]

[http://www.blogger.com/blogger.g?blogID=2311644141825422235]

[http://www.blogger.com/blogger.g?blogID=2311644141825422235]

[http://www.blogger.com/blogger.g?blogID=2311644141825422235]

[http://www.blogger.com/blogger.g?blogID=2311644141825422235]

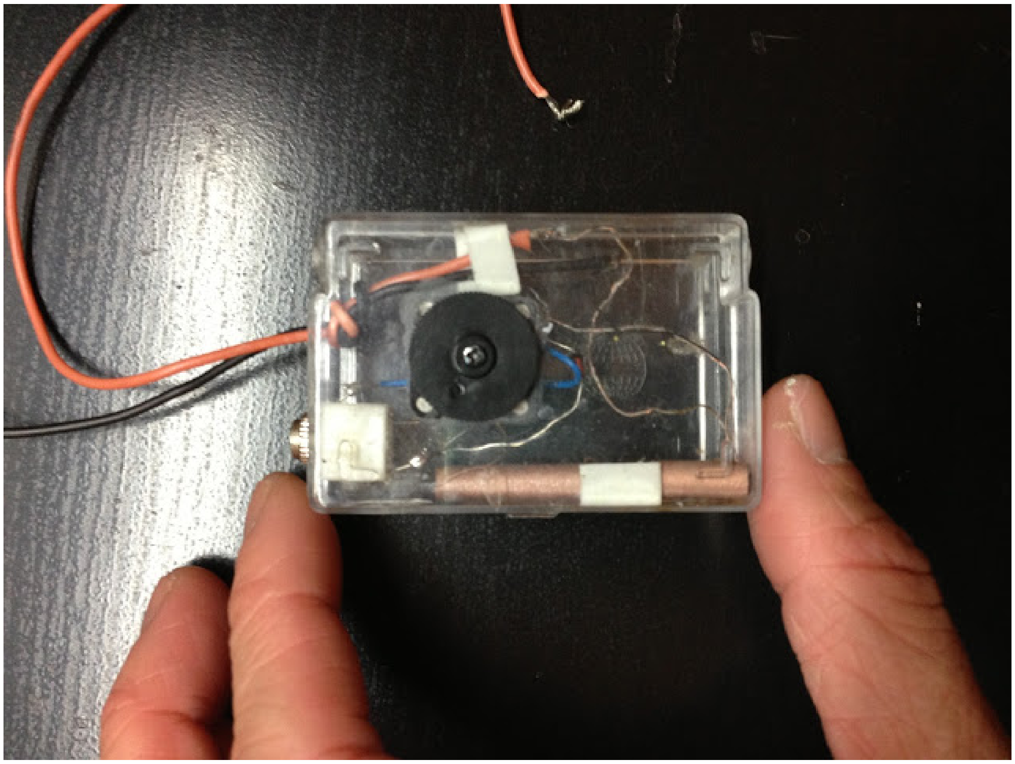

[http://2.bp.blogspot.com/-FYzx5aqFfVM/UVGOaN0HeBI/AAAAAAAAAxQ/XQtWPgoL_VQ/s1600/IMG_1098.JPG]

Parts you'll need

1. Variable Capacitor 360 pf - from an old radio.
2. Coil wound on a Ferrite Powder Rod - from an old radio e.g. old Walkman
3. A Crystal Earphone or a PC speaker with builtin amplifier.
4. Germanium Diode IN60
5. Resistor 100 Kohm
6. 3.5mm earphone jack
7. A small plastic box,
8. some wires .
9. Solder iron.
10. Antenna - 20 meters hanging vertically or horizontally at a high location (Warning : Lightning Hazard)
11. Ground - connect a wire to your metal water pipe.

[http://www.blogger.com/blogger.g?blogID=2311644141825422235]

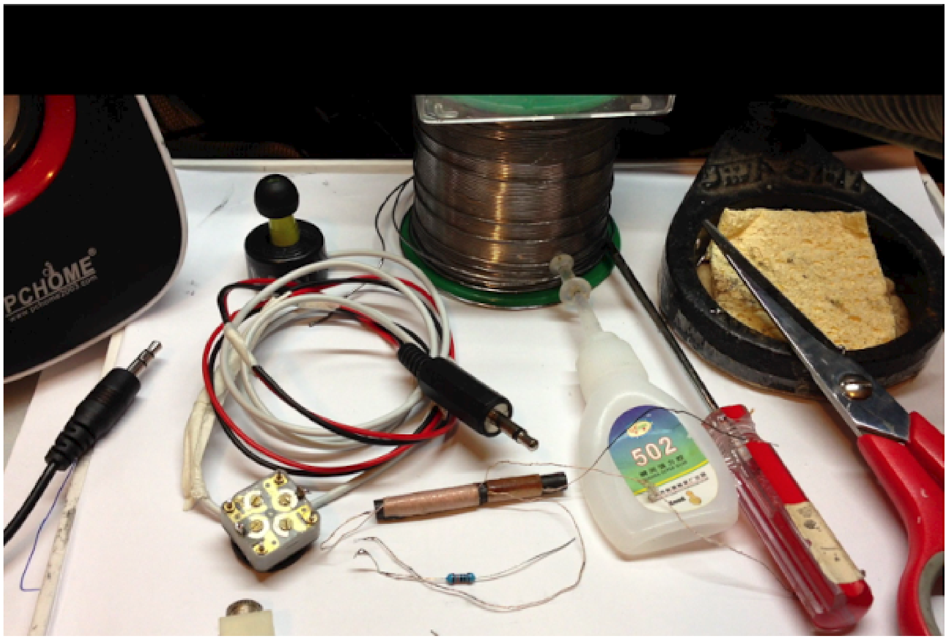

[http://4.bp.blogspot.com/-QqxK3OffVq4/UVLe8yWkFzI/AAAAAAAAAxo/4vtNVM1EUMY/s1600/parts.png]

You can get most of these from electronic ships in Shamshupo, except parts #1,2 & 3.
For the variable capacitor and the coil you can easily buy some old recycled walkman or old radios for HKD20-50 from the hawkers in Ip Liu Street in Sham Shui Po. The other way is to go to taobao.com to buy them online and ship to HK. The goods is RMB 5, but the shipping fees is minimum RMD 30, so not worth it unless you buy in great quantity.
For the Crystal Earphone, it's very hard to buy nowadays. Refer to my other post to make one yourselves. DIY Crystal Earphone for Crystal Radios [http://billydiy.blogspot.com/2013/03/build-crystal-earphone-for-crystal.html]

Remember to leave me some comments (at the comment box) at the bottom of this if you have read this blog.

How to make it

http://www.youtube.com/watch?v=6VsgXuNm13k [http://www.youtube.com/watch?v=6VsgXuNm13k]

Circuit Diagram

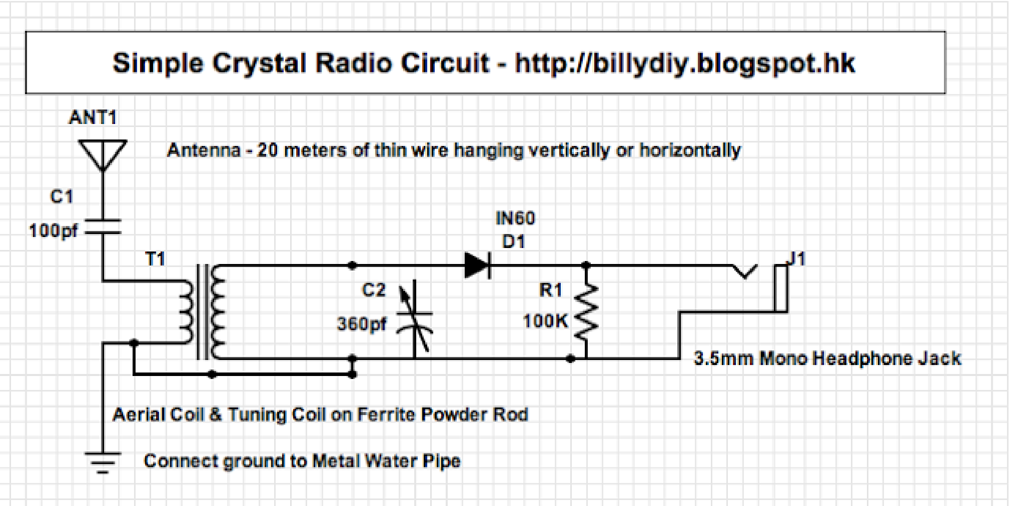
Simple Crystal Radio Circuit - http://billydiy.blogspot.hk
ANT1
Antenna - 20 meters of thin wire hanging vertically or horizontally
C1
100pf
T1
IN60
D1
J1
C2
360pf
R1
100K
3.5mm Mono Headphone Jack
Aerial Coil & Tuning Coil on Ferrite Powder Rod
Connect ground to Metal Water Pipe

[http://1.bp.blogspot.com/-C2qNgc8TvDg/UXFYJSUHmyI/AAAAAAAABDw/KaD3xmUTk-w/s1600/Screen+Shot+2013-04-19+at+10.43.18+PM.png]

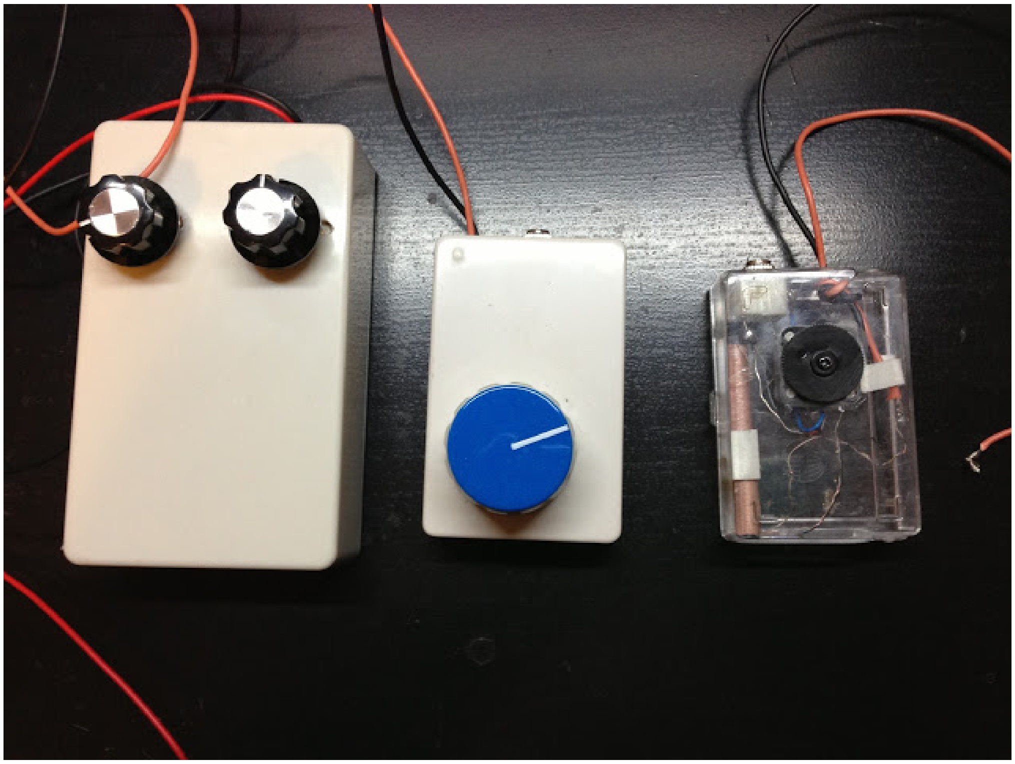

[http://1.bp.blogspot.com/-6Ns9sST5X4s/UVGPUo-lFmI/AAAAAAAAAxY/up4TGDU0jC0/s1600/IMG_1097.JPG]

Posted 22nd March 2013 by Billy

Solar Power Battery
Enter Your Zip Code & Connect. Pre-Screened Contractors Near You! homeadvisor.com

Prostate Cancer Warnings
Dr. Brownstein Reveals the 5 Signals when Prostate Cancer is Coming. w3.brownsteinhealth.com/Prostate

1 View comments

Ron 11 May 2016 at 18:41

Do you sell crystal radio kits,if you do I would love to buy a kit

Reply

Enter your comment...

Comment as: Karl Smirn ▼ **Sign out**

Publish **Preview** ☐ Notify me

2nd April

DIY Tiny portable crystal radio laboratory 135 x 70 x 30 mm

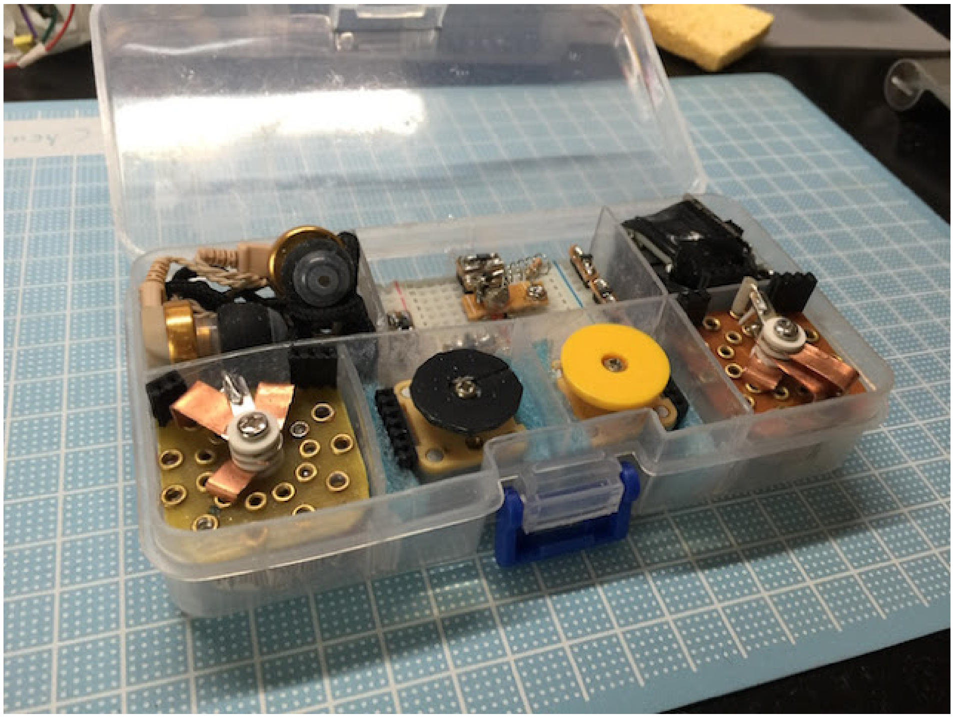

Inspired by skywave's traveling crystal set. I built this DIY Tiny (135 x 70 x 30 mm in size.) portable crystal radio laboratory.
I found a tiny sorting box made of low loss material (pp) and start to put things in:
- a couple of the DIY tiny multi-pole rotary switches (that I built) soldered in with the ferrite coils. I used 31mm ferrite core with100u permeability and Litz wire of 0.04 mm x 60 threads (see the picture below for the no. of turns).The shortest handle of the rotary switch selects the number of turns of the antenna coil. The longest handle selects the tapping point of the detector in the tank coil. The one in the middle selects the number of turns of the tank coil to match with the tuning capacitor
- a couple of 2x270pf plastic variable capacitor.
- a pair of 150 ohm hearing aid ear buds (total 300 ohm).
- a tiny 200K to 300 ohm audio transformer that I purchased from taboo.com soldered in with switches for the different way to connect , e.g. 200k to 300 ohm for my hearing aid ear buds, or just connect 200K in parallel with my piezoelectric ear bud.
The transform can be improved as its not ideal to have impedance fixed at 200k. I am trying to open it up to see if I can add additional taps without breaking the hair thin wires inside.
- A breadboard with the following different types of detectors
-- Crystal
-- BAT85
-- 1N60

-- 1N34A
-- 1SS86 x 2
-- HSM 2820 x 2
-- HSM 2850 x 2
-- 3SK143-Q (*3DQ) x 2
-- 33K resistor to work with 3DQ.
- Connecting wires (hidden under the breadboard).
- Tiny 50uA DC current meter (stilling trying to find one. So far I can only find a tiny 500uA meter that does not move much as the crystal radio current is so small). If I ever find one, it'll take the place currently used by the earbud.

I tried this in the hotel room when traveling and got good results.
- Tested using my 700mm umbrella loop antenna.
- Tested by hanging 3 meters of horizontal wire at the top of the curtain inside the room. [https://4.bp.blogspot.com/-Uq_TZea4wT4/VwBnupV99NI/AAAAAAAACeQ/OMB8y1i7TUsNOj0Eox2Ag9o-DGAV79HJw/s1600/IMG_6921.JPG]

Here are the pictures:

旅行矿机

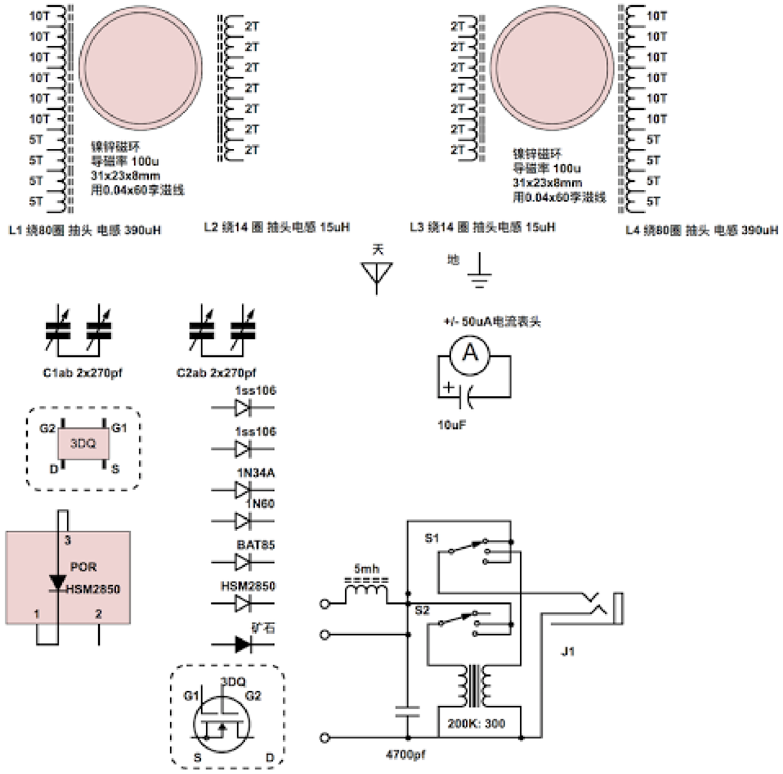
10T
10T
10T
10T
10T
10T
5T
5T
5T
5T
镍锌磁环
导磁率 100u
31x23x8mm
用0.04x60李滋线
2T
2T
2T
2T
2T
2T
2T
L1 绕80圈 抽头 电感 390uH
L2 绕14 圈 抽头电感 15uH
L3 绕14 圈 抽头电感 15uH
L4 绕80圈 抽头 电感 390uH
天
地
+/- 50uA电流表头
A
+
10uF
C1ab 2x270pf
C2ab 2x270pf
1ss106
1ss106
1N34A
1N60
BAT85
HSM2850
矿石
G2
G1
3DQ
D
S
3
POR
HSM2850
1
2
3DQ
G1
G2
S
D
5mh
S1
S2
J1
200K: 300
4700pf

[https://4.bp.blogspot.com/-4akx2SkEdjU/VwBstylVhDI/AAAAAAAACes/teuGgRr0 u9EyBOiGRm-BiyMI4ZpEH1OMQ/s1600/Screen%2BShot%2B2016-03-06%2Bat%2B10.59.31%2BPM.png]

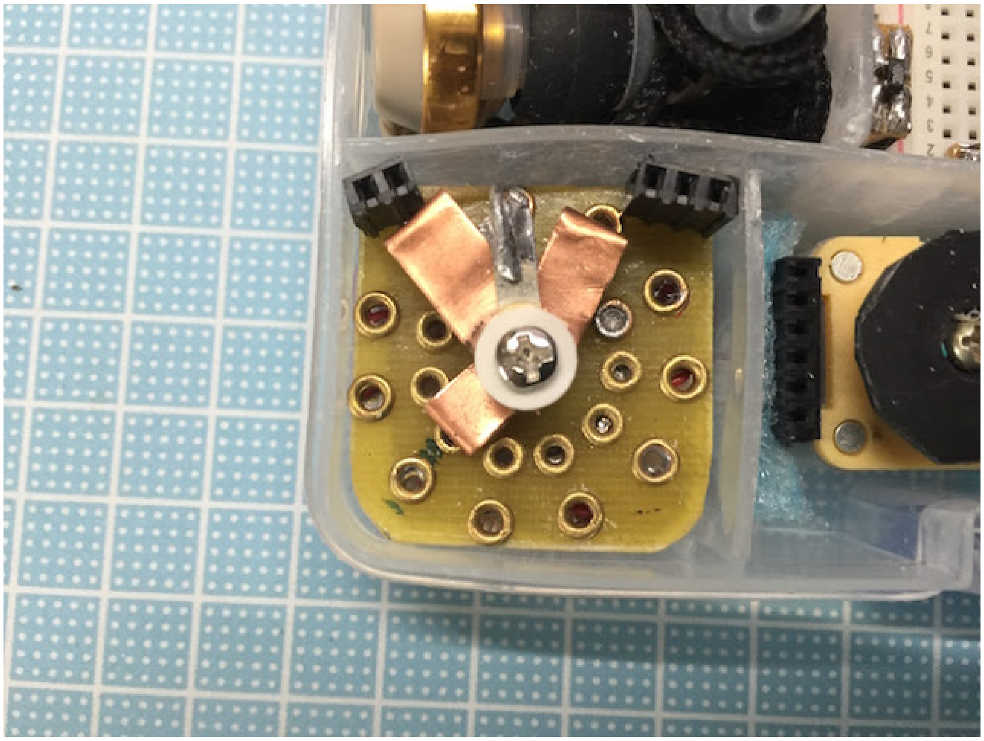

[https://1.bp.blogspot.com/-yylK17YOGV0/VwBnt1y4D_I/AAAAAAAACeA/WTuQCIOAYmcfpZ5ZTmJjAVaH72K2bpSmg/s1600/IMG_6917.JPG]

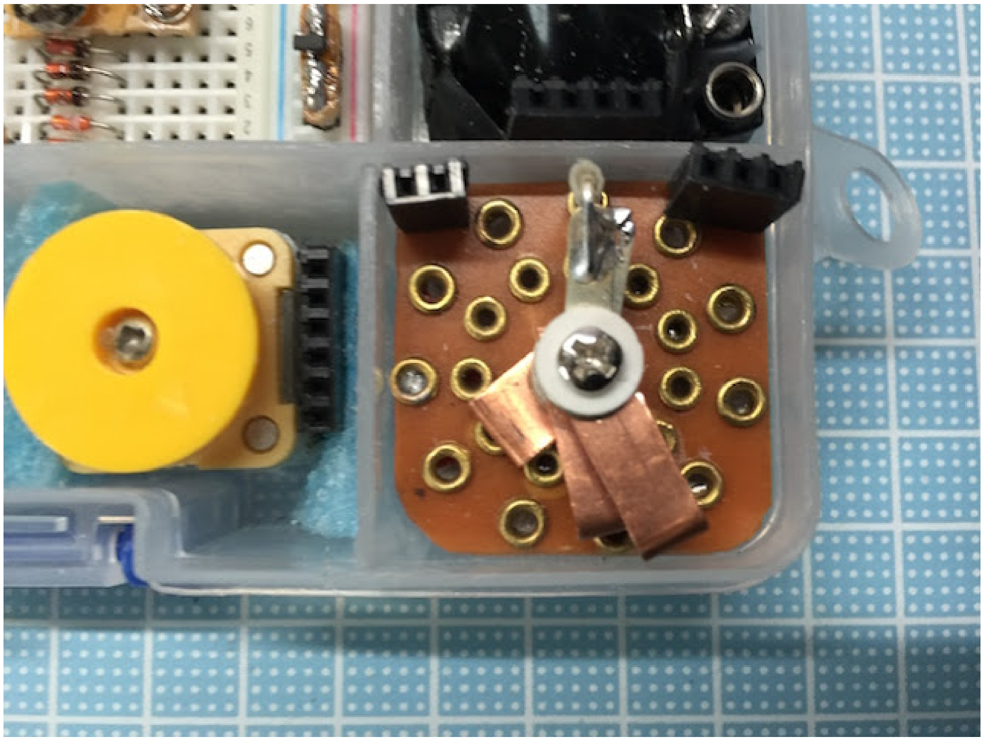

[https://3.bp.blogspot.com/-

XIfomc4Xom4/VwBnt3EiYLI/AAAAAAAACd8/Y5QgpHjMWM4hhB_NLiIxvjXAoicfQPd7g/s1600/IMG_6918.JPG]

[https://3.bp.blogspot.com/-M1ndSnQEU-k/VwBnuTtKiaI/AAAAAAAACeI/-l029n_uDjgBjMlk78nJ_7fn7st6j9-1A/s1600/IMG_6919.JPG]

[https://3.bp.blogspot.com/-

t1rJfevWnFo/VwBnukM43DI/AAAAAAAACeM/By6uiDuYMEwvISn5t6Soj9q_xSUR36IeA/s1600/IMG_6920.JPG]

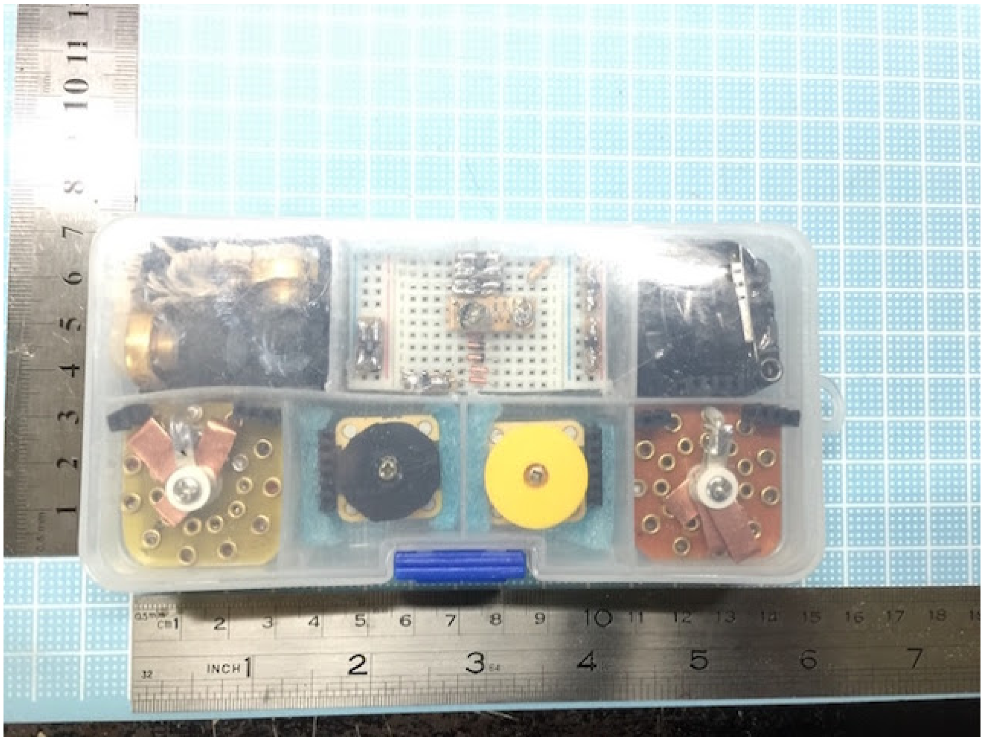

[https://2.bp.blogspot.com/-RlpVGlrkVbo/VwBnu2DM8qI/AAAAAAAACeU/4gvaXEIH8lkNs8P7d81F9r1Oc0BPOLV6Q/s1600/IMG_6922.JPG]

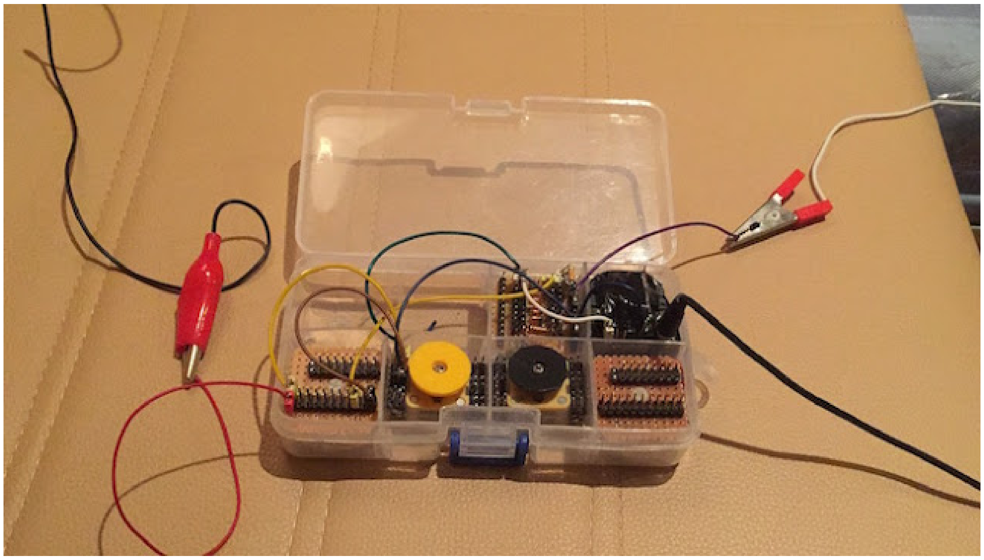

[https://1.bp.blogspot.com/-VAHDtVH_RwA/VwBn6fm2QcI/AAAAAAAACeY/OZU-FImODiUTo3P2Aa3ZMgJXHTJFq0Zsw/s1600/IMG_6538.JPG]

[https://2.bp.blogspot.com/-sh6TDwIRWf4/VwBoGMznkJI/AAAAAAAACec/fiF-jcUwiEcjDl7_WKHE2fvst41d4fTuw/s1600/IMG_6533.JPG]

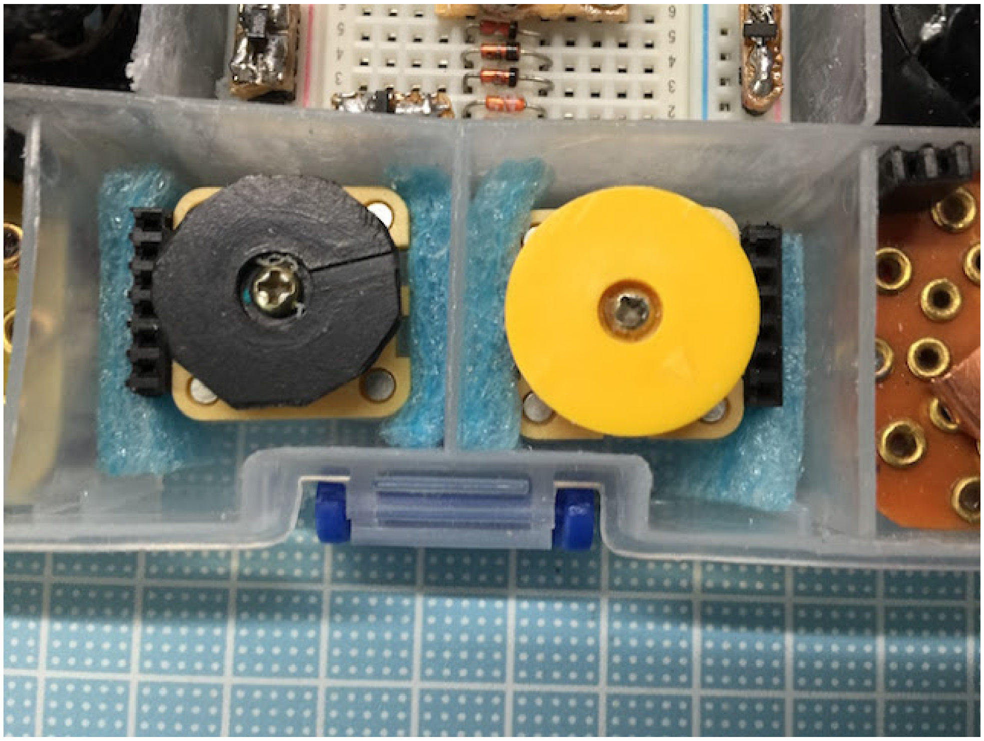

[https://1.bp.blogspot.com/-Re6TZmQrYU4/VwBntySygKI/AAAAAAAACeE/u-mA5R-jma4e2qgUjJd7plK5bF2anqABg/s1600/IMG_6916.JPG]

Posted 2nd April by Billy

0 Add a comment

Enter your comment...

Comment as: Unknown (Goo ▼ **Sign out**

Publish **Preview** ☐ Notify me

DK Eyewitness

CRYSTAL & GEM

Eyewitness
CRYSTAL & GEM

Apatite

Cut topazes

Danburite

Chalcedony

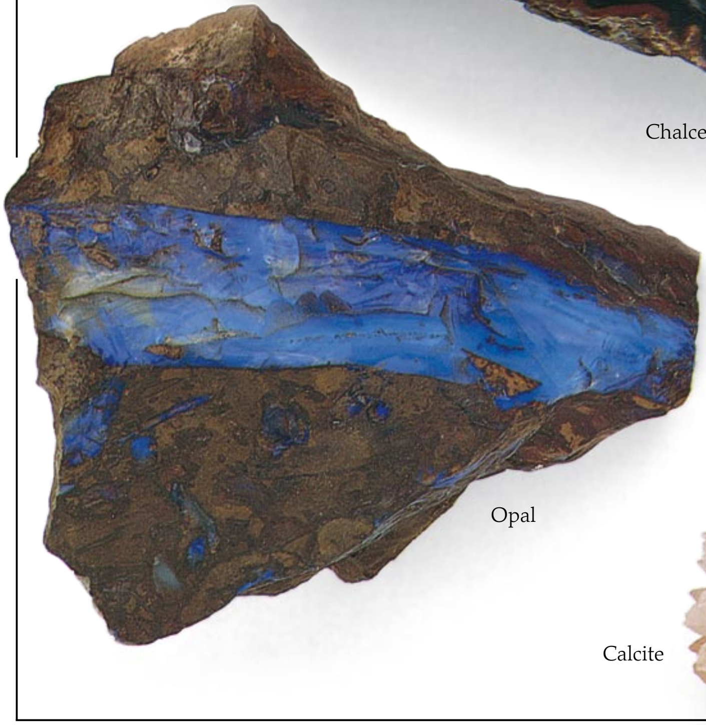

Opal

Calcite

Sphaerocobaltite

Eyewitness
CRYSTAL & GEM

Crocoite

Written by
Dr. R. F. SYMES and Dr. R. R. HARDING

Dumortierite bottle

Cut garnets

Cut tourmalines

Aragonite

Microcline

Meta-tobernite

DK
DK Publishing, Inc.

Malachite

Gold

Cut tourmaline

Cut topaz

Cut sapphire

LONDON, NEW YORK,
MELBOURNE, MUNICH, and DELHI

Project editor Louise Pritchard
Art editor Thomas Keenes
Senior editor Helen Parker
Senior art editors Julia Harris, Jacquie Gulliver
Production Louise Barrat
Picture research Cynthia Hole
Special photography Colin Keates ABIPP
(Natural History Museum)

Revised Edition
Managing editors Andrew Macintyre, Camilla Hallinan
Managing art editors Jane Thomas, Martin Wilson
Publishing manager Sunita Gahir
Category publisher Andrea Pinnington
Editors Angela Wilkes, Sue Nicholson
Art editor Catherine Goldsmith
Production Jenny Jacoby, Angela Graef
Picture research Marie Osborn, Kate Lockley
DTP designers Siu Ho, Andy Hilliard, Ronaldo Julien

U.S. editor Elizabeth Hester
Senior editor Beth Sutinis
Art director Dirk Kaufman
U.S. DTP designer Milos Orlovic
U.S. production Chris Avgherinos

This Eyewitness ® Guide has been conceived by
Dorling Kindersley Limited and Editions Gallimard

This edition published in the United States in 2007
by DK Publishing, Inc., 375 Hudson Street, New York, NY 10014

Copyright © 1991, © 2004, © 2007 Dorling Kindersley Limited

08 10 9 8 7 6 5 4 3
cd189 - 04/07

All rights reserved under International and Pan-American Copyright
Conventions. No part of this publication may be reproduced, stored in
a retrieval system, or transmitted in any form or by any means,
electronic, mechanical, photocopying, recording, or otherwise,
without the prior written permission of the copyright owner.

Published in Great Britain by Dorling Kindersley Limited.

A catalog record for this book is available from the Library of Congress.

ISBN 978-0-7566-3001-0 (HC) 978-0-7566-0663-3 (Library Edition)

Color reproduction by Colourscan, Singapore
Printed in China by Toppan Printing Co., (Shenzhen) Ltd.

Discover more at

www.dk.com

Crocoite

Agate

Tourmaline

Mother of pearl

Agate

Contents

6
What is a crystal?
8
A world of crystals
10
Natural beauty
12
Crystals—outside...
14
...and inside
16
The color of crystals
18
Identification
20
Natural growth
22
Good habits
24
Discovery—recovery
26
Growing from seed
28
Crystals at work
30
Good vibrations
32
Quartz
34
Diamond
36
Corundum
38
Beryl
40
Opal

Amethyst

42
Other gemstones
48
Collectors' items
50
Stones for carving
52
Precious metals
54
Animal and vegetable
56
What is it worth?
58
Making them sparkle
60
Lore and legends
62
Crystals at home
64
Did you know?
66
Identifying gemstones
68
Find out more
70
Glossary
72
Index

What is a crystal?

CRYSTALS ARE ASSOCIATED WITH PERFECTION, transparency, and clarity. Many crystals fit these ideals, especially those cut as gemstones, but most are neither perfect nor transparent. Crystals are solid materials in which the atoms are arranged in a regular pattern (pp. 14–15). Many substances can grow in characteristic geometric forms enclosed by smooth plane surfaces. They are said to have crystallized, and the plane surfaces are known as faces. The word *crystal* is based on the Greek word *krystallos*, derived from *kryos*, meaning icy cold. In ancient times it was thought that rock crystal, a colorless variety of quartz, was ice that had frozen so hard it would never melt.

STATES OF MATTER
A material can exist as a solid, a liquid, or a gas depending on its temperature. Water is made of atoms of hydrogen and oxygen bound together to form molecules. In the vapor (steam) the molecules move about vigorously; in the liquid they move slowly; in the solid (ice) they are arranged in a regular order and form a crystalline solid. These ice crystals are about 450 times their real size.

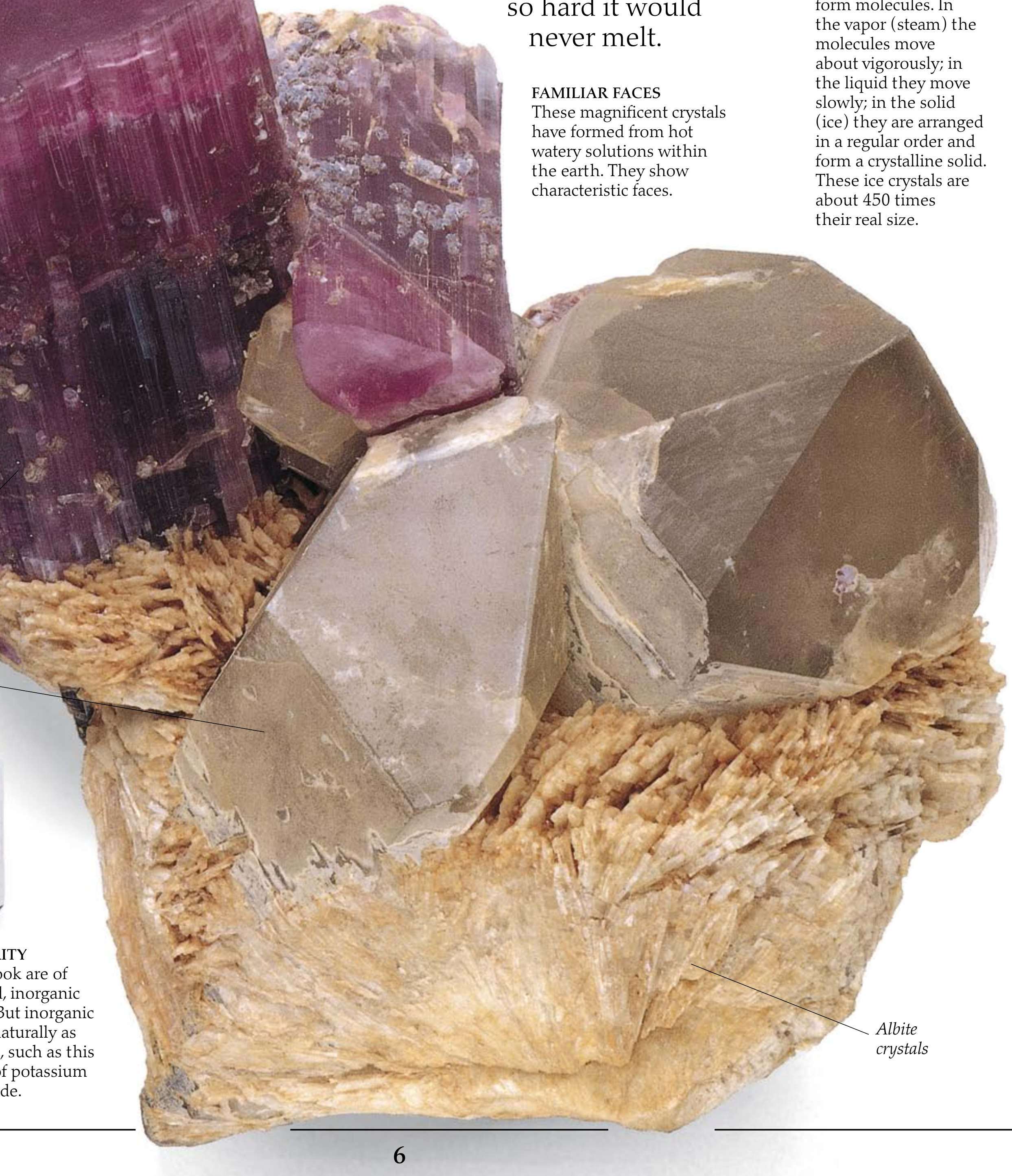

FAMILIAR FACES
These magnificent crystals have formed from hot watery solutions within the earth. They show characteristic faces.

CRYSTAL MINORITY
Most crystals in this book are of naturally occurring, solid, inorganic materials called minerals. But inorganic compounds not found naturally as minerals also form crystals, such as this artificially grown crystal of potassium magnesium sulfide.

MASSIVE MINERAL
Crystals only grow large and perfect in the right conditions. Most grow irregularly and the faces are often difficult to distinguish. This specimen of the mineral scapolite consists of a mass of small, poorly formed crystals. Minerals in this form are described as massive.

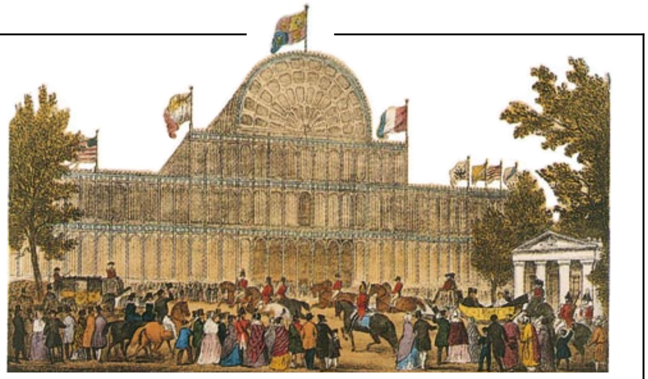

GLASS HOUSE
The Crystal Palace was built for the Great Exhibition of London of 1851, but was destroyed by fire in 1936. The roof and outer walls were made of nearly 300,000 panes of glass – not crystals.

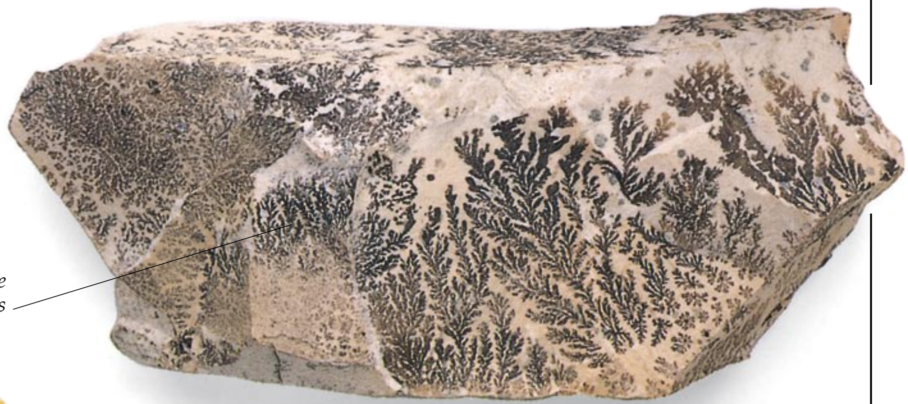

CRYSTAL LINING
These fern-like growths look like a plant but are in fact crystalline growths of the mineral pyrolusite. Such crystals are called dendrites (p. 21) and are often found lining joints and cracks in rocks.

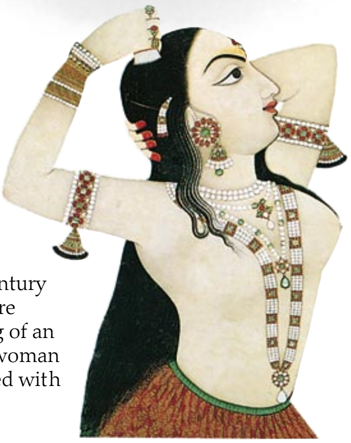

18th-century miniature painting of an Indian woman bedecked with jewelry

Cut heliodor (pp. 38–39)

Cut aquamarine (pp. 38–39)

GEM OF A CRYSTAL
Most gemstones are natural crystals chosen for their beauty, durability, and, in many instances, rarity. They are usually cut and polished (pp. 58–59). Crystals with the same composition and properties as naturally occurring minerals can now be grown artificially (pp. 26–27) and cut as gemstones.

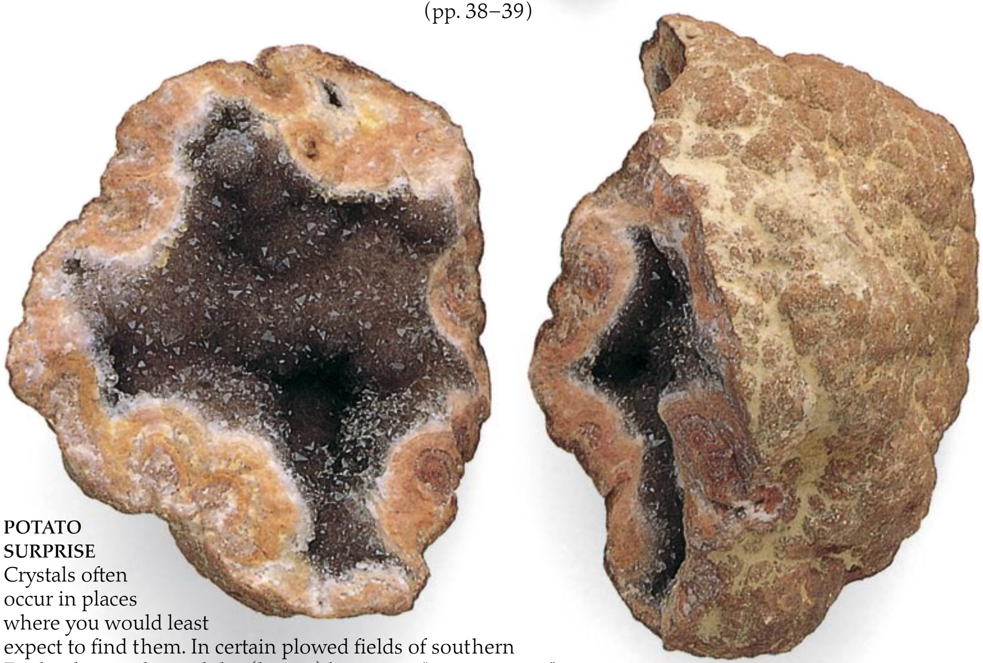

POTATO SURPRISE
Crystals often occur in places where you would least expect to find them. In certain plowed fields of southern England, irregular nodules (lumps) known as "potato stones" are found. When broken open, they often reveal sparkling crystals.

MOST IRREGULAR
Some of the objects which we know as "crystal" are glass and are not truly crystalline. Glass has little structure, as it is cooled too quickly for the atoms to arrange themselves into a regular order, and is said to be amorphous.

A world of crystals

Crystals are all around us. We live on a crystal planet in a crystal world. The rocks which form the earth, the moon, and meteorites – pieces of rock from space – are made up of minerals and virtually all of these minerals are made up of crystals. Minerals are naturally occurring crystalline solids composed of atoms of various elements. The most important of these are oxygen, silicon, and six common metallic elements including iron and calcium. Crystalline particles make up mountains and form the ocean floors. When we cross the beach we tread on crystals. We use them at home (pp. 62–63) and at work (pp. 28–29); indeed, crystals are vital to today's technology.

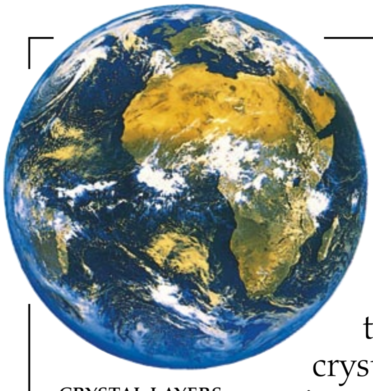

CRYSTAL LAYERS
The earth is formed of three layers: the crust, the mantle, and the core. These are made mostly of solid rock-forming minerals. Some rocks, such as pure marble and quartzite, are made of just one mineral, but most are made of two or more.

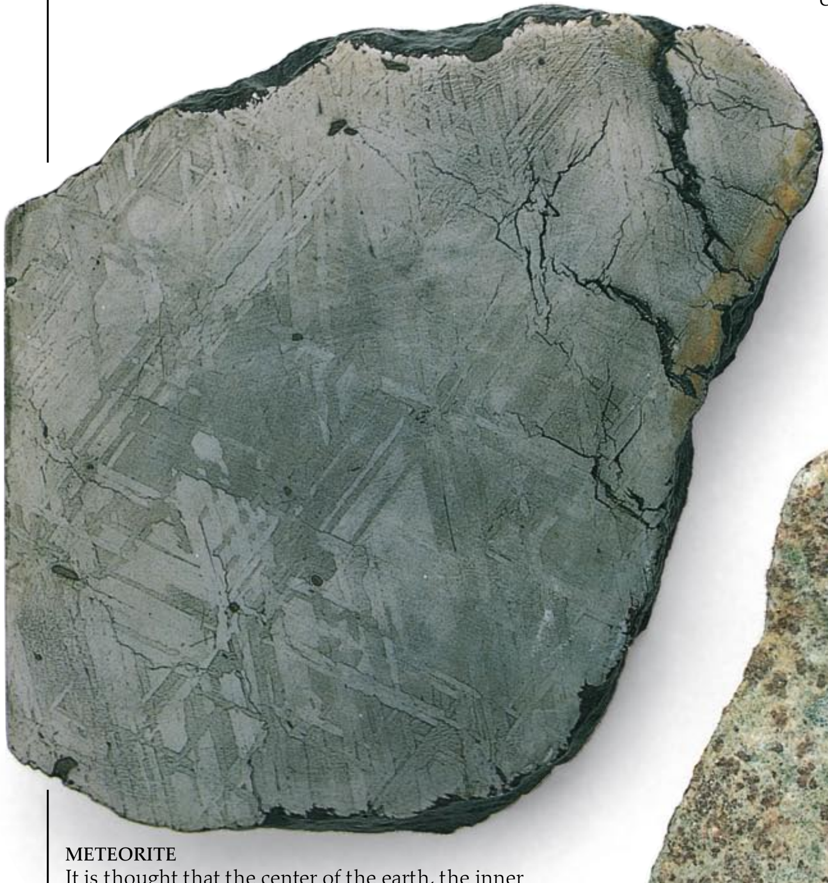

METEORITE
It is thought that the center of the earth, the inner core, may be similar in composition to this iron meteorite. It has been cut, polished, and acid-etched to reveal its crystalline structure.

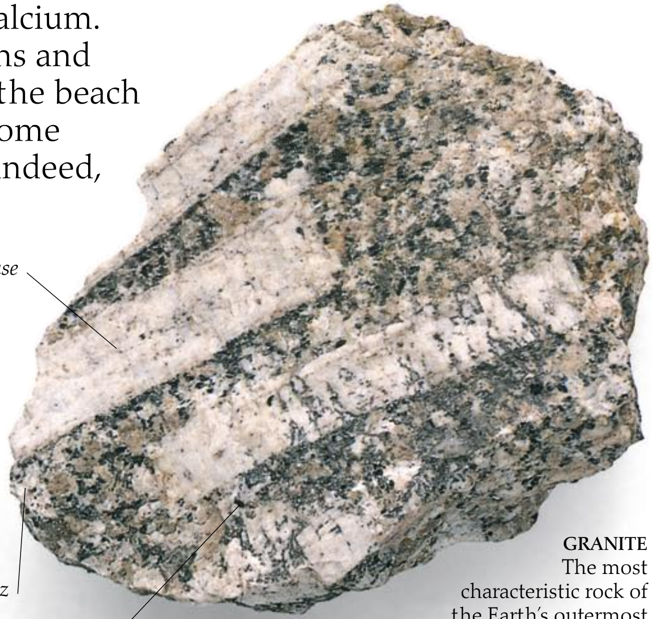

GRANITE
The most characteristic rock of the Earth's outermost layer, the continental crust, is granite. It consists mainly of the minerals quartz, feldspar, and mica. This specimen shows very large crystals of the feldspar mineral orthoclase, with small crystals of quartz and biotite mica.

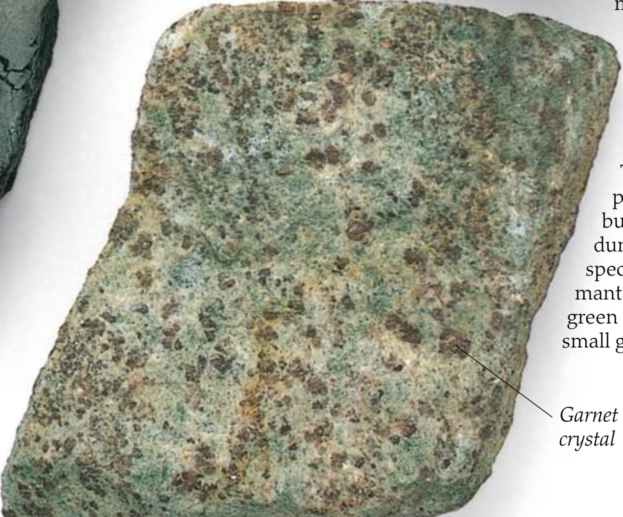

ECLOGITE
The earth's upper mantle is probably mostly peridotite but other rocks include dunite and eclogite. This specimen, originally from the mantle, is eclogite containing green pyroxene and small garnets.

LIQUID ROCK
Molten lava from inside the earth can erupt from volcanoes such as the Kilauea volcano, Hawaii, shown here. When the lava cools, minerals crystallize and it becomes a solid rock.

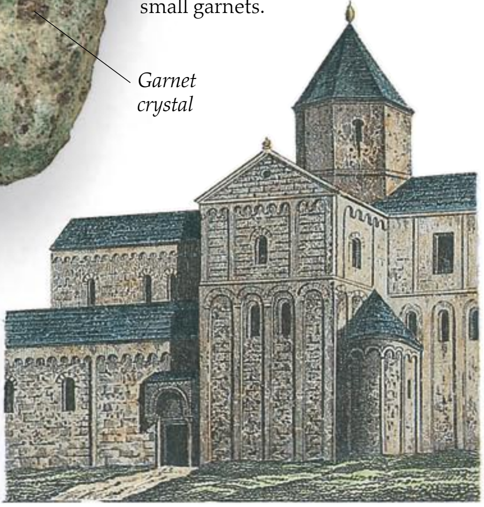

CRYSTAL STRENGTH
Most buildings are made of crystals. Both natural rock and artificial materials are mostly crystalline, and the strength of cement depends on the growth of crystals.

DOWN TO DUST
Pebbles, sand, and the greater part of soil are all formed from eroded rocks. Eventually, they will be eroded even further to form dust in the air (p. 32). Like the rocks they come from, these familiar things are all made up of crystals.

Feldspar crystal

Basalt pebble

Quartzite pebbles

Quartz sand grains

Soil

CRYSTAL CAVE
Fine stalactites and stalagmites form the spectacular scenery in these grottoes of Giita in Lebanon.

DRIP BY DRIP
Stalagmites and stalactites are mostly made of calcite crystals. This group of stalagmites grew upward from the floor of an abandoned mine as water, rich in calcium carbonate, dripped down from above.

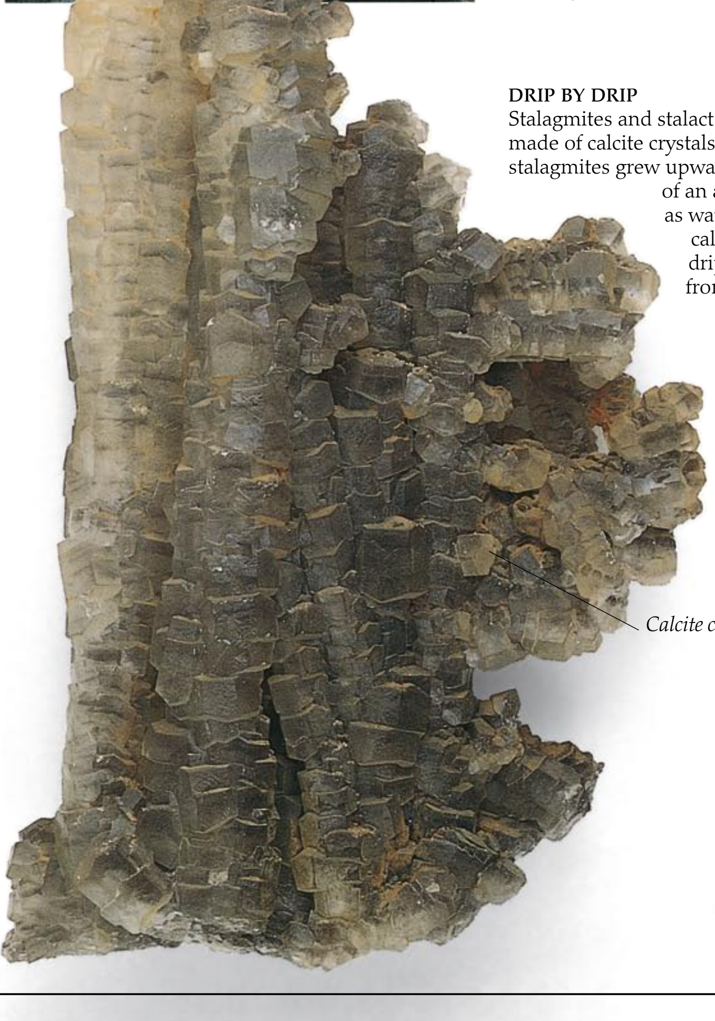

Organic crystals

Crystals do not only grow in rocks. The elements that make up most rock-forming minerals are also important to life on earth. For example, minerals such as calcite and apatite crystallize inside plants and animals.

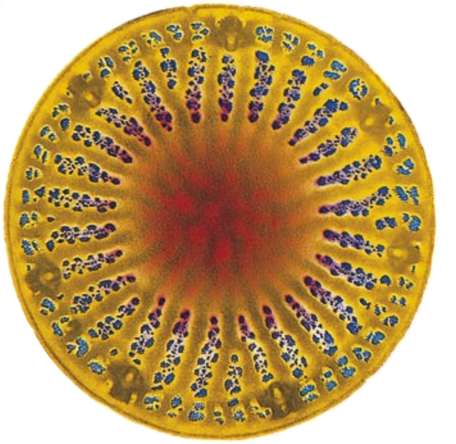

MICROCRYSTALS
This microscope picture of a diatom, *Cyclotella pseudostelligera*, shows a symmetrical (even) structure. Diatoms are microscopic algae whose cell walls are made up of tiny silica crystals.

ANIMAL MINERAL
Gallstones sometimes form inside an animal's gall bladder. This gallstone from a cow has exactly the same crystalline composition as struvite, a naturally occurring mineral.

STRESSFUL
Adrenaline is a hormone, a substance produced by the body to help it cope with stress. This greatly enlarged picture of adrenaline shows it is crystalline.

HUMAN APATITE
Bones contain tiny crystals of the mineral apatite. They make up the skeleton in vertebrate mammals – those that have a backbone, such as humans and horses. This is a human humerus (upper arm bone).

Natural beauty

Well-formed crystals are objects of great beauty and extreme rarity. Conditions have to be just right for them to grow (pp. 20–21) and any later changes in conditions must act to protect rather than destroy them. Even if they do grow and survive, many are destroyed by people during mining and other activities. Survivors are therefore of great interest. The crystals shown are about 60 percent of their real size.

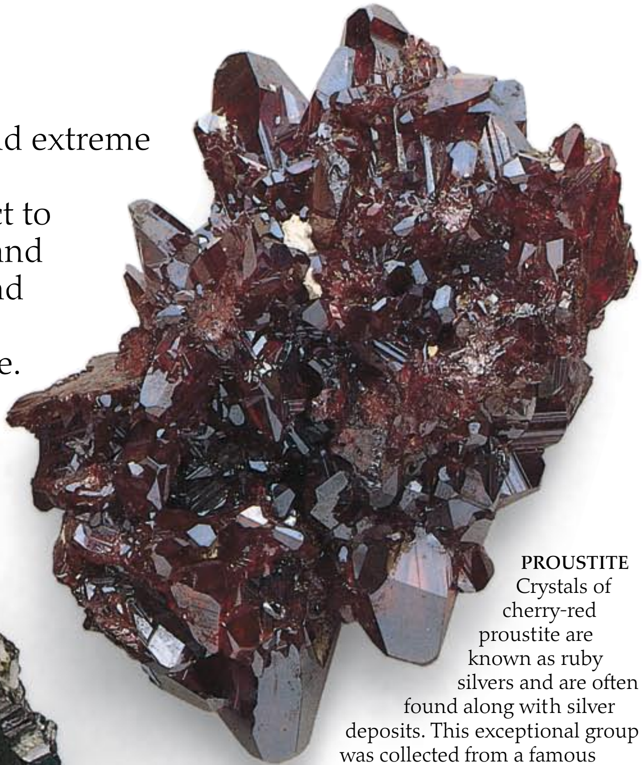

PROUSTITE
Crystals of cherry-red proustite are known as ruby silvers and are often found along with silver deposits. This exceptional group was collected from a famous silver mine area at Chanarcillo, Copiapo, Chile. The mines were extensively worked between 1830 and 1880.

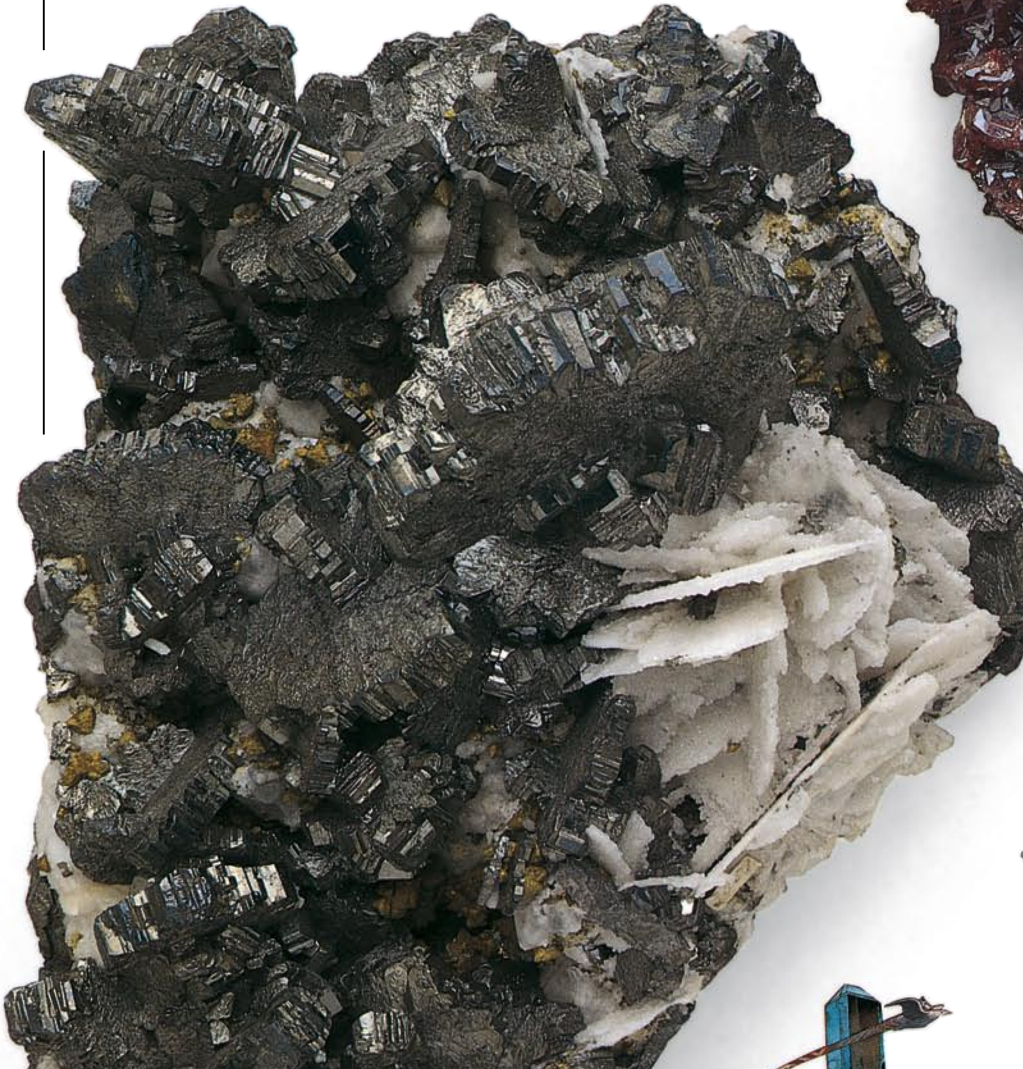

BOURNONITE
These magnificent bright-gray "cogwheel" crystals were collected from the Herodsfoot lead mine in Cornwall, England. Between 1850 and 1875 this mine produced bournonite crystals of a quality still unsurpassed elsewhere.

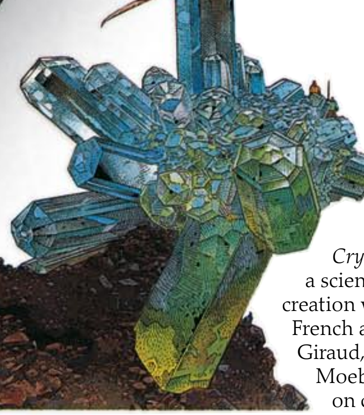

Crystal Dream a science fiction creation which the French artist Jean Giraud, known as Moebius, based on crystal shapes

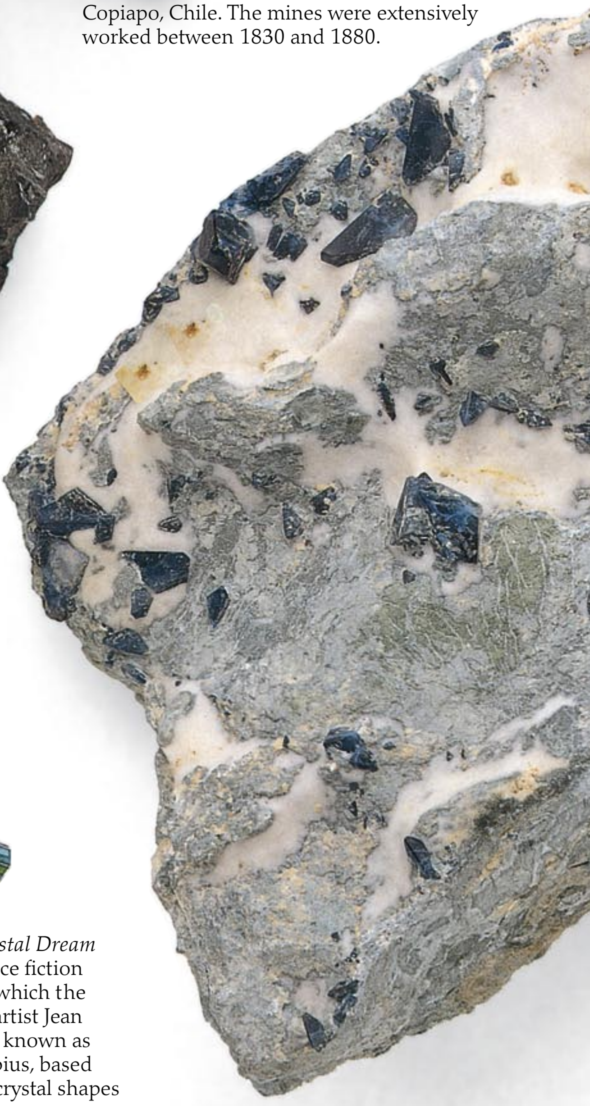

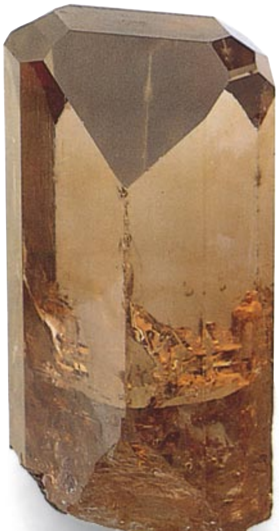

TOPAZ
This perfect topaz crystal was one of many wonderful crystals that were found in the last century close to the Urulga River in the remote areas of the Borshchovochnyy Mountains in Siberia. Most were yellowish brown and some weighed up to 22 lb (10 kg).

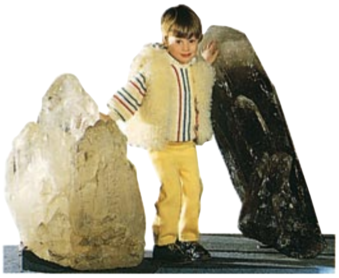

Giant rock crystal and smoky quartz crystal, as found inside cavities in certain rocks, especially in Brazil

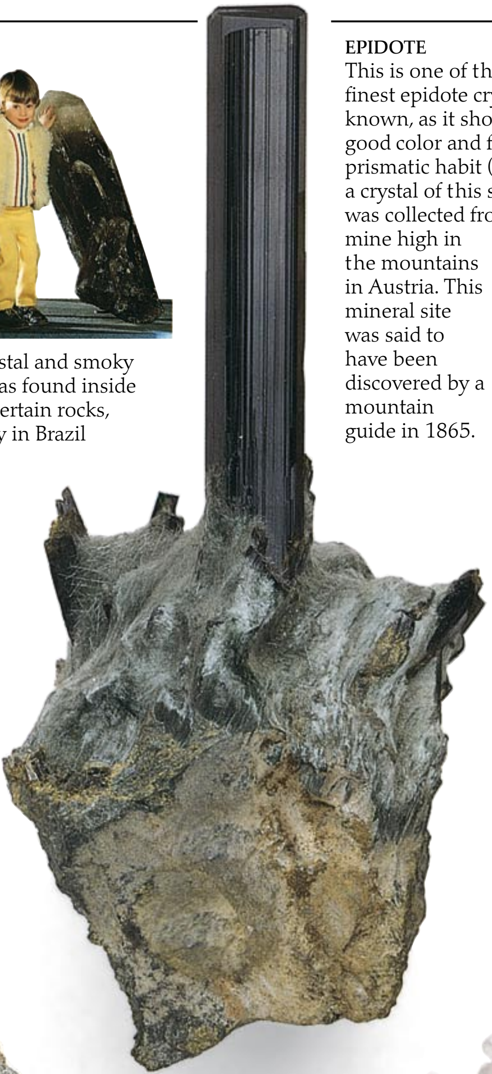

EPIDOTE
This is one of the finest epidote crystals known, as it shows good color and fine prismatic habit (p.23) for a crystal of this species. It was collected from a small mine high in the mountains in Austria. This mineral site was said to have been discovered by a mountain guide in 1865.

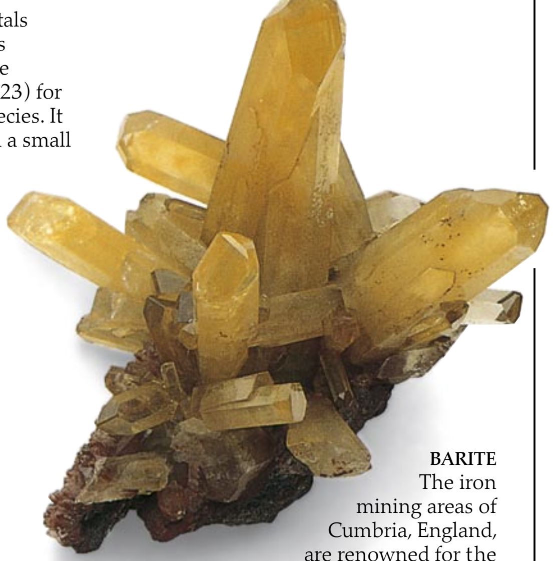

BARITE
The iron mining areas of Cumbria, England, are renowned for the quality of their barite crystals. The crystals display a range of colors, and each color comes mostly from one mine. These golden-yellow crystals came from the Dalmellington mine, Frizington, where many fine specimens were collected during the 19th century.

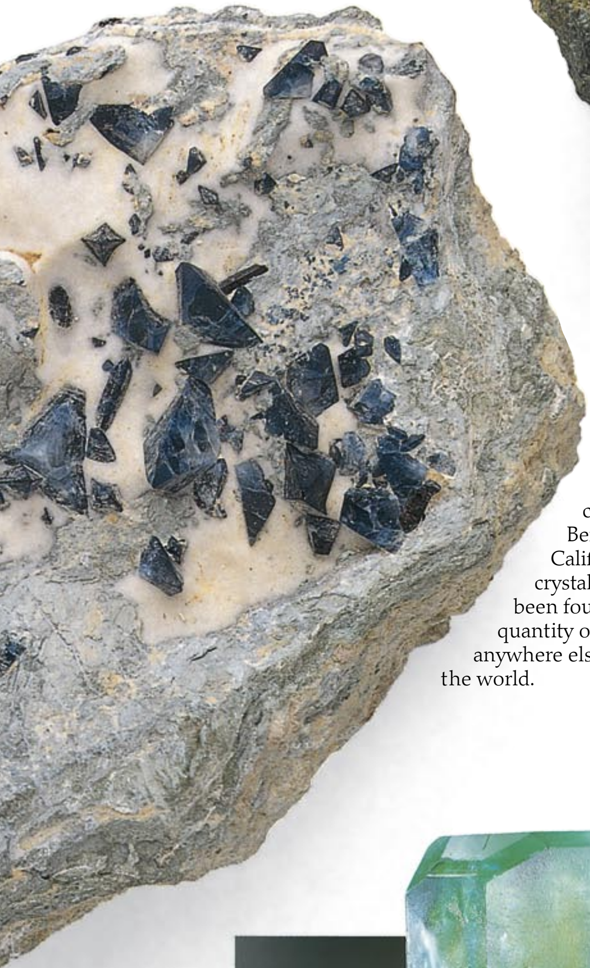

BENITOITE
These triangular-shaped, sapphire-blue crystals of benitoite (p. 49) were found close to the San Benito River in California. Such crystals have not been found in this quantity or quality anywhere else in the world.

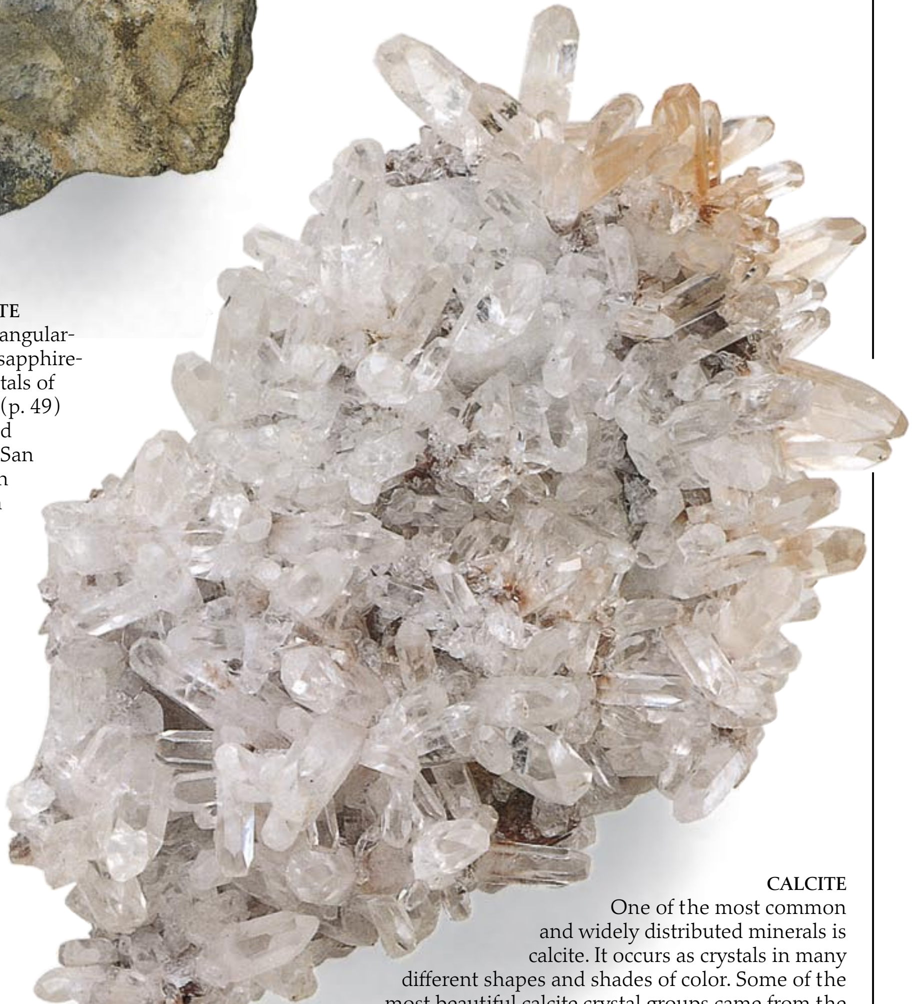

CALCITE
One of the most common and widely distributed minerals is calcite. It occurs as crystals in many different shapes and shades of color. Some of the most beautiful calcite crystal groups came from the Egremont iron mining area of Cumbria, England, in the late 19th century. This typical example consists of many fine colorless crystals, some of which are slightly tinged with red.

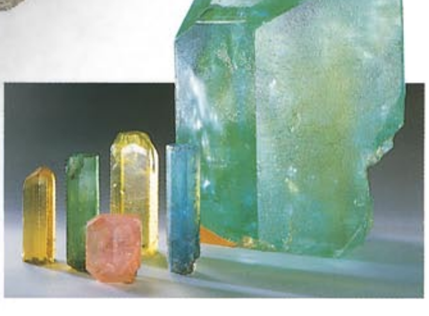

Beautifully formed beryl crystals from various parts of the world

Crystals–outside...

A WELL-FORMED CRYSTAL has certain regular or symmetrical features. One feature is that sets of faces have parallel edges. Another feature of many crystals is that for every face, there is a parallel face on the opposite side. Crystals may have three types of symmetry. If a crystal can be divided into two, so that each half is a mirror image of the other, the line that divides them is called a "plane of symmetry." If a crystal is rotated around an imaginary straight line and the same pattern of faces appears a number of times in one turn, then the line is an "axis of symmetry." Depending on how many times the pattern appears, symmetry around an axis is described as twofold, threefold, fourfold, or sixfold. If a crystal is entirely bounded by pairs of parallel faces then it has a "center of symmetry."

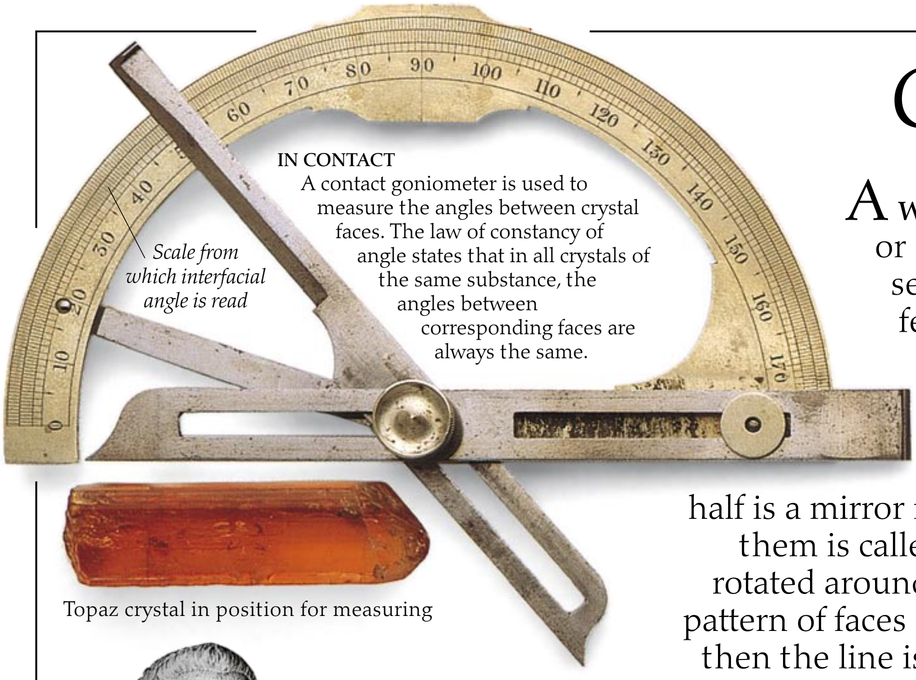

IN CONTACT
A contact goniometer is used to measure the angles between crystal faces. The law of constancy of angle states that in all crystals of the same substance, the angles between corresponding faces are always the same.

Scale from which interfacial angle is read

Topaz crystal in position for measuring

Romé de l'Isle (1736-90), who established the law of constancy of angle first proposed by the scientist Steno in 1669

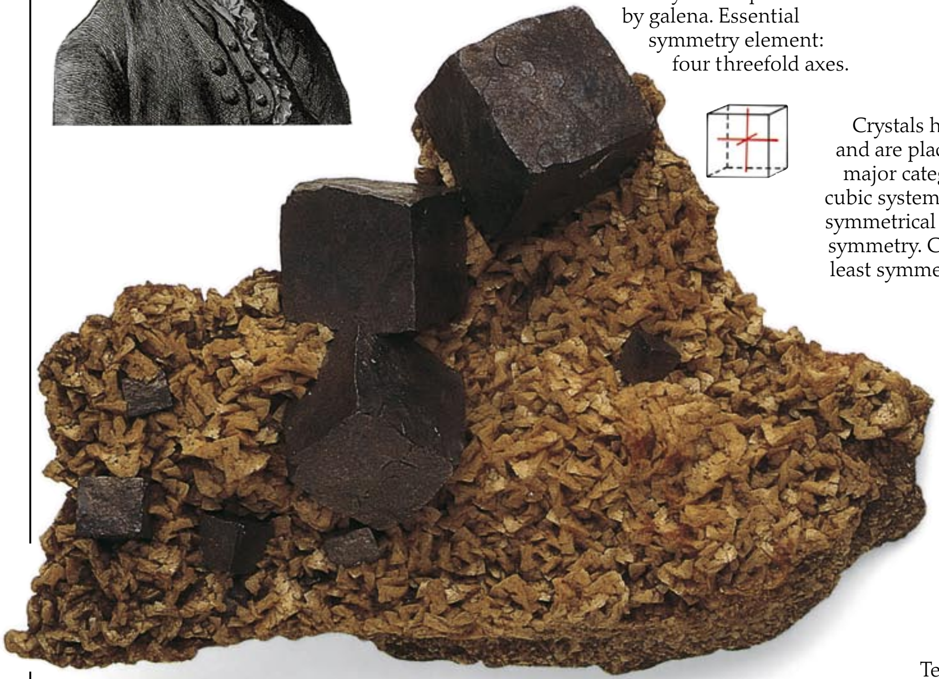

Cubic system represented by galena. Essential symmetry element: four threefold axes.

SEVEN SYSTEMS
Crystals have differing amounts of symmetry and are placed, according to this, in one of seven major categories called systems. Crystals in the cubic system have the highest symmetry. The most symmetrical have 9 planes, 12 axes, and a center of symmetry. Crystals in the triclinic system have the least symmetry with only a center of symmetry or no symmetry at all.

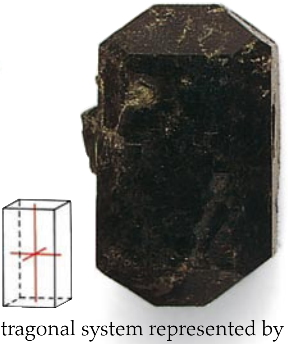

Tetragonal system represented by idocrase. Essential symmetry element: one fourfold axis.

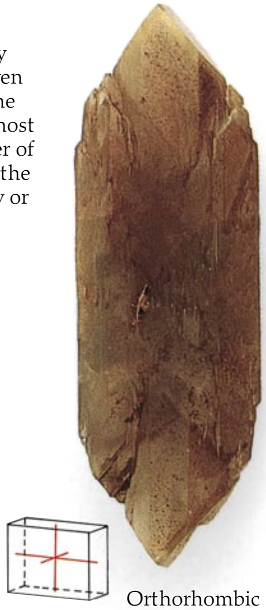

Orthorhombic system represented by barite. Essential symmetry element: three twofold axes.

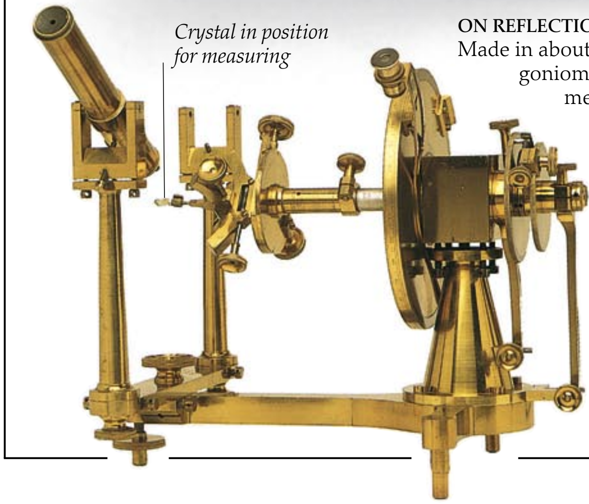

Crystal in position for measuring

ON REFLECTION
Made in about 1860, this optical goniometer is designed to measure the interfacial angles of small crystals by the reflection of light from their faces. The crystal is rotated until a reflection of light is seen from two adjacent faces. The angle between the two faces is read off the graduated circle on the right.

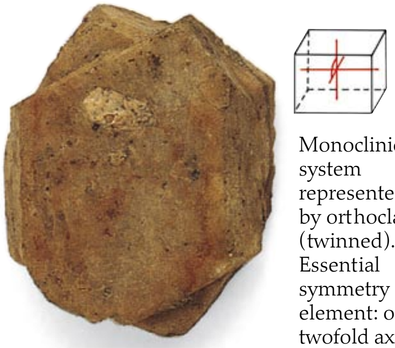

Monoclinic system represented by orthoclase (twinned). Essential symmetry element: one twofold axis.

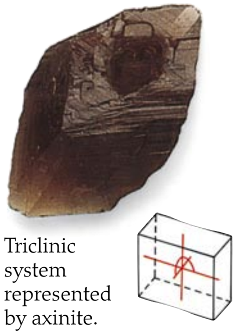

Triclinic system represented by axinite. No axis of symmetry.

Hexagonal system represented by beryl. Essential symmetry: one sixfold axis.

SAME BUT DIFFERENT
Some crystallographers (studiers of crystals) consider the trigonal system part of the hexagonal system. Both systems have the same set of axes, but the trigonal has only threefold symmetry. This is seen in the terminal faces.

Trigonal system represented by calcite. Essential symmetry: one threefold axis.

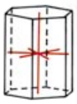

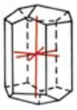

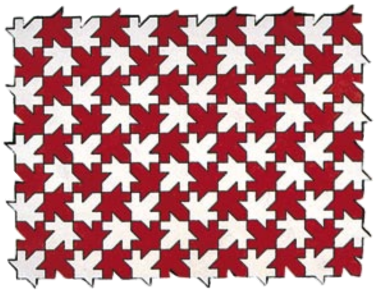

DESIGNED FOR SYMMETRY
This maple leaf design is one of 13 made to commemorate the 13th Congress of the International Union of Crystallography, held in Canada in 1981. The repetitive designs were based on the elements of crystal symmetry.

Triclinic model

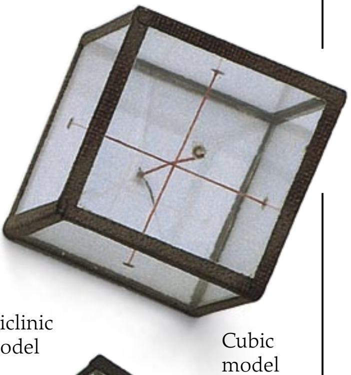

Cubic model

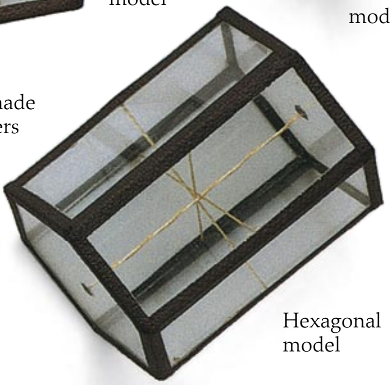

Hexagonal model

MODEL CRYSTALS
Crystal models were made to help crystallographers understand symmetry. These glass models were made in about 1900 in Germany. They contain cotton threads strung between the faces to show axes of rotation.

Form

Crystals of the same mineral may not look alike. The same faces on two crystals may be different sizes and therefore form different-shaped crystals. Crystals may also grow with a variation of "form." Shown here are three forms found in the cubic crystal system, illustrated by pyrite.

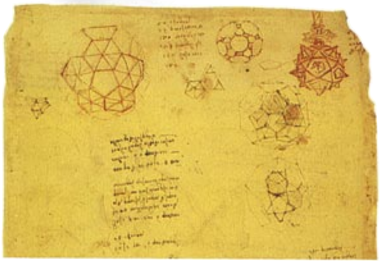

Studies of the transformation of geometrical bodies from Leonardo da Vinci's sketchbook

CUBE
A form of six square faces that make 90° angles with each other. Each face intersects one of the fourfold axes and is parallel to the other two.

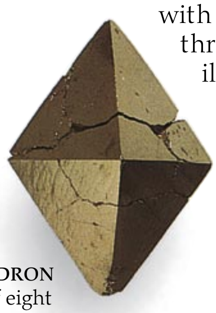

OCTAHEDRON
A form of eight equilateral triangular faces, each of which intersects all three of the fourfold axes equally.

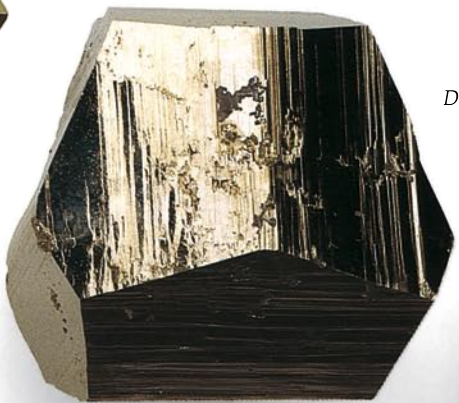

PYRITOHEDRON
This form consists of 12 five-sided faces. It is also known as a pentagonal dodecahedron.

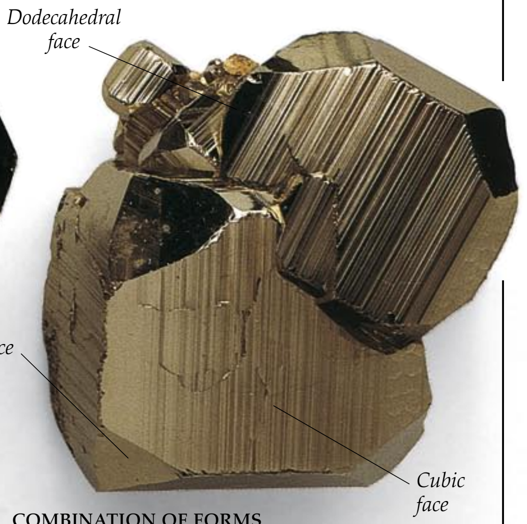

COMBINATION OF FORMS
These crystals show cubic faces combined with octahedral faces with poorly developed dodecahedral faces blending into the cubic faces.

Below: Diagram to show the relationship between different cubic forms

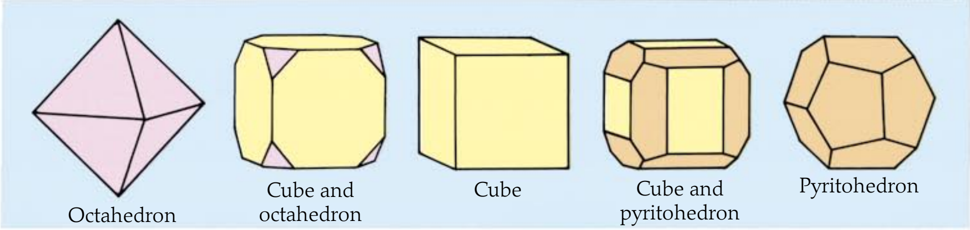

…and inside

THE INTERNAL ATOMIC STRUCTURE of crystals determines their regular shape and other properties. Each atom has its own special position and is tied to others by bonding forces. The atoms of a particular mineral always group in the same way to form crystals of that mineral. In early crystallography, the study of crystals, one of the most significant deductions was made by R. J. Hauy (p. 15) in 1784. In 1808, English chemist J. Dalton defined his theory that matter was built up from tiny particles called atoms. In 1895, German physicist W. Röntgen discovered X-rays, and in 1912, Laue (p. 15) realized that X-rays might help determine the arrangement of atoms within a solid. This was the start of our understanding of the inside of crystals.

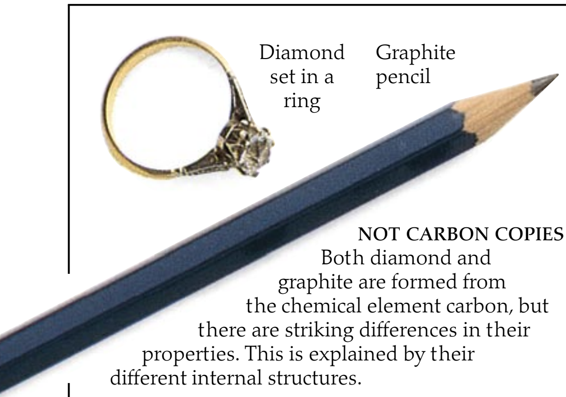

NOT CARBON COPIES
Both diamond and graphite are formed from the chemical element carbon, but there are striking differences in their properties. This is explained by their different internal structures.

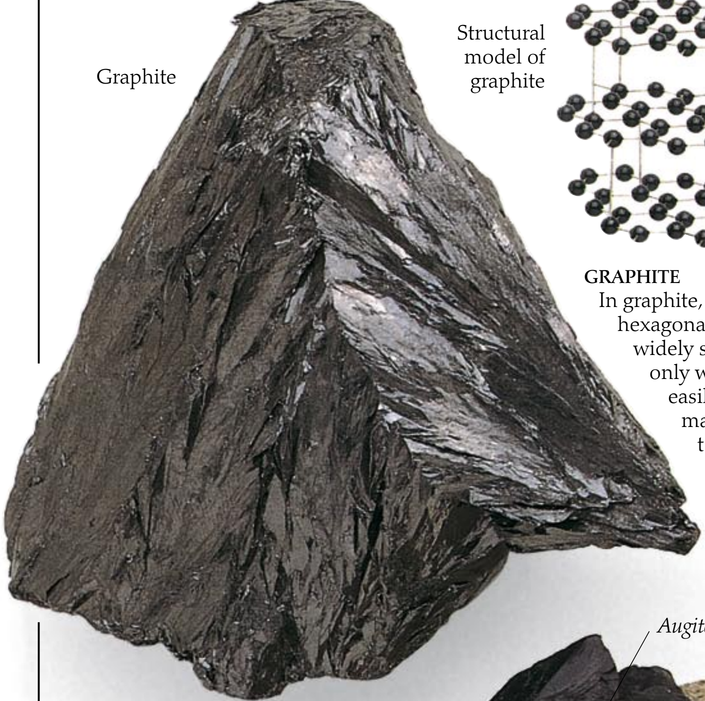
Graphite

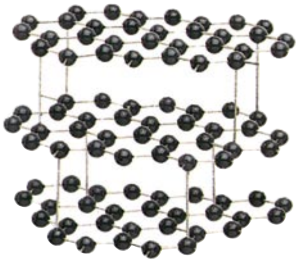
Structural model of graphite

GRAPHITE
In graphite, carbon atoms are linked in a hexagonal (six-sided) arrangement in widely spaced layers. The layers are only weakly bonded and can slip easily over one another, making graphite one of the softest minerals.

Diamond crystal

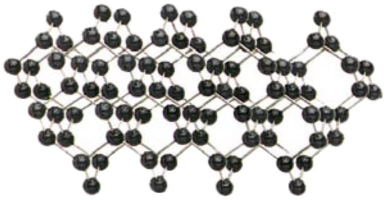
Structural model of diamond

DIAMOND
In diamond, each carbon atom is strongly bonded to four others to form a rigid compact structure. This structure makes diamond much harder than graphite.

Augite crystal

AUGITE
An important group of silicate minerals is the pyroxenes, including augite. Their internal structure is based on a single chain of SiO_4 tetrahedra.

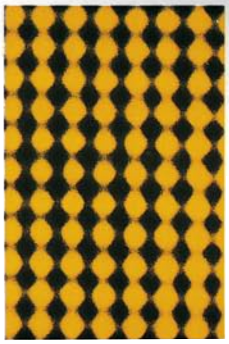

GOLD ATOMS
Crystalline solids have a complex latticework of atoms. This photograph shows the atomic lattice of gold magnified millions of times. Each yellow blob represents an individual atom.

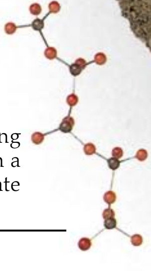
Model showing SiO_4 tetrahedra in a single-chain silicate

ACTINOLITE
Silicate minerals, present in all common rocks apart from limestone, have a basic unit of a tetrahedron (four faces) of one silicon and four oxygen atoms (SiO_4). Actinolite, a member of a group of minerals known as amphiboles, has a structure based on a double chain of these tetrahedra.

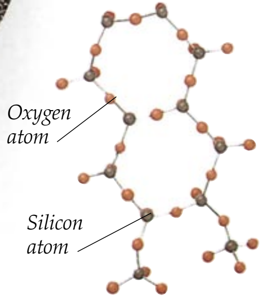

Model showing SiO_4 tetrahedra in a double-chain silicate

BERYL
In some silicate minerals, the internal structure is based on groups of three, four, or six SiO_4 tetrahedra linked in rings. Beryl (pp. 38–39) has rings made of groups of six tetrahedra.

MAX VON LAUE (1879–1960)
Laue showed with X-ray photographs that crystals were probably made of planes of atoms.

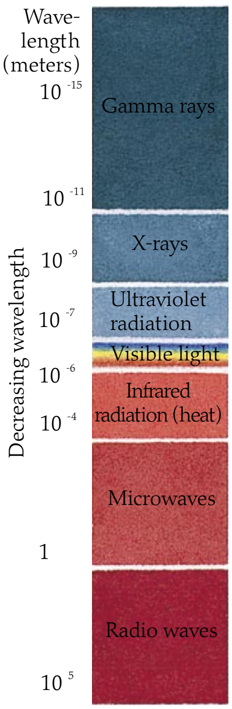

ELECTRO-MAGNETIC WAVES
X-rays are part of the electromagnetic radiation spectrum. All radiations can be described in terms of waves, many of which, such as light, radio, and heat, are familiar. The waves differ only in length and frequency. White light, which is visible to the human eye, is composed of electromagnetic waves varying in wavelength between red and violet in the spectrum (p. 16), but these visible rays are only a fraction of the whole spectrum.

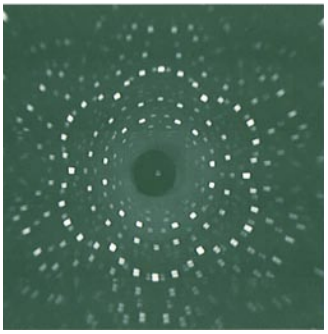

X-RAY PHOTO
This Laue photograph shows the diffraction, or splitting up, of a beam of X-rays by a beryl crystal. The symmetrical pattern is related to the hexagonal symmetry of the crystal.

Cleavage

Some crystals split along well-defined planes called cleavage planes which are characteristic for all specimens of that species. Cleavage forms along the weakest plane in the structure and is direct evidence of the orderly arrangement of atoms.

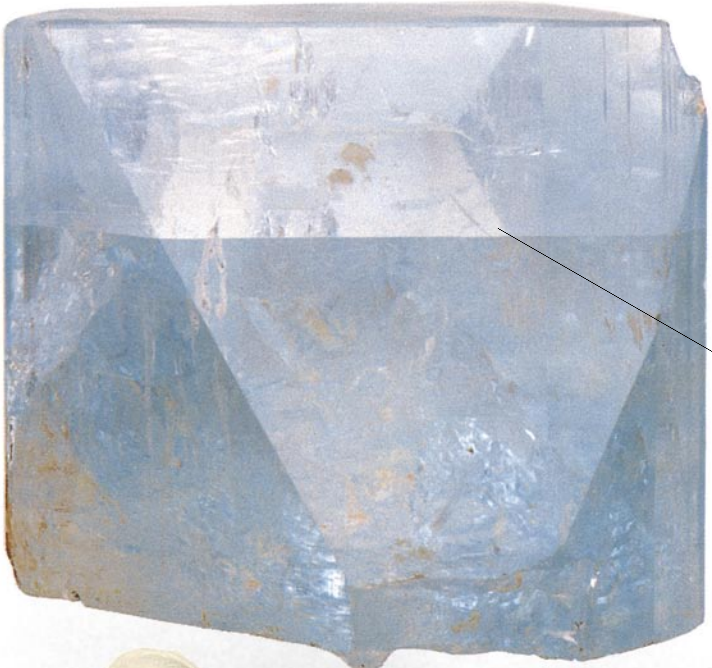

TOPAZ
This fine blue topaz crystal from Madagascar shows a perfect cleavage. Topaz is one of a group of silicates with isolated SiO_4 groups in their structure.

Thin cleavage flakes

MICA
The micas are a group of silicate minerals which have a sheet structure. The atomic bonding perpendicular (at right angles) to the sheet structure is weak, and cleavage occurs easily along these planes.

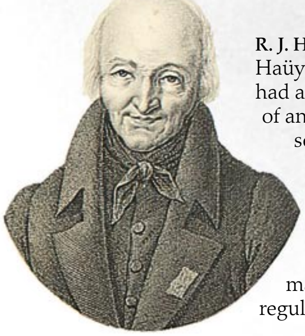

R. J. HAÜY (1743–1822)
Haüy realized that crystals had a regular shape because of an inner regularity. He had seen how calcite often fractures along cleavage planes into smaller diamond shapes (rhombs) and decided crystals were built up by many of these small, regularly stacked blocks.

QUARTZ
The structure of quartz is based on a strongly bonded, three-dimensional network of silicon and oxygen atoms. Crystals do not cleave easily but show a rounded, concentric fracture known as conchoidal.

The color of crystals

The color of a crystal can be its most striking feature. The causes of color are varied, and many minerals occur in a range of colors. Something looks a particular color largely due to your eye and brain reacting to different wavelengths of light (p. 15). When white light (daylight) falls on a crystal, some of the wavelengths may be reflected, and some absorbed. If some are absorbed, those remaining will make up a color other than white because some of the wavelengths that make up white light are missing. Sometimes light is absorbed and re-emitted without changing and the mineral will appear colorless.

MOONSTONES
The most familiar gem variety of the feldspar minerals is moonstone (p. 45). The white or blue sheen is caused by layers of tiny crystals of albite within orthoclase.

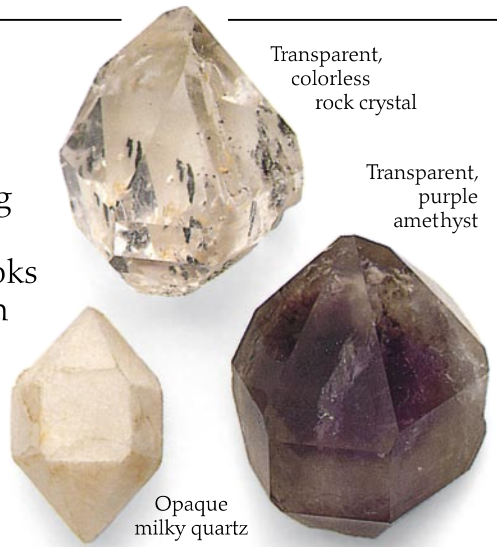
Transparent, colorless rock crystal
Transparent, purple amethyst
Opaque milky quartz

SEE-THROUGH OR OPAQUE
Crystals can be transparent (they let through nearly all the light and can be seen through), translucent (they let some light through but cannot be seen through clearly), or opaque (they do not let any light through and cannot be seen through at all). Most gemstones are transparent but can be colored or colorless.

Idiochromatic

Some minerals are nearly always the same color because certain light-absorbing atoms are an essential part of their crystal structure. These minerals are described as idiochromatic. For example, copper minerals are nearly always red, green, or blue according to the nature of the copper present.

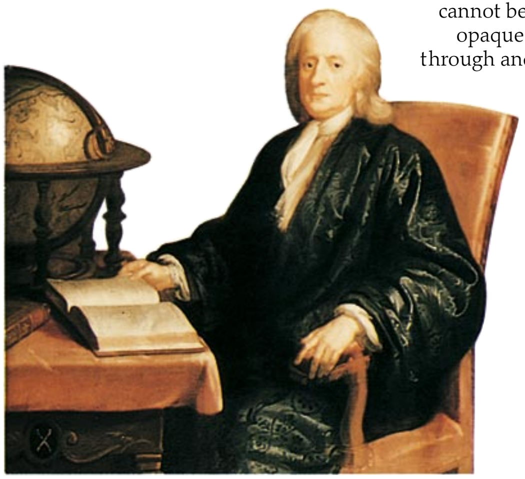

ISAAC NEWTON (1642–1727)
Sir Isaac Newton was an English scientist who achieved great fame for his work on, among other things, the nature of white light. He discovered that white light can be separated into seven different colors, and followed this with an explanation of the theory of the rainbow.

The colors known as the spectrum, produced by dispersion (scattering) of white light in a prism

SULFUR
Sulfur is an idiochromatic mineral and normally crystallizes in bright yellow crystals. These are often found as encrusting masses around volcanic vents and fumaroles (p. 20).

AZURITE
Azurite is a copper mineral which is always a shade of blue – hence the term azure blue. In ancient times it was crushed and used as a pigment.

Allochromatic

A large number of minerals occur in a wide range of colors caused by impurities or light-absorbing defects in the atomic structure. For example, quartz, diamond, beryl, and corundum can be red, green, yellow, and blue. These minerals are described as allochromatic.

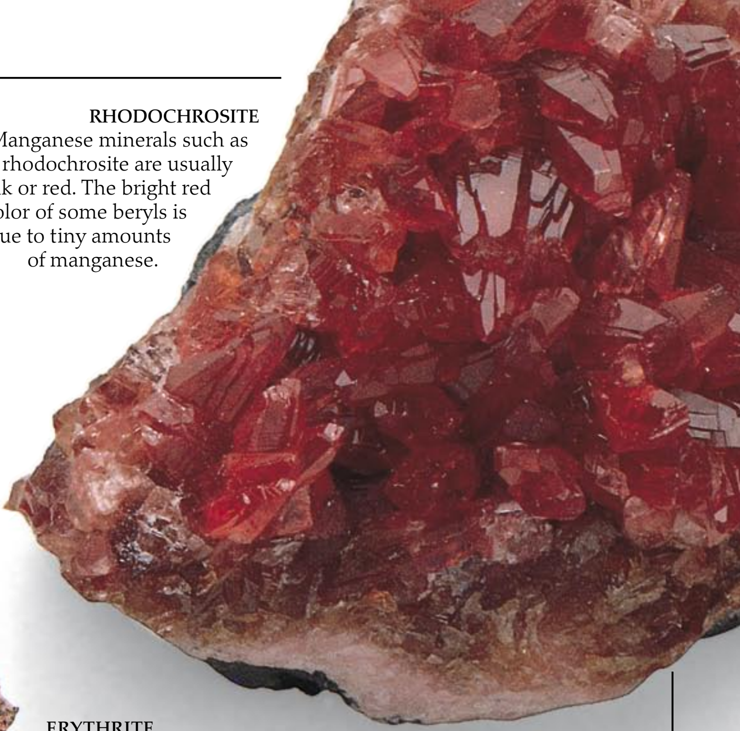

RHODOCHROSITE
Manganese minerals such as rhodochrosite are usually pink or red. The bright red color of some beryls is due to tiny amounts of manganese.

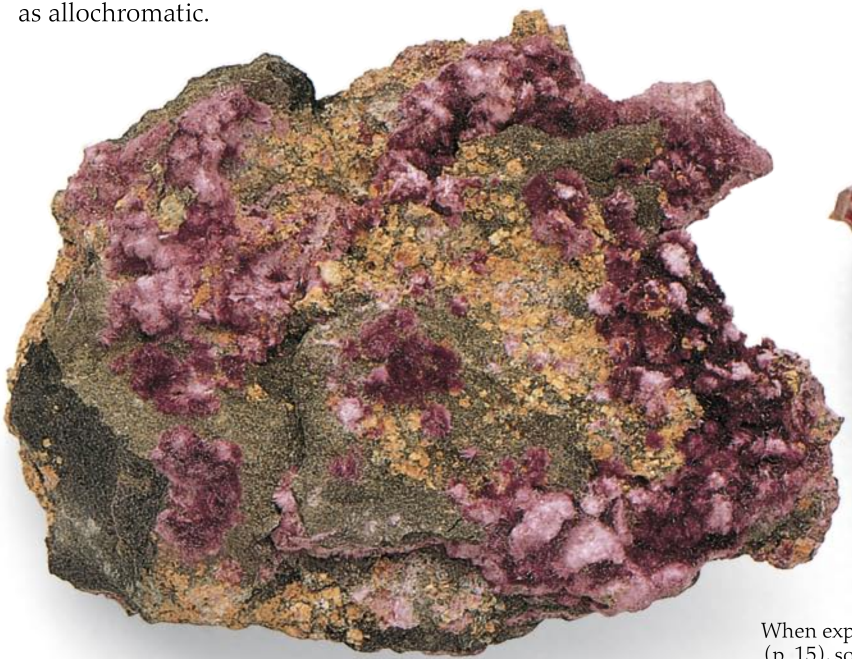

ERYTHRITE
Cobalt minerals such as erythrite are usually pink or reddish. Trace amounts of cobalt may color normally colorless minerals.

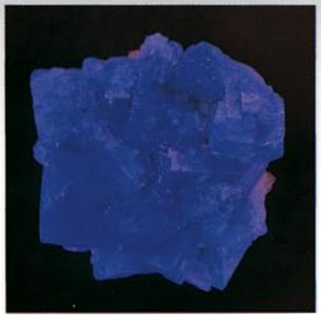

FLUORITE
When exposed to invisible ultraviolet light (p. 15), some minerals emit visible light of various colors. This is called fluorescence, usually caused by foreign atoms called activators in the crystal structure. The fluorescent color of a mineral is usually different from its color in daylight. This fluorite crystal is green in daylight.

Play of colors

The color in some minerals is really a play of colors like that seen in an oil slick or a soap bubble. This may be produced when the light is affected by the physical structure of the crystals, such as twinning (p. 21) or cleavage planes (p. 15), or by the development during growth of thin films. Microscopic "intergrowths" of plate-like inclusions (p. 21) also interfere with the light.

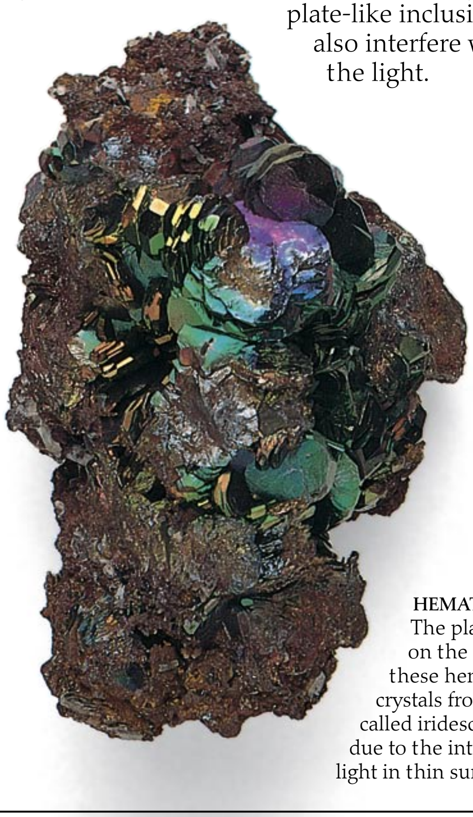

HEMATITE
The play of colors on the surface of these hematite crystals from Elba is called iridescence. It is due to the interference of light in thin surface films.

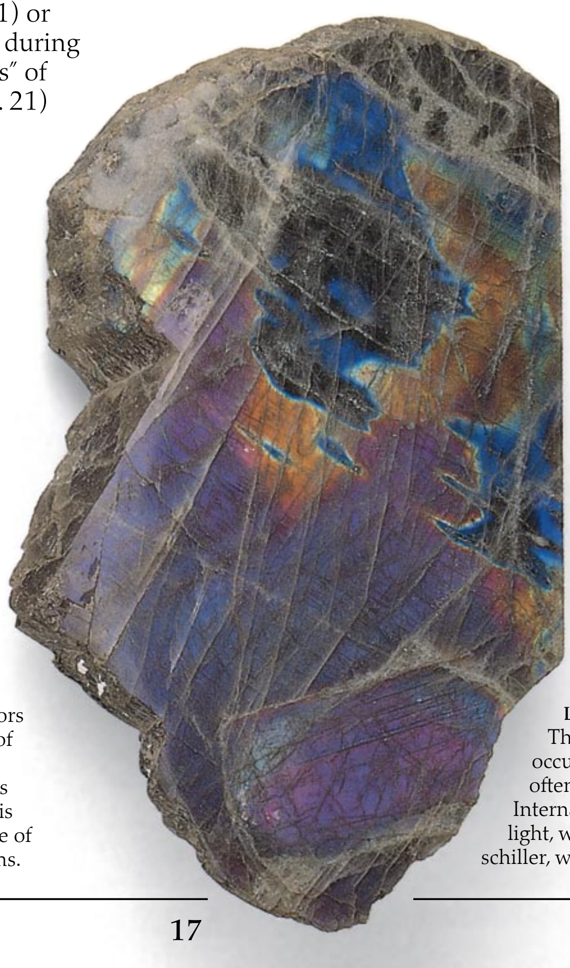

LABRADORITE
The feldspar mineral labradorite can occur as yellowish crystals, but more often it forms dull gray crystalline masses. Internal twinning causes interference of light, which gives the mineral a sheen, or schiller, with patches of different colors.

SALT
A space in the atomic structure of a crystal, caused by a missing atom, can form a color center. Coloration of common salt is thought to be caused by this.

Identification

Sherlock Holmes, the fictional master of criminal investigation and identification, looks for vital clues with the help of a hound

"What is it?" This is the first question to ask about a mineral, crystal, or gemstone. In order to identify a crystal it is necessary to test its properties. Most minerals have fixed or well-defined chemical compositions and a clearly identifiable crystal structure (pp. 14–15). These give the mineral a characteristic set of physical properties. Color (pp. 16–17), habit (pp. 22–23), cleavage (p. 15), and surface features can be studied using a hand lens, but in most cases this is not enough for positive identification. Other properties such as hardness and specific gravity (SG) can be studied using basic equipment, but more complex instruments are needed to fully investigate optical properties, atomic structure, and chemical composition.

SPOT THE DIFFERENCE
These two gemstones look almost identical in color, yet they are two different minerals: a yellow topaz (*left*), and a citrine (*right*).

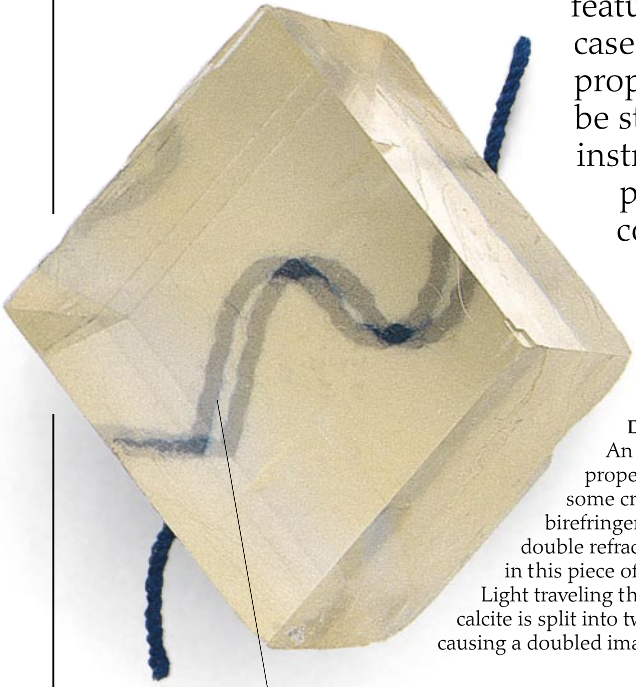

SEEING DOUBLE
An important property of some crystals is birefringence, or double refraction, as in this piece of calcite. Light traveling through the calcite is split into two rays, causing a doubled image.

Doubled image of wool seen through calcite

Orthoclase
SG = 2.6
Galena
SG = 7.4

WEIGHING IT UP
Specific gravity is a basic property. It is defined as the ratio of the weight of a substance compared to that of an equal volume of water If W1 = weight of specimen in air, and W2 = its weight in water, then W1 divided by W1-W2 = SG. The two crystals shown are of similar size but their SG differs considerably. This reflects the way the atoms are packed together.

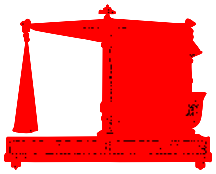

Chemical beam balance being used to determine specific gravity

Hardness

The property of hardness is dependent upon the strength of the forces holding atoms together in a solid. A scale of hardness on which all minerals can be placed was devised by F. Mohs in 1822. He selected 10 minerals as standards and arranged them in order of hardness so that one mineral could scratch only those below it on the scale. Intervals of hardness between the standard minerals are roughly equal except for that between corundum (9) and diamond (10).

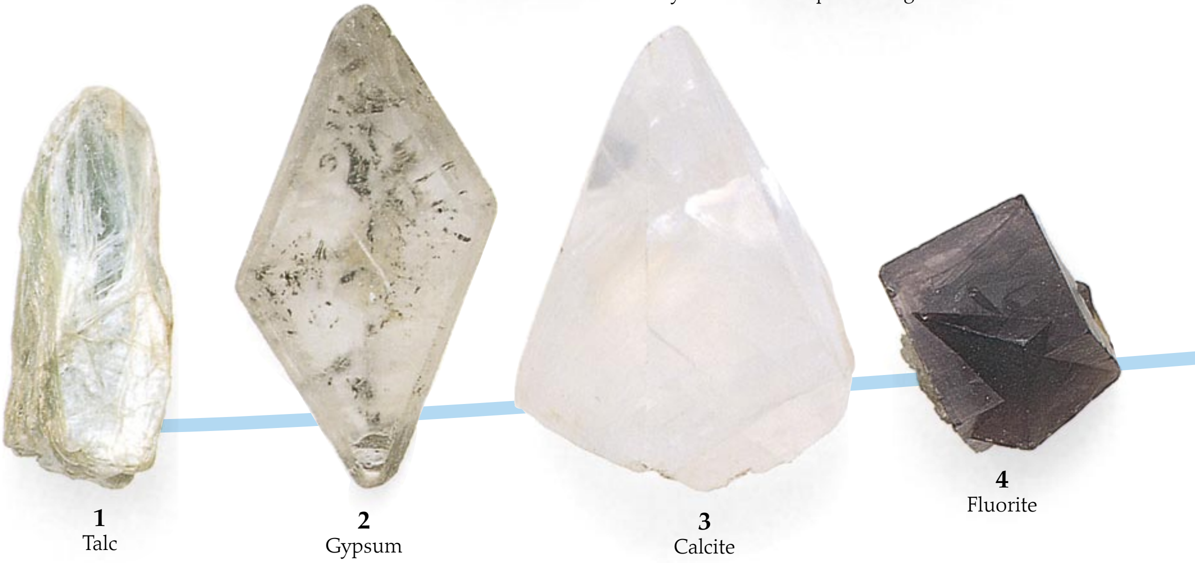

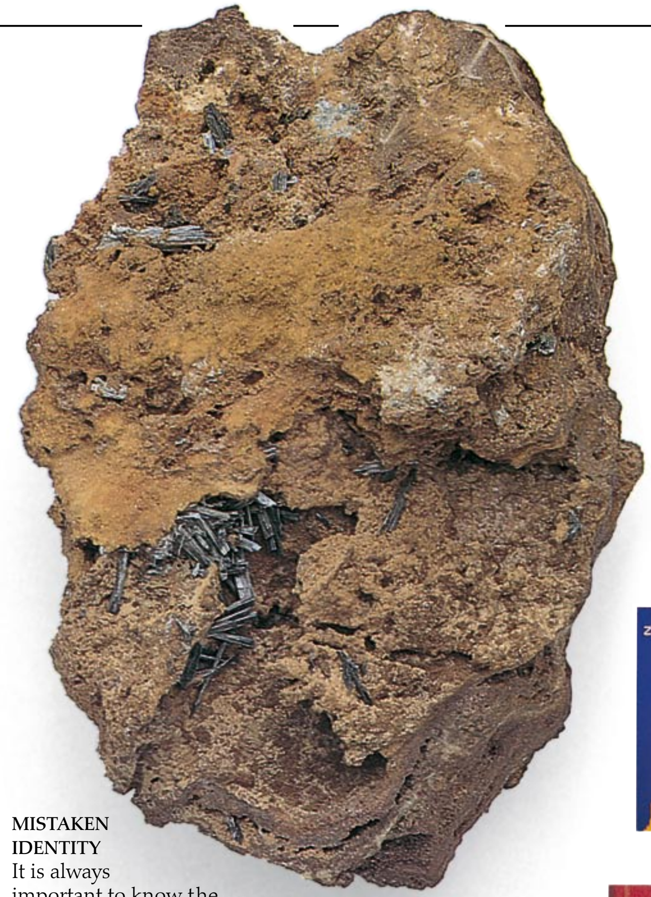

PROBING ABOUT
A modern technique called electron probe microanalysis was used to investigate the specimen on the left. In a scanning electron microscope (SEM) equipped with a special analysis system, a beam of electrons was focused on the specimen, producing a characteristic X-ray spectrum (below).

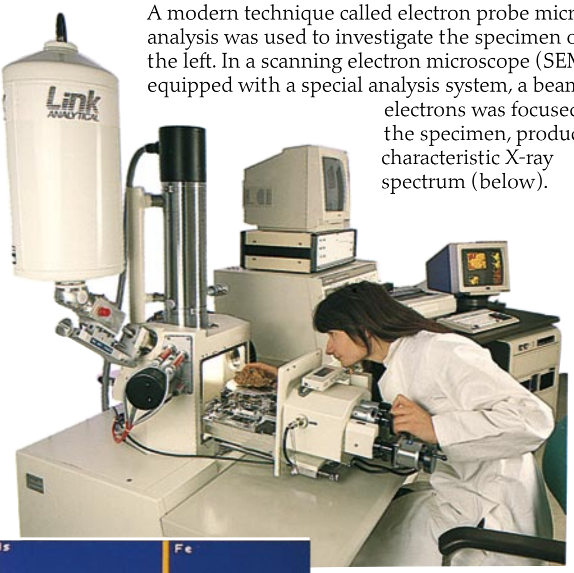

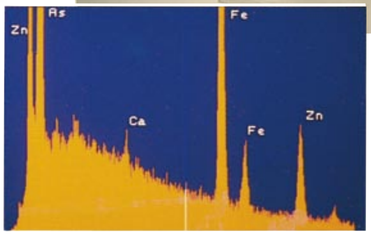

The X-ray spectrum showing large peaks for iron (Fe), arsenic (As), calcium (Ca), and zinc (Zn)

10
Diamond

MISTAKEN IDENTITY
It is always important to know the chemical composition of a crystal or mineral, and modern techniques can reveal some surprising results. These small blue-gray crystals on limonite were shown by X-ray methods to be the mineral symplesite (hydrated iron arsenate). However, further analysis showed that they unexpectedly contain some calcium and zinc as well.

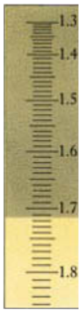

Spinel
RI: 1.71

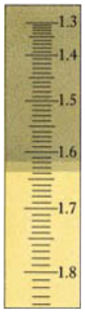

Tourmaline
RI: 1.62 and 1.64

SHADOW PLAY
Refractive index (RI) is a mineral's refracting ability – that is, its ability to bend a beam of light – and is useful in identification. It can be measured, along with birefringence, with a refractometer. A light is made to pass through the stone, and one or two shadow edges form on a scale depending on whether the gem is singly or doubly refractive. The position of the shadow gives the RI.

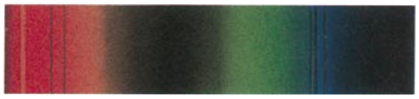

Ruby colored by chromium

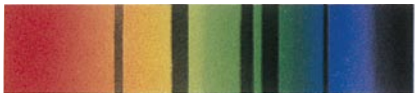

Almandine garnet colored by iron

ABSORBED IN STONE
A spectroscope is often used to distinguish between gemstones of a similar color. Light enters through a slit and separates into its spectrum of colors (p. 16). If a gemstone is put between the light source and the slit, dark bands appear in the spectrum, where wavelengths have been absorbed by the stone.

9
Corundum

Diamond

8
Topaz

7
Quartz

6
Orthoclase

5
Apatite

Sapphire

Chrysoberyl

Topaz

Garnet

Amethyst

Peridot

Opal

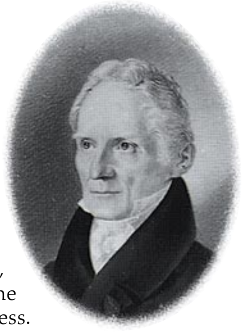

MOHS (1773-1839)
Friedrich Mohs was professor of mineralogy at Graz, and later Vienna, Austria. While at Graz he developed the scale of hardness.

Natural growth

CRYSTALS GROW as atoms arrange themselves, layer by layer, in a regular, three-dimensional network (pp. 14–15). They can form from a gas, liquid, or solid and usually start growing from a center or from a surface. Growth continues by the addition of similar material to the outer surfaces until the supply stops. It is rare to find a perfect crystal. Temperature, pressure, chemical conditions, and the amount of space all affect growth. It is estimated that in an hour, millions and millions of atoms arrange themselves layer by layer across a crystal face. With this number it is not surprising that defects occur.

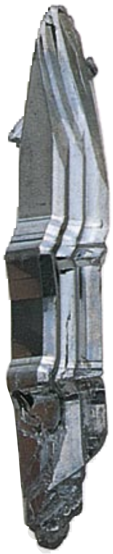

TWISTED AROUND
Crystals can be bent or twisted like this stibnite. This may be because they were bent by some outside force during growth.

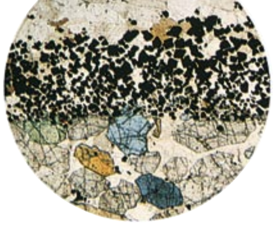

CRYSTAL LAYERS
This magnified image, called a photomicrograph, shows the layers of different crystals in a thin section of magmatic rock.

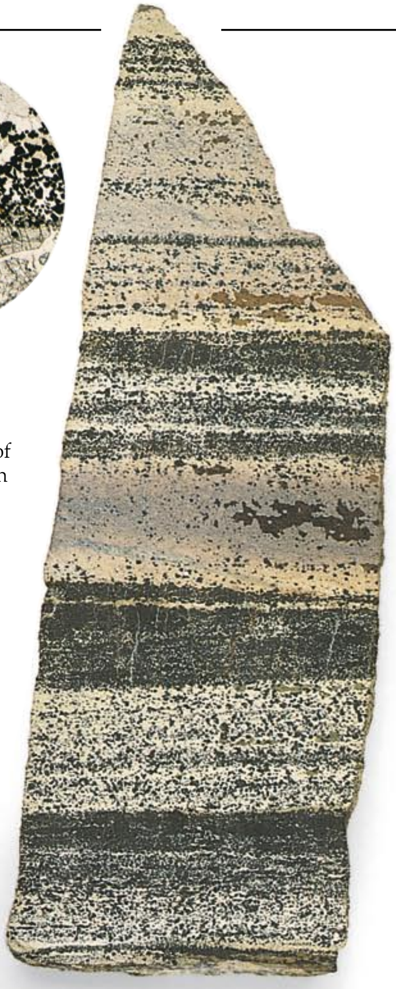

SETTLING DOWN
As magma (the molten rock below the earth's surface) cools, so crystals of various minerals form. Some magmatic rock forms in layers, as different rock-forming minerals settle and crystallize at different times.

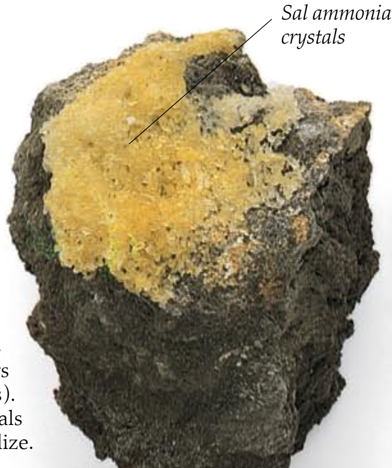

MINERAL SPRINGS
Hot watery solutions and gases containing minerals, such as sal ammoniac (ammonium chloride), sometimes reach the earth's surface through hot springs and fumaroles (gas vents). Here, the minerals may crystallize.

IN THE POCKET
Holes in rocks often provide space in which crystals can grow. Cavities containing fine gemquality crystals are known as gem pockets. This gem pocket at Mt. Mica, Maine, was discovered in 1979.

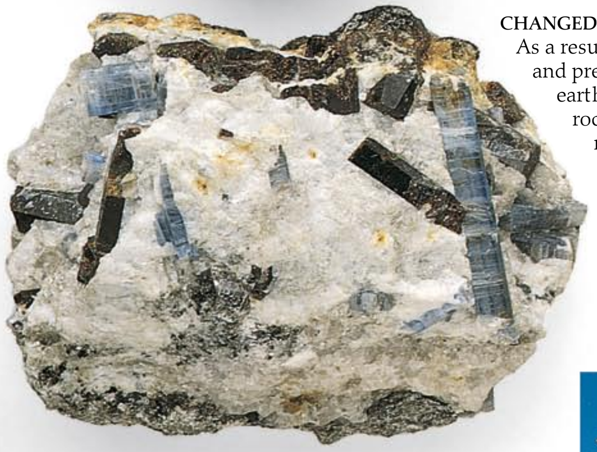

CHANGED BY FORCE
As a result of the high temperatures and pressures deep within the earth's crust, minerals in solid rock can recrystallize, and new minerals form. This process is known as metamorphism. The blue kyanite and brown staurolite crystals in this specimen have been formed in this way.

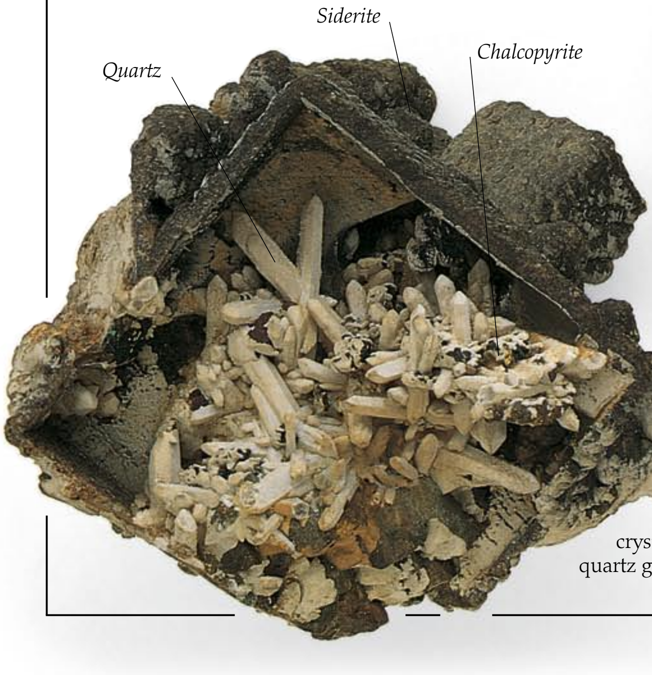

TAKING SHAPE
Many minerals crystallize from watery solutions. We only see the final product but can often work out a sequence of events. In this specimen, a fluorite crystal grew first, and was coated with siderite. The fluorite was later dissolved and removed, but the coating of siderite kept the characteristic cubic shape of the fluorite crystal. Lastly, crystals of chalcopyrite and quartz grew inside the hollow cube.

BUILDING BLOCKS
Skyscrapers are built in a similar way to crystals – by adding layer upon layer to the same symmetrical shape.

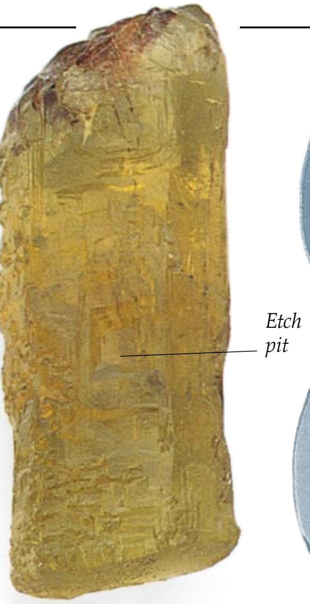

BERYL ETCHING
Solutions or hot gases may dissolve the surface of certain crystals after growth, as in this beryl. Regularly shaped hollows known as etch pits are formed. Their shape is related to the internal atomic structure.

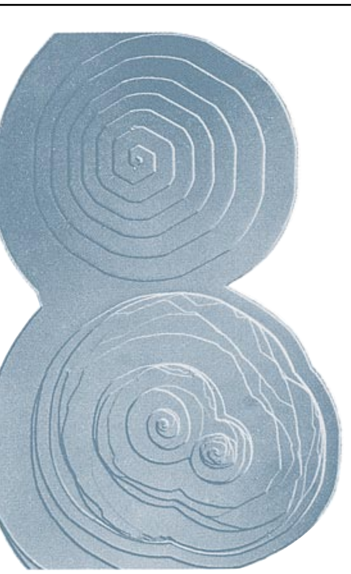

SPIRALING AROUND
Crystal faces are rarely flat, due to a variety of growth defects. This magnified image of the surface of a crystal shows the atoms forming a continuous spiral, instead of layers across the crystal face.

Twinning

During crystallization, two crystals of the same mineral may develop in such a way that they are joined at a common crystallographic plane. Such crystals are known as twinned crystals. The apparent line of contact between the two parts is known as the twin plane.

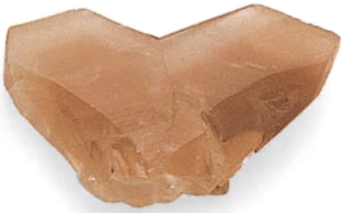

BUTTERFLY TWINS
This simple type of twin is known as a butterfly contact twin crystal because of its resemblance to butterfly wings. This example is calcite.

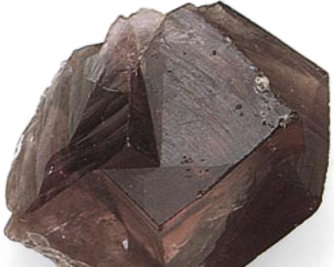

GROWING UP TOGETHER
When the two parts of twin crystals grow into each other, they are known as penetration twins. The example shown is a twin of purple fluorite.

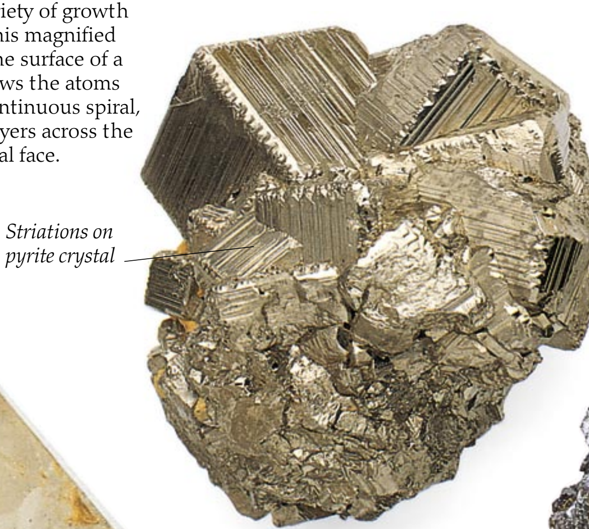

FORM COMPETITION
Many crystals have parallel lines called striations running along or across their faces. These are usually caused when two forms (p. 13) try to grow at the same time.

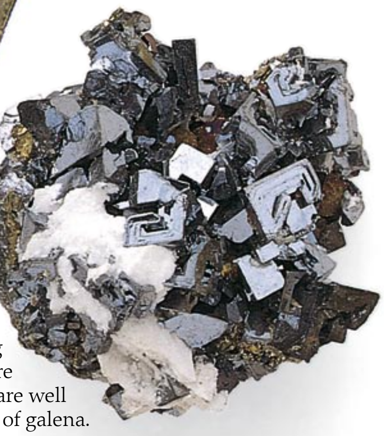

AT THE HOP
Some crystals tend to build up more quickly along the edges of the faces than at the centers, producing cavities in the faces. These are known as hopper crystals and are well illustrated here by crystals of galena.

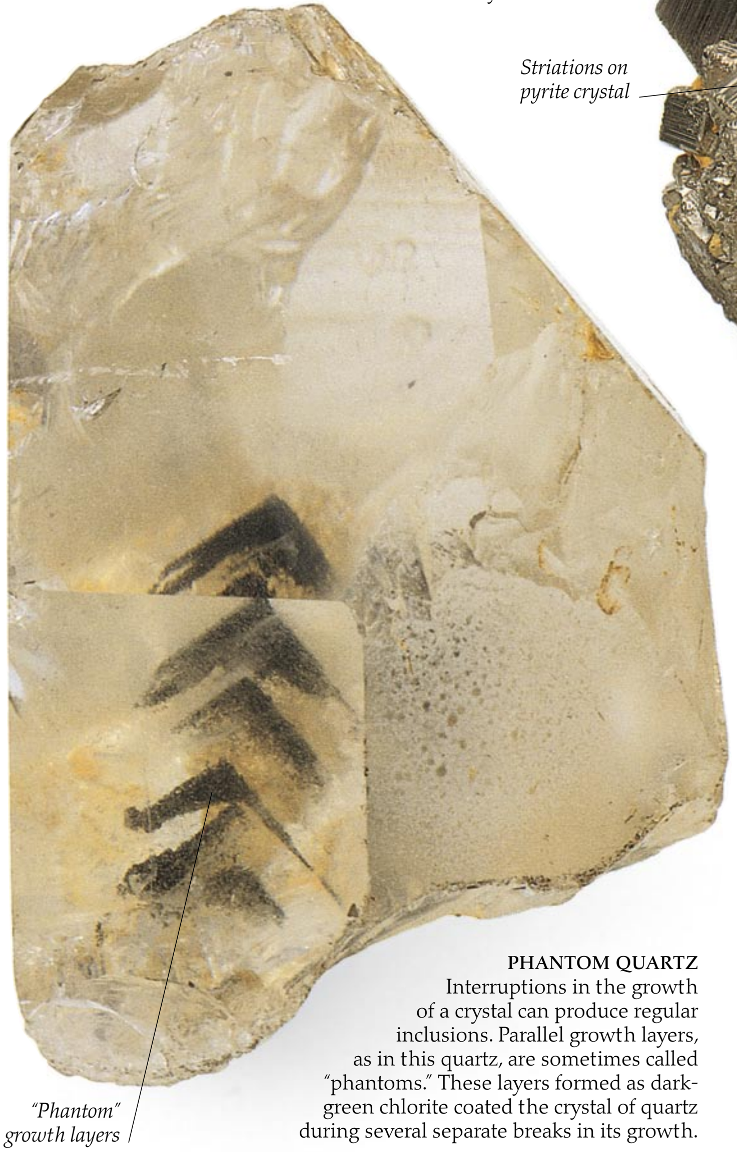

PHANTOM QUARTZ
Interruptions in the growth of a crystal can produce regular inclusions. Parallel growth layers, as in this quartz, are sometimes called "phantoms." These layers formed as dark-green chlorite coated the crystal of quartz during several separate breaks in its growth.

CRYSTAL ENCLOSURE
During growth, a crystal may enclose crystals of other minerals, commonly hematite, chlorite, and tourmaline. These are known as inclusions.

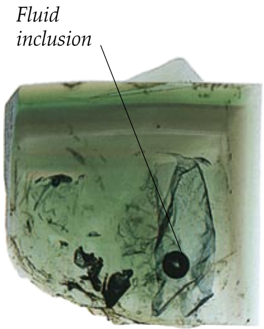

A fluorite crystal containing inclusions of ancient mineral-forming fluids

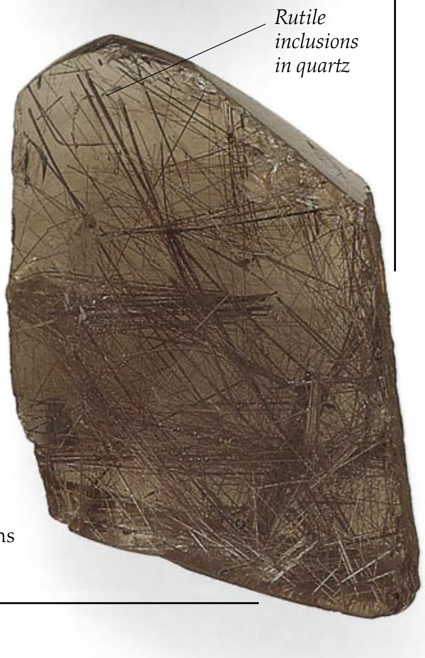

Good habits

The general shape of crystals is called their habit and is an important part of crystallography. Crystal habit is useful in identification and in well-formed crystals may be so characteristic of a particular mineral that no other feature is needed to identify it. The forms (p. 13) or group of forms that are developed by an individual crystal are often what give it a particular habit. As crystals grow, some faces develop more than others, and it is their relative sizes that create different shapes. Most minerals tend to occur in groups of many crystals rather than as single crystals and rarely show fine crystal shapes. These are called aggregates.

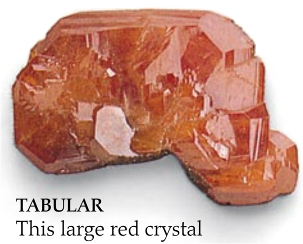

TABULAR
This large red crystal of wulfenite comes from the Red Cloud mine in Arizona. Its habit is known as tabular. Such crystals are often extremely thin. Wulfenite belongs to the tetragonal crystal system.

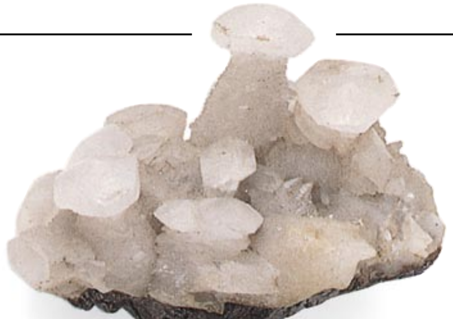

TWO FORMS
These "mushrooms" show two forms of calcite crystals: The "stems" are scalenohedrons and have eight of twelve triangular faces. The "caps" are formed by rhombohedra in parallel position. This group comes from Cumbria, England.

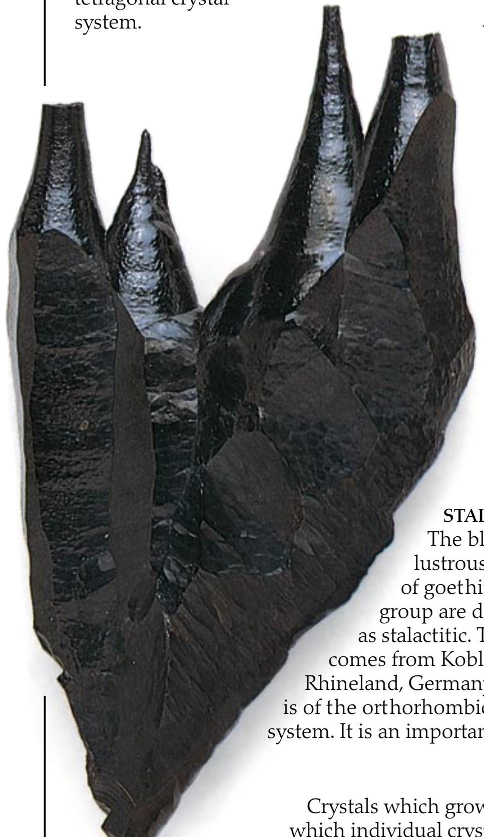

STALACTITIC
The black, lustrous aggregates of goethite in this group are described as stalactitic. The group comes from Koblenz, Rhineland, Germany. Goethite is of the orthorhombic crystal system. It is an important iron ore.

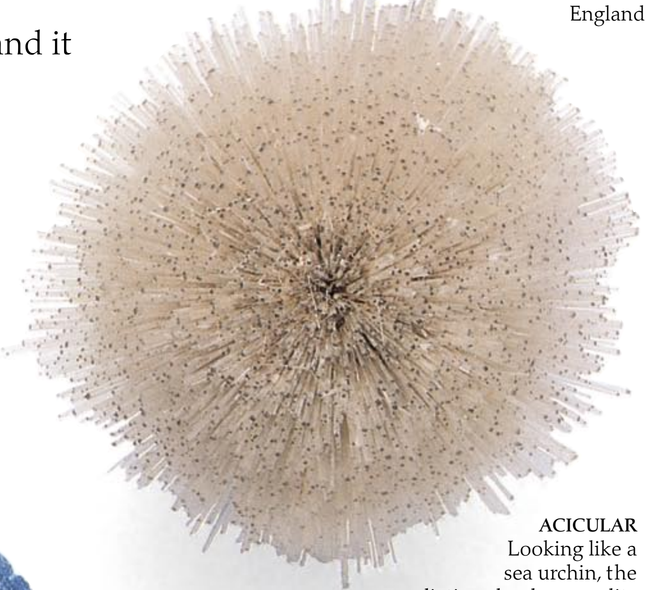

ACICULAR
Looking like a sea urchin, the radiating slender mesolite crystals in this aggregate are described as acicular, meaning needle-like. They are very fragile and, like needles, can pierce your skin. This group comes from Bombay, India.

MASSIVE
Crystals which grow in a mass, in which individual crystals cannot be clearly seen, are described as massive. Dumortierite is a rare mineral which is usually massive like this piece from Bahia, Brazil.

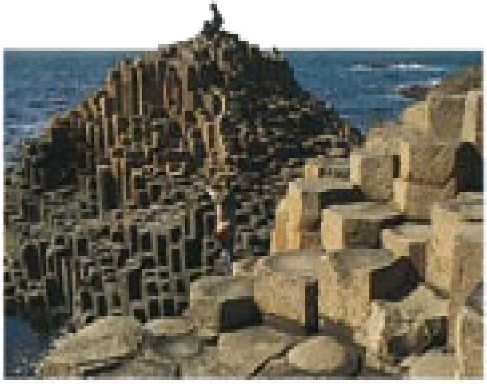

CRYSTAL-SHAPED
The Giant's Causeway near Portrush in County Antrim, Northern Ireland, looks like a collection of hexagonal crystals. However, the phenomenon is not crystal growth but jointing due to contraction as the basaltic lava cooled.

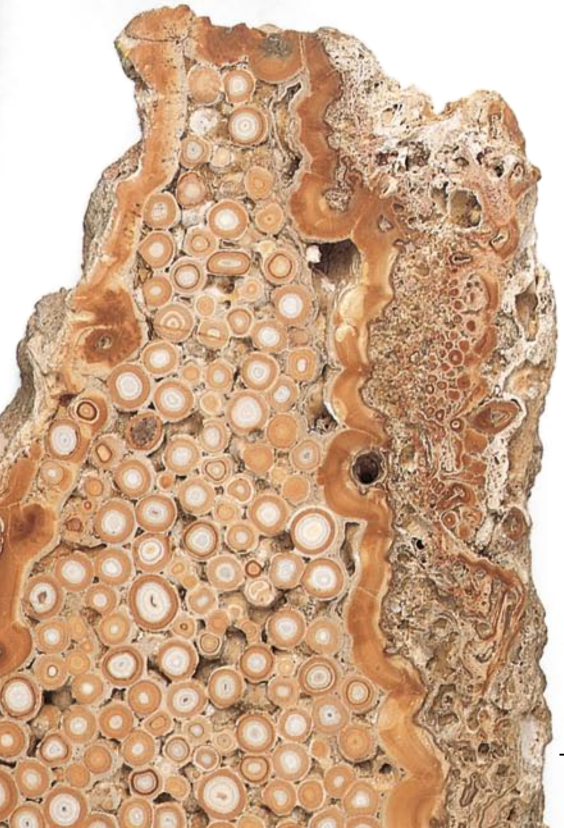

PISOLITIC
This polished slab of limestone from Czechoslovakia is described as pisolitic. Pisolites are round pea-sized aggregates of crystals built of concentric layers, in this case of calcium carbonate.

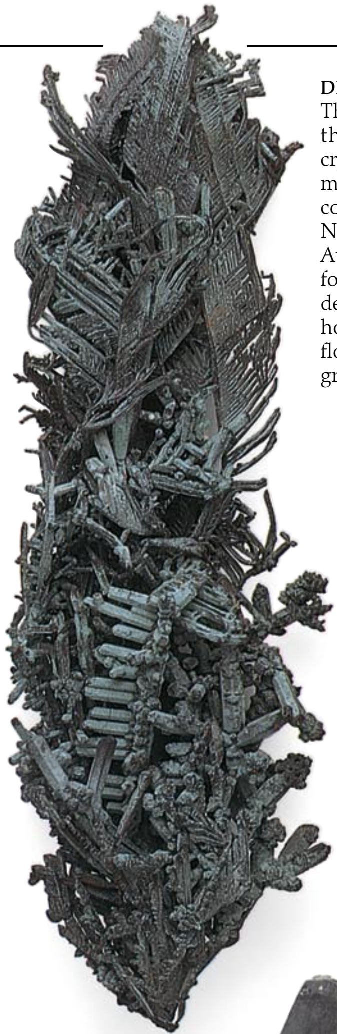

DENDRITIC
The term used to describe the habit of these copper crystals is dendritic, meaning tree-like. They come from Broken Hill, New South Wales, Australia. Copper often forms in hydrothermal deposits (p. 24), filling holes in some basaltic lava flows, but is also found as grains in sandstones.

PRISMATIC
Beryl crystals are mostly found in granite pegmatites (p. 25) and can grow to be very large. Those illustrated are prismatic – they are longer in one direction than the other. They were found in 1930 in a quarry in Maine and were over 30 ft (9 m) long.

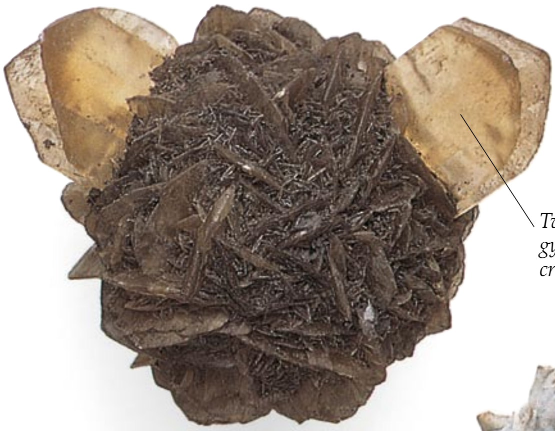

Twinned gypsum crystal

LENTICULAR
Twinned (p. 21) clear crystals of gypsum form the "ears" on this mass of lenticular crystals from Winnipeg, Canada. Lenticular means shaped like a lentil or lens, from the Latin *lenticula*, a lentil.

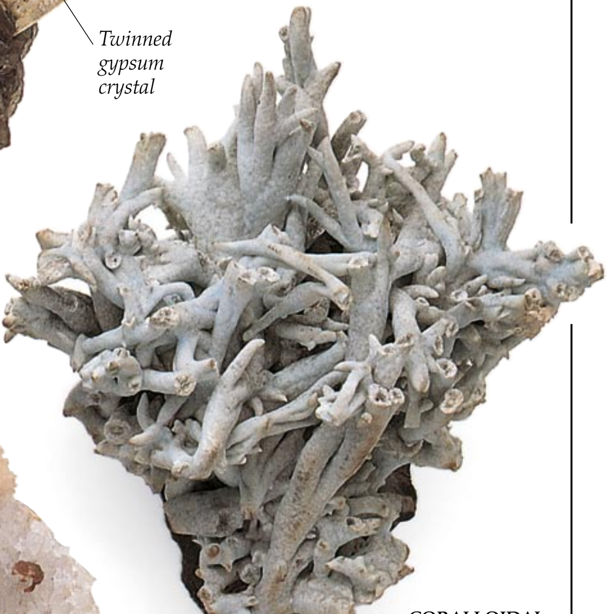

CORALLOIDAL
Aggregated crystals that look like coral are said to have a coralloidal habit. This mass of pale-green aragonite crystals came from Eisenberg, Styria, Austria.

Bladed hornblende crystal

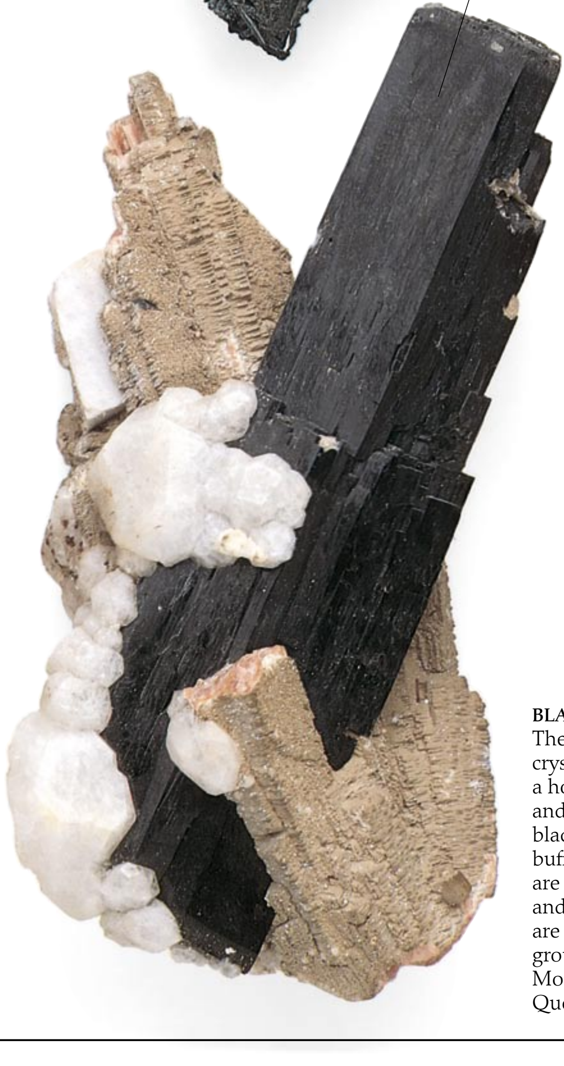

Globular caldite crystal aggregate

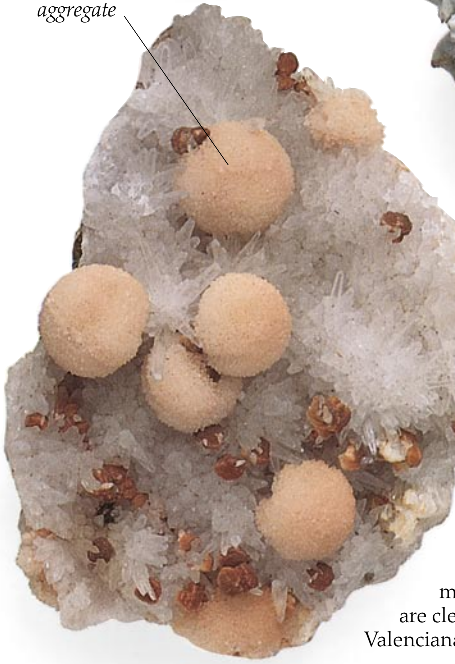

GLOBULAR
These aggregated crystals of calcite look a bit like scoops of ice cream and are described as globular, meaning spherical. The other crystals are clear quartz, and the group came from Valenciana mine, Guanajuato, Mexico.

BLADED
The prismatic black crystal in this group is a hornblende crystal and an example of a bladed crystal. The buff-colored crystals are prismatic serandite and the white crystals are analcime. The group was found at Mont-St.-Hilaire, Quebec, Canada.

QUARTZ IN A CAVE
Crystal growth is influenced by the physical and chemical conditions at the time. Many good crystals grow in cavities which can vary in size from small potato stones (p. 7) to huge caves, as shown in this 19th-century impression of a quartz grotto.

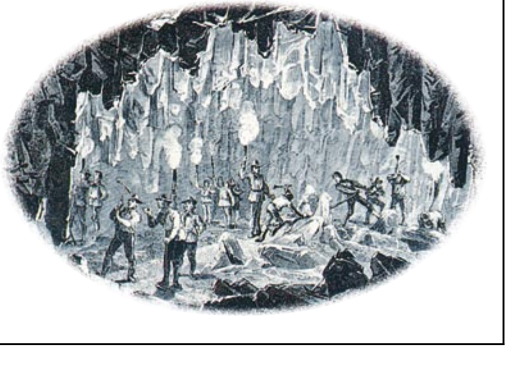

Discovery – recovery

The search for mineral deposits including metals and gemstones has been going on since prehistoric times. Some minerals, such as copper, occur in great quantity; others, such as silver, gold, and diamond, are found in much smaller quantities but fetch higher prices. If mining is to be profitable, large quantities of the mineral must occur in one area and be relatively easy to extract, either by surface quarrying, panning, or dredging, or, if necessary, by deep mining. Minerals from which useful metals such as copper, iron, and tin are extracted are called ores.

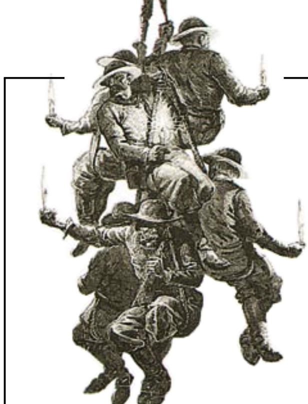

19th-century engraving or miners descending the shaft at Wieliczka salt mine, Poland

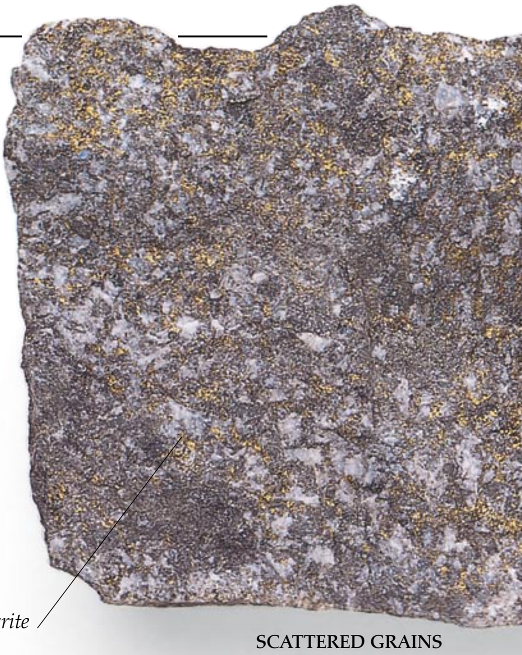

SCATTERED GRAINS
Rocks of less than one percent ore are worked today by open-pit quarrying. The ore, such as this chalcopyrite copper ore, occurs as small grains scattered through the rock body. The whole rock has to be worked, a huge amount of gangue, or waste, is produced, and an enormous hole is left.

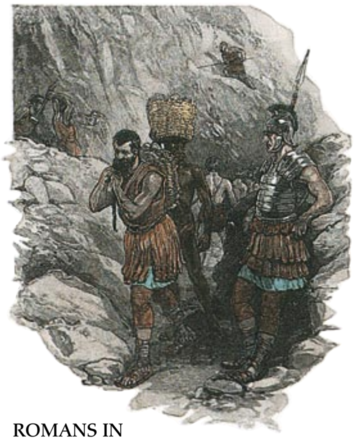

ROMANS IN CORNWALL
The Romans knew of the rich tin deposits in Cornwall, England. Mining techniques have improved since then, but the ore still has to be crushed and separated from the gangue minerals and then refined.

R.K.MICHELL
& Co
MARAZION
REFINED

Ingot of refined Cornish tin, produced in 1860

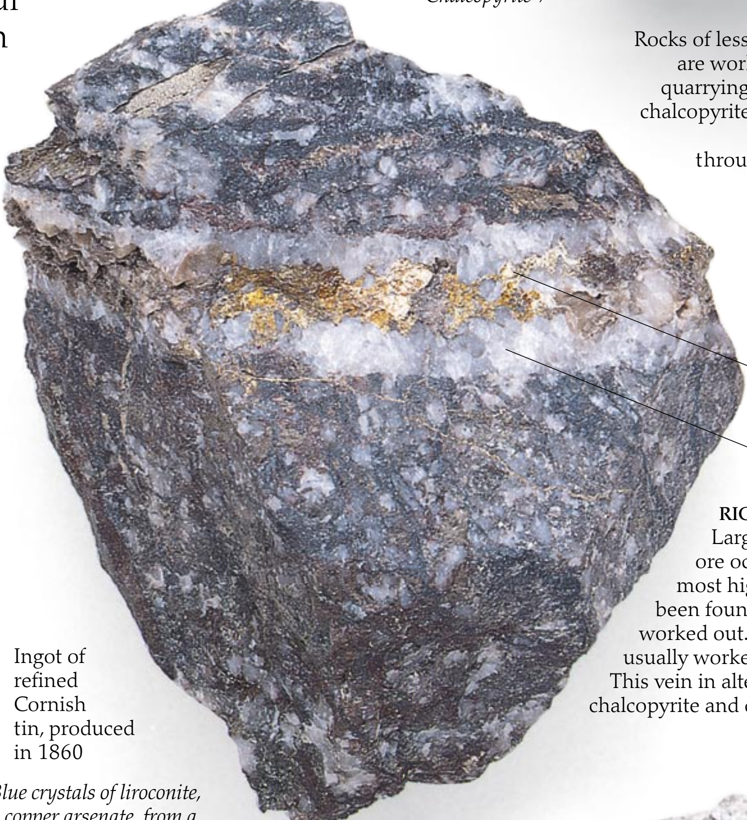

RICH VEIN
Larger concentrations of ore occur in veins, but most high-grade ores have been found and in many cases worked out. Ores in veins are usually worked by deep mining. This vein in altered granite contains chalcopyrite and quartz.

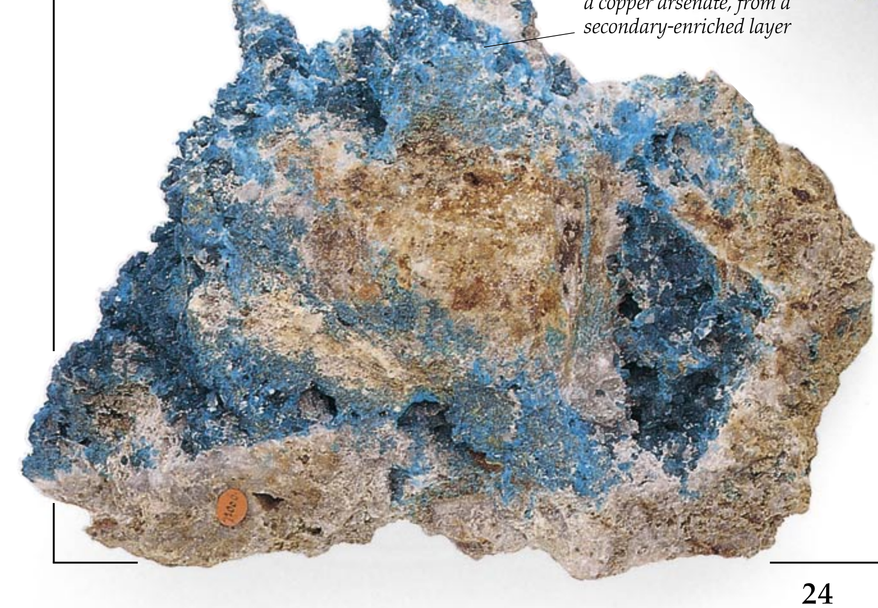

Blue crystals of liroconite, a copper arsenate, from a secondary-enriched layer

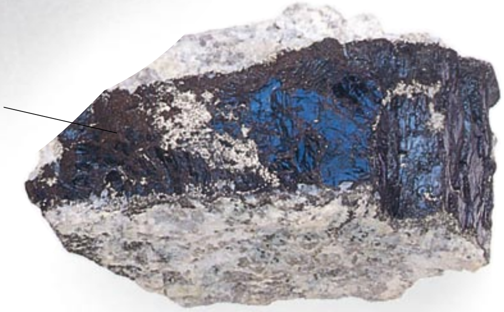

Vein of covellite, a copper sulfide, from a secondary sulfide enrichment layer

GRADUAL IMPROVEMENT
The natural process of "secondary alteration and enrichment" can improve relatively low-grade ores to higher concentrations. Groundwaters filter down through the upper layers of rock and carry elements downward. These are redeposited in lower layers which are thus enriched. Enriched layers in copper deposits may contain azurite, malachite, and sometimes liroconite, or sulfide minerals such as bornite, chalcocite, and covellite.

SWIRLING WATERS
Panning is a simple method of separating minerals. Light gangue minerals are washed away by the swirling action of water in a metal or wood pan, leaving the wanted minerals behind. This technique is often used to sort out gem-rich river gravels in areas such as Myanmar (formerly Burma) and Thailand.

Panning for gold in the Irrawaddy River, Myanmar. The prospector looks for the glint of gold grains within the waste.

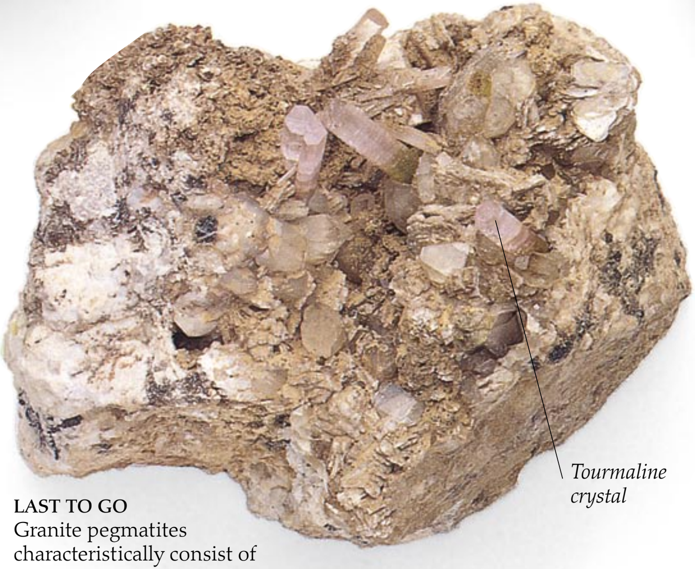

LAST TO GO
Granite pegmatites characteristically consist of large crystals and are the source of many fine gems, including tourmaline (p. 43), topaz (p. 42), and beryl (pp. 38–39). They are formed by the crystallization of the last fluids left after most of the granite has solidified.

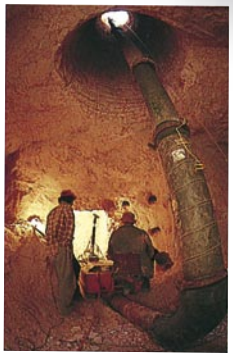

DOWN UNDER
Much mining activity takes place underground. This is the Coober Pedy opal mine in South Australia – a source of fine white opals (pp. 40–41).

SMALLER THAN SOME
These beryl crystals measure about 8 x 6 in (20 x 14 cm) but are small compared to some crystals found in pegmatites.

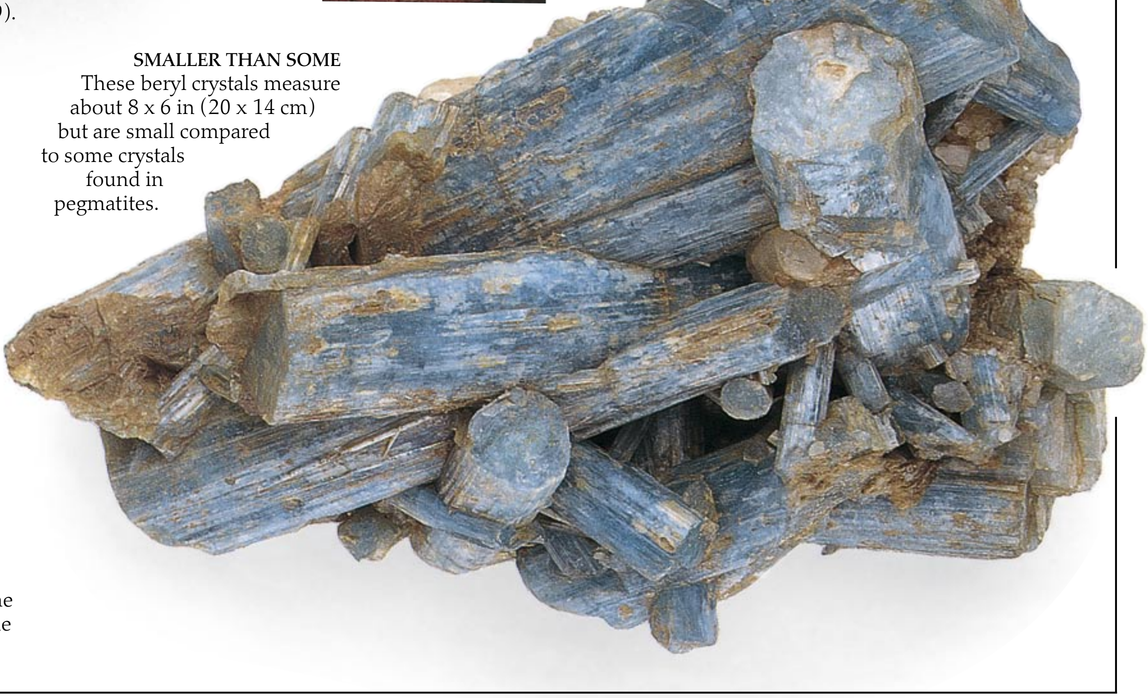

ON THE SURFACE
The Argyle mine in Western Australia is the largest diamond producer in the world. The diamonds are mined by surface-quarrying.

MELTDOWN
Natural bismuth does occur, but artificial crystal groups, like this one, are often produced by melting and then cooling the metal. Bismuth is used in solders, electric fuses, and pigments.

Growing from seed

SCIENTISTS HAVE TRIED TO MAKE CRYSTALS like those found in the earth's crust for well over a century. Natural crystals often contain impurities or are flawed in some way (pp. 20–21), but synthetic ones can be made flawless. They can also be made to grow a particular shape and size for specific needs. In recent times a range of artificial crystals has become important to modern technology. Grown crystals are built into almost every electronic or optical device made today. The need for a huge amount of perfect crystals has led to more and more synthetic crystals being made, and it could be said that future developments in electronics will depend on the development of crystal-growing techniques.

DRAWN OUT
Pure silicon does not occur naturally, so crystals are made artificially for a variety of uses (p. 28). Quartz sand is heated to produce nearly pure silicon. In one process a seed crystal on the end of a rotating rod is dipped into the melt and slowly removed; this is known as "drawing a crystal."

IN A FLUX
Many emeralds are produced by the flux-fusion technique. A powder made of the components of emerald is heated in a crucible with a solid known as a flux. The flux melts, the powder dissolves, and the mixture is left to cool and form crystals. This method is extremely slow. It takes several months for a crystal to grow.

Cut synthetic emerald

Synthetic emerald crystal

VOYAGE OF DISCOVERY
Crystal growing is important enough for experiments to be done in space. Here, astronaut George Nelson is photographing a protein-crystal-growth experiment on board the space shuttle *Discovery* in 1988.

Melt technique

Excellent crystals may be grown by slow cooling or evaporation of a supersaturated solution (no more will dissolve) of a salt such as halite, alum, or ammonium dihydrogen phosphate (ADP). In the experiment shown, powdered ADP containing a small chrome-alum impurity has been completely dissolved in boiling water and then cooled.

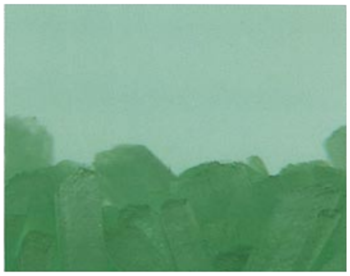

The liquid cools rapidly. Stubby, cloudy prismatic (p. 23) crystals form.

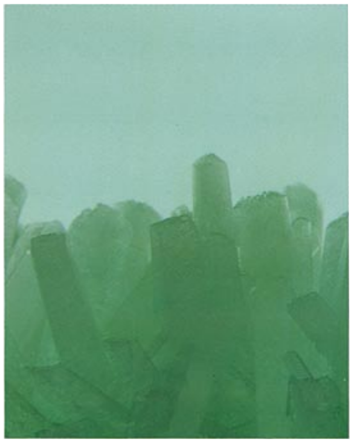

The crystals grow more slowly, allowing them to become clearer.

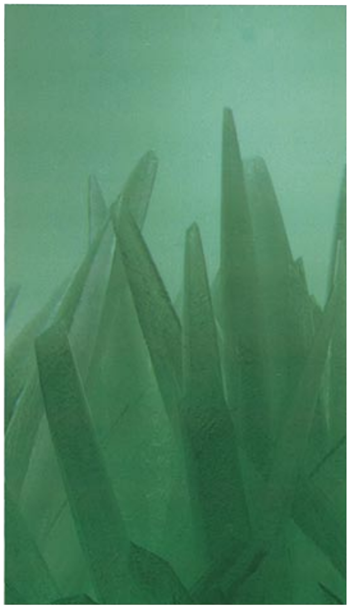

At room temperature, crystals still grow slowly due to evaporation.

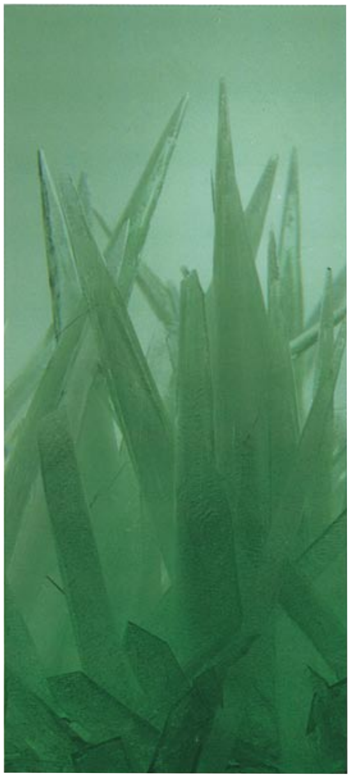

Cooling stops, but evaporation continues. The crystals slowly grow.

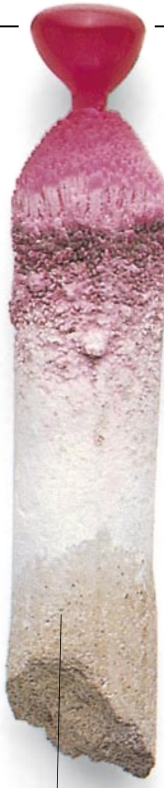

FIRE BOULES
The flame-fusion technique was pefected by French mineral expert A. Verneuil in about 1900. It is used to make spinel (p. 46), rutile (p. 57), and corundum (pp. 36–37). Powdered material is fed through a flame to fuse into liquid drops which drip onto a support. By gradually pulling the support from the heat, a single crystal, or boule, is formed.

Support for growing crystal

Synthetic sapphire boule

Two halves of synthetic ruby boule

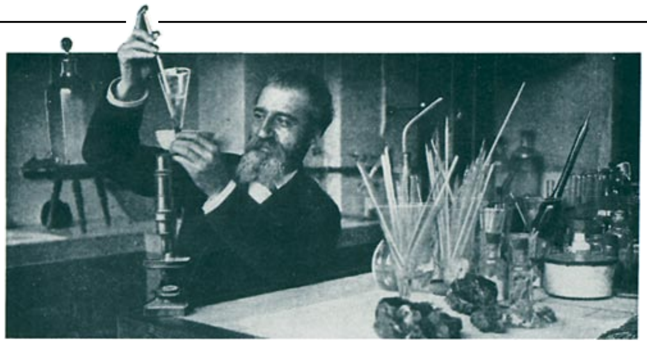

HENRI MOISSAN (1852-1907)
French chemist Henri Moissan tried to produce artificial diamonds in iron crucibles at the Edison workshops in Paris.

EUREKA!
In 1970, the General Electric Company announced the laboratory creation of gem-quality diamonds, two of which are shown here.

Synthetic rubies produced in a crucible

GROWN IN SIZE
The French chemist Frémy was the first to grow gem-quality crystals of a reasonable size, in 1877. He discovered a method of making rubies by melting the necessary materials together and fusing them in a porcelain crucible at very high temperatures.

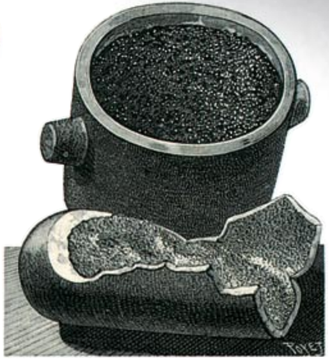

1890 crucible containing a mass of small gemstones

ABRASIVE CHARACTER
The artificial material carborundum (silicon carbide) is produced by the fusion of coke and sand heated in electrically fired furnaces. It is nearly as hard as diamond – 9.5 on Mohs' hardness scale (pp. 18–19) – and is therefore mostly used as an abrasive.

Hexagonal carborundum crystal

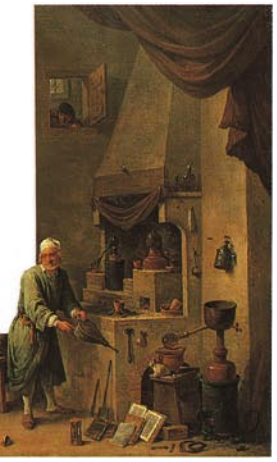

GOLD FEVER
Over the centuries many people have tried to find a way to change nonprecious metals into gold. A complicated process if this detail of *The Alchemist at Work* by David Teniers the Elder (1582-1649) is anything to go by.

Crystals at work

CRYSTALS PLAY AN IMPORTANT PART in this age of rapid technological and social change. Although the basic understanding of crystals was developed before the 20th century, it was only in the latter part of the 20th century that crystal technology became so important. Crystals are now used in control circuits, machines, electronics, communications, industrial tools, and medicine. We also use crystals when shopping – in credit cards. From the crystal laboratory (pp. 26–27) has come the silicon chip, ruby laser rods, and the many forms of diamonds for tools. New crystals are continually being developed for specific purposes.

DIAMOND WINDOW
The properties of diamond have led to it being used in space where it has to withstand extreme conditions. Diamond was used in this window for an infrared radiometer experiment on the Pioneer Venus probe. It had to withstand a temperature of 840°F (250°C) near the surface of the planet Venus.

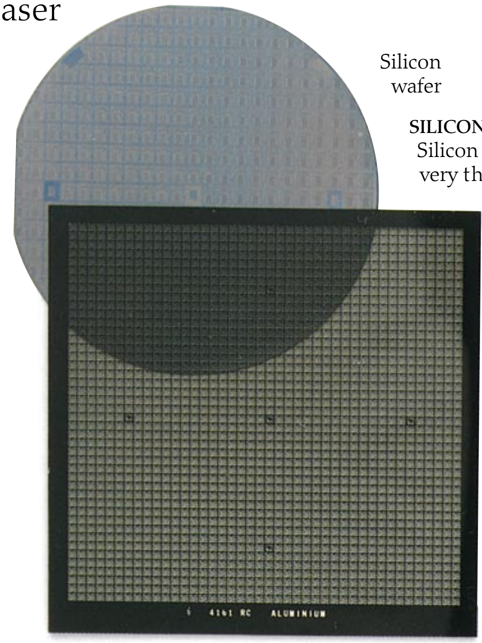

Silicon wafer

SILICON SLICE
Silicon chips are made from very thin slices called wafers cut from artficial crystals of pure silicon (p. 26). The wafers are etched with electronic circuits. one for each chip. The circuit patterns are transferred on to a wafer from a piece of film called a matrix.

Silicon chip matrix

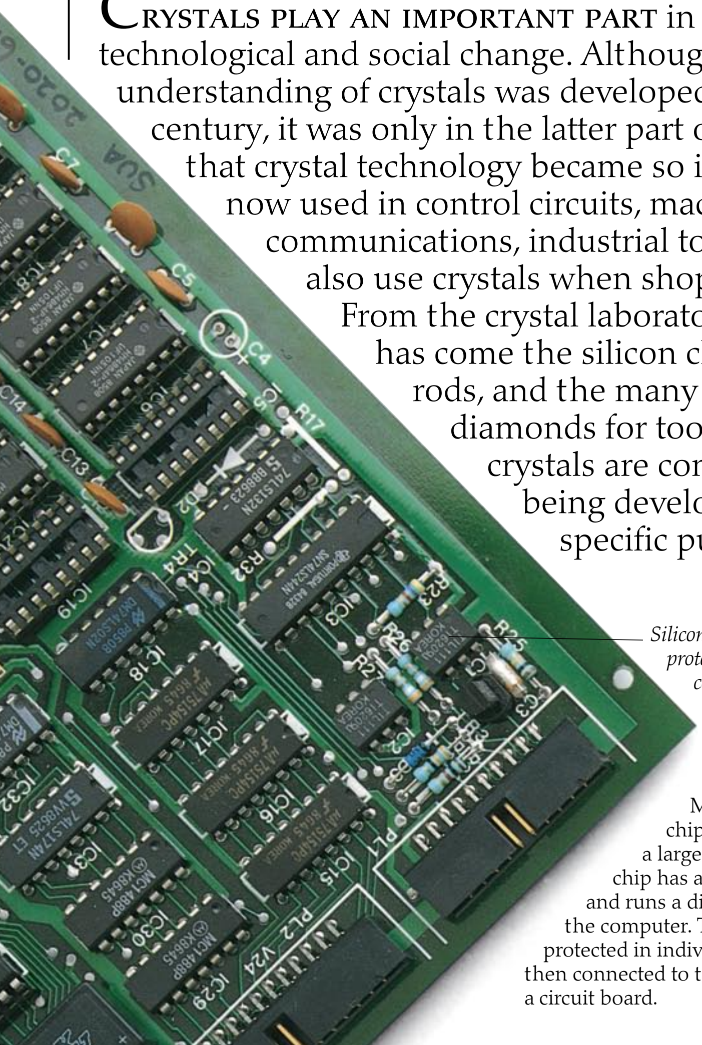

Silicon chip in protective covering

CIRCUIT BOARD
Many different chips are needed in a large computer. Each chip has a different circuit and runs a different part of the computer. The chips are protected in individual cases, then connected to the others on a circuit board.

Location of silicon chip

SMART CARDS
There is a tiny built-in mini computer on a silicon chip in each of these "smart cards." When the card is inserted into a reading device, the chip makes contact with an electrical connector that reads the information on the card. Smart cards are used for identity cards, driver's licenses, and as tickets on public transportation.

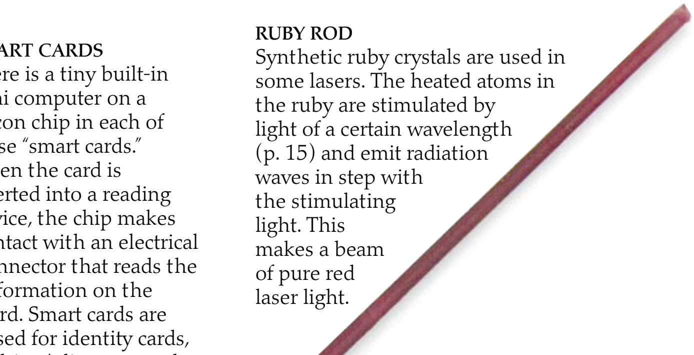

RUBY ROD
Synthetic ruby crystals are used in some lasers. The heated atoms in the ruby are stimulated by light of a certain wavelength (p. 15) and emit radiation waves in step with the stimulating light. This makes a beam of pure red laser light.

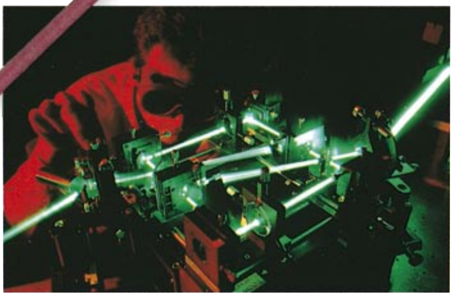

LASERS
This scientist is experimenting with laser beams. Laser beams can be focused to very small points, generating intense heat. This is put to use in welding, drilling, and surgery.

Diamond tools

Diamonds are used in a vast number of jobs mainly because they are so hard. They are used in sawing, drilling, grinding, and polishing – from quarrying stone to performing delicate eye surgery – and come in a large range of sizes, shapes, and strengths. Natural and synthetic diamonds are used, but more than 80 percent of industrial diamonds are synthetic.

A surgeon using a diamond-bladed scalpel in delicate eye surgery

DIAMOND SCALPEL
As well as being hard, diamond does not corrode. This property is one reason why diamonds are used in surgery. This surgical scalpel contains a blade made from a natural diamond.

Diamond blade

DRILL BITS
Diamond-tipped drills are used for drilling all types of rock. They are used for drilling oil wells and in prospecting for metals and minerals. Some bits contain diamonds set in the surface. The diamonds are different shapes for different purposes. Other bits are covered with tiny pieces of diamond grit, or abrasive.

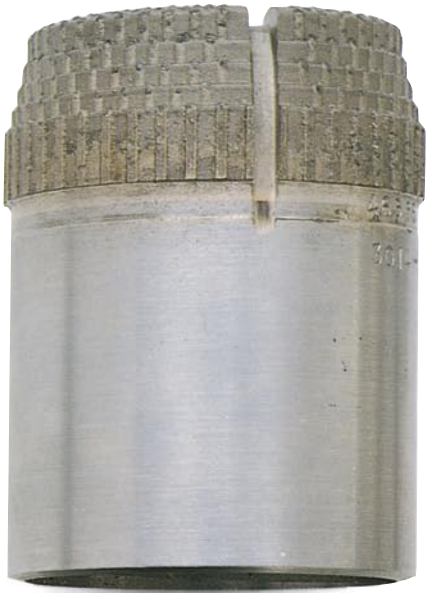

Drill bit containing surface-set natural diamonds

Drill bit covered with synthetic diamond abrasive

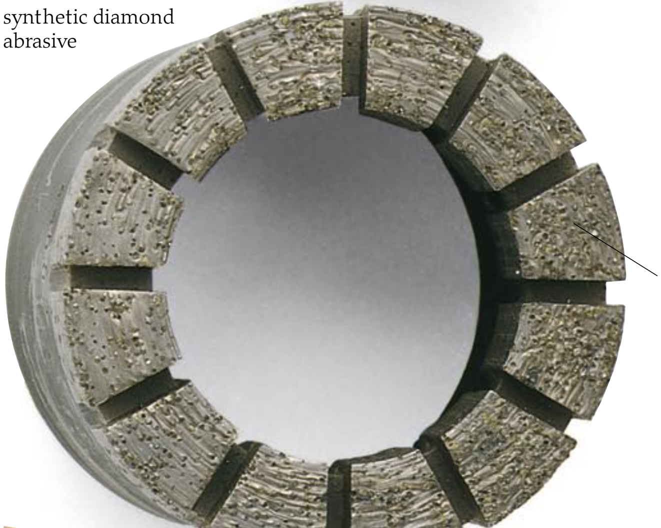

Cutting segment containing synthetic diamond abrasive

DIAMOND GRIT
Grit and powders are made from synthetic diamonds or poor-quality natural stones. They are used for polishing and grinding.

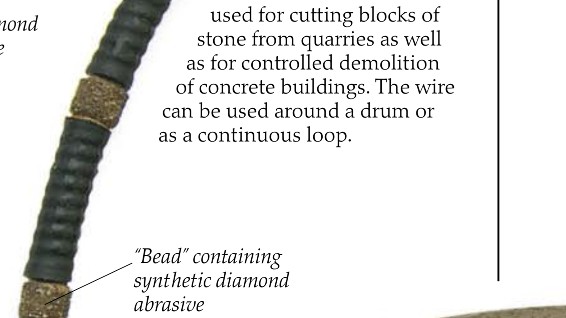

DIAMOND WIRE
Cutting with a diamond wire keeps the loss of material to a minimum. Wires can be used for cutting blocks of stone from quarries as well as for controlled demolition of concrete buildings. The wire can be used around a drum or as a continuous loop.

"Bead" containing synthetic diamond abrasive

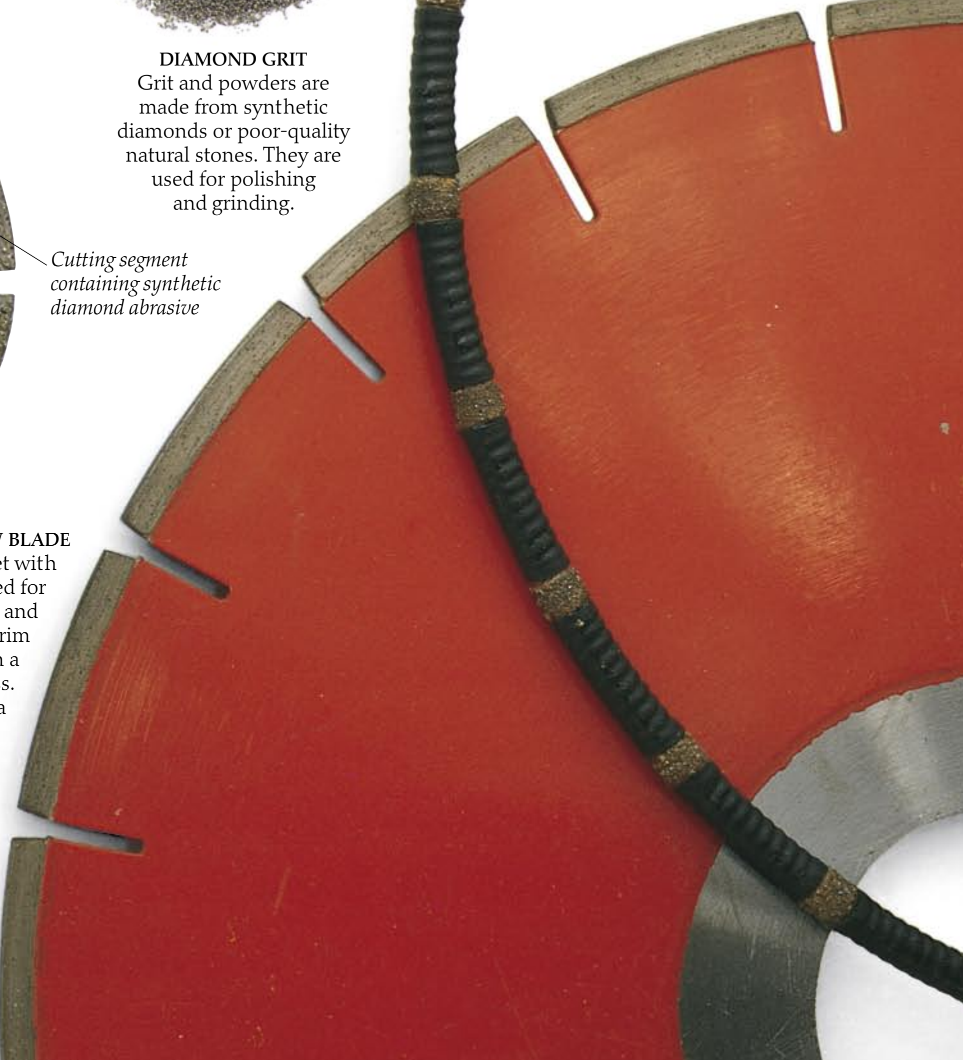

SAW BLADE
Saws set with diamonds are used for cutting glass, ceramics, and rocks. The blades have a rim of industrial diamonds in a "carrier" such as brass. This rim is bonded to a steel disk. As the blade cuts, the carrier wears away rapidly and exposes new diamonds.

Cutting an opening for a window in brickwork using a diamond saw

Good vibrations

QUARTZ IS ONE OF THE MOST COMMON MINERALS in the earth's crust. It is widely distributed as veins (p. 24) and is associated with major mineral deposits. It is one of the chief materials in granite and is also the main component of sand and sandstone. As quartzite and sandstone, it is used extensively for building and in the manufacture of glass and ceramics. One of the most interesting properties of quartz crystals is piezoelectricity. This can be used to measure pressure, and quartz crystal oscillators provide fixed, very stable frequency control in radios and televisions (an oscillator is something that vibrates). The piezoelectric effect of crystals is also used in gas igniters. When a crystal is "squeezed," an electric charge is produced as a spark which lights the gas. Because it so often forms perfect crystals, quartz is also used in crystal healing.

PAST FAVORITE
Quartz crystals from Brazil were important material for electronics before synthetic crystals were grown. Large quartz crystals are still found there, as demonstrated by this local miner, or *garimpeiro*.

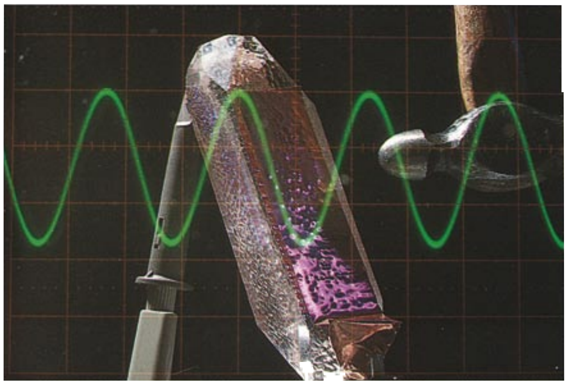

WAVES OF ENERGY
Quartz crystals are used in electronics. They can change a mechanical force, such as a blow from a hammer, into electrical energy, shown here as a wave-form on an oscilloscope screen.

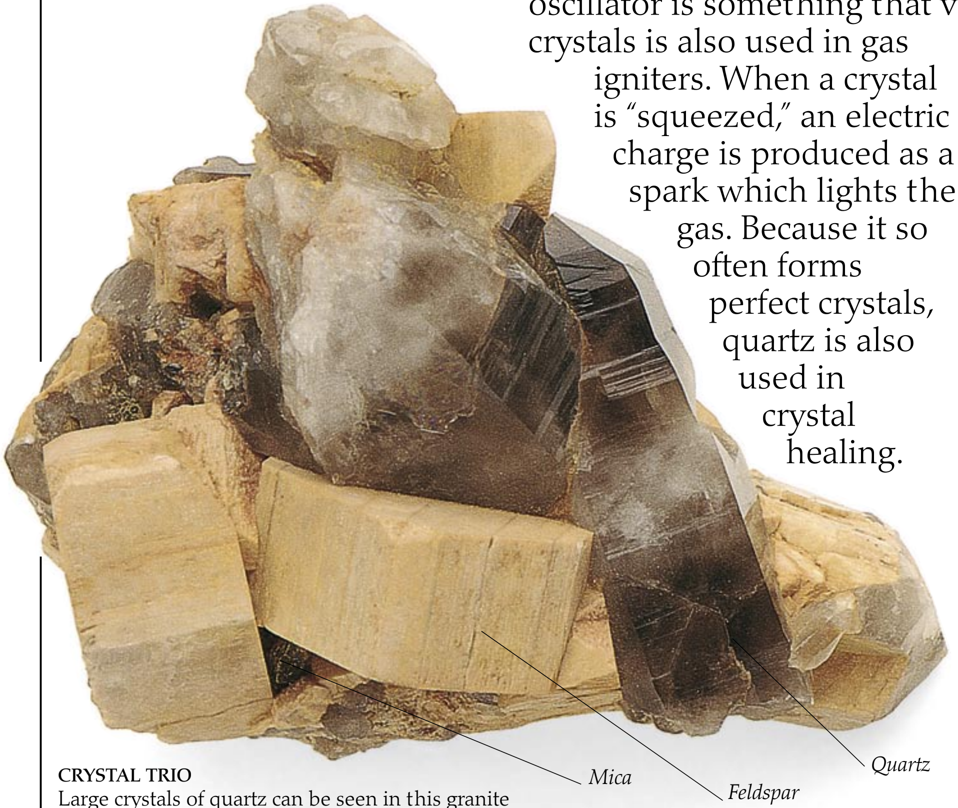

CRYSTAL TRIO
Large crystals of quartz can be seen in this granite pegmatite crystal group (p. 25). Fine crystals of the other two major components of granitic rocks, feldspar and mica, are also here.

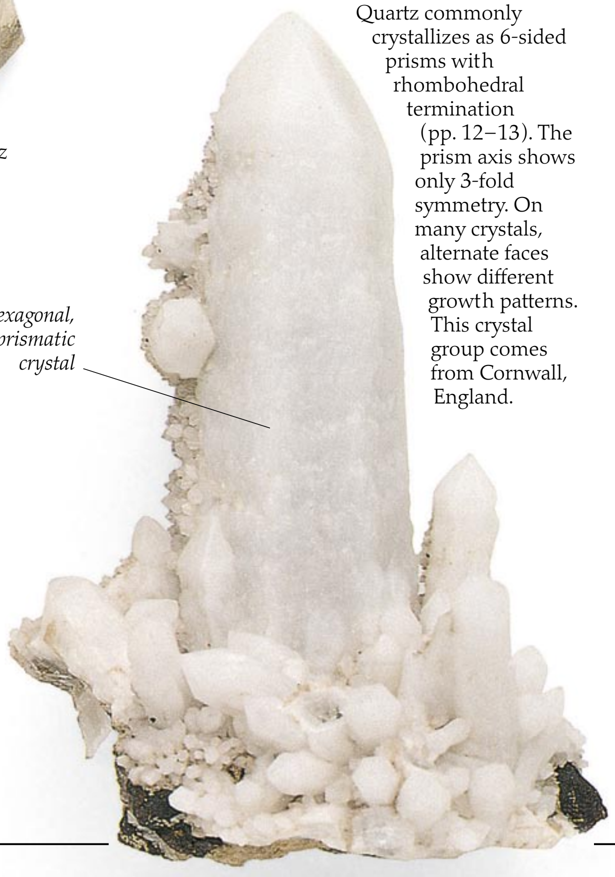

ENGLISH PRISM
Quartz commonly crystallizes as 6-sided prisms with rhombohedral termination (pp. 12–13). The prism axis shows only 3-fold symmetry. On many crystals, alternate faces show different growth patterns. This crystal group comes from Cornwall, England.

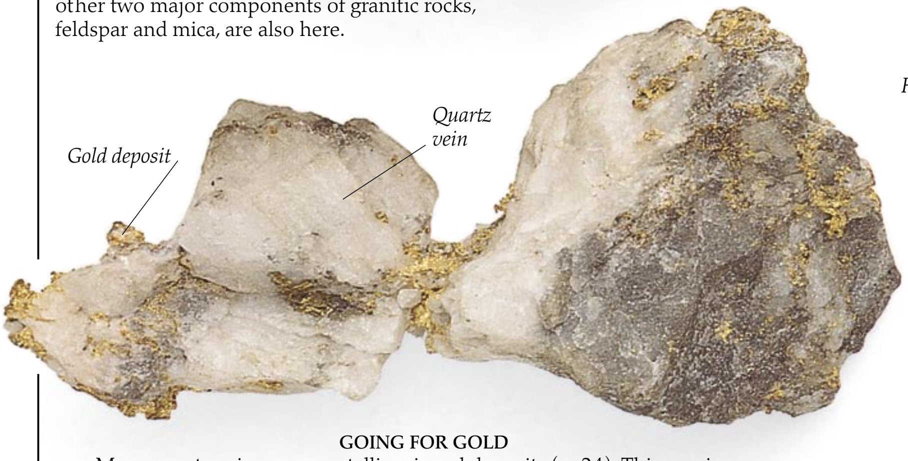

GOING FOR GOLD
Many quartz veins carry metallic mineral deposits (p. 24). This specimen contains gold. It came from St. David's mine, Gwynedd, Wales, an area famous for the extraction of British gold. The quartz and gold were both deposited by hydrothermal (hot, watery) fluids. In mining practice, the quartz would be considered a "gangue," or unwanted mineral.

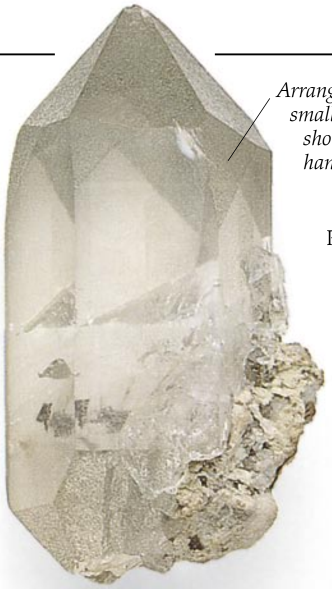

Arrangement of small faces shows left-handedness

Left-handed quartz crystal

Right-handed quartz crystal

AMBIDEXTROUS
In a quartz crystal, silicon and oxygen atoms are joined in the shape of a tetrahedron (a four-sided triangular pyramid). These tetrahedra are connected in a spiral arrangement, like a spiral staircase, and can be left- or right-handed. It is this structure which accounts for the piezoelectricity of quartz.

Crystal pendant thought by some to help with healing

Crystal healing

The laying on of stones is an ancient art. It is thought that as light reflects off the crystals and stones, the electromagnetic field of the body – the aura – absorbs energy. The receiver can then become aware of mental and emotional causes of physical disease, and heal.

CRYSTAL CLEAR
Rock crystals from groups such as this one from Arkansas are highly prized for their beauty and clarity and are often used for crystal healing.

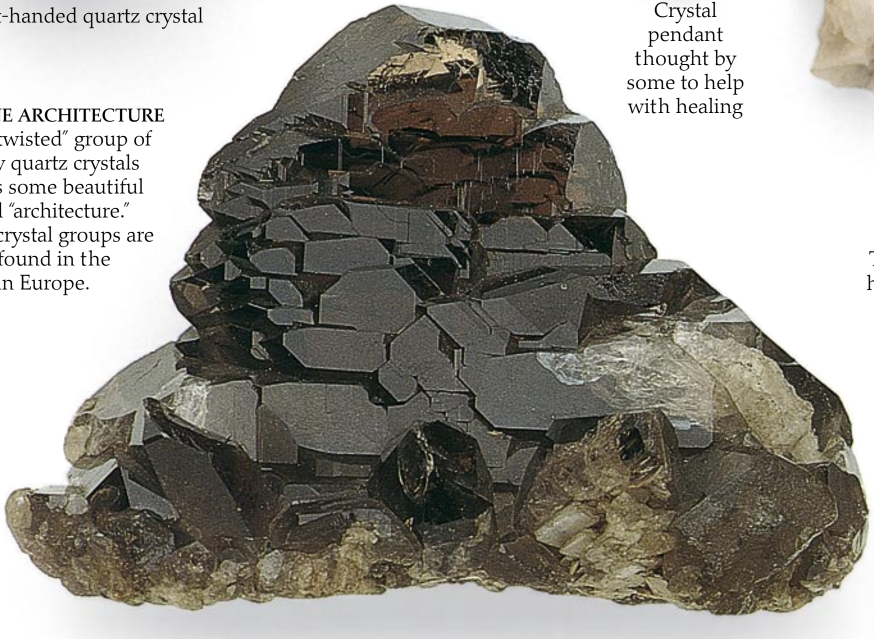

ALPINE ARCHITECTURE
This "twisted" group of smoky quartz crystals shows some beautiful crystal "architecture." Such crystal groups are often found in the Alps, in Europe.

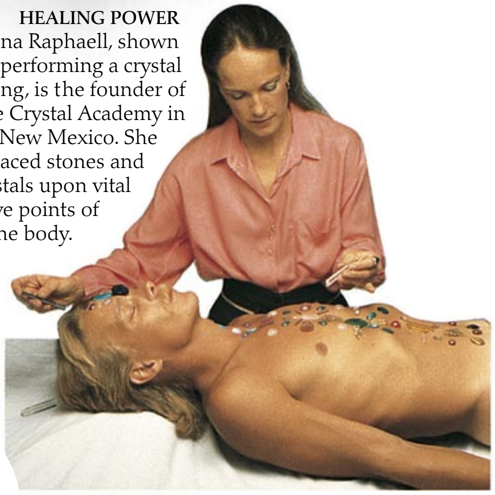

HEALING POWER
Katrina Raphaell, shown here performing a crystal healing, is the founder of the Crystal Academy in Taos, New Mexico. She has placed stones and crystals upon vital nerve points of the body.

PURE NECESSITY
To meet the demand for pure, flawless quartz crystals necessary for making oscillator plates, synthetic crystals like this one are now grown by a hydrothermal process (p. 26).

Piezoelectricity

Piezoelectricity was discovered by the brothers Pierre and Jacques Curie in 1880. They discovered that pressure on a quartz crystal causes positive and negative charges to be produced across the crystal. It was later found that an alternating electrical charge placed on a piezoelectric crystal could cause the crystal to vibrate. This is the basis of the use of quartz as oscillator plates to control radio waves and keep time.

Jacques and Pierre Curie with their parents

Quartz crystal slice

WATCH PIECE
This microthin quartz crystal slice is used to keep time in a quartz watch. The photograph is greatly enlarged.

SPLIT-SECOND TIMING
The crystal slice in a quartz watch vibrates more than 30,000 times each second. It is this regularity of vibration which makes it a good timekeeper.

Quartz

Quartz is silicon dioxide. It occurs as individual crystals and fine-grained masses in a large variety of forms, patterns, and colors. If conditions are right, giant crystals can grow (Brazil is famous for these). The largest recorded rock crystal was about 20 ft (6 m) long and weighed more than 48 tons (44,000 kg). Other sources of fine quartz include the Swiss Alps, the USA, and Madagascar. Quartz is tough and has no cleavage (p. 15), making it ideal for carving and cutting, and it is very popular as a gemstone. The name quartz usually refers to individual crystals or coarse-grained aggregates while the fine-grained materials are called chalcedonies or jaspers.

QUARTZ CRYSTAL
Crystal system: trigonal; hardness: 7; specific gravity: 2.65.

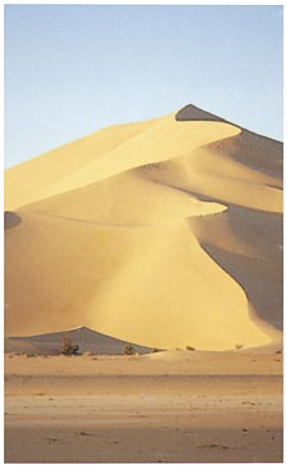

DUNES AND DUST
As quartz is relatively hard and common, it forms the major part of sand and also of dust in the air. Dust can therefore damage gems of 6 or less on Mohs' hardness scale (pp. 18–19).

Single crystals

The best-known single crystals of quartz are colorless rock crystal, purple amethyst, rose quartz, smoky quartz, and yellow citrine. These transparent crystals often occur in large enough pieces to be cut as gemstones.

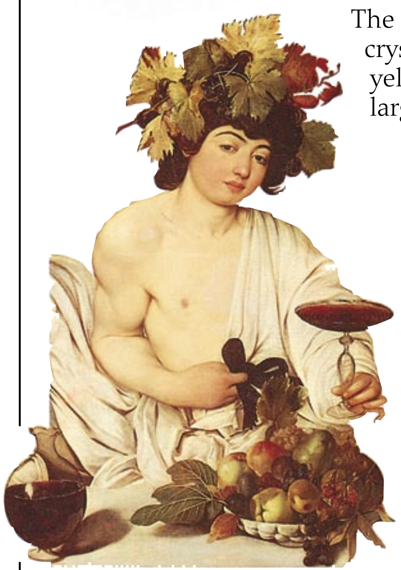

***BACCHUS* BY CARAVAGGIO**
A 16th-century French verse tells how Bacchus, the god of wine, declares in a rage that the first person he passes will be eaten by tigers. This turns out to be a beautiful maiden called Amethyst. The goddess Diana quickly turns Amethyst into a white stone to save her from the tigers. Regretting his anger, Bacchus pours red wine over the stone as an offering to Diana, so turning the stone purple.

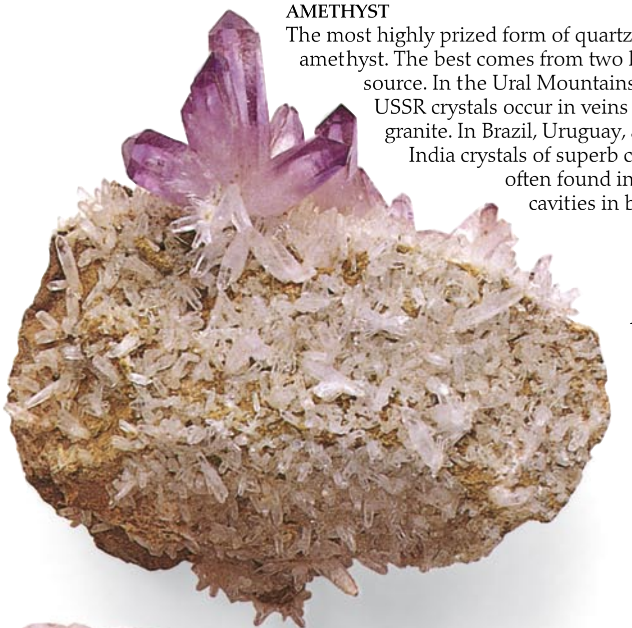

AMETHYST
The most highly prized form of quartz is amethyst. The best comes from two kinds of source. In the Ural Mountains in the USSR crystals occur in veins in granite. In Brazil, Uruguay, and India crystals of superb color are often found in cavities in basalt.

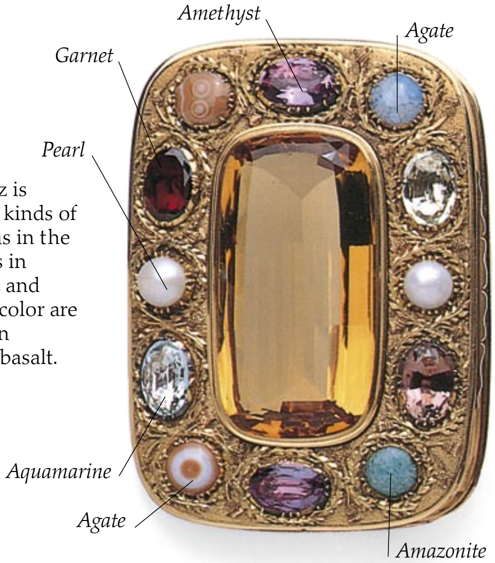

RARE BEAUTY
This 19th-century gold box is set with a superb rare citrine surrounded by a garnet (p. 44), an amazonite, two pearls (p. 55), two aquamarines (p. 39), three agates, and three amethysts.

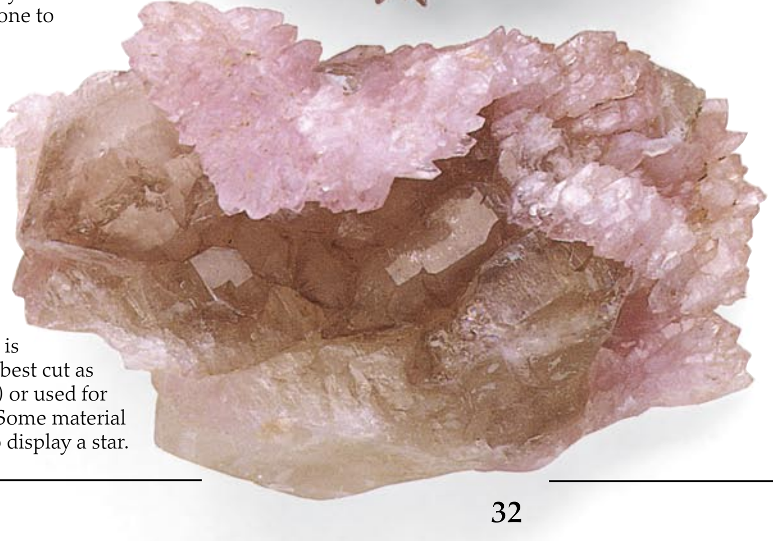

ROSE QUARTZ
Single rose quartz crystals are very rare and most rose quartz is massive. It is best cut as cabochons (pp. 58-59) or used for carvings and beads. Some material can be polished to display a star.

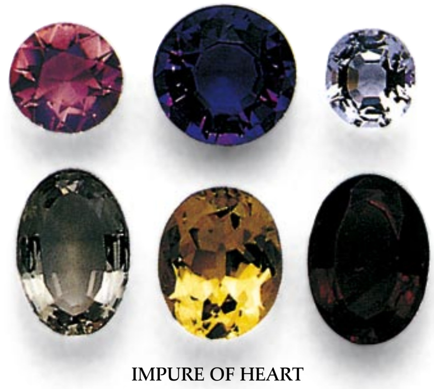

IMPURE OF HEART
Colorless rock crystal is the purest form of quartz, the many other colors being caused by impurities. Amethyst and citrine contain iron, rose quartz contains titanium and iron, and smoky quartz contains aluminum.

Massive

There are several massive varieties of quartz which are composed of very tiny grains or fibers. Chalcedony – such as carnelian, chrysoprase, and agate – and jasper are distinguished by the different arrangements of these grains. Tiger's-eye and hawk's-eye form when tiny fibers of asbestos are replaced by quartz and iron oxides.

AGATE
The quartz grains in chalcedony are arranged in layers and their buildup is clearly visible in the different colored layers of agate. In this specimen they crystallized progressively toward the middle of a cavity in lava.

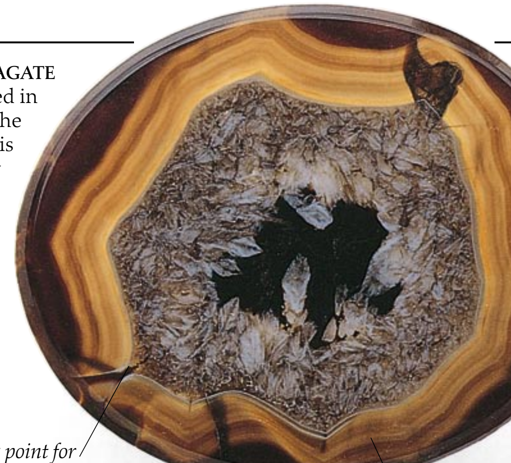

TIGER'S-EYE
Originally this vein of tiger's-eye contained silky blue crystals of asbestos. These were dissolved by solutions which deposited quartz and iron oxides in their place. The structure of the tiny fibers of asbestos was exactly reproduced by the quartz, and this gives rise to the light reflection or the "cat's-eye."

Polished tiger's-eye showing the cat's-eye effect called chatoyancy

A tiger shows why tiger's-eye is so named

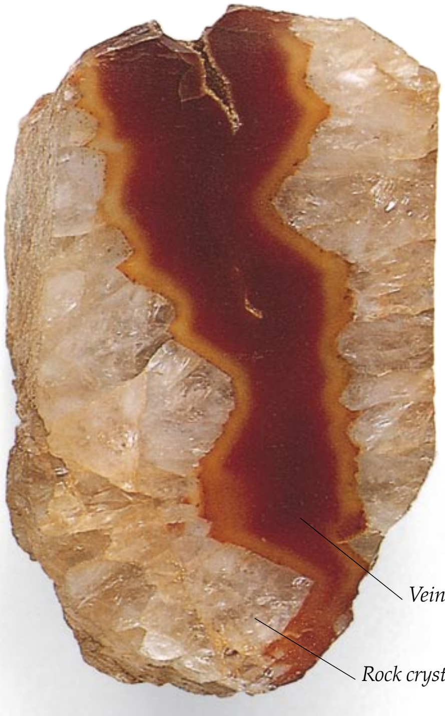

CARNELIAN
Carnelian is the name given to translucent (p. 16) orange-red chalcedony. Most specimens are the result of heat-treating a less attractive chalcedony. The treatment turns iron-bearing minerals into iron oxides which give the more desirable orange-red colors.

JASPER
The interlocking quartz crystals in jasper are arranged in a random mass. They are mixed with colorful impurities, making the stone opaque (p. 16).

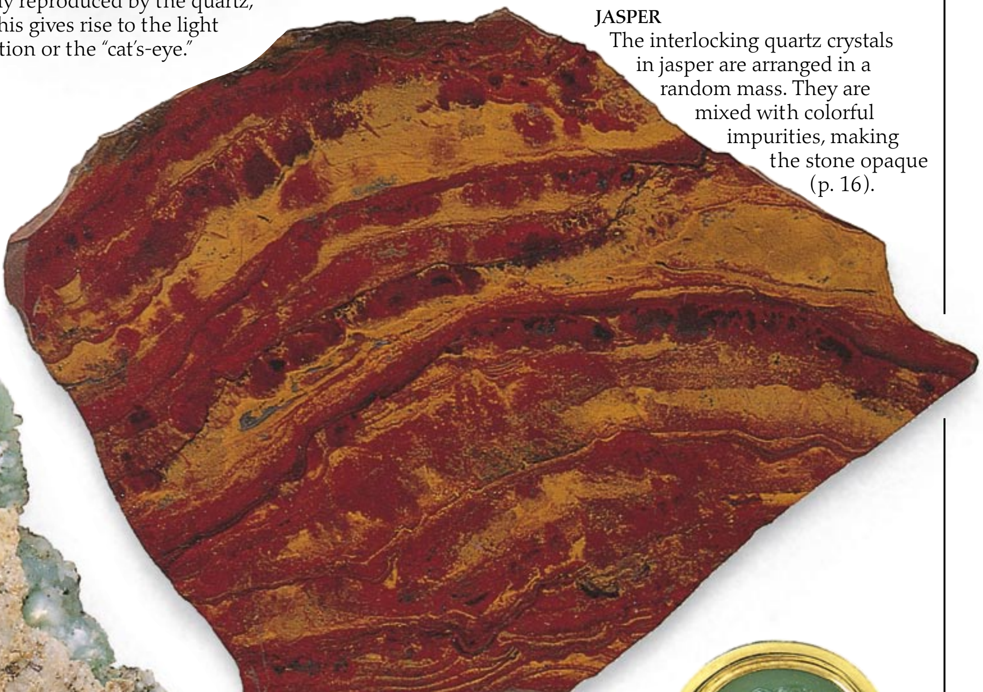

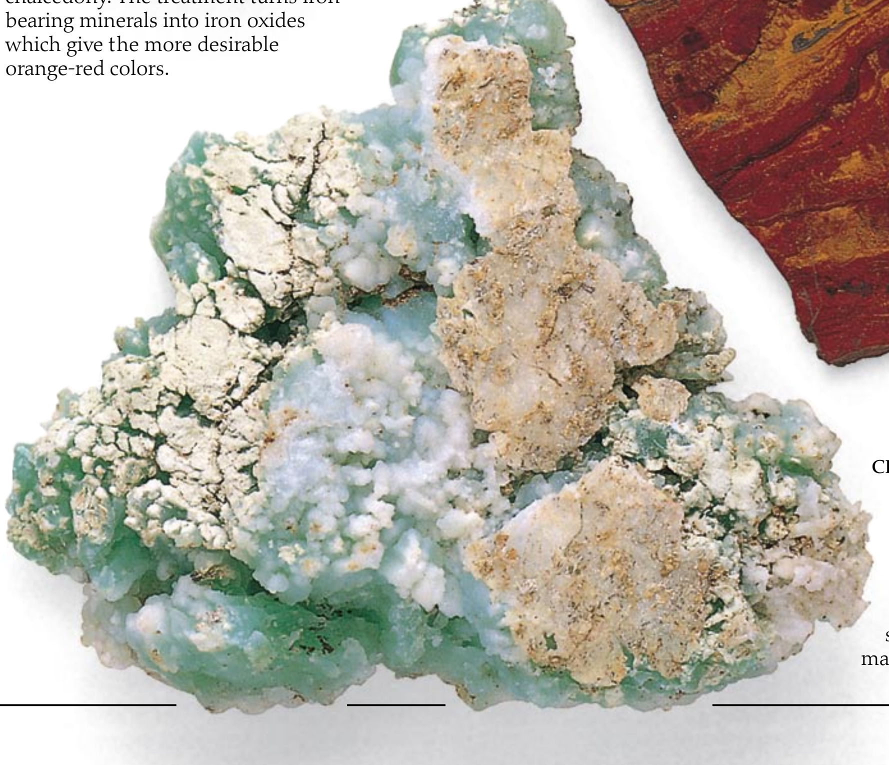

CHRYSOPRASE
At its finest, chrysoprase is a vibrant green and the most valuable of the chalcedonies. It has been used in ornament and decorative patterns since prehistoric times. A recent source of some of the best material is Queensland, Australia.

Chrysoprase cameo set in gold

Diamond

DIAMOND CRYSTAL
Crystal system: cubic; hardness: 10; specific gravity: 3.5.

THE WORD DIAMOND is derived from the Greek word *adamas*, meaning "unconquerable," given to the stone because of its supreme hardness. Diamond is made of pure carbon and has an immensely strong crystal structure (p. 14). It is this which makes it the hardest of all minerals. Evidence suggests that diamonds were formed up to 125 miles (200 km) deep within the earth, and some stones may be as much as three billion years old. Diamonds were first discovered over 2,000 years ago and came mainly from river gravel in India. In 1725, they were found in Brazil, which remained the major source until production in South Africa became significant in 1870. Today, about 20 countries produce diamonds. The top producer is Australia, which supplies a quarter of the world's needs, mainly for industrial purposes (p. 29). Diamond has great luster and fire, properties which are best revealed in the brilliant cut (p. 58).

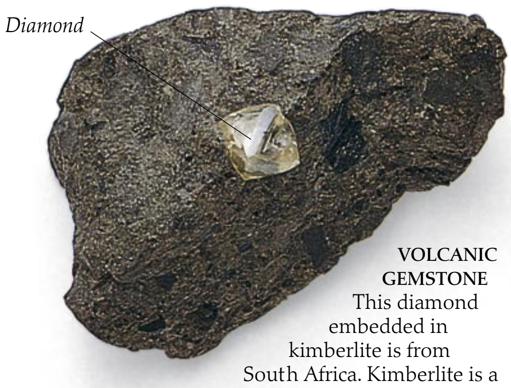

VOLCANIC GEMSTONE
This diamond embedded in kimberlite is from South Africa. Kimberlite is a volcanic rock that was first discovered in the Kimberley region of South Africa.

DIAMOND RUSH
In 1925 some very rich alluvial deposits were discovered at Lichtenburg, South Africa. The government decided to allocate claims (areas of land to mine) on the outcome of a race. So, on August 20, 1926, 10,000 miners lined up and had to race about 218 yards (200 m) to stake their claims.

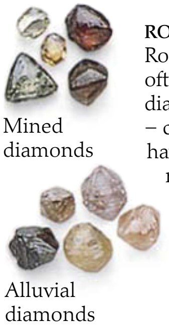

ROUGH DIAMONDS
Rough diamonds mined from kimberlites often have lustrous crystal faces; alluvial diamonds – those recovered from gravel – can be dull. This is because they may have been carried long distances in rough water with other rocks and gravel.

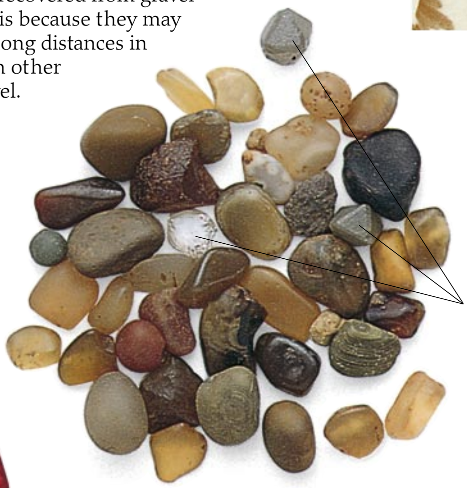

SPOT THE DIAMONDS
Diamond-bearing gravel is the result of one of nature's sorting processes. Seriously flawed or fractured stones are more likely to be broken up and eroded away, so a high proportion of the diamonds found in gravel are of gem quality.

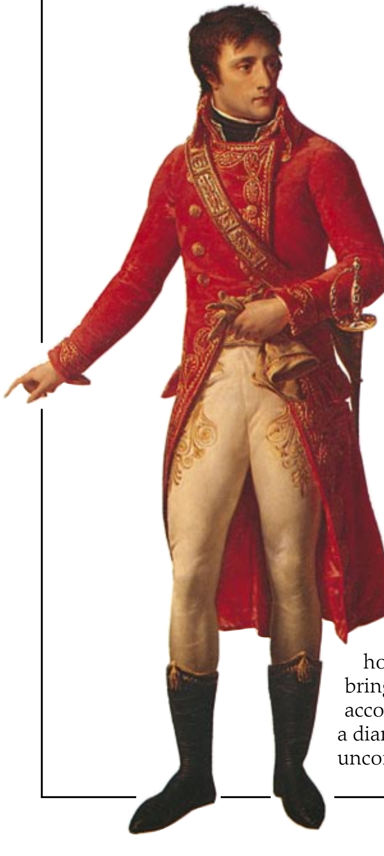

UNCONQUERABLE BELIEF
Napoleon Bonaparte is depicted here as First Consul wearing a sword set with the Regent diamond. He hoped the diamond would bring him victory in battle; according to an ancient belief, a diamond made its wearer unconquerable.

RICH MIX
Conglomerate rock is a mixture of different sizes of rounded pebbles and mineral grains which have been deposited from water and cemented together. This specimen from the west coast of South Africa is particularly rich in diamonds.

Famous diamonds

Diamonds of exceptional beauty and rarity are highly prized. Some have long, recorded histories and others have inspired fantastic legends. Most belong to the rich and famous.

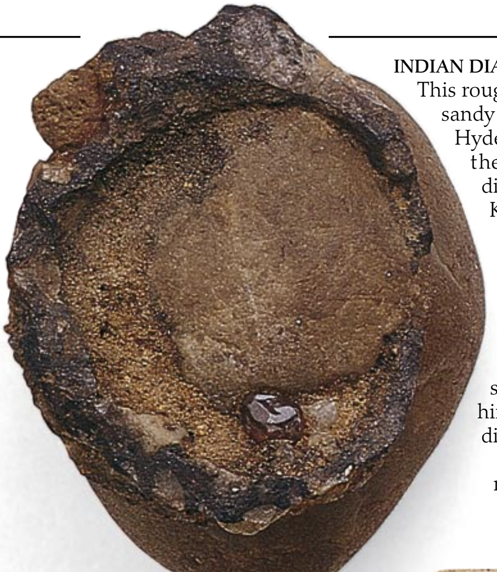

INDIAN DIAMOND
This rough diamond is embedded in a sandy conglomerate found near Hyderabad in India. This area was the source of many famous large diamonds such as the Koh-i-noor and the Regent.

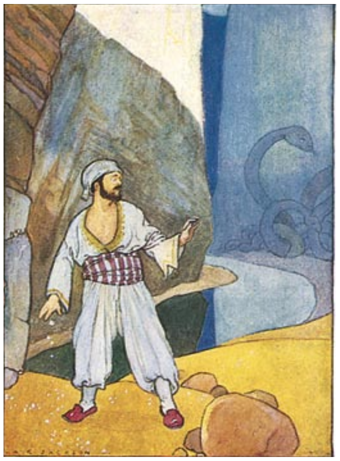

VALLEY OF DIAMONDS
Sindbad was once stranded in the legendary Valley of Diamonds. On the valley floor were diamonds guarded by snakes. Sindbad escaped by tying himself to meat thrown down by a diamond collector. As intended by the collector, a bird carried the meat out of the valley stuck with diamonds – and Sindbad!

BUTTERFLY BROOCH
This butterfly brooch is set with over 150 diamonds.

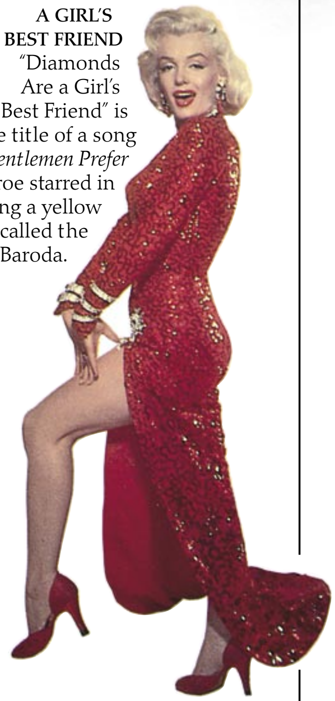

A GIRL'S BEST FRIEND
"Diamonds Are a Girl's Best Friend" is the title of a song from the film *Gentlemen Prefer Blondes*. Marilyn Monroe starred in the film wearing a yellow diamond called the Moon of Baroda.

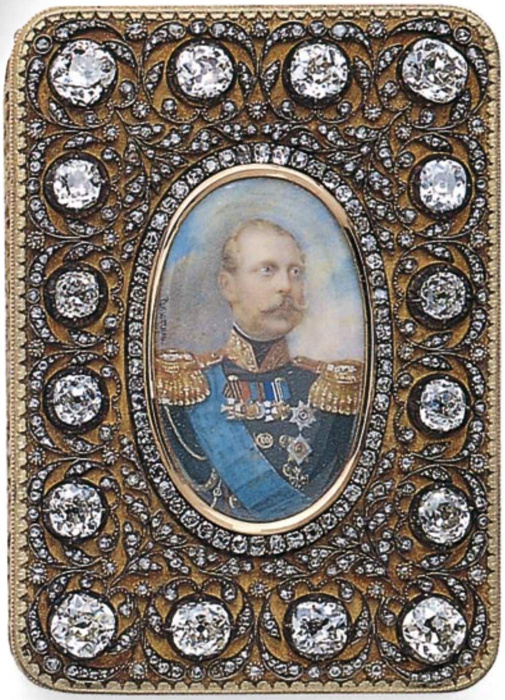

MURCHISON SNUFFBOX
This gold box set with diamonds bears a portrait of Czar Alexander II of Russia. It was presented in 1867 by the czar to Sir Roderick Murchison, the second director of the British Geological Survey, in recognition of Sir Roderick's geological work in Russia.

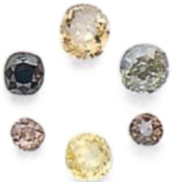

BRILLIANT COLORS
Most natural diamonds are near-colorless; truly colorless ones are rare. A few stones are also found of all colors in the spectrum (p. 16) and good-quality ones are known as fancies.

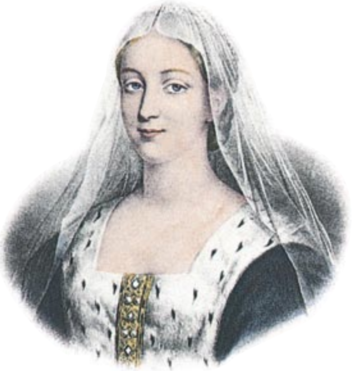

AGNÈS SOREL (c 1422-1450)
Agnès Sorel, the mistress of the French king Charles VII, was the first commoner in France to break the law made by Louis IX in the 13th century decreeing that only kings and nobles could wear diamonds.

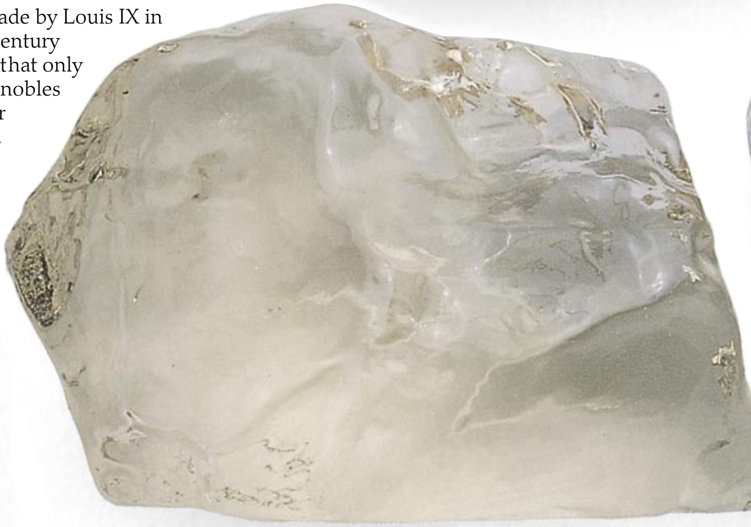

PREMIER DIAMOND
In 1905 the Cullinan crystal was found in the Premier diamond mine in the Transvaal, South Africa. It weighed 3,106 carats and is still the largest diamond ever found. This replica shows its actual size. In 1908 it was cut into 9 large and 96 lesser stones. The two largest, Cullinan I and II, are in the British crown jewels (p. 46).

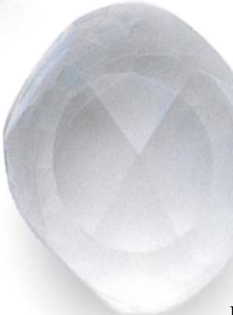

THE JEWEL IN THE CROWN
The Koh-i-noor (mountain of light) is claimed to be the oldest large diamond. It was probably found in India and after belonging to Mogul kings was presented to Queen Victoria in 1850. Its cut, shown in this replica, was unimpressive, so it was recut (p. 58). Today, it is in the British crown jewels.

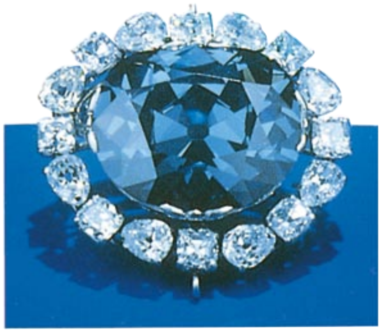

BLUE HOPE
The Hope has a reputation for bringing bad luck, but the sinister stories are untrue. It is 45.52 carats and is now in the Smithsonian Institution, Washington D.C.

Corundum

Ruby and sapphire are varieties of the mineral corundum, an aluminum oxide. Only true red stones are called rubies, and the term *sapphire* on its own indicates a blue stone. Other colors are described as sapphire, that is, yellow sapphire and pink sapphire. Corundum is next to diamond in hardness, so gem crystals are resistant to wear. It is pleochroic, which means the color of a stone varies when it is viewed in different directions. Most gem crystals are recovered from gravel, and the most famous sources are Myanmar (formerly Burma), Kashmir, and Sri Lanka. Today, Australia is the largest producer of blue and golden sapphires. Other producers include Thailand and countries in East Africa.

CORUNDUM
Crystal system: trigonal; hardness: 9; specific gravity: 3.96–4.05.

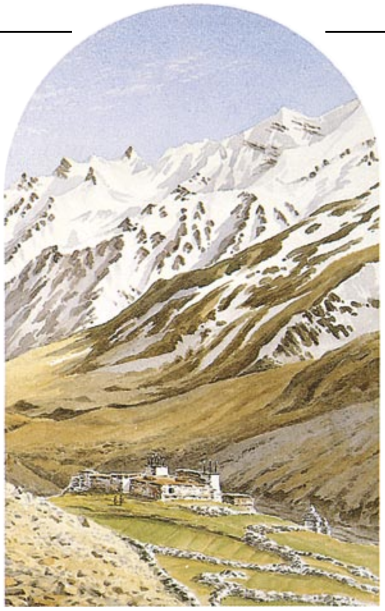

SOURCE REVEALED
A famous source of fine sapphires is in a valley in the Kanskar range of the Himalayas in Kashmir. It is said the source was only revealed after a landslide in about 1881.

Twin sapphire crystals

Sapphire intergrown with tourmaline

KASHMIR BLUE
Kashmir has a reputation for producing sapphires of the finest blue, like these two examples. The term *Kashmir blue* is often used to describe sapphires of this color from other parts of the world.

RUSKIN'S RUBY
This Myanmar ruby crystal was presented to the Natural History Museum in London by the philosopher John Ruskin in December 1887. It is about 162 carats. Its deep red color is the most admired color for a ruby and is sometimes described as "pigeon's blood" red.

MYANMAR CRYSTAL
Most of the highest-quality rubies come from the Mogok region in Myanmar, and this fine crystal embedded in calcite is a good example. Rubies from Myanmar, Pakistan, and Afghanistan are often found in calcite.

Flattened prism of fine-quality ruby from the Mogok district of upper Myanmar

BAZAAR DEALING
This 1930 photograph shows ruby dealers in a Mogok bazaar. Gem-quality corundum is rare, and ruby is the most valuable variety of all. Good quality stones can fetch even higher prices than diamonds of the same size.

IN A SPINEL
Sapphires are often found with spinels (p. 46) in gem gravel but they are rarely found in their source rock. This sapphire from Myanmar has grown with crystals of spinel. The crystals grew in conditions of high temperature and pressure.

QUEEN CAMEO
These gems were mounted on a rock crystal pitcher in about 1660. The pitcher was rediscovered in 1985. In the center is a ruby cameo of the head of Queen Elizabeth I of England.

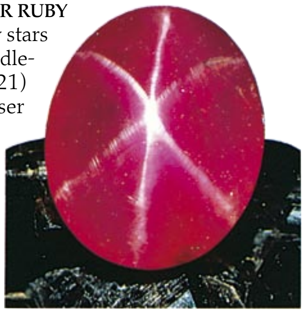

STAR RUBY
Some rubies display stars caused by fine needle-like inclusions (p. 21) of rutile. The Rosser Reeves ruby is one of the finest. It weighs 138.7 carats and belongs to the Smithsonian Institution, in Washington D.C.

BLUE BUDDHA
The figure representing a Buddha on this pin is carved from a fine sapphire crystal.

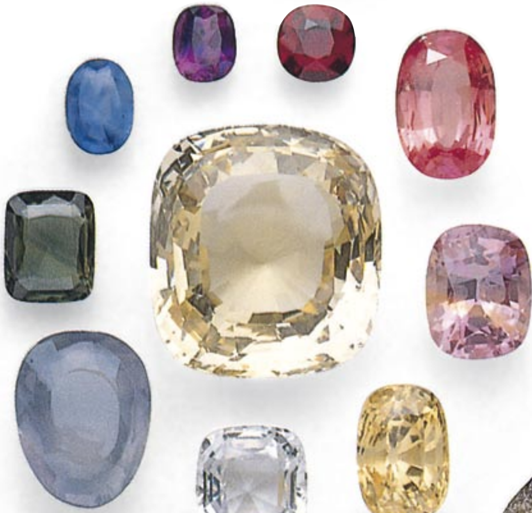

RICH COLORS
Pure corundum is colorless, but tiny amounts of impurities are responsible for a wide range of rich colors: chromium (red), iron (yellow and green), and iron and titanium (blue).

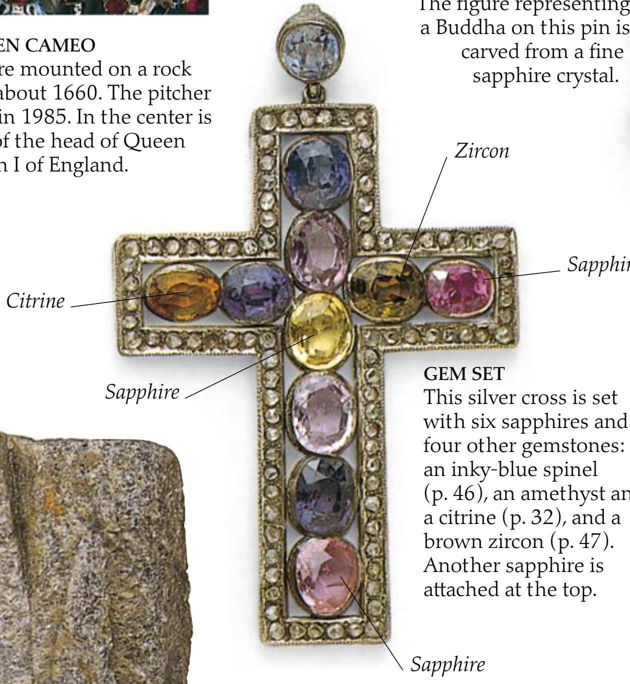

GEM SET
This silver cross is set with six sapphires and four other gemstones: an inky-blue spinel (p. 46), an amethyst and a citrine (p. 32), and a brown zircon (p. 47). Another sapphire is attached at the top.

HARD AS ROCK
Most natural corundum is not of gem quality. It is opaque and gray or brown like this crystal from Madagascar. Because of its hardness this material is used on industrial tools, and in the past it was used for fashioning other softer stones.

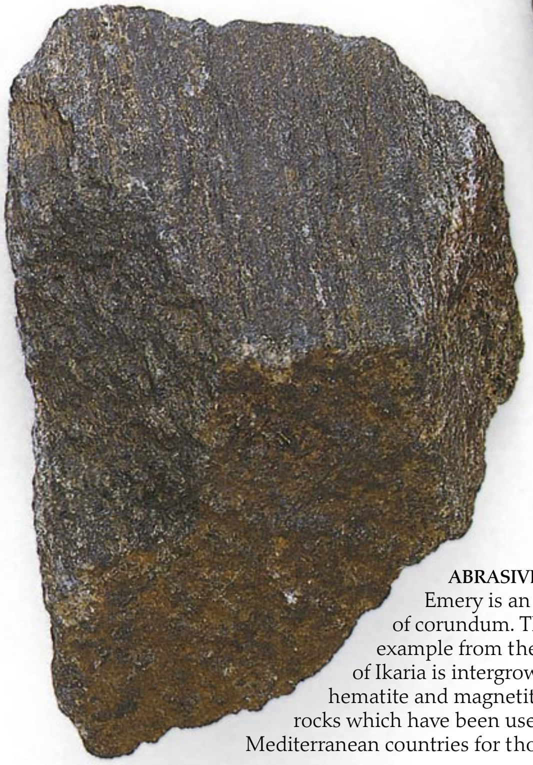

ABRASIVE ROCK
Emery is an impure type of corundum. This fine-grained example from the Greek island of Ikaria is intergrown with black hematite and magnetite. It is typical of rocks which have been used as abrasives in Mediterranean countries for thousands of years.

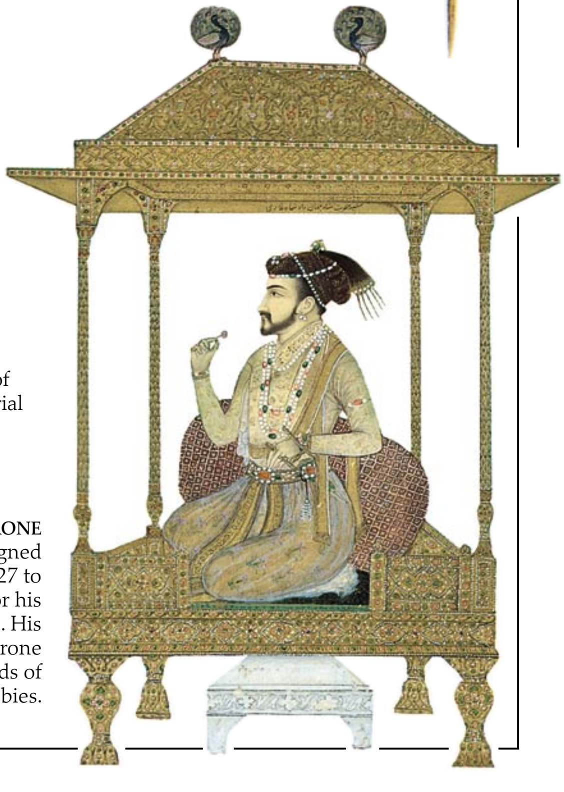

PEACOCK THRONE
Shah Jahan, who reigned over Persia from 1627 to 1658, was famous for his knowledge of gems. His legendary Peacock Throne was set with hundreds of gems including 108 rubies.

Beryl

Beryls are popular as gems because of their fine colors and resistance to wear. The most well-known varieties are emerald (green) and aquamarine (blue-green). Yellow beryl is known as heliodor, and pink beryl is morganite. The name beryl can be traced to Greek, Roman, and possibly Sanskrit sources, and it is highly likely that aquamarine and heliodor were known in prehistoric times. Beryl is found in pegmatites (p. 25) and granites and in its massive, opaque, non–gem form can occur in crystals weighing many tons. The recordholder is a crystal which was found in Madagascar weighing 36 tons and measuring about 60 ft (18 m) long.

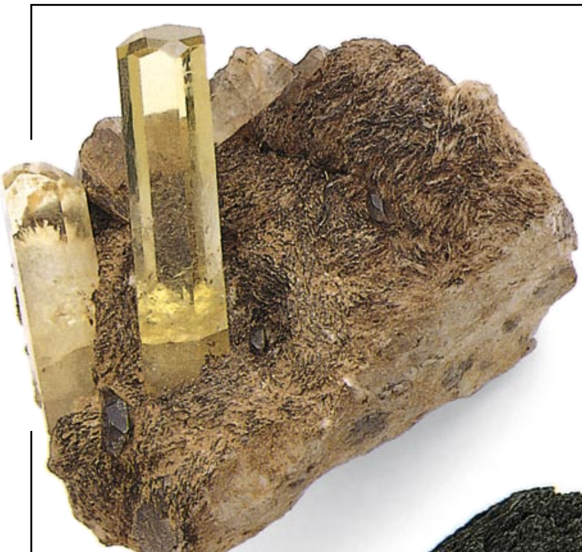

BERYL CRYSTAL
Crystal system: hexagonal; hardness: 7.5; specific gravity 2.63–2.91.

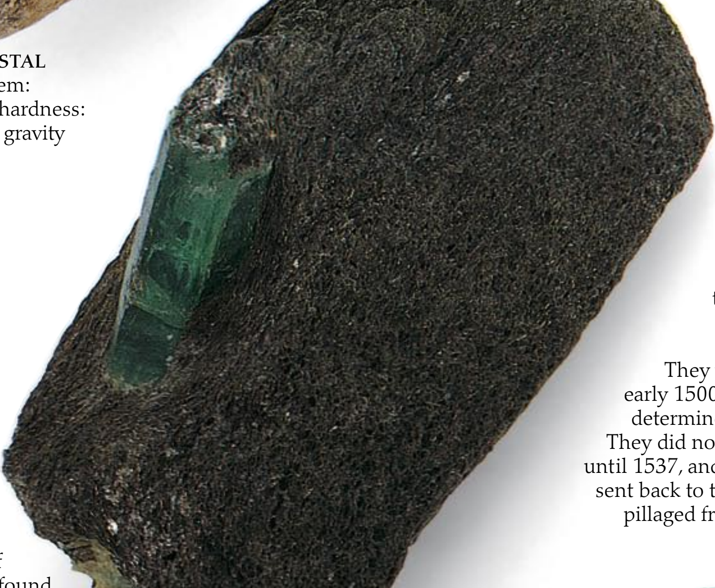

SOVIET HOST
A typical source of emeralds is mica schist. This kind of source was found in the early 1800s in the Ural Mountains, USSR. Many of these crystals have mica and amphibole inclusions (p. 21), the same rock as the host.

SPANISH SPOILS
The Chibcha Indians of Colombia mined emeralds which, through trade, reached the Incas of Peru and the Aztecs of Mexico. They were seen there in the early 1500s by the Spanish, who determined to find their source. They did not find the Chivor mine until 1537, and most of the emeralds sent back to the Spanish court were pillaged from the Incas' treasure.

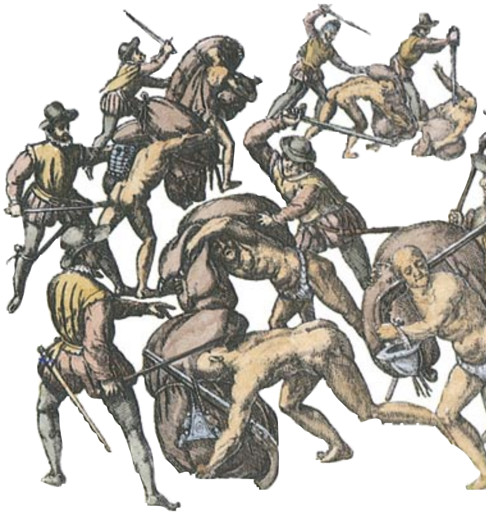

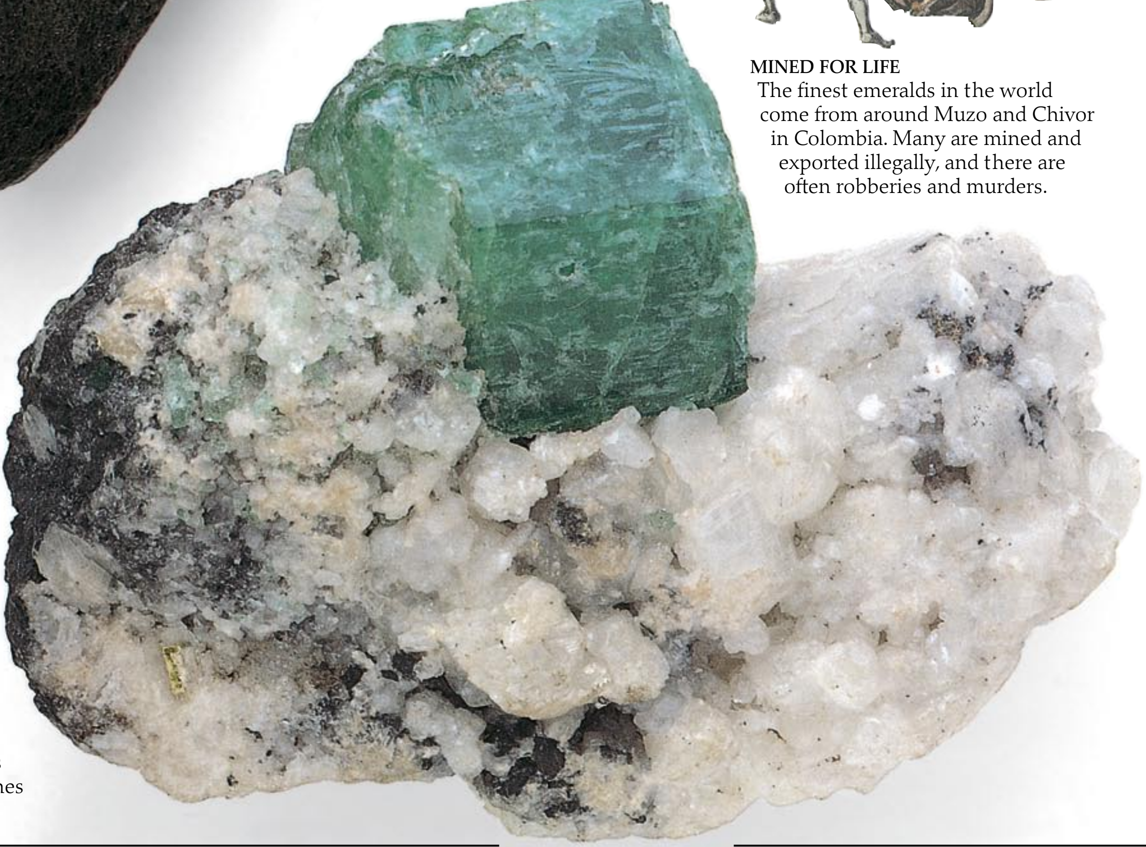

MINED FOR LIFE
The finest emeralds in the world come from around Muzo and Chivor in Colombia. Many are mined and exported illegally, and there are often robberies and murders.

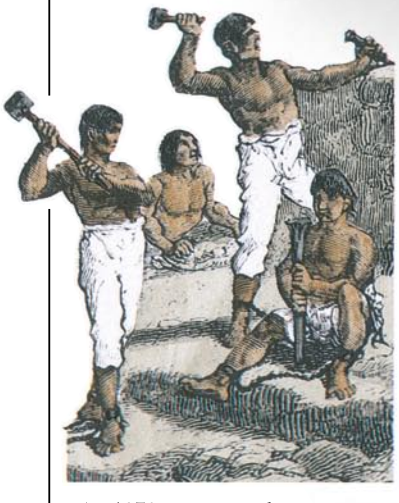

An 1870 engraving showing convicts working the Colombian emerald mines

SECOND-CLASS CRYSTAL
A few emeralds are still found in Egypt. They come from an area of granite, schist, and serpentine. Most crystals are blue-green with many inclusions and do not compare in quality with the best Colombian emeralds.

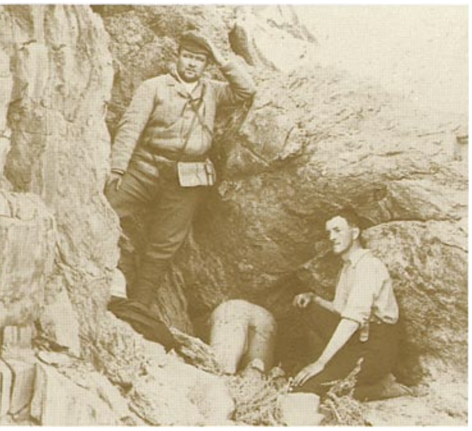

ANCIENT ACCESS
Emeralds were mined near the Red Sea in Egypt from about 1500 B.C. The mines were rediscovered in 1816 by the French adventurer Cailliaud, but attempts at mining were not a success. This old entrance to one of Cleopatra's mines was discovered in about 1900.

FINE CUT
This exceptionally fine, 911-carat, cut aquamarine belongs to the Smithsonian Institution in Washington, D.C.

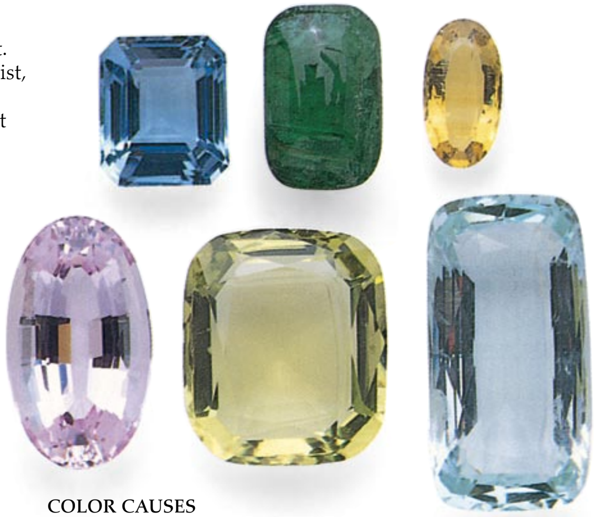

COLOR CAUSES
Pure beryl is colorless. The spectacular reds and pinks are caused by manganese, the blues and yellows by iron. The beautiful emerald green is due to tiny amounts of chromium or vanadium.

SEA-GREEN
Aquamarine means "sea water," which accurately describes its color. It has a range from pale green to blue, caused by varying amounts of different forms of iron. It is relatively common, with sources in many countries; the main source is Brazil.

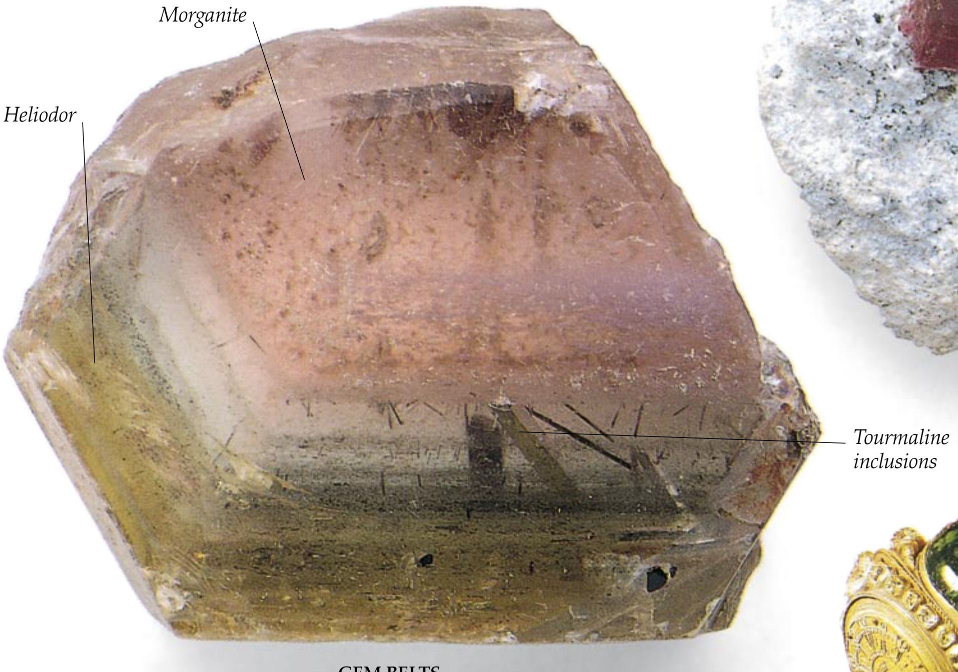

GEM BELTS
This large crystal of beryl is made up of zones of the varieties morganite and heliodor. This gem-quality specimen comes from Brazil, but other sources of these varieties include California, Madagascar, and Pakistan.

DRY RED
Red beryl is extremely rare and is the only natural beryl that contains no water in its crystal structure. It occurs in "dry" volcanic rocks in the western USA. This fine specimen comes from the Wah Wah Mountains in Utah.

TURKISH DELIGHT
The Topkapi Palace Treasury in Istanbul, Turkey, contains many pieces with fine emeralds. The hilt of this 18th-century dagger holds three, plus one cut as a hexagonal slice and hinged over a small watch at the end.

Opal

THE POPULARITY OF OPAL has risen and fallen over the centuries. The ancient Romans used it as a symbol of power but since then, at different times, it has been considered to be unlucky. The Aztecs mined opal over 500 years ago in Central America; this area is still an important source, especially of fire opal, from Mexico. Australia is the top producer of both black and white opals; they were first discovered there in the 1870s. Opal is one of the few noncrystalline gems. It has a tendency to crack and chip, especially under extreme temperature changes or after a hard knock. The exciting flashes of color shown by precious opal are best displayed in cabochons, but Mexican fire opals are usually cut as brilliants or step cuts (p. 58).

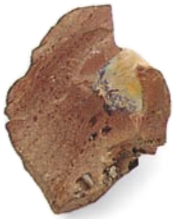

OPAL
Crystal system: amorphous or poorly crystalline; hardness: 5.5–6.5; specific gravity: 1.98–2.25.

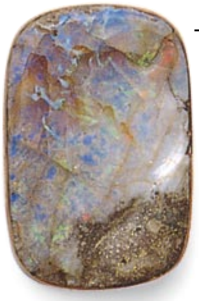

ROMAN SOURCE
This piece of white opal comes from Cervenica, Czechoslovakia, the source used by the Romans. This region was once part of Hungary, and the opal from here is usually described as Hungarian.

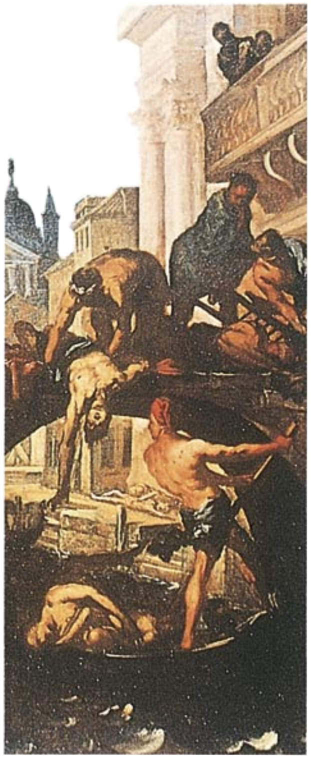

THE PLAGUE OF VENICE
This detail of a painting by Antonio Zanchi depicts the Black Death of the 14th century. The people in Venice, Italy, noticed that opals became brilliant when a wearer caught the disease and dulled when the person died. This reinforced the belief that opals were unlucky.

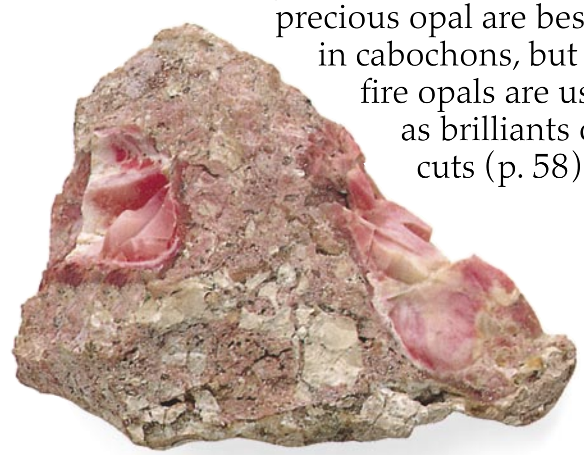

NONFLASHY
Nonprecious opal without flashes of color is called potch opal. Rose opal is potch, but its striking color has led to its being used in beads and decorative jewelry. This specimen comes from France; there are other sources in Idaho.

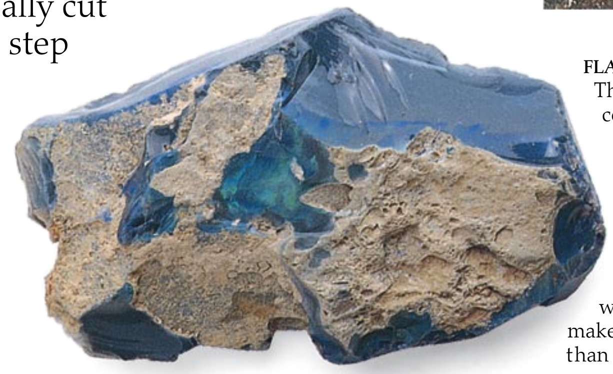

FLASH OF LIGHTNING
The finest black opal comes from Lightning Ridge, New South Wales, Australia. Against its dark body the color flashes are quite dramatic and this attraction, coupled with its greater rarity, make it more valuable than white opal.

AUSTRALIAN FAIR
All the major Australian opal deposits occur in sedimentary rocks in the Great Artesian Basin. Famous mines include White Cliffs, Lightning Ridge, and Coober Pedy. A popular way of marketing the opal is in doublets and triplets (p. 56).

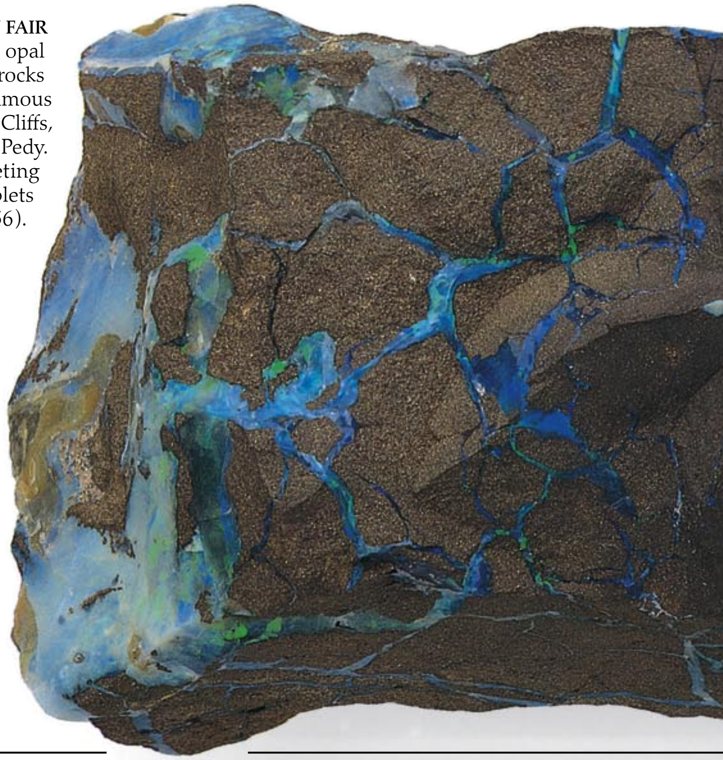

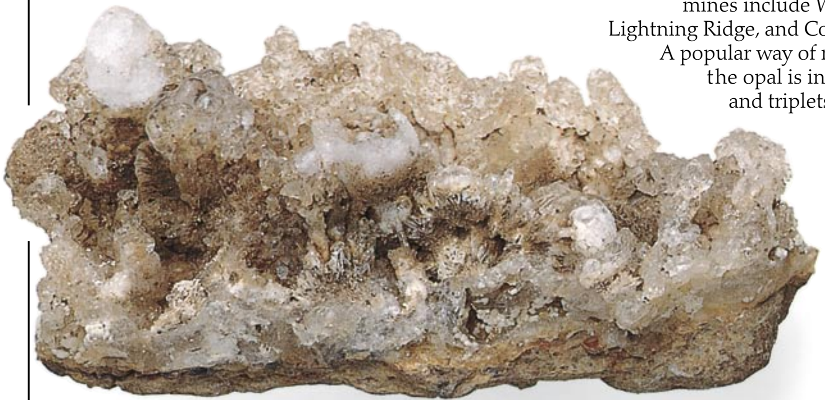

GLASSY LOOK
Clear, glassy-looking opal (hyalite) occurs in cavities in volcanic lavas. This example is from Bohemia (Czechoslovakia) but there are several other sources. If similar material shows a play of color, it is known as water opal. Another kind of opal, called hydrophane, is opaque but appears colorless in water.

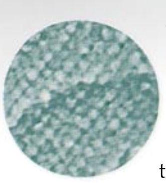

Greatly enlarged photograph of precious opal, showing the ordered silica spheres in the structure that cause the play of colors

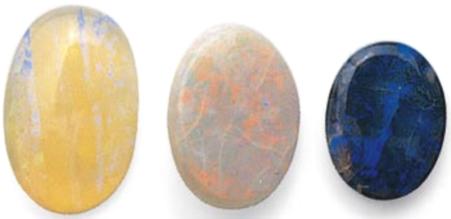

FLASHING LIGHTS
Precious opal displays flashes of different colors. The colors depend on the size of the silica spheres in the structure. Opal with a background color of gray, blue, or black is called black. Others are called white.

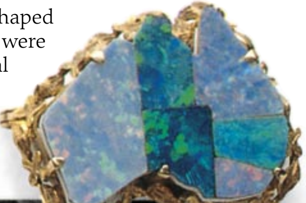

ON THE MAP
"Prospector's brooches," shaped like the map of Australia, were made to mark the arrival on the gemstone market of Australian opal. This one is thought to have been made around 1875.

MOVED IN
The opal localities in Australia are extremely hot, so when mines are exhausted, those excavations near the surface are adapted to provide relatively cool and pleasant living conditions.

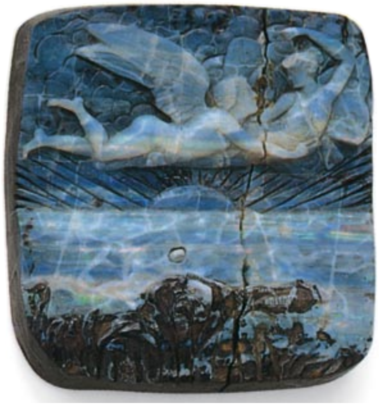

Opal cameo *Dawn with Cupid and Psyche* now in the Natural History Museum, London

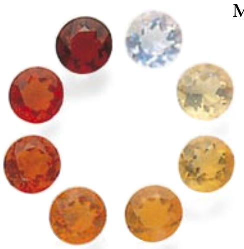

MEXICAN FIRE
Mexico has long been famous for its fire opal. This is a nearly transparent variety, still showing flashes of color. It ranges from yellow to orange and red.

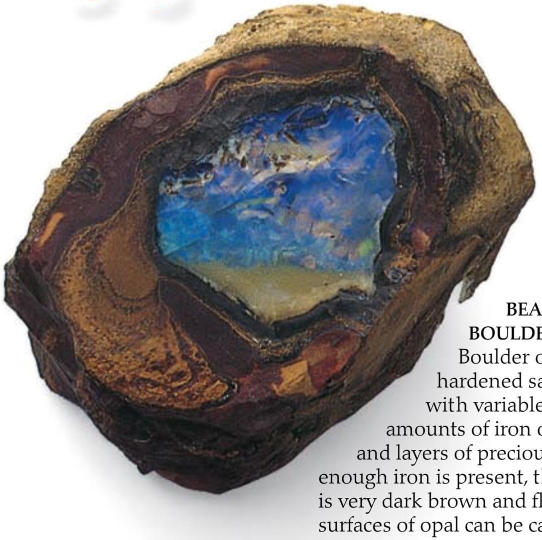

BEAUTIFUL BOULDER
Boulder opal is hardened sandy clay with variable amounts of iron oxides and layers of precious opal. If enough iron is present, the rock is very dark brown and flat surfaces of opal can be carved into beautiful cameos.

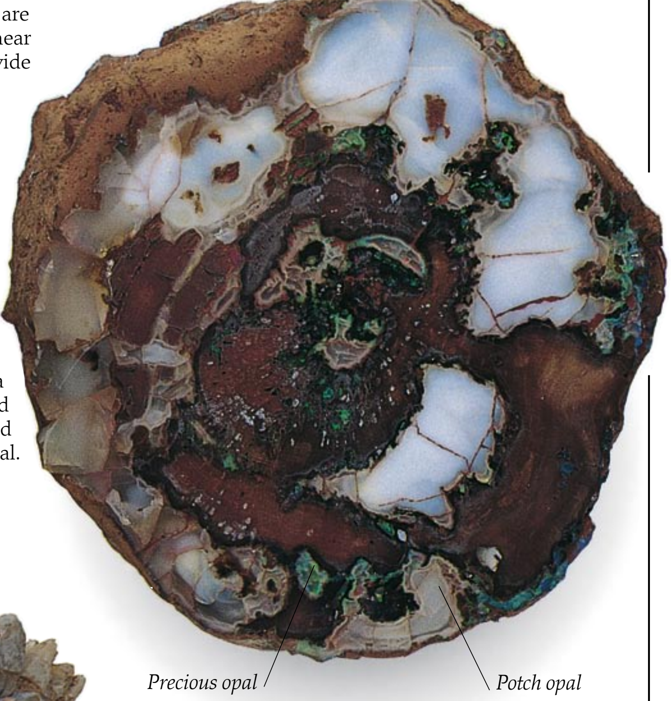

PRECIOUS FOSSIL
Opal often replaces the tissues of wood, and bones and shells of ancient animals, in a kind of fossilization. It grows bit by bit in place of the original material. This piece of wood from Nevada has been replaced by precious and potch opal.

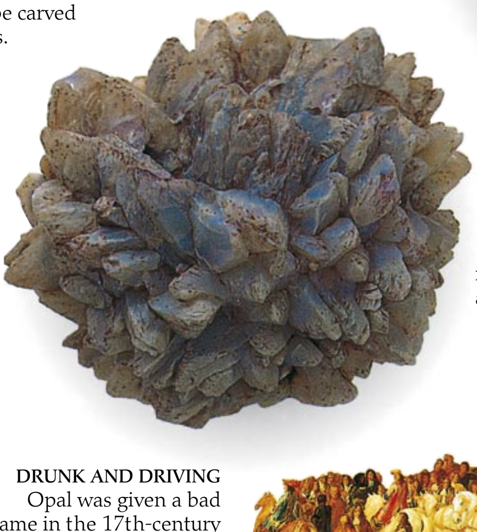

OPAL FRUIT
This was an aggregate (p. 22) of radiating crystals of glauberite. It has been completely replaced by precious opal. This kind of opal is found in Australia and is popularly known as pineapple opal.

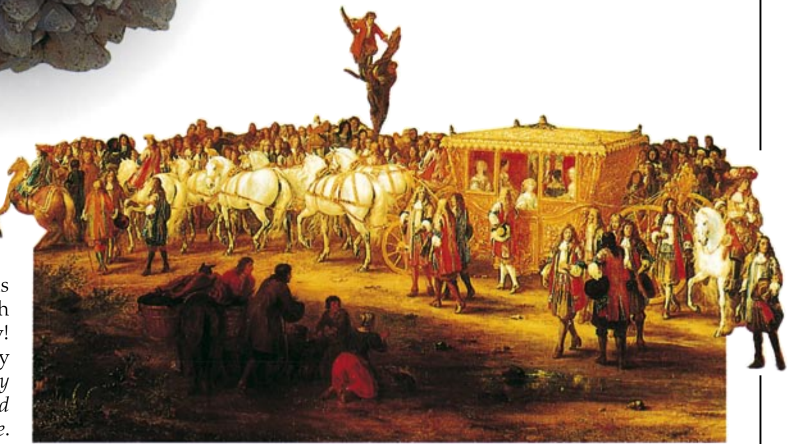

DRUNK AND DRIVING
Opal was given a bad name in the 17th-century court of the French king Louis XIV. He named his coaches after gemstones. The driver of the *Opal* was usually drunk, and the coach was considered unlucky! This is a detail of a picture by Van der Meulen called *Entry into Arras by Louis XIV and Marie Thérèse*.

Other gemstones

THE PROPERTIES OF A GEMSTONE are said to be beauty, rarity, and durability, and these standards have been applied to many species. As well as quartz (pp. 32–33), diamond (pp. 34–35), ruby and sapphire (pp. 36–37), beryl (pp. 38–39), and opal (pp. 40–41), gemstones to be seen in jewelers' shops include topaz, tourmaline, garnet, peridot, and many others. Some species, such as kunzite, sphene, and fluorite, are top soft or rare to be in general circulation as gemstones and are cut only for collectors on the basis of their beauty and rarity.

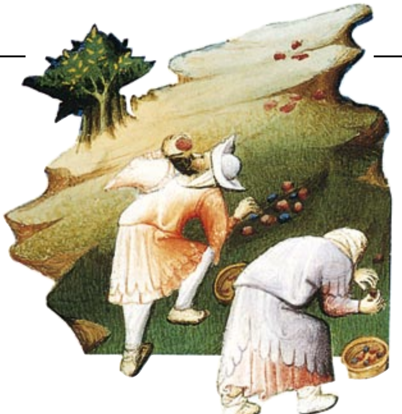

MARVELOUS GEMS
"Fishing for pearls and gathering turquoises" from *The Book of Marvels* by Marco Polo.

Topaz

The mineral we know as topaz was only given the name in the first half of the 18th century. Prior to that its history is not clear. The name topaz is said to come from *Topazius*, the Greek name for Zabargad, an island in the Red Sea. This island is, however, a source of what we now call peridot (p. 45).

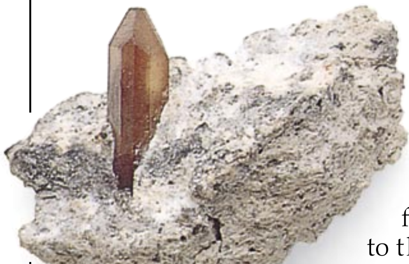

TOPAZ CRYSTAL
Crystal system: orthorhombic; hardness: 8; specific gravity: 3.52–3.56.

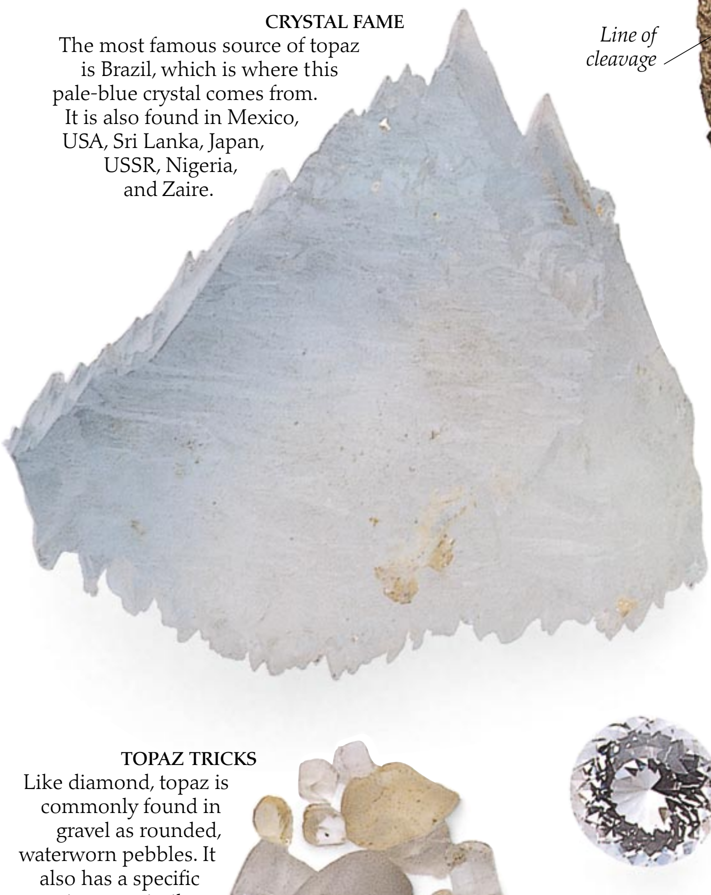

CRYSTAL FAME
The most famous source of topaz is Brazil, which is where this pale-blue crystal comes from. It is also found in Mexico, USA, Sri Lanka, Japan, USSR, Nigeria, and Zaire.

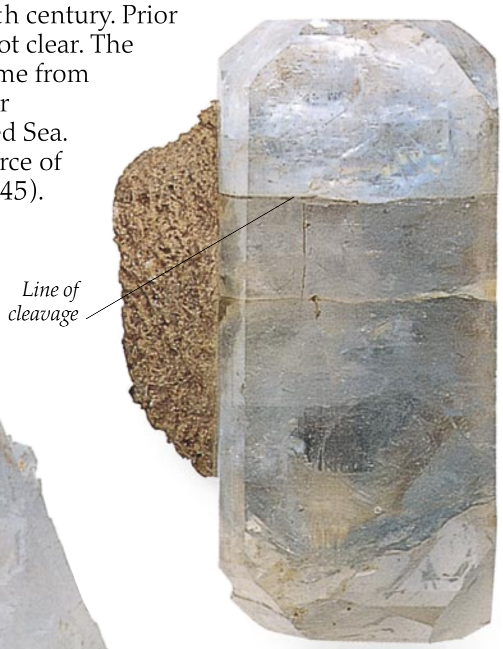

NEEDS PROTECTION
Although it is very hard, topaz can be broken easily because it has one direction of perfect cleavage (p. 15). A line of cleavage can be seen clearly in this crystal. If the topaz is used in jewelry, the setting must therefore provide protection from accidental knocks.

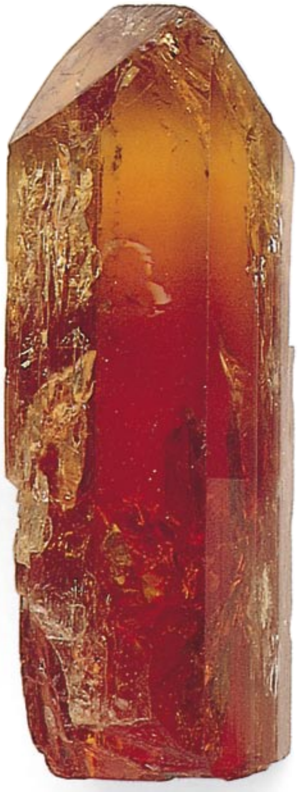

ONE OF THE BEST
The finest golden topaz crystals come from the Ouro Preto area in Brazil, and this wedge-shaped prism is typical of them. Some may show color zoning from golden brown to pink. Gems containing a hint of pink are often called imperial topaz.

TOPAZ TRICKS
Like diamond, topaz is commonly found in gravel as rounded, waterworn pebbles. It also has a specific gravity very similar to that of diamond and this has led to some premature celebrations.

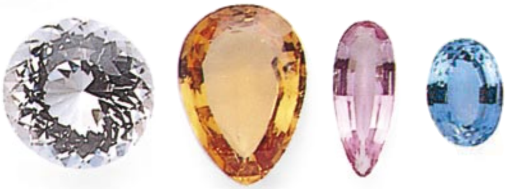

WATER COLORS
Topaz is an aluminum silicate containing about 20 percent water and fluorine. Those crystals with more water than fluorine are golden brown or, rarely, pink; those with more fluorine than water are blue or colorless.

BRAZILIAN PRINCESS
This topaz was cut in 1977 and weighed 21,327 carats. The largest cut stone today is 36,853 carats.

Tourmaline

Tourmaline is a mineral with a complex chemistry. It crystallizes as prisms with flat or wedge-shaped terminations. Every crystal has a different structure at each end, sometimes indicated by different colors. This gives tourmaline an unusual electrical property. If a crystal is gently warmed, one end becomes positively charged and the other negatively charged, which is the reason for its tendency to attract dust.

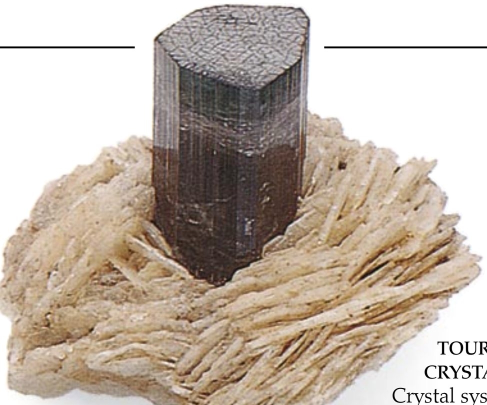

TOURMALINE CRYSTAL
Crystal system: trigonal; hardness: 7–7.5; specific gravity: 3–3.25.

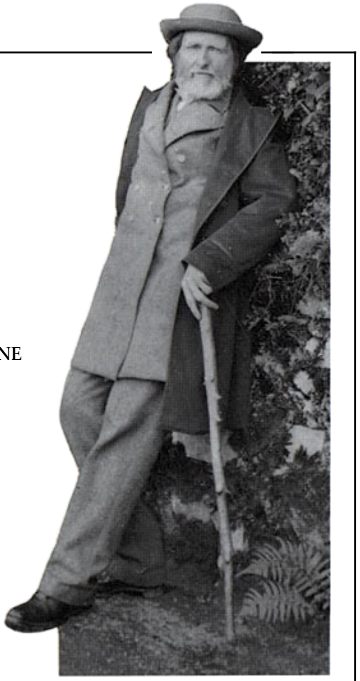

MEDICINE OR MINERAL
Philosopher John Ruskin, shown here in 1885, wrote "the chemistry of tourmaline is more like a medieval doctor's prescription than the making of a respectable mineral!"

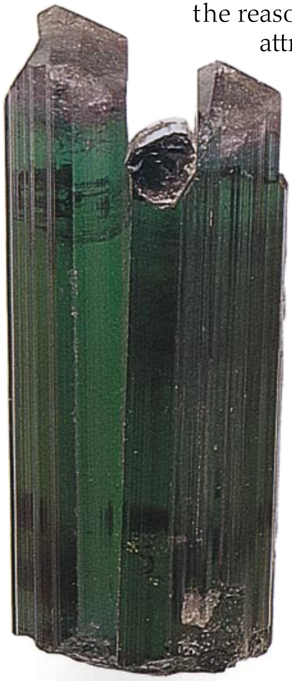

BLACK AND GREEN
Tourmaline is pleochroic, which means it is a different color when viewed down different axes (pp. 12–13). These green crystals would be almost black if viewed down the long axis.

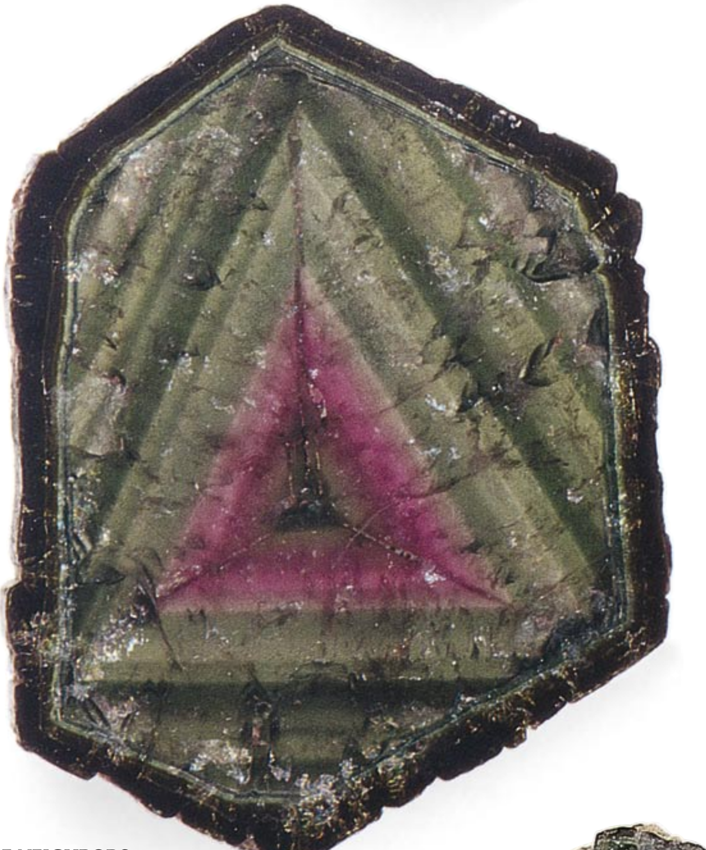

FRAMED UP
This slice across a tourmaline prism shows typical three-fold symmetry and a triangular cross-section. The zones of color illustrate how the crystal was built up in layers, each layer being a different phase of crystallization. The final shape of the crystal is controlled by the last phase, which in this case formed a hexagonal "frame."

The growth rings seen in some crystals are similar to those of tree trunks

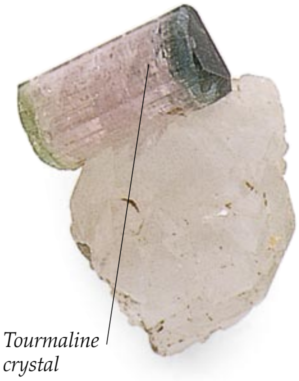

CLOSE NEIGHBORS
This tourmaline is unusual in being attached to its neighbor (quartz) by a prism face. The pink prism crystallized first, then green tourmaline formed the terminations.

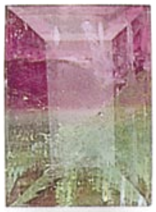

Cut stone showing the two colors of watermelon tourmaline

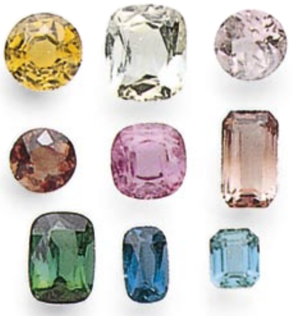

MULTICOLORED
Tourmaline shows the greatest color range of any gemstone, and some crystals are themselves more than one color. Watermelon tourmaline has pink cores and green outer zones.

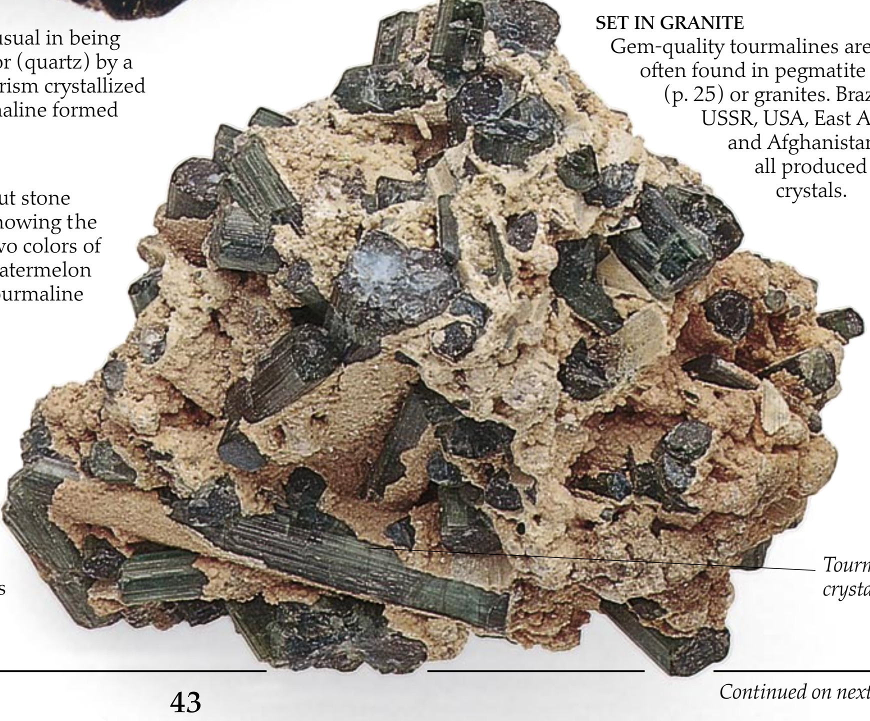

SET IN GRANITE
Gem-quality tourmalines are most often found in pegmatite veins (p. 25) or granites. Brazil, USSR, USA, East Africa, and Afghanistan have all produced fine crystals.

Continued on next page

Continued from previous page

Garnet

Garnet is the name of a family of chemically related minerals that includes almandine, pyrope, spessartine, grossular, and andradite. They can all be found as gemstones, the almandine-pyrope group being the most widely used. Because of the different chemical compositions, garnet occurs in most colors other than blue. Sources of gem-quality material include Czechoslovakia, South Africa, USA, Australia, Brazil, and Sri Lanka.

Ring set with almandine garnet

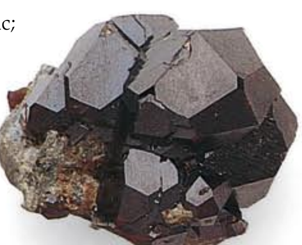

GARNET CRYSTALS
Crystal system: cubic; hardness: 6.5–7.5; specific gravity: 3.52–4.32.

Cut demantoid garnet

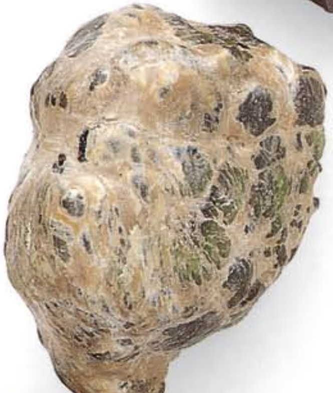

DEMANTOID
Emerald-green demantoid is the most prized of all the garnets. The finest stones come from the Ural Mountains in the USSR.

Cut pyrope garnet

PYROPE
The deep-red garnet, pyrope, was popular in the 19th century. Most stones came from Bohemia.

Spessartine cabochon

SPESSARTINE
The beautiful orange colors of spessartine are caused by manganese. Spessartine is not often seen in jewelry, as gem-quality crystals are rare.

FRUITY NAME
"Garnet" may come from *pomum granatum*, Latin for pomegranate. The gem color of the almandine-pyrope group is similar to that of the pomegranate.

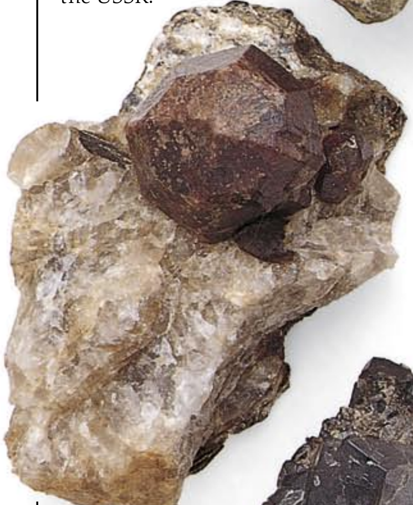

ALMANDINE
Garnet commonly crystallizes as icositetrahedrons, like these almandine crystals. Almandine usually has a very deep color so it is often cut as cabochons (p. 59) with the backs hollowed out to make it more transparent.

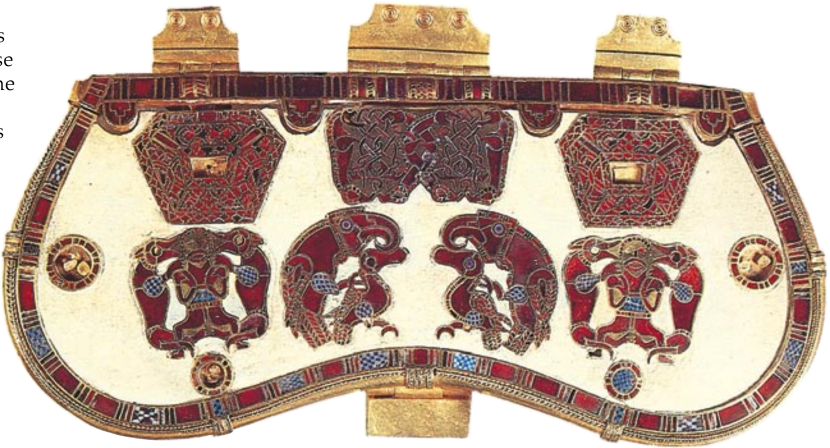

FIT FOR A KING
This fine 7th-century purse lid was among many garnet-set pieces found in an Anglo-Saxon royal burial ship in Sutton Hoo, Suffolk, England. The quality of the workmanship of all the pieces indicates the high status of their owner.

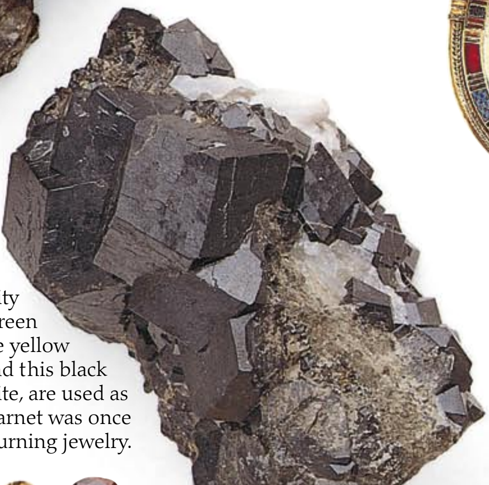

ANDRADITE
Most andradite garnet does not occur as gem-quality crystals. Only green demantoid, the yellow variety topazolite, and this black variety, melanite, are used as gemstones. Black garnet was once used as mourning jewelry.

Cut grossular garnets

COLOR TRACES
Brilliant-green grossular contains trace amounts of vanadium, while the yellow and red stones contain iron. The red variety is known as hessonite.

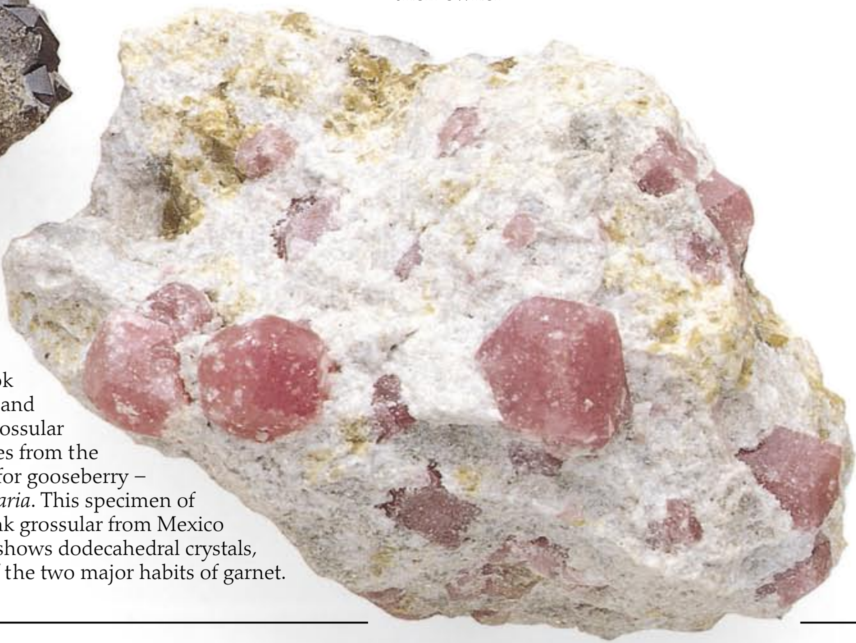

GROSSULAR
Some grossular garnet is said to look like gooseberries and the name grossular probably comes from the Latin for gooseberry – *grossularia*. This specimen of pink grossular from Mexico clearly shows dodecahedral crystals, one of the two major habits of garnet.

Peridot

Peridot is a French word and may derive from the Arabic *faridat*, meaning a gem. It is the gem variety of the mineral olivine, a magnesium and iron silicate that is common in volcanic rocks.

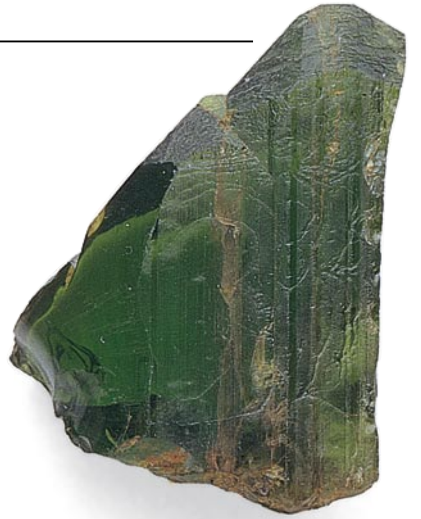

PERIDOT CRYSTAL
Crystal system: orthorhombic; hardness: 6.5; specific gravity: 3.22–3.40.

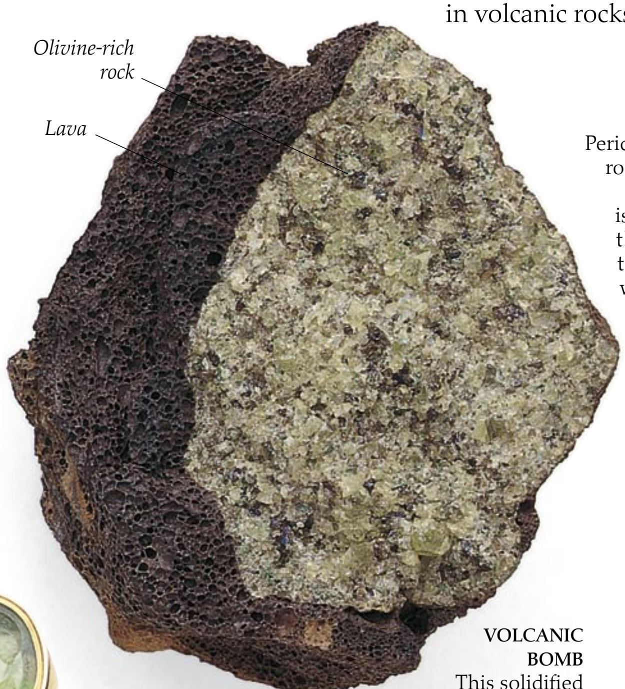

VOLCANIC BOMB
This solidified lava contains fragments of rock rich in olivine. The lava came from deep within the earth, carrying the rock with it, and the whole piece was ejected through a volcano as a volcanic bomb.

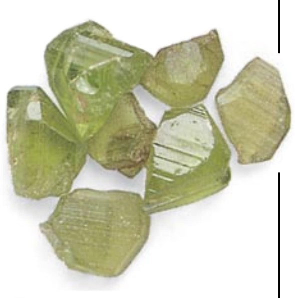

ISLAND GEM
Peridot usually occurs in rocks intergrown with other minerals. The island of Zabargad in the Red Sea is one of the few places in the world where crystals with distinct faces, such as these, are found.

Cut peridot from Arizona

Cut peridot from Myanmar

Cut peridot from Norway

PERIDOT SUPPLIERS
The largest peridots come from Zabargad and Myanmar, but Arizona, Hawaii, and Norway have also supplied fine gems.

Ring set with peridot

NAME CHANGE
There have been peridot mines on the island of Zabargad in the Red Sea for a long time. The stones from here were known by the ancient Greeks as *topazos* (p. 42).

Moonstone

Moonstone is the best-known feldspar gem. Feldspars are common in rocks, but rarely of gem quality. There are two main groups: one which is rich in potash and includes the moonstones; one which is rich in soda and calcium and includes sunstone. They range in hardness from 6–6.5 and in specific gravity from 2.56–2.76.

SUNSTONE
The bright spangles in sunstones are reflections from tiny dark-red flakes of hematite. Some are arranged parallel, giving extra brightness in some directions.

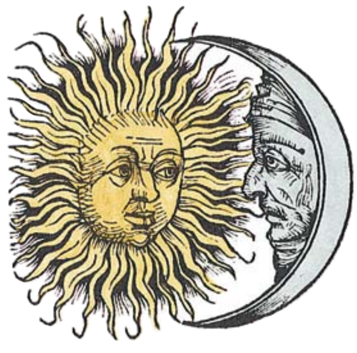

The sun and the moon – appropriate names for these two stones

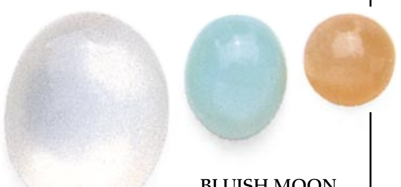

BLUISH MOON
Most moonstones are colorless with a silvery or bluish sheen, but some varieties may be steely gray, orange pink, yellow, or pale green. The gray stones particularly may show good cat's-eyes (p. 59).

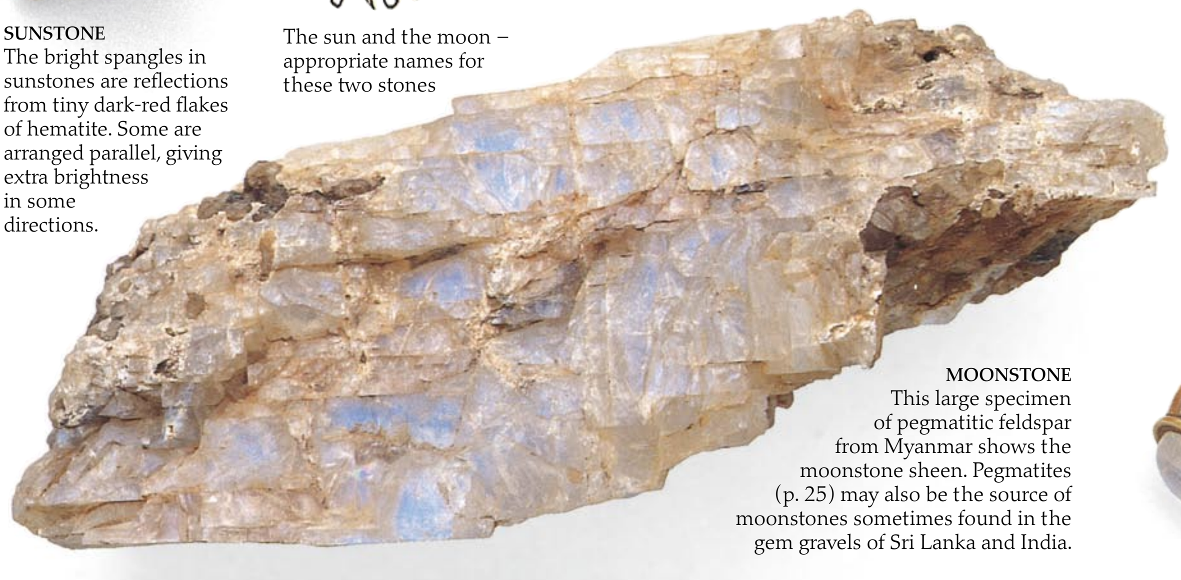

MOONSTONE
This large specimen of pegmatitic feldspar from Myanmar shows the moonstone sheen. Pegmatites (p. 25) may also be the source of moonstones sometimes found in the gem gravels of Sri Lanka and India.

Pin set with sunstone

Ring set with moonstone

Continued on next page

Continued from previous page

Spinel

The most beautiful red and blue spinels can rival ruby and sapphire in their richness. Until the 19th century, red spinels were called balas rubies, which led to some confusion. The scientist Romé de l'Isle (p. 12) was the first to distinguish clearly between true ruby and red spinel. The term *balas* may relate to a source of these stones in Balascia, now called Badakhshan, in Afghanistan.

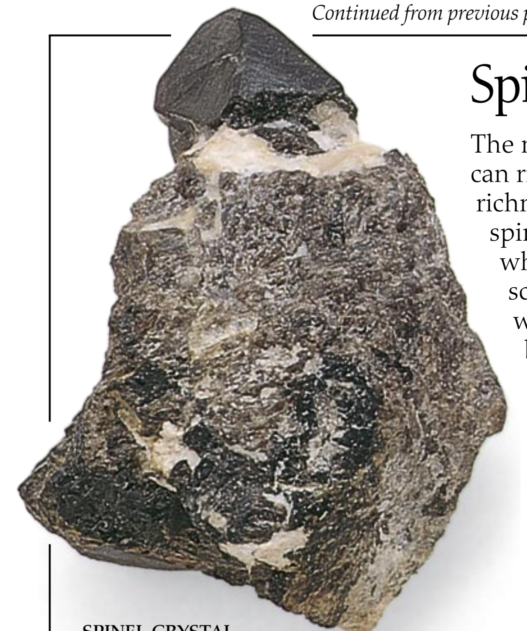

SPINEL CRYSTAL
Crystal system: cubic; hardness: 8; specific gravity: 3.5–3.7.

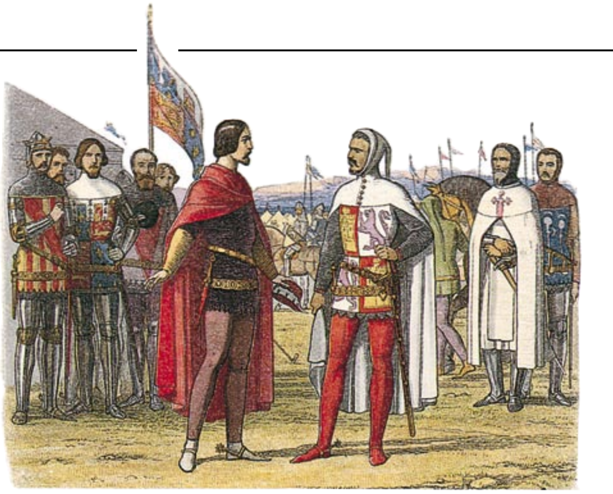

PRINCELY REWARD
The Black Prince, son of King Edward III of England, helped Pedro the Cruel, King of Castile, Spain, to win the Battle of Najera in 1367. He was rewarded with a balas ruby, now in the British imperial state crown.

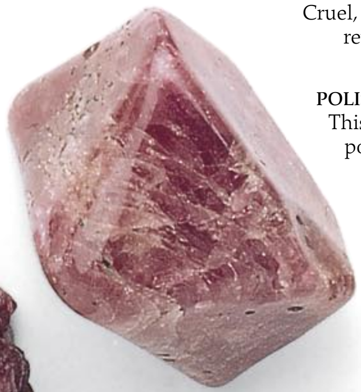

POLISHED OVER
This crystal has been polished to remove surface blemishes, but it still has its original octahedral (eight-faced) shape.

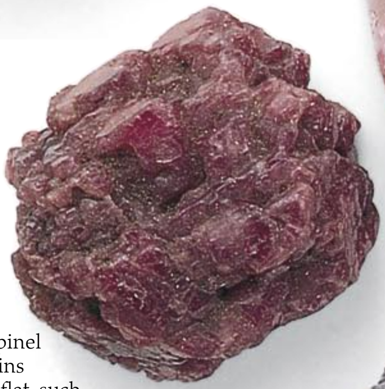

SMALL DISTORTION
Spinel usually crystallizes as octahedra. This specimen is a crystal aggregate (p. 22) of small distorted octahedra in parallel growth. Spinel often occurs as twins (p. 21), and when flat, such crystals are called macles.

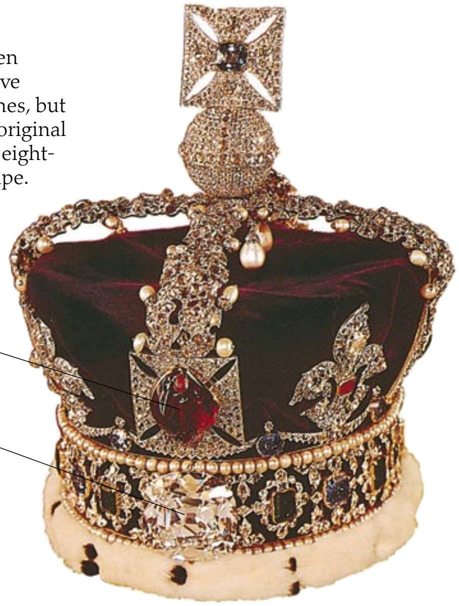

LYING IN STATE
The Black Prince's ruby in the British imperial state crown is really a 170-carat spinel, once called a balas ruby. It is mounted above another famous stone, the Cullinan II diamond (p. 35). The Timur ruby, which belongs to England's Queen Elizabeth II, is also a spinel.

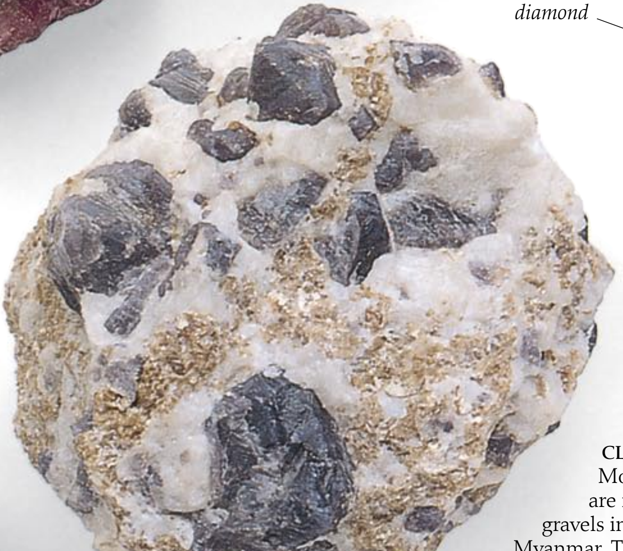

REFORMED CHARACTER
This specimen from Lake Baikal, USSR, contains octahedra of blue spinel set in a matrix, or network, of white calcite and shiny muscovite mica. Originally it was probably an impure limestone which completely recrystallized under moderate temperature and pressure.

CLOSE TO HOME
Most gem spinels are recovered from gravels in Sri Lanka and Myanmar. These Myanmar fragments are not very worn, indicating that they have not traveled far from their source.

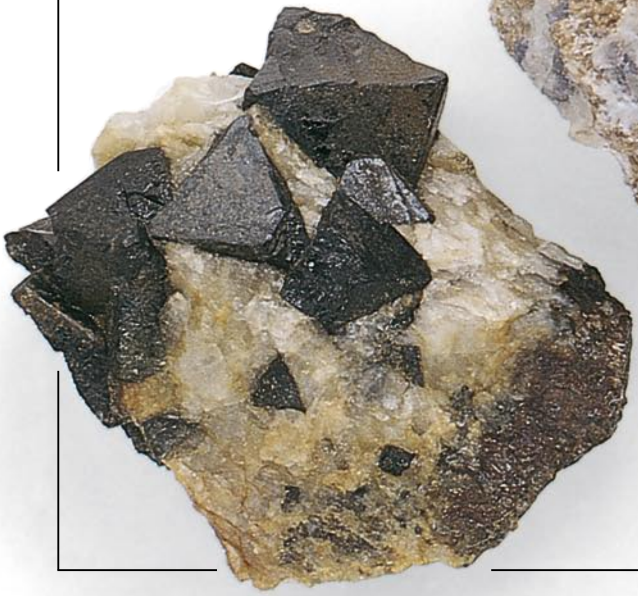

THORNY CRYSTALS
These beautifully crystallized octahedra from Bodenmais in Germany are gahnite, a zinc-rich variety of spinel. They show the typical triangular-shaped crystal faces of spinel which may account for the possible derivation of its name from the Latin *spina*, meaning thorn.

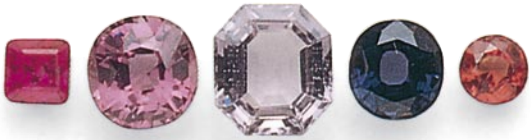

CRYSTAL COLORS
Pure spinel is colorless. The beautiful reds and pinks are due to small amounts of chromium in the crystals. Blues and greens are caused by iron and, rarely, by zinc.

Zircon

The name zircon comes from the Arabic *zargoon*, meaning vermilion or golden-colored. Sri Lanka has been a source of zircons for 2,000 years, but today stones also come from Thailand, Australia, and Brazil. Colorless zircon looks like diamond in luster and fire and is used as a diamond simulant (p. 57), but it is softer and may look "sleepy" due to inclusions (p. 21) and double refraction (p. 19).

The color of red hyacinths may account for red zircon once being known as hyacinth

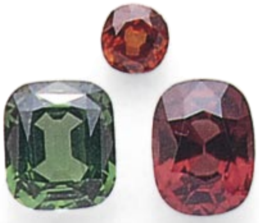

NATURAL COLORS
Zircon is zirconium silicate, colorless when pure, but found in a wide range of colors in nature because of different impurities.

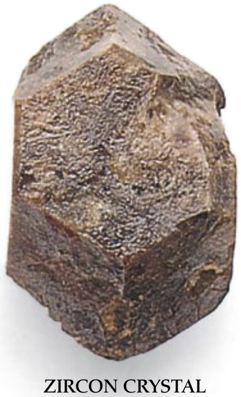

ZIRCON CRYSTAL
Crystal system: tetragonal; hardness: 6–7.5; specific gravity: 4.6–4.7.

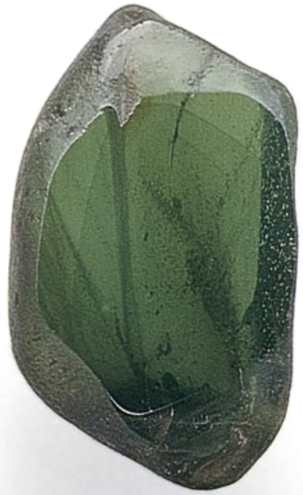

RADIOACTIVE
This exceptionally large pebble from Sri Lankan gem gravels shows a typical zircon color. Some zircons contain so much uranium and thorium that the radioactivity of these elements breaks down the crystal structure and the stone becomes amorphous, or noncrystalline.

HEAT TREATMENT
Colorless, blue, and golden zircons can be produced by heating red-brown crystals. Heating in an oxygen-free atmosphere produces blue zircon; heating in air, i.e. with oxygen, produces a golden color. Some colorless stones are produced by both methods. These colors may fade but can be restored by reheating.

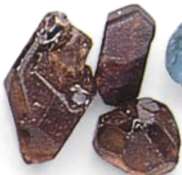

Natural brown zircon crystals

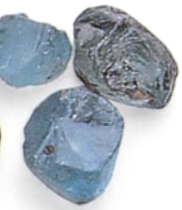

Heat-treated blue zircon crystals

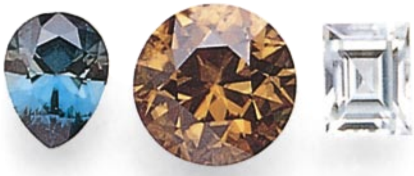

Stones cut from heat-treated zircon

Chrysoberyl

Gem chrysoberyl is exceeded in hardness only by diamond and corundum. The yellow, green, and brown colors are caused by iron or chromium. There are three varieties: clear yellow-green gems; cat's-eye, or cymophane, usually cut as cabochons to display the "eye" effect (p. 59); and alexandrite, famous for its dramatic color change. Sri Lanka and Brazil are sources for all three, but the best alexandrites come from the USSR.

ALEXANDRITE
Alexandrite was discovered in the Ural Mountains, USSR, on the birthday of Czar Alexander II in 1830, hence the name. They appear a deep green in daylight and red in artificial light, matching the Russian imperial colors.

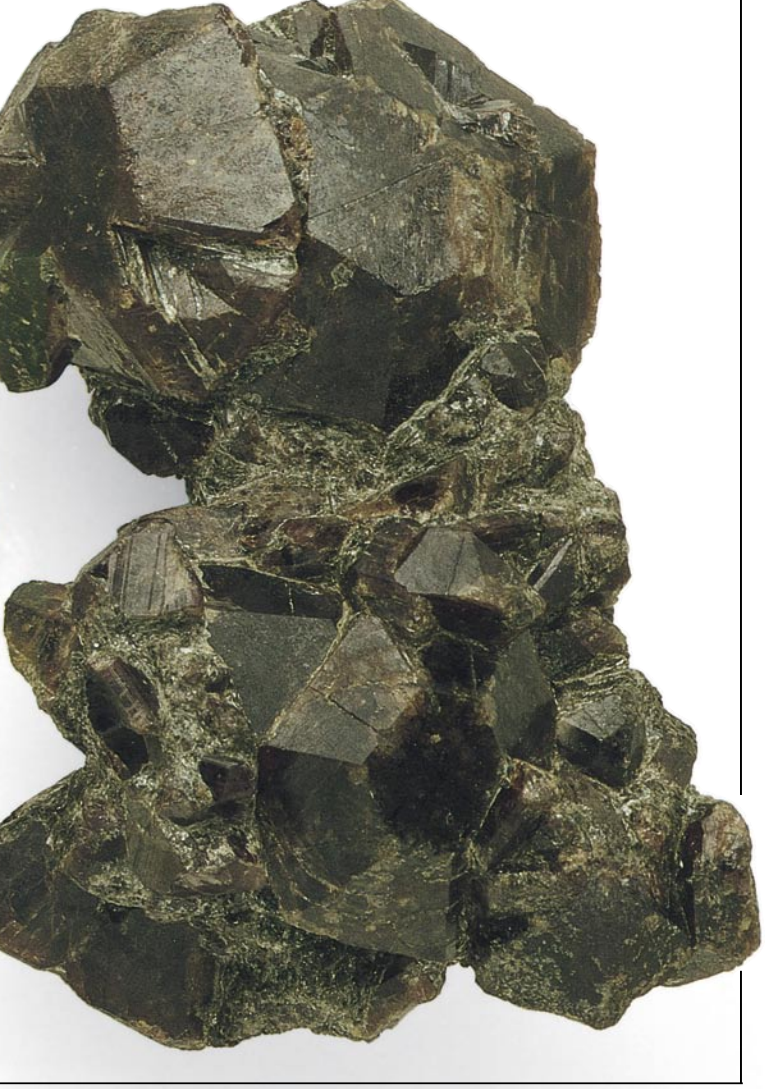

Cut yellow chrysoberyl

Cut alexandrite

POPULAR IN PORTUGAL
Yellow-green chrysoberyls were found in Brazil in the 18th century. These became very popular in Portuguese and Spanish jewelry.

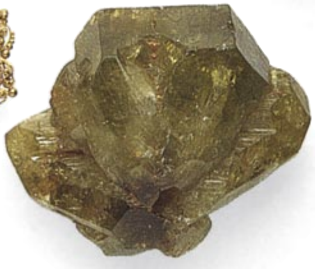

CHRYSOBERYL CRYSTAL
Crystal system: orthorhombic; hardness:8.5; specific gravity: 3.68–3.78.

Collectors' items

With over 3,000 species of minerals to choose from, the potential number of gems would appear at first to be very large. But crucial factors such as hardness (pp. 18–19), durability, and rarity reduce the number of commercial gems to a few dozen. Many people collect rarities that are not in general circulation. They may seek the rare colors or exceptional sizes of common gems, or cut examples of minerals too soft or fragile to wear in jewelry. For example, blende and sphene are available in reasonable quantities but are too soft for constant wear. Benitoite is durable enough to be worn but is too rare to be generally available.

AXINITE
The most beautiful wedge-shaped crystals of brown axinite come from Bourg d'Oisans in France and display flashes of gray and violet in different directions. Although they were once extremely rare, crystals are being recovered more regularly from the Sri Lanka gem gravels.

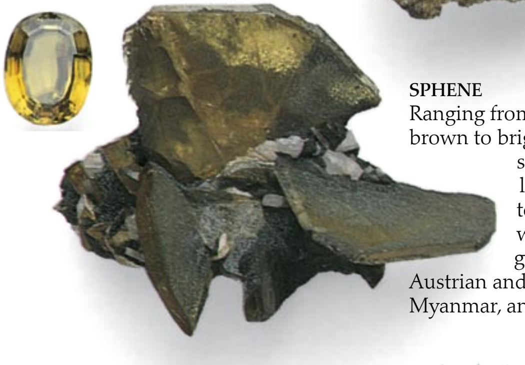

SPHENE
Ranging from golden yellow brown to bright emerald green, sphene has great luster and fire but is too soft for general wear. The finest gems come from the Austrian and Swiss Alps, Myanmar, and California.

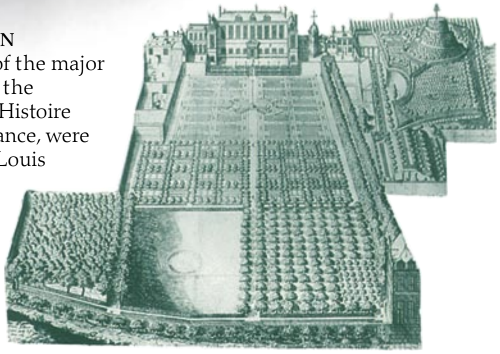

FRENCH COLLECTION
The first specimens of the major mineral collection in the Muséum National d'Histoire Naturelle in Paris, France, were brought together in Louis XIII's pharmacy and botanical gardens. This engraving shows the gardens as they were in 1636, seven years before the king died.

ALPINE EXPERTS
Many fine crystals are collected from crystal-lined clefts in the Alps by people known as *strahlern. Strahlern* are experienced mountaineers but also talented mineral collectors, usually with great knowledge of a particular Alpine area.

TANZANITE
The purplish blue gem variety of the mineral zoisite is tanzanite. It was found in northern Tanzania in 1967 and is remarkable for its displays of rich blue, magenta, and yellowish gray. Many crystals are greenish gray and are heat-treated to the more attractive blue.

DANBURITE
The mineral danburite is named after Danbury, Connecticut, where it was first found as colorless crystals in a pegmatite (p. 25). Fine yellow stones come from Madagascar and Myanmar, and colorless stones from Japan and Mexico.

CORDIERITE
Fine, gem-quality cordierite comes from Sri Lanka, Myanmar, Madagascar, and India. Cordierite is exceptional in showing very strong pleochroism (p.36), from deep purplish blue in one direction, to pale yellowish gray in another. This pleochroism was used by the Vikings for navigating their long boats (p. 60) and has also led to the crystals being called water sapphires.

BENITOITE

The color of benitoite crystals can be compared with that of fine sapphires (pp. 36–37). They display similar fire to diamonds (pp. 34–35), but remain collector's stones because of their rarity. Benitoite comes from localities in San Benito County, California, after which it is named (p. 11).

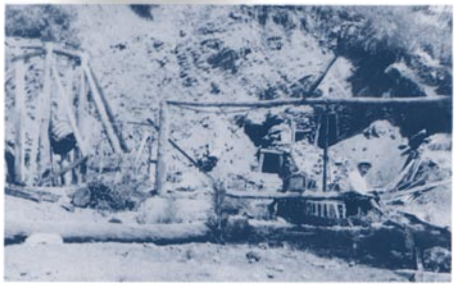

The San Benito mine in 1914 showing the open cut and an ore bucket on the left

BLENDE

The popular name for sphalerite, the world's major source of zinc, is blende. Normally it is opaque gray to black, but gem-quality reddish brown, yellow, and green crystals come from Mexico and Spain. The rich colors are very attractive, but the stones are too soft to be used in jewelry.

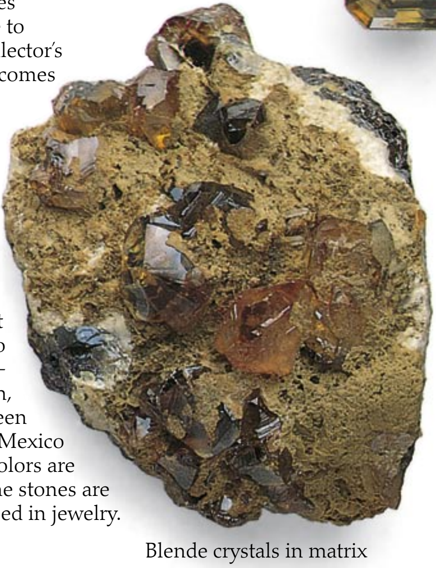

Blende crystals in matrix

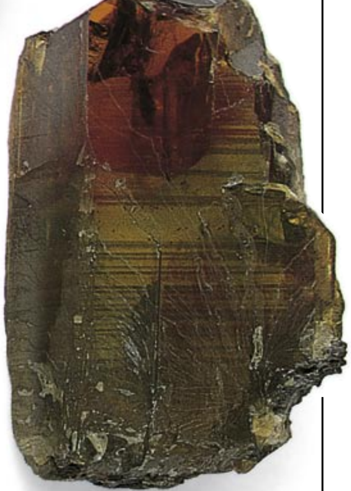

Rough blende crystal

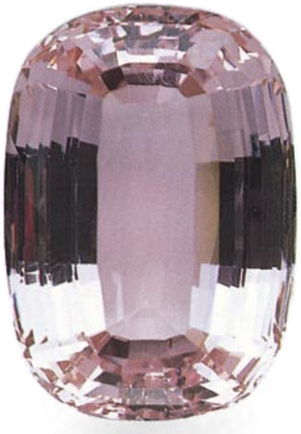

Cut kunzite

Cut pale-green spodumene

Kunzite crystal

SPODUMENE

Magnificent crystals of spodumene come from Brazil, California, and Afghanistan. Fine gems weighing hundreds of carats have been cut from pale-green and yellow crystals, and from the pink variety kunzite, named after G. F. Kunz. Small crystals of a rare, emerald-green variety called hiddenite are found in North Carolina and Sri Lanka.

George Frederick Kunz, an author on gems, who worked for the New York jewelers Tiffany's early this century

GOLDNEY GROTTO

Precious stones and corals are among items collected to cover the walls and pillars of this underground grotto. It was built between 1737 and 1764 near Bristol, England.

SCAPOLITE

Myanmar and East Africa are sources of scapolite gems. They occur in pastel shades of yellow, pink, purple, and fine cat's-eyes (p. 59).

SINHALITE

Originally traded as peridot from Sri Lanka, sinhalite was proved in 1952 to be a new species. Mineralogists in the British Museum named it after an old name for Sri Lanka – Sinhala.

FIBROLITE

This bluish violet stone of 19.84 carats is fibrolite, a rare variety of the mineral sillimanite. It comes from Myanmar and is one of the largest in the world. Andalusite has the same chemical composition – aluminum silicate – but a different structure. The gem-quality stones show bold red and green pleochroic colors. Fine examples come from Brazil and Sri Lanka.

Cut fibrolite

Cut andalusite

Stones for carving

Microcrystalline rocks and minerals have been used in decoration for thousands of years. The best known are the jades, lapis lazuli, and turquoise, and there are many more which are suitable for carving work. Ancient civilizations such as the Egyptians, Chinese, and Sumerians used jade, lapis, and turquoise to make jewelry. South American Indians and the Maoris of New Zealand have been carving turquoise and jade for centuries.

MALACHITE
Malachite is a vivid green copper mineral. It is often found as kidney-shaped masses surrounded by bands of color. It is 4 on the hardness scale (pp. 18–19) and has a specific gravity of 3.8. Zaire, Zambia, Australia, and the USSR are the main sources.

TURQUOISE TRADITION
Traditional Indian jewelry has been made for thousands of years in the southwestern USA, where most turquoise is still produced.

Turquoise

The word *turquoise* comes from the French *pierre turquoise*, meaning stone of Turkey; in the past most turquoise was sold in Turkey. It occurs in nodules and veins of green or blue. Copper makes it blue; iron makes it green. It has a specific gravity of 2.6–2.9 and a hardness of 5–6.

Lapis lazuli

Lapis lazuli is not a single mineral but a rock consisting of blue lazurite with variable amounts of calcite and pyrite. The best, from Afghanistan, consists mostly of lazurite and is deep blue. It is 5.5 on Mohs' hardness scale and has a specific gravity of 2.7–2.9. There are other sources in the Soviet Union and Chile.

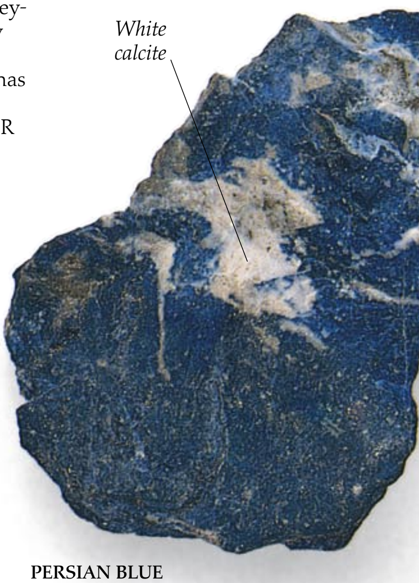

PERSIAN BLUE
The name lapis lazuli is derived from the Persian word *lazhward*, meaning blue. The blue color is caused by sulfur which is an essential part of its composition.

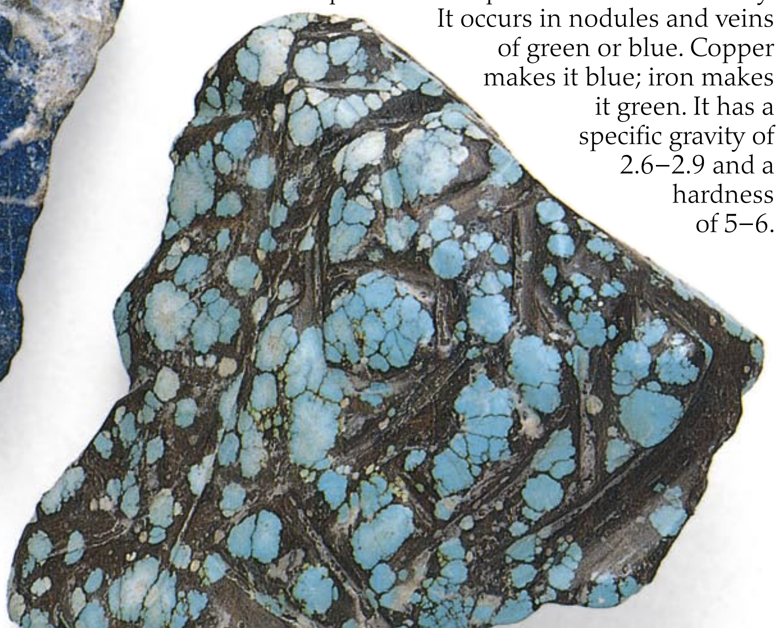

NATURAL MOSAIC
Turquoise is rare in large masses and is more often found forming mosaics. The finest blue turquoise comes from Iran (previously called Persia), where it has been used in decoration for almost 6,000 years.

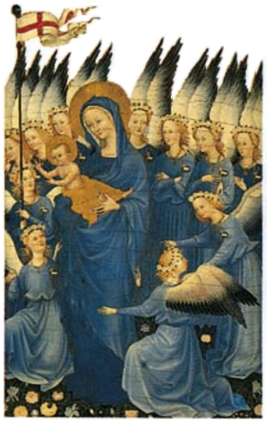

MEDIEVAL PAINTING
In medieval times lapis lazuli was crushed and purified to make the paint pigment ultramarine. It was used to paint the Wilton Diptych, a detail of which is shown here. This famous altarpiece is now in the National Gallery in London, England.

POPULAR JEWEL
Lapis lazuli has been used extensively for beads and other pieces of jewelry.

Blue Persian turquoise engraved and inlaid with gold

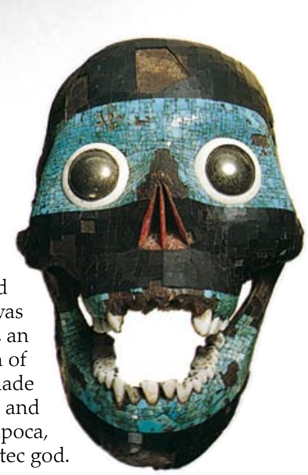

ANCIENT SKULL
This mask, shaped around a human skull, was made by the Aztecs, an ancient civilization of Central America. It is made of turquoise and lignum, and may represent Tezcatlipoca, an important Aztec god.

Jade

The Spanish conquerors of Mexico believed that the Indians' green stones would cure kidney ailments. They called them kidney stones or *piedras de hijadas* and from this the word *jade* was derived. In Europe, the name was then given to material of the same color and hardness which was imported from China. It was only in 1863 that they were proved to be two different minerals now called jadeite and nephrite.

LIFE JACKET
The ancient Chinese believed that jade had the power to give life and used it to try to preserve the dead. They linked plates of nephrite around the corpse to make a funeral suit. This one belonged to a princess of the 2nd century B.C. and is linked with gold.

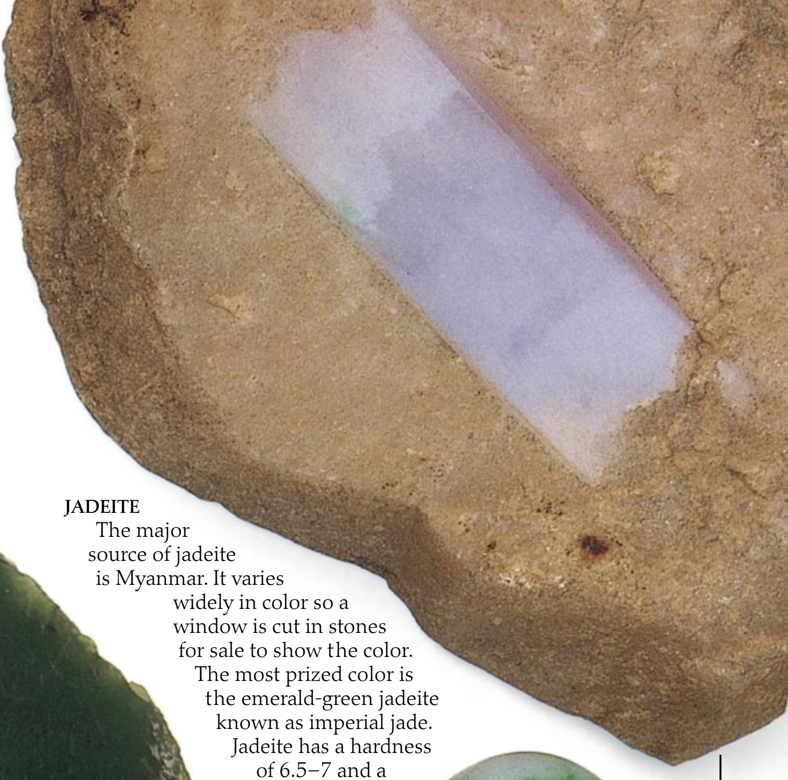

JADEITE
The major source of jadeite is Myanmar. It varies widely in color so a window is cut in stones for sale to show the color. The most prized color is the emerald-green jadeite known as imperial jade. Jadeite has a hardness of 6.5–7 and a specific gravity of 3.3–3.5.

Jadeite fashioned into a ball

NEPHRITE
Nephrite is made of interlocking grains, making the stone tough. The "greenstone" used by the Maoris of New Zealand is nephrite. It has a hardness of 6.5 and a specific gravity of 2.9–3.1.

Nephrite snail designed by the famous Russian jeweler Fabergé

CHINESE CAMEL
This nephrite camel was carved in China. White and cream nephrites contain very little iron. More iron causes the spinach-green stones of the USSR, Canada, and New Zealand and the black jade of South Australia.

Other stones

Many other stones are popular for carving, mainly because of their color. These include malachite, serpentine, blue john, and rhodonite as well as marbles and alabaster.

RHODONITE
The bright pink color of rhodonite (*rhodo* means pink) is caused by manganese. It has a hardness of about 6 and is used for carving and inlays. Sources of gem material include USSR, Canada, and Australia.

SERPENTINE
The patterns in serpentine often look like snake-skin, and carvers can use these to create works of art. Some is soft and easy to carve, but the yellow-green variety bowenite, a favorite of Chinese carvers, is harder – up to 6 on Mohs' scale.

BLUE JOHN
The distinctive purple and pale yellow banded fluorite called blue john comes from Derbyshire in England. It is fragile, so is usually bonded with resins to make it easier to work and harder-wearing.

19th-century blue john vase

Precious metals

GOLD, SILVER, AND PLATINUM are crystalline, but single crystals are rarely found. Gold and silver were among the earliest metals worked, over 5,000 years ago. Platinum was first noted in 1735 as a white metal used by the Chibcha Indians of Colombia, and today it is more valuable than gold and silver. All three metals are useful because they are relatively soft and easy to work. They are difficult to destroy and have high SGs (p. 18).

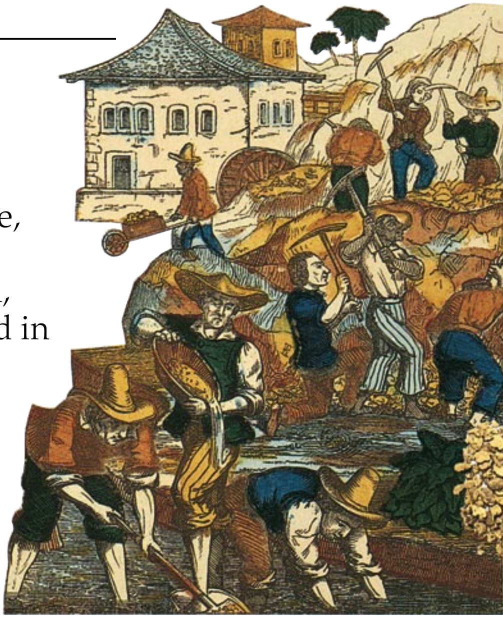

CALIFORNIA GOLD RUSH
The desire for gold has driven people to inhabit areas of the earth from the frozen Arctic to the scorching desert. Gold seekers rushed to California in 1848 and many became rich. Most of the gold was recovered from placer deposits by panning (p. 25).

WELL PLACED
These are deposits of erosion debris from gold-bearing rock. Small particles of gold can be recovered from placer deposits by washing away the sand and gravel.

Gold

Gold is used as a standard against which wealth is measured. Pure gold is a dense (SG=19.3) but soft (H=2.5–3) metal. Before it can be used it has to be refined, and for most uses it is alloyed with other metals to make it harder. Purity of gold for jewelry is measured in carats, pure gold being 24 carats.

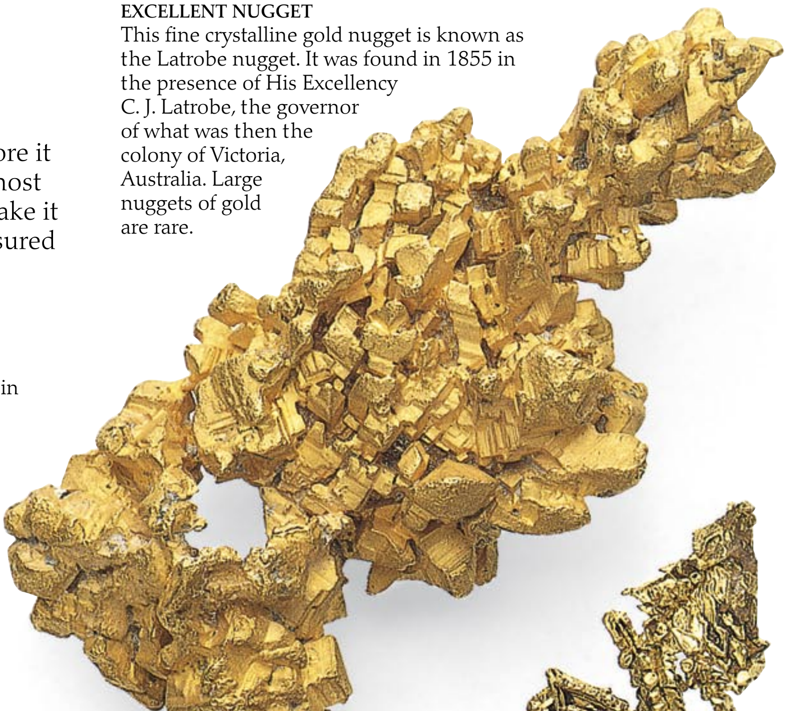

EXCELLENT NUGGET
This fine crystalline gold nugget is known as the Latrobe nugget. It was found in 1855 in the presence of His Excellency C. J. Latrobe, the governor of what was then the colony of Victoria, Australia. Large nuggets of gold are rare.

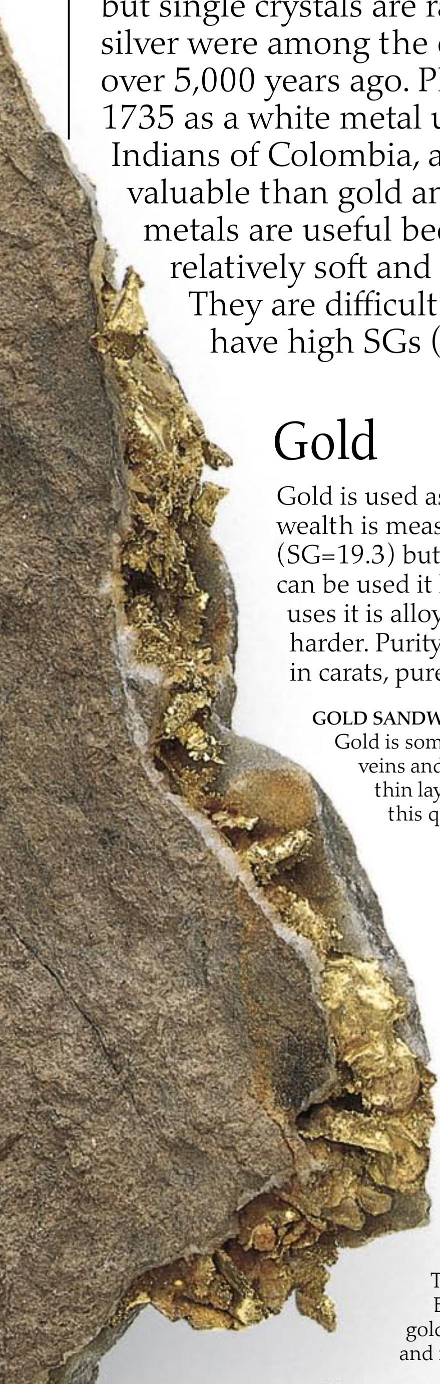

GOLD SANDWICH
Gold is sometimes found concentrated in veins and associated with quartz. A wafer-thin layer of crystalline gold can be seen in this quartz vein from New Zealand.

WORTH ITS WEIGHT
The famous golden Buddha of Bangkok is 5.5 tonnes of solid gold. It is worth over $50 million and is the most valuable religious item in the world.

BUILT ON GOLD
Between 1700 and 1900, the Asante kingdom dominated the area of Africa now known as Ghana. Its power was mostly founded upon its gold resources. Gold dust was the currency for internal trade. This lion ring is from the Asante kingdom:

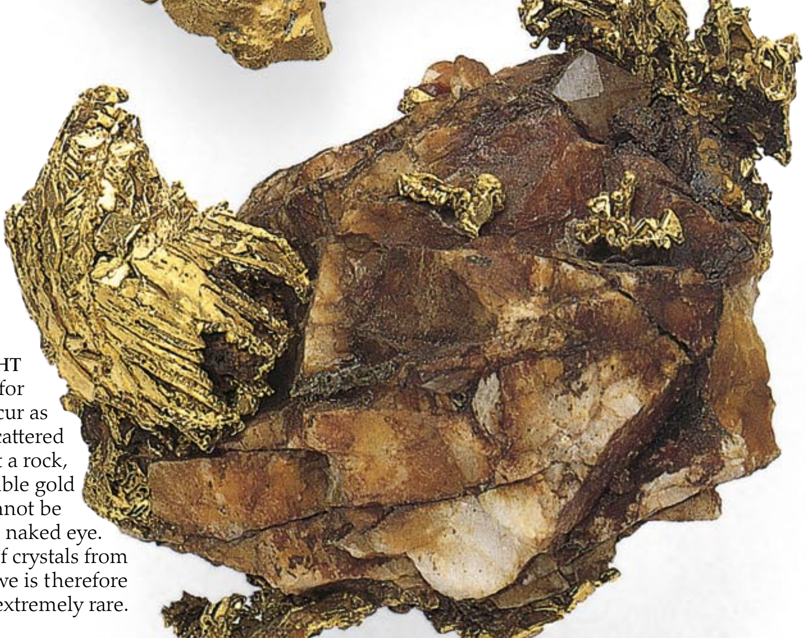

RARE SIGHT
It is usual for gold to occur as fine grains scattered throughout a rock, or as invisible gold that cannot be seen by the naked eye. This group of crystals from Zimbabwe is therefore extremely rare.

Platinum

Platinum plays a key role in modern technology. It is used as a standard weight, for surgical instruments, and of course in jewelry. The name is derived from *platina*, meaning little silver. Platinum is often found in granules or small nuggets in placer deposits. There are major deposits in the USSR, Canada, and S. Africa but most are of very low concentration.

ROUNDED
Platinum is quite soft (H=4–4.5) so it is unusual to find sharp crystals. These cubic crystals come from Sierra Leone (West Africa).

CROWN OF PLATINUM
The crown of Queen Elizabeth the Queen Mother is made of platinum. It is part of the British Crown Jewels.

RICH LAYER
This piece of platinum-bearing pyroxenite comes from a layer of igneous rock in South Africa called the Merensky reef. This layer is only about 12 in (30 cm) thick but is very rich in platinum.

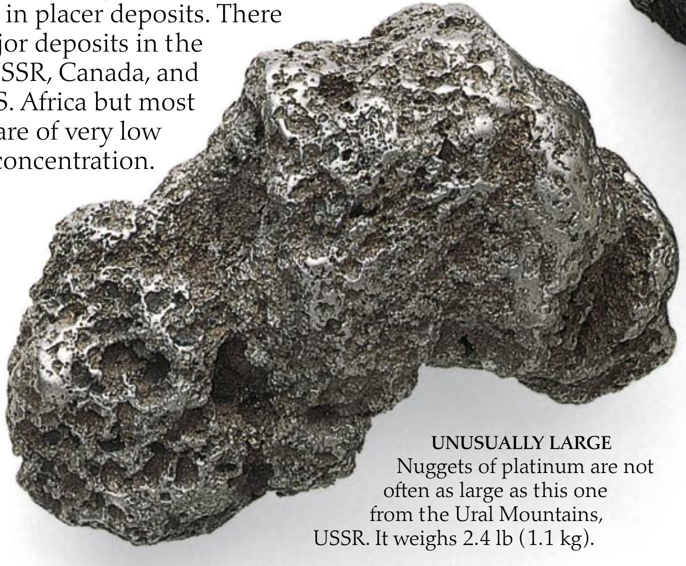

UNUSUALLY LARGE
Nuggets of platinum are not often as large as this one from the Ural Mountains, USSR. It weighs 2.4 lb (1.1 kg).

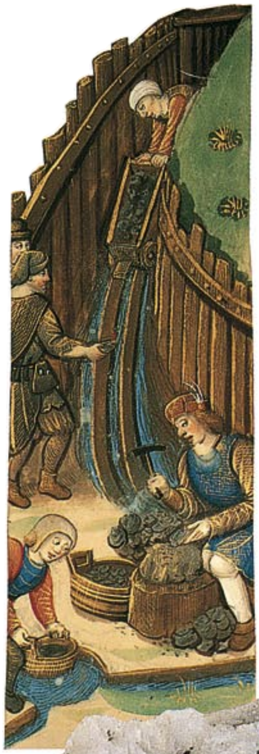

MEDIEVAL MINE
In medieval times the area around Sainte Marie, Alsace, France, was one of the richest silver-mining areas in Europe. This illustration from a medieval manuscript shows miners removing silver ore from one of the mines.

Silver

Crystals of silver are rare, but cubic crystals are occasionally found. Silver usually occurs massive or as thick wiry aggregates. In medieval times, silver was more valuable than gold. It was the main metal used for money and was also used for fine metalwork, having a hardness of only 2.5–3. Today, metallic silver is used in electronics, silver plating, and jewelry, and a huge amount is used in the photographic industry (p. 63).

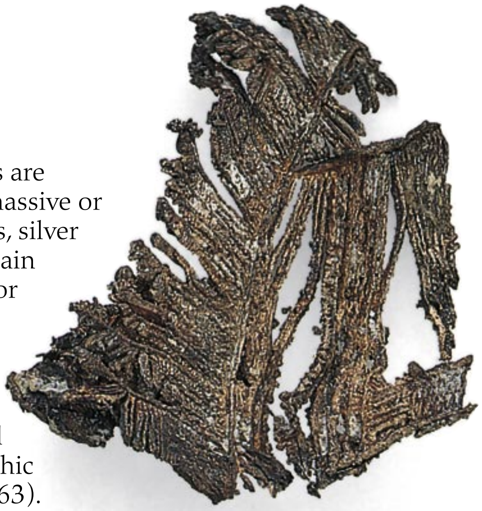

IN NEED OF A POLISH
This dendritic growth (p. 23) of silver crystals is slightly tarnished. It comes from the Huantajaya mines, in Chile.

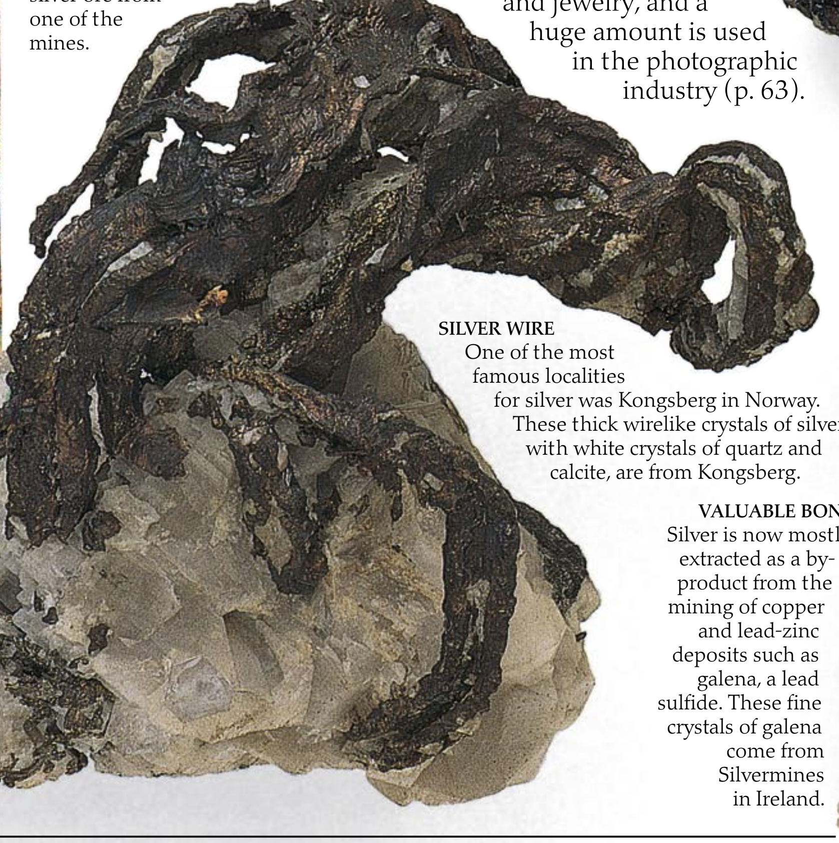

SILVER WIRE
One of the most famous localities for silver was Kongsberg in Norway. These thick wirelike crystals of silver, with white crystals of quartz and calcite, are from Kongsberg.

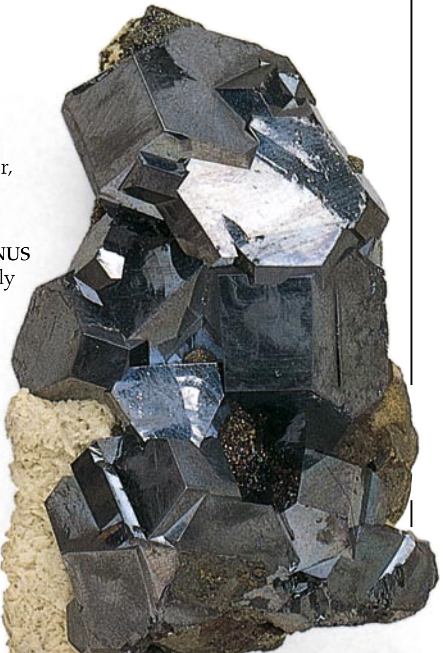

VALUABLE BONUS
Silver is now mostly extracted as a by-product from the mining of copper and lead-zinc deposits such as galena, a lead sulfide. These fine crystals of galena come from Silvermines in Ireland.

Animal and vegetable

GEMS DERIVED FROM ANIMALS AND PLANTS are described as organic. They include amber, jet, coral, pearl, and shell. They are not as hard (4 or less) or as dense (1.04 – amber; 2.78 – pearl) as gemstones but have been popular for thousands of years because of their beauty. Beads of shell and amber have been found in ancient graves dating from 2000 B.C. Pearls have long had great value as symbols of beauty and purity. The Roman emperor Julius Caesar is said to have paid the equivalent of about $250,000 for a single pearl.

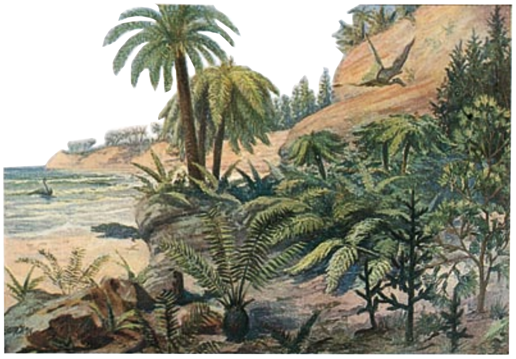

PREHISTORIC GEMS
In the Jurassic period, 160 million years ago, dinosaurs and other giant reptiles lived among the trees, which produced amber and jet.

Jet and amber

Jet and amber both come from trees. Jet is a fine-grained black rock formed over millions of years from rotted and compressed trees, in a similar way to coal. Amber is the fossilized resin, or sap, of trees that lived as much as 300 million years ago.

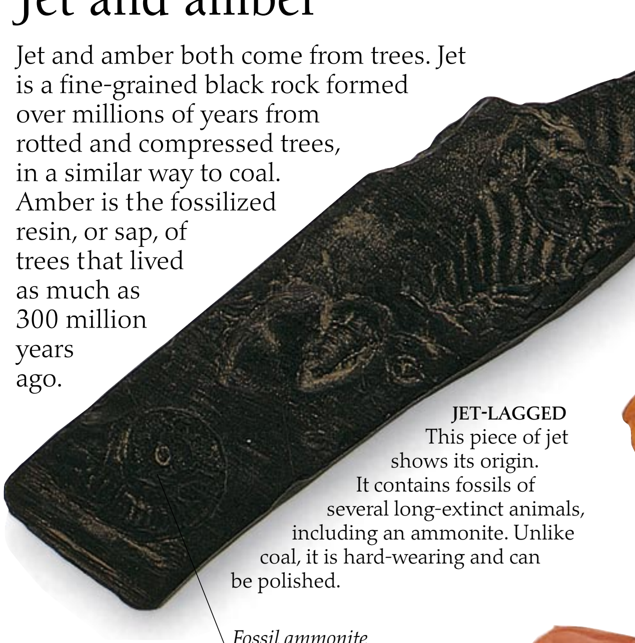

JET-LAGGED
This piece of jet shows its origin. It contains fossils of several long-extinct animals, including an ammonite. Unlike coal, it is hard-wearing and can be polished.

Fossil ammonite

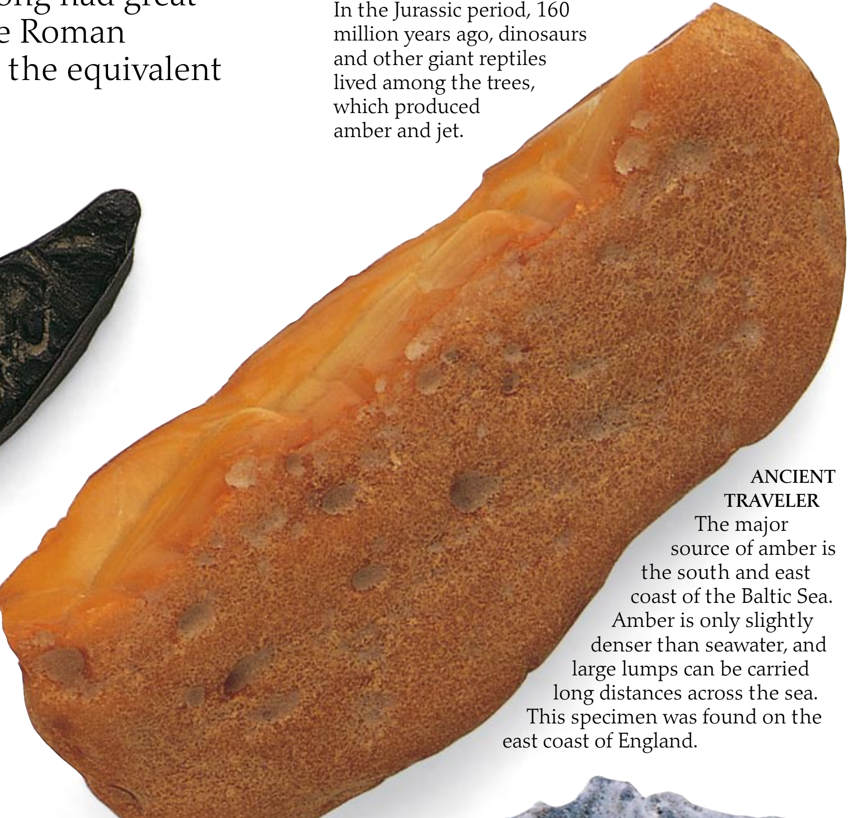

ANCIENT TRAVELER
The major source of amber is the south and east coast of the Baltic Sea. Amber is only slightly denser than seawater, and large lumps can be carried long distances across the sea. This specimen was found on the east coast of England.

Coral

Coral is a skeleton of calcium carbonate made by colonies of soft-bodied animals which live in tropical or subtropical waters. The range of colors, from black to blue to cream to red, is due to different growth conditions and organic contents.

Carving from Mediterranean coral of a monkey clinging to a branch

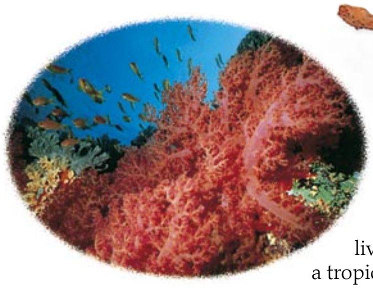

Coral living in a tropical sea

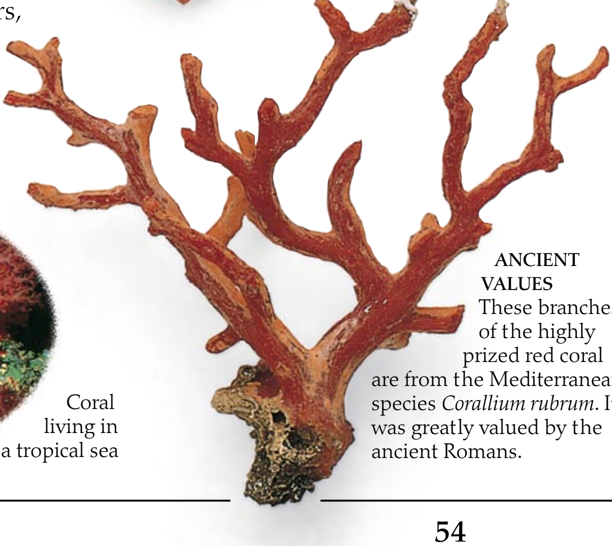

ANCIENT VALUES
These branches of the highly prized red coral are from the Mediterranean species *Corallium rubrum*. It was greatly valued by the ancient Romans.

NECKLACE MATERIAL
This blue coral comes from the species *Heliopora caerulea* which grows in the seas around the Philippines. It is often cut into beads for necklaces.

Pearl and shell

The sheen on pearls and the inside of some shells is caused by light reflecting on tiny plates of calcium carbonate called nacre. Pearls, found in some marine and freshwater shells, form when a foreign body such as a sand grain becomes lodged inside the shell. To stop the irritation, the animal slowly surrounds the grains with nacre.

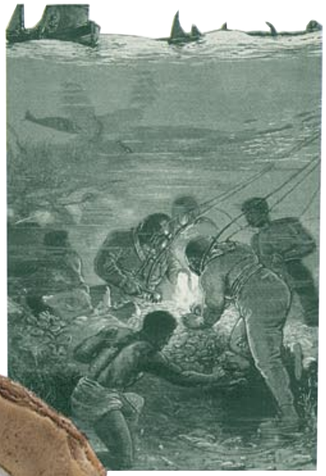

OYSTER CATCHERS
For more than 2,000 years the Persian Gulf has supplied the most beautiful natural pearls. The oysters (*Pinctada vulgaris*) used to be recovered by teams of divers. Today many pearls are cultured. An irritant is put into the oysters and the shells are farmed for their pearls.

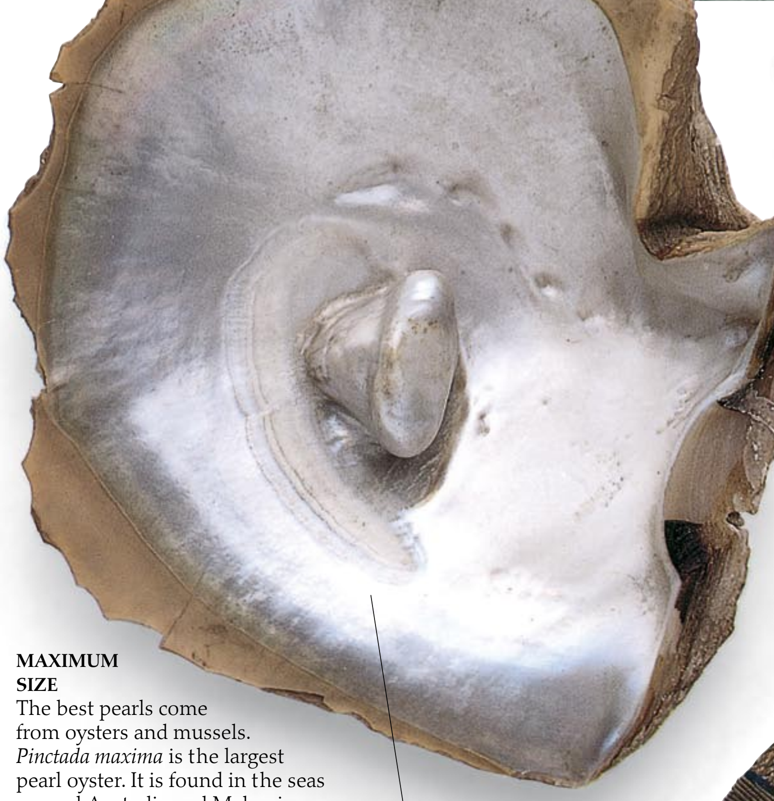

MAXIMUM SIZE
The best pearls come from oysters and mussels. *Pinctada maxima* is the largest pearl oyster. It is found in the seas around Australia and Malaysia.

Mother-of-pearl

PEARLS OF COLOR
Pearls come in many colors including blackish, golden yellow, pink, cream, and white.

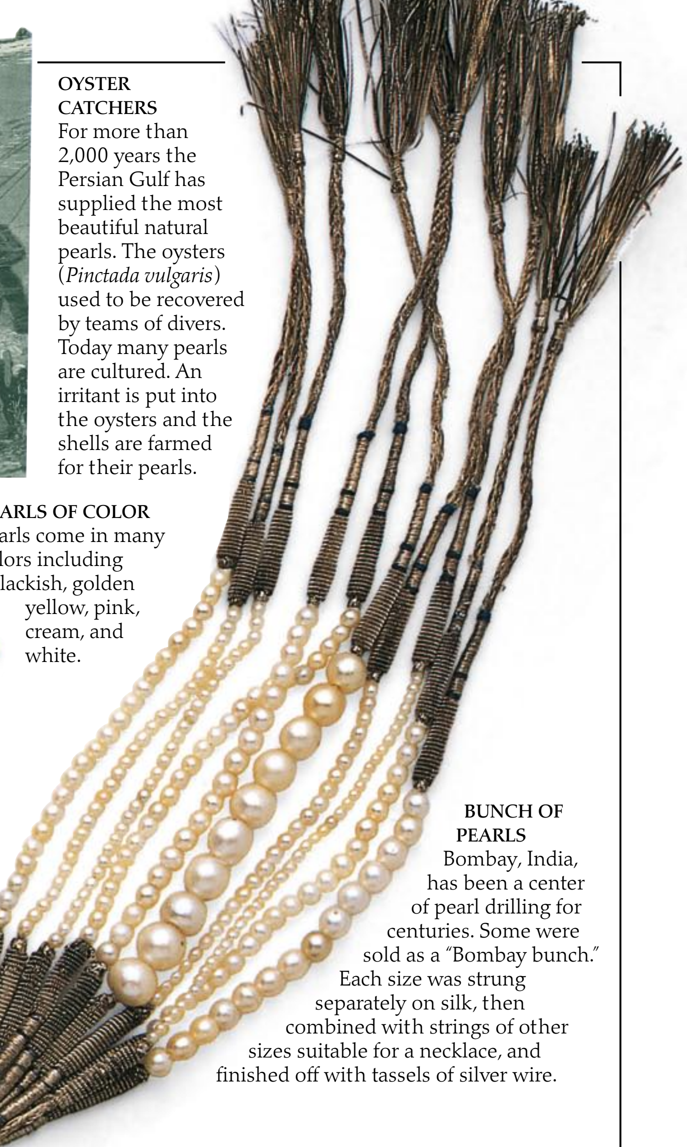

BUNCH OF PEARLS
Bombay, India, has been a center of pearl drilling for centuries. Some were sold as a "Bombay bunch." Each size was strung separately on silk, then combined with strings of other sizes suitable for a necklace, and finished off with tassels of silver wire.

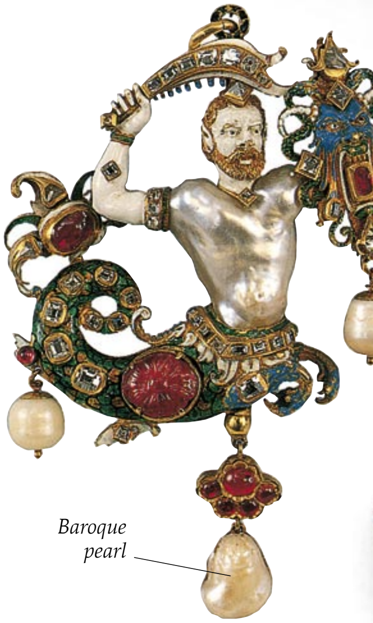

CANNING JEWEL
Irregularly shaped pearls are called baroque pearls. There are four, including a magnificent one forming the body, in the Canning Triton jewel. It was probably made in south Germany in the late 16th century and is now in the Victoria and Albert Museum in London.

Baroque pearl

Pillbox with an abalone lid

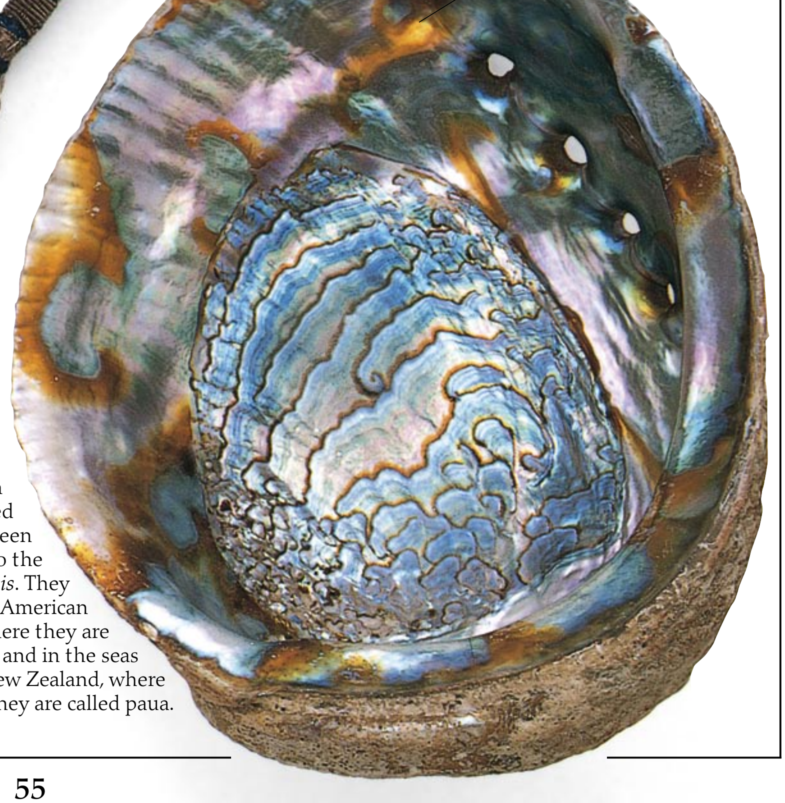

Iridescent nacre

SHELL SHINE
Shells with brightly colored blue and green nacre belong to the genus *Haliotis*. They are found in American waters, where they are called abalone, and in the seas around New Zealand, where they are called paua.

What is it worth?

THE MARKET VALUE OF GEMS plays a large part in persuading people to buy them. Fashions change, so that gems that were priceless in the 19th century may no longer be so, and vice versa. Farther back in time, lapis lazuli, turquoise, agates, and emerald were prized. From the fifth to the 14th century A.D. stones were not often cut because this was thought to destroy their magic. Since then, however, more and more stones have been cut to enhance their beauty, and they are traded to satisfy people's desire to display their style or wealth. Since medieval times diamond, ruby, pearl, emerald, and sapphire have been popular, while the popularity of species such as topaz, garnet, and aquamarine has gone up and down.

WHAT PRICE?
Gem values can vary widely, even within one species, according to color, clarity, and cut. This 57.26-carat sapphire is of such extraordinary size and fine color that it can only be valued if it changes hands.

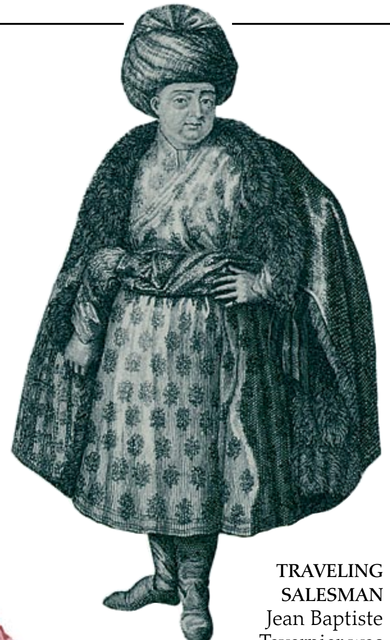

TRAVELING SALESMAN
Jean Baptiste Tavernier was a remarkable Frenchman who lived in the 17th century. He traveled in Europe and Asia, buying and selling gems. His accounts of his travels are in such detail that they are used to research the origins of famous diamonds.

Carob pod

1-carat ruby

Carob seed

WEIGHT IN BEANS
The seeds of the carob tree are of remarkably constant weight and were used for centuries as a standard for comparing the weights of precious stones. Later, a standard weight similar to that of the carob seed was used, and it was called the carat. For some time the weight was slightly different in different trading centers, but early this century a metric standard of 0.007 oz (0.2 g) was agreed internationally.

PRICELESS
Painite is priceless. Only three crystals are known, and none have yet been on the open market. It was found in Myanmar by the gem dealer A.C.D. Pain and, when confirmed as a new species, was named after him.

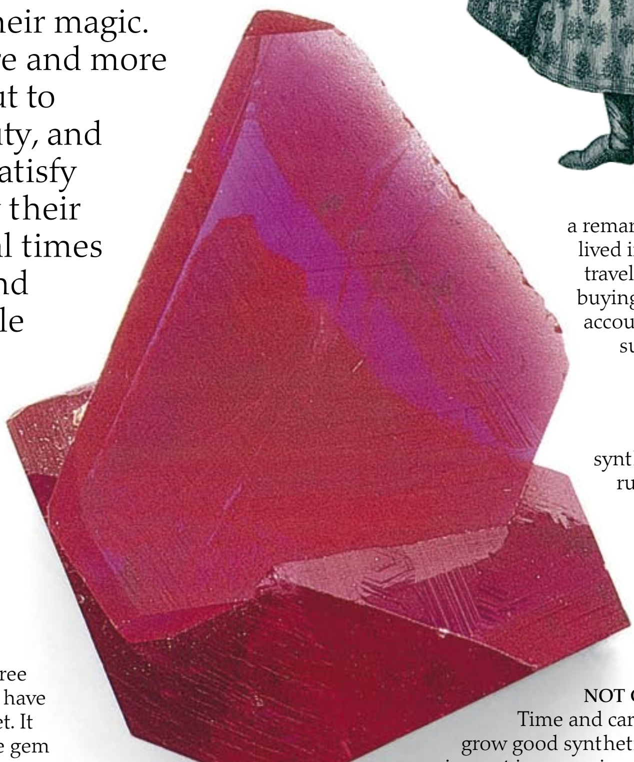

Synthetic ruby crystals

Cut synthetic ruby

NOT CHEAP
Time and care are needed to grow good synthetic crystals, and the equipment is expensive. Stones cut from such crystals are therefore not cheap, but natural stones of similar color still cost 10 to 100 times more.

Chipped glass of a GTD

DOUBLED UP
A stone made of two materials glued or fused together is known as a doublet. Triplets have three layers – the central one being colored glass or cement. The most popular doublet is the garnet-topped doublet (GTD). The top is a thin piece of colorless garnet which is more durable than the glass base which provides the color. These cheaper stones are sometimes sold by disreputable traders as rubies, sapphires, or emeralds.

THEFT PREVENTION
This 1910 photograph shows Chinese sorters in the ruby mines of Mogok, Myanmar (pp. 36–37). The sorters were made to wear wire helmets to stop them from stealing rubies by hiding them in their mouths while they were sorting.

SHORT-LASTING
Red glass has long been used to imitate ruby, but its luster and shape soon deteriorate with age. In contrast, ruby, with its greater hardness and toughness, keeps these qualities far longer.

STONE SUBSTITUTES
Diamond (pp. 34–35) is the most frequently imitated stone in jewelry, as there are many people who cannot afford the real thing. The oldest imitations are glass and rock crystal, but in the 20th century synthetic garnets and then cubic zirconia became available. This selection of diamond simulants is arranged clockwise in order of increasing fire. The five stones GGG to synthetic rutile have more fire than diamond but cannot compare in their luster, hardness, or durability.

Fluorite

Quartz

Topaz

Synthetic sapphire

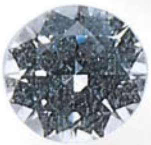
Synthetic spinel

YAG (yttrium aluminum garnet)

Glass

Zircon

GGG (gadolinium gallium garnet)

Cubic zirconia

Lithium niobate

Strontium titanate

Synthetic rutile

Diamond

JUST LIKE THE REAL THING
This advertisement was for Faulkner's Celebrated Diamonds. They were advertised as "The only perfect substitute for diamonds of first water."

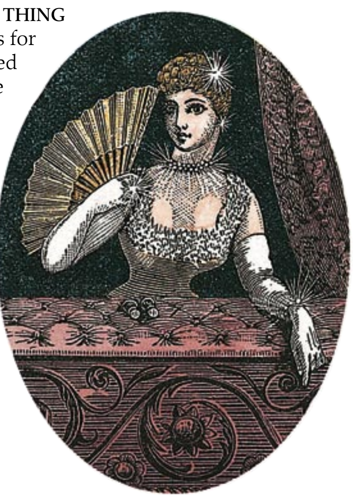

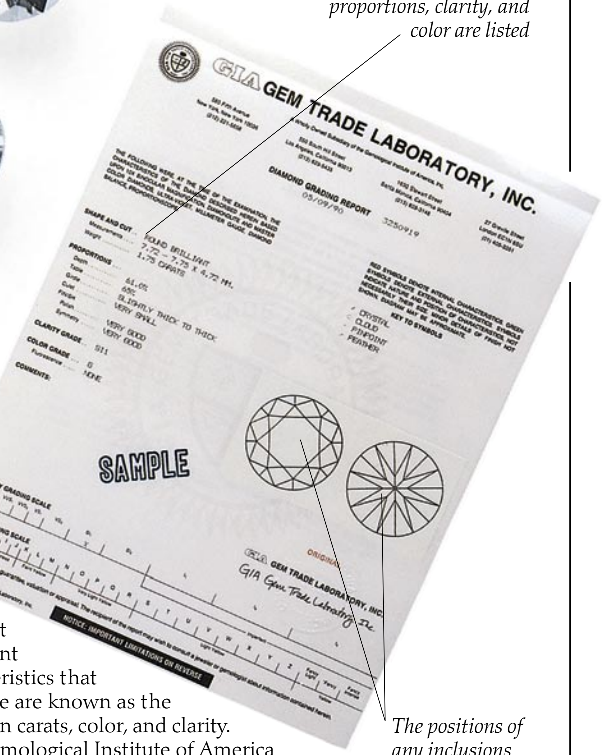
GIA GEM TRADE LABORATORY, INC.

DIAMOND GRADING REPORT
05/09/90
3250919

THE FOLLOWING WERE, AT THE TIME OF THE EXAMINATION, THE CHARACTERISTICS OF THE DIAMOND DESCRIBED HEREIN BASED UPON 10X BINOCULAR MAGNIFICATION, DIAMONDLITE AND MASTER COLOR DIAMONDS, ULTRAVIOLET, MILLIMETER GAUGE, DIAMOND BALANCE, PROPORTIONSCOPE.

SHAPE AND CUT ... ROUND BRILLIANT
Measurements ... 7.72 – 7.75 X 4.72 MM.
Weight ... 1.75 CARATS
PROPORTIONS ...
Depth ... 61.0%
Table ... 65%
Girdle ... SLIGHTLY THICK TO THICK
Culet ... VERY SMALL
Finish ...
Polish ... VERY GOOD
Symmetry ... VERY GOOD
CLARITY GRADE ... SI1
COLOR GRADE ... G
Fluorescence ... NONE
COMMENTS:

RED SYMBOLS DENOTE INTERNAL CHARACTERISTICS. GREEN SYMBOLS DENOTE EXTERNAL CHARACTERISTICS. SYMBOLS INDICATE NATURE AND POSITION OF CHARACTERISTICS, NOT NECESSARILY THEIR SIZE. MINOR DETAILS OF FINISH NOT SHOWN. DIAGRAM MAY BE APPROXIMATE.

KEY TO SYMBOLS
CRYSTAL
CLOUD
PINPOINT
FEATHER

SAMPLE

GIA CLARITY GRADING SCALE
GIA COLOR GRADING SCALE
D E F G H I J K L M N O P Q R S T U V W X Y Z

ORIGINAL
GIA GEM TRADE LABORATORY, INC.
GIA Gem Trade Laboratory Inc

NOTICE: IMPORTANT LIMITATIONS ON REVERSE

Measurements are given in millimeters and the details of proportions, clarity, and color are listed

Clarity and color are graded on a scale. This diamond is graded SI_1 (small inclusion) for clarity and G (near colorless) for color

The positions of any inclusions are indicated on the outlines of the crown and pavilion

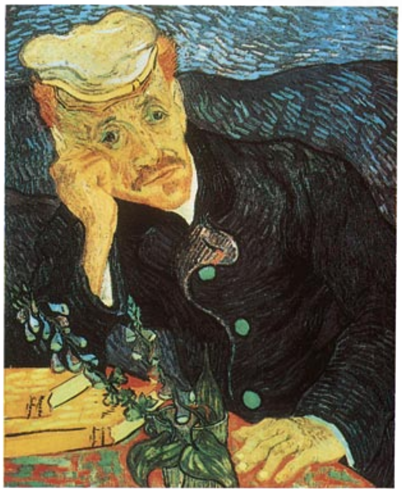

MILLION-DOLLAR MAN
The most valuable stones on the market in the past ten years have been diamonds, rubies, and emeralds, but even the most expensive diamond was only a fraction of the record price paid for Van Gogh's *Portrait of Dr. Gachet*. This was sold at Christie's, New York City, in May 1990 for $82.5 million.

FOUR Cs
Two diamonds that look alike at first can have very different values. Four characteristics that determine their value are known as the four Cs: cut, weight in carats, color, and clarity. In the 1930s, the Gemological Institute of America established its own system of grading and issued certificates for individual diamonds containing information about the four Cs. Different systems were introduced in Europe, but in 1975 the grading procedure started to be standardized.

MADE TO MATCH
Opal (pp. 40–41) is imitated in several ways. The flamboyant Slocum stone is made of glass and is much cheaper than true opal. Gilson opal is grown in the laboratory over many months and is intermediate in value. True opal is graded according to body color and the play of colors, with black opal being the rarest and most expensive.

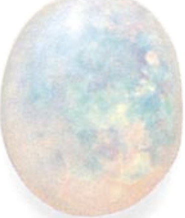
Polystyrene latex

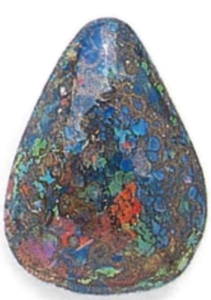
Real black opal

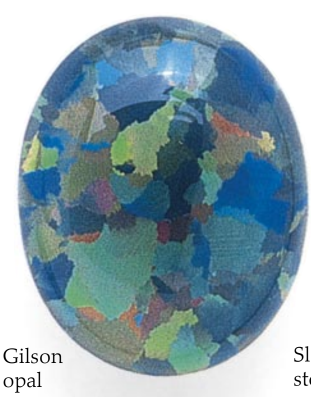
Gilson opal

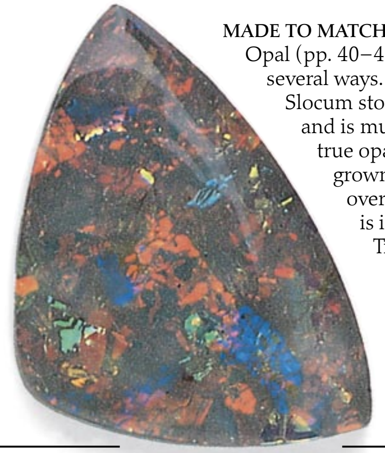
Slocum stone

Making them sparkle

SOME ROUGH CRYSTALS are beautifully shaped and have breathtaking luster and color, but most are worn down or have other surface imperfections. A skilled stonecutter and polisher, called a lapidary, can turn these stones into objects of beauty and value by using their individual qualities in the right way. Beads and cabochons are the oldest cuts for such materials as lapis lazuli and turquoise (p. 50), and carnelian and agate (p. 33). Brilliant-cut diamonds are known from the 17th century, and brilliant is the most popular cut for diamonds today.

BRILLIANT-CUT RUTILE
In 1919 Marcel Tolkowsky set down the precise angles and proportions of a brilliant cut to give the best combination of sparkle, brilliance, and fire.

ROSE-CUT SMOKY QUARTZ
The rose cut dates from the 14th century. Most stones have a flat base and a shallow or deep dome-shaped top covered in triangular facets.

TABLE-CUT AMETHYST
The table cut is derived from the diamond octahedron by sawing off the top above what becomes a square or rectangular girdle.

STEP-CUT QUARTZ
The step cut has many rectangular facets. It is particularly suitable for strongly colored gems, such as emerald and tourmaline.

Cutting a brilliant

A lapidary first studies a rough stone with a powerful lens called a loupe. This shows the direction of grain and the flaws. The stone is marked to show where it should be sawed and the facets are then ground.

1 **ROUGH CHOICE**
A rough crystal is chosen for cutting.

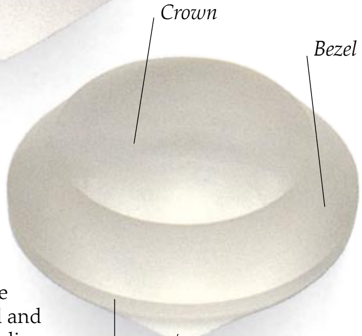

2 **SAWED IN TWO**
The crystal is sawed to remove the top pyramid and rounded by grinding against another diamond, called bruting. (The cutting sequence is shown here with models.)

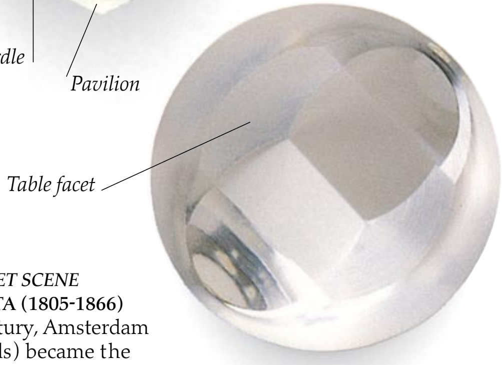

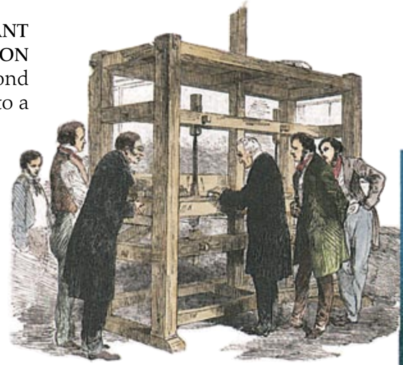

BRILLIANT TRANSFORMATION
The Koh-i-noor diamond (p. 35) was recut into a brilliant in 1852. Here the Duke of Wellington makes the first facet, watched by Mr. Voorsanger, the cutter from Amsterdam.

The Koh-i-noor cut as a brilliant

3 **FACETING BEGINS** *left*
The stone is mounted on a stick called a dop, and the first facet – the flat table – is ground on a cast-iron wheel, called a scaife, followed by more facets.

4 **FACETED** *left*
Four "bezel" facets are ground between the table and the girdle. The stone is turned over and four facets ground on the pavilion. Four more facets are ground on the crown and on the pavilion, and the "culet" facet on the base.

5 **FINISHED OFF** *right*
The stone is finished by the brillianteerer, who adds 24 facets above and 16 below the girdle. The standard brilliant has 57 facets or 58 with a culet.

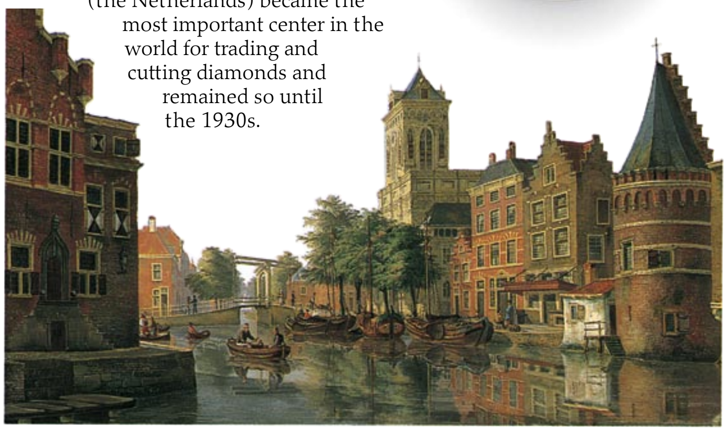

AMSTERDAM STREET SCENE
BY PETRUS BERETTA (1805-1866)
In the 17th century, Amsterdam (the Netherlands) became the most important center in the world for trading and cutting diamonds and remained so until the 1930s.

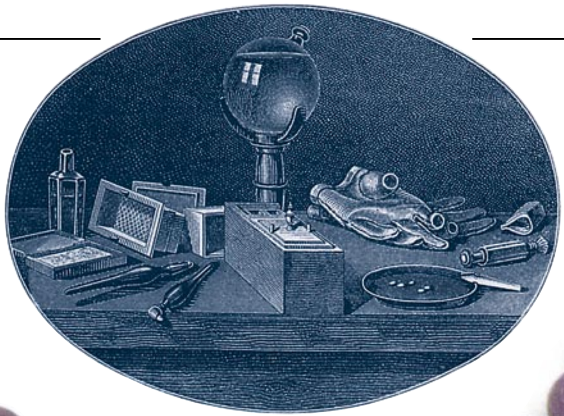

Diamond cutter's table of the 1870s

Agate

Amethyst

OLD FASHION
Beads can be made out of relatively soft materials and are among the oldest forms of fashioned gems. They are made today using automatic machines.

CUTTING IT FINE
These craftsmen from Perth, Australia, are cutting diamonds from the Argyle mine (p. 25). Diamond-cutting tools remained virtually unchanged from the 15th century until the development of automatic machines in the 1970s and, later, lasers.

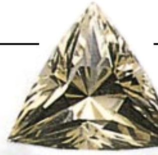

Triangular-cut citrine

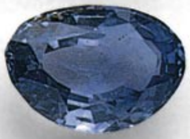

Irregular-cut sapphire

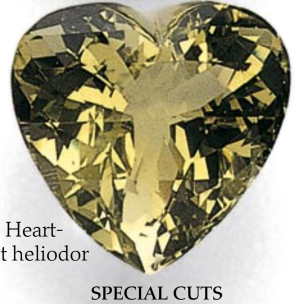

Heart-cut heliodor

SPECIAL CUTS
If a stone is very rare, a special cut may be developed to keep as much weight as possible. Or stones may be given unusual cuts for special occasions.

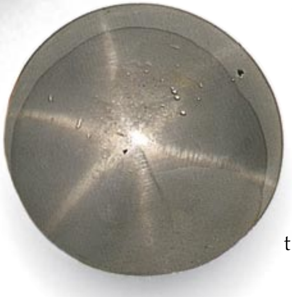

LIGHT DISPLAY
Star sapphires and rubies (p. 37) must be cut as cabochons if the tiny needles of rutile inside are to reflect the light and display the star.

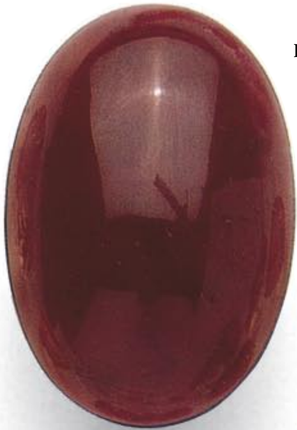

HARDHEADED
Hard-wearing opaque materials such as this garnet are often cut into rounds or ovals with plain, curved surfaces. These are called cabochons – a French word derived from the Latin *cabo*, meaning a head.

Cabochons of quartz, cut to display the cat's-eye effect and set in a brooch

OVER AND OVER
Rocks and minerals can be polished by tumbling them in a drum with water and grit. The kind of grit depends on the material to be polished; it ranges from coarse grits to start with, to fine grits for finishing.

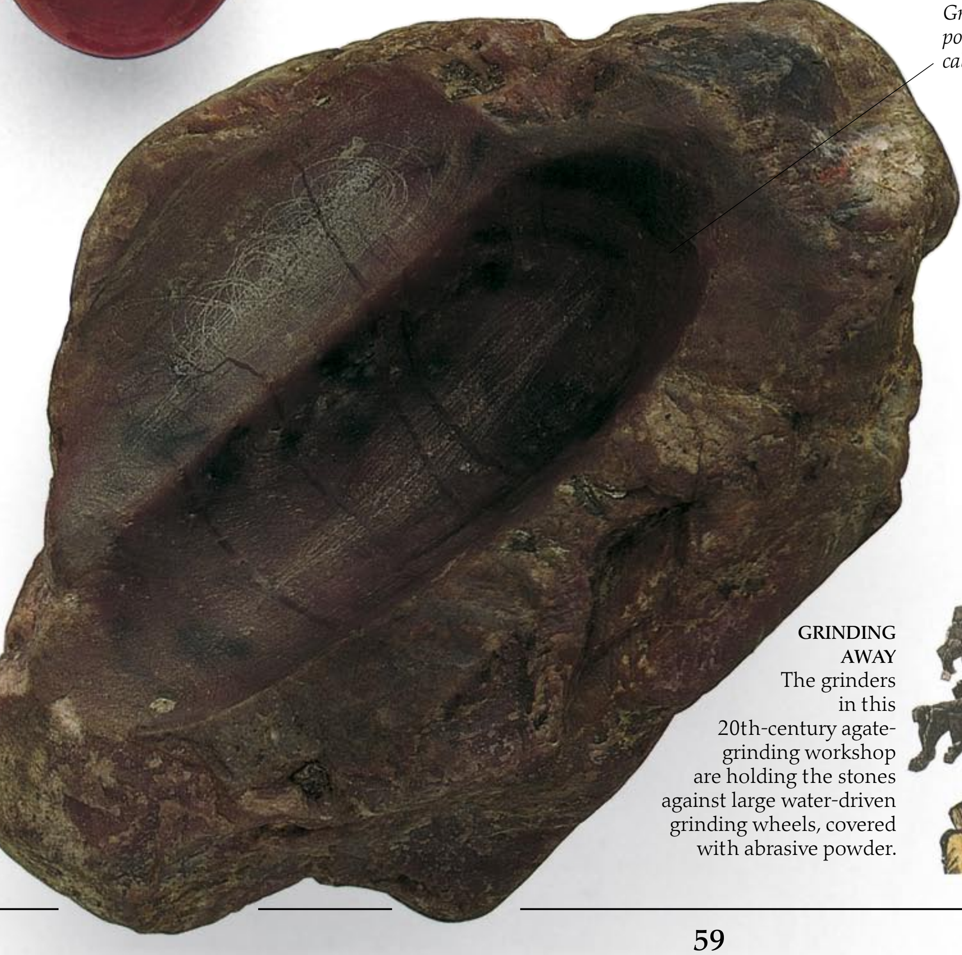

Groove caused by polishing garnet cabochons

INDIAN POLISH
This massive corundum was used in India in the 19th century for fashioning and polishing garnets. Continual use has caused tiny chips to split off, eventually leaving grooves.

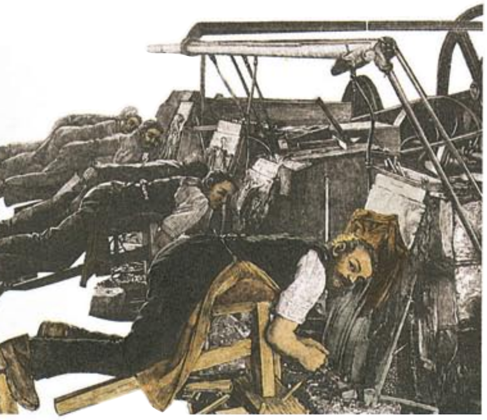

GRINDING AWAY
The grinders in this 20th-century agate-grinding workshop are holding the stones against large water-driven grinding wheels, covered with abrasive powder.

Lore and legends

Superstition, myth, and romance have been attached to crystals since ancient times. In Persian mythology the world is said to stand on a giant sapphire, the reflection of which colors the skies blue. Emeralds were once thought to blind snakes, and wondrous medicinal properties have been given to diamonds. Rubies were symbols of power and romance in the Middle Ages, and ladies often presented jewels to their knights as love tokens. Many specific stones are said to have been put under a terrible curse and stories are told of how some, such as the Hope diamond, bring disaster to their owners.

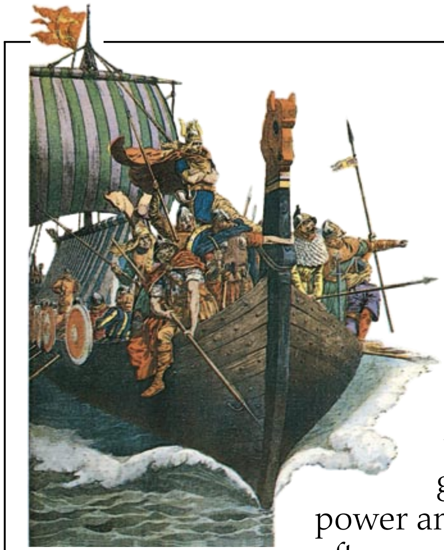

CRYSTAL COMPASS
Some sources say that the Vikings navigated with transparent cordierite crystals. When held up to sunlight and rotated, cordierite darkens and changes color. Using this property, the Vikings would have been able to work out compass directions.

BRAINY BIRDS
The mystery of how a bird migrates may have been solved. It is thought that a tiny magnetic crystal in its brain detects the earth's magnetic field.

Iron filings attracted to magnetite fan out to follow magnetic lines of force

POWER TO REPEL
Some pieces of magnetite, a naturally occurring iron oxide, are magnetic. They are known as lodestones, and people in ancient times believed they had special powers. For example, it is said that Alexander the Great, a famous general of the third century B.C., gave lodestones to his soldiers to repel evil spirits.

Magnetite crystals

SOBERING INFLUENCE
Amethyst was supposed to have many useful powers in the 15th century, but its traditional virtue was its ability to cure drunkenness. One way in which it may have gained this reputation is that drinking vessels were sometimes made of amethyst. In these, water would look like red wine, but could be drunk with no intoxicating effect!

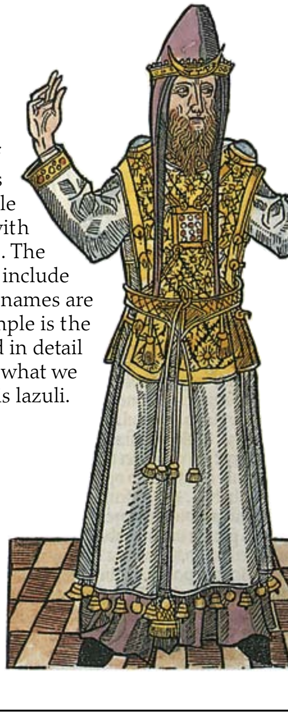

HOLY ROWS
The original breastplate of the High Priest of Israel is described in the Bible (Exodus 28: 15–30) as set with four rows of three stones. The stones are all named and include familiar species, but some names are misleading. An example is the sapphire. This is described in detail elsewhere and is evidently what we now know as lapis lazuli.

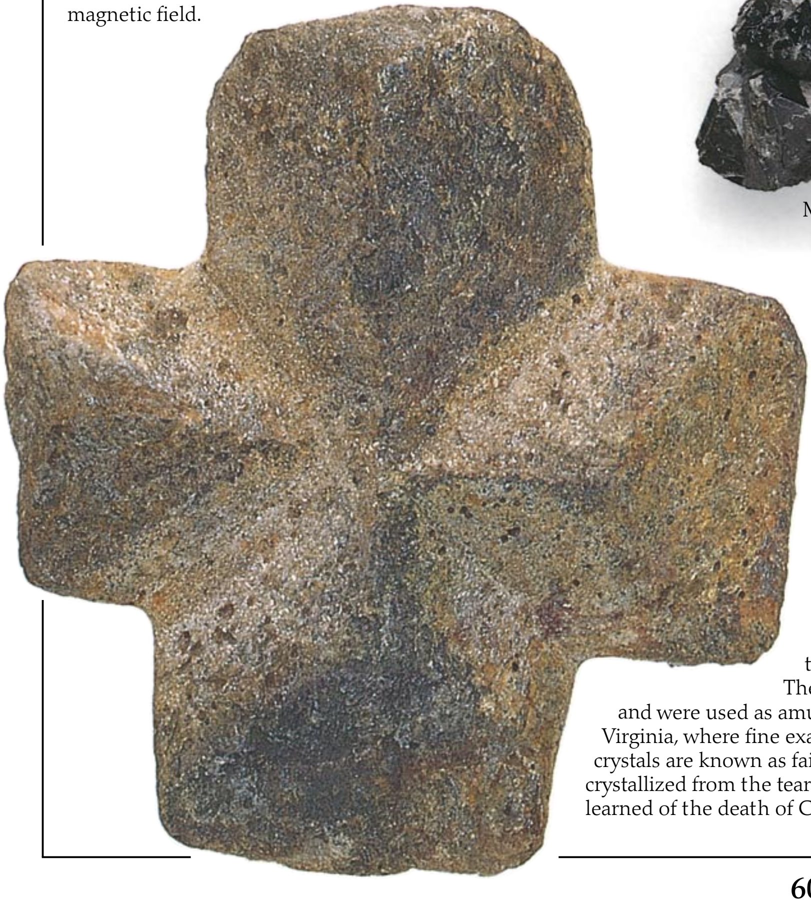

STONE TEARS
Crystals of staurolite sometimes twin at right angles to form crosses. These used to be called *lapis crucifer* and were used as amulets at baptisms. In Patrick County, Virginia, where fine examples have been found, these twin crystals are known as fairy stones. The legend is that they crystallized from the tears of fairies, who cried when they learned of the death of Christ.

Crystal gazing

Crystal balls have been used for telling the future since Greek and Roman times. The fortuneteller looks at the polished surface of the ball intently until he or she can no longer focus on the ball but sees instead a curtain of mist. This enables the gazer to see "visions" in the ball, imagined in response to questions asked by the person whose fortune is being told.

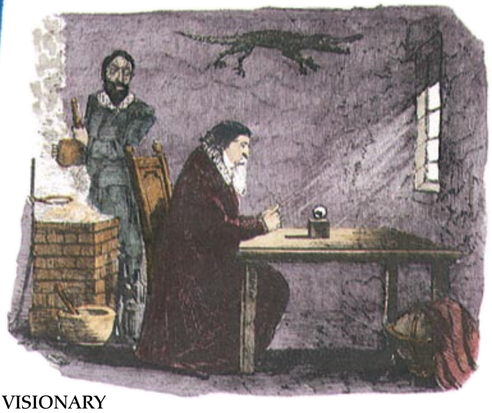

VISIONARY
John Dee was a charlatan of the 16th century but a favorite of Queen Elizabeth I of England. He conducted many crystal gazings in which a colleague, Kelley, interpreted "visions" in the crystal.

FOR THE FUTURE
Quartz (pp. 30–33) is the most popular material for crystal balls, but other materials which have a shiny surface, such as mirrors, polished steel, and even the surface of water, have been known to produce a satisfactory effect.

EASTERN SEER
Crystal balls have been found from cultures in the Far East, the Americas, and Europe. This one comes from China.

Birthstones

The dedicating of a special stone to each month of the year was first suggested in the 1st century A.D., and was linked to the twelve stones in the breastplate of the High Priest of Israel. The custom of wearing such stones became popular in the 18th century. It started in Poland and then spread throughout Europe and the rest of the world.

January – Garnet

February – Amethyst

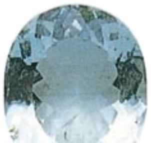

March – Aquamarine

April – Diamond

May – Emerald

June – Pearl

July – Ruby

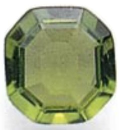

August – Peridot

September – Sapphire

October – Opal

November – Topaz

December – Turquoise

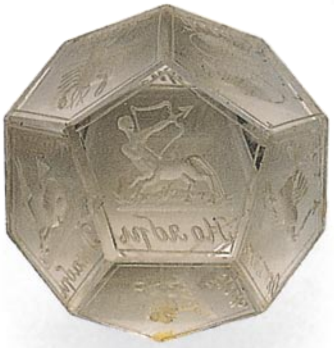

Rock crystal engraved with the twelve signs of the zodiac

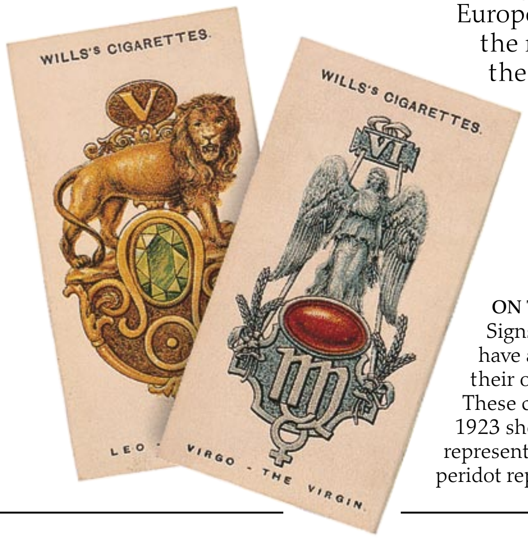
WILLS'S CIGARETTES.
V
LEO
WILLS'S CIGARETTES.
VI
VIRGO - THE VIRGIN.

ON THE CARDS
Signs of the zodiac have also been given their own gemstones. These cigarette cards of 1923 show carnelian representing Virgo and peridot representing Leo.

CHANGING MONTHS
The gem species which represent each month have varied considerably throughout history. The Roman, Arabian, Jewish, and Russian cultures all favored different combinations. The group shown here is the most popular today.

Crystals at home

MANY EVERYDAY OBJECTS in the home are crystalline. There are ice crystals in the freezer, salt and sugar crystals in the food cupboard and in food itself, crystals of vitamin C and aspirin in the medicine cabinet, tartrate crystals in the wine bottle, and silicon crystal chips in the refrigerator and washing machine. The TV, telephone, radio, and camera work because of crystals, the house is built of materials which are mostly crystalline, and outside, bikes and cars stand slowly rusting – crystallizing!

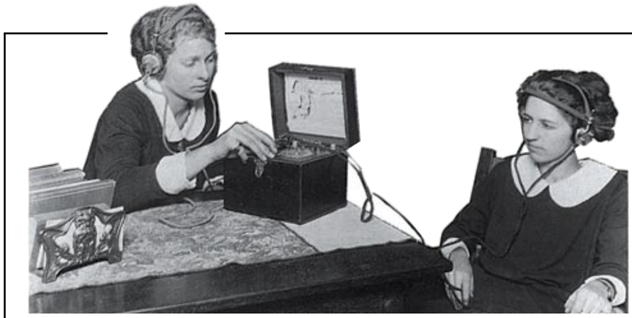

BY A WHISKER
"Hello. This is New York. Here is the news." This might be what these women are listening to using a crystal set. In this early form of radio, operators moved a thin copper wire, sometimes known as a cat's whisker, against a galena crystal to pick up radio waves. Crystal sets became popular in the 1920s when public broadcasting began.

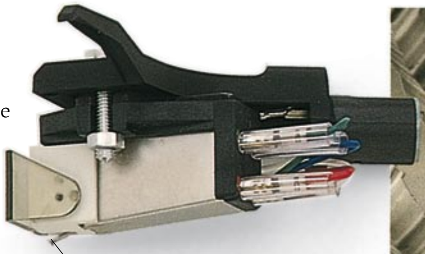

FOR THE RECORD
In some record players, there are two crystals. The stylus is made of hard-wearing diamond or corundum, and a piezoelectric crystal (p. 31) in the cartridge converts vibrations received from the record into an electrical charge.

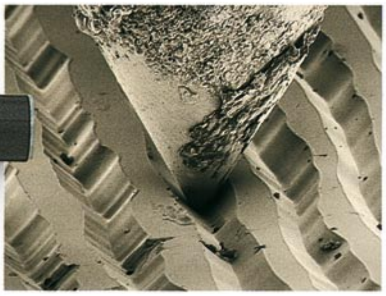

Enlarged photograph of a diamond stylus traveling through the groove on a stereo record

SPOONFUL OF SUGAR
Over 100 million tons of sugar are crystallized every year in refineries. Sugar is extracted as a liquid solution from raw sugar cane or beet, then converted into sugar crystals. The silver spoon holding these sugar crystals is itself a mass of silver crystals.

Liquid crystal display

CRYSTAL DISPLAY
The displays in many calculators are liquid crystal displays (LCDs). Liquid crystals are not truly crystalline. They flow like a liquid but have molecules arranged like those of crystals and some properties of crystals. Power rearranges the molecules so that they reflect or absorb light and show dark or light.

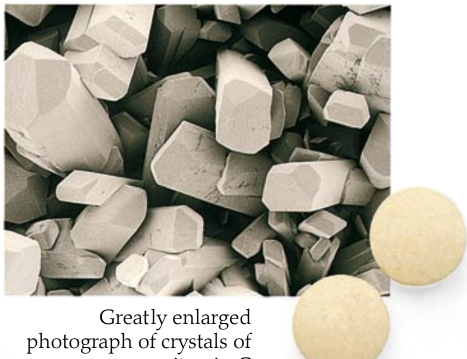

Greatly enlarged photograph of crystals of vitamin C

VITAL INTAKE
These tablets contain crystals of ascorbic acid, or vitamin C. Ascorbic acid is a white crystalline substance present in plants, especially citrus fruits, tomatoes, and green vegetables. Vitamins are essential to us in small quantities. Most cannot be produced by the body, and therefore have to be taken in through food or tablets.

Collecting

Crystal collecting can be enjoyed without spending too much money. Crystals can be collected in the field, bought, or exchanged with friends and dealers. They are usually fragile and should be carefully stored. Those found in the field should be kept with details of where they were found and, if possible, their host rock.

PRECIOUS STONES
Many people own jewelry made with precious or semiprecious stones. This silver brooch contains diamonds, a blue sapphire, and a pearl.

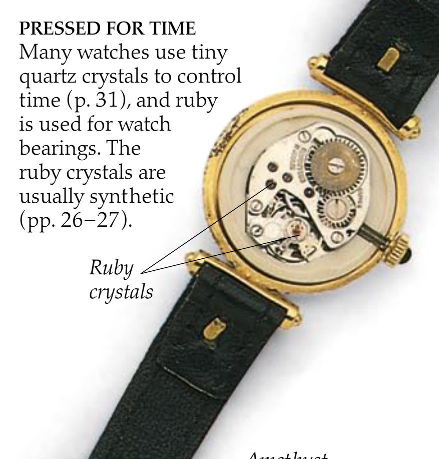

PRESSED FOR TIME
Many watches use tiny quartz crystals to control time (p. 31), and ruby is used for watch bearings. The ruby crystals are usually synthetic (pp. 26–27).

Ruby crystals

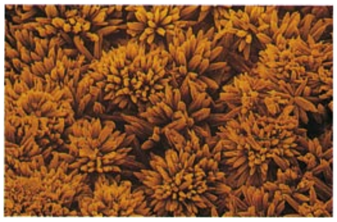

Photo enlargement of needle-like crystals in kettle "fur"

KETTLE FUR
Crystals can be found in your kettle. Even after purification processes there are still some harmless minerals dissolved in tap water. These crystallize and coat the inside of your kettle when the water is boiled.

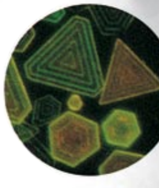

Photo enlargement of silver nitrate crystals on a photographic film

ON FILM
When a photograph is taken, an image is recorded on light-sensitive material by the action of light. Most film uses light-sensitive crystals of silver salts. The photographic industry is one of the largest users of silver.

Hand lens for studying the major features of crystals

Wulfenite crystal

IN MINIATURE
A popular method of collecting and storing crystals is as "micromount" specimens a few millimeters in size. They take up little space, and fine crystal groups of rare and unusual minerals can be collected.

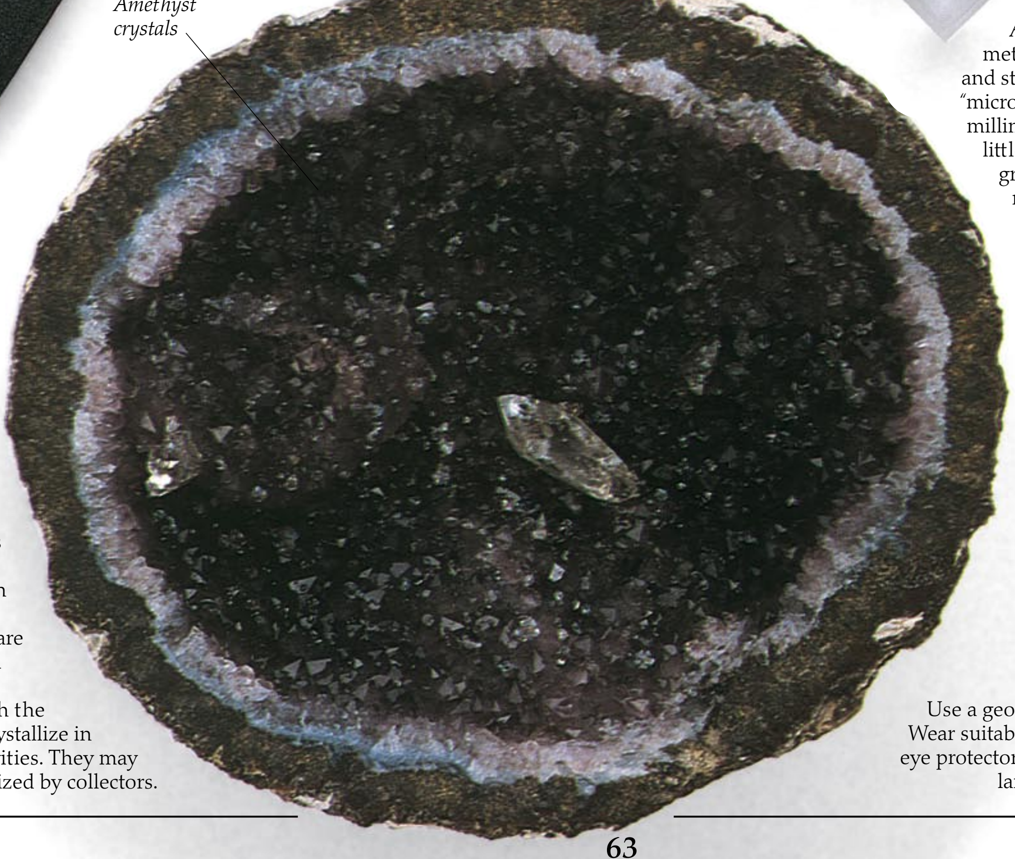

Amethyst crystals

CAVITY FILLERS
Crystal-lined geodes are often found within basaltic lava flows. They are formed from fluids which filter through the rocks and crystallize in available cavities. They may be highly prized by collectors.

FIELDWORK
Use a geological hammer to collect in the field. Wear suitable clothing and boots, a hard hat, and eye protector. Ask permission to collect on private land and always observe the local laws.

Did you know?

AMAZING FACTS

In ancient times, the Persians thought the Earth rested on a giant sapphire and that the blue of the heavens was its reflection. Others thought the sky was a sapphire in which the Earth was set.

Sapphire and diamond pendant

Sapphire-blue is a color associated with the qualities of harmony, trust, and loyalty. This is the reason why women in many countries choose sapphires as the stones in their engagement rings.

The name "garnet" comes from the Latin for pomegranate, a fruit which has bright red, garnetlike seeds. In fact, the color of garnets varies from violet-red to the deepest burgundy.

Pomegranate

Diamonds are sometimes found under the sea. People trawl for diamonds off the coast of Namibia in southern Africa. The latest techniques involve large offshore ships pumping gravel containing diamonds up to the surface.

Since diamond was discovered in kimberlite rock, it has been extracted on a massive scale. Over 25 tons of rock have to be blasted for every finished carat—0.007 oz (0.2 g)—of diamond mined.

In medieval times, people who could afford it wore a diamond jewel to protect them from the plague. The Ancient Greeks thought that diamonds could protect people from poisons.

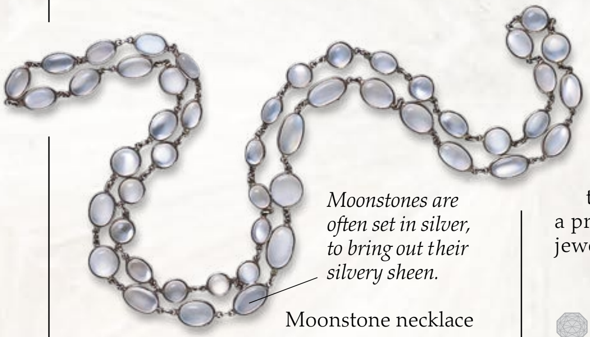

Moonstones are often set in silver, to bring out their silvery sheen.

Moonstone necklace

People once thought that a moonstone's opalescent luster waxed and waned just like the Moon, so moonstones have always been worn by worshipers of the Moon.

Topaz crystals can be over 3 ft (1 m) long and weigh hundreds of pounds. The name "topaz" is thought to come from the Sanskrit word *tapas*, which means "fire."

The Ancient Greeks thought that amber was the hardened rays of a sunset and it was considered sacred to the Sun god Apollo. Amber can produce an electric charge when rubbed. In fact, the word "electricity" comes from the Greek word for amber, *elektron*.

Hawksbill turtle

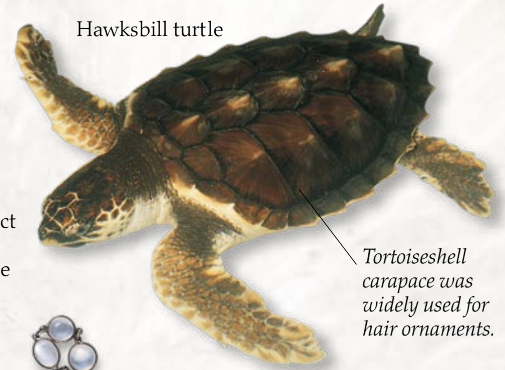

Tortoiseshell carapace was widely used for hair ornaments.

Tortoiseshell doesn't come from a tortoise at all, but from a turtle. It is made from the shell of the rare Hawksbill turtle, which is now a protected species. Most "tortoiseshell" in jewelry today is made of plastic.

Six-rayed star sapphires were once thought to provide the best protection from evil. The three crossing arms of the star were meant to represent faith, hope, and destiny.

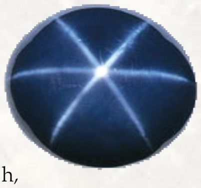

Star sapphire

To the Egyptians, the intense blue of lapiz lazuli meant that it was a heavenly stone. They often used it on statues of their gods and in burial masks, to protect them in the next life.

The Arabians thought that pearls were the tears of gods. Although cultivating pearls (making them grow by putting irritants into oysters) is much faster than waiting for natural pearls to form, it can still take a very long time—up to 4 years.

Diamond dredging boats

QUESTIONS AND ANSWERS

Q **How long have people been mining for gemstones?**

A Jewelry has been found in burial sites dating back thousands of years. Ancient jewelry is rare, but some Ancient Egyptian pieces have survived. These are often made of gold set with gems such as turquoise, lapis lazuli, and carnelian.

Q **What are potato stones?**

A Potato stone is a name given to a geode, a hollow ball of rock encrusted inside with amethyst or other crystals. The crystals form when silica-rich liquids seep into gas bubbles in lava as it cools to form volcanic rock. Also known as thunder eggs, geodes are highly prized.

Potato stones

Q **Why are gemstones so precious?**

A Gemstones are valuable because of their natural beauty, their durability, their rarity, and the way in which they are cut and polished. There are 3,000 different kinds of mineral, but only about 100 could be classed as gemstones, making them very rare.

Q **Where do diamonds come from?**

A Diamond forms at extremely high temperatures and pressure 50 miles (80 km) or more below the ground in the Earth's mantle. When diamonds were first discovered, 2,000 years ago, they were mainly found in river gravel. These days, most diamond is mined from kimberlite rock. Australia is the main producer today, but diamonds are also mined in Africa, Russia, Brazil, and the US.

Q **What are seed pearls ?**

A Pearls vary in size, the smallest of all being seed pearls. Pearls are not weighed in carats like other gemstones, but in grains. One grain = 0.002 oz (0.05 g). Seed pearls weigh less than 0.25 of a grain.

Scarab beetle good-luck charm, found in Tutankhamun's tomb

Q **Why were children often given coral jewelry in the past?**

A Coral was thought to protect the person wearing it from evil. For this reason, children were often given coral necklaces and bracelets to keep them healthy and safe from harm.

Q **Why are emeralds green?**

A The characteristic green color of emeralds come from tiny amounts of chromium and vanadium.

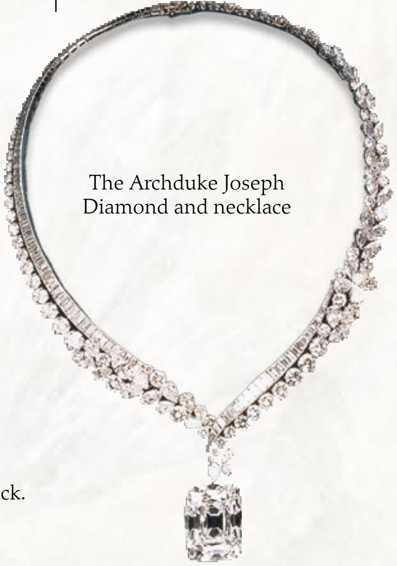

The Archduke Joseph Diamond and necklace

Q **Why are gemstones cut and polished?**

A Cutting and polishing gemstones transforms them into jewels. It maximizes the amount of light they reflect so that they sparkle and shine.

Q **Why are organic gems often carved rather than cut into facets?**

A Organic gems, such as coral, ivory, amber, jet, and pearl are softer than mineral gems and are often opaque. For this reason, they are often carved and polished. Light cannot shine through them, so it is pointless to cut them into facets to increase their brilliance.

Q **What is the connection between rubies and emery boards?**

A Rubies are one of the most expensive gems. Ruby is a variety of the mineral corundum, which is second in hardness to diamond. Emery is an impure form of corundum and has been used as an abrasive for thousands of years.

Cut ruby

Q **Which famous ruby isn't really a ruby at all?**

A Many crown jewels around the world contain massive red gemstones called spinels, which people mistook for rubies. The huge Black Prince's Ruby in the British Imperial State Crown is in fact a spinel, given to the Black Prince in 1367 by the king of Spain as a token of his appreciation.

Record Breakers

MOST VALUABLE GEMSTONE:
Diamonds are the most precious gems, famed for their fiery beauty, as well as for being the hardest mineral on Earth.

BIGGEST DIAMOND:
The largest single rough diamond ever mined was the Cullinan diamond, found in 1905 in South Africa. It weighed 3,106 carats and was cut into 9 large and 96 smaller stones.

CUTTING MARATHON:
It took three polishers, working 14 hours a day, eight months to cut and polish the Cullinan 1 diamond. The stone is now set in the British Imperial Scepter.

PRICELESS GEM:
Only three crystals of painite are known, and none of them has ever been sold, making them priceless.

BIGGEST BERYL CRYSTAL:
A beryl crystal found in Madagascar weighed 39 tons (36 metric tons) and was 60 ft (18 m) long.

Identifying gemstones

Emerald lizard

To THE UNTUTORED EYE, many gemstones look alike. They are often similar in color and are cut in much the same way. Here is a guide to the color and characteristics of some of the most popular gemstones, both mineral and organic.

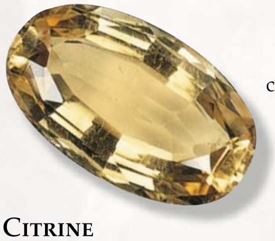

Mixed-cut citrine with the orange tinge often seen in this gem

CITRINE
Citrine is a yellow or golden form of quartz. Natural citrine (its name comes from the word "citrus") is pale yellow, but it is extremely rare.

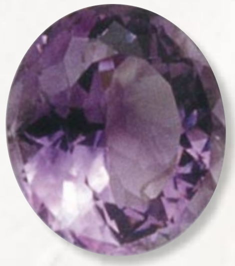

Oval mixed-cut amethyst with a typical purplish violet color

AMETHYST
Amethysts are purple, lilac, or mauve quartz crystals. They often have distinctive internal markings and a blue or reddish tinge when seen from different angles.

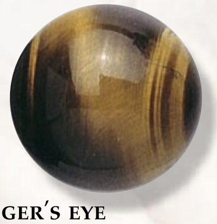

Tiger's eye, cut and polished to show up its stripes

TIGER'S EYE
Tiger's eye is a variety of chalcedony, a type of quartz composed of tiny fibers. It has a waxy appearance and is black with yellow and golden-brown stripes.

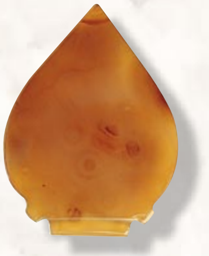

Typical reddish orange polished stone from India

CARNELIAN
Also called cornelian, this is a translucent, reddish-orange or brown form of chalcedony.

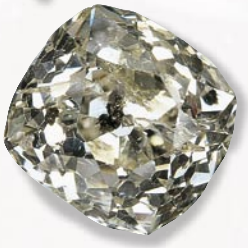

Colorless, brilliant-cut diamond with black inclusions

DIAMOND
Made of pure carbon, diamond is exceptionally hard and shines brightly. It occurs in many colors, but the most popular variety is pure and colorless.

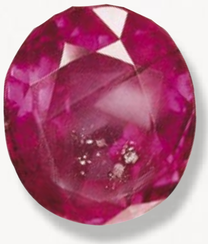

Cushion mixed-cut ruby in bright red

RUBY
One of the most expensive gems, the classic ruby is a rich red, but it can vary in color from pink to brown. Rubies are second in hardness only to diamonds.

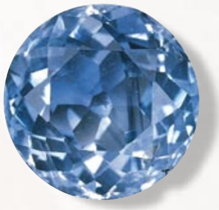

Pale blue Sri Lankan sapphire

SAPPHIRE
The most valuable sapphires are a clear shade of deep blue, but they can also be yellow, green, pink, or colorless. Like ruby, sapphire is a type of corundum.

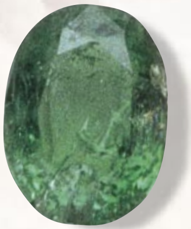

Bluish-green emerald with many tiny fissures and internal markings

EMERALD
A variety of beryl, emeralds are a rich green. Only the finest gems are transparent and flawless. Most emeralds have flaws known as a *jardin*, the French for "garden."

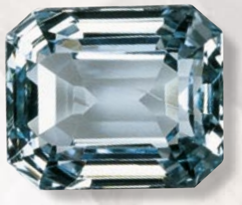

Octagonal step-cut aquamarine with a slight greenish tinge

AQUAMARINE
Another variety of beryl, aquamarine ranges in color from pale sea-green to dark blue. It can appear to change color when viewed from different angles.

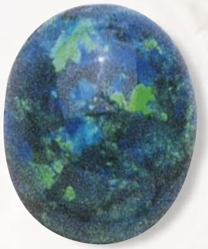

An opal displaying flashes of green and blue

OPAL
Opal is known for its iridescence and flashes of color. Iridescent opal with a dark background is called black opal. "Potch opal" is opaque, without any iridescence.

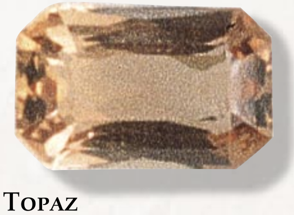

Salmon-pink colored topaz

TOPAZ
Topaz occurs in several different colors, ranging from deep golden yellow (known as sherry topaz) and pink to blue and green. Natural pink stones are extremely rare.

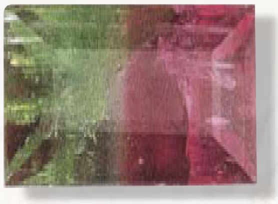

Watermelon tourmaline

TOURMALINE
Tourmalines come in a range of colors, but they all have the same crystal structure. Watermelon tourmaline is so called because of its pink and green coloring.

Pyrope (garnet) cut as an oval

Garnet

There are several different gemstones in the garnet group. The most popular for jewelry are the pyrope, which is blood-red, and the deep red almandine.

Octagonal mixed-cut peridot

Peridot

Peridot is an olive or bottle-green color with a distinctive waxy luster. It has strong double refraction, which means you can often see a doubling of the back facets.

A gray moonstone

Moonstone

The moonstone derives its name from its blue-white sheen, like moonshine. Some varieties are gray, yellow, pink, or green.

Octagonal mixed-cut with a vitreous luster

Spinel

The most popular spinel is ruby-red, but it also occurs in blue and yellow. Until the 19th century, red spinels were known as Balas rubies, perhaps after their source.

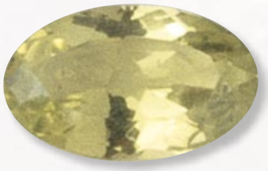

Cut yellow chrysoberyl

Chrysoberyl

Chrysoberyl is renowned for its golden color. One variety, called alexandrite, appears to change color from green to light red when seen in artificial light.

Colorless zircon produced by heating a reddish-brown stone

Zircon

Pure zircon is colorless and resembles diamond, but it is more likely to be golden brown. Many zircons are heat-treated to produce blue or colorless gemstones.

Translucent jadeite with black inclusions

Jade

Two different minerals, jadeite and nephrite, are recognized as jade. The finest quality jadeite is emerald green. Nephrite varies in color from cream to olive green.

Polished rock speckled with pyrite

Lapis lazuli

Prized for its intense dark blue, lapis lazuli is a rock made up of several minerals. Specks or streaks of pale pyrite and calcite are often visible against the blue.

Stone cut and polished as a cabochon

Turquoise

Turquoise is valued for its intense color, which varies from blue–green to bright blue. Opaque, it is usually cut and polished into rounded beads or cabochons.

ORGANIC GEMSTONES

Unlike most other gemstones, which are mineral in origin, organic gems are derived from plants and animals. Amber, jet, coral, pearl, ivory, and shell are all organics. These materials are not stones and they are not as hard and durable as mineral gems. Instead of being cut into facets like most other gemstones, they are usually polished or carved.

Carved jet with a finely wrought rose at the center

Jet

Jet is a fine-grained rock formed from fossilized wood. Black or very dark brown, it is opaque with a velvety luster. Once popular for mourning jewelry, it is often faceted and polished to a high shine.

Transparent golden brown beads that have been faceted

Amber necklace

Amber

Amber is formed from the hardened resin of trees. Transparent or translucent, it is usually a golden orange color, but it can also be a rich dark red, and occasionally contains insects and plants.

Intricate red coral carving showing a monkey climbing a tree

Coral

Coral is made of the remains of coral polyps. It can be pink, red, white, or blue. Naturally dull, it shines when polished.

A roughly spherical pearl suitable to be used as a bead

Pearl

Pearls are formed in shellfish and have a characteristic iridescent sheen. They vary in color from white and cream with a hint of pink to brown, or even black.

Find out more

THE BEST WAY TO find out more about crystals and gemstones is to visit museums. Natural history and geological museums usually have extensive rock and mineral displays and are an invaluable source of information on how crystals are formed and what they look like in their natural state. Many of them also have good gemstone collections. There are many different places where you can see how precious gems have been used in jewelry. Here are some suggestions for interesting places to visit, as well as a list of useful websites that will provide you with plenty more information.

CRYSTALS AND GEMSTONES
In most natural history and geological museums, such as the Earth Galleries at the Natural History Museum in London, England, there are wonderful displays of cut gemstones and famous jewels, specially lit to show them to their best effect. There are also excellent examples of crystals and gemstones in their natural state, often still embedded in their matrix (host) rock.

WHERE GEMSTONES ARE FOUND
Some mineral gemstones, such as quartz and garnet, are found all over the world. Others, such as diamonds, are far more rare. Where gemstones are found depends on particular geological conditions. This map shows the main locations around the world of 12 of the most popular and highly prized gems. If you visit any of these areas, you may be able to visit mines or see samples of the gemstones in local galleries, museums, and shops.

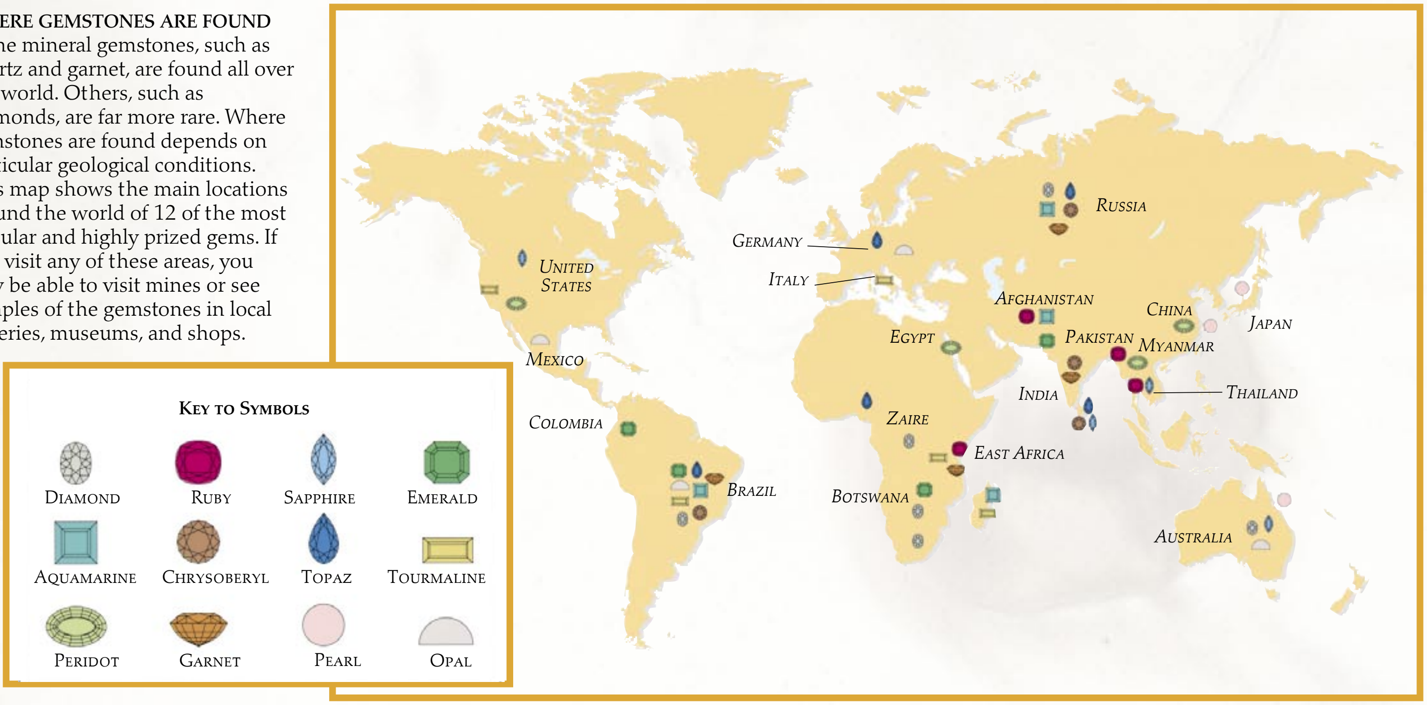
United States
Mexico
Colombia
Brazil
Germany
Italy
Russia
Afghanistan
Egypt
Pakistan
China
Japan
Myanmar
Thailand
India
Zaire
East Africa
Botswana
Australia
Key to Symbols
Diamond
Ruby
Sapphire
Emerald
Aquamarine
Chrysoberyl
Topaz
Tourmaline
Peridot
Garnet
Pearl
Opal

GEM COLLECTIONS
There are often private collections of gemstones on display in museums. You could start a gem collection of your own. Look for specimens on beaches, riverbanks, and hillsides. Clean your finds with water and let them dry, then arrange them in empty matchboxes or small cardboard trays. Try taking them to your local museum for help in identifying them.

Cut gemstones that form part of the Mathews collection in London, England

USEFUL WEB SITES

- Learn about U.S. geography through games and pictures, as well as great hands-on activities you can try at home: **www.usgs.gov/education/**
- See pictures of more than 1,000 different types of minerals: **webmineral.com/specimens.shtml**
- You can see great gems from the Smithsonian Gem and Mineral Collection at: **www.gimizu.de/sgmcol/**
- Games and activities about minerals and mining are a great source of information: **www.womeninmining.org/**

Places to visit

THE NATIONAL MUSEUM OF NATURAL HISTORY AT THE SMITHSONIAN INSTITUTION
10th Street and Constitution Avenue, NW,
Washington, D.C. 20560
This impressive collection is home to some famous gems, including the Hope Diamond, the largest diamond in the world.

HENRY FORD MUSEUM
20900 Oakwood Boulevard
Dearborn, MI 48124-4088
The museum houses a large collection of American-made jewelry up to 300 years old.

CARNEGIE MUSEUM OF NATURAL HISTORY
4400 Forbes Avenue
Pittsburgh, PA 15213
More than 350 minerals are on display in the Hillman Hall of Minerals and Gems.

THE FIELD MUSEUM
1400 S. Lake Shore Drive
Chicago, IL 60605-249
Discover the basics about minerals and gems.

JEWELS AND JEWELRY
A museum of decorative arts is a good place to see how gemstones have been set in jewelry over the ages. In addition to historic pieces, modern jewelry in styles such as Art Deco are well worth looking at. To see examples of early jewelry, try visiting the Ancient Egyptian section of a museum. If you travel abroad, visit local craft museums to see samples of ethnic jewelry.

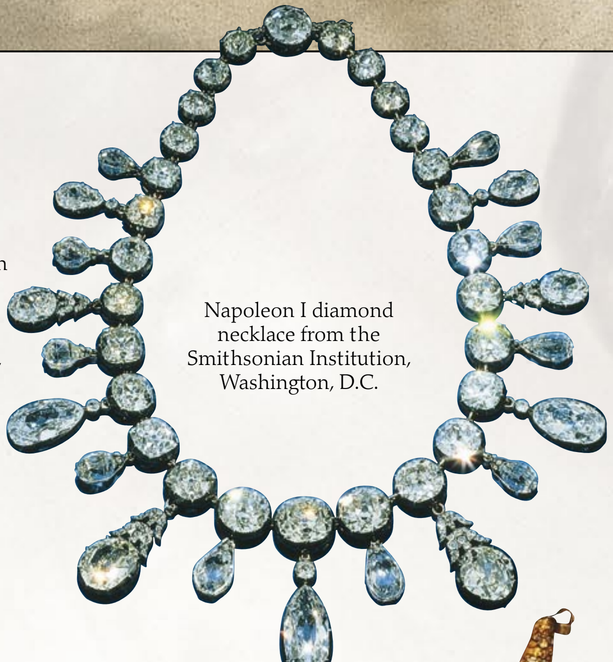
Napoleon I diamond necklace from the Smithsonian Institution, Washington, D.C.

TOPKAPI PALACE TREASURY
If you go to Istanbul in Turkey, it is worth visiting the Topkapi Palace Museum, which has a magnificent collection of Colombian emeralds set in jewelry and other accessories. As well as the dagger below, the collection includes a golden throne studded with tourmaline, a carved emerald snuffbox, and vases carved from exquisite green jadeite, a type of jade.

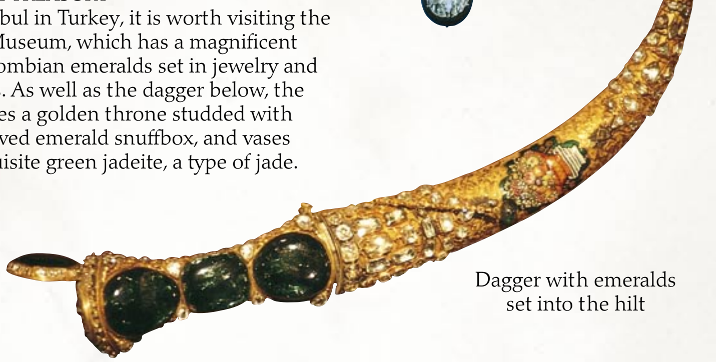
Dagger with emeralds set into the hilt

CROWN JEWELS
Impressive examples of famous gemstones set in gold and silver are on display in the Crown Jewels of France, Britain, and Austria. At the Louvre in Paris, France, you can see the coronation crowns of Napoleon and Louis XV, as well as other magnificent crowns, scepters, and swords, some dating back as far as the Middle Ages. Also on view is the Regent, one of the purest diamonds in the world, which was worn by Louis XV at his coronation in 1722. The British Crown Jewels are on display at the Tower of London and include many jewels that are still used in state ceremonies today, such as St. Edward's Crown, the crown of England, which is only used for coronations, and the Imperial State Crown, which contains the famous Black Prince's Ruby.

Charlemagne (742–814), king of the Franks

Golden scepter set with gemstones, made for Charles V in 1380

Gold set with precious jewels

Crown of Empress Eugénie on display at the Louvre

The Tower of London Education Center in London, where schoolchildren can try on replicas of the crown jewels and royal cloaks, as well as armor.

Glossary

ALLOCHROMATIC A term meaning "other-colored." It describes gems that are naturally colorless but are colored by tiny amounts of impurities.

ALLUVIAL DEPOSITS Weathered fragments of rock that have been carried along in rivers and streams and deposited elsewhere.

AMORPHOUS Without a regular internal atomic structure or external shape.

ASTERISM The star effect seen in some gemstones, such as rubies and sapphires, when they are cut into cabochons.

BIREFRINGENCE (DR) Double refraction, a property of some crystals in which light passing through them is split into two rays.

Brilliant cut diamond

BRILLIANT CUT The most popular cut for diamonds and many other stones. The standard brilliant has 57 facets, or 58 if the gem is cut with a flat face at the base.

CABOCHON A type of cut in which a gemstone is cut into a round or oval with a plain, domed upper surface.

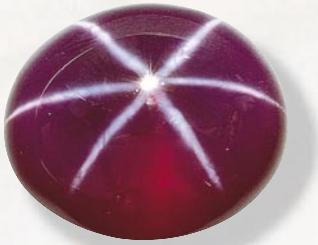

Star ruby cut into a cabochon

CARAT The standard measure of weight for gemstones. One metric carat equals 0.2 g.

CHATOYANCY The tiger's eye effect shown by some stones, such as chalcedony, when they are cut into cabochons.

CLEAVAGE The way in which a crystal splits apart along certain well-defined planes according to its internal structure.

COMPOSITION The fixed or well-defined chemical makeup of a mineral.

COMPOUND A chemical compound is made of two or more elements joined together chemically, which can only be separated by heat or great pressure.

CORE The area of iron and nickel at the center of the Earth.

CRUST The thin outermost layer of the Earth.

CRYSTAL A naturally occurring solid with a regular internal structure and smooth external faces.

CRYSTALLINE Having a crystal structure.

CUT The way in which a gemstone is cut into a number of flat faces called facets, or rounded and polished.

Calcite crystal

DENDRITES Fernlike growths of crystals found lining the cracks and joints in rocks.

DICHROIC A term used to describe a gem that appears to be two different colors when viewed from different directions.

DIFFRACTION The splitting of white light into its constituent colors.

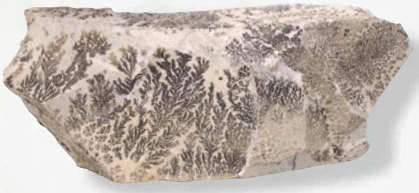

Dendrites of the mineral pyrolusite

DOUBLET A composite stone made of two pieces cemented or glued together.

DURABILITY The capacity to last for a long time without wearing out.

EROSION The wearing away of the surface of the land and rocks by a moving medium, such as water, ice, or the sea.

FACET One flat surface of a cut gemstone.

FACETING Cutting and polishing gems into flat surfaces called facets.

FIRE A term used for dispersed light. A gem with strong fire is unusually bright.

FLUORESCENCE Colored light that radiates from a mineral when it is exposed to invisible ultraviolet light.

Fluorescent crystal

GEMSTONE A decorative mineral or organic substance prized for its beauty, durability, and rarity.

GEODE A cavity within a rock that is lined with crystals that grow toward the center.

GIRDLE The widest part around the middle of a cut stone, where the top half (the crown) and the bottom half (the pavilion) meet.

HABIT The shape in which a crystal naturally occurs. Crystal habit is a key factor in identifying minerals.

IDIOCHROMATIC A term used to describe minerals, such as sulfur, whose color is part of their chemical composition.

INCLUSIONS Material (usually a mineral) trapped within another mineral.

INTERGROWN When two or more minerals grow together and interlock.

Idiochromatic sulfur

IRIDESCENCE A rainbowlike play of colors on the surface of a mineral, similar to a film of oil on water.

LAPIDARY A craftsman who is skilled at cutting gemstones to obtain the best optical effect.

LAVA Magma from within the Earth that erupts to the surface from volcanoes.

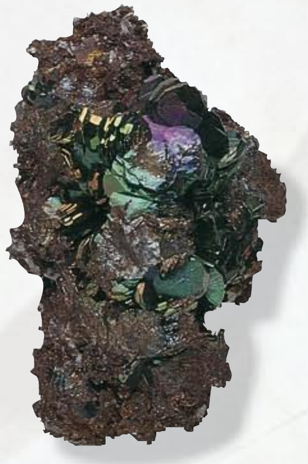

Iridescent hematite crystals

LODESTONE A piece of magnetite, a naturally occurring magnetic iron oxide.

LUSTER The way in which a mineral shines. Its luster is affected by the way in which light reflects off the surface of the mineral.

MAGMA Molten rock deep below the Earth's surface.

MANTLE The layer of the Earth between its core and its crust.

MASSIVE A term used to describe minerals that have no particular shape.

MATRIX A term for the main body of a rock.

METAMORPHOSE To undergo recrystallization in a solid rock, leading to a change in mineral compostion and texture. In rocks, this is usually caused by the effects of heat and temperature.

Magma

MICROCRYSTALLINE A mineral structure in which the crystals are too small to be seen with the naked eye.

MINERAL A naturally occurring inorganic solid with regular characteristics, such as crystal structure and chemical composition.

MIXED CUT A gemstone cut in which the facets above and below the girdle follow different styles, usually a brilliant cut above and a step cut below.

MOHS' SCALE A scale devised by the Austrian mineralogist Friedrich Mohs that measures the hardness of minerals according to what they are able to scratch, on a scale from 1 to 10.

NACRE Tiny platelets of calcium carbonate that create the soft sheen on pearls and inside some seashells as they reflect light.

Nacre inside shell

OPALESCENCE Milky blue form of iridescence.

OPAQUE Does not let light pass through it.

ORGANIC GEM A gem made by, or derived from, one or more living organisms.

PEGMATITE Igneous rocks containing very large crystals, which have formed from the very last water-rich magma to crystallize.

PENDELOQUE CUT Lozenge-shaped cut often used for flawed gems.

PHANTOMS Regular inclusions that occur within a crystal, such as parallel growth layers.

PIZOELECTRICITY A property of quartz crystals. Pressure on a crystal creates positive and negative charges across the crystal.

PLEOCHROIC A term used to describe a gemstone that looks as if it is two or more different colors when viewed from different directions.

PRISMATIC A term used to describe a crystal that is "pencil-like," with elongated crystals.

PROPERTY A characteristic of a mineral, crystal or gemstone, such as its color or habit.

REFRACTIVE INDEX (RI) A measure of how light rays slow down and bend as they enter a gemstone. One of the properties used to identify gems.

RESIN A sticky substance from certain plants.

RHOMB A shape rather like a lopsided cube.

RIVER GRAVELS Deposits of minerals that have been broken away from their host rock and washed downstream.

ROCK A combination of mineral particles. Some rocks contain a variety of minerals and some only one. Rocks may be inorganic, chemical, or biological in origin.

ROUGH A term for the natural state of a rock or crystal before it is cut or polished.

SCHILLER Sheen or iridescence.

SPECIFIC GRAVITY (SG) A property of minerals that is defined by comparing the weight of a mineral with the weight of an equal volume of water.

SPECTROSCOPE An instrument used to identify different gemstones. It reveals the bands of light that a gemstone absorbs.

Coral,
an organic gem

Selenite

STEP CUT A rectangular or square-shaped gemstone cut with several facets parallel to the edges of the stone. It is generally used for colored stones.

STRIATION Parallel lines, grooves or scratches in a mineral.

SYMMETRY, AXIS OF An imaginary straight line through a crystal. If the crystal were rotated about this line, the same pattern of faces would occur a number of times in a full turn.

SYNTHETIC GEMSTONE An artificial stone made in a laboratory that has a chemical composition and properties similar to the natural gemstone from which it is copied.

Step-cut ruby

TABLE CUT A type of step cut with a square table facet and girdle and parallel square facets.

TRANSLUCENT Material that allows some light to pass through it.

TRANSPARENT Material that allows light to pass right through it; able to be seen through.

TWINNED CRYSTALS Two crystals of the same mineral that are joined together at a common plane, known as the twin plane.

VEIN An infilled joint, fissure, or fault. Veins are often made of minerals.

VITREOUS A term used for the glasslike quality of some gemstones. It is used to describe a gem's luster.

Twinned calcite crystals

Index

A

adrenaline, 9
agate, 32, 33, 56, 58, 59
albite, 6,16
Alexander II, Czar of Russia, 35, 47
alexandrite, 47, 67
almandine, 67
aluminum, 32, 37, 42, 49
amazonite, 32
amber, 54, 64, 65, 67
amethyst, 16, 19, 32, 48, 60, 61, 63, 65, 66
amphibole, 38
andalusite, 49
apatite, 9, 19
aquamarine, 7, 32, 38, 39, 56, 61, 66
aragonite, 23
Argyle diamonds, 25, 59
asbestos, 33
aspirin, 62
atoms, 6, 7, 8, 14–15, 18, 20, 21
augite, 14
axinite, 48
azurite, 16, 24

BC

barite, 11, 12
basalt, 9, 32
benitoite, 11, 48, 49
beryl, 11, 13, 15, 17, 21, 23, 25, 38–39, 65, 66
birefringence, 18
bismuth, 26
Black Prince, 46
blende, 48, 49
blue john, 51
Bonaparte, Napoleon, 34
bornite, 24
bournonite, 10
bowenite, 51
cabochons, 32, 44, 47, 58, 66
calcite, 9, 11, 13, 15, 18, 21, 22, 23, 36, 50, 67
calcium, 8, 9, 19, 22, 45, 54, 55
Canning jewel, 55
carats, 52, 56
carbon, 14, 34, 66
carborundum, 27
carnelian, 33, 48, 61, 65, 66
cat's-eye, 45, 47, 49
chalcedony, 32, 33, 66
chalcopyrite, 20, 24
chatoyancy, 33
chromium, 19, 37, 39, 46,47
chrysoberyl, 19, 47, 67
chrysoprase, 33
citrine, 18, 32, 37, 58, 66
cleavage, 15,18, 42
cobalt, 17
collecting, 48–49, 63
copper, 16, 23, 24, 50, 53
coral, 54, 65, 67
cordierite, 48, 60
cornelian (*see* carnelian)
corundum, 17,18, 19, 27, 36-37, 59, 62, 65, 66
covellite, 24
Crown jewels, 35, 46, 53
crystal growth, 20–21, 26, 27; atomic structure, 14–15, 17, 18, 21; form, 13, 21, 22; organic, 9; symmetry, 12–13, 15, 30, 43; uses, 28–29, 30, 31, 61, 62–63
crystallography, 13, 14
crystal gazing, 61
crystal healing, 31
Crystal Palace, 7
crystal set, 62
Curie, Jacques & Pierre, 31
cut stones, 65
cymophane, 47

D

Dalton, 14
danburite, 48
de L'Isle, Romé, 12
Dee, John, 61
dendrites, 7, 23, 53
diamond, 14, 17, 18, 19, 24, 25, 34–35, 56, 57, 58, 60, 61, 64, 66; colored, 66; famous, 35, 46, 58, 65; substitutes, 27, 29, 57; uses, 28–29, 62, 63
doublets, 40, 56
dumortierite, 22

EF

eclogite, 8
electricity, 64
electromagnetic spectrum, 15, 16, 19
emerald, 26, 38–39, 56, 58, 60, 61, 65, 66
emery, 37, 65
epidote, 11
erythrite, 17
Fabergé, 51
feldspar, 8, 9, 16, 17, 30, 45
fibrolite, 49
flourescence, 17
fluorite, 17, 18, 21, 42, 57
fossilization, 41, 54

G

gahnite, 46
galena, 12, 18, 21, 53, 62
garnet, 8, 19, 32, 42, 44, 56, 57, 61, 64, 67
gem collecting, 48–49; cutting, 58–59; polishing, 58–59
geodes, 63, 65
Giant's Causeway, 22
glass, 7, 30, 57
glauberite, 41
goethite, 22
gold, 14, 24, 27, 30, 32, 52
goniometer, 12
granite, 8, 23, 24, 30, 39, 43
graphite, 14
greenstone, 51
gwindel, 31
gypsum, 18, 23

HIJ

halite, 26
hardness, 18, 27, 32, 34, 51
Hauy, R.J., 14, 15
hawk's-eye, 33
heliodor, 7, 38, 39, 59
hematite, 17, 21, 37, 45
hessonite, 44
High Priest's breastplate, 60
hoppers, 21
hornblende, 23
hyalite, 40
ice, 6, 62
idocrase, 12
inclusions, 17, 21, 37, 38, 39, 47, 57, 67
International Union of Crystallography, 13
iridescence, 17
iron, 8, 19, 24, 32, 37, 39, 44, 46, 50
ivory, 65
jade, 50, 51, 67
jadeite, 51, 67
jasper, 32, 33
jet, 54, 65, 67

KL

kimberlite, 34, 64
Kunz, George, 49
kunzite, 42, 49
kyanite, 42, 49
labradorite, 17
lapis lazuli, 50, 56, 58, 60, 64, 65, 67
laser beams, 28, 58
law of constancy of angle, 12
lazurite, 50
lead, 10, 53
liquid crystal display, 62
liroconite, 24
lodestones, 60
Louis XIII of France, 48
Louis XIV of France, 41

M

magnetite, 37, 60
malachite, 24, 50, 51
manganese, 17, 39, 44, 51
melanite, 44
mesolite, 22
meteorites, 8
mica, 8, 15, 30, 38, 46
micromounts, 63
mining, 65
Mohs, Friedrich, 18, 19
Moissan, Henri, 27
Monroe, Marilyn, 35
moonstone, 16, 45, 64, 67
morganite, 38, 39
museum collection, 68–69
mythology, 60

NOP

nacre, 55
nephrite, 51, 67
Newton, Isaac, 16
olivine, 45
opal, 19, 25, 40–41, 57, 61, 66
organic gems, 65, 67
orthoclase, 8, 12, 16, 18, 19
oxygen, 6, 8, 14, 15, 31
painite, 56, 65
panning, 25, 52
Peacock throne, 37
pearl, 32, 54, 55, 56, 61, 64, 65, 67; cultivated, 64; seed, 65
pegmatites, 23, 25, 30, 38, 43, 45, 48
peridot, 8, 19, 42, 45, 49, 61, 67
piezoelectricity, 30, 31, 62
platinum, 52, 53
pleochroism, 36, 43, 49
potato stones, 7, 23, 65
potch opal, 66
prousite, 10
pyrite, 13, 21, 50, 67
pyrope, 67
pyroxenes, 8,14, 53

QR

quartz, 6, 8, 11, 15, 16, 17, 19, 20, 21, 23, 26, 30–33, 57, 58, 66; watch, 31, 63
quartzite, 8, 9, 30
refraction, 18, 19, 47
rhodochrosite, 17
rhodonite, 51
rock crystal, 6, 16, 31, 32
ruby, 19, 36–37, 46, 56, 59, 60, 61, 63, 65, 66
Ruskin, John, 36, 43
rutile, 27, 37, 57

S

salt, 17, 62
sapphire, 19, 36–37, 46, 56, 57, 58, 61, 63, 64, 66
scapolite, 7,49
serandite, 23
serpentine, 39, 51
shell, 54, 55
siderite, 20
silica, 9
silicates, 14–15
silicon, 8, 14, 15, 26, 31; chips, 28, 62
sillimanite, 49
silver, 10, 24, 52, 53, 62, 63
sinhalite, 49
smart cards, 28
Sorel, Agnès, 35
specific gravity, 18
spectroscope, 19
spectrum, 16, 19
sphalerite, 49
sphene, 42, 48
spinel, 19, 27, 37, 46, 57, 65, 67
spodumene, 49
stalactites, 9, 22
stalagmites, 9
star ruby, 37, 59
star sapphire, 59, 64
staurolite, 20, 60
Steno, 12
stibnite, 20
sugar, 20
sulfur, 16
sunstone, 45
Sutton Hoo, 44
synthetic crystals, 6, 7, 26–27, 56, 57, 63
symplesite, 19

TUV

talc, 18
tanzanite, 48
Tavernier, Jean Baptiste, 56
tiger's-eye, 33, 66
tin, 24
titanium, 32, 37
topaz, 11, 12, 15, 18, 19, 25, 42–43, 56, 57, 61, 64, 66
Topkapi dagger, 39
tortoiseshell, 64
tourmaline, 6, 19, 21, 25, 36, 42–43, 58, 66
triplets, 40, 56
turquoise, 50, 56, 58, 61, 65, 67
twinning, 17, 21, 23, 46, 60
uranium, 47
vanadium, 39
Verneuil, Auguste, 27
Victoria, Queen of England, 35
vitamin C, 62
van Laue, Max, 14, 15

WXZ

watermelon tourmaline, 66
Wilton Diptych, 50
wulfenite, 22, 63
X-rays, 14, 15, 19
zinc, 19, 46, 49, 53
zircon, 37, 47, 57, 67

Acknowledgments

Dorling Kindersley would like to thank: Peter Tandy at the Natural History Museum for his expert advice and help; Karl Shone for additional photography (pp. 28-29, 62-63); De Beers Industrial Diamond Division for the loan of diamond tools (p. 29); Gemmological Association of Great Britain for the gem certificate (p. 57); Keith Hammond for the loan of the beryl crystal (p. 21); Nancy Armstrong for the loan of the prospector's brooch (p. 41); Jane Parker for the index.

For this edition, the publisher would also like to thank: Dr Wendy Kirk for assisting with revisions; Claire Bowers, David Ekholm–JAlbum, Sunita Gahir, Joanne Little, Nigel Ritchie, Susan St Louis, Carey Scott, and Bulent Yusef for the clip art; David Ball, Neville Graham, Rose Horridge, Joanne Little, and Sue Nicholson for the wall chart.

The publisher would like to thank the following for their kind permission to reproduce their images:

Picture credits c=center; b=bottom; l=left; r=right; t=top

Alamy Images: vario images GmbH & Co KG 28bl; **Peter Amacher**: 48cl; **Ancient Art and Architecture Collection**: 9cl; **Archives Pierre et Marie Curie**: 31bc; **Art Directors & TRIP**: 69cr; **Aspect Picture Library/Geoff Tompkinson**: 28br; **Dr Peter Bancroft**: 45cr; **Bergakademie Freiberg**: 12cl, 19br; **Bibliotheca Ambrosiana, Milan**: 13cr; **Bibliotheque St. Die**: 53cl; **Bridgeman Art Library, London / New York**: Egyptian National Museum, Cairo, Egypt 65tc, 18tl, 52tr, 58bl; Bibliotheque Nationale, Paris: 42tr; **Paul Brierley**: 16cr; **F. Brisse, "La Symetrie Bidimensionnelle et le Canada", Smithsonian Institution, Washington DC**: 37tr, 39t, 42br, 69tr; **Canadian Mineralogist,** 19, 217-224 (1981): 13tc; **British Geological Survey**: 63br; **A. Bucher/Fondation M.S.A.**:30tl; **Gordon Bussey**: 62tl,/Bibl. Magazin, Paris: 60br; © **Christie's Images Ltd**: 27br, 64tl, 65cb; **Christie's, New York**: 57cl; **Bruce Coleman/Michael Freeman** 52c; **Lawrence H. Conklin**: 49bl; **Corbis**: 64b, 69br; **Crown copyright:** 46cr, 53tl, 58cr; **De Beers**: 29bl, 29tc, 34cr, 35br; **Dorling Kindersley/Eric Crichton**: 47tc, **DK Picture Library: Natural History Museum** 66c; **e.t. archive**: 41br; **Mary Evans Picture Library**: 7tr, 15tl, 23br, 24cl, 38/39c, 38bl, 46tr, 57tr, 58c, 61tr, 61bl; **Fondation M.S.A.**: 11 tl; **Michael Freeman**: 25tc, 37c; **Grukker & Zn, Netherlands**: 28tr; **Robert Harding Picture Library**: 44tr, 51tl; **Harvaard Mineralogical Museum**: 20cl; **Harvey Caphin Alaemeda, New Mexico**: 50tr; **Ernst A. Heiniger**: 39br; **Michael Holford**: 7cl, 37br, 44cr, 50br; **Image Bank/Lynn M. Stone**: 33cr,/Lionel ASY-Schwart: 54bl; **India Office**: 36tr, 56tr; **Kobal Collection**: 35cr; **Kodak Ltd**: 63c; **Kunsthistoriches Museum, Vienna. Photo**: **Courthault Institute of Art**: 40tr; **Lauros-Giraudon**: 34bl; **S.E. Little**: Octopus card Ltd 28clb; **Mansell Collection**: 15bl, 35cl, 45c, 54tr; **Moebius/Exhibition Bijoux Cailloux Fous, Strasbourg:** 10bc; **Museum national, d'Histoire Naturelle, Paris**: 48cr; **Museum of Manking**: 52bl; **National Gallery**: 50bl; **Natural History Museum**: 15tr, 19tr, 19c, 33br, 40bc, 51br, 68tl/Frank Greenaway FRPS: 11 bl, 21bc,/P.Krishna, SS Jiang and A.R. Land: 21tc,/ Harry Taylor ABIPP 31br; **National Portrait Gallery, London**: 16c, 43tr; **Northern Island Tourist Board**: 22bl; **Perham's of West Paris, Maine**: 23tr; **Phototake, NYC/Yoav Levy**: 30cr; **Katrina Raphaell**: 31cr; **Réunion Des Musées Nationaux Agence Photographique:** Musee de Louvre 69bc, 69bl; **Ann Ronan Picture Library**: 27tc, 27cr, 55tc; **Royal Geographical Society**: 36bl, 39tc, 56bl; **S. Australian Dept of Mines and Energy/B.Sowry**: 41tc; **Science Photo Library**: 9tr, 14bl, /Dr Jeremy Burgess: 6tr, 62c, 63tr, /ESA/ PLI:8tl, /John Howard: 43cr, /Peter Menzel: 25c, /NASA: 26c, /David Parker: 9br, /Soames Summerhays: 8bl; **Brian Stevenson & Co**: 25bl, 59tc; **Stockphotos**: 20br; **R. Symes:** 20tr; **Uffizi, Florence. Photo**: **Giraudon**: 32cl; **Victoria and Albert Museum**: 51cr, 55bl; **Werner Forman Archive**: 61cl; **Peter Woloszynski**: 49cr; **Zefa** /Leidmann: 32tr, /Luneski: 60cl.

Illustrations by Thomas Keenes.

Wall chart: PunchStock: Photodisc c.

Jacket credits:
Front: B: Lawrence Lawry /Science Photo Library; Cr: Francis G Mayer /Corbis.

All other images © Dorling Kindersley.

For further information see:
www.dkimages.com

Eyewitness
CRYSTAL & GEM

DR R.F. SYMES & DR R.R. HARDING
in association with
THE NATURAL HISTORY MUSEUM

Be an eyewitness to the stunning natural beauty of crystals of every size, shape and colour – and their remarkable uses in everything from surgeons' scalpels to silicon chips.

See
a crystal's hidden layers

Find out
how to make stones sparkle

Discover
where to find crystals in your home

Discover more at
www.dk.com

Jacket image *Front:* Getty Images: C Squared Studios (c).

rexresearch.com

Dr Ibrahim KARIM

Biogeometry

Ad
MAKE AN IMPACT
Earn Your Credential Online
Alliant International University
LEARN MORE

fetchingstock
shutterstock

The Science of BioGeometry

by

Dr Ibrahim Karim

**©1997
BioGeometry Energy Systems Ltd. & Dr. Ibrahim karim.**

Balancing the activities of daily life, achieving harmony with our inner and outer environments, humanizing modern technology, integrating science and spirituality, and discovering the unified scientific reality behind all religions is the work of the science of BioGeometry®.

Just two decades ago everybody was afraid that the end of life on earth would come as a result of a Third world War. We thought that the global extinction of life on earth would come from nuclear warfare and our carelessness towards our environment, endangering plant and animal life; the ozone hole and global warming causing natural disasters everywhere. However, the real potential danger is one that we are barely aware lurking just around the corner.

The age of information carries with it the potential to the global extinction of our civilization. We are continuously increasing the amount of carrier waves needed for the wireless technology of modern communication in the earth’s atmosphere every day. These electromagnetic waves are thousands of times stronger than the level used in the communication in our body cells. The problem is not the saturation of the earth's atmosphere through quantity, but also a detrimental quality. Even people who avoid using high technology are not immune. No one is immune even if one lives at the far end of the world or on top of the Himalayas, because these are carrier waves with penetrating properties. It is like trying to move a breeze through a storm, our immune systems are continuously trying to correct the distortion in the transfer of inner information in our body; very soon the threshold will be reached when a total collapse of our body defenses will take place.

BioGeometry® is a science that deals with the Energy of Shape; it uses shapes, colours, motion, orientation and sound to produce a vibrational quality that balances energy fields. BioGeometrical shapes are two or three-dimensional shapes specially designed to interact with the earth's energy fields to produce balancing effects on multiple levels on biological systems. They were developed and patented by Dr. Ibrahim F. Karim, D.Sc. in Cairo, Egypt, during research since 1968.

To understand the effects of BioGeometrical shapes on the human energy system, we have to recognize that the human body has an energy field around it, which has its own north-south axis. As we move around, the angle formed between our individual axis and that of the earth is constantly changing, and this in turn either strengthens or weakens our energy field. The positive range is very small, vulnerable most of the time, and is a major factor affecting our health and well being. BioGeometrically balanced energy, however, considerably strengthens our energy fields to such an extent that we are not detrimentally affected by changes of orientation. In fact, it appears to cancel obvious energy interactions predicted by currently accepted physical laws.

Research in BioGeometry® was and still is mainly dedicated to the development of a new form of architecture that would enhance the human biological system and give a new meaning to the concept of "Home". To upgrade the energy quality of existing homes and cancel the potentially harmful effects of unchecked energy fields due to the architectural design, furniture layout, electrical wiring and modern appliances, specially designed decorative elements are strategically placed to neutralize negative energy and add a positive quality to it. BioGeometrical shapes when designed or engraved on jewelry have shown positive effects on the body's energy field, and reduce considerably the potential health hazards caused by cellular phones, computers and all other modern appliances.

The impact of geometrical shapes on human energy systems has always been universally recognized. This awareness gradually disappeared, and our "modern" approach is to consider these ancient forms either as symbolic art without function, or attribute them to magical practices. Although the modern development and practices are not directly derived from Ancient Egypt, a very advanced know-how in this field is evident when we analyze and use the shapes that they have developed thousands of years ago.

They must have been able to interact with nature in a more advanced way than we do today, based on the study of the vibrational properties of the geometrical shapes they used in their monuments, art, statues, amulets, and many other aspects of life. The effects of energy went beyond the Pyramid shape; the Ancient Egyptians used it in a very practical way in all aspects of their life. Unlike our modern energy forms that are highly amplified states of energies occurring in nature, this science dealt with forces on a natural level as they occur in nature. That means that this Ancient Egyptian science was more like a language that they used to establish two-way communication with nature. This two-way information flow process was used to establish total harmony in all actions between man and nature, to acquire a deeper knowledge about anything by accessing information about it's working principles on the energy level, and most important of all to influence any action in nature by manipulating its energy patterns to achieve the intended results.

Pythagoras was the first to introduce to the western world the ancient Egyptian way of correlating musical qualities with quantifiable, numerical values. The golden ratio of 1.618 expressed as the ultimate proportion of harmony, beauty and spirituality was used in the design of sacred buildings in Ancient Architecture to produce spiritual energy that facilitated connectivity with spiritual realms through resonant prayer. Popular among spiritually significant shapes are pyramids and hemispheres (e.g. the domes, that are the basis of religious buildings, be it a mosque, a church or a synagogue). These particular shapes are energy emitters; they are shapes that produce a type of penetrating carrier wave which Chaumery and De Belizal named, negative green, (which acts as carrier, like radio waves that carry sound information). The vibrational quality of negative green gives it very strong communication properties, which facilitate resonance with higher realms in prayer. Negative green turned out to have other properties, however, which make it very harmful under continuous exposure. Considerable research into this type of energy has been done by Dr. Karim, and the different components have been identified.

A revival of the Ancient Design Criteria or Canons in modern architecture was attempted by the Swiss pioneer of the modern architecture, Le Corbusier, with his "modular" system, which comprised two scales of dimensions based on the golden ratio.

Dr. Ibrahim F. Karim has done extensive research and found that BioGeometrical shapes have three primary vibrational qualities:

1) Negative green.
2) A higher harmonic of ultra-violet.
3) A higher harmonic of gold.

Only shapes, which produce energy fields with all three components, are BioGeometrical.

The effect of BioGeometrical energy on health is not specific and not precisely predictable. It appears to amplify and balance the energy fields of the body on all levels, and thereby give the body greater power to heal itself. The healing process resulting from a strengthening and balancing of the immune system manifests differently from one person to the other; certain results, however, have been repeatedly observed.

BioGeometrical shapes balance the body energies on different levels; positive effects are usually felt on the emotional, mental, spiritual as well as the physical level. They have been found to be effective over a very broad range, including the protection against harmful radiation emanating from the earth (believed to be a major cause of cancer) and different types of man-made pollution. For specific healing purposes, research is being conducted in collaboration with medical doctors in the science of BioSignatures®, which deals with energy of shape in relation to specific functions of the body organs.

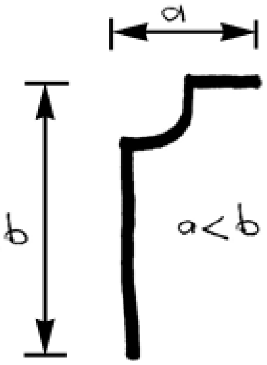

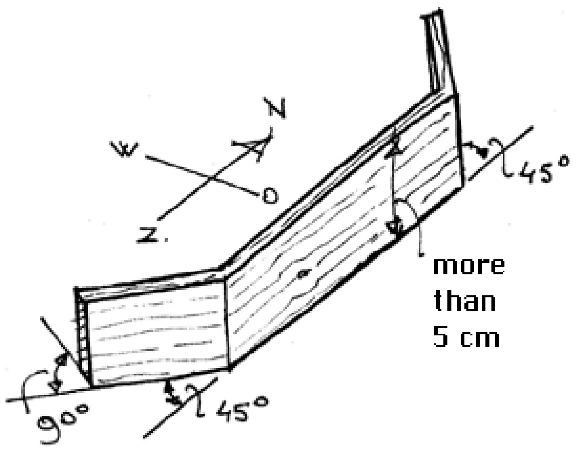

http://users.pandora.be/wouterhagens/biogeometry/index.html ---
http://www.biogeometry.org ---

The BioGeometrical Systems Institute Company

The BioGeometrical Systems Institute Company was founded in 1993 by Dr. Ibrahim Karim, D.Sc., and Mrs. Rawya Karim, M.A. It was envisioned as a design center to research and implement the new science of BioGeometry, developed & patented by Dr. Karim. The BioGeometrical Energy system Ltd. is a multi-disciplinary organization working in the following fields:

1. A BioGeometrical design language used in architectural, planning decoration, industrial design and accessories.
2. Energy adjustment on all levels and protection from harmful earth radiation ranging from a building to a whole city.
3. The development and use of BioSignatures to support traditional and alternative medicine.
4. Environmental application and protection from side effects of modern technology. For example electronic and cellular devices
5. Agriculture research in organic planting eliminating pesticides and prolonging shelf life.

---

http://www.lightconnectiononline.com/Archive/sep04_article2.htm ---

BioGeometry: The Egyptian Art of Energy Balancing

by

Leanna Joyner

BioGeometry has been called the Egyptian version of Feng Shui but its unique methods and effects are not found in any other system. BioGeometry combines previously hidden methods used in the Inner Temple Sciences of Ancient Egypt with modern European subtle energy detection tools. Developed by renowned Architect and energy researcher Dr. Ibrahim Karim of Cairo Egypt, BioGeometry has been shown to have astonishing benefits for humans, animals, plants, and the environment.

BioGeometry utilizes Nature's own design language of geometric form, color, sound and motion to balance the Energy in all living systems. It also features profound methods to transmute harmful environmental energies into beneficial ones. Robert Gilbert Ph.D., an expert in Sacred Geometry and the author of the only English language texts on BioGeometry, identifies BioGeometry as an essential key to understanding the Living Energies which every spiritual tradition uses for Healing and Spiritual Development.

“Working directly with Dr. Karim and his experts, I found they had unique insights into practical energy work. I'm still in awe of how clearly they explain energy principles which are usually thought to be very mysterious; when you work with the Egyptians, one piece of essential knowledge after another just seems to fall into place. For example, virtually every tradition knows about the Golden radiance which appears in the energy field of great Saints and Masters, but only this group of Egyptians can show you simple methods anyone can use to detect that energy, create it with methods including movement and geometric forms, and also practically apply it to balance and heal a variety of energy problems. These Egyptian methods, now that they are finally becoming public, are going to have a powerful impact on how we understand and work with healing energies,” Dr. Gilbert explained.

Although BioGeometry makes no claims, multiple scientific and medical studies suggest its potential to create powerful energy effects. Just a few examples:

The Egyptian National Hepatitis C Research Project under Dean Taha Khalifa M.D. at Al Azhar University's Pharmaceutical School announced publicly that the best results in the entire study were enjoyed by those given the appropriate BioGeometry patterns to wear. Although Dr. Karim insists that BioGeometry is simply a supportive modality that he makes no medical claims for, the publicity led to Dr. Karim becoming a celebrity with his own television show on holistic health and spiritual development.

The Egyptian National Research Centre discovered that simple BioGeometry forms could stop the replication of Bacteria, a result so astonishing that they created a special project to study the effect of Geometric Forms on Living Systems under the leadership of Dr. Karim.

A project evaluated by an official of the Egyptian Department of Agriculture found that chickens grew healthier and faster in an environment which was energy balanced by BioGeometry methods than they did with antibiotics and growth hormones.

Professor Peter Mols of Wageningen Agriculture University in Holland found that BioGeometry energy methods could be used in place of pesticides and artificial fertilizers to grow healthy organic crops.

The Swiss village of Hemberg suffered both human and animal health problems after a powerful Microwave antenna array was placed in the center of town. Dr. Karim balanced the electromagnetic pollution problem for the entire area, leading to such a dramatic improvement in living conditions (and the return of birds and other animals which had fled the area) that he was given the Key to the City by the Mayor and featured on national Swiss news programs.

”In my personal experience, I've used BioSignatures (precise BioGeometry energy patterns linked to different parts of the body) to help free me from the chronic neck and spine pain I suffered from a severe car accident,” said Dr. Gilbert. “We get incredible reports all the time from students regarding the power of the BioGeometry methods, but we don't make them public. Frankly, many just seem unbelievable if you don't understand how BioGeometry works. Also, we want everyone to understand that we don't make any medical claims whatsoever; we simply teach the principles and methods and leave it to the students to find out for themselves how powerful BioGeometry can be in many cases.”

BioGeometry has only been taught publicly for a short period of time. But already several hundred Americans have been trained in the Art - first by Dr. Karim and now by Dr. Gilbert (Dr. Karim no longer teaches Foundation trainings in BioGeometry, and rarely teaches at all.) Students have applied BioGeometry to a wide range of fields including conventional and alternative Health Care, Environmental Restoration, Architecture and Interior Design, Feng Shui, Spiritual Practices, Landscape Design, Agriculture and Gardening, and Graphic Design.

“Basically, BioGeometry teaches practical methods to work with the fundamental energies of Creation. It can be applied to virtually any field of work, and blends with any methods one currently uses. In collaboration with Dr. Karim, we've set up the Foundation Training so that we can teach all the essential BioGeometry methods to detect, create, and apply beneficial energies in just one week. We also make sure students understand what the Egyptians call Energy Alchemy, how to transmute harmful environmental energies into helpful ones. But perhaps most importantly, BioGeometry doesn't ask for students to take anything on faith. Rather it provides the tools to directly detect and monitor energies which shape the world around us,” said Dr. Gilbert. “I'm deeply grateful to the Egyptians who have made these secrets public for the first time. It's having a profound effect on a lot of people's lives.”

Dr. Robert J. Gilbert is founder of Vesica: Spirit and Science Resources and the first non-Egyptian ever granted permission to teach certified trainings in BioGeometry. Vesica is offering a rare BioGeometry Foundation Training in San Diego from October 9 -15, 2004. More information on BioGeometry, the foundation training and registration is available online at www.vesica.org

---

http://www.amcham-egypt.org/Publications/BusinessMonthly/August%2099/Followup.asp ---

Those companies embracing the growing trend of custom designed, built-from-scratch workplaces might want to give Dr. Ibrahim Karim a call --- that is unless they want to risk spending millions to produce an office full of creativity-destroying energies.

According to Karim, a proper office requires more than just comfortable chairs and soft lighting. As inventor of a science he calls biogeometry, Karim claims he can manipulate energy fields to create an office environment that produces healthy, happy, creative workers.

“We can transform any building into an energy balancing and protective second skin,” he said. “You can raise the quality of life of [the employees]. They’ll feel more energized, more creative.”
Karim, a Zurich-trained architect and consultant to the Ministry of Scientific Research, said that the average office worker in Egypt or anywhere else toils away amid “architectural pollution” that can wreak havoc on both his body and mind. Each employee, Karim said, is like a “breeze trying to move through a heavy storm” of disruptive energy and electronic noise. “In any office, in any city, you’ll find a mess.”

But biogeometry, Karim claimed, is the answer. After “measuring the interaction of energy fields” in a room or building, Karim will recommend changes in room color, furniture shape or orientation. Specially designed picture frames can be used to emit balancing energy. “Stand in front of one, and you’ll feel better within five minutes,” Karim said. Electronic devices such as computers or cellular phones inherently produce negative energy, but they can be rendered less disruptive by engraving special symbols into their sides.

On the surface, biogeometry seems to borrow liberally from feng shui, an Asian design philosophy that attempts to manipulate energy through shape, color and decor. But Karim said biogeometry is based in modern physics, not feng shui, which anyway has become watered-down and faddish. “The old masters are gone, and the essence has been forgotten,” Karim said.
There’s another difference. Feng shui is an accepted part of Asian architectural practice, and a feng shui consultant is brought in at the start of construction projects in many Asian cities. Biogeometry isn’t quite there yet in the eyes of the local architectural or business community. Karim hasn’t been contracted to apply his philosophies that often in Egypt. He recently worked on energy adjustment at a Montessori school in Maadi, and he’s building his own re-sort in Ain Sukhna where everything from the furniture to the landscaping will be designed to give an energy boost. “We’re working on all levels,” he said. “It’s a health spa where the most important activity is just being there.”

Karim gets a far more receptive reaction overseas. He’s spent several months working on energy modification at the industrial design building at Eindhoven in Holland, where he seems to be regarded as a genuine visionary. Most likely, Karim’s local profile will rise once Egyptians realize that the foreigners are taking him seriously.

Among Karim’s architectural peers, reactions to his theories ranges from enthusiastic support to polite bemusement. Karim’s peers seem to give him a lot of leeway – partly because his father, Sayed Karim, is a legendary Cairo Uni-versity architecture professor who helped design Medinat Nasr, and partly because Karim carries the fervor of a true believer, offering free seminars and classes for curious architects and designers. Even architects who consider Karim a nut consider him a sincere nut.

“Give the guy a chance,” said Cairo University architecture and urban planning professor Sayed Ettouney. “You know somebody is a crook when they’re after money or power. But he’s offering his knowledge as a service. He’s spending his own money on this.”

Egypt’s medical establishment reacted less charitably when Karim ventured into their field by claiming he could regulate the body’s energy levels and heal the sick by using special “biosignature” symbols hung from necklaces or bracelets. In February, Karim appeared on the popular interview show Talk of the City, along with patients who said they’d been cured by his biosignatures. The response was immediate and twofold: first, he was denounced in print by the Doctors’ Syndicate and the minister of health; second, his Maadi office was swamped by desperate crowds seeking miracle cures.

The controversy, and the fear that his office might be overrun, led Karim to cut back on seminars and medical treatments months ago. The gate of his Maadi villa now bears a large sign turning away medical cases with the blunt pronouncement, “This is an architecture office.” But the calls keep coming.

“We get calls every day from people who think he can help them have children or cure their polio,” said Karim’s office director. “It’s like a he’s a fugitive. We have to hide him.”

Like any good magician, Karim is reluctant to get into the specifics of his work. Ask him to describe the basics of what shapes or colors create a healthy office energy level, and he'll say he doesn't want to get into "ready-made solutions." Ask just how he discovered that this particular squiggly line can help cure liver failure or fight depression, and he'll smile and say he worked on it for many years. Even his own students and devotees are stumped.

"We asked where these come from, and he says, 'Guys, that's a very big subject. Let's not get distracted,'" said Mu-hammed Samir, a young architect who took several of Karim's free training seminars. But despite the absence of explanations, Samir believes that Karim is genuinely on to something. "People have been treated successfully. The re-sults are in front of us," he said. "Where it comes from, we don't know and he won't say."

---

http://www.vesica.org ---

BioGeometry: The Egyptian Science of Energy --- An Interview with Dr. Robert J. Gilbert

[Excerpts]

What is BioGeometry?

BioGeometry is remarkable for how vast the possibilities are for applying it in virtually any field you can name! Because BioGeometry deals with the fundamental energies of Creation, those which are essential to human health and higher consciousness, it can be used to support and empower many different approaches to energy work. What I find particularly valuable are the tools which BioGeometry offers so that anyone can work tangibly with different energies rather than "assume" the energies are present or just follow someone else's techniques and hope for the best! BioGeometry offers methods to directly detect, measure, create, and transmute specific qualities of energy, so that you really know the effects you are creating with your energy work. It really is a huge step forward in the human race's ability to understand and work responsibly with energy fields. And I believe this is just the beginning; many of those who train in BioGeometry will use this knowledge to further advance humanity's work with energy into even more advanced levels which most of us can't even imagine today.

BioGeometry is a powerful science which shows how to use shapes, patterns, colors, sounds, and motion to create specific energy effects. It clearly identifies core principles as well as providing a wide range of practical techniques. BioGeometry does not ask for the student to take anything on faith, rather it provides the tools to directly detect and monitor the energies which shape the world around us. BioGeometry reveals publicly for the first time many of the ways which spiritual energy works in the world to create life, health, and consciousness, information which was previously restricted to small groups of spiritual initiates. Put simply, BioGeometry provides critical and unprecedented insights into the use of living energy for beneficial purposes; its simple but extremely powerful techniques can be used by anyone to help humans, animals, plants, and/or the environment.

Can you give us any concrete examples of how BioGeometry has been applied in specific fields or projects?

Absolutely. BioGeometry has been studied and applied in a wide range of fields and projects, including studies at Egyptian and European Universities. Here is just a short list of successful studies:

The Egyptian National Research Centre determined that BioGeometry forms could stop the replication of Bacteria, a result which they stated in their report that they could not explain according to modern scientific theory.

Al Azhar University in Cairo conducted a national Hepatitis C Research Project under the sponsorship of the Egyptian Government, in which they evaluated a wide range of Pharmaceuticals and treatments that might help Hepatitis C sufferers. One of the methods evaluated was BioGeometry, with a particular group in the study given only a medallion with particular BioSignatures on it to correct their energy imbalances. At the end of 6 months, Dr. Taha Khalifa who headed the study announced on national Egyptian television that the most successful results in the whole study (with 90% normalization of enzyme levels on average, and some subjects with complete viral clearance) were enjoyed by the patients who were in the BioGeometry group.

A project evaluated by an official of the Egyptian Department of Agriculture found that chickens grew up healthier and faster in an environment which was energy balanced by BioGeometry methods than they did with Antibiotics and Growth Hormones.

Prof. Peter Mols of the Wageningen Agriculture University in Holland found that BioGeometry energy methods could be used in place of Pesticides and Artificial Fertilizers to grow healthy organic crops with a greatly extended shelf life.

A 1998 agricultural project on the Red Sea headed by Adel Ammar discovered that crops could actually grow better in salt water than in fresh water if the salt water was treated with BioGeometry energy methods.

This is not even to mention the tremendous amount of anecdotal evidence from people who report having all types of different conditions helped by BioGeometry methods; even a Reuters New Service article carried by CNN.com and ABCNews.com contained such reports. Of course we do not make any medical claims whatsoever for BioGeometry; no system that uses subtle energies can make such claims, since the dominant medical and scientific model asserts that these subtle energies don't even exist. So we just tell people to learn the system and make up their own minds from the results they get!

Another article from Vesica.org ---

What the Ancient Egyptians Knew... And Modern Scientists Can't Explain

CAIRO EGYPT, August 5 1993: A team of researchers at the Egyptian National Research Centre watches as an Egyptian Architect places a simple geometric form over a culture of bacteria. Reputed to be a master of the lost energy science of Ancient Egypt, the Architect has been brought to the Centre to test his ability to create specific energy effects through the use of shapes, sounds, colors, and movements. The researchers are certain that they will easily discredit such 'nonsense' by designing controlled experiments to test his abilities.

To their amazement, the Architect's claim to be able to stop the replication of the bacteria through the subtle energy emitted by a simple geometric form proves true. In the report of the experiment released by the Head of the Department of Microbial Chemistry, he notes "quite a noticeable cessation of the growth of Saccharomyses cerevisiae was observed... the essence of this type of interaction seems to surpass explanation through our traditional and classical knowledge in basic science."

After this success, other scientists and researchers took up the challenge of testing the Egyptian Architect's energy methods; in every case they are confounded by the powerful effects:

Al Azhar University in Cairo conducted a national Hepatitis C Research Project in which they evaluated a wide range of Pharmaceuticals and alternative treatments. One of the methods evaluated was a medallion which held hundreds of tiny energy patterns termed 'BioSignatures', which the Architect had determined would balance the energy of Hepatitis C sufferers. At the end of 6 months, Dr. Taha Khalifa, head of the study, announced on national Egyptian television that the most successful results in the whole study — with 90% normalization of enzyme levels on average, and some subjects with complete viral clearance— were enjoyed by the patients whose only treatment had been wearing the medallion.

Professor Peter Mols of Wageningen Agriculture University in Holland found that BioGeometry energy methods could be used in place of Pesticides and Artificial Fertilizers to grow healthy organic crops with a greatly extended shelf life.

An agricultural project on the Red Sea headed by Adel Ammar discovered that crops could actually grow better with salt water irrigation than in fresh water if the salt water was first treated with these subtle energy methods.

Biogeometric Carpet:

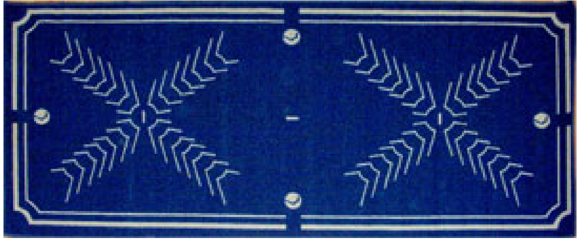

Burns Sign:

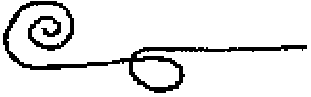

Car Protection Signs:

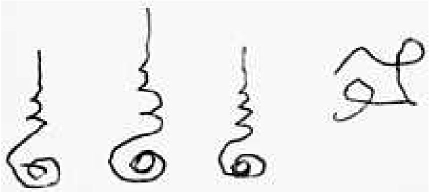

Biosignatures:

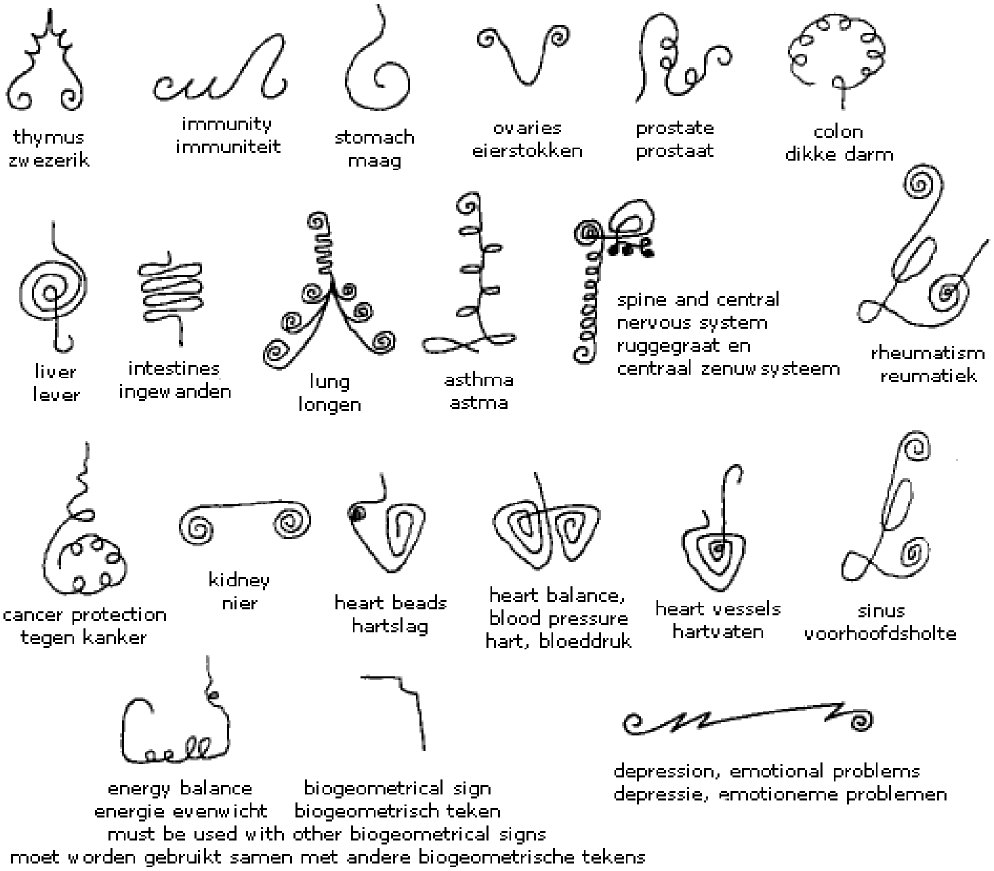
thymus
zwezerik
immunity
immuniteit
stomach
maag
ovaries
eierstokken
prostate
prostaat
colon
dikke darm
liver
lever
intestines
ingewanden
lung
longen
asthma
astma
spine and central
nervous system
ruggegraat en
centraal zenuwsysteem
rheumatism
reumatiek
cancer protection
tegen kanker
kidney
nier
heart beats
hartslag
heart balance,
blood pressure
hart, bloeddruk
heart vessels
hartvaten
sinus
voorhoofdsholte
energy balance
energie evenwicht
biogeometrical sign
biogeometrisch teken
must be used with other biogeometrical signs
moet worden gebruikt samen met andere biogeometrische tekens
depression, emotional problems
depressie, emotioneme problemen

Biogeometry Shapes & Methodology for Biological Energy Balancing

WO2005062691

7-14-2005

KARIM IBRAHIM FAHMY (EG)

Application number: WO2003EG00017 20031231
Priority number(s): WO2003EG00017 20031231
Also published as: AU2003287943 (A1)

Abstract ~ Geometric shapes and techniques, along with the methodology behind them, invented to produce a qualitative energy balancing effect on biological systems, be it human beings or otherwise. This results in a positive effect on biological functions, either preventative or healing, and protection from environmental energy disturbances. The methodology introduced by the inventor is the discovery of the qualitative energy components of earth energy power-spots {specifically those components are qualitative higher harmonics of the frequencies of gold and ultra-violet, and a carrier wave in the negative green of the energy spectrum} and development of geometric shapes and techniques to reproduce and amplify that energy balancing quality. This methodology has been termed/copyrighted "BioGeometry" by the inventor. The geometric shapes and techniques work either directly through proximity to the biological system or from within its surrounding environment, or by superimposing on a radiating energy source that reaches the biological system. The two techniques are: (1) Using two of the shape (Fig. 4) as geometric polarizers and orienting them in relation to a central axis to produce the energy balancing quality as in (Fig 7) and termed "BioGeometry Dial", (2) Using shape (Fig. 4) as an individual geometric polarizer and aligning samples of materials from the biological system itself or its environment to produce the energy balancing effect (Fig. 8) and termed "Energy Balancing Wheel" when in two-dimensional form and "Energy Balancing Sphere" when in three-dimensional form.

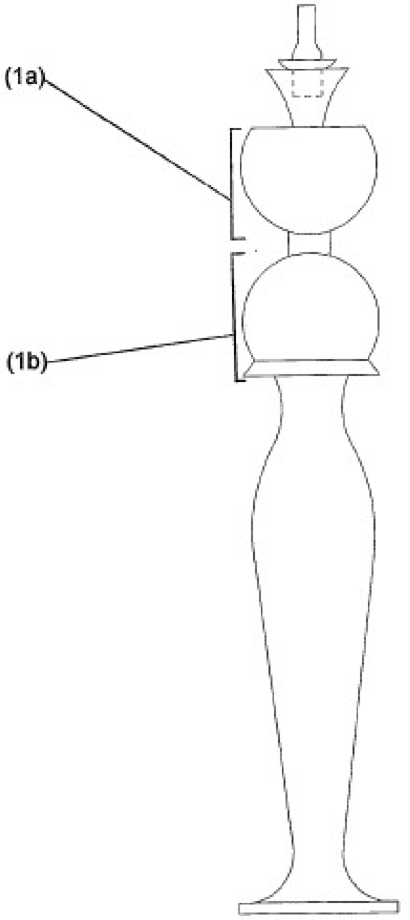
(1a)
(1b)

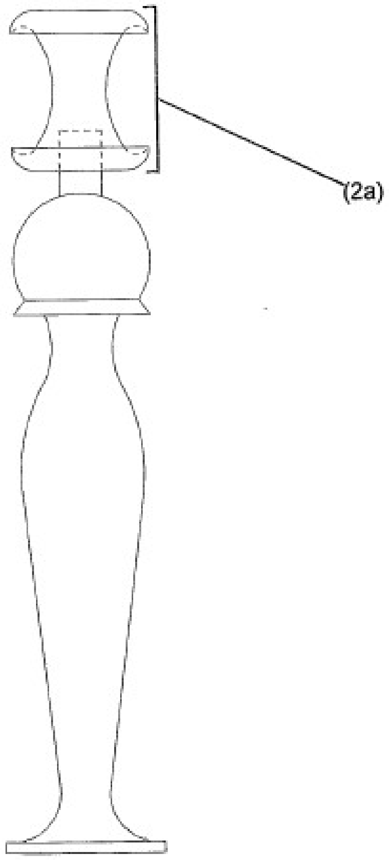
(2a)

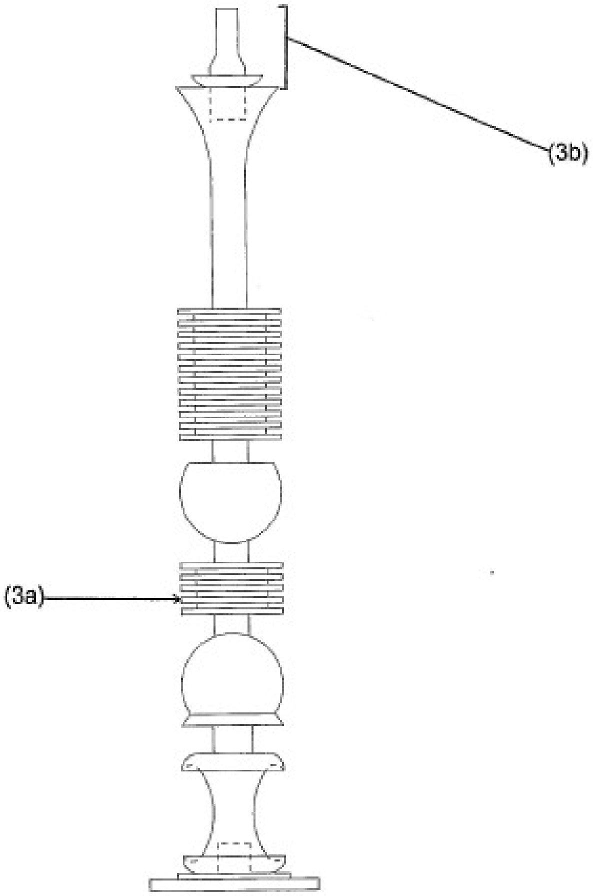
(3b)
(3a)

Fig.(5)

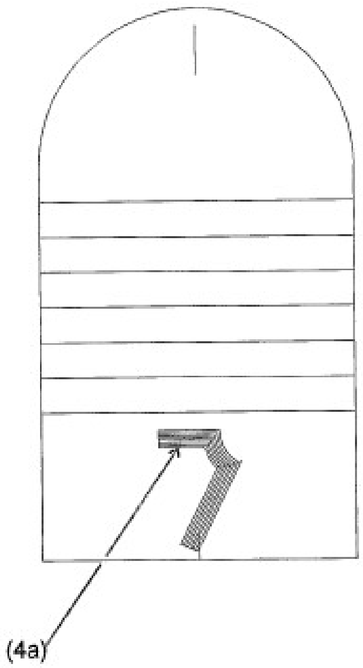

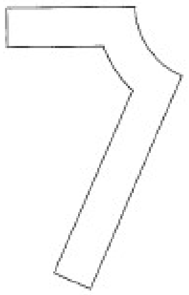

Fig. (6)

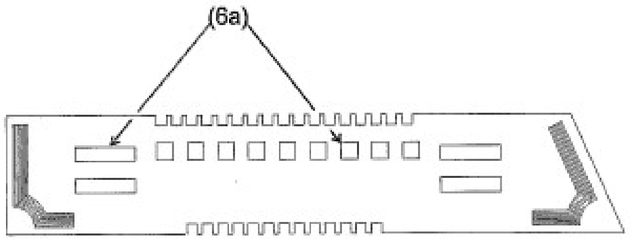

Fig. (8)

Fig. (7)

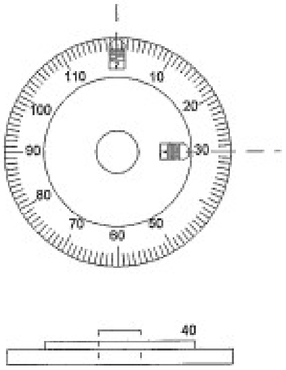
110
10
20
100
30
90
80
70
60
50
40

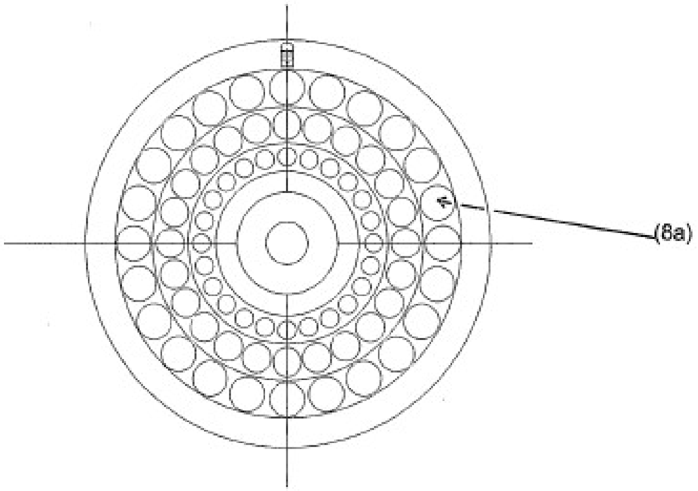

Biosignatures: Linear Diagrams for Biological Energy Balancing

WO2005062690
7-14-2005
KARIM IBRAHIM FAHMY (EG)

Application number: WO2003EG00016 20031231
Priority number(s): WO2003EG00016 20031231
Also published as: AU2003287942 (A1)

Abstract ~ Linear diagrams (Figures 1 to 247) tracing and corresponding to the functional paths of subtle energy flow within the organs of the human body to restore proper flow and energy balancing through resonance. This results in a positive effect on biological functions, either preventative or healing, and protection from environmental energy disturbances. The diagrams are activated by the body's peripheral energy field, which is channeled through the linear diagrams, and enters into resonance with the functions of the organs they correspond to. The energy balancing effect and restoration of proper energy flow within the organs can be introduced by the diagrams either by proximity to the human body, or by superimposing on a radiating energy source that reaches the biological system. These linear diagrams have been termed and copyrighted "BioSignatures" by the inventor.

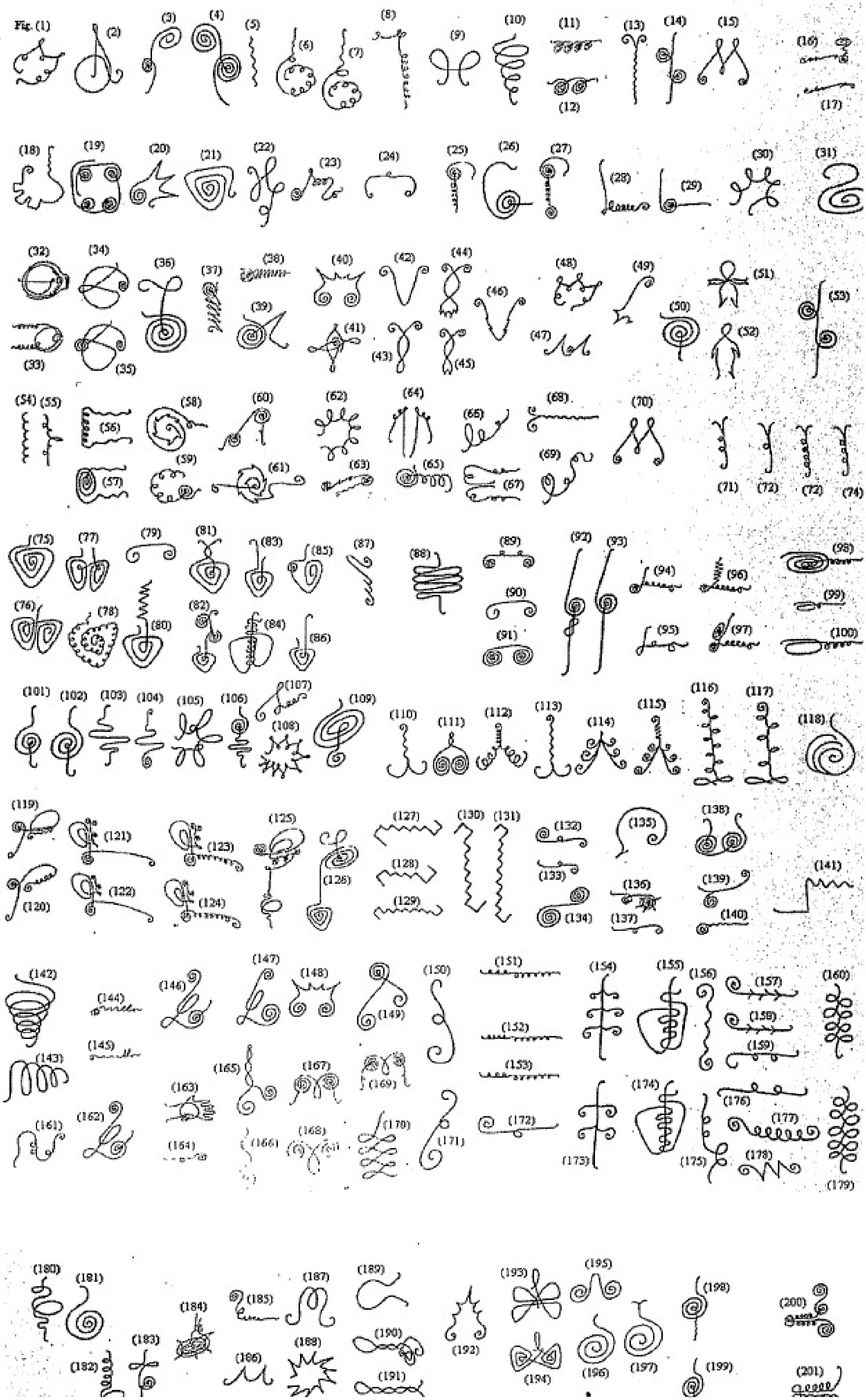
Fig. (1)
(2)
(3)
(4)
(5)
(6)
(7)
(8)
(9)
(10)
(11)
(12)
(13)
(14)
(15)
(16)
(17)
(18)
(19)
(20)
(21)
(22)
(23)
(24)
(25)
(26)
(27)
(28)
(29)
(30)
(31)
(32)
(33)
(34)
(35)
(36)
(37)
(38)
(39)
(40)
(41)
(42)
(43)
(44)
(45)
(46)
(47)
(48)
(49)
(50)
(51)
(52)
(53)
(54)
(55)
(56)
(57)
(58)
(59)
(60)
(61)
(62)
(63)
(64)
(65)
(66)
(67)
(68)
(69)
(70)
(71)
(72)
(72)
(74)
(75)
(76)
(77)
(78)
(79)
(80)
(81)
(82)
(83)
(84)
(85)
(86)
(87)
(88)
(89)
(90)
(91)
(92)
(93)
(94)
(95)
(96)
(97)
(98)
(99)
(100)
(101)
(102)
(103)
(104)
(105)
(106)
(107)
(108)
(109)
(110)
(111)
(112)
(113)
(114)
(115)
(116)
(117)
(118)
(119)
(120)
(121)
(122)
(123)
(124)
(125)
(126)
(127)
(128)
(129)
(130)
(131)
(132)
(133)
(134)
(135)
(136)
(137)
(138)
(139)
(140)
(141)
(142)
(143)
(144)
(145)
(146)
(147)
(148)
(149)
(150)
(151)
(152)
(153)
(154)
(155)
(156)
(157)
(158)
(159)
(160)
(161)
(162)
(163)
(164)
(165)
(166)
(167)
(168)
(169)
(170)
(171)
(172)
(173)
(174)
(175)
(176)
(177)
(178)
(179)
(180)
(181)
(182)
(183)
(184)
(185)
(186)
(187)
(188)
(189)
(190)
(191)
(192)
(193)
(194)
(195)
(196)
(197)
(198)
(199)
(200)
(201)

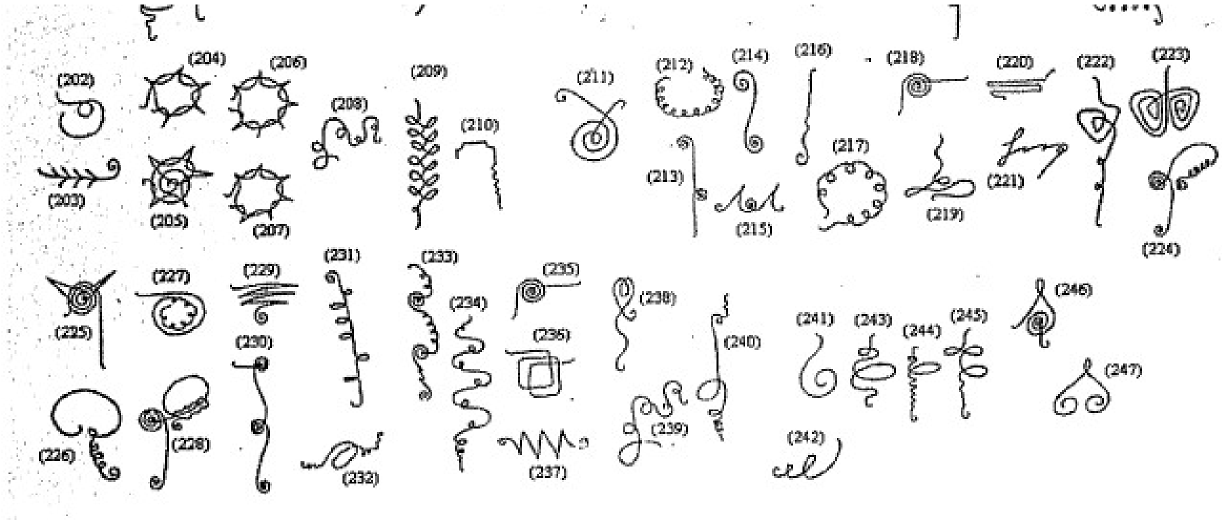
(202)
(203)
(204)
(205)
(206)
(207)
(208)
(209)
(210)
(211)
(212)
(213)
(214)
(215)
(216)
(217)
(218)
(219)
(220)
(221)
(222)
(223)
(224)
(225)
(226)
(227)
(228)
(229)
(230)
(231)
(232)
(233)
(234)
(235)
(236)
(237)
(238)
(239)
(240)
(241)
(242)
(243)
(244)
(245)
(246)
(247)

Bio-Geometrical Shapes

EG22820
9-30-2003
KARIM IBRAHIM FAHMY

Classification: - international: A61B1/00; A61B1/00; (IPC1-7): A61B1/00
Application number: EG19990001427 19991110
Priority number(s): EG19990001427 19991110

Bio-Geometrical Shapes

EG22227
11-30-2002
IBRAHIM FAHMY KARIM

Classification: - international: A61B1/00; A61B1/00; (IPC1-7): A61B1/00; - european:
Application number: EG19990000645 19990602
Priority number(s): EG19990000645 19990602

---

http://biogeometry.com

Products are available in the following countries:

Nozom El Taqa El Handasseya Agent of BioGeometry® Energy Systems Ltd.IBC-BVI
97 Road 9 Maadi, Cairo - Egypt
Tel: (+2)02 3786152
Fax: (+2)02 3590505
Email: (Egypt & Arab World): BioGeometrycentera@hotmail.com
Email: (US & Europe): BioGeometryeg@hotmail.com

Dr. Robert J.Gilbert
Vesica Science and Spirit Resources.
PO Box: 27
Asheville, NC 28802
Tel: 828-257-2600 // 828-296-8324
E-mail: info@vesica.org
www.vesica.org

Michael Maley, PhD, LP and Tricia Mageli-Maley, MA, LP
Creating Healthy Spaces,
Minnetonka, MN
Tel: (952) 542 9480
E-mail: maleys@creatinghealthyspaces.com
www.creatinghealthyspaces.com

Feelgood's Apotheke Witikon
Witikonerstrasse 295
PO Box: 778
CH: 8053 Zurich ,
Switzerland
E-mail: apo@feelgoods.ch
www.feelgoods.ch

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

Set 7. DX receiver with two tuned circuits

Back to the index

This receiver has been built in three parts, each part will be described on a separate page:

1- The antenna tuner unit
2- The detector unit
3- The transformer unit

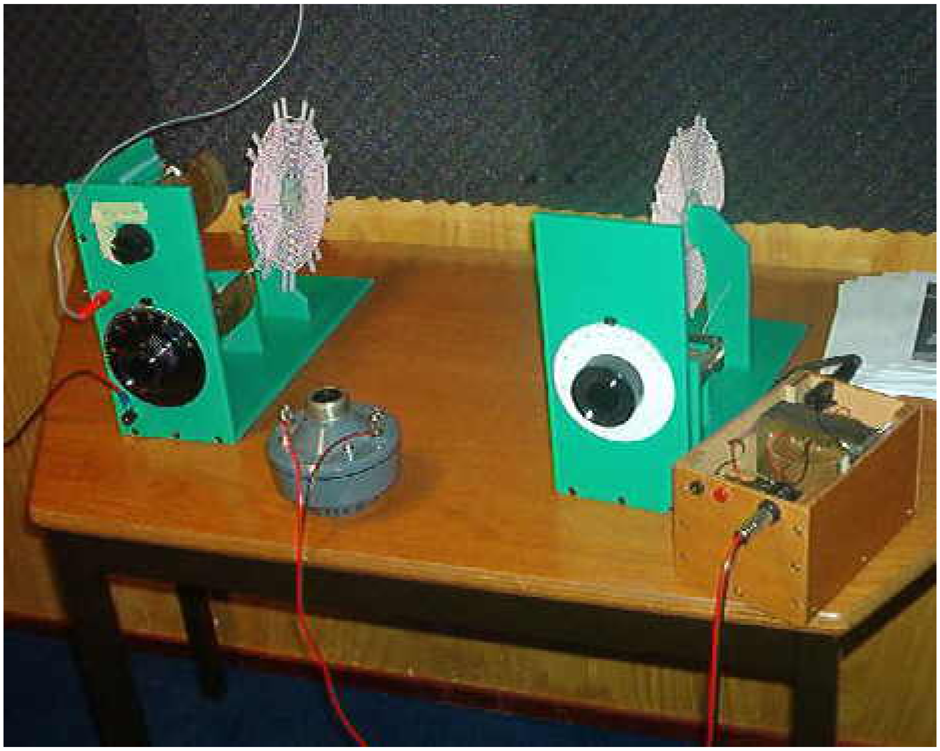

In this receiver I tried to obtain the best possible selectivity and sensitivity, by using a proper circuit and good quality materials.

Before I started building this receiver, I performed many experiments in order to obtain the best possible Q factor in the tuned circuits.
This resulted in circuits with a (unloaded) Q above 1000 for the most part of the mediumwave band.

Experiments en measurements carried out with the antenna tuner unit are described here.

The loudspeaker used is a driver unit from a hornloudspeaker, for more information on the sensitivity of this driver unit, see the loudspeakertest.

For a good performance, the coils of the antenna tuner unit and the detector unit must be placed at a distance of 42 cm.
More information and measurements on this topic, you will find here.

The sensitivity of the receiver.

To measure the sensitivity of the receiver, the receiver was connected to a signal generator.
The signal generator was AM modulated with 1000 Hz, this 1000 Hz was a square wave with a modulation depth of 100 %.

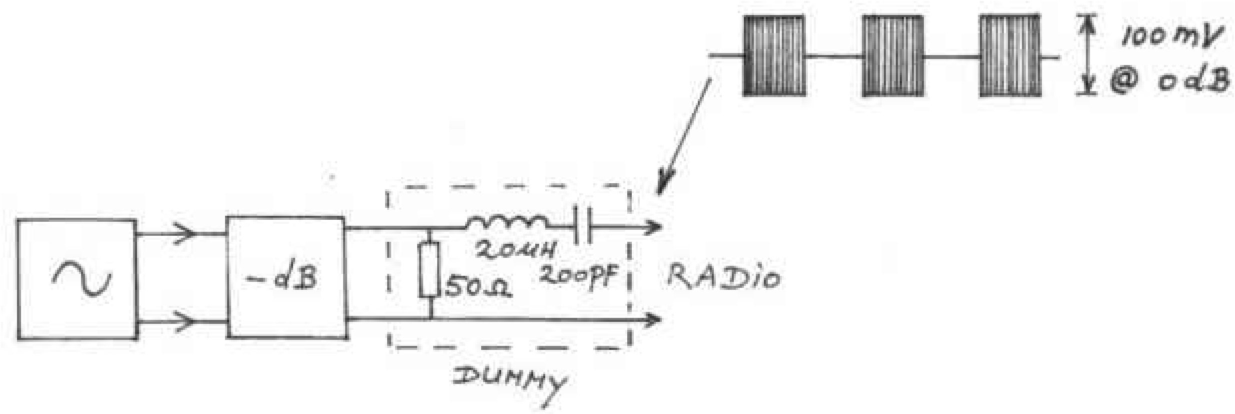

Between signal generator and receiver, I connected a variable attenuator and a dummy antenna.

The output voltage of the dummy antenna was 100 mVpp with the attenuator at 0 dB.
This 100 mVpp was measured without connecting the receiver.
When we connect the receiver to the dummy antenna, the output voltage of the dummy antenna will increase, this happens because the coil and capacitor in the dummy antenna then form a part of the resonant antenna unit circuit.

Attenuator setting	Voltage out of dummy antenna (unloaded)	Audio
0 dB	100 mV pp	1600 mV pp
-10 dB	32 mV pp	540 mV pp
-20 dB	10 mV pp	120 mV pp
-30 dB	3.2 mV pp	10 mV pp
-40 dB	1.0 mV pp	< 1 mV pp
-43 dB	0.7 mV pp	sound just audible.

In the last column the amplitude of the detected audiosignal is given, the audio signal is loaded with the transformer unit with speaker.
With the attenuator at -43dB the 1000 Hz test tone was just audible.
The (unloaded) antenna voltage is then 0.7mVpp or 0.25mVeff.
For listening to a normal audio signal, the antenna signal must however be stronger.

With this receiver I took part in national and international receiving contests.

In 2005 this receiver was the 1st price winner in the "Elmer Memorial Crystal Radio DX Contest".

For loglists of the received stations, click here.

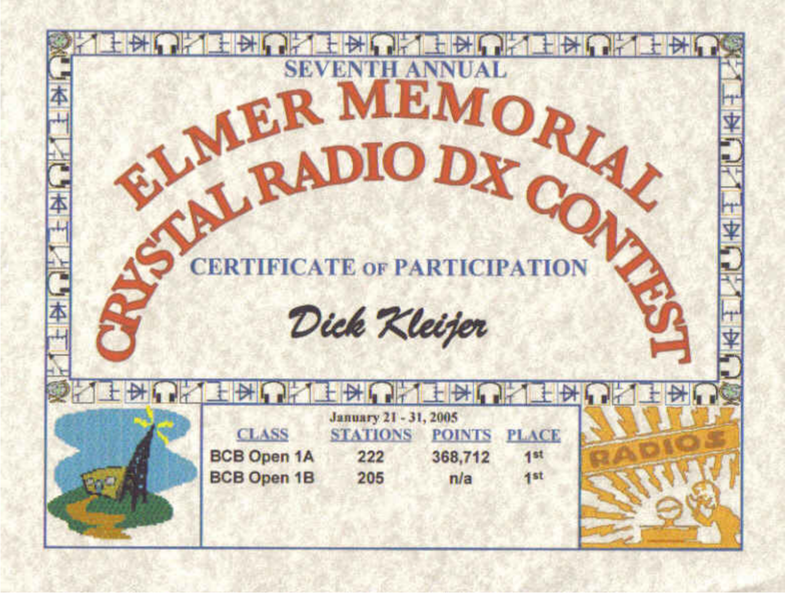

SEVENTH ANNUAL

ELMER MEMORIAL

CRYSTAL RADIO DX CONTEST

CERTIFICATE OF PARTICIPATION

Dick Kleijer

January 21 - 31, 2005

CLASS	STATIONS	POINTS	PLACE
BCB Open 1A	222	368,712	1st
BCB Open 1B	205	n/a	1st

Back to the index

(19) **United States**
(12) **Patent Application Publication**
Dinwiddie et al.
(10) **Pub. No.: US 2017/0117714 A1**
(43) **Pub. Date: Apr. 27, 2017**

(54) **POWER RECEIVER FOR EXTRACTING POWER FROM ELECTRIC FIELD ENERGY IN THE EARTH**

(71) Applicant: **Earth Energies, Inc.**, Johns Creek, GA (US)

(72) Inventors: **John Dinwiddie**, Cary, NC (US); **Terry L. Wright**, Suwanee, GA (US); **David R. Ames**, Johns Creek, GA (US)

(21) Appl. No.: **15/397,281**

(22) Filed: **Jan. 3, 2017**

Related U.S. Application Data

(63) Continuation-in-part of application No. 14/509,772, filed on Oct. 8, 2014, now Pat. No. 9,564,268.

(60) Provisional application No. 61/889,894, filed on Oct. 11, 2013.

Publication Classification

(51) **Int. Cl.**

H02J 4/00	(2006.01)
H02J 7/35	(2006.01)
H02J 3/38	(2006.01)

(52) **U.S. Cl.**
CPC ***H02J 4/00*** (2013.01); ***H02J 3/385*** (2013.01); ***H02J 7/35*** (2013.01)

(57) **ABSTRACT**

A resonant transformer connected between a ground terminal and elevated terminal draws current from the earth's electric field through a primary winding of the transformer. An impulse generator applies a high voltage impulse to the primary winding of the resonant transformer to cause current to flow from the ground terminal through the primary winding. The flow of current through the primary winding of the resonant transformer induces a current in the secondary winding, which may be converted and filtered to a usable form, e.g. 60 Hz AC or DC.

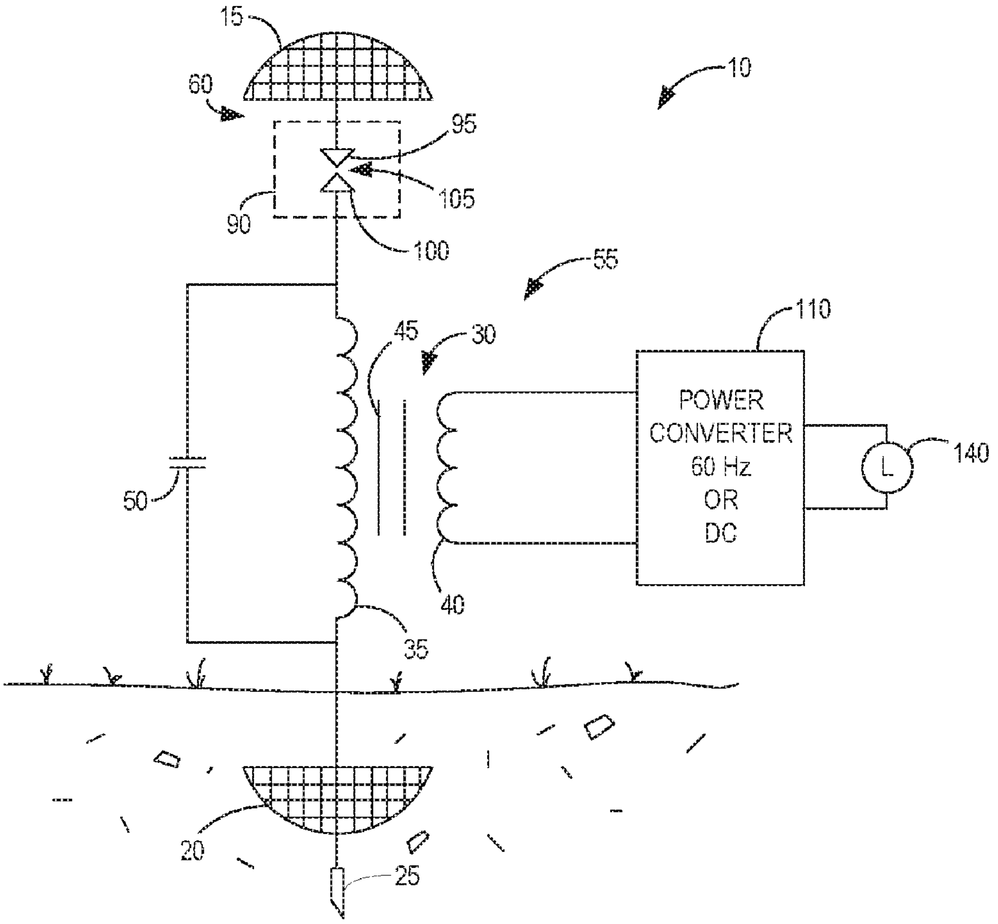

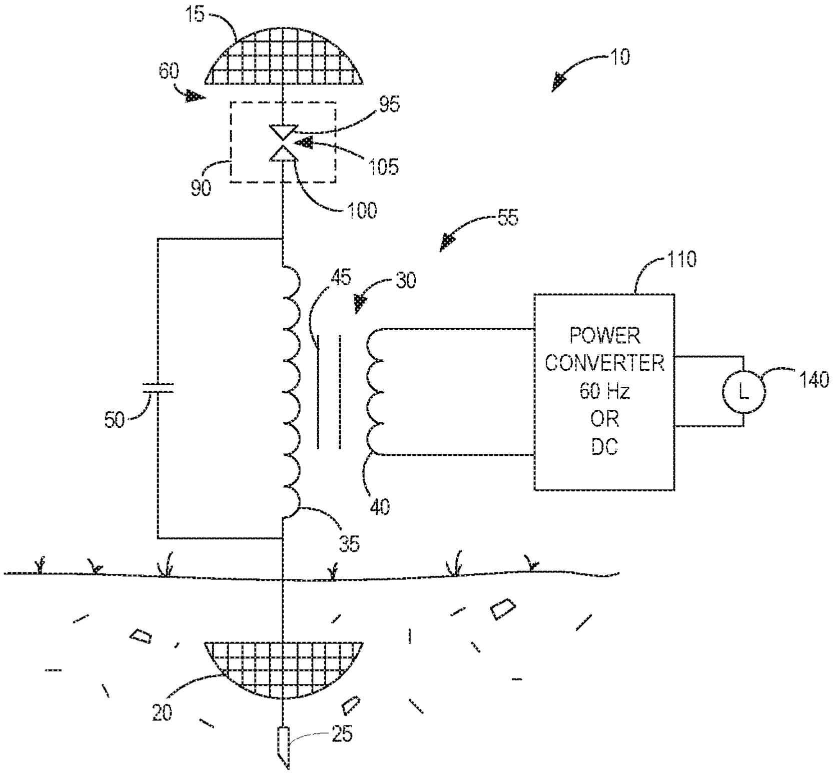
POWER
CONVERTER
60 Hz
OR
DC
L

FIG 1

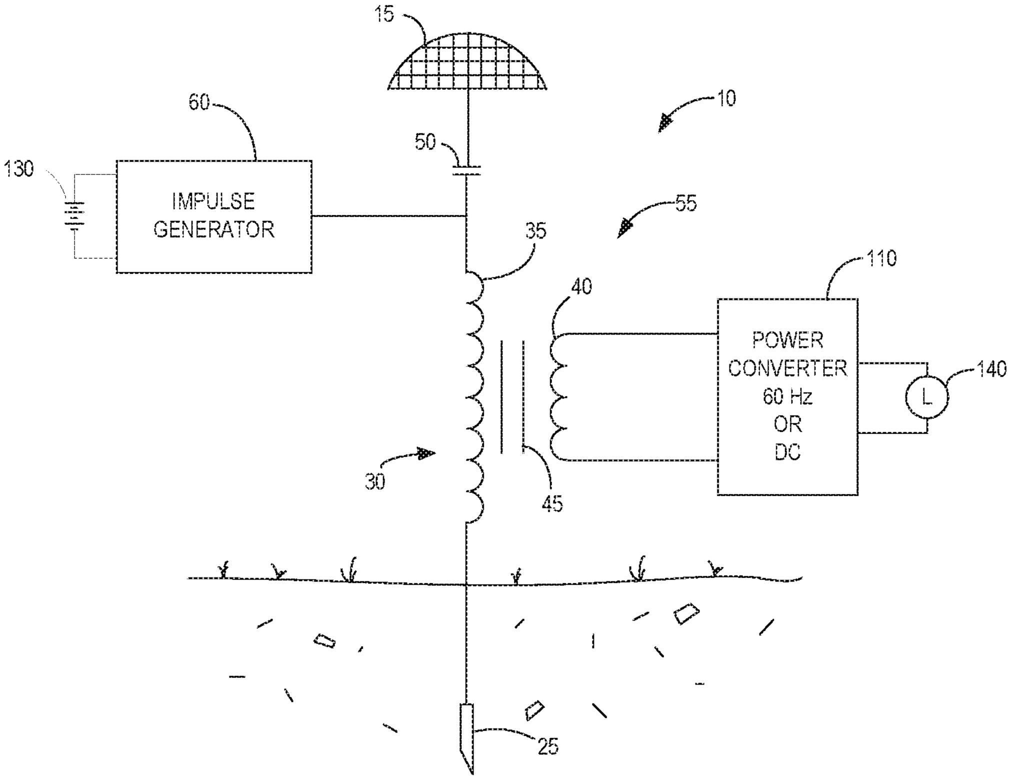
IMPULSE GENERATOR
POWER CONVERTER 60 Hz OR DC
L

FIG 2

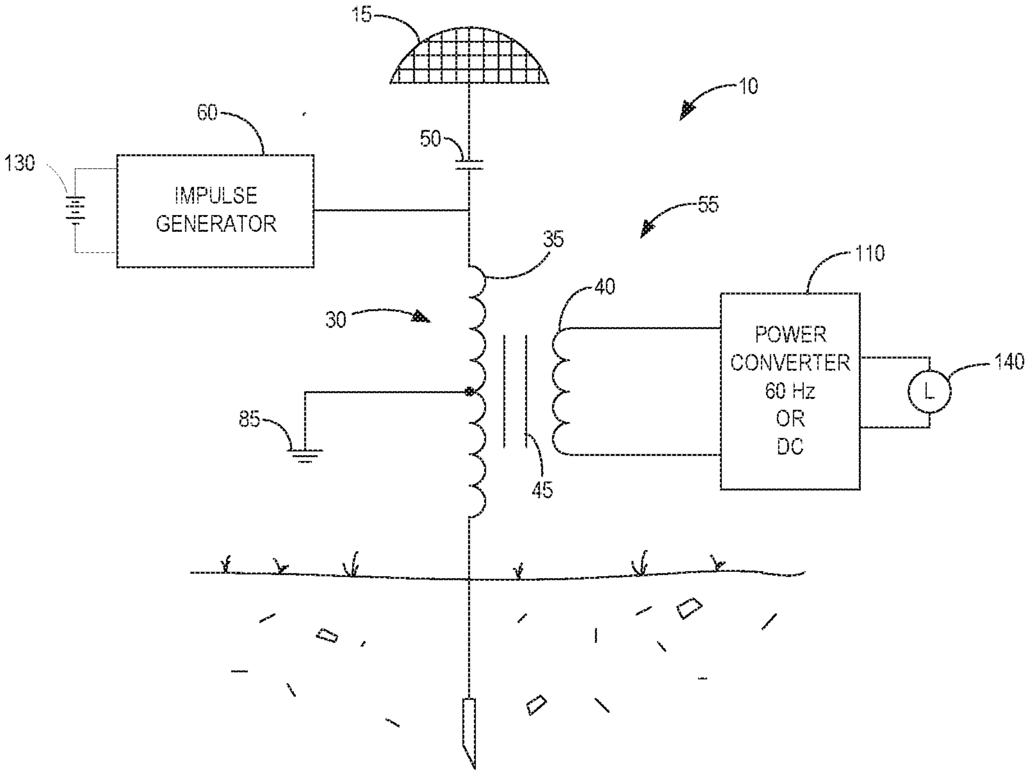
IMPULSE GENERATOR
POWER CONVERTER 60 Hz OR DC
L

FIG 3

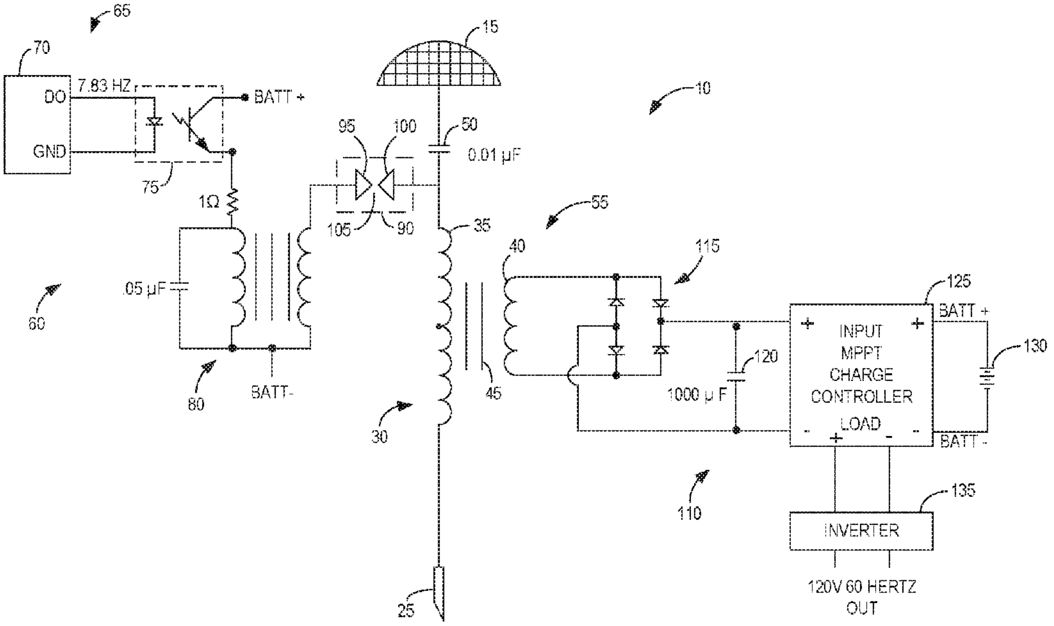
65
70
DO
7.83 HZ
GND
75
BATT +
1Ω
.05 μF
60
80
BATT-
15
50
0.01 μF
10
95
100
105
90
35
55
40
30
45
25
115
120
1000 μ F
125
INPUT
MPPT
CHARGE
CONTROLLER
LOAD
+
-
BATT +
BATT -
130
110
135
INVERTER
120V 60 HERTZ
OUT

FIG 4

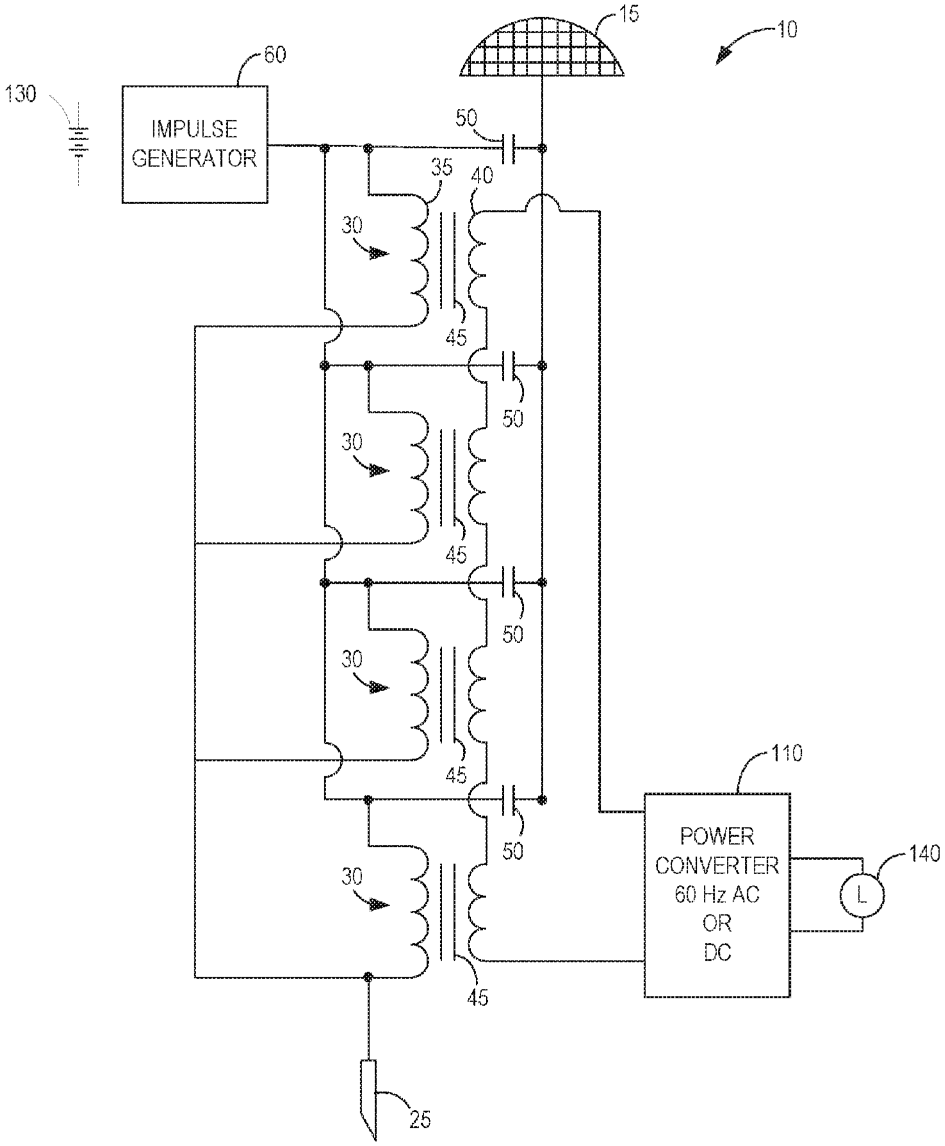
IMPULSE GENERATOR
POWER CONVERTER 60 Hz AC OR DC
L

FIG 5

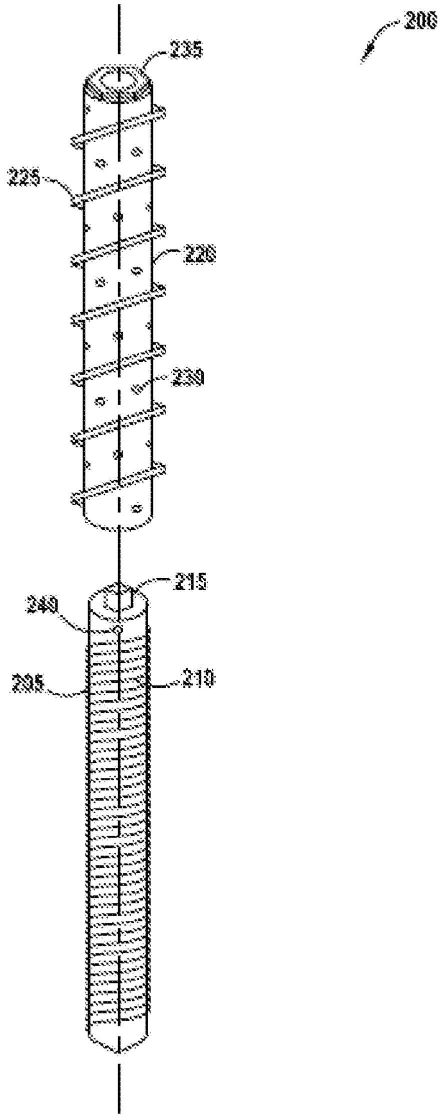
200
235
225
220
230
215
240
205
210

FIG 6A

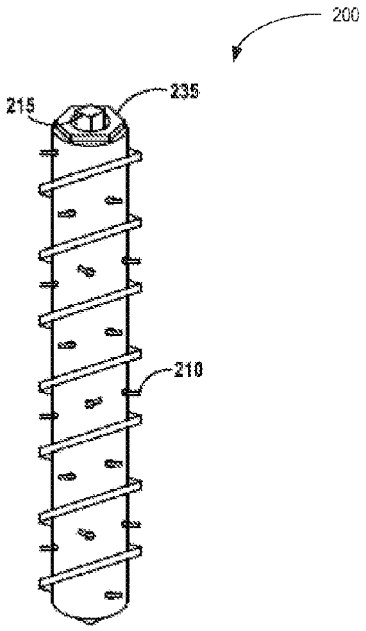
200
215
235
210

FIG 6B

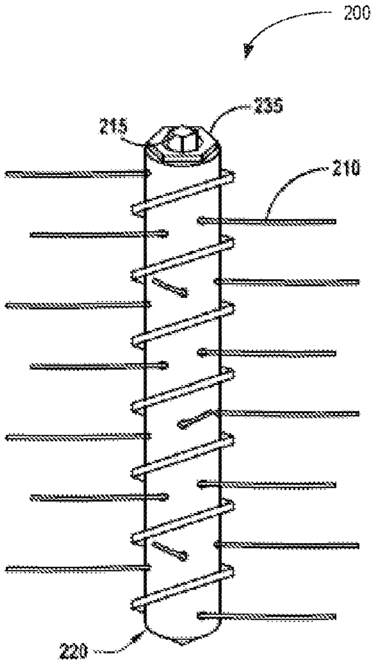
200
215
235
210
220

FIG 6C

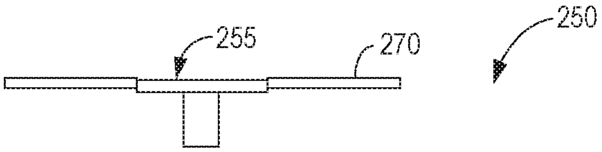

FIG 7A

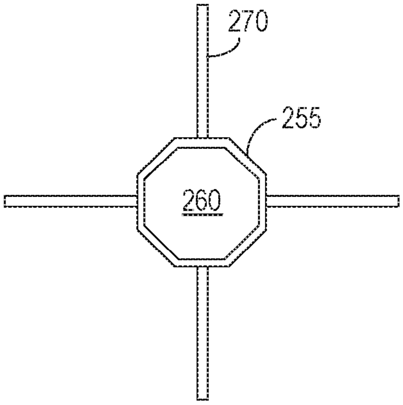

FIG 7B

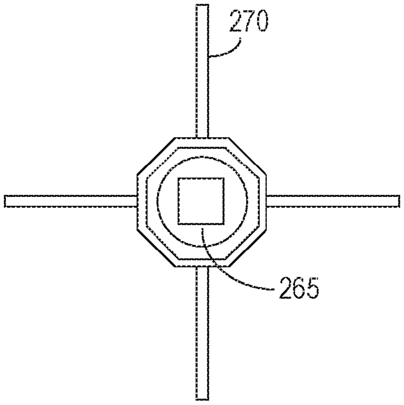

FIG 7C

POWER RECEIVER FOR EXTRACTING POWER FROM ELECTRIC FIELD ENERGY IN THE EARTH

[0001] This application is a continuation-in-part of prior U.S. application Ser. No. 14/509772 filed 8 Oct. 2014 which claims the benefit of U.S. Provisional Application No. 61/889894 filed 11 Oct. 2013, the disclosures of all of which are incorporated by reference herein in their entirety.

TECHNICAL FIELD

[0002] The present invention relates generally to renewable energy, and more particularly to methods and apparatus for extracting energy from subsurface electrical fields beneath the earth's surface.

BACKGROUND

[0003] The earth and the ionosphere cavity may be viewed as a global electric circuit. Electrical currents are constantly flowing within the earth and its atmosphere. Within the earth, the majority of the earth's energy is carried by extremely low frequency (ELF) and ultralow frequency (ULF) waves in the 0-200 Hz frequency range. The earth's rotating magnetic field and positive lightning are two energy sources that sustain the ELF/ULF waves within the earth and the atmosphere.

[0004] A great deal of research has been devoted to studying the electric field present in the earth's ionosphere cavity. Joseph M. Crawley, the "Fair Weather Atmosphere as a Power Source", Proceedings ESA Annual Meeting on Electrostatics 2011; O. Jefimenko, "Operation of Electric Motors from Atmospheric Electric Field," American Journal of Physics, Vol. 39, Pgs. 776-779, 1971; M. L. Breuer, "Usability of Tapping Atmospheric Charge as a Power Source," Renewable Energy, Vol. 28, Pgs. 1121-1127, 2003. Numerous attempts have been made in the past to extract electrical energy from the earth's atmosphere. For example, U.S. Pat. No. 1,540,998 to Plauson describes a system for converting atmospheric electrical energy into usable power. These past attempts have been successful in producing only small amounts of power from the electrical field in the earth's ionosphere cavity. The modest success of these experiments compared to results from other renewable energy sources, such as solar and wind, has tempered further research and prevented widespread use of the electric field in the ionosphere cavity as an energy source.

SUMMARY

[0005] The present invention relates to a power receiver for extracting power from electric fields beneath the earth's surface. In embodiments of the present disclosure, a resonant transformer connected to a ground terminal draws current from the earth's electric field through the primary winding of the transformer. Current flow through the resonant transformer is induced by applying a high voltage impulse to the primary winding. The flow of current through the primary winding of the resonant transformer induces a current in the secondary winding, which may be converted and filtered to a usable form, e.g. 60 Hz AC or DC.

[0006] In some embodiments of the power receiver, the resonant frequency of the resonant transformers is below 200 Hz.

[0007] In some embodiments of the power receiver, the resonant transformer comprises a ferro-resonant transformer.

[0008] In some embodiments, the power receiver further comprises an elevated terminal.

[0009] In some embodiments of the power receiver, the primary winding of the resonant transformer is connected between the ground terminal and elevated terminal.

[0010] In some embodiments of the power receiver, the elevated terminal comprises an upper capacitive plate coupled to the earth's ionosphere cavity.

[0011] In some embodiments of the power receiver, the impulse generator comprises the upper capacitive plate and a spark gap connected between the upper capacitive plate and the primary winding of the resonant transformer. The spark gap comprises a pair of electrodes separated by a gap and configured to generate a spark when a voltage difference between the electrodes reaches a predetermined level.

[0012] In some embodiments of the power receiver, the impulse generator comprises a pulse generator for generating low voltage pulses, a step-up transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses, and a spark gap connected between the step-up transformer and the primary winding of the resonant transformer to generate a spark responsive to the high voltage impulses from the step-up transformer.

[0013] In some embodiments of the power receiver, the impulse generator comprises a pulse generator for generating low voltage pulses, and a step-up transformer connected to the primary winding of the resonant transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses.

[0014] In some embodiments of the power receiver, the impulse generator comprises a solid state spark generator.

[0015] In some embodiments of the power receiver, the resonant transformer includes a capacitor connected in parallel with the primary winding.

[0016] In some embodiments of the power receiver, the resonant transformer includes a capacitor connected in series with the primary winding between the impulse generator and the elevated terminal.

[0017] In some embodiments, the power receiver comprises multiple resonant transformers having primary windings connected in parallel between the ground terminal and the elevated terminal.

[0018] In some embodiments of the power receiver, the resonant transformers have different resonant frequencies.

[0019] In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are all below 200 Hz.

[0020] In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

[0021] Another embodiment of the power receiver comprises a resonant circuit connected to a ground terminal disposed below the surface of the earth, an impulse generator for generating and applying a high voltage electrical impulse to the resonant circuit to induce current flow from the ground terminal through the resonant circuit, and a power conversion circuit connected to the resonant circuit to convert electrical current flowing through the resonant circuit to a desired form. The resonant circuit has a resonant frequency below 200 Hertz.

[0022] In some embodiments of the power receiver, the resonant circuit comprises a resonant transformer having a primary winding, a secondary winding, and resonant capacitor connected in series with the primary winding.

[0023] In some embodiments of the power receiver, the resonant circuit comprises multiple resonant transformers having primary windings connected in parallel to the ground terminal.

[0024] In some embodiments of the power receiver, the resonant transformers have different resonant frequencies.

[0025] In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are all below 200 Hz.

[0026] In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

[0027] Other embodiments of the disclosure comprise a ground terminal for a power receiver. In one embodiment, the ground terminal comprises a ground shaft configured for insertion beneath the surface of the earth, a hollow cylinder surrounding the ground shaft and having a plurality of openings, and a plurality of ground wires connected at one end to the ground shaft. The ground wires are wound around the ground shaft and have free ends protruding through respective openings in the hollow shaft so that rotation of the ground shaft relative to the hollow cylinder causes the ground wires to extend radially into the earth.

[0028] Other embodiments of the disclosure comprise methods of extracting power from the earth. In one embodiment, the method comprises applying a high voltage impulse to resonant circuit coupled to a ground terminal disposed beneath the surface of the earth to initiate resonance in the resonant circuit and induce the flow of current from the ground terminal to the resonant circuit, and converting the current flowing from the ground terminal into the resonant circuit into a useful form.

[0029] In some embodiments of the method, the resonant circuit comprises a resonant transformer including a primary winding coupled to the ground terminal and a second winding coupled to a power converter, and applying a high voltage impulse to resonant circuit comprises applying a high voltage impulse to the primary winding of the resonant transformer.

[0030] In some embodiments of the method, applying a high voltage impulse to the primary winding of the resonant transformer comprises applying an impulse in the range to 10,000 to 40,000 volts to primary winding of the transformer.

BRIEF DESCRIPTION OF THE DRAWINGS

[0031] FIG. **1** illustrates a first exemplary embodiment of a power receiver.

[0032] FIG. **2** illustrates a second exemplary embodiment of a power receiver.

[0033] FIG. **3** illustrates a third exemplary embodiment of a power receiver.

[0034] FIG. **4** illustrates a fourth exemplary embodiment of a power receiver.

[0035] FIG. **5** illustrates a fifth exemplary embodiment of a power receiver.

[0036] FIG. **6**A is an exploded perspective view of an exemplary ground antenna array for the power receiver.

[0037] FIG. **6**B is a perspective view of an assembled ground antenna array before being deployed.

[0038] FIG. **6**C is a perspective view of an assembled ground antenna array after being deployed.

[0039] FIG. **7**A is a side view of an insertion tool for installing the ground antenna array.

[0040] FIG. **7**B is a top view of the insertion tool for installing the ground antenna array.

[0041] FIG. **7**C is a bottom view of the insertion tool for installing the ground antenna array.

DETAILED DESCRIPTION

[0042] Referring now to the drawings, a power receiver for extracting energy from the earth's electric field are illustrated and indicated generally by the numeral **10**. Various embodiments of the power receiver **10** are described and similar reference numbers are used throughout the description to indicate similar components.

[0043] The power receiver **10** converts energy in the ELF/ULF waves to useful form, e.g. 60 Hz AC or DC. The power receiver **10** is essentially a resonance circuit that resonates at the natural resonance frequencies in the earth's electric field. These resonance frequencies, known as Schumann resonance frequencies, occur at 7.83 Hz, 14.3 Hz, 20.8 Hz, 27.3 Hz, and 33.8 Hz. A high voltage impulse initiates resonance within the power receiver **10**. In the resonant mode, the impedance of the power receiver **10** is reduced to near zero thus inducing ground currents to flow into the power receiver **10** where the ground currents are converted to useful form.

[0044] FIG. **1** illustrates a first embodiment of the power receiver **10**. The power receiver **10** comprises a resonant transformer **30** connected between an elevated terminal **15** and ground terminal **25**. In this embodiment, the elevated terminal **15** is capacitively coupled to electric fields within the earth's ionosphere cavity and functions as an upper capacitive plate. A lower capacitive plate **20** is connected to the ground terminal **25** beneath the surface of the earth.

[0045] The resonant transformer **30** comprises a primary winding **35**, secondary winding **40**, ferromagnetic core **45**, and capacitor **50**. One end of the primary winding **35** is connected to the lower capacitive plate **20** and ground terminal **25**. The opposite end of the primary winding **35** is connected via a spark gap **90** to the elevated terminal **15**. The capacitor **50** is connected in parallel with the primary winding **35** of the resonant transformer **30** to form an LC circuit **55** with a resonance frequency range of between about 0.1 and 200 Hz. In a preferred embodiment, the resonant transformer has a Q of about 10 or greater and resonance frequency in the range of about 0.1-200 Hertz. For example, the resonant transformer **30** may have a resonance frequency of about 7.83 Hz, the fundamental Schumann resonance frequency. The secondary winding **40** of the resonant transformer **30** is connected to a power converter **110** as will be hereinafter described in greater detail. The power converter **110** converts the energy extracted from the earth's electric field by the power receiver **10** into a usable form for driving a load **140**.

[0046] The elevated terminal/upper capacitive plate **15** comprises an insulated, dish-shaped plate with a large radius of curvature. The capacitance and resistance of the elevated terminal is chosen for receiving broadband electric field frequencies in the 0-200 Hz range. The upper capacitive plate **15** is sized to maximize to the extent practical coupling with the electric field in the earth's ionosphere cavity.

[0047] The lower capacitive plate **20** is also a dish-shaped plate with a large radius of curvature. One function of the lower capacitive plate **20** is to collect charge from the earth's ground currents and provide an instantaneous source of current as hereinafter described. The capacitance and resistance of the lower capacitive plate **20** is selected to promote the flow of current from the ground with minimal losses.

[0048] The spark gap **90** connected between the elevated terminal **15** and resonant transformer **30** comprises a pair of electrodes **95**, **100** separated by an evacuated air gap **105**. Electrode **95** is connected to the upper capacitive plate **15**. Electrode **100** is connected to the resonant transformer **30**. The spark gap **105** prevents electrical discharge from the upper capacitive plate **15** to the earth's ionosphere cavity. The spark gap **90** in combination with the elevated terminal **15** function as an impulse generator that applies a high voltage impulse of about 10,000-40,000 volts to the primary winding **35** to initiate resonance in the transformer **30**.

[0049] In operation, the capacitive coupling of the upper capacitive plate **15** induces a high voltage operating current in the upper capacitive plate **15**. The upper capacitive plate is connected to a first electrode **95** to the spark gap **90**. When the voltage difference between the electrodes **95** and **100** reaches a threshold, a spark forms across the electrodes **95**, **100** and a high voltage impulse is applied to the primary winding **35** of the resonant transformer **30**. This high voltage impulse initiates resonance within the transformer **30**.

[0050] In resonant mode, the impedance of the resonance transformer is reduced to nearly zero allowing current to flow from the capacitive plate **20** and ground terminal **25** through the primary winding **35** of the transformer **30**, which in turn induces current in the secondary winding **40**. Power converter **110** converts the current flowing through the secondary winding **40** into a usable form for driving a load **140**. The transformer **30** will continue to resonate for a short period of time. By providing high voltage impulses to the primary winding **35** of the resonant transformer **30** at periodic intervals, it is possible to maintain a continuous flow of current from the earth into the resonant transformer **30**, thus producing a continuous supply of power.

[0051] FIG. **2** discloses a second embodiment of the primary receiver **10**. This embodiment includes a resonant transformer **30** connected between an elevated terminal **15** and ground terminal **25**. The resonant transformer **30** comprises a primary winding **35**, secondary winding **40**, ferromagnetic core **45** and a high voltage capacitor **50**. One end of the primary winding **35** is connected to the ground terminal **25**. The opposite end of the primary winding **35** is connected to the elevated terminal **15**. The capacitor **50** has a capacitance of about 0.01 micro-farads. In contrast to the previous embodiment, capacitor **50** is connected in series with the primary winding **35** and elevated terminal **15** and forms a LC circuit **55** with a Q of about 10 or greater and a resonance frequency in the range of about 0.1 to 200 Hz. In a preferred embodiment, the resonance frequency of the transformer **30** is 7.83 Hz, the fundamental Schumann resonance frequency. An impulse generator **60** is connected between the primary winding **35** of the resonant transformer **30** and the series capacitor **50** and applies a high voltage impulse to the primary winding **35** of the resonant transformer **30**. A battery **130** or other external power source supplies power to the impulse generator **60**. As previously described, the high voltage impulse applied by the impulse generator **60** initiates resonance within the resonant transformer **30** inducing current flow from the ground terminal **25** into the primary winding **35** of the resonant transformer **30**. The flow of current from the ground terminal **25** through the primary winding **35** induces current in the secondary winding **40**. Power converter **110** converts the electrical energy in the current flowing through the primary winding **40** into a usable form.

[0052] In contrast to the first embodiment, it is not required to capacitively couple the elevated terminal **15** in the second embodiment to the earth's ionosphere cavity. Rather, the elevated terminal **15** in this embodiment provides lightning protection and dissipates some of the energy flowing into the power receiver **10** to the earth's ionosphere cavity. Also, in contrast to the first embodiment, the capacitor **50** is connected in series between the primary winding **35** of the transformer **30** and the elevated terminal **15**. Those skilled in the art will appreciate that the capacitor **50** could also be connected in parallel rather than series with the primary winding **35** as shown in FIG. **1**. Another difference is that the impulse generator **60** has an external power source. The amount of energy generated by the power receiver **10**, however, is far greater than the energy needed to generate high voltage impulses. The first embodiment does not require an external power source to generate high voltage impulses.

[0053] FIG. **3** illustrates a third embodiment of the power receiver **10**. This embodiment is essentially the same as the embodiment shown in FIG. **2**. The main difference is that a center tap of the primary winding **35** in the resonant transformer **30** is connected to an electrical ground **85**. It should be appreciated that the electrical ground **85** may be different than the earth ground. When the center tap of the resonant transformer **30** is grounded at a distance away from the ground terminal **25** (e.g. 50 ft to 100 ft), the power receiver **10** becomes a transmitter via the ground loop formed.

[0054] FIG. **4** illustrates the power receiver **10** of FIG. **2** in greater detail. The power receiver includes a resonant transformer **30** connected between a ground terminal **25** and an elevated terminal **15**. The ground terminal **25** may comprise a ⅝-inch×8-foot copper ground rod, such as the ERICO 615880UPC. The elevated terminal **15** may comprise a 90% copper mesh formed into a hemisphere with a radius of about 9 inches. The elevated terminal **15** may be elevated at a height of approximately 6 feet above the ground.

[0055] The resonant transformer **30** includes a primary winding **35**, secondary winding **40**, ferromagnetic core **45** and series capacitor **50** configured as previously described. The resonant transformer **30** may have a Q of about 10 and a resonance frequency in the range of about 0.1 to 200 Hz. The resonant transformer **30** may be made using an Allanson transformer (part #1530BP120R) connected in series with a 0.01 micro-farad capacitor, such as the Condensor Products high voltage capacitor (part #TC 103-17-125). The resonant transformer **30** is used in a step-down configuration. The center tap of the resonant transformer **30** may optionally be connected to a ground.

[0056] An impulse generator **60** is connected between the primary winding **35** of the resonant transformer **30** and the series capacitor **50** and applies a high voltage impulse in the range of about 10,000 to 40,000 volts to the primary winding of the transformer **30**. A battery **130** or other external power source supplies power to the impulse generator **60**. The power converter **110** connects to the secondary winding **40**

of the resonant transformer **30** for converting current in the secondary winding of the transformer to a useful form.

[0057] The impulse generator **60** comprises a pulse generator **65** for generating low voltage pulses, a step-up transformer **80** for converting the low voltage pulses from the pulse generator **65** to high voltage pulses, and a spark gap **90** for generating sparks responsive to the high voltage pulses from the step-up transformer **80**.

[0058] The pulse generator **65** comprises a square wave generator **70**, such as a Sinometer VC2002 function signal generator, and solid state relay **75**. The square wave generator **70** generates a digital pulse stream. In one embodiment, the digital pulse stream generates a square waveform with a frequency of about 7.83 Hz. The frequency of the digital pulse stream is selected to match the resonance frequency of the transformer **30**, though such is not necessarily required. The pulse stream output from the square wave generator **70** is applied to the solid state relay **75**. The solid state relay **75** is connected between a battery or other power source and a first winding of the step-up transformer **80**. The battery may comprise a 12 V, 7.0 A/H sealed lead acid battery, such as the ELB 1270A by Lithonia Lighting. The solid state relay **75** functions as a switch that is activated responsive to the waveform from the square wave generator **70** to provide a continuous stream of low voltage pulses from the battery to the first winding of the step-up transformer **80**. A 1 ohm resistor is connected between the solid state relay **75** and step-up transformer **80**.

[0059] The step-up transformer **80** may comprise a Transco 15 kV, 30 mA neon sign transformer (part #S15612). The step-up transformer **80** converts the low voltage pulses from the pulse generator **65** to high voltage pulses that are applied to the spark gap **90**. The step-up transformer has a 0.5 micro-farad capacitor connected in parallel with the primary winding of the step-up transformer **80**. The step-up transformer produces pulses at the output of about 30,000 to 40,000 volts.

[0060] The spark gap **90** comprises a pair of electrodes **95**, **100** separated by an air gap **105**. A suitable spark gap electrode pair is the Information Unlimited SPARK05 ¼-inch×1-inch tungsten electrodes. As previously described, when the voltage potential between the electrodes **95**, **100** reaches a threshold, a spark forms between the electrodes **95**, **100** and supplies a nearly instantaneous, high voltage impulse to the primary winding **35** of the resonant transformer **30**. This high voltage impulse initiates resonance in the resonant transformer **30** inducing current flow from the ground terminal **25** through the primary winding **35** of the resonant transformer **30**.

[0061] The power converter **110** comprises a bridge rectifier **115**, filter capacitor **120**, charge controller **125**, and inverter **135**. A suitable rectifier is the Micro Commercial Components 10 amp, 1000 volt bridge rectifier (Part #GBJL 1010). The bridge rectifier **115** converts the AC current flowing through the secondary winding **40** of the resonant transformer to a DC current. A filter capacitor **120** removes unwanted frequencies from the DC current. A suitable capacitor **120** is Cornell Dubilier 1000uF 450VDC capacitor (part #383LX102M450N082). The filter capacitor **120** has a capacitance of about 1000 micro-farads. The DC current is input to the charge controller **125**. The charge controller **125** may, for example, comprise a maximum power point tracking (MPPT) charge controller, such as a Tracer 4215 BN MPPT Solar Charge Controller, which is commonly used in solar power generating systems. The charge controller **125** applies a small amount of energy to a battery **30** to charge the battery **130**. As previously noted, the battery **130** serves as a power source for the impulse generator **60**. The remaining current is supplied to an inverter **135**, which converts the DC current to an AC current with a desired voltage and frequency, e.g., 120 volts/60 Hz AC. A suitable inverter **135** is the 1500 W Pure Sine power inverter (AIMS) (part #PWRI1500125).The power converter **110** as shown in FIG. **4** may be utilized in the embodiment shown in FIGS. **1**, **2** and **3**.

[0062] FIG. **5** illustrates a power receiver **10** according to another embodiment. The power receiver **10** comprises a plurality of resonant transformers **30** connected between a ground terminal **25** and elevated terminal **15**. Each of the resonant transformers **30** comprises a primary winding **35**, secondary winding **40**, ferromagnetic core **45** and series capacitor **50**. The primary windings **35** of the resonant transformers **30** are connected in parallel. The secondary windings **40** are connected in series with the power converter **110**. An impulse generator **60** applies a high voltage impulse to the primary windings **35** of the resonant transformers **30**. A battery **130** or other external power source supplies power to the impulse generator **60**. The power converter **110** converts the current in the power converter circuit to a usable form for driving a load **140**.

[0063] In one embodiment, each of the resonant transformers **30** shown in FIG. **5** is configured to have a different resonant frequency. In one embodiment, the resonant transformers **30** are configured to resonate at frequencies of 7.83 Hz, 14.8 Hz, 20.3 Hz and 26.8 Hz respectively. Additional resonant transformers **30** could be added to operate at other resonance frequencies.

[0064] FIGS. **6A-6C** illustrate a high quality ground antenna array **200** which may be used as a ground terminal **25**. The ground antenna array **200** comprises a generally cylindrical ground shaft **205** disposed with a hollow cylinder **220** and a plurality of reinforced, heavy gauge ground wires **210** attached at one end to the ground shaft **205**. The ground shaft **205** and ground wires **210** should be highly conductive and have low resistance to supply current from the ground to the power receiver **10**. In one embodiment, the ground wires **210** may be copper or other highly-conductive metal. The end of the ground shaft may be pointed to facilitate insertion into the earth. A connection port on the ground shaft **220** is provided to electrically connect the ground antenna array **220** to the resonant transformer **30**.

[0065] The hollow cylinder **220** has external threads **25** to facilitate insertion into the ground. A rotator nut **235** is fixedly secured to the top end of the hollow shaft **220**. A square shaft **215** protrudes from the top end of the ground shaft **205** into the opening in the rotator nut **235**. FIG. **6B**. A tool **250**, shown in FIG. **7**, engages with the rotator nut **235** and square shaft **215** during insertion of the ground antennas array **200** into the ground as will be hereinafter described.

[0066] The insertion tool **250** is shown in FIG. **7**. The insertion tool **250** includes a tool body **255** having a first socket **260** on one side to fit the rotator nut **235** on the hollow cylinder **220** and a second socket **265** on the other side to fit the square shaft **215** on the ground shaft **205**. Arms **270** extend from the outer periphery of the tool body **255** for manually or mechanically turning the insertion tool **250**.

[0067] Before the antenna array **200** is deployed, the ground wires **210** are wound around the ground shaft **205** with the free ends protruding slightly from respective openings **230** in the hollow cylinder **220** to a distance not to exceed one half (½) the depth of the external threads **225** on the hollow cylinder **220**. FIG. **6**B illustrates the ground antenna array **200** before deployment. FIG. **6**C illustrates the ground antenna array in a deployed configuration.

[0068] Installation of the ground antenna array **200** is performed in two stages. In the first stage, a hole slightly smaller in diameter than the threads **235** of the hollow cylinder **220** is drilled into the Earth to a depth matching the length of the hollow cylinder **220** or slightly longer. The hole is filled with water and the water is allowed to soak into the soil. After the ground is softened, the hollow cylinder **220** is rotated using the insertion tool **250** to insert the ground antenna array **200** into the ground. The first socket **260** of the insertion tool **250** is engaged with the rotator nut **230** and the insertion tool **250** is turned by hand or a mechanized rotating shaft fitted and attached to the tool arms **270** to thread the ground assembly into the hole. During the initial insertion of the ground antenna array **200**, the ground shaft **205** is fixed to the hollow shaft **220** and rotates with the hollow shaft. The hollow cylinder **220** is rotated until it reaches the full depth of the hole.

[0069] Once the ground antenna array **200** has been fully inserted into the earth, the insertion tool **250** is flipped over and the second socket **265** of the insertion tool **250** is engaged with the square shaft **215**. The insertion tool **250** is turned by hand or a mechanized rotating shaft fitted and attached to the tool arms **270** to rotate the ground shaft **205**. During the second phase, the ground shaft **205** rotates freely inside the hollow cylinder **220**. Rotation of the ground shaft **205** causes the reinforced ground wires **210** to extend radially into the earth. The ground shaft **220** is rotated until the ground wires are fully extended. The ends of the ground wires may be sharpened to aid in the extension of the ground wires during the second phase.

[0070] After the ground antenna array **200** is deployed, a connection cable **280** is attached to a connection port **240** on the ground shaft **220** to electrically connect the ground antenna array **220** to the resonant transformer **30** in the power receiver **10**.

What is claimed is:

1. A power receiver for extracting electrical energy from the earth's electric field, said power receiver comprising:

a resonant transformer connected to a ground terminal disposed below the surface of the earth;

an impulse generator for generating and applying a high voltage electrical impulse to a primary winding of the resonant transformer to induce current flow from the ground terminal through the primary winding of the transformer; and

a power conversion circuit connected to a secondary winding of the resonant transformer to convert electrical current flowing through the secondary winding to a desired form.

2. The power receiver of claim **1** wherein a resonant frequency of the resonant transformers is below 200 Hz.

3. The power receiver of claim **1** wherein the resonant transformer comprises a ferro-resonant transformer.

4. The power receiver of claim **1** further comprising an elevated terminal, and wherein the primary winding of the resonant transformer is connected between the ground terminal and elevated terminal.

5. The power receiver of claim **1** wherein the impulse generator comprises:

a pulse generator for generating low voltage pulses;

a step-up transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses;

a spark gap connected between the step-up transformer and the primary winding of the resonant transformer to generate a spark responsive to the high voltage impulses from the step-up transformer.

6. The power receiver of claim **1** wherein the impulse generator comprises a solid state spark generator.

7. The power receiver of claim **1** wherein said resonant transformer includes a capacitor connected in parallel with the primary winding.

8. The power receiver of claim **1** wherein said resonant transformer includes a capacitor connected in series with the primary winding between the impulse generator and the elevated terminal.

9. The power receiver of claim **1** comprising multiple resonant transformers having primary windings connected in parallel between the ground terminal and the elevated terminal.

10. The power receiver of claim **9** wherein the resonant transformers have different resonant frequencies.

11. The power receiver of claim **10** wherein the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

12. A power receiver for extracting electrical energy from the earth's electric field, said power receiver comprising:

a resonant circuit connected to a ground terminal disposed below the surface of the earth, said resonant circuit having a resonant frequency below 200 Hertz

an impulse generator for generating and applying a high voltage electrical impulse to the resonant circuit to induce current flow from the ground terminal through the resonant circuit; and

a power conversion circuit connected to the resonant circuit to convert electrical current flowing through the resonant circuit to a desired form.

13. The power receiver of claim **12** wherein the impulse generator comprises:

a pulse generator for generating low voltage pulses;

a step-up transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses;

a spark gap connected between the step-up transformer and resonant circuit to generate a spark responsive to the high voltage impulses from the step-up transformer.

14. The power receiver of claim **12** wherein the resonant circuit comprises a resonant transformer having a primary winding, a secondary winding, and resonant capacitor connected in parallel with the primary winding.

15. The power receiver of claim **12** wherein the resonant circuit comprises a resonant transformer having a primary winding, a secondary winding, and resonant capacitor connected in series with the primary winding.

16. The power receiver of claim **12** wherein the resonant circuit comprises multiple resonant transformers having primary windings connected in parallel to the ground terminal.

17. The power receiver of claim **16** wherein the resonant transformers have different resonant frequencies.

18. The power receiver of claim **17** wherein the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

* * * * *

(No Model.)

G. F. DIECKMANN.

ELECTRIC EARTH BATTERY.

No. 329,724. Patented Nov. 3, 1885.

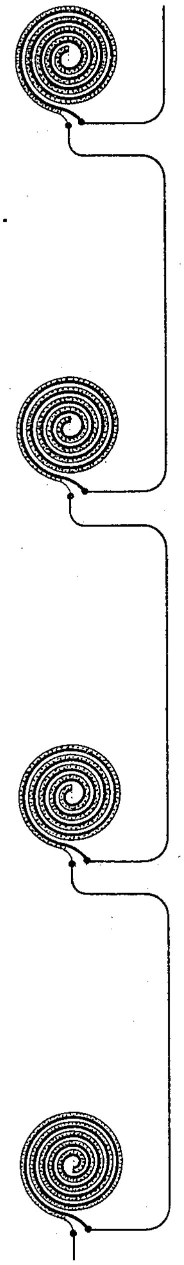

Fig. 1.

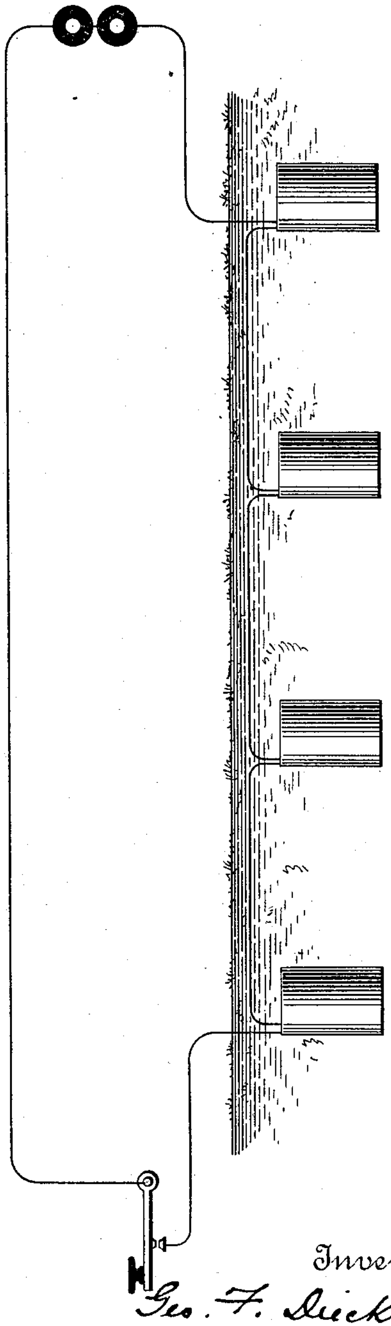

Fig. 2.

Witnesses

Geo. W. Breck.

Chas. J. Maguire

Inventor

Geo. F. Dieckmann

By his Attorneys

Curtis & Crocker

N. PETERS, Photo-Lithographer, Washington, D. C.

UNITED STATES PATENT OFFICE.

GEORGE F. DIECKMANN, OF NEW YORK, N. Y., ASSIGNOR OF ONE-FOURTH TO PAUL GMEHLIN, OF SAME PLACE.

ELECTRIC EARTH-BATTERY.

SPECIFICATION forming part of Letters Patent No. 329,724, dated November 3, 1885.

Application filed June 15, 1885. Serial No. 168,690. (No model.)

To all whom it may concern:

Be it known that I, GEORGE F. DIECKMANN, a citizen of Germany, residing in the city, county, and State of New York, have invented a new and useful Electric Earth-Battery, of which the following is a specification.

Earth-batteries, consisting of electrodes buried in the earth, as heretofore constructed, have not been capable of giving an electromotive force greater than that obtainable from a single couple—which is too small for practical purposes—because, if a number of couples or elements were used and connected in series all the couples stood in the same electrolyte and short-circuited one another. By my invention I am enabled to bury a series of couples in the same body of earth and connect them together in series, so as to obtain from such a battery an indefinitely high electro-motive force sufficient for charging storage-batteries, operating electric bells, telegraphic and other purposes, &c.

To this end my invention consists in arranging the electrodes in such a way that the resistance between the electrodes of each couple is small, while the resistance between the couples is relatively large, and connecting the negative electrode of each with the positive electrode of the next by a low-resistance conductor, so that the short-circuiting effect of the couples upon each other may be indefinitely reduced, and the effect is practically the same as if the couple were immersed in separated electrolytes. To arrange the relative resistances so as produce this effect I place the two electrodes of each couple comparatively near together, and the separate couples relatively far apart, so that the resistance between one electrode of one couple and the same electrode of the next couple is very much greater than the resistance between the two electrodes of the same couples, and consequently the short-circuiting effect is reduced comparatively to nothing.

My invention is illustrated in the accompanying drawings, Figure 1 representing my arrangement of electrodes in plan, and Fig. 2 showing them in elevation buried in the earth.

In my battery any electrodes may be used, and they may be of any desired size and shape; but in order to make the internal resistance of the battery as low as possible and obtain a cheap construction, I prefer to form each element or couple of zinc and copper plates or sheets separated the proper distance by a sheet of felt, asbestus, or other suitable non-conducting material and rolled up in a spiral form, as clearly represented in the drawings, by which means I obtain a compact element having a large surface with a small distance between the plates, and consequently a very low resistance. The zinc plate being the one most acted upon should be made the thicker, as I have indicated. Having constructed any desired number of these elements according to the electro-motive required, I bury them in the earth at a proper depth in any desired positions, but relatively far apart, so that the distance between them is great compared with the distance between the two electrodes of any one couple. The negative plate of each couple is then connected with the positive plate of the next through the series, as represented—and the terminals of the battery thus formed are connected with the circuit to be supplied, a telegraph-circuit, as represented, for example. The wires connecting the couples should preferably be insulated where they come near together below ground, to prevent short-circuiting there. I have shown them entirely buried in the earth; but of course they may be laid above ground. The distance apart that the couples should be placed depends upon the resistance of the telegraphic or other working circuit. If its resistance be very high, then the couples should be much farther apart, in order that the difference of potential between one electrode of one couple and the same electrode of the next couple may be high compared with the difference of potential between the two electrodes of the same couple. Similarly, the end couples of the battery should be so far apart with reference to the resistance of the working-circuit that the resistance between them through the earth is great compared with the resistance of the working-circuit, so that the working-circuit will not be short-circuited; or, to express it more accurately, the earth-resistance between the end couples should be so high that the fall of potential through it is great compared with the fall through the working-circuit. This effect is obtained by making the distance between the end couples correspondingly greater than that between adjacent couples—by arranging the couples in a straight line, for example. By thus properly proportioning the resistance between the couples to the circuit-resistance and internal resistance of the couples, the short-circuiting effect may be reduced more or less to an insignificant amount, and any desired electro-motive force obtained from the battery.

I claim as my invention—

1. An earth-battery consisting of a number of couples or elements buried in the earth and connected in series, the two electrodes of each couple being placed near together, while the couples or elements are placed comparatively far apart, whereby the short-circuiting effect of the couples upon each other is avoided, and an indefinitely high electro-motive force obtained.

2. An earth-battery consisting of a number of couples or elements composed of sheet-electrodes rolled up in a spiral form and buried in the earth and connected in series, the couples being arranged comparatively far apart, whereby the short-circuiting effect of the couples upon each other is avoided.

June 12, 1885.

GEORGE F. DIECKMANN.

Witnesses:

Francis B. Crocker,
Chas. J. Maguire.

(No Model.)

M. EMME.

GROUND GENERATOR OF ELECTRICITY.

No. 495,582. Patented Apr. 18, 1893.

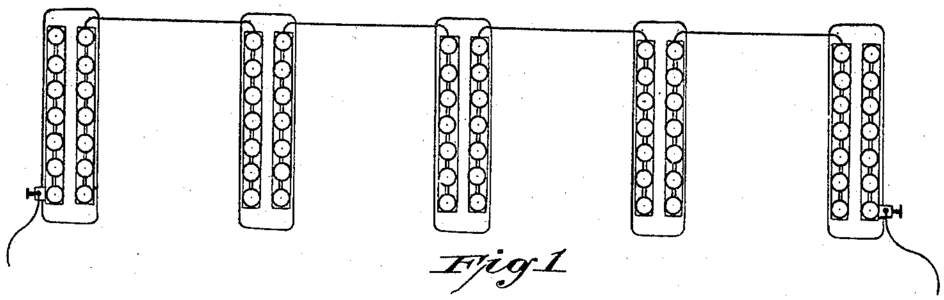

Fig 1

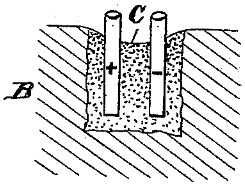

Fig 2

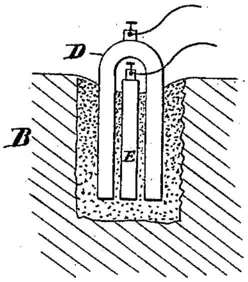

Fig 3

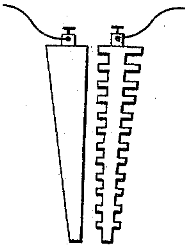

Fig 4

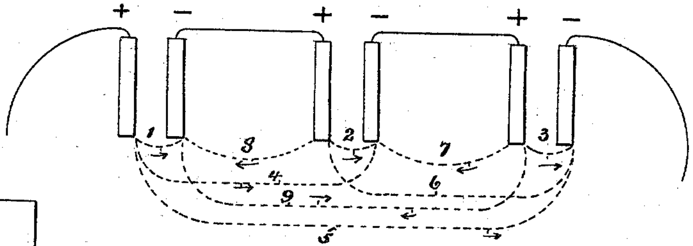

Fig 5

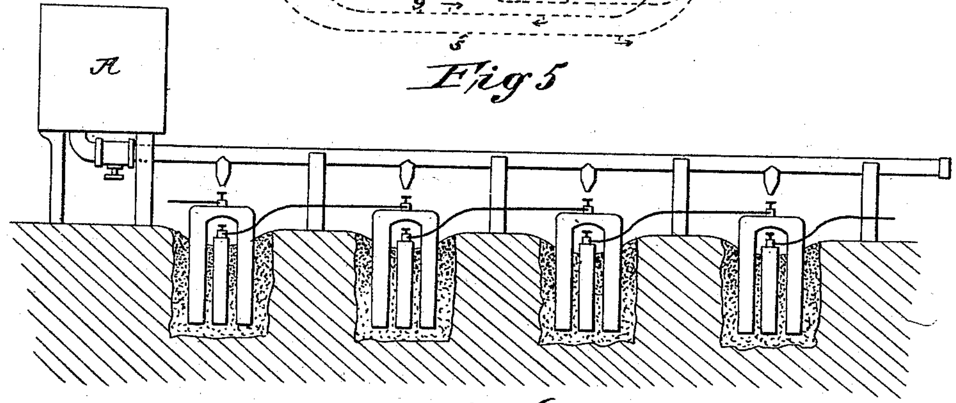

Fig. 6.

Attest;
C. W. Benjamin,
H. G. Tuchman

Inventor;
Michael Emme
by Read & Price
his Attys,

THE NORRIS PETERS CO., PHOTO-LITHO., WASHINGTON, D. C.

UNITED STATES PATENT OFFICE.

MICHAEL EMME, OF OAKLAND, CALIFORNIA.

GROUND GENERATOR OF ELECTRICITY.

SPECIFICATION forming part of Letters Patent No. 495,582, dated April 18, 1893.

Application filed August 29, 1892. Serial No. 444,391. (No model.)

To all whom it may concern:

Be it known that I, MICHAEL EMME, a citizen of France, residing at Oakland, in the county of Alameda and State of California, have invented certain new and useful Improvements in Ground Generators of Electricity; and I do hereby declare the following to be a full, clear, and exact description of the invention, such as will enable others skilled in the art to which it appertains to make and use the same.

My invention relates to chemical generators of electricity, a prepared body of earth being used as the support and exciting medium for the electrodes or elements. In a generator constructed in accordance with my invention, any desired number of elements may be assembled in the same piece of ground and coupled in series or multiple series to produce the electro-motive force desired. I find that if a series of galvanic couples be inserted in a body of ground so that a straight line will pass transversely through the several couples and the space between the several couples be made large comparatively to the distance between the two elements composing the couples, the couples may be joined for series in the same manner as if they were contained in independent vessels. In order, however, to attain the best results it is necessary to prepare the body of soil immediately adjacent to the two elements composing a couple in a manner which will be hereinafter fully described.

The several features of novelty of the invention will be hereinafter more fully described in this specification and definitely indicated in the claims appended to this specification.

In the accompanying drawings Figure 1 illustrates an earth generator in which the several couples are arranged in series. Figs. 2 and 3 show a cross section of ground containing one couple. Fig. 4 shows wedge shaped electrodes. Fig. 5 shows the co-operation of the internal circuits of the several couples. Fig. 6 shows the means for supplying an excitant or moistening medium periodically to the several couples.

Any character of soil may be adapted for use with my ground generator by saturating that portion of it surrounding the several elements with a suitable solution rich in oxygen, chlorine, bromine, iodine or fluorine, or with a solution of a salt of an alkali. As elements I prefer to use iron as a positive electrode and hard pressed coke carbon for the negative electrode. The positive electrode is preferably a U-shaped bar of iron round in cross section. The two limbs of the U straddle a rod of carbon. The iron should be soft wrought iron. Cast iron also can be used, but I find that cast iron gives a little less electro-motive-force, probably by reason of the percentage of carbon and other impurities contained therein. Magnesium also yields excellent results producing with carbon a voltage of 2.25. Zinc, aluminium or any metal with which the ground and its contained salts or exciting medium will develop electrolytic action may be used with varying results.

In carrying my invention into practice I level a piece of ground of sufficient area to contain the generator. For instance, for three hundred positive elements each twenty inches long and two inches in diameter bent as indicated in Fig. 3, and three hundred negative elements fifteen inches long and three inches in diameter, the length of the piece of ground should be about one hundred feet and its width about three feet. I dig forty-three holes, at a distance of thirty inches apart from center to center, in a line as indicated in Fig. 1. Each hole is ten inches wide and thirty inches broad and of a depth sufficient to contain the elements. The loose soil dug from the ground is mixed with a proper salt or acid to render the generator active. For instance, if the ground is a vegetable mold commercial concentrated nitric acid should be added in sufficient quantity to saturate the soil, and peroxide of manganese or pyrolusite should be mixed with the mass. If the soil be of a sandy character hydrochloric acid or carbonate of soda or potash may be substituted for the materials just mentioned. Should the soil be of a clayey character hydrochloric or sulphuric acid and chloride of sodium may be used, the salt being dissolved in water and poured in the hole before the acid is mingled with the soil. The bottom of the hole is moistened with water and the prepared soil mixed with water to the consistency of a thick paste is then introduced, the composite elements being immersed in it. The several groups of elements thus arranged may be connected in series by conductors as shown in Fig. 1, though the generator will act without any external conductors except the terminal wires. A generator constructed as above described will yield 53.85 volts and fifty-six ampères, developing a total of 3,015.60 watts, or about four horse-power. By increasing the number of cells the capacity of the generator may be correspondingly increased to any desired horse-power. The couples may be joined in simple series or in multiple series. The prepared body of soil should be periodically moistened preferably with the acid with which it was treated when first prepared for action, and in a plant constructed for continuous action I prefer to provide a reservoir as indicated at A in Fig. 6 and run a pipe of a material not attacked by the acid throughout the plant providing nozzles over the several couples so that they may be moistened when desired. With provisions of this nature the soil may be kept in a substantially uniform condition and the generator rendered continuously serviceable. Any accumulation of oxides or other products of the reaction between the prepared soil and the elements may be removed and a clean metal surface exposed for chemical action by raising the positive electrode and then forcing it back into place again. The carbon may be cleansed by simply turning it without lifting it from its place.

I find that the period of activity of the generator during which no addition of salt or acid is required increases with the length of service. For example, during the first day of use the exciting medium should be added after ten hours of work, after which it will yield twenty-six hours of service, and then after another addition of excitant it will yield service for two days, and so on.

In Fig. 1 is represented a generator the several electrodes of which are composed of rods grouped in multiple and arranged as indicated in Fig. 2 where B represents the solid ground and C the mass of prepared soil.

In Fig. 3 is shown a preferable shape of electrodes, D representing an iron electrode and E a carbon.

A wedge shaped electrode as shown in Fig. 4 permits easy withdrawal from the soil and the corrugation increases the extent of active surface.

In Fig. 5 are indicated the internal circuits of the generator. The several couples may be interconnected by conductors placed exterior to the ground as indicated by F or the conductors may be dispensed with, (although inferior results are obtained by such an arrangement) as indicated at G.

The dotted lines Fig. 5 indicate the directions of electro-motive-force in the internal circuits, 1, 2, 3, 4, 5, 6 being in one direction and 7, 8 and 9 in an opposite direction. There is a resultant or differential electro-motive-force which develops current in the external circuit.

I am not prepared to state the cause of constancy of action in the generator constructed as hereinbefore described, but I ascribe it to the mutual assistance of the several elements of the series, each positive element finding a conducting medium to all of the negative elements of the series as indicated by dotted lines in Fig. 5, the solid ground acting as a conducting medium. Thus the internal resistance of the generator is lowered and the output of the generator consequently increased. The fact that such internal circuits actually exist may be demonstrated by the interposition of a voltmeter in the ground between successive pairs of elements; the whole body of earth in the neighborhood of the generator seems to be permeated with stream lines of electric energy. This fact probably assists in some way to prevent polarization of the negative electrodes as the battery acts with great constancy.

Having thus described my invention, what I claim as new, and desire to secure by Letters Patent, is—

1. A generator of electricity comprising a series of galvanic couples each couple being embedded in an electrolytic medium composed of earth mixed to a pasty consistency with an exciting solution the spaces between the several couples being occupied by unprepared earth.

2. A ground generator of electricity comprising a series of galvanic couples embedded in an electrolytic medium composed of earth mixed to a pasty consistency with a suitable excitant and filled in holes formed at intervals in the earth.

3. A ground generator of electricity comprising a series of galvanic couples embedded in the earth and connected in series relation, the distance between the several couples being large compared to the distance between the elements of any couple, and a reservoir for periodically supplying a moistening medium and maintaining the action of the generator.

In testimony whereof I affix my signature in presence of two witnesses.

MICHAEL EMME.

Witnesses:
LINCOLN SONNTAG,
CHAS. SONNTAG.

US008357461B1

(12) **United States Patent**
Jackson

(10) **Patent No.:** **US 8,357,461 B1**
(45) **Date of Patent:** **Jan. 22, 2013**

(54) **EARTH ELECTROMOTIVE FORCE GENERATION METHOD AND APPARATUS**

(76) Inventor: **John R. Jackson**, Littleton, CO (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 595 days.

(21) Appl. No.: **12/618,506**

(22) Filed: **Nov. 13, 2009**

Related U.S. Application Data

(60) Provisional application No. 61/114,178, filed on Nov. 13, 2008.

(51) **Int. Cl.**
H01M 14/00 (2006.01)
H01M 2/00 (2006.01)
H01M 10/44 (2006.01)

(52) **U.S. Cl.** **429/47**; 429/7; 429/50

(58) **Field of Classification Search** 429/7, 9, 429/47, 50, 61, 67, 68, 90, 121, 122, 123, 429/149
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

160,152	A	2/1875	Bryan	
495,582	A	4/1893	Emme	
3,288,648	A *	11/1966	Jones	429/442
4,457,988	A *	7/1984	Ryeczek	429/47
5,605,715	A *	2/1997	Giardina et al.	156/151
5,681,662	A *	10/1997	Chen et al.	428/607
5,797,452	A *	8/1998	Martin	166/110
2006/0240297	A1*	10/2006	Takeda et al.	429/22

* cited by examiner

Primary Examiner — Philip Tucker
Assistant Examiner — Brian R Slawski
(74) *Attorney, Agent, or Firm* — Leyendecker and Lemire LLC; Kurt Leyendecker

(57) **ABSTRACT**

A method and apparatus for producing electricity using Earth electromotive force is disclosed herein. The method comprises placing a first electrode into a first section of the Earth, typically a hydrocarbon bearing formation. Next, a second electrode is placed into a second section of the Earth, typically in or proximal a water zone. Further, the first and second electrodes are electrically coupling to a boost converter assembly. The boost converter assembly is adapted to convert an input voltage derived from the first and second electrodes to an output voltage higher than the input voltage. An apparatus and electrical system comprising the various components for harvesting Earth energy are further disclosed.

20 Claims, 4 Drawing Sheets

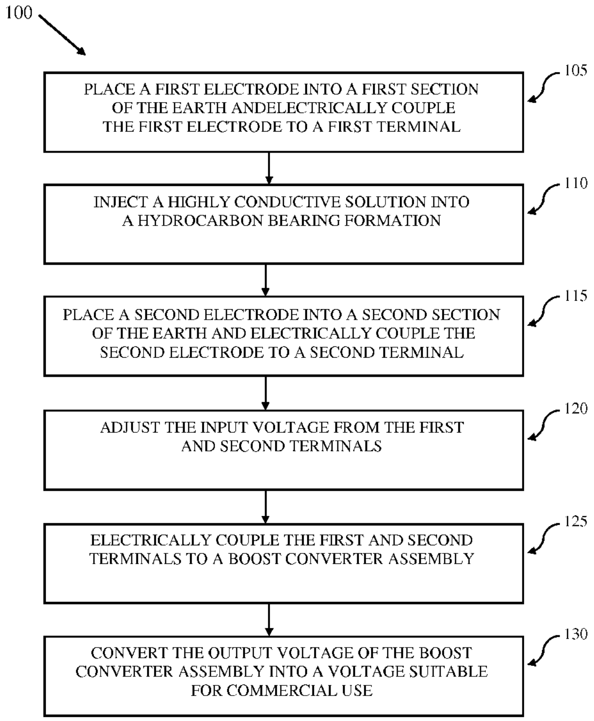
100
PLACE A FIRST ELECTRODE INTO A FIRST SECTION OF THE EARTH ANDELECTRICALLY COUPLE THE FIRST ELECTRODE TO A FIRST TERMINAL
105
INJECT A HIGHLY CONDUCTIVE SOLUTION INTO A HYDROCARBON BEARING FORMATION
110
PLACE A SECOND ELECTRODE INTO A SECOND SECTION OF THE EARTH AND ELECTRICALLY COUPLE THE SECOND ELECTRODE TO A SECOND TERMINAL
115
ADJUST THE INPUT VOLTAGE FROM THE FIRST AND SECOND TERMINALS
120
ELECTRICALLY COUPLE THE FIRST AND SECOND TERMINALS TO A BOOST CONVERTER ASSEMBLY
125
CONVERT THE OUTPUT VOLTAGE OF THE BOOST CONVERTER ASSEMBLY INTO A VOLTAGE SUITABLE FOR COMMERCIAL USE
130

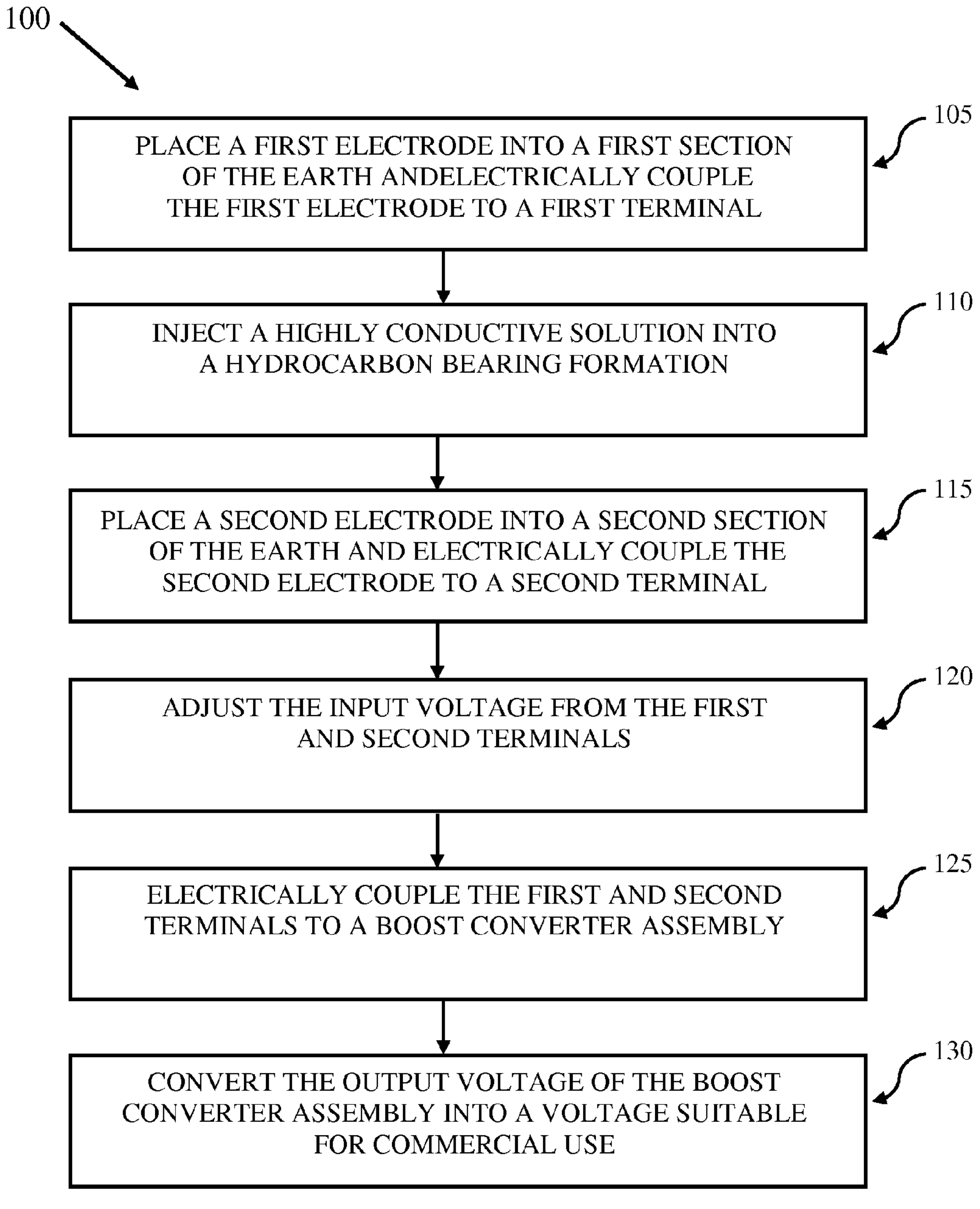
100
PLACE A FIRST ELECTRODE INTO A FIRST SECTION OF THE EARTH ANDELECTRICALLY COUPLE THE FIRST ELECTRODE TO A FIRST TERMINAL
105
INJECT A HIGHLY CONDUCTIVE SOLUTION INTO A HYDROCARBON BEARING FORMATION
110
PLACE A SECOND ELECTRODE INTO A SECOND SECTION OF THE EARTH AND ELECTRICALLY COUPLE THE SECOND ELECTRODE TO A SECOND TERMINAL
115
ADJUST THE INPUT VOLTAGE FROM THE FIRST AND SECOND TERMINALS
120
ELECTRICALLY COUPLE THE FIRST AND SECOND TERMINALS TO A BOOST CONVERTER ASSEMBLY
125
CONVERT THE OUTPUT VOLTAGE OF THE BOOST CONVERTER ASSEMBLY INTO A VOLTAGE SUITABLE FOR COMMERCIAL USE
130

FIG. 1

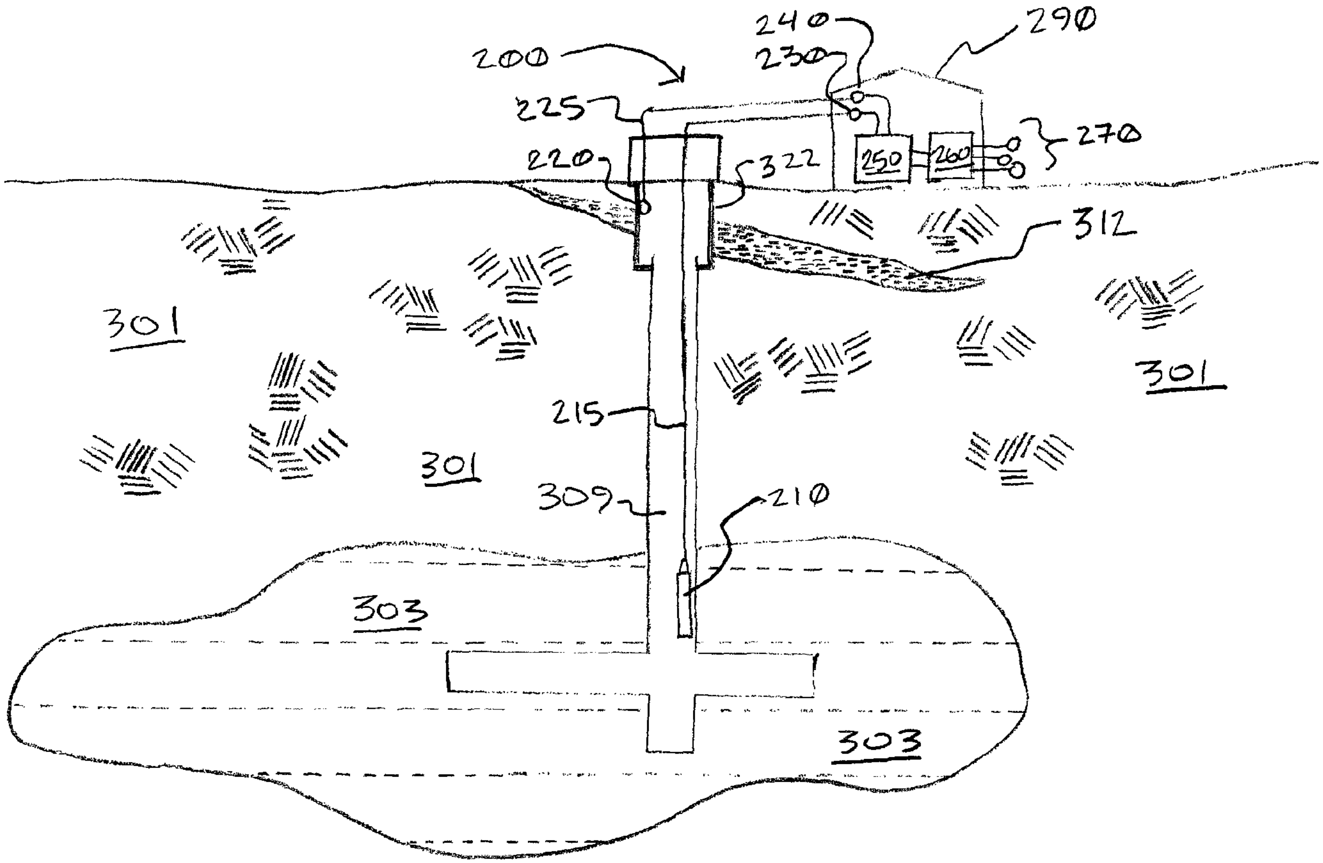

FIG. 2

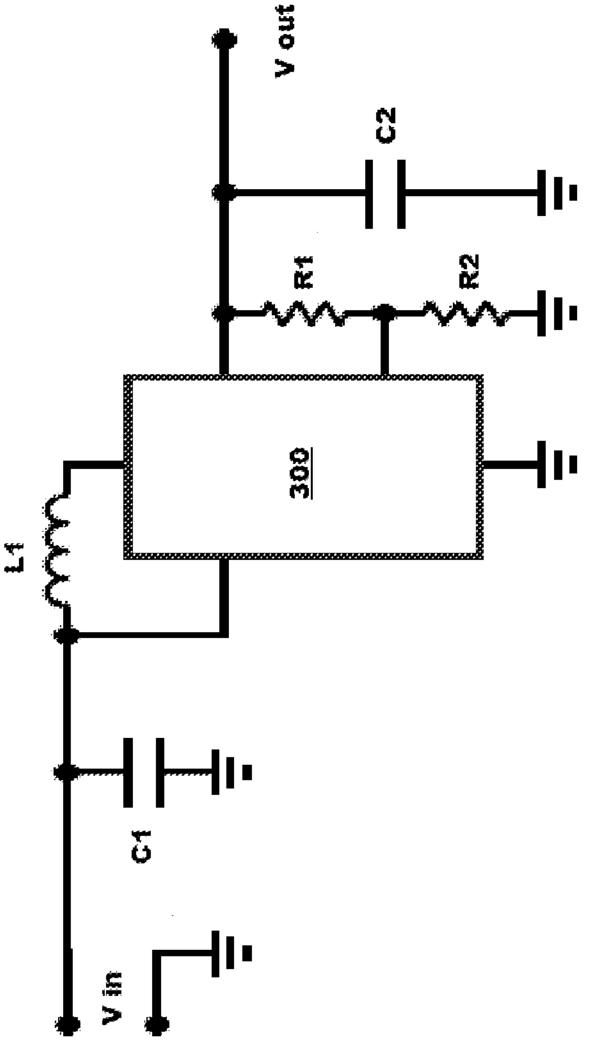
V out
C2
R1
R2
300
L1
C1
V in

FIG. 3

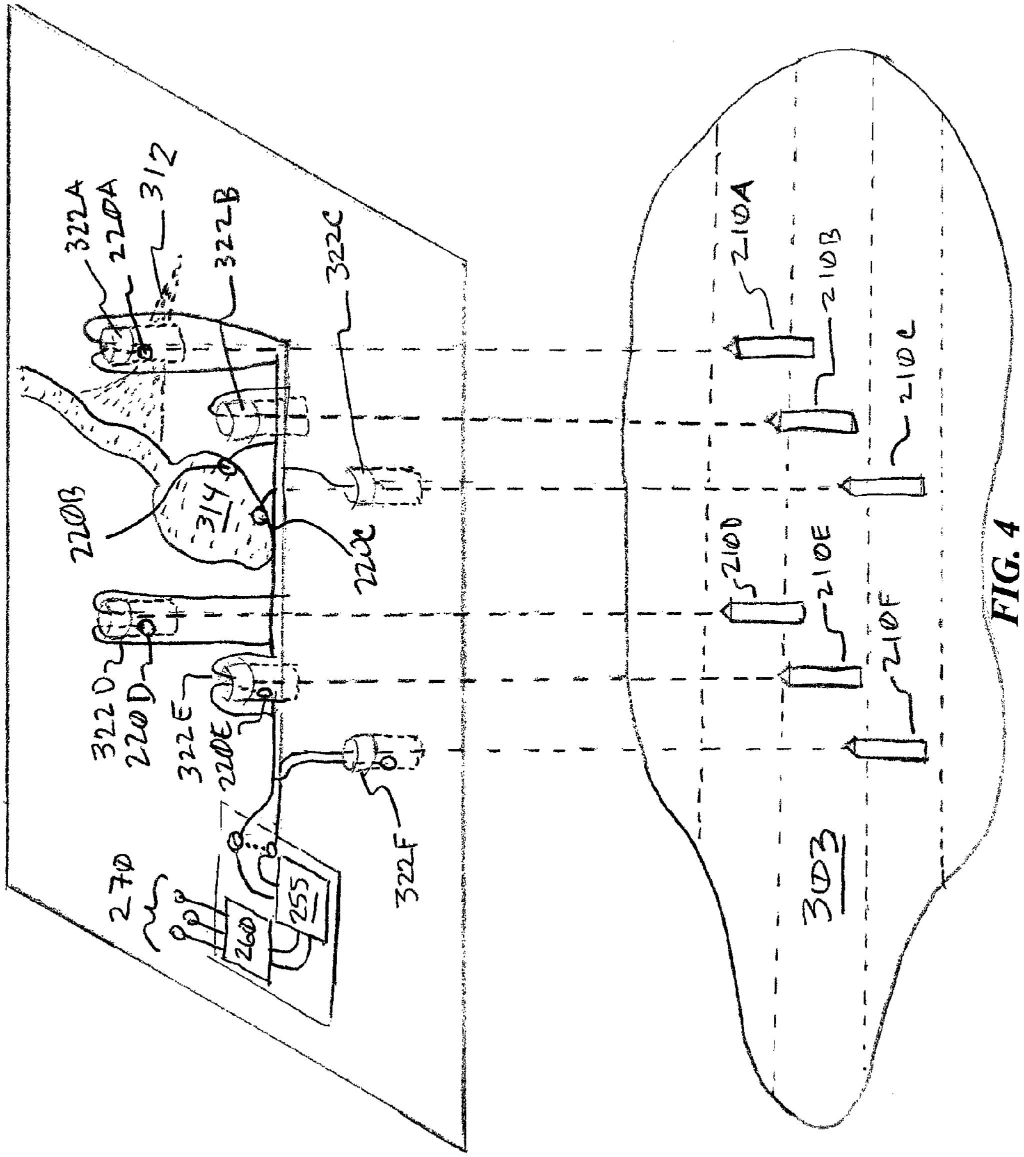
322F
210F

FIG. 4

EARTH ELECTROMOTIVE FORCE GENERATION METHOD AND APPARATUS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. Provisional Patent Application No. 61/114,178 filed Nov. 13, 2008, the full disclosure of which is incorporated herein by reference. The aforementioned provisional patent application has the title "Earth Electromotive Generator Apparatus and Method Based Upon Natural Earth Currents," and has the same named inventor as the present application.

FIELD OF THE INVENTION

The present invention relates generally to harvesting energy from the Earth and producing electricity therefrom. More particularly, the present invention relates to providing electricity from electromotive force of the Earth.

BACKGROUND

New energy sources are always desired as the world's energy demand continues to rise. Much of the world's energy is produced from extracting oil and gas from hydrocarbon bearing formations contained within the Earth's strata. Hydrocarbons in the form of oil and gas are usually extracted by drilling wells into these hydrocarbon bearing formations subsurface deep into the Earth's strata.

Often energy producing hydrocarbons remain trapped inside hydrocarbon bearing formations after drilling and practical methods of extraction have been exhausted. These hydrocarbons can provide a source of electrons for use in alternative energy methods. Moreover, in many modern extraction efforts, hundreds of deep wells and steel or metallic casings remain after extraction of the oil and gas has been completed thereby providing access to the hydrocarbon bearing formations.

Certain types of Earth batteries are generally known in the prior art, but lack a means to provide electricity for commercial purposes. Other means of harvesting Earth energy are not know in the prior art. Heretofore, there is a need in the art to improve methods and apparatus to provide a commercially viable energy source from Earth electromotive force generation.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a flow chart describing an exemplary method of the invention for producing electricity from Earth electromotive force according to an embodiment of the invention

FIG. **2** is a side plan view of an apparatus for producing electricity from Earth electromotive force according to an embodiment of the invention.

FIG. **3** is a schematic diagram for an exemplary boost converter circuit of an apparatus for producing electricity from Earth electromotive force according to an embodiment of the invention.

FIG. **4** is an overhead perspective view of a plurality of electrodes from the apparatus creating an electrical system for producing electricity from Earth electromotive force according to an embodiment of the invention.

DETAILED DESCRIPTION

Embodiments of the method and apparatus for providing electricity using Earth electromotive force comprise placing electrodes within sections of the Earth's strata to take advantages of discontinuities therein and boosting the small voltage potential between the electrodes. Embodiments include using discontinuities between two sections of the Earth's strata such as, but are not limited to, (i) a hydrocarbon bearing formation (ii) a water zone, (iii) a geothermal brine formation, (iv) a geothermal hot water formation, (v) a geothermal steam formation, (vi) a basaltic dyke, and (vii) a granite formation.

By extending the electrodes with low resistance conductors to two of the aforementioned sections of the Earth's strata, the discontinuities resulting from redox (reduction-oxidation) chemical reactions creating variations in current and ion flow within the Earth's strata can create a small voltage potential between the two electrodes. With modern boost circuitry, the small voltage potential between the electrodes can then be boosted, converted, and aggregated to provide a voltage source suitable for commercial use.

An exemplary embodiment comprises placing a first electrode into a hydrocarbon bearing formation and a second electrode proximal or directly in a water zone. The first and second electrodes are electrically coupled to a first and second terminal. The first and second terminals are likewise electrically coupled to a boost converter assembly. The boost converter assembly is adapted to convert an input voltage from the first and second terminals into an output voltage higher than the input voltage.

Typically, the first electrode is placed deep below the Earth's surface through the overburden to reach the hydrocarbon bearing formation. Hence, an elongated conductor may be required to allow the first electrode to reach the hydrocarbon bearing formation. In some embodiments, a highly conductive fluid is injected into the hydrocarbon bearing formation. Moreover, when a plurality of first and second electrodes are placed throughout sections of the Earth's strata with discontinuities therein, the power output of the energy source harvested from the Earth can be greatly increased.

Terminology

The terms and phrases as indicated in quotation marks (" ") in this section are intended to have the meaning ascribed to them in this Terminology section, applied to them throughout this document, including in the claims, unless clearly indicated otherwise in context. Further, as applicable, the stated definitions are to apply, regardless of the word or phrase's case, to the singular and plural variations of the defined word or phrase.

The term "or" as used in this specification and the appended claims is not meant to be exclusive; rather the term is inclusive, meaning "either or both."

References in the specification to: "one embodiment"; "an embodiment"; "another embodiment"; "an alternative embodiment"; "one variation"; "a variation"; and similar phrases mean that a particular feature, structure, or characteristic described in connection with the embodiment or variation, is included in at least an embodiment or variation of the invention. The phrase "in one embodiment," "in one variation," or similar phrases, as used in various places in the specification, are not necessarily meant to refer to the same embodiment or the same variation.

The term "couple" or "coupled," as used in this specification and the appended claims, refers to either an indirect or direct connection between the identified elements, components or objects. Often the manner of the coupling will be related specifically to the manner in which the two coupled elements interact.

The term "hydrocarbon bearing formation," as used in this specification and the appended claims, refers to any type of formation including one or more hydrocarbon containing layers. Hydrocarbon bearing formations may also contain one or more non-hydrocarbon layers with the one or more hydrocarbon containing layers. The hydrocarbons contained in hydrocarbon bearing formations are typically, but not necessarily, oil and gas.

The term "overburden," as used in this specification and the appended claims, refers to any type of material that lies above a hydrocarbon bearing formation. Non-limiting examples of materials in the overburden include rock and soil.

An Exemplary Method of Providing Electricity using Earth Electromotive Force

FIG. **1** is a flow chart illustrating an exemplary method for producing electricity from Earth electromotive force. Method **100** generally describes producing electricity from Earth electromotive force where discontinuities in the Earth's strata enable redox chemical reactions creating variations in current and ion flow within the Earth's strata. However, some operations will be described according to an exemplary implementation of method **100** where a hydrocarbon formation and a water zone are used as discontinuities within the Earth's strata.

As shown in block **105**, a first electrode is placed within a first section of the Earth. The first section of the Earth can, and typically is, a hydrocarbon bearing formation. The hydrocarbon bearing formation is typically subsurface hundreds or thousands of feet below the Earth's surface. Modern drilling techniques can provide bores or wells as deep as 29,000 feet subsurface to reach various hydrocarbon bearing formations. However, a bore or well is more typically in the range of 5,000-8,000 feet below the Earth's surface. The first electrode is electrically coupled to a first terminal (block **105**). As can be deduced, electrically coupling the first electrode within a hydrocarbon bearing formation 5,000-8,000 feet below the Earth's surface can present a challenge. An elongated insulated conductor, typically a wire or cable, typically accomplishes the electrical coupling. However, other variations of the elongated insulated conductor are contemplated as will be described later in this specification.

Additionally, when the first electrode is placed into a hydrocarbon bearing formation, a highly conductive solution can be injected into the hydrocarbon bearing formation (block **110**). The highly conductive solution can comprise any suitable fluid. Typically, but not necessarily, the solution acts as an electrolyte in which ions made readily flow. In one variation, the highly conductive solution is salt water comprising sulfuring acid. The percentage of sulfuric acid can vary depending on the specific implementation, but is typically between 5-20%. Similarly, the concentration of salt in the water of the highly conductive solution is typically 20% or greater. Moreover, a powder of conductive metal may be added to and suspended in the highly conductive solution. For instance, a powder of zinc can be used to increase oxidation reactions thereby increasing the number of electrons available to the anode.

Next, as shown in block **115**, a second electrode is placed within a second section of the Earth. The second section of the Earth can, and typically is, a water zone. The second electrode can be placed in or proximal the water zone. Water zones include, but are not limited to, surface water zone or groundwater zone. Surface water zones include lakes (man-made or natural), rivers and streams, whereas groundwater zones include unconfined and confined aquifers. Moreover, it is often advantageous to electrically couple the second electrode to a metal casing of a shallow well bore, particularly if the metal casing is proximal or within a water zone such as a groundwater zone. Metal casings of the shallow well typically comprise the first section of an oil and gas well from the Earth's surface to a few hundred feet into the overburden.

With both the first and second electrodes properly placed and installed, an input voltage (typically very small at approximately less than 2 Volts) will be present at the first and second terminals. The input voltage can be measured to determine whether it is within a desired operating range. Hence, the input voltage at the first and second terminal can be adjusted (block **120**). In some implementations of method **100**, it is desirable to adjust the input voltage to approximately 900 mVolts prior for an effective boost converter stage.

The input voltage can be adjusted in any number of ways. For example, the first electrode can be moved up or down in the hydrocarbon bearing formation. In some implementations, the first electrode can be moved horizontally through the hydrogen bearing formation as some wells have horizontal well portions extending approximately 2,000 feet. Additionally, the second electrode can be moved closer or further from the water zone. Moreover, the size of either or both of the first and second electrodes can be increased or decreased. Also, the resistance of the elongated insulated conductor to the first terminal can be increased, but such a modification may not be desirable in some implementations as it would increase power loss prior to the boost converter stage.

Next, as shown in block **125**, the first and second terminals are electrically coupled to a boost converter assembly. Essentially, the boost convertor assembly comprises electronic circuitry to boost the unusable low input direct current (DC) voltage, typically lower than 1 Volt, to a more workable or desirable higher DC voltage, typically between 3-5 Volts. Additional DC-DC converters can boost the voltage to an output voltage of approximately 12-15 Volts.

As next described in block **130**, the operation of converting the output voltage of the boost converter assembly into a voltage suitable for commercial use is performed. Alternating current (AC) voltage is generally desired for commercial application. Typically, the output voltage is fed into a DC-AC inverter, which converts the approximately 12-15 Volts DC into an suitable AC power source of the required frequency and voltage (e.g., 60 Hertz 110-120 Volts AC).

It is to be appreciated that method **100** and variations thereof can be performed where the first and second section of the Earth are other than a hydrogen bearing formation and a water zone. The first and second sections can be areas of the Earth where discontinuities resulting from redox chemical reactions create variations in current flow between the first and second sections of the Earth such as, but not limited to, a geothermal brine formation, a geothermal hot water formation, a geothermal steam formation, a basaltic dyke, and a granite formation.

Embodiments of an Apparatus for Providing Electricity using Earth Electromotive Force

An embodiment of an apparatus for providing electricity using Earth electromotive force is illustrated in FIG. **2**. Apparatus **200** comprises a first electrode **210**, a first extended conductor **215**, a first terminal **230**, a second electrode **220**, a second conductor **225**, a second terminal **240**, and a boost converter assembly **250**. Apparatus **200** typically also comprises a DC-AC inverter assembly **260** having an AC output **270**.

Still referring to FIG. **2**, an oil and gas well **309** comprises a metal casing **322** of a shallow well bore at the Earth's surface. The oil and gas well **309** continues through an overburden **301** and into a hydrocarbon bearing formation **303**. As illustrated within the hydrocarbon bearing formation **303**, the oil and gas well **303** can have horizontal portions. Additionally, a groundwater zone **312** is shown running through a portion of the metal casing **322**.

Typically, but not necessarily, the first electrode **210** is an anode in an embodiment. As illustrated, the first electrode **210** is placed within the hydrocarbon bearing formation **303**. The hydrocarbons contained within the hydrocarbon bearing formation **303** have an excess supply of electrons. Also as previously described, the highly conductive solution can be added to the hydrocarbon bearing formation **303** to mix with the hydrocarbons thereby increasing conductivity and creating an electrolyte solution. The hydrocarbons are oxidized and lose electrons which are accumulated by the first electrode **210** acting as the anode. The first electrode **210** comprises material adapted to gain electrons such as, but not limited to zinc. Moreover, the first electrode **210** can be an elongated rod, typically metallic, between 1-2 inches in diameter and 50-100 feet long. However, it is to be appreciated that the first electrode can vary substantially between implementations and can comprise a number of different shapes and materials.

The first electrode **210** is connected and electrically coupled to a first extended conductor **215**. The first extended conductor **215** is typically insulated and of low resistance. Moreover, the first extended conductor **215** is electrically coupled to the first terminal **230**. The first extended conductor **215** can be an insulated wire or cable. However, given the extensive depth of the oil and gas well **309**, alternative first extended conductors can be utilized. For example, the first extended conductor **215** can be an elongated non-conductive tube comprising a highly conductive solution therein to establish conductivity between the first electrode **210** and the first terminal **230**.

Typically, but not necessarily, the second electrode **220** is a cathode in an embodiment. Still referring to FIG. **2**, the second electrode **220** is connected and electrically coupled to the metal casing **322** of the shallow well bore. Further, the metal casing **322** is in contact with a water zone, specifically the groundwater zone **312**. Water molecules and materials/chemicals in the overburden **301** in contact with the second electrode **220** can gain electrons (naturally having a lack of electrons) thereby creating a reduction reaction. Hence, the water molecules and materials/chemicals in the overburden **301** adopt electrons from the second electrode **220** acting as a cathode. The second electrode **220** comprises material adapted to lose electrons such as, but not limited to copper. The second electrode **220** is electrically coupled to the second terminal **240** through the second conductor **225**. It is pertinent to note that other embodiments comprise utilizing natural Earth currents associated with discontinuities in otherwise homogeneous Earth strata differently than as described above.

Importantly to apparatus **200**, the boost converter assembly **250** receives an input voltage from the first and second terminals **230** & **240** and converts the input voltage to an output voltage higher than the input voltage. The electronic circuitry of the boost converter assembly **250** boosts the unusable low voltages from the first and second terminals **230** & **240** to a more workable higher voltage. The low input DC voltage, typically lower than 1 Volt, can be boosted to a more workable higher DC voltage, typically between 3-5 Volts. Additional DC-DC converters can boost the voltage to an output voltage of approximately 12-15 Volts.

Still referring to FIG. **2**, the DC-AC inverter assembly **260** receives the DC voltage from the output of the boost converter assembly **250**. The DC-AC inverter assembly **260** converts the DC voltage (approximately 12-15 Volts DC) from the output of the boost converter assembly **250** into the AC output **270**. The AC output **270** output voltage from the DC-AC inverter assembly **260** comprises a suitable AC power source such as, but not limited to, a 60 Hertz 110-120 Volts AC output. Additionally, the boost converter assembly **250** and DC-AC inverter assembly **260**, as well as other electronic components of the apparatus **200** can be contained within a weatherproof but **290** proximal the metal casing **322** of the shallow well bore and oil and gas well **309**.

FIG. **3** is a schematic diagram for an exemplary boost converter circuit of an apparatus for producing electricity from Earth electromotive force according to an embodiment of the invention. Boost converter circuit comprises terminals for receiving a DC input voltage (V in). Referring momentarily back to FIG. **2**, the input voltage from the first and second terminals **230** & **240** may be received by one or more boost converter circuits. In one embodiment, the boost converter circuit (FIG. **3**) comprises a first capacitor C**1** having a value of 4.7 μF, an first inductor L**1** having a value of 4.7 μH, a first resistor having a value of 1.02 Mega Ohms, a second resistor having a value of 604 kilo Ohms, second capacitor having a value of 10 μF, and a DC-DC converter **300**. The DC-DC converter **300** is a synchronous, fixed frequency, step-up DC-DC converter contained within an integrated chip. Additionally, the boost converter circuit has a DC output voltage (V out). Again referring momentarily back to FIG. **2**, the output voltage and terminals from the boost converter assembly **250** illustrated into DC-AC inverter assembly **260** may be the DC output voltage(s) of one or more boost converter circuits. One or more boost converter circuits in a parallel array along with associated wiring and electrical components comprise an embodiment of the boost converter assembly **250**.

It is to be appreciated that the use of the solid state electronics in the one or more boost converter circuits enables the small voltage potential created from discontinuities between the two sections of the Earth's strata to be converted into a voltage suitable for commercial use.

FIG. **4** is an overhead perspective view of a plurality of electrodes from the apparatus creating an electrical system for producing electricity from Earth electromotive force according to an embodiment of the invention. An embodiment of the electrical system for producing electricity using Earth electromotive force comprises a plurality of first electrodes **210***a*, **210***b*, **210***c*, **210***d*, **210***e*, & **210***f*, a plurality of second electrodes **220***a*, **220***b*, **220***c*, **220***d*, **220***e*, & **220***f* and a multi-circuit boost converter assembly **255**.

In one embodiment, each first electrode **210***a*, **210***b*, **210***c*, **210***d*, **210***e*, & **210***f* is an anode and electrically coupled to a negative terminal through an extended conductor adapter to allow each first electrode **210***a*, **210***b*, **210***c*, **210***d*, **210***e*, & **210***f* to be placed within the hydrocarbon bearing formation **303**. Each second electrode **220***a*, **220***b*, **220***c*, **220***d*, **220***e*, & **220***f* is a cathode and electrically coupled to a positive terminal through a conductor to various areas. As illustrated, second electrodes **210***a*, **220***d*, **220***e*, & **220***f* are connected and electrically coupled to their metal casings **322***a*, **322***d*, **322***e*, & **322***f* some of which are in contact with groundwater zone **312**, whereas second electrodes **210***b* & **210***c* are placed in a lake **314**.

Still referring to FIG. **4**, the multi-circuit boost converter assembly **255** is very similar to the aforementioned boost converter assembly except that is adapted to receive each negative terminal from the plurality of first electrodes **210***a*, **210***b*, **210***c*, **210***d*, **210***e*, & **210***f*, and each positive terminal from the plurality of second electrodes **220***a*, **220***b*, **220***c*, **220***d*, **220***e*, & **220***f* into a boost converter circuit (FIG. **3**). The multi-circuit boost converter assembly **255** then converts an input voltage from each boost converter circuit to an output voltage higher than the input voltage and combines the output voltage from each boost converter circuit thereby providing a higher power output. Additionally, the output voltage can be converted to the AC output **270** suitable for commercial use by converting it with the DC-AC inverter assembly **260**.

It is pertinent to note that may variations of the electrical system exist. For instance, each of the plurality of second electrodes may be coupled to a single metal casing. Moreover, tens or hundreds of first and second electrodes can exist in a large oil field significantly increasing to power source of the electrical systems for producing electricity from Earth electromotive force. For example, using 100 boost converter circuits in the multi-circuit boost converter assembly can produce approximately 50 watts output at 3.3 Volts DC.

Alternate Embodiments and Variations

The various embodiments and variations thereof illustrated in the accompanying figures and/or described above are merely exemplary and are not meant to limit the scope of the invention. It is to be appreciated that numerous variations to the invention have been contemplated as would be obvious to one of ordinary skill in the art with the benefit of this disclosure. All variations of the invention that read upon the appended claims are intended and contemplated to be within the scope of the invention.

For example, the first electrode is typically an anode and the second electrode is typically a cathode of the apparatus. However, in some alternative embodiments, it is contemplated that the apparatus for generating power via Earth electromotive force may be used in a recharging battery configuration whereby the anode would be a positive terminal and the cathode would be a negative terminal. Hence, in such alternative embodiments, the cathode would send current back to an external generator.

I claim:

1. A method for producing electricity using Earth electromotive force, the method comprising:

placing a first electrode into a first section of the Earth, the first electrode being electrically coupled to a first terminal;

placing a second electrode into a second section of the Earth, the second electrode being electrically coupled to a second terminal; and

electrically coupling the first and second terminals to a boost converter assembly, the boost converter assembly adapted to convert an input voltage of approximately 2 volts or less from the first and second terminals to an output voltage of 3 to 5 volts, the boost converter assembly comprising a plurality of boost converter circuits operatively configured in parallel wherein each boost converter circuit comprises a synchronous step-up DC-DC converter contained within an integrated chip.

2. The method of claim **1**, wherein the first section of the Earth is a hydrocarbon bearing formation.

3. The method of claim **1**, wherein the second section of the Earth is a water zone.

4. The method of claim **2** further comprising,

injecting a highly conductive solution into the hydrocarbon bearing formation.

5. The method of claim **4**, wherein the highly conductive solution comprises salt water and sulfuric acid.

6. The method of claim **4**, wherein the highly conductive solution includes a powder of conductive metal.

7. The method of claim **1** further comprising,

adjusting the input voltage at the first and second terminals.

8. The method of claim **7**, wherein said adjusting the input voltage comprises moving the first and second electrodes a distance between each other.

9. The method of claim **7**, wherein said adjusting the input voltage comprises adjusting a voltage potential between the first and second electrodes until the voltage potential is approximately 900 mVolts.

10. The method of claim **1**, wherein said placing the first electrode comprises lowering the first electrode connected to a low resistance insulated conductor into a man-made bore on the Earth's surface to a distance greater than approximately 5,000 feet, and said placing the second electrode comprises connecting the second electrode to a metal casing of a shallow well bore.

11. The method of claim **1** further comprising,

converting the output voltage of the boost converter assembly into a standard AC voltage adapted for commercial use.

12. The method of claim **1**, wherein the first and second section of the Earth are in areas of the Earth where discontinuities resulting from redox chemical reactions create variations in current flow between the first and second sections of the Earth.

13. An apparatus for producing electricity using Earth electromotive force, the apparatus comprising:

a first electrode, the first electrode being an anode and installed in a first section of the Earth having an abundance of electrons relative to other parts of the Earth's strata;

a first extended conductor, the first extended conductor being insulated and electrically coupled to the first electrode;

a first terminal, the first terminal being electrically coupled to the first extended conductor;

a second electrode, the second electrode being a cathode and installed in a second section of the Earth having a lack of electrons relative to other parts of the Earth's strata;

a second terminal, the second terminal being electrically coupled to the second electrode; and

a boost converter assembly, the boost converter assembly adapted to convert an input voltage of approximately 2 volts or less from the first and second terminals to an output voltage of 3 to 5 volts, the boost converter assembly comprising a plurality of boost converter circuits operatively configured in parallel wherein each boost converter circuit comprises a synchronous step-up DC-DC converter contained within an integrated chip.

14. The apparatus of claim **13**, wherein (i) the first section of the Earth is a hydrocarbon bearing formation and (ii) the second section of the Earth is a water source.

15. The apparatus of claim **14**, wherein the water source is one of a lake and a groundwater flow area.

16. The apparatus of claim **13**, wherein the first electrode is an elongated conductive rod comprising copper and having a diameter between 1 and 2 inches and a length between 50 to 100 feet.

17. The apparatus of claim **13**, wherein the first extended conductor is a shielded wire having a length greater than 1,000 feet.

18. The apparatus of claim **13**, wherein the first extended conductor is an elongated non-conductive tube comprising a highly conductive solution therein.

19. An electrical system for producing electricity using Earth electromotive force, the electrical system comprising:

a plurality of first electrodes, each first electrode being an anode and electrically coupled to a negative terminal through an extended conductor adapted to allow each first electrode to be placed within a hydrocarbon bearing formation;

a plurality of second electrodes, each second electrode being a cathode and electrically coupled to a positive terminal through a conductor to a one of (1) a metal casing of a shallow well bore proximal a water zone and (2) a water zone; and

a multi-circuit boost converter assembly, the multi-circuit boost converter assembly comprising a plurality of boost converter circuits operatively configured in parallel, the multi-circuit boost converter assembly adapted to receive each negative terminal from the plurality of first electrodes and each positive terminal from the plurality of second electrodes, each boost converter circuit being conductively coupled with each other boost converter circuit in parallel and adapted to-convert an input voltage to an output voltage higher than the input voltage, each boost converter circuit comprising a synchronous step-up DC-DC converter contained within an integrated chip.

20. The electrical system of claim **19**, wherein the multi-circuit boost converter assembly further comprises,

a DC-DC converter assembly, the DC-DC converter assembly adapted to receive the higher power output from multi-circuit boost converter assembly and boost the output voltage to approximately 12 to 15 Volts, and

a DC-AC inverter assembly, DC-AC inverter assembly adapted to receive the output voltage to approximately 12 to 15 Volts from the DC-DC inverter assembly and convert the output voltage to a standard 120 Volts AC output.

* * * * *

764

UNITED STATES PATENT OFFICE.

EMIL JAHR, OF BERLIN, GERMANY.

METHOD OF UTILIZING ELECTRICAL EARTH-CURRENTS.

SPECIFICATION forming part of Letters Patent No. 690,151, dated December 31, 1901.

Application filed July 24, 1900. Serial No. 24,680. (No specimens.)

To all whom it may concern:

Be it known that I, EMIL JAHR, a subject of the King of Prussia, German Emperor, residing at Berlin, in the Kingdom of Prussia and German Empire, have invented certain new and useful Improvements in Methods of Utilizing Electrical Earth-Currents, (for which I have applied for patents in England, No. 9,971, dated May 30, 1900; in France, No. 289,156, dated June 2, 1900; in Belgium, No. 118,752, dated June 2, 1900; in Germany, dated May 26, 1900, and in Russia, dated June 5, 1900,) of which the following is a specification.

It is known that continuous electric currents flow through the solid and liquid portions of the earth's crust. Lamont (*Der Erdstrom und der Zusammenhang desselben mit dem Erdmagnetismus*, Leipsic, 1862) showed the existence of such currents by thrusting two metal plates at a certain distance from each other into the ground in the direction of the magnetic or astronomical meridian and connecting them above the ground by means of a wire including a galvanometer. The stronger currents of this kind have been shown to flow from south to north. Later observations (Weinstein, *Electrotechnische Zeitschrift*, 1898, p. 794) show that this so-called "earth-current" possesses a considerable uniformity of current strength and pressure. From the fact that this earth-current flows from south to north it may be concluded that its pressure will increase if starting from a given point in the south the distance between this and the northern point is increased, provided the connecting-lead be of little resistance. Notwithstanding that these facts have been known the earth-current has never been utilized, because in order to obtain a sufficient current the distance between the north and south points must be so great that the cost of an installation is prohibitive. I have discovered that the south to north direction of the earth-current is maintained only when the plates are of the same metal. When zinc is used for the southern plate and iron or carbon for the northern one, the current flows through the lead connecting them from north to south. All the common metals behave relatively similarly—that is, the current flows always from that plate whose position in the ordinary electrical potential series is near the negative end, (palladium.) I have also discovered that the current produced is of highest tension when the two metals are most widely separated from each other in the electrical potential series and that which is nearer to the positive end (zinc) of this series is thrust into earth or water at the more northerly point, while that which is nearer the negative end (palladium) of the series is thrust into earth or water at the southerly point. In such an arrangement the metals are not appreciably chemically corroded, even when they are in earth saturated with water, and are connected together by lead for a long time.

My invention relates, therefore, to a method for utilizing the earth-current, consisting in thrusting into the earth or water on the earth two electrodes, of which the one has a position north of the other and is of a material nearer to the positive end (zinc) of the electrical potential series than is the material of the other electrode and connecting the two above ground by a lead. It is advantageous that the northerly more electropositive electrode should be thrust more deeply into the earth or water than the southerly electrode, for then the current is stronger. The best result as to strength of current and pressure is obtained when this difference in depth is such that a line joining the two electrodes is in the direction of the magnetic dip. The best effect, therefore, is obtained when these conditions are combined—namely, when the two electrodes are approximately in the magnetic meridian and are of materials widely separated in the electrical potential series and the northerly and more electropositive electrode is buried at a greater depth than that at which the southerly electrode is buried and that a line joining them has the inclination of the magnetic dip. The current can be tapped from the lead joining the electrodes and utilized in any well-known manner.

What I claim, and desire to secure by Letters Patent of the United States, is—

1. A method of utilizing electrical earth-currents consisting in placing electrodes occupying different positions in the electrical potential series into the earth or water so that the electrode which is nearer the negative end of the series is at the more southerly point,

while the electrode which is nearer to the positive is at the more northerly point, and connecting the two electrodes together by a lead, the northern electrode being inserted deeper into the earth or water than the southern one, substantially as described.

2. A method of utilizing electrical earth-currents consisting in placing electrodes occupying different positions in the electrical potential series into the earth or water so that the electrode which is nearer the negative end of the series is at the more southerly point, while the electrode which is nearer to the positive is at the more northerly point, and connecting the two electrodes together by a lead, the insertion of the electrodes into the earth or water at such heights that a line connecting the same is situated in the direction from north to south, and also in the direction of the magnetic dip, substantially as described.

In testimony whereof I have hereunto set my hand in the presence of two witnesses.

EMIL JAHR.

Witnesses:

WOLDEMAR HAUPT,

HENRY HASPER.

United States Patent [19]

Ryeczek

[11] **Patent Number: 4,457,988**

[45] **Date of Patent: Jul. 3, 1984**

[54] **EARTH BATTERY**

[76] Inventor: **John J. Ryeczek,** R.D. #1, Box 190C, Point Marion, Pa. 15474

[21] Appl. No.: **384,781**

[22] Filed: **Jun. 4, 1982**

[51] **Int. Cl.³** **H01M 14/00**

[52] **U.S. Cl.** .... **429/47;** 204/105 R; 175/64

[58] **Field of Search** 429/10, 47, 13

[56] **References Cited**

U.S. PATENT DOCUMENTS

155,209	9/1874	Snow	429/47
182,802	10/1896	Cerpaux	429/47
329,724	11/1885	Dieckmann	429/47
495,582	4/1893	Emme	429/47
728,381	5/1903	Emme	429/47
3,288,648	11/1966	Jones	429/13

Primary Examiner—Donald L. Walton

Attorney, Agent, or Firm—Webb, Burden, Robinson & Webb

[57] **ABSTRACT**

An earth battery which includes a first electrode which is a veined material located in the earth, a second electrode electrically connected to the surface of the earth and disposed in a hole which extends from the surface of the earth into the first electrode, an electrolyte disposed in the hole and in contact with the first and second electrodes, and means for electrically connecting the first electrode to the surface of the earth. A seam of coal is preferred as the first electrode. A method of mining metals located in the earth using the earth battery and a method of drilling holes in the earth are also disclosed.

21 Claims, 6 Drawing Figures

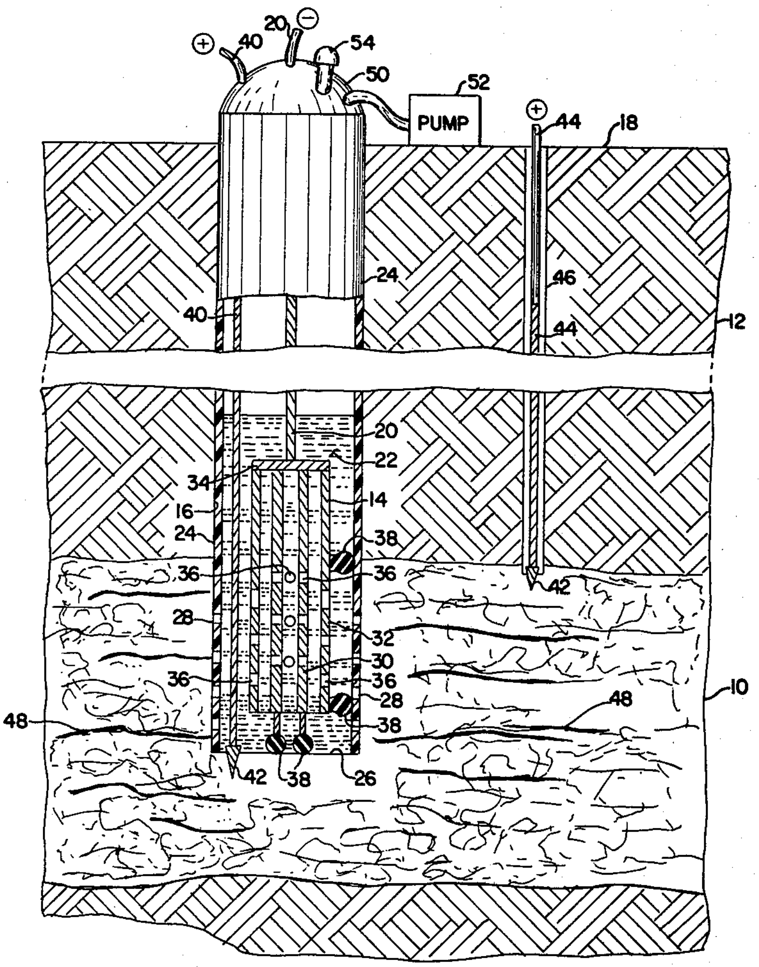

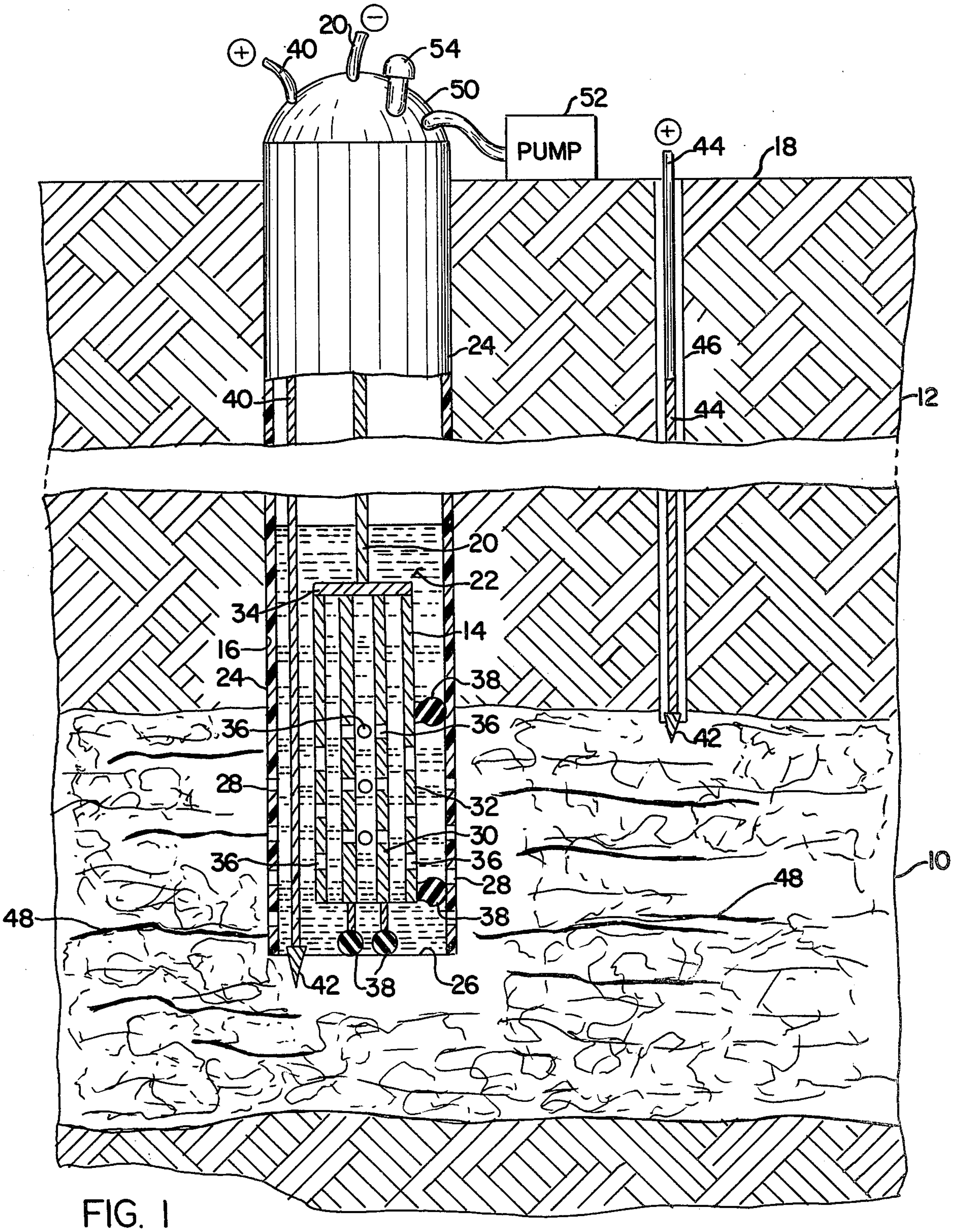
+
−
PUMP
+

FIG. 1

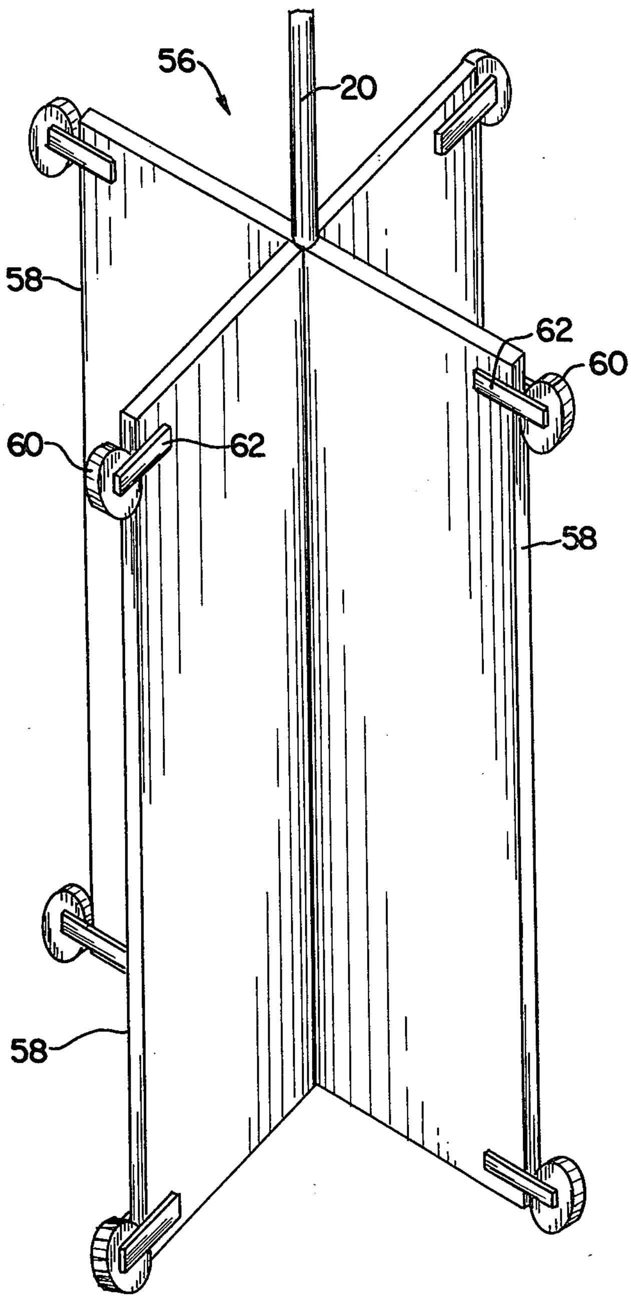
56
20
58
62
60
60
62
58
58

FIG. 2

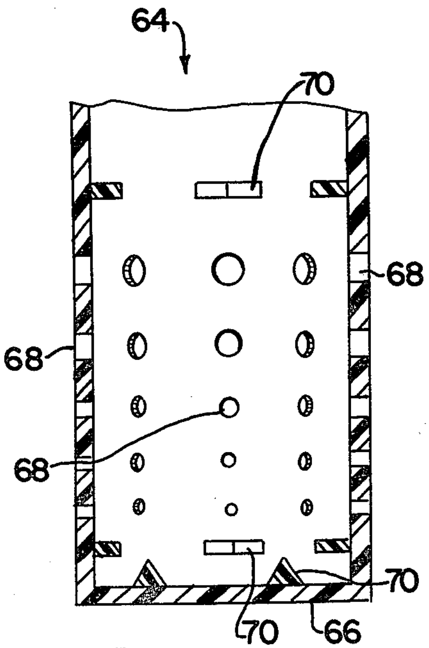
64
70
68
68
68
70
70
66

FIG. 3

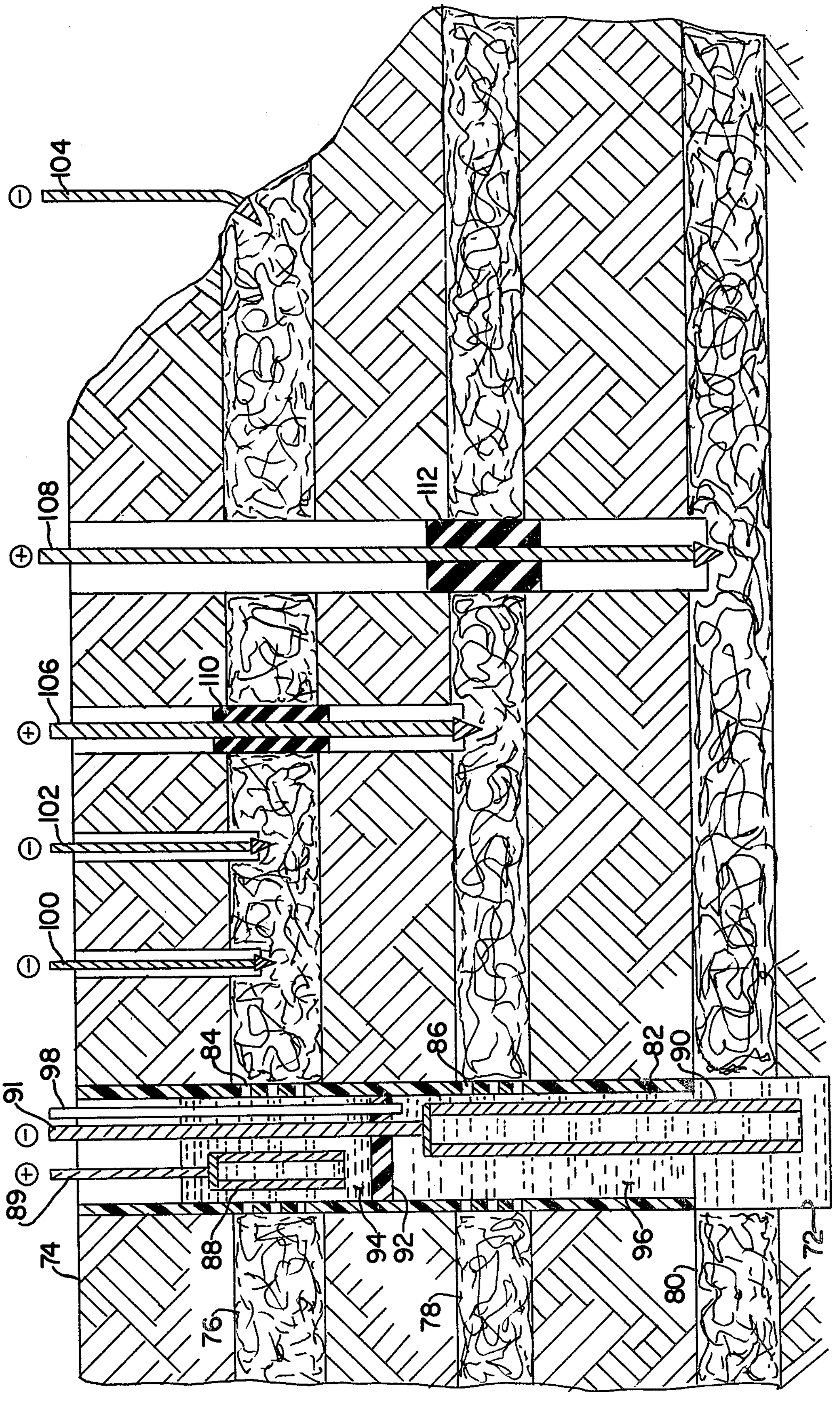

⊖ 104
⊕ 108
⊕ 106
⊖ 102
⊖ 100
⊖ 91
⊕ 89

FIG. 4

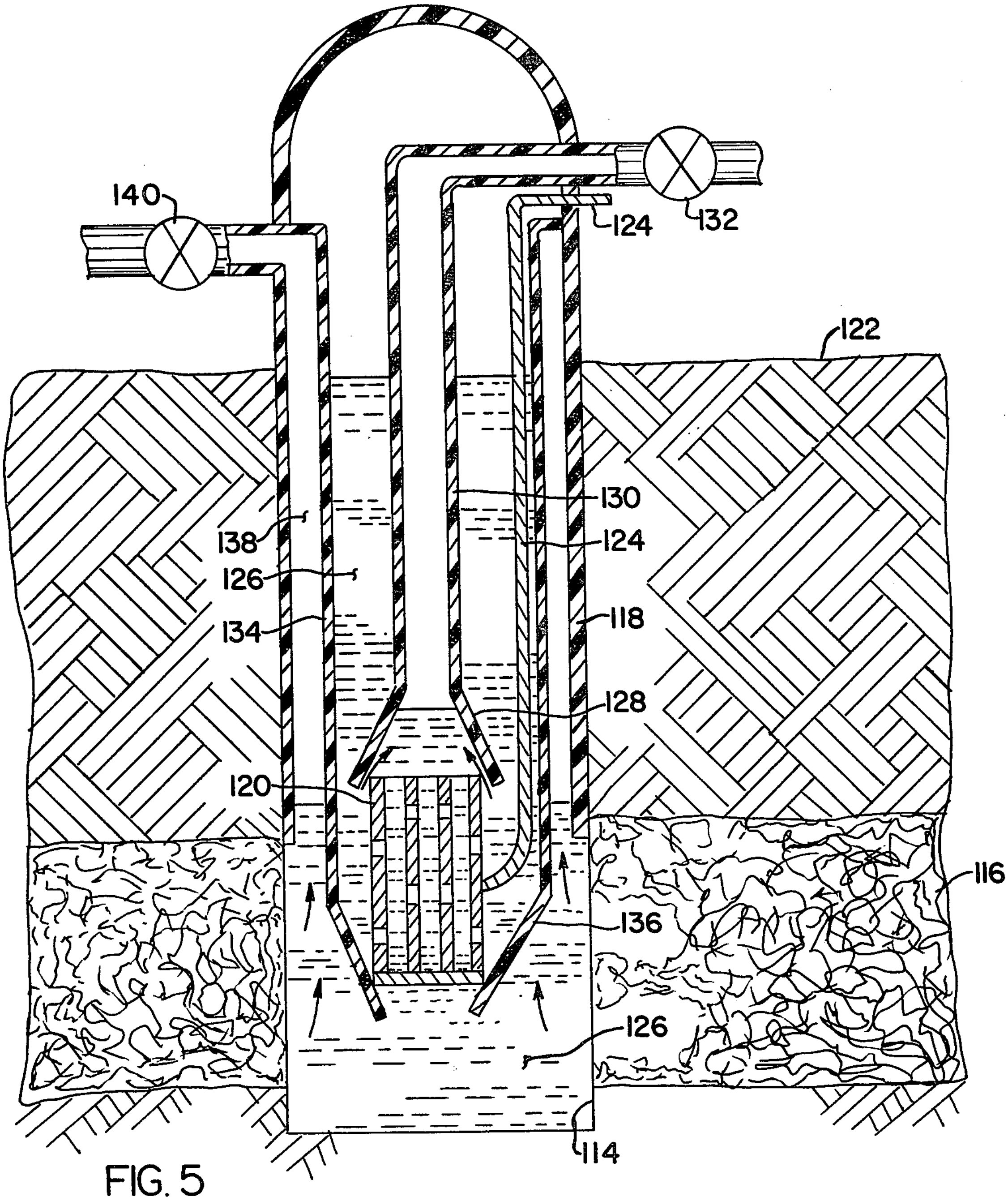
140
124
132
122
130
138
124
126
134
118
128
120
116
136
126
114

FIG. 5

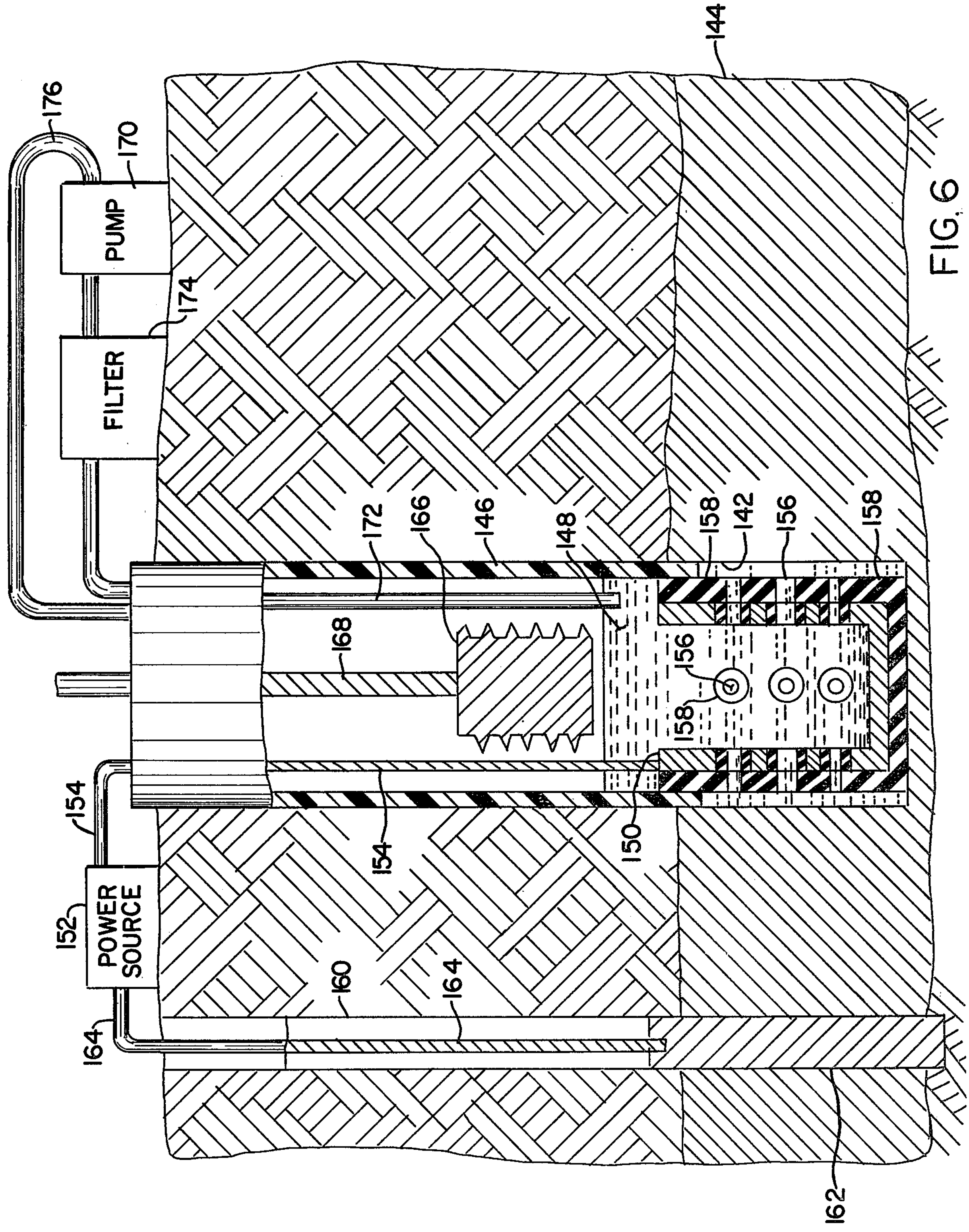
PUMP
FILTER
POWER SOURCE

FIG 6

EARTH BATTERY

DESCRIPTION

1. Field of the Invention

This invention pertains to earth batteries or electric wells and, more particularly, to a battery in which one electrode is a veined material located in the earth.

2. Background Art

Batteries which convert the chemical energy contained in their active materials directly into electrical energy by means of an oxidation-reduction electrochemical reaction are well-known in the art. Such batteries, in their most basic sense, include a pair of spaced apart electrodes and an electrolyte in the space between the electrodes. During the electrochemical reaction electrons are transferred from one electrode to the other and produce an electrical current when an external circuit is connected between the electrodes.

The electrode which gives up electrons during the reaction, and is oxidized in the process, is referred to as the anode. The other electrode, which is capable of accepting electrons and is an oxidizing material for the overall process, is referred to as the cathode. The electrolyte, generally an aqueous solution, provides the necessary internal ionic conductivity for the flow of electrons released at the anode. The electrolyte may also be a solid material.

The anode, cathode and the electrolyte may be made of a variety of materials, as is well known in the art. The only requirement for the anode and cathode is that one material must give up electrons at a faster rate than the other material, and thereby acquire a positive charge with respect to the other material. In one common battery arrangement, one electrode is made of carbon or copper, the other electrode is made of zinc and the electrolyte is sulphuric acid (H_2SO_4).

Various arrangements of these batteries, and the various combinations of materials which may be used for the electrodes and the electrolyte is discussed in Fink, *Standard Handbook for Electrical Engineers,* Eleventh Ed. 1978, pages 11-92 to 11-141, and is incorporated herein by reference.

It is also known in the art to use the earth as an electrolyte in a battery. Typically, two spaced electrodes are buried in the ground, and the ground, by itself or moistened with a chemical, functions as the electrolyte. See U.S. Pat. Nos. 155,209; 160,152; 182,802; 211,322; 329,724; 495,582; and 728,381. In the process disclosed in U.S. Pat. No. 3,278,335, an oil bearing formation forms a portion of an electrolyte and electrical energy is produced through the growth of bacteria in the formation. However, in each of these prior art patents, the electrodes are the typical zinc, carbon or copper plates or rods which are inserted into the earth, and the earth is used solely as an electrolyte.

It is an object of the present invention to provide a battery in which a portion of the earth is one electrode thereof.

It is a further object of the present invention to provide a battery in which one electrode is a veined material in the earth.

Another object of the present invention is to provide such a battery which is capable of generating electricity as well as capable of storing electrical energy.

It is yet a further object of the present invention to utilize such a battery in the mining of veined materials located in the earth.

SUMMARY OF THE INVENTION

Therefore, I have invented an electric battery which includes a first and second spaced electrode and an electrolyte in the space between the electrodes and in contact therewith, in which the first of the electrodes is a veined material located in the earth. In one embodiment, the second electrode is disposed in a hole which extends from the surface of the earth into the first electrode and is electrically connected to the surface of the earth, an electrolyte is disposed in the hole and in contact with the electrodes, and the first electrode is electrically connected to the surface of the earth. The battery may also include a casing disposed within the hole and surrounding the second electrode, such that the casing contains the electrolyte and permits contact of the electrolyte only with the first electrode but not with the surrounding earth. The first electrode is electrically connected to the earth's surface by way of one or more cables anchored thereto and located either within the hole or outside of the hole. The gases generated during the battery operation may be recovered and used as desired.

The first electrode is preferably a veined material located in the earth and may be a seam or vein of coal, graphite, magnetite, nickeline, sphalerite, arsenic, siderite, gold, silver, or copper.

I have also invented a method of mining metals located in the earth using the earth battery of the present invention. The method includes the steps of drilling a hole from the earth's surface into the metal to be mined, placing an electrolyte in the hole and in contact with the metal, positioning a recovery electrode within the hole and at least partially immersed in the electrolyte, making electrical contact with the metal, passing an electric current through the metal, electrolyte and recovery electrode such that the metal will be deposited on the recovery electrode, and removing the deposited metal from the recovery electrode.

I have also invented a method of drilling holes in the earth using the present invention. The method includes the steps of predrilling a small hole into the earth for a desired distance, placing a quantity of acid into the small hole such that the acid will seep into and soften the area around the small hole, and then drilling a hole through the area softened by the acid. The acid may be neutralized prior to the final drilling step.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a partial cross section taken through the earth and through one embodiment of an earth battery in accordance with the present invention;

FIG. **2** is a perspective view of one embodiment of the second electrode;

FIG. **3** is a cross section of the bottom portion of one embodiment of the casing;

FIG. **4** is a cross section taken through the earth showing several alternative arrangements of the earth battery;

FIG. **5** is a cross section taken through the earth showing one embodiment of a gas recovery mechanism for the earth battery; and

FIG. **6** is a partial cross section taken through the earth showing one embodiment of a metal recovery apparatus in accordance with the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

An earth battery in accordance with the present invention is shown, partially in cross section, in FIG. **1**. The earth battery includes a first electrode **10** which is a veined material located within the earth **12**. A second electrode **14** is disposed within a hole **16** which extends from the surface **18** of the earth **12** and into the first electrode **10**. The second electrode **14** is electrically connected to the surface **18** of the earth **12** via cable **20**. An electrolyte **22** is disposed in the hole **16**, is in contact with the first electrode **10** and the second electrode **14**, and fills the space in the hole **16** therebetween. The electrolyte **22** is shown in FIG. **1** as a fluid, but it is to be understood that a solid electrolyte may also be utilized.

The earth battery shown in FIG. **1** also includes a casing **24** disposed within the hole **16**, adjacent the earth **12** and the first electrode **10** and surrounding the second electrode **14**. The casing **24** is preferably a hollow tube made of plastic or other insulating material and is adapted to prevent the electrolyte **22** from seeping out into the surrounding earth **12**. As shown in FIG. **1**, the casing extends from the earth's surface **18** to the bottom **26** of the hole **16**, but does not cover said bottom **26**. The casing **24** includes a plurality of openings **28** in the area where the casing is adjacent the first electrode **10**. These openings **28** permit the electrolyte **22** to contact and flow into the first electrode **10** but not into the surrounding earth **12**. The hole **16** and the casing **24** are preferably of circular cross section.

The second electrode **14** may be constructed in a variety of shapes. As shown in FIG. **1**, the second electrode **14** includes inner and outer cylindrical sleeves **30**, **32** mounted to a top plate **34**. A plurality of openings **36** through the sleeves **30**, **32** permit the electrolyte **22** to flow completely around all the surfaces of the second electrode **14**. In order to prevent the second electrode from touching the first electrode **10** or the casing **24**, and thereby shorting out the entire battery, insulating spacers **38** may be mounted on the sides or on the bottom of the second electrode **14**.

The first electrode **10** is electrically connected to the surface **18** of the earth **12** by means of a cable which extends into the earth and is physically anchored to the first electrode. As shown in FIG. **1**, cable **40** may be located within hole **16** and anchored to the bottom **26** of the hole by a spike **42** or the like. Alternatively, cable **44** may be located outside of hole **16** in a smaller hole **46** which passes into the first electrode **10**. Cable **44** is similarly anchored with spike **42**. A plurality of cables, spaced around or within the hole **16**, may also be used.

The electrolyte **22** will seep into and cause stress fractures **48** in the first electrode **10**. Cable **44** located outside of hole **16** is preferably anchored within the extent of the electrolyte **22** seepage. Over a period of time, the area of electrolyte seepage, and hence the effective size of the first electrode **10** in the battery, will slowly increase. It can be appreciated that the earth battery of the present invention can be permitted to grow as large as desired by allowing the electrolyte **22** to continue to seep into the first electrode **10**. The casing **24** is preferably sealed at the top by a lid **50** or the like. The rate of seepage of the electrolyte **22** into the first electrode may be regulated by a vacuum pump **52** connected to the inside of the casing **24**.

Various gases will be generated by the earth battery, both from the electrochemical reaction involved in the battery itself and the gases released from the first electrode **10** as the stress fractures **48** are formed. For example, if the first electrode **10** is a seam of coal, methane will be released, and if the second electrode is made of zinc and the electrolyte is sulphuric acid, oxygen and hydrogen will be generated during the electrochemical reaction of the battery. The casing lid **50** may include a pressure relief valve **54** to release these generated gases built up in the casing **24** when the pressure exceeds a predetermined value.

The earth battery described above operates in the exact same manner as a conventional battery and generates electrical energy from the electrochemical reaction of the electrodes and the electrolyte. The first electrode **10** is selected to be a veined material located in the earth and may include seams or layers of such materials as coal, graphite, magnetite, nickeline, sphalerite, arsenic, siderite, gold, silver and copper. The second electrode is selected from a material suitable for use with the particular veined material used as is well known in the art. A suitable electrolyte is selected to function with the electrodes chosen as is well known in the art.

In a preferred embodiment, a seam of coal is used as the first electrode, the second electrode is made of zinc, and sulphuric acid is used as the electrolyte. As an alternative, pickling acid, which is usually a waste by-product of a metal manufacturing plant, can be utilized as the electrolyte. Furthermore, a solid electrolyte, such as ammonium chloride (NH_4Cl) or zinc chloride ($ZnCl$) may be used. The use of an acidic electrolyte should not cause any environmental problems if the seam of coal, or other veined material, is located beneath the surrounding water table level. If pollution of the water table is of concern, an alkaline electrolyte, such as potassium hydroxide, may be used.

The earth battery described above may be used as a source of electrical energy at a variety of locations which have underground veins of suitable electrode material. The earth battery has the advantages that it is located primarily underground and out of the way, the functioning portions remain at a constant temperature, and the battery has the potential of growing to a rather large size. The battery of the present invention may be used as an electrical energy source, but may also be used to store electrical energy from a variety of sources such as solar cells, windmills, lightning, or the excess capacity from electrical utility plants.

An alternate embodiment of a second electrode **56** is shown in FIG. **2** and includes four rectangular plates **58** joined together along one edge to form an X-shaped structure. While only four plates **58** are shown in FIG. **2**, more or fewer plates may be utilized as desired. Second electrode **56** also includes a plurality of glass wheels **60**, each of which are mounted to the plates **58** by a bracket **62**. The glass wheels **60** are insulating and function to support and space second electrode **56** from the casing and the first electrode in a similar manner to the spacers **38** discussed above in connection with FIG. **1**.

An alternate arrangement for the casing is shown in FIG. **3**. Casing **64** shown in FIG. **3** is similar to casing **24** discussed above but further includes a bottom casing plate **66** which seals casing **64** and prevents the electrolyte from flowing downwardly. Openings **68** in casing **64** vary in size inversely with distance of each opening from the surface of the earth. The smallest opening is located nearest the bottom plate 66 and openings 68 gradually increase in size as the distance from the bottom plate increases. Since the pressure of the electrolyte fluid increases toward the bottom plate 66 from the weight of the fluid above, the rate of flow of the electrolyte through openings 68 and into the first electrode will remain constant.

A plurality of sharp protrusions 70 are located along the inner surface of casing 64 near openings 68 and on the bottom plate 66. When the earth battery of the present invention is no longer in use, it may be desirable to neutralize the acidic electrolyte liquid used. Baking soda, or any other acid neutralizing substance, may be placed into a glass or other breakable container and dropped into the casing. The breakable container will strike the sharp protrusions 70, break apart, and release the acid neutralizing substance into the electrolyte.

The earth battery of the present invention may utilize more than one veined material in the earth as electrodes and the materials may be the same or may be different from each other. In the configuration shown in FIG. 4, hole 72 passes from the earth's surface 74 through three veins of material, referred to as first electrode 76, first electrode 78 and first electrode 80. Casing 82 is disposed within hole 72 and extends to the top of first electrode 80. Openings 84 extend through casing 82 into first electrode 78 and openings 86 extend through casing 82 into first electrode 78. Second electrode 88 is disposed within casing 82 adjacent first electrode 76 and is connected to the earth's surface 74 via cable 89. Second electrode 90 is disposed within casing 82 adjacent both first electrode 78 and first electrode 80 and is connected to the earth's surface 74 via cable 91. As shown in FIG. 4, second electrode 90 is a single electrode which reacts electrochemically with both first electrode 78 and first electrode 80.

In the arrangement shown in FIG. 4, first electrode 76 is a negative material with respect to both first electrode 78 and first electrode 80. To prevent a shorting out of the earth battery, the interior of casing 82 is divided by a plug 92. In this manner, one earth battery is formed by first electrode 76, second electrode 88, and electrolyte 94, and a second, separate earth battery is formed by first electrode 78, first electrode 80, second electrode 90, and electrolyte 96. To replenish the supply of electrolyte 96, a pipe 98 extends from the earth's surface 74 through the plug 92.

Since hole 72 passes completely through first electrode 76 and first electrode 78, electrical connection between these electrodes and the earth's surface 74 cannot be made within hole 72. Cables 100 and 102 extend down and are attached to first electrode 76. While only two cables 100 and 102 are shown, it is to be understood that fewer or more cables may be attached to one such electrode. As shown in FIG. 4, first electrode 76 is exposed to the earth's surface, such as on a hill, and may be electrically contacted directly by cable 104 without the necessity of passing through the earth. First electrode 78 is electrically connected to the earth's surface by cable 106 and first electrode 80 is electrically connected to the earth's surface by cable 108. To avoid short circuits, the cables should not electrically contact the other veins of material through which they pass. Cable 106 is shown surrounded by heavy insulation 110 in the vicinity of first electrode 76. Cable 108 is shown physically spaced from first electrode 76 and surrounded by heavy insulation 112 in the vicinity of first electrode 78. Ideally, the entire length of each cable would be encased in an insulating material, except where electrical contact is made at the ends.

The relative electrical potential of each electrode shown in FIG. 4 is shown as being either positive or negative. The various electrodes could be operated as separate batteries, or could be connected in parallel or in series as is desired.

As an example of the arrangement shown in FIG. 4, first electrodes 78 and 80 could each be a seam of coal, which is considered to be a relatively positive material. Second electrode 90 may be made of zinc, a negative material when compared with coal. First electrode 76 may be a seam of magnetite (Fe_3S_4), nickeline (NiAs), or sphalerite (ZnS), each considered to be a negative material. Second electrode 88 may be made of carbon, a positive material when compared with the magnetite, nickeline or sphalerite. The particular arrangement used will be dictated solely by the various veined materials actually encountered within the earth.

An arrangement of an earth battery of the present invention which includes apparatus for collecting the gases generated during battery operation is shown in FIG. 5. The earth battery includes hole 114, passing through first electrode 116, and casing 118 which extends adjacent hole 114 as described above. Second electrode 120 is identical to the second electrode 14 shown in FIG. 1, except that it is turned upside down, and is electrically connected to the earth's surface 122 by cable 124. Electrolyte 126 is located in hole 114 between first electrode 116 and second electrode 120.

A funnel-shaped trap 128 is disposed above second electrode 120 to collect the gases generated therein. The trap 128 is connected to pipe 130 which carries the gas through valve 132 and then out of casing 118. Sleeve 134 is disposed within and spaced from casing 118 and terminates near the bottom of hole 114 in an inwardly directed, cone-shaped deflector 136. Deflector 136 is open to allow electrolyte 126 to flow between the interior and exterior of sleeve 134. Gases which are either generated by or are released from first electrode 116 are directed by deflector 136 into the space 138 between sleeve 134 and casing 118, and then out of casing 118 through valve 140.

If valves 132 and 140 were open, the gases generated would continuously escape. If valves 132 and 140 were closed, the pressure of the generated gases would increase, and eventually force electrolyte 126 up into the area between pipe 130 and sleeve 134, as is shown in FIG. 5. When the valves are opened, electrolyte 126 will drop back down to its original level and aid in expelling the generated gases.

The particular gases generated will depend on the exact materials involved. If the first electrode were coal, the second electrode zinc, and the electrolyte sulphuric acid, then oxygen would be generated at the second electrode, hydrogen would be generated at the first electrode, and methane would be released from the first electrode due to the fracturing of the coal from the electrolyte. The recovered gases could be used as is or could be burned to generate electric energy.

The earth battery of this invention can also be easily used to recover metals located in the earth. If a metal, such as gold, copper, or silver, is used as the first electrode, and a direct current is passed through the battery, the metal will be removed from the first electrode and plated onto the second electrode. By withdrawing the second electrode from the battery and removing the plated metal, the mining operation is completed.

Any of the earth battery arrangements discussed above could be utilized in the mining of such a metal. FIG. **6** shows an arrangement particularly adapted for mining a vein of metal. Hole **142** extends into the vein of metal **144** to be mined, casing **146** extends down to the top of the vein of metal, and electrolyte **148** is deposited in hole **142**. A recovery electrode **150** is placed into hole **142** and is at least partially immersed in electrolyte **148**. The recovery electrode **150** is connected to a direct current power source **152** by cable **154**.

The recovery electrode **150** shown in FIG. **6** is cup-shaped, open at the top, and includes a plurality of openings **156** passing therethrough. The outside of the recovery electrode **150** and the surface of openings **156** are covered with insulation **158**. Only the inner surface of the recovery electrode **150** is in contact with electrolyte **148**, and will be plated with the metal **144**.

To make electrical connection between the power source **152** and the vein of metal **144**, hole **160** is passed through the vein of metal, a liquid, electrically conductive material **162** is poured into hole **160**, and cable **164** is connected between the power source **152** and material **162**, preferably before liquid material **162** has solidified. By passing a direct current through the vein of metal **144**, electrolyte **148** and recovery electrode **150**, the metal will be plated on the recovery electrode as is known in the art.

The plated metal may be recovered by withdrawing the recovery electrode **150** from hole **142** and thereafter removing the metal. Alternatively, the metal could be removed with the recovery electrode **150** in place within hole **142**. Shown in FIG. **6** is a threaded auger **166** which is as wide as the inner width of the recovery electrode. The auger **166** may be rotated via shaft **168** and lowered into the recovery electrode **150** to thereby scrape off any metal plated thereon. The metal will then be floating in electrolyte **148** in small pieces. By removing electrolyte **148** via pump **170** through an outlet pipe **172**, the metal may be recovered by passing electrolyte **148** through a filter **174**. The filtered electrolyte may then be returned to the battery via an inlet pipe **176**.

This method can be used to recover natural veined metals located in the earth or metals buried in the earth such as gold bars, coins, or the like, which are otherwise inaccessible to conventional mining techniques. This method is particularly useful for mining paper thin seams of gold which cannot be economically mined using conventional mining techniques.

When the earth battery is no longer in use and the casing, second electrode, and other materials are removed from the hole, the ground surrounding the hole will be relatively soft due to the action of the acidic electrolyte. If a larger hole were desired, such as to reach oil deposits beneath the vein of material previously used in the earth battery, such drilling could be accomplished relatively early because the ground is already softened. This method of drilling may be useful in passing a large hole through a particularly hard rock formation. By merely drilling a small hole through the rock formation, pouring in a quantity of an acid solution, and letting the rock be softened by the acid, a larger hole may be easily drilled through the formation. It may be desirable to neutralize the acid remaining in the rock before the final drilling step. This may be done by pouring in an acid neutralizing solution, such as a baking soda solution.

Having described presently the preferred embodiments of this invention, it is to be understood that it may be otherwise embodied within the scope of the appended claims.

I claim:

1. In an electric battery including a pair of spaced electrodes and an electrolyte in the space between the electrodes and in contact therewith, the improvement comprising: one of said electrodes being a solid veined material located in the earth.

2. The battery of claim **1** wherein the veined material is selected from the group consisting of coal, graphite, magnetite, nickeline, sphalerite, arsenic, siderite, gold, silver, or copper.

3. The battery of claim **1** wherein the veined material is a seam of coal.

4. An earth battery comprising:

(a) a first electrode which is a solid veined material located in the earth;

(b) a second electrode electrically connected to the surface of the earth and disposed in a hole which extends from the surface of the earth into the first electrode;

(c) an electrolyte disposed in said hole and in contact with said first and second electrodes, and

(d) means for electrically connecting said first electrode to the surface of the earth.

5. The earth battery of claim **4** further including a casing disposed within the hole and surrounding the second electrode, wherein said casing contains the electrolyte and permits contact of the electrolyte with the first electrode but not with the surrounding earth.

6. The earth battery of claim **5** wherein the casing is a plastic tube which includes a plurality of openings in the area where the tube is adjacent the first electrode.

7. The earth battery of claim **6** wherein the electrolyte is fluid and wherein the openings vary in size inversely with their distance from the surface of the earth, such that the flow of electrolyte into the first electrode is at a uniform rate.

8. The earth battery of claim **4** wherein the means for electrically connecting the first electrode to the surface of the earth is a cable anchored to the first electrode.

9. The earth battery of claim **8** wherein the cable is anchored to the first electrode inside of said hole.

10. The earth battery of claim **8** wherein the cable is anchored to the first electrode outside of said hole.

11. The earth battery of claim **4** wherein the means for electrically connecting the first electrode to the surface of the earth is a plurality of cables anchored to the first electrode.

12. The earth battery of claim **11** wherein said plurality of cables are located outside of said hole.

13. The earth battery of claim **4** wherein the first electrode further includes one or more additional and separate veins of material located in the earth.

14. The earth battery of claim **4** further including means for recovering the gases generated during use of the battery.

15. The earth battery of claim **14** wherein the means for recovering gases includes a funnel-shaped trap located above the second electrode to collect the gases generated by the second electrode and wherein the means for recovering gases further includes a sleeve disposed within but narrower than the hole and adapted to direct the gases released by the first electrode into the space between the hole and the sleeve.

16. The earth battery of claim **4** further including a regulating means for controlling the flow of electrolyte into the first electrode.

17. The earth battery of claim **16** wherein the regulating means is a vacuum pump.

18. The earth battery of claim **4** wherein the first electrode is a vein of material selected from the group consisting of coal, graphite magnetite, nickeline, sphalerite, arsenic, siderite, gold, silver, or copper.

19. The earth battery of claim **18** wherein the first electrode is a seam of coal.

20. An earth battery comprising:

(a) a vein of coal located in the earth;

(b) an electrode disposed in a hole which extends through the vein of coal;

(c) a liquid electrolyte located in the space between the coal and the electrode;

(d) a casing disposed within the hole and surrounding the electrode, said casing directing the liquid electrolyte into the coal; and

(e) an electric cable anchored to the coal and extending to the earth's surface.

21. The earth battery of claim **20** wherein the electrode disposed in the hole is made of zinc and the electrolyte is an acid solution.

* * * * *

W. D. SNOW

Earth-Batteries for Generating Electricity.

No. 155,209. Patented Sept. 22, 1874.

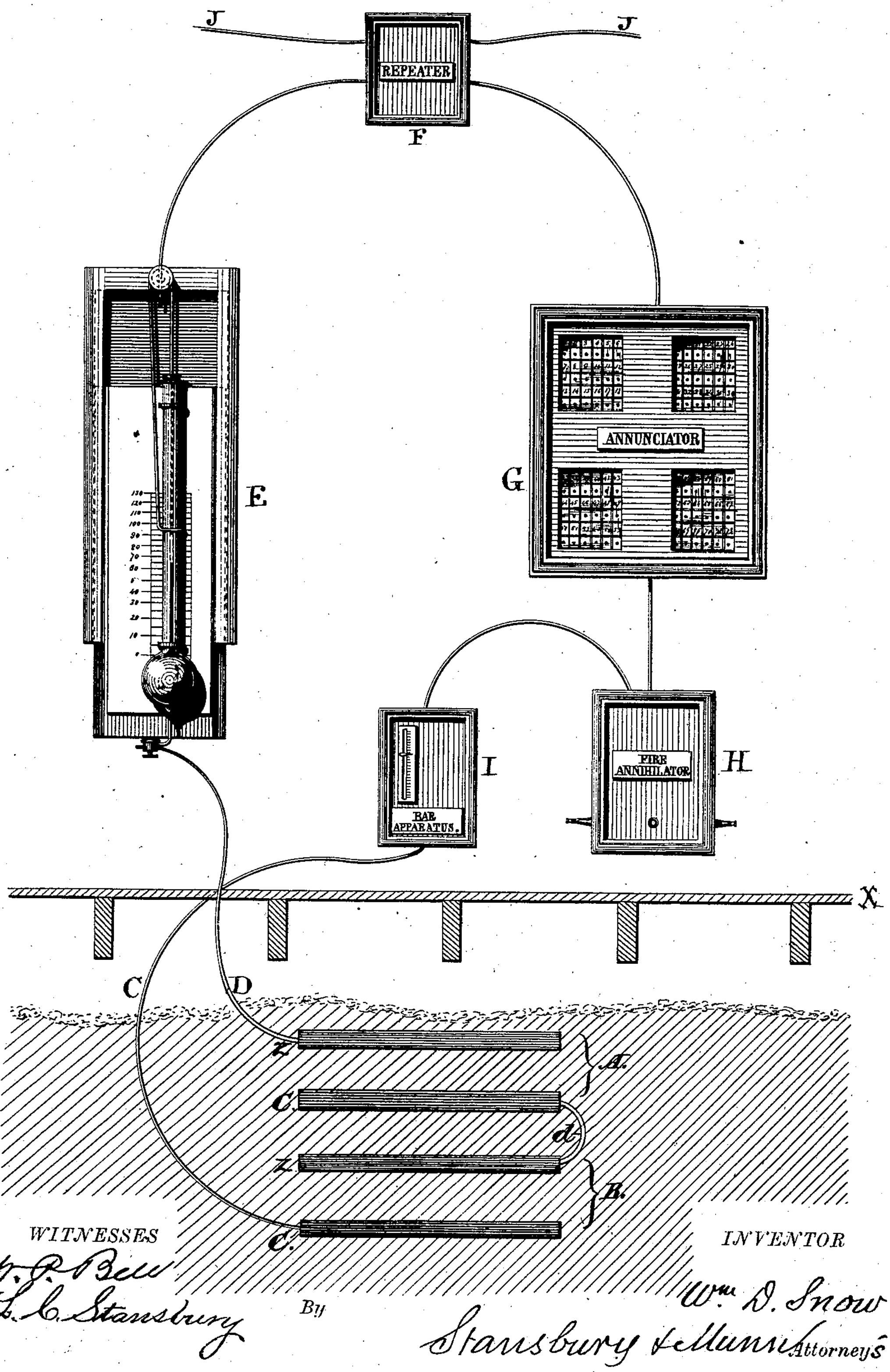
J
J
REPEATER
F
E
G
ANNUNCIATOR
I
BAR APPARATUS.
H
FIRE ANNIHILATOR
X
C
D
z
c
A
d
z
B
c

WITNESSES

W. P. Bell
L. C. Stansbury

INVENTOR

Wm. D. Snow

By

Stansbury & Munn *Attorneys*

THE GRAPHIC CO. PHOTO-LITH. 39 & 41 PARK PLACE, N.Y.

UNITED STATES PATENT OFFICE.

WILLIAM D. SNOW, OF BROOKLYN, NEW YORK.

IMPROVEMENT IN EARTH-BATTERIES FOR GENERATING ELECTRICITY.

Specification forming part of Letters Patent No. **155,209**, dated September 22, 1874; application filed June 6, 1874.

To all whom it may concern:

Be it known that I, WILLIAM D. SNOW, of Brooklyn, in the State of New York, have invented a certain Improvement in Earth-Batteries for Electro-Magnetic Alarms, &c.; and I do hereby declare the following to be a full and correct description of the same, reference being had to the accompanying drawing, which represents a series of electro-magnetic signaling or alarm apparatus placed in a house and operated by a current derived from an earth-battery, as hereinafter more fully set forth.

The object of my invention is to avoid the employment of artificial electric batteries for the purposes of house-signaling, fire and burglar alarms, and for the operation of other circuits.

In the common systems employed for these purposes a local or house battery is employed, requiring to be kept in working order by the occupants of the building in which it is placed, who are usually entirely ignorant of its structure and uses; or else the street line of wires is carried into the house, and thus subjected to all the contingencies to which the house itself is liable.

I obviate the trouble, annoyance, and danger to public safety incident to the systems referred to, by employing the current from an earth-battery without the intervention of any cups or vessels containing acids or other agents for the decomposition of zinc or other metals, commonly called a "battery." This I effect by making use of an earth-battery formed by burying the electrodes in the earth under the house at the depth of permanent moisture, so that the action of the dampness of the earth upon the plates will develop a sufficient permanent current for the required purposes. From these buried plates I run a telegraphic circuit in the house to connect with thermostats, signaling apparatus, fire and other alarms, annunciators, gas-lighters, or repeaters, according to the nature of the use to which the electric current is to be applied.

Some attempts have been made to use earth-batteries for various purposes. Such earth-batteries were made by simply burying a single set of elements in the earth, and proved of little or no practical value, on account of the very feeble intensity of the current generated. I have discovered, however, that if a series of elements be used in the earth, coupled together so as to form an earth-battery of any required size, this objection is obviated, and a constant current of considerable intensity is obtained, capable of operating steadily quite an amount of apparatus.

In the drawing, A and B mark two series of electrodes, which may be of zinc and carbon, as marked C and Z, buried in the earth under the house in which the circuit is to be used, the wire *d* being a connection between the dissimilar elements of the two series; C D, the wires for conducting the current, which may be made to complete the circuit by passing through E, an automatic thermostat; F, a repeater, connected with the street-wires J J; G, an annunciator; H, a fire-annihilator; or I, a signal-writing-bar apparatus, either one or more of these, each operated by the current in any of the usual modes.

What I claim is—

1. The improved earth-battery, consisting of a series of elements buried in the earth and connected together to form a battery, substantially as and for the purposes set forth.

2. The combination of two or more electric currents or circuits, when one of these is a current derived from an earth-battery, substantially as described, and is employed to operate the other or others.

The above specification of my said invention signed and witnessed at Washington this 30th day of May, A. D. 1874.

W. D. SNOW.

Witnesses:
W. P. BELL,
CHAS. F. STANSBURY.

US 20150280518A1

(19) **United States**

(12) **Patent Application Publication**

Stupac

(10) **Pub. No.: US 2015/0280518 A1**

(43) **Pub. Date: Oct. 1, 2015**

(54) **GENERATING ELECTRICITY FROM THE EARTH**

(71) Applicant: **Omer Stupac**, Utica, NY (US)

(72) Inventor: **Omer Stupac**, Utica, NY (US)

(21) Appl. No.: **14/229,873**

(22) Filed: **Mar. 29, 2014**

Publication Classification

(51) **Int. Cl.**
H02K 7/18 (2006.01)

(52) **U.S. Cl.**
CPC *H02K 7/18* (2013.01)

(57) **ABSTRACT**

A system for generating electricity from the geomagnetic field and rotation of the earth is presented. The earth rotates through the geomagnetic field to form a potential difference between first and second terrestrial charged bodies. The first and second charged bodies are spaced apart in a direction substantially normal to the earth's surface. The system comprises an electrical load, a first current path, and a second current path. The electrical load has a supply input and a ground output. The first current path is coupled to the first charged body and to the supply input of the electrical load, for supplying the potential difference to the electrical load. The second current path is coupled to the ground output of the electrical load and to the second charged body. As a result of this arrangement, an electric current related to the potential difference is generated in the electrical load.

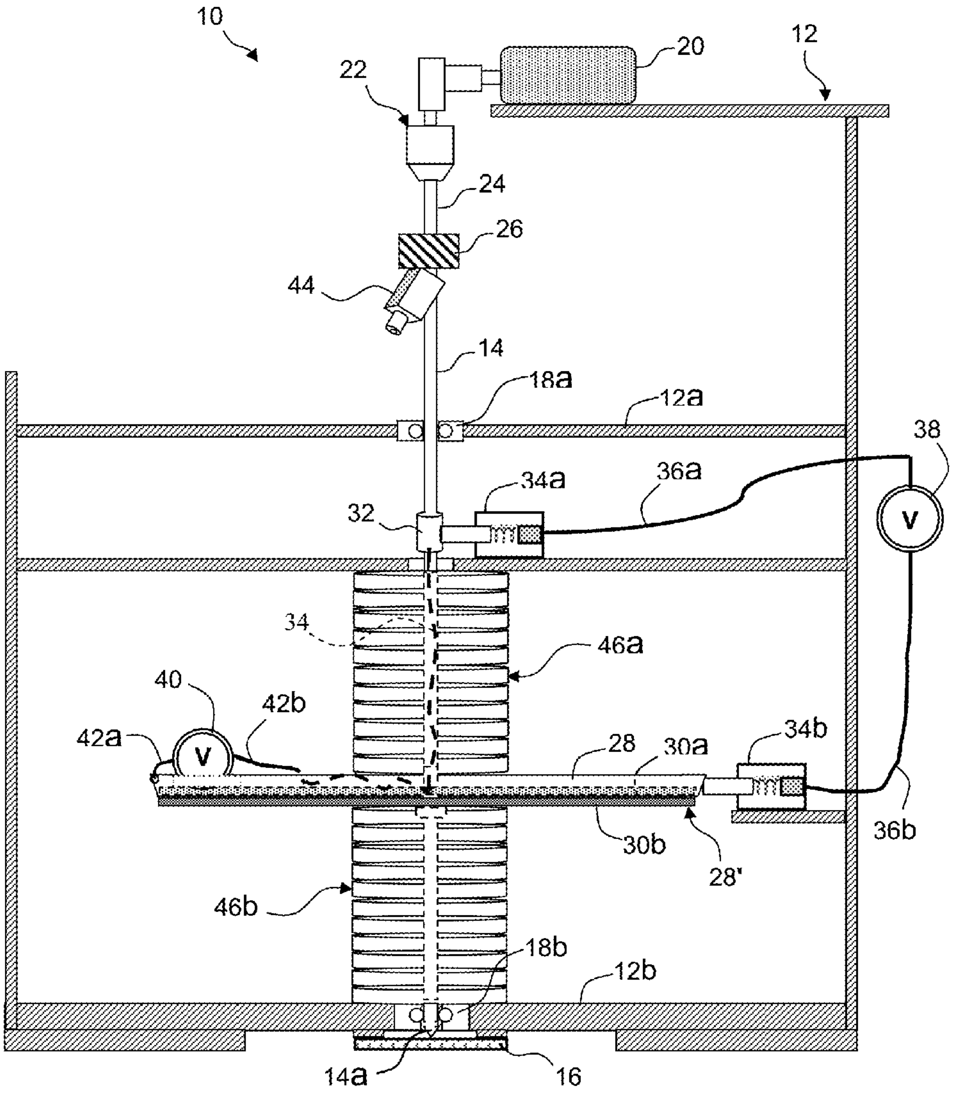

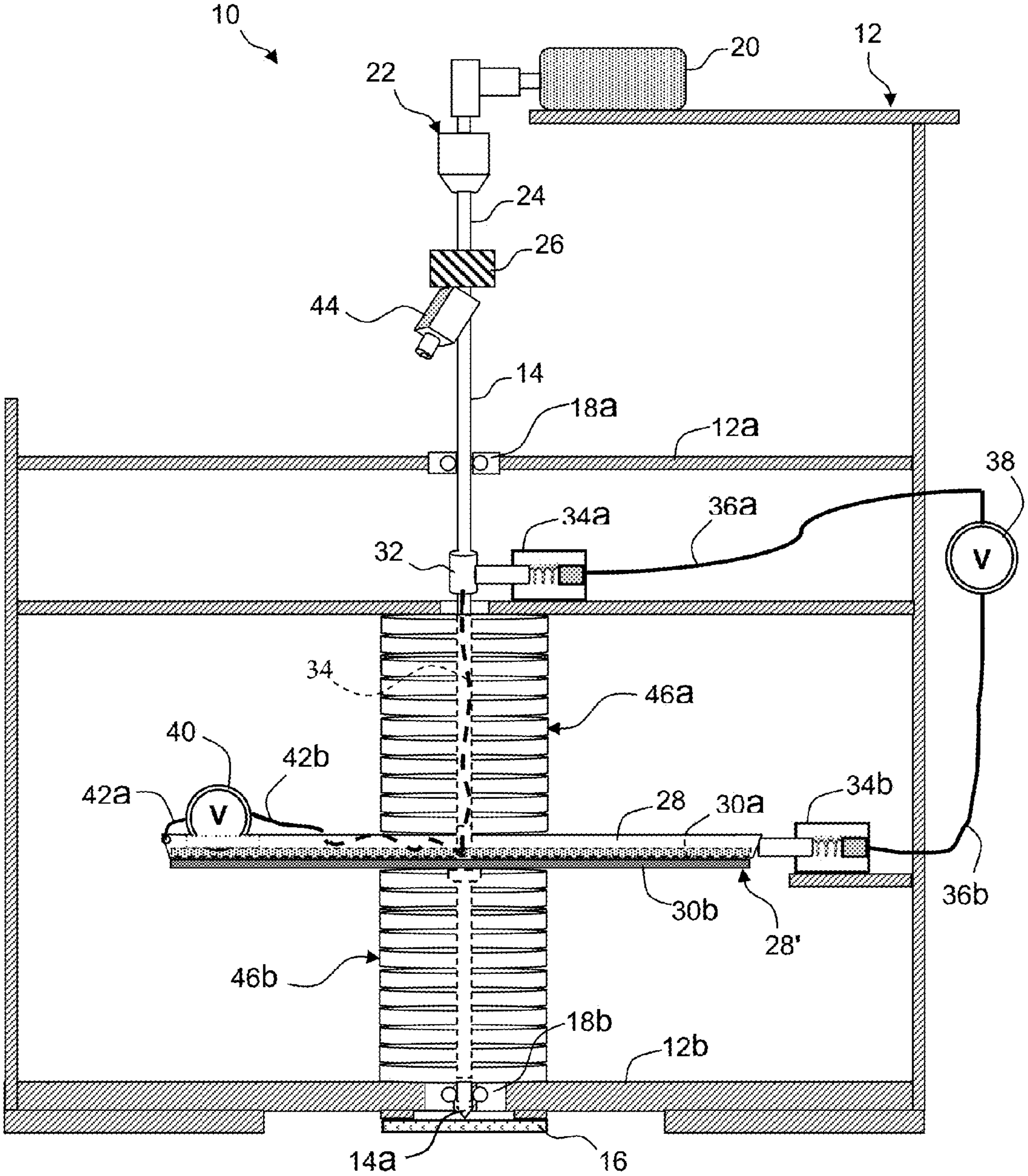
V
V

FIG. 1

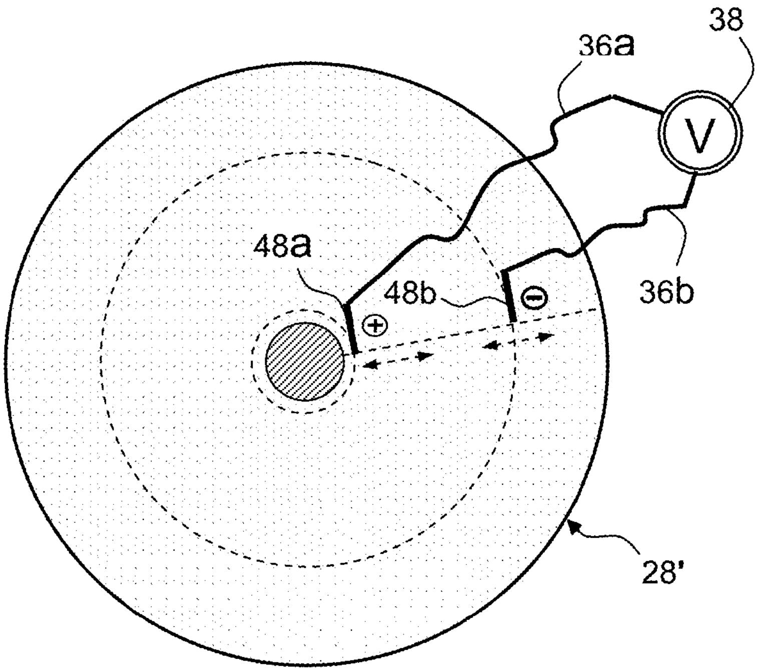
38
36a
V
48a
48b
⊕
⊖
36b
28'

FIG. 2

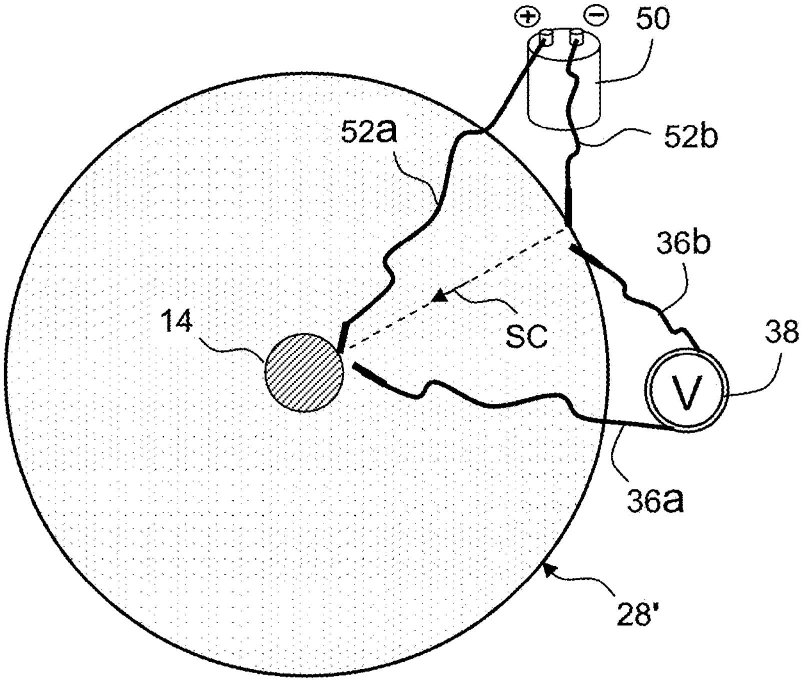
⊕
⊖
50
52a
52b
36b
14
SC
38
V
36a
28'

FIG. 3

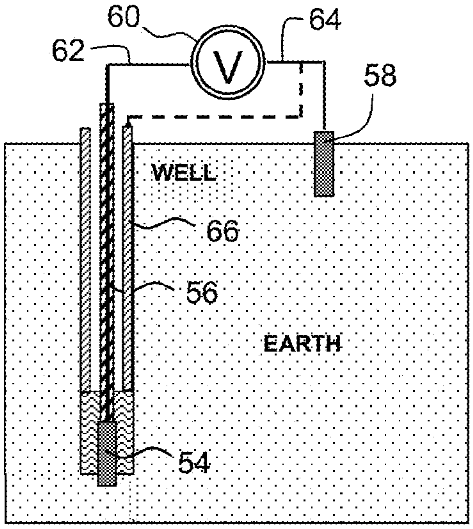
60
64
62
58
WELL
66
56
EARTH
54

FIG. 4

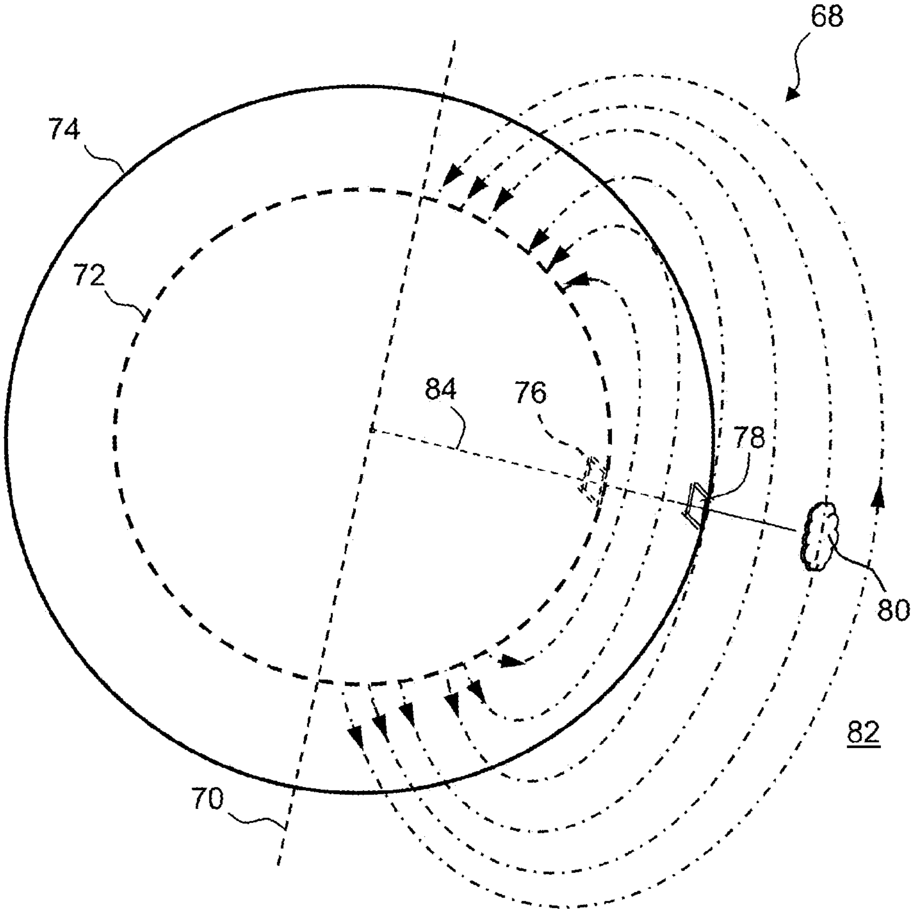
68
74
72
84
76
78
80
82
70

FIG. 5

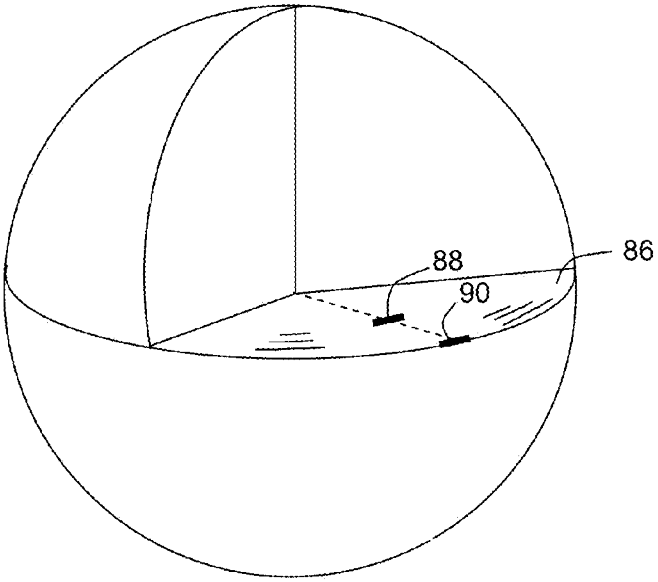

FIG. 6

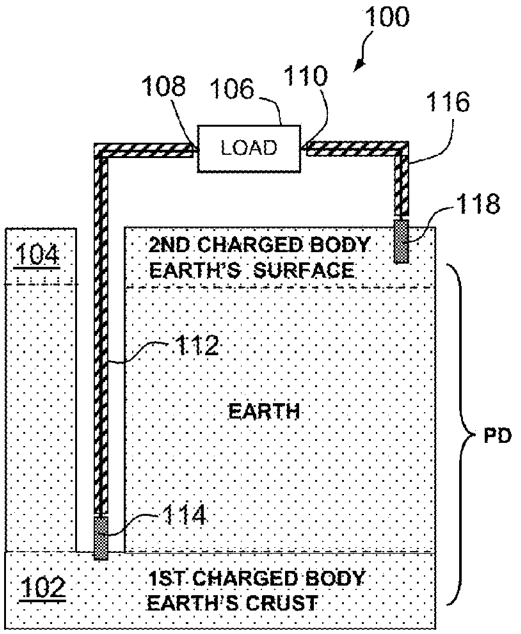
100
108
106
110
116
LOAD
118
104
2ND CHARGED BODY
EARTH'S SURFACE
112
EARTH
PD
114
102
1ST CHARGED BODY
EARTH'S CRUST

FIG. 7

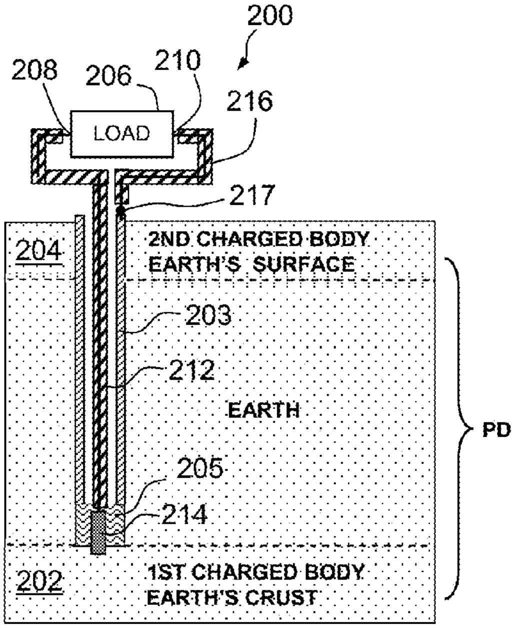
200
208
206
210
216
LOAD
217
204
2ND CHARGED BODY
EARTH'S SURFACE
203
212
EARTH
PD
205
214
202
1ST CHARGED BODY
EARTH'S CRUST

FIG. 8

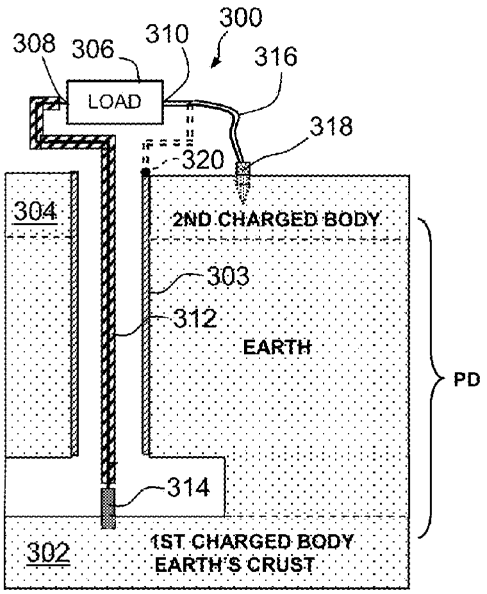
300
308
306
310
316
LOAD
318
320
304
2ND CHARGED BODY
303
312
EARTH
PD
314
302
1ST CHARGED BODY
EARTH'S CRUST

FIG. 9

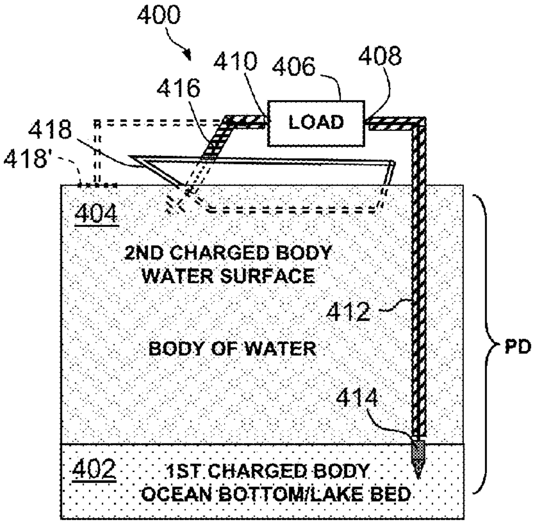
400
410
406
408
416
LOAD
418
418'
404
2ND CHARGED BODY
WATER SURFACE
412
BODY OF WATER
PD
414
402
1ST CHARGED BODY
OCEAN BOTTOM/LAKE BED

FIG. 10

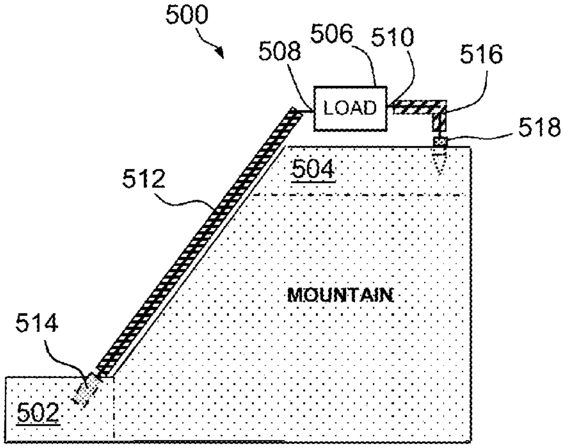
500
508
506
510
516
LOAD
518
504
512
MOUNTAIN
514
502

FIG. 11

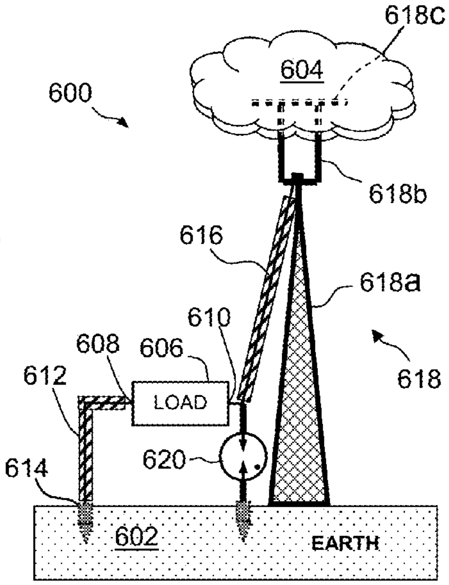
618C
604
600
618b
616
618a
610
608
606
612
LOAD
618
614
620
602
EARTH

FIG. 12

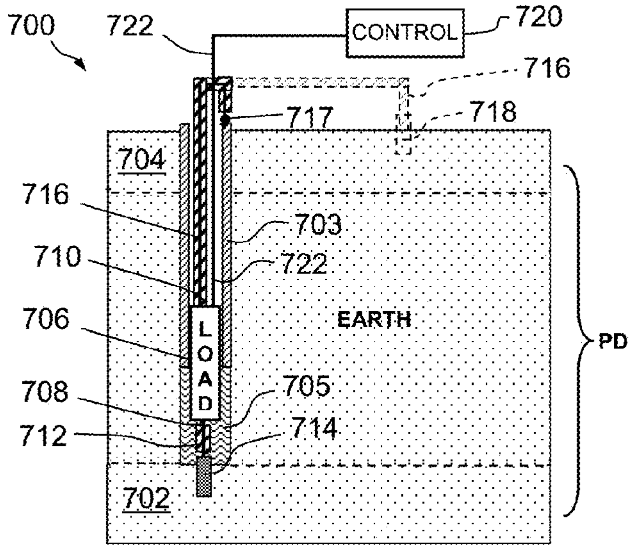
700
722
CONTROL
720
716
717
718
704
716
703
710
722
EARTH
706
L O A D
PD
708
705
712
714
702

FIG. 13

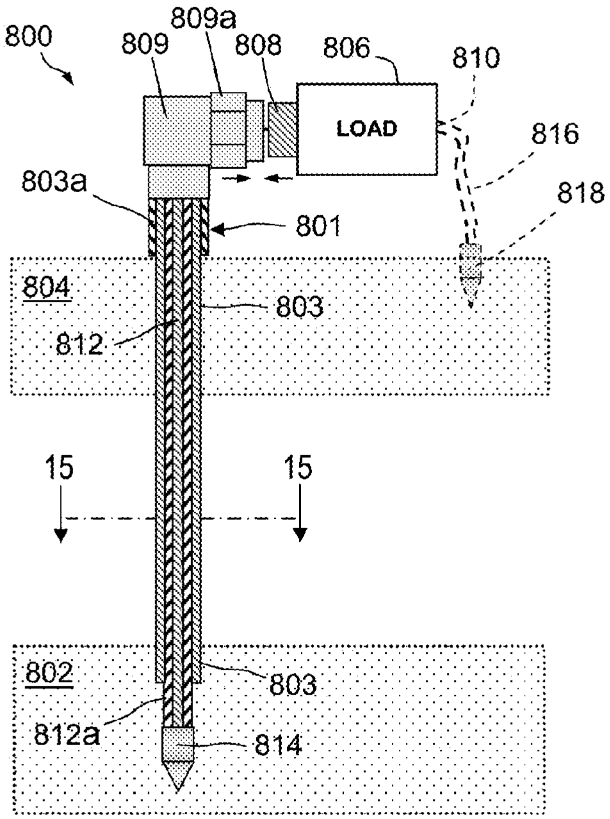
800
809
809a
808
806
LOAD
810
816
818
803a
801
804
803
812
15
15
802
803
812a
814

FIG. 14

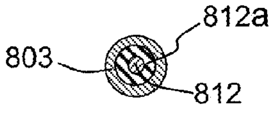
812a
803
812

FIG. 15

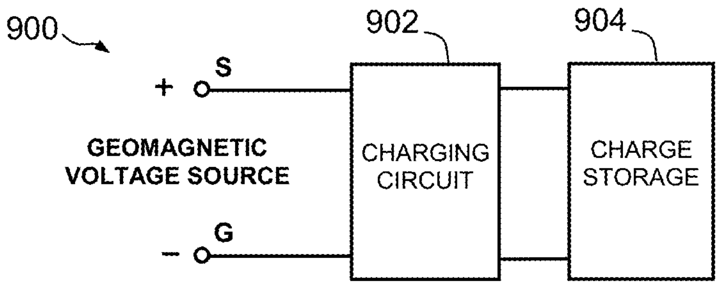
900
902
904
+ S
GEOMAGNETIC VOLTAGE SOURCE
− G
CHARGING CIRCUIT
CHARGE STORAGE

FIG. 16

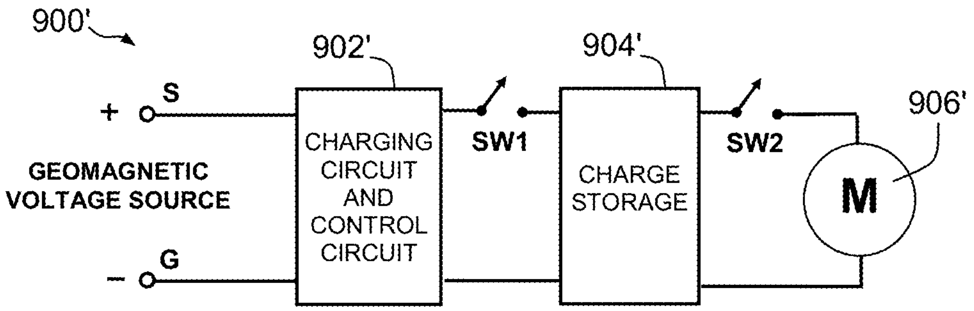
900'
902'
904'
906'
+ S
GEOMAGNETIC VOLTAGE SOURCE
− G
CHARGING CIRCUIT AND CONTROL CIRCUIT
SW1
CHARGE STORAGE
SW2
M

FIG. 17

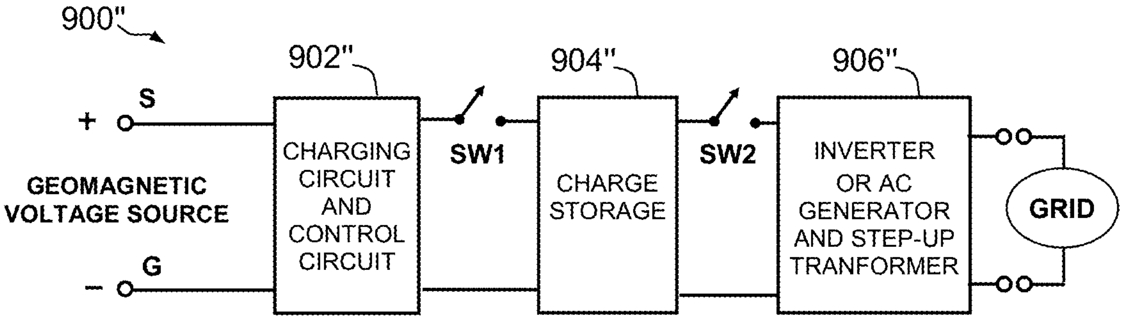
900"
902"
904"
906"
+ S
GEOMAGNETIC VOLTAGE SOURCE
− G
CHARGING CIRCUIT AND CONTROL CIRCUIT
SW1
CHARGE STORAGE
SW2
INVERTER OR AC GENERATOR AND STEP-UP TRANFORMER
GRID

FIG. 18

GENERATING ELECTRICITY FROM THE EARTH

BACKGROUND OF THE INVENTION

[0001] 1. Field of Invention

[0002] The present invention relates generally to renewable electrical energy sources, and more particularly to apparatus and methods for generating electricity from the earth's magnetic field and rotation.

[0003] 2. Background Art

[0004] The concepts of electromagnetic induction, Faraday's law, Faraday's disk, and the Lorentz force on charges are all well-known. These concepts have been applied to the earth, which is modeled as a large conductive sphere rotating through its own magnetic field ("geomagnetic field"). In this regard, the earth has been compared to a Faraday disk or homopolar generator. See, for example: W. M. Elsasser, The Earth as a Dynamo, Scientific American, Inc., Vol. 198, Issue 5, May 1, 1958, p. 44, 45; and C. R. Carrigan and D. Gubbins, The Source of the Earth's Magnetic Field, Scientific American, Inc., Vol. 240, Issue 2, Feb. 1, 1979, p. 118, 199-20. In making this comparison, the assumption has been carried over that the electricity to be obtained from the earth is from electromagnetic induction. Laboring under this assumption, it has been asserted that electricity from the geomagnetic field is not practicably retrievable, because the electricity gathering circuit would have to be stationary relative to the rotating earth (i.e., have a stationary frame of reference). This assertion is apparently based on a theory that the currents induced in the Faraday disk (or earth) and in the circuit employed to retrieve the electricity, would cancel each other out if both were moving through the magnetic field. Another theory is that a static electric field is created in the Faraday disk (or earth) that neutralizes the induced electric field, resulting in a zero voltage reading in the rotational frame of reference.

[0005] Further to this point, U.S. Patent App. Pub. No. 2004/0027022 (Feb. 12, 2004) to Weir, states that no voltage is generated by a Faraday type homopolar dynamo in the rotational reference frame of the dynamo (Weir, Par. [0010]). Weir even attempted to measure such voltage by mounting a voltmeter onto the rotating disk and connecting one probe to a shaft rotating the disk and the other probe to the rim of the disk. Weir states that the voltmeter read zero volts when the disk was rotating. (Weir Par. [0011].) No details are given regarding the sensitivity of the voltmeter or the test setup. In a book by Thomas Valone, Ph.D, P.E., entitled The Homopolar Handbook, published by Integrity Research Institute, Washington, D.C., 3rd Edition, October 2001 ("Valone I"), it is stated that an electrostatic field (created in response to the Lorentz force) neutralizes the electric field induced on the Faraday disk ("effective electric field"), and thus no voltage can be detected with a meter that rotates with the disk (Valone I, p. 7 & FIG. **8**). Valone discusses his voltage measurement on page 24 and again at page 63. He indicates that his voltmeter was set to indicate a voltage reading only if the voltage was greater than 15 millivolts. Thus, the sensitivity of Valone's experiment may have been limited to 15 millivolts, presumably because the induced electric field (obtained in the stationary reference frame) was expected to be 100 millivolts (see Valone I, p. 63).

[0006] Valone reminds us that Michael Faraday, in 1831, compared his Faraday disk to the earth and the earth's rotation through its magnetic field. Valone states that Faraday, thinking that an induced voltage could be measured on his rotating disk, attempted to measure a like voltage in the earth (i.e., in rivers and streams) and apparently was unsuccessful. (Valone I, p. 61.) Valone characterizes Faraday's thinking as a mistake. Valone presumably mentions Faraday's unsuccessful attempt at measuring voltage in the earth as support for his position that a voltage cannot be measured in the rotating reference frame. In an earlier paper, entitled "The Homopolar Generator: Tesla's Contribution," Proceedings of the International Tesla Symposium, January 1986 ("Valone II"), Valone specifically stated that "we can't draw power from the earth's Homopolar generator while rotating with it." (Valone II, p. 273.)

[0007] Notwithstanding the view that a voltage is unobtainable in the rotating frame of reference in a Faraday system, proposals were made in the 1980s for a space shuttle to drag a tether (an insulated conductor) through space across the geomagnetic field to generate electric power onboard the shuttle. Current and a voltage potential difference would be induced on the tether's conductor as it passed through the magnetic field lines of the earth. The current return path (or ground path) would be the charged plasma medium of the ionosphere, which would exist between the outer wall of the space shuttle and the distal end of the tether (or an outer wall of a satellite attached to the distal end). A load (e.g., electronic equipment) inside the space shuttle would be connected between the proximal end of the tether and the outer wall of the shuttle (i.e., the return or ground path). A tether experiment was actually performed on a space shuttle mission in 1996, according to a NASA article (www.istp.gsfc.nasa.gov/Education/wtether.html). The voltage generated on the tether was 3500 volts, apparently measured in the rotational frame of reference. It is suggested that the voltage may have been measured at the distal end of the tether with instruments aboard a satellite connected to the distal end. The voltage level at the load inside the shuttle is not mentioned. In this experiment, the voltage applied to the intended load (equipment aboard the shuttle) was supplied by the tether's conductor and the return or ground path for the load was, according to NASA, the ionospheric plasma between the shuttle and the satellite.

[0008] An alternative design to the above-mentioned space tether was purportedly developed by or for the U.S. Army in 1988 and patented in U.S. Pat. No. 4,923,151 (1990) to Roberts. Roberts discloses a coaxial tether (Roberts, FIGS. **2** and **3**) about 12.5 miles long, comprising an outer conductor **20**, a magnetic shield conductor **26**, and a center conductor **30**. Outer conductor **20** is electrically connected at one end to a satellite **12** and at the other end to an orbiting shuttle **10** (Roberts, FIGS. **1** and **2**). Center conductor **30** is magnetically shielded from the geomagnetic field by conductor **26**, which is made of a high permeability metal such as soft iron. Center conductor **30** is electrically insulated from conductors **20** and **26** by insulation **28**. Center conductor **30** is also electrically connected at one end to satellite **12** (at point **32**) and at the other end to a load **36** inside shuttle **10** (through a switch **34**). Load **36** is grounded via a return current path including the outer surface **40** of shuttle **10**, conductor **20**, and the outer surface **22** of satellite **12**. The tether passes through the geomagnetic field at an orientation substantially normal to the earth and its magnetic field lines. As a result, a potential difference ("PD") is produced between shuttle **10** and satellite **12**. Conductor **26** magnetically shields center conductor **30**, so no induced current flows in conductor **30**. Conductor **30** assumes the charge potential of satellite **12** and applies the potential difference PD across load **36** (once switch **34** is closed), causing current to flow in load **36**. Conductor **20** replaces the ionosphere as the return current path, purportedly making the circuit more stable. Thus, conductor **20** functions both as a Faraday conductor (i.e., a conductor on which currents are induced from a magnetic field) and as the return current path. In this case, the Faraday conductor (conductor **20**) does not function as the supply current path (conductor **30** does). In the previously discussed space tether design, the Faraday conductor (i.e., the tether's only conductor) did serve as the supply current path and did not serve as the return current path (the ionosphere).

[0009] For over 40 years, the inventor named herein has been working on an idea for generating electricity from the earth's rotation and magnetic field. The argument that electricity is generated in the earth is supported by some natural phenomena, including: the decreasing speed of the earth's rotation, the continued molten state of the earth's interior, volcanic and geyser activity, the earth's magnetic field, and lightning. The earth uses its primary kinetic energy, which originally started its rotation, to electrically charge itself by electromagnetic induction. Electric energy flowing toward the center of the earth is transformed into magnetic and heat energy, sustaining the earth's magnetic field and the molten state of the earth's interior. The balance in pressure between the surface and the molten interior is maintained through the loss of heat energy at the surface by heat radiation and volcanic and geyser activity. As a result of the work performed by the earth, and possibly because of a counter electromagnetic force (Lenz's law), the earth's kinetic energy is decreasing and its rotation is slowing.

[0010] The inventor herein began thinking about his idea while observing thunderstorms. In 1979, during one of his experiments, he was struck by a branch of lightning and luckily survived with no significant injury. He decided to continue experimentation, but in a safe manner. Upon arrival to the United States, the inventor herein applied for a grant and technical support from the National Institute of Standards and Technology (NIST) on May 18, 1995. His application was assigned an Evaluation Request (ER) Number 32042 and a Control Number 015351. The application was ultimately rejected with an explanation that the magnetic field rotates with the earth and thus currents could not be induced in a conductor installed on the earth. NIST's rejection of the idea dissuaded the inventor herein from pursuing the idea further, until recently when the inventor started thinking about the Faraday disk and "Faraday's paradox."

[0011] Faraday's paradox is the non-intuitive result that electric currents are induced in the Faraday disk even when the magnet rotates with the disk. It is non-intuitive because many would assume that the magnetic field rotates with the rotating magnet and disk, in which case currents would not be induced in the disk. Faraday conducted three tests to investigate this phenomenon. First, he rotated the disk and kept the magnet stationary and observed a current induced in the disk. Second, he kept the disk stationary and rotated the magnet and observed no induced current in the disk. Third, he rotated the disk and magnet together and observed a current in the disk. From these tests, Faraday concluded that the magnetic field remained stationary when the magnet was rotated around its own magnetic axis. Similarly, the earth (representing a rotating disk and magnet) should produced a stationary magnetic field through which the earth rotates, thus inducing generally radial electric currents in the earth and on conductors substantially aligned with the earth's radius. Since Faraday, there has been a long-standing debate and experimentation about Faraday's paradox and whether or not the magnetic field rotates with the magnet. For example, see Valone I, pp. 2-4 and 37. Notwithstanding the theoretical debate and investigation, it is believed that Faraday was correct and that currents are induced by the geomagnetic field in the earth, in the surrounding atmosphere, and in conductors appropriately located on the earth and in the atmosphere.

[0012] Nikola Tesla disclosed a conductor extending into the atmosphere and normal to the earth's surface in U.S. Pat. No. 685,957 (Tesla). Rather than seek to induce currents in the conductor from the geomagnetic field, Tesla sought to collect radiant energy (or particles charged by radiant energy) primarily from the sun. Tesla somehow suspended a conductive plate P and connected it to the distal end of the conductor (Tesla, FIG. **1**). The proximal end of the conductor was connected to a capacitor C. The other end of capacitor C was connected to a secondary conductor, which is, in turn, connected to a conductive plate P' grounded in the earth. Plate P was the collector of radiant energy charged particles. The charge from plate P charged capacitor C. A voltage threshold device (d) and a load R are connected across capacitor C to form a circuit. Device (d) closes the circuit when the charge on capacitor C exceeds a predetermined voltage, causing the voltage to be applied to load R. In this patent, Tesla does not seek to obtain the potential difference between two "terrestrial charged bodies" (hereinafter defined).

[0013] Herein, the term "terrestrial" is used in a broad sense, to include the whole or any part of the earth and its surrounding atmosphere, including but not limited to: the earth's inner and outer cores, mantle, crust, any layer or level thereof, and its surface; all bodies of water; air; clouds; ice; vegetation; minerals and other resources such as oil; rocks including basalt; or any other composition or thing of the earth and its atmosphere. As used herein, the term "terrestrial charged body" (or its plural form) means a terrestrial body, region, layer, zone, other volume or area, any composition or thing, with or without specifically defined or fixed boundaries, having an electric charge represented by a voltage potential.

OBJECTS AND SUMMARY OF THE INVENTION

[0014] It is therefore an object of the present invention to overcome the problems and limitations associated with the prior art and to fulfill a long-felt but unsolved need for a renewable source of energy.

[0015] It is another object of the present invention to provide a clean and renewable source of electrical energy.

[0016] It is a further object of the present invention to generate electricity from the earth's magnetic field and rotation.

[0017] It is yet another object of the present invention to provide a reliable source of electricity from the earth's magnetic field and rotation.

[0018] It is yet a further object of the present invention to provide a source of electricity from the earth's magnetic field and rotation that can be easily implemented.

[0019] It is still another object of the present invention to provide a source of electricity from the earth's magnetic field and rotation that can be implemented at a low-cost.

[0020] These and other objects are attained in accordance with the present invention, wherein there is provided a system for generating electricity from the geomagnetic field and rotation of the earth. The earth rotates through the geomagnetic field to form a potential difference between first and second terrestrial charged bodies. The first and second terrestrial charged bodies are spaced apart in a direction substantially normal to the earth's surface. The system comprises an electrical load, an electrically insulated first current path, and a second current path. The electrical load has a supply input and a ground output. The first current path is coupled to the first terrestrial charged body and to the supply input of the electrical load, for supplying the potential difference to the electrical load. The second current path is coupled to the ground output of the electrical load and to the second terrestrial charged body. As a result of this arrangement, an electric current related to the potential difference is generated in the electrical load.

[0021] In one particular embodiment, the first terrestrial charged body includes a first portion of the earth located below the earth's surface and the second terrestrial charged body includes a second portion of the earth located at or near the earth's surface. The earth contains an opening extending between the first and the second portions of the earth. In this embodiment, the first current path includes (i) an electrically insulated conductor extending at least partially through the opening and (ii) a conductive contact member coupled to the electrically insulated conductor for establishing electrical contact between the first portion of the earth and the electrically insulated conductor. In a more specific embodiment, the opening is a well bore or mine shaft, and the first portion of the earth includes a bottom of a well or mine. The electrically insulated conductor extends through the bore or mine shaft, and the conductive contact member establishes electrical contact between the bottom of the well or mine and the electrically insulated conductor.

[0022] In another embodiment, the first terrestrial charged body may be a portion of the earth at or near the surface of the earth, and the second terrestrial charged body may be a cloud above the surface of the earth. In yet another embodiment, the first and second terrestrial charged bodies may be first and second charged portions of the earth, respectively, where both portions are located at or near the surface of the earth. In this latter embodiment, the first portion of the earth is at a first elevation and the second portion of the earth is at a second and different elevation. For example, the first elevation may be at or near the bottom of a mountain or cliff and the second elevation may be at an elevated position on the mountain or cliff. In still another embodiment, first terrestrial charged body is a portion of the earth at the bottom of a body of water, such as the ocean, and the second terrestrial charged body is the water at or near the surface of the body of water.

[0023] In any of the embodiments, the electrical load may include a rechargeable energy source, including a rechargeable battery or battery pack and a charging circuit coupled to the rechargeable battery or pack. Alternatively, or in addition to batteries, the rechargeable energy source may include a charging capacitor or a bank of charging capacitors along with an appropriate charging circuit for the capacitor(s). The capacitor or capacitors may be ultracapacitors. The electrical load may further include an electric motor coupled to the rechargeable energy source. The electric motor may be a DC motor, which may be mechanically coupled to an AC generator. In operation, the DC motor would actuate the AC generator, resulting in the AC generator generating AC electric power. Alternatively, the electric load may include an inverter for converting DC electric power to AC electric power. In either case (AC generator or inverter), the AC electric power can be stepped up in voltage and applied to an electric power grid.

[0024] A method of generating electricity from the geomagnetic field and rotation of the earth is also within the scope of the present invention. As indicated earlier, the earth rotates through the geomagnetic field to form a potential difference between first and second terrestrial charged bodies. The first and second terrestrial charged bodies are spaced apart in a direction substantially normal to the earth's surface. The method comprises the steps of: (a) providing an electrically insulated first current path having first and second electrically exposed ends; (b) coupling the first end of the first current path to the first terrestrial charged body to establish electrical communication between the first current path and the first terrestrial charged body; (c) coupling the second end of the first current path to a supply input of an electrical load to supply the potential difference to the electrical load; (d) providing a second current path having first and second ends; (e) coupling the first end of the second current path to a ground output of the electrical load; and (f) coupling the second end of the second current path to the second terrestrial charged body to establish electrical communication between the electrical load and the second terrestrial charged body. As a result of these steps, an electric current related to the potential difference is generated in the electrical load.

BRIEF DESCRIPTION OF THE DRAWING

[0025] Further objects of the present invention will become apparent from the following description of the preferred embodiments, with reference to the accompanying drawing, in which:

[0026] FIG. **1** is a diagram of a test setup used to demonstrate the operation of a Faraday disk and the generation of electricity within the rotational frame of reference;

[0027] FIGS. **2** is a diagram of an experiment showing a rotating Faraday disk with a voltmeter having lead wires and test leads for measuring voltages at different points along the disk's radius;

[0028] FIG. **3** is a diagram of an experiment illustrating a battery analogy of the Faraday disk and the concept of short circuit (SC) current on the disk;

[0029] FIG. **4** is a diagram of a field experiment with one embodiment of the present invention, implemented in conjunction with a water well;

[0030] FIG. **5** is a diagram of the earth rotating through its magnetic field and showing (schematically) three terrestrial charged bodies spaced apart from each other in a direction normal to the earth's surface;

[0031] FIG. **6** is a diagram of the earth partially dissected to reveal a portion of a flat circular plane through the earth's equator, presented as an analogy of a Faraday disk;

[0032] FIG. **7** is a diagram of a first embodiment of the present invention, wherein the terrestrial charged bodies are separated portions of the earth;

[0033] FIG. **8** is a diagram of a second embodiment of the invention, which is an implementation utilizing a well;

[0034] FIG. **9** is a diagram of a third embodiment of the invention, which is an implementation utilizing a mine and a mine shaft;

[0035] FIG. **10** is a diagram of a fourth embodiment of the invention, which is an implementation utilizing a body of water, wherein one terrestrial charged body is the bottom of the body of water and another terrestrial charged body is a surface of the body of water;

[0036] FIG. **11** is a diagram of a fifth embodiment of the invention, which is an implementation utilizing different elevations of the earth's surface, such as associated with a mountain or cliff;

[0037] FIG. **12** is a diagram of a sixth embodiment of the invention, which is an implementation utilizing the atmosphere of the earth, wherein one terrestrial charged body is a cloud and another terrestrial charged body is a surface of the earth;

[0038] FIG. **13** is a diagram of a seventh embodiment of the invention, which is an implementation utilizing a well having a bore casing and an electrical load (which may include a submersible pump) is located at least partially within the bore casing;

[0039] FIG. **14** is a diagram of an eighth embodiment of the invention, which employs coaxial conductors to carryout the invention;

[0040] FIG. **15** is a cross-sectional view of the coaxial conductors of FIG. **14**, taken along line **15-15** in FIG. **14**;

[0041] FIG. **16** is a block diagram of one example of an electrical load of the present invention, which is a rechargeable energy source including a charging circuit and a charge storage device, such as a rechargeable battery(s) or storage capacitor(s);

[0042] FIG. **17** is a block diagram of another example of an electrical load of the invention, which includes a charging and control circuit, charge storage device, and an electric motor; and

[0043] FIG. **18** is a block diagram of a further example of an electrical load of the present invention, including a charging and control circuit, a charge storage device, and an inverter or AC generator with step-up transformer.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

[0044] Inspired by the book "The Homopolar Handbook" by Thomas Valone (Valone I), the named inventor built a Faraday disk to perform a few original experiments. A direct consequence of the experiments is the extraction and use of one part of the existing electrical energy of the earth. FIG. **1** illustrates a Faraday disk apparatus **10** built by the inventor herein. Apparatus **10** includes a frame **12** made of 2×6 wooden boards, connected together with screws. An axle **14** is made of a ⅝ inch diameter, 3-foot long steel threaded rod. Axle **14** has a sharpened lower end **14***a*, which rests on a steel plate **16**. Axle **14** is rotatably supported by two bearing assemblies **18***a* and **18***b*. Bearing assembly **18***a* is mounted in an upper cross member **12***a* of frame **12** and bearing assembly **18***b* is mounted in a bottom cross member **12***b* of frame **12**. A drill **20** is connected, via a chuck **22**, to a short rod **24**, which is, in turn, connected to one end of an insulator block **26**. The upper end of axle **14** is connected to the other end of insulator block **26**. Rod **24**, block **26** and axle **14** rotate together as a unit. The speed of drill **20** is manually controlled to control the rotational speed of axle **14**. Insulator **26** electrically isolates axle **14** from the drill to eliminate any chance of electric current from drill **20** finding its way into axle **14**.

[0045] As further shown in FIG. **1**, the actual Faraday disk comprises a 17-inch diameter round aluminum baking pan **28** and two 16-inch diameter (and ¼-inch thick) steel disks **30***a* and **30***b*. Pan **28** and disks **30***a*, **30***b* are fixed together and balanced on axle **14**, using washers and steel hex nuts threaded on axle **14**. Disk **30***a* is fixed in and against the bottom of pan **28** and disk **30***b* is fixed directly underneath and against pan **28** (collectively referred to as disk assembly **28'**). Axle **14** and disk assembly **28'** rotate together, almost like a spinning top. In order to reduce electrical resistance associate with axle **14** above disk assembly **28'**, a copper ring **32** is clamped to axle **14** and an insulated copper wire **34** is connected between ring **32** and disk assembly **28'**. Two spring-loaded graphite brush contacts **34***a* and **34***b* are installed. Brush contact **34***a* is adjacent to and in contact with copper ring **32** and brush contact **34***b* is adjacent to and in contact with the rim of disk assembly **28'**. Contact **34***a* and a lead wire **36***a* electrically connect ring **32** to one input of a digital voltmeter **38**. Contact **34***b* and a lead wire **36***b* electrically connect the rim of disk assembly **28'** to a second input of voltmeter **38**. Voltmeter **38** is mounted outside of frame **12** and is held stationary. Another digital voltmeter **40** is mounted on disk assembly **28'** and rotates with assembly **28'**. Voltmeter **40** is electrically connected, via a short lead wire **42***a*, to the rim of disk assembly **28'** and is electrically connected, via a longer lead wire **42***b*, to axle **14**. A digital camera **44** is fixed on axle **14** (just below insulator block **26**) to record the digitally displayed readings of voltmeter **40**.

[0046] Again referring to FIG. **1**, twenty-two cylindrical welding magnets, each five inches in diameter and each with a pull strength of 75 lbs, are arranged to form two magnetic cylinders **46***a* and **46***b* (11 magnets each). Cylinders **46***a* and **46***b* are separated from each other by disk assembly **28'**. Magnetic cylinders **46***a*, **46***b* are installed in a way that allows them to be held stationary or to be rotated with disk assembly **28'**. In a series of tests, axle **14** was rotated at a speed of about 200 RPM and voltmeter **38** measured and read a generated potential difference of about 300 millivolts. No difference in voltage generation was noted between tests where magnetic cylinders **46***a*, **46***b* were held stationary or rotated (Faraday's paradox). During the 200 RPM tests, voltmeter 40 read 1 to 3 millivolts, supporting the view that a voltage potential difference is generated and measurable on the rotating frame of reference. These tests are referred to collectively as "Experiment A." The difference in readings between voltmeter **38** and voltmeter **40** will be explained below.

[0047] Referring now to FIG. **2**, a diagram of a rotating disk experiment is shown. Disk assembly **28'** and voltmeter **38** are used in this experiment. Lead wires **36***a* and **36***b* of voltmeter **38** include test leads **48***a* and **48***b*, respectively. Leads **48***a*, **48***b* are placed on rotating disk assembly **28'** in a sliding engagement to establish dynamic electrical contact with assembly **28'**. As disk assembly **28'** was rotated, voltage measurements were taken at different points along the radius of the assembly (i.e., at different imaginary concentric circles on disk assembly **28'**). Voltmeter **38** measured the potential difference between the points of contact. The positive side of the potential difference is toward the center, as indicated in FIG. **2**, due to the Lorentz force pushing electrons radially out to the rim. The measured potential difference increased as the distance between the contact points increased. This experiment is referred to hereinafter as "Experiment B."

[0048] An understanding as to why there is a difference in electric potential measured by voltmeter **38** and voltmeter **40** (FIG. **1**) was gained in another experiment, illustrated in FIG. **3**. In this experiment, an attempt was made to model the electrodynamics on a rotating Faraday disk. This was done by applying a potential difference across disk assembly **28'** (between axle **14** and the rim), while disk assembly **28'** remained stationary (as did voltmeter **38**). This setup was intended to mimic the case where the voltmeter and disk are rotating in the same frame of reference. The potential difference was applied across disk assembly **28'** using a 6-volt battery **50**. This modeled the potential difference generated on a rotating disk by electromagnetic induction. Insulated copper wires **52***a* and **52***b* were connected to the terminals of battery **50** and to axle **14** and the rim of disk assembly **28'**, as shown in FIG. **3**. The positive side of battery **50** is connected at axle **14** and the negative side is connected at the rim to model the charge displacement due to induction. Lead wires **36***a,* **36***b* of voltmeter **38** were also connected to disk assembly **28'**, in the vicinity of the connections of wires **52***a,* **52***b,* respectively. The voltage (or potential difference) measured by voltmeter **38** was 50 millivolts. A reading of 6 volts would be expected in an open circuit measurement. On the other hand, a reading of 50 millivolts would suggest a near short circuit (considering the inherent resistance in disk assembly **28'**). Voltmeter **38** did measure a short circuit current, which is indicated in FIG. **3** by arrow SC. This experiment is referred to hereinafter as "Experiment C."

[0049] Based on Experiments A, B and C, it was concluded that the rotation of disk assembly **28'** in the magnetic field (created by magnetic cylinders **46***a,* **46***b*—FIG. **1**) caused the disk assembly to become charged and develop a potential difference in the radial direction. The potential difference increased as the distance between the center of disk assembly **28'** and a measurement point along the radius of assembly **28'** increased. This potential difference was created by the Lorentz force pushing electrons out to the rim, causing an induced current to flow radially outward. In response to the induced potential difference, electrons seek to establish equilibrium on the disk (as it is a conductor) and begin to flow in a direction opposite of the induced current. This opposite flow of electrons is referred to as the short circuit current and is indicated in FIG. **3** by arrow SC. It is believed that this short circuit current is flowing at all times on a rotating Faraday disk. This may be the reason why a voltage is measureable on the rotating frame of reference (voltmeter on the disk), albeit a smaller voltage than the induced voltage measured in the stationary frame of reference. A rotating Faraday disk (i.e., disk assembly **28'**) behaves like a battery (or multiple batteries connected in series) with one pole at the center of the disk (axle **14**) and the other pole at the rim. While rotation in the magnetic field causes the disk to be charged, a short circuit simultaneously occurs. The disk loses its charge immediately when it stops rotating. A Faraday disk (and thus disk assembly **28'**) can be thought of as comprising a number of imaginary annular rings, each behaving like a circular-shaped battery and the batteries are connected in series. Thus, the largest potential would occur at the rim, where all of the annular rings (batteries) of the disk are in series and in the circuit.

[0050] The concepts studied in the lab in Experiments A, B and C were then applied in the field, with respect to the earth's rotation and magnetic field. FIG. **4** illustrates an experiment that was conducted at two different drilled water wells, one 117 feet deep and the other 210 feet deep. The wells were drilled at construction sites of new homes in a suburban area near Utica, N.Y. At the time of the experiments, the wells were not in use. There were no electric power lines or transmitting antennas within a great distance around the construction sites. A copper pipe **54**, one-half inch in diameter and two feet long, was connected to an insulated copper wire **56**, to ensure a good connection with the earth at the bottom of the well. Pipe **54** and wire **56** were lowered to the bottom of each well. Pipe **54** made direct contact with the earth at the bottom of each well. The bottom of each well is considered a first charged body (defined above). Another one-half inch diameter and two-foot long copper pipe **58** was hammered into the ground at the surface, about 6 feet from the well. The surface and earth surrounding pipe **58** is considered a second charged body. A digital voltmeter **60** was used to measure the difference in electric potential between the bottom of the well and the surface of the ground (i.e., between the first and the second charged bodies). Voltmeter **60** has test lead wires **62** and **64**. Test lead wire **62** is connected to insulated wire **56** and test lead wire **64** is connected to copper pipe **58**. Steady measurements of **592** millivolts were obtained in the 117-foot well and 858 millivolts were obtained in the 210-foot well. Short circuit current between the surface and the bottom of the well was also measured with meter **60**. As the distance between the surface and bottom of the well increased, the difference in electric potential increased.

[0051] A variation of the setup in FIG. **4** was also tested at the 210-foot well (indicated in dash lines). The well includes a metal bore casing **66** that extends substantially down the well bore. The variation is that casing **66** was used as part of the circuit instead of copper pipe **58**. Test lead wire **64** was connected directly to casing **66**. The voltage measurements were consistent with those obtained in the original tests for the 210-foot well. It is believed that the use of casing **66** as part of the circuit, instead of a ground pipe (like pipe **58**), may yield more consistent voltage readings. The tests performed at both wells, as described in this and the previous paragraph, are collectively referred to hereinafter as "Experiment D." In the tests of Experiment D, wire **56** was an insulated 14-Gauge Stranded THHN Cable, Model No. 22955957, supplied by the Southwire Company, Carrollton, Ga., featuring conductors made of annealed (soft) copper, rated at 600 Volts, and insulated with tough, heat- and moisture-resistant PVC. Voltmeter **60** was a Mini Pocket MultiMeter, Model DM110, by Extech Instruments Corp., Nashua, N.H., with 0.1 millivolt DC voltage resolution and 0.01 milliamp DC current resolution. Another voltmeter was also used to re-confirm the readings. That meter was a Cen-Tech Seven Function Digital Multimeter, Model 98025, supplied by Harbor Freight Tools Co., Camarillo, Calif., with milliamp and millivolts resolutions.

[0052] The earth is an oblate spheroid with an equatorial radius of 6378.16 km. The earth and its surrounding atmosphere rotate within the earth's magnetic field (i.e., the geomagnetic field). The speed of a point at the earth's equator is about 463.8 m/sec. In the Utica, N.Y. area, where Experiment D was performed, the speed is about 300 m/sec. By comparison, the speed of sound is 340.29 m/sec. The earth's magnetic field is estimated to be about 0.35 to 0.65 gauss at the earth's surface. By comparison, a strong refrigerator magnet has a field strength of 100 gauss. The geomagnetic field is generally dipolar as it has two poles, the north and south magnetic poles, and the field lines resemble those of a dipolar magnet. FIG. **5** illustrates the geomagnetic field on one side of the earth's magnetic axis **70**. Generally, the earth can be considered a conductor having the following five parts: (1) the solid inner core, which is an excellent conductor; (2) the molten outer core, which is an excellent conductor; (3) the mantle, much of which is a good conductor; (4) the crust, which is a relatively good conductor; and (5) the atmosphere, which can be a good or poor conductor.

[0053] To help with the understanding of the results obtained in Experiment D and with the operation of the present invention, consider the following theory. (This theory should not be construed as defining or limiting the scope of the present invention in any way, and it should not affect the validity of the invention if proven to be incorrect. It is presented in the hope that it may aid in the understanding of the invention.) One should picture the earth and the surrounding (rotating) atmosphere as an endless number of concentric spheres, from the earth's center to the earth's surface, and from the earth's surface to the outer reaches of the surrounding, rotating atmosphere. FIG. **5** depicts two spheres associated directly with the earth itself. A sphere **72** represents any level or layer under the earth's surface between the center and the surface. For example, sphere **72** could be (generally) a junction between the earth's crust and its mantle, a junction between the earth's mantle and the outer core, or a particular layer within the earth's crust or mantle. A sphere **74** represents the earth's surface. Alternatively, one can picture the earth as an endless number of disks, with the largest diameter disk cutting through the equator (see FIG. **6**) and the smallest diameter disks approaching the poles.

[0054] As shown in FIG. **5**, the geomagnetic field (showing one side of the field) comprises a plurality of magnetic field lines **68** disposed generally in a dipolar arrangement about magnetic axis **70**. As the earth rotates through the geomagnetic field, it behaves similar to a rotating Faraday disk, in that electrons dislocate outward from the center to the surface. This causes a potential gradient to be established along the radius of the earth and into the atmosphere, with the positive side toward the center and the negative side toward the surface. A potential difference is measured between any two designated spheres of the earth and atmosphere (e.g., spheres **72** and **74** in FIG. **5**), and the potential difference increases with the spacing between the designated spheres. At the same time, electrons travel back toward the center of the earth as the electromagnetic system seeks equilibrium, and this causes a permanent short circuit current to flow in the earth's interior.

[0055] Continuing with the theory, induction also exists in the atmosphere, especially in the clouds. Clouds are connected to the earth's surface by air that is generally a poor conductor. Some short circuit current through dry air may exist, but a big difference in electric potential develops between clouds and the earth's surface. During storms, the air between the earth's surface and clouds, and between clouds at different levels (representing different spheres), become saturated with water and form columns that are relatively good conductors. An intense short circuit current occurs through such conductors. When the amount of current becomes so intense that the conductor cannot accept it any longer, because of its properties (air and water), it simply burns out. The flash and explosion (i.e., lightning and thunder) are results of the burning out of the overloaded conductor. Most likely, a similar event happened to NASA's tether experiment, when an almost 20 km long insulated wire burned out by an estimated 1 amp of current at 3500 volts (see NASA tether experiment article, www.istp.gsfc.nasa.gov/Education/wtether.html).

[0056] Further with the theory, the earth behaves like one huge battery (similar to Faraday's disk), constantly being charged by the rotation within its own magnetic field. One pole of the battery is at the earth's center and the other pole is at the earth's surface or at some designated point in the atmosphere. One can also imagine the earth as an endless number of sphere-shaped batteries connected in series. At the same time, a permanent short circuit in the earth exists, like the battery in Experiment C above.

[0057] It is not possible or practicable to obtain the electric potential between the center of the earth and its surface using Faraday's approach, because stationary brush contacts cannot be placed at the center and the surface (i.e., a stationary frame of reference for the collection circuit cannot be established). However, it is possible to obtain an electric potential from the short circuit current between different designated spheres reachable from the earth's surface (e.g., Experiment D).

[0058] Referring again to FIG. **5**, sphere **72** contains a designated terrestrial charged body **76** and sphere **74** (earth's surface) contains a designated terrestrial charged body **78**. A cloud **80** in the atmosphere **82** is also designated as a terrestrial charged body. Terrestrial charged bodies **76**, **78** and **80** are spaced apart from each other in a direction substantially normal to the earth's surface. An imaginary axis **84** represents the alignment of the charged bodies in a direction normal to the surface. Due to electromagnetic induction, a potential difference is formed between terrestrial charged bodies **76** and **78**, between charged bodies **76** and **80**, and between charged bodies **78** and **80**. A terrestrial charged body is that amount (e.g., area or volume) of earth, water, minerals, rocks, clouds, etc. sufficient to define, accumulate or hold a charge which can produce a measurable potential difference with another terrestrial charged body.

[0059] Referring to FIG. **6**, there is shown a model of the earth with a portion removed to reveal an imaginary plane or disk **86**. Disk **86** cuts through the equator. FIG. **6** illustrates an analogy between the earth's interior and a Faraday disk (see earlier discussion above). In this example, an interior section or region of the earth's crust or mantle serves as a first terrestrial charged body **88**, and the earth's surface serves as a second terrestrial charged body **90**. Compare FIG. **6** with FIG. **2**. The rotation of the earth through its magnetic field causes terrestrial bodies **88** and **90** to become charged and a potential difference to be form between them, as in Experiment B (FIG. **2**).

[0060] Referring now to FIG. **7**, a system **100** for generating electricity from the geomagnetic field and rotation of the earth is shown as a first embodiment of the present invention. As explained, a potential difference (PD) is formed between terrestrial charged bodies—here terrestrial charged bodies **102** and **104**. In this embodiment, body **102** is underground earth (e.g., rock, soil, and/or minerals, etc.) at the bottom of an opening or hole (man-made or natural) and body **104** is the surface soil of the earth. Charged bodies **102**, **104** are spaced apart in a direction normal to the earth's surface. System **100** comprises an electrical load **106** having an input **108** and an output **110**. A first insulated conductor **112** is coupled at one end to charged body **102** via a contact member **114** and is connected at its other end to load input **108**. Contact member **114** may be, e.g., a short (e.g., 2 ft long) piece of copper or aluminum pipe. Conductor **112** is insulated to avoid direct electrical contact with the earth, except for charged body **102**. Conductor **112** is soldered or otherwise electrically connected to contact member **114**. The other end of conductor **112** is either directly connected to load input **108** (as shown) or is coupled to input **108** via a jumper wire and/or a connector. In FIG. **7**, a first current path is defined between load **106** and body **102** by contact member **114** and conductor **112**. The first current path supplies the PD to load **106**. A second insulated conductor **116** is coupled at one end to charged body **104** via a contact member **118** and is connected at the other end to load output **110**. Contact member **118** may be the same type as contact member **114**. Conductor **116** may be the same type and model as conductor **112**. A second current path is defined between load **106** and body **104** by conductor **116** and contact member **118**. A return current path occurs between charged bodies **104**, **102** via the intervening earth. Thus, a complete electrical circuit is established from charged body **102**, contact member **114**, conductor **112**, load **106**, conductor **116**, contact member **118**, charged body **104**, and the return current path back to charged body **102**. An electric current related to the PD is generated in load **106**. Load **106** may comprise a charging circuit and a charge storage device, a motor, and/or an AC generator or inverter, as discussed later. Contact members **114**, **118** may be any form of electrical conductor, including but not limited to a plate, disk, sphere, cone, rod, cylinder (hollow or solid), or sharp pointed member or members.

[0061] Referring now to FIG. **8**, a system **200** for generating electricity from the geomagnetic field and rotation of the earth is shown as a second embodiment of the present invention. A potential difference PD is formed between spaced apart terrestrial charged bodies **202** and **204**. Body **202** is underground earth at the bottom of a well and body **204** is the surface soil of the earth. The well includes a drilled wellbore extending from the earth's surface to the well bottom. A metal casing **203** extends down into the bore. At the bottom of the well is a supply of water **205**. System **200** comprises an electrical load **206** having an input **208** and an output **210**. A first insulated conductor **212** is coupled at one end to charged body **202** via a contact member **214** and is connected at its other end to load input **208**. Contact member **214** may be submerged in water **205**. Conductor **212** is insulated to avoid direct electrical contact with metal casing **203** and the earth, except for charged body **202**. A first current path is defined between load **206** and body **202** by conductor **212** and contact member **214**. This current path supplies the PD to load **206**. A second insulated conductor **216** is coupled at one end to charged body **204** via a connection (e.g., a solder joint) **217** to casing **203**. Conductor **216** is connected at its other end to load output **210**. A second current path is defined between load **206** and body **204** by conductor **216**, connection **217** and casing **203**. A return current path occurs between charged bodies **204**, **202** via casing **203** and the earth (and/or water **205**). Thus, a complete electrical circuit is established from charged body **202**, contact member **214**, conductor **212**, load **206**, conductor **216**, connection **217**, casing **203**, charged body **204**, and the return current path back to charged body **202**. An electric current related to the PD is generated in load **206**. In this embodiment, it is preferred that the bottom of the well be at least **400** feet deep relative to the surface.

[0062] Referring now to FIG. **9**, a system **300** for generating electricity from the geomagnetic field and rotation of the earth is shown as a third embodiment of the present invention. A potential difference PD is formed between spaced apart terrestrial charged bodies **302** and **304**. In this embodiment, body **302** is underground earth at a bottom of a mine and body **304** is the surface soil of the earth. The mine contains a drilled or excavated opening or mineshaft extending from the earth's surface to the bottom. A metal frame **303** may be present in the shaft. System **300** comprises an electrical load **306** having an input **308** and an output **310**. A first insulated conductor **312** is coupled at one end to charged body **302** via a contact member **314** and is connected at its other end to load input **308**. Contact member **314** makes direct contact with the bottom of the mine and, in some cases, may slightly penetrate the bottom. Conductor **312** is insulated to avoid direct electrical contact with frame **303** and/or the earth (except for charged body **302**). A first current path is defined between load **306** and body **302** by conductor **312** and contact member **314**. The first current path supplies the PD to load **306**. A second insulated conductor **316** is coupled at one end to charged body **304** via a contact member **318** and is connected at its other end to load output **310**. Contact member **320** has been driven into the earth to make good electrical contact with body **304**. Alternatively, conductor **316** may be coupled to body **304** via a connection (e.g., a solder joint) **320** to metal frame **303** (if any). A second current path is defined between load **306** and body **304** by conductor **316** and contact member **318** (or by conductor **316**, connection **320** and frame **303**). A return current path occurs between charged bodies **304**, **302** via the earth or, alternatively, via frame **303** and the earth. Thus, a complete circuit is established from charged body **302**, contact member **314**, conductor **312**, load **306**, conductor **316**, contact member **318** (or connection **320** and frame **303**), charged body **304**, and the return current path back to charged body **302**. An electric current related to the PD is generated in load **306**.

[0063] Referring now to FIG. **10**, a system **400** for generating electricity from the geomagnetic field and rotation of the earth is shown as a fourth embodiment of the present invention. A potential difference PD is formed between spaced apart terrestrial charged bodies **402** and **404**. In this embodiment, charged body **402** is the bottom or bed of a body of water (e.g., the earth's crust, basalt, sand, clay, sediments or other settled materials, a change in salinity, etc.). Charge body **404** is the water at or near the surface of the body of water. The body of water is preferably a salt water system such as an ocean, sea, ocean bay, strait, estuary, lake, and the like. System **400** comprises an electrical load **406** having an input **408** and an output **410**. A first insulated conductor **412** is coupled at one end to charged body **402** via a contact member **414** and is connected at its other end to load input **408**. Contact member **414** makes direct contact with the bottom and, in some cases, may penetrate the bottom. Conductor **412** is insulated to avoid direct electrical contact with the body of water. A first current path is defined between load **406** and body **402** by conductor **412** and contact member **414**. The first current path supplies the PD to load **406**. A second insulated conductor **416** is coupled at one end to charged body **404** via a metal boat **418**, which serves as a contact member. Conductor **416** is connected at its other end to load output **410**. Conductor **416** is directly connected to boat **418**, as represented by the electrical ground symbol. An alternative contact member may be a thin metal (e.g., copper or aluminum) plate **418'**, which floats on the water. In the latter case, conductor **416** is directly connected to plate **418'**. A second current path is defined between load **406** and body **404** by conductor **416** and boat **418** (or by conductor **416** and plate **418'**). A return current path occurs between charged bodies **404**, **402** via the body of water. Thus, a complete circuit is established from charged body **402**, contact member **414**, conductor **412**, load **406**, conductor **416**, boat **418** or other contact member, charged body **404**, and the return current path back to charged body **402**. An electric current related to the PD is generated in load **406**.

[0064] Referring now to FIG. **11**, a system **500** for generating electricity from the geomagnetic field and rotation of the earth is shown as a fifth embodiment of the present invention. A potential difference PD is formed between spaced apart terrestrial charged bodies **502** and **504**. In this embodiment, both charged bodies **502**, **504** are portions of the earth at or near the earth's surface, but body **504** is at a higher elevation than body **502**. This may be the case, for example, at a mountain or cliff System **500** comprises an electrical load **506** having an input **508** and an output **510**. A first insulated conductor **512** is coupled at one end to charged body **502** via a contact member **514** and is connected at its other end to load input **508**. Contact member **514** is driven into the earth to establish good electrical contact with charged body **502**. Conductor **512** is insulated to avoid direct electrical contact with the earth. A first current path is defined between load **506** and body **502** by conductor **512** and contact member **514**. The first current path supplies the PD to load **506**. A second insulated conductor **516** is coupled at one end to charged body **504** via a contact member **518** and is connected at its other end to load output **510**. Member **518** has been driven into the earth to make good electrical contact with body **504**. A second current path is defined between load **506** and body **504** by conductor **516** and contact member **518**. A return current path occurs between charged bodies **502**, **504** via the intervening earth. Thus, a complete circuit is established from charged body **502**, contact member **514**, conductor **512**, load **506**, conductor **516**, contact member **518**, charged body **504**, and the return current path back to charged body **502**. An electric current related to the PD is generated in load **506**.

[0065] Referring now to FIG. **12**, a system **600** for generating electricity from the geomagnetic field and rotation of the earth is shown as a sixth embodiment of the present invention. A potential difference PD is formed between spaced apart terrestrial charged bodies **602** and **604**. In this embodiment, charged body **604** is a cloud or a stream of clouds passing overhead and charged body **602** is a portion of the earth at or near the earth's surface. A preferred location for system **600** is at a high elevation, such as a mountain top, where low lying clouds are a frequent occurrence. As air is generally a poor conductor, the flow of short circuit current between clouds and the earth's surface is very limited, resulting in the buildup of a huge difference in electrical potential between the clouds and the earth's surface. THIS EMBODIMENT IS VERY DANGEROUS DUE TO LIGHTNING STRIKES AND STATIC ELECTRICITY. ANYONE WHO BUILDS THIS EMBODIMENT, WHETHER AUTHORIZED BY THE INVENTOR HEREIN OR NOT, ASSUMES ALL RISK OF PERSONAL INJURY OR DEATH AND ASSUMES ALL LIABILITY TO OTHERS INJURED OR KILLED AS A RESULT OF THE CONSTRUCTION AND/OR OPERATION OF THIS EMBODIMENT.

[0066] Again referring to FIG. **12**, system **600** comprises an electrical load **606** having an input **608** and an output **610**. A first insulated conductor **612** is coupled at one end to charged body **602** via a contact member **614** and is connected at its other end to load input **608**. Contact member **614** has been driven into the earth to make good electrical contact with body **602**. A first current path is defined between load **606** and body **602** by conductor **612** and contact member **614**. The first current path supplies the PD to load **606**. A second insulated conductor **616** is coupled at one end to charged body **604** via a contact apparatus **618** and is connected at its other end to load output **610**. Apparatus **618** includes a non-conductive tower **618***a*, an upwardly extending metal bracket **618***b*, and a metal collector plate **618***c*. Tower **618***a* is preferably built from a non-conductive material such as, for example, wood, fiberglass, fiberglass reinforced plastic, and the like. Tower **618***a* may be supported by non-conductive guy wires such as Phyllistran guys or Polygon structural fiberglass guys. Alternatively, tower **618***a* may be electrically insulated from the ground using an electrically insulating barrier between the tower and the ground. Conductor **616** is electrically connected to metal bracket **618***b*. Conductor **616** is insulated to avoid electrical contact with the surrounding atmosphere. Bracket **618***b* and collector plate **618***c* establish electrical coupling with the electric charge in cloud **604**. Plate **618***c* is a metal sheet of metal, such as copper or aluminum. A second current path is defined between load **606** and body **604** by conductor **616**, metal bracket **618***b*, and collector plate **618***c*. A return current path occurs between charged bodies **604**, **602**, via the atmosphere or moisture or precipitation in the atmosphere. Thus, a complete circuit is established from charged body **602**, contact member **614**, conductor **612**, load **606**, conductor **616**, bracket **618***b*, collector plate **618***c*, charged cloud **604**, and the return current path back to charged body **602**. An electric current related to the PD is generated in load **606**.

[0067] As mentioned, the hazards of lightning are obviously a major concern in system **600**. Lightning protection schemes are not infallible and no representation is made herein that system **600** is safe from the hazards of lightning strikes or static electricity. In system **600**, a surge arrester device **620** is installed between load output **610** and ground (FIG. **12**). Device **620** is shown as a gas-filled surge arrester, but the preferred arrester is a spark gap rated for lightning strike voltages. Preferably, conductor **616** is a high voltage stranded single conductor cable, such as used to connect lightning rods to grounding rods, but covered with insulation having a high dielectric strength to withstand lightning strikes.

[0068] An alternative embodiment to system **600** (not shown) is to replace apparatus **618** with a hydrogen- or helium-filled balloon (e.g., a weather balloon). The balloon would have a conductor such as copper or aluminum foil wrapped around it or a metal plate or band mounted on it. In this embodiment, insulated conductor **616** would be electrically connected to the conductor on the balloon. This embodiment would otherwise be the same as system **600**. The warnings and disclaimers made above with respect to system **600** are equally applicable to this embodiment.

[0069] Referring now to FIG. **13**, a system **700** for generating electricity from the geomagnetic field and rotation of the earth is shown as a seventh embodiment of the present invention. This embodiment concerns a well and is fundamentally the same as the second embodiment, except here the electrical load is located down in the wellbore. A potential difference PD is formed between terrestrial charged bodies **702** and **704** (i.e., bottom of well and earth surface). The well includes a drilled bore and a metal casing **703** extending down into the bore. At the bottom of the well is a supply of water **705**. An electrical load **706** has an input **708** and an output **710**. A first insulated conductor **712** is coupled at one end to charged body **702** via a contact member **714** and is connected at its other end to load input **708**. A first current path is defined between load **706** and body **702** by conductor **712** and contact member **714**. The first current path supplies the PD to load **706**. A second insulated conductor **716** is coupled at one end to charged body **704** via an electrical connection (e.g., a solder joint) **717** to metal casing **703**. The other end of conductor **716** is connected to load output **710**. A second current path is defined between load **706** and body **704** by conductor **716**, connection **717** and casing **703**. A return current path occurs between charged bodies **704**, **702**, via casing **703** and the earth (and/or water **705**). Thus, a complete electrical circuit is established from charged body **702**, contact member **714**, conductor **712**, load **706**, conductor **716**, connection **717**, casing **703**, charged body **704**, and the return current path back to charged body **702**. An electric current related to the PD is generated in load **706**. In the case where casing **703** is not made of electrically conductive material, an alternative coupling arrangement with charged body **704** is: conductor **716** is connected to a contact member **718**, which is driven into the ground at the earth's surface.

[0070] Load **706** may contain a charge storage system and a wellbore submersible pump controlled by a control unit **720**. The charge storage system stores electricity generated by system **700** and supplies the stored electricity to the pump when called upon by control unit **720**. Control unit **720** communicates with the pump via a control cable **722**. Control unit **720** includes an auto-control function that controls the pump automatically. Unit **720** switches on the power to the pump by means of a pressure switch actuated in response to a pressure valve or by means of a float level switch actuated by a float. An example of a submersible pump powered by a renewable electrical energy source is disclosed in U.S. Pat. No. 7,837,450, which is incorporated herein by reference.

[0071] Referring now to FIGS. **14** and **15**, a system **800** is another embodiment of the present invention. This embodiment is characterized by a coaxial cable **801** installed between two spaced apart, substantially parallel, terrestrial charged bodies **802** and **804**. Cable **801** includes a copper (or copper-clad steel) center conductor **812** insulated by a dielectric insulating material **812***a*, such as cellular polyurethane, polyurethane foam or solid polyethylene. Cable **801** further includes an outer shield conductor **803**, which may be a woven braid of bare or tinned copper wires, a wrap of aluminum foil tape, a solid copper or aluminum tube, or a combination thereof. Shield conductor **803** may be insulated or left uninsulated. In FIGS. **14** and **15**, conductor **803** is insulated generally by a jacket **803***a* made of plastic, such as polyvinyl chloride (PVC). When cable **801** is installed, jacket **803***a* is preferably removed between charged bodies **802** and **804**, as shown in FIG. **14**. System **800** comprises an electrical load **806** having a supply input and a ground output made available at a threaded female coaxial connector **808**. Cable **801** has a finished proximal end with a threaded male coaxial connector **809**. Connector **809** is considered male because it has a stinger wire, which is to be inserted into connector **808**. The stinger wire is connected to or an extension of center conductor **812**. Connector **809** includes a threaded slip ring **809***a* for threading connector **809** onto connector **808**.

[0072] Once connectors **808** and **809** are connected together, center conductor **812** is in electrical contact with the supply input of load **806** and shield conductor **803** is in electrical contact with the ground output of load **806**. Center conductor **812** is electrically coupled to charged body **802** via a metal contact member **814**. Conductor **812** delivers the potential difference between charged bodies **802**, **804** to load **806**. Shield conductor **803** is electrically coupled to charged body **804** by direct contact as shown. Thus, the ground output of load **806** is electrically coupled to charged body **804**. Conductor **803** is also coupled to charged body **802** by direct contact as shown. Alternatively, conductor **803** may be in direct contact with a portion of the earth defined to be outside of charged body **802** (thus making indirect contact with body **802**). A return current path occurs between charged bodies **802**, **804** primarily through shield conductor **803**. Using conductor **803** as the return current path improves return path performance by making the path easier to implement and maintain, and it operates more predictably because it is not dependent on the variable properties and conditions of earth, water, atmospheric, and other natural or in situ return current paths. In system **800**, a complete electrical circuit is established from charged body **802**, contact member **814**, center conductor **812**, load **806** (via connectors **809**/**808**), shield conductor **803** (via connectors **808**/**809**), charged body **804**, and shield conductor **803** back to charged body **802**. An electric current related to the PD between charged bodies **802**, **804** is generated in load **806**.

[0073] As shown in FIG. **14**, an alternative or additional ground output **810** is provided in electrical load **806**. An insulated conductor **816** is connected to output **810** and to a contact member **818**. In this case, shield conductor **803** may be decoupled from the original ground output, for example, by operation of an automatic impedance sensing switch internal to load **806**. Optionally, both outputs (**810** and original output via **803**) may be in operation together. In connection with output **810**, the return current path remains primarily through shield conductor **803**. The complete electrical circuit is from charged body **802**, contact member **814**, center conductor **812**, load **806** (via connectors **809**/**808**), conductor **816**, contact member **818**, charged body **804**, and shield conductor **803** back to charged body **802**.

[0074] Referring now to FIG. **16**, there is shown one example of an electrical load **900** of the present invention. A supply input S and a ground output G of load **900** correspond to, for example, load input **108** and load output **110**, respectively (e.g., see FIG. **7**). The potential difference between a pair of terrestrial charged bodies, obtained in accordance with the present invention, is presented across terminals S and G of load **900**. The apparatus of the present invention that obtains this potential difference (examples of which are shown in FIGS. **7**-**14**) is referred to in FIGS. **16**-**18** as a geomagnetic voltage source. Load **900** includes a charging circuit **902** and a charge storage system **904**. The specifics of charging circuit **902** depend upon the type of charge storage system used. System **904** may be, for example, a rechargeable battery or series of batteries or a charging capacitor or bank of charging capacitors. A basic charging circuit for a rechargeable battery (or series of batteries) may be a positive going diode in series with a source resistor, and both in series with the rechargeable battery (or series of batteries). A basic charging circuit for a capacitor (or bank of capacitors in series) may be an inductor in series with the capacitor (or bank of capacitors). The latter circuit typically includes a switching element or elements to cause the geomagnetic voltage source to charge the inductor in a first switched state and to cause the charged inductor to charge the capacitor or capacitors in a second switched state. If capacitors are used for the charge storage elements, a preferred form of capacitor is the ultracapacitor. Ultracapacitors and their charging circuits are disclosed in the following references: M. S. W. Chan, K. T. Chau, and C. C. Chan, Effective Charging Method for Ultracapacitors, Journal of Asian Electric Vehicles, Vol. 3, No. 2, December 2006, pp. 771-76; and Maxwell Technologies Application Note, Document 1008981 Rev 1, December 2005, pp. 2-3. The disclosures of these references are incorporated herein by reference. See also U.S. Pat. No. 7,170,260 to Thrap (Maxwell Technologies), entitled, Rapid Charger for Ultracapacitors, which is incorporated herein by reference.

[0075] Referring now to FIG. **17**, there is shown another example of an electrical load **900'** of the present invention. Load **900'** has a supply input S and a ground output G, as with load **900**. A geomagnetic voltage source of the present invention is applied to terminals S and G. Load **900'** includes a charging and control circuit **902'**, a charge storage system **904'**, and an electric motor **906'**. Blocks **902'** and **904'** contain the same types of charging circuits and charge storage elements as in blocks **902** and **904** (FIG. **16**), respectively. Circuit **902'** further includes a control circuit or IC chip that controls the operation of switches SW**1** and SW**2** in FIG. **17**. In a first operating stage, switch SW**1** is closed and switch SW**2** is open, and the geomagnetic voltage source and circuit **902'** charge the storage element or elements of charge storage system **904'**. In a second operating stage, switch SW**1** is open (or sometimes remains closed) and switch SW**2** is closed, and system **904'** powers electric motor **906'**. Motor **906'** is typically a DC motor, but may be an AC motor if, e.g., an inverter is installed between the output of system **904'** and the motor. It is understood that, in an alternative embodiment, load **900'** may simply consist of motor **906'**, which would be directly driven by the geomagnetic voltage source.

[0076] Referring now to FIG. **18**, there is shown a further example of an electrical load **900"** of the present invention. Load **900"** has a supply input S and a ground output G, as with loads **900** and **900'**. A geomagnetic voltage source of the present invention is applied to terminals S and G. Load **900"** includes a charging and control circuit **902"**, a charge storage system **904"**, and an inverter or AC generator with step-up transformer **906"**. Blocks **902"** and **904"** contain the same types of circuits and charge storage elements as in blocks **902'** and **904'** (FIG. **17**), respectively. In a first operating stage of load **900"**, switch SW**1** is closed and switch SW**2** is open, and the geomagnetic voltage source and circuit **902"** charge the storage element or elements of charge storage system **904"**. In a second operating stage, switch SW**1** is open (or sometimes remains closed) and switch SW**2** is closed, and system **904"** powers inverter or AC generator with step-up transformer **906"**. If an inverter, block **906"** directly converts the DC voltage from block **904"** to AC voltage and steps up the voltage to a desired level. If an AC generator, block **906"** includes a DC motor for driving the AC generator and a step-up transformer raises the voltage to a desired level. In the latter case, block **904"** drives the DC motor, which in turn drives the AC generator. The desired level of AC voltage may be at the mains level so that it can be applied to the electric power grid, for contributing power to the grid. FIG. **18** shows the output of block **906"** being applied to an electric power grid.

[0077] The phrases, "earth's surface" and "the surface of the earth," when speaking about orientation or relative position (e.g., "a direction normal to the earth's surface"), mean the general or averaged plane of the surface if observed locally or the general tangent of the surface if observed from space. In addition, these phrases are to be construed broad enough to include the surface of any body of land, water or ice or any geological formation such as, e.g., a mountain top or slope, or a valley, whether above or below sea level.

[0078] In connection with the first through fifth and seventh embodiments of the present invention (FIGS. **7-11** and **13**), the insulated conductors (**112**, **116**, **212**, **216**, **312**, **316**, etc.) may be a 14-Gauge insulated Stranded THHN Cable, Model No. 22955957, supplied by Southwire Co., Carrollton, Ga.

[0079] While the preferred embodiments of the invention have been particularly described in the specification and illustrated in the drawing, it should be understood that the invention is not so limited. Many modifications, equivalents and adaptations of the invention will become apparent to those skilled in the art without departing from the spirit and scope of the invention, as defined in the appended claims.

What is claimed is:

1. A system for generating electricity from the geomagnetic field and the rotation of the earth, the earth rotating through the geomagnetic field to form a potential difference between first and second terrestrial charged bodies, the first and the second terrestrial charged bodies being spaced apart in a direction substantially normal to the earth's surface, said system comprising:

(a) an electrical load having a supply input and a ground output;

(b) a first current path, coupled to the first terrestrial charged body and to the supply input of said electrical load, for supplying the potential difference between the first and the second terrestrial charged bodies to said electrical load; and

(c) a second current path coupled to the ground output of said electrical load and to the second terrestrial charged body,

whereby an electric current related to the potential difference is generated in said electrical load.

2. The system of claim **1**, wherein the first terrestrial charged body includes a first portion of the earth located below the earth's surface and the second terrestrial charged body includes a second portion of the earth located at about or near the earth's surface, and wherein the earth contains an opening extending between the first and the second portions of the earth,

said first current path including (i) an electrically insulated conductor extending at least partially through the opening and (ii) a conductive contact member coupled to the electrically insulated conductor for establishing electrical contact between the first portion of the earth and the electrically insulated conductor.

3. The system of claim **2**, wherein said electrical load is located within the opening, between the first and the second portions of the earth.

4. The system of claim **2**, wherein the opening is a well bore or mine shaft, and wherein the first portion of the earth includes a bottom of a well or mine shaft, said electrically insulated conductor extending through the bore or mind shaft and the conductive contact member establishing electrical contact between the bottom of the well or mine shaft and the electrically insulated conductor.

5. The system of claim **4**, wherein the well bore or mine shaft is lined with a metal casing through which said electrically insulated conductor extends, the metal casing being in contact with the second portion of the earth and said second current path being in contact with the metal casing, such that said second current path is in electrical communication with the second portion of the earth.

6. The system of claim **5**, wherein said electrical load is located within the well bore or mine shaft and includes a rechargeable energy source and a pump, the electric current generated in the electrical load charging the rechargeable energy source and powering the pump.

7. The system of claim **1**, wherein the first terrestrial charged body is a portion of the earth at about or near the surface of the earth and the second terrestrial charged body is a cloud above the surface of the earth.

8. The system of claim **1**, wherein the first and the second terrestrial charged bodies are first and second portions of the earth, respectively, located at about or near the surface of the earth, the first portion of the earth being at a first elevation and the second portion of the earth being at a second elevation, the first elevation being different than the second elevation.

9. The system of claim **8**, wherein the first elevation of the first portion of the earth is at about or near the bottom of a mountain or cliff, and wherein the second elevation of the second portion of the earth is at a position on the mountain or cliff elevated from the bottom of the mountain or cliff.

10. The system of claim **1**, wherein the first terrestrial charged body is a portion of the earth at the bottom of a body of water, and wherein the second terrestrial charged body is water at about or near the surface of the body of water.

11. The system of claim **10**, wherein the body of water is a body of saltwater selected from the group consisting essentially of: ocean, sea, strait, saltwater canal, saltwater bay, saltwater lake, and saltwater river.

12. The system of claim **1**, wherein said first current path is substantially unshielded from the geomagnetic field.

13. The system of claim **1**, wherein said electrical load includes a rechargeable energy source including a rechargeable battery and a charging circuit coupled to the rechargeable battery.

14. The system of claim **13**, wherein said electrical load further includes an electric motor coupled to the rechargeable battery.

15. The system of claim **1**, wherein said electrical load includes a rechargeable energy source including one or more capacitors and a charging circuit coupled to the one or more capacitors.

16. The system of claim **15**, wherein the one or more capacitors are one or more ultracapacitors.

17. The system of claim **15**, wherein said electrical load further includes an electric motor coupled to the one or more capacitors.

18. A method of generating electricity from the geomagnetic field and the rotation of the earth, the earth rotating through the geomagnetic field to form a potential difference between first and second terrestrial charged bodies, the first and the second terrestrial charged bodies being spaced apart in a direction substantially normal to the earth's surface, said method comprising the steps of:

(a) providing a first current path having first and second electrically exposed ends;

(b) coupling the first end of the first current path to the first terrestrial charged body to establish electrical communication between the first current path and the first terrestrial charged body;

(c) coupling the second end of the first current path to a supply input of an electrical load to supply the potential difference to the electrical load;

(d) providing a second current path having first and second ends;

(e) coupling the first end of the second current path to a ground output of the electrical load; and

(f) coupling the second end of the second current path to the second terrestrial charged body to establish electrical communication between the electrical load and the second terrestrial charged body,

whereby an electric current related to the potential difference is generated in the electrical load.

19. The method of claim **18**, wherein step (a) includes providing a coaxial cable having a center conductor, insulation surrounding the center conductor, and an outer shield surrounding the insulation,

step (b) including coupling one end of the center conductor to the first terrestrial charged body, and

step (c) including coupling another end of the center conductor to the supply input of the electrical load, whereby the first current path includes the center conductor of the coaxial cable.

20. The method of claim **19**, further comprising the steps of:

(g) coupling a first end of the outer shield to the first terrestrial charged body; and

(h) coupling a second end of the outer shield to the second terrestrial charged body,

whereby a return current path is established between the first and the second terrestrial charged bodies.

21. The method of claim **20**, wherein step (f) includes coupling the second end of the second current path to the second end of the outer shield, to establish electrical communication between the electrical load and both the outer shield and the second terrestrial charged body.

22. The method of claim **20**, wherein the center conductor of the coaxial cable is substantially unshielded from the geomagnetic field.

23. A method of generating electricity from the geomagnetic field and the rotation of the earth, the earth rotating through the geomagnetic field to form a potential difference between first and second charged portions of the earth, the first and the second charged portions being spaced apart in a direction substantially normal to the earth's surface, said method comprising the steps of:

(a) providing a first conductor having first and second ends;

(b) coupling the first end of the first conductor to the first charged portion, to establish electrical communication between the first conductor and the first charged portion;

(c) coupling the second end of the first conductor to a supply input of an electrical load, to supply the potential difference to the electrical load;

(d) providing a second conductor having first and second ends;

(e) coupling the first end of the second conductor to a ground output of the electrical load; and

(f) coupling the second end of the second conductor to the second charged portion to establish electrical communication between the electrical load and the second charged portion,

whereby an electric current related to the potential difference is generated in the electrical load.

* * * * *

US009712036B2

(12) **United States Patent**
Stupac

(10) **Patent No.: US 9,712,036 B2**
(45) **Date of Patent: Jul. 18, 2017**

(54) **GENERATING ELECTRICITY FROM THE EARTH**

(71) Applicant: **Omer Stupac**, Utica, NY (US)

(72) Inventor: **Omer Stupac**, Utica, NY (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 396 days.

(21) Appl. No.: **14/229,873**

(22) Filed: **Mar. 29, 2014**

(65) **Prior Publication Data**

US 2015/0280518 A1 Oct. 1, 2015

(51) **Int. Cl.**

G01R 31/00	(2006.01)
H02K 53/00	(2006.01)
H02K 99/00	(2014.01)

(52) **U.S. Cl.**
CPC ***H02K 53/00*** (2013.01); ***H02K 99/10*** (2016.11)

(58) **Field of Classification Search**
CPC H02K 57/003; H02K 53/00
USPC 324/96, 76, 76.11, 76.66
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

155,209	A	*	9/1874	Snow	H01M 14/00 429/47
406,968	A		7/1889	Tesla	
685,957	A		11/1901	Tesla	
2,181,601	A	*	11/1939	Jakosky	G01V 3/22 324/357
3,185,877	A		5/1965	Sears	
3,245,836	A	*	4/1966	Agruss	H01M 6/36 429/103
3,288,648	A	*	11/1966	Jones	H01M 8/00 166/248
3,471,394	A	*	10/1969	Sudrabin	G01N 27/301 204/435
4,502,028	A	*	2/1985	Leake	H01P 5/04 324/601
4,678,554	A	*	7/1987	Oppitz	A01G 7/04 204/290.11
4,923,151	A	*	5/1990	Roberts	B64G 1/648 244/158.2
5,451,825	A		9/1995	Strohm	

(Continued)

OTHER PUBLICATIONS

Southwire Co., Southwire 500 ft. 14-Gauge Stranded THHN Wire—Black, Model 22955957, Mar. 10, 2014, p. 1, www.homedepot.com, USA.

(Continued)

Primary Examiner — Giovanni Astacio-Oquendo
(74) *Attorney, Agent, or Firm* — Lawrence P. Trapani, Esq.

(57) **ABSTRACT**

A system for generating electricity from the geomagnetic field and rotation of the earth is presented. The earth rotates through the geomagnetic field to form a potential difference between first and second terrestrial charged bodies. The first and second charged bodies are spaced apart in a direction substantially normal to the earth's surface. The system comprises an electrical load, a first current path, and a second current path. The electrical load has a supply input and a ground output. The first current path is coupled to the first charged body and to the supply input of the electrical load, for supplying the potential difference to the electrical load. The second current path is coupled to the ground output of the electrical load and to the second charged body. As a result of this arrangement, an electric current related to the potential difference is generated in the electrical load.

17 Claims, 7 Drawing Sheets

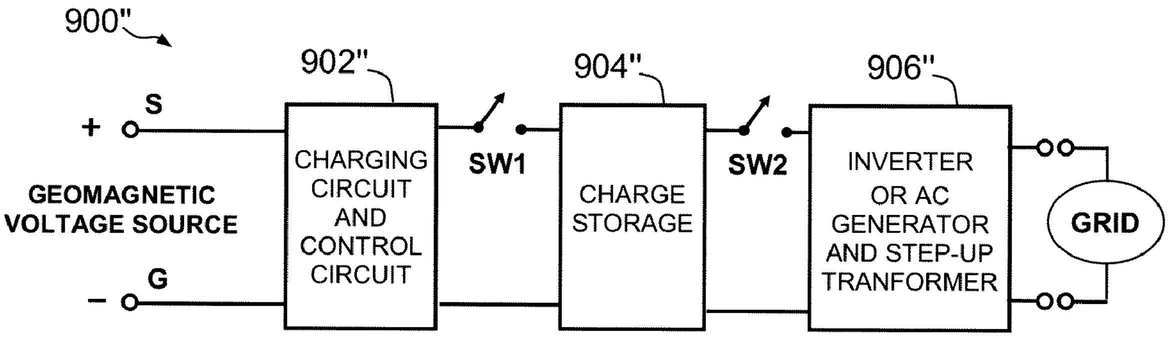
900"
902"
904"
906"
+
S
GEOMAGNETIC VOLTAGE SOURCE
G
−
CHARGING CIRCUIT AND CONTROL CIRCUIT
SW1
CHARGE STORAGE
SW2
INVERTER OR AC GENERATOR AND STEP-UP TRANFORMER
GRID

(56) References Cited

U.S. PATENT DOCUMENTS

6,051,905	A *	4/2000	Clark	H02K 1/2706 310/178
6,297,730	B1 *	10/2001	Dickinson	H01R 13/6625 307/3
7,170,260	B2	1/2007	Thrap	
7,837,450	B2	11/2010	Moreland	
2003/0232599	A1 *	12/2003	Dostert	H04B 3/56 455/66.1
2004/0027022	A1	2/2004	Weir	

OTHER PUBLICATIONS

Chan, Marco S.W. et al., Effective Charging Method for Ultracapacitors, Journal of Asian Electric Vehicles, vol. 3, No. 2, Dec. 2006, pp. 771-775.

Maxwell Technologies, Charging of Ultracapacitors, Application Note, Document 1008981 Rev 1, Dec. 2005, pp. 1-5, Maxwell Technologies, Inc. San Diego, CA.

GHOST7584 (Forum Member), Free Electricity From Earth Rotation, SciForums.com/Free-Electricity-From-Earth-Rotation-t-44434.html, Dec. 4, 2008 (purported date of post, Jan. 25, 2005), pp. 1-6, web.archive.org/web/20081204124044/http://www.sciforums.com/Free-Electricity-From-Earth-Rotation-t-44434.html.

Valone, Thomas, The Homopolar Handbook, A Definitive Guide to Faraday Disk and N-Machine Technologies, 3rd Edition, Oct. 2001, pp. 2-4, 6-8, 24, 37, 49-53, 59-64, and 77-92, Integrity Research Institute, Washington, DC.

Stern, Dr. David P., #25c. The Space Tether Experiment, www-istp.gsfc.nasa.gov/Education/wtether.html, Nov. 2001, pp. 1-3, NASA.

Nichelson, Oliver, Tesla's Fuelless Generator, Jan. 1991, pp. 1-10, home.earthlink.net/~drestinblack/generator.htm.

Stupac, Omer, Production of Electrical Energy from Electromagnetic Induction Using Earth's Rotation, Submission to Office of Technology Innovation, National Institute of Standards and Technology (NIST), May 15, 1995, pp. 1-8, Ridgewood, NY.

NIST, Energy-Related Inventions Program, Report of First-Stage Evaluation (Response to Submission), Aug. 1995, pp. 1-3, NIST, Geithersburg, MD.

National Bureau of Standards, National Engineering Laboratory (NIST) Brochure, The NBS/DOE Energy-Related Inventions Program, Synopsis of Evaluation Process, U.S. Department of Commerce, Washington, DC, Reprinted Jan. 1980, pp. 1-4.

Valone, Thomas, The Homopolar Generator: Tesla's Contribution (from book, Harnessing the Wheelwork of Nature: Tesla's Science of Energy, by Thomas Valone, Dec. 31, 2004, pp. 271-88, Integrity Research Institute, Washinton, DC, Reprinted from Proceedings of International Tesla Symposium, Jan. 1986, pp. 6-29.

Carrigan, Charles R. and Gubbins, David, The Source of the Earth's Magnetic Field, Scientific American, Inc., vol. 240, Issue 2, Feb. 1, 1979, p. 118-128 & 130.

Elsasser, W.M., The Earth as a Dynamo, Scientific American, Inc., vol. 198, Issue 5, May 1, 1958, p. 44-48.

* cited by examiner

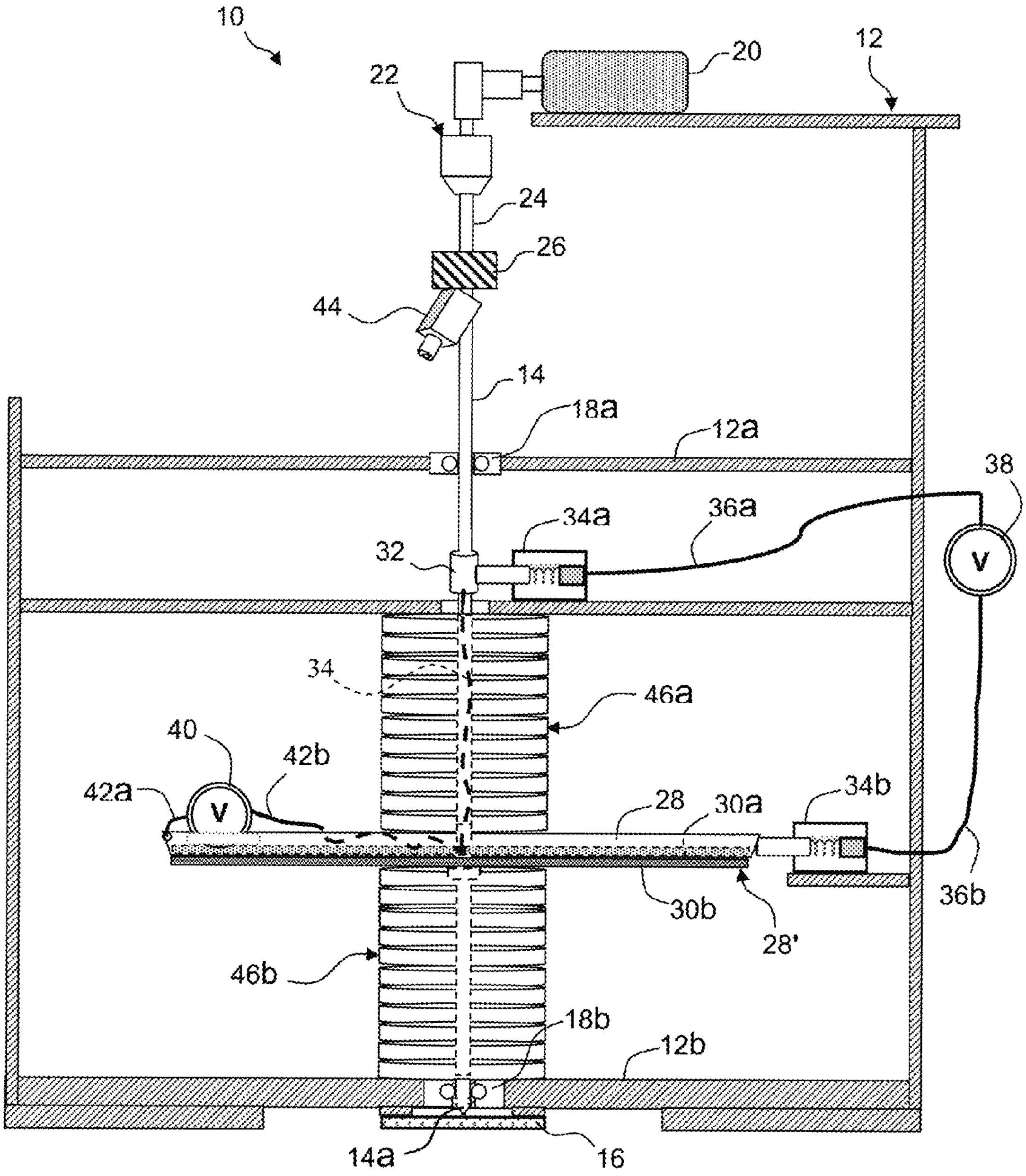
V
V

FIG. 1

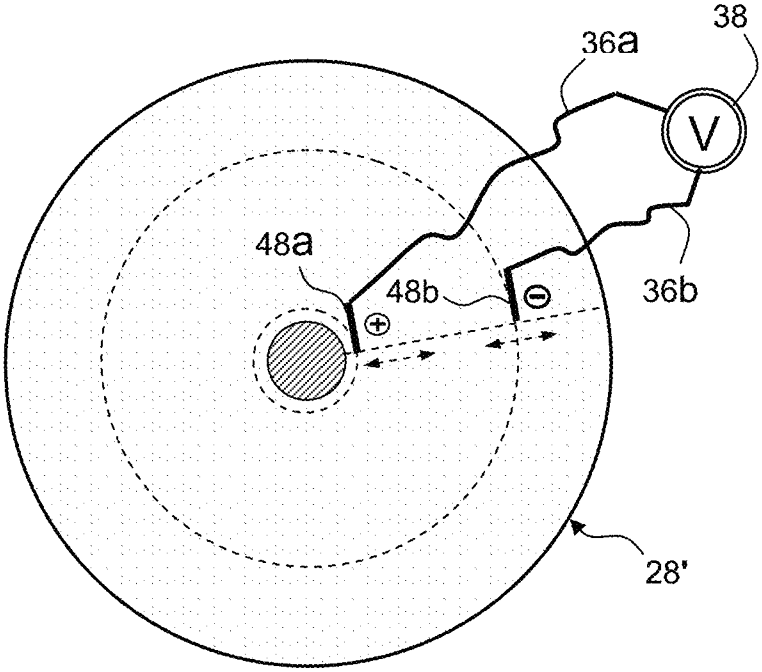
38
36a
V
48a
48b
⊕
⊖
36b
28'

FIG. 2

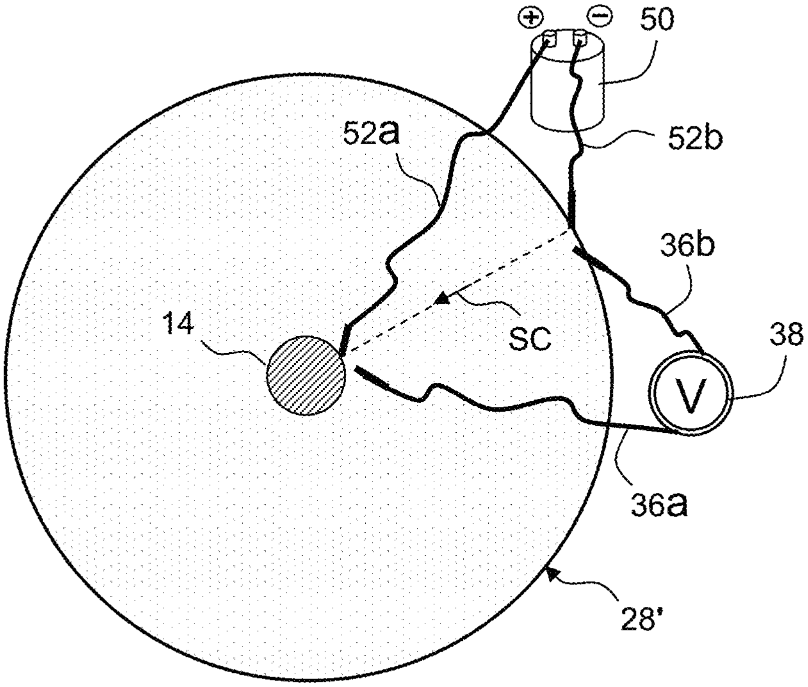
⊕
⊖
50
52a
52b
36b
14
SC
38
V
36a
28'

FIG. 3

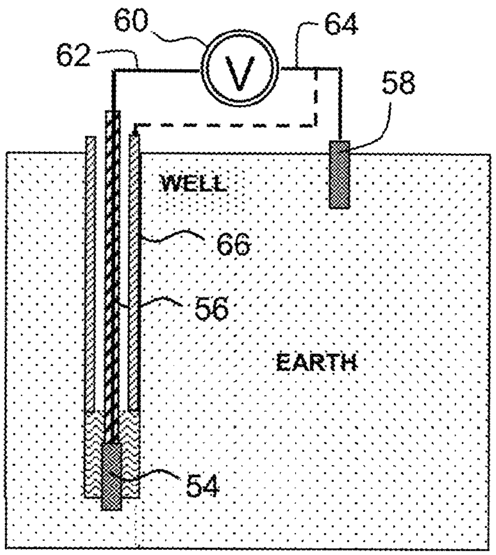
60
64
62
V
58
WELL
66
56
EARTH
54

FIG. 4

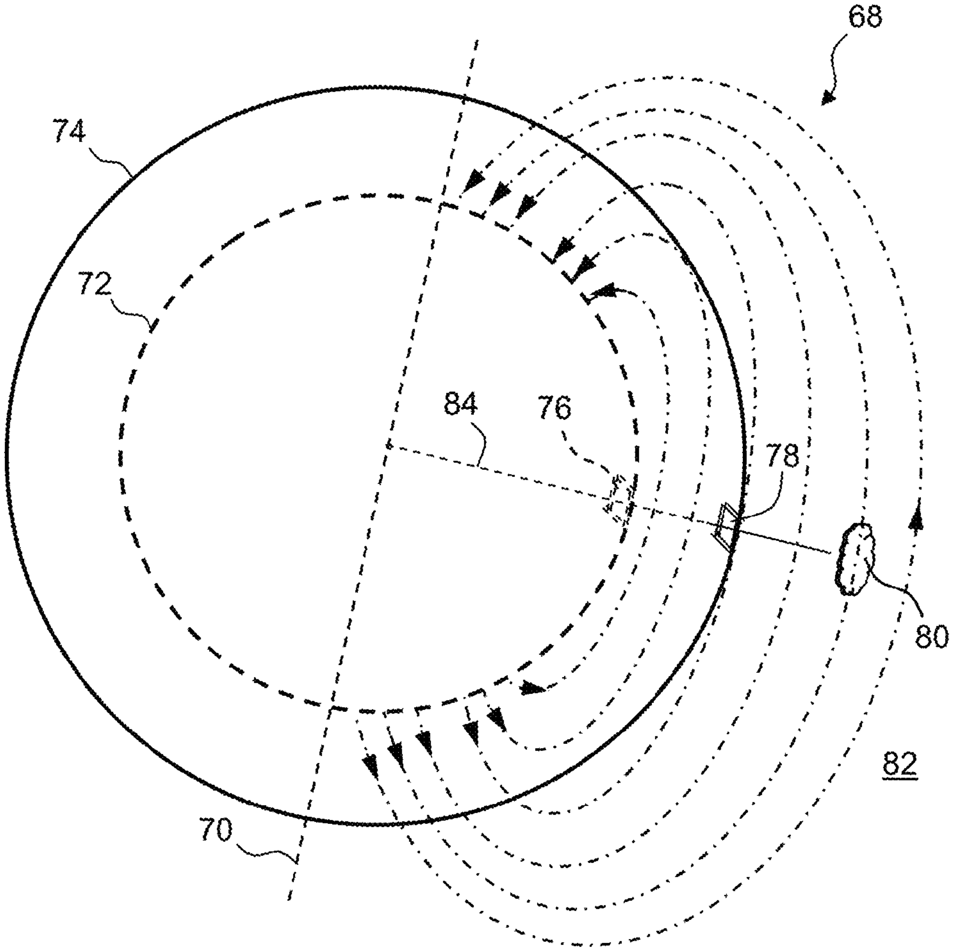
68
74
72
84
76
78
80
82
70

FIG. 5

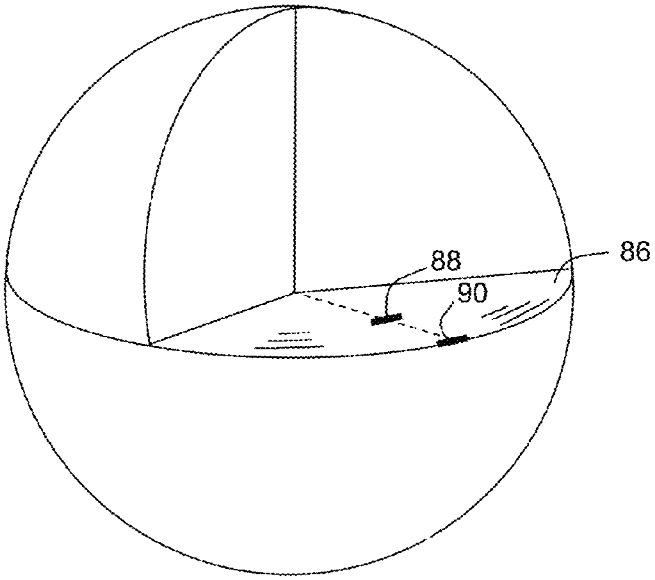

FIG. 6

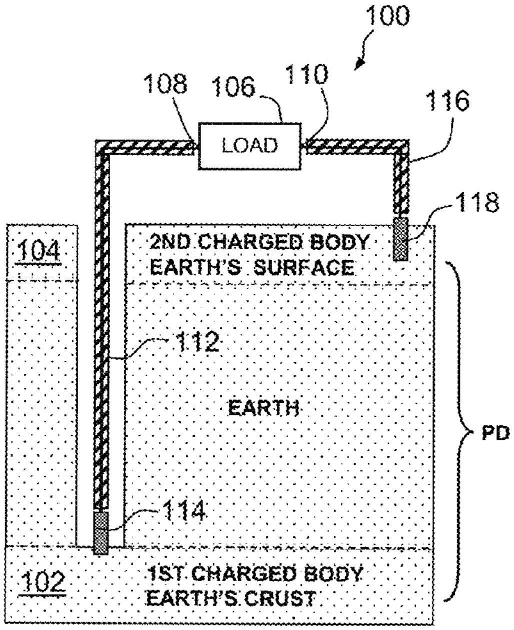
100
108
106
110
116
LOAD
118
104
2ND CHARGED BODY
EARTH'S SURFACE
112
EARTH
PD
114
102
1ST CHARGED BODY
EARTH'S CRUST

FIG. 7

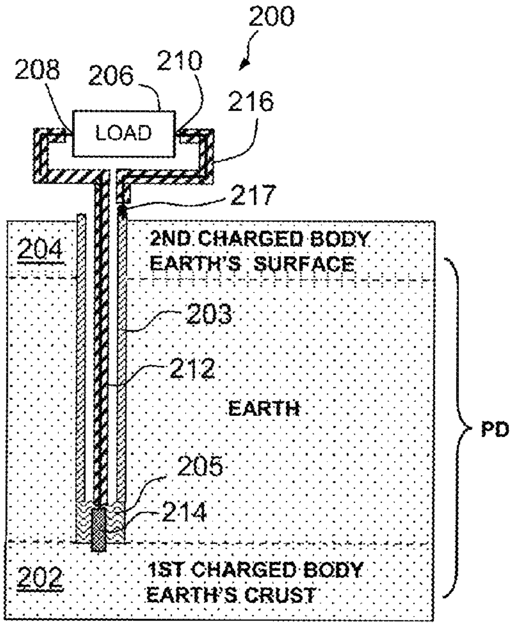
200
208
206
210
216
LOAD
217
204
2ND CHARGED BODY
EARTH'S SURFACE
203
212
EARTH
PD
205
214
202
1ST CHARGED BODY
EARTH'S CRUST

FIG. 8

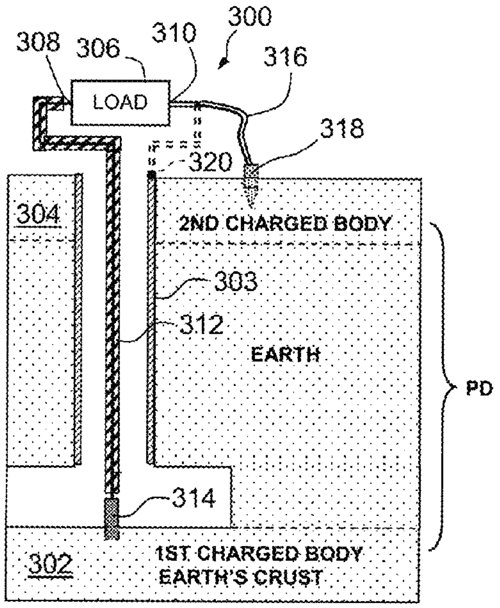
308
306
310
300
316
LOAD
318
320
304
2ND CHARGED BODY
303
312
EARTH
PD
314
302
1ST CHARGED BODY
EARTH'S CRUST

FIG. 9

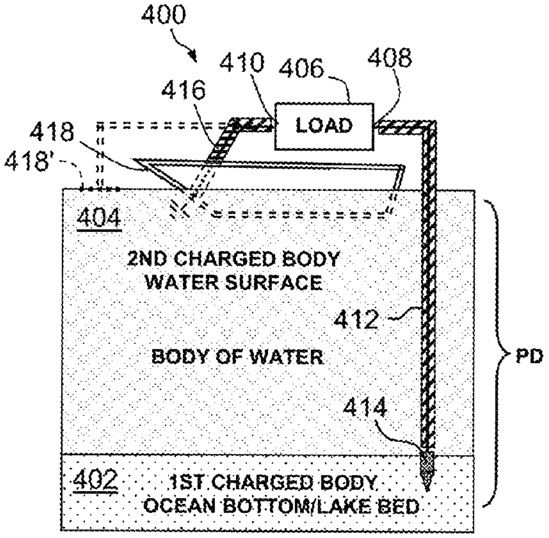
400
410
406
408
416
LOAD
418
418'
404
2ND CHARGED BODY
WATER SURFACE
412
BODY OF WATER
PD
414
402
1ST CHARGED BODY
OCEAN BOTTOM/LAKE BED

FIG. 10

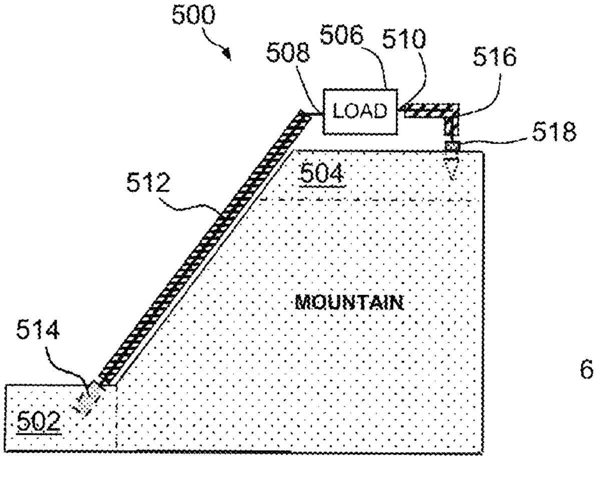
500
508
506
510
516
LOAD
518
512
504
MOUNTAIN
514
502

FIG. 11

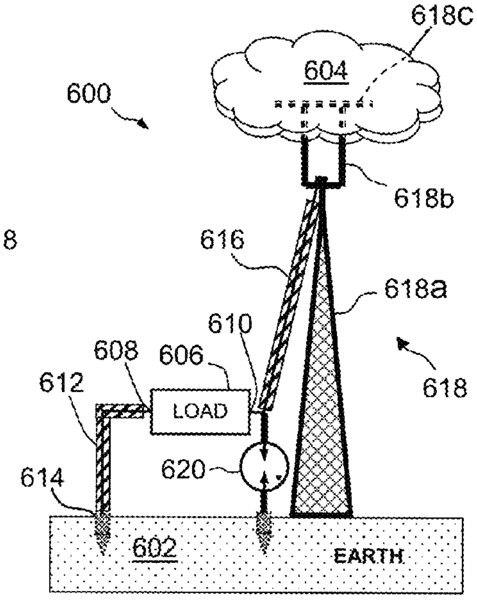
618C
604
600
618b
616
618a
610
608
606
612
LOAD
618
620
614
602
EARTH

FIG. 12

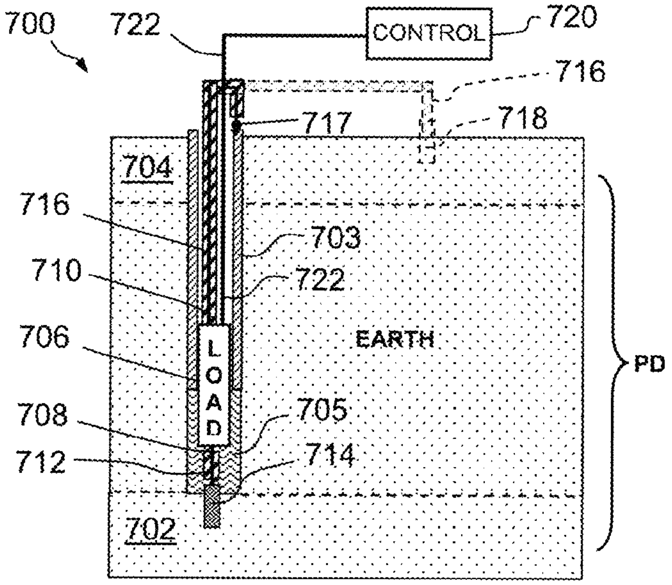
700
722
CONTROL
720
716
717
718
704
716
703
710
722
706
L
O
A
D
EARTH
PD
708
705
712
714
702

FIG. 13

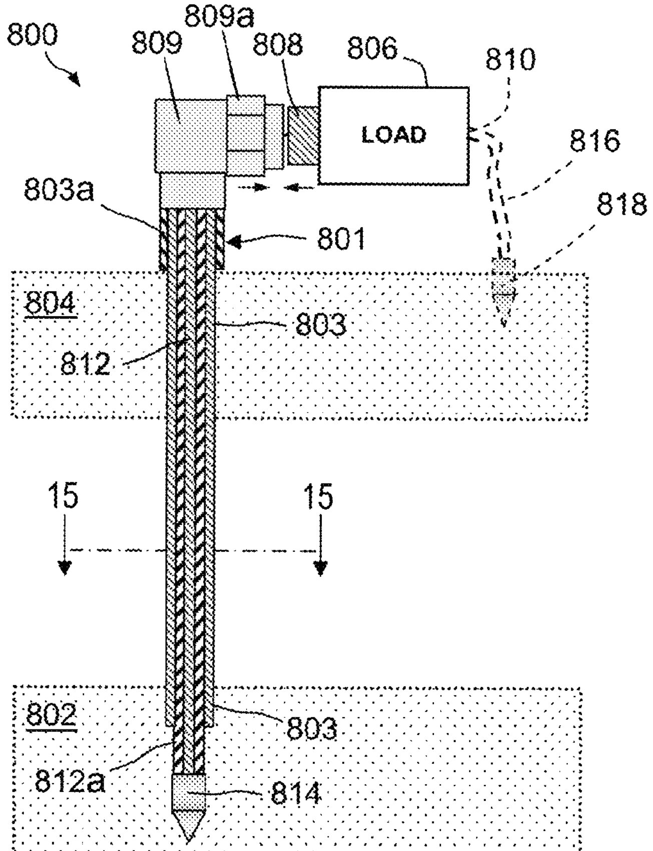
800
809
809a
808
806
LOAD
810
816
818
803a
801
804
803
812
15
15
802
803
812a
814

FIG. 14

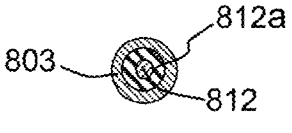
812a
803
812

FIG. 15

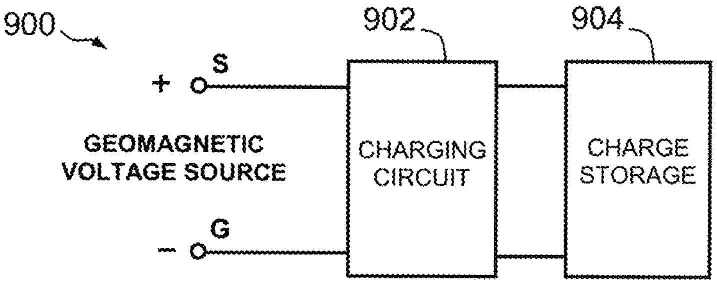
900
902
904
+ S
GEOMAGNETIC VOLTAGE SOURCE
− G
CHARGING CIRCUIT
CHARGE STORAGE

FIG. 16

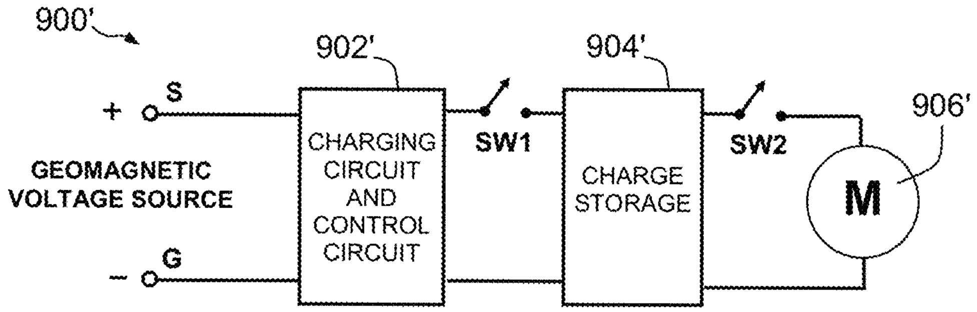
900'
902'
904'
906'
+ S
GEOMAGNETIC VOLTAGE SOURCE
− G
CHARGING CIRCUIT AND CONTROL CIRCUIT
SW1
CHARGE STORAGE
SW2
M

FIG. 17

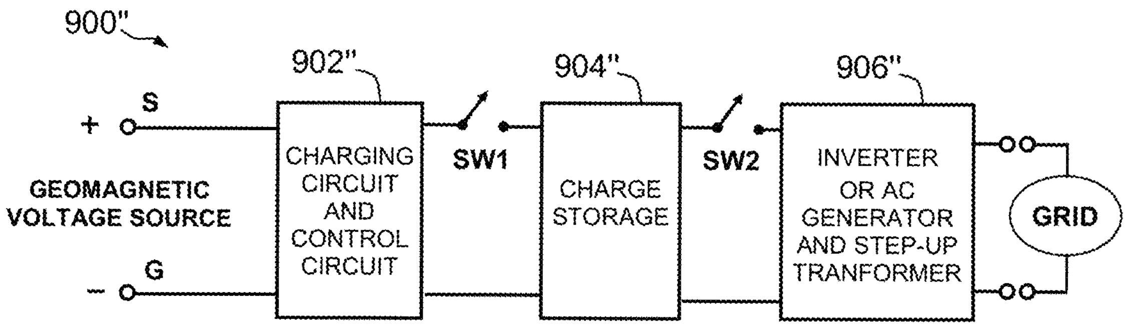
900"
902"
904"
906"
+ S
GEOMAGNETIC VOLTAGE SOURCE
− G
CHARGING CIRCUIT AND CONTROL CIRCUIT
SW1
CHARGE STORAGE
SW2
INVERTER OR AC GENERATOR AND STEP-UP TRANFORMER
GRID

FIG. 18

GENERATING ELECTRICITY FROM THE EARTH

BACKGROUND OF THE INVENTION

Field of Invention

The present invention relates generally to renewable electrical energy sources, and more particularly to apparatus and methods for generating electricity from the earth's magnetic field and rotation.

Background Art

The concepts of electromagnetic induction, Faraday's law, Faraday's disk, and the Lorentz force on charges are all well-known. These concepts have been applied to the earth, which is modeled as a large conductive sphere rotating through its own magnetic field ("geomagnetic field"). In this regard, the earth has been compared to a Faraday disk or homopolar generator. See, for example: W. M. Elsasser, The Earth as a Dynamo, Scientific American, Inc., Vol. 198, Issue 5, May 1, 1958, p. 44, 45; and C. R. Carrigan and D. Gubbins, The Source of the Earth's Magnetic Field, Scientific American, Inc., Vol. 240, Issue 2, Feb. 1, 1979, p. 118, 199-20. In making this comparison, the assumption has been carried over that the electricity to be obtained from the earth is from electromagnetic induction. Laboring under this assumption, it has been asserted that electricity from the geomagnetic field is not practicably retrievable, because the electricity gathering circuit would have to be stationary relative to the rotating earth (i.e., have a stationary frame of reference). This assertion is apparently based on a theory that the currents induced in the Faraday disk (or earth) and in the circuit employed to retrieve the electricity, would cancel each other out if both were moving through the magnetic field. Another theory is that a static electric field is created in the Faraday disk (or earth) that neutralizes the induced electric field, resulting in a zero voltage reading in the rotational frame of reference.

Further to this point, U.S. Patent App. Pub. No. 2004/0027022 (Feb. 12, 2004) to Weir, states that no voltage is generated by a Faraday type homopolar dynamo in the rotational reference frame of the dynamo (Weir, Par. [0010]). Weir even attempted to measure such voltage by mounting a voltmeter onto the rotating disk and connecting one probe to a shaft rotating the disk and the other probe to the rim of the disk. Weir states that the voltmeter read zero volts when the disk was rotating. (Weir Par. [0011].) No details are given regarding the sensitivity of the voltmeter or the test setup. In a book by Thomas Valone, Ph.D, P.E., entitled The Homopolar Handbook, published by Integrity Research Institute, Washington, D.C., 3rd Edition, October 2001 ("Valone I"), it is stated that an electrostatic field (created in response to the Lorentz force) neutralizes the electric field induced on the Faraday disk ("effective electric field"), and thus no voltage can be detected with a meter that rotates with the disk (Valone I, p. 7 & FIG. 8). Valone discusses his voltage measurement on page 24 and again at page 63. He indicates that his voltmeter was set to indicate a voltage reading only if the voltage was greater than 15 millivolts. Thus, the sensitivity of Valone's experiment may have been limited to 15 millivolts, presumably because the induced electric field (obtained in the stationary reference frame) was expected to be 100 millivolts (see Valone I, p. 63).

Valone reminds us that Michael Faraday, in 1831, compared his Faraday disk to the earth and the earth's rotation through its magnetic field. Valone states that Faraday, thinking that an induced voltage could be measured on his rotating disk, attempted to measure a like voltage in the earth (i.e., in rivers and streams) and apparently was unsuccessful. (Valone I, p. 61.) Valone characterizes Faraday's thinking as a mistake. Valone presumably mentions Faraday's unsuccessful attempt at measuring voltage in the earth as support for his position that a voltage cannot be measured in the rotating reference frame. In an earlier paper, entitled "The Homopolar Generator: Tesla's Contribution," Proceedings of the International Tesla Symposium, January 1986 ("Valone II"), Valone specifically stated that "we can't draw power from the earth's Homopolar generator while rotating with it." (Valone II, p. 273.)

Notwithstanding the view that a voltage is unobtainable in the rotating frame of reference in a Faraday system, proposals were made in the 1980s for a space shuttle to drag a tether (an insulated conductor) through space across the geomagnetic field to generate electric power onboard the shuttle. Current and a voltage potential difference would be induced on the tether's conductor as it passed through the magnetic field lines of the earth. The current return path (or ground path) would be the charged plasma medium of the ionosphere, which would exist between the outer wall of the space shuttle and the distal end of the tether (or an outer wall of a satellite attached to the distal end). A load (e.g., electronic equipment) inside the space shuttle would be connected between the proximal end of the tether and the outer wall of the shuttle (i.e., the return or ground path). A tether experiment was actually performed on a space shuttle mission in 1996, according to a NASA article (www.istp.gsfc.nasa.gov/Education/wtether.html). The voltage generated on the tether was 3500 volts, apparently measured in the rotational frame of reference. It is suggested that the voltage may have been measured at the distal end of the tether with instruments aboard a satellite connected to the distal end. The voltage level at the load inside the shuttle is not mentioned. In this experiment, the voltage applied to the intended load (equipment aboard the shuttle) was supplied by the tether's conductor and the return or ground path for the load was, according to NASA, the ionospheric plasma between the shuttle and the satellite.

An alternative design to the above-mentioned space tether was purportedly developed by or for the U.S. Army in 1988 and patented in U.S. Pat. No. 4,923,151 (1990) to Roberts. Roberts discloses a coaxial tether (Roberts, FIGS. 2 and 3) about 12.5 miles long, comprising an outer conductor 20, a magnetic shield conductor 26, and a center conductor 30. Outer conductor 20 is electrically connected at one end to a satellite 12 and at the other end to an orbiting shuttle 10 (Roberts, FIGS. 1 and 2). Center conductor 30 is magnetically shielded from the geomagnetic field by conductor 26, which is made of a high permeability metal such as soft iron. Center conductor 30 is electrically insulated from conductors 20 and 26 by insulation 28. Center conductor 30 is also electrically connected at one end to satellite 12 (at point 32) and at the other end to a load 36 inside shuttle 10 (through a switch 34). Load 36 is grounded via a return current path including the outer surface 40 of shuttle 10, conductor 20, and the outer surface 22 of satellite 12. The tether passes through the geomagnetic field at an orientation substantially normal to the earth and its magnetic field lines. As a result, a potential difference ("PD") is produced between shuttle 10 and satellite 12. Conductor 26 magnetically shields center conductor 30, so no induced current flows in conductor 30. Conductor 30 assumes the charge potential of satellite 12 and applies the potential difference PD across load 36 (once switch 34 is closed), causing current to flow in load 36. Conductor 20 replaces the ionosphere as the return current path, purportedly making the circuit more stable. Thus, conductor 20 functions both as a Faraday conductor (i.e., a conductor on which currents are induced from a magnetic field) and as the return current path. In this case, the Faraday conductor (conductor 20) does not function as the supply current path (conductor 30 does). In the previously discussed space tether design, the Faraday conductor (i.e., the tether's only conductor) did serve as the supply current path and did not serve as the return current path (the ionosphere).

For over 40 years, the inventor named herein has been working on an idea for generating electricity from the earth's rotation and magnetic field. The argument that electricity is generated in the earth is supported by some natural phenomena, including: the decreasing speed of the earth's rotation, the continued molten state of the earth's interior, volcanic and geyser activity, the earth's magnetic field, and lightning. The earth uses its primary kinetic energy, which originally started its rotation, to electrically charge itself by electromagnetic induction. Electric energy flowing toward the center of the earth is transformed into magnetic and heat energy, sustaining the earth's magnetic field and the molten state of the earth's interior. The balance in pressure between the surface and the molten interior is maintained through the loss of heat energy at the surface by heat radiation and volcanic and geyser activity. As a result of the work performed by the earth, and possibly because of a counter electromagnetic force (Lenz's law), the earth's kinetic energy is decreasing and its rotation is slowing.

The inventor herein began thinking about his idea while observing thunderstorms. In 1979, during one of his experiments, he was struck by a branch of lightning and luckily survived with no significant injury. He decided to continue experimentation, but in a safe manner. Upon arrival to the United States, the inventor herein applied for a grant and technical support from the National Institute of Standards and Technology (NIST) on May 18, 1995. His application was assigned an Evaluation Request (ER) Number 32042 and a Control Number 015351. The application was ultimately rejected with an explanation that the magnetic field rotates with the earth and thus currents could not be induced in a conductor installed on the earth. NIST's rejection of the idea dissuaded the inventor herein from pursuing the idea further, until recently when the inventor started thinking about the Faraday disk and "Faraday's paradox."

Faraday's paradox is the non-intuitive result that electric currents are induced in the Faraday disk even when the magnet rotates with the disk. It is non-intuitive because many would assume that the magnetic field rotates with the rotating magnet and disk, in which case currents would not be induced in the disk. Faraday conducted three tests to investigate this phenomenon. First, he rotated the disk and kept the magnet stationary and observed a current induced in the disk. Second, he kept the disk stationary and rotated the magnet and observed no induced current in the disk. Third, he rotated the disk and magnet together and observed a current in the disk. From these tests, Faraday concluded that the magnetic field remained stationary when the magnet was rotated around its own magnetic axis. Similarly, the earth (representing a rotating disk and magnet) should produced a stationary magnetic field through which the earth rotates, thus inducing generally radial electric currents in the earth and on conductors substantially aligned with the earth's radius. Since Faraday, there has been a long-standing debate and experimentation about Faraday's paradox and whether or not the magnetic field rotates with the magnet. For example, see Valone I, pp. 2-4 and 37. Notwithstanding the theoretical debate and investigation, it is believed that Faraday was correct and that currents are induced by the geomagnetic field in the earth, in the surrounding atmosphere, and in conductors appropriately located on the earth and in the atmosphere.

Nikola Tesla disclosed a conductor extending into the atmosphere and normal to the earth's surface in U.S. Pat. No. 685,957 (Tesla). Rather than seek to induce currents in the conductor from the geomagnetic field, Tesla sought to collect radiant energy (or particles charged by radiant energy) primarily from the sun. Tesla somehow suspended a conductive plate P and connected it to the distal end of the conductor (Tesla, FIG. 1). The proximal end of the conductor was connected to a capacitor C. The other end of capacitor C was connected to a secondary conductor, which is, in turn, connected to a conductive plate P' grounded in the earth. Plate P was the collector of radiant energy charged particles. The charge from plate P charged capacitor C. A voltage threshold device (d) and a load R are connected across capacitor C to form a circuit. Device (d) closes the circuit when the charge on capacitor C exceeds a predetermined voltage, causing the voltage to be applied to load R. In this patent, Tesla does not seek to obtain the potential difference between two "terrestrial charged bodies" (hereinafter defined).

Herein, the term "terrestrial" is used in a broad sense, to include the whole or any part of the earth and its surrounding atmosphere, including but not limited to: the earth's inner and outer cores, mantle, crust, any layer or level thereof, and its surface; all bodies of water; air; clouds; ice; vegetation; minerals and other resources such as oil; rocks including basalt; or any other composition or thing of the earth and its atmosphere. As used herein, the term "terrestrial charged body" (or its plural form) means a terrestrial body, region, layer, zone, other volume or area, any composition or thing, with or without specifically defined or fixed boundaries, having an electric charge represented by a voltage potential.

OBJECTS AND SUMMARY OF THE INVENTION

It is therefore an object of the present invention to overcome the problems and limitations associated with the prior art and to fulfill a long-felt but unsolved need for a renewable source of energy.

It is another object of the present invention to provide a clean and renewable source of electrical energy.

It is a further object of the present invention to generate electricity from the earth's magnetic field and rotation.

It is yet another object of the present invention to provide a reliable source of electricity from the earth's magnetic field and rotation.

It is yet a further object of the present invention to provide a source of electricity from the earth's magnetic field and rotation that can be easily implemented.

It is still another object of the present invention to provide a source of electricity from the earth's magnetic field and rotation that can be implemented at a low-cost.

These and other objects are attained in accordance with the present invention, wherein there is provided a system for generating electricity from the geomagnetic field and rotation of the earth. The earth rotates through the geomagnetic field to form a potential difference between first and second terrestrial charged bodies. The first and second terrestrial charged bodies are spaced apart in a direction substantially normal to the earth's surface. The system comprises an electrical load, an electrically insulated first current path, and a second current path. The electrical load has a supply input and a ground output. The first current path is coupled to the first terrestrial charged body and to the supply input of the electrical load, for supplying the potential difference to the electrical load. The second current path is coupled to the ground output of the electrical load and to the second terrestrial charged body. As a result of this arrangement, an electric current related to the potential difference is generated in the electrical load.

In one particular embodiment, the first terrestrial charged body includes a first portion of the earth located below the earth's surface and the second terrestrial charged body includes a second portion of the earth located at or near the earth's surface. The earth contains an opening extending between the first and the second portions of the earth. In this embodiment, the first current path includes (i) an electrically insulated conductor extending at least partially through the opening and (ii) a conductive contact member coupled to the electrically insulated conductor for establishing electrical contact between the first portion of the earth and the electrically insulated conductor. In a more specific embodiment, the opening is a well bore or mine shaft, and the first portion of the earth includes a bottom of a well or mine. The electrically insulated conductor extends through the bore or mine shaft, and the conductive contact member establishes electrical contact between the bottom of the well or mine and the electrically insulated conductor.

In another embodiment, the first terrestrial charged body may be a portion of the earth at or near the surface of the earth, and the second terrestrial charged body may be a cloud above the surface of the earth. In yet another embodiment, the first and second terrestrial charged bodies may be first and second charged portions of the earth, respectively, where both portions are located at or near the surface of the earth. In this latter embodiment, the first portion of the earth is at a first elevation and the second portion of the earth is at a second and different elevation. For example, the first elevation may be at or near the bottom of a mountain or cliff and the second elevation may be at an elevated position on the mountain or cliff. In still another embodiment, first terrestrial charged body is a portion of the earth at the bottom of a body of water, such as the ocean, and the second terrestrial charged body is the water at or near the surface of the body of water.

In any of the embodiments, the electrical load may include a rechargeable energy source, including a rechargeable battery or battery pack and a charging circuit coupled to the rechargeable battery or pack. Alternatively, or in addition to batteries, the rechargeable energy source may include a charging capacitor or a bank of charging capacitors along with an appropriate charging circuit for the capacitor(s). The capacitor or capacitors may be ultracapacitors. The electrical load may further include an electric motor coupled to the rechargeable energy source. The electric motor may be a DC motor, which may be mechanically coupled to an AC generator. In operation, the DC motor would actuate the AC generator, resulting in the AC generator generating AC electric power. Alternatively, the electric load may include an inverter for converting DC electric power to AC electric power. In either case (AC generator or inverter), the AC electric power can be stepped up in voltage and applied to an electric power grid.

A method of generating electricity from the geomagnetic field and rotation of the earth is also within the scope of the present invention. As indicated earlier, the earth rotates through the geomagnetic field to form a potential difference between first and second terrestrial charged bodies. The first and second terrestrial charged bodies are spaced apart in a direction substantially normal to the earth's surface. The method comprises the steps of: (a) providing an electrically insulated first current path having first and second electrically exposed ends; (b) coupling the first end of the first current path to the first terrestrial charged body to establish electrical communication between the first current path and the first terrestrial charged body; (c) coupling the second end of the first current path to a supply input of an electrical load to supply the potential difference to the electrical load; (d) providing a second current path having first and second ends; (e) coupling the first end of the second current path to a ground output of the electrical load; and (f) coupling the second end of the second current path to the second terrestrial charged body to establish electrical communication between the electrical load and the second terrestrial charged body. As a result of these steps, an electric current related to the potential difference is generated in the electrical load.

BRIEF DESCRIPTION OF THE DRAWING

Further objects of the present invention will become apparent from the following description of the preferred embodiments, with reference to the accompanying drawing, in which:

FIG. **1** is a diagram of a test setup used to demonstrate the operation of a Faraday disk and the generation of electricity within the rotational frame of reference;

FIGS. **2** is a diagram of an experiment showing a rotating Faraday disk with a voltmeter having lead wires and test leads for measuring voltages at different points along the disk's radius;

FIG. **3** is a diagram of an experiment illustrating a battery analogy of the Faraday disk and the concept of short circuit (SC) current on the disk;

FIG. **4** is a diagram of a field experiment with one embodiment of the present invention, implemented in conjunction with a water well;

FIG. **5** is a diagram of the earth rotating through its magnetic field and showing (schematically) three terrestrial charged bodies spaced apart from each other in a direction normal to the earth's surface;

FIG. **6** is a diagram of the earth partially dissected to reveal a portion of a flat circular plane through the earth's equator, presented as an analogy of a Faraday disk;

FIG. **7** is a diagram of a first embodiment of the present invention, wherein the terrestrial charged bodies are separated portions of the earth;

FIG. **8** is a diagram of a second embodiment of the invention, which is an implementation utilizing a well;

FIG. **9** is a diagram of a third embodiment of the invention, which is an implementation utilizing a mine and a mine shaft;

FIG. **10** is a diagram of a fourth embodiment of the invention, which is an implementation utilizing a body of water, wherein one terrestrial charged body is the bottom of the body of water and another terrestrial charged body is a surface of the body of water;

FIG. **11** is a diagram of a fifth embodiment of the invention, which is an implementation utilizing different elevations of the earth's surface, such as associated with a mountain or cliff;

FIG. **12** is a diagram of a sixth embodiment of the invention, which is an implementation utilizing the atmosphere of the earth, wherein one terrestrial charged body is a cloud and another terrestrial charged body is a surface of the earth;

FIG. **13** is a diagram of a seventh embodiment of the invention, which is an implementation utilizing a well having a bore casing and an electrical load (which may include a submersible pump) is located at least partially within the bore casing;

FIG. **14** is a diagram of an eighth embodiment of the invention, which employs coaxial conductors to carryout the invention;

FIG. **15** is a cross-sectional view of the coaxial conductors of FIG. **14**, taken along line **15-15** in FIG. **14**;

FIG. **16** is a block diagram of one example of an electrical load of the present invention, which is a rechargeable energy source including a charging circuit and a charge storage device, such as a rechargeable battery(s) or storage capacitor (s);

FIG. **17** is a block diagram of another example of an electrical load of the invention, which includes a charging and control circuit, charge storage device, and an electric motor; and

FIG. **18** is a block diagram of a further example of an electrical load of the present invention, including a charging and control circuit, a charge storage device, and an inverter or AC generator with step-up transformer.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Inspired by the book "The Homopolar Handbook" by Thomas Valone (Valone I), the named inventor built a Faraday disk to perform a few original experiments. A direct consequence of the experiments is the extraction and use of one part of the existing electrical energy of the earth. FIG. **1** illustrates a Faraday disk apparatus **10** built by the inventor herein. Apparatus **10** includes a frame **12** made of 2×6 wooden boards, connected together with screws. An axle **14** is made of a ⅝ inch diameter, 3-foot long steel threaded rod. Axle **14** has a sharpened lower end **14***a*, which rests on a steel plate **16**. Axle **14** is rotatably supported by two bearing assemblies **18***a* and **18***b*. Bearing assembly **18***a* is mounted in an upper cross member **12***a* of frame **12** and bearing assembly **18***b* is mounted in a bottom cross member **12***b* of frame **12**. A drill **20** is connected, via a chuck **22**, to a short rod **24**, which is, in turn, connected to one end of an insulator block **26**. The upper end of axle **14** is connected to the other end of insulator block **26**. Rod **24**, block **26** and axle **14** rotate together as a unit. The speed of drill **20** is manually controlled to control the rotational speed of axle **14**. Insulator **26** electrically isolates axle **14** from the drill to eliminate any chance of electric current from drill **20** finding its way into axle **14**.

As further shown in FIG. **1**, the actual Faraday disk comprises a 17-inch diameter round aluminum baking pan **28** and two 16-inch diameter (and ¼-inch thick) steel disks **30***a* and **30***b*. Pan **28** and disks **30***a*, **30***b* are fixed together and balanced on axle **14**, using washers and steel hex nuts threaded on axle **14**. Disk **30***a* is fixed in and against the bottom of pan **28** and disk **30***b* is fixed directly underneath and against pan **28** (collectively referred to as disk assembly **28'**). Axle **14** and disk assembly **28'** rotate together, almost like a spinning top. In order to reduce electrical resistance associate with axle **14** above disk assembly **28'**, a copper ring **32** is clamped to axle **14** and an insulated copper wire **34** is connected between ring **32** and disk assembly **28'**. Two spring-loaded graphite brush contacts **34***a* and **34***b* are installed. Brush contact **34***a* is adjacent to and in contact with copper ring **32** and brush contact **34***b* is adjacent to and in contact with the rim of disk assembly **28'**. Contact **34***a* and a lead wire **36***a* electrically connect ring **32** to one input of a digital voltmeter **38**. Contact **34***b* and a lead wire **36***b* electrically connect the rim of disk assembly **28'** to a second input of voltmeter **38**. Voltmeter **38** is mounted outside of frame **12** and is held stationary. Another digital voltmeter **40** is mounted on disk assembly **28'** and rotates with assembly **28'**. Voltmeter **40** is electrically connected, via a short lead wire **42***a*, to the rim of disk assembly **28'** and is electrically connected, via a longer lead wire **42***b*, to axle **14**. A digital camera **44** is fixed on axle **14** (just below insulator block **26**) to record the digitally displayed readings of voltmeter **40**.

Again referring to FIG. **1**, twenty-two cylindrical welding magnets, each five inches in diameter and each with a pull strength of 75 lbs, are arranged to form two magnetic cylinders **46***a* and **46***b* (11 magnets each). Cylinders **46***a* and **46***b* are separated from each other by disk assembly **28'**. Magnetic cylinders **46***a*, **46***b* are installed in a way that allows them to be held stationary or to be rotated with disk assembly **28'**. In a series of tests, axle **14** was rotated at a speed of about 200 RPM and voltmeter **38** measured and read a generated potential difference of about 300 millivolts. No difference in voltage generation was noted between tests where magnetic cylinders **46***a*, **46***b* were held stationary or rotated (Faraday's paradox). During the 200 RPM tests, voltmeter 40 read 1 to 3 millivolts, supporting the view that a voltage potential difference is generated and measurable on the rotating frame of reference. These tests are referred to collectively as "Experiment A." The difference in readings between voltmeter **38** and voltmeter **40** will be explained below.

Referring now to FIG. **2**, a diagram of a rotating disk experiment is shown. Disk assembly **28'** and voltmeter **38** are used in this experiment. Lead wires **36***a* and **36***b* of voltmeter **38** include test leads **48***a* and **48***b*, respectively. Leads **48***a*, **48***b* are placed on rotating disk assembly **28'** in a sliding engagement to establish dynamic electrical contact with assembly **28'**. As disk assembly **28'** was rotated, voltage measurements were taken at different points along the radius of the assembly (i.e., at different imaginary concentric circles on disk assembly **28'**). Voltmeter **38** measured the potential difference between the points of contact. The positive side of the potential difference is toward the center, as indicated in FIG. **2**, due to the Lorentz force pushing electrons radially out to the rim. The measured potential difference increased as the distance between the contact points increased. This experiment is referred to hereinafter as "Experiment B."

An understanding as to why there is a difference in electric potential measured by voltmeter **38** and voltmeter **40** (FIG. **1**) was gained in another experiment, illustrated in FIG. **3**. In this experiment, an attempt was made to model the electrodynamics on a rotating Faraday disk. This was done by applying a potential difference across disk assembly **28'** (between axle **14** and the rim), while disk assembly **28'** remained stationary (as did voltmeter **38**). This setup was intended to mimic the case where the voltmeter and disk are rotating in the same frame of reference. The potential difference was applied across disk assembly **28'** using a 6-volt battery **50**. This modeled the potential difference generated on a rotating disk by electromagnetic induction. Insulated copper wires **52***a* and **52***b* were connected to the terminals of battery **50** and to axle **14** and the rim of disk assembly **28'**, as shown in FIG. **3**. The positive side of battery **50** is connected at axle **14** and the negative side is connected at the rim to model the charge displacement due to induction. Lead wires **36***a*, **36***b* of voltmeter **38** were also connected to disk assembly **28'**, in the vicinity of the connections of wires **52***a*, **52***b*, respectively. The voltage (or potential difference) measured by voltmeter **38** was 50 millivolts. A reading of 6 volts would be expected in an open circuit measurement. On the other hand, a reading of 50 millivolts would suggest a near short circuit (considering the inherent resistance in disk assembly **28'**). Voltmeter **38** did measure a short circuit current, which is indicated in FIG. **3** by arrow SC. This experiment is referred to hereinafter as "Experiment C."

Based on Experiments A, B and C, it was concluded that the rotation of disk assembly **28'** in the magnetic field (created by magnetic cylinders **46***a*, **46***b*—FIG. **1**) caused the disk assembly to become charged and develop a potential difference in the radial direction. The potential difference increased as the distance between the center of disk assembly **28'** and a measurement point along the radius of assembly **28'** increased. This potential difference was created by the Lorentz force pushing electrons out to the rim, causing an induced current to flow radially outward. In response to the induced potential difference, electrons seek to establish equilibrium on the disk (as it is a conductor) and begin to flow in a direction opposite of the induced current. This opposite flow of electrons is referred to as the short circuit current and is indicated in FIG. **3** by arrow SC. It is believed that this short circuit current is flowing at all times on a rotating Faraday disk. This may be the reason why a voltage is measureable on the rotating frame of reference (voltmeter on the disk), albeit a smaller voltage than the induced voltage measured in the stationary frame of reference. A rotating Faraday disk (i.e., disk assembly **28'**) behaves like a battery (or multiple batteries connected in series) with one pole at the center of the disk (axle **14**) and the other pole at the rim. While rotation in the magnetic field causes the disk to be charged, a short circuit simultaneously occurs. The disk loses its charge immediately when it stops rotating. A Faraday disk (and thus disk assembly **28'**) can be thought of as comprising a number of imaginary annular rings, each behaving like a circular-shaped battery and the batteries are connected in series. Thus, the largest potential would occur at the rim, where all of the annular rings (batteries) of the disk are in series and in the circuit.

The concepts studied in the lab in Experiments A, B and C were then applied in the field, with respect to the earth's rotation and magnetic field. FIG. **4** illustrates an experiment that was conducted at two different drilled water wells, one 117 feet deep and the other 210 feet deep. The wells were drilled at construction sites of new homes in a suburban area near Utica, N.Y. At the time of the experiments, the wells were not in use. There were no electric power lines or transmitting antennas within a great distance around the construction sites. A copper pipe **54**, one-half inch in diameter and two feet long, was connected to an insulated copper wire **56**, to ensure a good connection with the earth at the bottom of the well. Pipe **54** and wire **56** were lowered to the bottom of each well. Pipe **54** made direct contact with the earth at the bottom of each well. The bottom of each well is considered a first charged body (defined above). Another one-half inch diameter and two-foot long copper pipe **58** was hammered into the ground at the surface, about 6 feet from the well. The surface and earth surrounding pipe **58** is considered a second charged body. A digital voltmeter **60** was used to measure the difference in electric potential between the bottom of the well and the surface of the ground (i.e., between the first and the second charged bodies). Voltmeter **60** has test lead wires **62** and **64**. Test lead wire **62** is connected to insulated wire **56** and test lead wire **64** is connected to copper pipe **58**. Steady measurements of **592** millivolts were obtained in the 117-foot well and 858 millivolts were obtained in the 210-foot well. Short circuit current between the surface and the bottom of the well was also measured with meter **60**. As the distance between the surface and bottom of the well increased, the difference in electric potential increased.

A variation of the setup in FIG. **4** was also tested at the 210-foot well (indicated in dash lines). The well includes a metal bore casing **66** that extends substantially down the well bore. The variation is that casing **66** was used as part of the circuit instead of copper pipe **58**. Test lead wire **64** was connected directly to casing **66**. The voltage measurements were consistent with those obtained in the original tests for the 210-foot well. It is believed that the use of casing **66** as part of the circuit, instead of a ground pipe (like pipe **58**), may yield more consistent voltage readings. The tests performed at both wells, as described in this and the previous paragraph, are collectively referred to hereinafter as "Experiment D." In the tests of Experiment D, wire **56** was an insulated 14-Gauge Stranded THHN Cable, Model No. 22955957, supplied by the Southwire Company, Carrollton, Ga., featuring conductors made of annealed (soft) copper, rated at 600 Volts, and insulated with tough, heat- and moisture-resistant PVC. Voltmeter **60** was a Mini Pocket MultiMeter, Model DM110, by Extech Instruments Corp., Nashua, N.H., with 0.1 millivolt DC voltage resolution and 0.01 milliamp DC current resolution. Another voltmeter was also used to re-confirm the readings. That meter was a Cen-Tech Seven Function Digital Multimeter, Model 98025, supplied by Harbor Freight Tools Co., Camarillo, Calif., with milliamp and millivolts resolutions.

The earth is an oblate spheroid with an equatorial radius of 6378.16 km. The earth and its surrounding atmosphere rotate within the earth's magnetic field (i.e., the geomagnetic field). The speed of a point at the earth's equator is about 463.8 m/sec. In the Utica, N.Y. area, where Experiment D was performed, the speed is about 300 m/sec. By comparison, the speed of sound is 340.29 m/sec. The earth's magnetic field is estimated to be about 0.35 to 0.65 gauss at the earth's surface. By comparison, a strong refrigerator magnet has a field strength of 100 gauss. The geomagnetic field is generally dipolar as it has two poles, the north and south magnetic poles, and the field lines resemble those of a dipolar magnet. FIG. **5** illustrates the geomagnetic field on one side of the earth's magnetic axis **70**. Generally, the earth can be considered a conductor having the following five parts: (1) the solid inner core, which is an excellent conductor; (2) the molten outer core, which is an excellent conductor; (3) the mantle, much of which is a good conductor; (4) the crust, which is a relatively good conductor; and (5) the atmosphere, which can be a good or poor conductor.

To help with the understanding of the results obtained in Experiment D and with the operation of the present invention, consider the following theory. (This theory should not be construed as defining or limiting the scope of the present invention in any way, and it should not affect the validity of the invention if proven to be incorrect. It is presented in the hope that it may aid in the understanding of the invention.) One should picture the earth and the surrounding (rotating) atmosphere as an endless number of concentric spheres, from the earth's center to the earth's surface, and from the earth's surface to the outer reaches of the surrounding, rotating atmosphere. FIG. **5** depicts two spheres associated directly with the earth itself. A sphere **72** represents any level or layer under the earth's surface between the center and the surface. For example, sphere **72** could be (generally) a junction between the earth's crust and its mantle, a junction between the earth's mantle and the outer core, or a particular layer within the earth's crust or mantle. A sphere **74** represents the earth's surface. Alternatively, one can picture the earth as an endless number of disks, with the largest diameter disk cutting through the equator (see FIG. **6**) and the smallest diameter disks approaching the poles.

As shown in FIG. **5**, the geomagnetic field (showing one side of the field) comprises a plurality of magnetic field lines **68** disposed generally in a dipolar arrangement about magnetic axis **70**. As the earth rotates through the geomagnetic field, it behaves similar to a rotating Faraday disk, in that electrons dislocate outward from the center to the surface. This causes a potential gradient to be established along the radius of the earth and into the atmosphere, with the positive side toward the center and the negative side toward the surface. A potential difference is measured between any two designated spheres of the earth and atmosphere (e.g., spheres **72** and **74** in FIG. **5**), and the potential difference increases with the spacing between the designated spheres. At the same time, electrons travel back toward the center of the earth as the electromagnetic system seeks equilibrium, and this causes a permanent short circuit current to flow in the earth's interior.

Continuing with the theory, induction also exists in the atmosphere, especially in the clouds. Clouds are connected to the earth's surface by air that is generally a poor conductor. Some short circuit current through dry air may exist, but a big difference in electric potential develops between clouds and the earth's surface. During storms, the air between the earth's surface and clouds, and between clouds at different levels (representing different spheres), become saturated with water and form columns that are relatively good conductors. An intense short circuit current occurs through such conductors. When the amount of current becomes so intense that the conductor cannot accept it any longer, because of its properties (air and water), it simply burns out. The flash and explosion (i.e., lightning and thunder) are results of the burning out of the overloaded conductor. Most likely, a similar event happened to NASA's tether experiment, when an almost 20 km long insulated wire burned out by an estimated 1 amp of current at 3500 volts (see NASA tether experiment article, www.istp.gsfc.nasa.gov/Education/wtether.html).

Further with the theory, the earth behaves like one huge battery (similar to Faraday's disk), constantly being charged by the rotation within its own magnetic field. One pole of the battery is at the earth's center and the other pole is at the earth's surface or at some designated point in the atmosphere. One can also imagine the earth as an endless number of sphere-shaped batteries connected in series. At the same time, a permanent short circuit in the earth exists, like the battery in Experiment C above.

It is not possible or practicable to obtain the electric potential between the center of the earth and its surface using Faraday's approach, because stationary brush contacts cannot be placed at the center and the surface (i.e., a stationary frame of reference for the collection circuit cannot be established). However, it is possible to obtain an electric potential from the short circuit current between different designated spheres reachable from the earth's surface (e.g., Experiment D).

Referring again to FIG. **5**, sphere **72** contains a designated terrestrial charged body **76** and sphere **74** (earth's surface) contains a designated terrestrial charged body **78**. A cloud **80** in the atmosphere **82** is also designated as a terrestrial charged body. Terrestrial charged bodies **76**, **78** and **80** are spaced apart from each other in a direction substantially normal to the earth's surface. An imaginary axis **84** represents the alignment of the charged bodies in a direction normal to the surface. Due to electromagnetic induction, a potential difference is formed between terrestrial charged bodies **76** and **78**, between charged bodies **76** and **80**, and between charged bodies **78** and **80**. A terrestrial charged body is that amount (e.g., area or volume) of earth, water, minerals, rocks, clouds, etc. sufficient to define, accumulate or hold a charge which can produce a measurable potential difference with another terrestrial charged body.

Referring to FIG. **6**, there is shown a model of the earth with a portion removed to reveal an imaginary plane or disk **86**. Disk **86** cuts through the equator. FIG. **6** illustrates an analogy between the earth's interior and a Faraday disk (see earlier discussion above). In this example, an interior section or region of the earth's crust or mantle serves as a first terrestrial charged body **88**, and the earth's surface serves as a second terrestrial charged body **90**. Compare FIG. **6** with FIG. **2**. The rotation of the earth through its magnetic field causes terrestrial bodies **88** and **90** to become charged and a potential difference to be form between them, as in Experiment B (FIG. **2**).

Referring now to FIG. **7**, a system **100** for generating electricity from the geomagnetic field and rotation of the earth is shown as a first embodiment of the present invention. As explained, a potential difference (PD) is formed between terrestrial charged bodies—here terrestrial charged bodies **102** and **104**. In this embodiment, body **102** is underground earth (e.g., rock, soil, and/or minerals, etc.) at the bottom of an opening or hole (man-made or natural) and body **104** is the surface soil of the earth. Charged bodies **102**, **104** are spaced apart in a direction normal to the earth's surface. System **100** comprises an electrical load **106** having an input **108** and an output **110**. A first insulated conductor **112** is coupled at one end to charged body **102** via a contact member **114** and is connected at its other end to load input **108**. Contact member **114** may be, e.g., a short (e.g., 2 ft long) piece of copper or aluminum pipe. Conductor **112** is insulated to avoid direct electrical contact with the earth, except for charged body **102**. Conductor **112** is soldered or otherwise electrically connected to contact member **114**. The other end of conductor **112** is either directly connected to load input **108** (as shown) or is coupled to input **108** via a jumper wire and/or a connector. In FIG. **7**, a first current path is defined between load **106** and body **102** by contact member **114** and conductor **112**. The first current path supplies the PD to load **106**. A second insulated conductor **116** is coupled at one end to charged body **104** via a contact member **118** and is connected at the other end to load output **110**. Contact member **118** may be the same type as contact member **114**. Conductor **116** may be the same type and model as conductor **112**. A second current path is defined between load **106** and body **104** by conductor **116** and contact member **118**. A return current path occurs between charged bodies **104**, **102** via the intervening earth. Thus, a complete electrical circuit is established from charged body **102**, contact member **114**, conductor **112**, load **106**, conductor **116**, contact member **118**, charged body **104**, and the return current path back to charged body **102**. An electric current related to the PD is generated in load **106**. Load **106** may comprise a charging circuit and a charge storage device, a motor, and/or an AC generator or inverter, as discussed later. Contact members **114**, **118** may be any form of electrical conductor, including but not limited to a plate, disk, sphere, cone, rod, cylinder (hollow or solid), or sharp pointed member or members.

Referring now to FIG. **8**, a system **200** for generating electricity from the geomagnetic field and rotation of the earth is shown as a second embodiment of the present invention. A potential difference PD is formed between spaced apart terrestrial charged bodies **202** and **204**. Body **202** is underground earth at the bottom of a well and body **204** is the surface soil of the earth. The well includes a drilled wellbore extending from the earth's surface to the well bottom. A metal casing **203** extends down into the bore. At the bottom of the well is a supply of water **205**. System **200** comprises an electrical load **206** having an input **208** and an output **210**. A first insulated conductor **212** is coupled at one end to charged body **202** via a contact member **214** and is connected at its other end to load input **208**. Contact member **214** may be submerged in water **205**. Conductor **212** is insulated to avoid direct electrical contact with metal casing **203** and the earth, except for charged body **202**. A first current path is defined between load **206** and body **202** by conductor **212** and contact member **214**. This current path supplies the PD to load **206**. A second insulated conductor **216** is coupled at one end to charged body **204** via a connection (e.g., a solder joint) **217** to casing **203**. Conductor **216** is connected at its other end to load output **210**. A second current path is defined between load **206** and body **204** by conductor **216**, connection **217** and casing **203**. A return current path occurs between charged bodies **204**, **202** via casing **203** and the earth (and/or water **205**). Thus, a complete electrical circuit is established from charged body **202**, contact member **214**, conductor **212**, load **206**, conductor **216**, connection **217**, casing **203**, charged body **204**, and the return current path back to charged body **202**. An electric current related to the PD is generated in load **206**. In this embodiment, it is preferred that the bottom of the well be at least **400** feet deep relative to the surface.

Referring now to FIG. **9**, a system **300** for generating electricity from the geomagnetic field and rotation of the earth is shown as a third embodiment of the present invention. A potential difference PD is formed between spaced apart terrestrial charged bodies **302** and **304**. In this embodiment, body **302** is underground earth at a bottom of a mine and body **304** is the surface soil of the earth. The mine contains a drilled or excavated opening or mineshaft extending from the earth's surface to the bottom. A metal frame **303** may be present in the shaft. System **300** comprises an electrical load **306** having an input **308** and an output **310**. A first insulated conductor **312** is coupled at one end to charged body **302** via a contact member **314** and is connected at its other end to load input **308**. Contact member **314** makes direct contact with the bottom of the mine and, in some cases, may slightly penetrate the bottom. Conductor **312** is insulated to avoid direct electrical contact with frame **303** and/or the earth (except for charged body **302**). A first current path is defined between load **306** and body **302** by conductor **312** and contact member **314**. The first current path supplies the PD to load **306**. A second insulated conductor **316** is coupled at one end to charged body **304** via a contact member **318** and is connected at its other end to load output **310**. Contact member **320** has been driven into the earth to make good electrical contact with body **304**. Alternatively, conductor **316** may be coupled to body **304** via a connection (e.g., a solder joint) **320** to metal frame **303** (if any). A second current path is defined between load **306** and body **304** by conductor **316** and contact member **318** (or by conductor **316**, connection **320** and frame **303**). A return current path occurs between charged bodies **304**, **302** via the earth or, alternatively, via frame **303** and the earth. Thus, a complete circuit is established from charged body **302**, contact member **314**, conductor **312**, load **306**, conductor **316**, contact member **318** (or connection **320** and frame **303**), charged body **304**, and the return current path back to charged body **302**. An electric current related to the PD is generated in load **306**.

Referring now to FIG. **10**, a system **400** for generating electricity from the geomagnetic field and rotation of the earth is shown as a fourth embodiment of the present invention. A potential difference PD is formed between spaced apart terrestrial charged bodies **402** and **404**. In this embodiment, charged body **402** is the bottom or bed of a body of water (e.g., the earth's crust, basalt, sand, clay, sediments or other settled materials, a change in salinity, etc.). Charge body **404** is the water at or near the surface of the body of water. The body of water is preferably a salt water system such as an ocean, sea, ocean bay, strait, estuary, lake, and the like. System **400** comprises an electrical load **406** having an input **408** and an output **410**. A first insulated conductor **412** is coupled at one end to charged body **402** via a contact member **414** and is connected at its other end to load input **408**. Contact member **414** makes direct contact with the bottom and, in some cases, may penetrate the bottom. Conductor **412** is insulated to avoid direct electrical contact with the body of water. A first current path is defined between load **406** and body **402** by conductor **412** and contact member **414**. The first current path supplies the PD to load **406**. A second insulated conductor **416** is coupled at one end to charged body **404** via a metal boat **418**, which serves as a contact member. Conductor **416** is connected at its other end to load output **410**. Conductor **416** is directly connected to boat **418**, as represented by the electrical ground symbol. An alternative contact member may be a thin metal (e.g., copper or aluminum) plate **418'**, which floats on the water. In the latter case, conductor **416** is directly connected to plate **418'**. A second current path is defined between load **406** and body **404** by conductor **416** and boat **418** (or by conductor **416** and plate **418'**). A return current path occurs between charged bodies **404**, **402** via the body of water. Thus, a complete circuit is established from charged body **402**, contact member **414**, conductor **412**, load **406**, conductor **416**, boat **418** or other contact member, charged body **404**, and the return current path back to charged body **402**. An electric current related to the PD is generated in load **406**.

Referring now to FIG. **11**, a system **500** for generating electricity from the geomagnetic field and rotation of the earth is shown as a fifth embodiment of the present invention. A potential difference PD is formed between spaced apart terrestrial charged bodies **502** and **504**. In this embodiment, both charged bodies **502**, **504** are portions of the earth at or near the earth's surface, but body **504** is at a higher elevation than body **502**. This may be the case, for example, at a mountain or cliff System **500** comprises an electrical load **506** having an input **508** and an output **510**. A first insulated conductor **512** is coupled at one end to charged body **502** via a contact member **514** and is connected at its other end to load input **508**. Contact member **514** is driven into the earth to establish good electrical contact with charged body **502**. Conductor **512** is insulated to avoid direct electrical contact with the earth. A first current path is defined between load **506** and body **502** by conductor **512** and contact member **514**. The first current path supplies the PD to load **506**. A second insulated conductor **516** is coupled at one end to charged body **504** via a contact member **518** and is connected at its other end to load output **510**. Member **518** has been driven into the earth to make good electrical contact with body **504**. A second current path is defined between load **506** and body **504** by conductor **516** and contact member **518**. A return current path occurs between charged bodies **502**, **504** via the intervening earth. Thus, a complete circuit is established from charged body **502**, contact member **514**, conductor **512**, load **506**, conductor **516**, contact member **518**, charged body **504**, and the return current path back to charged body **502**. An electric current related to the PD is generated in load **506**.

Referring now to FIG. **12**, a system **600** for generating electricity from the geomagnetic field and rotation of the earth is shown as a sixth embodiment of the present invention. A potential difference PD is formed between spaced apart terrestrial charged bodies **602** and **604**. In this embodiment, charged body **604** is a cloud or a stream of clouds passing overhead and charged body **602** is a portion of the earth at or near the earth's surface. A preferred location for system **600** is at a high elevation, such as a mountain top, where low lying clouds are a frequent occurrence. As air is generally a poor conductor, the flow of short circuit current between clouds and the earth's surface is very limited, resulting in the buildup of a huge difference in electrical potential between the clouds and the earth's surface. THIS EMBODIMENT IS VERY DANGEROUS DUE TO LIGHTNING STRIKES AND STATIC ELECTRICITY. ANYONE WHO BUILDS THIS EMBODIMENT, WHETHER AUTHORIZED BY THE INVENTOR HEREIN OR NOT, ASSUMES ALL RISK OF PERSONAL INJURY OR DEATH AND ASSUMES ALL LIABILITY TO OTHERS INJURED OR KILLED AS A RESULT OF THE CONSTRUCTION AND/OR OPERATION OF THIS EMBODIMENT.

Again referring to FIG. **12**, system **600** comprises an electrical load **606** having an input **608** and an output **610**. A first insulated conductor **612** is coupled at one end to charged body **602** via a contact member **614** and is connected at its other end to load input **608**. Contact member **614** has been driven into the earth to make good electrical contact with body **602**. A first current path is defined between load **606** and body **602** by conductor **612** and contact member **614**. The first current path supplies the PD to load **606**. A second insulated conductor **616** is coupled at one end to charged body **604** via a contact apparatus **618** and is connected at its other end to load output **610**. Apparatus **618** includes a non-conductive tower **618***a*, an upwardly extending metal bracket **618***b*, and a metal collector plate **618***c*. Tower **618***a* is preferably built from a non-conductive material such as, for example, wood, fiberglass, fiberglass reinforced plastic, and the like. Tower **618***a* may be supported by non-conductive guy wires such as Phyllistran guys or Polygon structural fiberglass guys. Alternatively, tower **618***a* may be electrically insulated from the ground using an electrically insulating barrier between the tower and the ground. Conductor **616** is electrically connected to metal bracket **618***b*. Conductor **616** is insulated to avoid electrical contact with the surrounding atmosphere. Bracket **618***b* and collector plate **618***c* establish electrical coupling with the electric charge in cloud **604**. Plate **618***c* is a metal sheet of metal, such as copper or aluminum. A second current path is defined between load **606** and body **604** by conductor **616**, metal bracket **618***b*, and collector plate **618***c*. A return current path occurs between charged bodies **604**, **602**, via the atmosphere or moisture or precipitation in the atmosphere. Thus, a complete circuit is established from charged body **602**, contact member **614**, conductor **612**, load **606**, conductor **616**, bracket **618***b*, collector plate **618***c*, charged cloud **604**, and the return current path back to charged body **602**. An electric current related to the PD is generated in load **606**.

As mentioned, the hazards of lightning are obviously a major concern in system **600**. Lightning protection schemes are not infallible and no representation is made herein that system **600** is safe from the hazards of lightning strikes or static electricity. In system **600**, a surge arrester device **620** is installed between load output **610** and ground (FIG. **12**). Device **620** is shown as a gas-filled surge arrester, but the preferred arrester is a spark gap rated for lightning strike voltages. Preferably, conductor **616** is a high voltage stranded single conductor cable, such as used to connect lightning rods to grounding rods, but covered with insulation having a high dielectric strength to withstand lightning strikes.

An alternative embodiment to system **600** (not shown) is to replace apparatus **618** with a hydrogen- or helium-filled balloon (e.g., a weather balloon). The balloon would have a conductor such as copper or aluminum foil wrapped around it or a metal plate or band mounted on it. In this embodiment, insulated conductor **616** would be electrically connected to the conductor on the balloon. This embodiment would otherwise be the same as system **600**. The warnings and disclaimers made above with respect to system **600** are equally applicable to this embodiment.

Referring now to FIG. **13**, a system **700** for generating electricity from the geomagnetic field and rotation of the earth is shown as a seventh embodiment of the present invention. This embodiment concerns a well and is fundamentally the same as the second embodiment, except here the electrical load is located down in the wellbore. A potential difference PD is formed between terrestrial charged bodies **702** and **704** (i.e., bottom of well and earth surface). The well includes a drilled bore and a metal casing **703** extending down into the bore. At the bottom of the well is a supply of water **705**. An electrical load **706** has an input **708** and an output **710**. A first insulated conductor **712** is coupled at one end to charged body **702** via a contact member **714** and is connected at its other end to load input **708**. A first current path is defined between load **706** and body **702** by conductor **712** and contact member **714**. The first current path supplies the PD to load **706**. A second insulated conductor **716** is coupled at one end to charged body **704** via an electrical connection (e.g., a solder joint) **717** to metal casing **703**. The other end of conductor **716** is connected to load output **710**. A second current path is defined between load **706** and body **704** by conductor **716**, connection **717** and casing **703**. A return current path occurs between charged bodies **704**, **702**, via casing **703** and the earth (and/or water **705**). Thus, a complete electrical circuit is established from charged body **702**, contact member **714**, conductor **712**, load **706**, conductor **716**, connection **717**, casing **703**, charged body **704**, and the return current path back to charged body **702**. An electric current related to the PD is generated in load **706**. In the case where casing **703** is not made of electrically conductive material, an alternative coupling arrangement with charged body **704** is: conductor **716** is connected to a contact member **718**, which is driven into the ground at the earth's surface.

Load **706** may contain a charge storage system and a wellbore submersible pump controlled by a control unit **720**. The charge storage system stores electricity generated by system **700** and supplies the stored electricity to the pump when called upon by control unit **720**. Control unit **720** communicates with the pump via a control cable **722**. Control unit **720** includes an auto-control function that controls the pump automatically. Unit **720** switches on the power to the pump by means of a pressure switch actuated in response to a pressure valve or by means of a float level switch actuated by a float. An example of a submersible pump powered by a renewable electrical energy source is disclosed in U.S. Pat. No. 7,837,450, which is incorporated herein by reference.

Referring now to FIGS. **14** and **15**, a system **800** is another embodiment of the present invention. This embodiment is characterized by a coaxial cable **801** installed between two spaced apart, substantially parallel, terrestrial charged bodies **802** and **804**. Cable **801** includes a copper (or copper-clad steel) center conductor **812** insulated by a dielectric insulating material **812***a*, such as cellular polyurethane, polyurethane foam or solid polyethylene. Cable **801** further includes an outer shield conductor **803**, which may be a woven braid of bare or tinned copper wires, a wrap of aluminum foil tape, a solid copper or aluminum tube, or a combination thereof. Shield conductor **803** may be insulated or left uninsulated. In FIGS. **14** and **15**, conductor **803** is insulated generally by a jacket **803***a* made of plastic, such as polyvinyl chloride (PVC). When cable **801** is installed, jacket **803***a* is preferably removed between charged bodies **802** and **804**, as shown in FIG. **14**. System **800** comprises an electrical load **806** having a supply input and a ground output made available at a threaded female coaxial connector **808**. Cable **801** has a finished proximal end with a threaded male coaxial connector **809**. Connector **809** is considered male because it has a stinger wire, which is to be inserted into connector **808**. The stinger wire is connected to or an extension of center conductor **812**. Connector **809** includes a threaded slip ring **809***a* for threading connector **809** onto connector **808**.

Once connectors **808** and **809** are connected together, center conductor **812** is in electrical contact with the supply input of load **806** and shield conductor **803** is in electrical contact with the ground output of load **806**. Center conductor **812** is electrically coupled to charged body **802** via a metal contact member **814**. Conductor **812** delivers the potential difference between charged bodies **802**, **804** to load **806**. Shield conductor **803** is electrically coupled to charged body **804** by direct contact as shown. Thus, the ground output of load **806** is electrically coupled to charged body **804**. Conductor **803** is also coupled to charged body **802** by direct contact as shown. Alternatively, conductor **803** may be in direct contact with a portion of the earth defined to be outside of charged body **802** (thus making indirect contact with body **802**). A return current path occurs between charged bodies **802**, **804** primarily through shield conductor **803**. Using conductor **803** as the return current path improves return path performance by making the path easier to implement and maintain, and it operates more predictably because it is not dependent on the variable properties and conditions of earth, water, atmospheric, and other natural or in situ return current paths. In system **800**, a complete electrical circuit is established from charged body **802**, contact member **814**, center conductor **812**, load **806** (via connectors **809/808**), shield conductor **803** (via connectors **808/809**), charged body **804**, and shield conductor **803** back to charged body **802**. An electric current related to the PD between charged bodies **802**, **804** is generated in load **806**.

As shown in FIG. **14**, an alternative or additional ground output **810** is provided in electrical load **806**. An insulated conductor **816** is connected to output **810** and to a contact member **818**. In this case, shield conductor **803** may be decoupled from the original ground output, for example, by operation of an automatic impedance sensing switch internal to load **806**. Optionally, both outputs (**810** and original output via **803**) may be in operation together. In connection with output **810**, the return current path remains primarily through shield conductor **803**. The complete electrical circuit is from charged body **802**, contact member **814**, center conductor **812**, load **806** (via connectors **809/808**), conductor **816**, contact member **818**, charged body **804**, and shield conductor **803** back to charged body **802**.

Referring now to FIG. **16**, there is shown one example of an electrical load **900** of the present invention. A supply input S and a ground output G of load **900** correspond to, for example, load input **108** and load output **110**, respectively (e.g., see FIG. **7**). The potential difference between a pair of terrestrial charged bodies, obtained in accordance with the present invention, is presented across terminals S and G of load **900**. The apparatus of the present invention that obtains this potential difference (examples of which are shown in FIGS. **7-14**) is referred to in FIGS. **16-18** as a geomagnetic voltage source. Load **900** includes a charging circuit **902** and a charge storage system **904**. The specifics of charging circuit **902** depend upon the type of charge storage system used. System **904** may be, for example, a rechargeable battery or series of batteries or a charging capacitor or bank of charging capacitors. A basic charging circuit for a rechargeable battery (or series of batteries) may be a positive going diode in series with a source resistor, and both in series with the rechargeable battery (or series of batteries). A basic charging circuit for a capacitor (or bank of capacitors in series) may be an inductor in series with the capacitor (or bank of capacitors). The latter circuit typically includes a switching element or elements to cause the geomagnetic voltage source to charge the inductor in a first switched state and to cause the charged inductor to charge the capacitor or capacitors in a second switched state. If capacitors are used for the charge storage elements, a preferred form of capacitor is the ultracapacitor. Ultracapacitors and their charging circuits are disclosed in the following references: M. S. W. Chan, K. T. Chau, and C. C. Chan, Effective Charging Method for Ultracapacitors, Journal of Asian Electric Vehicles, Vol. 3, No. 2, December 2006, pp. 771-76; and Maxwell Technologies Application Note, Document 1008981 Rev 1, December 2005, pp. 2-3. The disclosures of these references are incorporated herein by reference. See also U.S. Pat. No. 7,170,260 to Thrap (Maxwell Technologies), entitled, Rapid Charger for Ultracapacitors, which is incorporated herein by reference.

Referring now to FIG. **17**, there is shown another example of an electrical load **900'** of the present invention. Load **900'** has a supply input S and a ground output G, as with load **900**. A geomagnetic voltage source of the present invention is applied to terminals S and G. Load **900'** includes a charging and control circuit **902'**, a charge storage system **904'**, and an electric motor **906'**. Blocks **902'** and **904'** contain the same types of charging circuits and charge storage elements as in blocks **902** and **904** (FIG. **16**), respectively. Circuit **902'** further includes a control circuit or IC chip that controls the operation of switches SW1 and SW2 in FIG. **17**. In a first operating stage, switch SW1 is closed and switch SW2 is open, and the geomagnetic voltage source and circuit **902'** charge the storage element or elements of charge storage system **904'**. In a second operating stage, switch SW1 is open (or sometimes remains closed) and switch SW2 is closed, and system **904'** powers electric motor **906'**. Motor **906'** is typically a DC motor, but may be an AC motor if, e.g., an inverter is installed between the output of system **904'** and the motor. It is understood that, in an alternative embodiment, load **900'** may simply consist of motor **906'**, which would be directly driven by the geomagnetic voltage source.

Referring now to FIG. **18**, there is shown a further example of an electrical load **900"** of the present invention. Load **900"** has a supply input S and a ground output G, as with loads **900** and **900'**. A geomagnetic voltage source of the present invention is applied to terminals S and G. Load **900"** includes a charging and control circuit **902"**, a charge storage system **904"**, and an inverter or AC generator with step-up transformer **906"**. Blocks **902"** and **904"** contain the same types of circuits and charge storage elements as in blocks **902'** and **904'** (FIG. **17**), respectively. In a first operating stage of load **900"**, switch SW**1** is closed and switch SW**2** is open, and the geomagnetic voltage source and circuit **902"** charge the storage element or elements of charge storage system **904"**. In a second operating stage, switch SW**1** is open (or sometimes remains closed) and switch SW**2** is closed, and system **904"** powers inverter or AC generator with step-up transformer **906"**. If an inverter, block **906"** directly converts the DC voltage from block **904"** to AC voltage and steps up the voltage to a desired level. If an AC generator, block **906"** includes a DC motor for driving the AC generator and a step-up transformer raises the voltage to a desired level. In the latter case, block **904"** drives the DC motor, which in turn drives the AC generator. The desired level of AC voltage may be at the mains level so that it can be applied to the electric power grid, for contributing power to the grid. FIG. **18** shows the output of block **906"** being applied to an electric power grid.

The phrases, "earth's surface" and "the surface of the earth," when speaking about orientation or relative position (e.g., "a direction normal to the earth's surface"), mean the general or averaged plane of the surface if observed locally or the general tangent of the surface if observed from space. In addition, these phrases are to be construed broad enough to include the surface of any body of land, water or ice or any geological formation such as, e.g., a mountain top or slope, or a valley, whether above or below sea level.

In connection with the first through fifth and seventh embodiments of the present invention (FIGS. **7-11** and **13**), the insulated conductors (**112**, **116**, **212**, **216**, **312**, **316**, etc.) may be a 14-Gauge insulated Stranded THHN Cable, Model No. 22955957, supplied by Southwire Co., Carrollton, Ga.

While the preferred embodiments of the invention have been particularly described in the specification and illustrated in the drawing, it should be understood that the invention is not so limited. Many modifications, equivalents and adaptations of the invention will become apparent to those skilled in the art without departing from the spirit and scope of the invention, as defined in the appended claims.

What is claimed is:

1. A method of generating electricity from the geomagnetic field and rotation of the earth, the earth rotating through the geomagnetic field to form a potential difference between first and second earth-bound terrestrial charged bodies, the first and the second earth-bound terrestrial charged bodies being spaced apart in a direction substantially normal to the earth's surface, said method comprising the steps of:

(a) using a system including (i) an electrical load having a supply input and a return output, (ii) a first current path having a first end coupled to the supply input of the electrical load, and (iii) a second current path having a first end coupled to the return output of the electrical load;

(b) coupling a second end of the first current path to the first earth-bound terrestrial charged body and coupling a second end of the second current path to the second earth-bound terrestrial charged body, such that the electrical load is electrically coupled to the first and the second earth-bound terrestrial charged bodies; and

(c) generating in the electrical load an electrical current related to the potential difference formed between the first and the second earth-bound terrestrial charged bodies by the earth's rotation through the geomagnetic field.

2. The method of claim **1**, wherein the first earth-bound terrestrial charged body includes a first portion of the earth located below the earth's surface and the second earth-bound terrestrial charged body includes a second portion of the earth located at about or near the earth's surface, and wherein the earth contains an opening extending between the first and the second portions of the earth,

said first current path including (i) an electrically insulated conductor extending at least partially through the opening and (ii) a conductive contact member coupled to the electrically insulated conductor for establishing electrical contact between the first portion of the earth and the electrically insulated conductor.

3. The method of claim **2**, wherein the electrical load is located within the opening, between the first and the second portions of the earth.

4. The method of claim **2**, wherein the opening is a bore or shaft of a well or mine, the well or mine including a bottom, and wherein the first portion of the earth includes the bottom of the well or mine, said electrically insulated conductor extending through the bore or shaft and the conductive contact member establishing electrical contact between the bottom of the well or mine and the electrically insulated conductor.

5. The method of claim **4**, wherein the bore or shaft is lined with a metal casing, the metal casing being in contact with the second portion of the earth and the second current path being in contact with the metal casing, such that the second current path is in electrical communication with the second portion of the earth.

6. The method of claim **1**, wherein the first and the second earth-bound terrestrial charged bodies are first and second portions of the earth, respectively, located at about or near the surface of the earth, the first portion of the earth being at a first elevation and the second portion of the earth being at a second elevation, the first elevation being different than the second elevation.

7. The method of claim **1**, wherein the first earth-bound terrestrial charged body is a portion of the earth at a bottom of a body of water, and wherein the second earth-bound terrestrial charged body is water at about or near the surface of the body of water.

8. The method of claim **7**, wherein the body of water is a body of saltwater.

9. The method of claim **1**, wherein the electrical load includes a rechargeable energy source, said method further comprising the step of

(c) charging the rechargeable energy source with the electric current generated in step (c).

10. The method of claim **9**, wherein the electrical load further includes an electric pump, said method further comprising the step of

(d) powering the electric pump with the rechargeable energy source.

11. The method of claim **9**, wherein the electrical load includes an electric motor, said method further comprising the step of

(d) powering the electric motor with the rechargeable energy source.

12. The method of claim **1**, wherein the electrical load includes an electric pump, said method further comprising the step of

(c) powering the electric pump with the electric current generated in step (c).

13. The method of claim **1**, wherein the electrical load includes an electric motor, said method further comprising the step of

(c) powering the electric motor with the electric current generated in step (c).

14. The method of claim **1**, wherein the electrical load includes a rechargeable energy source including a rechargeable battery and a charging circuit coupled to the rechargeable battery.

15. The method of claim **14**, wherein the electrical load further includes an electric motor coupled to the rechargeable battery.

16. The method of claim **1**, wherein the electrical load includes a rechargeable energy source including a one or more capacitors and a charging circuit coupled to the one or more capacitors.

17. The method of claim **16**, wherein the one or more capacitors are one or more ultracapacitors.

* * * * *

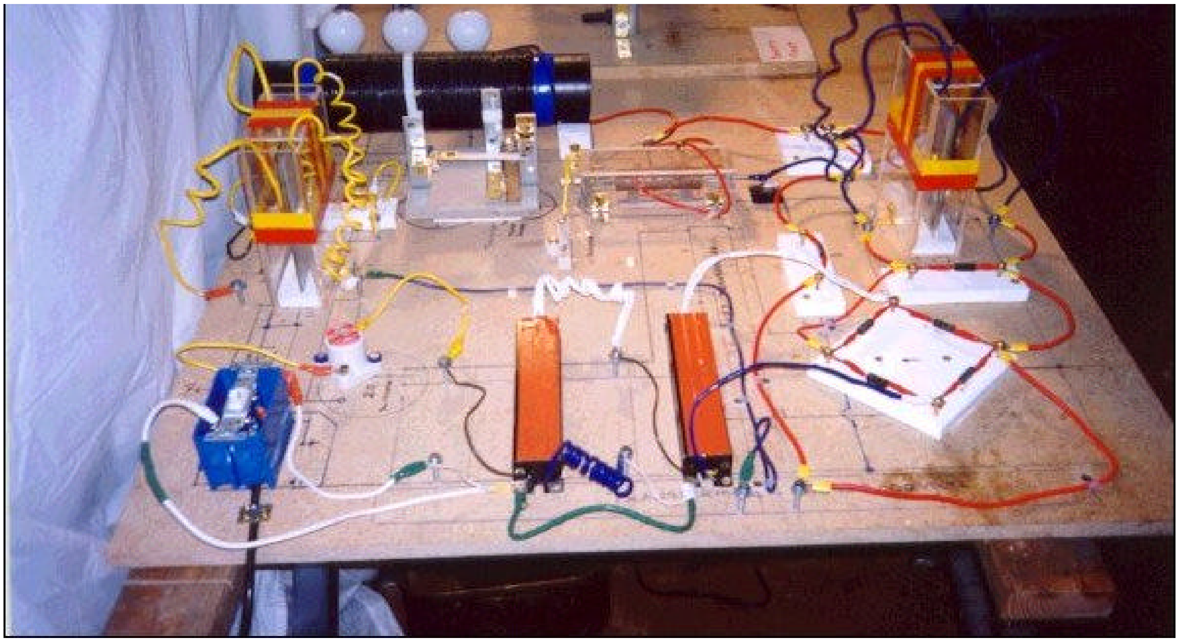

Efficient Power Supply Suitable For Inductive Loads

Engineering Report Gray Pat. no. 4,595,975

by

Mr. Gary Magratten

26901 Ridge Rd.
Willits CA 95490
ph: 707-459-1435
fax:707-459-9298

Table of Contents

Introduction

Circuit Section 1. Battery to the Transformer

Circuit Section 2. Transformer to Full Wave Bridge.

Circuit Section 3. From Full Wave Bridge; to the Capacitor#1; to the Conversion Element Switching Tube (CSET); to the Load; to the Capacitor #2 to the Battery.

Subsection 3A. From the Full Wave Bridge to the Capacitor#1
Subsection 3B. From Capacitor#1 to the Conversion Element Switching Tube. (CSET)
Subsection 3C. From the CSET to the Load to Capacitor#2
Subsection 3D. From Capacitor #2 to the Battery.
Circuit Section 4. Battery Charging Circuit.
Circuit Section 5. Chopper - Vibrator Circuit
Circuit Section 6. Low-Voltage Switching Circuit
Subsection 6A. Mechanical Switching
Subsection 6B. Solid State Switching
Subsection 6C. Laser Diode – Photo transistor Switching

Circuit Section 7. Conversion Element Switching Tube [CSET]

Circuit Section 8. Load.

Circuit Section 9. Spark Overshoot Device.

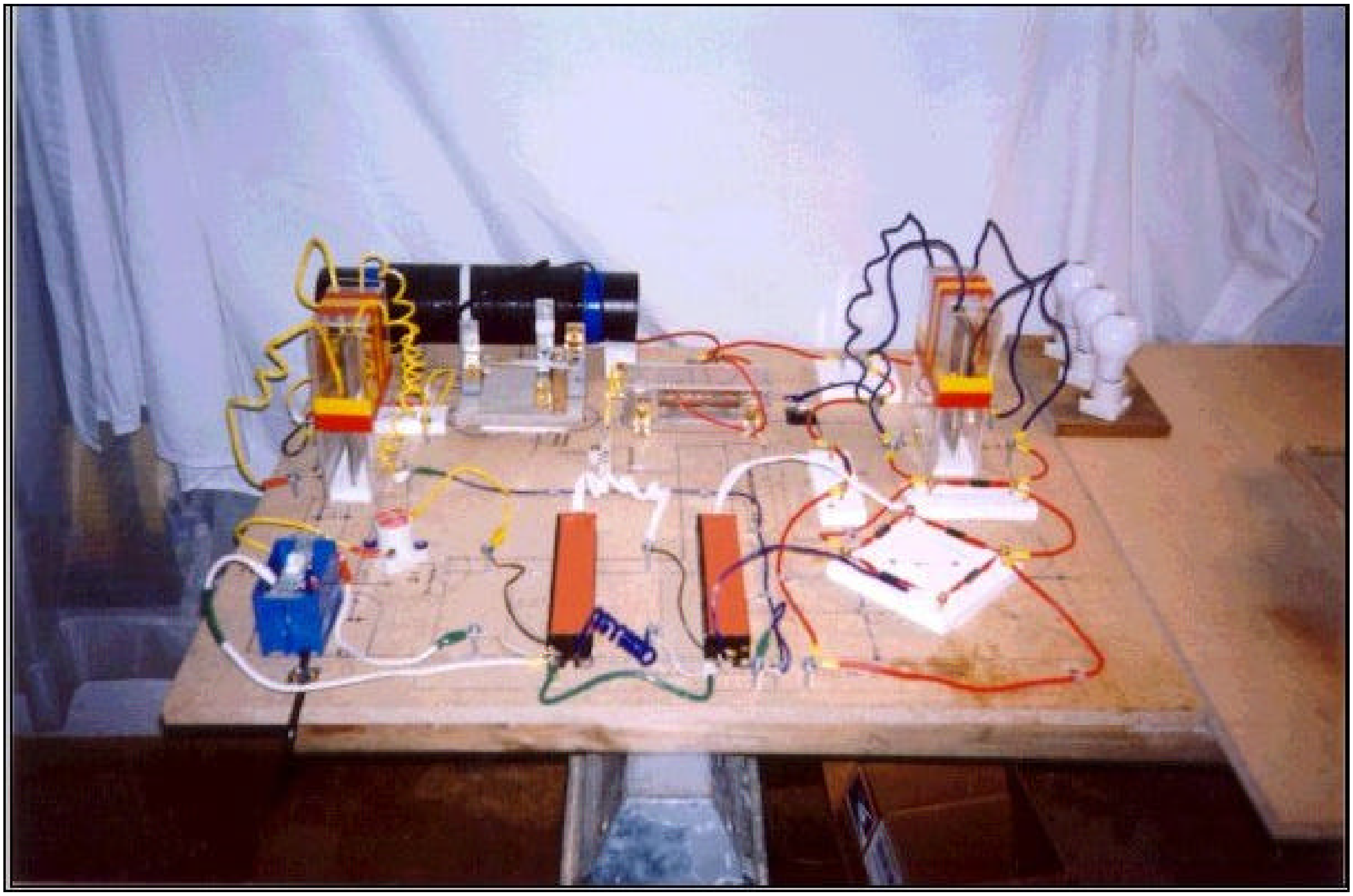

INTRODUCTION

The purpose of this report is to provide the circuit analysis, working drawings, component specifications and component manufactures for Mr. Edwin Gray's patent number 4,595,975 **-** Efficient Power Supply Suitable for Inductive Loads.

This information is for **SCIENTIFIC RESEARCH** only in order to more fully understand the scientific principles demonstrated by Gray's pat no. 4,595,975.

In the spirit of open information and cooperation between scientist, researchers and business professionals, this information is open to the public and may be reproduced and distributed by anyone for the purpose of providing scientific research information

WARNING! Gray's Circuit pat. no. 4,595,975 contains a very HIGH POWER CAPACITOR

that is DANGEROUS! Experimentation with this circuit requires at least the

formal education of a licensed electrician. If you perform research with this circuit,

YOU DO SO AT YOUR OWN RISK. Follow all standard safety precautions and use a proper laboratory to conduct testing.

If you follow the instructions, the circuit will work. I have confirmed this with actual test runs. It is my sincere hope that clean, efficient means to generate electricity will evolve from this information for the benefit of everyone. Mr Gary Magratten

Engineering Report Gray Pat. no. 4,595,975 Working Drawing no. 1

A.

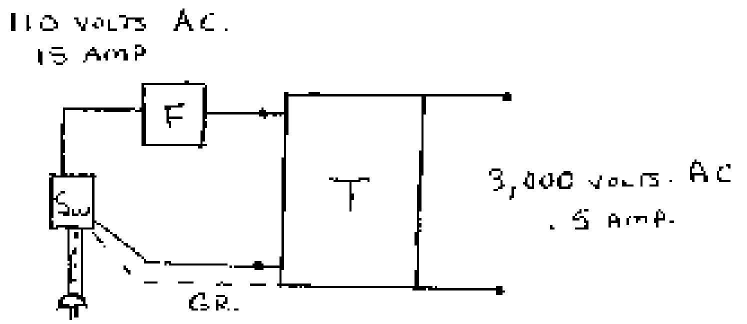
110 VOLTS AC.
15 AMP
F
T
Sw
GR.
3,000 VOLTS. AC
.5 AMP.

B.

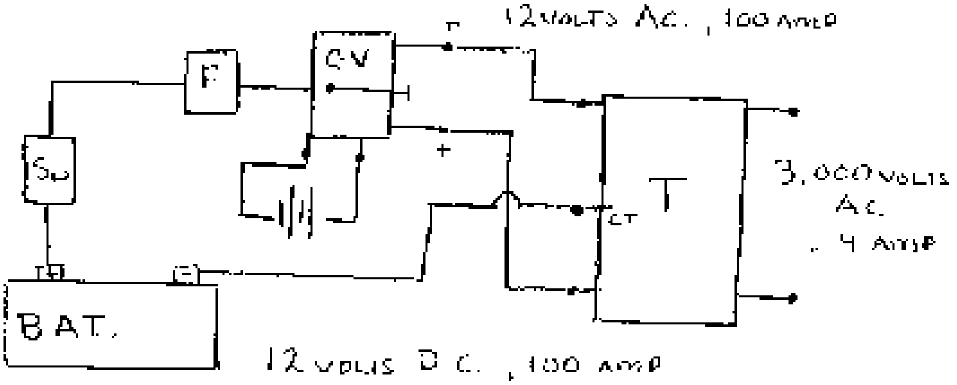
12 VOLTS AC., 100 AMP
F
CV
+
Sw
CT
T
3,000 VOLTS
AC.
.4 AMP
+
−
BAT.
12 VOLTS D.C. ,100 AMP

C.

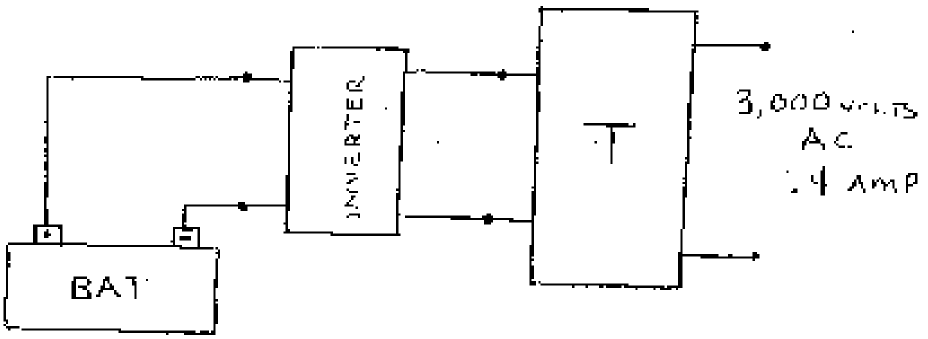
INVERTER
T
3,000 VOLTS
AC.
.4 AMP
+
−
BAT

Circuit Section 1. Battery to transformer:

The batteries should be deep cycle golf cart batteries, 12 volts and under 200 A hrs Use true deep cycle batteries. Automotive batteries are designed for high current for a short period of time. Deep cycle batteries provide steady current and can handle repeated recharging. A deep cycle battery rated at 200 A hrs should deliver 10 amps of current for 20 hrs.

There are three ways to accomplish this section of the circuit. Working drawing no.1 show all three ways. First, and easiest for research is the use of 110 AC house current at 15 amps to the primary of a transformer and a 3,000 volt, .5 amp AC secondary. The second is a 12 volt battery, DC, to a chopper-vibrator which produces 12 volt AC for the transformer. The design of the chopper-vibrator will be described separately later. This requires a special transformer with a 12 volt AC at 100 amps to the primary. The primary side also requires a center tap. The secondary is 3,000 volts and .5 amps AC. The third, is to use a deep cycle battery 12 volts, 100 amps and an inverter to form 110 AC at 11 amps. The transformer should then be rated at 110 volts at II amps on the primary and
3,000 volts and .4 amps.
Include a switch and a fuse into the circuit as shown on the working drawings.

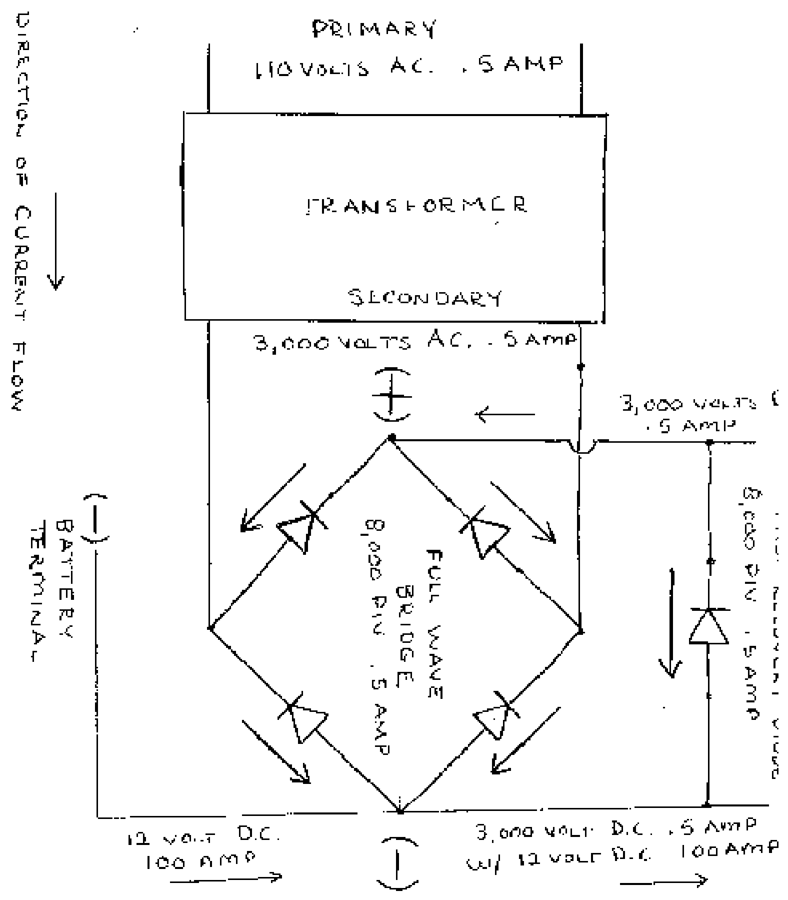
PRIMARY
110 VOLTS AC. .5 AMP
TRANSFORMER
SECONDARY
3,000 VOLTS AC. .5 AMP
DIRECTION OF CURRENT FLOW
(+)
3,000 VOLTS
.5 AMP
(–) BATTERY TERMINAL
FULL WAVE BRIDGE 8,000 PIV .5 AMP
8,000 PIV .5 AMP
12 VOLT D.C. 100 AMP
(–)
3,000 VOLT D.C. .5 AMP
W/ 12 VOLT D.C. 100 AMP

Engineering Report Gray Pat. no 4,595,975

Working Drawing no. 2

Circuit #2 - Transformer to Full Wave Bridge:

From the secondary of the transformer there is 3,000 volts and .5 amps alternating current. The full wave bridge converts the alternating current to direct current pulses. Position the diodes of the bridge for a positive terminal close to the transformer and the negative terminal away from the battery. To build a full wave bridge you can use four diodes or a one piece full wave bridge unit. The rating of the diodes is 8,000 PIV (peak inverse voltage) and .5 amp forward current.

Engineering Report Gray Pat. no. 4,595,975 Working Drawing no. 3A.

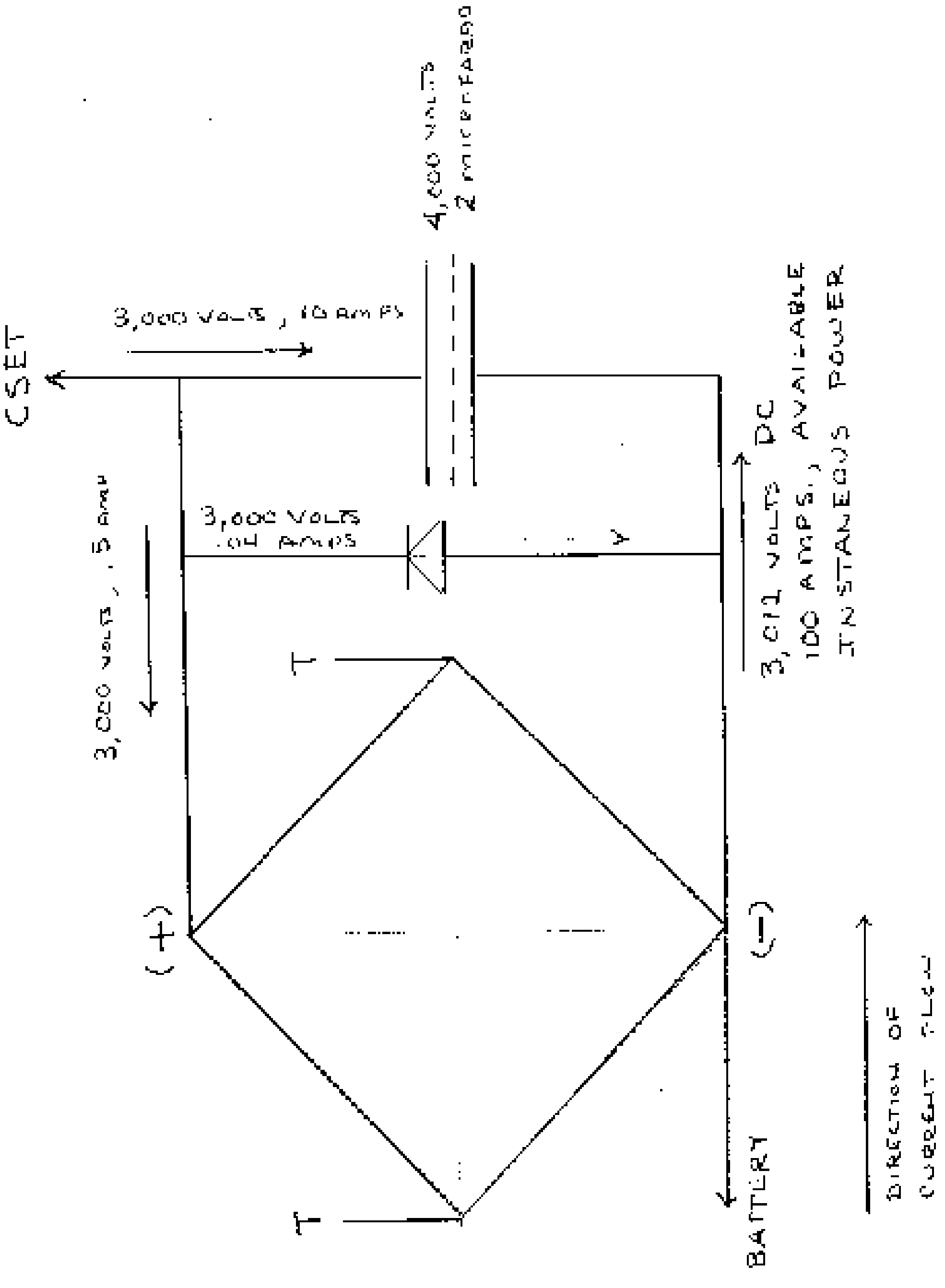
4,000 VOLTS
2 MICROFARADS
3,000 VOLTS, 10 AMPS
CSET
3,000 VOLTS, .5 AMP
3,000 VOLTS
.04 AMPS
3,012 VOLTS DC
100 AMPS., AVAILABLE
INSTANEOUS POWER
T
T
(+)
(-)
BATTERY
DIRECTION OF
CURRENT FLOW

Circuit Subsection 3A: - Full Wave Bridge to the

Capacitor #1:

In subsection 3A the full wave bridge can be treated as a direct current source with a positive terminal and a negative terminal. The voltage available is 3,000 volts at .5 amps. The positive terminal of the full wave bridge is connected to the positive side of Capacitor #1.

The negative terminal of the full wave bridge is connected to the negative side of Capacitor #1. The positive terminal of the capacitor is directly connected to the high-voltage anode of the CSET.

The negative terminal of the battery is connected to the negative terminal of the full wave bridge. Then the negative terminal of the full wave bridge is connected to the negative side of Capacitor #1.

THIS IS VERY IMPORTANT. THE 3,OOOvolts

.5 amp DIRECT CURRENT IS COUPLED WITH THE l2volt, 100 amps AVAILABLE FROM THE

BATTERY. THIS GIVES A 300 Kva CHARGE TO CAPACITOR #1. THIS IS VERY HIGH POWER

and DANGEROUS!

I found this difficult to understand but the technical discussion from Mr. Richard Hackenburger (Mr. Gray’s electrical engineer clearly shows ZS6Kva of instantaneous power available from Capacitor #1.) Careful examination of the wiring diagram will confirm this statement. Pulses of direct current are built up (technically called RAMPING) on capacitor #1 to be available for the arc in the CSET.

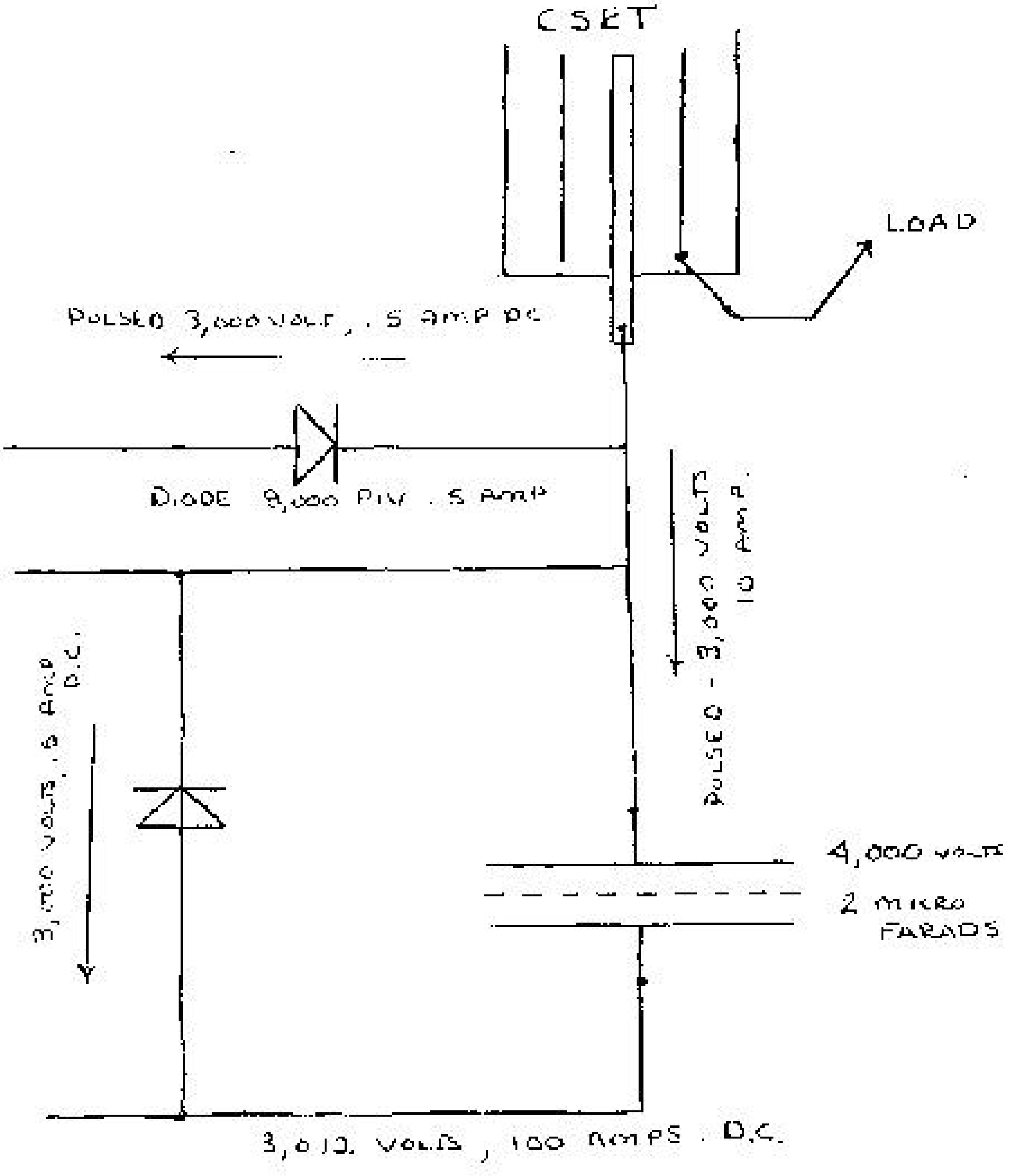
CSET
LOAD
PULSED 3,000 VOLT, .5 AMP DC
DIODE 8,000 PIV .5 AMP
PULSED - 3,000 VOLT 10 AMP.
3,000 VOLTS .5 AMP D.C.
4,000 VOLT
2 MIKRO FARADS
3,012 VOLTS, 100 AMPS. D.C.

Capacitor #1 to the high-voltage anode of the CSET. The positive side of Capacitor # 1 is coupled directly to the high-voltage anode of the CSET. Use #10 stranded copper wire with standard insulation and electrical tape to prevent contact with this HIGH POWER section of the circuit. The high-voltage anode is 3/16” zinc plated steel rod. This will be discussed in detail in the CSET section.

Engineering Report Gray Pat. no. 4,595,975   Working Drawing no. 3B

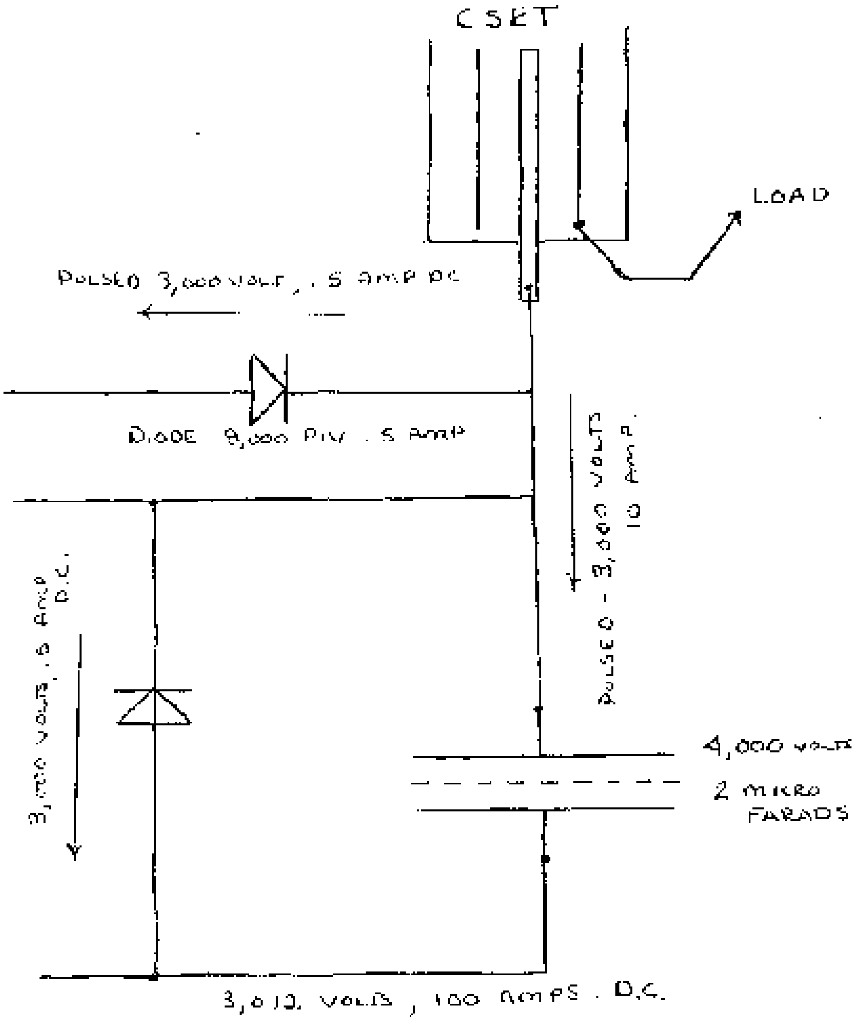
CSET
LOAD
PULSED 3,000 VOLT, .5 AMP DC
DIODE 8,000 PIV .5 AMP
PULSED - 3,000 VOLTS
10 AMP.
3,000 VOLTS, .5 AMP
D.C.
4,000 VOLTS
2 MIKRO
FARADS
3,012 VOLTS, 100 AMPS. D.C.

Circuit Subsection 3C. From the collector grids of the CSET to the Load to the plate of Capacitor #2.

THIS SECTION 15 UNUSUAL AND THEREFORE DIFFICULT TO

COMPREHEND, PLEASE HAVE PATIENCE!

The wave form of this section of the circuit is very unusual. This may be what was termed “Cold Electricity”. I have to do laboratory test to verify this explanation. Circuit Subsection 3C extends from the collector plates in the CSET to the Load to the Capacitor #2.

The LOAD and the CSET will be treated a separate subjects. The collector plates are exposed to the high-voltage anode 3,000 volt positive potential when the circuit switch on the low voltage side of the CSET is open. The 3,000 volt positive potential draws electrons to the collector plates from the circuit.

When the switch is quickly closed then reopened [.00005 sec] an ARC forms in the spark gap between the low voltage anode’s available current and the high voltage anode 3,000 volt positive potential. When an ARC is formed it is commonly accepted knowledge that the atmosphere is ionized to form positive ions and negative ions[free electrons].

When the ARC occurs, The Electrons Give Up Quanta or Photons of Electromagnetic Radiation. This is evidenced by the flame like discharge of visible light. This release of EMR or Radiant Event induces a current in the collector plates. This is the Photoelectric Effect to which the discovery is accredited to A, Einstien,

The ARC charges the copper tubes(collector grids 34A and 34 B) at the end near the spark gap.

The charge moves down the tube to the wire that is located at the end of the this slowly, and step by step.

Once again the switch is open and the 3,000 volt positive potential returns to the high voltage anode, and the process begins again.

Think of the circuit that extends from the collector plates through the LOAD to the Capacitor #2 plate as a column of fluid in a pipe, that is first stressed or subjected to suction, then compressed by pressure, in a repetitive manner. in the same manner the circuit is stressed then compressed with short duration, high-voltage, high-current energy spikes.

A single energy spike is capable of repelling like charged electromagnets with great force. This is how the motor was powered.

As the instantaneous energy spike travels through the load an EMF is produced about the inductor, be they electromagnets or a transformer.

When the energy spike ceases, the EMF is absorbed back into the inductor and on to the Capacitor #2 which is a coupling capacitor in that it draws current from the positive terminal of the battery thus recharging the battery.

CAPACITOR #2 to BATTERY.

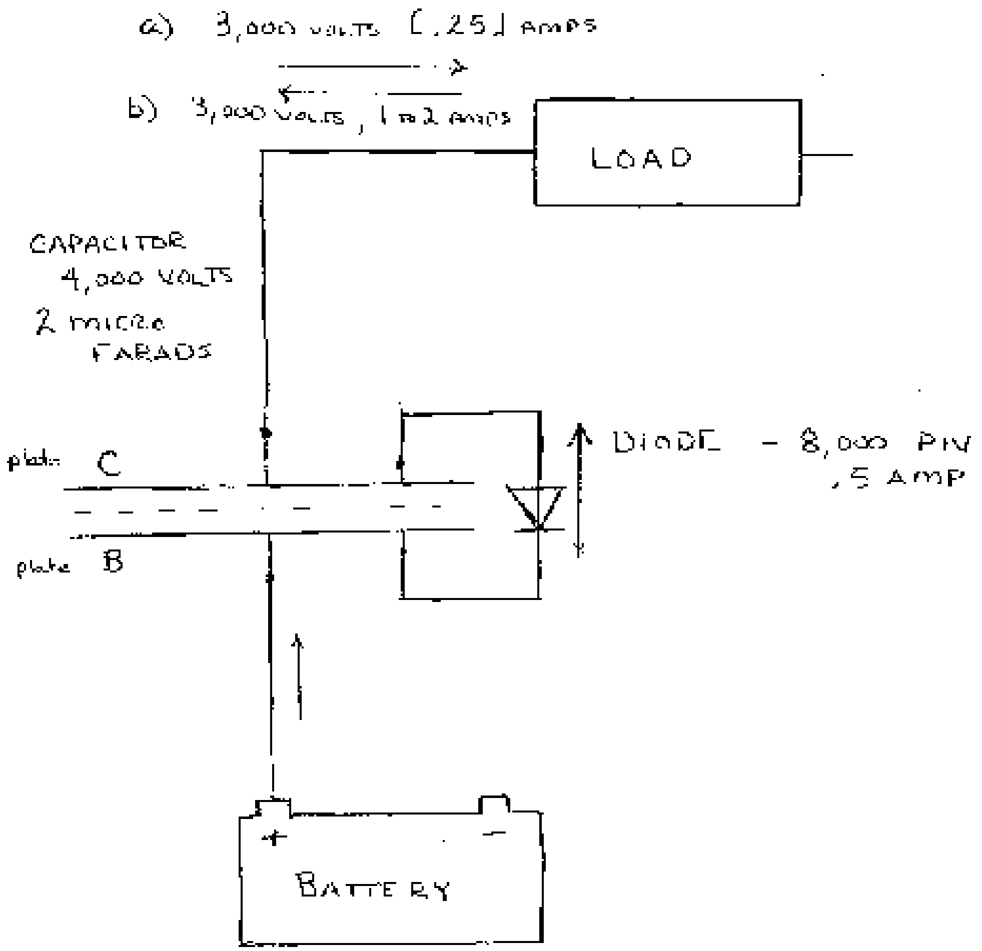
a) 3,000 volts [.25] amps
b) 3,000 volts, 1 to 2 amps
LOAD
CAPACITOR
4,000 VOLTS
2 MICRO
FARADS
plate C
plate B
DIODE - 8,000 PIV
.5 AMP
+
−
BATTERY

Circuit Subsection 3D: From Capacitor #2 to the Battery.

This section of the circuit extends from the plate of Capacitor #2 to the positive terminal of the battery. According to the patent the CEME is transferred through Capacitor #2 to recharge the battery.

I am going to propose a slight change in the design. This is the only place where I deviate from the patent and the schematic as originally presented. [I would also like to state that during my efforts to reconstruct this circuit, I wasted a many hours trying to second guess the information and take short cuts. The message here is to follow the patent and the schematic, the device does work.)

The change in the design is the introduction of a diode from one plate of the capacitor to the other with the forward current flow in the direction of the LOAD. **I have NOT TESTED THIS YET**. Here is the reason for the proposed change in the design. Capacitor #2 is a coupling capacitor in that it conducts alternating-current but blocks the transmission of steady direct current. For the sake of analysis let us say that pulsed direct current can be treated the same as alternating current with respect to the capacitor#2 in that the changing value of voltage and current are coupled through the dielectric.

When the low-voltage switch of the commutator is open, no power is flowing in the circuit. The high-voltage anode is at a charged state of 3,000 volts. This draws electrons to the collector plates from the plate of the capacitor #2.

The deficit of electrons on Plate C of the capacitor#2 becomes a positive potential. The plate B of capacitor #2 becomes negatively charged by drawing electrons from the positive terminal of the battery**. This is the same way a battery charger works.**

By introducing a high current diode from plate B to Plate C of Capacitor#2, a one-way-path is created for the current drawn from the positive terminal of the battery.

When the commutator switch is closed and power is flowing in the circuit an **ARC** is created from the low voltage anode to the high voltage anode. This **ARC** [EMR or Radiant Energy] is absorbed by the collector grids 34a and 34b.

This pulse of power induces a spike of energy [**high-voltage, high-current, short duration pulse** I through the wire back to plate C of capacitor #2, which had a positive potential.

The power pulse arrives on the plate C and distributes electrons on the plate. These electrons can travel no further than the plate because the diode prevents the transfer of electrons to plate B.

The commutator switch is opened and the high-voltage potential returns to the high-voltage anode of the CSET.

The cycle repeats itself and electrons are drawn to the collector plates by the high positive potential of the high- voltage anode. The introduction of a **HIGH CURRENT DIODE** from plate B to plate C creates a one-way-path for current to be drawn from the positive terminal of the battery to allow the battery to be recharged.

Engineering Report Gray Pat. no. 4,595,975 Working Drawing no. 4

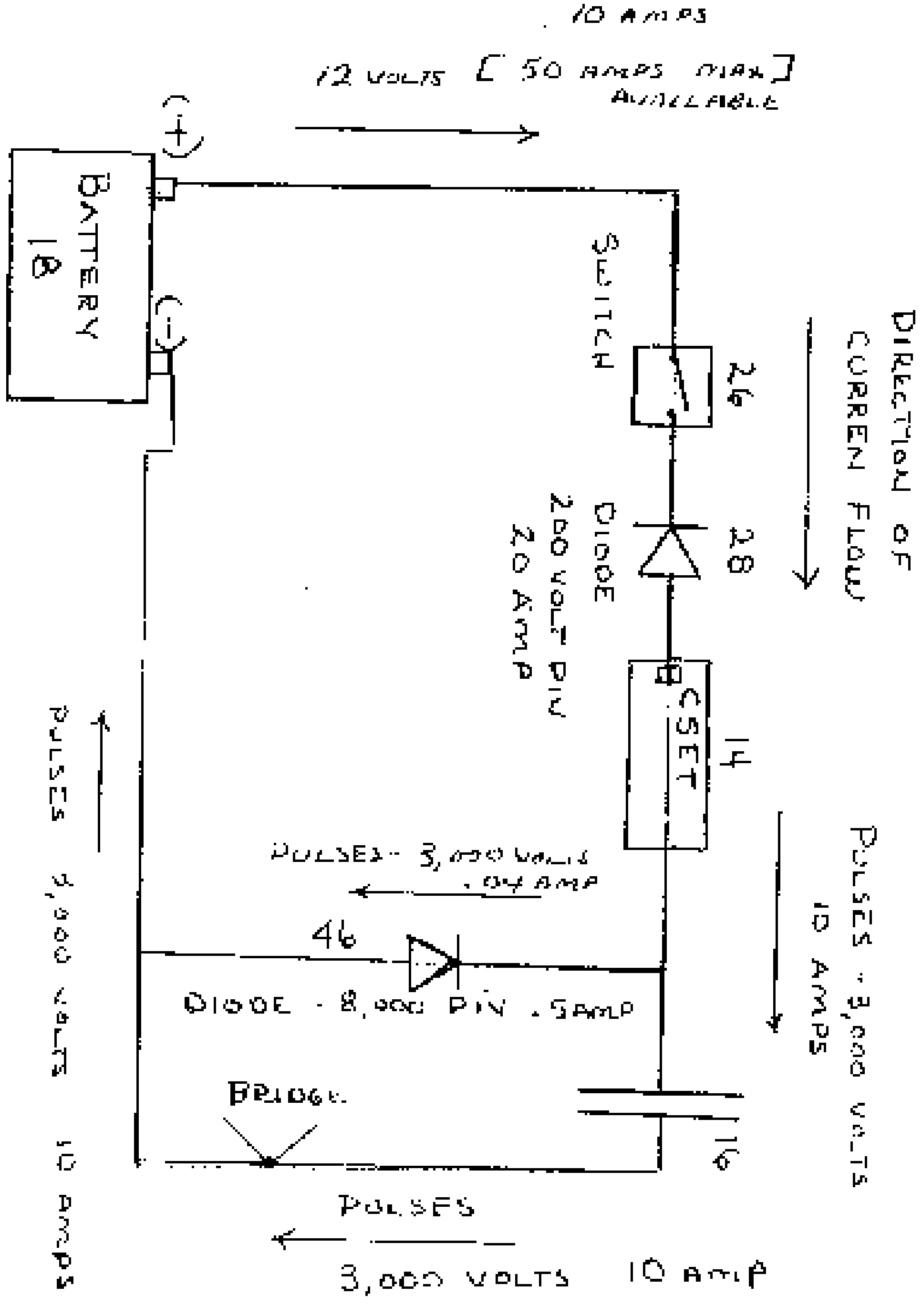
10 AMPS
12 VOLTS [50 AMPS MAX]
AVAILABLE
(+)
BATTERY
18
(-)
SWITCH
26
DIRECTION OF
CURRENT FLOW
DIODE
200 VOLT PIV
20 AMP
28
C SET
14
PULSES - 3,000 VOLTS
10 AMPS
PULSES - 3,000 VOLTS
.04 AMP
46
DIODE - 8,000 PIV .5AMP
BRIDGE
16
PULSES
3,000 VOLTS 10 AMP
PULSES 3,000 VOLTS 10 AMPS

Circuit Section 4: Battery Charging Circuit.

THIS SECTION IS NOT EXPLAINED BY THE PATENT. SIMPLY TRACING THE CURRENT FLOW WIll VERIFY THIS ANALYSIS.

The battery charging circuit extends from the positive terminal of the battery through the commutator; through the triode; to the low voltage anode; across the spark gap; to the high voltage anode; to DIODE 46; to the negative terminal of the battery.

IT IS CRITICAL TO THE UNDERSTANDING OF HOW THIS CIRCUIT WORKS TO COMPREHEND THE DIRECTION OF CURRENT FLOW IN THIS SECTION OF THE CIRCUIT! THE CURRENT FLOW IS FROM THE POSITIVE TERMINAL OF THE BATTERY THROUGH THIS CIRCUIT TO THE NEGATIVE TERMINAL OF THE BATTERY! THIS IS OPPOSITE THE NORMAL DIRECTION OF A DIRECT ***CURRENT CIRCUIT. THINK OF IT THE SAME AS A BATTERY CHARGER.***

Current is drawn from the positive terminal of the battery through the TRIODE through the CSET to the high -voltage anode and then to the plate of capacitor#l.

The 3,000 volt positive potential overcomes the 12 volt positive potential at the terminal of the battery. Most of the current is delivered to the plate of capacitor#l.

The 12 volt negative potential provided by the negative terminal of the battery draws a small portion of the current through DIODE 46 back to the battery to provide a pulsed recharging current. The current directed back to the battery through **DIODE 46** is probably only .04 amps.

A MUCH LARGER PULSE IS CREATED ON THE BACK PLATE OF CAPACITOR #1 THAT IS WIRED BACK TO THE NEGATIVE TERMINAL OF THE BATTERY TO PRODUCE A RECHARGING CURRENT.

When **ARC OVER** occurs at the spark gap in the CSET, a pulse of current is directed to the high-voltage positive potential developed on the plate of capacitor #1 facing the CSET.

Capacitor #1 is a coupling capacitor in that it transfers the power pulse. This pulse of current repels electrons from the other plate of capacitor#1 forming a current pulse back to the battery of about 10 amps [depending on the current available from the low voltage anode at the time of ARC OVER.]

This Is the Primary Means By Which the Battery is Recharged While Driving the

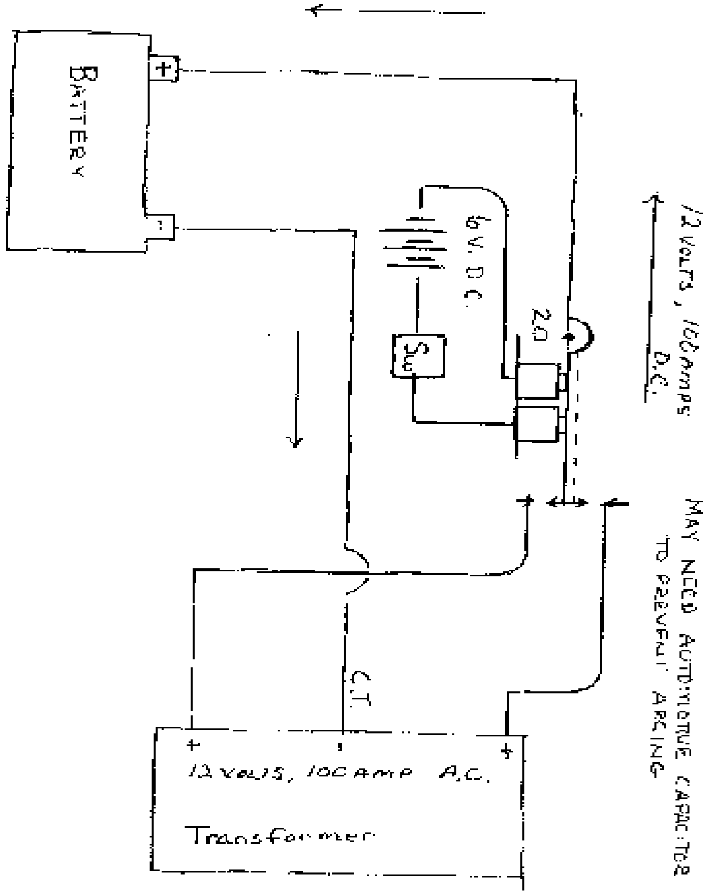
BATTERY
+
-
6V. D.C.
SW
2Ω
12 VOLTS, 100 AMPS D.C.
MAY NEED AUTOMOTIVE CAPACITOR TO PREVENT ARCING
C.T.
+
+
12 VOLTS, 100 AMP A.C.
Transformer

CHOPPER - VIBRATOR CIRCUIT

Load Circuit.

Circuit Section 5. Chopper-Virator Circuit.

If you use a 12 volt battery for the primary energy source to drive the transformer, then the 12 volts direct current must be changed to 12 volts alternating current as an alternating current is necessary to power an inductor.

A simple chopper- vibrator constructed from a common alarm bell available at any hardware store can be used.

An Alarm bell consist of two small electromagnets and a spring held lever. The contacts make and break the circuit to switch the circuit at a very low frequency. Please look at the working drawing, Circuit Section 5WD. As stated before, you can power the transformer with 110 house current or a 12 volt battery to an inverter and then on to the transformer.

If you use 12 volts DC and the chopper-vibrator to form 12 volts AC , you must use a transformer that has a center tap on the primary side. Johnson Electric Coil Co. is familiar with the design. Ask to talk with Ms. Beth Bockes, design engineer.

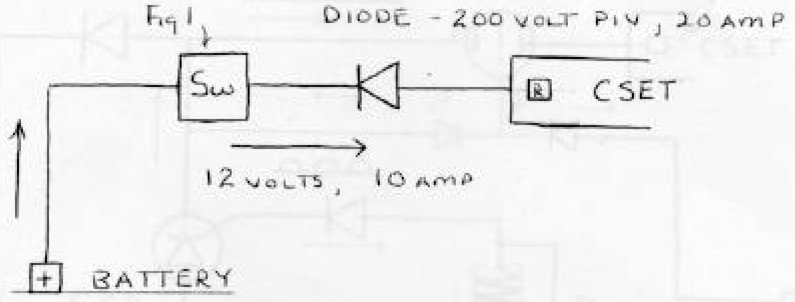
Fig 1
DIODE - 200 VOLT PIV, 20 AMP
Sw
R
CSET
12 VOLTS, 10 AMP
+
BATTERY

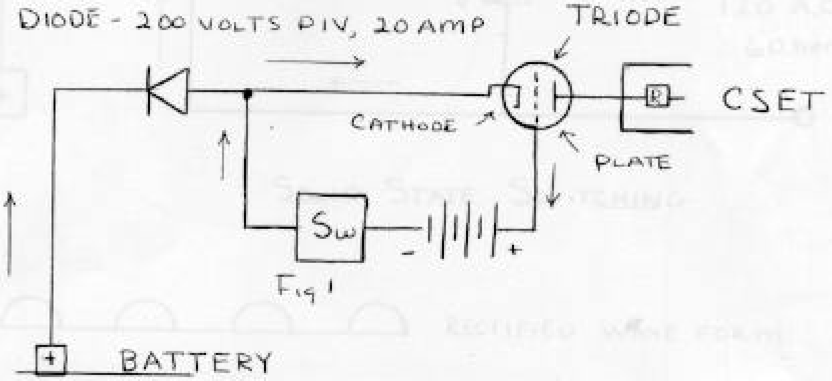
DIODE - 200 VOLTS PIV, 20 AMP
TRIODE
R
CSET
CATHODE
PLATE
Sw
-
+
Fig 1
+
BATTERY

Fig 1

Switch

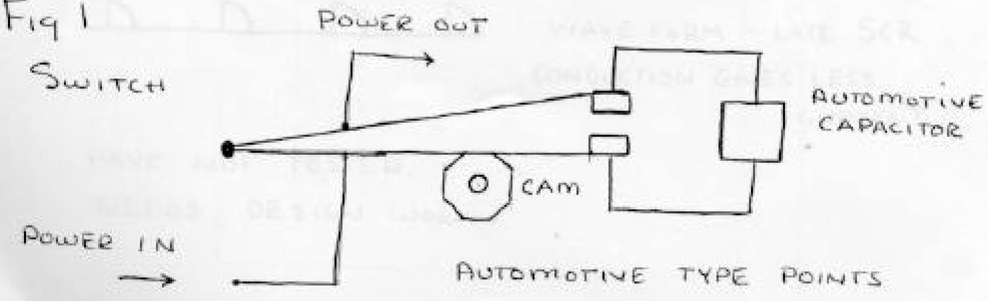
POWER OUT
AUTOMOTIVE CAPACITOR
CAM
POWER IN
AUTOMOTIVE TYPE POINTS

DURATION .0005 SEC.
FREQUENCY 120 PULSES / SEC.
MOTOR RPM 3600

This circuit creates the short duration [.0005 sec.] current pulse necessary for the formation of the ARC in the CSET. There are many ways to accomplish this.

You can use a TRIODE or a FAST-RECOVERY DIODE.

If you use a fast-recovery diode, you can place mechanical switching in the circuit line from the positive terminal of the battery to the diode.

If you use a triode, you must control the grid of the triode with a separate circuit.

MAKE SURE THE TRIODE OR THE DIODE IS INSTALLED IN THE DIRECTION OF THE CURRENT FLOW WHICH IS FROM THE POSITIVE TERMINAL OF THE BATTERY TO THE CSET. With a TRIODE, the cathode goes to the pos. terniinal of the battery and the plate goes to the CSET. With a high current fast recovery DIODE, the cathode goes to the pos. terminal of the battery and the anode goes to the CSET.

If the intended purpose of the LOAD is a Gray type motor, then one **PULSE** of power is required to repel like pole electromagnets arranged slightly offset from each other.

If the intended purpose of the LOAD is to run a direct current system, the LOAD could be a deep cycle battery from which current is drawn to power a circuit.

If the LOAD is standard type AC, 60 hertz, lights and electrical equipment then a step-down transformer that provides narrow pulsed direct current of a duration around .0005 sec. at a frequency of 120 pulses per sec. is suggested.

THE KEY HERE IS TO DESIGN THE WAVE FORM TO MEET THE NEEDS OF THE EQUIPMENT YOU

ARE TRYING TO POWER.

I have only run a mechanical switch made of a simple electric motor, a cam and a set of points connected to a diode. I performed two successful test runs.

MANY PEOPLE HAVE ASKED,” WHY DON’T YOU TRY THIS?”. Here is why.

FROM THE MAGNITUDE OF THE POWER DISCHARGE IN THE CSET, I REALIZED THAT SAFETY REQUIRES ANY FUTHER TEST RUNS BE DONE IN IN A LABORATORY WITH QUALIFIED HELP.

I am negotiating with a laboratory now, but because of the cost it may be months before the laboratory, test equipment and qualified personnel are organized and funded.

This is HIGH -VOLTAGE, HIGH-CURRENT PULSES OF ELECTRICAL POWER CONVERTED TO **ELECTROMAGNETIC RADIATION** of unknown frequency and therefore DANGEROUS. All known methods of generating electrical energy are dangerous but we know what we are working with. This is research, PROCEED CAREFULLY and SAFELY.

Here are three ways to a accomplish the switching. The first is a mechanical switch consisting of an octagonal cam (nut) mounted on a motor shaft and a set of automotive points.

If you place the mechanical switching in the circuit line then an automotive type capacitor across the points may be helpful to prevent arcing. If you use mechanical switching to control the grid of a triode, then the capacitor is probably not necessary.

Circuit Subsection 6A. A mechanical switch consisting of a motor, a octagonal cam a set of automotive points. See working drawing. To achieve the .0005 sec

Engineering Report Gray Pat. no. 4,595,975 Working Drawing no. 6B

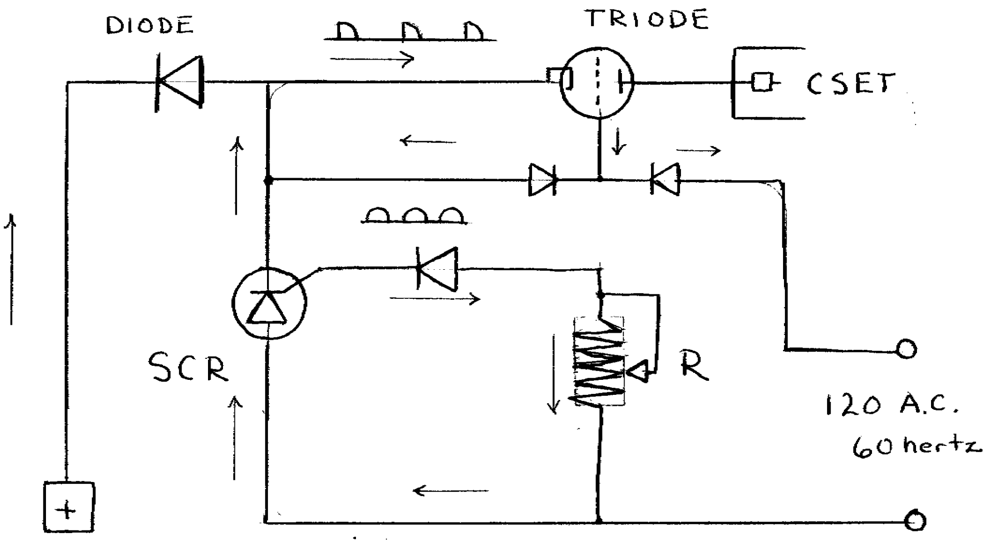

SOLID STATE SWITCHING

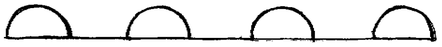
RECTIFIED WAVE FORM

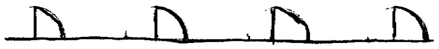
WAVE FORM - LATE SCR
CONDUCTION GIVES LESS
CURRENT

HAVE NOT TESTED.
NEEDS DESIGN WORK.

Circuit Subsection 6B. Solid State Switching.

THIS HAS NOT BEEN TESTED. This circuit controls the grid of the triode with a SCR, a diode, a variable resister and a low current AC power source. The AC is rectified to form a pulsed DC by the diode. The SCR clips the pulse to create a narrow pulsed direct current wave form. The variable resistor defines the width

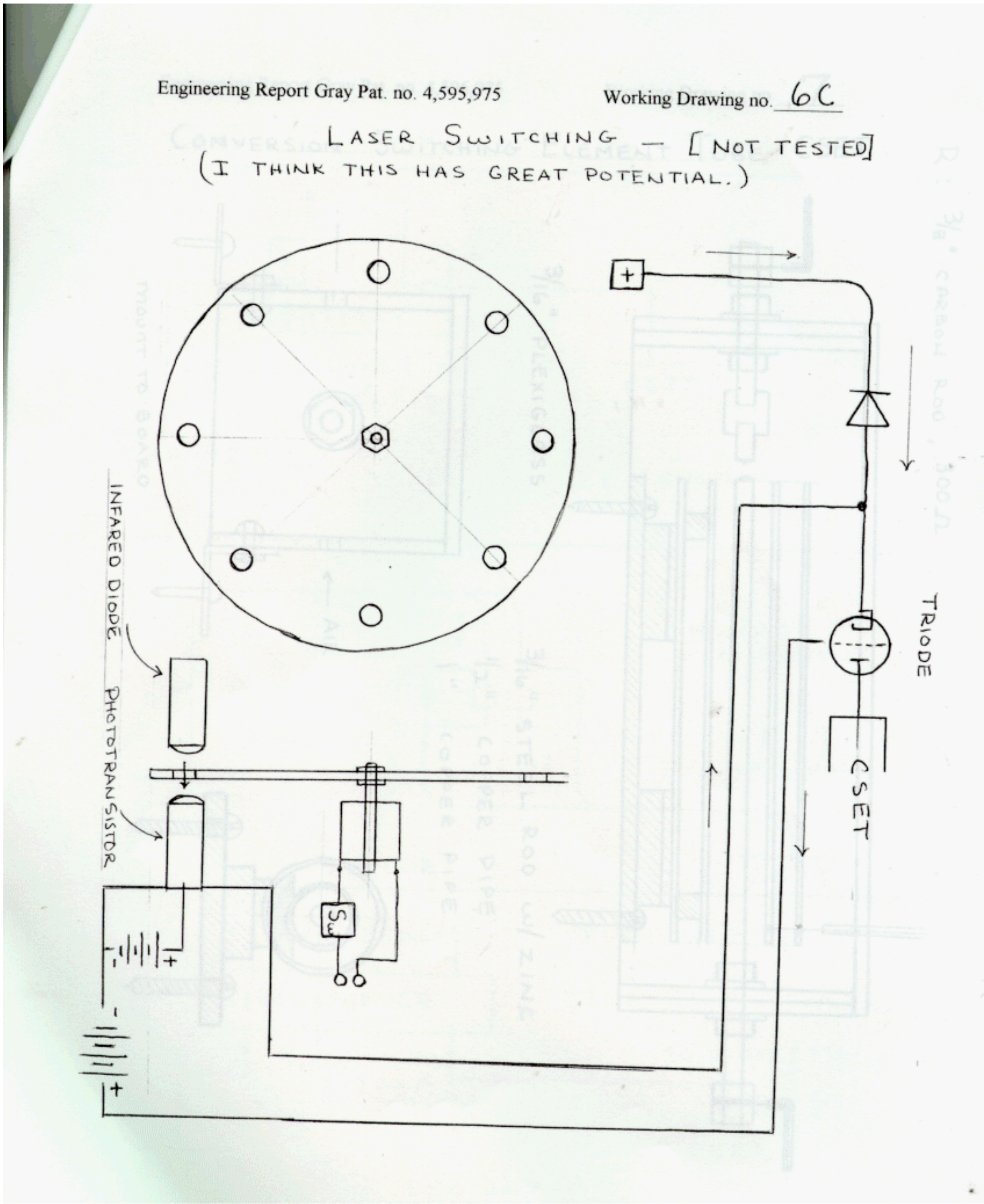
Engineering Report Gray Pat. no. 4,595,975
Working Drawing no. 6C
LASER SWITCHING — [NOT TESTED]
(I THINK THIS HAS GREAT POTENTIAL.)
+
TRIODE
CSET
INFARED DIODE
PHOTOTRANSISTOR
S
+
-
-
+

Circuit Subsection 6C. Laser Switching.

THIS HAS NOT BEEN TESTED.

This switching circuit consist of a infared light emitting diode and a phototransistor.

They are positioned on the opposite sides of a revolving disk powered by a motor. Located on the disk are holes 318” in diameter equally positioned 30 degrees apart at a 4” radius on the disk. A motor rotates the disk at 1800 rpms (approximately). The holes on the disk allow the infared beam of light to trigger the base of the phototransistor. The diameter of the hole, the position on the disk and the rpm of the motor can be adjusted to create any pulsed DC

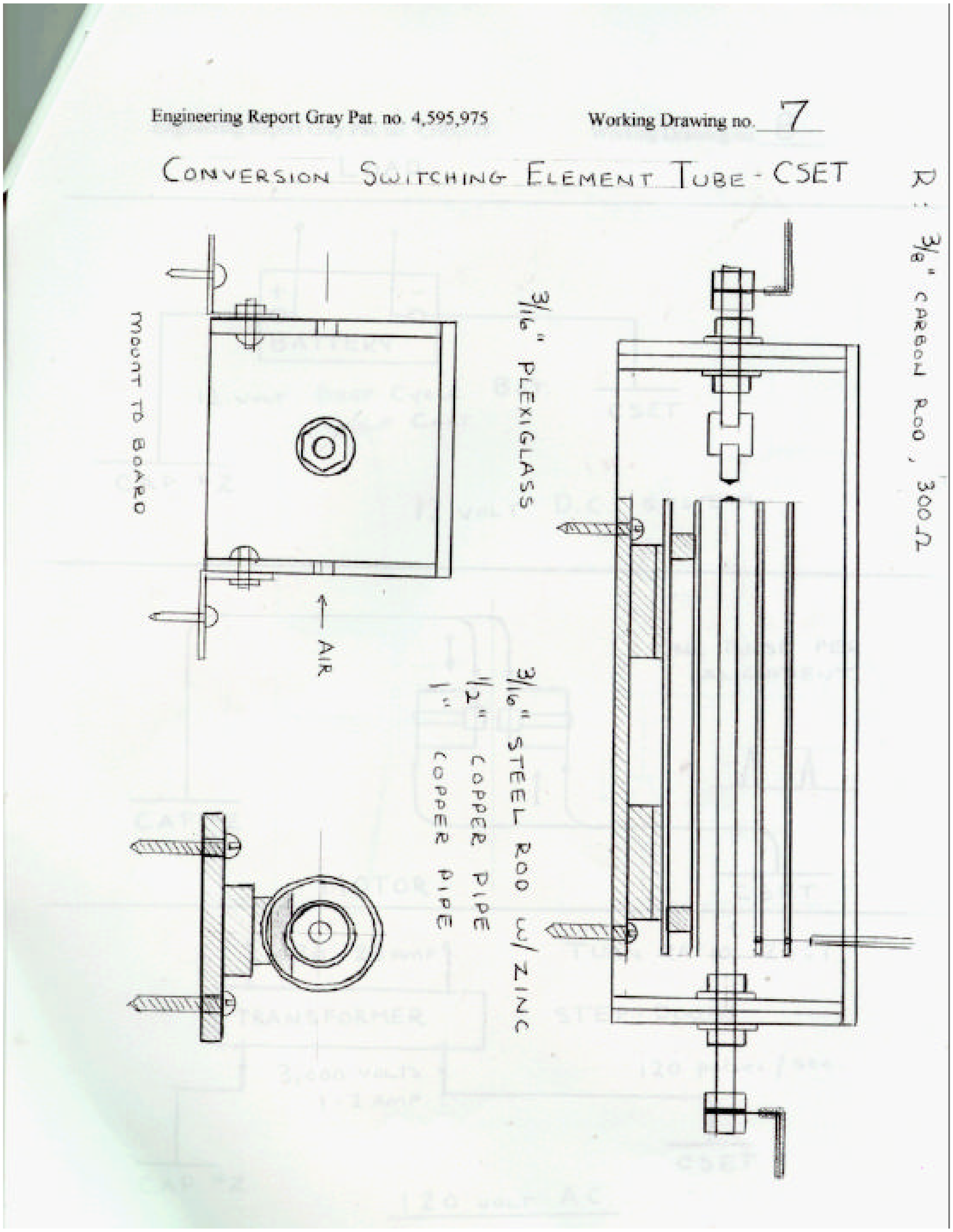
Engineering Report Gray Pat. no. 4,595,975 Working Drawing no. 7

CONVERSION SWITCHING ELEMENT TUBE - CSET

Circuit Section 7. Conversion Element Switching Tube. CSET

Before I explain the mechanics of the C SET, I would like to explain the principles that make the operation of the CSET possible. If we clearly understand how something works then the mechanics become very easy.

When the switch on the low voltage side of the CSET is open, the high voltage anode has a 3,000 volt positive potential. This draws electrons to the collector plates 34a and 3 4b.

When the switch is closed, current from the low voltage anode jumps the spark gap to the high voltage anode forming an ARC. An electric ARC does two things. First, it ionizes the air molecules to form positive and negative ions. Negative ions are free electrons. The high voltage positive potential anode picks up these free electrons. They are delivered to the negative terminal of the battery in the form of recharging pulses of power.

This gives the circuit a GAIN or increase in electrical energy. The second thing is absorbed by the collector plates in the CSET. This is called the photoelectric effect [see A. Einstein - photoelectric effect]. Quanta or photons of EMIR transfer their energy to the electrons in the copper collector plates. The pulse of EMR creates a pulse of electric power that travels across the collector plates to the wire terminal at the end of the plates. The wire transmits the power pulse to the load. At the same time, a BUNCH of electrons from the low voltage anode hit the high voltage anode and travel across the steel rod.

This negates the 3,000 volt positive potential on the high voltage anode thus repelling the electrons that were drawn to the collector plates. The combined effect is the creation of an instantaneous high-voltage, high-current energy pulse delivered to the LOAD.

The CSET requires AIR for the ionization of the atmosphere. Look at Gray’s motor patent and you will find that an air supply was delivered to insure proper conditions for arc-over.

In a technical discussion issued by Richard Hackenberger, Mr. Gray’s electrical engineer,

Mr. Hackenberger states that the disassociation of air molecules to form positive ions and negative ions is a key principle by which the gain in the circuit occurs. Quantum physics explains that when an electron is freed from its energy shell a quanta of EMR is released. [see Quantum Physics - Bohr I. It may be that the CSET was an vacuum tube.

The possibility will be included in future testing. All evidence points to the importance of AIR to form the ARC necessary to produce the EMR that induces the current in the collector plates.

The construction of the CSET is fairly simple. I consist of a plexiglass housing that allows air to reach the spark gap.

The high voltage anode and the low voltage anode are 3/16” steel rod, zinc plated. The collector plates are ½” by 4” copper pipe and I” by 4” copper pipe.

Spacers are scrap plexiglass held in place with super glue. I drilled holes in the collector grids. [

The collector grids may function better if there are no holes.}

The resistive element in the low voltage anode was not incorporated in the tests. I am going to include a resistive element made of 3/8” carbon rod as shown in the patent drawings of the CSET.

I am designing the resistive element for 300 ohms based on the assumption that there is 50 amps available from the positive terminal of the battery that can be drawn in a reverse current.

A 300 ohm resistor would limit the current to a 10 amp range. I am hoping that this will help bring the magnitude of the discharge into a containable range for the size of the CSET elements.

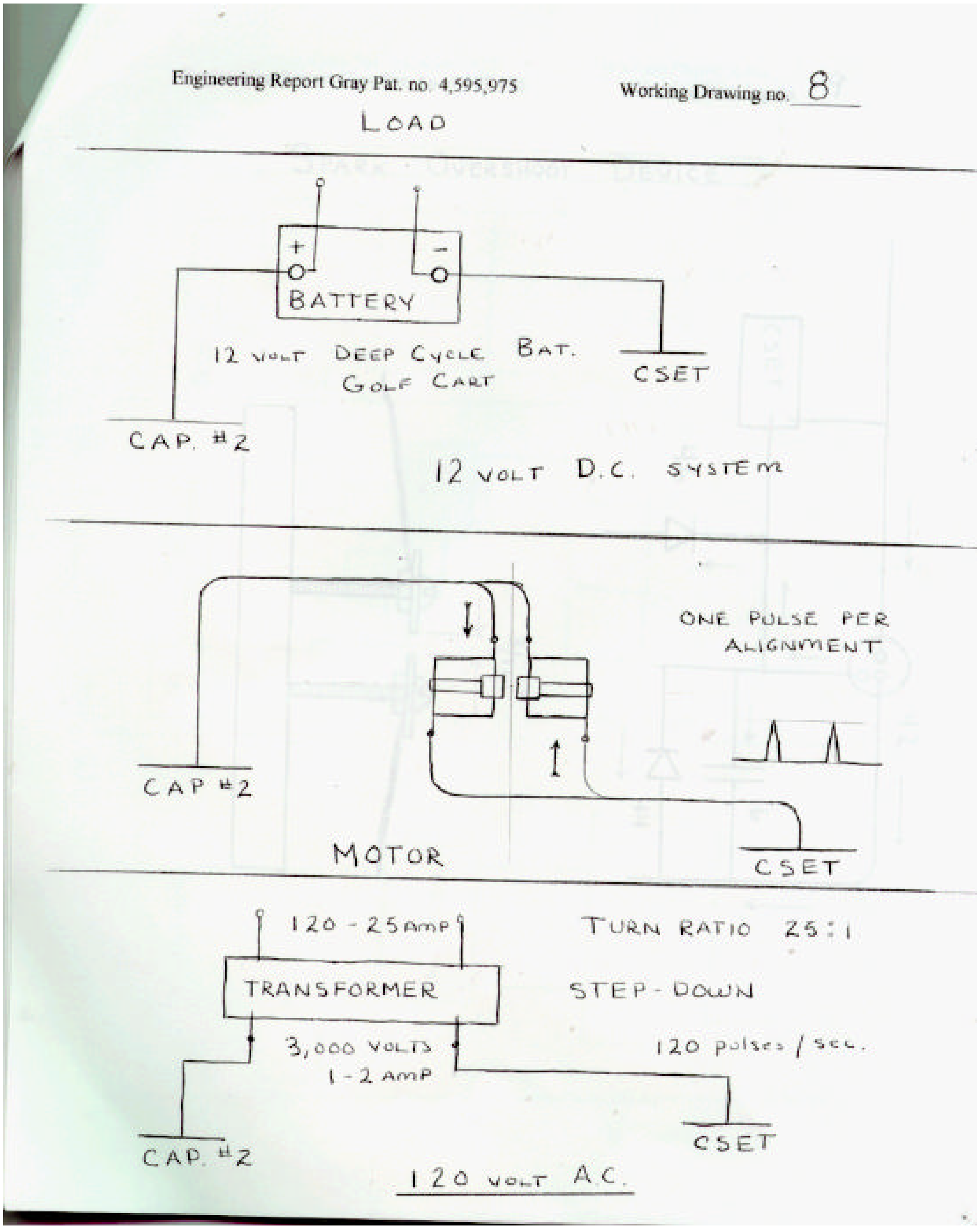
Engineering Report Gray Pat. no. 4,595,975
Working Drawing no. 8
LOAD
+
-
BATTERY
12 VOLT DEEP CYCLE BAT.
GOLF CART
CSET
CAP. #2
12 VOLT D.C. SYSTEM
ONE PULSE PER ALIGNMENT
CAP #2
MOTOR
CSET
120 - 25AMP
TURN RATIO 25:1
TRANSFORMER
STEP-DOWN
3,000 VOLTS
1-2 AMP
120 pulses / sec.
CAP. #2
CSET
120 VOLT A.C.

Circuit Section 8. Load

The Load or actually the means by which the Load (being the equipment to be powered by the circuit) is transferred is dependent upon the requirements of the equipment being powered. I would like to quote the Abstract from pat.4,595,975.

"When adapted to present day direct-current or alternating-current devices the LOAD could be a BATTERY or CAPACITOR to enhance the productivity of electrical energy."

Consider three different types of LOADS. A pulsed DC motor, a direct current circuit of 12 volts and an 110 AC circuit of lights or standard AC, 60 hertz equipment. In the case of a pulsed DC motor such as the one designed by Mr. Gray, one energy pulse repels like pole electromagnets arranged in a slightly offset manner.

Therefore the motor can be driven directly from the pulsed DC generated by the CSET if the pulse is timed with the alignment of the electromagnets. [please see motor patent,I]

If the Load is a 12 volt DC circuit, then a battery could be placed in circuit current energy spikes from the CSET to 110 volt, 120 pulses per sec may work. [haven’t tried it yet] I am trying to drive a motor I designed based upon Gray’s principles.

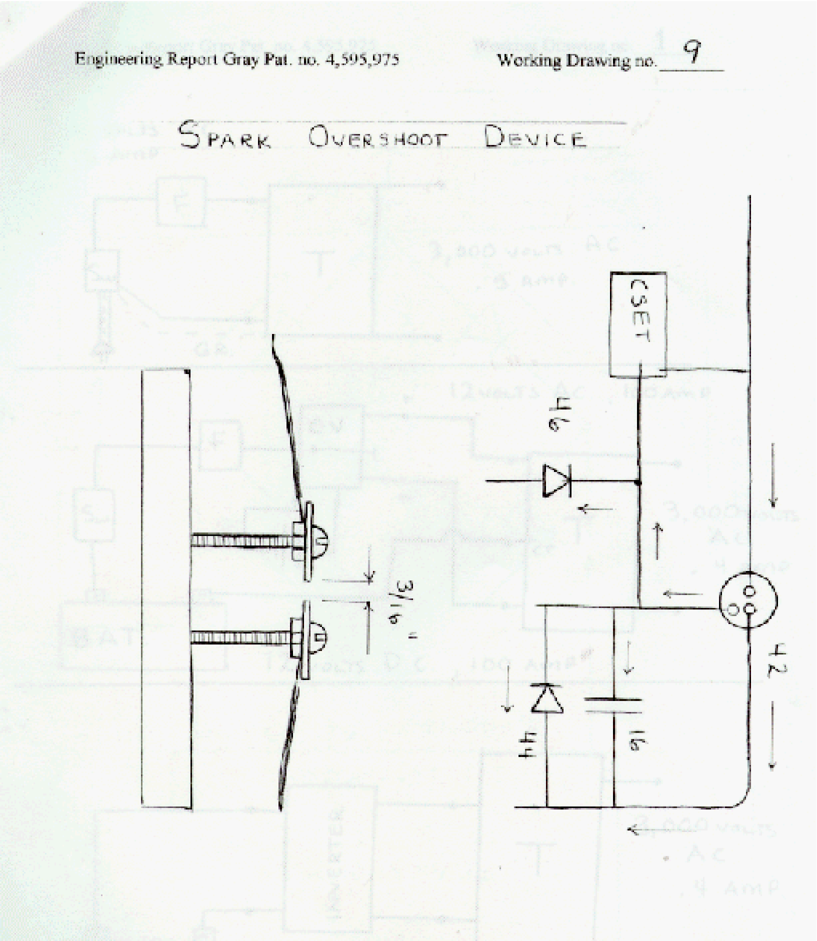

Circuit Section 9: Spark Overshoot Device

The spark overshoot device is a safety value for the CSET. If the energy pulse exceeds the design parameters of maximum allowable voltage, then arc over occurs and the excess energy is recycled back into the high-voltage circuit to be adsorbed by the capacitor #l and the battery. Spark gap is 3/16”, See Working Drawings.

Patent Rights:

Efficient Power Supply Suitable for Inductive Loads was patented by Mr. Edwin Gray in 1986. Mr. Gray passed away in 1989. Rumour has it [from his wife, he disappeared. To my knowledge he still holds the rights to the patent ? If you are interested in manufacturing this device you should try to seek permission from any controlling interest of the patent. We all are aware of the fact that we now stand at the edge of a world war and/or terrorist attacks motivated in part by the need for energy. It is definitely in the best interest of the United States of America to develop new energy sources as quickly as possible. Technological information travels so fast today that if we in the United States take a day off, an other country will bring this to market by tomorrow. 1 personally think that there is a social responsibility to bring this to market.

If someone does hold the patent rights, then it would seem that a fair and

June 17, 2003. Patent rights expire in 17 years unless other legal action is taken.

Parts List and United States America suppliers

TRANFORMER:

Johnson Electric Coils Co. 1-800-826-9741, 1-715-627-4367 821 Watson Street,

Antigo, Wisconson 54409-2753
contact: Ms. Beth Bockes ext. 309
part: a) 110 volt, 15 amp primary / 3,000 volt .5amp secondary
b) 12 volt, 100 amp primary / 3,000 volt .4 amp secondary

DIODES & BRIDGE. HIGH VOLTAGE:

HVCA 1-732-938-4499

P.O.Box 2245 , Farmingdale, N.J. 07727
part: fast recovery diodes; UX-FOB 8,000 PIV, .5 amp
ultra fast bridge; 2HVFWB8KBUF 8,000 Ply, .5 amp

CAPACITORS:

By Cap 1-800-322-9227 contact: Ken

5505 North Walcut Ave., Chicago III. 60640

part: Capacitor #1 - 4,000 volts , 2 microfarads

Capactor #2 - [talk to Ken J still running calcs.

TRIODE:

Parts Express 1-800-338-0531

725 Pleasant Valley Drive, Springboro, OH 45066-1158
part: Svetlana 5V81 1-10, part no. 072-5 19
HIGH CURRENT DIODE: Mouser 1-800-346-6873
1000 N. Main Street, Mansfield TX 76063-15 14
part: 526-NTE6240 ultra fast, dual center tap, carbon cathode, 200 volt PIV, 20 amp forward current, 150 amp surge.
also (laser diode and photo transistor for laser switch- soon to be designed)

Resistor: 300 ohms - 3/8” carbon rod

Batteries: Deep Cycle Golf Cart

Low-Voltage Switching: motor with octagonal cam, automotive points, duration .0005 sec., 3600 rpm on motor - 120 pulses per Sec.

CSET: 3/16” plexiglass, 3/16” steel rod zinc plate, 3/16” spark gap, ½” by 4” copper pipe, U’ by 4” copper pipe, spacers- scrap plexiglass

CONCLUSION

This Engineering Report would never have been possible if it were not for the genius of Peter Lindemann,D. Sc.. His research connecting the work of Nicola Tesla and Edwin Gray is a brilliant contribution to the application of theoretical physics for the solution of world energy problems.

Those who break new ground are rarely acknowledged by the established scientific or higher education community until long after research becomes history. Kepler, Columbus, Einstien, Rutherford, Tesla, Gray and many others painfully encountered resistance to new research. The message here is not to be discouraged. If you know that the basic theory is correct, then never give up

To those of you who are doing scientific research on Gray's patents, congratulations, you are working on the actual generation of electrical power based on Quantum Physics. Here are some areas that need consideration;

BREMSSTRAIILUNG: “breaking radiation” The radiation emitted by electrons slowed down in matter. [free electrons entering a steel high-voltage anode] [Bunching]

PAIR PRODUCTION: The formation of a positron and an electron when electromagnetic energy interacts with matter.

PHOTOIONIZATION: The ionization of of a gas by light or other electromagnetic radiation; the photons must possess enough energy to detach one or more outer electrons from the gas atoms.

PHOTOCONDUCTION: The absorption of EMR by electrons in matter whereby the electrons are brought to the range of energy levels at which they move freely to conduct electricity.

BREAKDOWN: When electrons are accelerated to high velocity by the electric field at the [spark gap] and produce other free electrons by ionization collision with atoms. These free electrons are similarly accelerated by the field and in turn cause other ionizations. The avalanche process leads to a very large current.

Study the work of Tesla, Gray, Dr. Lindemann, Plank, Einstien and Bohr. Let us all work hard to make this technology available to everyone in a useful and constructive manner.

Engineering Report Gray Pat no. 4,595,975
Conclusion
Engineering Report Gray Pat, no. 4,595,975
Mr Gary Magratten

This information by Mr Gary Magratten and is reproduced by energy21 with many thanks to the fore mention Gary Magratten

Additional information on this device can be found at
http://www.Fortunecity.com/greenfield/bp/16/the_answer.htm

If you have found this article of interest you may care to check out related articles on the Edwin Gray devices on the Encyclopedia of Free energy on CD where you find also other devices like this described as well in greater detail

ENERGY21
The Encyclopedia of free energy on CD
click here for details

Complete ordering details for the FREE ENERGY ENCYCLOPEDIA can be found at
http://angelfire.com/ak/egel/ver2.html
or
http://www.Fortunecity.com/greenfield/bp/16/ver2.html

The basic ordering information though is as follows:

Make Payment Please in the following amounts and in these currencies only

The Encyclopedia of Free Energy
Cost $35 Dollars US (Cash)
$45 Euro Dollars (cash)
$50 Dollars Australian (Australian Cheque) or Australian Money order also acceptable(for Australia resident customers only)
$55 New Zealand dollars
$50 Canadian Dollars

This payment includes airmail postage to your address worldwide.
Please send order and payment to
Geoff Egel
18 Sturt Street
Loxton 5333
South Australia Australia

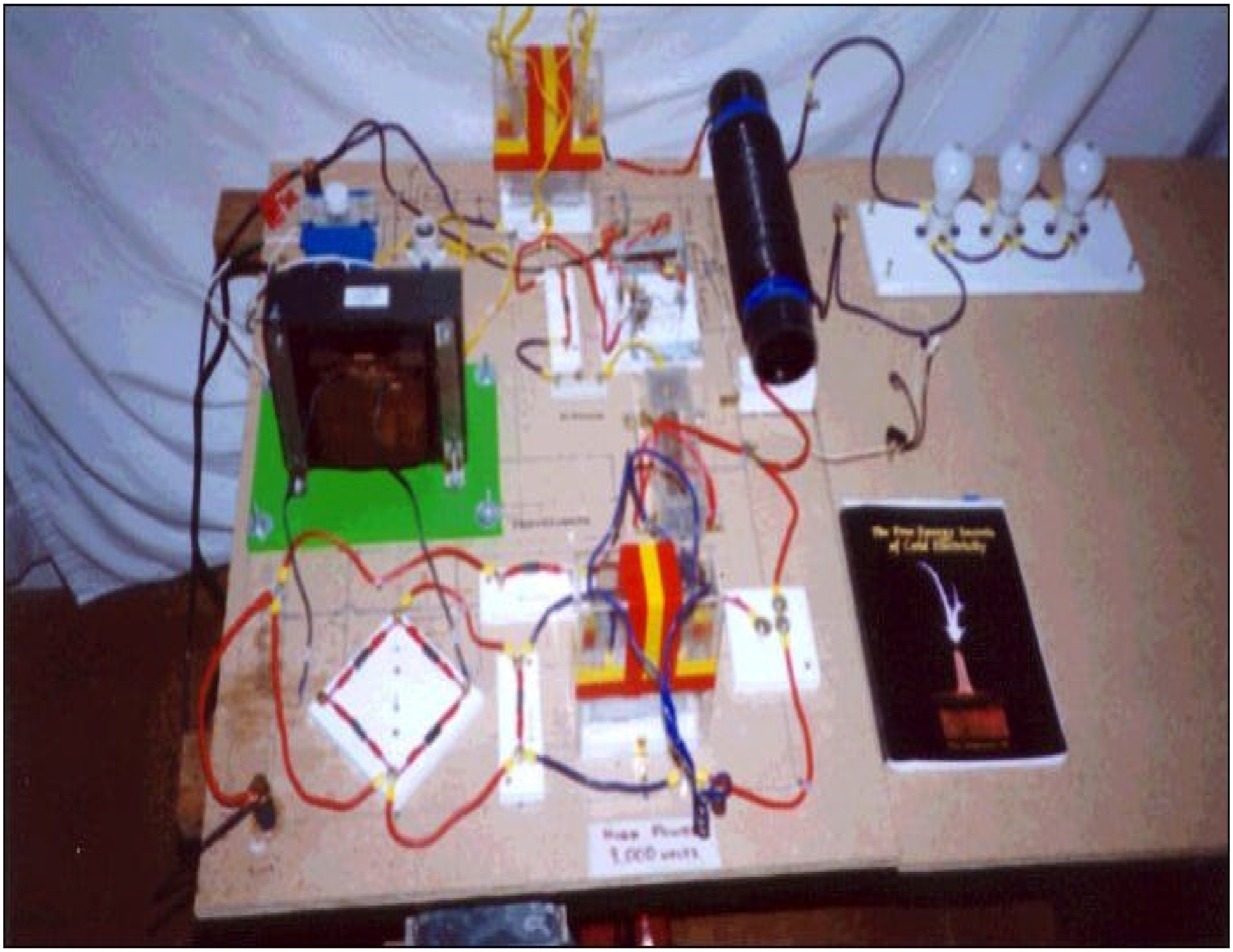
High Power
1,000 watts

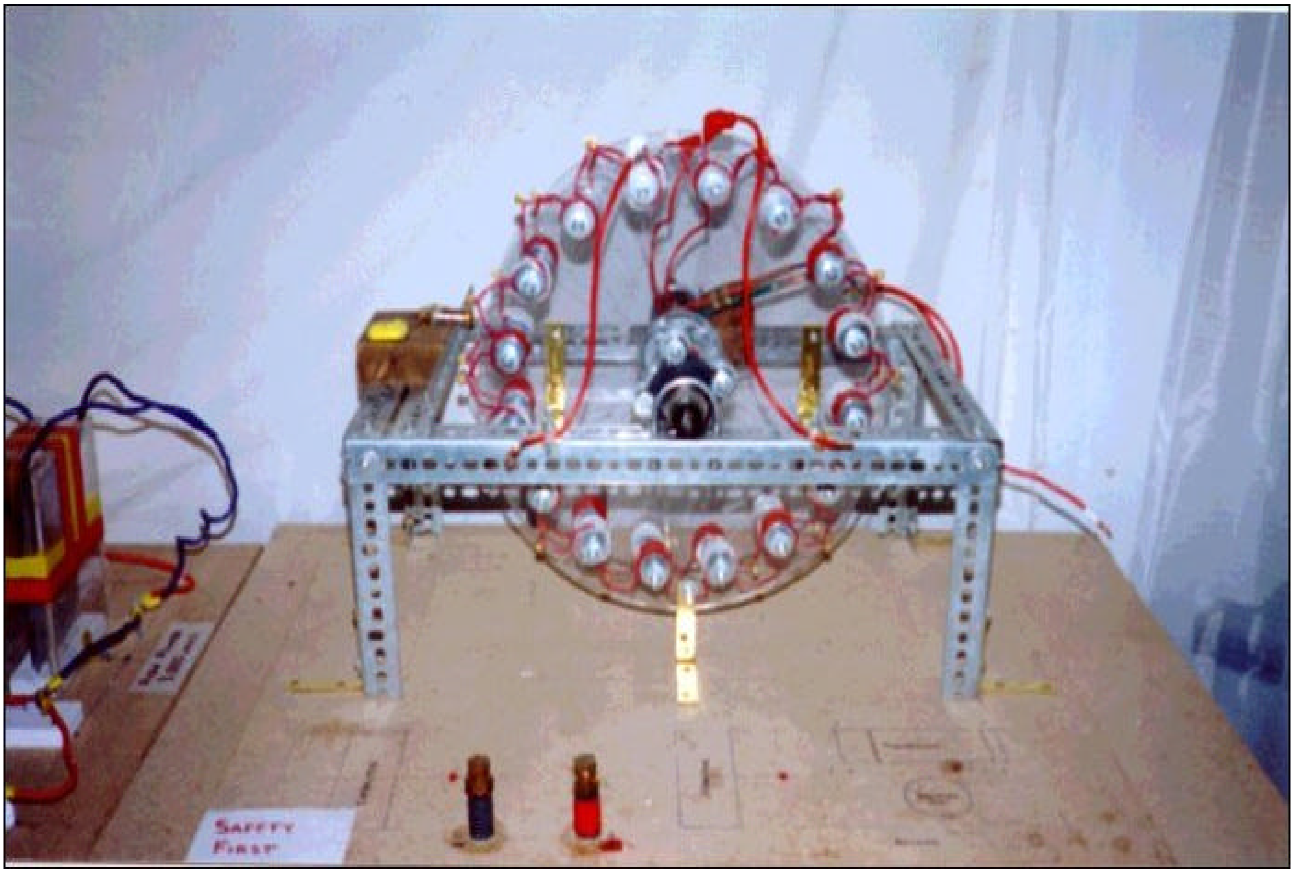
SAFETY
FIRST

All New Electronics Self-Teaching Guide, Third Edition

Harry Kybett and Earl Boysen

WILEY

Wiley Publishing, Inc.

All New Electronics Self-Teaching Guide, Third Edition

All New Electronics Self-Teaching Guide, Third Edition

Harry Kybett and Earl Boysen

WILEY

Wiley Publishing, Inc.

All New Electronics Self-Teaching Guide, Third Edition

Published by
Wiley Publishing, Inc.
10475 Crosspoint Boulevard
Indianapolis, IN 46256
www.wiley.com

Copyright © 2008 Wiley Publishing, Inc., Indianapolis, Indiana

Published simultaneously in Canada

ISBN: 978-0-470-28961-7

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Legal Department, Wiley Publishing, Inc., 10475 Crosspoint Blvd., Indianapolis, IN 46256, (317) 572-3447, fax (317) 572-4355, or online at http://www.wiley.com/go/permissions.

Limit of Liability/Disclaimer of Warranty: The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation warranties of fitness for a particular purpose. No warranty may be created or extended by sales or promotional materials. The advice and strategies contained herein may not be suitable for every situation. This work is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If professional assistance is required, the services of a competent professional person should be sought. Neither the publisher nor the author shall be liable for damages arising herefrom. The fact that an organization or Website is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or Website may provide or recommendations it may make. Further, readers should be aware that Internet Websites listed in this work may have changed or disappeared between when this work was written and when it is read.

For general information on our other products and services or to obtain technical support, please contact our Customer Care Department within the U.S. at (800) 762-2974, outside the U.S. at (317) 572-3993 or fax (317) 572-4002.

Library of Congress Cataloging-in-Publication Data is available from the publisher.

Trademarks: Wiley, the Wiley logo, and are related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. Wiley Publishing, Inc., is not associated with any product or vendor mentioned in this book.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

To my wonderful wife Nancy.
Thanks for wandering through life side by side with me.

About the Author

Earl Boysen is an engineer who, after 20 years working in the computer chip industry, decided to slow down and move to a quiet town in the state of Washington. Boysen is the co-author of three other books: *Electronics for Dummies* (Indianapolis: Wiley, 2005), *Electronics Projects for Dummies* (Indianapolis: Wiley, 2006), and *Nanotechnology for Dummies* (Indianapolis: Wiley, 2005). He lives with his wife, Nancy, in a house they built together, and finds himself busy as ever writing books and running two technology-focused Web sites. His site, `www.BuildingGadgets.com`, focuses on electronics circuits and concepts. The other site, `www.understandingnano.com`, provides clear explanations of nanotechnology topics. Boysen holds a masters degree in Engineering Physics from the University of Virginia.

Credits

Executive Editor
Carol Long

Development Editor
Kevin Shafer

Technical Editor
Rex Miller

Production Editor
Eric Charbonneau

Copy Editor
Mildred Sanchez

Editorial Manager
Mary Beth Wakefield

Production Manager
Tim Tate

Vice President and Executive Group Publisher
Richard Swadley

Vice President and Executive Publisher
Joseph B. Wikert

Project Coordinator, Cover
Lynsey Stanford

Proofreader
Sossity Smith

Indexer
Johnna VanHoose Dinse

Contents

Acknowledgments **xv**

Introduction **xvii**

Chapter 1 **DC Review and Pre-Test** **1**
Current Flow 1
Ohm's Law 4
Resistors in Series 6
Resistors in Parallel 7
Power 8
Small Currents 11
The Graph of Resistance 12
The Voltage Divider 14
The Current Divider 17
Switches 20
Capacitors in a DC Circuit 22
Summary 29
DC Pre-Test 30

Chapter 2 **The Diode** **35**
Understanding Diodes 36
The Diode Experiment 40
Diode Breakdown 55
The Zener Diode 58
Summary 65
Self-Test 66
  Answers to Self-Test 68

Chapter 3 **Introduction to the Transistor** **71**
Understanding Transistors 72
The Transistor Experiment 87

The Junction Field Effect Transistor 97
Summary 100
Self-Test 101
Answers to Self-Test 103

Chapter 4 The Transistor Switch 107
Turning the Transistor on 108
Turning the Transistor off 114
Why Transistors are Used as Switches 117
The Three-Transistor Switch 127
Alternative Base Switching 131
Switching The Jfet 137
The Jfet Experiment 138
Summary 142
Self-Test 142
Answers to Self-Test 145

Chapter 5 AC Pre-Test and Review 147
The Generator 148
Resistors in AC Circuits 152
Capacitors in AC Circuits 154
The Inductor in an AC Circuit 156
Resonance 158
Summary 160
Self-Test 161
Answers to Self-Test 162

Chapter 6 AC in Electronics 165
Capacitors in AC Circuits 165
Capacitors and Resistors in Series 167
The High Pass Filter Experiment 174
Phase Shift of an RC Circuit 180
Resistor and Capacitor in Parallel 186
Inductors in AC Circuits 190
Phase Shift for an RL Circuit 197
Summary 198
Self-Test 199
Answers to Self-Test 203

Chapter 7 Resonant Circuits 207
The Capacitor and Inductor in Series 208
The Output Curve 218
Introduction to Oscillators 233
Summary 236
Self-Test 237
Answers to Self-Test 239

Chapter 8 Transistor Amplifiers 241
Working with Transistor Amplifiers 242

The Transistor Amplifier Experiment 251
A Stable Amplifier 252
Biasing 256
The Emitter Follower 265
Analyzing an Amplifier 271
The JFET as an Amplifier 275
The Operational Amplifier 284
Summary 288
Self-Test 288
Answers to Self-Test 290

Chapter 9 Oscillators 293
Understanding Oscillators 293
Feedback 304
The Colpitts Oscillator 308
Optional Experiment 314
The Hartley Oscillator 314
The Armstrong Oscillator 316
Practical Oscillator Design 316
Simple Oscillator Design Procedure 318
Optional Experiment 320
Oscillator Troubleshooting Checklist 320
Summary and Applications 325
Self-Test 326
Answers to Self-Test 326

Chapter 10 The Transformer 329
Transformer Basics 329
Transformers in Communications Circuits 338
Summary and Applications 343
Self-Test 343
Answers to Self-Test 345

Chapter 11 Power Supply Circuits 347
Diodes in AC Circuits Produce Pulsating DC 348
Level DC (Smoothing Pulsating DC) 358
Summary 373
Self-Test 374
Answers to Self-Test 375

Chapter 12 Conclusion and Final Self-Test 377
Conclusion 377
Final Self-Test 378
Answers to Final Self-Test 387

Appendix A Glossary 393

Appendix B List of Symbols and Abbreviations 397

Appendix C Powers of Ten and Engineering Prefixes 401

Appendix D Standard Composition Resistor Values 403
Appendix E Supplemental Resources 405
Appendix F Equation Reference 409
Appendix G Schematic Symbols Used in This Book 413
Index 417

Acknowledgments

I want to first thank Harry Kybett for authoring the original version of this book many years ago. It was an honor to update such a classic book in the electronics field. Thanks also to Carol Long for bringing me on board with the project, and Kevin Shafer for his able project management of the book. My appreciation to Rex Miller for his excellent technical editing, and to Mildred Sanchez for handling all the mechanics of spelling and grammar in a thorough copy edit. Thanks to the people at Wiley, specifically Liz Britten for coordinating the creation of all the diagrams required and Eric Charbonneau for keeping the whole thing on schedule. Finally, thanks to my wife, Nancy Muir, for her advice and support throughout the writing of this book.

Introduction

The rapid growth of modern electronics is truly a phenomenon. All of the things you see in the marketplace today that utilize electronics either did not exist before 1960, or were crude by today's standards. Some of the many examples of modern electronics in the home include the small (but powerful) pocket calculator, the personal computer, the portable MP3 player, the DVD player, and digital cameras. Many industries have been founded, and older industries have been revamped, because of the availability and application of modern electronics in manufacturing processes, as well as in electronics products themselves.

Modern electronics *is* based on the transistor and its offspring — the integrated circuit (IC) and the microprocessor. These have short-circuited much of traditional electronic theory, revolutionized its practice, and set the whole field off on several new paths of discovery. This book is a first step to help you begin your journey down those paths.

What This Book Teaches

The traditional way of teaching electronics is often confusing. Too many students are left feeling that the real core of electronics is mysterious and arcane, akin to black magic. This just is not so. In fact, while many areas of our lives have become almost unbelievably complex, the study and practice of electronics in industry and as a hobby has surprisingly been made much simpler. *All New Electronics Self-Teaching Guide, Third Edition*, takes advantage of this simplicity and covers only those areas you actually need in modern electronics.

This book is for anyone who has a basic understanding of electronics concepts, but who wants to understand the operation of components found in the most common discrete circuits. The chapters in this book focus on circuits that are the building blocks for many common electronics devices, and on the very few important principles you need in working with electronics.

The arrangement and approach of this book is completely different from any other book on electronics in that it uses a ‘‘question-and-answer’’ approach to lead you into simple, but pertinent, experiments. This book steps you through calculations for every example in an easy-to-understand fashion, and you do not need to have a mathematical background beyond first-year algebra to follow along. In addition, this book omits the usual chapters on semiconductor physics, because you don’t need these in the early stages of working with electronics.

Electronics is a very easy technology, which anyone can understand with very little effort. This book focuses on how to apply the few basic principles that are the basis of modern electronic practice. Understanding the circuits composed of discrete components and the applicable calculations discussed in this book is useful not only in building and designing circuits, but it also helps you to work with ICs. That’s because ICs use miniaturized components such as transistors, diodes, capacitors, and resistors that function based on the same rules as discrete components (along with some specific rules necessitated by the extremely small size of IC components).

How This Book Is Organized

This book is organized with sets of problems that challenge you to think through a concept or procedure, and then provides answers so you can constantly check your progress and understanding. Specifically, the chapters in this book are organized as follows:

- *Chapter* 1: *DC Review and Pre-Test* — This chapter provides a review and pre-test on the basic concepts, components, and calculations that are useful when working with direct current (DC) circuits.
- *Chapter* 2: *The Diode* — Here you learn about the diode, including how you use diodes in DC circuits, the main characteristics of diodes, and calculations you can use to determine current, voltage, and power.
- *Chapter* 3: *Introduction to the Transistor* — In this chapter, you learn about the transistor and its use in circuits. You also discover how bipolar junction transistors (BJTs) and junction field effect transistors (JFETs) control the flow of electric current.

- *Chapter* 4: *The Transistor Switch* — This chapter examines the most simple and widespread application of a transistor: switching. In addition to learning how to design a transistor circuit to drive a particular load, you also compare the switching action of a JFET and a BJT.
- *Chapter* 5: *AC Pre-Test and Review* — This chapter examines the basic concepts and equations for alternating current (AC) circuits. You discover how to use resistors and capacitors in AC circuits, and learn related calculations.
- *Chapter* 6: *AC in Electronics* — This chapter looks at how resistors, capacitors, and inductors are used in high pass filters and low pass filters to pass or block AC signals above or below a certain frequency.
- *Chapter* 7: *Resonant Circuits* — This chapter examines the use of capacitors, inductors, and resistors in circuits called bandpass filters and band-reject filters, which pass or block AC signals in a band of frequencies centered around the resonant frequency of the circuit. You also learn how to calculate the resonance frequency and bandwidth of these circuits. This chapter also introduces you to how to use resonant circuits in oscillators.
- *Chapter* 8: *Transistor Amplifiers* — Here you explore the use of transistor amplifiers to amplify electrical signals. In addition to examining the fundamental steps used to design BJT-based amplifiers, you learn how to use JFETs and operational amplifiers (op-amps) in amplifier circuits.
- *Chapter* 9: *Oscillators* — This chapter introduces you to the oscillator, a circuit that produces a continuous AC output signal. You learn how an oscillator works, and step through the procedure to design and build an oscillator.
- *Chapter* 10: *The Transformer* — In this chapter, you discover how AC voltage is converted by a transformer to a higher or lower voltage. You learn how a transformer makes this conversion, and how to calculate the output voltage that results.
- *Chapter* 11: *Power Supply Circuits* — In this chapter, you find out how power supplies convert AC to DC with a circuit made up of transformers, diodes, capacitors, and resistors. You also learn how to calculate the values of components that produce a specified DC output voltage for a power supply circuit.
- *Chapter* 12: *Conclusion and Final Self-Test* — This chapter enables you to check your understanding of the topics presented in this book through the use of a final self-test that allows you to assess your overall knowledge of electronics.

In addition, this book contains the following appendixes for easy reference:

- *Appendix A: Glossary* — This glossary provides key electronics terms and their definitions.
- *Appendix B: List of Symbols and Abbreviations* — This appendix provides a handy reference of commonly used symbols and abbreviations.
- *Appendix C: Powers of Ten and Engineering Prefixes* — This guide lists prefixes that are commonly used in electronics, along with their corresponding values.
- *Appendix D: Standard Composition Resistor Values* — This appendix provides standard resistance values for the carbon composition resistor, the most commonly used type of resistor.
- *Appendix E: Supplemental Resources* — This appendix provides references to helpful Web sites, books, and magazines.
- *Appendix F: Equation Reference* — This is a quick guide to commonly used equations, along with chapter and problem references showing you where they are first introduced in this book.
- *Appendix G: Schematic Symbols Used in This Book* — This appendix provides a quick guide to schematic symbols used in the problems found throughout the book.

Conventions Used In This Book

As you study electronics you will find that there is some variation in terminology and the way that circuits are drawn. Here are two conventions followed in this book that you should be aware of:

- The discussions in this book use V to stand for voltage, versus E, which you will see used in some other books.
- In all circuit diagrams in this book, intersecting lines indicate an electrical connection. (Some other books use a dot at the intersection of lines to indicate a connection.) If a semicircle appears at the intersection of two lines, it indicates that there is no connection. See Figure 9.5 for an example of this.

How to Use This Book

This book assumes that you have some knowledge of basic electronics such as Ohm's law and current flow. If you have read a textbook or taken a course on electronics, or if you have worked with electronics, you probably have the prerequisite knowledge. If not, you should read a book such as *Electronics for Dummies* (Indianapolis: Wiley, 2005) to get the necessary background for this book. You can also go to the author's Web site, `www.BuildingGadgets.com`, and use the Tutorial links to find useful online lessons in electronics. In addition, Chapters 1 and 5 allow you to test your knowledge and review the necessary basics of electronics.

Note that you should read the chapters in order, because often later material depends on concepts and skills covered in earlier chapters.

All New Electronics Self-Teaching Guide, Third Edition, is presented in a self-teaching format that allows you to learn easily and at your own pace. The material is presented in numbered sections called *problems*. Each problem presents some new information and gives you a question to answer, or an experiment to try. To learn most effectively, you should cover up the answers with a sheet of paper and try to answer each question on your own. Then, compare your answer with the correct answer that follows. If you miss a question, correct your answer and then go on. If you miss many in a row, go back and review the previous section, or you may miss the point of the material that follows.

Be sure to try to do all of the experiments. They are very easy and help reinforce your learning of the subject matter. If you don't have the equipment to do an experiment, simply reading through it will help you to better understand the concepts it demonstrates.

When you reach the end of a chapter, evaluate your learning by taking the Self-Test. If you miss any questions, review the related parts of the chapter again. If you do well on the Self-Test, you're ready to go on to the next chapter. You may also find the Self-Test useful as a review before you start the next chapter. At the end of the book, there is a Final Self-Test that allows you to assess your overall learning.

Go through this book at your own pace. You can work through this book alone, or you can use it in conjunction with a course. If you use the book alone, it serves as an introduction to electronics, but is not a complete course. For that reason, at the end of the book are some suggestions for further reading and online resources. Also, at the back of the book is a table of symbols and abbreviations, which are useful for reference and review.

Now you're ready to learn *electronics*!

CHAPTER

1

DC Review and Pre-Test

Electronics cannot be studied without first understanding the basics of electricity. This chapter is a review and pre-test on those aspects of direct current (DC) that apply to electronics. By no means does it cover the whole DC theory, but merely those topics that are essential to simple electronics. This chapter will review the following:

- Current flow
- Potential or voltage difference
- Ohm's law
- Resistors in series and parallel
- Power
- Small currents
- Resistance graphs
- Kirchhoff's voltage and current laws
- Voltage and current dividers
- Switches
- Capacitor charging and discharging
- Capacitors in series and parallel

Current Flow

1 Electrical and electronic devices work because of an electric current.

Question

What is an electric current?

____

Answer

An electric current is a flow of electric charge. The electric charge usually consists of negatively charged electrons. However, in semiconductors, there are also positive charge carriers called *holes*.

2 There are several methods that can be used to generate an electric current.

Question

Write at least three ways an electron flow (or current) can be generated.

____

____

____

Answers

The following is a list of the most common ways to generate current:

Magnetically — The induction of electrons in a wire rotating within a magnetic field. An example of this would be generators turned by water, wind, or steam, or the fan belt in a car.

Chemically — Involving electrochemical generation of electrons by reactions between chemicals and electrodes (as in batteries).

Photovoltaic generation of electrons — When light strikes semiconductor crystals (as in solar cells).

Less common methods to generate an electric current include the following:

Thermal generation — Using temperature differences between thermocouple junctions. Thermal generation is used in generators on spacecraft that are fueled by radioactive material.

Electrochemical reaction — Occurring between hydrogen, oxygen, and electrodes (fuel cells).

Piezoelectrical — Involving mechanical deformation of piezoelectric substances. For example, piezoelectric material in the heels of shoes power LEDs that light up when you walk.

3 Most of the simple examples in this book will contain a battery as the voltage source. As such, the source provides a potential difference to a circuit that will enable a current to flow. An *electric current* is a flow of electric charge. In the case of a battery, electrons are the electric charge, and they flow from the terminal that has an excess number of electrons to the terminal that has a deficiency of electrons. This flow takes place in any complete circuit that is connected to battery terminals. It is this difference of charge that creates the potential difference in the battery. The electrons are trying to balance the difference.

Because electrons have a negative charge, they actually flow from the negative terminal and return to the positive terminal. We call this direction of flow *electron flow*. Most books, however, use current flow, which is in the opposite direction. It is referred to as *conventional current flow* or simply *current flow*. In this book, the term conventional current flow is used in all circuits.

Later in this book, you will see that many semiconductor devices have a symbol that contains an arrowhead pointing in the direction of conventional current flow.

Questions

A. Draw arrows to show the current flow in Figure 1-1. The symbol for the battery shows its polarity.

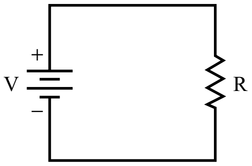

Figure 1-1

B. What indicates that a potential difference is present?

____

C. What does the potential difference cause?

____

D. What will happen if the battery is reversed?

____

Answers

A. See Figure 1-2.

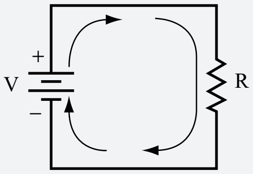

Figure 1-2

B. The battery symbol indicates that a difference of potential, also called *voltage*, is being supplied to the circuit.

C. Voltage causes current to flow if there is a complete circuit present, as shown in Figure 1-1.

D. The current will flow in the opposite direction.

Ohm's Law

4 Ohm's law states the fundamental relationship between voltage, current, and resistance.

Question

What is the algebraic formula for Ohm's law?

Answer

$V = I \times R$

This is the most basic equation in electricity, and you should know it well. Note that some electronics books state Ohm's law as $E = IR$. E and V are both symbols for voltage. This book uses V throughout. Also, in this formula, resistance is the opposition to current flow. Note that larger resistance results in smaller current for any given voltage.

5 Use Ohm's law to find the answers in this problem.

Questions

What is the voltage for each combination of resistance and current values?

A. R = 20 ohms I = 0.5 amperes V = ____________

B. R = 560 ohms I = 0.02 amperes V = ____________

C. R = 1000 ohms I = 0.01 amperes V = ____________

D. R = 20 ohms I = 1.5 amperes V = ____________

Answers

A. 10 volts

B. 11.2 volts

C. 10 volts

D. 30 volts

6 You can rearrange Ohm's law to calculate current values.

Questions

What is the current for each combination of voltage and resistance values?

A. V = 1 volt R = 2 ohms I = ____________

B. V = 2 volts R = 10 ohms I = ____________

C. V = 10 volts R = 3 ohms I = ____________

D. V = 120 volts R = 100 ohms I = ____________

Answers

A. 0.5 amperes

B. 0.2 amperes

C. 3.3 amperes

D. 1.2 amperes

7 You can rearrange Ohm's law to calculate resistance values.

Questions

What is the resistance for each combination of voltage and current values?

A. V = 1 volt I = 1 ampere R = ____________

B. V = 2 volts I = 0.5 ampere R = ____________

C. V = 10 volts I = 3 amperes R = ____________

D. V = 50 volts I = 20 amperes R = ____________

Answers

A. 1 ohm

B. 4 ohms

C. 3.3 ohms

D. 2.5 ohms

8 Work through these examples. In each case, two factors are given and you must find the third.

Questions

What are the missing values?

A. 12 volts and 10 ohms. Find the current.

____

B. 24 volts and 8 amperes. Find the resistance.

____

C. 5 amperes and 75 ohms. Find the voltage.

____

Answers

A. 1.2 amperes

B. 3 ohms

C. 375 volts

Resistors in Series

9 Resistors can be connected in series. Figure 1-3 shows two resistors in series.

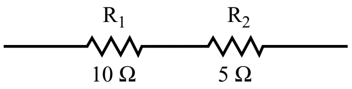

Figure 1-3

Question

What is their total resistance?

Answer

$R_T = R_1 + R_2 = 10$ ohms $+ 5$ ohms $= 15$ ohms

The total resistance is often called the *equivalent series resistance*, and is denoted as R_{eq}.

Resistors in Parallel

10 Resistors can be connected in parallel, as shown in Figure 1-4.

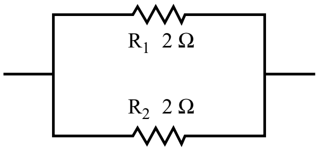

Figure 1-4

Question

What is the total resistance here?

Answer

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} = \frac{1}{2} + \frac{1}{2} = 1;\ \text{thus } R_T = 1 \text{ ohm}$$

R_T is often called the *equivalent parallel resistance*.

11 The simple formula from problem 10 can be extended to include as many resistors as desired.

Question

What is the formula for three resistors in parallel?

Answer

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

You will often see this formula in the following form:

$$R_T = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}}$$

12 In the following exercises, two resistors are connected in parallel.

Questions

What is the total or equivalent resistance?

A. $R_1 = 1$ ohm $R_2 = 1$ ohm $R_T =$ ____________

B. $R_1 = 1000$ ohms $R_2 = 500$ ohms $R_T =$ ____________

C. $R_1 = 3600$ ohms $R_2 = 1800$ ohms $R_T =$ ____________

Answers

A. 0.5 ohms

B. 333 ohms

C. 1200 ohms

Note that R_T is always smaller than the smallest of the resistors in parallel.

Power

13 When current flows through a resistor, it dissipates power, usually in the form of heat. Power is expressed in terms of watts.

Question

What is the formula for power?

____

Answer

There are three formulas for calculating power:

$$P = VI \text{ or } P = I^2R \text{ or } P = \frac{V^2}{R}$$

14 The first formula shown in problem 13 allows power to be calculated when only the voltage and current are known.

Questions

What is the power dissipated by a resistor for the following voltage and current values?

A. $V = 10$ volts $I = 3$ amperes $P =$ ______________

B. $V = 100$ volts $I = 5$ amperes $P =$ ______________

C. $V = 120$ volts $I = 10$ amperes $P =$ ______________

Answers

A. 30 watts

B. 500 watts, or 0.5 kilowatts

C. 1200 watts, or 1.2 kilowatts

15 The second formula shown in problem 13 allows power to be calculated when only the current and resistance are known.

Questions

What is the power dissipated by a resistor given the following resistance and current values?

A. $R = 20$ ohm $I = 0.5$ ampere $P =$ ______________

B. $R = 560$ ohms $I = 0.02$ ampere $P =$ ______________

C. $V = 1$ volt $R = 2$ ohms $P =$ ______________

D. $V = 2$ volt $R = 10$ ohms $P =$ ______________

Answers

A. 5 watts

B. 0.224 watts

C. 0.5 watts

D. 0.4 watts

16 Resistors used in electronics generally are manufactured in standard values with regard to resistance and power rating. Appendix D shows a table of standard resistance values. Quite often, when a certain resistance value is needed in a circuit, you must choose the closest standard value. This is the case in several examples in this book.

You must also choose a resistor with the power rating in mind. You should never place a resistor in a circuit that would require that resistor to dissipate more power than its rating specifies.

Questions

If standard power ratings for carbon composition resistors are 1/4, 1/2, 1, and 2 watts, what power ratings should be selected for the resistors that were used for the calculations in problem 15?

A. For 5 watts________________

B. For 0.224 watts_____________

C. For 0.5 watts_______________

D. For 0.4 watts_______________

Answers

A. 5 watt (or greater)

B. 1/4 watt (or greater)

C. 1/2 watt (or greater)

D. 1/2 watt (or greater)

Most electronics circuits use low power carbon composition resistors. For higher power levels (such as the 5 watt requirement in question A), other types of resistors are available.

Small Currents

17 Although currents much larger than 1 ampere are used in heavy industrial equipment, in most electronic circuits, only fractions of an ampere are required.

Questions

A. What is the meaning of the term *milliampere*?

____

B. What does the term *microampere* mean?

____

Answers

A. A milliampere is one-thousandth of an ampere (that is, 1/1000 or 0.001 amperes). It is abbreviated mA.

B. A microampere is one-millionth of an ampere (that is, 1/1,000,000 or 0.000001 amperes). It is abbreviated μA.

18 In electronics, the values of resistance normally encountered are quite high. Often, thousands of ohms and occasionally even millions of ohms are used.

Questions

A. What does kΩ mean when it refers to a resistor?

____

B. What does MΩ mean when it refers to a resistor?

____

Answers

A. Kilohm (k = kilo, Ω = ohm). The resistance value is thousands of ohms. Thus, 1 kΩ = 1,000 ohms, 2 kΩ = 2,000 ohms, and 5.6 kΩ = 5,600 ohms.

B. Megohm (M = mega, Ω = ohm). The resistance value is millions of ohms. Thus, 1 MΩ = 1,000,000 ohms, and 2.2 MΩ = 2,200,000 ohms.

19 The following exercise is typical of many performed in transistor circuits. In this example, 6 V is applied across a resistor and 5 mA of current is required to flow through the resistor.

Question

What value of resistance must be used and what power will it dissipate?

R = ______________ P = ______________

Answer

$$R = \frac{V}{I} = \frac{6\text{ V}}{5\text{ mA}} = \frac{6}{0.005} = 1200\text{ ohms} = 1.2\text{ k}\Omega$$

$$P = V \times I = 6 \times 0.005 = 0.030\text{ watts} = 30\text{ mW}$$

20 Now, try these two simple examples.

Questions

What is the missing value?

A. 50 volts and 10 mA. Find the resistance.

____

B. 1 volt and 1 MΩ. Find the current.

____

Answers

A. 5 kΩ

B. 1 μA

The Graph of Resistance

21 The voltage drop across a resistor and the current flowing through it can be plotted on a simple graph. This graph is called a *V-I curve*.

Consider a simple circuit in which a battery is connected across a 1 kΩ resistor.

Questions

A. Find the current flowing if a 10 V battery is used.

____

B. Find the current when a 1 V battery is used.

____

C. Now find the current when a 20 V battery is used.

____

Answers

A. 10 mA

B. 1 mA

C. 20 mA

22 Plot the points of battery voltage and current flow from problem 21 on the graph shown in Figure 1-5, and connect them together.

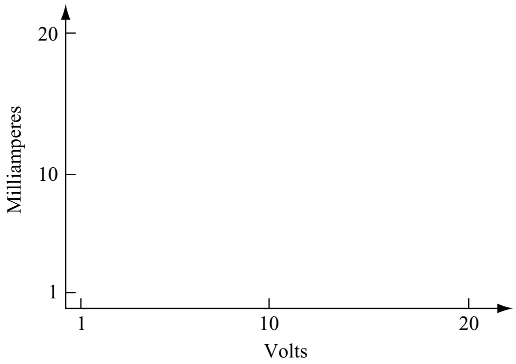

Figure 1-5

Answer

You should have drawn a straight line, as in the graph shown in Figure 1-6.

Sometimes you need to calculate the slope of the line on a graph. To do this, pick two points and call them A and B.

For point A let V = 5 volts and I = 5 mA

For point B let V = 20 volts and I = 20 mA

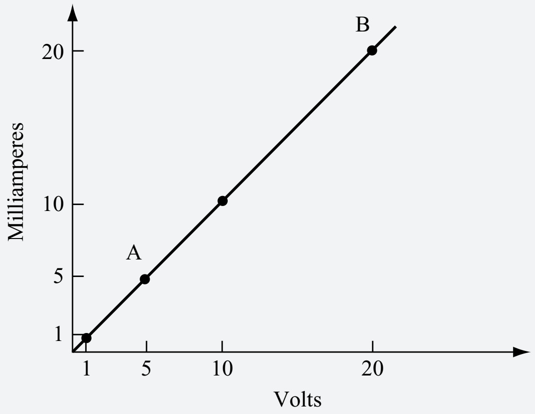

Figure 1-6

The slope can be calculated with the following formula:

$$\text{Slope} = \frac{V_B - V_A}{I_B - I_A} = \frac{20\text{ volts} - 5\text{ volts}}{20\text{ mA} - 5\text{ mA}} = \frac{15\text{ volts}}{15\text{ mA}} = \frac{15\text{ volts}}{0.015\text{ A}} = 1\text{ k}\Omega$$

In other words, the slope of the line is equal to the resistance.

Later, you will learn about V-I curves for other components. They have several uses, and often they are not straight lines.

The Voltage Divider

23 The circuit shown in Figure 1-7 is called a *voltage divider*. It is the basis for many important theoretical and practical ideas you encounter throughout the entire field of electronics.

The object of this circuit is to create an output voltage (V_0) that you can control based upon the two resistors and the input voltage. Note that V_0 is also the *voltage drop* across R_2.

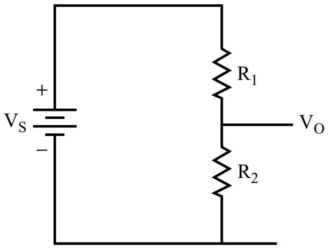

Figure 1-7

Question

What is the formula for V_0?

Answer

$$V_o = V_S \times \frac{R_2}{R_1 + R_2}$$

Note that $R_1 + R_2 = R_T$, the total resistance of the circuit.

24 A simple example will demonstrate the use of this formula.

Question

For the circuit shown in Figure 1-8, what is V_0?

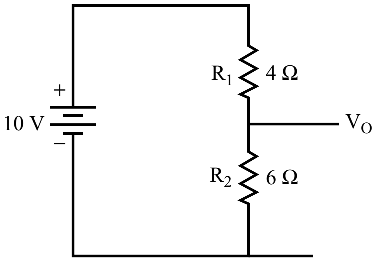

Figure 1-8

Answer

$$\begin{aligned} V_O &= V_S \times \frac{R_2}{R_1 + R_2} \\ &= 10 \times \frac{6}{4+6} \\ &= 10 \times \frac{6}{10} \\ &= 6 \text{ volts} \end{aligned}$$

25 Now, try these problems.

Questions

What is the output voltage for each combination of supply voltage and resistance?

A. $V_S = 1$ volt $R_1 = 1$ ohm $R_2 = 1$ ohm $V_0 =$ ______________

B. $V_S = 6$ volts $R_1 = 4$ ohms $R_2 = 2$ ohms $V_0 =$ ______________

C. $V_S = 10$ volts $R_1 = 3.3$ kΩ $R_2 = 5.6$ kΩ $V_0 =$ ______________

D. $V_S = 28$ volts $R_1 = 22$ kΩ $R_2 = 6.2$ kΩ $V_0 =$ ______________

Answers

A. 0.5 volts

B. 2 volts

C. 6.3 volts

D. 6.16 volts

26 The output voltage from the voltage divider is always less than the applied voltage. Voltage dividers are often used to apply specific voltages to different components in a circuit. Use the voltage divider equation to answer the following questions.

Questions

A. What is the voltage drop across the 22k resistor for question D of problem 25?

____

B. What total voltage do you get if you add this voltage drop to the voltage drop across the 6.2k resistor?

____

Answers

A. 21.84 volts

B. The sum is 28 volts.

Note that the voltages across the two resistors add up to the supply voltage. This is an example of Kirchhoff's voltage law (KVL), which simply means that the voltage supplied to a circuit must equal the sum of the voltage drops in the circuit. In this book, KVL will often be used without actual reference to the law.

Also notice that voltage drop across a resistor is proportional to the resistor's value. Therefore, if one resistor has a greater value than another in a series circuit, the voltage drop across the higher value resistor will be greater.

The Current Divider

27 In the circuit shown in Figure 1-9, the current splits or divides between the two resistors that are connected in parallel.

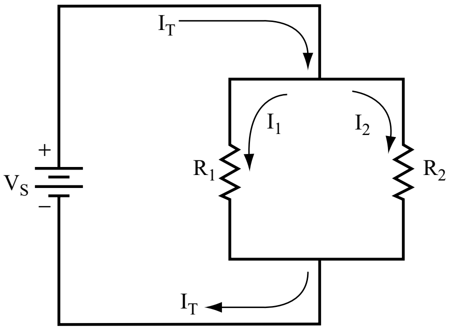

Figure 1-9

I_T splits into the individual currents I_1 and I_2, and then these recombine to form I_T.

Question

Which of the following relationships are valid for this circuit?

A. $V_S = R_1I_1$

B. $V_S = R_2I_2$

C. $R_1I_1 = R_2I_2$

D. $I_1/I_2 = R_2/R_1$

Answer

All of them are valid.

28 When solving current divider problems, follow these steps:

1. Set up the ratio of the resistors and currents.

   $I_1/I_2 = R_2/R_1$

2. Rearrange the ratio to give I_2 in terms of I_1:

   $$I_2 = I_1 \times \frac{R_1}{R_2}$$

3. From the fact that $I_T = I_1 + I_2$, express I_T in terms of I_1 only.
4. Now, find I_1.
5. Now, find the remaining current (I_2).

Question

The values of two resistors in parallel and the total current flowing through the circuit are shown in Figure 1-10. What is the current through each individual resistor?

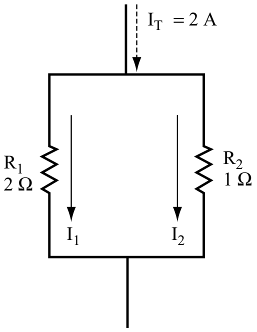

Figure 1-10

Answers

Working through the steps as shown:

1. $I_1/I_2 = R_2/R_1 = 1/2$
2. $I_2 = 2I_1$
3. $I_T = I_1 + I_2 = I_1 + 2I_1 = 3I_1$
4. $I_1 = I_T/3 = 2/3\ \text{A}$
5. $I_2 = 2I_1 = 4/3\ \text{A}$

29 Now, try these problems. In each case, the total current and the two resistors are given. Find I_1 and I_2.

Questions

A. $I_T = 30\ mA$, $R_1 = 12\ k\Omega$, $R_2 = 6\ k\Omega$

B. $I_T = 133\ mA$, $R_1 = 1\ k\Omega$, $R_2 = 3\ k\Omega$

C. What current do you get if you add I_1 and I_2?

Answers

A. $I_1 = 10\ mA$, $I_2 = 20\ mA$

B. $I_1 = 100\ mA$, $I_2 = 33\ mA$

C. They add back together to give you the total current supplied to the parallel circuit.

Note that question C is actually a demonstration of Kirchhoff's current law (KCL). This law simply stated says that the total current entering a junction in a circuit must equal the sum of the currents leaving that junction. This law will also be used on numerous occasions in later chapters. KVL and KCL together form the basis for many techniques and methods of analysis that are used in the application of circuit analysis.

Also, notice that the current through a resistor is inversely proportional to the resistor's value. Therefore, if one resistor is larger than another in a parallel circuit, the current flowing through the higher value resistor will be the smaller of the two. Check your results for this problem to verify this.

30 You can also use the following equation to calculate the current flowing through a resistor in a two-branch parallel circuit:

$$I_1 = \frac{(I_T)(R_2)}{(R_1 + R_2)}$$

Question

Write the equation for the current I_2.

Check the answers for the previous problem using these equations.

Answer

$$I_2 = \frac{(I_T)(R_1)}{(R_1 + R_2)}$$

Note that the current through one branch of a two-branch circuit is equal to the total current times the resistance of the opposite branch, divided by the sum of the resistances of both branches. This is an easy formula to remember.

Switches

31 A *mechanical switch* is a device that completes or breaks a circuit. The most familiar use is that of applying power to turn a device on or off. A switch can also permit a signal to pass from one place to another, prevent its passage, or route a signal to one of several places.

In this book we deal with two types of switches. The first is the simple on-off switch, also called a *single pole single throw* switch. The second is the *single pole double throw* switch. The circuit symbols for each are shown in Figure 1-11.

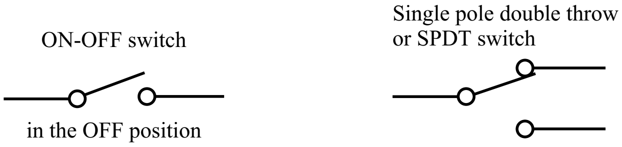

Figure 1-11

Two important facts about switches must be known.

- A closed (or ON) switch has the total circuit current flowing through it. There is *no* voltage drop across its terminals.
- An open (or OFF) switch has *no* current flowing through it. The full circuit voltage appears between its terminals.

The circuit shown in Figure 1-12 includes a closed switch.

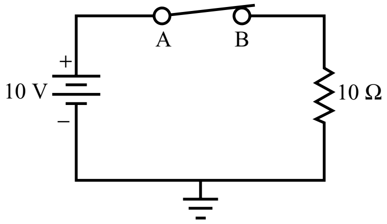

Figure 1-12

Questions

A. What is the current flowing through the switch?

____

B. What is the voltage at point A and point B with respect to ground?

____

C. What is the voltage drop across the switch?

____

Answers

A. $\frac{10 \text{ V}}{10 \text{ ohms}} = 1$ ampere

B. $V_A = V_B = 10$ V

C. 0 V (There is no voltage drop because both terminals are at the same voltage.)

32 The circuit shown in Figure 1-13 includes an open switch.

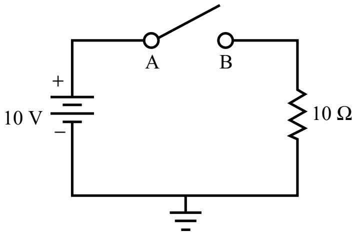

Figure 1-13

Questions

A. What is the voltage at point A and point B?

____

B. How much current is flowing through the switch?

____

C. What is the voltage drop across the switch?

____

Answers

A. $V_A = 10$ V; $V_B = 0$ V

B. No current is flowing because the switch is open.

C. 10 V. If the switch is open, point A is the same voltage as the positive battery terminal and point B is the same voltage as the negative battery terminal.

33 The circuit shown in Figure 1-14 includes a single pole double throw switch. The position of the switch determines whether lamp A or lamp B is lit.

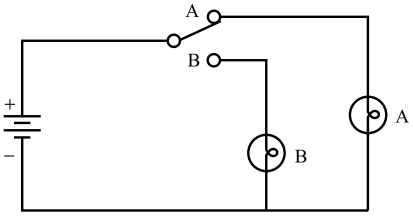

Figure 1-14

Questions

A. In the position shown, which lamp is lit?

B. Can both lamps be lit simultaneously?

Answers

A. Lamp A.

B. No, one or the other must be off.

Capacitors in a DC Circuit

34 Capacitors are used extensively in electronics. They are used in both alternating current (AC) and DC circuits. Their main use in DC electronics is to become charged, hold the charge, and, at a specific time, release the charge.

The capacitor shown in Figure 1-15 charges when the switch is closed.

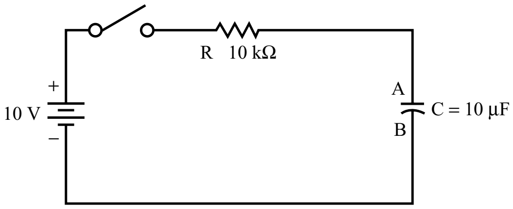

Figure 1-15

Question

To what final voltage will the capacitor charge?

Answer

It will charge up to 10 V. It will charge up to the voltage that would appear across an open circuit located at the same place where the capacitor is located.

35 How long does it take to reach this voltage? This is a most important question, with many practical applications. To find the answer you have to know the *time constant* (τ) of the circuit.

Questions

A. What is the formula for the time constant of this type of circuit?

B. What is the time constant for the circuit shown in Figure 1-15?

C. How long does it take the capacitor to reach 10 V?

D. To what voltage level does it charge in one time constant?

Answers

A. $\tau = R \times C$

B. $\tau = 10\ \text{k}\Omega \times 10\ \mu\text{F} = 10{,}000\ \Omega \times 0.00001\ \text{F} = 0.1$ seconds. (Convert resistance values to ohms and capacitance values to farads for this calculation.)

C. Approximately 5 time constants, or about 0.5 seconds.

D. 63 percent of the final voltage, or about 6.3 V.

36 The capacitor will not begin charging until the switch is closed. When a capacitor is uncharged or discharged, it has the same voltage on both plates.

Questions

A. What will be the voltage on plate A and plate B of the capacitor in Figure 1-15 before the switch is closed?

B. When the switch is closed, what will happen to the voltage on plate A?

C. What will happen to the voltage on plate B?

D. What will be the voltage on plate A after one time constant?

Answers

A. Both will be at 0 V if the capacitor is totally discharged.

B. It will rise towards 10 V.

C. It will stay at 0 V.

D. About 6.3 V.

37 The *capacitor charging graph* in Figure 1-16 shows for how many time constants a voltage must be applied to a capacitor before it reaches a given percentage of the applied voltage.

Questions

A. What is this type of curve called?

B. What is it used for?

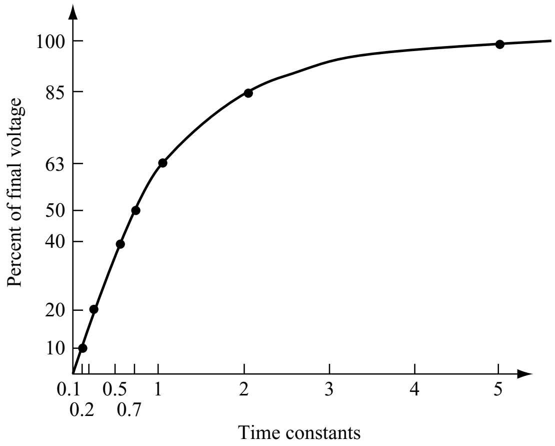

Figure 1-16

Answers

A. It is called an *exponential* curve.

B. It is used to calculate how far a capacitor has charged in a given time.

38 In the following, examples a resistor and a capacitor are in series. Calculate the time constant, how long it takes the capacitor to fully charge, and the voltage level after one time constant if a 10 V battery is used.

Questions

A. R = 1 kΩ C = 1000 μF

B. R = 330 kΩ C = 0.05 μF

Answers

A. $\tau = 1$ second; charge time = 5 seconds; $V_C = 6.3$ V

B. $\tau = 16.5$ ms; charge time = 82.5 ms; $V_C = 6.3$ V (The abbreviation ms indicates milliseconds.)

39 The circuit shown in Figure 1-17 uses a double pole switch to create a discharge path for the capacitor.

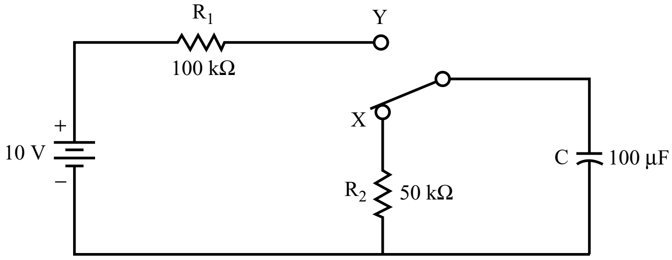

Figure 1-17

Questions

A. With the switch in position X, what is the voltage on each plate of the capacitor?

B. When the switch is moved to position Y, the capacitor begins to charge. What is its charging time constant?

C. How long does it take to fully charge the capacitor?

Answers

A. 0 V

B. $\tau = R \times C = (100\ k\Omega)\ (100\ \mu F) = 10$ secs

C. Approximately 50 seconds

40 Suppose that the switch shown in Figure 1-17 is moved back to position X once the capacitor is fully charged.

Questions

A. What is the discharge time constant of the capacitor?

B. How long does it take to fully discharge the capacitor?

Answers

A. $\tau = R \times C = (50\ k\Omega)\ (100\ \mu F) = 5$ seconds (discharging through the 50 kΩ resistor)

B. Approximately 25 seconds

The circuit powering a camera flash is an example of a capacitor's ability to store charge and then discharge upon demand. While you wait for the flash unit to charge, the camera is using its battery to charge a capacitor. Once the capacitor is charged, it holds that charge until you click the shutter button, causing the capacitor to discharge, which powers the flash.

41 Capacitors can be connected in parallel, as shown in Figure 1-18.

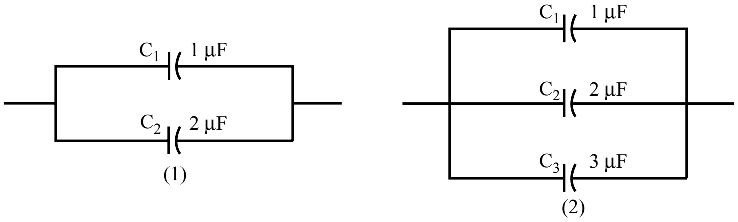

Figure 1-18

Questions

A. What is the formula for the total capacitance?

____

B. What is the total capacitance in circuit 1?

____

C. What is the total capacitance in circuit 2?

____

Answers

A. $C_T = C_1 + C_2 + C_3 + \cdots + C_N$

B. $C_T = 1 + 2 = 3\ \mu F$

C. $C_T = 1 + 2 + 3 = 6\ \mu F$

In other words, the total capacitance is found by simple addition of the capacitor values.

42 Capacitors can be placed in series, as shown in Figure 1-19.

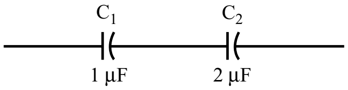

Figure 1-19

Questions

A. What is the formula for the total capacitance?

B. In Figure 1-19, what is the total capacitance?

Answers

A. $\frac{1}{C_T} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots + \frac{1}{C_N}$

B. $\frac{1}{C_T} = \frac{1}{1} + \frac{1}{2} = 1\frac{1}{2} = \frac{3}{2}$; thus $C_T = \frac{2}{3}$

43 In each of these examples the capacitors are placed in series. Find the total capacitance.

Questions

A. $C_1 = 10$ μF $C_2 = 5$ μF

B. $C_1 = 220$ μF $C_2 = 330$ μF $C_3 = 470$ μF

C. $C_1 = 0.33$ μF $C_2 = 0.47$ μF $C_3 = 0.68$ μF

Answers

A. 3.3 μF

B. 103.06 μF

C. 0.15 μF

Summary

The few simple principles reviewed in this chapter are those you need to begin the study of electronics. Following is a summary of these principles:

- The basic electrical circuit consists of a source (voltage), a load (resistance), and a path (conductor or wire).
- The voltage represents a charge difference.
- If the circuit is a complete circuit, then electrons will flow in what is called current flow. The resistance offers opposition to current flow.
- The relationship between V, I, and R is given by Ohm's law:

  $$V = I \times R$$

- Resistance could be a combination of resistors in series, in which case you add the values of the individual resistors together to get the total resistance.

  $$R_T = R_1 + R_2 + \cdots + R_N$$

- Resistance could be a combination of resistors in parallel, in which case you find the total by using the following formula:

  $$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots + \frac{1}{R_N} \quad \text{or} \quad R_T = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots + \frac{1}{R_N}}$$

- The power delivered by a source is found by using the following formula:

  $$P = VI$$

- The power dissipated by a resistance is found by using the following formula:

  $$P = I^2R \quad \text{or} \quad P = \frac{V^2}{R}$$

- If you know the total applied voltage, V_S, the voltage across one resistor in a series string of resistors is found by using the following voltage divider formula:

  $$V_1 = \frac{V_S R_1}{R_T}$$

- The current through one resistor in a two resistor parallel circuit with the total current known is found by using the current divider formula:

  $$I_1 = \frac{I_T R_2}{(R_1 + R_2)}$$

- Kirchhoff's voltage law (KVL) relates the voltage drops in a series circuit to the total applied voltage.

  $V_S = V_1 + V_2 + \cdots + V_N$

- Kirchhoff's current law (KCL) relates the currents at a junction in a circuit by saying that the sum of the input currents equals the sum of the output currents. For a simple parallel circuit, this becomes the following where I_T is the input current:

  $I_T = I_1 + I_2 + \cdots + I_N$

- A switch in a circuit is the control device that directs the flow of current or, in many cases, allows that current to flow.
- Capacitors are used to store electric charge in a circuit. They also allow current or voltage to change at a controlled pace. The circuit time constant is found by using the following formula:

  $\tau = R \times C$

- At one time constant in an RC circuit, the values for current and voltage have reached 63 percent of their final values. At five time constants, they have reached their final values.
- Capacitors in parallel are added to find the total capacitance.

  $C_T = C_1 + C_2 + \cdots + C_N$

- Capacitors in series are treated the same as resistors in parallel to find a total capacitance.

$$\frac{1}{C_T} = \frac{1}{C_1} + \frac{1}{C_2} + \cdots + \frac{1}{C_N} \quad \text{or} \quad C_T = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots + \frac{1}{C_N}}$$

DC Pre-Test

The following problems and questions will test your understanding of the basic principles presented in this chapter. You will need a separate sheet of paper for your calculations. Compare your answers with the answers provided following the test. You will find that many of the problems can be worked more than one way.

Questions 1-5 use the circuit shown in Figure 1-20. Find the unknown values indicated using the values given.

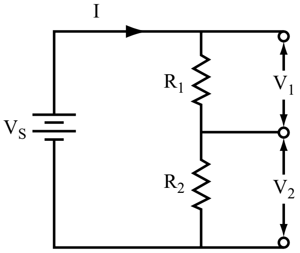

Figure 1-20

1. $R_1 = 12$ ohms, $R_2 = 36$ ohms, $V_S = 24$ V

   $R_T =$ ______________ , $I =$ ______________

2. $R_1 = 1$ KΩ, $R_2 = 3$ KΩ, $I = 5$ mA

   $V_1 =$ ______________ , $V_2 =$ ______________ , $V_S =$ ______________

3. $R_1 = 12$ kΩ, $R_2 = 8$ kΩ, $V_S = 24$ V

   $V_1 =$ ______________ , $V_2 =$ ______________

4. $V_S = 36$ V, $I = 250$ mA, $V_1 = 6$ V

   $R_2 =$ ______________

5. Now, go back to problem 1. Find the power dissipated by each resistor and the total power delivered by the source.

   $P_1 =$ ______________ , $P_2 =$ ______________ , $P_T =$ ______________

Questions 6-8 will use the circuit shown in Figure 1-21. Again, find the unknowns using the given values.

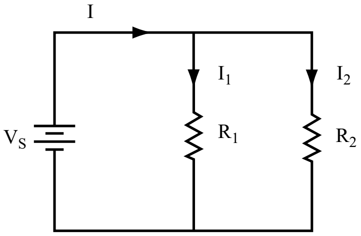

Figure 1-21

6. $R_1 = 6$ kΩ, $R_2 = 12$ kΩ, $V_S = 20$ V

   $R_T =$ ______________ , $I =$ ______________

7. $I = 2$ A, $R_1 = 10$ ohms, $R_2 = 30$ ohms

   $I_1 =$ ______________ , $I_2 =$ ______________

8. $V_S = 12$ V, I = 300 mA, $R_1 = 50$ ohms

$R_2 =$ ____________ , $P_1 =$ ____________

9. What is the maximum current that a 220 ohm resistor can safely have if its power rating is 1/4 watts?

$I_{MAX} =$ ____________

10. In a series RC circuit the resistance is 1 kΩ, the applied voltage is 3 V, and the time constant should be 60 μsec.

A. What is the required value of C?

C = ____________

B. What will be the voltage across the capacitor 60 μsec after the switch is closed?

$V_C =$ ____________

C. At what time will the capacitor be fully charged?

T = ____________

11. In the circuit shown in Figure 1-22, when the switch is at position 1, the time constant should be 4.8 ms.

A. What should be the value of resistor R_1?

$R_1 =$ ____________

B. What will be the voltage on the capacitor when it is fully charged, and how long will it take to reach this voltage?

$V_C =$ ____________ , T = ____________

C. After the capacitor is fully charged, the switch is thrown to position 2. What is the discharge time constant and how long will it take to completely discharge the capacitor?

$\tau =$ ____________ , T = ____________

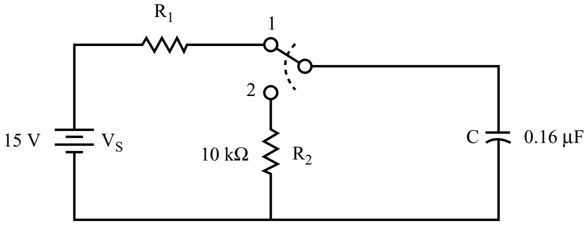

Figure 1-22

12. Three capacitors are available with the following values:

$C_1 = 8\ \mu F$; $C_2 = 4\ \mu F$; $C_3 = 12\ \mu F$.

A. What is C_T if all three are connected in parallel?

$C_T =$ ____________

B. What is C_T if they are connected in series?

$C_T =$ ____________

C. What is C_T if C_1 is in series with the parallel combination of C_2 and C_3?

$C_T =$ ____________

Answers to DC Pre-Test

If your answers do not agree with those provided here, review the problems indicated in parentheses before you go on to the next chapter. If you still feel uncertain about these concepts, go to a Web site such as `www.BuildingGadgets.com` and work through DC tutorials listed there.

It is assumed that Ohm's law is well known, so problem 4 will not be referenced.

1.		$R_T = 48$ ohms, $I = 0.5$ A	(problem 9)
2.		$V_1 = 5$ V, $V_2 = 15$ V, $V_S = 20$ V	(problems 23 and 26)
3.		$V_1 = 14.4$ V, $V_2 = 9.6$ V	(problem 23 and 263)
4.		$R_2 = 120$ ohms	(problems 9 and 23)
5.		$P_1 = 3$ W, $P_2 = 9$ W, $P_T = 12$ W	(problems 9 and 13)
6.		$R_T = 4$ kΩ, $I = 5$ mA	(problem 10)
7.		$I_1 = 1.5$ A, $I_2 = 0.5$ A	(problem 289)
8.		$R_2 = 200$ ohms, $P_1 = 2.88$ W	(problem 10 and 13)
9.		$I_{MAX} = 33.7$ mA	(problems 13, 15 and 16)
10.	A.	$C = 0.06\ \mu F$	(problems 34 and 35)
	B.	$V_C = 1.9$ V	
	C.	$T = 300\ \mu sec$	(34–38)

11.	A.	$R_1 = 30\ k\Omega$	(problems 35, 39 and 40)
	B.	$V_C = 15$ V, T = 24 ms	
	C.	$\tau = 1.6$ ms, T = 8.0 ms	(39–40)
12.	A.	24 μF	(problems 41 and 42)
	B.	2.18 μF	
	C.	5.33 μF	(42–43)

CHAPTER 2

The Diode

The main characteristic of a diode is that it conducts electricity in one direction only. Historically, the first vacuum tube was a diode; it was also known as a *rectifier.* The modern diode is a semiconductor device. It is used in all applications where the older vacuum tube diode was used, but it has the advantages of being much smaller, easier to use, and less expensive.

A *semiconductor* is a crystalline material that, depending on the conditions, can act as a conductor (allowing the flow of electric current) or an insulator (preventing the flow of electric current). Techniques have been developed to customize the electrical properties of adjacent regions of semiconductor crystals, which allow the manufacture of very small diodes, as well as transistors and integrated circuits.

When you complete this chapter you will be able to do the following:

- Specify the uses of diodes in DC circuits
- Determine from a circuit diagram whether a diode is forward or reverse biased
- Recognize the characteristic V-I curve for a diode
- Specify the knee voltage for a silicon or a germanium diode
- Calculate current and power dissipation in a diode
- Define diode breakdown
- Differentiate between zeners and other diodes
- Determine when a diode can be considered ‘‘perfect’’

Understanding Diodes

1 Silicon and germanium are semiconductor materials used in the manufacture of diodes, transistors, and integrated circuits. Semiconductor material is refined to an extreme level of purity, and then minute, controlled amounts of a specific impurity are added (a process called *doping*). Based on which impurity is added to a region of a semiconductor crystal, that region is said to be *N type* or *P type.* In addition to electrons (which are negative charge carriers used to conduct charge in a conventional conductor), semiconductors contain positive charge carriers called *holes.* The impurities added to an N type region increases the number of electrons capable of conducting charge, while the impurities added to a P type region increase the number of holes that are capable of conducting charge.

When a semiconductor chip contains an N doped region adjacent to a P doped region, a *diode junction* (often called a *PN junction*) is formed. Diode junctions can also be made with either silicon or germanium. However, silicon and germanium are never mixed when making PN junctions.

Question

Which diagrams in Figure 2-1 show diode junctions?

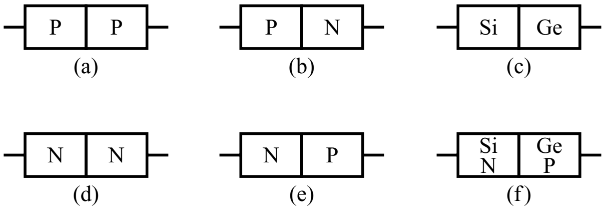

Figure 2-1

Answer

Diagrams (b) and (e) only

2 In a diode, the P material is called the *anode.* The N material is called the *cathode.*

Question

Identify which part of the diode shown in Figure 2-2 is P material and which part is N material.

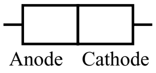

Figure 2-2

Answer

The anode is P material, the cathode is N material.

3 Diodes are useful because electric current will flow through a PN junction *in one direction only*. Figure 2-3 shows the direction in which the current flows.

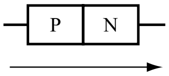

Figure 2-3

The circuit symbol for a diode is shown in Figure 2-4. The arrowhead points in the direction of current flow. While the anode and cathode are indicated here, they are not usually indicated in circuit diagrams.

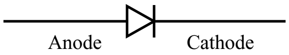

Figure 2-4

Question

In a diode, does current flow from anode to cathode, or cathode to anode?

Answer

Current flows from anode to cathode.

4 In the circuit shown in Figure 2-5, an arrow shows the direction of current flow.

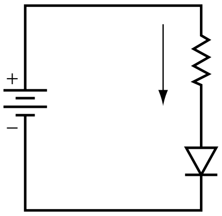

Figure 2-5

Questions

A. Is the diode connected correctly to permit current to flow?

____

B. Notice the way the battery and the diode are connected. Is the anode at a higher or lower voltage than the cathode?

____

Answers

A. Yes.

B. The anode is connected to the positive battery terminal, and the cathode is connected to the negative battery terminal. Therefore, the anode is at a higher voltage than the cathode.

5 When the diode is connected so that the current is flowing, it is said to be *forward biased*. In a forward biased diode, the anode is connected to a higher voltage than the cathode, and current is flowing. Examine the way the diode is connected to the battery in Figure 2-6.

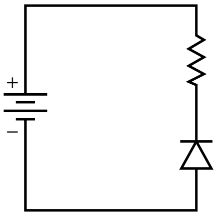

Figure 2-6

Question

Is the diode forward biased or not? Give the reasons for your answer.

____

Answer

No, it is not forward biased. The cathode is connected to the positive battery terminal and the anode is connected to the negative battery terminal. Therefore, the cathode is at a higher voltage than the anode.

6 When the cathode is connected to a higher voltage level than the anode, the diode cannot conduct. In this case, the diode is said to be *reverse biased*.

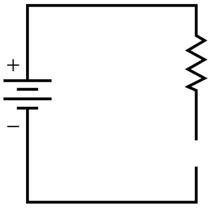

Figure 2-7

Question

Draw a reverse biased diode in the circuit shown in Figure 2-7.

Answer

Your drawing should look something like Figure 2-8.

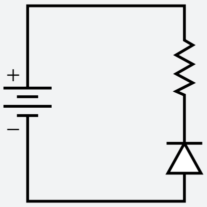

Figure 2-8

7 In many circuits, the diode is often considered to be a *perfect* diode to simplify calculations. A perfect diode has zero voltage drop in the forward direction, and conducts no current in the reverse direction.

Question

From your knowledge of basic electricity, what other component has zero voltage drop across its terminals in one condition, and conducts no current in an alternative condition?

Answer

The mechanical switch. When closed, it has no voltage drop across its terminals, and when open, it conducts no current.

8 A forward biased perfect diode can thus be compared to a closed switch. It has no voltage drop across its terminals and current flows through it.

A reverse biased perfect diode can be compared to an open switch. No current flows through it and the voltage difference between its terminals equals the supply voltage.

Question

Which of the switches shown in Figure 2-9 performs like a forward biased perfect diode?

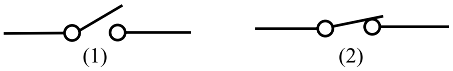

Figure 2-9

Answer

Switch (2) represents a closed switch and, like a forward biased perfect diode, allows current to flow through it. There is no voltage drop across its terminals.

The Diode Experiment

9 If you have access to electronic equipment, you may wish to perform the simple experiment described in the next few problems. If this is the first time you have tried such an experiment, get help from an instructor or someone who is familiar with electronic experiments.

If you do not have access to equipment, do not skip this exercise. Read through the experiment and try to picture or imagine the results. This is sometimes called ''dry-labbing'' the experiment. You can learn a lot from this exercise, even though it is always better to actually perform the experiment. This advice also applies to the other experiments that are given in many of the following chapters.

The objective of this experiment is to plot the V-I curve (also called a *characteristic curve*) of a diode, which shows how current flow through the diode varies with the applied voltage. As shown in Figure 2-10, the I-V curve for a diode demonstrates that if very little voltage is applied to a diode, current will not flow. However, once the applied voltage exceeds a certain value, the current flow increases quickly.

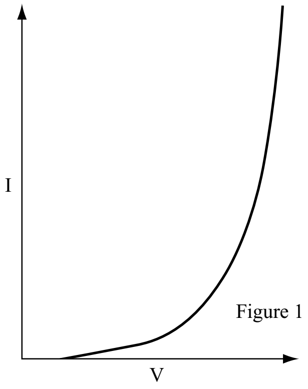

Figure 2-10

In performing this experiment, you will gain experience in the following:

- Setting up a simple electronic experiment
- Measuring voltage and current
- Plotting a graph of these

Set up the circuit shown on Figure 2-11. The circled A and V designate meters. The ammeter will measure current, and the voltmeter will measure voltage in the circuit.

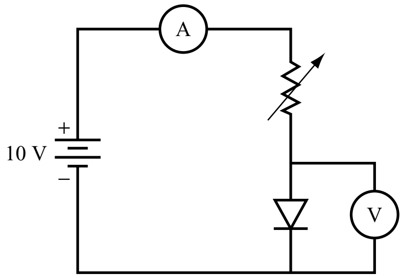

Figure 2-11

You can use two multimeters set to either measure current or voltage. Because you need to change the resistance at each step in the experiment, try using a 1 megohm potentiometer and measure its resistance value after each adjustment. If you don't have a 1 megohm potentiometer, but do have a collection of resistors with values ranging from a few hundred ohms to 1 megohm, you can simply change resistors at each step. Carefully check your circuit against the diagram, especially the direction of the battery and the diode.

Once you have checked your circuit, follow these steps:

1. Set R to its highest value and record it in the following table.

R	V	I
ohms	**volts**	**mA**

2. Close the switch and measure I and V. Record them in the table.
3. Lower the value of R a little to get a different reading of I.
4. Measure and record I and V again.
5. Continue in this fashion, taking as many readings as possible. There will suddenly come a point when V will not increase, but I will increase very rapidly. STOP.

NOTE **If V gets very large — above 3 or 4 volts — and I remains very small, then the diode is backward. Reverse it and start again.**

6. Graph the points recorded in the table, using the blank graph in Figure 2-12. Your curve should look like the one in Figure 2-10.

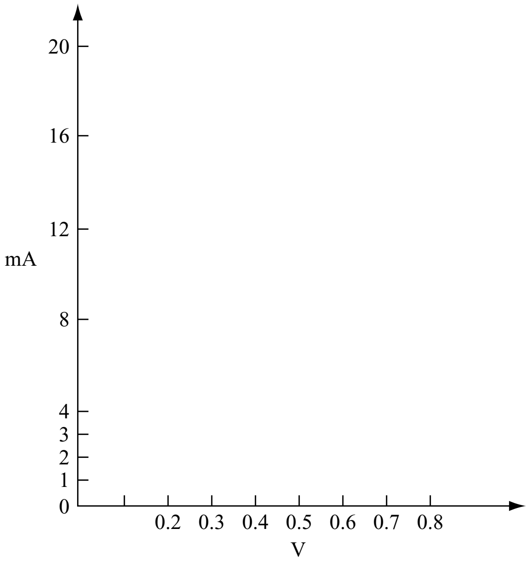

Figure 2-12

10 The measurements in the following table were taken using a commercial 1N4001 diode.

R	V	I
ohms	**volts**	**mA**
1 MΩ	0.30	0.02
220 kΩ	0.40	0.05
68 kΩ	0.46	0.14
33 kΩ	0.50	0.26
15 kΩ	0.52	0.50
10 kΩ	0.55	0.80
6.8 kΩ	0.56	1.20
4.7 kΩ	0.60	2.00
3.3 kΩ	0.62	2.80
2.2 kΩ	0.64	4.20
1.5 kΩ	0.65	5.50

R	V	I
ohms	**volts**	**mA**
1.0 kΩ	0.67	8.40
680 Ω	0.70	12.00
470 Ω	0.70	18.00
330 Ω	0.71	23.00

Further reductions in the value of R will cause very little increase in the voltage, but will produce large increases in the current.

Figure 2-13 is the V-I curve generated using the measurements shown in the preceding table.

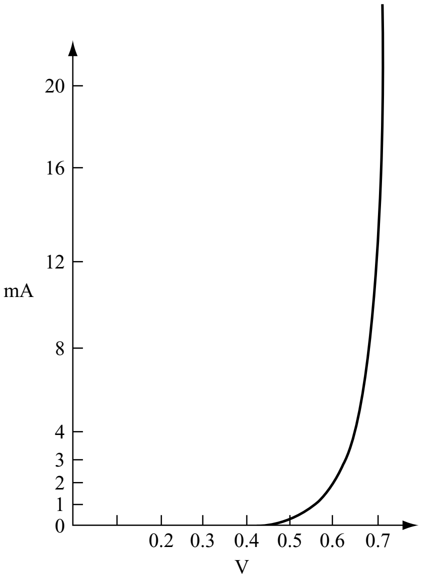

Figure 2-13

The V-I curve (or diode characteristic curve) is repeated in Figure 2-14 with three important regions marked on it.

The most important region is the *knee region*. This is not a sharply defined changeover point, but it occupies a very narrow range of the curve, where the diode resistance changes from high to low.

The ideal curve is shown for comparison.

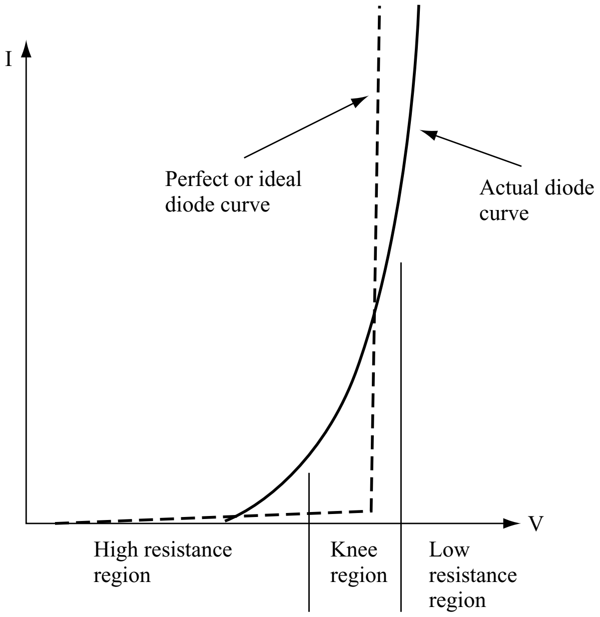

Figure 2-14

For the diode used in this problem, the knee voltage is about 0.7 V, which is typical for a silicon diode. This means (and your data should verify this) that at voltage levels below 0.7 V, the diode has such a high resistance that it limits the current flow to a very low value. This characteristic knee voltage is sometimes referred to as a *threshold voltage*. If you use a germanium diode, the knee voltage is about 0.3 V.

Question

What is the knee voltage for the diode you used?

____

Answer

If you use silicon, the knee voltage would be approximately 0.7 V; with germanium, it would be approximately 0.3 V.

11 The knee voltage is also a *limiting voltage*. That is, it is the highest voltage that can be obtained across the diode in the forward direction.

Questions

A. Which has the higher limiting voltage, germanium or silicon?

____

B. What happens to the diode resistance at the limiting or knee voltage?

____

Answers

A. Silicon, with a limiting voltage of 0.7 V, is higher than germanium, which has a limiting voltage of only 0.3 V.

B. It changes from high to low.

NOTE **You will be using these knee voltages in many later chapters as the voltage drop across the PN junction when it is forward biased.**

12 Refer back to the diagram of resistance regions in Figure 2-14.

Question

What happens to the current when the voltage becomes limited at the knee?

____

Answer

It increases rapidly.

13 For any given diode, the knee voltage will not be exactly 0.7 V or 0.3 V. Rather, it will vary slightly. But when using diodes in practice (that is, imperfect diodes), you can make two assumptions:

- The voltage drop across the diode is either 0.7 V or 0.3 V.
- Excessive current is prevented from flowing through the diode by using the appropriate resistor in series with the diode.

Questions

A. Why are imperfect diodes specified here?

____

B. Would you use a high or low resistance to prevent excessive current?

____

Answers

A. All diodes are imperfect, and the 0.3 or 0.7 voltage values are only approximate. In fact, in some later problems, it is assumed that the voltage drop across the diode, when it is conducting, is 0 V. This assumes, then, that as soon as you apply any voltage above 0, current flows in an ideal resistor (that is, the knee voltage on the V-I curve for an ideal diode is 0 volts).

B. Generally use a high resistance. However, the actual resistance value can be calculated given the applied voltage and the maximum current the diode can withstand.

14 Calculate the current through the diode in the circuit shown in Figure 2-15, using the steps in the following question.

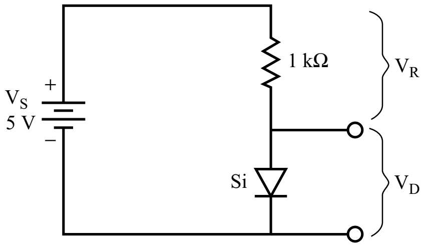

Figure 2-15

Questions

A. The voltage drop across the diode is known. It is 0.7 V for silicon and 0.3 V for germanium. (''Si'' near the diode means it is silicon.) Write down the diode voltage drop.

V_D = ____________________

B. Find the voltage drop across the resistor. This can be calculated using $V_R = V_s - V_D$. This is taken from KVL, which was discussed in Chapter 1.

$V_R =$ ______

C. Calculate the current through the resistor. Use $I = V_R/R$.

$I =$ ______

D. Finally, determine the current through the diode.

$I =$ ______

Answers

You should have written these values.

A. 0.7 V

B. $V_R = V_s - V_D = 5\text{ V} - 0.7\text{ V} = 4.3\text{ V}$

C. $I = \dfrac{V_R}{R} = \dfrac{4.3\,\text{V}}{1\,\text{k}\Omega} = 4.3\,\text{mA}$

D. 4.3 mA (In a series circuit the same current runs through each component.)

15 In practice, when the battery voltage is 10 V or above, the voltage drop across the diode is often considered to be 0 V instead of 0.7 V.

The assumption here is that the diode is a perfect diode, and the knee voltage is at 0 V rather than at a threshold value that must be exceeded. As discussed later, this assumption is often used in many electronic designs.

Questions

A. Calculate the current through the silicon diode shown in Figure 2-16.

$V_D =$ ______

$V = V_s - V_D =$ ______

$I = \dfrac{V_R}{R} =$ ______

$I_D =$ ______

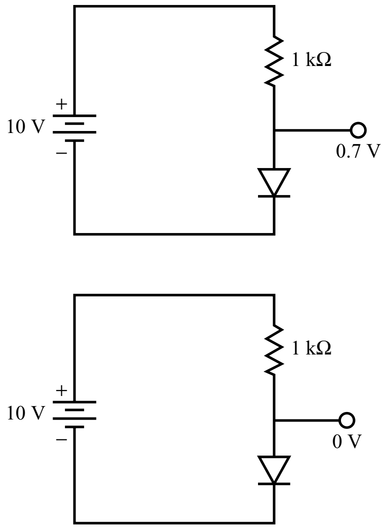

Figure 2-16

B. Calculate the current through the perfect diode shown in Figure 2-16.

$V_D =$ ______________________________

$V_R = V_s - V_D =$ ____________________

$I = \frac{V_R}{R} =$ ___________________________

$I_D =$ _______________________________

Answers

A. 0.7 V; 9.3 V; 9.3 mA; 9.3 mA

B. 0 V; 10 V; 10 mA; 10 mA

16 The difference in the values of the two currents found in problem 15 is less than 10 percent of the total current. That is, 0.7 mA is less than 10 percent of 10 mA. Many electronic components have a plus or minus 5 percent tolerance in their nominal values. This means that a 1 k resistor can be anywhere from 950 ohms to 1,050 ohms, meaning that the value of current through a resistor can vary plus or minus 5 percent.

Because of slight variance in component values, calculations are often simplified if the simplification does not change values by more than 10 percent. Therefore, a diode is often assumed to be perfect when the supply voltage is 10 V or more.

Questions

A. Examine the circuit in Figure 2-17. Is it safe to assume that the diode is perfect?

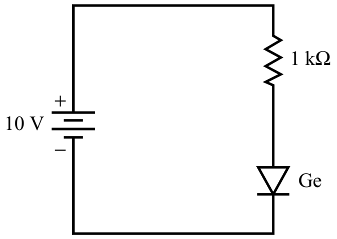

Figure 2-17

B. Calculate the current through the diode.

Answers

A. Yes, it can be considered a perfect diode.

B. I = 10 mA

17 When a current flows through a diode, it causes heating and power dissipation, just as with a resistor. The power formula for resistors is $P = V \times I$. This same formula can be applied to diodes to find the power dissipation. To calculate the power dissipation in a diode, you must first calculate the current as shown previously. The voltage drop in this formula is assumed to be 0.7 V for a silicon diode, even if you considered it to be 0 V when calculating the current.

For example, a silicon diode has 100 mA flowing through it. Determine how much power the diode dissipates.

$$P = (0.7\,\text{V})(100\,\text{mA}) = 70\,\text{mW}$$

Question

Assume a current of 2 amperes is flowing through a silicon diode. How much power is being dissipated?

Answer

$P = (0.7\,V)\,(2\,A) = 1.4\,\text{watts}$

18 Diodes are made to dissipate a certain amount of power, and this is quoted as a maximum power rating in the manufacturer's specifications of the diode.

Assume a silicon diode has a maximum power rating of 2 watts. How much current can it safely pass?

$$P = V \times I$$
$$I = \frac{P}{V}$$
$$= \frac{2\,\text{watts}}{0.7\,V}$$

2.86 A (rounded off to two decimal places)

Provided the current in the circuit does not exceed this, the diode will not overheat and burn out.

Question

Suppose the maximum power rating of a germanium diode is 3 watts. What is its highest safe current?

Answer

$$I = \frac{3\,\text{watts}}{0.3\,V} = 10\,A$$

19 Answer the following questions for another example.

Questions

A. Would a 3 watt silicon diode be able to carry the current calculated for the germanium diode for problem 18?

B. What would be its safe current?

Answers

A. No, 10 amperes would cause a power dissipation of 7 watts, which would burn up the diode.

B. $I = \frac{3}{0.7} = 4.3\,A$

Any current less than this would be safe.

20 The next several examples concentrate on finding the current through the diode. Look at the circuit shown in Figure 2-18.

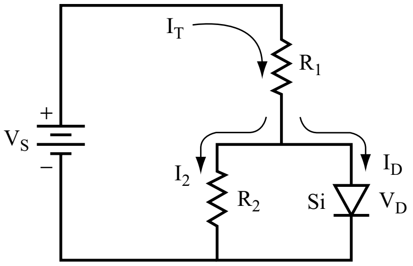

Figure 2-18

The total current from the battery flows through R_1, and then splits into I_2 and I_D. I_2 flows through R_2 and I_D through the diode.

Questions

A. What is the relationship between I_T, I_2, and I_D?

B. What is the value of V_D?

____

Answers

A. Remember KCL, $I_T = I_2 + I_D$

B. $V_D = 0.7$ V

21 To find I_D, it is necessary to go through the following steps because there is no way of finding I_D directly.

1. Find I_2. This is done using $V_D = R_2 \times I_2$.
2. Find V_R. For this use $V_R = V_s - V_D$ (KVL again).
3. Find I_T, the current through R_1. Use $V_R = I_T \times R_1$.
4. Find I_D. This is found by using $I_T = I_2 + I_D$ (KCL again).

To find I_D in the circuit shown in Figure 2-19, go through these steps, and then check your answers.

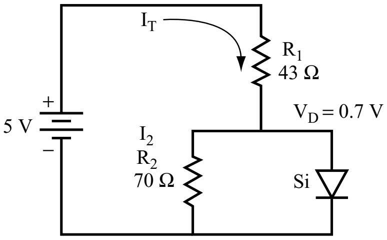

Figure 2-19

Questions

A. $I_2 =$

____

B. $V_R =$

____

C. $I_T =$

____

D. I_D =

____

Answers

A. $I_2 = \dfrac{V_D}{R_2} = \dfrac{0.7\,V}{R_2} = \dfrac{0.7\,V}{70\,\text{ohms}} = 0.01\,A = 10\,mA$

B. $V_R = V_S - V_D = 5\ V - 0.7\ V = 4.3\ V$

C. $I_T = \dfrac{V_R}{R_1} = \dfrac{4.3\,V}{43\,\text{ohms}} = 0.1\,A = 100\,mA$

D. $I_D = I_T - I_2 = 100\ mA - 10\ mA = 90\ mA$

22 For this problem, refer to your answers in problem 21.

Question

What is the power dissipation of the diode in problem 21?

____

Answer

$P = V_D \times I_D = (0.7\ V)(90\ mA) = 63$ milliwatts

23 To find the current in the diode for the circuit shown in Figure 2-20, answer these questions in order:

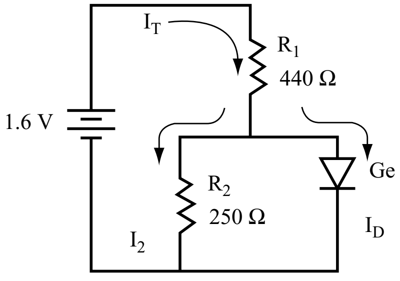

Figure 2-20

Questions

A. $I_2 =$ ____________________________

B. $V_R =$ ____________________________

C. $I_T =$ ____________________________

D. $I_D =$ ____________________________

Answers

A. $I_2 = \frac{0.3}{250} = 1.2\,\text{mA}$

B. $V_R = V_S - V_D = 1.6 - 0.3 = 1.3\ \text{V}$

C. $I_T = \frac{V_R}{R_1} = \frac{1.3}{440} = 3\,\text{mA}$

D. $I_D = I_T - I_2 = 1.8\ \text{mA}$

If you are going to take a break soon, this is a good stopping point.

Diode Breakdown

24 Earlier, you read that if the experiment is not working correctly, then the diode is probably in backward. If you place the diode in the circuit backward — as shown on the right in Figure 2-21 — then almost no current flows. In fact, the current flow is so small, it can be said that no current flows. The V-I curve for a reversed diode looks like the one shown on the left in Figure 2-21.

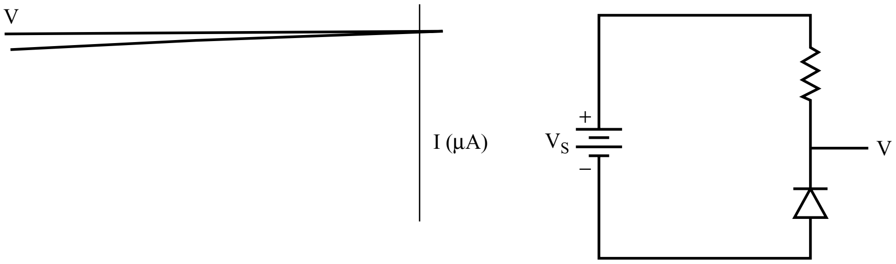

Figure 2-21

The V-I curve for a perfect diode would show zero current for all voltage values. But for a real diode, a voltage is reached where the diode ''breaks down'' and the diode allows a large current to flow. The V-I curve for the diode breakdown would then look like the one in Figure 2-22.

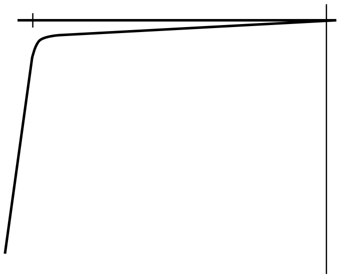

Figure 2-22

If this condition continues, the diode will burn out. You can avoid burning out the diode, even though it is at the breakdown voltage, by limiting the current with a resistor.

Question

The diode in the circuit shown in Figure 2-23 is known to break down at 100 V, and it can safely pass 1 ampere without overheating. Find the resistance in this circuit that would limit the current to 1 ampere.

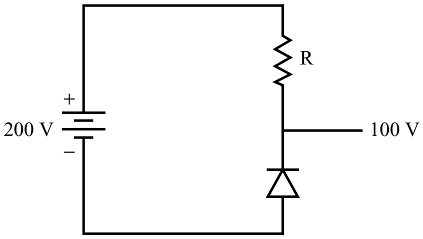

Figure 2-23

Answer

$V_R = V_s - V_D = 200\,\text{V} - 100\,\text{V} = 100\,\text{V}$

Since 1 ampere of current is flowing, then:

$$R = \frac{V_R}{I} = \frac{100\,\text{V}}{1\,\text{A}} = 100\,\text{ohms}$$

25 All diodes will break down when connected in the reverse direction if excess voltage is applied to them. The breakdown voltage, which is a function of how the diode is made, varies from one type of diode to another. This voltage is quoted in the manufacturer's datasheet.

Breakdown is not a catastrophic process and does not destroy the diode. If the excessive supply voltage is removed, the diode will recover and operate normally. You can use it safely many more times, provided the current is limited to prevent the diode from burning out.

A diode will always break down at the same voltage, no matter how many times it is used.

The breakdown voltage is often called the *peak inverse voltage* (PIV) or the *peak reverse voltage* (PRV). Following are the PIVs of some common diodes:

DIODE	PIV
1N4001	50 V
1N4002	100 V
1N4003	200 V
1N4004	400 V
1N4005	600 V
1N4006	800 V

Questions

A. Which can permanently destroy a diode, excessive current or excessive voltage?

____

B. Which is more harmful to a diode, breakdown or burnout?

____

Answers

A. Excessive current. Excessive voltage will not harm the diode if the current is limited.

B. Burnout. Breakdown is not necessarily harmful, especially if the current is limited.

The Zener Diode

26 Diodes can be manufactured so that breakdown occurs at lower and more precise voltages than those just discussed. These types of diodes are called *zener diodes*, so named because they exhibit the ‘‘Zener effect’’ — a particular form of voltage breakdown. At the zener voltage, a small current will flow through the zener diode. This current must be maintained to keep the diode at the zener point. In most cases, a few milliamperes are all that is required. The zener diode symbol and a simple circuit are shown in Figure 2-24.

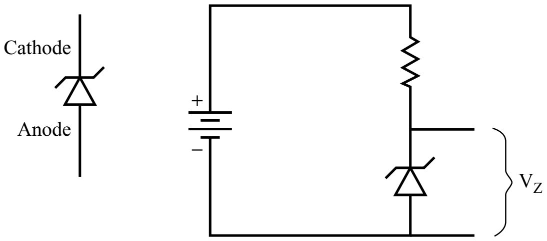

Figure 2-24

In this circuit, the battery determines the applied voltage. The zener diode determines the voltage drop (labeled V_z) across it. The resistor determines the current flow. Zeners are used to maintain a constant voltage at some point in a circuit.

Question

Why are zeners used for this purpose, rather than ordinary diodes?

____

Answer

Because zeners have a precise breakdown voltage

27 Examine an application in which a constant voltage is desirable — for example, a lamp driven by a DC generator. In this example, when the generator is turning at full speed, it puts out 50 V. When it is running more slowly, the voltage can drop to 35 V. You want to illuminate a 20 V lamp with this generator. Assume the lamp draws 1.5 A. The circuit is shown in Figure 2-25.

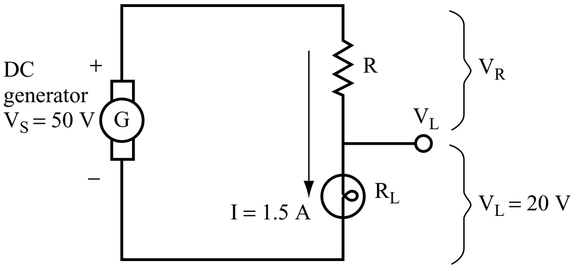

Figure 2-25

You need to determine a suitable value for the resistance. Follow these steps to find a suitable resistance value:

1. Find R_L, the lamp resistance. Use
$$R_L = \frac{V_L}{I}$$
2. Find V_R. Use $V_S = V_R + V_L$.
3. Find R. Use
$$R = \frac{V_R}{I}$$

Questions

Work through these steps, and write your answers here.

A. $R_L =$ ____

B. $V_R =$ ____

C. $R =$ ____

Answers

A. $R_L = \frac{20\text{ V}}{1.5\text{ A}} = 13.33$ ohms

B. $V_R = 50\text{ V} - 20\text{ V} = 30\text{ V}$

C. $R = \frac{50\,V - 20\,V}{1.5\,A} = \frac{30\,V}{1.5\,A} = 20\,\text{ohms}$

28 Assume now that the 20 ohm resistor calculated in problem 27 is in place, and the voltage output of the generator drops to 35 V, as shown in Figure 2-26. This is similar to what happens when a battery gets old. Its voltage level decays and it will no longer have sufficient voltage to produce the proper current. This results in the lamp glowing less brightly, or perhaps not at all. Note, however, that the resistance of the lamp stays the same.

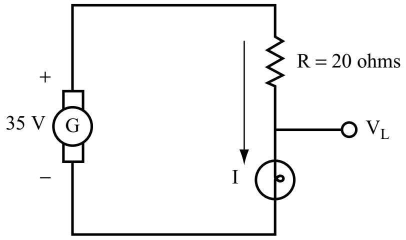

Figure 2-26

Questions

A. Find the total current flowing. Use

$$I_T = \frac{V_S}{(R + R_L)}$$

I_T =

B. Find the voltage drop across the lamp. Use $V_L = I_T \times R_L$

V_L =

C. Have the voltage and current increased or decreased?

Answers

A. $I_T = \frac{35\,V}{20\,\Omega + 13.3\,\Omega} = \frac{35\,V}{33.3\,\Omega} = 1.05\,A$

B. $V_L = 1.05\ A \times 13.3\ \Omega = 14\ V$

C. Both have reduced in value.

29 In many applications, a lowering of voltage across the lamp (or some other component) may be unacceptable. This can be prevented by using a zener diode, as shown in the circuit in Figure 2-27.

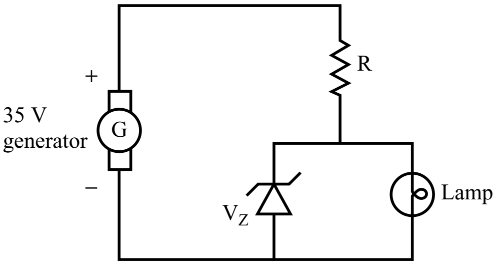

Figure 2-27

If you choose a 20 V zener (that is, one that has a 20 V drop across it), then the lamp will always have 20 V across it, no matter what the output voltage is from the generator (provided, of course, that the output from the generator is always above 20 V).

Questions

Say that the voltage across the lamp is constant, and the generator output drops.

A. What happens to the current through the lamp?

____

B. What happens to the current through the zener?

____

Answers

A. The current stays constant because the voltage across the lamp stays constant.

B. The current decreases because the total current decreases.

30 To make this circuit work and keep 20 V across the lamp at all times, you must find a suitable value of R. This value should allow sufficient total current to flow to provide 1.5 amperes required by the lamp, and the small amount required to keep the diode at its zener voltage. To do this, you start at the ''worst case'' condition. (''Worst case'' design is a common practice in electronics. It is used to ensure that equipment will work under the most adverse conditions.) The worst case here would occur when the generator is putting out only 35 V. Figure 2-28 shows the paths that current would take in this circuit.

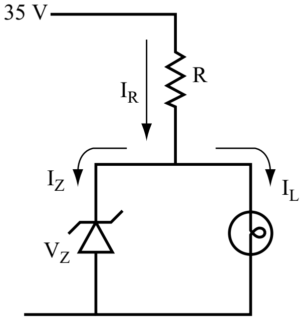

Figure 2-28

Find the value of R which will allows 1.5 A to flow through the lamp. How much current will flow through the zener diode? You can choose any current you like, provided it is above a few milliamperes, and provided it will not cause the zener diode to burn out. In this example assume that the zener current I_z is 0.5 A.

Questions

A. What is the total current through R?

I_R = ____________________

B. Calculate the value of R.

R = ____________________

Answers

A. $I_R = I_L + I_Z = 1.5\text{ A} + 0.5\text{ A} = 2\text{ A}$

B. $R = \dfrac{(V_S - V_Z)}{I_R} = \dfrac{(35\text{ volts} - 20\text{ volts})}{2\text{ A}} = 7.5\text{ ohms}$

A different choice of I_z here would produce another value of R.

31 Now take a look at what happens when the generator is supplying 50 V, as shown in Figure 2-29.

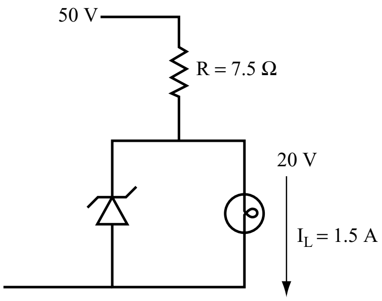

Figure 2-29

Because the lamp still has 20 V across it, it will still draw only 1.5 A. But the total current and the zener current will change.

Questions

A. Find the total current through R.

$I_R =$

B. Find the zener current.

$I_Z =$

Answers

A. $I_R = d\frac{(V_S - V_Z)}{R} = \frac{(50 - 20)}{7.5} = 4\text{ A}$

B. $I_Z = I_R - I_L = 4 - 1.5 = 2.5\text{ A}$

32 Although the lamp voltage and current remain the same, the total current and the zener current both changed.

Questions

A. What has happened to I_T (I_R)?

B. What has happened to I_Z?

Answers

A. I_T has increased by 2 A.

B. I_Z has increased by 2 A.

Note that the increase in I_T flows through the zener diode and not through the lamp.

---

33 The power dissipated by the zener diode changes as the generator voltage changes.

Questions

A. Find the power dissipated when the generator voltage is 35 V.

B. Now find the power when the generator is at 50 V.

Answers

A. $P_Z = V \times I = (20 \text{ volts})\ (0.5\text{ A}) = 10$ watts

B. $P_Z = V \times I = (20 \text{ volts})\ (2.5\text{ A}) = 50$ watts

If you use a zener diode with a power rating of 50 watts or more, it will not burn out.

34 Use Figure 2-30 to answer the following question.

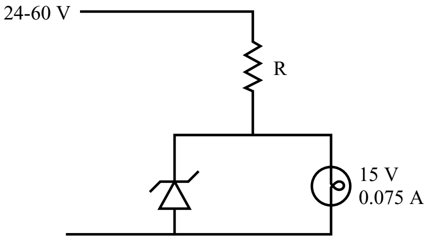

Figure 2-30

Question

For the circuit shown in Figure 2-30, what power rating should the zener diode have? The current and voltage ratings of the lamp are given.

Answer

At 24 volts, assuming a zener current of 0.5 A:

$$R = \frac{9}{0.575} = 15.7\,\text{ohms}$$

At 60 V:

$$I_R = \frac{45}{15.7} = 2.87\text{A; therefore } I_Z \approx 2.8\text{ A}$$

$$P_Z = (15\text{ volts})(2.8\text{ A}) = 42\text{ watts}$$

Summary

Semiconductor diodes are used extensively in modern electronic circuits. The main advantages of semiconductor diodes are:

- They are very small.
- They are rugged and reliable if properly used. You must remember that excessive reverse voltage or excessive forward current could damage or destroy the diode.
- Diodes are very easy to use, as there are only two connections to make.
- They are inexpensive.
- They can be used in all types of electronic circuits, from simple DC controls to radio and TV circuits.
- They can be made to handle a wide range of voltage and power requirements.
- Specialized diodes (which have not been covered here) can perform particular functions, which no other components can handle.
- Finally, as you will see in the next chapter, diodes are an integral part of transistors.

All of the many uses of semiconductor diodes are based on the fact they conduct in *one direction only*. Diodes are often used for the following:

- Protecting circuit components from voltage spikes
- Converting AC to DC

- Protecting sensitive components from high-voltage spikes
- Building high speed switches
- Rectifying radio frequency signals

Self-Test

The following questions test your understanding of this chapter. Use a separate sheet of paper for your diagrams or calculations. Compare your answers with the answers that follow the test.

1. Draw the circuit symbol for a diode, labeling each terminal.

2. What semiconductor materials are used in the manufacture of diodes?

3. Draw a circuit with a battery, resistor, and a forward biased diode.

4. What is the current through a reverse biased perfect diode?

5. Draw a typical V-I curve of a forward biased diode. Show the knee voltage.

6. What is the knee voltage for silicon?

   Germanium?

7. In the circuit shown in Figure 2-31, $V_S = 10$ V and $R = 100$ ohms. Find the current through the diode, assuming a perfect diode.

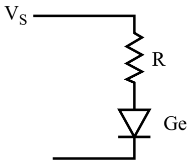

Figure 2-31

8. Calculate question 7 using these values: $V_S = 3$ V and R = 1 kΩ.

9. In the circuit shown in Figure 2-32, find the current through the diode.

$V_S = 10$ V
$R_1 = 10$ kΩ
$R_2 = 1$ kΩ

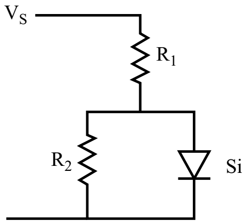

Figure 2-32

10. In the circuit shown in Figure 2-33, find the current through the zener diode.

$V_S = 20$ V
$V_Z = 10$ V
$R_1 = 1$ kΩ
$R_2 = 2$ kΩ

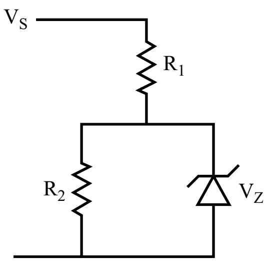

Figure 2-33

11. If the supply voltage for question 10 increases to 45 V, what is the current in the zener diode?

---

12. What is the maximum power dissipated for the diode in questions 10 and 11?

Answers to Self-Test

If your answers do not agree with those given below, review the problems indicated in parentheses before you go on to the next chapter.

1.	See Figure 2-34.	(problem 3)
	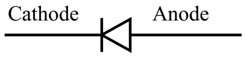 **Figure 2.34**	
2.	Germanium and silicon	(problem 1)
3.	See Figure 2-35.	(problem 4)
	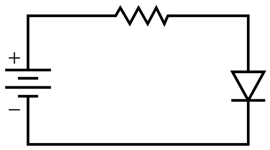 **Figure 2.35**	
4.	There is zero current flowing through the diode.	(problem 6)

5.	See Figure 2-36.	(problem 9 and 10)

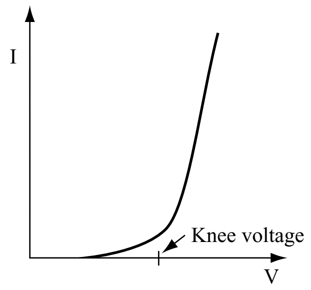

Figure 2.36

6.	Si = 0.7 V; Ge = 0.3 V (These are approximate.)	(problem 10)
7.	$I_D = 100$ mA	(problem 14)
8.	As $V_S = 3$ V, do not ignore the voltage drop across the diode. Thus, $I_D = 2.7$ mA	(problem 14)
9.	Ignore V_D in this case. Thus, $I_D = 0.3$ mA. If V_D is not ignored, $I_D = 0.23$ mA.	(problem 21)
10.	$I_Z = 5$ mA	(problem 31)
11.	$I_Z = 30$ mA	(problem 31)
12.	The maximum power will be dissipated when I_Z is at its peak value of 30 mA. Therefore, $P_Z(MAX) = 0.30$ W.	(problem 33)

CHAPTER 3

Introduction to the Transistor

The transistor is undoubtedly the most important modern electronic component because it has enabled great and profound changes in electronics and in our daily lives since its discovery in 1948.

This chapter introduces the transistor as an electronic component that acts similarly to a simple mechanical switch, and, in fact, it is actually used as a switch in many modern electronic devices. A transistor can be made to conduct or not conduct an electric current — exactly what a mechanical switch does.

Most transistors used in electronic circuits are *bipolar junction transistors* (BJT), commonly referred to as *bipolar transistors*, *junction field effect transistors* (JFET), or *metal oxide silicon field effect transistors* (MOSFET). This chapter (along with Chapters 4 and 8) illustrates how BJTs and JFETs function and how they are used in electronic circuits. Because JFETs and MOSFETs function in similar fashion, MOSFETs are not covered here.

An experiment in this chapter will help you to build a simple one-transistor switching circuit. You can easily set up this circuit on a home workbench. You should take the time to obtain the few components required and actually perform the experiment of building and operating the circuit.

In Chapter 4, you will continue to study switching designs and the operation of the transistor as a switch. In Chapter 8, you learn how a transistor can be made to operate as an amplifier. In this mode, the transistor produces an output that is a magnified version of an input signal, which is useful because many electronic signals require amplification. These chapters taken together present an easy introduction to how transistors function and how they are used in basic electronic circuits.

When you complete this chapter, you will be able to do the following:

- Describe the basic construction of a bipolar junction transistor (BJT).
- Describe the basic construction of a junction field effect transistor (JFET).
- Specify the relationship between base and collector current in a BJT.
- Specify the relationship between gate voltage and drain current in a JFET.
- Calculate the current gain for a BJT.
- Compare the transistor to a simple mechanical switch.

Understanding Transistors

1 The diagrams in Figure 3-1 show some common transistor cases (also called *packages*). The cases protect the semiconductor chip on which the transistor is built, and provide leads that can be used to connect it to other components. For each transistor, the diagrams show the lead designations and how to identify them based on the package design. Transistors can be soldered directly into a circuit, inserted into sockets, or inserted into breadboards. When soldering, you must take great care, because transistors can be destroyed if overheated. A heat sink clipped to the transistor leads between the solder joint and the transistor case can reduce the possibility of overheating. If you're using a socket, you can avoid exposing the transistor to heat by soldering the connections to the socket before inserting the transistor.

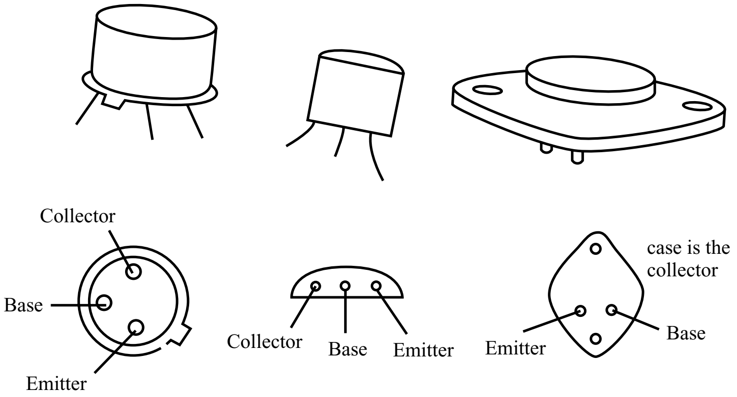

Figure 3-1

Questions

A. How many leads are there on most transistors?

____

B. Where there are only two leads, what takes the place of the third lead?

____

C. What are the three leads or connections called?

____

D. Why should you take care when soldering transistors into a circuit?

____

Answers

A. Three.

B. The case can be used instead, as indicated in the diagram on the right side of Figure 3-1. (This type of case is used for power transistors.)

C. Emitter, base, and collector.

D. Excessive heat can damage a transistor.

2 You can think of a bipolar junction transistor as functioning like two diodes, connected back-to-back, as illustrated in Figure 3-2.

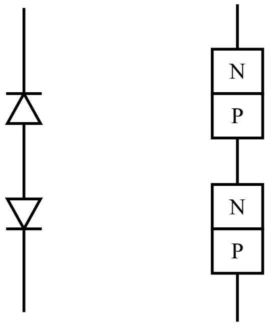

Figure 3-2

However, in the construction process, one very important modification is made. Instead of two separate P regions as shown in Figure 3-2, only one very thin region is used, as shown in Figure 3-3.

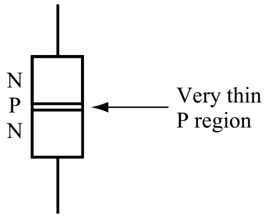

Figure 3-3

Question

Which has the thicker P region, the transistor shown in Figure 3-3 or two diodes connected back-to-back?

Answer

Two diodes. The transistor has a very thin P region.

3 Because two separate diodes wired back-to-back share two thick P regions, they will not behave like a transistor. The reason for this takes you into the realm of semiconductor physics, so it won't be covered in this book.

Question

Why don't two diodes connected back-to-back function like a transistor?

Answer

The transistor has one thin P region while the diodes share two thick P regions.

4 The three terminals of a transistor (the base, the emitter, and the collector) are connected as shown in Figure 3-4.

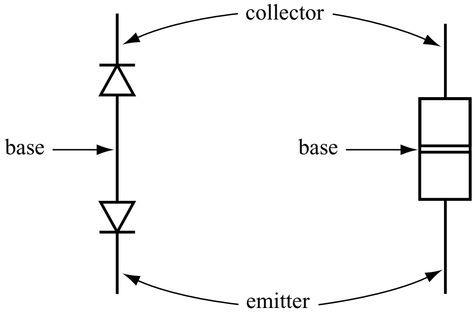

Figure 3-4

When talking about a transistor as two diodes, you refer to the diodes as the *base-emitter diode*, and the *base-collector diode*.

The symbol used in circuit diagrams for the transistor is shown in Figure 3-5, with the two diodes and the junctions shown for comparison. Because of the way the semiconductor materials are arranged, this is known as an NPN transistor.

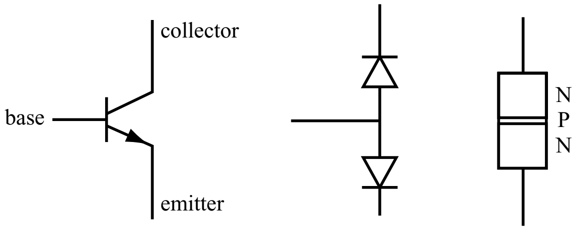

Figure 3-5

Question

Which transistor terminal includes an arrowhead?

Answer

The emitter

5 It is also possible to make transistors with a PNP configuration, as shown in Figure 3-6.

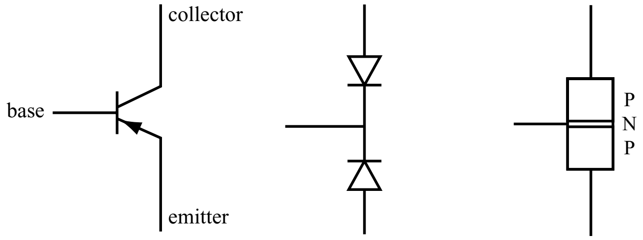

Figure 3-6

Both NPN and PNP type transistors can be made from either silicon or germanium.

Questions

A. Draw a circuit symbol for both an NPN and a PNP transistor. (Use a separate sheet of paper for your drawings.)

B. Which of the transistors represented by these symbols might be silicon?

____

C. Are silicon and germanium ever combined in a transistor?

____

Answers

A. See Figure 3-7.

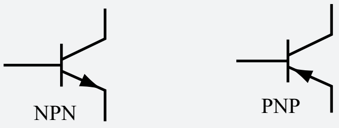

Figure 3-7

B. Either or both could be silicon. (Either or both could also be germanium.)

C. Silicon and germanium are *not* mixed in any commercially available transistors. However, researchers are attempting to develop ultrafast transistors that contain both silicon and germanium.

6 Take a look at the simple examples using NPN transistors in this and the next few problems.

If a battery is connected to an NPN transistor as shown in Figure 3-8, then a current will flow in the direction shown.

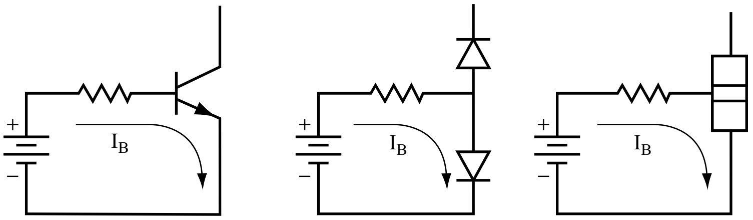

Figure 3-8

The current, flowing through the base-emitter diode, is called *base current* and is represented by the symbol I_B.

Question

Would base current flow if the battery were reversed? Give a reason for your answer.

____

Answer

Base current would not flow because the diode would be back-biased.

7 In the circuit in Figure 3-9, you can calculate the base current using the techniques covered in Chapter 2.

Question

Find the base current in the circuit shown in Figure 3-9. (*Hint:* Do not ignore the 0.7 V drop across the base-emitter diode.)

I_B = ____

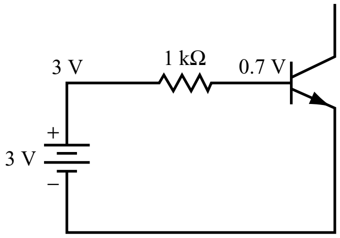

Figure 3-9

Answer

Your calculations should look something like this:

$$I_B = \frac{(V_S - 0.7\,V)}{R} = \frac{(3 - 0.7)}{1\,k\Omega} = \frac{2.3\,V}{1\,k\Omega} = 2.3\,mA$$

8 For the circuit shown in Figure 3-10, because the 10 volts supplied by the battery is much greater than the 0.7 V diode drop, you can consider the base-emitter diode to be a perfect diode, and thus assume the voltage drop is 0 V.

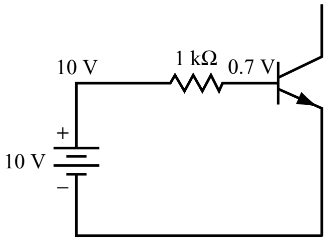

Figure 3-10

Question

Calculate the base current.

$I_B =$ ______________________________

Answer

$$I_B = \frac{(10 - 0)}{1\,k\Omega} = \frac{10}{1\,k\Omega} = 10\,mA$$

9 Look at the circuit shown in Figure 3-11.

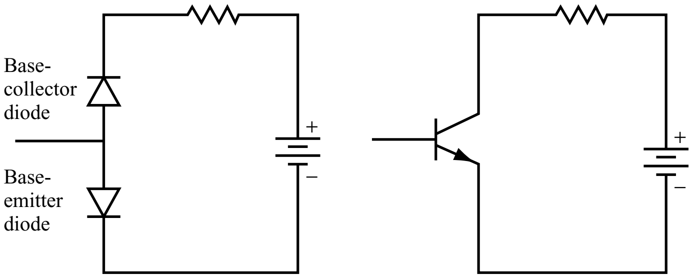

Figure 3-11

Question

Will current flow in this circuit? Why or why not?

Answer

It will not flow because the base-collector diode is reverse-biased.

10 Examine the circuit shown in Figure 3-12. Notice that batteries are connected to both the base and collector portions of the circuit.

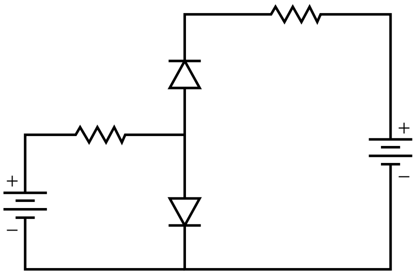

Figure 3-12

When you connect batteries to both the base and the collector portions of the circuit, currents flowing through the circuit demonstrate a key characteristic of the transistor. This characteristic is sometimes called *transistor action — if base current flows in a transistor, collector current will also flow*.

Examine the current paths shown in Figure 3-13.

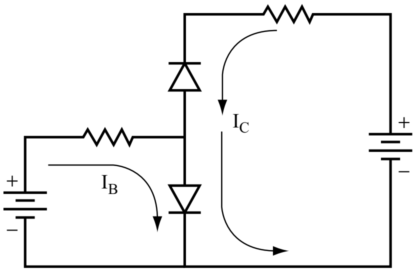

Figure 3-13

Questions

A. What current flows through the base-collector diode?

B. What current flows through the base-emitter diode?

C. Which of these currents causes the other to flow?

Answers

A. I_C (the collector current).

B. I_B and I_C. Note that both of them flow through the base-emitter diode.

C. Base current causes collector current to flow.

No current flows along the path shown by the dotted line in Figure 3-14 from the collector to the base.

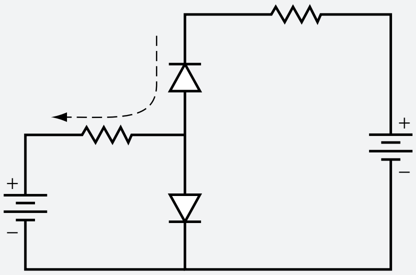

Figure 3-14

11 Up to now, you have studied the NPN bipolar transistor. PNP bipolar transistors work in the same way as NPN bipolar transistors, and can also be used in these circuits.

There is, however, one important circuit difference, which is illustrated in Figure 3-15. The PNP transistor is made with the diodes oriented in the reverse direction from the NPN transistor.

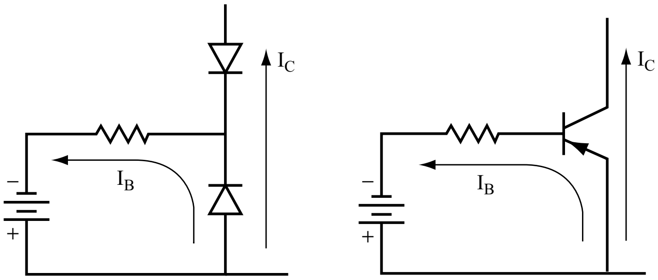

Figure 3-15

Questions

Compare Figure 3-15 with Figure 3-13. How are the circuits different relative to the following?

A. Battery connections:

B. Current flow:

Answers

A. The battery is reversed in polarity.

B. The currents flow in the opposite direction.

12 Figure 3-16 shows the battery connections necessary to produce both base current and collector current in a circuit that uses a PNP transistor.

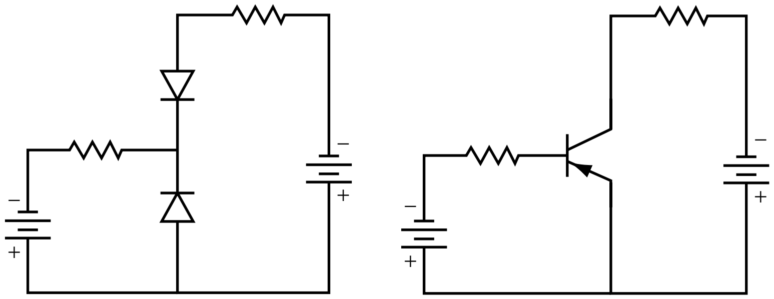

Figure 3-16

Question

In which direction do these currents circulate — clockwise or counterclockwise?

Answers

Base current flows counterclockwise.
Collector current flows clockwise.

As stated earlier, NPN and PNP bipolar transistors work in much the same way: Base current causes collector current to flow in both. The only significant difference in using a PNP versus an NPN bipolar transistor is that the polarity of the supply voltage (for both the base and collector sections of the circuit) is reversed. To avoid confusion, bipolar transistors used throughout the rest of this book are NPNs.

13 Consider the circuit in Figure 3-17. It uses only one battery to supply voltage to both the base and the collector portions of the circuit. The path of the base current is shown in the diagram.

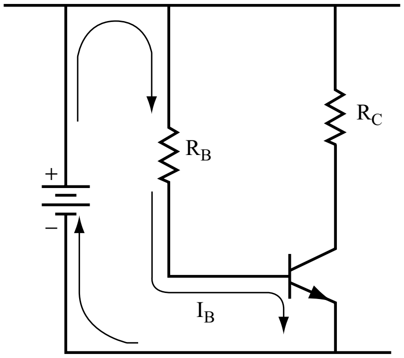

Figure 3-17

Questions

A. Name the components through which the base current flows.

B. Into which terminal of the transistor does the base current flow?

C. Out of which transistor terminal does the base current flow?

D. Through which terminals of the transistor does base current not flow?

Answers

A. The battery, the resistor R_B, and the transistor

B. Base

C. Emitter

D. Collector

14 Take a moment to recall the key physical characteristic of the transistor.

Question

When base current flows in the circuit shown in Figure 3-17, what other current will flow, and which components will it flow through?

Answer

Collector current will flow through the resistor R_C and the transistor.

15 In Figure 3-18 the arrows indicate the path of the collector current through the circuit.

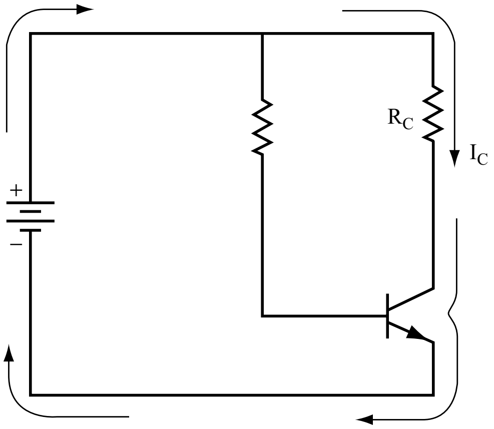

Figure 3-18

Questions

A. List the components through which the collector current flows.

B. What causes the collector current to flow?

Answers

A. The resistor R_C, the transistor, and the battery.

B. Base current. (Collector current doesn't flow unless base current is flowing.)

16 It is a property of the transistor that the *ratio of collector current to base current is constant*. The collector current is always much larger than the base current. The ratio of the two currents is called the *current gain* of the transistor and is represented by the symbol β, or beta. Typical values of β range from 10 to 300.

Questions

A. What is the ratio of collector current to base current called?

____

B. What is the symbol used to represent this?

____

C. Which is larger — base or collector current?

____

D. Look back at the circuit in problem 13. Will current be greater in R_B or in R_C?

____

Answers

A. Current gain

B. β

C. Collector current is larger.

D. The current is greater in R_C, because it is the collector current.

NOTE **The β introduced here is referred to in manufacturers' specification sheets as h_{FE}. Technically it is referred to as the static or DC β. For the purposes of this chapter, it is called β. Discussions on transistor parameters in general, which are well covered in many textbooks, will not be covered here.**

17 The mathematical formula for current gain is as follows:

$$\beta = \frac{I_C}{I_B}$$

where:

I_B = base current

I_C = collector current

The equation for β can be rearranged to give $I_C = \beta\, I_B$. From this, you can see that if no base current flows, no collector current flows. Also, if more base current flows, more collector current flows. This is why it's said that the ''base current controls the collector current.''

Question

Suppose the base current is 1 mA and the collector current is 150 mA. What is the current gain of the transistor?

Answer

150

18 Current gain is a physical property of transistors. You can find its value in the manufacturers' published data sheets, or you can determine it by experimenting.

In general, β is a different number from one transistor part number to the next, but transistors with the same part number have β values within a narrow range of each other.

One of the most frequently performed calculations in transistor work is to determine the values of either collector or base current, when β and the other current are known.

For example, suppose a transistor has 500 mA of collector current flowing and you know it has a β value of 100. Find the base current. To do this, use the following formula:

$$\beta = \frac{I_C}{I_B}$$

$$I_B = \frac{I_C}{\beta} = \frac{500\,\text{mA}}{100} = 5\,\text{mA}$$

Questions

Calculate the following values:

A. $I_C = 2$ A, $\beta = 20$. Find I_B.____________________

B. $I_B = 1$ mA, $\beta = 100$. Find I_C.____________________

C. $I_B = 10$ μA, $\beta = 250$. Find I_C.____________________

D. $I_B = 0.1$ mA, $I_C = 7.5$ mA. Find β.____________________

Answers

A. 0.1 A, or 100 mA

B. 100 mA

C. 2500 μA, or 2.5 mA

D. 75

19 This problem serves as a summary of the first part of this chapter. You should be able to answer all these questions. Use a separate sheet of paper for your drawing and calculations.

Questions

A. Draw a transistor circuit utilizing an NPN transistor, a base resistor, a collector resistor, and one battery to supply both base and collector currents. Show the paths of I_B and I_C.

B. Which current controls the other?

____

C. Which is the larger current, I_B or I_C?

____

D. $I_B = 6$ μA, $\beta = 250$. Find I_C.____________________

E. $I_c = 300$ μA, $\Omega = 50$. Find I_B.____________________

Answers

A. See Figure 3-17 and Figure 3-18.

B. I_B (base current) controls I_C (collector current).

C. I_C

D. 1.5 mA

E. 6 mA

The Transistor Experiment

20 The objective of the following experiment is to find β of a particular transistor by measuring several values of base current with the corresponding values of collector current. Next, divide the values of collector current by the values of the base current to determine β. The value of β will be almost the same for all the measured values of current. This demonstrates that β is a constant for a transistor.

While the circuit is set up, measure the collector voltage for each current value. This demonstrates (experimentally) some points that are covered in future problems. As you perform the experiment, observe how the collector voltage V_c drops toward 0 V as the collector current increases.

If you do not have the facilities for setting up the circuit and measuring the values, just read through the experiment. If you do have the facilities, you will need the following equipment and supplies.

- One 9 V transistor radio battery (or a lab power supply).
- One multimeter set to measure current on a scale that will measure at least 100 μA.
- One multimeter set to measure current on a scale that will measure at least 10 mA.
- One multimeter set to measure voltage, on a scale that will measure at least 10 V.
- One resistor substitution box, or a 1 MΩ potentiometer, or assorted resistors with the values listed in the following table. If you use a potentiometer, you'll need another multimeter to read its resistance after each adjustment.

R_B
1 MΩ
680 kΩ
470
330

R_B
270
220
200
180
160
150
120
110
100

- One 1 kΩ resistor.
- One transistor, preferably NPN.
- One breadboard.

Almost any small commercially available transistor will do for this experiment. The measurements given in this book were obtained using a 2 N3643. If only a PNP is available, then simply reverse the battery voltage and proceed as described.

Set up the circuit shown in Figure 3-19 on the breadboard. Using breadboards allows you to easily connect components in a circuit that you can then disassemble when testing is complete. You can then use the components and breadboard in the construction of your next circuit.

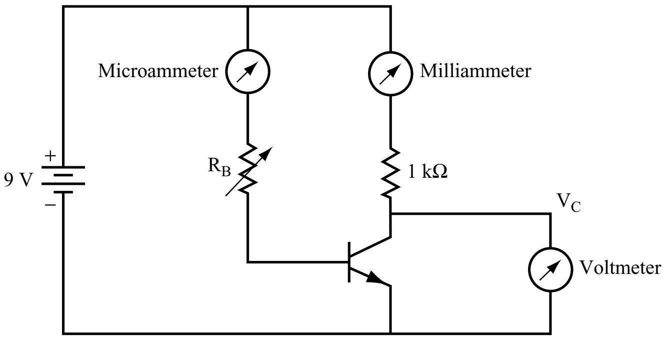

Figure 3-19

Follow these steps, recording your measurements in the following blank table.

R_B	I_B	I_C	V_C	β

1. Set R_B to its highest value.
2. Measure and record I_B
3. Measure and record I_C.
4. Measure and record V_C. This voltage is sometimes referred to as the *collector-emitter voltage* (V_{CE}), because it is taken across the collector-emitter leads if the emitter is connected to ground or the negative of the power supply.
5. Lower the value of R_B enough to produce a different reading of I_B.
6. Measure and record the new values for R_B, I_B, I_C, and V_C.
7. Lower R_B again and get a new I_B.
8. Measure and record the new values for R_B, I_B, I_C, and V_C again.
9. Repeat Steps 7 and 8 until $V_C = 0$ V.
10. Further reductions in the value of R_B will increase I_B, but will not affect the values of I_C or V_C.

Check the numbers in your table to make sure you got a consistent pattern. Then compare your measurements with the ones shown in the following table.

R_B	I_B	I_C	V_C	β
1 MΩ	9 μA	0.9 mA	8.1 volts	100
680 kΩ	13	1.3	7.7	100
470	19	1.9	7.1	100
330	27.3	2.8	6.2	103
270	33.3	3.3	5.7	99
220	40	4.1	5.0	103
200	45	4.5	4.5	100
180	50	5	4.0	100
160	56	5.6	3.4	100
150	60	6	3	100
120	75	7.5	1.5	100
110	82	8.0	1.0	98
100	90	9	0.3	100

The measurements shown here were obtained by conducting this experiment very carefully. Precision resistors and a commercial 2 N3643 transistor were used. With ordinary +/− 5 percent tolerance resistors and a transistor chosen at random, you could get different results. Don't worry if your results are not as precise as those listed here.

Calculation of β results in values at or very close to 100 for each measurement, which agrees with the manufacturer's specification of β = 100.

For each value of I_B and its corresponding value of I_C in your experiment, calculate the value of β ($\beta = I_C/I_B$). The values will vary slightly but will be close to an average. (Excessively low and high values of I_B may produce values of β, which will be quite different. Ignore these for now.) Did you get a consistent β? Was it close to the manufacturer's specifications for your transistor?

21 In the experiment, you measured the voltage level at the collector (V_C) and recorded your measurements. Now, examine how to determine the voltage at the collector, when it's not possible to measure the voltage level.

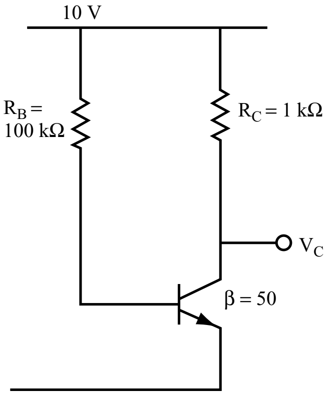

Figure 3-20

Use the values shown in the circuit in Figure 3-20 to complete these steps.

1. Determine I_C.
2. Determine the voltage drop across the collector resistor R_C. Call this V_R.
3. Subtract V_R from the supply voltage to calculate the collector voltage.

Here is the first step:

1. To find I_C, you must first find I_B.

$$I_B = \frac{10\,\text{V}}{100\,\text{k}\Omega} = 0.1\,\text{mA}$$

$$I_C = \beta \times I_B = 50 \times 0.1\,\text{mA} = 5\,\text{mA}$$

Now, perform the next two steps.

Questions

2. $V_R =$ ______________
3. $V_C =$ ______________

Answers

2. To find V_R:

   $V_R = R_C \times I_C = 1\ \text{k}\Omega \times 5\ \text{mA} = 5\ \text{V}$
3. To find V_C:

   $V_C = V_S - V_R = 10\ \text{V} - 5\ \text{V} = 5\ \text{V}$

22 Determine parameters for the circuit shown in Figure 3-20, using the value of $\beta = 75$.

Questions

Calculate the following:

A. $I_C =$ ____________________

B. $V_R =$ ____________________

C. $V_C =$ ____________________

Answers

A. $I_B = \frac{10\,V}{100\,k\Omega} = 0.1\,mA$

$I_C = 75 \times 0.1\,mA = 7.5\,mA$

B. $V_R = 1\,k\Omega \times 7.5\,mA = 7.5\,V$

C. $V_C = 10\,V - 7.5\,V = 2.5\,V$

23 Determine parameters for the same circuit, using the values of $R_B = 250\,k\Omega$ and $\beta = 75$.

Questions

Calculate the following:

A. $I_C =$ ____________________

B. $V_R =$ ____________________

C. $V_C =$ ____________________

Answers

A. $I_B = \frac{10\,V}{250\,k\Omega} = \frac{1}{25}\,mA$

$I_C = 75 \times \frac{1}{25}\,mA = 3\,mA$

B. $V_R = 1\,k\Omega \times 3\,mA = 3\,V$

C. $V_C = 10\,V - 3\,V = 7\,V$

24 From the preceding problems, you can see that you can set V_C to any value by choosing a transistor with an appropriate value of β, or by choosing the correct value of R_B.

Now, consider the example shown in Figure 3-21. The objective is to find V_C. Use the steps outlined in problem 21.

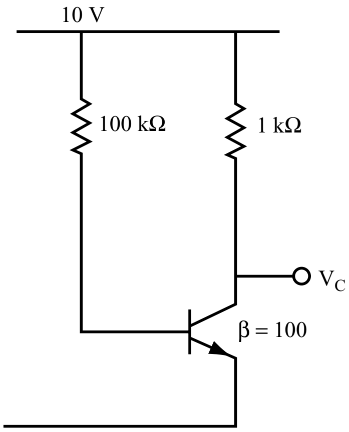

Figure 3-21

Questions

Calculate the following:

A. $I_B =$ ____________________

$I_C =$ ____________________

B. $V_R =$ ____________________

C. $V_C =$ ____________________

Answers

Your results should be as follows:

A. $I_B = \frac{10\,\text{V}}{100\,\text{k}\Omega} = 0.1\,\text{mA}$

$I_C = 100 \times 0.1\text{mA} = 10\,\text{mA}$

B. $V_R = 1\ \text{k}\Omega \times 10\ \text{mA} = 10\ \text{V}$

C. $V_C = 10\ \text{V} - 10\ \text{V} = 0\ \text{V}.$

Here, the base current is sufficient to produce a collector voltage of 0 volts and the maximum collector current possible, given the stated values of the collector resistor and supply voltage. This condition is called *saturation*.

25 Look at the two circuits shown in Figure 3-22 and compare their voltages at the point labeled V_C.

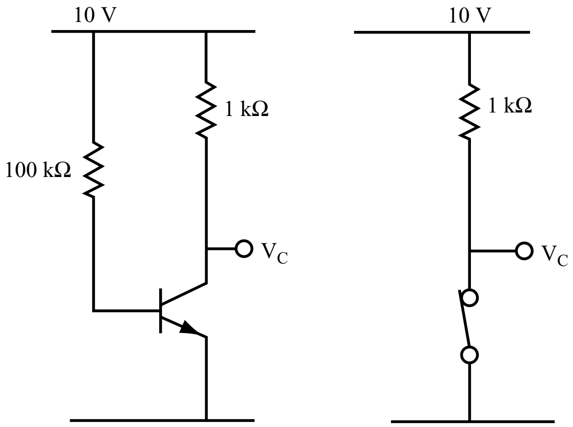

Figure 3-22

Consider a transistor that has sufficient base current and collector current to set its collector voltage to 0 V. Obviously, this can be compared to a closed mechanical switch. Just as the switch is said to be ON, then the transistor is also said to be ''turned on,'' or just ON.

Questions

A. What can you compare a turned on transistor to?

____

B. What is the collector voltage of an ON transistor?

____

Answers

A. A closed mechanical switch

B. 0 V

26 Now, compare the circuits shown in Figure 3-23.

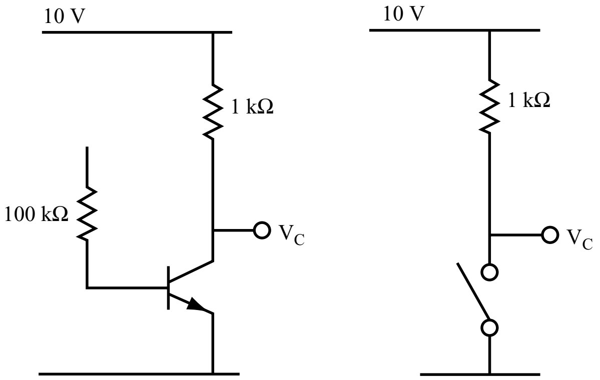

Figure 3-23

Because the base circuit is broken (that is, it is not complete), there is no base current flowing.

Questions

A. How much collector current is flowing?

____

B. What is the collector voltage?

____

C. What is the voltage at the point V_C in the mechanical switch circuit?

____

Answers

A. None.

B. Because there is no current flowing through the 1 kΩ resistor, there is no voltage drop across it, so the collector is at 10 V.

C. 10 V, because there is no current flowing through the 1 kΩ resistor.

27 From problem 26, it is obvious that a transistor with no collector current is similar to an open mechanical switch. For this reason, a transistor with no collector current and its collector voltage at the supply voltage level is said to be ''turned off,'' or just OFF.

Question

What are the two main characteristics of an OFF transistor?

____

Answer

It has no collector current, and the collector voltage is equal to the supply voltage.

28 Now, calculate the following parameters for the circuit in Figure 3-24 and compare the results to the examples in problems 26 and 27. Again, the objective here is to find V_C.

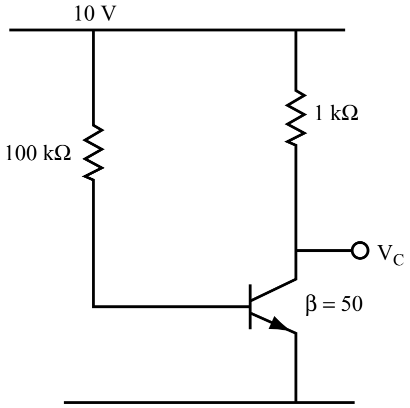

Figure 3-24

Questions

A. $I_B =$ ____

$I_C =$ ____

B. $V_R =$ ____

C. $V_C =$ ____

Answers

A. $I_B = \dfrac{10\,\text{V}}{100\,\text{k}\Omega} = 0.1\,\text{mA}$

$I_C = 50 \times 0.1\text{mA} = 5\text{mA}$

B. $V_R = 1\ \text{k}\Omega \times 5\ \text{mA} = 5\ \text{V}$

C. $V_C = 10\ \text{V} - 5\ \text{V} = 5\ \text{V}$

NOTE **The output voltage in this problem is exactly half the supply voltage. This condition is very important in AC electronics and is covered in Chapter 8.**

The Junction Field Effect Transistor

29 Up to now, the only transistor described was the bipolar junction transistor (BJT). Another common transistor type is the junction field effect transistor (JFET). The JFET, like the BJT, is used in many switching and amplification applications. The JFET is preferred when a high input impedance circuit is needed. The BJT has a relatively low input impedance as compared to the JFET. Like the BJT, the JFET is a three terminal device. The terminals are called the source, drain, and gate. They are similar in function to the emitter, collector, and base, respectively.

Questions

A. How many terminals does a JFET have, and what are these terminals called?

____

B. Which terminal has a function similar to the base of a BJT?

____

Answer

A. Three, called the source, drain, and gate.

B. The gate has a control function similar to that of the base of a BJT.

30 The basic design of a JFET consists of one type of semiconductor material with a channel made of the opposite type of semiconductor material running through it. If the channel is N material, it is called an N-channel JFET; if it is P material, it is called a P-channel. Figure 3-25 shows the basic layout of N and P materials, along with their circuit symbols. Voltage on the gate controls the current flow through the drain and source by controlling the effective width of the channel, allowing more or less current to flow. Thus, the voltage on the gate acts to control the drain current, just as the voltage on the base of a BJT acts to control the collector current.

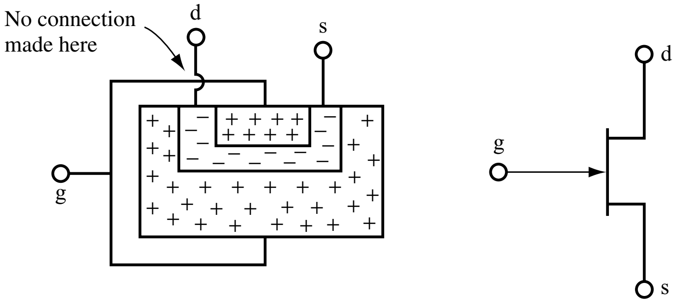

N-channel JFET

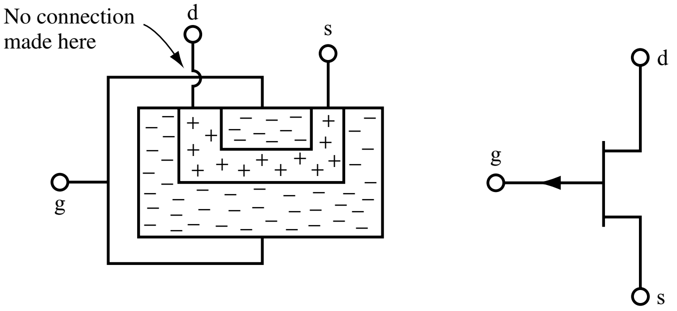

P-channel JFET

Figure 3-25

Questions

A. Which JFET would use electrons as the primary charge carrier for the drain current?

B. What effect does changing the voltage on the gate have on the operation of the JFET?

Answers

A. N-channel, because N material uses electrons as the majority carrier.

B. It changes the current in the drain. The channel width is controlled electrically by the gate potential.

31 To operate the N-channel JFET, apply a positive voltage to the drain with respect to the source. This allows a current to flow through the channel. If the gate is at 0 V, the drain current is at its largest value for safe operation, and the JFET is in the ON condition. When a negative voltage is applied to the gate, the drain current is reduced. As the gate voltage becomes more negative, the current lessens until cutoff, which occurs when the JFET is in the OFF condition. Figure 3-26 shows a typical biasing circuit for the N-channel JFET. For a P-channel JFET, you must reverse the polarity of the bias supplies.

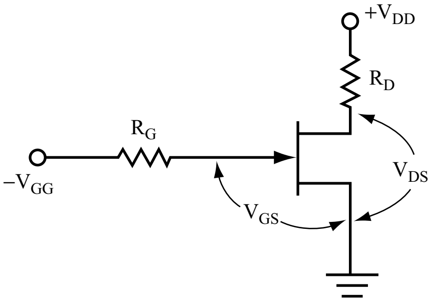

Figure 3-26

Question

How does the ON-OFF operation of a JFET compare to that of a BJT?

Answer

The JFET is ON when there are 0 volts on the gate, whereas you turn the BJT ON by applying a voltage to the base. You turn the JFET OFF by applying a voltage to the gate, and the BJT is OFF when there are 0 volts on the base. The JFET is a ''normally ON'' device, but the BJT is considered a ''normally OFF'' device. Therefore you can use the JFET (like the BJT) as a switching device.

32 When the gate to source voltage is at 0 V ($V_{GS} = 0$) for the JFET shown in Figure 3-26, the drain current is at its maximum or saturation value. This means that the N-channel resistance is at its lowest possible value, in the range of 5 to 200 ohms. If R_D is significantly greater than this, the N-channel resistance, r_{DS}, is assumed to be negligible.

Questions

A. What switch condition would this represent, and what will be the drain to source voltage (V_{DS})?

B. As the gate becomes more negative with respect to the source, the resistance of the N-channel increases until the cutoff point is reached. At this point the resistance of the channel is assumed to be infinite. What condition will this represent, and what will be the drain to source voltage?

C. What does the JFET act like when it is operated between the two extremes of current saturation and current cutoff?

Answers

A. Closed switch, $V_{DS} = 0$ V, or very low value

B. Open switch, $V_{DS} = V_{DD}$

C. A variable resistance

Summary

At this point, it's useful to compare the properties of a mechanical switch with the properties of both types of transistors, as summarized in the following table.

SWITCH	BJT	JFET
OFF or open		
No current	No collector current	No drain current
Full voltage across terminals	Full supply voltage between collector and emitter	Full supply voltage between drain and source
ON or closed		
Full current	Full circuit current	Full circuit current
No voltage across terminals	Collector to emitter voltage is 0 V	Drain to source voltage is 0 V

The terms ON and OFF are used in digital electronics to describe the two transistor conditions you just encountered. Their similarity to a mechanical switch is useful in many electronic circuits. In Chapter 4,you learn about the transistor switch in more detail. This is the first step toward an understanding of digital electronics. In Chapter 8,you examine the operation of the transistor when it is biased at a point falling between the two conditions, ON and OFF. In this mode, the transistor can be viewed as a variable resistance, and utilized as an amplifier.

Self-Test

The questions here test your understanding of the concepts presented in this chapter. Use a separate sheet of paper for your drawings or calculations. Compare your answers with the answers provided following the test.

1. Draw the symbols for an NPN and a PNP bipolar transistor. Label the terminals of each.
2. In Figure 3-27, draw the paths taken by the base and collector currents.

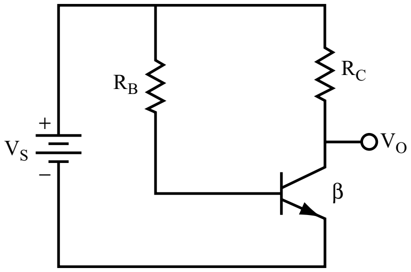

Figure 3-27

3. What causes the collector current to flow?

____

4. What is meant by the term *current gain*? What symbol is used for this? What is its algebraic formula?

____

Use the circuit in Figure 3-27 to answer questions 5 through 10.

5. Assume that the transistor is made of silicon. Set $R_B = 27\ k\Omega$ and $V_s = 3$ V. Find I_B.

____

6. If $R_B = 220\ k\Omega$ and $V_S = 10$ V. Find I_B.

7. Find V_O when $R_B = 100\ k\Omega$, $V_S = 10$ V, $R_C = 1\ k\Omega$, and $\beta = 50$.

8. Find V_O when $R_B = 200\ k\Omega$, $V_S = 10$ V, $R_C = 1\ k\Omega$, and $\beta = 50$.

9. Now use these values to find V_O: $R_B = 47\ k\Omega$, $V_S = 10$ V, $R_C = 500$ ohms, and $\beta = 65$.

10. Use these values to find V_O: $R_B = 68\ k\Omega$, $V_S = 10$ V, $R_C = 820$ ohms, and $\beta = 75$.

11. Draw the symbols for the two types of junction field effect transistors and identify the terminals.

12. What controls the flow of current in both a JFET and a BJT?

13. In the JFET common source circuit shown in Figure 3-28, add the correct polarities of the power supplies and draw the current path taken by the drain current.

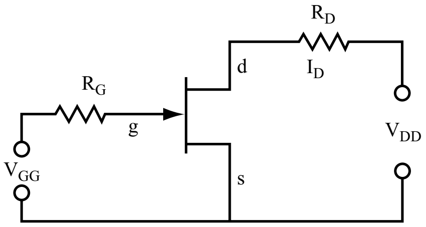

Figure 3-28

14. When a base current is required to turn a BJT ON, why is there no gate current for the JFET in the ON state.

15. Answer the following questions for the circuit shown in Figure 3-29.

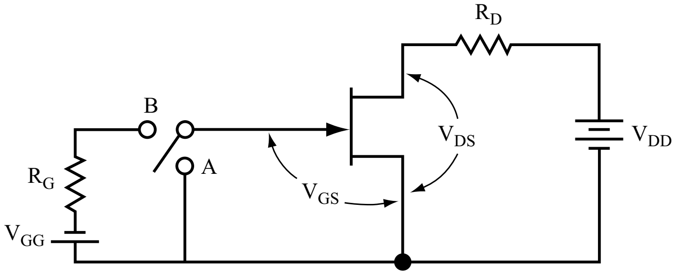

Figure 3-29

A. If the switch is at position A, what will the drain current be, and why? ______________________

B. If the switch is at position B, and the gate supply voltage is of sufficient value to cause cutoff, what will the drain current be, and why? ______________________

C. What is the voltage from the drain to the source for the two switch positions? ______________________

Answers to Self-Test

If your answers do not agree with those given below, review the problems indicated in parentheses before you go on to the next chapter.

---

1. See Figure 3-30. (problems 4, 5)

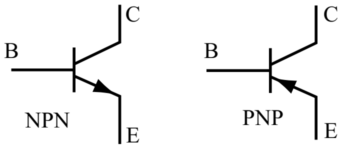

Figure 3.30

2.	See Figure 3-31.	(problems 13,15)

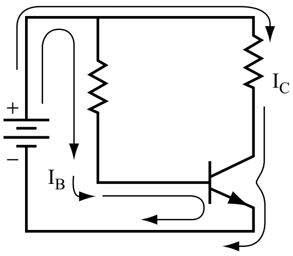

Figure 3.31

3.	Base current.	(problem 15)
4.	Current gain is the ratio of collector current to base current. It is represented by the symbol β. β = I_C/I_B.	(problems 16, 17)
5.	$I_B = \frac{(V_S - 0.7)}{R_B} = \frac{(3V - 0.7V)}{27k\Omega} = \frac{2.3V}{27k\Omega} = 85\mu\ A$	(problem 7)
6.	$I_B = \frac{10V}{220k\Omega} = 45.45\mu A$	(problem 7)
7.	5 V	(problem 21–24)
8.	7.5 V	(problem 21–24)
9.	3.1 V	(problem 21–24)
10.	1 V	(problem 21–24)
11.	See Figure 3-32.	(problem 30)

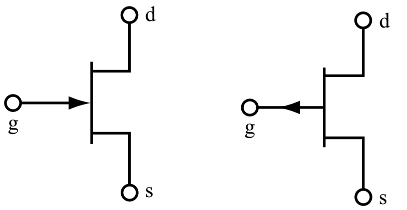

N-channel P-channel

Figure 3.32

12.		The voltage on the gate controls the flow of drain current, which is similar to the base voltage controlling the collector current in a BJT.	(problem 30)
13.		See Figure 3-33.	(problem 31)
		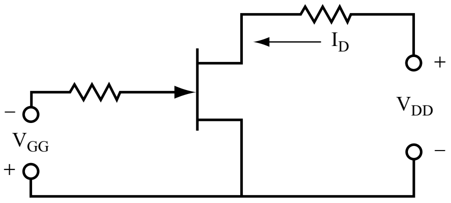 **Figure 3.33**	
14.		The JFET is a high-impedance device and does not draw current from the gate circuit. The BJT is a relatively low-impedance device and does, therefore, require some base current to operate.	(problem 29)
15.	A.	The drain current will be at its maximum value. In this case, it equals V_{DD}/R_D because you can ignore the drop across the JFET. The gate to source voltage is 0 V, which reduces the channel resistance to a very small value close to 0 ohms.	(problem 3-32)
	B.	The drain current now goes to 0 A because the channel resistance is at infinity (very large), which does not allow electrons to flow through the channel.	
	C.	At position A, V_{DS} is approximately 0 V. At position B, $V_{DS} = V_{DD}$.	

CHAPTER 4

The Transistor Switch

Transistors are everywhere. You can't avoid them as you move through your daily tasks. For example, almost all industrial controls, and even your MP3 player, stereo, and television may use transistors as switches.

In Chapter 3, you saw how a transistor can be turned ON and OFF, similar to a mechanical switch. Computers work with Boolean algebra, which uses only two logic states — TRUE and FALSE. These two states are easily represented electronically by a transistor that is ON or OFF. Therefore, the transistor switch is used extensively in computers. In fact, the logic portions of microprocessors (the brains of computers) consist entirely of transistor switches.

This chapter introduces the transistor's simple and widespread application–switching, with emphasis on the bipolar junction transistor (BJT).

When you complete this chapter, you will be able to do the following:

- Calculate the base resistance, which turns a transistor ON and OFF.
- Explain how one transistor will turn another ON and OFF.
- Calculate various currents and resistances in simple transistor switching circuits.
- Calculate various resistances and currents in switching circuits, which contain two transistors.
- Compare the switching action of a junction field effect transistor (JFET) to a BJT.

Turning the Transistor on

1 Start by examining how to turn a transistor ON by using the simple circuit shown in Figure 4-1. In Chapter 3, R_B was given, and you had to find the value of collector current and voltages. Now, do the reverse. Start with the current through R_C and find the value of R_B that will turn the transistor ON and permit the collector current to flow.

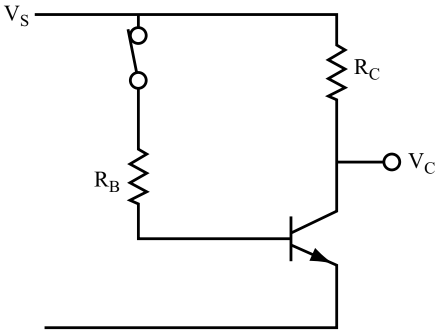

Figure 4-1

Question

What current values do you have to know in order to find R_B?

Answer

The base and collector currents

2 In this problem circuit, a lamp can be substituted for the collector resistor. In this case, R_C (the resistance of the lamp) is referred to as the *load*, and I_C (the current through the lamp) is called the *load current*.

Questions

A. Is load current equivalent to base or collector current?

B. What is the path taken by the collector current discussed in problem 1? Draw this path on the circuit.

Answers

A. Collector current

B. See Figure 4-2. In this figure, note that the resistor symbol has been replaced by the symbol for an incandescent lamp.

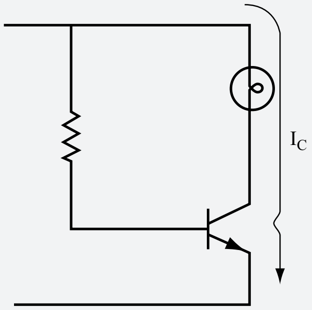

Figure 4-2

3 For the transistor switch to perform effectively as a CLOSED switch, its collector voltage must be at the same voltage as its emitter voltage. In this condition, the transistor is said to be turned ON.

Questions

A. What is the collector voltage when the transistor is turned ON?

____

B. What other component does an ON transistor resemble?

____

Answers

A. The same as the emitter voltage, which, in this circuit, is 0 volts

B. A closed mechanical switch

NOTE **In actual practice, there is a very small voltage drop across the transistor from the collector to the emitter. This is really a saturation voltage and is the smallest voltage drop that can occur across a transistor when it is ON as "hard" as possible. The discussions in this chapter consider this voltage drop to be a negligible value and, therefore, the collector voltage is said to be 0 V. For a quality switching transistor, this is a safe assumption.**

4 The circuit in Figure 4-3 shows a lamp with a resistance of 240 ohms in place of R_C.

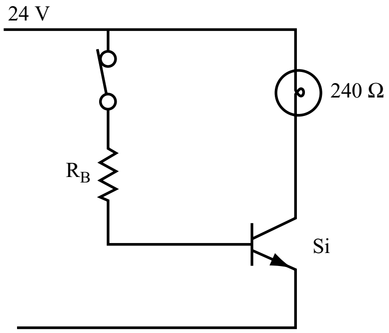

Figure 4-3

This figure shows the supply voltage and the collector resistance. Given these two values, using Ohm's law, you can calculate the load current (also called the *collector current*) as follows:

$$I_L = I_C = \frac{V_S}{R_C} = \frac{24\text{ V}}{240\text{ ohms}} = 100\text{ mA}$$

Thus, 100 mA of collector current must flow through the transistor to fully illuminate the lamp. As you learned in Chapter 3, collector current will not flow unless base current is flowing.

Questions

A. Why do you need base current?

____

B. How can you make base current flow?

____

Answers

A. To enable collector current to flow, so that the lamp will light up

B. By closing the mechanical switch in the base circuit

5 You can calculate the amount of base current flowing. Assume that $\beta = 100$. (Normally, you would look this up in the manufacturer's datasheet for the transistor you are using.)

Questions

A. What is the value of the base current I_B?

____

Answer

$$I_B = \frac{I_C}{\beta} = \frac{100\,\text{mA}}{100} = 1\,\text{mA}$$

6 The base current flows in the direction shown in Figure 4-4. Base current flows through the base-emitter junction of the transistor as it does in a forward-biased diode.

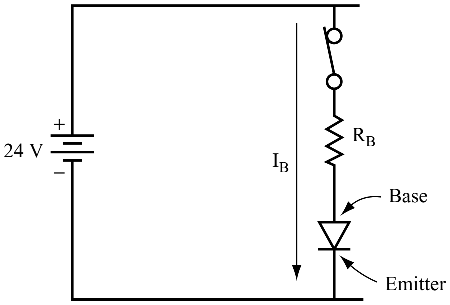

Figure 4-4

Questions

A. What is the voltage drop across the base-emitter diode?

____

B. What is the voltage drop across R_B?

____

Answers

A. 0.7 V, because it is a silicon transistor

B. 24 V if the 0.7 is ignored; 23.3 V if it is not

7 The next step is to calculate R_B. The current flowing through R_B is the base current I_B, and you determined the voltage across it in problem 6.

Question

Calculate R_B.

Answer

$$R_B = \frac{23.3\text{ V}}{1\text{ mA}} = 23,300\,\text{ohms}$$

The final circuit, including the calculated current and resistance values, is shown in Figure 4-5.

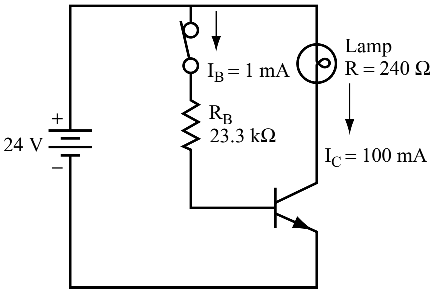

Figure 4-5

8 Use the following steps to calculate the values of I_B and R_B needed to turn a transistor ON:

1. Determine the required collector current.
2. Determine the value of β.
3. Calculate the required value of I_B from the results of steps 1 and 2.
4. Calculate the required value of R_B.
5. Draw the final circuit.

Now, assume that $V_S = 28$ V, that you are using a lamp requiring 50 mA of current, and that $\beta = 75$.

Questions

A. Calculate I_B.

B. Determine R_B.

Answers

A. The collector current and β were given. Thus:

$$I_B = \frac{I_C}{\beta} = \frac{50\,\text{mA}}{75} = 0.667\,\text{mA}$$

B. $R_B = \dfrac{28\,\text{V}}{0.667\,\text{mA}} = 42\,\text{k}\Omega$

This calculation ignores V_{BE}.

9 Now, assume that $V_S = 9$ V, that you are using a lamp requiring 50 mA of current, and that $\beta = 75$.

Question

Calculate R_B.

Answer

$R_B = 31.1\,\text{k}\Omega$

In this calculation, V_{BE} is included.

10 In practice, if the supply voltage is much larger than the 0.7 V drop across the base-emitter junction, you can simplify your calculations by ignoring the 0.7 V drop and assume that all the supply voltage appears across the base resistor R_B. (Most resistors are only accurate to within $+/-5$ percent of their stated value anyway.) If the supply voltage is less than 10 volts, however, you shouldn't ignore the 0.7 V drop across the base-emitter junction.

Questions

Calculate R_B for the following problems, ignoring the voltage drop across the base-emitter junction, if appropriate.

A. A 10 V lamp that draws 10 mA. $\beta = 100$.

B. A 5 V lamp that draws 100 mA. β = 50.

Answers

A. $$I_B = \frac{10\,\text{mA}}{100} = 0.1\,\text{mA}$$

$$R_B = \frac{10\,\text{V}}{0.1\,\text{mA}} = 100\,\text{k}\Omega$$

B. $$I_B = \frac{100\,\text{mA}}{50} = 2\,\text{mA}$$

$$R_B = \frac{(5\,\text{V} - 0.7\,\text{V})}{2\,\text{mA}} = \frac{4.3\,\text{V}}{2\,\text{mA}} = 2.15\,\text{k}\Omega$$

Turning the Transistor off

11 Up to now, you have concentrated on turning the transistor ON, thus making it act like a closed mechanical switch. Now you focus on turning it OFF, thus making it act like an open mechanical switch. If the transistor is OFF, no current flows through the load (that is, no collector current flows).

Questions

A. When a switch is open, are the two terminals at different voltages, or at the same voltage?

B. When a switch is open, does current flow?

C. For a transistor to turn OFF and act like an open switch, how much base current is needed?

Answers

A. At different voltages, the supply voltage and ground voltage.

B. No.

C. The transistor is OFF when there is no base current.

12 You can be sure that there is no base current in the circuit shown in Figure 4-6 by opening the mechanical switch.

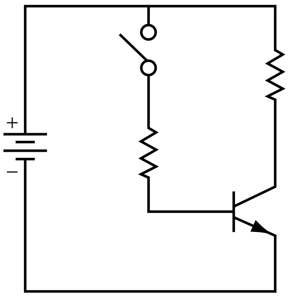

Figure 4-6

To ensure that the transistor remains off when the base is not connected to the supply voltage you add a resistor (labeled R_2 in Figure 4-7) to the circuit. The base of the transistor is connected to ground or 0 volts through this resistor. Therefore, no base current can possibly flow.

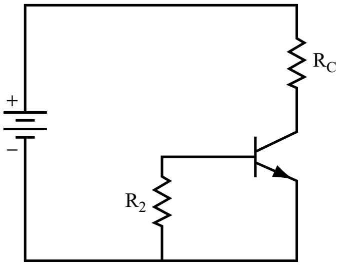

Figure 4-7

Questions

A. Why will current not flow from the supply voltage to the base-emitter junction?

____

B. How much current flows from collector to base?

C. Why will current not flow from collector to base through R_2 ground?

D. Why is the transistor base at 0 volts when R_2 is installed?

Answers

A. There is no current path from the supply voltage through the base-emitter junction. Thus, there is no base current flowing.

B. None at all.

C. The internal construction of the transistor prevents this, because the collector-to-base junction is basically a reverse-biased diode.

D. Because there is no current through R_2, there is no voltage drop across R_2 and, therefore, the transistor base is at ground (0 volts).

13 Because no current is flowing through R_2, you can use a wide range of resistance values. In practice, the values you find for R_2 will be between 1 kΩ and 1 MΩ.

Question

Which of the following resistor values would you use to keep a transistor turned off? 1 ohm, 2 kΩ, 10 kΩ, 20 kΩ, 50 kΩ, 100 kΩ, 250 kΩ, 500 kΩ.

Answer

They would all be suitable except the 1 ohm, because the rest are all above 1 kΩ and below 1 MΩ.

14 Figure 4-8 shows a circuit using both R_1 and R_2. Note that the circuit includes a two-position switch that you can use to turn the transistor ON or OFF.

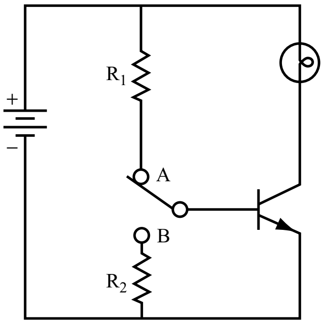

Figure 4-8

Questions

A. As shown in Figure 4-8, is the transistor ON or OFF?

____

B. Which position, A or B, will cause the collector current to be 0 amperes?

____

Answers

A. ON — the base-to-emitter diode is forward biased. Therefore, base current can flow.

B. Position B — the base is tied to ground. Therefore, no base current can flow, and the transistor is OFF.

Why Transistors are Used as Switches

15 You can use the transistor as a switch (as you saw in the previous problems) to perform simple operations such as turning a lamp current on and off. Although often used between a mechanical switch and a lamp, there are other uses for the transistor.

Following are a few other examples that demonstrate the advantages of using a transistor in a circuit as a switch:

- *Example 1* — Suppose you have to put a lamp in a dangerous environment, such as a radioactive chamber. Obviously, the switch to operate the lamp must be placed somewhere safe. You can simply use a switch outside the chamber to turn the transistor switch ON or OFF.

- *Example 2* — If a switch controls equipment that requires large amounts of current, then that current must flow through the wires that run between the switch and the lamp. Because the transistor switch can be turned ON or OFF using low voltages and currents, you can connect a mechanical switch to the transistor switch using small, low-voltage wire and, thereby, control the larger current flow. If the mechanical switch is any distance from the equipment you're controlling, using low-voltage wire will save you time and money.
- *Example 3* — A major problem with switching high current in wires is that the current induces interference in adjacent wires. This can be disastrous in communications equipment such as radio transceivers. To avoid this, you can use a transistor to control the larger current from a remote location, reducing the current needed at the switch that is located in the radio transceiver.
- *Example 4* — In mobile devices (such as a radio controlled airplane), using transistor switches minimizes the power, weight, and bulk required.

Question

What features mentioned in these examples make using transistors as switches desirable?

Answer

The switching action of a transistor can be directly controlled by an electrical signal, as well as by a mechanical switch in the base circuit. This provides a lot of flexibility to the design and allows for simple electrical control. Other factors include safety, reduction of interference, remote switching control, and lower design costs.

16 The following examples of transistor switching demonstrate some other reasons for using transistors:

- *Example 1* — You can control the ON and OFF times of a transistor very accurately, while mechanical devices are not very accurate. This is most important in applications such as photography, where it is necessary to expose a film or illuminate an object for a precise period of time. In these types of uses, transistors are much more accurate and controllable than any other device.

- *Example 2* — A transistor can be switched ON and OFF millions of times a second, and will last for many years. In fact, transistors are one of the longest lasting and most reliable components known, while mechanical switches usually fail after a few thousand operations.
- *Example 3* — The signals generated by most industrial control devices are digital. These control signals can be simply a high or low voltage, which is ideally suited to turning transistor switches ON or OFF.
- *Example 4* — Modern manufacturing techniques allow for the miniaturization of transistors to such a great extent that many of them (even hundreds of millions) can be fabricated into a single silicon chip. Silicon chips on which transistors (and other electronic components) have been fabricated are called *integrated circuits* (ICs). ICs are little flat, black plastic components built into almost every mass-produced electronic device, and are the reason that electronic devices continue to get smaller and lighter.

Question

What other features, besides the ones mentioned in the previous problem, are demonstrated in the examples given here?

Answer

Transistors can be accurately controlled, have high-speed operation, are reliable, have a long life, are very small, have low power consumption, can be manufactured in large numbers at low cost, and are extremely small.

17 At this point, consider the idea of using one transistor to turn another one ON and OFF, and using the second transistor to operate a lamp or other load. (This idea is explored in the next section of this chapter.)

If you have to switch many high current loads, then you can use one switch that controls several transistors simultaneously.

Questions

A. With the extra switches added, will the current that flows through the main switch be more or less than the current that flows through the load?

B. What effect do you think the extra transistor will have on the following?

1. Safety

____

2. Convenience to the operator

____

3. Efficiency and smoothness of operation

____

Answers

A. Less current will flow through the main switch than through the load.

B.
1. It increases safety and allows the operator to stay isolated from dangerous situations.
2. Switches can be placed conveniently close together on a panel, or in the best place for an operator, rather than the switch position dictating operator position.
3. One switch can start many things, as in a master lighting panel in a television studio or theater.

18 This problem reviews your understanding of the concepts presented in problems 15, 16, and 17.

Questions

Indicate which of the following are good reasons for using a transistor as a switch:

A. To switch equipment in a dangerous or inaccessible area on and off

B. To switch very low currents or voltages

C. To lessen the electrical noise that might be introduced into communication and other circuits

D. To increase the number of control switches

E. To use a faster, more reliable device than a mechanical switch

____

Answers

A, C, and E.

19 Many types of electronic circuits contain multiple switching transistors. In this type of circuit, one transistor is used to switch others ON and OFF. To illustrate how this works, again consider the lamp as the load and the mechanical switch as the actuating element. Figure 4-9 shows a circuit that uses two transistors to turn a lamp on or off.

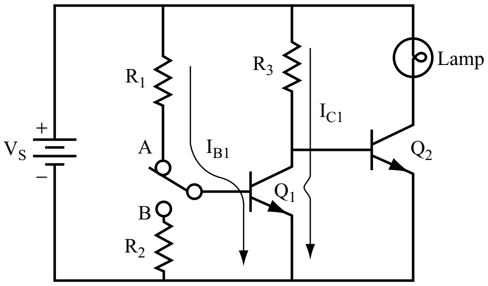

Figure 4-9

When the switch is in position A, Q_1's base-emitter junction is forward biased. Therefore, base current (I_{B1}) flows through R_1 and through Q_1's base-emitter diode, turning the transistor ON. This causes the collector current (I_{C1}) to flow through Q_1 to ground, and the collector voltage drops to 0 volts, just as if Q_1 was a closed switch. Because the base of Q_2 is connected to the collector of Q_1, the voltage on the base of Q_2 also drops to 0 volts. This ensures that Q_2 is turned OFF and the lamp remains unlit.

Now, flip the switch to position B, as shown in Figure 4-10. The base of Q_1 is tied to ground, or 0 volts, turning Q_1 OFF. Therefore, no collector current can flow through Q_1. A positive voltage is applied to the base of Q_2 and Q_2's emitter-base junction is forward biased. This allows current to flow through R_3 and the emitter-base junction of Q_2, which turns Q_2 ON, allowing collector current (IC_2) to flow and the lamp is illuminated.

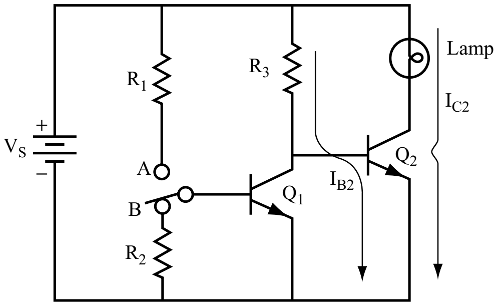

Figure 4-10

Now that you have read the descriptions of how the circuit works, try to answer the following questions. First assume the switch is in position A, as shown in Figure 4-9.

Questions

A. What effect does I_{B1} have on transistor Q_1?

____

B. What effect does turning Q_1 ON have on

1. Collector current I_{C1}?

   ____

2. Collector voltage V_{C1}?

   ____

C. What effect does the change to V_{C1} covered in the previous question have on

1. The base voltage of Q_2?

   ____

2. Transistor Q_2 (that is, is it ON or OFF)?

   ____

D. Where does the current through R_3 go?

____

E. In this circuit is the lamp on or off?

____

Answers

A. I_{B1}, along with a portion of V_S (0.7 volts if the transistor is silicon), turns Q_1 ON.

B. (1) I_{C1} flows; (2) V_{C1} drops to 0 V.

C. (1) base of Q_2 drops to 0 V; (2) Q_2 is OFF.

D. I_{C1} flows through Q_1 to ground.

E. Off.

20 Now, assume the switch is in the B position, as shown in Figure 4-10, and answer these questions.

Questions

A. How much base current I_{B1} flows into Q_1?________________

B. Is Q_1 ON or OFF?________________

C. What current flows through R_3?________________

D. Is Q_2 ON or OFF?________________

E. Is the lamp on or off?________________

Answers

A. None

B. OFF

C. I_{B2}

D. ON

E. On

21 Refer back to the circuit in Figure 4-9 and 4-10. Now, answer these questions assuming the supply voltage is 10 V.

Questions

A. Is the current through R_3 ever divided between Q_1 and Q_2? Explain.

____

____

B. What is the collector voltage of Q_2 with the switch in each position?

____

C. What is the collector voltage of Q_1 with the switch in each position?

____

Answers

A. No. If Q_1 is ON, all the current flows through it to ground as collector current. If Q_1 is OFF, all the current flows through the base of Q_2 as base current.

B. In position A, 10 volts because it is OFF.

Inposition B, 0 volts because it is ON.

C. In position A, 0 volts because it is ON.

In position B, the collector voltage of Q_1 equals the voltage drop across the forward-biased base-emitter junction of Q_2, because the base of Q_2 is in parallel with the collector of Q_1. The voltage drop across the forward-biased base-emitter junction will not rise to 10 V, but can only rise to 0.7 V if Q_2 is made of silicon.

22 Now calculate the values of R_1, R_2, and R_3 for this circuit. The process is similar to the one you used before, but you have to expand it to deal with the second transistor. Similar to the steps you used in problem 8. Follow these steps to calculate R_1, R_2, and R_3:

1. Determine the load current I_{C2}.
2. Determine β for Q_2. Call this β_2.
3. Calculate I_{B2} for Q_2. Use $I_{B2} = I_{C2}/\beta_2$.
4. Calculate R_3 to provide this base current. Use $R_3 = V_S/I_{B2}$.
5. R_3 is also the load for Q_1 when Q_1 is ON. Therefore, the collector current for Q_1 (I_{C1}) will have the same value as the base current for Q_2, as calculated in step 3.
6. Determine β_1, the β for Q_1.
7. Calculate the base current for Q_1. Use $I_{B1} = I_{C1}/\beta_1$.
8. Find R_1. Use $R_1 = V_s/I_{B1}$.
9. Choose R_2. For convenience, let $R_2 = R_1$.

Continue to work with the same circuit shown in Figure 4-11. Use the following values:

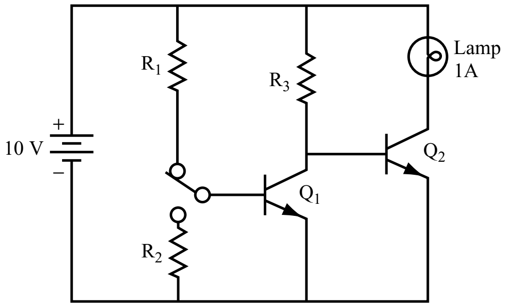

Figure 4-11

- A 10-volt lamp that draws 1 ampere; therefore $V_S = 10$ volts, $I_{C2} = 1$ A.
- $\beta_2 = 20$, $\beta_1 = 100$

Ignore any voltage drops across the transistors.

Questions

Calculate the following:

A. Find I_{B2} as in step 3.

I_{B2} = __________

B. Find R_3 as in step 4.

R_3 = __________

C. Calculate the load current for Q_1 when it is ON as shown in step 5.

I_{C1} = __________

D. Find the base current for Q_1.

I_{B1} = __________

E. Find R_1 as in step 8.

R_1 = __________

F. Choose a suitable value for R_2.

R_2 = __________

Answers

The following answers correspond to the steps.

A. **1.** I_{C2} is given as 1 A.

2. $\beta_2 = 20$ (given). This is a typical value for a transistor that would handle 1 A.

3. $I_{B2} = \frac{1\text{ A}}{20} = 50\text{ mA}$

B. **4.** $R_3 = \frac{10\text{ Volts}}{50\text{ mA}} = 200\ \Omega$

Note that the 0.7 V base-emitter drop has been ignored.

C. **5.** $I_{C1} = I_{B2} = 50$ mA

D. **6.** $\beta_1 = 100$

7. $I_{B1} = \frac{50\,\text{mA}}{100} = 0.5\,\text{mA}$

E. **8.** $R_1 = \frac{10\,\text{V}}{0.5\,\text{mA}} = 20\,\text{k}\Omega$

Again, the 0.7 V drop is ignored.

F. **9.** For convenience choose a value for R_2 that is the same as R_1, or 20 kΩ. This reduces the number of different components in the circuit. The fewer different components you have in a circuit, the less components you have to keep in your parts bin. You could, of course, choose any value between 1 kΩ and 1 MΩ.

23 Following the same procedure, and using the same circuit shown in Figure 4-11, work through this example. Assume that you are using a 28-volt lamp that draws 560 mA, and that $\beta_2 = 10$ and $\beta_1 = 100$.

Questions

Calculate the following:

A. I_{B2} = _______________

B. R_3 = _______________

C. I_{C1} = _______________

D. I_{B1} = _______________

E. R_1 = _______________

F. R_2 = _______________

Answers

A. 56 mA

B. 500 ohms

C. 56 mA

D. 0.56 mA

E. 50 kΩ

F. 50 kΩ by choice

The Three-Transistor Switch

24 The circuit shown in Figure 4-12 uses three transistors to switch a load on and off. In this circuit Q_1 is used to turn Q_2 ON and OFF, and Q_2 is used to turn Q_3 ON and OFF. The calculations are similar to those you performed in the last few problems, but a few additional steps are required to deal with the third transistor. Use this circuit diagram to determine the answers to the following questions.

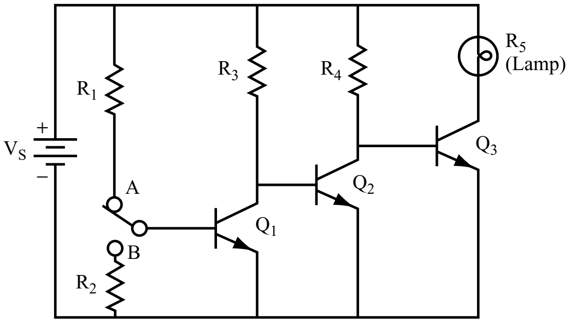

Figure 4-12

Questions

If the switch is in position A:

A. Is Q_1 ON or OFF?

B. Is Q_2 ON or OFF?

C. Where is current through R_4 flowing?

D. Is Q_3 ON or OFF?

Answers

A. ON

B. OFF

C. Into the base of Q_3

D. ON

25 Now use the same circuit as in problem 24.

Questions

If the switch is in position B:

A. Is Q_1 ON or OFF?

B. Is Q_2 ON or OFF?

C. Where is the current through R_4 flowing?

D. Is Q_3 ON or OFF?

E. Which switch position turns the lamp on?

F. How do the on/off positions for the switch in the three-transistor switch differ from the on/off positions for the switch in the two-transistor switch circuit?

Answers

A. OFF.

B. ON.

C. Through Q_2 to ground.

D. OFF.

E. Position A.

F. The positions are opposite. Therefore, if a circuit controls lamps with two transistors, and another circuit controls lamps with three transistors, flipping the switch that controls both circuits would change which lamps (or which other loads) are on.

26 Work through this example using the same equations you used for the two-transistor switch in problem 22. The steps are similar, but with a few added steps, as shown here:

1. Find the load current. This is often given.
2. Determine the current gain of Q_3. This is β_3 and usually it is a given value.
3. Calculate I_{B3}. Use $I_{B3} = I_{C3}/\beta_3$.
4. Calculate R_4. Use $R_4 = V_S/I_{B3}$.
5. Assume $I_{C2} = I_{B3}$.
6. Find β_2. Again this is a given value.
7. Calculate I_{B2}. Use $I_{B2} = I_{C2}/\beta_2$.
8. Calculate R_3. Use $R_3 = V_S/I_{B2}$.
9. Assume $I_{C1} = I_{B2}$.
10. Find β_1.
11. Calculate I_{B1}. Use $I_{B1} = I_{C1}/\beta_1$.
12. Calculate R_1. Use $R_1 = V_s/I_{B1}$.
13. Choose R_2.

For this example, use a 10-volt lamp that draws 10 amperes. Assume the βs of the transistors are given in the manufacturer's datasheets as $\beta_1 = 100$, $\beta_2 = 50$, and $\beta_3 = 20$. Now work through the steps, checking the answers for each step as you complete it.

Questions

Calculate the following:

A. I_{B3} = ____________

B. R_4 = ____________

C. I_{B2} = ____________

D. R_3 = ____________

E. I_{B1} = ____________

F. R_1 = ____________

G. R_2 = ____________

Answers

The answers here correspond to the steps.

A. **1.** The load current is given as 10 A.

2. β_3 is given as 20.

3. $I_{B3} = \frac{I_{C3}}{\beta_3} = \frac{10\text{ A}}{20} = 0.5\text{ A} = 500\text{ mA}$

B. **4.** $R_4 = \frac{10\text{ volts}}{500\text{ mA}} = 20\text{ ohms}$

C. **5.** $I_{C2} = I_{B3} = 500\text{ mA}$

6. β_2 is given as 50.

7. $I_{B2} = \frac{I_{C2}}{\beta_2} = \frac{500\text{ mA}}{50} = 10\text{ mA}$

D. **8.** $R_3 = \frac{10\text{ V}}{10\text{ mA}} = 1\text{ k}\Omega$

E. **9.** $I_{C1} = I_{B2} = 10\text{ mA}$

10. β_1 is given as 100.

11. $I_{B1} = \frac{I_{C1}}{\beta_1} = \frac{10\text{ mA}}{100} = 0.1\text{ mA}$

F. **12.** $R_1 = \frac{10\text{ V}}{0.1\text{ mA}} = 100\text{ k}\Omega$

G. **13.** R_2 can be chosen to be 100 kΩ also.

27 Determine the values in the same circuit for a 75-volt lamp that draws 6 A. Assume that $\beta_3 = 30$, $\beta_2 = 100$, and $\beta_1 = 120$.

Questions

Calculate the following values using the steps in problem 26:

A. I_{B3} = ________________

B. R_4 = ________________

C. I_{B2} = ________________

D. R_3 = ________________

E. I_{B1} = ________________

F. R_1 = ________________

G. R_2 = ________________

Answers

A. 200 mA

B. 375 Ω

C. 2 mA

D. 37.5 kΩ

E. 16.7 μA

F. 4.5 MΩ

G. Choose $R_2 = 1$ MΩ

Alternative Base Switching

28 In the examples of transistor switching, the actual switching was performed using a small mechanical switch placed in the base circuit of the first transistor. This switch has three terminals, and switches from position A to position B. (This is a single pole double throw switch.) This switch does not have a definite ON or OFF position as does a simple ON-OFF switch.

Question

Why couldn't a simple ON-OFF switch with only two terminals have been used with these examples?

____

Answer

An ON-OFF switch is either open or closed, and cannot switch between position A and position B, as shown earlier in Figure 4-12.

29 If you connect R_1, R_2, and a switch together as shown in Figure 4-13, you can use a simple ON-OFF switch with only two terminals. (This is a single pole single throw switch.)

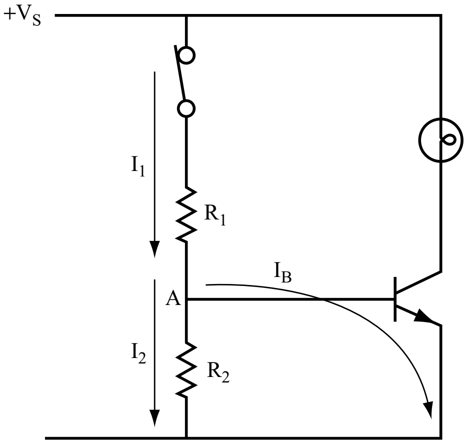

Figure 4-13

Questions

A. When the switch is open, is Q_1 ON or OFF?

____

B. When the switch is closed, is the lamp on or off?

____

Answers

A. OFF

B. On

30 When the switch is closed, current flows through R_1, However, at point A in Figure 4-13, the current divides into two paths. One path is the base current I_B, and the other is marked I_2.

Question

How could you calculate the total current I_1?

____

Answer

$I_1 = I_B + I_2$

31 The problem now is to choose the values of both R_1 and R_2 so that when the current divides, there will be sufficient base current to turn Q_1 ON.

Question

Consider this simple example. Assume the load is a 10-volt lamp that needs 100 mA of current, and $\beta = 100$. Calculate the base current required.

$I_B =$ ____________

Answer

$$I_B = \frac{100\,\text{mA}}{100} = 1\,\text{mA}$$

32 After the current I_1 flows through R_1, it must divide, and 1 mA of it becomes I_B. The remainder of the current is I_2. The difficulty at this point is that there is no unique value for either I_1 or I_2. In other words, you could assign them almost any value. The only restriction is that both must permit 1 mA of current to flow into the base of Q_1.

You must make an arbitrary choice for these two values. Based on practical experience, it is common to set I_2 to be 10 times greater than I_B. This split makes the circuits work reliably, and keeps the calculations easy:

$I_2 = 10\,I_B$

$I_1 = 11\,I_B$

Question

In problem 31 you determined that $I_B = 1$ mA. What is the value of I_2?

____

Answer

$I_2 = 10$ mA

33 Now it is possible to calculate the value of R_2. The voltage across R_2 is the same as the voltage drop across the base-emitter junction of Q_1. Assume that the circuit uses a silicon transistor, so this voltage is 0.7 V.

Questions

A. What is the value of R_2?

B. What is the value R_1?

Answers

A. $R_2 = \dfrac{0.7\,\text{V}}{10\,\text{mA}} = 70\,\text{ohms}$

B. $R_2 = \dfrac{(10\,\text{V} - 0.7\,\text{V})}{11\,\text{mA}} = \dfrac{9.3\,\text{V}}{11\,\text{mA}} = 800\,\text{ohms (approximately)}$

You can ignore the 0.7 V in this case, which would give $R_1 = 910$ ohms.

34 The resistor values you calculated in problem 33 ensure that the transistor turns ON, and that the 100 mA current (I_C) you need to illuminate the lamp flows through the lamp and the transistor. The labeled circuit is shown in Figure 4-14.

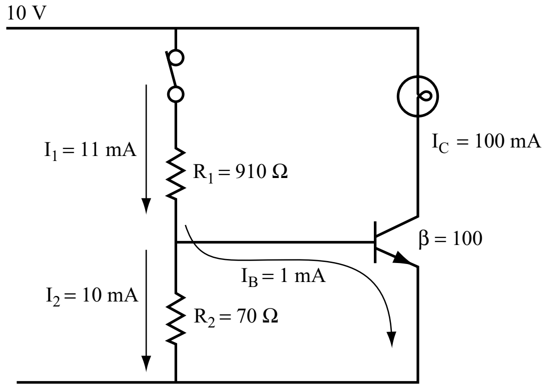

Figure 4-14

Questions

For each of the following lamps, perform the same calculations you used in the last few problems to find the values of R_1 and R_2.

A. A 28-volt lamp that draws 56 mA. $\beta = 100$

____

B. A 12-volt lamp that draws 140 mA. $\beta = 50$

____

Answers

A. $I_B = \dfrac{56\,\text{mA}}{100} = 0.56\,\text{mA}$

$I_2 = 5.6\,\text{mA}$

$R_2 = \dfrac{0.7\,\text{V}}{5.6\,\text{mA}} = 125\,\text{ohms}$

$R_1 = \dfrac{28\,\text{V}}{6.16\,\text{mA}} = 4.5\,\text{k}\Omega$

B. $R_2 = 25\,\text{ohms}$

$R_1 = 400\,\text{ohms}$

35 The arbitrary decision to make the value of I_2 10 times the value of I_B is obviously subject to considerable discussion, doubt, and disagreement. *Transistors are not exact devices*; *they are not carbon copies of each other*. In general, any transistor of the same type will have a different β from any other because of the variance in tolerances found in component manufacturing. This leads to a degree of inexactness in designing and analyzing transistor circuits. The truth is that if you follow exact mathematical procedures, it will complicate your life. In practice, a few ‘‘rules of thumb’’ have been developed to help you make the necessary assumptions. These rules lead to simple equations that provide workable values for components that you can use in designing circuits.

The choice of $I_2 = 10I_B$ is one such rule of thumb. Is it the only choice that will work? Of course not. Almost any value of I_2 that is at least 5 times larger than I_B will work. Choosing 10 times the value is a good option for three reasons:

- It is a good practical choice. It will always work.
- It makes the arithmetic easy.
- It's not overly complicated, and doesn't involve unnecessary calculations.

Question

In the example from problem 32, $I_B = 1$ mA and $I_2 = 10$ mA. Which of the following values could also work efficiently for I_2?

A. 5 mA

B. 8 mA

C. 175 mA

D. 6.738 mA

E. 1 mA

Answer

Choices A, B, and D. Value C is much too high to be a sensible choice, and E is too low.

36 Before you continue with this chapter, answer the following review questions.

Questions

A. Which switches faster, the transistor or the mechanical switch?

____

B. Which can be more accurately controlled?

____

C. Which is the easiest to operate remotely?

____

D. Which is the most reliable?

____

E. Which has the longest life?

____

Answers

A. The transistor is much faster.

B. The transistor.

C. The transistor.

D. The transistor.

E. Because transistors have no moving parts they have a much longer operating lifetime than a mechanical switch. A mechanical switch will fail after several thousand operations, while transistors can be operated several million times a second and last for years.

Switching The JFET

37 The use of the junction field effect transistor (JFET) as a switch is discussed in the next few problems. You may want to review problems 29 through 32 of Chapter 3 where this book introduced the JFET.

The JFET is considered a ''normally on'' device, which means that with 0 volts applied to the input terminal (called the *gate*), it is ON and current can flow through the transistor. When you apply a voltage to the gate, the device conducts less current because the resistance of the drain to source channel increases. At some point, as the voltage increases, the value of the resistance in the channel becomes so high that the device ''cuts off'' the flow of current.

Questions

A. What are the three terminals for a JFET called, and which one controls the operation of the device?

____

B. What turns the JFET ON and OFF?

____

Answers

A. Drain, source, and gate, with the gate acting as the control.

B. When the gate voltage is zero (at the same potential as the source), the JFET is ON. When the gate to source voltage difference is high, the JFET is OFF.

The JFET Experiment

38 The objective of the following experiment is to determine the drain current that will flow when a JFET is fully ON and the gate voltage needed to fully shut the JFET OFF, using the circuit shown in Figure 4-15. You can change the gate voltage (actually, the voltage difference between the gate and the source, or V_{GS}) by adjusting the potentiometer, while measuring the resulting current flow through the transistor from drain to source (I_D). When the JFET is OFF, I_D is at zero; when the JFET is fully ON, I_D is at its maximum (called I_{DSS}).

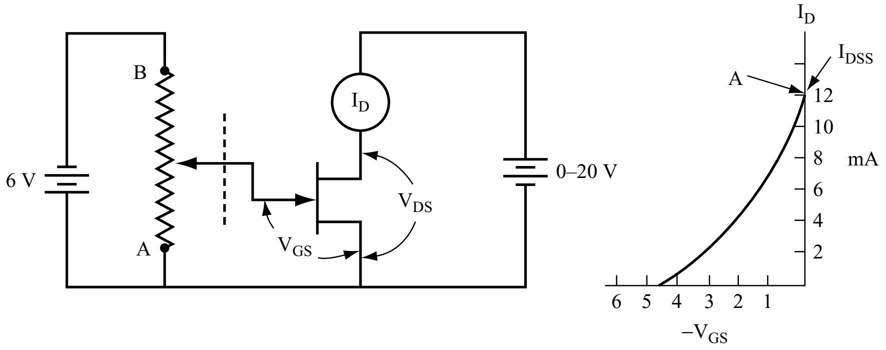

Figure 4-15

You will need the following equipment and supplies:

- One 6 V battery pack.
- One 12 V battery pack (or a lab power supply).
- One multimeter set to measure current.
- One multimeter set to measure voltage.
- One JFET.
- One potentiometer. Any value of potentiometer will work. If you have a selection available, use a 10 kΩ potentiometer.
- One breadboard.

If you do not have what you need to set up the circuit and measure the values, just read through the experiment.

Follow these steps to complete the experiment:

1. Set up the circuit shown in Figure 4-15 on the breadboard with the power supply for the drain portion of the circuit set to 12 V.

2. Set the potentiometer to 0 ohms and take a reading of V_{GS} and I_D and record the readings in the following table.

V_{GS}	I_D

3. Increase the potentiometer resistance and take another set of readings and record them. Repeat this until I_D drops to 0 mA.
4. 4 Plot the recorded points on a graph as shown on the right-hand side of Figure 4-15, with I_D as the vertical axis and V_{GS} as the horizontal axis. Draw a curve through the points.

With the potentiometer set at point A (0 ohms), the voltage from the gate to the source is zero ($V_{GS} = 0$). The current that flows between the drain and source terminals of the JFET at this time is at its maximum value, and is called the *saturation current* (I_{DSS}).

On the graph shown in the right side of Figure 4-15, the saturation point is indicated at A. This graph shows the transfer curve, and is a characteristic curve that is included on JFET datasheets.

NOTE **If you are using a lab power supply for the drain portion of the circuit, you should have it set to 12 volts. One property of the saturation current is that when V_{GS} is set at zero, and the transistor is fully ON, the current doesn't drop as long as the value of V_{DS} is above a few volts. If you have an adjustable power supply, you can determine the value of V_{DS} at which I_D starts to drop by starting with the power supply set at 12 volts. Watch the value of I_D as you lower the power supply voltage until you see I_D start to decrease.**

Questions

Using the transfer curve shown, answer the following:

A. With $V_{GS} = 0$, what is the value of drain current?

B. Why is this value called the drain saturation current?

C. What is the gate to source cutoff voltage for the curve shown?

D. Why is this called a cutoff voltage?

Answers

A. 12 mA on the graph.

B. The word saturation is used to indicate that the current is at its maximum.

C. Approximately −4.2 V on the graph.

D. It is termed a cutoff voltage because at this value the drain current goes to 0 A.

39 Now, look at the circuit shown in Figure 4-16. Assume that the JFET has the transfer characteristic shown by the curve in Figure 4-15.

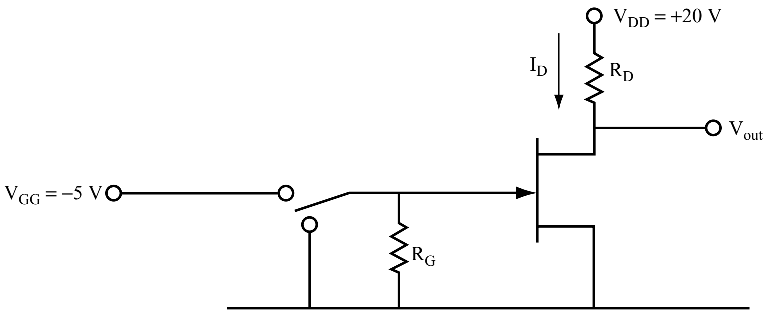

Figure 4-16

When the gate is connected to ground, the drain current will be at 12 mA. Assuming that the drain to source resistance is negligible, you can calculate the required value for R_D using the following formula:

$$R_D = \frac{V_{DD}}{I_{DSS}}$$

If you know the drain to source voltage, then you can include it in the calculation.

$$R_D = \frac{(V_{DD} - V_{DS})}{I_{DSS}}$$

Question

What should the value of R_D be for the I_{DSS} shown at Point A in the curve?

Answer

$$R_D = \frac{20\,\text{V}}{12\,\text{mA}} = 1.67\,\text{k}\Omega$$

40 For the JFET circuit shown in Figure 4-16, assume that $V_{DS} = 1$ V when the I_D is at saturation.

Questions

A. What is the required value of R_D?

B. What is the effective drain to source resistance (r_{DS}) in this situation?

Answers

A. $R_D = \dfrac{(20\,\text{volts} - 1\,\text{volt})}{12\,\text{mA}} = 1583\,\text{ohms}$

B. $r_{DS} = \dfrac{V_{DS}}{I_{DSS}} = \dfrac{1\,\text{volt}}{12\,\text{mA}} = 83\,\text{ohms}$

NOTE **You can see from this calculation that R_D is 19 times greater than r_{DS}. Thus, ignoring V_{DS} and assuming that $r_{DS} = 0$ does not greatly affect the value of R_D. The 1.67 kΩ value is only about 5 percent higher than the 1583 ohm value for R_D.**

41 Now, turn the JFET OFF. From the curve shown in Figure 4-15, you can see that a cutoff value of −4.2 V is required. Use a gate to source value of −5 V to ensure that the JFET is in the ''hard OFF'' state. The purpose of resistor R_G is to ensure that the gate is connected to ground while you flip the switch between terminals, changing the gate voltage from one level to the other. Use a large value of 1 MΩ here to avoid drawing any appreciable current from the gate supply.

Question

When the gate is at the −5 V potential, what is the drain current and the resultant output voltage?

Answer

$I_D = 0$ A and $V_{out} = V_{DS} = 20$ V, which is V_{DD}

Summary

In this chapter, you learned about the transistor switch and how to calculate the resistor values required to use it in a circuit.

- You worked with a lamp as the load example, because this provides an easy visual demonstration of the switching action. All of the circuits shown in this chapter will work when you build them on a breadboard, and the voltage and current measurements will be very close to those in the text.
- You have not yet learned all there is to transistor switching. For example, you haven't found out how much current a transistor can conduct before it burns out, what maximum voltage a transistor can sustain, or how fast a transistor can switch ON and OFF. You can learn these things from the datasheet for each transistor model, so these are not covered here.
- When you use the JFET as a switch, it will not switch as fast as a BJT, but it does have certain advantages relating to its large input resistance. The JFET does not draw any current from the control circuit in order to operate. Conversely, a BJT will draw current from the control circuit because of its lower input resistance.

Self-Test

These questions test your understanding of the concepts introduced in this chapter. Use a separate sheet of paper for your diagrams or calculations. Compare your answers with the answers provided.

For the first three questions, use the circuit shown in Figure 4-17. The objective is to find the value of R_B that will turn the transistor ON. As you may know, resistors are manufactured with ''standard values.'' After you have calculated an exact value, choose the nearest standard resistor value from Appendix D.

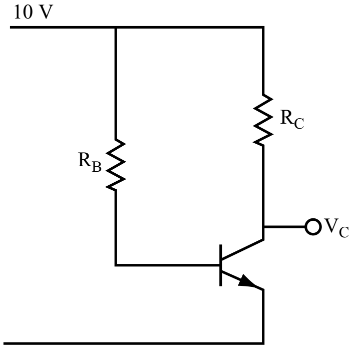

Figure 4-17

1. $R_C = 1\Omega$, $\beta = 100$, $R_B =$ ____________
2. $R_C = 4.7\ k\Omega$, $\beta = 50$. $R_B =$ ____________
3. $R_C = 22\ k\Omega$, $\beta = 75$. $R_B =$ ____________

For questions 4–6, use the circuit shown in Figure 4-18. Find the values of R_3, R_2, and R_1 that ensure that Q_2 is ON or OFF when the switch is in the corresponding position. Calculate the resistors in the order given. After you have found the exact values, again choose the nearest standard resistor values.

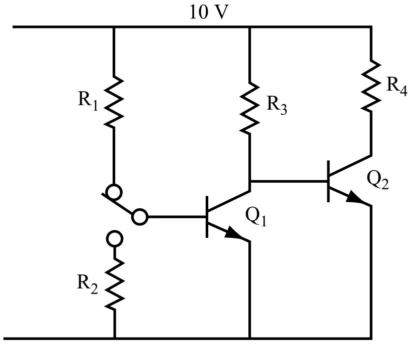

Figure 4-18

NOTE You should be aware that rounding off throughout a problem, or rounding off the final answer, could produce slightly different results.

4. $R_4 = 100$ ohms, $\beta_1 = 100$, $\beta_2 = 20$.

$R_3 =$ ___________ $R_1 =$ ___________ $R_2 =$ ___________

5. $R_4 = 10$ ohms, $\beta_1 = 50$, $\beta_2 = 20$.

$R_3 =$ ___________ $R_1 =$ ___________ $R_2 =$ ___________

6. $R_4 = 250$ ohms, $\beta_1 = 75$, $\beta_2 = 75$.

$R_3 =$ ___________ $R_1 =$ ___________ $R_2 =$ ___________

For questions 7–9, find the values of the resistors in the circuit shown in Figure 4-19 that ensure that Q_3 will be ON or OFF when the switch is in the corresponding position. Then, select the nearest standard resistor values.

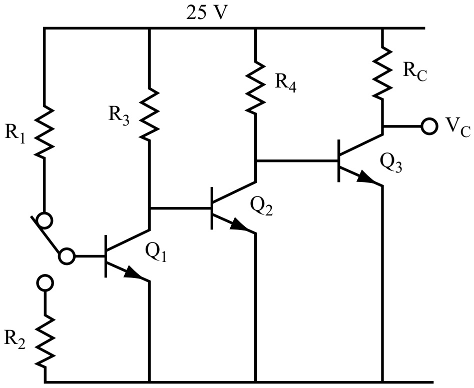

Figure 4-19

7. $R_C = 10$ ohms, $\beta_3 = 20$, $\beta_2 = 50$, $\beta_1 = 100$.

$R_4 =$ ___________ $R_2 =$ ___________

$R_3 =$ ___________ $R_1 =$ ___________

8. $R_C = 28$ ohms, $\beta_3 = 10$, $\beta_2 = 75$, $\beta_1 = 75$.

$R_4 =$ ___________ $R_2 =$ ___________

$R_3 =$ ___________ $R_1 =$ ___________

9. $R_C = 1$ ohm, $\beta_3 = 10$, $\beta_2 = 50$, $\beta_1 = 75$.

$R_4 =$ ___________ $R_2 =$ ___________

$R_3 =$ ___________ $R_1 =$ ___________

Questions 10–12 use the circuit shown in Figure 4-20. Find values for R_1 and R_2 that ensure that the transistor turns ON when the switch is closed and OFF when the switch is open.

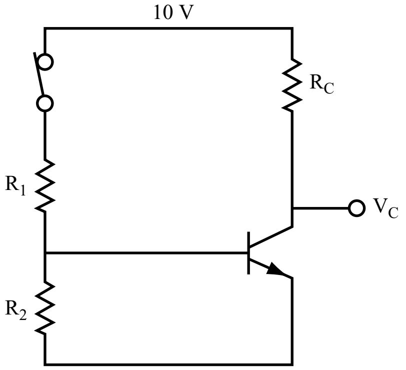

Figure 4-20

10. $R_C = 1\ k\Omega$, $\beta = 100$.

R_1 = ____________ R_2 = ____________

11. $R_C = 22\ k\Omega$, $\beta = 75$.

R_1 = ____________ R_2 = ____________

12. $R_C = 100\ \Omega$, $\beta = 30$.

R_1 = ____________ R_2 = ____________

13. An N-channel JFET has a transfer curve with the following characteristics. When $V_{GS} = 0$ V, the saturation current, (I_{DSS}) is 10.5 mA; the cutoff voltage is −3.8 V. With a drain supply of 20 V, design a biasing circuit that switches the JFET from the ON state to the OFF state.

Answers to Self-Test

The exercises in this Self-Test show calculations that are typical of those found in practice, and the odd results you sometimes get are quite common. Thus, choosing a nearest standard value of resistor is a common practice. If your answers do not agree with those given here, review the problems indicated in parentheses before you go on to the next chapter.

1.	100 kΩ	(problem 8)
2.	235 kΩ. Choose 240 kΩ as a standard value.	(problem 8)
3.	1.65 MΩ. Choose 1.6 MΩ as a standard value.	(problem 8)
4.	$R_3 = 2k\Omega$; $R_1 = 200\ k\Omega$; $R_2 = 200\ k\Omega$. Use these values.	(problem 22)

5.	$R_3 = 200$ ohms; $R_1 = 10$ kΩ; $R_2 = 10$ kΩ. Use these values.	(problem 22)
6.	$R_3 = 18.8$ kΩ. Choose 18 kΩ as a standard value.	(problem 22)
	$R_1 = 1.41$ MΩ. Choose 1.5 MΩ as a standard value.	
	Select 1 MΩ for R_2.	
7.	$R_4 = 200$ ohms; $R_3 = 10$ kΩ; $R_2 = 1$ MΩ; $R_1 = 1$ MΩ. Use these values.	(problem 26)
8.	$R_4 = 280$ ohms. Choose 270 ohms as a standard value.	(problem 26)
	$R_3 = 21$ kΩ. Choose 22 kΩ as a standard value.	
	$R_2 = 1.56$ MΩ. Choose 1.5 or 1.6 MΩ as a standard value.	
	$R_1 = 1.56$ MΩ. Choose 1.5 or 1.6 MΩ as a standard value.	
9.	$R_4 = 10$ ohms. Choose 10 ohms as a standard value.	(problem 26)
	$R_3 = 500$ ohms. Choose 510 ohms as a standard value.	
	$R_2 = 37.5$ kΩ. Choose 39 kΩ as a standard value.	
	$R_1 = 37.5$ kΩ. Choose 39 kΩ as a standard value.	
10.	$R_2 = 700$ ohms. Choose 680 or 720 ohms as a standard value.	(problems 31–33)
	$R_1 = 8.45$ kΩ. Choose 8.2 kΩ as a standard value.	
	If 0.7 is ignored, then $R_1 = 9.1$ kΩ.	
11.	$R_2 = 11.7$ kΩ. Choose 12 KΩ as a standard value.	(problems 31–33)
	$R_1 = 141$ kΩ. Choose 140 or 150 kΩ as a standard value.	
12.	$R_2 = 21$ ohms. Choose 22 ohms as a standard value.	(problems 31–33)
	$R_1 = 273$ ohms. Choose 270 ohms as a standard value.	
13.	Use the circuit shown in Figure 4-16. Set the gate supply at a value slightly more negative than −3.8 V. A value of −4 V would work. Make resistor $R_G = 1$ MΩ. Set R_D at a value of (20 volts)/(10.5 mA), which calculates a resistance of 1.9 kΩ. You can wire a standard resistor of 1 kΩ in series with a standard resistor of 910 ohms to obtain a resistance of 1.91 kΩ.	(problems 39 and 41)

CHAPTER

5

AC Pre-Test and Review

You need to have some basic knowledge of alternating current (AC) to study electronics. To understand AC, you basically have to understand sine waves.

A *sine wave* is simply a shape, like waves in the ocean. Sine waves in electronics are used to represent voltage or current moving up and down in magnitude. In AC electronics, some signals or power sources (such as the house current provided at a wall plug) are represented by sine waves. The sine wave shows how the voltage moves from 0 volts to its peak voltage, and back down through 0, its negative peak voltage at 60 cycles per second, or 60 Hertz (Hz).

The sound from a musical instrument also consists of sine waves. When you combine sounds (such as all the instruments in an orchestra), you get complex combinations of many sine waves, at various frequencies.

The study of AC starts with the properties of simple sine waves, and continues with an examination of how electronic circuits can generate or change sine waves.

This chapter discusses the following:

- The generator
- The sine wave
- Peak-to-peak and RMS voltages
- Resistors in AC circuits
- Capacitive and inductive reactance
- Resonance

The Generator

1 In electronic circuits powered by direct current, the voltage source is usually a battery or solar cell, which produces a constant voltage and a constant current through a conductor.

In electronic circuits or devices powered by alternating current, the voltage source is usually a *generator*, which produces a regular output waveform, such as a sine wave.

Question

Draw one cycle of a sine wave.

Answer

See Figure 5-1.

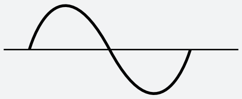

Figure 5-1

2 A number of electronic instruments are used in the laboratory to produce sine waves. For purposes of this discussion, the term *generator* means a sine wave source. These generators allow you to adjust the voltage and frequency by turning a dial or pushing a button. These instruments are called by various names, generally based on their method of producing the sine wave, or their application as a test instrument. The most popular generator at present is called a *function generator*. It actually provides a choice of functions or waveforms, including a square wave and a triangle wave. These waveforms are useful in testing certain electronic circuits.

The symbol shown in Figure 5-2 represents a generator. Note that a sine wave shown within a circle designates an AC sine wave source.

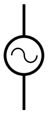

Figure 5-2

Questions

A. What is the most popular instrument used in the lab to produce waveforms?

B. What does the term AC mean?

____

C. What does the sine wave inside a generator symbol indicate?

____

Answers

A. The function generator.

B. Alternating current, as opposed to direct current.

C. The generator is a sine wave source.

3 Some key parameters of sine waves are indicated in Figure 5-3. The two axes are voltage and time.

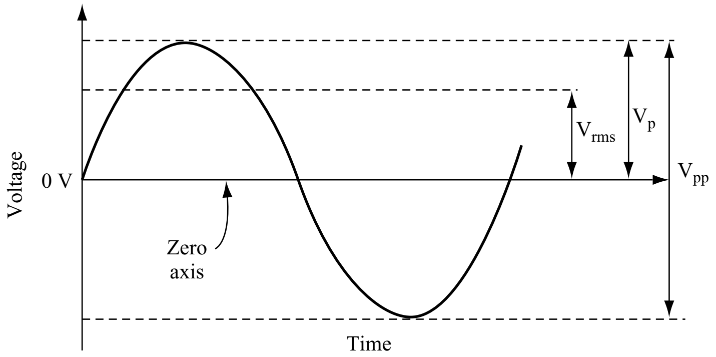

Figure 5-3

The *zero axis* is the reference point from which all voltage measurements are made.

Questions

A. What is the purpose of the zero axis?

____

B. What is the usual point for making time measurements?

____

Answers

A. It is the reference point from which all voltage measurements are made.

B. Time measurements can be made from any point in the sine wave, but usually they are made from a point at which the sine wave crosses the zero axis.

4 The three most important voltage or amplitude measurements are the *peak* (p), *peak-to-peak* (pp), and the *root mean square* (rms) voltages.

The following equations show the relationship between p, pp, and rms voltages for sine waves. The relationships between p, pp, and rms voltages differ for other waveforms (such as square waves).

$$V_p = \sqrt{2} \times V_{rms}$$

$$V_{pp} = 2V_p = 2 \times \sqrt{2} \times V_{rms}$$

$$V_{rms} = \frac{1}{\sqrt{2}} \times V_p = \frac{1}{\sqrt{2}} \times \frac{V_{pp}}{2}$$

Note the following:

$$\sqrt{2} = 1.414$$

$$\frac{1}{\sqrt{2}} = 0.707$$

Question

If the pp voltage of a sine wave is 10 volts, find the rms voltage.

Answer

$$V_{rms} = \frac{1}{\sqrt{2}} \times \frac{V_{pp}}{2} = 0.707 \times \frac{10}{2} = 3.535\,\text{V}$$

5 Calculate the following for a sine wave.

Question

If the rms voltage is 2 volts, find the pp voltage.

Answer

$$V_{pp} = 2 \times \sqrt{2} \times V_{rms} = 2 \times 1.414 \times 2 = 5.656\,\text{V}$$

6 Calculate the following for a sine wave.

Questions

A. $V_{pp} = 220$ volts. Find V_{rms}.

B. $V_{rms} = 120$ volts. Find V_{pp}.

Answers

A. 77.77 volts

B. 340 volts (This is the common house current supply voltage; 340 $V_{pp} = 120$ V_{rms}.)

7 There is a primary time measurement for sine waves. The duration of the complete sine wave is shown in Figure 5-4 and referred to as a *cycle*. All other time measurements are fractions or multiples of a cycle.

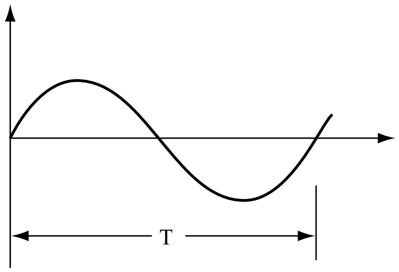

Figure 5-4

Questions

A. What is one complete sine wave called?

B. What do you call the time it takes to complete one sine wave?

C. How is the frequency of a sine wave related to this time?

D. What is the unit for frequency?

E. If the period of a sine wave is 0.5 ms, what is its frequency? What is the frequency of a sine wave with a period of 40 μsec?

F. If the frequency of a sine wave is 60 Hz, what is its period? What is the period of sine waves with frequencies of 12.5 kHz and 1 MHz?

Answers

A. A cycle

B. The period, T

C. $f = 1/T$

D. Hertz (Hz) is the standard unit for frequency. One Hertz equals one cycle per second.

E. 2 kHz, 25 kHz

F. 16.7 ms, 80 μsec, 1 μsec

8 Choose all answers that apply.

Question

Which of the following could represent electrical AC signals?

A. A simple sine wave

B. A mixture of many sine waves, of different frequencies and amplitudes

C. A straight line

Answers

A and B

Resistors in AC Circuits

9 Alternating current is passed through components, just as direct current is. Resistors interact with alternating current just as they do with direct current.

Question

Suppose an AC signal of 10 V_{pp} is connected across a 10 ohm resistor. What is the current through the resistor?

Answer

Use Ohm's law.

$$I = \frac{V}{R} = \frac{10\,V_{pp}}{10\,\text{ohms}} = 1\,A_{pp}$$

Because the voltage is given in pp, the current is a pp current.

10 An AC signal of 10 V_{rms} is connected across a 20 ohm resistor.

Question

Find the current.

Answer

$$I = \frac{10\,V_{rms}}{20\,\text{ohms}} = 0.5\,A_{rms}$$

Because the voltage was given in rms, the current is rms.

11 You apply an AC signal of 10 V_{pp} to the voltage divider circuit shown in Figure 5-5.

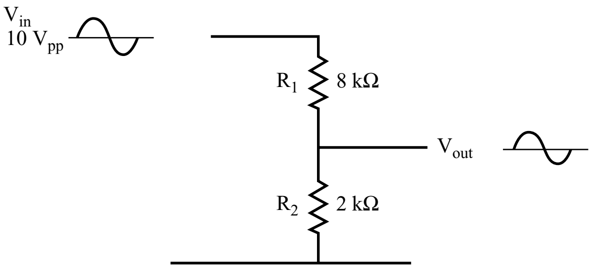

Figure 5-5

Question

Find V_{out}.

Answer

$$V_{out} = V_{in} \times \frac{R_2}{(R_1 + R_2)} = 10 \times \frac{2k\Omega}{(8k\Omega + 2k\Omega)} = 10 \times \frac{2}{10} = 2V_{pp}$$

Capacitors in AC Circuits

12 A capacitor opposes the flow of an AC current.

Questions

A. What is this opposition to the current flow called?

B. What is this similar to in DC circuits?

Answers

A. Reactance

B. Resistance

13 Just as with resistance, you determine reactance by using an equation.

Questions

A. What is the equation for reactance?

B. What does each symbol in the equation stand for?

C. How does the reactance of a capacitor change as the frequency of a signal increases?

Answers

A. $X_C = \dfrac{1}{2\pi fC}$

B. X_C = the reactance of the capacitor in ohms

f = the frequency of the signal in hertz.

C = the value of the capacitor in farads.

C. The reactance of a capacitor decreases as the frequency of the signal increases.

14 Assume the capacitance is 1 μF and the frequency is 1 kHz.

Question

Find the capacitor's reactance. (*Note*: $1/(2\pi) = 0.159$, approximately.)

Answer

$$X_C = \frac{1}{2\pi fC}$$

f = 1 kHz = 10^3 Hz

C = 1 μF = 10-6 F

Thus,

$$X_C = \frac{0.159}{10^3 \times 10^{-6}} = 160\,\text{ohms}$$

15 Now, perform these two simple calculations. In each case, find X_{C1}; the capacitor's reactance at 1 kHz, and X_{C2}; the capacitors reactance at the frequency specified in the question.

Questions

Find X_{C1} and X_{C2}:

A. C = 0.1 μF, f = 100 Hz.__________

B. C = 100 μF, f = 2 kHz.__________

Answers

A. At 1 kHz, $X_{C1} = 1600$ ohms; at 100 Hz, $X_{C2} = 16{,}000$ ohms

B. At 1 kHz, $X_{C1} = 1.6$ ohms; at 2 kHz, $X_{C2} = 0.8$ ohms

A circuit containing a capacitor in series with a resistor (as shown in Figure 5-6) functions as a voltage divider.

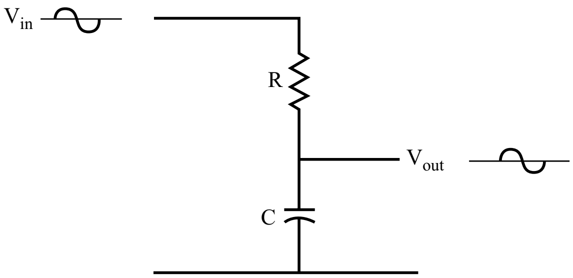

Figure 5-6

While this voltage divider provides a reduced output voltage, just like a voltage divider using two resistors, there's a complication. If you view the output and input voltage waveforms on an oscilloscope, you see that one is shifted away from the other. The two waveforms are said to be ''out of phase.'' *Phase* is an important concept in understanding how certain electronic circuits work. In Chapter 6, you learn about phase relationships for some AC circuits. You will also encounter this again when you study amplifiers.

The Inductor in an AC Circuit

16 An *inductor* is a coil of wire, usually wound many times around a piece of soft iron. In some cases, the wire is wound around a nonconducting material.

Questions

A. Is the AC reactance of an inductor high or low? Why?

____

B. Is the DC resistance high or low?

____

C. What is the relationship between the AC reactance and the DC resistance?

____

D. What is the formula for the reactance of an inductor?

____

Answers

A. Its AC reactance (X_L), which can be quite high, is a result of the electromagnetic field that surrounds the coil and induces a current in the opposite direction to the original current.

B. Its DC resistance (r), which is usually quite low, is simply the resistance of the wire that makes up the coil.

C. None.

D. $X_L = 2\pi fL$, where L = the value of the inductance in henrys. Using this equation, you can expect the reactance of an inductor to increase as the frequency of a signal passing through it increases.

17 Assume the inductance value is 10 henrys (H) and the frequency is 100 Hz.

Question

Find the reactance.

____

Answer

$X_L = 2\pi fL = 2\pi \times 100 \times 10 = 6280$ ohms

18 Now, try these two examples. In each case, find X_{L1}; the reactance of the inductor at 1 kHz and X_{L2}; the reactance at the frequency given in the question.

Questions

A. L = 1 mH (0.001 H), f = 10 kHz

____

B. L = 0.01 mH, f = 5 MHz

Answers

A. $X_{L1} = 6.28 \times 10^3 \times 0.001 = 6.28$ ohms

$X_{L2} = 6.28 \times 10 \times 10^3 \times 0.001 = 62.8$ ohms

B. $X_{L1} = 6.28 \times 10^3 \times 0.01 \times 10^{-3} = 0.0628$ ohms

$X_{L2} = 6.28 \times 5 \times 10^6 \times 0.01 \times 10^{-3} = 314$ ohms

A circuit containing an inductor in series with a resistor functions as a voltage divider, just as a circuit containing a capacitor in series with a resistor does. Again, the relationship between the input and output voltages is not as simple as a resistive divider. The circuit is discussed in Chapter 6.

Resonance

19 Calculations in previous problems demonstrate that capacitive reactance decreases as frequency increases, and that inductive reactance increases as frequency increases. If a capacitor and an inductor are connected in series, there will be one frequency at which their reactance values are equal.

Questions

A. What is this frequency called?

B. What is the formula for calculating this frequency? You can find it by setting $X_L = X_C$ and solving for frequency.

Answers

A. The resonant frequency

B. $2\pi fL = 1/(2\pi fC)$. Rearranging the terms in this equation to solve for f yields the following formula for the resonant frequency (f_r):

$$f_r = \frac{1}{2\pi\sqrt{LC}}$$

20 If a capacitor and an inductor are connected in parallel, there will also be a resonant frequency. Analysis of a parallel resonant circuit is not as simple as for a series resonant circuit. The reason for this is that inductors always have some internal resistance, which complicates some of the equations. However, under certain conditions, the analysis is similar. For example, if the reactance of the inductor in ohms is more than 10 times greater than its own internal resistance (r), the formula for the resonant frequency is the same as if the inductor and capacitor were connected in series. This is an approximation that you will use often.

Questions

For the following inductors, determine if the reactance is more or less than 10 times its internal resistance. A resonant frequency is provided.

A. $f_r = 25$ kHz, $L = 2$ mH, $r = 20$ ohms

B. $f_r = 1$ kHz, $L = 33.5$ mH, $r = 30$ ohms

Answers

A. $X_L = 314$ ohms, which is more than 10 times greater than r.

B. $X_L = 210$ ohms, which is less than 10 times greater than r.

NOTE **Chapter 7 discusses both series and parallel resonant circuits. At that time, you'll learn many useful techniques and formulas.**

21 Find the resonant frequency (f_r) for the following capacitors and inductors when they are connected both in parallel and in series. Assume r is negligible.

Questions

Determine f_r:

A. $C = 1$ μF, $L = 1$ henry

B. $C = 0.2$ μF, $L = 3.3$ mH

Answers

A. $f_r = \dfrac{0.159}{\sqrt{10^{-6} \times 1}} = 160\,\text{Hz}$

B. $f_r = \dfrac{0.159}{\sqrt{3.3 \times 10^{-3} \times 0.2 \times 10^{-6}}} = 6.2\,\text{kHz}$

22 Now, try these two final examples.

Questions

Determine f_r:

A. C = 10 μF, L = 1 henry

____

B. C = 0.0033 μF, L = 0.5 mH

____

Answers

A. f_r = 50 Hz (approximately)

B. f_r = 124 kHz

Understanding resonance is important to understanding certain electronic circuits, such as filters and oscillators.

Filters are electronic circuits that either block a certain band of frequencies, or pass a certain band of frequencies. One common use of filters is in circuits used for radio, TV, and other communications applications. *Oscillators* are electronic circuits that generate a continuous output without an input signal. The type of oscillator that uses a resonant circuit produces pure sine waves. (You learn more about oscillators in Chapter 9.)

Summary

Following are the concepts presented in this chapter:

- The sine wave is used extensively in AC circuits.
- The most common laboratory generator is the function generator.
- $V_p = \sqrt{2} \times V_{rms}$, $V_{pp} = 2\sqrt{2} \times V_{rms}$

- $f = 1/T$
- $I_{pp} = \frac{V_{pp}}{R}, \; I_{rms} = \frac{V_{rms}}{R}$
- Capacitive reactance is calculated as follows:

  $$X_C = \frac{1}{(2\pi fC)}$$

- Inductive reactance is calculated as follows:

  $$X_L = 2\pi fL$$

- Resonant frequency is calculated as follows:

  $$f_r = \frac{1}{2\pi\sqrt{LC}}$$

Self-Test

The following problems test your understanding of the basic concepts presented in this chapter. Use a separate sheet of paper for calculations if necessary. Compare your answers with the answers provided following the test.

1. Convert the following peak or peak-to-peak values to rms values:

A. $V_p = 12$ V, $V_{rms} =$ ____________________

B. $V_p = 80$ mV, $V_{rms} =$ ____________________

C. $V_{pp} = 100$ V, $V_{rms} =$ ____________________

2. Convert the following rms values to the required values shown:

A. $V_{rms} = 120$ V, $V_p =$ ____________________

B. $V_{rms} = 100$ mV, $V_p =$ ____________________

C. $V_{rms} = 12$ V, $V_{pp} =$ ____________________

3. For the given value, find the period or frequency:

A. $T = 16.7$ ms, $f =$ ____________________

B. $f = 15$ kHz, $T =$ ____________________

4. For the circuit shown in Figure 5-7, find the total current flow and the voltage across R_2, (V_{out}).

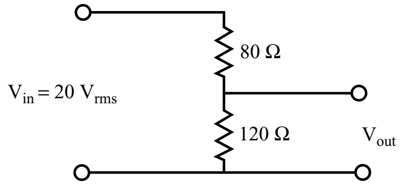

Figure 5-7

5. Find the reactance of the following components:

A. $C = 0.16\ \mu F$, $f = 12\ kHz$, $X_C =$ ____________________

B. $L = 5\ mH$, $f = 30\ kHz$, $X_L =$ ____________________

6. Find the frequency necessary to cause each reactance shown:

A. $C = 1\ \mu F$, $X_C = 200$ ohms, $f =$ ____________________

B. $L = 50\ \mu F$, $X_L = 320$ ohms, $f =$ ____________________

7. What would be the resonant frequency for the capacitor and inductor values given in A and B of question 5 if they were connected in series?

____

8. What would be the resonant frequency for the capacitor and inductor values given in A and B of question 6 if they were connected in parallel? What assumption would you have to make?

____

Answers to Self-Test

If your answers do not agree with those provided here, review the problems indicated in parentheses before you go on to the next chapter.

1.	A.	$8.5\ V_{rms}$	(problems 4–6)
	B.	$56.6\ V_{rms}$	
	C.	$35.4\ V_{rms}$	
2.	A.	$169.7\ V_p$	(problems 4–6)
	B.	$141.4\ mV_p$	
	C.	$33.9\ V_{pp}$	
3.	A.	60 Hz	(problem 7)
	B.	66.7 µsec	
4.	$I_T = 0.1\ A_{rms}$, $V_{out} = 12\ V_{rms}$		(problems 9–11)

5.	A.	82.9 ohms	(problems 14 and 17)
	B.	942.5 ohms	
6.	A.	795.8 Hz	(problems 14 and 17)
	B.	1.02 kHz	
7.	5.63 kHz		(problem 19)
8.	711.8 Hz. Assume the internal resistance of the inductor is negligible.		(problem 20)

CHAPTER 6

AC in Electronics

Certain types of circuits are found in most electronic devices used to process alternating current (AC) signals. One of the most common of these, *filter circuits*, is covered in this chapter. Filter circuits are formed by resistors and capacitors (RC), or resistors and inductors (RL). These circuits (and their effect on AC signals) play a major part in communications, consumer electronics, and industrial controls.

When you complete this chapter you will be able to do the following:

- Calculate the output voltage of an AC signal after it passes through a high pass RC filter circuit.
- Calculate the output voltage of an AC signal after it passes through a low pass RC circuit.
- Calculate the output voltage of an AC signal after it passes through a high pass RL circuit.
- Calculate the output voltage of an AC signal after it passes through a low pass RL circuit.
- Draw the output waveform of an AC or combined AC-DC signal after it passes through a filter circuit.
- Calculate simple phase angles and phase differences.

Capacitors in AC Circuits

1 An AC signal is continually changing, whether it is a pure sine wave or a complex signal made up of many sine waves. If such a signal is applied to one plate of a capacitor, it will be induced on the other plate. To express this another way, a capacitor will ''pass'' an AC signal as illustrated in Figure 6-1.

NOTE Unlike an AC signal, a DC signal is blocked by a capacitor. Equally important is the fact that a capacitor is *not* a short circuit to an AC signal.

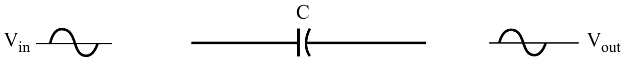

Figure 6-1

Questions

A. What is the main difference in the effect of a capacitor upon an AC signal versus a DC signal?

____

B. Does a capacitor appear as a short or an open circuit to an AC signal?

____

Answers

A. A capacitor will pass an AC signal, while it will not pass a DC voltage level.

B. Neither

2 A capacitor will, in general, oppose the flow of an AC current to some degree. This opposition to current flow, as you saw in Chapter 5, is called the *reactance* of the capacitor.

Reactance is similar to resistance, except that the reactance of a capacitor changes when you vary the frequency of a signal. The reactance of a capacitor can be calculated by a formula introduced in Chapter 5.

Question

Write the formula for the reactance of a capacitor.

____

Answer

$$X_C = \frac{1}{2\pi fC}$$

3 From this formula, you can see that the reactance changes when the frequency of the input signal changes.

Question

If the frequency increases, what happens to the reactance?

Answer

It decreases.

If you had difficulty with these first three problems, you should review the examples in Chapter 5.

Capacitors and Resistors in Series

4 For simplicity, consider all inputs at this time to be pure sine waves. The circuit shown in Figure 6-2 shows a sine wave as the input signal to a capacitor.

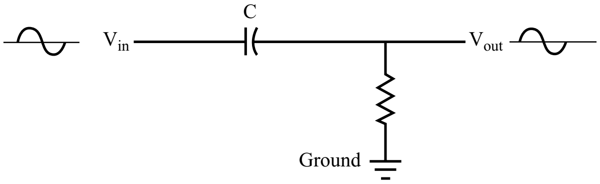

Figure 6-2

Question

If the input is a pure sine wave, what is the output?

Answer

A pure sine wave

5 The output sine wave has the same frequency as the input sine wave. A capacitor cannot change the frequency of the signal. But, remember, with an AC input, the capacitor behaves in a manner similar to a resistor in that the capacitor does have some level of opposition to the flow of alternating current. The level of opposition depends upon the value of the capacitor and the frequency of the signal. Therefore, the output amplitude of a sine wave will be less than the input amplitude.

Question

With an AC input to a simple circuit like the one described here, what does the capacitor appear to behave like?

Answer

It appears to have opposition to alternating current similar to the behavior of a resistor.

6 If you connect a capacitor and resistor in series (as shown in Figure 6-3), the circuit functions as a voltage divider.

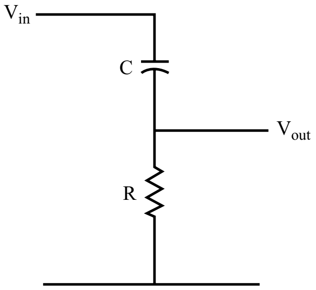

Figure 6-3

Question

What formula would you use to calculate the output voltage for a voltage divider formed by connecting two resistors in series?

Answer

$$V_{out} = V_{in} \times \frac{R_2}{R_1 + R_2}$$

7 You can calculate a total resistance to the flow of electric current for a circuit containing two resistors in series.

Question

What is the formula for this total resistance?

Answer

$R_T = R_1 + R_2$

---

8 You can also calculate the total opposition to the flow of electric current for a circuit containing a capacitor and resistor in series. This parameter is called *impedance*, and you can calculate it using the following formula:

$$Z = \sqrt{X_C^2 + R^2}$$

where:

Z = the impedance of the circuit in ohms

X_C = the reactance of the capacitor in ohms

R = the resistance of the resistor in ohms

Questions

Use the following steps to calculate the impedance of the circuit, and the current flowing through the circuit shown in Figure 6-4.

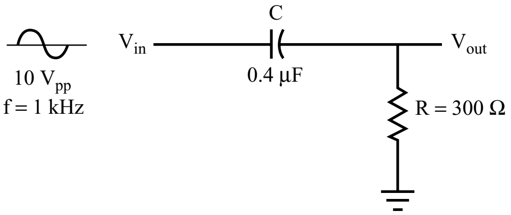

Figure 6-4

A.

$$X_C = \frac{1}{2\pi fC} =$$

B.

$$Z = \sqrt{X_C^2 + R^2} =$$

C.

$$I = \frac{V}{Z} =$$

Answers

A. 400 ohms

B. 500 ohms

C. 20 mA_{pp}

9 Now, for the circuit shown in Figure 6-4, calculate the impedance and current using the values provided.

Questions

A. C = 530 μF, R = 12 ohms, V_{in} = 26 V_{pp}, f = 60 Hz

B. C = 1.77 μF, R = 12 ohms, V_{in} = 150 V_{pp}, f = 10 kHz

Answers

A. Z = 13 ohms, I = 2 A_{pp}

B. Z = 15 ohms, I = 10 A_{pp}

10 You can calculate V_{out} for the circuit shown in Figure 6-5 with a formula similar to the formula used in Chapter 5 to calculate V_{out} for a voltage divider composed of two resistors.

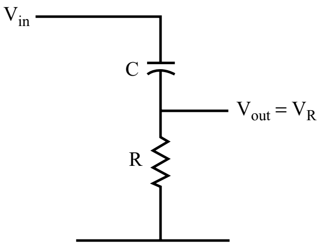

Figure 6-5

The formula to calculate the output voltage for this circuit is as follows:

$$V_{out} = V_{in} \times \frac{R}{Z}$$

Questions

Calculate the output voltage in this circuit using the component values and input signal voltage and frequency listed on the circuit diagram shown in Figure 6-6.

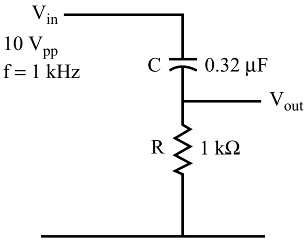

Figure 6-6

A. Find X_C:

B. Find Z:

C. Use the formula to find V_{out}:

Answers

A. $X_C = 500$ ohms (rounded off)

B. $Z = 1120$ ohms (rounded off)

C. $V_{out} = 8.9\ V_{pp}$

11 Now, find V_{out} for the circuit in Figure 6-5 using the given component values, signal voltage, and frequency.

Questions

A. $C = 0.16\ \mu F$, $R = 1\ k\Omega$, $V_{in} = 10\ V_{pp}$, $f = 1\ kHz$

____

B. $C = 0.08\ \mu F$, $R = 1\ k\Omega$, $V_{in} = 10\ V_{pp}$, $f = 1\ kHz$

____

Answers

A. $V_{out} = 7.1\ V_{pp}$

B. $V_{out} = 4.5\ V_{pp}$

NOTE **Hereafter, you can assume that the answer is a peak-to-peak value if the given value is a peak-to-peak value.**

12 The output voltage is said to be *attenuated* in the voltage divider calculations as shown in the calculations in problems 10 and 11. Compare the input and output voltages in problems 10 and 11.

Question

What does *attenuated* mean?

____

Answer

To reduce in amplitude or magnitude (that is, V_{out} is smaller than V_{in}).

13 When you calculated V_{out} in the examples in problems 10 and 11, you first had to find X_C. However, X_C changes as the frequency changes, while the resistance remains constant. Therefore, as the frequency changes, the impedance Z changes, and so also does the amplitude of the output voltage V_{out}.

If V_{out} is plotted against frequency on a graph, the curve looks like that shown in Figure 6-7.

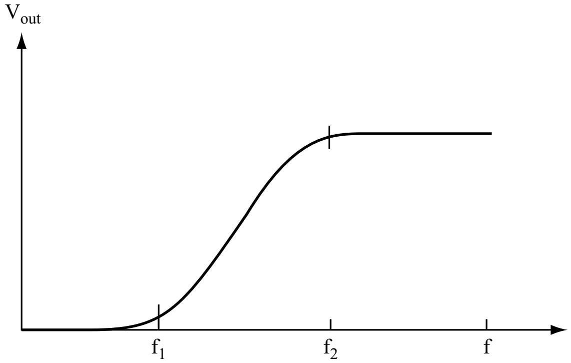

Figure 6-7

The frequencies of f_1 (at which the curve starts to rise) and f_2 (where it starts to level off) depend on the values of the capacitor and the resistor.

Questions

Calculate the output voltage for the circuit shown in Figure 6-8 for frequencies of 100 Hz, 1 kHz, 10 kHz, and 100 kHz.

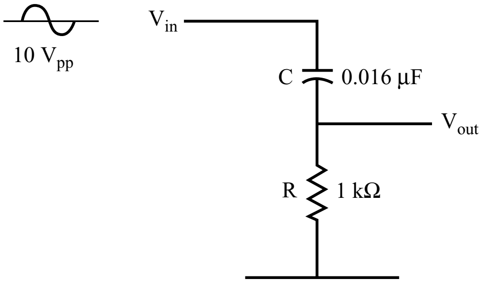

Figure 6-8

A. 100 Hz:

B. 1 kHz:

C. 10 kHz:

D. 100 kHz:

E. Plot these values for V_{out} against f and draw a curve to fit the points. Use a separate sheet of paper to draw your graph.

Answers

A. $V_{out} = 0.1$ V

B. $V_{out} = 1$ V

C. $V_{out} = 7.1$ V

D. $V_{out} = 10$ V

E. The curve is shown in Figure 6-9.

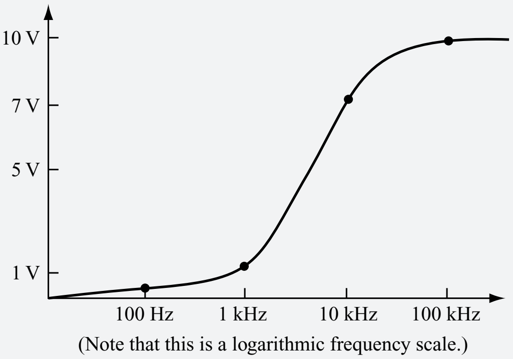

Figure 6-9

NOTE You can see that V_{out} is equal to V_{in} for the highest frequency, and at nearly zero for the lowest frequency. You call this type of circuit a *high pass circuit* because it will pass high frequency signals with little attenuation and block low frequency signals.

The High Pass Filter Experiment

14 The objective of the following experiment is to determine how V_{out} changes as the frequency of the input signal changes for the high pass filter circuit shown in Figure 6-8. You will also calculate X and Z for each frequency value to show the relationship between the output voltage and the impedance.

You will need the following equipment and supplies:

- One 1 kΩ resistor.
- One 0.016 μF capacitor.
- One function generator.
- One oscilloscope. (You can substitute a multimeter and measure V_{out} in rms voltage rather than peak-to-peak voltage.)
- One breadboard.

If you do not have what you need to set up the circuit and measure the values, just read through the experiment.

Follow these steps to complete the experiment:

1. Set up the circuit shown in Figure 6-8 on the breadboard.
2. Set the function generator to generate a 10 V_{pp}, 25 Hz sine wave to be used for V_{in}, measure V_{out} and record the reading in the following table.

f_{in}	X_C	Z	V_{OUT}
25 Hz			
50 Hz			
100 Hz			
250 Hz			
500 Hz			
1 kHz			
3 kHz			
5 kHz			
7 kHz			
10 kHz			
20 kHz			
30 kHz			
50 kHz			
100 kHz			

3. Increase the frequency of V_{in} to the value of f_{in} shown in the next row of the preceding table, measure V_{out} and record the reading in the table. Repeat this until you have entered the value of V_{out} in the last row.

4. Calculate the values of X_C and Z for each row and enter them in the preceding table.
5. Plot V_{out} vs f_{in} with the voltage on the vertical axis and the frequency on the X axis (use graph paper with a log scale for the X axis). The curve should have the same shape as the curve in Figure 6-9, but don't worry if your curve is shifted slightly to the right or left.

Question

What would cause the curve to be moved slightly to the right or the left?

Answer

You may have used a resistor and capacitor with slightly different values than those used to produce the curve shown in Figure 6-10. Such variation in resistor and capacitor values are to be expected, given the manufacturing tolerance allowed for standard components.

Your values should be close to those shown in the following table, and the curve should be very similar to Figure 6-10.

f_{in}	X_C	Z	V_{OUT}
25 Hz	400 kΩ	400 kΩ	0.025 V
50 Hz	200 kΩ	200 kΩ	0.05 V
100 Hz	100 kΩ	100 kΩ	0.1 V
250 Hz	40 kΩ	40 kΩ	0.25 V
500 Hz	20 kΩ	20 kΩ	0.5 V
1 kHz	10 kΩ	10 kΩ	1 V
3 kHz	3.3 kΩ	3.5 kΩ	2.9 V
5 kHz	2 kΩ	2.2 kΩ	4.47 V
7 kHz	1.4 kΩ	1.7 kΩ	5.8 V
10 kHz	1 kΩ	1.414 kΩ	7.1 V
20 kHz	500 Ω	1.12 kΩ	8.9 V
30 kHz	330 Ω	1.05 kΩ	9.5 V
50 kHz	200 Ω	1.02 kΩ	9.8 V
100 kHz	100 Ω	1 kΩ	10 V

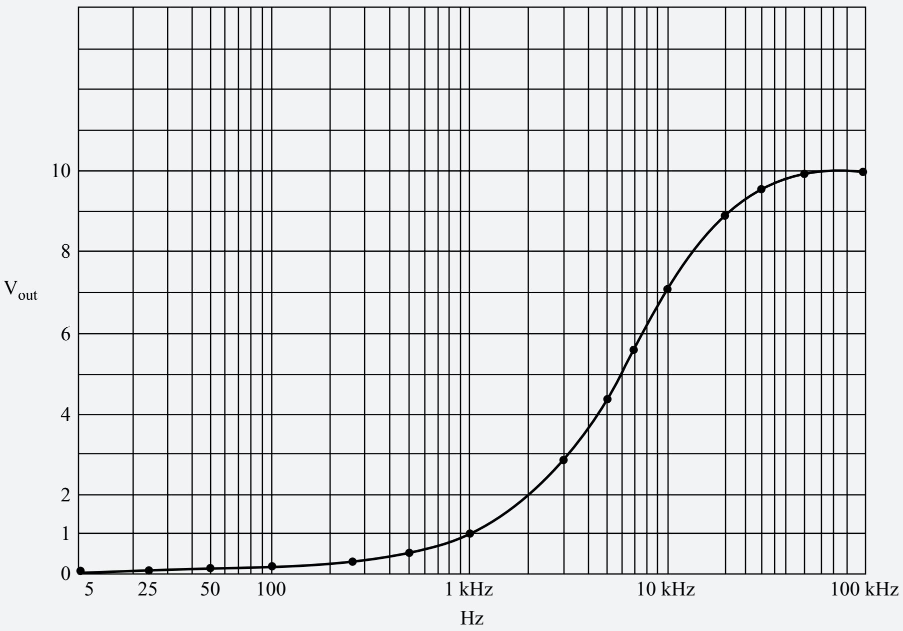

Figure 6-10

Question

What would cause your curve to be moved slightly to the right or the left of the curve shown in Figure 6-10?

Answer

Slightly different values for the resistor and capacitor that you used versus the resistor and capacitor used to produce the curve in Figure 6-10. Variations in resistor and capacitor values are to be expected, given the tolerance allowed for standard components.

15 The circuit shown in Figure 6-11 is used in many electronic devices.

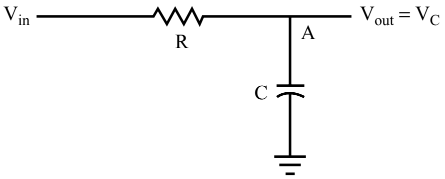

Figure 6-11

For this circuit, you measure the output voltage across the capacitor, instead of across the resistor (between point A and ground).

The impedance of this circuit is the same as that of the circuit used in the last few problems. It will still behave like a voltage divider, and you can calculate the output voltage with an equation that is similar to the one you used for the high pass filter circuit discussed in the last few problems. However, by switching the positions of the resistor and capacitor to create the circuit shown in Figure 6-11, you switch which frequencies will be attenuated and which will not be attenuated, making the new circuit a low pass filter whose characteristics you will explore in the next few problems.

Questions

A. What is the impedance formula for the circuit?

____

B. What is the formula for the output voltage?

____

Answers

A. $Z = \sqrt{X_C^2 + R^2}$

B. $V_{out} = V_{in} \times \frac{X_C}{Z}$

16 Refer to the circuit shown in Figure 6-11 and the following values:

$V_{in} = 10\ V_{pp}, f = 2\,kHz$

$C = 0.1\ \mu F,\ R = 1\,k\Omega$

Questions

Find the following:

A. X_C:

____

B. Z:

C. V_{out}:

Answers

A. 795 ohms

B. 1277 ohms

C. 6.24 V

17 Once again refer to the circuit shown in Figure 6-11 to answer the following question.

Question

Calculate the voltage across the resistor using the values given in problem 16, along with the calculated impedance value.

Answer

$$V_R = V_{in} \times \frac{R}{Z} = 10 \times \frac{1000}{1277} = 7.83V_{pp}$$

18 Use the information from problems 16 and 17 to answer the following question.

Question

What is the formula to calculate V_{in} using the voltages across the capacitor and the resistor?

Answer

The formula is $V_{in}^2 = V_C^2 + V_R^2$.

19 V_{out} of the circuit shown in Figure 6-11 changes as the frequency of the input signal changes. Figure 6-12 shows the graph of V_{out} versus frequency for this circuit.

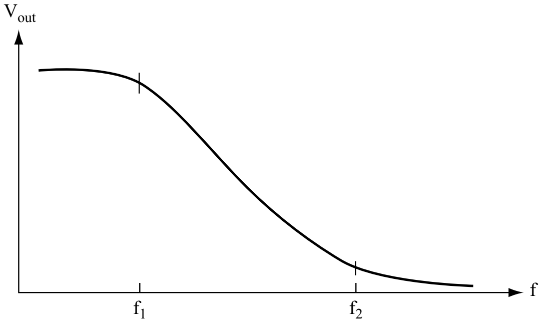

Figure 6-12

Question

What parameters determine f_1 and f_2?

Answer

The values of the capacitor and the resistor

NOTE You can see in Figure 6-12 that V_{out} is large for the lowest frequency, and nearly zero for the highest frequency. You call this type of circuit a *low pass circuit* because it will pass low frequency signals with little attenuation while blocking high frequency signals.

Phase Shift of an RC Circuit

20 In both of the circuits shown in Figure 6-13, the output voltage is different from the input voltage.

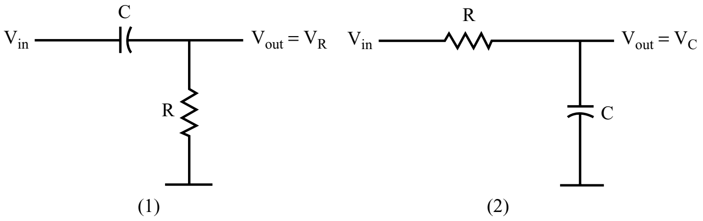

Figure 6-13

Question

In what ways do they differ?

Answer

The signal is attenuated, or reduced. The amount of attenuation depends upon the frequency of the signal.

21 The voltage is also changed in another way. The voltage across a capacitor rises and falls at the same frequency as the input signal, but it does not reach its peak at the same time, nor does it pass through zero at the same time. You can see this when you compare the V_{out} curves to the V_{in} curves in Figure 6-14.

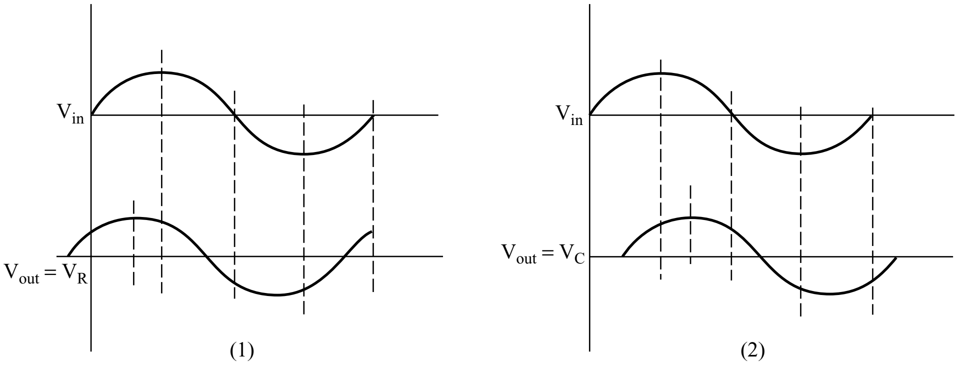

Figure 6-14

NOTE **The numbered graphs in Figure 6-14 are produced by the corresponding numbered circuits in Figure 6-13.**

Questions

A. Examine graph (1). Is the output voltage peak displaced to the right or the left?

B. Examine graph (2). Is the output voltage peak displaced to the right or the left?

Answers

A. To the left

B. To the right

22 The output voltage waveform in graph (1) of Figure 6-14 is said to *lead the input voltage waveform*. The output waveform in graph (2) is said to *lag the input waveform*. The amount that V_{out} leads or lags V_{in} is measured in degrees. There are 90 degrees between the peak of a sine wave and a point at which the sine wave crosses zero volts. You can use this information to estimate the number of degrees V_{out} is leading or lagging V_{in}. The difference between these two waveforms is called a *phase shift* or *phase difference*.

Questions

A. What is the approximate phase shift of the two waveforms shown in the graphs?

____

B. Do you think that the phase shift depends on the value of frequency?

____

C. Will an RC voltage divider with the voltage taken across the capacitor produce a lead or a lag in the phase shift of the output voltage?

____

Answers

A. Approximately 35 degrees.

B. It does depend upon frequency because the values of the reactance and impedance depend upon frequency.

C. A lag as shown in graph (2).

23 The current through a capacitor is out of phase with the voltage across the capacitor. The current leads the voltage by 90 degrees. The current and voltage across a resistor are in phase, that is, they have no phase difference.

Figure 6-15 shows the vector diagram for a series RC circuit. θ is the phase angle by which V_R leads V_{in}. ϕ is the phase angle by which V_C lags V_{in}.

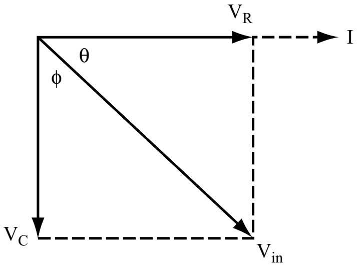

Figure 6-15

NOTE Although the voltage across a resistor is in phase with the current through the resistor, both are out of phase with the applied voltage.

You can calculate the phase angle from this formula:

$$\tan\theta = \frac{V_C}{V_R} = \frac{1}{2\pi fRC} = \frac{X_C}{R}$$

As an example, calculate the phase angle when 160 Hz is applied to a 3.9 kΩ resistor in series with a 0.1 μF capacitor.

$$\tan\theta = \frac{1}{2 \times \pi \times 160 \times 3.9 \times 10^3 \times 0.1 \times 10^{-6}} = 2.564$$

You can calculate the inverse tangent of 2.564 on your calculator and find that the phase angle is 68.7 degrees, which means that V_R leads V_{in} by 68.7 degrees. This also means that V_C lags the input by 21.3 degrees.

In electronics, the diagram shown in Figure 6-15 is called a *phasor diagram*, but the mathematics involved is the same as for vector diagrams, with which you should be familiar.

Question

Sketch a phasor diagram using the angles θ and ϕ resulting from the calculations in this problem. Use a separate sheet of paper for your diagram.

Answer

See Figure 6-16. Note that the phasor diagram shows that the magnitude of V_C is greater than V_R.

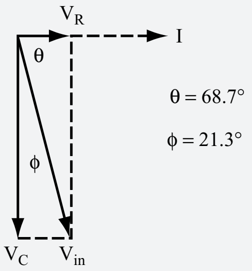
V_R
I
θ
θ = 68.7°
ϕ = 21.3°
ϕ
V_C
V_{in}

Figure 6-16

24 Using the component values and input signal shown in Figure 6-17, find the following.

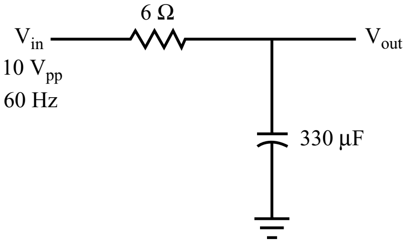
6 Ω
V_{in}
10 V_{pp}
60 Hz
V_{out}
330 μF

Figure 6-17

Questions

A. X_C:

____

B. Z:

____

C. V_{out}:

____

D. V_R:

E. The current flowing through the circuit:

F. The phase angle:

Answers

A. $X_C = \frac{1}{2\pi fC} = 8\,\text{ohms}$

B. $Z = \sqrt{8^2 + 6^2} = 10\,\text{ohms}$

C. $V_{out} = V_C = V_{in} \times \frac{X_C}{Z} = 8\,\text{V}$

D. $V_R = V_{in} \times \frac{R}{Z} = 6\,\text{V}$

E. $I = \frac{V}{Z} = \frac{10\,V_{pp}}{10\,\Omega} = 1\,\text{amp}$

F. $\tan\theta = \frac{X_C}{R} = \frac{8\,\Omega}{6\,\Omega} = 1.33.$
Therefore, $\theta = 53.13$ degrees.

25 For the circuit shown in Figure 6-18, calculate the following parameters.

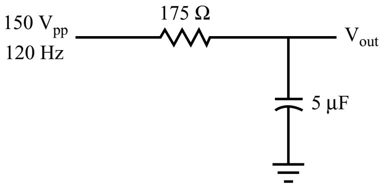

Figure 6-18

Questions

A. X_C:

B. Z:

C. V_{out}:

D. V_R:

E. The current flowing through the circuit:

F. The phase angle:

Answers

A. $X_C = 265$ ohms

B. $Z = \sqrt{175^2 + 265^2} = 317.57\,\Omega$

C. $V_C = 125$ V

D. $V_R = 83$ V

E. $I = 0.472$ A

F. $\tan\theta = \dfrac{265\,\Omega}{175\,\Omega} = 1.5$

Therefore, $\theta = 56.56$ degrees.

Resistor and Capacitor in Parallel

26 The circuit in Figure 6-19 is a common variation on the low pass filter circuit introduced in problem 15.

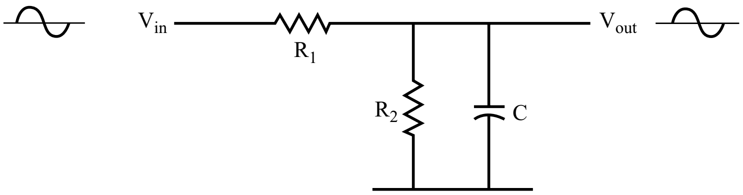

Figure 6-19

Because a DC signal will not pass through the capacitor, this circuit functions like the circuit shown in Figure 6-20 for DC input signals.

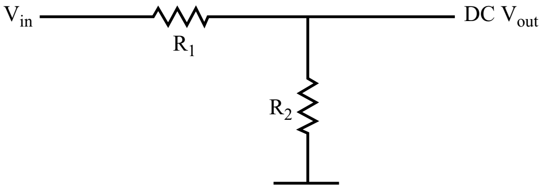

Figure 6-20

An AC signal will pass through both the capacitor and R_2. You can treat the circuit as if it had a resistor with a value of r, where r is the parallel equivalent of R_2 and X_C, in place of the parallel capacitor and resistor. This is shown in Figure 6-21.

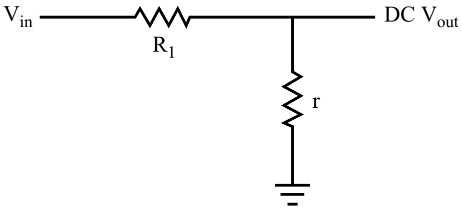

Figure 6-21

Calculating the exact parallel equivalent (r) is very complicated and beyond the scope of this book. However, to demonstrate the usefulness of this circuit let's make a major simplification. Consider a circuit where X_C is only about one tenth the value of R2 or less. This circuit has many practical applications, because it will attenuate the AC and the DC differently.

The following example will help to clarify this. For the following circuit, calculate the AC and DC output voltages separately.

For the circuit shown in Figure 6-22, you can calculate the AC and DC output voltages separately by the steps outlined in the following questions.

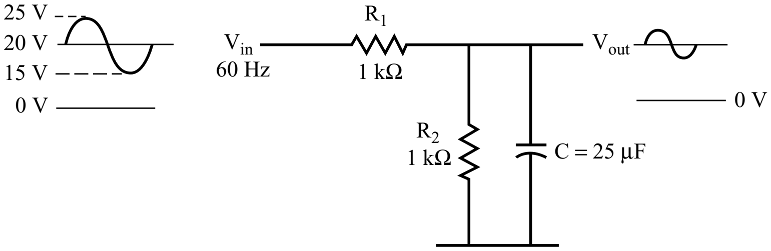

Figure 6-22

Questions

A. Find X_C. Check that it is less than one tenth of R_2.

$X_C =$ ______________________________

B. For the circuit in Figure 6-22, determine which circuit components DC signals will flow through. Then use the voltage divider formula to find DC V_{out}.

DC $V_{out} =$ ______________________________

C. For the circuit in Figure 6-22 determine which circuit components AC signals will flow through. Then use the voltage divider formula to find AC V_{out}.

AC $V_{out} =$ ______________________________

D. Compare the AC and DC input and output voltages.

____

Answers

A. $X_C = 106$ ohms and $R_2 = 1000$ ohms, so X_C is close enough to one tenth of R_2.

B. The portion of the circuit that a DC signal passes through is shown in Figure 6-23.

$$V_{out} = 20 \times \frac{1\,\text{k}\Omega}{1\,\text{k}\Omega + 1\,\text{k}\Omega} = 10\,\text{V}$$

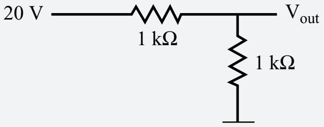

Figure 6-23

C. The portion of the circuit that an AC signal passes through is shown in Figure 6-24.

$$V_{out} = 10 \times \frac{106}{\sqrt{(1000)^2 + (106)^2}} = 1.05\,V$$

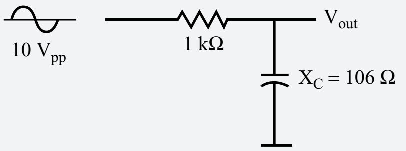

Figure 6-24

D. Figure 6-25 shows the input waveform on the left and the output waveform on the right. You can see from the waveforms that the DC voltage has dropped from 20 V to 10 V, and that the AC voltage has dropped from 10 V to 1.05 V.

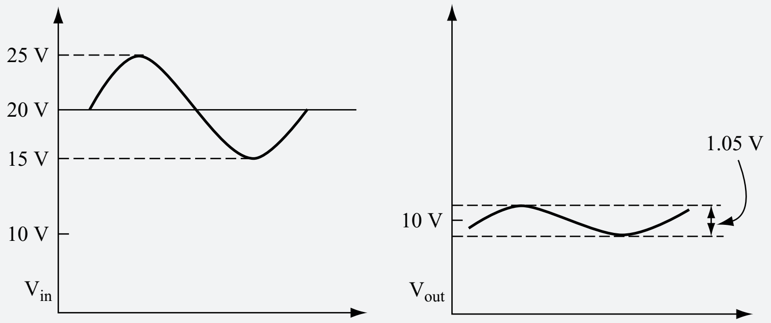

Figure 6-25

27 Figure 6-26 shows two versions of the circuit discussed in problem 26 with changes to the value of the capacitor or the frequency of the input signal. The input voltage in both cases is 20 V_{pp}. Use the same steps shown in problem 26 to find and compare the output voltages with the input voltages for the two circuits shown in Figure 6-26.

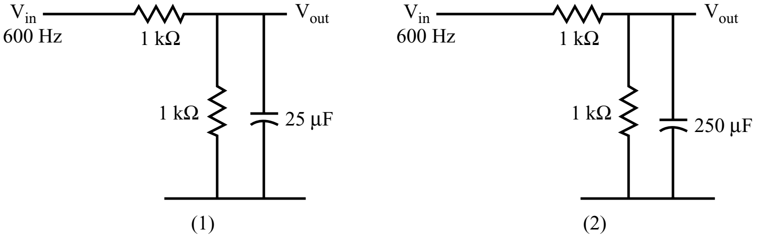

(1) (2)

Figure 6-26

Questions

1. **A.** $X_C =$ ________________________________
   B. DC $V_{out} =$ ________________________________
   C. AC $V_{out} =$ ________________________________
   D. Attenuation: ________________________________
2. **A.** $X_C =$ ________________________________
   B. DC $V_{out} =$ ________________________________
   C. AC $V_{out} =$ ________________________________
   D. Attenuation: ________________________________

Answers

1. **A.** $X_C = 10.6$ ohms.
   B. DC $V_{out} = 10$ V.
   C. AC $V_{out} = 0.1$ V.
   D. Here, the DC attenuation is the same as the example in problem 26, but the AC output voltage is reduced because of the higher frequency.
2. **A.** $X_C = 10.6$ ohms.
   B. DC $V_{out} = 10$ V.
   C. AC $V_{out} = 0.1$ V.
   D. The DC attenuation is still the same, but the AC output voltage is reduced because of the larger capacitor.

Inductors in AC Circuits

28 Figure 6-27 shows a voltage divider circuit using an inductor, rather than a capacitor.

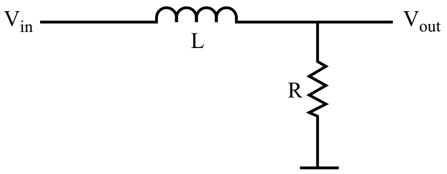

Figure 6-27

As with previous problems, consider all the inputs to be pure sine waves. Like the capacitor, the inductor cannot change the frequency of a sine wave, but it can reduce the amplitude of the output voltage.

The simple circuit, as shown in Figure 6-27, opposes current flow.

Questions

A. What is the opposition to current flow called?

____

B. What is the formula for the reactance of the inductor?

____

C. Write out the formula for the opposition to the current flow for this circuit.

____

Answers

A. Impedance.

B. $X_L = 2\pi fL$.

C. $Z = \sqrt{X_L^2 + R^2}$

In many cases, the DC resistance of the inductor is very low, so assume that it is 0 ohms. For the next two problems make that assumption in performing your calculations.

29 You can calculate the voltage output for the circuit shown in Figure 6-28 with the voltage divider formula.

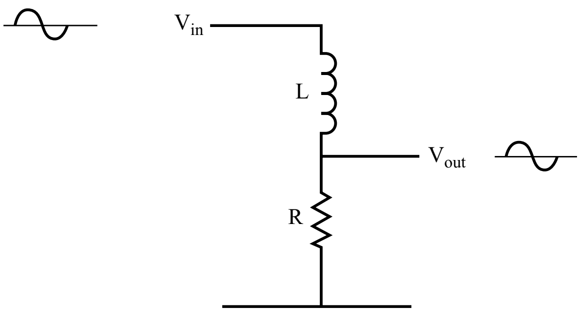

Figure 6-28

Question

What is the formula for V_{out}?

Answer

$$V_{out} = V_{in} \times \frac{R}{Z}$$

30 Find the output voltage for the circuit shown in Figure 6-29.

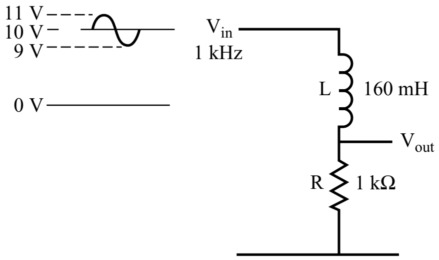

Figure 6-29

Use the steps in the following questions to perform the calculation.

Questions

A. Find the DC output voltage. Use the DC voltage divider formula.

DC V_{out} = ________________

B. Find the reactance of the inductor.

X_L = ________________

C. Find the AC impedance.

Z = ________________

D. Find the AC output voltage.

AC V_{out} = ________________

E. Combine the outputs to find the actual output. Draw the output waveform and label the voltage levels of the waveform on the blank graph in Figure 6-30.

Figure 6-30

Answers

A. DC $V_{out} = 10\,V \times \dfrac{1\,k\Omega}{1\,k\Omega + 0} = 10\,V$

B. $X_L = 1$ kΩ (approximately).

C. $Z = \sqrt{1^2 + 1^2} = \sqrt{2} = 1.414\,k\Omega$

D. AC $V_{out} = 2V_{pp} \times \dfrac{1\,k\Omega}{1.414\,k\Omega} = 1.414\,V_{pp}$

E. The output waveform is shown in Figure 6-31.

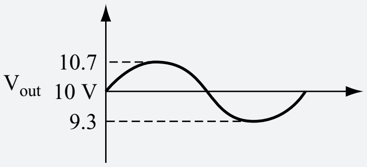

Figure 6-31

31 For the circuit shown in Figure 6-32, the DC resistance of the inductor is large enough that you should include that value in your calculations.

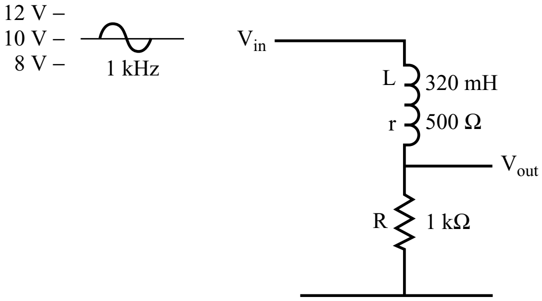

Figure 6-32

Questions

For the circuit shown in Figure 6-32, calculate the DC and AC output voltages, using the steps listed in problem 30.

A. DC V_{out} = ____________________

B. X_L = ____________________

C. Z = ____________________

D. AC V_{out} = ____________________

E. Draw the output waveform and label the voltage levels of the waveform on the blank graph in Figure 6-33.

Figure 6-33

Answers

A. $$\text{DC } V_{out} = 10\text{V} \times \frac{1\,\text{k}\Omega}{(1\,\text{k}\Omega + 500\,\Omega)} = 6.67\,\text{V}$$

NOTE **The 500 Ω DC resistance of the inductor has been added to the 1 kΩ resistor value in this calculation.**

B. $X_L = 2\ \text{k}\Omega$,

C. $Z = \sqrt{1.5^2 + 2^2} = 2.5\,\text{k}\Omega$

NOTE **The 500 Ω DC resistance of the inductor has been added to the 1 kΩ resistor value in this calculation.**

D. AC $V_{out} = 1.6\ V_{pp}$,

E. See Figure 6-34.

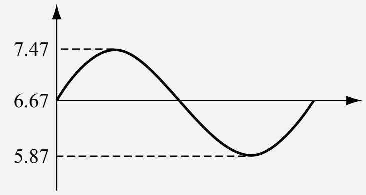

Figure 6-34

32 To calculate V_{out}, in problems 30 and 31, you also had to calculate X_L. However, because X_L changes with the frequency of the input signal, the impedance and the amplitude of V_{out} also change with the frequency of the input signal. Plotting the output voltage V_{out} against frequency results in the curve shown in Figure 6-35.

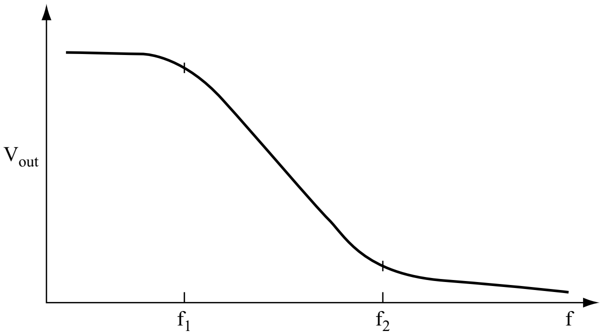

Figure 6-35

The values of the inductor and resistor determine the frequency at which V_{out} starts to drop (f_1) and the frequency at which V_{out} levels off (f_2).

The curve in Figure 6-35 shows that using an inductor and resistor in a circuit such as the one shown in Figure 6-29 produces a low pass filter similar to the one discussed in problems 15 through 19.

Question

What values will control f_1 and f_2?

Answer

The values of the inductor and the resistor

33 You can also create a circuit as shown in Figure 6-36, in which the output voltage is equal to the voltage drop across the inductor.

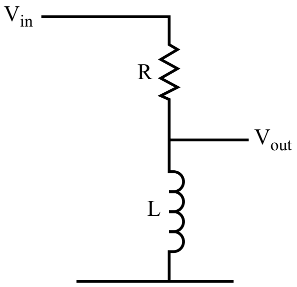

Figure 6-36

Questions

A. What formula would you use to find V_{out}?

B. If you plot the output voltage versus the frequency, what would you expect the curve to be? Use a separate sheet of paper to draw your answer.

Answers

A. $V_{out} = V_{in} \times \frac{X_L}{Z}$

B. See Figure 6-37.

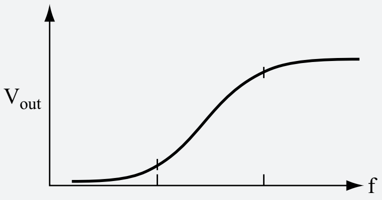

Figure 6-37

The curve in Figure 6-37 demonstrates that using an inductor and resistor in a circuit, such as the one shown in Figure 6-36, produces a high pass filter similar to the one discussed in problems 6 through 13.

Phase Shift for an RL Circuit

34 Filter circuits that use inductors (such as those shown in Figure 6-38) produce a phase shift in the output signal, just as filter circuits containing capacitors do. You can see the shifts for the circuits shown in Figure 6-38 by comparing the input and output waveforms shown below the circuit diagrams.

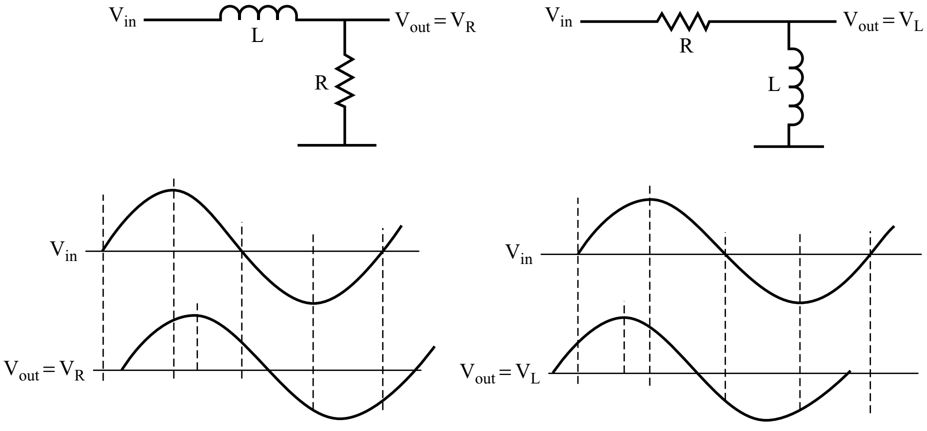

(1) (2)

Figure 6-38

Question

In which circuit does the output voltage lead the input voltage?

Answer

In graph (1), the output voltage lags the input voltage, and in graph (2) the output voltage leads.

35 Figure 6-39 shows a vector diagram for both the circuits shown in Figure 6-38. The current through the inductor lags the voltage across the inductor by 90 degrees.

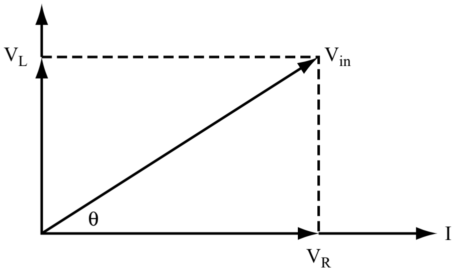

Figure 6-39

The phase angle is easily found:

$$\tan\theta = \frac{V_L}{V_R} = \frac{X_L}{R} = \frac{2\pi fL}{R}$$

Question

Calculate the phase angle for the circuit discussed in problem 30.

Answer

45 degrees

36 Refer to the circuit discussed in problem 31.

Question

Calculate the phase angle.

Answer

$$\tan\theta = \frac{X_L}{R} = \frac{2\,k\Omega}{1.5\,k\Omega} = 1.33$$

Therefore $\theta = 53.1$ degrees.

Summary

This chapter has discussed the uses of capacitors, resistors, and inductors in voltage divider and filter circuits. You learned how to determine the following:

- The output voltage of an AC signal after it passes through a high pass RC filter circuit
- The output voltage of an AC signal after it passes through a low pass RC circuit
- The output voltage of an AC signal after it passes through a high pass RL circuit
- The output voltage of an AC signal after it passes through a low pass RL circuit
- The output waveform of an AC or combined AC-DC signal after it passes through a filter circuit
- Simple phase angles and phase differences

Self-Test

These questions test your understanding of this chapter. Use a separate sheet of paper for your calculations. Compare your answers with the answers provided following the test.

For questions 1–3, calculate the following parameters for the circuit shown in each question.

A. X_C

B. Z

C. V_{out}

D. I

E. tan θ and θ

1. Use the circuit shown in Figure 6-40.

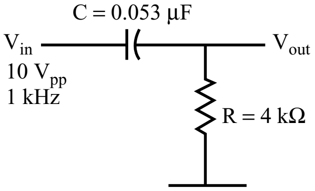

Figure 6-40

A. ____________________

B. ____________________

C. ____________________

D. ______________________

E. ______________________

2. Use the circuit shown in Figure 6-41.

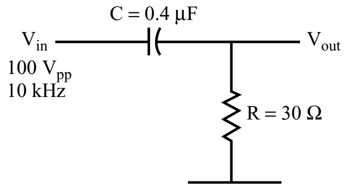
C = 0.4 μF
V_{in}
100 V_{pp}
10 kHz
V_{out}
R = 30 Ω

Figure 6-41

A. ______________________

B. ______________________

C. ______________________

D. ______________________

E. ______________________

3. Use the circuit shown in Figure 6-42.

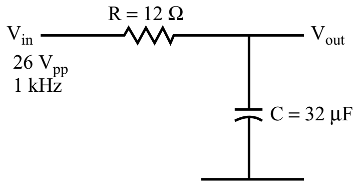
R = 12 Ω
V_{in}
26 V_{pp}
1 kHz
V_{out}
C = 32 μF

Figure 6-42

A. ______________________

B. ______________________

C. ______________________

D. ______________________

E. ______________________

For questions 4–6, calculate the following parameters for the circuit shown in each question.

A. X_C

B. AC V_{out}

C. DC V_{out}

4. Use the circuit shown in Figure 6-43.

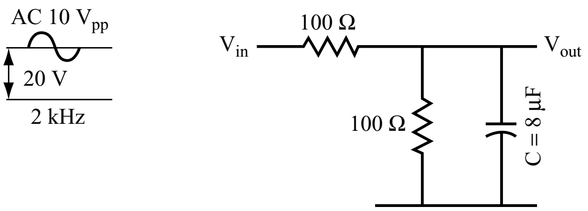
AC 10 V$_{pp}$
20 V
2 kHz
V$_{in}$
100 Ω
V$_{out}$
100 Ω
C = 8 μF

Figure 6-43

A. ______________________

B. ______________________

C. ______________________

5. Use the circuit shown in Figure 6-44.

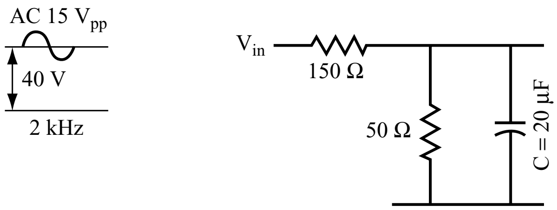
AC 15 V$_{pp}$
40 V
2 kHz
V$_{in}$
150 Ω
50 Ω
C = 20 μF

Figure 6-44

A. ______________________

B. ______________________

C. ______________________

6. Use the circuit shown in Figure 6-45.

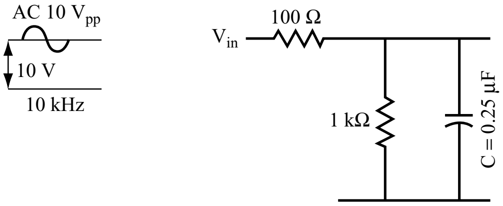
AC 10 V$_{pp}$
10 V
10 kHz
V$_{in}$
100 Ω
1 kΩ
C = 0.25 μF

Figure 6-45

A. ______________________

B. ______________________

C. ______________________

For questions 7–9, calculate the following parameters for the circuit shown in each question.

A. DC V_{out}

B. X_L

C. Z

D. AC V_{out}

E. tan θ and θ

7. Use the circuit shown in Figure 6-46.

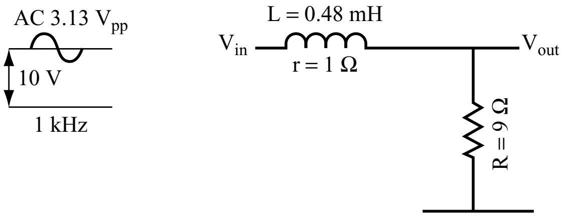
AC 3.13 V_{pp}
10 V
1 kHz
L = 0.48 mH
V_{in}
r = 1 Ω
V_{out}
R = 9 Ω

Figure 6-46

A. ______________________

B. ______________________

C. ______________________

D. ______________________

E. ______________________

8. Use the circuit shown in Figure 6-47.

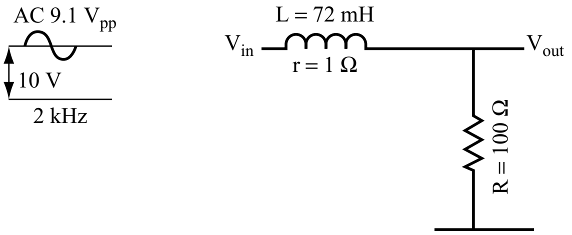
AC 9.1 V_{pp}
10 V
2 kHz
L = 72 mH
V_{in}
r = 1 Ω
V_{out}
R = 100 Ω

Figure 6-47

A. ______________________

B. ______________________

C. ______________________

D. ______________________

E. ______________________

9. Use the circuit shown in Figure 6-48.

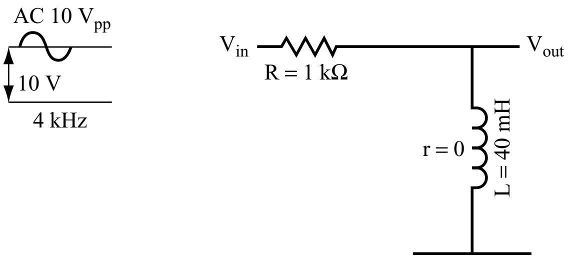

Figure 6-48

A. ______________________

B. ______________________

C. ______________________

D. ______________________

E. ______________________

Answers to Self-Test

If your answers do not agree with those provided here, review the applicable problems in this chapter before you go on to the next chapter.

1.	A.	3 kΩ	problems 8, 9, 10, 23
	B.	5 kΩ	
	C.	8 V	
	D.	2A	
	E.	36.87 degrees	
2.	A.	40 ohms	problems 8, 9, 23
	B.	50 ohms	

	C.	60 V	
	D.	2A	
	E.	53.13 degrees	
3.	A.	5 ohms	problems 8, 9, 23
	B.	13 ohms	
	C.	10 V	
	D.	2A	
	E.	22.63 degrees	
4.	A.	10 ohms	problems 26 and 27
	B.	1 V	
	C.	10 V	
5.	A.	4 ohms	problems 26 and 27
	B.	0.4 V	
	C.	10 V	
6.	A.	64 ohms	problems 26 and 27
	B.	5.4 V	
	C.	9.1 V	
7.	A.	9 V	problems 28–30, 35
	B.	3 ohms	
	C.	10.4 ohms	
	D.	2.7 V	
	E.	16.7 degrees	
8.	A.	10 V	problems 28–30, 35
	B.	904 ohms	
	C.	910 ohms	
	D.	1 V	
	E.	83.69 degrees	

9.	A.	0 V	problems 28–30, 35
	B.	1 kΩ	
	C.	1.414 kΩ	
	D.	5 V	
	E.	45 degrees	

CHAPTER

7

Resonant Circuits

You have seen how the inductor and the capacitor each present an opposition to the flow of an AC current, and how the magnitude of this reactance depends upon the frequency of the applied signal.

When inductors and capacitors are used together, in series or in parallel, a useful phenomenon called *resonance* occurs. Resonance is the frequency at which the reactance of the capacitor and the inductor is equal.

In this chapter, you learn about some of the properties of resonant circuits, and concentrate on those properties that lead to the study of oscillators (which is touched upon in the last few problems in this chapter, and covered in more depth in Chapter 9).

After completing this chapter you will be able to do the following:

- Find the impedance of a series LC circuit.
- Calculate the series LC circuit's resonant frequency.
- Sketch a graph of the series LC circuit's output voltage.
- Find the impedance of a parallel LC circuit.
- Calculate the parallel LC circuit's resonant frequency.
- Sketch a graph of the parallel LC circuit's output voltage.
- Calculate the bandwidth and the Q of simple series and parallel LC circuits.
- Calculate the frequency of an oscillator.

The Capacitor and Inductor in Series

1 Many electronic circuits contain a capacitor and an inductor placed in series, as shown in Figure 7-1.

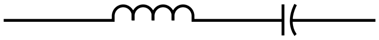

Figure 7-1

You can combine a capacitor and an inductor in series with a resistor to form voltage divider circuits, such as the two circuits shown in Figure 7-2. A circuit that contains resistance (R), inductance (L), and capacitance (C) is referred to as an RLC circuit. Note that although the order of the capacitor and inductor differs in the two circuits shown in Figure 7-2, in fact, they have the same effect on electrical signals.

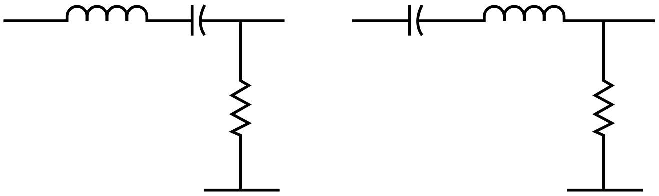

Figure 7-2

To simplify your calculations in the next few problems you can assume that the small DC resistance of the inductor is much less than the resistance of the resistor R, and you can, therefore, ignore DC resistance in your calculations.

When you apply an AC signal to the circuits in Figure 7-2, both the inductor's and the capacitor's reactance value depends on the frequency.

Questions

A. What formula would you use to calculate the inductor's reactance?

____

B. What formula would you use to calculate the capacitor's reactance?

____

Answers

A. $X_L = 2\pi fL$

B. $X_C = \dfrac{1}{2\pi fC}$

2 You can calculate the net reactance (X) of a capacitor and inductor in series using the following formula:

$X = X_L - X_C$

You can calculate the impedance of the RLC circuits shown in Figure 7-2 using the following formula:

$Z = \sqrt{R^2 + X^2}$

In the formula, keep in mind that X^2 is $(X_L + X_C)^2$.

Calculate the net reactance and impedance for an RLC series circuit, such as those shown in Figure 7-2, with the following values:

$f = 1\ \text{kHz}$

$L = 100\ \text{mH}$

$C = 1\ \mu\text{F}$

$R = 500\ \text{ohms}$

Questions

Follow these steps to calculate the following:

A. Find X_L:

____

B. Find X_C:

____

C. Use $X = X_L - X_C$ to find the net reactance:

____

D. Use $Z = \sqrt{X^2 + R^2}$ to find the impedance:

____

Answers

A. $X_L = 628$ ohms

B. $X_C = 160$ ohms

C. $X = 468$ ohms (inductive)

D. $Z = 685$ ohms

3 Calculate the net reactance and impedance for an RLC series circuit, such as those shown in Figure 7-2, using the following values:

$f = 100$ Hz
$L = 0.5$ H
$C = 5\ \mu F$
$R = 8$ ohms

Questions

Follow the steps outlined in problem 2 to calculate the following parameters:

A. $X_L =$ ____________________

B. $X_C =$ ____________________

C. $X =$ ____________________

D. $Z =$ ____________________

Answers

A. $X_L = 314$ ohms

B. $X_C = 318$ ohms

C. $X = -4$ ohms (capacitive)

D. $Z = 9$ ohms

By convention, the net reactance is negative when it is capacitive.

4 Calculate the net reactance and impedance for an RLC series circuit, such as those shown in Figure 7-2, using the following values.

Questions

A. $f = 10$ kHz, $L = 15$ mH, $C = 0.01\ \mu F$, $R = 494$ ohms

$X =$ ____________________

$Z =$ ____________________

B. f = 2 MHz, L = 8 μH, C = 0.001 μF, R = 15 ohms

X = ____________________

Z = ____________________

Answers

A. X = −650 ohms (capacitive), Z = 816 ohms

B. X = 21 ohms (inductive), Z = 25.8 ohms

5 For the circuit shown in Figure 7-3, the output voltage is the voltage drop across the resistor.

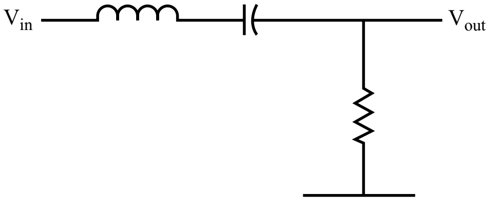

Figure 7-3

In problems 1 through 4, the net reactance of the series inductor and capacitor changes as the frequency changes. Therefore, as the frequency changes, the voltage drop across the resistor changes, and so also does the amplitude of the output voltage V_{out}.

If you plot V_{out} against frequency on a graph for the circuit shown in Figure 7-3, the curve looks like the one shown in Figure 7-4.

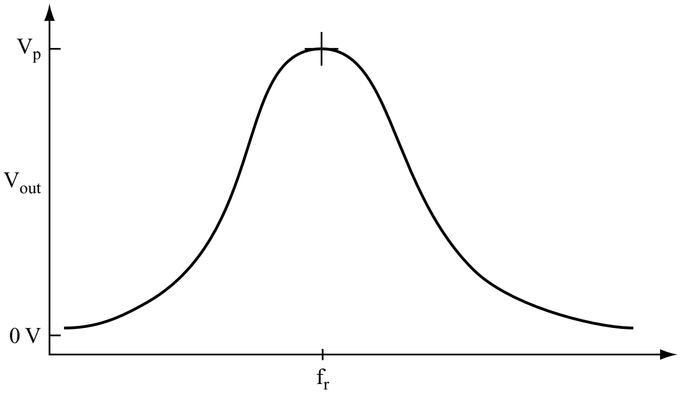

Figure 7-4

The maximum output voltage (or peak voltage) shown in this curve, V_p, is slightly less than V_{in}. This slight attenuation of the peak voltage from the input voltage is because of the DC resistance of the inductor.

The output voltage peaks at a frequency, f_r, where the net reactance of the inductor and capacitor in series is at its lowest value. At this frequency, there is very little voltage drop across the inductor and capacitor. Therefore, most of the input voltage is applied across the resistor, and the output voltage is at its highest value.

Question

Under ideal conditions, if X_C were 10.6 ohms, what value of X_L results in a net reactance (X) of 0 for the circuit shown in Figure 7-3?

Answer

$X = X_L - X_C = 0$, therefore:

$X_L = X_C + X = 10.6\ \Omega + 0 = 10.6\ \Omega$

6 You can find the frequency at which $X_L - X_C = 0$ by setting the formula for X_L equal to the formula for X_C and solving for f:

$$2\pi fL = \frac{1}{2\pi fC}$$

Therefore,

$$f_r = \frac{1}{2\pi\sqrt{LC}}$$

where f_r is the *resonant frequency* of the circuit.

Question

What effect does the value of the resistance have on the resonant frequency?

Answer

It has no effect at all.

7 Calculate the resonant frequency for the circuit shown in Figure 7-3 using the capacitor and inductor values given in the following questions.

Questions

A. C = 1 μF, L = 1 mH

f_r = ____________________

B. C = 16 μF, L = 1.6 mH

f_r = ____________________

Answers

A. $f_r = \dfrac{1}{2\pi\sqrt{1\times 10^{-3}\times 1\times 10^{-6}}} = 5.0\,\text{kHz}$

B. $f_r = \dfrac{1}{2\pi\sqrt{16\times 10^{-6}\times 1.6\times 10^{-3}}} = 1\,\text{kHz}$

8 Calculate the resonant frequency for the circuit shown in Figure 7-3 using the capacitor and inductor values given in the following questions.

Questions

A. C = 0.1 μF, L = 1 mH

f_r = ____________________

B. C = 1 μF, L = 2 mH

f_r = ____________________

Answers

A. f_r = 16 kHz

B. f_r = 3.6 kHz

9 For the RLC circuit shown in Figure 7-5, the output voltage is the voltage drop across the capacitor and inductor.

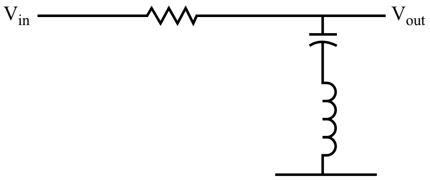

Figure 7-5

If V_{out} is plotted on a graph against the frequency for the circuit shown in Figure 7-5, the curve looks like that shown in Figure 7-6.

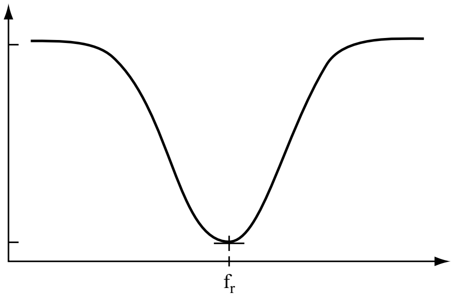

Figure 7-6

The output voltage drops to its minimum value at the resonant frequency for the circuit, which you can calculate with the formula provided in problem 6. At the resonant frequency, the net reactance of the inductor and capacitor in series is at a minimum. Therefore, there is very little voltage drop across the inductor and capacitor, and the output voltage is at its minimum value.

Questions

What would you expect the minimum output voltage to be?

Answers

0 V, or close to it

10 You can connect the capacitor and inductor in parallel, as shown in Figure 7-7.

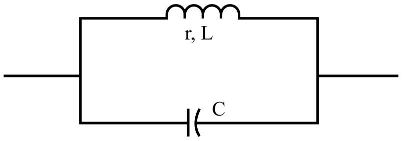

Figure 7-7

You can calculate the resonance frequency of this circuit using the following formula:

$$f_r = \frac{1}{2\pi\sqrt{LC}}\sqrt{1 - \frac{r^2C}{L}}$$

In this formula, r is the DC resistance of the inductor. However, if the reactance of the inductor is equal to or more than 10 times the DC resistance of the inductor, you can use the following simpler formula. This is the same formula that you used in problems 7 and 8 for the series circuit.

$$f_r = \frac{1}{2\pi\sqrt{LC}}$$

The symbol Q is equal to X_L/r. Therefore, you can use this simple equation to calculate f_r if Q is equal to, or greater than, 10.

Questions

A. Which formula should you use to calculate the resonant frequency of a parallel circuit if the Q of the coil is 20?

____

B. If the Q is 8?

____

Answers

A. $f_r = \frac{1}{2\pi\sqrt{LC}}$

B. $f_r = \frac{1}{2\pi\sqrt{LC}}\sqrt{1 - \frac{r^2C}{L}}$

NOTE Here is another version of the resonance frequency formula that is helpful when Q is known:

$$f_r = \frac{1}{2\pi\sqrt{LC}}\sqrt{\frac{Q^2}{1+Q^2}}$$

11 You can calculate the total opposition (impedance) of an inductor and capacitor connected in parallel to the flow of current by using the following formulas for a circuit at resonance:

$Z_p = Q_2r$, if Q is equal to or greater than 10

$Z_p = \frac{L}{rC}$, for any value of Q

At resonance, the impedance of an inductor and capacitor in parallel is at its maximum.

You can use an inductor and capacitor in parallel in a voltage divider circuit, as shown in Figure 7-8.

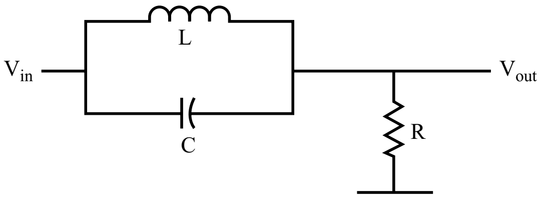

Figure 7-8

If V_{out} is plotted against frequency on a graph for the circuit shown in Figure 7-8, the curve looks like that shown in Figure 7-9.

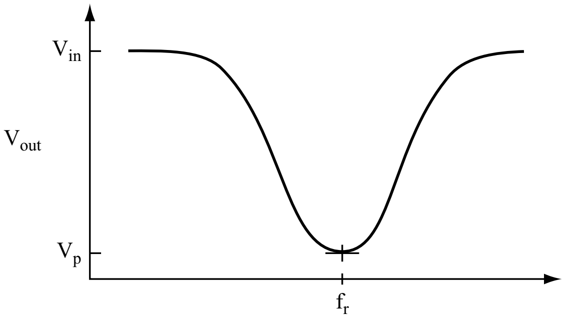

Figure 7-9

Questions

A. What would be the total impedance formula for the voltage divider circuit at resonance?

____

B. What is the frequency called at the point where the curve is at its lowest point?

____

C. Why is the output voltage at a minimum value at resonance?

____

Answers

A. $Z_T = Z_p + R$

NOTE **The relationship shown by this formula is only true at resonance. At all other frequencies, Z_T is a complicated formula or calculation found by considering a series r, L circuit in parallel with a capacitor.**

B. The parallel resonant frequency

C. The output voltage is at its lowest value at the resonant frequency. This is because the impedance of the parallel resonant circuit is at its highest value at this frequency.

12 For the circuit shown in Figure 7-10, the output voltage equals the voltage drop across the inductor and capacitor.

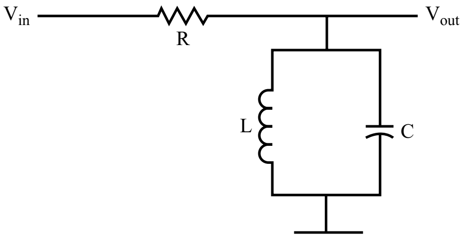

Figure 7-10

If V_{out} is plotted on a graph against frequency for the circuit shown in Figure 7-10, the curve looks like that shown in Figure 7-11. At the resonance frequency, the impedance of the parallel inductor and capacitor is at its maximum value. Therefore, the voltage drop across the parallel inductor and capacitor (which is also the output voltage) is at its maximum value.

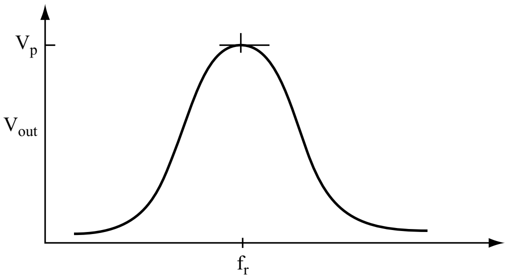

Figure 7-11

Question

What formula would you use to calculate the resonant frequency?

____

Answer

$$f_r = \frac{1}{2\pi\sqrt{LC}} \text{ if Q is equal to or greater than 10}$$

$$f_r = \frac{1}{2\pi\sqrt{LC}}\sqrt{1 - \frac{r^2C}{L}} \text{ if Q is less than 10}$$

13 Find the resonant frequency in these two examples, where the capacitor and the inductor are in parallel. (Q is greater than 10.)

Questions

A. L = 5 mH, C = 5 μF

f_r = ____________________________

B. L = 1 mH, C = 10 μF ________________

f_r = ____________________________

Answers

A. f_r = 1 kHz (approximately)

B. f_r = 1600 Hz (approximately)

The Output Curve

14 Now it's time to look at the output curve in a little more detail. Take a look at the curve shown in Figure 7-12 for an example.

An input signal at the resonant frequency, f_r, passes through a circuit with minimum attenuation and with its output voltage equal to the peak output voltage, V_p, shown on this curve.

The two frequencies f_1 and f_2 are ''passed'' almost as well as f_r is passed. That is, signals at those frequencies have a high output voltage, almost as high as the output of a signal at f_r. The graph shows this voltage as V_x.

Signals at frequencies f_3 and f_4 have a very low output voltage.

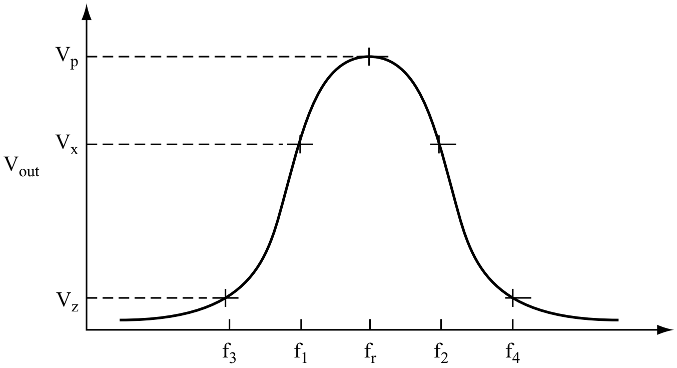

Figure 7-12

These two frequencies are not passed, but are said to be *blocked* or *rejected* by the circuit. This output voltage is shown on the graph as V_z.

The output or frequency response curve for a resonant circuit (series or parallel) has a symmetrical shape for a high value of Q. You can make the assumption that the output curve is symmetrical when Q is greater than 10.

Questions

A. What is meant by a frequency that is passed?

____

B. Why are f_1 and f_2 passed almost as well as f_r?

____

C. What is meant by a frequency that is blocked?

____

D. Which frequencies shown on the previous output curve are blocked?

____

E. Does the output curve shown appear to be symmetrical? What does this mean with regard to the circuit?

____

Answers

A. It appears at the output with minimum attenuation.

B. Because their frequencies are close to f_r.

C. It has a low output voltage.

D. f_3 and f_4 (as well as all frequencies below f_3 and above f_4).

E. It does appear to be symmetrical. This means that the coil has a Q greater than 10.

15 Somewhere between f_r and f_3, and between f_r and f_4, there is a point at which frequencies are said to be either passed or reduced to such a level that it is effectively blocked. The dividing line is at the level at which the power output of the circuit is half as much as the power output at peak value. This happens to occur at a level that is 0.707, or 70.7 percent of the peak value. For the output curve shown in problem 14, this occurs at a voltage level of 0.707 V_p. The two corresponding frequencies taken from the graph are called the *half power frequencies* or *half power points*. These are common expressions used in the design of resonant circuits and frequency response graphs.

If a certain frequency results in an output voltage that is equal to or greater than the half power point, it is said to be passed or accepted by the circuit. If it is lower than the half power point, it is said to be blocked or rejected by the circuit.

Question

Suppose $V_P = 10$ V. What is the minimum voltage level of all frequencies that are passed by the circuit?

Answer

$V = 10\ V \times 0.707 = 7.07\ V$

(If a frequency has an output voltage above 7.07 V, we would say it is passed by the circuit.)

16 Assume the resonant frequency in a circuit is 5 V. Another frequency has an output of 3.3 V.

Question

Is this second frequency passed or blocked by the circuit?

Answer

$V = V_p \times 0.707 = 5 \times 0.707 = 3.535$ V
3.3 V is less than 3.535 V, so this frequency is blocked.

17 In these examples, find the voltage level at the half power points.

Questions

A. $V_p = 20$ V ____________________

B. $V_p = 100$ V ____________________

C. $V_p = 3.2$ V ____________________

Answers

A. 14.14 V

B. 70.70 V

C. 2.262 V

18 Although this discussion started off by talking about the resonance frequency, a few other frequencies have been introduced. In fact, at this point, the discussion is dealing with a band or a range of frequencies.

There are two frequencies that correspond to the half power points on the curve. Assume these frequencies are f_1 and f_2. The difference you find when you subtract f_1 from f_2 is very important, because this gives the range of frequencies that are passed by the circuit. This range is called the *bandwidth* of the circuit, and can be calculated using the following equation.

$$BW = f_2 - f_1$$

All frequencies within the bandwidth are passed by the circuit, while all frequencies outside the bandwidth are blocked. A circuit with this type of output (such as the circuit shown in Figure 7-10) is referred to as a *bandpass filter*.

Question

Indicate which of the following pairs of values represent a wider range of frequencies, or, in other words, the wider bandwidth.

A. $f_2 = 200$ Hz, $f_1 = 100$ Hz

B. $f_2 = 20$ Hz, $f_1 = 10$ Hz

Answer

The bandwidth is wider for the frequencies given in A.

When playing a radio, you listen to one station at a time, not to the adjacent stations on the dial. Thus, your radio tuner must have a narrow bandwidth so that it can select only the frequency of that one station.

The amplifiers in a television set, however, must pass frequencies from 30 Hz up to about 4.5 MHz, which requires a wider bandwidth. The application or use to which you'll put a circuit determines the bandwidth that you should design the circuit to provide.

19 The output curve for a circuit that passes a band of frequencies around the resonance frequency (such as the curve shown in Figure 7-13) was discussed in the last few problems.

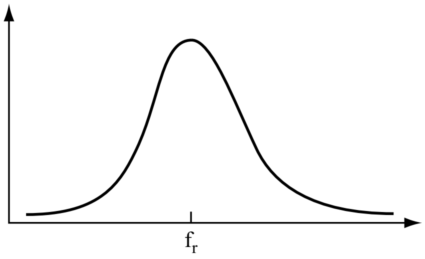

Figure 7-13

The same principles and equations apply to the output curve for a circuit that blocks a band of frequencies around the resonance frequency, as is the case with the curve shown in Figure 7-14.

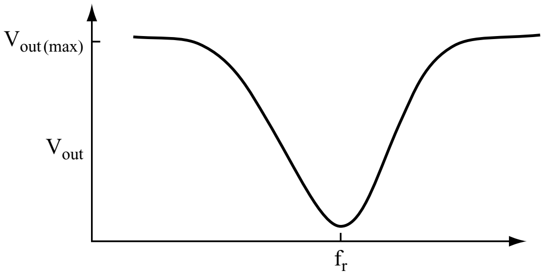

Figure 7-14

A circuit with this type of output (such as the circuit shown in Figure 7-8) is called a *notch filter*, or *band-reject filter*.

Questions

A. What points on the curve shown in Figure 7-14 would you use to determine the circuit's bandwidth?

____

B. Would the output voltage at the resonant frequency be above or below these points?

____

Answers

A. The half power points (0.707 $V_{out(max)}$)

B. The output voltage at the resonant frequency is the minimum point on the curve, which is below the level for the half power points.

20 You can find the bandwidth of a circuit by measuring the frequencies (f_1 and f_2) at which the half power points occur, and then using the following formula:

$$BW = f_2 - f_1$$

Or, you can calculate the bandwidth of a circuit using this formula:

$$BW = \frac{f_r}{Q}$$

where:

$$Q = \frac{X_L}{R}$$

The formula used to calculate bandwidth indicates that, for two circuits with the same resonant frequency, the circuit with the larger Q will have the smaller bandwidth.

When you calculate Q for a circuit containing a capacitor and inductor in series (such as that shown in Figure 7-15) use the total DC resistance — the sum of the DC resistance (r) of the inductor and the value of the resistor (R) — to calculate Q.

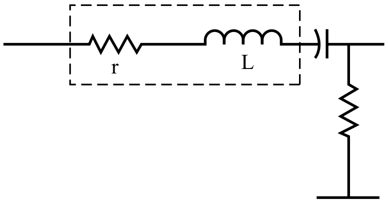

Figure 7-15

When you calculate Q for a circuit containing an inductor and capacitor in parallel, as with the circuit shown in Figure 7-16, you do not include the value of the resistor (R) in the calculation. The only resistance you use in the calculation is the DC resistance (r) of the inductor.

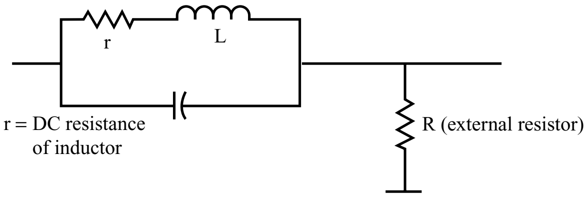

Figure 7-16

When you calculate Q for a circuit containing an inductor, capacitor, and resistor in parallel, as with the two circuits shown in Figure 7-17, include the value of the resistor (R) in the calculation.

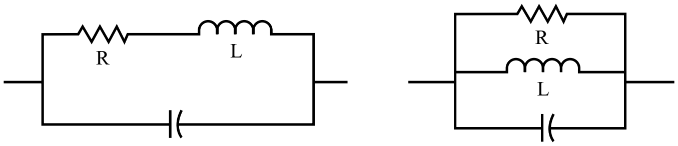

Figure 7-17

Questions

For the circuit shown in Figure 7-18, all the component values are provided in the diagram. Find f_r, Q, and BW.

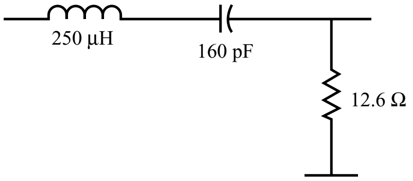

Figure 7-18

$f_r =$ ____

Q = ____

BW = ____

Answers

$$f_r = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{250 \times 10^{-6} \times 160 \times 10^{-12}}} = 796\,\text{kHZ}$$

$$Q = \frac{X_L}{R} = \frac{2\pi fL}{R} = \frac{2\pi \times 796\,\text{kHz} \times 250\,\mu\text{H}}{12.6\,\Omega} = 99.2$$

$$BW = \frac{f_r}{Q} = \frac{796\,\text{kHz}}{99.2} = 8\,\text{kHz}$$

21 Use the circuit and component values shown in Figure 7-19 to answer the following questions.

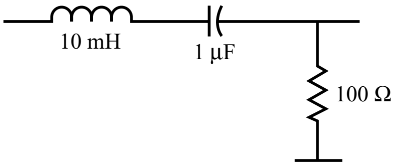

Figure 7-19

Questions

Find f_r, Q, and BW. Then, on a separate sheet of paper, draw an output curve showing the range of frequencies that are passed and blocked.

f_r = ________________________________
Q = ________________________________
BW = ________________________________

Answers

$f_r = 1590$ Hz; Q = 1; BW = 1590 Hz
The output curve is shown in Figure 7-20

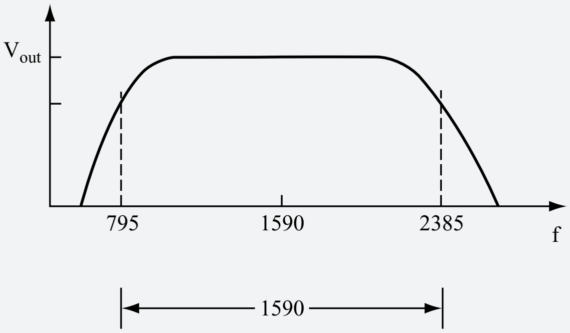

Figure 7-20

22 Use the circuit and component values shown in Figure 7-21 to answer the following questions.

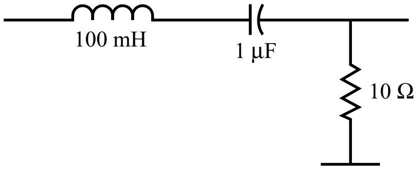

Figure 7-21

Questions

Find f_r, Q, and the bandwidth for this circuit. Then draw the output curve on a separate sheet of paper.

f_r = ____________________

Q = ____________________

BW = ____________________

Answers

$f_r = 500$ Hz; Q = 31.4, BW = 16 Hz

The output curve is shown in Figure 7-22

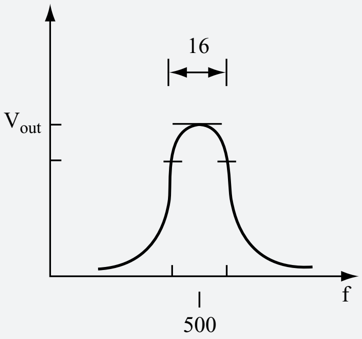

Figure 7-22

23 Use the circuit shown in Figure 7-23 for this problem. In this case, the resistor value is 10 ohms. However, the inductor and capacitor values are not given.

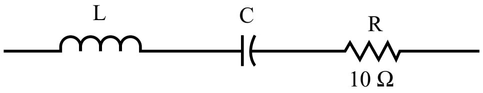

Figure 7-23

Questions

Find the bandwidth and the values of L and C required to give the circuit a resonant frequency of 1200 Hz and a Q of 80.

BW =
L =
C =

Answers

BW = 15 Hz
L = 106 mH
C = 0.166 μF
You can check these values by using the values of L and C to find f_r.

24 Use the circuit shown in Figure 7-23 for this problem. In this case, the resistor value is given as 10 ohms. However, the inductor and capacitor values are not given.

Questions

Calculate the values of Q, L, and C required to give the circuit a resonant frequency of 300 kHz with a bandwidth of 80 kQ.

Q =
L =
C =

Answers

Q = 3.75
L = 20 μH
C = 0.014 μF

25 A circuit that only passes (or blocks) a narrow range of frequencies is called a *high Q* circuit. Figure 7-24 shows the output curve for a high Q circuit.

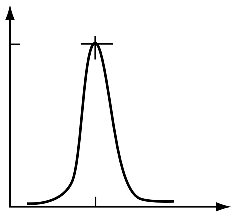

Figure 7-24

Because of the narrow range of frequencies it passes, a high Q circuit is said to be very *selective* in the frequencies it passes.

A circuit that passes (or blocks) a wide range of frequencies is called a *low Q circuit*. Figure 7-25 shows the output curve for a low Q circuit.

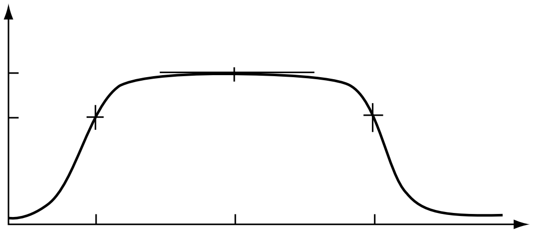

Figure 7-25

Recall the discussion in problem 18 (comparing the bandwidths of radio tuners and television amplifiers) to help you answer the following questions.

Questions

A. Which is the more selective, the radio tuner or the television amplifier?

____

B. Which would require a lower Q circuit, the radio tuner or the television amplifier?

____

Answers

A. The radio

B. The television amplifier

26 The inductor and capacitor in Figure 7-26 are connected in parallel, rather than in series. However, you can use the same formulas you used for the series circuit in problem 20 to calculate f_r, Q, and the bandwidth for parallel LC circuits.

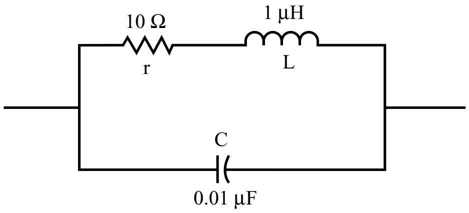

Figure 7-26

Questions

Find f_r, Q, and the bandwidth for the circuit shown in Figure 7-26.

A. f_r = ______________________

B. Q = ______________________

C. BW = ______________________

Answers

A. $f_r = 1.6$ MHz

B. $X_L = 10$ ohms, so $Q = 10/0.01 = 100$ (Note that the only resistance here is the small DC resistance of the inductor.)

C. BW = 16 kHz (This is a fairly high Q circuit.)

27 In the last few problems, you learned how to calculate f_r, BW, and Q, for a given circuit, or conversely, to calculate the component values that would produce a circuit with specified f_r, BW, and Q values.

Once you know the resonant frequency and bandwidth for a circuit, you can sketch an approximate output curve. With the simple calculations listed in this problem, you can draw a curve that is accurate to within 1 percent of its true value.

The curve that results from the calculations used in this problem is sometimes called the *general resonance curve*.

You can determine the output voltage at several frequencies by following these steps:

1. Assume the peak output voltage V_p at the resonant frequency f_r to be 100 percent. This is point A on the curve shown in Figure 7-27.
2. The output voltage at f_1 and f_2 is 0.707 of 100 percent. On the graph, these are the two points labeled B in Figure 7-27. Note that $f_2 - f_1 = \text{BW}$. Therefore, at half a bandwidth above and below f_r, the output is 70.7 percent of V_p.
3. At f_3 and f_4 (the two points labeled C in Figure 7-27), the output voltage is 44.7 percent of V_p. Note that $f_4 - f_3 = 2$ BW. Therefore, at 1 bandwidth above and below f_r, the output is 44.7 percent of maximum.
4. At f_5 and f_6 (the two points labeled D in Figure 7-27), the output voltage is 32 percent of V_p. Note that $f_6 - f_5 = 3$ BW. Therefore, at 1.5 bandwidth above and below f_r, the output is 32 percent of maximum.

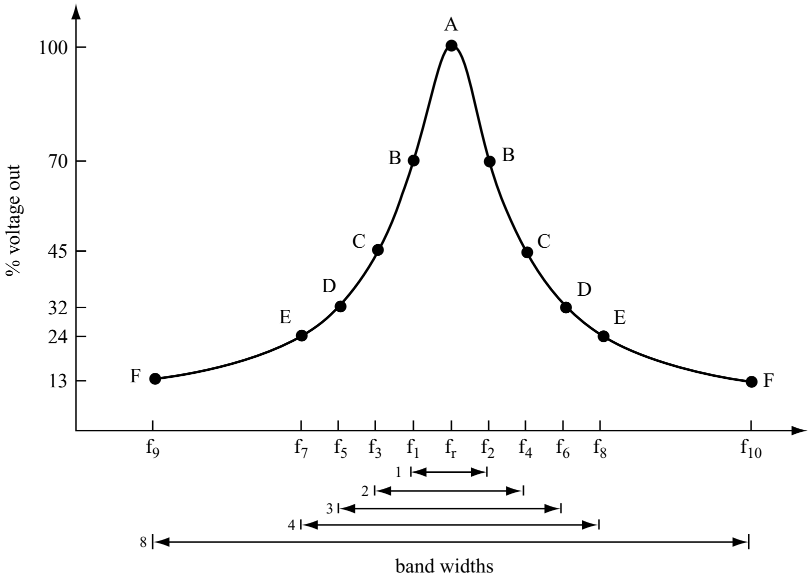

Figure 7-27

5. At f_7 and f_8 (the two points labeled E in Figure 7-27), the output voltage is 24 percent of V_p. Note that $f_8 - f_7 = 4$ BW. Therefore, at 2 bandwidth above and below f_r, the output is 24 percent of maximum.
6. At f_{10} and f_9 (the two points labeled F in Figure 7-27), the output is 13 percent of V_p. Note that $f_{10} - f_9 = 8$ BW. Therefore, at 4 bandwidth above and below $f_{r,}$, the output is 13 percent of maximum.

Questions

Calculate f_r, X_L, Q, and the BW for the circuit shown in Figure 7-28.

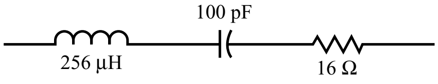

Figure 7-28

$f_r =$ ______________________________
$X_L =$ ______________________________
Q = ______________________________
BW = ______________________________

Answers

$f_r = 1$ MHz
$X_L = 1607$ ohms

$Q = 100$
$BW = 10$ kHz

28 Now, calculate the frequencies that correspond with each percentage of the peak output voltage listed in steps 1 through 6 of problem 27. (Refer to the graph in Figure 7-27 as needed.)

Questions

A. At what frequency will the output level be maximum?

____

B. At what frequencies will the output level be 70 percent of V_p?

____

C. At what frequencies will the output level be 45 percent of V_p?

____

D. At what frequencies will the output level be 32 percent of V_p?

____

E. At what frequencies will the output level be 24 percent of V_p?

____

F. At what frequencies will the output level be 13 percent of V_p?

____

Answers

A. 1 MHz

B. 995 kHz and 1005 kHz (1 MHz − 5 kHz and + 5 kHz)

C. 990 kHz and 1010 kHz

D. 985 kHz and 1015 kHz

E. 980 kHz and 1020 kHz

F. 960 kHz and 1040 kHz

29 You can calculate the output voltage at each frequency in the answers to problem 28 by multiplying the peak voltage by the related percentage for each frequency.

Question

Calculate the output voltage for the frequencies given here, assuming that the peak output voltage is 5 volts.

A. What is the output voltage level at 995 kHz?

B. What is the output voltage level at 980 kHz?

Answer

A. V = 5 V × 0.70 = 3.5 V

B. V = 5 V × 0.24 = 1.2 V

Figure 7-29 shows the output curve generated by plotting the frequencies calculated in problem 28, and the corresponding output voltages calculated in this problem.

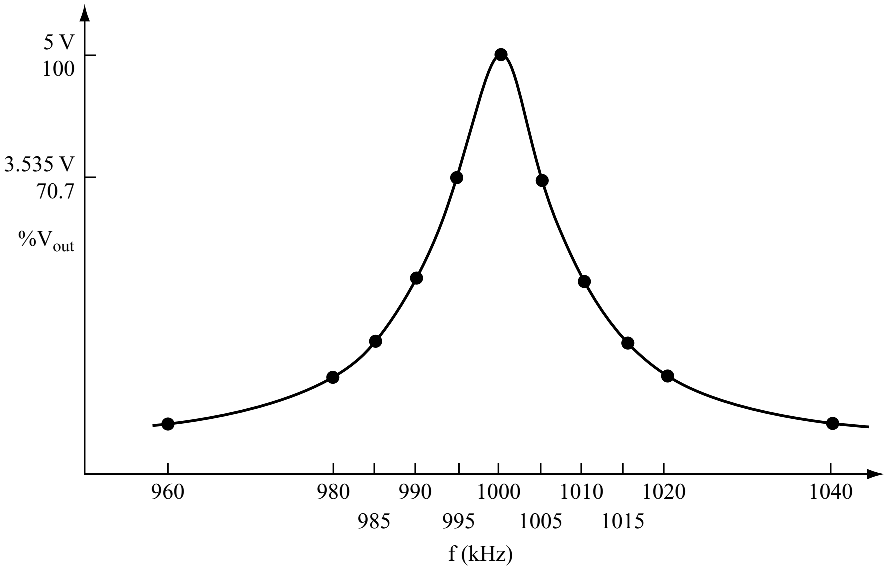

Figure 7-29

Introduction to Oscillators

In addition to their use in circuits used to filter input signals, capacitors and inductors are used in circuits called *oscillators*.

Oscillators are circuits that generate waveforms at particular frequencies. Many oscillators use a tuned parallel LC circuit to produce a sine wave output. This section is an introduction to the use of parallel capacitors and inductors in oscillators.

30 When the switch in the circuit shown in drawing (1) of Figure 7-30 is closed, current flows through both sides of the parallel LC circuit in the direction shown.

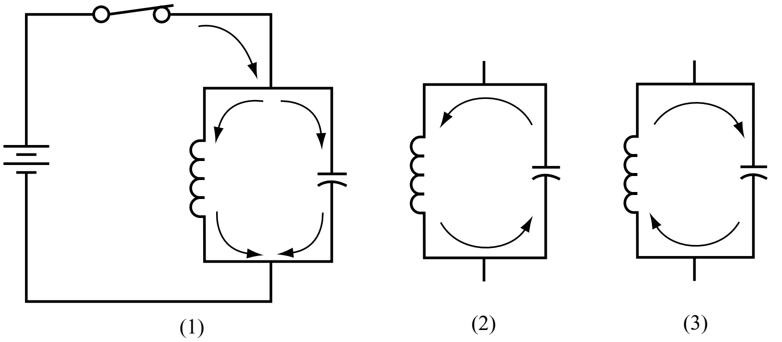

Figure 7-30

It is difficult for the current to flow through the inductor initially, because the inductor opposes any changes in current flow. Conversely, it is easy for the current to flow into the capacitor initially because, with no charge on the plates of the capacitor, there is no opposition to the flow.

As the charge on the capacitor increases, the current flow in the capacitor side of the circuit decreases. However, more current flows through the inductor. Eventually the capacitor is fully charged, so current stops flowing in the capacitor side of the circuit, and a steady current flows through the inductor.

Question

When you open the switch, what happens to the charge on the capacitor?

Answer

It discharges through the inductor. (Note the current direction, shown in drawing (2) of Figure 7-30.)

31 With the switch open, current continues to flow until the capacitor is fully discharged.

Question

When the capacitor is fully discharged, how much current is flowing through the inductor?

Answer

None.

32 Because there is no current in the inductor, its magnetic field collapses. The collapsing of the magnetic field induces a current to flow in the inductor, and this current flows in the same direction as the original current through the inductor (remember that an inductor resists any change in current flow), which is shown in drawing (2) of Figure 7-30. This current now charges the capacitor to a polarity that is opposite from the polarity that the battery induced.

Question

When the magnetic field of the inductor has fully collapsed, how much current will be flowing?

Answer

None.

33 Next, the capacitor discharges through the inductor again, but this time the current flows in the opposite direction, as shown in drawing (3) of Figure 7-30. The change in current direction builds a magnetic field of the opposite polarity. The magnetic field stops growing when the capacitor is fully discharged.

Because there is no current flowing through the inductor, its magnetic field collapses and induces current to flow in the direction shown in drawing (3) of Figure 7-30.

Question

What do you think the current generated by the magnetic field in the inductor will do to the capacitor?

Answer

It charges it to the original polarity.

34 When the field has fully collapsed, the capacitor stops charging. It now begins to discharge again, causing current to flow through the inductor in the direction shown in drawing (2) of Figure 7-30. This ‘‘seesaw’’ action of current will continue indefinitely.

As the current flows through the inductor, a voltage drop occurs across the inductor. The magnitude of this voltage drop will increase and decrease as the magnitude of the current changes.

Question

What would you expect the voltage across the inductor to look like when you view it on an oscilloscope?

Answer

A sine wave

35 In a perfect circuit this oscillation continues and produces a continuous sine wave. In practice, a small amount of power is lost in the DC resistance of the inductor and the other wiring. As a result, the sine wave gradually decreases in amplitude and dies out to nothing after a few cycles, as shown in Figure 7-31.

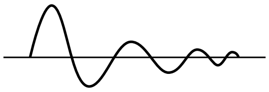

Figure 7-31

Question

How might you prevent this fade out?

Answer

By replacing a small amount of energy in each cycle

This lost energy can be injected into the circuit by momentarily closing and opening the switch at the correct time. (See drawing (1) of Figure 7-30.) This would sustain the oscillations indefinitely.

An electronic switch (such as a transistor) could be connected to the inductor as shown in Figure 7-32. Changes in the voltage drop across the inductor would turn the electronic switch on or off, thereby opening or closing the switch.

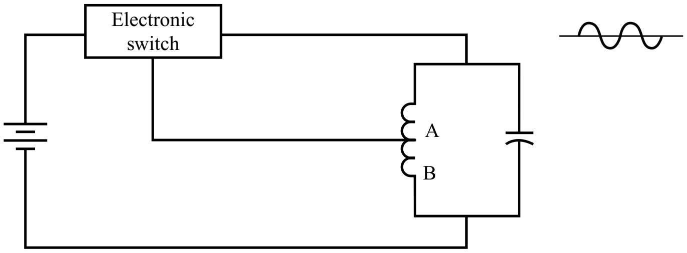

Figure 7-32

The small voltage drop across the few turns of the inductor (also referred to as a *coil*), between point B at the end of the coil and point A about halfway along the coil is used to operate the electronic switch. These points are shown in Figure 7-32.

Using a small part of an output voltage in this way is called feedback because the voltage is ''fed back'' to an earlier part of the circuit to make it operate correctly.

When you properly set up such a circuit, it produces a continuous sine wave output of constant amplitude and constant frequency. This circuit is called an *oscillator*. You can calculate the frequency of the sine waves generated by an oscillator with the following formula for determining resonant frequency:

$$f = \frac{1}{2\pi\sqrt{LC}}$$

The principles you learned in the last few problems are used in practical oscillator circuits, such as those presented in Chapter 9.

Summary

In this chapter, you learned about the following topics related to resonant circuits:

- How the impedance of a series LC circuit and a parallel LC circuit changes with changes in frequency.

- At resonant frequency for a parallel LC circuit, the impedance is at its highest, whereas for a series LC circuit impedance is at its lowest.
- The concept of bandwidth allows you to easily calculate the output voltage at various frequencies and draw an accurate output curve.
- The principles of bandpass filters and notch (or band-reject) filters.
- The fundamental concepts integral to understanding how an oscillator functions.

Self-Test

These questions test your understanding of the concepts covered in this chapter. Use a separate sheet of paper for your drawings or calculations. Compare your answers with the answers provided following the test.

1. What is the formula for the impedance of a series LC circuit?

2. What is the formula for the impedance of a series RLC circuit (a circuit containing resistance, inductance, and capacitance)?

3. What is the relationship between X_C and X_L at the resonant frequency?

4. What is the voltage across the resistor in a series RLC circuit at the resonant frequency?

5. What is the voltage across a resistor in series with a parallel LC circuit at the resonant frequency?

6. What is the impedance of a series circuit at resonance?

7. What is the formula for the impedance of a parallel circuit at resonance?

8. What is the formula for the resonant frequency of a circuit?

9. What is the formula for the bandwidth of a circuit?

10. What is the formula for the Q of a circuit?

Questions 11–13 use a series LC circuit. In each case, the values of the L, C, and R are given. Find f_r, X_L, X_C, Z, Q, and BW. Draw an output curve for each answer.

11. L = 0.1 mH, C = 0.01 μF, R = 10 ohms

12. L = 4 mH, C = 6.4 μF, R = 0.25 ohms

13. L = 16 mH, C = 10 μF, R = 20 ohms

Questions 14 and 15 use a parallel LC circuit. No R is used; r is given. Find f_r, X_L, X_C, Z, Q, and BW.

14. L = 6.4 mH, C = 10 μF, r = 8 ohms

15. L = 0.7 mH, C = 0.04 μF, r = 1.3 ohms

16. Use the output curve shown in Figure 7-33 to answer the following questions.

A. What is the peak value of the output curve?

B. What is the resonant frequency?

C. What is the voltage level at the half power points?

D. What are the half power frequencies?

E. What is the bandwidth?

F. What is the Q of the circuit?

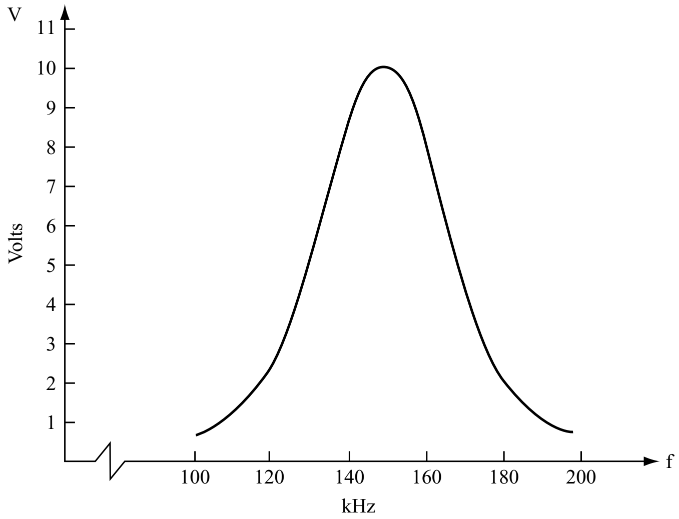

Figure 7-33

Answers to Self-Test

If your answers do not agree with those given here, review the problems indicated in parentheses before you go on to the next chapter.

1.	$Z = X_L - X_C$	(problem 2)
2.	$Z = \sqrt{(X_L - X_C)^2 + R^2}$	(problem 2)
3.	$X_L = X_C$	(problem 5)
4.	Maximum output	(problem 5)
5.	Minimum output	(problem 11)
6.	Z = minimum. Ideally, it is equal to the resistance.	(problem 5)
7.	$Z = \frac{L}{Cr}$	(problem 10)

		In this formula, r is the resistance of the coil.	
8.		$f_r = \frac{1}{2\pi\sqrt{LC}}$	(problem 6)
9.		$BW = \frac{f_r}{Q}$	(problem 20)
10.		$Q = \frac{X_L}{R}$	(problem 20)
		or	
		$\frac{X_L}{r}$	
		To draw the output curves for Questions 11–13, use the graph in Figure 7-27 as a guide and insert the appropriate bandwidth and frequency values.	(problems 21–29)
11.		$f_r = 160$ kHz, $X_L = X_C = 100$ ohms, $Q = 10$, $BW = 16$ kHz, $Z = 10$ ohms	(problems 21–29)
12.		$f_r = 1$ kHz, $X_L = X_C = 25$ ohms, $Q = 100$, $BW = 10$ Hz, $Z = 0.25$ ohms	(problems 21–29)
13.		$f_r = 400$ Hz, $X_L = X_C = 40$ ohms, $Q = 2$, $BW = 200$ Hz, $Z = 20$ ohms	(problems 21–29)
14.		$f_r = 600$ Hz, $X_L = 24$ ohms, $X_C = 26.5$, $Q = 3$, $BW = 200$ Hz, $Z = 80$ ohms	(problems 21–29)
		Because Q is not given, you should use the more complicated of the two formulas shown in Problem 10 to calculate the resonant frequency.	
15.		$f_r = 30$ Hz, $X_L = 132$ ohms, $X_C = 132$, $Q = 101.5$, $BW = 300$ Hz, $Z = 13.4$ ohms	(problems 21–29)
16.			
	A.	10.1 V	(problems 27 and 28)
	B.	148 kHz	(problems 27 and 28)
	C.	$10.1 \times 0.707 = 7.14$ V	(problems 27 and 28)
	D.	Approximately 135 kHz and 160 kHz (not quite symmetrical)	(problems 27 and 28)
	E.	$BW = 25$ kHz	(problems 27 and 28)
	F.	$Q = \frac{f_r}{BW} =$ about 5.9	(problems 27 and 28)

CHAPTER 8

Transistor Amplifiers

Many of the AC signals you'll work with in electronics are very small. For example, the signal that an optical detector reads from a DVD disk cannot drive a speaker, and the signal from a microphone's output is too weak to send out as a radio signal. In cases such as these, you must use an amplifier to boost the signal.

The best way to demonstrate the basics of amplifying a weak signal to a usable level is by starting with a one-transistor amplifier. Once you understand a one-transistor amplifier, you can grasp the building block that makes up amplifier circuits used in electronic devices such as cellphones, MP3 players, and home entertainment centers.

Many amplifier circuit configurations are possible. The simplest and most basic of amplifying circuits are used in this chapter to demonstrate how a transistor amplifies a signal. You will also see the steps for designing an amplifier.

The emphasis in this chapter is on the BJT, just as it was in Chapters 3 and 4 (which dealt primarily with the application of transistors in switching circuits). Two other types of devices used as amplifiers are also examined: the JFET (introduced in Chapters 3 and 4) and an integrated circuit called the *operational amplifier* (*op-amp*).

When you complete this chapter, you will be able to do the following:

- Calculate the voltage gain for an amplifier.
- Calculate the DC output voltage for an amplifier circuit.
- Select the appropriate resistor values to provide the required gain to an amplifier circuit.

- Identify several ways of increasing the gain of a one-transistor amplifier.
- Distinguish between the effects of a standard one-transistor amplifier and an emitter follower circuit.
- Design a simple emitter follower circuit.
- Analyze a simple circuit to find the DC level out and the AC gain.
- Design a simple common source (JFET) amplifier.
- Analyze a JFET amplifier to find the AC gain.
- Recognize an op-amp and its connections.

Working with Transistor Amplifiers

1 In Chapter 3, you learned how to turn transistors ON and OFF. You also learned how to calculate the value of resistors in amplifier circuits to set the collector DC voltage to half the power supply voltage. To review this concept, examine the circuit shown in Figure 8-1.

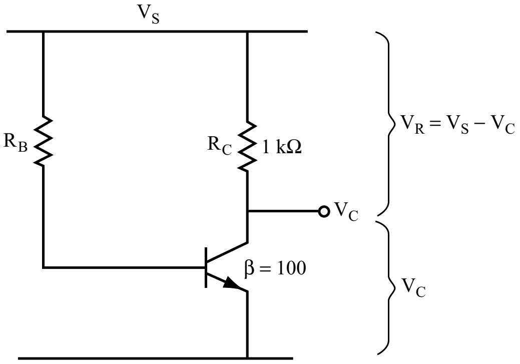

Figure 8-1

Use the following steps to find the value of R_B that will set the collector DC voltage (V_C) to half the supply voltage (V_S).

1. Find I_C by using the following:

$$I_C = \frac{V_R}{R_C} = \frac{V_S - V_C}{R_C}$$

2. Find I_B by using the following:

$$I_B = \frac{I_C}{\beta}$$

3. Find R_B by using the following:

$$R_B = \frac{V_S}{I_B}$$

Questions

Find the value of R_B that will set the collector voltage to 5 V, using Steps 1–3 and the following values for the circuit:

$V_S = 10\,V, \quad R_C = 1\,k\Omega, \quad \beta = 100$

A. $I_C =$ ________________________

B. $I_B =$ ________________________

C. $R_B =$ ________________________

Answers

A. $I_C = \frac{5\,V}{1\,k\Omega} = 5\,mA$

B. $I_B = \frac{5}{100} = 0.05\,mA$

C. $R_B = \frac{10\,V}{0.05\,mA} = 200\,k\Omega$

2 You have seen that using a 200 kΩ resistor for R_B gives an output level of 5 V at the collector. This procedure of setting the output DC level is called *biasing*. In problem 1, you biased the transistor to a 5 V DC output.

Use the circuit shown in Figure 8-1 and the formulas given in problem 1 to answer the following questions.

Questions

A. If you decrease the value of R_B, how do I_B, I_C, V_R, and the bias point V_C change?

____

B. If you increase the value of R_B how do I_B, I_C, V_R, and V_C change?

____

Answers

A. I_B increases, I_C increases, V_R increases, and so the bias point V_C decreases.

B. I_B decreases, I_C decreases, V_R decreases, and so the bias point V_C increases.

3 In problem 2 you found that changing the value of R_B, in the circuit shown in Figure 8-1 changes the value of I_B.

The transistor amplifies slight variations in I_B. Therefore, the amount I_C that fluctuates is β times the change in value in I_B.

The variations in I_C cause changes in the voltage drop V_R across R_C. Therefore, the output voltage measured at the collector also changes.

Questions

For the circuit shown in Figure 8-1, calculate the following parameters when $R_B = 168\ \text{k}\Omega$ and $V_S = 10$ V.

A. $I_B = \dfrac{V_S}{R_B} =$

____

B. $I_C = \beta I_B =$

____

C. $V_R = I_C R_C =$

____

D. $V_C = V_S - V_R =$

____

Answers

A. $I_B = \dfrac{10\ \text{V}}{168\ \text{k}\Omega} = 0.059\ \text{mA}$

B. $I_C = 100 \times 0.059 = 5.9\ \text{mA}$

C. $V_R = 1\ \text{k}\Omega \times 5.9\ \text{mA} = 5.9\ \text{V}$

D. $V_C = 10\ \text{V} - 5.9\ \text{V} = 4.1\ \text{V}$

4 Use the circuit shown in Figure 8-1 to answer the following questions when $V_S = 10$ V.

Questions

Calculate V_C for each of the following values of R_B.

A. 100 kΩ

____

B. 10 MΩ

____

C. 133 kΩ

____

D. 400 kΩ

____

Answers

A. $I_B = 0.1$ mA, $I_C = 10$ mA, $V_C = 0$ V

B. $I_B = 1$ μA, $I_C = 0.1$ mA, $V_C = 10$ V (approximately)

C. $I_B = 0.075$ mA, $I_C = 7.5$ mA, $V_C = 2.5$ V

D. $I_B = 0.025$ mA, $I_C = 2.5$ mA, $V_C = 7.5$ V

5 The values of I_C and V_C that you calculated in problems 1 and 4 are plotted on the graph on the left side of Figure 8-2. The straight line connecting these points on the graph is called the *load line*.

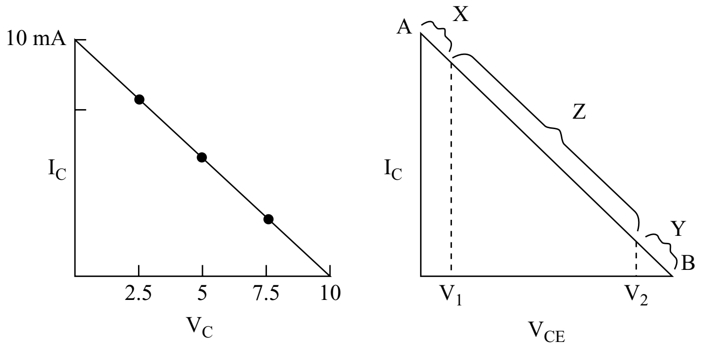

Figure 8-2

The axis labeled V_C represents the voltage between the collector and the emitter of the transistor, and not the voltage between the collector and ground. Therefore, this axis should correctly be labeled V_{CE}, as shown in the graph on the right of the figure. (For this circuit, $V_{CE} = V_C$ because there is no resistor between the emitter and ground.)

Questions

A. At point A in the graph on the right, is the transistor ON or OFF?

B. Is it ON or OFF at point B?

Answers

A. ON because full current flows, and the transistor acts like a short circuit. The voltage drop across the transistor is 0 V.

B. OFF because essentially no current flows, and the transistor acts like an open circuit. The voltage drop across the transistor is at its maximum (10 V, in this case).

6 Point A on the graph shown in Figure 8-2 is called the *saturated point* (or the *saturation point*) because it is at that point that the collector current is at its maximum.

Point B on the graph shown in Figure 8-2 is often called the *cutoff* point because, at that point, the transistor is OFF and no collector current flows.

In regions X and Y, the gain (β) is not constant, so these are called the *nonlinear regions*. Note that $\beta = I_C/I_B$. Therefore, β is the slope of the line shown in the graph.

As a rough guide, V_1 is about 1 V, and V_2 is about 1 V less than the voltage at point B.

Question

What is the value of V_{CE} at point B?

Answer

$V_{CE} = V_S$, which is 10 V in this case.

7 In region Z of the graph shown in Figure 8-2, β (that is, the slope of the graph) is constant. Therefore, this is called the *linear region*. Operating the transistor in the linear region results in an output signal that is free of distortion.

Question

Which values of I_C and V_C would result in an undistorted output in the circuit shown in Figure 8-1?

A. $I_C = 9$ mA, $V_C = 1$ V

B. $I_C = 1$ mA, $V_C = 9$ V

C. $I_C = 6$ mA, $V_C = 4.5$ V

Answer

C is the only one. A and B fall into nonlinear regions.

8 If you apply a small AC signal to the base of the transistor after it has been biased, the small voltage variations of the AC signal (shown in Figure 8-3 as a sine wave) cause small variations in the base current.

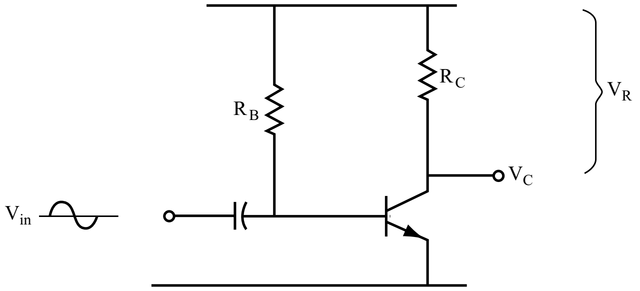

Figure 8-3

These variations in the base current will be amplified by a factor of β, and will cause corresponding variations in the collector current. The variations in the collector current, in turn, will cause similar variations in the collector voltage.

The β used for AC gain calculations is different from the β used in calculating DC variations. The AC β is the value of the common emitter AC forward current transfer ratio, which is listed as h_{fe} in manufacturers' transistor datasheets. Use the AC β whenever you need to calculate the AC output for a given AC input or to determine an AC current variation. Use the DC β to calculate the base or collector DC current values. It is important that you know which β to use and remember that one is used for DC and the other is used for AC variations. The DC β is sometimes called h_{FE} or β_{dc}.

As V_{in} increases, the base current increases, which causes the collector current to increase. An increase in the collector current increases the voltage drop across R_C, which causes V_C to decrease.

NOTE The capacitor shown at the input blocks DC (infinite reactance) and easily passes AC (low reactance). This is a common isolation technique used at the input and output of AC circuits.

Questions

A. If the input signal decreases, what happens to the collector voltage?

B. If you apply a sine wave to the input, what waveform would you expect at the collector?

Answers

A. The collector voltage, V_C, increases.

B. A sine wave, but inverted as shown in Figure 8-4.

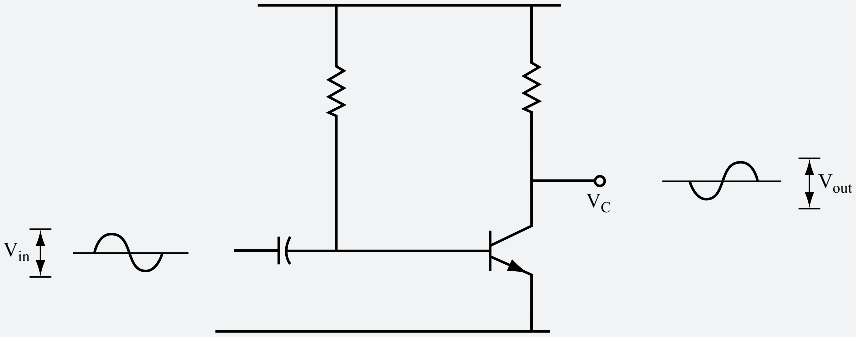

Figure 8-4

9 Figure 8-4 shows the input and output sine waves for an amplifier circuit.

The input voltage V_{in} is applied to the base. (Strictly speaking, it is applied across the base-emitter diode.) The voltage variations at the collector are centered on the DC bias point V_C, and they will be larger than variations in the input voltage. Therefore, the output sine wave is larger than the input sine wave (that is, amplified).

This amplified output signal at the collector can be used to drive a load (such as a speaker).

To distinguish these AC variations in output from the DC bias level, you indicate the AC output voltage by V_{out}. In most cases, V_{out} is a peak-to-peak value.

Questions

A. What is meant by V_C?

____

B. What is meant by V_{out}?

____

Answers

A. Collector DC voltage, or the bias point

B. AC output voltage

The ratio of the output voltage to the input voltage is called the *voltage gain* of the amplifier.

$$\text{Voltage gain} = A_V = \frac{V_{out}}{V_{in}}$$

To calculate the voltage gain of an amplifier, you can measure V_{in} and V_{out} with an oscilloscope. Measure peak-to-peak voltages for this calculation.

10 For the circuit shown in Figure 8-4 you can calculate the voltage gain using the following formula:

$$A_V = \beta \times \frac{R_L}{R_{in}}$$

In this equation:

- R_L = is the *load resistance*. In this circuit, the collector resistor R_C is the load resistance.
- R_{in} = the *input resistance* of the transistor. You can find R_{in} (often called h_{ie}) on the manufacturers' data or specification sheets. In most transistors, input resistance is about 1 kΩ to 2 kΩ.

You can find V_{out} by combining these two voltage gain equations:

$$A_V = \frac{V_{out}}{V_{in}} \quad \text{and} \quad A_V = \beta \times \frac{R_L}{R_{in}}$$

$$\text{Therefore, } \frac{V_{out}}{V_{in}} = \beta \times \frac{R_L}{R_{in}}$$

Solving this for V_{out} results in the following equation. Here, the values of $R_{in} = 1\ k\Omega$, $V_{in} = 1\ mV$, $R_C = 1\ k\Omega$, and $\beta = 100$ were used to perform this sample calculation.

$$\begin{aligned} V_{out} &= V_{in} \times \beta \times \frac{R_L}{R_{in}} \\ &= 1\,mV \times 100 \times \frac{1\,k\Omega}{1\,k\Omega} \\ &= 100\,mV \end{aligned}$$

Questions

A. Calculate V_{out} if $R_{in} = 2\ k\Omega$, $V_{in} = 1\ mV$, $R_C = 1\ k\Omega$, and $\beta = 100$.

____

B. Find the voltage gain in both cases.

____

Answers

A. $V_{out} = 50\ mV$

B. $A_V = 100$ and $A_V = 50$

This simple amplifier can provide voltage gains of up to about 500. But it does have several faults that limit its practical usefulness.

- Because of variations in β between transistors, V_C changes if the transistor is changed. To compensate for this, you will have to adjust R_B.
- R_{in} or h_{ie} varies greatly from transistor to transistor. This variation, combined with variations in β, means that you cannot guarantee the gain from one transistor amplifier to another.
- Both R_{in} and β change greatly with temperature, hence the gain is very temperature-dependent. For example, a simple amplifier circuit like that discussed in this problem was designed to work in the desert in July. It would fail completely in Alaska in the winter. If the amplifier worked perfectly in the lab, it probably would not work outdoors on either a hot or cold day.

NOTE An amplifier whose gain and DC level bias point change as described in this problem is said to be *unstable*. For reliable operation, an amplifier should be as stable as possible. In later problems, you will see how to design a stable amplifier.

The Transistor Amplifier Experiment

In the following experiment, you build a simple transistor amplifier circuit, measure the output voltage and gain, and determine how the output voltage and gain changes with temperature.

You will need the following equipment and supplies:

- A transistor
- A 250 kΩ potentiometer (that is, a variable resistor)
- A 1 kΩ resistor
- A 10 kΩ resistor
- A 0.1 μF capacitor
- A 9 V transistor radio battery or a lab type power supply
- A breadboard
- A signal generator or a sine wave oscillator
- An oscilloscope

If you do not have what you need to set up the circuit and measure the values, just read through the experiment.

Follow these steps to complete the experiment:

1. Set up the circuit shown in Figure 8-5 on the breadboard.

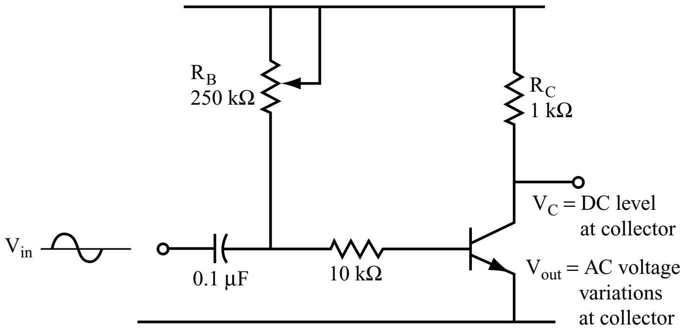

R_B
250 kΩ
R_C
1 kΩ
V_C = DC level at collector
V_{in}
0.1 μF
10 kΩ
V_{out} = AC voltage variations at collector

Figure 8-5

2. Adjust the potentiometer so that the collector DC level is between 4.5 V and 5 V.
3. Connect the signal generator to the input and adjust the frequency to produce a sine wave output of 1 kHz.

4. Adjust the level of the signal generator so that the output of the transistor at the collector, viewed on an oscilloscope, is not distorted. Make sure the output looks like a sine wave and not like the waveform shown in Figure 8-6.

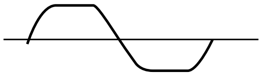

Figure 8-6

5. Measure the pp sine wave input and output of the circuit.
6. Find the voltage gain from the formula $A_V = V_{out}/V_{in}$.
7. Now, heat the transistor while it is in the circuit by placing a soldering iron near it for 15–30 seconds. Note the changes in the DC level at the collector, and the changes in the pp level of the output sine wave.
8. Repeat steps 2–7 with other transistors. Note the differences in gain.

You will discover that you cannot always guarantee in which way the outputs will change. This is because the relative values of β and R_{in} may be different for various transistors, and they may change at different rates as the temperature changes.

It is very important to understand that the bias point and the gain can (and will) change with changes in temperature. Obviously, this limits the practical usefulness of this circuit. To design a usable stable amplifier you must make a few changes, as covered in the next few problems.

A Stable Amplifier

11 You can overcome the instability of the transistor amplifier discussed in the first ten problems of this chapter by adding two resistors to the circuit. Figure 8-7 shows an amplifier circuit to which resistors R_E and R_2 have been added. R_2, along with R_1 (labeled R_B in the previous circuits), ensures the stability of the DC bias point.

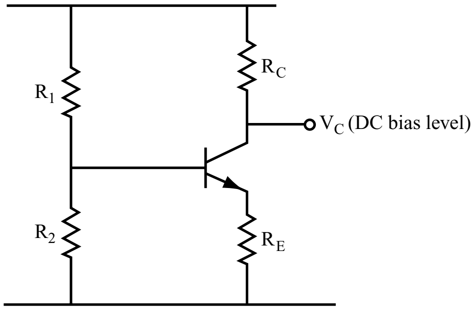

Figure 8-7

By adding the emitter resistor $R_{E,}$ you ensure the stability of the AC gain.

The labels in Figure 8-8 identify the DC currents and voltages present in the circuit. These parameters are used in the next several problems.

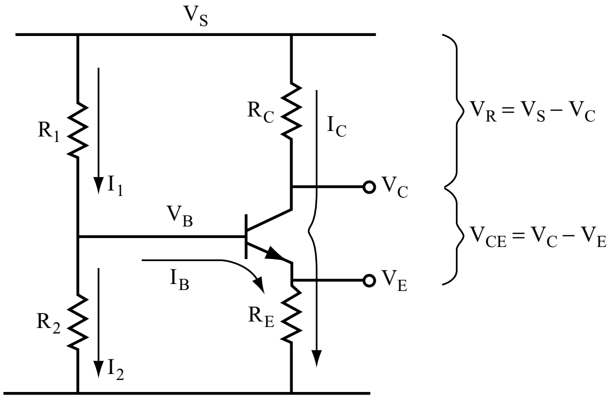

Figure 8-8

Question

In designing an amplifier circuit and choosing the resistor values, there are two goals. What are they?

Answer

A stable DC bias point, and a stable AC gain

12 Look at the gain first. The gain formula for the circuit shown in Figure 8-8 is as follows:

$$A_V = \frac{V_{out}}{V_{in}} = \frac{R_C}{R_E}$$

This is a slight variation on the formula shown in problem 10. (The complex mathematical justification for this is not important right here.) Here, the AC gain has been made independent of the transistor β and input resistance.

As these two parameters vary with temperature, and vary from transistor to transistor, you now have a method of setting the AC gain so that it will be constant regardless of all these variations.

Questions

Use the circuit shown in Figure 8-8 with $R_C = 10\ k\Omega$ and $R_E = 1\ k\Omega$ to answer the following questions.

A. What is the AC voltage gain for a transistor if its $\beta = 100$?

B. What is the gain if $\beta = 500$?

Answers

A. 10

B. 10

13 This problem provides a couple of examples that will help you understand how to calculate voltage gain and the resulting output voltage.

Questions

A. Calculate the voltage gain (A_V) of the amplifier circuit shown in Figure 8-8 if $R_C = 10\text{ k}\Omega$ and $R_E = 1\text{ k}\Omega$. Then, use A_V to calculate the output voltage if the input signal is 2 mV_{pp}.

B. Calculate the voltage gain if $R_C = 1\text{ k}\Omega$ and $R_E = 250$ ohms. Then, use A_V to calculate the output voltage if the input signal is 1 V_{pp}.

Answers

A. $$A_V = \frac{R_C}{R_E} = \frac{10\,\text{k}\Omega}{1\,\text{k}\Omega} = 10$$

$$V_{\text{out}} = 10 \times V_{\text{in}} = 20\,\text{mV}$$

B. $$A_V = \frac{1\,\text{k}\Omega}{250\,\text{ohms}} = 4$$

$$V_{\text{out}} = 4\,V_{\text{pp}}$$

Although the amplifier circuit shown in Figure 8-8 produces stable values of voltage gain, it does not produce high values of voltage gain. For various reasons, this circuit is limited to voltage gains of 50 or less. Later, this chapter discusses an amplifier circuit that can produce higher values of voltage gain.

14 Before you continue, look at the current relationships in the amplifier circuit shown in Figure 8-8 and an approximation that is often made. You can calculate the current flowing through the emitter resistor with the following equation:

$I_E = I_B + I_C$

In other words, the emitter current is the sum of the base and the collector currents.

I_C is much larger than I_B. You can, therefore, assume that the emitter current is equal to the collector current.

$I_E = I_C$

Question

Calculate V_C, V_E, and A_V for the circuit shown in Figure 8-9 with $V_S = 10$ V, $I_C = 1$ mA, $R_C = 1$ kΩ, and $R_E = 100$ ohms.

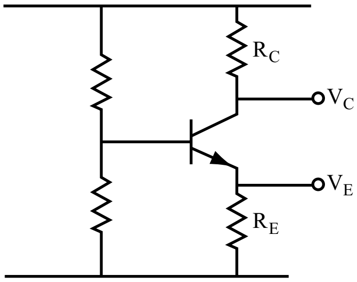

Figure 8-9

Answers

$V_R = 1\,k\Omega \times 1\,mA = 1\,V$

$V_C = V_S - V_R - 10 - 1 = 9\,V$

$V_E = 100\,ohms \times 1\,mA = 0.1\,V$

$$A_V = \frac{R_C}{R_E} = \frac{1\,k\Omega}{100\,ohms} = 10$$

15 For this problem, use the circuit shown in Figure 8-9 with $V_S = 10$ V, $I_C = 1$ mA, $R_C = 2$ kΩ, and $R_E = 1$ kΩ.

Question

Calculate V_C, V_E, and A_V.

Answer

$$V_R = 2\,\text{k}\Omega \times 1\,\text{mA} = 2\,\text{V}$$

$$V_C = 10 - 2 = 8\,\text{V}$$

$$V_E = 1\,\text{k}\Omega \times 1\,\text{mA} = 1\,\text{V}$$

$$A_V = \frac{R_C}{R_E} = \frac{2\,\text{k}\Omega}{1\,\text{k}\Omega} = 2$$

16 For this problem, use the circuit shown in Figure 8-9 with $V_S = 10$ V and $I_C = 1$ mA.

Questions

Find V_C, V_E, and A_V for the following values of R_C and R_E.

A. $R_C = 5$ kΩ, $R_E = 1$ kΩ

B. $R_C = 4.7$ kΩ, $R_E = 220$ ohms

Answers

A. $V_R = 5\,\text{V}$, $V_C = 5\,\text{V}$, $V_E = 1\,\text{V}$, $A_V = 5$

B. $V_R = 4.7\,\text{V}$, $V_C = 5.3\,\text{V}$, $V_E = 0.22\,\text{V}$, $A_V = 21.36$

Biasing

17 In this problem, you see the steps used to calculate the resistor values needed to bias the amplifier circuit shown in Figure 8-10.

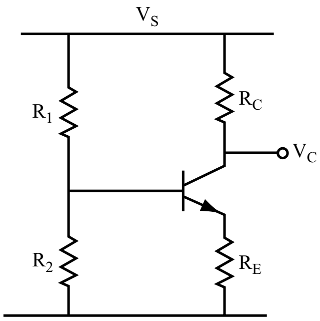

Figure 8-10

You can determine values for R_1, R_2, and R_E that bias the circuit to a specified DC output voltage and a specified AC voltage gain by using the following steps.

Read the following procedure and the relevant formulas first, and then you will walk through an example.

1. Find R_E by using the following:

   $$A_V = \frac{R_C}{R_E}.$$

2. Find V_E by using the following:

   $$A_V = \frac{V_R}{V_E} = \frac{V_S - V_C}{V_E}.$$

3. Find V_B by using the following:

   $$V_B = V_E + 0.7\,V$$

4. Find I_C by using the following:

   $$I_C = \frac{V_S - V_C}{R_C}.$$

5. Find I_B by using the following:

   $$I_B = \frac{I_C}{\beta}$$

6. Find I_2 where I_2 is $10I_B$. (Refer to the circuit shown in Figure 8-8.) This is a convenient rule of thumb that is a crucial step in providing stability to the DC bias point.
7. Find R_2 by using the following:

   $$R_2 = \frac{V_B}{I_2}.$$

8. Find R_1 by using the following:

   $$R_1 = \frac{V_S - V_B}{I_2 + I_B}.$$

9. Steps 7 and 8 might produce nonstandard values for the resistors, so choose the nearest standard values.
10. Use the voltage divider formula to see if the standard values you chose in step 9 result in a voltage level close to V_B found in step 3. (''Close'' means within 10 percent of the ideal.)

This procedure produces an amplifier that works, and results in a DC output voltage and AC gain that are close to those specified at the beginning of the problem.

Questions

Find the values of the parameters specified in each question for the circuit shown in Figure 8-11 if $A_V = 10$, $V_C = 5V$, $R_C = 1\ k\Omega$, $\beta = 100$, and $V_S = 10\ V$.

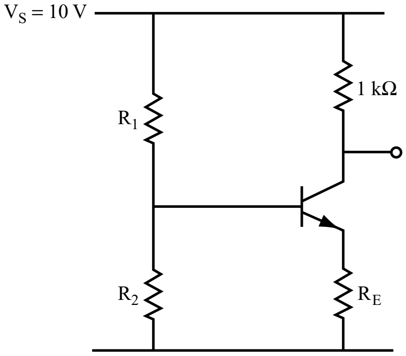

Figure 8-11

Work through steps 1–10, referring to the steps in this problem for formulas as necessary.

1. Find R_E.

   $$A_V = \frac{R_C}{R_E}. \text{ So } R_E = \frac{R_C}{A_V} = \frac{1\,k\Omega}{10} = 100\,\text{ohms}.$$

2. V_E = ____________________________
3. V_B = ____________________________
4. I_C = ____________________________
5. I_B = ____________________________
6. I_2 = ____________________________

7. R_2 = ____________________

8. R_1 = ____________________

9. Choose the standard resistance values that are closest to the calculated values for R_1 and R_2.

R_1 = ____________________; R_2 = ____________________.

10. Using the standard resistance values for R_1 and R_2, find V_B.

V_B = ______________________________________

Answers

You should have found values close to these:

1. 100 ohms
2. 0.5 V
3. 1.2 V
4. 5 mA
5. 0.05 mA
6. 0.5 mA
7. 2.4 kΩ
8. 16 kΩ
9. 2.4 kΩ and 16 kΩ are standard values (they are 5 percent values). Alternative acceptable values would be 2.2 kΩ and 15 kΩ.
10. With 2.4 kΩ and 16 kΩ, $V_B = 1.3$ V. With 2.2 kΩ and 15 kΩ, $V_B = 1.28$ V. Either value of V_B is within 10 percent of the 1.2 V calculated for V_B in step 3.

Figure 8-12 shows an amplifier circuit using the values you calculated in this problem for R_1, R_2, and R_E.

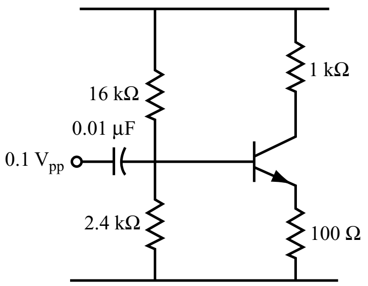

Figure 8-12

18 Follow the steps in problem 17 to answer the following questions.

Questions

Find the values of the parameters specified in each question for the circuit shown in Figure 8-11 if $A_V = 15$, $V_C = 6$ V, $\beta = 100$, $R_C = 3.3$ kΩ, and $V_S = 10$ V.

1. R_E = ______________________
2. V_E = ______________________
3. V_B = ______________________
4. I_C = ______________________
5. I_B = ______________________
6. I_2 = ______________________
7. R_2 = ______________________
8. R_1 = ______________________
9. Choose the standard resistance values that are closest to the calculated values for R_1 and R_2.

   R_1 = ______________________; R_2 = ______________________.
10. Using the standard resistance values for R_1 and R_2, find V_B.

    V_B = ______________________

Answers

These are the values you should have found:

1. 220 ohms
2. 0.27 V
3. 0.97 V (You can use 1 V if you wish.)
4. 1.2 mA
5. 0.012 mA
6. 0.12 mA
7. 8.3 kΩ
8. 68.2 kΩ
9. These are very close to the standard values of 8.2 kΩ and 68 kΩ.
10. 1.08 using the standard values. This is close enough to the value of V_B calculated in question 3.

19 The AC voltage gain for the circuit discussed in problem 18 was 10. Earlier, you learned that the maximum practical gain of the amplifier circuit shown in Figure 8-11 is about 50.

However, in problem 10, you learned that AC voltage gains of up to 500 are possible for the amplifier circuit shown in Figure 8-4. Therefore, by ensuring the stability of the DC bias point, the amplifier has much lower gain than is possible with the transistor amplifier circuit shown in Figure 8-4.

You can make an amplifier with stable bias points without giving up high AC voltage gain by placing a capacitor in parallel with the emitter resistor, as shown in Figure 8-13.

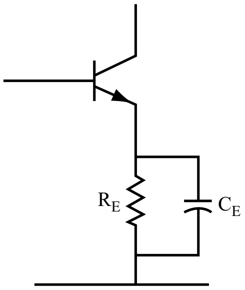

Figure 8-13

If the reactance of this capacitor for an AC signal is significantly smaller than R_E, the AC signal passes through the capacitor rather than the resistor. Therefore, the capacitor is called an *emitter bypass capacitor*. The AC signal ''sees'' a different circuit from the DC, which is blocked by the capacitor and must flow through the resistor. Figure 8-14 shows the different circuits seen by AC and DC signals.

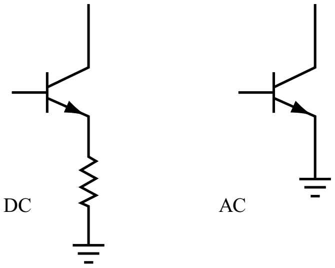

Figure 8-14

The AC voltage gain is now very close to that of the amplifier circuit discussed in problems 1–10.

Questions

A. What effect does the emitter bypass capacitor have on an AC signal?

____

B. What effect does the emitter bypass capacitor have on the AC voltage gain?

____

C. What is the AC voltage gain formula with an emitter bypass capacitor included in the circuit?

Answers

A. It makes the emitter look like a ground, and effectively turns the circuit into the circuit in Figure 8-4.

B. It increases the gain.

C. The same formula used in problem 10:

$$A_V = \beta \times \frac{R_C}{R_{in}}$$

20 You can use the circuit shown in Figure 8-13 when you need as much AC voltage gain as possible. When the volume of AC voltage gain is your priority, predicting the actual amount of gain is usually not important, so the fact that the equation is inexact is unimportant. If you need an accurate amount of gain, then you must use a different type of amplifier circuit that produces lower amounts of gain.

You can find the value of the capacitor C_E using the following steps:

1. Determine the lowest frequency at which the amplifier must operate.
2. Calculate X_C with this formula:

$$X_C = \frac{R_E}{10}$$

3. Calculate C_E with the following formula using the lowest frequency at which the amplifier must operate (determined in step 1):

$$X_C = \frac{1}{2\pi fC}$$

For the following question, use the circuit shown in Figure 8-12, with an emitter bypass capacitor added, as shown in Figure 8-15

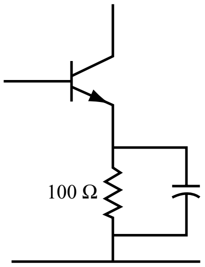

Figure 8-15

Questions

Follow the previous steps to calculate the value of C_E required if the lowest operating frequency of the amplifier is 50 Hz.

1. 50 Hz is the lowest frequency at which the amplifier must operate.
2. $X_C =$ ______________________
3. $C_E =$ ______________________

Answers

$X_C = 10$ ohms

$C_E = 320\ \mu F$ (approximately)

The AC voltage gain formula for an amplifier with an emitter bypass capacitor (Circuit 2 of Figure 8-16) is the same as the AC voltage gain formula for the amplifiers discussed in problems 1–10, where the emitter is directly connected to ground (Circuit 1 of Figure 8-16).

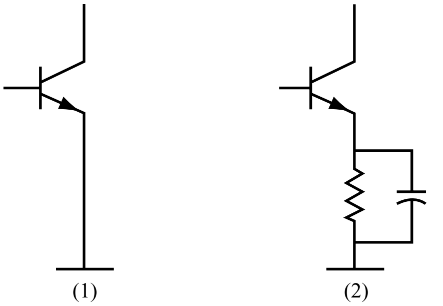

Figure 8-16

The AC voltage gain formula for an amplifier is as follows:

$$A_V = \beta \times \frac{R_C}{R_{in} + R_E}$$

(R_C is being used instead of R_L, because the collector resistor is the total load on the amplifier.)

- *Circuit 1* — Here, $R_E =$ zero, so the AC voltage gain formula is as follows:

  $$A_V = \beta \times \frac{R_C}{R_{in}}.$$

- *Circuit 2* — Here, R_E = zero for an AC signal, because the AC signal is grounded by the capacitor and R_E is out of the AC circuit. Thus, the AC voltage gain formula is as follows:

  $$A_V = \beta \times \frac{R_C}{R_{\text{in}}}.$$

21 To obtain even larger voltage gains, two transistor amplifiers can be *cascaded*. That is, you can feed the output of the first amplifier into the input of the second amplifier. Figure 8-17 shows a two-transistor amplifier circuit, also called a *two-stage amplifier*.

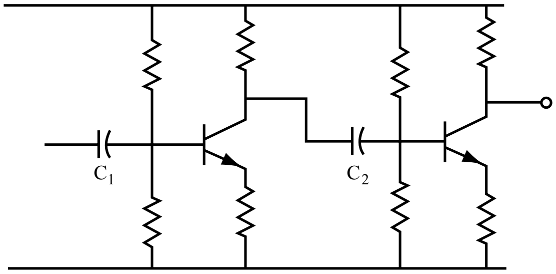

Figure 8-17

You find the total AC voltage gain by multiplying the individual gains. For example, if the first amplifier has an AC voltage gain of 10, and the second has an AC voltage gain of 10, then the overall AC voltage gain is 100.

Questions

A. Suppose you cascade an amplifier with a gain of 15 with one that has a gain of 25. What is the overall gain?

____

B. What is the overall gain if the individual gains are 13 and 17?

____

Answers

A. 375

B. 221

22 Two stage amplifiers can achieve very large AC voltage gains if each amplifier uses an emitter bypass capacitor.

Question

What is the total AC voltage gain if each stage of a two-transistor amplifier has a gain of 100?

Answer

10,000

The Emitter Follower

23 Figure 8-18 shows another type of amplifier circuit.

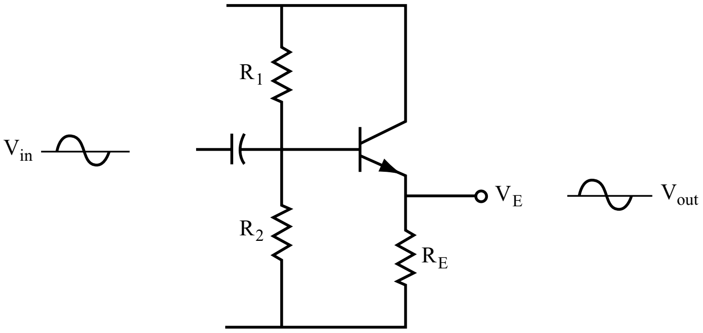

Figure 8-18

Question

How is the circuit shown in Figure 8-18 different from the amplifier circuit discussed in problems 11–18?

Answer

There is no collector resistor, and the output signal is taken from the emitter.

24 The circuit shown in Figure 8-18 is called an *emitter follower* amplifier. (In some cases, it is also called the *common collector* amplifier.)

The output signal has some interesting features:

- The peak-to-peak value of the output signal is almost the same as the input signal. In other words, the circuit gain is slightly less than 1, though in practice it is often considered to be 1.
- The output signal has the same phase as the input signal. It is not inverted. In fact, the output is simply considered to be the same as the input.
- The amplifier has a very high input resistance. Therefore, it draws very little current from the signal source.
- The amplifier has a very low output resistance. Therefore, the signal at the emitter appears to be emanating from a battery or signal generator with a very low internal resistance.

Questions

A. What is the voltage gain of an emitter follower amplifier?

____

B. Is the output signal inverted?

____

C. What is the input resistance of the emitter follower amplifier?

____

D. What is its output resistance?

____

Answers

A. 1

B. No

C. High

D. Low

25 The example in this problem demonstrates the importance of the emitter follower circuit. The circuit shown in Figure 8-19 contains a small AC motor with 100 ohms resistance that is driven by a 10 V_{pp} signal from a generator. The 50 ohm resistor labeled R_G is the internal resistance of the generator.

In this circuit, only 6.7 Vpp is applied to the motor; the rest of the voltage is dropped across R_G.

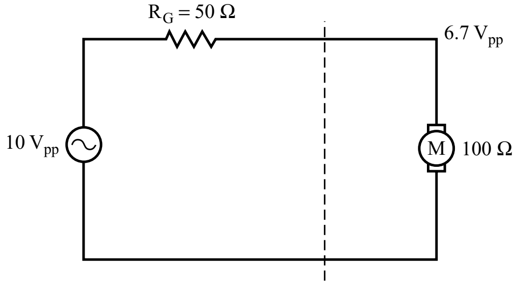

Figure 8-19

Figure 8-20 shows the same circuit, with a transistor connected between the generator and the motor in an emitter follower configuration.

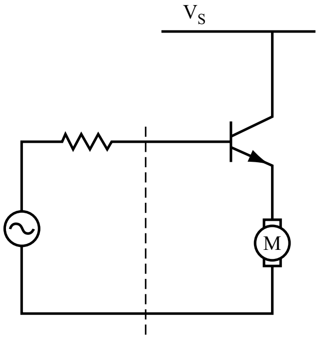

Figure 8-20

You can use the following formula to calculate the approximate input resistance of the transistor.

$R_{in} = \beta \times R_E = 100 \times 100\,\Omega = 10{,}000\,\Omega$ (assuming that $\beta = 100$)

The 10 V_{pp} from the generator is divided between the 10,000 ohm input resistance of the transistor and the 50 ohm internal resistance of the generator. Therefore, there is no significant voltage drop across R_G, and the full 10 V_{pp} is applied to the base of the transistor. The emitter voltage remains at 10 V_{pp}.

Also, the current through the motor is now produced by the power supply and not the generator, and the transistor looks like a generator with a very low internal resistance.

This internal resistance (R_O) is called the *output impedance* of the emitter follower. You can calculate it using this formula:

$$R_O = \frac{\text{internal resistance of generator}}{\beta}$$

for the circuit shown in Figure 8-20, if $R_G = 50$ ohms and $\beta = 100$, $R_O = 0.5$ ohms. Therefore, the circuit shown in Figure 8-20 is effectively a generator with an internal resistance of only 0.5 ohms driving a motor with a resistance of 100 ohms. Therefore, the output voltage of 10 V_{pp} is maintained across the motor.

Questions

A. What is the emitter follower circuit used for in this example?

B. Which two properties of the emitter follower are useful in circuits?

Answers

A. To drive a load that could not be driven directly by a generator

B. High input resistance and its low output resistance

26 The questions in this problem apply to the emitter follower circuit discussed in problems 23–25.

Questions

A. What is the approximate gain of an emitter follower circuit?

B. What is the phase of the output signal compared to the phase of the input signal?

C. Which has the higher value, the input resistance or the output resistance?

D. Is the emitter follower more effective at amplifying signals or at isolating loads?

Answers

A. 1

B. The same phase

C. The input resistance

D. Isolating loads

27 You can design an emitter follower circuit using these steps:

1. Specify V_E. This is a DC voltage level, which is usually specified as half the supply voltage.
2. Find V_B. Use $V_B = V_E + 0.7$ V.
3. Specify R_E. Often this is a given factor, especially if it is a motor or other load that is being driven.
4. Find I_E by using the following:

   $$I_E = \frac{V_E}{R_E}$$

5. Find I_B by using the following:

   $$I_B = \frac{I_E}{\beta}$$

6. Find I_2 by using $I_2 = 10I_B$.
7. Find R_2 by using the following:

   $$R_2 = \frac{V_B}{I_2}.$$

8. Find R_1 by using the following:

   $$R_1 = \frac{V_S - V_B}{I_2 + I_B}.$$

   Usually, I_B is small enough to be dropped from this formula.
9. Choose the nearest standard values for R_1 and R_2.
10. Check that these standard values give a voltage close to V_B. Use the voltage divider formula.

A simple design example illustrates this procedure. Use the values shown in the circuit in Figure 8-21 for this problem.

Questions

Work through Steps 1–10 to find the values of the two bias resistors.

1. V_E = __________
2. V_B = __________
3. R_E = __________

4. I_E = __________
5. I_B = __________
6. I_2 = __________
7. R_2 = __________
8. R_1 = __________
9. The nearest standard values are as follows:
   R_1 = __________ and R_2 = __________
10. V_B = __________

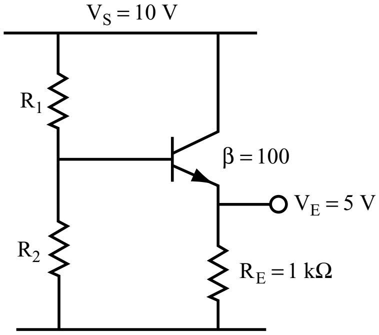

Figure 8-21

Answers

Your answers should be close to these values.

1. 5 V (This was given in Figure 8-21.)
2. 5.7 V
3. 1 kΩ (This was given in Figure 8-21.)
4. 5 mA
5. 0.05 mA
6. 0.5 mA
7. 11.4 kΩ
8. 7.8 kΩ
9. The nearest standard values are 8.2 kΩ and 12 kΩ.

10. The standard resistor values result in $V_B = 5.94$ V. This is a little higher than the V_B calculated in step 2, but it is acceptable.

V_E is set by the biasing resistors. Therefore, it is not dependent upon the value of R_E. Almost any value of R_E can be used in this circuit. The minimum value for R_E is obtained by using this simple equation:

$$R_E = \frac{10\,R_2}{\beta}$$

Analyzing an Amplifier

28 Up to now, the emphasis has been on designing a simple amplifier and an emitter follower. This section shows how to ''analyze'' a circuit that has already been designed. To ''analyze'' in this case means to calculate the collector DC voltage (the bias point) and find the AC gain. This procedure is basically the reverse of the design procedure.

Start with the circuit shown in Figure 8-22.

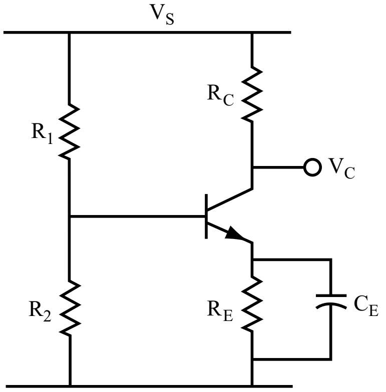

Figure 8-22

Here are the steps you use to analyze a circuit:

1. Find V_B by using the following:

$$V_B = V_S \times \frac{R_2}{R_1 + R_2}.$$

2. Find V_E by using $V_E = V_B - 0.7$ V.

3. Find I_C by using the following:

$$I_C = \frac{V_E}{R_E}.$$

Note that $I_C = I_E$.

4. Find V_R by using $V_R = R_C \times I_C$.
5. Find V_C by using $V_C = V_S - V_R$. This is the bias point.
6. Find A_V by using the following:

$$A_V = \frac{R_C}{R_E} \text{ or } A_V = \beta \times \frac{R_C}{R_{in}}$$

When you use the second formula, you must find the value of R_{in} (or h_{ie}) on the manufacturer's datasheets for the transistor.

Use the circuit shown in Figure 8-23 for following questions. For these questions, use $\beta = 100$, $R_{in} = 2$ kΩ and the values given in the circuit drawing.

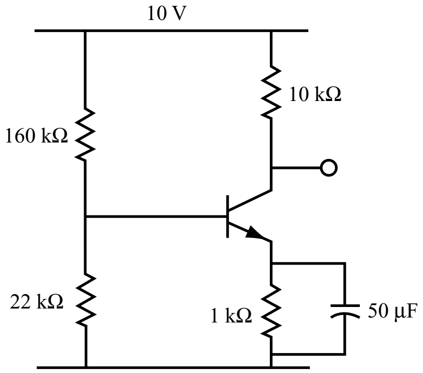

Figure 8-23

Questions

Calculate V_B, V_E, I_C, V_R, V_C, and A_V using steps 1–6 of this problem.

1. V_B = __________
2. V_E = __________
3. I_C = __________
4. V_R = __________
5. V_C = __________
6. A_V = __________

Answers

1. $V_B = 10 \times \dfrac{22\,\text{k}\Omega}{160\,\text{k}\Omega + 22\,\text{k}\Omega} = 1.2\,\text{V}$
2. $V_E = 1.2 - 0.7 = 0.5$ V
3. $I_C = \dfrac{0.5\,\text{V}}{1\,\text{k}\Omega} = 0.5\,\text{mA}$
4. $V_R = 10\ \text{k}\Omega \times 0.5\ \text{mA} = 5$ V
5. $V_C = 10$ V − 5 V = 5 V (This is the bias point.)
6. With the capacitor:

   $A_V = 100 \times \dfrac{10\,\text{k}\Omega}{2\,\text{k}\Omega} = 500$ (a large gain)

   Without the capacitor:

   $A_V = \dfrac{10\,\text{k}\Omega}{1\,\text{k}\Omega} = 10$ (a small gain)

29 You can determine the lowest frequency the amplifier will satisfactorily pass by following these simple steps:

1. Determine the value of R_E.
2. Calculate the frequency at which $X_C = R_E/10$. Use the capacitor reactance formula. (This is one of those ''rules of thumb'' that *can* be mathematically justified, and gives reasonably accurate results in practice.)

Questions

For the circuit shown in Figure 8-23, find the following.

A. R_E = __________

B. f = __________

Answers

A. $R_E = 1$ kΩ (given in the circuit diagram)

B. So, we will set $X_C = 100$ ohms, and use this formula:

$$X_C = \frac{1}{2\pi f C}$$

$$100\,\text{ohms} = \frac{0.16}{f \times 50\,\mu\text{F}} \quad \text{since}\ 0.16 = \frac{1}{2\pi}$$

So:

$$f = \frac{0.16}{100 \times 50 \times 10^{-6}} = 32\,\text{Hz}$$

30 For the circuit shown in Figure 8-24 follow the steps given in problems 28 and 29 to answer the following questions.

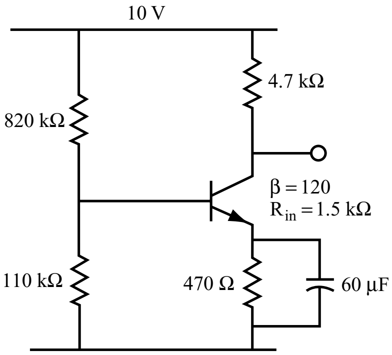

Figure 8-24

Questions

1. V_B = __________
2. V_E = __________
3. I_C = __________
4. V_R = __________
5. V_C = __________
6. With capacitor: A_V = __________

   Without capacitor: A_V = __________
7. Low frequency check: f = __________

Answers

Your answers should be close to these.

1. 1.18 V
2. 0.48 V
3. 1 mA
4. 4.7 V
5. 5.3 V (bias point)
6. With capacitor: 376
   Without capacitor: 10
7. 57 Hz (approximately)

The JFET as an Amplifier

31 Chapter 3 discussed the junction field effect transistor (JFET) in problems 29–32 and Chapter 4 discussed JFET in problems 37–41. You may wish to review these problems before answering the questions in this problem. Figure 8-25 shows a typical biasing circuit for a JFET.

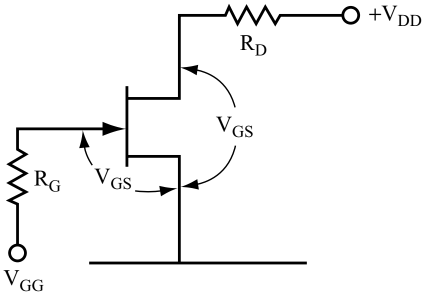

Figure 8-25

Questions

A. What type of JFET is depicted in the circuit?

B. What value of V_{GS} would you need to turn the JFET completely ON?

C. What drain current flows when the JFET is completely ON?

D. What value of V_{GS} would you need to turn the JFET completely OFF?

E. When a JFET is alternately turned completely ON and OFF in a circuit, what type of component are you using the JFET as?

Answers

A. N-channel JFET.

B. $V_{GS} = 0$ V to turn the JFET completely ON.

C. Drain saturation current (I_{DSS}).

D. V_{GS} should be a negative voltage for the N-channel JFET to turn it completely OFF. The voltage must be larger than or equal to the cutoff voltage.

E. The JFET is being used as a switch.

32 You can use a JFET to amplify AC signals by biasing the JFET with a gate to source voltage about halfway between the ON and OFF states. You can find the drain current that flows in a JFET biased to a particular V_{GS} by using the following equation for the transfer curve:

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)^2$$

In this equation, I_{DSS} is the value of the drain saturation current and $V_{GS(off)}$ is the gate to source voltage at cutoff. Both of these are indicated on the transfer curve shown in Figure 8-26.

For the transfer curve shown in Figure 8-26, $I_{DSS} = 12$ mA and $V_{GS(off)} = -4$ V. Setting the bias voltage at $V_{GS} = -2$ V returns the following value for the drain current:

$$I_D = 12\,\text{mA} \times \left(1 - \frac{-2}{-4}\right)^2 = 12\,\text{mA} \times (0.5)^2 = 3\,\text{mA}$$

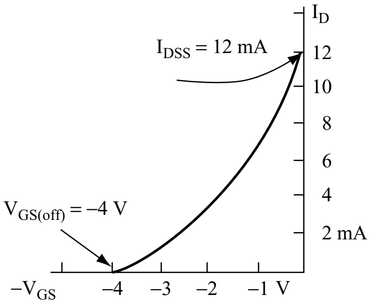

Figure 8-26

Questions

Calculate the drain current for the following:

A. $V_{GS} = -1.5$

B. $V_{GS} = -0.5$ volts.

Answers

A. 4.7 mA

B. 9.2 mA

NOTE **Datasheets give a wide range of possible I_{DSS} and $V_{GS(off)}$) values for a given JFET. You may have to resort to actually measuring these with the method shown in Chapter 4, problem 38.**

33 For the circuit shown in Figure 8-25, you choose the value of the drain to source voltage, V_{DS}, then calculate the value of the load resistor, R_D, by using the following equation:

$$R_D = \frac{(V_{DD} - V_{DS})}{I_D}$$

For this problem, use $I_D = 3$ mA, and a drain supply voltage (V_{DD}) of 24 V. Calculate the value of R_D that results in the specified value of V_{DS}; this is also the DC output voltage of the amplifier.

Question

Calculate the value of R_D that will result in $V_{DS} = 10$ V.

Answer

$$R_D = \frac{(V_{DD} - V_{DS})}{I_D} = \frac{(24\,V - 10V)}{3\,mA} = \frac{14\,V}{3\,mA} = 4.67\,k\Omega$$

34 The circuit shown in Figure 8-27 (which is referred to as a *JFET common source amplifier*) applies a 0.5 V_{pp} sine wave to the gate of the JFET, and produces an amplified sine wave output from the drain.

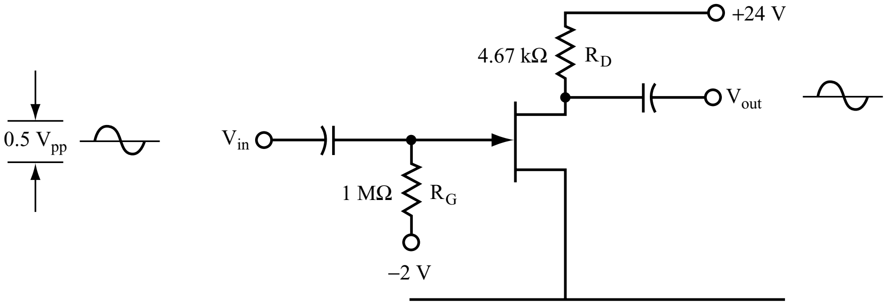

Figure 8-27

The input sine wave is added to the −2 V bias applied to the gate of the JFET. Therefore, V_{GS} varies from −1.75 to −2.25 V.

Question

Calculate I_D, using the formula in problem 32, for the maximum and minimum values of V_{GS}.

Answer

For $V_{GS} = -1.75\,V$, $I_D = 3.8\,mA$.
For $V_{GS} = -2.25\,V$, $I_D = 2.3\,mA$.

35 As the drain current changes, V_{RD} (the voltage drop across resistor R_D) also changes.

Question

For the circuit shown in Figure 8-27, calculate the values of V_{RD} for the maximum and minimum values of I_D you calculated in problem 34.

Answer

For $I_D = 3.8\,\text{mA}, V_{RD} = 3.8\,\text{mA} \times 4.67\,\text{k}\Omega = 17.7\,\text{V}$
For $I_D = 2.3\,\text{mA}, V_{RD} = 2.3\,\text{mA} \times 4.67\,\text{k}\Omega = 10.7\,\text{V}$

This corresponds to a 7 V_{pp} sine wave.

36 As the voltage drop across R_D changes, the output voltage also changes.

Question

For the circuit shown in Figure 8-27, calculate the values of V_{out} for the maximum and minimum values of V_{RD} you calculated in problem 35.

Answer

For $V_{RD} = 17.7\,\text{V},\ V_{out} = V_{DD} - V_{RD} = 24\,\text{V} - 17.7\,\text{V} = 6.3\,\text{V}$
For $V_{RD} = 10.7\,\text{V},\ V_{out} = V_{DD} - V_{RD} = 24\,\text{V} - 10.7\,\text{V} = 13.3\,\text{V}$

Therefore, the output signal is a 7 V_{pp} sine wave.

37 Table 8-1 shows the results of the calculations made in problems 34–36 including the DC bias point.

Table 8-1

V_{GS}	I_D	V_{RD}	V_{OUT}
−1.75 V	3.8 mA	17.7 V	6.3 V
−2.0 V	3.0 mA	14.0 V	10.0 V
−2.25 V	2.3 mA	10.7 V	13.3 V

Question

What are some characteristics of the AC output signal?

Answer

The output signal is a 7 V_{pp} sine wave with the same frequency as the input sine wave. Note that, as the input voltage on V_{GS} increases (toward

0 V), the output decreases. As the input voltage decreases (becomes more negative), the output voltage increases. This means that the output is 180 degrees out of phase with the input.

38 You can calculate the AC voltage gain for the amplifier discussed in problems 34–37 by using the following formula:

$$A_v = \frac{-V_{\text{out}}}{V_{\text{in}}}$$

The negative sign in this formula indicates that the output signal is 180 degrees out of phase from the input signal.

Question

Calculate the AC voltage gain for the amplifier discussed in problems 34–37.

Answer

$$\mathrm{A_V} = \frac{-7\,\mathrm{V_{pp}}}{0.5\,\mathrm{V_{pp}}} = -14$$

39 You can also calculate the AC voltage gain by using the following formula:

$$\mathrm{A_v} = -(\mathrm{g_m})(\mathrm{R_D})$$

In this equation, $\mathrm{g_m}$ is the *transconductance* and is a property of the JFET. It is also called the *forward transfer admittance*. A typical value for $\mathrm{g_m}$ is usually provided for JFETs in the manufacturers' datasheet. You can also use the data in Table 8-1 to calculate $\mathrm{g_m}$ using the following formula:

$$g_m = \frac{\Delta \mathrm{I}_D}{\Delta \mathrm{V}_{GS}}$$

In this equation, Δ indicates the change or variation in $\mathrm{V_{GS}}$ and the corresponding drain current. The unit for transconductance is mhos.

Questions

A. Using the data from Table 8-1, what is the value of $\mathrm{g_m}$ for the JFET used in the amplifier?

____

B. What is the corresponding AC voltage gain?

____

Answers

A. $gm = \dfrac{1.5\,\text{mA}}{0.5\,\text{V}} = 0.003\,\text{mhos}$

B. $A_v = -(0.003)(4670) = -14$, the same result you found in problem 38.

40 Design a JFET common source amplifier using a JFET with $I_{DSS} = 14.8$ mA and $V_{GS(off)} = -3.2$ V. The input signal is 40 mV_{pp}. The drain supply is 24 V.

Questions

A. Determine the value of V_{GS} that will bias the JFET at a voltage near the middle of the transfer curve.

____

B. Calculate the drain current when V_{GS} is at the value determined in step A, using the formula in problem 32.

____

C. Choose a value of V_{DS} and calculate the value of R_D using the formula in problem 33.

____

D. Calculate the maximum and minimum values of V_{GS} that result from the input signal, and the corresponding values of drain current using the procedure in problem 34.

____

E. Calculate the maximum and minimum values of V_{out} that result from the input signal using the procedures in problems 35 and 36.

____

F. Calculate the gain of the amplifier.

____

Answers

A. $V_{GS} = -1.6\ V$

B. $I_D = 3.7\ mA$

C. For $V_{DS} = 10\ V$,

$$R_D = \frac{14\,V}{3.7\,mA} = 3780\,\text{ohms}$$

D. V_{GS} will vary from −1.58 to −1.62 V. Use the formula to calculate values of drain current. I_D will vary from 3.79 to 3.61 mA.

E. V_{RD} will vary from 14.3 to 13.6 V. Therefore, V_{out} will vary from 9.7 to 10.4 V.

F. $A_v = \frac{-0.7}{0.04} = -17.5$

41 Use the results of problem 40, question D, to answer the following question.

Question

Calculate the transconductance of the JFET and the AC voltage gain using the formulas discussed in problem 39.

Answer

$$g_m = \frac{\Delta I_D}{\Delta V_{GS}} = \frac{0.18\,mA}{40\,mV} = 0.0045\,\text{mhos}$$

$$A_v = -(g_m)(R_D) = -(0.0045)(3780) = -17$$

This is very close to the value you found in problem 40, question F.

42 Figure 8-28 shows a JFET amplifier circuit that uses one power supply, rather than separate power supplies for the drain and gate used in the amplifier discussed in problems 34–41.

The DC voltage level of the gate is zero because the gate is tied to ground through R_G. Therefore, the voltage drop across R_S becomes the gate to source voltage. To design the circuit, you must find values for both R_S and R_D. Use the same bias point for this problem as you used for the amplifier discussed in problems 34–41: $V_{GS} = -2$ V and $I_D = 3$ mA. Follow these steps:

1. Calculate R_S, using the following formula, recognizing that $V_{RS} = V_{GS}$:

$$R_S = \frac{V_{RS}}{I_D} = \frac{V_{GS}}{I_D}$$

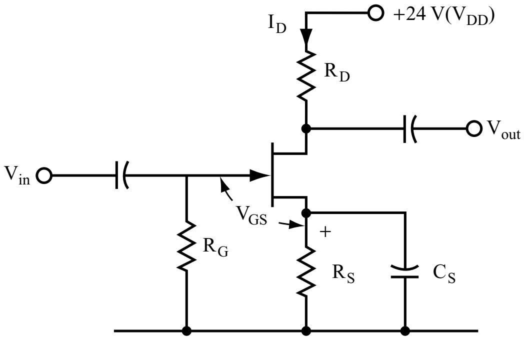

Figure 8-28

2. Calculate R_D using the following formula, using $V_{DS} = 10$ V, the same value you used for the amplifier discussed in problems 34–41:

$$R_D = \frac{(V_{DD} - V_{DS} - V_S)}{I_D}$$

3. Calculate X_{CS} using the following formula:

$$X_{CS} = \frac{R_S}{10}$$

Then, calculate C_S using the following formula:

$$X_{CS} = \frac{1}{2\pi f C_S}$$

4. Calculate the peak-to-peak output voltage using the procedures shown in problems 34–36.
5. Calculate the AC voltage gain using this formula:

$$A_V = \frac{-V_{out}}{V_{in}}$$

NOTE Choose the value of C_S so that its reactance is less than 10 percent of R_S at the lowest frequency you need to amplify. The DC load for the JFET is R_D plus R_S. The AC load is R_D only, because C_S bypasses the AC signal around R_S, which keeps the DC operating point stable. The use of C_S reduces the gain slightly because you now use a smaller R_D to calculate the AC voltage swings at the output.

Questions

A. What is the value of R_S?

B. What is the value of R_D?

C. What is the value of C_S? Assume f = 1 kHz.

D. Calculate the peak-to-peak V_{out} for $V_{in} = 0.5\ V_{pp}$.

E. What is the voltage gain?

Answers

A. $R_S = \dfrac{2\,V}{3\,mA} = 667\,ohms$

B. $R_D = \dfrac{12\,V}{3\,mA} = 4\,k\Omega$

C. $X_{CS} = 66.7$ ohms, $C_S = 2.4\ \mu F$

D. The AC drain current will still vary from 3.8 to 2.3 mA, as in problem 37. The voltage across R_D is now 6 V_{pp}, because R_D is 4 kΩ. The output voltage is also 6 V_{pp}.

E. $A_V = \dfrac{-6}{0.5} = -12$ The gain is 12.

The Operational Amplifier

43 The operational amplifier (op-amp) in use today is actually an integrated circuit (IC). This means that the device has numerous transistors and other components constructed on a very small silicon chip. These IC op-amps are much smaller and, therefore, more practical than an amplifier with equivalent performance that is made with discrete components.

It is possible to purchase op-amps in different case configurations. Some of these configurations are the TO metal package, the flat pack, and the DIP package. You will also find 2 op-amps (dual) or 4 op-amps (quad) in a single IC.

Their size, low cost, and wide range of applications have made op-amps so common today that they are thought of as a circuit device or component in and of themselves, even though a typical op-amp may contain 20 or more transistors in its design. The characteristics of op-amps very closely resemble those of an ideal amplifier. Following are these characteristics:

- High input impedance (does not require input current)
- High gain (used for amplifying small signal levels)
- Low output impedance (not affected by the load)

Questions

A. What are the advantages of using op-amps?

____

B. Why are op-amps manufactured using IC techniques?

____

Answers

A. Small size, low cost, wide range of applications, high input impedance, high gain, and low output impedance.

B. Because of the large numbers of transistors and components that are required in the design of an op-amp, they must be constructed on a single, small silicon chip using IC manufacturing techniques to be of a reasonable size.

44 Figure 8-29 shows the schematic symbol for an op-amp.

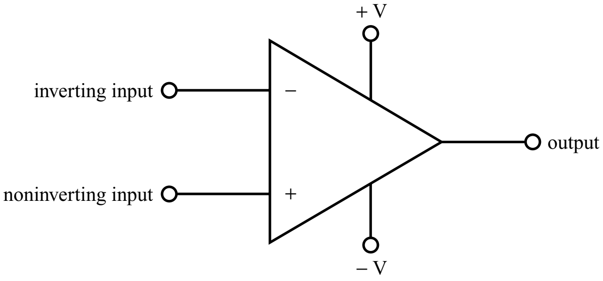

Figure 8-29

An input at the inverting input results in an output that is 180 degrees out of phase with the input. An input at the non-inverting input results in an output that is in phase with the input. Both positive and negative voltage supplies are required, and the datasheet will specify their values for the particular op-amp you are using. Datasheets usually contain circuit diagrams showing how you should connect external components to the op-amp for specific applications. These circuit diagrams (showing how a particular op-amp can be used for various applications) can be very useful to the designer or the hobbyist.

Questions

A. How many terminals does the op-amp require, and what are their functions?

B. How is the output related to the input when the input is connected to the inverting input?

Answers

A. 5 — two input terminals, one output terminal, two power supply terminals.

B. The output is 180 degrees out of phase with the input.

45 Figure 8-30 shows a basic op-amp circuit. The input signal is connected to an inverting input, as indicated by the negative sign. Therefore, the output signal will be 180 degrees out of phase with the input.

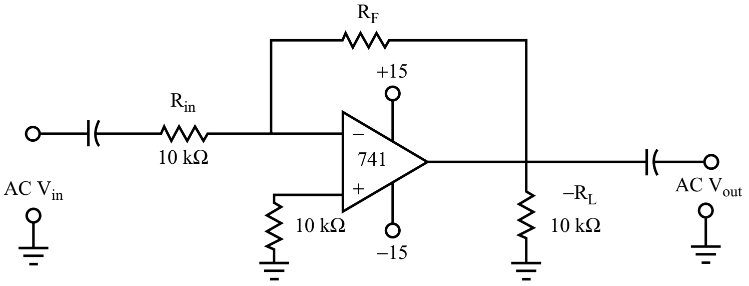

Figure 8-30

You can find the AC voltage gain for the circuit using the following equation:

$$A_v = \frac{-R_F}{R_{in}}$$

Resistor R_F is called a *feedback resistor*, because it forms a feedback path from the output to the input. Many op-amp circuits use a feedback loop. Because the op-amp has such a high gain, it is easy to saturate it (at maximum gain) with small voltage differences between the two input terminals. The feedback loop allows the operation of the op-amp at lower gains, allowing a wider range of input voltages. When designing a circuit, you can choose the value of the feedback resistor to achieve a specific voltage gain. The role of the capacitors in the diagram is to block DC voltages.

Questions

A. Calculate the value of R_F that would give the amplifier an AC voltage gain of 120.

____

B. Calculate A_C V_{out} if AC V_{in} is 5 mV_{rms}.

____

Answers

A. $R_F = 120 \times 10\,k\Omega = 1.2\,M\Omega$

B. $V_{out} = 120 \times 5\,mV = 0.6\,V_{rms}$

The output signal is inverted with respect to the input signal.

46 Use the op-amp circuit shown in Figure 8-30 to build an amplifier with an output voltage of 12 V_{pp}, an AC voltage gain of 50, and with $R_{in} = 6.8\,k\Omega$.

Questions

A. Calculate the value of R_F.

____

B. Calculate the value of V_{in} required to produce the output voltage specified earlier.

____

Answers

A. $R_F = 50 \times 6.8\,k\Omega = 340\,k\Omega$

B. $V_{in} = \dfrac{12\,V_{pp}}{50} = 0.24\,V_{pp}$ or $0.168\,V_{rms}$

Summary

This chapter introduced the most common types of amplifiers in use today: the common emitter BJT, the common source JFET, and the op-amp. At best, this chapter has only scratched the surface of the world of amplifiers. In fact, there are many variations and types of amplifiers. Still, the terminology and design approach you learned here should give you a basic foundation for further study.

Following are the key skills you gained in this chapter:

- How to design a simple amplifier when the bias point and the gain are specified
- How to do the same for an emitter follower
- How to analyze a simple amplifier circuit

Self-Test

These questions test your understanding of the material presented in this chapter. Use a separate sheet of paper for your diagrams or calculations. Compare your answers with the answers provided following the test.

1. What is the main problem with the amplifier circuit shown in Figure 8-1?

____

2. What is the gain formula for that circuit?

____

3. Does it have a high or low gain?

Use the circuit shown in Figure 8-31 for questions 4–8.

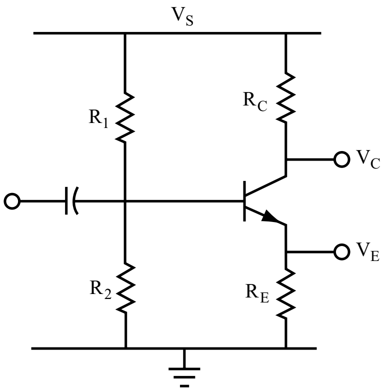

Figure 8-31

4. Design an amplifier so that the bias point is 5 V and the AC voltage gain is 15. Assume $\beta = 75$, $R_{in} = 1.5\ k\Omega$, $V_S = 10$ V, and $R_C = 2.4\ k\Omega$. Add capacitor C_E to the circuit and calculate a suitable value to maintain maximum AC voltage gain at 50 Hz. What is the approximate value of this gain?

---

5. Repeat question 4 with these values: $V_S = 28$ V, $\beta = 80$, $R_{in} = 1\ k\Omega$, and $R_C = 10\ k\Omega$. The bias point should be 14 V, and the AC voltage gain 20.

---

6. Repeat question 4 with these values: $V_S = 14$ V, $\beta = 250$, $R_{in} = 1\ k\Omega$, and $R_C = 15\ k\Omega$. The bias point should be 7 V, and the AC voltage gain 50.

---

7. Design an emitter follower amplifier given that $V_S = 12$ V, $R_E = 100$ ohms, $\beta = 35$, $V_E = 7$ V, $R_C = 0$ ohms.

---

8. Design an emitter follower amplifier given that $V_S = 28$ V, $R_E = 100$ ohms, $\beta = 35$, $V_E = 7$ V, $R_C = 0$ ohms.

In questions 9–11, the resistance and β values are given. Analyze the circuit to find the bias point and the gain.

9. $R_1 = 16\ k\Omega$, $R_2 = 2.2\ k\Omega$, $R_E = 100$ ohms, $R_C = 1\ k\Omega$, $\beta = 100$, $V_S = 10$ V

10. $R_1 = 36\ k\Omega$, $R_2 = 3.3\ k\Omega$, $R_E = 110$ ohms, $R_C = 2.2\ k\Omega$, $\beta = 50$, $V_S = 12$ V

11. $R_1 = 2.2\ k\Omega$, $R_2 = 90\ k\Omega$, $R_E = 20$ ohms, $R_C = 300\ k\Omega$, $\beta = 30$, $V_S = 50$ V

12. The circuits from questions 4 and 5 are connected to form a two-stage amplifier. What is the gain when there is an emitter bypass capacitor for both transistors? When the capacitor is not used in either of them?

13. Design a JFET amplifier using the circuit shown in Figure 8-27. The characteristics of the JFET are $I_{DSS} = 20$ mA and $V_{GS(off)} = -4.2$ V. The desired value of V_{DS} is 14 V. Find the value of R_D.

14. If the transconductance of the JFET used in question 13 is 0.0048 mhos, what is the voltage gain?

15. If the desired output is 8 V_{pp} for the JFET of questions 13 and 14, what should the input be?

16. Design a JFET amplifier using the circuit in Figure 8-28. The JFET characteristics are $I_{DSS} = 16$ mA and $V_{GS(off)} = -2.8$ V. Using a V_{DS} of 10 V, find the values of R_S, C_S, and R_D.

17. If the input to the JFET in question 16 is 20 mV_{pp}, what is the AC output voltage and what is the gain?

18. For the op-amp circuit shown in Figure 8-30, what is the output voltage if the input is 50 mV and the feedback resistor is 750 kΩ?

Answers to Self-Test

If your answers do not agree with those provided here, review the problems in parentheses before you go on to the next chapter.

1.	Its bias point is unstable, and its gain varies with temperature. Also, you cannot guarantee what the gain will be.	(problem 10)
2.	$A_V = \beta \times \frac{R_C}{R_{in}}$	(problem 10)
3.	Usually the gain is quite high.	(problem 10)
For Numbers 4–6, suitable values are given. Yours should be close to these.		
4.	$R_1 = 29$ kΩ, $R_2 = 3.82$ kΩ, $R_E = 160$ ohms, $C_E = 200$ μF, $A_V = 120$	(problems 17–22)
5.	$R_1 = 138$ kΩ, $R_2 = 8$ kΩ, $R_E =$ 500 ohms, $C_E = 64$ μF, $A_V = 800$	(problems 17–22)
6.	$R_1 = 640$ kΩ, $R_2 = 45$ kΩ, $R_E =$ 300 ohms, $C_E = 107$ μF, $A_V = 750$	(problems 17–22)
7.	$R_1 = 8$ kΩ; $R_2 = 11.2$ kΩ	(problem 27)
8.	$R_1 = 922$ ohms; $R_2 = 385$ ohms	(problem 27)
9.	$V_C = 5$ V, $A_V = 10$	(problems 28–30)
10.	$V_C = 6$ V, $A_V = 20$	(problems 28–30)
11.	$V_C = 30$ V, $A_V = 15$	(problems 28–30)
12.	When the capacitor is used, $A_V = 120 \times 800 = 96{,}000$.	(problems 17–22)
	When the capacitor is not used, $A_V = 15 \times 20 = 300$.	
13.	Use $V_{GS} = -2.1$ V, then $I_D = 5$ mA, $R_D = 2$ kΩ.	(problems 31–33)
14.	$A_v = -9.6$ mV_{pp}	(problem 39)
15.	$V_{in} = 83$ mV_{pp}	(problem 38)

16.	Use $V_{GS} = -1.4$ V, then $I_D = 4$ mA.	(problem 42)
	$R_S = 350$ ohms	
	$C_S = 4.5\ \mu F$ (assume $f = 1$ kHz)	
	$R_D = 3.15\ k\Omega$	
17.	V_{GS} varies from -1.39 to -1.41 V, I_D varies from 4.06 to 3.94 mA, V_{out} will be 400 mV_{pp}, $A_v = \frac{-400}{20} = -20$	(problem 42)
18.	$A_v = -75$, $V_{out} = 3.75$ V	(problem 45)

CHAPTER 9

Oscillators

This chapter introduces you to oscillators. An *oscillator* is a circuit that produces a continuous output signal. There are many types of oscillator circuits that are used extensively in electronic devices. Oscillators can produce a variety of different output signals, such as sine waves, square waves, or triangle waves.

When the output signal of an oscillator is a sine wave of constant frequency, the circuit is called a *sine wave oscillator*. Radio and television signals are sine waves transmitted through the air, and the 120 V AC from the wall plug is a sine wave, as are many test signals used in electronics.

This chapter introduces three basic sine wave oscillators. They all rely on resonant LC circuits as described in Chapter 7 to set the frequency of the sine wave.

When you complete this chapter, you will be able to do the following:

- Recognize the main elements of an oscillator.
- Differentiate between positive and negative feedback.
- Specify the type of feedback that causes a circuit to oscillate.
- Specify at least two methods of obtaining feedback in an oscillator circuit.
- Understand how resonant LC circuits set the frequency of an oscillator.
- Design a simple oscillator circuit.

Understanding Oscillators

1 An oscillator can be divided into three definite sections: (1) an amplifier; (2) the feedback connections; and (3) the components that set frequency.

The amplifier replaces the switch in the basic oscillator circuit, introduced in problem 35 of Chapter 7 (Figure 7-32).

Question

Draw an oscillator circuit, and label the parts. Use a separate sheet of paper for your drawing.

Answer

See Figure 9-1.

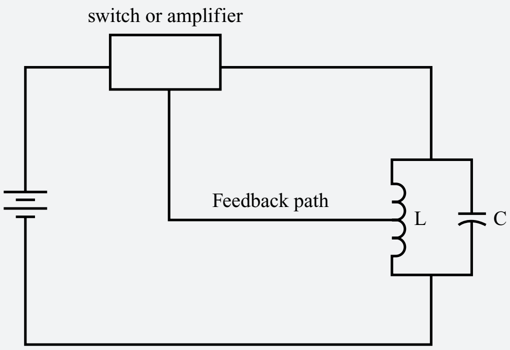

Figure 9-1

2 When you connect the output of an amplifier to its input, you get feedback. If the feedback is ‘‘out of phase’’ with the input, as shown in Figure 9-2, then the feedback is *negative*.

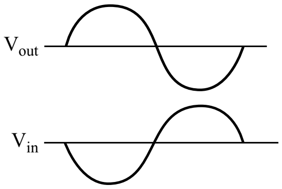

Figure 9-2

When the signal from the collector is fed back to the base of the transistor through a feedback resistor (R_f), as in the circuit shown in Figure 9-3, the feedback signal is out of phase with the input signal. Therefore, the feedback is negative.

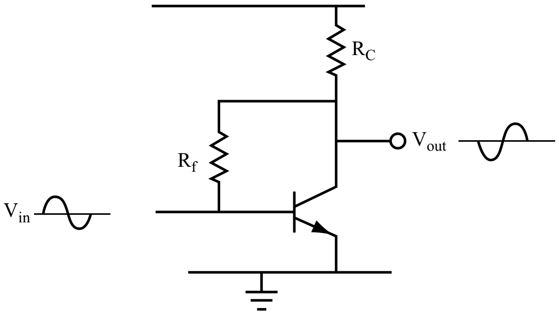

Figure 9-3

Negative feedback is used to stabilize the operation of an amplifier by doing the following:

- Preventing the DC bias point and gain of an amplifier from being affected by changes in temperature
- Reducing distortion in amplifiers, thereby improving the quality of the sound

Questions

A. Why would feedback signals be used in quality audio amplifiers?

____

B. What kind of feedback do they have?

____

Answers

A. To reduce distortion

B. Negative feedback

3 If the feedback from the output is in phase with the input, as shown in Figure 9-4, the circuit's feedback is *positive*.

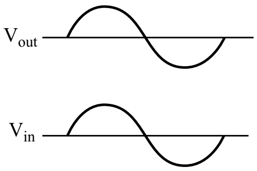

Figure 9-4

In the circuit shown in Figure 9-5, the collector of the second transistor is connected to the base of the first transistor. Because the output signal at the collector of the second transistor is in phase with the input signal at the base of the first transistor, this circuit has positive feedback.

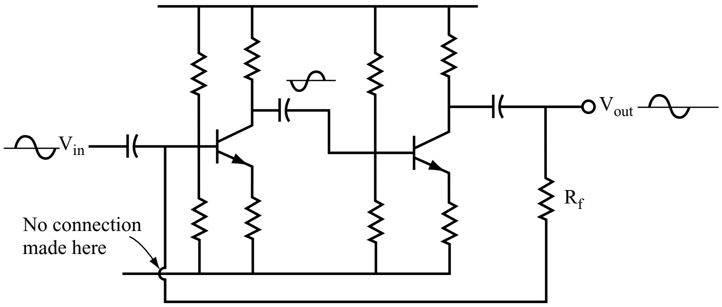

Figure 9-5

Positive feedback can cause an amplifier to oscillate even when there is *no external input*.

Questions

A. What type of feedback is used to stabilize an amplifier?

____

B. What type of feedback is used in oscillators?

____

C. What parts of an amplifier do you connect to produce feedback?

____

Answers

A. Negative feedback

B. Positive feedback

C. Connect the output of an amplifier to its input.

4 The amplifier shown in Figure 9-6 is the same type of amplifier that was discussed in problems 11–18 of Chapter 8. It is called a *common emitter amplifier*.

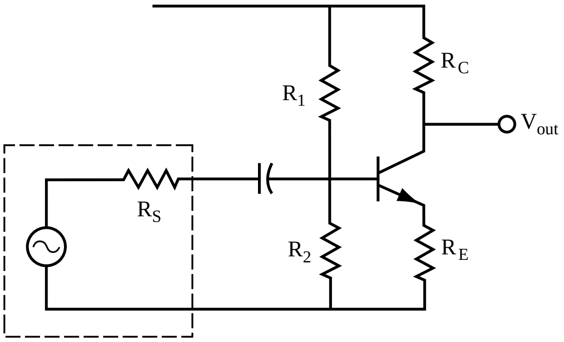

Figure 9-6

Questions

A. What effect would negative feedback have on this amplifier?

____

B. What effect would positive feedback have on this amplifier?

____

Answers

A. Stabilize it, reduce gain, and reduce distortion.

B. Cause it to oscillate.

5 In the circuit shown in Figure 9-6, an input signal that is applied to the base will be amplified.

Questions

A. What is the basic formula for an amplifier's voltage gain?

____

B. What is the voltage gain formula for the amplifier circuit shown in Figure 9-6?

____

Answers

A. $A_V = \beta \times \dfrac{R_L}{R_{in}}$

B. $A_V = \dfrac{R_L}{R_E} = \dfrac{R_C}{R_E}$

(as discussed in problem 12 of Chapter 8)

6 In the circuit shown in Figure 9-7, an input signal is applied to the emitter of the transistor instead of the base. This circuit is called a *common base amplifier*.

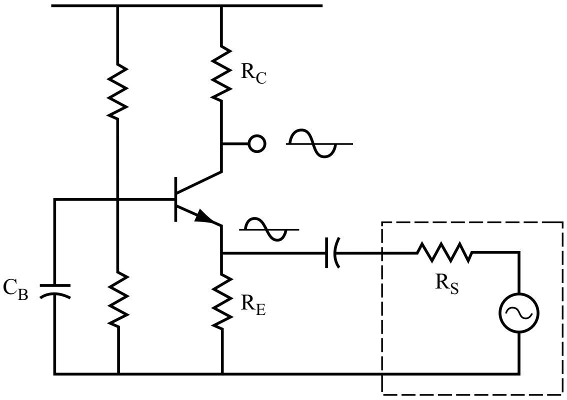

Figure 9-7

NOTE When you apply a signal to the emitter, it changes the voltage drop across the base-emitter diode, just as an input signal applied to the base does. Therefore, a signal applied to the emitter changes the base current and the collector current, just as if you had applied a signal to the base.

The voltage gain formula for this type of amplifier can be simplified because the input impedance to the amplifier is so low when the signal is fed into the emitter that you can discount it. This results in the following voltage gain formula for the common base amplifier:

$$A_V = \frac{R_L}{R_S}$$

R_S is the output resistance or impedance of the source or generator. It is also called the *internal impedance* of the source.

Question

What is the voltage gain formula for the circuit shown in Figure 9-7?

Answer

$A_V = \frac{R_L}{R_S} = \frac{R_C}{R_S}$, ($R_C$ is the load in this circuit)

7 Notice that the input and output sine waves in Figure 9-7 are in phase. Although the signal is amplified, it is not inverted.

Questions

A. What happens to the input signal to the amplifier when you apply it to the emitter instead of the base?

B. Is the input impedance of the common base amplifier high or low compared to the common emitter amplifier?

C. What is the gain formula for the common base amplifier?

Answers

A. Amplified and not inverted

B. Low

C. $A_V = \frac{R_L}{R_S} = \frac{R_C}{R_S}$

8 Figure 9-8 shows an amplifier circuit with a parallel inductor and capacitor connected between the collector of the transistor and ground. A parallel inductor and capacitor circuit is sometimes called a *tuned* (or *resonant*) *load*.

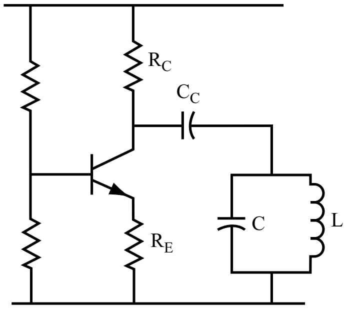

Figure 9-8

In this circuit, the inductor has a very small DC resistance, which could pull the collector DC voltage down to near 0 V. Therefore, you include capacitor C_C in the circuit to allow AC signals to pass through the LC circuit while preventing the collector DC voltage from being pulled down to 0 V.

Questions

A. What term would you use to describe the load in this circuit?

____

B. Does the circuit contain all three components of an oscillator at this point?

____

Answers

A. Resonant or tuned

B. No, the feedback connections are missing.

NOTE **The circuit in Figure 9-8 does not have an input signal either to the emitter or to the base. By adding a feedback connection to a parallel LC circuit, you provide an input signal to the emitter or base, as explained later in this chapter.**

9 Write the voltage gain formulas for the following circuits. Refer to the circuits and voltage gain formulas in problems 4–6, if necessary.

Questions

A. Common emitter circuit____________________________

B. Common base circuit____________________________

Answers

A. $A_V = \dfrac{R_C}{R_E}$

B. $A_V = \dfrac{R_C}{R_S}$

10 You can use common emitter and common base amplifier circuits in oscillators, and in each case, you would usually also include an extra capacitor.

In a common emitter amplifier, you can add a capacitor (C_E) between the emitter and ground, as discussed in problems 19 and 20 of Chapter 8.

In a common base circuit, you can add a capacitor (C_B) between the base and the ground, as is shown in Figure 9-7.

Question

What is the general effect in both cases?

____

Answer

An increase in the gain of the amplifier.

The gain is increased to the point where you can consider it ''large enough'' to use the amplifier as an oscillator. When these capacitors are used in either a common emitter or common base amplifier, it is not usually necessary to calculate the gain of the amplifier.

11 An LC circuit has a resonance frequency that you can determine using the methods discussed in problems 6–12 of Chapter 7. When you use an LC circuit in an oscillator, the output signal of the oscillator will be at the resonance frequency of the LC circuit.

Question

What is the formula for the oscillation (or resonant) frequency?

____

Answer

$$f_r = \frac{1}{2\pi\sqrt{LC}}$$

In practice, the actual measured frequency is never quite the same as the calculated frequency. The capacitor and inductor values are not exact, and other stray capacitances in the circuit affect the frequency. When you need to set an exact frequency, use an adjustable capacitor or inductor.

12 Figure 9-9 shows the parallel LC circuit connected between the collector and the supply voltage, rather than between the collector and ground (as in Figure 9-8).

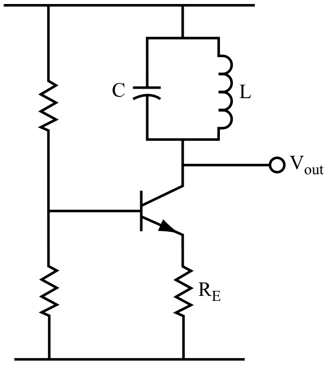

Figure 9-9

You can use this circuit and the circuit shown in Figure 9-8 to selectively amplify one frequency far more than others.

Questions

A. What would you expect this one frequency to be?

____

B. Write the formula for the impedance of the circuit at the resonance frequency.

____

C. What is the AC voltage gain at this frequency?

____

Answers

A. The resonance frequency

B. $Z = \frac{L}{C \times r}$

where r is the DC resistance of the coil

C. $A_V = \frac{Z}{R_E}$

13 Because of the very low DC resistance of the coil, the DC voltage at the collector is usually very close to the supply voltage (V_S). In addition, the AC output voltage positive peaks can exceed the DC level of the supply voltage. With large AC output, the positive peaks can actually reach $2V_S$, as shown in Figure 9-10.

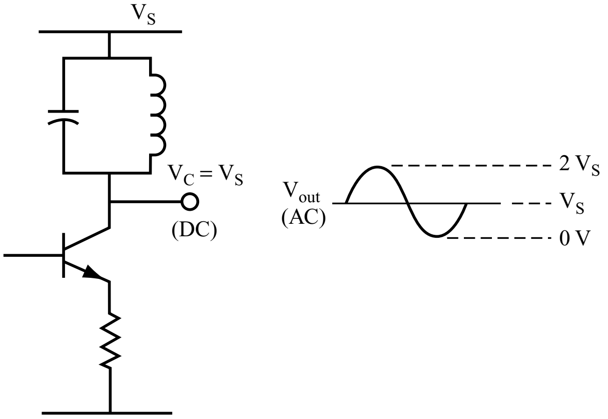

Figure 9-10

Question

Indicate which of the following is an accurate description of the circuit in Figure 9-10:

A. Oscillator

B. Tuned amplifier

C. Common base circuit

D. Common emitter circuit

Answer

B

Feedback

14 To convert an amplifier into an oscillator, you must connect a portion of the output signal to the input. This feedback signal must be in phase with the input signal to induce oscillations.

Figure 9-11 shows three methods you can use to provide a feedback signal from a parallel LC circuit. Each is named for its inventor.

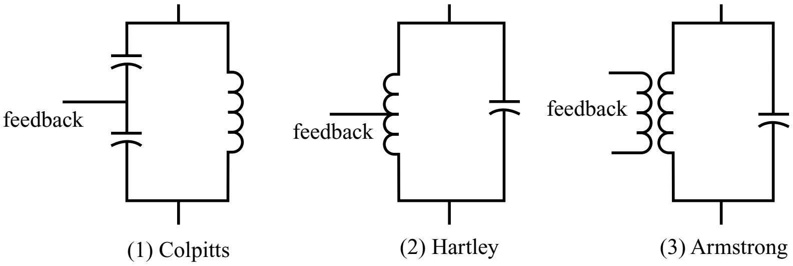

Figure 9-11

In the *Colpitts* method, the feedback signal is taken from a connection between two capacitors that form a voltage divider. In the *Hartley* method, the feedback signal is taken from a tap partway down the coil. Therefore, an inductive voltage divider determines the feedback voltage. The *Armstrong* method uses a step down transformer (an inductor with an extra coil with fewer turns than the main coil). In all three of these methods, between one-tenth and a half of the output must be used as feedback.

Questions

A. Where is the feedback taken from in a Colpitts oscillator?

B. What type of oscillator uses a tap on the coil for the feedback voltage?

C. What type does not use a voltage divider?

____

Answers

A. A capacitive voltage divider

B. Hartley

C. Armstrong

15 The output voltage appears at one end of the parallel LC circuit shown in Figure 9-12, and the other end is effectively at ground. The feedback voltage V_f is taken between the junction of the two capacitors.

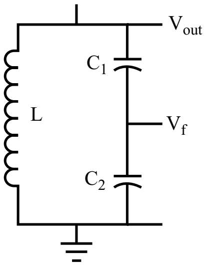

Figure 9-12

Question

Using the voltage divider formula, what is V_f?

____

Answer

$$V_f = \frac{V_{out}X_{C2}}{(X_{C1} + X_{C2})}$$

which becomes

$$V_f = \frac{V_{out}C_1}{(C_1 + C_2)}$$

16 To find the resonance frequency in this circuit, first find the equivalent total capacitance C_T of the two series capacitors. You then use C_T in the resonance frequency formula.

Questions

A. What is the formula for C_T?

____

B. What is the resonance frequency formula for the Colpitts oscillator?

____

Answers

A. $C_T = \frac{C_1 C_2}{C_1 + C_2}$

B. $f_r = \frac{1}{2\pi\sqrt{LC_T}}$,
if Q is equal to or greater than 10

NOTE **If Q is less than 10, you can use one of the following two formulas to calculate the resonance frequency for a parallel LC circuit:**

$$f_r = \frac{1}{2\pi\sqrt{LC}}\sqrt{1 - \frac{r^2C}{L}} \text{ or } f_r = \frac{1}{2\pi\sqrt{LC}}\sqrt{\frac{Q^2}{1+Q^2}}$$

17 Figure 9-13 shows a parallel LC circuit in which the feedback voltage is taken from a tap N_1 turns from one end of a coil, and N_2 turns from the other end.

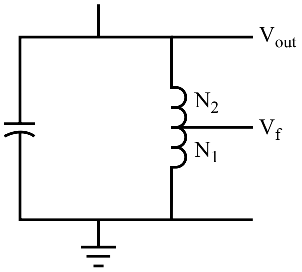

Figure 9-13

You can calculate the feedback voltage with a voltage divider formula that uses the number of turns in each part of the coil.

$$V_f = V_{out} \times \frac{N_1}{N_1 + N_2}$$

The manufacturer should specify N_1 and N_2.

Questions

A. Who invented this feedback method?

____

B. When you divide V_f by V_{out}, what should the result be?

____

Answers

A. Hartley

B. Between one-tenth and one-half

18 Figure 9-14 shows a parallel LC circuit in which the feedback voltage is taken from the secondary coil of a transformer. The formula used to calculate the output voltage of a secondary coil is covered in problem 6 of Chapter 10.

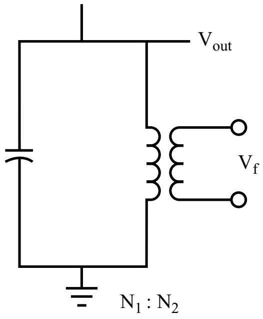

Figure 9-14

Question

Who invented this type of oscillator?

____

Answer

Armstrong

19 For each of the feedback methods described in the last few problems, the voltage fed back from the output to the input is a fraction of the total output voltage ranging between one-tenth and one-half of V_{out}.

To ensure oscillations, the product of the feedback voltage and the amplifier voltage gain must be greater than 1.

$A_V \times V_f > 1$

It is usually easy to achieve this because A_v is much greater than 1.

No external input is applied to the oscillator. Its input is the small part of the output signal that is fed back. If this feedback is of the correct phase and amplitude, the oscillations start spontaneously and continue as long as power is supplied to the circuit.

The transistor amplifier amplifies the feedback signal to sustain the oscillations, and converts the DC power from the battery or power supply into the AC power of the oscillations.

Questions

A. What makes an amplifier into an oscillator?

____

B. What input does an amplifier need to become an oscillator?

____

Answers

A. A resonant LC circuit with feedback of the correct phase and amount

B. None. Oscillations will happen spontaneously if the feedback is correct.

The Colpitts Oscillator

20 Figure 9-15 shows a Colpitts oscillator circuit, the simplest of the LC oscillators to build.

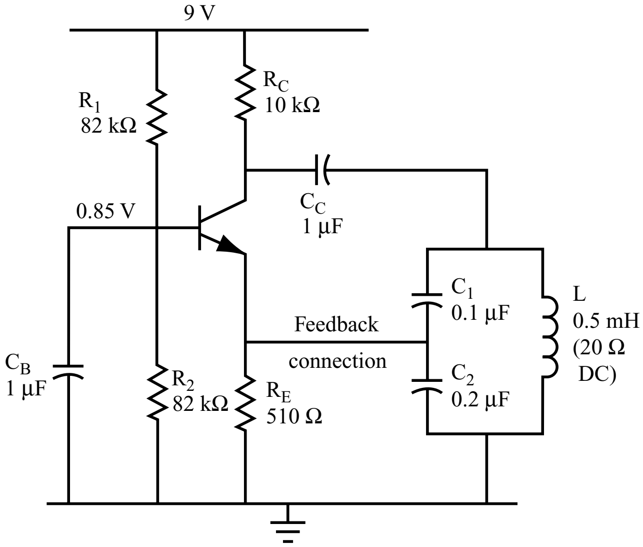

Figure 9-15

The feedback signal is taken from the capacitive voltage divider and fed to the emitter. This connection provides a feedback signal to the emitter in the phase required to provide positive feedback.

In this circuit, the reactance of capacitor C_B is low enough for the AC signal to pass through it, rather than passing through R_2. Capacitor C_E should have a reactance, X_{CB}, of less than 160 ohms at the oscillation frequency. If R_2 happens to be smaller than 1.6 kΩ, choose a value of X_{CB} that is less than one-tenth of R2.

Question

For the circuit shown in Figure 9-15, what is your first estimate for C_B? Assume that f_r is equal to 1 kHz and that X_c equals 160 ohms.

Answer

$$X_{CB} = 160\,\text{ohms} = \frac{1}{2\pi f C_B} = \frac{1}{2 \times \pi \times 10^3 \times C_B}$$

Therefore, $C_B = 1\ \mu F$; larger values of C_B also work.

21 Use the Colpitts oscillator component values shown in Figure 9-15 to answer the following questions.

Questions

A. What is the effective total capacitance of the two series capacitors in the tuned circuit?

C_T = ______________________________

B. What is the oscillator frequency?

f_r = ______________________________

C. What is the impedance of the tuned circuit at this frequency?

Z = ______________________________

D. What fraction of the output voltage is fed back?

V_f = ______________________________

E. What is the reactance of C_B at the frequency of oscillation?

X_{CB} = ______________________________

Answers

A. $C_T = 0.067\ \mu F$

B. Because Q is not known, use the formula that includes the resistance of the coil (see problem 16):

$f_r = 26.75$ kHz

If you use the calculated value of f_r to calculate Q, as in problem 20 of Chapter 8, you find that $Q = 4.2$. Therefore, it is appropriate to use the formula that includes the resistance of the coil to calculate f_r.

C. Using

$$Z_T = \frac{L}{rC},\ Z_T = 373 \text{ ohms}.$$

D. Using a voltage divider with the capacitor values,

$$V_f = V_{out}\frac{C_1}{(C_1 + C_2)} = \frac{V_{out}}{3}.$$

E. X_{CB} = about 6 ohms, which is a very good value (much less than the 8200 ohm value of R_2)

22 Figure 9-16 shows a Colpitts oscillator circuit that uses a different method for making feedback connections between the parallel LC circuit and the transistor.

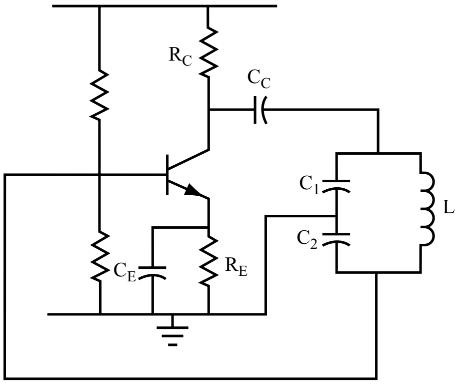

Figure 9-16

Question

List the differences between this circuit and the one shown in Figure 9-15.

Answer

The feedback is connected to the base instead of the emitter, and the ground is connected to the center of the capacitive voltage divider. The capacitor C_E has been added. (This connection provides a feedback signal to the base in the correct phase to provide positive feedback.)

23 In the circuit shown in Figure 9-16, capacitor C_E should have a reactance of less than 160 ohms at the oscillation frequency. If the emitter resistor R_E is smaller in value than 1.6 kΩ, then C_E should have a reactance that is less than $R_E/10$ at the oscillation frequency.

Question

If you use an emitter resistor of 510 ohms in a 1 kHz oscillator, what value of capacitor should you use for C_E?

Answer

$$X_C = \frac{510}{10} = \frac{1}{2\pi f C_E} = \frac{0.16}{10^3 \times C_E}$$

So, $C_E = 3.2\ \mu F$. Thus, you should use a capacitor larger than 3 μF.

24 Figure 9-17 shows a Colpitts oscillator with the parallel LC circuit connected between the collector and the supply voltage. As with the circuits shown in Figures 9-15 and 9-16, this circuit provides a feedback signal to the transistor (in this case, the emitter) in the correct phase to provide positive feedback.

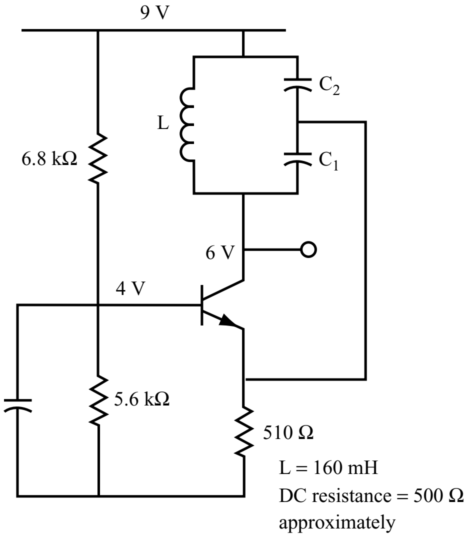

L = 160 mH
DC resistance = 500 Ω approximately

Figure 9-17

The following table shows possible values you might use for C_1 and C_2 in the circuit shown in Figure 9-17.

C_1	C_2	C_T	F_R
0.01 µF	0.1 µF		
0.01 µF	0.2 µF		
0.01 µF	0.3 µF		
0.1 µF	1 µF		
0.2 µF	1 µF		

Questions

A. Calculate C_T and f_r for each row of the preceding table.

B. Does increasing C_2 while holding C_1 constant increase or decrease the resonance frequency?

____

C. What effect does increasing C_1 have on the resonance frequency?

____

D. What is the condition that results in the highest possible resonance frequency?

____

E. What would be the highest resonance frequency if C_1 is fixed at 0.01 μF and C_2 can vary from 0.005 μF to 0.5 μF?

____

Answers

A. The values of C_T and f_r are shown in the following table.

C_1	C_2	C_T	F_R
0.01 μF	0.1 μF	0.009 μF	4.19 kHz
0.01 μF	0.2 μF	0.0095 μF	4.08 kHz
0.01 μF	0.3 μF	0.0097 μF	4.04 kHz
0.1 μF	1 μF	0.09 μF	1.33 kHz
0.2 μF	1 μF	0.167 μF	0.97 kHz

B. Increasing C_2 decreases the resonance frequency, and, therefore, decreases the output frequency of the oscillator.

C. Increasing C_1 will also decrease the resonance frequency and the output frequency of the oscillator.

D. When C_T is at its lowest possible value.

E. When C_2 is 0.005 μF, C_T will be 0.0033 μF, which is its lowest possible value. Therefore, the frequency is at the highest possible value, or about 6.9 kHz. The lowest frequency occurs when C_2 is at its highest setting of 0.5 μF.

Optional Experiment

If you have an oscilloscope, you may want to build the oscillator shown in Figure 9-17 and see how closely your measurements of the frequency agree with those you have just calculated. If you are within 20 percent, then the answers are satisfactory. In some cases, the waveform may be distorted.

The Hartley Oscillator

25 Figure 9-18 shows a Hartley oscillator circuit. In this type of circuit, the feedback is taken from a tap on the coil.

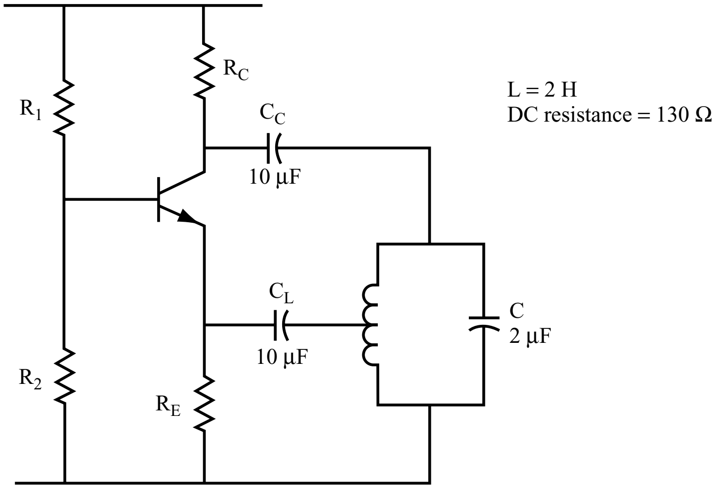

Figure 9-18

Capacitor C_L stops the emitter DC voltage from being pulled down to 0 V through the coil. C_L should have a reactance of less than $R_E/10$, or less than 160 ohms at the oscillator frequency.

Questions

Work through the following calculations:

A. What is the resonance frequency?

f_r = ____________________

B. What is the approximate impedance of the load?

Z = ____________________

C. What missing information prevents you from calculating the fraction of the voltage drop across the coil that is fed back to the emitter?

____

Answers

A. 80 Hz (approximately)

B. 7.7 kΩ (approximately)

C. The number of turns in the coil and the position of the tap are not known.

Figure 9-19 shows a Hartley oscillator with the parallel LC circuit connected between the collector and the supply voltage. As with the circuit shown in Figure 9-18, this circuit provides a feedback signal to the emitter from a tap in the coil, in the correct phase to provide positive feedback.

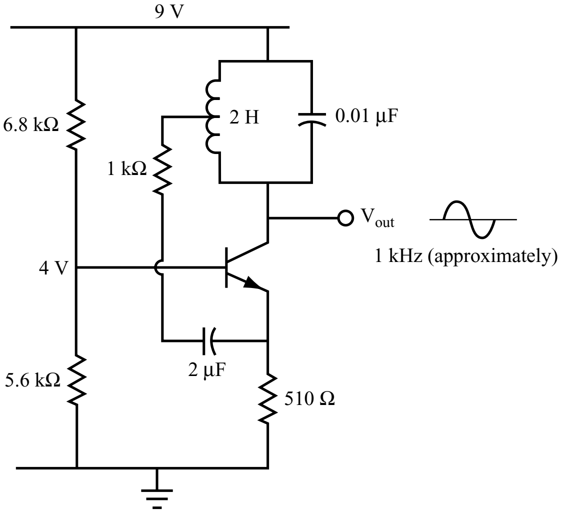

Figure 9-19

For a Colpitts oscillator it is easy to choose the two capacitors to provide both the desired frequency and the desired amount of feedback. If the circuit does not oscillate, it is relatively simple to change the capacitors until it does oscillate, and then adjust the values slightly to get the desired frequency.

Because a Hartley oscillator uses a tapped coil, trying out different feedback voltages is not quite so easy. The feedback ratio cannot be altered, because it is impossible to make another tap or change the tap on the coil.

The Armstrong Oscillator

The Armstrong oscillator shown in Figure 9-20 is somewhat more difficult to design and build.

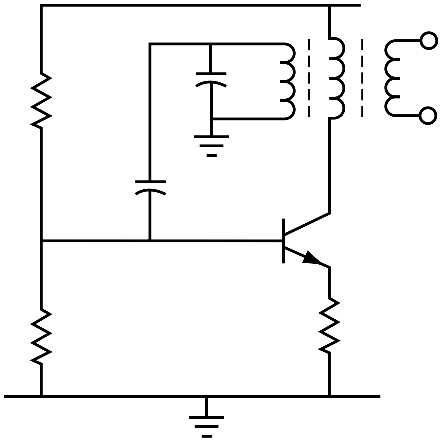

Figure 9-20

Here, the oscillations depend more on the extra winding on the coil than on any other factor.

Because of the large variety of transformers and coils available, it is almost impossible to give you a simple procedure for designing an Armstrong oscillator. Instead, the manufacturer specifies the number of turns required on the coils, which guarantees that the oscillator will work in its most common operation, at high radio frequencies.

Because of the practical difficulties, the Armstrong and its variations are not explored any further.

Practical Oscillator Design

26 This section briefly covers some practical problems with oscillators, and then presents a simple design procedure.

Before you proceed, review the important points of this chapter by answering the following questions.

Questions

A. What three elements must an oscillator have present to work?

____

B. What determines the frequency of an oscillator's output signal?

____

C. What provides the feedback?

____

D. How many feedback methods for oscillators have we presented?

____

E. What do you need to start the oscillations once the circuit has been built?

____

Answers

A. An amplifier, a resonant LC circuit (or some other frequency determining components), and feedback.

B. The frequency of the output signal is the same as the resonance frequency.

C. A voltage divider on the resonant circuit.

D. Three — the Colpitts, Hartley, and the Armstrong.

E. Nothing — the oscillations should start spontaneously if the component values in the circuit are correct.

The main practical problem in building oscillators is selecting the coil. For mass production, a manufacturer can specify and purchase the exact coil required. But in a lab or workshop, where you are building only a single circuit, it is often difficult or impossible to find the exact inductor specified in a circuit design. What usually happens is that you use the most readily available coil, and design the rest of the circuit around it. This presents three possible problems:

- You may not know the exact value of the inductance.
- The inductance value may not be the best for the desired frequency range.
- The coil may or may not have tap points or extra windings, and this may cause a change in the circuit design. For example, if there are no taps, then you cannot build a Hartley oscillator.

Because Colpitts is the easiest oscillator to make work in practice and provides an easy way around some of the practical difficulties, let's focus on that oscillator.

You can use almost any coil when building a Colpitts oscillator, provided it is suitable for the frequency range you want. For example, a coil from the tuner section of a television set would not be suitable for a 1 kHz audio oscillator because its inductance value is outside the range best suited to a low frequency audio circuit.

Simple Oscillator Design Procedure

27 Following is a simple step-by-step procedure for the design of a Colpitts oscillator. The Colpitts will work over a wide frequency range. (A Hartley can be built from a very similar set of steps.)

By following this procedure, you can build an oscillator that works in the majority of cases. There is a procedure you can use that guarantees that the oscillator will work, but it is far more complex.

If you are not actually building the oscillator, use the assumed values provided here, and draw the circuits using your calculated values.

Follow these steps:

1. Choose the frequency of the oscillator output signal.
2. Choose a suitable coil. This step presents the greatest practical difficulty. Some values of coil are often not available, so you must use whatever is readily available. Fortunately, you can use a wide range of inductance values and still obtain the desired resonance frequency by adjusting the value of the capacitor.
3. If you know the value of the inductance, calculate the capacitor value using this formula:

   $$f_r = \frac{1}{2\pi\sqrt{LC}}$$

   Use this value of capacitor for C_1 in the next steps.
4. If you don't know the inductance value, choose any value of capacitance and call this C_1. This may produce a frequency considerably different from what you require. However, at this stage, the main thing is to get the circuit oscillating. You can adjust values later.
5. Choose a capacitor C_2 that is between 3 and 10 times the value of C_1. Figure 9-21 shows the two capacitors and the coil connected in a parallel circuit with the two capacitors acting as a voltage divider.

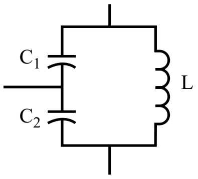

Figure 9-21

At this point, stop and make some assumptions. Suppose you need a frequency of 10 kHz and have a coil with a 16 mH inductance.

Questions

A. What approximate value of C_1 do you need?

____

B. What value of C_2 do you need?

____

Answers

A. $C_1 = 0.016\ \mu F$

B. $C_2 = 0.048\ \mu F$ to $0.16\ \mu F$

28 Now, continue with the design procedure by following the next steps.

6. Design an amplifier with a common emitter gain of about 20. Choose a collector DC voltage that is about half the supply voltage. The main point to keep in mind here is that the collector resistor R_C should be about one-tenth the value of the impedance of the LC circuit at the resonance frequency. This is often a difficult choice to make, especially if you don't know the coil value. Usually, you have to make an assumption, so R_C is an arbitrary choice.
7. Assemble the circuit shown in Figure 9-15.
8. Calculate the value of C_C. Do this by making X_C 160 ohms at the desired frequency. This is another ''rule of thumb'' that happens to work, and you can justify it mathematically. Use the following formula:

$$C_C = \frac{1}{2\pi f_r X_C}$$

Question

Substitute the values given so far into the formula to calculate C_C.

____

Answer

$$C_C = \frac{1}{2\pi \times 10\ \text{kHz} \times 160\ \Omega} = 0.1\ \mu F$$

29 Now, complete one last step.

9. Calculate the value of C_B. Again, choose a value so that X_C is 160 ohms at the desired frequency.

Question

What is the value of C_B?

Answer

$C_B = 0.1\,\mu F$

30 Continue the design procedure steps.

10. Apply power to the circuit and look at the output signal on an oscilloscope. If the output signal is oscillating, check the frequency. If the frequency varies significantly from the desired frequency, then change C_1 until you get the desired frequency. Change C_2 to keep the ratio of the capacitance values about the same as discussed in step 5. C_2 will affect the output level.
11. If the circuit does not oscillate, go through the steps outlined in the troubleshooting checklist that follows.

Optional Experiment

Once the oscillator is working, if you want to take it further, you can change the circuit to connect the feedback signal to the base instead of the emitter, as shown in Figure 9-16. To build this oscillator, you also need to calculate C_E. Assume that X_C is 160 ohms at the desired frequency.

Oscillator Troubleshooting Checklist

If an oscillator does not work, most often the trouble is with the feedback connections. A little experimenting (as outlined in steps 2 to 6 of the following checklist) should produce the right results. This is especially true when you use an unknown coil that may have several taps or windings. However, you should try each of the following steps if you're having trouble.

1. Ensure that C_B, C_C, and C_E are all large enough to have a reactance value less than 160 ohms. Ensure that C_E is less than one-tenth of R_E.

2. Check the C_1/C_2 ratio. It should be between 3:1 and 10:1.
3. Swap out C_1 and C_2. They may be connected to the wrong end of the LC circuit.
4. Check that you made the feedback connection to and from the correct place.
5. Check both ends of the LC circuit to see that they are connected to the correct place.
6. Check the DC voltage level of the collector, base, and emitter.
7. Check the capacitor values of the LC circuit. If necessary, try some other values until the circuit oscillates.
8. If none of the previous actions produce oscillations, check to see if any of the components are defective. The coil may be opened or shorted. The capacitor may be shorted. The transistor may be dead, or its β may be too low. Check the circuit wiring carefully.

In most cases, one or more of these steps produces oscillations.

When an oscillator is working it may still have one or two main faults, including the following:

- *A distorted output waveform* — This can happen when C_B, C_C, or C_E are not low enough in value, or when an output amplitude is too high.
- *Output level too low* — When this happens, the sine wave is usually very ''clean'' and ''pure.'' In a Colpitts oscillator, changing the ratio of C_1 and C_2 often helps raise the output level. If not, you can use another transistor as an amplifier after the oscillator as discussed in Chapter 8, problem 21.

31 Now, work through a design example. Design and build an oscillator with an output frequency of 25 kHz using a coil with a value of 4 mH, and address each of the steps in problems 27–30 as described in these questions.

Questions

1. The value of f_r is given as 25 kHz.
2. L is given as 4 mH.
3. Use the formula to find C_1.

   C_1 = ____________________
4. You do not need this step.
5. Choose C_2.

   C_2 = ____________________
6. The procedure to design amplifiers is shown in Chapter 8.
7. The circuit is shown in Figure 9-22.

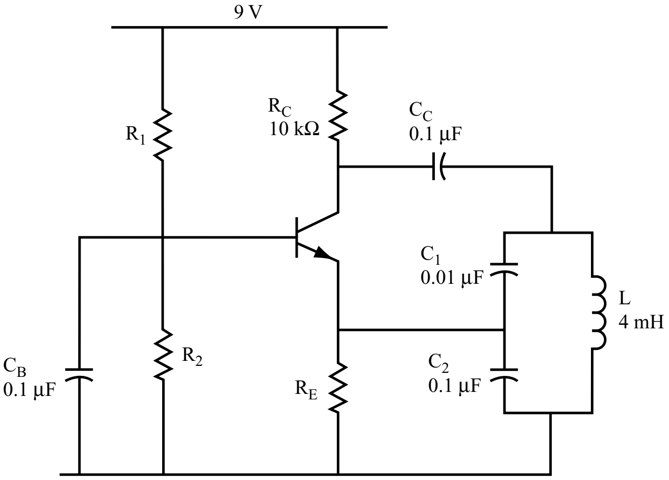

Figure 9-22

8. Find C_C.

C_C = ____________________

9. Find C_B.

C_B = ____________________

Answers

$C_1 = 0.01\ \mu F$
$C_2 = 0.1\ \mu F$
$C_C = 0.047\ \mu F$ (use $0.1\ \mu F$)
$C_B = 0.047\ \mu F$ (use $0.1\ \mu F$)

Steps 10–11 are the procedure you use to ensure that the oscillator works. If you build this circuit, go through steps 10–11. You don't need to do them if you didn't actually build the circuit.

32 Figure 9-22 shows the circuit designed in problem 31.

Measurements of the output signal of this oscillator confirm a frequency very close to 25 kHz.

Question

Find the impedance of the LC circuit at resonance. Note that r (the DC resistance of the inductor) is 12 ohms.

____

Answer

$$Z = \frac{L}{C \times r} = \frac{4 \times 10^{-3}}{0.01 \times 10^{-6} \times 12} = 33\,\text{k}\Omega \text{ (approximately)}$$

Note that this is about three times the value used for R_C instead of being 10 times the value of R_C, as suggested in step 6 of problem 28.

33 If you wish, work through this second oscillator design example. Design an oscillator with an output frequency of 250 kHz using a coil with a value of 500 μH.

Questions

1. $f_r = 250$ kHz
2. $L = 500\ \mu H = 0.5$ mH
3. Find C_1.
   $C_1 =$ ____________________
4. You do not need this step.
5. Find C_2.
   $C_2 =$ ____________________
6. Use the same amplifier as in the last example.
7. The circuit is shown in Figure 9-23.
8. Find C_C.
   $C_C =$ ____________________
9. Find C_B.
   $C_B =$ ____________________

Answers

$C_1 = 0.0008\ \mu F$; therefore, choose a standard value of 0.001 μF
$C_2 = 0.0047\ \mu F$, which is a standard value
$C_B = C_C = 0.004\ \mu F$ (minimum)

34 The circuit you designed in problem 33 is shown in Figure 9-23.

Measurements of the output signal of this oscillator confirm a frequency close to 250 kHz.

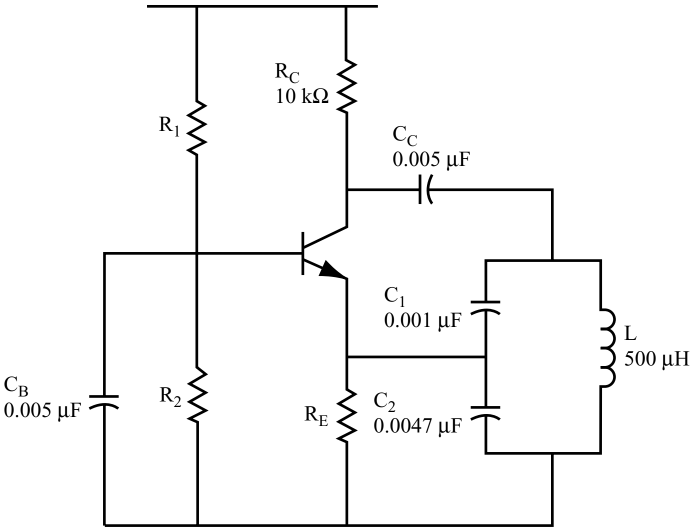

Figure 9-23

Question

Find the impedance of the LC circuit at resonance. Note that r (the DC resistance of the inductor) is 20 ohms.

Answer

Z = 30 kΩ

This is about 3 times the value of R$_C$ rather than 10 times the value of R$_C$, as suggested in step 6 of problem 28.

35 Figure 9-24 shows several other oscillator circuits. Calculate the expected output frequency for each circuit and build as many as you want to. Check the measured oscillator output frequency against the calculated values for each circuit you build.

Questions

What is the output frequency for each circuit?

A. f =

B. f =

C. f =

D. f =

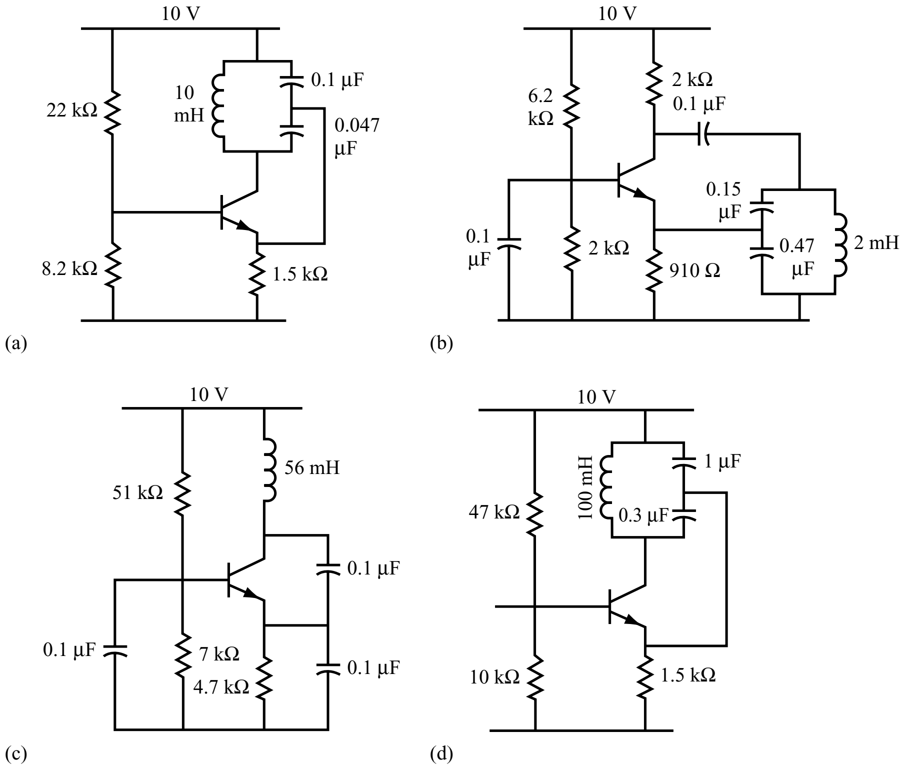
10 V
22 kΩ
10 mH
0.1 μF
0.047 μF
8.2 kΩ
1.5 kΩ
(a)
10 V
6.2 kΩ
2 kΩ
0.1 μF
0.15 μF
0.1 μF
2 kΩ
910 Ω
0.47 μF
2 mH
(b)
10 V
56 mH
51 kΩ
0.1 μF
0.1 μF
7 kΩ
4.7 kΩ
0.1 μF
(c)
10 V
100 mH
1 μF
0.3 μF
47 kΩ
10 kΩ
1.5 kΩ
(d)

Figure 9-24

Answers

A. 8.8 kHz

B. 10 kHz

C. 3 kHz

D. 1 kHz

Summary and Applications

In this chapter, we covered the following topics related to oscillators:

- The main elements that make up an oscillator.
- How to differentiate between positive and negative feedback.
- The type of feedback that causes a circuit to oscillate.
- Two methods of obtaining feedback in an oscillator circuit.
- How resonant LC circuits set the frequency of an oscillator.

You also practiced designing a simple oscillator circuit to solidify your understanding of its elements and operation.

Self-Test

These questions test your understanding of the concepts and equations presented in this chapter. Use a separate sheet of paper for your diagrams or calculations. Compare your answers with the answers provided following the test.

1. What are the three sections that are necessary in an oscillator?

   ____

2. What is the difference between positive and negative feedback?

   ____

3. What type of feedback is required in an oscillator?

   ____

4. What is the formula for the frequency of an oscillator?

   ____

5. Draw the circuit for a Colpitts oscillator.
6. Draw the circuit for a Hartley oscillator.
7. Draw the circuit for an Armstrong oscillator.
8. Problems 27–30 give a design procedure for oscillators. How well do the circuits in problem 35 fulfill the criteria for that procedure? In other words, check the values of V_f, A_V (for a common emitter amplifier), C_1/C_2 ratio, R_C/Z ratio, and the frequency.

   A. ______________________________

   B. ______________________________

   C. ______________________________

   D. ______________________________

9. For the circuit shown in Figure 9-23, calculate the values of C_1, C_2, C_C, and C_B for an oscillator with an output frequency of 10 kHz using a 100 mH coil.

Answers to Self-Test

If your answers do not agree with those provided here, review the problems indicated in parentheses before you go on to the next chapter.

1.		An amplifier, feedback, and a resonant load	(problem 1)
2.		Positive feedback is "in phase" with the input, and negative feedback is "out of phase" with the input.	(problems 2–3)
3.		Positive feedback	(problem 3)
4.		$f_r = \frac{1}{2\pi\sqrt{LC}}$	(problem 11)
5.		See Figure 9-15.	(problem 20)
6.		See Figure 9-18.	(problem 25)
7.		See Figure 9-20.	(problem 25)
8.	A.	$V_f = \frac{0.047}{0.147}$	(problems 27–30)
		A_V cannot be calculated.	
		$C_1/C_2 = 0.047/0.1 = 0.47$	
		Z cannot be calculated because r is unknown.	
		$f_r = 8.8$ kHz (approximately)	
	B.	$V_f = \frac{0.15}{0.62}$	
		$A_V = 2.2$ (approximately)	
		$C_1/C_2 = 1/3$ (approximately)	
		Z cannot be calculated.	
		$f_r = 10$ kHz (approximately)	
	C.	$V_f = \frac{0.1}{0.2}$	
		A_v cannot be calculated.	
		$C_1/C_2 = 1$	
		Z cannot be calculated.	
		$f_r = 3$ kHz	
	D.	$V_f = \frac{0.3}{1}$	
		A_v cannot be calculated.	
		$C_1/C_2 = 0.3$	
		Z cannot be calculated.	
		$f_r = 1$ kHz (approximately)	
9.		$C_1 = 0.0033\ \mu F$; $C_2 = 0.01\ \mu F$; $C_B = C_C = 0.1\ \mu F$	(problems 26–30)

CHAPTER 10

The Transformer

Transformers are used to ''transform'' an AC voltage to a higher or lower level. When you charge up your cellphone, you are using a transformer to reduce the 120 volts supplied by the wall outlet to the 5 volts or so needed to charge your cellphone's battery. In fact, most electrical devices that you plug into wall outlets use transformers to reduce power coming from an outlet to that required by the electrical components in the device.

You can also use transformers to increase voltage. For example, some of the equipment used to manufacture integrated circuits requires thousands of volts to operate. Transformers are used to increase the 240 volts supplied by the power company to the required voltage.

When you complete this chapter, you will be able to do the following:

- Recognize a transformer in a circuit.
- Explain and correctly apply the concepts of *turns ratio* and *impedance matching*.
- Recognize two types of transformer.
- Do simple calculations involving transformers.

Transformer Basics

1 Consider two coils placed very close to each other, as shown in Figure 10-1. If you apply an AC voltage to the first (or primary) coil, the alternating current flowing through the coil creates a fluctuating magnetic field that surrounds the coil. As the strength and polarity of this magnetic field changes, it induces an alternating current and a corresponding AC voltage in the second (or secondary) coil. The AC signal induced in the secondary coil is at the same frequency as the AC signal applied to the primary coil.

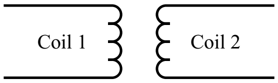

Figure 10-1

Both transformer coils are usually wound around a core made of a magnetic material such as iron or ferrite to increase the strength of the magnetic field.

Questions

A. When the two coils are wound around the same core, are they connected electrically?

____

B. What type of device consists of two wire coils wound around an iron or ferrite core?

____

C. If you apply an AC voltage to the terminals of the primary coil, what occurs in the secondary coil?

____

Answers

A. No.

B. A transformer.

C. An alternating current is induced in the secondary coil, which produces an AC voltage between the terminals of the secondary coil.

2 A transformer is only used with alternating currents. A fluctuating magnetic field (such as that generated by alternating current flowing through a primary coil) is required to induce current in a secondary coil. The stationary magnetic field generated by direct current flowing through a primary coil will not induce any current or voltage in a secondary coil.

When a sine wave signal is applied to a primary coil, you can observe a sine wave of the same frequency across the secondary coil, as shown in Figure 10-2.

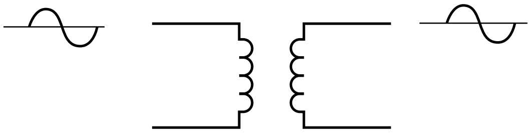

Figure 10-2

Questions

A. What will be the difference in frequency between a signal applied to a primary coil and the signal induced in a secondary coil?

____

B. What will be the voltage difference across a secondary coil if 10 V DC is applied to the primary coil?

____

Answers

A. No difference. The frequencies will be the same.

B. Zero volts. When a DC voltage is applied to the primary coil, there is no voltage or current induced in the secondary coil. You can summarize this by saying that DC does not pass through a transformer.

3 You can compare the output waveform measured between the terminals of the secondary coil to the output waveform measured between the terminals of the primary coil. If the output goes positive when the input goes positive, as shown in Figure 10-3, then they are said to be *in phase*.

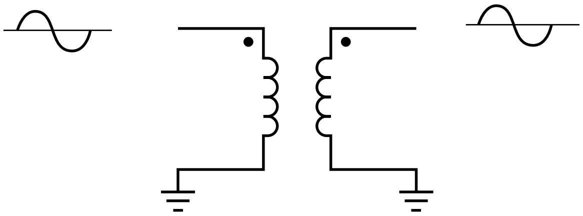

Figure 10-3

The dots on the coils in Figure 10-3 indicate the corresponding end of each coil. If one coil is reversed, then the output will be inverted from the input. The output is said to be *out of phase* with the input, and a dot is placed at the opposite end of the coil.

Question

In Figure 10-4, the output sine wave is out of phase with the input sine wave. Place a dot in the correct location in the secondary coil to show that it is out of phase.

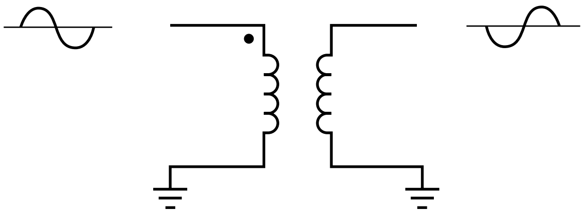

Figure 10-4

Answer

The dot should be at the lower end of the right coil.

4 The transformer shown in the right side of Figure 10-5 has three terminals. The additional terminal, in the middle of the coil, is called a *center tap*.

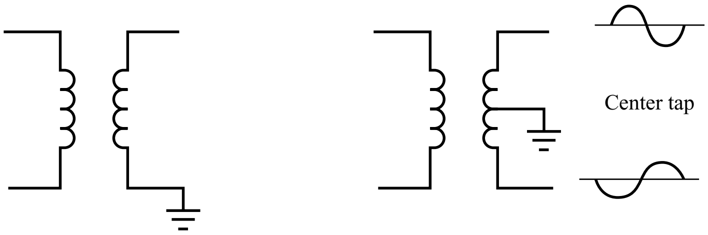

Figure 10-5

Question

What is the difference between the two output waveforms shown for the transformer on the right side of Figure 10-5?

____

Answer

The two waveforms are 180 degrees out of phase. That is, the positive peak of the upper output occurs at the same time as the negative peak of the lower waveform.

5 In a transformer, the output voltage from the secondary coil is directly proportional to the number of turns of wire in the secondary coil. If you increase the number of turns of wire in the secondary coil, a larger output voltage is induced across the secondary coil. If you decrease the number of turns of wire in the secondary coil, a smaller output voltage is induced across the secondary coil.

Question

How does increasing the number of turns of wire in a secondary coil affect the output voltage across the secondary coil?

Answer

It increases the output voltage across the secondary coil.

6 The number of turns in the primary and secondary coils are shown in Figure 10-6 as N_p and N_s.

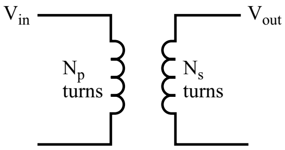

Figure 10-6

Question

The ratio of the input to output voltage is the same as the ratio of the number of turns in the primary coil to the number of turns in the secondary coil. Write a simple formula to express this.

Answer

$$\frac{V_{in}}{V_{out}} = \frac{N_p}{N_s}$$

NOTE **The ratio of primary turns to secondary turns is called the *turns ratio* (TR):**

$$TR = \frac{N_p}{N_s} = \frac{V_{in}}{V_{out}}$$

7 Use the formula from problem 6 to answer the following question.

Question

Calculate the output voltage of a transformer with a 2 to 1 (2:1) turns ratio when you apply a 10 V_{pp} sine wave to the primary coil.

____

Answer

$$\frac{V_{in}}{V_{out}} = \frac{N_p}{N_s} = TR$$

$$V_{out} = V_{in}\frac{N_s}{N_p} = V_{in} \times \frac{1}{TR}$$

$$V_{out} = V_{in} \times \frac{1}{TR} = 10 \times \frac{1}{2} = 5\ V_{pp}$$

8 Use the input voltage and turns ratio for a transformer to answer the following questions.

Questions

Calculate V_{out} in the following:

A. $V_{in} = 20\ V_{pp}$, turns ratio = 5:1.

$V_{out} =$ ____________________________

B. $V_{in} = 1\ V_{pp}$, turns ratio = 1:10.

$V_{out} =$ ____________________________

C. $V_{in} = 100\ V_{rms}$. Find V_{out} when the primary and secondary coil have an equal number of turns.

$V_{out} =$ ____________________________

Answers

A. 4 V_{pp} (This is called a *step down transformer*.)

B. 10 V_{pp} (This is called a *step up transformer*.)

C. 100 V_{rms} (This is called an *isolation transformer*, which is used to separate or isolate the voltage source from the load electrically.)

9 Almost all electronic equipment that is operated from 120 V AC house current requires a transformer to convert the 120 V AC to a more suitable, lower voltage. Figure 10-7 shows a transformer that steps down 120 V AC to 28 V AC.

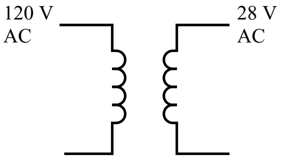

Figure 10-7

Question

Calculate the turns ratio for this transformer.

Answer

$$\text{TR} = \frac{N_p}{N_s} = \frac{120}{28} = 4.3{:}1$$

10 Figure 10-8 shows an oscilloscope trace of the output waveform from the 28 V secondary coil.

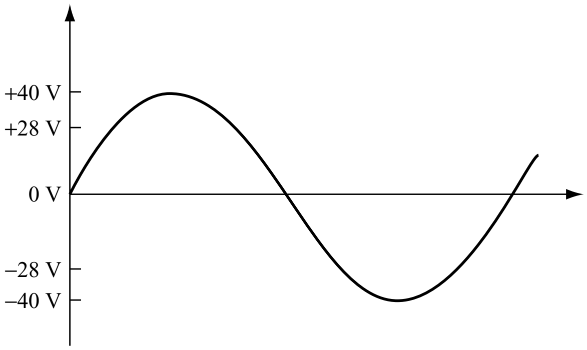

Figure 10-8

Questions

A. Is 28 V a peak-to-peak or an rms value?

B. What is the peak-to-peak value of the 28 V across the secondary coil?

____

Answers

A. rms

B. $2 \times 1.414 \times 28 = 79.184$ V

11 Like the 28 V transformer output value, the 120 V wall plug value is an rms measurement.

Question

What is the peak-to-peak value of the voltage from the wall plug?

____

Answer

Approximately 340 volts

12 The actual voltage measured across the secondary coil of a transformer depends upon where and how you make the measurement. Figure 10-9 illustrates different ways to measure voltage across a 20 V_{pp} secondary coil that has a center tap.

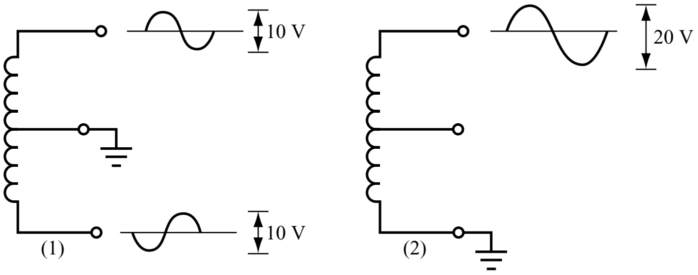

Figure 10-9

If the center tap is grounded as shown in diagram (1) of Figure 10-9, then there is 10 V_{pp} AC between each terminal and ground. You can see that the two output waveforms in diagram (1) are out of phase (180 degrees out of phase, in this case) by comparing the two sine waves shown next to the two terminals. If the bottom terminal is grounded as it is in diagram (2) of Figure 10-9 and the center tap is not used, then there is 20 V_{pp} between the top terminal and ground.

Questions

A. Assume a center-tapped secondary coil is rated at 28 V_{rms} referenced to the center tap. What is the rms voltage output when the center tap is grounded?

____

B. Assume the 28 V_{rms} is the total output voltage across the entire secondary winding. What will be the output voltage between each end of the coil and the center tap?

____

C. Assume the output voltage of a center-tapped secondary coil is 15 V_{rms} between each end of the coil and the center tap. What is the peak-to-peak output voltage when the center tap is not connected?

____

Answers

A. 28 V_{rms} between each end of the coil and the center tap

B. 14 V_{rms} (one half of the total V_{out})

C. When the center tap is not connected, the output is 30 V_{rms}. Therefore, $V_{pp} = 2 \text{ X } 1.414 \text{ X } 30 = 84.84$ volts

13 When the magnetic field induces an AC signal on the secondary coil, there is some loss of power. The percentage of power out of the transformer versus the input power is called the *efficiency* of the transformer. For the sake of this discussion, assume the transformer has an efficiency of 100 percent. Therefore, the output power of the secondary coil equals the power into the primary coil.

Power in = Power out, or $P_{in} = P_{out}$

However, P = VI. Therefore, the following is true:

$$V_{in}I_{in} = V_{out}I_{out}$$

You can rearrange this to come up with the following formula:

$$\frac{I_{out}}{I_{in}} = \frac{V_{in}}{V_{out}} = TR$$

Questions

A. What would be the input current for a transformer if the input power was 12 watts at a voltage of 120 V_{rms}?

____

B. What would be the transformer's output voltage if the turns ratio was 5:1?

____

C. What would be the output current?

____

D. What would be the output power?

____

Answers

NOTE In AC power calculations, you must use the rms values of current and voltage.

A. $I_{in} = \frac{P_{in}}{V_{in}} = \frac{12}{120} = 0.1\,A_{rms}$

B. $V_{out} = \frac{V_{in}}{TR} = \frac{120}{5} = 24\,V_{rms}$

C. $I_{out} = I_{in}\,(TR) = 0.1 \times 5 = 0.5\ A_{rms}$

D. $P_{out} = V_{in}I_{out} = 24 \times 0.5 = 12$ watts (same as the power in)

Transformers in Communications Circuits

14 In communications circuits, an input signal is often received via a very long interconnecting wire, usually called a *line*, which normally has an impedance of 600 ohms. A typical example is a telephone line between two cities.

Question

Communications equipment works best when connected to a load that has the same impedance as the output of the equipment. What output impedance should communications equipment have?

Answer

600 ohms output impedance, to be connected to a 600 ohms line

15 Because most electronic equipment does not have a 600 ohm output impedance, a transformer is often used to connect such equipment to a line. Often, the transformer will be built into the equipment for convenience. The transformer is used to ''match'' the equipment to the line, as shown in Figure 10-10.

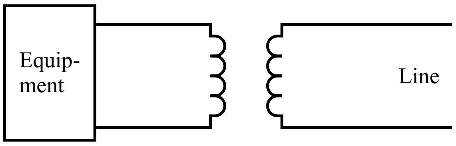

Figure 10-10

To work correctly, the output of the transformer secondary coil should have a 600 ohm impedance to match the line. The output impedance of the transformer (measured at the secondary winding) is governed by two things. One of these is the output impedance of the equipment.

Question

What would you expect the other governing factor to be?

Answer

The turns ratio of the transformer. (The DC resistance of each coil has no effect, and you can ignore it.)

16 Figure 10-11 shows a signal generator with an output impedance of Z_G connected to the primary coil of a transformer. A load impedance of Z_L is connected to the secondary coil.

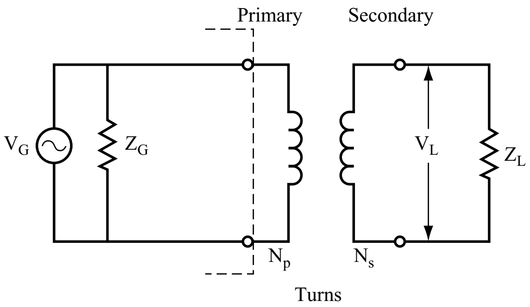

Figure 10-11

You know that $P_{in} = P_{out}$ and that $P = V^2/Z$. Therefore, you can write an equation equating the power of the generator to the power of the load in terms of V and Z, as shown here:

$$\frac{{V_G}^2}{Z_G} = \frac{{V_L}^2}{Z_L}$$

You can rearrange this equation to give the ratio of the voltages, as shown here:

$$\frac{Z_G}{Z_L} = \left(\frac{V_G}{V_L}\right)^2$$

And since $V_G = V_{in}$ and $V_L = V_{out}$ and $V_{in}/V_{out} = N_p/N_s$, the following is true:

$$\frac{Z_G}{Z_L} = \left(\frac{V_{in}}{V_{out}}\right)^2 = \left(\frac{N_p}{N_s}\right)^2 = (TR)^2$$

Therefore, the ratio of the input impedance to the output impedance of a transformer is equal to the square of the turns ratio. As you will see in question A, you can determine the turns ratio for a transformer that will match impedances between a generator and a load. In this way, the generator ‘‘sees’’ an impedance equal to its own impedance, and the load also ‘‘sees’’ an impedance equal to its own impedance.

For the following problem, a generator has an output impedance of 10 kΩ and it produces a 10 V_{pp} (3.53 V_{rms}) signal. It will be connected to a 600 ohm line.

Questions

A. To properly match the generator to the line, what turns ratio is required?

B. Find the output voltage across the load.

C. Find the load current and power.

Answers

A. $\text{TR} = \sqrt{\dfrac{Z_G}{Z_L}} = \sqrt{\dfrac{10{,}000\,\Omega}{600\,\Omega}} = \dfrac{4.08}{1}$ or 4.08:1

B. $V_L = \dfrac{V_G}{\text{TR}} = \dfrac{10}{4.08} = 2.45\,V_{pp}$, which is 0.866 V_{rms}

C. $P_{in} = \dfrac{{V_G}^2}{Z_G} = \dfrac{(3.53)^2}{10{,}000} = 1.25\,\text{mW}$

NOTE **For the power calculation, you must use the RMS value of the voltage.**

$I_{in} = \dfrac{P_{in}}{V_{in}} = \dfrac{1.25\,\text{mW}}{3.53\,V_{rms}} = 0.354\,\text{mA}_{rms}$, which is 1 mA_{pp}

$I_L = I_{in}\,(\text{TR}) = 0.354 \times 4.08 = \mathbf{1.}445\ \text{mA}_{rms}$, which is $\mathbf{4.}08\ \text{mA}_{pp}$

$P_L = \dfrac{{V_L}^2}{Z_L} = \dfrac{(0.866)^2}{600} = 1.25\,\text{mW};$

which is the same as the input power. This circuit is shown in Figure 10-12.

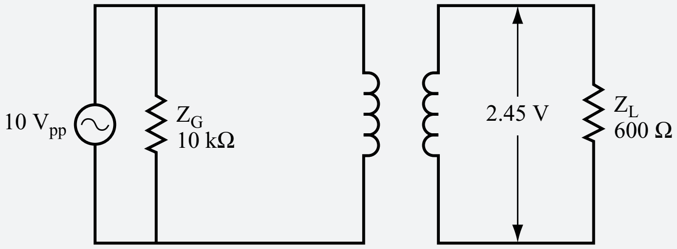

Figure 10-12

NOTE The generator now sees 10 kΩ when it looks toward the load, rather than the actual 600 ohm load. By the same token, the load now sees 600 ohms when it looks toward the source. This condition allows the optimum transfer of power to take place between the source and the load. In practice, however, the optimum condition as calculated here rarely exists. Because it may be impossible to obtain a transformer with a turns ratio of 4.08:1, you would have to select the closest available value, which might be a turns ratio of 4:1. The difference in the turns ratio affects the conditions at the load side, but only slightly.

17 In this problem, you use a transformer to match a generator to a load.

Questions

A. What turns ratio is required to match a generator that has a 2 kΩ output to a 600 ohm line?

____

B. If the generator produces 1 V_{pp}, what is the voltage across the load?

____

Answers

A. TR = 1.83

B. $V_L = 0.55\ V_{pp}$

18 In this problem you use a transformer to match a generator to a 2 kΩ load.

Questions

A. What turns ratio is required to match a 2 kΩ load with a source that has an output impedance of 5 kΩ?

____

B. If the load requires a power of 20 mW, what should the source be? (First, find the voltage across the load.)

____

C. What are the primary and secondary currents and the power supplied by the source to the primary side of the transformer?

____

Answers

A. TR = 1.58

B. $V_L = \sqrt{P_L \times Z_L} = \sqrt{20\ \text{mW} \times 2\ \text{k}\Omega} = 6.32\ V_{rms}$; and
$V_G = V_L \times TR = 6.32\ V_{rms} \times 1.58 = 10\ V_{rms}$

C. $I_L = 3.16\ \text{mA}_{rms}$, $I_p = 2\ \text{mA}_{rms}$, $P_{in} = 20\ \text{mW}$

Summary and Applications

In this chapter, you learned about the following topics related to transformers:

- The principles that allow an AC signal to be induced in a secondary coil
- How the AC voltage across the secondary coil can be stepped up or down depending upon the turns ratio of the transformer
- The use of a center tap to produce various voltages from a transformer
- The use of transformers to match impedances between a generator and a load
- That transformers can cause the output signal to be inverted (out of phase) from the input signal

Self-Test

These questions test your understanding of the material in this chapter. Use a separate sheet of paper for your diagrams or calculations. Compare your answers with the answers provided following the test.

1. How is a transformer constructed?

2. What type of signal is used as an input to a transformer?

3. If a sine wave is fed into a transformer shown in Figure 10-13, what does the output waveform look like?

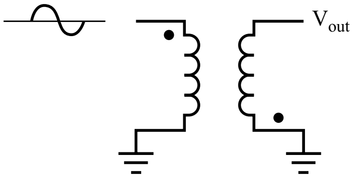

Figure 10-13

4. What is meant by the term *turns ratio*?

5. If $V_{in} = 1\ V_{pp}$ and $TR = 2$, what is the output voltage?

V_{out} __________

6. $V_{in} = 10\ V_{pp}$ and $V_{out} = 7\ V_{pp}$, what is the turns ratio?

TR = __________

7. In the center-tapped secondary winding shown in Figure 10-14, the voltage between points A and B may be expressed as $V_{A\text{-}B} = 28\ V_{pp}$. What is the voltage between C and A?

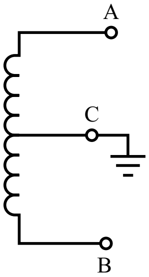

Figure 10-14

8. In the center-tapped secondary winding shown in Figure 10-14, the voltage between points B and C is $V_{B\text{-}C} = 5\ V_{rms}$. What is the peak-to-peak voltage between A and B?

9. If $I_{in} = 0.5\ A_{rms}$ and $I_{out} = 2.0\ A_{rms}$, what is the turns ratio?

10. Is the transformer in problem 9 a step up or a step down transformer?

____

11. If $Z_L = 600$ ohms and $Z_G = 6$ kΩ, find the turns ratio.

TR = ______________________________

12. If $Z_L = 1$ kΩ and the turns ratio is 10:1, what is the generator impedance?

$Z_G =$ ______________________________

Answers to Self-Test

If your answers do not agree with those given here, review the frames indicated in parentheses before you go on to the next chapter.

1.	Two coils of wire wound around a magnetic core (such as iron or ferrite)	(problem 1)
2.	An AC voltage — DC does not work.	(problem 2)
3.	An inverted sine wave	(problem 3)
4.	The ratio of the turns in the primary winding to the number of turns in the secondary coil	(problem 6)
5.	$V_{out} = 0.5$ V	(problem 7)
6.	TR = 1.43:1	(problem 7)
7.	$V_{C\text{-}A} = 14\ V_{pp}$	(problem 12)
8.	$V_{A\text{-}B} = 14.14\ V_{pp}$	(problem 12)
9.	TR = 4:1	(problem 13)
10.	It is a step down transformer. The voltage is lower (stepped down) in the secondary coil than in the primary coil if the current in the secondary coil is higher than the current in the primary coil. This maintains the same power on either side of the transformer.	(problem 13)
11.	TR = 3.2:1	(problem 16)
12.	$Z_G = 100$ kΩ	(problem 16)

CHAPTER 11

Power Supply Circuits

A power supply is incorporated into many electronic devices. Power supplies take the 120 V AC from a wall plug and convert it to a DC voltage, providing power for all types of electronic circuits.

Power supply circuits are very simple in principle, and those shown in this chapter have been around for many years. Because power supplies incorporate many of the features covered in this book, they make an excellent conclusion to your study of basic electronics.

Diodes are a major component in power supplies. Learning how AC signals are affected by diodes is fundamental to your understanding of how power supplies work. Therefore, this chapter begins with a brief discussion of diodes in AC circuits.

Note that throughout this chapter, diagrams show how AC signals are affected by diodes and other components in power supply circuits. If you have an oscilloscope, you can breadboard the circuits and observe these waveforms yourself.

When you complete this chapter, you will be able to do the following:

- Describe the function of diodes in AC circuits.
- Identify at least two ways to rectify an AC signal.
- Draw the output waveforms from rectifier and smoothing circuits.
- Calculate the output voltage from a power supply circuit.
- Determine the appropriate component values for a power supply circuit.

Diodes in AC Circuits Produce Pulsating DC

1 You can use diodes for several purposes in AC circuits, where their characteristic of conducting in only one direction is useful.

Questions

Assume that you apply +20 V DC at point A of the circuit shown in Figure 11-1.

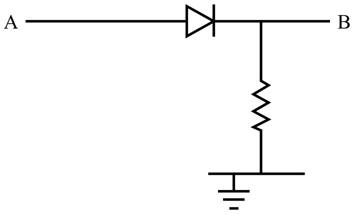

Figure 11-1

A. What is the output voltage at point B?

B. Suppose that you now apply +10 V DC at point B. What is the voltage at point A?

Answers

A. 20 V DC (Ignore, for now, the voltage drop of 0.7 V across the diode.)

B. 0 V (The diode is reverse biased.)

2 Figure 11-2 shows the circuit in Figure 11-1 with a 20 V_{pp} AC input signal centered at +20 V DC.

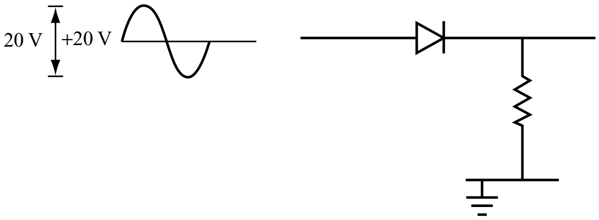

Figure 11-2

Questions

A. What are the positive and negative peak voltages of the input signal?

B. What is the output waveform of this circuit?

Answers

A. Positive peak voltage is 20 V + 10 V = 30 V.

Negative peak voltage is 20 V − 10 V = 10 V.

B. The diode is always forward biased, so it always conducts. Thus, the output waveform is exactly the same as the input waveform.

3 Figure 11-3 shows a circuit with 20 V_{pp} AC input signal centered at 0 V DC.

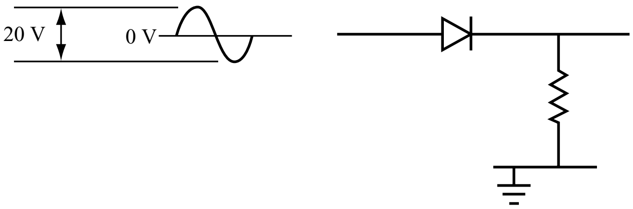

Figure 11-3

Questions

A. What are the positive and negative peak voltages of the input signal?

B. Draw the output waveform on the blank graph provided in Figure 11-4. Remember that only the positive portion of the input waveform passes through the diode.

Figure 11-4

Answers

A. Positive peak voltage is +10 V. Negative peak voltage is −10 V.

B. See Figure 11-5.

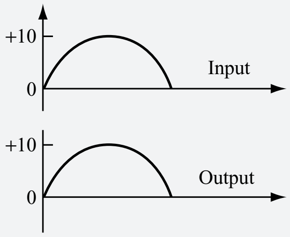

Figure 11-5

4 When the input is negative, the diode in the circuit shown in Figure 11-3 is reverse biased. Therefore, the output voltage remains at 0 V.

Question

Figure 11-6 shows the input waveform for the circuit shown in Figure 11-3. On a separate sheet of paper, draw the output waveform.

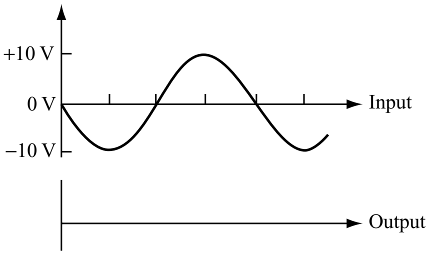

Figure 11-6

Answer

See Figure 11-7.

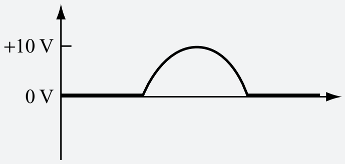

Figure 11-7

5 Figure 11-7 shows the output waveform, of the circuit shown in Figure 11-3, for one complete cycle of the input waveform.

Question

Now, draw the output waveform for three complete cycles of the input waveform shown in Figure 11-6. Use a separate sheet of paper for your drawing.

Answer

See Figure 11-8.

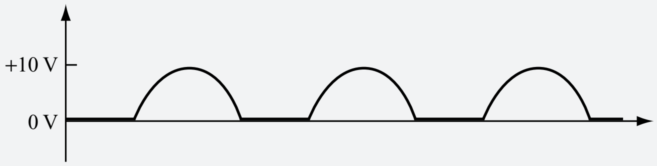

Figure 11-8

6 When the diode is connected in the opposite direction, it is forward biased and, therefore, conducts current when the input signal is negative. In this case, the diode is reverse biased when the input signal is positive. Therefore, the output waveform is inverted from the output waveform shown in Figure 11-8.

Question

On a separate sheet of paper, draw the output waveform for three input cycles, assuming that the diode is connected in the opposite direction from the diode shown in Figure 11-3.

Answer

See Figure 11-9.

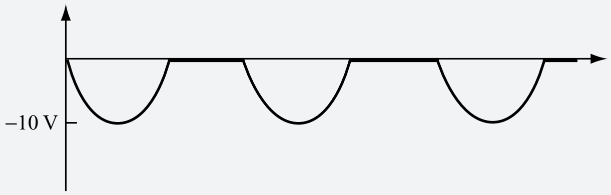

Figure 11-9

7 Figure 11-10 shows a circuit with a 20 V_{pp} AC input signal centered at −20 V DC.

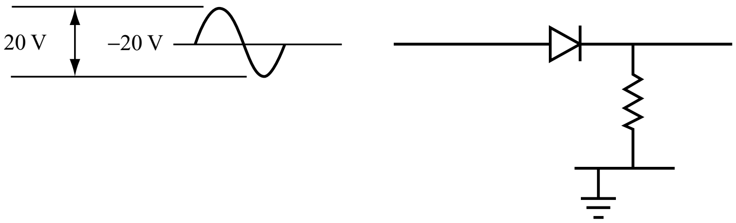

Figure 11-10

Questions

A. When is the diode forward biased?

B. What is the output voltage?

Answers

A. Never, because the voltage that results from adding the AC and DC signals ranges from −10 V to −30 V. Therefore, the diode is always reverse biased.

B. A constant 0 V

8 As you have seen, a diode passes either the positive or negative portion of an AC voltage waveform, depending on how you connect it in a circuit. Therefore, the AC input signal is converted to a pulsed DC output signal, a process called *rectification*. A circuit that converts either the positive or negative portion of an AC voltage waveform to a pulsed DC output signal is called a *half-wave rectifier*.

Question

Refer to the output waveforms shown in Figures 11-8 and 11-9. Do these waveforms represent positive DC voltage pulses or negative DC voltage pulses?

Answer

The waveform In Figure 11-8 represents positive pulses of DC voltage. The waveform in Figure 11-9 represents negative pulses of DC voltage.

9 The circuit shown in Figure 11-11 shows a diode connected to the secondary coil of a transformer.

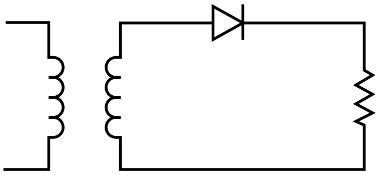

Figure 11-11

Questions

A. How does the diode affect the AC signal?

____

B. Draw the waveform of the voltage across the load for the circuit shown in Figure 11-11 if the secondary coil of the transformer has a 30 V_{pp} AC output signal centered at 0 V DC. Use a separate sheet of paper for your drawing.

Answers

A. The AC signal is rectified.

B. See Figure 11-12. This type of circuit (called a *half-wave rectifier*) produces in an output waveform containing either the positive or negative portion of the input waveform.

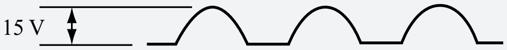

Figure 11-12

10 Figure 11-13 shows the waveforms at each end of a center-tap transformer secondary coil. Diode D_1 rectifies the waveform shown at point A and diode D_2 rectifies the waveform shown at point B.

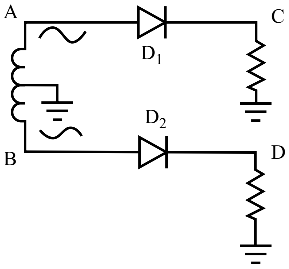

Figure 11-13

Questions

A. Which diode conducts during the first half of the cycle?

B. Which diode conducts during the second half of the cycle?

C. Draw the input waveforms (points A and B), and underneath draw each output waveform (points C and D). Use a separate sheet of paper for your drawing.

Answers

A. During the first half of the cycle, D_1 is forward biased and conducts current. D_2 is reverse biased and does not conduct current.

B. During the second half of the cycle, D_2 is forward biased and conducts current. D_1 is reversed biased and does not conduct current.

C. See Figure 11-14.

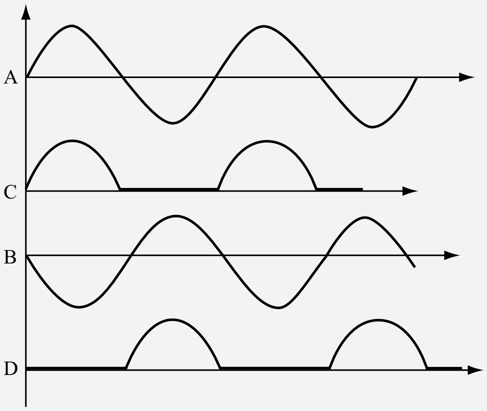

Figure 11-14

11 Figure 11-15 shows a circuit in which diodes connected to the ends of a center-tap transformer are connected to ground through a single resistor. The output voltage waveforms from both diodes are therefore applied across one load resistor. This type of circuit is called a *full-wave rectifier*.

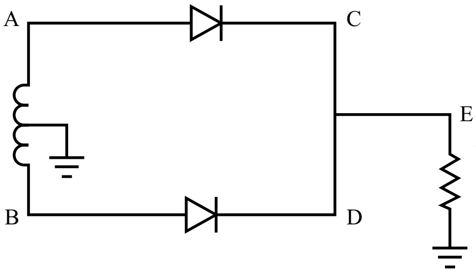

Figure 11-15

Question

On a separate sheet of paper, draw the waveform representing the voltage at point E in the circuit shown in Figure 11-15. (This waveform is a combination of the waveforms at points C and D shown in Figure 11-14.)

Answer

See Figure 11-16.

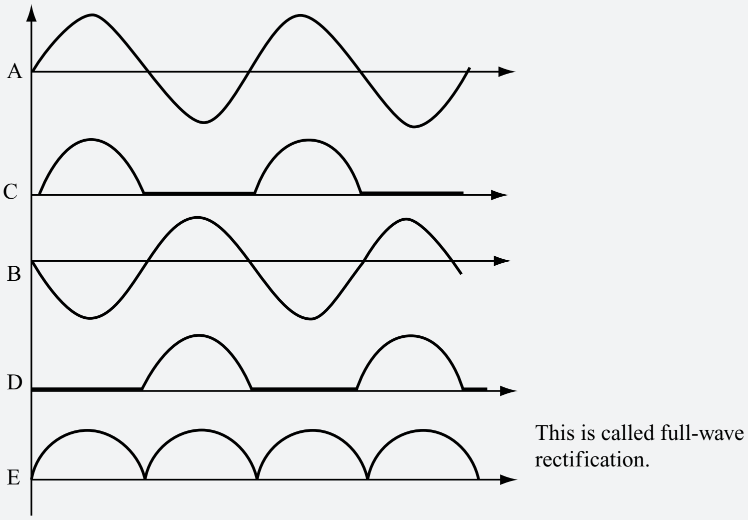

Figure 11-16

12 Full-wave rectification of AC allows a much ''smoother'' conversion of AC to DC than does half-wave rectification.

Figure 11-17 shows a full-wave rectifier circuit that uses a transformer with a two terminal secondary coil, rather than a center-tapped secondary coil.

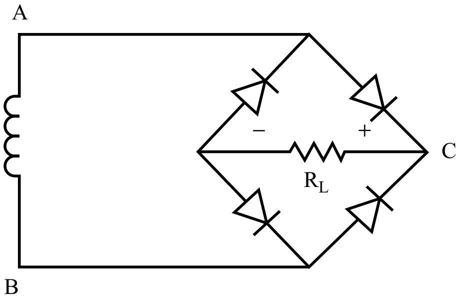

Figure 11-17

Question

How does this circuit differ from the circuit shown in Figure 11-15?

Answer

This circuit has no center tap on the secondary coil, and it uses four diodes.

13 Figure 11-18 shows the direction of current flow when the voltage at point A is positive.

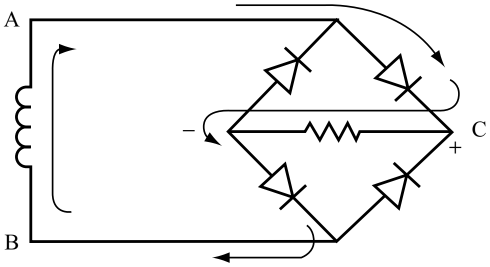

Figure 11-18

Figure 11-19 shows the direction of current flow when the voltage at point B is positive.

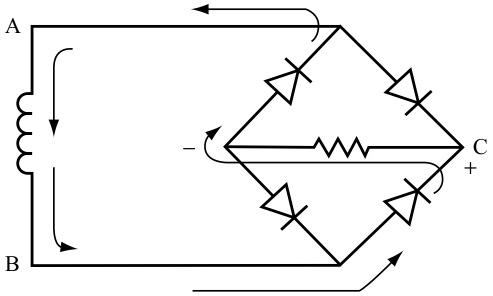

Figure 11-19

Notice that the direction of current through the load resistor is the same in both cases.

Question

A. Through how many diodes does the current travel in each conduction path?

B. Draw the voltage waveform at point C. Use a separate sheet of paper for your drawing.

Answer

A. Two diodes in each case

B. See Figure 11-20.

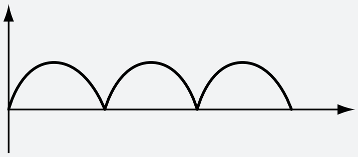

Figure 11-20

What this chapter has explored to this point is how AC is turned into pulsating DC. In fact, rectified AC is often called *pulsating DC*. The next step in your understanding of power supplies is to learn how you turn pulsating DC into level DC.

Level DC (Smoothing Pulsating DC)

14 A basic power supply circuit can be divided into four sections, as shown in Figure 11-21.

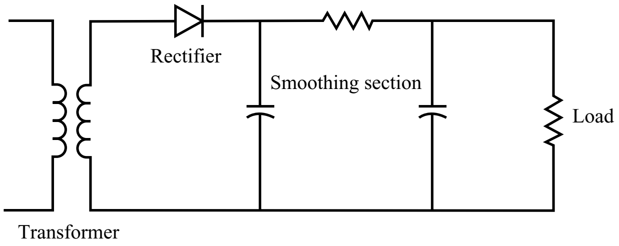

Figure 11-21

Questions

A. If you use a center-tap transformer in a power supply, how many diodes would you need to produce a full-wave rectified output?

____

B. Will the power supply circuit shown in Figure 11-21 result in full- or half-wave rectification?

____

C. What type of output will the rectifier section of the power supply circuit shown in Figure 11-21 produce?

____

Answers

A. Two

B. Half-wave

C. Pulsating DC

15 The function of the smoothing section of a power supply circuit is to take the pulsating DC (PDC) and convert it to a ''pure'' DC with as little AC ''ripple'' as possible. The smoothed DC voltage, shown in the illustration on the right in Figure 11-22, is then applied to the load.

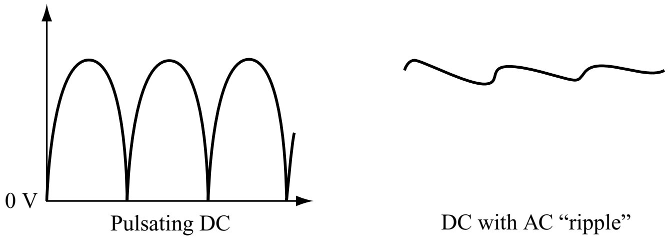

Figure 11-22

The load (which is ''driven'' by the power supply) can be a simple lamp or a complex electronic circuit. Whatever load you use, it will require a certain voltage across its terminals and will draw a current. Therefore the load will have a resistance.

Usually the voltage and current required by the load (and, hence, its resistance) are known, and you must design the power supply to provide that voltage and current.

To simplify the circuit diagrams, you can treat the load as a simple resistor.

Questions

A. What does the smoothing section of a power supply do?

____

B. What is connected to a power supply, and what can you treat it like?

____

Answers

A. The smoothing section converts the pulsating DC to a ''pure'' DC.

B. A load such as a lamp or an electronic circuit is connected to a power supply. In most cases you can treat the load as you do a resistor.

16 Figure 11-23 shows a power supply circuit with a resistor as the load.

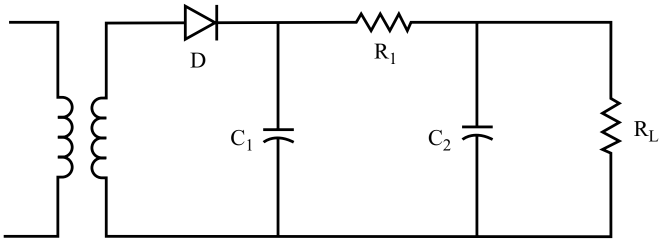

Figure 11-23

Questions

Look at the circuit shown in Figure 11-23 and answer these questions.

A. What type of secondary coil is used?

____

B. What type of rectifier is used?

____

C. What components make up the smoothing section?

____

D. What output would you expect from the rectifier section?

____

Answers

A. A secondary coil with no center tap

B. A single diode half-wave rectifier

C. A resistor and two capacitors (R_1, C_1, and C_2)

D. Half-wave pulsating DC

17 Figure 11-24 shows the output waveform from the rectifier portion of the power supply circuit shown in Figure 11-23.

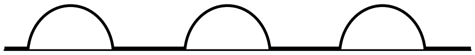

Figure 11-24

This waveform is the input to the smoothing section of the power supply circuit. Use one of the DC pulses (shown in Figure 11-25) to analyze the effect of the smoothing section on the waveform.

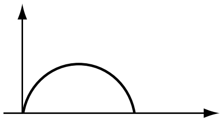

Figure 11-25

As the voltage level of the DC pulse rises to its peak, the capacitor C_1 is charged to the peak voltage of the DC pulse.

When the input DC pulse drops from its peak voltage back to 0 volts, the electrons stored on capacitor C_1 discharge through the circuit. This maintains the voltage across the load resistor at close to its peak value, as shown in Figure 11-26. Note that the DC pulse to the right of the diode stays at the peak voltage, even though V_{in} drops to zero.

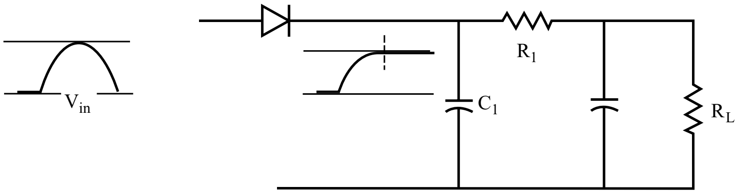

Figure 11-26

Question

What discharge path is available for the capacitor C_1?

Answer

The diode is not conducting, so the capacitor cannot discharge through the diode. The only possible discharge path is through R_1 and the load R_L.

18 If no further pulses pass through the diode, the voltage level drops as the capacitor discharges, resulting in the waveform shown in Figure 11-27.

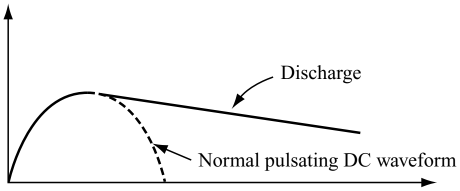

Figure 11-27

However, if another pulse passes through the diode before the capacitor is discharged, the resulting waveform looks like that shown in Figure 11-28.

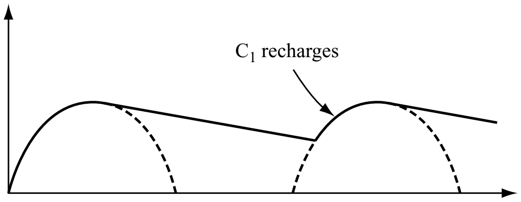

Figure 11-28

The capacitor only discharges briefly before the second pulse recharges it to peak value. Therefore, the voltage applied to the load resistor only drops a small amount.

Applying further pulses can produce this same recharging effect again and again. Figure 11-29 shows the resulting waveform.

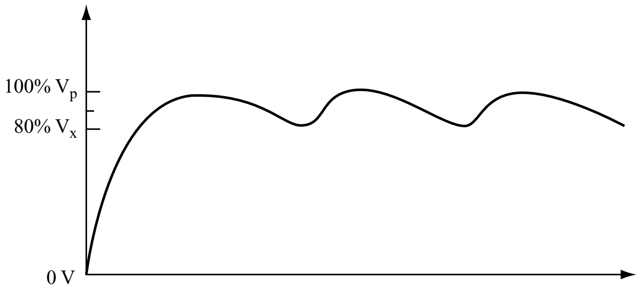

Figure 11-29

The waveform in Figure 11-29 has a DC level with an AC ripple, which varies between V_p and V_x. If you choose values of C_1, R_1, and R_L that produce a discharge time constant for C_1 equal to about 10 times the duration of an input pulse, V_x will be about 80 percent of V_p.

If the discharge time you select is greater than 10 times the duration of an input pulse, the smoothing effect minimizes the AC ripple. A time constant of 10 times the pulse duration results in practical design values that are used throughout this chapter.

NOTE The smoothing section of a power supply circuit is sometimes referred to as a *low pass filter*. Though such a circuit can function as a low pass filter, in the case of a power supply circuit converting AC to DC, it is the release of electrons by the capacitor that is primarily responsible for leveling out the pulsating DC. For that reason, this discussion uses the term *smoothing section*.

Question

Estimate the average DC output level of the waveform shown in Figure 11-29.

Answer

About 90 percent of V_p

19 The output from the secondary coil of the circuit shown in Figure 11-30 is a 28 V_{rms}, 60 Hz sine wave. For this circuit, you need 10 V DC across the 100 ohm load resistor.

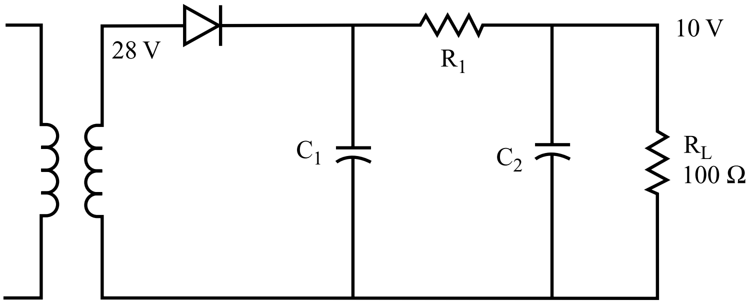

Figure 11-30

Question

What is the peak voltage out of the rectifier?

Answer

The transformer secondary coil delivers 28 V_{rms}, so

$$V_p = \sqrt{2} \times V_{rms} = 1.414 \times 28V_{rms} = 39.59V$$

or about 40 V.

20 Figure 11-31 shows the waveform after the diode has rectified the sine wave for the halfway rectifier circuit shown in Figure 11-30.

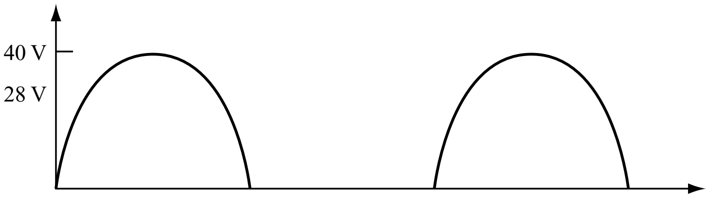

Figure 11-31

Question

Calculate the duration of one pulse.

Answer

60 Hz represents 60 cycles (that is, *wavelengths*) in 1 second. Therefore, one wavelength lasts for 1/60 second.

1/60 second = 1000/60 milliseconds = 16.67 ms

Therefore, the duration of a pulse, which is half a wavelength, is 8.33 ms.

21 The average DC voltage at point B in the circuit shown in Figure 11-30 is about 90 percent of the peak value of the sine wave from the secondary coil, or $V_B = 0.9 \times 40\text{ V} = 36\text{ V}$. R_1 and R_L act as a voltage divider to reduce the 36 V DC level to the required 10 V DC at the output.

Question

Using the voltage divider formula, calculate the value of R_1 that will result in 10 V DC across the 100 ohm load resistor.

Answer

$$V_{out} = \frac{V_{in}R_L}{(R_1 + R_L)}$$

$$10 = \frac{36 \times 100}{(R_1 + 100)}$$

Therefore, $R_1 = 260$ ohms

22 Figure 11-32 shows the half-wave rectifier circuit with the 260 ohm value you calculated for R_1.

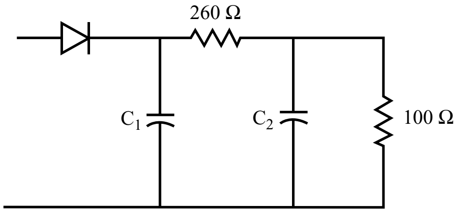

Figure 11-32

Now, choose a value for C_1 that produces a discharge time through the two resistors equal to 10 times the input wave duration.

Questions

A. How long should the discharge time constant be for the circuit in Figure 11-32? Refer to problems 18 and 20.

B. Given the time constant, calculate the value of C_1.

Answers

A. The time constant should be ten times the pulse duration (8.33 ms), so:

$\tau = 10 \times 8.33 \text{ ms} = 83.3 \text{ ms}$ or 0.083 seconds

B. $\tau = R \times C = (R_1 + R_L) \times C_1 = 360 \times C_1$

Therefore, $0.0833 = 360 \times C_1$ or $C_1 = 230\ \mu\text{F}$

23 Figure 11-33 shows voltage waveforms at various points in the half-wave rectifier circuit.

Questions

A. What happens to the DC output voltage between points B and C in this circuit?

B. What happens to the waveform between points A and C in the circuit?

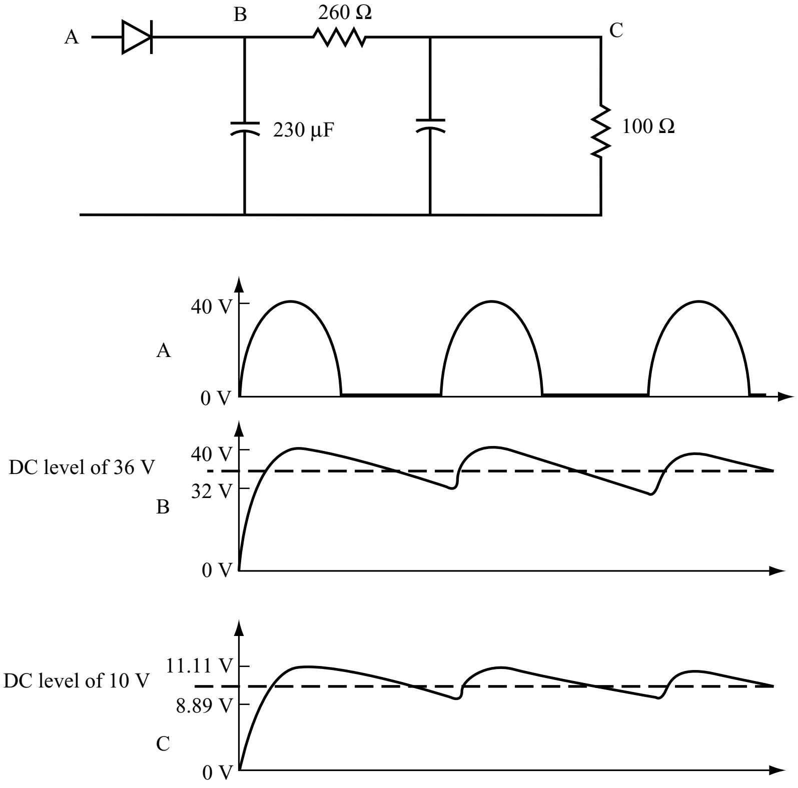
A
B
260 Ω
C
230 µF
100 Ω
40 V
A
0 V
DC level of 36 V
40 V
32 V
B
0 V
DC level of 10 V
11.11 V
8.89 V
C
0 V

Figure 11-33

Answers

A. The voltage has been reduced from 36 V to 10 V.

B. The waveform has changed from pulsating DC to a 10 V DC level with an AC ripple.

24 In most cases, the level of the AC ripple is still too high, and further smoothing is required. Figure 11-34 shows the portion of the half-wave rectifier circuit that forms a voltage divider using R_1 and the parallel combination of R_L and C_2. This voltage divider reduces the AC ripple and the DC voltage level.

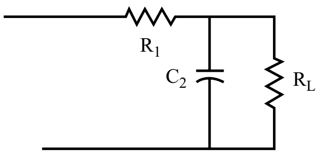

Figure 11-34

Choose a value for C_2 that causes the capacitor's reactance (X_{C2}) to be equal to or less than one tenth of the resistance of the load resistor. C_2, R_1, and R_2 form an AC voltage divider. As discussed in problem 26 of Chapter 6, choosing such a value for C_2 simplifies the calculations for an AC voltage divider circuit containing a parallel resistor and capacitor.

Questions

A. What should the value of X_{C2} be?

B. What is the formula for the reactance of a capacitor?

C. What is the frequency of the AC ripple?

D. Calculate the value of the capacitor C_2.

Answers

A. $X_{C2} = R_L/10 = 100/10 = 10\ \Omega$ or less

B. $X_C = \dfrac{1}{2\pi fC}$

C. 60 Hz. This is identical to the frequency of the sine wave output from the transformer's secondary coil.

D. Solving the reactance formula for C results in the following:

$$C_2 = \frac{1}{2\pi f X_c} = \frac{1}{2 \times \pi \times 60\,\text{Hz} \times 10\,\Omega} = 265\,\mu\text{F}$$

25 Figure 11-35 shows the half-wave rectifier circuit with all capacitor and resistor values.

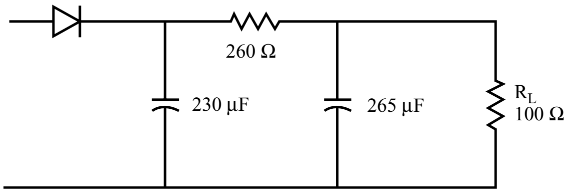

Figure 11-35

Because X_{c2} is one tenth of R_L, you can ignore R_L in AC voltage divider calculations. The resulting AC voltage divider circuit is shown in Figure 11-36.

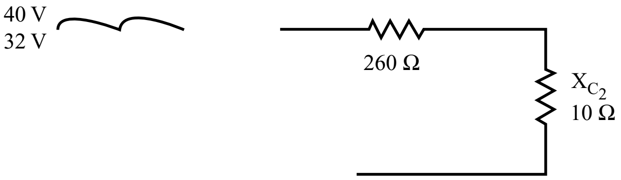

Figure 11-36

Questions

A. What is the peak-to-peak voltage at the input to the AC voltage divider?

B. Find the AC ripple output across R_L using the AC voltage divider formula discussed in problem 26 of Chapter 6.

Answers

A. $V_{pp} = V_p - V_x = 40\text{ V} - 32\text{ V} = 8\text{ V}_{PP}$

B. $$\text{AC V}_{out} = (\text{AC V}_{in}) \times \frac{X_{C2}}{\sqrt{X_{C2}{}^2 + R_1{}^2}}$$

$$\text{AC V}_{out} = 8 \times \frac{10}{\sqrt{(10^2 + 260^2)}} = 0.31\text{ V}_{pp}$$

NOTE **This result means that the addition of C_2 lowers the AC ripple shown by curve C of Figure 11-33, with peak values of 11.11 and 8.89, to values of 10.155 and 9.845 V. This represents a lower ripple at the output. Hence, C_2 aids the smoothing of the 10 V DC at the output.**

26 You can apply the calculations you performed for a half-wave rectifier circuit in the last few problems to a full-wave rectifier circuit. In the next few problems, you calculate the values of R_1, C_1, and C_2 required to provide

10 volts DC across a 100 ohm load for a full-wave rectifier circuit with a 28 V_{rms} sine wave supplied by the secondary coil of a transformer.

Figure 11-37 shows the output waveform from the rectifier section of the circuit.

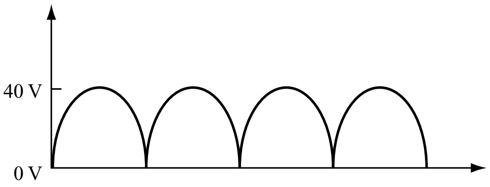

Figure 11-37

Figure 11-38 shows the waveform that results from using a smoothing capacitor

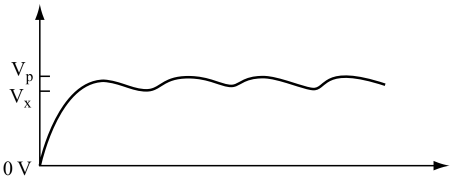

Figure 11-38

If C_1's discharge time constant is ten times the period of the waveform, V_x is about 90 percent of V_p. The average DC level is about 95 percent of V_p.

Questions

A. What is the average DC level for the half-wave rectifier at point B in Figure 11-33?

____

B. What is the average DC level for the waveform in Figure 11-38 given that $V_p = 40$ V?

____

C. Why does a full-wave rectifier have a higher average DC level than a half-wave rectifier?

____

Answers

A. 36 V, which was 90 percent of V_p

B. 38 V, which is 95 percent of V_p

C. The slightly higher values occur because the capacitor does not discharge as far with full-wave rectification, and, as a result, there is slightly less AC ripple. Therefore, V_x is higher and the average DC level is higher.

27 You can use the method for calculating the value of R_1 for a half-wave rectifier (see problem 21) to calculate the value of R_1 for a full-wave rectifier.

Question

Calculate the value of R_1 when $R_L = 100\ \Omega$, $V_{in} = 38$ V, and the required voltage across R_L is 10 V.

Answer

$$V_{out} = 10\text{V} = \frac{V_{in}R_L}{(R_1 + R_L)} = \frac{38 \times 100}{(R_1 + 100)}$$

Therefore, $R_1 = 280$ ohms.

28 You can also use the method for calculating the value of C_1 for a half-wave rectifier (see problem 22) to calculate the value of C_1 for a full-wave rectifier.

Question

Calculate the value of C_1.

Answer

With a time constant of $\tau = 83.3$ ms and a discharge resistance of $R_1 + R_L = 380$ ohms, $C_1 = 220\ \mu\text{F}$.

29 You can use the voltage divider equation to find the amount of AC ripple across the load resistor for a full-wave rectifier with $R_1 = 280\ \Omega$ and $R_L = 100\ \Omega$. For $V_p = 40$ V, the calculation results in 10.52 V. For $V_x = 36$ V, the calculation results in 9.47 V. Therefore, the voltage levels at the load resistor vary between 10.52 V and 9.47 V, with an average DC level of 10 V. You can reduce the AC ripple by adding a second capacitor in parallel with the load resistor.

Questions

Use the method for calculating the value of C_2 for the half-wave rectifier in problem 24.

A. Calculate the reactance of the second capacitor (C_2).

____

B. Calculate the value of C_2. (Note that the frequency of the AC ripple for the full-wave rectifier is 120 Hz.)

____

Answers

A. The reactance should be one tenth (or less) of the load resistance. Therefore, it should be 10 ohms or less.

B. $$C_2 = \frac{1}{2\pi f X_C} = \frac{1}{2 \times \pi \times 120\,\text{Hz} \times 10\,\Omega} = 135\,\mu\text{F}$$

30 The AC ripple at the first smoothing capacitor ranges from 36 V to 40 V. The AC ripple at the load ranges from 9.47 V to 10.52 V when there is only one capacitor in the circuit.

Question

Calculate the upper and lower values of the AC ripple at the output if you use a second capacitor with a value of 135 μF in parallel with the load resistor. You can use the same formulas as those for the half-wave rectifier in problem 25. Note that X_{C2} is 10 Ω from problem 29; $R_1 = 280$ from problem 27; and AC $V_{in} = V_p - V_x = (40\text{ V} - 36\text{ V}) = 4\ V_{pp}$.

____

Answer

$$\text{AC V}_{\text{out}} = (\text{AC V}_{\text{in}}) \times \frac{X_{C2}}{\sqrt{(R_1{}^2 + X_{C2}{}^2)}} = 0.143\,\text{V}_{\text{pp}}$$

The result of approximately 0.14 V_{pp} means that the output will now vary from 10.07 to 9.93 V. This shows that the second capacitor lowers the ripple significantly. The AC ripple is less than half of the ripple shown for the half-wave rectifier in problem 25. In other words, a full-wave rectifier produces a smoother DC output than a half-wave rectifier.

31 Figure 11-39 shows a full-wave rectifier circuit with an output voltage of 5 V across a 50 Ω load resistor. Use the following steps to calculate the values of the other components.

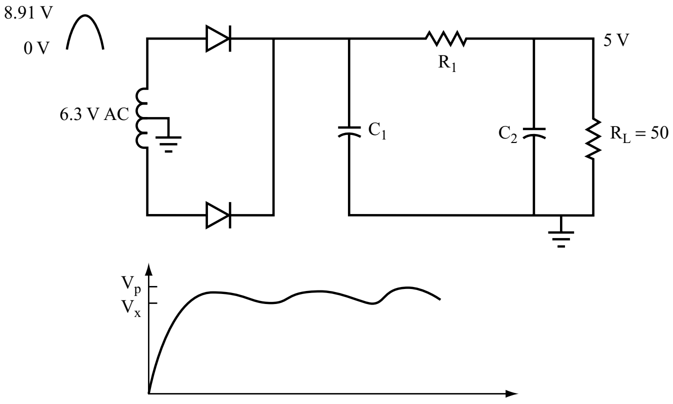

Figure 11-39

Questions

A. What are V_p, V_x, and the DC level at the first capacitor?

B. Calculate the value of R_1 required to make the DC level at the output 5 V.

C. Calculate the value of C_1.

D. Calculate the value of C_2.

E. What is the amount AC ripple at the output?

F. Draw the final circuit showing the calculated values. Use a separate sheet of paper for your drawing.

Answers

A. $V_p = 6.3 \times \sqrt{2} = 8.91\,V$, $V_x = 90$ percent of $V_p = 8.02\,V$

The DC level is 95 percent of V_p, which is 8.46 V.

B. About 35 ohms.

C. 980 μF.

D. Using $X_{c2} = 5$ ohms and 120 Hz, $C_2 = 265\,\mu F$.

E. At the input to the smoothing section, the AC variation is 8.91 to 8.02, or 0.89 V_{pp}. Using the AC voltage divider equation with $R_1 = 35$ ohms and $X_{c2} = 5$ ohms, AC V_{out} equals about 0.13 V_{pp}. Therefore, the AC variation at the output is 5.065 to 4.935 V, a very small AC ripple.

F. See Figure 11-40.

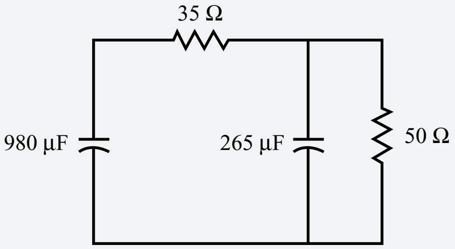

Figure 11-40

Using the simple procedure shown here always produces a working power supply circuit. This is not the only design procedure you can use for power supplies, but it is one of the simplest and most effective.

Summary

This chapter introduced the following concepts and calculations related to power supplies:

THE EFFECTS OF DIODES ON AC SIGNALS

- Methods of rectifying an AC signal
- Half-wave and full-wave rectifier circuit designs
- The calculations you can use to determine component values for half-wave and full-wave rectifier power supply circuits

Self-Test

These questions test your understanding of the information presented in this chapter. Use a separate sheet of paper for your diagrams or calculations. Compare your answers with the answers provided following the test.

In questions 1 through 5, draw the output waveform of each circuit. The input is given in each case.

1. See Figure 11-41.

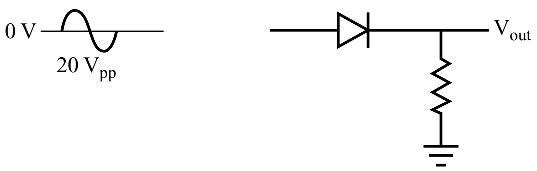

Figure 11-41

2. See Figure 11-42.

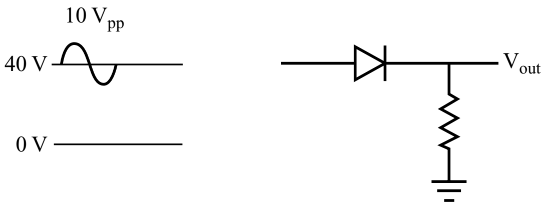

Figure 11-42

3. See Figure 11-43.

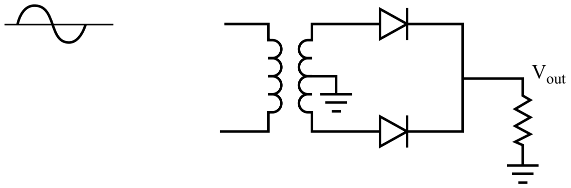

Figure 11-43

4. See Figure 11-44.

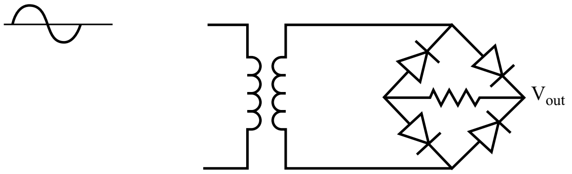

Figure 11-44

5. See Figure 11-45.

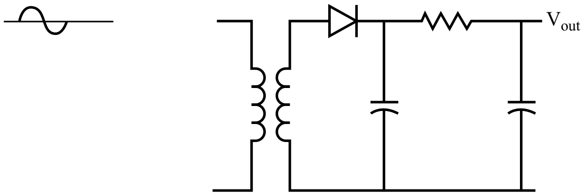

Figure 11-45

6. In the circuit shown in Figure 11-46, 100 V_{rms} at 60 Hz appears at the secondary coil of the transformer; 28 V DC with as little AC ripple as possible is required across the 220 ohm load. Find R_1, C_1, and C_2. Find the approximate AC ripple.

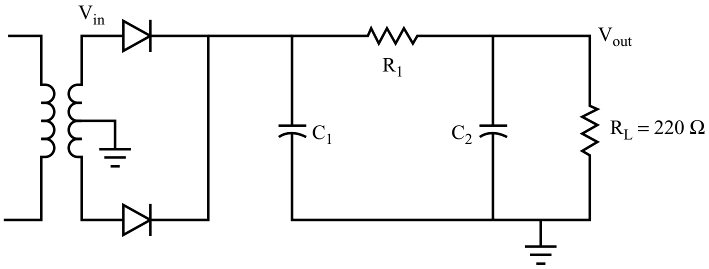

Figure 11-46

Answers to Self-Test

If your answers do not agree with those given here, review the problems indicated in parentheses.

1. See Figure 11-47. (problems 1–5)

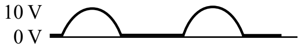

Figure 11-47

2. See Figure 11-48. (problem 2)

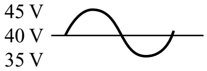

Figure 11-48

3. See Figure 11-49. (problem 11)

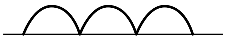

Figure 11-49

4. See Figure 11-50. (problem 13)

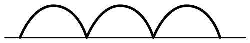

Figure 11-50

5. See Figure 11-51. (problems 15–18)

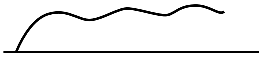

Figure 11-51

6. $R_1 = 833$ ohms, $C_1 = 79\ \mu F$, let $X_{c2} = 22$ ohms and then $C_2 = 60\ \mu F$ (problems 26–30)

$$\text{AC}\,V_{out} = 14 \times \frac{22}{\sqrt{(22^2 \times 833^2)}} = 0.37\,V_{pp}$$

CHAPTER 12

Conclusion and Final Self-Test

In this book, you have discovered basic concepts and formulas that will provide a foundation for your studies in modern electronics, whether you become a dedicated hobbyist, or go on to study electrical or electronics engineering.

Conclusion

Having read this book, you should now know enough to read intermediate-level electronics books and articles intelligently, to build electronics circuits and projects, and to pursue electronics to whatever depth and for whatever reason you wish. Specifically, you should now be able to do the following:

- Recognize all the important, discrete electronics components in a schematic diagram.
- Understand how circuits that use discrete components work.
- Calculate the component values needed for circuits to function efficiently.
- Design simple circuits.
- Build simple circuits and electronics projects.

To see how much you have learned, you may want to take the final self-test at the end of this chapter. It tests your comprehension of the concepts and formulas presented throughout this book.

Once you complete the following self-test and feel confident that you have mastered the information in this book, refer to Appendix E for additional resources for further learning, including the following:

- Books such as *The Art of Electronics* by Paul Horowitz and Winfield Hill (New York: Cambridge University Press, 1989) provide a great next step in further electronics study.
- You can buy books that step you through electronics projects, such as *Electronics Projects For Dummies* by Earl Boysen and Nancy Muir (Indianapolis: Wiley, 2006), to help you get set up and get hands-on practice.
- You can browse Web sites for electronic project ideas.
- Earl Boysen's Web site, `www.buildinggadgets.com`, provides tips, ideas, and links to a variety of great online resources.

NOTE **For those interested in more serious study, you should be aware that there is a difference between the path you take to become an electrician (or technician) and an electrical (or electronics) engineer. Training for electronics technicians is available in military trade schools, public and private vocational schools, and in many high schools. Engineers are required to understand the mathematical details in more depth, and must take at least a four-year curriculum at an accredited college or university.**

Whatever your goal, you can feel confident that this book has given you a solid grounding for your future studies. Wherever you go in electronics, good luck!

Final Self-Test

This final test allows you to assess your overall knowledge of electronics. Answers and review references follow the test. Use a separate sheet of paper for your calculations and drawings.

1. If R = 1 MΩ and I = 2 μA, find the voltage.

2. If V = 5 V and R = 10 kΩ, find the current.

3. If V = 28 V and I = 4 A, find the resistance.

4. If 330 ohms and 220 ohms are connected in parallel, find the equivalent resistance.

5. If V = 28 V and I = 5 mA, find the power.

6. If the current through a 220 ohm resistor is 30.2 mA, what is the power dissipated by the resistor?

7. If the power rating of a 1000 ohm resistor is 0.5 watts, what will be the maximum current that should safely flow through the resistor?

8. If a 10 ohm resistor is in series with a 32 ohm resistor and the combination is across a 12 volt supply, what is the voltage drop across each resistor, and what will the two voltage drops add up to?

9. A current of 1 ampere splits between 6 ohm and 12 ohm resistors in parallel. Find the current through each.

10. A current of 273 mA splits between 330 ohm and 660 ohm resistors in parallel. Find the current through each resistor.

11. If $R = 10\ k\Omega$ and $C = 1\ \mu F$, find the time constant.

12. If $R = 1\ M\Omega$ and $C = 250\ \mu F$, find the time constant.

13. Three capacitors of 1 μF, 2 μF, and 3 μF are connected in parallel. Find the total capacitance.

14. Three capacitors of 100 μF, 220 μF, and 220 μF are connected in series. Find the total capacitance.

15. Three capacitors of 22 pF, 22 pF, and 33 pF are connected in series. Find the total capacitance.

16. What is the knee voltage for a germanium diode?

17. What is the knee voltage for a silicon diode?

18. In the circuit shown in Figure 12-1, $V_S = 5$ V, R = 1 kΩ, find the current through the diode, I_D.

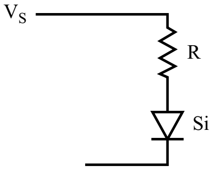

Figure 12-1

19. For the circuit in question 15, V = 12 V and R = 100 ohms. Find I_D.

20. For the circuit shown in Figure 12-2, $V_S = 100$ V, $R_1 = 7.2$ kΩ, $R_2 = 4$ kΩ, and $V_Z = 28$ V. Find the current through the zener diode, I_Z.

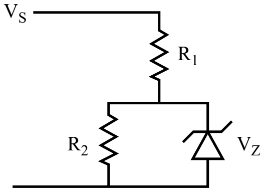

Figure 12-2

21. For the circuit in Figure 12-2, $V_S = 10$ V, $R_1 = 1$ kΩ, $R_2 = 10$ kΩ, and $V_Z = 6.3$ V. Find I_Z.

22. Use the circuit shown in Figure 12-3 to answer this question.

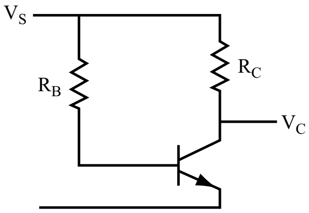

Figure 12-3

Find the DC collector voltage, V_C, if $V_S = 28$ V, $\beta = 10$, $R_B = 200$ kΩ, $R_C = 10$ kΩ.

23. Again using the circuit shown in Figure 12-3, find R_B if $V_S = 12$ V, $\beta = 250$, $R_C = 2.2$ kΩ, $V_C = 6$ V.

24. Using the circuit shown in Figure 12-3, find β if $V_S = 10$ V, $R_B = 100$ kΩ, $R_C = 1$ kΩ, $V_C = 5$ V.

25. What are the three terminals for a JFET called, and which one controls the operation of the JFET?

When is the drain current of a JFET at its maximum value?

26. Use the circuit shown in Figure 12-4 to answer this question.

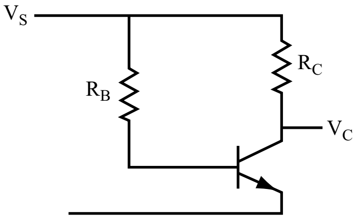

Figure 12-4

Find the value of R_B required to turn the transistor ON if $V_S = 14$ V, $R_C = 10$ kΩ, $\beta = 50$.

27. Again, using the circuit shown in Figure 12-4, find the value of R_B required to turn the transistor ON if $V_S = 5$ V, $R_C = 4.7$ kΩ, $\beta = 100$.

28. Use the circuit shown in Figure 12-5 to answer this question.

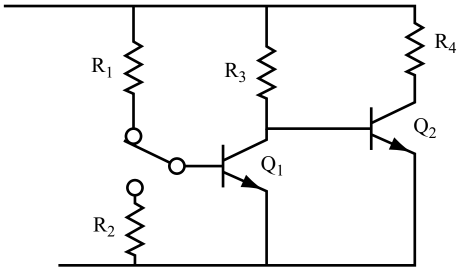

Figure 12-5

Find the values of R_1, R_2, and R_3 that will enable the switch to turn Q_2 ON and OFF, if $V_S = 10$ V, $\beta_1 = 50$, $\beta_2 = 20$, $R_4 = 2.2$ kΩ.

---

29. Again, using the circuit shown in Figure 12-5, find the values of R_1, R_2, and R_3 that will enable the switch to turn Q_2 ON and OFF if $V_S = 28$ V, $\beta_1 = 30$, $\beta_2 = 10$, $R_4 = 220$ Ω.

---

30. An N-channel JFET has a drain saturation current of $I_{DSS} = 14$ mA. If a 28 V drain supply is used, calculate the drain resistance, R_D.

---

31. Draw one cycle of a sine wave.

---

32. Mark in V_{pp}, V_{rms}, and the period of the waveform on your drawing for question 31.

---

33. If $V_{pp} = 10$ V, find V_{rms}.

---

34. If $V_{rms} = 120$ V, find V_{pp}.

---

35. If the frequency of a sine wave is 14.5 kHz, what is the period of the waveform?

---

36. Find the reactance X_C for a 200 μF capacitor when the frequency is 60 Hz.

---

37. Find the value of the capacitance that gives a 50 ohm reactance at a frequency of 10 kHz

---

38. Find the inductive reactance X_L for a 10 mH inductor when the frequency is 440 Hz.

---

39. Find the value of the inductance that has 100 ohms reactance when the frequency is 1 kHz.

40. Find the series and parallel resonant frequency of a 0.1 μF capacitor and a 4 mH inductor that has negligible internal resistance.

41. For this question refer to the circuit shown in Figure 12-6.

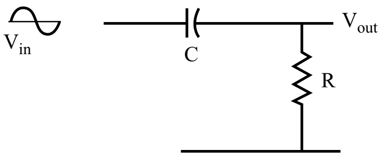

Figure 12-6

Find X_C, Z, V_{out}, I, tan θ, and θ if $V_{in} = 10\ V_{pp}$, f = 1 kHz, C = 0.1 μF, R = 1600 ohms.

42. Again, using the circuit shown in Figure 12-6, find X_C, Z, V_{out}, I, tan θ, and θ if $V_{in} = 120\ V_{rms}$, f = 60 Hz, C = 0.33 μF, R = 6 kΩ.

43. For this question, refer to the circuit shown in Figure 12-7.

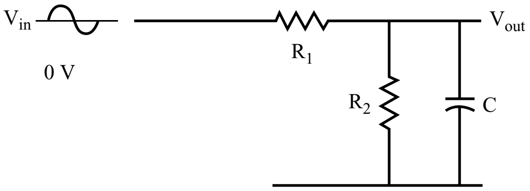

Figure 12-7

Find X_C, AC V_{out}, and DC V_{out} if $V_{in} = 1\ V_{pp}$ AC, riding on a 5 V DC level; f = 10 kHz; R_1 = 10 kΩ; R_2 = 10 kΩ; C = 0.2 μF.

44. Again, using the circuit shown in Figure 12-7, find X_C, AC V_{out}, and DC V_{out} if $V_{in} = 0.5\ V_{pp}$ AC, riding on a 10 V DC level; f = 120 Hz; R_1 = 80 ohms; R_2 = 20 ohms; C = 1000 μF.

45. For this question, refer to the circuit shown in Figure 12-8.

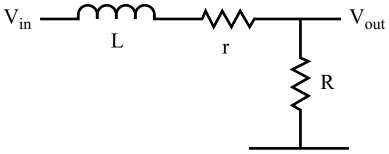

Figure 12-8

In this circuit, $V_{in} = 10\ V_{pp}$ AC, riding on a 5 V DC level; f = 1 kHz; L = 10 mH; r = 9 ohms; R = 54 ohms. Find AC V_{out}, DC V_{out}, X_L, Z, tan θ, and θ.

---

46. In the circuit shown in Figure 12-9, L = 1 mH, C = 0.1 μF, and R = 10 ohms.

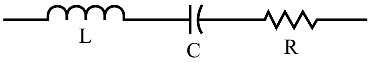

Figure 12-9

Find f_r, X_L, X_C, Z, Q, and the bandwidth.

---

47. In the circuit shown in Figure 12-10, L = 10 mH, C = 0.02 μF, and r = 7 ohms.

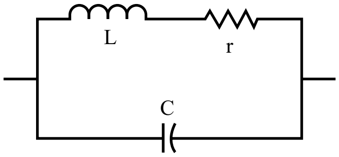

Figure 12-10

Find f_r, X_L, X_C, Z, Q, and the bandwidth.

---

48. If the voltage across the resonant circuit of question 47 is at a peak value of 8 V at the resonant frequency, what is the voltage at the half power points?

What are the half power frequencies?

---

49. Use the amplifier circuit shown in Figure 12-11 to answer this question.

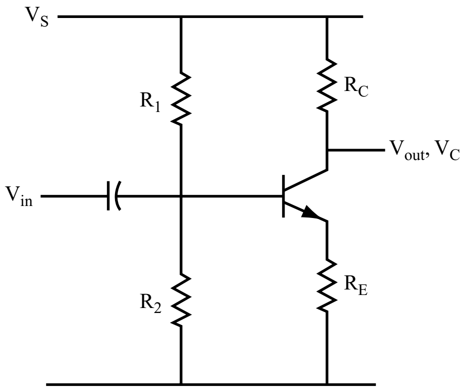

Figure 12-11

Find the values of R_1, R_2, and R_E that will provide the amplifier with a voltage gain of 10. Use $V_S = 28$ V, $R_C = 1$ kΩ, $\beta = 100$.

---

50. Again, using the circuit shown in Figure 12-11, find the values of R_1, R_2, and R_E that will provide the amplifier a voltage gain of 20. Use $V_S =$ 10 V, $R_C = 2.2$ kΩ, $\beta = 50$.

---

51. Using the circuit shown in Figure 12-11, how would you modify the amplifier in question 50 to obtain a maximum gain? Assume that the lowest frequency it has to pass is 50 Hz.

---

52. Using the JFET amplifier circuit shown in problem 42 of Chapter 1, with a bias point of $V_{GS} = -2.8$ V, a drain current of $I_D = 2.7$ mA, and $V_{DS} =$ 12 V, find the values of R_S and R_D.

---

53. If the transconductance of the JFET used in question 52 is 4000 μmhos, what will the AC voltage gain be?

---

54. A certain op-amp circuit uses an input resistance of 8 kΩ to an inverting input. In order for the op-amp circuit to have a gain of 85, what should the value of the feedback resistance be?

---

55. If the input to the op-amp circuit of question 54 is 2 mV, what is the output?

56. What is an oscillator?

57. Why is positive feedback rather than negative feedback necessary in an oscillator?

58. What feedback method is used in a Colpitts oscillator?

59. What feedback method is used in a Hartley oscillator?

60. Draw the circuit of a Colpitts oscillator.

61. Draw the circuit of a Hartley oscillator.

62. What is the formula used to calculate the output frequency of an oscillator?

63. Draw the circuit symbol for a transformer with a center tap.

64. Name the two main coils used on a transformer.

65. What is the equation that shows the relationship between the input voltage, the output voltage, and the number of turns in each coil of a transformer?

66. What is the equation that shows the relationship between the turns ratio and the currents in the primary and secondary coils of the transformer?

____

67. What is the equation that shows the relationship between the impedance of the primary coil, the impedance of the secondary coil, and the number of turns in each coil of a transformer?

____

68. What are the two main uses for transformers?

____

69. Draw a simple half-wave rectifier circuit with a smoothing filter at the output.

____

70. Draw a simple full-wave rectifier circuit using a center tap transformer and a smoothing filter at the output.

____

71. Given a 10 V_{rms} input to a full-wave rectified power supply, calculate the values of R_1, C_1 and C_2 (see Figure 12-12) that results in a 5 V DC output across a 50 ohm load.

____

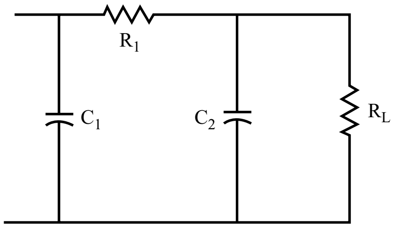

Figure 12-12

Answers to Final Self-Test

The references in parentheses to the right of the answers give you the chapter and problem number where the material is introduced so that you can easily review any concepts covered in the test.

1.	V = 2 V	(Chapter 1, problem 5)
2.	I = 0.5 mA	(Chapter 1, problem 6)
3.	R = 7 ohms	(Chapter 1, problem 7)
4.	132 ohms	(Chapter 1, problem 10)
5.	P = 140 milliwatts or 0.14 watts	(Chapter 1, problems 13 and 14)
6.	0.2 W	(Chapter 1, problems 13 and 15)
7.	22.36 mA	(Chapter 1, problems 13 and 16)
8.	2.86 V, 9.14 V, 12 V	(Chapter 1, problems 23 and 26)
9.	2/3 A through the 6 ohm resistor; 1/3 A through the 12 ohm resistor	(Chapter 1, problem 28 or 29)
10.	91 mA through the 660 ohm resistor; 182 mA through the 330 ohm resistor	(Chapter 1, problem 28 or 29)
11.	$\tau = 0.01$ seconds	(Chapter 1, problem 34)
12.	$\tau = 250$ seconds	(Chapter 1, problem 34)
13.	6 μF	(Chapter 1, problem 40)
14.	52.4 μF	(Chapter 1, problem 41)
15.	8.25 μF	(Chapter 1, problem 41)
16.	Approximately 0.3 V	(Chapter 2, problem 10)
17.	Approximately 0.7 V	(Chapter 2, problem 10)
18.	$I_D = 4.3$ mA	(Chapter 2, problem 14)
19.	$I_D = 120$ mA	(Chapter 2, problem 14)
20.	$I_Z = 3$ mA	(Chapter 2, problem 31)
21.	$I_Z = 3.07$ mA	(Chapter 2, problem 31)
22.	$V_C = 14$ V	(Chapter 3, problems 21–24)
23.	$R_B = 1.1$ MΩ	(Chapter 3, problems 21–24)
24.	$\beta = 50$	(Chapter 3, problems 21–24)
25.	Drain, source, and gate, with the gate acting to control the JFET	(Chapter 3, problem 29)
26.	$R_B = 500$ kΩ	(Chapter 4, problems 8)
27.	$R_B = 470$ kΩ	(Chapter 4, problems 4–8)

28. $R_3 = 44\ \text{k}\Omega$, $R_1 = 2.2\ \text{k}\Omega$, $R_2 = 2.2\ \text{k}\Omega$	(Chapter 4, problems 19–23)
29. $R_3 = 2.2\ \text{k}\Omega$, $R_1 = 66\ \text{k}\Omega$, $R_2 = 66\ \text{k}\Omega$	(Chapter 4, problems 19–23)
30. $R_D = 2\ \text{k}\Omega$	(Chapter 4, problem 39)
31. See Figure 12-13.	(Chapter 5, problem 7)

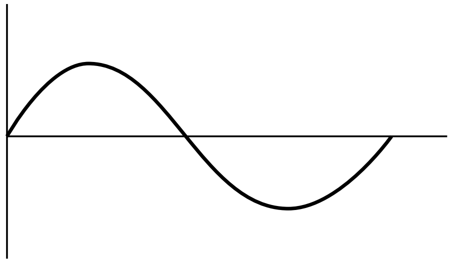

Figure 12-13

32. See Figure 12-14.	(Chapter 5, problems 3 and 7)

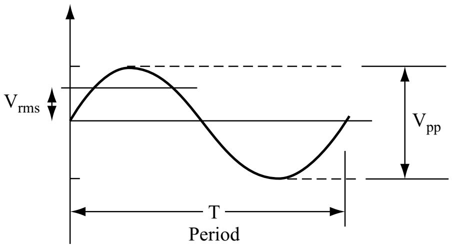

Figure 12-14

33. 3.535 V	(Chapter 5, problem 4)
34. 340 V	(Chapter 5, problem 5)
35. 69 μsec	(Chapter 5, problem 7)
36. 13.3 ohms	(Chapter 5, problem 14)
37. 0.32 μF	(Chapter 5, problem 14)
38. 27.6 ohms	(Chapter 5, problem 17)
39. 16 mH	(Chapter 5, problem 17)
40. 8 kHz	(Chapter 5, problems 19 and 21)
41. $X_C = 1.6\ \text{k}\Omega$, $Z = 2263$ ohms, $V_{out} = 7.07$ V, $I = 4.4$ mA, $\tan\theta = 1$, $\theta = 45$ degrees	(Chapter 6, problems 10 and 23)

42. $X_C = 8\ \text{k}\Omega$, $Z = 10\ \text{k}\Omega$, $V_{out} = 72$ V, I = 12 mA, $\tan\theta = 1.33$, $\theta = 53.13$ degrees	(Chapter 6, problems 10 and 23)
43. $X_C = 80$ ohms, AC $V_{out} = 8$ mV, DC $V_{out} = 2.5$ V	(Chapter 6, problem 26)
44. $X_C = 1.33$ ohms, AC $V_{out} = 8.3$ mV, DC $V_{out} = 2$ V	(Chapter 6, problem 26)
45. $X_L = 62.8$ ohms, Z = 89 ohms, AC $V_{out} = 6.07$ V, DC $V_{out} = 4.3$ V, $\tan\theta = 1$, $\theta = 45$ degrees	(Chapter 6, problems 31 and 35)
46. $f_r = 16$ kHz, $X_L = 100$ ohms, $X_C = 100$ ohms, Z = 10 ohms, Q = 10, BW = 1.6 kΩ	(Chapter 7, problems 2, 6, and 20)
47. $f_r = 11{,}254$ Hz, $X_L = X_C = 707$ ohms, $Z = 71.4\ \text{k}\Omega$, Q = 101, BW = 111 Hz	(Chapter 7, problems 10, 11, and 20)
48. $V_{hp} = 5.656$ V, $f_{1hp} = 11{,}198$ Hz, $f_{2hp} = 11{,}310$ V	(Chapter 7, problem 27)
49. Your values should be close to the following: $R_E = 100$ ohms, $V_C = 14$ V, $V_E = 1.4$ V, $V_B = 2.1$ V, $R_2 = 1.5\ \text{k}\Omega$, $R_1 = 16.8\ \text{k}\Omega$	(Chapter 8, problem 17)
50. $R_E = 110$ ohms, $V_C = 5$ V, $V_E = 0.25$ V, $V_B = 0.95$ V, $R_2 = 2.2\ \text{k}\Omega$, $R_1 = 18.1\ \text{k}\Omega$	(Chapter 8, problem 17)
51. The gain can be increased by using a capacitor to bypass the emitter resistor R_E; $C_E = 300\ \mu\text{F}$ (approximately).	(Chapter 8, problem 20)
52. $R_S = 1.04\ \text{k}\Omega$, $R_D = 3.41\ \text{k}\Omega$	(Chapter 8, problem 42)
53. $A_v = -13.6$	(Chapter 8, problem 39)
54. $R_F = 680\ \text{k}\Omega$	(Chapter 8, problem 45)
55. $V_{out} = 170$ mV and is inverted	(Chapter 8, problem 45)
56. An *oscillator* is a circuit that emits a continuous sine wave output, without requiring an input signal. Other types of oscillators exist that do not have sine wave outputs, but they are not discussed in this book.	(Chapter 9, introduction)
57. Positive feedback causes the amplifier to sustain an oscillation or sine wave at the output. Negative feedback causes the amplifier to stabilize, which reduces oscillations at the output.	(Chapter 9, problems 2 and 3)
58. A capacitive voltage divider	(Chapter 9, problem 14)

59. An inductive voltage divider (Chapter 9, problem 14)

60. See Figure 12-15. (Chapter 9, problem 24)

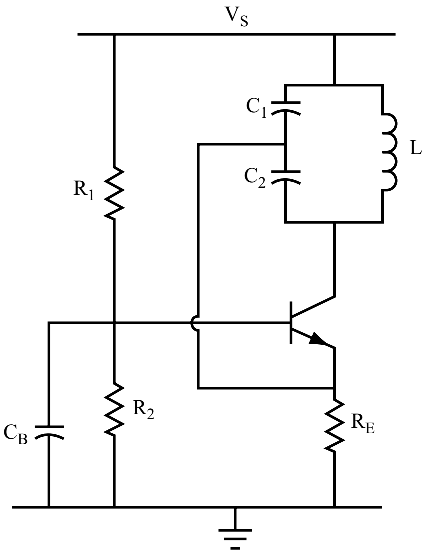
V_S
C_1
C_2
L
R_1
R_2
C_B
R_E

Figure 12-15

61. See Figure 12-16. (Chapter 9, problem 25)

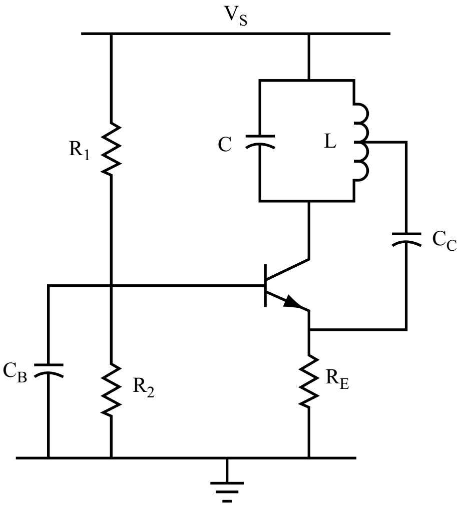
V_S
R_1
C
L
C_C
C_B
R_2
R_E

Figure 12-16

62. $f_r = \dfrac{1}{2\pi\sqrt{LC}}$ (Chapter 9, problem 11)

63. See Figure 12-17. (Chapter 10, problem 4)

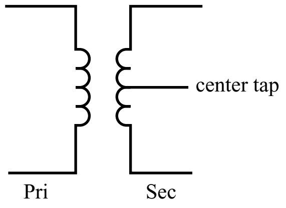

Figure 12-17

64. Primary and secondary (Chapter 10, problem 2)

65. $V_{in}/V_{out} = V_P/V_S = N_P/N_S = TR$ (Chapter 10, problem 6)

66. $I_{out}/I_{in} = I_S/I_P = N_P/N_S = TR$ (Chapter 10, problem 13)

67. $Z_{in}/Z_{out} = (N_P/N_S)^2$, or impedance ratio, is the square of the turns ratio (Chapter 10, problem 16)

68. They are used for stepping up or stepping down an AC voltage, and to match impedances between a generator and a load (Chapter 10, introduction)

69. See Figure 12-18. (Chapter 11, problem 14)

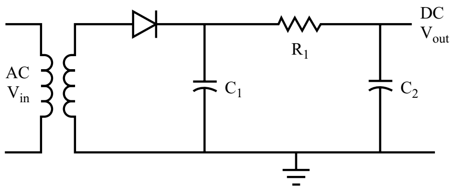

Figure 12-18

70. See Figure 12-19. (Chapter 11, problem 31)

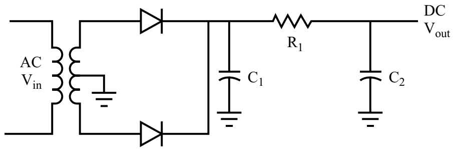

Figure 12-19

71. $R_1 = 84$ ohms, $C_1 = 622\ \mu F$, $C_2 = 265\ \mu F$ (Chapter 11, problems 26–29)

APPENDIX A

Glossary

Ampere (A) The unit of measurement of electric current.

Amplifier Electronic device or circuit that produces an output signal with greater power, voltage, or current than that provided by the input signal.

Capacitance (C) The capability of a component to store an electric charge when voltage is applied across the component, measured in farads.

Capacitor A component that stores electric charge when voltage is applied to it, that can return the charge to a circuit in the form of electric current when the voltage is removed.

Discrete components Individual electronic parts such as resistors, diodes, capacitors, and transistors.

Diode A component that conducts current in one direction only.

Farad (F) The unit of measurement of capacitance.

Feedback A connection from the output of an amplifier back to the input, where a portion of the output voltage is used to control, stabilize, or modify the operation of the amplifier.

Frequency (f) Number of cycles of a waveform that occur in a given time period, measured in hertz (cycles per second).

Ground Zero volts. This is the arbitrary reference point in a circuit from which all voltage measurements are made.

Henry (H) The unit of measurement of inductance.

Impedance (Z) Total opposition (resistance and reactance) of a circuit to AC current flow, measured in ohms.

Inductance (L) The property of a component that opposes any change in an existing current, measured in henrys.

Inductor A coil of wire whose magnetic field opposes changes in current flow when the voltage across the coil is changed.

Integrated circuit (IC) Electronic component in the form of a very small silicon chip in which numerous transistors and other components have been built to form a circuit.

Kirchhoff's laws A set of formulas that form the basis for DC and AC circuit analysis, including: *current law (KCL):* The sum of all currents at a junction equals zero; and *voltage law (KVL):* The sum of all voltages in a loop equals zero.

Ohm (Ω) The unit of measurement of resistance.

Ohm's law A formula used to calculate the relationship between voltage, current, and resistance, expressed as $V = IR$. Also expressed as $E = IR$.

Operational amplifier (op-amp) An integrated circuit, multi-stage amplifier. An op-amp is much smaller and, therefore, more practical than an equivalent amplifier made with discrete components.

Oscillator An electronic circuit that produces a continuous output signal such as a sine wave or square wave.

Phase angle For a signal, the angle of lead or lag between the current waveform and the voltage waveform.

Phase shift The change in phase of a signal as it passes through a circuit, such as in an amplifier.

Power The expenditure of energy over time. Power is measured in watts.

Reactance (X) The degree of opposition of a component to the flow of alternating current (AC), measured in ohms. There are two types of reactance: capacitive reactance (X_C) exhibited by capacitors and inductive reactance (X_L) exhibited by inductors.

Rectification The process of changing alternating current (AC) to direct current (DC).

Resistance (R) The degree of opposition of a component to the flow of electric current, measured in ohms.

Resistor A component whose value is determined by the amount of opposition it has to the flow of electric current.

Semiconductor A material that has electrical characteristics of a conductor or an insulator, depending on how it is treated. Silicon is the semiconductor material most commonly used in electronic components.

Transformer A component that transforms an input AC voltage to either a higher level (step up transformer) or a lower level (step down transformer) AC voltage.

Transistor, BJT A bipolar junction transistor (BJT) is a semiconductor component that can either be used as a switch or an amplifier. In either case, a small input signal controls the transistor, producing a much larger output signal.

Transistor, JFET A junction field effect transistor (JFET), like the bipolar junction transistor, can be used either as a switch or an amplifier.

Transistor, MOSFET Metal oxide silicon field effect transistor (MOSFET); like the BJT and JFET, a MOSFET can be used either as a switch or an amplifier. The MOSFET is the most commonly used transistor in integrated circuits.

Turns ratio (TR) The ratio of the number of turns in the primary or input winding of a transformer to the number of turns in the secondary or output winding.

Volt (V) The unit of measurement for the potential difference that causes a current to flow through a conductor.

Watt (W) Unit of electric power dissipated as heat when 1 amp of current flows through a component that has 1 volt applied across it.

Zener A particular type of diode that will flow current at a definite reverse-bias voltage level.

APPENDIX B

List of Symbols and Abbreviations

The following table lists common symbols and abbreviations.

SYMBOL/ABBREVIATION	MEANING
A	Ampere
AC	Alternating current
A_V	AC voltage gain
β (beta)	Current gain
BW	Bandwidth
C	Capacitor
DC	Direct current
F	Farad
g_m	Transconductance
f	Frequency
f_r	Resonant frequency
H	Henry

SYMBOL/ABBREVIATION	MEANING
Hz	Hertz
I	Electric current
I_B	Base current
I_C	Collector current
I_D	Drain current of a FET; also current through a diode
I_{DSS}	Saturation current
L	Inductor
LC	Inductor-capacitor circuit
N_P	Number of turns in a primary coil
N_S	Number of turns in a secondary coil
Ω	Ohms
P	Power
Q	Transistor, also the Q value of a resonant circuit
R	Resistor
R_{in}	Input resistance of a transistor
r	DC resistance of an inductor
T	Period of a waveform
τ	Time constant
TR	Turns ratio
θ	Phase angle
V	Voltage
V_C	Voltage at the collector of a transistor
V_{DD}	Drain supply voltage
V_E	Voltage at the emitter of a transistor
V_{GG}	Gate supply voltage
V_{GS}	Gate to source voltage
$V_{GS(off)}$	Gate to source cutoff voltage
V_{in}	AC voltage of an input signal
V_{out}	AC output voltage

SYMBOL/ABBREVIATION	MEANING
V_P	Peak voltage
V_{PP}	Peak-to-peak voltage
V_{rms}	Root mean square voltage
V_S	Supply voltage
W	Watts
X_C	Reactance of a capacitor
X_L	Reactance of an inductor
Z	Impedance

APPENDIX C

Powers of Ten and Engineering Prefixes

The following table shows powers of the number 10, decimal equivalents, prefixes used to denote the value, symbols used, and an example.

POWER	DECIMAL	PREFIX	SYMBOL	EXAMPLE
10^9	1,000,000,000	Giga-	G	Gohms
10^6	1,000,000	Mega-	M	Mohms
10^3	1,000	Kilo-	k	kohms
10^{-3}	0.001	Milli-	m	mA
10^{-6}	0.000,001	Micro-	μ	μA
10^{-9}	0.000,000,001	Nano-	n	nsec
10^{-12}	0.000,000,000,001	Pico-	p	pF

APPENDIX

D

Standard Composition Resistor Values

The most commonly-used type of resistor is the carbon composition resistor with a ± 5 percent tolerance and either a 1/4 or 1/2 watt power rating, The standard resistance values for this type of resistor are listed here (in ohms). You should be able to purchase resistors at any of these values through the online distributors listed in Appendix E. Power resistors are available with fewer resistance values, which you can find in various suppliers' catalogs.

NOTE In the following, "k" represents kilo-ohms, so 7.5 k translates into 7,500 ohms. Similarly, "M" stands for megohms, so a value of 3.6 M represents 3,600,000 ohms.

2.2	24	270	3.0 k	33 k	360 k
2.4	27	300	3.3 k	36 k	390 k
2.7	30	330	3.6 k	39 k	430 k
3	33	360	3.9 k	43 k	470 k
3.3	36	390	4.3 k	47 k	510 k
3.6	39	430	4.7 k	51 k	560 k
3.9	43	470	5.1 k	56 k	620 k

4.3	47	510	5.6 k	62 k	680 k
4.7	51	560	6.2 k	68 k	750 k
5.1	56	620	6.8 k	75 k	820 k
5.6	62	680	7.5 k	82 k	910 k
6.2	68	750	8.2 k	91 k	1.0 M
6.8	75	820	9.1 k	100 k	1.2 M
7.5	82	910	10 k	110 k	1.5 M
8.2	91	1.0 k	11 k	120 k	1.8 M
9.1	100	1.1 k	12 k	130 k	2.2 M
10	110	1.2 k	13 k	150 k	2.4 M
11	120	1.3 k	15 k	160 k	2.7 M
12	130	1.5 k	16 k	180 k	3.3 M
13	150	1.6 k	18 k	200 k	3.6 M
15	160	1.8 k	20 k	220 k	3.9 M
16	180	2.0 k	22 k	240 k	4.7 M
18	200	2.2 k	24 k	270 k	5.6 M
20	220	2.4 k	27 k	300 k	6.8 M
22	240	2.7 k	30 k	330 k	8.2 M

APPENDIX E

Supplemental Resources

The following list of books, magazines, tutorials, and electronics suppliers may be of interest to the reader who is seeking more knowledge regarding either basic electronics concepts, reference material for circuit design, or the supplies needed to build circuits.

Web Sites

- *Building Gadgets* (`www.buildinggadgets.com/`) — An electronics reference site maintained by Earl Boysen (one of the authors of this book). There are lots of handy links to electronics tutorials, discussion forums, suppliers, and interesting electronics projects here.
- *All About Circuits* (`www.allaboutcircuits.com/`) — Includes an online book on electronics theory and circuits, as well as discussion forums on electronics projects, microcontrollers, and general electronics issues.
- *Electronics Tutorials* (`www.electronics-tutorials.com/`) — Tutorials on a wide range of electronics topics.
- *Williamson Labs* (`www.williamson-labs.com/`) — Tutorials on electronics components and circuits. Many of these tutorials include

animated illustrations that may help you understand how each circuit functions.

- *Electro Tech online* (`www.electro-tech-online.com/`) — Discussion forum on electronics projects and general electronics issues.
- *Electronics Lab* (`www.electronics-lab.com/forum/`) — Discussion forum on project design, electronics theory, and microcontrollers along with a collection of a few hundred interesting projects.
- *Discover Circuits* (`www.discovercircuits.com/`) — A collection of thousands of electronics circuits.

Books

- *Electronics For Dummies, First Edition, by Gordon McComb and Earl Boysen (Indianapolis: Wiley, 2005)* — A good book to start with; provides an introduction to electronics concepts, components, circuits, and methods.
- *The Art of Electronics, Second Edition, by Paul Horowitz and Winfield Hill. (New York: Cambridge University Press, 1989)* — Useful reference book for designing circuits as well as understanding the functionality of existing circuits.
- *2008 ARRL Handbook for Radio Communications (Newington, Connecticut: American Radio Relay League, 2008)* — While this is intended for ham radio enthusiasts, it is also a useful reference book for understanding circuit design.
- *Electronics Projects For Dummies, First Edition, by Earl Boysen and Nancy C. Muir (Indianapolis: Wiley, 2006)* — Provides an introduction and hands-on practice in building electronic circuits and projects.

Magazines

- *Everyday Practical Electronics Magazine* (`www.epemag.wimborne.co.uk/`) — Information on new components for hobbyists and projects with focus on circuits using discrete components.

- *Nuts and Volts Magazine* (`nutsvolts.com/`) — Information on new components for hobbyists and projects with focus on circuits using microcontrollers.
- *EDN Magazine* (`www.edn.com/`) — Articles on new components/ designs for the engineering community.

Suppliers

This section shows retail stores and online distributors.

Retail Stores

- *Radio Shack* (`www.radioshack.com/`) — Has stores in most U.S. cities that carry electronic components.
- *Fry's Electronics* (`www.frys.com/`) — Has stores in nine states that carry electronics components.

Online Distributors

- *Jameco Electronics* (`www.jameco.com/`) — A medium-sized distributor that carries most of the components you'll need with a reasonably sized catalog that it's easy to find components in.
- *Mouser Electronics* (`www.mouser.com/`) — A large distributor that carries a wide range of components with a nice ordering system on their Web site that lets you put together separate orders for different projects, very handy if you're planning multiple projects. They also do a very good job of packaging, clearly labeling components for shipment.
- *Digi-key* (`www.digikey.com/`) — Another large distributor with a very broad selection of components. Digi-key may carry components that are hard to find at smaller suppliers.

APPENDIX F

Equation Reference

The following table provides a quick reference to common equations.

PARAMETER		EQUATION	CHAPTER REFERENCE
Bandwidth		$BW = \frac{f_r}{Q}$	Chapter 7, problem 20
Capacitance			
	Parallel Capacitance	$C_T = C_1 + C_2 + \cdots + C_N$	Chapter 1, Summary
	Series Capacitance	$\frac{1}{C_T} = \frac{1}{C_1} + \frac{1}{C_2} + \cdots + \frac{1}{C_N}$, or $C_T = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots + \frac{1}{C_N}}$, or $C_T = \frac{C_1 C_2}{C_1 + C_2}$ for two capacitors	Chapter 1, Summary

PARAMETER		EQUATION	CHAPTER REFERENCE
Frequency		$f = \frac{1}{T}$	Chapter 5, problem 7
	Resonance Frequency (series LC circuit)	$f_r = \frac{1}{2\pi\sqrt{LC}}$	Chapter 7, problem 6
	Resonance Frequency (parallel LC circuit)	$f_r = \frac{1}{2\pi\sqrt{LC}}\sqrt{1 - \frac{r^2C}{L}}$ if Q is less than10 $f_r = \frac{1}{2\pi\sqrt{LC}}$, if $Q \geqslant 10$	Chapter 7, problem 10
Gain			
	Voltage Gain	$A_V = \frac{V_{out}}{V_{in}}$	Chapter 8, problem 9
	Current Gain	$\beta = \frac{I_C}{I_B}$	Chapter 3, problem 17
Impedance		$Z = \sqrt{X_C{}^2 + R^2}$	Chapter 6, problem 8
Phase Shift			
	Phase Angle (RC circuit)	$\tan\theta = \frac{X_C}{R} = \frac{1}{2\pi fRC}$	Chapter 6, problem 23
	Phase Angle (LC circuit)	$\tan\theta = \frac{X_L}{R} = \frac{2\pi fL}{R}$	Chapter 6, problem 35
Q Value		$Q = \frac{X_L}{R}$	Chapter 7, problem 20
Resistance			
	Parallel Resistance	$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots + \frac{1}{R_N}$, or $R_T = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots + \frac{1}{R_N}}$, or $R_T = \frac{R_1R_2}{R_1 + R_2}$ for two resistors	Chapter 1, Summary

PARAMETER		EQUATION	CHAPTER REFERENCE
	Series Resistance	$R_T = R_1 + R_2 + \cdots + R_N$	Chapter 1, Summary
Power		$P = VI$, or $P = I^2R$, or $P = \frac{V^2}{R}$	Chapter 1, Summary
Reactance			
	Capacitive Reactance	$X_C = \frac{1}{2\pi fC}$	Chapter 5, problem 13
	Inductive Reactance	$X_L = 2\pi fL$	Chapter 5, problem 16
Time Constant		$\tau = RC$	Chapter 1, Summary
Turns Ratio		$TR = \frac{N_p}{N_s}$	Chapter 11, problem 7
Voltage			
	Ohm's Law (DC)	$V = IR$	Chapter 1, Summary
	Ohm's Law (AC)	$V = IZ$	Chapter 6, problem 8
	Voltage divider	$V_1 = \frac{V_S R_1}{R_T}$	Chapter 1, Summary
	Peak-to-Peak Voltage (sine wave)	$V_{pp} = 2V_p = 2 \times \sqrt{2} \times V_{rms}$ $= 2.828 \times V_{rms}$	Chapter 5, problem 4
	RMS Voltage (sine wave)	$V_{rms} = \frac{1}{\sqrt{2}} \times V_p = \frac{1}{\sqrt{2}} \times \frac{V_{pp}}{2}$ $= 0.707 \times \frac{V_{pp}}{2}$	Chapter 5, problem 4
	Transformer output voltage	$V_{out} = \frac{V_{in} N_s}{N_p}$	Chapter 11, problem 7

APPENDIX

G

Schematic Symbols Used in This Book

The following table shows schematic symbols used in this book.

COMPONENT		SYMBOL
Battery		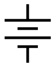
Capacitor		
Diodes		
	Diode	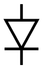
	Zener Diode	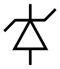
Generator (DC)		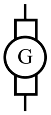

COMPONENT		SYMBOL
Ground		
Inductor		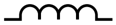
Lamp		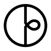
Meters		
	Meter	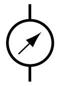
	Ammeter	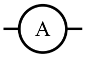
	Voltmeter	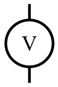
Motor		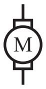
Operational Amplifier		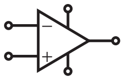
Resistors		
	Resistor	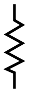
	Variable Resistor (Potentiometer)	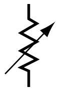
Signal Generator (Sine Wave)		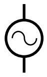

COMPONENT		SYMBOL
Switches		
	Single-pole, Single-throw (SPST) Switch	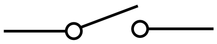
	Single-pole, Double-throw (SPDT) Switch	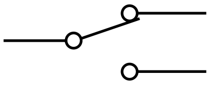
Transformers		
	Transformer	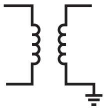
	Center Tap Transformer	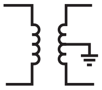
Transistor		
	NPN BJT	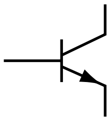
	PNP BJT	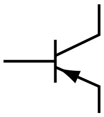
	N-channel JFET	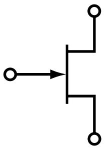
	P-channel JFET	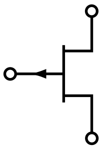

Index

Index

A

AC (alternating current), 22
  circuits
    capacitors, 154–156, 165–167
    inductor, 156–158
    inductors, 190–197
    resistors, 152–154
  diodes in AC circuits, 347–358
  ripple, 359
  signal, capacitor, 187
  signals, 152
    capacitors passing, 166
  sine waves and, 147
AC reactance of an inductor, 156
algebraic formula for Ohm's law, 3
alternative base switching, 131–137
ammeter, 41
amperes
  microampere, 11
  milliampere, 11
amplifiers
  analyzing, 271–275
  common base amplifier, 298
  common collector amplifier, 265
  common emitter amplifier, 296–297
  feedback, 294
  input resistance, 265
  JFET as, 275–284
  lowest frequency passed, 273
  operational amplifier, 284–288
  output resistance, 265
  parallel inductor and capacitor, 299
  resonant load, 299
  stability, 252–256
  tuned load, 299
  two-stage, 264
  voltage gain, 249
anode, 36
  forward biased diode, 38
  voltage, 38
Armstrong method, 304
Armstrong oscillator, 316
attenuation, 181
  f_r, 218

B

ß (beta), 84–85
  finding of transistor, 87–97
band of frequencies, 221
band-reject filter, 222
bandpass filter, 221
bandwidth of circuits, 221
  calculating, 223
base-collector diode, 75
  current flow, 79

base current
  collector current, 84
  flow, 110
  need for, 110
base-emitter diode, 75
  current flow, 80
  voltage drop, 111
base-emitter junction, current flow from supply voltage to, 115
batteries, 2, 3
  voltage, 13
biasing, 243, 256–265
bipolar junction transistors. *See* BJT (bipolar junction transistors)
bipolar transistors, 71
BJT (bipolar junction transistors), 71
  terminals, 97
Boysen, Earl, 378
break down of diode, 57
BW, 221

C

capacitors
  AC, 22
  AC circuits, 154–156
  capacitor charging graph, 24
  DC circuit, 22–28
  DC signal, 187
  emitter bypass capacitor, 261
  and inductor in series, 208–218
  level of opposition, 168
  passing AC signals, 166
  reactance, 155, 166
  resistors and, 186–190
  in series, 167–174
  voltage, 24, 181
cascading transistor amplifiers, 264
cathode, 36
  voltage, 38
center tap, 332
characteristic curve of a diode, 41
chemically generating electric current, 2
circuits
  bandwidth, 221
    calculating, 223
  breaking, 20
  completing, 20
  RLC, 208
  time constant, 23
  transistors, soldering, 72
coils, 236
  number of turns, 333
  oscillators, 317
  reversed, 331
  transformers, 329–330
collector, current flow to base, 116
collector current
  base current and, 84
  calculating, 110
collector-emitter voltage, 89
collector voltage, ON transistor, 109
Colpitts method, 304
Colpitts oscillator circuit, 308–313
  feedback connections, 310
common base amplifier, 298
  oscillators and, 301
common collector amplifier, 265
common emitter amplifier, 296–297
  oscillators and, 301
communications circuits, transformers in, 338–343
conventional current flow, 3
current
  closed switches, 20
  definition, 2
  KCL (Kirchhoff's current law), 19
  open switches, 20
current divider, 17–20
current flow, 1–4, 3
  direction, 81
  magnetic field and, 234
current gain, 84
  formula, 85
currents, small currents, 11–12
cutoff point, 246
cutoff voltage, 140

D
DC circuit, capacitors in, 22–28
DC (direct current), 1
  level, 358–373
  pre-test, 30–33
  pulsating, diodes in AC circuits, 347–358
  pure, 359
  resistance, 156
  signal, capacitor, 187
  smoothing pulsating, 358–373
diode
  in AC circuits, 347–358
  forward biased, 38
  perfect diode, 39
  reverse biased, 39
diode junction, 36
diodes, 35
  backward in circuit, 55
  base-collector diode, 75
  base-emitter diode, 75
  break down, 57
  experiment, 40–45
  germanium, 36
  power dissipated, 51
  silicon, 36
  V-I curve, 41
    perfect diode, 56
  zener diode, 58–65
direction of current flow, 81
doping, 36
dry-labbing, 40

E
efficiency of transformer, 337
electric current
  definition, 2
  generating, 2
    chemically, 2
    electrochemical reaction, 2
    magnetically, 2
    photovoltaic generation of electrons, 2
    piezoelectrical, 3
    thermal generation, 2
electrochemical reaction, 2
electron flow, 3
Electronics Projects For Dummies, 378
emitter bypass capacitor, 261
emitter follower, 265–271
  output impedance, 265
emitter voltage, ON transistor, 109
exponential curve, 25

F
f_3, signal output voltage, 218
f_4, signal output voltage, 218
feedback, 304–308
  Armstrong method, 304
  Colpitts method, 304
  feedback resistor, 287
  Hartley method, 304
  negative, 294–295
  positive, 295–296
  voltage, calculating, 307
formulas, for power, 8–9
forward biased diode, 38
  closed switches and, 40
forward transfer admittance, 280
f_r, attenuation, 218
frequencies
  band of, 221
  blocked by circuit, 219
  range of, 221
frequency response graphs, half power frequencies, 219
fuel cells, 2
full-wave rectifier, 355
function generator, 148

G
gain
  accurate amounts, 262
  op-amp, 287
general resonance curve, 229
generators, 2, 148
  function generator, 148

germanium, diodes, 36
graph of resistance, 12–14

H

half power frequencies, 219
half power points, 219
half-wave rectifier, 352
Hartley method, 304
Hartley oscillator, 314–315
high pass filter experiment, 174–180
high Q circuit, 227
Hill, Winfield, 378
holes, 2, 36
Horowitz, Paul, 378

I

I_B, 80
  calculating, 112–113
  value, 111
I_c, 80
  load current, 108
IC (integrated circuit), 284
  op-amps, 285
I_{DSS} (saturation current), 139
impedance
  calculating, 169
  inductor and capacitor, calculating, 215
  internal impedance, 299
in phase, 331
inductor
  AC circuits, 156–158
  and capacitor in series, 208–218
  sine waves, 191
  voltage divider, 190
input resistance, 249
internal impedance, 299

J

JFET common source amplifier, 278
JFET (junction field effect transistors), 71, 97, 241
  as amplifier, 275–284
  amplifying AC signals, 276
  biasing circuit, 275
  experiment, 138–142
  forward transfer admittance, 280
  gate voltage, 99
  N-channel, 97
    operating, 99
  P-channel, 97, 99
  semiconductor material, 97
  switches, 137
  terminals, 97
  transconductance and, 280
junction field effect transistors. *See* JFET (junction field effect transistors)

K

KCL (Kirchhoff's current law), 19
  KVL and, 19
kilohm, 11
knee region, 44
knee voltage, 45, 46
KVL (Kirchoff's voltage law), 16
  KCL and, 19

L

LC circuit, 301
level DC, 358–373
limiting voltage, 45
line, 338
load, R_C, 108
load current, calculating, 110
load line, 245
load resistance, 249
low power carbon composition resistors, 10
low Q circuit, 228

M

magnet field, current flow and, 234
magnetically generating electric current, 2
mechanical switches, 20
  voltage drop, 40
megohm, 12

metal oxide silicon field effect transistors. *See* MOSFET
microampere, 11
milliampere, 11
MOSFET (metal oxide silicon field effect transistors), 71
Muir, Nancy, 378

N
N-channel JFET, 97
N type region, 36
negative feedback, 294–295
nonlinear regions, 246
notch filter, 222
NPN transistor, 75–76, 82

O
OFF transistor
  characteristics, 96
  resistor, 115
  resistor values, 116
Ohm’s law, 4–6
ON transistor
  collector voltage, 109
  turning on, 108–114
op-amp (operational amplifier), 241, 284–288
  gain, 287
operational amplifier, 284–288
oscillators, 233–236, 293–303
  amplifier, 293
  Armstrong oscillator, 316
  coil, 317
  Colpitts oscillator circuit, 308–313
  components, 293
  design, 316–320
  external input, 308
  feedback connections, 293
  Hartley oscillator, 314–315
  sine wave and, 235
  troubleshooting, 320–325
out of phase, 331
output impedance, 265
output signal, peak-to-peak value, 265
output voltage
  attenuated, 172
  calculationg, 168
  determining, 229
  resonant frequency and, 214
  waveform
    lagging, 182
    leading, 182

P
P type region, 36
peak (p) voltage, 150
peak-to-peak (pp) voltage, 150
perfect diode, 39
  V-I curve, 55
phase
  in, 331
  out of, 331
phase shift, RL circuit, 197–198
phasor diagram, 184
photovoltaic generation of electrons, 2
piezoelectrical generation of electric current, 3
PIV (peak inverse voltage), 57
PN junction, 36
  direction of flow, 36
PNP transistor, 75–76, 80, 82
positive feedback, 295–296
power, 8–10
  formula for, 8–9
  watts and, 8
power ratings, 10
power supply circuits, 347
  (peak-to-peak) voltage,D
    150
primary time measurement, sine waves, 151
PRV (peak reverse voltage), 57

Q
Q, 219
  calculating, 224
  high Q circuit, 227
  low Q circuit, 228

R
R_1, biasing and, 257
radio bandwidth, 222
range of frequencies, 221
R_B, 82
  calculating, 111–112, 113
  current values, 108
  turning transistor ON, 108
  voltage drop, 111
R_C, 84
  turning transistor ON, 108
RC circuit, phase shift, 180–186
R_E, biasing and, 257
reactance, 154
  capacitors, 166
rectification, 352
  full-wave rectifier, 355
  half-wave rectifier, 352
rectifier, 35
resistance, 154
  calculating total, 169
  graph of resistance, 12–14
resistors
  AC circuits, 152–154
  capacitors and, 186–190
  low power carbon composition, 10
  OFF transistor, 115
  in parallel, 7–8
  power ratings, 10
  in series, 6–7
  in series, 167–174
  values, OFF transistor, 116
resonance, 158–160
resonant circuits
  capacitor and inductor in series, 208–218
  half power freqencies, 219
  output curve, 218–232
resonant frequency, net reactance of inductor and capacitor and, 214
reverse biased diode, 39
  open switches and, 40
R_F (feedback resistor), 287
R_{IN} (input resistance), 249
RL circuit, phase shift, 197–198
R_L (load resistance), 249
RLC circuit, 208
  impedance, calculating, 209
  reactance, calculating, 209
rms (root mean square) voltage, 150
root mean square (rms) voltage, 150

S
saturation, 94, 140
saturation current (I_{DSS}), 139
saturation point, 246
semiconductor, 35
series
  capacitors in, 167–174
  resistors in, 167–174
silicon
  diodes, 36
  transistor, 111
sine wave, 147
  cycle, 148, 151
  duration, 151
  inductor, 191
  as input, 167
  oscillation and, 235
  as output, 167
  p voltage, 150
  pp voltage, 150
  primary time measurement, 151
  producing, 148
  rms voltage, 150
  signals, primary coils, 330
  zero axis, 149
single pole double throw switch, 20
single pole single throw switch, 20
small currents, 11–12
smoothing pulsating DC, 358–373
solar cells, 2
switches, 20–22
  alternative base switching, 131–137
  closed, 20
  controlling, 136

  JFET, 137
  mechanical, 20
  open, 20
    current flow, 114
    terminal voltage, 114
  remote operation, 136
  single pole double throw, 20
  single pole single throw, 20
  speed, 136
  transistors as, 117–126
    three-transistor switch, 127–131

T

television bandwidth, 222
The Art of Electronics, 378
thermal generation of electrical current, 2
three-transistor switch, 127–131
threshold voltage, 45
time constant of circuit, 23
total capacitance, 28
transconductance, 280
transformers
  alternating currents, 330
  coils, 329–330
  in communications circuits, 338–343
  efficiency, 337
  output voltage, 333
transistor, 71
  base-collector diode, 75
  base-emitter diode, 75
  cases, 72
  current gain, 85
  experiment, 87–97
  lead designations, 72
  leads, 73
  no collector current, 95
  NPN, 75–76
  OFF
    characteristics, 96
    resistor, 115
    resistor values, 116
  P region thickness, 74
  packages, 72
  PNP, 75–76
  silicon, 111
  soldering to circuit, 72, 73
  switching, alternative base switching, 131–137
  turning OFF, 114–117
  turning ON, 108–114
  used as switches, 117–126
    three-transistor switch, 127–131
  voltage, between collector and emitter, 246
transistor action, 79
transistor amplifiers, 242–250
  cascading, 264
  emitter follower, 265–271
  experiment, 251–252
  gain, 249
  stability, 252–256
two-stage amplifier, 264

V

V-I curve, 41
  perfect diode, 56
V_{ce}, 89
V_{CE}, 246
voltage
  capacitors, 24, 181
  cutoff, 140
  knee voltage, 45, 46
  KVL (Kirchhoff's voltage law), 16
  limiting voltage, 45
  PIV (peak inverse voltage), 57
  PRV (peak reverse voltage), 57
  threshold voltage, 45
  voltage divider, 14–17
    capacitor in series, 168
    inductors, 190
    output voltage calculation, 168
    resistor in series, 168
  Zener effect, 58

voltage gain, 249
voltmeter, 41
V_{out}, 170
  plotting on graph, 214
  resonance frequency, 217
V_Z, 219

W
watts, 8
waveforms, oscillators, 233

Z
zener diode, 58–65
Zener effect, 58

United States Patent [19]

Gundlach

[11] **Patent Number: 4,595,852**

[45] **Date of Patent: Jun. 17, 1986**

[54] **ELECTROSTATIC GENERATOR**

[76] Inventor: **Robert W. Gundlach,** 2434 Turk Hill Rd., Victor, N.Y. 14564

[21] Appl. No.: **599,199**

[22] Filed: **Apr. 11, 1984**

[51] **Int. Cl.^{4}** **H02N 1/00**

[52] **U.S. Cl.** **310/309**

[58] **Field of Search** 310/308-310

[56] **References Cited**

U.S. PATENT DOCUMENTS

2,610,994	9/1952	Bosch et al.	310/309
2,644,903	7/1953	Lehrer	310/307
2,739,248	3/1956	Meier	310/309
2,756,352	7/1956	Felici et al.	310/308 X

Primary Examiner—Donovan F. Duggan

[57] **ABSTRACT**

A novel induction-type electrostatic generator is disclosed in which moving electrodes are transported back and forth between two Faraday cages. Each moving electrode, once enclosed by a cage, is contacted to the cage, then separated by more than the sparking distance before making electrical connection to the opposite moving electrode or to ground. In this way, charge opposite in polarity but proportional in magnitude to the charge on each cage, is induced in each moving electrode and carried into and imparted onto the next cage, progressively elevating the cage potentials until the corona or arcing limits are reached. Initiation of charging is spontaneous, deriving from inevitable statistical imbalances in nature.

5 Claims, 4 Drawing Figures

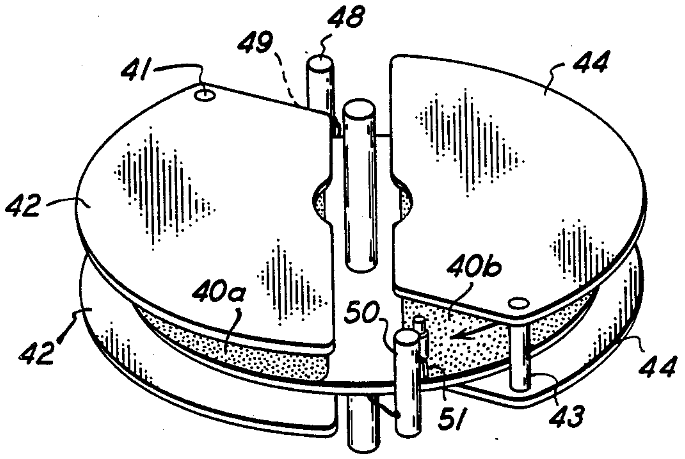

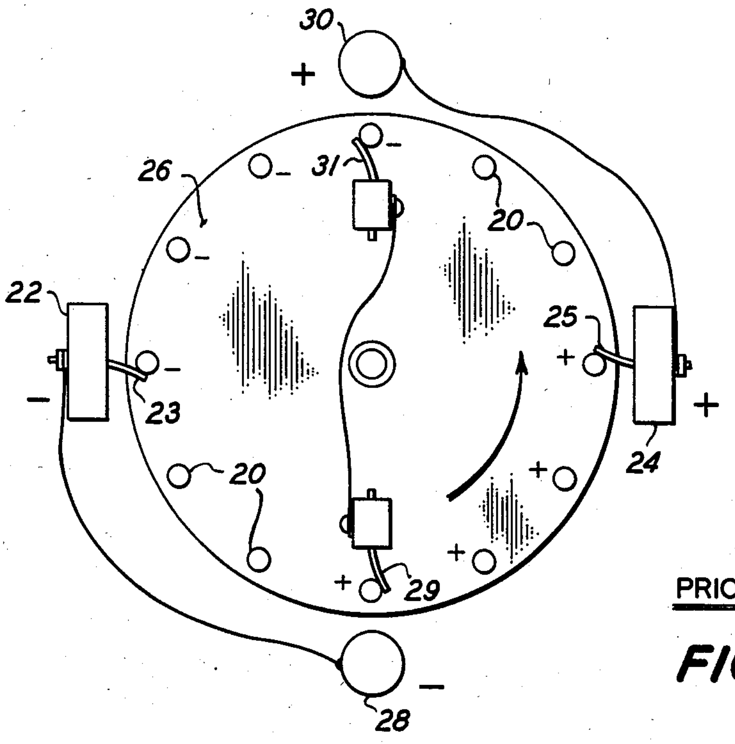
30
+
31
26
20
22
25
23
20
24
29
28
+
-

PRIOR ART

FIG. 1

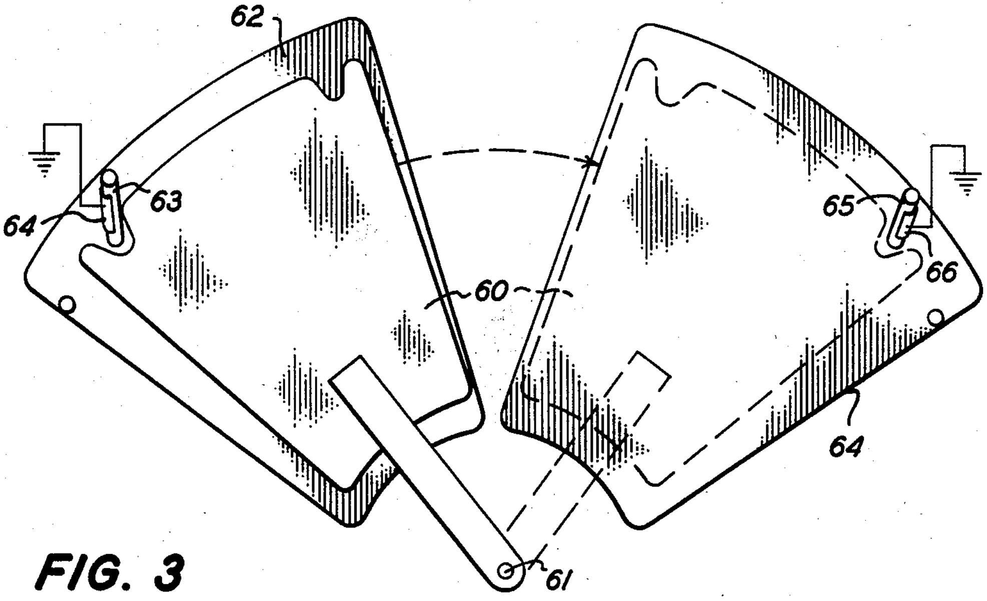
62
63
64
60
65
66
64
61

FIG. 3

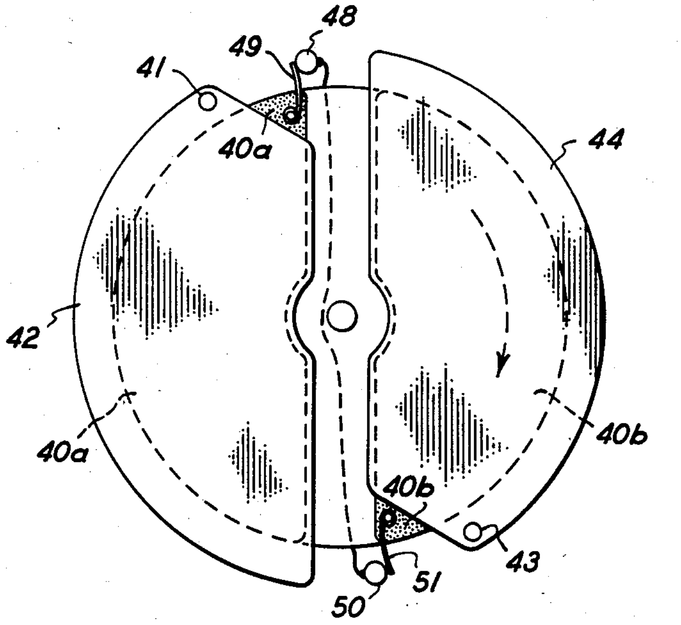

FIG. 2a

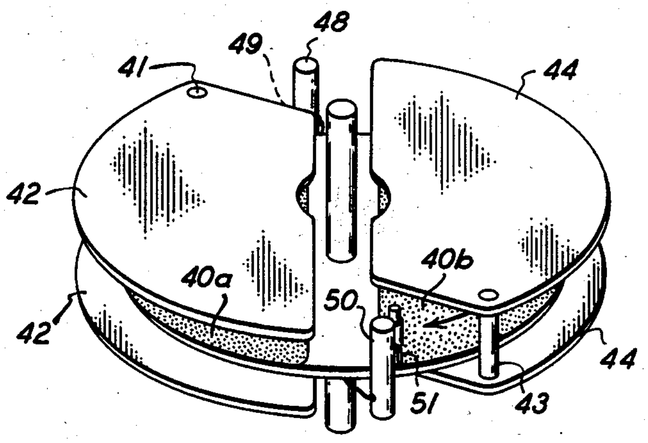

FIG. 2b

ELECTROSTATIC GENERATOR

BACKGROUND OF THE INVENTION

This invention relates to electrostatic generators, and more particularly to an improved induction-type electrostatic generator that is efficient, compact, and inexpensive, in which electrodes are transported back and forth between two Faraday cages.

The Greeks were the first to record observations of static electricity several thousand years ago. They discovered that amber, when rubbed vigorously with fur or fabrics under dry conditions, takes on electrostatic charge. In fact, the Greek word for amber, "electron", has become the root of many of our words relating to electricity.

It was not until about 1660 that the first electrostatic generator was invented. Otto von Guericke of Magdeburg, Germany showed electrostatic sparks could be produced by spinning and rubbing a dry sulfur globe. In this device, high potentials were developed by friction, or triboelectric charging. More than 200 years passed before James Wimshurst invented the first reliable induction-type generator, called the "influence machine" at the time (1878). This involved two insulating discs sliding against each other as they rotated in opposite directions on a common axis. Attached to these oppositely rotating plates were metal sectors into which charges are induced and carried to a capacitor for storage. The metal sectors continue inducing opposite charges in each other and carrying the charges to the appropriate ends of the capacitor, gradually building up the potentials of the capacitor until sparking voltages are reached. Wimshurst built many such machines, some with oppositely rotating discs more than two meters in diameter. The only substantial application of the Wimshurst generator appears to have been to operate the early X-ray tubes.

Then, in 1937, Robert Van de Graaff invented a belt-type generator in which an insulating belt, sometimes carrying conducting elements, transported charges into a giant conducting sphere, or "Faraday cage", to generate, in some cases, over a million volts. Van de Graaff charged the moving belt by spraying corona charge onto it or, alternatively, by frictionally charging it.

In the 1960's Prof. A. D. Moore designed an induction-type generator for instructional purposes. This is the simplest of the prior art generators to understand, and works purely by induction, requiring neither corona nor frictional charging means to produce the potential differences. FIG. **1** is a schematic diagram of Moore's "Dirod" generator, which he described in his book "Electrostatics", (Doubleday, Garden City, NJ, 1968; a Doubleday Anchor Book). In this generator, rods $\mathbf{20}_{(1 \rightarrow N)}$ are rotated between two collector electrodes, **22** and **24**. Each rod makes momentary electrical contact to collectors **22** and **24** through flexible conducting brushes **23** and **25**, respectively, as insulating disk **26** is rotated counterclockwise about its center. Just above and below rotating disc **26** are supported conducting rods **28** and **30**, called "inductors", electrically connected to collector electrodes **22** and **24**, respectively. At the point of nearest approach to each inductor, each of the rods $\mathbf{20}_n$ is momentarily contacted, one at a time, to ground or to the opposite rod $\mathbf{20}_{(n+N/2)}$ through flexible conducting brushes **29** and**31**, respectively. Any incipient charge or potential difference between collector electrodes **22** and **24** will induce opposite charges in those rods $\mathbf{20}_n$ and $\mathbf{20}_{(n+N/2)}$ adjacent the inductors **28** and **30**. For example, if we assume that, of the many billions of electrons in each of the collector electrodes, there is a slight excess in collector **22**, then collector **24** will be slightly positive relative to collector **22**. Inductor **28** will therefore induce positive charges in each rod $\mathbf{20}_n$ as it is grounded in its proximity to inductor **28**. At the same time inductor **30** induces negative charges in each rod $\mathbf{20}_{(n+N/2)}$ when grounded next to inductor **30**. As the charged rods separate from inductor rods **30** and **28**, their potentials increase so that part of the induced charges flow off to the collector plates **22** and **24**, increasing their negative and positive potentials, respectively. As the potentials of the collector electrodes increase, so the charges induced in rods $\mathbf{20}_n$ by inductors **28** and **30** increase, quickly driving the collector electrodes to their maximum values. Once charging begins it will be seen how the potential difference grows at a progressively increasing rate. The statistical imbalance of nature insures that the initial imbalance can always be depended on. (To understand how the instability builds on itself, we need only understand why a sharpened pencil cannot be balanced long on its point; the restoring force is negative, and is proportional to the degree of imbalance.) Professor Moore's Dirod generator routinely produces 60 to 70 kV, depending on the spacing of the rods **20** from inductors **28** and **30**, since air breakdown (arcing) occurs at about 30 kV per cm for gaps larger than a few millimeters. This generator is not entirely satisfactory for manufacturing as an educational toy or for use for demonstrations because it is relatively complicated and expensive to produce and complex to understand.

SUMMARY OF THE INVENTION

Accordingly, the present invention provides a simplified electrostatic generator for an educational toy. Also, the present invention provides an electrostatic generator that is more efficient, more affordable, more rugged, and easier to understand for teaching purposes and experimental uses than prior art devices.

The foregoing and other features of the instant invention will be more apparent from a further reading of the specification and claims and from the drawings, in which:

FIG. **1** is a schematic plan view of a prior art generator.

FIG. *2a* is a schematic plan view of the preferred embodiment of the present invention.

FIG. *2b* is an isometric view of the preferred embodiment of the present invention.

FIG. **3** shows an alternative embodiment of the present invention that employs a single mobile electrode.

While the invention will be described hereinafter in connection with a preferred embodiment and an alternative embodiment, it will be understood that it is not intended to limit the invention to those embodiments. On the contrary, it is intended to cover all alternatives, modifications, and equivalents as may be included within the spirit and scope of the invention as defined by the appended claims.

DESCRIPTION OF THE PREFERRED EMBODIMENT

In the pesent invention the induction-type generator has been simplified and made much more efficient by combining each pair of collector and inductor electrodes into a single Faraday cage, as shown in FIGS. 2*a* and 2*b*. As Faraday taught, and Van de Graaff put to use, by enclosing a charged object in a conducting cage, the potential of the whole cage is increased as if all the charges were placed onto the cage, itself. In fact, if the enclosed charged object is conducting and makes conductive contact with the cage, all charges do, indeed, flow from the charged object onto the surrounding "Faraday cage".

In this example only two rotor electrodes **40***a* and **40***b* are shown. In fact, any even number can be used, within practical space limitations. An odd number could be used if external ground is provided during charge induction. With an even number of rotor electrodes, contacts **48** and **50** can be connected to each other, instead of to external ground, to induce electrons from one rotor electrode to the other. Rotor electrode **40***a* is transported by clockwise rotation from one cage **42**, to the other cage **44**, making sequential contact first with the surrounding, or enclosing case, through contact post **43**, then, after adequate separation from contact post **43**, making momentary contact with ground or the opposite rotor electrode **40***b* through contact posts **50** and **48**, whereby charges opposite to those on the enclosing cage are induced on each rotor plate and then transported to the opposite cage through brushes **49** and **51**. This results in a progressive build-up of opposite electrostatic potentials on collector cages **42** and **44**. Assuming thin, flat rotor plates **40**, the capacitance of each rotor plate **40** to the shell or cage **42** when the plate is totally enclosed by the cage is given by:

$$C = \frac{K\epsilon_0 A}{t}$$

where C is capacitance, K is the dielectric constant of the medium (K=1 for air), ϵ_o is the permittivity of space ($\epsilon_o = 9\times10^{-12}$), A is the area of the surface of the rotor plate, and t is the distance between the rotor plate **40** and the inner surface of the shell **42**. Each plate carries a charge q out of one shell and into the other shell every half rotation, equal to:

$$q = CV.$$

It is possible, then, to calculate the current that can be supplied by this generator for any specified values of rotor plate area A, distance t between the rotor plate **40** and the parallel surface of the cage **42**, and speed of rotation of the rotor plates.

$$i = \frac{dq}{dt} = 4\omega CV$$

where ω is the spin velocity in rotations per second. For example, an area of 100 cm^2 on each face of the rotor plate ($A = 2\times10^{-2}m^2$), and a plate-to-cage spacing of 1 cm ($t = 10^{-2}m$), and a spin velocity of ten rotations per second, gives a generator current of:

$$i = 40\ CV = 40 \times \left(\frac{9\times10^{-12}\times2\times10^{-2}}{10^{-2}}\right) V = 7.2\times10^{-10}\ V \text{ amperes}$$

Clearly, then, current is very low for very low voltages, but builds exponentially with time, the initial voltage build-up can be expedited by placing a charged object on one of the cages. Once the maximum potential of, say, 20 kV is reached, the generator of this example of the present invention would produce a current of:

$$i = 1.4\times10^{-5} \text{ amperes, or about } 14\ \mu A$$

FIG. **3** shows another embodiment of the present invention. Here a single mobile electrode **60** is mounted on a hinge **61** and swings reciprocally between Faraday cages **62** and **64**. In this figure, the top plate of each cage **62** and **64** has been removed for clarity. Movable electrode **60** enters either cage **62** or **64**, makes conducting contact with said cage, separates from said conducting contact by more than the sparking distance (1 cm for every 30 kV), and then makes the electrical contact to ground while it is still enclosed by said cage. During grounding, charges opposite to the charge on the enclosing cage are induced in electrode **60**, and these are transported by movable electrode **60** into and deposited onto the opposite Faraday cage. Swinging contacts **63** and **65** prevent electrical contact between movable plate **60** and ground during entry of plate **60** to each cage, while they insure contact with conducting edges **64** and **66** during exit, while plate **60** is still fully enclosed within the associated Faraday cage. Since the charges induced on movable plate **60** are proportional to the charges on the enclosing cage, repeating the cyclic process produces progressively increasing potentials until corona or sparking limits are reached.

In summary, the advantages of the electrostatic generator of the present invention are its elegant simplicity for construction, for providing a simple understanding for the student, its ruggedness of design for portability without damage, affordability of cost, and its high efficiency derived by maximizing the rotor area and thus capacitance, for a given spacing, the spacing determining the inherent limitation of voltage imposed by dielectric breakdown of air or other fluid medium in which the generator is operated. If we wish to develop potentials of 30 kV operating in air, for example, the spacing of rotors **40** to cages **42** must be at least one cm, since the dielectric breakdown of air for such gaps is 30 kV/cm.

While the principles described herein have been described by the example of FIGS. 2*a*, 2*b*, and **3**, it should be understood that the concepts disclosed can be embodied in various mechanical designs without departing from the teachings of this invention and may come within the scope of the following claims.

What is claimed is:

1. A spontaneous non-frictional electrostatic induction-type generator, comprising:

(a) two Faraday cages:

(b) electrode means, substantially surrounded by and mounted for rotary movement relative to said Faraday cages in order to enable said Faraday cages to serve alternately as both the inductor and collector electrodes: and

(c) flexible contact means comprising a single set of brushes mounted to said rotary electrode means so as to extend by centrifugal force into electrical contact with said Faraday cages, then, after separating from each cage by more than the sparking distance, momentarily contacting each said flexible contact means to ground or to opposite ends of a conductor connecting said flexible contact means to each other to progressively charge said Faraday cages.

2. The spontaneous non-frictional electrostatic induction-type generator of claim **1**, wherein said electrode means comprises an even number of movable electrodes, and flexible contact means mounted for centrifugal extension to make momentary electrical contact separately to each cage; then, after separating from each cage by more than the sparking distance, momentarily connecting said movable elctrodes to each other.

3. The method of claim **1**, wherein an even number of rotary movable electrodes is provided, and means, mounted for extension by centrifugal force, is provided for contacting said rotary movable electrodes separately to each Faraday cage, then, after separating from each cage by more than the sparking distance, momentarily connecting said rotary movable electrodes to each other.

4. The spontaneous electrostatic induction-type generator of claim **1**, wherein said electrode means comprises a plurality of movable electrodes.

5. A method of generating high electrostatic potentials using a spontaneous, non-frictional electrostatic induction-type generator comprising the steps of:

(a) providing two Faraday cages;

(b) providing electrode means substantially surrounded by and mounted for rotary movement relative to said Faraday cages; and

(c) providing flexible contact means mounted to said rotary electrode means so as to extend by centrifugal force into electrical contact with each cage separately, then, after separating from each cage by more than the sparking distance, momentarily contacting to ground or to a neutral conductor, in order to progressively charge said Faraday cages.

* * * * *

US 20030209635A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2003/0209635 A1**

St. Clair (43) **Pub. Date: Nov. 13, 2003**

(54) **ELECTRIC DIPOLE MOMENT PROPULSION SYSTEM**

(76) Inventor: **John Quincy St. Clair**, San Juan, PR (US)

Correspondence Address:
John St. Clair
Hyperspace Research Institute
52 Kings Court, 4A
San Juan, PR 00911 (US)

(21) Appl. No.: **10/142,239**

(22) Filed: **May 9, 2002**

Publication Classification

(51) **Int. Cl.^{7}** **B64G 1/24**; F02K 1/00; F03H 3/00; F02K 5/00; F02K 7/00; F02K 11/00

(52) **U.S. Cl.** **244/164**

(57) **ABSTRACT**

This invention relates to a spacecraft propulsion system utilizing a rotating octagon of trapezoidal electrically charged flat panels to create an electric dipole moment that generates lift on the hull. On the interior side of each panel are electrostatically charged rods which produce a planar electric field that emerges from holes in the panel to form an ellipsoidal potential energy bubble on the outside of the hull. The rotating hull dipole moment generates a magnetic moment which, together with the magnetic field gradient developed by the rotating electric field of the electrostatically charged panels, produces said lift force. The potential energy field is enhanced by using a double cladding of hull material with different ranges of permittivities.

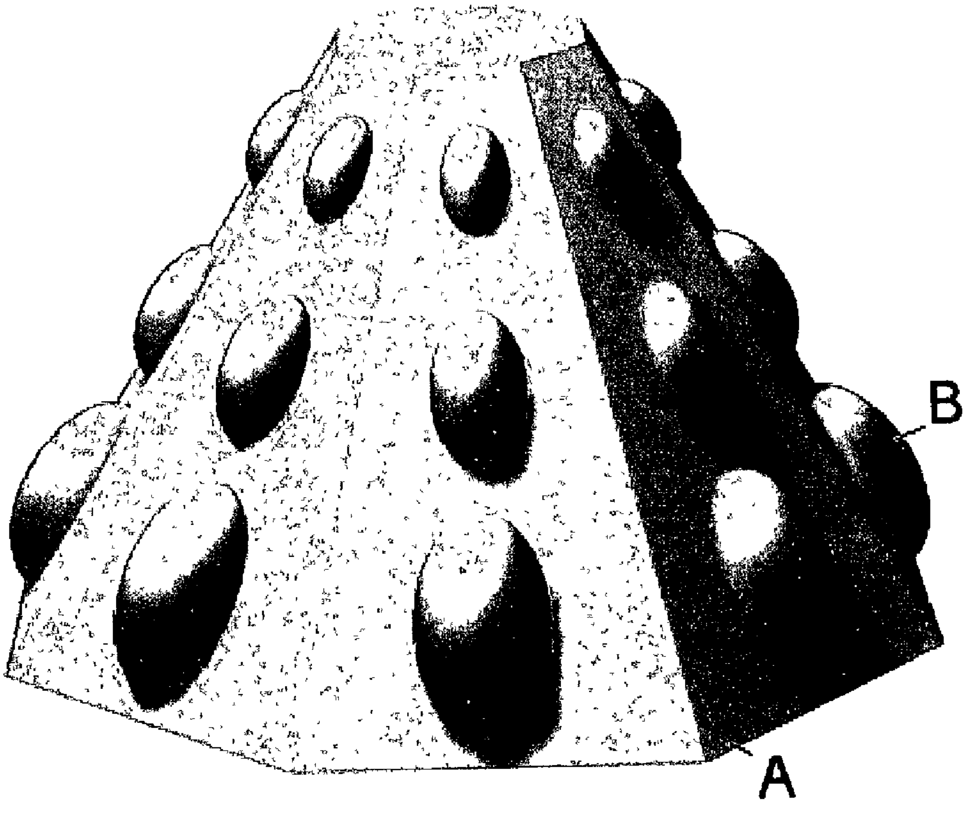

Figure 1

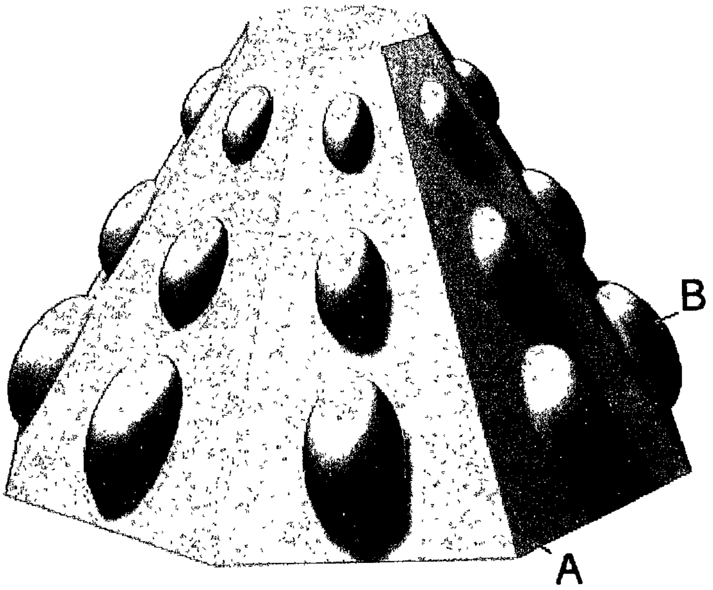

Figure 2

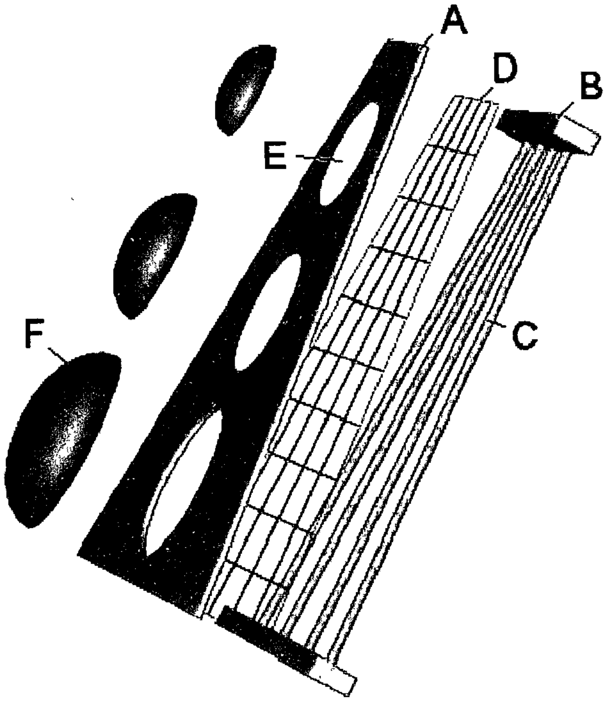

Figure 3

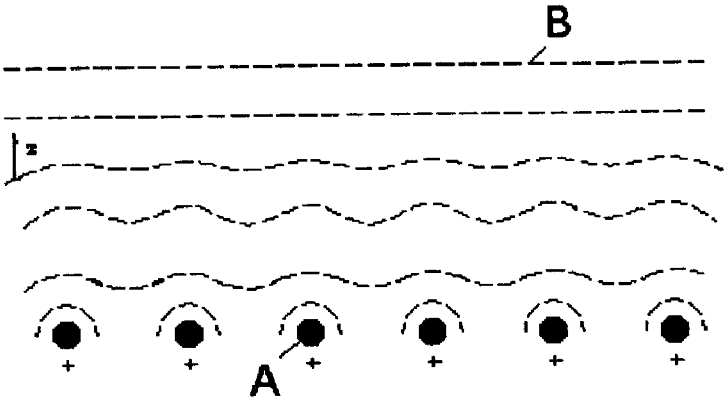

Figure 4

$$\phi = -E_0 z$$

Figure 5

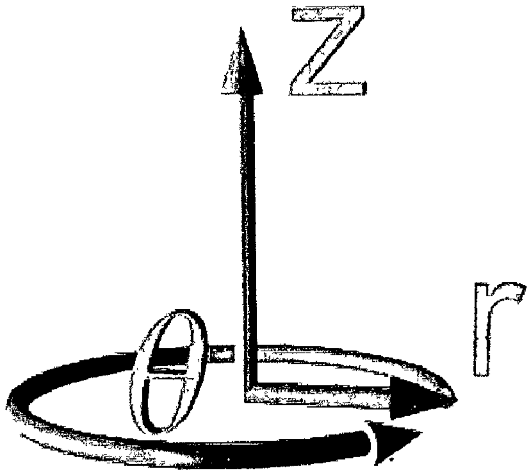

Figure 6

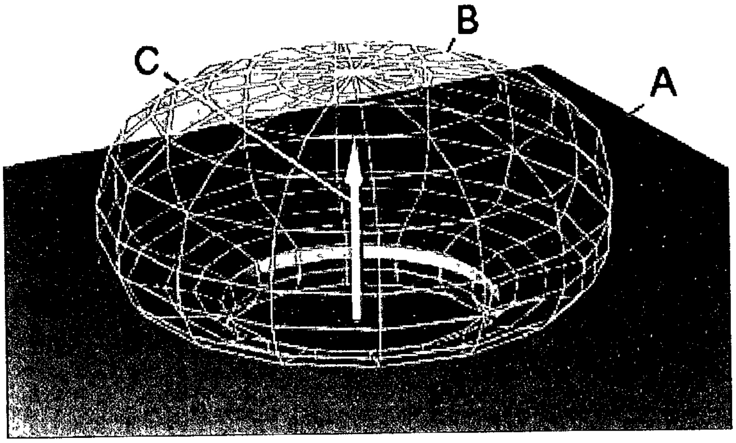

Figure 7

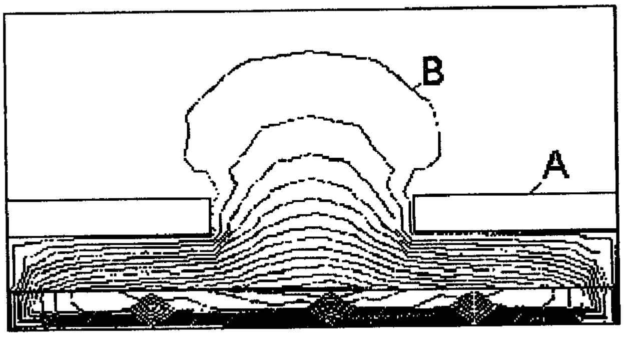

Figure 8

$$\bar{y}_{\text{ellipse}} = \frac{4a}{3\pi}$$

Figure 9

$$q = \varepsilon E \pi a^2$$

Figure 10

$$\text{coul} = \frac{\text{coul}^2}{m^2 n} \frac{n}{\text{coul}} m^2$$

Figure 11

$$p = q\bar{y} = \varepsilon E \pi a^2 \frac{4}{3} \frac{a}{\pi} = \frac{4}{3} \varepsilon E a^3$$

Figure 12

$$\mu = p v = \text{coul}\, m \frac{m}{\text{sec}} = \frac{\text{coul}}{\text{sec}} m^2 = \text{amp}\, m^2$$

Figure 13

$$\mu^z = p_r x^r \omega^z$$

Figure 14

$$F = \nabla(\mu \cdot B)$$

Figure 15

$$(\nabla \times B)_r = \frac{1}{c^2} \frac{\partial E_r}{\partial t}$$

Figure 16

$$\frac{\partial B_z}{r\partial\theta} - \frac{\partial B_\theta}{\partial z} = \frac{\partial E_r}{\partial t}\frac{1}{c^2} = \frac{\partial B_z}{r\partial\theta}$$

Figure 17

$$E_r = E_0 e^{i\omega t}$$

Figure 18

$$B_z = \frac{v}{c^2} E_r$$

Figure 19

$$\frac{dB_z}{dz} = \frac{E_r}{c^2}\frac{dv}{dz}$$

Figure 20

$$\frac{dB}{dz} = \frac{E_r}{c^2}\frac{dv}{dz}$$

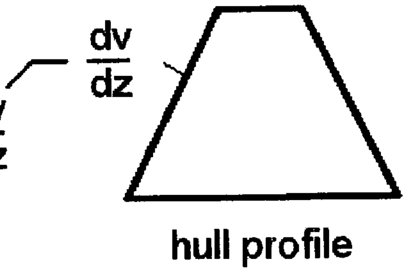

Figure 21

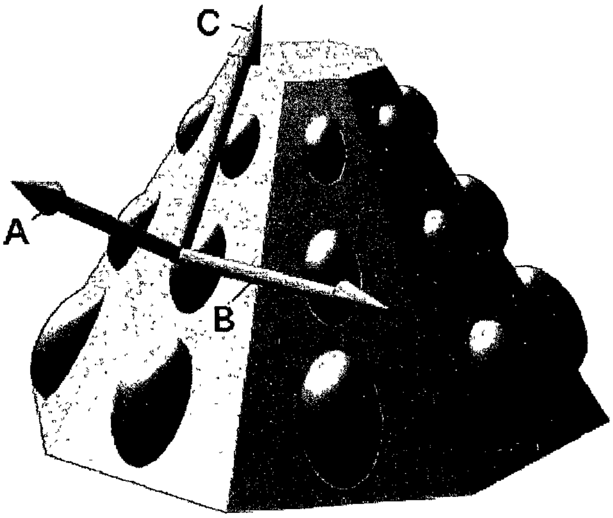

Figure 22

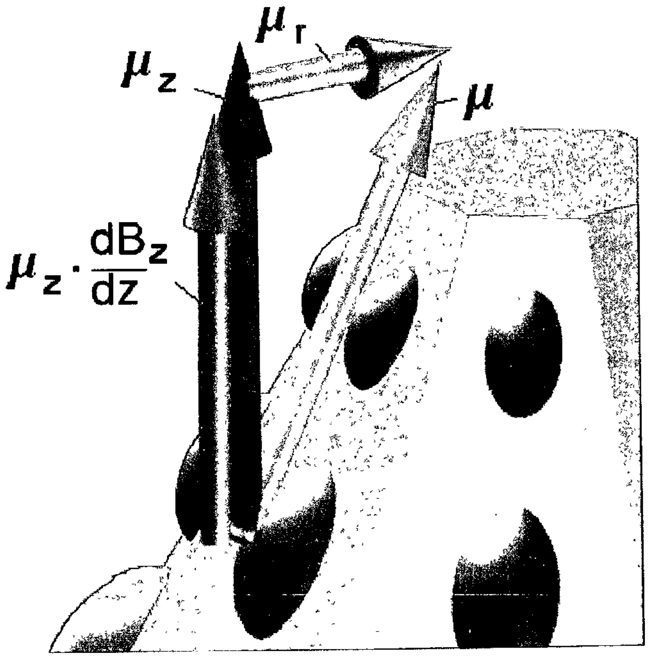
μ_r
μ_z
μ
$\mu_z \cdot \frac{dB_z}{dz}$

Figure 23

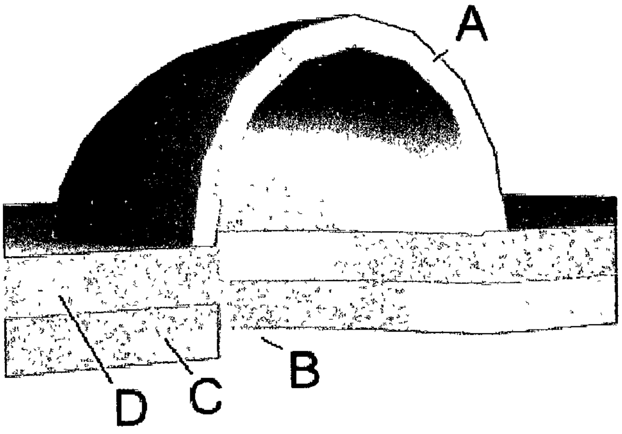

Figure 24

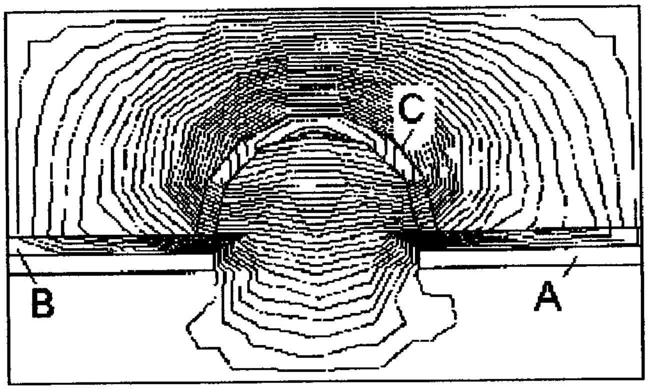

ELECTRIC DIPOLE MOMENT PROPULSION SYSTEM

BRIEF SUMMARY OF THE INVENTION

[**0001**] The invention, which is the object of my present application, is a spacecraft utilizing trapezoidal electrostatically charged flat plate panels which form a pyramidal hull. A panel contains three holes each of which produces a potential energy ellipsoidal bubble that creates an electric dipole moment. The rotation of the hull generates a magnetic moment and a magnetic field gradient in the vertical direction that produces a lift force on the spacecraft.

REFERENCE PAPERS

[**0002**] *Gravitation,* John Archibald Wheeler

[**0003**] Levitron, Hones, U.S. Pat. No. 5,404,062.

[**0004**] *Classical Electrodynamics,* John David Jackson, Chapter 3.12, *Mixed Boundary Conditions; Conducting Plan with a Circular Hole.*

BACKGROUND OF THE INVENTION

[**0005**] I was reading Dr. Jackson's book about the formation of a potential energy bubble by a hole in a conducting plane, a picture of which is shown on page 134 of his book. It turns out that this bubble creates an electric dipole moment from which it is possible to get a magnetic moment. Then I realized that a rotating tilted hull having a velocity gradient would produce a magnetic field gradient in the vertical direction. This combination produces a lift force on the spacecraft.

[**0006**] I did quite a number of computer simulations of the bubble using an electromagnetic simulation software program. I found that the hole in the plate doesn't produce a very large potential energy bubble. So I created another simulation where the hole protrudes out of the plate in an ellipsoidal shape. This produces a much larger field. I then found that using a double cladding, in which each layer has a different permittivity, confined the field to the outside of the hull for even better results.

[**0007**] The planar potential energy is created by a grid of electrically charged wires or rods running the length of each panel. As shown by Feynman, the circular potential energy from each rod very quickly sums to form a flat sheet of energy which emerges from the hole to form said potential energy bubble.

SUMMARY OF THE INVENTION

[**0008**] The invention relates to a spacecraft utilizing a rotating octagon of trapezoidal electrically charged flat plate panels to form a hull in the shape of a pyramid. Each panel has three protruding ellipsoidal bubbles that produce an electric dipole moment from a planar potential energy field created by a group of charged rods parallel to the panel. Because the panels are tilted and the hull is rotating, there is a tangential velocity gradient in the vertical direction. This creates the magnetic moment. Because the hull rotates, the radial electric field produces a magnetic field gradient in the vertical direction. This combination of magnetic moment and magnetic field gradient produces a lift force on the hull of the spacecraft.

[**0009**] On the underside of each panel is a group of high voltage electrically charged rods which run parallel to the panel. These wires or rods produce a planar electrical potential field underneath the holes in the panel. This potential energy field then bubbles out of the holes in the panel to create a large ellipsoidal potential energy field above the hull. The potential energy bubble carries an electric dipole moment which when rotated with the hull generates a magnetic moment in the vertical direction.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[**0010**] Not Applicable.

A BRIEF DESCRIPTION OF THE DRAWINGS

[**0011**] **FIG. 1**. Perspective view of electric dipole moment spacecraft.

[**0012**] **FIG. 2**. Perspective exploded view of one panel with the ellipsoidal domes, flat hull panel with three holes, the charged rod grid and the planar potential energy field.

[**0013**] **FIG. 3**. Flat potential energy field produced by electrically charge wire rods.

[**0014**] **FIG. 4**. The potential in terms of the electric field of the rods times distance from rods.

[**0015**] **FIG. 5**. Cylindrical coordinates $\{t, r, \theta, z\}$.

[**0016**] **FIG. 6**. Ellipsoidal potential energy field emerging from hole in plate. This produces an electric dipole moment.

[**0017**] **FIG. 7**. Computer simulation of potential energy bubble.

[**0018**] **FIG. 8**. Centroid of an ellipse needed to calculate dipole moment.

[**0019**] **FIG. 9**. The charge of the hole.

[**0020**] **FIG. 10**. The units are charge.

[**0021**] **FIG. 11**. The dipole moment p.

[**0022**] **FIG. 12**. The magnetic moment in terms of the dipole moment and hull surface velocity.

[**0023**] **FIG. 13**. The magnetic moment in tensor notation.

[**0024**] **FIG. 14**. The lift force on the hull is equal to the gradient of the dot product of the magnetic moment and magnetic field.

[**0025**] **FIG. 15**. Maxwell's equation for the electric field changing with time.

[**0026**] **FIG. 16**. The components of Maxwell's equation.

[**0027**] **FIG. 17**. The radial electric field rotates with the angular velocity of the hull.

[**0028**] **FIG. 18**. The magnetic field is equal to the velocity of the hull times the electric field of the hull divided by the speed of light squared.

[**0029**] **FIG. 19**. The magnetic field gradient is proportional to the tangential velocity gradient along the panel.

[**0030**] **FIG. 20**. The hull profile needed to get the velocity gradient.

[0031] **FIG. 21**. The three vectors of the electric dipole moment (A), the tangential velocity of the hull (B), and the magnetic moment (C).

[0032] **FIG. 22**. The magnetic moment components in the radial and vertical direction. The dot product of the vertical magnetic moment with the magnetic field gradient is the lift force.

[0033] **FIG. 23**. Cross section of dome showing two layer cladding with different permittivities to enhance potential energy field.

[0034] **FIG. 24**. Computer simulation of potential energy field around dome.

DETAILED DESCRIPTION OF THE INVENTION

[0035] 1. Referring to **FIG. 1**, the spacecraft is a rotating octagon of trapezoidal electrostatically charged flat panels which form a closed hull (A). Each panel has three ellipsoidal domes (B) of varying size centrally located along the major length of the panel. The purpose of the dome is to create a large ellipsoidal potential energy bubble over the hull which develops an electrical dipole moment. Because the hull is rotating, a magnetic moment is created in the vertical direction. A magnetic field gradient created by the rotating electric field on the hull in combination with said magnetic moment produces a lift force on the hull.

[0036] 2. Referring to **FIG. 2**, the trapezoidal hull panel (A) contains three ellipsoidal holes (E). A group of wires or rods (C) running parallel to and just underneath the panel are electrically charged to a high voltage at the end terminals (B). Said rods produce a planar potential energy field (D) just under the holes in the panel. The field emerges from the holes in the shape of an ellipsoidal bubble and is amplified by an ellipsoidal dome (F) on the outside of the hull.

[0037] 3. Referring to **FIG. 3**, the group of parallel rods (A) are given a linear charge λ in units of charge per meter. The electric field E developed by the rod is the linear charge divided by the circumference of a circle of radius r around the wire times the permittivity ϵ of space. The analysis of this arrangement shows that within a few grid width spacings, the potential energy field ϕ has become planar (B) in the z-direction given by the equation in **FIG. 4**.

[0038] 4. The forthcoming analysis is done in cylindrical coordinates, **FIG. 5**.

[0039] 5. Referring to **FIG. 6**, the ellipsoidal potential energy (B) emerges through the hole in the panel plate (A). In doing so it creates an electrical dipole moment (C) shown by the arrow normal to the hole area. The theoretical analysis matched a computer simulation of this emergence of the bubble (B) through plate (A), as shown in **FIG. 7**.

[0040] 6. Because the bubble has the shape of an ellipse, the centroid of the bubble would be four thirds the radius a divided by π as given in **FIG. 8**. The electric dipole moment is then given as the charge times said centroid. The charge of the hole is equal to the permittivity times the electric field emerging from the hole times the area of the hole as seen in the equation of **FIG. 9** with the units of charge, **FIG. 10**.

[0041] 7. The electric dipole moment is said centroid times said charge, as seen in the equation given in **FIG. 11** and having units of coulomb-meter. Notice that the electric dipole moment p times a velocity v is equal to a magnetic moment μ which is what creates the lift force on the hull as shown in the next **FIG. 12**.

[0042] 8. The rotating hull creates the electric dipole moment velocity so that the entire hull develops a magnetic moment. In tensor notation, **FIG. 13**, the magnetic moment is in the vertical direction because there is a radial component of the electric dipole moment times the velocity. The velocity is the radius r in the radial direction times the angular velocity in the z-direction.

[0043] 9. The force on the hull is the gradient of the dot product of the magnetic moment with the magnetic B field given in equation **FIG. 14**.

[0044] 10. By electrically charging the hull of the vehicle, a radial electric field is produced. By rotating the hull, the radial electric field changes with time. Thus Maxwell's equations will involve the curl of the magnetic field in the radial direction because the radial electric field is varying with time. Maxwell's equation is seen in **FIG. 16**.

[0045] 11. The cross product involves the magnetic field in the theta direction which is zero. The electric field varies with time as shown in **FIG. 17**. Substituting the derivative of the electric field and integrating with respect to angle theta gives the vertical magnetic field as the tangential velocity times the radial field divided by the speed of light squared (**FIG. 18**).

[0046] 12. Now the force on the hull is the gradient of the magnetic moment times the magnetic field. In the equation for the magnetic field, the only available variable to work with in order to get a gradient is the velocity. To me this was the surprise in this invention. The velocity had to be a function of the height of the hull (**FIG. 19**) which meant that the hull had to be in the shape of a pyramid, **FIG. 20**. Using eight flat sides keeps the radial electric field pointing in the same direction in each panel. Each panel has three domes to produce the magnetic moment for a total of 24 magnetic moment generators.

[0047] 13. Referring to **FIG. 21**, the electric dipole moment (A) points in the radial direction, the rotating hull produces a tangential velocity (B), and the result is a magnetic moment (C) along the panel.

[0048] 14. Because the magnetic moment is parallel to the panel, there are vertical and radial components of the magnetic moment. The vertical magnetic moment creates the dot product with the magnetic field gradient, which is equal to the lift force, as depicted in **FIG. 22**.

[0049] 15. **FIG. 23** shows a cross-section of the dome (A) and the plate hole (B) with double cladding to enhance the field as determined by computer simulation. The upper cladding (D) has a low relative permittivity in the range of 2 to 40, and the lower layer has a high relative permittivity in the range of 1200 to 4000. This dome and cladding configuration creates a much larger electric dipole moment compared to a hole in the plate, comparing **FIG. 23** to that of **FIG. 7**. The wavy lines are the equi-potential energy lines from the dome (C) and the upper layer (B) and the lower level (A).

What I claim as my invention is:

1. A spacecraft having a rotating octagon of trapezoidal electrostatically charged flat panels which form a closed hull in the shape of a pyramid.

2. Said panels each having three holes covered by three ellipsoidal domes of varying size centrally located along the major length of the panel.

3. A grid of high voltage electrostatically charged rods located on the interior side of each panel such that a planar potential field is produced parallel to and under each hole of said panel.

4. Said domes and holes in panel produce an ellipsoidal potential energy field carrying an electric dipole moment on the outside of the hull.

5. Said rotating hull together with electric dipole moment produce a magnetic moment in the vertical direction.

6. Said rotating charged hull produces an electric field in the radial direction which generates a corresponding magnetic field gradient in the vertical direction proportional-to the velocity gradient of the sloping panels of the hull.

7. Combined magnetic moment and magnetic field produce lift on the hull.

8. Dome material and surface layer cladding having different permittivities which enhance said electric dipole moment.

9. Use of magnetic vortex generator to bring in low density hyperspace energy in order to increase the field strengths as described in my patent applications Magnetic Vortex Wormhole Generator and Magnetic Vortex Generator.

* * * * *

US 20060070371A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2006/0070371 A1**

St. Clair (43) **Pub. Date: Apr. 6, 2006**

(54) **ELECTRIC DIPOLE MOMENT PROPULSION SYSTEM**

(76) Inventor: **John Quincy St. Clair**, San Juan, PR (US)

Correspondence Address:
JOHN ST. CLAIR
52 KINGS COURT, 4A
SAN JUAN, PR 00911 (US)

(21) Appl. No.: **10/958,436**

(22) Filed: **Oct. 5, 2004**

Publication Classification

(51) **Int. Cl.**
F03H 1/00 (2006.01)

(52) **U.S. Cl.** **60/203.1**

(57) **ABSTRACT**

A spacecraft propulsion system utilizing a rotating octagon of trapezoidal electrically charged flat panels to create an electric dipole moment that generates lift on the hull.

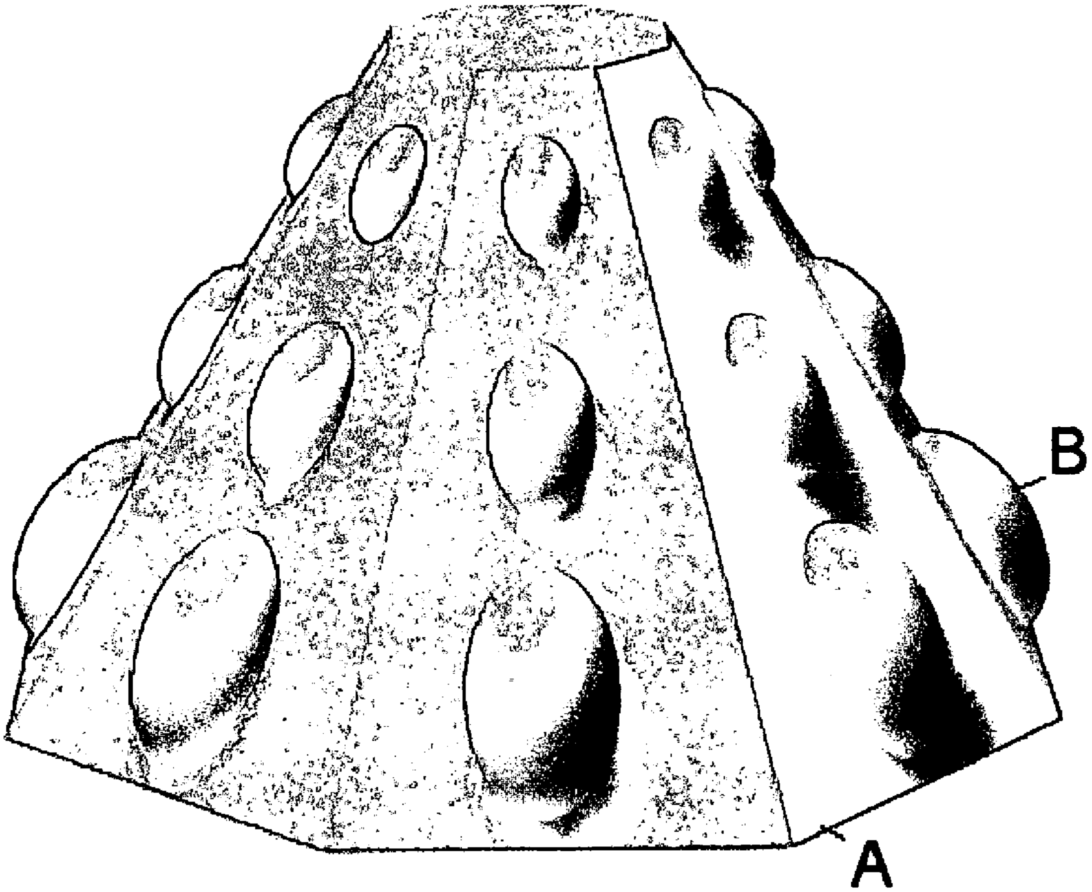

Figure 1

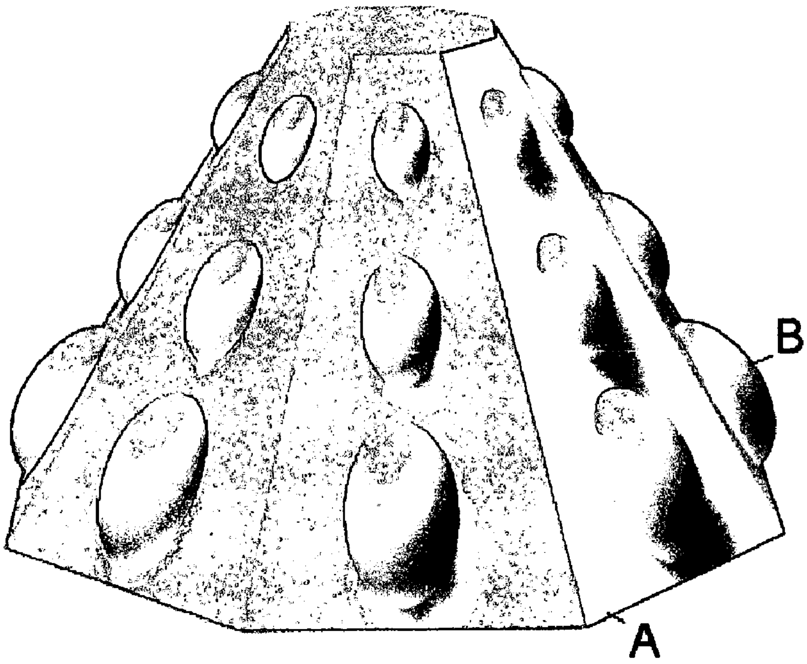

Figure 2

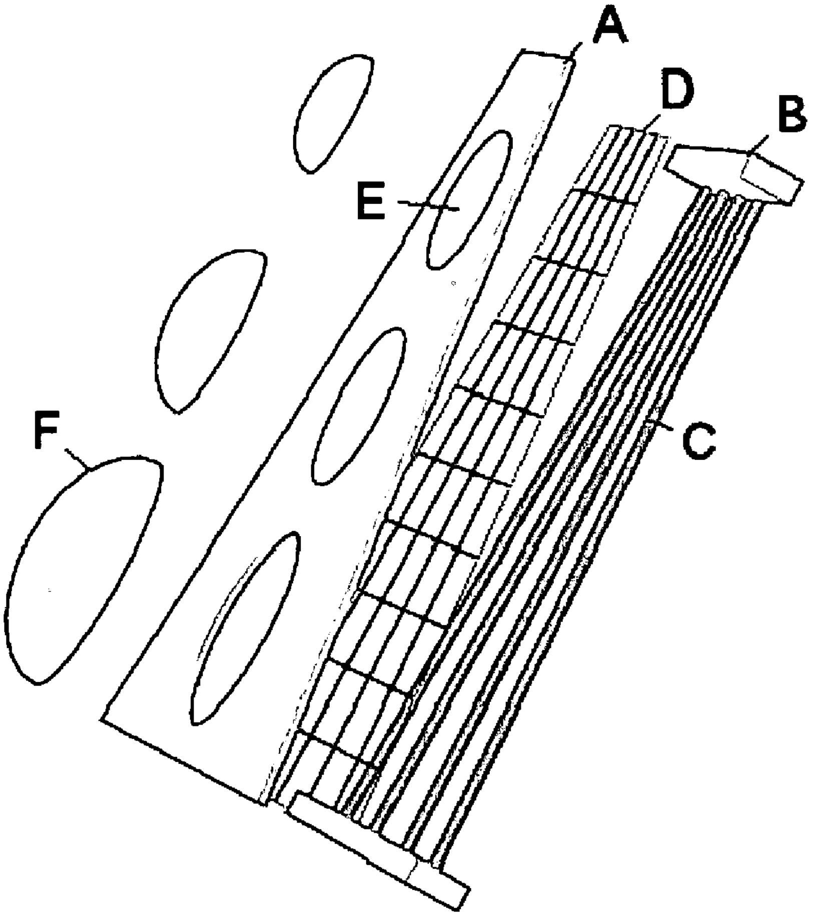

Figure 3

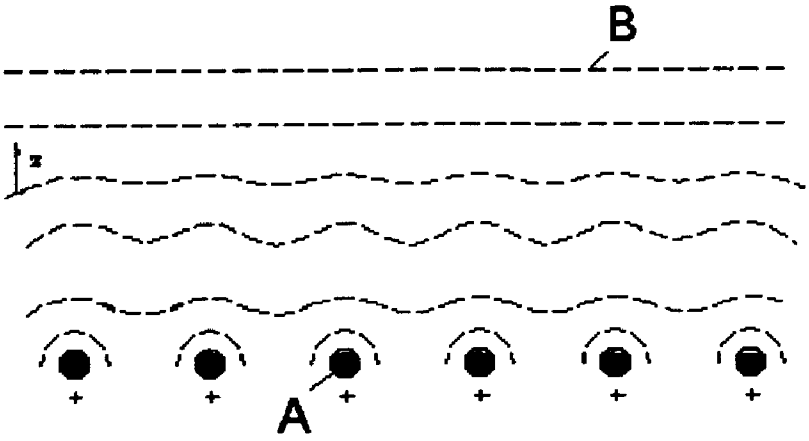

Figure 4

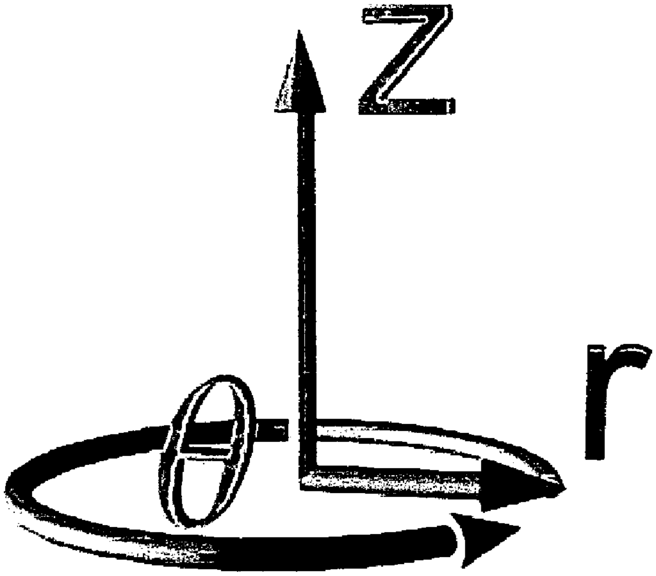

Figure 5

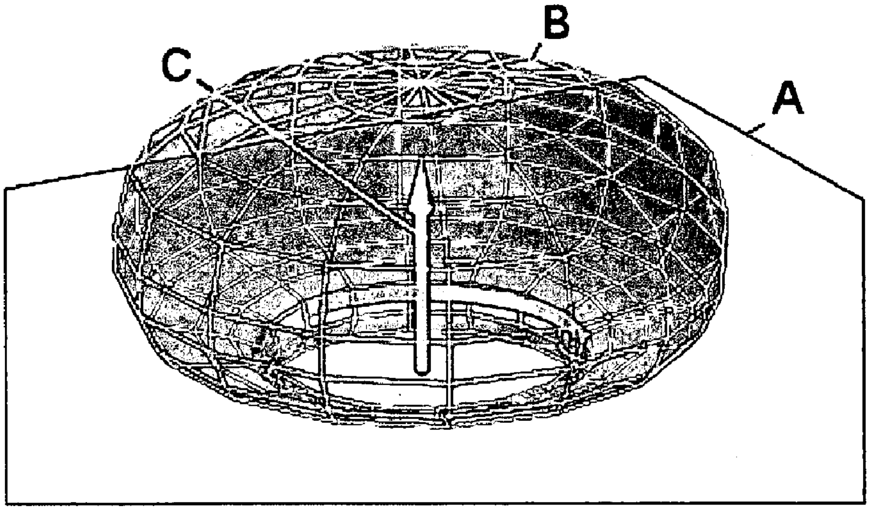

Figure 6

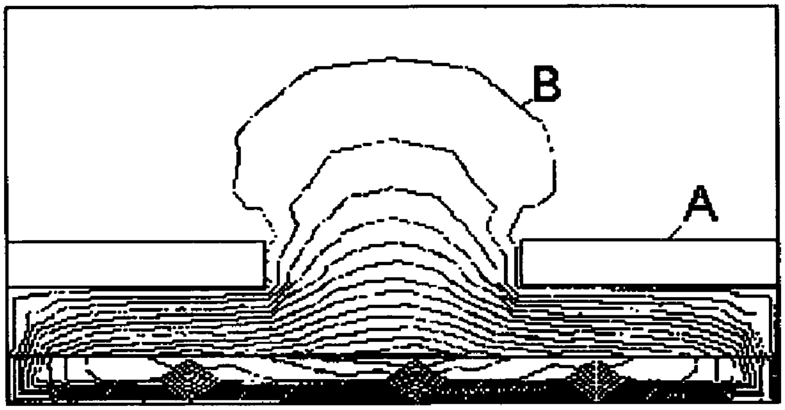

Figure 7

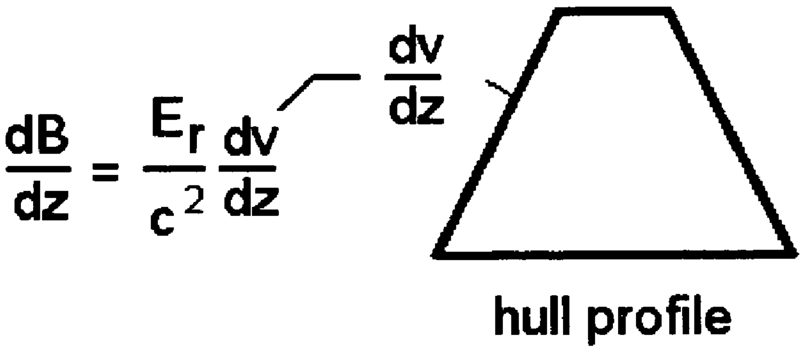
$\frac{dB}{dz} = \frac{E_r}{c^2}\frac{dv}{dz}$
$\frac{dv}{dz}$
hull profile

Figure 8

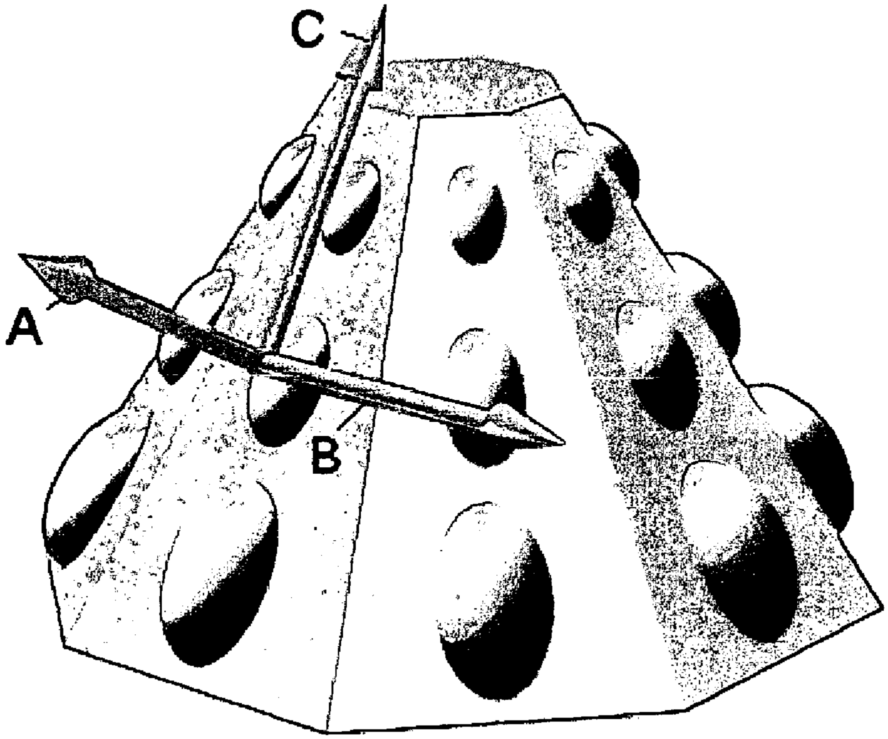

Figure 9

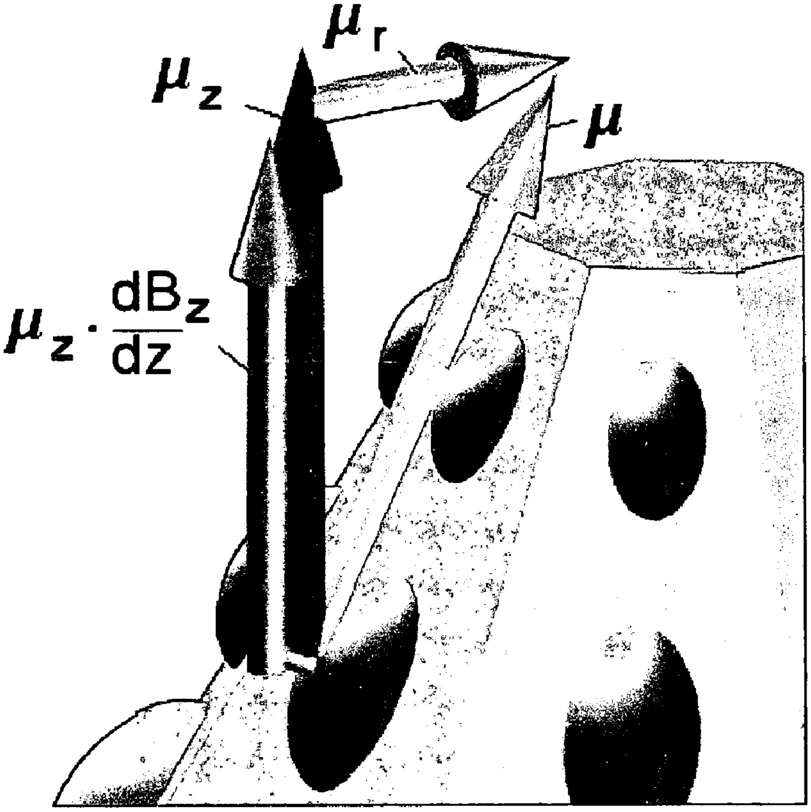
μ_r
μ_z
μ
$\mu_z \cdot \frac{dB_z}{dz}$

Figure 10

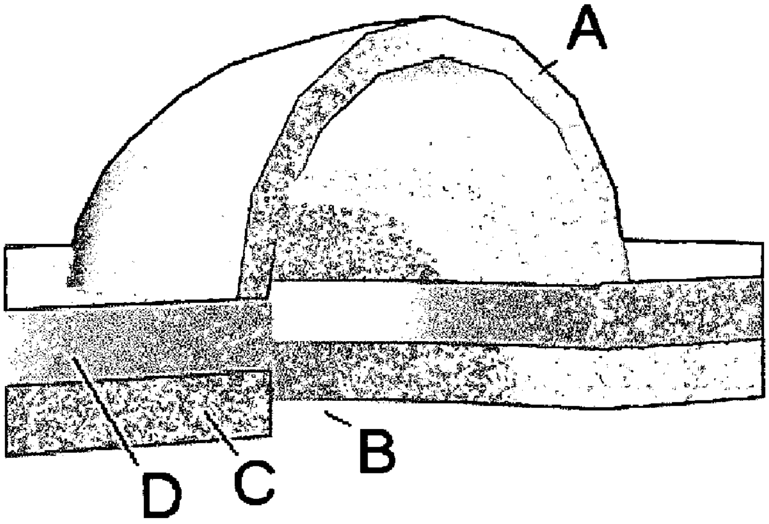

Figure 11

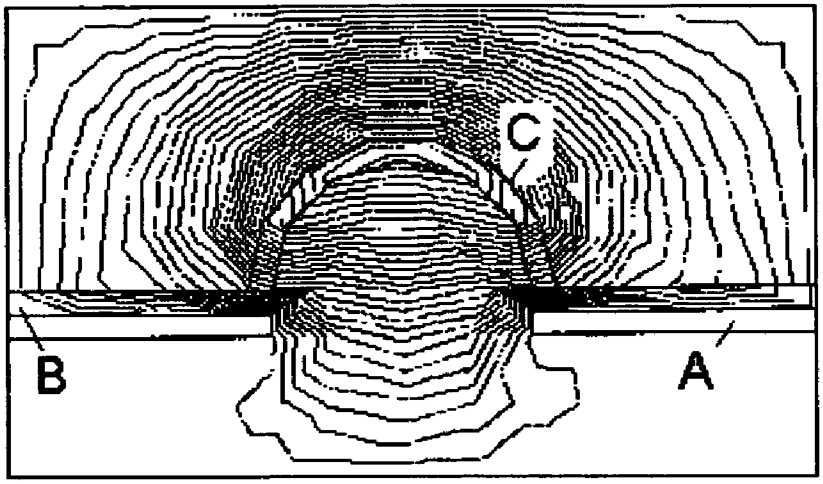

ELECTRIC DIPOLE MOMENT PROPULSION SYSTEM

BRIEF SUMMARY OF THE INVENTION

[0001] The invention is a spacecraft utilizing trapezoidal electrostatically charged flat plate panels which form a pyramidal hull. A panel contains three holes each of which produces a potential energy ellipsoidal bubble that creates an electric dipole moment. The rotation of the hull generates a magnetic moment and a magnetic field gradient in the vertical direction that produces a lift force on the spacecraft.

BACKGROUND OF THE INVENTION

[0002] It is known from electrodynamics that a hole in a conducting plane forms a potential energy bubble. This bubble creates an electric dipole moment from which it is possible to develop a magnetic moment. A rotating tilted hull produces a velocity gradient that generates a magnetic field gradient in the vertical direction. This combination produces a lift force on the spacecraft. A very large potential energy bubble is produced provided that the hole protrudes out of the plate in an ellipsoidal shape. Furthermore, a double cladding, in which each layer around the hole has a different permittivity, confines the field to the outside of the hull for even better results.

[0003] The planar potential energy is created by a grid of electrically charged wires or rods running the length of each panel. A circular potential energy from each rod very quickly sums to form a flat sheet of energy which emerges from the hole to form the potential energy bubble.

SUMMARY OF THE INVENTION

[0004] The invention relates to a spacecraft utilizing a rotating octagon of trapezoidal electrically charged flat plate panels to form a hull in the shape of a pyramid. Each panel has three protruding ellipsoidal bubbles that produce an electric dipole moment from a planar potential energy field created by a group of charged rods parallel to the panel. Because the panels are tilted and the hull is rotating, there is a tangential velocity gradient in the vertical direction. This creates the magnetic moment. Because the hull rotates, the radial electric field produces a magnetic field gradient in the vertical direction. This combination of magnetic moment and magnetic field gradient produces a lift force on the hull of the spacecraft.

[0005] On the underside of each panel is a group of high voltage electrically charged rods which run parallel to the panel. These wires or rods produce a planar electrical potential field underneath the holes in the panel. This potential energy field then bubbles out of the holes in the panel to create a large ellipsoidal potential energy field above the hull. The potential energy bubble carries an electric dipole moment which when rotated with the hull generates a magnetic moment in the vertical direction.

A BRIEF DESCRIPTION OF THE DRAWINGS

[0006] **FIG. 1**. Perspective view of electric dipole moment spacecraft.

[0007] **FIG. 2**. Perspective exploded view of one panel with the ellipsoidal domes, flat hull panel with three holes, the charged rod grid and the planar potential energy field.

[0008] **FIG. 3**. Planar view of flat potential energy field produced by electrically charge wire rods.

[0009] **FIG. 4**. Perspective view of cylindrical coordinates $\{r, \theta, z\}$.

[0010] **FIG. 5**. Perspective view of ellipsoidal potential energy field emerging from hole in plate which produces an electric dipole moment.

[0011] **FIG. 6**. Planar view of field lines of potential energy bubble emerging from plate hole.

[0012] **FIG. 7**. Planar view of sloping hull profile needed to get a velocity gradient.

[0013] **FIG. 8**. Perspective view of hull showing electric dipole moment, the tangential velocity of the hull, and the magnetic moment.

[0014] **FIG. 9**. Perspective view of the magnetic moment components in the radial and vertical direction whereby the lift force is generated by the dot product of the vertical magnetic moment with the magnetic field gradient.

[0015] **FIG. 10**. Perspective view of cross section of dome showing two layer cladding with different permittivities to enhance potential energy field.

[0016] **FIG. 11**. Planar view of enhanced potential energy field with two layer cladding.

DETAILED DESCRIPTION OF THE INVENTION

[0017] 1. Referring to **FIG. 1**, the spacecraft is a rotating octagon of trapezoidal electrostatically charged flat panels which form a closed hull (A). Each panel has three ellipsoidal domes (B) of varying size centrally located along the major length of the panel. The purpose of the dome is to create a large ellipsoidal potential energy bubble over the hull which develops an electrical dipole moment. Because the hull is rotating, a magnetic moment is created in the vertical direction. A magnetic field gradient created by the rotating electric field on the hull in combination with the magnetic moment produces a lift force on the hull.

[0018] 2. Referring to **FIG. 2**, the trapezoidal hull panel (A) contains three ellipsoidal holes (E). A group of wires or rods (C) running parallel to and just underneath the panel are electrically charged to a high voltage at the end terminals (B). The rods produce a planar potential energy field (D) just under the holes in the panel. The field emerges from the holes in the shape of an ellipsoidal bubble and is amplified by an ellipsoidal dome (F) on the outside of the hull.

[0019] 3. Referring to **FIG. 3**, the group of parallel rods (A) are given a linear charge λ in units of charge per meter. The electric field E developed by the rod is the linear charge divided by the circumference of a circle of radius r around the wire times the permittivity ϵ of space. The analysis of this arrangement shows that within a few grid width spacings, the potential energy field ϕ due to the electric field E_0 has become planar (B) in the z-direction given by the equation

$$\phi=-E_0 z$$

[0020] 4. Referring to **FIG. 4**, the following analysis is done in cylindrical coordinates {r,θ,z}.

[0021] 5. Referring to **FIG. 5**, the ellipsoidal potential energy (B) emerges through the hole in the panel plate (A). In doing so it creates an electrical dipole moment (C) shown by the arrow normal to the hole area.

[0022] 6. Referring to **FIG. 6**, the bubble (B) emerges through plate (A).

[0023] 7. Because the bubble has the shape of an ellipse, the centroid y of the bubble would be four thirds the radius a divided by π as given by

$$\overline{y}_{ellipse} = \frac{4}{3}\frac{a}{\pi}$$

[0024] The electric dipole moment is then given as the charge q times the centroid y. The charge of the hole is equal to the permittivity E times the electric field E emerging from the hole times the area of the hole of radius a

$$q = \varepsilon_0 E\pi a^2 \frac{\text{coul}^2}{\text{m}^2\ \text{newton}}\frac{\text{newton}}{\text{coul}}\ \text{m}^2 = \text{coul} = \text{charge}$$

[0025] 8. The electric dipole moment p is the centroid y times the charge q

$$p = qy = \varepsilon_0 E\pi a^2 \frac{4}{3}\frac{a}{\pi} = \frac{4}{3}\varepsilon_0 E a^2\ \text{coul}\cdot\text{meter}$$

[0026] The electric dipole moment p times the hull velocity v is equal to a magnetic moment μ which is what creates the lift force on the hull

$\mu = pv$ amp·m^2

[0027] 9. The rotating hull creates the electric dipole moment velocity so that the entire hull develops a magnetic moment. In tensor notation, the magnetic moment μ is in the vertical z-direction because there is a radial component of the electric dipole moment times the velocity. The velocity is the radius r in the radial direction times the angular velocity ω in the z-direction

$\mu^z = p_r x^r \omega^z$

[0028] 10. The force F on the hull is the gradient of the dot product of the magnetic moment μ with the magnetic B field

$F = \nabla(\mu \cdot B)$

[0029] 11. By electrically charging the hull of the vehicle, a radial electric field is produced. By rotating the hull, the radial electric field changes with time. Thus Maxwell's equations will involve the curl of the magnetic field in the radial direction because the radial electric field is varying with time

$$(\nabla \times B)_r = \frac{1}{c^2}\frac{\partial E_r}{\partial t}$$

[0030] 12. The cross product involves the magnetic field in the theta direction which is zero

$$\frac{1}{r}\frac{\partial B_z}{\partial \theta} - \frac{\partial B_\theta}{\partial z} = \frac{1}{c^2}\frac{\partial E_r}{\partial t} = \frac{1}{r}\frac{\partial B_z}{\partial \theta}$$

[0031] Substituting the derivative of the electric field E

$E_r = E_0 e^{i\omega t}$

[0032] and integrating with respect to angle theta gives the vertical magnetic field B as the tangential velocity v times the radial electric field E divided by the speed of light c squared

$$B_z = \frac{v}{c^2}E_r$$

[0033] 13. The force on the hull is the gradient of the magnetic moment μ times the magnetic field B. In the equation for the magnetic field, the only available variable to work with in order to get a gradient of the magnetic field comes from the velocity.

$$\frac{dB_z}{dz} = \frac{E_r}{c^2}\frac{dv}{dz}$$

[0034] 14. Referring to **FIG. 7**, because the hull is in the shape of a pyramid, the velocity is a function of the height z of the hull. Using eight flat sides keeps the radial electric field pointing in the same direction in each panel. Each panel has three domes to produce the magnetic moment for a total of 24 magnetic moment generators.

[0035] 15. Referring to **FIG. 8**, the electric dipole moment (A) points in the radial direction, the rotating hull produces a tangential velocity (B), and the result is a magnetic moment (C) along the panel.

[0036] 16. Referring to **FIG. 9**, because the magnetic moment is parallel to the panel, there are vertical and radial components of the magnetic moment. The vertical magnetic moment creates the dot product with the magnetic field gradient, which is equal to the lift force.

[0037] 17. **FIG. 10** shows a cross-section of the dome (A) and the plate hole (B) with double cladding to enhance the field. The upper cladding (D) has a low relative permittivity in the range of 2 to 40, and the lower layer has a high relative permittivity in the range of 1200 to 4000.

[0038] 18. Comparing **FIG. 11** to **FIG. 7**, this dome and cladding configuration creates a much larger electric dipole moment compared to a hole in the plate. The wavy lines are the equi-potential energy lines from the dome (C) and the upper layer (B) and the lower level (A).

I claim:

1. A spacecraft propulsion system comprising:

a rotating octagon of trapezoidal electrostatically charged flat panels which form a closed sloping hull in the shape of a pyramid;

panels each having three holes covered by three ellipsoidal domes of varying size centrally located along the major length and axis of each panel; and

a grid of high voltage electrostatically charged rods located on the interior side of each panel such that a planar potential field is produced parallel to and under each panel hole.

2. The domes, holes and rotating charged hull of method 1 producing:

an ellipsoidal potential energy field emerging from the holes and generating an electric dipole moment on the outside of the hull;

a magnetic moment in the vertical direction due to the rotating electric dipole moment; and

a rotating electric field in the radial direction which generates a corresponding magnetic field gradient in the vertical direction proportional to the velocity gradient of the sloping panels of the hull.

3. A lift force on the spacecraft hull generated by:

the magnetic moment times the gradient of the magnetic field in the vertical direction; and

a dual surface layer hull cladding having different permittivities which enhance the electric dipole moment whereby the upper cladding has a low relative permittivity in the range of 2 to 40, and the lower layer has a high relative permittivity in the range of 1200 to 4000.

* * * * *

US 20060038081A1

(19) **United States**
(12) **Patent Application Publication** (10) **Pub. No.: US 2006/0038081 A1**
St. Clair (43) **Pub. Date: Feb. 23, 2006**

(54) ELECTRIC DIPOLE SPACECRAFT

(76) Inventor: **John Quincy St. Clair**, San Juan, PR (US)

Correspondence Address:
JOHN ST. CLAIR
52 KINGS COURT, 4A
SAN JUAN, PR 00911 (US)

(21) Appl. No.: **10/912,621**

(22) Filed: **Aug. 4, 2004**

Publication Classification

(51) **Int. Cl.**
B64G 1/40 (2006.01)
(52) **U.S. Cl.** **244/171.5**

(57) ABSTRACT

This invention is a rotating spacecraft that produces an electric dipole on four rotating spherical conducting domes perturbing a uniform spherical electric field to create a magnetic moment interacting with the gradient of a magnetic field that generates a lift force on the hull.

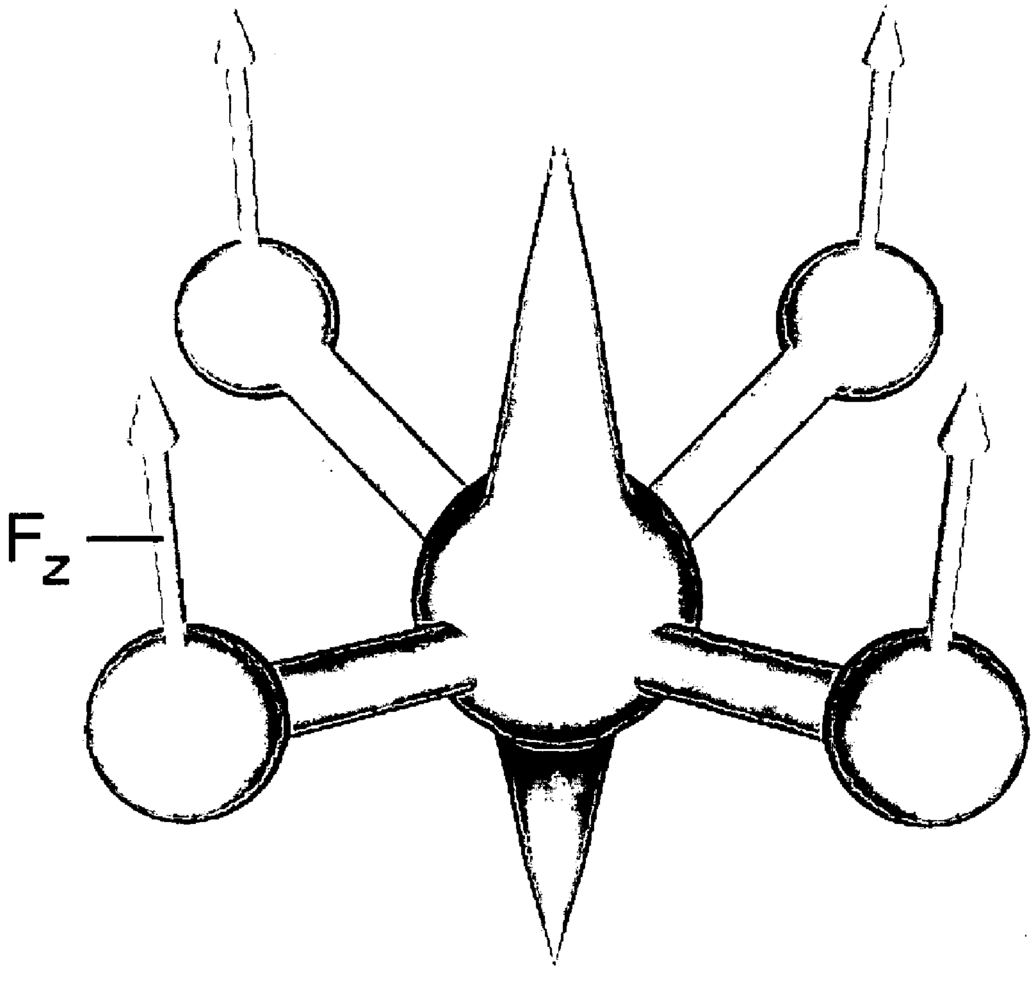

Figure 1

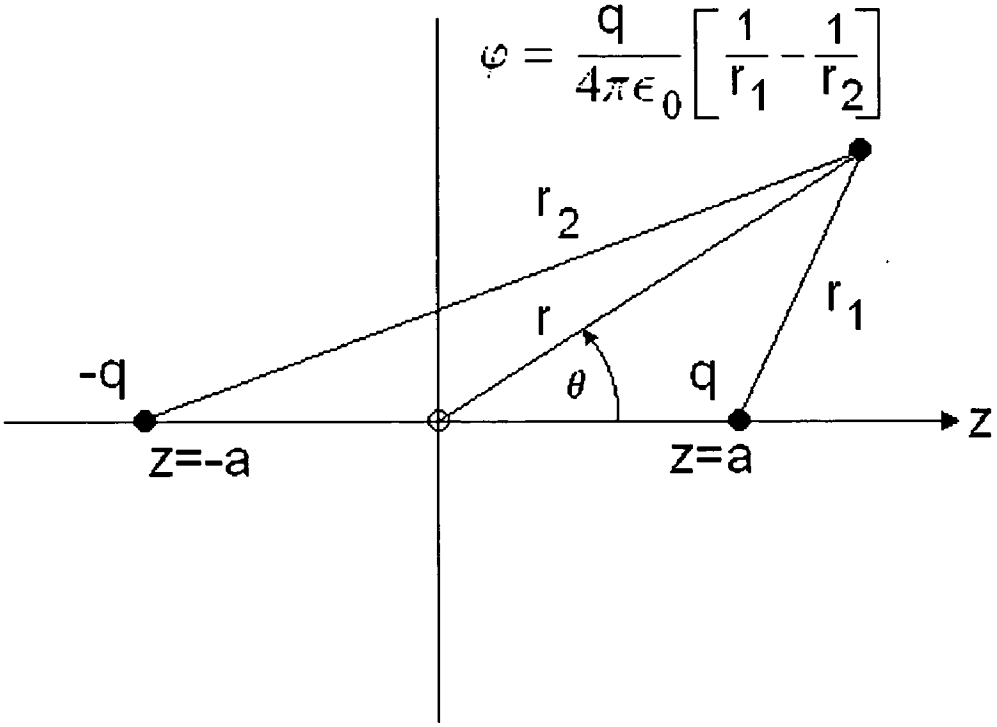
$\varphi = \frac{q}{4\pi\epsilon_0}\left[\frac{1}{r_1} - \frac{1}{r_2}\right]$
r_2
r
r_1
θ
-q
q
z=-a
z=a
z

Figure 2

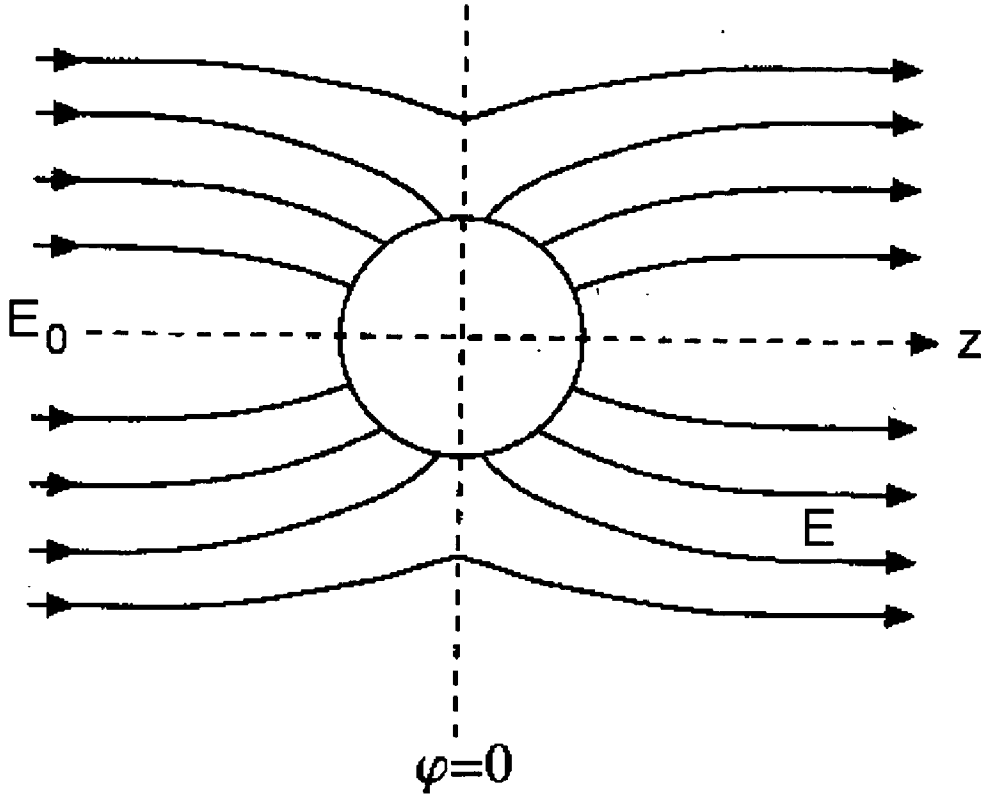
E_0
z
E
φ=0

Figure 3

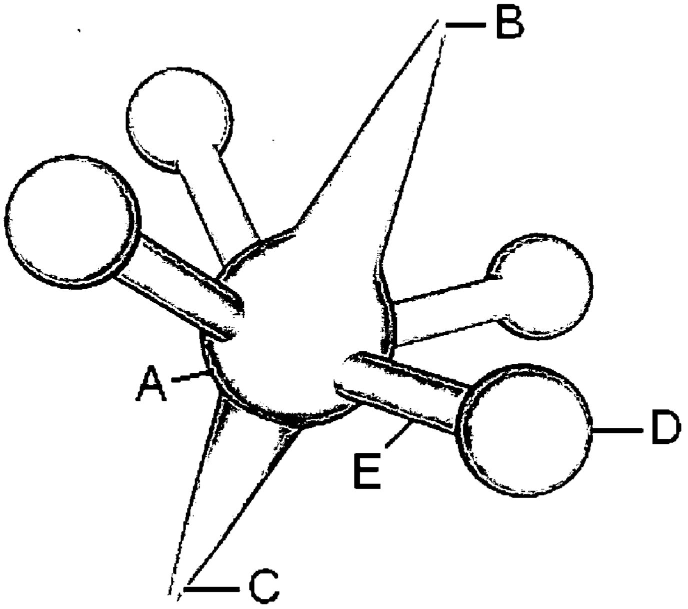

Figure 4

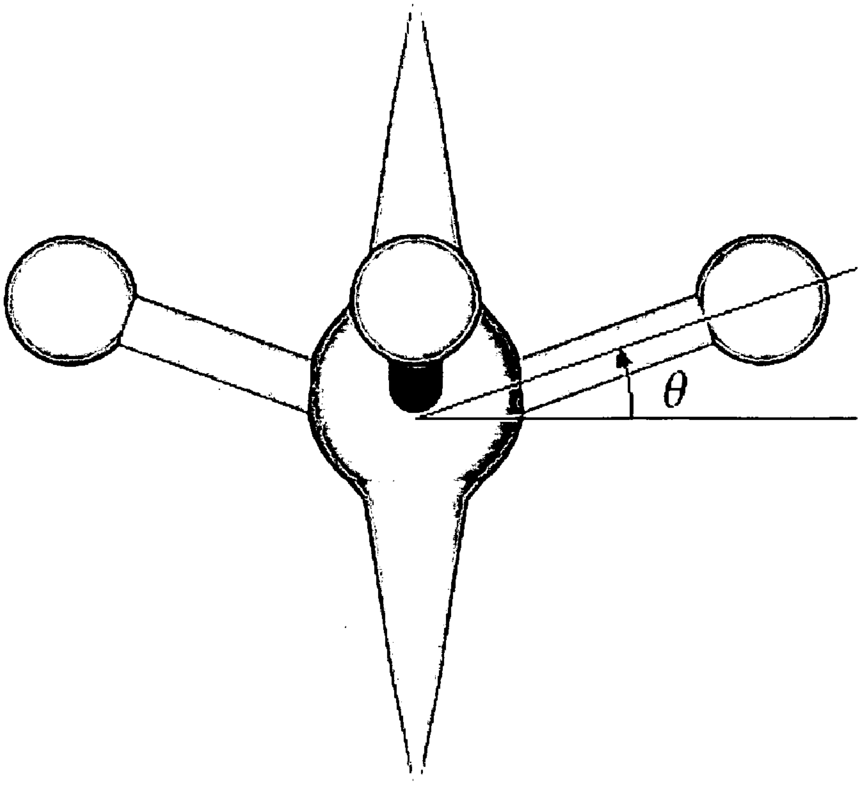

Figure 5

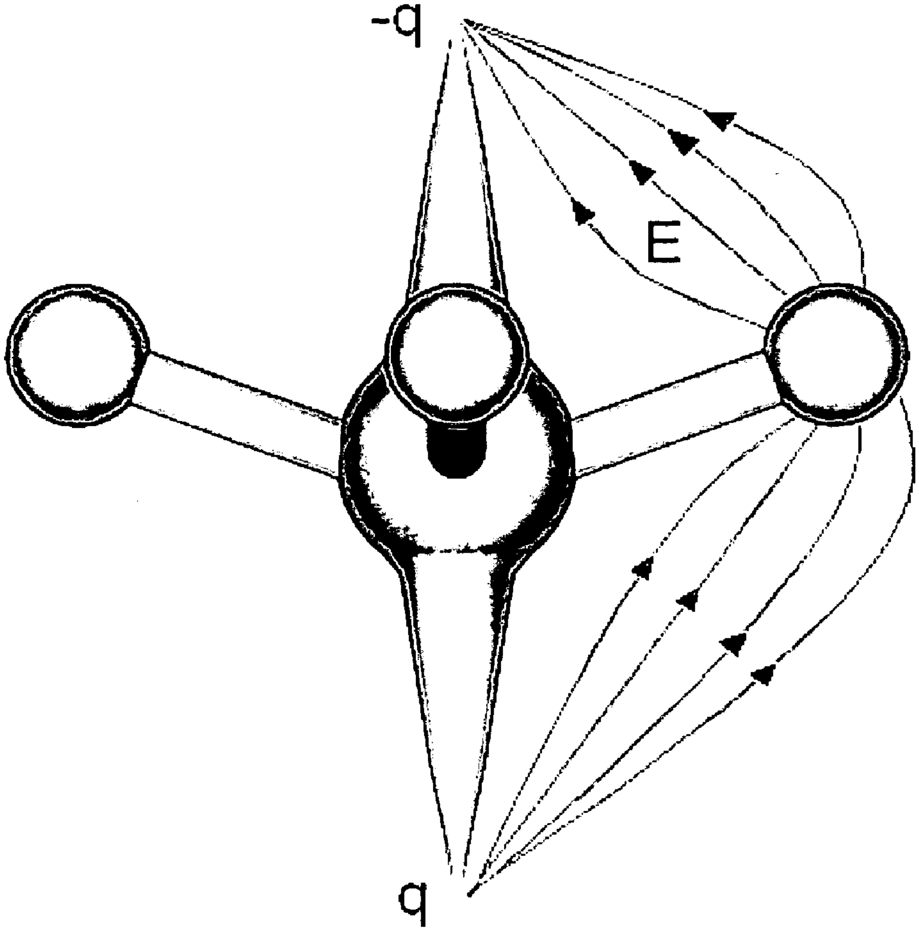

Figure 6

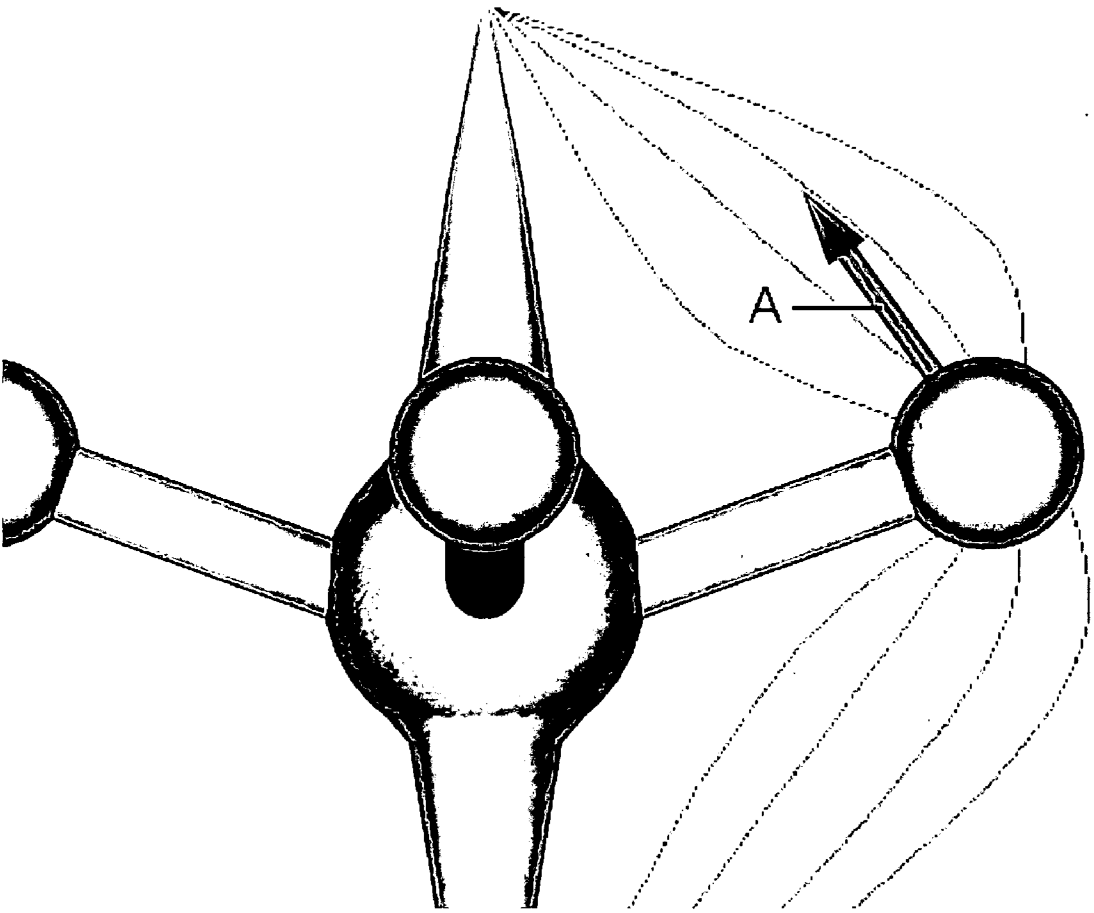

Figure 7

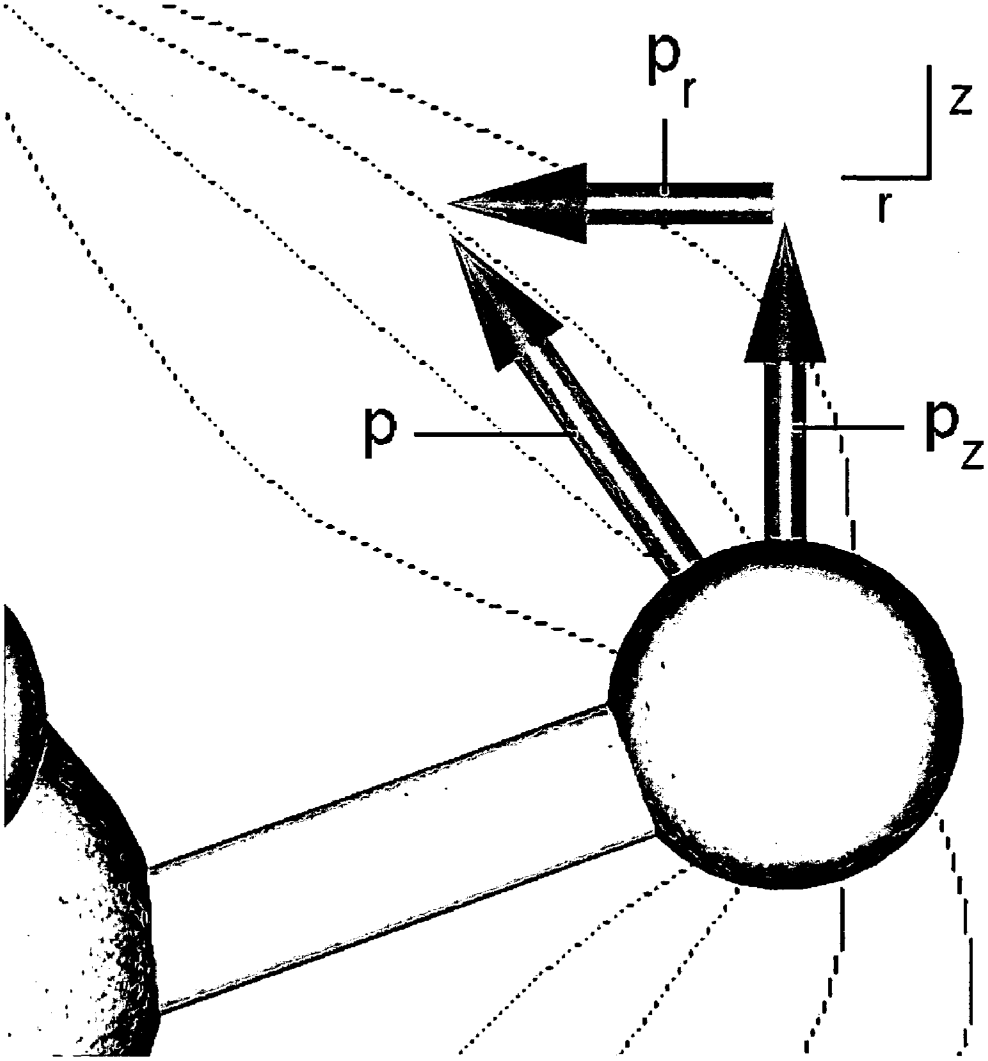

Figure 8

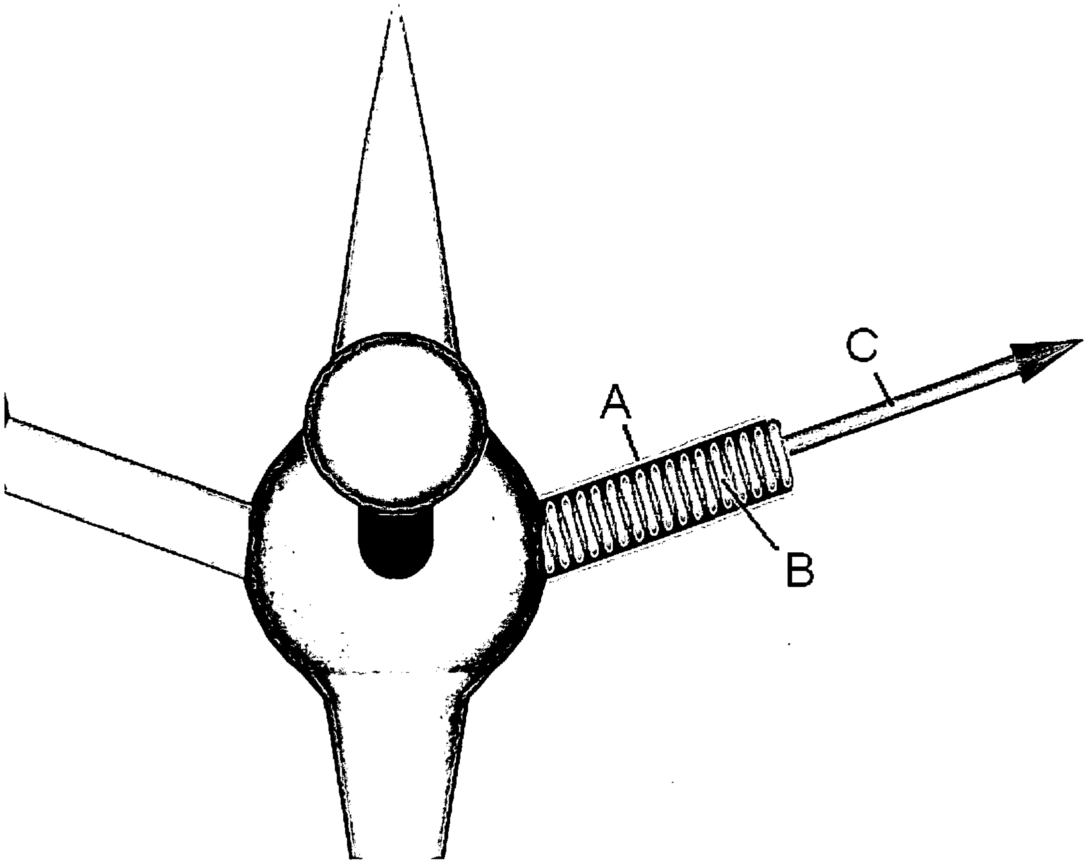

Figure 9

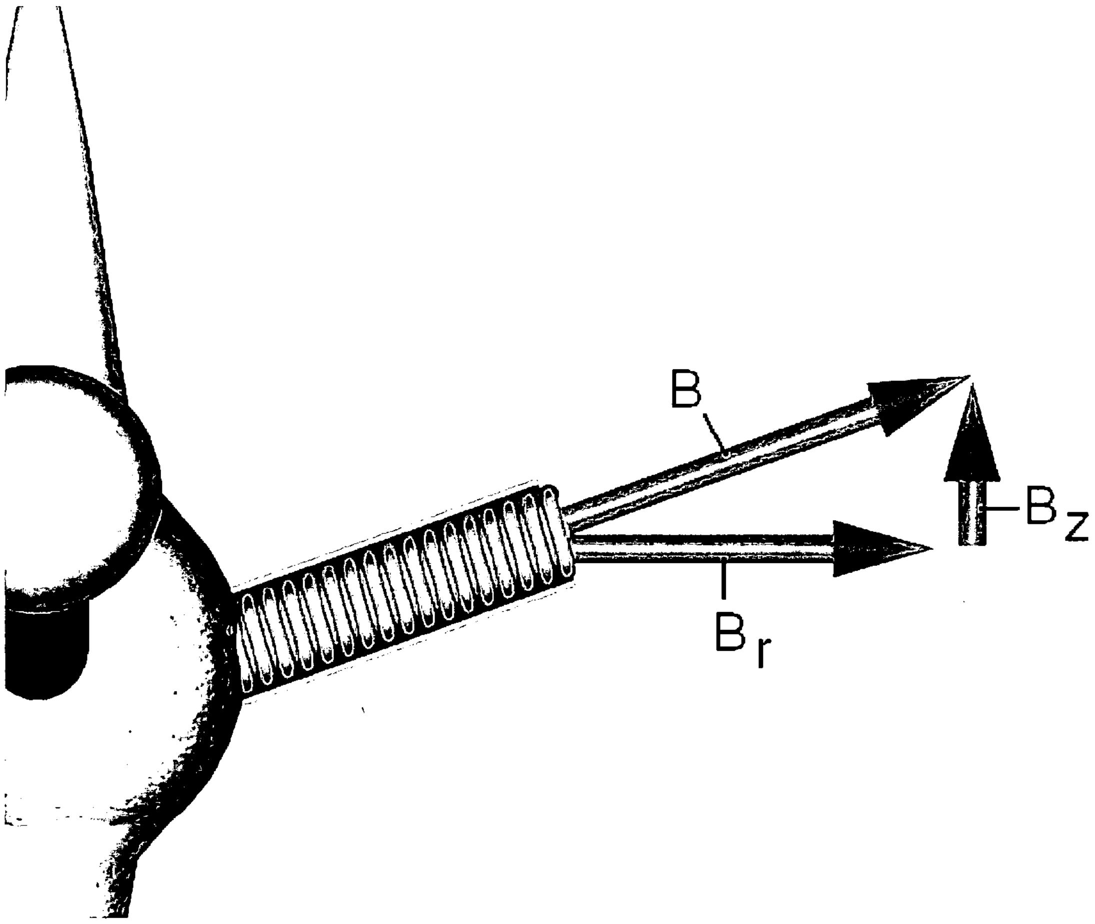

Figure 10

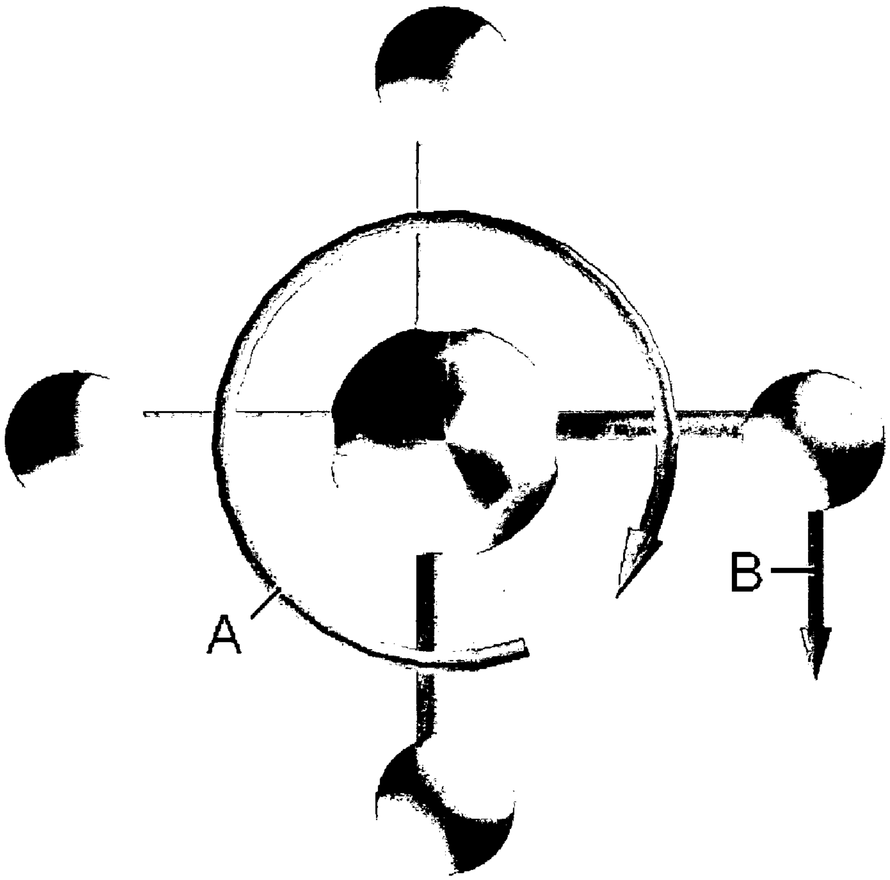

Figure 11

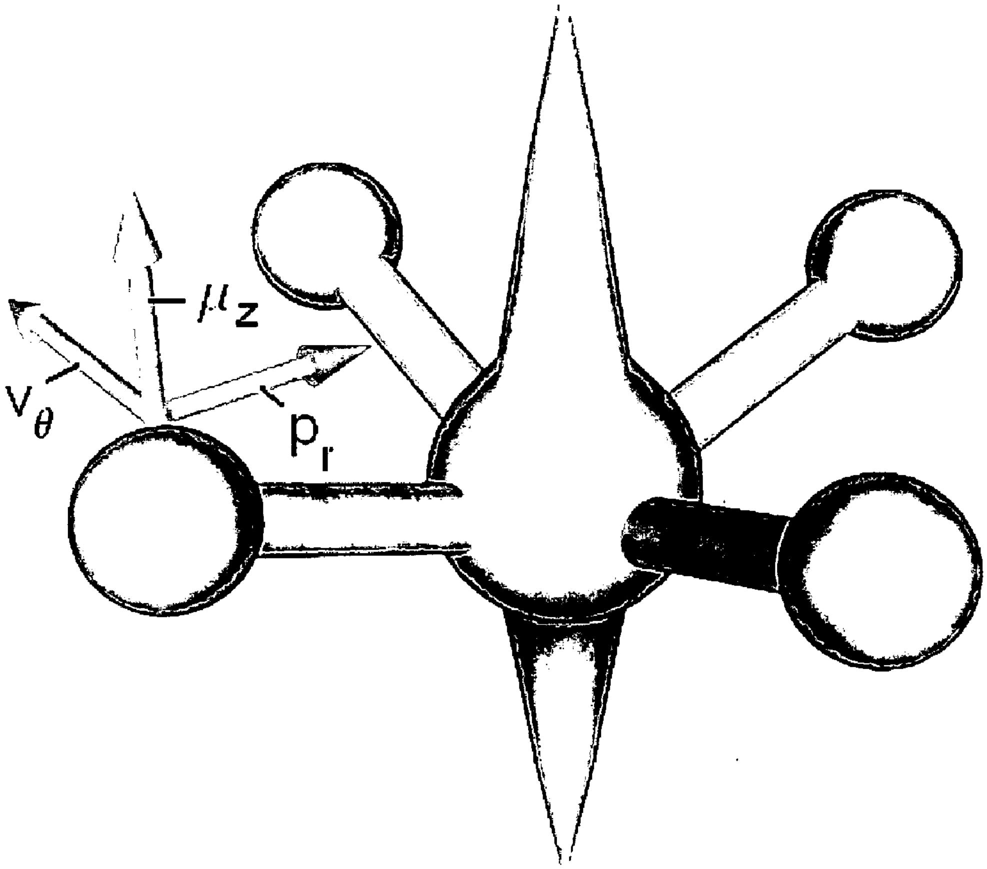
μ_z
v_θ
p_r

Figure 12

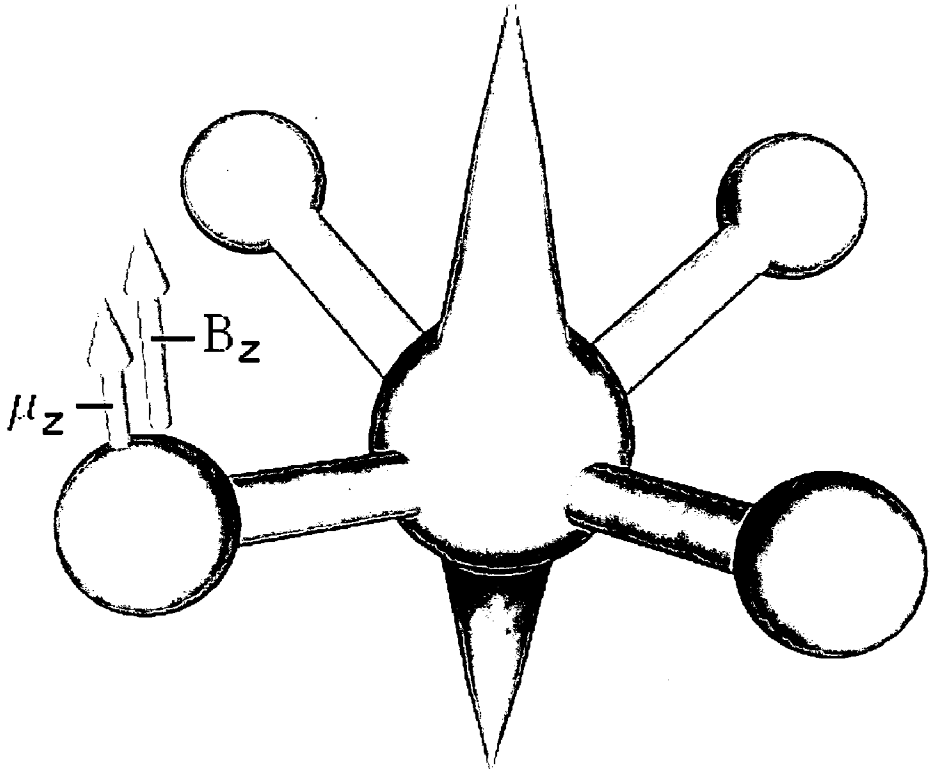

Figure 13

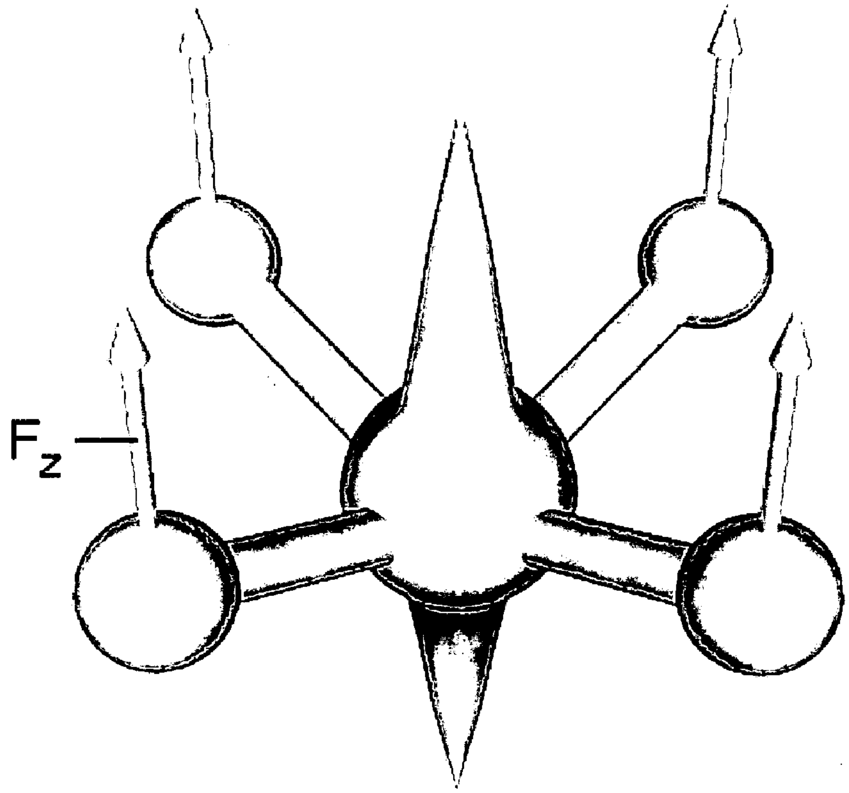

Figure 14

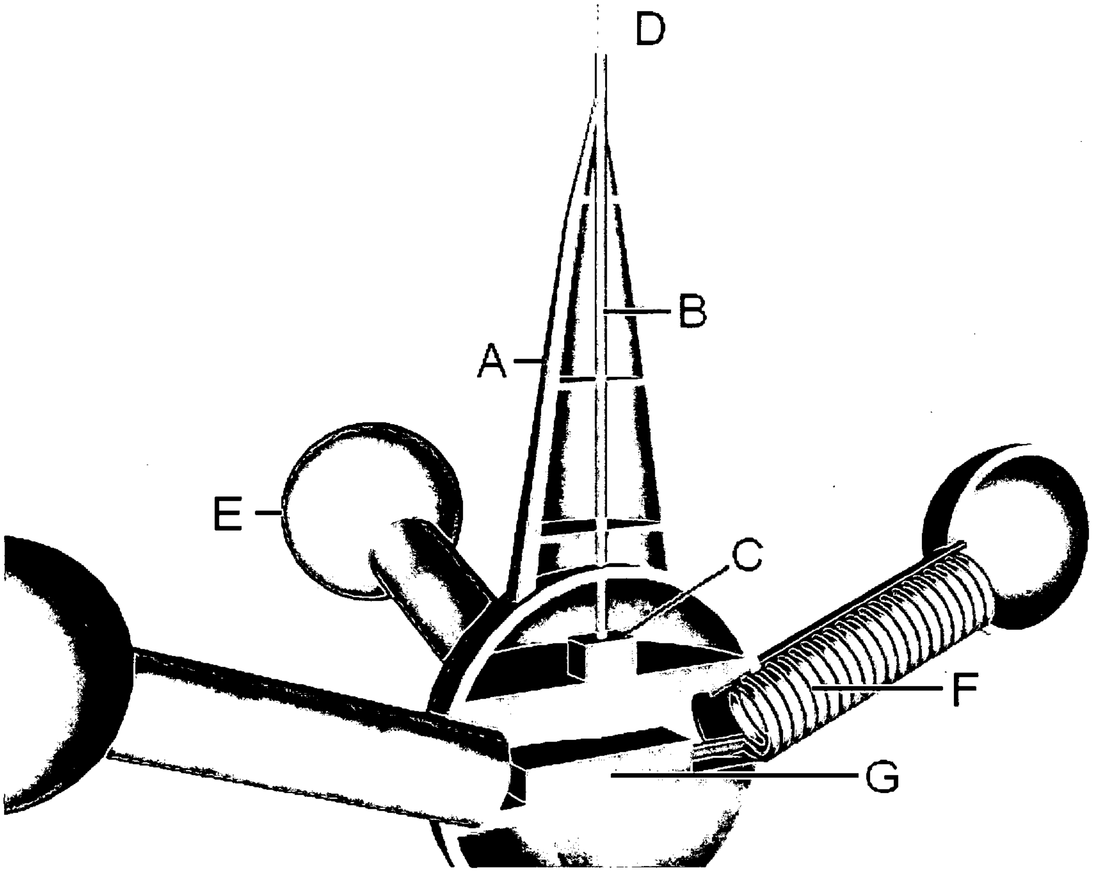

ELECTRIC DIPOLE SPACECRAFT

BRIEF SUMMARY OF THE INVENTION

[0001] This invention is a rotating spacecraft that utilizes four spherical conducting domes perturbing a uniform electric field in order to create a lift force by means of a magnetic moment times the gradient of a magnetic field.

BACKGROUND OF THE INVENTION

[0002] An electric dipole p is two electrical charges of opposite sign {q, −q} separated by a distance a.

$$p=qa=\text{coulomb}\cdot\text{meter}$$

[0003] If this dipole is moving with a velocity v, it produces a magnetic moment μ.

$$\mu = pv = qav = \text{coulomb meter}\frac{\text{meter}}{\text{sec}} = \frac{\text{coulomb}}{\text{sec}}\text{meter}^2 = IArea$$

[0004] The magnetic moment is equal to a current I circulating around an area. The magnetic field B has units of kilogram per second per charge coulomb.

$$B = \frac{\text{kg}}{\text{sec } coul}$$

[0005] The gradient of the magnetic field in the vertical direction z has units of

$$\frac{\text{dB}}{dz} = \frac{\text{kg}}{\text{sec } coul \text{ meter}}$$

[0006] This gradient interacting with a magnetic moment creates a force F measured in newtons.

$$F_z = \mu\frac{\text{dB}}{dz} = \frac{coul\ \text{m}^2}{\text{sec}}\frac{\text{kg}}{\text{sec } coul\ m} = \frac{\text{kg } m}{\text{sec}^2} = \text{newton}$$

[0007] In terms of vectors, the force is equal to

$$F=\nabla(\mu\cdot B)$$

which is the gradient ∇ of the dot product (·) of the magnetic moment with the magnetic field. This means that the magnetic moment has to be aligned with the field. The lift force on the spacecraft would then be the magnetic moment in the vertical z-direction μ_z times the magnetic field in the z-direction B_z. For constant magnetic moment, the gradient affects the magnetic field only, resulting in the same force equation

$$F_z=\nabla(\mu_z\cdot B_z)=\mu_z\nabla B_z$$

[0008] Referring to **FIG. 1**, the electric dipole has a positive charge q located on the z-axis at a distance a from the origin of the graph. A second negative charge −q is located at a distance −a from the origin. The positive charge produces an electrostatic potential ϕ_1 at a radius r_1 equal to the charge q divided by 4π times the permittivity of space ϵ_0

$$\varphi_1 = \frac{q}{4\pi\varepsilon_0}\frac{1}{r_1}$$

where the permittivity is linear capacitance, measured in farads per meter. The electrostatic potential has units of volts

$$\varphi = \frac{coul}{\left(\frac{\text{farad}}{\text{meter}}\right)}\frac{1}{\text{meter}} = \frac{coul}{\text{farad}} = \text{volt}$$

because the charge in coulombs held by a capacitor is equal to the capacitance, measured in farads, times the capacitor voltage. Because the second charge has the opposite sign, the potential ϕ_2 at a radius r_2 to the same point in space is

$$\varphi_2 = \frac{-q}{4\pi\varepsilon_0}\frac{1}{r_2}$$

[0009] The total potential ϕ at some point in space is equal to the sum of the two potentials, or

$$\varphi = \varphi_1 + \varphi_2 = \frac{q}{4\pi\varepsilon_0 r_1} - \frac{q}{4\pi\varepsilon_0 r_2} = \frac{q}{4\pi\varepsilon_0}\left[\frac{1}{r_1} - \frac{1}{r_2}\right]$$

[0010] As seen in the diagram, the point of space is a distance r from the origin. Using the law of cosines, radius r_1 can be written as

$$r_1 = (r^2 + a^2 - 2ar\cos(\theta))^{\frac{1}{2}} = r\left(1 + \left(\frac{a}{r}\right)^2 - 2\left(\frac{a}{r}\right)\cos(\theta)\right)^{\frac{1}{2}} = r(1 - 2xt + t^2)^{\frac{1}{2}}$$

where t is the ratio of the charge location over the radius, and x is cos(θ). The potential for positive charge q_1 can be written

$$\varphi_1 = \frac{q}{4\pi\varepsilon_0}\frac{1}{r_1} = \frac{q}{4\pi\varepsilon_0}\frac{1}{r}(1 - 2xt + t^2)^{-1/2}$$

[0011] Dropping the factor $q/4\pi\epsilon_0 r$, the square root can be expressed in terms of the Legendre polynomial P_n cos(θ) of the nth power

$$g(t, x) = (1 - 2xt + t^2)^{-1/2} = \sum_{n=0}^{\infty} P_n(x)t^n$$

where the absolute value of t is less than one. The polynomial coefficients of t^n can be obtained by using the binomial theorem to expand the generating function g(t,x) as

$$(1-2xt+t^2)^{-1/2}=\sum_{n=0}^{\infty}\frac{(2n)!}{2^{2n}(n!)^2}(2xt-t^2)^n$$

which evaluates to

$$\frac{0!}{2^0(0!)^2}(2xt-t^2)^0+\frac{2!}{2^2(1!)^2}(2xt-t^2)^1+\frac{4!}{2^4(2!)^2}(2xt-t^2)^2$$

$$1t^0+xt^1+\left(\frac{3}{2}x^2-\frac{1}{2}\right)t^2+\text{order}(t^3)$$

[0012] The first three Legendre polynomials are therefore

$$P_0=1$$

$$P_1=x$$

$$P_2=\frac{1}{2}(3x^2-1)$$

[0013] The electrostatic potential for both charges of the electric dipole is

$$\varphi=\frac{q}{4\pi\varepsilon_0}\frac{1}{r}\left\{\left(1-2\left(\frac{a}{r}\right)\cos(\theta)+\left(\frac{a}{r}\right)^2\right)^{-1/2}-\left(1+2\left(\frac{a}{r}\right)\cos(\theta)+\left(\frac{a}{r}\right)^2\right)^{-1/2}\right\}$$

[0014] The potential can be evaluated in terms of the Legendre polynomials as

$$\begin{aligned}\varphi&=\frac{q}{4\pi\varepsilon_0 r}\left[\sum_{n=0}^{\infty}P_n(\cos(\theta))\left(\frac{a}{r}\right)^n-\sum_{n=0}^{\infty}P_n(\cos(\theta))(-1)^n\left(\frac{a}{r}\right)^n\right]\\&=\frac{2q}{4\pi\varepsilon_0 r}\left[P_1(\cos(\theta))\left(\frac{a}{r}\right)+P_3(\cos(\theta))\left(\frac{a}{r}\right)^3+\dots\right]\end{aligned}$$

[0015] The first and most dominant term when the radius is much greater than location a is equal to

$$\varphi=\frac{2aq}{4\pi\varepsilon_0}\frac{P_1(\cos(\theta)0}{r^2}$$

which is the electric dipole potential and 2aq is the dipole moment

p=2aq

[0016] Now imagine a constant electric field E_0 which is perturbed by a conducting sphere of radius a. The unperturbed electrostatic potential outside the sphere would the negative of the electric field times the radius times the Legendre polynomial, or

$\Phi_1=-E_0rP_1$

[0017] The electrostatic potential perturbed by the charges is the voltage E_0a times the radius a times the a of the dipole moment times the Legendre polynomial divided by the radius squared

$$\varphi_2=E_0aaa\frac{P_1}{r^2}=E_0a^3\frac{P_1}{r^2}$$

[0018] The total potential outside the sphere is the sum of the two potentials equal to

$$\varphi=-E_0rP_1+E_0a^3\frac{P_1}{r^2}=-E_0P_1\left(r-\frac{a^3}{r^2}\right)=-E_0rP_1\left(1-\left(\frac{a}{r}\right)^3\right)$$

[0019] Referring to **FIG. 2**, the previously uniform electric field is shown perturbed by the neutral conducting sphere. The center of the sphere is taken as the origin and the z-axis is oriented parallel to the original uniform field.

[0020] The electric field induces a surface charge density σ on the sphere equal to the negative of the permittivity of space times the gradient of the electrostatic potential

$$\sigma=-\varepsilon_0\frac{\partial\varphi}{\partial r}|_{r=a}=3\varepsilon_0E_0\ \cos(\theta)$$

[0021] The electric field also induces an electric dipole moment on the sphere equal to the

$$p=\frac{qa}{r}\cdot\frac{2a^2}{r}=4\pi\varepsilon_0E_0a^3$$

with units of coulomb-meter. If this sphere is rotating around a central axis at a velocity v, it will create a magnetic moment μ equal to the dipole moment times the velocity.

μ=pv

with units of ampere-meter^2.

SUMMARY OF THE INVENTION

[0022] As shown in the preceding background section, a neutral conducting sphere placed in a uniform electric field will generate a magnetic moment when rotated around a central axis. The electric field can be created by two points charges of opposite sign separated by a distance between them.

[0023] Referring to **FIG. 3**, the spacecraft has a spherical cabin (A) to which are attached cone-shaped electrostatic towers (B,C) above and below the cabin along the direction of travel in the z-direction. Because the electric field goes from the positive charge to the negative charge, the tip of the lower tower has a positively charged electrode, and the upper tower has a negatively charged electrode. Four equally-spaced neutral conducting spheres (D) are connected to the cabin by non-conducting tubes (E). The tubes make an angle with the cabin such that the distance (CD) is greater than distance (DB). The angle θ of the tube with respect to the cabin can be seen in side view **FIG. 4**.

[**0024**] Referring to **FIG. 5**, the charges create a uniform spherical field between the towers. The conducting spheres perturb this field such that the electric field (E) points toward the upper tower in a manner similar to that shown previously in **FIG. 2**.

[**0025**] Referring to **FIG. 6**, the conducting sphere produces an electric dipole moment (A) pointing at an angle toward the upper tower.

[**0026**] Referring to **FIG. 7**, by the law of addition of vectors, the electric dipole $\bar{p}$ can be represented by two orthogonal vectors pointing in the vertical z-direction p_z and in the inward radial direction p_r.

[**0027**] Referring to **FIG. 8**, the hollow tube (A) connecting the cabin with the conducting sphere contains a spiral-wound electrical solenoid (B) which produces a magnetic field (C). This magnetic field B can be decomposed into two orthogonal vectors pointing in the vertical z-direction B_z and in the outward radial direction B_r as shown in **FIG. 9**.

[**0028**] Referring to top-view **FIG. 10**, the spacecraft has a clockwise angular velocity ω (A) which gives the conducting sphere a velocity v as shown by the vector (B). By the right-hand rule of physics, the angular velocity vector points in the negative z-direction. The angular velocity in the z-direction crossed with the radius r in the radial direction produces a velocity v in the clockwise θ-direction using cylindrical coordinates {r,θ,z}.

$$v_\theta = w_z \times r_r = -\omega r$$

[**0029**] Referring to **FIG. 11**, the negative radial dipole moment p_r crossed with the negative velocity v_θ of the sphere produces a positive magnetic moment μ_z in the z-direction.

$$\mu_z = p_r \times v_\theta = (-p_r)(-v_\theta) = pv$$

[**0030**] Referring to **FIG. 12**, the magnetic field B_z in the vertical z-direction is dotted with the magnetic moment μ_z in the z-direction to produce a force F_z in the vertical z-direction on each conducting sphere **(FIG. 13)**.

$$F_z = \nabla(\mu_z \cdot B_z) = \mu_z \nabla B_z$$

[**0031**] The magnetic field that is produced by the solenoid actually curves away and around. Thus there is a gradient of the field in the z-direction.

[**0032**] The force can also be expressed in tensor notation. The magnetic B field in the vertical direction is part of an electromagnetic 4×4 matrix Faraday tensor F

$$F^{\alpha}_{\beta} = \begin{array}{r|cccc|} & \beta = t & r & \theta & z \\ \alpha = t & 0 & 0 & 0 & 0 \\ \alpha = r & 0 & 0 & B_z & 0 \\ \alpha = \theta & 0 & -B_z & 0 & 0 \\ \alpha = z & 0 & 0 & 0 & 0 \end{array}$$

which shows that the magnetic field is located in slot $F^r{}_\theta$ of the Faraday tensor. In tensor notation the subscripts and superscripts have to match up on both sides of the equation. Matching subscripts and superscripts on the same side of the equation cancel. In this case, the electric dipole moment is in the radial direction p_r. The velocity can be represented as a time derivative of the θ-coordinate x^θ or

$$v^\theta = \frac{\partial x\theta}{\partial t}$$

[**0033**] Thus the force component in the z-direction becomes

$$F_z = p_r v^\theta F^r_{\theta,z} = (-p)(-v)\frac{\partial B_z}{\partial z} = pv\frac{\partial B_z}{\partial z}$$

where the angular and radial tensor components cancel and comma-z (, z) represents differentiation of the magnetic field in the z-direction.

[**0034**] The spacecraft design also has an inherent motion control system for moving in various directions. If the magnetic field of one solenoid arm is increased or decreased, the force on that sphere will be increased or decreased. Thus the spacecraft can turn in a particular direction.

A BRIEF DESCRIPTION OF THE DRAWINGS

[**0035**] **FIG. 1**. Electric dipole.

[**0036**] **FIG. 2**. Uniform electric field perturbed by electric dipole.

[**0037**] **FIG. 3**. Perspective view of spacecraft.

[**0038**] **FIG. 4**. Angle of solenoid tube.

[**0039**] **FIG. 5**. Electric field perturbed by conducting sphere.

[**0040**] **FIG. 6**. Electric dipole generated by conducting sphere.

[**0041**] **FIG. 7**. Orthogonal vector components of electric dipole.

[**0042**] **FIG. 8**. Magnetic field produced by tube arm solenoid.

[**0043**] **FIG. 9**. Orthogonal vector components of magnetic field.

[**0044**] **FIG. 10**. Angular velocity of hull.

[**0045**] **FIG. 11**. Magnetic moment produced by radial electric dipole and sphere velocity.

[**0046**] **FIG. 12**. Dot product of the magnetic moment with the magnetic field.

[**0047**] **FIG. 13**. Vertical lift force on all four conducting spheres.

[**0048**] **FIG. 14**. Perpective view of spacecraft interior.

DETAILED DESCRIPTION OF THE INVENTION

[**0049**] 1. Referring to the cut-away view **FIG. 14**, the construction of the spacecraft is a thin-wall insulating thermoplastic having a dielectric constant in the range of 20 kilovolts per millimeter (A). An insulated electrode (B) runs from the cabin power supply and high-voltage transformer

(C) to the tip of each tower (D). The four spheres (E) are silver plated to make them conducting. The tube solenoids (F) are driven by a direct current power supply (G).

[0050] 2. The present model uses 3D computer design software and stereolithography fabrication techniques to create the thin-wall, low-weight, hollow structure of the hull. The computer model is sliced into many thin horizontal slices. A laser, mounted on an x-y table, draws out the slice on a table immersed in a bath of liquid polymer. Due to its sensitivity to the light, the liquid polymerizes. The table is then lowered a few thousandths of an inch more and the process is repeated. Thus making hollow spherical and conical shapes is extremely easy to do. Parts can be designed and stored in *.STL stereolithography files for transmission by Internet e-mail to the service bureau machine shop which sends the finished parts back the next day by express mail.

I claim:

1. A spacecraft comprising:

a. a spherical cabin;

b. an electrostatic conical tower mounted on top of item (**1***a*), supporting a vertically-mounted negatively-charged insulated electrode at the tip of the tower;

c. an electrostatic conical tower mounted on the bottom of item (**1***a*), supporting a vertically-mounted positively-charged insulated electrode at the tip of the tower;

d. a vertical electric dipole created by items (**1***b*) and (**1***c*);

e. a high-voltage transformer to drive item (**1***d*), mounted in item (**1***a*);

f. four tubular arms, mounted at 90° around and extending at an angle from item (**1***a*);

g. four solenoids, each of which is mounted axially inside item (**1***f*);

h. a direct current power supply to drive item (**1***g*);

i. four silver-plated conducting spheres, each of which is mounted on the end of item (**1***f*);

2. an electrostatic lift system that:

a. produces a uniform spherical electric field by means of item (**1***d*) which envelopes item (**1***i*);

b. produces a perturbed electric field due to the presence of item (**1***i*);

c. produces an electric dipole moment in the direction of item (**1***b*) due to items (**2***a*) and (**2***b*);

d. produces a vertical magnetic moment due to the clockwise angular velocity of item (**1***a*) combined with item (**2***c*);

e. produces a vertical lift force on item (**1***i*) due to item (**2***d*) combined with the magnetic field gradient in the vertical direction produced by item (**1***g*); and

f. creates a motion control system by varying the current to item (**1***g*) in order to increase or decrease the effect of item (**2***e*) on a particular item (**1***i*).

* * * * *

Nov. 12, 1957

L. R. CRUMP

POWERING ELECTRICAL DEVICES WITH ENERGY
ABSTRACTED FROM THE ATMOSPHERE

Filed March 12, 1954

2,813,242

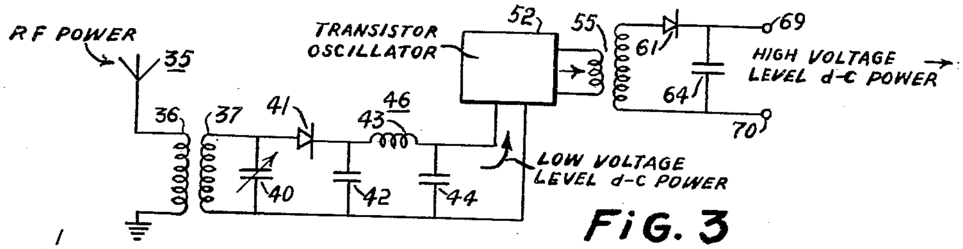
RF POWER
35
36
37
41
43
46
40
42
44
TRANSISTOR OSCILLATOR
52
55
61
64
69
70
HIGH VOLTAGE LEVEL d-c POWER
LOW VOLTAGE LEVEL d-c POWER

FiG. 3

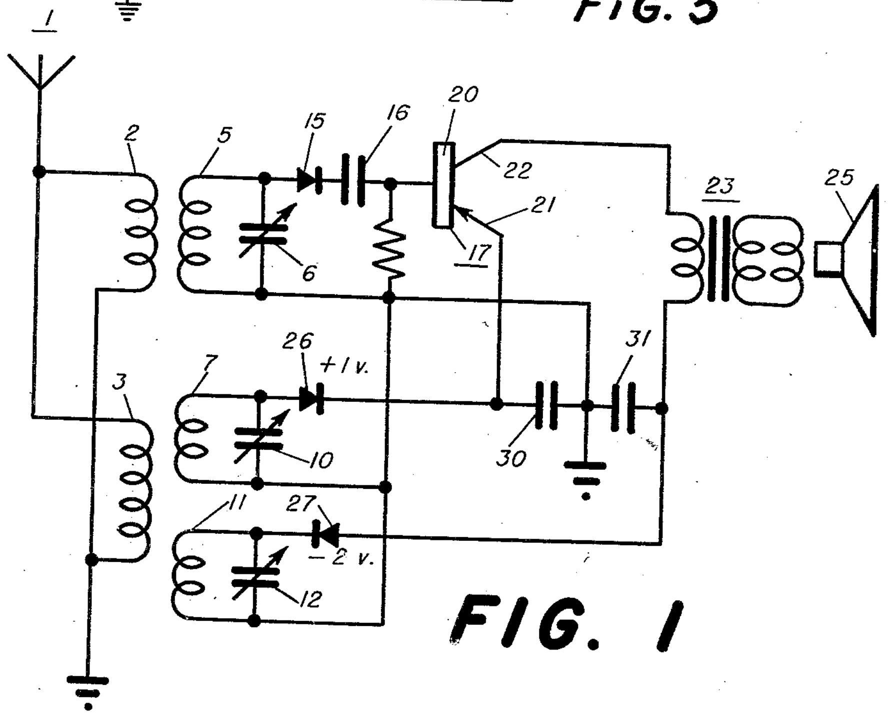
1
2
5
15
16
20
22
21
17
6
23
25
3
7
26
+1 v.
10
31
30
11
27
−2 v.
12

FIG. 1

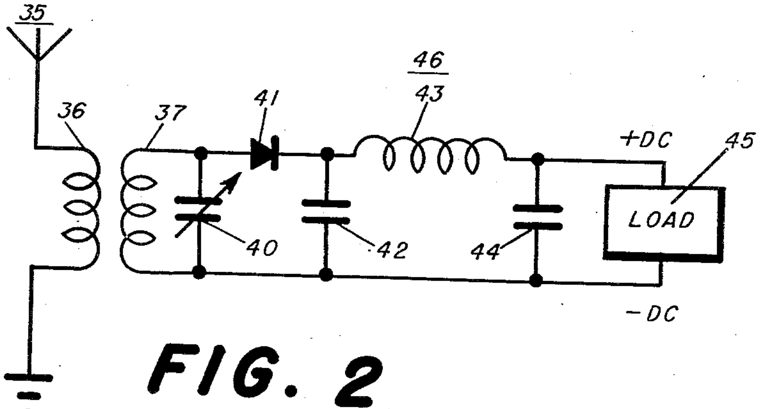
35
36
37
41
46
43
40
42
44
+DC
45
LOAD
−DC

FIG. 2

INVENTOR
Lloyd R. Crump

BY W.E. Thibodeau + A.W. Dew

ATTORNEYS

United States Patent Office

2,813,242

Patented Nov. 12, 1957

2,813,242

POWERING ELECTRICAL DEVICES WITH ENERGY ABSTRACTED FROM THE ATMOSPHERE

Lloyd R. Crump, Silver Spring, Md.

Application March 12, 1954, Serial No. 415,986

1 Claim. (Cl. 321—2)

(Granted under Title 35, U. S. Code (1952), sec. 266)

The invention described herein may be manufactured and used by or for the Government for governmental purposes without payment to me of any royalty thereon.

This invention relates to the convenient and economical provision of power for the operation of electronic circuits and devices using transistors, and of other electrical devices having modest power requirements.

A great advantage of transistors, and a major reason for their enthusiastic reception since their introduction a few years ago, is the fact that they will operate satisfactorily with very low supply voltages and currents. One milliwatt or even less is sufficient to power a transistor in many applications. Various batteries have been developed to provide, in a minimum of space, the relatively minute amounts of power needed by transistors.

My invention provides methods and means that permit transistor circuits, and also other low-powered electrical devices, to be economically and conveniently operated without any batteries whatever, and indeed without any power supply whatever as power supplies are ordinarily conceived.

The invention centers around my discovery that it is practicable to construct operative transistor circuits that are able to abstract from the atmosphere sufficient electromagnetic energy to provide all necessary supply voltages and currents for their own operation. Circuits and devices powered according to my invention will operate indefinitely without any local power source whatever.

I have successfully constructed and demonstrated such circuits. For example, I have constructed a batteryless transistor radio receiver on which I have listened to either nearby or distant broadcast stations as desired, using either headphones or a loudspeaker; this receiver has been powered entirely by electromagnetic energy abstracted from the atmosphere.

From the successful operation of this receiver, and from other experimental work, it becomes clear that, by the methods and means of the invention, a great variety of practical and useful transistor circuits can be powered entirely by energy abstracted from the atmosphere.

Furthermore, as will become apparent below, my invention is applicable to the powering of other electrical devices requiring relatively small amounts of power.

An object of the present invention is to provide methods and means for powering transistor circuits entirely from radiofrequency energy abstracted from the atmosphere.

Another object is to provide methods and means for powering remote radio receivers, low-powered radio transmitters, and other low-powered electrical devices, with energy received by radio from a master station, so that no local power supplies are needed by the devices and so that the powering or non-powering of the remote device is under the control of the master station.

A further object is to provide methods and means for powering transistor circuits and other low-powered electrical devices with radiofrequency energy received from one or more remote radio transmitters.

Other objects, aspects, uses, and advantages of the invention will become apparent from the following description and from the drawing.

Figure 1 is a schematic diagram of a transistor radio receiver in which all necessary power is supplied by energy abstracted from the atmosphere in accordance with the invention.

Figure 2 is a schematic diagram showing a general application of the invention to provide direct-current power to a load.

Figure 3 is a schematic diagram of a system for obtaining a high energy D.-C. source at a high voltage level using energy abstracted from the atmosphere.

Referring to Figure 1, a receiving antenna **1** is connected to antenna coupling coils **2** and **3**, the other ends of which are connected to ground. A parallel resonant circuit consisting of coil **5** and capacitor **6** is coupled to coil **2**. A second parallel resonant circuit consisting of coil **7** and capacitor **10** is coupled to coil **3**. A third parallel resonant circuit consisting of coil **11** and capacitor **12** is also coupled to coil **3**.

Coil **5** and capacitor **6** are tuned to the frequency of a radio transmitter from which it is desired to receive information—for instance, an amplitude-modulated standard broadcast station. The signal received from this transmitter need not be strong. The signal is detected by diode **15** to obtain an audio-frequency information signal. This audio signal is coupled through a capacitor **16** and is amplified by a circuit that includes a transistor **17** having a base **20**, an emitter **21**, and a collector **22**. The amplified audio output of the transistor is coupled through an audio transformer **23** to an electroacoustical transducer, preferably a permanent-magnet dynamic loudspeaker **25** as shown.

The novelty of the invention lies largely in the method and means by which the necessary direct-current power is supplied to the emitter and collector circuits of transistor **17**. This method and means will now be described.

Coil **7** and capacitor **10**, and also coil **11** and capacitor **12**, are tuned to receive radio signals of relatively high strength. It does not matter whether these signals contain information. These power signals are rectified by diodes **26** and **27** to provide direct-current power that is filtered by capacitors **30** and **31**. The D.-C. power thus obtained is utilized to power the transistor **17**.

In the circuit shown, two tuned circuits (coil **7** and capacitor **10**, and coil **11** and capacitor **12**) are tuned to power signals and the D.-C. voltages obtained from each are connected in series. The tuned power circuits may be tuned to the same or different power signals. Under certain circumstances it may be desirable to use more than two tuned power circuits and to tune them to more than two power signals; in this way power can be obtained from several signals and combined. On the other hand, if a strong power signal is available, a single tuned power circuit may suffice to give the needed D.-C. power.

Even weak information signals can be received successfully. A plurality of transistor amplifier stages can be used if desired, or other circuits such as superheterodyne circuits can be used. It is merely necessary that a sufficiently strong power signal or signals be available to provide the small amount of D.-C. needed to power the transistors.

If the information signal happens to be strong, it can be used as the power signal; all of the tuned circuits (coil **5** and capacitor **6**, coil **7** and capacitor **10**, coil **11** and capacitor **12**) are tuned to the information signal.

Engineers who have observed my invention in operation have been surprised at the unexpectedly good results obtained, even with readily available power signals of quite moderate strength. For instance, sufficient power for satisfactory operation of a loudspeaker at low volumes is

readily obtained from a 5-kilowatt standard broadcast station 5 miles away, using only an indoor antenna to pick up the power signal as well as information signals. In typical operation under these conditions a D.-C. voltage of about 2.5 to 3 volts is obtained between the emitter and the collector, at a current of about 250 microamperes; D.-C. power input to the transistor is thus of the order of 0.5 to 1 milliwatt. So far as I am aware, no one has ever before discovered and demonstrated the practicability of this method of powering a radio receiver.

Because existing broadcast stations within a radius of a number of miles provide adequate power signals, the invention is readily practicable with existing power signals in almost any location in or near any city in the United States.

Although I have described a transistor radio receiver powered by my invention, it will be readily apparent that the invention is applicable to the powering of any transistor circuit using one or a number of transistors, and to the powering of other devices requiring relatively small amounts of power. For instance, sensitive electromechanical, electrochemical, or electrothermal devices can be operated by the method of the invention.

Referring to Figure 2, which shows a more general embodiment of my invention, an antenna **35** picks up radiofrequency energy from the atmosphere. This energy flows through coil **36**, which is coupled to a tuned circuit consisting of coil **37** and capacitor **40**. The radiofrequency voltage across capacitor **40** is rectified by diode **41** and filtered by a low-pass filter **46** consisting of capacitors **42** and **44** and choke coil **43**. The resulting D.-C. voltage is applied to a load **45**.

In the practice of my invention, larger amounts of power can be obtained for short periods of time by storing received energy in a suitable energy storage device. Stored energy may then be withdrawn at intervals at a more rapid rate than that at which it was received and put into the storage device. In this way the invention can be used to provide short pulses of relatively very high electrical energy. This result can be readily obtained by charging a relatively large capacitor with direct current and then discharging the capacitor rapidly into a load when desired. This rapid discharge can be initiated automatically when the voltage across the capacitor reaches a certain level, or it can be initiated when a transistor radio receiver receives a certain information signal.

Higher voltages can be obtained with the invention by means of well known devices for raising D.-C. voltages as shown in Figure 3. The D.-C. voltage output from the capacitor **44** can be used to power a low frequency transistor oscillator **52** whose A.-C. output is raised to a higher voltage level by the transformer **55**. This relatively high A.-C. voltage can then be rectified by a diode **61** and fed to a capacitor **64** to provide a high energy D.-C. source at a relatively high voltage level at the terminals **69** and **70**. If desired, energy can now be withdrawn from the capacitor **64** at intervals in short pulses of high energy at a high voltage level. Pulsed radio transmission is one of the possible uses for this form of the invention. Other uses would be to provide a single relatively powerful pulse needed to actuate an electrothermal or electromechanical device.

As has been indicated above, in many locations and particularly anywhere in or near most American cities, power signals normally present in the atmosphere are readily available for the easy and convenient practice of the invention. However, the invention also has important applications in systems in which the necessary power signal is generated and transmitted specifically for the operation of the particular system. Such systems can, for example, comprise a master station transmitting all the power that is needed for hundreds or thousands of fixed or mobile transistor receivers or other remote devices over a range of many miles. This eliminates the need for hundreds or thousands, as the case may be, of local power supplies. At the same time, such a system has the advantage that all of the remote devices can be simultaneously activated or deactivated at the will of the master station, simply by starting or stopping the transmission of the power signal. In such systems it will often be advantageous to use power signals of frequencies sufficiently high to permit the use of resonant receiving antennas of small physical dimensions for signal pickup at the remote devices. In addition to the power signal, the master station may transmit an information signal on the same or a different carrier.

Certain types of devices powered entirely by received radio waves are of course well known. The well-known "crystal set" of the early days of radio, which used a diode rectifier to demodulate an amplitude-modulated radiofrequency signal, is an outstanding example of such a device. My invention is readily distinguishable from such prior devices, however. In typical prior devices a modulated radiofrequency signal is applied to a diode to obtain unidirectional half-wave pulses whose amplitudes vary with modulation. These pulses are integrated by means of a capacitor to obtain a unidirectional signal the amplitude of which follows the audiofrequency modulation envelope. If the radiofrequency signal is received with sufficient strength the audio signal may have sufficient power to operate headphones or similar utilization device without power amplification; but the signal is utilized for its information content, rather than to supply non-information-containing power.

My invention, on the other hand, entails the utilization of received radiofrequency energy to supply power to at least one pair of circuit points (across capacitor **31** in Fig. 1, for example), such circuit points requiring power solely for its power content and not for any information or modulation it may contain. In other words, my invention entails the utilization of radiofrequency energy to supply power that would otherwise have to be supplied by batteries, generator, or other local power source.

It will be apparent that the embodiments shown are only exemplary and that various modifications can be made in construction and arrangement within the scope of the invention as defined in the appended claim.

I claim:

An electrical device for obtaining a high energy D.-C. source at a high voltage level using energy abstracted from the atmosphere, said device comprising in combination: resonant means for receiving radio waves, first rectifier means for converting said radio waves into first direct current energy, first capacitor means for storing said first direct current energy, an oscillator powered by said direct current energy, said oscillator producing an A.-C. output, transformer means for raising said A.-C. output to an increased voltage level, second rectifier means for converting the A.-C. output of increased voltage level from said transformer into second direct current energy, and second capacitor means for storing said second direct current energy.

References Cited in the file of this patent

UNITED STATES PATENTS

1,540,998	Plauson	June 9, 1925
2,056,824	Cawley	Oct. 6, 1936
2,553,734	Adler	May 22, 1951
2,569,687	Pan	Oct. 2, 1951
2,647,957	Mallinckrodt	Aug. 4, 1953

OTHER REFERENCES

Turner: "A crystal receiver with transistor amplifier," Radio and Television News, January 1950.

US006404089B1

(12) **United States Patent**
Tomion

(10) **Patent No.:** **US 6,404,089 B1**
(45) **Date of Patent:** **Jun. 11, 2002**

(54) **ELECTRODYNAMIC FIELD GENERATOR**

(76) Inventor: **Mark R. Tomion**, 2938 Ferguson Corners Rd., Geneva, NY (US) 14456

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/621,152**

(22) Filed: **Jul. 21, 2000**

(51) **Int. Cl.**^{7} **H02K 1/00**; H02K 5/00

(52) **U.S. Cl.** .... **310/162**; 290/1 A; 290/1 R; 244/23 A; 244/23 C

(58) **Field of Search** .... 310/162–168, 310/171, 112, 308–309, 74, 68 R, 90, 118, 216, 261, 254, 89, 258; 290/1 R, 1 A, 49; 244/1 N, 130, 62, 23 C, 205, 23 A

(56) **References Cited**

U.S. PATENT DOCUMENTS

1,889,208	A	*	11/1932	Masterson et al.	310/126
2,949,550	A		8/1960	Brown	
3,071,705	A		1/1963	Coleman et al.	
3,177,654	A		4/1965	Gradecak	
3,464,207	A		9/1969	Okress	
3,620,484	A	*	11/1971	Schoppe	244/130
3,662,554	A	*	5/1972	Broqueville	244/1 N
3,774,865	A	*	11/1973	Pinto	244/23 C
4,733,099	A		3/1988	Hutson, Jr.	
4,789,801	A		12/1988	Lee	
4,900,965	A		2/1990	Fisher	
5,291,734	A	*	3/1994	Sohnly	60/202
5,382,833	A		1/1995	Wirges	

FOREIGN PATENT DOCUMENTS

GB	2312709	5/1997
WO	WO85/03053	7/1985

* cited by examiner

Primary Examiner—Tran Nguyen
(74) *Attorney, Agent, or Firm*—Stephen B. Salai, Esq.; Brian B. Shaw, Esq.; Harter, Secrest & Emery LLP

(57) **ABSTRACT**

This device is a brushless high-voltage electrical generator, requiring suitable means of input rotary torque, for purposes of producing a very-high-energy external electrodynamic field or continuous quasi-coherent DC corona or arc discharge of uniform current density which completely encloses the machine's conductive housing. This housing is divided into distinct electrical sections and contains a flat conductive rotor which electrically links separate negative and positive housing sections and upon which a plurality of toroidal generating coils are rotatably mounted. Circular arrays of stationary permanent magnets are affixed within the housing which induce a constant DC voltage within said coils upon their rotation. The primary voltage so-generated is electrostatically impressed across the rotor such that great quantities of electronic charge may be transported between the opposite polarity sections of the housing, in such a manner that a much higher secondary voltage is caused to appear across interposed neutral sections thereof, and the resulting external breakdown current once initiated is independent of the generating coils' ampacity. Ancillary mechanical, electrical, an/or electronic features may be attached upon or within the housing to aid in harnessing and controlling the useful effects associated with the external dynamic electric field produced by the device.

48 Claims, 9 Drawing Sheets

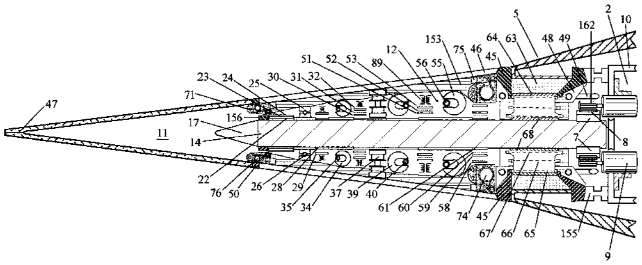

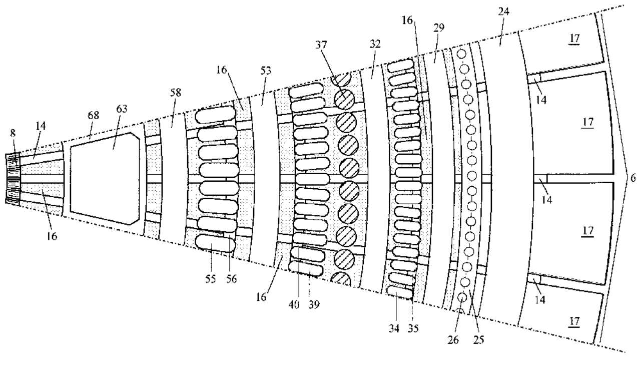

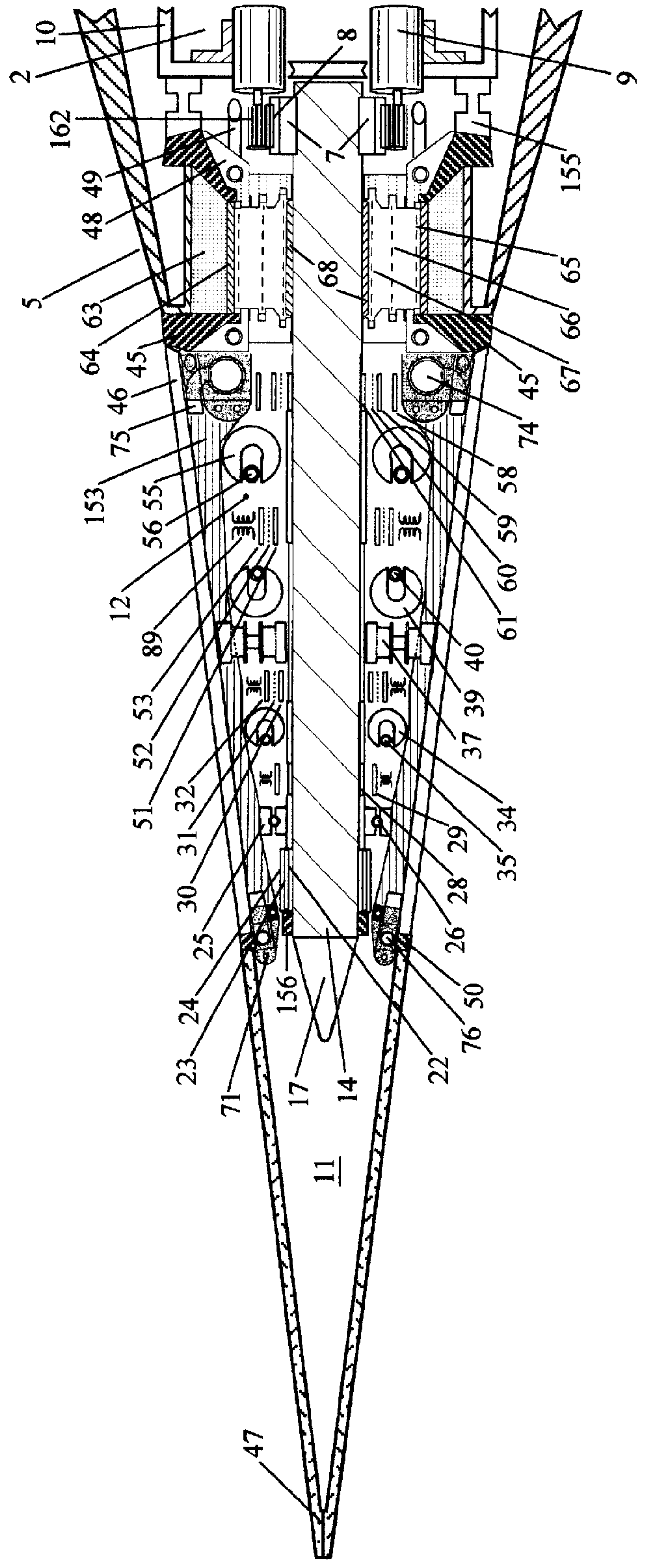

Fig. 1

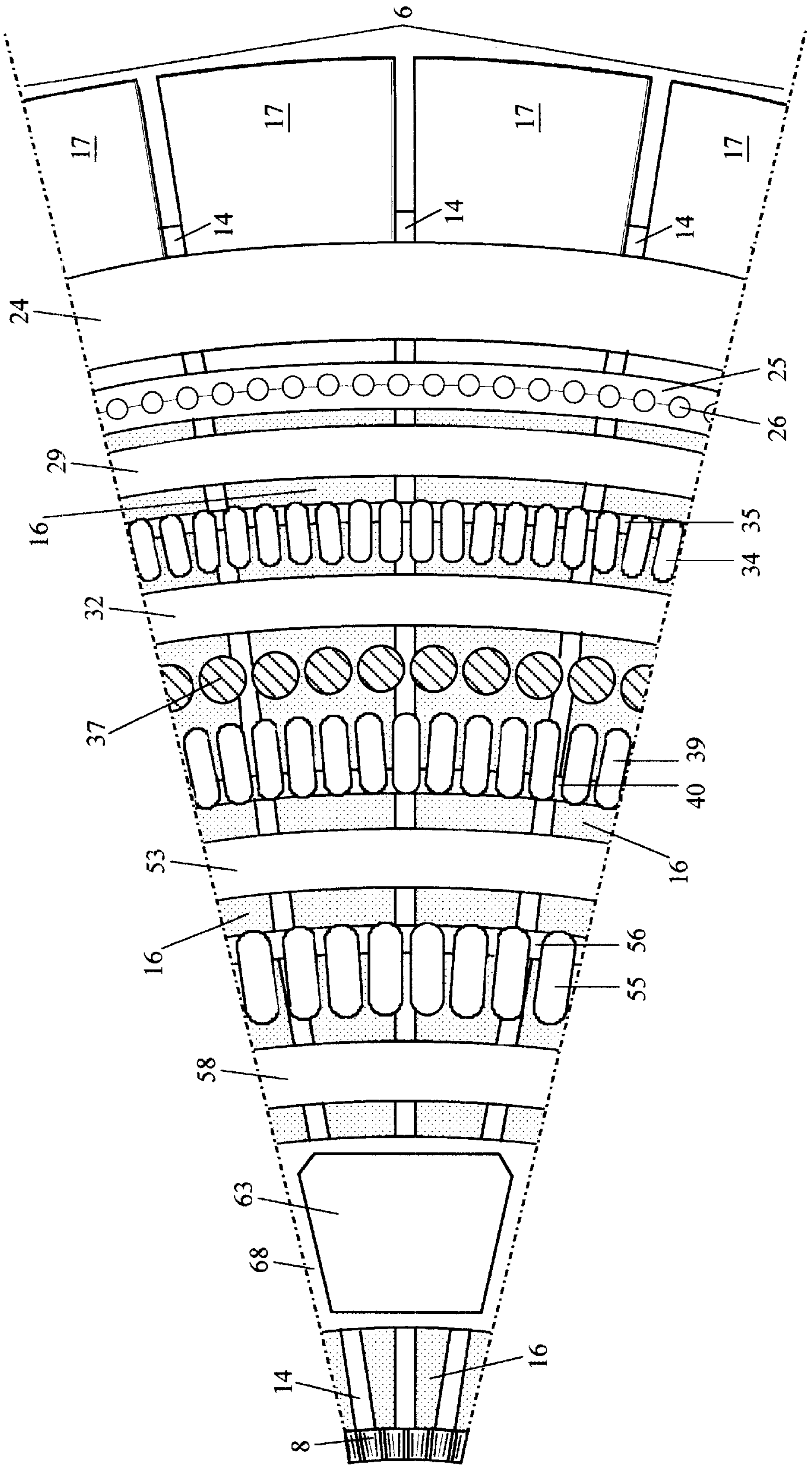

Fig. 2

Fig. 3

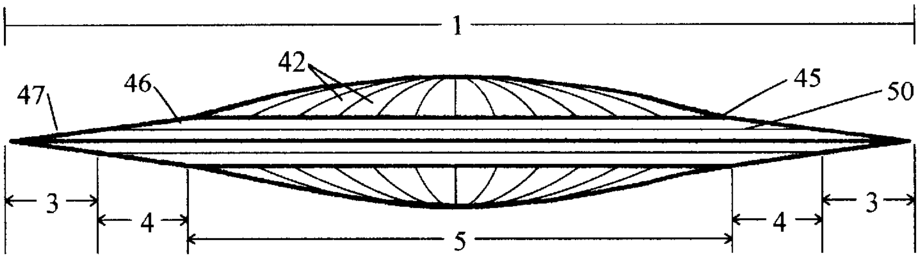
1
42
47
46
45
50
3
4
5
4
3

Fig. 4

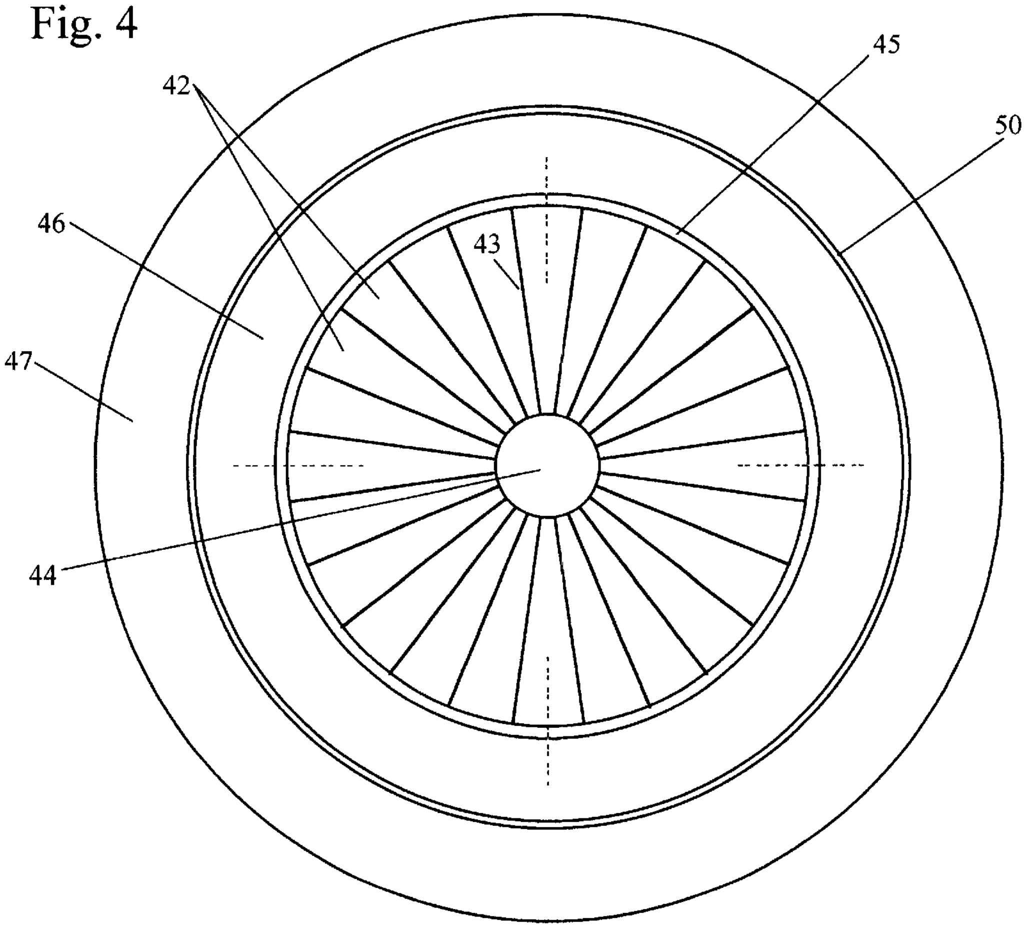
45
42
50
46
43
47
44

Fig. 5

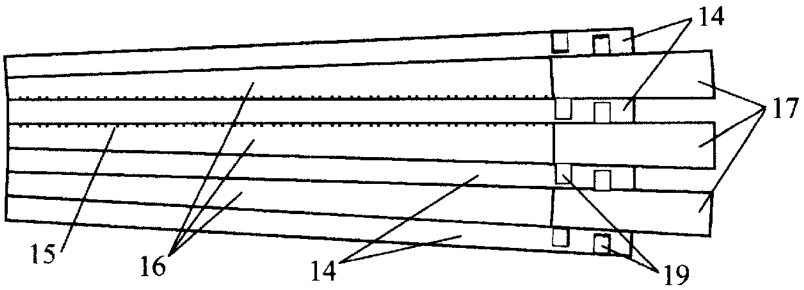
14
17
15
16
14
19

Fig. 6a

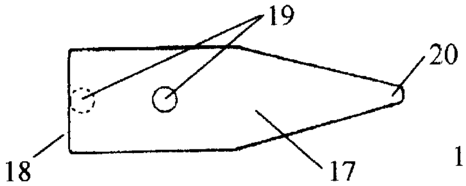

Fig. 6b

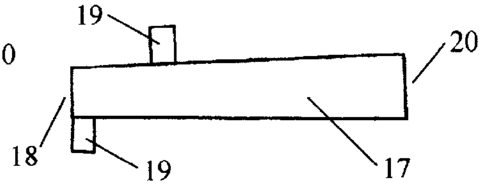
19
20
18
19
17

Fig. 7

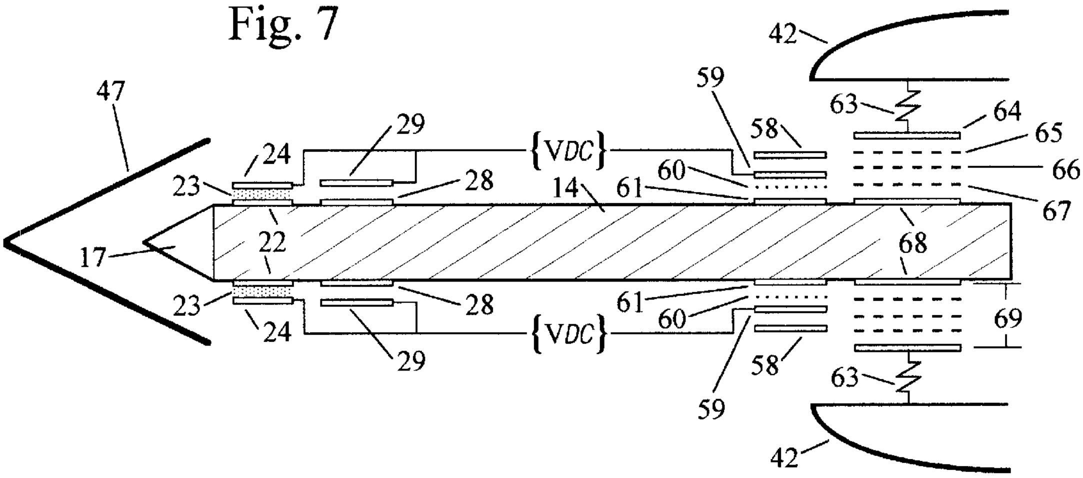
42
47
59
63
64
58
65
29
{VDC}
24
66
60
23
28
14
61
67
22
68
17
22
68
23
28
61
69
60
24
{VDC}
29
58
63
59
42

Fig. 8

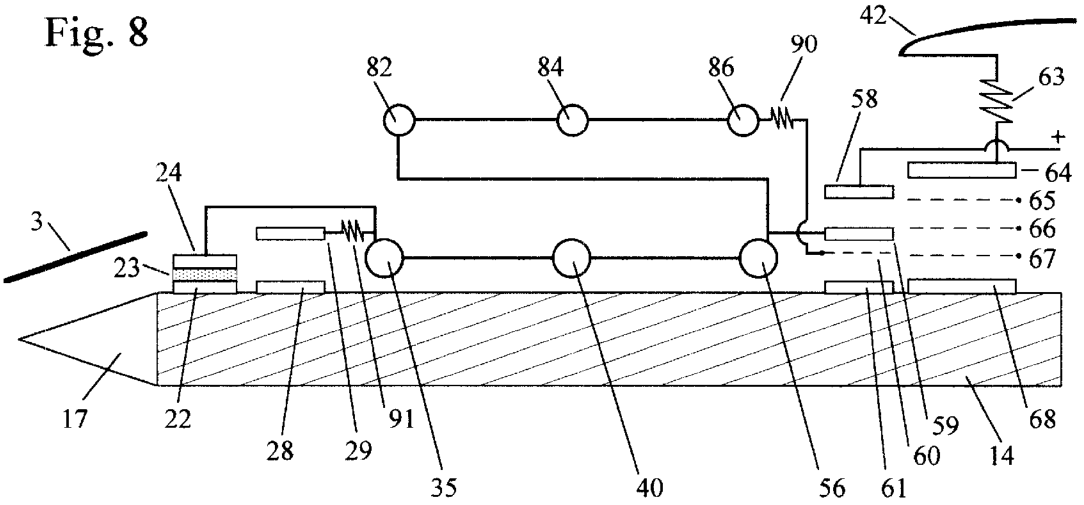
+

Fig. 9

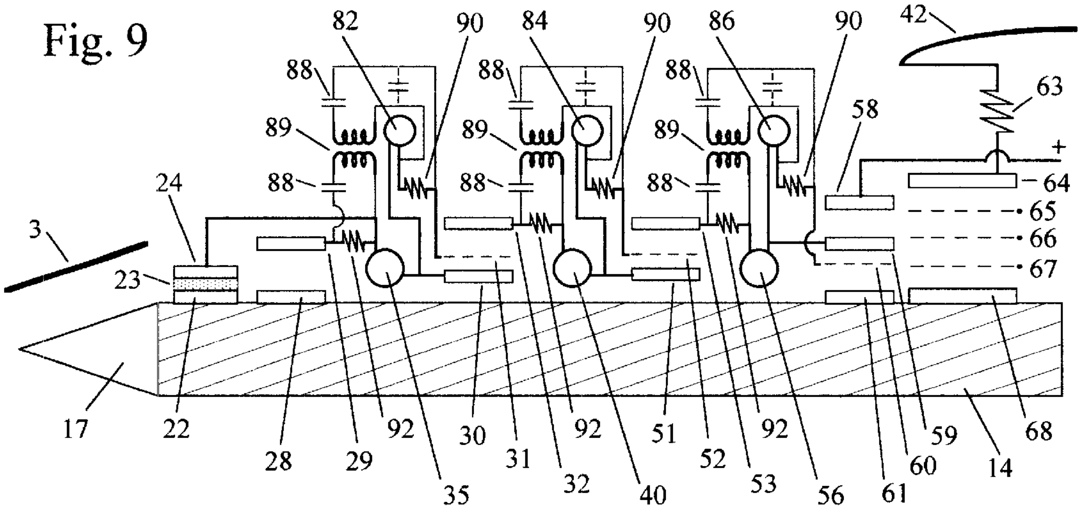
+

Fig. 10

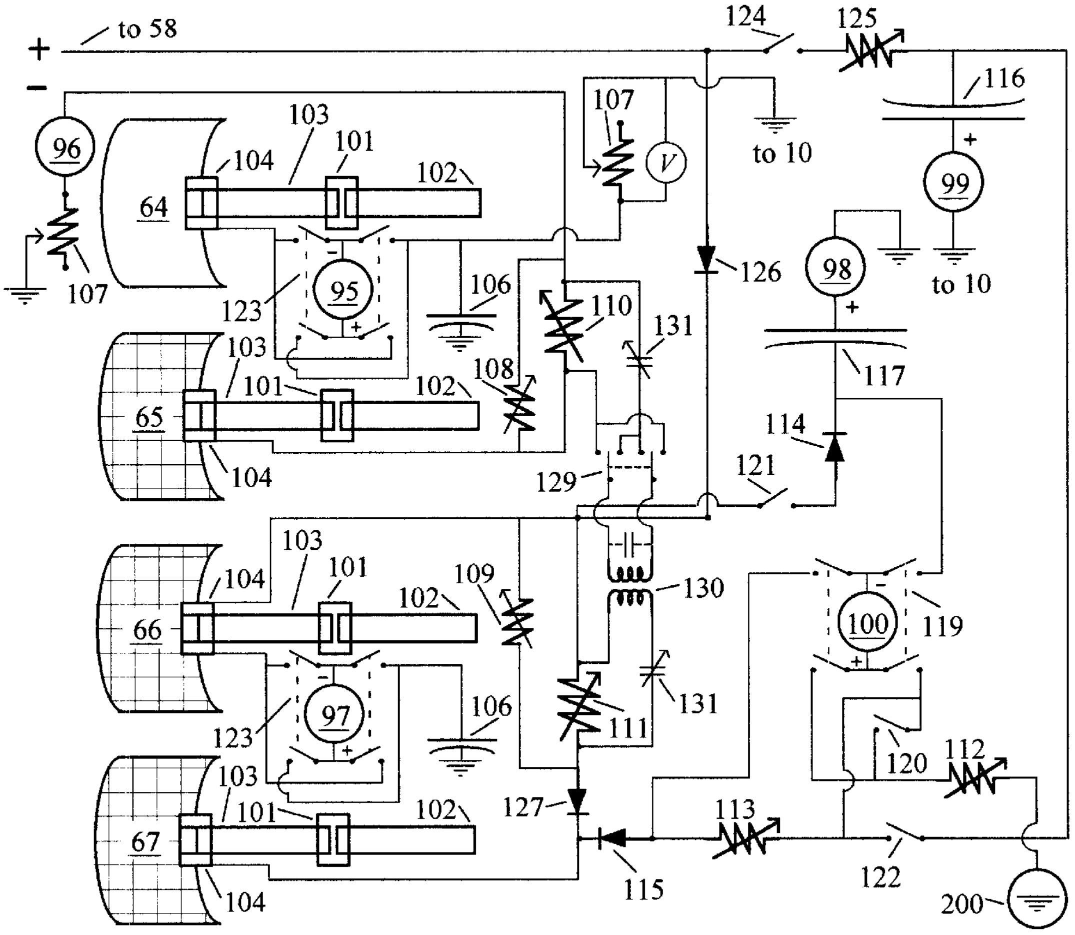
+
–
to 58
V
to 10
to 10

Fig. 11

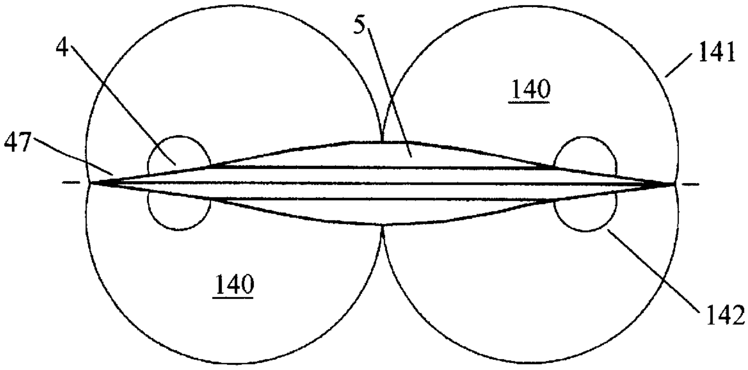
5
4
141
140
47
140
142

Fig. 12

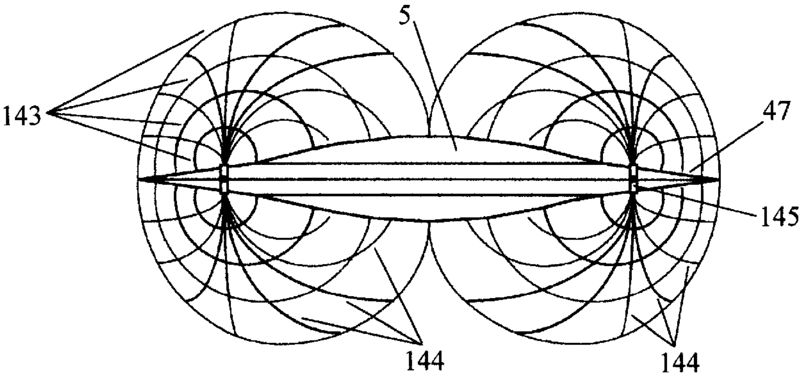
5
143
47
145
144
144

Fig. 13

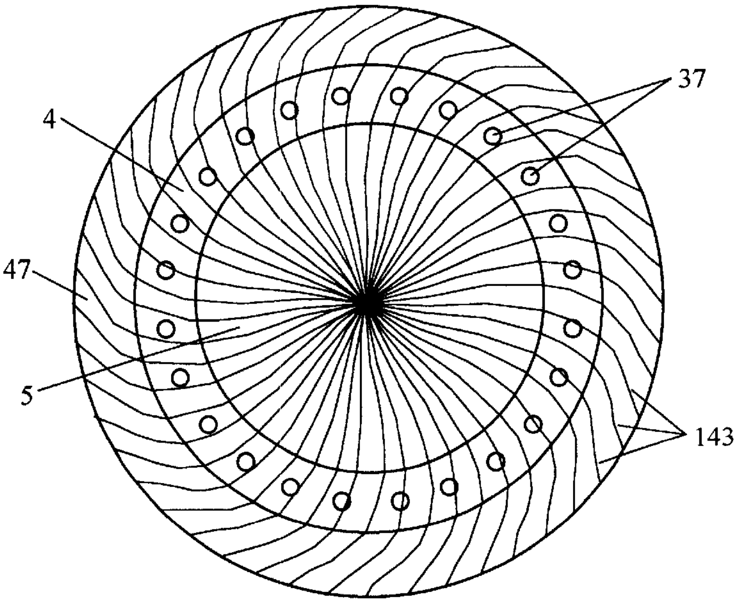

Fig. 14

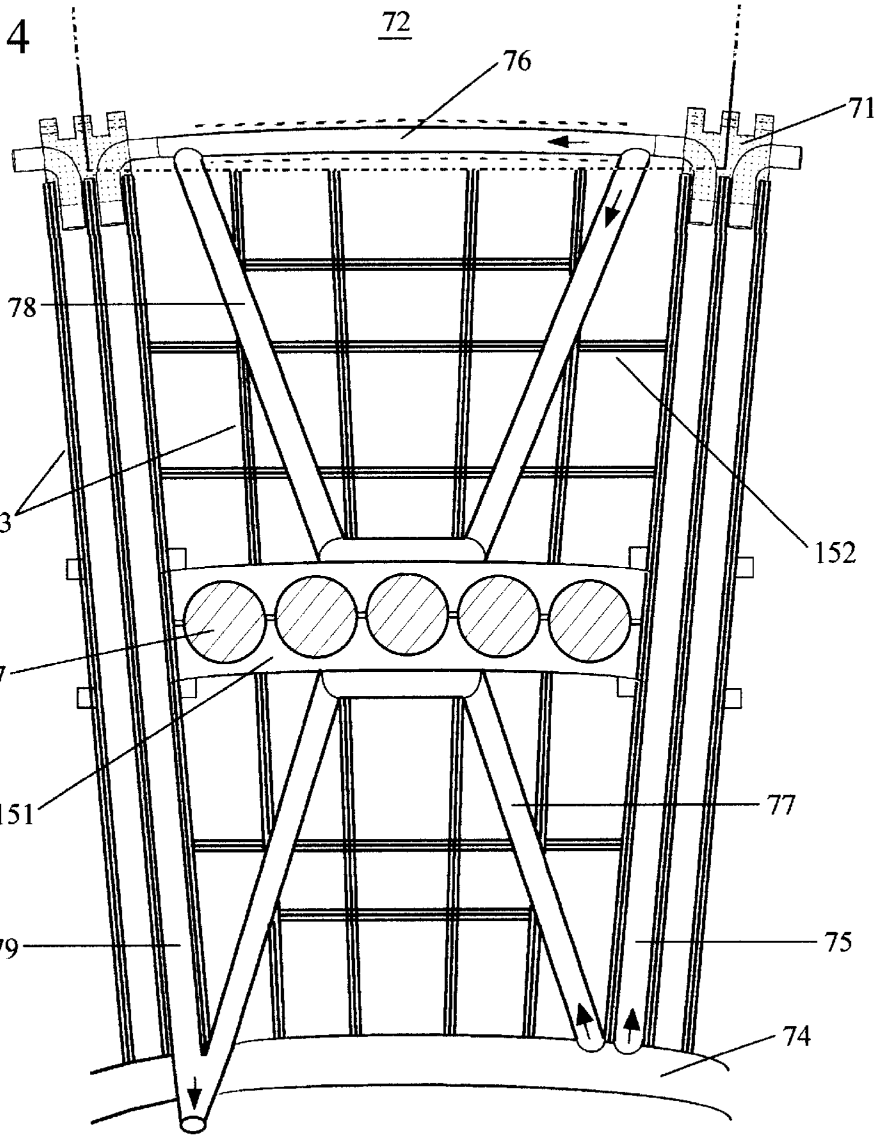

Fig. 15

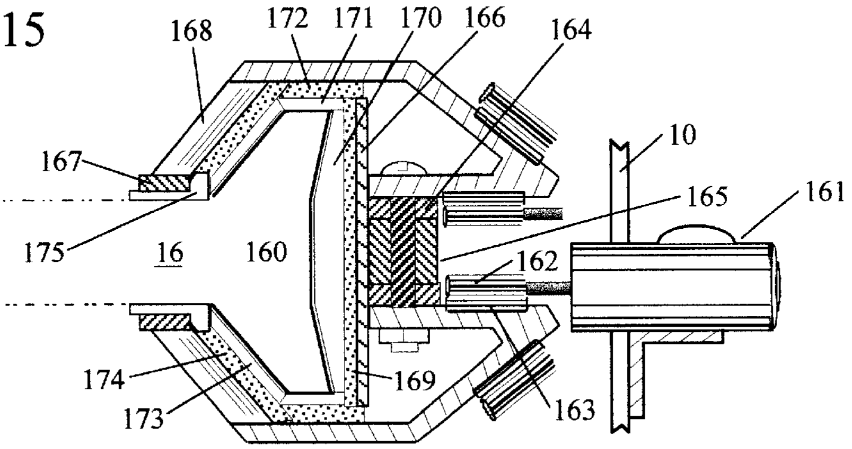

Fig. 16

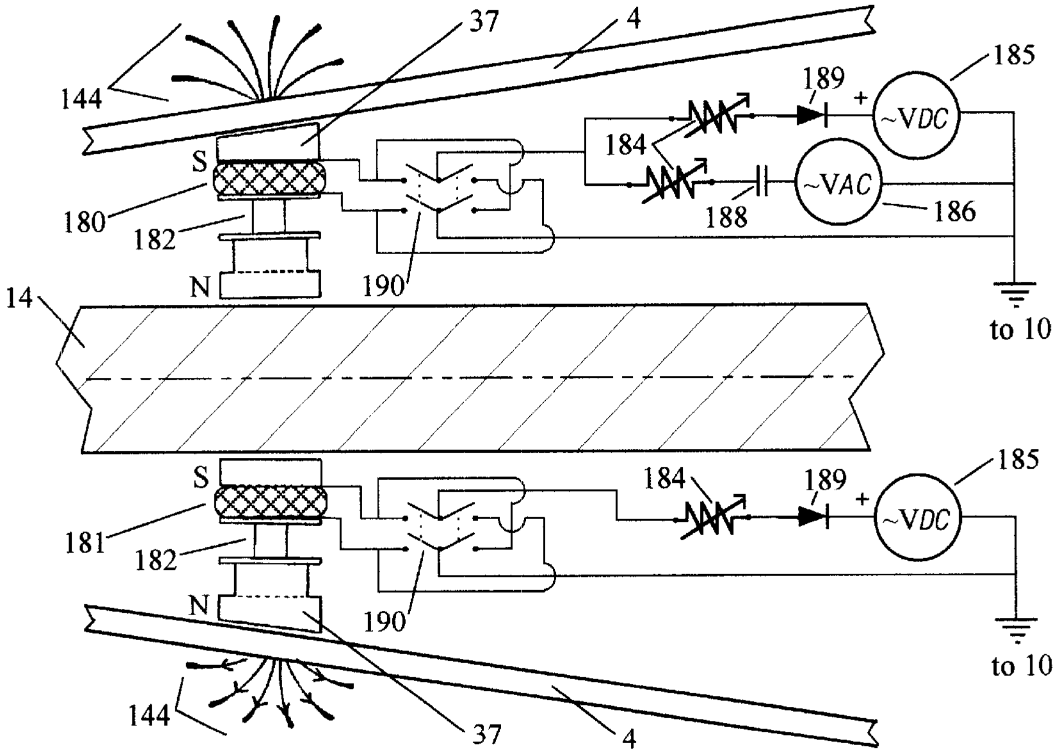
+
~VDC
S
~VAC
N
to 10
S
+
~VDC
N
to 10

ELECTRODYNAMIC FIELD GENERATOR

FIELD OF THE INVENTION

The present invention is related in general to rotary direct current (DC) electrical generators which incorporate permanent magnet fields, and in particular to similar electromechanical generating devices whose principal purpose is to produce a useful continuous high-voltage DC corona discharge. This invention also pertains to the field of aerospace vehicles which are capable of propulsion through the utilization of an ultra-high-voltage corona or arc discharge of a special type and form.

DESCRIPTION OF RELATED ART

In order to provide background information so that the present invention may be completely understood and appreciated in its proper context, reference is made hereinbelow to a number of related art patents. These cited references contain certain similarities to the present invention, primarily related to the objective of producing electrically-developed thrust. However, key differences and limitations with regard to achieving this objective are in evidence.

The related art disclosed in U.S. Pat. No. 2,949,550 proposes three attendant objects which are quite similar to certain of the objects of the present invention: (i) to provide an apparatus for the direct conversion of electrical potential to usable kinetic energy; (ii) to provide such an apparatus having a hollow body or housing which contains a source of high electrical potential; and (iii) to provide a self-propelling vehicular apparatus which includes a pair of electrically conductive body portions joined by an insulative portion, whereby said conductive portions constitute electrodes. A further similarity is that the vehicular embodiment proposed preferably be of a circular disc shape somewhat thicker in its center than at its edges. The device set forth in this Patent just cited is noticeably different from the present invention, however, in that a positive voltage is applied to the housing periphery and a negative voltage is applied at the central axis. The proposed arrangement is furthermore limited to operation within a gaseous dielectric medium, upon which it is reliant for the production of the claimed motive force, and is to be operated at a potential gradient less than that which would produce a visible corona. This device moreover makes no provisions for the production and/or extraction of vast quantities of useful thermal kinetic energy, as the present invention does.

The related art disclosed in U.S. Pat. No. 3,071,705 is based on three empirical principles of electrostatics which also figure prominently in explaining the actual shape of the discharge current field that will be produced by the present invention: (i) electrostatic lines of force tend to concentrate on the surface of a charged conductor in those places with the smallest radius of curvature; (ii) they are normal to the surface from which they emanate; and (iii) they do not cross but will bend under the influence of another charged body. The device set forth in this Patent just cited is in fact quite similar to that of the Patent cited immediately before, in that relatively massive positive atmospheric ions and entrained neutral air molecules are repelled from the positive electrode toward the negative electrode along a linear axis—thereby inducing an air flow or "electric wind" which represents the fundamental source of thrust. It is therefore once again limited to operation in a gaseous dielectric medium, and does not make use of a flow of negative and greatly less massive electron current in the course of thrust production—which may achieve comparatively enormous acceleration and even relativistic velocity and specific impulse at a similar given level of applied potential gradient—as the present invention does.

The related art disclosed in U.S. Pat. No. 3,177,654 actually addresses the issue of producing electrically-developed thrust which is not limited to atmospheric operation, and the stated primary object thereof is to provide a propulsion system for enabling controlled flight in the atmosphere which may continue extensively into space without changing the system's basic operation. One embodiment thereof also proposes a circular disc-shaped vessel wherein high-voltage electrode elements ionize atmospheric gases by corona discharge, and a pulsed electromagnet arrangement causes the ejection of the resultant plasma from the propulsion chamber—producing thrust by reaction. This device furthermore addresses the issue of producing local thrust differential(s) on the body or housing thereof, for purposes of flight maneuvering and/or directional control, by providing for a plurality of such pulsed discharge propulsion units arranged in concentric rings (within the discoid body) and having individual divergent nozzles. As such, however, the device set forth in this Patent just cited actually constitutes an ionized jet propulsion method, requiring the use of either a gaseous medium like air or the onboard storage and release of relatively massive ionizable reaction material for any operational capability in space. Moreover, the pulsed mode of operation used, which limits the ionizing potential gradient to a value of less than breakdown field intensity, significantly restricts the level of power produced as thrust as compared to the continuous DC discharge field current produced by the present invention—at voltages which may be greatly in excess of breakdown intensity. Finally, this said device relies upon an unspecified source of generated electrical power in operation which must be assumed would have to be nuclear if conventional in nature, for said device to avoid otherwise having to carry a massive and range-limiting onboard store of fuel, and not upon convenient and protracted-use permanent magnets of high energy density as the prime energy source (as the present invention does).

U.S. Pat. No. 3,464,207 discloses a quasi-corona aerodynamic vehicle which bears only a limited relationship to the present invention, but which is referenced nonetheless as being further indicative of the predominant trend in conceptualization in relation to devices intended for producing electrically-developed thrust. This device also relies upon pulsed corona discharge operation between spaced electrodes at less than breakdown field intensity, thereby greatly constraining the propulsive power obtained. It is furthermore effective only in a gaseous dielectric medium and is based upon an electric wind pressure gradient arising of an admitted incompletely-understood mechanism. This same device evidences a great deal of theoretical complexity and construction material specificity, which might seem to indicate a narrow range of feasibility. Such attributes are not uncommon in the field of electrical aerospace propulsion, and are, to a degree, shared even by the present invention. Nevertheless, it is interesting to note that sizeable thrust in the form of an aerodynamic pressure gradient or lift of up to 30 lbs./ft^2 of external electrode area or more is asserted to be obtainable from said same device by utilizing asymmetrical electrodes or electrode arrays, rather than using electrodes or electrode arrays of approximately equal area (as the present invention does)—which therefore present a uniform cross section of discharge current conduction.

PCT Application WO 85/03053 discloses a flying apparatus whose primary similarity to the present invention is that it incorporates a hollow annular flywheel which provides gyroscopic stability and encircles a relatively stationary central cabin and/or payload area. The device set forth therein also provides free-rolling means to center and stabilize the flywheel or rotor within a discoid housing, and addresses the issue that horizontal flight maneuvering and/or directional control may be accomplished by altering the local angular acceleration of the rotor—as does the present invention. This said other device, however, relies upon the questionable concept that the very large angular momentum of the rotating flywheel is somehow able to compensate the potential energy caused by gravitation and that the overall weight of the apparatus may thereby be annulled. Vertical lift must then be provided for said device in operation by atmospheric thermal updraft means, and no electrical discharge means are utilized to produce a force of lift or thrust.

The device set forth in UK Patent GB 2 312 709 A is a flying craft which is essentially disc- or saucer-shaped and which has a central axial post electrode and a spaced enclosing ring electrode whereby an electrical arc may be struck between the two, to either ignite fuel introduced via nozzles or greatly heat an exhausted air flow produced by fan means. Thus, this device actually provides hybrid jet or turbo-fan means for producing propulsive thrust, and utilizes direct electrical discharge not as the primary source of thrust but to enhance thrust produced by other means. To this end, said device must therefore operate solely in a gaseous atmosphere and/or carry onboard a range-limiting store of combustible fuel. This same device does, however, utilize powerful magnetic fields to impart a rotational and de-randomizing aspect to the arc discharge, thereby promoting a more uniform distribution of the arc discharge field's energy, as does the present invention.

In summary, a number of similar or significant electrical, magnetic, and mechanical features of the several related art devices have been examined but, whatever the precise merits and advantages thereof, none of said devices achieve the correct combination of such features required to fulfill the purposes of the present invention. In this respect it may be stated that none of these related art devices attain the distinction of being able to operate extensively in the vacuum of space without the onboard storage of relatively massive reaction material for the production of thrust. Furthermore it would appear that none of these same devices utilize a charged conductive housing comprising suitably disposed electrodes and/or electrode arrays to create and control an asymmetrical electromotive field of force on itself without expelling a reaction mass, whether in a gaseous dielectric medium or the vacuum of space, to achieve a propulsive object as the present invention does. Finally, it is observed that none of the aforementioned related art devices can develop this type of useful thrust-producing electrodynamic field in a manner which allows more than one objective function to be fulfilled thereby, such as the production of useable thermal kinetic energy and/or the development of potential signal communications capability.

SUMMARY OF THE INVENTION

The present invention is a brushless permanent magnet electrical generator which produces a useful ultra-high-energy external electrodynamic field, or continuous quasi-coherent DC corona or arc discharge of uniform current density, that completely surrounds and encloses the machine's symmetrically disc-shaped conductive housing and tends in consequence toward a balanced bihemitoroidal shape.

The invention incorporates a uniform arrangement of stationary electromagnetic armatures which may be employed as a group to impart a holistic and de-randomizing rotational force to the external discharge current that acts to inhibit the formation of concentrated arc streamer or channel phenomena, and thus to assist in rendering the electrodynamic field quasi-coherent in nature. The arrangement of electromagnetic armatures in said generator acting as a group simultaneously also contribute a portion of the machine's required input rotary torque upon the current-carrying rotor assembly thereof.

This invention provides a permanent magnet high-voltage generator wherein the primary DC voltage induced within its rotating coils is electrostatically impressed across the conductive rotor assembly thereof, in a brushless manner utilizing vacuum tube cutoff bias techniques, such that a secondary DC corona or arc discharge current is thereby initiated about the housing thereof which is orders of magnitude larger than the operating current through the said coils.

The said high-voltage primary generator of the invention then has a housing or hull which is divided into separate negative (emitter), neutral, and positive (collector) electrical sections, such that the continuous external flow of such discharge current is best facilitated and conducted internally by the proximal rotor, and wherein suitable means to circulate one or more form(s) of liquid coolant through any one or more of said housing sections may be employed for purposes of extracting an extremely high level of recoverable thermal energy (arising from such discharge current)—which is intended principally for use in the large-scale production of electrical power and/or distilled or desalinated water.

This invention as an electrodynamic field generator also provides an appreciable and practical level of electrically-developed thrust, from the relativistic impulse of an induced ultra-high-voltage external DC arc discharge current incident upon its positive (collector) housing sections, which may be rendered variably non-isometric in nature and therefore directionally propulsive by means of selectively controlling: (i) the resistance of the brushless electrical linkages between the positive housing sections and the proximal current-carrying rotor; and/or (ii) the magnetic flux output of the various stationary electromagnetic armatures.

The invention as an ultra-high-energy field generator furthermore presents a unique yet practicable method of producing a variable electromagnetic and/or gravimetric signal, by means of which useful information may be both transmitted to and received by a separate similar device, in further harnessing the effects associated with the device's external electromotive field. The relationship of said field effects to quantum potential vacuum fluctuation theory just now being studied, which may allow the devising of such a signal having an unconventional nature, is referred to again below.

In any event, the electrodynamic field generator described may be feasibly manufactured and marketed in various sizes according to need or desire in pursuit of its remarkable objects; may be operated either in a gaseous dielectric medium or in the vacuum of space; and may incorporate ancillary mechanical, electrical, and/or electronic features upon or within its housing to accomplish all the objects set forth hereinabove.

Briefly, in accordance with the preferred embodiment of the invention, first provided is a flat annular segmented rotor of very high conductive ampacity which is rendered capacitive in nature by the attachment of suitably paired electrode rings. Alternating in placement amidst such capacitive ring pairs are also attached on the rotor a series of toroidal generating coils of very large mean radius relative to the cross-sectional radius thereof. The resultant composite rotor assembly is then rotatably mounted within a sectional conductive housing. The potential energy inherent in the fields of ring-shaped stationary permanent magnets, which are affixed in concentric circular arrays within the housing, is utilized to electrically polarize the rotor across its radial annular width by means of the "primary" DC voltage created upon relative rotational motion between such coils and magnets. The thereby polarized rotor is electrically linked on one hand (about its positive inner circumference) to two axially-central collector housing sections, by electrostatic induction via a system of plane-parallel electrode elements, and on the other hand (about its negative outer circumference) to two radially-peripheral emitter housing sections across an evacuated space charge chamber. The rotor's subsequent transportation and storage of electronic charge between the proximal opposite-polarity sections of the housing causes a much higher "secondary" voltage to appear across interposed neutral sections thereof and results in a continuous ultra-high-energy DC discharge current flow about the housing externally and across the rotor internally.

BRIEF DESCRIPTION OF THE DRAWINGS

The benefits and advantages attendant to the invention and its objects as discussed hereinabove are best illuminated through the following drawings and description.

FIG. **1** shows a side sectional view of the invention's peripheral and principal working structure, comprising a rotor assembly as well as frame-mounted elements.

FIG. **2** shows a top sectional view of a 30° radial sector of that same working structure.

FIG. **3** shows a side view of the invention's symmetrical housing or hull configuration.

FIG. **4** shows a top view or bottom view of that same housing/hull configuration.

FIG. **5** depicts a small radial sector of the basic rotor assembly construction, in top view, details of the rotor segments, inter-segment insulators, and field emitters.

FIG. **6***a* and **6***b* show side and top views, respectively, of one of the invention's conductive rotor field emitters.

FIG. **7** shows a simplified schematic representation of the invention's two "primary induction ring" electrode arrays, in relation to the rotor electrical circuit.

FIG. **8** shows a schematic of one symmetrical half of the primary power system circuit of the invention, in a "single-stage" rotor construction embodiment.

FIG. **9** shows a schematic of one symmetrical half of that same primary power system circuit, in a "three-stage" rotor construction embodiment.

FIG. **10** shows a schematic of the field voltage control system circuit of the invention.

FIG. **11** shows a simplified diagram of the electrodynamic field "envelope" produced by the invention, around a side view of the invention's housing/hull.

FIG. **12** shows the diagram of FIG. **11**, with the addition of lines indicating the electric and magnetic field vectors comprising that same electrodynamic field envelope.

FIG. **13** shows a top or bottom view corresponding to the side view of FIG. **12**.

FIG. **14** shows a top or bottom view of a radial sector of the invention's neutral hull section, showing the housing cooling and/or heat transfer system arrangements.

FIG. **15** shows a detail of an embodiment of the preferred central rotor segment attachment means and rotor drive assembly.

FIG. **16** shows a schematic of the stationary armature power circuit of the invention.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Method of Operation

The invention disclosed herein is an Electrodynamic Field ("EDF") Generator. The primary challenge involved in fulfilling the objects of the present invention is how to generate and maintain a multi-megavolt potential difference across the exterior sectional housing which is not immediately neutralized due to uncontrolled direct arc discharge. The first part of the problem in this respect is that it is not possible to produce this kind of voltage directly within a traditional generating coil arrangement, so the voltage applied to the charged housing sections must be achieved by special electrostatic means. Secondly, the discharge current must be produced and conducted internally in such a manner that: (i) only a tiny percentage thereof is actually conducted by the generating coils (within their very limited ampacity compared to that of the rotor); and (ii) the negative and positive housing sections are each charged faster than they are discharged until an enormous equilibrium field potential difference is established which is a sum of roughly equal negative and positive housing section voltages.

Having thus stated the principal parts of the overall problem, it is perhaps best to examine and explain the means to its solution in the reverse order. Therefore, it is deemed of fundamental import to realize that the EDF Generator's housing configuration as proposed will demonstrate a significant static capacitance, despite its unusual geometry, given the fact that the negative and positive sections of its housing are to have essentially equal surface areas. This simple precept ensures two very important consequences: (i) that the voltage across the housing will attain a value representing equal negative and positive potentials, in reliance upon the absence of a fixed ground reference, given induced surface charge densities which are sufficient to cause a continuous 'leakage' current between the opposite-polarity sections thereof; and (ii) that a leakage current between the opposite-polarity housing sections will tend to assume a relatively uniform cross section and current density in response to a calculable level of non-uniform surface charge density between non-parallel but equal-area 'plates'. Thus, it is possible to determine using traditional formulas just how much electron charge must be stripped from the positive housing sections, and stored in ballast capacitors, in order to establish a positive surface charge density which corresponds to the desired housing or hull potential difference.

Next, the Generator must, and will, be able to charge these positive housing sections (by stripping them of native electrons) far faster than the positive surface charge so-created is neutralized by corona or arc discharge from the negative housing sections. Because these discharge current electrons will ultimately attain a transit time approaching zero, this in turn means that it must be "easier" per volt to strip negative charge from the positive sections—using a field induction system which is comprised of "primary electrode arrays"—than it is to deliver this charge to the negative sections across a vacuum chamber using relatively massive rotor-mounted field emitters.

It can be shown that the potential difference across said primary arrays will tend to be equal and opposite in operation as proposed to that across such vacuum chamber, in part due to the Faraday shielding principle, and the negative 'plate' voltage of the capacitive housing will naturally tend toward numeric equality with that of the positive sections by electrostatic induction. Therefore, the stationary electron-emitting primary cathodes which link the positive sections of the housing to the rotor are to have a much lower surface work function (and higher emissivity) than the rotating electron-emitting elements which link the rotor with the negative housing sections. This criterion is achieved by thorium-impregnating these cathodes (to stimulate their emissivity) and by not so-impregnating the rotor field emitters.

The relationship just described establishes what electrical engineers refer to as an 'instantaneous charge differential' with respect to these two sets of components as operated at equal but opposite charging voltages or potential differences, and the ratio of their respective electron emissivities will thus define the maximum ratio by which the said secondary field voltage (across the exterior housing) can be expected to exceed that in evidence across the stationary internal primary electrode arrays. This ratio will hereinafter be referred to as the "primary voltage expansion ratio", and the instantaneous charge differential concept as well as its relationship to the said voltage expansion ratio will be explained in further detail in a pertinent section to follow. It is important to note at this point, however, that for this voltage expansion to be realized it must and will be supported by ballast capacitors having suitable total capacitance as outlined hereinabove, which provide a corresponding force of electrostatic induction.

Finally, a method of impressing as large a primary DC voltage across the rotor as may practicably be developed in the device's principal generating system must and will be employed, without such "Primary Power System" circuit conducting more than a minute percentage of the discharge current thereby initiated. To do so, and in a manner that requires no brushes, vacuum tube cutoff bias technique is utilized in the present invention: whereby such primary voltage is electrostatically induced upon the rotor between two sets of capacitive electrode ring pairs (in vacuo), yet the current flow across these rotor-mounted induction electrode arrays—and through the principal generating circuit itself—is absolutely minimized by the application of a control grid voltage which is substantially negative with respect to the corresponding cathodes.

At a certain relative potential, the electrostatic force exerted by such control grid(s) on rotor circuit cathode-emitted electrons is sufficient to all but completely cut off the flow of DC current across these rotor induction arrays; yet the primary DC voltage is transmitted to the large conductive rotor segments such that a comparatively huge charging current between the opposite-polarity housing sections is initiated and maintained (across the rotor)—which subsequently expresses the expanded secondary voltage as a discharge current about the housing. Thus, an analogy could be made in describing the EDF Generator as a DC equivalent for boosting voltage by electrostatic induction means to the AC transformer (for boosting AC voltage by electromagnetic induction means).

General Form of the Invention

Accordingly, referring to FIGS. **1** through **4**, the present invention constitutes a novel form of brushless permanent magnet electrical generator for producing a useful ultra-high-energy external electrodynamic field, or continuous quasi-coherent DC corona or arc discharge of essentially uniform current density, and as such the Generator fundamentally comprises and incorporates:

[A] a flat conductive rotor **(6)** principally formed of a large number of evenly-spaced individual metallic conductors or segments **(14)** and an equal number of conductive electron-emitting elements or field emitters **(17)**, radially arranged in an annular flywheel configuration rather like the rotatable deck of a carousel, which are all connected in parallel electrically by the attachment of a thin metallic electrode ring **(68)** about the positive inner circumference of each of the two major flat faces of the annular flywheel thus formed and by the attachment of another such ring **(22)** about the negative outer circumference of each such flat rotor face;

[B] a principal electrical generating system mounted on said rotor and used to create a very high primary DC voltage within a plurality of rotating toroidal field coils **(35)(40)(56)** by electromagnetic induction with an equal number of circular arrays of ring-shaped stationary permanent magnets **(34)(39)(55)** (which are positioned adjacent and concentric thereto), such that an internal primary discharge current may thereby be initiated between the rotor **(6)** and a housing electron "Emitter Ring" **(47)**—across an evacuated space charge chamber **(11)**—which is orders of magnitude larger than the operating current through said field coils;

[C] a conductive exterior housing **(1)** divided into a plurality of distinct electrical sections—negative **(3)**, neutral **(4)**, and positive **(5)**—which may be charged from the internal rotor by electrostatic induction and/or by such primary discharge current (as the case may be) and within an evacuated induction compartment **(12)** of which said rotor **(6)** as an assembly, including the said generating coils attached thereto, is rotatably mounted;

[D] a plurality of individual systems of stationary plane-parallel electrode elements **(64)**-**(67)** which electrically link the polarized rotor **(6)** by electrostatic induction about its positive inner circumference (within said evacuated compartment **(12)**) to two axially-central electron collector housing sections **(5)**, and a plurality of similar individual systems of rotor-mounted plane-parallel electrode rings and/or radial electrode elements which may be utilized to connect a principal portion of the individual DC voltage outputs of said rotating field coils in opposed series-parallel across such composite rotor assembly;

[E] this composite rotor assembly **(6)** as a means for the transportation and capacitive storage of electronic charge between opposite-polarity sections of the housing **(5)** and **(3)** (which are proximal to the inner and outer rotor circumferences respectively), such that a vastly higher secondary voltage is caused to appear across interposed neutral housing sections **(4)**, resulting in a continuous ultra-high-energy DC discharge current about the housing externally which is also conducted internally across the rotor segments **(14)** that extend across the neutral housing portion and between the positive sections;

[F] a large number of stationary electromagnetic armatures **(37)** in a uniform circular arrangement located within the neutral electrical region **(4)** of the housing which are to be operated principally on variable DC voltage, but which may also accept limited pulsed unidirectional or alternating voltage input, and which are employed as a group to impart an orderly rotational magnetic vector moment to the external discharge current that acts to inhibit the formation of concentrated arc streamer or channel phenomena and thus to assist in rendering such current quasi-coherent in nature; and

[G] the said principal electrical generating or Primary Power System as operated and used in such a manner that a minor pulsed unidirectional or alternating voltage is superimposed upon the high-energy DC rotor current, and the external discharge current therefore acquires a discrete AC power factor, for purposes of modulating the electrodynamic characteristics of the external field during normal operation and/or for purposes of exploring the invention's potential signal communications capability in relation to intriguing new theories in high-energy quantum physics and relativity.

In a "single-stage" rotor electrical circuit embodiment of the Generator, as intended for use in an electrical and/or thermal energy output application, any such minor AC voltage present in said rotor circuit receives only a modest and single level of amplification—in a manner generally associated with a single multi-electrode vacuum tube. In a "three-stage" rotor electrical circuit embodiment of the Generator, as intended for use in a propulsive and/or communications application, any such minor AC voltage may receive a substantial and multiple level of amplification—in a manner generally associated with a multiple number of sequentially-coupled multi-electrode vacuum tubes.

FIG. **1** shows a side sectional view of the general form of the invention, as if cut through the central axis of the circular structure. Only one side of the tapered peripheral portion of the structure is shown, the overall Generator being symmetrical about the axis along which the section shown is cut. FIG. **2** shows a top sectional view of the invention, as if cut through just above the rotor assembly at the "equator", with only about a 30° radial sector of the circular structure shown (the remaining 330° being identical around the Generator). In each case, identical reference numbers are used to indicate identical elements. FIGS. **3** and **4** are side and top or bottom views of the Generator (respectively), showing the outward appearance of the invention and the external field it produces. The following discussion of the general form of the invention will be in reference to FIGS. **1** to **4**, unless otherwise noted.

The entire apparatus of the invention is contained within a circular conductive exterior hull or housing **(1)**, which is tapered from its vertical centerline ("axis") to its rim (or "equator"). The overall housing is radially symmetrical around the axis, and bilaterally symmetrical about the equator. This housing's two central one-base spheric-zone-shaped sections **(5)** will be positively charged, its two intermediary conical ring-shaped sections **(4)** will be electrically neutral, and the convergent outermost or peripheral conical ring-shaped sections **(3)** will be negatively charged. Preferably, the surface areas of the outer **(3)** and inner **(5)** housing sections are equal, and specific details of the housing or hull design are given in a section to follow.

Each of the housing's two symmetrical positive sections **(5)**, or hull positive "zones", should be divided into equal-area radial sectors **(42)** which are equal in number to the number of like sections into which the device's superstructure is divided. Thus, in the preferred embodiment of the Generator these two positive zones **(5)** will each be divided into 36 such sectors, and isolated top and bottom, yielding a total of 72 individually controllable discharge collector areas. The positive zone sectors **(42)** jointly comprise what may hereinafter be referred to as the device's "Field Hub", and should be composed of a conductive refractory metal or high-temperature structural alloy such as #310 stainless steel. These positive sectors are insulated from sector-to-sector by insulating "partitions" **(43)** of a suitable ceramic material (such as Cordierite), and must be supported by an insulating understructure (not shown).

At least one hull "neutral ring" surface layer **(46)** is also to be affixed on a housing frame and composed of one or more insulating ceramic material/s (such as Cordierite, Zirconia, and/or Kézite, as necessary). Said neutral hull layer/s may be applied in the form of plates and/or tiles to nonmetallic "deck" sheeting (not shown separately) supported by superstructure beams **(153)**. This decking and its supporting beams are both preferably composed of a carbon composite material, and will thus be fairly conductive. Therefore, both must be isolated from the positive Field Hub by a dielectric buffer material **(45)** and from the hull's Emitter Ring **(47)**, comprised of two negative housing sections **(3)**, by an auxiliary dielectric buffer **(50)**. The Generator's neutral housing section superstructure is shown in detail in FIG. **14**, as is a piping system of coolant tubes or "thermal conduits" **(74)**-**(79)** for maintaining the hull neutral sections **(4)** and superstructure—as well as the enclosed Primary Power System—at an acceptable operating temperature. Such an intercooler system would utilize a cryogenic coolant such as liquid air or nitrogen when the EDF Generator is operated in the atmosphere for purposes of thermal energy output, and will be explained in greater detail in a section to follow.

The negative hull plates which form each section **(3)** of the Generator's peripheral Emitter Ring **(47)**, to which the rotor **(6)** distributes the discharge current forming the exterior field envelope, should be composed of high-purity aluminum which is copper-clad and then nickel-plated to prevent erosion at high field voltage and current levels. The abutting edges of all of the adjacent such plates must be welded one to another, and affixed to any necessary neutral hull section hardware, such that a hard vacuum may be drawn within the peripheral chamber **(11)** thereby formed. The induction compartment **(12)** enclosed between the two neutral hull sections **(4)** is contiguous with the peripheral or space charge chamber **(11)**, and thus is equally evacuated.

A pressurizable central chamber **(2)** within the housing **(1)**, located between the two hull positive zones **(5)** thereof, constitutes the location of the Generator's control cabin and of its payload area when used as an aerospace vessel. The interior **(11)**-**(12)** of the negative **(3)** and neutral **(4)** housing sections (respectively) is separately sealed, and must be as highly evacuated as possible to promote the efficiency of the various rotating and stationary electrode arrays and to prevent their destructive failure. The main vacuum seals **(155)** and the rotor's vacuum chamber charge buffers **(156)** are shown in FIG. **1**.

Inside the evacuated induction compartment **(12)** of the housing **(1)**, and surrounding the central chamber **(2)**, is located the composite rotor assembly **(6)** in the form of a flat annular flywheel. In the preferred embodiment, 180 individual conductors or segments **(14)** of very high collective ampacity are employed in the rotor, as will be seen in the discussion of FIG. **5** below, as are a like number of inter-segment insulators or segment "separators" **(16)**. The preferred material for the rotor segments is high-purity copper which is certified oxygen-free, to resist the production of free oxygen radicals at high operating temperatures which would degrade the performance of the various electrode arrays and cause an undesirable increase in vacuum chamber **(11)** temperature.

The inner ends of the rotor segments (**14**) and segment separators (**16**) are mechanically connected via a retaining and lockdown section thereof to one or more ring gear(s) (**8**), which run/s entirely around the central aperture of the rotor assembly. Suitable insulating attachment hardware (**7**), as described in detail in the following section (and shown in FIG. **15**), will be used to effect a rigid mechanical union of the segments and separators to the said ring gear(s) and associated drive means such that the gear(s) and all conductive components of such drive means are electrically isolated from the energized segments. The ring gear(s) (**8**) may be engaged by one or more pinion or drive gear/s (**162**) powered by a motor (**9**) or plurality thereof, which may be of various types (such as electrical or hydraulic), causing the rotor to rotate around the central chamber (**2**). The rotor is supported near its outer edge by two ballraces (**25**) mounted to the housing superstructure and to the rotor, and having bearing balls (**26**), allowing it to spin freely.

As can be seen in either FIG. **1** or **7**, the rotor segments (**14**) are electrically connected together at each end by continuous inner (**68**) and outer (**22**) electrode rings, and also at intermediate points by induction cathode rings (**28**) and induction anode rings (**61**). The inner electrode rings (**68**) constitute the primary induction ring array anode elements, as will be described below. The outer electrode rings (**22**) are each covered by a dielectric layer (**23**) which is in turn covered by an anode ring (**24**), and each combination of rings (**22**) and (**24**) separated by dielectric layer (**23**) forms a housing charge ballast capacitor whose function will be explained in detail below. Induction anode rings (**29**) and cathode rings (**59**) are used to electrostatically induce corresponding negative and positive voltages on rotor-mounted rings (**28**) and (**61**), respectively, thereby electrically energizing the rotor segments as will also be described in further detail below.

FIG. **5** shows details of the basic rotor construction. Wedge-shaped ceramic pieces or segment separators (**16**) equal in number to the conductive rotor segments (**14**) are to be positioned uniformly between the individual segments, as are the rotor's peripheral field emitters (**17**). To form the base rotor construction, the segments, separators, and field emitters must be equal in number, preferably using 180 of each. The said segment separators are actually more structural than insulative in purpose, and not only stiffen the entire rotor to prevent its distortion at elevated temperature and rotational speed but also provide a nonconductive base for mounting all other rotor assembly components. Each of the ceramic separators (**16**) is bonded to its adjacent segments (**14**) by refractory adhesive layers (**15**).

Two important criteria for selecting a ceramic segment separator material are that the compound chosen be readily machinable like steel and provide very high physical strength yet require no firing. These parameters limit the choice made almost to a solitary substance: magnesium silicate ($MgSiO_3$). When suitably formed, this material can be milled, drilled, and tapped like steel (using low-speed tungsten carbide tools) and approaches the physical strength of dry-pressed porcelain. It is ready for use directly after shaping and requires no kiln hardening.

This latter characteristic is very important as it is extremely difficult to produce fired ceramic pieces with the exacting dimensional control necessary in this case, and it eliminates the tendency of thicker fired pieces to develop micro-cracks which may potentially cause destructive failure of the component under high centrifugal loading. Surfacing can be done with abrasive paper or grinding wheels, and it is very important that this material be machined or worked perfectly dry as it is fairly porous and any lubricant used tends to significantly alter its dielectric properties. Magnesium silicate of this type is available from CeramTec NA as CeramTec designation AlSiMag 222, and will withstand constant working temperature of over 1300° C.

The insulating separators (**16**) are shorter than the conductive segments (**14**), allowing for the field emitters (**17**) to be fastened outboard of the separators. The field emitters are thus located around the outer circumference of the rotor assembly and should be formed of a sintered refractory composite, such as tungsten-copper, selected for its relative surface work function and superior arc-erosion characteristics. As can be seen in FIGS. **5**, **6***a*, and **6***b*, the field emitters are fastened mechanically and electrically to the conductive segments (**14**) through locator pins (**19**), which fit into mating holes recessed into the ends of the segments. These emitters (**17**) taper to a somewhat rounded edge at the tips (**20**) which form the outermost periphery of the rotor (as an assembly), and are squared-off at their bases (**18**) to fit tightly against the insulating separators (**16**). From the top view of FIG. **6***b*, it can be seen that the field emitters follow the tapering of the segment separators (**16**), increasing slightly in width from their bases (**18**) to the tips (**20**). The conductive segments (**14**), in contrast, are of constant width over their length.

Overall Description of the Primary Power System Structure

Rotor Drive and Positioning Assemblies

Referring now to FIG. **15**, a split-frame centrifuge-style carrier assembly reminiscent of the rotatable mounting arrangement for the turntable of a carousel may be employed to firmly suspend, center, and rotatably mount the composite rotor assembly about its inner circumference within the evacuated induction compartment of the housing, and such carrier assembly should preferably be constructed of a weldable nonmagnetic metal alloy which is suitably insulated from the electrically energized rotor assembly components.

The rotatable carrier assembly may be mechanically-affixed and/or welded to one or more large metal ring gear(s) which may be comprised of like sections (**163**) to facilitate construction and which in turn may be supported, driven, and dynamically braked by one or more low-voltage DC motor-generator unit(s) (**161**) (similar to those used in large electric trains) and drive pinion gear/s (**162**). Such ring gear(s) may additionally be supported by pinion gear and pillow block bearing assemblies (if necessary—not shown in FIG. **15**). Such drive motor-generators(s) (**161**) and said pillow block bearing assemblies (if any) must be separately and nonrotatably affixed to the relatively stationary housing superstructure.

Referring still to FIG. **15**, it can be seen that in the preferred embodiment the inner end of each rotor segment (**14**) and segment separator (**16**) is flared outward into a retaining "fantail" (**160**). These retaining fantails are preferably enclosed and prepared for connection into a driven assembly using a system of: [i] insulating inner (**171**) and outer (**173**) ceramic thermal spacers; [ii] axial (**172**) and radial (**174**) nylon (or equivalent) load buffers; and [iii] insulating displacement (**170**) and lockdown (**175**) ceramic bushings. The retaining fantails are then secured into a driven assembly by two identical carrier half-structures comprising frame struts and members (**168**) and retaining ring sections (**167**), using: [i] mounting ring sections (**164**) each having two ring gear sections (**163**) attached thereto;

[ii] a corresponding number of mounting ring spacers **(165)** each having a very stiff metallic flexor plate **(166)** attached thereto; and [iii] nylon (or equivalent) torsion buffers **(169)**. The drive pinion gears **(162)** allow a preferred total of 32 like and evenly-spaced DC drive motor-generator units **(161)**, which are mounted to the Generator's ground frame **(10)**, to apply input rotational torque (or counter-rotational braking torque) to the composite rotor assembly by engaging a total of four ring gears (comprised of sections **(163)**) which are mounted to the rotor's described drive carrier assembly.

Rolling bearing assemblies should be employed to further center and stabilize the composite rotor about its outer periphery within the evacuated induction compartment of the housing, as a readily feasible mechanical alternative for that purpose to 'zero-friction' electromagnetic positioning systems (of the type normally associated with mag-lev train technology) which are of a much greater complexity.

Such rolling bearing assemblies may comprise two complimentary pairs of stationary and rotating circular-groove raceways **(25)** about the outer periphery of the composite rotor assembly, as well as an appropriate number of bearing balls **(26)** including uniform ball separation means, and these bearing assemblies may be composed of nonmagnetic stainless steel and/or a specialized ball bearing ceramic (such as silicon nitride (Si_3N_4)) which may run unlubricated in a very hot electrically-charged environment

The Rotor and Associated Elements

There are a number of electrical structures attached to the base rotor construction, which rotate with the segments, separators, and field emitters as a composite assembly. It will be understood that the structures mounted on the rotor and those interacting with the rotor are symmetrical above and below the incorporating device's equator, as well as either continuous or repeating in placement around the device in the plane of the rotor segments, so that a reference to an element at one point above the rotor also applies to the similar elements below and around the rotor.

Referring to FIGS. **1** and **2**, and in the preferred three-stage embodiment of the invention suitable for an aerospace vessel (as shown schematically in FIG. **9**), the following elements of the principal electrical generating means are mounted upon the annular rotor. Generally starting at the outer circumference of the rotor and working inward, and, for each structure, starting at the rotor and working upward, these are:

[a] the ballast capacitor, comprising a negative ring element **(22)** attached directly to the rotor, a dielectric layer **(23)** affixed thereto, and a positive ring element **(24)**;

[b] the bearing ballrace **(25)**, within which the bearing balls **(26)** roll between two identical raceway halves—one on the rotor, and one mounted to the vessel's stationary structural frame. The ballrace **(25)** and bearing balls **(26)** support the rotor peripherally and are structurally mechanical, not taking part in the electrical operation of the invention;

[c] the voltage-induction diode ring array, in the form of a cathode ring **(28)** which is mounted on and electrically connected to the rotor segments **(14)** and an anode ring **(29)** which is affixed plane-parallel to the cathode ring **(28)** and insulated from it by a suitable supporting structure using insulating posts, pins, or brackets in or on the segment separators **(16)**. Although the resulting combination could be thought of as a 2-element or diode vacuum-tube construct and is similarly named, it is more accurately a "cold" parallel plate capacitor of high value and appreciable AC conductance which preferably exhibits a modest circuit DC leakage current;

[d] the outer field coil **(35)**, helically wound and supported upon a continuous coil support or structural core which is preferably nonferromagnetic (not shown). There are preferably two windings on the core which together comprise such coil, as will be discussed in the circuit section below and as is shown in FIG. **9**—the major portion or field winding **(81)**, and a minor portion or bias winding **(82)**;

[e] the outer voltage-transfer triode ring array, in the form of cathode **(30)**, control grid element **(31)**, and anode **(32)** rings or elements, affixed plane-parallel or radially-concentric above the rotor segments **(14)** and insulated from them and from each other by similar such supporting structure (using insulating posts, pins, or brackets) in or on the segment separators **(16)**. This combination also comprises a triode vacuum-tube construct with controllably variable AC conductance;

[f] the center field coil **(40)**, helically wound and supported upon a continuous coil support or core of the type described. There are again two windings on the core, as will be discussed in the circuit section below (and as is shown in FIG. **9**)—the major portion or field winding **(83)**, and a minor portion or bias winding **(84)**;

[g] the inner voltage-transfer triode ring array, in the form of cathode **(51)**, control grid element **(52)**, and anode **(53)** rings or elements, affixed plane-parallel or radially concentric above the segments and insulated from them and from each other by the described supporting structure(s). This combination again comprises a triode vacuum-tube construct with controllably variable AC conductance;

[h] the inner field coil **(56)**, helically wound and supported upon such continuous coil support or core. There are once again two windings on the core, as will be discussed in the circuit section below and as is shown in FIG. **9**—the major portion or field winding **(85)**, and a minor portion or bias winding **(86)**;

[i] the voltage-induction triode ring array, in the form of an anode ring **(61)** which is mounted on and electrically connected to the rotor segments **(14)** as well as a cathode ring **(59)** and control grid element ring **(60)** affixed plane-parallel to the anode ring **(61)** and insulated from it and from each other by the said supporting structure(s). The resulting combination comprises a triode vacuum-tube construct with controllably variable AC conductance;

[j] the primary induction anode ring **(68)**, which is also mounted on and electrically connected to the rotor segments **(14)** and which provides the positive element of the stationary plane-parallel electrode systems used to induce the exterior field; and

[k] the ring gear(s) **(8)** or other appropriate means for allowing a drive mechanism to spin the rotor.

Field Coils

Referring now to FIG. **1**, the field coils **(35)(40)(56)** which comprise the major rotating portion of the primary DC voltage generating means are formed of conductive insulated magnet wire multi-layer toroidally wound upon cores (not shown) whose permeability should generally be minimized, and main portions of the DC voltage outputs of the three field coils comprising each of two like groups thereof—one above the plane of the rotor and one below—are sequentially connected via a plurality of individual systems of rotating plane-parallel electrode rings (or alternatively, modular radial electrode element systems or tubes) in the preferred three-stage rotor emodiment of the invention shown. The number of wire turns comprising each such field coil should in general be maximized (in a practical manner), in either the single- or three-stage embodiments.

As is depicted more clearly in FIG. **8**, each such field coil may be compound wound or comprise two or more independent voltage generating portions, with either all or the major portion of each coil constituting the field winding **(81)(83)** or **(85)** or principal DC voltage generating portion of that coil as a whole, and the principal or field winding portions of the field coils comprising either of the said two like groups thereof may be hooked in series one to another by direct physical connection in the single-stage rotor embodiment of the Generator shown. Each of the two like sets of three such field windings thus formed are then connected on their respective sides of the rotor between the plate resistor **(91)** of the diode array anode ring **(29)** and the inner triode array cathode ring **(59)**, while observing appropriate conventions of polarity assignment.

The remaining minor portion (if any) of each field coil may comprise one or more independent bias and/or control winding(s), or secondary DC voltage generating portion(s) of that coil as a whole, with the largest such secondary winding if any constituting the control grid bias winding of that coil. As shown in FIGS. **8** and **9**, one such bias winding **(82)(84)** or **(86)** is provided for each field coil and is positioned in the schematic(s) immediately thereabove. In the single-stage rotor embodiment of FIG. **8**, these bias windings are also directly connected in series one to another between the inner triode array cathode ring **(59)** and the corresponding grid resistor **(90)** of that same array's control grid **(60)**, with the negative end of each bias winding toward the said grid resistor.

In the three-stage embodiment of FIG. **9**, the field winding of each said field coil is connected in series between the plate resistor **(92)** attached to the anode ring of the outerlying electrode system which is adjacent to that coil and the cathode ring of the innerlying electrode system adjacent thereto, while observing appropriate conventions of polarity assignment. The bias winding (if any) of each said field coil is then connected between the cathode element **(30)(51)** or **(59)** of such innerlying electrode system and the grid resistor **(90)** attached to the negative control grid **(31)(52)** or **(60)** (respectively) of that same electrode system, while observing appropriate conventions of polarity assignment, such that said bias winding is in parallel with the cathode ring or element of that electrode system and may therefore provide an AC and/or DC voltage bias with respect thereto in the manner associated with electron vacuum tubes.

Rotating Electrode Arrays

Electron vacuum tube design, construction, and operating methods are employed to impress the principal portion of each field coil group's combined series DC voltage output across the Generator's rotor in an opposed parallel arrangement (comprising two like field winding circuits or rotor subcircuits), such that the rotor segments become electrically polarized, although the rotor is not considered uniformly energized electrically until a finite DC Primary Power System circuit current is established against the series capacitance and high bias voltage of the several rotor-mounted electrode systems.

Two special pairs of highly-capacitive electrode rings **(22)(24)** seperated by a dielectric media **(23)** are affixed upon the polarized rotor **(6)** near the outer negative periphery thereof, and supplied with a portion of the Generator's positive primary DC output voltage, in order to accomplish the storage of a predetermined amount of ballast electron charge which is sufficient to enable a desired much-higher secondary voltage to appear across the neutral sections of the housing (given the static capacitance and surface charge density characteristics of the housing as a whole in operation).

Electron vacuum tube design, construction, and operating methods may be employed to induce and/or modulate a substantial DC bias upon the plate currents of any or all of the various rotating plane-parallel and/or radial electrode systems or "arrays" employed in the present invention. Each rotating three-electrode system or triode array so-employed should be constructed in such a manner that it exhibits a minimum design amplification factor (μ) equal to 4.0, with reference to any AC voltage or signal present in its electrical circuit and as a function of its engineered relative electrode spacing(s).

Electron vacuum tube design, construction, and operating methods may furthermore be employed to induce, modulate, and/or amplify a minor pulsed unidirectional or alternating voltage upon the high-energy DC rotor current, and therefore on the field envelope current as well, as a means by which the electrodynamic field produced by the Generator may potentially be used (as a form of antenna) for purposes of transmitting and receiving a variable electromagnetic and/or gravimetric resonant frequency signal either to or from itself or a separate similar device (as the case may be).

In light of this teaching, it is contemplated that one possible type of communications signal which might be investigated employing the EDF Generator would couple the electric, magnetic, and gravitic forces to utilize the wave mechanics of quantum potential vacuum fluctuations ($\Delta E_q/\Delta t$), which are believed by some theoretical physicists to propagate at c^2 in order to explain the uniform operation of gravity and entropy in two space-time continuums which are coincident but completely out-of-phase. Should this quantum gravity theory prove correct, the transmission delay of a gravimetrically-coupled EM signal at this wave speed would be only 0.10525 seconds per light-year.

In any event, it should be noted that the rotor interstage coupling transformers **(89)** which are depicted in FIGS. **1** and **9** (but are not shown in FIG. **2**) would be employed only in a preferred three-stage rotor embodiment, and mounted (one each) immediately above the rotating diode and transfer triode arrays (**(28)**–**(29)**, **(30)**–**(32)**, and **(51)**–**(53)** respectively). These transformers may preferentially take the form of continuous toroid coils on either powdered iron or nonmagnetic cores (like the field coils), having single-layer primary and secondary windings, or of one or more pairs of balanced toroidal arc-section coils and cores of similar composition which are uniformly distributed above said rotating arrays. Each transformer **(89)** uses two DC blocking capacitors **(88)**.

It may be appreciated from FIGS. **8** and **9** that each bias winding or series group thereof, its associated triode array, and a corresponding output plate resistor **(91)** or **(92)** together comprise one stage of amplification of any AC signal voltage which may be present across that triode's grid resistor **(90)** as an input. In this way, electromagnetic waves of one form or another from outside the external field created by the Generator may be both detected, from their influence on the waveform(s) of the field envelope current, and amplified from within the device (when used as an aerospace vessel).

The Interaction of the Rotor and Frame-mounted Elements

Magnetic Rings

A plurality of circular arrays of stationary permanent magnets essentially comprise the stationary portion of the primary DC voltage generating means, and each such magnet is preferably to be formed as a C-shaped ring composed of axially round stock of a metallic ferromagnetic substance which exhibits a very high residual induction (such as the alloy family commonly known as Alnico). The number of individual magnetic rings comprising each such array should in general be maximized (in a practical manner). Preferably, each magnet array has a rigid and annular nonmagnetic round core (not shown) which is roughly equal in minor diameter to the inside diameter of the constitutent rings, such that the rings may be easily strung thereupon rather like beads on an abacus wire.

Referring once again to FIGS. **1** and **2**, each of the several magnetic ring arrays **(34)(39)(55)** may be installed in sections which are preferably equal in number to the number of like sections into which the incorporating device's superstructure is divided. The rotating field coil **(35)(40)** or **(56)** associated with each such circular array of individual magnetic rings must be respectively concentric and adjacent thereto, and is basically positioned on both the radial and axial centerlines of the flux gaps of that array's constituent rings. As power is generated thereby, these rings may need periodic replacing.

The maximum length of the flux gap of each such C-shaped magnetic ring must of course be just slightly larger than the minor outside diameter of its associated field coil, and should be roughly equal to or just slightly less than the formed inside diameter of that magnetic ring. In the preferred embodiment these ring flux gaps are oriented horizontally, although they may also be readily oriented vertically if so desired. Each magnetic ring may also be mounted within the superstructure of the housing such that a comparatively small thin axially-polarized wafer of a highly-coercive permanent ferromagnetic substance (like sintered Ferrite 5 ($BaO.6Fe_2O_3$)) is positioned in the hollow center of that ring magnet, with such wafer's magnetic poles facing the like poles thereof, so as to act as a leakage flux blocking or reducing mechanism with respect thereto. These optional components would be most easily installed by insertion into a corresponding slot cut into each of the described mounting cores, and are recommended for use in very large devices.

Also arranged concentrically around the evacuated induction compartment **(12)** of the Generator, but fixedly mounted within the vessel's structure instead of rotating with the rotor assembly, are the following elements of the principal electrical generating means. Again, they are discussed in order from the outer circumference of the vessel inward, as shown in FIGS. **1** and **2**, and a reference to an element at one point above the rotor also applies to the similar elements below and around the rotor:

[a] the stationary half of the ballrace **(25)**;

[b] the outer array **(34)** of stationary permanent magnets, which should be mounted on a nonmagnetic core and in the preferred embodiment incorporates a maximum of 900 such ring magnets equally spaced around the vessel superstructure;

[c] the circular group of stationary electromagnetic armatures **(37)** or "variable inductor" array, as used to impart a rotational magnetic force upon both the external discharge current and the internal rotor assembly (which is proximal thereto). Each such array preferably comprises a maximum of 180 such armatures, and is to be approximately centered within the neutral region of the housing so that its constituent armatures are axially parallel to the rotor's axis of rotation;

[d] the center array **(39)** of stationary permanent magnets, which should be mounted on a nonmagnetic core and in the preferred embodiment incorporates a maximum of 720 such ring magnets equally spaced around the vessel superstructure;

[e] the inner array of **(55)** stationary permanent magnets, which should be mounted on a nonmagnetic core and in the preferred embodiment incorporates a maximum of 576 such ring magnets equally spaced around the vessel superstructure; and

[f] the stationary anode ring **(58)**, located adjacent and plane-parallel to the rotating cathode **(59)** of the inner electrostatic-induction triode array, which develops an induced positive voltage that is made available to the control circuitry shown in FIG. **10** as described below.

FIGS. **8** and **9** show two embodiments of the Primary Power System. The "Field Induction System" or stationary plane-parallel electrode arrangement(s) employed to induce the exterior field is/are basically the same in either case, and the voltage control system of FIG. **10** may be used with either embodiment.

The simpler single-stage rotor system of FIG. **8** is primarily intended for applications wherein the EDF Generator would be employed to produce useful electricity or heat, the latter being extracted from a principal liquid hull coolant circulated through the primary thermal conduits **(48)** (shown in FIG. **1**) that encircle the field power resistors **(63)**. Two thermal exchange or main service manifolds (not shown) would provide ground-based support to such a Generator from an associated utility or physical plant, as well as a remarkable level of electrical or thermal output from such Generator thereto. Each such manifold would connect directly to a single circular center sector of one of the two positive housing zones. Each of these two center sectors **(44)**, one or either of which is depicted in FIG. **4**, must in this case be composed of a nonconductive material.

Description of the Primary Power System Circuit

Referring again to FIG. **8**, the respective outer, center, and inner field windings **(81)(83)(85)** are directly connected in series, and the voltage thus generated is applied to the inner induction triode array cathode ring **(59)** and (through plate resistor **(91)**) to the outer induction diode array anode ring **(29)** and to the ballast capacitor anode ring **(24)**. The respective outer, center, and inner bias windings **(82)(84)(86)** are also connected in series, and the voltage thus generated is applied between the inner induction triode array cathode ring **(59)** and (through grid resistor **(90)**) to the control grid **(60)** of that same array. This places said control grid in parallel with that cathode electrically.

The turns numbers of the bias windings **(82)(84)(86)** and the grid resistor **(90)** values are chosen to apply a bias voltage to control grid **(60)** such that the inner induction triode array is biased very nearly to current cutoff. This results in a very high voltage being induced on the rotor segments **(14)** between outer induction cathode ring **(28)** and inner induction anode ring **(61)**, but at a very low series field coil current due to the near-cutoff bias of the triode array. It is expected that the voltage on the rotor segments between these two rings, in a preferred embodiment using this arrangement, would be approximately 8,000 volts for a four-foot diameter prototype, or about 2,000 volts per foot of housing diameter. It is also anticipated that series field coil current could be limited to fractional amperage in smaller such devices and to single-digit amperage in the largest.

The preferred embodiment shown in FIG. **9** is a three-stage rotor system which is considered to be much more useful for applications wherein the EDF Generator may be employed as an aerospace vessel, in that it allows energy/wave-function signals from outside the vessel to be detected and amplified in the manner previously described.

In this embodiment, instead of the field windings **(81)(83)(85)** being directly connected together in series, intermediary voltage-transferring triode vacuum tube constructs are used to effect an indirect series connection between the said field windings. The voltage thus generated is again applied to the inner induction triode array cathode ring **(59)** and (through plate resistors **(92)**) to the outer induction diode array anode ring **(29)** and to the ballast capacitor anode ring **(24)**. The bias windings **(82)(84)(86)** are in this case used to provide an independent bias voltage to the control grids **(31)(52)(60)** of the outer, center, and inner rotating triode arrays, respectively, in parallel with their cathodes.

As in the single-stage rotor embodiment described above, the turns numbers of the bias windings **(82)(84)(86)** and the grid resistor **(90)** values are chosen to apply a bias voltage to control grids **(31)(52)** and **(60)** respectively such that each corresponding triode array is biased very nearly to current cutoff. This results in a very high voltage being induced on the rotor segments **(14)** between outer induction cathode ring **(28)** and inner induction anode ring **(61)**, again at a very low or fractional current due to the near-cutoff bias of the triode arrays. It is expected that the voltage on the rotor segments between these two rings, in a preferred embodiment using this arrangement, would be approximately 1,500 volts per foot of vessel diameter.

In general, the stationary anode rings **(58)** and all of the rotating electrode rings except the primary anode rings **(68)** should be composed of a nonmagnetic structural nickel alloy (such as Inconel 600). All rotating control grid element wires should be composed of an alloy such as that commonly known as nichrome. The rotating cathodes may be composed of modestly-thoriated tungsten, however, if and only as necessary to establish a small DC Primary Power System current if the observed 'dark discharge' current values are deemed insufficient, particularly in the three-stage embodiment wherein amplification of a minor AC signal voltage may be desired for communications purposes. The ceramic material Titania, with a nominal dielectric constant of 85, may preferably be used to form the posts, pins, or brackets which support the various rotating electrodes.

It is important to observe that both Primary Power System (rotor) field winding circuits as configured in both the single- and three-stage Generator embodiments will have an inherent AC series resonant frequency, and operation at that frequency will maximize the series field winding AC line currents (within given circuit resistor constraints) as well as the dependent stage plate voltage drops (across resistor(s) **(91)** or **(92)** respectively) which allow stage AC signal reproduction and amplification. As in traditional related practice, the series resonant frequency will be that at which the given circuit's inductive and capacitive reactances are approximately equal and opposite. Thus, the series resonant frequency of either of said Generator rotor circuit embodiments will be largely contingent upon the total field winding inductance and coil core permeability, as the various rotor electrode arrays and the plate resistor(s) **(91)** or **(92)** must have quite specific fixed relative design values of capacitance and resistance (respectively)—as a function of total field winding voltage—for proper overall DC and AC circuit performance.

Each Primary Power System control grid winding subcircuit as configured in the three-stage rotor embodiment may then be readily tuned for stage AC parallel resonance at the said inherent series resonant frequency, thereby minimizing stage grid subcircuit line current while maximizing the stage grid voltage drop (across each grid resistor **(90)**) which determines the level of net stage AC signal voltage gain or amplification obtained. As in traditional related practice, the parallel resonant frequency will once again be that at which the given subcircuit's inductive and capacitive reactances are approximately equal and opposite. This condition may be achieved by connecting an optional capacitor of suitable value if desired across the secondary winding (or control grid resistor side) of each rotor stage coupling transformer **(89)**, as indicated in FIG. **9**.

The Field Induction System

Before commencing a technical discussion of the stationary induction electrode systems which will be used to charge the positive housing sections, it will be necessary to further elucidate certain aspects of the Primary Power System electrical circuit so that a few of its key operational characteristics are more clearly understood.

Referring now to FIG. **7**, each of the Generator's two identical groups of rotating field coils (one above and one below the rotor) are in essence connected in simple series across the rotor segments using pairs of capacitive electrode rings, thus forming a single series loop with respect to the rotor. The two such series loops are at the same time, however, connected in mutually-opposed electrostatic parallel with respect to the rotor segment circuit leg common to both such loops.

Based on applicable principles of proper circuit resolution and series DC capacitance, such an opposed series-parallel circuit may readily be configured such that equal and opposite-polarity voltages will result upon the rotor-mounted induction cathode **(28)** and anode **(61)** rings in the absence of a fixed ground reference. To this end, the rotor assembly itself is to remain completely without direct earth or chassis ground in operation, and a corresponding segment polarization will continue as the circuit is duly energized by a finite DC conduction current.

When a primary DC rotor current is also established to the housing's Emitter Ring **(47)** and an external discharge current in turn occurs to the positive zone sectors **(42)**, the two systems of primary induction electrodes **(64)**–**(68)** also complete two simple series loops back to the rotor and the two such external field envelope subcircuits in parallel constitute the EDF Generator's "Field Induction System" circuit. Thus, the total field discharge current equals the primary rotor current, and the rotor segment circuit leg is common to both of the major system circuits described.

As is apparent from the section on Method of Operation, it is imperative that the charged housing sections also have no direct earth or chassis ground, being instead of a 'floating-ground' reference with respect to the described primary rotor circuit. The cathodes **(64)** of these two primary electrode systems may have a chassis ground reference at a very high resistive bias (as indicated in FIG. **10**), across which any actual cathode charge imbalance or potential in operation may be detected and measured.

Therefore, in light of the Faraday shielding principle and the said equal and opposite voltages which occur at opposite ends of the rotor segments (**14**), the potential difference between the rotor field emitters (**17**) and the inner surfaces of the two negative hull sections (or Emitter Ring (**47**)) will tend to become equal (and opposite) in operation to that across the primary electrode systems. And thus, a suitably vast and engineered differential in the electron emissivity of the field emitters (**17**) and the said ground-referenced primary cathodes (**64**) will define the primary voltage expansion ratio which may be obtained (in the manner first outlined in the section on Method of Operation). This topic will be discussed in more detail in a section to follow.

Moreover, a basic field-envelope-circuit generator design is disclosed which thus far, in accordance with Faraday's and Lenz's Laws, can produce a very high DC primary voltage with minimized circuit AC current inductance losses. The Primary Power System's DC voltage is not readily available, however, for purposes of normally-usable electrical output. Brushless electrode rings (**58**) must once again be employed to extract any portion of this primary voltage for off-rotor use. The Primary Power System schematics shown in FIGS. **8** and **9** in conjunction with FIG. **10** best illustrate the use of these two additional stationary anode rings (**58**), one suspended adjacent to each inner rotor induction cathode ring (**59**), which allow a large electrostatically-induced positive DC voltage to be 'picked off'. This positive voltage may then be used as a source of low power within an external circuit, provided that circuit is negatively-grounded and only to the Generator's chassis, and is therefore really intended for auxiliary onboard usage only.

It can be seen thus far, and especially in FIG. **7**, that there are essentially two major systems involved in creating the voltages which must be applied to the positive and negative housing or hull sections in order to form the electrodynamic field that will surround the Generator in operation. These two systems are:

[A] the Primary Power System, which generates a large DC voltage by electromagnetic induction and then electrostatically induces a major portion of that voltage upon the rotor segments (**14**) between two pairs of induction anode (**61**) and cathode (**28**) rings which are conductively affixed thereto (about the inner and outer circumferences of the rotor assembly, respectively); and

[B] the Field Induction System, which utilizes the potential difference which then exists between the positive inner circumference of the rotor—and its two attached primary anode rings (**68**)—and two separate sets of highly electron-emissive stationary cathodes (**64**) which are maintained at chassis ground and which are circularly-arranged near and electrically-connected to the periphery of each central collector housing section (or "zone") so that they may jointly charge their respective zone sectors (**44**) and/or (**42**) positively by stripping them of native electrons.

Each of these primary electrode array cathodes (**64**) connect directly to a field power resistor (**63**), which in turn connects to one positive radial zone sector (**42**) of the central housing sections. Each such cathode (**64**) is also vertically aligned inboard with a group of three stationary plane-parallel grid electrode elements and one of the two innermost rotating induction anode rings (**68**) conductively-affixed to the rotor segments (**14**) (about their positive inner circumference). Each such group of three grid elements is comprised of a control grid (**65**), an accelerator grid (**66**), and a suppressor grid (**67**), with one such combined grid element assembly per radial zone sector (**42**).

Together with the said primary induction anode ring (**68**), each such 5-electrode group constitutes what is termed a "unit pentode array" ((**69**) as seen in FIG. **7**). The unit pentode arrays (**69**) and their associated frame-mounted power resistors are thus arranged in two ring-shaped sets, located between the rotor and the two positive housing sections, which are referred to throughout as the "primary induction ring rays".

The potential difference between the unit pentode array cathodes (**64**) and rotor primary anode rings (**68**) is then referred to as the "primary array voltage", as the unit pentode arrays of each primary induction ring array are connected in parallel electrically (between the field envelope and rotor), and the number of unit pentode arrays should be equal to the number of like sections into which the Generator's superstructure is divided.

Electron vacuum tube design, construction, and operating methods may be employed to induce and/or modulate a substantial DC bias upon the plate current(s) of the several unit pentode arrays, which electrically link the polarized rotor to the central housing sections, for purposes of rendering the external electrodynamic field variably non-isometric and therefore directionally propulsive. Each such "stationary" 5-electrode system or unit pentode array (**69**) (whose anode element actually rotates) should be constructed in such a manner that it exhibits a minimum design amplification factor (μ) equal to 12.0, again with reference to any AC voltage or signal present in the Field Induction System circuit and as a function of its engineered relative electrode spacing(s).

The rotating primary induction anode rings (**68**), stationary cathodes (**64**), and various grid element wires which are associated with the stationary multi-electrode systems should be composed of tungsten, due to the extremely heavy conduction current across these primary arrays. Ceramic posts, pins, or brackets composed of a specialized material such as Titania may once again be used to support all of the various stationary electrodes and grid elements, and each unit pentode array's grid wires should be vertically aligned (or mutually-shading) similarly to those in a "beam power pentode" vacuum tube.

As generally depicted in FIG. **1**, a primary thermal conduit (**48**) with an imbedded coolant-carrying core (**49**) encircles and mechanically supports each unit pentode array, its power resistor (**63**), and that resistor's enclosing dielectric buffer (**45**). These thermal conduits and dielectric buffers will be examined in further detail hereinafter.

The operation of the Field Induction System is regulated by a central Field Voltage Control System which is modular in nature, as illustrated in FIG. **10**, and which is therefore mechanically and electrically connected to the stationary cathodes and grid elements of each unit pentode array (**69**). In this regard, the auxiliary stationary anode rings (**58**) described above are to serve as the source of positive voltage applied to the accelerator grids (**66**) of said unit pentode arrays, and such Field Voltage Control System is designed to directly accept this voltage for that purpose. The Field Induction and Voltage Control Systems will be discussed further in sections to follow.

Zone Sector Construction

Referring now generally to FIGS. **3** and **4**, one major goal in dividing each of the two hull positive zones or sections (**5**) which jointly comprise the Generator's Field Hub into a large number of radial sectors (**42**) in electrical parallel is to limit the external field current reaching each particular power resistor and unit pentode array combination to a fairly uniform level, while somewhat reducing field current return eddy losses as well. However, each radial sector (**42**) is thereby also given the capability of effecting a local thrust production differential (in propulsive three-stage rotor devices), by virtue of its electrical isolation, as it is then possible to vary the local resistance presented to the electrodynamic field with respect to any given radial sector so that a substantial measure of navigational control in the z-axis may thus be achieved by varying the proportional field current conducted by that sector.

Because the inherent resistance of each such sector itself will be negligible even at an extremely elevated operating temperature, in part due to the very low temperature coefficient of resistivity of the proposed steel hull plate material, the voltage drop across the radial length of each sector will also be negligible. Thus, the determination of an appropriate sector thickness becomes an entirely structural consideration. A secondary design goal with respect to zone sectorization is therefore to select a uniform cross-sectional area for the radial zone sectors (**42**) which yields a Field Hub total conduction mass approaching that of the device's rotor segments (**14**).

The hull plating which forms each positive zone (or section (**5**)) of the Generator's Field Hub should preferably be composed of stainless steel or a refractory metal, and divided into thirty-six equal-area radial zone sectors (**42**) of truncated wedge shape and a circular center zone sector (**44**). As stated earlier, in the single-stage Generator embodiment these two center zone sectors must be of a nonconductive material or construction, and should also comprise approximately 4 to 5% of the total area of their respective zones to allow the connection of the exchange or main service manifolds described above. In the three-stage embodiment such center sectors (**44**) must instead be conductive as just prescribed for the radial zone sectors (**42**), and should each comprise about 1% of the said zone area. Each such positive radial sector should be formed of a single piece of the hull plating material in such a manner that its cross-sectional area in the direction of its radial length is maintained at a very uniform value.

The uniform thickness of each center zone sector (**44**) should be equal to that of the radial sectors as measured at the minor (inner) arc width thereof, and the radial sectors must be insulated from the respective center sector where they meet that center sector's periphery (except in single-stage rotor devices). A suitable thin ceramic insulating partition (**43**) may therefore be installed between each adjacent pair of radial zone sectors (**42**), and around each center sector (**44**) in three-stage devices, and the uniform flat thickness of these sector insulators should in general be minimized.

In three-stage devices, four radial sectors (**42**) of each positive housing zone which correspond to cardinal points separated by 90 degrees of the hull's circumference, and which divide the Field Hub into four equal quadrants (as indicated with phantom lines in FIG. **4**), should be connected in parallel with the center sector of that positive zone at the maximum positive voltage end of those four sectors' respective power resistors (**63**).

The parallel resistance of the four conductors which therefore connect each center zone sector (**44**) (in three-stage devices) to its respective network of power resistors should then be such that that sector's per unit surface area level of field current conduction is slightly higher on the average than that of its adjacent radial sectors (**42**), under uniform zero-signal field bias conditions as hereinafter described.

The four unit pentode arrays (**69**) of each primary induction ring array which correspond to the power resistors so-connected to that primary array's center zone sector (**44**) would generally not be utilized individually in the active field current bias modulation related to the production of local thrust differentials (in three-stage rotor devices), being instead reserved primarily for potential involvement in signal communications transmission and/or reception activities.

Power Resistors

As best seen in FIG. **1**, the cathode element (**64**) of each plane-parallel electrode system or unit pentode array which is used to link the rotor assembly with a positive housing collector section or radial zone sector (**42**) is electrically connected to such positive section or sector by means of a ceramic resistor block (**63**), of a certain low dielectric constant and high volume resistivity, which becomes a poor but effective conductor at the elevated operating voltage and temperature of such housing collector sections and which ensures an adequate external field current voltage drop between such section or sector (**42**) and said cathode element (**64**) (which is chassis-grounded as shown in FIG. **10**).

Referring still to FIG. **1**, each said ceramic resistor block or "power resistor" (**63**) is to be entirely jacketed by a hollow ceramic thermal conduit (**48**), including a dielectric buffer (**45**) composed of a ferroelectric material having a certain high dielectric constant and very high volume resistivity which does not become a conductor even at the elevated operating voltage and temperature levels of said power resistors. These thermal conduits (**48**) mechanically connect the power resistors to a flattened inner surface of such positive hull section or radial sector (**42**) and a form of very-high-temperature heat transfer fluid such as liquid sodium may be circulated therein so that the proper resistor temperature is maintained to ensure an optimum cathode potential with respect to ground.

A suitable manifolding, pump, and heat exchanger system (not shown) may be used to extract or transfer an extremely high level of recoverable thermal energy from the power resistors (**63**), arising from the field current voltage drop associated therewith, by means of such primary heat transfer fluid circulated through the thermal conduit cores (**49**) in the manner normally associated with nuclear power plants. Similarly, the rate at which such primary hull coolant is circulated may be used to directly regulate and limit the external field current to a value within the safe ampacity of the rotor assembly.

It is important to point out that each of the Generator's power resistors (**63**) and its associated dielectric buffer (**45**) and primary thermal conduit (**48**) may be designed, as an assembly, to incorporate a fixed positive reduced-voltage tap (not shown), whereby a main source of onboard or output DC or AC electrical power may be provided either for or from the Generator within a limited-duty circuit which is grounded directly to the primary cathodes. Provisions should and may also be provided for biasing, damping, and/or inductively coupling any primary conduit connecting sections, to control or harness electromotive currents induced in the highly-conductive primary coolant. Onboard electrical resistive heating means should and may be incorporated for purposes of pre-heating the primary coolant and, in turn, the power resistors and dielectric buffers.

Referring briefly to FIG. **11**, it is also important to note that the Generator's power resistors must be in a conductive state before the housing potential difference reaches breakdown field intensity, as figured along a minimum semicircular arc trajectory (**142**) across the hull neutral sections (**4**), and the flow of field discharge current commences. Therefore, the actual maximum dielectric constant of the resistor material used in very-high-voltage single-stage rotor devices, and in ultra-high-voltage three-stage rotor devices as well, may be calculated at about k=11.

As is best seen in FIG. **14**, the recovery thermal conduits (**79**) of the secondary hull coolant system which assist in supporting the entire tapered peripheral portion of the housing must pass inboard between the power resistor assemblies referred to above, and therefore the resistors (**63**) must assume a complex tapered shape to provide the necessary clearance. As generally depicted in FIGS. **1** and **2**, they will have a trapezoidal cross section at the 'top' or outward-facing end (with radial length somewhat greater than minor circumferential width) and a rectangular cross section at the bottom or cathode end (with circumferential width much greater than radial length). In this regard, their engineered shape should still maintain a uniform conduction cross section exactly analogous to that of the preferred positive zone sector construction described above.

Dielectric Buffers

Referring once again to FIG. **1**, it is imperative to electrically insulate the power resistors (**63**)—due to the extreme voltage drop that will occur across them in operation—to prevent direct discharge to other nearby structural components from taking place. However, not only is this a problem due to the thickness of insulation which would be required at any 'ordinary' dielectric value, but the power resistors' projected operating temperature range of 600 to 700+° C. completely precludes the use of almost all known dielectrics. Fortunately, the search for high-temperature dielectrics for use in aerospace electronics applications (particularly multi-layer capacitors) has resulted in the development of a small number of exotic materials potentially suitable for use as a space charge buffer (**156**) or dielectric encapsulant (**45**) for the Generator's power resistors (**63**).

In the application contemplated herein, the volume resistivity of the resistor encapsulant is somewhat secondary to its dielectric constant k at high temperature, because no opposing contact voltage is present (as in a capacitor) to encourage circuit conduction losses within the material. Therefore, the prime consideration is the dielectric's ability to exhibit a very large k value at over 600° C. This places the material squarely in the realm of a very small select group of Class III high transition temperature ferroelectrics. These materials, such as Tantalum-modified Lanthanum Titanate and Lead Ytterbium Tantalate, do not begin to demonstrate significant dielectric strength at less than 300 to 450° C. One such compound, Sodium Bismuth Titanate ($Na_{0.5}Bi_{4.5}Ti_4O_{15}$), actually exhibits an amazing peak k value of about 3100 near the middle of the said power resistor operating temperature range: at 655° C. (1,202° F.). This would seem to indicate that this compound is eminently suited for use in forming the necessary rotor space charge buffer pieces (**156**) and/or resistor dielectric buffers (**45**) in this application.

It should be stressed that great care in quality control must be exercised in formulating and sintering this or any similar such special compound during manufacture of the dielectric buffer pieces. An absolute minimum of impurities must be assured, the density of the constituent powder maximized, and each piece finally certified as being free of the smallest physical defect prior to use in order to attain the desired operating performance. It is also hoped that these crucial components may each be compaction die formed and sintered simultaneously with a respective power resistor (**63**), and perhaps even with the corresponding primary thermal conduit (**48**) as well.

The core (**49**) of the thermal conduit associated with each power resistor (**63**) should be constructed of a refractory metal or alloy tubing (such as molybdenum) which is jacketed with high-purity Alumina or a comparable material formed directly thereupon. Thus, each such primary thermal conduit becomes an extremely strong structural member which may also be used to support the stationary electrodes of the corresponding unit pentode array, while circulating the preferred liquid sodium heat transfer fluid or coolant around the outward faces of that power resistor to recover or remove its excess heat.

Each power resistor (**63**), its dielectric buffer (**45**), and associated primary thermal conduit (**48**) together comprise one power resistor assembly, and each such assembly in combination with its corresponding unit pentode array is considered for present purposes as forming the constituent component group of which the primary induction ring arrays and Field Induction System are comprised.

Extensive calculations reveal that a suitable power resistor material in most sizes of the single-stage embodiment of the Generator would be a ceramic material with the chemical formula $MgO.SiO_2$, which is available from CeramTec NA as CeramTec designation Steatite 357. A suitable power resistor material in most sizes of the preferred three-stage embodiment of the Generator would be a somewhat different ceramic material with the chemical formula $2MgO.2Al_2O_3.5SiO_2$, which is available from CeramTec NA as CeramTec designation Cordierite 547.

Overall Description of the Field Induction System Circuit

Before an overview of the EDF Generator's Field Induction System electrical circuit is presented, a somewhat comprehensive examination of the nature of the housing or hull configuration should be made, as in this application the hull itself is an active part of that circuit. Its electrostatic characteristics as a function of such configuration will therefore have a fundamental influence upon the nature of the external field envelope discharge current itself.

Hull Design

The derivation of the Generator's design hull configuration, as best seen in FIG. **3**, comprises a long and complicated story. This should not be taken to imply that the final and rather exacting composite shape selected is the only one that 'works' properly. It merely reflects the preferred embodiment of the inventor's conceptualization of a fundamentally mechanical device which happens to entail both tremendous electrical complexity and inspiring aeronautical implications.

Should the proposed vessel eventually be used to explore the possibility of transcending the time/light barrier, extremely precise calculations regarding its spatial displacement and charge/mass ratio will no doubt become necessary. Therefore, one of the principal reasons for the chosen hull design is that the total vessel displacement may readily be calculated using traditional formulas from analytic geometry for the volumes of the two truncated right conic sections and two one-base spheric zones, and the volume of the remaining cylindrical central chamber region.

Using just these formulas, a vessel hull configuration has therefore been developed which may be linearly expanded (as a scalar function of the hull radius) to virtually any size with no significant loss of accuracy. The Table of Dimensions which follows and the attendant schematic hull diagram of FIG. **3**, upon which all of the subsequent Detailed Calculations quoted are based, illustrates the application of this design technique to a theoretical vessel hull only 48" in diameter. It must be emphasized that the construction of an Electrodynamic Field Generator this small might not prove entirely practical in reality, due strictly to mechanical limitations, and the inventor believes this prototype model to be in fact the smallest such machine that could be built. This original hull size was deliberately selected, however, to encourage the development of the maximum possible accuracy with respect to contingent specifications for the relative size and positioning of the Generator's principal components, prior to the actual construction of larger machines.

The surface areas of the negative **(3)** and positive **(5)** hull sections are set equal to each other, and for two very important reasons: [i] to give the vessel hull a significant theoretical capacitance (despite its unusual geometry); and [ii] to render the two "driving" or discharge current portions of the field envelope uniform in cross section. Essential considerations which are related to these fundamental design criteria are best expressed in the following electrostatic formula for uniform field intensity in a DC capacitor which exhibits a continuous leakage current:

$$E^2 = V\sigma/d\epsilon_o,$$

where V=equilibrium potential difference in volts,

σ=average surface charge density of one plate,

d=plate separation in meters,

ϵ_o=universal electrostatic constant,

and E=uniform field intensity in volts/meter.

This equation plays an important supporting role in the invention's Method of Operation, as previously discussed in that section above.

The volumes of the peripheral right conic sections of hull volume surrounding the Field Hub and and those of the one-base spheric zones comprising the Field Hub itself are also set equal in a preferred embodiment, as this criterion is postulated to promote both the structural integrity of the hull in any attempt to breach light speed (and establish a stable Kerr metric space warp about the vessel) as well as the gravimetric stability of the central chamber/cabin area in light of the relativistic mass effects involved while doing so.

As the case may be, to achieve this equal-area/equal-volume design, it was found necessary to employ two crucial interrelated constants (in addition to the ratio ⅕ used to radially "step off" the hull's interior space): the Hull Area Constant and the Polar Hull Constant. The Area Constant is (as the name implies) necessary to achieve the equal-area part of the overall design solution, and merely dictates the small deviation of the neutral **(4)** and negative **(3)** housing sections' ring radii from the stated preferred design increment of the hull's radial width.

The other requisite design factor, the Polar Hull Constant, is much more complex in nature. This value specifies the maximum height of the spheric hull zones **(5)** as a function of their arc-shape, and the corresponding contingent Polar Volume Differential is absolutely essential to achieving the equal-volume aspect of the preferred design. It is important to note that certain allowances must be made in practice for a finite but electrically-significant peripheral edge 'thickness' (and therefore surface area) at the equator of the device's negative housing Emitter Ring **(47)**, but will in general only minimally modify the "pure" design-model values obtained using the mathematical formulas given in the Table of Dimensions to follow.

Hull Configuration

Referring now to FIGS. **3** and **4**, the exterior housing or hull **(1)** is designed specifically to have a circular disc shape the thickness of which along the vertical axis of symmetry (that also constitutes the rotor's axis of rotation) is relatively minor as compared to the diameter along its radial centerline plane, and which tapers gently from a maximum thickness along such vertical centerline to a very minor thickness at the radial periphery **(47)** of the housing.

The tapered disc shape of the hull is postulated to promote the formation and maintenance of a stable bihemitoroidal corona or arc discharge field based upon certain empirical principles of electrostatics and upon laboratory experiments performed by the inventor involving concentric circular electrodes. Such an electrodynamic field may be characterized electrostatically in this application by an axial positive polar electric field contiguous with a biplanar negative electric field, with each field of one given polarity tending strongly to become oriented parallel to a field of opposite respective polarity but similar intensity.

The exterior housing or hull **(1)** is therefore divided into two axially-central positive electron collector sections **(5)**, as well as: [i] either a single radially-peripheral negative electron emitter ring **(47)** which is biplanar in cross section or two or more such ring sections **(3)** which are uniplanar in cross section having separate planar orientations; and [ii] two dielectric neutral sections **(4)** which are interposed between said positive and negative sections such that each such section of a given polarity is spatially separated from each section of the opposite respective polarity.

The total external surface area of the housing's negative Emitter Ring **(47)** or sections **(3)** is essentially equal to that of the positive collector housing sections **(5)** or zones, such that the overall housing configuration thereby acquires a significant theoretical static capacitance by design as intended. An estimated base value for such hull capacitance 'C' may be obtained by applying the surface area 'A' of the hull's two sets of "plates", as figured using the formulas in the Table of Dimensions below, to the standard formula for parallel-plate capacitance ($C=\epsilon_o A/d$). When plate separation 'd' is taken to be the simple average of the longest **(141)** and shortest **(142)** purely-semicircular arc current trajectories across the hull, as depicted in FIG. **11**, this estimate of hull capacitance is approximately 13 to 14 mmf for an EDF Generator 48" in diameter.

Field Composition

Referring to FIG. **11**, the external discharge current once initiated may be expanded as a function of the operative field voltage to fill two hemitoroidal spatial volumes of rotation **(140)** whose outer perimeters **(141)** are defined by a semicircular electron arc trajectory across the full housing radius in a plane perpendicular to the housing surface, or from the outermost point on the negative periphery thereof to the respective positive centerpoint of the housing lying on the vertical centerline thereof, and whose inner boundaries

(142) are defined by a similar such trajectory across the hull's neutral ring or sections (4) from any innermost point on the negative periphery to the closest point thereto lying on the respective positive collector section (5).

The physical volume (140) of the external discharge current as described has a uniform cross-sectional area in the radial direction of electron conduction which corresponds to any such semicircular arc trajectories between points on the housing's negative Emitter Ring (47) and the respective positive collector housing sections (5) and therefore, by the definition of a conductor, such discharge current must assume a uniform current density. While very difficult to effectively describe in verbal terms, the idealized mathematical configuration of the Generator's heat- and thrust-producing field envelope region (as a function of its housing geometry) is visually easy to grasp and is clearly illustrated in FIG. 11.

As indicated in FIGS. 12 and 13, stationary electromagnetic armatures (37) are uniformly distributed into two circular arrays (145) containing equal numbers thereof which are positioned on opposite sides of the rotor from one another, and which are each located concentric to the rotor's axis of rotation midway between the negative Emitter Ring (47) and positive housing sections (5). These armature arrays (145) may be employed as two separate groups to impart a holistic but attenuating rotational magnetic vector moment upon the external discharge current, reaching through the thin neutral sections (4) of the housing, and thus may be used to modulate the field envelope's kinetic and electrical characteristics in a variety of ways useful to the invention and its objects.

FIGS. 12 and 13 also show the electric field vectors (143) as described above, and magnetic field vectors (144) as produced and modulated by the armatures (37) comprising the two armature or variable inductor arrays (145). FIG. 13 specifically depicts the regular lateral bending or displacement of the radially-impinging electron trajectories (143), under the influence of the inductor arrays (145), which is postulated to have a unitizing effect on the driving portions (140) of the field envelope.

It can therefore be seen that the device contemplated herein may be operated in such a manner that its external breakdown discharge current field achieves a uniform conductive cross section and current density, as well as a holistic and orderly rotational aspect which subsequently acts to assist in rendering such current quasi-coherent. And thus, for purposes of this device and its applications, such a discharge current field should be thought of as a special qualified type and form of DC corona or arc discharge—having significant propulsive potential (to be further examined)—which then merits technical description as an electrodynamic field.

In this respect, it may for now be emphasized that a useful level of electrically-developed thrust may be realized from the relativistic impulse of the electrodynamic field current electrons incident upon the positive collector housing sections and, postulating the operation of the device in the vacuum of space (with a physical field current volume as described), the impulse velocity of such incident field electrons may be raised to as high as 99.99% of the speed of light and more with the corresponding relativistic mass thereof then equal to 69 times their rest mass or indeed much more.

Primary Voltage Expansion Ratio

In keeping with the logic presented at the opening of the Field Induction System section above, the primary array voltage will tend to equal only one-half (½) of the DC voltage expressed across the rotor segments (or "rotor voltage"). Therefore, to develop an extraordinary multi-megavolt 'secondary' or field voltage across the Generator's housing of the type just described, the primary array cathodes' electron emissivity (expressed as a certain factor) must exceed that of the rotor field emitters by a ratio which is at least as large as the ratio of the desired field voltage to half the observed said rotor voltage. The latter such ratio is then thought of as the primary voltage expansion ratio in this application, as outlined in the Method of Operation section hereinabove.

A difference in the instantaneous rate at which these two said sets of parts will tend to discharge under the impetus of a given equivalent voltage or potential difference is thereby created. This instantaneous charge differential enables the primary cathodes to strip native electrons from the positive housing sections (5) thousands of times faster than such charge can reach the negative emitter housing sections (3) from the rotor field emitters. Thus, an instantaneous charge differential is perhaps best described as a cumulative charge imbalance that is known to operate (under certain conditions) in capacitive and/or thermoelectric circuit arrangements and to express itself as an elevation of the applied voltage or potential difference therein.

To actually achieve and support such a condition requires the use of two housing charge ballast capacitors whose negative 'plate' elements or rings are affixed directly to the rotor segments (about their outer perimeter) and which are immediately able to store an amount of such charge that is sufficient to produce equilibrium collector surface charge densities resulting in the desired field voltage. This average required housing charge density is calculated using the traditional formula given in the Hull Design section above.

The voltage applied to the positive rings of these ballast capacitors is taken from the rotor's outer field coil anode ring connections and therefore represents a large portion of the Primary Power System's generated field winding voltage. It should be noted that this voltage must actually be passed to such capacitors beneath the rotor ballrace assemblies in the form of recessed conductive traces on one or more of the segment separators. This same technique must also be employed to connect electrical elements of the Primary Power System lying outboard of the variable inductor arrays (145) with corresponding elements lying inboard thereof.

In any event, these crucial ballast capacitors are therefore properly thought of as belonging to both the Primary Power System and Field Induction System circuits. They provide the necessary electric force means by which a given primary voltage expansion ratio (as defined) may be supported and realized using the engineered instantaneous charge differential concept described. Because of the complexity of this aspect of the Generator's operation, the specific example below will illustrate the relevant principles using the respective temperature emissivity factors of the chosen field emitter and primary cathode materials with respect to the aforementioned 4ft. prototype model.

This temperature emissivity factor, $e^{-\phi/kT}$, where 'e' is the base of natural logarithms 2.71828 . . . , 'k' is Boltzman's constant, and 'T' is the absolute temperature in °K, derives from the famous Richardson-Dushman equation: a formula for the correct thermionic emission current density 'J' of a clean metal cathode in vacuum having surface work function ϕ. And, expressed in mathematical terms, the principle of ratio equivalence defined above may be stated as follows:

$$\text{Primary Voltage Expansion Ratio} \approx \text{cathode } e^{-\phi/kT}/\text{emitter } e^{-\phi/kT}.$$

Using field emitters of sintered 0.68/0.32 tungsten/copper composite and thorium-adsorbed tungsten primary cathodes, a 4 ft. diameter three-stage-rotor prototype unit of the invention's design would achieve a primary voltage expansion ratio equal to about 12,106 given a field emitter $\phi=4.408$ eV and primary cathodes of $\phi=3.639$ eV with both component sets at a temperature of 948° K (or 675° C.) as figured per the above equation. The field voltage expressing this ratio is verified in the Detailed Calculations hereinbelow.

It is important to note that this equation can readily be solved for the required primary cathode ϕ where the projected mean operating temperature of the primary cathodes is given (as it would be in operation), and where the field emitters' ϕ is also fixed and known but their actual operating temperature within a range of possible values must be experimentally or theoretically obtained.

The external breakdown discharge current once initiated and sustained in this manner is generally limited only by the engineered design characteristics of the unit pentode arrays which comprise the Field Induction System. In particular, the negative DC bias voltage applied to the unit pentode array control grids **(65)** must be sufficient to limit the total primary induction ring array current to a value within the safe operating ampacity of the rotor. For purposes of this text, the term "full-power" as used in relation to the Primary Power System and its incorporating device shall be taken to indicate a corresponding rotor speed wherein a full nominal field envelope voltage is achieved, maintained, or exceeded. Such an operational state is depicted in FIGS. **11**, **12**, and **13**.

Field Current Bias

In a simple single-stage rotor embodiment of the Generator used as a "Thermal Power Unit", no active biasing of the current in either of the two current-carrying portions **(140)** of the field envelope (which are otherwise symmetrical) need take place at the primary arrays, and the current comprising each such driving portion of the field is essentially equal. In the preferred three-stage rotor embodiment intended for propulsive units, the field current is actively biased or proportionally shunted by the primary arrays between such field envelope portions **(140)** to render the impulse thrust developed thereby mutually non-isometric along the Generator's vertical centerline. Several important considerations directly related thereto are discussed below, and reference should generally be made to FIG. **7** in this section.

The Generator's maximum full-power total field current must be equal to the rated DC parallel ampacity of the rotor segments, as figured using traditional methods. An operational minimum full-power field current in any propulsive three-stage rotor device used as an "Impulse Drive Unit" should then be considered as equal to either one-half (½) of the so-rated rotor ampacity or a level of total rotor current sufficient to allow the Generator to produce a net vertical thrust equal to its own weight, whichever is the lesser.

Given the minimum engineered amplification factor of the unit pentode arrays **(69)** earlier stated, this minimum full-power field current may in all likelihood be maintained with an average value of primary array negative bias voltage (as applied to the stationary control grids **(65)** shown in FIG. **7**) which is less than or equal to about one-half (½) of its specified design-maximum value. If no unidirectional or alternating voltage component of the Field Induction System current is present, this said average negative bias voltage value constitutes the primary arrays' "zero-signal" bias voltage and the primary arrays may be said to be in a zero-signal field bias condition (as mentioned earlier hereinabove).

Each of the two said driving portions **(140)** of the Generator's field envelope may be referred to as a "field hemitorus" because of its postulated shape, and the maximum individual field hemitorus current at any given value of operating rotor current should at all times be limited to a level which is less than or equal to two-thirds (⅔) of that operating rotor current. Peak net field thrust will then equal one-third the total isometric thrust. Assuming the said given amplification factor, this maximum advisable hemitorus level of proportional total field current may in all likelihood be maintained with an average value of negative bias voltage applied to the corresponding primary array's control grids **(65)** which is equal to approximately one-third (⅓) of its specified design-maximum value (or less proportionally at a rotor current level which is less than rated ampacity).

Due to the lag time in heating either of the two power resistor **(63)** networks through which an increasing field hemitorus current must return, the negative bias voltage applied to the corresponding primary array's control grids **(65)** may be briefly reduced to as little as one-sixth (⅙) of its specified design-maximum value (as a standard minimum operating control grid voltage) or less proportionally at a rotor current level which is less than rated ampacity.

The minimum individual field hemitorus current at any given value of operating rotor current should at all times be maintained at a level equal to or greater than one-third (⅓) of that operating rotor current. Assuming the said given amplification factor, this minimum advisable hemitorus level of proportional total field current may in all likelihood be maintained with an average value of negative bias voltage applied to the corresponding primary array's control grids **(65)** which is equal to approximately two-thirds (⅔) of its specified design-maximum value (or less proportionally at a rotor current level which is less than rated ampacity).

Due to the lag time in cooling either of the two power resistor **(63)** networks through which a decreasing field hemitorus current must return, the negative bias voltage applied to the corresponding primary array's control grids **(65)** may be briefly increased to as much as five-sixths (⅚) of its specified design-maximum value (as a standard maximum operating control grid voltage) or less proportionally at a rotor current level which is less than rated ampacity.

It is important to note that the current passed by the unit pentode arrays **(69)** will depend to a great degree on the accelerator grid **(66)** voltage rather than on the overall primary array potential difference, just as in any standard vacuum tube which employs a screen or accelerator grid whose positive voltage is less than the applied plate voltage. It therefore becomes evident that this positive grid **(66)** voltage may also be modulated in a manner similar to the negative control grid **(65)** voltage. This type of 'duplex' signal-handling or control voltage response capability allows the unit pentode arrays (of either primary induction ring array) to amplify a resonant frequency communications signal while simultaneously controlling an independent level of current-contingent thrust.

In any event, it can be seen here in light of the Law of Conservation of Momentum that a sizeable force of thrust will be developed by each field hemitorus **(140)** from the relativistic impulse of its constituent electrons incident upon the collector housing sections **(5)**, and that suitable means have been provided for rendering such mutually-opposed y-axis thrust vectors variably non-isometric.

A less obvious but very important distinction which should be made in regard to the EDF Generator's production of such electrically-developed thrust is certainly this: it is not necessary that the Generator or in particular its positive hull sections (**5**) (or Field Hub) provide the work required to move the field current charge against the field potential gradient. Per classical electric field theory, the work may be done by the charge itself in being repelled along the potential gradient. In the first case, work is 'performed' and in the second it is 'recovered'. Furthermore, it can be shown mathematically in light of the preceding distinction that any net recoil force experienced by the negative hull Emitter Ring (**47**) is strictly Newtonian in nature, compared to the relativistic impulse thrust produced at the Field Hub, and that the peak value of such recoil would amount to no more than about $^{3}/_{1,000}$ of one percent of the peak net value of the said thrust.

Therefore, the Electrodynamic Field Generator's input rotary torque drive means will provide the work-energy required to establish and maintain the rotor's rotation and the resultant field potential gradient that is induced in the manner described, and the electron charge comprising field hemitorus current will then do the work required to effect its passage toward the positive Field Hub. The resulting collisions are almost completely inelastic, so momentum and kinetic energy are conserved independently, and the field current's gained kinetic energy is recovered almost entirely as heat.

Field Voltage Control System

Accordingly, it can also be seen that the interactive factors or features which result in the EDF Generator's exhibited field voltage as a function of its rotor voltage are much more complicated in nature than the simple turns ratio which defines an AC transformer's secondary voltage in terms of its primary voltage. And because of the extremely high primary voltage expansion ratios which will be operative in large devices, small fluctuations in the Generator's stationary primary array electrode voltages can conceivably cause very large and undesirable fluctuations in the exhibited field voltage. Therefore, a relatively simple but effective Field Voltage Control System is provided which may be used to monitor, regulate, and adjust all individual primary cathode and grid voltages to optimum specified and/or interim operating values such that rotor ampacity is not exceeded, and this system anticipates the use of an associated onboard computer system (to render its various interactive functions automatic).

The embodiment of such a control system as shown in FIG. **10** is intended merely to provide a logical circuit model which is illustrative of the principles of the invention, and of certain principles of appropriate circuit resolution which should be applied within such a system in light of the following aspects of its teaching, and not to preclude the use of other possible embodiments of such control system circuitry.

The cathode elements (**64**) of each plane-parallel electrode system or unit pentode array (used to link the rotor assembly to the positive housing sections) are to have a surface work function ϕ which is significantly lower and/or an operating temperature which is significantly higher than that of the rotor field emitters (**17**), and will therefore tend to exhibit a comparatively much-higher electron emissivity, such that the ratio of their respective temperature emissivity factors (which are each represented by $e^{-\Phi/kT}$) has a very determinate effect upon the respective exhibited ratio of the device's external discharge field voltage to the primary DC voltage generated internally.

Referring to FIG. **10**, these primary array cathodes (**64**) must be composed of tungsten or another refractory metal which may also be either impregnated with thorium oxide or thorium metal submonolayer-adsorbed to achieve such lowered comparative surface work function, and, to calibrate their emissivity relative to that of the rotor field emitters (**17**) as a means of defining and determining the device's external field voltage relative to its primary array potential difference (or half the rotor voltage).

The Field Voltage Control System as shown is basically designed to directly accept the rather high positive induced voltage of the stationary anode rings (**58**) (shown in FIGS. **8** and **9**) for the purpose of providing the primary accelerator grids (**66**) with their standard DC operating voltage. Due to the linear scalability of the Generator's overall design, the nominal (or zero-signal) value of such voltage relative to chassis ground (**10**) is expected to be approximately +362 to +483 volts per foot of hull diameter in three-stage and single-stage Generator embodiments (respectively). An isolation diode (**126**) and switch or relay (**124**) prevent such positive voltage from being neutralized by negative current (s) elsewhere in the control circuit.

Any one or more of the stationary electrode elements (**64**)–(**67**) comprising each said unit pentode array may be artificially cooled, as a means of further regulating the device's external field voltage, and/or subjected to modest variable DC control voltages at times to further assist in the interim adjustment of that unit pentode array's current.

A very high value capacitor (**116**) or a plurality of separate high value capacitors connected into a parallel or series-parallel matrix thereof may be employed to store the displacement charge caused within the Primary Power System by the charging of the rotor-mounted ballast capacitors, as part of such Field Voltage Control System. Said displacement charge holding capacitor(s) (**116**) may hereinafter be referred to as "ballast compensating capacitor(s)".

A very high value capacitor (**117**) or similar such capacitor matrix may also be employed to prevent the grounding of positive housing section potential by ambient ionization charge, from operation in air during the device's start-up period before such sections are fully enclosed (by the external field hemitorus currents), as a further part of this Field Voltage Control System. Such bulk electron storage capacitor(s) (**117**) may hereinafter be referred to as "ambient charge capacitor(s)".

The negative plates of these ambient charge capacitors are to be isolated from ground during rotor spin-up, in order to effect a net negative charge storage, and so their positive plate charge is supplied by a dedicated common variable DC voltage supply (**98**) which is oppositely connected to the ground frame. Each ambient charge capacitor (**117**) employed during the device's start-up period may be selectively switch- or relay-connected by its negative terminal to an accelerator grid element (**66**) of one or more of said plane-parallel electrode systems, in order to effect and control the storage of such ionization charge. Each such start-up capacitor (**117**) may also be selectively switch- or relay-connected by its negative terminal via a variable resistor and/or diode to a suppressor grid element (**67**) of one or more of said plane-parallel electrode systems, in order to effect and control the distribution of onboard stored negative charge reserves from the suppressor grid(s) (**67**) to the rotor, as a standard means of regulating the external field voltage (especially in the vacuum of space).

Said ambient charge capacitor(s) (**117**) may also be similarly connected by the negative terminal(s) thereof either to a separate source of true earth ground or to one or more superconductive current storage ring/s (**200**) (including charge deposition and retrieval means), for purposes of removing any excess ambient electron charge above the bulk storage capability of such capacitor or capacitor matrix which may be produced from normal or full-power operation in air—or other gaseous dielectric media.

Such superconductive current storage ring(s) would be used exclusively in large three-stage rotor propulsive models of the Generator, and thus constitute/s the means by which further expendable onboard charge reserves may be accumulated. The working core of each superconductive storage ring would preferably be a large but relatively thin torus (of small cross section) composed of solid mercury metal immersed in liquid helium, and as such would have a known ampacity measured in the hundreds of thousands of amps.

A linearly-actuated control rod (**103**) or similar mechanism may be used to effect a mechanically variable thermal junction between each primary array cathode or grid element (**64**)–(**67**) and a cold thermal mass of the same or similar refractory metal, maintained at the reservoir temperature of a secondary cryogenic housing coolant such as liquid air or nitrogen. The plate elements and dielectric media of each capacitor or capacitor matrix employed in the Field Voltage Control System could also be incorporated (in a suitable manner) into the coolant vessel which contains said cold thermal mass. This System's diodes would generally not be rated for mounting at such a low temperature.

The voltage control circuit shown in FIG. **10** is modular in nature, in that one such unit is provided for each of the 72 unit pentode arrays associated with the Field Induction System (in the preferred embodiment). All measurements of primary electrode voltage(s), and of positive housing section voltage, are to be made with respect to the Generator's ground frame (**10**) (as generally depicted in FIG. **1**): the metallic structural shell of the device's central chamber to which the rotor carrier assembly drive motor(s) should be mounted. This grounding method is referred to throughout as "chassis ground".

One thermal control rod (**103**) is provided for each primary cathode and grid element (**64**)–(**67**) which variably engages both a control rod socket (**104**) attached thereto and a thermal junction (**101**), which is attached to a cold thermal mass (**102**) maintained at the reservoir temperature of the secondary (cryogenic) housing coolant, by means of a linear actuator (not shown). Such control rod (**103**) may then be used to regulate that electrode's temperature to approximately the same temperature as the associated power resistor (at 675° C. ±55°). In this manner, small adjustments in cathode (**64**) emissivity may be made as necessary during operation and the various grids may be closely matched in temperature to that of their respective cathode(s) to ensure the accuracy of mutual primary electrode voltage balance. Any deviation of cathode (**64**) voltage from a chassis ground potential may be detected and measured across a very-high-value resistor (**107**), connected between said cathode(s) and chassis ground, which is preferably of a large carbon type having multiple fixed taps (and across which an output voltage may exist).

A dedicated common variable DC voltage supply (**95**) is provided for all primary cathodes (**64**) as a means of ensuring an optimum cathode potential with respect to ground, despite fluctuating power resistor temperatures (in three-stage rotor devices where field hemitorus current is variably biased). This supply is shunted with a double-pole/double-throw switch or relay (**123**) to enable either a positive or negative interim control voltage to be applied to the cathodes (**64**), in order to hold their true potential as close to ground as possible. A similar voltage supply (**97**) and switch/relay (**123**) is also provided for all primary accelerator grids (**66**) as a means of modulating the fixed level of applied positive grid voltage (from stationary anode ring (**58**)) and therefore the level of field hemitorus current, independent of the level of applied control grid (**65**) bias. In both cases, these DC supplies are isolated from chassis ground by blocking capacitors (**106**) to inhibit any net charge loss or accumulation from accruing to the ground frame (**10**) during operation (in propulsive units which are not connected to true earth ground).

A dedicated common variable DC voltage supply (**96**) is provided for all primary control grids (**65**) as well. This voltage supply is not capacitor-isolated from chassis ground in this case, so that the primary control grids (**65**) and cathodes (**64**) may have a common direct ground reference as they would in a typical vacuum tube circuit arrangement. The said control grid supply (**96**) should also have a resistor (**107**) of the same type as that provided for the associated cathodes (**64**), connected between such supply's positive terminal and chassis ground. These two criteria further ensure the accuracy of mutual primary electrode voltage balance, and the nominal (or zero-signal) value of such DC control grid voltage relative to chassis ground is expected to be approximately −36 to −48 volts per foot of hull diameter in three-stage and single-stage Generator embodiments (respectively). A variable bypass resistor (**108**) may be adjusted to ensure a very low and relatively constant DC operating grid current despite fluctuations in the value of the AC input signal resistor (**110**), as discussed below.

Ballast compensating capacitors (**116**) are provided whereby negative outrush charge from the stationary anode rings (**58**) and the positive plates of the rotor ballast capacitors may be purged from the Primary Power System during rotor spin-up, to prevent stage electrode array voltage ratings from otherwise being greatly exceeded, via these same two anode rings (one per primary induction ring array), connecting switch/s or relay/s (**124**), and resistor/s (**125**). During the Generator's spin-up and 'run' periods, each ballast discharge switch/relay (**122**) is normally open; switch/relays (**124**) are 'run-open'. Upon rotor de-spin, such ground-restoring charge must be proportionately returned to each stationary anode ring via switch/relays (**124**) and resistors (**125**), and to the Primary Power System as a whole via the discharge switch/relays (**122**) and suppressor grid discharge shunts (with variable resistors (**113**) and diodes (**115**)), for the most part after vacuum chamber current to the hull Emitter Ring has essentially ceased.

Then, such normalizing charge may be 'dumped' across the described suppressor grid discharge shunts at a rate sufficient to ensure that the entire Primary Power System returns to a ground potential (by brief direct surge rotor-shorting to the outer induction array anode rings). Thus, resistors (**113**) provide a variable time constant for the discharging of the said compensating capacitors (**116**). An a independent common variable DC voltage supply (**99**) whose negative terminal is connected to the ground frame (**10**) ensures that the proper total ballast compensating charge described may be stored against the high positive applied potential of the stationary anode rings (**58**).

The ambient charge capacitors (**117**) absorb ionization charge (arising from operation in air) which perforce must fall into the potential well(s) established by the positive housing sections at least until the field envelope's two hemitorus currents are fully formed, and which would otherwise tend to ground the desired positive housing section voltage. Therefore, the negative plate voltage of these storage capacitors (**117**) is applied from another dedicated DC supply (**100**), which is similar to that (**99**) provided for the compensating capacitors (**116**), by means of a common dpdt switch/relay (**119**). None of these DC supplies (**95**)–(**100**) are to have internal load-line diodes.

The variable output of this "dynamic compensating" voltage supply (**100**) is to be approximately the same as that of the ambient charge capacitors' positive voltage supply (**98**), thereby allowing the said capacitors (**117**) to be gradually saturated during rotor spin-up with electrons collected at the accelerator grids via storage shunt inrush current switch/relays (**121**) and diodes (**114**). During the Generator's spin-up and run periods, the "control gate" switch/relays (**119**) are normally closed such that the negative pole of supply (**100**) is connected to the negative plates of these same capacitors. In the Generator's run and de-spin periods, however, switch/relays (**121**) are normally open.

Any excess ionization charge above these capacitors' joint storage capability which might lower the 'run' Field Induction System circuit voltage must then be either grounded off or deposited in said superconductive storage ring(s) (**200**) (in ungrounded propulsive Generators designed for maximum ambient charge storage), via the same said throw of control gate switch/relay (**119**) in conjunction with shunt switch/relay (**120**) and resistor (**112**). In this case, the storage shunt switch/relays (**121**) remain run-closed. Otherwise, the slow 'run' release and distribution of negative charge reserves to the rotor from the storage capacitors (**117**) is effected by the gradual synchronized relaxation of the voltage/s from supplies (**98**) and (**100**) across switch/relay (**119**) and discharge resistor (**113**), with shunt switchs (**120**) and (**121**) open, while diodes (**114**) and (**115**) ensure that the said capacitors (**117**) may only be discharged in the manner described.

This discharge of stored charge reserves actually takes place from the suppressor grids (**67**), which may serve as low-power electron emitters given their proximity to the primary anode rings (**68**). An isolation diode (**127**) prevents such negative charge from neutralizing accelerator grid circuit voltage. It should be noted that the suppressor grids may be held at ground potential as they would in a typical beam power pentode circuit arrangement, or at a very low if not negligible negative voltage for purposes of effecting such controlled low-power charge release, by means of [i] a minor negative voltage induced thereon by the corresponding proximal primary anode ring/s (proportional to their capacitance relative thereto); [ii] the combined DC voltage drop across the paralleled resistors (**109**) and (**111**) connected to their respective accelerator grids; and/or [iii] the voltage of the dynamic compensating supply (**100**) relative either to earth ground and resistor (**112**) or to chassis ground and the relative opposing voltage of the ambient charge storage capacitor's supply (**98**).

The bypass resistor (**109**) provided between each accelerator grid (**66**) and its associated suppressor grid (**67**) returns the inevitable stationary anode ring (**58**) and accelerator grid (**66**) DC circuit current to the rotor from the suppressor grids (**67**), and simply assists in maintaining each suppressor grid at a DC potential as close to true ground as possible. The suppressor grid resistor (**111**) which is paralleled with said bypass resistor (**109**), however, is analogous to a plate resistor on behalf of the corresponding inner (rotor) induction array in both the single- and three-stage rotor embodiments, in that any rotor circuit final-stage AC output voltage variations will be reproduced across it once field envelope circuit current commences. Likewise, any AC signal oscillations from an outside (remote) source which are induced upon the field envelope current will immediately act as a detected first-stage AC input for rotor circuit amplification, and again will be reproduced across the said grid resistor (**111**).

Therefore, any rotor or field circuit AC potential difference across this grid resistor (**111**) provides an output signal voltage that may be: [i] received by an actual communications console or a simple AC control-voltage operating circuit in the Generator's central cabin; and/or [ii] reapplied either in- or out-of-phase with the voltage across the input signal resistor (**110**) of the associated control grid (**65**) as a further AC signal voltage amplification or suppression stage (respectively). An input AC signal or control voltage may also be applied across the control grid resistor (**110**) from such an onboard communications console or control-voltage operating circuit, as would generally apply to three-stage and single-stage devices, respectively. This grid resistor is variable for circuit tuning purposes related to resonant frequency communications (if desired).

It can be seen from FIGS. **8** and **9**, however, that the AC output amplification stage (or inner induction array) of the Primary Power System and its input stage (or outer induction array) are directly coupled by the rotor segments (**14**) in both single-stage and multi-stage rotor embodiments. Thus, the rotor electrical circuit as configured will be subject to continuous positive (or regenerative) feedback of any AC signal voltage present either in the field current or on the rotor, as described above. To provide a desired or requisite level of signal amplitude suppression or compensation, negative (or inverse) feedback may be applied to the rotor circuit using the feedback coupling relays (**129**) and stage transformers (**130**) shown in FIG. **10**.

Because plate voltage and grid voltage are typically 180° out of phase in a beam pentode arrangement such as the Generator's primary arrays, inverse feedback is achieved when the voltage fed back to the control grid (**65**) has the same waveform and phase as the plate or anode (**68**) voltage. In related traditional practice, an inverse feedback signal applied to the input (control grid) current decreases AC amplifier input impedance and distortion; an inverse feedback signal proportional to the output (plate/anode) current raises the output impedance of such amplifier. Therefore, an inverse feedback AC signal proportional to the current through resistor (**111**) which is applied in-phase with the AC line current across resistor (**110**) decreases the rotor input stage impedance yet also reduces the amplitude of the AC output current. A loss of total amplification (through a fractional stage 'gain') results, but is accompanied by a decrease in signal distortion.

The stage feedback transformers (**130**) ensure voltage reference isolation between such stationary amplification or suppression stage and the corresponding rotor stages, when AC signal feedback across the control grid resistors (**110**) is necessary or desired. A double-pole/double-throw switch/relay (**129**) may be used to decouple each such transformer from the input signal resistor (**110**) side of its circuit connections, and/or to apply either positive (regenerative) or negative (inverse) feedback to the rotor AC amplification stage(s), and tunable blocking capacitors (**131**) isolate the transformers from circuit DC voltages. The coupling transformers (**130**), isolation relays (**129**), and blocking capacitors (**131**) are therefore only necessary in Generator embodiments not having a central AC control-voltage operating circuit with suitable inverse feedback characteristics and in three-stage Generator embodiments which may be used in a communications capacity. In the latter case, an optional capacitor across the secondary or control grid resistor (**110**) side of each such transformer (in parallel with the said resistor) may once again be employed in parallel resonance fashion, and is indicated in FIG. **10**.

During continuous operation of an EDF Generator in the vacuum of space, the exhibited DC field voltage will tend to rise gradually as the total amount of charge contained in the field envelope circuit is gradually reduced through unavoidable electron leakage losses (primarily at the hull Emitter Ring peripheral edge). To compensate for this effect, a small amount of stored ambient charge must therefore be continually released at the suppressor grids (**67**) into the rotor return current. The rate at which this gradual discharge must be effected is dependent upon the observed field leakage rate, as only a rough approximation of this leakage may be pre-calculated. It is for this reason that at least one superconductive storage ring (**200**)—containing a tremendous quantity of reserve ambient charge in the form of a continuous zero-loss current—must be included in any ungrounded Impulse Drive unit used in space exploration, as the capacity of such storage ring determines the vessel's effective range of operation.

It is postulated that the external field current's energy density will be sufficiently enormous that it exhibits a virtual self-evacuating operational condition above a certain ultrahigh field voltage in propulsive three-stage EDF Generators, and such a condition is essential to the successful atmospheric ground launch thereof Such kinetic occlusion of a significant portion of the otherwise available ambient ionization charge may also occur once the field envelope is fully formed even in ground-based single-stage devices operated at breakdown field intensity. During continuous operation of such ground-based devices in a gaseous atmosphere, field leakage losses which could cause an undesired gradual field voltage rise if they exceed available ambient charge accretion may be compensated by bypass-shunted charge from earth ground. In this case, the positive pole of supply (**100**) is connected at low power via dpdt switch/relays (**119**) to earth ground, in order to avoid a limited spin-up/run/de-spin repeating duty cycle.

General Method of Construction

The actual construction of an Electrodynamic Field Generator of any given size is relatively straightforward, whether of a manned propulsive embodiment or not, and proceeds in fact from the inside out. Referring generally to FIG. **1**, one of the most practical features of the construction process is that virtually the entire instrumentation and payload compartment or "Central Chamber" (**2**) may be outfitted first, before any portion of the Primary Power System is assembled or even necessarily designed. In the preferred three-stage embodiment to be used as an aerospace Impulse Drive Unit, the central chamber should obviously incorporate low-density materials wherever possible.

Construction next proceeds with the Generator's ground frame (**10**): the metallic structural shell which provides an enclosing framework for the central chamber to which all ancillary inboard electrical equipment may be grounded. This structural framework should be as strong, lightweight, and nonmagnetic as possible, and may preferably be of a welded tubular design using stainless steel or a suitable titanium alloy.

Once the ground frame and enclosed central chamber are completed, including the installation of a preferred total of 32 high-torque DC motor-generator rotor drive units (in four sets of eight), the two rotor mounting frame sections and attached ring gear/s (**8**) of the "Carrier Assembly" may be built and dynamically balanced (using a temporary peripheral spacing "jig"). Then, construction of the composite "Rotor Assembly" may commence with the laying out of equal numbers of copper segments (**14**), ceramic segment separators (**16**), and refractory composite field emitters (**17**) which are all of exactly equal weight by type, the latter components providing the nonconductive base to which other rotating electrical hardware may be attached.

When the three sets of said principal large rotor pieces (which preferably number 180 each) are bound together with electrode rings silver-soldered to the segments, and clamped between the two halves of the centrifuge-style carrier assembly, the assembly of the "Primary Power System" itself may begin. Once thin insulative rotor surface "facing", the aforementioned recessed and conductive separator traces, and the rotor ballraces (**25**) have all been affixed to the base rotor assembly, and the field coils, electrode arrays, and other rotating components have been added, the fabrication of the vessel's "Structural Intercooler System" may proceed. When this secondary thermal conduit (hull coolant) system is completed and has been pressure-tested, the "Primary and Magnetic Arrays" may be installed—including the 5-armature variable inductor array sections depicted in FIG. **14**. The primary (induction ring) arrays here are comprised of the unit pentode arrays (**69**) as individual assemblies each having an adjunct assembly comprising a field power resistor (**63**), dielectric buffer (**45**), and primary thermal conduit (**48**).

Finally, once the rotor has satisfied nonenergized operational mechanical clearance and dynamic balancing criteria (using a temporary outer rotor bearing support "jig") and a final complete thermal conduit system pressure test has been made, the "Outer Hull Components" may be installed. Thus, the basic vessel construction may be completed in a total of seven distinct steps or stages. Due to the tremendous weight of the finished rotor (as the device's most massive single assembly), an operating characteristic of utmost importance is the actual rotor speed required to maintain a specification field voltage which is figured using the method outlined in the Detailed Calculations below. This will determine the rotor's operating angular momentum, which in propulsive units must be sufficiently large to provide gyroscopic stability against the use of variably-imbalanced isometric thrust (produced by field hemitorus current) but not so large that the vessel's navigational and maneuvering characteristics are thereby rendered sluggish.

It can be shown mathematically that the nominal rotor speed for a 4 ft. diameter theoretical design prototype would be roughly equal to or somewhat less than that of a typical small electric motor. However, as the size of the device increases, it will be necessary to reduce rotor rpm due to the extreme increase in the rotor's 'tip' speed and therefore the centrifugal forces to which it is subjected. Sufficient information is contained in the basic specifications which underlie this disclosure to allow engineers to reduce rotor speed as necessary in larger devices following the course of an intensive program of small device tesing; e.g., the projected nominal rotor rpm for a 20 ft. Thermal Unit is 1045.

Structural Intercooler System

The secondary thermal conduit sections which carry hull coolant or heat transfer fluid through the Generator's tapered "Drive Ring" (or the combined volume of the negative and neutral housing sections) must pass between the power resistor and unit pentode array assemblies, and so they must conform in outside dimension to the available space within the supporting framework of the neutral hull construction. The most unusual feature of these secondary conduits is that they also comprise (of necessity) the principle Drive Ring load-bearing members, and together they form the EDF Generator's structural intercooler system as illustrated in FIG. **14**.

These secondary hull coolant conduits should be fashioned in a manner similar to the primary thermal conduits described hereinabove, and in the preferred embodiment would be formed in four types or varieties: [i] the initial outbound coolant supply runs **(75)**, called "headers"; [ii] the circumferential heat transfer sections **(76)** (to which the Emitter Ring hull plates are fastened), called "peripheral shunts"; [iii] the inner **(77)** and outer **(78)** intermediary coolant runs, or "transfer links"; and [iv] the radial coolant returns **(79)** or "recovery lines". Within the Drive Ring, each of these conduit types may be fashioned as single sections which connect one to another in 'series-parallel' branch systems or zones. To wit, one header **(75)** from a reservoir coolant manifold **(74)** connects to one peripheral shunt **(76)** which in turn connects to one recovery line **(79)**.

Thus, a single coolant zone is illustrated in FIG. **14**, the arrows indicating the direction of coolant flow. A molded ceramic support bracket **(71)** having integral elbows and a set of mounting "knuckles" may be used to secure the inner corners of each two adjacent negative hull plates **(72)**, by means of a matching mounting ring welded to each inner plate corner and a retaining pin (not shown). Each Emitter Ring hull plate **(72)** is to have an angle-stock thermal transfer channel welded along but just inside its inner edge (as indicated by dashed lines), which encloses the corresponding peripheral shunt **(76)**.

These coolant headers, shunts, transfer links, and recovery lines are preferably formed of molybdenum tubing, and would take advantage of the high thermal conductivity and electrical resistance of a thin Alumina coating thereon to absorb excess heat from the single-stage Drive Ring hull and (in particular) the Emitter Ring hull plates. In single-stage rotor devices which are used as Thermal Power Units, the recovery lines then exit the Generator housing through two exchange manifolds (as stated earlier), from a preferred total of 72 individual coolant zones, to an external heat exchanger. Just how effective this method of cooling the Drive Ring as a whole will prove to be during operation in air, using the preferred liquid air or nitrogen coolant, is somewhat difficult to project. Liquid helium could alternatively be used in this respect to ensure adequate hull cooling, but would entail significantly increased piping and pumping difficulties in doing so.

It is expected, however, that the intercooler system described will be satisfactorily effective during the operation of three-stage devices in the vacuum of space, and is principally designed with this mode of usage in mind. Accordingly, it is believed that an Impulse Drive Unit operated in space will be able to purge its excess heat by actually circulating the primary conduit liquid sodium or equivalent through the secondary thermal conduit system (instead of the cryogenic coolant). In this case, the intercooler structure is used to transfer heat to the Emitter Ring hull plates as a radiative heat sink. This cooling method would provide the necessary means of liberating the Generator from permanent ground-based support to achieve free-ranging operation in space.

Neutral Housing Section Construction

In order to establish and withstand the action of the field envelope discharge current, the surface of the neutral sections of the Generator's housing or hull (or "Neutral Ring") must be constructed entirely of specialized nonconductive material(s) which should be applied in two layers to an underlying superstructure that is actually conductive in nature. It is proposed that overlapping ceramic 'tiles' be used in this regard, in a manner rather like that employed on the NASA space shuttles, which are adhesed to an underlying deck of structural nonmetallic sheet stock While the intercooler system conduits (described in the preceding section) comprise the principal load-bearing members which support the Drive Ring portion of the hull, a system of beams **(153)** and struts **(152)** that are also composed of the same nonmetallic structural material as the deck or a similar and compatible material should be utilized to further strengthen and 'unitize' the Drive Ring superstructure such that an internal hard vacuum may be drawn and maintained prior to operation. These secondary load-bearing members are also depicted in FIG. **14**, and in a preferred embodiment these members and the base deck material itself would be composed of a carbon composite commonly referred to incorrectly as graphite.

Sophisticated carbon composites similar to those found in golf-club shafts and racing-bicycle frames were first developed to make use of their special combination of extreme strength, rigidity, and lightness in aerospace applications. These materials are generally made from polyacrylonitrile (PAN) carbon fibers which are heated under tension to drive off the non-carbon portion. The individual fibers are about 7 microns in diameter and approach 300 kpsi in tensile strength. They may then be pulled into a rope-like 'tow' or woven into a fabric-like 'matte' before being coated with either epoxy or polyester resin. Therefore, carbon composites are eminently suited to forming rods, tubes, and sheets, and display excellent vibration damping characteristics. The final product is, however, still very expensive per pound.

To simplify the vast diversity in specialized grades of PAN carbon composite for purposes of this application, this material should be considered as coming in two basic density grades: low-density "deck" grade (at about 35% of the specific gravity of sheet steel), and high-density "beam" grade (at about 65% of the specific gravity of structural steel). It is also available in two standard temperature grades, depending on the bonding agent used: a 350° F. rating (lo-temp) epoxy-bonded material and a 750° F. rating (hi-temp) polyamide-bonded composition. The hi-temp grade only should be employed in the Neutral Ring, but the lo-temp grade may be used to build the Central Chamber.

Thus, a low-density/hi-temp grade of carbon composite laminate should be used as the basic deck material in construction of the Neutral Ring hull sections. To wit, this deck layer itself is initially affixed to the vessel superstructure by solvent welding and/or mechanical means. The PAN carbon deck, a middle layer of ceramic substrate plates and an outer skin of exolayer tiles together then comprise the composite Neutral Ring hull. Several factors enter into the selection of an appropriate ceramic material for the substrate layer plates. Of primary concern here is that the material of choice have a very high volume resistivity and very low thermal conductivity at temperatures from 300–500° C. It is also important that the material have a very low thermal expansion coefficient, as the substrate layer plates must be bonded to the deck layer with a refractory-class adhesive. Therefore, the substrate material should also be non-vitreous and have relatively high porosity to promote adequate adhesion. One of the Cordierite class of compounds, bearing CeramTec designation 447, has relatively low hardness and flexural strength but should function admirably in Thermal

Power Units as the center layer of a bonded composite construction given its other highly desirable characteristics. Impulse Drive Units will require an alternate substrate plate material as discussed below.

The selection of a ceramic material for use as the exolayer dielectric thermal tiles of the EDF Generator's housing or hull is contingent upon the operative field voltage, and therefore different materials must be specified for Thermal and Impulse Drive Units. To prevent substantial skin conduction losses, the minimum dielectric value k for the exolayer tile material used in Thermal Power Units has been calculated (using the peak field voltage) at approximately k=9. Therefore, the material of choice is a Zirconia composition bearing CeramTec designation 848, with a minimum (high-frequency) dielectric constant of 28. Even though this material has a comparatively high thermal expansion coefficient, it has very low thermal conductivity and is very dense and hard. Moreover, it exhibits exceptional flexural strength and resistance to fracture.

The minimum dielectric value k for the exolayer tile material used in Impulse Drive Units has been calculated at approximately k=107. Therefore, the material of choice is the same titanate composition prescribed for the dielectric buffers (and described in some detail in a previous section). Known chemically as Sodium Bismuth Titanate (and commercially as Kézite), this highly unusual compound is technically a piezoelectric material, but is also ferroelectric in that its dielectric constant actually increases on temperature rise to a peak low frequency value of about 3,100 at its 655° C. curie point. Like most ferroelectrics, its tensile and flexural strengths are quite low yet it is very dense and exceedingly hard. It does, however, exhibit a thermal expansion coefficient which is very high for a ceramic material.

The expected higher flexural stresses upon the Neutral Ring of an Impulse Drive Unit (compared to a Thermal Power Unit) may prove sufficient to crack the substrate plates were they composed of the Cordierite material described above. The Zirconia compound mentioned earlier should therefore be substituted for Cordierite 447 as the substrate plate material in Impulse Drive Units. This would likely stiffen the composite Neutral Ring hull adequately to prevent such stress cracking, while effecting a much closer match in thermal expansion coefficient to the titanate tiles. In both Thermal and Drive Units, the exolayer tiles should be applied to the substrate plates with a refractory-class adhesive in a mosaic overlay pattern, such that the edges of each exolayer tile are evenly spaced as far as possible from the underlying edges of the nearest substrate plates.

Because the underlying deck layer of the Neutral Ring hull will be quite conductive when composed of carbon composite sheeting, and referring now to FIG. **1**, it becomes imperative to use an auxiliary dielectric buffer **(50)** between the inner edges of the Emitter Ring **(47)** hull plates and the outer edge of the said deck layer to prevent an otherwise enormous deck leakage current from occuring. Although it might seem that the much-cheaper and nearly-as-strong fiberglass composites available (which are nonconductive) would tend to minimize any such leakage current condition, their use would not prevent the possible build-up of unacceptable levels of static charge within the working induction compartment **(12)**. Therefore, the use of conductive deck material ensures that the Faraday shielding principle protects Neutral Ring components from transient static voltages because the presence of a finite deck leakage current means the deck layer is a charged conductor, whereupon any net static charge must be external.

A secondary concern is the actual operating temperature of the contact areas where the deck layer abuts to both the auxiliary **(50)** and primary **(45)** dielectric buffers, which must be kept below the 750° F. temperature rating of the carbon composite. It can be shown mathematically that the deck leakage current can be held to a thoroughly acceptable fractional amperage per cm^2 of such contact area if this condition is met. In Thermal Power Units, the use of a secondary cryogenic hull coolant should keep the auxiliary buffers **(50)** and outer deck contact areas at a temperature which is under the said rating. Doing so at the primary buffer **(45)** contact area locations may well require the use of auxiliary cryogenic thermal conduits in both Thermal and Drive Unit models of the EDF Generator. The auxiliary dielectric buffers in any given Generator should be composed of the same material used in that device's exolayer tiles, and should be of about half the radial thickness of the primary buffers employed therein.

While the specific materials discussed above should not be construed as critical to the EDF Generator's proper construction, their properties and characteristics are highly indicative of the rather involved considerations which must be undertaken to ensure that the neutral sections of the vessel hull are physically suited to optimizing a given size device's performance and minimizing its housing maintainance or repair requirements.

Stationary Electromagnetic Armatures

Because the EDF Generator's stationary electromagnetic armatures have no bearing on the device's rotor or field voltages, and no direct connection therewith, they are properly thought of as being part of neither the Primary Power System or Field Induction System circuits but rather a separate ancillary subsystem intended to optimize the invention's overall performance and efficiency in a number of important ways. To this end, their main purpose is to create and independently control a variable level of both rotor torque and field current rotational force, as necessary or desired.

As depicted in FIG. **14**, the individual stationary armatures **(37)** (or "variable inductors") may be installed in 5-armature groups or array sections within the Neutral Ring of the hull, using notched and mating supports **(151)** which firmly clamp the armatures about a reduced diameter center section of each armature core. These armatures and a basic control circuit arrangement for them are shown in FIG. **16**, with the polarities indicated being those which would result in a clockwise rotation of the rotor as viewed from above. This armature power distribution system may be manually adjusted and/or automatically controlled (using a computer and/or other circuitry).

In both single-stage and three-stage embodiments of the Generator, each of the two electromagnetic armature arrays should be wired in parallel from one or more separate source(s) **(185)** of low-voltage DC current which is/are chassis-grounded and which is/are common to the armatures **(37)** of that array. No variable regulation or biasing of individual armature DC current(s) in single-stage rotor or Thermal Power Unit devices is necessary. The desired individual biasing of armature DC current in three-stage rotor devices, for propulsion-related reasons explained below, may be simply accomplished through a variable resistor **(184)** associated with each armature and its DC power supply subcircuit(s). A minor unidirectional or alternating voltage component of such armature supply current may be added across or in parallel with such resistor **(184)**, using an AC power source **(186)** and/or suitable control interface of a conventional nature, in either single-stage or three-stage Generator embodiments as desired.

Each stationary armature **(37)** incorporates two separate electrical coils **(180)**–**(181)**, to be operated principally on DC voltage, which are formed of conductive insulated magnet wire multi-layer wound upon a single ferromagnetic core. Each such armature core also incorporates a connecting center section or "flux reductor" **(182)** between said separate coils which is of a significantly reduced cross-sectional area with respect to the balance thereof, and which is designed to saturate magnetically when the respective core flux density of either of said coils is approximately equal to or greater than half saturation value. Pure annealed iron or low-carbon steel should be used as the armature core material, due to the high permeability, superior intrinsic induction, and low hysteresis thereof.

The inner or "flux initiator" coils **(181)** of these electromagnetic armatures (or those closest to the rotor) may be used as a means of creating axial magnetic fields of force **(144)** which are perpendicular to the plane of the current-carrying rotor segments **(14)**. Thus, when powered collectively at a common DC current value, the said inner armature coils **(181)** in conjunction with their respective cores may serve as a source of uniform yet variable secondary input rotary torque (or in a torque-assist capacity) for the invention in the course of its normal operation as an electrical generator.

The flux initiator coils **(181)** are so-named because their as-wound polarity relative to the fixed outward direction of the primary DC rotor current establishes both the direction of rotor rotation and the required winding direction and polarity of the outer (and similar) armature coils **(180)**. The said inner coils **(181)** are energized by a variable DC power supply **(185)**, through an output variable resistor **(184)** and a filter diode **(189)** which blocks induced positive AC half-cycles (if any) from crossing the negatively-grounded DC supply **(185)**. Such power supply **(185)** must, however, be sufficiently rated to carry any corresponding induced unidirectional negative current impulses to ground, and may be of a solid-state or rotary induction design.

The axial armature-produced fields **(144)** described may also be individually and variably superimposed on the rotor to develop a selectively controllable force of rotational torque upon its current-carrying segments **(14)**, through variable regulation of the principal DC current in individual armature inner coils **(181)**, such that a locally-unbalanced force of angular acceleration may be applied to the rotor and therefore a directional force of horizontal thrust may indirectly be imparted to the housing for lateral flight maneuvering in propulsive three-stage rotor devices.

The outer coils **(180)** of the stationary electromagnetic armatures, or those closest to the field envelope, may be used as a means of creating axial magnetic fields of force **(144)** which in this case are essentially perpendicular to the described field current arc trajectories. This situation is best illustrated in FIG. **12**, which shows the hull location of the two variable inductor arrays **(145)** (or circular groups of armatures **(37)**). Thus, when powered collectively at a common DC current value (including any minor unidirectional or alternating component thereof), these outer armature coils **(180)** in conjunction with their respective ferromagnetic cores serve to impart an attenuating yet holistic and orderly rotational vector moment upon the external discharge current. The general effect of such applied magnetic rotational force, which may be used to modify or modulate the field's electrodynamic characteristics, is further depicted in FIG. **13**.

Referring again to FIG. **16**, and to continue in the same vein, it is postulated that the quasi-coherent aspects of the field envelope may actually be optimized through the proper control of the amplitude, frequency, and/or phase relationships between any minor AC voltage component of the field envelope current and the said minor AC current (if any) collectively supplied by AC power source **(186)** to the outer or "transflection" coils **(180)** (of each variable inductor array). This consideration applies to both single- and three-stage embodiments of the Generator, although it is decidedly more important in the case of the latter, and may also allow the engineered reduction of field radiative emissions.

These transflection coils **(180)** are so-named because they may be used to impart a variable transverse deflectionary force upon all of the radially-impinging field current electrons. In the region(s) of the field envelope shown in FIG. **11** between the surface of each neutral hull section **(4)** and the corresponding field envelope inner boundary **(142)**, this magnetic force may in fact be large enough to produce a continuous annular flow of "displacement charge" current. In any event, and as shown in FIG. **16**, such outer armature coils **(180)** are also energized by a variable DC power supply **(185)** which is similar in type to that of the inner coils **(181)**, through an output variable resistor **(184)** and filter diode **(189)**. In addition, a variable AC power source **(186)** may impress a minor unidirectional or AC voltage upon the DC current supplied to the said transflection coils **(180)**, either across or in parallel with the output resistor(s) **(184)** of their respective DC power supply subcircuit(s), through separate output variable resistors **(184)** and DC blocking capacitors **(188)** connected to said AC power source.

The axial magnetic fields **(144)** produced by the armatures **(37)** may also be individually and variably superimposed on the field envelope to develop a selectively controllable rotational force upon the described field hemitorus current, through variable regulation of the principal DC current in individual outer (or transflection) coils **(180)**, such that a locally-unbalanced force of angular acceleration may be applied to such hemitorus current. Any such angular force imbalance will cause a corresponding change in the local field current density and therefore a secondary directional force of variable thrust in the z-axis may indirectly be imparted to the housing in propulsive devices, yielding a further measure of navigational attitude control over non-isometric thrust produced by the Generator which may otherwise be largely unilinear.

Given the nature of the flux reductor **(182)** as described above, it can be seen that either the inner or the outer armature core sections may be collectively or individually operated at any flux density above half of saturation level without significantly affecting the operating flux density of the respective opposite core sections, provided both corresponding sets of armature coils (**(181)** and **(180)** respectively) are powered by DC current of the same relative polarity. Thus, the base collective full-power operational level of flux density in the inner armature core sections of a three-stage rotor Impulse Drive Generator should be roughly half saturation level, in order to facilitate the achievement of temporary local rotor torque imbalances for propulsive purposes as described above. Such inner core section base flux density for single-stage rotor or Thermal Power Unit devices may be any desired in a range between 0 and 100% of saturation.

As shown in FIG. **16**, means have nevertheless been provided for reversing the fundamental DC polarity of either the inner **(181)** or outer **(180)** coil of any given armature **(37)**, without changing the given polarity of such coil's DC power supply **(185)**, using a dpdt relay **(190)**. It can be seen that this simple feature allows a counter-polarity magnetomotive force (or counter-mmf) to be applied to either wound core section of any given armature, with respect to the core polarity of that armature's opposite wound section. This feature, when taken in conjunction with the described nature of the flux reductor(s), enables the outer armature core sections in both single- and three-stage devices to be operated at nearly any flux density within their output capability which might be required to optimize the desired quasi-coherent properties of the field envelope, despite the relative polarity and/or flux density of the inner armature core sections. In both single- and three-stage devices, these same two features also allow a substantial level of braking or counter-rotational torque to be applied to the Generator's rotor —by any or all of the inner armature core sections— while maintaining such an optimum level of flux density in the outer armature core sections: even if the two such opposite core sections are operated at opposite respective polarities.

Detailed Calculations for Preferred Embodiment

It is important to note that all of the underlying specifications pertaining to the size and/or positioning of Primary Power System components located within the hull's neutral section region (or "Neutral Ring") are given in terms of fractional increments of the hull radius (R_h) and/or the radial width of the Neutral Ring (C_v–C_a). This means that these components may be easily scaled in exact proportion to the hull, for any selected radius thereof, with no significant error or modification. However, an upper limit of vessel size utilizing the technology represented by the EDF Generator probably exists at a hull diameter not greatly above 100 feet, due to inordinate DC Field Voltage concerns (as that voltage is calculated using the method below).

Table of Dimensions	
Hull Radius = R_h	
Hull Volume Constant (C_v)	$C_v = \frac{1}{5}R_h$
Hull Area Constant (C_a)	$C_a = 0.012919C_v$
Drive Ring Radius (r_f)	$r_f = 2C_v$
Emitter Ring Radius (r_{neg})	$r_{neg} = C_v + C_a$
Neutral Ring Radius (r_{neut})	$r_{neut} = C_v - C_a$
Field Hub Radius (r_z)	$r_z = 3C_v$ (Note: $R_h = r_f + r_z$)
Polar Hull Constant (h_z)	$h_z = 0.1421245r_z$
Polar Volume Differential (X_h)	$X_h = \frac{(r_z^2 - h_z^2)}{2h_z}$
Radial Hull Constant (R_s)	$R_s = h_Z + X_h$
Area of Positive Zone (A_z)	$A_z = 2\pi R_s h_z$
Area of Negative Section (A_n)	$A_n = \frac{s(C + c)}{2}$, where: $s = \sqrt{(C_v + C_a)^2 + (\tan\alpha(C_v + C_a))^2}$, $C = 2\pi R_h$, and $c = 2\pi(R_h - (C_v + C_a))$, and Note: $A_z = A_n$.
Volume of Central Section (V_c)	$V_c = \pi r_z^2 h_f$, where $h_f = r_f(\tan\alpha)$, and α is the said Emitter Ring angle.
Volume of Positive Zone (V_z)	$V_z = \frac{2}{3}\pi R_s^3 - \pi\frac{R_s^2 X_h - X_h^3}{3}$
Vol. of Drive Ring Section (V_f)	$V_f = \frac{\pi h_f(R_h^2 + R_h r_z + r_z^2)}{3} - \pi r_z^2 h_f$ Note: $V_z = V_f$
Total Hull Volume (V_t)	$V_t = 2(V_z + V_f) + 2V_c$

The variable 'α' equals the displacement angle of the Emitter Ring negative sections (with respect to the hull's horizontal centerline plane), which is preferably 7.5° to 8° each.

Field and Rotor Voltages

Part A: Model Unit Field Intensity

[1] The Marginal Envelope DC Field Intensity for all EDF Generator models regardless of type should be equal to the breakdown dielectric strength of air or vacuum at 3 million (3×10^6) volts/meter, as measured along a full pure semicircular arc drawn from the center point on the surface of either hull Positive Zone (**5**) to any exterior point on the vessel's design hull configuration which lies on the horizontal centerline plane thereof. This distance shall be referred to as the "Drive Field Perimeter" (**141**), as shown in FIG. **11**, and is taken to represent the longest continuous DC field current trajectory.

[2] For purposes of these Specifications, the "Drive Field Boundary" (**142**) shall represent the distance measured along a full pure semicircular arc from any point on the outer peripheral edge of either hull Positive Zone (**5**) to the closest point thereto which lies on the inner peripheral edge of the vessel's corresponding negative hull section (**3**) (as shown in FIG. **11**), and is taken to represent a shortest such DC drive current trajectory.

Part B: Thermal Power Units

[1] The Nominal Field Voltage (nom.V_f) should be numerically equal to 1.5 million (1.5×10^6) times the vessel R_h as measured in feet. This value is also equal to the Thermal Unit's standard design Primary Voltage Expansion Ratio of 750 times its Specification DC Primary Array Voltage, as a function of the charge stored within the Primary Power System's two Field Ballast Capacitors and of the Unit's primary cathode emissivity.

[2] The Specification Field Voltage (spec.V_f) should be equal to the Marginal Envelope DC Field Intensity times the distance comprising the Drive Field Perimeter, or (as presently calculated under the above directive) 0.9666 times the Nominal Field Voltage. This value shall constitute the Unit's standard operating value of field envelope voltage.

[3] The Peak (design operating maximum) Field Voltage (max.V_f) shall be equal to 110% of the Specification value thereof, and should never be exceeded in operation.

[4] The Nominal Rotor Voltage (nom.V_r) would be numerically equal to one-third the product of 1,000 (1×10^3) times the vessel R_h, as measured in inches, as a gauge of the basic full-power DC generator value of rotor rotational speed, in a 'single-stage' device.

[5] The Specification Rotor Voltage (spec.V_r) should be numerically equal to 0.25776% of 1.5 million (1.5×10^6) times the vessel R_h as measured in feet, and is also equal to 0.9666 times the Nominal Rotor Voltage. This value would constitute the standard operating value of DC rotor voltage, as a function of the engineered Primary Power System output at a nominal design rotor rotational speed.

[6] The Peak (design operating maximum) Rotor Voltage (max.V_r) shall be equal to 110% of the Specification value thereof, and should likewise never be exceeded.

[7] The Specification DC Primary Array Voltage (spec.ΔV_p) will be equal to half (½) of the Specification Rotor Voltage, and affects all primary array electrode spacings.

Part C; Impulse Drive Units

Unlike the case with Thermal Power Units, it is not possible to specify in one step a Nominal Field Voltage for Impulse Drive Units which is a linear function of the hull radius R_h; the DC Field Voltage which must be produced to achieve the chosen design goal for net linear thrust output, equal to five-thirds the Earth's pull of gravity, is jointly proportional to the rotor ampacity which rises as a function of the square of increases in R_h and to the vessel weight (which rises as a function of the cube of R_h increases).

Therefore, a Marginal Field Voltage is initially to be computed for Impulse Drive Units which reflects a projected theoretical operating value of DC Field Voltage required to produce a vessel acceleration approximately equal to 1.67 'g' outside of a proximal gravity field. This Marginal Field Voltage is to be calculated using the "specific impulse" of the vessel concerned (in nt-sec), and for purposes of the application contemplated herein this term is hereby defined as being equal to five times the vessel weight (in newtons) divided by the rotor ampacity times the number of electrons per coulomb.

[1] Accordingly, the Marginal Field Voltage shall then be equal to the Unit's specific impulse (F_{dt}) times the speed of light (c) divided by the unit electron charge (q).

[2] An estimated nominal Field Voltage will also be computed whereby a nominal value for the final electron speed of Drive Field current is assigned which is in fact a linear function of R_h, as applicable to vessels of from 4 to 100 ft. in diameter only.

[3] The Nominal DC Field Voltage (nom.V_f) should then be equal to the simple average of the Marginal Field Voltage and the said estimated nominal Field Voltage.

[4] The Specification Field Voltage (spec.V_f) should equal the Nominal DC Field Voltage (nom.V_f) divided by an assigned engineering design constant equal to 0.982826.

[5] The Peak (design operating maximum) Field Voltage (max.V_f) shall be equal to 110% of the Specification value thereof, and should never be exceeded in operation.

[6] The Nominal Rotor Voltage (nom.V_r) would be numerically equal to one-fourth the product of 1,000 (1×10^3) times the vessel R_h as measured in inches, as a gauge of the basic full-power DC generator value of rotor rotational speed, in a 'three-stage' device.

[7] The Specification Rotor Voltage (spec.V_r) should be numerically equal to 0.19332% of 1.5 million (1.5×10^6) times the vessel R_h as measured in feet, and is also equal to 0.9666 times the Nominal Rotor Voltage. This value would constitute the standard operating value of DC rotor voltage, as a function of the engineered Primary Power System output at a nominal design rotor rotational speed.

[8] The Peak (design operating maximum) Rotor Voltage (max.V_r) shall be equal to 110% of the Specification value thereof, and should likewise never be exceeded.

[9] The Specification DC Primary Array Voltage (spec.ΔV_p) will be equal to half (½) of the Specification Rotor Voltage, and affects all primary array electrode spacings.

[10] The design Primary Voltage Expansion Ratio for each Impulse Drive Unit should then be equal to the ratio of Specification Field Voltage to Specification DC Primary Array Voltage, and is once again a function of the ballast capacitor charge and of the primary cathode emissivity as in the case with Thermal Power Units.

Example of Specific Unit Voltage Values

Four-foot diameter model

Section A: Thermal Power Units

A 4 ft. EDF Generator constructed for use as a Thermal Power Unit would have specific Rotor and Field Voltages, according to the preceding instructions, as follows:

[1] The Nominal Rotor Voltage (nom.V_r) will be equal to 8,000.

[2] The Specification Rotor Voltage (spec.V_r) will be equal to 7,732.8.

[3] The Spec. DC Primary Array Voltage (spec.ΔV_p) will be equal to 3,866.4.

[4] The Nominal Field Voltage (nom.V_f) will be equal to 3,000,000.

[5] The Specification Field Voltage (spec.V_f) will be equal to 2,899,800.

The corresponding Primary Voltage Expansion Ratio (for this and any single-stage Thermal Power Unit) will thus be 750:1, and in this case is equal to 2,899,800/3,866.4.

Section B: Impulse Drive Units

A 4 ft. diameter three-stage-rotor EDF Generator constructed for use as an Impulse Drive Unit would have specific Rotor and Field Voltages, according to the preceding instructions, as follows:

[1] The Nominal Rotor Voltage (nom.V_r) will be equal to 6,000.

[2] The Specification Rotor Voltage (spec.V_r) will be equal to 5,799.6.

[3] The Spec. DC Primary Array Voltage (spec.ΔV_p) will be equal to 2,899.8.

[4] Marginal Field Voltage: The 4 ft. Drive Unit's Marginal Field Voltage, which represents a projected minimum standard operating value of field envelope voltage, is based upon the vessel's specific impulse (F_{dt})—given its calculated design rotor ampacity of I_{max}=38,160 amps and estimated weight of 79.2 lbs/ft^3—which is computed as follows:

$$F_{dt}=\frac{5(9.8\text{ m/sec}^2)V_t(79.2\text{ lbs/ft}^3)}{(3.816\times10^4\text{ coul/sec})(6.25\times10^{18})(2.2\text{ lbs/kg})}$$

where V_t=2.5068 ft, as calculated by the formulas in the Table of Dimensions above using an Emitter Ring displacement angle (α) of 7.5°. Therefore,

$$F_{dt}=\frac{5(9.8\text{ m/sec}^2)(90.245\text{ kg})}{23.85\times10^{22}\text{ electrons/sec}}=1.8541\times10^{-20}\text{ nt-sec/electron}^*$$

*This impulse represents the instantaneous force of thrust produced by each Field current electron's collision with the Field Hub (equivalent here to: kg-m/sec/electron).

Therefore, the Marginal Field Voltage is equal to:

$$\frac{F_{dt}c}{q}=\frac{(1.8541\times10^{-20}\text{ nt-sec})(299.7925\times10^6\text{ m/sec})}{1.6\times10^{-19}\text{ coul}}=34.7403\times10^6\text{ volts (nt-m/coul)}$$

[5] Estimated Nominal Field Voltage: The 4 ft. Impulse Drive Unit's estimated nominal field voltage, which represents an alternate but extremely accurate projection of

the minimum standard operating value of field envelope voltage, is based upon a nominal value for the final electron speed of Drive Field current (ΔV_e) which is an assigned linear function of the vessel R_h and is calculated as follows:

$$\Delta V_e=[0.999+(2R_h-4)(1.03125\times10^{-5})]c,$$

where R_h is in feet, for vessels from 4 to 100 ft. in diameter.

Therefore, $\Delta V_e=[0.999+0(1.03125\times10^{-5})]c=0.999$ c.

The purpose of this text subsection is to verify that a goal-requisite level of Field current thrust may be produced by the Electrodynamic Field Generator, at an operating DC Field Voltage approximately equal to that stated above, within the ampacity of the rotor and at an assigned value of Field current electron velocity (for this model) equal to 0.999c. In this manner, it is hoped that unnecessarily high operating Drive Field Voltage may be avoided, by virtue of design (due to the finite level of unavoidable uncertainty in the relativistic values involved), without sacrificing device performance.

(a) The amount of thrust (force) required to produce an acceleration of 1 g, thus exactly compensating the weight of the Impulse Drive Unit or its normal acceleration due to Earth's gravity, may be calculated by Newton's Second Law as follows:

F=mg, where m=est. design mass* of 90.245 kg, and $g=9.8$ m/sec^2.

*Note: The estimated design weight of any given Impulse Drive Unit is 79.2 lbs/ft^3.

Thus, F=884.4 nt.

(b) Since impulse equals change of momentum, the electron thrust developed by the Drive Field equals the total relativistic mass of Field current times the incident current velocity (as the final electron velocity equals zero). The amount of impulse thrust which is equivalent to 1 g in this example is 884.4 nt.

Therefore, let $\Sigma M_t \Delta V_e=884.4$ nt.

(c) The design goal for net linear thrust output of any given Impulse Drive Unit is 16.333 m/sec^2 at the minimum standard operating value of Field Voltage and rated rotor ampacity (I_{max}) Therefore, given the considerations of Field Current Bias values (as were outlined in an earlier section), total isometric thrust at Marginal Field Voltage must equal 5 g. The largest current available to effect a Field impulse equivalent to 1 g is then equal to:

$$I_g=I_{max}/5=7{,}632 \text{ amps.}$$

(d) Since Σ equals the total number of electrons comprising a Drive current of I_g, we find:

$$\Sigma=I_g(6.25\times10^{18})=4.77\times10^{22} \text{ electrons/sec.}$$

(e) Therefore, from [b] above:

$$884.4\text{nt}/\Sigma=M_t\Delta V_e, \text{ and } M_t\Delta V_e=1.8541\times10^{-20} \text{ kg-m/sec/electron.}$$

(f) Letting $\Delta V_e=0.999c$, where $c=299.7925\times10^6$ m/sec, we calculate $M_t=619.08\times10^{-31}$ kg, or about 67.96 m_o, where m_o=the electron rest mass of 9.11×10^{-31} kg.

(g) If m_i equals the relativistic mass equivalent of the gained kinetic energy of each Drive Field electron, then $M_t=m_o+m_i$.

Therefore, $m_{i=Mt}-m_o=619.08\times10^{-31}-9.11\times10^{-31}=609.97\times10^{-31}$ kg.

(h) Each Drive Field electron's gained kinetic energy (E_q) equals $m_i\,c^2$. Using a precise value of 8.98755×10^{16} for c^2, we obtain $E_q=(6.0997\times10^{-29})\,(8.98755\times10^{16})=54.8214\times10^{-13}$ joules.

(i) Here, E_q also equals $W=q(\text{est.}V_{nf})$, where q=electron chg. @1.6×10^{-19} coul., and est.V_{nf}=est. nominal Field Voltage.

Therefore, $E_q/q=\text{est.}V_{nf}$, and est.$V_{nf}=34.2633\times10^6$ volts (jouleslcoul).

Thus, we find that the estimated nominal Field Voltage required to develop requisite thrust is in fact well within the thrust goal parameters discussed above, at 98.63% of the projected Marginal value (calculated in the preceding subsection B[4]).

[6] The Nominal DC Field Voltage (nom.V_f) should therefore be equal to the simple average of the Marginal Field Voltage projected earlier and the estimated nominal Field Voltage Oust computed), at 34.5018×10^6 volts.

[7] The Specification Field Voltage (spec.V_f) should then be equal to the Nominal DC Field Voltage divided by the said design constant 0.982826, at 35.1047×10^6 volts. This value shall constitute the Unit's standard operating value of field envelope voltage.

[8] The design Primary Voltage Expansion Ratio for this model Impulse Drive Unit is therefore equal to $(35.1047\times10^6)/2{,}899.8=12{,}106{:}1$, as was referred to earlier.

From the foregoing description it will be observed that the present invention provides a remarkable level of electrical and thermal output in a lower-voltage corona discharge induction embodiment, which is intended for use in a utility or physical plant application. However, this invention also provides the potential for an appreciable level of net impulse thrust in a higher-voltage arc discharge induction embodiment intended as an electrically-propulsive aerospace vessel. It can be seen as well that the Electrodynamic Field Generator and the electromotive field of force produced thereby will moreover fulfill the stated objective of NASA's premier Breakthrough Propulsion Physics Program: to discover a method whereby "a vehicle can create and control an asymmetric force on itself without expelling a reaction mass" and which satisfies conservation laws in the process. Such a vessel could perhaps be made to attain a velocity commensurate with its collector field electron collision velocity, and affords hope that true interstellar travel might therefore become possible as well. While there is due cause for some concern as to the EDF Generator's production of high-frequency radiative emissions coincident with its production of heat and impulse thrust using the new means described, it is postulated that the electromagnetic characteristics of the device's field envelope can be suitably tuned in the manner(s) oulined hereinabove such that the field envelope is rendered largely opaque to such very-high-energy emissions—by promoting continuous and adsorptive Compton effect interactions thereof with the field's impinging drive current electrons. When used in association with a ground-based utility or physical plant, a non-propulsive thermal and/or electrical power output model of the Generator may also be enclosed within a suitable Faraday "cage" or shielding structure to further minimize radiative emissions which are undesirable. Neverthless, the said field envelope tuning principles (related to controlling the amplitude, frequency, and/or phase relationships amid field AC voltage components) may moreover permit the development of the Electrodynamic Field Generator as a novel signal communications means. In pursuit of a new and more complete quantum gravity theory, experiments may be conducted involving the gravimetric coupling of two or more separate such devices: whose field envelopes will by design be able to exhibit a relativistic and oscillatory apparent mass effect that is artificially generated—at desired resonant frequencies—by using alternating electromagnetic induction methods of the type described hereinabove to modulate electrodynamic field properties which are normally of constant polarity and magnitude or random frequency.

Accordingly, it is to be understood that the embodiments of the invention herein described are merely illustrative of the application of the principles of the invention. Reference herein to details of the illustrated embodiments are not intended to limit the scope of the claims, which themselves recite those features regarded as essential to the invention.

What is claimed is:

1. An electrodynamic field generator comprising:

(a) a disc-shaped hull having upper and lower surfaces separated by a distance at an axis of the disc and tapering to an edge at a perimeter of the disc, enclosing a volume therebetween, the upper and lower surfaces each being divided into a positive section surrounding the axis of the disc, a negative section around the periphery of the disc, and a neutral section therebetween; the positive and negative section surfaces being of conductive material and the neutral section surfaces being of insulative material;

(b) a rotor having an outer circumference, an inner circumference surrounding an open central aperture, and an annular body therebetween, with the rotor body having flat upper and lower surfaces; the rotor being rotatably mounted within the volume of the hull and centered upon the axis thereof, an inner diameter and outer diameter of the rotor being chosen such that the rotor extends across at least all of the neutral portion of the hull volume and at least part of each positive hull section;

(c) an evacuated induction compartment and peripheral space charge chamber comprising at least the portion of the hull volume within which the rotor is mounted;

(d) a principal generating system and corresponding rotor electrical circuit comprising two assemblies, one associated with the upper surface of the rotor and one with the lower surface thereof, each assembly in turn comprising:

(i) an outer induction electrode array comprising a flat cathode ring mounted upon and electrically coupled to the rotor, adjacent to an outer circumference thereof, and a flat anode ring mounted plane-parallel to, but insulated from, the cathode ring;

(ii) an inner induction electrode array comprising a flat anode ring mounted upon and electrically coupled to the rotor, a flat cathode ring, and a grid ring element, the cathode and grid being mounted plane-parallel to, but insulated from, the anode and each other; the inner induction ring array being mounted to the rotor closer to the inner circumference thereof than the outer induction ring array;

(iii) at least one generating coil mounted to the rotor and extending toroidally around the rotor concentric to either circumference thereof, located between the outer induction ring array and the inner induction ring array, and electrically coupled to the cathode of the inner induction array and to the anode of the outer induction array; and

(iv) a plurality of stationary permanent magnets mounted to the hull, such that the magnetic field from each of the permanent magnets intercepts a generating coil; such that when the rotor is rotated, the magnetic fields from the stationary magnets induce an electrical voltage in each generating coil which results in a DC rotor circuit output current;

(e) a field induction system comprising two assemblies, one associated with the upper surface of the rotor and one with the lower surface thereof, each assembly in turn comprising:

(i) at least one plane-parallel electrode array comprising:

1) a primary anode ring mounted to the rotor adjacent to the inner circumference thereof, between the inner circumference and the inner induction ring array;

2) at least one primary cathode mounted to, but electrically separated from, each positive section of the hull; and

3) at least one stationary grid element located between each primary cathode and the corresponding primary anode ring;

(ii) a primary cathode resistor electrically coupled between each primary cathode and the corresponding positive section of the hull, each cathode resistor being encapsulated by a dielectric buffer; and

(iii) a principal hull coolant and heat transfer system, thermally coupled to each primary cathode resistor;

(f) at least one housing charge ballast capacitor, electrically coupled to the principal electrical generating system circuit and to the rotor;

(g) a plurality of field emitters which are mounted around the outer circumference of the rotor and electrically coupled thereto; and

(h) a motor drive mounted to the hull, having a mechanical output coupled to the rotor for rotation thereof.

2. The electrodynamic field generator of claim **1**, in which:

(i) the principal electrical generating circuit induces a primary DC voltage between the outer circumference of the rotor and the inner circumference thereof when the rotor is rotated by the motor drive, resulting in a unidirectional rotor current which distributes electron charge from the outer rotor circumference to the negative hull sections; and

(ii) ii) the field induction system utilizes a portion of the primary DC rotor voltage to establish an external discharge current which expresses a very large expansion of that voltage, within the electric field between the negative and positive sections of the hull, the measure of the expanded field voltage being defined by an engineered instantaneous charge differential between the rotor field emitters and the primary cathodes which is supported by the electric force in one or more housing charge ballast capacitors.

3. The electrodynamic field generator of claim **1** in which the rotor comprises a plurality of conductive radial segments and the same number of nonconductive segment separatorss, with the individual segments and separators uniformly alternating in placement.

4. The electrodynamic field generator of claim **3**, in which the conductive rotor segments are made of copper.

5. The electrodynamic field generator of claim **3**, in which each generating coil is mounted to the segment separators.

6. The electrodynamic field generator of claim **3** in which the anode of each outer induction array of the principal generating system is mounted to the segment separators, as are the cathode and grid element of each inner induction array.

7. The electrodynamic field generator of claim **1** in which each inner induction array constitutes a triode vacuum-tube construct, the control grid voltage of which biases that array nearly to current cutoff.

8. The electrodynamic field generator of claim **1** in which each of the two assemblies of principal generating system and corresponding rotor electrical circuit comprise a set of three generating coils, arranged across the annular body of the rotor, with each set of coils being sequentially coupled electrically.

9. The electrodynamic field generator of claim **1** in which each generating coil is compound wound, having primary and bias windings.

10. The electrodynamic field generator of claim **1**, in which each generating coil further comprises a nonmagnetic core.

11. The electrodynamic field generator of claim **1**, in which each generating coil further comprises a powdered ironi core.

12. The electrodynamic field generator of claim **1** in which the stationary permanent magnets are C-shaped, and are arranged into a circular body with a flux gap and wherein the magnetic field passes through the flux gap, and the generating coils are positioned to pass through the flux gaps of the magnets in a uniform manner during rotation of the rotor.

13. The electrodynamic field generator of claim **1**, in which the magnets are made of an alnico family alloy.

14. The electrodynamic field generator of claim **1** in which a housing charge ballast capacitor is mounted upon the rotor between each outer induction array and the rotor's outer circumference, each ballast capacitor having an inner conductive ring electrically connected to the rotor, an outer conductive ring plane-parallel to the inner ring and electrically connected to the outer induction array anode, and a dielectric layer between its inner and outer rings.

15. The electrodynamic field generator of claim **1** in which the field emitters are made of a sintered refractory composite material which is electrically conductive, and have a shape such that they taper from an inner end or surface having a thickness equal to a thickness of the rotor's outer circumference to a slightly blunted edge at an outer end forming an outermost circumference of the rotor and field emitters.

16. The electrodynamic field generator of claim **9** in which each assembly of the principal generating system and corresponding rotor electrical subcircuit comprises one or more AC subcircuitry stages, each stage being mounted to the rotor for rotation therewith and in turn comprising:

an AC input, or grid resistor;

one or more bias windings electrically coupled to the AC input;

a triode array electrically coupled to the AC input and to that input's corresponding bias winding or windings, comprising a cathode, an anode, and a control grid; and

an output, or plate resistor, electrically coupled to the anode of that triode array or to the anode of a successive electrode array.

17. The electrodynamic field generator of claim **16** in which a plurality of AC stages are electrically coupled in series, the input of an outermost stage closest to the outer circumference of the rotor being coupled to the outer induction array anode and the output of an innermost stage closest to the inner rotor circumference being coupled to the control grid of the inner induction array.

18. The electrodynamic field generator of claim **17** in which each assembly of the principal generating system, and the corresponding rotor electrical subcircuit, comprises three AC stages.

19. The electrodynamic field generator of claim **1** in which the plane-parallel electrode arrays comprising the field induction system each in turn comprise a primary cathode, three grid elements, and an anode ring and therefore constitute pentode vacuum-tube constructs.

20. The electrodynamic field generator of claim **1**, in which a voltage control system is electrically or thermally coupled to at least one stationary electrode of each plane-parallel array comprising the field induction system such that the total discharge current is limited to a value within the total ampacity of the rotor segments.

21. The electrodynamic field generator of claim **20**, in which the operating electrode temperature is controlled by movable thermal adjusting rods that may be thermally coupled to one or more of the cathodes or grids and to a cryogenic heat sink.

22. The electrodynamic field generator of claim **20**, in which the voltage control system is further utilized to apply inverse AC signal feedback to the discharge current such that rotor circuit AC voltage amplitude is limited to a desired or requisite value.

23. The electrodynamic field generator of claim **1** in which each positive section of the hull is divided into a plurality of radial sectors, each sector being separated from adjoining sectors by an insulating partition.

24. The electrodynamic field generator of claim **23** in which each assembly of the field induction system comprises a plurality of plane-parallel electrode arrays, with each array being electrically coupled to a radial sector of a positive hull section.

25. The electrodynamic field generator of claim **23**, in which each positive hull section is divided into 36 radial sectors.

26. The electrodynamic field generator of claim **23** further comprising a center sector of each positive hull section, located at the axis of the hull, these sectors each being electrically isolated from the radial sectors of the respective positive hull section.

27. The electrodynamic field generator of claim **26**, in which each center sector is made of an insulating material.

28. The electrodynamic field generator of claim **26**, in which single radial sectors at four cardinal points of each positive hull section are electrically connected in parallel with a conductive center sector of that positive hull section.

29. The electrodynamic field generator of claim **1**, in which the principal hull-coolant system comprises one or more primary thermal conduits conveying a high-temperature heat transfer fluid and incorporates a heater to pre-heat that heat transfer fluid and, in turn, the cathode resistors, dielectric buffers, and primary cathodes.

30. The electrodynamic field generator of claim **29**, in which the high-temperature heat transfer fluid is liquid sodium.

31. The electrodynamic field generator of claim **1** in which the primary cathode material is selected from a group consisting of pure tungsten metal and tungsten, as well as alloys thereof, which is or have been impregnated with thorium.

32. The electrodynamic field generator of claim **1**, in which the primary cathode resistors are made of a ceramic material which becomes electrically conductive at an elevated operating temperature.

33. The electrodynamic field generator of claim **32**, in which the ceramic cathode resistor material is selected from a group consisting of the steatite and cordierite families of compounds.

34. The electrodynamic field generator of claim **1**, in which the dielectric buffer encapsulating each primary cathode resistor is made of sodium bismuth titanate.

35. The electrodynamic field generator of claim **1**, further comprising a secondary hull coolant or heat transfer system which in turn comprises one or more secondary thermal conduits conveying a heat transfer fluid and which is thermally coupled to the neutral and negative sections of the hull.

36. The electrodynamic field generator of claim **35**, in which the heat transfer fluid is a cryogenic coolant.

37. The electrodynamic field generator of claim **36**, in which the heat transfer fluid is a liquefied gas selected from the group consisting of liquid air, liquid nitrogen, and liquid helium.

38. The electrodynamic field generator of claim **1** further comprising a plurality of stationary electromagnetic armatures arranged in two circular groups or arrays within the hull volume, one array above the upper surface of the rotor and one below the lower surface thereof, each armature comprising:

a ferromagnetic body or core having a first end section adjacent to the rotor, a second end section adjacent to a neutral section of the hull, and a length therebetween which is of a reduced cross-sectional area with respect to either end section; and

at least one electrical winding wound on either end section of the core, wherein an electric current in the winding induces a magnetic field in the core which is superimposed on the rotor or on the external electric field.

39. The electrodynamic field generator of claim **38** in which each of the armatures has two windings, comprising:

a first winding on the first end section, wherein a current in the first winding causes a magnetic field to be superimposed on the rotor; and

a second winding on the second end section, wherein a current in the second winding causes a magnetic field to be superimposed on the external field.

40. The electrodynamic field generator of claim **38**, in which the armature core material is selected from a group consisting of pure annealed iron and low-carbon steels.

41. The electrodynamic field generator of claim **38**, further comprising a DC power supply electrically coupled to at least one winding of each stationary armature.

42. The electrodynamic field generator of claim **38**, further comprising an AC power source electrically coupled to at least one winding of each stationary armature.

43. The electrodynamic field generator of claim **1**, in which the mechanical output of one or more motor drives is coupled to the rotor through a gear on each motor drive's mechanical output meshing with one or more driven ring gears about the inner circumference of the rotor.

44. The electrodynamic field generator of claim **1** further comprising a pair of rolling bearing assemblies or ballraces for supporting the rotor, one around the upper surface of the rotor and one around the lower surface thereof with both ballraces being mounted near the rotor's outer circumference; each ballrace comprising a first race half or raceway mounted on the rotor and a second raceway mounted to the hull opposite the first raceway, defining a bearing ball passage therebetween, and a plurality of bearing balls contained in the ball passage.

45. The electrodynamic field generator of claim **1**, further comprising a central control cabin inside the hull within the aperture or inner circumference of the rotor.

46. The electrodynamic field generator of claim **1**, in which the cathode resistors each comprise a fixed positive reduced-voltage tap whereby an onboard or external electrical load may be connected between the fixed taps and the primary cathodes.

47. The electrodynamic field generator of claim **1**, in which the heat transfer system is thermally coupled to an external heat exchanger.

48. The electrodynamic field generator of claim **1**, comprising one or more superconductive current storage rings utilized to deposit and retrieve an expendable store of ambient electron charge during extended operation in the vacuum of space.

* * * * *

US 20050230525A1

(19) **United States**
(12) **Patent Application Publication**
Paterro
(10) **Pub. No.: US 2005/0230525 A1**
(43) **Pub. Date: Oct. 20, 2005**

(54) **CRAFT WITH MAGNETICALLY CURVED SPACE**

(76) Inventor: **Von Friedrich C. Paterro**, Makati City (PH)

Correspondence Address:
OLIFF & BERRIDGE, PLC
P.O. BOX 19928
ALEXANDRIA, VA 22320 (US)

(21) Appl. No.: **10/812,018**

(22) Filed: **Mar. 30, 2004**

Publication Classification

(51) **Int. Cl.^{7}** **B64C 15/02**; B64C 29/00

(52) **U.S. Cl.** .... **244/23 B**; 244/23 C

(57) **ABSTRACT**

A craft can navigate through a magnetically curved space provided by an electromagnetically charged hull. The hull contains a plurality of sectionalized and independent electromagnetic plates forming portions of the hull exterior wall that can be adaptively energized to have a desired polarity. When combined together, these sectionalized plates can provide an ultra high frequency electromagnetic field of a strength that can magnetically curve the space around the craft. This enables the craft to expeditiously travel through air, space or water. A novel energy generating system using a combination of jet engines and high frequency oscillators is also provided.

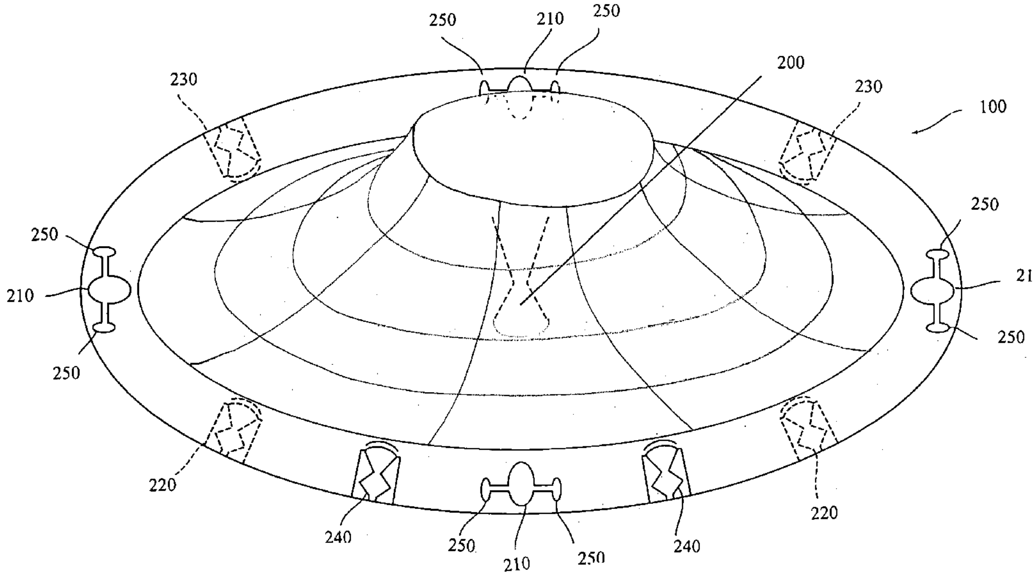

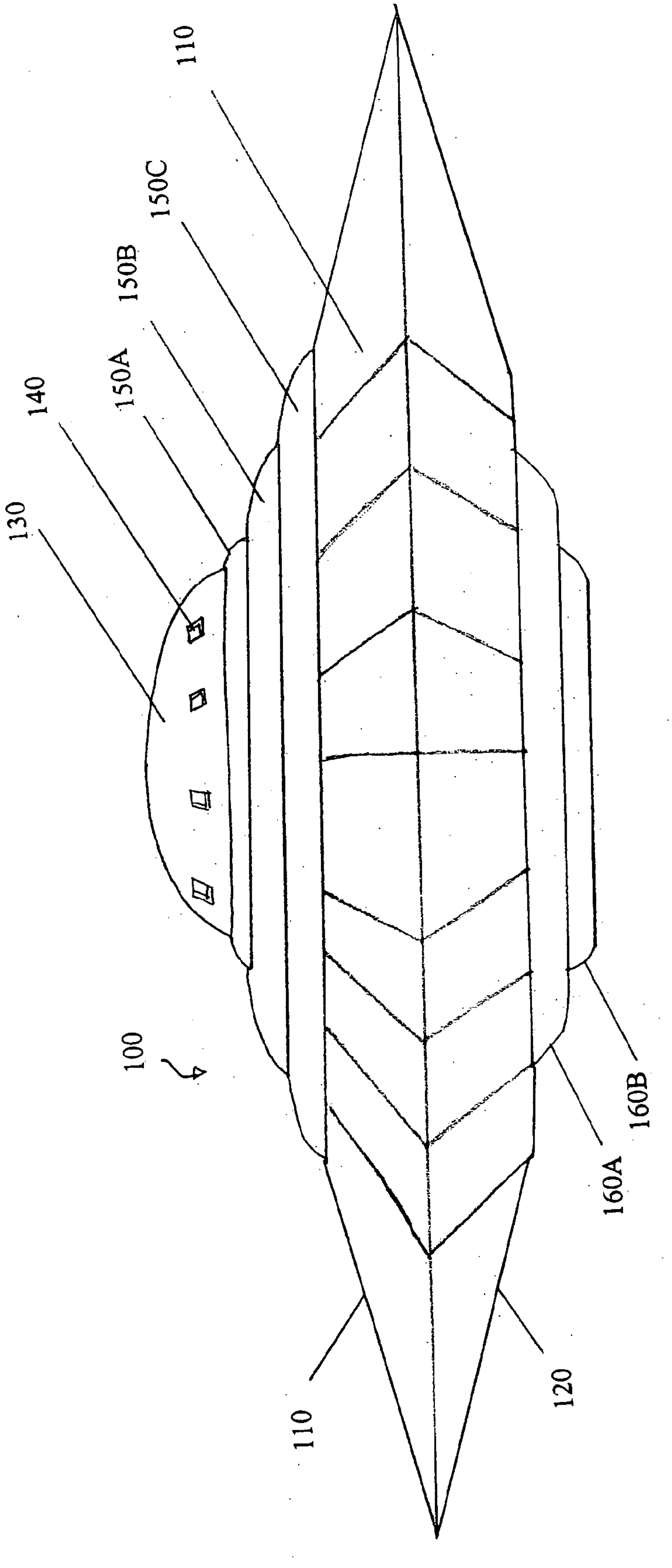
110
150C
150B
150A
140
130
100
160B
160A
120
110

FIG. 1

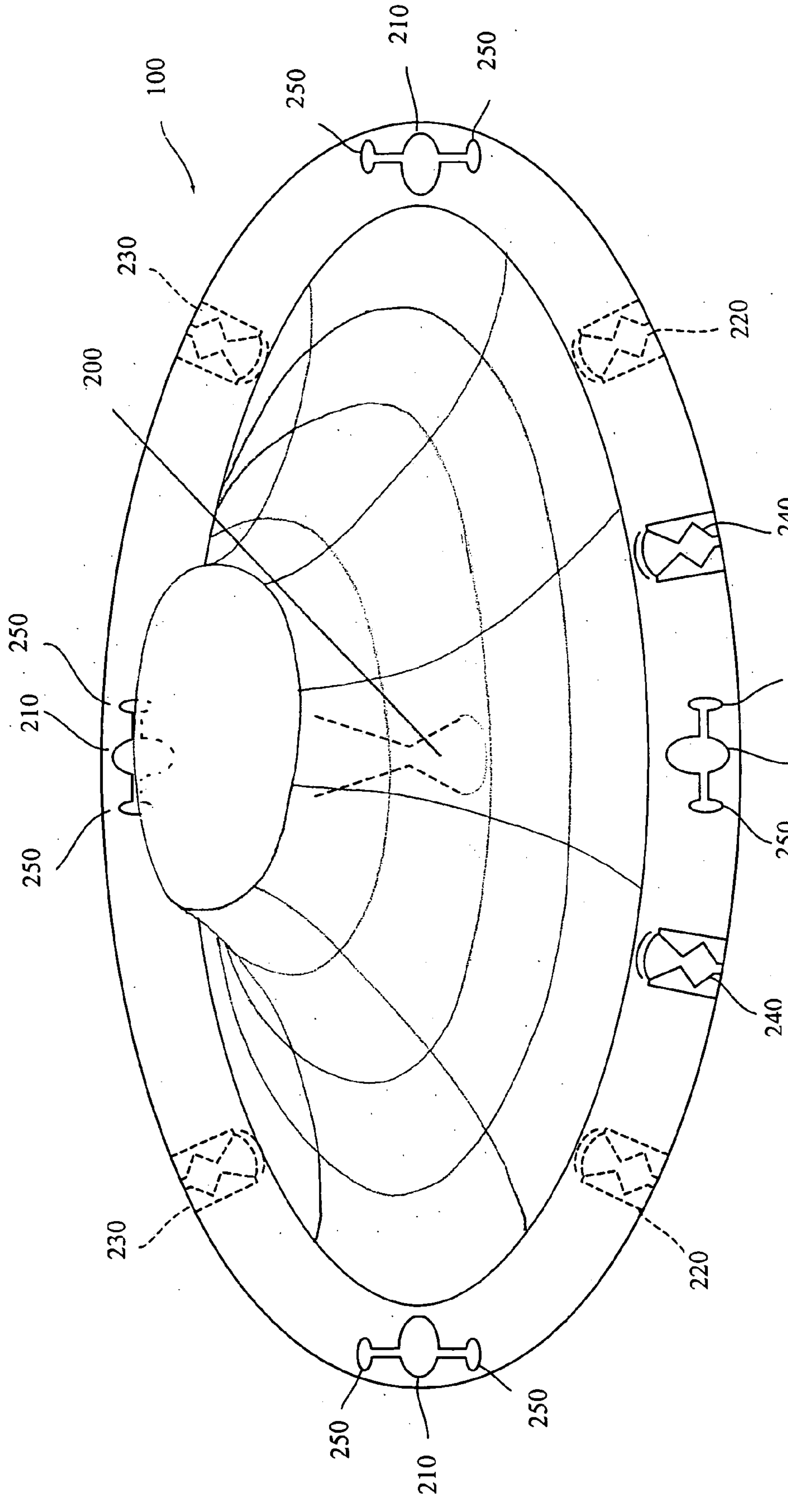

FIG. 2

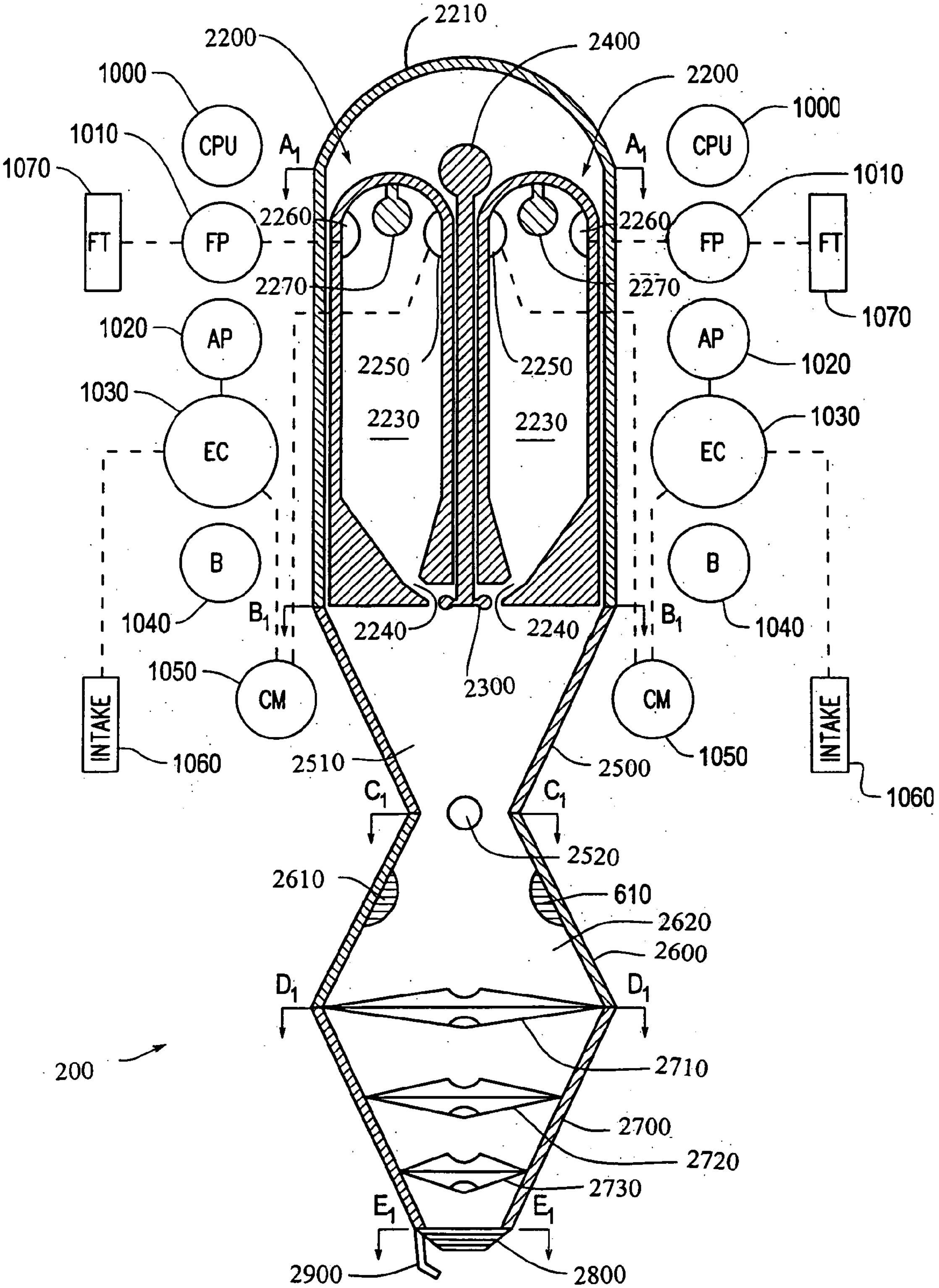
CPU
CPU
A1
A1
FT
FP
FP
FT
AP
AP
EC
EC
B
B
B1
B1
CM
CM
INTAKE
INTAKE
C1
C1
D1
D1
E1
E1

FIG. 3

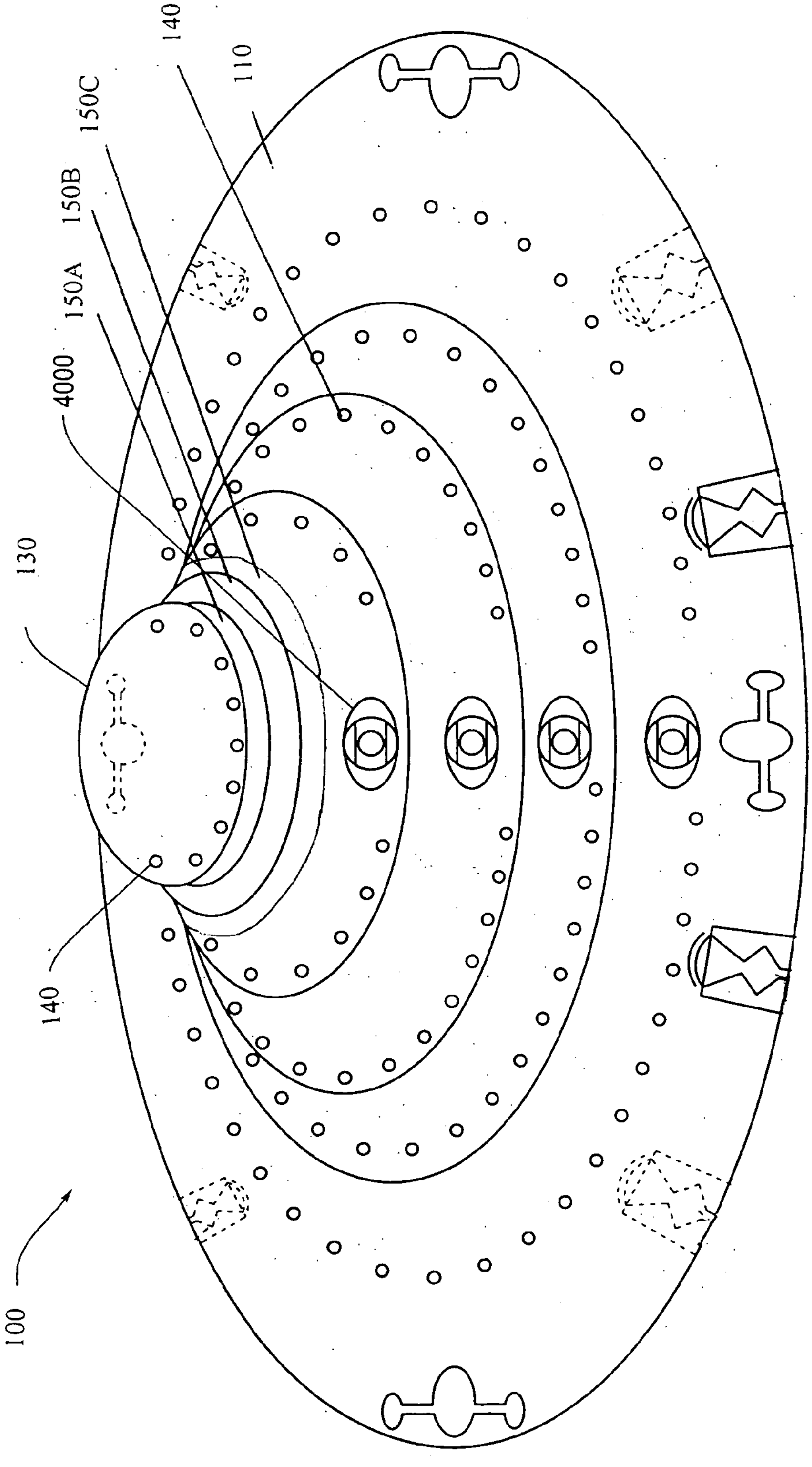
140
110
150C
150B
150A
4000
130
140
100

FIG. 4

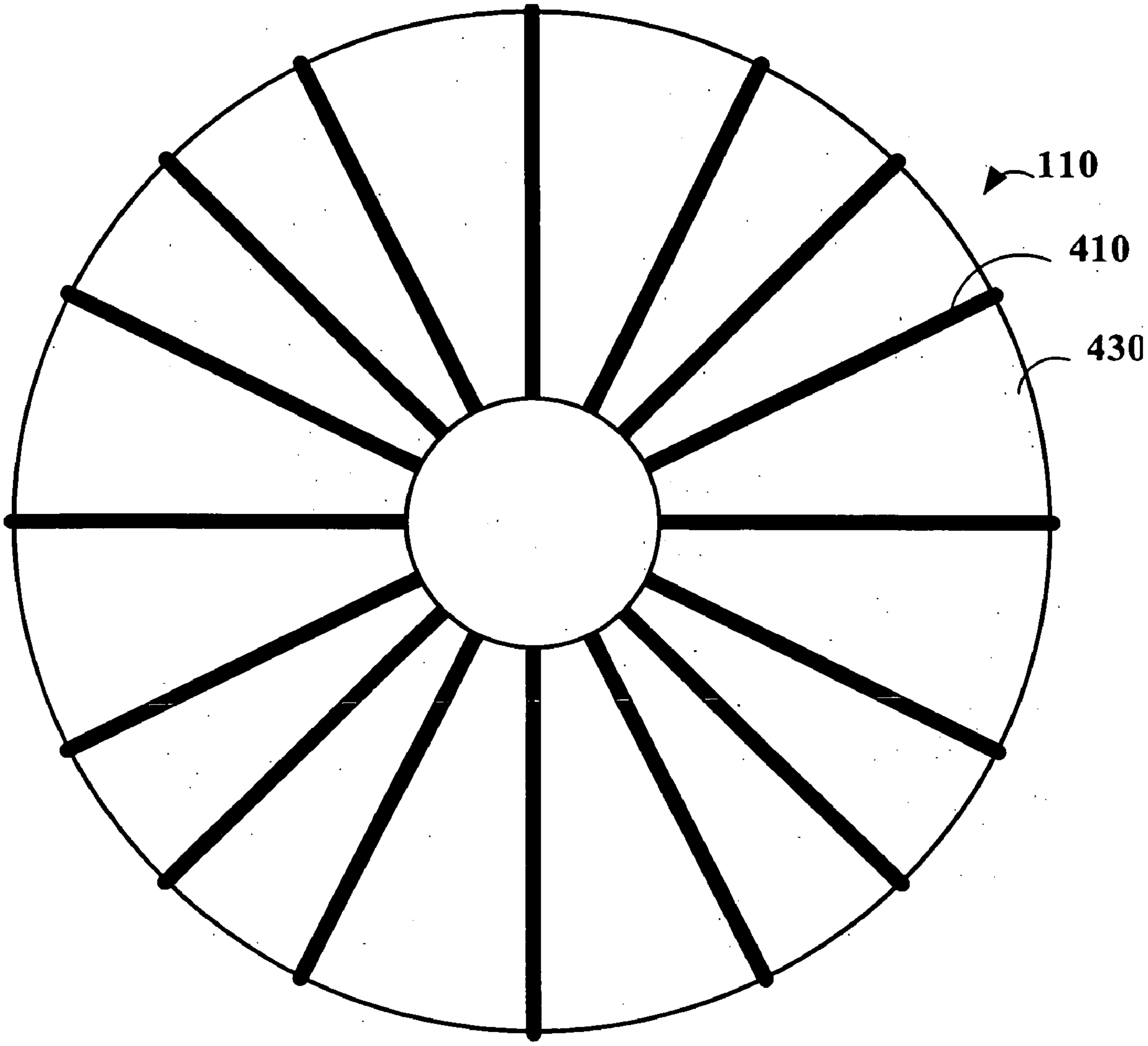

FIG. 5A

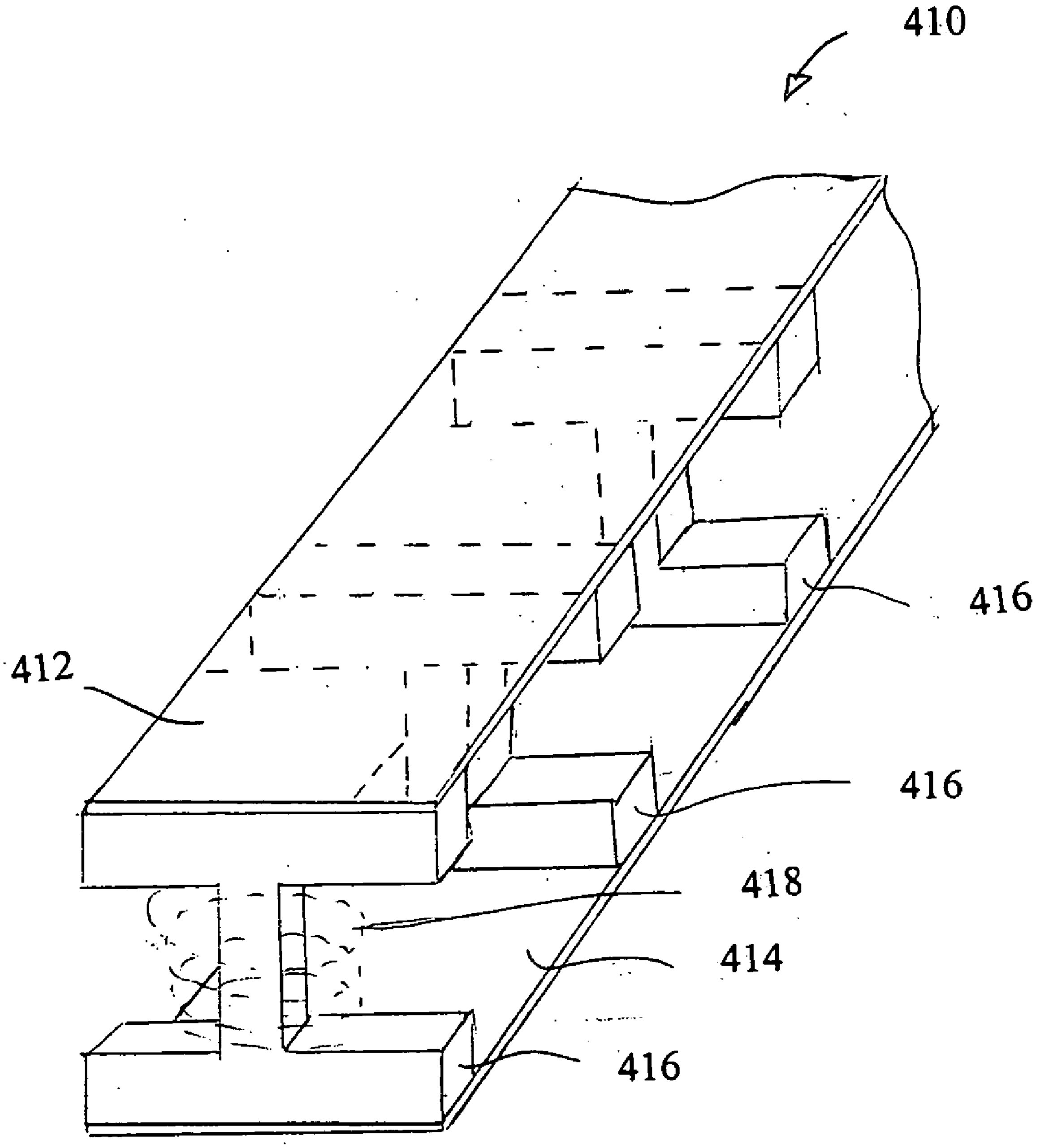
410
416
412
416
418
414
416

FIG. 5B

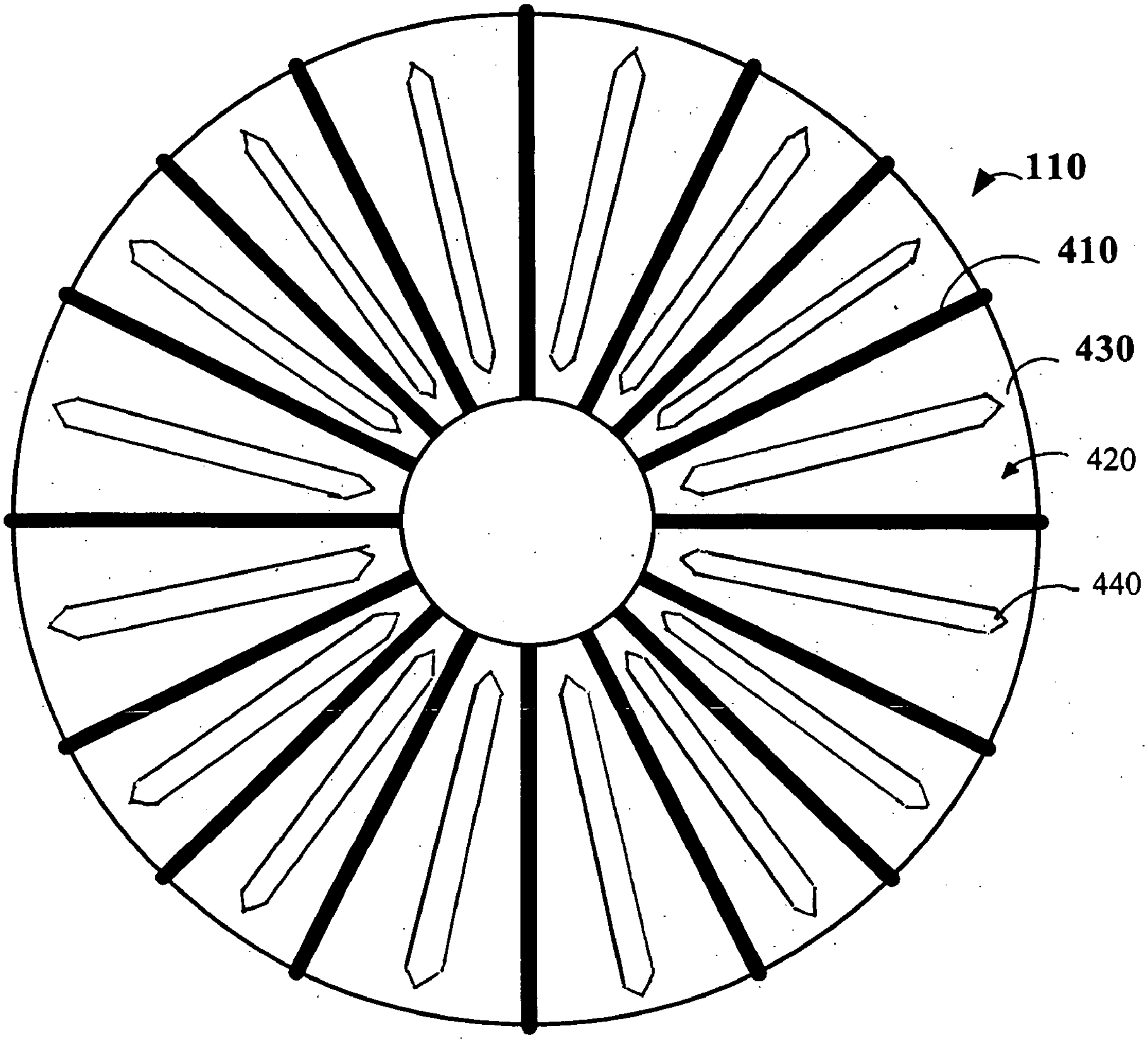
110
410
430
420
440

FIG. 5C

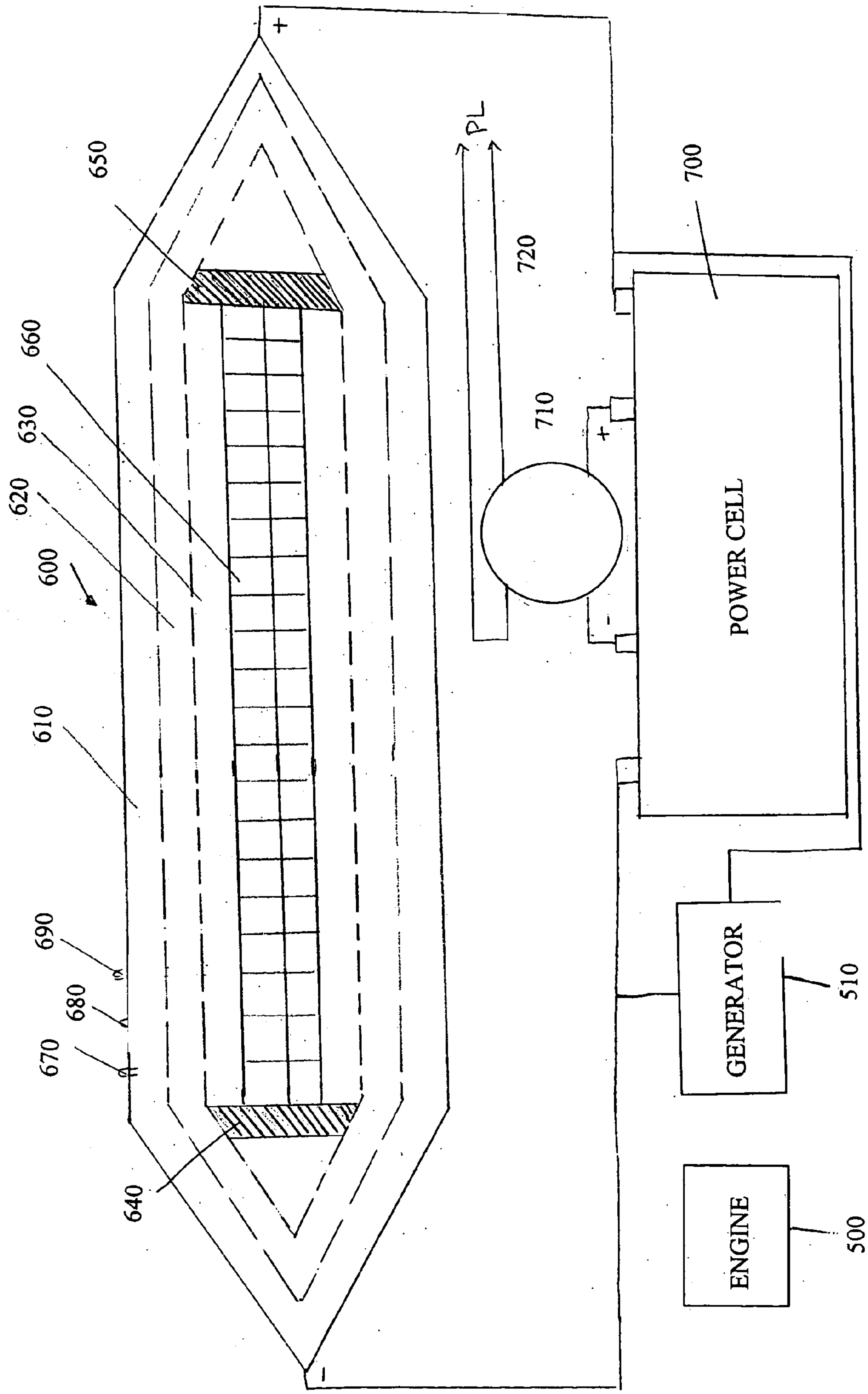
+
-
PL
+
-
POWER CELL
GENERATOR
ENGINE

FIG. 6

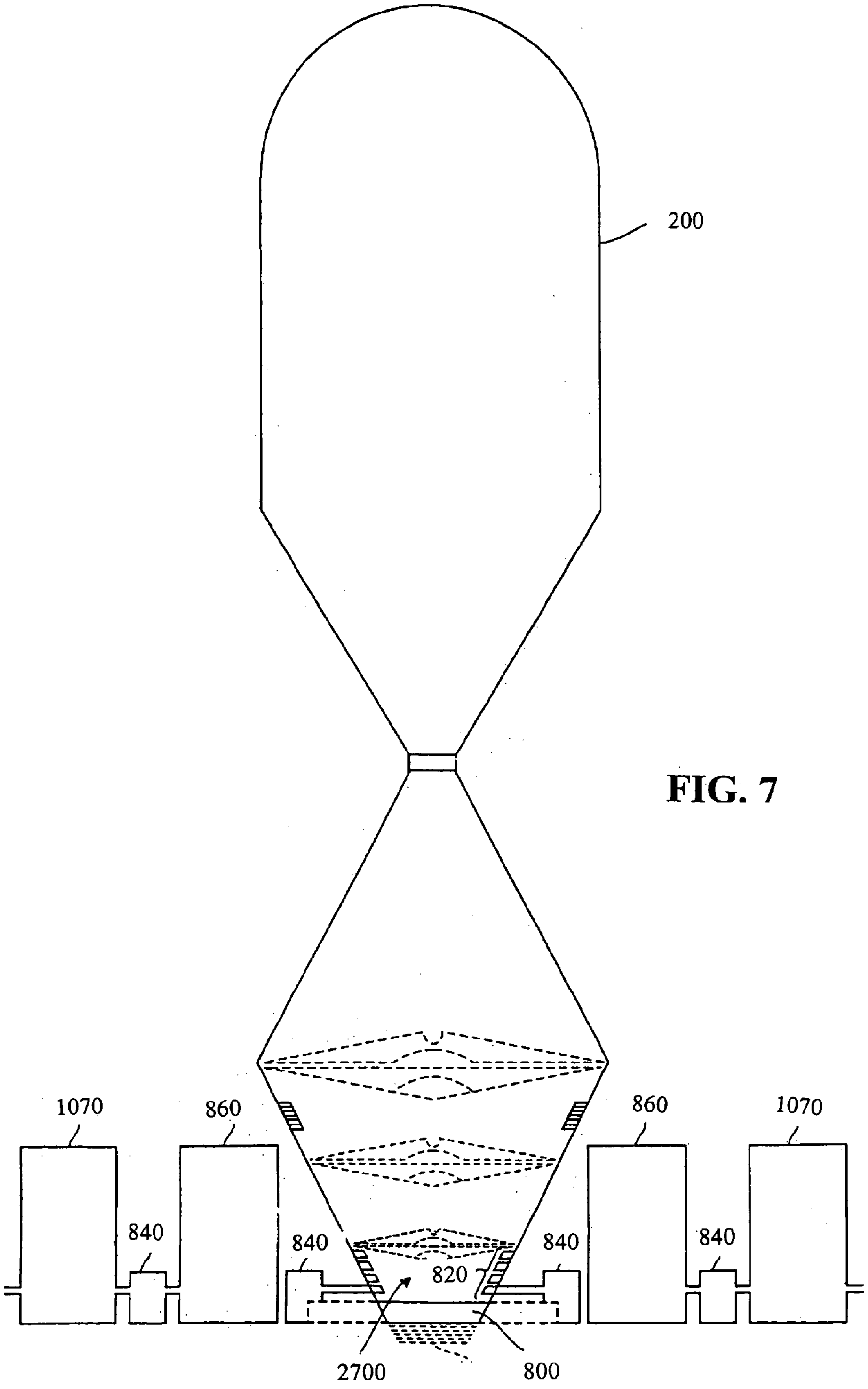
200
1070
860
840
840
840
820
860
1070
840
2700
800

FIG. 7

CRAFT WITH MAGNETICALLY CURVED SPACE

BACKGROUND OF THE INVENTION

[0001] 1. Field of Invention

[0002] This invention relates to a craft that can be propelled through a magnetically curved space. More particularly, the inventive craft has a hull built of electromagnets that generate a controllable electromagnetic field around the craft capable of reducing pressure forces acting on the craft. The electromagnetic field may be powered by use of one or more ultra high frequency oscillators.

[0003] 2. Description of Related Art

[0004] Numerous aircraft and spacecraft have undergone several innovations to increase the power and efficiency of their engines, fuel economy, speed, load capacity, and improved hull designs. However, all flying craft of today are still subjected to nature's ever present atmospheric pressures coupled with the natural effects of gravity, which adversely limit speed, load factor, size and design of hull, and much more, including flyability.

[0005] There is a need for an efficient, economical, bigger and faster, safe and versatile flying craft. In particular, there is a need for a craft capable of flying within the earth's atmosphere and beyond in which the effects of atmospheric pressure are reduced to a minimum or perhaps cancelled, allowing the flying craft to fly at extreme speeds without the danger of being torn apart and, in the process, reducing the full effects of Earth's gravitational forces.

[0006] There also is a need for a submersible craft capable of handling great pressures acting on the craft, allowing it to travel to great depths or be propelled more efficiently.

[0007] There also is a need for an improved energy generation system for such craft.

SUMMARY OF THE INVENTION

[0008] In exemplary embodiments, Applicant has overcome long felt needs for improved flight efficiencies by inventing a novel craft capable of flying within the earth's atmosphere and into the depths of outer space at extreme speeds, or traveling in the depths of the sea, along pathways in a magnetically curved space using a powered hull.

[0009] In exemplary embodiments, the craft can be aircraft, spacecraft and sea craft. In a preferred embodiment, the hull is saucer-shaped.

[0010] In various exemplary embodiments, the hull includes a plurality of electromagnetic plates to provide the magnetically curved space.

[0011] In various exemplary embodiments, the electromagnetic plates extend radially around the hull.

[0012] In various exemplary embodiments, the electromagnetic plates are powered by jet engine-driven generators and/or high frequency oscillators.

[0013] Very strong electromagnetic fields are generated around the hull of the craft to magnetically curve or bend the space in close proximity to the hull of the craft, such as by pushing all or part of the atmospheric particles or matter contained in this space away from the craft. This decreases the atmospheric drag against the hull, which increases the craft's aerodynamic efficiency and reduces the effect of atmospheric pressure on the craft. This magnetic curving of the space in the vicinity of the craft also shields the craft from excess heat, such as when reentering a planet's atmosphere, by creating a barrier around the hull that deflects such atmospheric particles, so as to preferably keep the hull free from contact with the particles, reducing heat transfer to the craft and reducing pressures acting on the craft. As such, extreme heat is not generated at the hull. Instead, generated heat acts on particles remote from the craft and is pushed away by the radiated field so as to provide a replenishable frontal barrier of such particles.

[0014] This phenomenon of curving "space" can be likened to several known or determinable phenomenon. For example, a hovercraft in operation can be observed to have a curved "space" at the immediate area below the craft where an apron is installed around the periphery. This space extends between the craft and the surface of land and/or water and maintains the hovercraft in a levitated position. Similarly, known MAGLEV trains magnetically curve the space between the track and the train to push the train slightly off the ground so as to form an air bearing. That is, the train and track collectively push the space therebetween to achieve slight levitation of the train. Forward propulsion is achieved through manipulated switching of the generated field. Other visible examples of curving of "space" can be found. For example, if you move a strong electromagnetic field towards a body of water, a crater-like opening in the vicinity of the magnetic field displaces a certain volume of water, acting to curve the space in close proximity to the magnetic field. Similarly, this curving of space can be observed if one directs a beam of light towards an ultra-strong electromagnetic field. There will be a slight bending of light as it passes through areas of strong magnetic effect. These examples visualize the effect achieved by aspects of the invention.

[0015] The invention similarly pushes away atmospheric particles around its hull continuously or selectively to keep the hull free from contact with such particles to form a curved space where particles in the close vicinity of the craft are pushed or displaced away so as to provide a "space" around the craft that allows increased efficiency through a reduction of aerodynamic drag, as well as a reduced pressure force acting on the hull.

[0016] In other exemplary embodiments, Applicant has overcome long felt needs for improved navigation efficiencies by providing upper and lower hull surfaces formed from electromagnetic walls that can be sectionalized and separately adjusted to a desired polarity. For example, when both upper and lower hull surfaces are adjusted to a same polarity, a magnetic levitation effect can be realized, similar to that used in Japan's MAGLEV trains, to supplement the craft's vertical thrust and achieve an efficient hover operation. In another example, the polarity can be selected to be the same as the earth's magnetic field (i.e., the Earth's closest hemispheric field) so as to provide a supplemental repulsion force that can be used to propel the craft, particularly once it has left Earth's atmosphere. This magnetic field control can also be used to adaptively adjust the electromagnetic radiation profile emitting from the craft's hull for various purposes.

[0017] Propulsion for the craft may come from any of several sources, including steam power created by nuclear energy, such as disclosed in Applicant's U.S. Pat. No. 6,290,184 entitled Flying Craft With Air and Water Propulsion System, one or more quantum jet turbine engines as disclosed in Applicant's co-pending U.S. patent application Ser. No. ______ (Attorney Docket 102901) filed concurrently herewith, and/or other conventional propulsion sources, such as jets or rockets. Any of the above may be supplemented with a compound exhaust system, such as disclosed in Applicant's U.S. Pat. No. 6,367,739, which increases the efficiency of the propulsion source. The disclosures of U.S. Pat. No. 6,290,184, U.S. Pat. No. 6,367,739, and co-pending U.S. patent application Ser. No. ______ (Attorney Docket 102901) are hereby incorporated herein by reference in their entireties.

[**0018**] The craft is preferably double-walled. The inner wall is much thinner than the outer wall and made up of strong composite materials, such as alloys of iron, cobalt, aluminum, nickel and copper. Similarly, the outer wall is also made up of strong composite materials, preferably of the same composition. However, the outer wall is much thicker and in sectionalized plates. The walls are joined by a core that is powered to generate an electromagnetic force.

[**0019**] The hull of the inventive craft is thus a very strong electromagnet, which can be amplified to radiate an ultra high frequency magnetic field around it. The field can be powered by a plurality of, preferably 33, inboard independent high frequency field generators that are directly coupled to multiple, preferably 11, quantum jet turbine engines, each preferably with three, independent high frequency generators each. The quantum jet engines use the resultant exhaust gas flow to drive the inboard generators to provide an electrical power source, rather than serving as a propulsion source.

[**0020**] The amount of electricity available to meet the electrical needs in the production of the ultra high frequency magnetic field may be augmented by electricity produced by high frequency oscillators (cavity resonators) in vacuum. Such oscillators may be installed in between deep lower circular cavities of the bottom side of the outer wall sectionalized plates. Such oscillators may also take the form of oversized ring-shaped triode high frequency oscillators suitably installed, such as three located at a periphery of the bottom of a cockpit and two located at a periphery of a top of an exhaust system.

[**0021**] When the craft is in operation, an ultra high frequency magnetic field will be radiated by its hull, causing atmospheric particles in the space in the vicinity of the hull to curve or bend. This curved space allows the craft to be propelled expeditiously along the pathways of its magnetically curved space at extreme speeds, greatly reducing or canceling the effects of atmospheric pressures against the craft's hull and increasing its inertia. This can produce beneficial effects against the pull of gravity, allowing the spaceship to perform greatly beyond the allowable limits of conventional craft. With this invention's capability of curving magnetically the atmospheric particles immediately around the craft, the craft's maneuverability can thus be improved by effective reduction of aerodynamic drag.

[**0022**] By further using an efficient compound exhaust system, such as the one described in Applicant's U.S. Pat. No. 6,367,739, the inventive craft is able to greatly improve fuel economy, load capacity and other beneficial factors that limit current production craft.

[**0023**] During outer space travel, the inventive craft may deactivate all secondary high frequency generators installed in the lower perimeters of the propulsion system in order to fully use all expended superheated steam for propulsion only, making every drop of fuel more efficiently used, and allowing the spaceship to cover wide areas for every gallon of fuel. The inventive craft may also use super-chilled air and water (H_2O) as a propulsion source when operating within the earth's atmosphere.

[**0024**] During sea craft use, the inventive craft is capable of operating at substantial depths due to the electromagnetic field providing a magnetically curved space that shield's and protects the craft's hull from the full effect of the extreme pressures that would otherwise act on the craft. This allows the craft to operate at depths otherwise unobtainable.

BRIEF DESCRIPTION OF THE DRAWINGS

[**0025**] The invention will be described with reference to the following drawings wherein:

[**0026**] **FIG. 1** shows a side view of a saucer-shaped craft according to an exemplary embodiment of the invention;

[**0027**] **FIG. 2** shows an underside of the craft of **FIG. 1** and placement of individual jet engines and directional control nozzles according to exemplary aspects of the invention;

[**0028**] **FIG. 3** shows an exemplary quantum jet engine according to embodiments of the invention;

[**0029**] **FIG. 4** shows an exemplary passenger version of a flying craft showing a cockpit;

[**0030**] FIGS. **5**A-C show a partial cross-sectional bottom view of a sectionalized part of the outer wall of the hull of the craft of **FIG. 1**, a partial perspective cut-away view of the plates making up substantial portions of the hull of the **FIG. 1** craft, and a cross-sectional bottom view of the outer wall of the **FIG. 1** craft showing optional oscillators provided in deep cavities between adjacent plates;

[**0031**] **FIG. 6** shows a cross-sectional view of a triode high frequency oscillator according to exemplary embodiments of the invention; and

[**0032**] **FIG. 7** shows an alternative exemplary exhaust design, in which compressors are provided to recycle at least part of the exhaust gases for recirculation.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

[**0033**] The invention will be described with reference to **FIGS. 1-7**, which show an exemplary saucer-shaped craft design. However, the shape and configuration of the hull is not limited to such designs, but may encompass other hull shapes. **FIG. 1** shows a craft **100** that is saucer-shaped including top and bottom hull halves **110**, **120**, and cockpit **130** with one or more viewing windows **140**, which may be automatically opened and closed. External high frequency oscillators **150**A-C and **160**A-B are also shown, with three spaced oscillators **150**A-C being provided on the top hull **110** around cockpit **130** and two oscillators **160**A-B being provided on the bottom hull **120** around the craft's exhaust system. The high frequency oscillators **150**, **160** may be formed from tri-cylindrical triode tubes, often called cavity resonators, as better described below with reference to **FIG. 6**. The high frequency oscillators **150**, **160** supplement the production of electricity for the craft and empower the hull **110**, **120** to radiate an ultra high frequency magnetic field. A plurality of electromagnetic plates **410** may be provided along portions of the hull exterior as better described below with reference to **FIG. 5**. In various exemplary embodiments, the craft can serve as a flying craft.

[0034] **FIG. 2** shows exemplary placement of propulsion engines and navigation control equipment used in connection with operation of the craft. **FIG. 2** is a skeletal view of the inventive craft of **FIG. 1**. In particular, multiple combustion power plants, such as quantum jet turbine engines, are spaced about the hull, each with an assigned task. This includes central engine **200**, fixed vertical engines **210**, two sets of reversible vertical and forward thrust engines **220**, two sets of reversible vertical and backward thrust engines **230**, two sets of fixed quantum jet turbine engines **240** tasked to produce forward thrust for propulsion, and various directional control nozzles **250** tasked to work as navigational guides. Additional directional control nozzles may be provided, such as at the lower rim of cockpit **130** (unshown).

[0035] The main and/or central engine **200** is preferably an octa-quantum jet propulsion system having eight individual engines sealed inside one compound exhaust system tasked to function as a fixed central vertical quantum jet turbine engine. Four permanent (fixed) vertical engines **210**, preferably four quantum engine units each containing four individual quantum jet turbine engines serviced by an independent compound exhaust system, are spaced around the periphery of the craft and tasked to produce additional vertical lifting power. The reversible engines **220**, **230** are tasked to work as reversible vertical/forward thrust engines (**220**) or reversible vertical/backward thrust engines (**230**). That is, engines **220**, **230** are pivotal through about 90° of rotation to adjust the direction of thrust. Engines **230** in addition to providing backward thrust for a reverse flight configuration, also serve as a brake mechanism used to slow the craft when operating with forward speed.

[0036] Exemplary craft mobility will now be discussed, referring back to the various Figures. For take-off from Earth, the two (2) reversible vertical and forward thrust engines **220** can be set for vertical work. Likewise, the two (2) reversible vertical and backward thrust engines **230** can be set for vertical work. The four (4) reversible engines will augment the five (5) permanent vertical engines (main vertical engine **200** and four fixed vertical engines **210**) to make nine (9) total vertical engines to provide adequate take-off propulsion. Such engines can also be used for landing.

[0037] Upon take-off, all the nine (9) vertical engines will be set at full power to raise the craft **100** to its desired altitude for horizontal flight. Upon reaching this altitude, the four (4) reversible engines **220**, **230** can be idled while the five (5) vertical engines **200**, **210** increase power to maintain a hover configuration. At this time, the four reversible engines **220**, **230** can be returned to a designated position. In the process, the reversible vertical and backward thrust engines **230** can be shut off while the two reversible vertical and forward thrust engines **220** can be set for forward thrust together with the fixed forward thrust engines **240**. At this time, the engines **220** will be oriented to face the forward movement direction. Thereafter, the four (4) engines **210** can be fully synchronized and powered accordingly to propel the craft with a desired flight profile and speed in a horizontal flight configuration. At this time, the high intensity electromagnetic fields may be generated by the various sectionalized plates **410** (**FIG. 1**). Thus will result in magnetic curving of the space immediately around the hull, thereby creating a magnetically curved pathway wherein the craft can travel at extreme speeds and be spared from the adverse effects of strong atmospheric pressure.

[0038] To bank left, the craft's fixed right vertical engine **210** will increase power while the fixed left vertical engine **210** will decrease power. To bank right, the opposite can occur. Righting to a horizontal position can be achieved by returning the low side engine to the power level of the opposite high side engine. To dive, the front vertical engine **210** will decrease in power as compared to the back vertical engines **210** and/or **240**. Then, while diving, the front vertical engine **210** can be powered just enough to sustain a desired angle of dive. To climb, the front vertical engine **210** can be increased in power while the power of the back vertical engine **210** is decreased, and the craft will climb with its front vertical engine **210** sustaining power enough to support the desired angle of climb. To return to normal, the front vertical engine **210** can be decreased in power, while the back vertical engine **210** is increased in power.

[0039] For increased mobility, the craft can be equipped with multiple directional control nozzles, such as nozzles **250** shown in **FIG. 2**. The directional control nozzles are installed around the hull and control the direction of the spaceship while in flight by guiding it to whichever direction is desired.

[0040] An exemplary embodiment of a quantum jet engine is illustrated in **FIG. 3**, which shows multiple quantum jet turbines housed in a common air-tight sealed exhaust system. The jet engines are preferably sized and symmetrically arranged within the exhaust system as shown, so as to provide a commonly and centrally oriented gas exhaust flow path.

[0041] The jet engines do not take in air directly from the atmosphere as in conventional jet engines. Rather, air or oxygen are received through sealed feed lines from efficient and independent on-board air compressors on the craft or externally provided for the engine. The air compressors may receive and transfer to the quantum jet engines air/oxygen received from either a remote storage tank or a remote air intake separate from the sealed jet turbine engines. The incoming air may be filtered as desired. This puts an end to the numerous problems associated with conventional jet engine designs that are prone to sucking large objects into their jet engine intakes.

[0042] In particular, **FIG. 3** shows a quantum jet turbine engine system **200** including multiple separate quantum jet engines **2200** housed within a single, common sealed exhaust system, preferably made up of sections A1, B1, C1, D1 and E1. Each quantum jet engine **2200** is housed in section Al of the exhaust system and includes an outer casing **2210** having a sealed, airtight top and converging lower walls **2220** defining a combustion chamber **2230** therebetween. Each quantum jet engine **2200** further includes a combustion exit orifice **2240**.

[0043] Within each combustion chamber **2230** are located one or more air nozzles **2250**. Air nozzles **2250** are operably connected to an electric air compressor **1030** through a suitable airtight, sealed feed line (unshown). Flow from the compressor **1030** to air nozzles **2250** may be enhanced by air pump **1020** provided in-line between compressor **1030** and air nozzles **2250**. Electric air compressors **1030** may receive air/oxygen from a suitable remote source, such as an on-board storage tank or through shown intake **1060**, which is in communication with the atmosphere but is provided remote from the jet engines **2200**. Suitable filtering may be provided at or between the intake **1060** and electric compressors **1030** to prevent large objects from entering the system.

[**0044**] Also within each combustion chamber **2230** are located one or more fuel nozzles **2260**. Fuel nozzles **2260** are operably connected to an on-board fuel storage tank **1070** through a suitable airtight, sealed feedline (unshown). Flow from the tank **1070** may be enhanced by a fuel pump **1010** provided in-line between the tank **1070** and fuel nozzles **2260**.

[**0045**] A spark generator **2270** is also provided within the combustion chamber **2230** of each jet engine **2200**. Spark generators **2270** may receive electrical power from one or more on-board batteries **1040**, or from generator **2400** provided within the common exhaust system. Generator **2400** may be operably connected through a shaft or other structure to a turbine **2300** having one or more turbine blades placed in the exit path of the combustion exit orifices **2240** as shown. Upon generation of combustion gases exiting the various jet engines **2200** through orifices **2240**, rotation of turbine **2300** will occur, which can be used with known and conventional structure to generate electrical energy from generator **2400**. Electrical output from generator **2400** may be electrically connected to batteries **1040** for recharging purposes and/or may be used to power various auxiliary devices, such as the hull's electromagnetic plates, preferably by providing power to a common power grid. Other accessories, such as processor **1000**, fuel pump **1010**, air pump **1020**, electric air compressors **1030**, and cooling mechanisms **1050** may also be powered by this electrical output.

[**0046**] During operation, quantum jet turbine engines **2200** are started by activating battery power to both the air and fuel pumps **1020**, **1010**, respectively. Upon reaching suitable operating pressures, a desired amount of air and fuel will be fed to combustion chambers **2230** while spark generators **2270** are electrically activated. Upon initial ignition, processor **1000** can cut off battery current and simultaneously activate the main electric air compressors **1030**, while simultaneously activating the fuel and air pumps and other electrical devices by way of current flowing from generator **2400**, which is suitably sized to power all required electrical devices.

[**0047**] The inventive quantum jet turbine system is extremely versatile and adaptable to a multitude of possible fuel sources, such as high grade kerosene, high grade diesel fuel, alcohol, or other solid or liquid fossil fuels. It can also operate on a mixture, such as a 70/30 mix of high grade (distilled) alcohol (C_2H_6O, C_2H_50H, or CH_3OH) plus distilled purified water (H_2O), which results in an efficient, safe and more environmentally friendly fuel that can be smokeless. Other applications may use a 50/50 mixture of alcohol and water, or may use 100% purified water alone (or with superchilled air) as a steam-powered version, that are completely environmentally friendly solutions that do not rely on fossil fuels.

[**0048**] For example, in one exemplary design, it is possible to use a fuel mix of 70% high grade alcohol (C_2H_6O) plus 30% distilled water (H_2O), considering the physical properties of both compounds wherein alcohol has a low boiling point of about 375° F. (197.2° C.) and distilled water has a boiling point of 212° F. (100° C.). Both compounds should be distilled to make them more efficient in achieving faster conversion from liquid to gaseous state, due to the pure substances having no other minerals or deposits that are not combustible and could solidify and produce nozzle clogging or contamination to the combustion chamber walls **2210**, which can cause maintenance problems.

[**0049**] Most alcohols and water mix well. As such, the combination is suitable as a mixture. When this fuel mix is fed to the combustion chambers **2230** and ignited by spark generators **2270**, the alcohol portion of the mix burns easily, raising the temperature inside the combustion chambers **2230** to over 100° C. in a very short time. Thus, expanded gases from the burnt alcohol will start moving at extreme speeds. Likewise, the water portion of the mix (30%) will be rapidly heated and boiled into steam at 100° C., at which time it also expands and moves at great speeds through the combustion chambers **2230** towards exit orifices **2240** where the accelerating and expanding gases pass across turbine **2300**. This generates electrical power from generator **2400** used to continue operation of all electrical accessories.

[**0050**] The exiting combustion gases enter an upper gas reaction area **2510** formed from converging walls **2500** of exhaust section **B1**. In this section, the exiting gases further expand and develop high pressure and temperature, ever continuously expanding and rushing toward automatic adjustable gas entry point **2520** where the exiting gases then enter a lower gas reaction area **2620** formed by diverging walls **2600** of exhaust section **C1**. In lower gas reaction area **2620**, the exiting gases further increase in pressure and temperature and enter the first stage of a multiple stage compound exhaust system **2700** provided at section **D1** of the exhaust system. As shown, there are three stages formed by stage sections **2710**, **2720** and **2730**. Continued flow paths of the exiting gases develop multiple action and reaction forces, acting to further extract kinetic force from the gases and further providing thrust force to propel the jet and associated craft upward. A suitable exemplary multiple stage compound exhaust system is the **3**-stage compound exhaust system disclosed in U.S. Pat. No. 6,367,739, the subject matter of which is hereby incorporated herein by reference in its entirety.

[**0051**] Upon exiting from compound exhaust system **2700**, exiting combustion gases are received by thrust vector nozzle **2800**, which can be suitably controlled to direct the exiting gases in a desired thrust vector that may be other than in axial alignment with the exhaust system. Owing to the sealed intake structure, such a jet engine will operate with reduced sound level than that typically found on conventional jet engines that include a large open-mouth intake system. If additional sound reducing properties are desired, a conventional sound cancellation device **2900** can be installed to the end of the exhaust as known in the art. Additional details on the operation of the quantum jet engines can be found in Applicant's incorporated co-pending U.S. patent application Ser. No. ______ (Attorney Docket 102901).

[**0052**] Although shown above to describe operation of main quantum jet turbine engine system **200**, the other jet turbine engines **210**, **220**, **230** and **240** can have the same general quantum jet engine configuration, albeit preferably smaller in scale.

[**0053**] **FIG. 4** illustrates an exemplary passenger craft according to the invention showing cockpit **130**, wide automatic windows **140** provided not only around cockpit **130**, but also spaced around upper hull **110**, and wide automatic doors **310**. The design features of this craft make it very versatile in that it can be applied to all military and civilian models, specifically military craft, bulk cargo carriers, fire fighters, passenger flying craft, tankers, transporters, space shuttles, and others. It also has applications other than flying craft, such as a submersible craft capable of navigating at great depths in the sea.

[**0054**] FIGS. 5A-C illustrate the preferred double-walled hull structure. **FIG. 5A** is a cross-sectional bottom view of a sectionalized part of upper hull **110**, with the inner hull wall removed, showing the internal radial structure of the electromagnetic plates **410**. Lower hull **120** is similar, but reversed. **FIG. 5B** illustrates a partial perspective cut-away view of the electromagnetic plates **410** making up a portion of the hull. **FIG. 5C** is a cross-sectional bottom view of an assembled outer hull wall, with the inner hull wall missing, to show the radial internal structure of the electromagnetic plates **410** and optional high frequency oscillators **440** provided in deep spaces **420** between adjacent radial extending plates **410**.

[**0055**] Upper hull **110**, which extends around cockpit **130**, includes a series of sectionalized electromagnetic plates **410**, shown in exemplary radial form, and conventional wall surfaces **430** provided between adjacent plates. The combination of conventional wall surfaces and plates **410** form the outer hull of the hull **110**. As better shown in **FIG. 5B**, each electromagnetic plate **410** includes an outer wall **412**, an inner wall **414**, and magnetic cores **416**, each with a plurality of coil windings **418** wrapped around it.

[**0056**] The illustrative electromagnetic sectionalized plates **410** can come in many sizes and/or shapes to accommodate this. For example, each could be 1 foot wide by 200 feet long or 1 meter wide by 200 meters long, and extend radially from the periphery of the upper cockpit to the lower edge of the craft hull. Alternatively, the sectionalized plates could be installed perpendicularly or peripherally around the upper and lower hull sections.

[**0057**] In an exemplary embodiment, the consolidated sectionalized upper walls together comprise a large amount of the surface area of the upper hull, such as about fifty percent (50%). However, when higher electromagnetic fields are generated, the surface area can be reduced. It is also desirable to have an outer peripheral rim of the upper and lower hulls **110**, **120** taper, so as to lock or mate with a similar outer wall on the corresponding mating hull. This will make the hull airtight.

[**0058**] The double-walled hull structure of inner and outer plate walls **414**, **412** is preferably made of a strong composite material with alloys of iron, cobalt, aluminum, nickel, and/or copper so as to provide a good electromagnetic property. Walls **430** may be made of similar materials. Plate walls **412**, **414** may be riveted, welded or otherwise fixed to cores **416** as known in the art.

[**0059**] The electromagnetic cores **416** are made up of strong composite materials, preferably with alloys of iron and nickel (PERMALOY), which are suitable for use in generation of a high electromagnetic field. The cores **416** preferably are I-shaped, with a large number of turns of high frequency, insulated electro-wire wrapped around each "I" continuously from end to end. The cores preferably have a depth of about two (2) feet or more from top to bottom, and are connected to a power source, such as the power grid in **FIG. 6**, preferably in parallel, for better and failsafe performance. This minimizes the possibility of malfunctions that may occur if plates **410** are connected to a power source in series. By making the cores **416** have a depth of about 2 feet or more, deep cavities **420** (**FIG. 5C**) can be provided between the upper and lower walls, which can house one or more oscillators (cavity resonators) **440** as shown in **FIG. 5C**. Such cavities may be insulated or the like to protect the oscillators from the coil wires **418** or the like.

[**0060**] The highly magnetized inner wall **414** is much thinner and preferably made up of much larger metal plates. The highly magnetized outer wall **412** is preferably much thicker and made up of sectionalized metal plates with the deep cavities **420**. That is, oscillators **440** can be provided in the spaces defined between the inner and outer walls **414**, **412** and adjacent plate sections **410**.

[**0061**] Coils **418** are wrapped around cores **416** with countless turns of heavy duty electrical wire to produce strong and high frequency electromagnets by infusing high electrical energy in the complete circuit. The size and number of sectionalized electromagnetic plates around the upper and lower hulls **110**, **120** should be powerful enough to enable the spaceship to radiate an ultra high frequency magnetic field around its hull and, in the process, magnetically curve the atmospheric particles in the area immediately around the craft. This will allow the craft to navigate expeditiously in any direction, greatly minimizing, if not totally canceling, the atmospheric pressure and drag acting on its hull during extreme speeds.

[**0062**] Likewise, the inner wall radiates the same high frequency magnetic field around its perimeter, and in the process magnetically curves the space around it. This is envisioned to counter weightlessness, allowing passengers, equipment, cargo and other movable objects to remain upright and firm during extreme speeds and difficult maneuvers.

[**0063**] All of the magnetic plates around the entire hull can be interconnected and operated independently or as a whole. For example, circuit terminals can be manipulated to engage or disengage a plate from power or reverse its polarity. Moreover, the sectionalized plates **410** are preferably independently isolated and insulated from each other.

[**0064**] All of the sectionalized plates may be controlled by complex multi-circuit panels run by efficient inboard computers, which direct all the necessary applications required, such as the electrical energy needed by the individual sectionalized metal plates to change or maintain each plate's polarity ("S" and "N" magnetic poles), either singly, by pairs or for an entire section of any side of the hull. For example, the hull can radiate a high frequency magnetic field with a desired "S" magnetic pole or an "N" magnetic pole at any given time using the isolated and insulated sectionalized metal plates, either singly, by pairs or by section. Polarity can be changed by introducing a large electrical current flowing from a positive terminal to a negative terminal. Polarity can be switched back by introducing a large electrical current flowing from the negative terminal to the positive terminal.

[0065] This, for example, allows the craft to intelligently decide at anytime and place when to shift to "S" pole magnetic field and "N" pole magnetic field, considering that the Earth is a giant magnet with a strong magnetic field around it and lines of force running between the poles that curve high into the atmosphere, even where the pull of gravity is weak. Earth's huge magnetic field differs in intensity in direct proportion to its distance from either the south pole or the north pole. Because same polarity magnets repel, this craft, while in operation around the earth's atmosphere, can have the capacity to radiate an ultra high intensity magnetic field of the same polarity, yet much stronger, than Earth's magnetic pole prevalent in a particular area. This may be beneficial to help propel the craft outward, particularly when gravitational forces have been overcome or are weak.

[0066] Although the craft's capabilities to interplay with the Earth's great magnetic field do not in totality remove the effects of the Earth's pull of gravity, the sum total of the repulsion forces of all of the craft's electromagnetic plates acted on by the Earth's magnetic field can constitute a great effect in the flyability of the craft.

[0067] The various sectionalized plates **410** can also be interchanged at will in various combinations (of "N" or "S" poles or combinations) or singly to selectively magnetically curve the space around specific portions of the hull and reduce the effects of atmospheric pressure against it. This produces unlimited advantages in the craft's ability to fly expeditiously beyond a conventional aircraft. This can be used to radiate an ultra high intensity magnetic field with an "S" magnetic pole on the entire lower section of the outer-wall and, likewise, radiate an ultra high intensity magnetic field on the upper section of the outer wall with an "N" magnetic pole or vice versa, if needed. For example, when both upper and lower hull plates **410** are adjusted to a same polarity, such as an "S" magnetic pole, a magnetic levitation effect can be realized, similar to that used in Japan's MAGLEV trains, to supplement the craft's vertical thrust and achieve an efficient hover operation. Also, when reentering into Earth's atmosphere, the lower plates can be activated to provide a forward magnetic shield that protects the front surfaces of the craft from the intense heat and pressure of reentry by the magnetically curved space it creates.

[0068] **FIG. 6** illustrates an exemplary electrical power source for the craft. The electrical requirements of the spaceship will be met by two (2) sources:

> **[0069]** 1. Electricity generated by a series of engines **500**, preferably eleven (11) quadro quantum jet turbine engines, each equipped to power three (3) inboard high frequency generators **510**, for a total of thirty-three (33) generators **510**.
>
> **[0070]** 2. Electricity generated by a series of vacuum high frequency oscillators **600**, preferably of the modified KLYSTRON type, which can take the form of ring shaped oscillators such as oscillators **150**A-C in **FIG. 1**, oscillators **160**A-B in **FIG. 1** or the elongated oscillators **440** in **FIG. 5C**. However, any form of oscillator can be used.

[0071] The engines **500** are preferably of the type described in **FIG. 3**, although any engine source could be used to power generators **510**. Although shown as inboard, generators **510** could also be external and of conventional types. Power generated by generators **510** can be fed to a central power grid, such as central ultra high frequency power cell **700** shown, through electrical lines. The engines **500** and generators **510** are suitably located within the confines of the flying craft.

[0072] In preferred embodiments, the main thrust engine **200** and other thrust engines **210**, **220**, **230** and **240** may each have one or more generators to assist in electricity generation (in addition to the **33** dedicated engines and generators **500**, **510**). At start up, one of the generators may power the main engine **200**. However, once powered, secondary generators may be activated to take over all the electrical requirements of the flying craft and all battery power can be cut off. At this stage, the remaining thrust engines **210**, **220**, **230** and **240** may be started using power generated by the generators. All excess power will be provided to the central power cell **700**. Power from power cell **700** will then energize all or various combinations of the sectionalized electromagnetic plates to produce a desired ultra high intensity magnetic field around the craft that will magnetically curve the space around the spaceship.

[0073] As shown, the oscillator **600** uses three (3) hollow tubular cavities made up of one (1) external hollow zinc tube **610** with two (2) inner independent hollow copper tubes **620** and **630**. Electrons oscillate in each of the three (3) tubes at a rate of more than one (1) billion cycles per second, and are moved by a magnet (unshown) or other source to make an electric current.

[0074] Also shown are a cathode plate **640** which is negatively charged, and an anode plate **650** which is positively charged. A grid **660** is provided between the plates through which electrons oscillate from the cathode plate **640** to the anode plate **650**. Direct current is supplied to the oscillator **600** and alternating current is produced. Although magnetism is preferred because of its availability on board, other sources of energy can be used to move electrons in order to produce an electric current, including friction, pressure, chemical action and light.

[0075] Three (3) suction air tubes **670**, **680** and **690** pump out the air inside the oscillator **600**, preferably to keep it in total vacuum. Power generated by oscillator **600** is fed to the power cell **700** by electrical lines. Power from power cell **700** can be amplified by a high frequency transistorized power amplifier **710**, which can feed power to a power line **720** that forms a power grid that provides power to all accessory devices of the flying craft.

[0076] The first tube **610** is zinc and electrons from the atmosphere travel into this tube. Because zinc gives up electrons freely relative to copper, the electrons will pass from the zinc tube to the copper tubes **620**, **630**, which become negatively charged. Electrons are particles of negative electricity, and are found in the atmosphere, within any object or on its surface, and also in a vacuum. In a vacuum, it is rather simple to produce a current of electrons. The mere heating of a filament in vacuum can produce a swarm of electrons. As they "boil off" from the filament, they can be induced to flow across the vacuum toward a positive connector from a power cell connected to the filament. Thus, the positive connector at some distance away will lure the electrons toward itself because positive charges attract negative charges.

[0077] All the high frequency oscillators (HFO) of this invention will be inter-connected to each other, such that the anode plate **640** of one oscillator **600** is connected to the grid **660** of a second one, and so on and so forth (in series), yielding much greater amplification and delivering the sum total of all their electric current to power cell **700**.

[0078] With the capacity of this invention to radiate an ultra high frequency magnetic field, radar could become obsolete, for it cannot penetrate through the magnetic field and give away the signature of the spaceship. It could thus serve as a "stealth" flying craft. Because of its ability to change its field generation at will, it can also rapidly change any magnetic profile it does give off.

[0079] The inventive craft and propulsion sources may use a compound exhaust system as in Applicant's U.S. Pat. No. 6,367,739. The thrust generating source may use, for example, water (H_2O) and/or liquid hydrogen (H_2) as fuel. In the **FIG. 7** embodiment which uses such a fuel source, the compound exhaust system **2700** may be provided with an automatic retractable air tight locking device **800** at the periphery of the exhaust chamber that will trap expanding gases. Multiple vacuum-locking doors **820** may be provided above locking device **800** around the exhaust chamber that are in communication with multiple efficient compressors **840**, which are capable of sucking out the trapped gases and pumping them into multiple condensers **860** that convert the gases back to a liquid state. Thereafter, the converted liquid is pumped back to a fuel tank **1070** to be recycled and later fed to the thrust source **200**. The vacuum-locking doors **820** are preferably located about one foot above the air tight locking device **800**.

[0080] In operation, gases pass from the thrust source **200** through the three (or more) stages of the compound exhaust system **2700** to the exhaust chamber where the gases are contained by the locking device **800**. Thereafter, the multiple vacuum-locking doors **820** may be opened to allow sucking of the gases by the multiple compressors **840** to the multiple condensers **860** where the gases are converted back to their liquid state and thereafter pumped back to the fuel tank **1070**.

[0081] The locking device **800** at the periphery of the exhaust chamber may be openable to allow maintenance. The locking device **800** may also be openable to allow the superheated exiting gases to be released to the atmosphere. This may form a thick cloud of vapors, which may be a harmless byproduct or a desirable stealth mechanism that can at least partially conceal the craft.

[0082] In another embodiment, the craft can be used for interplanetary flight. The hull in such an embodiment would be designed to withstand the pressures and temperatures encountered when exiting or entering Earth's atmosphere. Additionally, the hull would be pressurized. Optionally, when traveling through space, the exhaust gases can be vented back into the craft by a vacuum system powered by compressors so that the exhaust gases can be recycled.

[0083] When used for interplanetary flight, it may be desirable to rotate the craft about its axis at a predetermined speed while in motion to create an artificial gravity. This can be achieved using the various thrust sources, by the electromagnetic plates, or other source.

[0084] In another exemplary embodiment, the inventive craft can expeditiously operate under water, similar to the craft in Applicant's U.S. Pat. No. 6,290,184. However, because of the electromagnetic hull, it has far greater capacity to dive to the deepest depths of the ocean and withstand the extreme pressures present at such depths. This is because with the craft's ability to radiate an ultra-strong electromagnetic field around its hull, it can neutralize the destructive effects of such extreme pressures by manipulating the magnetic field to magnetically curve the space along its path to keep the full force of the pressures from reaching and acting on the hull.

[0085] This invention can well be the most environmentally friendly amongst all the flying crafts of today because of its capacity to utilize water (H_2O) and/or chilled air (O_2) as fuel. It also can travel at extreme speeds, much faster than any conventional aircraft of today, and can make radar obsolete because of its capacity to radiate an ultra high frequency magnetic field around its hull and magnetically curve the space around it. This escapes the harsh effects of atmospheric pressure against its hull and repels radar signals. This invention can well become the "Ultimate Spaceship".

[0086] While specific aspects of the invention have been described with respect to preferred embodiments of the invention, these are not intended to be limiting. Various modifications can be made without departing from the scope of the appended claims.

What is claimed is:

1. A craft, comprising:

an exterior hull defining an exterior craft surface of a predetermined surface area;

at least one propulsion source; and

an electrical generator,

wherein the hull includes a plurality of sectionalized electromagnetic plates, each of the plurality of electromagnetic plates including an outer wall and an inner wall fixedly provided on an electromagnetic core, the electromagnetic core including at least one turn of coil operatively coupled to the electrical generator to provide an electromagnetic force to each plate of a given magnitude and polarity, the electromagnetic plates collectively forming a substantial portion of the exterior craft surface to provide a controllable electromagnetic field around at least portions of the hull to magnetically curve the space adjacent the hull.

2. The craft according to claim 1, wherein the electromagnetic plates extend around the periphery of the hull of the craft.

3. The craft according to claim 1, wherein the electromagnetic plates extend radially around the hull of the craft.

4. The craft according to claim 1, wherein the electromagnetic plates are isolated from each other and are independently changeable between N and S polarity.

5. The craft according to claim 4, wherein at least one of the electromagnetic plates can be selectively adjusted to a same polarity as the Earth's closest magnetic pole.

6. The craft according to claim 1, wherein the craft is a flying craft.

7. The flying craft according to claim 6, wherein the flying craft has a saucer-shape with an upper hull portion and a lower hull portion.

8. The flying craft according to claim 7, wherein the at least one propulsion source includes a main central fixed vertical jet engine, a plurality of secondary fixed vertical jet engines provided around the periphery of the hull, at least one forward-oriented jet engine and at least one backwards-oriented jet engine.

9. The flying craft according to claim 8, further comprising a plurality of directional control nozzles around the hull.

10. The flying craft according to claim 6, wherein the flying craft has a saucer-shape with an upper hull portion and a lower hull portion, one or more sectionalized electromagnetic plates being provided on each of the upper hull portion and the lower hull portion, the collective polarity of the electromagnetic plates on the upper hull portion being the same as the collective polarity of the electromagnetic plates on the lower hull portion at least when the vehicle is hovering.

11. The craft according to claim 1, wherein the electrical generator includes at least one high frequency oscillator.

12. The craft according to claim 11, wherein the high frequency oscillator is in the form of one or more rings externally provided around the periphery of the hull.

13. The flying craft according to claim 11, wherein the high frequency oscillator is in the form of a long tube and is provided in a cavity defined between adjacent electromagnetic plates.

14. The craft according to claim 1, wherein the propulsion source is a sealed quantum jet engine having a compound exhaust system.

15. The craft according to claim 1, wherein the craft is a submersible water craft.

16. The craft according to claim 1, wherein the craft is a submersible craft.

17. A saucer-shaped flying craft, comprising:

a saucer-shaped exterior having a hull defined by an upper hull portion and a lower hull portion of predetermined surface areas;

a propulsion source including at least a main, central, fixed vertical jet engine, a plurality of secondary fixed vertical jet engines provided around the periphery of the hull, at least one forward-oriented jet engine and at least one backwards-oriented jet engine; and

an electrical generator,

wherein one or more of the upper and lower hull portions includes a plurality of sectionalized electromagnetic plates, each of the plurality of electromagnetic plates including an outer wall and an inner wall fixedly provided on an electromagnetic core, the electromagnetic core including at least one turn of coil operatively coupled to the electrical generator to provide an electromagnetic force to each plate of a given magnitude and polarity, the electromagnetic plates collectively forming a substantial portion of the exterior craft surface to provide an electromagnetic field around at least portions of the upper and/or lower hull portions to magnetically curve the space adjacent the hull.

18. The saucer-shaped flying craft according to claim 17, wherein one or more sectionalized electromagnetic plates are provided on each of the upper hull portion and the lower hull portion, and the collective polarity of the electromagnetic plates on the upper hull portion are the same as the collective polarity of the electromagnetic plates on the lower hull portion at least when the vehicle is hovering.

19. The saucer-shaped flying craft according to claim 17, wherein the electromagnetic plates extend radially around the hull of the craft.

20. The saucer-shaped flying craft according to claim 19, further comprising a cockpit located centrally on the upper hull portion, the electromagnetic plates on the upper hull portion extending from substantially a periphery of a cockpit to substantially an outer periphery of the upper hull portion.

21. The saucer-shaped flying craft according to claim 17, wherein a collective exterior surface area of the sectionalized electromagnetic plates is about one-half of the surface area of the hull.

22. The saucer-shaped flying craft according to claim 17, wherein the electrical generator includes at least one high frequency oscillator.

23. The saucer-shaped flying craft according to claim 22, wherein the high frequency oscillator is in the form of one or more rings externally provided around the periphery of the hull.

24. The saucer-shaped flying craft according to claim 22, wherein the high frequency oscillator is in the form of a long tube and is provided in a cavity defined between adjacent electromagnetic plates.

25. The saucer-shaped flying craft according to claim 17, wherein the electromagnetic plates are isolated from each other and are independently changeable between N and S polarity.

26. The saucer-shaped flying craft according to claim 25, wherein at least one of the electromagnetic plates can be selectively adjusted to a same polarity as the Earth's closest magnetic pole.

27. The saucer-shaped flying craft according to claim 16, wherein the propulsion source operates using a non-fossil fuel source.

28. A saucer-shaped flying craft, comprising:

a saucer-shaped exterior having a hull defined by an upper hull portion and a lower hull portion of predetermined surface areas, a cockpit being substantially centrally located on the upper hull portion;

a propulsion source including at least a main central fixed vertical jet engine, a plurality of secondary fixed vertical jet engines provided around the periphery of the hull, at least one forward-oriented jet engine and at least one backwards-oriented jet engine; and

an electrical generator that includes at least one ring-shaped high frequency oscillator externally provided on the periphery of the hull, wherein at least one of the upper and lower hull portions includes a plurality of sectionalized electromagnetic plates, each of the plurality of electromagnetic plates including an outer wall and an inner wall fixedly provided on an electromagnetic core, the electromagnetic core including at least one turn of coil operatively coupled to the electrical generator to provide an electromagnetic force to each plate of a given magnitude and polarity, the electromagnetic plates extending radially on the upper hull portion from substantially the cockpit to an outer periphery of the upper hull portion, the electromagnetic plates collectively forming a substantial portion of the exterior craft surface to provide an electromagnetic field around at least portions of the upper and/or lower hull portions to magnetically curve the space adjacent the hull.

* * * * *

Manuscript submitted to: Volume 3, Issue 2, 184-200.

AIMS Energy DOI: 10.3934/energy.2015.2.184

Received date 29 December 2014, Accepted date 14 April 2015, Published date 24 April 2015

Research article

Analysis of Passive RF-DC Power Rectification and Harvesting Wireless RF Energy for Micro-watt Sensors

Antwi Nimo*, Tobias Beckedahl, Thomas Ostertag and Leonhard Reindl

University of Freiburg, Department of Microsystems Engineering - IMTEK. Laboratory for Electrical Instrumentation. Georges-Köhler-Allee 106, 79110 Freiburg, Germany

* **Correspondence:** Email: antwi.nimo@imtek.de or reindl@imtek.de; Tel: +49-761-203-7221; Fax: +49-761-203-7222.

Abstract: In this paper, analytical modeling of passive rectifying circuits and the harvesting of electromagnetic (EM) power from intentionally generated as well as from ubiquitous sources are presented. The presented model is based on the linearization of rectifying circuits. The model provides an accurate method of determining the output characteristics of rectifying circuits. The model was verified with Advance Design System (ADS) Harmonic balance (HB) simulations and measurements. The results from the presented model were in agreement with simulations and measurements. Consequently design considerations and trade-off of radio frequency (RF) harvesters are discussed. To verify the exploitation of ambient RF power sources for operation of sensors, a dual-band antenna with a size of ~$\lambda/4$ at 900 MHz and a passive dual-band rectifier that is able to power a commercial Thermo-Hygrometer requiring ~1.3 V and 0.5 MΩ from a global system for mobile communications (GSM) base station is demonstrated. The RF power delivered by the receiving dual-band antenna at a distance of about 110 m from the GSM base station ranges from −27 dBm to −50 dBm from the various GSM frequency bands. Additionally, wireless range measurements of the RF harvesters in the industrial, scientific and medical (ISM) band 868 MHz is presented at indoor conditions.

Keywords: Ambient RF energy harvesting; Schottky diode rectifier; dual-band RF harvester; rectenna; RF to DC rectification model; wireless power transmission

---

1. Introduction

Recent advances in lowering power demands of wireless remote sensors has afforded the ability to power these sensors with micro-watt power levels. In situations where periodic change of the batteries is not desired, powering the sensors through wireless power sources provides an effective alternative. As a result, wireless RF energy harvesting from ambient sources or from dedicated transmitters has gained research interest for use in these sensor systems. The major challenge in passive wireless energy harvesting is the efficient transformation of the input EM power levels that are less than −10 dBm to the voltage and current demands of the direct current (DC) load. A passive RF to DC power harvester must be optimized at each stage of the circuit design for true implementation of this energy harvesting approach. An RF power harvester consists of a receiving antenna, an impedance matching/transformation network, DC power conditioning and or management and the wireless sensor to be powered. Figure 1 shows a schematic of a passive wireless RF energy harvester. The rectifier may be a single diode or diode-connected transistor, or a cascade of rectifiers. In-between the antenna and the rectifier is an impedance matching or transformation network. The matching network consists of passive components providing inductive or capacitive reactance. The load represents the DC storage or transformation elements and the sensor to be powered.

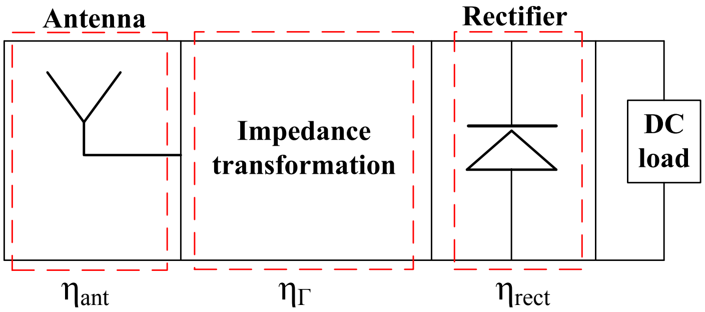

Figure 1. Block diagram of a passive wireless RF energy harvester. η_{ant}, η_{Γ} and η_{rect} are the efficiencies of the receiving antenna, impedance transformation network and the rectifier respectively.

The design of fully passive RF harvesters lies in the accurate modeling of the complete rectifier circuit so that its output performance can be predicted. Harrison et al. [1] used Ritz-Galerkin analysis to present the DC output characteristics of a single diode rectifier with some restrictions. The model neglects the diode junction capacitance and therefore the output DC voltage dependence on the operating frequency. A closed-form equation was presented in [2] that approximates a diode rectifier efficiency and input impedance. In [3,4], models that predict the conversion efficiency of an N-stage full-wave rectifier was presented. However, the DC load current must be determined or maintained to satisfy the constraints in the model. HB simulations using Agilent ADS software can be used to determine the output DC voltage of a rectifier circuit, however it is computationally intensive and do not easily provide the global view of the circuit response [1].

In this paper, a linearized model is presented to determine the output behavior of rectifying circuits. The model assumes steady state conditions and may be applied at various input power levels and connected DC load resistance without neglecting the impedance matching effects and the dynamic influence of the diode impedance on the rectifying process.

With the appropriate design of a rectifier circuit, ambient electromagnetic power harvesters may be realized. In our previous work [5,6], wireless RF power transfer were presented in the ISM bands using intentionally generated sources. In [7,8] and [9], ambient RF power harvesters from television (TV) transmitting base stations at the frequency band of (450 to 770) MHz were presented. In [7] for example, the TV base station transmitting about 960 kW was able to power sensors requiring up to 60 µW at a distances of 4.1 km. In [8] and [9], it was shown that a single TV base station can provide multiple TV channels for harvesting. However published work using ambient TV base station signals relies on antenna sizes in the range of 1.5 A4 size or larger for harvesting. These antenna sizes corresponds to at least half of the operating wavelength at (450 to 770) MHz. A rectenna that is larger than A4 size hinder practical integration of the system into modern wireless sensors. In [10], a survey was made to prove the possibility of harvesting RF energy from ambient sources such as GSM. GSM base station transmitters may be the most available in urban environments due to the immense availability of cellular network providers.

In this paper, an ISM band 868 MHz and ambient GSM wireless RF harvester is presented. Planar antenna sizes as small as 6 cm × 8 cm ($\sim\lambda/4$ at 900 MHz) may be used with the rectifier for ubiquitous RF harvesting. The rectifier is designed to operate with dual-band frequency. Using this approach, ubiquitous RF power sources in the range of −27 dBm to −50 dBm and from different frequencies can be utilized for continuous operation of micro-watt sensors.

This paper is structured as follows. Section 2 presents the rectifying circuits and the linearized analytical model. Section 3 presents the verification of the model and its comparison to ADS software HB simulations and measured results. Section 4 presents the design consideration and trade-off when realizing RF power harvesters. The design considerations are based on the presented model. Section 5 presents harvesting RF power from intentionally generated sources at indoor conditions and from ubiquitous sources using a GSM base station at outdoor conditions. Conclusions are given in section 6.

2. Analysis of rectifying circuits

2.1. Efficiency versus voltage sensitivity

Efficiency η is defined as in (1), where P_L is the output DC power and P_A is the input RF power. The voltage sensitivity γ is given by (2), where V_L is the output DC voltage.

$$\eta = \eta_{ant} \times \eta_{\Gamma} \times \eta_{rect} = \frac{P_L}{P_A} \tag{1}$$

$$\gamma = \frac{V_L}{P_A} \tag{2}$$

The current state of the art of low power sensors would require a DC voltage supply of about 1 V and DC current of 10 µA. Therefore, it is necessary to specify the output voltage and current separately rather than the product, the DC output power, when considering RF energy harvesters for remote sensors [4,11]. The delivered input RF power P_A from the receiving antenna must be transformed to that minimum voltage required to energize the sensor. Although the efficiency is widely presented in literature, its practical use in ambient RF energy harvesting can be limited. This is because an RF harvester can deliver an output DC power more than what is required by a remote sensor but may be unable to energize the sensor due to its specific voltage and input current demands.

Therefore, voltage sensitivity of a passive RF harvester may be more important than the output efficiency when powering DC loads from micro-watt RF power levels, as such more emphasis is placed on the γ rather than η in this paper.

2.2. RF diode equivalent circuit model

As shown in Figure 1, an impedance transformation network is a sub-system of a wireless RF harvester. The impedance transformation network filters higher order impedances of the diode at the operating frequency. For this reason, only the diodes fundamental impedance at the frequency of operation may be required for its accurate modeling in RF energy harvesting applications. A packaged Schottky diode may be modeled as a nonlinear junction resistance R_j shunted by a non-linear junction capacitance C_j. V_j is the voltage loss at the rectifying junction. R_S, L_S and C_P are the bulk resistance, packaging inductance and capacitance respectively [12]. A bare diode may also be modeled as a voltage controlled junction current source in parallel with a controlled capacitance (see Figure 2) and a parasitic bulk resistance [13]. I_{DC} is the voltage controlled current source from the rectifying junction. Without neglecting the bulk resistance and the package parasitics, the controlled current source is treated as a DC source in parallel with a conductance G and a susceptance B. R_D and X_D are the parallel resistance and reactance of the diode respectively. I_{DC}, R_D and X_D as in the linearized model are variables which depend on the diode parameters, input RF power P_A, the connected DC load and the reflection coefficient at the input of the rectifier.

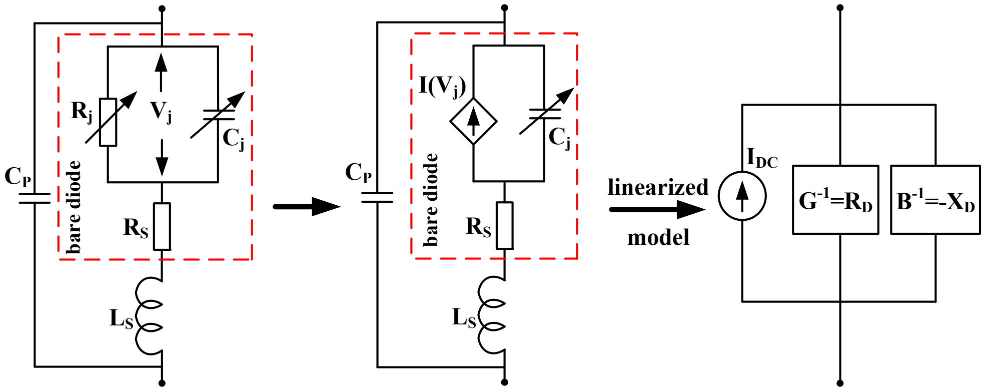

Figure 2. Schottky diode equivalent circuit models.

From Figure 2 (Left), the admittance of a Schottky diode may be written as in (3), where Y_{diode} is the diode admittance.

$$Y_{diode} = G + jB = \left(\frac{-jR_j \dfrac{1}{\omega C_j}}{R_j - \dfrac{j}{\omega C_j}} + R_S + j\omega L_S \right)^{-1} + j\omega C_P \tag{3}$$

Equation (3) can be expanded into the conductance G and susceptance B. The expression for C_j [13] is given by (4). Where Q is the charge stored at the rectifying junction.

$$C_j = \frac{\partial Q}{\partial V_j} \tag{4}$$

R_j can be found from Richardson's equation [14], which describes the current voltage relationship at the rectifying junction of a Schottky diode. The current voltage relationship at the rectifying junction is given by (5).

$$I = I_S\left[exp\left(\frac{qV_j}{\eta_0 KT}\right) - 1\right] \tag{5}$$

I_S is the diode reverse saturation current, $\alpha = \frac{KT}{q}$. α is the thermal voltage. η_0 is the diode ideality factor. q is the charge of an electron. T is the temperature and K is the Boltzmann constant. Since I is an infinite series with higher order values, there are higher order values of the diode junction resistance R_j.

$$R_j = \left(\frac{\partial I}{\partial V_j}\right)^{-1} = \frac{\eta_0 \alpha}{I_S + I} \tag{6}$$

I varies with changes in the input RF power level P_A, the losses in the impedance transformation network, rectifier impedance or connected DC load. It is easily deduced from (6) that the magnitude of I also affects the junction resistance R_j and consequently the overall impedance of the diode. Due to this, the input impedance of a diode must always be measured at a specified RF power level and DC load conditions.

The effect of a Schottky diode equivalent circuit parameters on the parallel resistance $(R_D = 1/G)$ is presented. This allow for direct comparison of each diode parasitic parameter on the R_D. The diode parasitic parameters which were used in calculating the diode impedance are shown in Table 1. V_{BV} is the reverse breakdown voltage of the Schottky diode.

Table 1. The Spice parameters of HSMS-286 Schottky diode.

I_S (nA)	I (nA)	C_j (pF)	R_S (Ω)	L_S (nH)	C_P (pF)	V_{BV} (V)
50	2	0.18	6	2	0.08	7

Figure 3(Top-left) shows the effect of the saturation current I_S on the diode rectifying junction voltage losses V_j, for various values of I. At I in the range of 5 µA, the V_j across both the I_S = 0.5 µA and the I_S = 50 nA approaches zero. When the I passing through the diode rectifying junction is greater than about 0.05 mA, the efficiency of rectification is skewed to the one with I_S = 0.5 µA, since the peak V_j of the diode with I_S = 50 nA is much higher at this condition. The results show that the smaller the I that goes through a Schottky diode rectifying junction, the smaller the V_j across it. Therefore the η_{rect} of a Schottky diode increases with high input voltage instead of high input current. Hence maximum voltage transformation at the input of the diode rectifier is always necessary for its efficient operation. The voltage transformation of the input RF power will reduce the amount of current passing through the diode rectifying junction.

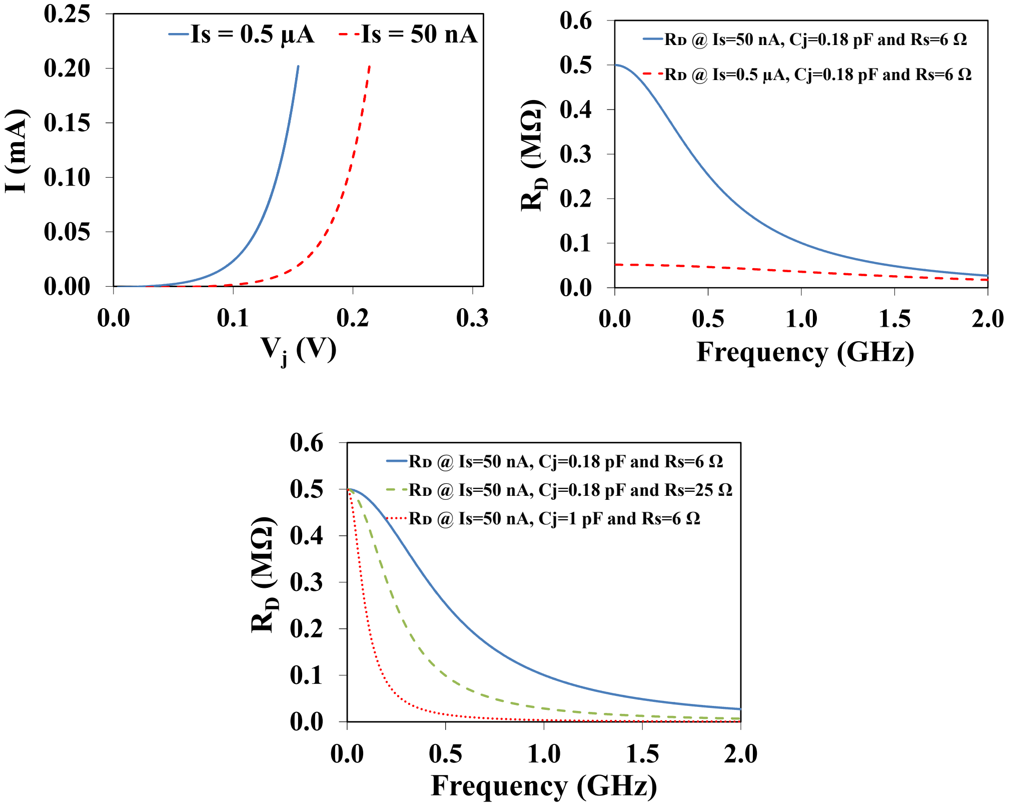

Figure 3. The effect of the changes in the equivalent circuit parameters of a Schottky diode on the parallel resistive impedance, R_D. (Top-left) Effect of the saturation current I_S on the V_j at 25 °C for different currents I passing through the rectifying junction. (Top-right) Schottky diode parallel resistive impedance R_D as a function of different I_S. (Bottom) Schottky diode parallel resistive impedance R_D as a function of different bulk resistances R_S, and junction capacitances C_j.

Figure 3(Top-right) and Figure 3(Bottom) show the calculated effect of I_S, R_S and C_j on the parallel resistive impedance R_D of a diode. From Figure 3(Top-right), as the I_S is increased from 50 nA to 0.5 µA, the R_D reduces. From Figure 3(Bottom), as the junction capacitance C_j is increased from 0.18 pF to 1 pF, the R_D of the diode reduces. By increasing R_S from 6 Ω to 25 Ω, the R_D reduces as well. The results shows that a Schottky diode with a low I_S, R_S and C_j will have high R_D. The results also show that at ultra-high frequency (UHF), the R_D of a diode is in the range of a few kilo-ohms.

2.3. RF power harvesting circuits

Figure 4 shows a fabricated RF harvester realized at 930 MHz on an FR4 printed circuit board using off-the-shelf discrete components. V_S is the input RF voltage. R_A is the RF source resistance. R_L is the DC load resistor. The impedance transformation network is realized with an inductive coil resonator and a chip capacitor. The unloaded quality factor Q_u of an impedance transformation element is used as the performance parameter for selection a component. Q_u is given by Equation (7), where X_{Coil} and R_{Coil} are the net series reactive and resistive impedance of the inductive transformation element respectively. As can be seen in Figure 3, the R_D of a Schottky diode at about 900 MHz is in the range of a few kilo-ohms. The voltage transformation that can be achieved at the input of a Schottky diode is directly proportional to the $\sqrt{R_D}$ at resonance [5] as in (8). Where V_D is the voltage at the input of the diode rectifier. When the RF voltage amplification at the input of the diode is limited due to a lower value of R_D, the rectifier efficiency is skewed toward diodes with saturation current I_S in the micro-amp range (see Figure 3(Top-left)), since more current I passes through the rectifying junction. Hence the Avago Techologies HSMS-285 series of diodes with I_S = 3 µA were used at 930 MHz to realize the RF harvester.

$$Q_u = \frac{X_{Coil}}{R_{Coil}} \tag{7}$$

$$\frac{V_D}{V_S} \propto \sqrt{\frac{R_D}{R_A}} \tag{8}$$

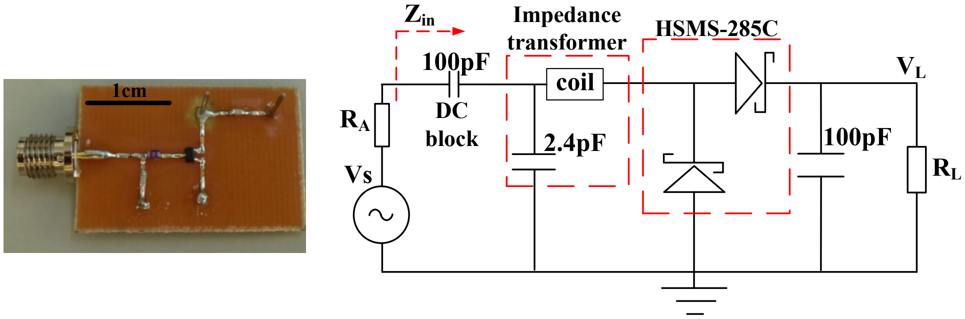

Figure 4. PCB of the realized RF harvester at 930 MHz using a HSMS-285C diode voltage doubler. The HSMS-285C has the following spice parameters I_S = 3 µA, C_j = 0.18 pF and R_S = 25. *Coil* = 38.5 nH at 900 MHz with a Q_u of 69. The chip capacitors have Q_u of about 1000 at 900 MHz.

2.4. Linearized analysis of the RF harvester

The linearized impedance diagram of the RF rectifier shown in Figure 4 is presented in Figure 5. I_{DC} is the same for each diode since two identical diodes are assumed. I_{DC} may be found from (10). P_{in} is the power absorbed at Z_{in} plane as in (9). Where Γ is the reflection coefficient at the Z_{in} plane (see Figure 4). N is the number of doubler stages.

$$P_{in} = P_A\left(1-|\Gamma|^2\right) \tag{9}$$

$$I_{DC} = \frac{P_{in}}{NV_D} \tag{10}$$

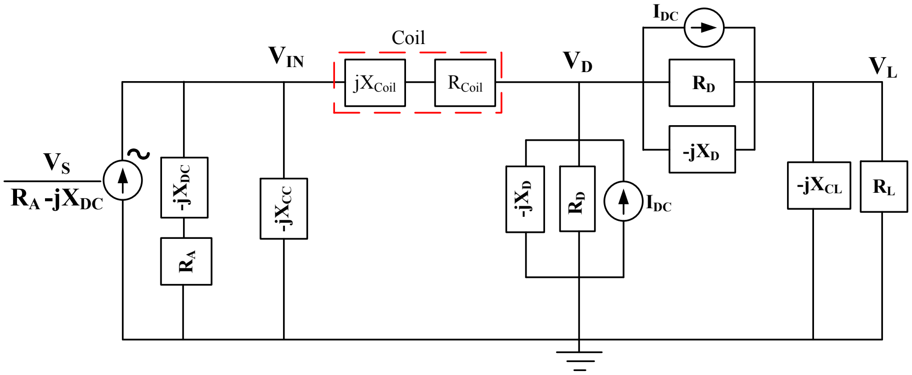

Figure 5. Linearized model of the RF power harvester at 930 MHz. X_{DC} is the reactance of DC block capacitor. X_{CC} is the reactance of 2.4 pF capacitor. R_{Coil} and X_{Coil} are the series resistive and reactive impedance of the coil respectively. X_{CL} is the reactance of the 100 pF capacitor. R_L is the load resistor.

After the linearized model is setup as in Figure 5, the steady state output voltage V_L is found using nodal analysis. From Figure 5, (11), (12) and (13) may be used to determine the voltages at node V_{IN}, V_D and V_L.

$$\frac{V_S}{R_A - jX_{DC}} = V_{IN}\left[\frac{1}{R_A - jX_{DC}} + \frac{1}{-jX_{CC}} + \frac{1}{R_{Coil} + jX_{Coil}}\right] - V_D\left[\frac{1}{R_{Coil} + jX_{Coil}}\right] \tag{11}$$

$$0 = -V_{IN}\left[\frac{1}{R_{Coil} + jX_{Coil}}\right] + V_D\left[\frac{1}{R_{Coil} + jX_{Coil}} + \frac{2}{-jX_D} + \frac{2}{R_D}\right] - V_L\left[\frac{1}{-jX_D} + \frac{1}{R_D}\right] \tag{12}$$

$$I_{DC} = -V_D\left[\frac{1}{R_D} + \frac{1}{-jX_D}\right] + V_L\left[\frac{1}{-jX_{CL}} + \frac{1}{-jX_D} + \frac{1}{R_D} + \frac{1}{R_L}\right] \tag{13}$$

From (11) to (13), the minima of V_L occurs when I_{DC} in the doubler diodes limits to zero. $I_{DC} \approx 0$ implies the diodes are not rectifying at the input RF power level. Hence the output voltage V_L approaches (14). The alternating current (AC) output voltage when $I_{DC} = 0$ is given by V_{L_AC}.

$$V_L \rightarrow V_{L\_AC} \tag{14}$$

The maxima of the output voltage V_L approaches (15). For N number of doublers, the upper limit of the V_L is increased by a factor N.

$$V_L \rightarrow 2NV_D \tag{15}$$

2.5. Linearized analysis applied to multipliers and multiple input matching circuits

The RF harvester presented for dual-band applications is realized with a coupled matching to provide the multiple resonant frequencies. The presented dual-band harvester which can harvest from

GSM-900 or LTE-800 and GSM-2160 is shown in Figure 6 using a modified voltage quadrupler. The capacitors at the input of the diode chain are connected in parallel, however the capacitors are connected in series at the output of the diode chain. To achieve the additional bands of frequency operation, each input branch into a diode voltage doubler is tuned to a specific frequency; ω_1 (at 935 MHz) and ω_1 (at 2.2 GHz). The resonant frequencies can be approximated by (16). The employed diodes are the HSMS-285C Schottky diodes.

$$\omega_1 \approx \frac{1}{\sqrt{C_{C1}Coil_1}} \text{ and } \omega_2 \approx \frac{1}{\sqrt{C_{C2}Coil_2}} \tag{16}$$

By replacing the diodes in Figure 6 with its linearized model (see Figure 5), its output DC characteristics can be analyzed with the same approach presented in Section 2.4.

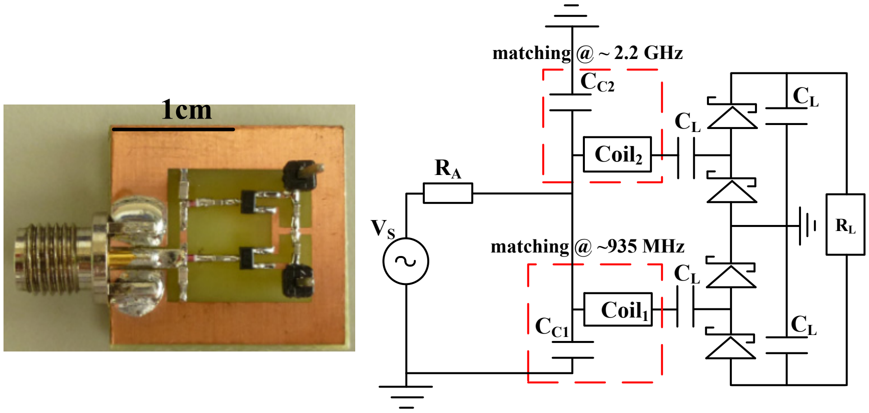

Figure 6. Picture and circuit layout of the dual-band RF harvester. The harvester is matched at 935 MHz and 2.2 GHz. Schottky diodes are HSMS-285x series. C_{C1} = 2.7 pF, $Coil_1$ = 39 nH; $Coil_1$ Q_U @ 900 MHz = 88, C_{C2} =0.8 pF, $Coil_2$ = 2.14 nH; $Coil_2$ Q_U @ 1.7 GHz = 35, C_L =100 pF.

3. Results of RF to DC power transmission

The presented linearized model, ADS HB simulations and measured voltage sensitivity for the harvesters presented in Figure 4 and Figure 6 are shown in Figure 7. For the harvester presented in Figure 4, a DC voltage of 0.4 V across R_L of 0.5 MΩ is measured at −20 dBm input RF power. At −14 dBm input RF power, the harvester as shown in Figure 4 delivers a measured DC voltage of 1 V across a 0.5 MΩ R_L. The harvester achieves a peak efficiency of 24% and 41% at −20 dBm and −10 dBm input RF power respectively for about 6 kΩ R_L when operating at 930 MHz.

For the harvester shown in Figure 6, a DC voltage of 0.43 V across a 0.5 MΩ R_L is measured at −20 dBm input RF power at 935 MHz. At −14.4 dBm input RF power, the harvester as shown in Figure 6 delivers a measured DC voltage of 1 V across a 0.5 MΩ R_L. At 5 dBm input RF power, a DC voltage of 6 V is measured across a 0.5 MΩ R_L at 935 MHz. The harvester achieves a peak efficiency of 16% and 35% at −20 dBm and −10 dBm input RF power respectively for about 10 kΩ R_L when operating at 935 MHz. At 2.2 GHz, an output DC voltage of 0.4 V is measured across a 0.5 MΩ R_L at −10 dBm input RF power. At 2.2 GHz, a peak efficiency of 4% is measured at

−10 dBm input RF power for about 10 kΩ R_L.

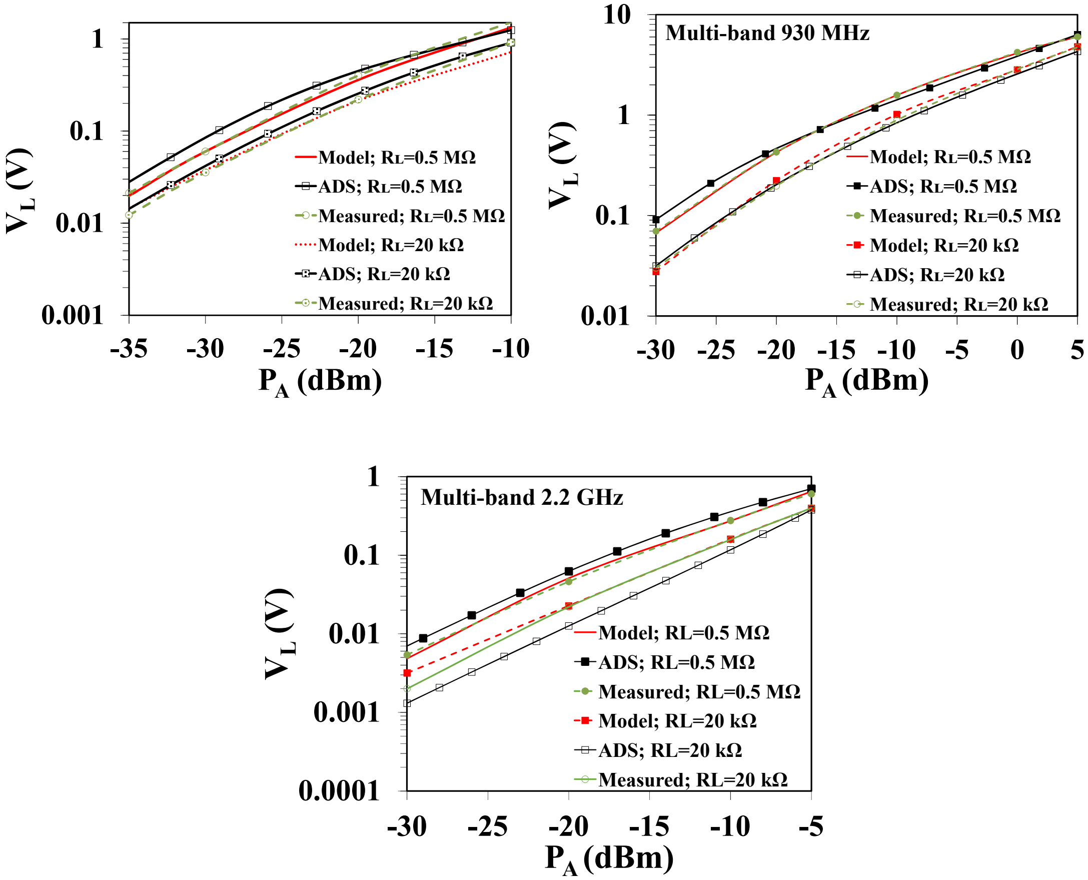

Figure 7. Presented linearized model as compared to Agilent ADS HB simulations and measurements. V_L is the output DC voltage and P_A is the input RF power. (Top-left) Results of the RF harvester as shown in Figure 4. (Top-right) Results of the dual-band RF harvester as shown in Figure 6 at 930 MHz. (Bottom) Results of the dual-band RF harvester as shown in Figure 6 at ~2.2 GHz.

From the results shown in Figure 7, the maximum calculated deviation between measurements and the presented linearized model is 15%. The 15% maximum difference is as a result of calculating the diode R_D and I_{DC}. I_{DC} is dependent on P_A, Γ at Z_{in}, R_D and R_L, as a result any variation such as component tolerances in these parameters affects the estimation of the output DC voltage. R_D is also a function of the R_L and P_A.

4. Circuit Design Considerations

4.1. *Circuit performance parameters*

The maximum DC load power from the rectifier may be approximated by (17). Where P_{Lmax} is the maximum power delivered to the R_L. The DC load resistance that results in the maximum efficiency from the RF harvester is given by (18). R_{Lmax} is the DC load resistance R_L that results in the maximum η. For Figure 4, R_{Lmax} is ~6 kΩ. For Figure 6, R_{Lmax} is ~10 kΩ. From (8) and (17), the diode R_D and peak voltage into a diode rectifier V_D, must be high as possible for high voltage sensitivity of the RF harvester.

$$P_{L\,max} \approx \frac{(2NV_D)^2}{2R_L} \tag{17}$$

$$R_{L\,max} \approx 2NR_D \tag{18}$$

From the linearized transfer functions that describe the performance of an RF harvester, a maxima of P_L (P_{Lmax}), R_L (R_{Lmax}), η and γ may be found for any RF harvester topology.

4.2. Rectifying diode selection

The harvester shown in Figure 6 is realized to cover the LTE-800 MHz or GSM 900 MHz and GSM 2110 MHz frequency bands (see Figure 8). Hence the realized harvester loaded quality factor ($\propto \sqrt{R_D/R_A}$) is low at 935 MHz and 2100 MHz.

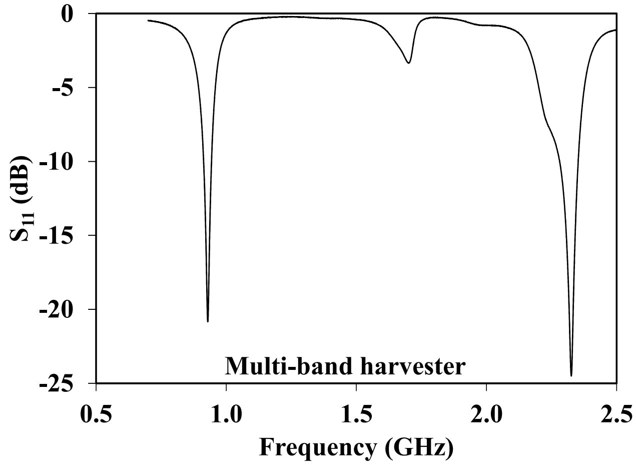

Figure 8. Measured open circuit S_{11} of the dual-band RF harvester. The input RF power level is −15 dBm. The 3-dB return loss frequency bandwidth is 120 MHz around 930 MHz and 200 MHz around 2.2 GHz.

For RF harvesting applications which require such broad-band operations, diodes with low R_D should be considered. Diodes with lower C_j and R_S are always preferable in RF energy harvesting applications. R_S consumes ohmic power, hence the lower the better. C_j in the range of 1 pF reduces the diodes R_D as well as its switching speed. The action of C_j dominates at sufficiently higher diode operating frequencies resulting in lower attainable R_D. Since diodes are more sensitive to high voltage at input power levels less than −10 dBm, a low R_D results in limited voltage transformation. As a result more current goes through the diode and the diodes voltage losses V_j tend to be maximum.

Hence at low diode input R_D, the efficiency of rectifying RF signals is skewed towards the diodes with I_S in the micro-amp range (such as HSMS-285x series), since they have lower peak V_j.

Impedance transformation components with high unloaded quality factor Q_u are always preferable in RF energy harvesting applications. The question always arise whether to use discrete components or transmission lines on a PCB. The unloaded quality factor of a transmission line may be described by (19) [15]. Where F_0 is the transmission line frequency in mega-hetz, A_0 is the attenuation in dB/100 feet at F_0 and VF is the velocity factor.

$$Q_u = \frac{2.7743\, F_0}{A_0\, VF} \tag{19}$$

As an example, an inductive or capacitive transmission line on an FR4 PCB has VF of 0.48 and A_0 of 132 dB/100 feet at 1 GHz. Using (19) and the values of an FR4 transmission line at 1 GHz gives a Q_u of 44. Hence up to 1 GHz, discrete components may be preferred to transmission lines for impedance transformation since chip inductors and capacitors can provide Q_u of up to 80 and 1000 at 1 GHz respectively. However, since inductive transmission lines quality factor is directly proportional to the operating frequency, it may be preferred to discrete inductors at frequencies above 3 GHz.

For RF harvesting applications with input power levels greater than −10 dBm, diodes may be cascaded as in multipliers. This is due to the possible constant output voltage or power of a single rectifier stage after a certain threshold input RF power. The appropriate number of diode doubler stages that strikes a good balance between the losses V_j across the diodes in the multiplier and the power saturation of the rectifier stage must be made. Equation (20) may be used as a guide in choosing the number of rectifier voltage doubler stages N and the peak voltage or power handling capabilities of the rectifier.

$$RF\ power \begin{Bmatrix} V_D << V_j ; N = 1 \\ V_j << V_D < V_{BV} ; N \\ V_D \approx V_{BV} ; N + 1 \end{Bmatrix} \tag{20}$$

The maximum DC voltage that can be outputted by a diode is determined by the reverse breakdown voltage V_{BV}. The maximum output DC voltage V_L of a single diode rectifier is $\sim V_{BV}/2$ [2]. Hence the maximum output DC voltage of a diode voltage doubler is $\sim V_{BV}$ and for a multiplier, it is $\sim NV_{BV}$.

5. Wireless RF energy harvesting

5.1. Wireless RF energy transfer in the ISM band 868 MHz.

At 868 MHz, the commercially available Huber+Suhner SPA 860 (1308.17.0058) planar antenna [16] was used as a transmitter. The Huber+Suhner SPA 860 has a gain G_{it} of 8 dBi at 868 MHz and measures 20 cm × 20 cm in size. The antenna presented by Homg-Dean Chen et al. [17] was used to receive the wireless signals at 868 MHz. The antenna by Homg-Dean Chen et al. (see Figure 11) is a dual-band antenna with a 10-dB return loss bandwidth of 90 MHz around 900 MHz and 218 MHz around 1800 MHz. It has a gain G_{ir} of 1 dBi at around 900 MHz and 3 dBi around 1800 MHz.

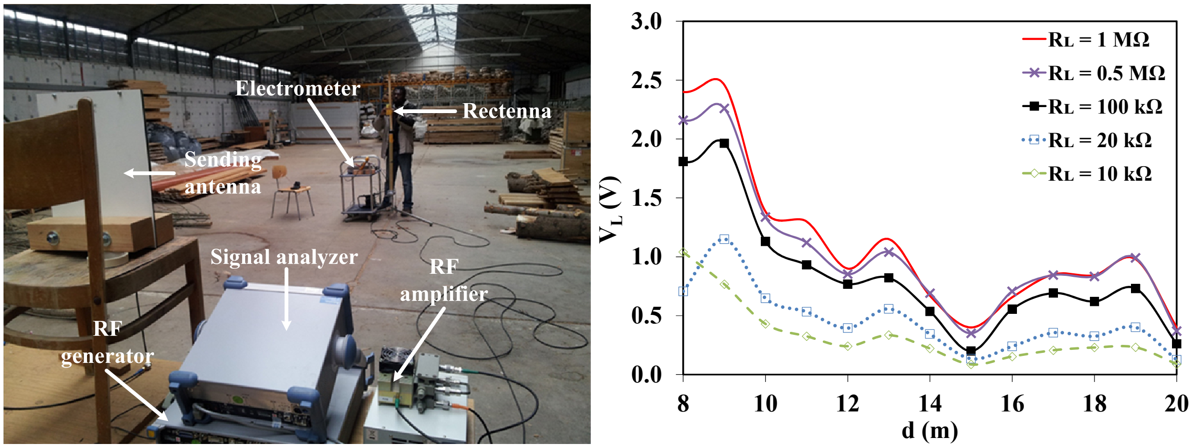

Figure 9. (Left) Picture of the experimental setup for the wireless range measurements at 868 MHz. (Right) Rectenna (antenna + rectifier) receiving range performance by sending P_t of 1 W at 868 MHz. Transmitting antenna size is 20 cm × 20 cm with G_{it} of 8 dBi. Receiving rectenna size is 6 cm × 8 cm. *d* is the distance between transmitter and rectenna.

The size of the antenna is 6 cm × 8 cm, $\lambda/4$ at 868 MHz. The dual-band rectifier as presented in Figure 6 is used with the antenna for harvesting at 868 MHz. The transmitted RF power P_t was 1 W at 868 MHz. The experimental setup and the wireless range measurements were performed in a warehouse as can be seen in Figure 9 (Left). The sending and the receiving antennas were arranged ~1.5 m above the ground. The rectenna delivered DC voltage was measured with the Keithley 6514 electrometer.

Figure 9(Right) shows the range measurements at 868 MHz. A harvested DC voltage of 0.4 V is measured across a 0.5 MΩ R_L at the distance of 20 m. At 13 m, 1 V DC voltage is measured across a 0.5 MΩ R_L. The far-field results at 868 MHz show that the received power generally degrades with $1/d^2$ as in the Friis equation (21) [18]. Where P_r and P_t are the transmitted and received power respectively. λ is the wavelength of the RF signal. ξ as in (21) is an *additional* parameter in the Friis equation which account for reflections and multi-path propagations of the wireless signals. It can be seen from Figure 9 (Right) that the measured output DC power is modulated by the constructive and destructive interferences along $1/d^2$, since there are farther distances where the delivered power is greater than a nearer range. This is evident at a distance *d* of 15 m, where a DC output voltage of 0.5 V is harvested as compared to 1 V harvested at *d* of 19 m for 0.5 MΩ R_L. At 15 m there is destructive interference from to the reflected signals to the direct signals, hence the harvested power is lower than that predicted by the original Friis formula, however, at 19 m, there is constructive interference and the harvested power is higher than that predicted by the original Friis formula.

$$P_r = \xi P_t G_{ir} G_{it} \left(\frac{\lambda}{4\pi d} \right)^2 \tag{21}$$

5.2. Harvesting from ubiquitous GSM signals

To investigate the possible use of ambient RF power sources for operation of remote micro-watt sensors. The RF power levels harvested by a ~$\lambda/4$ receiving antenna at 930 MHz was first measured at a distance of about 110 m from a GSM base station. The receiving antenna that was used for this experiment is the dual-band antenna by Homg-Dean Chen et al. [17].

Due to reflections, multi-path propagation and the fact that the transmitting antennas from the GSM base station point at different directions (see Figure 11), the power delivered by the dual-band receiving antenna was not estimated by the Friis formula [18] as in Equation 21. The power delivered by the receiving antenna were made by on-site measurements using a Rohde & Schwarz PR100 RF monitor. From the balcony of a building, about a distance of 110 m from a GSM transceiver base station, the antenna was positioned to harvest the ambient RF power. Figure 10 shows the average (over 5 minutes) harvested RF power levels of the dual-band receiving antenna at a distance of about 110 m from the GSM base station.

The GSM base station is located in the city of Freiburg, Germany at 48.0087,7.829022 coordinates. Table 2 shows the transmitting parameters of the GSM base station.

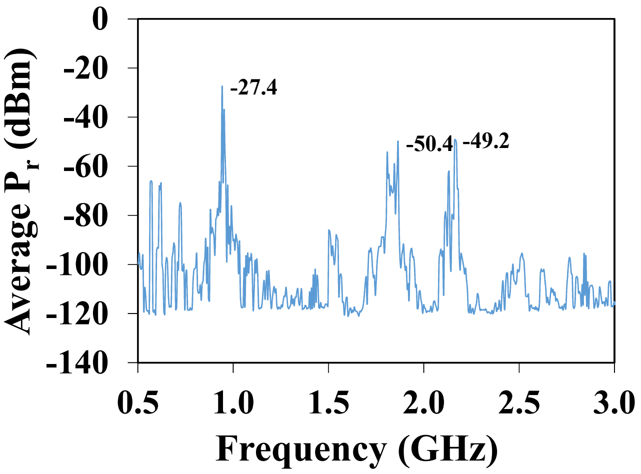

Figure 10. Average (over 5 minutes) ambient RF power delivered by the dual-band antenna at a distance of about 110 m from a GSM base station. Receiving antenna size is 6 cm × 8 cm (~$\lambda/4$ at 930 MHz) with a gain of 1 dBi at 930 MHz and 3 dBi at 1800 MHz.

Table 2. Operating parameters of a GSM base station in Freiburg, Germany.

Transmitter	1	2	3	4	5	6
Frequency (MHz)	935	935	2160	2160	2160	811
Transmitting power (W)	102	102	32	32	32	30
Number of channels	1	1	2	2	2	2
Antenna gain (dBi)	15.7	15.7	18.0	18.0	18.0	15.8

Figure 10 shows that the harvested power levels by the dual-band receiving antenna, from the various frequency bands of a GSM base station ranges from −27 dBm to −50 dBm at a distance of about 110 m. Figure 11 shows the output DC voltage from the dual-band rectifier and the dual-band [17] receiving antenna. Figure 11 shows a harvested output DC voltage of about 2.3 V across a ~0.5 MΩ resistor is achieved from the several GSM power sources at a distance of 110 m.

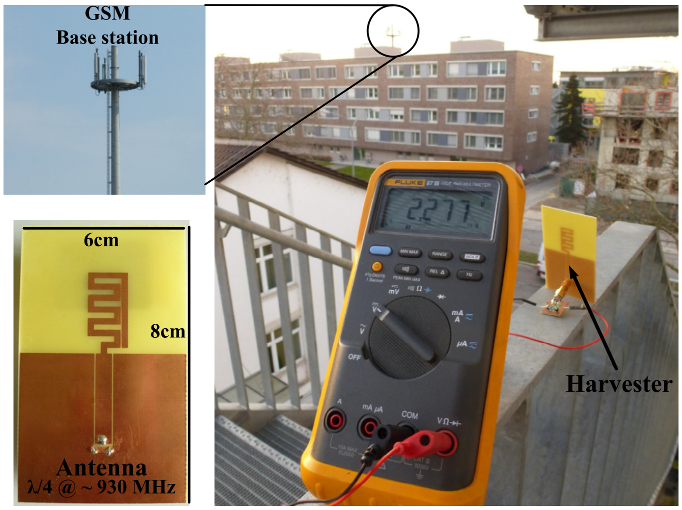

Figure 11. Setup for measuring the power delivered by the GSM base station across a resistive load. The RF to DC rectifier is the circuit as presented in Figure 6 and the antenna is from Homg-Dean Chen et al. [17]. R_L is ~0.5 MΩ. V_L is about 2.3 V.

The delivered DC power by the rectenna is the sum of all the harvested RF signals in the multiple transmitting frequencies (811 MHz, 935 MHz and 2160 MHz) of the base station. The harvested DC power can be approximate by (22) [8]. Where C_1 up to C_N are the transmitting channels. For GSM signals, a channel covers 200 kHz. f_1 and f_2 are the lower and upper band of the RF frequency respectively. f is the frequency. Using this approach, ambient RF power sources in the range of −27 dBm to −50 dBm and from different frequencies can be utilized to power sensors requiring ~1.3 V DC across a 0.5 MΩ and still maintain a ~λ/4 rectenna size.

$$P_{DC} \propto \sum_{C_1}^{C_N} \int_{f_1}^{f_2} P_t \delta f \tag{22}$$

The RF harvester was also loaded with a commercially available TFA 30.5014.02 Thermo-Hygrometer. The Thermo-Hygrometer is normally powered by an LR44 battery requiring 1.3 V to 1.5 V at 3 µA for its continuous operation. It can be seen that the RF harvester was also able to power the sensor through startup-mode and steady state operations (see Figure 12). Since the Thermo-Hygrometer has an input resistance of ~0.5 MΩ and consumes ~1.3 V, the theoretical distance at which the Thermo-Hygrometer may still be powered by the presented RF harvester is about 220 m using the GSM base station.

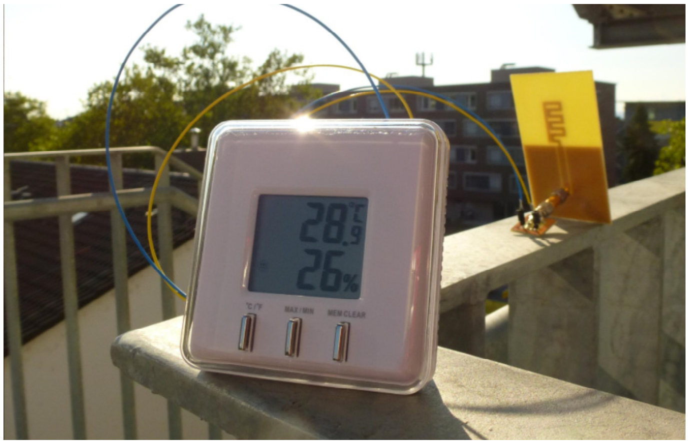

Figure 12. Using ambient RF GSM power to operate Thermo-Hygrometer sensor. The RF to DC rectifier is as presented in Figure 6.

6. Conclusions

By using a linearized model of a rectifier, a complete description of an RF harvester can be made without compromise. This analysis is confirmed by measured results and HB computer aided simulations for different RF harvester circuit topologies. It is shown that an RF harvester can be optimized using the equations from a linearized RF harvester model. This provides a unique advantage over computer aided simulations which gives little insight into the design trade-off and the effect of the component parameters on the performance of the harvester. The theory presented in this paper could be used for broad design considerations before a specific RF harvester circuit topology is adapted. The theory has been proved by using low cost off-the-shelf components. However the model is equally applicable to custom made designs based on complementary metal-oxide-semiconductor (CMOS) process. It has been shown that using a dual-band RF harvester, ambient RF power sources in the range of −27 dBm to −50 dBm from different frequency bands can be used to provide perpetual energy for operation of wireless remote sensors requiring 1.3 V at 0.5 MΩ. This has been achieved by using rectenna size of ~$\lambda/4$ at 930 MHz. In the ISM 868 MHz frequency band, an operation range of 19 m is demonstrated for DC loads requiring 1 V at 0.5 MΩ from a single transmitted signal at indoor conditions.

Acknowledgments

This work is part of the graduate program GRK 1322 Micro Energy Harvesting at IMTEK, University of Freiburg, funded by the German Research Foundation (DFG).

Conflict of Interest

There is no conflict of interest.

References

1. Harrison RG, Le Polozec X (1994) Nonsquarelaw behavior of diode detectors analyzed by the Ritz-Galerkin method. *Microw Theory Tech IEEE Trans* 42: 840–846.
2. McSpadden JO, Fan I, Chang K (1998) Design and experiments of a high-conversion-efficiency 5.8-GHz rectenna. *Microw Theory Tech IEEE Trans* 46: 2053–2060.
3. Curty J-P, Joehl N, Krummenacher F, et al. (2005) A model for μ-power rectifier analysis and design. *Circuits Syst Regul Pap IEEE Trans* 52: 2771–2779.
4. Barnett RE, Liu J, Lazar S (2009) A RF to DC voltage conversion model for multi-stage rectifiers in UHF RFID transponders. *Solid-State Circuits IEEE* 44: 354–370.
5. Nimo A, Grgić D, Reindl LM (2012) Optimization of passive low power wireless electromagnetic energy harvesters. *Sensors* 12: 13636–13663.
6. Nimo A, Grgić D, Reindl LM (2012) Ambient electromagnetic wireless energy harvesting using multiband planar antenna. *Systems Signals Devices (SSD), 9th International Multi-Conference on* 1–6.
7. Sample A, Smith JR (2009) Experimental results with two wireless power transfer systems. *Proceedings of the 4th international conference on Radio and wireless symposium*, San Diego, CA, USA, 16–18.
8. Mikeka C, Arai H (2011) Design issues in radio frequency energy harvesting system. *Sust Energ Harvesting Technologies—Past Present and Future*, InTech.
9. Vyas RJ, Cook BB, Kawahara Y, et al. (2013) E-WEHP: A batteryless embedded sensor-platform wirelessly powered from ambient digital-TV signals. *Microw Theory Tech IEEE Trans* 61: 2491–2505.
10. Pinuela M, Mitcheson PD, Lucyszyn S (2013) Ambient RF energy harvesting in urban and semi-urban environments. *Microw Theory Tech IEEE Trans* 61: 2715–2726.
11. Joe J, Chia M, Marath A, et al. (1997) Zero bias schottky diode model for low power, moderate current rectenna. *DETS'97 Proc*.
12. Watson HA (1969) Schottky-barrier Devices. *Microwave semiconductor devices and their circuit applications*, New York; Maidenhead: McGraw-Hill.
13. Maas SA (2003) Harmonic-Balance Analysis and Related Methods. *Nonlinear microw RF circuits*, Boston, MA: Artech House.
14. Sah C-T (1991) P/N and Other Junction Diodes. *Fundamentals of solid-state electronics*, Singapore; River Edge, NJ: World Scientific.
15. Audet J (2006) Q Calculations of L-C Circuits and Transmission Lines: A Unified Approach. *QEX Mag* 43–51.
16. HUBER+SUHNER (2010) Rfid reader antenna: Spa 860/65/9/0/v (1308.17.0005).
17. Chen H, Chen W, Cheng Y, et al. (2003) Dualband meander monopole antenna. *Antennas Propagation Society Int Symp IEEE* 3: 48–51.
18. Friis HT (1946) A Note on a Simple Transmission Formula. *Proc IRE* 34: 254– 256.

© 2015, Antwi Nimo, et al., licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)

ELECTRONICA

para todos

TOMO 1

Analógica

SALVAT

ELECTRONICA
para todos

TOMO 1

Analógica

SALVAT

Entender la electrónica

Energía, comunicaciones, tratamiento, control: la electrónica es uno de los cimientos del mundo contemporáneo. Vamos a explorar este fascinante universo con muchas explicaciones y pocas fórmulas, con palabras sencillas (pero correctas) para los que empiezan y algo de profundización para quien desee aumentar sus conocimientos.

Esta primera sección le introduce gradualmente **a la electrónica analógica**, es decir, electricidad y electrónica en general.

Presentaremos los principales **dispositivos**, por ejemplo el transistor, y **circuitos**, como los amplificadores o los receptores de radio.

Veremos también cómo leer un **esquema** eléctrico y entender, al menos a grandes rasgos, su funcionamiento.

Esta sección sirve además como **base** para comprender mejor las demás (electrónica digital, componentes, herramientas y técnicas, aplicaciones, proyectos).

Energía

Comunicaciones y tratamiento

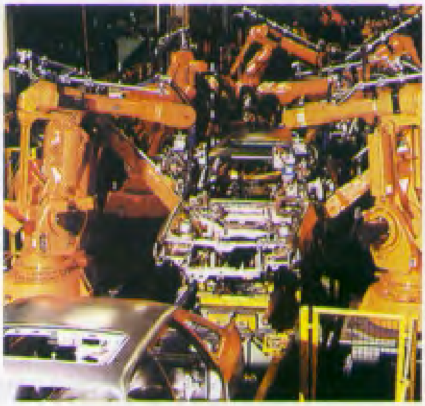

Control

LA CORRIENTE FLUYE COMO EL AGUA

- Hay ciertas **similitudes** entre el comportamiento de la corriente eléctrica y el de una corriente de agua, aunque desde un punto de vista físico son cosas muy distintas.
- Por ejemplo, la corriente fluye por los cables, igual que el agua **fluye** por las tuberías. Para hacerlas circular es necesaria una **fuerza** que las empuje.
- Ambas pueden transportar **energía**, por ejemplo para hacer girar una lavadora o la rueda de un molino.
- Esta comparación resulta útil para **ilustrar** intuitivamente algunos comportamientos de la corriente eléctrica, pero no hay que exagerarla.

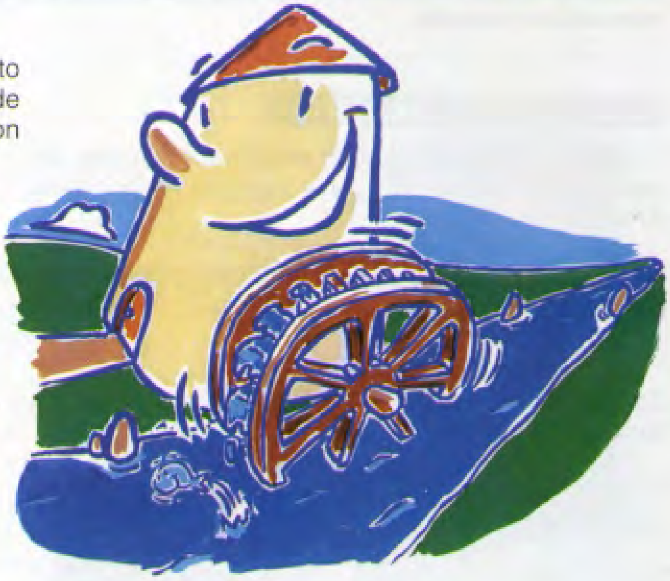

Como una corriente de agua, también la corriente eléctrica transporta energía que puede realizar un trabajo.

Circuitos eléctricos

La corriente fluye por los circuitos, transportando energía

La figura muestra un **circuito**: una pila conectada con dos cables a una bombilla. La pila proporciona energía a la bombilla.

La pila se comporta como una bomba que **empuja** a la corriente a pasar por los cables, atravesando la bombilla.

Esto ocurre sólo si el circuito está **cerrado**, es decir, no tiene interrupciones: la corriente debe poder salir de un extremo ("polo") de la pila y entrar de nuevo por el otro.

Con un circuito **abierto**, es decir, interrumpido, no pasa la corriente. Así, por ejemplo, una linterna tiene un interruptor para abrir y cerrar el circuito.

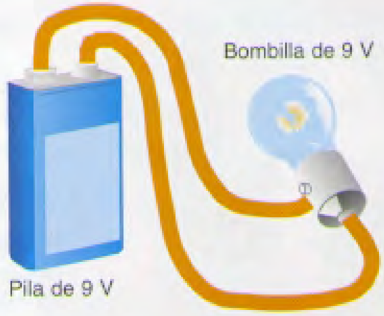

Procedente de la pila, la corriente recorre un circuito cerrado y transfiere energía a la bombilla.

EL ESQUEMA: UN CIRCUITO IDEAL

La figura muestra el **esquema eléctrico** del mismo circuito, es decir, su representación reducida a lo más esencial.

En un esquema los **componentes** (pila, bombilla, cables) están representados por símbolos convencionales, de modo que la lectura sea sencilla e inmediata.

El esquema permite así **describir** el circuito de un modo comprensible y reproducible.

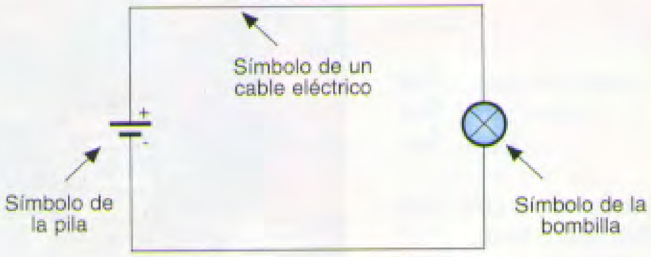

Un esquema eléctrico representa lo más básico de un circuito de un modo sencillo y esencial.

DEL POSITIVO AL NEGATIVO

El símbolo '+' indica el **polo positivo** de la pila. En este circuito la corriente fluye siempre en la misma dirección.

Por **convenio** la corriente sale del polo positivo de la pila y entra de nuevo por el negativo, aunque en realidad... ¡es al contrario!

De hecho, la corriente está formada por partículas con carga negativa, los **electrones**.

Pero el convenio está ya muy arraigado, y **en la práctica** no cambia nada.

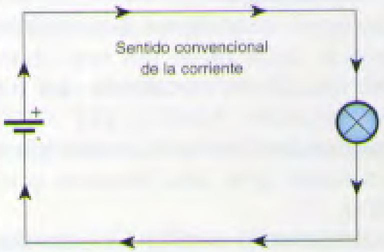

Convencionalmente, la corriente va desde el polo positivo al negativo.

Tensión y corriente

La tensión es la fuerza que empuja a la corriente a circular

Igual que el agua no se mueve sin una bomba o un **desnivel** (diferencia de altitud), tampoco la corriente circula sola.

La fuerza que la empuja se llama **tensión**, o diferencia de potencial, y se mide en **voltios** (abreviado V), que toman su nombre de Alessandro Volta, inventor de la pila eléctrica.

La instalación doméstica tiene normalmente una tensión de **220 V** (220 voltios), valor suficientemente alto como para ser peligroso.

La **baja tensión** de una pila (normalmente por debajo de los 12 V) y de muchos circuitos electrónicos suele ser, en cambio, totalmente inocua.

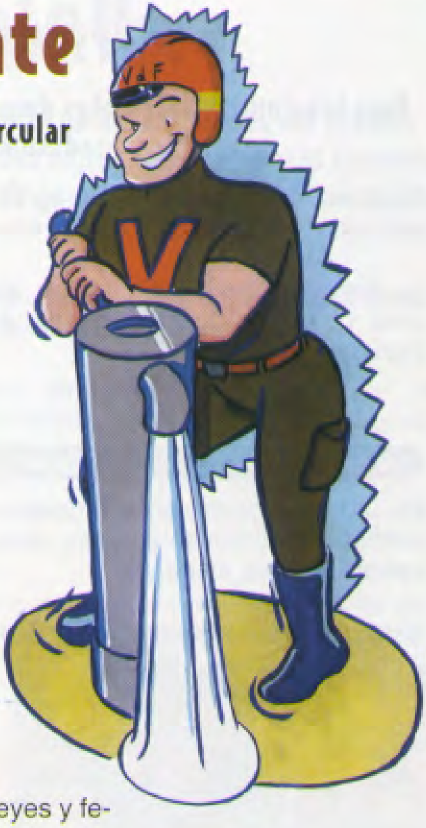

Para moverse, la corriente necesita ser empujada por la fuerza de la tensión.

¿CUÁNTA CORRIENTE PASA?

La corriente se mide en **amperios** (abreviado A), nombre tomado de André-Marie Ampère, descubridor de numerosas leyes y fenómenos eléctricos.

La placa que lleva todo electrodoméstico informa de la tensión de trabajo (normalmente 220 V) y la corriente absorbida.

Por ejemplo, una bombilla halógena requiere cerca de **1 A** (un amperio), mientras que un calentador de agua eléctrico puede utilizar incluso 10 A.

Aunque conectados a la misma **tensión** de 220 V, los dos dispositivos utilizan corrientes diferentes, según el trabajo desarrollado.

CORRIENTE CONTINUA, CORRIENTE ALTERNA

La corriente producida por una pila se mueve en un solo sentido y se llama **corriente continua**, abreviada "c.c", "dc" o "=".

La corriente de una vivienda se mueve alternativamente en los dos sentidos: es una **corriente alterna**.

Se abrevia "c.a", "ac", "~" o " ≈ ". Obviamente es alterna también la tensión que la empuja.

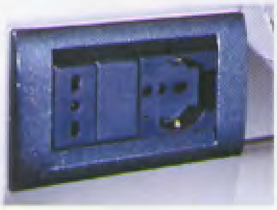

Aquí hay 220 Vca, es decir, 220 V alternos.

VARTA
ACCU
rechargeable

Esta pila tiene una tensión de 9 Vcc, es decir, 9 V continuos.

Unidades de medida

Para tensiones y corrientes demasiado grandes o demasiado pequeñas, voltios y amperios no son adecuados: se utilizan múltiplos y submúltiplos

Medir en metros el Atlántico o las patas de las hormigas no es **práctico**: para esto utilizamos kilómetros, milímetros y otros múltiplos y submúltiplos.

Lo mismo ocurre en electrónica. Afortunadamente, basta con aprender unos pocos **símbolos** que después se utilizan siempre de la misma manera.

LA TABLA DE LOS MULTIPLICADORES

En la siguiente tabla se muestran los principales **prefijos** multiplicadores, con ejemplos reales de aplicaciones en electrónica.

Hemos omitido otros prefijos de los **extremos** de la gama, como tera (10^{12}) o femto (10^{-15}), de uso bastante raro.

Los prefijos multiplicadores más utilizados en electrónica. En los ejemplos se utilizan algunas medidas (hertzios, ohmios, faradios) que veremos más adelante en el curso.

MULTIPLICADOR	SÍMBOLO	VALOR	EJEMPLOS TÍPICOS Y LECTURA
giga	G	10^9 (mil millones)	3 GHz (tres gigahertzios)
mega	M	10^6 (un millón)	1 MΩ (un megaohmio)
kilo	k	10^3 (mil)	2 kV (dos kilovoltios)
milli	m	10^{-3} (una milésima)	13 mA (trece miliamperios)
micro	µ	10^{-6} (una millonésima)	4 µsec (cuatro microsegundos)
nano	n	10^{-9} (una mil millonésima)	6 nA (seis nanoamperios)
pico	p	10^{-12} (milésima de mil millonésima)	2 pF (dos picofaradios)

Aquí puede haber 380 kVca.

FLUKE
240.0 V~
OFF
V
COM

Herramientas y datos técnicos separan los decimales con un punto.

CÓMO SE ESCRIBEN LOS VALORES

Normalmente escribimos 3 km o 140 cm, con un **espacio** entre el número y la unidad de medida, por lo que lo correcto es escribir 12 V o 25 mA.

Sin embargo, se acepta también la práctica contraria, es decir, 12V o 25mA, especialmente en los **esquemas** eléctricos (por motivos de espacio y claridad).

Para los decimales se utiliza a menudo el **punto** separador anglosajón, por ejemplo 13.24 V, en lugar de la coma (13,24 V).

Resistencia y resistores

La corriente que pasa por un circuito no depende sólo de la tensión que la empuja, sino también de la resistencia que los electrones encuentran en su movimiento.

Un conductor perfecto no pone ningún obstáculo a la corriente. Sin embargo, en la práctica, los materiales ofrecen una cierta **resistencia** al paso de los electrones.

Con la misma tensión (es decir, fuerza que la empuja), la corriente efectiva que **circula** depende de la resistencia del circuito, exactamente lo mismo que ocurre con el agua.

En este curso vamos a evitar todo tratamiento **matemático**, por otra parte no esencial para una eficaz compresión; sólo le pedimos al lector que soporte un par de sencillas fórmulas fundamentales.

Aumentando la resistencia, la corriente se reduce.

LA LEY DE OHM

La relación aparentemente trivial descubierta por Georg Simon Ohm en 1.827 es la **base** de toda la electrónica:

I = V / R

es decir: la corriente (I) se obtiene **dividiendo** la tensión (V) por la resistencia (R).

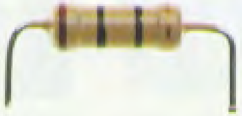

Todo resistor tiene un cierto valor de resistencia, expresado en ohmios y escrito normalmente con un código de colores.

La resistencia se mide en **ohmios** (símbolo Ω) con sus correspondientes múltiplos y submúltiplos: en particular, el kiloohmio (kΩ, 1.000 ohmios) y el megaohmio (MΩ, un millón de ohmios).

Se pueden adquirir, y se utilizan con mucha frecuencia, **resistores** (comúnmente llamados "resistencias"): dispositivos construidos especialmente para ofrecer una resistencia muy precisa.

LAS RESISTENCIAS EN LOS CIRCUITOS

En los circuitos eléctricos, las resistencias puede estar representadas con dos **símbolos** diferentes: el "europeo" y el "americano".

El primero, que utilizaremos principalmente durante el curso, es un sencillo **rectángulo** del que salen los dos terminales, el segundo es una línea en zigzag.

Existen normativas oficiales para los **símbolos** eléctricos, aunque en la práctica no son siempre seguidas al pie de la letra.

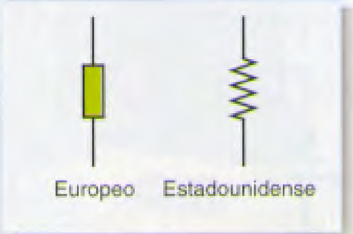

Símbolos de las resistencias.

Calcular la corriente

Con la ley de Ohm es fácil calcular la corriente que pasa por un circuito; veamos algunos ejemplos

En el **esquema** mostrado en la figura, una pila de 12 voltios está conectada con dos cables a una resistencia de 6 ohmios. ¿Cuánta corriente pasa?

El cálculo es muy sencillo. La ley de Ohm dice que hay que **dividir** la tensión (en voltios) por la resistencia (en ohmios):

$$I = V / R$$

En nuestro caso V es una tensión de **12 V** y R es un resistor de **6 Ω**, así:

$$I = 12 / 6 = 2$$

Es decir, pasa una corriente de **2 amperios**. Muchos problemas en proyectos electrónicos se resuelven aplicando esta fórmula.

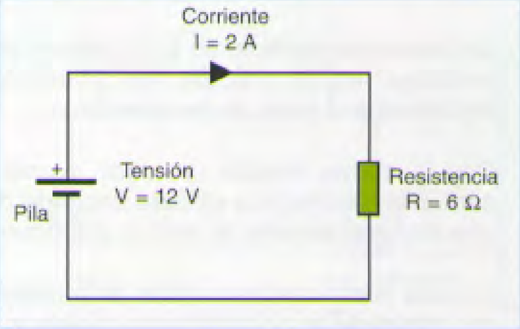

Para saber cuánta corriente pasa, se divide la tensión por la resistencia.

TODO ESTÁ EN PROPORCIÓN

Matemáticamente hablando, la ley de Ohm dice que la corriente es **proporcional** a la tensión e inversamente proporcional a la resistencia. En la práctica, esto significa que duplicando la tensión (24 voltios en lugar de 12) se **duplica** también la corriente:

$$I = 24\ V / 6\ \Omega = 4\ A$$

Esto es razonable: cuanto más se empuja, más corriente pasa. Duplicando la resistencia, la corriente se **divide** por la mitad:

$$I = 12\ V / 12\ \Omega = 1\ A$$

Es como cerrar parcialmente un grifo: saldrá menos agua.

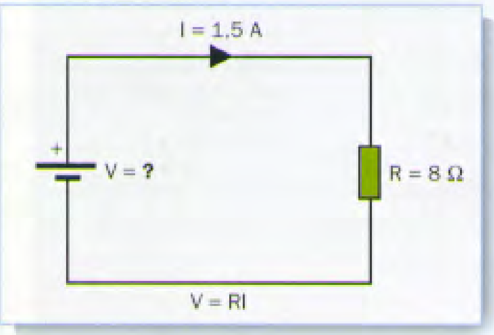

¿Qué tensión es necesaria para hacer pasar 1,5 amperios por una resistencia de 8 ohmios? V = RI, es decir, 8 Ω x 1,5 A = 12 V.

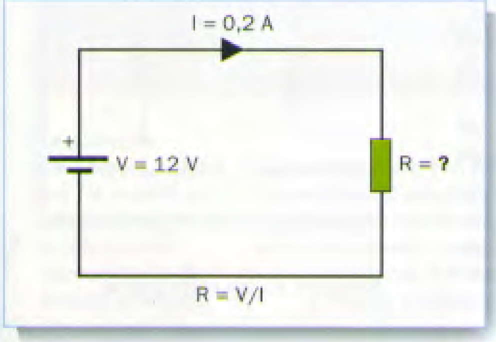

¿Cuánta resistencia es necesaria para hacer pasar 0,2 amperios (es decir 200 mA) si la tensión es de 12 voltios? R = V/I, es decir, 12 V / 0,2 A = 60 Ω.

DEMOS LA VUELTA A LA LEY DE OHM

La ley de Ohm se puede escribir también de otras dos maneras, perfectamente **equivalentes** a la que acabamos de ver, que sirven en el caso de que la corriente sea un dato conocido:

$V = RI$ (es decir R x I)

Significa que la **tensión** se obtiene multiplicando la corriente por la resistencia.

$$R = V / I$$

Quiere decir que la resistencia se obtiene **dividiendo** la tensión entre la corriente. Las figuras ilustran estas dos variantes de la fórmula.

Potencia eléctrica

La corriente eléctrica transporta energía que puede convertirse a otras formas, como luz, calor, sonido o movimiento

El paso de la corriente por una resistencia causa un **calentamiento**: la energía eléctrica se convierte en calor. Así funcionan, por ejemplo, tostadoras, planchas, estufas, hornos y calentadores de agua eléctricos.

Otros dispositivos **convierten** la energía eléctrica en luz (bombilla), sonido (altavoz), energía mecánica (motor eléctrico), etc.

También en estos casos, una parte de la energía eléctrica utilizada (se suele decir "absorbida" o "consumida") se **dispersa** como calor, es decir, se "disipa".

Una tostadora contiene una resistencia que convierte la energía eléctrica en energía térmica, es decir, en calor.

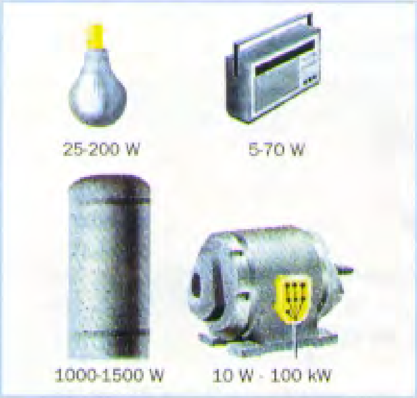

Potencias típicas empleadas por los distintos dispositivos eléctricos.

POTENCIA: WATIOS

La energía se mide en **julios** (símbolo J), que toman su nombre del científico inglés James Prescott Joule. Pero no es ésta la unidad de medida más interesante.

Como en los automóviles, es preferible conocer la **potencia**, es decir, la energía convertida en cada segundo.

La potencia se mide en **watios** (símbolo W), nombre proveniente del ingeniero escocés James Watt, que perfeccionó la máquina de vapor.

Por ejemplo, una **estufa** eléctrica convierte cerca de 2 KW (dos kilowatios, dos mil watios) en calor, mientras que el motor de un utilitario produce más o menos 40 KW de energía mecánica, y mucho más calor.

CALCULAR LA POTENCIA

La sencilla **ley de Joule** permite conocer la potencia eléctrica absorbida por un dispositivo. Basta con multiplicar la tensión en sus extremos por la corriente que lo atraviesa:

$W = VI$

Por ejemplo, si en los extremos de una plancha hay 220 V y la corriente que lo atraviesa es de 2 A, la potencia absorbida es de:

$220 \times 2 = 440$ W

es decir, 440 watios, más que suficiente para mantenerla a la temperatura necesaria para su utilización.

Un ordenador personal normal absorbe cerca de 300 W en el momento de su encendido y después unos 100 W durante su funcionamiento normal.

Resistencias no deseadas

En el mundo real todos los componentes, cables incluidos, ofrecen una cierta resistencia al paso de la corriente: es necesario tener esto en cuenta

El esquema eléctrico es una representación **ideal** de un circuito: cuando se pasa a la práctica hay siempre pequeñas diferencias.

Por ejemplo, los filamentos y los cables ofrecen una cierta resistencia. Según la ley de Ohm, cuando son atravesados por la corriente se crea en sus extremos una **tensión** (o "diferencia de potencial"). Esto significa que, cuando pasa la corriente, la tensión en los dos extremos de un cable no es nunca idéntica: se tiene una pequeña **caída de tensión**.

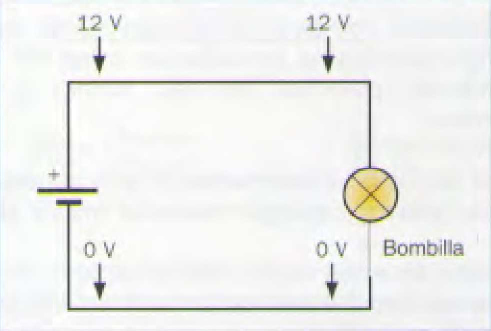

Si los cables fueran conductores ideales, no se tendría caída de tensión.

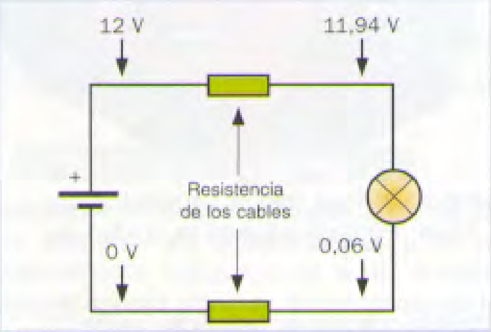

A causa de la resistencia de los cables, la tensión en la bombilla es inferior a la de la batería.

LOS CABLES SE CALIENTAN

Si un cable es atravesado por corriente, y por lo tanto entre sus extremos hay una tensión, según la ley de Joule una determinada **potencia** se transforma en calor.

Pongamos un ejemplo práctico: una **estufa** de 2,2 KW, es decir, 2.200 W. Si la tensión es de 220 V, la corriente absorbida es de 10 A (220 x 10, que da un total de 2.200).

Conectémosla a la toma con un alargador y supongamos que éste tiene una resistencia de 100 mΩ (miliohmios, milésimas de ohmio), es decir, 0,1 ohmios.

Según la ley de Ohm, multiplicamos 10 A por 0,1 ohmios y obtenemos 1 V: ésta es la **caída de tensión** sobre el alargador.

En los extremos del alargador hay por tanto 1 V, pero el alargador es atravesado por una corriente de 10 A y por tanto la **potencia** disipada por el cable es de 1 x 10 = 10 W.

La potencia disipada en calor por los cables es uno de los problemas del transporte de la energía eléctrica.

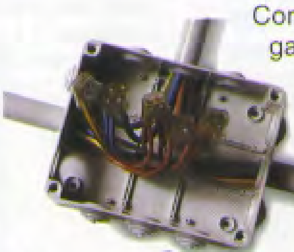

Una abrazadera con los tornillos mal apretados tiene una resistencia excesiva y puede causar calentamientos, averías e incluso incendios.

La corriente no sale por los cables

Igual que una tubería de agua bien sellada, un circuito electrónico no tiene pérdidas: lo que entra, sale.

Los coches que entran en la autopista deben después **salir** por alguna parte (aunque no necesariamente enteros...).

Lo mismo ocurre con los **electrones** que componen la corriente: si en un cable (perfectamente aislado del exterior) entran 3 amperios por un extremo, por el otro saldrán exactamente los mismos 3 amperios.

Esta característica de la corriente hace más fácil la comprensión del **funcionamiento** de los circuitos y la lectura de los esquemas eléctricos.

UNA CURIOSIDAD: CONTAR LOS ELECTRONES

En la práctica no se cuentan los electrones, sino que se utilizan los **culombios** (símbolo C), llamados así por el físico francés Charles de Coulomb.

Un culombio equivale a la **carga eléctrica** de un gran número de electrones, tantos que es necesario un número con 19 ceros para contarlos.

El flujo de electrones en movimiento forma la corriente. Igual que una corriente de agua se mide en litros por segundo, una corriente eléctrica se mide en culombios por segundo (C/s) o **amperios**.

Así, hemos vuelto al punto de partida: la corriente y los amperios. Pero es esencial entender que "un amperio" significa "un **flujo** de una cantidad de electrones por segundo".

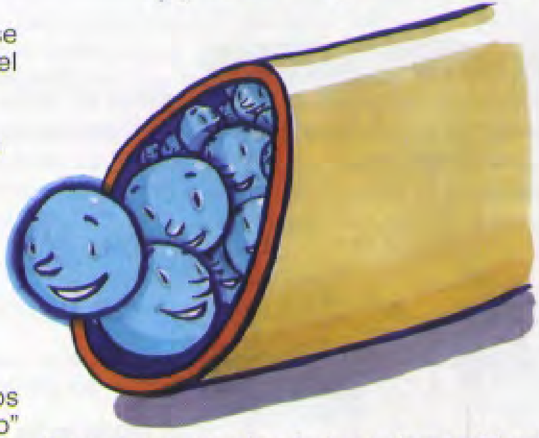

Se puede pensar en los electrones de un cable como en pequeñas esferas dentro de un tubo: empujándolas hacia dentro por un lado salen las mismas por el otro.

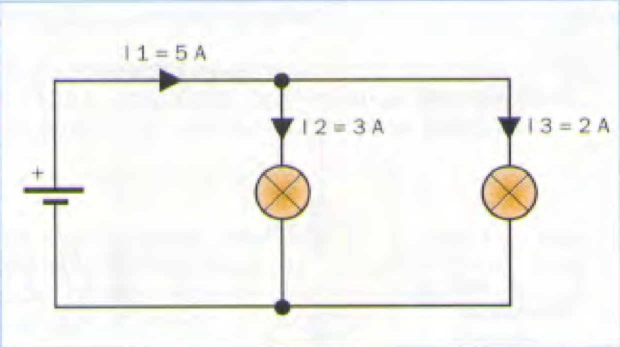

La suma de las dos corrientes que van a las bombillas (I2 e I3) es igual a la corriente I1 que entra.

CUANDO UN CABLE SE DIVIDE

¿Qué sucede en un **empalme** entre dos cables? Lo mismo que ocurre en una bifurcación de una carretera: el total de los coches que salen, sea cual sea el destino que elijan, es igual al de los coches que entran.

Así tenemos la primera **ley de Kirchhoff** fundamental: la suma de las corrientes que salen de un nodo (es decir, un empalme, una unión) es igual a la suma de las que entran.

De este modo, la corriente no se **pierde**: su total es constante.

Resistencias en serie

Dos o más resistencias en serie son atravesadas por la misma corriente: sus valores se suman

¿Qué sucede si dos resistencias están conectadas **en serie**, es decir, una después de otra? Dado que la corriente no puede "escapar", deberá atravesarlas a ambas.

Como los electrones que entran por una parte deben salir obligatoriamente por la otra, ambas resistencias son atravesadas por la **misma corriente** y ambas la obstaculizan.

Las bombillas que se utilizan en las decoraciones de navidad están generalmente conectadas en serie, una después de otra en el mismo cable.

LAS TENSIONES SE SUMAN

Cuando pasa corriente, en los extremos de cada resistencia habrá una **tensión** que dependerá de su valor. Recordemos la ley de Ohm:

V = RI

Se dice además que el paso de la corriente en una resistencia causa una **diferencia de potencial** (es decir, una tensión eléctrica) en sus extremos.

Observando la figura, si una corriente de 3 amperios atraviesa una resistencia de 4 Ω (R1) y una de 2 Ω (R2), **en sus extremos** se encuentran respectivamente las dos tensiones V1 y V2:

V1 = R1 x I = 4 Ω x 3 A = 12 V
V2 = R2 x I = 2 Ω x 3 A = 6 V

Como las tensiones (igual que las altitudes) se **suman**, en total hay 18 V. Para que circule esta corriente de 3 A se necesitan exactamente 18 V.

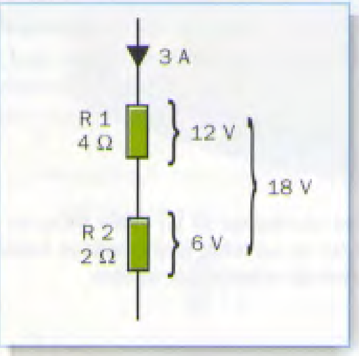

La tensión en los extremos de cada resistor depende de la corriente que lo atraviesa (igual en ambos) y de su resistencia.

RESISTENCIA TOTAL

Tomemos por un momento el circuito **por dentro**, ignorando el hecho de que esté formado por dos resistencias en serie.

En sus extremos hay 18 V y la corriente que circula es de 3 A, por tanto podemos aplicar la ley de Ohm para conocer la resistencia **global**:

R = V / I = 18 V / 3 A = 6 Ω

Estos 6 Ω no son otra cosa que la **suma** de los valores de las dos resistencias en serie: 2 Ω + 4 Ω, es decir, dos o más resistencias en serie simplemente se suman.

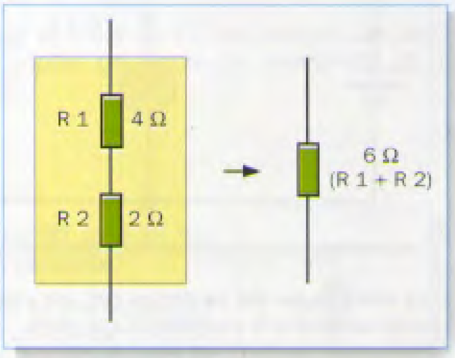

Dos resistencias en serie equivalen a una resistencia de valor igual a su suma.

Resistencias en paralelo

En el caso de las resistencias en paralelo, es la tensión lo que se mantiene igual; la corriente se divide entre las dos bifurcaciones

La otra posibilidad de conexión es **en paralelo**: ambas resistencias se conectan a los mismos puntos. Una conexión en paralelo típica es la de los aparatos eléctricos domésticos conectados a un enchufe con tres entradas (denominado comúnmente "ladrón"): a cada uno de ellos le llega la **misma** tensión de 220 V.

A la salida de estos enchufes triples los dispositivos están en paralelo: la tensión es igual para todos.

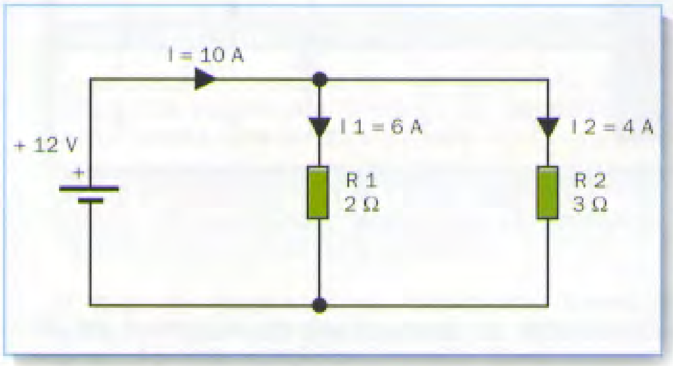

Conexión en paralelo: la corriente que atraviesa cada resistencia puede calcularse por separado.

UNA TENSIÓN, VARIAS CORRIENTES

En una conexión en paralelo, la tensión es igual para todos, pero la corriente no. En la práctica, cada dispositivo es **independiente** de los demás.

Observando el esquema vemos que la misma tensión de 12 V se aplica a dos resistencias R1 y R2, respectivamente de 2 Ω y 3 Ω. Con la ya habitual ley de Ohm averiguamos las dos **corrientes**:

I1 = V / R1 = 12 V / 2 Ω = 6 A
I2 = V / R2 = 12 V / 3 Ω = 4 A

Así, la **corriente total** que debe proporcionar la batería es de 6 + 4 = 10 A.

FÓRMULA DEL PARALELO

Si consideramos las dos resistencias en paralelo como una "caja negra" y fingimos que es una sola resistencia, ¿cuál es su **valor**?

Dado que conocemos la tensión (12 V) y la corriente total (10A), aplicamos de nuevo la **ley de Ohm**, prácticamente la regla universal de la electrónica:

R = V / I = 12 V / 10 A = 1,2 Ω

Existe una **fórmula**, mostrada en la figura (y que puede obtenerse a partir de la ley de Ohm), para conocer rápidamente el valor del paralelo de dos resistencias R1 y R2:

En nuestro caso podemos verificarla **fácilmente**:

R = (2 x 3) / (2 + 3) = 6 / 5 = 1,2 Ω

Una nota práctica: en estos ejemplos utilizamos por sencillez ohmio (Ω) y amperio (A), pero en los circuitos reales de baja potencia es más probable encontrar **kiloohmios** (KΩ) y **miliamperios** (mA).

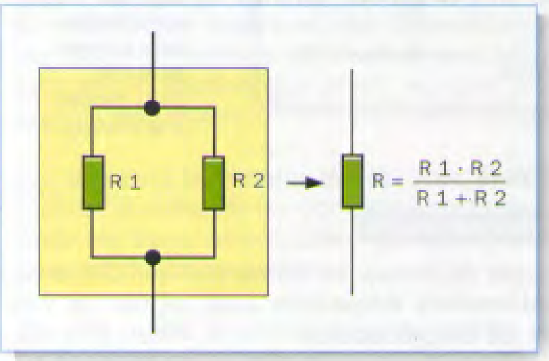

Fórmula para calcular la resistencia global de dos resistores en paralelo.

Divisores

El comportamiento de las resistencias en serie permite subdividir una tensión como se quiera

Poniendo una cajonera de 60 cm de altitud **encima** de otra de 120 cm, la altitud total desde el suelo (el punto de referencia) es de 180 cm.

Lo mismo ocurre con las **tensiones**: 6 V en serie con 12 V dan un total de 18 V, como hemos visto en el caso de las resistencias en serie.

Normalmente, en electrónica se toma como referencia la **masa**, que a menudo coincide con el terminal negativo de la batería u otra fuente de alimentación (energía que hace funcionar el circuito).

Podemos medir tensiones **absolutas**, es decir, con respecto al punto elegido como masa, o **relativas**, esto es, entre dos puntos del circuito.

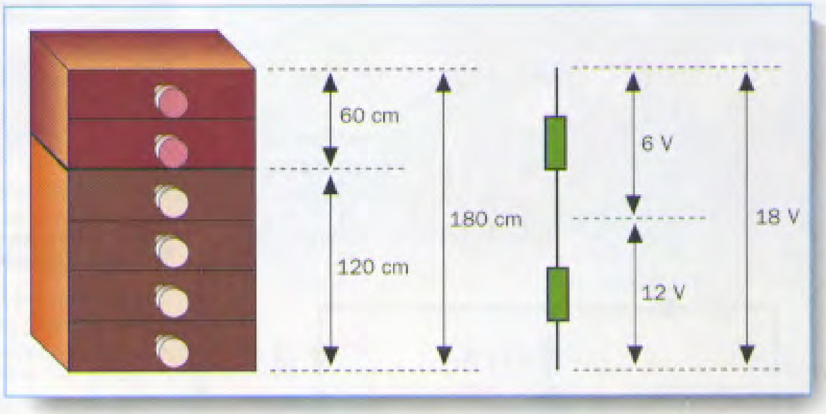

Altura y tensión son valores relativos, aunque normalmente se toma un "punto cero" de referencia.

SUBDIVIDIR UNA TENSIÓN

Dos resistencias en serie pueden utilizarse para obtener una **fracción** de la tensión total: es suficiente con mirar el esquema bajo otro punto de vista.

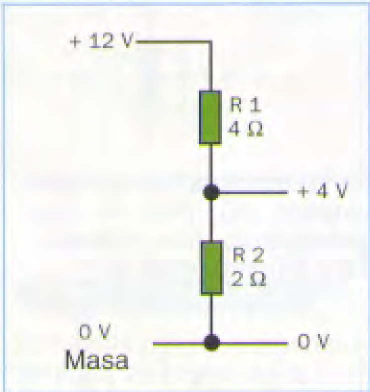

Divisor realizado con dos resistencias, para obtener un tercio de la tensión a la entrada.

Observando la figura, dividimos la tensión de **alimentación** (12 V) por la resistencia total ($4\ \Omega + 2\ \Omega = 6\ \Omega$):

$I = V / R = 12\ V / 6\ \Omega = 2\ A$

La **corriente** que atraviesa ambas resistencias es por tanto de 2 A. En los extremos de la resistencia de abajo (R2) hay:

$V = R2 \times I = 2\ \Omega \times 2\ A = 4\ V$

Como un extremo de la resistencia está conectada a masa (0 V, por definición), el otro extremo lo está a 4 V: un tercio de la tensión total. Así, tenemos un **divisor** de tensión.

Observemos que el valor de R2 es también un tercio de la resistencia total: **un tercio** de resistencia da un tercio de tensión.

POTENCIÓMETROS

El principio del divisor se utiliza, por ejemplo, en los **potenciómetros** empleados para regular el volumen de los amplificadores.

Se trata de resistencias con una toma intermedia dirigida mecánicamente, por ejemplo por una manivela giratoria, de modo que se obtiene un **divisor variable**.

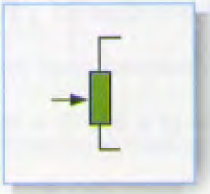

Un potenciómetro es una resistencia con una toma intermedia móvil: se comporta como dos resistencias en serie, que forman un divisor.

Corriente alterna

Haciendo circular la corriente en los dos sentidos en lugar de en uno solo, es más fácil transportar energía o información

La **tensión** de la instalación doméstica (220 V), y por tanto la corriente que circula en electrodomésticos y bombillas, no es continua.

Se mueve alternativamente hacia delante y hacia atrás, cambiando continuamente de intensidad de un modo similar al movimiento de un **péndulo**.

Como se puede ver en la figura, la tensión varía continuamente en el tiempo entre distintos valores positivos y negativos, siguiendo una curva matemática muy común llamada **sinusoide**.

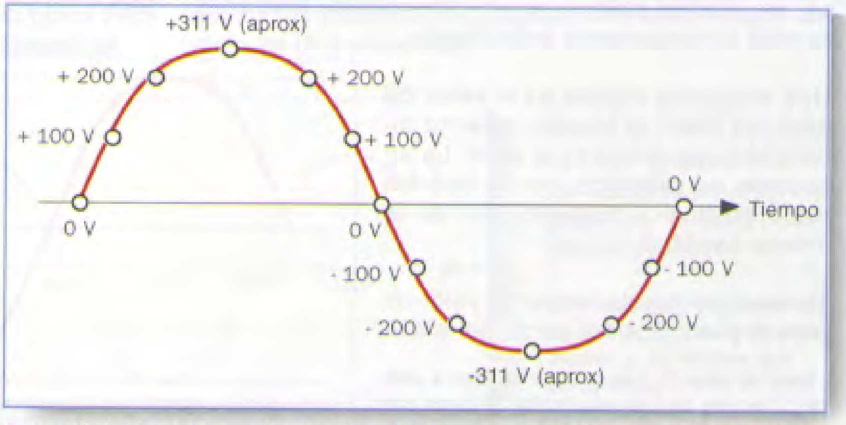

Forma de onda de la tensión de red utilizada en la instalación doméstica: es una sinusoide.

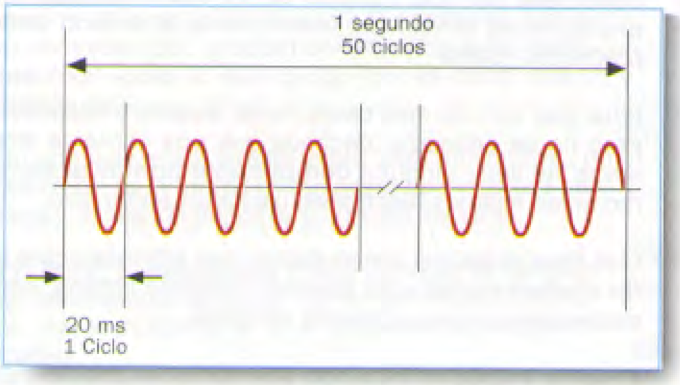

Cada segundo, la corriente circula hacia delante y hacia atrás 50 veces.

FRECUENCIA Y PERÍODO

El comportamiento mostrado se repite siempre igual. No lo vemos porque se produce con bastante rapidez: **50 ciclos por segundo**.

La **frecuencia** es eso mismo, el número de ciclos por segundo; se mide en hercios (símbolo Hz), nombre procedente del físico alemán Heinrich Hertz.

Así, la tensión de red española tiene una frecuencia de **50 Hz**. Otros países, como los Estados Unidos, emplean 60 Hz por motivos históricos.

Cada ciclo dura 1/50 de segundo, es decir 20 milisegundos (ms): este es el **período** de la onda, al término del cual se repite idéntica.

USOS DE LA CORRIENTE ALTERNA

De la corriente alterna se puede cambiar la tensión (a misma potencia transportada) con un dispositivo llamado **transformador**.

Ello permite, por ejemplo, **reducir** los 220 V de casa para obtener las bajas tensiones (por ejemplo 12 V) utilizadas por muchos dispositivos electrónicos.

En lugar de energía, una corriente alterna puede transportar **información**, como la señal producida por un micrófono, que reproduce la evolución de una onda sonora.

La corriente alterna y los transformadores simplifican el transporte de la energía eléctrica, que se produce a tensiones elevadas, incluso a cientos de kilovoltios.

Valor de la corriente alterna

Si tensión y corriente varían continuamente, ¿cómo se puede tener un valor exacto?

El **valor instantáneo** de una tensión o corriente alterna, es decir, el medido en un determinado momento, cambia continuamente en el tiempo.

Una referencia estable es el **valor de pico**, es decir, la tensión máxima alcanzada con respecto al cero. La sinusoide es simétrica, por lo que los picos positivo y negativo son de la misma amplitud.

También se puede indicar el valor de **pico a pico** (pp), es decir, del máximo positivo al máximo negativo, que es obviamente el doble del valor de pico.

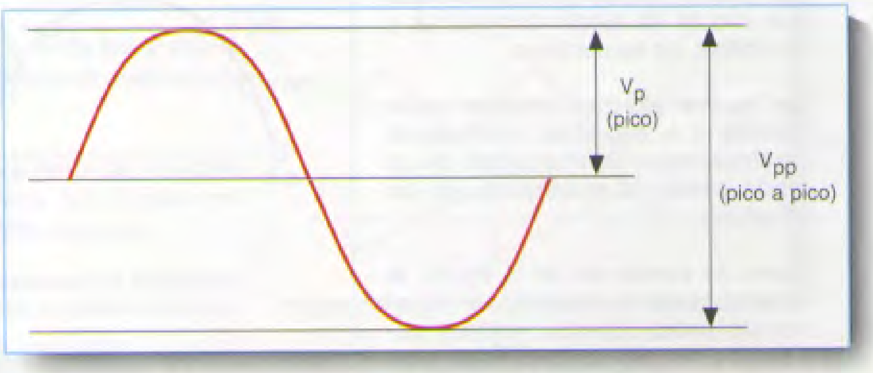

Valor de pico (V_p) es el valor de pico a pico (V_{pp}) en una tensión sinusoidal. Si fuera una corriente utilizaríamos los símbolos I_p y I_{pp}.

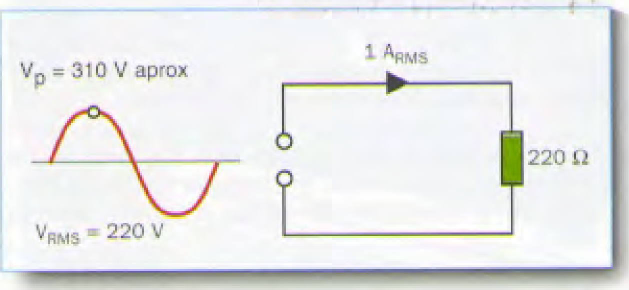

Esta tensión alterna produce en la resistencia una disipación de 220 W, la misma que se tendría con una tensión continua de 220 V.

VALOR MEDIO Y VALOR EFICAZ

Dado que las dos mitades de la onda eléctrica son exactamente simétricas, como media la tensión (o corriente) es ¡**cero**!

Igual que un columpio oscila hacia delante y hacia atrás pero no se desplaza, decimos que una corriente sinusoidal no tiene ninguna **componente continua**: es como si los mismos electrones oscilaran en su sitio.

Con fines prácticos, sin embargo, nos interesa sobre todo el **efecto** producido por una corriente alterna, comparándola con una corriente continua.

El **valor eficaz** (o "rms") de una corriente alterna es el valor que debería tener una corriente continua para producir el mismo calentamiento en una resistencia. Lo mismo ocurre con la tensión.

CUALQUIER NÚMERO

Para una **sinusoide** se da normalmente el valor eficaz (rms). Cuando decimos 220 V_{ac}, entendemos 220 V_{eff} (o 220 V_{rms}).

Para obtener el valor de pico basta con multiplicar por la raíz cuadrada de 2 ($\sqrt{2}$), es decir 1,41 aproximadamente. Así, para los 220 V domésticos tenemos:

$$V_p = V_{rms} \times \sqrt{2} = 220\text{ V} \times 1{,}41 = 310\text{ V aprox.}$$

Por otra parte, 220 V_{rms} tienen el mismo efecto (por ejemplo, calentamiento de una resistencia) que 220 V continuos, por lo que es este el **valor** que en realidad nos interesa.

Ángulos y fase

Es útil considerar el ciclo de una sinusoide como si se tratase de la rotación de una rueda

Tomando un punto de una rueda y midiendo su **altura** (con respecto a los ejes) mientras la rueda gira a velocidad constante se obtiene... una onda sinusoidal.

Esta estrecha relación entre rotación y sinusoide resulta muy cómoda para indicar un **punto** exacto del ciclo de una sinusoide.

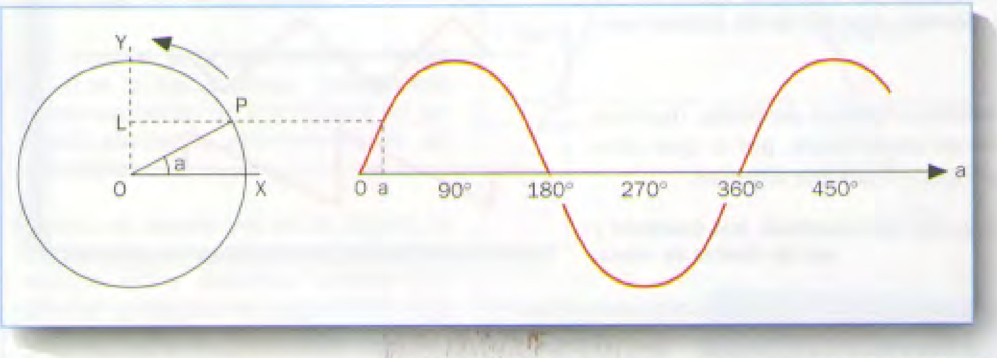

En lugar de utilizar fracciones de ciclo, se utilizan los **grados** normales sexagesimales: un giro, y por tanto un ciclo, son 360 grados.

Haciendo girar P y tomando su altura (OL) al variar el ángulo a, se obtiene una sinusoide.

RADIANES

En lugar de los grados se utiliza habitualmente otra medida del ángulo de rotación: los radianes. Un giro completo (360 grados) corresponde a **2π radianes** (dos veces la letra griega "pi", es decir, más o menos 6,28).

En consecuencia, **medio ciclo** son 180 grados, o bien (que es lo mismo) π radianes. La sinusoide pasa por el cero al principio y a mitad del ciclo.

Por tanto, la **frecuencia** puede expresarse también en radianes por segundo: 50 Hz son 50 x 2π, es decir, aproximadamente 314 radianes por segundo (rad/s).

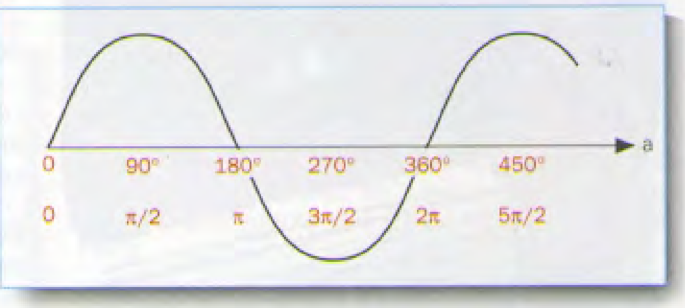

Correspondencia entre grados y radianes.

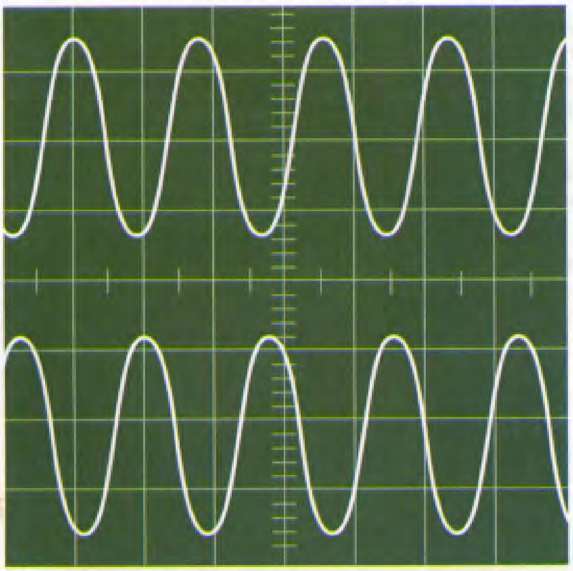

FASE Y DESFASE

Dos sinusoides pueden tener la misma frecuencia pero pasar por el cero en distintos momentos: se dice entonces que no tienen la misma **fase**, o que están "desfasados".

La diferencia se mide en ángulos, o en radianes. En particular, si la diferencia es de 180 grados (π radianes), están en **contrafase** y aparecen una como la imagen especular de la otra.

Pero si la diferencia es de sólo 90 grados (π/2 radianes) más o menos, se dice que las dos ondas están en **cuadratura**: una alcanza el máximo mientras que la otra pasa por el cero, y viceversa.

Dos tensiones sinusoidales en contrafase: la segunda es idéntica pero inversa, es decir, con el signo cambiado.

Formas de onda

El comportamiento de las resistencias en serie permite subdividir una tensión como se quiera

La sinusoide es una **forma de onda** común, no sólo en electrónica, sino también en el mundo físico: es sinusoidal, por ejemplo, una nota "pura", como la de un diapasón.

También lo es, como veremos, la forma de onda **base** con la que es posible construir las demás (combinando convenientemente varias sinusoides).

No obstante, existen **otras** infinitas formas de onda, muchas de las cuales son recurrentes en electrónica, por lo que conviene saberlas reconocer. La figura muestra algunas.

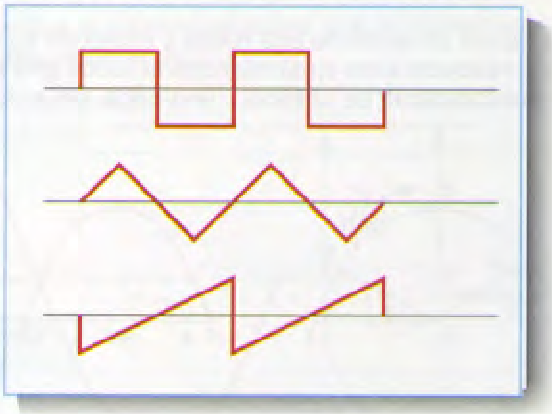

Tres formas de onda características: una onda cuadrada, una triangular y una de dientes de sierra.

ONDAS CUADRADAS

Una onda cuadrada cambia **bruscamente** su valor a intervalos regulares. Por ejemplo, la tensión puede alternar su valor entre +12 V y -12 V.

Las ondas cuadradas se encuentran normalmente trabajando con los **circuitos digitales**, porque utilizan dos únicos valores de tensión.

Es interesante observar que el valor eficaz (RMS) de una onda cuadrada es **igual** a su valor de pico: la tensión, realmente, está siempre al máximo y no pasa nunca (al menos idealmente) por valores intermedios.

COMPONENTE CONTINUA

Si una tensión varía periódicamente entre +4 V y +12 V, ¿se trata de una tensión **alterna** o no? Después de todo, el signo es siempre positivo.

En casos como este es útil considerar la onda como suma de dos **componentes**: una continua y otra alterna.

Si la onda tiene forma simétrica (por ejemplo sinusoidal o cuadrada), la **componente continua** (u "offset") es el valor medio: entre 4 V y 12 V la media es (4 + 12) / 2, es decir 8 V.

Este valor representa el **cero** con respecto al que medimos la tensión alterna: varía entre 8 V - 4 V y 8 V + 4 V; por tanto tiene un valor de 4 V_p, es decir, 4 V de pico.

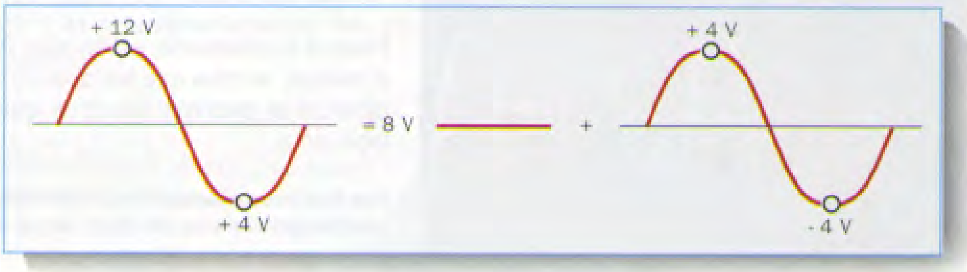

Toda forma de onda con el valor medio distinto de cero puede ser considerada como suma de una componente continua y una alterna.

Capacidad y condensadores

El condensador es un dispositivo capaz de acumular energía eléctrica

Retomando la comparación entre corriente eléctrica y corriente de agua, un condensador es como un **depósito** donde se pueden acumular los electrones. Un condensador sencillo está formado por dos **placas** metálicas, o "armaduras": las cargas de signo opuesto (positivo y negativo) de las dos placas se atraen, concentrándose sobre las propias placas.

El efecto se potencia si entre las placas se encuentra material **dieléctrico**, es decir, que tiende a polarizarse asumiendo carga negativa en un lado y positiva en el otro.

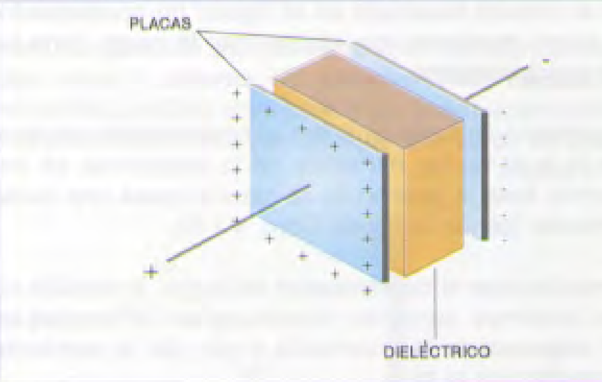

Dos placas metálicas y material dieléctrico (si no hay otro, el mismo aire) forman un condensador capaz de acumular carga eléctrica.

CARGA Y CAPACIDAD

La capacidad de un condensador para acumular carga eléctrica se mide en **faradios** (símbolo F), medida que toma su nombre del físico británico Michael Faraday. Un faradio es una capacidad considerable; normalmente se utilizan los **submúltiplos** microfaradio (μF), nanofaradio (nF) y picofaradio (pF).

El número de electrones acumulados, es decir, la carga eléctrica, depende tanto de la capacidad como de la **tensión** aplicada:

$$Q = V \times C$$

Es decir: Q, la carga medida en culombios, se obtiene multiplicando la tensión por la **capacidad** del condensador.

El símbolo del condensador representa las placas metálicas. Algunos están polarizados, es decir, tienen un lado positivo y otro negativo.

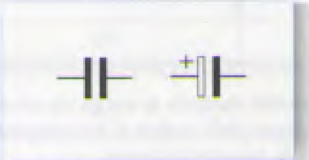

USOS DE LOS CONDENSADORES

El condensador es un componente **fundamental** de los circuitos electrónicos y sus aplicaciones son bastante numerosas.

Además de la acumulación de energía, citemos por ejemplo la **separación** de la componente continua de la alterna, y la realización de circuitos de retardo (temporizadores).

Existen condensadores de todo tipo para las numerosas aplicaciones de este útil componente.

Carga y descarga

La carga de un condensador requiere un cierto tiempo y tiene un desarrollo característico

En el circuito mostrado en la figura, un condensador se carga mediante una resistencia: la carga tiene lugar **gradualmente**.

Al principio, la **tensión** sobre el condensador es cero, por lo que, en los extremos de la resistencia se encuentra toda la tensión de la batería: pasa una cierta corriente (según la ley de Ohm, V / R).

A medida que el condensador se carga, la tensión en sus extremos aumenta; disminuye así la tensión en los extremos de la resistencia y con ella la **corriente** de carga.

Así, la tensión en el condensador sube cada vez más despacio: en teoría el tiempo de carga completa es **infinito**.

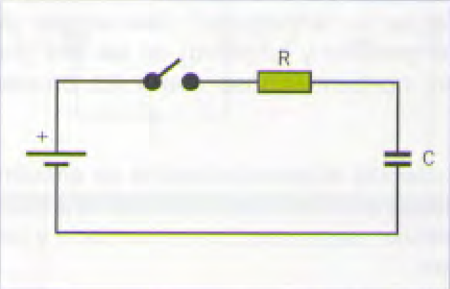

Cerrando el interruptor, el condensador empieza a cargarse, al principio rápido, pero después cada vez más despacio.

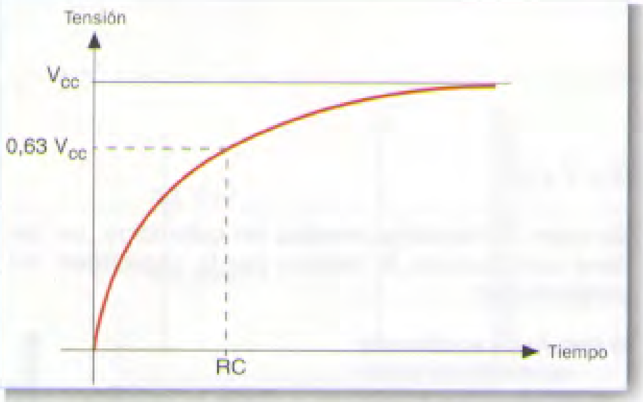

Tensión durante la carga de un condensador. La constante de tiempo (RC) indica el tiempo necesario para alcanzar el 63% de la tensión aplicada.

CONSTANTE DE TIEMPO

Multiplicando el valor de la resistencia por el del condensador se obtiene un tiempo en **segundos**. Por ejemplo, para R = 10 kΩ y C = 200 µF:

$T = R \times C = 10.000 \times 0{,}0002 = 2$ segundos

Esta constante RC, o **constante de tiempo**, indica el tiempo necesario para cargar el condensador al 63% de la tensión aplicada.

Así, es posible, por ejemplo, construir circuitos **temporizadores** de características previsibles con una cierta precisión.

DESCARGA RÁPIDA

Si la resistencia tiene un valor muy bajo, por ejemplo es un **cortocircuito** entre los dos terminales del condensador, la corriente inicial es considerable.

Se tiene así una chispa o una "detonación", según la cantidad de energía almacenada, **energía** que depende de la capacidad y del cuadrado de la tensión:

$E = (C \times V^2) / 2$

Un condensador cargado a alta tensión puede ser bastante **peligroso**, conservando una energía potencialmente letal incluso en un aparato desconectado durante varios días.

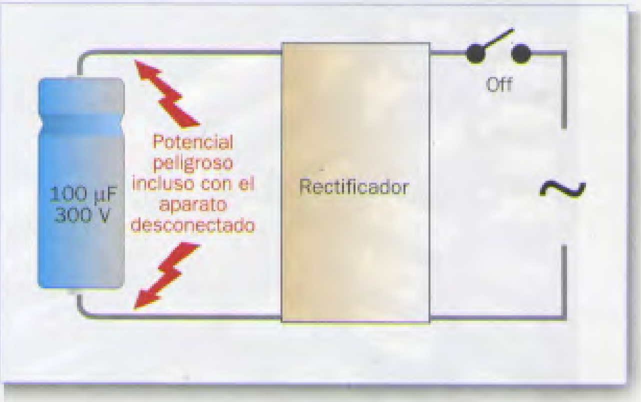

¡Atención a los condensadores cargados!

Acoplamiento capacitivo

Un condensador permite separar la componente alterna de la continua

Una **membrana elástica** situada en un tubo impide el paso continuo del agua, pero le permite pequeños desplazamientos hacia delante y hacia atrás. Del mismo modo, un condensador bloquea la corriente continua pero, cargándose alternativamente en los dos sentidos, deja pasar la **alterna**. Un **acoplamiento capacitivo** entre dos circuitos sirve para dejar pasar solamente la componente alterna, por ejemplo una señal de audio, como muestra la figura.

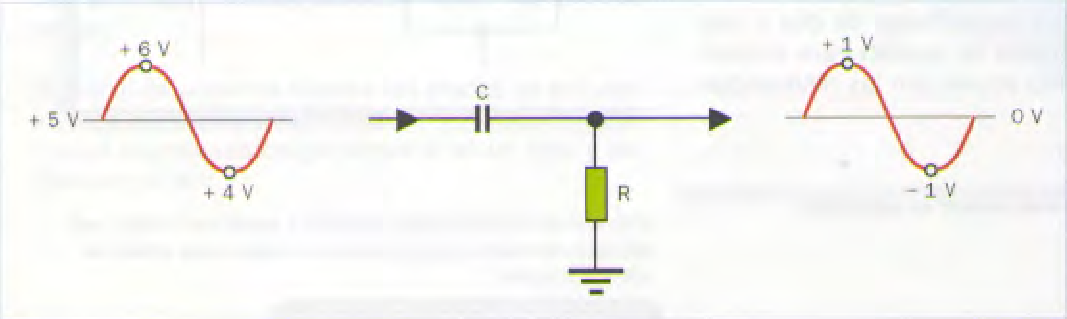

Un condensador es un aislante para la componente continua y un conductor (más o menos bueno) para la alterna.

FILTRADO

Intercambiando resistencia y condensador se obtiene el efecto opuesto: la corriente **continua** pasa, mientras que la alterna es cortocircuitada a masa. Se habla de "filtrado" o "desacoplamiento".

El comportamiento de un circuito de corriente continua se puede examinar con facilidad simplemente **ignorando** todos los condensadores, como si no existieran. En corriente alterna, sin embargo, en una primera aproximación se los puede considerar como simples **cables** de conexión.

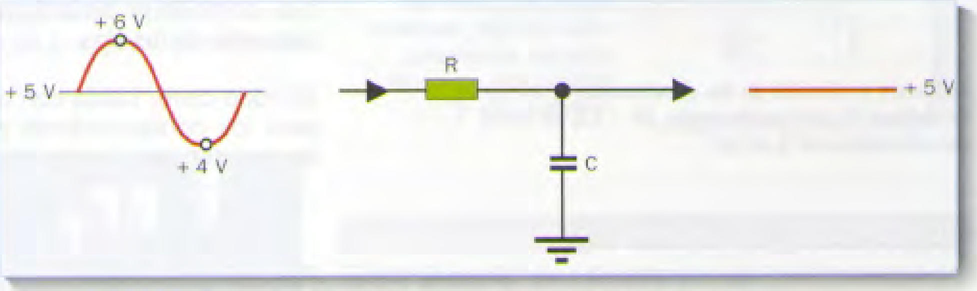

Un filtro RC *(resistencia-condensador) elimina la componente alterna, al menos para valores altos de los componentes y de la frecuencia.*

Los controles de tono aprovechan (también) la impedancia de los condensadores a distintas frecuencias.

FRECUENCIA E IMPEDANCIA

En realidad, un condensador no es nunca un conductor perfecto para la corriente alterna, sino que ofrece una cierta resistencia, o más exactamente **impedancia**. Ésta depende del valor del condensador y de la **frecuencia** de trabajo: cuanto mayores sean estos valores, menor será la impedancia y, por tanto, mejor la conducción de la corriente.

Veremos cómo se puede aprovechar esta dependencia de la frecuencia para construir **filtros**, es decir, circuitos que dejan pasar ciertas frecuencias (por ejemplo, los agudos de la música) mejor que otras (por ejemplo, los bajos).

Serie y paralelo

Las combinaciones de condensadores funcionan... al contrario que las de las resistencias

Poner dos condensadores en **paralelo** significa, en la práctica, conectar las placas de uno con las de otro.

Es como tener un solo condensador con el **doble** de superficie útil de las placas, y por tanto con el doble de capacidad.

Dicho de otro modo, las capacidades de dos o más condensadores en paralelo se **suman** para obtener la capacidad total (como ocurre con las resistencias en serie).

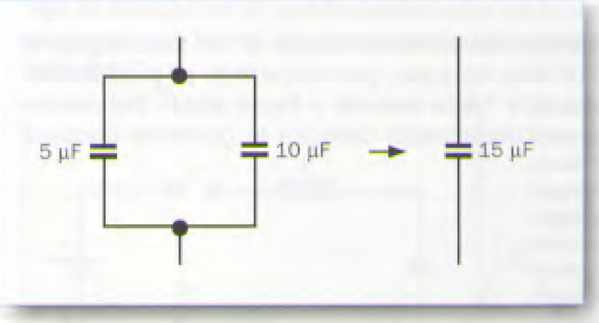

Dos condensadores en paralelo suman su capacidad.

CONDENSADORES EN SERIE

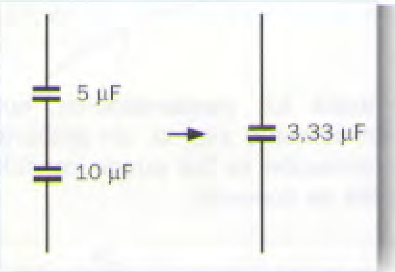

Aplicando la fórmula de los condensadores en serie, se obtiene el valor equivalente: 50 / 15, es decir, aproximadamente 3,33 µF.

La relación con las resistencia se aplica también en el caso de dos condensadores en serie; la **fórmula** es la misma que la utilizada para las resistencias en paralelo:

$C = (C1 \times C2) / (C1 + C2)$

donde, naturalmente, C1 y C2 son los dos condensadores en serie y C es la capacidad **equivalente** que, como se ve en la figura, es inferior a la del más pequeño de los dos.

En todo caso, basta con recordar que las fórmulas para los condensadores son **recíprocas** a las de las resistencias (intercambiar "serie" con "paralelo").

CONDENSADORES VARIABLES

Igual que existen las resistencias variables (potenciómetros), existen condensadores de valor **modificable** manualmente.

Estos se llaman simplemente "condensadores variables" o, si se regulan sólo con un destornillador, **trimmers capacitivos**.

Como veremos, el uso principal de los condensadores variables se da en los **circuitos resonantes**, es decir, capaces de resonar (como un diapasón o una cuerda de guitarra) a una cierta frecuencia.

Por ejemplo, un condensador variable puede utilizarse para **sintonizar** una radio de modo que reciba una determinada emisora.

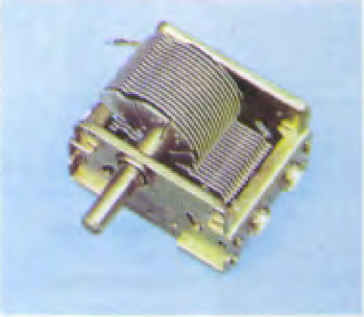

Un condensador variable: la capacidad cambia aumentando o disminuyendo la superficie de las placas.

Bobinas o Inductancias

Desplazándose por un conductor, la corriente genera un campo magnético, que se puede potenciar enrollando el cable sobre sí mismo

Fue el físico Hans Christian Ørsted quien descubrió en 1.820 que un cable recorrido por una corriente continua mueve la aguja de una brújula como si fuera un pequeño **imán**.

Dicho de otra forma, una corriente de electrones produce un **campo magnético** tan fuerte como intensa es la corriente.

Este fenómeno, llamado **electromagnetismo**, fue objeto de muchos estudios fundamentales para el desarrollo de la electrotecnia primero y la electrónica después.

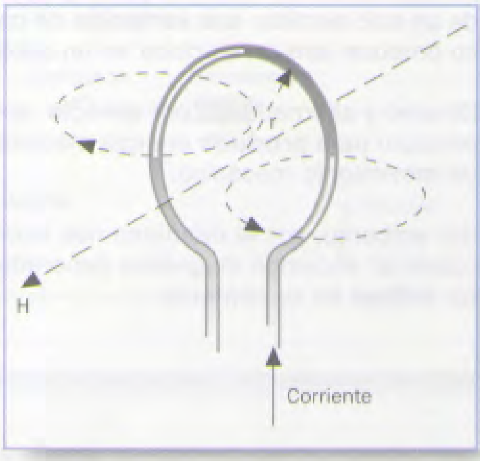

En el interior de una espira recorrida por corriente, el campo magnético es más intenso.

BOBINAS

Si el cable realiza un recorrido **circular**, es decir, una espira, los campos magnéticos en su interior se suman, aumentando de intensidad.

Superponiendo varias espiras, cada una de ellas refuerza la intensidad del campo: un dispositivo de este tipo se llama **bobina o inductor**.

El efecto se amplifica posteriormente si en el centro de la bobina se pone un **núcleo** de material "ferromagnético", por ejemplo de hierro.

El campo magnético del dispositivo así obtenido, llamado también **electroimán**, puede ser muy intenso: algunas grúas los utilizan para levantar toneladas de materiales metálicos.

Los relés se realizan con una bobina potenciada por un núcleo metálico.

NORTE Y SUR

Como los imanes normales, también los electroimanes (o "solenoides") tienen una **polaridad**: el campo magnético tienen un "polo norte" y un "polo sur".

Dos polos distintos (por ejemplo norte y sur) se **atraen**, mientras que dos polos iguales (como norte y norte) se **repelen**.

Como la polaridad depende de la **dirección** de la corriente, es posible controlarla: ello permite, como veremos, realizar motores y otras máquinas eléctricas.

Un inductor recorrido por una corriente se convierte en un* electroimán*: en sus extremos se forman dos polos magnéticos norte y sur.

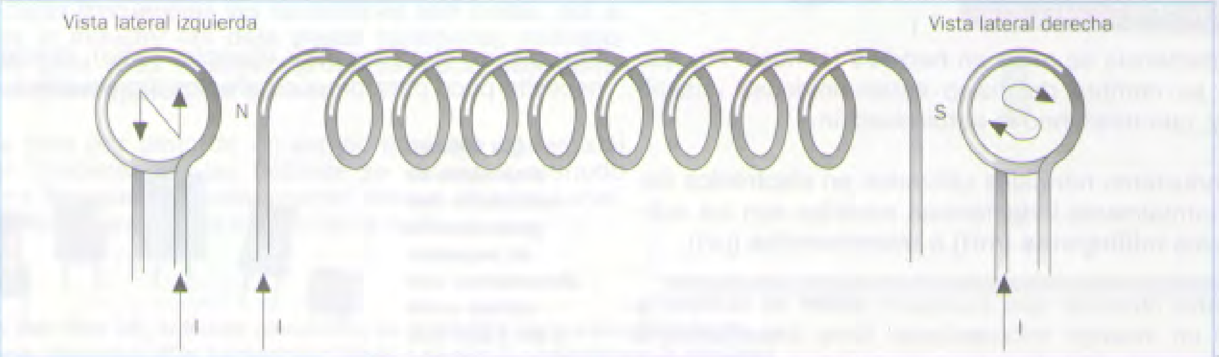

Inductancia

Un inductor tiende a oponerse a las variaciones de la corriente que lo recorre

La relación entre corriente y campo magnético no es de un solo sentido: una **variación** de campo magnético produce tensión eléctrica en un cable.

Dinamo y alternadores, por ejemplo, aprovechan este principio para **producir** energía eléctrica partiendo de un movimiento mecánico.

Sin embargo, por el momento nos interesa un efecto colateral: el campo magnético generado por un inductor **influye** en su corriente.

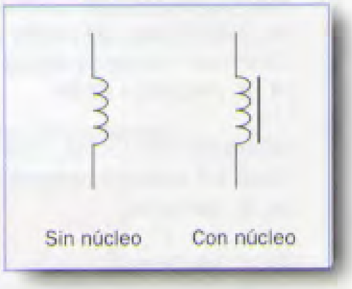

Símbolo ***del inductor en los esquemas eléctricos, con y sin el posible núcleo ferromagnético.***

INERCIA ELÉCTRICA

Aumentando la **corriente** que pasa por un inductor aumenta de intensidad el campo magnético que produce dicha corriente.

Pero este aumento es una **variación** del campo, que a su vez genera una tensión en los extremos del propio inductor ("autoinducción").

La tensión es de signo contrario a la inicial, es decir, se **opone** al aumento de corriente. Recíprocamente, tratando de disminuir la corriente, el campo magnético tiende a reforzarla.

Así, un inductor actúa como un volante, oponiéndose a las **variaciones bruscas** de corriente. La corriente varía en cualquier caso, pero más despacio.

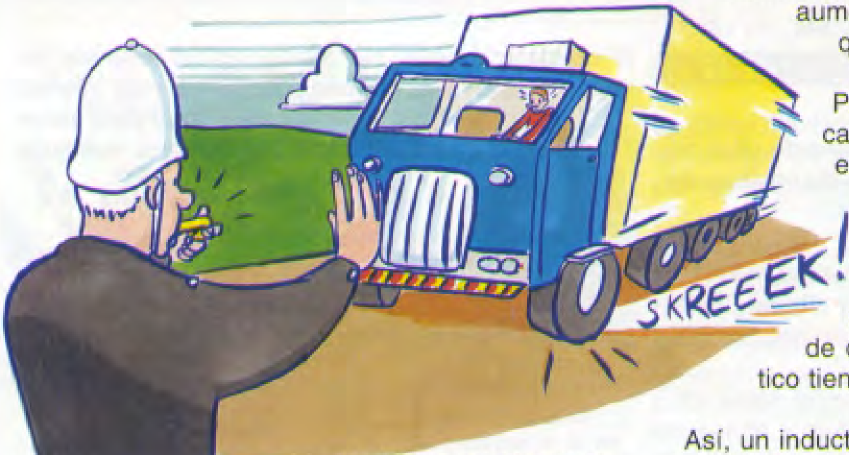

La autoinductancia provoca un efecto similar a la inercia mecánica: un vehículo pesado es difícil de acelerar y difícil de frenar.

UNIDADES DE MEDIDA

La inductancia se mide en **henrios** (símbolo H), que toma su nombre del físico estadounidense Joseph Henry, que descubrió la autoinducción.

Los inductores normales utilizados en electrónica tienen normalmente inductancias medidas con los submúltiplos **milihenrios** (mH) o **microhenrios** (μH).

Se debe observar que cualquier **cable** se comporta como un inductor microscópico: tiene importancia a frecuencias elevadas (por ejemplo, radio), donde se necesita poco para producir efectos significativos.

Las bobinas empleadas en electrónica son generalmente de pequeñas dimensiones, con valores entre 1 μH y 100 mH.

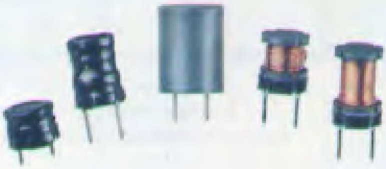

Carga de un inductor

Al igual que los condensadores, también los inductores pueden cargarse... pero de otro modo

Aplicando tensión a un inductor, **inicialmente** no circula corriente alguna: la autoinducción produce una tensión igual y contraria.

Gradualmente, sin embargo, esta tensión se reduce y la corriente puede aumentar **progresivamente**, en teoría hasta el infinito.

Lo que significa que el inductor se **carga**, acumulando energía en forma de campo magnético: esta característica se aprovecha en diversas aplicaciones.

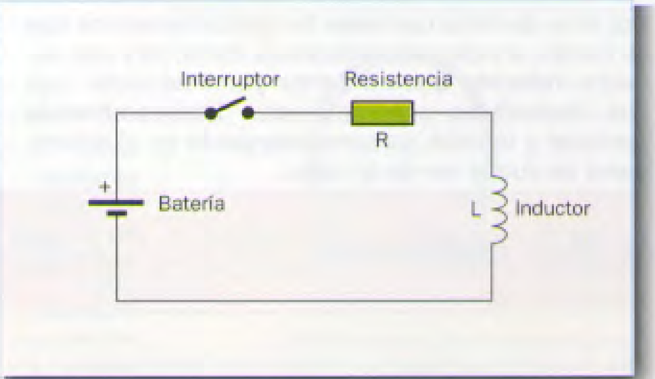

Cerrando el interruptor, el inductor L empieza a cargarse: la corriente que lo recorre va aumentando.

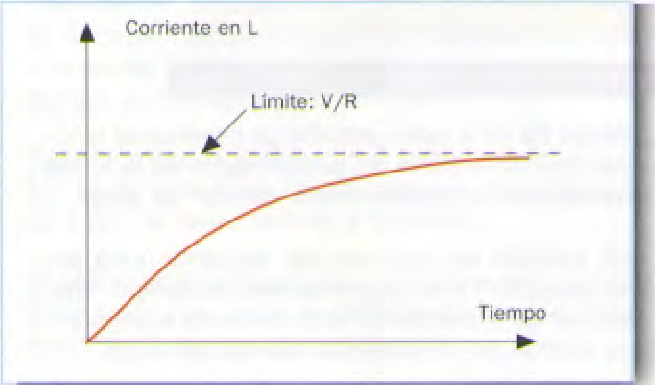

Corriente en un circuito RL en función del tiempo: se debe observar que no interesa la tensión en el inductor, sino la corriente que lo atraviesa.

CURVA DE CARGA

Como un circuito de **resistencia** cero no existe, consideremos una situación realista como la mostrada anteriormente: un inductor se carga mediante una resistencia.

La corriente tendería a aumentar linealmente, pero se ve **limitada** gradualmente por la resistencia, como se observa en la figura adjunta.

Obviamente, no puede superar el máximo permitido por la ley de Ohm, es decir, V / R, donde R puede también ser la resistencia del mismo **cable** con el que está enrollado el inductor.

FILTROS LR Y OTROS USOS

Como un inductor se opone a las variaciones de la tensión aplicada, y por tanto de la corriente que lo recorre, su comportamiento cambia con la **frecuencia**.

A bajas frecuencias las variaciones son lentas, por lo que el inductor las **deja pasar** fácilmente, mientras que para altas frecuencias se comporta casi como una **interrupción** del circuito.

Se trata por tanto de un comportamiento opuesto al del condensador: las bobinas se utilizan a menudo para **bloquear** la componente alterna de una señal, especialmente en el campo de la radio.

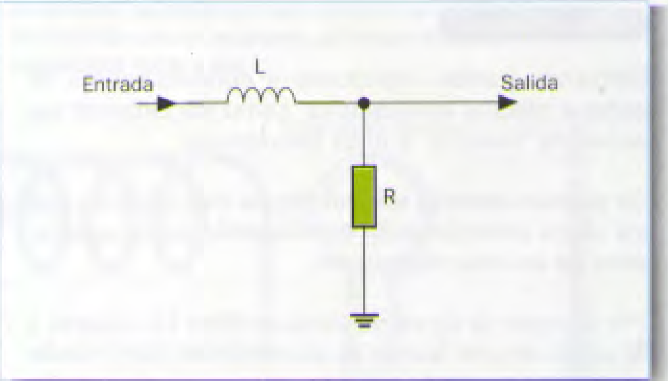

En este filtro LR, la bobina obstaculiza las corrientes que varían rápidamente, es decir, alternas a altas frecuencias: tiende a nivelar la corriente que lo atraviesa.

Bobinas y condensadores

Estos dos dispositivos, aunque bastante distintos entre sí, tienen un comportamiento curiosamente recíproco

Por muy distintos que sean los principios en los que se basan, e incluso su estructura física, hay una estrecha **relación** entre inductor y condensador. Los dos dispositivos son simétricos: **intercambiando** corriente y tensión, su comportamiento es el mismo, como se puede ver en la tabla.

La tabla se refiere a componentes **ideales**; los principios básicos valen de igual forma para los componentes reales empleados en electrónica.

Simetría ***entre bobina y condensador, intercambiando tensión y corriente.***

Condensador	Bobina
Acumula energía como campo eléctrico	Acumula energía como campo magnético
Con una corriente constante, la tensión sube linealmente	Con una tensión constante, la corriente sube linealmente
Se opone a las variaciones de tensión	Se opone a las variaciones de corriente
Se comporta como aislante para la componente continua	Se comporta como conductor para la componente continua
Se comporta como conductor a altas frecuencias	Se comporta como aislante a altas frecuencias
Corriente excesiva "de cierre" si se pone en cortocircuito	Corriente excesiva "de apertura" si se pone en cortocircuito

Las chispas que se ven cuando se interrumpe un cable son debidas a los inductores presentes en el circuito, o a la inductancia de los cables.

TENSIÓN EXCESIVA EN LA APERTURA

La última fila de la tabla describe un interesante fenómeno, que en el caso del condensador es la fuerte **corriente** que se obtiene descargándolo de golpe.

Si un inductor es recorrido por corriente (está por tanto "cargado") y se abre el circuito, el campo magnético tiende a **mantener** esta corriente aumentando la tensión en los extremos del componente.

Ésta **sube** bruscamente a valores muy elevados, causando a menudo una chispa que impide la interrupción durante un breve tiempo.

CIRCUITOS LC

Utilizando **juntos** inductores y condensadores se obtiene efectos interesantes, como los circuitos capaces de "resonar" a altas frecuencias.

Se pueden además realizar **filtros** más eficaces que los vistos anteriormente, combinando las características de los dos dispositivos.

Por ejemplo, la figura muestra un **filtro LC** situado a la salida de una fuente de alimentación para nivelar la tensión de salida: es más eficaz que un sencillo filtro RC y tiene una caída de tensión menor.

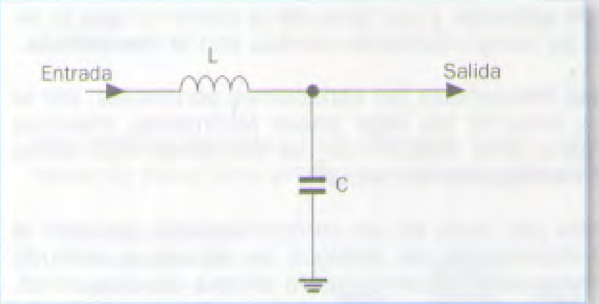

FILTRO LC: ***las irregularidades (componente alterna) son bloqueadas por el inductor y cortocircuitadas a masa por el condensador.***

Transformadores

Se puede transferir energía de un inductor a otro mediante un campo magnético variable

Hemos visto que el paso de corriente por un inductor (bobina o solenoide) produce un **campo magnético**.

También ocurre lo contrario: la **variación** de un campo magnético externo, como el movimiento de un imán, genera tensión en los extremos de la bobina.

El fenómeno (ya citado en la lección 6) se llama **inducción**; es importante recordar que un campo magnético constante no tiene efecto, son las variaciones lo que cuenta.

Inducción *magnética: moviendo un imán se genera una tensión en los extremos de un inductor.*

INDUCCIÓN MUTUA

Cuando una **corriente alterna** circula por una bobina, crea un campo magnético continuamente variable.

Si al lado se encuentra un segundo inductor, el campo magnético variable **induce** en este último una tensión alterna: es un caso de "inducción mutua".

La energía eléctrica se **transforma** así en energía magnética, y después de nuevo en energía eléctrica en la segunda bobina.

La inducción mutua permite* transferir *energía sin contacto eléctrico, aprovechando el campo magnético, siempre que la tensión sea alterna.

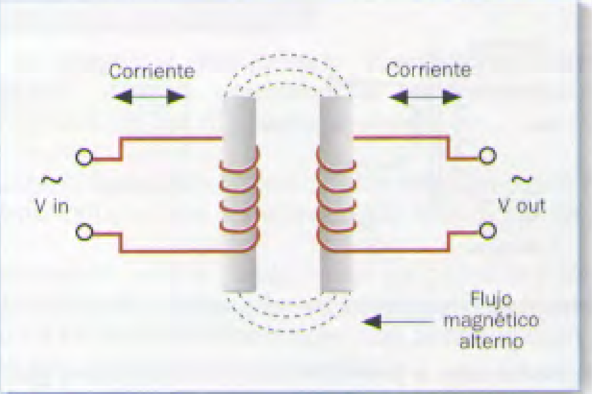

TRANSFORMADORES: DEVANADOS Y NÚCLEO

Para transferir energía de un modo eficaz basta una estrecha **unión magnética** entre las dos bobinas: en teoría, todo el campo magnético producido en la primera debe alcanzar a la segunda.

Esto se consigue utilizando un **núcleo** de material ferromagnético para formar un "circuito magnético" cerrado, de modo que el flujo magnético permanezca en su interior.

En torno al núcleo se encuentran los dos devanados de cable, es decir, los inductores. El dispositivo obtenido se llama **transformador.**

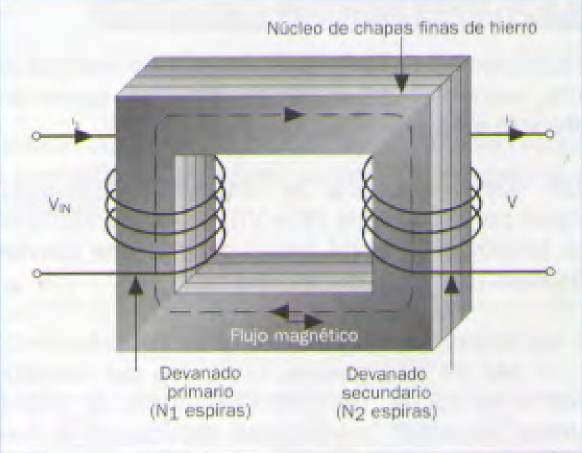
Núcleo de chapas finas de hierro
V_{IN}
V
Flujo magnético
Devanado primario (N_1 espiras)
Devanado secundario (N_2 espiras)

Estructura de un* transformador*: dos inductores enrollados en el mismo núcleo.

Relación de transformación

Un transformador puede subir o bajar la tensión con la misma potencia

Los dos devanados de un transformador se suelen denominar **primario** (al que se aplica la tensión) y **secundario** (del que se toma corriente).

La definición es arbitraria: el transformador es por su naturaleza **reversible**; aplicando tensión al secundario, se recupera en el primario.

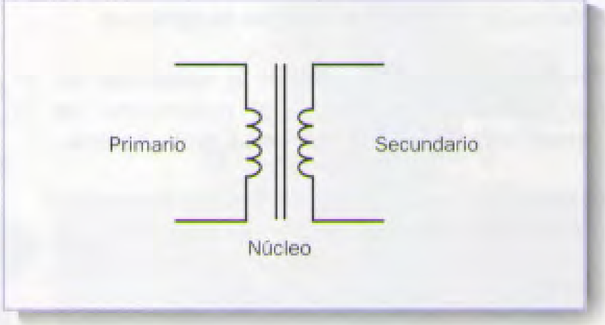

Símbolo de un transformador: las barritas centrales representan el núcleo ferromagnético.

ESPIRAS Y TENSIÓN

Si los dos devanados tienen el mismo número de espiras, la **tensión** que se mide en el secundario es igual a la aplicada al primario. Sin embargo, si el secundario tiene (por ejemplo) **la mitad de espiras** que el primario, en sus extremos habrá la mitad de tensión.

Del mismo modo, si el secundario tiene **tres veces** el número de espiras del primario, nos encontraremos con una tensión igual a tres veces la aplicada.

Un transformador es por tanto capaz de cambiar el valor de una tensión alterna, es decir, **transformarla** para obtener el valor deseado.

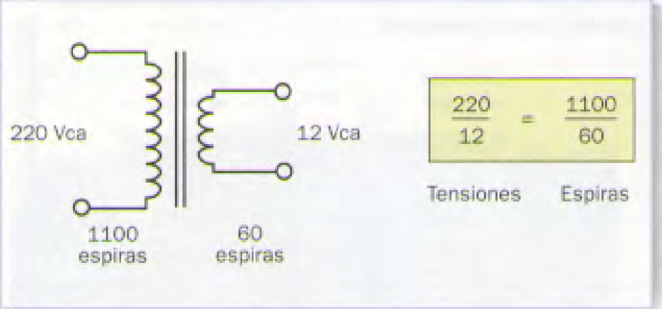

La relación entre la tensión primaria y la secundaria es igual a la relación entre el número de espiras de los dos devanados.

LA POTENCIA PERMANECE CONSTANTE

Naturalmente, como no es posible crear energía de la nada, aumentar la tensión no significa aumentar la **potencia eléctrica**.

Dado que la potencia se obtiene multiplicando la tensión por la corriente (W = VI), si el secundario tiene una tensión alta podrá hacer circular una **corriente** más baja.

Un transformador es por tanto el equivalente eléctrico de un **par de engranajes**, como los del cambio de marcha del coche. También este último, en realidad, permite aumentar la velocidad reduciendo la fuerza (par) y viceversa, pero no cambia la **potencia** del motor.

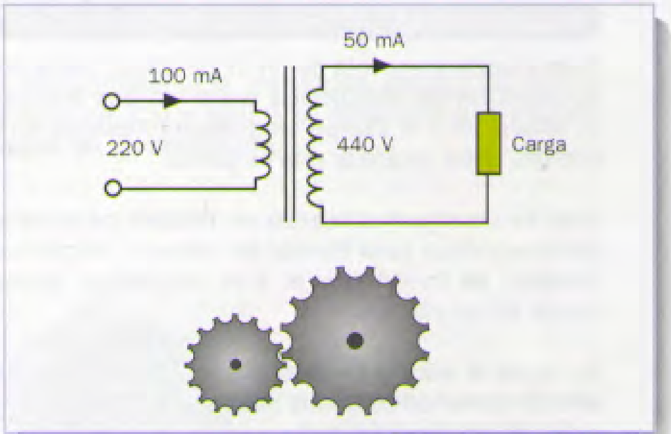

La tensión sube de 220 V a 440 V, pero la potencia es constante: 22 W tanto en el primario como en el secundario, donde la corriente es menor.

Variantes y polaridad

No todos los transformadores están constituidos por dos devanados distintos

Los devanados de un transformador pueden tener **tomas intermedias**; ello ocurre frecuentemente en el secundario para disponer de varias tensiones.

Si por ejemplo un secundario de 500 espiras tiene una toma exactamente en la mitad, es como tener **dos devanados** de 250 espiras en serie.

La tensión estará, por tanto, igualmente **dividida** entre las dos mitades, por ejemplo 6 V en la primera mitad y 6 V en la segunda, dando un total de 12 V en los extremos.

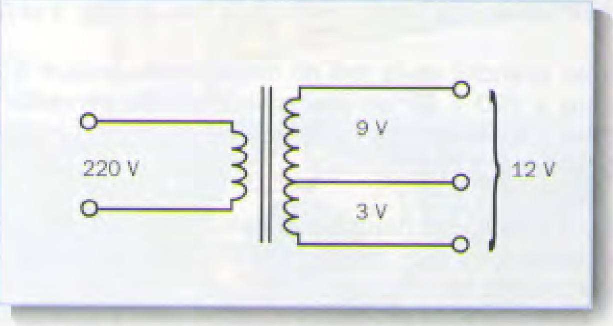

Una toma intermedia ***permite obtener varias tensiones del devanado secundario***

AUTOTRANSFORMADORES

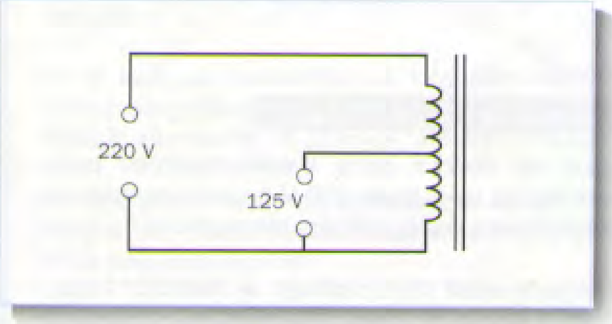

Un autotransformador es un transformador realizado con un solo devanado

No es necesario que primario y secundario estén **aislados**, aunque normalmente es conveniente. Incluso pueden ser el mismo devanado.

El campo magnético variable actúa también sobre el **mismo** inductor que lo ha producido (autoinducción).

El devanado entero puede hacer de primario y una parte de él de secundario, o viceversa (ver figura): esto es un **autotransformador**, que puede utilizarse si no hay problemas de aislamiento.

Por el contrario, nada impide realizar transformadores con **muchos devanados** independientes, por ejemplo para obtener diversas tensiones todas aisladas entre ellas.

POLARIDAD Y FASE

Cuando el sentido de los devanados es importante, se marcan con un **punto** los extremos que tienen la misma polaridad.

Hablar de polaridad con tensión alterna puede parecer extraño, pero en cada **instante** dado la tensión tiene un valor bien definido y de medida exacta.

Se puede también utilizar el término **fase**: en realidad una inversión de polaridad equivale a un desfase de 180 grados.

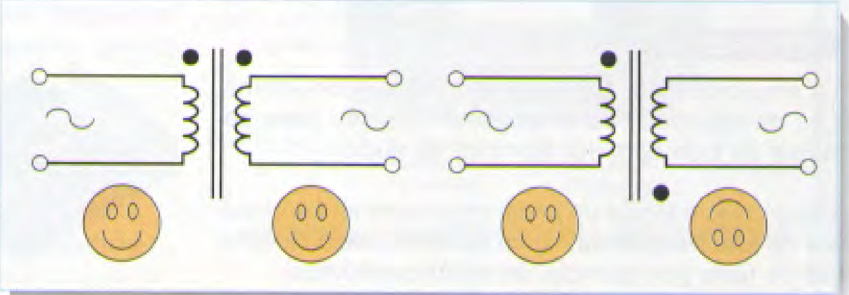

Un transformador puede utilizarse para obtener una tensión alterna* en contrafase, *es decir, desfasada en 180 grados.

Uso de los transformadores

Los transformadores son importantes tanto en electrónica como para la distribución de la energía eléctrica

Una primera aplicación consiste, precisamente, en la **transformación** de la energía eléctrica de una tensión a otra.

Ello es esencial en la red de distribución, porque largos trechos a 220 V serían impracticables: las elevadas corrientes y la resistencia de los cables causarían **pérdidas** excesivas en forma de calor.

Así, se utilizan los transformadores para obtener **altísimas tensiones** (por ejemplo 380 KV) y por tanto corrientes proporcionalmente más bajas.

Cerca de los puntos de utilización, la tensión se **reduce** progresivamente (siempre con transformadores) hasta los 220 V que entran en las casas.

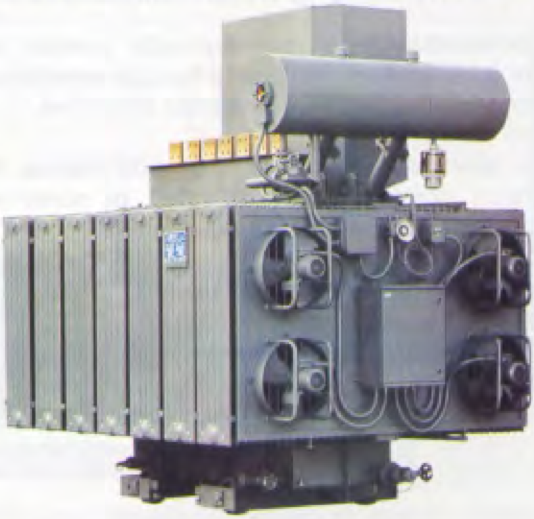

Un transformador de gran potencia, utilizado en las centrales de distribución de la red eléctrica.

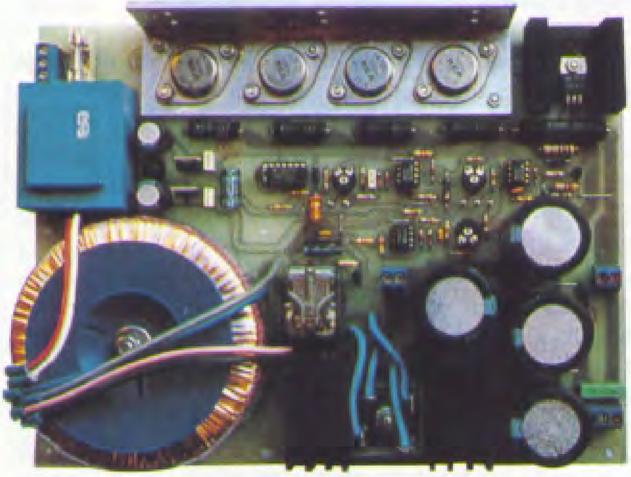

El transformador de alimentación **separa eléctricamente** *los circuitos internos de la red eléctrica, además de transformar la tensión.*

AISLAMIENTO Y ALIMENTACIÓN

En casa se utilizan otros transformadores para **reducir** más la tensión de 220 V a los bajos valores empleados por los dispositivos electrónicos.

Gran parte de estos últimos utilizan al menos un transformador, **externo** y visible (por ejemplo centralitas telefónicas) o **interno**, dentro del propio dispositivo.

Este transformador desarrolla también la importante función de **aislar** el dispositivo de la red eléctrica, evitando descargas y cortocircuitos.

El componente que desarrolla estas dos tareas (reducir de 220 y aislar) se suele llamar **transformador de alimentación**.

OTROS USOS

La sinusoide de la red no es la única tensión alterna. Los transformadores se emplean también para **señales** de todo tipo, por ejemplo, de audio.

Desarrollan además un papel importante en los circuitos de **alta frecuencia**, como transmisores o receptores de radio (por ejemplo, en teléfonos móviles).

Las dimensiones de los transformadores de las fuentes de alimentación conmutadas, se reducen notablemente por la frecuencia empleada (decenas de Khz)

Diodos

Nos encontramos por fin con el primer auténtico componente electrónico: el diodo, que deja pasar corriente en una sola dirección

Un diodo es una vía de sentido único para la corriente, que puede fluir sólo desde el terminal llamado **ánodo (A)** hacia el denominado **cátodo (K)** y no al revés.

El símbolo del diodo, con forma de **flecha**, indica el sentido del recorrido. Como es habitual, por convenio la corriente se desplaza del polo positivo al negativo.

Aplicando tensión **inversa**, es decir, con el positivo en el cátodo, el diodo se comporta como un aislante: no pasa nada de corriente (al menos en teoría).

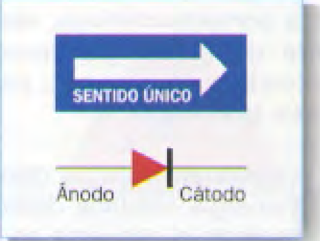

Símbolo *del diodo; la corriente puede pasar sólo en el sentido de la flecha.*

COMPONENTES NO LINEALES

Aplicando una tensión alterna **sinusoidal** a un circuito compuesto por resistencias, inductores y condensadores, todas las corrientes del circuito serán sinusoidales.

Los tres componentes citados se llaman por ello **lineales**, al menos en su comportamiento ideal. Los diodos, sin embargo, son no lineales, como muchos otros dispositivos electrónicos.

Esto no es un defecto, sino justamente la característica que los hace **útiles** en un gran número de aplicaciones.

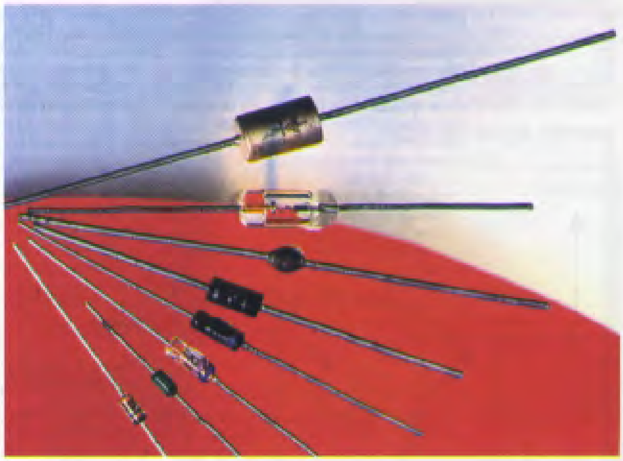

Varios modelos de diodos, la banda indica el cátodo, como la placa en el modelo transparente

RECTIFICADOR

El primer uso del diodo consiste en **rectificar** la corriente alterna, dejando pasar solamente la mitad (una semionda) positiva.

La figura muestra cómo el diodo **bloquea** la semionda negativa, obteniendo a la salida una corriente que, cuando pasa, se desplaza siempre en el mismo sentido. En la mitad del tiempo en que la tensión a la entrada del rectificador es negativa, el diodo se **polariza inversamente** (es decir, con el positivo en el cátodo) y no pasa corriente alguna.

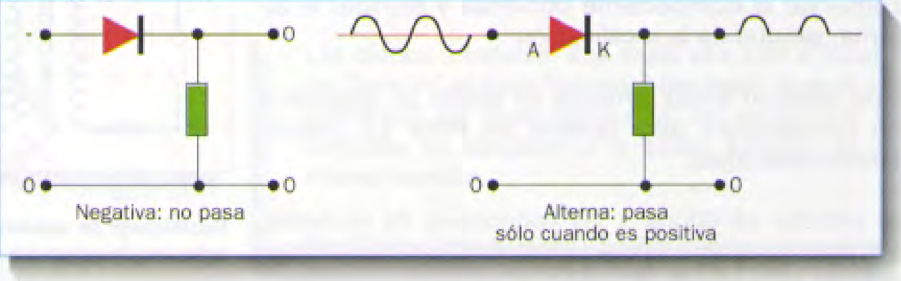

Un rectificador de diodo deja pasar sólo la semionda positiva; durante la negativa la tensión a la salida es cero.

Alimentadores

Filtrando la salida de un rectificador de diodos, se puede obtener una corriente casi continua

El circuito rectificador de la página anterior hace pasar por la resistencia una corriente sólo **parcialmente** continua, a veces definida "pulsante".

Para obtener una auténtica corriente continua, sería necesario **acumular** parte de la energía proporcionada cuando el diodo conduce (deja pasar), para disponer de ella cuando esté bloqueado.

Esto puede hacerse con un **condensador**, el dispositivo ideal para conservar energía eléctrica durante breves períodos; la figura muestra el esquema de un rectificador "de semionda".

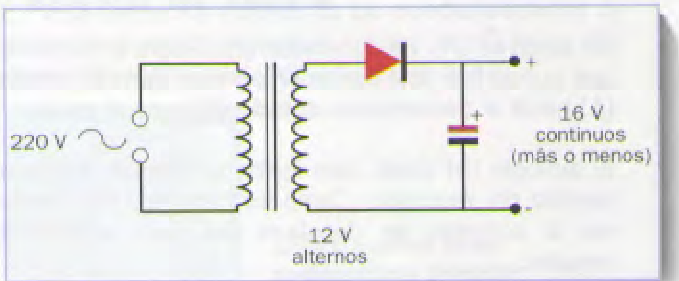

El condensador funciona como depósito para nivelar las ondulaciones de la tensión a la salida.

UNA RESERVA DE CORRIENTE

El condensador funciona como un lago alimentado por un torrente: durante la crecida el lago se llena; cuando el torrente está seco el **lago** sigue proporcionando agua al valle.

Durante la semionda positiva, la corriente que pasa por el diodo **carga** rápidamente el condensador; durante la negativa es este último el que proporciona corriente en la carga.

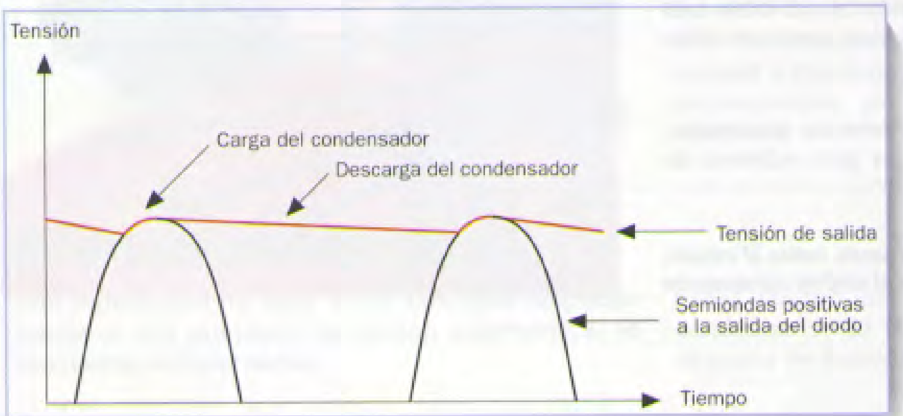

Como el condensador se carga con los **picos**, la tensión máxima a la salida es más o menos 1,41 veces la eficaz (rms) de la alterna a la entrada (menos aproximadamente 1 V para el diodo, como veremos).

En los picos positivos el condensador se carga, mientras que en el resto del tiempo se descarga proporcionando corriente en la carga, es decir, para su utilización.

FILTRAR MEJOR

Nivelar la salida significa esencialmente filtrarla para conservar la **componente continua** y eliminar la alterna, residuo de la rectificación.

Una solución eficaz consiste en añadir un inductor y un condensador para obtener un **filtro LC** (inductancia-capacidad).

El inductor se opone a las variaciones de corriente, actuando también como acumulador de energía e **integrando** eficazmente la misión de los condensadores.

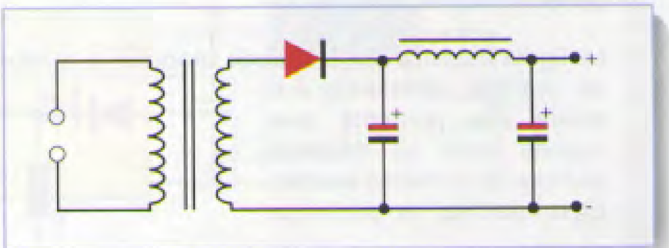

Rectificador de semionda con filtro LC: el efecto combinado de inductor y condensadores permite obtener una corriente casi continua.

Rectificadores de onda completa

Una mejora del circuito consiste en aprovechar ambas semiondas

Durante la **semionda negativa** de la tensión alterna en la entrada, por el diodo del circuito de la página anterior no pasa corriente.

Ello significa que el **transformador** se aprovecha sólo a la mitad; además, por sus devanados pasa una componente continua, cosa que no interesa demasiado.

Esta situación, similar a la de un camión que vuelve vacío tras cada entrega, puede resolverse **modificando** el circuito del rectificador.

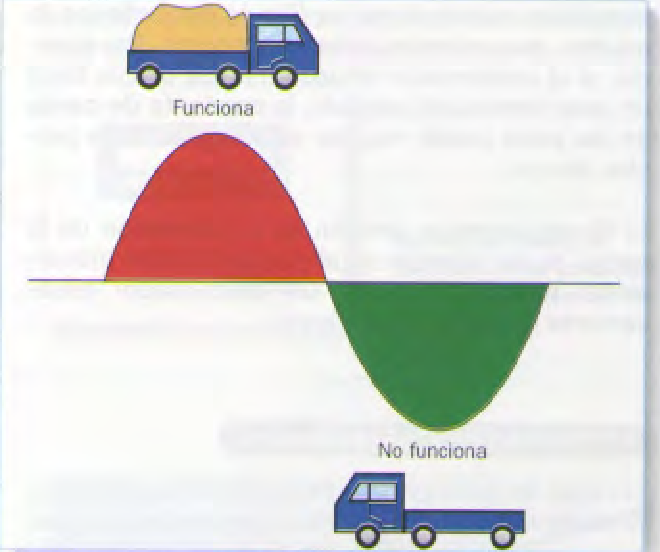

El rectificador de semionda funciona solamente durante la semionda positiva

TOMA CENTRAL

Utilizando un transformador con el doble de espiras y una toma en el centro del devanado, y dos diodos, se obtiene el rectificador de **doble semionda** mostrado en la figura.

Durante una semionda funcionan la mitad superior del devanado y el diodo D1, durante la semionda **opuesta** funcionan la mitad inferior y el diodo D2.

El condensador se carga así **dos veces** por ciclo; entre otras ventajas, la interferencia a la salida es de 100 Hz en lugar de 50, lo que facilita su filtrado. La toma central del transformador constituye el punto de **referencia**, es decir la masa, con respecto al cual "giran" los dos extremos.

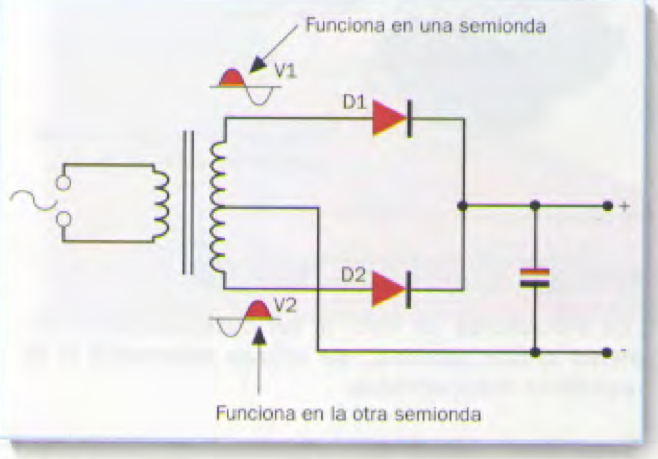

***Rectificador de doble semionda:** cuando V1 es positiva funciona D1, pero cuando V1 es negativa es positiva sin embargo V2 y, por lo tanto, funciona D2.*

PUENTE DE DIODOS

Se puede evitar el doble devanado utilizando cuatro diodos **de puente**: dos conducirán durante la semionda positiva, los otros dos durante la negativa; es la técnica más difundida.

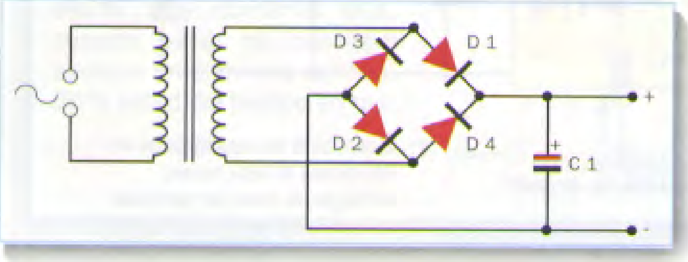

Los diodos conducen **dos cada vez**, con el efecto de "invertir" el transformador (es decir, invertir su polaridad) en cada semionda, de modo que la corriente se desplaza a la salida siempre en el mismo sentido.

***Rectificador de onda entera:** durante la semionda positiva conducen D1 y D2, durante la negativa D3 y D4.*

Alimentadores en la práctica

Algunas consideraciones prácticas que el diseñador de un alimentador no puede ignorar

Condensadores e inductores de valor infinito permitirían teóricamente un filtrado **perfecto**; en la práctica, naturalmente, esto es imposible. Por ejemplo, si el condensador situado tras los diodos tiene un valor demasiado elevado, la **corriente de carga** en los picos puede resultar excesiva para los propios diodos.

El filtrado depende también de la **absorción** de la carga: si por ejemplo el alimentador debe proporcionar pocos mA, basta un condensador relativamente pequeño (y viceversa).

En la práctica, los valores de los **condensadores** de filtro varían desde pocos µF a miles de µF, estos últimos empleados para corrientes elevadas.

Las dimensiones de los condensadores dependen de la corriente que el alimentador debe proporcionar y del grado de filtrado necesario.

FILTROS ANTI-INTERFERENCIAS

Un buen alimentador debe bloquear las interferencias procedentes de la red y no debe **generarlas** a su vez; se utilizan por tanto filtros adecuados entre la red y el primario.

Este enchufe incorpora un filtro anti-interferencias de tipo LC.

Como deben filtrar **frecuencias elevadas** y no los 50 Hz de la red, sus componentes son de pequeñas dimensiones. En general, se añaden también **pequeños** condensadores (por ejemplo de 100 pF) en paralelo a los diodos, con el fin de que no produzcan interferencias de radio en cada cambio de semionda.

ALIMENTADORES ESTABILIZADOS

Los **inductores** de filtro, al ser costosos (especialmente si son grandes), se utilizan solamente si es realmente indispensable.

Se prefieren normalmente los **alimentadores estabilizados**, que permiten obtener una tensión estable e independiente de las variaciones de red.

Estos utilizan un circuito de regulación entre el condensador de filtro y la carga: los estudiaremos después de haber introducido otros componentes.

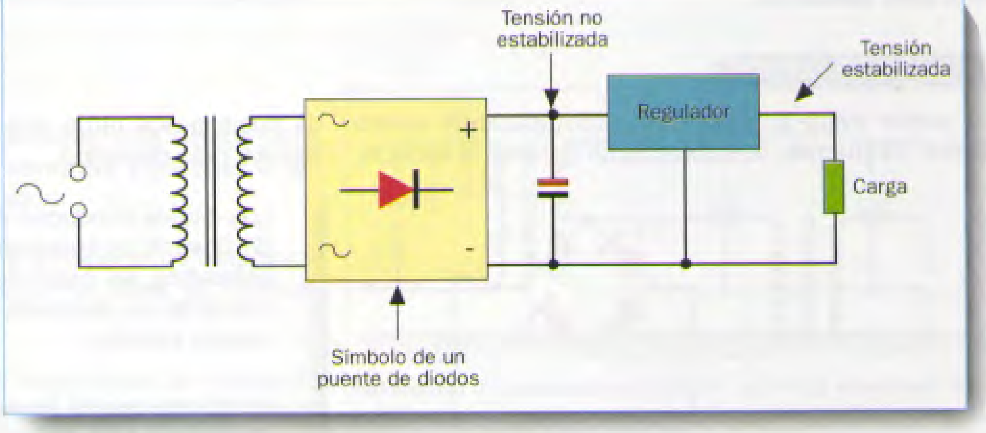

Un alimentador estabilizado proporciona una tensión constante, independientemente de las variaciones de red y de la corriente pedida por la carga.

Semiconductores

La microelectrónica ha sido posible gracias a las peculiares características de estos materiales

La miniaturización de los dispositivos electrónicos deriva de su construcción: no están formados por varios fragmentos, sino por un único cristal de silicio u otro material **semiconductor**. Los semiconductores, como el silicio o el germanio, son sustancias **a mitad de camino** entre los conductores (por ejemplo el cobre) y los aislantes (cerámica). Simplemente no ofrecen **resistencia** al paso de la corriente: en estos materiales se producen complejos fenómenos físicos, que ofrecen oportunidades especiales.

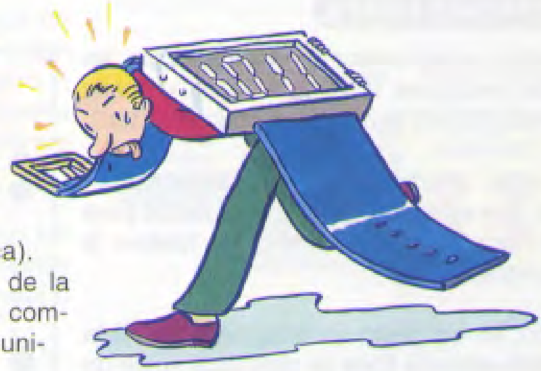

Si se utilizaran tubos de vacío (válvulas), un reloj de bolsillo ocuparía un armario y consumiría tanto como una vivienda

UNA RETÍCULA PERFECTA

Los materiales están formados por **átomos**, que en los sólidos cristalinos están dispuestos de un modo regular y repetitivo.

Los átomos contienen un cierto número de **protones**, partículas cargadas positivamente pero inmóviles en el núcleo del propio átomo. Otros tantos **electrones** con carga negativa giran en torno al núcleo y, entre otras cosas, permiten al átomo unirse más o menos estrechamente a los situados cerca de él.

En los semiconductores, cada átomo está **unido** a cuatro átomos adyacentes, en una estructura regular y bastante estable.

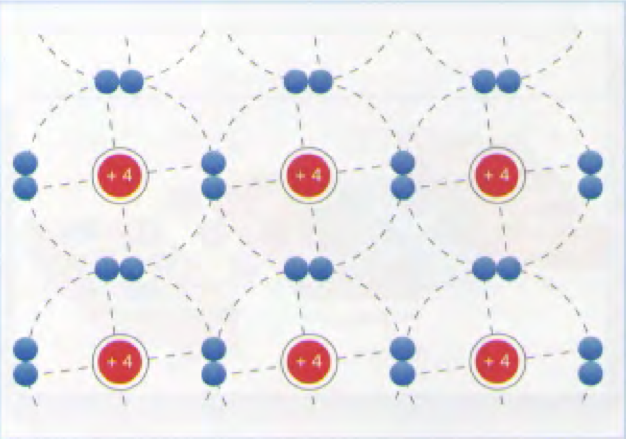

En un cristal de silicio los átomos son tetravalentes, es decir, forman cuatro enlaces con los situados en las proximidades.

ELECTRONES Y HUECOS

Mientras en los metales los electrones son relativamente **libres** de moverse, y pueden por tanto formar una corriente, en los semiconductores están ocupados uniendo los átomos.

Por efecto de la temperatura, sin embargo, de vez en cuando un electrón abandona su puesto y se va "a dar una vuelta", dejando un **hueco** positivo.

Como estas cargas libres pueden transportar una corriente se les llama **portadoras.**

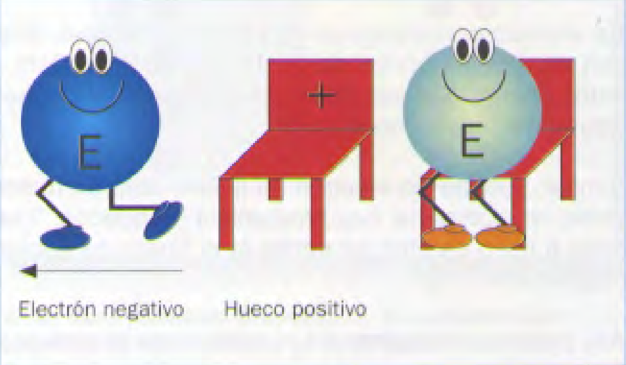

Electrón negativo

Hueco positivo

Dopaje

Pequeñas impurezas en la estructura cristalina cambian el comportamiento de los semiconductores

AÑADIR ELECTRONES LIBRES

Arrancar un electrón de un enlace requiere **energía**; en consecuencia, los electrones libres son pocos y puede pasar poca corriente (de ahí el nombre "semiconductor").

El discurso cambia añadiendo algún átomo pentavalente, es decir, con **cinco** electrones dispuestos para formar enlaces (por ejemplo, antimonio, fósforo o arsénico, los "donantes").

El quinto electrón, en un cierto sentido, no tiene nada que hacer: es relativamente **libre** de desplazarse de un átomo al otro, haciendo de portador de corriente.

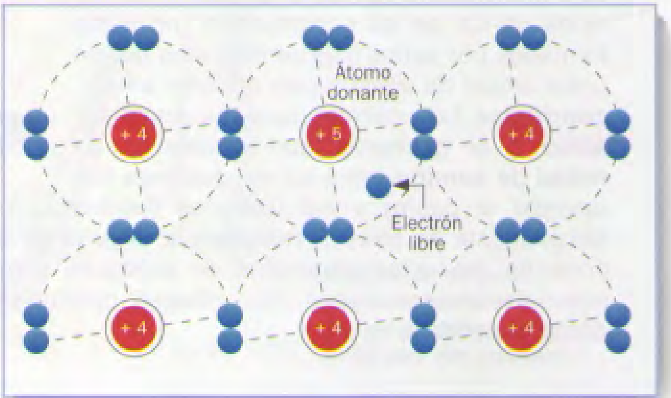

Un átomo pentavalente tiene un electrón desocupado, relativamente libre de circular por el material

DOPAJE DE TIPO N Y P

El añadido de impurezas controladas se llama "dopaje". Cuando deja disponibles portadores **negativos** (electrones), se denomina de tipo N.

Otros materiales (por ejemplo, boro, galio e indio, los "receptores") son trivalentes, es decir, tienen un electrón disponible de menos: proporcionan lagunas **positivas**. Se llaman de tipo P.

Así, tenemos **tres** materiales distintos: el silicio puro (casi aislante), el silicio dopado de tipo N y el de tipo P.

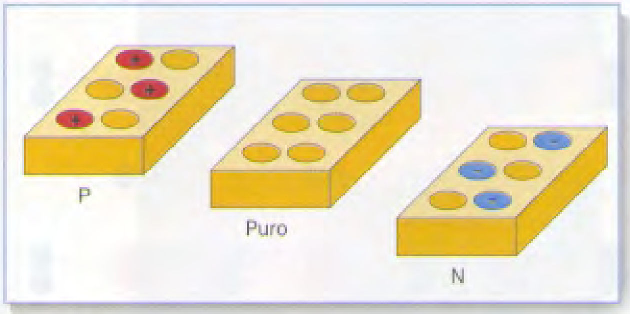

El silicio dopado N y el dopado P son los ladrillos fundamentales en la electrónica; el silicio puro sirve de base.

TAMBIÉN LOS HUECOS SE MUEVEN

En el silicio con dopaje de tipo N los portadores libres son electrones, en los de tipo P son huecos. Pero si estos últimos son sencillas **ausencias** de electrones, ¿qué hacen para moverse?

Simple: cuando un electrón se libera, deja un hueco; como en el material hay predominio de huecos, "cae" poco a poco en otro: es **como si** el hueco se hubiera desplazado.

Así, podemos **imaginar** los huecos como cargas positivas móviles.

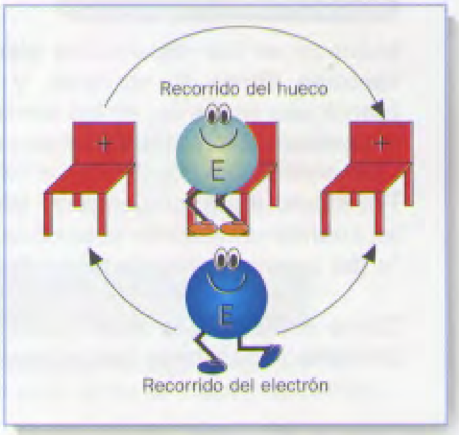

Cuando un electrón cambia de sitio, es como si un sitio libre (hueco) se desplazase en la dirección opuesta

El diodo de unión

Una unión entre las regiones P y N puede ser recorrido por la corriente en un único sentido

Un diodo se consigue dopando un mismo cristal de silicio de dos maneras diferentes: por una parte con impurezas de tipo N y por otra con impurezas de tipo P. Los dos extremos del cristal forman respectivamente el cátodo (N) y el ánodo (P). La zona de contacto entre las dos partes se llama unión PN.

Muchos componentes electrónicos activos, como los transistores, están basados en una o varias uniones entre regiones (áreas) de silicio de tipo N y P, más o menos dopadas

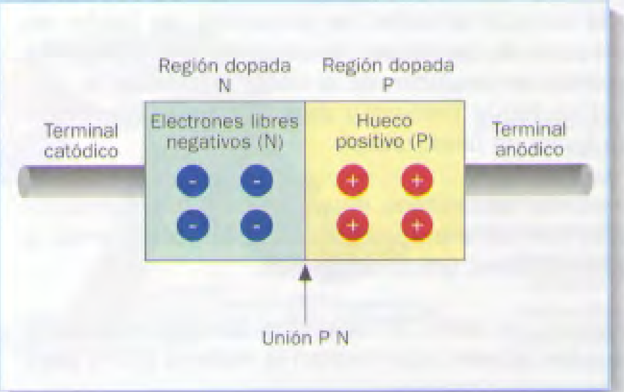
Región dopada N
Región dopada P
Terminal catódico
Electrones libres negativos (N)
Hueco positivo (P)
Terminal anódico
Unión P N

Un diodo de unión consiste en dos zonas adyacentes de silicio, con dopaje opuesto.

EL DIODO DE CONDUCCIÓN

Polarizando el diodo de modo **directo**, es decir, con el positivo sobre el ánodo (zona P) y el negativo en el cátodo (zona N), pasa corriente.

De hecho, los electrones de la zona N, empujados por los procedentes del exterior, se dirigen hacia la **unión**. Lo mismo hacen los huecos por el otro lado.

En la unión, los dos tipos de portadores se **encuentran** y se neutralizan recíprocamente (los electrones ocupan los huecos), dejando así sitio para nuevos portadores.

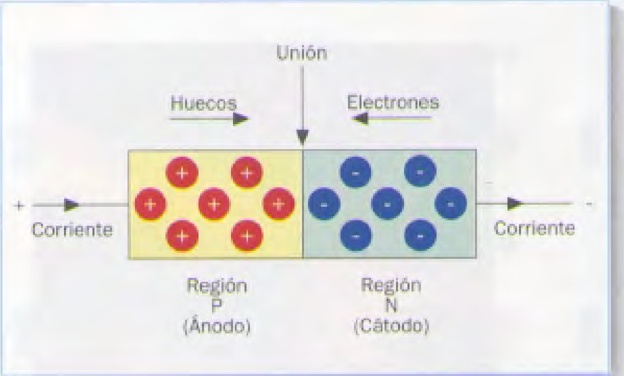
Unión
Huecos
Electrones
+
Corriente
Corriente
-
Región P (Ánodo)
Región N (Cátodo)

Cuando el diodo conduce, en cada una de las dos zonas hay portadores en abundancia: electrones en la N y huecos en la P

POLARIZACIÓN INVERSA: TODOS PARADOS

Invirtiendo la polaridad de la tensión aplicada, los portadores son atraídos hacia los **extremos** y en la zona central no queda ninguna carga dispuesta a moverse.

La corriente es prácticamente **nula**, excepto una pequeña pérdida debida a las parejas hueco-electrones creadas por efecto térmico.

Es como si fuera un bloqueo, que en realidad toma el nombre de **barrera de potencial**: las cargas no pueden atravesarla.

Naturalmente, hemos simplificado un poco; los fenómenos físicos reales son un poco más **complejos**, pero el principio básico es el descrito.

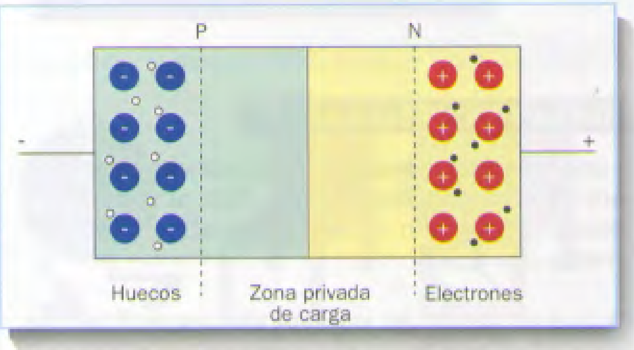
P
N
-
+
Huecos
Zona privada de carga
Electrones

En un diodo polarizado* inversamente*, los portadores se acumulan en los extremos y no están disponibles para transportar corriente.

Producción de los semiconductores

Los diodos y los demás componentes electrónicos vienen... de las piedras

El silicio es uno de los componentes más habituales de la corteza terrestre: se encuentra de hecho en gran parte de las **rocas**. No obstante, los fenómenos descritos se producen en el silicio **cristalino**: el material se funde primero y después enfría, de modo que forma un único cristal.

Como los fenómenos electrónicos están condicionados por las impurezas, el silicio se **refina** antes a un elevadísimo grado de pureza.

Después se corta en delgadas láminas circulares llamadas **obleas**, que forman la materia prima para los sucesivos procesos.

El silicio usado en electrónica procede de gruesos cristales purificados con precisión.

REALIZACIÓN DE LAS REGIONES P Y N

Para fabricar un diodo se puede partir de una **base** de silicio ya dopado, por ejemplo de tipo P, y depositarla sobre una fina capa de silicio de tipo N (proceso epitasial).

Otros semiconductores se producen haciendo que las impurezas se **difundan** en los puntos deseados de la capa básica (proceso por difusión).

En todo caso es necesario una **precisión** muy elevada, porque la concentración de los dopantes es fundamental para el buen funcionamiento de los dispositivos producidos.

Difusión *de las impurezas dopantes en las obleas de silicio (colgadas arriba).*

PRUEBA Y ENCAPSULAMIENTO

En una sola oblea encuentran su sitio **un gran número** de componentes idénticos que, al término del proceso, se dividen y son controlados automáticamente uno a uno.

Los que no entran en las espirales específicas, se descartan; el porcentaje de componentes buenos se llama **rendimiento** de la producción.

Encapsulados por separado en contenedores especiales de vidrio, resina, cerámica, y conectados a los terminales, pueden quedar finalmente sometidos a una posterior prueba final.

Un diodo común es el resultado de un largo, complejo y delicado proceso. El bajo coste deriva del elevado rendimiento de la producción.

Transistores

Estos versátiles dispositivos son la base de toda la electrónica moderna

El **premio Nobel** de física otorgado en 1956 a Bardeen, Brattain y Shockley fue muy merecido.

Su invención del **transistor de unión** (BJT: bipolar junction transistor) hizo posible la miniaturización de los dispositivos electrónicos y, posteriormente, de los circuitos integrados.

Hasta ese momento los dispositivos electrónicos se llevaban a cabo con los incómodos tubos de vacío, o **válvulas** termoiónicas, cuya miniaturización era bastante problemática.

El primer transistor *fue creado por el grupo de Shockley.*

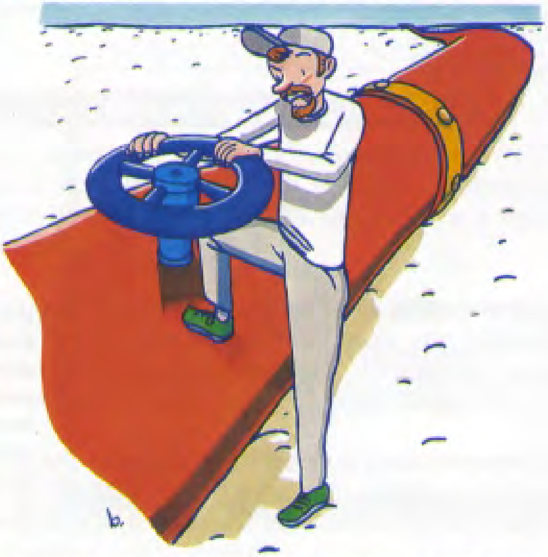

Un transistor permite controlar *una corriente elevada con un pequeño esfuerzo, es decir, con una corriente muy inferior.*

COMPONENTES ACTIVOS

A diferencia de los componentes "pasivos" hasta ahora tratados, el transistor es un componente activo, es decir, puede **controlar** una corriente.

Es esencialmente una especie de **válvula de regulación**, en la que una pequeña corriente controla otra más elevada.

También los **relés** (sección Digital de la lección 3) controlan una corriente elevada con una débil, pero son dispositivos electromecánicos: relativamente voluminosos, lentos y necesitados de corriente.

Además, el transistor no se limita a abrir y cerrar el paso, sino que puede **variar** de una manera continua (analógica) la corriente que lo atraviesa, un poco como una resistencia variable.

TIPOS DE TRANSISTORES

Además de los transistores de unión, existe entre otro tipo: los transistores de **efecto de campo** (FET: Field-Effect Transistor).

Ambos tipos se utilizan con mucha frecuencia, los primeros (BJT) sobre todo en electrónica analógica y los segundos (FET) especialmente en los circuitos integrados digitales y en los controles de potencia.

Hay también otros dos dispositivos electrónicos más o menos especializados (como el diodo controlado, SCR), pero prácticamente **derivan** todos ellos del transistor.

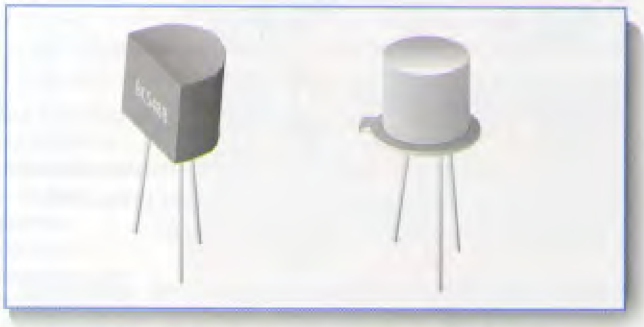

Encapsulados habituales de transistores *de baja potencia.*

Colector, base, emisor

Los transistores están disponibles en dos polaridades: NPN y PNP

El símbolo del transistor de unión muestra sus **tres terminales**, cuyos nombres (colector, base, emisor) derivan de la construcción y del funcionamiento físico.

Los transistores más utilizados son del tipo **NPN**, normalmente empleados en circuitos con alimentación positiva con respecto a masa.

La **flecha** indica, como en los diodos, el sentido convencional de la corriente (del positivo al negativo) que recorre el terminal emisor durante el funcionamiento normal.

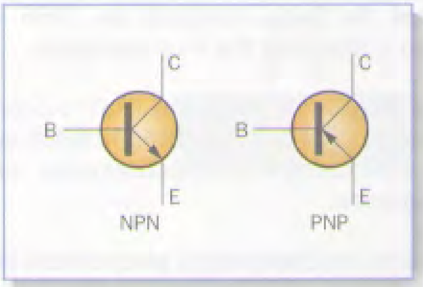

Símbolos de los transistores NPN y PNP: C=colector, B=base, E=emisor.

CONTROLAR UNA CORRIENTE

Los transistores se utilizan habitualmente en la denominada configuración de **emisor común**, simplificada en la figura.

En este caso, la **corriente de base** (de la base al emisor) controla la corriente de colector (del colector al emisor), que es la que enciende la bombilla.

Si no pasa corriente de la base, no pasa tampoco del colector y la bombilla permanece apagada: el transistor está en **corte**, es decir, abierto.

Una pequeña corriente de base lo pasa a **conducción**: entre colector y emisor pasa corriente, como si existiera una resistencia de bajo valor.

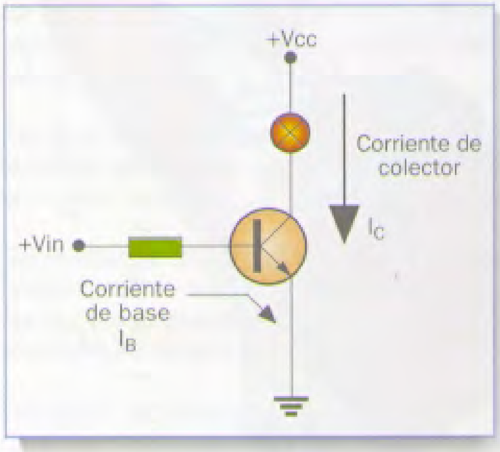

Una débil corriente de base I_B controla la fuerte corriente de colector I_C.

GANANCIA

Si una corriente de base de 1 mA ocasiona el paso de una corriente de colector de 150 mA, se dice que el transistor tiene una **ganancia** de 150 veces.

Obviamente no se **crea** corriente de la nada: la energía viene proporcionada por la pila o por la fuente de alimentación.

El transistor actúa como el **servofreno** del automóvil, que permite aplicar una fuerte presión a los frenos con un ligero esfuerzo del pie.

La ganancia puede aprovecharse, como veremos, para **amplificar** (aumentar el nivel) tensiones alternas como las señales de audio.

La recepción y transformación en imágenes de las transmisiones de televisión aprovecha la amplificación de los transistores.

Interruptores de transistores

El transistor puede utilizarse para abrir o cerrar un circuito

Si la corriente de base es lo bastante elevada, el transistor se **satura**: el colector está casi conectado con el emisor.

Por tanto, en una primera aproximación la conexión colector-emisor puede verse como un interruptor **cerrado**.

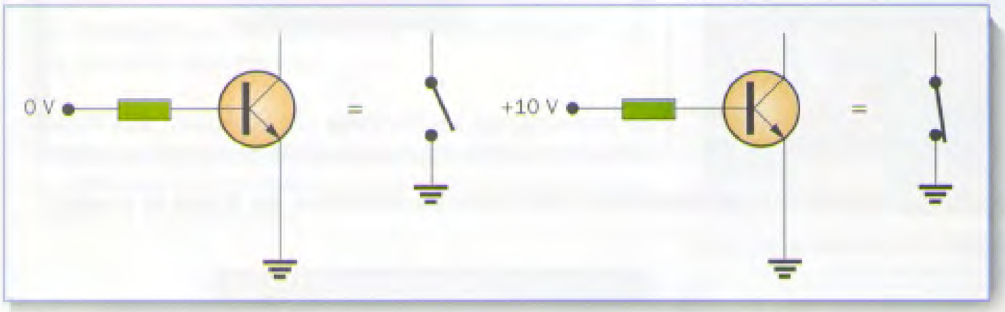

Alternando así entre corte y saturación, el transistor puede utilizarse como si fuese un **interruptor**.

Los dos estados extremos del transistor (corte y saturación) corresponden a un interruptor abierto o cerrado.

INVERSIÓN LÓGICA CON UN TRANSISTOR

La figura muestra un sencillo **inversor lógico** (NOT) realizado con un transistor y dos resistencias: aplicando tensión en la entrada, la salida va a masa, y viceversa.

En realidad, si a la entrada hay tensión (valor lógico 1), por la resistencia de base R1 pasa una pequeña corriente que envía a **conducción** al transistor, que a su vez conecta la salida a masa (valor lógico 0).

Pero si la entrada está a masa (0) no pasa corriente de base y el transistor está en **corte**: el colector no absorbe corriente y la resistencia de colector tiene la salida alta a V_{cc} (1).

Los primeros **circuitos integrados** de transistores se realizaban de un modo similar a este.

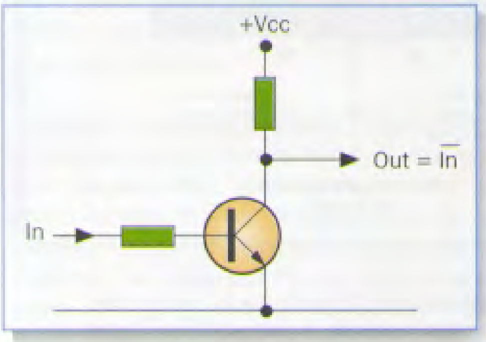

Inversor (NOT) realizado con un transistor: una tensión a la entrada elimina tensión a la salida, y viceversa.

CAÍDAS DE TENSIÓN

La conexión base-emisor se comporta como un **diodo**: significa que hay una caída de tensión (de más o menos 0,65 V).

Si la tensión aplicada es **inferior** a este valor, no pasa corriente por la base y el transistor no conduce.

En saturación, en cambio, el colector no está del todo conectado al emisor: hay una **tensión de saturación** de unas décimas de voltio.

Si el transistor controla corrientes fuertes, se origina un **calentamiento** del componente, cosa que hay que tener en cuenta.

Los transistores de potencia pueden llegar a disipar cientos de vatios y requieren ser enfriados con disipadores especiales.

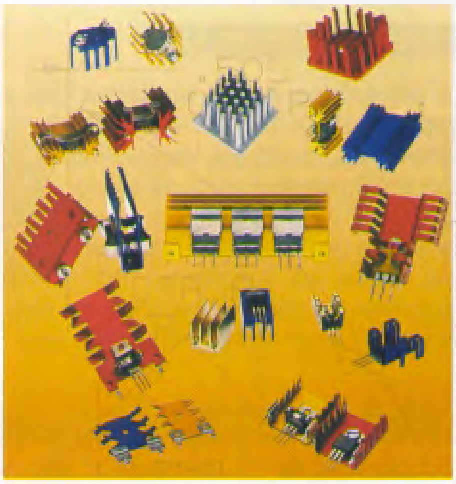

Estructura de un transistor

Como el diodo, el transistor se realiza con un único cristal de silicio

Un transistor habitual BJT se fabrica con **tres regiones**, la central dopada como P y las dos de los extremos como N (en el caso del transistor NPN).

Hay además dos **uniones**: base-emisor y base-colector, cada uno de los cuales es similar al de un diodo normal.

Las tres regiones están conectadas al exterior con los tres terminales C, B, E. La región central, la **base**, es más sutil que las otras dos.

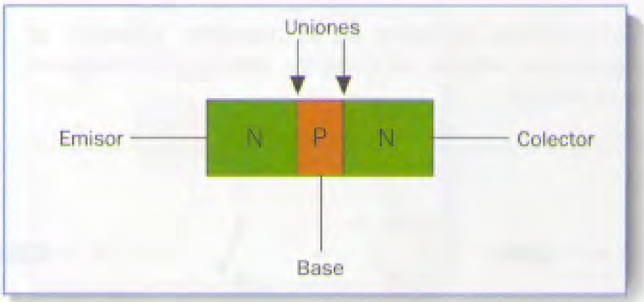
Uniones
Emisor
N
P
N
Colector
Base

Un transistor BJT está formado por dos uniones, con la base en el centro.

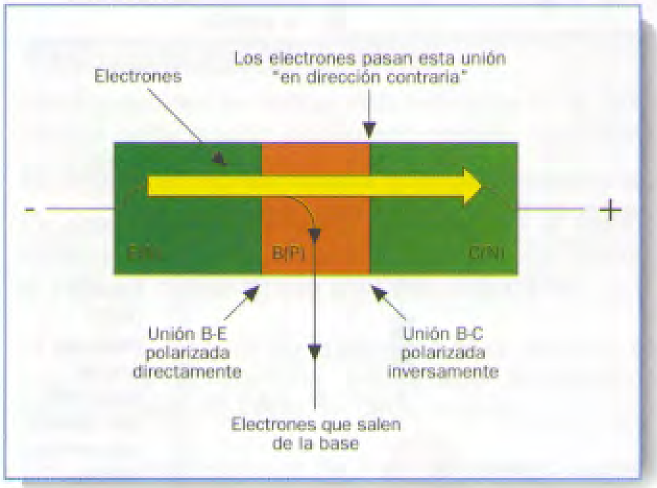
Electrones
Los electrones pasan esta unión "en dirección contraria"
-
+
E(N)
B(P)
C(N)
Unión B-E polarizada directamente
Unión B-C polarizada inversamente
Electrones que salen de la base

La corriente de colector está formada por los portadores procedentes del emisor mediante la unión E-B, que logran atravesar la base.

FUNCIONAMIENTO SIMPLIFICADO

En el funcionamiento normal, la tensión en el colector es siempre **mayor** que en la base, lo que significa que la unión base-colector está inversamente polarizada.

Por el colector no puede entrar **corriente**, exactamente igual que en un diodo en el que se aplique el positivo al cátodo.

Sin embargo, si se hace pasar corriente por el empalme base-emisor, los **portadores** procedentes del emisor (electrones, ya que el emisor es una región N) se dirigen hacia la base.

Deberían recombinarse con los huecos (anulándose como portadores), pero la base es muy **estrecha**, por lo que muchos logran alcanzar el colector, que los "recoge" formando una corriente.

CONSTRUCCIÓN PLANA

La producción de un transistor con los terminales dispuestos en ángulo recto sería **complicada** y, sobre todo, resultaría impracticable realizar más de uno en el mismo cristal de silicio.

Se utiliza por tanto una disposición **plana**, es decir, con todos los terminales en la misma cara; el transistor se obtiene dopando de distinta forma los diversos estratos del cristal.

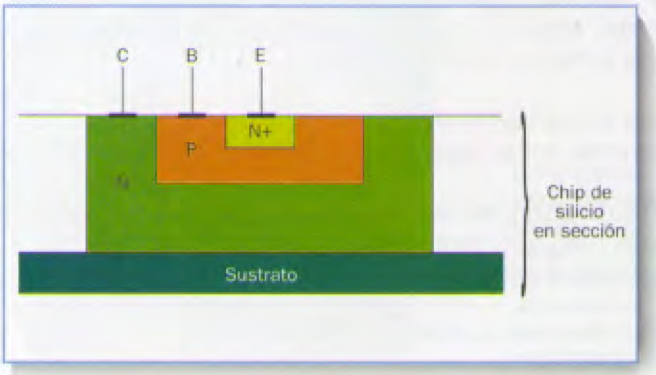
C
B
E
N+
P
N
Sustrato
Chip de silicio en sección

Sección de un transistor plano: la región de emisor se dopa abundantemente (n+) para proporcionar varios portadores.

Amplificación

Un transistor permite aumentar la amplitud de una tensión alterna

La señal eléctrica procedente de un micrófono es muy débil y requiere ser **amplificada** tanto para poder ser grabada en cinta como (con mayor razón) para controlar un altavoz.

Los **transistores** son adecuados para desarrollar esta tarea, gracias a su capacidad de amplificar las variaciones de la corriente de base.

Pero funcionan con corrientes continuas, mientras que nosotros queremos amplificar una **tensión alterna**; afortunadamente el problema se resuelve fácilmente con la incorporación de algunos componentes.

TENSIÓN Y CORRIENTE

En primer lugar, ¿cómo se cambia una tensión en una corriente? Sencillísimo: con una **resistencia**. La tensión provoca el paso de corriente en la resistencia.

Y, ¿cómo se cambia una corriente en una tensión? ¡También con una resistencia! El paso de la corriente produce **tensión** en sus extremos. Así pues, si podemos realizar un amplificador de corriente, lo podemos transformar fácilmente en un **amplificador de tensión**, añadiendo dos resistencias como se muestra en la figura.

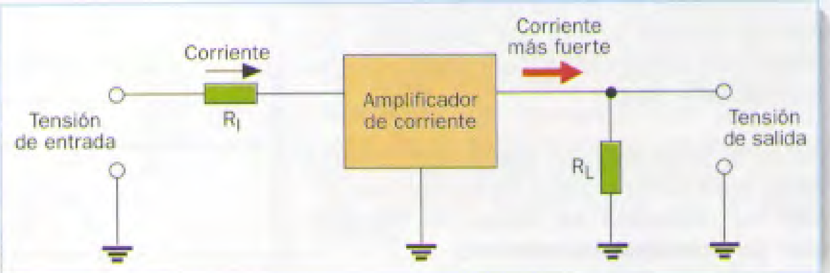

Cómo utilizar un amplificador de corriente para obtener una tensión más amplia de la original.

NO SÓLO UN TRANSISTOR

Aunque el transistor sea propio de un amplificador de corriente, no se puede utilizar **directamente** como amplificador, sin otros componentes.

Esto funciona sólo cuando la corriente de base es positiva (en el caso de un NPN), es decir, lleva la unión base-emisor a **conducción**. Además, es necesario también superar los 0,65 V mínimos de base.

Para señales alternas, es necesario un amplificador que funcione también cuando la tensión (y por tanto la corriente) de entrada sea **negativa**.

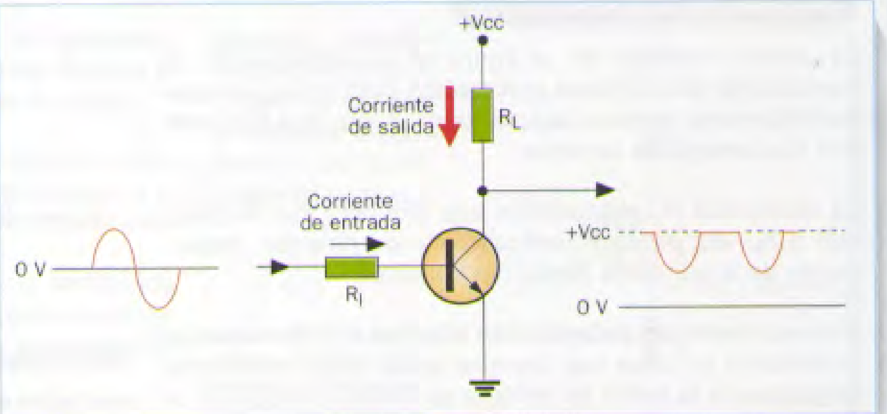

Un transistor así conectado amplificaría sólo las semiondas positivas (observe que la salida tiene signo opuesto).

Polarización

Una corriente continua de fondo resuelve el problema de las semiondas negativas

Para amplificar también tensiones negativas, se puede hacer pasar **siempre** una cierta corriente: la corriente de "polarización".

La débil corriente de entrada se **suma** a esta corriente fija, provocando variaciones de más o de menos, pero sin poner nunca a cero la corriente o hacerla negativa.

También en la salida (es decir, en el colector) habrá siempre corriente, cuyas variaciones serán más amplias que las de entrada, por efecto de la amplificación.

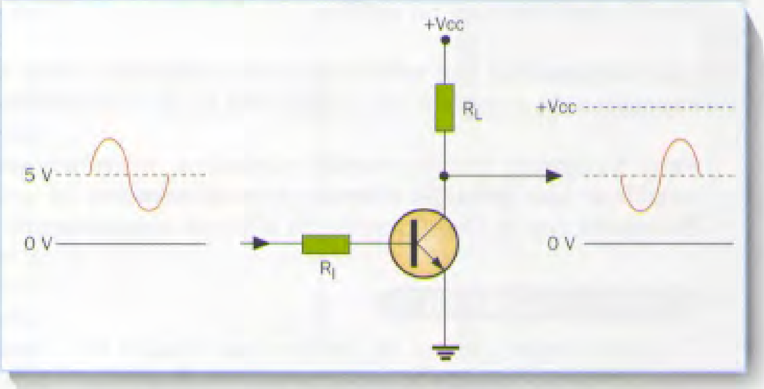

Si la entrada no baja nunca por debajo de 0,65 V, la corriente de base se desplaza siempre en el mismo sentido y el amplificador puede funcionar.

ACOPLAMIENTO CAPACITIVO

La **componente continua** que se añade a la señal de entrada puede ser producida, por ejemplo, con un sencillo divisor resistivo entre alimentación y masa.

Pero de la señal interesa sólo la componente alterna: para evitar el paso de corrientes continuas no deseadas se utiliza un **condensador** de valor bastante elevado.

Esta forma de conexión entre dos circuitos, que aprovecha un condensador para dejar pasar la componente alterna y **bloquear** la continua, toma el nombre de "acoplamiento capacitivo".

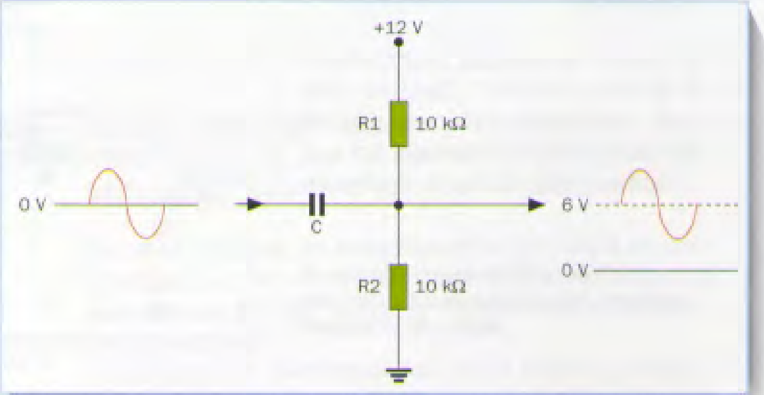

En los dos lados de un* acoplamiento capacitivo *hay la misma tensión alterna, pero la tensión continua es distinta.

UN AMPLIFICADOR PRIMITIVO

El circuito mostrado en la figura es un amplificador de transistores que funciona con **ambas** semiondas, aunque decididamente es poco aceptable en la práctica por motivos que enseguida veremos.

La resistencia R_P proporciona una corriente **fija** de base, que a su vez produce una corriente de colector, dependiente de la ganancia (Beta) del transistor.

Esta **corriente de polarización** añadida por R_P hace que la corriente de base sea siempre positiva: las semiondas negativas de la señal de entrada se limitan a reducirla un poco.

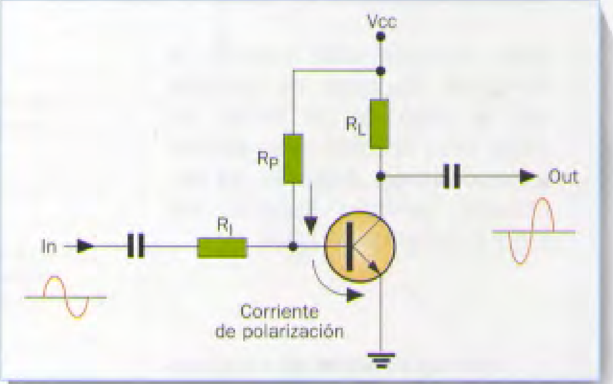

Polarización mínima: ***R_P proporciona una corriente fija de base, también en ausencia de señal.***

Punto de trabajo

Un transistor amplificador debe estar lejos tanto de la saturación como del corte

Para que un amplificador funcione correctamente, la salida debe ser **libre de moverse** en ambas direcciones.

Si por ejemplo la señal amplificada tiene una **amplitud** de 2 V de pico, debe haber al menos 2 V de margen por arriba y por abajo.

Idealmente, la tensión en el colector debería estar más o menos **a medio camino** entre alimentación y masa, para poderse mover libremente tanto hacia arriba como hacia abajo.

Debe haber espacio para que la tensión de salida pueda aumentar y disminuir, para reproducir correctamente la señal amplificada.

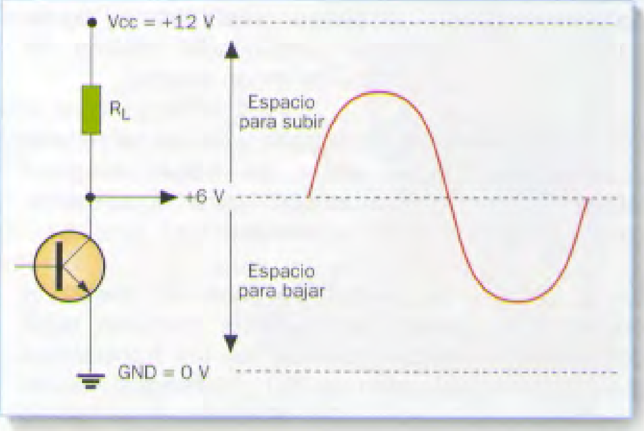

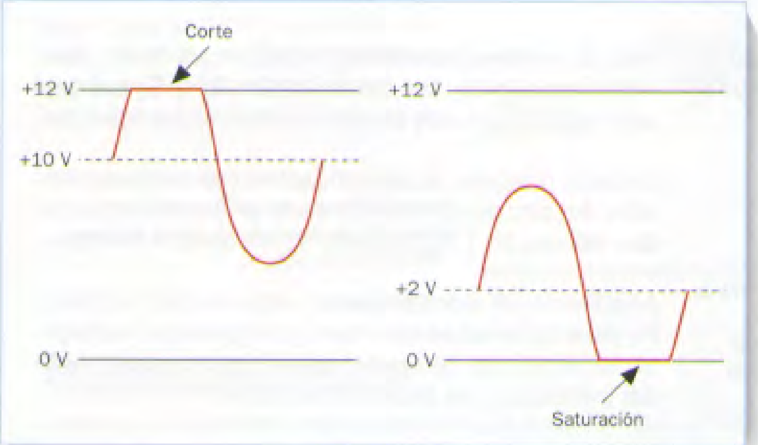

LA GANANCIA ES IRREGULAR

La tensión continua de salida depende de la corriente de colector, que a su vez es la de base multiplicada por la **ganancia de corriente** del transistor. Por desgracia, este último valor, indicado como Beta (ß) o h_{FE}, es extremadamente **variable** tanto de un ejemplar a otro como con la temperatura.

Ello significa que el **punto de trabajo** del transistor (es decir, la corriente de salida) puede desplazarse hacia la saturación o hacia el corte, quitando espacio a la señal.

Una tensión en reposo demasiado alta o demasiado baja no deja espacio para la señal, cuyos picos quedan cortados.

AUTOESTABILIZACIÓN

El mismo transistor puede resolver el problema, añadiendo una resistencia al **emisor** como muestra la figura (lo que hace también superfluo el de base).

Si la tensión de emisor **baja** por debajo de la de base (menos de 0,65 V), el transistor conduce más y la aumenta; si sube, conduce de menos y la disminuye.

Dada una cierta tensión sobre la base, la tensión de emisor es por tanto **estable**, y en consecuencia es estable también la corriente que atraviesa la resistencia R_E.

Pero la misma corriente (salvo la de base, relativamente pequeña) atraviesa también R_L, por lo que es estable la **corriente de colector**, y por tanto la tensión relativa.

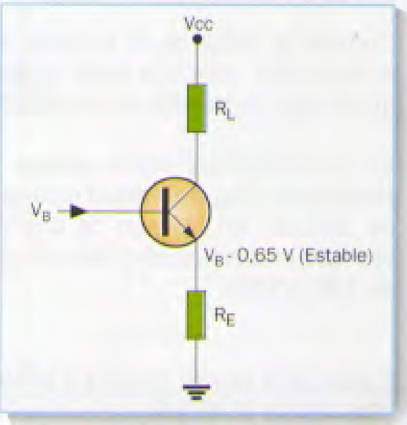

Una resistencia de emisor estabiliza la corriente en la carga, y por tanto la tensión de colector.

Amplificador clásico

El amplificador de emisor común emplea cuatro resistencias

La tensión de polarización de base V_B se obtiene normalmente con un **divisor resistivo**: la figura muestra el circuito del amplificador clásico de emisor común, polarizado de modo estable.

Los **condensadores** de entrada y de salida, para el acoplamiento de la señal, no tienen ninguna influencia en la polarización: para la corriente continua es como si no existieran.

En el cálculo del divisor se tiene en cuenta la **corriente de base**, normalmente bastante débil debido a la elevada ganancia de los transistores (normalmente del orden de 200...500 veces).

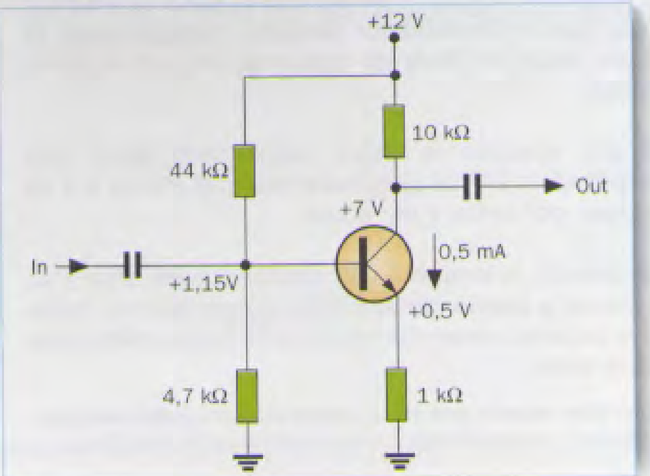

Amplificador de transistores polarizado en modo estable: la salida tiene amplio espacio para moverse en ambas direcciones.

AMPLIFICACIÓN

La tensión de base se refleja exactamente igual en el emisor (0,65 V más bajo), así la **corriente** de R_E depende de la señal de entrada.

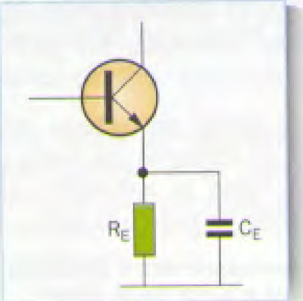

Un condensador aumenta la ganancia en corriente alterna, es decir, para la señal, pero la hace poco previsible.

Pero la **misma corriente** (o mejor casi toda) atraviesa también R_L: las tensiones en R_E y R_L son por tanto proporcionales al valor de las resistencias.

En otras palabras, la **amplificación** (ganancia en tensión) del circuito es más o menos la relación entre los dos valores: R_L / R_E, es decir, 10 en nuestro ejemplo.

Añadiendo un **condensador** entre emisor y masa, R_E para la señal es casi cero y la ganancia es teóricamente infinita (en la práctica depende de la beta del transistor y es bastante variable).

INVERSIÓN DE POLARIDAD

Cuando la señal de la entrada va hacia el positivo, el transistor conduce más: la corriente de colector aumenta y la tensión en el colector **disminuye**.

En consecuencia, como hemos visto, la señal a la salida tiene una **polaridad opuesta** con respecto a la de entrada: en el caso de una onda repetitiva (por ejemplo una sinusoide), resulta, por tanto, desfasada en 180 grados.

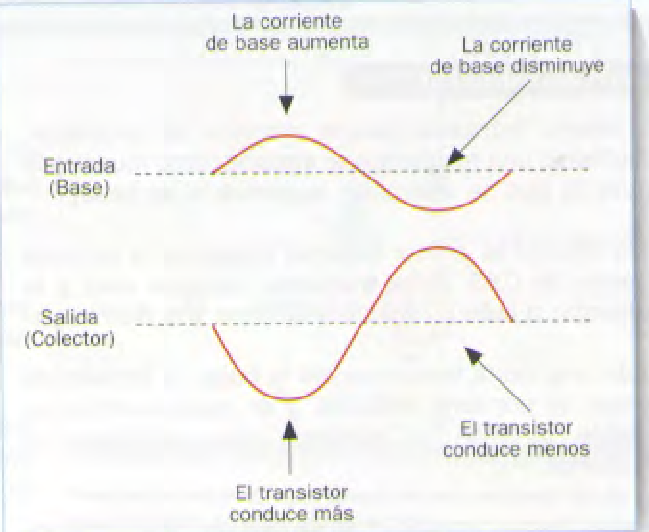
La corriente de base aumenta
La corriente de base disminuye
Entrada (Base)
Salida (Colector)
El transistor conduce menos
El transistor conduce más

El signo de la tensión alterna a la salida se invierte con respecto a la de entrada.

Capacidad y desfase

La corriente alterna de condensadores e inductores no está en fase con la tensión

Aplicando una tensión alterna **sinusoidal** a una resistencia pasa una corriente que sigue fielmente el recorrido de la tensión.

En los momentos en los que la tensión es cero, la corriente es cero, mientras que durante los **picos** de la sinusoide la corriente alcanza el valor máximo.

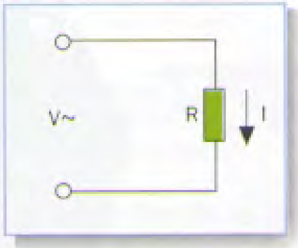

La corriente en una resistencia está en fase con la tensión: picos y pasos por el cero coinciden

En todo instante se puede aplicar la ley de Ohm: I = V / R, es decir, la corriente se obtiene dividiendo la tensión por la resistencia. Se dice que corriente y tensión están **en fase**.

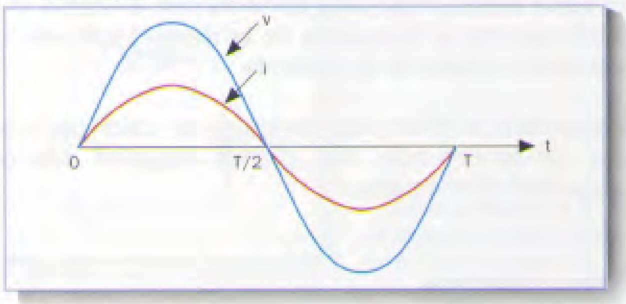

CORRIENTE EN UN CONDENSADOR

Aplicando una tensión **continua** a un condensador, una vez que está cargado ya no circula corriente alguna.

Sin embargo, cuando la tensión **cambia** sí pasa corriente, porque el condensador debe cargarse o descargarse; cuanto más rápido es el cambio de la tensión, mayor es la corriente.

Observando una sinusoide, se ve que **alterna** cambios veloces (por ejemplo, al paso por el cero), con instantes de relativa estabilidad (en los picos).

La corriente alterna en un **condensador** será por tanto mínima cuando la tensión es máxima... ¡y máxima cuando la tensión es cero!

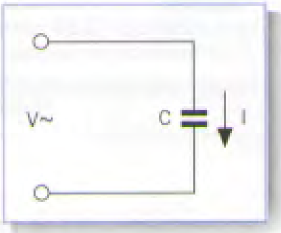

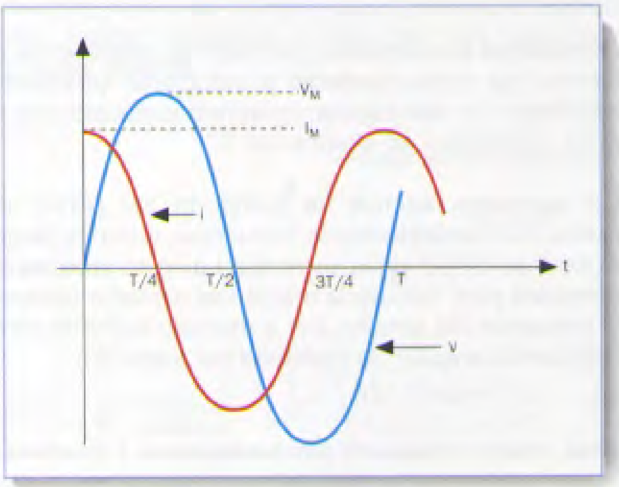

La corriente en un condensador es máxima cuando la tensión se mueve rápidamente, es decir, cerca del cero.

ANTICIPACIÓN Y RETRASO

En la figura central se ve que la corriente alterna en un condensador es también una sinusoide, pero **anticipada 90°** (un cuarto de ciclo) con respecto a la tensión.

Lo mismo ocurre con las **bobinas**, sólo que en este caso la corriente (siempre sinusoidal) está retrasada 90° con respecto a la tensión.

Es fácil recordarlo, considerando que los inductores tienen "inercia", es decir, reaccionan siempre con un cierto **retraso** a los cambios de tensión.

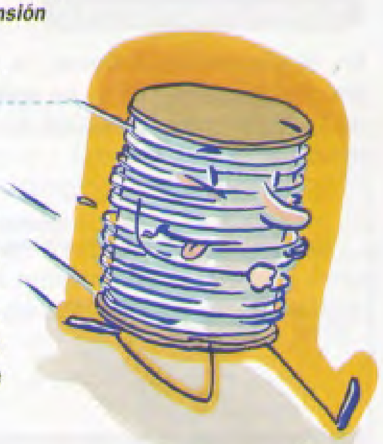

Reactancia

La corriente alterna en condensadores e inductores depende de la frecuencia

Si una tensión alterna varía muy despacio, es decir, tiene **baja frecuencia**, un condensador conectado a ella se deberá cargar y descargar con menos frecuencia: por tanto la corriente será menor.

El **valor eficaz** (rms) de la corriente que atraviesa el condensador dependerá pues de la frecuencia de la tensión aplicada: cuanto mayor es la frecuencia, mayor es la corriente.

Dependerá también, naturalmente, del **valor** del mismo condensador: un condensador más grande requerirá más corriente para cargarse y descargarse.

REACTANCIA

Considerando siempre los valores eficaces (rms), podemos aplicar la **ley de Ohm** y dividir la tensión por la corriente.

Obtenemos la reactancia, es decir la "resistencia en alterna" de condensadores e inductores (elementos reactivos). La **reactancia**, igual que la resistencia, se mide en ohmios; su símbolo es X.

Un elemento reactivo se comporta así como una resistencia **variable** con la frecuencia, además de con el desfase de 90° de la corriente. La figura muestra las **fórmulas** para calcular la reactancia; se debe observar la presencia del término 2πf, a menudo indicado como "frecuencia angular" ω (radianes por segundo).

	Condensador	**Inductor**
Reactancia en ohmios	$1 / \omega C$ o bien $1/2\pi fC$	ωL o bien $2\pi fL$

Leyenda:

π = número pi (3.14 aprox.)

f = frecuencia en herzios

C = capacidad en faradios

L = inductancia en henrios

Cómo calcular la reactancia para condensadores e inductores.

EJEMPLO DE CÁLCULO

En el circuito de la figura, la tensión de red (220 V sinusoidales a 50 Hz) se ha aplicado a un condensador de 1 µF (es decir, 0,000001 F). La reactancia a 50 Hz del condensador es:

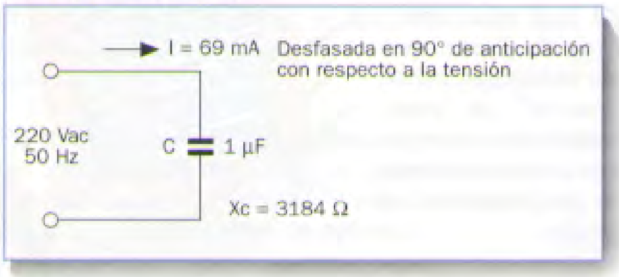

$X_C = 1/2\pi fC = 1 / (6{,}28 \times 50 \times 0{,}000001) = 3184\ \Omega$

La **corriente**, según la ley de Ohm, es por tanto:

$I = 220 / 3184 = 0{,}069\ A = 69\ mA$

Hay sin embargo una diferencia importante con respecto a una resistencia: pasa corriente, pero no hay ninguna **disipación** de potencia: la energía se mueve simplemente hacia delante y hacia atrás.

Un condensador de 1 µF al que se le aplica una tensión sinusoidal de 50 Hz tiene una reactancia de 3184 Ω aproximadamente.

Impedancia

En los circuitos reales, la combinación de resistencias y elementos reactivos hace las cosas más complejas

Un circuito como el mostrado (R y C en serie) no es tan sencillo como parece: reactancia y resistencia no se pueden **sumar**.

La corriente estaría en fase con la tensión si únicamente estuviera la resistencia, **desfasada** 90° si solamente estuviera el condensador.

Como la corriente los atraviesa a ambos, es necesario considerar **conjuntamente** la amplitud (RMS) y la fase: todos los cálculos deberán tenerse en cuenta.

El valor resultante de los dos componentes no es ni una resistencia pura ni una reactancia pura, sino una combinación de ambas: se llama **impedancia**.

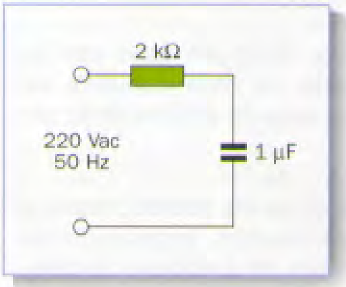

***Este circuito no es solamente resistivo ni solamente reactivo: tiene una cierta* impedancia.**

NÚMEROS COMPLEJOS

Nos referimos a una útil herramienta matemática: los números complejos, formados por dos partes distintas: la parte real y la **imaginaria**.

Se pueden representar como un **punto** en el plano, identificado por dos coordenadas. Por ejemplo, el número 3 + 4i está en posición 3 horizontal y 4 vertical (forma cartesiana). Esta representación se presta especialmente a indicar la **parte resistiva** y la **parte reactiva** de una impedancia.

Como alternativa, se pueden indicar con un valor absoluto y un ángulo (forma polar), útiles por ejemplo para indicar amplitud y fase de una tensión o corriente alterna.

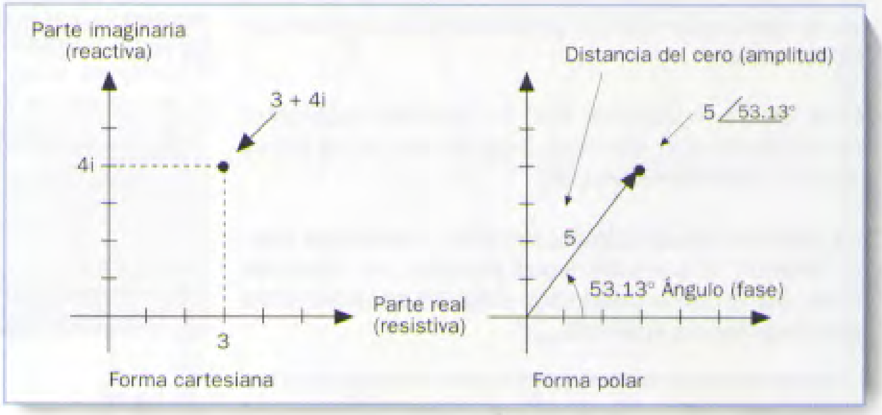

El número 3 + 4i* puede representarse *en forma cartesiana (dos coordenadas) o en forma polar (amplitud y ángulo).

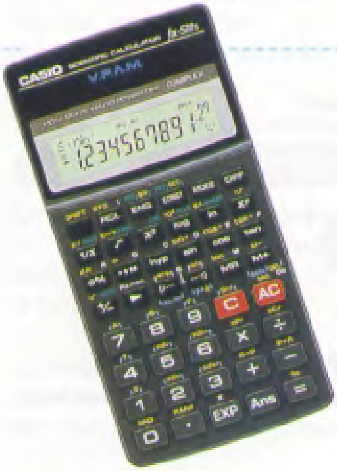

CÁLCULOS EN LA PRÁCTICA

No profundizaremos en los cálculos; observemos sin embargo que, representadas impedancia y reactancia (o amplitud y fase) en un solo número complejo, se cumple nuestra **ley de Ohm** de siempre.

La diferencia es que los cálculos se desarrollan con números complejos, y el resultado (por ejemplo, la corriente) será un **número complejo** que indica amplitud y desfase con respecto a la tensión de alimentación.

Muchas calculadoras científicas pueden trabajar* con números complejos, *y convertirlos a las dos formas.

Potencia y refase

Para el máximo rendimiento, conviene que la red eléctrica vea una carga resistiva

Muchos dispositivos conectados a la red, como motores y transformadores, pueden tener un comportamiento más o menos **reactivo** (normalmente inductivo) en determinadas condiciones.

La corriente no estará por tanto **en fase** con la tensión aplicada; una parte de esta corriente se moverá simplemente hacia delante y hacia atrás sin producir potencia.

Circulando en la resistencia de los cables, causará sin embargo **pérdidas** de energía, especialmente molestas (para las potencias en juego) en instalaciones industriales y líneas de distribución.

Los transformadores pueden causar* desfase *entre corriente y tensión, dependiendo de la carga conectada al secundario.

POTENCIA APARENTE Y POTENCIA REAL

Si corriente y tensión no están en fase, la potencia ya no es simplemente el **producto** de la tensión por la corriente.

Este último representa sólo la "potencia aparente" de un dispositivo eléctrico, indicada no ya en vatios sino en **voltamperios** (VA).

La potencia consumida eficazmente, o **potencia real**, es inferior; la corriente sigue pasando en cualquier caso, por lo que la instalación eléctrica se dimensiona para la potencia aparente.

La relación entre potencia real y potencia aparente (W / VA) se llama **factor de potencia**: su valor ideal es uno.

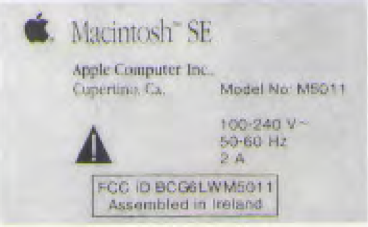

Macintosh™ SE

Apple Computer Inc.
Cupertino, Ca.

Model No: M5011

100-240 V~
50-60 Hz
2 A

FCC ID BCG6LWM5011
Assembled in Ireland

Los cables de la instalación se dimensionan para la potencia aparente (en voltamperios); la real que se paga es inferior.

REFASE

El factor de potencia se indica también como "cos φ", es decir, el coseno (un operador matemático) del **desfase** φ, que se lee "fi".

Este último no es otra cosa que la **diferencia de fase** entre tensión y corriente; si están en fase (φ = 0), significa que la carga es puramente resistiva: cos φ vale 1.

Para reducir al mínimo el desfase, se suele añadir un componente reactivo de signo opuesto: para compensar un inductor, por ejemplo en un motor, se pone un **condensador de refase**.

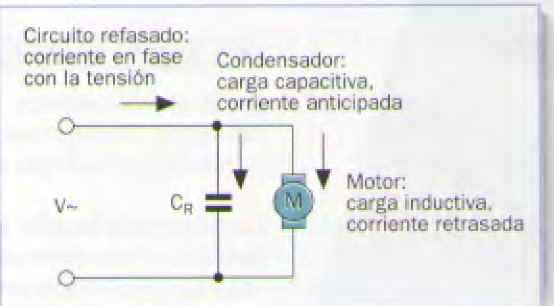

Los condensadores de refase evitan excesivos desfases entre tensión y corriente.

Seguidor de emisor

Ilustramos algunas configuraciones clásicas de empleo de los transistores

La figura muestra un seguidor de emisor, un circuito elemental, pero muy utilizado.

Aplicando en la base una cierta tensión, en el emisor se recuperará **la misma tensión** menos la caída base-emisor (0,65 V más o menos).

Si la tensión a la salida fuera más baja, aumentaría la corriente de base y por tanto la de colector, **subiendo** así la tensión de emisor (y viceversa).

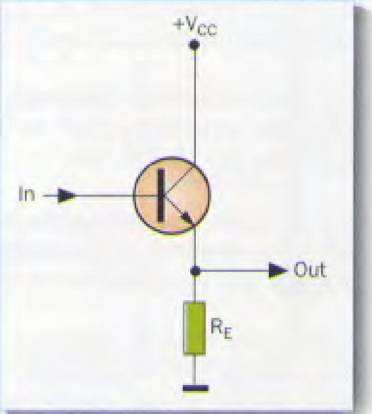

Seguidor de emisor: la tensión de emisor sigue a la de base, quedando siempre inferior a los aproximados 0,65 V.

GANANCIA DE CORRIENTE

La tensión a la salida sigue fielmente la de entrada: si varía en un voltio, también la otra hace lo mismo. En otras palabras, la **ganancia de tensión** es 1; podría parecer que el circuito no es útil.

Pero a la salida está disponible una **corriente** más elevada que la proporcionada a la entrada: hay una ganancia de corriente, y por tanto de potencia. Un seguidor de emisor puede controlar una **carga** que requiera una corriente inaceptable para la fuente de la que se dispone.

Como el seguidor de emisor, la dirección asistida sigue fielmente los movimientos del conductor, pero reduce el esfuerzo necesario.

REDUCCIÓN DE IMPEDANCIA

Si una salida tiene una elevada resistencia interna, basta poca corriente para que la tensión decaiga: se dice que tiene una **alta impedancia**.

Un seguidor de emisor tiene sin embargo una **baja impedancia** de salida y una alta impedancia de entrada: es decir, proporciona mucha corriente pero absorbe poca.

Si una fuente de alta impedancia está conectada a una carga de baja impedancia (como una resistencia de bajo valor), se crea un **divisor**: el nivel de la señal se reduce.

El seguidor de emisor, funcionando como **reductor** de impedancia, permite resolver el problema: amplificando la corriente es como si se redujera la resistencia interna de la fuente.

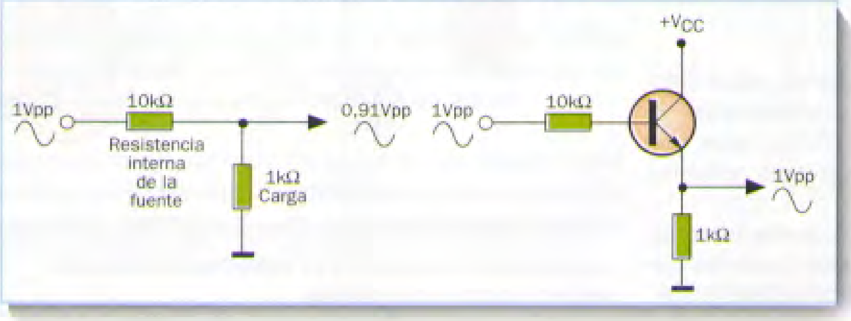

El seguidor de emisor permite adaptar una fuente de alta impedancia a una carga de baja impedancia

Control de la carga

La corriente de colector puede activarse o desactivarse, o bien regularse de un modo continuo

El circuito mostrado en la figura emplea un transistor como **interruptor** para encender o apagar la bombilla.

Hay una importante **ganancia de corriente**, ya que la corriente absorbida por la base es inferior a la requerida por la bombilla.

Hay también una **ganancia de tensión**, aunque menor: en el ejemplo, una variación de 5 V en la entrada produce una variación de 12 V a la salida.

Para cerrar bien el transistor, la **corriente de base**, multiplicada por la ganancia (ß), debe ser al menos igual (mejor si es mayor) que la requerida por la bombilla.

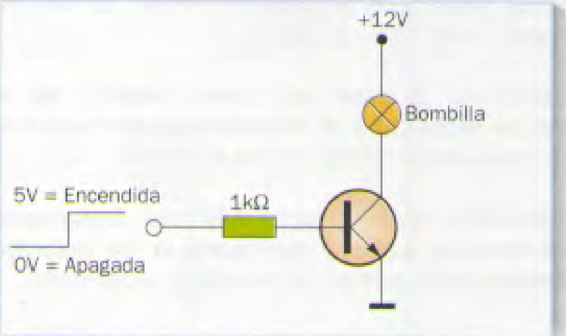

El transistor se cierra completamente (mínima caída de tensión) sólo si la corriente de base es suficiente.

CONTROL LINEAL Y DISIPACIÓN

La disipación en el interruptor es muy baja; pero las cosas cambian si no nos limitamos a conmutar, sino que por ejemplo queremos variar de **modo continuo** la luminosidad de la bombilla.

El transistor se comportará como una **resistencia variable**, por la que se desplazará la corriente de la bombilla, y en cuyos extremos habrá una cierta tensión: por tanto deberá disipar potencia (V x I).

El peor caso, es decir, la máxima disipación (y calentamiento) se tiene cuando la tensión de colector está **a mitad de camino** entre masa y alimentación.

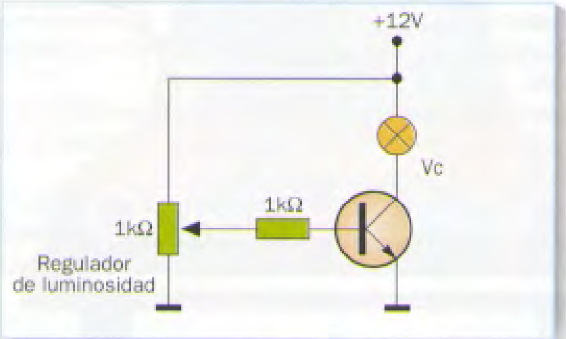

Regulador de luminosidad; en el peor de los casos (V_c = 6 V), el transistor disipa una potencia similar de la bombilla.

TRANSISTORES COMBINADOS

Para aumentar la ganancia, y por tanto reducir la corriente de base requerida, es posible utilizar **dos transistores** uno detrás de otro, incluso en el mismo contenedor.

La configuración **Darlington**, bastante difundida, puede también adquirirse como componente ya preparado: la ganancia es elevada (1000 o más, que para los transistores de potencia es un valor notable).

Por el contrario, el Darlington no cierra perfectamente, es decir, tiene una **tensión de saturación** bastante elevada (incluso de más de 2 V), a menudo no deseable.

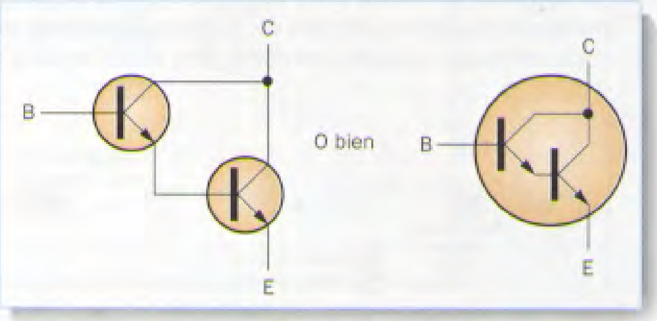

Dos transistores en cascada en configuración* Darlington*, para alta ganancia de corriente.

Transistores complementarios

A menudo resulta útil emplear transistores PNP y NPN en el mismo circuito

Aunque se empleen con más frecuencia los transistores NPN, nada impide utilizar transistores del tipo **PNP** complementario.

Resultan útiles, por ejemplo, en circuitos en los que la tensión de alimentación es **negativa**, como el interruptor mostrado en la figura.

La corriente (convencional, del positivo al negativo) sigue siempre el **sentido** de la flecha indicada en el transistor, como se ve por las flechas de la figura.

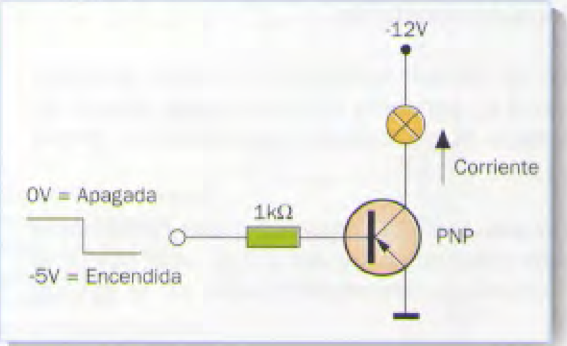

Utilizando un transistor PNP, corrientes y tensiones se invierten con respecto a las de un NPN.

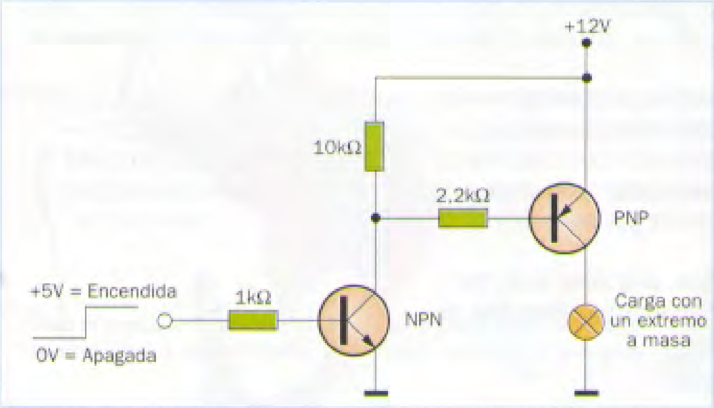

Interruptor en el positivo: el transistor PNP es el interruptor principal, controlado por el NPN.

INTERRUPTOR SOBRE LA ALIMENTACIÓN

A menudo resulta útil interrumpir la corriente en el **lado caliente** de una carga, en lugar de en el conectado a masa; en este caso el clásico transistor con emisor a masa no sirve.

Dándole la vuelta un momento al esquema y suponiendo que la alimentación sea la masa, podemos darnos cuenta de que se puede utilizar un transistor **PNP**, terminando en el esquema anterior.

Se puede controlar la **base** con un interruptor normal NPN, que tome corriente de la "alimentación" del PNP, es decir, en realidad de la masa, como muestra la figura.

CIRCUITOS COMPLEMENTARIOS

Combinando así un NPN y PNP, es posible realizar circuitos que se comporten de un modo **simétrico**, tanto hacia alimentación como hacia masa.

La figura muestra el principio de un seguidor de emisor **complementario**, en el que se basan por ejemplo las etapas finales de los amplificadores de audio.

La versión mostrada tiene sin embargo un defecto (en el campo analógico): una **inseguridad** en el centro, debida a la caída base-emisor de los transistores; veremos cómo se utilizan circuitos de corrección adecuados.

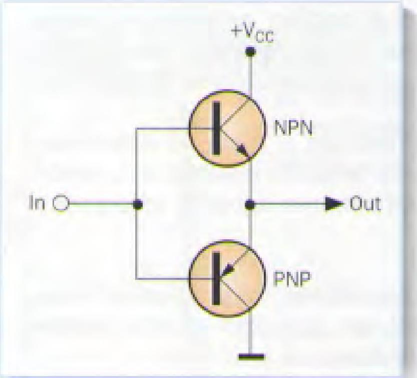

Seguidor de emisor complementario, capaz de trabajar en ambas polaridades de la señal de entrada.

Generadores de corriente constante

Un transistor se puede utilizar para mantener constante el valor de una corriente

El seguidor de emisor puede verse como un generador de **tensión constante**.

El intento de **reducir** la tensión de emisor aumenta la tensión B-E, por tanto la corriente de base y en consecuencia la de colector, que sube la propia tensión.

Aplicando esta tensión constante a una **resistencia** obtenemos (según la ley de Ohm) una corriente también constante, dependiente sólo de la tensión de base.

Hemos visto en la lección 11 que esta técnica se puede utilizar para **polarizar** de un modo estable los amplificadores de transistores.

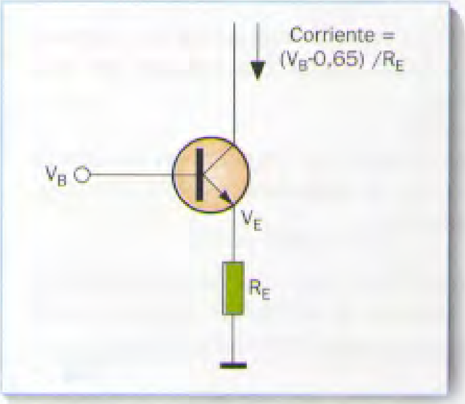

Generador de corriente constante: *la corriente en R_E depende de V_E, que a su vez es V_B – 0,65 V más o menos.*

LÍMITES E IMPRECISIONES

Obviamente no es posible hacer pasar la corriente estable, si la **carga** no lo permite: por ejemplo, si hay una resistencia de valor demasiado elevado.

En condiciones normales, el circuito mantiene la corriente tanto más **estable** (al variar la carga) cuanto más elevada es la ganancia del transistor.

Por la resistencia RE pasa además la **corriente de base**: si la ß es baja, se convierte en un porcentaje significativo de la corriente total e interfiere con la regulación.

CONDENSADORES Y CORRIENTES INVERSAS

El circuito mostrado en la figura, un seguidor de emisor con un gran condensador como carga, ilustra dos posibles **problemas**.

Conectando la entrada al positivo, se tiene una fuerte **corriente de arranque**: el condensador descargado es como un cortocircuito que puede dañar el transistor.

Poniendo después la entrada a masa, el condensador cargado aplica una **tensión inversa** a la unión emisor-base, que normalmente aguanta sólo pocos voltios.

Aparte de posibles protecciones, los problemas no se producen si el control del transistor cambia **lentamente** en lugar de a saltos.

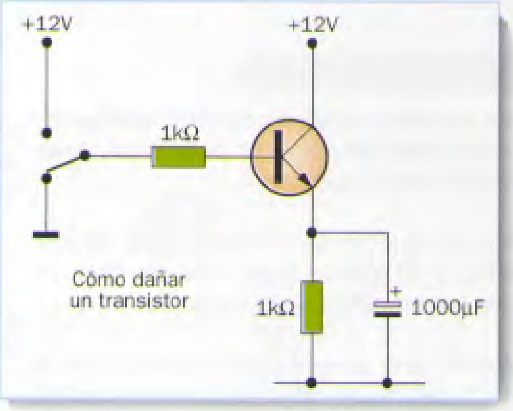

Los condensadores *pueden causar picos de corriente y tensiones inversas en los circuitos conmutados rápidamente.*

Diodos Zener

Una limitación de los diodos puede llegar a ser una característica interesante para regular una tensión

Cuando la **tensión inversa** aplicada a un diodo se aproxima al límite de ruptura, empieza a pasar corriente en el "sentido prohibido" de la unión.

Esta corriente puede no ser deseable y dañar el componente, pero también se puede **utilizar** con otros fines.

Se debe a fenómenos físicos, llamados "efecto avalancha" y **efecto Zener**; este último puede controlarse dosificando con exactitud las impurezas en el silicio. Los **diodos Zener** aprovechan especialmente el segundo efecto, de modo que dejan pasar corriente inversa a partir de la tensión deseada, por ejemplo 5,6 V.

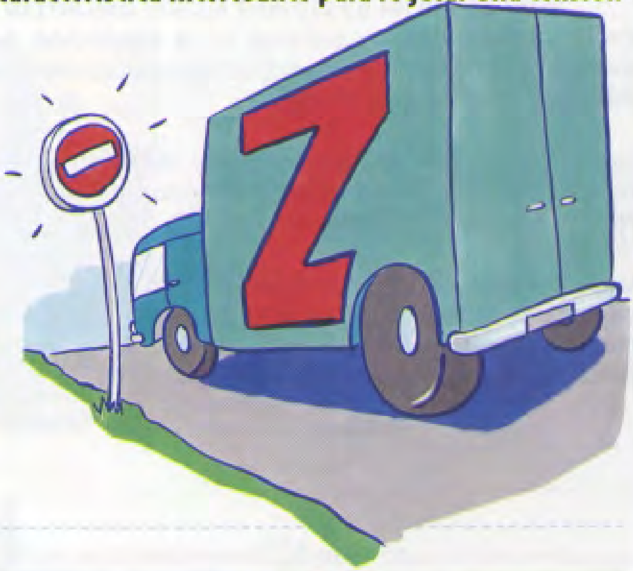

TENSIÓN Y CORRIENTE

En **conducción directa**, es decir, con el positivo aplicado en el ánodo, un diodo Zener se comporta como un diodo normal, dejando pasar la corriente.

Invirtiendo la tensión pasa, como siempre, sólo la minúscula **corriente de pérdida**, mientras la tensión aplicada se mantiene baja.

Cuando la tensión se aproxima al valor de Zener, la **resistencia** del dispositivo baja bastante y por tanto la corriente sube rápidamente, como muestra la figura.

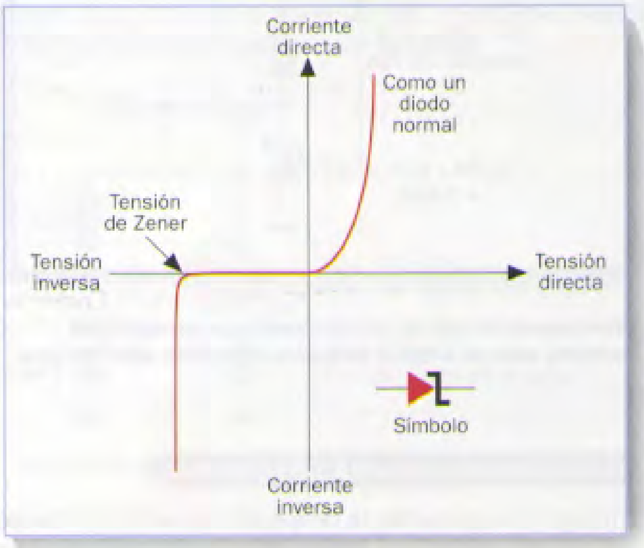

Cuando la tensión inversa alcanza el valor de "ruptura" Zener, la corriente aumenta bruscamente.

ESTABILIZADOR DE TENSIÓN

El uso principal de un diodo Zener es como **regulador** de tensión, polarizado inversamente con una resistencia en serie como muestra la figura.

La **tensión a la salida** es igual al valor de Zener: si fuera superior, el diodo Zener absorbería más corriente, provocando una mayor caída en la resistencia y reduciendo así la propia tensión.

El punto importante es que la tensión a la salida no se ve afectada (al menos teóricamente) por las **variaciones** de alimentación: ya sea de 10 o de 20 V, a la salida se tienen siempre los 4,7 V del diodo Zener.

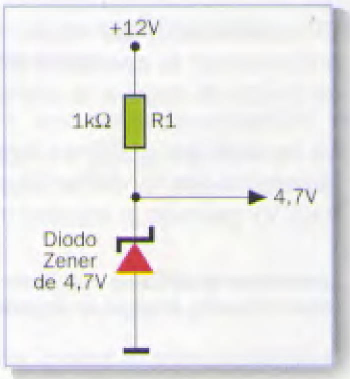

Generador de tensión estabilizada con una resistencia conectada en serie a un diodo Zener.

Alimentadores estabilizados

Zener y transistores permiten obtener una tensión independiente de la carga

En un alimentador, los cambios en la **absorción de corriente** por parte de la carga causan variaciones de tensión y zumbidos de fondo.

Supongamos por ejemplo que los devanados del transformador y de los diodos redirectores tienen en total una **resistencia interna** de 0,5 Ω.

Una variación de **carga** (corriente absorbida) de 5 A causa una variación de tensión de 5 x 0,5 = 2,5 V: inaceptable si el circuito está alimentado con una tensión estable, por ejemplo 5 V.

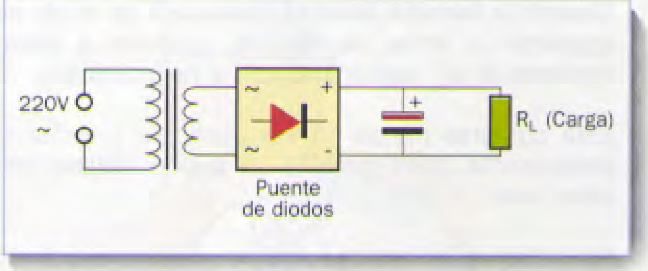

***Alimentador** no estabilizado, la tensión de salida cambia tanto al variar la de entrada, como la carga R_L.*

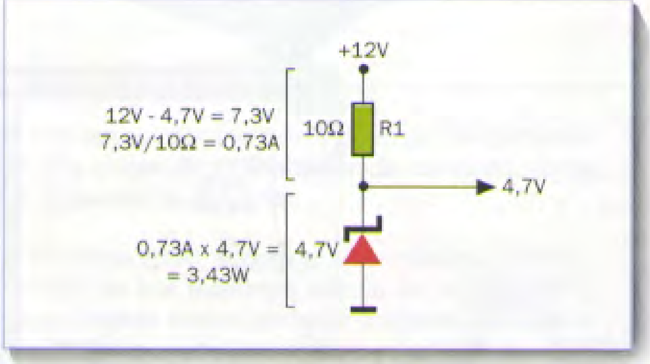

Disminuyendo el valor de la resistencia se proporciona más corriente, pero sin carga la disipación en el diodo sube bastante.

LÍMITES DEL DIODO ZENER

El sencillo estabilizador de la página anterior, formado por una resistencia y un diodo Zener, es aceptable para cargas muy **ligeras**.

Si la carga requiriese también únicamente 10 mA, en la resistencia R1 (de 1 KΩ, ver página anterior) caerían 10 V y la tensión de salida **descendería** a 2 V, muy por debajo del valor deseado.

Reducir R1, como en la figura, resolvería el problema, creando sin embargo otro: **en vacío**, es decir, en ausencia de carga, la corriente en el diodo (y por tanto su disipación) sería elevada.

UN TRANSISTOR PARA POTENCIAR

En lugar de aumentar la disipación en el diodo Zener, se puede utilizar un **transistor** como seguidor de emisor, tal y como muestra la figura.

El estabilizador de diodo Zener debe únicamente proporcionar la **corriente de base** del transistor, que se ocupa de regular la corriente principal.

La tensión de salida se **estabiliza** con la del diodo Zener, menos la caída base-emisor (más o menos 0,65 V): permite al transistor mantenerla constante.

***Alimentador estabilizado** con **Zener y transistor**: el primero estabiliza la tensión, el segundo proporciona corriente a la carga R_L.*

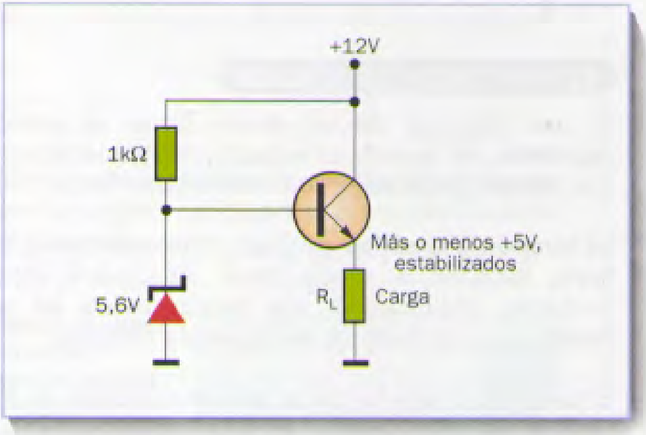

Disipación en el regulador

Los alimentadores estabilizados de tipo lineal derrochan mucha energía en forma de calor

El estabilizador de diodo Zener y transistor es de tipo **lineal**: se comporta como una resistencia variable, que regula la corriente de modo que se mantiene constante la tensión.

Como por el transistor regulador pasa corriente, y en sus extremos hay una tensión, **disipa potencia** (y por tanto se calienta).

Ello implica un derroche de energía y un **calentamiento** del propio transistor, que quizás requiera disipadores metálicos.

Veremos cómo los reguladores más complejos **de conmutación** resuelven el problema con una técnica radicalmente diferente.

El calor producido en el transistor regulador se extrae para mantener la unión a una temperatura aceptable

CAÍDA EN EL REGULADOR

Para **minimizar** la caída de tensión, y por tanto el calentamiento, la tensión de entrada debe ser baja, aunque siempre compatible con el funcionamiento del circuito.

La **tensión de entrada** ideal es aproximadamente de un par de voltios más que la de salida, suficiente para el estabilizador de diodo Zener y para la caída base-emisor.

Por desgracia, la tensión de red de la que proviene es variable (por ejemplo de 195 V a 245 V): el circuito debe funcionar **en el peor de los casos**, es decir, con una tensión mínima.

Para el calentamiento, el **peor** caso es sin embargo el contrario: máxima tensión de red, y máxima necesidad de corriente por parte de la carga.

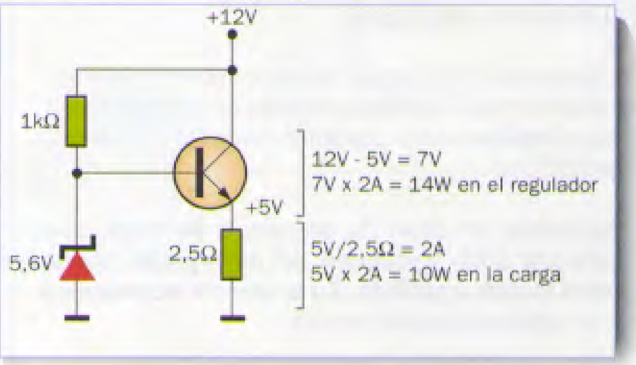

Disipación ***del transistor estabilizador en el peor de los casos: máxima tensión a la entrada, máxima corriente a la salida.***

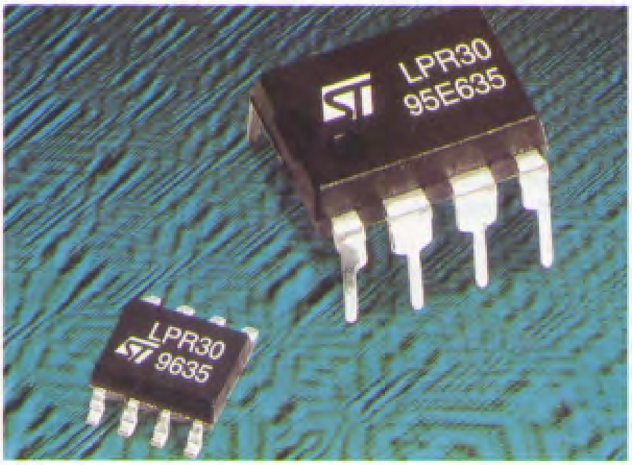

LÍMITES DE REGULACIÓN

La regulación al variar la carga es en cualquier caso **imperfecta**, a causa de la corriente de base y de la resistencia interna del transistor.

Lo ideal sería introducir una **retroalimentación**, es decir, controlar la tensión de salida y regularla en consecuencia; veremos a continuación cómo obtenerla.

Existen* reguladores de tensión *lineales en forma de circuitos integrados.

Protección de los alimentadores

Es necesario prevenir los problemas causados por corrientes excesivas, con el fin de que no ocasionen daños irreversibles

En caso de **cortocircuito**, el transistor regulador intenta subir la tensión dejando pasar más corriente, hasta ponerse en conducción completa.

Por tanto, pasa toda la corriente que la entrada puede proporcionar, causando normalmente la **destrucción** en pocos milisegundos del propio transistor.

Menos drástica, pero igualmente nociva, es simplemente una **sobrecarga**: con más corriente de la prevista, el transistor se calienta demasiado y su unión se daña.

El calentamiento causado por una corriente excesiva es el primer problema de los alimentadores estabilizados.

LIMITACIÓN DE CORRIENTE

Con una resistencia en serie y un segundo transistor se puede construir un **limitador** de corriente, como en el esquema de la figura.

Si la corriente en R_S supera el valor máximo previsto, la tensión en sus extremos manda en conducción a Q2, que **impide** enviar posterior corriente de base al regulador Q1.

La resistencia en serie R_S se calcula de modo que proporcione 0,65 V cuando es atravesado por la corriente **máxima** prevista. Su presencia empeora un poco la regulación de la tensión.

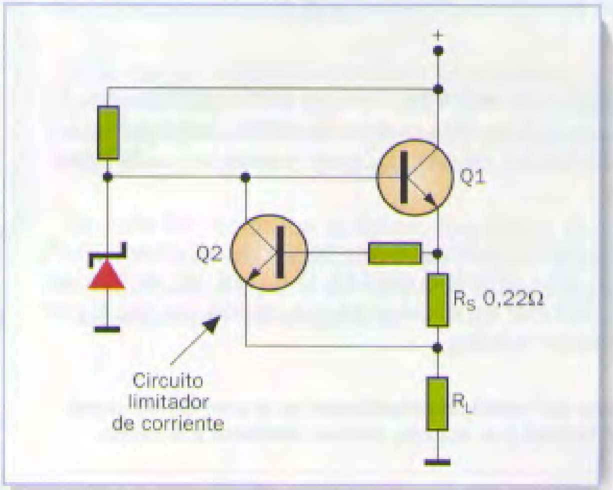

Circuito limitador de corriente

Limitación de corriente: ***con 3A, en los extremos de R_S hay 3 x 0,22 V = 0,66 V; Q2 conduce y le "roba" corriente en la base de Q1.***

LIMITACIÓN TÉRMICA

Aunque la corriente es limitada, el transistor regulador se **calienta** bastante; para evitar sobredimensionar su disipador, se puede añadir una protección basada en la temperatura.

Puede tratarse de un sencillo interruptor térmico (termostato) o de un circuito electrónico más refinado: debe en cualquier caso **apagar** el alimentador en caso de calentamiento del regulador.

Los alimentadores **de circuito integrado** están casi siempre protegidos tanto contra el cortocircuito como contra el calentamiento.

Resistencia sensible a la temperatura, utilizable por ejemplo en circuitos de protección contra el calentamiento.

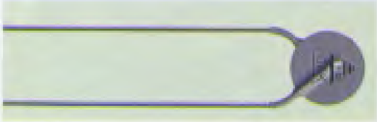

Transistores de potencia

Cuando entran en juego elevadas corrientes y potencias, se encuentran nuevos problemas

Incluso refrigerando perfectamente un transistor, hay un **límite** en la potencia máxima que puede disipar, límite indicado en los datos técnicos correspondientes (por ejemplo 100 W para un transistor de potencia).

El **producto** de la corriente que atraviesa el transistor por la tensión en sus extremos no puede superar este valor: 100 W permiten 5 A y 20 V, o bien 10 A y 10 V.

Un solo **impulso** de corriente puede ser también muy elevado, aunque muy breve: por ejemplo, podrían tolerarse 10 A y 20 V durante 300 µseg, aunque el producto dé 200 W.

SEGUNDA RUPTURA

No es solamente la temperatura de unión lo que plantea un **límite** a la corriente que puede atravesar un transistor BJT.

La corriente está también limitada por los **cables de unión** entre los terminales y la propia unión, que pueden calentarse y separarse.

Además, con el incremento de la tensión en los extremos del transistor se verifica un desagradable fenómeno denominado **segunda ruptura** (o second breakdown, en inglés). La corriente se concentra en "puntos calientes" de la unión, atrayendo más corriente y calentándola más de lo debido: es necesario mantener la corriente muy por **debajo** del valor calculado dividiendo la potencia por la tensión.

2N3055
MEXICO
9410

***Transistor BJT** de potencia en el típico encapsulado metálico TO-3, adecuado para una buena eliminación del calor.*

ÁREA DE TRABAJO

La figura resume los límites de los BJT mostrando una típica **zona operativa segura** (SOAR: safe operating area).

Los distintos colores del trazo indican las posibles **causas** de la limitación de la corriente tolerable de un modo continuado por el transistor: saliendo de esta zona, se daña.

Se observa el drástico efecto del **second breakdown**, bastante marcado por las altas tensiones; sin embargo hay en realidad una discreta tolerancia en las comparaciones de impulsos de corriente muy breves.

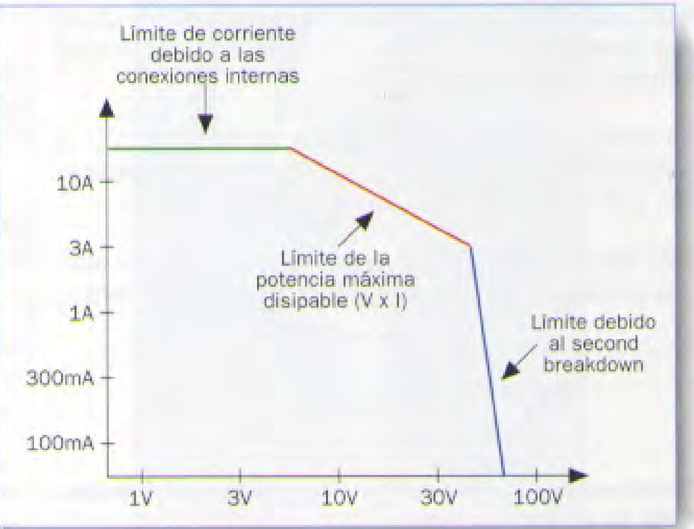

Zona operativa segura ***de un BJT de potencia: indica tensiones y corrientes que pueden estar presentes al mismo tiempo.***

Resistencia térmica

La temperatura eficaz de la unión depende de la eficacia de la eliminación del calor

La potencia disipable depende mucho de la capacidad de **eliminar** el calor, es decir, transportarlo lejos de la unión.

Para crear un flujo térmico es necesaria una **diferencia de temperatura**: cuanto más alta es esta diferencia, mayor será el calor transportado.

Se trata naturalmente de mejorar la conducción del calor, es decir, reducir la **resistencia térmica** entre la unión y el entorno externo.

Considerando la diferencia de temperatura como una tensión, la potencia disipada como una corriente y la resistencia térmica... como una resistencia, podemos aplicar la **ley de Ohm**.

DE LA UNIÓN AL ENTORNO

La resistencia térmica se expresa en °C/W o **grados centígrados por vatio**. Si por ejemplo es de 8 °C/W, una disipación de 15 W provocará una diferencia de 15 x 8 = 120 °C.

La unión del transistor (o de otro semiconductor) se encontrará a una temperatura **superior** de 120 °C con respecto a la temperatura ambiente.

Se considera el **peor caso** de temperatura ambiente, por ejemplo, 70 °C, y se añade la diferencia de temperatura causada por el paso del calor. La **unión** será por tanto de 70 °C + 120 °C = 190 °C, presumiblemente dentro del límite aceptable (alrededor de los 200÷250 °C para los transistores de silicio).

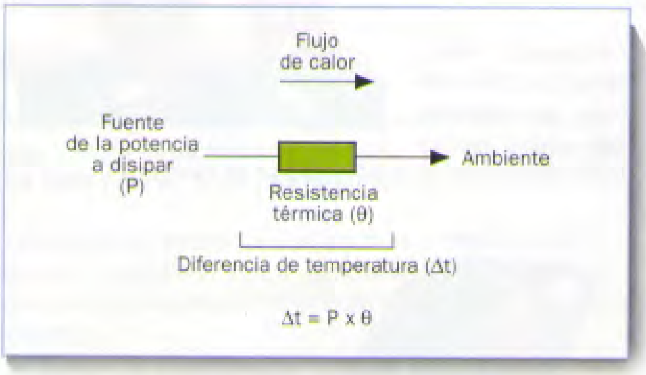
Flujo de calor
Fuente de la potencia a disipar (P)
Ambiente
Resistencia térmica (θ)
Diferencia de temperatura (Δt)
Δt = P x θ

La diferencia de temperatura (Δ_T), la potencia disipada (P) y la resistencia térmica (θ) están asociadas de un modo similar a la ley de Ohm.

DISIPADORES Y FLUJO DE AIRE

Para mantener baja la temperatura de unión incluso con potencias importantes, es necesario **reducir al mínimo** la resistencia térmica.

Se obtiene sumando la existente entre la unión y el encapsulado, entre el encapsulado y el posible disipador, y finalmente entre el disipador y el ambiente.

Esta última depende también del **movimiento** del aire: una circulación natural (convección) o forzada (viento).

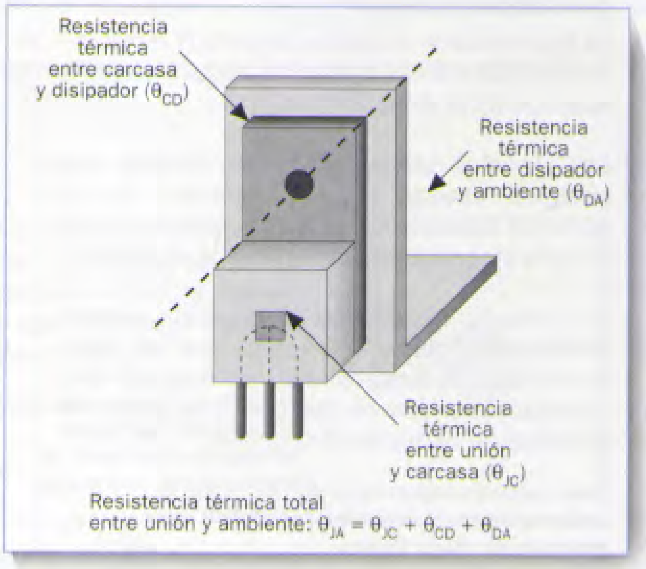
Resistencia térmica entre carcasa y disipador (θ_{CD})
Resistencia térmica entre disipador y ambiente (θ_{DA})
Resistencia térmica entre unión y carcasa (θ_{JC})
Resistencia térmica total entre unión y ambiente: $\theta_{JA} = \theta_{JC} + \theta_{CD} + \theta_{DA}$

Componentes de la resistencia térmica: con potencias elevadas, deben ser las tres muy bajas.

Transistores en paralelo

Para potencias elevadas, conviene repartir la carga entre varios dispositivos

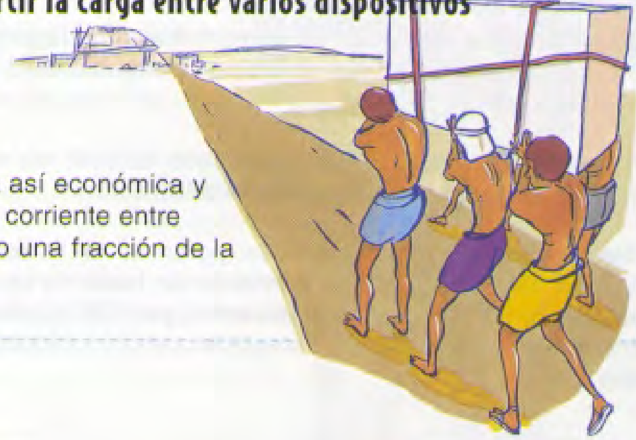

Consideraciones térmicas y el fenómeno second breakdown limitan la **corriente** controlable por un solo transistor.

Producir transistores de unión de gran potencia resulta así económica y prácticamente **poco conveniente**. Se puede **dividir** la corriente entre varios transistores, de modo que cada uno soporte sólo una fracción de la carga.

ESTABILIZACIÓN TÉRMICA

Conectando simplemente en **paralelo** dos o más transistores BJT, se darían pequeñas diferencias de producción en la ganancia (β) y en la tensión base-emisor (V_{BE}).

Uno de los transistores conduciría, y se calentaría, más que los demás; pero el calentamiento produce una **disminución** de V_{BE} y, por tanto, una mayor corriente de base (y de colector).

El transistor más caliente reclamaría así una parte **siempre mayor** de la corriente total, calentándose hasta destruirse... y los demás seguirían, uno cada vez.

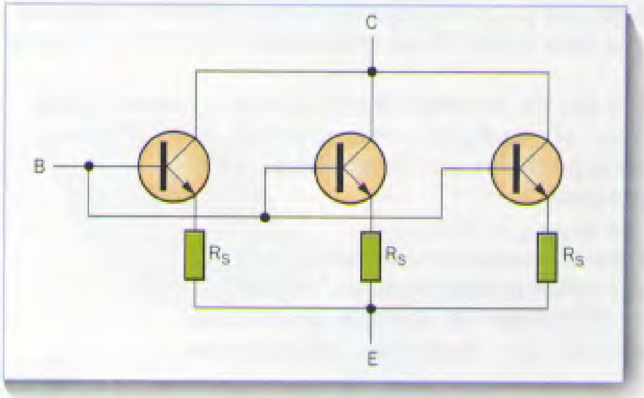

Las resistencias de estabilización impiden que por un transistor de unión pase mucha más corriente que por los demás.

MEJORES SOLUCIONES: IGBT Y MOSFET

Los transistores de unión no se utilizan ya para altas potencias: se **prefieren** los IGBT (Insulated-Gate Bipolar Transistor: transistor bipolar de puerta aislada) o los MOSFET.

Estos últimos no sufren la segunda ruptura y son **autoestabilizantes**: su tensión de "puerta" (el equivalente de la base) sube al aumentar la temperatura.

Se pueden por tanto conectar en paralelo **sin** resistencias adicionales, y consecuente disipación y caída de tensión, como veremos en la lección dedicada a ellos.

Para los coches eléctricos, que requieren cientos de amperios (en la foto una columna de recarga), se pueden utilizar muchos transistores MOSFET en paralelo.

Conmutadores de potencia

En su uso como interruptor, la presencia de corrientes elevadas plantea problemas específicos

En un transistor BJT utilizado como interruptor, la **potencia disipada** (V x I) es teóricamente siempre cero: realmente está abierta (I = 0) o cerrada (V = 0) .

En realidad, cuando el transistor está cerrado existe la disipación causada por la pequeña **tensión de saturación**.

Además, en los transistores de potencia la β está normalmente bastante baja, por lo que la **corriente de base** no es superflua. Debe asegurar la completa **saturación** en el peor de los casos, es decir, la mínima ganancia (β) y/o un exceso de corriente causado por una carga capacitiva.

TRANSICIONES PELIGROSAS

Si la potencia disipada en la saturación (cerrado) es baja, y la de corte (abierta) es nula, hay sin embargo una fase crítica: la de **transición**.

El paso de un estado a otro es más o menos rápido, pero en cualquier caso **gradual**: se tiene por tanto una cierta disipación de potencia.

Cuando el transistor está casi en corte, la tensión en sus extremos es elevada: existe riesgo de **second breakdown** incluso con corrientes relativamente bajas, especialmente si la carga es inductiva. Para evitar daños, la transición debe ser cuanto más **rápida** posible, tanto más si se repite con una elevada frecuencia.

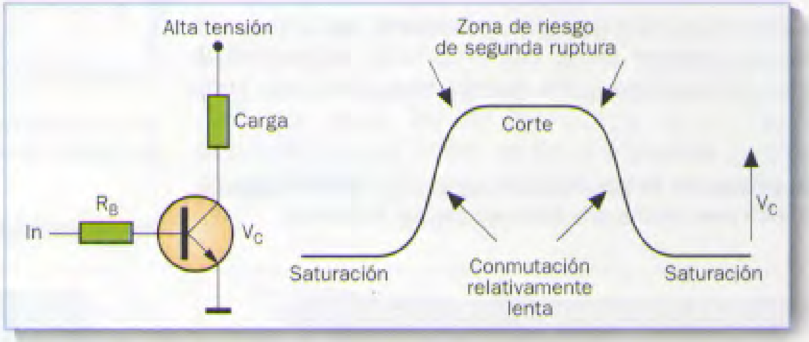
Alta tensión
Carga
R_B
In
V_C
Zona de riesgo de segunda ruptura
Corte
V_C
Saturación
Conmutación relativamente lenta
Saturación

En conmutación, el transistor debe estar durante el mínimo tiempo posible (idealmente cero) en la zona de segunda ruptura

VELOCIDAD DE CONMUTACIÓN

Las **capacidades internas** del transistor ralentizan el cierre y aumentan por tanto la disipación: para cargarlas es necesario enviar más corriente de la necesaria en la base.

En particular, el **efecto Miller** causado por la capacidad C-B (ver lección 11 de Componentes) limita la velocidad de conmutación, expresada en V/seg.

En la apertura, el transistor debe vaciarse de las cargas residuales antes de poderse abrir: también esto requiere tiempo.

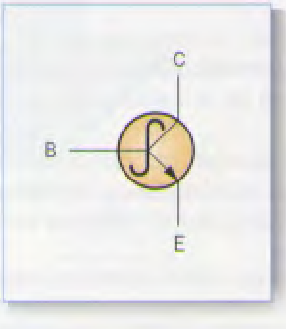

Los transistores Schottky no saturan completamente, de modo que no acumulen cargas y se abran a mayor velocidad.

Ondas que se repiten

Algunos tipos de señales cíclicas son comunes en los circuitos electrónicos; profundizamos en el tema introducido en la lección 4

La sinusoide se puede encontrar en muchos fenómenos naturales: representa el tipo más sencillo de **oscilación**.

Es sinusoidal, por ejemplo, el sonido producido haciendo vibrar un **diapasón** para afinar instrumentos musicales.

Su equivalente electrónico es un **circuito resonante**, en el que un condensador y una bobina intercambian continuamente energía, generando como decimos una sinusoide.

Recordemos que la forma y la frecuencia de las señales sinusoidales no cambia en circuitos compuestos sólo por componentes **lineales** (resistencias, bobinas, condensadores): cambia sólo la amplitud.

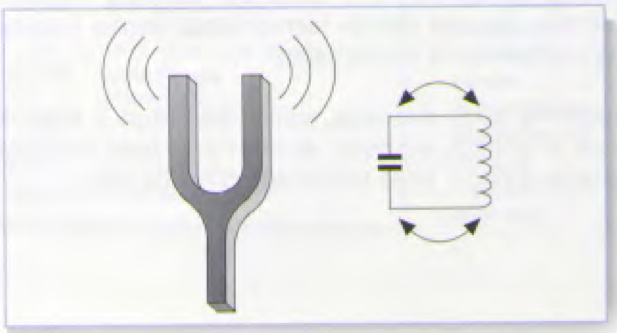

El diapasón y el circuito LC resuenan produciendo una sinusoide.

LA ONDA CUADRADA NO EXISTE

Son bastante comunes también las ondas cuadradas, en las que la tensión **se alterna** periódicamente entre dos valores bien definidos (por ejemplo L y H en los circuitos digitales).

A diferencia de la sinusoide, que existe realmente, una onda cuadrada es un **ideal**: la tensión no puede cambiar instantáneamente entre dos valores, sino que debe emplear un cierto tiempo.

En realidad, como también un simple cable tiene una inductancia mínima y una capacidad baja hacia tierra, los **frentes** de subida y bajada se ralentizan.

La figura muestra la diferencia entre una onda cuadrada teórica y un ejemplo **práctico** típico, incluidos los fenómenos que se pueden verificar en la relación entre frente de subida (o bajada) y zona constante.

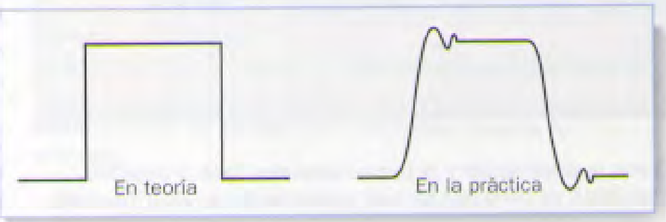

Las ondas "cuadradas" son sólo una simplificación teórica, la realidad es a menudo diferente, especialmente al crecer la frecuencia.

ONDAS CUADRADAS Y CIRCUITOS LINEALES

Más allá de los límites descritos anteriormente, las ondas cuadradas pueden **alterarse** de manera notable incluso en circuitos puramente lineales. La figura muestra un integrador primitivo (RC) y un **diferenciador** (CR): en ambos casos, la lenta carga del condensador produce una forma de onda diferente.

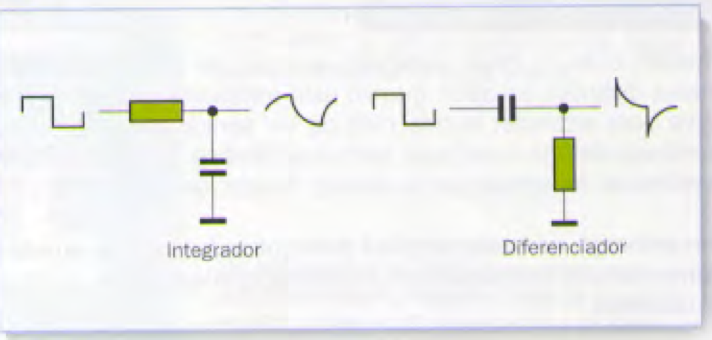

Un integrador atenúa las variaciones bruscas, mientras que un diferenciador las evidencia.

Serie de Fourier

Sumando sinusoides de frecuencia múltiple una de otra, se puede construir cualquier forma de onda

El matemático francés Joseph Fourier demostró que toda onda periódica puede considerarse como suma de una **serie** de ondas sinusoidales (serie de Fourier).

La primera sinusoide, llamada **fundamental**, tiene la misma frecuencia de la onda que se va a construir, mientras que las demás (**armónicas**) tienen frecuencia múltiple de la fundamental.

Según la onda deseada varían **amplitud** y **fase** de cada armónica, es decir, el nivel y la posición hacia delante o hacia atrás (retrasada o anticipada).

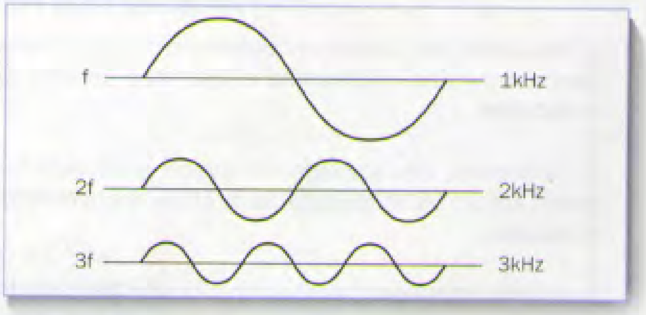

Ejemplo de frecuencia fundamental *(f), segunda armónica (2f) y tercera armónica (3f).*

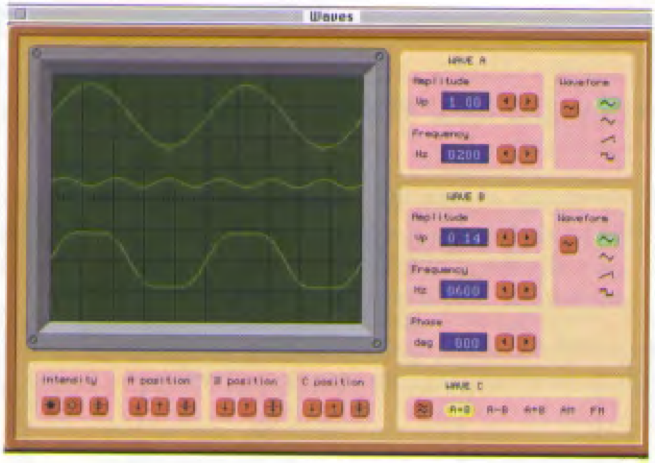

Vp 1.00
Hz 0200
Vp 0.14
Hz 0600
deg 000

Ya con la fundamental y la tercera armónica *(con la amplitud adecuada), se tiene una primera aproximación de onda cuadrada.*

CONSTRUIR UNA ONDA CUADRADA

Por ejemplo, una onda cuadrada se puede obtener sumando la fundamental con sus **armónicas dispares**: si la primera es de 100 Hz, se necesitan sinusoides de 300 Hz, 500 Hz, 700 Hz, etc.

La amplitud de las distintas armónicas es **decreciente**: a medida que sube la frecuencia, las armónicas se hacen más débiles.

Criterio empírico: cuanto más **angulosa** es la onda, más armónicas sinusoidales son necesarias para definirla por completo.

Para construir ángulos netos y precisos como los de una onda cuadrada ideal, se necesitaría un número **infinito** de armónicas.

OÍDO, FASE Y ARMÓNICAS

Nuestro oído (o, mejor, cerebro) es capaz de distinguir la **fase relativa** de dos ondas distintas, es decir, cuanto está retrasada una con respecto a la otra: nos sirve para entender la dirección de los sonidos. Percibe además el contenido armónico de una onda, que llamamos **timbre**: un sonido "dulce" tiene pocos armónicos, mientras que un sonido "sordo" tiene muchos.

Sin embargo, es bastante difícil distinguir la fase de las **armónicas** con respecto a la fundamental: dos ondas pueden percibirse iguales incluso aunque tengan una forma bastante diferente.

Decibelios

Para medir los sonidos y las señales que los representan, es más cómoda una escala logarítmica

El oído puede escuchar ruidos bastante débiles, como el murmullo del viento entre las hojas, y muy fuertes, como el aterrizaje de un avión: se dice que tiene una amplia **dinámica**. Sin embargo utiliza una **escala logarítmica**: es decir, aprecia las variaciones porcentuales del sonido en lugar de las variaciones absolutas.

Por ejemplo, un paso de 2 W a 3 W de audio, o de 200 a 300 W, se perciben como **incrementos** del mismo nivel, aunque la variación sea de 1 W en el primer caso y de 100 W en el segundo.

Intensidades sonoras indicativas en decibelios:

30 dB	susurro
40 dB	sala silenciosa
50 dB	lluvia de intensidad media
60 dB	conversación, lavavajillas
70 dB	aspiradora, tráfico
80 dB	despertador
– umbral de posible daño auditivo –	
90 dB	cortacésped
100 dB	sierra eléctrica
110 dB	concierto de rock, discoteca
120 dB	aterrizaje de un avión
130 dB	martillo neumático
140 dB	arma de fuego

La sensibilidad del oído es proporcional al nivel sonoro expresado en decibelios, es decir, al logaritmo del nivel real.

MEDIDAS EN DECIBELIOS

También en este caso, la relación entre dos señales (por ejemplo, la ganancia de un amplificador) se expresa en una medida logarítmica: el **decibelio**, de símbolo dB.

Aunque formalmente es submúltiplo del "belio" (que toma el nombre de A. G. Bell), el decibelio es la unidad de medida **estándar**. Como se puede ver en la figura, la **potencia** de la señal se duplica cada 3 dB (y se multiplica por diez cada 10 dB); por ejemplo, "aumentar 20 dB" significa multiplicar por 100 la potencia.

La **tensión** se duplica sin embargo cada 6 dB, y se multiplica por diez cada 20 dB: en realidad, duplicando la tensión se cuadruplica la potencia, que varía con el cuadrado de la tensión ($W = V^2R$).

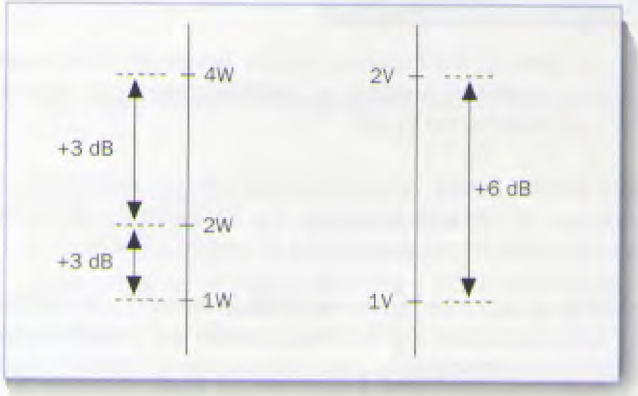

Sumar un valor en dB equivale a multiplicar tensión (o potencia).

VALORES EN dB ABSOLUTOS

El decibelio (dB) es una medida relativa, es decir, entre dos señales. Hay además unidades de medida absolutas en decibelios, que simplemente tienen un **punto de referencia**.

Por ejemplo, los **dBV** se refieren a 1 V rms: una señal de 1 V corresponde por definición a 0 dB, mientras que 2 V son +6 dB, 4 V son +12 dB, etc.

También se utiliza la medida de potencia **dBm**, referida a 1 mW con una cierta carga resistiva: normalmente 600 Ω para los circuitos de audio (aproximadamente 0,78 Vrms) y 50 Ω para los de radio (aproximadamente 0,22 Vrms).

SL-4001
75.8
SOUND LEVEL METER

Las herramientas de audio y para mediciones acústicas indican a menudo el nivel de la señal en dBm, o bien en dB con respecto a un punto de referencia.

Análisis con sinusoides

La serie de Fourier sugiere una técnica para calcular el comportamiento de un circuito con cualquier onda periódica

Conociendo el comportamiento de un circuito (lineal) con ondas sinusoidales en las distintas frecuencias, se puede prever su efecto sobre **cualquier** forma de onda.

Basta con **descomponer** la onda en sus componentes sinusoidales, considerar el efecto (en amplitud y fase) sobre cada una de ellas y al final "montarlas de nuevo" para obtener la salida. Esta técnica de análisis, denominada en el **dominio de la frecuencia**, simplifica notablemente el trabajo y se utiliza ampliamente en los proyectos de circuito.

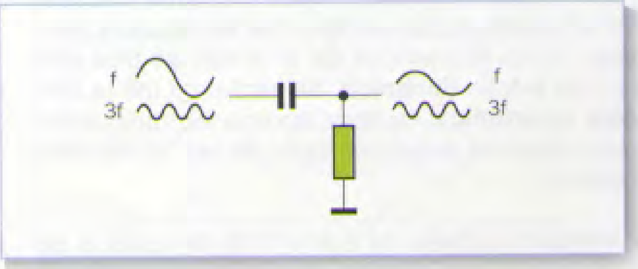

Para calcular el comportamiento de un circuito es cómodo considerar las distintas componentes **una cada vez.**

CIRCUITOS NO LINEALES

Por desgracia, los circuitos reales no están compuestos sólo por elementos lineales: **diodos y transistores**, por ejemplo, tienen un comportamiento no lineal.

En algunos casos ello no constituye un problema: por ejemplo, en un **amplificador** los transistores funcionan normalmente en su zona (más o menos) lineal.

En otras situaciones, como en el caso de un diodo rectificador, las sinusoides individuales pueden ser **modificadas** en su forma: la técnica de la descomposición ya no es suficiente.

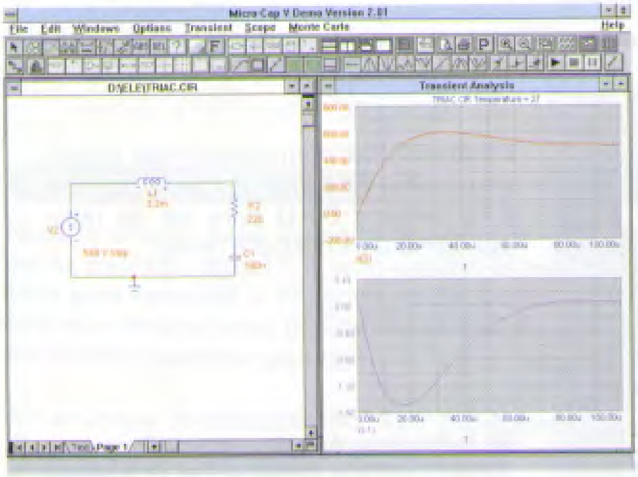
Micro Cap V Demo Version 2.01
File Edit Windows Options Transient Scope Monte Carlo Help
D:\ELE\TRIAC.CIR
Transient Analysis
L1
2.2m
R2
220
C1
V2
20.00u
40.00u
100.00u
T
Page 1

Análisis en el dominio del tiempo con uno de los distintos **programas** *derivados del conocido SPICE de la Universidad de Berkeley.*

SEÑALES NO REPETITIVAS

También en circuitos lineales, el método de la descomposición funciona solamente si la onda es **periódica**, es decir, se repite siempre igual a sí misma.

Impulsos sencillos, o en cualquier caso variaciones no periódicas, implican cálculos más complicados: se pasa al análisis en el **dominio del tiempo**.

En la práctica, el comportamiento del circuito se analiza **instante por instante**, utilizando intervalos de tiempo suficientemente breves.

Afortunadamente, hoy el análisis de circuitos está **automatizado** y, como veremos, al alcance de cualquier aficionado (aunque los conocimientos teóricos, como siempre, no vienen mal).

Filtros

Un circuito RC atenúa de distinta manera las señales según su frecuencia

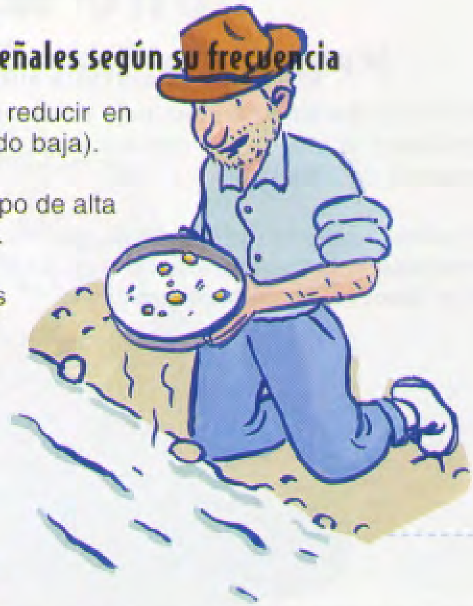

A menudo resulta conveniente bloquear, o en cualquier caso reducir en amplitud, las señales de frecuencia demasiado alta (o demasiado baja).

Se utilizan para ello **filtros** que, como el control de tono del equipo de alta fidelidad, atenúan más o menos una señal según su frecuencia.

Un grupo RC, como el mostrado en la figura de abajo, es el más sencillo ejemplo de filtro de **paso bajo** (low-pass): deja pasar las frecuencias bajas y atenúa las más elevadas.

De hecho, con una frecuencia de 0 Hz (continua), el condensador es un **aislante**, mientras que con una frecuencia infinita sería un **cortocircuito** hacia masa.

FRECUENCIA DE CORTE

Por debajo de una cierta frecuencia, el efecto **no es importante**. Sin embargo, las frecuencias altas se atenúan siempre mucho más.

Por exactitud, la atenuación es de **6 dB por octava**, es decir, 20 dB por década, es decir, la tensión de salida se divide a la mitad (-6 dB) cada vez que se duplica la frecuencia de entrada.

En un gráfico logarítmico, la atenuación en función de la frecuencia es una recta; idealmente la atenuación empezaría en un determinado punto: la **frecuencia de corte** del filtro.

En realidad el ángulo es redondeado y la frecuencia de corte se define como el punto en el que la atenuación es de **3 dB**, es decir, la tensión es 0,707 veces (1 / √2) de la de entrada.

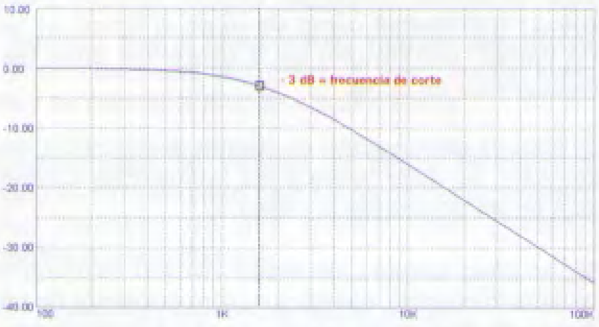

Respuesta en frecuencia *de un filtro de paso bajo RC: en horizontal la frecuencia, en vertical la atenuación en dB.*

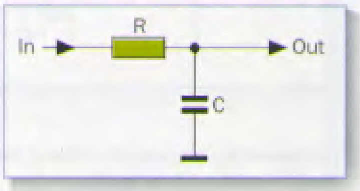

DESFASE

Los filtros no cambian sólo la atenuación, sino también la **fase** de la señal: a bajas frecuencias, donde el efecto del condensador no es importante, el desfase es cero.

A la frecuencia de corte el desfase es exactamente de **45**° (un octavo de ciclo) de retraso; aumentando aún más, el retraso se aproxima cada vez más a los 90°.

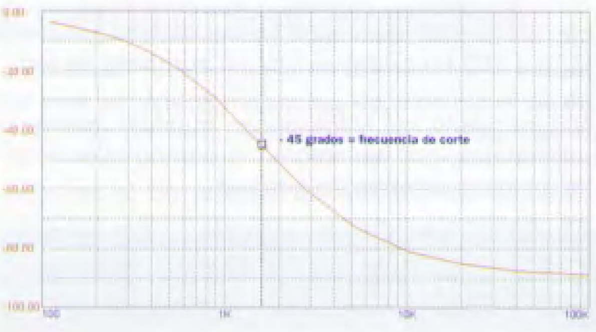

Desfase *de un filtro RC: en horizontal la frecuencia, en vertical los grados de retraso son respecto a la entrada.*

Tipos de filtros

Se pueden realizar filtros de distinto comportamiento, según la frecuencias a atenuar

El contrario del filtro de paso bajo es el de **paso alto** (high-pass): la atenuación crece cuando disminuye la frecuencia de entrada.

Se puede realizar **intercambiando** resistencia y condensador: la impedancia de este último aumenta con la disminución de la frecuencia.

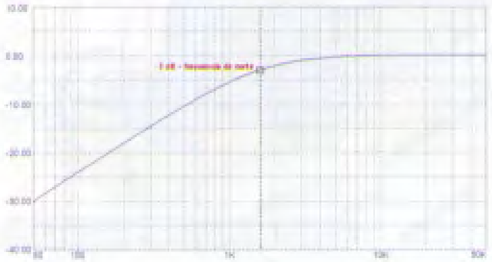

Se puede **variar** la frecuencia de corte de un filtro RC sustituyendo la resistencia por un potenciómetro, de modo que se pueda regular el valor.

Sin embargo, haciendo esto cambia también por desgracia la atenuación: es posible realizar filtros **regulables** de amplitud constante, pero con circuitos más complejos.

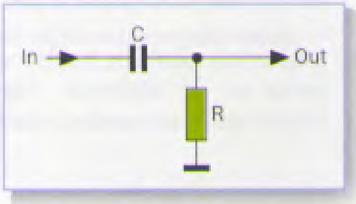

Un sencillo **filtro de paso alto** *y su respuesta en frecuencia, complementaria a la del filtro de paso bajo.*

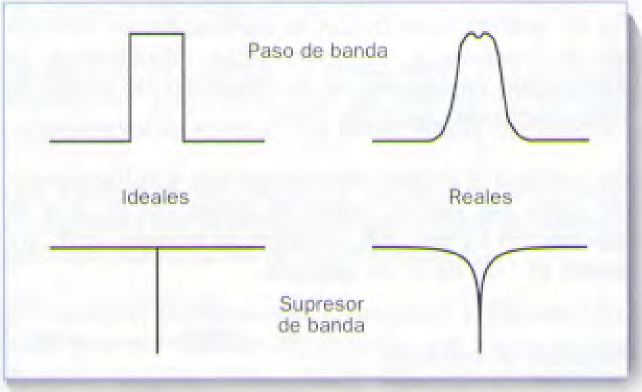

Respuesta en frecuencia *ideal y real de los filtros de paso de banda y supresor de banda.*

PASO DE BANDA Y SUPRESOR DE BANDA

A veces se desean solamente las señales en un determinado **rango** de frecuencias, por ejemplo entre 500 Hz y 3.800 Hz, y no en las más altas o más bajas.

Esto se puede conseguir con un filtro de **paso de banda** (bandpass) como se ve en la figura; el "ancho de banda" (bandwidth) es la diferencia entre las dos frecuencias de corte.

El filtro de **eliminación de banda** desarrolla la función opuesta, es decir, atenúa las frecuencias comprendidas en un determinado rango.

Si, en caso extremo, se desea eliminar sólo una determinada frecuencia, se utiliza un filtro **supresor de banda**: tiene una atenuación muy fuerte en un único punto.

EFECTOS DE LOS FILTROS

Un filtro de paso de banda puede utilizarse para reducir **interferencias** no deseadas: por ejemplo determinadas grabadoras baratas cortan las frecuencias altas para reducir el ruido.

Pero si las interferencias son de baja frecuencia, como un rumor de fondo, se podrá atenuar con un filtro de paso alto, mientras que un notch a 50 Hz o a 100 Hz podrá reducir un **zumbido** de red.

Naturalmente los filtros tienen muchas **aplicaciones** adicionales, por ejemplo en el campo telefónico y en las comunicaciones en general.

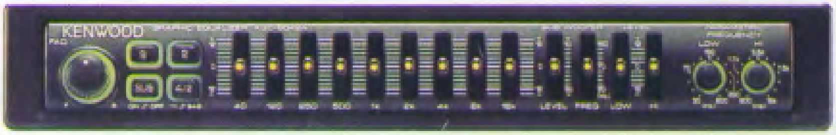

Un ecualizador *contiene un conjunto de filtros de paso de banda/eliminación de banda de atenuación o amplificación regulable.*

Filtros y formas de onda

Las formas de onda no sinusoidales cambian de forma atravesando un filtro

Hasta ahora hemos supuesto la aplicación a la entrada de un filtro una sola señal **sinusoidal**, midiendo su atenuación a la salida.

Empleando una forma de onda distinta de la simple sinusoide, probablemente será muy **distorsionada**.

Recordemos que una onda periódica puede **descomponerse** en una serie de sinusoides múltiplos de la frecuencia fundamental (serie de Fourier), cada una con una frecuencia y fase propias.

Cada una de estas armónicas, como la fundamental, tiene una frecuencia distinta, que por tanto se atenúa y desfasa con el filtro de un modo diferente.

EL PASO BAJO REDONDEA LOS ÁNGULOS

Igual que un filtro de paso bajo atenúa las frecuencias altas, las **armónicas** se atenúan más que la fundamental.

Eso significa que la onda resulta **menos angulosa**: una onda cuadrada, por ejemplo, será menos cuadrada y se parecerá más a una sinusoide.

Con mayor razón se atenúan los **picos**: se utiliza a menudo un filtro de paso bajo como protección contra breves impulsos de alta tensión (por ejemplo, descargas electrostáticas).

Un filtro de paso bajo* redondea *las formas de onda, amortiguando los picos de breve duración.

Se puede decir también que el filtro funciona como **integrador**: la salida responde lentamente a las variaciones de entrada, como si tuviera una cierta inercia.

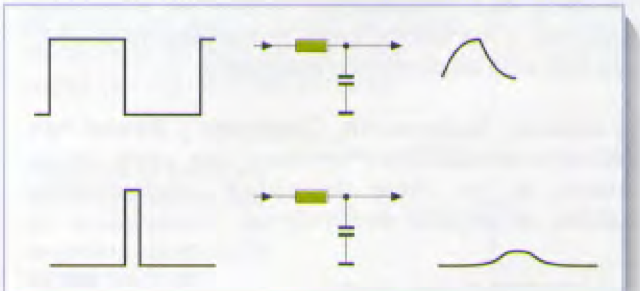

EL PASO ALTO RESALTA LAS VARIACIONES

Por el contrario, un filtro de paso alto atenúa la **fundamental** y las armónicas más bajas, dejando pasar las de orden superior (es decir, las de frecuencias más alta).

El efecto es el de resaltar las variaciones bruscas, haciendo que la forma de onda sea más **angulosa**: por ejemplo, una onda cuadrada se transforma en una serie de impulsos.

Un filtro de paso alto* resalta *las variaciones rápidas de la señal.

El filtro funciona como **diferenciador**: la salida es especialmente sensible a la velocidad de variación de la entrada.

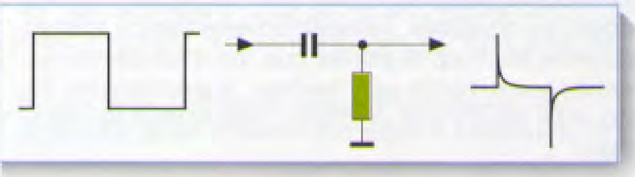

Filtros más eficaces

El uso de inductores y de amplificadores permite realizar filtros de corte más limpio

El filtro RC tiene dos defectos: empieza a atenuar **demasiado rápido**, mucho antes de la frecuencia de corte, y su efecto no es especialmente drástico.

Se puede mejorar con **varios filtros en cascada**, con el fin de que ninguno de ellos moleste al anterior; con tres sencillos filtros ("de primer orden") se tiene un filtro de tercer orden, que atenía 18 dB por octava.

No obstante, esta técnica no es especialmente eficaz con los filtros RC: presenta varios límites; para superarlos es necesario añadir **inductores**.

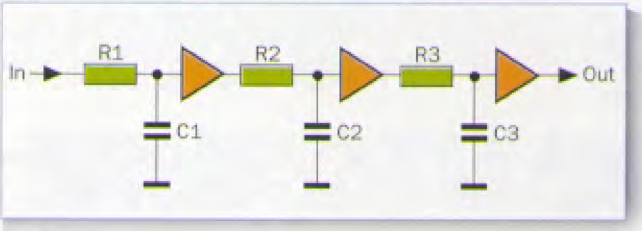

Filtro **multietapas**, *los triángulos representan amplificadores de corriente con impedancia de entrada alta (no molestan al filtro anterior).*

FILTROS EVOLUCIONADOS

El uso combinado de inductores y condensadores permite obtener filtros de características de corte **empinadas**, como el paso de banda o el supresor de banda mostrados en la figura

Se verifican fenómenos de **resonancia**, en los que la impedancia del circuito (o de una parte) cambia rápidamente en torno a una determinada frecuencia.

Los filtros RLC (resistencias, inductores, condensadores) son a menudo bastante **complejos** de diseñar, pero han sido ampliamente estudiados.

En especial, **Butterworth, Chebysev** y **Bessel** han analizado técnicas para optimizar una cierta característica de los filtros (linealidad, empinamiento, desfase) en perjuicio de las demás.

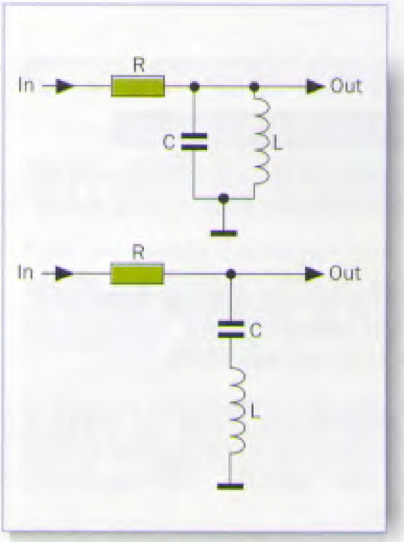

Filtros **RLC** *elementales: el primero es un paso de banda bastante estrecho, el segundo es un supresor de banda.*

FILTROS ACTIVOS

El uso de componentes activos permite **eliminar** los costosos e imprecisos inductores, realizando filtros de buenas características y dimensiones bastante reducidas.

Se utilizan ampliamente; los ilustraremos tras la introducción de los **amplificadores operacionales**, los circuitos integrados para todo que los han hecho sencillos y cómodos. También en este campo existen recientes técnicas **digitales** que, en distintos casos, permiten resultados aún mejores, especialmente en lo que se refiere a la precisión y capacidad de repetición.

Un amplificador operacional permite, entre otras cosas, la realización de filtros activos.

Clases de amplificadores

Examinamos las etapas finales para ilustrar los distintos modos de funcionamiento de los transistores

En el sencillo amplificador de transistores que hemos visto ya en varias ocasiones, la señal de entrada hace variar la **corriente de colector**.

En lugar de la habitual resistencia se puede también poner directamente una **carga**, por ejemplo un altavoz, como en la figura.

En un amplificador de **clase A** como el mostrado, transistor y carga siempre son atravesados por corriente: el transistor nunca se pone en corte.

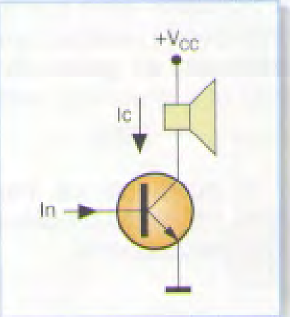

Amplificador de clase A: la corriente de colector I_C aumenta y disminuye, pero nunca es cero.

PROBLEMAS DE LA CLASE A

Es bastante poco práctico realizar un amplificador de potencia de clase A, debido a la excesiva **disipación** sobre transistor y altavoz debida al paso continuo de corriente. Con 50 V de alimentación igualmente repartidos y un altavoz de 4 Ω, pasarían también 6,25 A aún en **ausencia de señal** (corriente de reposo): ¡más de 156 W sobre transistor y altavoz!

La **potencia de audio** sería sólo la mitad: en realidad 50 V_{pp} (pico-pico) son 25 V_p (pico) y por tanto $25/\sqrt{2}$ = 17,7 V_{rms} más o menos, que sobre 4 Ω dan poco más de 78 W.

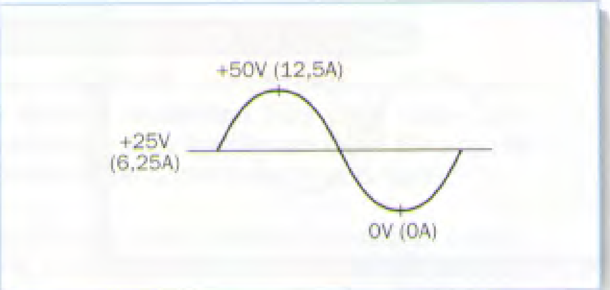

Tensión y corriente en un altavoz de 4 Ω, con V_{cc} = +50 V. La corriente media es la del eje (6,25 A).

PUSH-PULL Y CLASE B

Los problemas de disipación y corriente de reposo se pueden resolver empleando **dos transistores**, como en la configuración push-pull ("empuja-tira") mostrada en la figura.

El transistor superior **conduce** en las semiondas positivas, el inferior en las negativas: la corriente de reposo es cero, y por lo tanto también la disipación.

El condensador en serie con la carga bloquea la corriente continua: por el altavoz pasa solamente la señal de audio amplificada; el amplificador está definido de **clase B**.

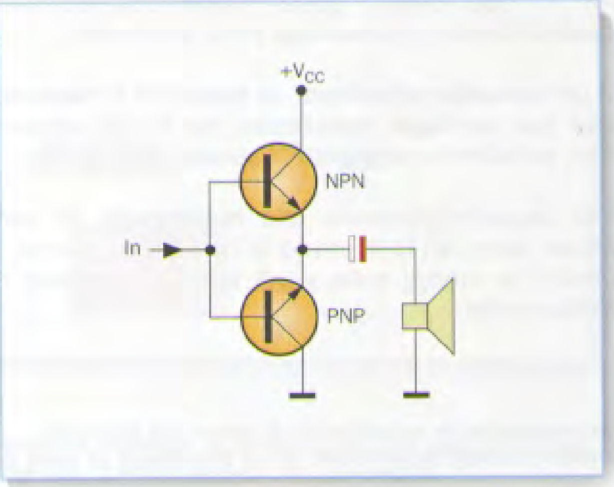

Amplificador push-pull de clase B con transistores complementarios (NPN y PNP): es perfectamente simétrico.

Distorsión y remedios

La señal de salida debe ser todo lo más parecida posible a la de entrada

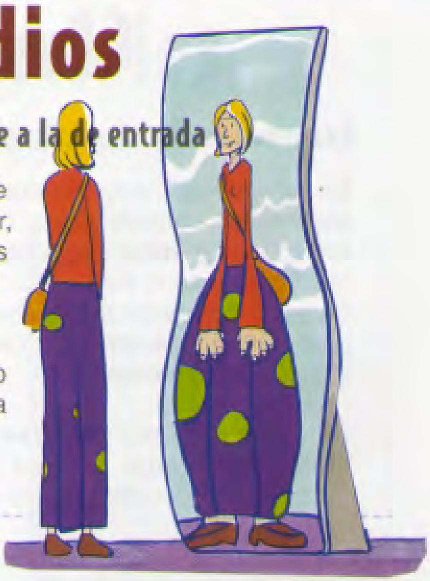

Como la ganancia (ß) de un transistor varía con la corriente (lección 11 de Componentes), la forma de onda puede resultar ligeramente alterada, es decir, **distorsionada**. También la **diferencia de ganancia** entre los dos transistores puede dar problemas, causando una diferencia correspondiente de corriente entre las dos semiondas.

Afortunadamente, como en el push-pull los transistores trabajan como **seguidor de emisor** (ver lección 13), estos problemas se reducen enviando a la base una corriente lo bastante elevada.

DISTORSIÓN DE CROSSOVER Y CLASE AB

No obstante, los transistores no conducen por media semionda exacta: la tensión base-emisor hace que, al paso por el cero de la onda ("crossover"), estén **ambos** en corte.

En ese breve período la carga es **libre**: ya no sigue la señal, especialmente si es inductiva (como un altavoz), y se tiene un escalón en la forma de onda.

Para evitarlo, se puede hacer pasar siempre un poco de corriente en ambos transistores, mucha menos que en la clase A (por ejemplo 30 mA en lugar de 6 A): se habla entonces de **clase AB**.

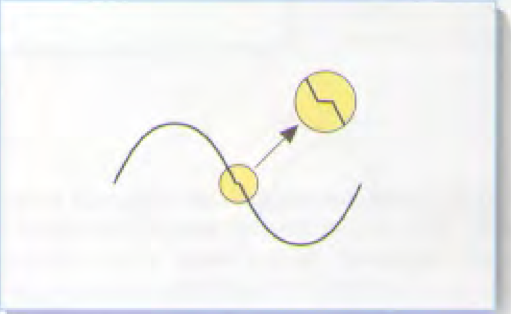

Distorsión de crossover *de clase B, en el cambio de semionda.*

RESISTENCIAS DE ESTABILIZACIÓN

Por desgracia no basta con aplicar una tensión entre las bases: la tensión B-E cambia con la **temperatura**, por lo que podría pasar demasiada corriente (calentamiento) o demasiado poca (distorsión).

SI un transistor se calienta, la tensión B-E necesaria para que conduzca **desciende**, por lo que conduce más, se calienta más... etc., hasta su destrucción.

Esto se evita poniendo una **resistencia** en cada emisor, como en la figura: si la corriente aumenta, la tensión de emisor sube y por tanto el transistor no conduce más.

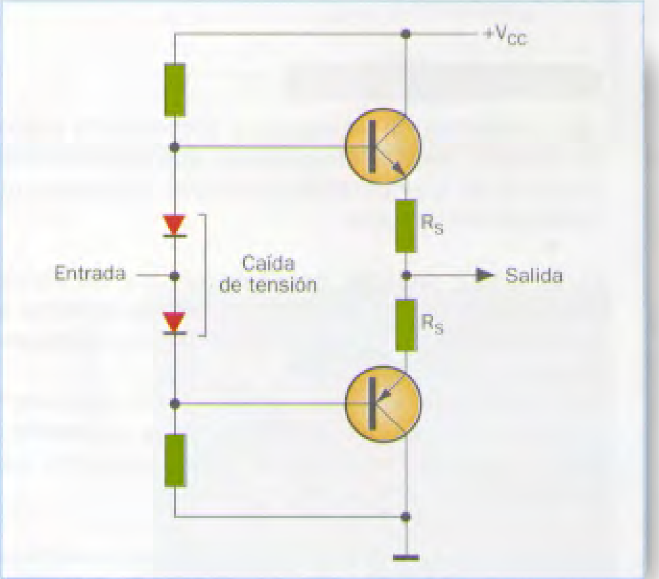

Las resistencias de estabilización R_S evitan una derivación térmica (runaway) incontrolable de los transistores de clase AB.

Cuidados adicionales para las etapas finales

Las etapas de salida pueden dañarse fácilmente: es necesario evitar riesgos

Un **cortocircuito** hacia masa puede pasarle decenas de amperios a un transistor de potencia, destruyéndolo en un instante.

Para **limitar** la corriente máxima se puede añadir un transistor que aproveche la caída de tensión en las resistencias de estabilización, como muestra la figura.

Si pasa **demasiada** corriente, la tensión en R1 hace conducir a T2, que le "roba" corriente de base a T1, impidiéndole conducir después.

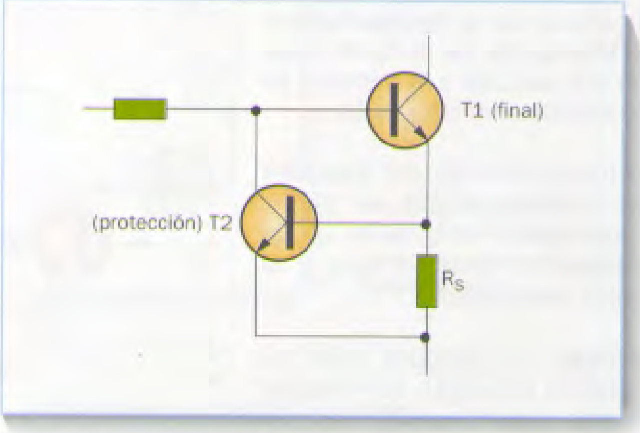

***Protecciones contra excesos de corriente** en un amplificador de clase AB (la mitad de abajo es simétrica).*

CONDENSADOR Y CERO CENTRAL

El condensador en serie con el altavoz debe tener una elevada **capacidad** para dejar pasar otras corrientes de baja frecuencia; debe ser además insensible a las inversiones de polaridad.

Para eliminar este molesto y costoso componente, basta con alimentar el circuito con dos tensiones simétricas (por ejemplo +25 V y –25 V), de modo que el punto central esté a **cero voltios** (masa) en reposo.

Sin embargo es esencial que esta tensión sea totalmente **estable**: para ello se proporciona el mismo sistema de corrección (retroalimentación) que reduce la distorsión a niveles muy bajos.

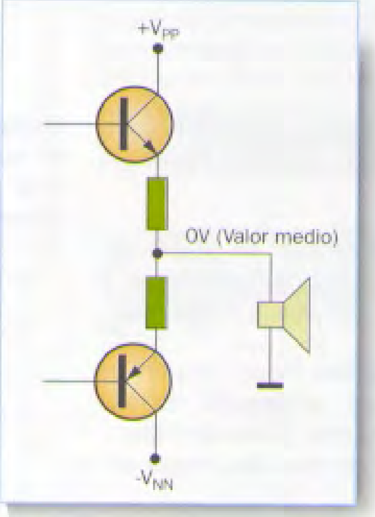

***Etapa final acoplada en continua:** por el altavoz no pasa corriente en reposo, porque la salida está a 0 V.*

OTRAS PROTECCIONES Y CUIDADOS

Normalmente se utiliza un **fusible** en serie con el altavoz para evitarle daños en caso de roturas del amplificador (por ejemplo, el transistor PNP entra en cortocircuito, y la salida va a –25 V).

Una **protección térmica** sobre los transistores finales puede evitar daños debidos a una dispersión insuficiente del calor (por ejemplo, fisuras de ventilación obstruidas). Finalmente, un **relé retardado** en serie con el altavoz lo mantiene aislado en el encendido, hasta que la tensión no esté de nuevo estabilizada en 0 V, evitando un "toc" audible y posibles daños al altavoz.

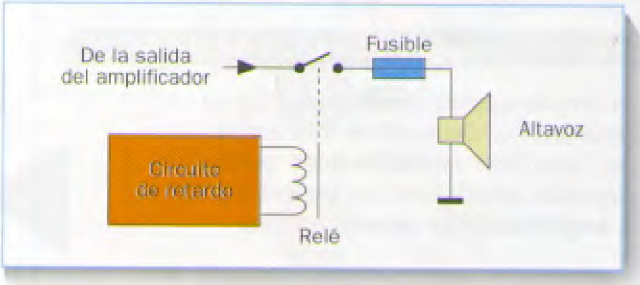

***La inserción retardada** de los altavoces sirve para conectarlos al final sólo cuando a la salida hay cero voltios.*

Amplificadores diferenciales

Volvemos a los circuitos de baja potencia para ilustrar una técnica fundamental de la electrónica moderna

El circuito de la figura amplifica la **diferencia** de tensión entre las dos señales a la entrada: es un amplificador diferencial.

El funcionamiento del transistor de la derecha Q2 es obvio: aumentando la tensión en la entrada I2, conduce más y la salida **desciende**.

También Q1 conduce más subiendo la entrada I1; aumenta por tanto la corriente en RE, y asimismo la tensión del punto común C.

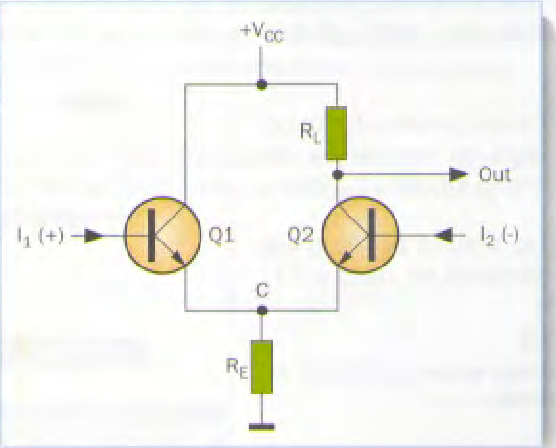

Subiendo C, se reduce sin embargo la tensión entre base y emisor de Q2, que por lo tanto conduce menos: la salida sube también.

Amplificador diferencial: ***las dos entradas tienen un efecto opuesto sobre la salida. I1 está en fase (positivo), I2 en contrafase (negativo).***

MODO COMÚN Y CMRR

Un amplificador diferencial debería ser insensible a las tensiones "de modo común", es decir, aplicadas **juntas** a las dos entradas.

Por desgracia, aumentando la tensión en ambas bases, aumenta la **corriente** en ambos transistores, y por lo tanto la caída sobre R_L: por ello la salida baja.

CMRR (Common Mode Rejection Ratio: proporción de rechazo en modo común), expresada normalmente en dB, es la **relación** entre la ganancia diferencial y la del modo común: cuanto más alta, mejor es.

Si la* CMRR *fuera infinita (caso ideal), con dos entradas idénticas no se tendría ninguna señal a la salida

Mejora aún más sustituyendo R3 por un generador de corriente constante (ver lección 13): la corriente total es fija y sólo puede **dividirse** de manera diferente entre los dos transistores.

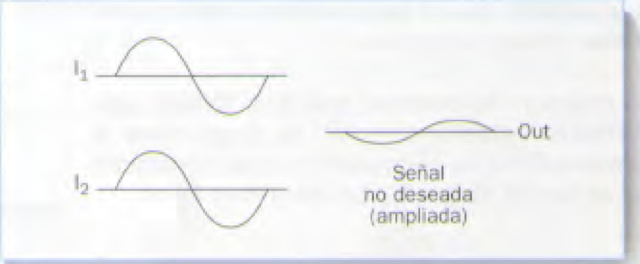

DENTRO DE LOS INTEGRADOS ANALÓGICOS

Los amplificadores diferenciales no se encuentran a menudo en los esquemas, pero son fundamentales en los integrados analógicos, en especial en los **amplificadores operacionales**.

En la práctica son amplificadores diferenciales casi perfectos, ya **esenciales** como los transistores... y a menudo más cómodos: hablaremos a fondo de ellos dentro de algunas lecciones.

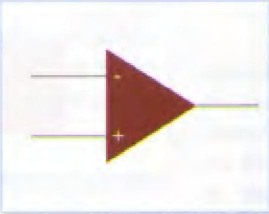

El símbolo de un* amplificador *operacional muestra dos entradas y una salida.

XD925SA1
92E622
X0925SA1
Z992A9622
A635
3V925

Retroalimentación

Reenviando a la entrada parte de la señal de salida, es posible mejorar notablemente el comportamiento de un amplificador

Hemos visto que la **ganancia** de un transistor, y por lo tanto de un amplificador, no puede preverse exactamente, pero depende de las tolerancias de producción.

Más grave es la variación de la ganancia según la corriente: causa, como hemos descrito en la lección 18, una alteración de la forma de onda, es decir, una **distorsión** de la señal.

En los años 20 Harold Black ideó una técnica para reducir al mínimo estos problemas: la **retroalimentación negativa**.

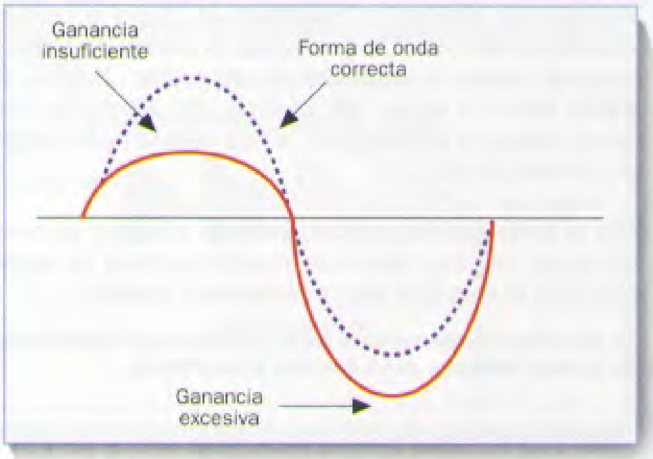

Un amplificador puede **distorsionar** *la señal aplicada a su entrada.*

COMPARAR Y CORREGIR

Si un amplificador tiene una ganancia similar a 10, la salida debe ser de diez veces la entrada, es decir, un **décimo** de la salida debe ser igual a la entrada.

Por tanto, se puede comparar un décimo de la señal de salida con la de entrada, utilizando un **amplificador diferencial**, como muestra la figura.

La posible diferencia, es decir, la **señal de error**, será amplificada por el propio amplificador, corrigiendo la salida hasta que no sea la deseada.

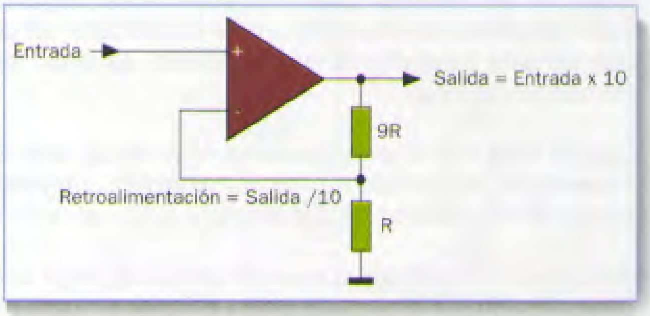

Retroalimentación negativa: *se amplifica la diferencia entre la salida deseada y la efectiva.*

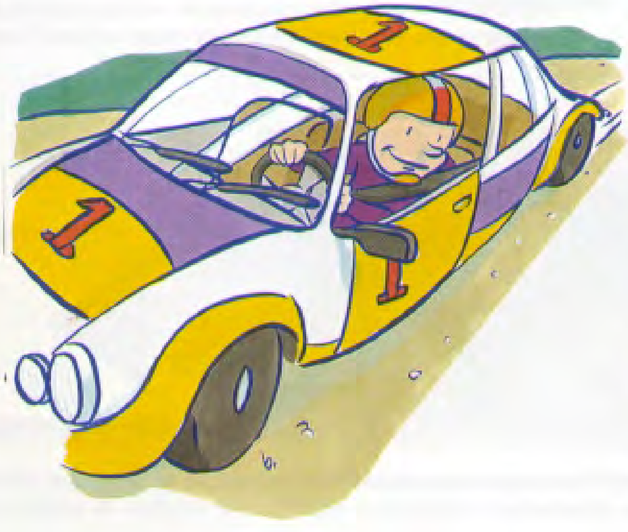

ANILLO ABIERTO, ANILLO CERRADO

Recapitulando, en un sencillo amplificador de **anillo abierto**, es decir, sin retroalimentación, la señal viaja unidireccionalmente desde la entrada hasta la salida.

En un amplificador de **anillo cerrado**, como en la figura de arriba, no se amplifica la señal, sino la diferencia entre la salida deseada y la medida con la retroalimentación.

La retroalimentación es **negativa**, es decir, actúa al revés: una tensión excesiva a la salida debe causar la reducción de la propia tensión (y viceversa).

El piloto controla donde va el coche (salida) y corrige con el volante (entrada): es un ejemplo de retroalimentación negativa.

Ventajas de la retroalimentación

La retroalimentación negativa no sirve sólo para reducir la distorsión, sino que cambia además las características de salida

Además de mejorar la linealidad (es decir, la correspondencia entre salida y entrada), la retroalimentación positiva reduce la **impedancia de salida**. Pero si la salida tiende a variar, por ejemplo por efecto de una fuerte **carga**, el amplificador "se da cuenta" y lo corrige en consecuencia.

Por lo tanto resulta **menos sensible** a las variaciones de carga: es decir, tiene una impedancia baja de salida (como si tuviera una menor resistencia interna).

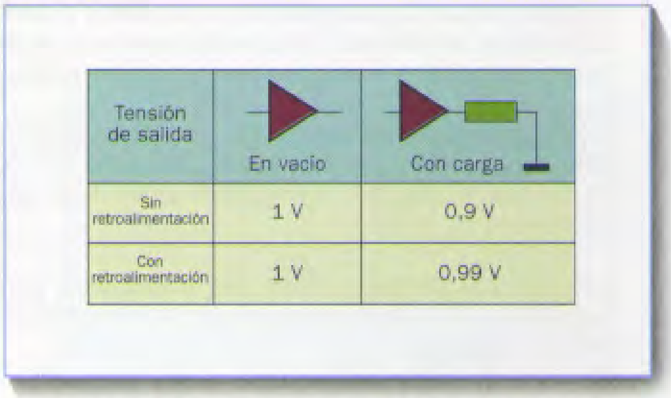

Tensión de salida	En vacío	Con carga
Sin retroalimentación	1 V	0,9 V
Con retroalimentación	1 V	0,99 V

***La retroalimentación negativa** hace la salida menos dependiente de la carga aplicada, por lo que baja la impedancia.*

SE NECESITA MUCHA GANANCIA

Para que las ventajas de la retroalimentación negativa sean tangibles, es necesario que el amplificador tenga una elevada ganancia **de anillo abierto**, es decir, sin retroalimentación.

Cuanto más alta es esta ganancia, más **eficaz** será la intervención del circuito de retroalimentación, y proporcionalmente elevados los beneficios.

Por ejemplo, es normal que un amplificador tenga una ganancia de 10.000 o más (anillo abierto) y que después sea **reducido** a 50 gracias a la retroalimentación. Las características de este amplificador serán, en primera aproximación, 200 veces (10.000 / 50) **mejores** que las de un circuito con ganancia 50, privado de retroalimentación.

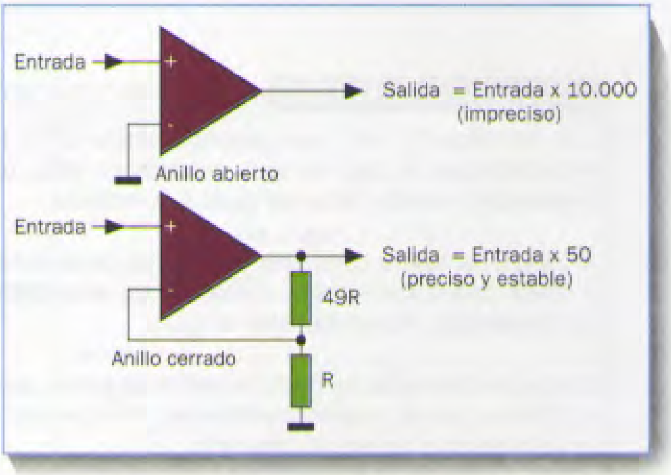
Entrada
+
-
Salida = Entrada x 10.000 (impreciso)
Anillo abierto
Entrada
+
-
Salida = Entrada x 50 (preciso y estable)
49R
Anillo cerrado
R

Una elevada ganancia** de anillo abierto **exalta las ventajas de la retroalimentación.

AMPLIFICADORES OPERACIONALES

Por lo que se ha dicho, parecería que el amplificador ideal debe tener entrada diferencial y ganancia **infinita** (se entiende de anillo abierto, es decir, sin retroalimentación).

Realmente es así, hasta el punto de que se **producen** eficazmente dispositivos que se aproximan a esta condición ideal: son los "amplificadores operacionales".

Estos representan los **ladrillos esenciales** de la moderna electrónica analógica, que examinaremos con detalles más adelante.

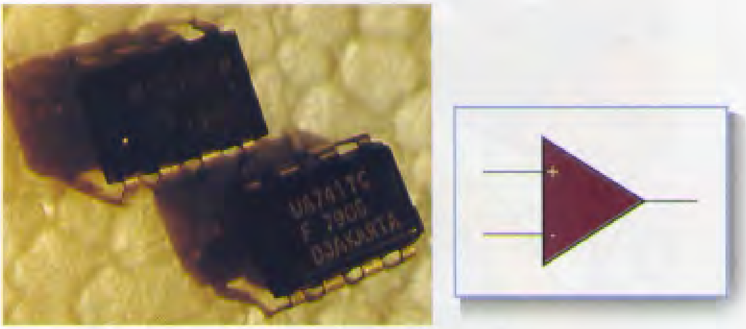

***Un amplificador operacional,** práctico componente de uso universal.*

Estructura de un amplificador

Un amplificador de audio utiliza distintas técnicas ilustradas en estas últimas lecciones

La figura muestra el **esquema de bloques** de un amplificador típico de audio de media potencia, como los utilizados en los equipos de alta fidelidad domésticos.

El **amplificador diferencial** compara la señal de entrada con una fracción de la de salida, para obtener la retroalimentación negativa.

El **preamplificador** tiene la misión de proporcionar gran parte de la ganancia total, y puede estar formado por varias etapas.

Finalmente, la etapa de **control** gobierna la **final**, que se ocupa de controlar la fuerte corriente requerida por la carga.

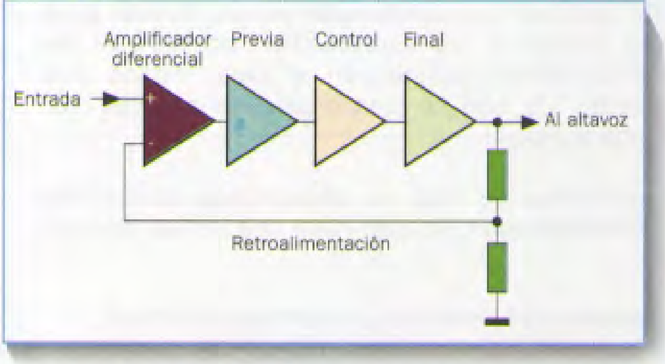

Amplificador diferencial
Previa
Control
Final
Entrada
Al altavoz
Retroalimentación

Esquema de bloques de un amplificador de audio típico.

CIRCUITOS YA VISTOS

En el esquema de amplificador mostrado en la figura se pueden distinguir las distintas **etapas** descritas en el esquema de bloques de la figura anterior.

El **amplificador diferencial** es similar a los ilustrados en la lección 18; por la máxima estabilidad, el condensador de la derecha hace que la retroalimentación en continua sea total (ganancia = 1).

El preamplificador utiliza un transistor en el clásico circuito de **emisor común**, descrito en la lección 11, el control y los finales están en configuración **push-pull** en clase AB, vista en la lección 18.

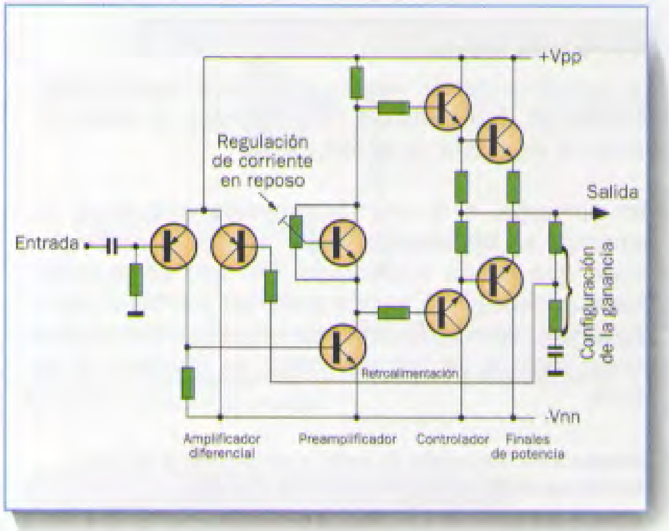

+Vpp
Regulación de corriente en reposo
Salida
Entrada
Configuración de la ganancia
Retroalimentación
-Vnn
Amplificador diferencial
Preamplificador
Controlador
Finales de potencia

Las distintas etapas de un amplificador en un esquema real (simplificado quitando las protecciones y los desacoplamientos.

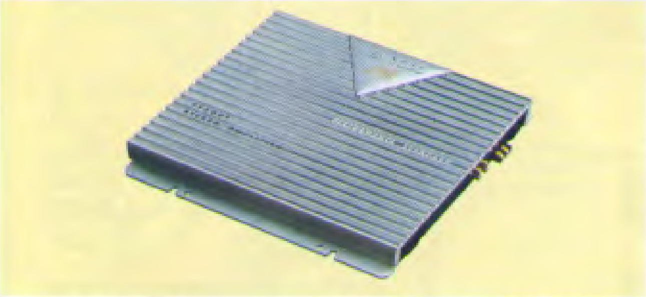

Cambia la técnica de realización, pero la estructura de un amplificador de audio es la misma que hace veinte años.

AMPLIFICADORES MODERNOS

El amplificador diferencial de transistores se sustituye frecuentemente por un **amplificador operacional** de circuito integrado más preciso.

Los **transistores finales** son a menudo del tipo MOSFET en lugar de unión (BJT): no cambia la estructura del circuito ni la clase de funcionamiento.

Finalmente existe la tendencia a encerrar una parte cada vez más grande del circuito en el interior de un solo **circuito integrado**, utilizando un único fragmento de silicio en lugar de componentes diferentes ("discretos").

Respuesta en frecuencia

La retroalimentación es fundamental para garantizar una respuesta "plana" en los amplificadores

La ganancia de un amplificador suele variar con la frecuencia de la señal: se puede trazar un gráfico de la **respuesta en frecuencia**, como en la figura.

En realidad, los condensadores que forman parte del circuito y las capacidades "ocultas" en los componentes contribuyen a crear **filtros** (ver lección 17), con un comportamiento también bastante complejo.

Las formas de onda no sinusoidales son por ello **distorsionadas**, ya que cada una de las distintas armónicas que las componen (ver lección 16) sufre una amplificación diferente.

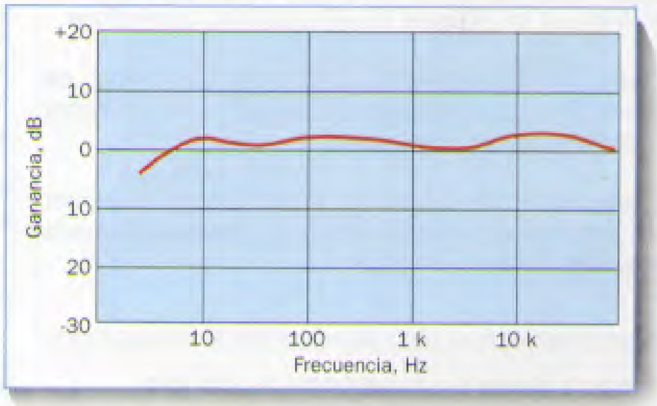

Respuesta en frecuencia ***de un amplificador; observe la escala logarítmica en decibelios (dB).***

NIVELAR CON LA RETROALIMENTACIÓN

La retroalimentación negativa resuelve muchos problemas, ya que **controla** con precisión la señal de salida: si es excesiva, la reduce.

Sin embargo, si a una determinada frecuencia la ganancia es **demasiado baja**, no hay retroalimentación que pueda aumentarla. Por eso es esencial que la ganancia sea lo más **superior** posible al valor requerido, con el fin de que la retroalimentación pueda realizar su trabajo, como se muestra en la figura.

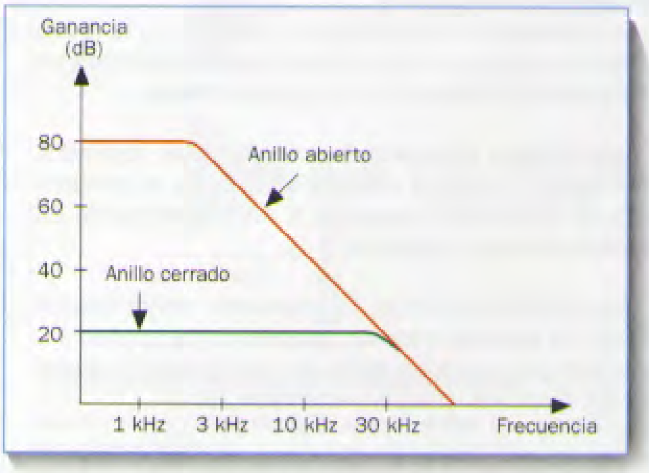

Respuesta en frecuencia de anillo abierto (rojo) y de anillo cerrado, es decir, con retroalimentación (verde).

RESPUESTA A MEDIDA

La retroalimentación negativa puede utilizarse, en lugar de para hacer uniforme la respuesta en frecuencia, para **alterarla** del modo deseado.

Si por ejemplo se añade a la retroalimentación un **filtro de paso alto RC**, como en la figura, a frecuencias bajas el porcentaje de la señal que vuelve a la entrada será inferior. Para compensar, el amplificador diferencial **aumentará** la señal de salida: se tendrá una mayor amplificación de los bajos.

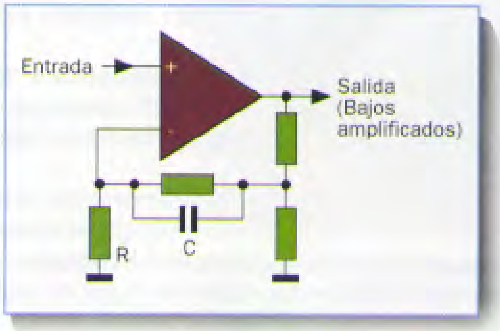

Proyectando con precisión el filtro sobre la retroalimentación, es posible obtener prácticamente **cualquier** curva de respuesta en frecuencia deseada.

Modificando el comportamiento en frecuencia de la retroalimentación, la del amplificador variará de un modo recíproco.

Osciladores

En muchas aplicaciones es necesario producir una tensión alterna de características controlables

La única **tensión alterna** que se puede encontrar es la de red, sinusoidal a 50 Hz más o menos; si se necesita una frecuencia distinta o forma de onda, es necesario producirla.

Un circuito capaz de generar una onda repetitiva, de cualquier frecuencia o forma, se llama **oscilador**, por analogía con la oscilación de un periodo.

Dado que la alimentación suele ser **continua**, un oscilador puede considerarse también un dispositivo capaz de convertir tensión continua en tensión alterna.

EFECTO LARSEN

Producir una oscilación "partiendo de cero" puede parecer difícil, pero en realidad basta un **amplificador** conectado a un micrófono y un altavoz.

Aproximando el micrófono al altavoz, se produce un fuerte **silbido** (efecto Larsen): se trata de una oscilación.

Hay además una retroalimentación (o "reacción") **positiva**: la salida se reenvía a la entrada, desde aquí se amplifica de nuevo, se envía otra vez a la entrada, etc.

Muchos osciladores actúan realmente así: **reenviando a la entrada** parte de la señal de salida.

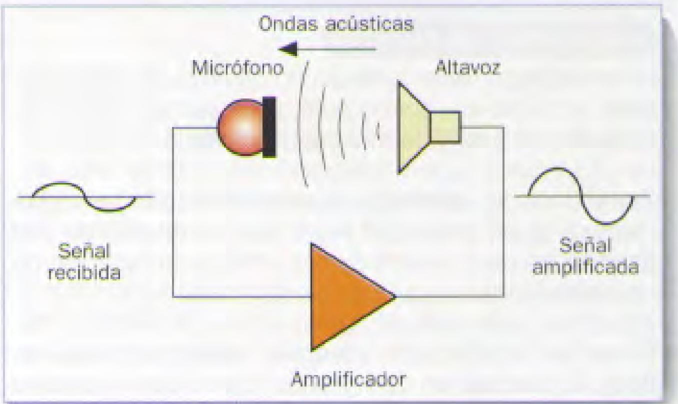

Reenviando a la entrada, con un micrófono, el sonido producido por el altavoz, se tiene una oscilación.

APLICACIONES DE LOS OSCILADORES

La **sirena** de un antirrobo y el "bip" de un reloj utilizan oscilaciones de frecuencia audible, normalmente entre los 800 y los 3.000 Hz, donde el oído es más sensible.

Un **transmisor de radio** o un teléfono móvil producen una onda electromagnética (lección 8 de Aplicaciones), partiendo de la señal eléctrica generada por un oscilador.

Numerosos circuitos **digitales**, del multiplexor de un display (lección 19 de Digital) a un ordenador personal, requieren una señal repetitiva.

Sin osciladores, gran parte de los dispositivos electrónicos no existirían. Nos ocupamos por ahora de los osciladores de **baja frecuencia.**

Los instrumentos musicales son osciladores: la onda de presión alterna es producida por la vibración de una cuerda, un estrangul, una columna de aire, etc.

Desfase de 360°

Para realizar un oscilador sinusoidal es necesario controlar la fase de la señal que vuelve a la entrada

Reenviar a la entrada parte de la salida **no basta** para crear un oscilador, de otro modo los amplificadores con retroalimentación negativa (lección 19) oscilarían.

Para los osciladores sinusoidales existe el "criterio de Barkhausen" que, cálculos matemáticos aparte, podemos resumir así: la señal de salida debe **recrearse exactamente** a sí misma.

Una vez vuelta a la entrada y reamplificada, debe tener **la misma amplitud**: la atenuación externa debe ser por tanto igual a la ganancia del amplificador.

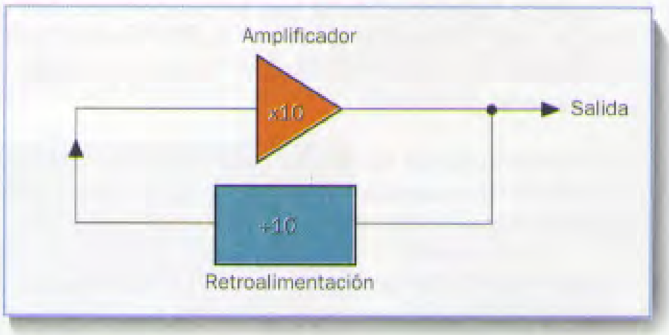

Oscilador ***compuesto por un amplificador y un circuito de retroalimentación positiva.***

FASE: UN GIRO ENTERO

La señal debe tener también un desfase de **cero grados**, es decir, un retardo nulo, pero eso es imposible: atravesar el amplificador requiere tiempo.

Tratándose sin embargo de una onda que se repite siempre igual, basta con llegar con un retardo de **360 grados**, es decir, exactamente similar a la duración de un ciclo.

Como un amplificador **inversor** vuelca la onda, es decir, la desfasa en 180 grados, basta que el circuito de retroalimentación introduzca los otros 180.

La figura muestra un oscilador construido con un **filtro RC de tres etapas**, calculado para introducir un retardo de 180° a la frecuencia deseada.

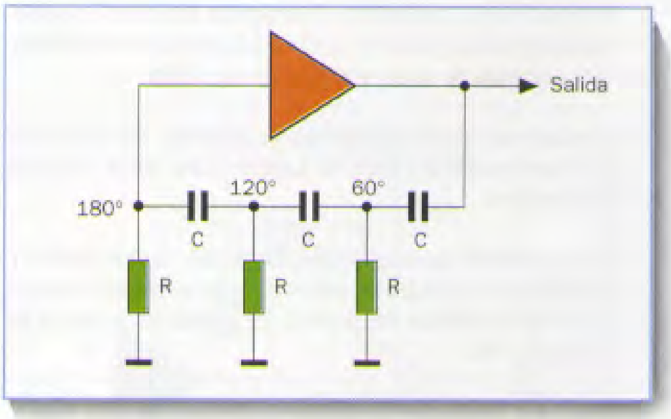

Oscilador sinusoidal: ***los tres circuitos RC desfasan la señal de 180° a la frecuencia deseada.***

AMPLITUD Y ESTABILIDAD

Un oscilador sinusoidal es **fundamental**: si además la ganancia global baja aunque sea muy poco por debajo de la unidad, la oscilación de detiene.

Si por otra parte la ganancia está ligeramente por encima de 1, la señal de salida se hace siempre más amplia hasta que el oscilador **distorsiona**, perdiendo la forma sinusoidal.

Por ello es necesario utilizar circuitos de **regulación automática** de la ganancia, como el que muestra la figura, bastante curioso.

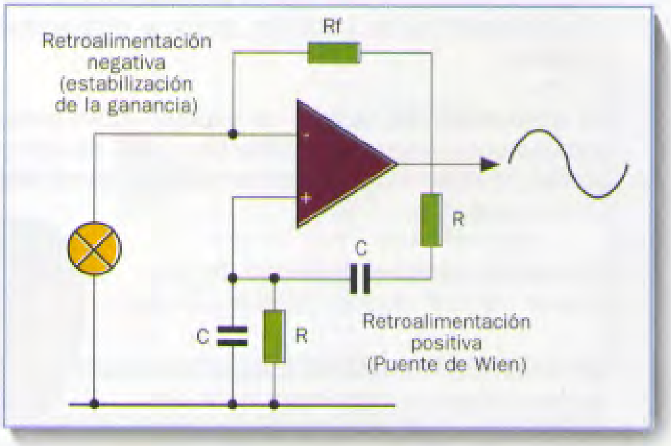
Retroalimentación negativa (estabilización de la ganancia)
Rf
-
+
R
C
C
R
Retroalimentación positiva (Puente de Wien)

Oscilador sinusoidal estabilizado: con señales fuertes la **bombilla** ***se enciende y su resistencia aumenta, reduciendo la ganancia.***

Otros osciladores

No siempre es necesario o conveniente una onda sinusoidal

Si no se desea una **nota pura**, es decir, una sinusoide, hay muchas otras técnicas para construir un oscilador.

Por ejemplo, el oscilador **trigger Schmidt** descrito en la lección 17 de Digital produce tanto una onda cuadrada como la señal que muestra la figura.

Recordemos su **funcionamiento**: C se carga mediante R; cuando la tensión alcanza el umbral del trigger Schmidt, la salida cambia de estado y el proceso se repite con polaridad opuesta.

Oscilador trigger **Schmidt***: produce dos ondas de distinta forma.*

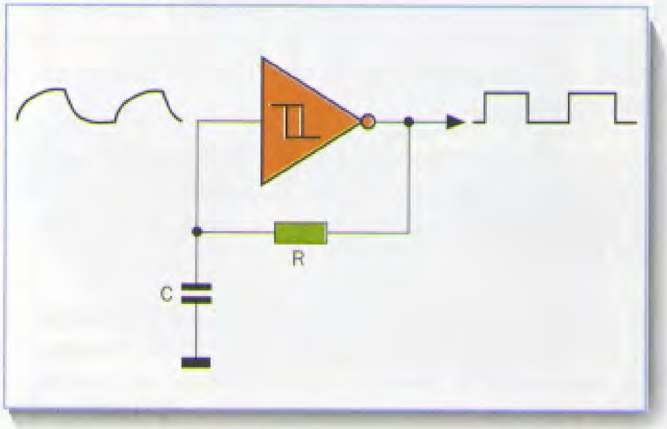

RAMPA Y TRIANGULAR

Con un generador de corriente constante, es posible cargar un condensador **linealmente**, es decir, de modo que la tensión aumente proporcionalmente con el paso del tiempo.

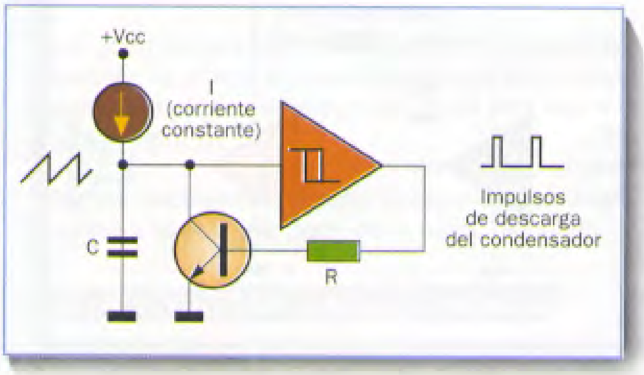

Añadiendo un circuito que **descargue** el condensador, por ejemplo un transistor y una resistencia, se puede producir una rampa (o "diente de sierra"), como en la figura.

Si además la **descarga** se efectuara a corriente constante, similar a la de la carga, se obtendría una forma de onda triangular.

Estos circuitos se realizan normalmente (y con más sencillez) con **amplificadores operacionales**, como veremos posteriormente en el curso.

Oscilador que produce una rampa, cargando un condensador a corriente constante y descargándolo rápidamente.

OSCILADORES INTEGRADOS

En la práctica, cuando se necesita un oscilador, se recurre casi siempre a un circuito integrado **especializado**, al menos a frecuencias relativamente bajas.

Por ejemplo, el difundido "**555**" contiene todos los componentes para la realización de osciladores de distinto tipo; por su popularidad les dedicaremos más adelante una lección entera. Todavía más cómodos son los osciladores **sinusoidales** ya fabricados, en algunos de los cuales la frecuencia puede controlarse también mediante una tensión externa (VCO: Voltage Controlled Oscillator, oscilador controlado mediante voltaje).

Producir una señal sinusoidal de buena calidad no es fácil: conviene utilizar un integrado ya preparado.

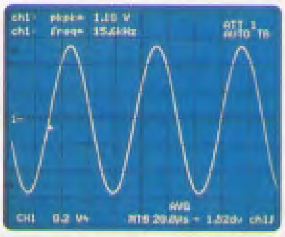

Autooscilaciones

Cuando la retroalimentación negativa se convierte en positiva, un amplificador se transforma en un oscilador

Un fenómeno similar al **efecto Larsen** se puede verificar también en el circuito de un amplificador de audio, si accidentalmente se crea una retroalimentación positiva.

Puede deberse, por ejemplo, a un **acoplamiento capacitivo**, como un cable de entrada que pasa cerca de un cable de salida, y capta parte de la señal que lleva.

Se provoca por tanto, especialmente si la ganancia es alta, una **autooscilación**, en general de frecuencia relativamente elevada (normalmente en campo ultrasónico, es decir, por encima de los 20 KHz). Esto puede **dañar** los tweeter (ver lección 17 de Aplicaciones) y el propio amplificador, además de producir una distorsión de la señal normal.

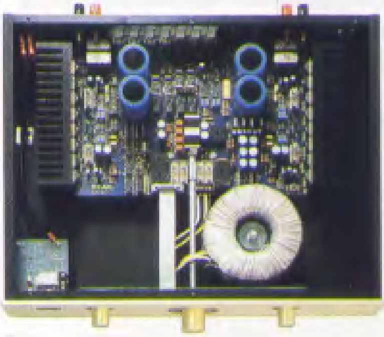

En los **cableados internos** *de un amplificador de audio, entradas y salidas se deben tener bien separadas para evitar oscilaciones*

ACOPLAMIENTOS IMPREVISTOS

Aunque los cables de entrada están lejos de la salida, y quizá apantallados, hay **otras vías** que pueden hacer retroceder a la señal y crear oscilaciones "residuales". Por ejemplo, la **absorción de corriente** de la etapa final causa variaciones de la tensión de alimentación que, si la misma alimenta las etapas anteriores, se pueden transferir a la señal.

Se evita el inconveniente **desacoplando** las primera etapas, es decir, filtrando con precisión la alimentación, de modo que permanezca estable.

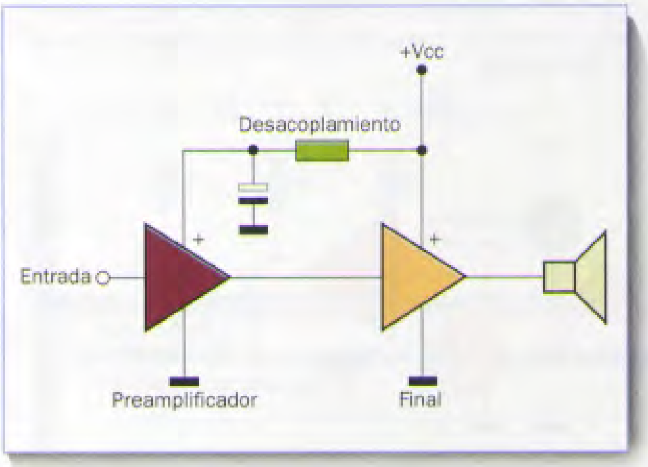

Desacoplamiento ***de la alimentación: filtrándola, se evita que pueda causar autooscilaciones.***

EL RETORNO DE MASA

Todavía más molesto es el acoplamiento causado por un retorno de masa **común** para la etapa final y las anteriores.

Como se ve en la figura, la corriente de la etapa final, **atravesando** la resistencia (y la inductancia) del cable de retorno, produce una débil tensión.

La masa de las etapas preamplificadoras es por tanto **desplazada**, obteniendo el mismo efecto que si se hubiera aplicado una señal a la entrada. Se evita el problema utilizando cables de retorno separados.

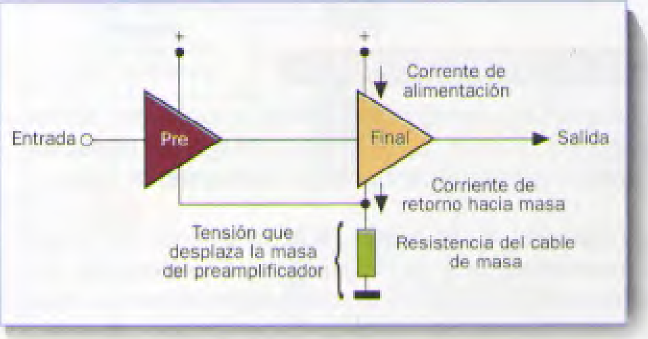
+
+
Corrente de alimentación
Entrada
Pre
Final
Salida
Corriente de retorno hacia masa
Tensión que desplaza la masa del preamplificador
Resistencia del cable de masa

Conexión equivocada de masa: la corriente del final desplaza la masa del preamplificador.

JFET

Los transistores de efecto de campo (FET: Field Effect Transistor) existen en dos versiones. Empezaremos por la de unión o JFET

Si el **principio** de funcionamiento de los transistores BJT clásicos es bastante complejo, es de los JFET es sin embargo sencillísimo.

Existe un **canal** semiconductor por el que discurre corriente, al lado del cual se encuentra una unión, similar a la de un diodo.

Polarizando **inversamente** esta unión, los portadores se alejan de la propia unión, deformando el canal como si fuera un tubo de goma (ver figura).

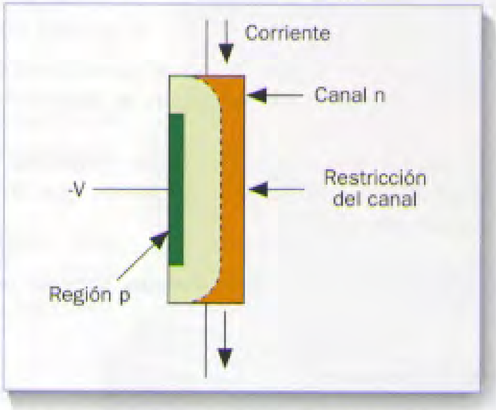

Aplicando tensión inversa más allá de la unión, el canal es restringido.

TERMINALES DE UN JFET

En lugar de emisor, base y colector hay **fuente** (O), **drenaje** (D) y **puerta** (G), como puede verse en la figura.

Fuente y drenaje son los extremos del canal, mientras que la puerta funciona como **electrodo de control**; hasta aquí hay un cierto parecido con el comportamiento del transistor BJT normal. También los JFET existen en dos polaridades: de **canal N** (similares a los transistores NPN) y de **canal P** (similares a los PNP).

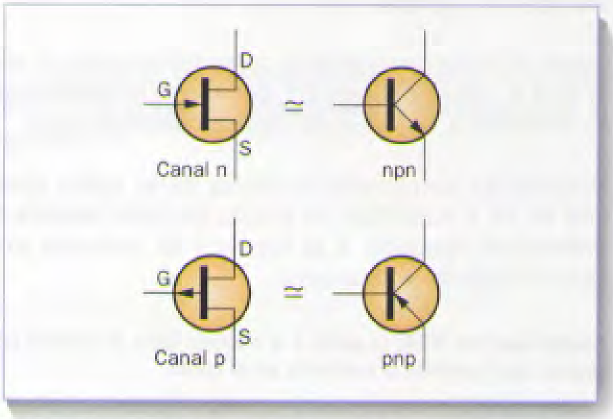

Símbolos de los JFET de canal N y P, y correspondencia (aproximada) con los transistores BJT.

CONTROL SIN CORRIENTE

La diferencia fundamental entre JFET y BJT es que los primeros se controlan con la **corriente** de base, mientras que en los segundos también cuenta la **tensión** de puerta.

La unión, representada con la flechita en el símbolo, está normalmente polarizada inversamente: lo que significa que **no pasa corriente** por la puerta.

Además, en **ausencia** de tensión en la puerta, por el canal pasa corriente, al contrario de lo que ocurre con los BJT normales.

Aplicando tensión inversa a la puerta, por ejemplo negativa en el caso de un JFET de canal N. se tiene un progresivo **vaciado** del canal, hasta su bloqueo completo.

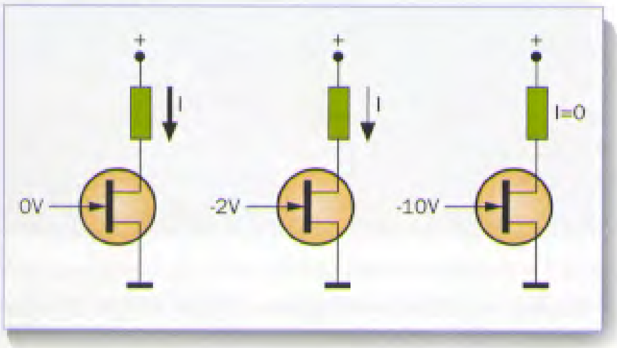

Aumentando la polarización inversa (opuesta a la flecha) en la puerta, se reduce la corriente en el canal hasta anularla.

Polarización de los JFET

Se puede hacer que un JFET produzca por sí solo la tensión inversa que necesita

El circuito de la figura es un **generador de corriente constante** que, a pesar de la simplicidad, funciona perfectamente.

Como no hay tensión alguna en la puerta, por el canal pasa corriente, que, atravesando la resistencia R, produce una **tensión**: el origen se hace positivo con respecto a masa.

Pero como la puerta está a masa, se encuentra **negativo** con respecto al origen: es como si hubiéramos aplicado una tensión negativa, que limita la corriente en el canal (a un valor que depende del JFET).

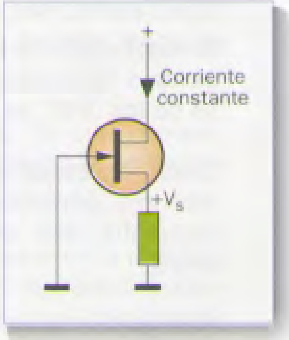

Generador de corriente constante de JFET:* se autopolariza, *haciendo que la puerta sea negativa con respecto al canal.

UN AMPLIFICADOR DE JFET

La figura muestra una etapa amplificadora completa, similar al clásico circuito de **emisor común** utilizado con los BJT (lección 11).

Desde el punto de vista de la corriente continua, la puerta está a masa mediante R1, por lo que la **polarización** es automática, como se ha descrito anteriormente.

Durante las semiondas positivas de la señal alterna que se va a amplificar, la puerta se hace **menos negativa** con respecto a la fuente y la corriente en el canal aumenta, y viceversa.

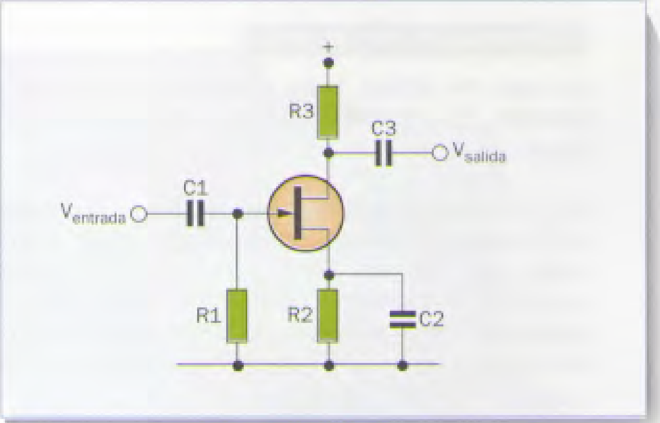

Amplificador de JFET; la señal a la entrada varía la tensión en la puerta, que controla la corriente en el canal.

ALTA IMPEDANCIA

Como la puerta no requiere corriente, el circuito examinado tiene una **impedancia de entrada** muy elevada: se comporta por tanto como una resistencia de alto valor (en la práctica, la entrada sólo ve R1).

Por el contrario, la **amplificación** es bastante baja y la impedancia de salida es relativamente alta, es decir, no puede proporcionar mucha corriente a la carga.

Los JFET son por tanto adecuados allí donde sea necesario mantener bajas las **corrientes** en juego, por ejemplo para no molestar una señal que se vaya a amplificar.

JFET y BJT tienen características en cierto sentido **complementarias**, que los hacen adecuados para tareas diversas, quizá en partes distintas del mismo circuito.

ENGLAND 761B
TL084CN

Muchos amplificadores integrados utilizan JFET en la etapa de entrada, y BJT en las etapas posteriores.

Transconductancia

Un nombre complicado para una cosa sencilla: la "ganancia" de los JFET

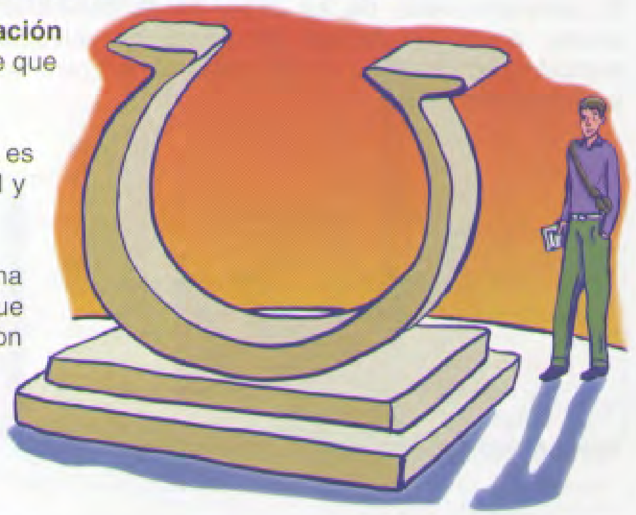

En los transistores de unión, o BJT, la ganancia es la **relación** entre una variación de la corriente de colector y la de base que la causa (lección 10).

En los JFET, al no haber **corriente** de puerta, la ganancia es la relación entre una variación de corriente (I) en el canal y la de la **tensión** (V) aplicada a la puerta.

Un I/V es el contrario de una resistencia, es por tanto una **conductancia**, mejor dicho, "transconductancia", porque está el JFET por medio; se mide en Siemens (S), que son ohmios invertidos (1/Ω).

Si por ejemplo (en determinadas condiciones) al variar la tensión de puerta de 1 V, la corriente en el canal varía en 4 mA, la transconductancia **g** es de 4 mA / 1 V = 4 mS.

PROBLEMAS DE LINEALIDAD

Si la señal a la entrada es amplia, el comportamiento del JFET no puede ya considerarse lineal: la ganancia **varía** (como por otra parte los BJT, lección 18), causando distorsiones.

Si la tensión inversa en la puerta supera un determinado valor (**pinch-off** o "cutoff"), el canal es interrumpido: la corriente baja a cero y no responde a posteriores variaciones de la propia tensión.

En el extremo opuesto, el canal no puede entrar en **conducción** mucho tiempo, porque la unión de puerta no se polariza nunca directamente.

Si es necesario un comportamiento lineal, se pueden adoptar las técnicas de **realimentación** descritas en la lección 19.

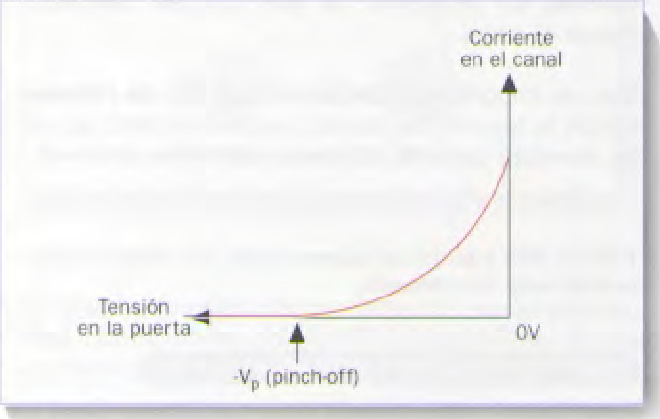

La curva no es lineal: la ganancia de un JFET *varía según la corriente del canal.*

USO DE LOS JFET

Aunque los JFET se utilicen como amplificadores donde sea necesaria una **impedancia alta** de entrada (es decir, poca absorción de corriente), tienen muchas otras aplicaciones.

Son especialmente útiles como **interruptores** y como **resistencias variables**, como veremos en la próxima lección.

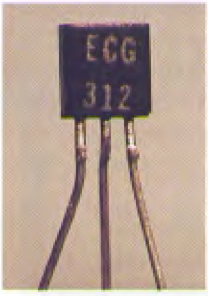

Sus primos MOSFET, además, se emplean en los controles de **potencia** y han monopolizado prácticamente la electrónica **digital**.

Los JFET tienen el aspecto de transistores normales: se reconocen solamente por las siglas.

Construcción y defectos

Los JFET se fabrican de modo similar a los transistores BJT, en un cristal de silicio

El comportamiento de los JFET es curiosamente similar al de los **tubos de vacío** o "válvulas termoiónicas", pero su construcción es la típica de los semiconductores.

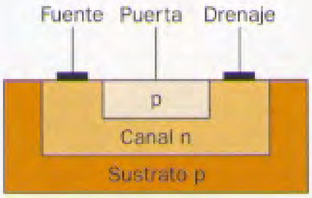

La figura muestra un ejemplo de **estructura** de un JFET de canal N, que puede ser también parte de un complejo circuito integrado.

Es bien visible el **canal** (dopado N) que conecta drenaje y fuente, con la puerta (dopada P) en medio y la unión correspondiente.

Un JFET es en teoría **simétrico**, pero en la práctica fuente y drenaje se construyen de manera diferente para optimizar el comportamiento en el uso normal (como ocurre con los BJT).

Estructura de un JFET de canal N visto en sección: a diferencia de los BJT, ninguna unión separa fuente y drenaje.

VARIABILIDAD DE PRODUCCIÓN

Los JFET tienden a ser bastante distintos de un ejemplar a otro del mismo modelo, incluso más que lo que ocurre con los transistores de unión.

En particular cambia, incluso en varios voltios, la **tensión de pinch-off**: la que bloquea completamente el canal.

Esto es importante, porque es el **punto de referencia** de la tensión de puerta: no importa tanto su valor absoluto como la diferencia con la de pinch-off.

Entre un JFET y el otro del mismo modelo, las características pueden variar notablemente.

TAMPOCO LOS JFET SON PERFECTOS

En teoría la corriente de puerta es cero, pero en la práctica hay una **mínima** corriente de pérdida, como en un diodo inversamente polarizado (ver lección 8 de Componentes).

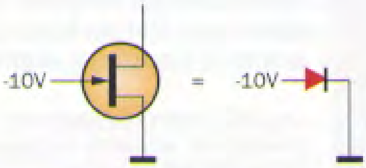

Además es difícil obtener una **conducción** perfecta del canal, entre fuente y drenaje comparable a la de los transistores BJT normales (entre colector y emisor).

Los JFET tienden a mantener en cualquier caso una cierta **resistencia**, que los hace inadecuados para fuertes corrientes (el problema se nota menos en los MOSFET).

La corriente inversa de puerta es muy baja, normalmente en torno al nA, pero varía con la temperatura.

MOSFET

Los FET de canal aislado son más prácticos y versátiles que los de unión

Es posible realizar un transistor de **efecto de campo** (FET: Field Effect Transistor) también sin la unión puerta-canal característica de los JFET.

Se obtiene **aislando** la puerta del canal con un sutil estrato de bióxido de silicio, material de resistencia elevadísima.

Un FET realizado con esta **tecnología MOS** (metal-óxido-semiconductor) toma el nombre de MOSFET: su estructura se muestra en la figura.

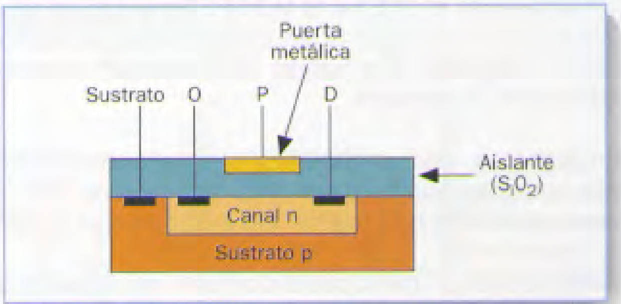

Estructura simplificada *de un MOSFET: la puerta metálica está muy próxima al canal conductor, pero perfectamente aislado de él.*

SÍMBOLO Y SUSTRATO

La figura muestra el **símbolo genérico** de un MOSFET, en el que se ha indicado claramente el aislamiento entre la puerta y el canal: tampoco en el diseño se tocan.

Hay en realidad un **cuarto terminal**, el del sustrato (B, body: cuerpo) en el que el dispositivo se ha realizado: forma una unión con el canal.

En las aplicaciones normales está a menudo **conectado** al origen, y a veces no está indicado; en algunos casos no puede ni tan siquiera accederse a él a través de un pin, al menos en este tipo de MOSFET.

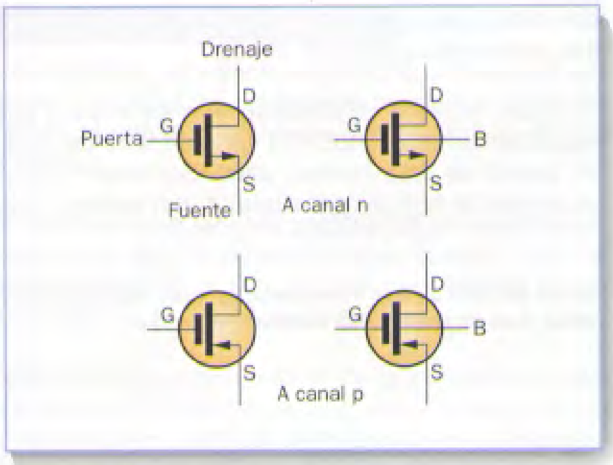

Algunos de los símbolos utilizados para los MOSFET de canal y de canal p; a la derecha la versión indicando la conexión al sustrato (B).

SENSIBILIDAD ELECTROSTÁTICA

Como la puerta está aislada, la corriente requerida por el MOSFET en este terminal es **casi nula**: no hay ni siquiera la mínima corriente de pérdida de la unión de los JFET.

En compensación, basta una mínima carga eléctrica (como la creada desplazando una bolsa de plástico) para **perforar** el sutilísimo estrato aislante, dañándolo irreparablemente.

Son necesarias, por tanto, **precauciones** especiales para manejar o transportar MOSFET: por ejemplo descargarse continuamente a tierra, o mejor utilizar tapetes y brazaletes conductores especiales.

No es necesaria la energía de un rayo: **la electricidad estática** *acumulada por el cuerpo es más que suficiente para perforar el aislante de un MOSFET.*

Vaciado e incremento

Existen dos tipos de MOSFET: uno reduce la corriente en el canal y el otro la aumenta

Consideremos el JFET y el MOSFET mostrados en la figura, ambos de canal N: en ambos casos, aplicando tensión **negativa** a la puerta se "obstruye" el canal, reduciendo la corriente.

Sin embargo, con el MOSFET se puede hacer también una cosa que para los JFET está vetada: aplicar tensión **positiva** a la puerta con respecto a la fuente.

Mientras el JFET haría pasar corriente por la unión de puerta, el MOSFET no hace otra cosa que **aumentar** la corriente en el canal: la puerta se queda aislada.

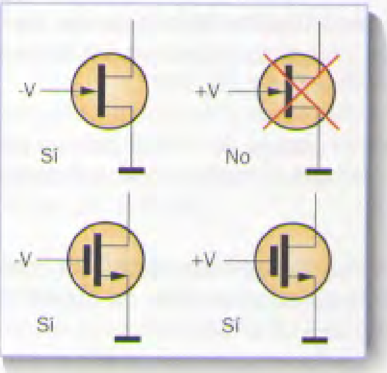

A diferencia del JFET, con el MOSFET es posible aplicar tensiones de ambas polaridades en la puerta.

CORRIENTE SÓLO BAJO PETICIÓN

El comportamiento de los JFET y del tipo de MOSFET descrito, se denomina **depletion** (vaciado): normalmente por el canal pasa corriente, aplicando tensión inversa a la puerta se reduce.

También existen MOSFET del tipo **enhancement** (incremento), por los que normalmente no pasa corriente; para que pase es necesario aplicar tensión a la puerta con polaridad directa.

La figura muestra el **símbolo** empleado para indicar explícitamente un MOSFET de este tipo, de canal N: el canal se interrumpe para demostrar que, en ausencia de tensión en la puerta, permanece abierto.

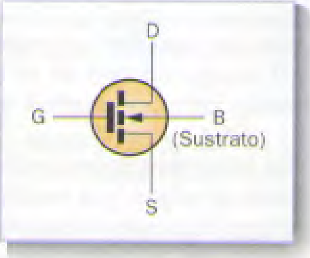

En los MOSFET de tipo incremento, no pasa corriente por el canal si no se proporciona tensión a la puerta.

UN SOLO DISPOSITIVO

Los MOSFET de los dos tipos son distintos, pero el **comportamiento** es similar, como se ve en el gráfico de la figura, que muestra la corriente en el canal al variar la tensión de puerta.

Difieren por la **tensión de umbral**: si es positiva, siempre refiriéndose a un MOSFET de canal N, no pasará corriente con 0 V de puerta. Se tratará por tanto del tipo incremento.

Una tensión de umbral negativa causará sin embargo el **paso de corriente** a 0 V de puerta, como en los JFET, y por lo tanto un comportamiento de tipo vaciado (se prefiere entonces el término "pinch-off" o "cutoff" para la tensión de umbral).

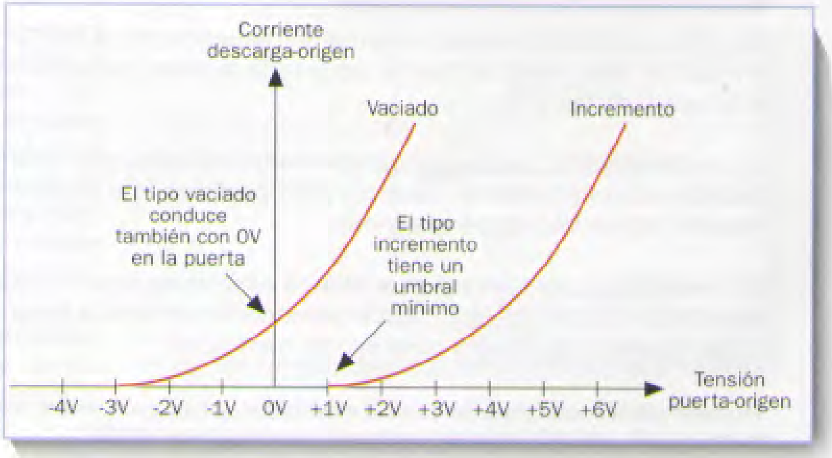

La diferencia entre MOSFET de tipo vaciado y de tipo incremento es la tensión de puerta a la que empiezan a conducir (pinch-off).

MOSFET de potencia

En muchas aplicaciones, estos dispositivos ofrecen interesantes ventajas con respecto a los transistores BJT

Como los transistores bipolares de unión (BJT) normales, también los MOSFET pueden funcionar como **interruptores**.

En particular, el tipo incremento corresponde (como un BJT) a un interruptor **normalmente abierto**: aplicando tensión, se cierra.

En el cierre, el MOSFET se comporta como una **resistencia de bajo valor**, sin ninguna tensión mínima de caída (como la tensión de saturación de los BJT).

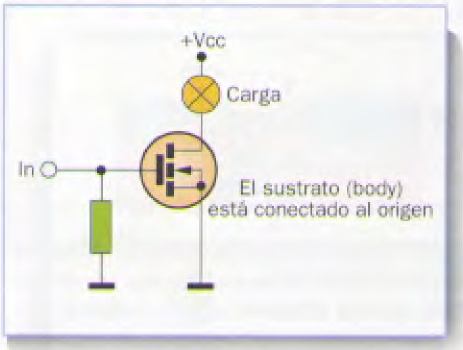

***Un MOSFET puede controlar directamente una carga:* cuando está en conducción plena, ofrece una baja resistencia.**

MENOS PROBLEMAS TÉRMICOS

Si un MOSFET se calienta debido a la corriente que atraviesa el canal, su resistencia **aumenta**; por la ley de Ohm, aumenta por tanto también la tensión en sus extremos.

Ello permite conectar varios MOSFET **en paralelo** sin resistencias de ecualización (ver lección 15): la corriente tenderá a fluir en los que tienen menor resistencia, es decir, los más fríos.

El mismo balance automático de la corriente se produce en el ámbito microscópico, en el interior del canal: los MOSFET son por tanto **inmunes** al desagradable efecto de segunda ruptura que afecta a los BJT.

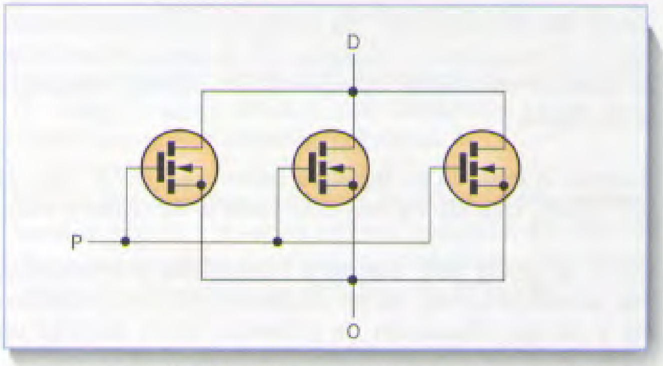

Conectando varios MOSFET* en paralelo, *se reduce la resistencia y se aumenta la corriente soportable.

CAPACIDAD DE PUERTA

Como los MOSFET no requieren ni siquiera **corriente** de puerta (si la tensión de puerta no varía), parecerían comportarse como interruptores perfectos y casi privados de disipación.

Pero como siempre, hay un límite: la puerta forma una **capacidad** relativamente elevada hacia el canal, es decir, tanto hacia el origen como (por desgracia) hacia la descarga.

Cargar esta capacidad requiere **corriente** y **tiempo**: durante este tiempo el canal es parcialmente conductor, en sus extremos hay tensión y en su interior circula corriente: el MOSFET disipa potencia.

SGS
ITALY
P362

Los MOSFET* de potencia *se emplean en numerosas aplicaciones, desde los amplificadores de audio hasta los mandos de los motores eléctricos.

CMOS

El uso de MOSFET complementarios ha permitido la realización de los modernos integrados digitales

La figura muestra dos versiones del clásico interruptor de MOSFET: la primera emplea un MOSFET de tipo incremento **de canal N**, comparable a un transistor NPN (ver lección 10).

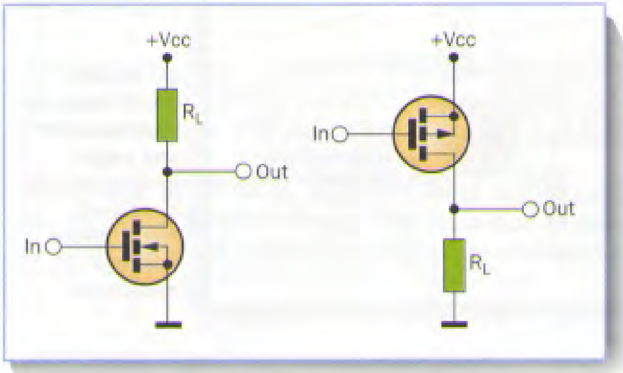

El segundo interruptor es perfectamente **simétrico**: utiliza un MOSFET de canal P y tiene el origen conectado (junto al sustrato) a la alimentación positiva en lugar de a masa.

Ambos circuitos pueden funcionar como **inversor**, pero tienen un inconveniente: la resistencia disipa la potencia cuando el MOSFET conduce.

Dos* inversores de MOSFET *con distinta polaridad: uno es la imagen especular del otro.

INTERRUPTORES COMPLEMENTARIOS

Como los dos MOSFET de la figura entran en conducción con polaridades opuestas de la señal de entrada, se pueden **combinar** obteniendo el inversor mostrado en la figura.

Cuando la entrada es **positiva** pone el MOSFET de canal N bajo, cuando es negativo pone el de canal P alto.

Como la salida está **siempre conectada** a una de las dos alimentaciones, no es necesario ninguna resistencia y no hay disipación de potencia, salvo durante un breve instante durante la conmutación.

Con dos MOSFET* complementarios *se realiza un inversor reducido a lo esencial, pero muy eficaz.

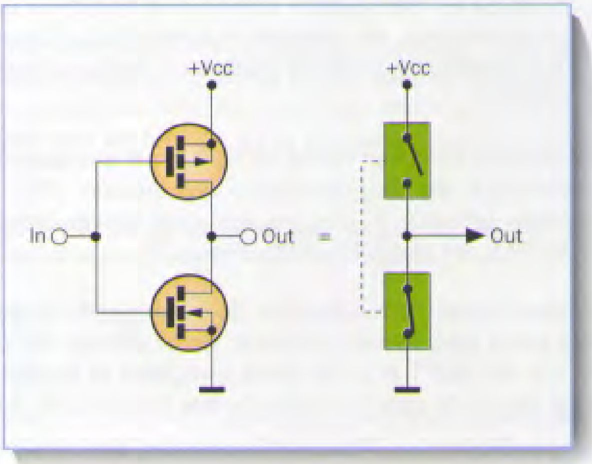

LÓGICA CMOS

Los inversores de **MOSFET complementarios**, o CMOS (Complementary MOS), se utilizan para realizar gran parte de los circuitos integrados digitales, desde sencillas puertas lógicas hasta complejos microprocesadores. Son compactos, rápidos, fácilmente **integrables** en cantidad, no requieren componentes pasivos y tienen una disipación bastante limitada (prácticamente nula, en ausencia de conmutaciones).

Sin embargo necesitan, en los pines de entrada y salida, **protecciones** adecuadas contra las cargas electrostáticas, como por ejemplo las mostradas en la figura.

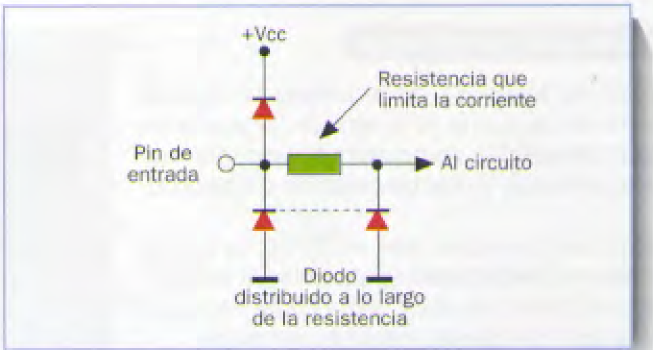
+Vcc
Resistencia que limita la corriente
Pin de entrada
Al circuito
Diodo distribuido a lo largo de la resistencia

Típicas* protecciones antiestáticas *de la serie 74HC; diodos y resistencias se realizan con el mismo silicio.

Interruptores analógicos

Los transistores MOSFET pueden controlar el paso de señales continuamente variables

Cuando un MOSFET está en **conducción** plena, la resistencia del canal es muy baja: se comporta como un interruptor cerrado (ver figura).

Cuando está en **corte** total, el aislamiento entre drenaje y fuente es bastante bueno, casi como si se tratara de un interruptor abierto.

En lugar de utilizar este interruptor para conmutar una carga, como en la lección anterior, se puede utilizar para dejar pasar o bloquear una **señal analógica**.

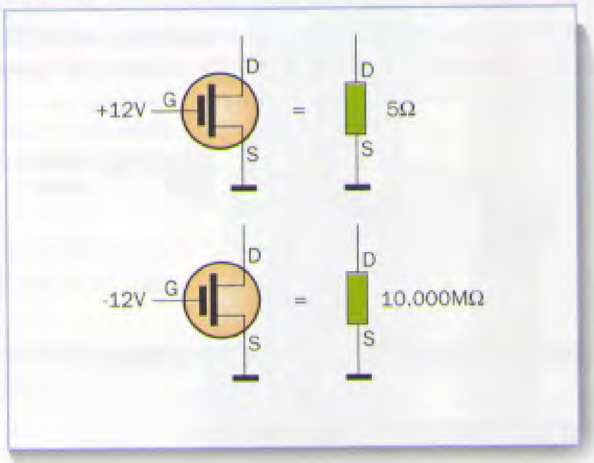

La resistencia del canal de un MOSFET varía drásticamente según la tensión aplicada entre puerta y fuente.

CONTROLAR UNA SEÑAL

La figura muestra una **aplicación** práctica: aplicando +12 V a la puerta, el MOSFET está en conducción y la señal pasa; aplicando -12 V, el canal se bloquea y no pasa prácticamente nada.

La **doble polaridad** de la tensión de control con respecto a masa (+12 V y -12 V) es necesaria para que la señal de entrada pueda ser tanto positiva como negativa.

El **sustrato** (body) está conectado a la tensión negativa, para evitar que la unión canal-sustrato, implícita en la estructura del MOSFET, pueda entrar en conducción.

Aunque no se indica explícitamente en la figura, el MOSFET es de tipo mejora.

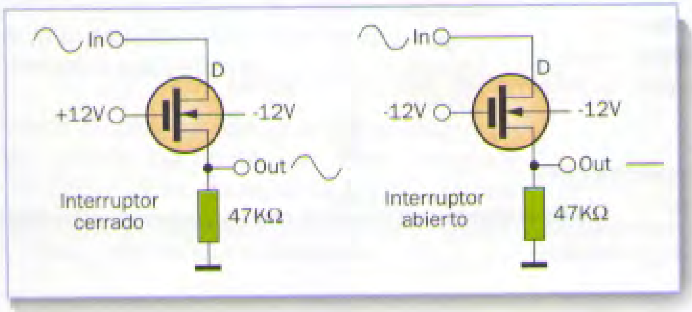

Interruptor analógico de MOSFET: la tensión en la puerta controla el paso de la señal.

LÍMITES DEL INTERRUPTOR DE MOSFET

Si los **picos positivos** de la señal se aproximan a los +12 V durante la conducción, es como si la tensión en la puerta se redujera: la resistencia del canal aumenta.

Pero si los **picos negativos** de la señal quedan por debajo de los -12 V durante el corte, es como si puerta se hiciera positiva: el MOSFET empieza a conducir y su resistencia baja.

Estos problemas se solucionan habitualmente empleando un circuito más sofisticado, con dos MOSFET (o JFET) **complementarios**, como se ha ilustrado en la lección de Componentes.

Conmutar con los MOSFET

Los interruptores analógicos de MOSFET se utilizan mucho, por ejemplo en las etapas de entrada de los ADC

El **muestreo y conservación** descrito en la lección 21 de Digital se puede realizar empleando un MOSFET como interruptor analógico, como muestra la figura.

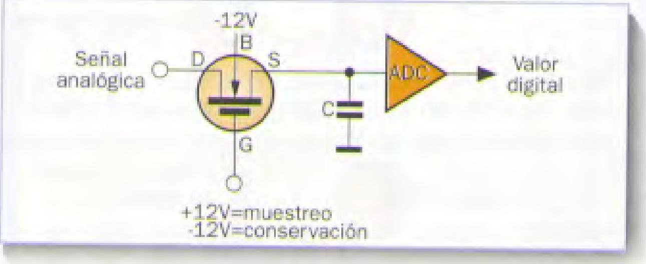

Cuando el MOSFET se comporta como un interruptor cerrado (puerta positiva), la entrada **carga** el condensador hasta llevarlo a su misma tensión.

Abriendo el MOSFET (puerta negativa), el condensador **mantiene** estable la tensión, de modo que el conversor analógico-digital la pueda medir cómodamente.

Un MOSFET como* muestreo y conservación *a la entrada de un conversor analógico-digital (ADC).

MULTIPLEXOR ANALÓGICO

Si el mismo ADC debe medir **por turno** el valor de la tensión en distintas entradas, se pueden utilizar varios interruptores para realizar un multiplexor analógico.

En la figura se muestra un multiplexor de **cuatro entradas**, realizado con otros tantos interruptores MOSFET, cuyas entradas de control (C) son controladas a su vez por un multiplexor digital.

Si la parte digital funciona entre 0 V y +5 V, se pueden utilizar también **drivers** para convertir las señales lógicas a los niveles requeridos por los conmutadores analógicos, por ejemplo -12 V y +12 V.

Multiplexor analógico: ***haciendo conducir un interruptor de MOSFET, se elige la entrada que se va a conectar a la salida.***

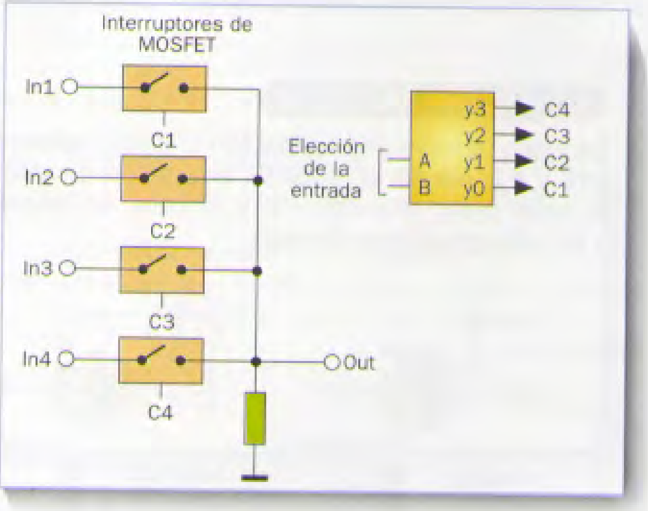

CONTROL DE VOLUMEN

Los interruptores analógicos tienen otras distintas **aplicaciones**; por ejemplo, la figura muestra un control de volumen primitivo para señales de audio.

Se trata en la práctica de un **potenciómetro de saltos**: los interruptores actúan como multiplexor para elegir la atenuación deseada.

Como todos los controles de volumen, también este es **logarítmico**: cada "salto" divide la tensión de la señal, es decir, la reduce en 6 dB (decibelios).

Sencillo* control de volumen *de cuatro posiciones, realizado con un multiplexor analógico.

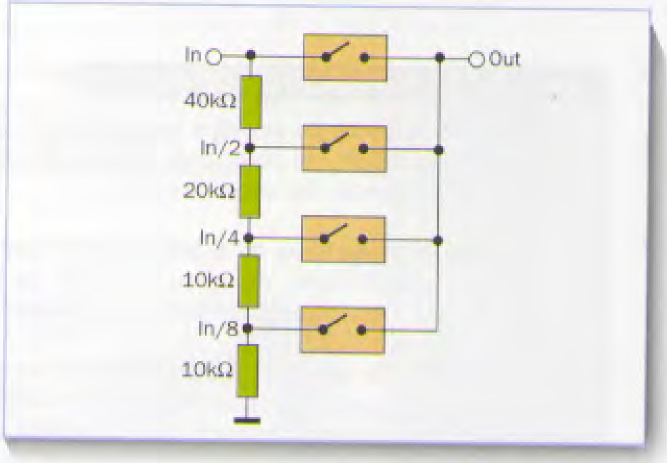

Problemas de los conmutadores

Como todos los dispositivos electrónicos, tampoco el interruptor de MOSFET está exento de inconvenientes

De manera ideal, la resistencia del canal debería ser **cero** con interruptor cerrado, sea cual sea la señal de entrada.

En realidad, variando la tensión de entrada durante la conducción se varía en consecuencia la tensión entre el canal y la puerta, es decir, justo la que **controla** la corriente en el canal.

La resistencia no es por tanto constante, lo que puede causar la **distorsión** de la señal si la salida está conectada a una impedancia relativamente baja, por ejemplo, una resistencia de valor no elevado.

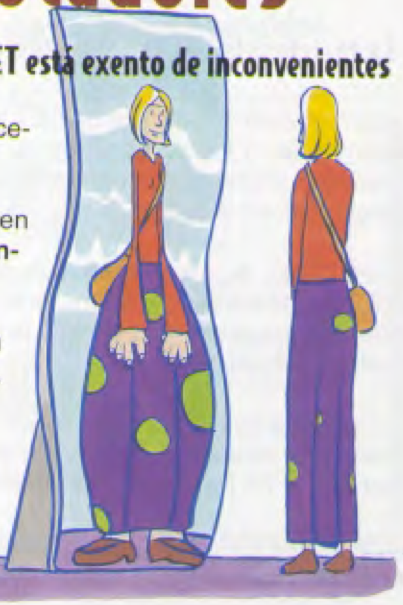

La variación de conducción del canal según el nivel de la señal puede causar distorsión.

PASO NO DESEADO

Si la carga es de **alta impedancia**, como una resistencia de valor elevado, la distorsión en la conducción es superflua... pero se verifica otro inconveniente.

La pequeña **capacidad** entre origen y descarga permite a las señales de frecuencia bastante elevada pasar parcialmente también cuando el interruptor está abierto.

Se puede resolver el dilema con **dos interruptores**, uno de los cuales se cierra mientras que el otro se abre, como en la figura: cuando Q1 está abierto, el segundo interruptor Q2 pone a masa toda señal no deseada.

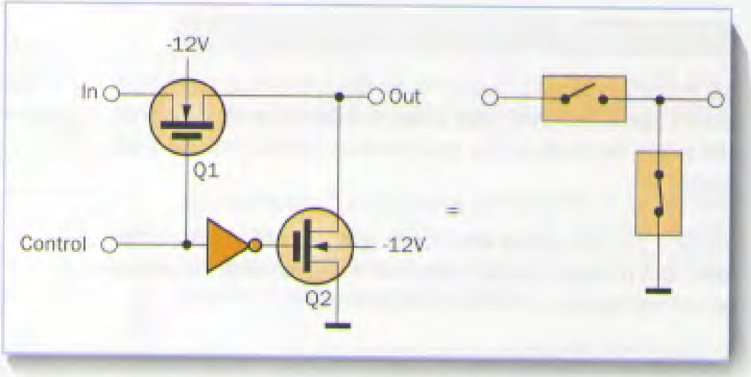

Dos interruptores aseguran alta impedancia en conducción (mínima distorsión) y baja en apertura (paso parcial mínimo).

ACOPLAMIENTO DE PUERTA

La capacidad más molesta es sin embargo la existente entre la puerta y el canal: hace que la señal de control (en la puerta) **interfiera** en lo que pasa en el mismo canal.

En el muestreo y conservación, por ejemplo, la carga adicional que pasa a través de esta capacidad puede **falsear** el valor muestreado, especialmente si se encuentra próximo al cero.

El problema se puede resolver empleando dos MOSFET de **polaridad opuesta** (P y N) controlados por tensiones complementarias: las dos interferencias se anularán recíprocamente, o casi.

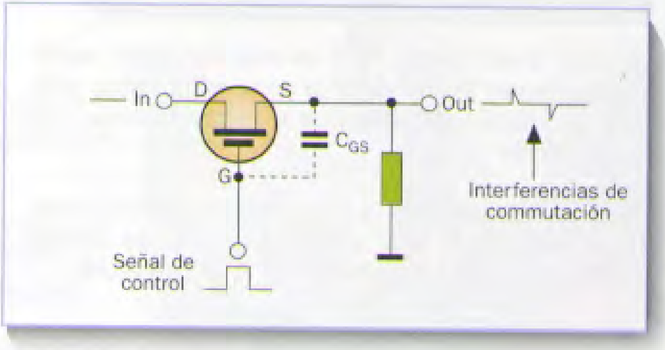

La capacidad entre puerta y canal causa una interferencia a la salida en cada transición de la señal de control.

FET como resistencias variables

La resistencia del canal puede controlarse variando la tensión de puerta, aunque sea con un límite

La corriente en el canal de un FET (de unión o MOS) depende de la tensión de puerta, pero variando la tensión en el drenaje, la corriente alcanza un **límite**, como se ve en la figura

Sin embargo, la **primera parte** de la curva tiene una evolución diferente: para tensiones de drenaje inferiores a algunas decenas de mV, la curva es prácticamente una línea recta.

Eso significa que la corriente en el canal es **proporcional** a la tensión entre drenaje y fuente: en otras palabras, el FET se comporta como una resistencia.

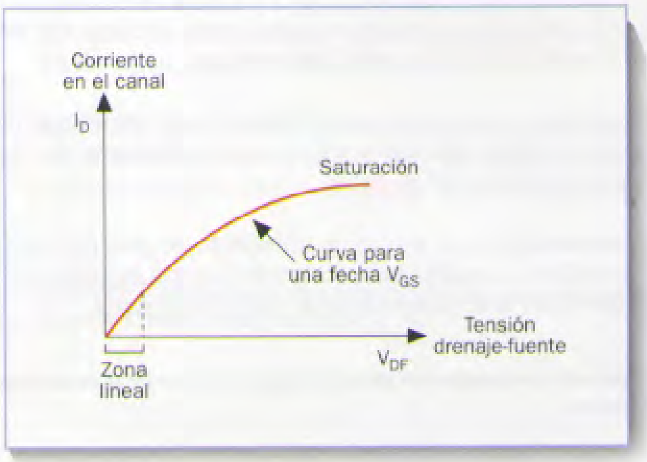

Para tensiones muy bajas entre drenaje y fuente, FET y MOSFET se comportan como resistencias puras.

RESISTENCIA CONTROLADA EN TENSIÓN

La resistencia del canal no es fija, sino que depende de la **tensión** aplicada a la puerta: aumentando la tensión, se reduce la resistencia (para un FET de canal N).

El FET se comporta así como una **resistencia variable**, cuyo valor no se controla mediante un mando, sino mediante la tensión de puerta.

La figura muestra cómo realizar un **potenciómetro** controlado en tensión, empleando un JFET y una tensión negativa de control, que determina el nivel de salida.

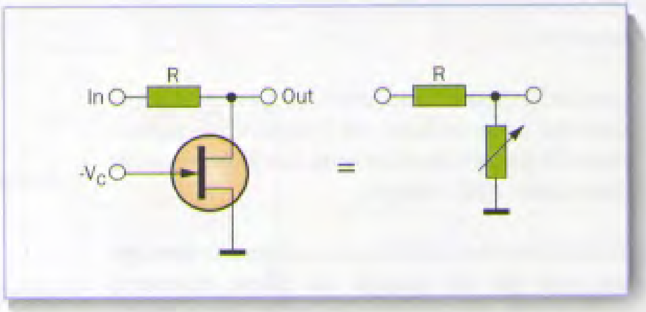

La tensión de control V_C varía la resistencia del canal, y por tanto el nivel de la señal a la salida.

LÍMITES DE LOS FET COMO RESISTENCIAS

La zona lineal de los FET es muy **limitada**: basta con aumentar ligeramente el nivel de la señal para que se distorsione, aunque existen técnicas de "linealización".

Además, la respuesta a la tensión de control **varía** de un ejemplar a otro, según la tensión de umbral (pinchoff) y depende también de la temperatura.

No obstante, con técnicas adecuadas de **compensación**, los FET encuentran aplicación también como resistencias variables, por ejemplo en el control automático de ganancia (AGC) de algunos amplificadores.

Un potenciómetro mecánico es más preciso y previsible que un FET, pero no puede controlarse mediante una tensión.

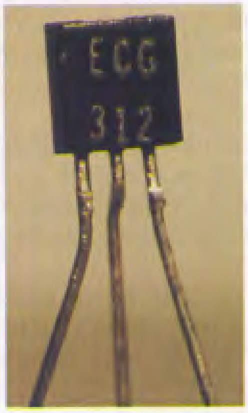

Circuitos resonantes

La combinación inductor-condensador conserva energía de un modo particular

Conectando un condensador cargado en paralelo a un inductor (como en la figura), empieza a **descargarse**, con una corriente "frenada" por el inductor pero progresivamente creciente.

Una vez que el condensador está descargado, la corriente no puede cesar de golpe: el inductor tiende a mantenerla (ver lección 6), haciendo que el condensador se **recargue** con polaridad opuesta.

En este momento el fenómeno se repite al revés: la energía se intercambia hacia delante y hacia atrás, provocando una **oscilación** a una frecuencia ("de resonancia") que depende de los valores de los componentes.

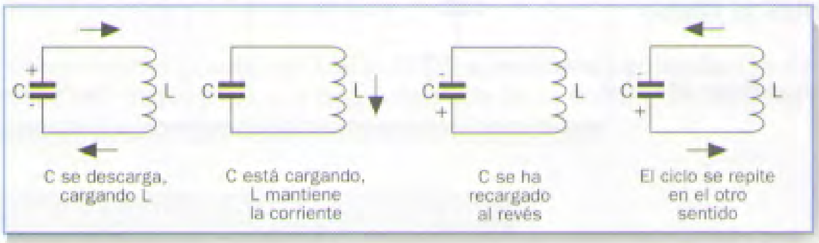
+
C
-
L
C se descarga, cargando L
C
L
C está cargando, L mantiene la corriente
-
C
+
L
C se ha recargado al revés
-
C
+
L
El ciclo se repite en el otro sentido

Circuito resonante: ***la carga del condensador se convierte en campo magnético, y después de nuevo en carga eléctrica.***

COMO UN PÉNDULO

La figura muestra un **equivalente mecánico** del circuito oscilante: un péndulo, donde la altura representa la tensión y la velocidad corresponde a la corriente.

Cuando el péndulo está parado en el punto más alto de su trayectoria tiene **una energía potencial**, equivalente a la carga eléctrica del condensador.

En el punto más bajo tiene **energía cinética** (es decir, de movimiento), correspondiente a la carga magnética del inductor: funciona por "inercia".

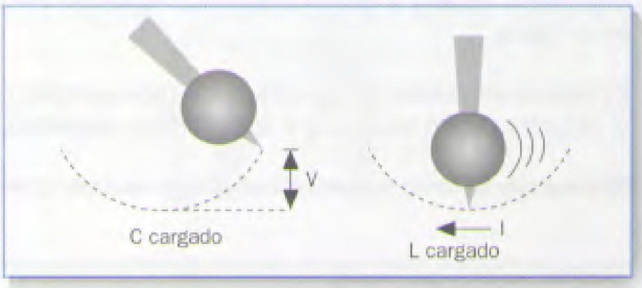

El péndulo oscila a una frecuencia determinada, cambiando continuamente la forma de acumulación de la energía.

NO VA AL INFINITO

Tanto en el caso mecánico (péndulo) como en el eléctrico (circuito LC), la oscilación se **amortigua**, es decir, decrece progresivamente de intensidad a causa de las inevitables pérdidas.

Se deben principalmente a la **resistencia en serie** del inductor, la del cable; en menor medida contribuyen las pérdidas del condensador, similares a una resistencia en paralelo.

Aparte de estas dispersiones, en cualquier caso, en el circuito oscilante no se tiene **disipación** de energía: si los componentes fueran perfectos, la oscilación mantendría por siempre la misma amplitud.

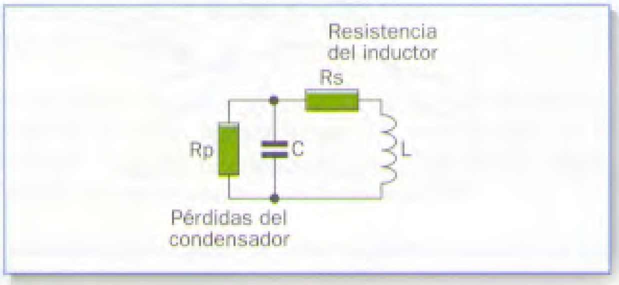

Las pérdidas del circuito oscilante se deben a disipaciones en las resistencias parásitas, inevitables con los componentes reales.

Serie y paralelo

Los circuitos resonantes pueden tener un comportamiento más o menos "brusco" a la frecuencia de resonancia

A la frecuencia de resonancia, un circuito LC tiene (teóricamente) **impedancia infinita**, es decir, se comporta como un circuito abierto: ya que mantiene la energía, no la absorbe.

La figura muestra un **filtro de paso de banda**: a frecuencias bajas el inductor tiene baja impedancia y la señal es atenuada; a frecuencias altas lo mismo ocurre con el condensador.

A la **frecuencia de resonancia** la impedancia sube bruscamente, como se ve en el gráfico de la respuesta en frecuencia, y la señal pasa sin atenuaciones.

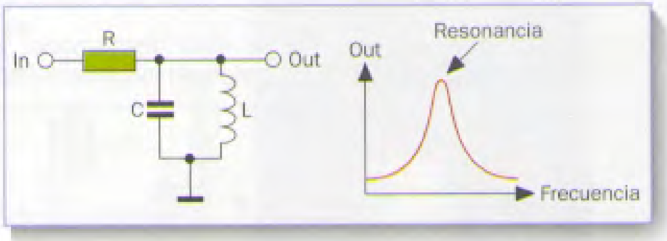

Filtro de paso de banda realizado con un circuito oscilante LC en paralelo.

CIRCUITO LC EN SERIE

Poniendo el inductor **en serie** con el condensador en lugar de en paralelo, se obtiene otro tipo de circuito oscilante: la impedancia es normalmente alta, pero se desmorona bruscamente a la frecuencia de resonancia.

Por tanto es posible realizar un filtro **supresor de banda** (ver lección 17), que deja pasar todas las frecuencias, salvo aquella a que el circuito LC resuena, como en la figura.

Es interesante observar que la tensión (por ejemplo) en el inductor puede ascender a valores muy **superiores** al de la tensión a la entrada; depende del "factor de calidad" del circuito.

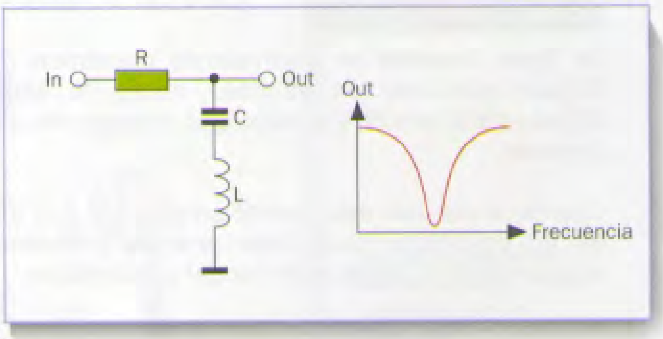

Filtro supresor de banda realizado con un circuito oscilante LC en serie.

FACTOR DE CALIDAD

Un circuito oscilante perfecto resuena **limpiamente** a una determinada frecuencia, pero las pérdidas resistivas hacen que el comportamiento de un circuito real sea más impreciso.

El factor de calidad **Q** determina lo brusca que es la resonancia, como se ve en la figura; a veces se añade una resistencia especial para "suavizar" la respuesta.

También las impedancias del **origen** y de la **carga** son importantes: la resistencia R en los datos altos contribuye a determinar el comportamiento del filtro.

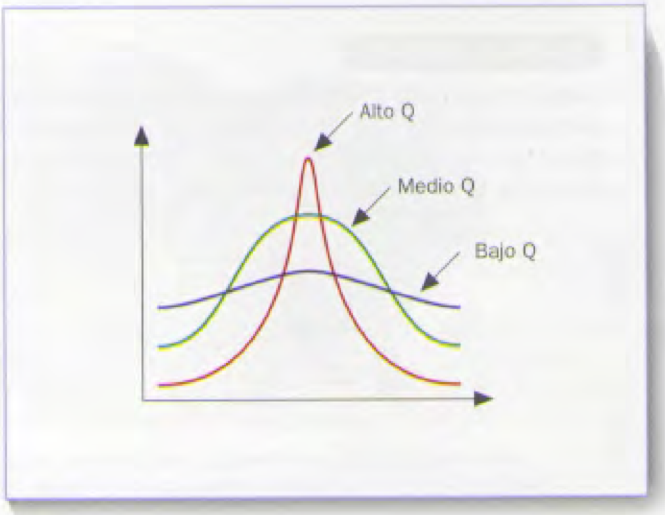

El factor de calidad Q se define como la longitud del pico (a -3 dB) dividida por la frecuencia de resonancia.

Osciladores

Un circuito resonante puede utilizarse para producir una tensión alterna a la frecuencia deseada

Para producir señales de **alta frecuencia**, como la "portadora" utilizada en las transmisiones de radio o televisión, no se utilizan normalmente los osciladores descritos en la lección 20.

Se emplean en su lugar circuitos resonantes **ajustados** a la frecuencia deseada, utilizando posiblemente un condensador o un inductor variable para regularla.

Un amplificador puede ser utilizado para **mantener** la oscilación del circuito LC, extrayendo una pequeña parte de la señal y emitiéndola de nuevo con mayor energía.

REALIMENTACIÓN DESDE EL INDUCTOR

En el circuito de la figura, el circuito ajustado constituye la propia carga del transistor amplificado; una **toma intermedia** extrae una parte de la señal para reenviarla a la entrada.

Para adaptar mejor las impedancias, la realimentación entra por el emisor y la base está a masa (para la señal) mediante C1: se trata de un amplificador **de base común**.

La toma intermedia en el inductor lo convierte en un autotransformador, desarrollando la misma tarea que un devanado independiente (ver lección 7): el circuito es conocido como oscilador **Hartley**.

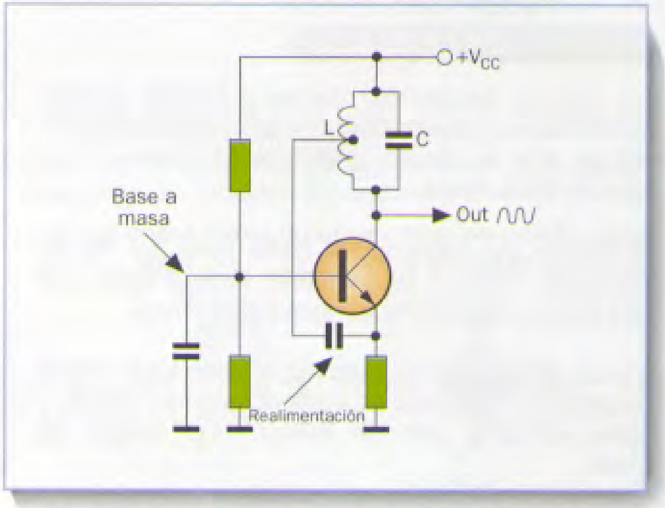

Oscilador LC Hartley con* autotransformador *para reenviar a la entrada parte de la señal de salida.

REALIMENTACIÓN DESDE EL CONDENSADOR

En lugar de una toma intermedia en el inductor, se puede utilizar... ¡una toma intermedia en el condensador! Se obtiene utilizando **dos condensadores** en serie, como en la figura.

Este **oscilador Colpitts** tiene la ventaja de requerir un inductor sencillo y económico sin tomas intermedias u otros devanados.

Vale la pena observar que capacidades e inductancias parásitas pueden **transformar** un amplificador en un oscilador Colpitts no deseado: hay que tener mucho cuidado en el proyecto y en la realización.

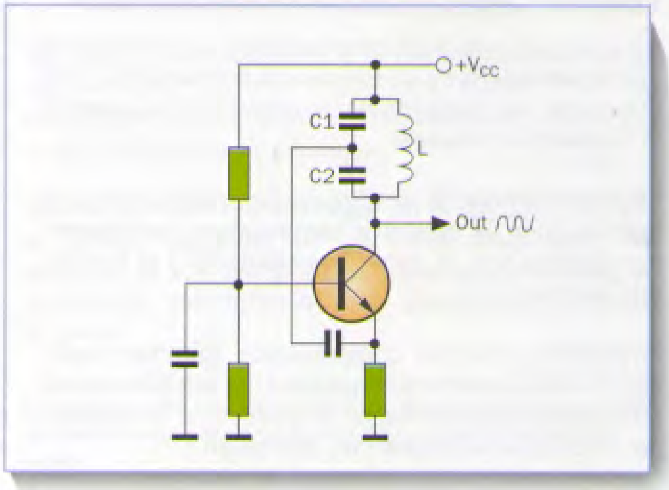

El oscilador Colpitts utiliza un* divisor capacitivo *para extraer una parte de la señal de salida.

Osciladores de cuarzo

Un cristal de cuarzo adecuadamente trabajado puede hacer las veces de resonador electromecánico

El cuarzo es un material **piezoeléctrico**: si se aplasta, produce una tensión; pero si, a la inversa, se le aplica una tensión, se deforma ligeramente.

Una fina lámina de este material, situada entre dos electrodos metálicos, tiene una **frecuencia de resonancia** precisa y estable, y un Q elevadísimo, como un circuito LC casi perfecto.

Por tanto es el dispositivo ideal para **controlar** la frecuencia de un oscilador; sin embargo es imposible variar la frecuencia producida, que depende sólo de las características mecánicas.

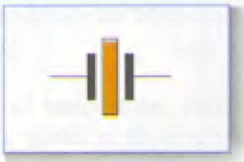

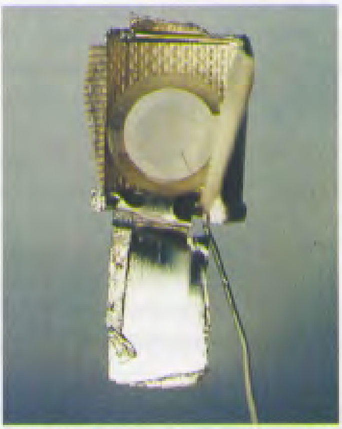

La estructura interna de un resonador de cristal de cuarzo, y su símbolo.

OSCILADORES DE CUARZO

Los clásicos osciladores Hartley y Colpitts pueden **modificarse** para funcionar con un cuarzo (cristal) en vez de con el circuito resonante LC, como en el ejemplo de la figura.

El inductor del esquema no forma parte del circuito resonante, sino que es un **"filtro"** para la señal, que deja pasar solamente la componente continua.

A pesar de la aparente sencillez, la puesta a punto de un oscilador de cuarzo es siempre bastante **crítica**, y debe realizarse para un determinado modelo de cristal.

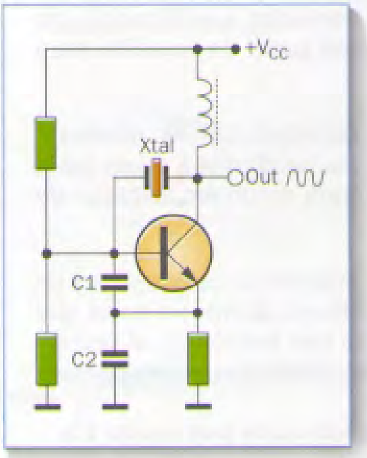

Oscilador Colpitts de cuarzo, es decir, controlado por un resonador piezoeléctrico de cristal de cuarzo.

ARMÓNICAS, TEMPERATURA Y REGULACIÓN

Un cuarzo puede funcionar también en el tercero o quinto **armónico** de su frecuencia fundamental, normalmente con la ayuda de un circuito LC ajustado a la frecuencia deseada.

La dependencia de la temperatura puede limitarse manteniendo el **cristal a temperatura constante**, por ejemplo con un pequeño calentador y el termostato asociado.

Finalmente, con un compensador (trimmer capacitivo) adecuadamente dispuesto, es posible aportar minúsculas variaciones en la frecuencia de trabajo, de modo que se **ajuste** con precisión.

El cuarzo es un mineral bastante común en la naturaleza, cuyas características electromecánicas nos resultan útiles.

Transmisores

Revisemos las principales técnicas utilizadas para transmitir información con señales de radio

Cada transmisión de radio tiene una **frecuencia** propia: hay, por tanto, una señal producto de un oscilador de alta frecuencia, que es amplificado y enviado a la antena transmisora.

Esta señal, que es una onda sinusoidal pura, se llama **portadora**, ya que tiene la misión de transportar la información que se desea transmitir.

La **información** puede consistir en una señal de audio (es el caso de la radio), una señal de vídeo (TV), datos en forma digital u otra cosa.

CÓDIGO MORSE

De por sí, la portadora no contiene información alguna: es necesario hacerla variar, es decir, **modularla**, ya que de otro modo es una simple oscilación fija de alta frecuencia.

El sistema más **sencillo** de modulación consiste en transmitir o no la propia portadora (on/off) a voluntad: a la salida se tendrá una señal sinusoidal, o bien nula.

La información está contenida en las **secuencias** de encendido y apagado, como en el código ideado por Samuel Morse y utilizado en todo el mundo durante muchos años (oficialmente hasta 1997).

a	•—	m	——	y	—•——
b	—•••	n	—•	z	——••
c	—•—•	o	———	1	•————
d	—••	p	•——•	2	••———
e	•	q	——•—	3	•••——
f	••—•	r	•—•	4	••••—
g	——•	s	•••	5	•••••
h	••••	t	—	6	—••••
i	••	u	••—	7	——•••
j	•———	v	•••—	8	———••
k	—•—	w	•——	9	————•
l	•—••	x	—••—	0	—————

Código Morse: *el "punto" es un encendido breve de la portadora y la "línea" es un encendido más largo.*

MODULACIÓN

La figura muestra el **esquema de principio** de un transmisor Morse: está formado por un oscilador, un interruptor y un amplificador.

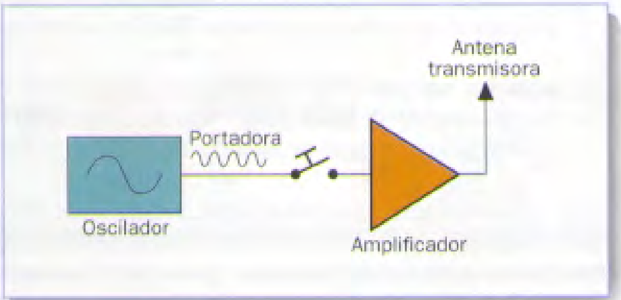

La **modulación** puede obtenerse interrumpiendo la señal entre oscilador y transmisor (como en el esquema), o controlando directamente la alimentación de este último.

Pero en general no se aconseja encender y apagar el **oscilador**, ya que esto suele requerir un breve período de arranque y estabilización.

Esquema de bloques del* transmisor *Morse: la presión de la tecla deja pasar la portadora del oscilador al amplificador.

Circuitos de radio

Los transmisores utilizan circuitos ajustados a la frecuencia deseada

Los amplificadores vistos hasta ahora no son adecuados para frecuencias altas, ya que tienen una frecuencia de corte bastante baja (algunas decenas o centenas de KHz).

Además, no conviene amplificar **todas** las frecuencias, incluyendo las no deseadas (por ejemplo, los armónicos debidas a la distorsión).

Se utilizan, por tanto, **amplificadores ajustados**, que emplean uno o varios circuitos resonantes, de modo que tienen una elevada ganancia sólo a la frecuencia deseada.

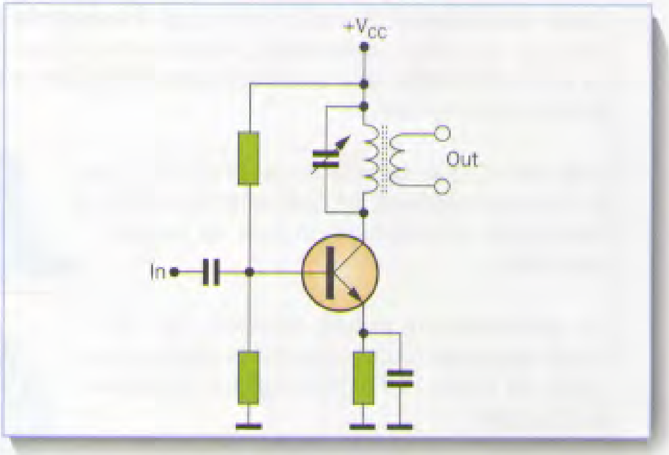

***Amplificador ajustado:** a la frecuencia de resonancia del circuito LC (primario del transformador) la ganancia es máxima.*

AMPLIFICADORES DE CLASE C

Especialmente en la etapa final, es decir, el último amplificador que controla la antena, están en juego **potencias elevadas**: también trabajando en clase B (ver lección 18), la disipación sería notable.

El circuito resonante de salida puede, sin embargo, **mantener** la oscilación recibiendo, además, un solo impulso de corriente por ciclo, como un columpio en el que basta un empujoncito en cada balanceo.

En los amplificadores en clase C, el transistor conduce sólo durante **una parte** del ciclo, por ejemplo aquella en la que la señal supera los 0,65 V necesarios para enviarla a conducción (figura).

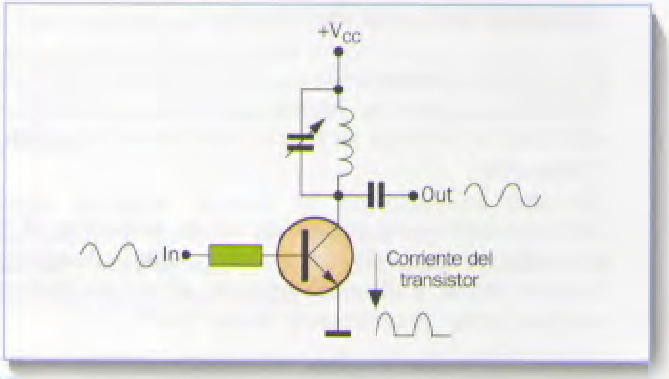

***Principio del amplificador** en clase C: en cada pico positivo a la entrada, el transistor "da un empujón" al circuito resonante.*

MULTIPLICADORES DE FRECUENCIA

Empujando un columpio cada **dos** oscilaciones en lugar de cada una mantiene de todas formas su propio movimiento regular, a una frecuencia doble de la de los empujones.

Del mismo modo, controlando un amplificador en clase C con una frecuencia similar a la mitad de la del circuito resonante, se **duplica** la frecuencia aplicada a la entrada.

En la figura aparece una técnica más **eficaz**: un transistor conduce en los picos positivos y el otro en los negativos, así el circuito resonante recibe dos impulsos por ciclo.

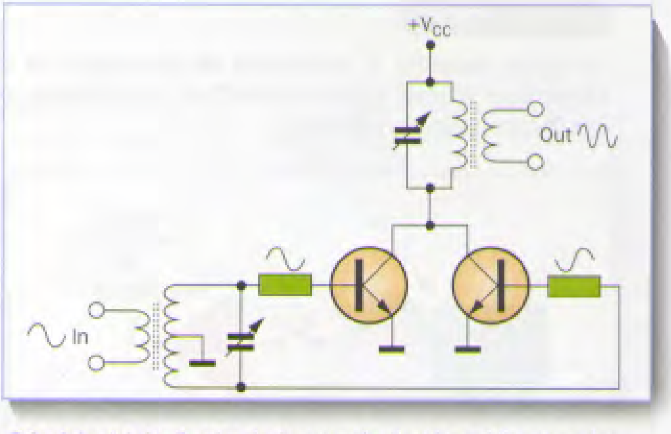

***Principio del duplicador** de frecuencia "push-push": por cada ciclo de la señal, el circuito resonante recibe dos "empujones".*

Modulación

Las dos técnicas principales de modulación consisten en variar la amplitud o la frecuencia de la portadora

Para transmitir una **señal analógica** en lugar de una simple frecuencia de impulsos es necesario que la portadora transporte una información continuamente variable.

La **modulación de amplitud** (AM) consiste en variar la amplitud de la portadora, de modo que siga fielmente el nivel de la señal moduladora, como en la figura.

Esta se puede obtener, por ejemplo, haciendo que la señal moduladora (por ej. audio) **controle** la tensión de alimentación de un amplificador y, por tanto, su potencia de salida.

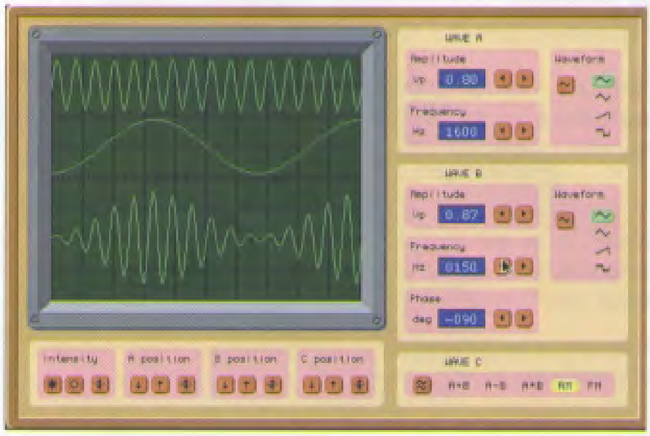
WAVE A
Amplitude Vp 0.80
Frequency Hz 1600
WAVE B
Amplitude Vp 0.87
Frequency Hz 0150
Phase deg -090
WAVE C: AM

Modulación de amplitud: ***el perfil de los picos de la portadora dibuja la señal a transmitir.***

MODULACIÓN DE FRECUENCIA

La modulación de amplitud tiene un inconveniente: es especialmente vulnerable a las **interferencias**; de hecho, cada variación de intensidad de la portadora se interpreta como señal. Resultados mejores se obtienen manteniendo constante la amplitud de la portadora, y variando, sin embargo, la **frecuencia** (FM), como muestra la figura.

Gracias a su mejor inmunidad a las interferencias, la modulación de frecuencia (FM) ha encontrado muchas aplicaciones en las **transmisiones radiofónicas** y para el sonido de los televisores.

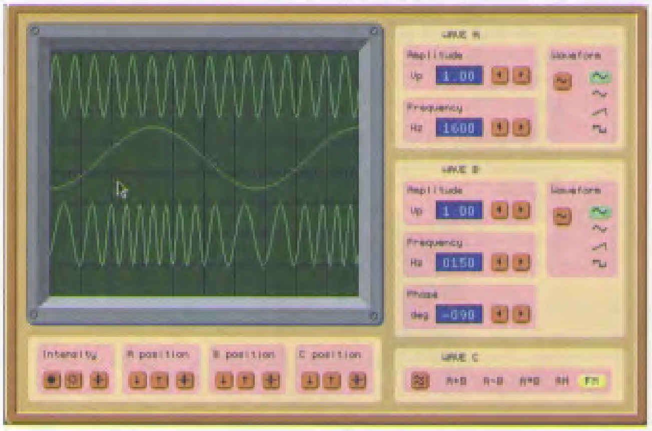
WAVE A
Amplitude Vp 1.00
Frequency Hz 1600
WAVE B
Amplitude Vp 1.00
Frequency Hz 0150
Phase deg -090
WAVE C: FM

Modulación de frecuencia: ***cuando la tensión de la señal sube, aumenta la frecuencia de la portadora, y viceversa.***

VCO Y VARACTOR

Hay diversas técnicas para obtener la modulación de frecuencia; la más sencilla consiste en variar directamente la frecuencia del **oscilador** que produce la portadora.

Para realizar un oscilador **controlado en tensión**, o VCO (voltage controlled oscillator), se puede utilizar un "varactor", un diodo especial de capacidad variable.

Cuando un varactor está polarizado inversamente, las cargas en ambos lados de la unión forman un **condensador variable**, cuya capacidad depende de la tensión aplicada.

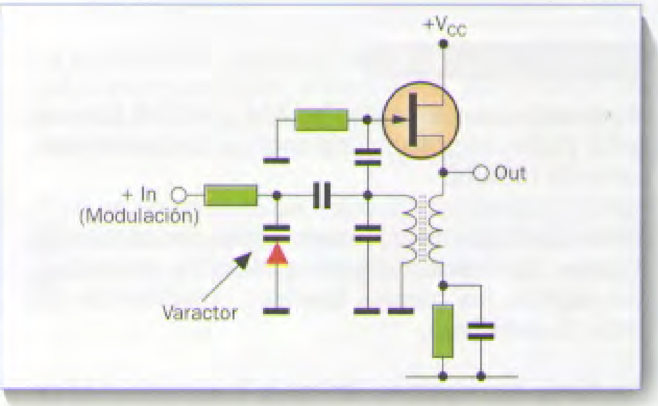

VCO: la tensión a la entrada hace variar la capacidad del varactor y, por tanto, la frecuencia de oscilación del Hartley de JFET.

Bandas laterales

Una transmisión de radio modulada no está nunca compuesta de una sola frecuencia

Se podría pensar que un transmisor de **modulación de amplitud** que funciona a 10 MHz produce una sola frecuencia, por ejemplo de diez megahertzios.

Esto es cierto mientras está presente **sólo la portadora**; nada más se modula con una señal (por ejemplo una sinusoide de 1 KHz) las cosas cambian.

En realidad, como la modulación es una operación **no lineal**, a la salida aparecen otras dos frecuencias, similares a la suma y a la diferencia de portadora y señal (ver figura).

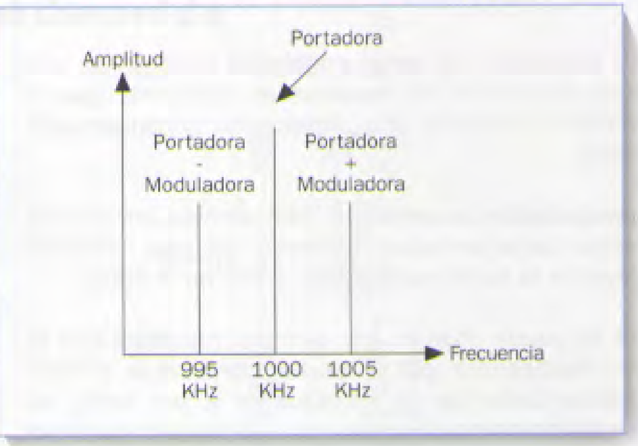

Espectro de las frecuencias contenidas en una señal de 1.000 KHz modulada en amplitud a 5 KHz, ambas sinusoidales.

ANCHO DE BANDA

Como la señal de entrada (moduladora) puede contener una cierta **gama de frecuencias**, variarán también los componentes de la señal de salida del transmisor AM.

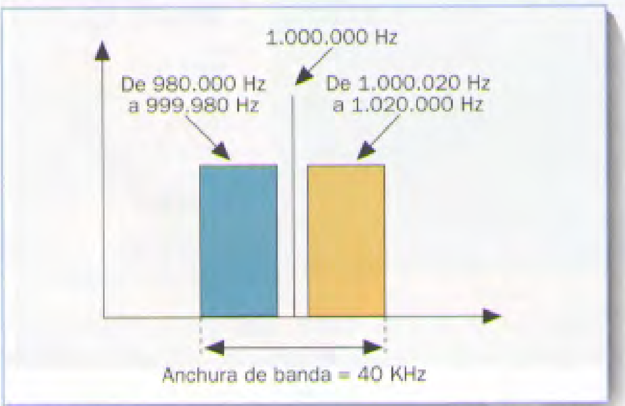

Si por ejemplo se trata de una **señal de audio** con frecuencias comprendidas entre 20 Hz y 20 KHz, en la señal transmitida estarán presentes frecuencias comprendidas en las dos bandas representadas en la figura.

Estas bandas toman el nombre de **bandas laterales**, y definen en conjunto la "anchura de banda", es decir, el espectro de frecuencias ocupado por el transmisor.

Bandas laterales y amplitud de banda ***para una portadora de 1 MHz modulada por señales comprendidas entre 20 Hz y 20 KHz.***

SEÑALES DIGITALES

Modulando una portadora (en AM o en FM) con una señal digital, se obtiene una anchura de banda teóricamente **infinita**.

En realidad, las **transiciones netas** de las señales digitales contienen un gran número de armónicos, que alargan las bandas laterales, interfiriendo con otros transmisores.

Por tanto, es necesario **limitar** la frecuencia máxima contenida en la señal, es decir, "redondear las esquinas", antes de utilizarla para modular un transmisor (en la práctica se utilizan técnicas bastante complejas).

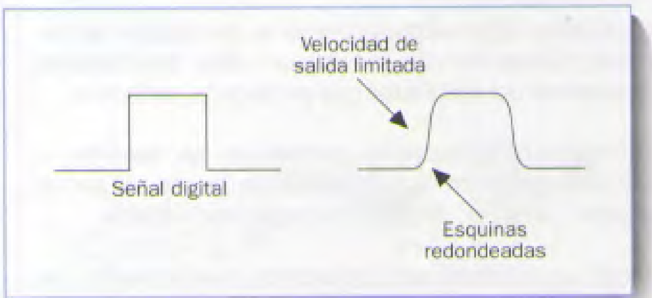

Antes de transmitir una señal digital, es necesario "suavizarlo" para limitar la banda ocupada.

Receptores

El dispositivo receptor debe ser capaz de seleccionar la frecuencia deseada, ignorando las demás

La **señal de radio** captada por la antena receptora no es la emitida por una sola emisora, sino la suma de numerosas señales procedentes de distintos transmisores.

El circuito de **sintonización** tiene la misión de seleccionar una sola frecuencia, de modo que el receptor sea sintonizado en la emisora que se desee recibir.

En otras palabras, las primeras etapas del receptor deben comportarse como un **filtro de paso de banda**, que deje pasar una determinada frecuencia, atenuando lo más posible las otras.

RESONANCIA VARIABLE

El modo más sencillo de sintonizar una frecuencia consiste en utilizar un **circuito resonante**, ajustado a la frecuencia deseada.

Regulando la frecuencia de resonancia, normalmente con un **condensador variable**, es posible obtener el efecto deseado, es decir, resaltar la señal deseada y atenuar las demás.

Si después las etapas amplificadores posteriores son **ajustadas** también a la misma frecuencia (ver lección 25), la selección resulta aún más limpia.

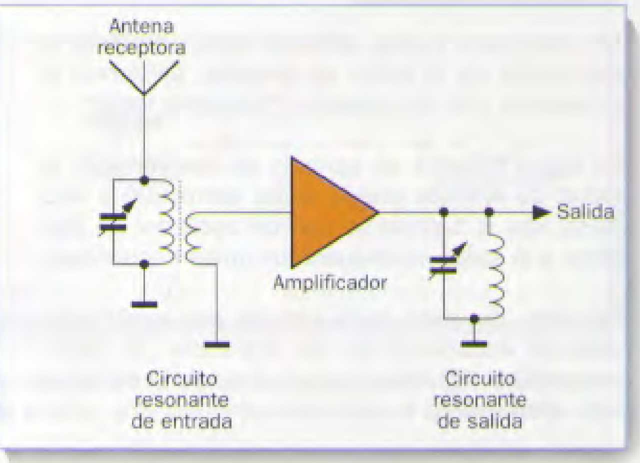

Los condensadores variables regulan la resonancia, de modo que se capte y amplifique sólo una estrecha banda de frecuencias.

Q E IMPEDANCIA

La **selectividad** de un circuito resonante, es decir, su capacidad de resonar en un campo muy estrecho de frecuencias, depende de su factor de calidad (Q, ver lección 24).

Cada **carga** a baja impedancia, por ejemplo, la entrada de un amplificador, se comporta (grosso modo) como una resistencia, "amortiguando" el circuito y reduciendo Q. Para limitar el inconveniente se utiliza una **toma intermedia** (o un segundo devanado, como en la figura de arriba) que interfiera lo menos posible en el circuito resonante, dejándolo libre para oscilar.

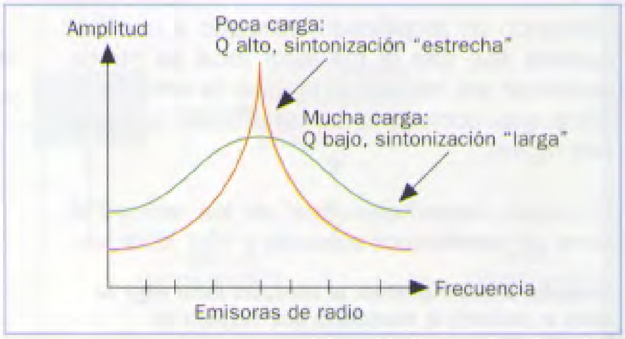

Reduciendo la carga en el circuito resonante aumenta la selectividad.

Conversión de frecuencia

Una característica de los circuitos no lineales permite mejorar notablemente las prestaciones de los receptores

Realizar varios circuitos resonantes, regulables con el mismo comando, no es práctico: especialmente a las altas frecuencias, el circuito resulta **difícil** de poner a punto.

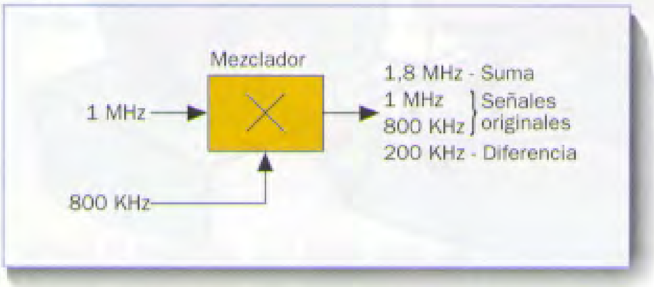
Mezclador
1 MHz
800 KHz
1,8 MHz - Suma
1 MHz
800 KHz
Señales originales
200 KHz - Diferencia

Los **mezcladores** ofrecen una oportunidad para solucionar el problema. Son circuitos no lineales, que contienen por ejemplo diodos, en los que se mezclan dos señales de distinta frecuencia.

En la salida no estarán sólo las dos señales originales, sino además las frecuencias de "batimiento", similares a la **suma** y a la **diferencia** de las principales (ver figura).

Un mezclador produce una intermodulación (o "batimiento"), extrayendo suma y diferencia de las frecuencias a la entrada.

OSCILADOR LOCAL

Un mezclador puede utilizarse para **convertir** la frecuencia de la señal de entrada, utilizando el batimiento con un oscilador ("oscilador local").

La figura muestra un ejemplo de **conversión**: la señal de entrada puede variar entre 900 y 950 MHz; tras el batimiento con un oscilador de 800 MHz, a la salida se encuentran varias frecuencias.

Un filtro de paso bajo elimina las superiores, dejando solamente las de diferencia, es decir, entre 100 y 150 MHz: la frecuencia de la señal ha sido **disminuida** en 800 MHz.

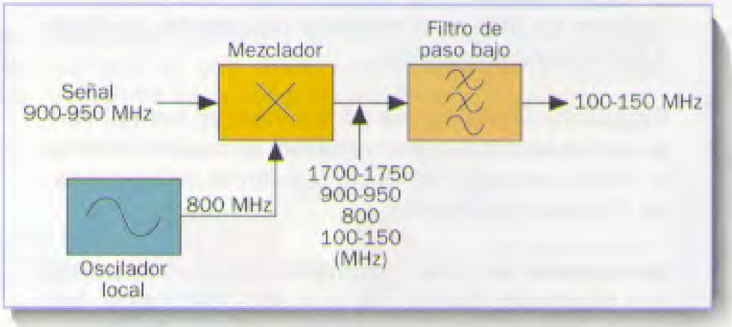
Señal
900-950 MHz
Mezclador
Filtro de
paso bajo
100-150 MHz
800 MHz
Oscilador
local
1700-1750
900-950
800
100-150
(MHz)

***Conversión de frecuencia** con un oscilador local, un mezclador que hace la suma y la diferencia, y un filtro de paso bajo.*

SUPERHETERODINO

Haciendo **variar** la frecuencia del oscilador local, se puede llevar la señal de entrada a la frecuencia deseada, lo que permite utilizar un interesante truquito.

Utilizando un amplificador ajustado a una frecuencia **fija**, con el oscilador local se puede sintonizar una frecuencia precisa de entrada: la única que, convertida, se amplificará después (ver figura).

El circuito "superheterodino" de las radios FM tiene un amplificador ajustado a 10,7 MHz, denominado de **"frecuencia intermedia"**, relativamente fácil de realizar.

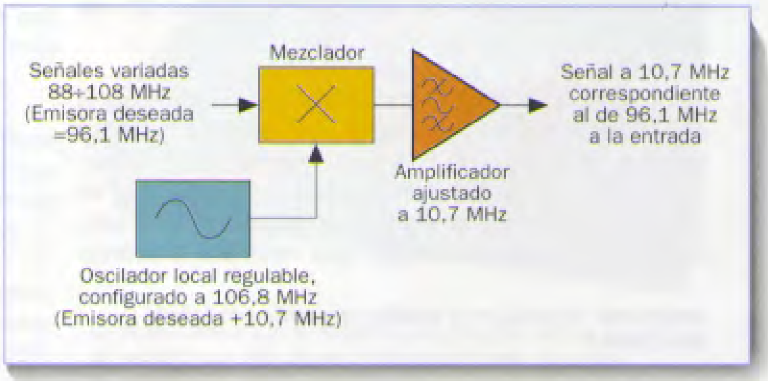
Señales variadas
88÷108 MHz
(Emisora deseada
=96,1 MHz)
Mezclador
Amplificador
ajustado
a 10,7 MHz
Señal a 10,7 MHz
correspondiente
al de 96,1 MHz
a la entrada
Oscilador local regulable,
configurado a 106,8 MHz
(Emisora deseada +10,7 MHz)

***Receptor superheterodino:** el oscilador local elige la señal a convertir la frecuencia que después se amplificará.*

Desmoduladores

Una vez seleccionada la frecuencia deseada, es necesario revelar la información contenida en ella

Si la señal de radio recibida está en **código Morse**, es decir, consiste en sencillos encendidos y apagados del transmisor, hay distintos modos para desmodularla, es decir, transformarla en una señal de audio.

Un sistema típico se muestra en la figura: se trata de hacer **batimiento** con una frecuencia muy próxima a la recibida (que suponemos ya convertida a una frecuencia fija).

Por ejemplo, si la señal es de 455 KHz, mezclándola con 454 KHz se obtendrá como diferencia una frecuencia de 1 KHz, es decir, un **silbido** perfectamente audible.

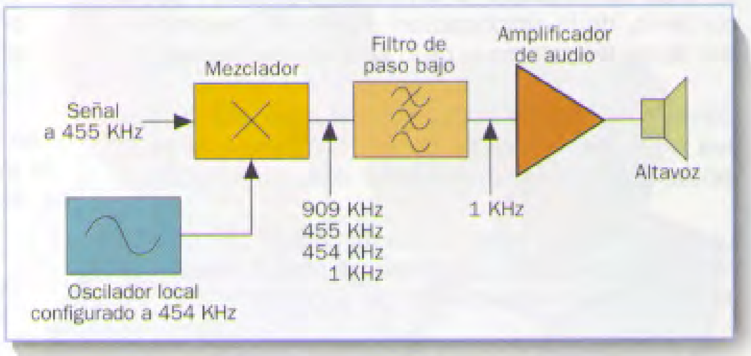
Señal a 455 KHz
Mezclador
Filtro de paso bajo
Amplificador de audio
Altavoz
909 KHz
455 KHz
454 KHz
1 KHz
1 KHz
Oscilador local configurado a 454 KHz

Desmodulador *para transmisiones Morse: el batimiento con el oscilador local produce una señal de frecuencia audible.*

DESMODULADORES AM

Para revelar la señal de audio contenida en una señal de radio modulada en **amplitud** (ver lección 35), se puede utilizar el sencillo circuito mostrado en la figura.

No es otra cosa que un **rectificador de semionda**, similar a los utilizados en las fuentes de alimentación: el condensador se carga con los picos y la salida sigue el perfil de la señal moduladora.

En los desmoduladores AM se utilizan normalmente **diodos de germanio** en lugar del silicio; tienen una menor caída de tensión, aunque las otras características (por ejemplo, corriente de pérdida) son peores.

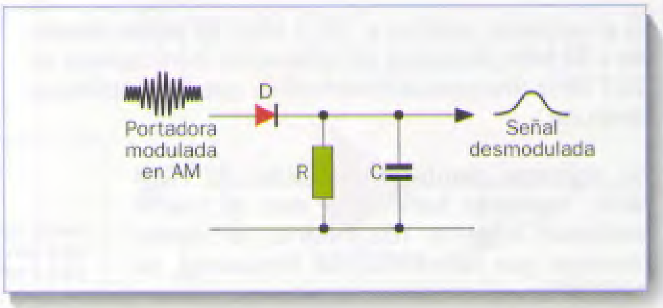
D
Portadora modulada en AM
R
C
Señal desmodulada

Desmodulador AM: *la tensión en el condensador sigue la amplitud de la señal de radio, que representa la señal de audio deseada.*

DESMODULADOR FM

Para revelar la modulación de frecuencia se utilizan varios sistemas, por ejemplo el basado en un **PLL** (Phase-Locked Loop: bucle bloqueado en fase).

Se trata de un circuito que regula un oscilador controlado en tensión (VCO), de modo que **siga** fielmente la frecuencia de la señal de entrada.

La tensión de control del VCO es, por tanto, proporcional a la **frecuencia** de la propia señal de radio: en otras palabras, es justo la señal de audio que se quería revelar.

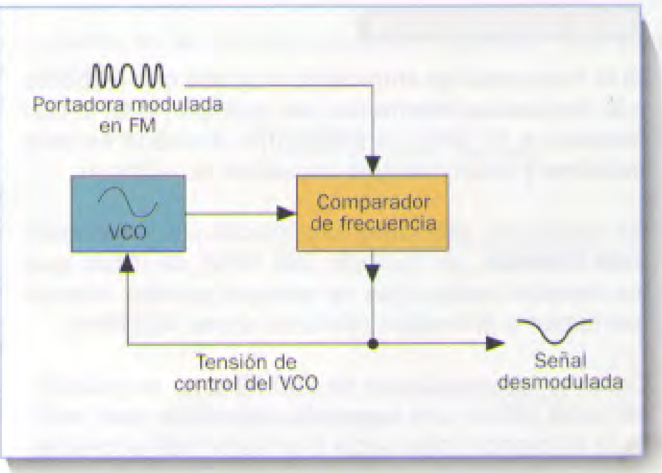
Portadora modulada en FM
VCO
Comparador de frecuencia
Tensión de control del VCO
Señal desmodulada

Desmodulador FM *con PLL: el VCO se queda "enganchado" a la señal de entrada, es decir, la sigue fielmente.*

Circuitos de radio

Veamos algunas soluciones circuitales empleadas en los receptores de radio para mejorar sus prestaciones

En lugar de mecánico, el control del oscilador local (y, por tanto, de la sintonización) puede ser electrónico; una de las técnicas es la de la **síntesis de frecuencia**.

Consiste en hacer que la frecuencia del oscilador local sea un **múltiplo exacto** de una frecuencia de referencia, obtenida con un oscilador de cuarzo preciso.

La figura muestra como se puede realizar con un PLL (ver página anterior) y un **divisor digital**: la relación de división controla la frecuencia del oscilador.

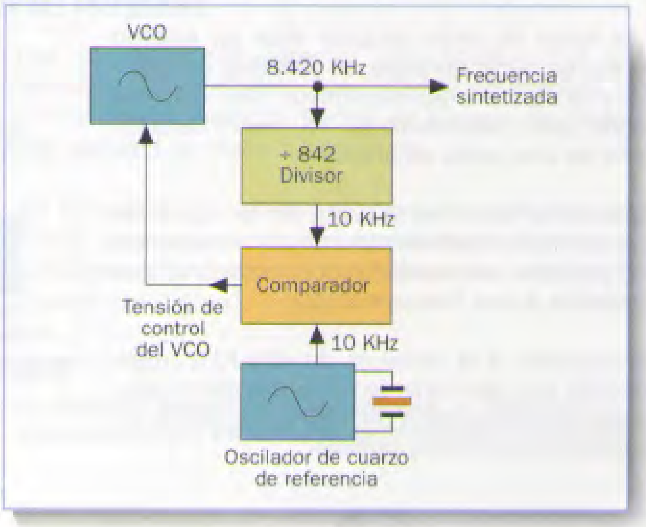
VCO
8.420 KHz
Frecuencia sintetizada
÷ 842
Divisor
10 KHz
Comparador
Tensión de control del VCO
10 KHz
Oscilador de cuarzo de referencia

Síntesis de frecuencia: *si el divisor divide por 842, el oscilador funcionará a 842 veces la frecuencia de referencia.*

FRECUENCIA IMAGEN

Si el oscilador local va a 102,7 MHz, la **señal deseada** a 92 MHz producirá por diferencia la frecuencia de 10,7 MHz (frecuencia intermedia), que se amplificará después.

No obstante, **también** una señal de 113,4 MHz, haciendo batimiento con el mismo oscilador local a 102,7 MHz, producirá (siempre por diferencia) la frecuencia de 10,7 MHz.

Esta "frecuencia imagen" se elimina de los **circuitos ajustados** de entrada, antes del mezclador, ya que de otro modo resulta totalmente imposible separarla de la señal en las etapas posteriores.

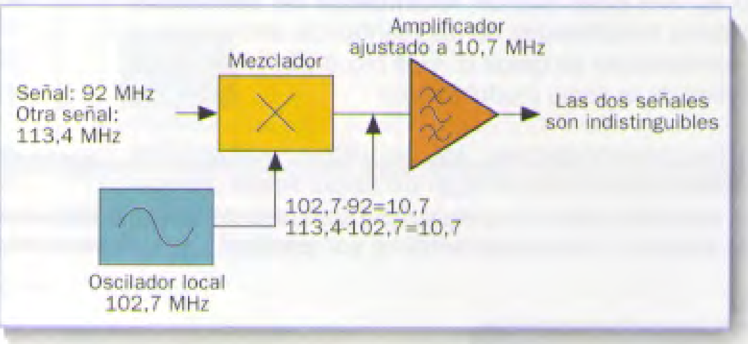
Señal: 92 MHz
Otra señal: 113,4 MHz
Mezclador
Amplificador ajustado a 10,7 MHz
Las dos señales son indistinguibles
102,7-92=10,7
113,4-102,7=10,7
Oscilador local 102,7 MHz

Son en realidad dos las frecuencias de entrada que se convierten a la frecuencia intermedia: una es deseada, pero la otra no.

DOBLE CONVERSIÓN

Si la frecuencia de entrada es muy alta con respecto a la frecuencia intermedia, por ejemplo, 2 GHz con respecto a 10 MHz, la imagen no deseada es muy **próxima** y prácticamente imposible de eliminar.

Es necesario, por tanto, una frecuencia intermedia **más elevada**, por ejemplo 200 MHz, de modo que los circuitos resonantes de entrada puedan atenuar fuertemente la imagen (distante ahora 400 MHz).

Como un desmodulador de 200 MHz no es práctico, se podrá utilizar una **segunda** conversión para reducir la frecuencia intermedia a un valor más accesible.

La doble conversión *se emplea, por ejemplo, en los receptores* **de satélite** *que funcionan a frecuencias muy elevadas.*

Antenas

La eficacia de una transmisión o recepción depende del correcto funcionamiento de estos dispositivos

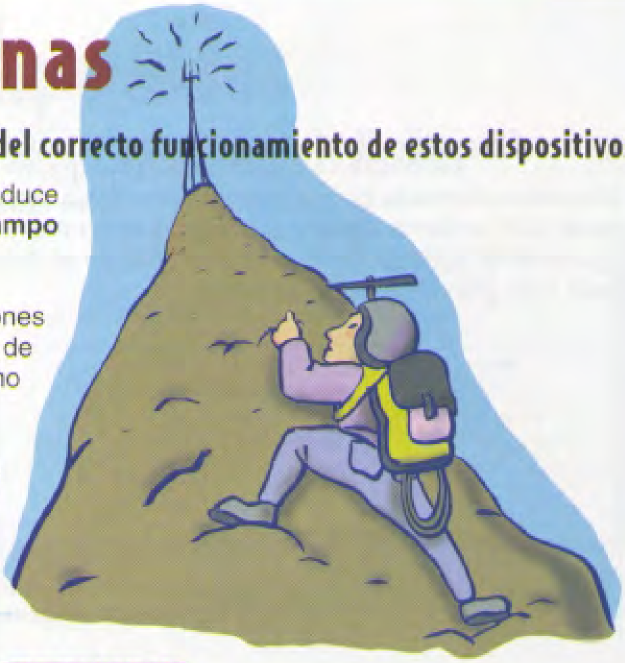

Una corriente alterna que circula por un cable no produce solamente un campo magnético sino, además, un **campo electromagnético**, es decir, una onda de radio.

Si se varía la longitud del cable, cuando sus dimensiones son próximas a la longitud de onda la energía irradiada de este modo aumenta notablemente: se comporta como **antena**.

Recordemos que la **longitud de onda** (λ) en el aire se obtiene dividiendo 300 por la frecuencia en MHz; por ejemplo, una onda de 100 MHz tiene una longitud de 300/100 = 3 metros.

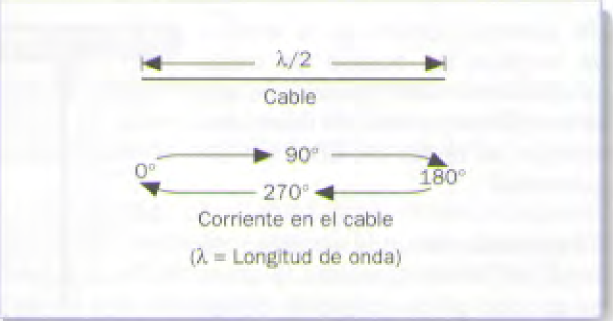

Resonancia *en un cable: la corriente da una vuelta y se encuentra de nuevo en el punto de partida justo a tiempo para* **el siguiente ciclo.**

RESONANCIA

La energía irradiada depende de la **corriente** que circula por el cable; a paridad de potencia, es máxima si el cable forma parte de un circuito resonante (ver lección 24).

También un simple cable puede resonar: basta que su longitud sea igual a la **mitad de la longitud de onda** (λ/2) o a un múltiplo. La señal empleará, para cruzar el cable, rebotar y volver al punto de partida, un tiempo similar a la duración de **un ciclo** de la onda eléctrica: estará, por tanto, en resonancia.

ONDAS ESTACIONARIAS

La corriente viaja a lo largo del cable (onda directa) y, cuando llega a su extremo, **vuelve hacia atrás**, produciendo una onda reflejada.

La suma de estas dos ondas en movimiento da lugar, en una antena resonante, a una onda aparentemente parada, llamada **onda estacionaria**.

En la práctica, en los extremos de un cable de longitud λ/2 la corriente es **cero** (la que va y la que vuelve se anulan), mientras que es máxima en el centro; ocurre lo contrario con la tensión.

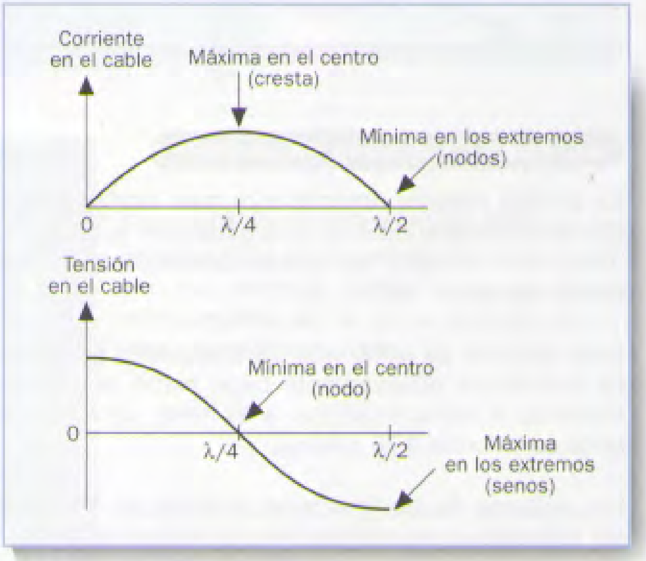

En un cable con una longitud de media onda se tiene una **onda estacionaria***: la tensión es máxima en los extremos, mientras que la corriente lo es en el centro.*

Tipos de antenas

El sencillo dipolo no es siempre la solución más práctica, especialmente en el mundo real

Un antena formada por un cable resonante a media onda (λ/2) se llama **dipolo**; su longitud es en realidad ligeramente superior porque la corriente en el cable viaja más plana que en el aire libre.

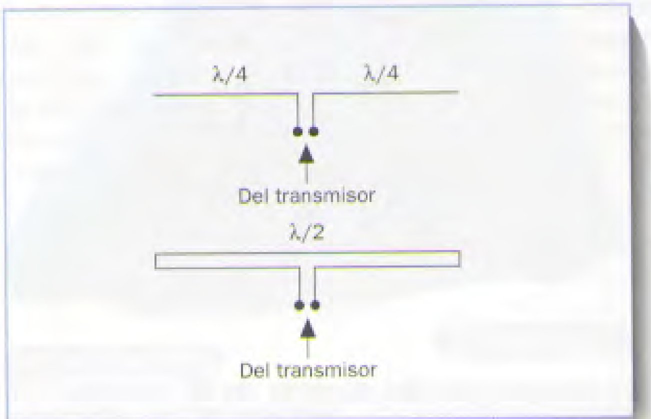

Dipolo sencillo y dipolo doblado, con la alimentación en el centro; en la práctica se utiliza a menudo un tubo metálico en lugar de cable.

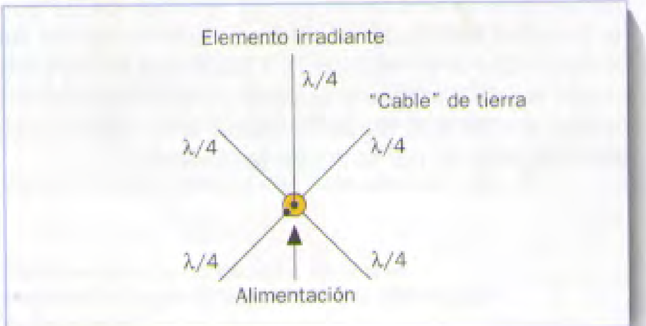

La señal que se transmite (o, en el caso de un receptor, el punto donde se toma) está **en el centro** del dipolo, donde la corriente es máxima.

En condiciones ideales, la antena (perfectamente ajustada) tiene una impedancia puramente **resistiva**, es decir, equivale a una resistencia que, en lugar de disipar energía, la transmite.

EFECTO DE TIERRA

Una antena **ideal** puede estar aislada en el espacio, pero las antenas reales deben trabajar próximas al terreno, que es conductor y modifica sus características.

Un ejemplo común es la antena de longitud **un cuarto de onda** (λ/4): la otra mitad viene dada por su imagen especular en tierra, por ejemplo el techo metálico de un automóvil.

En ausencia de un buen plano de tierra, se puede construir: la antena **ground plane** (plano de tierra) utiliza cuatro "cables" de un cuarto de onda para crearse su propia tierra.

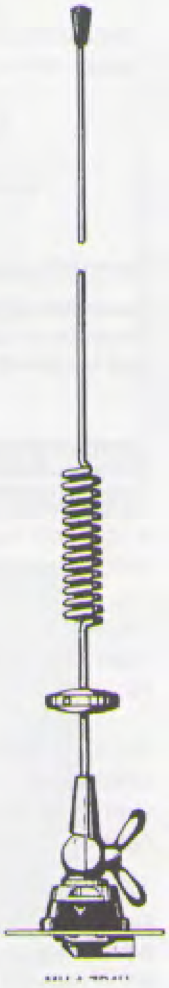

Antena en cuarto de onda con plano de tierra recreado artificialmente con cuatro cables horizontales (ground plane).

ANTENAS CARGADAS EN LA BASE

Es posible obtener antenas aún **más cortas** que un cuarto de onda añadiendo una bobina en la base para "cargarlas"; su valor se calibra para obtener la resonancia deseada.

Esta solución es adecuada para aparatos **portátiles** de frecuencia relativamente baja, como receptores-emisores o radiocomandos: a 27 MHz, un cuarto de onda sería unos 2,78 metros.

Las antenas de los **teléfonos móviles** de 900 MHz, sin embargo, pueden ajustarse sin añadir nada (λ/4 = 8 cm aprox.), pero es necesario acordarse de extenderlas de modo que funcionen correctamente.

En esta antena de automóvil, la bobina de carga se obtiene enrollando el mismo tubo metálico de la antena.

Antenas compuestas

Las interferencias pueden aprovecharse para obtener las características deseadas

La irradiación de un **dipolo perfecto**, situado lejos de tierra, es perfectamente circular; la señal es máxima en ángulo recto con respecto al cable, y disminuye hasta cero a lo largo del eje.

En la práctica, no obstante, es bastante difícil que se verifiquen estas condiciones ideales; la propia **tierra** influye notablemente, con absorciones y reflexiones, en el campo emitido por la antena.

La figura muestra el **cambio** en la irradiación de un dipolo de media onda, variando la altura a tierra: en el segundo caso, una parte de la energía se pierde hacia arriba.

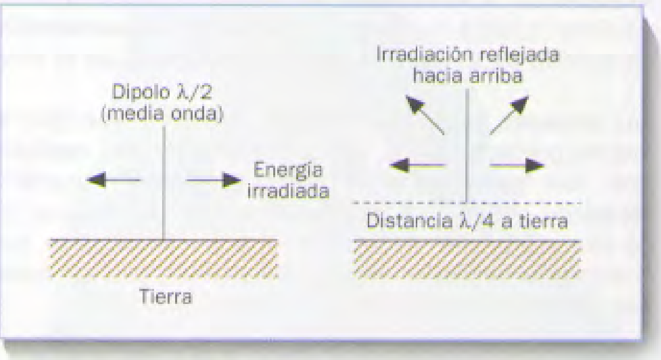

La distancia a tierra de un dipolo vertical influye en la distribución de la energía irradiada.

ANTENAS YAGI

Cualquier elemento conductor, incluso un trozo de cable, si se sitúa al lado de un dipolo **modifica** sus características de emisión (o de recepción).

Las antenas Yagi, por ejemplo las antenas receptoras de televisión normales, aprovechan este efecto para aumentar la eficacia de la antena, **concentrando** el campo en una única dirección.

Estas antenas utilizan trozos de cable (o tubo) como elementos directores para modificar su capacidad de **dirección**, y otros como **reflectores** para ignorar las señales procedentes de la dirección opuesta a la deseada.

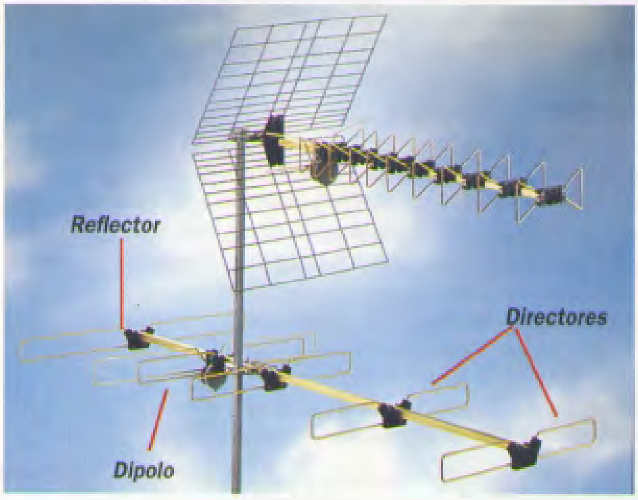

Las antenas Yagi para recepción de televisión tienen una elevada sensibilidad en la dirección en la que se apuntan, pero muy baja en las demás.

CONJUNTOS DE ANTENAS

En los transmisores, un sistema para mejorar la capacidad de dirección de la onda producida, es decir, concentrar su energía en la zona deseada, consiste en emplear **multitud** de dipolos.

Se puede tratar de una sencilla disposición **en línea**, de un retículo regular, o bien de una combinación más compleja.

Para reducir al mínimo los **efectos del terreno** y de las construcciones adyacentes, las antenas transmisoras se sitúan normalmente en la parte superior de torres especiales.

Cuando varias antenas se instalan en la misma torre es necesario calcular con precisión las posibles interferencias.

Más sobre las antenas

Las características de las antenas varían según su frecuencia y el uso al que están destinadas

Al aumentar la **frecuencia**, los efectos de interferencia se multiplican: a 2 GHz (2.000 MHz) basta un elemento de pocos milímetros, como un tornillo, para modificar las características de la antena.

Las **antenas para microondas** se construyen, por tanto, con mucha precisión, y no están conectadas con cables normales sino con "guías de onda" especiales, una especie de tubos metálicos rectangulares. Para disponer de máxima capacidad de dirección están dotadas de una **parábola**, que concentra la energía en la antena receptora o, en el caso del transmisor, alinea la energía emitida en la dirección deseada.

En una antena para **microondas**, *la parábola concentra sobre la antena las señales procedentes de la dirección a la que apunta.*

POLARIZACIÓN

Una onda electromagnética está polarizada con la misma **orientación**, horizontal o vertical, del dipolo que la emite (por precisión nos referimos a su componente eléctrica).

La antena receptora debe estar polarizada **del mismo modo**, ya que en otro caso la señal captada resulta mucho más débil (teóricamente nula), incluso con un ajuste perfecto.

Existen otras formas de polarización, como muestra la figura; la polarización **circular** se utiliza por ejemplo cuando la orientación relativa puede variar (por ejemplo, algunos satélites).

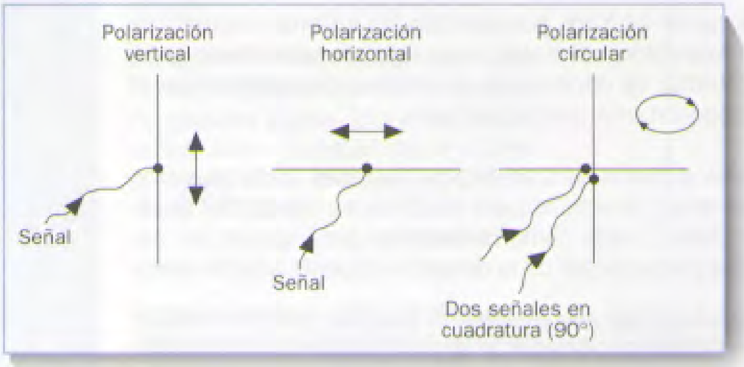

Distintos tipos de **polarización** *de la onda transmitida, según la disposición y la alimentación de los dipolos.*

CALIDAD DE LA ANTENA

Como en todo circuito resonante, el **factor de calidad** (Q) mide la capacidad de la antena de funcionar a una frecuencia muy precisa, con mínimas pérdidas.

En un transmisor, una antena de alto Q tendrá mayor **eficacia** y se calentará menos, mientras que en un receptor será más **selectiva** y contribuirá a aislar la frecuencia deseada.

Algunas antenas se construyen de modo que no son selectivas, es decir, que tienen una **sensibilidad similar** en cada frecuencia: son las empleadas para medir campos electromagnéticos.

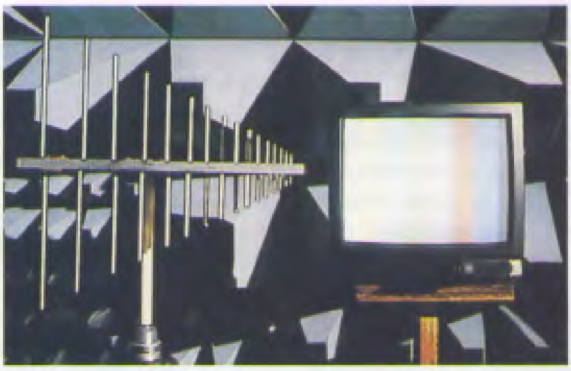

Una antena **log-periódica** *tiene ganancia (sensibilidad) constante en un campo de frecuencias bastante amplio.*

Altas frecuencias

A frecuencias elevadas el comportamiento de los circuitos es distinto del habitual

Al aumentar la frecuencia, los fenómenos **parásitos** (por ejemplo, capacidad e inductancia no deseadas) son cada vez más importantes, hasta el punto de condicionar el funcionamiento del circuito.

Por ejemplo, una capacidad parásita de 1 pF tiene una reactancia de aproximadamente 160 MΩ a la frecuencia de 1 KHz, y representa, por tanto, una molestia prácticamente **superflua**.

A 900 MHz (por ejemplo, muchos teléfonos móviles) la misma capacidad de 1 pF tiene sin embargo una reactancia de más o menos 177 Ω, por lo que ofrece una **ruta fácil** para la corriente, y no puede ser ignorada.

ELIMINACIÓN DE LOS PARÁSITOS

Como es evidentemente imposible eliminar del todo capacidades e inductancias no deseadas, sólo queda encontrar el modo de **compensarlas** introduciendo otros elementos en el circuito.

Se utilizan a menudo elementos reactivos de signo opuesto, por ejemplo inductores calculados para **resonar** con las capacidades parásitas; estas últimas se hacen, por tanto, parte del auténtico circuito.

En general se utilizan inductores o condensadores **variables** para ajustar el circuito a la frecuencia de trabajo, de modo que se tenga un funcionamiento óptimo.

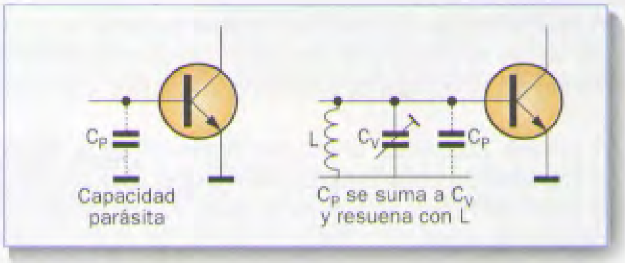
C_P
Capacidad parásita
L
C_V
C_P
C_P se suma a C_V y resuena con L

Para compensar la capacidad parásita C_P se puede hacer que forme parte del circuito resonante L-C_V.

CABLES COMO ANTENAS

A altas frecuencias, cada trozo de cable se convierte en una antena y puede, por tanto, **irradiar** ondas electromagnéticas, o bien recibirlas.

Un clásico ejemplo son las **conversaciones** de radioaficionados (a menudo distorsionadas) que a veces se escuchan por teléfono o en el equipo de alta fidelidad, especialmente con un viejo giradiscos.

La señal es **captada** de hecho por los cables que resuenan a la frecuencia de la onda, mientras que las uniones de los primeros pasos del amplificador funcionan como diodos detectores en modulación de amplitud.

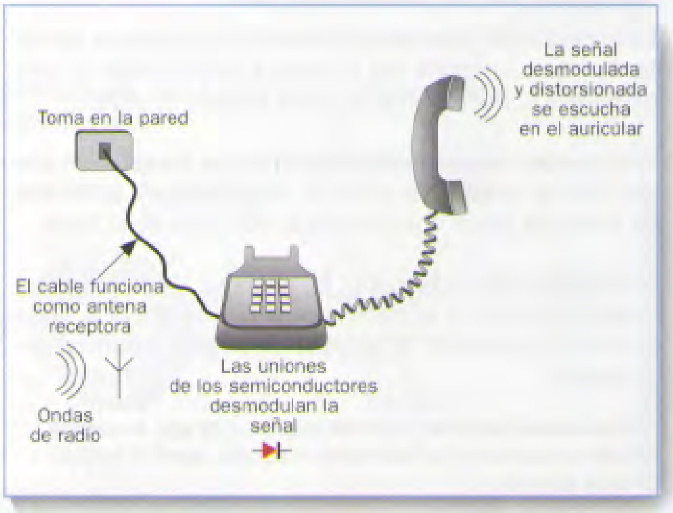
Toma en la pared
La señal desmodulada y distorsionada se escucha en el auricular
El cable funciona como antena receptora
Ondas de radio
Las uniones de los semiconductores desmodulan la señal

Procedencia de las voces misteriosas que se pueden oír en un teléfono estando cerca de transmisores de radio.

Líneas de transmisión

Una correcta adaptación de la impedancia elimina los efectos parásitos en los cables que llevan señales a alta frecuencia

La irradiación emitida por un cable es debida principalmente a las ondas estacionarias causadas por la **reflexión**: si el cable fuera infinito, la señal no se reflejaría.

Se puede simular un cable infinito utilizando un **par de cables** situados a una distancia constante, y cerrados en el extremo con una resistencia de un valor adecuado. Esta es una **línea de transmisión**, que tiene una "impedancia característica" precisa que depende de su construcción; si la resistencia tiene exactamente este valor, es como si la línea fuera infinita.

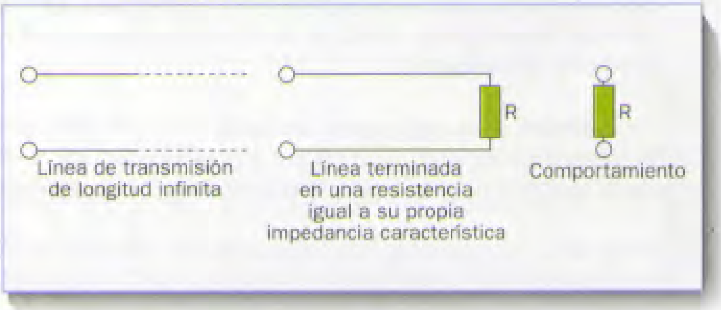
R
R
Línea de transmisión de longitud infinita
Línea terminada en una resistencia igual a su propia impedancia característica
Comportamiento

Una **línea de transmisión,** *infinita o terminada en su propia impedancia característica, aparece como una resistencia pura.*

CABLES COAXIALES Y GUÍAS DE ONDA

Un caso particular de línea es el **cable coaxial**; en él un cable va por el centro y el otro forma la pantalla externa, teniendo así la ventaja de tener una irradiación casi nula.

Los cables coaxiales utilizados habitualmente tienen una impedancia característica de 50 Ω, o bien de 75 Ω, que debe ser mantenida durante **todo el recorrido,** conectores incluidos.

A frecuencias aún más elevadas (GHz) se emplean las **guías de onda**, tubos metálicos especiales con forma rectangular realizados con una gran precisión mecánica.

Cable coaxial *de 50 Ω y conectores del tipo BNC, también con la misma impedancia característica.*

LÍNEAS EN CIRCUITO IMPRESO

La impedancia característica debería mantenerse **constante** hasta la salida del transistor amplificador (u otro dispositivo), por lo tanto ya en el circuito impreso.

Esto puede hacerse realizando líneas de transmisión con las mismas **pistas** de cobre (u otro material), junto con un plano de masa que recubre el otro lado de la base.

Calculando con precisión la anchura de las pistas, según el grosor y el tipo de material de soporte, esta técnica ("microstrip" o "stripline") asegura **óptimos** resultados.

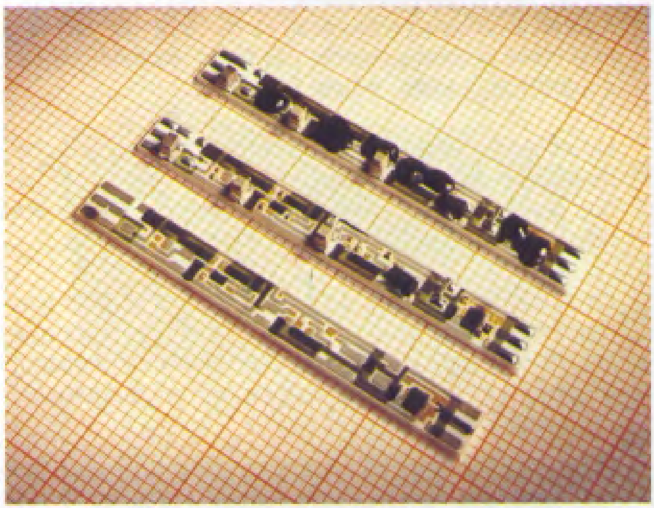

Prácticamente todos los circuitos modernos de alta frecuencia utilizan la técnica de las **microstrip** *realizadas sobre el propio circuito impreso.*

Adaptación de impedancia

Si la impedancia no coincide, es necesario introducir un circuito adaptador para evitar ondas estacionarias

En cada unión entre líneas donde haya una **inadaptación de impedancia**, es decir, donde las impedancias no sean idénticas, se tiene una reflexión parcial de la señal.

Esto es especialmente importante en un transmisor, donde la **señal reflejada** no se irradia y puede volver al amplificador y disipándose en calor.

La **relación de onda estacionaria** (SWR: standing wave ratio) indica la benignidad de la adaptación de impedancia: su valor ideal es uno, y se hace infinito en caso de reflexión total.

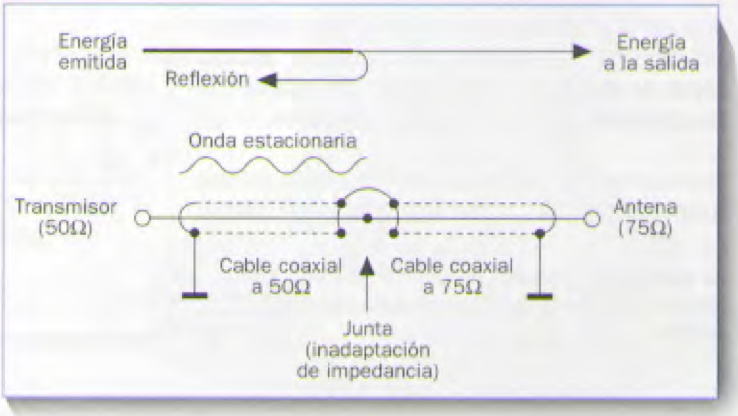
Energía emitida
Reflexión
Energía a la salida
Onda estacionaria
Transmisor (50Ω)
Antena (75Ω)
Cable coaxial a 50Ω
Cable coaxial a 75Ω
Junta (inadaptación de impedancia)

Una* inadaptación de impedancia *envía hacia atrás parte de la energía y crea una onda estacionaria en la línea.

ADAPTADORES DE ANTENA

Un **fragmento** de línea de transmisión presenta distintas características (resistivas y/o reactivas) a la entrada, según la impedancia en la que se cierra a la salida, y de su longitud.

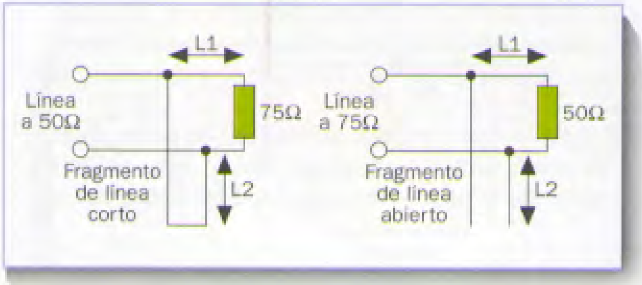
L1
Línea a 50Ω
75Ω
Fragmento de línea corto
L2
L1
Línea a 75Ω
50Ω
Fragmento de línea abierto
L2

Esta propiedad puede aprovecharse para **adaptar la impedancia** de la antena a la del cable, utilizando fragmentos de línea, abiertos o en cortocircuito en un extremo.

La longitud de los fragmentos y el punto de inserción en la línea no dependen sólo de las impedancias, sino también de la **longitud de onda**; de hecho, el sistema funciona sólo para una determinada frecuencia.

Adaptación de impedancia *con fragmentos de longitud calculada (L2), situados a una distancia exacta (L1).*

BALUN

A menudo la antena es **simétrica**, como en el caso del dipolo, mientras que la línea de transmisión es asimétrica, normalmente un cable coaxial. En este caso se utiliza un adaptador llamado **balun** (balanced-unbalanced: equilibrado-no equilibrado), que en las antenas receptoras (por ejemplo de televisión) se realiza generalmente en forma de transformador.

En **transmisión** se prefiere habitualmente utilizar fragmentos de línea o, para potencias superiores, estructuras metálicas de características eléctricas equivalentes.

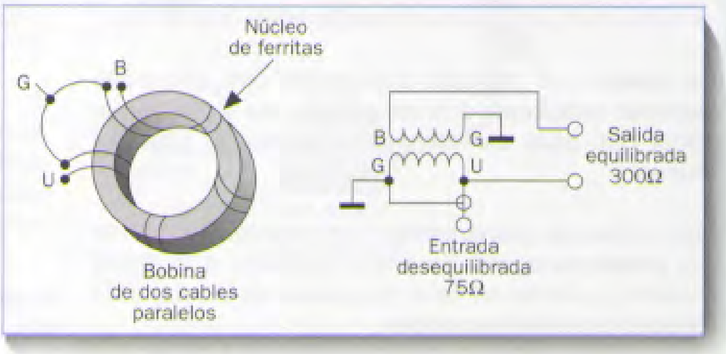
Núcleo de ferritas
G
B
U
Bobina de dos cables paralelos
B
G
G
U
Salida equilibrada 300Ω
Entrada desequilibrada 75Ω

Balun *de transformador, realizado con una bobina de dos cables: produce dos tensiones simétricas en contrafase.*

Técnicas para altas frecuencias

Algunos ejemplos de soluciones empleadas en circuitos que funcionan a frecuencias medias y elevadas

En los **transformadores** ajustados, las dos bobinas están dotadas de un condensador que los transforma en otros circuitos resonantes.

Una característica interesante de estos transformadores es que su anchura de banda depende también y sobre todo del **grado de acoplamiento** entre los dos inductores.

Si es estrecha, se tiene una respuesta en frecuencia con una **banda** relativamente ancha, útil por ejemplo en los amplificadores a frecuencia intermedia de los receptores de radio y de televisión.

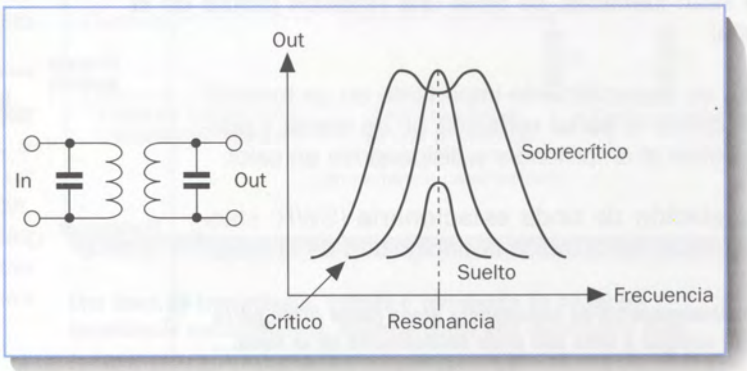

La anchura de banda ***de un transformador ajustado depende del grado de acoplamiento entre las bobinas.***

MEZCLADORES

Mientras a bajas frecuencias (por ejemplo en campo de audio) se trata de reducir al mínimo toda forma de **distorsión** y de influencia recíproca entre las señales, a veces en alta frecuencia ocurre lo contrario.

Por ejemplo, los **mezcladores** tienen justo la misión de causar una interacción entre dos señales, produciendo las frecuencias suma y resta (en campo de audio, el término “mezclador” indica, por tanto, un simple sumador).

Un mezclador ideal efectúa una **multiplicación** entre las dos señales; en la práctica no ocurre que sea un multiplicador perfecto, pero debe tener un comportamiento no lineal.

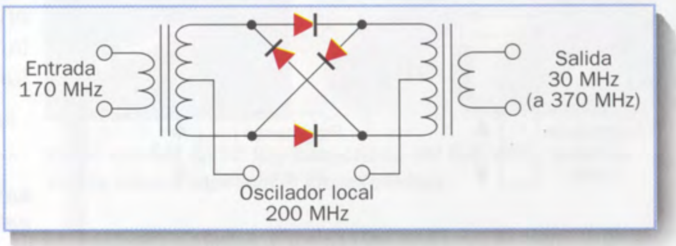

Modulador de anillo, ***aquí utilizado como mezclador para convertir una frecuencia en otra.***

BLINDAJES

Teniendo en cuenta la capacidad de las altas frecuencias de aprovechar toda capacidad mínima entre dos cables, es esencial que las distintas etapas de un circuito estén perfectamente **separadas** una de otra.

Es posible por ejemplo protegerlas con pequeñas **paredes metálicas**, dispuestas a su vez en un contenedor que aísle electromagnéticamente el todo del exterior.

Los cables de alimentación transitan habitualmente por **condensadores de paso**, soldados a la pared metálica, y tienen en serie inductores de bloque para detener las altas frecuencias.

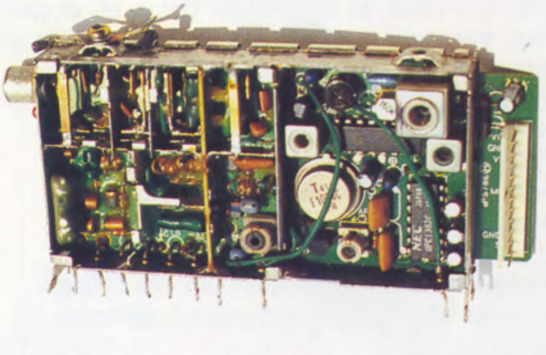

Un sintonizador ***de televisión funciona a cientos de MHz y está, por tanto, totalmente blindado.***

Optoelectrónica

La luz pertenece a la misma familia que las ondas de radio, tan sólo tiene una frecuencia diferente

Gran parte de las ondas electromagnéticas son **invisibles**, excepto las que rodean a los 500.000 GHz, ya que constituyen la luz visible.

La **frecuencia** determina su color, que varía en un espectro continuo, similar al del arco iris, del rojo hasta el violeta.

Normalmente se suele indicar la **longitud de onda** en lugar de la frecuencia; se mide en nanómetros (nm: milmillonésimas de metro).

Colores	**Longitud de onda**
Violeta	Inferior - 450 nm
Azul	450 - 500 nm
Verde	500 - 570 nm
Amarillo	570 - 590 nm
Naranja	590 - 610 nm
Rojo	610 - 700 nm
Blanco	Es la combinación de todas las longitudes de onda
Negro	Es la total absorción de todas las longitudes de onda

Nuestro ojo ve distintos colores según la* longitud de onda *de la luz que incide en él.

DIODOS QUE PRODUCEN LUZ

En un diodo normal de semiconductor recorrido por corriente directa (ver lección 8), electrones y huecos se **recombinan** en la parte central de la unión.

Esta recombinación puede causar la emisión de **radiación electromagnética** que, si la longitud de onda está dentro de la gama arriba citada, aparece como luz visible.

Los **LED** (ver lección 19 de Componentes) son diodos especialmente construidos para producir luz aprovechando este fenómeno; el material utilizado determina la longitud de onda producida, y por lo tanto su color.

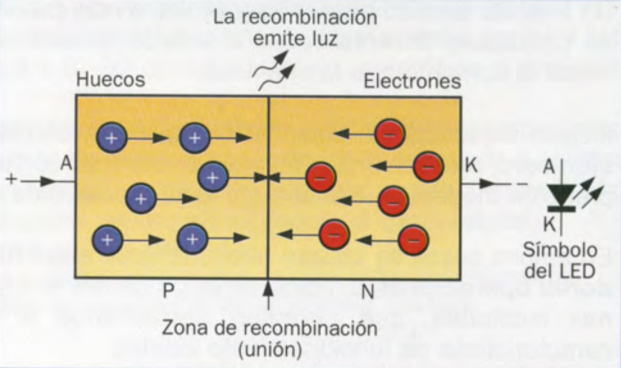
La recombinación emite luz
Huecos
Electrones
A
K
+
A
K
Símbolo del LED
P
N
Zona de recombinación (unión)

Los diodos luminosos (LED) se realizan con arseniuro de galio (GaAs) u otros materiales especiales.

VISIBLES E INVISIBLES

Antes los LED eran exclusivamente rojos, mientras que hoy se fabrican en una amplia gama de **colores**, que se extiende desde el infrarrojo hasta el azul y más allá.

La elevada velocidad de respuesta de los LED, privados de inercia térmica (a diferencia de las bombillas) permite **modular** fácilmente la luz emitida, variando la corriente en el diodo.

Las versiones **infrarrojas**, aunque invisibles al ojo humano, pueden enviar luz a un receptor adecuado, como de hecho ocurre en gran parte de los mandos a distancia.

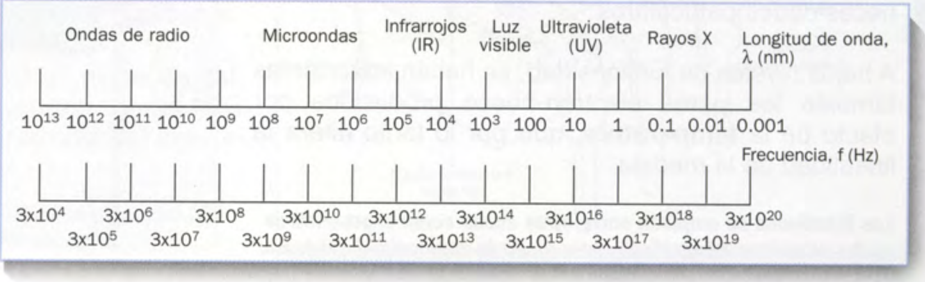
Ondas de radio
Microondas
Infrarrojos (IR)
Luz visible
Ultravioleta (UV)
Rayos X
Longitud de onda, λ (nm)
10¹³ 10¹² 10¹¹ 10¹⁰ 10⁹ 10⁸ 10⁷ 10⁶ 10⁵ 10⁴ 10³ 100 10 1 0.1 0.01 0.001
Frecuencia, f (Hz)
3x10⁴ 3x10⁵ 3x10⁶ 3x10⁷ 3x10⁸ 3x10⁹ 3x10¹⁰ 3x10¹¹ 3x10¹² 3x10¹³ 3x10¹⁴ 3x10¹⁵ 3x10¹⁶ 3x10¹⁷ 3x10¹⁸ 3x10¹⁹ 3x10²⁰

La luz visible no es otra cosa que una pequeña parte del amplio espectro de las ondas electromagnéticas.

Fotodiodos

Igual que existen dispositivos de semiconductor capaces de emitir luz, existen otros que desarrollan la función opuesta

Las **fotoresistencias** varían su resistencia en función de la luz recibida, pero el comportamiento no es en absoluto lineal y, sobre todo, la respuesta a las variaciones es bastante lenta.

Se comportan mucho mejor los **fotodiodos**, en cuya unión se producen pares electrón-hueco en número proporcional a la cantidad de luz recibida.

Resulta así una corriente proporcional a la intensidad (potencia luminosa) de la propia luz, entre otras cosas **lineal** en una gama bastante extensa de valores de luminosidad.

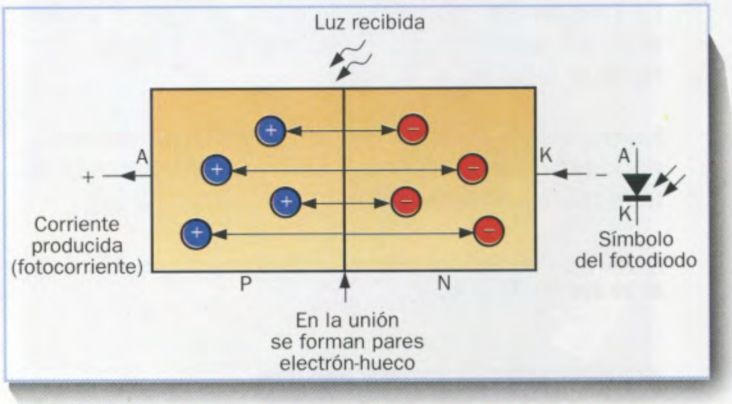
Luz recibida
+
A
+
−
K
A
−
K
Corriente producida (fotocorriente)
P
N
En la unión se forman pares electrón-hueco
Símbolo del fotodiodo

Fotodiodo: ***la energía de la luz aleja los electrones de los átomos de silicio dopado, dejando los correspondientes huecos.***

AMPLIFICAR LA SALIDA

Un método sencillo para utilizar un fotodiodo consiste en polarizarlo **inversamente** a tensión constante y medir la corriente que lo atraviesa.

Incluso dejándolo sin polarización, es decir, con **tensión cero**, produce una corriente proporcional, con una precisión moderada, a la energía luminosa recibida.

En ambos casos se utilizan habitualmente **amplificadores operacionales**, como veremos dentro de algunas lecciones, que permiten aproximarse a las características de funcionamiento ideales.

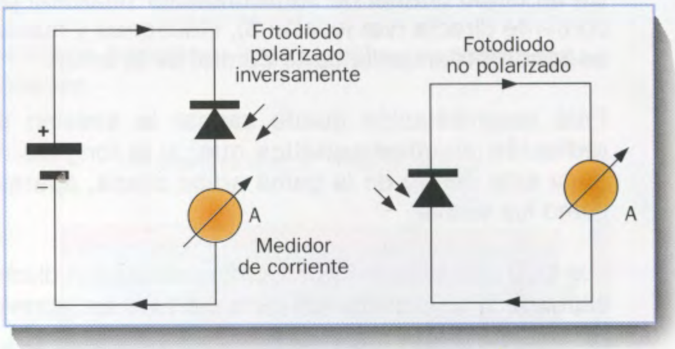

Medida ***de la luz recibida por un fotodiodo; en lugar de la herramienta se utiliza normalmente un amplificador de corriente.***

LÍMITES DE LOS FOTODIODOS

La corriente producida por un fotodiodo no es **constante** para todas las longitudes de onda; luces de intensidad similar, pero de distinto color, tendrán un efecto distinto. Naturalmente, cada modelo tiene una **curva de respuesta** propia a la longitud de onda de la luz a la entrada; es necesario elegir según las necesidades particulares.

A bajos niveles de luminosidad, se hacen importantes también los pares electrón-hueco producidos por efecto de la **temperatura**, que por lo tanto altera la linealidad de la medida.

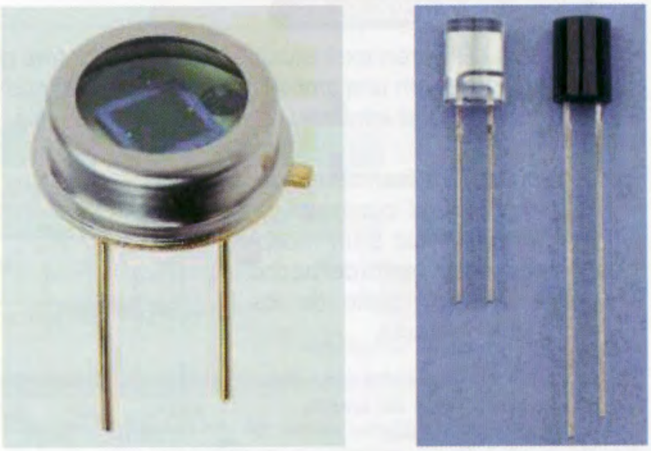

Los fotodiodos se emplean entre otras cosas como exposímetros en las máquinas fotográficas, para medir la luminosidad del tema.

Fototransistores y fotoacopladores

A veces no interesa tanto la linealidad como una elevada sensibilidad

La corriente producida por un fotodiodo es muy débil, pero el mismo **efecto fotoeléctrico** se obtiene iluminando la unión base de un transistor.

La **corriente** así obtenida produce una evidente variación de la corriente de colector, como si se hubiera introducido en la base desde el exterior.

Un fototransistor **amplifica,** por tanto, las variaciones de la luz que incide en él, aunque linealidad y dependencia de la temperatura son generalmente peores con respecto a los fotodiodos.

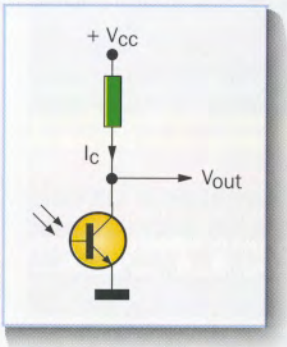

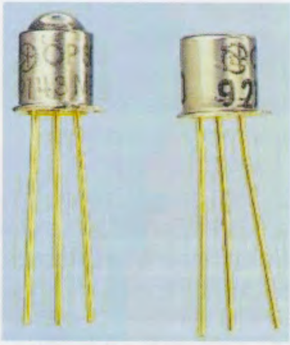

Fototransistor: *una pequeña variación de luz produce una variación relativamente elevada de I_C, y, por tanto, de V_{out}.*

AISLAMIENTO ÓPTICO

En muchas aplicaciones es necesario **aislar eléctricamente** un circuito con respecto a otro, es decir, evitar todo contacto, pero en cualquier caso hacer circular una señal entre ambos.

Ello se puede obtener fácilmente utilizando un **LED** y un **fototransistor**: la corriente en el primero produce luz, que a su vez causa el paso de corriente en el segundo.

Esta forma de aislamiento óptico es sólo aproximadamente **lineal** y es, por tanto, poco adecuada a señales analógicas, aunque hay técnicas para mejorarla; no obstante, se emplea mucho en el campo digital.

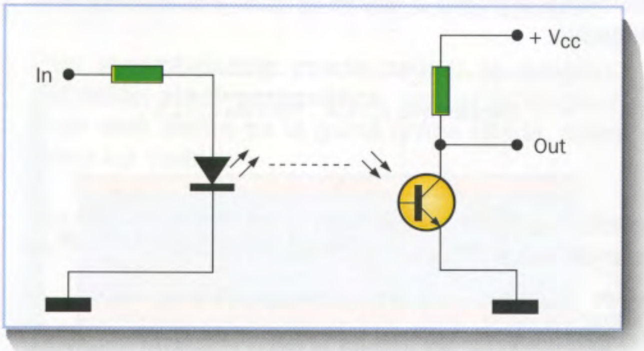

Aislamiento óptico *realizado con un LED y un fototransistor: entre los dos circuitos no hay contacto eléctrico.*

FOTOACOPLADORES

Los aisladores ópticos son de uso tan común que se fabrican normalmente con los dos componentes ya montados en el interior de un **único encapsulado** como integrado.

Gracias al uso de un material **aislante** transparente adecuado, algunos de estos dispositivos pueden resistir incluso a varios miles de voltios entre los dos circuitos.

En general es preferible emplearlos de modo **digital**, convirtiendo una eventual señal analógica con un ADC antes del fotodiodo, y reconstruyéndola tras el fototransistor con un DAC.

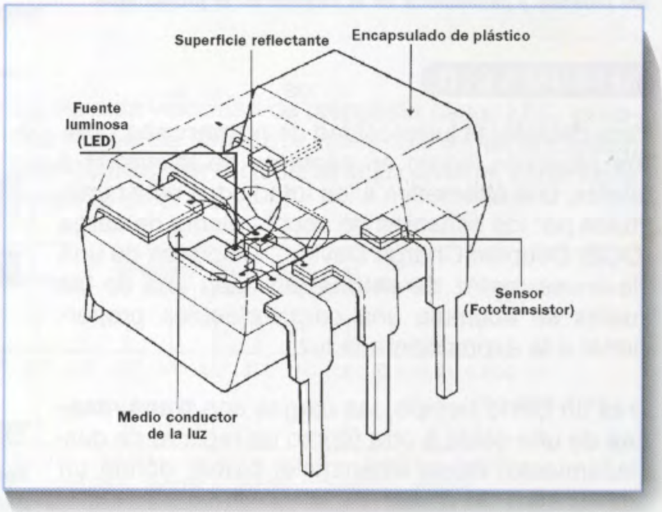

Estructura de un **fotoacoplador:** *a la derecha los pines del LED emisor, a la izquierda los del fototransistor receptor.*

Otros dispositivos optoelectrónicos

La luz está tomando un papel cada vez más importante en electrónica

Un **láser** es un dispositivo capaz de emitir luz "coherente", es decir, de una sola frecuencia y perfectamente en fase, gracias a una resonancia en el dispositivo emisor.

Esto puede llevarse a cabo con un tubo de descarga de un gas, o más sencillamente con un **diodo** láser especial, de potencia relativamente limitada pero práctico y económico.

La facilidad de **enfoque** con precisión de la luz láser la hace perfecta para la lectura de microscópicas variaciones superficiales, como en el caso de los discos compactos y los DVD.

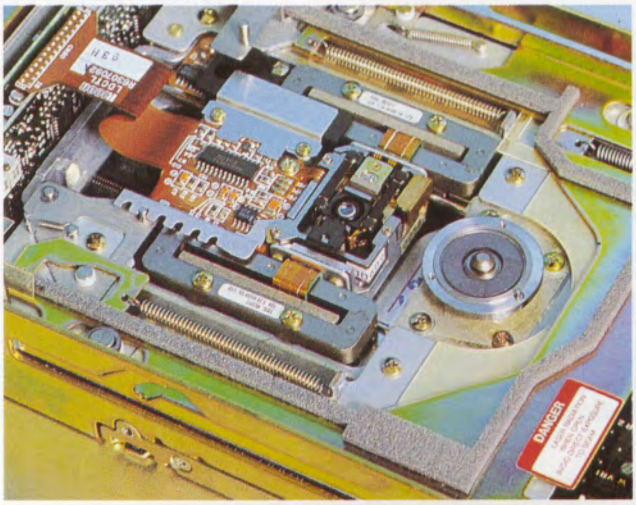

Los datos se imprimen en un CD en forma de minúsculas depresiones, leídas por un haz **láser** ***localizado por una óptica de precisión.***

FIBRAS ÓPTICAS

Los **cables coaxiales** sufren dos problemas: una discreta atenuación de la señal, y una banda de paso relativamente limitada.

Se puede enviar mucha más información empleando una **fibra óptica**, es decir, un fino tubo flexible de un material transparente, capaz de transportar una o varias ondas luminosas incluso a una gran distancia.

Presenta entre otras cosas la ventaja de no ser un conductor eléctrico, y de resultar, por tanto, totalmente **inmune** a interferencias y corrientes no deseadas.

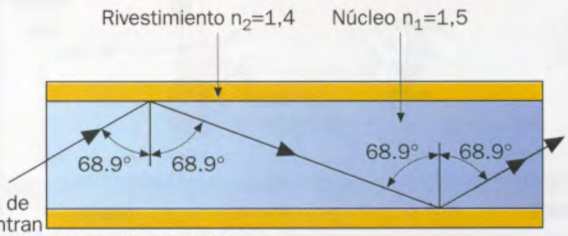

Los rayos de luz que entran en la fibra con un ángulo mayor del crítico (68,9° en este caso) se reflejan internamente a lo largo de la fibra.

En una **fibra óptica*****, la luz es reflejada continuamente por las paredes y permanece en el interior de la propia fibra.***

SENSORES CCD

Para detectar la luminosidad de **numerosos puntos** próximos, como en cámaras de televisión o telefax, una alternativa a los fotodiodos está constituida por los sensores de acoplamiento de carga (CCD: Coupled Charge Device). Disponen de una fila o una matriz de **celdas**, en cada una de las cuales se acumula una carga eléctrica proporcional a la exposición a la luz.

Tras un cierto tiempo, las cargas son **transvasadas** de una celda a otra (como un registro de desplazamiento) hasta alcanzar el borde, donde un amplificador las mide.

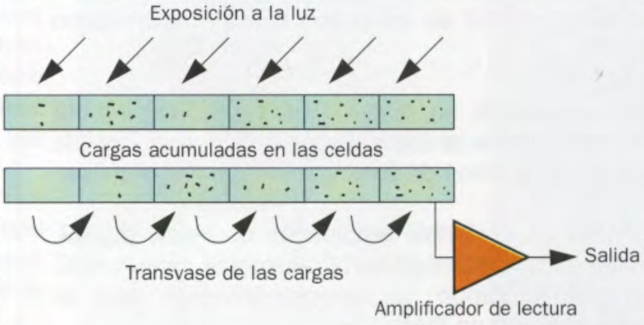

Sensor CCD: ***tras la exposición, las celdas son leídas otra vez; el proceso es lento, pero relativamente económico.***

Diodos controlados (SCR)

Estos curiosos componentes son capaces de "saltar" de un estado a otro

Conectando entre sí dos transistores **complementarios** (es decir, un NPN y un PNP), como en la figura, no hay paso de corriente.

En realidad, los dos transistores están ambos en **interdicción**, ya que no hay nada que haga pasar corriente en las uniones base-emisor.

No obstante, es interesante observar que **cada** transistor, si de algún modo se cerrara, proporcionaría abundante corriente a la base del otro.

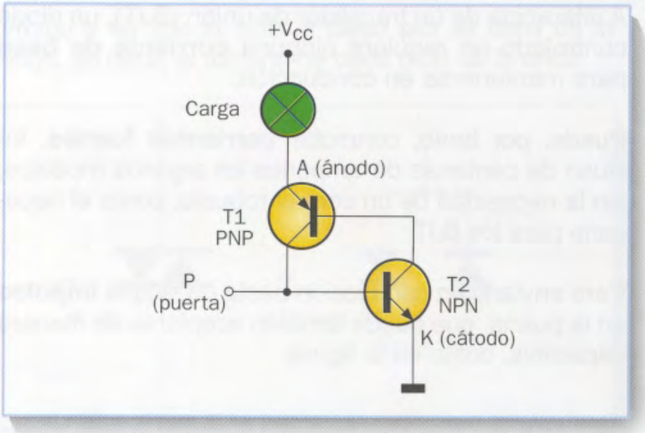

Dos transistores conectados de este modo no conducen... al menos hasta que llegue un estímulo externo.

CONDUCCIÓN AUTOSOSTENIDA

Enviando una pequeña corriente a la base de T2, este **empieza** a conducir; su conducción proporciona, por tanto, corriente a la base de T1, que a su vez refuerza la corriente de base de T2.

El circuito se comporta por lo tanto como un interruptor que se **cierra y permanece cerrado**; la tensión en sus extremos baja al mínimo valor suficiente para dejar pasar corriente de base.

El único modo de reabrir este interruptor consiste en **quitar tensión**, de modo que se interrumpa (aunque sólo por un instante) la corriente en los dos transistores.

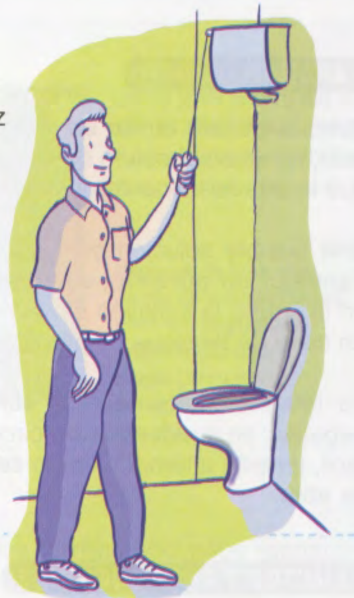

Una comparación irreverente: una vez activada, la carga (corriente) continúa hasta que hay agua (tensión) disponible.

CUATRO REGIONES PARA UN SCR

El dispositivo arriba descrito se realiza normalmente en **un solo chip**, utilizando cuatro regiones dopadas P-N-P-N, como en la figura: las dos regiones centrales son comunes a los dos transistores.

Toma el nombre de "thyristor" o **SCR** (Silicon Controlled Rectifier: rectificador controlado de silicio), ya que se comporta como un diodo que se activa bajo control.

Los tres terminales se llaman ánodo, cátodo y **puerta**: esta última es el electrodo de control que permite colocarlo en conducción.

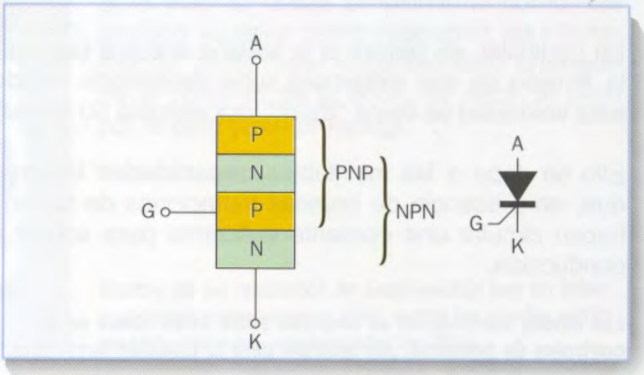

Estructura de un SCR y su símbolo: un diodo con un electrodo adicional de control, la puerta.

Circuitos con SCR

Las características de los diodos controlados les hacen útiles en varias circunstancias

A diferencia de un transistor de unión (BJT), un diodo controlado no requiere ninguna **corriente de base** para mantenerse en conducción.

Puede, por tanto, conmutar **corrientes fuertes**, incluso de centenas de amperios en algunos modelos, sin la necesidad de un control robusto, como el necesario para los BJT.

Para enviarlo en conducción basta un simple **impulso** en la puerta, que puede también acoplarse de manera capacitiva, como en la figura.

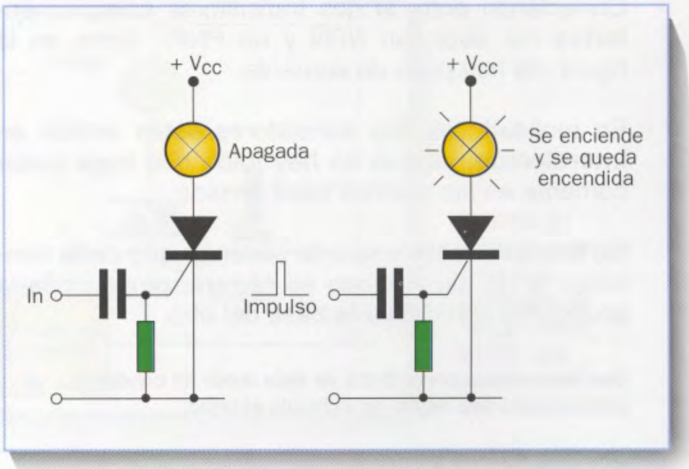

***Acoplamiento capacitivo:* un impulso en la puerta activa la conducción del SCR, que después se mantiene por sí solo.**

APAGAR UN SCR

Aunque es fácil enviar en conducción un diodo controlado, no es en absoluto sencillo **interrumpir** la corriente que lo atraviesa: no es posible hacerlo desde la puerta.

Una posible solución consiste en la utilización de un transistor **en paralelo**, como en el biestable mostrado en la figura; la solución es, sin embargo, poco práctica en caso de tensiones y corrientes fuertes.

Lo ideal es hacer que la **tensión** falte o se haga negativa: se pueden utilizar circuitos resonantes o, más fácil, tensión alterna: cuando se hace negativa, el SCR se abre.

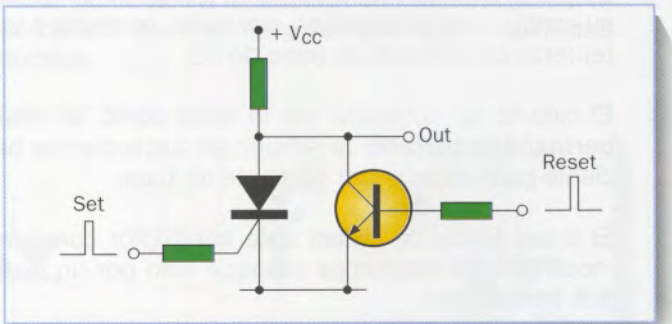

Flip-flop S-R: ***un impulso en la puerta hace conducir al SCR y otro en el transistor lo apaga "robándole" la corriente.***

LÍMITES Y PRECAUCIONES

Los diodos controlados tienen también una característica **no deseada**: la de tener el "gatillo fácil"; pueden cerrarse incluso sin haber recibido impulsos en la puerta.

En particular, se cierran si la tensión anódica (es decir, la tensión en sus extremos) sube demasiado **rápido**; esta velocidad se llama "dV/dt", por ejemplo 50 V/µseg.

Ello se debe a las inevitables **capacidades internas** que, en presencia de bruscas variaciones de tensión, hacen circular una corriente suficiente para activar la conducción.

Los diodos controlados se emplean entre otras cosas en los controles de potencia, por ejemplo para la tracción ferroviaria.

Triac

En corriente alterna, es cómodo disponer de un diodo controlado bidireccional

Un "triac" es similar a un SCR normal, pero puede funcionar con **ambas** polaridades, como si estuviera constituido por los dos SCR en antiparalelo mostrados en la figura.

La tensión entre puerta de cátodo, necesaria para activar la conducción, puede ser siempre positiva o, en muchos modelos, **negativa** cuando es negativa también la tensión anódica.

Un triac se utiliza normalmente para conmutar corrientes alternas, y se apaga sólo al **paso por el cero** de la tensión, es decir, al término de cada ciclo de la onda.

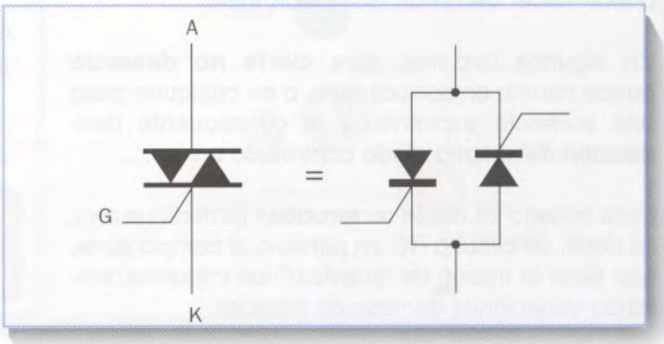

Un triac puede imaginarse como dos SCR, de los que uno está al revés; sin embargo el control está siempre entre puerta y cátodo.

CONTROL DE FASE

Un triac puede utilizarse como sencillo relé estático, para conmutar una fuerte corriente alterna, pero se puede también aprovechar para construir **reguladores** como los de las bombillas domésticas.

Basta activarlo, en cada ciclo, con un cierto **retraso**: permanecerá cerrado hasta el término del mismo ciclo, cuando la tensión pase por el cero.

Regulando el retraso, se podrá así **variar** la cantidad de energía proporcionada a la carga en cada ciclo, y por lo tanto (por ejemplo) la luminosidad de una bombilla de incandescencia.

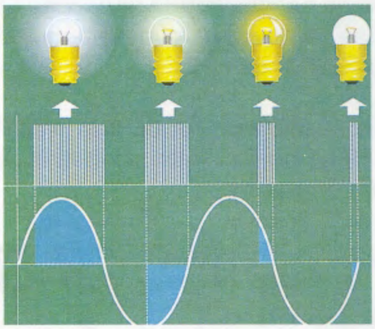

Variando la fase del punto de activación, se controla la energía proporcionada a la carga (área azul) durante cada ciclo.

ZERO CROSSING

La **brusca conmutación** de activación de un triac produce un impulso, que puede contener muchos armónicos de alta frecuencia (los reguladores de luz de mala calidad producen interferencias en radios y televisores).

Son, por tanto, necesarios **filtros** adecuados (generalmente de tipo LC, ver lección 17), que evitan este desagradable inconveniente; las normas actuales son, entre otras cosas, bastante severas en este punto.

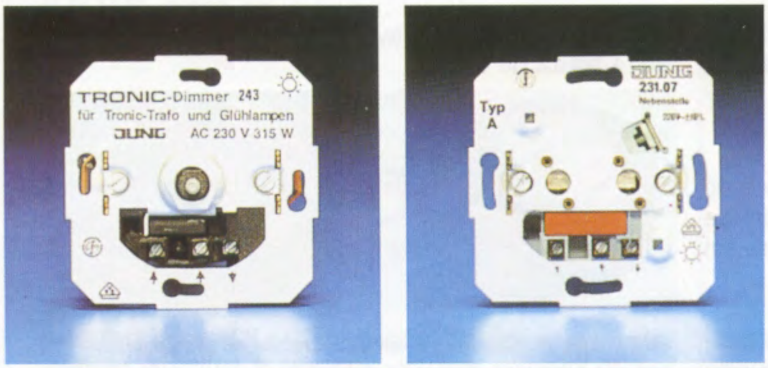

Si el triac se utiliza simplemente como relé estático, es decir, como interruptor, las interferencias se eliminan encendiéndolo **enseguida** al principio de la onda, es decir, al paso por el cero (zero crossing).

Dentro de un regulador de luminosidad hay un triac y algunos componentes más, entre los cuales están un inductor y un condensador de filtro.

Detalles sobre SCR y triac

Precauciones circuitales y versiones especiales relativas a estos dispositivos

Hemos visto que una **rápida variación** de tensión (es decir, una elevada dV/dt), causada por ejemplo por el apagado de un motor en otro lugar, puede hacer cerrar un SCR o un triac.

En algunos circuitos, este **cierre no deseado** puede causar un cortocircuito, o en cualquier caso una corriente excesiva, y la consecuente destrucción del mismo diodo controlado o triac.

Para evitarlo se utiliza un **snubber** (amortiguador), es decir, un circuito RC en paralelo al componente, que tiene la misión de "suavizar" los impulsos, evitando variaciones demasiado bruscas.

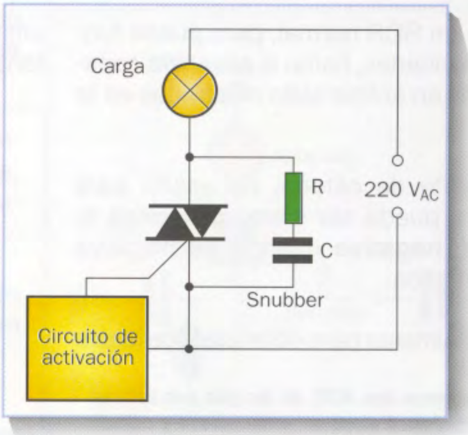

Snubber en paralelo a un triac para protegerlo de una excesiva dV/dt, que podría cerrarlo accidentalmente.

CONTROL Y AISLAMIENTO

Estos dispositivos se utilizan habitualmente para controlar circuitos alimentados con tensión relativamente **elevada**, por ejemplo, los 220 V de red.

Por otra parte, el circuito de pilotaje suele ser de **baja tensión**: por lo tanto es necesaria alguna forma de aislamiento entre las dos secciones.

Se puede utilizar, por ejemplo, un pequeño **transformador**, con la única misión de transmitir a la puerta el impulso procedente del circuito de activación, sin que haya continuidad eléctrica.

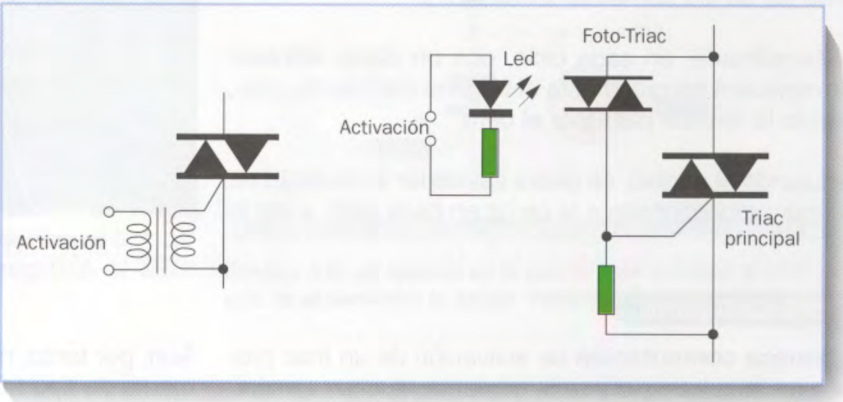

Activación con* transformador *y con* foto-triac*: dos modos de aislar los circuitos de control de la tensión aplicada a la carga.

FOTO-SCR Y TRIAC

Mejor todavía, se puede **aislar ópticamente** el triac (o el diodo controlado) de los circuitos que lo pilotan; sin embargo, existe el problema de transmitir energía suficiente para activarlo.

Una solución válida está constituida por SCR y triac **sensibles a la luz**: se activan cuando se enciende un LED montado en el mismo encapsulado, pero sin contacto eléctrico con ellos.

Pueden conmutar corrientes débiles, pero normalmente se utilizan a su vez para **controlar** la puerta de un dispositivo de mayor potencia, como se ilustra en la figura de arriba.

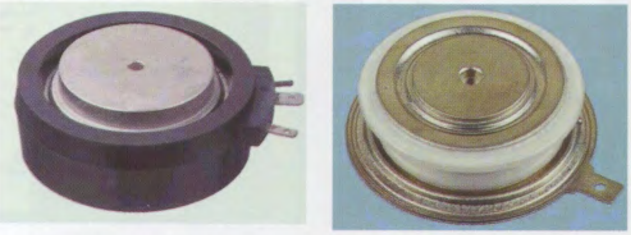

SCR y triac* de elevada potencia*, incluidos en encapsulados con una adecuada disipación del calor.

IGBT

Son transistores especiales, nacidos por exigencia de la electrónica de potencia

Los habituales transistores **bipolares** (BJT) tienen una tensión de saturación bastante baja incluso con corrientes fuertes, pero requieren corriente de base y tienen una ganancia escasa (ver lección 15).

Los **MOSFET**, por el contrario, no tienen estos problemas (ver lección 22), pero la caída de tensión en sus extremos sube rápidamente al aumentar la corriente y la temperatura.

Los **IGBT** (Insulated-Gate Bipolar Transistor: transistor bipolar de puerta aislada) reúnen las principales ventajas de los dos tipos antes citados, por lo que han encontrado mucha aplicación como conmutadores de potencia.

Un IGBT es, en un cierto sentido, un híbrido entre un MOSFET y un BJT, adecuado a su "nicho ecológico".

UN CURIOSO HÍBRIDO

La figura muestra el **símbolo** del IGBT que, como dicen las siglas, es un transistor BJT con una puerta aislada en lugar de la unión normal de base.

El **circuito equivalente** muestra que un IGBT está esencialmente formado por un MOSFET (de tipo enhancement) que controla un transistor bipolar PNP.

Naturalmente, en el encapsulado no hay en realidad dos componentes **distintos**, lo que ocurre es que está realizado con silicio de modo similar a un MOSFET; el esquema sirve sólo para dar una idea del comportamiento.

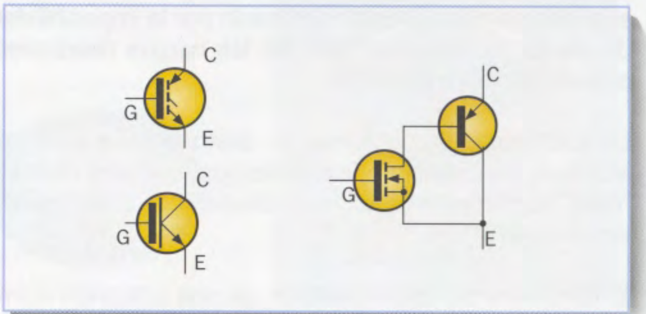

Símbolos ***de un IGBT, y el circuito equivalente (simplificado): se comporta como un MOSFET que controla un BJT.***

CONTROL DE LA TENSIÓN

Como un MOSFET, un IGBT se controla aplicando una **tensión** a la puerta: hace conducir el MOSFET del circuito equivalente, que a su vez proporciona corriente de base al transistor PNP.

A diferencia de un BJT no hay, por tanto, ninguna absorción de corriente de la puerta, lo que simplifica notablemente los circuitos de **control** de los interruptores de potencia, como con los MOSFET.

Es curioso observar que el **colector** de un IGBT (C en la figura de arriba) se corresponde en realidad con el emisor del transistor equivalente, que es un PNP montado al revés.

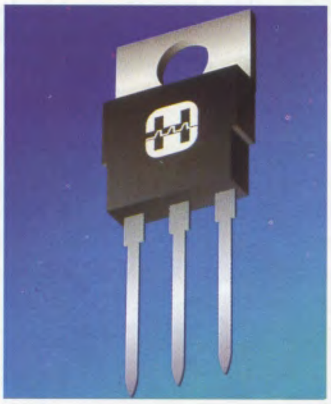

Un **IGBT** ***se presenta como un transistor normal de potencia.***

Virtudes y defectos de los IGBT

Son interruptores óptimos para tensiones altas y potencias medias

La figura muestra que la **ganancia** (transconductancia) de un IGBT, no baja al aumentar la corriente de colector como ocurre en los BJT, sino que varía aumentando.

La **caída de tensión**, a diferencia de lo que ocurre con los MOSFET, permanece baja también a corrientes elevadas, aunque siempre superior a una caída de diodo.

Los IGBT son también relativamente más fáciles de poner **en paralelo** entre sí con respecto a los BJT, ya que de todas formas no son tan cómodos y seguros como los MOSFET.

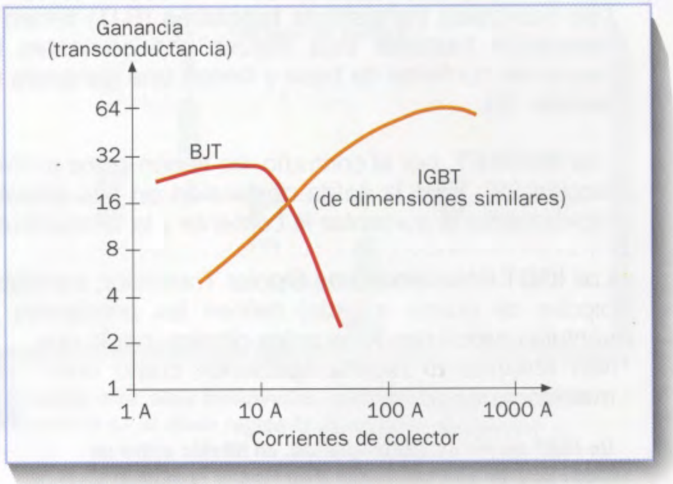

Mientras la ganancia de un transistor BJT baja rápidamente para corrientes fuertes, la de un IGBT sigue subiendo.

CARGAS RESIDUALES

Naturalmente, tampoco los IGBT son perfectos; uno de sus inconvenientes está constituido por la imposibilidad de **vaciar** la base del PNP de las cargas residuales, cuando se retira el control.

Un IGBT emplea, por tanto, un cierto **tiempo** para pasar de la conducción a la interdicción, con una molesta "cola" de corriente (que causa disipación), como puede verse en la figura.

El fenómeno se puede **reducir**, ya sea actuando sobre el control como utilizando modelos rápidos (pero más delicados en caso de exceso de corriente), pero no se puede eliminar del todo.

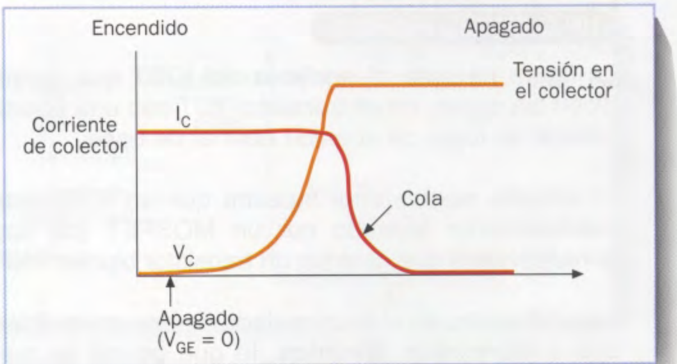

Es necesario un cierto tiempo para que la base se vacíe y la corriente de colector de un IGBT baje a cero.

USO DE LOS IGBT

Las ventajas de los MOSFET y los transistores bipolares se manifiestan de un modo cada vez más evidente con el aumento de la **tensión** de trabajo en los extremos del componente.

Los IGBT son por lo tanto ampliamente utilizados como **interruptores**, especialmente en tensiones bastante elevadas (incluso por encima de los 1.000 V), como por ejemplo en el control de motores eléctricos. Por distintos motivos, entre ellos la relativa sensibilidad a los cortocircuitos breves, el diseño de circuitos con IGBT es un poco más **delicado** que el de los circuitos equivalentes de MOSFET.

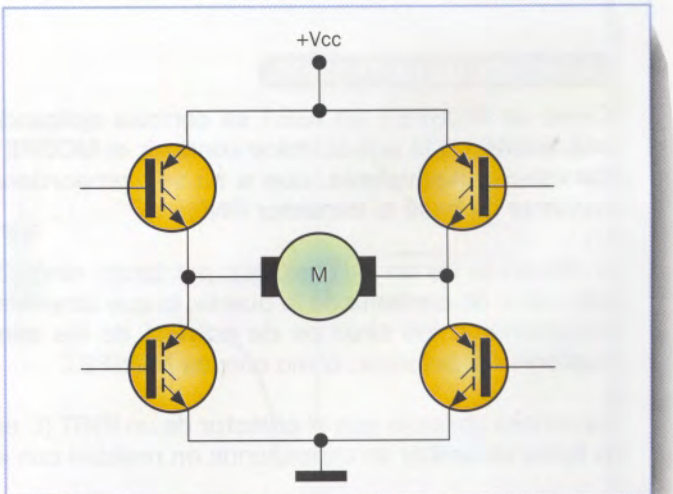

Motor controlado por un puente de IGBT: en las conmutaciones es necesario dejar tiempo para la "cola" de apagado.

GTO

Para potencias altas existen diodos controlados que se pueden apagar por la puerta

Un inconveniente de los SCR (diodos controlados, ver lección 30) está constituido por la necesidad de **interrumpir** la corriente que los atraviesa, cuando se desean enviar a interdicción.

Trabajando en **corriente continua**, su uso resulta bastante problemático; no obstante, los SCR soportan corrientes y tensiones elevadas, y son buenos conmutadores.

Por lo tanto se han desarrollado los GTO (Gate Turn-off Thrystor: tiristor de desactivación de puerta): son SCR en los que la corriente puede interrumpirse con un simple **impulso negativo** en la puerta.

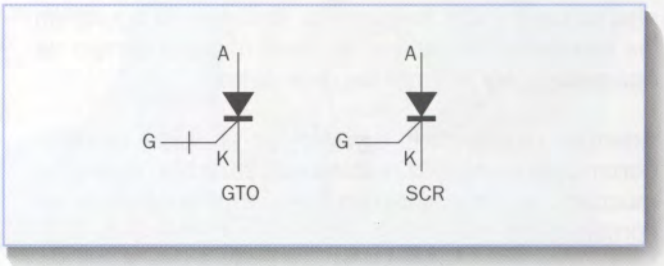

El símbolo de un GTO es muy parecido al de su primo que no puede interrumpirse, el SCR.

CONTROL DE PUERTA

A causa de los fenómenos especiales internos al dispositivo, el circuito que controla la puerta (es decir, el controlador) debe **respetar** una serie de características exigidas por el constructor.

Por ejemplo, la **velocidad de variación** de la corriente negativa de apagado (di/dt) no debe ser ni demasiado baja ni demasiado elevada; es necesario también tener en cuenta inductancias parásitas.

Además, el apagado requiere un cierto **tiempo** con una "cola" relativamente larga, al término de la cual conviene mantener una cierta corriente negativa en la puerta.

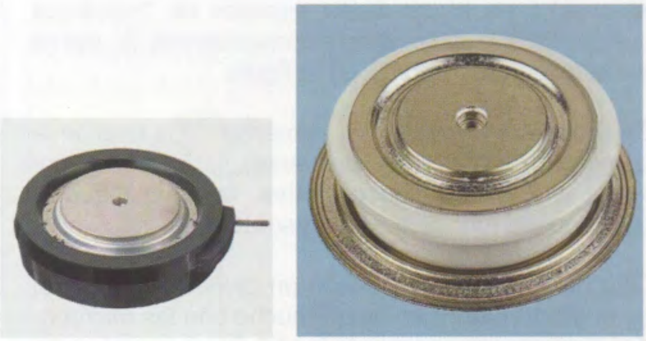

Los GTO se encuentran principalmente en los encapsulados habituales para semiconductores de elevada potencia.

USOS DE LOS GTO

La principal ventaja de estos dispositivos reside en la capacidad de soportar tensiones y corrientes muy **elevadas**, por ejemplo 4.000 V y 2.000 A.

Resultan, por tanto, especialmente útiles en situaciones en las que están en juego potencias bastante elevadas, como en el caso de la **tracción eléctrica** (especialmente en el campo ferroviario).

Para aplicaciones con potencias inferiores se prefiere generalmente emplear los **IGBT**, más manejables y veloces, cuyo control es notablemente más sencillo, especialmente mediante circuitos digitales.

Los GTO encuentran aplicación en situaciones en las que están en juego cientos o miles de KW, y no se requieren conmutaciones muy rápidas.

Diodos curiosos

Numerosos dispositivos "extraños" tienen parentesco con el diodo común

Los **diodos PIN** tienen una región I ("intrinsic": intrínseca) entre las regiones P y N, que produce numerosos portadores de ambas polaridades, ocasionando un cierto retardo en el "apagado" del diodo.

Esto hace posible el uso para conmutar tensiones alternadas de **alta frecuencia**, como en la figura: en las semiondas negativas, el diodo no tiene tiempo de apagarse y por lo tanto las deja pasar.

Además, regulando la corriente en el diodo, se comportan casi como una **resistencia variable**: el circuito mostrado es, por tanto, un atenuador controlado en corriente.

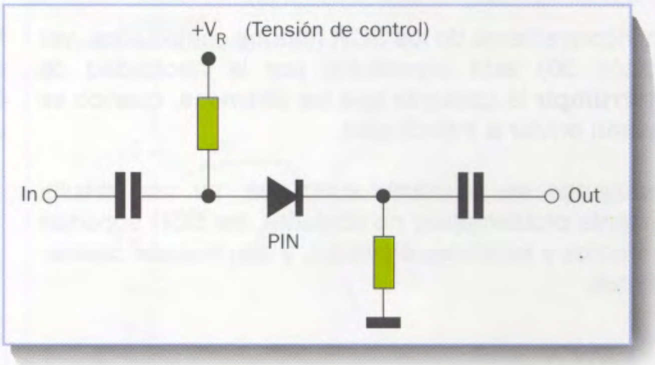

Atenuador para altas frecuencias, realizado con un diodo PIN: la corriente en el diodo regula la atenuación de la señal.

DIODOS TUNNEL Y GUNN

El diodo tunnel (por el que Leo Esaki recibió el Nobel) aprovecha un efecto físico especial de "mecánica cuántica", que tiene como consecuencia la **curva característica** mostrada en la figura.

En una determinada zona, aumentando la tensión la corriente disminuye (y viceversa), y se tiene, por tanto, una **resistencia negativa**, que hace que el diodo pueda funcionar como oscilador.

Otro diodo oscilador, basado en diversos principios, es el **diodo Gunn**: se utiliza mucho con las microondas (incluso por encima de los 50 GHz), por ejemplo en los radares antirrobo.

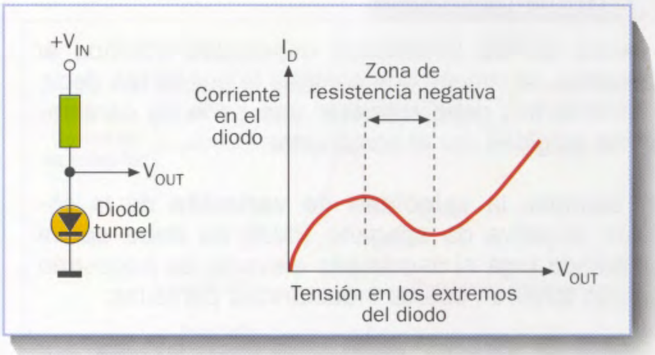

Característica de un diodo tunnel; observe la zona de resistencia negativa que les permite funcionar como oscilador.

DIAC

Los **diac**, o "diodos trigger", se comportan más o menos como los triac (ver lección 30), pero privados del electrodo de control, es decir, la puerta.

Cuando la tensión en sus extremos alcanza un cierto valor, positivo o negativo, empiezan **bruscamente** a conducir; para apagarlos es necesario que la corriente descienda por debajo de un mínimo.

Hace algún tiempo muy utilizados para generar un impulso neto para el control de SCR y triac, las actuales técnicas circuitales y características de los componentes los han dejado casi por completo **obsoletos**.

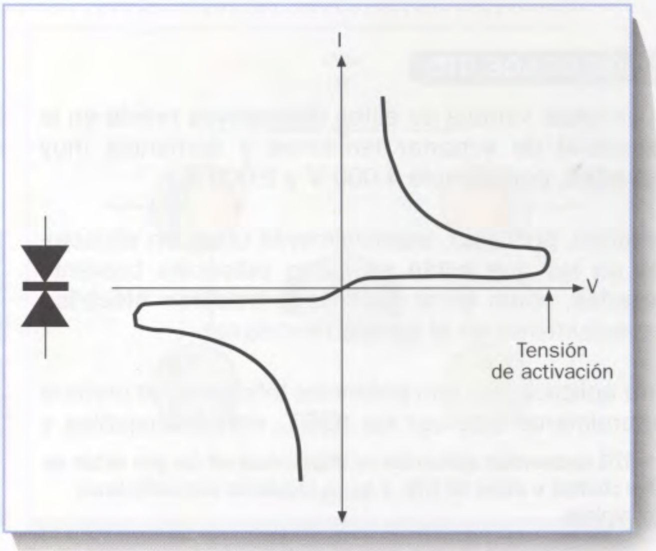

Símbolo del diac y su característica: cuando se alcanza la tensión de activación, conduce de golpe.

Circuitos integrados analógicos

Realizar un circuito entero en un solo chip de silicio ofrece numerosas ventajas

Si el paso de los tubos de vacío (válvulas) a los transistores ha hecho posible la difusión de la electrónica, los circuitos integrados la han **acelerado** a niveles que antes eran impensables.

Esto no ha ocurrido sólo en el campo digital, sino paralelamente también en las aplicaciones **analógicas**, aunque quizá de un modo menos vistoso.

De hecho, mientras en los circuitos digitales aumentaba, sobre todo, la cantidad de puertas lógicas integradas, en las analógicas se refinaba la **calidad**, es decir, las características ofrecidas por los propios circuitos.

CAJA DE MONTAJE

Una de las ventajas ofrecidas por la integración consiste en la posibilidad de disponer de **bloques funcionales** ya preparados, que se pueden combinar para realizar circuitos más complejos.

No siempre tienen características superiores a un circuito equivalente realizado con elementos **discretos** (es decir, separados), como transistores, resistencias o condensadores.

Sin embargo, ofrecen un **coste** notablemente inferior y también, para dimensiones bastante reducidas, menor consumo y mejor fiabilidad y uniformidad de producción.

Los circuitos integrados analógicos han permitido una notable reducción de precios en la electrónica de consumo.

ETAPAS HISTÓRICAS

El primer circuito integrado fue realizado por **Jack S. Kilby** en los laboratorios de Texas Instruments, y presentado oficialmente en 1959.

Alcanzaron gran popularidad los primeros **amplificadores operacionales** (introducidos en la próxima lección) como el µA709, realizado por Bob Widlar para Fairchild, y el posterior µA741.

Pero se necesitó bastante tiempo para que los integrados analógicos lograran afianzarse en el umbral de las **altas frecuencias**, donde hoy trabajan en colaboración con sus primos digitales.

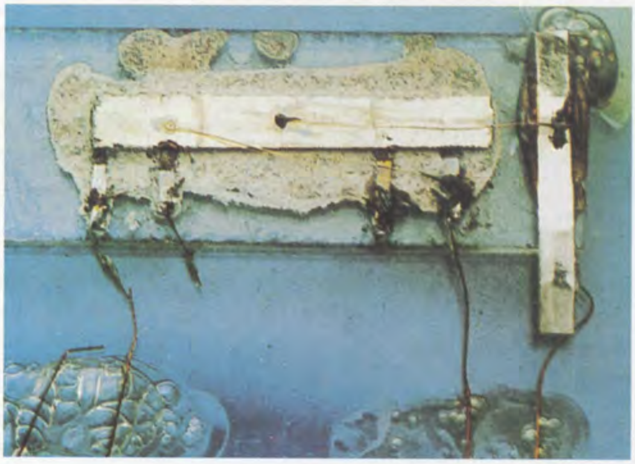

El primer circuito integrado era, para los estándares actuales, bastante voluminoso, pero inició una verdadera revolución.

Componentes pasivos

Un integrado analógico no contiene únicamente diodos y transistores

La producción de un circuito integrado (analógico o digital), es similar a la de los componentes **semiconductores** descritos en la lección 9, pero requiere más etapas.

De hecho, es necesario producir no sólo diodos y transistores, sino también **componentes pasivos** como condensadores, resistencias y, en cualquier caso, inductores.

Además, es necesario **conectar** entre ellos sus componentes, de modo que se construya el circuito completo, un poco como ocurre en las placas de circuito impreso.

Las conexiones ***en un integrado son análogas a las pistas de un circuito impreso, pero miles de veces más pequeñas.***

RESISTENCIAS INTEGRADAS

Como el silicio es un **semiconductor**, es decir, presenta una cierta resistencia al paso de la corriente, se puede utilizar directamente como resistencia una banda de este material.

La longitud, anchura, grosor y estado del estrato de silicio, determinarán el **valor** de resistencia del componente así obtenido.

El **aislamiento** con respecto al resto del circuito se obtiene interponiendo un estrato de silicio con estado opuesto, como en la figura (se crean, sin embargo, diodos parásitos inevitables).

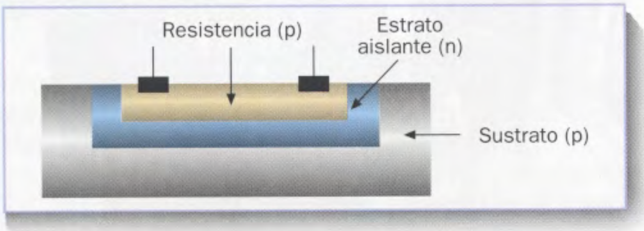

Una resistencia integrada, ***no es otra cosa que una banda de silicio aislada del resto del chip con dos uniones contrapuestas.***

CONDENSADORES INTEGRADOS

Pueden construirse pequeños condensadores utilizando la **capacidad de unión** de un diodo polarizado inversamente (formado por las cargas de los lados de la propia unión).

Hay diversas soluciones alternativas, más o menos **costosas** desde el punto de vista de trabajo, como por ejemplo el uso de un sutil estrato de pentóxido de tantalio entre dos estratos de silicio policristalino.

En todo caso, cuanto más elevada es la **capacidad**, más amplia es el área ocupada; como el espacio es valioso, en general se tratan de evitar valores superiores a algunas centenas de pF.

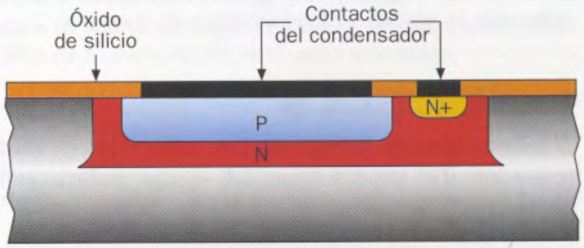

a) Condensador con unión

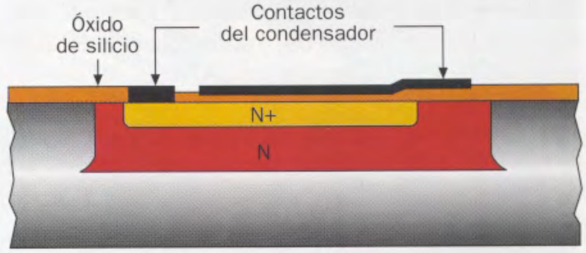

b) Condensador con óxido y metal

Dos ejemplos de **condensadores integrados*****: el segundo utiliza el óxido de silicio como dieléctrico entre las armaduras.***

Fotolitografía

Componentes y conexiones se realizan en fases sucesivas de trabajo

Para la realización de un circuito integrado se emplean técnicas **fotoquímicas**, más o menos similares a las empleadas para los circuitos impresos (ver lección 27 de Herramientas).

Una vez reproducida fotográficamente la imagen original (master) y desarrollada la fotoresistencia, una parte del chip resulta **enmascarada**, es decir, protegida.

La parte restante es sobre la que se desea trabajar, por ejemplo utilizando adecuadas sustancias de forma gaseosa para modificar la **distribución** del silicio, difundiendo impurezas en las partes expuestas.

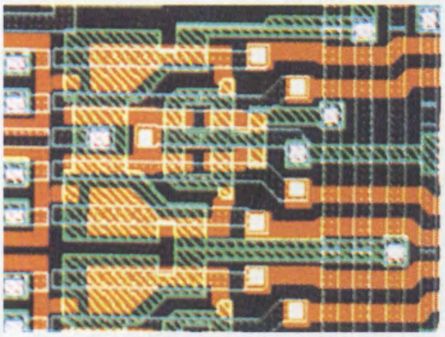

Las máscaras para los circuitos integrados se dibujan con ayuda de un programa adecuado, como los circuitos impresos.

CONEXIONES Y COSTE

Con la misma técnica es posible **depositar** materiales varios, por ejemplo un estrato aislante de óxido de silicio o un sutil conductor de aluminio (o, más recientemente, cobre).

En este último caso se formarán **conexiones eléctricas** entre distintos puntos del circuito, por ejemplo entre una resistencia y un transistor, como en un circuito impreso.

Con el incremento del **número** de operaciones fotoquímicas requeridas para la realización, aumentan el coste de proceso y la probabilidad de introducir defectos.

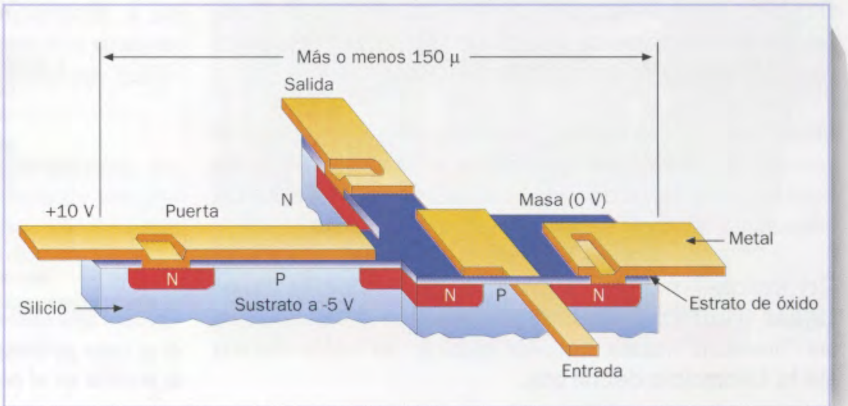

Este detalle de un integrado requiere tres etapas de trabajo: una para las zonas N, otra para el óxido aislante y otra para el aluminio.

FASES FINALES

En una oblea de silicio se realizan juntos varios chips, que después se **prueban**, separados y fijados a la mitad inferior del encapsulado plástico o cerámico.

Para conectar al mundo exterior el circuito integrado, algunas de sus pistas terminan en **áreas metálicas** de superficie relativamente amplia.

En estas áreas se produce la soldadura de los **cables de conexión**, que van desde el chip hasta los pines del encapsulado, que al final se sella, obteniéndose el componente terminado.

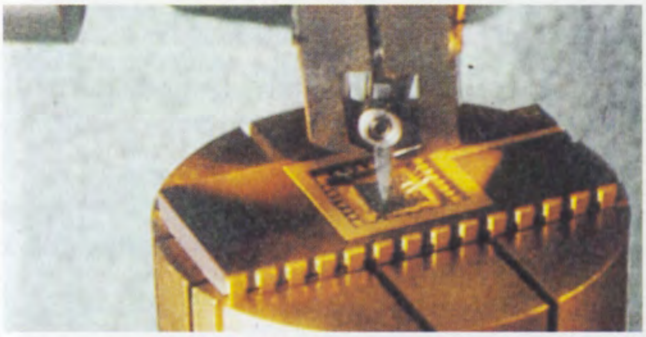

Soldadura de los cables de conexión, de los pines externos a las áreas especialmente dispuestas en el chip.

Problemas de los integrados

Los circuitos analógicos también sufren las tolerancias en los valores de los componentes

La **tasa de producción**, de la que depende buena parte del coste de fabricación, se ve influida por la complejidad del circuito y por el número de etapas requeridas.

Basta un **solo defecto** para descartar un chip entero, ya se trate de un defecto del material en ese punto de la oblea, o de una imperfección durante las fases de proceso posteriores.

Ocurre a menudo que, en las primeras fases de producción de un circuito innovador, la tasa es muy baja (incluso por debajo del 10%), pero está **compensada** por el elevado precio de venta.

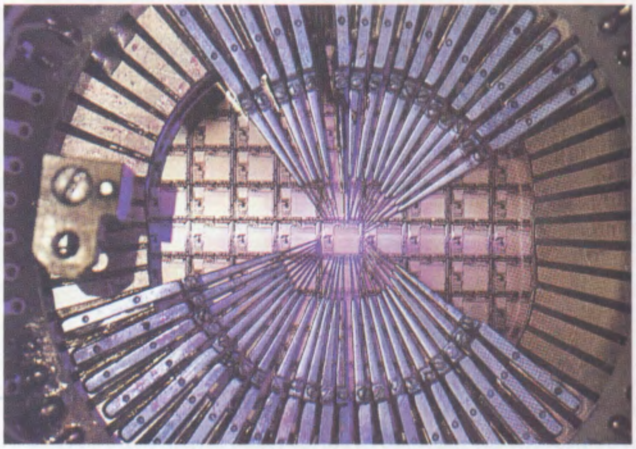

Prueba automática ***de uno de los numerosos chips realizados actualmente en una única oblea de silicio.***

PRECISIÓN DE LAS RESISTENCIAS

En un circuito integrado es difícil obtener resistencias de valor **determinado** con precisión: ello no provoca en general problemas en los circuitos digitales, pero resulta importante para los analógicos.

Una posible solución consiste en proyectar el circuito de modo que no importe el valor absoluto de las resistencias, sino sólo la **relación** entre pares de ellas, que es mucho más estable (ver figura).

En casos límite se recurre a la regulación láser (**laser trimming**) en fase de prueba: la resistencia es "limitada" hasta hacerle asumir un valor dentro de la tolerancia declarada.

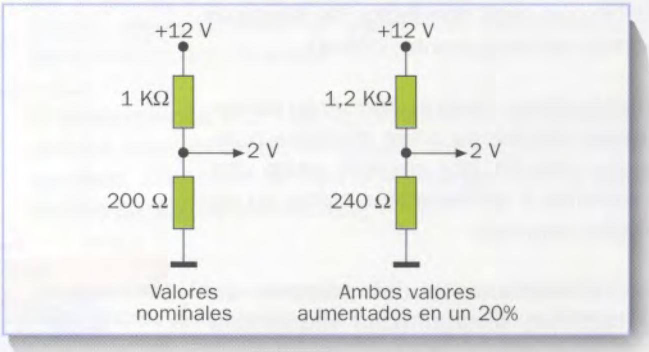

Si el valor de **ambas** ***resistencias varía en el mismo porcentaje, la tensión en el punto medio permanece constante.***

VERSIONES SELECCIONADAS

Para obtener circuitos con características mejor definidas, es decir, especificadas con **tolerancias** más ajustadas, no es siempre necesario incrementar costos para mejorar los procesos de producción.

Se pueden además **seleccionar**, dentro de la producción normal, los componentes que tengan las características requeridas y marcarlos con una sigla distinta.

Cuando se requiere una alta fiabilidad, los integrados son sometidos también a un proceso de envejecimiento previo (**burn-in**), seguido de una prueba posterior.

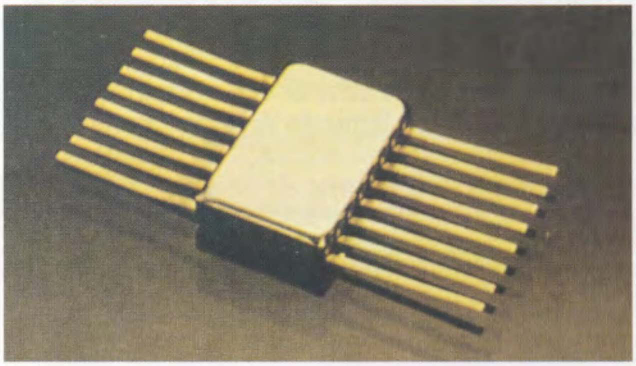

Los encapsulados cerámicos ***ofrecen mejores características térmicas y de duración, pero normalmente resultan más caros.***

Amplificadores operacionales

Son los ladrillos fundamentales de la electrónica analógica moderna

Realizar circuitos con **componentes sencillos** activos y pasivos (transistores, resistencias, condensadores, etc.) ofrece la máxima flexibilidad, pero no es la solución más práctica y conveniente.

Es necesario que el proyecto tenga en cuenta con precisión las **tolerancias** y los límites de cada componente sencillo, y de sus consecuencias en el comportamiento del circuito terminado.

Además, cada componente implica un **costo** y una ocupación de espacio, además de representar una posible fuente de funcionamiento erróneo en el caso de que se estropee: cuantos menos haya, mejor.

UN AMPLIFICADOR IDEAL

Como en la electrónica digital, la revolución en el campo analógico estuvo representada por la llegada de los **circuitos integrados**, que contienen un circuito completo en un chip de pocos milímetros cuadrados.

El integrado analógico fundamental es el **amplificador operacional** (operational amplifier u "op-amp"), un amplificador diferencial con una altísima ganancia y una también alta impedancia de entrada.

Se trata de un **componente** aparentemente sencillo que, como las puertas lógicas en el campo digital, permite realizar una gran variedad de circuitos de todo tipo.

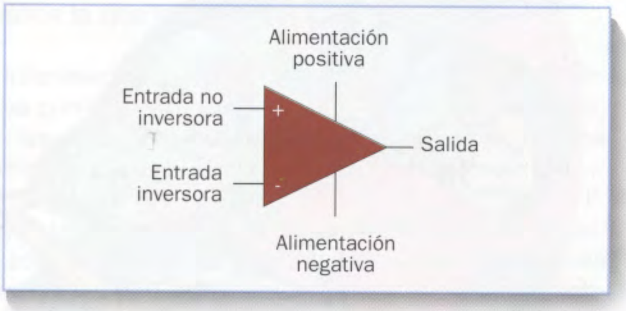

Símbolo de un amplificador operacional: amplifica la diferencia entre las tensiones en las dos entradas.

POLARIDAD DE LAS ENTRADAS

Un amplificador operacional tiene dos entradas: la marcada "+" es no **inversora**, es decir, la tensión a la salida tiene la misma fase de la aplicada a la entrada.

La marcada "-" es, sin embargo, **inversora**: la salida está invertida, es decir, desfasada en 180 grados, como en el amplificador común de un transistor ilustrado en la lección 11.

Sin embargo, un operacional (como se le llama familiarmente) no se utiliza directamente como en la figura, sino que se tienen que añadir algunos **componentes externos** para lograr que se comporte del modo deseado.

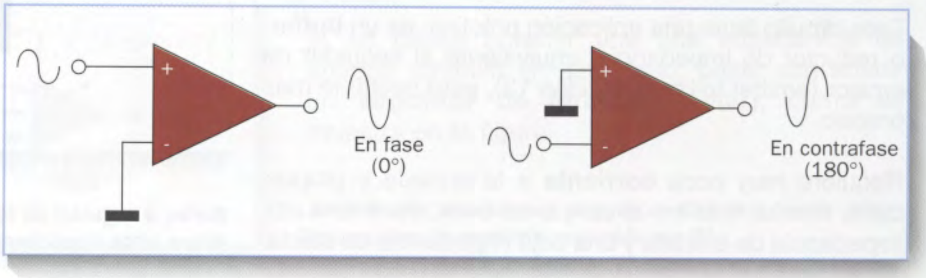

La entrada positiva no invierte la señal, pero la negativa sí la invierte; se amplifica la diferencia entre las dos.

Operacionales y realimentación

El comportamiento de un op-amp está determinado por la realimentación

Un amplificador operacional tiene una **ganancia** muy elevada, incluso de un millón de veces: no se aprovecha directamente, sino que se controla con la realimentación.

La figura muestra un amplificador no inversor con ganancia 100, obtenido con una **realimentación negativa** (ver lección 19), que lleva a la entrada una centésima de la señal a la salida.

Como las **ventajas** de la realimentación son más consistentes cuanto mayor es la ganancia del amplificador utilizado, con un operacional se obtienen características excelentes.

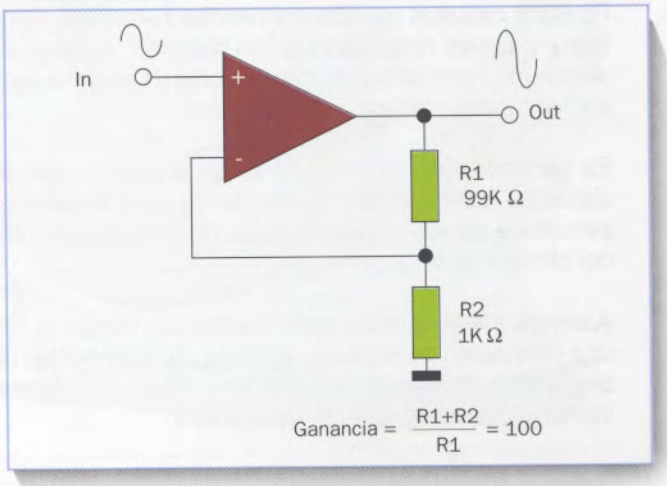

Llevando a la entrada negativa una centésima de la salida, se obtiene una ganancia 100, estable y precisa.

SALIDA GUIADA

Para entender bien el funcionamiento del amplificador no inversor, es necesario tener en cuenta que el operacional amplifica la **diferencia** entre sus dos entradas.

Por ejemplo, un aumento de tensión a la salida hará incrementar la tensión en la entrada inversora: el operacional moverá, por tanto, la tensión a la salida **en sentido opuesto**, bajándola.

En la práctica, en el caso de realimentación negativa (es decir, un retorno de la salida a la entrada inversora), el dispositivo actuará para mantener siempre las dos entradas **a la misma tensión**.

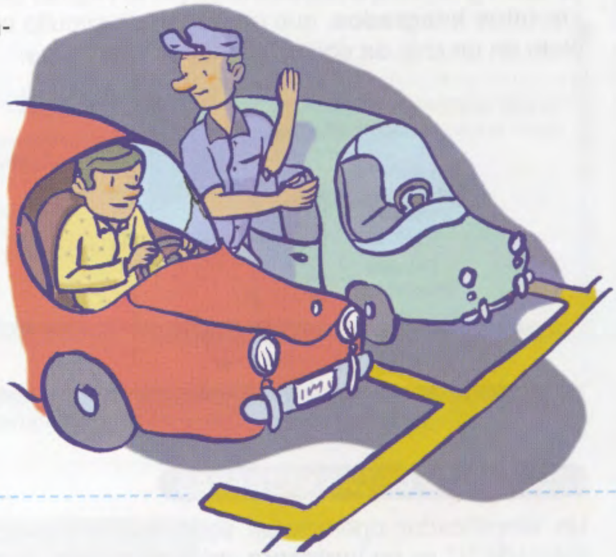

El operacional corrige continuamente, de modo que se mantenga la salida en el valor deseado.

SEGUIDOR DE TENSIÓN

En el caso límite de una realimentación del 100%, como el mostrado en la figura, la salida **sigue exactamente** la tensión en la entrada positiva: la ganancia es 1.

Este circuito tiene una aplicación práctica: es un **buffer**, o reductor de impedancia, equivalente al seguidor de emisor (emitter follower, lección 13), pero bastante más preciso.

Requiere muy poca **corriente** a la entrada y proporciona mucha más en la carga, es decir, tiene una alta impedancia de entrada y una baja impedancia de salida.

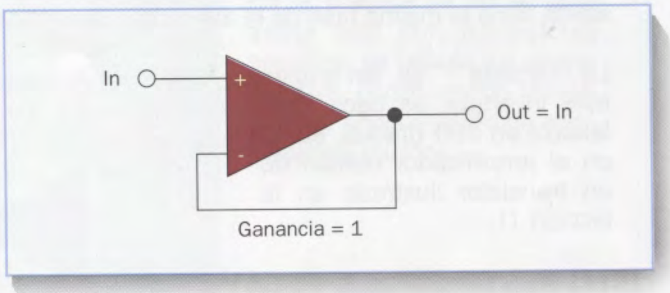

Buffer,* o seguidor de tensión *realizado con un operacional; no sirven otros componentes.

Amplificador inversor

Conectando a masa la entrada no inversora, se tiene un efecto interesante

La figura muestra un amplificador inversor, realizado con un operacional: se utilizan siempre **dos resistencias**, pero dispuestas de otro modo con respecto al circuito de la página anterior.

También en este caso tenemos una **realimentación negativa**, ya que la salida vuelve a la entrada no inversora (negativa) mediante R2.

El op-amp tratará, por tanto, de mantener las dos entradas al mismo nivel: como la entrada no inversora (positiva) está **a masa**, también en la negativa habrá cero voltios.

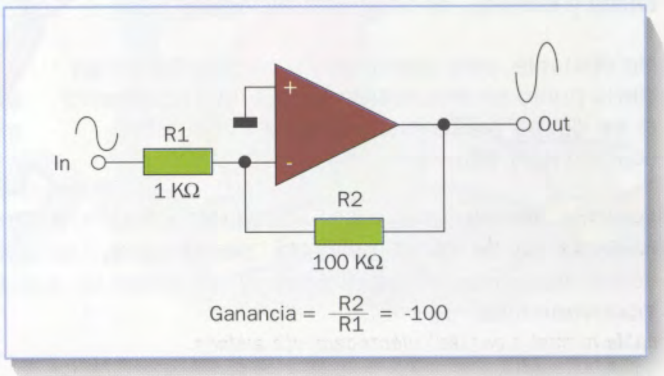

$$\text{Ganancia} = \frac{R2}{R1} = -100$$

Amplificador inversor *realizado con un operacional: cuando la entrada baja, la salida sube.*

MASA VIRTUAL

En la práctica, la entrada negativa se comporta como una "masa virtual", es decir, como si fuera un punto **conectado a masa**, aunque en realidad no lo es: es la salida la que la mantiene.

Es interesante observar que, como la entrada no absorbe corriente (suponiendo un operacional perfecto), por las dos resistencias R1 y R2 pasa **la misma corriente**.

Según la ley de Ohm, las tensiones de entrada y salida son, por tanto, proporcionales a las respectivas resistencias: la amplificación depende exclusivamente de la **relación** entre las dos.

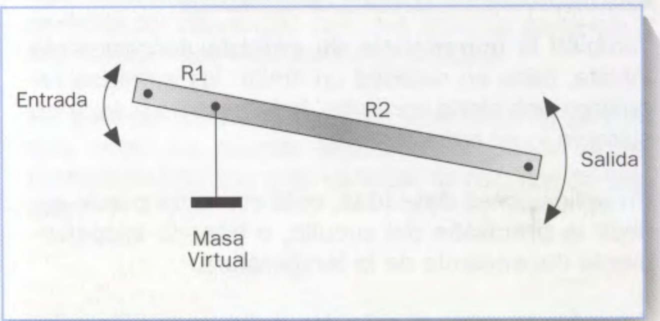

Un amplificador de masa virtual se comporta como una palanca, cuyo punto de apoyo es mantenido fijo por el operacional.

IMPEDANCIA DE ENTRADA

El amplificador no inversor (ver página anterior) tiene una **impedancia de entrada** muy alta, idealmente infinita, es decir, que no molesta a la señal aplicada.

El amplificador inversor tiene una impedancia de entrada exactamente **similar a R1**; la señal ve esta resistencia, dado que el otro extremo de R1 es como si estuviera conectado a masa.

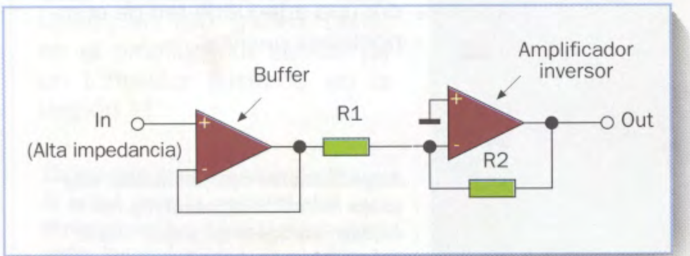

Si se desea obtener una alta impedancia de entrada también en este caso, basta con añadir un **seguidor de tensión** (buffer), como se muestra en la figura.

Amplificador inversor con entrada de alta impedancia: IC1 proporciona la corriente requerida por R1.

Límites y aplicaciones

Aunque no son perfectos, los operacionales son de enorme utilidad

La elevada **ganancia** de los op-amp permite aprovechar al máximo la realimentación, obteniendo una precisa correspondencia entre salida y entrada, es decir, una distorsión baja.

No obstante, esta ganancia no es **infinita**; en un cierto punto se encuentran límites, especialmente si se desea realizar circuitos con una amplificación muy alta.

Además, esto se produce al aumentar la **frecuencia** de la señal aplicada; ello implica, como veremos, otros límites y posibles inconvenientes.

Los operacionales no serán **perfectos**, *pero resultan muy cómodos y versátiles.*

CORRIENTE DE ENTRADA Y OFFSET

También la **impedancia de entrada**, teóricamente infinita, tiene en realidad un límite: las entradas requieren una cierta corriente, baja pero no nula, y no siempre igual entre las dos.

En aplicaciones delicadas, esta corriente puede reducir la **precisión** del circuito, o hacerlo excesivamente dependiente de la temperatura.

Además, es como si entre las entradas hubiera una pequeña **diferencia de tensión**, llamada "tensión de offset", que se traduce en un error a la salida (a veces se puede compensar con un trimmer resistivo).

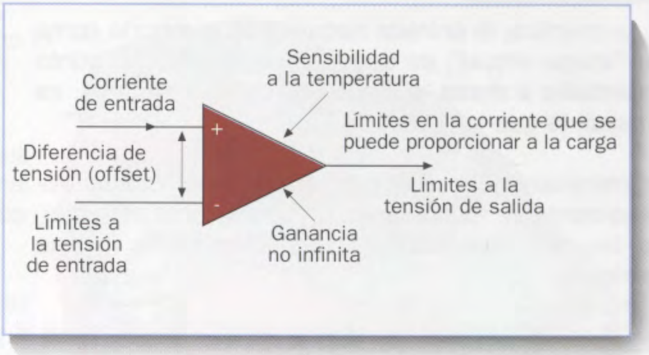
Corriente de entrada
Sensibilidad a la temperatura
Límites en la corriente que se puede proporcionar a la carga
Diferencia de tensión (offset)
+
-
Límites a la tensión de salida
Límites a la tensión de entrada
Ganancia no infinita

Un op-amp **real** *se aleja de su modelo ideal, pero no obstante es bastante menos imperfecto que un transistor.*

USOS DE LOS OPERACIONALES

Hasta ahora hemos visto algunos usos de los op-amp, como **amplificadores** de corriente continua; hay otros muchos, a partir naturalmente de la amplificación de señales alternas (por ejemplo, audio).

Pero estos **versátiles** dispositivos se prestan a innumerables aplicaciones más: generación de señales, filtros activos, producción de corriente constante, control de alimentadores, etc.

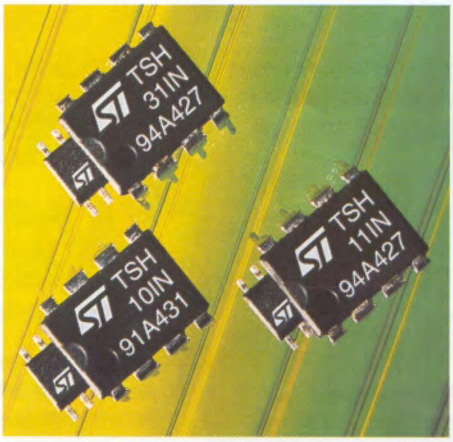
ST TSH 31IN 94A427
ST TSH 10IN 91A431
ST TSH 11IN 94A427

En la práctica ya no se **proyecta** casi con transistores sencillos: se utilizan los operacionales, construyendo alrededor una adecuada red de componentes pasivos.

Amplificadores operacionales: *con estos sencillos dispositivos, en el campo analógico se puede hacer prácticamente de todo.*

Operacionales en alterna

Los op-amp no requieren la complicada polarización de los transistores

Los circuitos tratados en la lección anterior (amplificadores inversores y no inversores) están **acoplados en continua**, es decir, conectando directamente un operacional al siguiente.

Eso puede ser útil en ciertos casos, pero no resulta deseable si se desea amplificar una **señal de audio**, o en cualquier caso una tensión alterna superpuesta a una tensión continua fija.

De hecho, **ambas** tensiones se verían amplificadas; con una ganancia elevada se tendría una saturación, es decir, la salida "chocaría" con la alimentación o la masa.

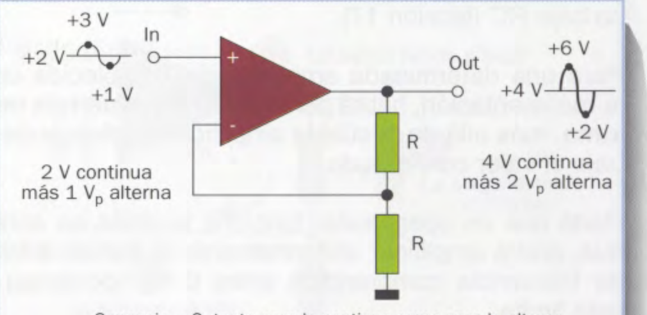
+3 V
+2 V
+1 V
In
Out
+6 V
+4 V
+2 V
R
R
2 V continua
más 1 V_p alterna
4 V continua
más 2 V_p alterna

Ganancia = 2, tanto para la continua como para la alterna

Un operacional **acoplado en continua** ***amplifica de manera indiferente tensiones continuas y alternas.***

SEPARAR LOS DOS CASOS

La tensión continua de entrada puede separarse fácilmente de la señal: basta con utilizar un **condensador**, como ya hemos visto en la lección 5 y en los amplificadores con BJT (por ejemplo en el lección 11).

Incluso con cero voltios continuos a la entrada, no obstante, la amplificación de la operación es molesta: de hecho se amplifica la **tensión de offset** (aunque sea mínima) del mismo dispositivo.

La figura muestra, por ejemplo, como una tensión de offset de 10 mV puede causar más de 10 V **no deseados** a la salida (se querrían 0 V), aun teniendo la entrada a masa.

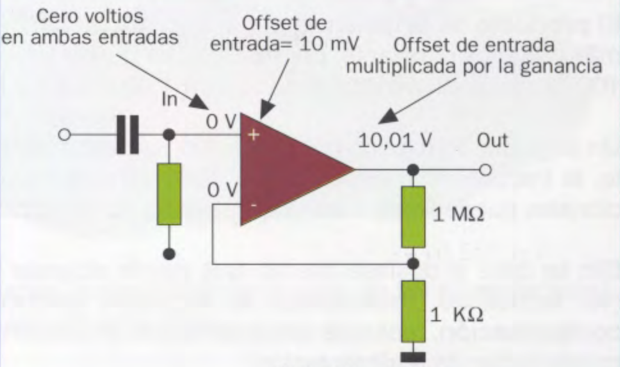
Cero voltios
en ambas entradas
Offset de
entrada= 10 mV
Offset de entrada
multiplicada por la ganancia
In
0 V
0 V
10,01 V
Out
1 MΩ
1 KΩ

Ganancia = 1001

Si la ganancia del circuito es alta, una pequeña tensión de **offset** ***desplaza notablemente la salida.***

DOS REALIMENTACIONES DISTINTAS

El problema se resuelve modificando la realimentación, de modo que el amplificador tenga **ganancia 1** en continua (es decir, no amplifique) y la ganancia deseada en alterna.

El amplificador de la figura se comporta como el anterior en lo que se refiere a las tensiones alternas, mientras que en continua el condensador **no existe**: puede verse como un cable interrumpido.

En consecuencia, se convierte en un sencillo **seguidor de tensión** (ver lección 33) para las tensiones continuas: si hay 10 mV de offset, habrá sólo 10 mV a la salida, más que aceptables.

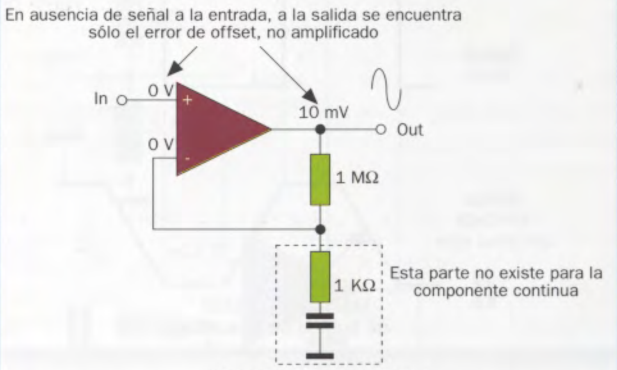
En ausencia de señal a la entrada, a la salida se encuentra
sólo el error de offset, no amplificado
In
0 V
0 V
10 mV
Out
1 MΩ
1 KΩ
Esta parte no existe para la
componente continua

Ganancia = 1 en continua
Ganancia = 1001 en alterna

Este amplificador tiene una ganancia de 1.000 **para la señal,** ***pero una ganancia de 1 para la tensión continua.***

Anchura de banda

Las características de un op-amp empeoran a frecuencias elevadas

A medida que la frecuencia de la señal aplicada a la entrada sube, la ganancia de un amplificador operacional **disminuye**, como si incorporase un filtro de paso bajo RC (lección 17).

Para una determinada amplificación establecida con la realimentación, habrá por tanto una **frecuencia máxima**, más allá de la cual la amplificación descenderá bajo el valor configurado.

Dado que un operacional funciona también en continua, podrá amplificar uniformemente la **banda** entera de frecuencia comprendida entre 0 Hz (continua) y este límite.

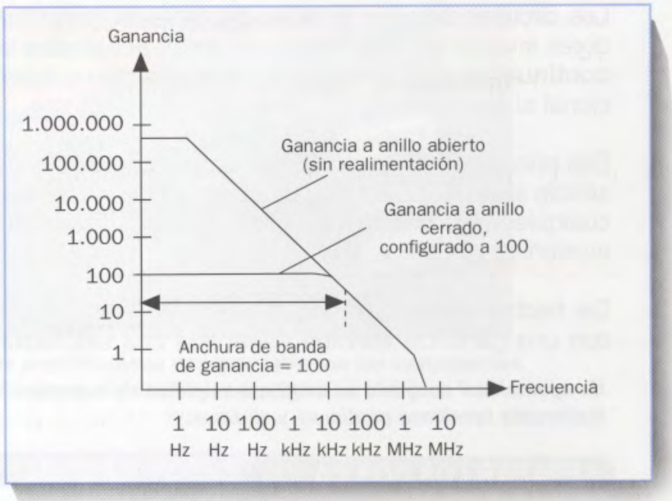

La amplificación configurada con la realimentación no se puede mantener cuando alcanza los límites de frecuencia del dispositivo.

ESTABILIDAD

El **producto** de la ganancia por la anchura de banda (G*BW o GBW) es más o menos constante: por ejemplo, se puede tener una ganancia de 100 hasta 30 KHz o una ganancia de 1.000 hasta 3 KHz.

Un seguidor de tensión (amplificación = 1) alcanza, por tanto, la frecuencia máxima posible, pero con algunos operacionales puede llegar a ser **inestable**, es decir, autooscilar.

Ello se debe al desfase interno, que puede alcanzar los 180° (ver lección 20, osciladores); es necesaria entonces una **compensación**, realizada únicamente con un pequeño condensador de realimentación.

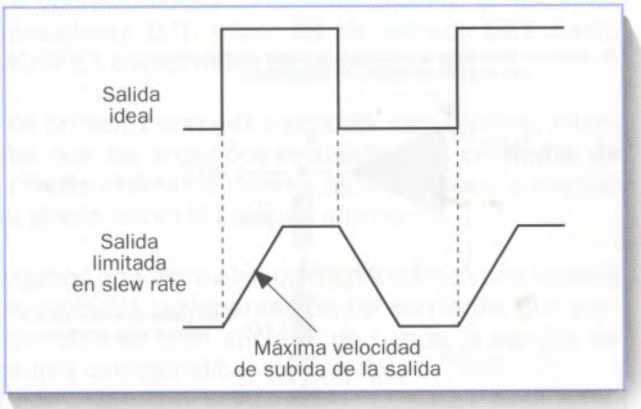
Salida ideal
Salida limitada en slew rate
Máxima velocidad de subida de la salida

El efecto de la* limitación de slew *rate no afecta sólo a las ondas cuadradas, sino a cada señal amplia, de frecuencia bastante elevada.

SLEW RATE

Entre las características de los operacionales ligadas a la degradación progresiva debida al aumento de la frecuencia, se encuentra el "slew rate", es decir, la **velocidad de subida** (y bajada) de la salida.

Se mide normalmente en **V/µs** (voltios por microsegundo) y puede alterar la forma de la señal, como en el ejemplo mostrado en la figura de al lado.

También una sinusoide sufre con este límite, especialmente en la zona que rodea al cero, que es más empinada cuanto mayores son la **frecuencia** y **amplitud** de la propia onda.

Impedancia cero

La salida de un operacional con realimentación tiene características muy buenas

Variando la **carga** aplicada a un operacional, es decir, la petición de corriente, la tensión a la salida tendería a reducirse por efecto de la resistencia interna.

La realimentación negativa, sin embargo, se da cuenta del error y hace que la salida "ponga más empeño" en mantenerse estable: en otros términos, se tiene una **baja impedancia** de salida (idealmente nula).

El beneficio es más marcado cuanto más alta es la ganancia en vacío, es decir, **a anillo abierto**, con respecto a la configurada con la realimentación.

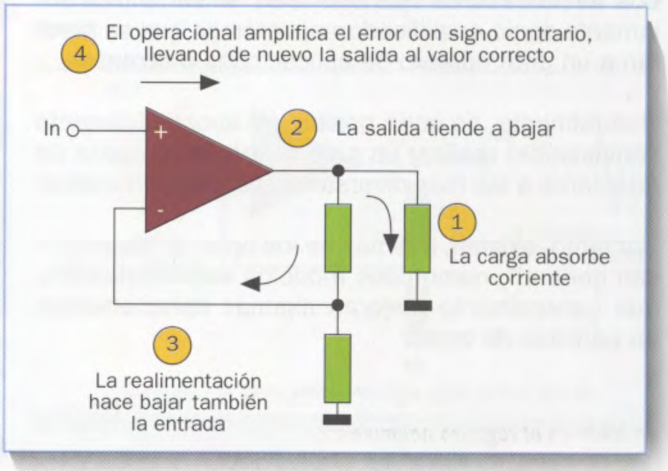
④ El operacional amplifica el error con signo contrario, llevando de nuevo la salida al valor correcto
In
+
-
② La salida tiende a bajar
① La carga absorbe corriente
③ La realimentación hace bajar también la entrada

La realimentación actúa de modo que el operacional* compense *las variaciones de carga, disminuyendo la impedancia de salida.

LÍMITES DE LA SALIDA

Como la impedancia de salida puede aproximarse al valor ideal de cero ohmios, la **corriente máxima** depende en cualquier caso de las características de los componentes internos del integrado.

Además, cuanto más se **aproxime** la salida a las alimentaciones (positiva o negativa), más limitada es la corriente que puede proporcionar a una carga conectada hacia la propia alimentación.

No obstante, para tensiones **alternas** cuyos picos se mantienen razonablemente lejanos de las alimentaciones, los operacionales constituyen una buena aproximación a un amplificador ideal.

ALIMENTACIÓN SENCILLA

Los amplificadores operacionales requieren normalmente **dos alimentaciones**: una positiva y otra negativa con respecto a masa, por ejemplo +12 V y –12 V.

Sin embargo, si se tienen que tratar sólo señales alternas, la "masa" hace simplemente la función de **referencia** de polarización; en la práctica, es un valor a medio camino entre la alimentación positiva y la negativa.

Por lo tanto, se puede utilizar **una sola** tensión de alimentación, y extraer un valor intermedio con un sencillo divisor, como en la figura: el condensador C hace que resulte a masa para la señal.

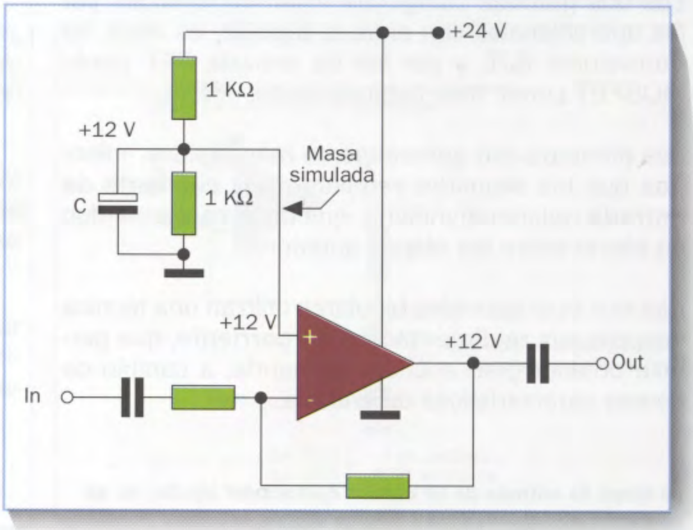
+24 V
1 KΩ
+12 V
Masa simulada
C
1 KΩ
+12 V
+
-
+12 V
In
Out

Amplificador que emplea una sola tensión de alimentación, con "masa" simulada con un divisor.

Tipos de operacionales

Existen numerosas variantes de estos versátiles componentes

Los amplificadores operacionales no se utilizan solamente como amplificadores; veremos que se prestan a un gran número de aplicaciones **diferentes**.

Naturalmente, no sería posible (ni económicamente conveniente) realizar un **solo dispositivo** capaz de adaptarse a las más diversas exigencias circuitales.

Por tanto, existen, además de los op-amp "de propósito general", numerosos modelos **especializados**, que generalmente mejoran algunas características en perjuicio de otras.

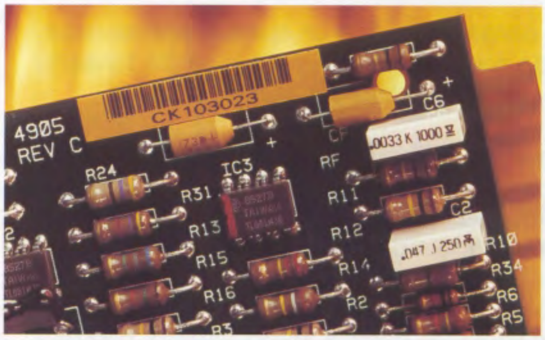

Los amplificadores operacionales son los componentes estándares del diseñador analógico.

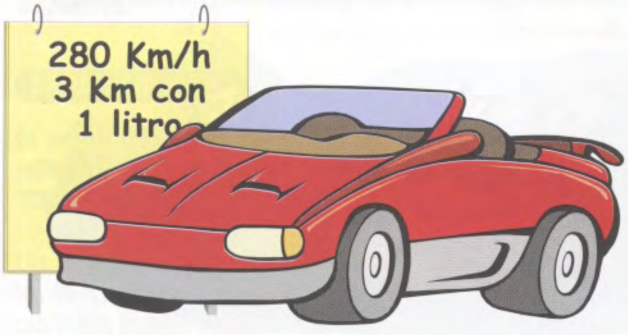

No se puede tener todo: para mejorar una característica normalmente se empeoran otras.

VELOCIDAD Y CONSUMO

Los operacionales **rápidos** son capaces de tratar señales de vídeo, o a frecuencias todavía superiores: la anchura de banda (a ganancia = 1) de algunos modelos es de bastantes centenas de MHz.

En el extremo opuesto encontramos los op-amp "micropower", es decir, de **muy bajo consumo**: son lentos, pero requieren muy poca corriente, cosa que los hace ideales para aplicaciones de batería.

Hay también operacionales "externos", capaces de trabajar a (relativamente) alta tensión, o bien de proporcionar alta **potencia** en la carga.

TIPOS DE ENTRADAS

Las dos grandes categorías están constituidas por los operacionales con entrada **bipolar**, es decir, de transistores BJT, y por los de entrada **FET** (tanto MOSFET como, más habitualmente, JFET).

Los primeros son generalmente más rápidos, mientras que los segundos requieren una **corriente de entrada** realmente mínima, que hace casi superfluo su efecto sobre las etapas anteriores.

Algunos operacionales bipolares utilizan una técnica denominada **realimentación de corriente**, que permite obtener gran anchura de banda, a cambio de peores características de entrada.

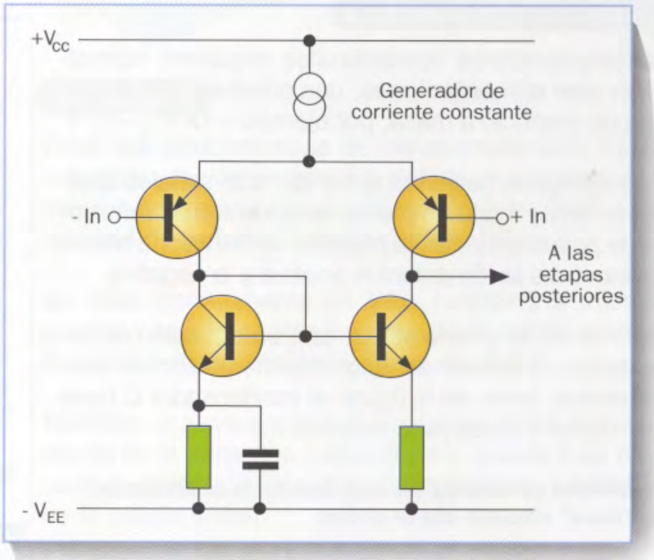

La etapa de entrada de un clásico operacional bipolar, no es otra cosa que un buen amplificador diferencial.

Reguladores de tensión

Permiten proporcionar una alimentación estable al resto del circuito

Piloto Automático

Hemos visto (en la lección 14) cómo se puede realizar un sencillo **alimentador estabilizado**, capaz de producir una tensión independiente de la carga.

Aun utilizando **componentes discretos**, como transistores y diodos Zener, sigue siendo difícil obtener buenas características; además, el circuito resulta bastante pesado.

Como la necesidad de disponer de tensiones estabilizadas es bastante común, existen a la venta numerosos circuitos integrados **especializados** para esta tarea.

SENCILLÍSIMOS DE UTILIZAR

La figura muestra un alimentador estabilizado, realizado con un típico **regulador de tres terminales**: un integrado que tiene simplemente una entrada, una salida y un terminal de masa.

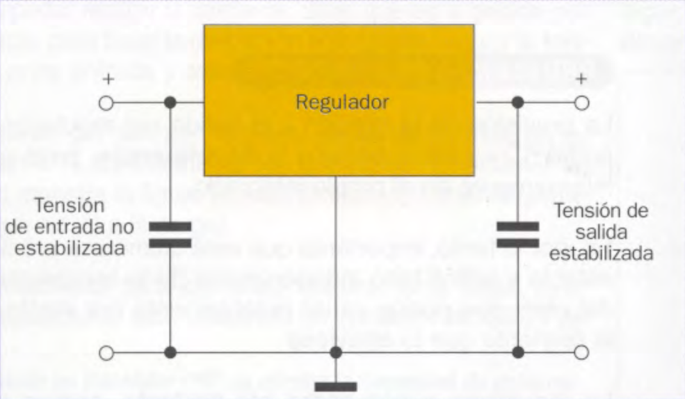
Regulador
+
+
Tensión de entrada no estabilizada
Tensión de salida estabilizada

Este integrado recibe a la entrada una tensión continua no estabilizada, procedente por ejemplo de un puente de diodos, y produce una tensión **regulada**, con buena precisión, al valor declarado por el fabricante del chip.

El circuito debe contrastar dos **causas de error** principales: las variaciones de tensión a la entrada, y las variaciones de absorción por parte de la carga.

Uso habitual de un regulador lineal de tres terminales: el integrado se encarga de estabilizar la tensión a la salida.

DESACOPLAMIENTO

Los dos **condensadores** a la entrada y salida, normalmente de capacidad media-alta (por ejemplo, 100 µF) son importantes para el correcto funcionamiento del circuito.

El primero nivela la entrada y proporciona los **picos de corriente** pedidos por el regulador; el segundo hace lo mismo con la carga, mientras que el integrado interviene. Para asegurar la estabilidad del regulador los picos deben tener **baja impedancia** con respecto a las demás frecuencias; se utilizan normalmente electrolíticos de tantalio, o dos tipos distintos (por ejemplo, electrolítico y cerámico) en paralelo.

Un electrolítico de aluminio o tantalio desacopla las bajas frecuencias, mientras que un condensador cerámico responde bien a las frecuencias altas.

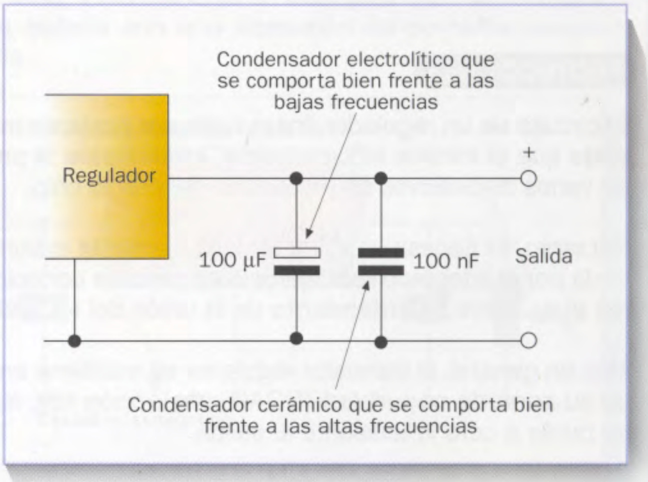
Condensador electrolítico que se comporta bien frente a las bajas frecuencias
Regulador
+
100 µF
100 nF
Salida
Condensador cerámico que se comporta bien frente a las altas frecuencias

Cómo funciona un regulador

A pesar de la simplicidad externa, el circuito es bastante refinado

La figura muestra el esquema simplificado de un regulador de tensión lineal: un **amplificador de error** (por ejemplo un operacional con resistencias adecuadas) compara la tensión de salida con la de referencia.

El amplificador pilota la base del **transistor regulador**, que varía la corriente de salida de modo que se mantengan las dos entradas del operacional a la misma tensión.

Una disminución de la tensión a la salida (por ejemplo para una petición de corriente mayor) causa, por tanto, una mayor **conducción** del transistor, que realza la misma salida al valor correcto.

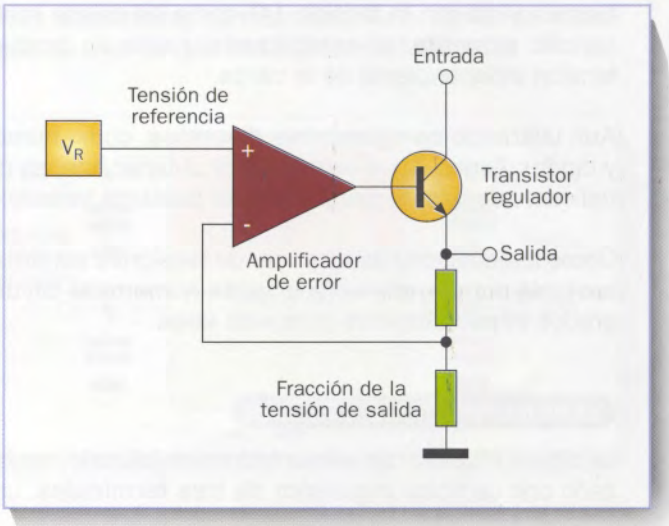

Dentro del integrado hay una referencia de tensión, un amplificador de error y un transistor regulador.

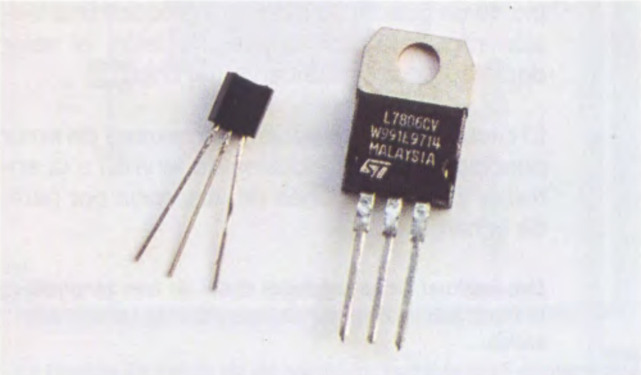

Los reguladores de tres terminales están disponibles para distintos valores de tensión de salida, corriente máxima y potencia disipable.

PROBLEMAS DE PRECISIÓN

La precisión de la tensión a la salida del regulador no podrá nunca ser superior a la de **referencia**, producida internamente en el propio integrado.

Es, por lo tanto, importante que esta última sea precisa, estable y sobre todo independiente de la **temperatura** del chip, que puede variar notablemente por efecto de la corriente que lo atraviesa.

La regulación nunca podrá ser **perfecta**, porque una mínima variación de la salida es, en cualquier caso, necesaria para producir la tensión de corrección en la salida del operacional.

PROTECCIONES

El circuito de un regulador lineal suele ser bastante más **complejo** que el mínimo indispensable, incluso para la presencia de varios dispositivos de protección del propio chip.

Por tanto, es necesario limitar tanto la **corriente** máxima suministrada por el integrado (por ejemplo en caso de cortocircuito), como el excesivo **calentamiento** de la unión del transistor.

Más en general, el transistor regulador se mantiene en el interior de su **zona de seguridad** (SOAR, ver lección 15), aun a costa de poner a cero la tensión a la salida.

Caída de tensión

Los reguladores lineales no rehuyen el problema de la disipación

La capacidad de regulación de la tensión de un integrado especializado es normalmente muy **superior** a la de los circuitos descritos en la lección 14.

No obstante se mantiene inalterado el problema de la disipación en el regulador, que depende de la inmutable **ley de Ohm**: es dada por la tensión en los extremos del integrado por la corriente que lo atraviesa.

Por otra parte, en los extremos del regulador debe existir, en cualquier caso, una cierta **tensión mínima**, únicamente para superar la tensión de saturación del transistor regulador.

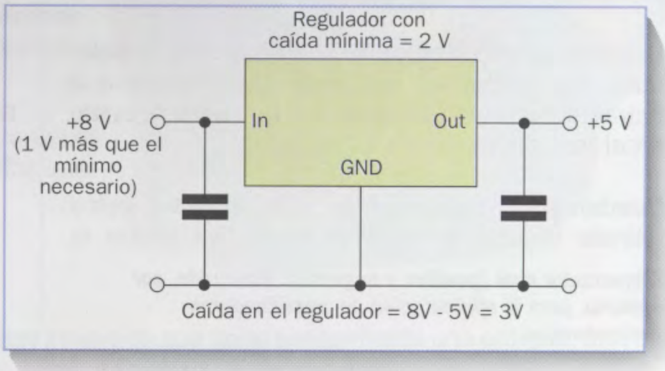
Regulador con caída mínima = 2 V
+8 V (1 V más que el mínimo necesario)
In
Out
GND
+5 V
Caída en el regulador = 8V - 5V = 3V

La tensión a la entrada debe ser siempre más alta que la de salida más la caída mínima, para asegurar una buena regulación.

REGULADORES DE BAJA CAÍDA

Al no poder reducir la corriente, dado que es la pedida por la carga, para bajar la disipación sólo queda reducir la **tensión** entre entrada y salida del regulador.

Se puede, por ejemplo, emplear un transistor **PNP** en lugar de NPN, controlándolo con una corriente hacia masa, como muestra la figura (es, sin embargo, corriente gastada, que no va a la carga).

Especialmente para corrientes relativamente bajas, existen reguladores que, utilizando un transistor **MOSFET** en lugar de un BJT, tienen una caída de tensión de algunas decenas de milivoltios.

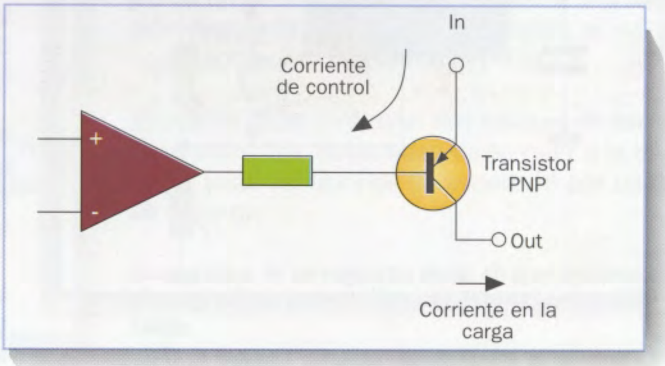
In
Corriente de control
+
-
Transistor PNP
Out
Corriente en la carga

Utilizando un transistor PNP se elimina la necesidad de gobernar la base con un transistor de 0,65 V más alta que la salida.

REGULADORES DE BAJO CONSUMO

En distintos casos, por ejemplo en las aplicaciones que funcionan con pilas o baterías recargables, es importante reducir al mínimo el **gasto** de energía, es decir, la que no llega a la carga.

Además de la caída en el regulador, es necesario también considerar la corriente absorbida para **alimentar** los circuitos internos, como el operacional y la fuente de tensión de referencia.

Algunos reguladores especializados se pueden **deshabilitar**, de modo que no proporcionen tensiones a la salida, y mantenerse en cualquier caso conectados a la batería con una absorción de corriente secundaria.

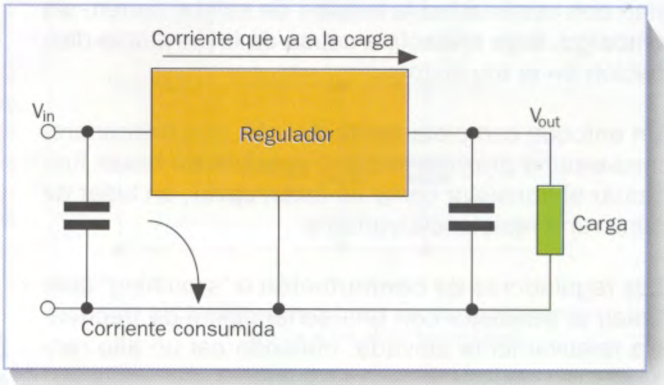
Corriente que va a la carga
V_{in}
Regulador
V_{out}
Carga
Corriente consumida

Si la carga absorbe poco, la corriente de alimentación del regulador puede constituir un importante consumo de energía.

Tipos de reguladores

Los reguladores lineales se producen en numerosas variantes

No todos los circuitos funcionan con tensión de alimentación positiva; existen, por tanto, reguladores con entrada y salida **negativas**, simétricos con respecto a los primeros.

Los reguladores **dobles** producen ambos la polaridad a la salida, y la requieren naturalmente a la entrada (hablamos siempre de las lineales), como en el ejemplo mostrado en la figura.

Pueden ser independientes o **acoplados**: estos últimos regulan ambos las tensiones según la más importante, normalmente la positiva, de modo que se mantengan exactamente simétricas.

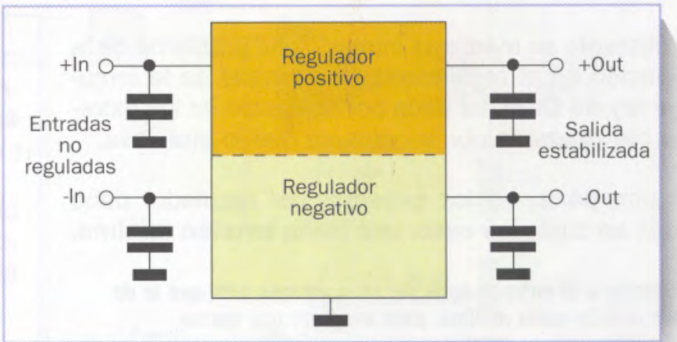

Alimentador dual *(positivo y negativo), adecuado, por ejemplo, para la alimentación de amplificadores operacionales.*

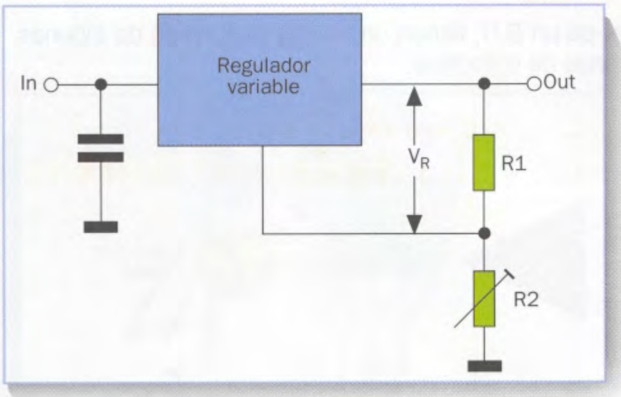

La tensión de salida de un regulador variable está determinada por el valor de dos resistencias (uno puede ser un trimmer).

REGULADORES VARIABLES

Además de los reguladores fijos, disponibles para los valores de tensión más comunes (por ejemplo 3,3 V, 5 V, 12 V), están disponibles reguladores lineales a tensión **regulable** (ajustable).

Permiten tanto obtener tensiones no previstas por los fijos (como 2,46 V o 7,3 V) como ajustar con mayor **precisión** la tensión de salida, utilizando un trimmer resistivo.

La figura muestra una configuración de uso típica de un regulador variable: **compara** la tensión V_R en los extremos de R1 con la interna de referencia.

ADEMÁS EL LINEAL

Los reguladores lineales se llaman así porque varían con continuidad la tensión de salida; tienen, sin embargo, **baja eficacia** a causa de la inevitable disipación en el transistor.

Un enfoque completamente distinto, que examinaremos en una próxima lección, consiste en hacer funcionar el transistor como un **interruptor**, en lugar de cómo una resistencia variable.

Los reguladores **de conmutación** o "switching" controlan el transistor con una señal digital de frecuencia relativamente elevada: obtienen así un alto rendimiento, es decir, bajas pérdidas.

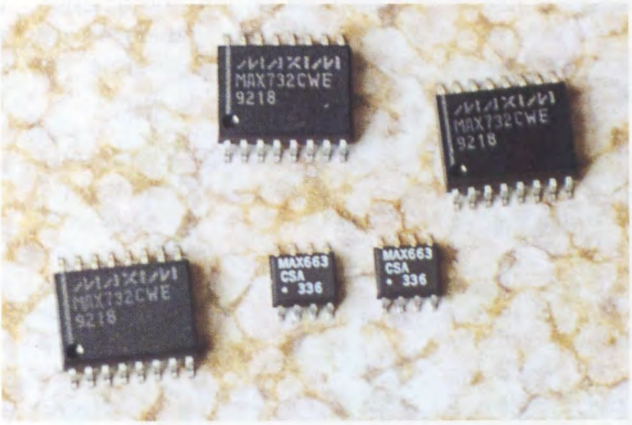

Los reguladores de tensión tienen ya dimensiones microscópicas, pero características siempre mejores.

Fusibles

Constituyen un sistema de protección primitivo, pero eficaz

Por la ley de Ohm, cuanto menor es la resistencia de un circuito mayor es la **corriente** que circula (a paridad de tensión); es necesario tener esto en cuenta en el proyecto de un circuito.

Si por ejemplo, a causa de la **avería** de un componente, la tensión de alimentación encuentra un recorrido de baja resistencia hacia masa, la corriente que resulta puede ser muy elevada.

Además de **dañar** otros componentes (como el transformador de alimentación), esta corriente puede fundir las pistas del circuito impreso, hacer "hervir" el cobre de los hilos o causar un incendio.

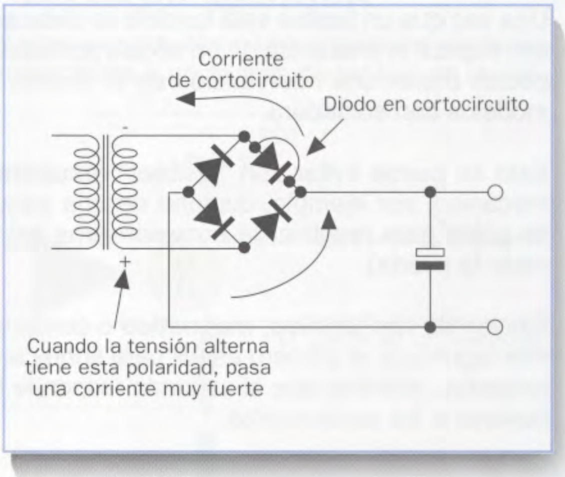
Corriente de cortocircuito
Diodo en cortocircuito
-
+
Cuando la tensión alterna tiene esta polaridad, pasa una corriente muy fuerte

Si un diodo del puente rectificador está en cortocircuito, una fuerte corriente atraviesa otro diodo y el transformador.

INTERRUMPIR EL CIRCUITO

Como la instalación doméstica reacciona a los cortocircuitos y a las sobrecargas, cortando la corriente (ver lección 1 de Aplicaciones) los dispositivos electrónicos están dotados a menudo de una **protección** local.

Un **fusible** no es otra cosa que un pequeño cable fino, protegido por un tubo de cristal, realizado de modo que se funda y corte al pasar una corriente superior a la especificada.

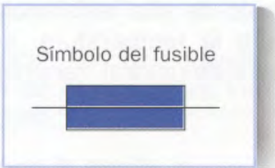

Esto ocurre porque el cable tiene una cierta resistencia (mínima, pero no nula), por lo que el paso de corriente lo **calienta**; por encima de una cierta temperatura se funde, interrumpiendo el circuito.

VELOCIDAD DE INTERVENCIÓN

Los fusibles no son todos iguales: además de por la corriente de intervención, difieren por el **tiempo de respuesta** a las sobrecargas, que depende de la inercia térmica del cable. El tipo **ultrarrápido** (FF) se utiliza para semiconductores delicados, el **rápido** (F) para uso normal, mientras que el **semiretardado** (T) soporta breves picos de exceso de corriente, por ejemplo al encendido.

Naturalmente la respuesta es **progresiva**: como se puede ver en la tabla, si la corriente supera sólo en el 20% el valor nominal (1,2 I_N), también el tipo ultrarrápido... ¡puede necesitar más de una hora!

Ultrarrápido (FF)

Especificaciones técnicas

I_N	$1{,}2I_N$	$1{,}5I_N$	$2{,}75I_N$		$4I_N$		$10I_N$
	Mín.	*Mín.*	*Mín.*	*Máx.*	*Mín.*	*Máx.*	
<125mA >100ma	>1 hora >1 hora	<30min.	50mseg 2mseg	15seg 50mseg	10mseg 1mseg	200mseg 15mseg	<5mseg <2mseg

Semiretardado (T)

Especificaciones técnicas

I_N	$2{,}1I_N$		$2{,}75I_N$		$4I_N$		$10I_N$	
	Mín.	*Máx.*	*Mín.*	*Máx.*	*Mín.*	*Máx.*	*Mín.*	*Máx.*
<125mA >100ma		2min 2min	200mseg 600mseg	10seg 10seg	40mseg 150mseg	3seg 3seg	10mseg 20mseg	300mseg 300mseg

Dos tipos distintos de fusibles: el semiretardado puede soportar breves excesos de corriente durante más tiempo sin fundirse.

Protecciones recuperables

Otras técnicas de protección no requieren componentes desechables

Una vez que un fusible está fundido se debe **sustituir**; ello implica la presencia de un zócalo portafusibles especial, o bien una intervención en el circuito para los modelos con soldadura.

Esto se puede evitar con "fusibles" **recuperables**; el mecánico, por ejemplo, dispone de una palanca que se pulsa para reactivar la conexión (tras haber eliminado la avería).

Existen de tipo **térmico, magnético** o combinado (termomagnético): el primero es útil para sobrecargas prolongadas, mientras que el segundo responde inmediatamente a los cortocircuitos.

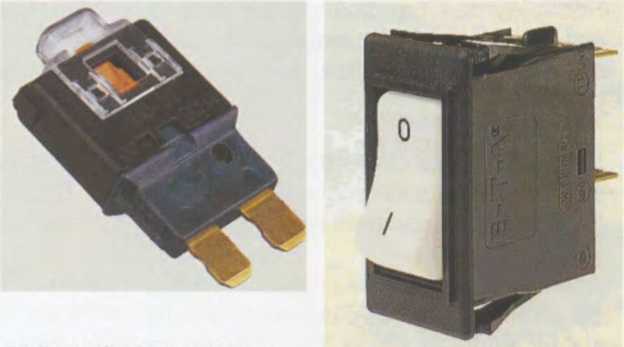

Fusibles térmicos recuperables, *el primero para coches, el segundo incluido en un interruptor.*

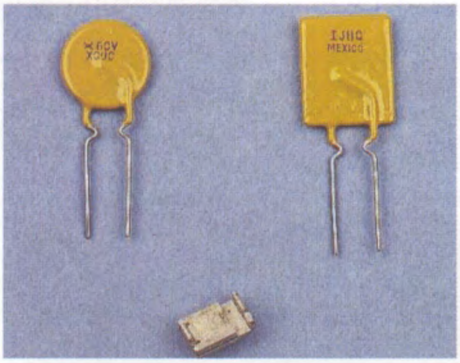

Los interruptores PTC poliméricos *interrumpen y recuperan automáticamente el circuito.*

POLÍMEROS PTC

Una solución interesante, debido a sus reducidas dimensiones y a que está privada de partes móviles, es la constituida por los **interruptores poliméricos** de coeficiente de temperatura positivo (PTC).

Presentan normalmente una resistencia muy baja, que sin embargo **sube bruscamente** al alcanzar una determinada temperatura, interrumpiendo prácticamente el circuito.

En este estado de alta resistencia, basta una débil corriente para mantener alta la temperatura: sólo **quitando tensión** baja por debajo el valor de umbral, recuperando la continuidad del circuito.

REPLIEGUE

En algunos alimentadores se utiliza una técnica especial de **protección**, denominada "repliegue": evita una excesiva disipación en el transistor regulador.

Es en la práctica una **limitación de corriente** (ver lección 14), pero con una interesante innovación: cuanto más baja la tensión a la salida por efecto de la carga, más bajo es el límite configurado por la misma corriente.

De este modo, cuanto más alta es la **tensión** en los extremos del transistor más baja es la corriente máxima que lo puede atravesar: la potencia permanece dentro de los límites tolerables del componente.

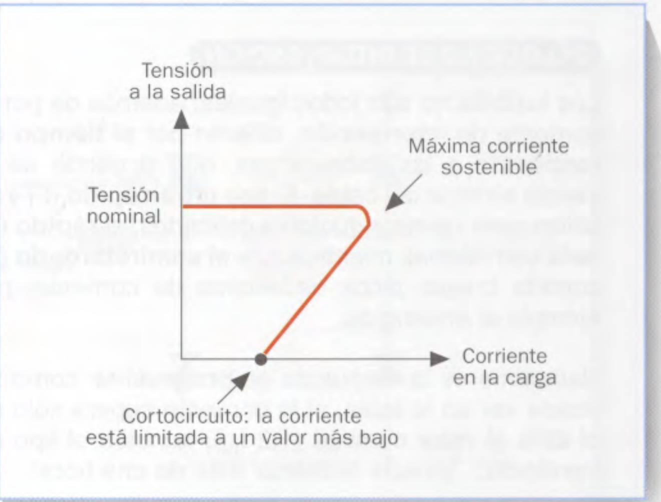

Limitando la corriente máxima según la* tensión *de salida, se controla la disipación sin tener que medir la temperatura.

Protecciones de los excesos de tensión

Una tensión elevada puede hacer tanto daño como una corriente excesiva

Cuanto más complicados son los componentes electrónicos, más crítica se hace la **tensión de alimentación**: por ejemplo, la tolerancia de los circuitos lógicos complejos está normalmente en torno al 5% más o menos.

El peor caso es, naturalmente, el de una tensión **demasiado alta**, que puede causar calentamiento y, subiendo aún más, la rotura del componente por superar el valor de ruptura.

Una tensión excesiva puede **depender** de un defecto interno (por ejemplo, la rotura de un transistor), o bien de una causa externa, como una regulación insuficiente correspondiente a un exceso de tensión en la red.

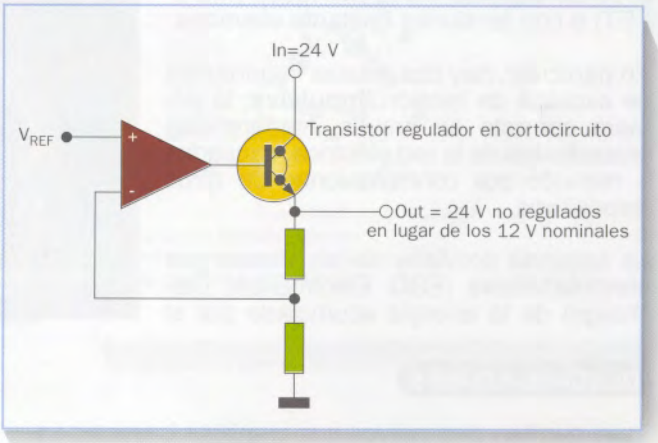
In=24 V
V_{REF}
+
-
Transistor regulador en cortocircuito
Out = 24 V no regulados en lugar de los 12 V nominales

Si el transistor regulador de una fuente de alimentación entra en cortocircuito, a la salida se encuentra la tensión completa de entrada.

PALANCA

Una técnica de protección drástica, pero eficaz, consiste en poner simplemente en **cortocircuito** hacia masa la tensión de alimentación, en caso de que supere el límite configurado.

Será el circuito de protección contra los excesos de corriente, por ejemplo un fusible, el que detecte el exceso e **intervenga**, evitando posibles daños a la fuente de alimentación.

Esta técnica se denomina "palanca", ya que equivale a tirar una palanca o una **barra de hierro** sobre los cables, poniéndolos en cortocircuito.

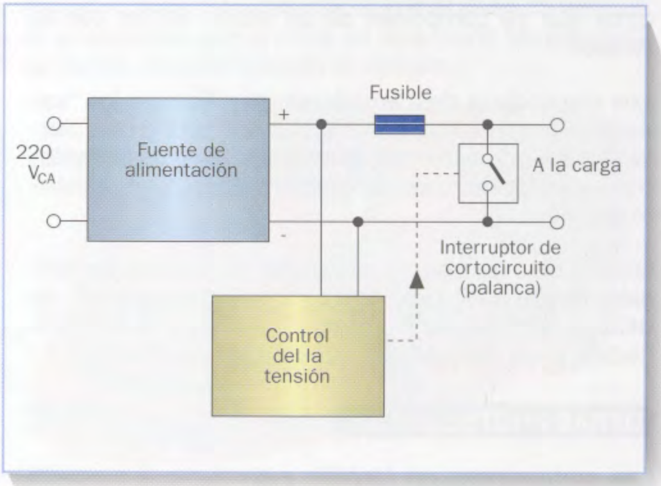
Fusible
+
-
220 V_{CA}
Fuente de alimentación
A la carga
Interruptor de cortocircuito (palanca)
Control del la tensión

Protección de palanca: en caso de exceso de tensión, la alimentación se conecta brutalmente a masa, haciendo saltar el fusible.

SCR Y PRECISIÓN

La figura muestra una sencilla protección de palanca, realizada con un **SCR** (ver lección 30), controlado por un diodo Zener, que conduce si la tensión supera el valor configurado.

Este circuito tiene la consideración de intervenir con mucha **rapidez**, evitando que el exceso de tensión pueda dañar los delicados circuitos lógicos situados en valle; puede además absorber corrientes muy elevadas.

En la práctica, sin embargo, un diodo Zener no es bastante **preciso**; para controlar el SCR se prefiere, por lo tanto, emplear circuitos integrados especiales, dotados de una referencia interna de tensión.

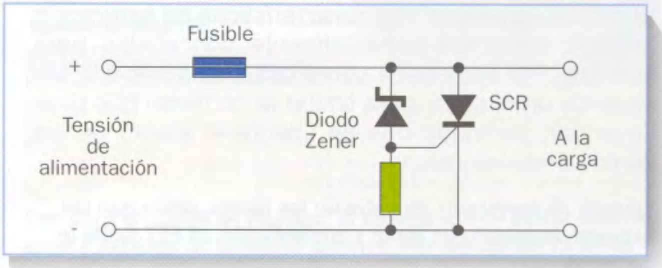
Fusible
+
-
Tensión de alimentación
Diodo Zener
SCR
A la carga

El diodo controlado (SCR) es dirigido por un dispositivo que detecta el exceso de tensión, en el caso más sencillo un diodo Zener.

Picos de tensión

Los excesos de tensión transitorios representan un peligro insidioso

Un **exceso de tensión** que dure pocos microsegundos no suele ser peligroso, pero puede llegar a serlo en el caso de circuitos especialmente delicados (MOSFET) o con tensiones bastante elevadas.

En particular, hay dos causas importantes de excesos de tensión **impulsiva**; la primera consiste en breves interferencias procedentes de la red eléctrica, causadas a menudo por conmutaciones de otros dispositivos.

La segunda proviene de las **descargas electrostáticas** (ESD: ElectroStatic Discharge) de la energía acumulada por el cuerpo humano, especialmente en los días de invierno,pero también con determinada indumentaria o calzado.

Un rayo que caiga relativamente cerca puede causar un breve, aunque amplio,* pico de tensión *en la red de alimentación.

DESCARGADORES

Igual que hay dispositivos que **cambian bruscamente** sus características al variar la temperatura, existen otros que se comportan de un modo similar con la tensión.

Los dispositivos descargadores, por ejemplo los "varistores", son normalmente aislantes, pero a una cierta tensión la resistencia se desmorona bruscamente; pueden además soportar **corrientes de pico** bastante elevadas.

Disponibles en una amplia **gama** de valores de tensión, desde pocos voltios hasta más o menos mil, se utilizan también para proteger las líneas de comunicación y los dispositivos conectados a ellas.

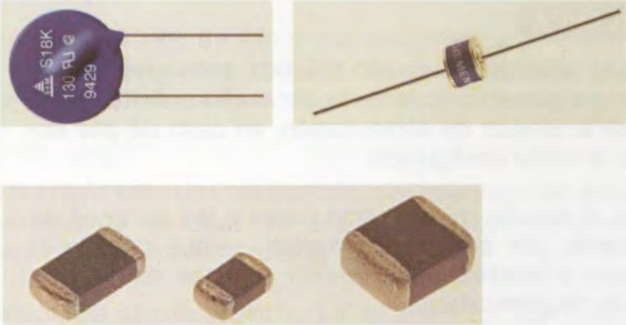

De izquierda a derecha,* un varistor, *un descargador de gas para altas tensiones y descargadores cerámicos de baja tensión para montaje superficial.

OTRAS PROTECCIONES

Las protecciones en tensión y corriente a menudo se **combinan**, especialmente en circuitos particularmente sujetos a descargar electrostáticas, como los chip contenidos en tarjetas de crédito y tarjetas telefónicas.

Es curioso observar que un FET, como puerta y origen de conexión, constituye un **limitador de corriente** natural, gracias a sus características de saturación (existen versiones especialmente concebidas para este fin). Por ello puede conectarse **en serie** a la entrada de un circuito, para limitar la corriente que puede entrar, de modo que se integre el efecto de las demás protecciones.

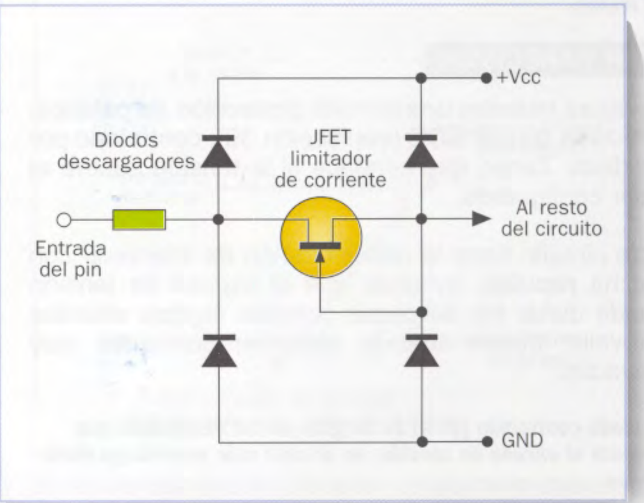

Ejemplo de protección de entrada*: los diodos descargan los excesos de tensión en masa y alimentación, el FET limita la corriente.*

Uso de los operacionales

Los op-amp son bastante más versátiles de lo que se podría pensar

Hasta hemos considerado los amplificadores operacionales en su aplicación más obvia, la que consiste en **amplificar una tensión** aplicada a la entrada.

No obstante, las características casi ideales de estos componentes les hacen adecuados a **otras aplicaciones** diversas, tanto en continua como en alterna, como veremos en esta y en otras lecciones.

En la práctica, cuando un diseñador debe realizar cualquier **circuito analógico**, parte casi siempre de un operacional y añade a su alrededor una adecuada red de componentes más sencillos.

ENTRADA EN CORRIENTE

Recordemos, de la lección 33, que por las dos resistencias R1 y R2 de un amplificador inversor pasa **la misma corriente**, y el punto común se comporta como si estuviera a masa (masa virtual, ver figura).

Ello significa que el operacional actúa (variando la tensión a la salida, y por lo tanto la caída de tensión en R2), de modo que se mantiene siempre a 0 V la entrada no inversora.

La tensión a la salida depende por tanto de la corriente que entra en el punto de masa virtual (o sale de él), y este punto es **visto** por la entrada como una sencilla conexión a masa, como si el operacional no existiera.

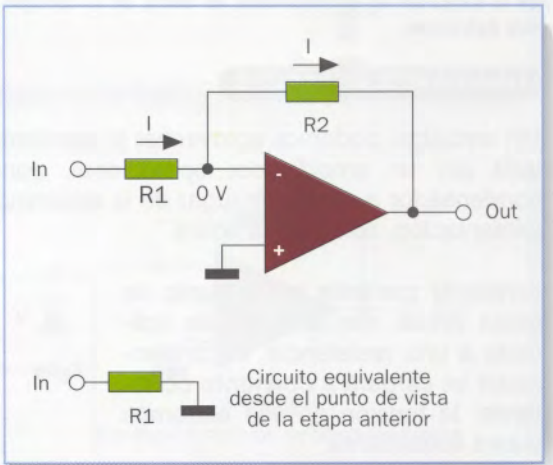

La tensión a la salida del* amplificador inversor *es proporcional a la corriente de la entrada (y tiene signo opuesto).

SUMADOR ANALÓGICO

En la figura se muestra un circuito capaz de **sumar** el valor de varias tensiones analógicas, teniendo en cuenta el signo correspondiente (positivo o negativo).

La característica notable de este circuito es que las entradas **no se influyen** entre ellas: cada una ve, de hecho, sólo la resistencia propia, como si por ahí hubiera una masa y nada más.

En realidad, cada tensión deja pasar una corriente **por la propia resistencia**, y las corrientes se suman en el punto de masa virtual; el op-amp convierte en tensión la corriente resultante.

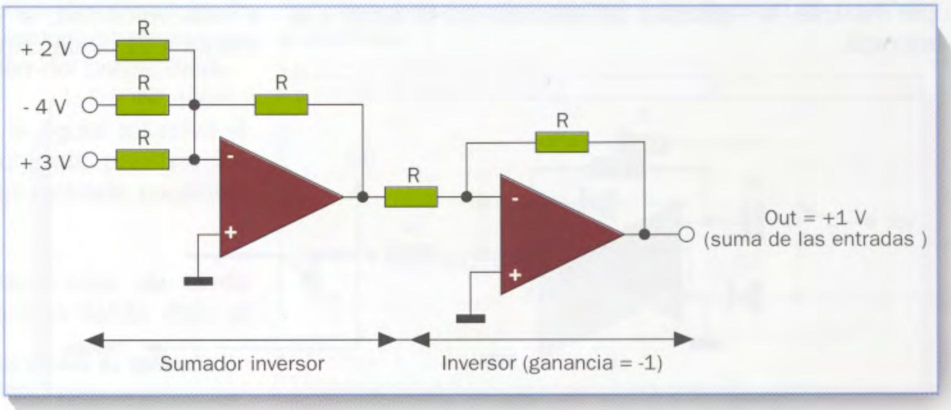

Sumador analógico: *la tensión de salida es igual a la suma de las de entradas; el segundo op-amp cambia sólo el signo.*

Integrar y diferenciar

Un op-amp puede controlar la corriente que carga un condensador

Como la tensión en los extremos de un condensador depende de la carga eléctrica acumulada entre las armaduras, para hacer **crecer linealmente** la tensión es necesario enviar electrones a un ritmo siempre igual.

Para ello se necesita una **corriente constante**, que sin embargo no se puede obtener utilizando sólo componentes pasivos, como por ejemplo el circuito RC mostrado en la figura.

De hecho, a medida que el condensador **se carga**, la tensión en los extremos de la resistencia disminuye; se reduce por tanto también la corriente que lo atraviesa, y la tensión cambia cada vez más despacio.

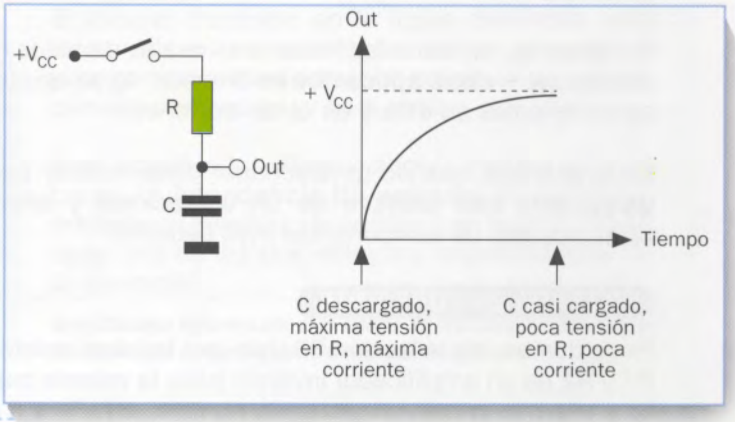

C descargado, máxima tensión en R, máxima corriente

C casi cargado, poca tensión en R, poca corriente

Una resistencia no permite obtener un funcionamiento* lineal *de la carga de un condensador, es decir, de la tensión en sus extremos.

UNA CARGA PERFECTA

Sin embargo, podemos aprovechar la **corriente** controlada por un amplificador operacional, poniendo el condensador a cargar en lugar de la resistencia de realimentación, como en la figura.

Como un extremo del condensador está a masa (virtual), la tensión de salida variará a una velocidad controlada por la tensión a la entrada: matemáticamente, el circuito es un **integrador**.

Enviando corriente en el punto de masa virtual, con una tensión aplicada a una resistencia, el condensador se cargará a **corriente constante**: la tensión en sus extremos subirá linealmente.

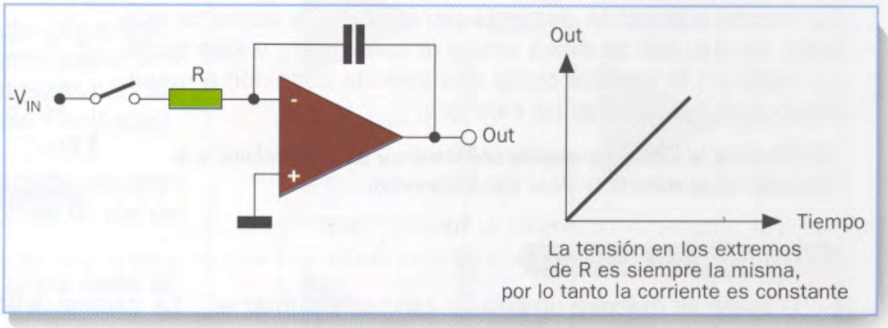

La tensión en los extremos de R es siempre la misma, por lo tanto la corriente es constante

Integrador: ***la tensión a la salida varía linealmente; la corriente en R, proporcional a V_{IN}, controla la velocidad de variación.***

MEDIR LAS VARIACIONES

La figura muestra el circuito complementario del integrador: el **diferenciador**, cuya salida indica, instante por instante, la velocidad de variación de la señal a la entrada.

La **corriente** en el condensador, y por tanto la tensión a la salida del operacional, es por tanto proporcional a esta velocidad; si la entrada no varía, no hay corriente y la salida permanece a cero.

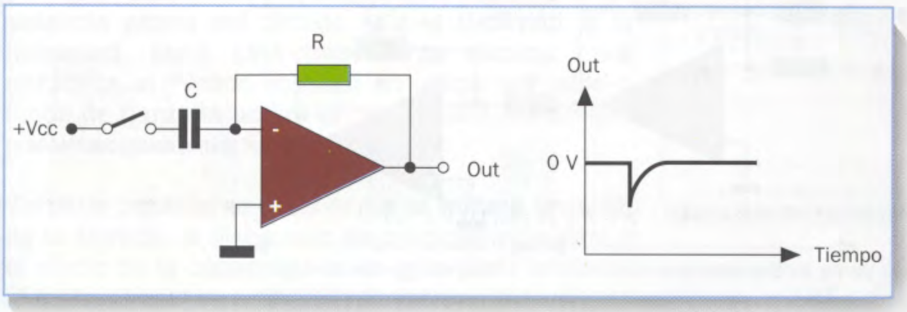

El circuito es **menos ideal** que el integrador descrito anteriormente, entre otras cosas por los inevitables tiempos de subida de la salida, y por una posible tendencia a la inestabilidad que puede requerir compensaciones.

La salida de un diferenciador indica la velocidad de subida (o bajada) de la tensión a la entrada, aquí con el signo cambiado.

Otros circuitos clásicos

Los op-amp permiten realizar varios "ladrillos de construcción"

La figura muestra un sencillo generador de **corriente constante**: el **op-amp** controla el FET, manteniendo estable la tensión en los extremos de la resistencia, y por tanto la corriente que lo atraviesa.

Como la corriente en la carga es igual a la de la resistencia, la tensión a la entrada del op-amp permite controlar esta corriente: el circuito es un **conversor tensión-corriente**.

Es aconsejable utilizar un JFET (o MOSFET) en lugar de un transistor normal BJT, para evitar la **corriente de base**, que atravesaría la resistencia pero no la carga, introduciendo un error.

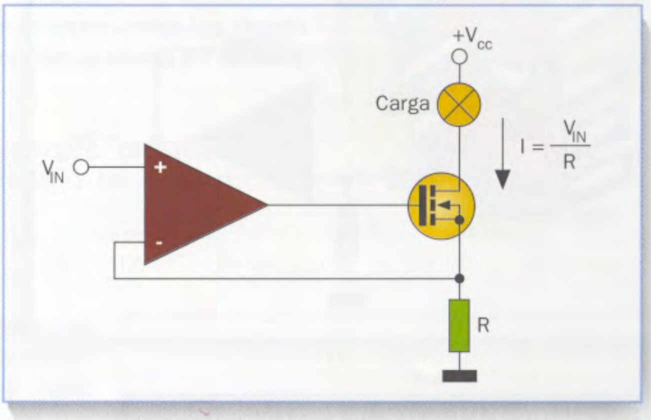

Conversor tensión-corriente: ***la corriente en la carga se mantiene estable en el valor configurado con la tensión a la entrada.***

DE CORRIENTE A TENSIÓN

Ya hemos visto que el amplificador inversor desarrolla la tarea opuesta: convierte una **corriente** a la entrada (masa virtual) en una tensión a la salida.

Ello resulta útil con dispositivos que deben trabajar con **tensión nula** en sus extremos, pero producen una cierta corriente, por ejemplo por efecto termoeléctrico o fotoeléctrico.

La figura muestra un **amplificador por fotodiodo** bastante sensible, que mantiene cero voltios en los extremos del mismo fotodiodo, haciéndole trabajar en condiciones óptimas.

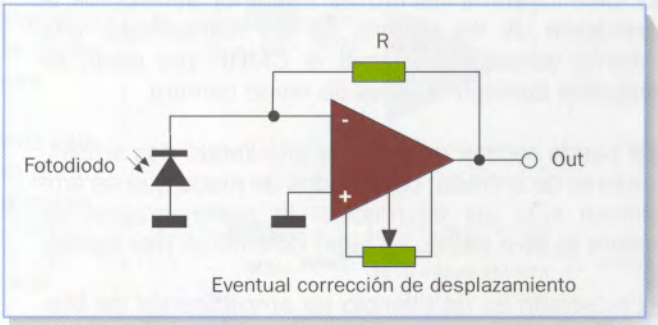

Conversor corriente-tensión, ***que amplifica la débil fotocorriente de un fotodiodo, mantenida sin tensión en sus extremos.***

RECTIFICADORES

Un rectificador normal de diodo tiene el defecto de la **caída de tensión**: no empieza a conducir hasta que no se alcanza la tensión de conducción del propio diodo.

El **rectificador de semionda** de la figura resuelve el problema con un op-amp: con entrada positivo, D2 conduce y la ganancia es -1; con entrada negativo, conduce D1 y la salida no sube.

Es igual de fácil obtener un rectificador **de onda completa**, por ejemplo sumando a la salida (con el sumador descrito en la página 145) la mitad de la señal a rectificar.

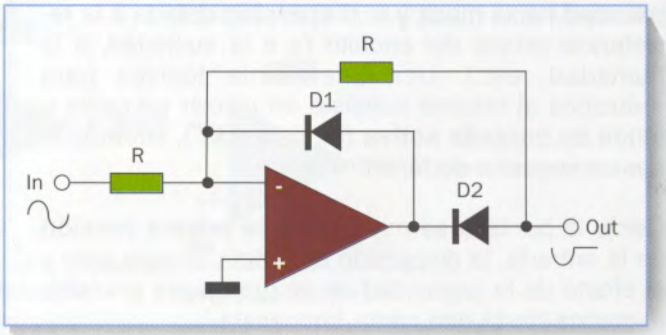

Rectificador de semionda***: la caída de los diodos es ignorada gracias a la realimentación del operacional.***

Amplificadores de precisión

Para medidas delicadas, es necesario reducir al mínimo las causas de error

Muchas señales a medir no están referidas a masa: es necesario amplificar la **diferencia de tensión** entre dos entradas, eliminando la posible tensión de un modo común (ver lección 18).

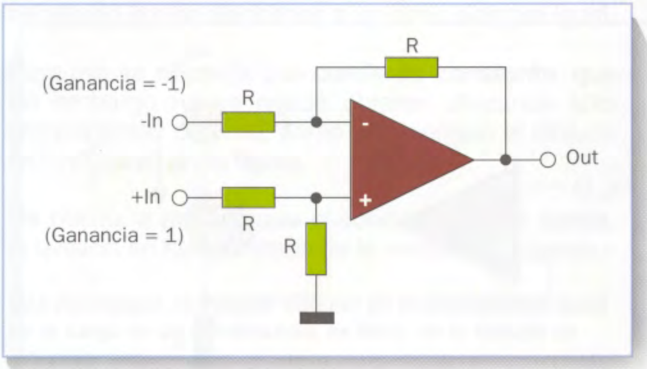

El circuito mostrado en la figura desarrolla esta tarea: como ambas entradas tienen ganancia 1 pero con **polaridad opuesta**, las posibles señales comunes se anulan unas a otras.

Este amplificador diferencial nota mucho, sin embargo, la **impedancia de entrada**: es necesario emplear un seguidor de tensión (ver lección 33) en cada una de las dos entradas, especialmente en la inversora.

Amplificador diferencial: ***a la salida se encuentra únicamente la diferencia entre las dos tensiones de entrada.***

AMPLIFICADORES DE INSTRUMENTACIÓN

El circuito arriba escrito es bastante sensible a la **precisión** de los valores de las resistencias: una mínima variación empeora el CMRR (es decir, se amplifica también la señal de modo común).

Se puede reducir este efecto añadiendo dos amplificadores de entrada, conectados de modo que se amplifican sólo las diferencias: la realimentación se refiere **al otro cable**, en lugar de a masa (ver figura).

El mostrado es un ejemplo de **amplificador de instrumentación** que, con la incorporación de un par de trimmers de ajuste, ofrece prestaciones más que buenas.

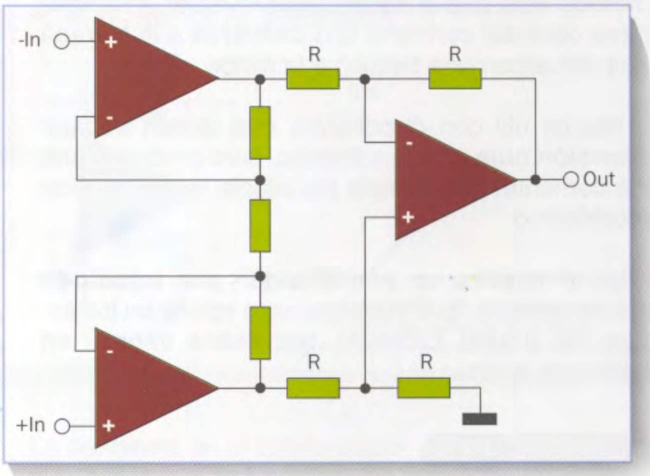

Amplificador de instrumentación: ***permite un óptimo CMRR sin la necesidad de componentes especialmente críticos.***

VIGILANCIA DE LAS ENTRADAS

Cuando se trabaja con corrientes extremadamente débiles, hay dos problemas en las entradas: la **capacidad** hacia masa y la **dispersión** debida a la resistencia propia del circuito (y a la suciedad, a la humedad, etc.). Una interesante técnica para reducirlos al mínimo consiste en utilizar un cable o funda de **pantalla activa** (o "vigilancia"), controlado por un seguidor de tensión.

Como la pantalla se mantiene **a la misma tensión** de la entrada, la dispersión es prácticamente cero y el efecto de la capacidad es en gran parte anulado, al menos hasta una cierta frecuencia.

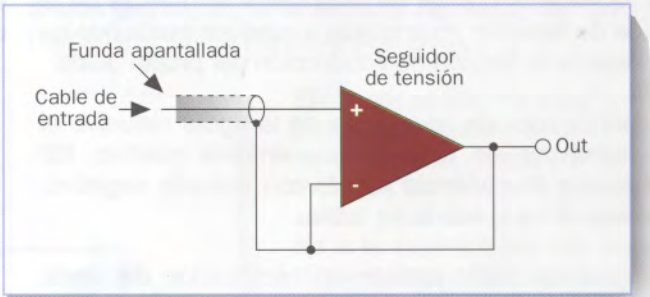

Pantalla activa ***en un cable de entrada especialmente delicado: no está a masa, pero sigue la misma señal.***

Comparadores

Viven en el confín entre el mundo analógico y el digital

Ya hemos tenido varias veces la necesidad de **decidir** si una tensión analógica es mayor o menor de un cierto valor de referencia.

Un dispositivo que desarrolla esta misión se llama **comparador**: tiene dos entradas analógicas y una salida digital, como muestra la figura.

Si la entrada **no inversora** (positivo) se encuentra a tensión más alta de la entrada **inversora** (negativo), la salida tiene valor lógico 1, en caso contrario tiene valor lógico 0.

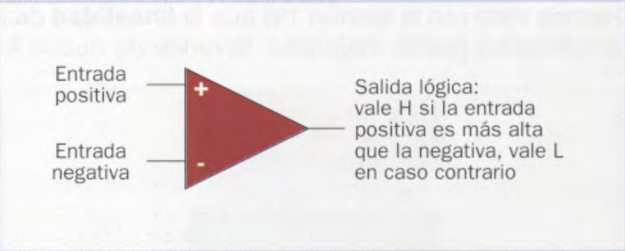
Entrada positiva
Entrada negativa
+
-
Salida lógica: vale H si la entrada positiva es más alta que la negativa, vale L en caso contrario

El símbolo del comparador es idéntico al del amplificador operacional... por buenos motivos.

OP-AMP DE ANILLO ABIERTO

Un comparador se puede realizar simplemente empleando un amplificador operacional **privado de realimentación**, es decir, funcionando a anillo abierto.

En realidad, como la **ganancia** es muy elevada, normalmente basta una fracción de milivoltio (de diferencia entre las dos entradas) para que la salida se desplace hasta la alimentación o hasta masa.

No obstante hay sutiles diferencias: los requisitos requeridos por un buen operacional no son los mismos que los de un buen comparador; existen por tanto integrados **especializados** para este último trabajo.

VELOCIDAD DE LOS COMPARADORES

Una característica importante de un comparador es su **velocidad de respuesta**, es decir, el tiempo necesario para cambiar el estado de la salida a continuación de una variación de entrada.

Depende también de la **amplitud** de la diferencia entre las dos entradas: cuanto mayor es esta diferencia, más rápido cambiará la salida de estado.

Eso se comprende fácilmente si se piensa en un comparador como en un **op-amp**: una amplia señal a la entrada produce una amplia señal a la salida, que sube por lo tanto más rápido.

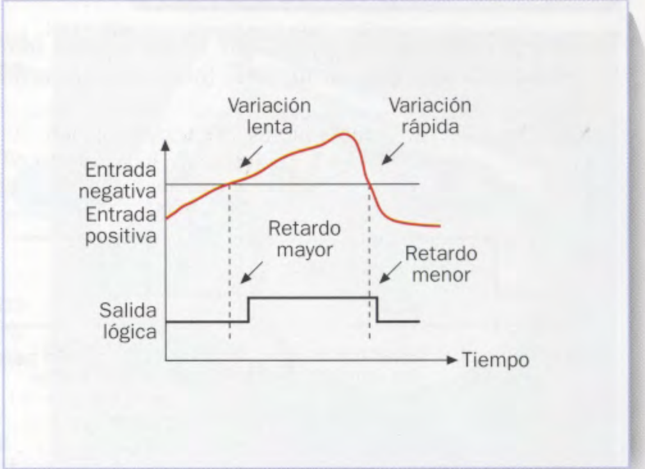

Si la diferencia entre las entradas es pequeña, la salida del comparador cambia de estado más despacio.

Realimentación positiva

Es posible hacer conmutar un comparador de un modo más limpio

Hemos visto (en la lección 19) que la **linealidad** de un amplificador puede mejorarse llevando de nuevo a la entrada una parte de la señal de salida con signo opuesto.

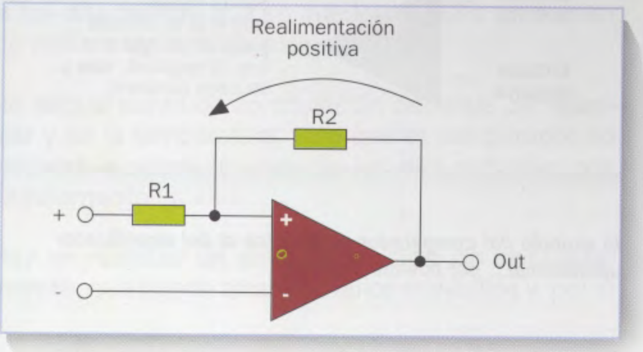

Se obtiene un efecto completamente distinto si se lleva de nuevo hacia atrás un porcentaje de la salida **con el mismo signo**: se tiene una realimentación (o "reacción") positiva en lugar de negativa.

Esto garantiza una **transición** (cambio de estado lógico) rápida y limpia de la salida, también para variaciones muy lentas de la diferencia entre las entradas.

***La realimentación positiva** consiste en llevar de nuevo a la entrada positiva (por tanto en fase) una parte de la señal de salida.*

TRIGGER SCHMITT

Con la realimentación positiva se obtiene un trigger Schmitt (ver lección 17 de Digital), es decir, un comparador con **dos umbrales distintos**: una para subir, otra para bajar.

El funcionamiento se ilustra en la figura: cuando la salida sube, **sube** también ligeramente la entrada positiva mediante la realimentación, aumentando la diferencia entre las dos entradas del comparador.

En la práctica, una vez que la conmutación se ha iniciado, también con una diferencia mínima entre las dos entradas, **se ayuda por sí solo**, garantizando un paso limpio y rápido al estado lógico opuesto.

***Trigger Schmitt:** la salida desplaza a la entrada, de modo que controla el comparador con una señal más amplia.*

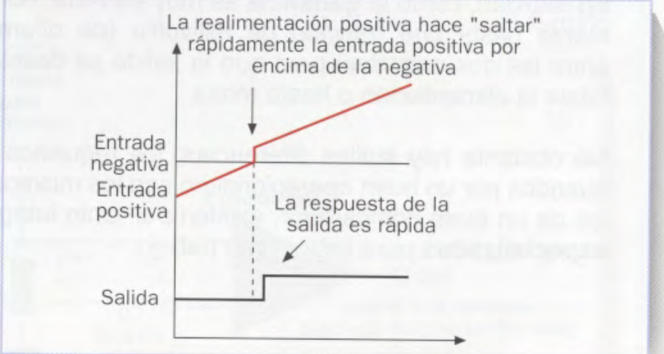

HISTÉRESIS: VIRTUDES Y DEFECTOS

Ya hemos ilustrado, en la lección 17 de Digital, cómo la presencia de dos umbrales (histéresis) permite eliminar **transiciones no deseadas** debidas, por ejemplo, a ruido eléctrico en una entrada.

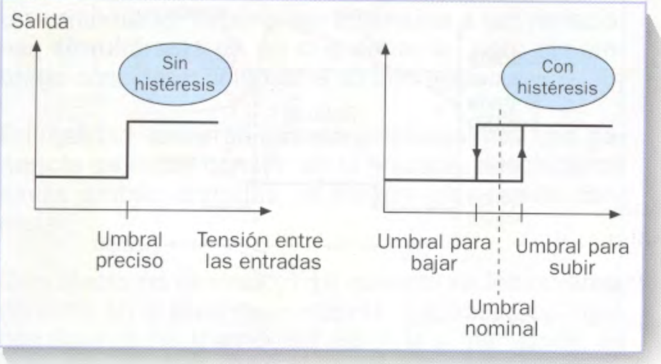

Se trata, no obstante, de un arma **de doble filo**: la histéresis reduce tanto la precisión como la posibilidad de seguir pequeñas variaciones de la señal de entrada.

La **zona vacía** entre umbrales puede ser inaceptable en aplicaciones que requieran la máxima precisión, como por ejemplo los ADC (ver lección 21 de Digital).

En un trigger Schmitt, la presencia de dos umbrales introduce una imprecisión en el comparador.

Usos de los comparadores

Estos componentes se emplean en numerosas situaciones

Una primera aplicación obvia es en los **conversores analógico-digitales** (ADC, lección21 de Digital), donde es necesario comparar la tensión de entrada con la de referencia.

Ya hemos observado que un comparador puede considerarse como un **ADC de un bit**, ya que indica si la tensión de entrada está por encima o por debajo de la de referencia.

En el comparador de un ADC son importantes tanto la precisión como la **velocidad**: esta última determina la frecuencia máxima de conversión.

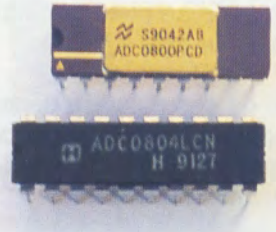

En un ADC, uno o varios comparadores constituyen el delicado punto de unión entre la sección analógica y la digital.

VENTANA Y PASO POR EL CERO

A veces es útil saber si una señal analógica está **comprendida** entre dos valores de referencia configurados; se pueden utilizar en este caso dos comparadores, como muestra la figura.

Esta configuración se llama **comparador de ventana**: la salida está al valor lógico 1 sólo si la tensión de entrada está entre los dos valores de referencia.

Si estos últimos están en torno al cero se tiene un detector de **paso por el cero**, útil por ejemplo para conmutar un triac al principio de una semionda (ver lección 30).

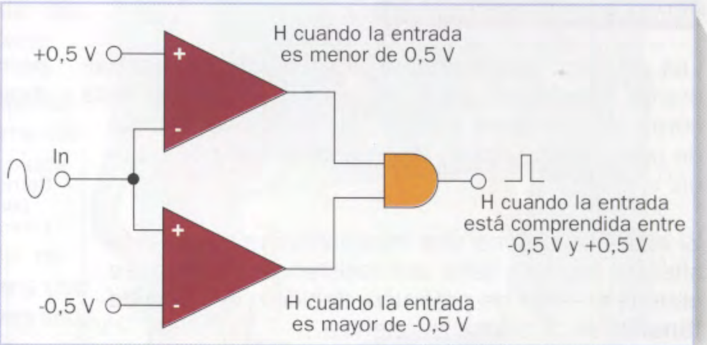
+0,5 V
+
-
H cuando la entrada es menor de 0,5 V
In
H cuando la entrada está comprendida entre -0,5 V y +0,5 V
+
-
-0,5 V
H cuando la entrada es mayor de -0,5 V

Detector de paso por el cero, realizado con un comparador de ventana.

ANALÓGICO Y DIGITAL

Los comparadores se utilizan también en campo puramente analógico, por ejemplo para la realización de **osciladores** con forma de onda no sinusoidal (ver lección 20).

También los circuitos exclusivamente **digitales** pueden emplear comparadores, por ejemplo para reconstruir la señal tras la alteración causada por una larga línea de transmisión (ver lección 31 de Digital).

Se utilizan mucho, entre otras cosas, en el interior de los **chips de memoria**, donde un fenómeno análogo se verifica a lo largo de las conexiones entre las distintas celdas (ver lección 22 de Digital).

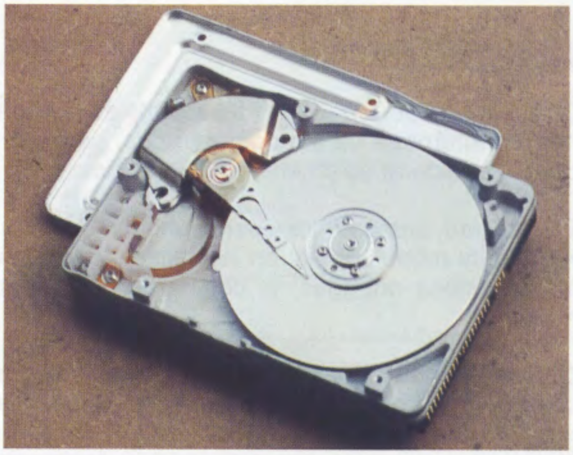

La débil señal leída por la cabeza de un disco duro se amplifica y envía a un comparador para reconstruir la información digital.

Límites de los comparadores

Como siempre, el componente perfecto no existe

El **comparador ideal** debería tener una precisión absoluta y un tiempo de conmutación cero; naturalmente los dispositivos que se pueden adquirir en comercio están muy lejos de esta situación.

No sólo el punto de conmutación depende del ejemplar y de la temperatura, sino que el comparador no empieza a conmutar cuando las dos entradas son **exactamente** iguales.

Hay en realidad un error: la **tensión de desplazamiento**, que puede alcanzar varios milivoltios y, por lo tanto, empeorar sensiblemente la precisión; puede, además, variar con la temperatura y con el tiempo.

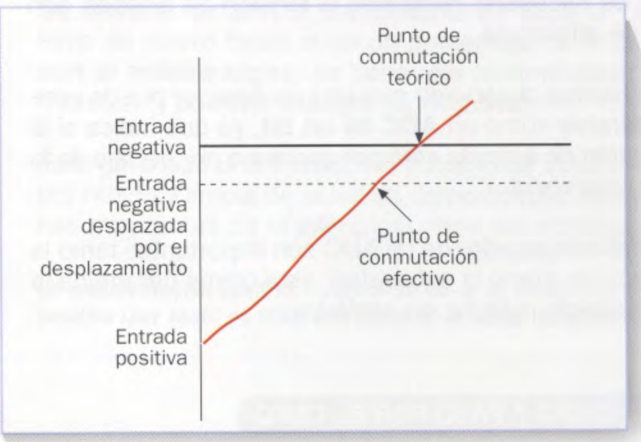

La* tensión de desplazamiento *desplaza el punto de conmutación del comparador, produciendo un error de precisión.

LÍMITES DE LA ENTRADA

Las entradas absorben una cierta **corriente**, normalmente mayor en los comparadores rápidos: esta corriente (que entre otras cosas puede ser distinta de una entrada a otra) es proporcionada por la señal a medir.

Si esta última tiene una impedancia relativamente alta, por ejemplo tiene una resistencia en serie, se obtiene un error de precisión, debido a la **caída de tensión** en la misma resistencia.

En algunas aplicaciones es necesario que las entradas puedan **desplazarse** hasta la alimentación, pero esto no es siempre posible: es necesario utilizar modelos especiales.

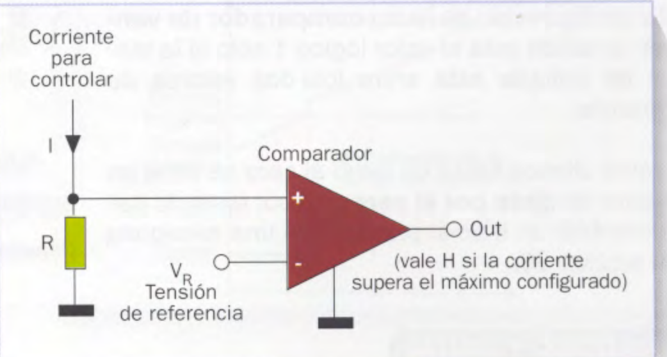

Para controlar la tensión sobre R, las entradas del comparador deben poder trabajar hasta la* alimentación negativa, *que está conectada a masa.

MODO COMÚN

Un comparador debería ser insensible a las variaciones **simultáneas** de las dos entradas, pero sin embargo considerar solamente su diferencia.

En realidad, como en los operacionales, hay una **ganancia** de modo común: en la práctica, desplazando juntas ambas entradas, el comparador puede conmutar.

Otro efecto no deseado, casi ausente en los op-amp, consiste en el **pico de corriente** que absorben algunos tipos de comparadores desde la alimentación, en el momento en el que la salida cambia de estado.

PORTUGAL
LM2903P
8913BR

A pesar de sus límites,* los comparadores *están entre los ingredientes fundamentales de muchos circuitos electrónicos.

Alimentadores de conmutación

La técnica "switching" elimina el regulador lineal

Como hemos visto en la lecciones 14 y 35, la principal desventaja de un alimentador estabilizado de tipo lineal reside en la **disipación** de potencia en el transistor regulador.

Es aparentemente inevitable, ya que depende de la **ley de Ohm**: de hecho es el producto de la tensión en los extremos del transistor por la corriente que lo atraviesa.

No obstante, existe una alternativa: **no regular** la corriente que atraviesa el transistor, sino limitarse a encenderla o apagarla completamente, alternando rápidamente entre los dos estados.

Si los PC utilizaran alimentadores lineales... ¡serían como radiadores!

CONMUTAR RÁPIDAMENTE

La figura muestra cómo se puede regular la luminosidad de una bombilla, maniobrando muy rápidamente (decenas de veces por segundo) un sencillo **interruptor**.

La luminosidad dependerá de la **corriente media**, es decir, de la relación entre el tiempo de encendido y de apagado en cada ciclo; la inercia térmica del filamento hará uniforme la luz.

Este es el principio de base de un **alimentador de conmutación** o "switching"; se observa que en el interruptor no hay disipación: está abierto (I=0) o cerrado (V=0).

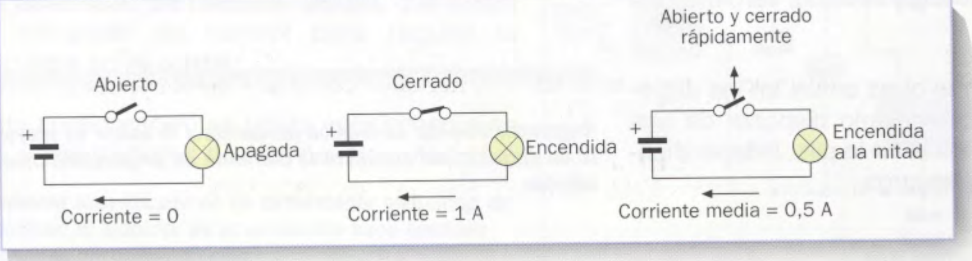

Una conmutación rápida del interruptor entre dos estados permite regular la corriente sin disipación de potencia.

INDUCTOR DE ACUMULACIÓN

Sustituyendo el interruptor con un transistor, bipolar o MOSFET, queda el problema de cómo **nivelar** la corriente producida, de modo que pueda utilizarla para alimentar otros circuitos.

Ello se resuelve acumulando energía en un **inductor**, que actúa como "volante": se carga cuando el interruptor (transistor) está cerrado, y mantiene la corriente cuando está abierto.

El **diodo** mostrado en la figura tiene la misión de cerrar el circuito en esta segunda fase, permitiendo al inductor proporcionar corriente a la carga.

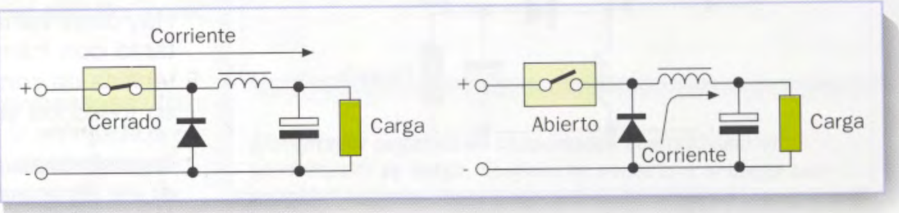

***Regulador switching:** el inductor acumula energía (como campo magnético) en la primera fase, y la restituye en la segunda.*

Ventajas de los switching

Son superiores a los reguladores tradicionales bajo distintos aspectos

En un regulador de conmutación, la **eficacia** o "rendimiento" (es decir, la relación entre potencia a la salida y potencia a la entrada) es muy elevada, incluso superior al 90%.

El uso de una **frecuencia** de conmutación relativamente elevada (normalmente, decenas o centenas de KHz) permite utilizar inductores de pequeñas dimensiones, reduciendo el rizado y el coste total.

Además, al no tener que limitar la caída de **tensión** en los extremos del transistor, se puede enviar a la entrada también una tensión relativamente alta y poco filtrada.

Un alimentador switching utiliza componentes de pequeñas dimensiones, con respecto a la potencia regulada (aquí 110 W y 250 W).

ELEVAR LA TENSIÓN

La figura muestra un regulador **step-up** (de elevación), cuya tensión a la salida es superior a la de entrada: aprovecha el exceso de tensión de apertura de un inductor (ver lección 6).

Cuando el transistor (aquí, un MOSFET) abre, el inductor trata de mantener la corriente, **aumentando** la tensión hasta enviar en conducción al diodo, cargando así el condensador de filtro.

Esta solución se utiliza entre otras cosas en los dispositivos **de pila**, donde es necesario disponer de una tensión constante, superior a la de la pila, independientemente de su estado de descarga.

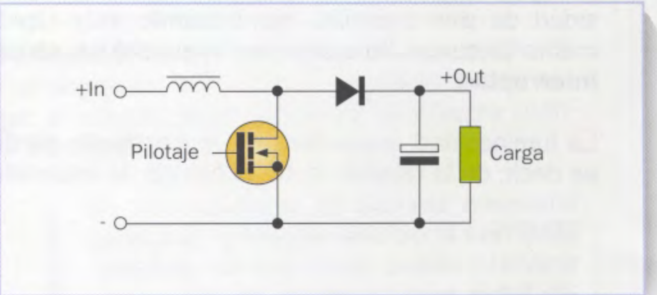

***Regulador step-up:** la tensión producida a la salida es mayor que la de entrada; por supuesto la corriente es proporcionalmente inferior.*

TENSIONES NEGATIVAS

Intercambiando de sitio el transistor y el inductor, se obtiene un regulador capaz de producir una tensión **de polaridad opuesta** con respecto a la disponible, y si se desea incluso con tensión superior.

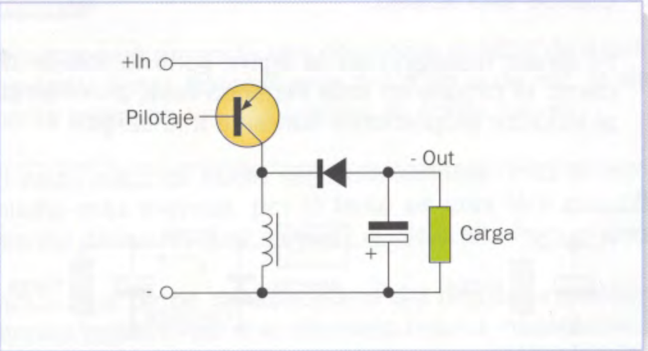

En el circuito mostrado en la figura, cuando el transistor deja de conducir, el inductor trata de **mantener** su propia corriente "succionándola" de la carga, produciendo así una corriente inversa.

Hay otras **variantes** de estos esquemas principales, tanto con transistores bipolares como MOSFET: la técnica de conmutación está ampliamente difundida en todos los sectores de la electrónica.

*Regulador capaz de obtener una tensión **negativa**, partiendo de una alimentación positiva.*

Alimentadores completos

En el mercado se encuentra una amplia selección de integrados de control

Para construir un alimentador estabilizado con un regulador switching, es necesaria alguna forma de **realimentación**, que examine el valor de la salida y la corrija en consecuencia.

Para variar la tensión es necesario modificar la **relación** entre el tiempo de apertura y de cierre (ciclo de servicio) del interruptor de transistor, o bien sobre la frecuencia de los impulsos, si estos tienen una duración fija.

La figura muestra un sistema de regulación primitivo de **comparador**: cuando la tensión a la salida alcanza el valor configurado, el transistor se apaga; cuando baja, se enciende de nuevo.

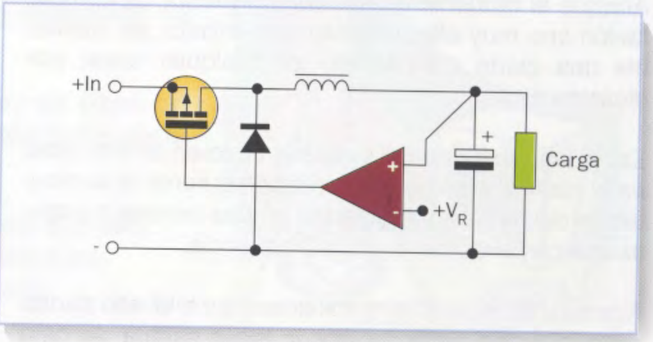

Un sencillo comparador, mejor con una pequeña histéresis, basta para realizar un alimentador switching.

SWITCHING DE TRANSFORMADOR

El circuito mostrado en la figura es, sin embargo, un switching de transformador, cuyo regulador se encuentra por el lado **de alta tensión** (220 V_{CA}), mientras que la carga está del lado aislado.

Actúa conmutando la corriente en el primario del transformador, teniendo entre otras cosas un **devanado de realimentación**, que utiliza el integrado de control para regular la corriente en la carga.

Esta configuración se utiliza comúnmente en los **alimentadores de red**, por ejemplo en los de televisores y ordenadores: es compacta, eficaz y económica.

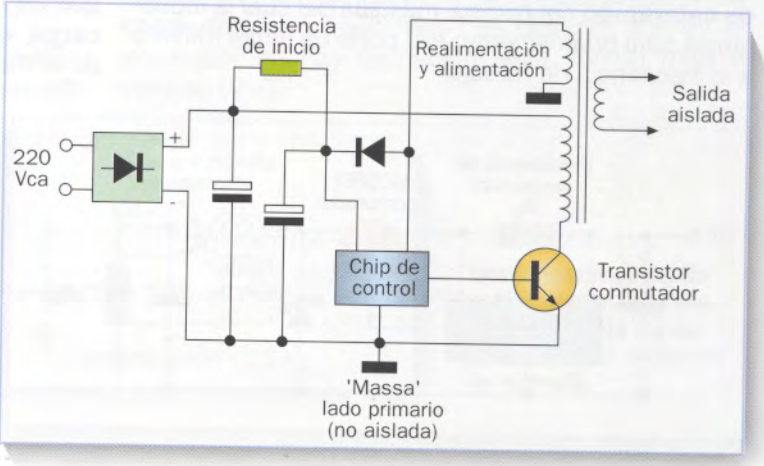

Esquema simplificado de un alimentador switching de televisor: el inductor de acumulación hace también las veces de transformador.

CAPACIDADES CONMUTADAS

Existe también una técnica para realizar alimentadores de conmutación **privados de inductores**, de modo que se reduzca posteriormente coste y rizado, aceptando alguna limitación.

Se denomina "de capacidades conmutadas", ya que consiste en conectar alternativamente uno o varios **condensadores** a distintos puntos del circuito, como se ilustra en la figura.

Este sistema no permite **corrientes** elevadas: se utiliza por ejemplo para obtener –5 V o +12 V en muchos circuitos digitales (partiendo de –5 V), evitando la necesidad de una alimentación independiente.

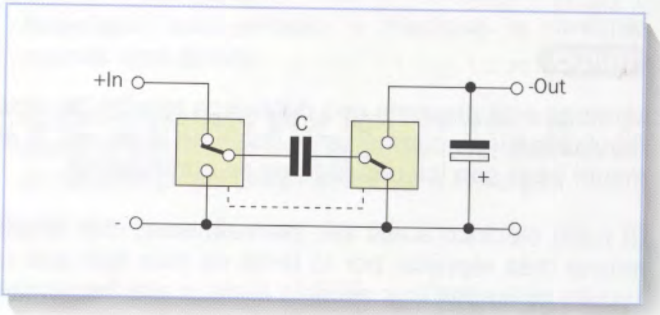

Alimentador negativo de capacidades conmutadas: el condensador se carga, después se conecta a la carga con polaridad opuesta, invirtiendo ambos desviadores.

Límites de los switching

En algunas situaciones puede ser conveniente volver al viejo sistema

Aunque el rendimiento de un alimentador de conmutación sea muy alto, teóricamente infinito, es inevitable una cierta **disipación**, en cualquier caso, por distintos motivos.

Están las **pérdidas** del inductor, tanto en el hilo como en el núcleo, y las que se producen durante la conmutación del transistor regulador; ambas tienden a acentuarse con el aumento de la frecuencia.

Además, la tensión en los extremos del **diodo** causa disipación: se prefieren por lo tanto diodos de baja caída y de rápida interrupción cuando la tensión cambia de signo (normalmente son de tipo Schottky).

Los alimentadores switching de potencia requieren, en cualquier caso, un cierto enfriamiento, normalmente con disipadores y ventiladores.

CORRIENTES

La corriente en el inductor no debe superar el valor de **saturación** del núcleo, más allá del cual la inductancia baja bruscamente; ello pone un límite mínimo a la frecuencia de trabajo.

Una limitación menos obvia consiste en el hecho de que un regulador switching no puede funcionar **sin carga**: el inductor debe descargar su energía cuando el transistor abre.

En consecuencia, es siempre necesario absorber, al menos, una cierta **corriente mínima**, de otro modo puede intervenir un circuito de protección para evitar excesos de tensiones a la salida.

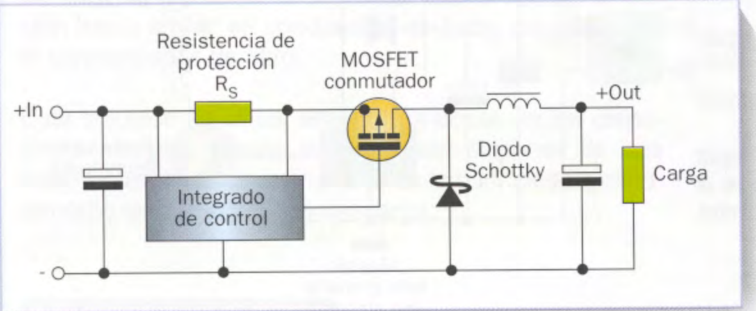

Como para los reguladores lineales, la corriente a la salida a menudo está limitada, midiéndola con una resistencia en serie.

RUIDO

Mientras está presente una debilísima tensión de rizado a la salida de un regulador lineal, normalmente doble de la de red, la situación es notablemente **peor** con los reguladores de conmutación.

El ruido eléctrico suele ser, normalmente, más amplio y de **frecuencia** mucho más elevada, por lo tanto es más fácil que cree problemas en circuito delicados (por ejemplo audio o alta frecuencia).

Incluso los **picos de absorción** del regulador pueden provocar interferencias posteriores: si la situación resulta inaceptable, sólo queda volver a utilizar un regulador lineal clásico, o combinar las dos técnicas.

Baterías

La energía química se puede convertir en energía eléctrica

Cuando Alejandro Volta puso a punto en 1799 su "aparato electromotor de columna", es decir, la primera **pila**, se hizo posible producir electricidad sin partir de un movimiento mecánico.

No obstante, la energía no se crea de la nada, sino que simplemente **cambia de forma**: una pila produce energía eléctrica a expensas de la energía química contenida en sus componentes.

Estos componentes pasan gradualmente de **un estado** a otro (conteniendo menor energía), hasta que ceden toda la parte de energía utilizable eléctricamente: la pila está entonces descargada.

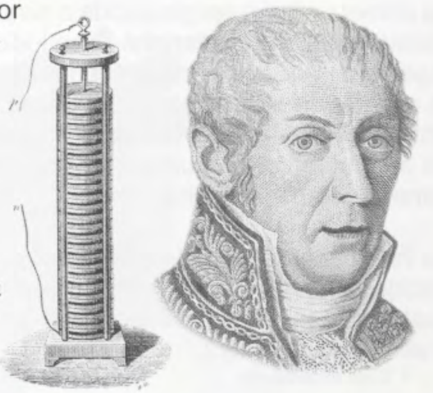

Alejandro Volta ***demostró que se puede producir corriente por vía electroquímica.***

ELECTROQUÍMICA

La figura muestra el principio de una **celda galvánica**, es decir, una pila elemental: dos electrodos están en contacto con distintas soluciones químicas (electrolitos), conectadas por un "puente salino".

En la celda se producen **dos reacciones** químicas: en el ánodo hay una oxidación (aumento de oxígeno), que libera electrones, mientras que en el cátodo se produce una reducción complementaria, que los absorbe.

Los electrones **circulan** por el circuito externo, produciendo una corriente eléctrica, mientras que en el interior se desplazan átomos con exceso o carencia de electrones (iones).

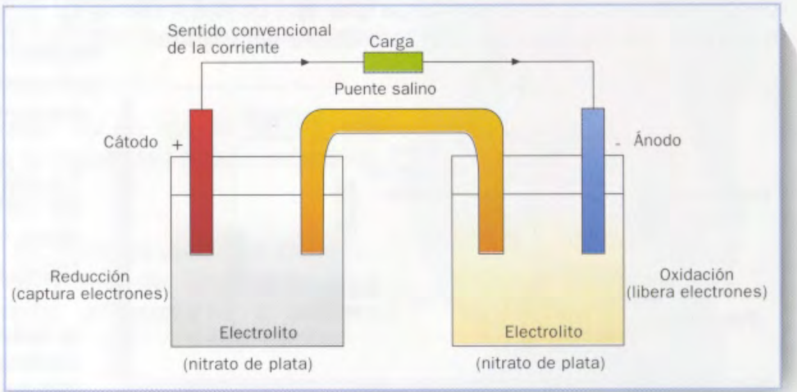
Sentido convencional de la corriente
Carga
Puente salino
Cátodo +
- Ánodo
Reducción (captura electrones)
Oxidación (libera electrones)
Electrolito
Electrolito
(nitrato de plata)
(nitrato de plata)

Celda electrovoltaica: ***al desplazamiento interno de iones (átomos cargados eléctricamente) le corresponde un desplazamiento externo de electrones.***

PILAS Y BATERÍAS

Aunque la pila de Volta estaba formada por varias celdas superpuestas (para tener tensión más alta), y era por tanto una **batería**, los términos "pila" y "batería" son sinónimos en el uso práctico. Cada tipo de pila tiene una **tensión de celda** distinta, dependiente del material utilizado para los dos electrodos; por ejemplo las pilas "alcalinas" comunes tienen 1,5 V por celda.

Esta tensión **baja** a medida que la celda se descarga, de un modo que depende también del tipo de reacción química en la que se basa el funcionamiento de la pila en cuestión.

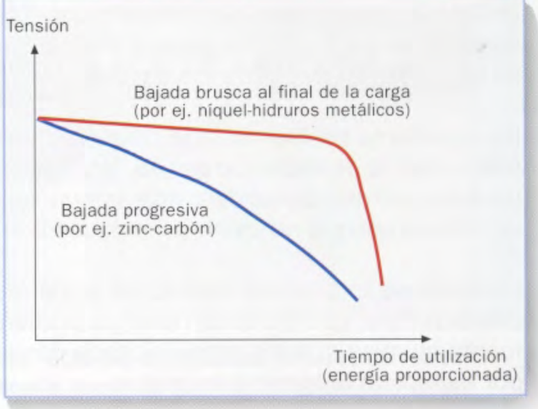

El progreso de la tensión durante la descarga de una batería depende del tipo de celda utilizada.

Baterías primarias

Una vez descargadas, estas fuentes de energía no son recuperables

La primera pila en ser producida a nivel industrial a gran escala fue el tipo **Leclanché**, introducida en 1868 por el ingeniero francés del mismo nombre, y utilizada durante más de un siglo. Denominada también **zinc-carbón** (Zn-C), utiliza un electrodo externo de zinc y uno interno de grafito, con cloruro de amonio y bióxido de manganeso como electrolitos.

La variante **alcalina** más reciente ofrece una mayor capacidad, aunque en las primeras versiones contuviera sustancias altamente tóxicas, como el mercurio (hoy en gran parte eliminado).

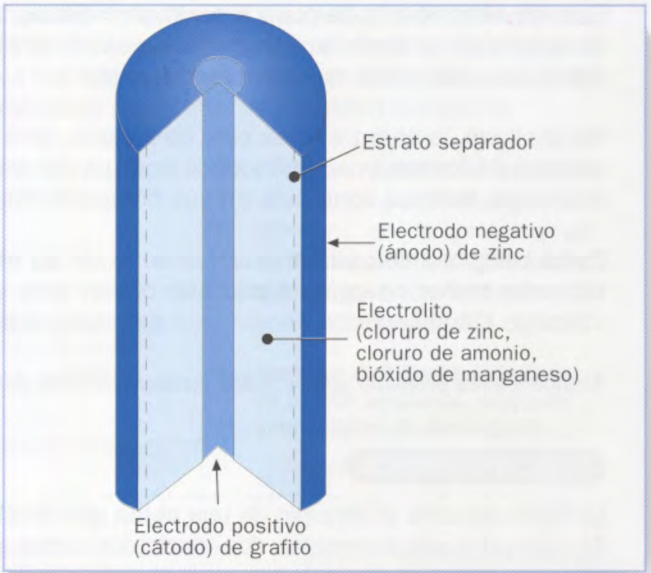
Estrato separador
Electrodo negativo (ánodo) de zinc
Electrolito (cloruro de zinc, cloruro de amonio, bióxido de manganeso)
Electrodo positivo (cátodo) de grafito

Estructura interna de una **pila de zinc-carbón*****: el ánodo central es un cilindro de grafito y el cátodo es el envoltorio de zinc.***

CAPACIDAD Y CORRIENTE

La capacidad de una pila se mide en **amperios-hora** (Ah): se obtiene multiplicando la corriente suministrada por la duración de la descarga (por ej. 100 mA x 10 h = 1 Ah), hasta alcanzar una tensión mínima fijada.

Son importantes también la **corriente máxima** y la resistencia interna, que indica cuánto baja la tensión aplicando una cierta carga, es decir, requiriéndole corriente a la pila.

En algunas aplicaciones, como la tracción eléctrica, se considera también la **densidad energética**, es decir, la cantidad de energía contenida en un kilo de peso, o en dm^3 de volumen.

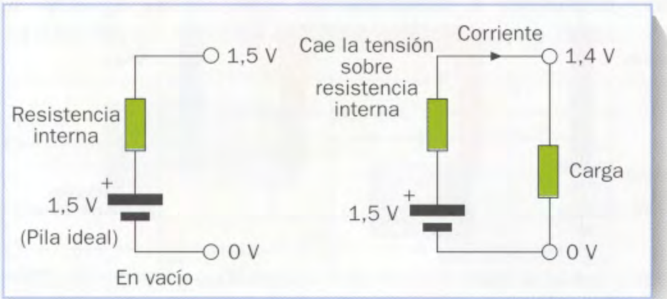
1,5 V
Resistencia interna
+
1,5 V
(Pila ideal)
0 V
En vacío
Cae la tensión sobre resistencia interna
Corriente
1,4 V
+
1,5 V
Carga
0 V

La resistencia interna, ***debida a las características electroquímicas de la pila, hace que la tensión de carga sea más baja que en vacío.***

LITIO Y PILAS DE COMBUSTIBLE

Entre los distintos modelos, las pilas de **litio** (en la versión no recargable) se utilizan mucho por su gran densidad energética, por ejemplo para la alimentación de la memoria CMOS de los ordenadores.

Es importante también la vida de almacenaje, que depende de la **autodescarga** de las pilas: todos los modelos pierden capacidad, más o menos rápidamente, incluso aunque no estén conectados a un circuito.

Un enfoque totalmente distinto es el de las **pilas de combustible**, que producen energía partiendo de una fuente química externa proporcionada continuamente, por ejemplo hidrógeno y oxígeno.

Baterías de litio ***no recargables: pueden alimentar una memoria CMOS durante años.***

Acumuladores

Las baterías secundarias son aquellas cuya carga es reversible

Tanto por motivos de coste como de contaminación, es más interesante poder **recargar** las pilas descargadas, en lugar de tener que tirarlas: se llaman entonces "baterías secundarias" o "acumuladores".

Históricamente, el tipo más difundido es el de **plomo-ácido**, utilizado en los automóviles y (en la versión sellada plomo-gel) para las baterías de antirrobo y de sistemas de emergencia o copia de seguridad.

Sin embargo este tipo resulta bastante "pesado", es decir, con escasa energía acumulable por unidad de peso: en los dispositivos portátiles se emplean habitualmente, por lo tanto, los acumuladores más cómodos de **níquel-cadmio** (NiCd).

UPPER LEVEL
LOWER LEVEL
YUASA
12N5-3B
YUASA BATTERY CO. LTD

Una batería de coche de plomo-ácido está compuesta de seis elementos de 2 V en serie, para un total de 12 V.

BATERÍAS MÁS EVOLUCIONADAS

Ambos tipos citados utilizan **metales pesados** (respectivamente, plomo y cadmio), tóxicos para el hombre y capaces de contaminar terrenos, vegetación, fauna y estratos acuíferos.

Además, las níquel-cadmio del tipo "sinterizado" (el de mayor capacidad) sufren el molesto **efecto memoria**, por lo que la capacidad puede reducirse antes de lo debido en caso de descargas parciales.

La continua búsqueda produce, por tanto, tecnologías siempre nuevas, como la de **níquel-hidruros metálicos** (NiMH) y la de **litio-iones**, muy utilizadas en los dispositivos portátiles, como ordenadores o teléfonos móviles.

SUPERCONDENSADORES

Un curioso tipo de pilas recargables... no son en realidad pilas: se trata de los "supercondensadores", es decir, condensadores especiales de **capacidad** elevadísima.

Se utilizan normalmente para **alimentar** memorias CMOS durante periodos relativamente breves, por ejemplo durante la noche, sin recurrir ni a baterías ni a memorias permanentes.

Son, en la práctica, electrolitos con una **corriente de pérdida** bastante baja, que permiten la conservación de la carga incluso durante varias horas (en condiciones favorables).

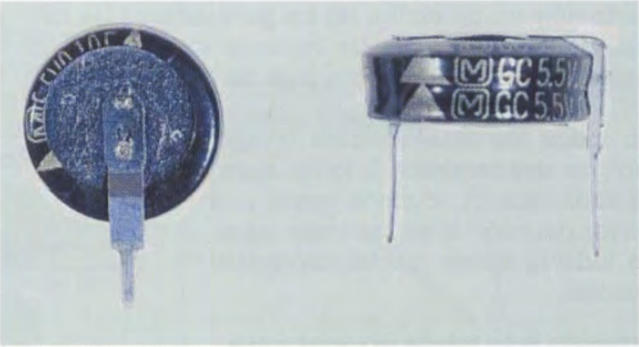

A pesar de sus reducidas dimensiones, la capacidad de un supercondensador puede ser también del orden del Faradio.

Cargador de baterías

Es necesario un circuito adecuado para cada tipo de celda

Los dispositivos de recarga para los acumuladores son fundamentalmente de dos tipos: el primero es el que se utiliza en **tampones**, es decir, para mantener una batería constantemente cargada y lista para su uso. Se aplica principalmente a las baterías de plomo-ácido (Pb) abiertas, o en la versión hermética (que no produce hidrógeno y puede estar en ambientes cerrados), cargándose normalmente a **tensión constante**.

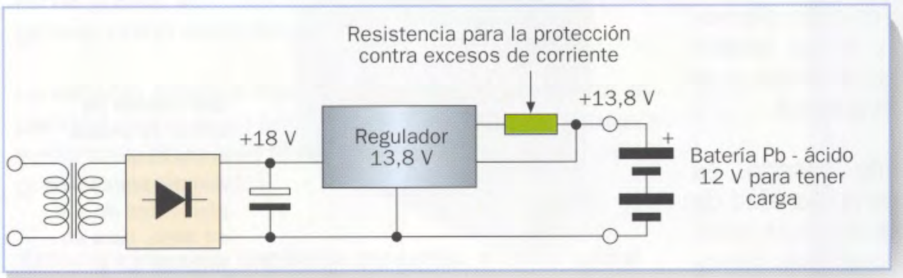
Resistencia para la protección contra excesos de corriente
+18 V
Regulador 13,8 V
+13,8 V
+
Batería Pb - ácido 12 V para tener carga

La figura muestra un ejemplo de circuito para recarga con tampón, en la práctica un **alimentador estabilizado**, que incluye un limitador de corriente para evitar daños.

Cargador de baterías de tampón de tensión constante, adecuado para baterías de plomo-ácido.

CARGA CÍCLICA

Las baterías de NiCd o NiMH se recargan solamente cuando están **descargadas** casi del todo, con el fin de mantenerlas en las mejores condiciones y obtener la máxima vida útil.

La recarga se produce normalmente **a corriente constante**, en general en torno a una décima de la capacidad de la batería; se puede lograr una cierta mejora utilizando impulsos en lugar de corriente continua.

El circuito debe apagarse, o pasar a una corriente débil de mantenimiento, cuando **finalice la carga**, para evitar calentamientos con posible expulsión de sustancias nocivas.

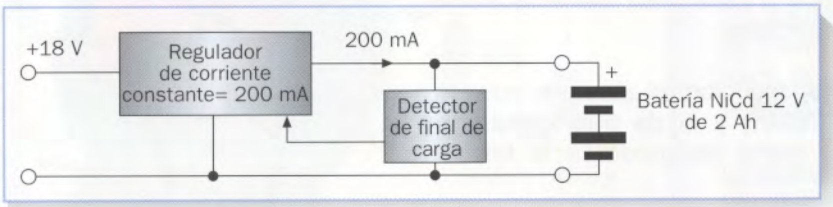
+18 V
Regulador de corriente constante= 200 mA
200 mA
Detector de final de carga
+
Batería NiCd 12 V de 2 Ah

***Cargador de baterías a corriente constante**: si falta el detector de final de carga, es necesario acordarse de apagarlo.*

VIDA DE LAS BATERÍAS

Las baterías recargables no son eternas: tras un cierto número de **ciclos de carga/descarga**, la capacidad disminuye, hasta hacerse insuficiente (o bien simplemente la batería deja de funcionar).

En todo caso, es importante **no tirar** pilas o acumuladores a la basura (o, peor todavía, al fuego), sino utilizar los contenedores especiales para su recogida, para evitar daños a la salud y al entorno.

No todos los acumuladores "prefieren" ser **descargados** a fondo antes de cada recarga; algunos tienen una mayor duración si se recargan cuando todavía tienen mucha capacidad residual.

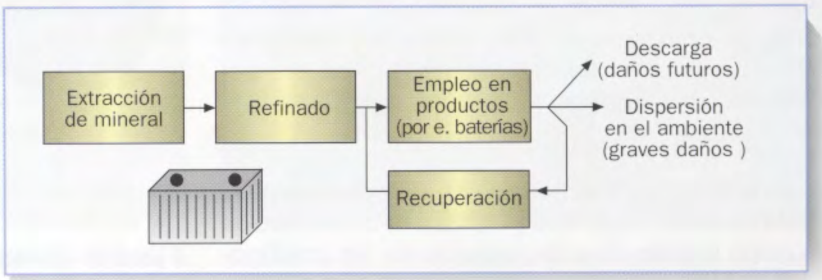
Extracción de mineral
Refinado
Empleo en productos (por e. baterías)
Descarga (daños futuros)
Dispersión en el ambiente (graves daños)
Recuperación

***La recogida de las baterías** de coches reduce la contaminación debida a este metal, bastante nocivo para el organismo humano.*

Amplificación de audio

Varios factores contribuyen a determinar la calidad del sonido

La calidad de reproducción de una instalación de audio depende sobre todo del **eslabón más débil** de la cadena, es decir, las cajas acústicas (ver lección 17 de Aplicaciones).

Naturalmente, también la electrónica tiene su importancia si se desea alcanzar la **alta fidelidad** (HIFI: High Fidelity) es decir, producir de manera ideal un sonido totalmente indistinguible del original.

El primer requisito es la **anchura de banda**: para no alterar el timbre de los sonidos, la amplificación debe ser uniforme (un buen amplificador está dentro de los ±0,5 dB o menos) entre al menos 20 Hz y 20 KHz.

DISTORSIÓN

Se habla de **distorsión armónica** cuando una forma de onda sinusoidal es alterada, produciendo armónicos a frecuencia múltiplo de la fundamental.

Ésta varía normalmente con la frecuencia y, sobre todo, con la **amplitud** de la señal aplicada: por encima de cierto nivel se obtiene un recorte de los picos de la señal.

Otro efecto de la distorsión es la **intermodulación**: una amplificador no lineal funciona como mezclador, generando desagradables sonidos (frecuencias suma y diferencia).

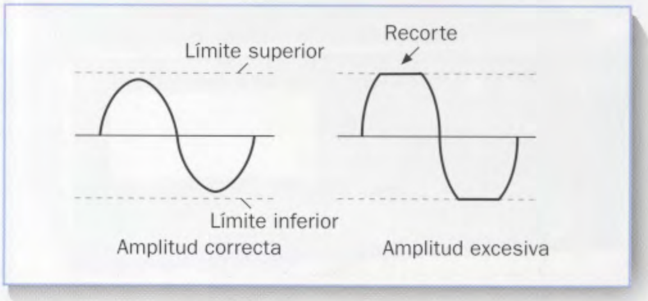
Límite superior
Recorte
Límite inferior
Amplitud correcta
Amplitud excesiva

Cuando la amplitud de la señal se aproxima a los límites del amplificador, primero se distorsiona y después se corta.

RUIDO

Sólo basta que una señal de audio atraviese el amplificador no alterado: es necesario también que no se le añadan señales no deseadas, es decir, **interferencias** de distinto tipo, o "ruido".

Una forma de interferencia más engañosa es la que no procede del exterior, sino que es generada internamente por los propios componentes, y después amplificada: se trata del **ruido intrínseco**.

Esto ocurre sobre todo en las primeras etapas, las más sensibles y sujetas a captar **interferencias**: campos eléctricos, magnéticos o incluso ondas de radio, si el proyecto no está suficientemente protegido.

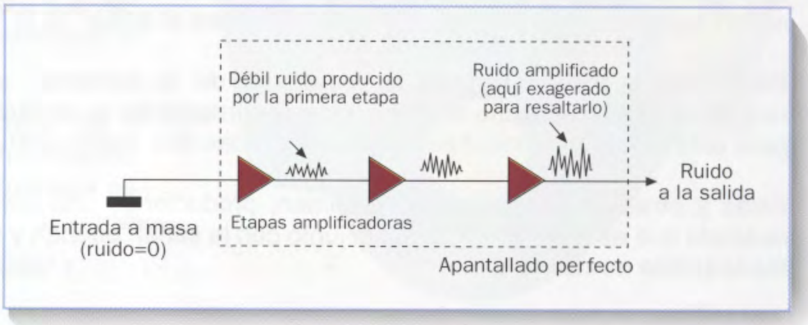
Débil ruido producido por la primera etapa
Ruido amplificado (aquí exagerado para resaltarlo)
Ruido a la salida
Entrada a masa (ruido=0)
Etapas amplificadoras
Apantallado perfecto

Un amplificador totalmente protegido produce en cualquier caso un zumbido a la salida, que proviene del ruido eléctrico de los componentes.

Tipos de ruido

Resistencias y semiconductores son fuentes de tensiones de ruido

Incluso una sencilla **resistencia** no alimentada produce un ruido eléctrico (ruido de Johnson), por efecto de la agitación térmica de los electrones en su interior.

Se distribuye uniformemente por todas las frecuencias y se llama **ruido blanco** (white noise): su amplitud depende de la banda de paso sobre la que se mide, además de la resistencia y la temperatura.

Cualquier resistencia produce una tensión de ruido, en sus extremos, que depende sólo del valor y aumenta con la temperatura.

Se da normalmente en **µV/√Hz**; por ejemplo, un ruido de 0,1 µV/√Hz medido en una banda de 10.000 Hz será de: 0,1 µV x 100 = 10 µV rms.

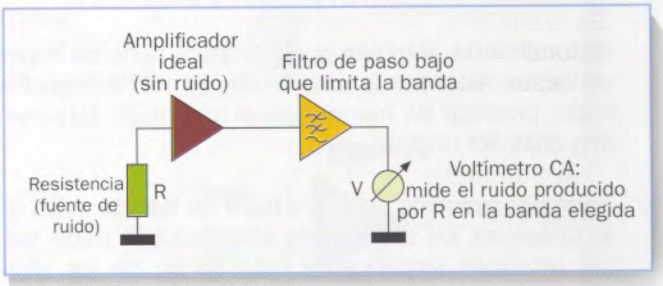
Amplificador ideal (sin ruido)
Filtro de paso bajo que limita la banda
Resistencia (fuente de ruido)
R
V
Voltímetro CA: mide el ruido producido por R en la banda elegida

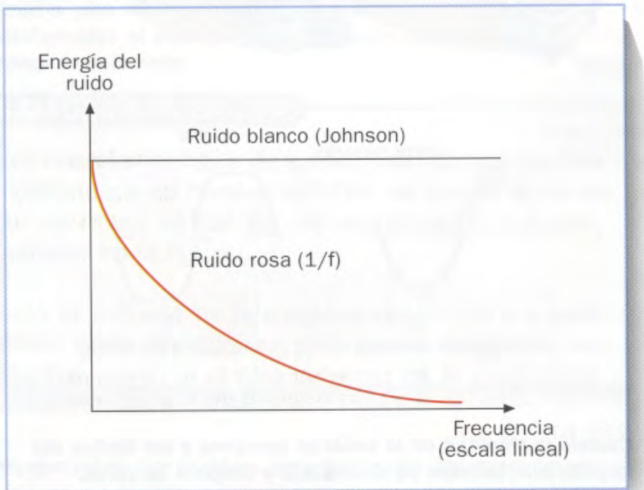

La cantidad de energía en el ruido blanco es constante a paridad de anchura de banda, la del ruido rosa baja con la frecuencia.

RUIDO 1/F

Haciendo pasar **corriente** por la misma resistencia, se añade un segundo tipo de ruido, que depende del tipo de construcción de la misma resistencia (y está presente en muchos otros componentes).

Se trata del **ruido rosa** (pink noise), o ruido "1/f", así llamado porque su energía decrece con la frecuencia: tiene similar energía por cada década de frecuencia.

Se puede indicar en **µV/V**, es decir, microvoltios por voltio aplicado (sobre una década de frecuencia), siempre rms: el ruido es de hecho de naturaleza casual, y se puede medir sólo estadísticamente.

RUIDO IMPULSIVO

Otro tipo de ruido, característico de los **semiconductores**, es el "shot noise", más molesto para el oído ya que se parece al gotear de la lluvia.

Es debido a **irregularidades estadísticas** de la corriente, a su vez relacionadas con el hecho de que la propia corriente no es un flujo continuo pero está formada de muchas cargas separadas (los electrones).

Estas y otras causas de ruido se suman, produciendo una señal no deseada que es amplificada también junto con la señal "buena", y se hace **inseparable** de esta última.

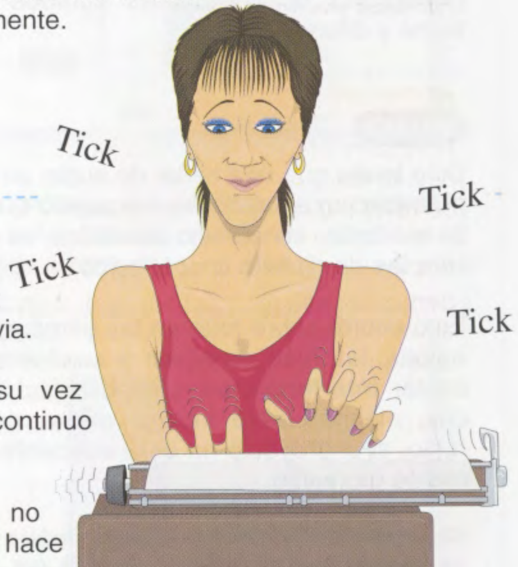

Ruido y dinámica

La máxima variación de amplitud posible es limitada

En un amplificador, el ruido se produce **en varios puntos**, desde las resistencias del circuito de entrada y las uniones de los semiconductores de las primeras etapas, hasta los transistores finales.

No obstante, las primeras son bastante más importantes, ya que el ruido por ellas generado es después **amplificado** por las etapas sucesivas, mientras que (por ejemplo) el de las finales suele ser superfluo.

El ruido se mide en la salida, pero se hace **referencia a él a la entrada**: por ejemplo, 1 mV a la salida de un amplificador con ganancia 10.000 corresponde a un ruido a la entrada de 0,1 mV.

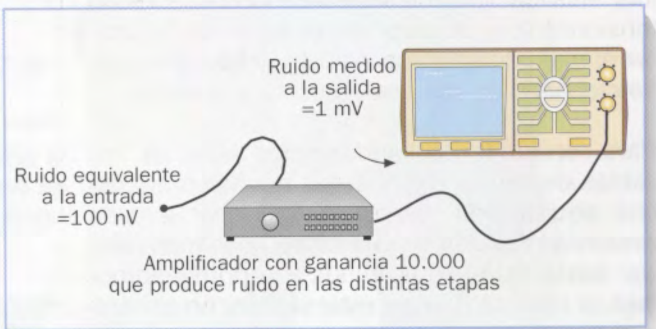
Ruido medido a la salida =1 mV
Ruido equivalente a la entrada =100 nV
Amplificador con ganancia 10.000 que produce ruido en las distintas etapas

Por comodidad, se supone un amplificador perfecto, al que se aplique a la entrada un ruido equivalente.

RELACIÓN S/N

Lo que normalmente importa, para la **percepción** acústica, no es la amplitud del ruido en sí, sino la relación existente entre la señal deseada y la no deseada (el ruido).

Un ruido de un microvoltio no será audible en una señal de 10 V, mientras que **molestará** bastante una señal de 10 μV, ya que será después amplificado junto a ella.

La relación **señal/interferencia** (S/N: signal/noise) indica, en dB, el porcentaje del ruido con respecto a la señal; el resultado depende también de la anchura de banda de la propia señal.

SINAD METER TYPE R 310 A
AF AMPL
ON VOLTS SINAD
VOLUME
FS VOLT
OUTPUT
INPUT
UNAOHM

Una sencilla herramienta para medir la relación señal/interferencia, incluida la distorsión.

DINÁMICA

Hemos visto que la amplitud **máxima** de la señal está limitada por la distorsión que se considera aceptable; la **mínima** está, sin embargo, limitada por el ruido, que ocultaría la propia señal.

La relación entre los dos límites es el **campo dinámico** o "dinámica", es decir, la máxima variación permitida a la señal, quedando íntegra y no distorsionada ni cubierta por el ruido.

Por ejemplo, si la amplitud máxima es de 10 V rms, y el ruido de fondo es de 1 mV rms (no profundizamos aquí en las condiciones de medida), la **dinámica** es de 10.000, es decir, de 80 dB.

LUDWIG VAN BEETHOVEN
HERBERT VON KARAJAN · LAZAR BERMAN

La reproducción de música clásica, con sus "fortissimo" y "pianissimo" requiere una dinámica muy elevada.

Circuitos de audio

En los preamplificadores se emplean varias soluciones circuitales

Las entradas para **micrófono**, que amplifican tensiones muy débiles, tienen especial tendencia al ruido: la amplitud de este último será proporcionalmente alta con respecto a la señal.

Para captar menos interferencias externas, los cables de los micrófonos son normalmente del tipo **equilibrado**, es decir, con dos señales simétricas con respecto a masa, y de impedancia bastante baja (600 Ω). El problema es menos notable con los **micrófonos amplificados** modernos, cuya primera etapa de amplificación (la más delicada) está situada directamente en el interior del mismo micrófono.

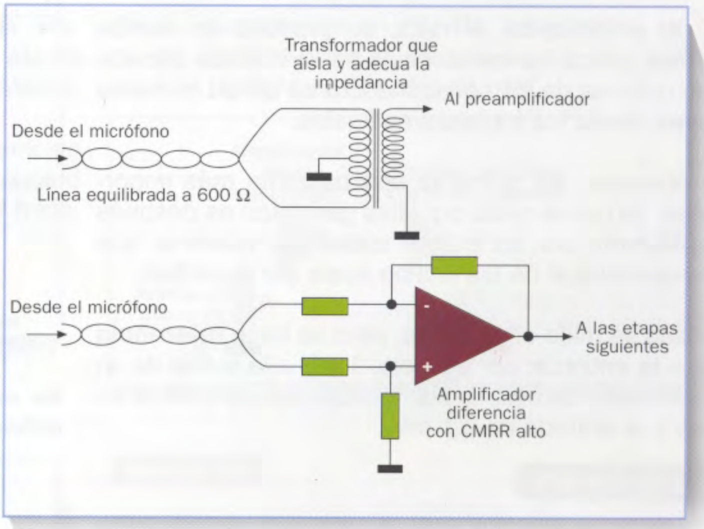
Transformador que aísla y adecua la impedancia
Al preamplificador
Desde el micrófono
Línea equilibrada a 600 Ω
Desde el micrófono
-
+
A las etapas siguientes
Amplificador diferencia con CMRR alto

***Entradas para micrófono:** el primero utiliza un transformador, el segundo un amplificador diferencial con op-amp de bajo ruido.*

SUMADOR CON OP-AMP

Para **mezclar** señales de audio, una vez que se han amplificado a un nivel aceptable, se puede aprovechar la masa virtual de un amplificador inversor, mostrado en la figura.

Como la entrada negativa se tiene siempre a cero voltios, cada entrada ve **solamente** su propia resistencia, como si el otro extremo de la resistencia misma estuviera conectada a masa.

Las distintas corrientes, sin embargo, se **suman** en la resistencia de realimentación Rf: el operacional produce, por tanto, una tensión a la salida proporcional a la suma de las señales de entrada.

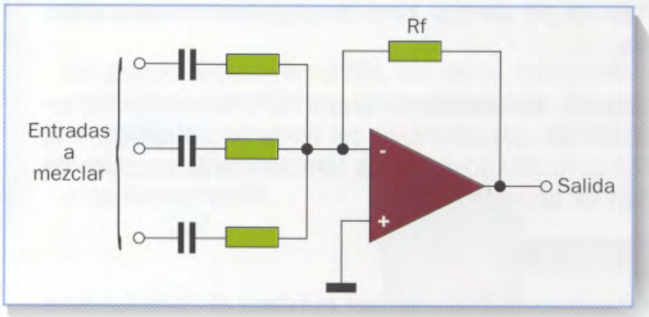
Rf
Entradas a mezclar
-
+
Salida

***Mezclador con masa virtual:** funciona como el sumador descrito en la lección 37, pero la ganancia en continua es **1**.*

AUDIO DIGITAL

Un enfoque totalmente distinto consiste en amplificador las señales a un nivel adecuado (para no añadir problemas de ruido) y después **digitalizarlos** con un ADC.

Los posteriores **procesos**, por ejemplo la mezcla de señales, serán después desarrollados por una CPU (o, más comúnmente, por un DSP), trabajan siempre en forma digital.

Estas técnicas digitales se van difundiendo, gracias a la versatilidad de proceso, y también a la posibilidad de **registrar** la señal infinitas veces, sin añadirles degradación o ruido a cada paso.

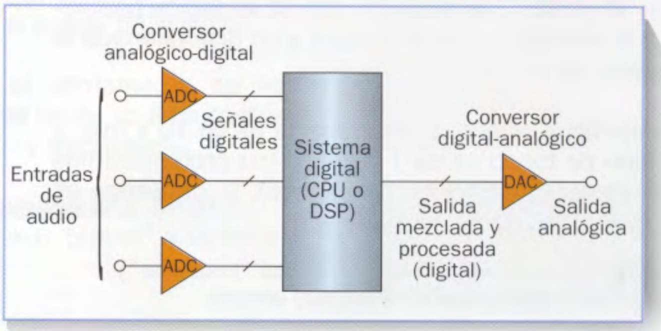
Conversor analógico-digital
ADC
ADC
ADC
Señales digitales
Entradas de audio
Sistema digital (CPU o DSP)
Conversor digital-analógico
DAC
Salida mezclada y procesada (digital)
Salida analógica

Para la mezcla digital de varias señales se ejecuta la suma aritmética, instante por instante, de los respectivos valores muestreados.

Filtros activos

Los operacionales permiten eliminar los incómodos inductores

El proyecto de filtros con tan sólo **componentes pasivos**, como los ilustrados en la lección 17, supone una operación bastante delicada y no siempre conveniente.

Necesitan, de hecho, **inductores** que ocupan espacio en el circuito (especialmente si son de valor alto), son costosos, captan interferencias magnéticas y mecánicas, y resultan difíciles de realizar con precisión.

Además, un filtro de orden elevado (es decir, muy sensible) requiere un gran número de componentes y es terriblemente sensible a las mínimas **tolerancias** de sus valores.

CONVERSIÓN DE IMPEDANCIA

Recordemos de la lección 6 que el comportamiento de inductores y condensadores es, en muchos aspectos, **recíproco**: si se logra "reinvertir" un condensador, se tendría un inductor.

Por ejemplo, en el circuito de la figura (conversor de impedancia negativa) la corriente a la entrada está **retrasada** con respecto a la tensión, como en un inductor, a pesar de que haya sólo un condensador.

No se comporta, sin embargo, como un inductor verdadero, ya que su conducta en **frecuencia** se queda como el del condensador C: a paridad de tensión, la corriente aumenta con el incremento de la frecuencia.

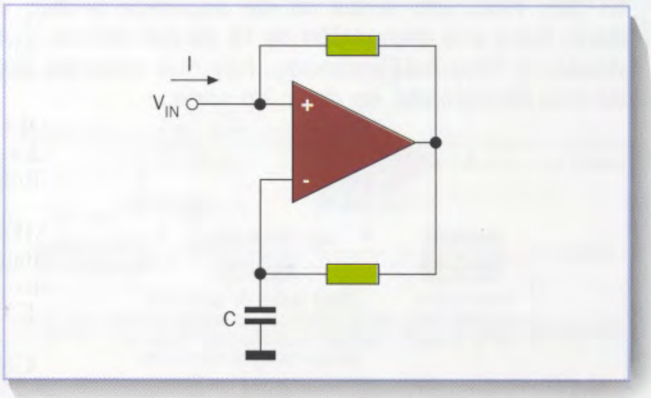

En este circuito la corriente a la entrada I responde con retraso a las variaciones de la tensión aplicada V_{IN}, como en los inductores.

INDUCTORES SIMULADOS

El circuito mostrado en la figura de al lado, también construido en torno a un solo amplificador operacional, es sin embargo un verdadero inversor de impedancia, o **girador**.

El circuito lo ve como un **inductor** auténtico y puede, por tanto, utilizarlo para la construcción de filtros, además de para obtener inductancias simuladas de elevadísimo valor. El inductor así obtenido **no es perfecto**: las resistencias empeoran Q y el comportamiento en alta frecuencia se ve afectado por los límites del integrado; no obstante, puede resultar útil en muchos casos.

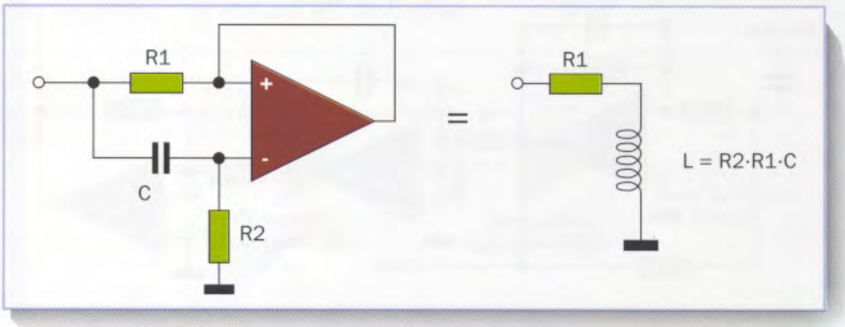

Un girador permite invertir la impedancia del condensador, ofreciendo el mismo comportamiento de un inductor.

Tipos de filtro activos

Algunas configuraciones circuitales son bastante utilizadas

Con tan sólo condensadores y amplificadores operacionales, es posible realizar **filtros activos**, es decir, filtros LC sin la L, ya que los inductores se simulan con giradores o con trucos similares.

Existen distintas soluciones circuitales, por ejemplo, la que toma el nombre de VCVS (Voltaged-Controlled Voltage Source: fuente de voltaje controlada por voltaje), utilizada en el filtro de **paso bajo** mostrado en la figura.

Cambiando oportunamente los valores de los componentes, es posible **optimizar** distintas características, como por ejemplo, la inclinación o linealidad, como para los filtros pasivos (ver lección 17).

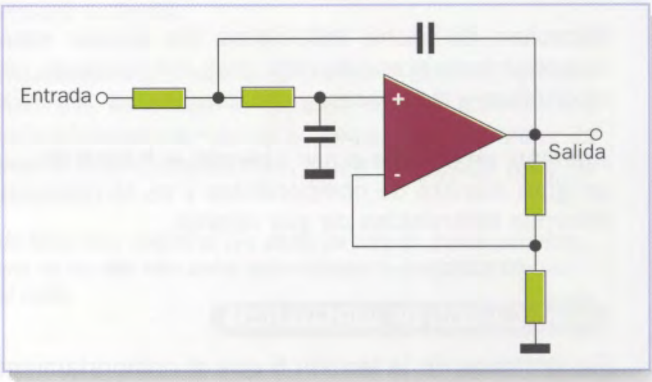

Filtro activo *de paso bajo de tipo VCVS: la realimentación elimina la necesidad de inductores realizados con hilo enrollado.*

FILTROS EN CASCADA

El filtro mostrado arriba es **de segundo orden**, es decir, tiene una atenuación de 12 dB por octava; si se desea un filtro más inclinado, hay que conectar más de uno en cascada, es decir, en serie.

A diferencia de lo que ocurre con los filtros pasivos, no sirven los buffer: la salida de un filtro activo es normalmente de **baja impedancia**, y resulta adecuada para controlar la entrada de la siguiente etapa.

El **cálculo** de un filtro de varias etapas es, sin embargo, bastante complicado; afortunadamente, como siempre en estos casos, están disponibles fórmulas ya preparadas para evitar complejos procesos matemáticos.

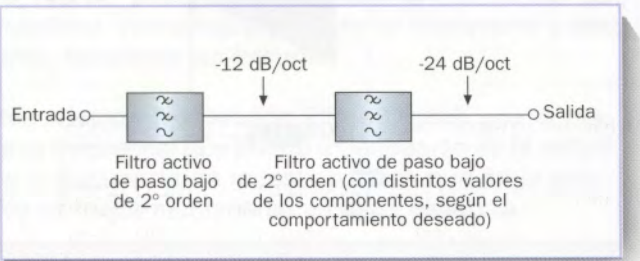
-12 dB/oct
-24 dB/oct
Entrada
Salida
Filtro activo de paso bajo de 2° orden
Filtro activo de paso bajo de 2° orden (con distintos valores de los componentes, según el comportamiento deseado)

Filtro de paso alto de cuarto orden (24 dB/octava) realizado con dos filtros de segundo orden conectados en cascada.

FILTROS MÁS COMPLEJOS

El filtro de **paso de banda** mostrado en la figura de abajo es del tipo llamado "bi-cuad"; como puede verse, está formado por dos integrados y un amplificador inversor. Ofrece la posibilidad, eligiendo los valores de los componentes, de configurar **por separado** la frecuencia central (máxima amplificación) y la banda de paso, es decir, elegir el Q deseado.

Hay **otros tipos** de filtros activos, más o menos complejos y adecuados a satisfacer distintos requisitos; las prestaciones aumentadas de los op-amp los dejan utilizables a frecuencias siempre más elevadas.

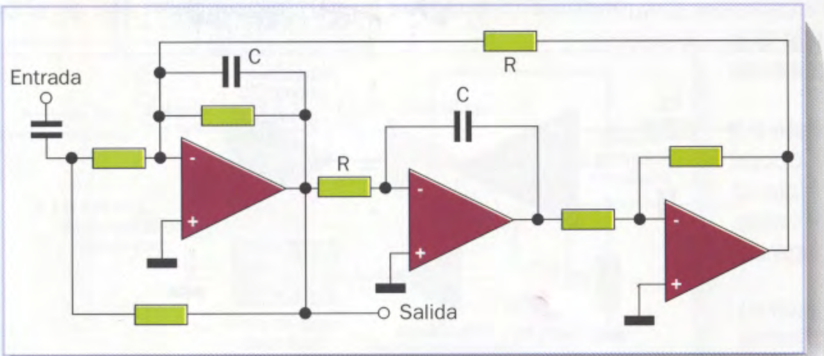

Filtro de paso de banda del tipo* bi-cuad*, realizado con tres amplificadores operacionales: R y C establecen la frecuencia de trabajo.

Filtros programables

Con pocas modificaciones, los filtros activos pueden controlarse eléctricamente

La **frecuencia** de corte (o la frecuencia central para los de paso de banda) depende del valor asignado a resistencias o condensadores: variándolo, se puede cambiar a voluntad, al menos dentro de ciertos límites.

Se puede controlar también eléctricamente, por ejemplo utilizando un JFET en lugar de una resistencia, que funcione como **resistencia variable** controlada por una tensión externa, como se describe en la lección 23.

La figura muestra un integrador **controlado en tensión**: con dos de estos, siempre que estén equilibrados, se puede, por ejemplo, controlar la frecuencia central de un filtro bi-cuad, y por lo tanto "ajustarlo".

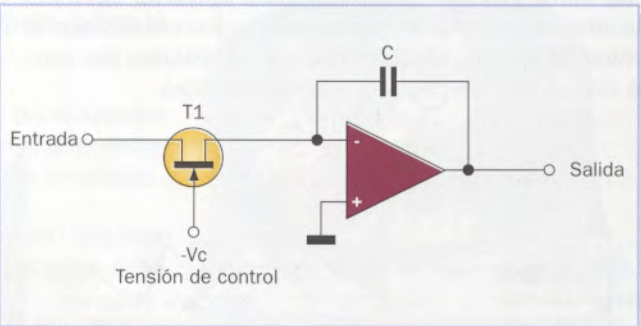

Integrador variable: ***la tensión de control determina el valor de resistencia del JFET, y por lo tanto la velocidad de integración.***

CAPACIDADES CONMUTADAS

El control en tensión es sin embargo **poco previsible**, estable y lineal; la señal no puede ser amplia; además, es difícil que dos JFET se comporten exactamente del mismo modo.

Se muestra en la figura un enfoque totalmente distinto: el pequeño condensador C1 es **cargado** por la tensión de entrada, y después rápidamente descargado por la otra parte, cargando parcialmente C2.

Efectuando esta operación a frecuencias mucho más elevadas de la señal, se puede **controlar** la velocidad de carga de C2 simplemente variando la frecuencia de la señal de control.

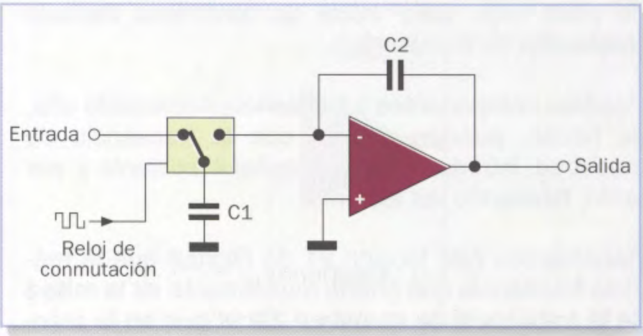

***Integrador de capacidad conmutada:** en cada ciclo del reloj de control, C1 transfiere un poco de carga a C2.*

FILTROS DIGITALES

La solución arriba descrita permite obtener un filtro de control digital, pero se puede dar otro paso en esta dirección y **eliminar** del todo los integrados analógicos.

Se puede **convertir** la señal en digital con un ADC, después tratarla con un microprocesador para ejecutar el filtro, y finalmente (si es necesario) reconvertirla en analógica mediante un DAC.

Según esta técnica, cada vez más difundida debido a los costes a la baja, se utilizan únicamente microprocesadores especializados: los **DSP** (Digital Signal Processor: procesador de señal digital), ilustrados en la lección 48 de Digital.

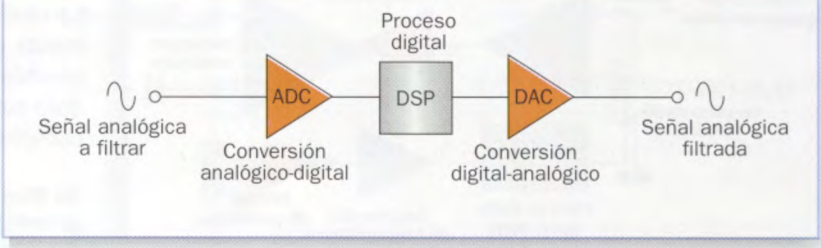
Proceso digital
ADC
DSP
DAC
Señal analógica a filtrar
Conversión analógico-digital
Conversión digital-analógico
Señal analógica filtrada

Filtro activo digital: ***el microprocesador especializado (DSP) funciona según números que representan la señal muestreada.***

Aplicaciones de los filtros

El filtrado no sirve solamente para modificar el timbre de la música

Los filtros pueden utilizarse para **eliminar** señales no deseadas, o bien para **resaltar** las deseadas, a pesar de las posibles interferencias creadas por parte de los demás componentes presentes.

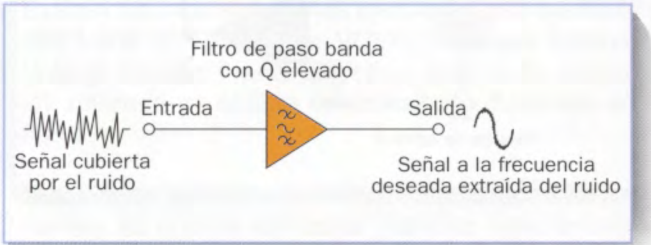
Filtro de paso banda
con Q elevado
Entrada
Salida
Señal cubierta
por el ruido
Señal a la frecuencia
deseada extraída del ruido

Por ejemplo, una nota a 2.400 Hz aparentemente enmascarada por el ruido, puede **aislarse** utilizando un filtro de paso de banda antes del amplificador, como en la figura.

Cuanto más alto sea el Q del filtro (es decir, cuanto más estrecha sea la banda), mayor resultará la **selectividad**, es decir, la capacidad de aislar la frecuencia deseada del ruido de fondo que la circunda.

Un filtro muy selectivo (es decir, de alto Q) puede permitir que se resalte una señal aparentemente enmascarada por el ruido.

CONVERSIÓN Y FILTROS

Cuando se efectúa una conversión analógico-digital, es importante filtrar la señal de entrada con un filtro de paso bajo, para evitar un fenómeno llamado **traslación** de frecuencia.

Posibles componentes a frecuencia demasiado alta, de hecho, pueden **chocar** con la frecuencia de muestreo, introduciendo una señal inexistente y, por tanto, falseando las lecturas.

Recordemos (ver lección 21 de Digital) que la máxima frecuencia que puede muestrearse es **la mitad** de la frecuencia de muestreo, de la que en la práctica es mejor mantenerse a buena distancia.

El filtro anti-traslación de un ADC evita un problema similar al de la frecuencia de imágenes en los receptores (ver lección 26).

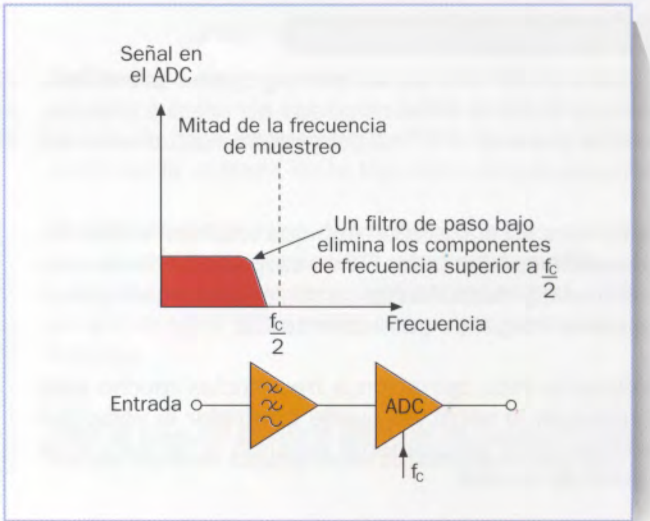
Señal en
el ADC
Mitad de la frecuencia
de muestreo
Un filtro de paso bajo
elimina los componentes
de frecuencia superior a $\frac{f_C}{2}$
$\frac{f_C}{2}$
Frecuencia
Entrada
ADC
f_C

REALIMENTACIÓN DE CONTROL

En muchos sistemas con realimentación, de un amplificador de audio al control hidráulico de la posición de corte de una herramienta, hay un riesgo de **inestabilidad**. Los distintos retardos eléctricos o mecánicos pueden causar fenómenos de **oscilación**, más o menos amortiguados pero en todo caso no deseables, ya que alteran la forma de onda o el movimiento deseados.

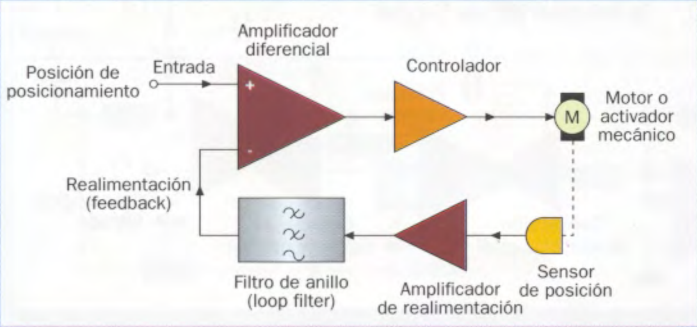
Amplificador
diferencial
Posición de
posicionamiento
Entrada
Controlador
Motor o
activador
mecánico
M
Realimentación
(feedback)
Filtro de anillo
(loop filter)
Amplificador
de realimentación
Sensor
de posición

La realimentación se debe **filtrar**, por tanto, de modo que se obtenga la máxima velocidad posible de respuesta del sistema, manteniéndolo sin embargo estable (el tema es bastante complejo, y se ha estudiado mucho).

Un filtro (adecuadamente calculado) permite garantizar la estabilidad en un sistema de anillo cerrado, es decir, con realimentación negativa.

Osciladores con op-amp

Los operacionales simplifican el circuito y mejoran sus características

La figura muestra el clásico oscilador de **trigger Schmitt** (ver lección 17 de Digital), pero realizado con un amplificador operacional en lugar de integrados digitales.

Las resistencias R2 y R3 introducen una **realimentación positiva** (ver lección 38), por lo que el op-amp se comporta en la práctica como un comparador con histéresis, es decir, un trigger Schmitt.

Este oscilador produce a la salida una **onda cuadrada**, mientras que el condensador C es cargado y descargado alternativamente por R1, según la salida sea positiva o negativa.

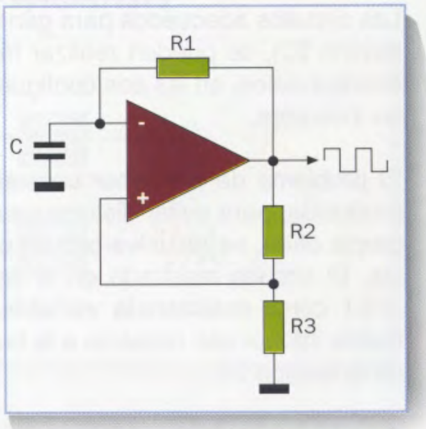

Oscilador *de trigger Schmitt: cuando la salida es alta, C se carga hasta alcanzar el umbral, y viceversa.*

FORMAS DE ONDA

El **funcionamiento** se ilustra con detalle en la figura: cuando la salida es alta, el condensador empieza a cargarse con la típica curva (inicialmente rápida, después cada vez más lenta).

Nada más lleva la carga la entrada negativa a **alcanzar** la positiva, la altísima amplificación del operacional hace que la salida cambie rápidamente de estado, haciéndose negativa.

No sólo el nivel de salida nuevo empieza a descargar C sino que, mediante R2, **desplaza** el nivel de umbral de modo que se establezca el límite inferior de la descarga; alcanzado este último, se iniciará un nuevo ciclo de carga.

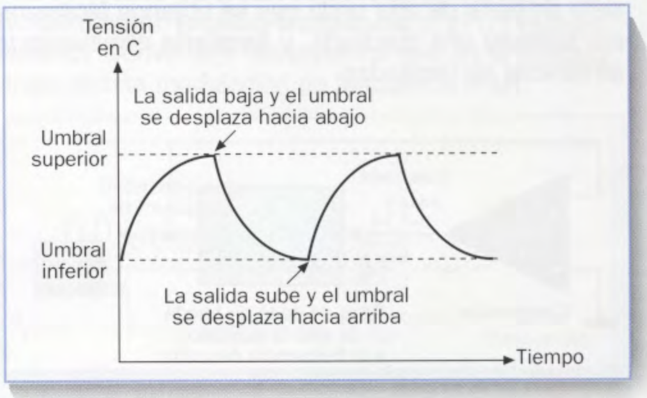

Cuando la salida cambia de estado, el* nivel de comparación *(entrada positiva) es desplazado por la parte opuesta.

ONDAS TRIANGULARES

Si se desea una salida triangular en lugar de cuadrada, es necesario cargar el condensador **a corriente constante**; esto se puede hacer utilizando el integrador descrito en la lección 37.

En el circuito mostrado al lado, el primer operacional funciona (como antes) como comparador con histéresis, mientras que el segundo es un **integrador**: sustituye el grupo RC del circuito de arriba.

El funcionamiento es similar al de antes, pero la salida del integrador sube **linealmente**; nada más alcanza el umbral, el comprador cambia de estado y el proceso se repite en el otro sentido.

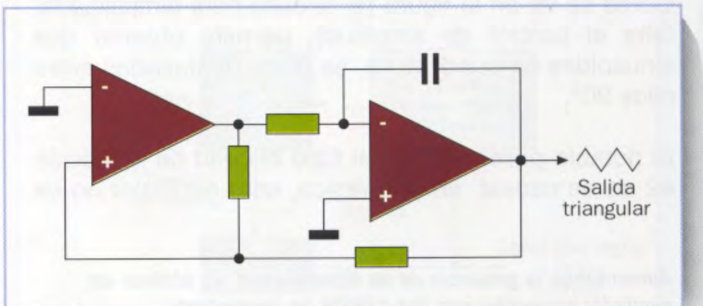

Oscilador con salida* triangular*: el segundo operacional carga C a corriente constante, por lo tanto su tensión sube linealmente.

Osciladores sinusoidales

Una baja distorsión es siempre deseable

Los circuitos adecuados para generar una **sinusoide** pura (ver lección 20), se pueden realizar fácilmente con amplificadores operacionales, en las dos configuraciones clásicas: inversora y no inversora.

El problema de mantener constante la **amplitud** de la onda producida, para evitar distorsiones que alteren la pureza de la propia onda, se resuelve con un control automático de ganancia. El circuito mostrado en la figura, por ejemplo, utiliza un JFET como **resistencia variable**, con un consumo notablemente inferior con respecto a la bombilla utilizada en el circuito de la lección 20.

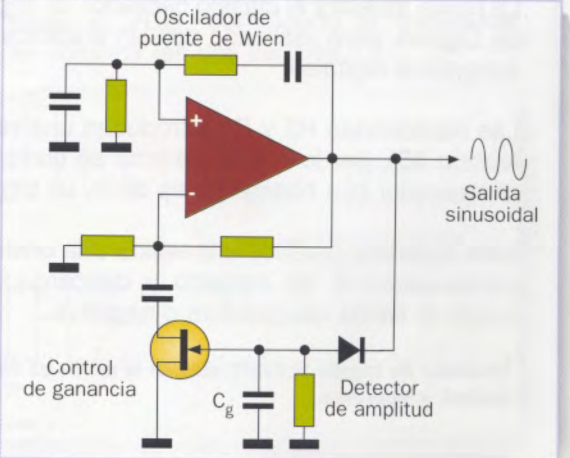

Oscilador sinusoidal con estabilizador de amplitud: la salida redirigida carga C_g, aumentando la resistencia del JFET y reduciendo así la ganancia.

SINUSOIDE DE UNA CUADRADA

Una técnica diferente para tener sinusoides puras consiste en partir de una onda que se obtenga fácilmente, por ejemplo una cuadrada, y **limpiarla** eliminando las armónicas no deseadas.

Esto se puede hacer con un filtro de **paso de banda** ajustado a la frecuencia deseada: si el Q es bastante alto, en su salida se encontrará prácticamente sólo la fundamental, es decir, una sinusoide.

La onda cuadrada puede producirse ¡**descuadrando** la propia sinusoide! La figura muestra cómo una conexión de anillo entre un comparador y un filtro forma un oscilador sinusoidal.

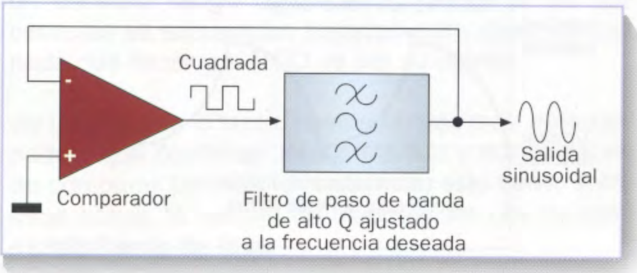

Oscilador sinusoidal: ***la onda cuadrada probada de los armónicos es la sinusoide deseada, que a su vez produce la cuadrada.***

OSCILADOR BI-CUAD

Otro sistema más consiste en utilizar **dos integradores**, cada uno de los cuales debe producir un desfase igual exactamente a 90 grados a la frecuencia de oscilación deseada.

Como se ve en la figura (el circuito está simplificado, falta el control de amplitud), permite obtener dos sinusoides en **cuadratura**, es decir, desfasados entre ellos 90°.

El notable **parecido** con el filtro bi-cuad de la lección 42 no es causal: en la práctica, este oscilador no es otra cosa que un filtro de paso de banda con la ganancia demasiado alta.

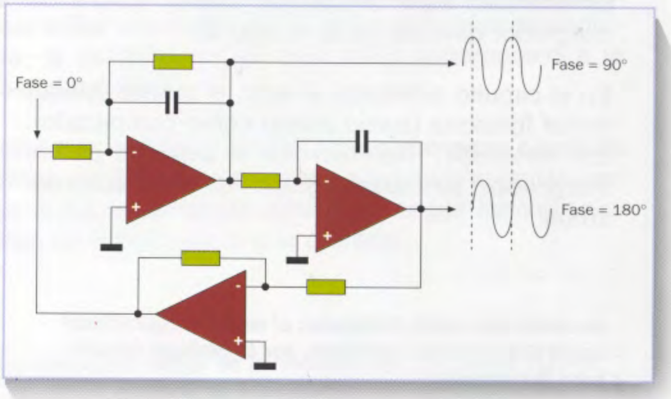

Aumentando la ganancia de un filtro bi-cuad, se obtiene un oscilador sinusoidal con dos salidas en cuadratura.

VCO y PLL

Los osciladores controlados en tensión tienen diversas aplicaciones

Casi todos los osciladores pueden modificarse fácilmente, de modo que se controle su **frecuencia** de oscilación con una **tensión** aplicada externamente.

Por ejemplo, se pueden utilizar las mismas **técnicas de control** descritas en la lección 42 a propósito de los filtros activos (en alta frecuencia se utilizan otros sistemas, ver lección 25).

Un circuito similar se denomina VCO (Voltage-Controlled Oscillator: oscilador controlado por tensión); una característica suya interesante es una buena **linealidad**, es decir, proporcionalidad entre tensión aplicada y frecuencia obtenida.

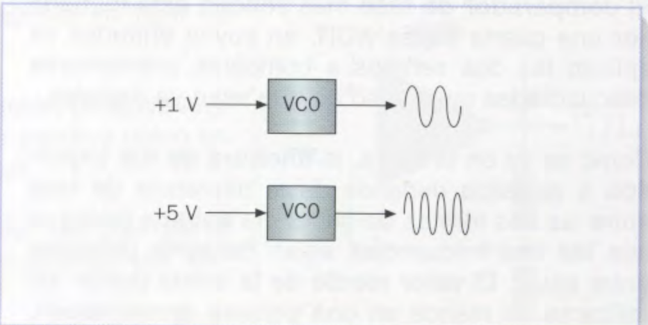
+1 V
VCO
+5 V
VCO

Un* VCO *es un oscilador cuya frecuencia depende de la tensión aplicada a la entrada de control.

USOS DE LOS VCO

Una primera aplicación de los VCO consiste en **codificar** una señal digital con dos notas distintas, de modo que se pueda transmitir a distancia (FSK, ver lección 20 de Digital).

Un VCO utilizado para este objetivo se denomina a menudo **conversor tensión-frecuencia** o VFC (Voltage-Frequency Converter): observemos que en la práctica se trata de una modulación de frecuencia (FM).

La misma técnica permite **transmitir** una señal analógica, es decir, una tensión variable, en forma de frecuencia, de modo que esté mucho menos sujeta a posibles interferencias de línea.

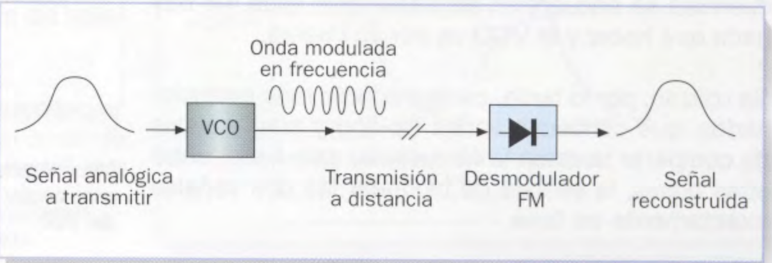
Onda modulada en frecuencia
VCO
Señal analógica a transmitir
Transmisión a distancia
Desmodulador FM
Señal reconstruída

Convirtiendo una tensión en una frecuencia, es más fácil enviar una señal analógica a largas distancias, o por radio.

BUCLE DE FASE BLOQUEADA

Una de las principales aplicaciones de los VCO es el anillo de bloqueo de fase o **PLL** (Phase-Locked Loop), en el que el oscilador se controla de modo que sigue una frecuencia externa.

Como se ve en la figura, la señal producida por el VCO se compara con la señal de entrada, modificando la tensión de control hasta alcanzar el **bloqueo**: el VCO marca el final.

Los PLL tienen numerosos usos, especialmente en el campo de la **radio**, por ejemplo en la desmodulación FM, en la síntesis de frecuencia (ver lección 26) y en la reconstrucción de la subportadora de televisores en color.

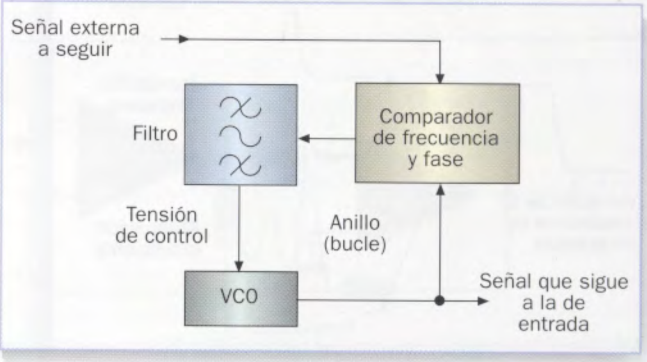
Señal externa a seguir
Filtro
Comparador de frecuencia y fase
Tensión de control
Anillo (bucle)
VCO
Señal que sigue a la de entrada

Un* PLL *hace que un oscilador controlado (*VCO*) siga fielmente una señal procedente del exterior.

Bloqueo de un PLL

Comparador y filtro son los componentes más importantes del circuito

El **comparador de fase** más sencillo está formado por una puerta lógica XOR, en cuyas entradas se aplican las dos señales a comparar, previamente descuadradas en el caso de que sean ya digitales.

Como se ve en la figura, la **anchura de los impulsos** a la salida depende de la diferencia de fase entre las dos formas de onda a la entrada (siempre que las dos frecuencias estén bastante próximas entre ellas). El **valor medio** de la salida puede así utilizarse, al menos en una primera aproximación, para controlar el VCO, corrigiendo la frecuencia hasta hacer coincidir las dos señales.

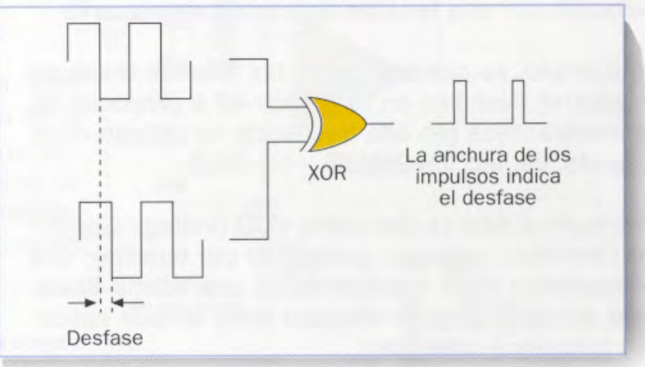
XOR
La anchura de los impulsos indica el desfase
Desfase

Comparador de fase ***realizado con un OR exclusivo, o XOR: el valor medio de la salida depende del desfase entre las señales.***

COMPARACIÓN DE FRECUENCIA

La comparación de fase tiene sentido si las frecuencias están muy **próximas**, por lo que bastan mínimos retoques en el VCO para desplazar su señal hacia delante o hacia atrás, hasta obtener el bloqueo deseado. La técnica arriba descrita tiene, sin embargo, un estrecho **rango de captura**, ya que si las frecuencias se encuentran alejadas entre ellas, no hay nada que hacer y el VCO va por su cuenta.

Se utilizan, por lo tanto, comparadores más evolucionados, que contienen varios flip-flop y son capaces de comparar también la frecuencia; esto tiene, entre otras cosas, la ventaja de bloquear las dos señales **exactamente en fase**.

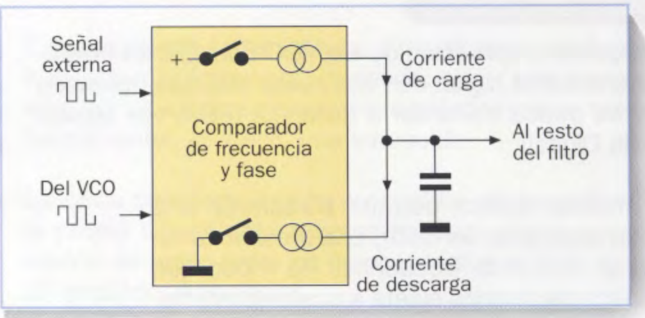
Señal externa
Del VCO
+
Comparador de frecuencia y fase
Corriente de carga
Al resto del filtro
Corriente de descarga

Los comparadores evolucionados utilizan una* bomba de carga *para cargar un condensador y obtener la tensión de control del VCO.

FILTRO DE ANILLO

El proyecto de un PLL es bastante delicado, porque el filtro necesario a la salida del comparador (filtro de bucle) puede causar fácilmente **inestabilidad**, interactuando con las características propias del circuito.

Sin entrar en detalles matemáticos, es necesario **conciliar** peticiones como el amplio rango de captura, la estabilidad y el bajo ruido superpuesto a la señal del VCO.

En otras palabras, el VCO debe **moverse** bastante despacio para ser insensible a las interferencias, pero lo bastante rápido cono para seguir las variaciones de la frecuencia a la entrada.

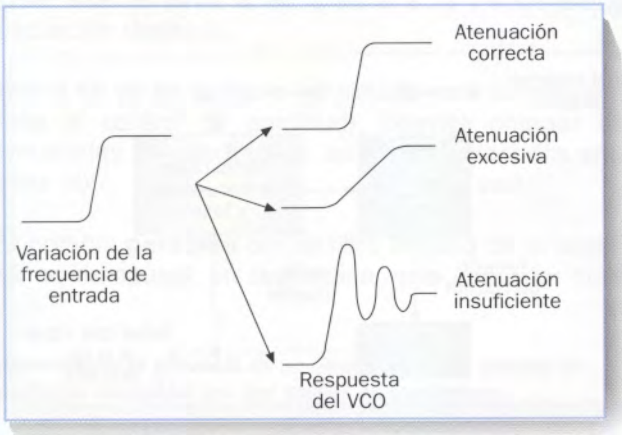
Atenuación correcta
Atenuación excesiva
Variación de la frecuencia de entrada
Atenuación insuficiente
Respuesta del VCO

La atenuación *influye en la respuesta del VCO de un PLL, en una rápida variación de la frecuencia de la señal a seguir.*

Integrados de audio

Existen diversos chips diseñados para determinadas aplicaciones

Sobre el silicio es fácil combinar circuitos sencillos para obtener funciones más complejas: ocultan, por tanto, siempre nuevos **chips a medida**, cada vez que la petición asegura una producción adecuada.

Por ejemplo, los **amplificadores de audio** integrados estaban antes limitados a potencias bastante bajas, mientras que hoy se encuentran chips capaces de controlar altavoces de potencia media-alta.

Esto se consigue integrando los **finales de potencia**, generalmente la tecnología MOSFET, con el controlador y el preamplificador, y resolviendo, naturalmente, los problemas térmicos y de convivencia entre los distintos circuitos.

REDUCCIÓN DE RUIDO

Hay varias técnicas para la **reducción de la banda** ocupada por la señal de audio, por ejemplo para el registro o la transmisión de bajo ruido (por ejemplo, Dolby, ver lección 11 de Aplicaciones).

Una de ellas consiste en reducir o eliminar los componentes de la señal que el oído **no puede escuchar**, porque están cubiertas por otros componentes de audio presentes en el mismo momento.

Aunque los integrados analógicos se utilizan normalmente en este campo, en los aparatos más evolucionados se prefiere convertir la señal en digital y utilizar un DSP (ver lección 48 de Digital).

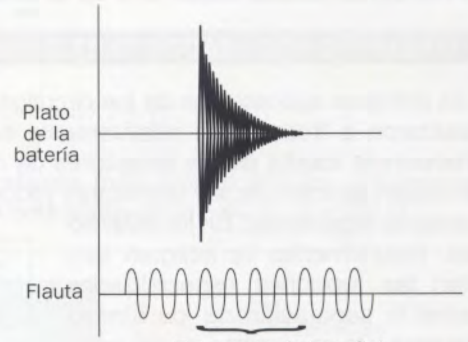

Esta parte de la señal esta cubierta por el ruido del plato, por lo tanto es inútil registrarla

En un fragmento musical, varios componentes del sonido no pueden ser percibidos por nuestro oído: se pueden eliminar para ahorrar banda.

SURROUND Y SIMILARES

Otros integrados de audio están dedicados a la reproducción de **efectos ambientales**, por ejemplo, la típica sonoridad de una sala de conciertos, o simplemente a la creación de un sonido más "rico".

En un ambiente grande, de hecho, las **múltiples reflexiones** en las paredes y las absorciones selectivas de los distintos componentes del sonido se traducen en retardos y desfases para el oído.

Reproduciendo exactamente (o construyendo) estos desfases, el que escucha recibirá las mismas **sensaciones** audibles: tendrá, por tanto, la sensación de encontrarse en un ambiente distinto del suyo.

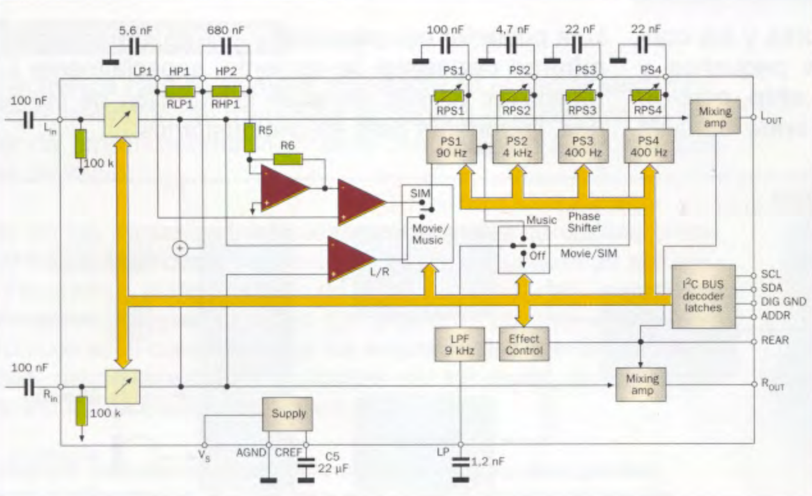
5,6 nF
680 nF
100 nF
4,7 nF
22 nF
22 nF
LP1
HP1
HP2
RLP1
RHP1
PS1
PS2
PS3
PS4
RPS1
RPS2
RPS3
RPS4
Mixing amp
L_{OUT}
100 nF
L_{in}
100 k
R5
R6
PS1 90 Hz
PS2 4 kHz
PS3 400 Hz
PS4 400 Hz
SIM
Movie/ Music
Music
Phase Shifter
Off
Movie/SIM
L/R
I²C BUS decoder latches
SCL
SDA
DIG GND
ADDR
REAR
LPF 9 kHz
Effect Control
100 nF
R_{in}
100 k
Mixing amp
R_{OUT}
Supply
V_S
AGND
CREF
C5 22 µF
LP
1,2 nF

Este integrado (STMicroelectronics) produce el efecto de surround con una serie de desfases controlados por la propia señal.

Integrados RF

La frecuencia máxima de trabajo está en continuo aumento

Análogamente a lo que ha ocurrido en campo digital, los integrados analógicos se han hecho cada vez más **rápidos** con el paso del tiempo, y esta tendencia no muestra signos de ralentización.

Si en los años 70 la **anchura de banda** de gran parte de los chips era apenas suficiente para las señales de audio, hoy existen integrados capaces de trabajar a frecuencias de varios GHz, es decir, miles de MHz.

La creación de nuevos chips, más económicos con respecto a los módulos prefabricados (ver lección 28 de Componentes), ha experimentado un fuerte empuje debido al notable desarrollo de la **telefonía móvil**.

Con el aumento de la* frecuencia de trabajo, *los integrados analógicos han "colonizado" nuevos territorios de frontera.

CIRCUITOS INTEGRADOS PARA RADIO/TV

Las primeras aplicaciones de los circuitos integrados se realizaron a frecuencias relativamente bajas, como la **frecuencia media** de los receptores de radio, y a continuación también de los televisores (donde el coste es bastante importante). En los televisores, naturalmente, se integran también las funciones especializadas, como la separación de los **sincronismos** y la generación de las señales de línea y de cuadro (ver lección 23 de Aplicaciones).

Los modernos integrados de radio pueden incluir, por ejemplo, un receptor FM completo de etapas a alta frecuencia, de sintonía **controlable digitalmente** mediante una conexión serie.

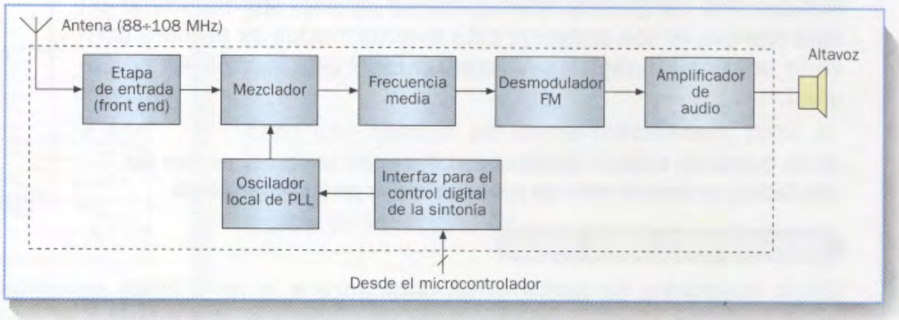
Antena (88÷108 MHz)
Etapa de entrada (front end)
Mezclador
Frecuencia media
Desmodulador FM
Amplificador de audio
Altavoz
Oscilador local de PLL
Interfaz para el control digital de la sintonía
Desde el microcontrolador

La integración de las etapas de alta* frecuencia *ha permitido la realización de radioreceptores completos en un solo chip.

TELÉFONOS MÓVILES

Subiendo con la frecuencia, los inductores y los condensadores se hacen cada vez más pequeños y pueden colocarse **directamente en el chip**, pero se hace más difícil evitar interferencias entre señales (diafonía).

No obstante, aquí también la tendencia se dirige hacia una **integración** cada vez mayor, con el objetivo del "teléfono móvil en un chip": reduciendo el número de integrados, se baja el coste total.

Una posterior complicación surge de la necesidad del **mínimo consumo** de corriente, especialmente en la recepción: de ello depende la duración de las baterías, importante para muchos usuarios.

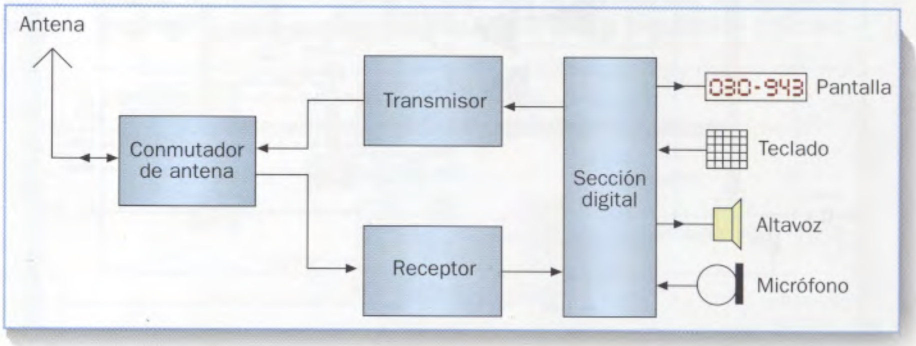
Antena
Conmutador de antena
Transmisor
Receptor
Sección digital
030-943
Pantalla
Teclado
Altavoz
Micrófono

Aunque sencillo en concepto, un* teléfono móvil *trabaja a frecuencia elevada: la integración completa plantea muchos problemas.

Coches e industria

Los integrados ya están difundidos incluso en entornos hostiles

La electrónica automovilística utiliza tanto tecnologías **digitales** como **analógicas**: estas últimas se emplean, por ejemplo, para la regulación de la energía producida por el alternador.

Los circuitos integrados para coche deben soportar una notable **fatiga** térmica y eléctrica, y deben además tener una fiabilidad superior que sus contrarios en la electrónica de consumo.

Es interesante observar que el uso de integrados **aumenta la fiabilidad** con respecto a los componentes discretos: se reduce el número de chips de silicio, y sobre todo el de interconexiones y de soldaduras.

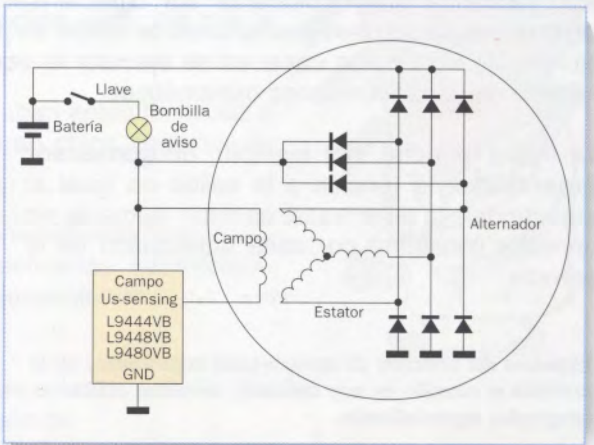

Un regulador para alternador se coloca sobre el motor del coche; por tanto, está sujeto a notables vibraciones y a cambios de temperatura.

LOS CONTROLADORES SON ANALÓGICOS

Como la electrónica digital domina ya incluso el campo industrial, es importante recordar que en realidad **todos** los circuitos, incluso los "lógicos", funcionan basándose en principios analógicos.

En particular, los **controladores** (ver lección 27 de Digital) son en realidad integrados analógicos a todos los efectos, como todos los circuitos de comunicación con el mundo exterior.

Igual que los chips para coches, también los industriales son normalmente capaces de funcionar en un rango de **temperaturas** más amplio que los integrados normales, habitualmente entre –40 °C y +125 °C.

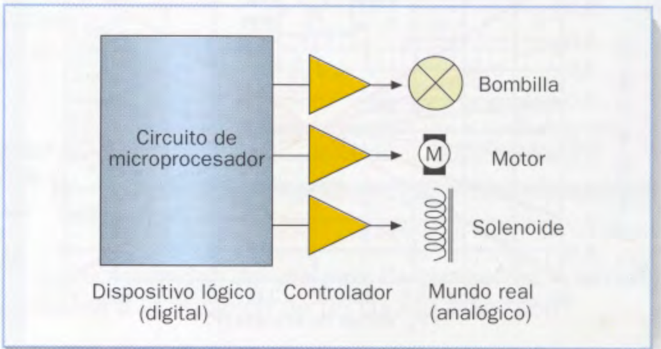

También de un modo totalmente digital, las técnicas analógicas son fundamentales para la interfaz con el entorno externo.

TELECOMUNICACIONES

La electrónica ha sido siempre importante en el sector telefónico: la propia invención del **transistor** se debió a la necesidad de disponer de amplificadores mejorados con respecto a los antiguos tubos de vacío.

Tanto en las centrales telefónicas como en las centralitas privadas y en los **teléfonos** domésticos, se han adoptado ya soluciones integradas, aunque (como en otros sectores) hay un uso creciente de tecnologías digitales. Un problema para los integrados telefónicos es el constituido por los **excesos de tensión momentáneos**, especialmente los causados por los rayos, directamente o por inducción electromagnética.

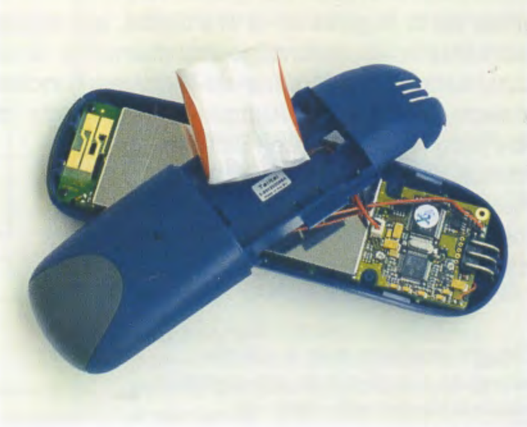

Un integrado para uso telefónico debe costar poco, pero debe ser muy robusto: la alimentación es irregular y está sujeta a excesos de tensión.

Matemáticas analógicas

No existen solamente los ordenadores digitales

Aprovechando adecuadamente las características **no lineales** de los semiconductores de unión, es posible realizar circuitos capaces de ejecutar el equivalente de varias funciones matemáticas.

La figura muestra, por ejemplo, un **conversor logarítmico**: la tensión a la salida es igual al logaritmo de la de entrada, es decir, aumenta en un valor constante por cada duplicación de la entrada.

Existe como integrado ya preparado, como otros circuitos de "cálculo", por ejemplo un **conversor rms** para medir el valor eficaz de una tensión alterna de cualquier forma (ver lección 4).

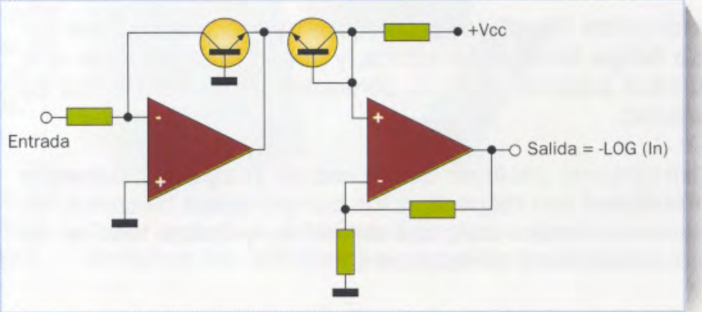

***Esquema del principio de un* conversor logarítmico*; en la práctica el circuito es muy delicado, conviene utilizar* integrados especializados.**

MULTIPLICADOR

Un circuito bastante utilizado en diversos campos es el **multiplicador**, que produce en su salida una tensión proporcional al producto de las dos tensiones aplicadas a las entradas.

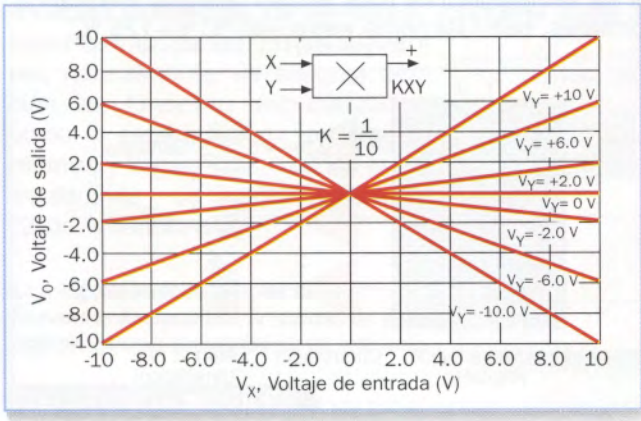

Puede, por tanto, funcionar como **control de amplitud** o de volumen, por ejemplo en un modulador AM (ver lección 25), aplicando a una entrada la portadora y a otra la señal moduladora.

Otro uso clásico es como **mezclador**, para la producción de frecuencia de batido (suma y diferencia) en las etapas de conversión, como se ilustra en la lección 26.

***La salida de un multiplicador (aquí el Motorola* MC1495*) es proporcional a* ambas entradas, es decir, a su producto.**

ORDENADORES ANALÓGICOS

Antes de la llegada de la era digital, se realizaron eficazmente calculadores completamente analógicos, constituido por una serie de **bloques funcionales** a conectar entre ellos. Aunque notablemente **imprecisos** para los estándares actuales, y sujetos a los inevitables errores de linealidad de los circuitos analógicos, tienen todavía algunas ventajas en los circuitos digitales.

El cálculo es de hecho prácticamente **inmediato** y continuo, y no está sujeto a errores de cuantización, lo que mejora la estabilidad en presencia de alimentación, por ejemplo en sistemas de regulación veloces.

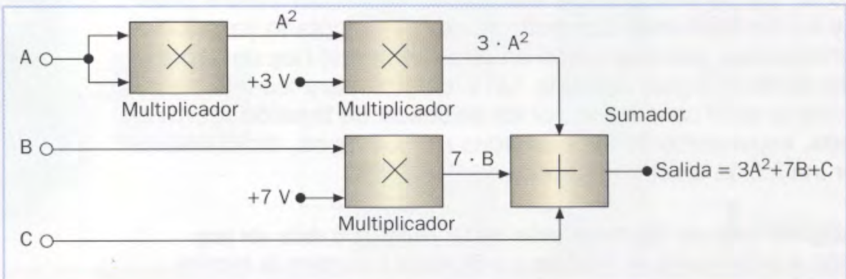

Circuito analógico para el cálculo inmediato y continuo de una* expresión matemática *(en este caso, un* polinomio*).

Válvulas termoiónicas

Los tubos de vacío, o tubos electrónicos, iniciaron la revolución electrónica

El **filamento** incandescente de una bombilla común no emite solamente luz, sino que produce, además, un cierto número de **electrones**, a los que el calor proporciona energía suficiente para separarse del hilo.

Este **efecto termoiónico** hace que alrededor del filamento encendido haya una "nube" cargada negativamente (estando hecha de electrones), la denominada "carga espacial".

Los electrones no pueden, sin embargo, **alejarse** tanto del hilo caliente, porque su ausencia hace positivo al mismo filamento, que por lo tanto los atrae nuevamente hacia sí.

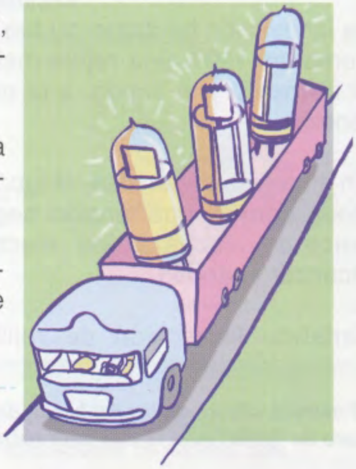

UN DIODO DE VACÍO

Aplicando tensión positiva, con respecto al filamento, a una placa metálica situada dentro del contenedor de cristal de la bombilla, esta **atrae** a los electrones producidos por el propio filamento.

Este fenómeno, descubierto por Edison, es aprovechado por Fleming en 1904 para construir el primer **diodo** de tubo electrónico, cuyo principio de funcionamiento se ilustra en la figura.

Si el **ánodo** (la placa) es positivo, la corriente pasa; pero si es negativo, al estar frío y no poder emitir electrones, no pasa corriente alguna.

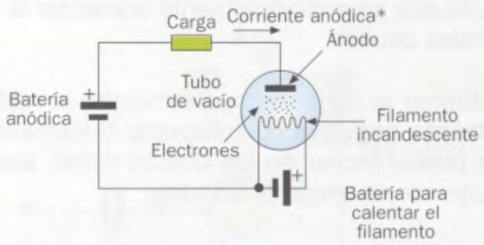

*Recordemos que, por convenio, la corriente se considera en sentido opuesto con respecto al movimiento real de los electrones.

Principio de un diodo de vacío: el flujo de electrones puede desplazarse solamente por el filamento hacia el ánodo, y no en sentido opuesto.

DETECTOR DE RADIO

El diodo permitió construir los primeros "detectores" auténticos, es decir, **desmoduladores** AM (ver lección 26), como el mostrado en el esquema, que es un receptor de radio sencillo, pero que funciona.

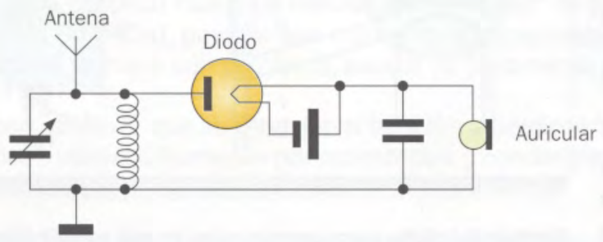

El diodo conduce durante las **semiondas positivas** de la señal de alta frecuencia, mientras que no conduce en las negativas; la corriente en el auricular sigue, por tanto, el perfil (modulación) de la señal de radio.

La señal en el auricular es, de este modo, bastante débil, ya que toda su energía procede de la señal de radio recibida por la antena: el circuito no proporciona ninguna **amplificación**.

Receptor de radio ***AM elemental, de diodo termoiónico; observe el símbolo del diodo, con el filamento y el ánodo esquematizados.***

El triodo

La incorporación de un tercer electrodo permite realizar un amplificador

La innovación fundamental fue introducida por Lee de Forest en 1907: una **rejilla** metálica interpuesta entre el filamento y el ánodo, a la que se le puede aplicar tensión.

En ausencia de tensión, el tubo se comporta como un diodo normal; una tensión **negativa** en la rejilla, sin embargo, rechaza los electrones, impidiéndoles alcanzar el ánodo.

Variando la tensión de rejilla, sin absorción de corriente, es posible **regular** la corriente anódica, de modo similar a lo que ocurre con el transistor MOSFET (ver lección 22).

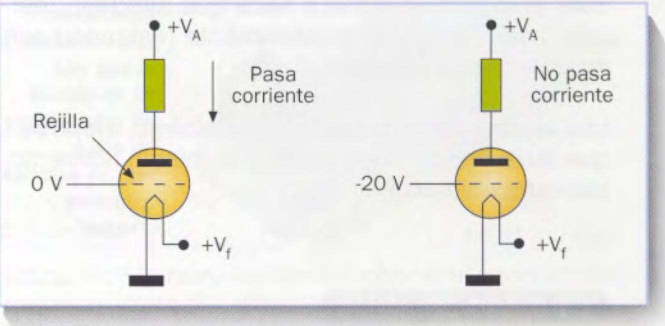

El nombre* válvula *deriva del hecho de que la tensión de rejilla hace de "grifo" para la corriente entre filamento y ánodo.

POLARIZACIÓN

La válvula de tres electrodos toma el nombre de **triodo**; la figura muestra una sencilla etapa amplificadora de triodo, el circuito que permitió finalmente aumentar la amplitud de las señales débiles.

Para obtener el mejor comportamiento, la rejilla debe ser **ligeramente negativa** con respecto al filamento, de modo que se pueda mover en los dos sentidos, aumentando o disminuyendo la corriente anódica.

Esta polarización se puede obtener con una sencilla **resistencia** (R_p en el esquema): la caída de tensión sobre él hace al filamento positivo con respecto a masa, y por lo tanto con respecto a la rejilla.

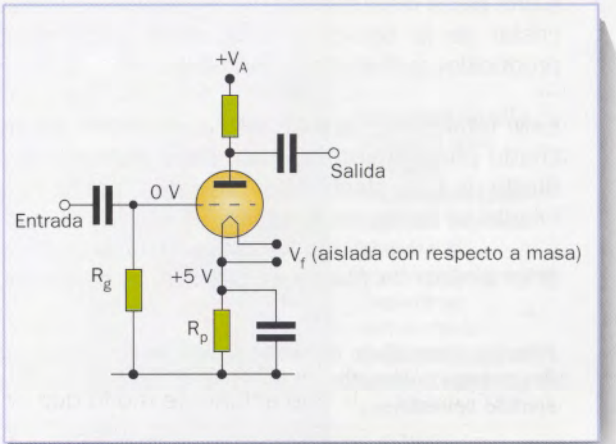

Amplificador de triodo,* autopolarizado *gracias a la resistencia situada entre el filamento y la masa.

VACÍO EN EL TUBO Y CÁTODO

Las primeras válvulas no eran muy **eficaces**, porque la emisión del filamento era relativamente escasa; además, las moléculas de aire residuales interferían con el paso de los electrones.

Una vez comprendidos los problemas, se resolvieron recubriendo el filamento de **tungsteno** para aumentar la emisión, y mejorar notablemente el vacío en el interior del tubo.

A continuación, el emisor de electrones fue separado eléctricamente del filamento y tomó el nombre de **cátodo**; permitía entre otras cosas utilizar corriente alterna para el filamento, sin introducir zumbidos.

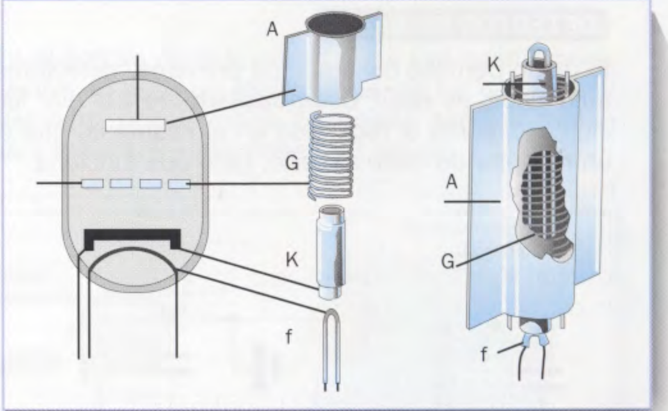

Símbolo del* triodo *y sus componentes en una válvula típica de los años 60: f=filamento, K=cátodo, G=rejilla, A=ánodo.

Válvulas más complejas

El número de electrodos fue aumentando progresivamente

Análogamente a la capacidad base-colector de los transistores, la **capacidad** entre rejilla y ánodo crea un recorrido no deseable de realimentación en el interior del tubo de vacío.

Para obviar este inconveniente, se incorporó una segunda rejilla para separarlos: la **rejilla de pantalla**, alimentada positivamente pero normalmente conectada a masa para la señal (con un condensador).

El tubo así obtenido, con cuatro electrodos, tomó el nombre de **tetrodo**; con la incorporación de una rejilla posterior para impedir la reemisión de electrones por parte del ánodo, se obtiene el **pentodo**.

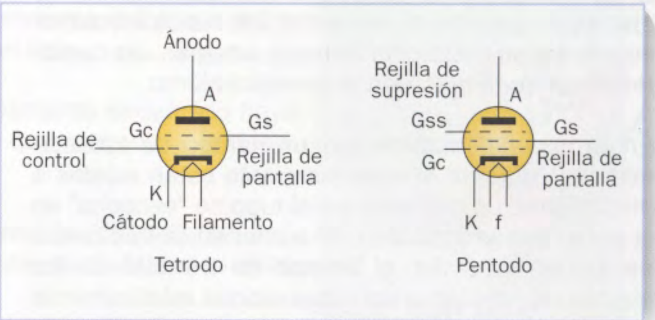

Tetrodo y pentodo: ***dos tipos de tubos electrónicos ampliamente utilizados para aplicaciones de distinto tipo.***

TUBOS MÚLTIPLES Y ESPECIALES

Para ahorrar y reducir la longitud de las conexiones, importante en alta frecuencia, se podían incluir **varias válvulas** (normalmente dos) en el mismo contenedor de cristal.

En paralelo con el aumento del **número de electrodos**, que llegó hasta 9 (n-odo), se introducían siempre nuevos tubos estudiados a medida para determinadas aplicaciones.

Por ejemplo, en los receptores de radio se utilizaba a menudo un **triodo-hexodo**, que desarrollaba las funciones de amplificador de entrada, oscilador local y conversor de frecuencia.

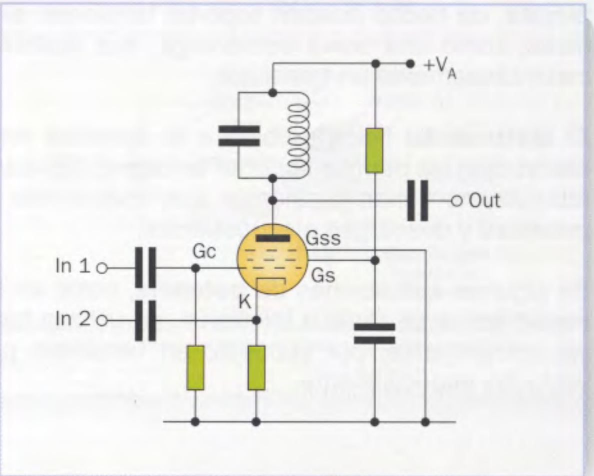

Mezclador por conversión, ***de pentodo: el circuito aprovecha la no linealidad del tubo para variar entre las dos señales; el filamento y la alimentación correspondiente están implícitos.***

EVOLUCIÓN DE LOS TUBOS DE VACÍO

La técnica fue mejorando, con tubos cada vez más fiables, con menos necesidad de energía y con **dimensiones** cada vez menores, de modo que se pudieron realizar dispositivos portátiles razonablemente prácticos.

El desarrollo continuó hasta los minúsculos "Nuvistor" de principios de los años 60 (RCA), pero en ese momento la superioridad de los **transistores**, aunque aún primitivos, estaba ya claramente definida.

Es curioso observar que se produjeron también auténticos "circuitos integrados": válvulas formadas por resistencias y condensadores **en el interior** del propio tubo.

PHILIPS
PL504
PL509
PY88
PL519
PY500
PCL86

Válvulas ***de distinto tipo: algunas tienen la conexión anódica arriba, lejos de los otros pines.***

Defectos y virtudes

Las válvulas no han desaparecido del todo: las hay en todas las casas

El **consumo** de corriente, debido a la necesidad de mantener caliente el filamento, los tubos electrónicos tienen una duración limitada; además, las características se alteran con el envejecimiento.

En el encendido necesitan un tiempo de calentamiento y, durante el funcionamiento están sujetas a **microfonía**: un golpecito en el tubo se "escucha" en la señal que lo atraviesa. Al aumentar la frecuencia de trabajo, además, el **tiempo de tránsito** de los electrones, debido a las dimensiones relativamente grandes (enormes con respecto a los transistores), crea numerosos problemas.

Algunos melómanos sostienen aún hoy que el sonido de un amplificador de audio de válvulas es, por principio, "mejor".

PUNTOS DE FUERZA

Por otra parte, las válvulas son inherentemente **robustas**, de hecho pueden soportar tensiones eléctricas, como una breve sobrecarga, que destruiría instantáneamente un transistor.

El **aislamiento** físico (debido a la distancia entre electrodos) les permite soportar tensiones elevadas; además, no tienen problemas con inversiones de polaridad y descargas electrostáticas.

En algunas aplicaciones **de potencia**, como en los transmisores de radio o televisión, se utilizan todavía comúnmente, por supuesto en versiones proyectadas especialmente.

Los tubos de alta potencia *para elevadas frecuencias (aquí algunos "klystron" grandes de transmisión) pueden requerir su enfriamiento mediante líquidos, como el motor de un coche.*

LAS VÁLVULAS HOY Y MAÑANA

El tubo electrónico más ampliamente utilizado es el **tubo catódico**, o CRT, empleado en los televisores y en los monitores de los PC, aunque la competencia de los cristales líquidos se hace cada vez más notable.

Incluso el horno de microondas utiliza una válvula: el **magnetrón** (ver lección 31 de Aplicaciones); es una aplicación habitual de alta frecuencia y alta potencia, donde los tubos de vacío se desenvuelven muy bien.

En el ámbito de la búsqueda una interesante perspectiva es la de los **microtubos**, en los que no hay filamento: pocos voltios de tensión anódica bastan para separar los electrones del cátodo.

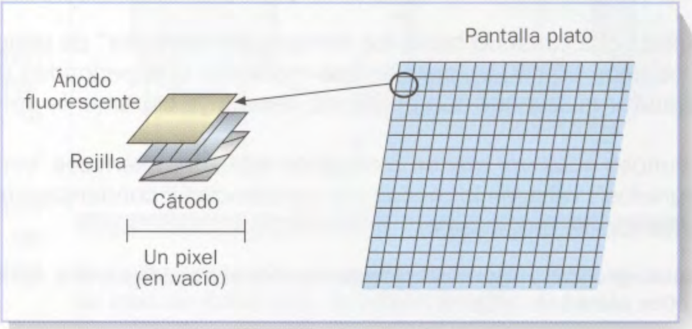

Las pantallas de microtubos *se han estudiado como posible alternativa a las demás tecnologías: en la práctica se trata de un conjunto de minúsculos CRT planos, sin filamento ni desviación.*

Microelectromecánica

Un chip puede contener también minúsculos dispositivos mecánicos

La industria de los circuitos integrados ha estimulado el desarrollo de **técnicas de elaboración** notablemente precisas, capaces de trabajar con resoluciones inferiores al micrón (milésima de milímetro). Nada impide utilizar las mismas técnicas para construir **dispositivos mecánicos**, bastante más pequeños que los que pueden obtenerse con las técnicas tradicionales a base de utensilios.

Normalmente un mismo chip puede contener tanto una parte micromecánica como una electrónica: este campo de aplicación, en fuerte desarrollo, toma por tanto el nombre de **MEMS** (microelectromechanics: microelectromecánica).

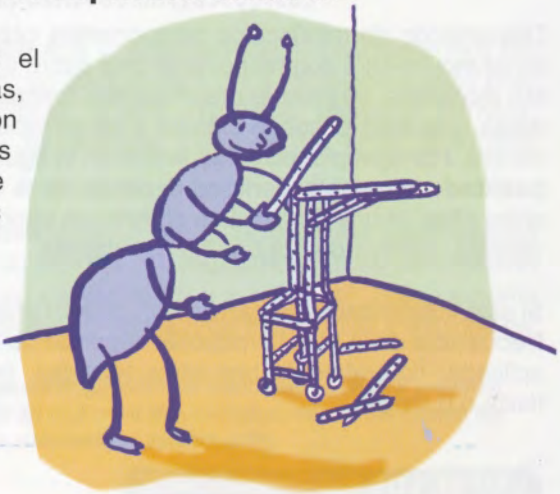

TÉCNICAS DE PRODUCCIÓN

Los microdispositivos pueden realizarse de varias maneras, normalmente basados en las ya bien probadas técnicas de **fotolitografía** utilizadas para la construcción de integrados (ver lección 32).

La figura ilustra la técnica del **estrato de pérdida**: un cierto material actúa como apoyo durante la elaboración, y después se elimina químicamente allí donde ocurre que un fragmento sea de movimiento libre.

Otro sistema (LIGA) consiste en realizar un **microsello** de polímero fotorresistivo, para después depositarlo sobre un estrato metálico que lo rellena, dando forma al fragmento deseado.

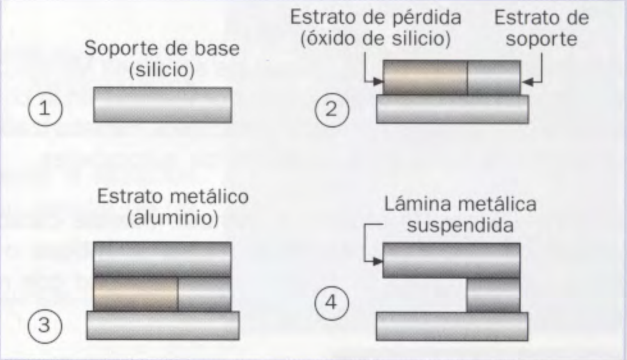

Realización de una lámina vibrante de aluminio* suspendida *en el vacío, con la técnica del estrato de pérdida.

MONTAJE Y CARCASA

En algunos casos es necesario **unir** fragmentos fabricados por separado, es decir, efectuar un "pegado" o fijación de precisión entre las superficies de dos componentes.

Según las características químicas de los materiales, se pueden pegar térmicamente, o bien se pueden utilizar técnicas **iónicas** que aseguran una atracción electrostática.

En la carcasa se hace habitualmente el **vacío**, con el fin de reducir la degradación química producto del paso del tiempo; a veces el montaje se ve complicado por la necesidad de una carcasa flexible en lugar de rígida.

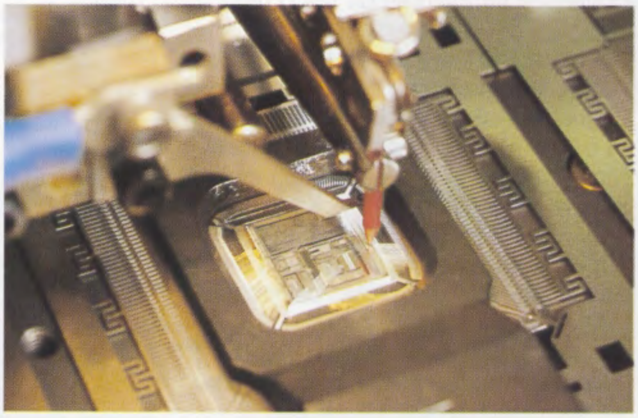

Microsensores

Los sensores miniaturizados abren nuevos campos de aplicación

Disponiendo de minúsculos componentes capaces de tener movimiento con respecto al chip sobre el que están montados, es posible medir débiles **fuerzas mecánicas** aplicadas desde el exterior a los propios componentes. Por ejemplo, como se ilustra en la figura, la **capacidad** entre dos superficies depende de la distancia entre ellas: si una es móvil se obtiene un condensador variable, controlado por una fuerza externa.

Si este condensador es parte de un circuito oscilante, la **frecuencia** de la señal producida depende de la fuerza aplicada; naturalmente hay otras técnicas, con base física o química.

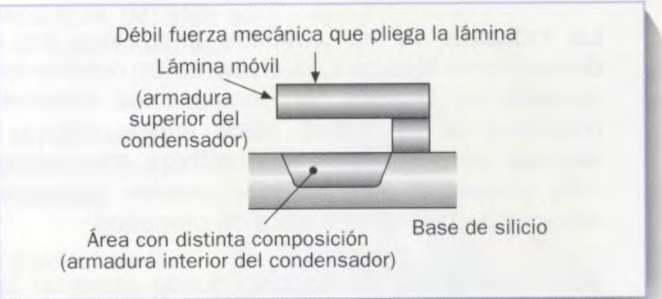
Débil fuerza mecánica que pliega la lámina
Lámina móvil
(armadura superior del condensador)
Base de silicio
Área con distinta composición
(armadura interior del condensador)

Conversor presión-frecuencia: ***el desplazamiento de la lámina varía la capacidad del condensador, que controla después un oscilador.***

USO DE LOS MICROSENSORES

Un interesante ejemplo de aplicación de los microsensores consiste en el lector de **huellas digitales**, que funciona como clave personal de acceso a lugares o datos que se desean proteger.

En otras aplicaciones se utilizan los sensores MEMS con partes móviles o flexibles, como por ejemplo en la construcción de **acelerómetros** y giroscopios miniaturizados, utilizados también en el sector de los automóviles.

Los microsensores ofrecen a menudo buenas características, también, para detectar **ondas acústicas** o vibraciones, con bajo coste y elevada fiabilidad con respecto a dispositivos más voluminosos.

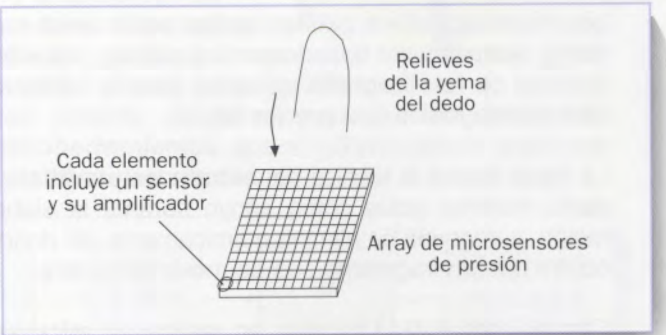
Relieves de la yema del dedo
Cada elemento incluye un sensor y su amplificador
Array de microsensores de presión

Un* array *rectangular de microsensores permite leer relieves y valles de la superficie de la yema de un dedo.

QUÍMICA Y MEDICINA

Las pequeñas dimensiones simplifican también la realización de **dispositivos electroquímicos**: una centralita completa para el control medioambiental podría estar contenida también en un solo chip.

En el campo **médico** se pueden diseñar sensores para su introducción en el interior del cuerpo, durante una consulta o bien permanentes (por ejemplo, un marcapasos), con una mínima molestia para el paciente.

Un aspecto no secundario de la miniaturización y de las técnicas utilizadas es la **reducción de costes**: los sensores microelectromecánicos abren nuevas vías a aplicaciones antes impensables.

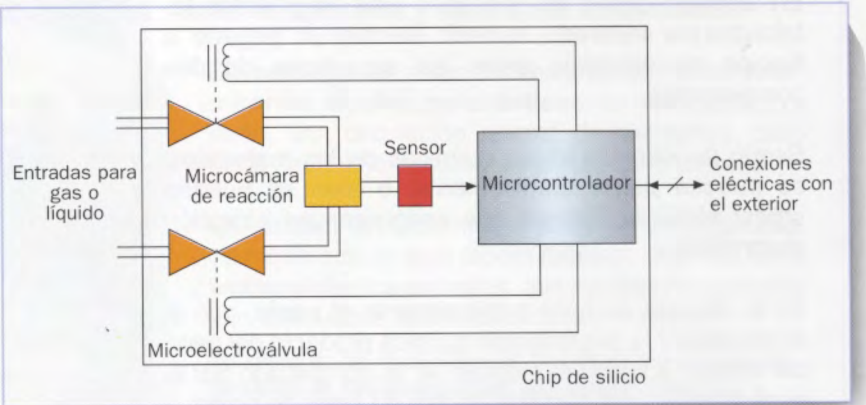
Entradas para gas o líquido
Microcámara de reacción
Sensor
Microcontrolador
Conexiones eléctricas con el exterior
Microelectroválvula
Chip de silicio

Esquema de bloques de un dispositivo microelectrónico-químico: es posible realizarlo en un solo chip con técnicas MEMS.

Microactivadores

Es posible producir fuerza mecánica con dispositivos microscópicos

Con las mismas técnicas de elaboración utilizadas para los sensores se pueden también obtener versiones miniaturizadas de los **activadores** convencionales, por ejemplo electroválvulas o relés.

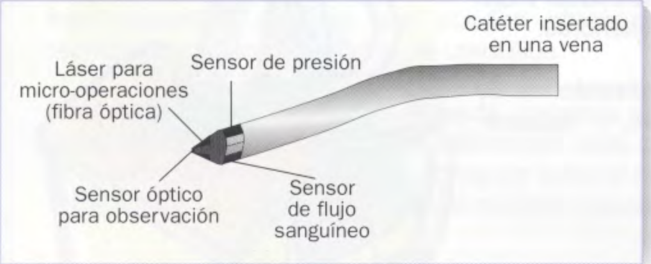
Catéter insertado en una vena
Láser para micro-operaciones (fibra óptica)
Sensor de presión
Sensor óptico para observación
Sensor de flujo sanguíneo

Las **fuerzas** que se obtienen son naturalmente muy débiles, pero en compensación los microactivadores pueden funcionar a una velocidad bastante elevada, gracias a su pequeña masa (y, por tanto, a su baja inercia).

Esto permite, por ejemplo, diseñar una electroválvula capaz de realizar **microdosificaciones** de precisión (útiles en campo médico y en otros), prácticamente imposibles de lograr con las tecnologías tradicionales.

En campo médico, sensores y activadores MEMS permiten funcionar en el ámbito microscópico, dentro del cuerpo del paciente, causando el mínimo daño.

MICROMOTORES

También pueden construirse los motores con microtecnología: tienen numerosas **aplicaciones**, desde la relojería a la medicina, incluyendo naturalmente las fabricaciones en campo electrónico.

Además de las habituales técnicas basadas en fuerzas electromagnéticas (ver lección 6 de Aplicaciones), se puede utilizar la atracción **electrostática** y la transmisión del movimiento a través de ondas acústicas.

Por ejemplo, se puede hacer que una **onda ultrasónica** cause la vibración de una microplataforma, que a su vez indica el movimiento circular de un rotor apoyado en ella.

©1993 UW-Madison
106009 10KV X40.0 .75mm

Micromotor electromagnético ***realizado con técnicas MEMS: el complejo entero tiene dimensiones milimétricas.***

MICROESPEJOS

Un interesante ejemplo de **nuevas aplicaciones** permitidas por la microelectromecánica es el array (matriz) de microespejos de proyección, cuyo funcionamiento se muestra en la figura. Cada uno de los microespejos, realizados en un único chip, controla un pixel de la imagen, y puede **girarse** en muy poco tiempo (pocos microsegundos), gracias a la atracción electrostática.

Cada pixel independiente puede encenderse o apagarse, girando rápidamente el espejo hacia delante y hacia atrás, y **variar** su luminosidad con un principio similar al de una fuente de alimentación de conmutación.

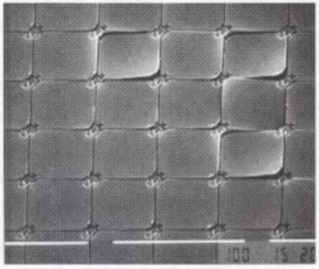

Array de microespejos ***para proyección (desarrollado por Texas Instruments): cada espejo controla un pixel.***

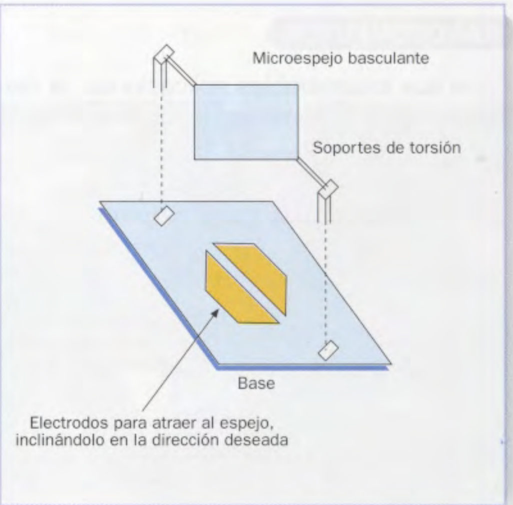
Microespejo basculante
Soportes de torsión
Base
Electrodos para atraer al espejo, inclinándolo en la dirección deseada

Nanotecnología

Por debajo de las dimensiones de los circuitos actuales, queda aún mucho espacio

Los circuitos integrados y los dispositivos microeletromecánicos (MEMS) se realizan con una **precisión** elevadísima, aprovechando al máximo las tecnologías actuales, como la fotolitografía.

No obstante, los límites teóricos están lejos: se trabaja en los alrededores del micrón (millonésima de metro), mientras que las moléculas individuales son del orden del **nanometro**, es decir, mil veces menores.

Es un poco como trabajar con una excavadora, moviendo **montones de arena**: cada granito es de unas dimensiones diminutas, y requiere tecnologías y herramientas completamente distintas.

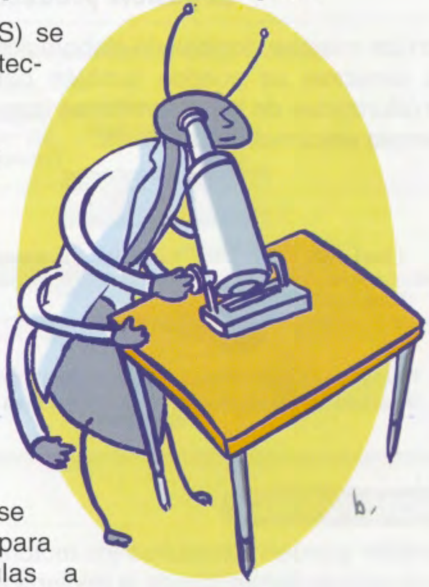

MÁQUINAS MOLECULARES

Las células de los organismos vivos contienen **nanomáquinas**, capaces de trabajar con gran precisión con moléculas individuales formando moléculas complejas (proteínas).

Se trata de los **ribosomas**, que realizan secuencias de instrucciones contenidas en otras moléculas (ARN), de un modo similar a una máquina de control numérico, que ejecuta su propio programa.

En la actualidad se pueden utilizar para obtener moléculas a medida, y en el futuro quizá sea posible utilizarlos para fabricar auténticas **nanoherramientas**, que se puedan utilizar realmente para la fabricación.

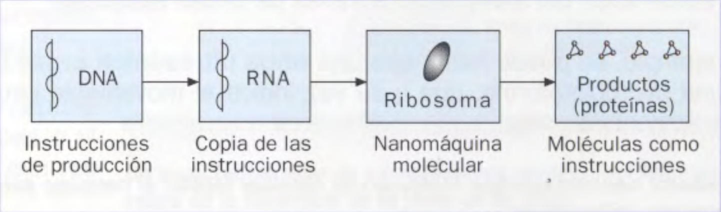
DNA
RNA
Ribosoma
Productos (proteínas)
Instrucciones de producción
Copia de las instrucciones
Nanomáquina molecular
Moléculas como instrucciones

El objetivo de la investigación es la realización de* nanomáquinas *capaces de funcionar sobre el modelo de las biológicas existentes.

NANOCOMPUTER

Entre sus innumerables aplicaciones, la nanotecnología permitiría la construcción de **circuitos integrados** (y otros dispositivos) bastante más pequeños, y por lo tanto más potentes y rápidos que los actuales.

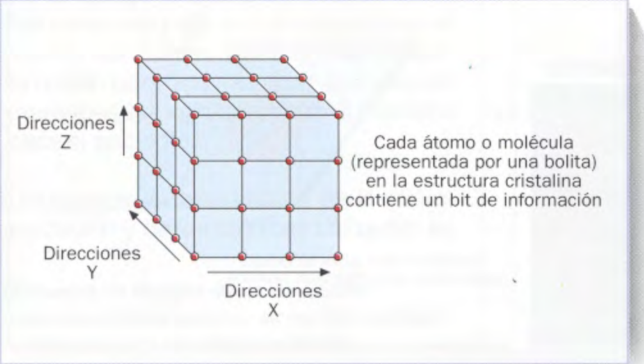
Direcciones Z
Direcciones Y
Direcciones X
Cada átomo o molécula (representada por una bolita) en la estructura cristalina contiene un bit de información

Sin embargo, en estas dimensiones se hacen determinantes los **efectos cuánticos**: no se tendría, por tanto, una circulación normal de electrones, pero sería necesario basarse en principios diferentes.

Se producirían, además, problemas de **fiabilidad**, de modo que un solo átomo fuera de lugar causaría funcionamientos erróneos, que se deberían prevenir con adecuados sistemas de autocorrección.

Un sueño de los diseñadores de hardware: la* nanomemoria cristalina *de tres dimensiones, con enorme capacidad y muy bajo tiempo de acceso.

Amplificadores de aislamiento

Transportan una señal analógica sin contacto eléctrico

No todos los dispositivos electrónicos pueden estar contenidos en **un solo chip** de silicio: a veces es necesaria una construcción modular con dos o varios integrados eléctricamente separados.

Es el caso, por ejemplo, de los "amplificadores de aislamiento", que tienen la misión de **transferir** una señal analógica de una entrada a una salida, formando parte sin embargo de circuitos distintos.

Se utilizan (entre otras cosas) en **aplicaciones Industriales**, donde es difícil mantener al mismo nivel las masas de los distintos aparatos, y puede haber tensiones o interferencias significativas entre ellos.

ALTA TENSIÓN
b.

AISLAMIENTO DE TRANSFORMADOR

Como se ilustra en la figura, un amplificador de aislamiento modula una portadora con la señal a transmitir, obteniendo una tensión alterna que un **transformador** lleva a la otra parte.

Una vez **desmodulada** la señal de un segundo circuito, se obtiene de nuevo la tensión original, sin contacto eléctrico ("galvánico") con la entrada y, en general, con una linealidad muy buena.

Para mantener separada la **alimentación** de las dos partes se utiliza después otro transformador, o incluso dos en el caso de que se deba aislar hacia la entrada en lugar de hacia la salida.

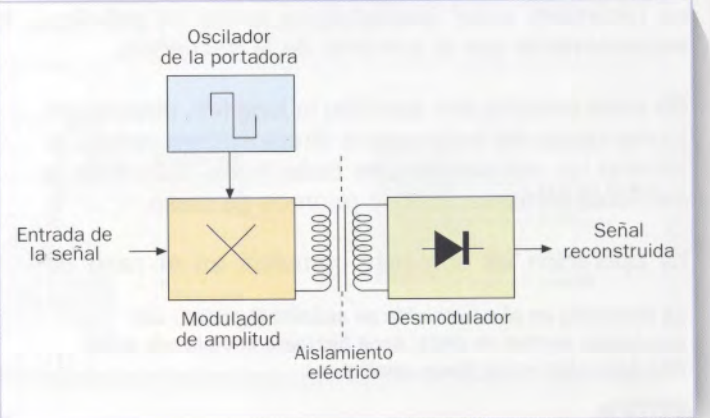

Esquema de bloques de un **amplificador de aislamiento***; también la alimentación se proporciona mediante* **transformadores.**

BANDA LIMITADA

La portadora en alterna limita la **frecuencia máxima** utilizable para la señal: la banda de paso va de cero, es decir, la tensión continua, hasta una fracción de la frecuencia de la misma portadora.

Además, los transformadores añaden voluminosidad y aumentan el **coste** de producción, una posible alternativa consiste en emplear una solución digital en lugar de analógica.

Se muestra en la figura un posible esquema, pero de todas formas tampoco es perfecto, porque la **cuantización** de la señal debida al ADC (ver lección 21 de Digital) introduce errores y ruido.

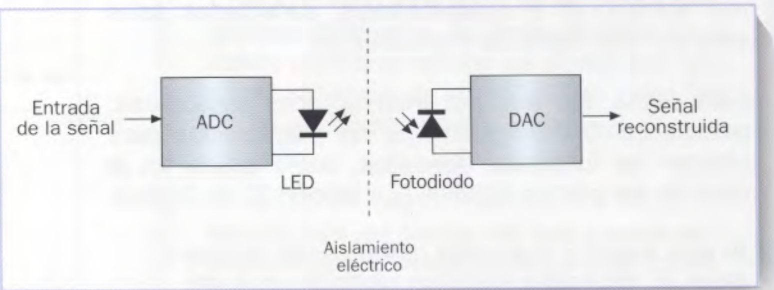

Amplificador de aislamiento con conversión analógico digital, **transferencia óptica** *y conversión recíproca.*

Diseñar integrados analógicos

No es siempre necesario hacer todo desde cero

Para aplicaciones a gran escala, se puede fabricar un circuito integrado analógico **a medida**; se habla entonces de ASIC (Application-Specific Integrated Circuit: circuito integrado específico de aplicación), como en el campo digital. Como alternativa al diseño desde el principio, bastante costosa, es posible utilizar **celdas analógicas**, es decir, dispositivos analógicos elementales para los que ya esté hecha la máscara de fabricación.

El trabajo, aun siendo complicado, se reduce en este caso a la **interconexión** entre los dispositivos existentes, y a la eventual incorporación o sustitución de algún componente.

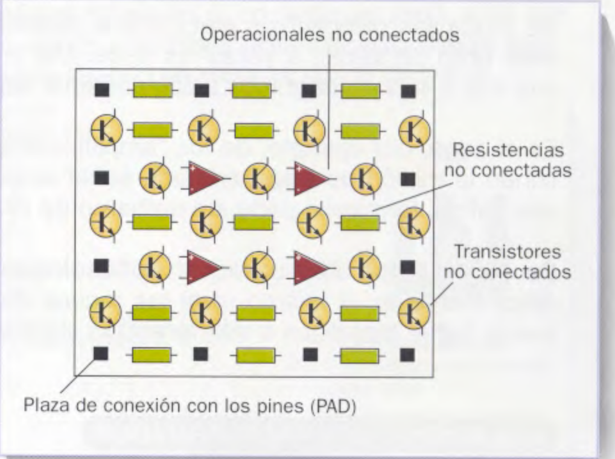
Operacionales no conectados
Resistencias no conectadas
Transistores no conectados
Plaza de conexión con los pines (PAD)

***Ejemplo de array analógico:* el diseñador debe solamente añadir el estrato de metal que conecta entre ellos los distintos componentes.**

SEÑAL MEZCLADA

La **puesta a punto** de un chip analógico suele ser más complicada con respecto a la de un chip digital: es necesario evitar interferencias entre las señales, especialmente con el aumento de la frecuencia.

Se debe estudiar con atención la longitud, disposición y orientación de los accesos directos, para reducir al mínimo los **acoplamientos** capacitivos, inductivos y mediante alimentaciones y retornos de masa.

La operación es aún más compleja en el caso de chips de señal mezclada, en los que las señales **digitales y analógicas** conviven en el mismo integrado: los primeros tienden a molestar a los segundos.

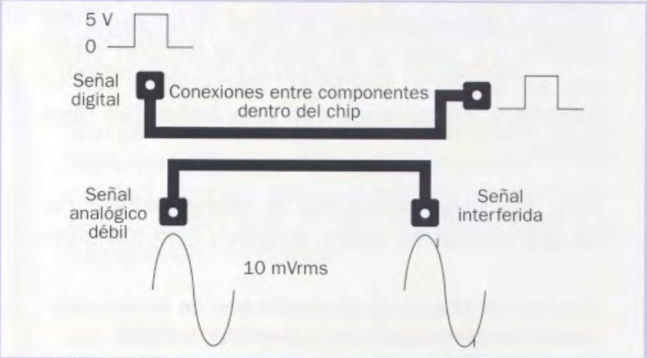
5 V
0
Señal digital
Conexiones entre componentes dentro del chip
Señal analógico débil
Señal interferida
10 mVrms

La presencia en el mismo chip de señales* digitales*, con empinados frentes de onda, hace bastante complicado evitar interferencias en las líneas analógicas.

FPAA

Se han propuesto también circuitos analógicos que pueden **fabricarse eléctricamente**, sobre el modelo de los chips digitales programables, llamados FPAA (Field Programmable Analog Array).

Contienen, en general, una serie de **macroceldas** relativamente complejas, como amplificadores, integradores, sumadores o multiplicadores, conectados entre ellos por una matriz de interconexión.

Esta última, formada por **interruptores analógicos**, permite combinar digitalmente las macroceldas para obtener las funciones deseadas, como ocurre en el caso de sus primos digitales (ver lección 32 de Digital).

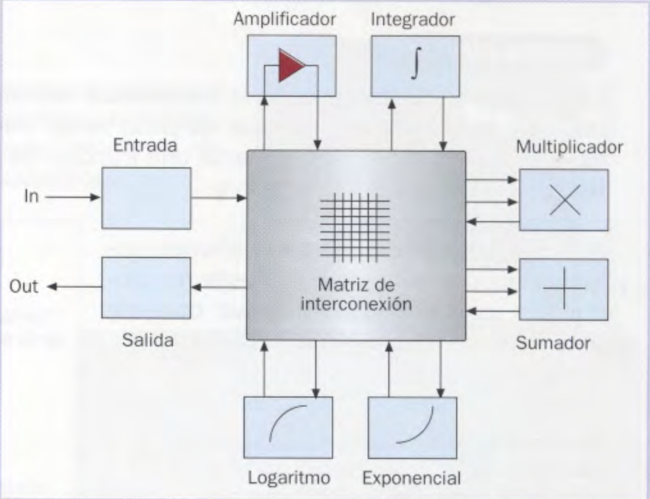
Amplificador
Integrador
Entrada
Multiplicador
In
Out
Matriz de interconexión
Salida
Sumador
Logaritmo
Exponencial

Un array analógico programable (FPAA) permite conectar a placer las macroceldas analógicas contenidas en el chip.

Síntesis de voz

El principio es analógico, la realización es normalmente digital

Los sistemas capaces de generar **voz artificial** encuentran numerosas aplicaciones: juguetes, sistemas de respuesta telefónica automáticos, publicidad en las emisoras de radio, sistemas de conducción de vehículos, etc.

Si el mensaje es **previsible**, se puede simplemente grabar; en caso contrario es necesario sintetizar (es decir, producir) cada vez los sonidos correspondientes a las palabras deseadas, o mejor a los "fonemas" que las componen.

Esto se puede hacer con varias técnicas, tanto **analógicas como digitales**; sin embargo, se suelen preferir estas últimas, porque aseguran una óptima capacidad de repetición a un coste reducido.

APARATO VOCAL

El sistema **LPC** (Linear Predictive Coding: codificación lineal predictiva, difundido por Texas Instruments), por ejemplo, imita las características de los órganos que, en el cuerpo humano, producen la voz.

Algunos sonidos, por ejemplo la letra C de "cima", se producen partiendo de **ruido blanco** (ver lección 41), obtenido a su vez restringiendo el paso del aire para acelerarla y crear turbulencia.

Los **pulmones** regulan el volumen cambiando la presión del aire, las **cuerdas vocales** vibran produciendo impulsos de presión, y la **cavidad oral** resuena según la forma que adquieren los músculos.

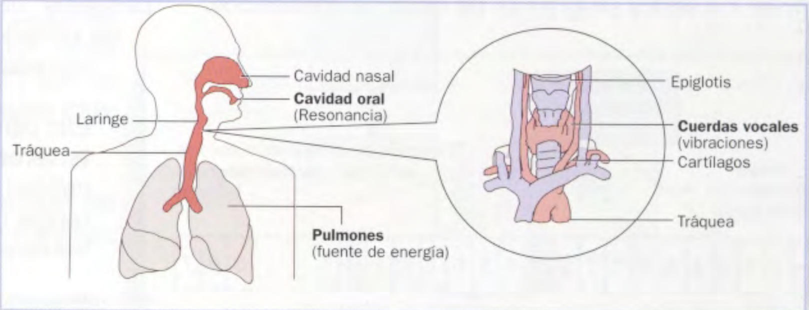
Cavidad nasal
Cavidad oral
(Resonancia)
Laringe
Tráquea
Pulmones
(fuente de energía)
Epiglotis
Cuerdas vocales
(vibraciones)
Cartílagos
Tráquea

Estructura del* aparato vocal humano*: las vocales son notas repetitivas, muchas consonantes son transiciones entre dos vocales.

CIRCUITO DE SÍNTESIS

La figura muestra el esquema de bloques de un **sintetizador vocal** que utiliza la técnica LPC: el efecto de las cuerdas vocales es simulado con un oscilador modulado a impulsos.

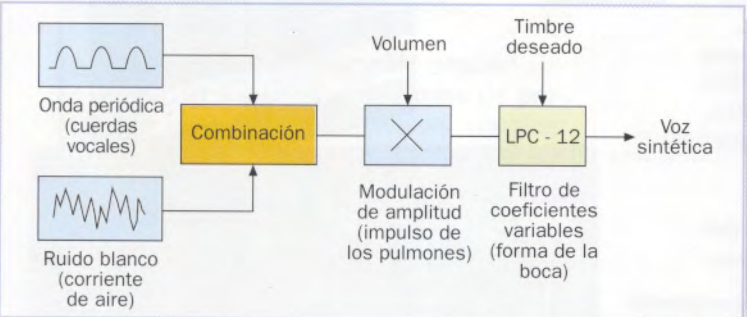
Onda periódica
(cuerdas
vocales)
Ruido blanco
(corriente
de aire)
Combinación
Volumen
Modulación
de amplitud
(impulso de
los pulmones)
Timbre
deseado
LPC - 12
Filtro de
coeficientes
variables
(forma de la
boca)
Voz
sintética

Las características del **filtro** (que representa la cavidad oral) se modifican periódicamente, de modo que se obtengan las resonancias deseadas, produciendo sonidos similares a los reales.

Como un filtro analógico no se puede **programar** fácilmente, es decir, modificarse al vuelo, conviene trabajar en digital, con técnicas similares a las utilizadas para los DSP (ver lección 48 de Digital).

Síntesis vocal *con técnica LPC: cada componente simula el comportamiento de una parte del aparato vocal humano.*

DAB: la radio digital

Las tecnologías analógicas se retiran de uno de sus sectores tradicionales

En muchos campos se produce un progresivo **cambio** de lo analógico hacia lo digital, pero es especialmente significativo en la radiodifusión, una de las primeras aplicaciones de la electrónica.

El **estándar europeo** de transmisión digital, el DAB (Digital Audio Broadcasting: emisión de audio digital) Eureka-147, resuelve algunos problemas de las transmisiones habituales FM, como la aglomeración y la sensibilidad a las interferencias. La transmisión analógica es también sensible a los **ecos**, es decir, a las señales reflejadas o, en cualquier caso, a los que han recorrido una mayor distancia; esto ocurre especialmente cuando se escucha desde un coche.

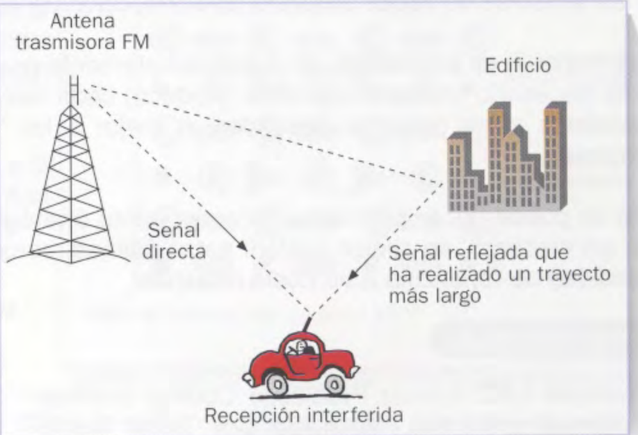
Antena trasmisora FM
Edificio
Señal directa
Señal reflejada que ha realizado un trayecto más largo
Recepción interferida

En la radiotransmisión FM, una señal **desfasada** *(a causa de un trayecto más largo), interfiere con el decodificador e introduce interferencias.*

FRECUENCIAS MÚLTIPLES

Un canal DAB en "modo 1" ocupa 1.536 MHz de anchura de banda, comprende 1.536 **portadores** separados (sistema "COFDM") y puede utilizarse para emitir 7 o varios programas de radio estereofónicos.

Cada **programa** se transmite con una serie de "paquetes" de datos, repartidos en un cierto número de portadores y conteniendo información para la corrección de los errores de recepción.

Ello permite una excelente **inmunidad a las interferencias**: se puede recibir una señal de alta calidad, también con una relación señal/interferencia de tan sólo 10 dB, prohibitivo para una transmisión FM clásica.

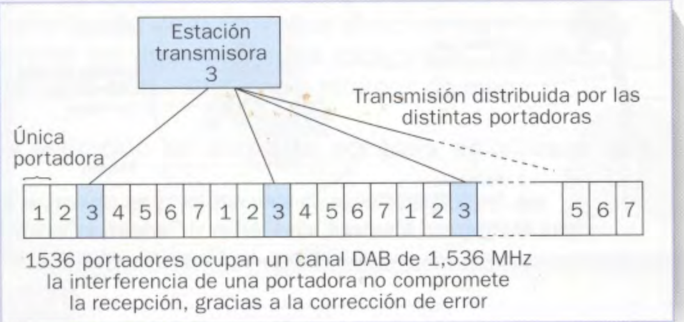
Estación transmisora 3
Transmisión distribuida por las distintas portadoras
Única portadora
1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2 3 5 6 7
1536 portadores ocupan un canal DAB de 1,536 MHz
la interferencia de una portadora no compromete
la recepción, gracias a la corrección de error

Una transmisión **DAB** *utiliza el sistema COFDM, que divide la información en varias bandas ortogonales.*

OTRAS VENTAJAS

La peculiar técnica de transmisión permite también **reducir la potencia** de los transmisores, con ventajas en consumo y salud, y transmitir a todas partes en la misma frecuencia (isofrecuencia).

La ocupación de banda es reducida, gracias a una **compresión de audio** ("MUSICAM", o bien MPEG 1/2 layer II) que elimina las frecuencias no audibles, por ejemplo porque están cubiertas por otras frecuencias.

Se puede añadir fácilmente **información digital**, por ejemplo de texto, sin interferir con la música, o directamente utilizar el sistema para transmitir imágenes de televisión.

DAB
GRUNDIG
DCR
HIGH POWER
GRUNDIG
RADIO
CD
TAPE
ANTENNE

Como el DAB resiste bien a ecos y señales múltiples, es especialmente adecuado para escuchar desde un **vehículo** *en movimiento.*

Cálculo circuital

Un análisis detallado de los circuitos requiere bastantes matemáticas

Al contrario de lo que se podría pensar, muchos dispositivos electrónicos (especialmente en campo digital) se pueden **diseñar** con pocos cálculos o ninguno, o bien aplicando fórmulas ya creadas.

Por otra parte, una buena **preparación teórica** no está de más, incluso en algunos casos se hace esencial, especialmente cuando se tenga que analizar con detalle el comportamiento de un circuito.

Hablaremos en esta lección de algunas **técnicas matemáticas**, mientras que en la próxima ilustraremos los cómodos programas informáticos de simulación, que permiten utilizarlas... sin conocerlas.

COMPONENTES Y CIRCUITOS

El objetivo principal del análisis de circuitos es el de determinar el **comportamiento** de un circuito electrónico, incluso bastante complejo, sin la necesidad de construirlo en la práctica.

En primer lugar, es necesario disponer de una **descripción matemática**, razonablemente aproximada, del funcionamiento de cada dispositivo físico (resistencia, condensador, transistor, etc.).

Después se deben considerar los efectos de la **interconexión** entre los distintos dispositivos que forman parte del circuito, de modo que se obtenga una descripción (también matemática) del dispositivo terminado.

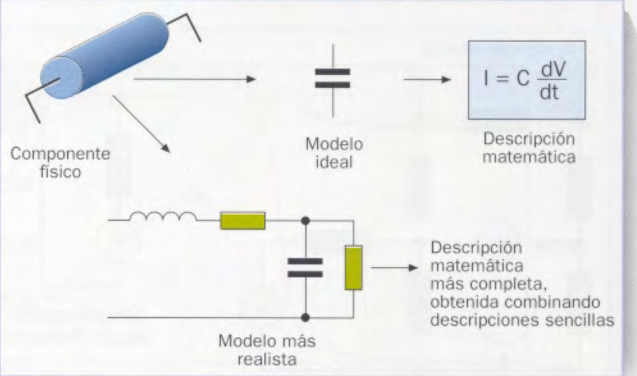

Componente físico
Modelo ideal
I = C dV/dt
Descripción matemática
Modelo más realista
Descripción matemática más completa, obtenida combinando descripciones sencillas

Un dispositivo físico puede describirse de un modo más o menos sencillo, según la aproximación deseada.

SINUSOIDE E IMPEDANCIA

Los circuitos menos complejos de estudiar son los formados por componentes de dos terminales **lineales** (bipolos), es decir, que no distorsionan una señal sinusoidal: resistencias, condensadores, inductores.

En otras palabras, se debe determinar la **impedancia** del circuito (ver lección 12) en la condición de "estado estacionario sinusoidal", es decir, una vez que se ha alcanzado una situación estable.

Una primera forma de análisis consiste en aplicar, se entiende matemáticamente, una tensión **sinusoidal** al dispositivo, calculando por lo tanto amplitud y fase de la corriente que resulta.

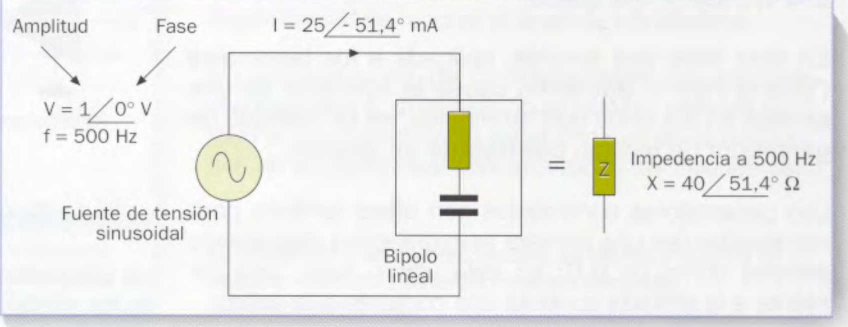

Amplitud
Fase
V = 1∠0° V
f = 500 Hz
Fuente de tensión sinusoidal
I = 25∠-51,4° mA
Bipolo lineal
=
Z
Impedencia a 500 Hz
X = 40∠51,4° Ω

Un circuito lineal, alimentado con una señal sinusoidal, presenta una "resistencia compleja", es decir, una impedancia.

Análisis de las redes

Herramientas matemáticas permiten simplificar el trabajo

Las **fuentes de tensión** (continua, sinusoidal u otra) están idealizadas con "generadores de tensión" teniendo resistencia interna nula; la eficaz se puede simular añadiendo una resistencia.

Recíprocamente, un **generador de corriente** ideal tiene impedancia infinita; para un dispositivo más realista (por ejemplo, un generador de corriente constante) se le añade una resistencia en paralelo.

Los generadores pueden corresponder a fuentes reales, como una fuente de alimentación, o utilizarse para **representar** otras partes del circuito desde el punto de vista de la que se está analizando.

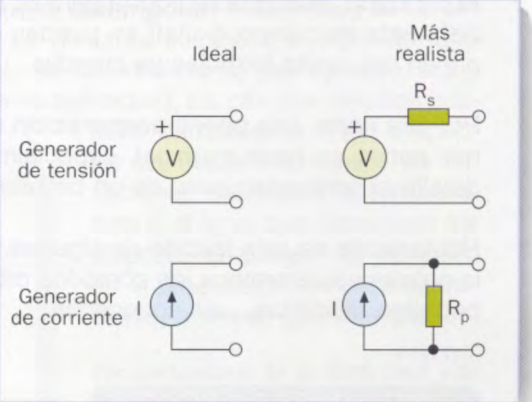

Generadores *de tensión y de corriente: el comportamiento de las fuentes de alimentación o de los generadores reales se simula con una resistencia adicional.*

REDES Y TEOREMAS

Varios teoremas ayudan a analizar el comportamiento de una **red** lineal, es decir, de un conjunto de componentes interconectados, tanto con respecto a una tensión continua como a una señal sinusoidal.

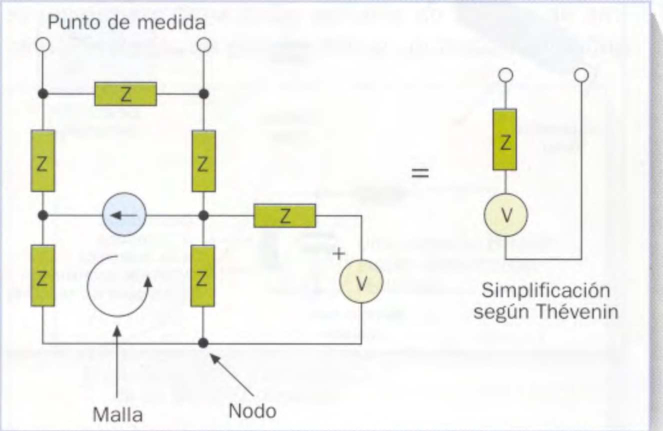

Una red está formada por **mallas** y los componentes se unen en los **nodos**; Kirchhoff demostró que la suma de las tensiones en una malla es cero, igual que la suma de las corrientes de un nodo.

Thévenin y Norton demostraron que una red se puede **simplificar** (dadas algunas condiciones) en un generador de tensión con una impedancia en serie, o de corriente con una impedancia en paralelo.

Simplificación de una red lineal, *que comprende incluso generadores de tensión o corriente (siempre que sean independientes uno de otro).*

CUADRIPOLOS

Un "cuadripolo", o circuito de **cuatro terminales**, se utiliza para representar dispositivos más complejos, como por ejemplo un amplificador lineal, dotado de una entrada y una salida.

En este caso una tensión, aplicada a los terminales ("puerta") de la izquierda, causa la aparición de una tensión en los otros dos terminales: es un ejemplo de generador de tensión **controlado** en tensión.

Los generadores controlados son útiles también para representar (en una primera aproximación) dispositivos **activos** como un BJT: en este último caso, una corriente a la entrada controla una corriente a la salida.

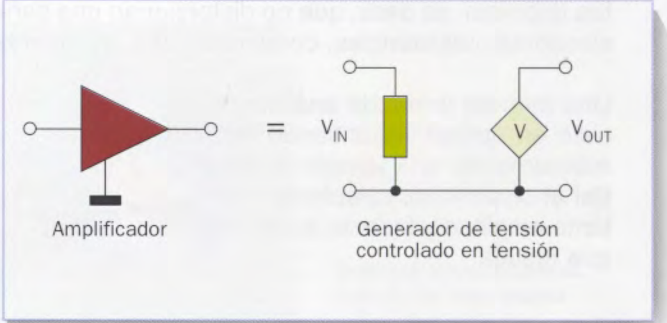

Los* generadores controlados *representan dispositivos dotados de dos puertas, normalmente una entrada y una salida.

Tipos de análisis

Un circuito puede examinarse en distintos aspectos

Aplicando exclusivamente **tensiones continuas** a un circuito lineal, éste tiende a alcanzar un estado estable, bastante fácil de determinar: basta con aplicar la ley de Ohm. En esta situación los **condensadores** pueden considerarse circuitos abiertos (interrumpidos), mientras que los inductores se comportan como sencillos cables de conexión (en cortocircuito).

A pesar de la simplicidad, este **análisis en continua** es útil para determinar el punto operativo del circuito, al que se somete.

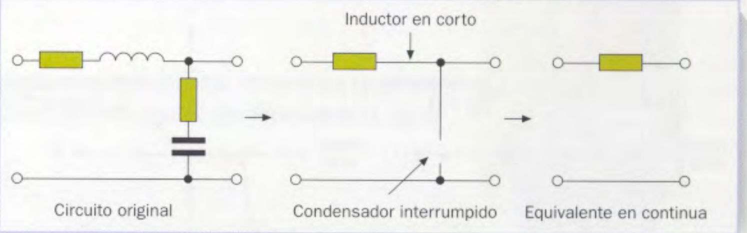

Desde el punto de vista de la continua, una vez alcanzado un estado estable, un circuito lineal se simplifica considerablemente.

SINUSOIDES Y FRECUENCIA

Utilizando, sin embargo, una fuente **sinusoidal** de amplitud fija, sus efectos (tensiones, corrientes y fases relativas) dependerán de la frecuencia de oscilación de la propia tensión.

Se puede, por tanto, describir el comportamiento del circuito en las distintas frecuencias, como cuando se examina un amplificador de audio: se habla de análisis **en el dominio de la frecuencia**.

Conociendo el comportamiento con una sinusoide, se puede determinar con una **forma de onda arbitraria**: basta con descomponerla en componentes sinusoidales y sumar los efectos asociados.

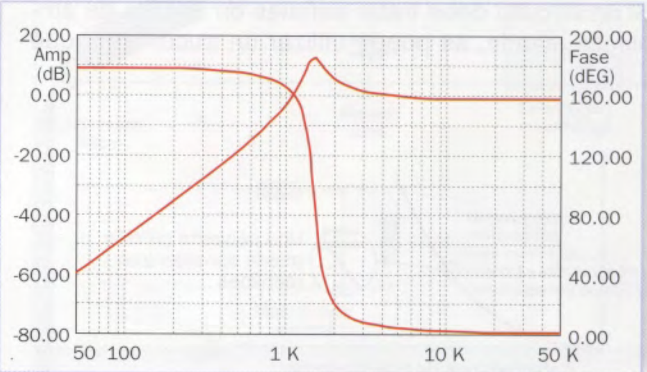

El análisis en el dominio de la frecuencia describe la variación de amplitud y de fase, en función de la frecuencia de la señal.

TRANSISTORES

Más complejo es el caso en el que al circuito se le aplica un **escalón de tensión** o, en cualquier caso, un “transitorio” entre dos estados, ya que se estabiliza gradualmente hacia un nuevo estado estable.

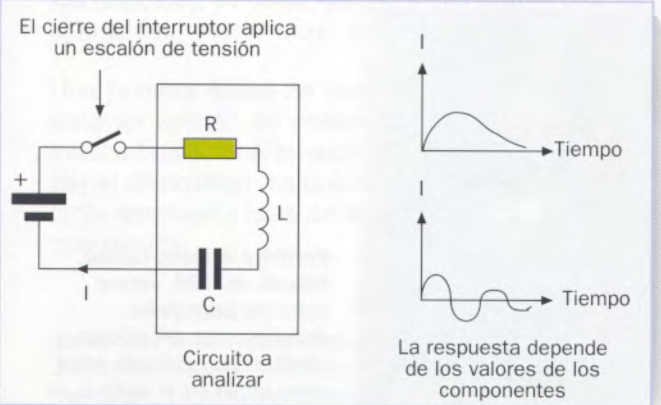

Circuito a analizar

La respuesta depende de los valores de los componentes

El análisis con detenimiento de la **transición** entre un estado y otro es, sin embargo, bastante complicado, porque es necesario resolver ecuaciones de tipo exponencial: se recurre al análisis infinitesimal.

En compensación, del comportamiento de un circuito en respuesta a un transitorio es posible **deducir** también el comportamiento en el dominio de la frecuencia: un escalón ideal contiene todas las frecuencias.

La respuesta a un escalón, aunque no siempre sencilla de calcular, contiene (matemáticamente hablando) informaciones útiles.

Elementos no lineales

Con los semiconductores, las cosas se complican un poco

Por desgracia, el comportamiento de un transistor (o incluso de un sencillo diodo) no es en absoluto **lineal**: a una duplicación de la tensión aplicada no le corresponde una duplicación de la corriente.

Como una buena **descripción** matemática de un semiconductor es bastante compleja, y depende también del modo de funcionamiento (por ejemplo, signo de las tensiones aplicadas), los cálculos resultan bastante complicados. Para empeorar las cosas intervienen las variaciones debidas a la **temperatura** y, al tener que considerar los dispositivos reales, la variación en la producción (también significativa) entre un ejemplar y otro.

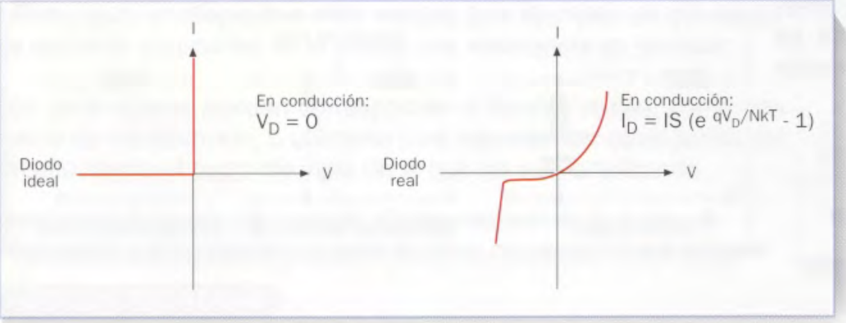

Comportamiento de un diodo **ideal** *y de su correspondiente real: la descripción matemática del segundo es mucho más compleja.*

UNA CÓMODA SIMPLIFICACIÓN

Si un circuito debe tratar señales en alterna de amplitud limitada, se puede utilizar un truco: fingir que los semiconductores se comportan de modo **lineal** para pequeñas variaciones de la señal.

En la práctica, se calcula el **punto operativo** en corriente continua y se considera que la señal hace desplazar muy poco este punto a lo largo de la curva característica del dispositivo.

En estas condiciones, la minúscula porción de curva considerada puede aproximarse a una recta: esta técnica toma el nombre de análisis para **pequeñas señales**.

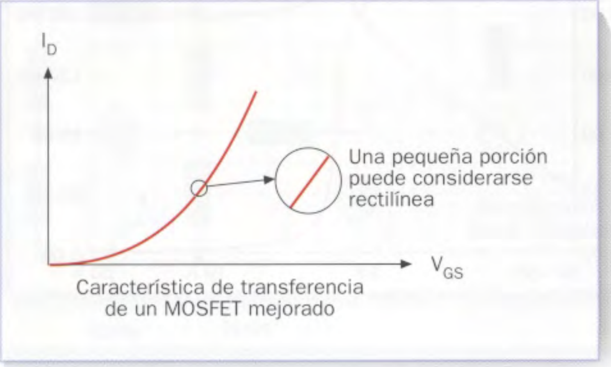

Una porción bastante **limitada** *de la curva (aquí la característica de un MOSFET) puede considerarse rectilínea.*

SISTEMAS Y ESTABILIDAD

A un nivel más alto, se aplican a los sistemas con **realimentación** otras técnicas de cálculo, por ejemplo a los de amplificadores operacionales: es importante, entre otras cosas, poder garantizar su estabilidad.

El hecho de que un circuito se estabilice más o menos rápidamente o, al contrario, tienda a **autooscilar**, puede determinarse según algunas características de las ecuaciones que lo describen ("ceros" y "polos").

Se han desarrollado métodos de análisis diversos para otros **campos de aplicación**, por ejemplo en alta frecuencia: cada técnica electrónica ha requerido el desarrollo de herramientas matemáticas adecuadas.

Gottfried Wilheim Leibniz, filósofo del 700, estuvo entre los principales descubridores del **cálculo infinitesimal**, *utilizado entre otras cosas en el análisis de los circuitos electrónicos.*

Simulaciones en el ordenador

Cálculos automatizados y pruebas virtuales facilitan el proyecto

El análisis circuital, al que hemos aludido en la lección 48, requiere un notable **compromiso**, tanto para la configuración de las ecuaciones que describen el circuito, como por el auténtico cálculo.

Tratándose de problemas bien definidos y de cálculos repetitivos, se pueden confiar a un **programa** para ordenador, dando así libertad al diseñador para que piense en las soluciones circuitales.

Un programa de simulación tiene también otra gran ventaja: permite **probar** los circuitos (al menos en una primera aproximación) sin construirlos físicamente... y sin riesgo de estropear componentes.

NACIMIENTO DE SPICE

Los programas de simulación se escribieron inicialmente para el diseño de los **circuitos integrados**, con el fin de reducir al mínimo la realización (lenta y bastante costosa) de chips de prueba.

Históricamente bastante **importante**, y aún hoy muy utilizado, es SPICE (Simulation Program with Integrated Circuit Emphasis: programa de simulación con énfasis en el circuito integrado), desarrollado a partir de 1970 a la Universidad de Berkeley en California. Evolucionado en versiones sucesivas, siempre ha permanecido de **distribución libre**; también se emplea como base de numerosos programas comerciales de ayuda al diseño.

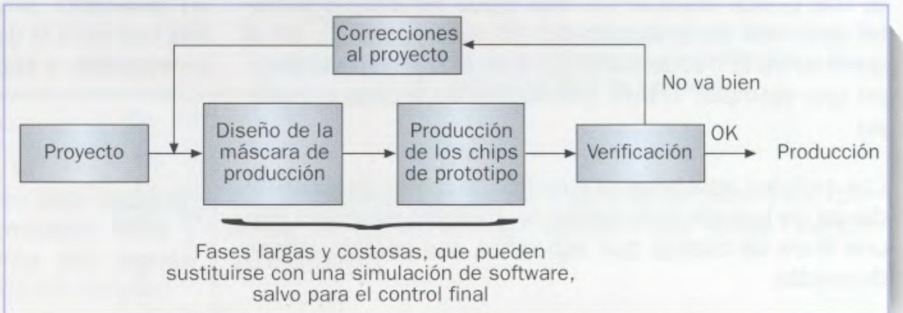
Correcciones al proyecto
No va bien
Proyecto
Diseño de la máscara de producción
Producción de los chips de prototipo
Verificación
OK
Producción
Fases largas y costosas, que pueden sustituirse con una simulación de software, salvo para el control final

El uso de un* programa de simulación *permite una drástica reducción de tiempos y costes de proyecto.

DEL ESQUEMA A LOS RESULTADOS

La figura ilustra las fases de una simulación con SPICE (o programas análogos) en primer lugar es necesario **describir** el circuito, listando los componentes utilizados y las conexiones relacionadas.

El programa simula cada componente con un **modelo** adecuado, es decir, una descripción idealizada pero razonablemente realista, y las ecuaciones relativas que describen su comportamiento.

Después analiza la **red** compuesta por varios modelos interconectados, resuelve las ecuaciones correspondientes (según el tipo de simulación requerida por el diseñador) y presenta los resultados de la forma deseada.

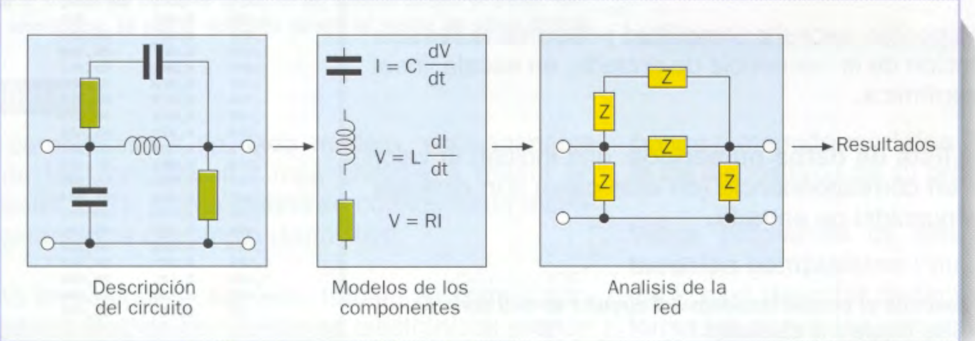
I = C dV/dt
V = L dI/dt
V = RI
Z
Z
Z
Z
Z
Resultados
Descripción del circuito
Modelos de los componentes
Analisis de la red

Fases de la* simulación circuital *con SPICE: los componentes se representan con modelos simplificados.

Netlist y análisis

La descripción del circuito es un sencillo archivo de texto

La lista de las conexiones, o "netlist", no es otra cosa que la transcripción, componente a componente, del **esquema eléctrico** del circuito cuyo comportamiento se desea simular.

Cada punto de interconexión del esquema, incluidas las entradas y las salidas, se numera y representa un **nodo** de la red circuital; un componente une dos o varios nodos.

La figura muestra un ejemplo de **transcripción** de un circuito en el lenguaje convencional de SPICE, que tiene varios convenios: por ejemplo, los componentes cuyo nombre empieza por "R" son resistencias.

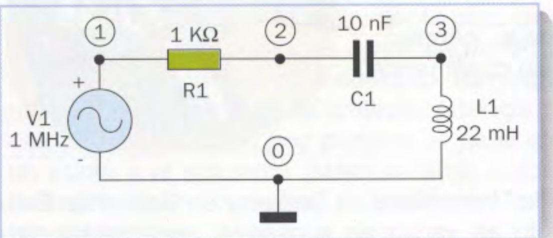

```
C1 3 2 10N
L1 3 0 22M
R1 1 2 1K
V1 1 0 AC 1 SIN (0 1 1e+6)
```

Cada línea de la **descripción** *contiene el nombre de un componente, los nodos a los que está conectado y sus características eléctricas.*

ANÁLISIS Y GENERADORES

SPICE puede efectuar los **tres tipos** de análisis circuital descritos en la lección 48: en continua (DC), en el dominio de la frecuencia (AC) y en el dominio del tiempo (por ejemplo, TRAN, generalmente la más compleja).

Las señales aplicadas al circuito proceden de **generadores** de tensión o de corriente, también descritos con una línea de código que especifica sus características deseadas.

Para el análisis en el dominio de la frecuencia se utiliza un generador sinusoidal de frecuencia variable, mientras que para la del dominio del tiempo se puede utilizar un **escalón**, o cualquier forma de onda.

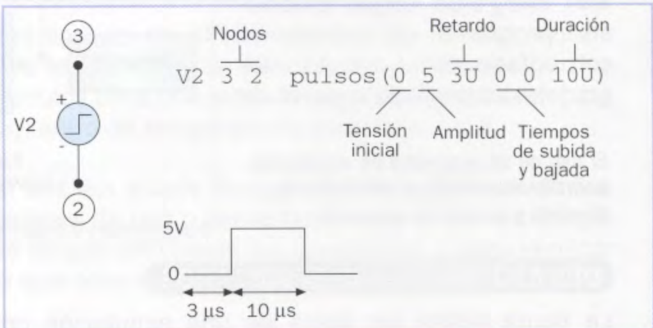

***Ejemplo de generador de impulso,** con la descripción de un escalón de tensión en el lenguaje de SPICE.*

RESULTADOS DEL ANÁLISIS

A la descripción del circuito se le añaden las **variables a examinar**, es decir, de qué valores se desea conocer el comportamiento al variar (según el tipo de análisis) la tensión, la frecuencia o el tiempo.

Por ejemplo, de un filtro se podrán necesitar **amplitud** y **fase** de la tensión en el nodo de salida, en función de la frecuencia de entrada, en escala lineal o (más probablemente) logarítmica.

El programa producirá una **lista de datos numéricos**, que indican el valor de las variables deseadas en correspondencia (en este caso) con distintas frecuencias de la tensión sinusoidal de entrada.

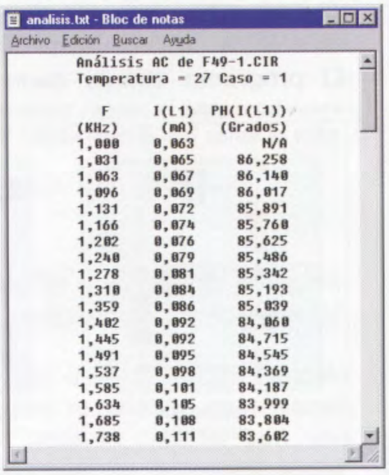

```
Análisis AC de F49-1.CIR
Temperatura = 27 Caso = 1

   F       I(L1)  PH(I(L1))
(KHz)       (mA)   (Grados)
1,000      0,063        N/A
1,031      0,065     86,258
1,063      0,067     86,140
1,096      0,069     86,017
1,131      0,072     85,891
1,166      0,074     85,760
1,202      0,076     85,625
1,240      0,079     85,486
1,278      0,081     85,342
1,310      0,084     85,193
1,359      0,086     85,039
1,402      0,092     84,060
1,445      0,092     84,715
1,491      0,095     84,545
1,537      0,098     84,369
1,585      0,101     84,187
1,634      0,105     83,999
1,685      0,108     83,804
1,738      0,111     83,602
```

***El archivo producido por SPICE describe el comportamiento del circuito en una serie de puntos de medida,** espaciados como requiera el diseñador.*

Gráficos y comodidad

Los modernos simuladores muestran los resultados de un modo fácilmente legible

Aunque SPICE sea un programa que lee un archivo de **texto** (la descripción del circuito y las especificaciones de medida) y produce otro con los resultados, es posible utilizarlo en modo gráfico.

Existen, de hecho, numerosos programas comerciales y no comerciales, que muestran los resultados de las simulaciones en una **pantalla gráfica** de alta definición, en forma de diagramas de distinto tipo.

Muchos de ellos permiten realizar el **trabajo completo**, desde el diseño del esquema hasta el examen de los resultados, sin ni siquiera ver el archivo de texto que contiene la descripción formal del circuito.

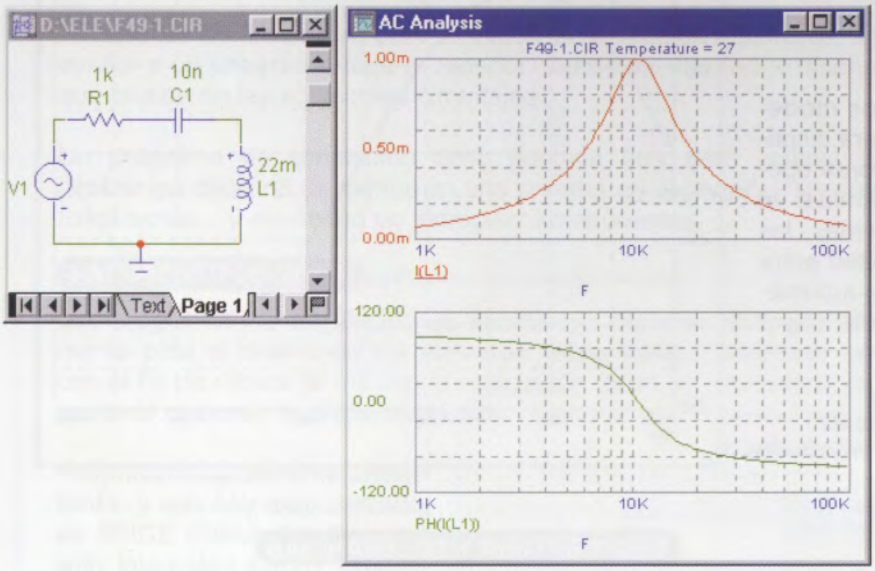

Los SPICE comerciales (aquí Micro-Cap V) permiten trabajar en un cómodo **entorno gráfico** ***de ventanas.***

DIGITALES E HÍBRIDOS

La simulación de un circuito digital no se basa generalmente en ecuaciones: resultaría demasiado lenta y compleja; lo que utiliza principalmente son **transiciones lógicas** y tiempos de tránsito de las señales.

No obstante, hay situaciones en las que los dispositivos analógicos funcionan en **estrecho contacto** con puertas lógicas, o en cualquier caso con integrados digitales: no se puede analizar el circuito entero con una sola de las dos técnicas.

Un buen simulador debe ser capaz de funcionar también en modo híbrido, o de **señal mezclada**, adecuando las técnicas de cálculo y simulación según los distintos componentes del circuito que están siendo examinados.

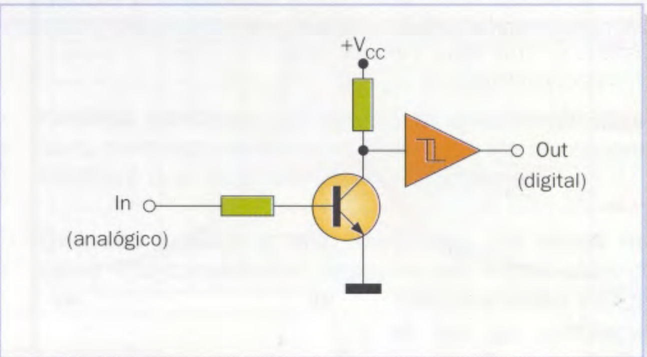

La parte izquierda de este circuito se simula desde el punto de vista **analógico*****, la parte derecha desde el punto de vista*** **digital.**

MODELOS

Las características de cada modelo, especialmente las de los componentes más complejos (como un transistor BJT), puede **definirse** configurando distintos parámetros del propio dispositivo.

Como en cualquier caso este trabajo es complicado, los fabricantes de componentes electrónicos proporcionan a menudo **modelos SPICE** ya preparados para los dispositivos de su producción.

Varios programas de simulación ofrecen, además **modelos compuestos** ("macromodelos" o "subcircuitos"), para describir de un modo sencillo dispositivos formados por varios componentes discretos.

Límites de las simulaciones

Un programa informático no es un dispositivo físico

En primer lugar es necesario **respetar las condiciones** requeridas por el tipo de simulación elegido: por ejemplo, un análisis en el dominio de la frecuencia varía normalmente sólo para pequeñas señales.

Es necesario, además, recordar que los **modelos** no son perfectos, especialmente para dispositivos complejos como los amplificadores operacionales; a veces es necesario cambiar el modelo según el tipo de análisis. Además, los cálculos se efectúan **por puntos**; el paso entre un punto y el siguiente (step) se elige automáticamente, pero conviene indicarlo para mayor precisión.

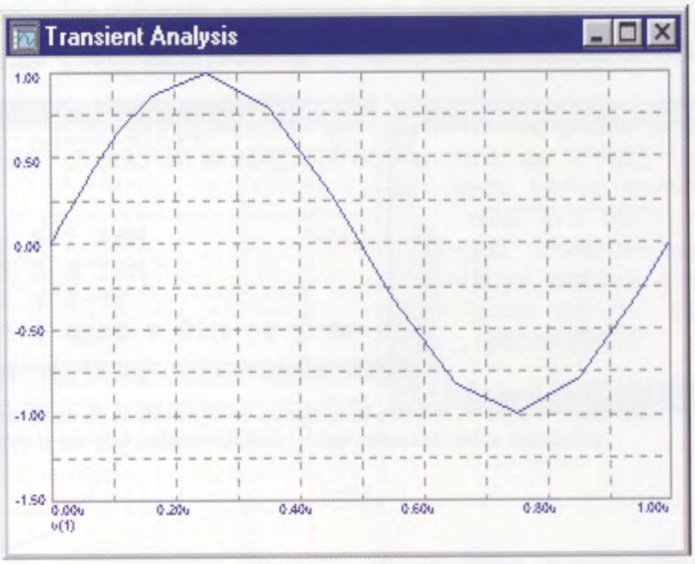

Una forma de onda alterada, como esta sinusoide distorsionada, puede ser indicativo de una inadecuada configuración de los parámetros de cálculo, o de visualización.

TEMPERATURA Y TOLERANCIAS

Los análisis de SPICE se efectúan normalmente a una temperatura de **27 °C**; se pueden efectuar los cálculos para una temperatura diferente, o bien para un rango de temperaturas completo.

El análisis puede también repetirse para **distintos valores** de uno o varios componentes, de modo que se tengan en cuenta las tolerancias o para verificar lo que ocurra al variar el mismo valor.

Un análisis "Monte Carlo", con referencia a la ruleta, se repite bastantes veces con **valores casuales** de los componentes (dentro de las tolerancias), para simular la distribución estadística entre los ejemplares de producción.

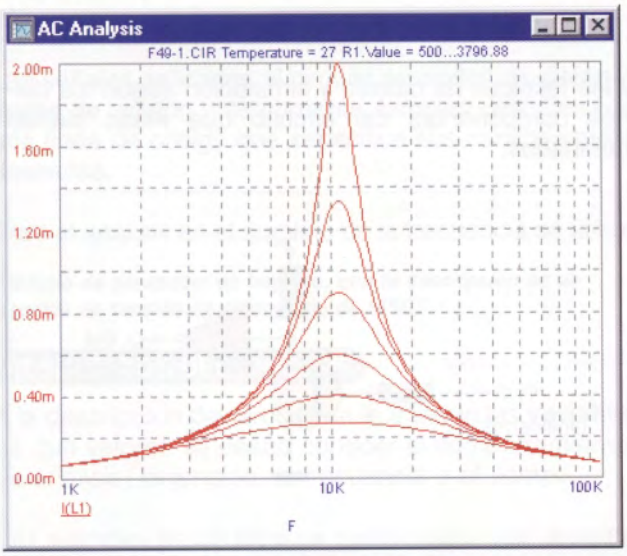

Esta familia de curvas indica cómo cambia el comportamiento del circuito de la página 194 (parte superior) al variar el valor de la resistencia R1.

SPICE NO BASTA

Los simuladores son herramientas bastante útiles, pero no son capaces de sustituir **completamente** la prueba práctica en un circuito físicamente construido, conectado a las herramientas de medida.

Es necesario, entre otras cosas, considerar que en el dispositivo real hay **componentes parásitos**, además de los dibujados en el esquema: por ejemplo, la capacidad entre resistencias adyacentes forma un condensador.
Aunque la experiencia sea de gran ayuda también en este caso, los mejores resultados se obtienen **combinando** un buen análisis circuital con una precisa verificación experimental.

A

A (amperio) 3
Abierto (circuito) 2
Acoplamiento capacitivo 19
Acoplamiento capacitivo (BJT) 42
Acumuladores 159
Adaptadores de antena 111
ADC a un bit (comparadores) 151
AGC (control de ganancia) 92
Ah (amperios-hora) 158
Alcalinas (pilas) 158
Alimentadores 32
Alimentadores estabilizados (Zener) 54
Alimentadores de conmutación 153
Altas frecuencias 109
Amperio 3
Amplificación (BJT) 38
Amplificador (BJT) 41
Amplificador (JFET) 82
Amplificador de audio 75
Amplificador de error (reguladores) 138
Amplificador inversor (op-amp) 131
Amplificador logarítmico 176
Amplificador no inversor (op-amp) 130
Amplificador para fotodiodos (op-amp) 147
Amplificadores 69
Amplificadores de aislamiento 185
Amplificadores de precisión (op-amp) 148
Amplificadores diferenciales 72, 148
Amplificadores operacionales 129
Análisis de las redes 190
Análisis de señal mezclada 195
Análisis digital 195
Análisis en alterna 191
Análisis en continua (estado estable) 191
Anchura de banda (filtros) 66
Anchura de banda (op-amp) 134
Anillo abierto/cerrado 73
Anillo de bloqueo de fase (PLL) 171
Ánodo (diodo) 29
Ánodo (válvulas) 177
Antenas 105
Anticipación (condensador) 45
Armónicos 62
Arrays analógicos 186
Arrays de antenas 107
ASIC (integrados a medida) 186
Atenuadores 12
Audio y ruido 161
Autodescarga (pilas) 158
Autoestabilización (BJT) 43
Autoinducción 22
Autooscilaciones 80
Autopolarización (JFET) 82
Autopolarización (válvulas) 178
Autotransformadores 27

B

Balun 111
Bandas laterales 100
Barrera de potencial 35
Base (BJT) 38
Baterías 157
Batimiento 103
Bessel (filtros) 68
Bi-cuad (filtros activos) 166
Bipolares (transistores) 37, 41, 49, 57
Bipolos lineales 189
BJT (transistores) 37, 41, 49, 57
Blindajes (altas frecuencias) 112
Bloqueo (PLL) 171
Bobinas 21
Body 85
Burn-in (circuitos integrados) 128
Butterworth (filtros) 68

C

Caída mínima (reguladores) 139
Cálculo circuital 189
Campo electromagnético 105
Campo magnético 21
Canal N/P (JFET) 81
Capacidad (baterías) 158
Capacidad (condensadores) 17
Capacidades conmutadas (switching) 155
Capacidades de puerta (MOSFET) 87
Capacidades internas (válvulas) 179
Carga (condensadores) 17
Cargas reactivas 48
Cargas residuales (IGBT) 122
Cátodo (diodo) 29
Cátodo (válvulas) 178
Cables coaxiales 110
CCD (sensores) 116
Celda electrovoltaica 157
Celdas analógicas 186
Cerrado (circuito) 2
Chebyshev (filtros) 68
Circuito resonante LC 61
Circuitos clásicos (BJT) 49
Circuitos clásicos (op-amp) 145
Circuitos de protección 141
Circuitos eléctricos 2
Circuitos integrados analógicos 125
Circuitos oscilantes 24
Circuitos resonantes 93
Clase A, clase B (amplificadores) 69
Clase AB (amplificadores) 70
Clase C (amplificadores) 98
CMOS (MOSFET complementarios) 88
CMRR (proporción de rechazo en modo común) 72, 148
Código Morse (recepción) 103
Código Morse (transmisión) 97
COFDM (radio digital) 188
Colector (BJT) 38
Colpitts 95
Conmutator analógico 89
Conmutatores de potencia (BJT) 60
Comparador de fase 172
Comparador de ventana 151
Comparadores 149
Compensación (op-amp) 134
Componente continua 16
Componentes activos 37
Componentes parásitos 196
Condensadores 17
Condensadores en paralelo y en serie 20
Condensadores de paso 112
Condensadores integrados 126
Condensadores variables 20
Conducción (BJT) 38
Conducción (diodo) 35
Contrafase 15
Control (audio) 75
Controladores digitales 175
Conversión de frecuencia 102
Conversor corriente-tensión (op-amp) 147
Conversor rms 176
Conversor tensión-corriente (op-amp) 147
Conversor tensión-frecuencia 171
Corriente alterna 3, 13
Corriente en vacío (reguladores) 139
Corriente de base y de colector (BJT) 38
Corriente de entrada (comparadores) 152
Corriente de pérdida (JFET) 84
Corriente eléctrica 1
Corriente mínima (switching) 156
Constante de tiempo 18
Construcción (BJT) 40
Construcción (JFET) 84
Cos φ (fi) 48
Criterio de Barkhausen 78
Cuadratura 15
Cuadripolos 190
Cuarzos 96
Culombio 17

D

DAB (radio digital) 188
Darlington 50
dB, dBm, dBV (decibelios) 63
De Forest 178
Decibelio 63
Densidad energética (baterías) 158

Depletion o vaciado (MOSFET) 86
Derivación térmica (amplificadores) 70
Desacoplamiento 80
Descarga (baterías) 157
Descargadores 144
Descargas electrostáticas 144
Descomposición (ondas) 64
Desfase (cargas reactivas) 48
Desfase (filtros) 65
Desmoduladores AM, FM 103
Devanados 25
di/dt (GTO) 123
Diac 124
Diagrama de irradiación (antenas) 107
Diferenciador 61
Diferenciadores (op-amp) 146
Dinámica 63, 163
Diodo de Fleming 177
Diodos 29
Diodos Gunn 124
Diodos parásitos (circuitos integrados) 126
Diodos PIN 124
Diodos tunnel 124
Diodos Zener 53
Dipolo 106
Directores (antenas) 107
Disipación 7
Disipación (alimentadores estabilizados) 55
Disipación (interruptor BJT) 50
Disipación (switching) 156
Disipadores 58
Distorsión (amplificadores) 70
Distorsión (audio) 70, 76, 161
Divisores 12
Doble conversión 104
Dominio de la frecuencia 64, 191
Dominio del tiempo 64
Dopaje (semiconductores) 34
Drenaje (JFET) 81
dV/dt (SCR) 118

E

Efecto Edison 177
Efecto Larsen 77
Efecto memoria (acumuladores) 159
Efecto Miller 60
Efecto piezoeléctrico 96
Efecto termoiónico 177
Efecto Zener 53
Efectos ambientales (audio) 173
Electromagnetismo 21
Electrones (semiconductores) 33
Elementos no lineales 64, 192
Eliminación de banda (filtros) 66
Emisor (BJT) 38
Emisor común 38
Encapsulamiento (semiconductores) 36
Energía acumulada (condensador) 18
Energía acumulada (inductor) 23
Enhancement o incremento (MOSFET) 86
Enmascarado (circuitos integrados) 127
Entrada en corriente (op-amp) 145
Entrada inversora y no inversora (op-amp) 129
Entradas para micrófono 164
ESD (descargas electrostáticas) 144
Espiras 21
Esquemas eléctricos 2
Estabilidad (op-amp) 134
Estabilización (amplificadores) 70
Estado estacionario sinusoidal 189
Eureka-147 (radio digital) 188
Excesos de tensión transitorios 144

F

F (faradio) 17
Factor de calidad 94
Factor de potencia 48
Familias de curvas 196
Faradio 17
Fase 15
Fenómenos parásitos (altas frecuencias) 109
Fiabilidad (circuitos integrados) 128
Fibras ópticas 116
Filamento (válvulas) 177
Filtro de anillo (PLL) 172
Filtro LC (alimentador) 30
Filtros antiinterferencias (alimentadores) 32
Filtros activos 165
Filtros de paso alto, de paso de banda, de eliminación de banda, supresor de banda 66
Filtros digitales 167
Filtros en cascada 68
Filtros LC 24
Filtros pasivos 65
Filtros RC 19
Filtros RL 23
Filtros RLC 68
Final (audio) 75
Forma cartesiana (números complejos) 47
Forma polar (números complejos) 47
Formas de onda 61
Foto-SCR y foto-triac 120
Fotoacopladores 115
Fotodiodos 114
Fotoresistencias 114
Fototransistores 115
FPAA (arrays analógicos) 186
Frecuencia 13
Frecuencia de corte (filtros) 65
Frecuencia de imagen 104
Frecuencia de resonancia 94
Frecuencia intermedia 102
Frecuencias moduladoras 100
Fuente (JFET) 81
Fuentes de tensión y corriente 190
Fundamental (onda) 62
Fusibles 141

G

Ganancia en modo común 72
Generador de corriente constante (BJT) 52
Generador de corriente constante (JFET) 82
Generador de corriente constante (op-amp) 147
Generadores (análisis de circuitos) 190
Girador 165
Ground plane o plano de tierra 106
GTO (transistores) 123
Guía de onda 110

H

H (henrio) 22
Hartley (osciladores) 95
Henrio 22
Hercio 13
Histéresis (comparadores) 150
Huecos (semiconductores) 33
Hz (hercio) 13

I

IGBT (transistores) 121
Impedancia 47
Impedancia característica (altas frecuencias) 110
Impedancia de salida (op-amp) 135
Inadaptación de impedancia 111
Inducción 25
Inducción mutua 25
Inductores 21
Integrador 61
Integradores (op-amp) 146
Integrados de audio 173
Integrados para coche 175
Integrados RF 174
Interdicción (BJT) 38
Intermodulación (audio) 161
Interruptor (BJT) 39
Interruptores (MOSFET) 87
Interruptores analógicos 89
Inversor (BJT) 39
Inversor (MOSFET) 88

J

JFET (transistores) 81
Joule (ley de) 7

K

Kilby (Jack S.) 125
Kirchhoff (leyes de) 9

L

Láser 116
Laser trimming o regulación de láser 128
LC en paralelo y en serie 94
Leclanché (pilas) 158
LED (diodos luminosos) 113
Ley de Joule 7
Ley de Ohm 5
Limitación de corriente (alimentadores estabilizados) 56
Limitador de corriente de FET 144
Líneas de transmisión 110
Litio (baterías) 158
Log-periódicas (antenas) 108
Lógica digital 88
LPC (síntesis vocal) 187
Longitud de onda (antenas) 105
Longitud de onda (luz) 113

M

Macroceldas analógicas 186
Macromodelos 195
Magnetrón 180
Mallas (redes) 190
Máquinas eléctricas 21
Masa virtual (op-amp) 131
Matemáticas analógicas 176
MEMS (microelectromecánica) 181
Mezclador (alta frecuencia) 102, 112
Microactivadores 183
Microelectromecánica 181
Microondas 108
Micropower (op-amp) 136
Microsensores 182
Microstrip 110
Modelos de los componentes 189
Modulación de amplitud y frecuencia 99
Modulación Morse 97
Monte Carlo (análisis) 196
MOS, MOSFET (transistores) 85
MOSFET complementarios 88
MOSFET de potencia 87
Muestreo y conservación 90
Multiplexor analógico 90
Multiplicador 176
Multiplicadores de frecuencia 98
MUSICAM (radio digital) 188

N

Nanotecnología 184
Negativo 2
Netlist 194
Níquel-cadmio e hidruro de níquel (acumuladores) 159
Nodos (redes) 190
NPN (transistores BJT) 38
Núcleo (inductores) 21
Núcleo (transformadores) 25
Números complejos 47

O

Oblea 36
Ohm (ley de) 5
Oído humano 63
Ondas cuadradas y triangulares 16
Op-amp 129
Optoelectrónica analógica 113
Orden de los filtros 68
Ordenadores analógicos 176
Oscilador bi-cuad 170
Oscilador local (conversores) 102
Oscilador triangular 169
Osciladores 77
Osciladores Colpitts y Hartley 95
Osciladores con operacionales 169
Osciladores de alta frecuencia 95
Osciladores de cuarzo 96
Osciladores de relajación y trigger Schmitt 79
Osciladores sinusoidales 78
Oxidación 157

P

Palanca 143
Pantalla activa 148
Parábola (antenas) 108
Parte real e imaginaria (números complejos) 47
Paso alto, paso de banda (filtros) 66
Paso bajo (filtros) 65
Paso parcial (interruptores analógicos) 92
Paso por el cero 151
Paso por el cero (triac) 119
Período 13
Pequeñas señales (análisis para) 192
Pila de Volta 157
Pilas 157
Pilas de combustible (baterías) 158
Pinch-off (JFET) 83
PLL (bloqueo de fase) 171
Plomo-ácido (acumuladores) 159
PNP (transistores BJT) 38
Polaridad 21
Polaridad (MOSFET) 85
Polaridad (transformadores) 27
Polarización (antenas) 108
Polarización (BJT) 42
Polarización inversa (diodo) 35
Poliméricos PTC 142
Portadora (transmisores) 97
Portadores (semiconductores) 33
Positivo 2
Potencia 7
Potencia aparente y real 48
Potenciómetros 12
Preamplificador 75
Primario (transformadores) 26
Problemas térmicos (BJT de potencia) 57
Proceso digital (audio) 164
Producción (circuitos integrados) 126
Producción (semiconducores) 36
Programas de análisis 193
Protección (amplificadores) 71
Protecciones 141
Protecciones (reguladores) 138
Protecciones recuperables 142
Punto de trabajo (BJT) 43
Puerta (JFET) 81
Puerta (SCR) 117
Push-pull (amplificadores) 69
Push-push (duplicadores de frecuencia) 98

Q

Q (factor de calidad) 94

R

Radianes 15
Radio digital 188
Rápidos (fusibles) 141
RC (constante de tiempo) 18
Reacción positiva 77
Reactancia 46
Realimentación 73
Realimentación (op-amp) 131
Realimentación de corriente (op-amp) 136
Realimentación negativa 73
Realimentación negativa (op-amp) 130
Realimentación positiva (comparadores) 150
Recarga (acumuladores) 159
Recarga con tampón (acumuladores) 160
Receptores 101
Reciprocidad UC 24
Recorte 161

Rectificadores 29, 31
Rectificadores (op-amp) 147
Redes lineales 190
Reducción de ruido 173
Reducción química 157
Refase 48
Reflector (antenas) 107
Reflexión (altas frecuencias) 110
Regiones (semiconductores) 35
Reguladores de tensión lineal 137
Reguladores de tres terminales 137
Reguladores dobles, negativos y variables 140
Rejilla (válvulas) 178
Rejilla de pantalla 179
Rejilla de supresión 179
Relación de transformación (transformadores) 26
Relación señal/interferencia 163
Rendimiento (semiconductores) 36
Repliegue 142
Resistencia 5
Resistencia controlada en tensión 91
Resistencia de conducción (MOSFET) 87
Resistencia negativa (diodos tunnel) 124
Resistencia térmica 58
Resistencias 5
Resistencias en paralelo 11
Resistencias en serie 10
Resistencias integradas 126
Resistencias no deseadas 8
Resonancia (filtros) 68
Resonancia (cuarzos) 96
Respuesta en frecuencia (audio) 76
Respuesta en frecuencia (filtros) 65
Retraso (inductor) 45
Reversibilidad (transformadores) 26
rms (valor eficaz) 14
Ruido 1/f, blanco, impulsivo, rosa 162
Ruido (audio) 161
Ruido (switching) 156

S

S (siemens) 83
S/N (relación señal/interferencia) 163
Saturación (BJT) 39
Schottky, transistores 60
SCR (diodos controlados) 117
Secundario (transformadores) 26
Seguidor de emisor 49
Seguidor de tensión (op-amp) 130
Segunda ruptura 57
Selectividad 101
Semiconductores 33
Semiretardados (fusibles) 141
Sensibilidad electrostática (MOSFET) 85
Sensores CCD 116
Señal mezclada 186
Serie de Fourier 62
Siemens (conductancia) 83
Silicio 36
Simulación circuital 193
Síntesis de frecuencia 104
Síntesis vocal 187
Sintonía 101
Sinusoide 13, 15, 61
Slew rate 134
Snubber 120
Soldadura 127
SPICE (simulación circuital) 194
Step-up (switching) 154
Stripline 110
Suma de sinusoides 62
Sumadores (op-amp) 145
Supercondensadores 159
Superheterodina 102
Supresor de banda 66
Sustrato (semiconductores) 85
SWR (ondas estacionarias) 111

T

Tasa de producción (circuitos integrados) 128
Tensión 3
Tensión de celda (baterías) 157
Tensión de offset (comparadores) 152
Tensión de offset (op-amp) 132
Tensión de referencia (reguladores) 138
Tensión excesiva en el cierre 24
Tensión excesiva en la apertura 24
Tensión inversa E-B 52
Tensión pulsante 30
Termomagnética (protección) 142
Tetrodo 179
Tiempo de conmutación (MOSFET) 87
Timbre 62
Tipos de op-amp 136
Transconductancia 83
Transistores BJT 37, 41, 49, 57
Transistores de potencia 57
Transformadores 25
Transformadores ajustados 112
Transformadores de alimentación y de señal 28
Transmisores 97
Triac 119
Trigger Schmitt (comparadores) 150
Triodo 178
Tubos de vacío 177

U

Ultrarápidos (fusibles) 141
Unidad de medida 4
Unión PN 35

V

V (voltio) 3
V/µs (velocidad de subida) 134
V/$\sqrt{\text{Hz}}$ (ruido) 162
Vaciado (JFET) 81
Valor de pico, eficaz, medio, de pico a pico 14
Válvulas termoiónicas 177
Varactor 99
Varistores 144
VCO (osciladores controlados) 79, 99, 171
VCVS (filtros activos) 166
Velocidad de respuesta (comparadores) 149
VFC (conversores tensión-frecuencia) 171
Vigilancia de las entradas 148
Voltamperio 48
Voltio 3

Y

Yagi (antenas) 107

Z

Zinc-carbón (baterías) 158
Zona lineal (FET) 91
Zona operativa segura (safe operating area) 57
Ω (ohmio) 5

Flamman y Lunamara

http://suspirando-estrellas.blogspot.com

http://flammancronno.blogspot.com

http://videosflamman.blogspot.com

MORE ELECTRONIC GADGETS FOR THE EVIL GENIUS

40 NEW BUILD-IT-YOURSELF PROJECTS

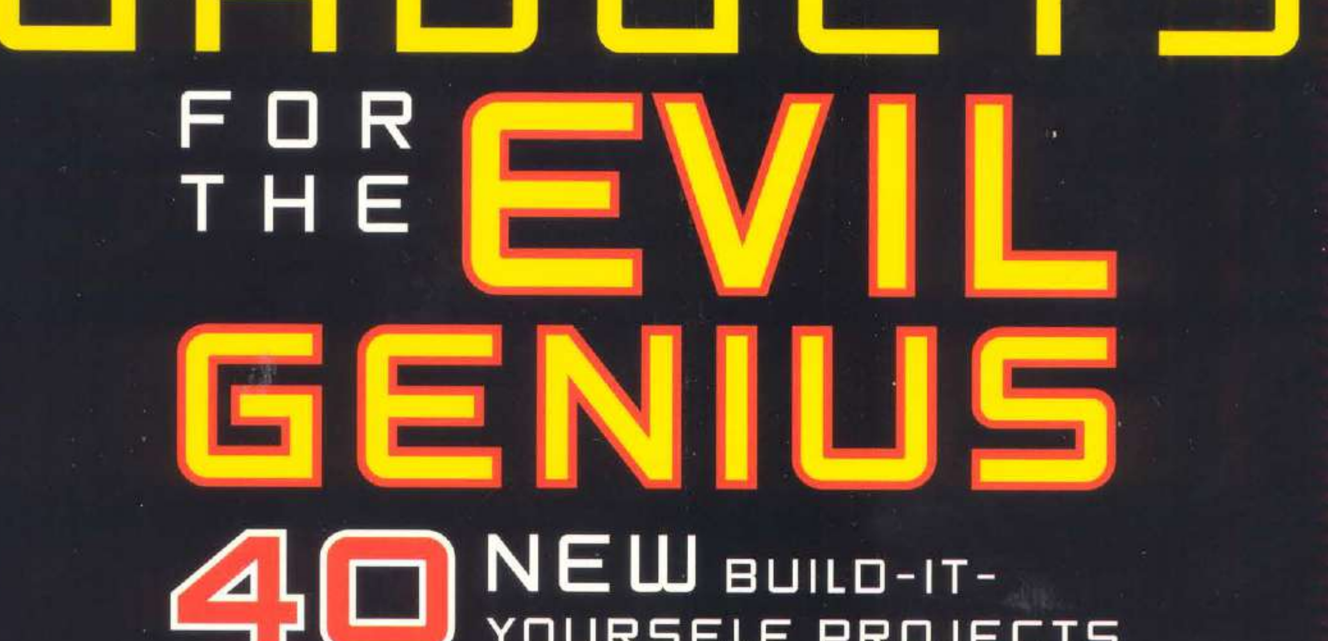

- **Every project includes a list of materials, sources for obtaining parts, schematics, documentation, and clear instructions**
- **Projects are fun and inexpensive, but challenging enough for an electronics wiz**
- **Build your own lasers, surveillance devices, ultrasonic generators, display and detection devices**

TAB
ELECTRONICS

BOB IANNINI

Contents

Chapter One Battery-Powered Infrared Pulsed Laser 1
Theory of Operation 2
Circuit Theory of Operation 2
Current Monitor 5
Charging Circuits 5
Chassis Assembly 5
Assembly Testing 7
Final Assembly 8
Special Note 8

Chapter Two High-Speed Laser Pulse Detector 11
Circuit Description 12
System Assembly 13
Assembly Testing 15
Final Assembly 16
Special Note 17
Special Note on Photo Detectors 17

Chapter Three Ultra-Bright Green Laser Project 19
Assembly Steps 20

Chapter Four 115 VAC, 5- to 50-Watt Pulsed Infrared Laser 25
Theory of Operation 25
Circuit Theory of Operation 26
Notes on SCR 28
Notes on the Storage Capacitor 28
Notes on Layout Wiring 28
Notes on Circuit Monitoring 30
Assembly Steps 30
Assembly Testing 33
Final Assembly Steps 35

Chapter Five Laser Property Guard 37
System Installation Suggestions 39

Chapter Six 30-Milliwatt, 980-Nanometer Infrared Laser 43
Assembly 44
Collimator and Final Assembly 47

Chapter Seven High-Voltage, High-Frequency Driver Module 51
Electrical and Mechanical Specifications 51
Circuit Description 52
Construction 52
Board Assembly Steps 52
Testing Steps 56

Chapter Eight Negative Ion Machine with Reaction-Emitting Rotor 59
Benefits of Negative Ions 59
Circuit Description 60
Assembly Steps 61
Notes 62

Chapter Nine Kirlian Imaging Project 65
Assembly Steps 66
Testing Steps for the Circuit 66
Notes 67

Chapter Ten Plasma Etching and Burning Pen 69
Driver Circuit Description 69
Drive Circuit Assembly 70
Driver Board Testing 71
Mechanical Assembly 72
Sample Demonstration 72

Chapter Eleven How to Electrify Objects and Vehicles 75
Supplementary Electrification Data and Information 76
Assembly of Suggested Power Sources 81
Circuit Description of Pulsed Shocker 81
Construction Steps for Pulsed Shocker 81
Notes Pulsed Shocker 85
Construction of Continuous Shocker 85

Chapter Twelve Electromagnetic Pulse (EMP) Gun 87
Basic Theory Simplified 87
Circuit Description 89
Circuit Assembly 89
Notes 89

Chapter Thirteen Microwave Cannon 93

Chapter Fourteen Induction Heater 103
Circuit Operation 103
Project Assembly 104
Project Testing 108
Using the Heater Project 108

Chapter Fifteen 50-Kilovolt Laboratory DC Supply 113
Basic Description 113
Circuit Operation 114
Assembly of the Driver Section 114
Pretesting the Driver Section 116
Assembly of the Multiplier Section 121
Final Testing 125
Applications 129

Chapter Sixteen Magnetic High-Impact Cannon 133
Theory of Operation 133
Circuit Theory 134
Assembly Steps 136
Circuit Testing 136
Assembly of the Completed System 137
Testing and Operation of the Cannon 143

Chapter Seventeen Vacuum Tube Tesla Coil Project 151
Basic Description 153
Assembly Steps 154
Testing and Operating 167
Notes 167

Chapter Eighteen Universal Capacitance Discharge Ignition (CDI) Driver 169
Circuit Description 169
Assembly Steps 170
Test Steps 175
Notes on Operation 175

Chapter Nineteen Long-range Telephone Conversation Transmitter 177
Circuit Description 177
Assembly of the Circuit Board 179
Notes 180

Chapter Twenty Line-Powered Telephone Conversation Transmitter 183
Circuit Description 183
Assembly of the Circuit Board 185
Notes 186

Chapter Twenty-One Remote Wireless FM Repeater 187
Circuit Description 187
Assembly 188
Notes 189

Chapter Twenty-Two Tracking and Homing Transmitter 191
Circuit Description 191
Assembly of the Circuit Board 192
Notes 194

Chapter Twenty-Three Snooper Phone Room-Listening Device 195
Circuit Operation 195

Construction Steps 196
Final Assembly 197
Testing Steps 198
Notes 199

Chapter Twenty-Four Long-Range FM Voice Transmitter 201
Circuit Description 201
Assembly Steps 202
Notes 204

Chapter Twenty-Five FM Pocket Radio and TV Disrupter 205
Circuit Theory 206
Assembly Steps 206

Chapter Twenty-Six Ozone Generator for Water Treatment 209
Introduction to the Benefits of Ozone 209
Ozone for Water Applications 210
Ways to Generate Ozone 210

Chapter Twenty-Seven Therapeutic Magnetic Pulser 213
Circuit Description 214
Assembly 215

Chapter Twenty-Eight Noise Curtain Generator 219
Circuit Description 220
Assembly Steps 220
Testing Steps 221

Chapter Twenty-Nine Mind-Synchronizing Generator 223
Information on Mind Control 223
Operating a Mind Machine 225
Circuit Theory 226
Assemble the Board 227

Chapter Thirty Alternative Health Multiwave Machine 231
Theory of Operation 231
Construction of the Multiwave Coil and Antenna 233
Operation 234

Chapter Thirty-One Mind Mangler 237
Circuit Operation 237
Construction Steps 238
Testing 239

Chapter Thirty-Two 500-Milligram Ozone Air Purification System 241
The Story of Ozone 241
Ozone for Treating Air 242
Construction Steps 242
Special Notes 244

Chapter Thirty-Three Invisible Pain-field Generator 249
Circuit Description 250
Construction 251

Chapter Thirty-Four Canine Controller 257
Device Description 257
Circuit Description 258
Circuit Assembly 260
Testing the Circuit 263

Chapter Thirty-Five Ultrasonic Phaser Pain-Field Generator 265
Basic Device Description 265
Circuit Theory 266
Construction Steps 267
Testing the Assembly 268
Basic Operating Instructions 269
Information on the System 272
A Word of Caution 272
General Information on Ultrasonics 272
Application Supplement 272

Chapter Thirty-Six Magnetic Distortion Detector 275
Circuit Description 276
Assembly Steps 277
Pretesting the Sense Head 278

Construction of the Control Box 279
Control Box Protesting 279
Applications 281
Optional Ultra-Sensitive Vibration and Tremor Detector 283
Chapter Thirty-Seven Body Heat Detector 285
Description 285
Body Heat Detector Modes 286
Circuit Description 286
Assembly Instructions 288
Operating Instructions 290
Final Notes 291
Chapter Thirty-Eight Ion and Field Detector 293
Circuit Description 293
Construction Steps 294
Applications 298
Chapter Thirty-Nine Light Saber Recycling Stand 301
Circuit Description 301
Assembly 302
Parts Fabrication 302
Operation 303
Chapter Forty Andromeda Plasma Lamp 309
Circuit Description 309
Assembly Board Wiring 310
Index 317

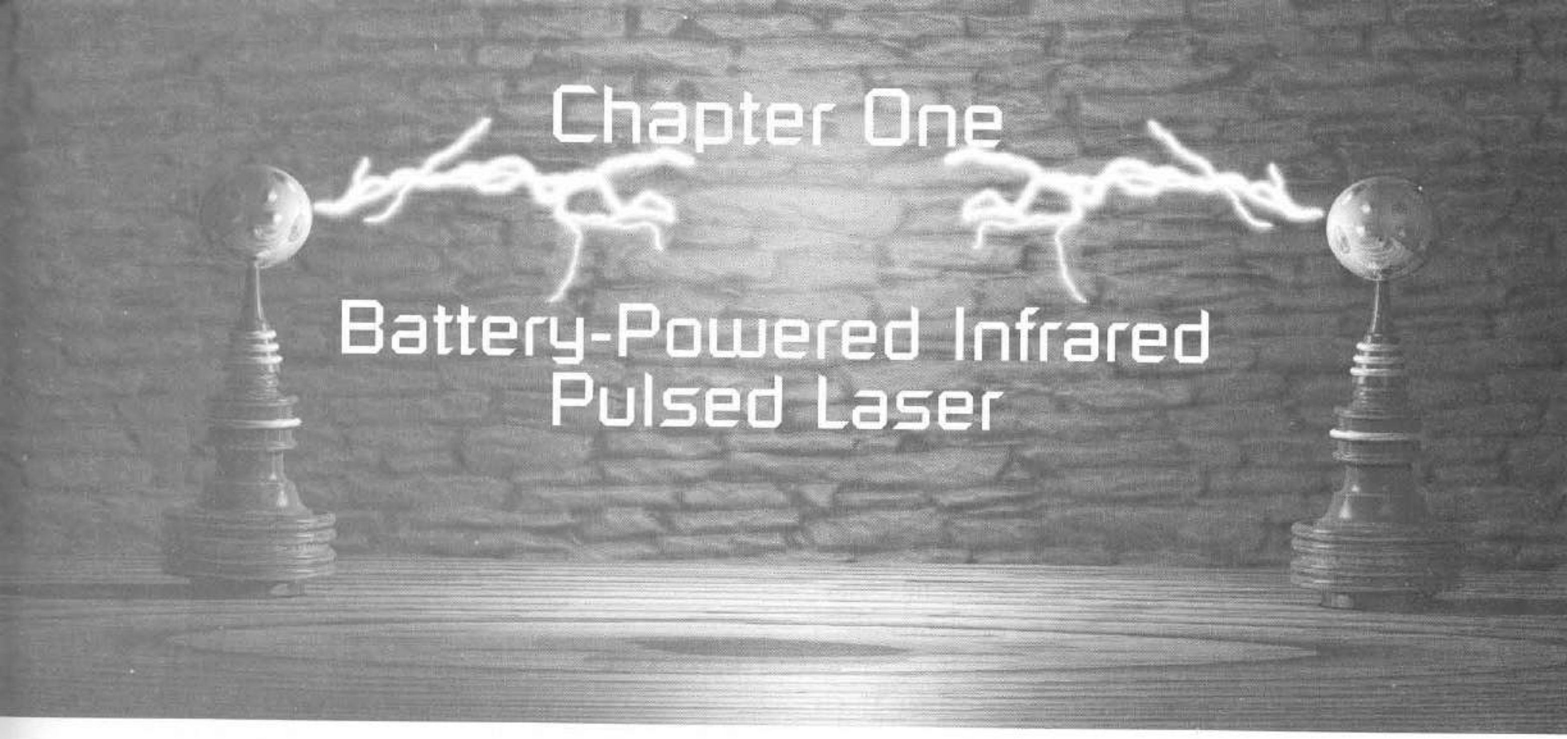

Chapter One

Battery-Powered Infrared Pulsed Laser

This project shows how to construct a low-powered, portable, solid-state, pulsed infrared laser. The system uses a gallium arsenide laser diode and provides pulse powers from 10 to 100 watts, depending on the diode used. The device operates from batteries and is completely self-contained. It may be built in a pistol, rifle, or a simple tubular configuration. The device is intended as a source of adjustable-frequency pulses from 10 to 1,000 repetitions per second of infrared energy at 900 nanometers

The system is shown in Figure 1-1 as being built on a perforated board assembly and combination copper ground plane that all fits into a tubular enclosure with a lens and holder for a collimator. The housing serves as the enclosure for the batteries and contains the control panel at its rear. It may be fitted with a handle that can hold an optional trigger switch when the device is designed in a gun configuration. A conventional sighting system is also easily adapted to the device.

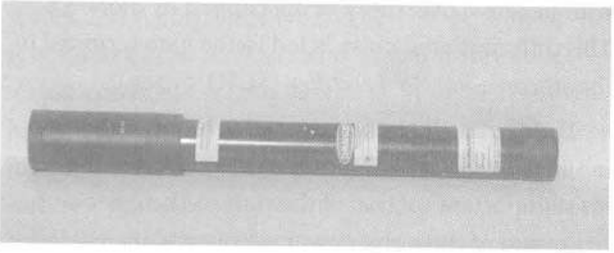

Figure 1-1 *Photograph of laser*

The laser is intended as a rifle-type simulated weapon, whose range can be several miles or as a long-range, laser-beam protection fence with a similar range of several miles. It is intended to be used with our high-speed laser pulse detector described in Chapter 2.

The laser, when assembled as shown, is a class 3B FDA certified device and requires the appropriate labeling and several included safety functions, as described in the assembly instructions. At no time should it be pointed at anyone without protective eyewear or at anything that could reflect the pulses. Never look into the unit when the power is on. The device is intended to be used for ranging, simulated weapons practice, intrusion detection, communications and signaling, and a variety of related scientific, optical experiments and uses.

This is an intermediate- to advanced-level project requiring electronic skills and basic electronic shop equipment. Expect to spend $100 to $150. All parts are readily available, with specialized parts obtainable through Information Unlimited (www.amazing1.com), and they are listed in Table 1-1.

Theory of Operation

A laser diode is nothing more than a three-layer device consisting of a *pn* junction of *n*-type silicon, a *p* type of gallium arsenide, and a third *p* layer of doped gallium arsenide with aluminum. The *n*-type material contains electrons that readily migrate across the *pn* junction and fill the holes of the *p*-type material. Conversely, holes in the *p* type migrate to the *n* type and join with electrons. This migration causes a potential hill or barrier consisting of negative charges in the *p*-type material and positive charges in the *n*-type material that eventually ceases growing when a charge equilibrium exists. In order for current to flow in this device, it must be supplied at a voltage to overcome this potential barrier. This is the forward voltage drop across a common diode. If this voltage polarity is reversed, the potential barrier is simply increased, assuring no current flow. This is the reversed bias condition of a common diode.

A diode without an external voltage applied to it contains electrons that move and wander through the lattice structure at a low, lazy average velocity as a function of temperature. When an external current at a voltage exceeding the barrier potential is applied, these lazy electrons increase their velocity so that some of them, by colliding, acquire a discrete amount of energy and become unstable. They eventually emit the acquired energy in the form of a photon after returning to a lower-energy state. These photons of energy are random both in time and direction; hence, any radiation produced is incoherent, such as that of an LED.

The requirement for coherent radiation is that the discrete packets of radiation must be in the form of a lockstep phase and in a definite direction. This demands two essential requirements: first, sufficient electrons at the necessary excited energy levels, and second, an optical resonant cavity capable of trapping these energized electrons to stimulate more electrons and give them direction. The amount of energized electrons is determined by the forward diode current. A definite threshold condition exists where the device emits laser light rather than incoherent light, such as in an LED. This is why the device must be pulsed with high current. The radiation from these energized electrons is reflected back and forth between the square-cut edges of the crystal that form the reflecting surfaces due to the index of refraction of the material and air.

The electrons are initially energized in the region of the *pn* junction. When these energized electrons drift into the *p*-type transparent region, they spontaneously liberate other photons that travel back and forth in the optical cavity interacting with other electrons, commencing laser action. A portion of the radiation traveling back and forth between the reflecting surfaces of these mirrors escapes and constitutes the output of the device.

Circuit Theory of Operation

Figure 1-2 shows the inverter section increasing the 12 volts of the portable battery pack to 200 to 300 volts performed by the circuit, which consists of a switching transistor (Q1) and step-up transformer (T1). Q1 conducts until saturated for a time until the base reverse biases and can no longer sustain it at an on state and Q1 turns off. This causes the magnetic field in its collector winding (COL) to collapse, thus producing a stepped-up voltage in the secondary (SEC) of proper phase. The variable resistor (R3) controls the charging current to the capacitor (C2), the transistor turnoff time, and consequently system power.

The stepped-up square wave voltage on the secondary of T1 is rectified by diode (D1) and integrated onto the storage capacitor (C3). The trigger circuit determines the pulse rep rate of the laser and uses a timer (I1) whose pulse rate is determined by the timing resistor and capacitor (R9 and C6). R9 is adjustable for changing the laser pulse rate. The output trigger pulse is differentiated by capacitor (C4) with negative overshoot being clipped by diode D2. This differentiated pulse is fed to the gate terminal of the *silicon controlled rectifier* (SCR) switch.

The discharge circuit generates the current pulse in the laser diode (LD1) and consequently is the most important section of the pulser. The basic configuration of the pulse power supply is shown in the system schematic. The current pulse is generated by the charging storage capacitor C3 being switched through the SCR and laser diode LD1. The rise time

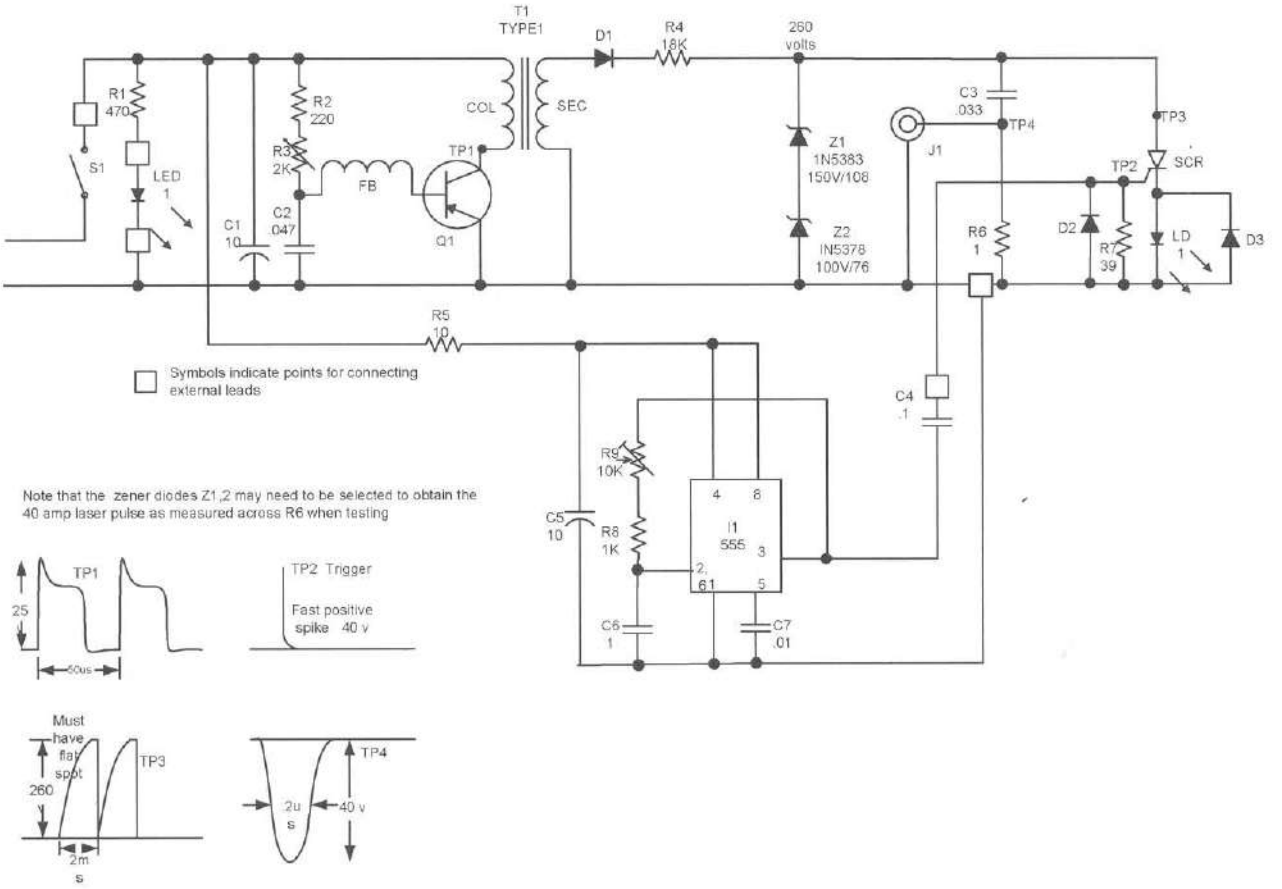
T1
TYPE1
D1
R4
18K
260
volts
R1
470
S1
LED
1
R2
220
R3
2K
COL
SEC
TP1
FB
Q1
C1
10
C2
.047
Z1
1N5383
150V/108
Z2
1N5378
100V/76
J1
C3
.033
TP4
TP3
TP2
SCR
R6
1
D2
R7
39
LD
1
D3
R5
10
Symbols indicate points for connecting external leads
C4
.1
R9
10K
R8
1K
C5
10
4
8
I1
555
3
2,
61
5
C6
1
C7
.01
Note that the zener diodes Z1,2 may need to be selected to obtain the 40 amp laser pulse as measured across R6 when testing
TP1
25
50us
TP2 Trigger
Fast positive spike 40 v
Must have flat spot
TP3
260
2m s
TP4
2u s
40 v

Figure 1-2 *Circuit schematic*

of the current pulse is usually determined by the SCR, while the fall time is determined by the capacitor value and the total resistance in the discharge circuit.

Figure 1-3 shows the typical anode voltage and current waveforms of the SCR during the current pulse through the diode laser. The peak current, pulse width, and voltage of the capacitor discharge circuit are related for various load and capacitance values. The peak laser current and charged capacitor voltage relationships are given for several different capacitor values and typical laser types. The voltage and current limits of the SCR are also shown. Short pulse widths provide less time for the SCR to turn on than longer pulse widths; therefore, the SCR impedance is higher and more voltage is required to generate the same current. Also shown are the current pulse waveforms for the three different values of the capacitance. The capacitor is charged to the same voltage in all three cases, that is, 400 volts.

In conventional SCR operation, the anode current, initiated by a gate pulse, rises to its maximum value in about 1 microsecond. During this time, the anode-to-cathode impedance drops from an open circuit to a fraction of an ohm. In injection laser pulsers, however, the duration of the anode-cathode pulse is much less than the time required for the SCR to turn on completely. Therefore, the anode-to-cathode impedance is at the level of 1 to 10 ohms throughout most of the conduction period.

The major disadvantage of the high SCR impedance is that it causes low circuit efficiency. For example, at a current of 40 amps, the maximum voltage would be across the SCR, while only 9 volts would be across the laser diode. These values represent very low circuit efficiency.

The efficiency of a laser array is greater due to its circuit impedance being more significant. Because the SCR is used unconventionally, many of the standard specifications such as peak current reverse

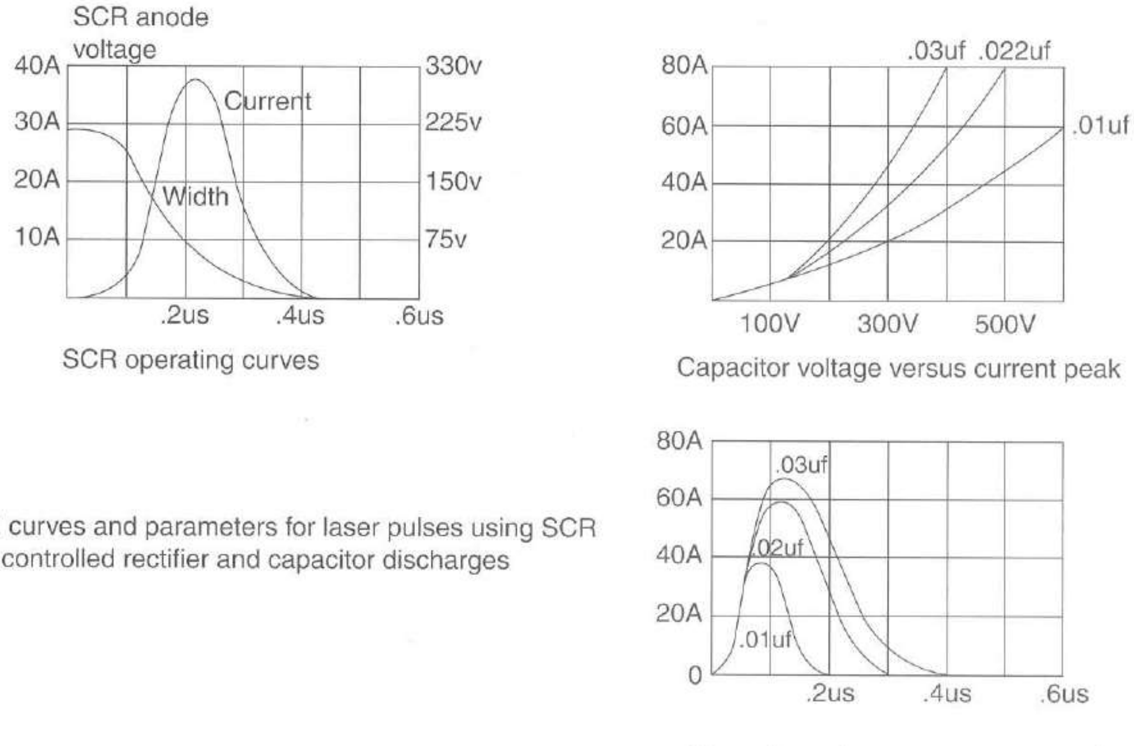

SCR operating curves

Capacitor voltage versus current peak

Typical curves and parameters for laser pulses using SCR silicon controlled rectifier and capacitor discharges

Capacitor value versus current peak

Figure 1-3 *SCR operating curves*

voltage, on-state forward voltage, and turn-off time are not applicable. In fact, it is difficult to select an SCR for a pulsing circuit on the basis of normally specified characteristics. The specifications important to laser pulser applications are the forward-blocking voltage and current rise time. A *use test* is the best and often the only practical method of determining the suitability of a particular SCR.

The voltage rating of the storage capacitor must be at least as high as the supply voltage. With the exception of ceramic types, most capacitors (metallized paper, mica, etc.) will perform well in this circuit. Ceramic capacitors have noticeably greater series resistance but are usable in slower speed pulsing circuits.

Lead lengths and circuit layout are very important to the performance of the discharge circuit. Lead inductance affects the rise time and peak value of the current, and it can also produce ringing and undershoot in the current waveform that can destroy the laser. A well-built discharge circuit might have a total lead length of only 1 inch and therefore an inductance of approximately 20 nanohenries. If the current rises to 75 amperes in 100 nanoseconds, the inductive voltage drop across this lead can be 15 volts. It is now obvious that if proper care is not taken in wiring the discharge circuits, high-inductive voltage drops will result.

A 1-ohm resistor in the discharge circuit will greatly reduce the current undershoot in single-diode lasers. Laser arrays usually have sufficient resistance to eliminate undershoot. The small resistance in the discharge circuit is also useful in monitoring the laser current, as described in the following section.

A clamping diode (D3) is added in parallel with the laser to reduce the current undershoot. Its polarity should be opposite that of the laser. Although the clamping diode is operated above its usual maximum current rating, the current undershoot caused by ringing is very short and the operating life of the diode is satisfactory.

Current Monitor

The current monitor in the discharge circuit provides a means of observing the laser current's waveform with an oscilloscope. A resistive-type monitor (R6) reduces circuit ringing and current undershoot, but the lead inductance of the resistor may cause a current reading that is higher than it actually is. A current transformer such as the Tektronix CT-2 can also be used to monitor the current and is not affected by lead inductance. Because the transformer does not respond to low-frequency signals, it should be used with fast waveforms that have a short pulse width and a fast fall time.

Charging Circuits

The second major section of the pulser is the charging circuit. This circuit charges the capacitor to the supply voltage during the time interval between the laser current pulses. It also isolates the supply voltage from the discharge circuit during the laser current pulse, thereby allowing the SCR to recover to the blocking state. Because the response times of the charging circuit are relatively long, lead lengths are not important and the circuit can be remotely located from the discharge circuit.

The simplest charging circuit is a resistor-cap combination. The resistor must limit the current to a value less than the SCR holding current, but it should be as low as practical, because this resistance also determines the charging time of the capacitor, C3.

Chassis Assembly

The following is a list of steps to construct the laser project.

1. Fabricate the copper chassis, as shown in Figure 1-4. This part is shown in Table 1-1 as being available.

2. Assemble the board as shown in Figure 1-5. Lay out and identify all the parts and pieces. Separate the components that go to the assembly board and fabricate the (PB1) perfboard from a 1.4 × 5-inch piece of .1 × .1 grid. Note the two holes for attachment to the copper chassis section from Figure 1-6.

3. Assemble the components as shown and observe the polarity of the capacitors and semiconductors.

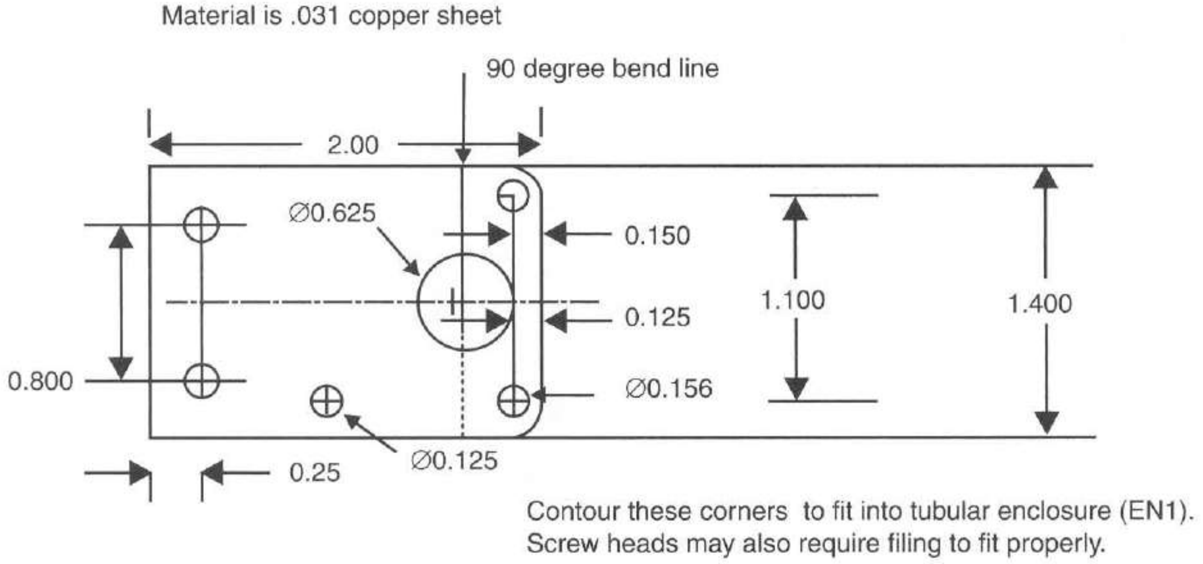
Material is .031 copper sheet
90 degree bend line
2.00
Ø0.625
0.150
0.125
1.100
1.400
0.800
Ø0.156
0.25
Ø0.125
Contour these corners to fit into tubular enclosure (EN1).
Screw heads may also require filing to fit properly.

Figure 1-4 *Fabrication of laser chassis*

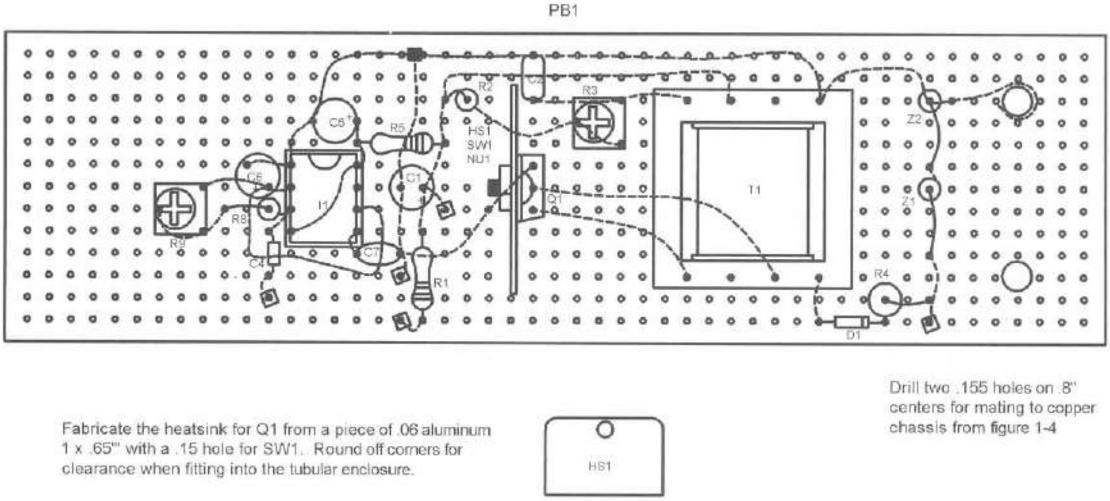

Fabricate the heatsink for Q1 from a piece of .06 aluminum 1 x .65" with a .15 hole for SW1. Round off corners for clearance when fitting into the tubular enclosure.

Drill two .155 holes on .8" centers for mating to copper chassis from figure 1-4

∼ Dashed lines are connections on underside of perforated circuit board using the component leads. These points can also be used to determine foil runs for those who wish to fabricate a printed circuit board.

• Small dots are holes used for component insertion.

■ Black squares are solder junctions

▣ Squares with dots are holes that may require drilling for component mounting and points for external connecting wires indicating strain relief points. It is suggested to drill clearance holes for the leads and solder to points beneath board.

Figure 1-5 *Assembly board parts identification*

4. Connect the components using their actual leads wherever possible. Follow the layout and avoid bare-wire bridges.
5. Figure 1-6 shows the wiring of the copper chassis (COPCHA1). This type of wiring is preferred for high-frequency, sensitive, low-noise circuitry. It is used in this circuit due to the high-current pulses with fast rise times.

   Start by forming the leads and soldering to the copper plate. You will need a good hot soldering iron to do this correctly. The components themselves serve as anchor points and will be self-supporting when soldered to one another.
6. Figure 1-7 shows the interconnection of the copper chassis and the perforated assembly board via the two screws and nuts. Attach and connect the external leads and components as shown.

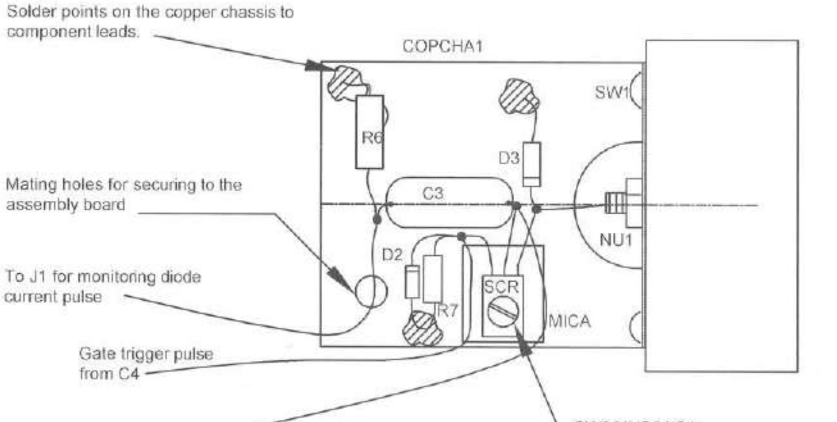

Solder points on the copper chassis to component leads.

Mating holes for securing to the assembly board

To J1 for monitoring diode current pulse

Gate trigger pulse from C4

170 volts from R4

SW2/NUS/MICA
Note that the SCR is insulated from the chassis by a mica or plastic washer and secured using a nylon screw.

Figure 1-6 *Assembly of laser chassis*

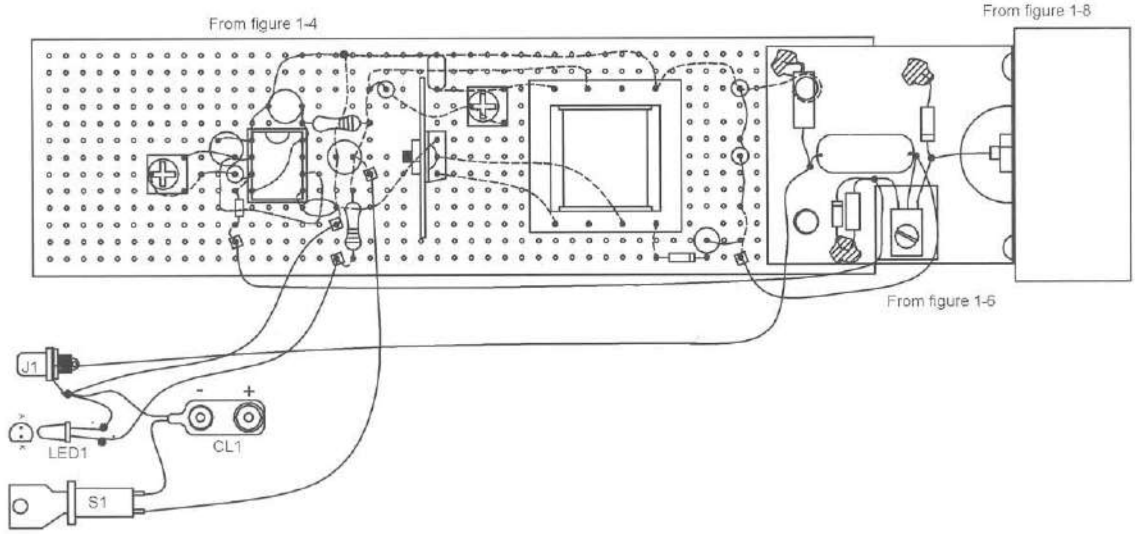

Figure 1-7 *Interconnecting the assembly board to the laser chassis*

7. Fabricate the LAHOLD1, as shown in Figure 1-8. Mount laser diode LD1 into the center hole and secure via a nut (NU1). Secure this assembly to the bent-down flange of the copper chassis. Use two 1/4-inch screws as shown.
8. A collimator adapter is shown and is optional.

Assembly Testing

The following steps will require basic testing equipment.

1. It is assumed at this point that the assembly is complete, as shown in Figure 1-7. Verify the circuit and preset the two trimpots R3 and R9 to midrange.
2. Obtain a 12-volt bench power supply and connect the 12 volts to the battery clip CL1. Turn on the key switch and note the LED coming on. A current meter on the bench supply should read approximately 300 milliamps.
3. Connect a scope to TP1, as shown in Figure 1-2, and adjust R3 for the waveshape as shown.
4. Connect the scope to TP2 and note a spiky waveform that verifies the trigger pulse to the SCR.
5. Connect the scope to TP3 and note the waveform, which is the voltage across the SCR. Note the flat spot indicating that capacitor C3 has reached full charge before the SCR is triggered. The top of this waveshape is affected by the pulse repetition rate and should support 500 repetitions per second when set by R9.
6. Connect the scope to TP4 and set it for a negative-going fast pulse, indicating the current waveform through the LD1 laser diode. If you have followed the instructions and your assembly of the copper chassis is secure with short leads, you should see a picture-perfect waveform as shown.

   Note that the current pulse is 40 amps and corresponds to 1 amp for every volt read at TP4. If you find the pulse is less, you can increase the voltage ratings on one of the zener diodes. The laser diode specified requires 40 amps for full output. Any value over 40 amps will greatly shorten the useful life of this part.
7. Verify the laser output by using a night vision scope or a video camera. You should see an indication of the output. Screw in LEN1 so it is flush with the holder and adjust for the smallest footprint at over 50 feet. Our lab model adjusted to about two turns below being flush with the holder for the best results. The output should be a sharp, narrow bar shape with the diode specified.

Material can be preferably aluminum or fabricated from a PVC rod.

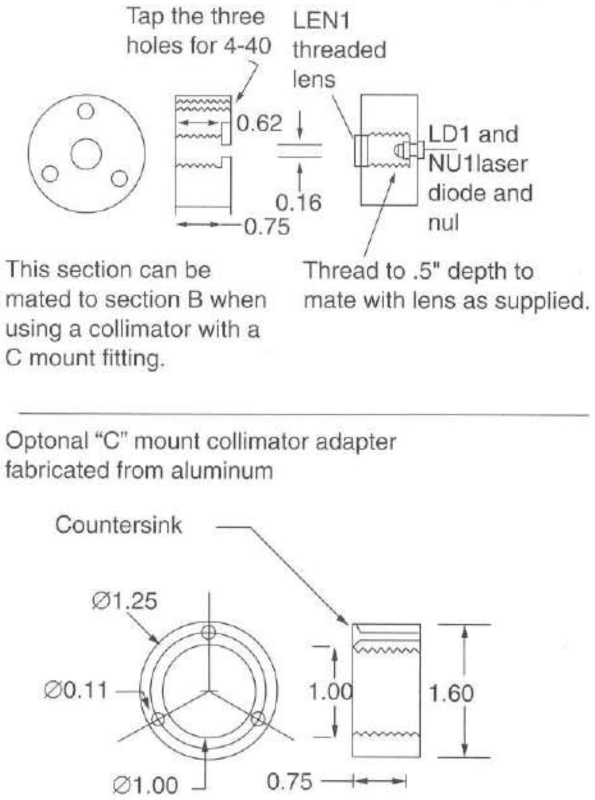

This section can be mated to section B when using a collimator with a C mount fitting.

Thread to .5" depth to mate with lens as supplied.

Optonal "C" mount collimator adapter fabricated from aluminum

Section B adapter for fitting to the lens holder when using the optional collimator

Three mounting holes of .11 on a 1.25" diameter reference 120 degrees are mated to those in above

Thread for C lens at 1" diameter 32 tpi to a minimum of .375"

Figure 1-8 *Laser diode and lens holder*

Final Assembly

Create the EN1 enclosure from a 12 × 1.5-inch inner diameter plastic or aluminum tube (see Figure 1-9). Enclose the assembly and label it as shown in Figure 1-10. You may want to construct a mounting block or another assembly to mechanically connect the laser to a good, sturdy video tripod for future use.

Special Note

This laser is rated at 10 to 20 watts and is not to be confused with a continuous-output device rated at this power level. The advantage of this system is that the pulses are optically equivalent to 10 to 20 watts, and a suitable detection system will see them at this power level.

The equivalent energy output is related to the repetition rate times the peak power times the pulse duration and is approximately equal to joules. A joule is a watt second and is energy. If that energy is released in 1 microsecond as a pulse, the power of that pulse is now 1 megawatt.

A very suitable detector for this pulsed laser is our laser light detector described in Chapter 2, and a useful application for it as a laser property security fence is described in Chapter 5. Note that a laser pointer is 5 milliwatts, meaning the pulsed laser is 5,000 times more detectable.

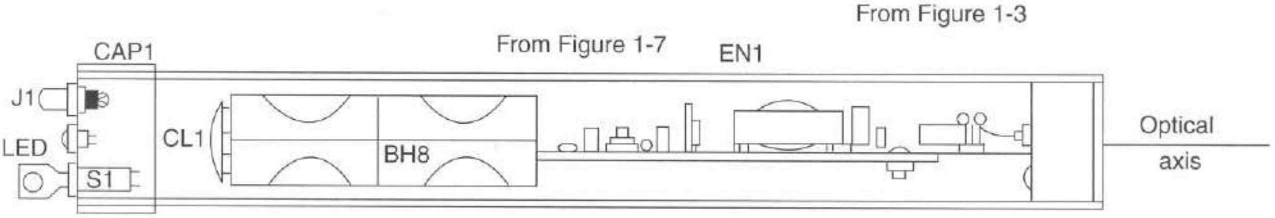

Figure 1-9 *X-ray view of enclosure*

CAUTION: Viewing of laser beam of direct reflections require the use of protective eye wear available through Information Unlimited at www.amazing1.com

Instruction

1. Turn key switch counterclockwise to verify unit is off. Note key is easily removed in the "off" position
2. Remove CAP1 and insert 8 AA batteries into holder.
3. Turn key clockwise and note LED lighting. If not, check the battery holder. Note key cannot be removed in the on position. Laser energy is infrared and can not be seen with the eye. You must use infrared detection devices such as our laser pulse detector or suitable night vision devices. Video cameras will often detect the laser.
4. Adjust collimator for desired effect if you are using.

Note that collimator will expand beam at near field but greatly reduce it at far field.

Compliance tests
1. Verify correct labels as shown
2. Key switch — nonremoval in off position
3. Beam indicator LED indicates beam emission

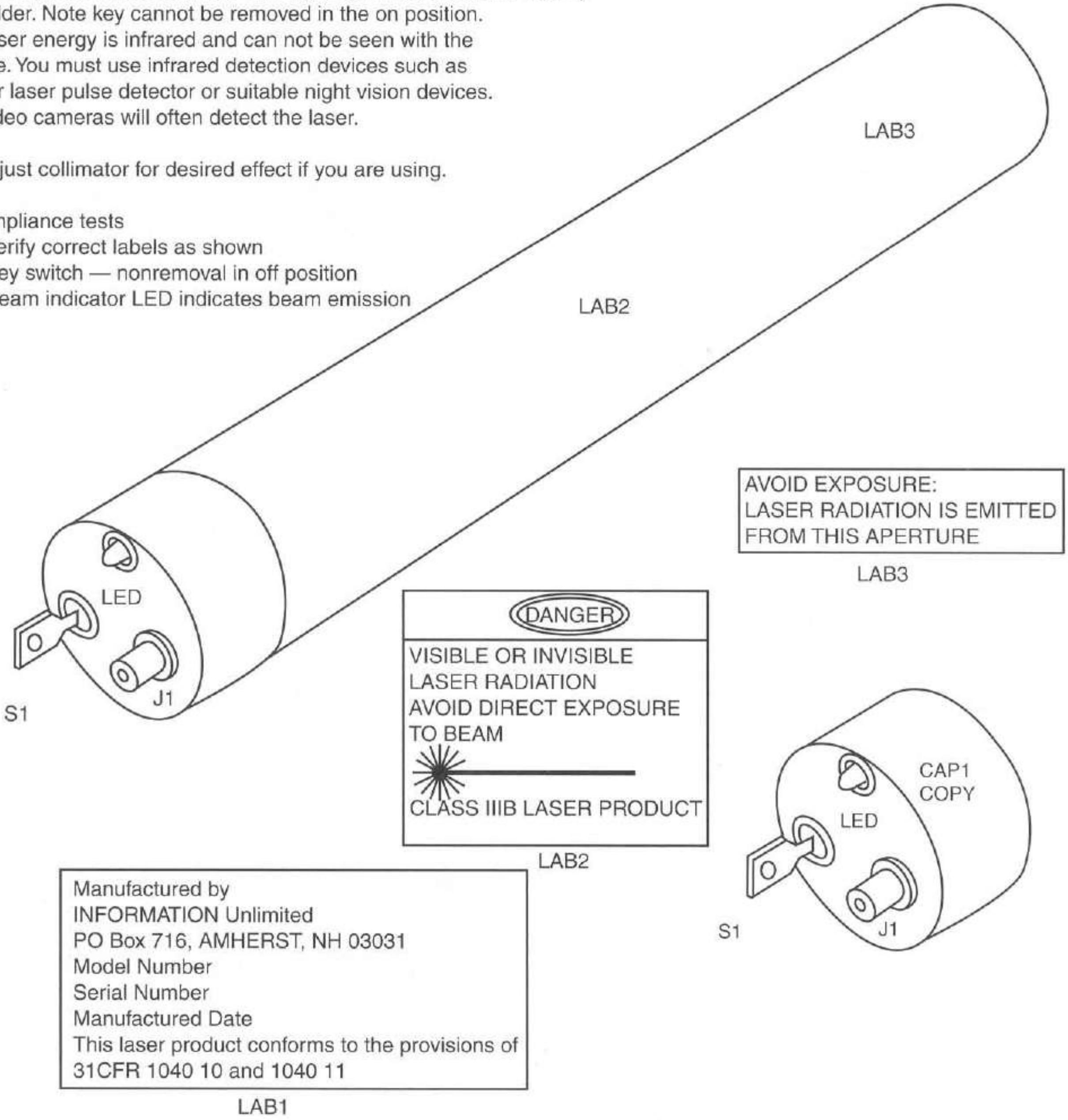

AVOID EXPOSURE:
LASER RADIATION IS EMITTED
FROM THIS APERTURE

DANGER
VISIBLE OR INVISIBLE
LASER RADIATION
AVOID DIRECT EXPOSURE
TO BEAM
CLASS IIIB LASER PRODUCT

Manufactured by
INFORMATION Unlimited
PO Box 716, AMHERST, NH 03031
Model Number
Serial Number
Manufactured Date
This laser product conforms to the provisions of
31CFR 1040 10 and 1040 11

Figure 1-10 *Final assembly view and labels*

Table 1-1 Battery-Operated Infrared Laser Parts List

Ref. #	*Description*	*DB Part #*
R1	470 ¼-watt resistor (yel-pur-br)	
R2	220 ¼-watt resistor (red-red-br)	
R3	2K horizontal trimmer resistor	
R4	18K, 3-watt metal oxide resistor	
R5	10-ohm, ¾-watt resistor (br-blk-blk)	
R6	1-ohm, ½-watt carbon resistor (br-blk-gold)	
R7	39 ¼-watt resistors (or-wh-blk)	
R8	1K ¼-watt resistor (br-blk-red)	
R9	10K horizontal trimmer resistor	
C1,5	Two 10-microfarad, 50-volt vertical electrolytic capacitors	
C2	.047-microfarad, 50-volt plastic capacitor 473	
C3	.033-microfarad, 250-volt polypropylene capacitor	
C4	.1-microfarad, 50-volt plastic capacitor	
C6	1-microfarad, 50-volt vertical electrolytic capacitor	
C7	.01-microfarad/50-volt plastic capacitor (103)	
D1,3	Two 1N4937 1-kilovolt, 1-amp diode	
D2	IN914 silicon diode	
LED1	Bright green *light-emitting diode* (LED)	
LD1	40-amp, 10- to 20-watt, 904-nanometer TO18 pack laser diode	DB# LD650
SCR1	C106M 600-volt, 4-amp *silicon-controlled rectifier* (SCR)	DB# C106M
Z1	100/76-volt, 5-watt zener diode 1N5378	
Z2	150/108-volt, 5-watt zener diode 1N5383	
Q1	MJE3055T NPN power transistor in TO220 case	
I1	555 *dual inline package* (DIP) timer	
T1	400-volt square-wave-switching transformer	DB# TYPE1PC
PB1	1.4 × 5-inch, .1 × .1 grid perforated board	
S1	Key switch with non-removable key in on position	DB# KEYSWSM
J1	RCA phono jack	
COPCHAS1	Copper chassis, created in Figure 1-4	DB# COPCHAS1
LAHOLD	Laser and lens holder, created in Figure 1-8	DB# LAHOLD1
LENS13	Threaded lens	DB# LENS13
EN1	1½-inch inner diameter × 1/32-inch wall × 12-inch plastic or aluminum housing tube	
CAP1,2	1⅝-inch plastic caps, reworked as shown in Figure 1-9	
BH8	Eight AA cell holder	
SW1	Five 6-32 × ¼-inch screws	
NU1	Four 6-32 hex nuts	
SW2	4-40 × ⅜-inch nylon screw	
MICA	Mica insulating washer	
NU2	4-40 hex nut	
HSINK	Heatsink tab for Q1	

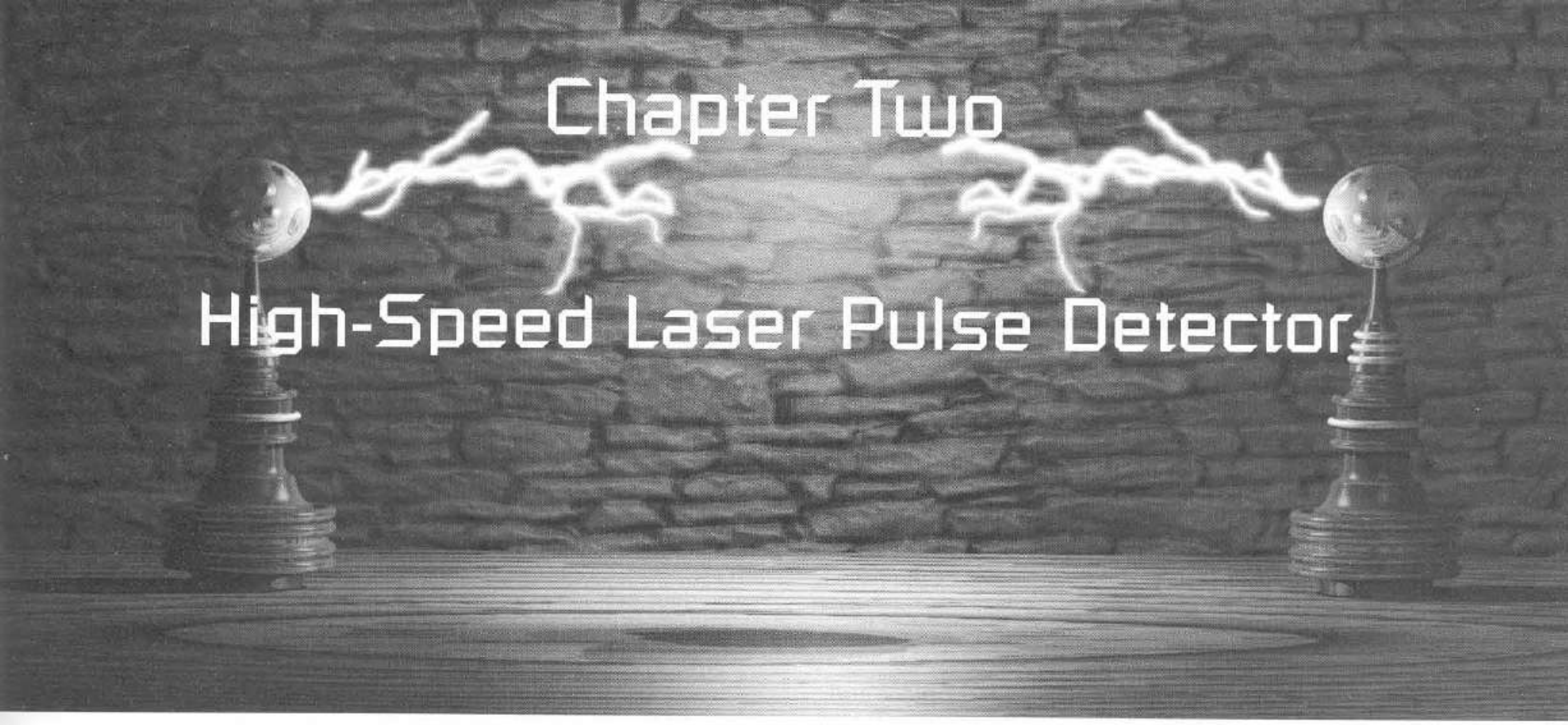

Chapter Two

High-Speed Laser Pulse Detector

This project shows how to build a specialized device capable of detecting fast, high-speed optical pulses. A suitable laser generating these pulses is described in Chapters 1 and 4 and can have a range, limited only by the curvature of the earth. Its uniqueness is its ability to resolve these pulses with rise times in the nanosecond range. This capability makes it possible to transmit large amounts of information within a small period of time.

The heart of this device is the actual front-end detector, in this case a PIN diode device. PIN comes from *positive negative* (PN) junction with intrinsic layer (I). The characteristics of this detector are low capacitance, high resistance, and low leakage when operating in the reverse bias mode. It may also be used as a photo cell when operated in the forward mode.

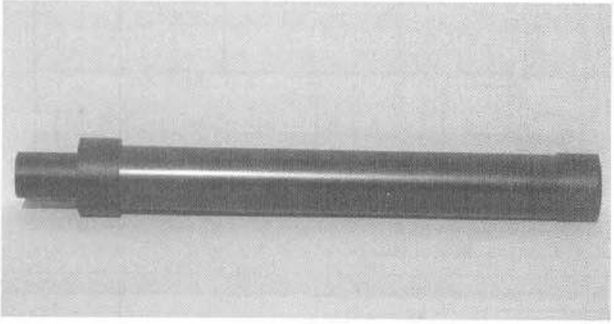

Figure 2-1 *The high-speed laser pulse detector*

The first part of the circuit is built on a piece of thin copper and uses discrete components. The builder may choose to build on a fully copper-clad circuit board if available, and a facsimile output can be set up via a phono jack or a BNC connector. A signal-conditioning circuit is shown built on a normal perforated circuit board and is designed to actuate a relay in the absence of the laser pulses. This mode of operation allows its use as a long-range laser perimeter or invisible fence intrusion detector, as described in Chapter 5.

Optics are shown for those wanting to operate at long ranges. Machined fittings are shown to mate with the 1.5-inch inner-diameter plastic enclosure used for the housing.

This is an intermediate- to advanced-level project requiring certain electronic skills, and one should expect to spend $35 to $70. All parts are readily available, with specialized parts obtainable through Information Unlimited (www.amazing1.com), and they are listed in Table 2-1.

Circuit Description

The circuit in Figure 2-2 utilizes a pin photodiode (LPIN) reverse biased by resistor R3. The reverse-biased photodiode produces a voltage across bias resistor R2. The response of the PIN photo detector is determined by the resistance of R2 and is relatively low due to the nanosecond rise time of the laser pulses detected. Capacitor C3 couples the positive-going detected laser pulses to the gate of a preamplifying *field effect transistor* (FET) Q1 where the negative-going pulse is amplified and fed into a second FET (Q2) for further amplification. NPN transistor Q3 again amplifies the positive-going pulse for Q2 to a negative pulse. This amplified signal is a reasonably good facsimile of the actual laser pulse and can be monitored or further processed via phono jack J1.

The negative-going pulses are also fed to a Schmidt discriminator that is in an untriggered state, providing that transistor Q4 continues to see these pulses. When the pulses are interrupted by an object in the laser path, Q4 turns on due to a positive bias set by trimpot resistor R15 and turns transistor Q5 off, regenerating the trigger state of the Schmidt discriminator. R15 allows the adjustment of a trigger threshold if required. The triggering time of the circuit also helps prevent false triggering such as flying birds, windblown debris, and insects.

The output of the Schmidt is DC coupled into the base of the emitter-follower transistor Q6 where the signal is inverted by transistor Q7. The negative-going level at the collector of Q7 now triggers timer I1 on for a preset time that controls an alarm or another function resulting from a qualified laser beam interruption. A relay (RE1) is shown in the circuit and can be used as normal open or normal closed contacts for control of external circuit functions. This is shown in the phasor pain field generator in Chapter 35.

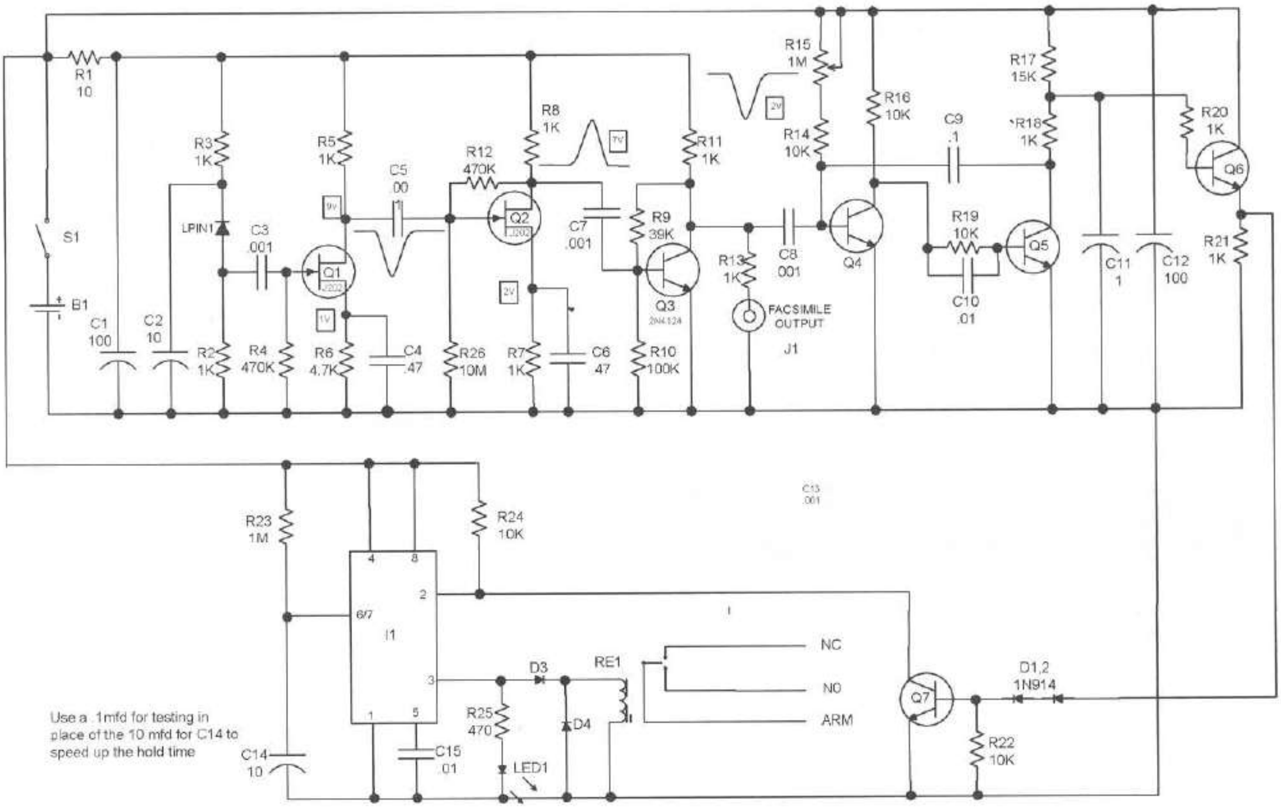
R1
10
S1
B1
C1
100
C2
10
R3
1K
LPIN1
R2
1K
C3
.001
R4
470K
R5
1K
9v
Q1
J202
1V
R6
4.7K
C4
.47
C5
.00
1
R12
470K
R26
10M
R8
1K
7V
Q2
J202
2V
R7
1K
C7
.001
C6
.47
R9
39K
Q3
2N4124
R10
100K
R11
1K
2V
R15
1M
R14
10K
R13
1K
C8
.001
FACSIMILE
OUTPUT
J1
Q4
R16
10K
C9
.1
R19
10K
C10
.01
Q5
R17
15K
R18
1K
C11
1
C12
100
R20
1K
Q6
R21
1K
C13
.001
R23
1M
R24
10K
4
8
2
6/7
I1
3
1
5
C14
10
C15
.01
R25
470
D3
LED1
D4
RE1
NC
NO
ARM
Q7
D1,2
1N914
R22
10K

Use a .1mfd for testing in place of the 10 mfd for C14 to speed up the hold time

Figure 2-2 *Circuit schematic*

System Assembly

1. Assemble the board, as shown in Figure 2-3. Lay out and identify all the parts and pieces. Separate the components that go to the assembly board and create the PB1 perfboard from a $1^1/_2 \times 5^1/_2$-inch piece of .1 × .1 grid. Note the two holes for attachment to the copper chassis section.
2. Assemble the components as shown and observe polarity of the capacitors and semiconductors.
3. Connect the components using their actual leads wherever possible. Follow the layout and avoid bare-wire bridges. Note that Figure 2-3 shows pieces of #20 bus lead used as VC+ and COMMON rails. This approach is easier for wiring as it gives a direct route for the component leads.
4. Create the copper chassis as shown in Figure 2-4. This part is listed as a fabricated piece in Table 2-1.
5. Figure 2-5 shows the wiring of the COPCHA1 copper chassis. This type of wiring is preferred for high-frequency and sensitive, low-noise circuitry. It is used in this circuit due to the fast-rise-time, low-level-voltage pulses.

   Start by forming the leads and soldering to the copper plate. You will need a good, hot soldering iron to do this correctly. The components themselves serve as anchor points and will be self-supporting when soldered to one another.
6. Figure 2-6 shows the connection of the copper chassis and the perforated assembly board via the two screws and nuts. Attach and connect the external leads and components as shown.

Use .031copper sheet

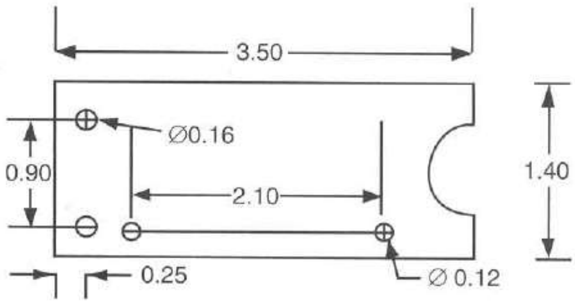

Figure 2-4 *Fabrication of the laser chassis*

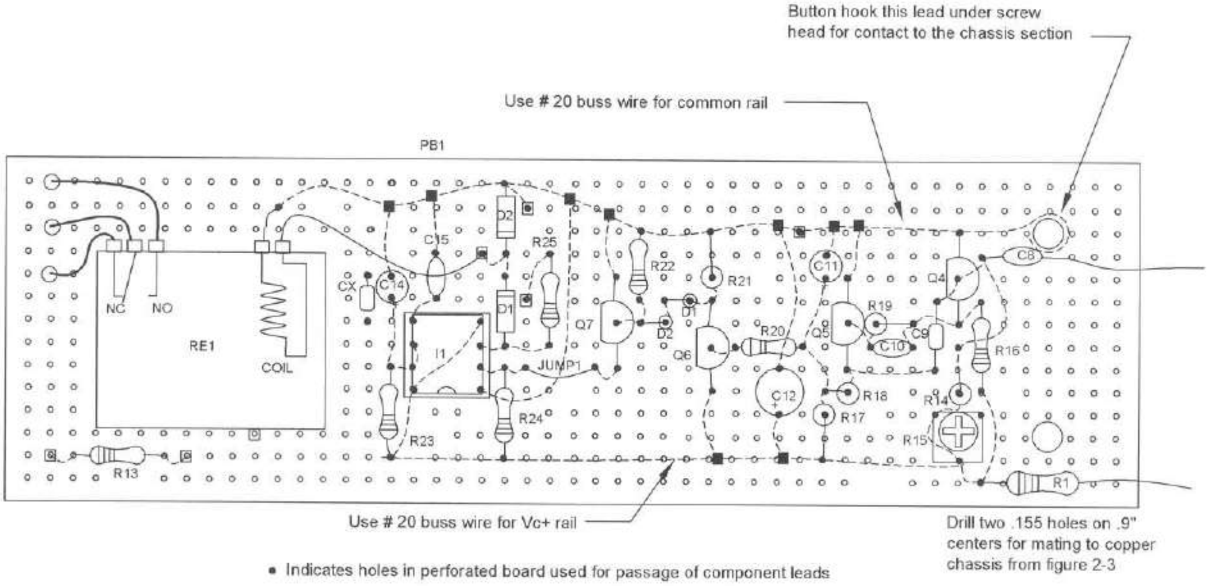

Button hook this lead under screw head for contact to the chassis section

Use # 20 buss wire for common rail

Use # 20 buss wire for Vc+ rail

Drill two .155 holes on .9" centers for mating to copper chassis from figure 2-3

- Indicates holes in perforated board used for passage of component leads

▣ Indicates holes in perforated board used for external leads connection points

■ Indicates component soldering points on (+) plus and (-) negative rails

Use a piece of #20 buss wire for rail leads indicated with heavy dashed lines

Figure 2-3 *Assembly board parts identification*

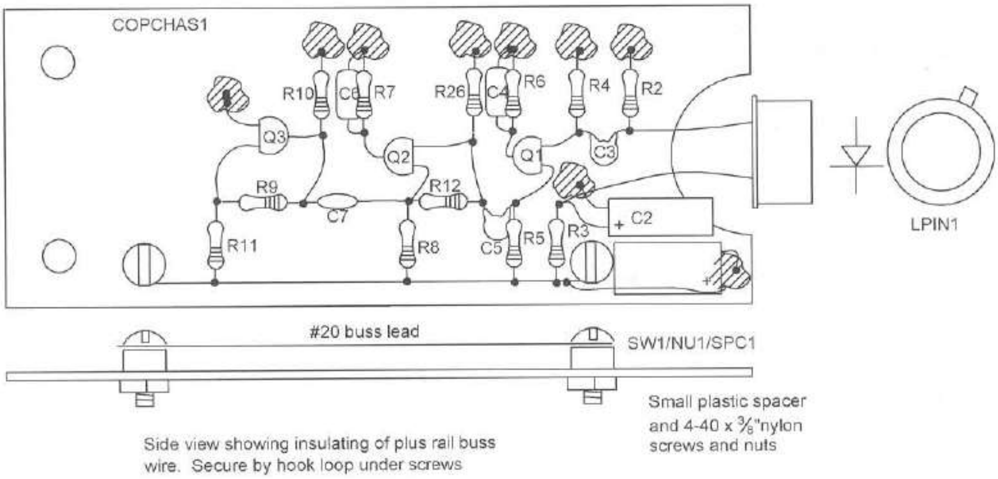
COPCHAS1
R10
C6
R7
R26
C4
R6
R4
R2
Q3
Q2
Q1
C3
R9
C7
R12
C2
+
R11
R8
C5
R5
R3
+
LPIN1
#20 buss lead
SW1/NU1/SPC1

Small plastic spacer and 4-40 x ⅜"nylon screws and nuts

Side view showing insulating of plus rail buss wire. Secure by hook loop under screws

Figure 2-5 *Assembly of the laser chassis*

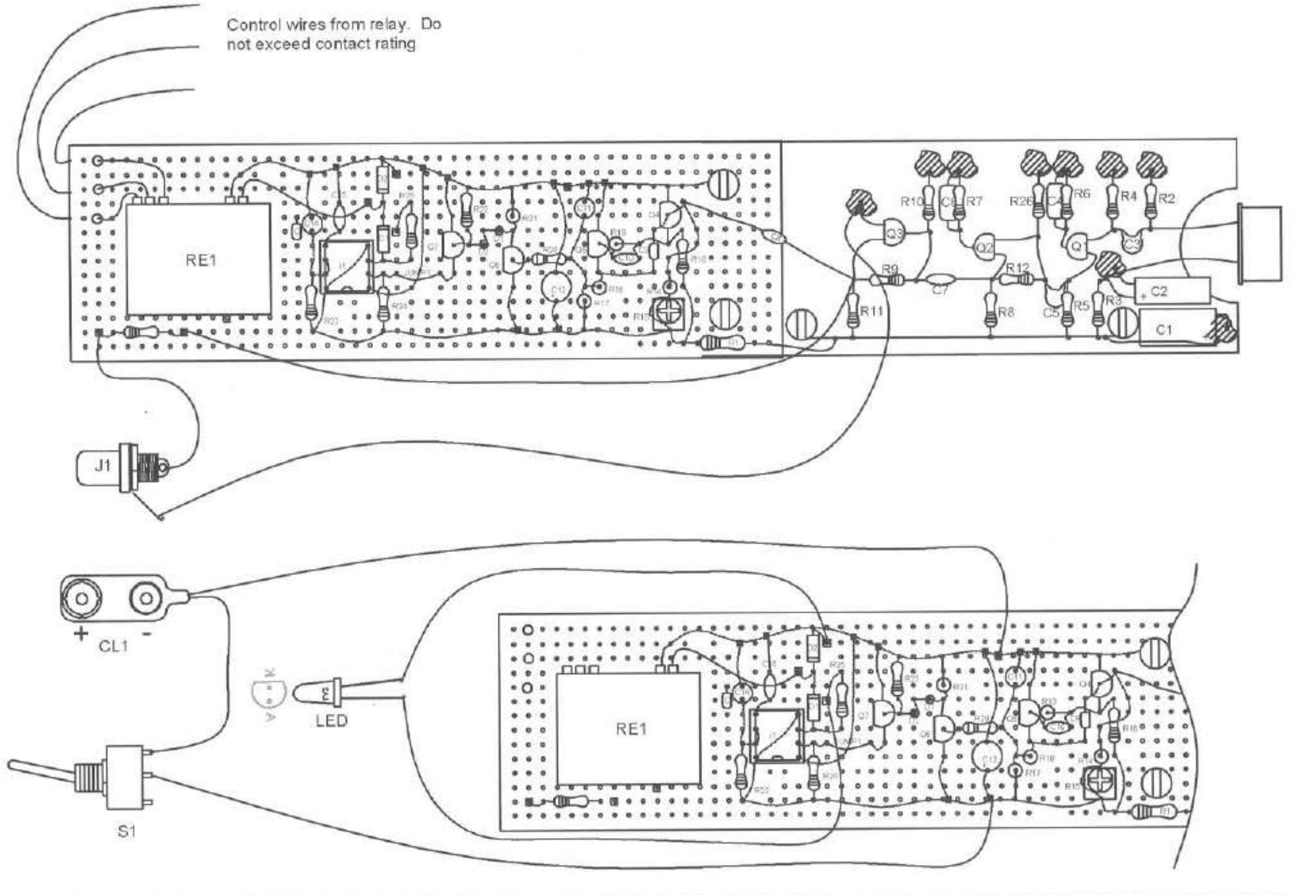
Control wires from relay. Do not exceed contact rating

RE1
R10
C6
R7
R26
C4
R6
R4
R2
Q3
Q2
Q1
R9
C7
R12
C2
R11
R8
C5
R5
R3
C1
J1
+
CL1
-
LED
RE1
S1

Figure 2-6 *Connection of board to chassis*

Assembly Testing

Testing of the laser receiver involves the use of a compatible laser transmitter to provide the necessary high-speed pulses. The laser described in Chapter 1 is recommended.

1. It is assumed at this point that the assembly is complete, as shown in Figure 2-6. Verify the circuit and preset the R15 trimpot to midrange. This test is performed *without* the circuitry installed and without a lens. You should temporarily replace C14 with a .1-microfarad value.
2. Obtain a 12-volt bench power supply and connect the 12 volts to the battery clip CL1. Turn on the S1 switch and note the LED coming on and flashing erratically. A current meter on the bench supply should read approximately 10 milliamps.
3. Measure the DC voltage levels, as indicated in Figure 2-2, noted in the squares. These values should be within 10 to 15 percent of those indicated. At this point, you will need an optical laser pulse transmitter, such as that described in Chapter 1. Any suitable laser or optical transmitter will work with a fast pulse rise of 1 usec or so. The detector, as shown, uses a silicon PIN diode with a spectral sensitivity peak at 960 nanometers. This is very close to the spectral output of the lasers described in Chapters 1 and 4.
4. Position the laser pulse generator source so that it is pointing in the direction of the laser receiver being tested. You will need to connect a scope at the drain of Q1 and then attempt to position the receiver so it is receiving the pulses and is secure in this position. This may be tricky and require patience. When the laser is perfectly aligned with the receiver, the signal will block the Q1 and produce an overloaded waveshape. Carefully adjust the receiver off axis until the scope indicates the negative-going facsimile pulse as shown. Continue through the circuit and observe the waveshapes as shown, again adjusting the receiver off axis to obtain the required waveshape.

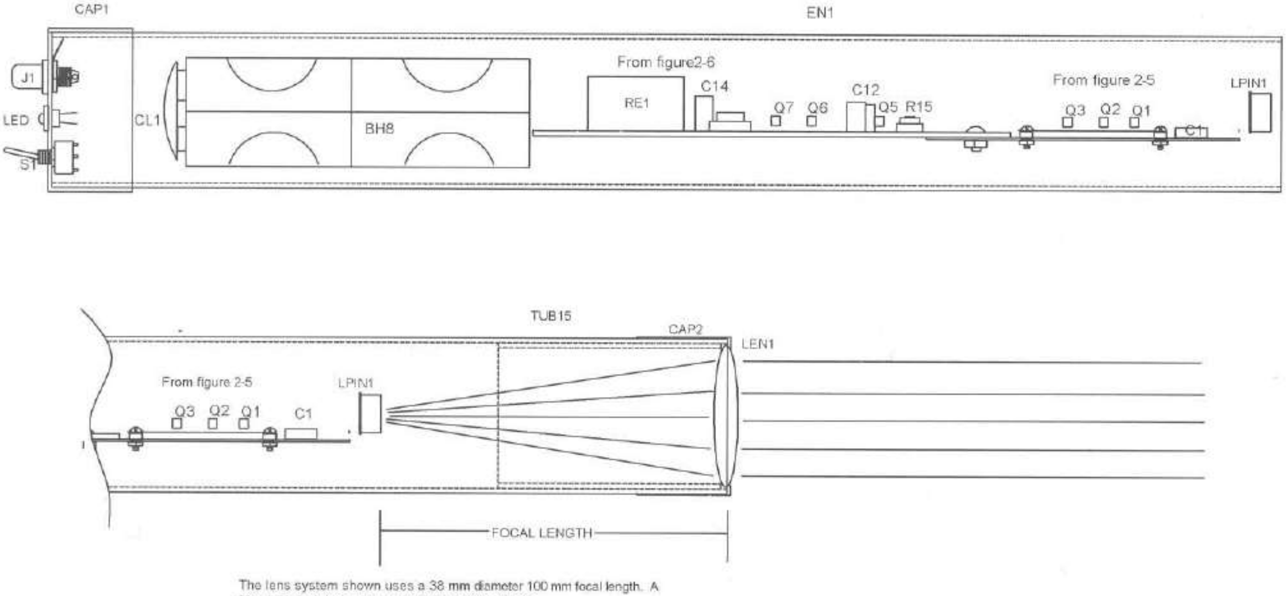

The lens system shown uses a 38 mm diameter 100 mm focal length. A 2" x 1.5" diameter tube (TUB15) is sleeved into the main enclosure EN1 and is positioned to abut to the lens on it's shoulders. A plastic cap (CAP2) retains the lens against TUB15. Acenter hole of over 1" is removed for the light aperture.

Figure 2-7 *X-ray view of enclosure*

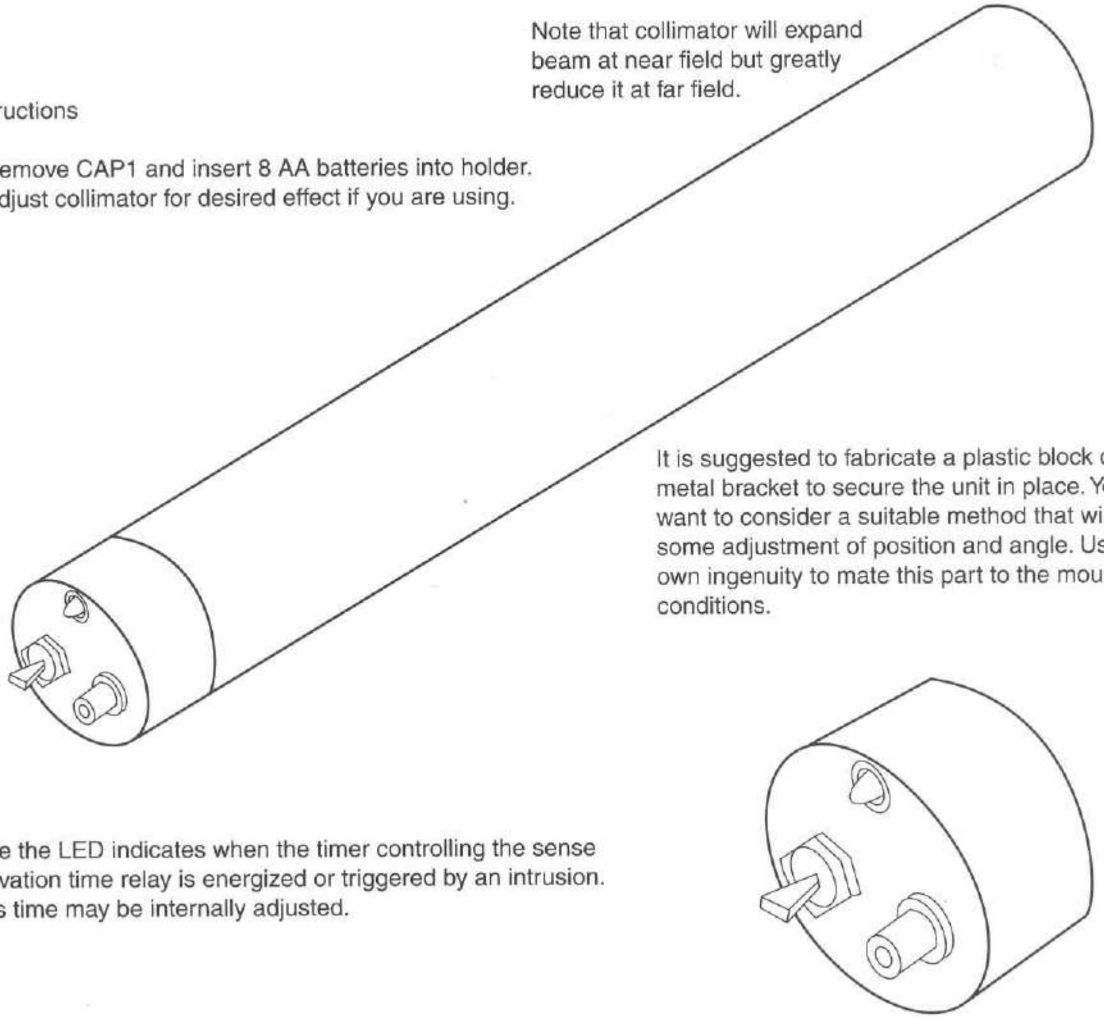

Instructions

1. Remove CAP1 and insert 8 AA batteries into holder.
2. Adjust collimator for desired effect if you are using.

Note that collimator will expand beam at near field but greatly reduce it at far field.

It is suggested to fabricate a plastic block or sturdy metal bracket to secure the unit in place. You may want to consider a suitable method that will allow some adjustment of position and angle. Use your own ingenuity to mate this part to the mounting conditions.

Note the LED indicates when the timer controlling the sense activation time relay is energized or triggered by an intrusion. This time may be internally adjusted.

Figure 2-8 *Final assembly view*

Note that the LED will be off if the circuit is working and will come on if you block the light with a solid object.

5. Verify that the discriminator works and you can set the sensitivity trip level via R15. Verify the timer trips when the light beam is broken, activating the relay for a preset time as determined by the value of C14. This feature now allows external circuitry to be activated when the laser beam is broken, powering alarms or deterrents. The relay as shown provides a normally closed and open circuit that can be selected by the user. Do not switch loads that exceed the relay contact rating.

Final Assembly

Fabricate the EN1 enclosure from a 15 × 1.5-inch ID plastic or aluminum tube. Enclose the assembly as shown in Figure 2-7. You may want to create a mounting block or another assembly to mechanically connect the laser to a good, sturdy video tripod for future use

The system as shown is enclosed in a long tube that includes the batteries. You may separate the batteries and relay control circuits from the actual detection section consisting of the copper chassis and optics. This approach will be a two-part system that is connected via a cable or cord.

Special Note

This laser pulse detector receiver, when used with a pulsed laser like that described in Chapter 1, can have a control range exceeding that allowed by the curvature of the earth when used with suitable optics. Alignment is about as critical as sighting a high-powered rifle and requires extreme mechanical stability once set. The alignment of a long-range system requires patience, perseverance, and a little black magic.

Chapter 5 describes how to use both the laser pulser from Chapter 1 and the above assembly in a multiple-reflection property protection device with an ultrasonic shock deterrent.

Special Note on Photo Detectors

Two photo detectors are referenced in Table 2-1. LPIN is the low-cost piece that will usually provide the necessary operating parameters for most hobbyist laser projects. LPINX is more expensive and has a much lower dark current rating. It is used for lower noise and more sensitive circuitry.

Table 2-1 Laser Pulse Detector Parts List

Ref. #	*Description*	*DB Part #*
R1	10 ohms, 1/4 watt (br-blk-blk)	
R2, 3, 5, 7, 8, 11, 13, 18, 20, 21	Ten 1K, 1/4-watt resistors (br-blk-red)	
R4, 12	Two 470K, 1/4-watt resistors (yel-pur-yel)	
R6	4.7K, 1/4-watt resistor (yel-blk-red)	
R9	39K, 1/4-watt resistor (or-wh-or)	
R10	100K, 1/4-watt resistor (br-blk-yel)	
R14, 16, 19, 22, 24	Five 10K, 1/4-watt resistors (br-blk-or)	
R15	1M horizontal trimmer resistor	
R17	15K, 1/4-watt resistor (br-blk-or)	
R23	1M, 1/4-watt resistor (br-blk-grn)	
R25	470-ohm, 1/4-watt resistor (yel-pur-br)	
R25	10M, 1/4-watt resistor (br-blk-blue)	
C1, 12	Two 100-microfarad, 50-volt vertical electrolytic capacitors	
C2, 14	Two 10-microfarad, 50-volt vertical electrolytic capacitors	
C3, 5, 7, 8	Four .001-microfarad, 100-volt disc capacitors	
C4, 6	Two .47-microfarad, 50-volt plastic capacitors	
C9, 13	Two .1-microfarad, 50-volt plastic capacitors	
C10, 15	Two .01-microfarad, 50-volt plastic capacitors (103)	
C11	1-microfarad, 50-volt vertical electrolytic capacitor	
Q1, 2	Two J202 N-channel FET transistors	
Q3	2N4124 *negative positive negative* (NPN) high-frequency transistor	
Q4,5,6,7	Four PN2222 NPN general-purpose transistors	
D1,2	Two IN914 silicon diodes	
D3,4	Two IN4007 1-kilovolt, 1-amp diodes	
I1	555 *dual inline package* (DIP) timer integrated circuit	
LED1	High-brightness *light-emitting diode* (LED)	
*LPIN1	Silicon PIN photodiode dark current less than .10 nano-amp (see text)	DB# PIN
*LPIN1X	Silicon photo diode dark current less than .02 nano-amp (see text)	DB# PINX
RE1	12-volt coil, 400-ohms, 120 VAC 15-amp contacts	DB# RE12115
COPCHAS1	1.4 × 3.5-inch .031 copper plate (see Figure 2-4)	
SW1/NU1	Four 4-40 × 3/8 nylon screws and metal hex nuts	
SPC1	Two 1/4-inch plastic spacers	
PB1	5½″ × 1½″ .1 × .1 grid perforated circuit board	
J1	Chassis mount RCA phono jack	
S1	Small *single pole, single throw* (SPST) toggle switch	
CL1	Battery clip	
BH8	Holder for 8 AA cell battery	
EN1	1 1/2-inch inner diameter × 1/32-inch wall plastic or aluminum tube	
CAP1, 2	Two 1 5/8″ plastic caps reworked as shown in Figure 2-8	
LEN1	38 mm × 100 mm focal length DCX glass lens	DB# 38100
RET1	Reworked 1 1/2-inch plastic cap for retaining lens	
TUB15	2-inch length of 1.5-inch outer diameter plastic or aluminum tube	

Chapter Three

Ultra-Bright Green Laser Project

This ultra-bright green laser project, as shown in Figure 3-1, is intended for the laser experimenter who desires to add a simple optical system that provides far-field focusing, beam collimation, or beam expanding.

Construction parts such as tubing, handles, brackets, and hardware are all readily available from most local hardware stores (see Table 3-1). Our suggested lab approach is to locate some telescopic tubing of metal or plastic, or a combination of both, and fabricate the laser and lens holders from a suitable sized plastic rod or equivalent. The center holes of the holders must be as true as possible or the optical system's integrity will be greatly degraded.

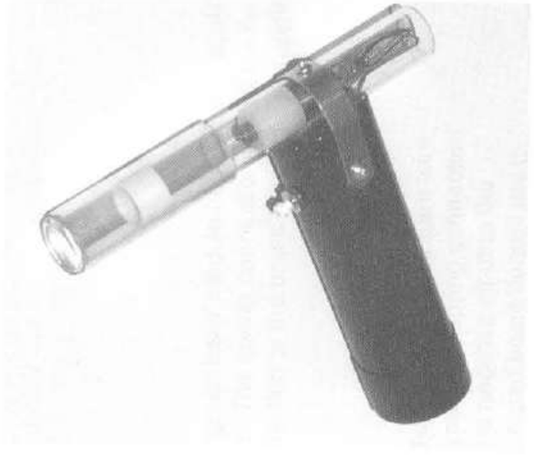

Figure 3-1 *Photo of green laser project*

The grip or handle is created from a 1-inch-thick piece of finished pine that has a section carefully removed for the battery holder and push-button switch. Passage holes for the power leads are drilled as shown. You may also use a suitable sized plastic box. A clear piece of Lexan (polycarbonate) sheet is formed with a flange at a 90-degree bend and is the surface for mounting the switch. This cover is screwed to the wooden handle and makes an attractive design if done with care and precision. The handle can be constructed with a contoured shape to match the curvature of the tube, and you may also add rings, fins, and other décor to make a space-age optical ray gun.

The operation of the laser allows usage without any optics, providing similar performance to a high-quality laser pointer. The addition of the lens system allows far-field focusing, close range expansion, and far-field collimation where the beam spot impact will be reduced by the collimating power of the system. As an example, the bare foot beam diameter (or operation without external optics) at 300 meters will be approximately a 50-centimeter diameter spot. With an ×10 collimator, the diameter can be reduced to 5 centimeters.

This is a class iiia beginner's laser project that uses a working laser module. Class iiia is a classification for lasers under 5 milliwatts of output power

Never point it at vehicles, aircraft, or directly at people. Expect to spend $40 to $50 for this very rewarding laser project.

Assembly Steps

1. Rework the laser module (LM1) as shown in Figure 3-2.
2. Wire the laser module with pushbutton switch (SW1) and battery clip (CL1) as shown in Figure 3-3.
3. Make final assembly as shown in Figure 3-4.
4. Connect batteries and activate SW1. Note a green impact point. Adjust lenses as shown for desired beam profile.

The green laser module has a built in push button switch that may be bypassed by adding a connection lead as shown. This added lead is the negative input. The spring connection is also the negative input but now requires activation of the built in switch. The positive input can be a mechanical connection to the brass housing by wrapping a bare lead around the slot. You may also carefully solder a lead as shown for this positive input point.

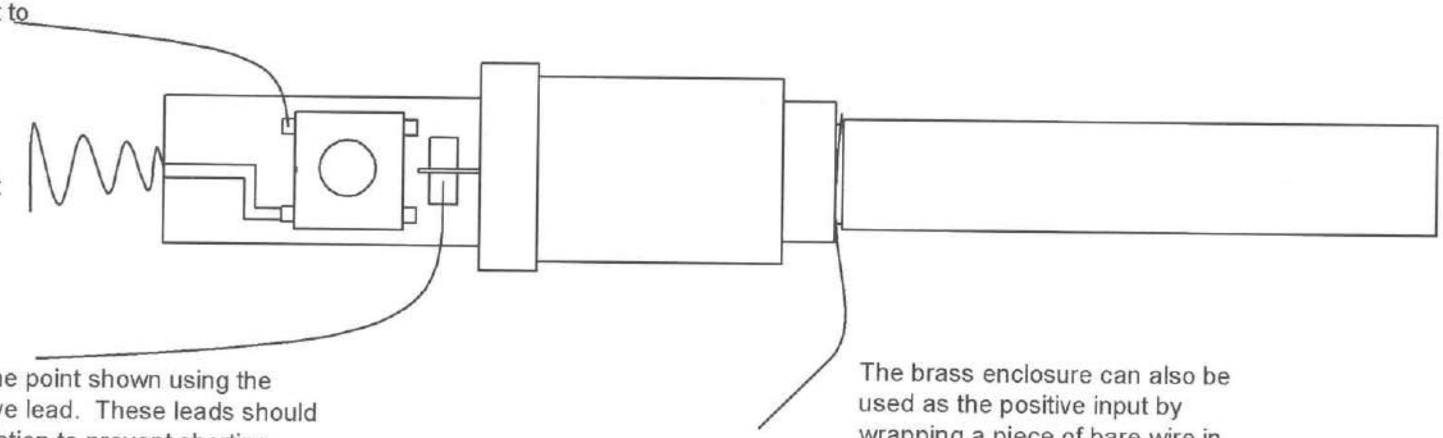

Solder a piece of #24 bare wire to bypass the spring connection for the negative input to the switch pad being careful not to damage the circuit board

The spring is the normal negative input point but will require activation of the built in push button switch

You may solder a lead to the point shown using the same method for the positive lead. These leads should be sleeved with some insulation to prevent shorting. You may also want to strain relieve these leads using silicon rubber etc.

The brass enclosure can also be used as the positive input by wrapping a piece of bare wire in the narrow slit as shown.

Figure 3-2 *Laser module rework*

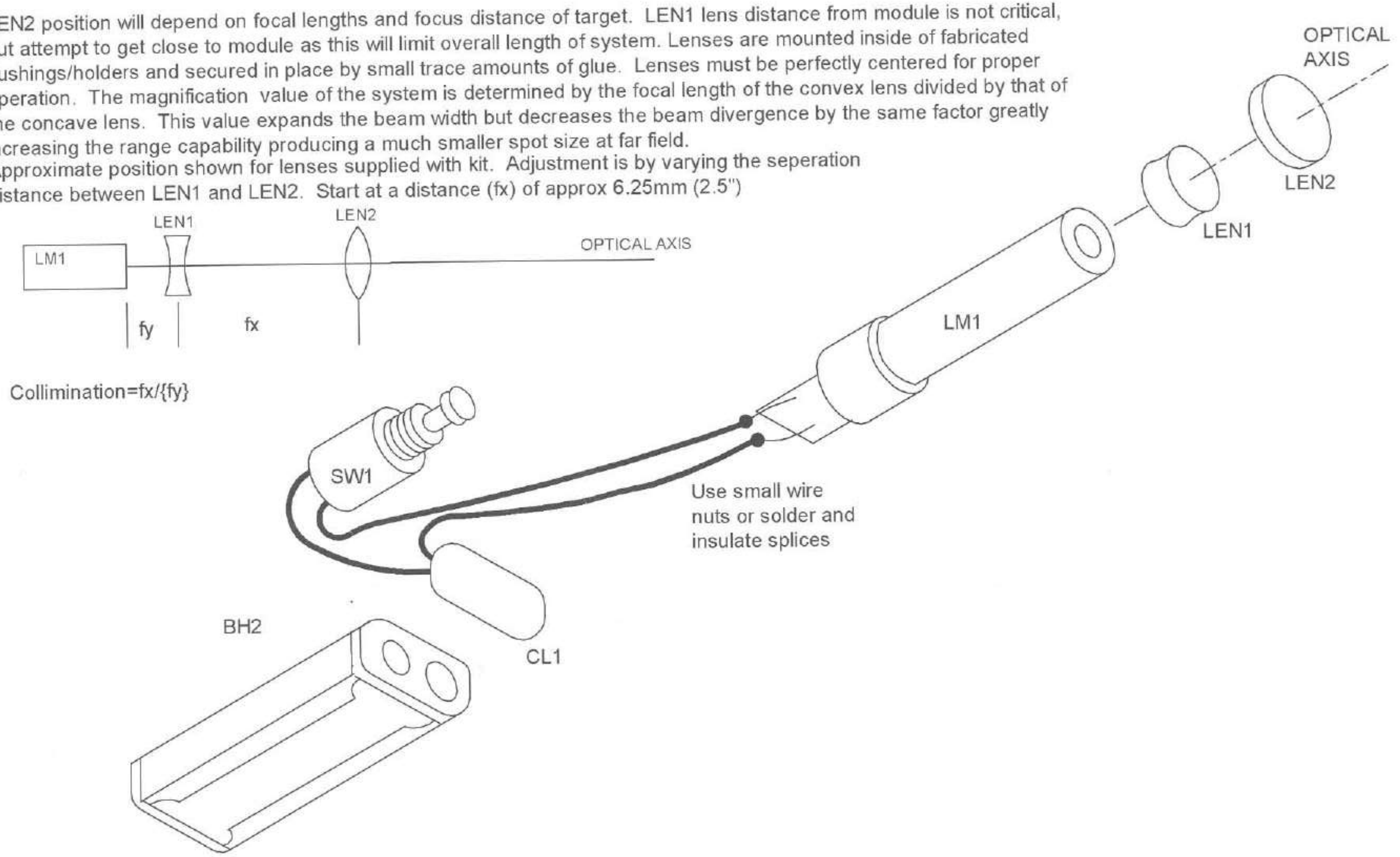

LEN2 position will depend on focal lengths and focus distance of target. LEN1 lens distance from module is not critical, but attempt to get close to module as this will limit overall length of system. Lenses are mounted inside of fabricated bushings/holders and secured in place by small trace amounts of glue. Lenses must be perfectly centered for proper operation. The magnification value of the system is determined by the focal length of the convex lens divided by that of the concave lens. This value expands the beam width but decreases the beam divergence by the same factor greatly increasing the range capability producing a much smaller spot size at far field.

Approximate position shown for lenses supplied with kit. Adjustment is by varying the seperation distance between LEN1 and LEN2. Start at a distance (fx) of approx 6.25mm (2.5")

Use small wire nuts or solder and insulate splices

The green laser module has a built in push button switch that may be bypassed by adding a connection lead as shown. This added lead is the negative input. The spring connection is also the negative input but now requires activation of the built in switch. The positive input can be a mechanical connection to the brass housing by wrapping a bare lead around the slot. You may also carefully solder a lead as shown for this positive input point.

Figure 3-3 *Ultra bright green isometric assembly*

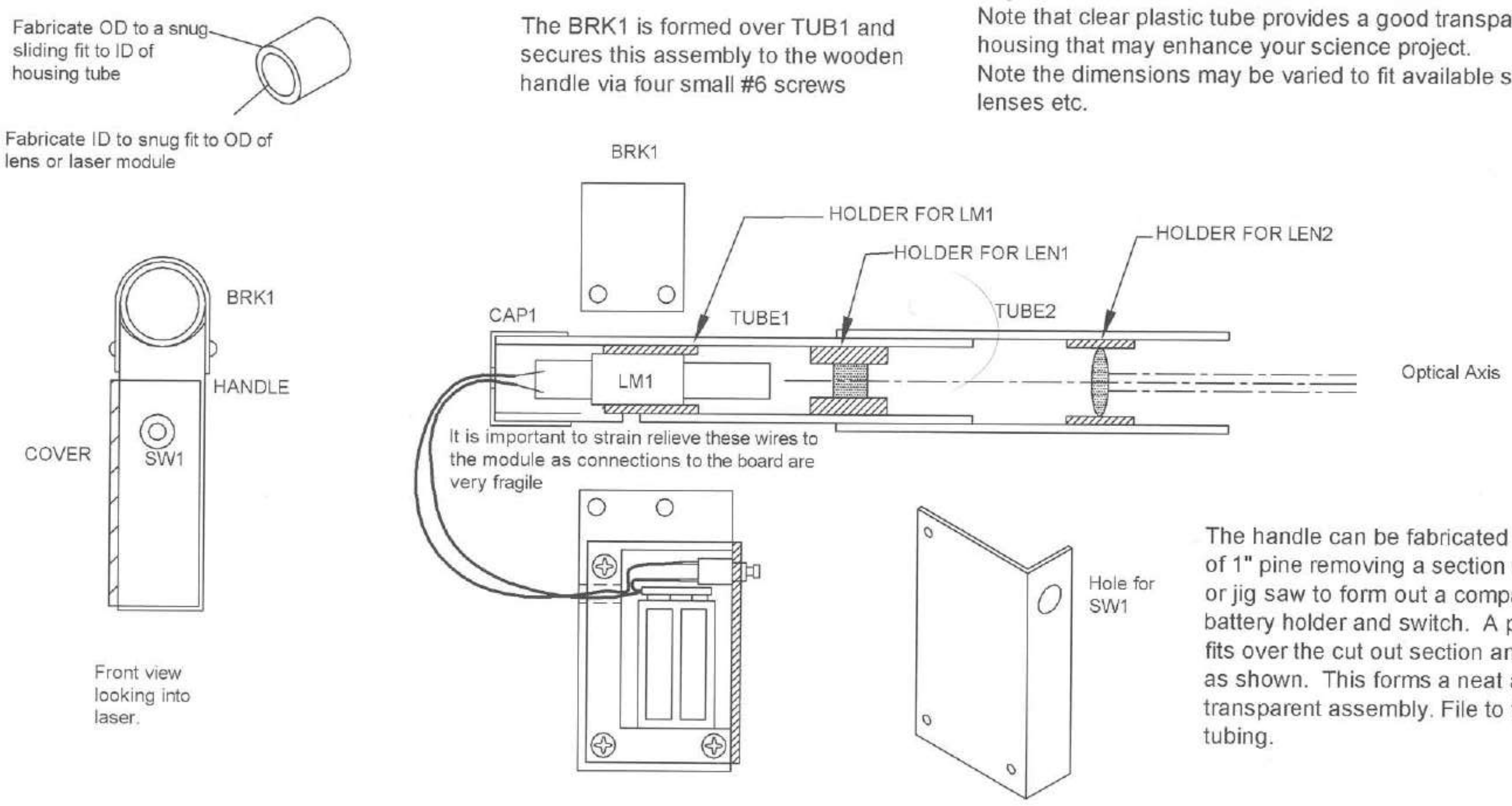

Holders for laser module and lenses may be made from fabricating wooden or plastic dowels to the required dimensions.

Fabricate OD to a snug sliding fit to ID of housing tube

Fabricate ID to snug fit to OD of lens or laser module

The BRK1 is formed over TUB1 and secures this assembly to the wooden handle via four small #6 screws

Tubing pieces should telescopically fit for easy adjustment of collimator. Start with around a 1" diameter section for TUBE1 and obtain a suitable size for TUBE2. You may ream out the larger tube or turn down the smaller for a suitable sliding fit. Note that clear plastic tube provides a good transparent housing that may enhance your science project. Note the dimensions may be varied to fit available sizes of lenses etc.

It is important to strain relieve these wires to the module as connections to the board are very fragile

The handle can be fabricated from a piece of 1" pine removing a section with a chisel or jig saw to form out a compartment for the battery holder and switch. A plastic cover fits over the cut out section and is fabricated as shown. This forms a neat appearing and transparent assembly. File to fit contour of tubing.

Clear piece of formable plastic such as lexan or equivalent. Fab as shown

Front view looking into laser.

Figure 3-4 *Ultra bright green optical laser gun suggested final design*

Table 3-1 Ultra-Bright Green Laser Parts List

Ref. #	*Description*	*DB Part #*
SW1	Pushbutton NO switch	
BH2AA	Two-cell AA battery holder	
CL1	Battery snap clip	
LM1	5-milliwatt green laser module	DB# LM-532-P5
LEN1	Concave lens .6 × -.75-inch focal length	DB# LE15
LENS	Convex lens .9 × 3.5-inch focal length	DB# LE2475
LAB1	Class 3a danger laser label	
LAB2	Certification label	
LAB3	Aperture label	
The mechanical parts you get yourself; they are not critical and those shown represent our finalized lab approach.		
TUBE1	1-inch OD thin-walled plastic tubing	
TUBE2	Tube for telescope over TUBE1	
HANDLE	2 × 4 × 1-inch piece of soft pine	
COVER	Piece of 1/16-inch clear plastic sheet	
BRK1	1/32 × 1/2-inch aluminum bracket	
LENS HOLDERS	Wood or plastic dowels	
SCREWS	Small wood	

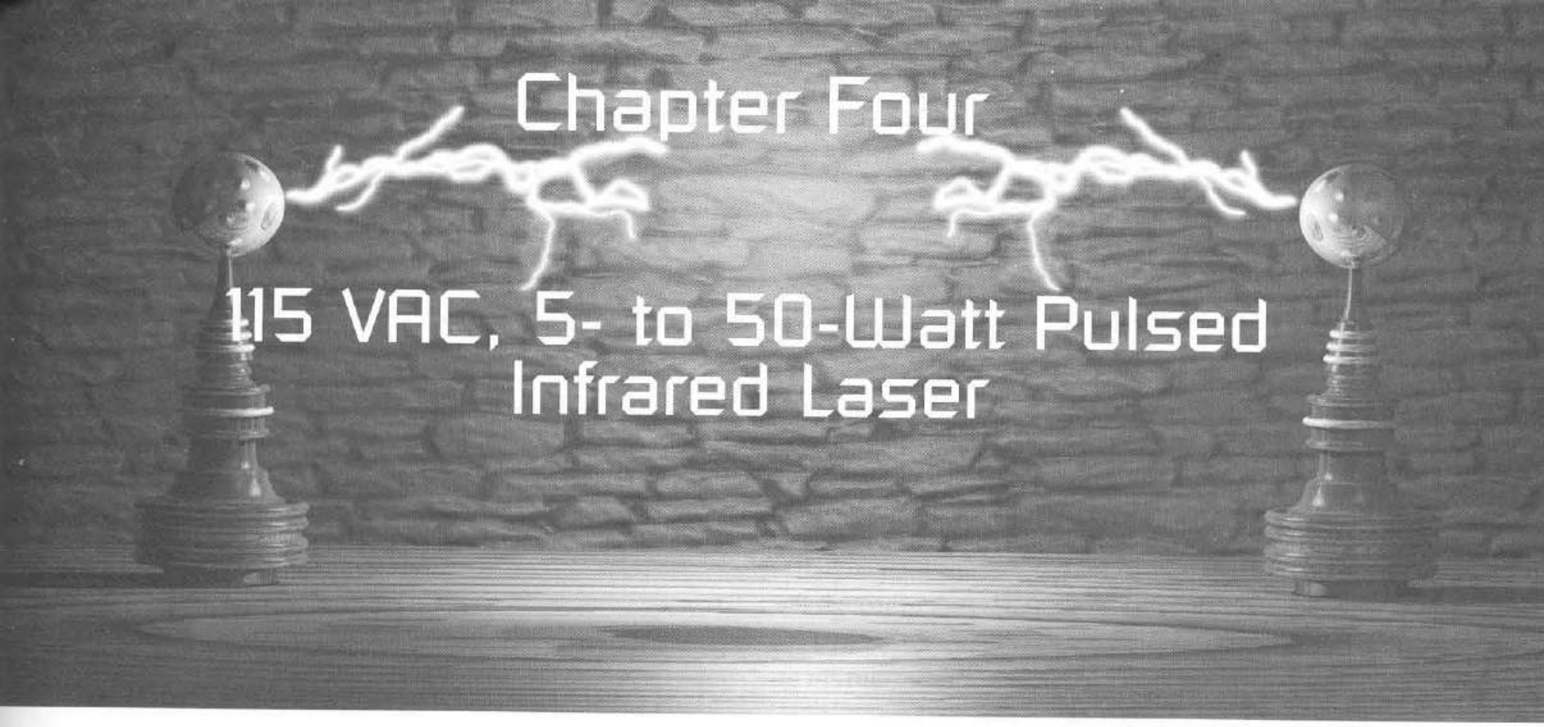

Chapter Four

115 VAC, 5- to 50-Watt Pulsed Infrared Laser

This project shows how to construct a laboratory-use, solid-state, pulsed infrared laser (see Figure 4-1). The system utilizes a gallium arsenide laser diode providing pulse powers of 10 to 100 watts depending on the diode used. The device operates from the 115 VAC line via an isolation transformer and is constructed in two parts. The laser is an excellent source of variable-frequency pulses adjustable from 10 to 2,500 repetitions per second of infrared energy.

The laser, when assembled as shown, is a class iiib device and requires the appropriate labeling and several included safety functions, as described in the assembly instructions. At no time should it be pointed at anyone without protective eyewear or at anything that could reflect these pulses. Also, never look into the unit when the power is on. It is intended to be used for ranging, simulated weapons practice, intrusion detection, communications and signaling, and a variety of related scientific, optical experiments and uses.

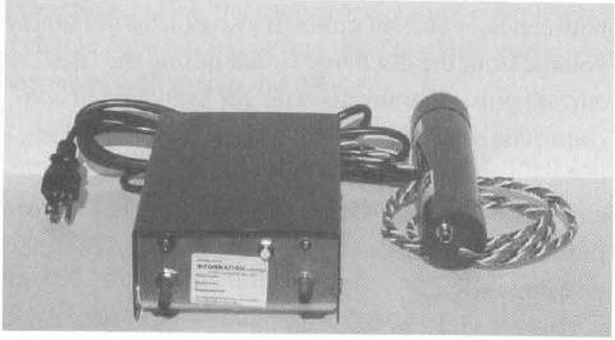

Figure 4-1 *IR laser*

This is an intermediate- to advanced-level project requiring electronic skills and basic electronic shop equipment. Expect to spend $100 to $150. All parts are readily available, with specialized parts obtainable through Information Unlimited (www.amazing1.com), and they are listed in Table 4-1.

Theory of Operation

A laser diode is nothing more than a three-layer device consisting of a *pn* junction of *n*-type silicon, a *p*-type of gallium arsenide, and a third *p* layer of doped gallium arsenide with aluminum. The *n*-type material contains electrons that readily migrate across the *pn* junction and fill the holes of the *p*-type material. Conversely, holes in the *p*-type migrate to the *n*-type and join with electrons. This migration causes a potential hill or barrier consisting of negative charges in the *p*-type material and positive charges in the *n*-type material that eventually cease growing when a charge equilibrium exists. In order for current to flow in this device, it must be supplied at a voltage to overcome this potential barrier. This is the forward voltage

Table 4-1

INFO #	*Pulse width*	*Package*	*Diodes*	*Peak current*	*Emitting area*	*Peak power*	*Duty factor*	*Beam symmetry*	*Spectral width*
LD660	200 ns	TO18	1	40 amps	9 × 1	12 watts	.1%	15 × 20	3.5 nm
LD780	200 ns	TO5	2	40 amps	21 × 1	20 watts	.1%	15 × 20	3.5 nm
LD1630	200 ns	TO5	2	40 amps	21 × 1	30 watts	.1%	15 × 20	3.5 nm

drop across a common diode. If this voltage polarity is reversed, the potential barrier is simply increased, assuring no current flow. This is the reversed bias condition of a common diode.

A diode without an external voltage applied to it contains electrons that move and wander through the lattice structure at a low, lazy average velocity as a function of temperature. When an external current at a voltage exceeding the barrier potential is applied, these lazy electrons now increase their velocity to where some, by colliding, acquire a discrete amount of energy and become unstable, eventually emitting this acquired energy in the form of a photon after returning to a lower energy state. These photons of energy are random both in time and direction; hence, any radiation produced is incoherent, such as that of a *light-emitting diode* (LED).

The requirements for coherent radiation are that the discrete packets of radiation must be in the form of a lockstep phase and in a definite direction. This demands two essential requirements; first, sufficient electrons at the necessary excited energy levels and, second, an optical resonant cavity capable of trapping these energized electrons for stimulating more and giving them direction. The amount of energized electrons is determined by the forward diode current. A definite threshold condition exists where the device emits laser light rather than incoherent light, such as in an LED. This is why the device must be pulsed with a high current. The radiation from these energized electrons is reflected back and forth between the square-cut edges of the crystal that form the reflecting surfaces due to the index of refraction of the material and air.

The electrons are initially energized in the region of the *pn* junction. When these energized electrons drift into the *p*-type transparent region, they spontaneously liberate other photons that travel back and forth in the optical cavity interacting with other electrons commencing laser action. A portion of the radiation traveling back and forth between the reflecting surfaces of these mirrors escapes and constitutes the output of the device.

Circuit Theory of Operation

AC power, as shown in Figure 4-2, is obtained via polarized plug CO1 through fuse FH1. Proper grounding of the green cord lead is a necessity to prevent an unnecessary shock hazard along with usage of grounded test and measuring equipment. S1 is a key switch type where the key can be removed only in the off position. Transformer T1 provides a one-to-one ratio and isolation from the power line, while step-down transformer T2 provides the low voltages necessary for the control circuits. The power indicator lamp consists of neon lamp NE1 and associated current-limiting resistor R1.

Diodes D1 and D2, along with capacitors C1 and C2, comprise a voltage doubler. The voltage across C1 and C2 is 1.4 × 230 or approximately 340 volts.

A major section of this laser pulser is the charging circuit. This circuit charges the pulse discharge capacitor (C10) to the supply voltage during the time interval between laser current pulses. It also isolates the supply voltage from the discharge circuit during the laser current pulse, thereby allowing the switching *silicon-controlled rectifier* (SCR) to recover to the blocking state. Because the response times of the charging circuit are relatively long, lead lengths are not important, and the circuit can be remotely located from the discharge circuit.

The simplest charging circuit is a resistor-capacitor combination. In this simple case, the resistor must limit the current to a value less than the SCR holding

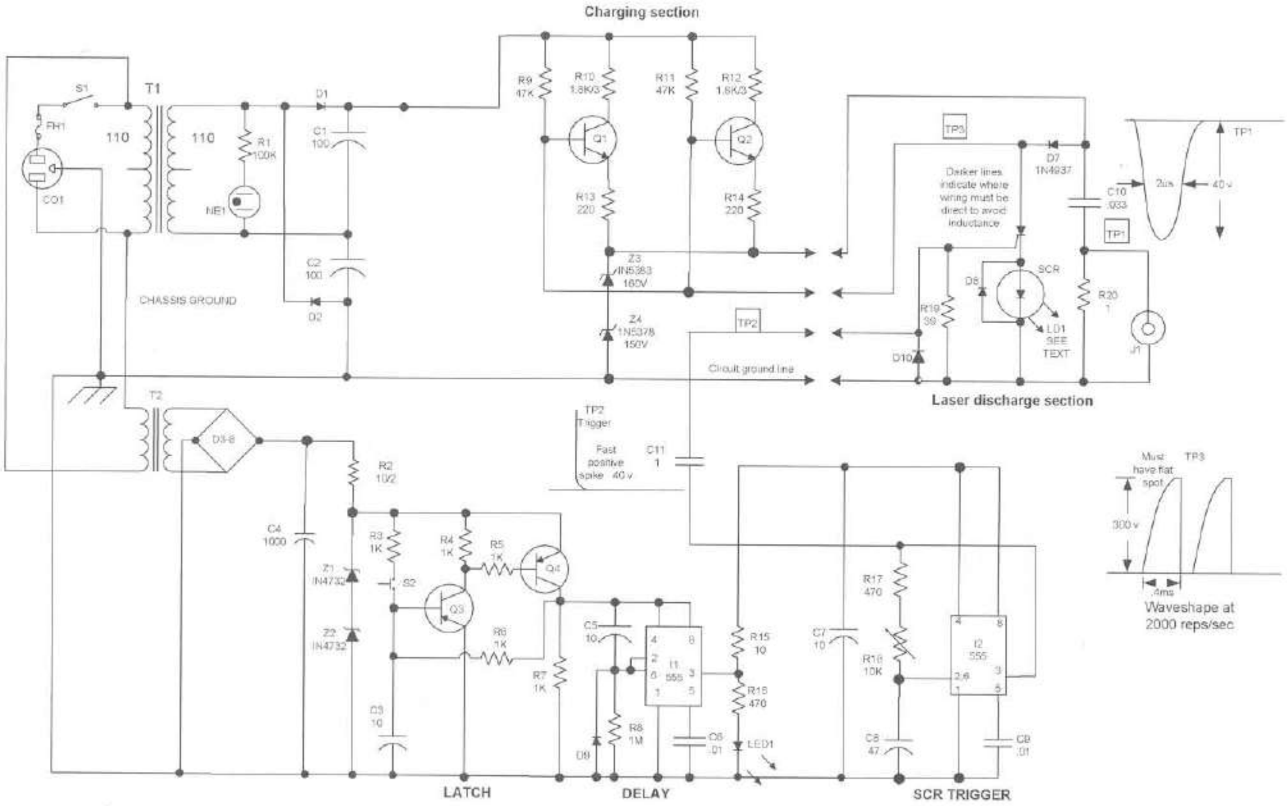
Charging section
S1
T1
FH1
110
110
CO1
R1
100K
NE1
D1
C1
100
C2
100
D2
CHASSIS GROUND
R9
47K
R10
1.8K/3
R11
47K
R12
1.8K/3
Q1
Q2
R13
220
R14
220
Z3
1N5383
160V
Z4
1N5378
150V
TP2
Circuit ground line
TP3
D7
1N4937
Darker lines indicate where wiring must be direct to avoid inductance
C10
.033
TP1
2us
40 v
TP1
SCR
D8
R19
39
R20
1
LD1
SEE
TEXT
J1
D10
Laser discharge section
T2
D3-8
R2
10/2
TP2
Trigger
Fast positive spike 40 v
C11
1
C4
1000
Z1
IN4732
Z2
IN4732
R3
1K
R4
1K
R5
1K
S2
Q3
Q4
R6
1K
R7
1K
C3
10
C5
10
I1
555
R8
1M
D9
C6
.01
R15
10
R16
470
LED1
C7
10
R17
470
R18
10K
C8
47
I2
555
C9
.01
Must have flat spot
TP3
300 v
.4ms
Waveshape at 2000 reps/sec
LATCH
DELAY
SCR TRIGGER

Figure 4-2 *Circuit schematic*

current but should be as low as practical because this resistance also determines the charging time of the capacitor C10, thus determining the available laser pulse repetition rate at full energy. For example, a resistance of 40 kilo-ohms limits the current to 10 milliamps from a 400-volt supply. This current value is just at the holding level of an average SCR. A time of almost 4 milliseconds is required to charge C10, a 0.033-microfarad capacitor, to the supply voltage in three time constants through a 40-kilo-ohm resistor. Therefore, the *pulse repetition rate* (PRR) of the pulsing circuit is limited to less than 250 Hz. If the PRR exceeds this value, the capacitor does not completely recharge between pulses and the peak laser current decreases with increasing PRR.

Varying the supply voltage may control the peak current in the discharge circuit, provided the PRR is low enough to allow the capacitor to fully recharge between current pulses. Both the supply voltage and the PRR determine the peak laser current. A considerable risk exists in increasing the supply voltage to compensate for an insufficient recharge time. If the PRR is decreased while the supply voltage is high, the capacitor again recharges completely, and the laser pulse current increases to a value that may damage or destroy the laser diode.

The needs for a variable voltage supply and the low PRR limit are the major disadvantages of the resistive-type charging network. Therefore, the limitations of the simple resistor drivers are that a resistor large enough to keep the current below the holding current of the SCR also limits the pulse repetition rate.

The frequency capability of the pulse power supply can be greatly improved with the charging circuits, as described. Capacitor C10 is charged to the supply voltage when the SCR is not conducting. Diode D7 is in the off state, because no current flows through it during this period. At the onset of a pulse trigger at the gate, the impedance of the SCR drops rapidly and capacitor C10 discharges toward ground potential. During this period, current is surging through diode D7, causing zero biasing of the current-control transistors (Q1, Q2) by shorting the bias resistors (R9, R11). The SCR now turns off when the current drops below its holding current. The voltage across capacitor

C10 then charges back to the supply voltage. During the capacitor charge cycle, diode D7 passes no current, and Q1 and Q2 are forward biased into the saturation region. Obviously, the charging time of C10 is now shorter due to being charged through the smaller-value resistors (R10, R12). The emitter resistors R13 and R14 balance the current through Q1 and Q2.

Zener diodes (Z3, Z4) are selected to regulate and provide the proper required current pulse for the laser diode in operation. The zener diodes provide a maximum charge voltage of 310 volts that produces 40 amps through laser diode LD1 for up to 2,500 laser pulses per second.

Other laser diodes may require a selection of other voltage-rated zener diodes. The builder may wish to provide a multiposition switch to select these various combinations of zener diodes externally via a front panel control.

The trigger circuit consists of an astable timer (I2) that derives its power through the output of timer I1, which provides a delay time required for FDA laser compliance. The trigger and delay circuitry, as well as the timers, are powered by 12 VDC from transformer T2 and rectifier diodes D3 through D6. Capacitor C4 filters the rectified wave and steps up the voltage to 15 VDC. Resistor R2 drops the 15 volts to the zener voltage of zener diodes Z1 and Z2 of regulated 12 VDC.

Operation of the laser requires activating pushbutton switch S2, which latches transistor Q4 upon supplying power to timer I1. The timer must time out to 10 seconds, as determined by resistor R8 and capacitor C5. The emission indicator LED1 comes on simultaneously with power to the trigger timer I2, which pulses the laser diode LD1. The laser pulse rate is determined by external pot R18.

Notes on SCR

In the conventional operation of an SCR, the anode current, initiated by a gate pulse, rises to its maximum value in about 1 microsecond. During this time, the anode-to-cathode impedance drops from open circuit to a fraction of an ohm. In injection laser pulsers, however, the duration of the anode-cathode pulse is much less than the time required for the SCR to turn on completely. Therefore, the anode-to-cathode impedance is at the level of 1 to 10 ohms throughout most of the conduction period. The major disadvantage of the high SCR impedance is that it causes low circuit efficiency. For example, at a current of 40 amps, the maximum voltage would be across the SCR, while only 9 volts would be across the laser diode, which represents very low circuit efficiency. The efficiency of a laser array is greater due to its circuit impedance being more significant.

Because the SCR is used unconventionally, many of the standard specifications such as peak current reverse voltage, on-state forward voltage, and turn-off time are not applicable. In fact, it is difficult to select an SCR for a pulsing circuit on the basis of normally specified characteristics. The specifications important to laser pulser applications are forward-blocking voltage and current rise time. A usage test is the best and many times the only practical method of determining the suitability of a particular SCR.

Notes on the Storage Capacitor

The voltage rating of the storage capacitor must be at least as high as the supply voltage. A high-quality metallized capacitor must be used for this part as peak currents are in the tens of amps.

Notes on Layout Wiring

Lead lengths and circuit layout are very important to the performance of the discharge circuit. Lead inductance affects the rise time and peak value of the current and can also produce ringing and undershoots in the current waveform that can destroy the laser. A well-built discharge circuit might have a total lead length of only 1 inch and therefore an inductance of approximately 20 nanohenries. If the current rises to 75 amperes in 100 nanoseconds, the inductive voltage drop will be $e = L\, di/dt$. If proper care is not taken in wiring the discharge circuits, high-inductance voltage drops will result.

A 1-ohm resistor (R20) in the discharge circuit will greatly reduce the current undershoot in single-diode lasers. Laser arrays usually have sufficient resistance

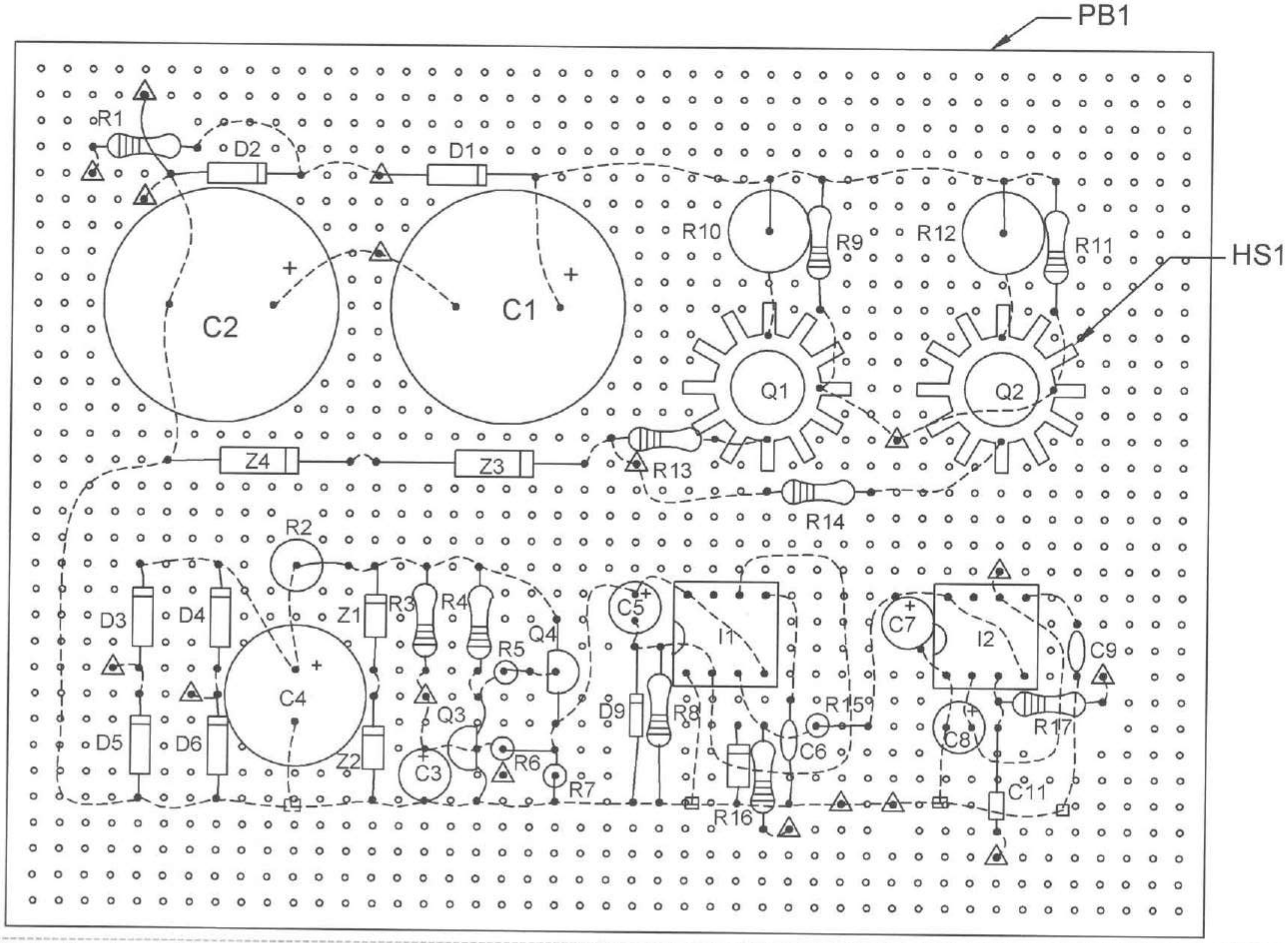
PB1
HS1
R1
D2
D1
C2
C1
R10
R9
R12
R11
Q1
Q2
Z4
Z3
R13
R14
R2
D3
D4
D5
D6
C4
Z1
Z2
R3
R4
R5
Q4
Q3
C3
R6
R7
C5
D9
R8
I1
R15
C6
R16
C7
I2
C9
C8
R17
C11

Figure 4-3 *Board assembly*

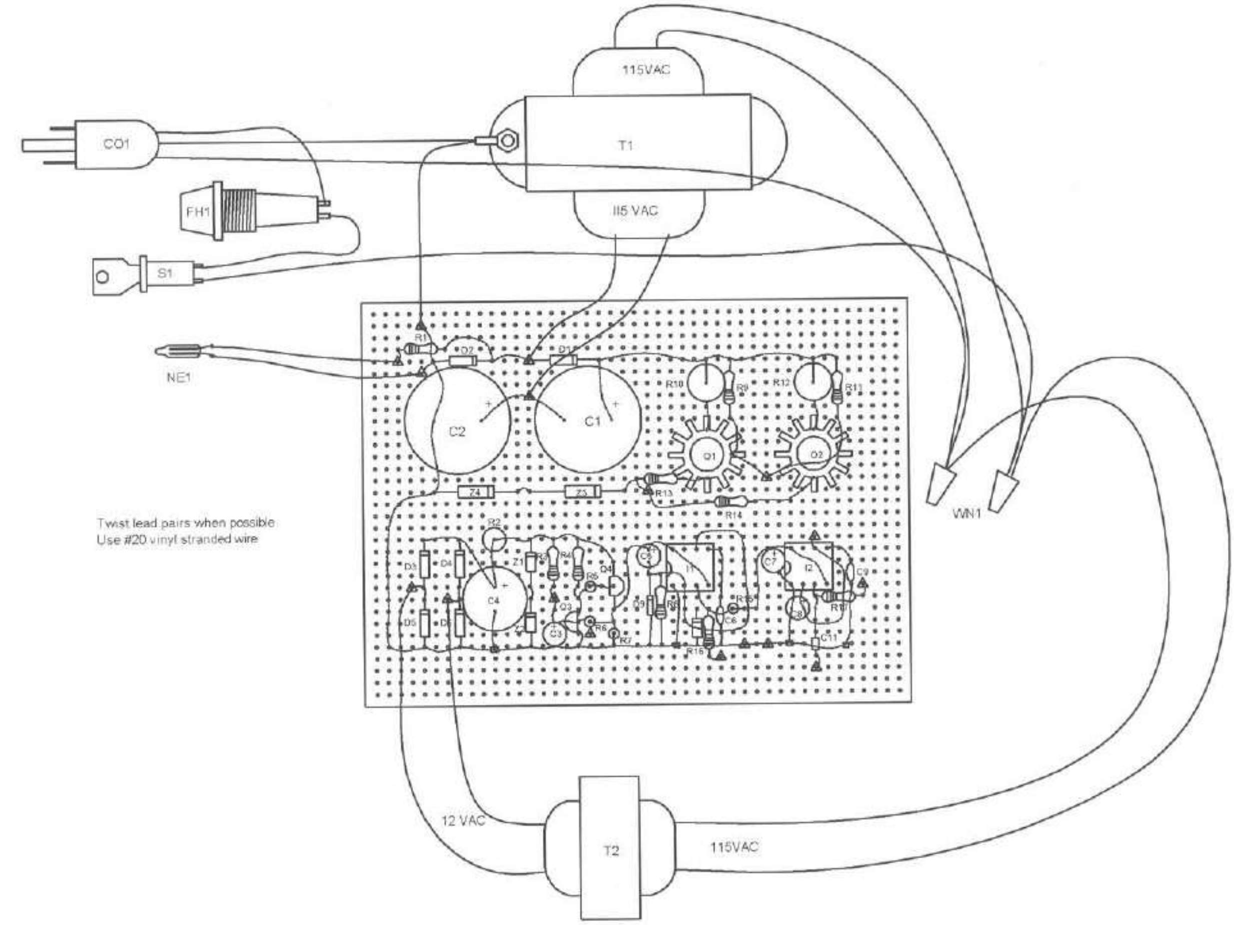
115VAC
T1
II5 VAC
CO1
FH1
S1
NE1
R1
D2
D1
C2
C1
R10
R9
R12
R11
Q1
Q2
Z4
Z3
R13
R14
WN1
Twist lead pairs when possible
Use #20 vinyl stranded wire
R2
D3
D4
D5
Z1
C4
Q3
Q4
R5
R6
R7
C3
I1
I2
C5
C6
C7
C9
C11
R16
12 VAC
T2
115VAC

Figure 4-4 *Board input power wiring.*

to eliminate an undershoot. The small resistance in the discharge circuit is also useful in monitoring the laser current, as described in the following section.

A clamping diode (D8) is added in parallel with the laser to reduce the current undershoot. Its polarity should be opposite that of the laser. Although the clamping diode is operated above its usual maximum current rating, the current undershoot caused by ringing is very short and the operating life of the diode is satisfactory.

Notes on Current Monitoring

The current monitor in the discharge circuit provides a means of observing the laser current waveform with an oscilloscope. A resistive-type monitor (R20) reduces circuit ringing and current undershoot, but the lead inductance of the resistor may cause a higher than actual current reading. A current transformer such as the Tektronix CT-2 can also be used to monitor the current and is not affected by lead inductance. Because the transformer does not respond to low-frequency signals, it should be used with fast, short-pulse-width, fast fall-time waveforms.

Assembly Steps

Assembly of the laser system is straight forward and will require basic tools and test equipment.

1. Lay out and identify all parts and pieces. Verify them with the parts list, and separate the resistors as they have a color code to determine their value. Colors are noted on the parts list.
2. Cut a piece of .1-inch grid perforated board to a size of 5.8 × 3.5 inches, as shown in Figure 4-3.

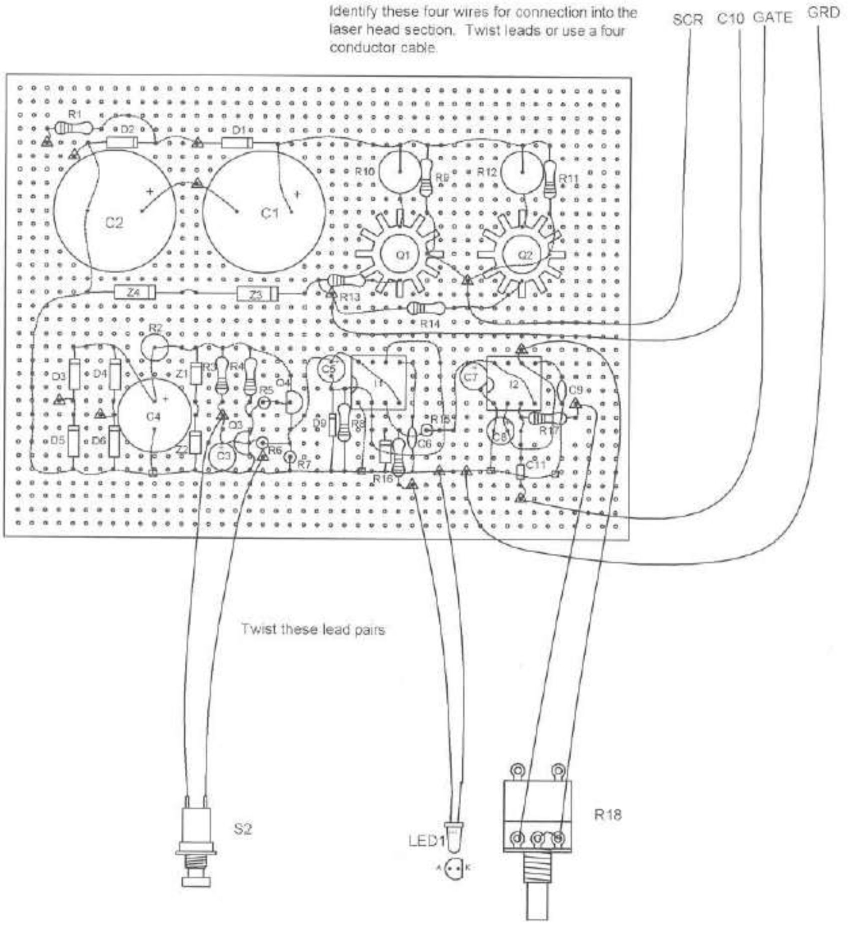
Identify these four wires for connection into the laser head section. Twist leads or use a four conductor cable.
SCR
C10
GATE
GRD
R1
D2
D1
R10
R9
R12
R11
C2
C1
Q1
Q2
Z4
Z3
R13
R14
R2
D3
D4
Z1
R3
R4
Q4
R5
C5
I1
C7
I2
C9
C4
D5
D6
Z2
C3
Q3
R6
R7
R8
C6
R15
C8
R17
C11
R16
Twist these lead pairs
S2
LED1
A
K
R18

Figure 4-5 *Board output and control wiring*

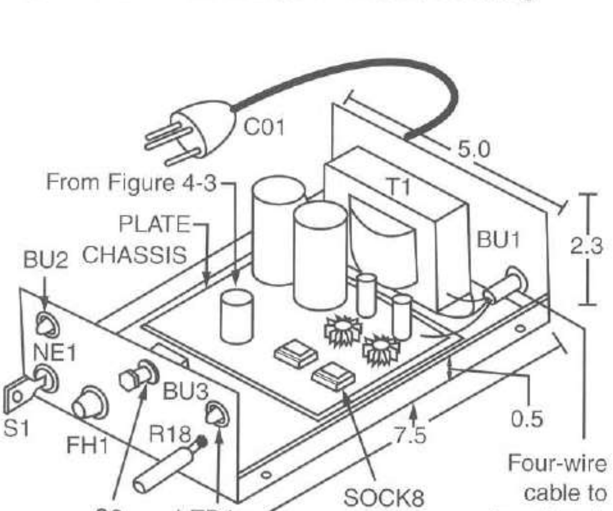
C01
From Figure 4-3
PLATE CHASSIS
BU2
T1
5.0
BU1
2.3
NE1
BU3
S1
FH1
R18
7.5
0.5
Four-wire cable to laser head section
SOCK8
S2
LED1

Fabricate the front panel holes as required for the components used.Trial fit before drilling.

Figure 4-6 *Fabrication and chassis assembly*

3. If you are building from a perforated board, insert the components, starting in the lower left-hand corner as shown. Pay attention to the polarity of the capacitors that have polarity signs and all semiconductors. Use eight-pin sockets for I1 and I2.

   Route the leads of the components as shown and solder as you go, cutting away unused wires. Attempt to use certain leads as the wire runs or use pieces of the #24 bus wire. Follow the dashed lines on the assembly drawing as these indicate the connection runs on the underside of the assembly board. The heavy dashed lines indicate the use of thicker #20 bus wire, as this is a high-current discharge path.

4. Attach the external leads as shown in Figure 4-4 to the external components. Note the chassis grounding to the green lead of the power cord (CO1) and the use of wire nuts (WN1) for the junction of the transformer's primary inputs.

5. Attach the external leads as shown in Figure 4-5 to the external components. Note the four leads that go to the laser head assembly. These

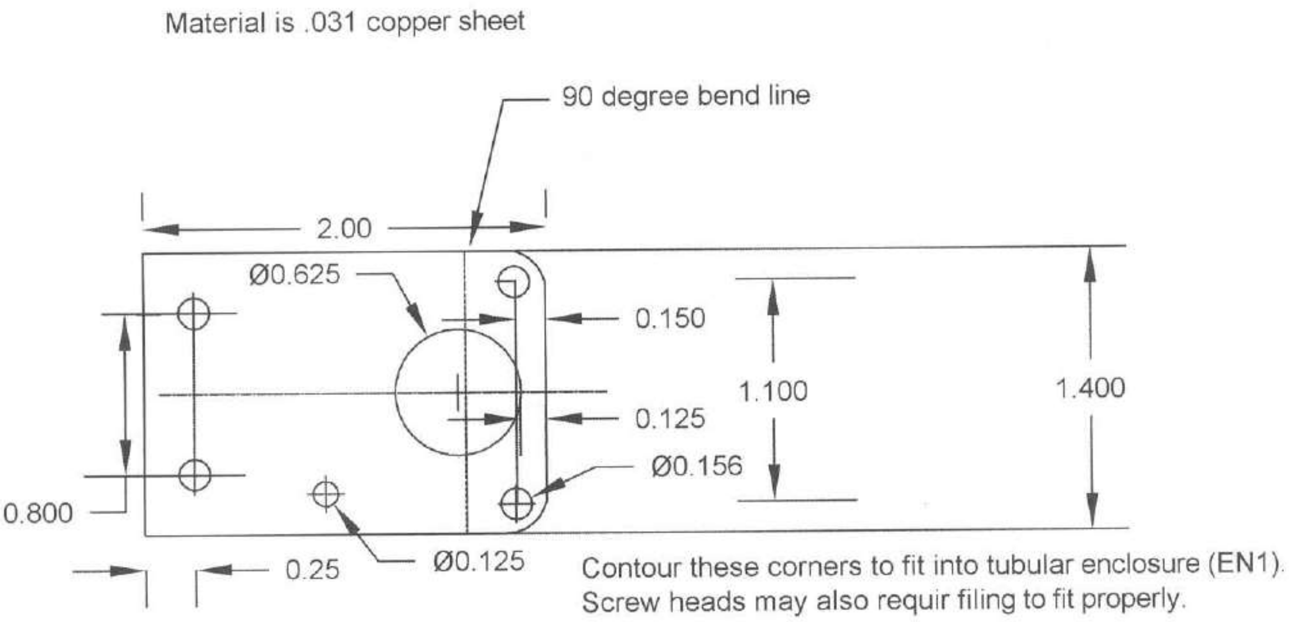
Material is .031 copper sheet
90 degree bend line
2.00
Ø0.625
0.150
1.100
1.400
0.125
Ø0.156
0.800
0.25
Ø0.125
Contour these corners to fit into tubular enclosure (EN1).
Screw heads may also requir filing to fit properly.

Figure 4-7 *Assembly of laser chassis*

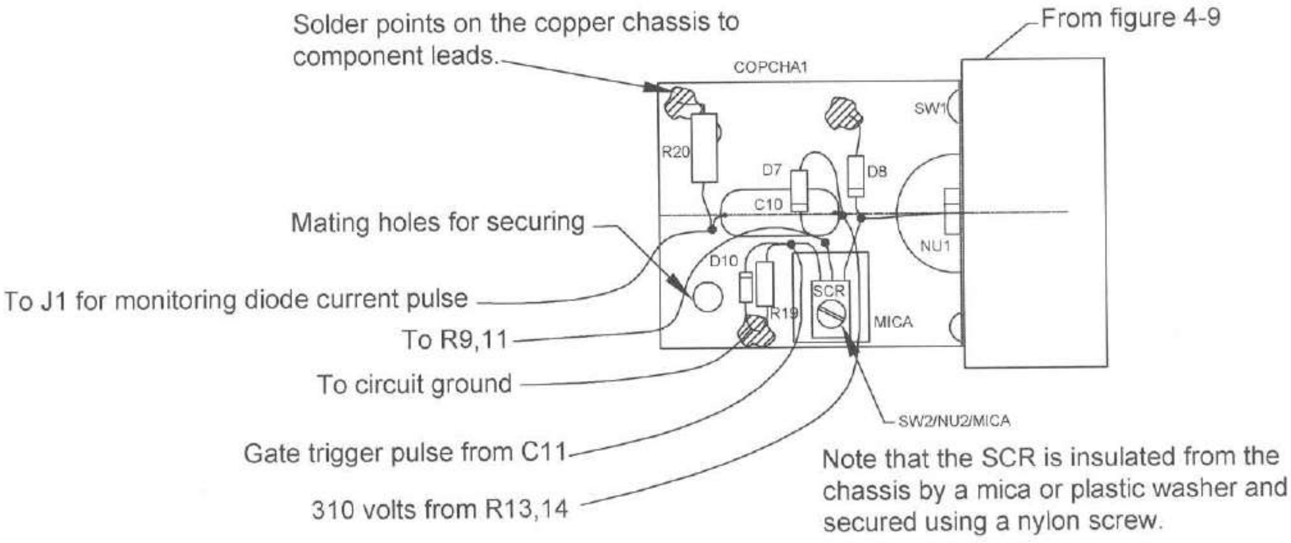
Solder points on the copper chassis to component leads.
From figure 4-9
COPCHA1
SW1
R20
D7
D8
C10
NU1
Mating holes for securing
D10
SCR
R19
MICA
To J1 for monitoring diode current pulse
To R9,11
To circuit ground
SW2/NU2/MICA
Gate trigger pulse from C11
310 volts from R13,14
Note that the SCR is insulated from the chassis by a mica or plastic washer and secured using a nylon screw.

Figure 4-8 *Fabricate the laser holder*

are suggested to be 24 to 36 inches in length as well as twisted and identified. You may use a four-conductor cable.

6. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.
7. Fabricate the chassis, as shown in Figure 4-6. Then position and drill the front and rear panel holes to fit the necessary external components. Use appropriate bushings for the power cord and wires to the laser head. Also, place an insulating piece (PLATE) under the assembly board to prevent contact with the metal chassis.
8. Fabricate the laser chassis, as shown in Figure 4-7, from a .031 piece of copper sheet metal.
9. Assemble the laser chassis, as shown in Figure 4-8. This section takes advantage of the ground plane effect of using the copper for

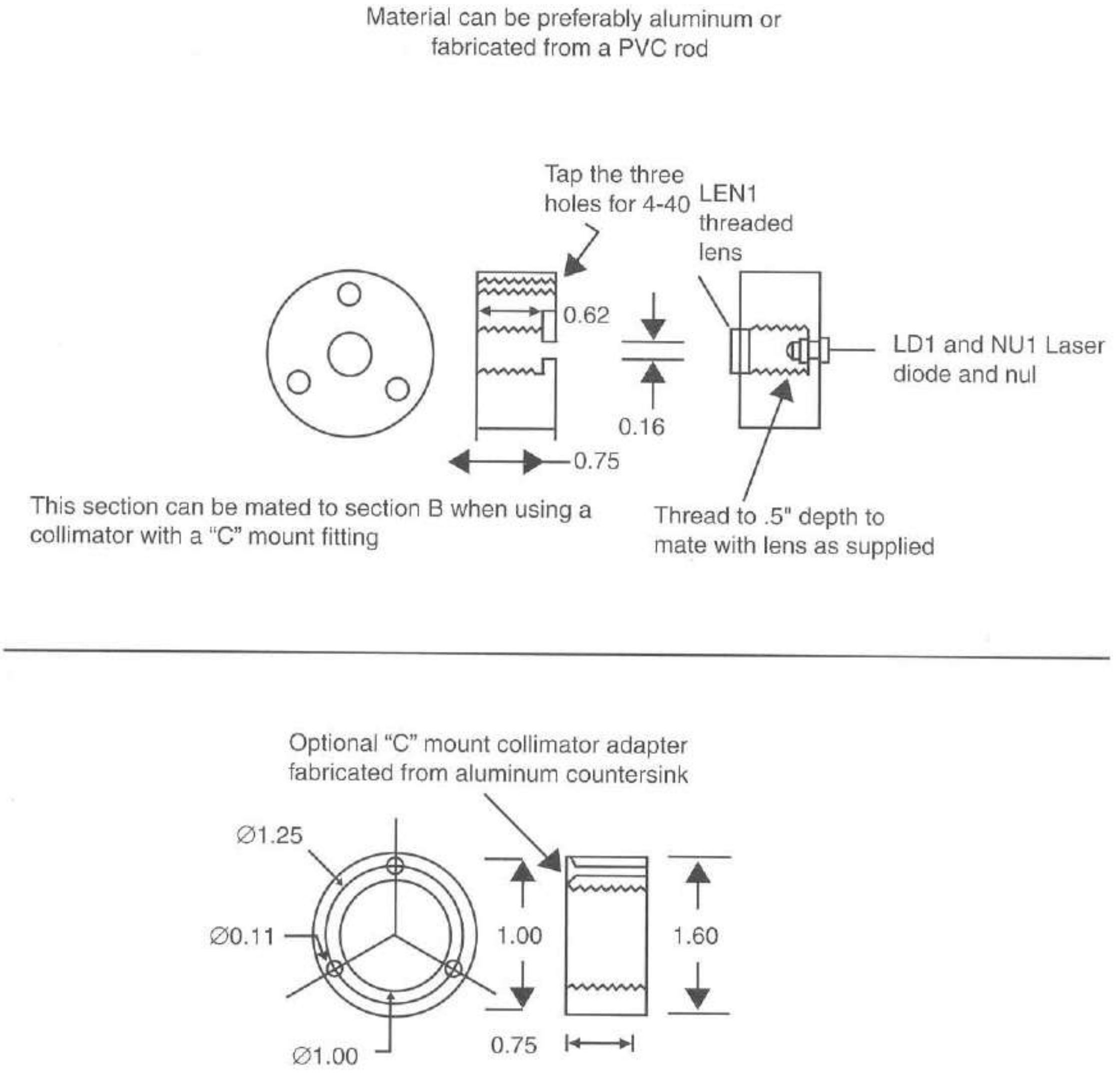
Material can be preferably aluminum or fabricated from a PVC rod
Tap the three holes for 4-40
LEN1 threaded lens
0.62
0.16
0.75
LD1 and NU1 Laser diode and nul
This section can be mated to section B when using a collimator with a "C" mount fitting
Thread to .5" depth to mate with lens as supplied
Optional "C" mount collimator adapter fabricated from aluminum countersink
Ø1.25
Ø0.11
Ø1.00
1.00
1.60
0.75

Section B adapter for fitting to the lens holder when using the optional collimator.

Three mounting holes of .11 on a 1.25" diameter 120 degres are mated to those in above.

Thread for "C" lens at 1" diameter 32 tpi to a minimum of .375".

Figure 4-9 *Laser diode and lens holder*

direct solder points. It is important to maintain a good waveshape because the switching currents are 40 amps and any undershoot is undesirable across the laser diode. Do not install the laser diode at this point.

10. Create the laser holder (LAHOLD), as shown in Figure 4-9. This assembly is attached to the bent-up lip of the chassis section by two small screws.

Assembly Testing

Testing will involve the use of an oscilloscope with a 100 MHz bandwidth and infrared detection equipment such as a night vision scope or camera.

1. It is assumed at this point that the electronic assembly is complete. Verify the circuit and preset the two trimpots, R3 and R9, to midrange.

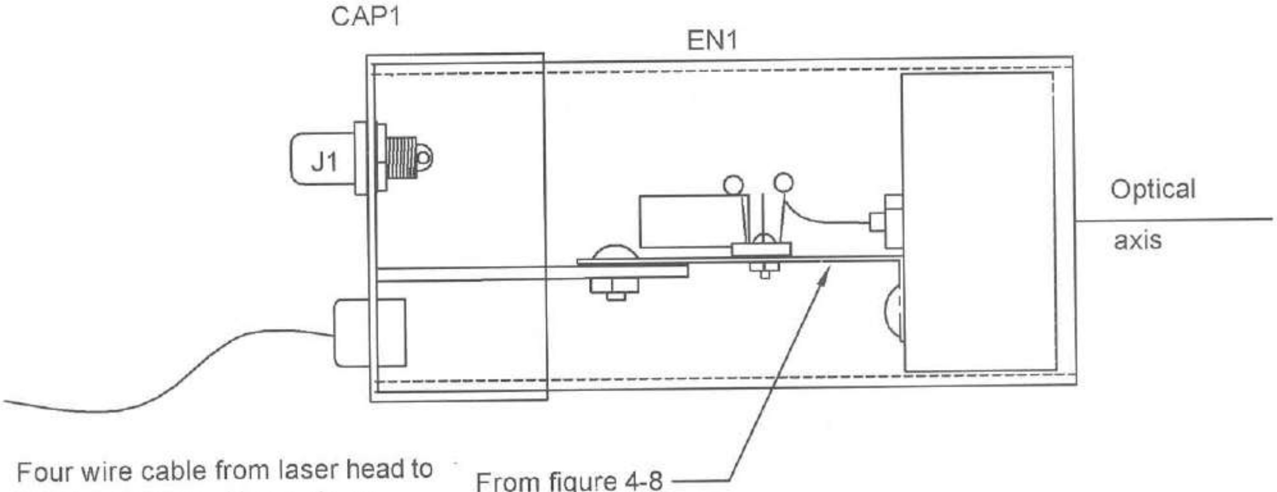

Four wire cable from laser head to control chassis. Use a clamp bushing to secure.

From figure 4-8

Figure 4-10 *Final head assembly*

2. Insert a .5-amp fuse into the fuse holder. Verify that the key switch is off and R18 is fully counterclockwise. Plug in a ballasted, 25-watt, 115-volt, 60 Hz power source. This is nothing more than the cord with a 25-watt lamp in series with the black lead. It protects the circuitry from the catastrophic currents of the power line should a major fault occur. Obtain a 1N4007 power diode and temporarily wire it in place of the laser diode LA1. The diode simulates the actual laser diodes and avoids costly replacement should a gross error exist. The actual laser diode is not to be wired in at this point. Turn on the key switch and note the NE1 coming on. The ballast lamp should not be lit.

13. Connect the scope to TP2 and note a spiky waveform that verifies the trigger pulse to the SCR, as shown in Figure 4-2.

14. Connect the scope to TP3 and note the waveform, which is the voltage across the SCR. Notice the flat spot indicating that capacitor C10 has reached a full charge before the SCR is triggered. The top of this waveshape is affected by the pulse repetition rate and should support 2,500 repetitions per second when set by R18 (see Figure 4-2).

15. Connect the scope to TP1 and set it up for a negative-going fast pulse, indicating the current waveform through the LD1 laser diode. If you have followed the instructions and your assembly of the copper chassis is secure with short leads, you should see a picture-perfect waveform, as shown.

    Note that the current pulse is 40 amps and corresponds to 1 amp for every volt read at TP1. If you find the pulse is less, you can increase the voltage ratings on one of the zener diodes, Z3 or Z4. The laser diode specified requires 40 amps for full output. Any value over 40 amps will greatly shorten the useful life of this part. The range of the pulse rep rate should be 200 to 2,500 pulses per second. The pulse amplitude should only vary several percent throughout this range. You will note that the amplitude will slightly decrease as the rep rate is increased.

16. At this point, you should be wearing your safety glasses. Remove the test diode D5 and connect the laser diode. Place an infrared indicator several inches from the diode and note an orange glow when the unit is turned on. Then recheck the pulse shape and amplitude.

17. Verify the laser output by using a night vision scope or a video camera. Screw in LENS13, so it is flush with the holder. Adjust it to find the smallest footprint at over 50 feet. Our lab model adjusted to about two turns below being flush with the holder for the best results. The output should be a sharp, narrow bar shape with this diode specified.

Please note that laser diodes are very easily destroyed by current pulses that are too large and overshoot. Pay attention to the correct current pulse shown in Figure 4-2.

The recommended laser diodes are available from Information Unlimited and are shown in Table 4-1.

Final Assembly Steps

Final assembly of head section as shown in Figure 4-10. For final assembly place the label as shown and verifying the compliance tests in Figure 4-11

Compliance Tests

1. Verify correct labels as shown.
2. Key switch — nonremoval in off position.
3. Beam indicator LED indicates beam emission.

Instruction

1. Turn key switch counterclockwise to verify unit is off. Note key is easily removed in the off position.
2. Plug unit into 115 vac.
3. Turn key clockwise and note NE1 lighting. Note key cannot be removed in the on position Laser energy is infrared and cannot be seen with the eye. You must use infrared detection devices such as our laser pulse detector or suitable night-vision devices. Video cameras will often detect the laser.
4. Adjust collimator for desired effect if you are using.

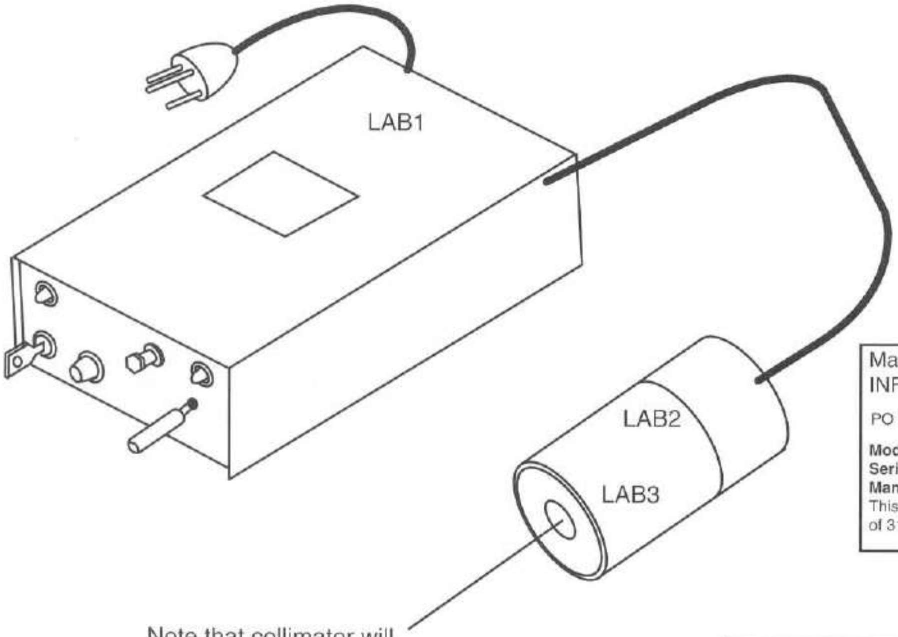

Note that collimator will expand beam at near field but greatly reduce it at far field.

Manufactured By
INFORMATION unlimited
PO Box 716, AMHERST, NH 03031
Model NUMBER
Serial Number
Manufactured Date
This laser product conforms to the provisions of 31CFR1040 10 and 1040 11

LAB1

AVOID EXPOSURE
LASER RADIATION IS EMITTED
FROM THIS APERTURE

LAB3

DANGER

VISIBLE OR INVISIBLE
LASER RADIATION
AVOID DIRECT EXPOSURE
TO BEAM

CLASS IIIB LASER PRODUCT

LAB2

Figure 4-11 *Isometric of final view*

Table 4-2 Parts List for Pulsed Laser

Ref #	*Description*	*DB Part #*
R1	100K, ¼ –watt resistor (br-blk-yel)	
R2	10-ohm, 1-watt resistor (br-blk-blk)	
R3, 4, 5, 6, 7	Two 1K ¼-watt resistors (br-blk-red)	
R8	1M, ¼-watt resistor (br-blk-gr)	
R9, 11	47K, ¼-watt resistor (yel-pur-or)	
R10, 12	1.8K, 3-watt mox resistor	
R13, 14	220, 1/2-watt resistor (red-red-br)	
R15	10-ohm, ¼-watt resistor (br-blk-blk)	
R16, 17	470-ohm, ¼-watt resistor (yel-pur-br)	
R18	10K, 10 mm potentiometer	
R19	39-ohm, ¼-watt resistor (or-wh-blk)	
R20	1-ohm, ½-watt carbon resistor (br-blk-gold)	
C1, 2	Two 330 mfd, 16-volt vertical electrolytic capacitor	
C4	1,000 mfd, 25-volt vertical electrolytic capacitor	
C3, 5, 7	10 mfd, 25-volt vertical electrolytic capacitor	
C6, 9	.01 mfd, 50-volt disc capacitor	
C10	.033 mfd, 400-volt polypropylene capacitor	
C11	.1 mfd 50-volt plastic capacitor	
D1, 2	1N4007 1KV, 1-amp rectifier diodes	
D3–6	Four 1N4001 50-volt 1-amp rectifier	
D7, 8	1N4937 fast recovery 1KV, 1-amp diode	
D9, 10	Two 1N914 signal diodes	
Z1, 2	1N4735 6.2-volt zener diode	
Z3	1N5383 150-volt, 5-watt zener diode	
Z4	1N5384 160-volt, 5-watt zener diode	
Q1, 2	Two 2N3439 NPN high-voltage transistors	
Q3	PN2907 PNP general-purpose transistor	
Q4	PN2222 NPN general-purpose transistor	
I1, 2	Two timer dip integrated circuits LM555	
LED1	Green LED	
NE1	Neon indicator lamp	
SCR1	12-amp, 800-volt SCR S8012R	DB# S8012R
LD1	Laser diode 40-amp, 10–20-watt, 904 nm TO18 package	DB# LD650
LAHOLD	Laser and lens holder fabricated as shown in Figure 4-9	
LENS13	Threaded lens	
T1	115/115 volt, 15-volt-amp isolation transformer	DB# TR115/115
T2	115/12 volt, 1-amp transformer	DB# 12DC/.1
J1	RCA phono jack	
CO1	Three-wire #18 power cord	
S1	Compliant key switch	DB# KEYSWSM
S2	Pushbutton no switch	
FH1	Panel-mount fuse holder	
FS1	1-amp fuse	
WN1	Two small Hi3 wire nuts	
PB1	4.8 × 3.5-inch .1 × .1 perforated circuit vector board	
SOCK8	Two 8-pin integrated circuit sockets	

Chapter Five

Laser Property Guard

This project, as shown in Figure 5-1, allows the user to protect a perimeter of over 1,000 meters. The system utilizes a laser transmitter, as described in Chapter 1, and a laser receiver, described in Chapter 2. The system can easily control any deterrent to an invader, such as flood lighting, alarms, man-trap systems, or our phaser pain field generator, as described in Chapter 35.

The operation of the laser property guard is simple. The laser beam is directed through the area to be protected, using front-surface mirrors located on sturdy and secure mounts and at the points necessary to obtain the required coverage (see Figure 5-2). The receiver is normally quiescent when receiving this beam. Once the beam is broken by an intrusion, the receiver now processes the interruption as a fault and powers the desired deterrent or alarm.

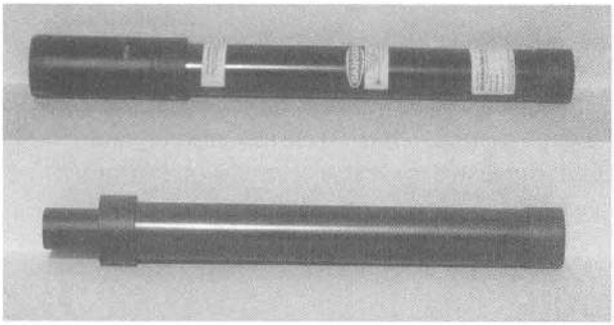

Figure 5-1 *Laser property guard protector components*

This project requires assembly of the laser devices, as described in Chapters 1 and 2. The setup requires basic mechanical skills and installation of the optical reflecting mirrors, and the hookup requires basic electrical skills. Expect to spend $25 to $50 after the assembly of the necessary components. All parts are readily available, with specialized parts obtainable through Information Unlimited (www.amazing1.com) and listed in Table 5-1.

The laser property guard system, as shown in Figure 5-3, utilizes a pulsed infrared laser to provide a perimeter path of programmed light pulses that, when interrupted, will sound an alarm, create an alert, or trigger a deterrent. The trigger fault signal actually triggers a timer or a latching circuit, which drives a switch for the external control of other functions. The timer mode allows the adjustment of the alarm period and automatically resets. The latch mode requires a manual reset.

Even though the laser pulses are rated in tens of watts, the overall energy over a period of 1 second is in the milliwatts. Installing the system at short ranges of only one to two reflection points is reasonably simple. Longer ranges requiring multiple reflections can be challenging.

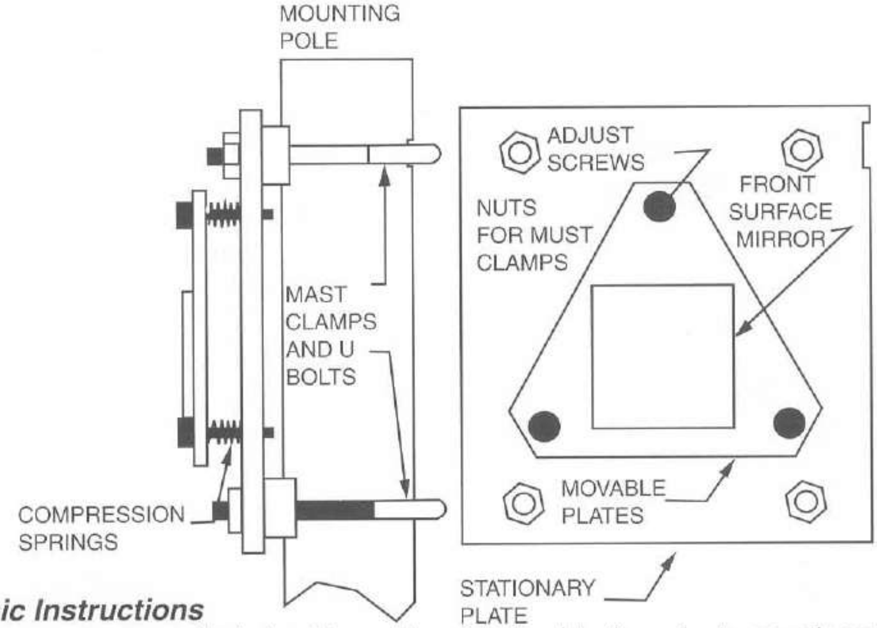

Basic Instructions

Your laser protector guard is designed to provide an intrusion detection perimeter around your home or target area. It uses a Class iiia pulsed diode laser to minimize but not eliminate any optical hazard. A low-liability sonic shock field generator such as that described in Chapter 35 is suggested as a deterrent. Caution: Check local laws for proper posting of this equipment on your property. Remember if you injure a criminal, even if he is robbing you, it can result in a stiff penalty in certain states for violating his rights.

System Setup

1. Obtain the following equipment listed in the latest Information Unlimited catalog.
   - ☐ Class iiia pulsed diode laser as described in Chapter 1.
   - ☐ Optical light detector and controller with built-in alarm as descrbed in Chapter 2.
   - ☐ Low-liability sonic shock pain field generator as described in Chapter 35.

2. Examine the area you want to protect. Attempt to position beam travel over level terrain to eliminate possible easy-to-sneak-under points.

3. Establish corner points of property where mirrors are to be positioned. Note that a clear view of adjacent mirrors, laser and receiver must be maintained in all weather conditions for reliable operation.

4. Determine the corner where you want to position the actual laser and laser detector at. Note: There should be access to these devices for powering and control.

5. Construct mirror mounts using your own ingenuity. Allow for adjustment and stability. Rough align mirror to approximate position.

6. Turn on laser and position to hit center of first mirror. Secure laser in place.

Adjust first mirror to hit center of second and repeat for remaining mirrors all the way back to the optical receiver. Secure in place as you go.

Special note: Large areas may require an optical collimator at the laser output to reduce beam width at longer ranges. The collimator can be a rifle scope, telescope, binocular, etc., positioned in axial alignment with the laser. Beam divergence will be reduced by the magnification factor of the system used, however the actual beam will be expanded by same factor. The net result is a smaller cross-section at longer ranges.

Figure 5-2 *Low-cost, adjustable mirror mount*

System Installation Suggestions

There are many ways to align an optical system of this type. The use of a visible laser module as listed in Table 5-1 can simplify the initial setup and positioning of the mirrors. Those fortunate to have night vision equipment will find this a considerable advantage.

- Determine the area you want to protect. Flat land is the most desirable, as gullies and hills are difficult to cover without using more mirrors.
- You must also consider the position of the laser transmitter and receiver, noting the power and connection into the alarm or the deterrent section. Weather exposure and convenience must also be taken into consideration, and mounting must be stable and somewhat adjustable.
- Determine the path that the light must travel to provide acceptable coverage. Always attempt to use as few mirrors as possible and consider the laser transmitter's and receiver's positions. If you are lucky, you may have some objects to mount the mirrors to. They must be stable and allow placement at the required height for your application.

  The mirrors used for this project must be the front-surface-coated type if multiple reflections are required. Regular household mirrors can be used up to two reflection points, but they will cause multiple reflections. A low, adjustable-cost mirror mount is shown in Figure 5-2. You may in certain cases use a good-quality automotive mirror, as they are easily adjustable.
- Evaluate the wildlife in your area and determine the best laser height that will provide minimal false alarms from the animal population. Obviously, deer will always be a problem, as their height may be close to that of the target intruder.
- Determine the deterrent you wish to use against the intruder. You can activate bright lights, sonic pain generators (described in Chapter 35), alarms, or just alerts so you can take further action. For more serious applications, you can activate shocking devices, such as those described in Chapter 11, nonlethal kinetic devices, or explosive charges for shock and surprise. Very serious deterrents can consist of deadly force, such as high-powered kinetic devices, shotgun traps, high-powered shocking devices, and buried explosives for lethality.

Caution: As man trapping is illegal in the United States, it may also be so in your country. Always check your laws before ever implementing a deterrent that can cause injury.

- The quality of the mirrors and their mounts cannot be overstressed when designing a high-integrity system. Here we discuss a low-cost system that may suffice in smaller installations but may be limited for use in larger, more reliable systems.

The initial alignment requires a visible, low-power laser to align the mirrors. The laser may have to be collimated for larger multiple reflection systems.

Optics on the laser light transmitter and receiver, as shown, may be sufficient for small systems. They do not have to be high-cost precision devices, as the only objective is to establish a properly aligned, position-stable beam spot.

Once the preliminary alignment is accomplished, you can put the laser transmitter in place and, with a night vision scope, recheck the alignment. The laser receiver, as constructed here, has an indicating *light-emitting diode* (LED). This function can be eliminated if required. The actual control of external functions is via a set of normally open or closed relay contacts that can handle 115 VAC at 10 amps AC.

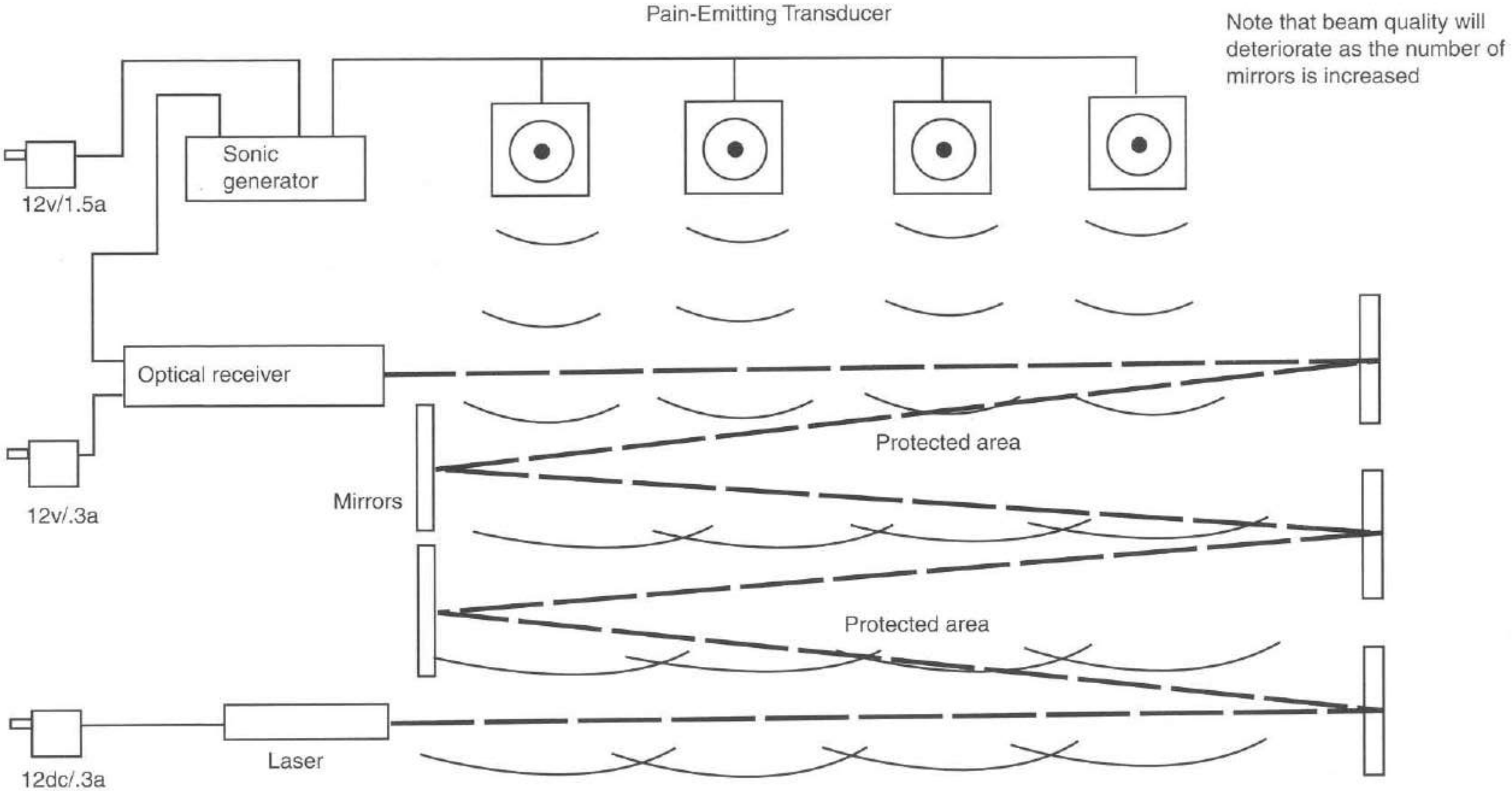
Pain-Emitting Transducer
Note that beam quality will deteriorate as the number of mirrors is increased
Sonic generator
12v/1.5a
Optical receiver
12v/.3a
Mirrors
Protected area
Protected area
Laser
12dc/.3a

Laser beam is reflected back and forth in the protected area via strategically placed mirrors. Once the beam is broken by an intruder the optical receiver detects and sends a level to the sonic generator where the transducers are activated producing an intolerable sonic pain field discouraging the intrusion.

Even though there is no lasting effect from exposure to the transducers, it is always best to check with local laws before implementing this system.

Figure 5-3 *Example of system layout*

Table 5-1 Laser Property Guard Parts List

Ref. #	*Description*	*DB Part #*
MIRROR	3 × 4-inch front surface mirrors	DB# MIR FRSUR
LRG4K	Pulsed infrared laser kit, as in Chapter 1	DB# LRG4K
LRG40	Assembled pulsed infrared laser, as in Chapter 1	DB# LRG40
LSD4K	Laser detector kit, as in Chapter 2	DB# LSD4K
LSD40	Assembled laser detector, as in Chapter 2	DB# LSD40
PPF4K	Pain field generator kit, as in Chapter 35	DB# PPF4K
PPF40	Assembled pain field generator, as in Chapter 35	DB# PPF40
SHOCK1K	Kit of shockers, as in Chapter 11	DB# PFSHK1K
SHOCK10	Assembled shockers, as in Chapter 11	DB# PFSHK10
LM650-3	Visible red laser module for alignment aid	DB# LM650- Description P3System

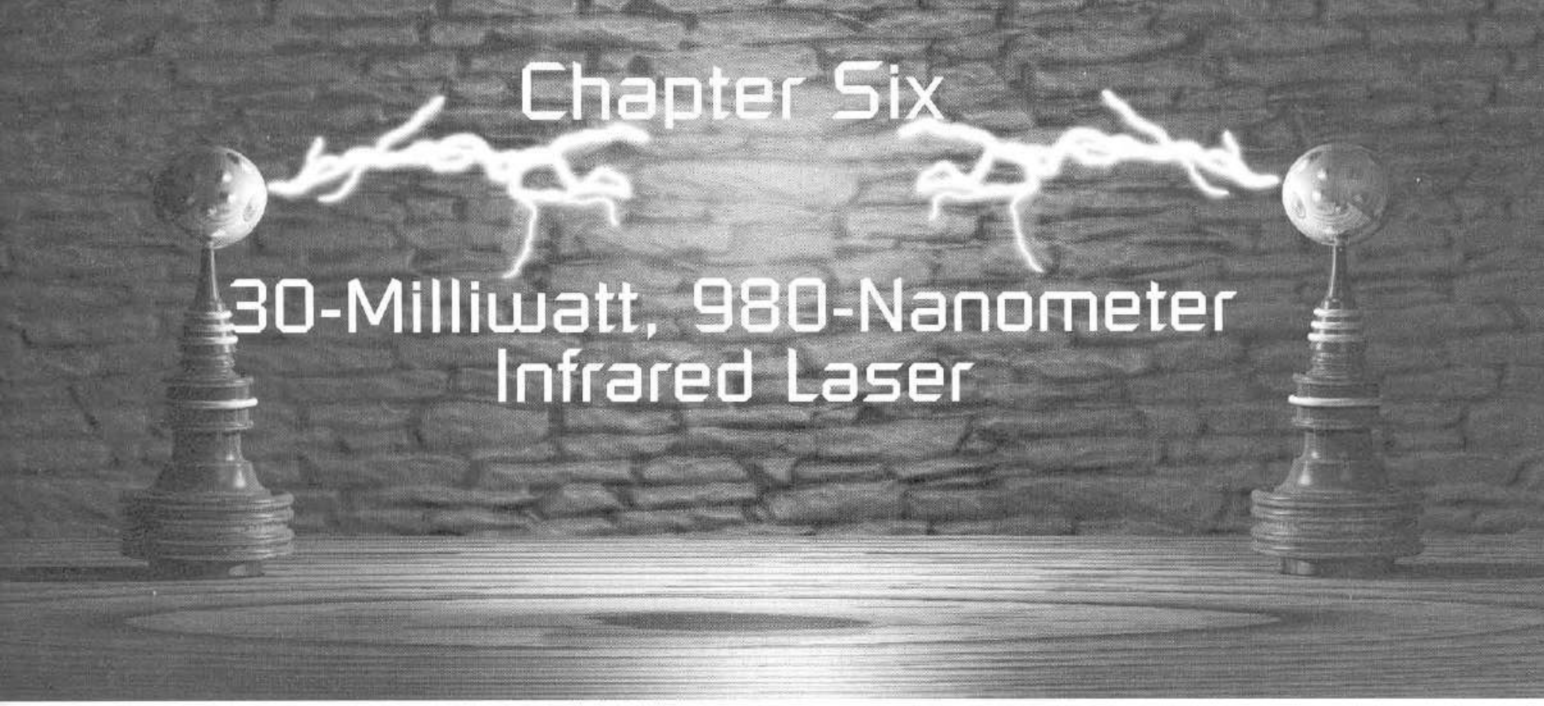

Chapter Six

30-Milliwatt, 980-Nanometer Infrared Laser

The laser project shown in Figure 6-1 is a continuous infrared device providing a continuous 30-milliwatt output at 980 nanometers. The laser, when built as shown, should comply with a class 3 laser-compliant system. The completed laser can be very optically dangerous, and eye protection must be worn to prevent direct or scattered reflections to the eyes.

The project can be used for many optical applications, including illuminating a surface, such as a window, for laser listening. Laser listening is were a laser beam illuminates a window and a sensitive receiver picks up the light reflections and processes them into audio signals similar to the picked-up voices that vibrate the illuminated window glass. A system such as this is described on www.amazing1.com. It should be noted that using a class 3b laser for this application can be an optical hazard to anyone in the target area.

Converting illumination using night vision cameras and spectrally compatible viewers is another potential application, again being fully aware of a possible optical hazard. Other applications are invisible target identification, voice and data communication, and optical signaling and experimenting.

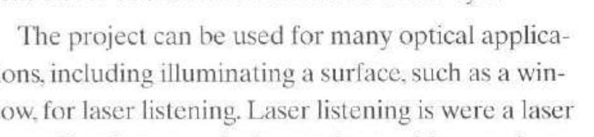

Figure 6-1 *The infrared laser module*

This is an intermediate-level project requiring basic electronic skills. Safety glasses are strongly advised, as the laser can cause eye damage if viewed directly or via direct reflections. Expect to spend $75 to $150. All parts are readily available, with specialized parts obtainable through Information Unlimited (www.amazing1.com), and they are listed in Table 6-1.

For those wanting to use this project as an illuminating laser for window bounce listeners, it is suggested that you refer to Chapter 13 in *Electronic Gadgets for the Evil Genius*, published by McGraw-Hill, ISBN 0-07-142609-4.

Circuit Description

Figure 6-2 shows the laser module LM1 and associated support circuitry. The system is powered by batteries B1 and B2 and is controlled by an internal key switch (S1) or external enabling control leads. A light-emitting diode (LED1) and a current-limiting resistor (R1) indicate the presence of laser emission.

The system will require S1 to be a key switch with a nonremovable key in the activated position if the assembly is integrated as a fully functional laser system. The power line or bench operation will require a

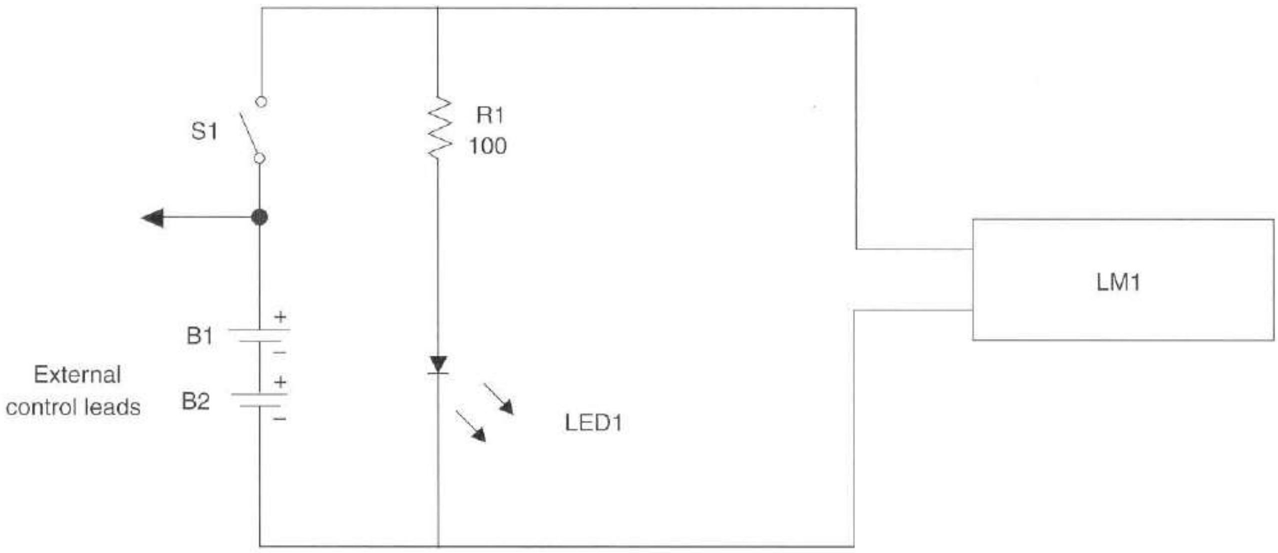

Laser current should be 50 to 60 milliamps at 3 volts input.

Warning: If you overvolt or reverse polarity on your laser module, you will ruin it!!

Assembly shows a battery-operated compliant laser system. You may wire direct to a 3-volt source that now contains the key switch and a built-in time delay necessary for compliance and safety.

Figure 6-2 *Circuit schematic*

key switch for external enabling and associated control circuitry that may require an emission delay and automatic resetting should a loss of power occur.

Assembly

1. Rework the laser module, as shown in Figure 6-3. This step requires care as it is easy to damage the module if one is not very careful.
2. Connect the soldered leads to a source of 3 volts, paying attention to the correct polarity. If you have a current meter, you can measure 40 milliamps.

   Output can be detected using infrared photosensitive paper or a night vision scope. Certain video cameras may also detect the laser output.

   Danger! Do not view directly or indirectly without safety glasses.
3. Fabricate the laser holder (HOLD1), as shown in Figure 6-4.
4. Fabricate the laser module tube (TUB1) from a piece of 3-inch × 1-inch OD × 1/16-inch wall plastic tubing, as shown in Figure 6-5.

   Fabricate the two spacing washers (WASH1, 2) from 1/16-inch plastic. Note the inner hole is 1 inch to allow for the sleeving onto TUB1 as shown. The outer diameter fits into the inner diameter of the main enclosure (EN1). It is these two spacer washers that position and secure TUB1 into the center of EN1. Space the two washers as shown.
5. Insert the laser module LM1 as shown and wire it up to the circuitry as shown. It is a good idea to leave enough lead length to allow one to pull out the battery holder for battery replacement by removing the rear cap (CAP1). Use a piece of foam to secure the battery holder in place.
6. Verify the proper circuit operation so that when the key switch is activated, the LED emits, indicating laser emission.

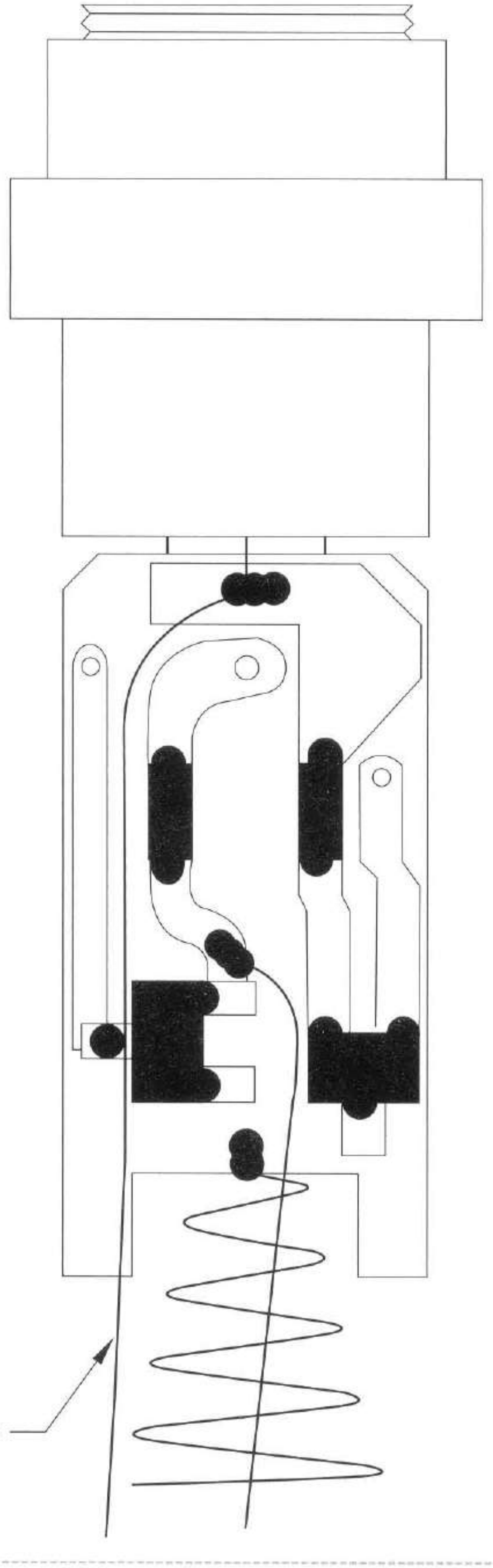

Very carefully attach #24 bare bus leads to the solder pads as shown. It is necessary to insulate the positive lead using a piece of insulation stripped from a piece of wire or insulated tubing.

Note that the brasshead of the laser is the positive connection and the spring the negative. Activation will require now depressing the on-board pushbutton switch. You may connect to these points if unsure of soldering to the small foils on the printed circuit board.

Warning: If you overvolt or reverse polarity on you laser module, you will ruin it!!

This lead must be insulated

Figure 6-3 *Laser module rework*

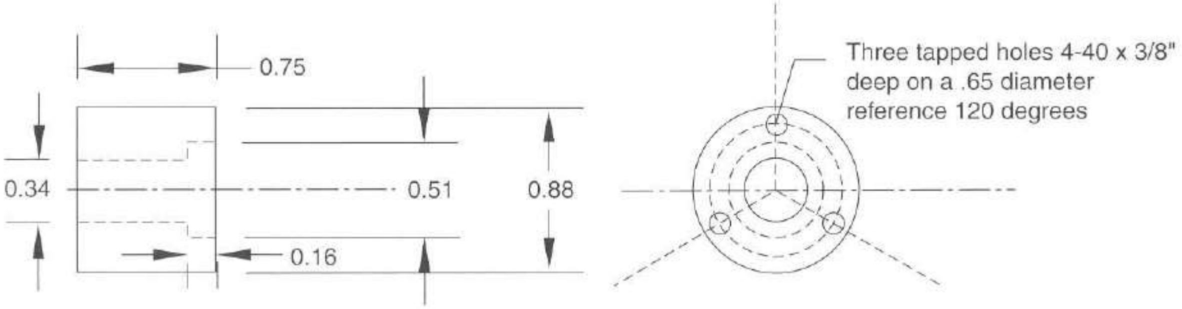
0.75
0.34
0.51
0.88
0.16
Three tapped holes 4-40 x 3/8" deep on a .65 diameter reference 120 degrees

HOLD1 Holder material: PVC

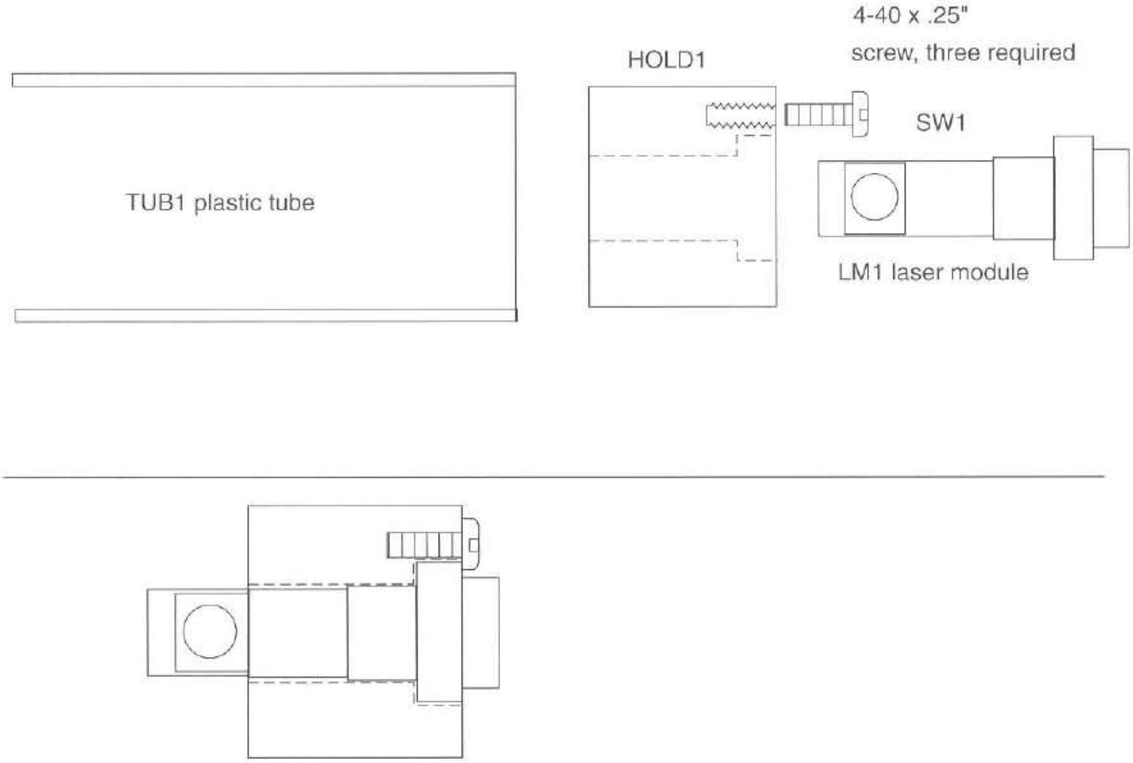
4-40 x .25" screw, three required
HOLD1
SW1
TUB1 plastic tube
LM1 laser module

X-ray view showing assembled LM1 laser module inserted into HOLD1 holder with three retaining screws SW1

Figure 6-4 *Creation of laser module holder and retaining washer*

This is our approach to assembling this laser project with external optics. The builder may use his own ingenuity noting that the objective is optical stability and safety.

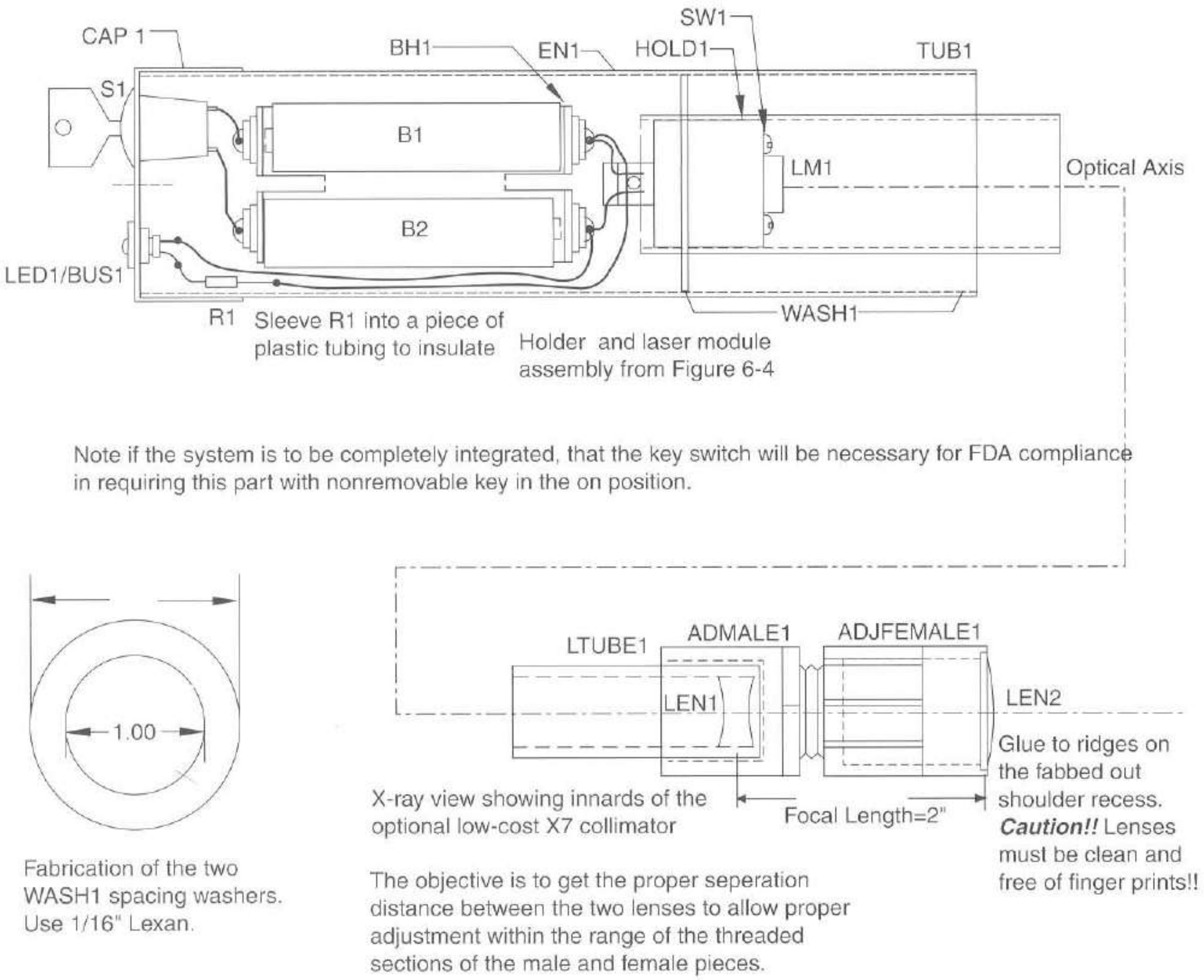

Sleeve R1 into a piece of plastic tubing to insulate

Holder and laser module assembly from Figure 6-4

Note if the system is to be completely integrated, that the key switch will be necessary for FDA compliance in requiring this part with nonremovable key in the on position.

X-ray view showing innards of the optional low-cost X7 collimator

Glue to ridges on the fabbed out shoulder recess. ***Caution!!*** Lenses must be clean and free of finger prints!!

Fabrication of the two WASH1 spacing washers. Use 1/16" Lexan.

The objective is to get the proper seperation distance between the two lenses to allow proper adjustment within the range of the threaded sections of the male and female pieces.

Figure 6-5 *X-ray of complete assembly*

Collimator and Final Assembly

1. Glue the larger lens LEN2 into ADFEMALE1 as shown, using silicon rubber glue or an equivalent. Clean the lens before gluing.
2. Insert LEN1 into the laser output end of the LTUB1 assembly as shown. Do not glue at this time.
3. Fully insert the LTUB1 into the 1-inch tube TUB1. Shim it with tape for a secure but removable fit.
4. Finish the assembling and label it as shown in Figure 6-6.
5. Point the laser at a target around 100 meters away and adjust the collimator to the smallest spot. This should be done in low light, preferably at night, using a night vision device and only looking at the beam impact point. Note

Instructions

1. Turn key switch full ccw. Note switch has several positions–only two are used in this version.
2. Remove CAP1 and insert 2 AA batteries into holder.
3. Turn on key two positions cw and note LED lighting. If not check the battery holder.
4. Adjust coliminator for desired effect.

Note that coliminator will expand the beam at near field but greatly reduce it at far field. Laser module may require beam adjustment over time using the special tool.

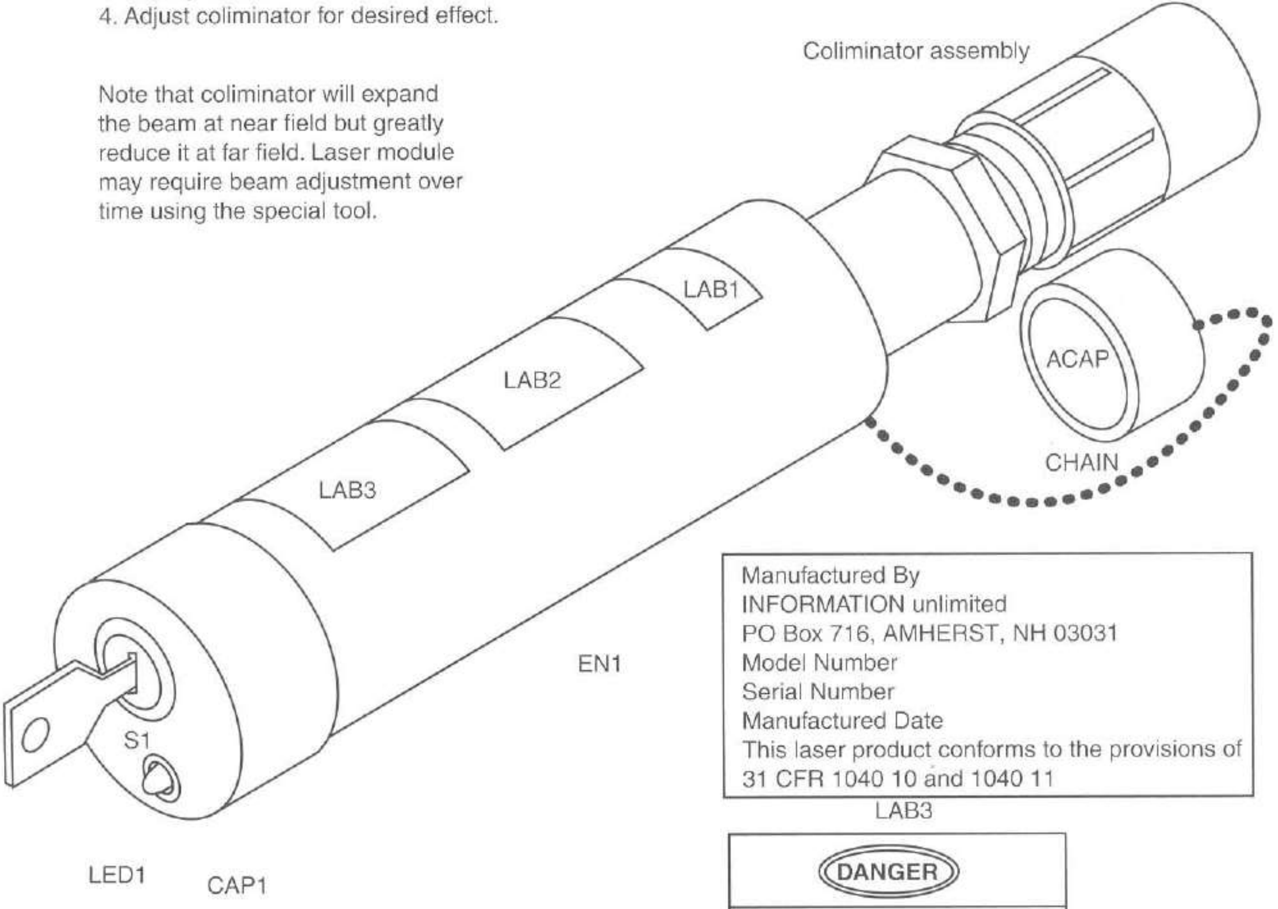

Manufactured By
INFORMATION unlimited
PO Box 716, AMHERST, NH 03031
Model Number
Serial Number
Manufactured Date
This laser product conforms to the provisions of
31 CFR 1040 10 and 1040 11

LAB3

DANGER

VISIBL AND/OR INVISIBLE
LASER RADIATION
AVOID DIRECT EXPOSURE
TO BEAM

30 Milliwatts @ 980 nm
CLASS 3B LASER PRODUCT

LAB2

AVOID EXPOSURE

INVISIBLE LASER RADIATION IS
EMITTED FROM THIS APERTURE

LAB1

Compliance test

1. Verify correct labels as shown
2. Key switch–nonremoval in off position
3. Beam indicator LED indicates after delay
4. Aperture cap included

WARNING: Do not under any circumstances point this device at people, vehicles, or especially aircraft. YOU WILL BE PROSECUTED IF CAUGHT!!!

Figure 6-6 *Final assembly view*

that the adjustment should be about 1/2 to 3/4 turns from being fully tight for a mechanically stable operation. If it is too far out, the assembly will be loose. You may have to use a wire brush on the threads and lube with dry teflon for smooth action. You may compensate the distance between lenses by changing the length of TUBE1. You may now glue LEN1 in place as noted.

The collimator will expand the beam width but decrease the divergence by the same ratio. This greatly reduces the far-field beam diameter.

Note that the laser module is at a wavelength of 980 nanometers and is invisible to the eye. Also, note that most of the mechanical parts used in this project can be obtained through most hardware stores.

Table 6-1 Infrared Laser Parts List

Ref. #	*Description*	*DB Part #*
R1	220-ohms, 1/4-watt resistor (red-red-br)	
LED1	High-brightness green LED	
LM1	30-milliwatt, 980-nanometer infrared laser module	DB# LM980-25
S1	Key switch with nonremovable key in "on" position	
BH1	Dual, side-by-side two-AA battery holder	
WR24	12-inch #24 bus wire	
WR22	24-inch #22 vinyl hookup wire	
HOLD1	Fabricated holder with washer for laser module	DB# HOLD1
SWI/NU1	Two 4-40, 1.25-inch machine screws and nuts	
TUB1	3 × 1-inch OD × .875-inch plastic or metal tube	
WASH1	Two fabricated plastic spacing washers, 1.5 OD with 1-inch center hole	
EN1	6-inch × 1.625 OD × 1.5-inch ID plastic or metal tube	
CAP1	1 5/8-inch plastic slip on cap	
Parts for low-cost collimator		
LEN1	6 mm × (-1) double-concave negative glass lens	DB# LENS13
LEN2	24 mm × 75 mm plano/convex glass lens	DB# LE2475
*LTUB1	625-inch ID × 2-inch schedule 40 PVC tube	
*ADMALE1	1/2-inch PVC schedule 40 slip fit to male thread, GENOVA 30405	
*ADFEMALE1	1/2-inch PVC schedule 40 slip fit to female threads, GENOVA 30305	
*ACAP1	Plastic cap and chain for aperture cap	
LAB1	Aperture label	
LAB2	Classification label	
LAB3	Certification label	

*Available in most hardware stores

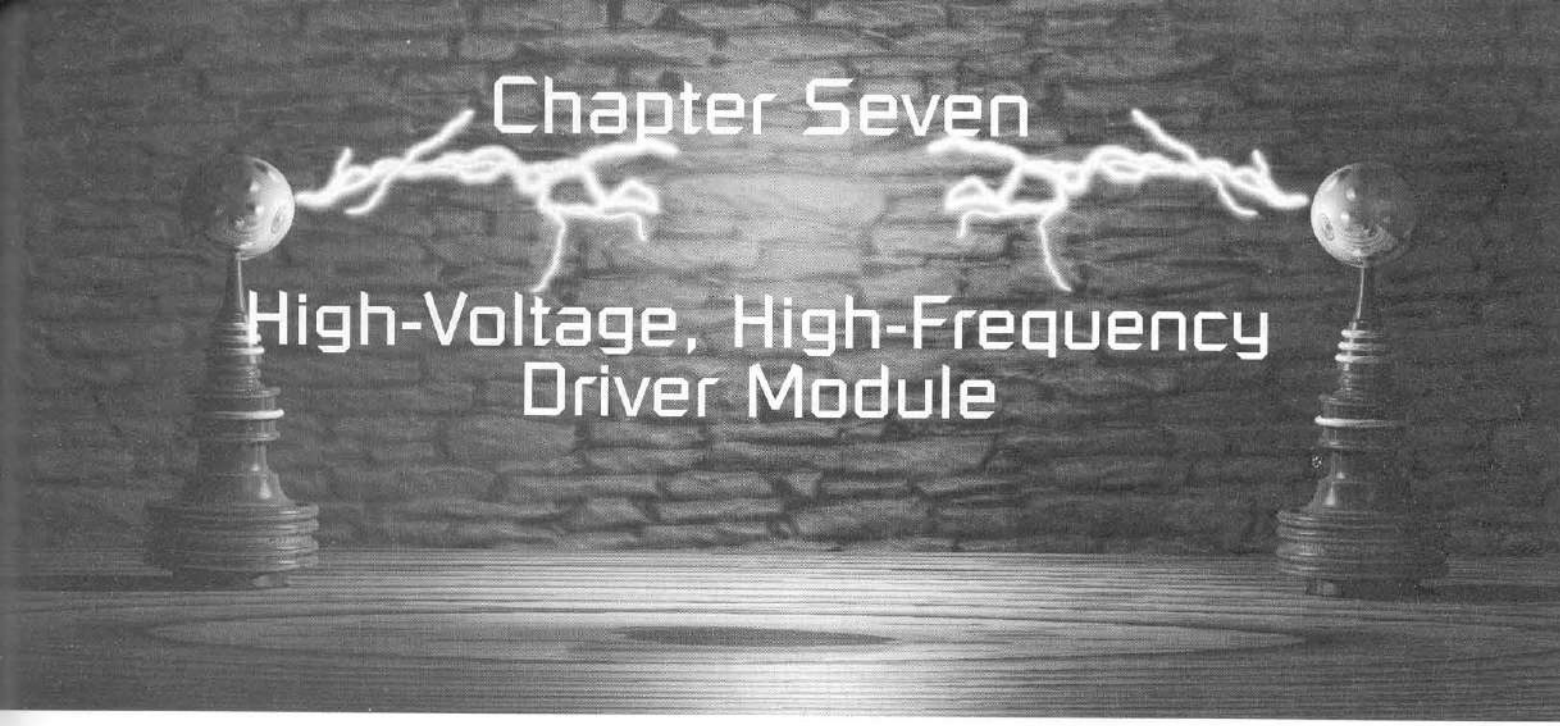

Chapter Seven

High-Voltage, High-Frequency Driver Module

Figure 7-1 shows that a universal, high-voltage, modular power supply project will provide you with many options. This "sweetheart" of a circuit has features that make it a priceless tool for the high-voltage researcher and experimenter.

The circuit operates on 11 to 15 volts DC, drawing 3 amps under a full load, allowing a portable battery or 115 volts AC via a converter. The output voltage is a 60 KHz, high-frequency current that is fully short-circuit protected. The high-frequency output also makes it possible to use low-energy voltage multiplier stacks for high-voltage DC sources, and it can also serve as an excellent plasma driver when used directly. The output current is fully adjustable via a control pot.

The unit is excellent for powering neon and all types of gas-filled vessels using one or two electrodes, or it can power objects simply by proximity. It easily retrofits to our voltage multipliers' modules that provide DC voltages up to 100 kilovolts and currents of up to .3 milliamperes. The current-limiting and control features make this combination an excellent choice for charging capacitors for low loss charging, utilizing 12-volt portable or 115 vac line operation. Also, the device is an excellent choice for powering large and small antigravity craft, ozone air purification, and other applications requiring a high-voltage current-controlled source. The module is shown built on both a rugged *printed circuit board* (PCB) and a vector perforated box. It is mounted in a plastic channel.

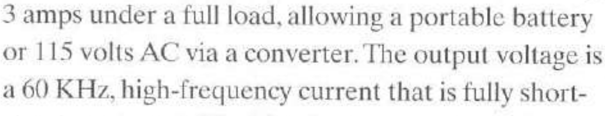

Figure 7-1 *Completed assembly module*

The module is used for several projects in this book. In order not to waste a lot of space, we show it as a common subassembly for the following chapters:

- Chapter 12, "Electro-magnetic (EMP) Gun"
- Chapter 32, "100-Milligram Ozone Machine"

Electrical and Mechanical Specifications

- Open-circuit voltage, 7,500 peak at 60 KHz
- Short-circuit current, 10 milliamperes short-circuit protected
- Input 11 to 15 volts DC at 3 amps fully loaded

- Adjustable current by duty-cycle-controlled pulse
- Compact size, 7 × 2$^{1}/_{8}$ × 1$^{1}/_{8}$ inches, weighing less than 5 ounces
- Easily retrofits to our voltage multipliers

This is an intermediate-level project requiring basic electronic skills and should cost between $25 and $50. All parts are readily available, with specialized parts obtainable through Information Unlimited (www.amazing1.com), and they are listed in Table 7-1.

High-voltage DC output is obtained from this module using a Cockcroft-Walton voltage multiplier with multiple stages of multiplication as required. Note that this method of obtaining high voltages was used in the first atom smasher that ushered in the nuclear age.

The multiplier section requires a high-voltage AC source for input supplied by the circuit transformer (T1) producing 6 to 8 kilovolts at approximately 60 KHz. You will note that this transformer is of a unique design, being owned by our company Information Unlimited. The part is very small, versatile, and lightweight for the power produced.

Circuit Description

The primary of T1 is current driven through inductor L1 and switched at the desired frequency by FET switch Q1. Capacitor C6 is resonated with the primary of T1 and zero voltage switches when the frequency is properly adjusted. (This mode of operation is very similar to class E operation.) The timing of the drive pulses to Q1 is therefore critical to obtain optimum operation (see Figure 7-2).

The drive pulses are generated by a 555 timer circuit (I1) connected as an astable multivibrator with a rep rate determined by the setting of trimpot R1 and the fixed-value timing capacitor C2.

The timer circuit is now turned on and off by a second timer, I2. This timer operates at a fixed frequency of 100 Hz but has an adjustable duty cycle (a ratio of on to off time) determined by the setting of control pot R10. I1 is now gated on and off with this controlled pulse now providing an adjustment of output power.

When the unit is interfaced with a DC voltage multiplier, an over-voltage protection spark gap is placed across the output and is easily set to break down at a preset voltage level for circuit protection. Even though the output is short-circuit protected against continuous overload, constant hard discharging of the output can cause damage and must be limited. A pulse-current-limiting resistor (R17) helps to protect the unit from these catastrophic current spikes.

Power input is controlled by switch S1, which is part of control pot R10. Actual power can be a small battery- capable of supplying up to 2 amps or a 12-volt, 2 to 3-amp converter for 115-volt operation.

Construction

The circuit is shown using the more challenging *perforated circuit board* often required for a science fair project. A *printed circuit board* is also individually available, requiring that you identify only the particular part and insert it into the respective holes as noted. The PCB is plainly marked with the part identification, and soldering is now very simple as you solder the component leads to the conductive metal traces on the underside of the board.

The perforated board approach is more challenging, as now the component leads must be routed and used as the conductive metal traces. We suggest that the builder closely follow the figures in this section and mark the actual holes with a pen before inserting the parts. Start from a corner, using it as a reference, and proceed from left to right. Note that the perforated board is the preferred approach for science projects, as the system looks more homemade.

Board Assembly Steps

If you are a beginner it is suggested to obtain our *GCAT1 General Construction Practices and Techniques*. This informative literature explains basic practices that are necessary in proper construction of electromechanical kits and is listed in Table 7-1.

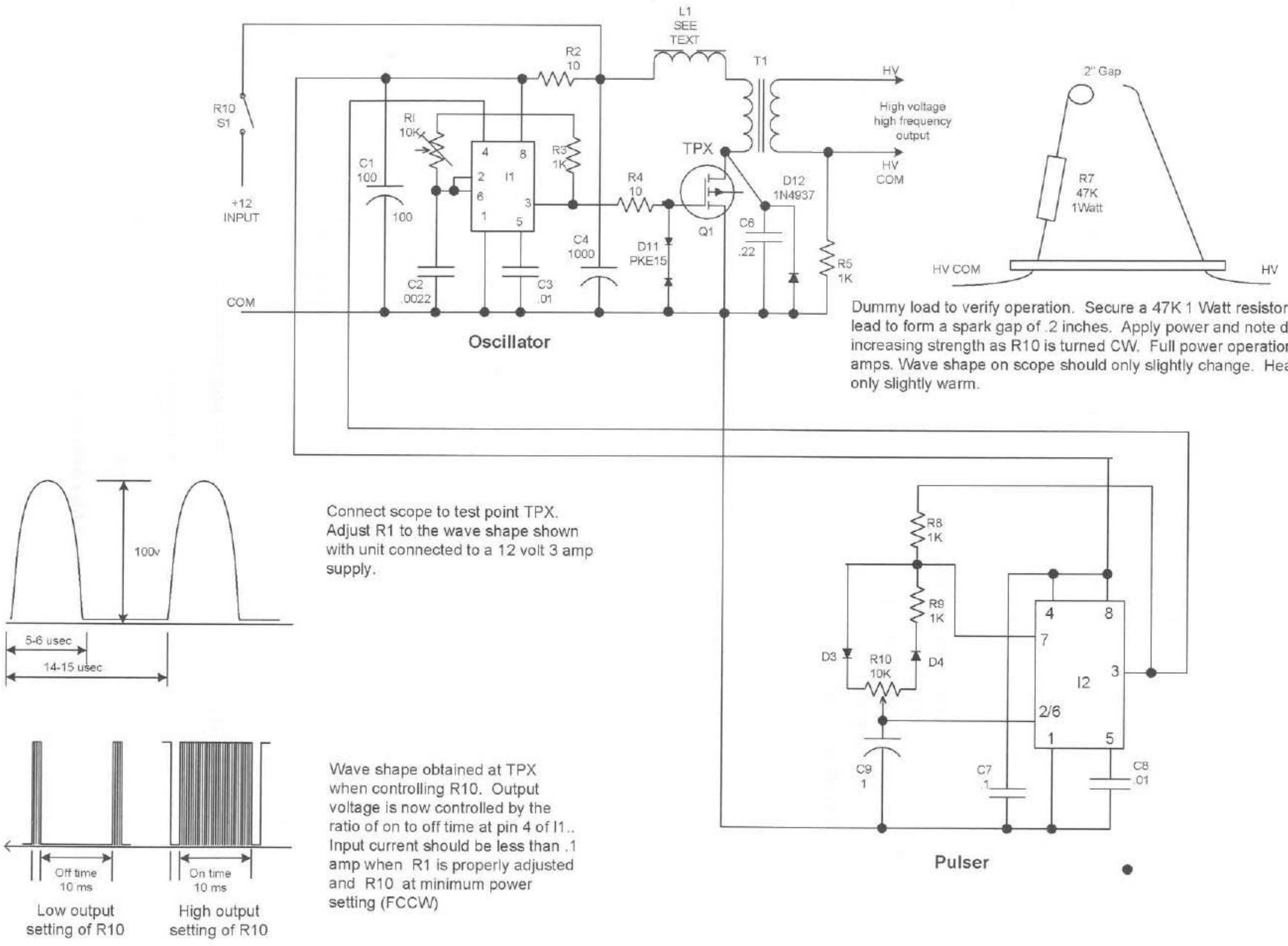
L1
SEE
TEXT
T1
R2
10
R10
S1
+12
INPUT
COM
RI
10K
C1
100
100
C2
.0022
I1
4
8
2
6
1
3
5
C3
.01
R3
1K
C4
1000
R4
10
D11
PKE15
TPX
Q1
C6
.22
D12
1N4937
R5
1K
HV
High voltage
high frequency
output
HV
COM
Oscillator
2" Gap
R7
47K
1Watt
HV COM
HV
100v
5-6 usec
14-15 usec
Off time
10 ms
On time
10 ms
Low output
setting of R10
High output
setting of R10
R8
1K
R9
1K
D3
R10
10K
D4
I2
4
8
7
3
2/6
1
5
C9
1
C7
.1
C8
.01
Pulser

Dummy load to verify operation. Secure a 47K 1 Watt resistor and a stiff wire lead to form a spark gap of .2 inches. Apply power and note discharge increasing strength as R10 is turned CW. Full power operation should draw 1.5 amps. Wave shape on scope should only slightly change. Heat sink should be only slightly warm.

Connect scope to test point TPX. Adjust R1 to the wave shape shown with unit connected to a 12 volt 3 amp supply.

Wave shape obtained at TPX when controlling R10. Output voltage is now controlled by the ratio of on to off time at pin 4 of I1.. Input current should be less than .1 amp when R1 is properly adjusted and R10 at minimum power setting (FCCW)

Figure 7-2 *Circuit schematic*

1. Lay out and identify all the parts and pieces. Verify them with the parts list, and separate the resistors, as they have a color code to determine their value. Colors are noted on the parts list.
2. Cut a piece of .1-inch grid perforated board to a size of 5 × 3 inches. Locate and drill the holes, as shown in Figure 7-3. An optional PCB is individually available (refer to Table 7-1).
3. Fabricate a metal heatsink for Q1 from a piece of .063 aluminum, at 1.5 × .75 inches, as shown in the inset of Figure 7-4.
4. Assemble L1 as shown in the Figure 7-4 inset. This part is individually available and listed in Table 7-1
5. If you are building from a perforated board, it is suggested that you insert components starting at the lower left-hand corner, as in Figure 7-5. Pay attention to the polarity of the capacitors with polarity signs and all semiconductors. Route the leads of the components as shown and solder as you go, cutting away unused wires. Attempt to use certain leads as the wire runs or use pieces of the #24 bus wire. Follow the dashed lines on the assembly drawing, as these indicate connection runs on the underside of the assembly board. The heavy dashed lines indicate the use of thicker #20 bus wire, as this is a high-current discharge path.

   Figures 7-6 and 7-7 show the available PCB, which is also listed as an available part in Table 7-1.
6. Attach external leads as shown in Figure 7-5.
7. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.

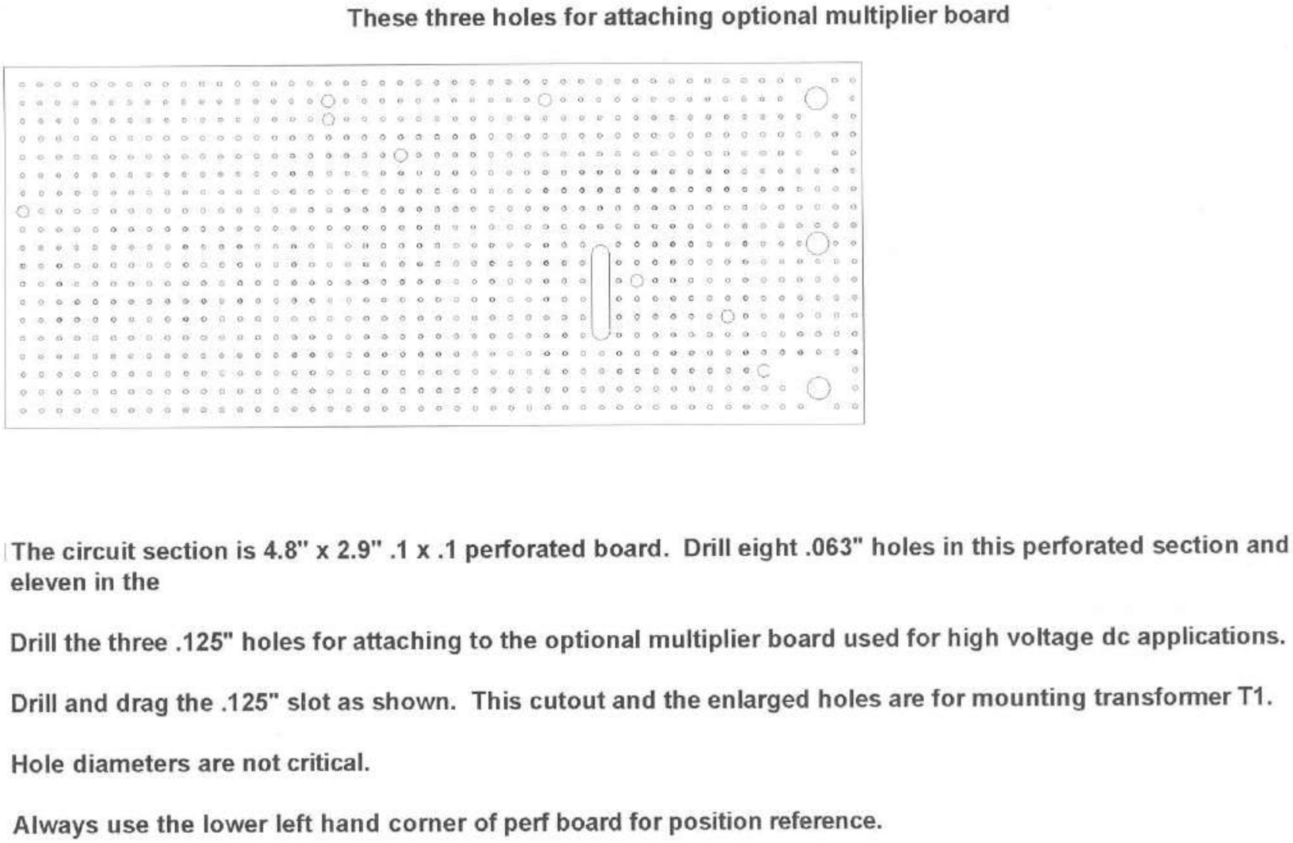

The circuit section is 4.8" x 2.9" .1 x .1 perforated board. Drill eight .063" holes in this perforated section and eleven in the

Drill the three .125" holes for attaching to the optional multiplier board used for high voltage dc applications.

Drill and drag the .125" slot as shown. This cutout and the enlarged holes are for mounting transfomer T1.

Hole diameters are not critical.

Always use the lower left hand corner of perf board for position reference.

Figure 7-3 *Perforated board*

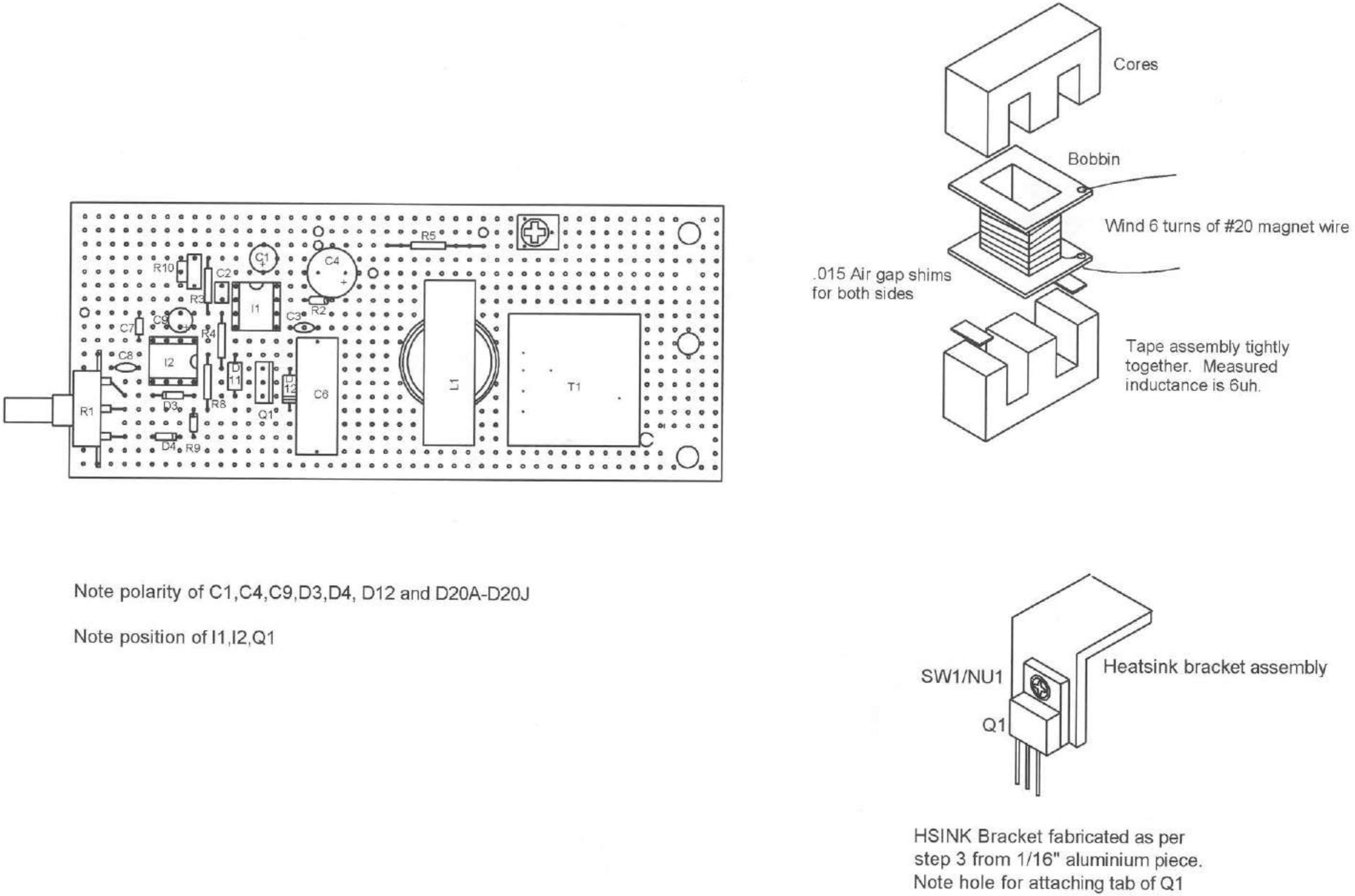
R1
R10
R3
R4
R8
R9
R2
R5
C1
C2
C3
C4
C6
C7
C8
C9
I1
I2
D3
D4
D11
D12
Q1
L1
T1
Cores
Bobbin
Wind 6 turns of #20 magnet wire
.015 Air gap shims for both sides
SW1/NU1
Q1
Heatsink bracket assembly

Note polarity of C1,C4,C9,D3,D4, D12 and D20A-D20J

Note position of I1,I2,Q1

Tape assembly tightly together. Measured inductance is 6uh.

HSINK Bracket fabricated as per step 3 from 1/16" aluminium piece. Note hole for attaching tab of Q1

Figure 7-4 *Parts assembly and fabrication*

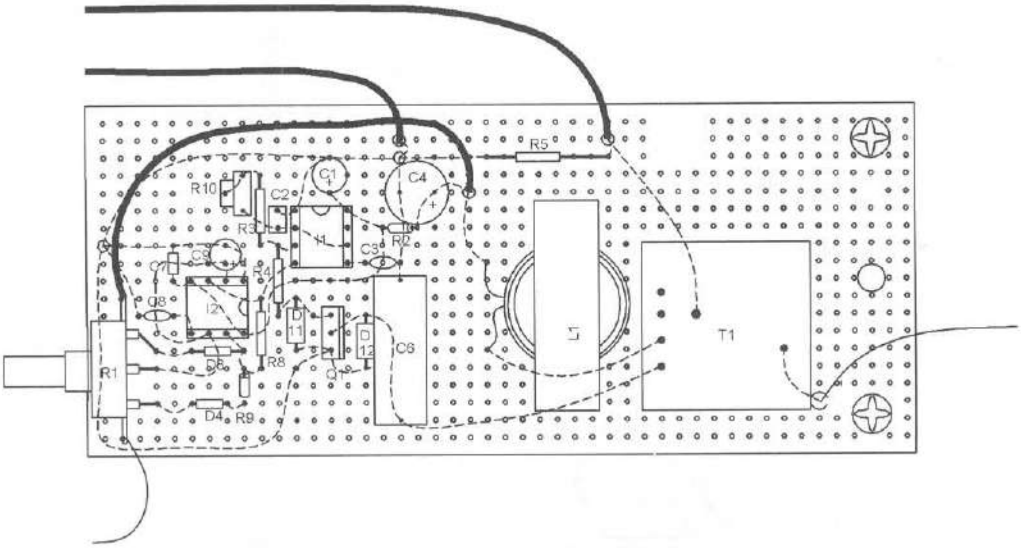

Thick dashed lines are direct connection runs beneath board of #20 buss wire (WR20BUSS) and are extended for the spark switch electrodes.

Thinner dashed lines are #24 buss wire (WR24BUSS) and component leads wherever possible.

Triangles are direct connection point junctions.

Solid black lines are external leads for input and output lines. Use red (WR20R) for +12 input. Use green (WR20G) for lifter - connection. Use black (WR20B) for com -12 input

Use 1/8-3/16 wide smooth globular solder joints for connections to C20A-J and D20A-J, R7 and HV output points. Thisis contrary to normal soldering but is necessary to prevent corona leakage.

Figure 7-5 *Wiring connections and external leads*

8. Fabricate a channel from a piece of $^1/_{16}$-inch plastic material. Place it in the assembly and secure it at the corners using silicon rubber adhesive. You may also enclose it in a suitable plastic box.

   This step may not be required when used for another project referenced in this book.

Testing Steps

This step will require basic electronic laboratory equipment including a 60 MHz oscilloscope.

1. Preset the trimpot R1 to midrange and R10 to full counterclockwise. Short out the output leads using a short clip lead.
2. Obtain a 47K, 1-watt resistor and construct the spark gap "dummy load," as shown in Figure 7-2.
3. Obtain a 12-volt DC, 3-amp power converter* or a 12-volt, 4-amp rechargeable battery.*

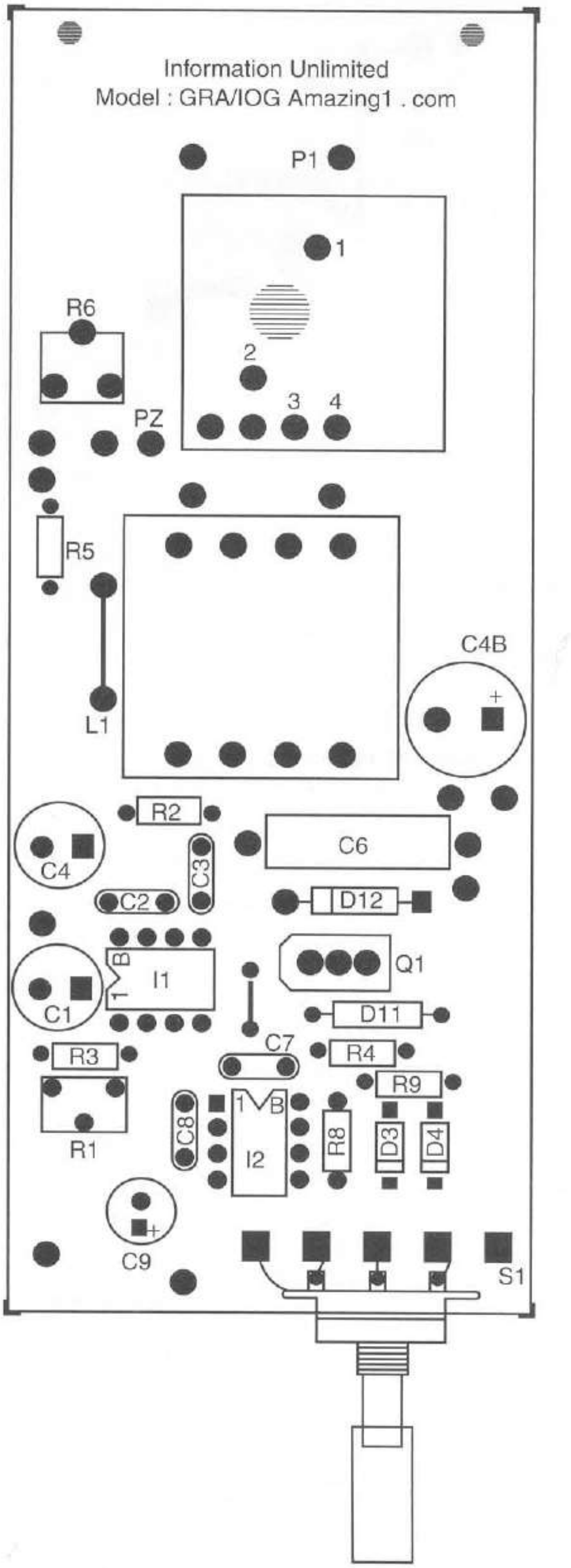

Figure 7-6 *PCB part identification*

4. Connect the input to the power converter and connect the scope, set to read 100 volts and with a sweep time of 5 usecs, to the drain of Q1.

5. Apply power and quickly adjust R1 to the waveshape shown in the inset of Figure 7-2. Note an input current dip of 100 milliamps or less. Check the heatsink of Q1, noting only warm to touch or slightly above 110 degree F.

   Note our laboratory-built units usually tune in with the pot set to 11 o'clock.

6. Remove the clip lead short in step 9 and connect the dummy load. Apply power and rotate R10 clockwise, noting the input current smoothly increasing to 1.5 amps and the spark increasing in energy. This control varies the ratio of off to on time and nicely controls the system current. Note that this system can easily provide 30 watts of usable power.

Table 7-1 High-Voltage, High-Frequency Driver Module Parts List

Ref. #	*Description*	*DB Part #*
R1	10K vertical trimpot	
R2, R4	Two 10-ohm, ¼-watt resistor (br-blk-blk)	
R3, 5, 8, 9	Four 1K, ¼-watt resistor (br-blk-red)	
R7	47K, 1-watt resistor (yel-pur-or)	
R10/S1	10K, 17 millimeter pot and switch	
C1	100-microfarad, 25-volt vertical electroradial leads	
C2	.0022-microfarad, 50-volt green plastic cap (222)	
C3,8	Two .01-microfarad, 50-volt disks (103)	
C4	1,000-microfarad, 25-volt vertical electrolytic capacitor	
C6	.22-microfarad, 250-volt metallized polypropylene	
C9	1-microfarad, 25-volt vertical electro cap	
C7	.1-microfarad, 50-volt cap, INFO#VG22	
D3, 4	Two IN914 silicon diodes	
D11	One PKE15, 15-volt transient suppressor	

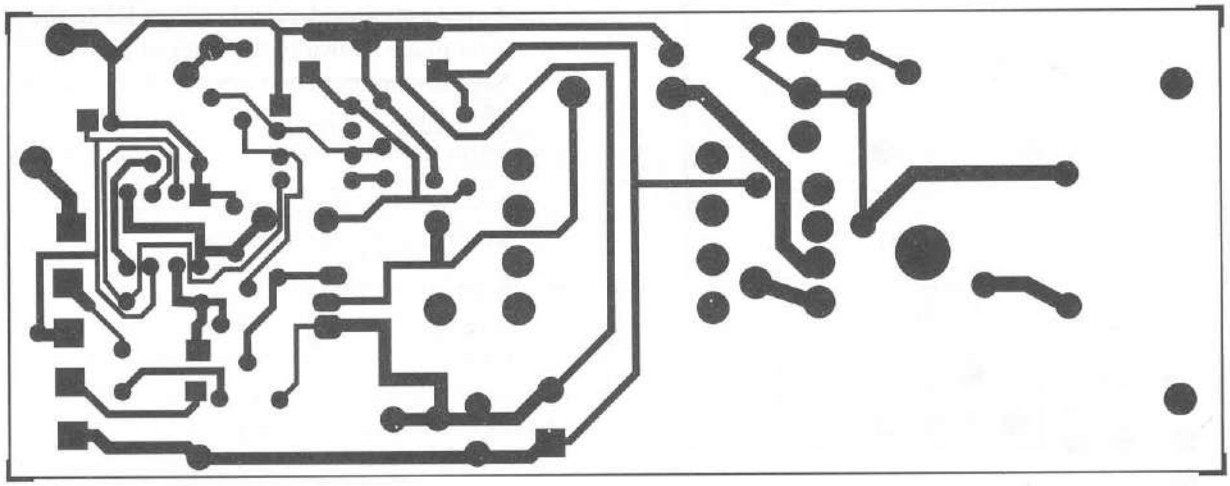

Figure 7-7 *PCB foil traces*

D12	1N4937 fast-switching 1-kilovolt diode	
Q1	IRF540 *metal-oxide-semiconductor field effect transistor* (MOSFET) transistor TO220	
I1, 2	Two 555 DIP timers	
T1	7-kilovolt, 10-milliamp mini-switching transformer	DB# IU28K089
L1	6 uH inductors (see text on assembly)	DB# IU6UH
PBOARD	5 × 2.8 × .1-inch grid perforated board; build to size per Figure 7-2	
PCGRA	Optional PCB	DB# PCGRA
WR20R	12-inch #20 vinyl red wire for positive input	
WR20B	12-inch #20 vinyl black wire for negative input	
WR20G	12-inch #20 vinyl green wire for output ground to craft return	
WR20KV	Four 20 kilovolt silicon high-voltage wires for output	
WR20BUSS	18 inches of #20 bus wire for spark gap and heavy leads (see Figure 7-5)	
WR24BUSS	12 inches of #24 bus wire for light leads (see Figure 1-5)	
HSINK	1.5 × 1-inch .063 AL plate created as per Figure 7-4	
SW1	One 6-32½ Philips screw	
NUT	Four 6-32 keep nuts	
12DC/7	115 VAC to 12 VDC, 3-amp converter	DB# 12DC/7
BAT12	12-volt, 4-amp. hour rechargeable battery	DB# BAT12
BC12K	Battery charger kit for above BAT12	DB# BC12K
GCAT1	*GENERAL CONSTRUCTION PRACTICES*	DB# GCAT1

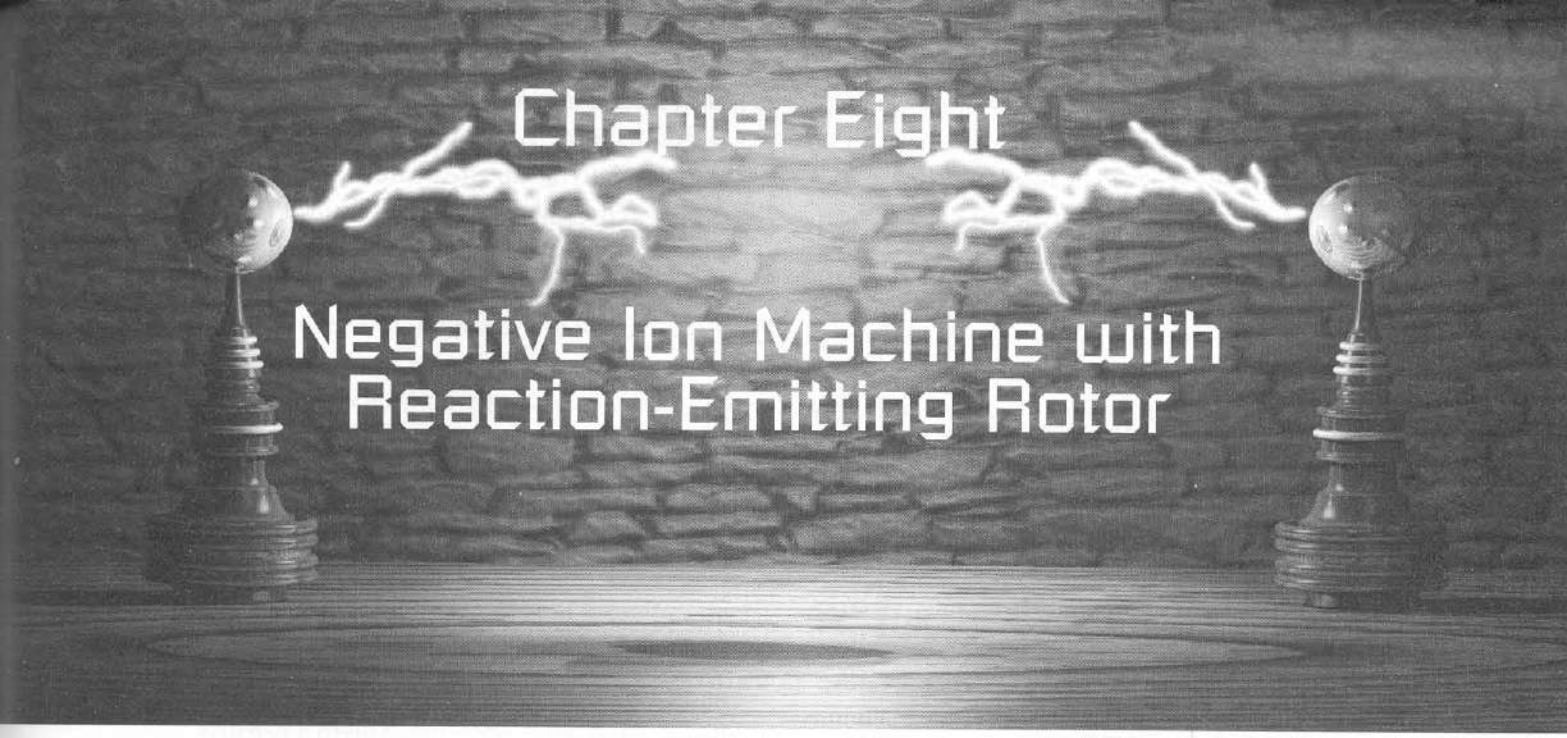

Chapter Eight

Negative Ion Machine with Reaction-Emitting Rotor

Figure 8-1 shows a new concept in negative ion generation. A spinning rotor powered by the emitted ions themselves generates copious amounts of beneficial ions. Large volumes of ions can now be produced, as the area circumscribed by the rotor greatly reduces the local charge flux. This feature allows many times more ions without ozone-producing voltages.

Negative ions are highly beneficial and cause a sense of well-being, as well as rid the air of pollen and other breathable irritants. This is an excellent device for people suffering from asthma and other bronchial problems. It is also a nice conversation piece with a desktop unit's spinning rotor silently reaching a relatively high rotational velocity. The device operates from 115 VAC/12 VDC or a battery. The project utilizes a 25-kilovolts negative output power module requiring 12 volts DC at 500 milliamps along with our proprietary rotating emitter.

This is a beginner- to intermediate-level project requiring basic electronic skills, and one should expect to spend $25 to $50. All the parts are readily available, with specialized parts obtainable from Information Unlimited (www.amazing1.com), and they are listed in Table 8-1.

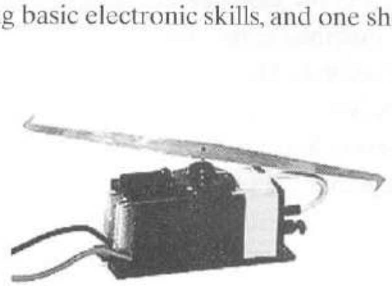

Figure 8-1 *The ion machine*

Precautions: Do not use this device in explosive or flammable atmospheres, as spark discharges may cause dangerous ignitions.

Benefits of Negative Ions

In the last several decades, a medical controversy has evolved pertaining to the beneficial effects of these minute electrical particles. As with any device that appears to affect people in a beneficial sense, there are those who sensationalize and exaggerate these claims as a cure for all aliments and ills. Such people manufacture and market these devices under these false pretenses and consequently give the products an adverse name. The Food and Drug Administration now steps in on these claims, and the product, along with its beneficial facets, goes down the tubes.

People are affected by the property of these particles, which increase the activity rate of cilia, whose function is to keep the trachea clean from foreign objects, thus enhancing oxygen intake and increasing the flow of mucus. This property neutralizes the effects of cigarette smoking, which slows down this activity of the cilia. Hay fever and bronchial asthma victims are greatly relieved by these particles, and burn and surgery patients are relieved of pain and heal faster.

Tiredness, lethargy, and general worn-out feelings are replaced by a sense of well-being and renewed energy. Negative ions destroy bacteria and purify the air with a country-air freshness. They also cheer people up by decreasing the serotonin content of the blood. As can be seen in countless articles and technical writings, negative ions are a benefit to humans and their environment.

Negative ions occur naturally from static electricity, the wind, waterfalls, crashing surf, cosmic radiation, radioactivity, and ultraviolet radiation. Positive ions are also produced from some of these phenomenon and both of them usually cancel each other out as a natural statistical occurrence. However, many manmade objects and devices have a tendency to neutralize the negative ions, thus leaving an abundance of positive ions, which create sluggishness and most of the opposite physiological effects of its negative counterpart.

One method for producing negative ions is by obtaining a radioactive source rich in beta radiation (negative electrons). Alpha and gamma emissions from this source produce positive ions that are neutralized electrically. The resulting negative ions are directed by electrostatic forces to the output exit of the device and are further dispersed by the action of a fan (this method has recently come under attack by the Bureau of Radiological Health and Welfare for the use of tritium or other radioactive salts). This approach appears to be the more hazardous of the two according to the Consumer Product Safety people. The second method is to produce a potential level of negative high-voltage electricity without ozone and allow the creation of negatively charged particles by the high voltage.

Your negative ion generator produces ions via a high-potential, low-current source of DC power. The high potential causes negative charges to be produced at the sharp emission points of the rotor blade. The reactionary force of the emitted particles now causes the rotor to spin, allowing these negative ions to escape into the air stream due to the high-charge density at the sharp, pointed rotor tips. (The point of a pin will have a much higher charge density than a larger diameter spheroid for the same potential.) Ozone is kept at a minimum by keeping the voltage relatively low while allowing a sufficient charge density lets the negative particles escape into the air.

Circuit Description

The circuit (MOD25KV) consists of a high-frequency, high-voltage oscillator being fed into a multistage Cockcroft-Walton multiplier. This energy is converted into a potential of up to a negative 25,000 volts. The high-frequency stage consists of a transistor connected as a simple oscillator where its collector drives the primary winding of the resonant transformer (T1).

The high-frequency output of T1 is fed into a voltage multiplier stage consisting of a diode string (D1 to D5) and a capacitor string (C1 to C5). A 12-volt DC wall adapter (T2) powers the unit. A separate ground is required for optimum performance and involves connecting the negative 12 volts to earth ground. This is done by connection to the AC receptacle plate or ground pin. This ground provides a virtual ground for the ions.

A charge concentration occurring at the ends of the rotor blade's points now produces ions, and reactionary forces move the rotor at a high speed. Ions are produced at a high rate as the effective emitter area is equal to that of the rotating rotor. Also, charge neutralization is minimal using this method and a high ion current is generated.

The output of the unit can be approximated by the fact that a Coulomb charge equals the current multiplied by the seconds. There are also 6.25 × E18 negative chargers per Coulomb and the maximum current to the emitters is 400 microamps. Therefore, the unit can produce 6.25 × E18 × 200 × E-6 = 2.5 × 10E15

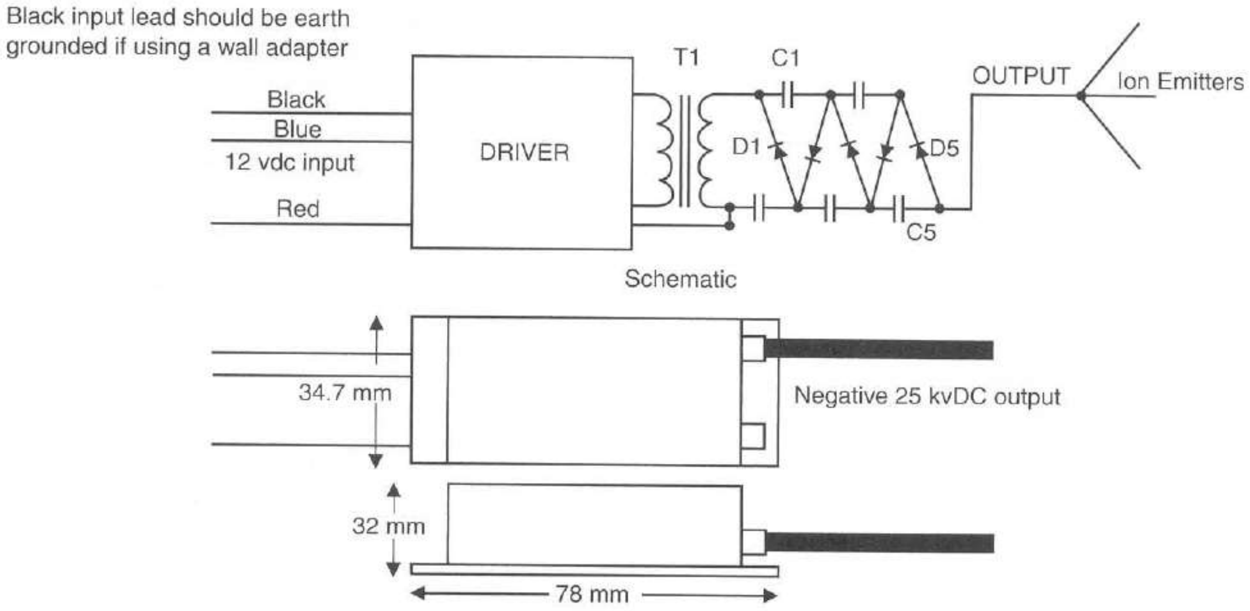
Black input lead should be earth grounded if using a wall adapter
Black
Blue
12 vdc input
Red
DRIVER
T1
C1
D1
D5
C5
OUTPUT
Ion Emitters
Schematic
34.7 mm
Negative 25 kvDC output
32 mm
78 mm

Function	Value	Notes	Other
Input Voltage	12 vdc		
Input Current (No load)			
Input Current (load)	400 ma		
Input Watt (No load)			
Input Watt (load)			
Output Volts	–20 kvDC		
Output Configuration	Single Output		
Internal Operating Frequency	35 KHz		

Use a 12 vdc/.5 amp DC wall adapter or other suitable power supply.
Always check entire system for any excessive heating.

Figure 8-2 *Circuit schematic*

negative charges per second. This is usually sufficient for a large room.

Assembly Steps

1. Lay out and identify all parts and pieces. Verify them with the parts list.
2. Carefully fabricate the rotor blade as shown in Figure 8-3 and note that it must be balanced for optimum rotational velocity and ion emission.
3. Position a common plastic pushpin (PIN1) and hot glue or epoxy to the module as shown. Glue is more forgiving if should you ever wish to replace the pin.
4. Carefully strip about $^3/_4$ of an inch of insulation from the output lead. Tin and solder it to the bottom of the actual pin. This connection point must allow complete clearance of the entire bearing section of the rotor assembly when in place.
5. Wire up the wall adapter (T2) and ground the lead using (WN1) wire nuts, as shown in Figure 8-4.

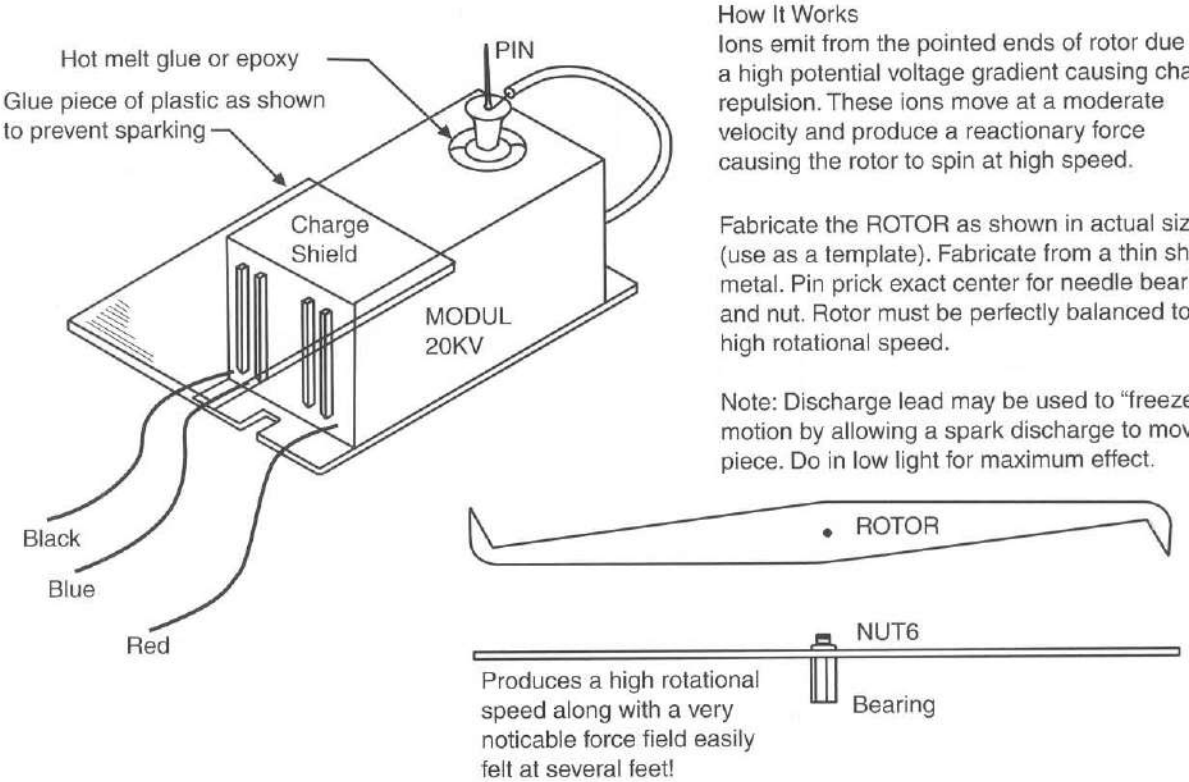
Hot melt glue or epoxy
Glue piece of plastic as shown to prevent sparking
PIN
Charge Shield
MODUL 20KV
Black
Blue
Red
ROTOR
NUT6
Bearing
Produces a high rotational speed along with a very noticable force field easily felt at several feet!

How It Works
Ions emit from the pointed ends of rotor due to a high potential voltage gradient causing charge repulsion. These ions move at a moderate velocity and produce a reactionary force causing the rotor to spin at high speed.

Fabricate the ROTOR as shown in actual size (use as a template). Fabricate from a thin sheet metal. Pin prick exact center for needle bearing stud and nut. Rotor must be perfectly balanced to reach high rotational speed.

Note: Discharge lead may be used to "freeze" rotor motion by allowing a spark discharge to moving piece. Do in low light for maximum effect.

Figure 8-3 *Isometric view of the overall unit*

6. Glue on the CS1 charge shield plate to prevent premature sparking to the metal heatsink of the module. Also, stick on the rubber feet.
7. The unit is perfectly usable. However, you may want to enclose it in some sort of plastic box. Always use good ventilation and make sure the box is large enough or about 1½ times the diameter of the rotor.
8. Find a suitable location, such as a nonmetallic table, to place the unit on. Place the rotor blade on the pin and verify it is unobstructed. Then connect the grounding lead to the receptacle mounting screw and plug in the T2 adapter. The rotor should begin rotating and quickly reach a high speed.

Notes

Ions are now being emitted into the air and can easily be detected by holding a fluorescent tube or a small neon indicating lamp (LA1) near the unit, noting a flashing. This must be done in the dark. You can also place a piece of paper near the base of the unit and note the force field generated that tries to pull the paper.

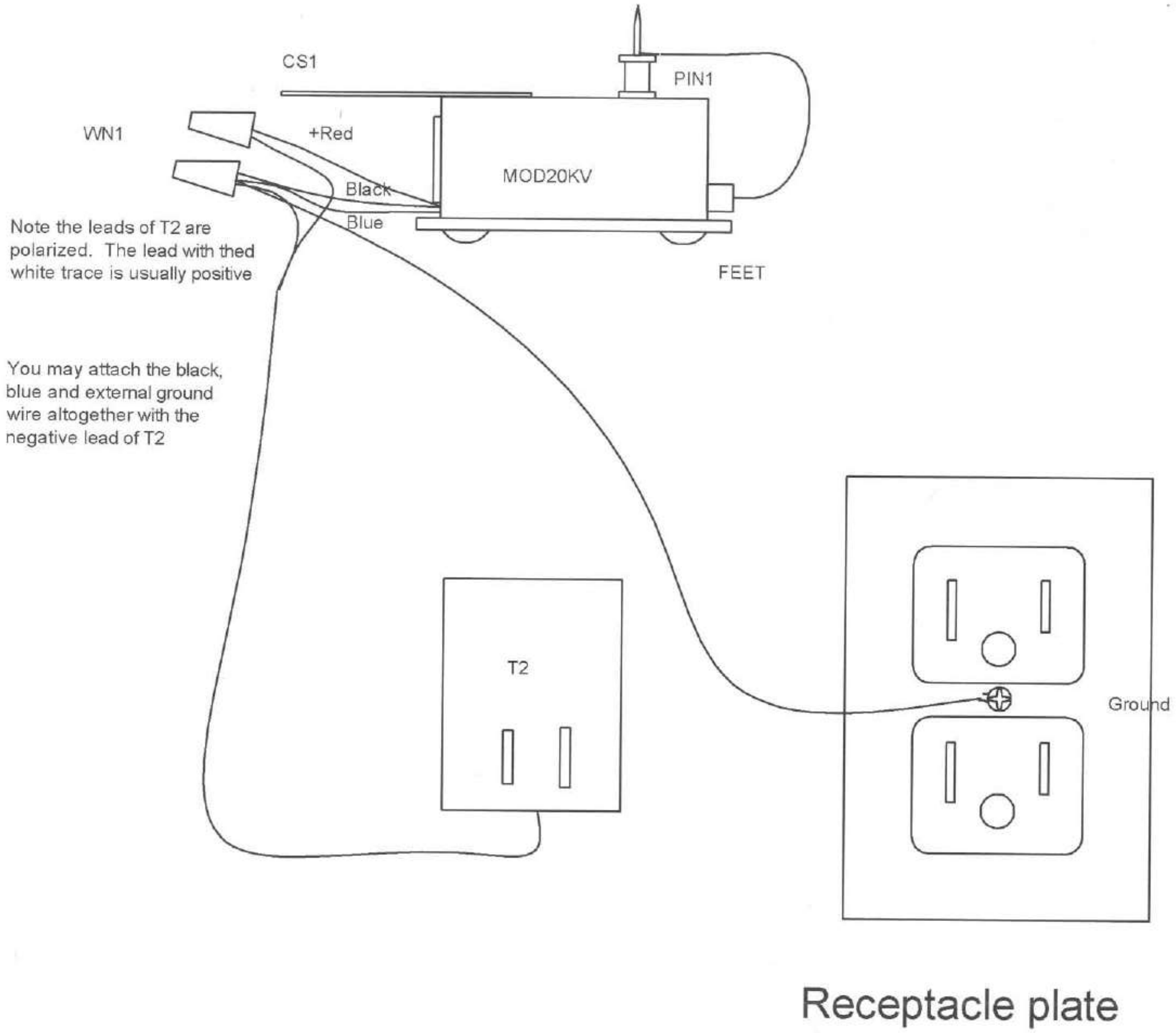
CS1
PIN1
WN1
+Red
MOD20KV
Black
Blue
FEET
T2
Ground
Receptacle plate

Note the leads of T2 are polarized. The lead with thed white trace is usually positive

You may attach the black, blue and external ground wire altogether with the negative lead of T2

Figure 8-4 *Side view showing power connections*

Table 8-1 Negative Ion Machine Parts List

Ref. #	*Description*	*DB Part #*
MODULE	Negative 20-kilovolt module reworked	DB# MOD20KV
T2	12-volt DC, 500-milliamp wall adapter	DB# 12DC/.5
ROTOR	Metal fabricated rotor	DB# ROTOR
PIN	Common plastic pushpin	
BEARING	#6-32 threaded standoff	
NUT6	6-32 hex nut	
CS1	2¼ × 1½-inch-thin, 1/16 to 1/8-inch-thick plastic	
WN1	Small wire nuts	
FEET	Four stick-on rubber feet	

Chapter Nine

Kirlian Imaging Project

Kirlian photography uses *cold electron emission* to expose a photographic film. Cold electron emission is the result of electrons accelerating in an electric field. These accelerated electrons ionize the air and recombine producing ultraviolet light that is mostly invisible to the naked eye. This phenomenon is commonly known as *corona*. Many living objects are claimed to emit this radiation that will vary according to the health and condition of the object being exposed. This in itself is now subject to scientific research in material analysis. A more "psychic application" can be seen in the practice of radiation being referred to as an aura and being dependent on other more metaphysical functions. This project describes a simple method of producing Kirlian images of small objects and is an introduction to those who want to explore this interesting field. Kirlian imaging ranges from the psychic to materials analysis. Often it is used for experiments in which it is claimed that objects all possess a unique aura signature. This depends on the subject's mental makeup and condition at the time. Basically, it is a form of corona discharge that occurs as a precursor to a dielectric breakdown by most conductive objects when placed in a high-frequency, high-voltage field. Corona is a form of electrical leakage occurring without actual air breakdown such as a spark discharge. A corona discharge possesses ultraviolet emissions that easily develop film and thermographic paper, producing an image print.

This is an intermediate-level project requiring basic electronic skills and one can expect to spend \$25 to \$50. All parts are readily available, with specialized parts obtainable through Information Unlimited (www.amazing1.com). The parts are listed in Table 9-1.

The system is simple and built from readily available parts (see Figure 9-1). The builder may deviate from what is shown as long as the basics are adhered to. The system requires a high-voltage, high-frequency module, as shown in Chapter 10 where it will be used to supply the burning plasma for the plasma pen project.

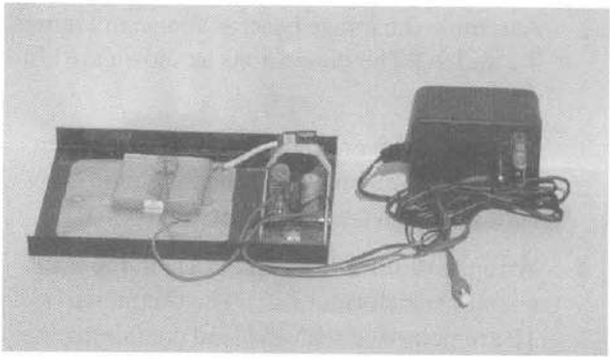

Figure 9-1 *A complete unit*

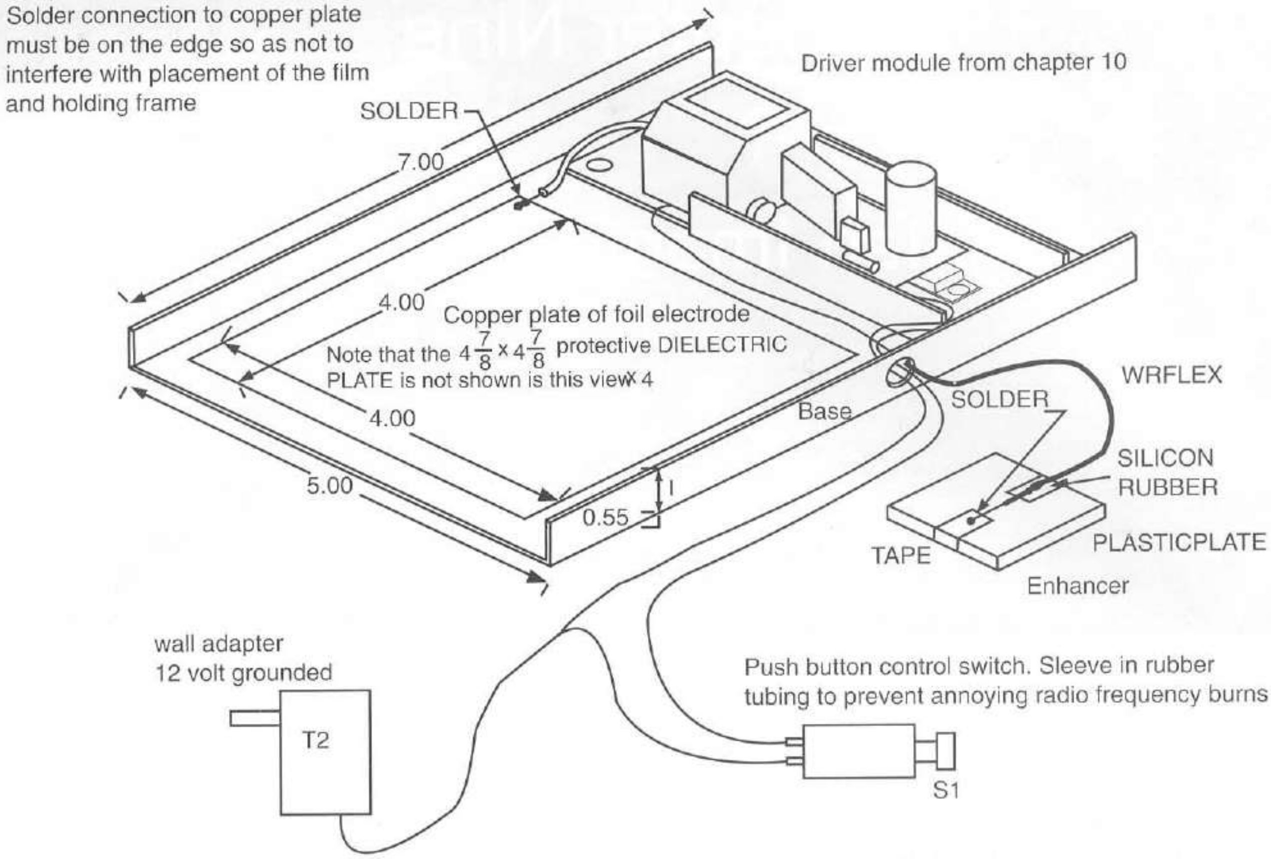
Solder connection to copper plate must be on the edge so as not to interfere with placement of the film and holding frame
SOLDER
7.00
Driver module from chapter 10
4.00
Copper plate of foil electrode
Note that the $4\frac{7}{8} \times 4\frac{7}{8}$ protective DIELECTRIC PLATE is not shown is this view× 4
4.00
5.00
0.55
Base
SOLDER
WRFLEX
SILICON RUBBER
PLASTICPLATE
TAPE
Enhancer
wall adapter
12 volt grounded
T2
Push button control switch. Sleeve in rubber tubing to prevent annoying radio frequency burns
S1

Figure 9-2 *Isometric of the image base and charge plate*

Assembly Steps

If you are a beginner, it is suggested to obtain our *GCAT1 General Construction Practices and Techniques*. This informative literature explains basic practices that are necessary in proper construction of electromechanical kits and is listed in Table 9-1.

1. Assemble the driver module, as shown in Figures 10-2 and 10-3 in Chapter 10. Verify proper operation before using.
2. Assemble the image base, as shown in Figures 9-2 and 9-3. The dimensions as shown are not critical.
3. Mount the driver module and solder the output lead as shown. Route the input and high-voltage return lead to the ENHANCER.
4. Wire the pushbutton switch (S1) to the wall adapter transformer (T2). The output wires of T2 are polarized with the lead containing the white trace, usually the positive. It is a good idea to cover the entire body of S1 except the plastic actuator; any contact to metal will result in small, very annoying burns. The wall adapter recommended has a third pin being an isolated ground for the high-frequency return.

Testing Steps for the Circuit

These steps usually will not require any special testing equipment.

1. Power up the unit and activate S1. Place your finger on the insulated plate and note a bluish corona discharge occurring under your finger. You may feel some heat or a tingling sensation. Try this for only up to 20 seconds.
2. Obtain some thermographic fax paper and place it on the dielectric plate. Place your finger on the paper and energize the circuit for 10 to 20 seconds. You will note an image on the paper.

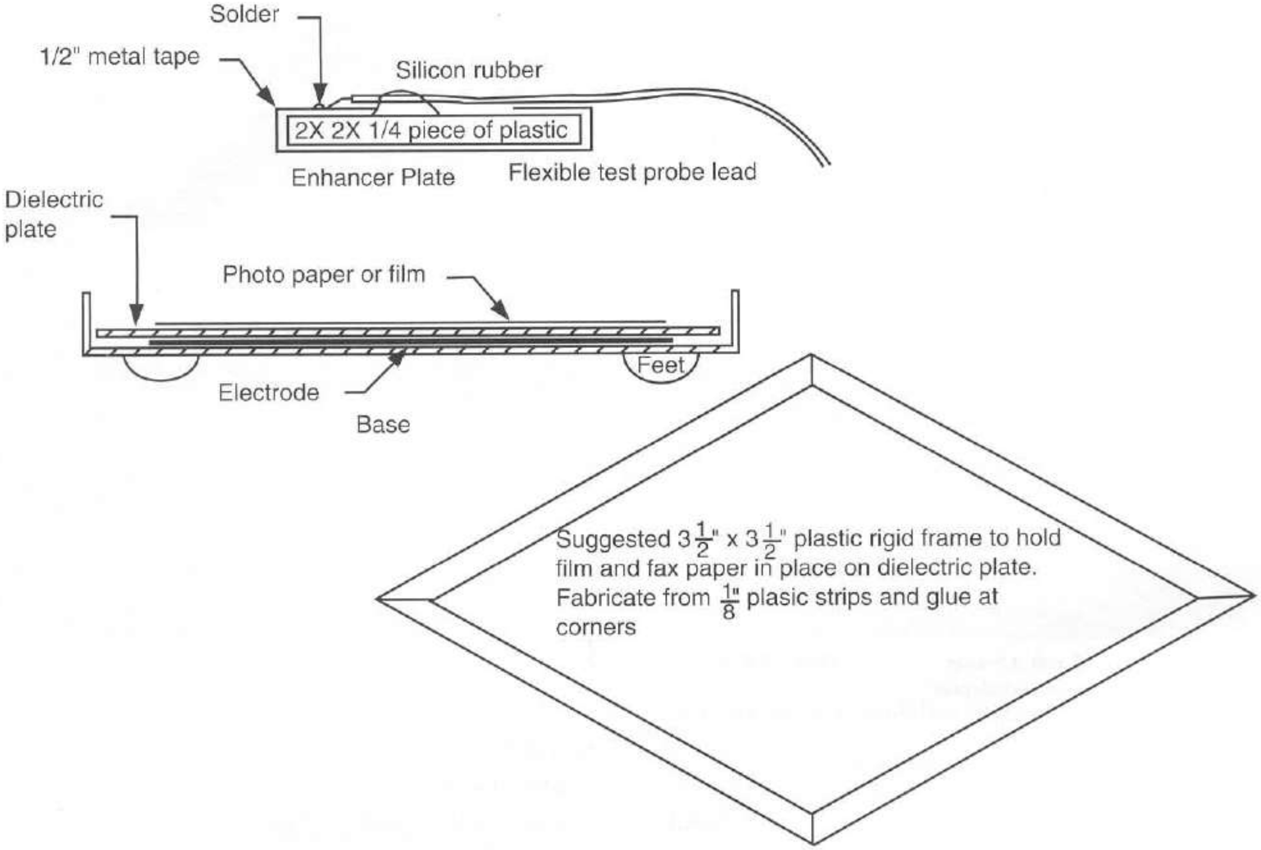

Suggested $3\frac{1}{2}$" x $3\frac{1}{2}$" plastic rigid frame to hold film and fax paper in place on dielectric plate. Fabricate from $\frac{1}{8}$" plasic strips and glue at corners

Figure 9-3 *Image base and charge plate*

Try it again with a metal object such as a coin on the dielectric plate and place the enhancer on the object. Push the button for 5 to 10 seconds and note an image forming.

3. You can upgrade your image to an actual photograph by using actual camera film. The old Polaroid 600 or other similar instant film will work. You must work entirely in the dark to avoid film exposure to any ambient light. The objective is to place the film on the dielectric plate along with the object to be photographed and the enhancer placed on top.

Notes

You will note that a few variables here can affect performance: the output voltage of the driver module and the dielectric plate's thickness and dielectric properties. The dielectric plate is important because it must form a high enough capacity between the object and the contacts to allow an adequate corona current to flow to create an image. The thinner the plastic dielectric plate, the more corona current. However, you must not allow a breakdown to occur by using material that is too thin, as this can produce a painful burn.

Many different sources of information, including some excellent books, are available on the applications, methods, and uses of Kirlian photography. You are encouraged to continue to investigate this unusual electronic phenomenon.

Table 9-1 Kirlian Imaging Parts List

Ref. #	*Description*	*DB Part #*
DRIVER	Driver module from Chapter 10	DB# MINIMAX7
BASE	6 × 7-inch plastic piece, formed as shown in Figure 9-2	
COPPER ELECTRODE	4 × 4-inch, thin piece of copper	
DIELECTRIC PLATE	5 × 5 × .02-inch piece of polypropylene sheet	DB# POLYP20
PLASTIC PLATE	2 × 2 × 1/4-inch plastic plate	
TAPE	Four inches of 1/2-inch solderable metal sticky tape	
FEET	Four sticky rubber feet	
T2	12-volt, 1.5-amp groundedadapter with third pin	DB# 12DC/1.5G
FRAME PIECES	Four strips of 1/8-inch plastic to fabricate the frame as shown in Figure 9-3	
WRFLEX	12 inches of flexible test lead wire	
S1	Push-button switch reworked as shown to prevent burns	
GCAT1	General construction practices	DB# GCAT1

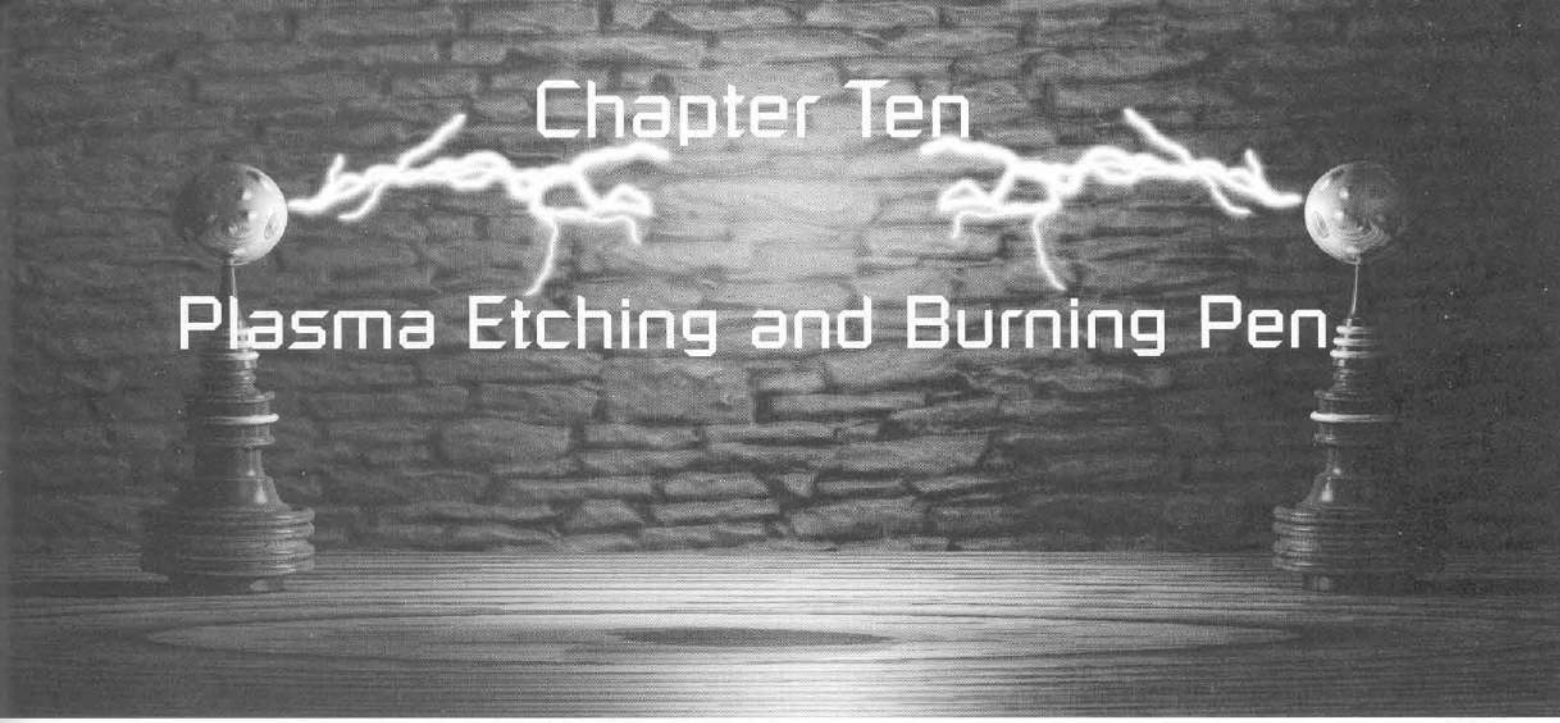

Chapter Ten

Plasma Etching and Burning Pen

A plasma pen uses high frequency, high voltage to burn or etch permanent markings into many different materials (see Figure 10-1). With a little practice, the user can create many intriguing designs of fine, lace-like etches. Names and other text can easily be etched into almost any organic material. Wood, paper, plastics, pumpkins, gourds, eggs, and leaves can all be marked with fine and fancy designs or inscriptions. The pen's other applications include Kirlian photography, aura enhancing, and magically lighting household fluorescent and neon tubes, all without any connecting wires, providing a great magic act. You can also use this for plasma experimentation or even removing warts.

This is an intermediate-level project requiring basic electronic skills. Expect to spend $25 to $35. All parts are readily available, with specialized parts obtainable through Information Unlimited (www.amazing1.com), and all of them are listed in Table 10-1.

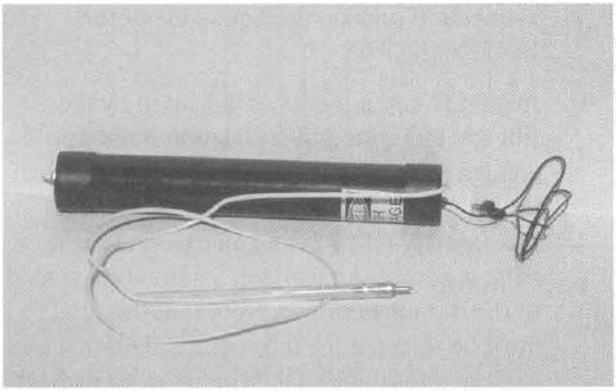

Figure 10-1 *Photo of assembly*

The system is shown in two parts: the power driver and the pen or etching section. These are connected together via a suitable, flexible, high-voltage wire. The power driver consists of the electronics module and the internal batteries for portable use. A jack is provided for external use from a 12 VDC wall adapter or another suitable source. This section is housed in a tubular enclosure with plastic end caps as shown.

Driver Circuit Description

The circuit is a frequency, high-voltage oscillator that consists of transistor Q1 connected as a simple oscillator where its collector drives the primary winding PR1 of the resonant transformer T1. Feedback is obtained via a second winding (FB) and fed to the base of Q1 through a current-limiting resistor (R1). Resistor R2 biases Q1 into conduction and initiates the oscillation (see Figure 10-2).

Capacitor C3 speeds up the turn-off time of Q1, while resistor R3 and capacitor C5 provide a filter to prevent oscillation at the self-resonant frequency of T1. Resonating capacitor C4 resonates the transformer

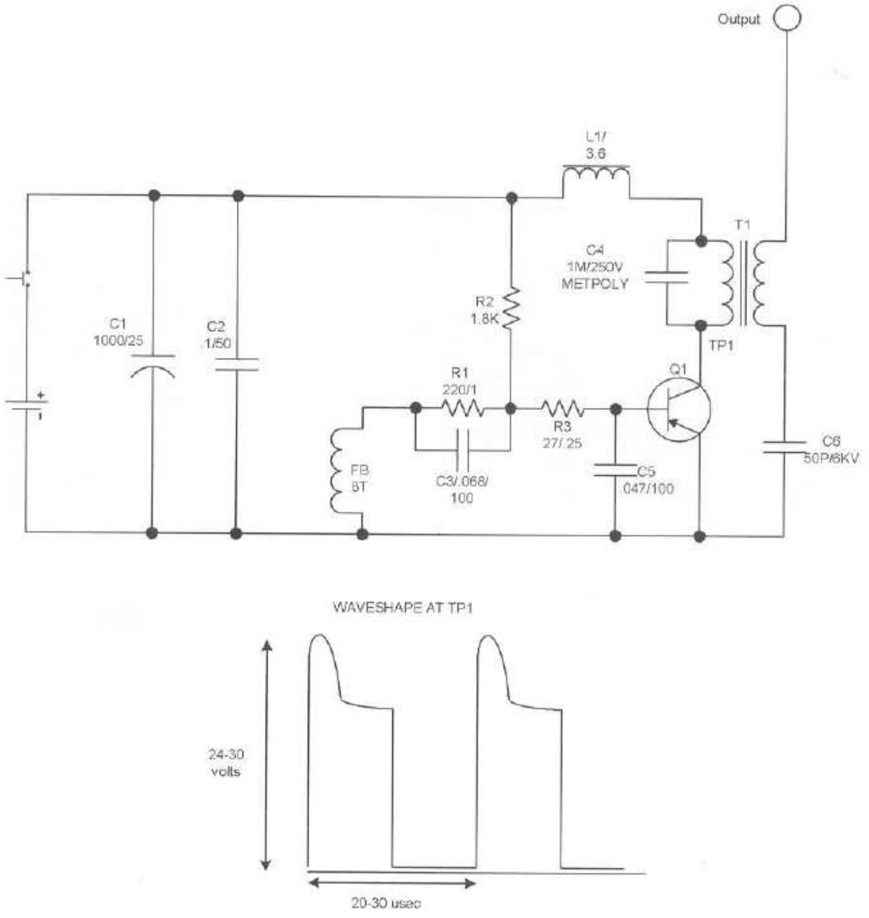

Figure 10-2 *Circuit schematic*

to around 35 KHz. The current-limiting inductor L1 limits the short-circuit current to a noncatastrophic value, and capacitors C1 and C2 bypass any signal to ground.

Driver Circuit Assembly

1. Lay out and identify all the parts and pieces. Verify them with the parts list, and separate the resistors as they have a color code to determine their value. Colors are noted on the parts list.
2. Obtain the available *printed circuit board* (PCB), as shown in Figure 10-3, or fabricate a piece of perforated circuit board to the PC as laid out on the PCB. Note the size of the PCB is 3½ × 1⅜ inches and contains the silk screening that shows the positioning of the mounted parts.
3. If you are building from a perforated circuit board, it is suggested that you insert the components starting in the lower left-hand corner. Pay attention to the polarity of the capacitors that have polarity signs and all semiconductors. Route the leads of the components as shown and solder as you go, cutting away unused wires. Attempt to use certain leads as the wire runs and follow the foil traces on the drawing as these indicate the connection runs on the underside of the assembly board.
4. Insert the components as indicated by the silk-screen printed identification numbers and compare this with the bill of materials. Attach three 6-inch, #22-20 leads as shown for the input power (P1, P2) and an external high-voltage return. Also attach a high-voltage lead to the output as shown. Note that this lead must be selected for the required length. Use a silicon, 20-kilovolt DC wire or an equivalent.

Note the ground lead that connects to the metal frame under the screw head as shown.

5. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.
6. Cut the metal bracket (BRK1) from a piece of .062 sheet aluminum and make the final assembly, as shown in Figure 10-4. Note that the metal tab of Q1 must be insulated from the bracket as shown on the mounting scheme. This piece is also the heatsink for Q1.

Driver Board Testing

This step will require basic electronic laboratory equipment. A 60 MHz oscilloscope is suggested.

1. Obtain a 12-volt DC source, preferably with a volt and ammeter. Leave the output leads as open circuit and apply power. Note a current draw of less than 250 milliamps. If you have a scope, observe the waveshape at the collector of Q1, as shown in Figure 10-2.
2. Contact the high-voltage output lead to the bracket and note that you can draw a 1/2-inch arc with the input current going to 1 amp. This completes this module.

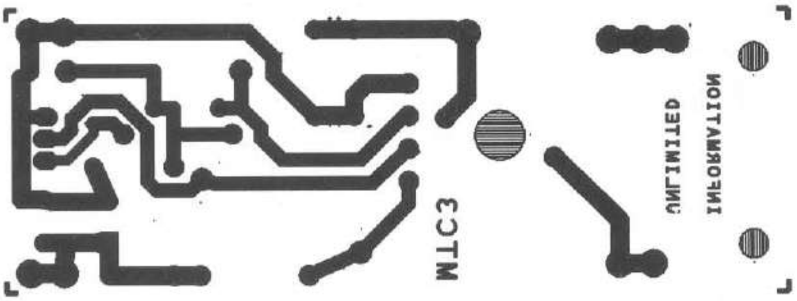
MTC3
INFORMATION
UNLIMITED

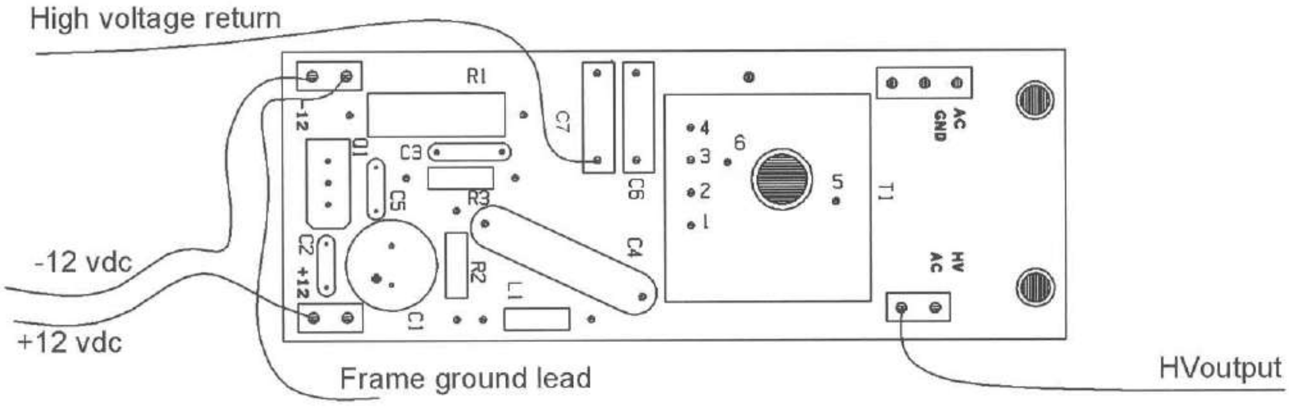
High voltage return
R1
C7
C6
-12
Q1
C3
C5
R3
C4
C2
+12
R2
L1
C1
4
3
2
1
6
5
T1
AC GND
HV AC
-12 vdc
+12 vdc
Frame ground lead
HVoutput

Figure 10-3 *Assembly board*

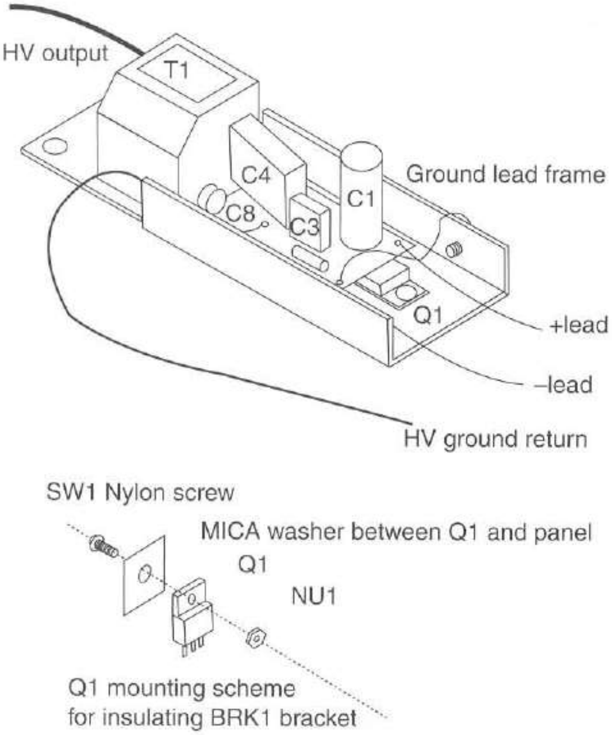
HV output
T1
C4
C8
C3
C1
Ground lead frame
Q1
+lead
–lead
HV ground return
SW1 Nylon screw
MICA washer between Q1 and panel
Q1
NU1
Q1 mounting scheme
for insulating BRK1 bracket

Figure 10-4 *Isometric of assembly*

Mechanical Assembly

You will need some basic cutting tools for the following steps.

1. Fabricate TUB2, as shown in Figure 10-5, and then make the caps CA1 and CA2 for passage of the output leads and for switch S1, jack J1, and indicator lamp LA1.
2. Insert the components and wire them up as shown. Note the current-limiting resistor R10. The exiting leads should be tied in a knot to relieve any strain.
3. Assemble the pen section as shown and solder the lead to the brass POINT. Shim the lead and POINT, and mechanically secure them in place.

Sample Demonstration

1. Obtain a 4 × 6-inch square piece of 1/4-inch masonite plywood, as shown in Figure 10-6. Moisten it with a sponge and place it on top of an approximately equal-size metal plate attached to the ground lead.
2. Contact the surface with the plasma pen and note the burning action. Experiment for the best results with other materials.

Note that your pen will light fluorescent lamps in your house just from its proximity to them, and it can also be used for Kirlian photography.

Switch S1 selects either internal batteries or external 12 volts, jack J1 is for the external 12 volts. LED indicates when the system is powered up. Resistor R1 limits the LED current. These components are mounted on the rear cap CAP2.

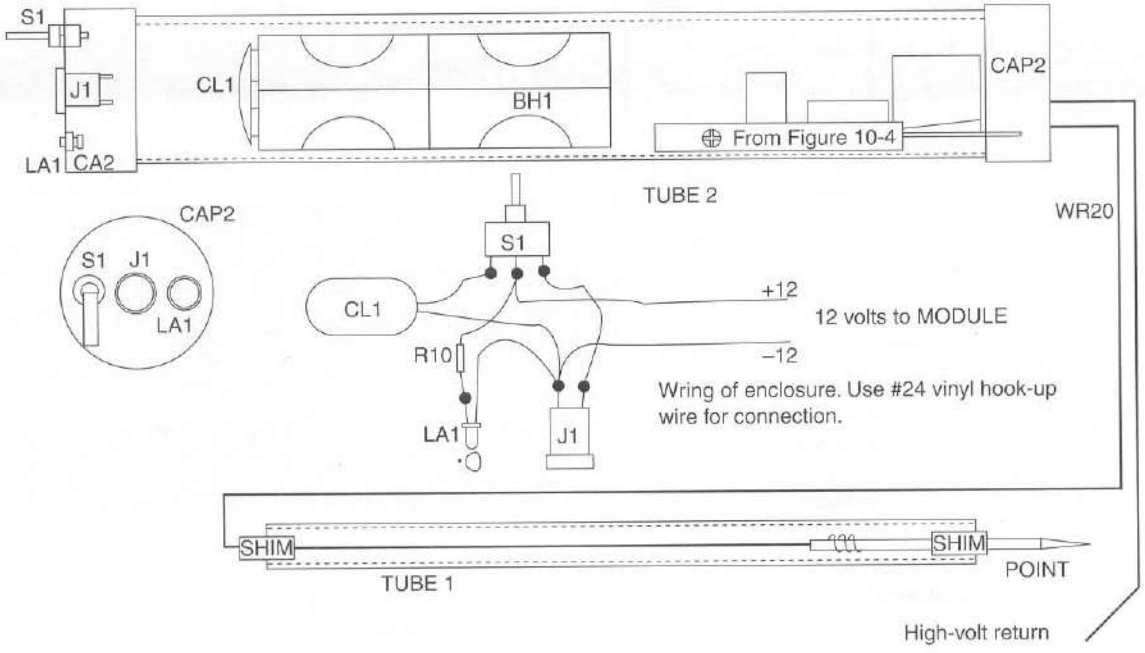
S1
J1
LA1
CA2
CL1
BH1
CAP2
From Figure 10-4
TUBE 2
WR20
CAP2
S1
J1
LA1
S1
CL1
+12
12 volts to MODULE
R10
−12
LA1
J1
SHIM
SHIM
TUBE 1
POINT
High-volt return

Wring of enclosure. Use #24 vinyl hook-up wire for connection.

The POINT can be a 1/5 to 1/4 brass rod that can be soldered to. The SHIMS can be rubber tubing that strain relieve and position the feed lead and burning point. TUBE1 can be a 3/8" diameter plastic tube.

Figure 10-5 *Final enclosure and wiring*

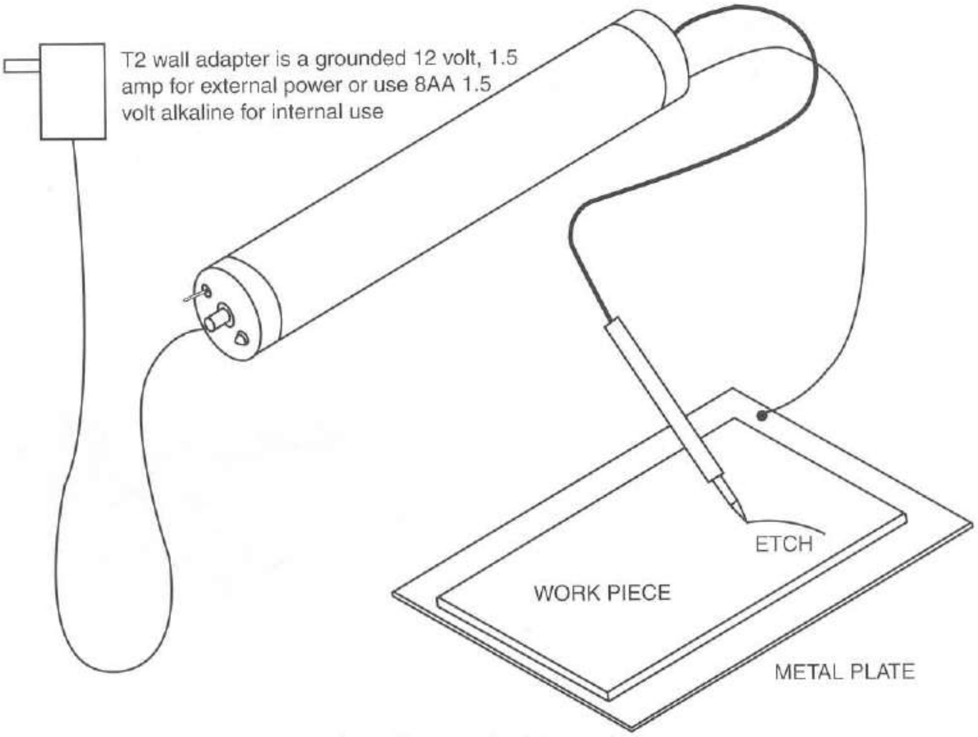
ETCH
WORK PIECE
METAL PLATE

T2 wall adapter is a grounded 12 volt, 1.5 amp for external power or use 8AA 1.5 volt alkaline for internal use

For optimum performance, use a piece of mansonite, thin wood, etc.
Place on top of metal plate connected to the high-voltage ground lead.
You may moisten the beginning area to start the etch.

Figure 10-6 *Sample demonstration*

Table 10-1 Plasma Etching and Burning Pen Parts List

Ref. #	*Description*	*DB Part #*
R1	220 ohms, 1 watt (red-red-br)	
R2	1.8K, 1/4-watt (br-gray-red)	
R3	27 ohms, 1/4-watt (red-pur-blk)	
C1	1,000-microfarad, 25-volt vertical electrolytic capacitor	
C2	.1-microfarad/50-volt small, plastic, capacitor	
C3	.068-microfarad, 50-volt plastic capacitor (683)	
C4	1-microfarad, 250-volt metallized foil large, blue, plastic capacitor	
C5	.047-microfarad, 100-volt plastic capacitor (473)	
C6, 7	Two 25-picofarad, 6-kilovolt ceramic capacitors or a single 50-picofarad, 6-kilovolt unit	
Q1	MJE3055T TO220 power tab NPN transistor	
T1	Special high-frequency transformer	DB# 28K089
WR20	24-inch #22-20 vinyl stranded hookup wire	
PCMTC	MTC3 printed circuit board or use a 3 1/2 × 1 1/2 × .1-inch grid perf board	DB# PCMTC
BRKT1	Create as shown in Figure 10-4	
TUBE2	12 × 2-inch ID schedule, 40 PVC tubing	
TUBE1	3/8 × 1/4-inch ID length of plastic tubing	
CA1, 2	Two mating flat-end caps for TUBE1	
POINT	1/8 × 3-inch brass rod filed to shape	
LA1	*Light-emitting diode* (LED), any color	
R10	220 1/4-watt resistor (red-red-br)	
CL1	Battery snap clip	
BH8	Eight-AA-cell battery holder	
S1	*Single pole, double throw* (SPDT) small toggle switch	
J1	Chassis-mounting DC power jack	
WR24	48-inch, #24 vinyl hookup wire	
WR20	24-inch, 20-kilovolt, silicon, high-voltage wire	
NEONX21	Assembled high-voltage module (see Figure 10-3)	DB# NEON21

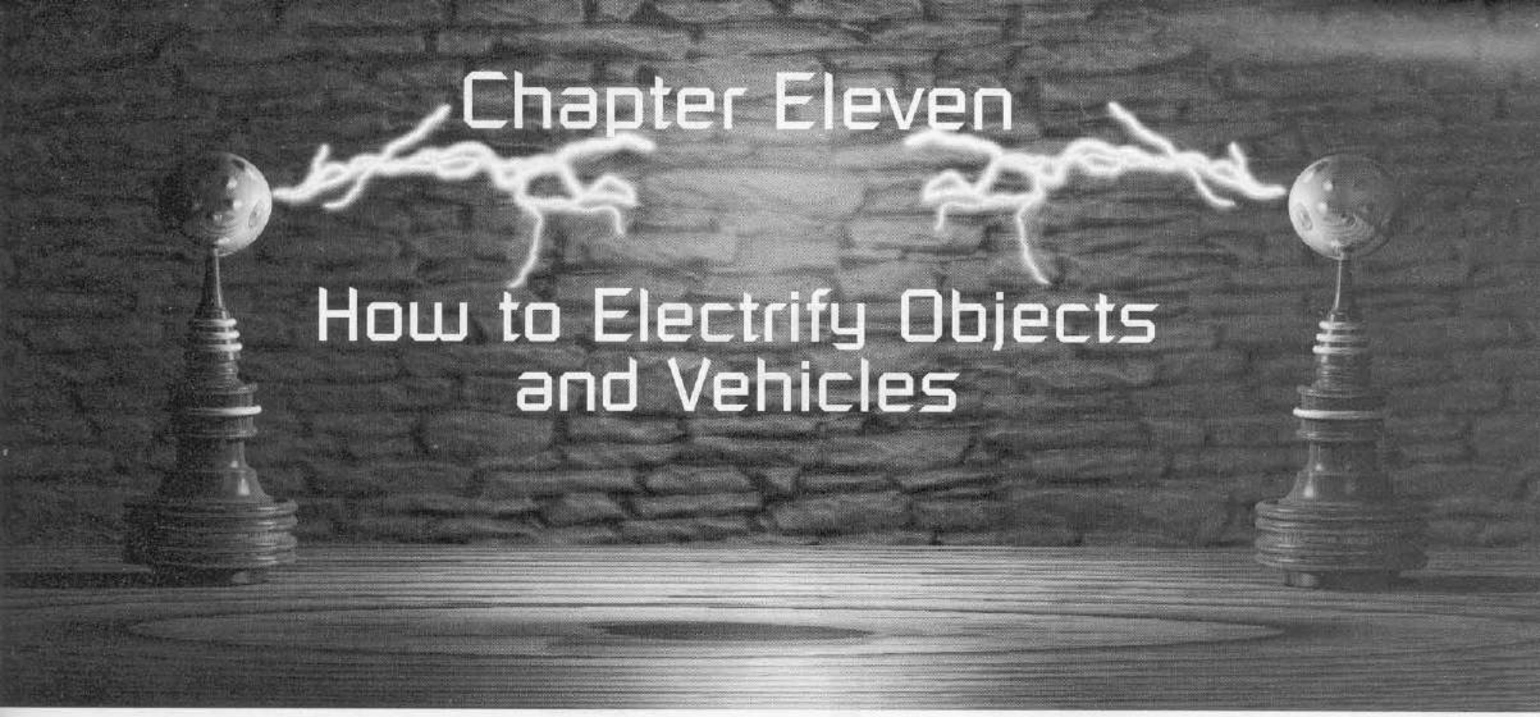

Chapter Eleven

How to Electrify Objects and Vehicles

Figure 11-1 shows two board-level devices capable of producing electrical shocks. The larger circuit generates low-energy, 20,000-volt pulses similar to an automotive ignition system. Although these pulses are nonlethal, they can still be very painful when contacted. The smaller circuit generates a steady 5-milliamp, low current at 2,000 volts. By itself, it is not lethal but again can produce a painful shock. If the output is allowed to charge a capacitor to a sufficient energy level, then the rules change and a potentially lethal situation exists. This level usually is any value over 25 joules as computed from W (joules) = half of C (capacitance in farads) $\times V^2$ (in volts) that the capacitor is allowed to charge too.

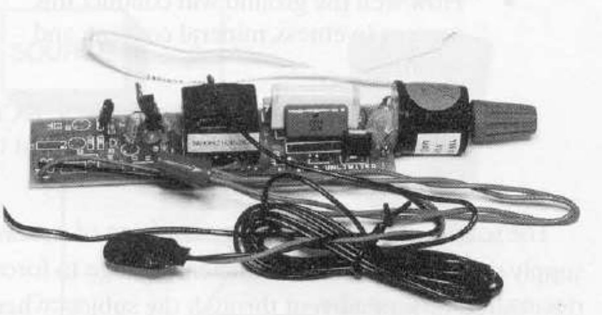

Figure 11-1 *Photograph of devices*

Caution: The use of electrification to prevent human intrusion is illegal. This application is legally referred to as "man trapping." Consult legal advice before using this if human contact is intended!

Warning: Never expose any person wearing a pacemaker or having a known physical condition to any form of electrical shock.

Data shows how to use several shocking systems with different types of outputs and applications. Even though the energy levels may be well below dangerous levels, secondary reactions can cause severe injury. Use caution if intended for human contact, and never connect the DC output version to a capacitor as a lethal charge may exist.

The following information is intended to serve as a guide to those contemplating using our equipment or other devices capable of producing a mild, severe, or possibly lethal electric shock. Any device of this nature, if used against people, may be considered a potential weapon. Remember that what may be a mild shock to one person could be harmful to another. Secondary effects, resulting from involuntary action, must also be considered.

The devices sold are intended for laboratory experiments or for use against animals. They should only be used for intimidation purposes to prevent an encounter. Human contact must be totally justified, as serious consequences can result from the illegality of their use.

Even though our electronic devices provide an average current below what is considered lethal to the average person, circumstances such as a person's heart condition or other physical parameters can influence both the immediate effect and after contact reaction.

Table 11-1 shows certain effects from various electrical currents. Remember that the indicated current must have an appropriate voltage behind it to flow through a given resistance, such as the human body. When working with sources that have storage capacitors (especially dangerous sources), one must take into account the high peak current that can result for a period of time and result in possible electrocution if this duration of time is sufficient.

Table 11-1 Electrical current scale

Voltage	*Degree of pain*	*Examples*
.01 to 1 ma	Tingling to annoying	Static electric shocks
1 to 5 ma	Annoying to painful	Spark plugs, TV picture tubes, trip current of GFI
5 to 20 ma	Painful to very painful	Oil burner ignition, bug killers, stun guns
20 to 50 ma	Very painful to possibly lethal	Neon sign transformer, old tube radios
50 to 100 ma	Possibly lethal to lethal	Low-powered transmitters, capacitor charges for lasers
100 to 500 ma	Lethal to deadly	Medium-powered transmitters, laboratory power supplies
500 ma to 1amp (+)	Deadly, with usually no second chance	Electric chair, 220-volt house current under certain circumstances

Supplementary Electrification Data and Information

The following information explains how to place an electrical charge on an object to protect against animal damage, intrusion, or theft, and the construction of several energy sources is referenced. These must be treated with caution and never be used on human beings without proper discretion due to potential injury or electrocution. Accidental contact must also be avoided. Many laws unfortunately are made to protect the criminal, and therefore any attempt made to protect oneself or property using these devices could result in a suit by the criminal if he or she can prove injury.

The methods for electrifying objects and areas will be shown as several examples that the user may adapt to his or her own needs. To successfully electrify, an object requires the following conditions to exist. Note that the following pertains to flowing electric current and not to static or super-high voltages. Figure 11-2 provides the basic explanation.

- The object of contact must be of a conductive nature such as metal, conductive paint, or wet ground.
- The subject (the intruder or animal) must supply the link to complete the circuit; that is, his or her physical extremities must come in contact with the object, along with a second common surface such as the ground, water, or another object anticipating simultaneous contact. No current flows unless contact is made to complete the circuit by the subject physically touching the object.
- The electric current flows from the +lead of SOURCE to the −lead. In doing so, it flows from the object to the point of contact by the subject to the ground and back to SOURCE via the −lead. This produces a current flow through the subject, developing an electrical shock. The severity of this shock is dependent on:
    - How well the ground will conduct this current (wetness, mineral content, and so on)
    - The electrical parameters of the SOURCE, that is, the voltage and available current to support this voltage

The source of electrical energy is a form of a power supply made to generate a sufficient voltage to force the desired flow of current through the subject when contact is made. This can take on many different forms depending on the applications and the desired results. If a system was built to kill a human being, a power supply with a current supply capability of 1/2 to 20 amps would be necessary to support a voltage high

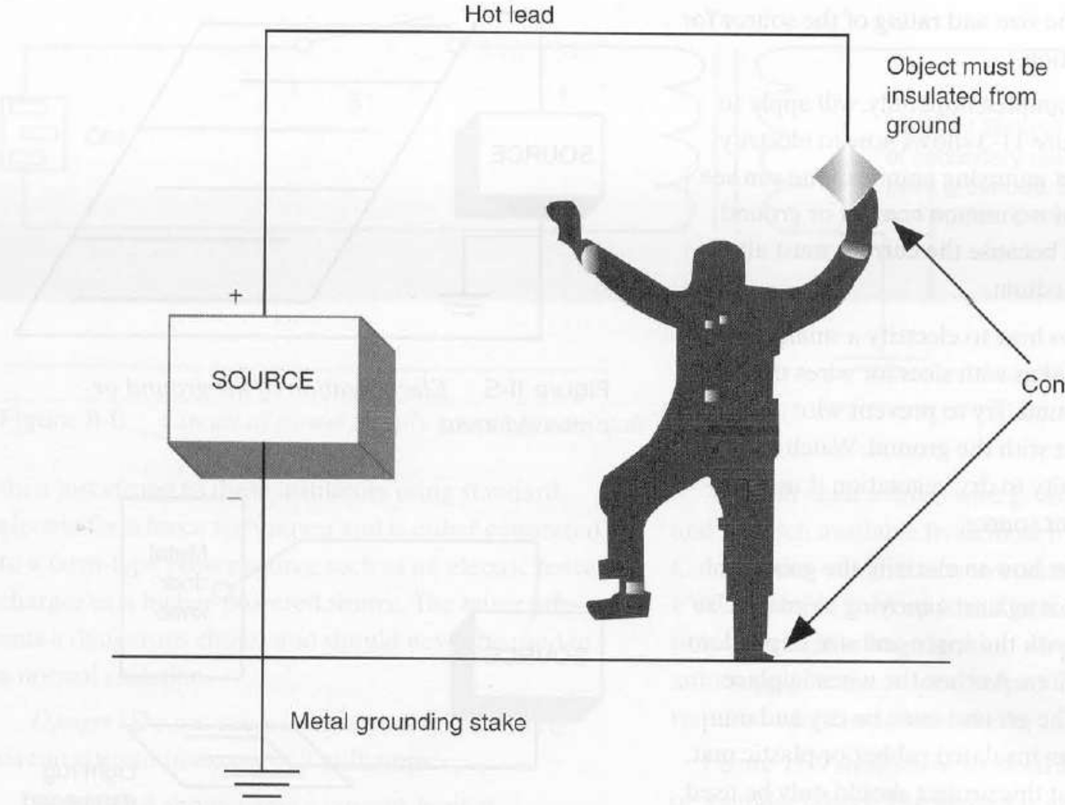
Hot lead
Object must be insulated from ground
Contact points
+
SOURCE
−
Metal grounding stake

Figure 11-2 *Electrification principle*

enough to cause this current to flow, usually several thousand volts. (These parameters are similar to those used in electric chairs.) Currents of more than $^1/_{10}$ of an amp are considered lethal to the average human under certain circumstances, $^1/_{100}$ of an amp produces intense pain, and $^1/_{1,000}$ of an amp is felt with no difficulty and is very annoying.

Special pulse-type energy sources, similar to capacitor discharge ignitions, can be used relatively safely as peak currents are high, yet the average is low. A source of this type is the pulsed model described in this project. This source is capable of generating the high-peak current pulses and maintaining a substantial voltage on the body of a vehicle in spite of leakages due to steel-belted tires and so on. The source is also relatively safe to exposure but will produce moderate shocks. With these parameters, one

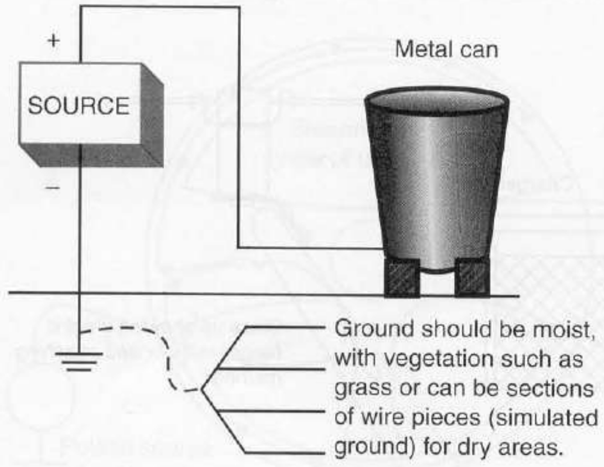
+
SOURCE
−
Metal can
Ground should be moist, with vegetation such as grass or can be sections of wire pieces (simulated ground) for dry areas.

Figure 11-3 *Electrification of a garbage can or similar object*

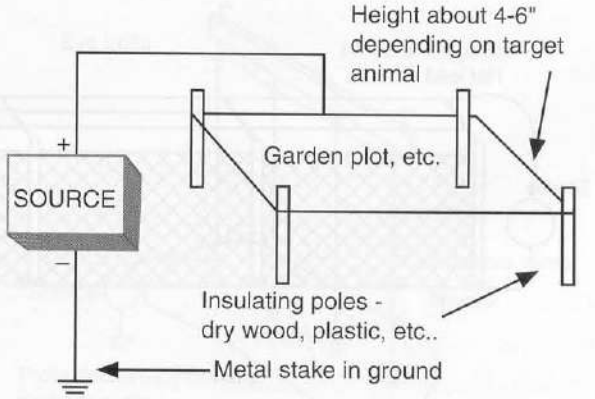
Height about 4-6" depending on target animal
Garden plot, etc.
+
SOURCE
Insulating poles - dry wood, plastic, etc..
Metal stake in ground

Figure 11-4 *Electrification of a small area*

can get an idea of the size and rating of the source for a particular application.

The following examples, hopefully, will apply to most situations. Figure 11-3 shows how to electrify garbage cans against annoying animals. One can see that the condition of a common contact or ground system is important because the current must also flow through this medium.

Figure 11-4 shows how to electrify a small garden area using plastic stakes with slots for wires that are pushed into the ground. Try to prevent wire droop and possible contact with the ground. Watch for fire hazards and proximity to dry vegetation if using a high, average current source.

Figure 11-5 shows how to electrify the ground of an area and protect it against annoying animals. Use #20 wire as shown with the space and size dependent on the animal and area. Anchor the wires in place with small stones. The ground must be dry and non-conductive, or use an insulated rubber or plastic mat. Please be aware that this project should only be used with a low-energy pulsed system or a hand-cranked magneto. Do not create this system if you have a heart condition.

Figure 11-6 shows a prank, but this is only to be done using a low-energy pulsed system or a hand-cranked magneto. The victim stands on a rug to open a door and zap! You turn the power on.

Figure 11-7 shows the electrification of a fence. This method is for educational use only, as serious injury can result either from the electric shock or the physical reaction. The intended use would be for military or other high-security situations.

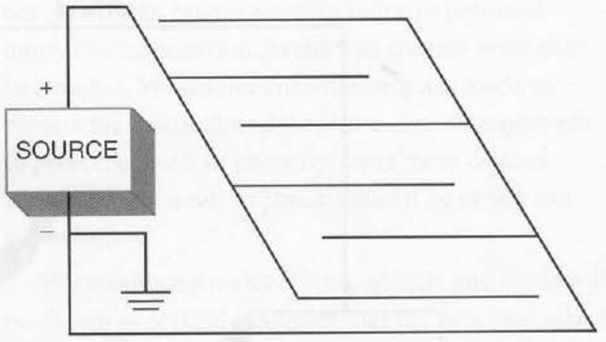

Figure 11-5 *Electrification of the ground or protected areas*

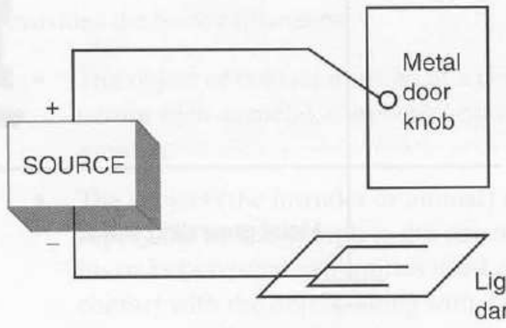

Figure 11-6 *Door knob prank*

The object here is to electrify the top-most wire of a chain-link fence, as anyone attempting to crawl over will no doubt come in contact with the top wire. This is a very effective method and can cause severe injury and even death from electrocution or physical injury from falling. The best method is to simply replace the original top-most wire with a standard, electrical farm fence insulator, as shown in Figure 11-7. The wire is

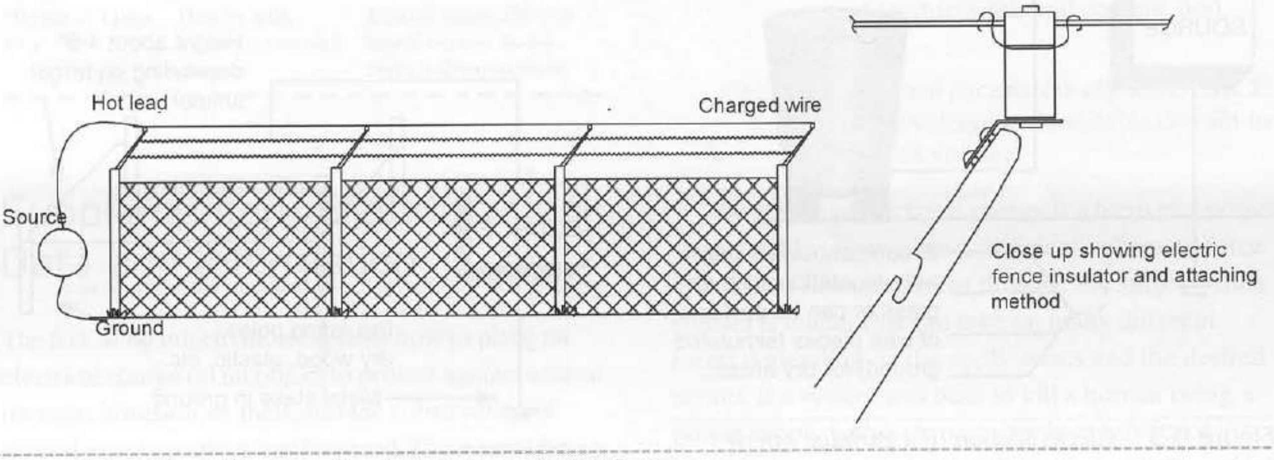
Hot lead
Source
Ground
Charged wire
Close up showing electric fence insulator and attaching method

Figure 11-7 *Electrification of a chain-link or similar fence*

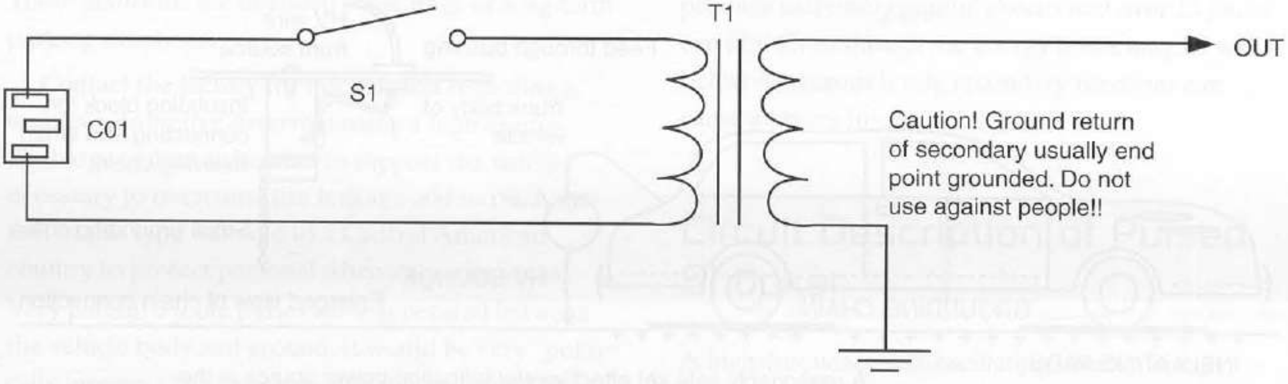

Caution! Ground return of secondary usually end point grounded. Do not use against people!!

Figure 11-8 *Circuit of power supply source for animals only*

then just strung to these insulators using standard, electric farm fence techniques and is either connected to a farm-type power source such as an electric fence charger or a higher-powered source. The latter presents a dangerous choice and should never be used in a normal situation.

Danger! Do not use a transformer with a short-circuit current in excess of 9 milliamps.

Figure 11-8 shows using a current-limited transformer powered by 115 VAC. The following are suggested power sources for the circuit shown in Figure 11-8.

4KV/009: T1 −4 KV, 9 ma, 115 VAC, current-limited floating secondary Database #4 KV/.03 A

6KV/02: T1 6.5 KV, 20 ma, 115 VAC, current-limited floating secondary Database #6 KV/.02 A

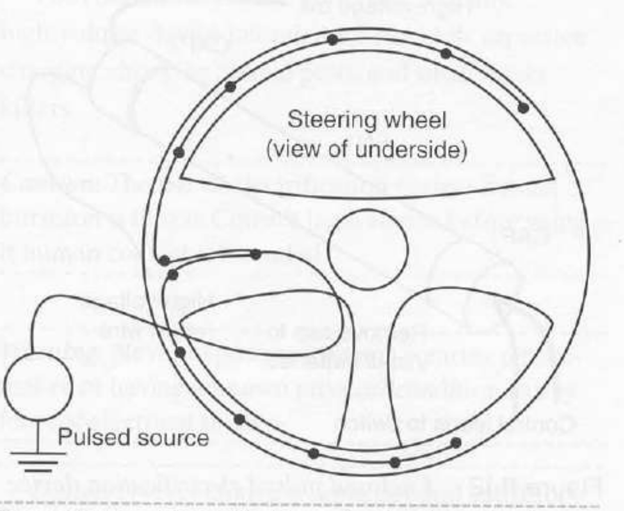

Figure 11-9 *Electrification of a steering wheel or similar object*

You will need a three-wire grounded power cord and a switch available from most hardware stores. Connect the −connection to common using standard PVC or plastic hookup wire. Connect the +connection via a high-voltage ignition wire available from an auto supply house. A weatherproof system is required when installed outside.

Figure 11-9 shows how to electrify a steering wheel or a similar object. The project is intended as a theft or unauthorized use deterrent. Use a low-energy, pulsed, high-voltage source.

The subject sits in the seat and activates the pressure switch to the source. When he grabs the steering wheel, he experiences a mild pulse-type shock between his hand and rear end. The voltage return is via metal springs in the seat. To create this project, please follow these steps:

1. Tape some thin, bare wire on the underside of the steering wheel. Secure it with small pieces of tape, wax, or silicon rubber.

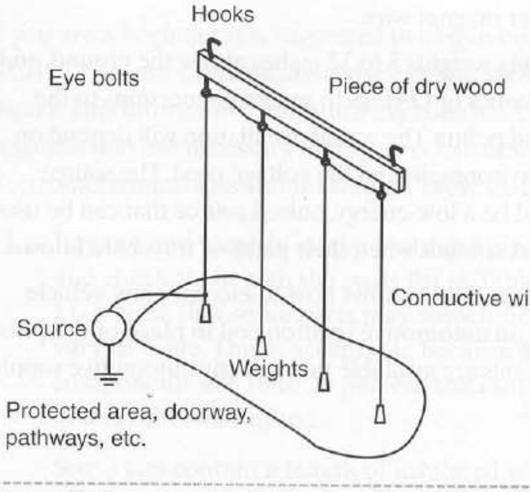

Figure 11-10 *Space or field electrification*

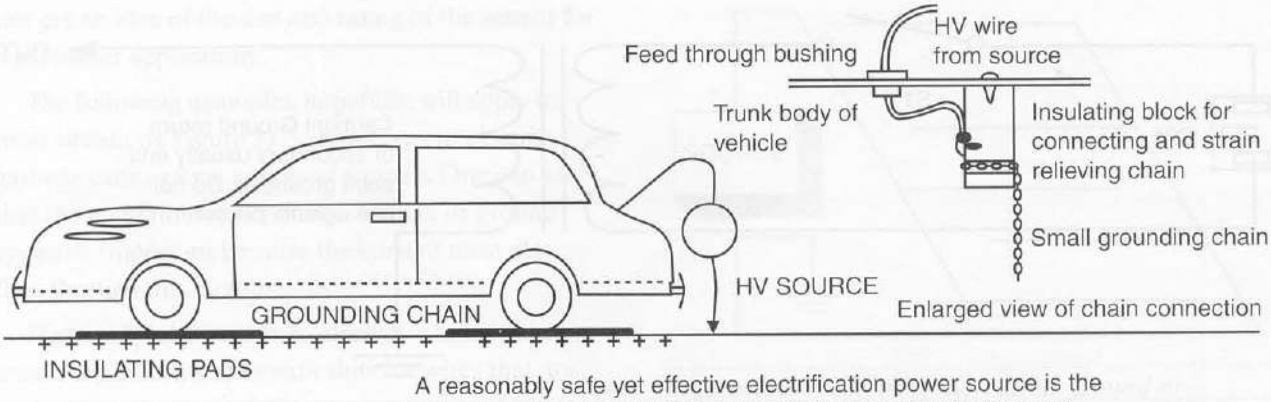
Feed through bushing
HV wire from source
Trunk body of vehicle
Insulating block for connecting and strain relieving chain
Small grounding chain
HV SOURCE
Enlarged view of chain connection
GROUNDING CHAIN
INSULATING PADS

A reasonably safe yet effective electrification power source is the pulsed version shown, with a standard automotive ignition coil replacing the small trigger coil. This approach lengthens the pulse to overcome the electrical capacity of the vehicle body and provides a better overall effect.

Figure 11-11 *Electrification of a vehicle or similar object*

2. Run a concealed connecting wire up the steering column to the bare wire. This wire should be insulated and connect to the output of the source.
3. Conceal an activation switch as desired and don't forget to de-energize it before entering the car. You may want to rig up a pressure-sensitive switch on the seat so when the thief gets in the unit it activates before he starts the car. Important caution: The unit must be activated without the car being in motion or serious accidents may result.
4. The use of our pulsed low-power source is suggested.

Figure 11-10 shows how to protect an area, doorway, or paths using space or field electrification. Hanging supports use ultra-thin steel wire for strength or weaker magnet wire.

Hang weights 8 to 12 inches above the ground, and use elastics or O-rings to secure connections to the ground points. The actual installation will depend on the environment and the voltage used. The source should be a low-energy, pulsed source that can be used against animals when their paths of travel are known.

Figure 11-11 shows how to electrify any vehicle using an automotive ignition coil in place of the pulse coil. Coils are available through any automotive supply house.

A remote switch must be mounted so that it is accessible from the outside of the vehicle. This may be actuated via an insulating object such as a plastic comb or something similar. An override switch may be installed to actuate this system while one is inside the vehicle in order to discourage certain encounters involving contact while driving or stopped at a light.

The suitable method for connecting to earth ground as the common terminal is accomplished via a grounding chain. This is shown strain relieved via a plastic insulating block. The user is asked to use his or her own ingenuity to create other methods of implementation.

Insulated platforms for wheels should be used if the tires have steel-belted radials or if roads are salty.

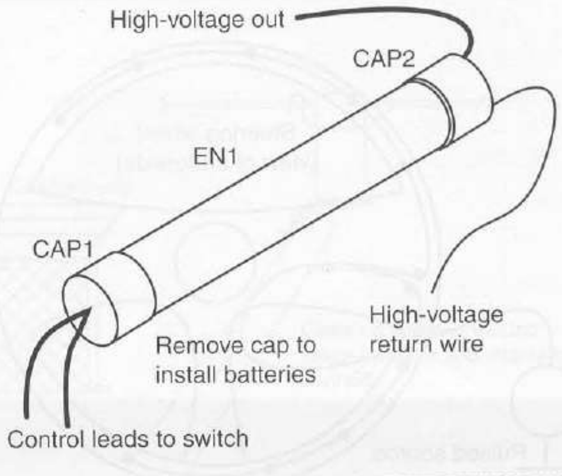
High-voltage out
CAP2
EN1
CAP1
Remove cap to install batteries
High-voltage return wire
Control leads to switch

Figure 11-12 *Enclosed pulsed electrification device*

These platforms are intended for storage or long-term parking situations.

Contact the factory for information regarding a serious automotive deterrent using a high-energy discharge system to be able to support the voltage necessary to overcome tire leakage and so on. A system of this type was sold to a Central American country to protect personal when in hostile areas. Very painful 5 joule pulses were generated between the vehicle body and ground. It would be very "politically incorrect" if it were used in this country. This system produces high-current pulses of several tens of amperes for short durations. This is necessary as most tires offer considerable leakage at these high voltages, and a steady-state supply would no doubt burn off the tires and be very lethal to the perpetrator.

Assembly of Suggested Power Sources

The *pulsed version* shown in Figure 11-12 can also be used as a high-voltage trigger for triggering spark gaps, flash tubes, chemical igniters, Kirklian photography, and for electrification in animal control. The unit is housed in a tubular enclosure and powered by built-in batteries or by an external 12 VDC wall adapter. The device produces 20,000-volt pulses for electrifying objects, fences, and so on. Dual-mode operation produces low-energy pulses of .02 joules at a repetition rate of 30 per second or medium-energy pulses of .2 joules at a repetition rate of 4 per second.

The *continuous version* generates a steady, high-voltage device intended for research, capacitor charging, shocking animal pests, and small insect killers.

Caution: The use of electrification against human intrusion is illegal. Consult legal advice before using if human contact is intended.

Warning: Never expose any person wearing a pacemaker or having a known physical condition to any form of electrical shock.

Danger! Never charge a capacitor and use it for shocking humans. Energy storage over 5 joules can produce extremely painful shocks and over 25 joules can kill! Even though the energy levels may be well below dangerous levels, secondary reactions can cause a severe injury.

Circuit Description of Pulsed Shocker

A high-frequency, self-oscillating inverter circuit comprised of switching transistor Q1 and stepup transformer T1 produces a high-voltage high frequency at the secondary winding. This high-voltage AC is rectified by diode D1 and charges up storage capacitor C3 or C4 through isolation resistor R3. When this voltage charges up to the breakdown potential of SIDAC (SID1), the energy stored in the capacitors is "dumped" into the primary of the high-voltage pulse transformer T2, producing a high-voltage pulse at the output terminals (see Figure 11-13).

The oscillator circuit utilizes a winding on T1 to produce the necessary positive feedback to the base of Q1 to sustain oscillation. Resistor R1 initiates Q1 to turn on while resistor R2 and C2 control the base current and operating point.

A charging circuit consisting of the current-limiting resistor R4 and rectifier diode D2 allows external charging of battery B1 when *nickle cadmium* (NiCad) or other rechargeable batteries are used.

Construction Steps for Pulsed Shocker

If you are a beginner it is suggested to obtain our *GCAT1 General Construction Practices and Techniques*. This informative literature explains basic practices that are necessary in proper construction of electromechanical kits and is listed on Table 11-2.

1. Lay out and identify all the parts and pieces, and check them with the parts list in Table 11-2. Note that some parts may sometimes vary in value. This is acceptable because all components are 10 to 20 percent tolerant unless otherwise noted.

   Some kits contain a length of insulated wire.

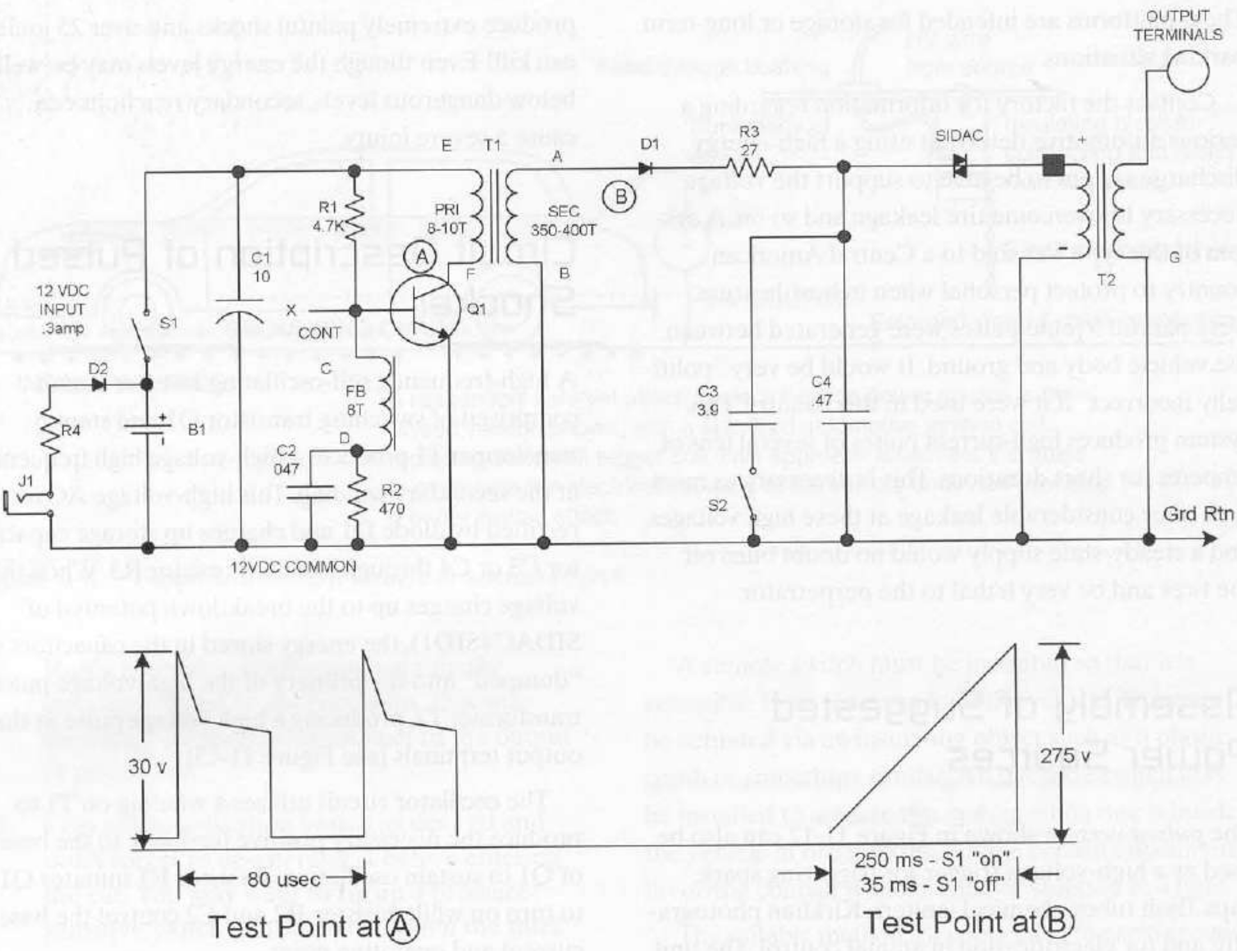
OUTPUT TERMINALS
E
T1
A
D1
R3
27
SIDAC
+
R1
4.7K
PRI
8-10T
SEC
350-400T
B
C1
10
F
B
T2
G
12 VDC INPUT .3amp
S1
X
CONT
Q1
D2
C
FB
8T
C3
3.9
C4
47
R4
B1
C2
.047
D
J1
R2
470
S2
Grd Rtn
12VDC COMMON
A
30 v
80 usec
Test Point at A
275 v
250 ms - S1 "on"
35 ms - S1 "off"
Test Point at B

Figure 11-13 *Circuit schematic*

This must be cut, stripped, and tinned according to where it is used.

2. Assemble the board as shown in Figure 11-14. Insert the components into the board holes as shown, proceeding from left to right. Note the polarity of the components.

Note: Certain leads of the actual components will be used for connecting points and circuit runs. Do not cut or trim them at this time. It is best to temporarily fold the leads over to secure the individual parts from falling out of the board holes for now.

---

3. Verify the wiring, proper components, and polarity to the schematics. Check for cold solder, excess solder, and bare-wire bridge shorts. Now the unit is ready to test.
4. Test the unit as per the following:

A. Position the bare end of the ground return lead (GRD RTN) to allow a $^1/_2$- to $^3/_4$-inch air gap between the output pins of T2.

B. Connect 12 volts from a bench power supply or use eight AA NiCads at 1.25 volts each or eight AA alkalines at 1.5 volts for 12 volts.

C. Note with S2 open a fast pulsing action will produce a thin, bluish discharge. This can cause a very mild electrical shock and could be safely used within reason as a prank.

D. Note a thick, slow pulsing discharge when S2 is closed. This can produce a painful shock and is intended for use against animals.

E. The current draw with the unit properly operating should be approximately 250 milliamps.

F. Check the power tabs of both Q1 and SIDAC. These should be cool to warm to the touch.

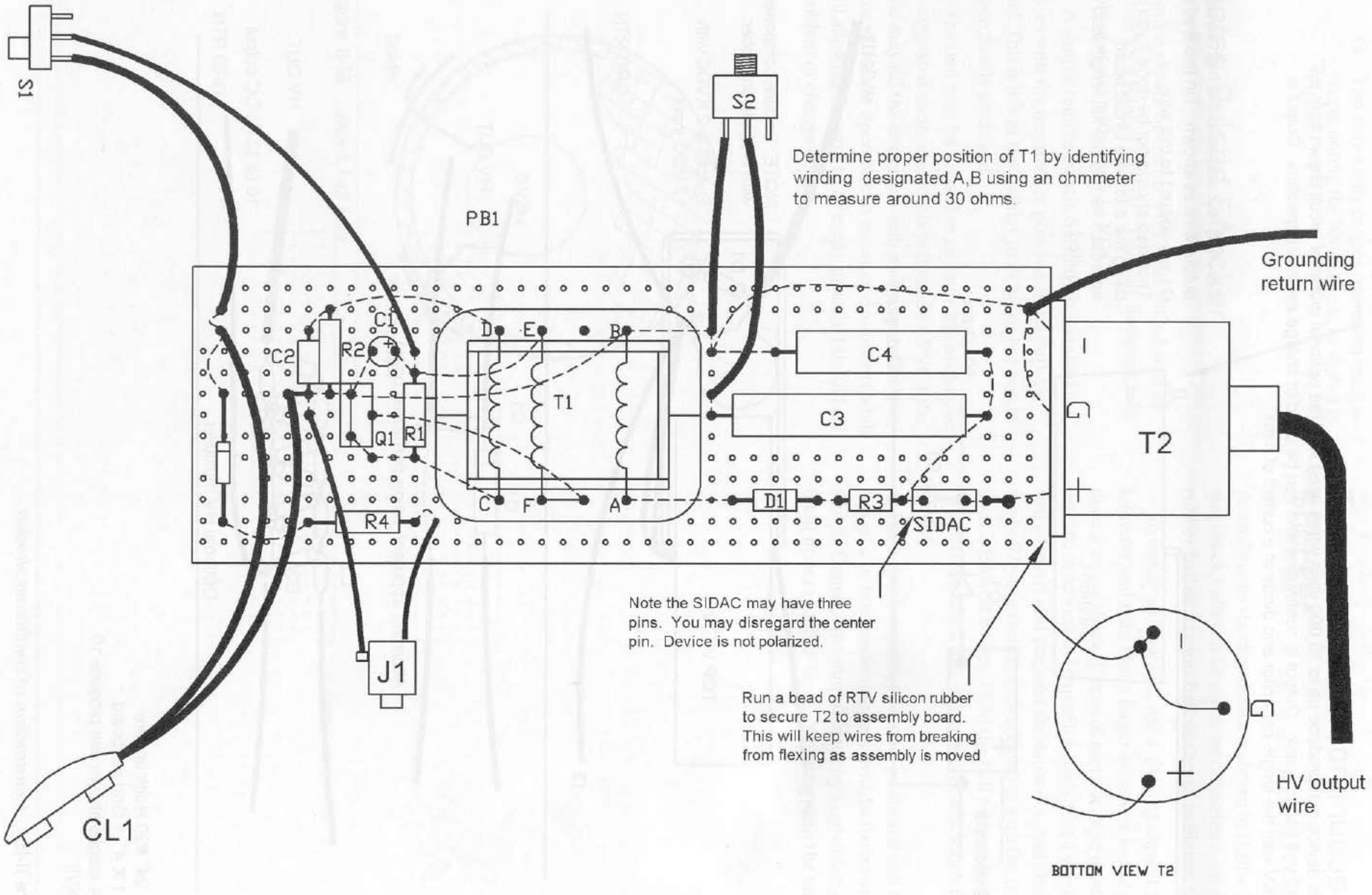
S1
S2
PB1
Grounding return wire
C1
C2
R2
Q1
R1
D
E
B
T1
C
F
A
C4
C3
R4
J1
D1
R3
SIDAC
T2
G
+
-
CL1
HV output wire
BOTTOM VIEW T2

Determine proper position of T1 by identifying winding designated A,B using an ohmmeter to measure around 30 ohms.

Note the SIDAC may have three pins. You may disregard the center pin. Device is not polarized.

Run a bead of RTV silicon rubber to secure T2 to assembly board. This will keep wires from breaking from flexing as assembly is moved

Figure 11-14 *Assembly board*

This device can produce up to 10,000 volts when used with the optional multiplier circuit shown from a 9 to 12 volt battery pack. Output is continuous and can be used to charge external capacitors. Output is 1.5KV with the single capacitor and diode at a current of 5 ma.

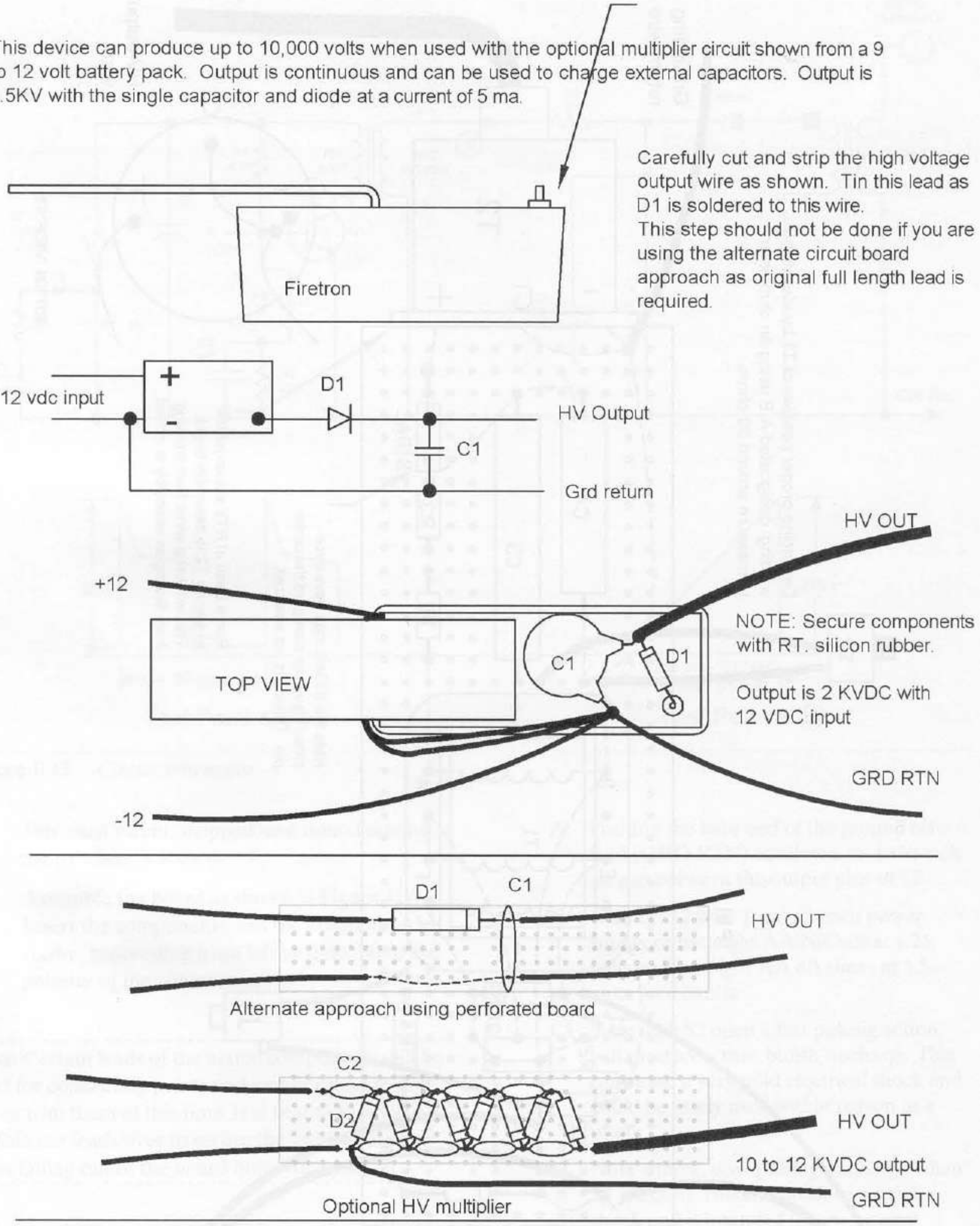
Firetron
12 vdc input
+
-
D1
HV Output
C1
Grd return
HV OUT
+12
TOP VIEW
C1
D1
-12
GRD RTN
D1
C1
HV OUT
C2
D2
HV OUT
GRD RTN

Carefully cut and strip the high voltage output wire as shown. Tin this lead as D1 is soldered to this wire.
This step should not be done if you are using the alternate circuit board approach as original full length lead is required.

NOTE: Secure components with RT. silicon rubber.

Output is 2 KVDC with 12 VDC input

Alternate approach using perforated board

10 to 12 KVDC output

Optional HV multiplier

WR1 24" #20 Hook up wire
PB1 1 X 4" .1 Grid perfboard
Note 4 stages shown can produce 10 to12 KV!!

Figure 11-15 *Construction of continuous shocker*

G. You may verify proper operation using a scope, noting the waveshapes as shown in Figure 11-13.

Notes Pulsed Shocker

Input is shown operating with eight NiCad cells in series for a total of 10 VDC. The unit, however, reliably operates within 8 to 14 VDC.

A simple remote switch, S1, may be installed in series with the negative or positive of the battery lead. This is left to the builder to fit his or her needs when finally packaged.

The unit may be housed in any suitable enclosure. A suggested method is to place it inside a PVC tube You may fill the enclosure with melted paraffin wax. This will allow operation in moist environments while still allowing the unit to be easily unsealed should a problem or change occur

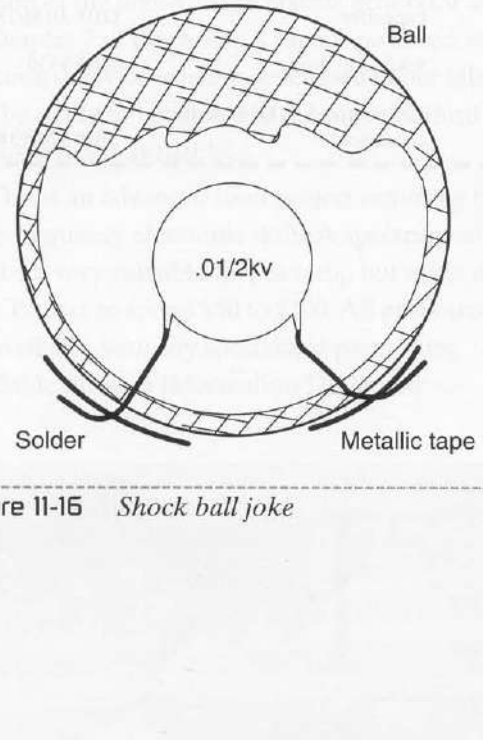

Figure 11-16 *Shock ball joke*

Construction of Continuous Shocker

A continuous shocker device is shown in Figure 11-15, and shock balls can be made by obtaining a small rubber ball or a tennis ball, as shown in Figure 11-16.

To create a shock ball, use a .01-amp capacitor at 2 kilovolts and note that a larger capacitor would produce a worse shock. Pieces of metallic tape must be used to sandwich the capacitor leads. Cut a slit in the ball to insert and position the capacitor, and charge the ball by contacting a metallic strip with the continuous 2-kilovolt source. Hold the ball without touching the metallic strips, toss it to a buddy, and watch the results.

For a shock wand application, position a unit inside a plastic or nonconductive tube with batteries and a switch. Connect the output leads by sandwiching them under metallic tape for contact with the target subject.

Table 11-2 Parts List for Shockers

Ref. #	*Description*	*DB Part #*
R1	4.7K, $^{1}/_{4}$-watt carbon film resistor (yel, pur, red)	
R2	470 ohm, $^{1}/_{4}$-watt carbon film resistor (yel, pur, br)	
R3	27-ohm, $^{1}/_{4}$-watt carbon film resistor (red, pur, blk)	
R4	100-ohm, $^{1}/_{4}$-watt carbon film resistor (br, blk, br)	
C1	10-microfarad, 25-volt electrolytic capacitor vertical mount	
C2	.047-microfarad, 50-volt polyester capacitor	
C3	3.9 to 4-microfarad, 350-volt polyester capacitor	
C4	.47-microfarad, 250-volt polyester capacitor	
Q1	MJE3055 NPN transistor TO220	
D1, 2	Two IN4007 1-kilovolt rectifier diode	
SIDAC	300-volt Sidactor switch	SIDAC
T1	400-volt switching square-wave transformer	DB# TYPE1PC
T2	25-kilovolt pulse transformer	DB# CD25B
S1, 2,	Two *single pole, single throw* (SPST) 3-amp toggle switch or equivalent	
PB1	5 × 1.5 × .1-inch grid perforated circuit board	
PCLITE	Optional *printed circuit board* (PCB)	DB# PCLITE
WR20B	36-inch #20 vinyl stranded hookup wire, black	
WR20R	36-inch #20 vinyl stranded hookup wire, red	
WRHV20	12-inch 20-kilovolt wire	
CL1	Battery clip	
CAP1, 2	Two $1^{5}/_{8}$-inch plastic caps	
EN1	12 inches of $1^{5}/_{8}$ × $1^{1}/_{2}$ × 12 plastic tube	
BH1	Eight-AA-cell battery holder	
J1	3.5-millimeter mono jack	
12DC/.3	Optional 12 VDC, .3-amp wall adapter	DB# 12DC/.3
Parts for the continuous version		
FIRETRON	Firetron 2,000-volt module	DB# FIRETRON
C1	.01-microfarad, 2,000-volt disc capacitor	DB# .01M/2KV
D1	6-kilovolt diode	DB# VG6
C2	270-picofard, 3,000-volt disc capacitor	DB# 270P/3KV

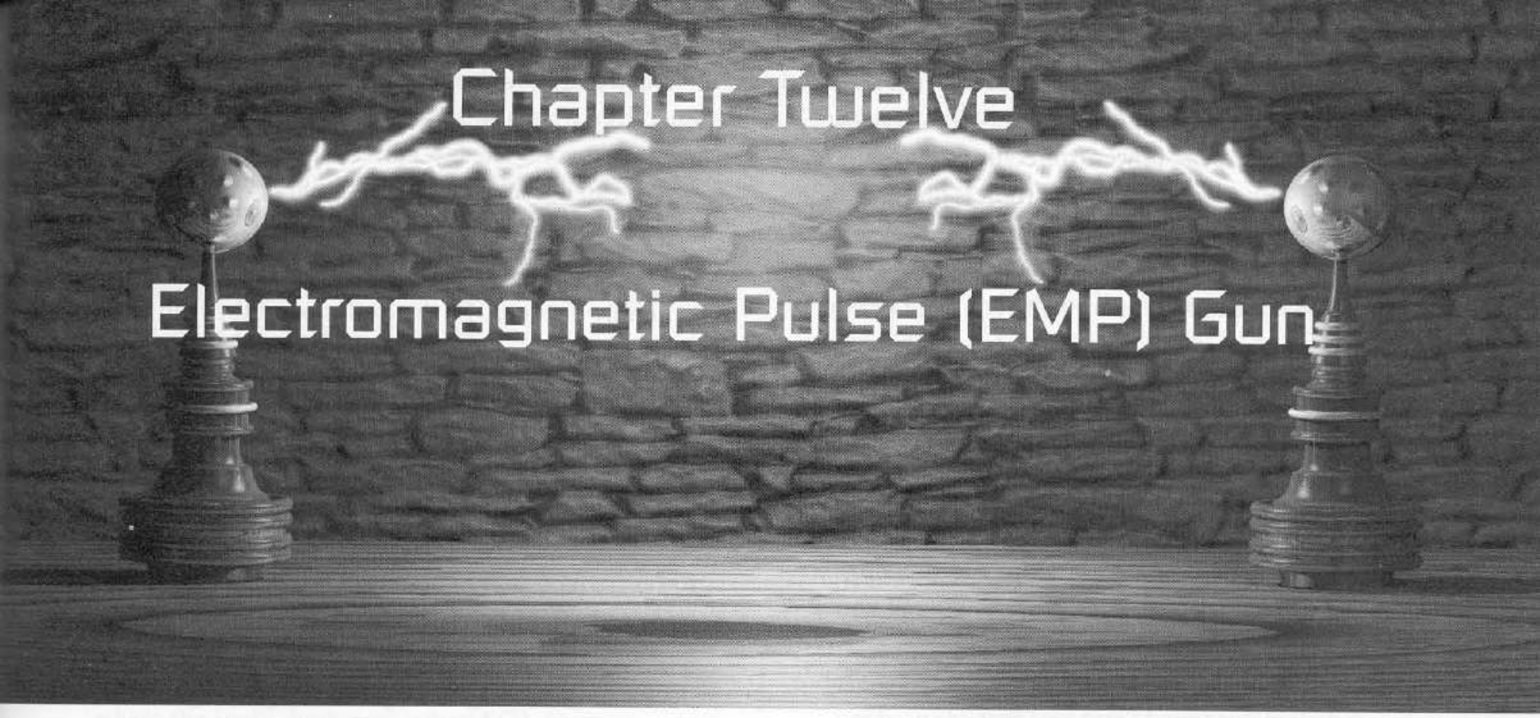

Chapter Twelve

Electromagnetic Pulse (EMP) Gun

Figure 12-1 shows the construction of a low-power pulse gun that will provide kilowatts of peak power at frequencies up to 100 MHz with harmonics. The unit shown is battery powered and uses a modified version of the high-voltage plasma generator shown in Chapter 7 of this book. A higher powered, more functional device is being developed in our labs and will be ready to be copied at the time the third book of this series is available.

This is an advanced level project requiring basic high-frequency electronic skills. A spectrum analyzer can be a very valuable tool in setup but is not necessary. Expect to spend $50 to $100. All parts are readily available, with any specialized parts being available through Information Unlimited (www.amazing1.com) as listed in the parts list at the end of this chapter.

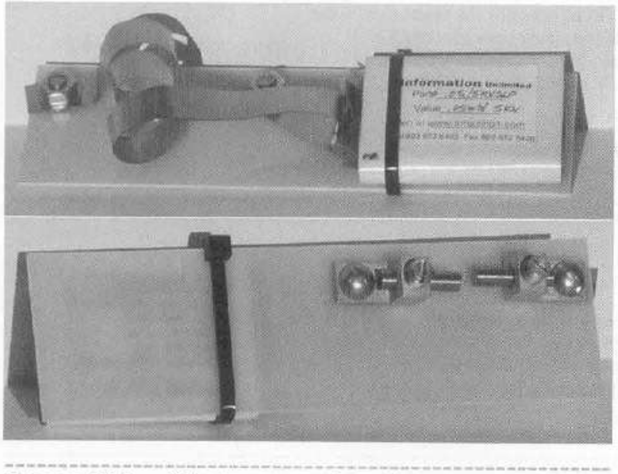

Figure 12-1 *Electromagnetic pulse gun*

Basic Theory Simplified

The ability of a signal to disrupt sensitive circuitry requires several properties. Most microprocessors consist of *field effect transistors* (FET) operating at very low voltages. Once these voltages are exceeded, catastrophic failure becomes imminent. Forgiveness to an overvoltage fault is practically nonexistent due to a microthin metal oxide between controlling elements. Any overvoltage generated across these elements consequently produces permanent damage or, in some less severe cases, causes deprogramming. To generate these damaging voltages from an external source requires a wave that can produce a standing wave of energy across circuit board traces, components, and other key points. The external signal energy therefore must be high enough for the circuitry geometry to be a significant part of energy at the wavelength. Therefore microwaves having fast rise times (high Fourier equivalents) and burst duration help to maximize the effect. The energy required is great, as it must be sufficient to enhance damage. A good measure would be the quotient of energy/wavelength. A high-power pulse of microwaves can be generated in several ways.

Explosive flux compression-driven virtual cathode oscillators and their generic relatives can produce gigawatts of peak power from only hundreds of kilojoules. This is where a seed current is pulsed into an inductor and at its peak is compressed by a shaped explosive charge thereby trapping the flux and creating a source of high energy. The coil must be compressed both along its axis and along its radius using a high-detonation velocity explosive such as the cyclotrimethyltrinitramine, its derivatives PETN, or some other equivalent energetic explosive. This trapped flux now produces an energy gain that is conditioned into the final peak-powered pulse of microwave power (HEPM). Flux compression like that of a nuclear initiation requires precise timing of the explosive chargers. For flux compression, Krytron switches or similar can be used instead of the more radiation-hardened Sprytrons that are used in nuclear initiations where ionizing radiation is produced from the inherent fissionable materials.

A virtual cathode oscillator can also easily be energized from a small Marx impulse generating 200 to 400 kilovolts. The fast current rise and high peak power can produce a powerful burst of microwaves.

Other methods include exploding wires where energy is allowed to flow into a LCR circuit and then is rapidly disrupted by the explosion of the feed wire as it vaporizes at near the peak injection current. A very fast and energetic pulse is produced, which is capable of generating an *electromagnetic pulse* (EMP).

Microwave pulses are excellent candidates for damaging sensitive electronic circuitry. But much lower frequencies are better for disrupting power grids and other similarly sized systems, as now the lengths of the conductive elements are more conducive to generating the high standing voltage waves. Obviously more energy is now required, as breakers, switches, and transformers require more energetic pulses

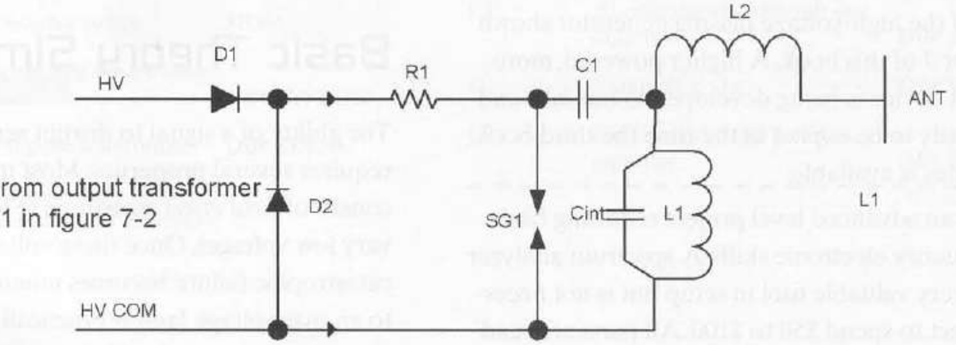

The ideal peak power of this circuit can be approximated by evaluating the product of the charge voltage across C1 x the peak current as determined by Ipk = Epk x the square root of the capacitance of C1/the total inductance of the discharge circuit. This expression implies no resistance (R) in the discharge circuit that now is only in an ideal situation. In all actuality the resistance in the real world will now be a factor in where e to the -(pi/square root of L/C) / R must be a factor in the above ideal condition. This now implies a damped waveform.

Diodes D1,2 are 10 kv 10 ma fast recovery
Resistor R1 is three 47k 1 watt resistors in series and isolate the diodes from the dv/dt of the discharge.
L2 inductor tunes out the capacitive reactance of the antenna at the desired resonant frequency

Capacitor C1 is a "slapper" capacitor. These are used to produce a very fast rise time necessary to detonate initiators necessary for initiating high explosive. Its claim to fame is the high peak current discharge current. The capacitor is constructed as strip lines with the connection leads exiting a common end.

Figure 12-2 *Circuit schematic*

Circuit Description

The circuit shown in Figure 12-2 shows a simple method of obtaining a high-power pulse with Fourier equivalents above 100 MHz. Even though the power and frequency are relatively low, close range effects are possible on many target circuits.

The project uses a high-frequency plasma source, which is converted to a direct current charging source and is short circuited functionally by the use of lossless reactive ballasting. This means that the capacitor can be charged without the use of an energy-robbing resistor, as only the real part of the complex current is seen by the battery supply.

The modified source now supplies a current charge to the reservoir capacitor (C1) to a value where the spark discharges across SG1. Current through L1 rapidly rises and rings along with the circuit and lumped capacitance (Cint). Spark gap SG1 must turn off and allow the energy to circulate in the discharge in order to generate a resonant peak of power that is now coupled into the system emitter. Experimentation of the gap settings is necessary to obtain an optimum effect.

Circuit Assembly

1. Fabricate a piece of Lexan plastic or G10 circuit board, as shown in Figure 12-3, to a 7 × 2 inch plate for PL1. Locate the two 1/4-inch spark-gap-holder (SGH1, 2) screws holes as shown. The holes are spaced 2 3/4 inches from one another.
2. Form a three-turn 1-inch diameter coil (L1) from 1/2-inch sheet copper or #14 solid copper wire, as shown in Figure 12-4. Note the leads as attached to the capacitor (C1) and the end gap holder.
3. Connect an output port at the junction of C1 and L1 as shown. This point is now considered the output and can be connected to the radiating emitter.
4. Connect it to the converted plasma driver in the manner shown in Chapter 7 to the generator (see the schematic Figure 12-2). Note the added diodes D1 and D2 for converting the output to direct current
5. You will note a coil (L2) connected in series with the output lead. This inductor tunes out the capacitive reactance of the lead and the end capacitance. Experiment using a radio wave or absorption wave meter to determine the resonant frequency of L1/C1. Select a value for L2 to provide maximum radiation from a distance.
6. Experiment on various electronic devices and observe the effects at various distances.

Notes

A low-cost spectrum analyzer would be a great aid in setting up this system. Please note this is a low-power device intended as an introductory project for those desiring to experiment with shock pulse and EMP research. Several more functional and sophisticated devices will be featured in the next book in this series.

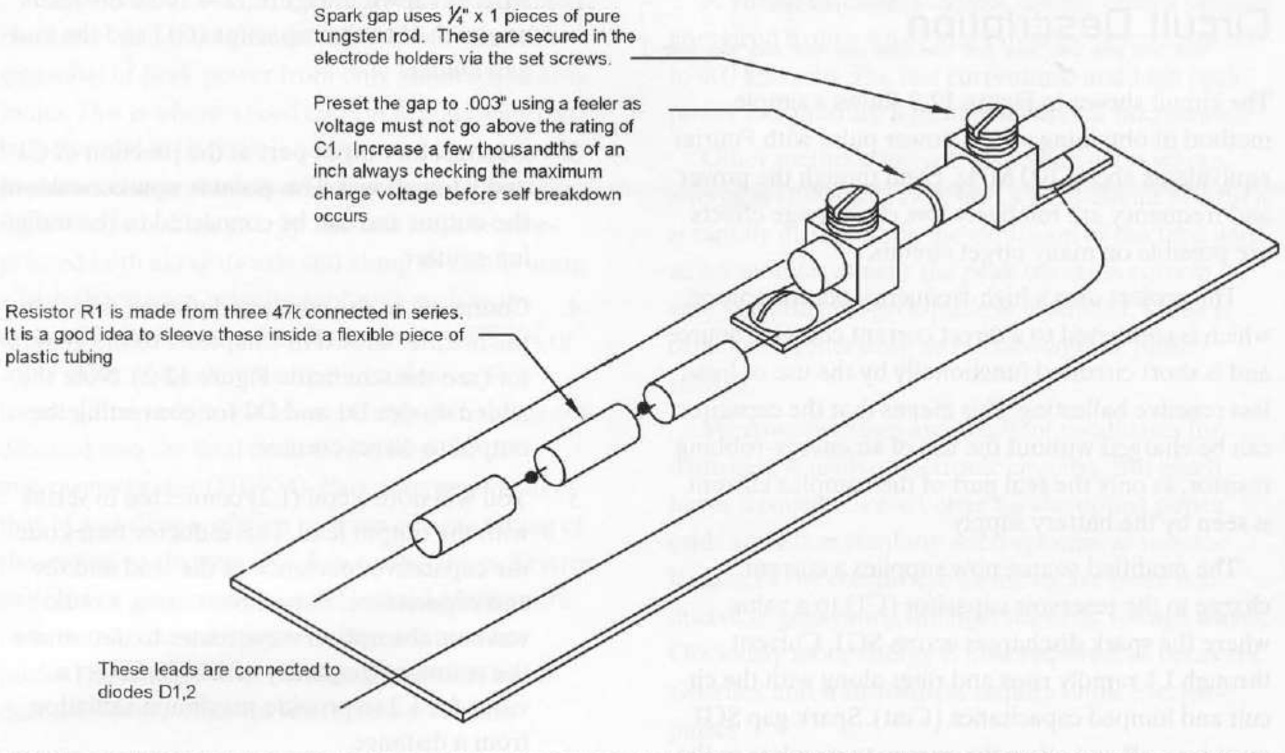
Spark gap uses ¼" x 1 pieces of pure tungsten rod. These are secured in the electrode holders via the set screws.
Preset the gap to .003" using a feeler as voltage must not go above the rating of C1. Increase a few thousandths of an inch always checking the maximum charge voltage before self breakdown occurs
Resistor R1 is made from three 47k connected in series. It is a good idea to sleeve these inside a flexible piece of plastic tubing
These leads are connected to diodes D1,2

Figure 12-3 *Generator board bottom view*

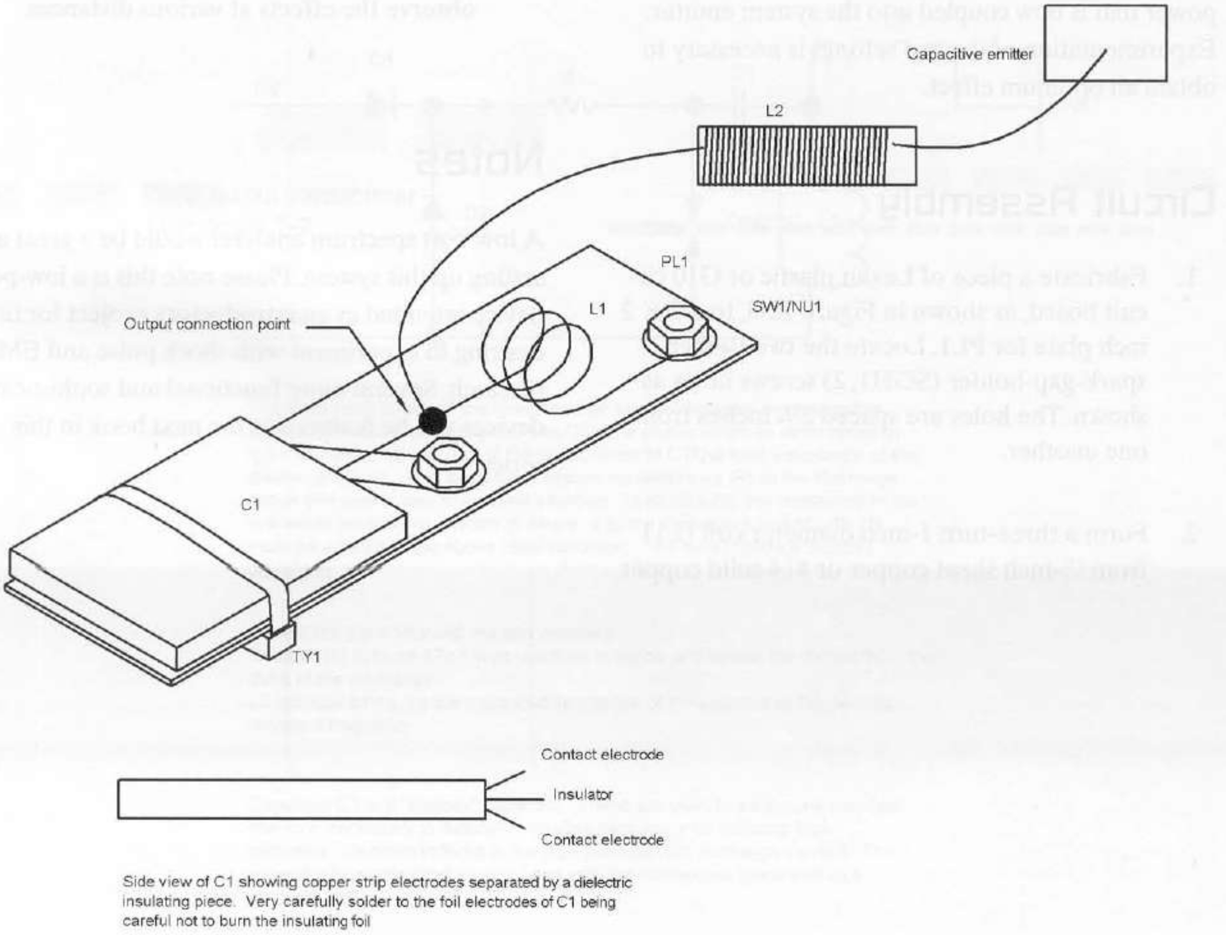
Capacitive emitter
L2
PL1
L1
SW1/NU1
Output connection point
C1
TY1
Contact electrode
Insulator
Contact electrode

Side view of C1 showing copper strip electrodes separated by a dielectric insulating piece. Very carefully solder to the foil electrodes of C1 being careful not to burn the insulating foil

Figure 12-4 *Generator board bottom view*

Table 12-1 EMP Gun Parts List

Ref. #	*Description*	*DB part #*
R1	Three 47K, 1-watt resistors (yel-pur-or) connected in series	
D1, D2	Two 16 KV, 10 ma fast-recovery high-voltage rectifiers	DB# VG16
C1	.05 mfd, 5 KV stripline capacitor	DB#.05M/5KV
L1	Inductor 3-turn, 1-inch diameter use #12 copper wire	
L2	Inductor wind, as directed in text	
ELECTRODES	Two ½ × 1 inch pure tungsten electrodes	DB# TUNG141B
PL1	7 × 2 inch × .063 Lexan or G10 plastic	
TYE1	10-inch nylon tie wraps	
SW1/NU1	Two ¼ × 1 inch brass screw and nuts	
GRAVDRIV1K	Plasma driver kit	DB# GRAVDRIV1K
GRAVDRIV10	Plasma driver assembly	DB# GRAVDRIV10

Chapter Thirteen

Microwave Cannon

This project shows how to build a powerful microwave transmitter intended to be a useful laboratory resource or as a device for disrupting bothersome stereos, boom boxes, and other annoyances perpetrated by inconsiderate hip-hop miscreants. This chapter was prepared by Dr. Barney Vincellete, who holds several doctorate degrees in the electrical sciences. It is a very useful tutorial involving the design of a microwave system, going into basic mathematical details for those who are interested. The system can be built from the parts of a microwave oven.

Warning: The project uses dangerously high voltages and can be a radiation hazard at close range if not used with protective shielding. This is an advanced project intended for experienced builders and must not be attempted by anyone unfamiliar with high voltages or high-power radio frequency circuitry.

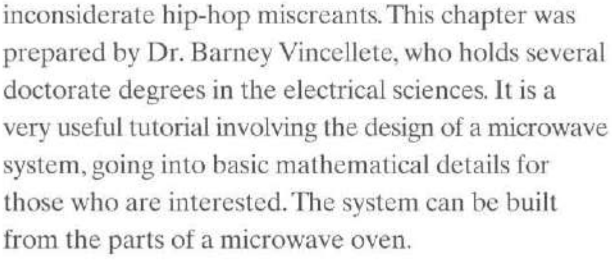

Figure 13-1 *Microwave cannon*

The microwave cannon consists of three parts: the parabolic reflector horn, the electric field focusing lens, and the magnetron. The horn focuses the magnetic oscillations that circumscribe the magnetron antenna into a narrow beam in order to concentrate the radiation into a two-dimensional space. It also prepares the electric field oscillations for the lens to concentrate the radiation in the third dimension. Because the electric field and magnetic field are orthogonal, calculations upon them and the modifications that we will perform upon them can be done on one without altering the other.

In designing the horn, we will assign the Cartesian coordinate system with the z axis along the longitudinal axis of the horn; that is, the z axis will occupy the center of the shaft of radiation that will be fired into your cultural tyrants' stereos. The Poynting vector will be on the z axis, and the x axis will be directed horizontally to the left and right of the z axis. The x axis will, during the calculations for the shape of the horn, be given an absolute value; that is, it will be endowed with a positive value whether it increases to the left or right of the z axis. This is because its values will be the square root of x squared, and it will be symmetrical to the left and right of the z axis. The origin will be located at the vertex of the parabola that will define the sides of the horn.

The y axis will be in the vertical direction to complete the basic vectors in our discourse and calculations. It will pass through the z axis at the mouth of the horn, rather than through the vertex, as the x axis does. It too will have only positive values because, due to its symmetry about the z axis, there is no need to consider negative values of y (see Figure 13-2).

The dimensions of the sides of the horn will be performed first. In building the wooden mold upon which the sheet copper will be formed, an adjustment will be made to correct for the flare that can be seen in Figure 13-3.

The horn will be made of 20-inch-wide sheet copper that can be purchased from a roofing supply house. Because the dimensions of the parabolic sides shown in Figure 13-2 are fixed by the top and bottom pieces of copper that will be soldered to the copper side pieces, the top and bottom pieces will be cut to a parabolic shape with an extra 1/4-inch margin on the sides that will be bent into a ridge that will be soldered to the side piece. The actual parabola will have a maximum width of 19.5 inches, or the maximum value of x will be 9.75 inches. Also, for structural rigidity, a half-inch lip will be bent along the front edge of the top and bottom pieces.

The parabola will contain the magnetron antenna aligned parallel to the y axis and positioned at the focal point of the parabola. The focal point will be one-eighth of the wavelength of the magnetron frequency, which, when the wavelength is computed by dividing the speed of light by the frequency or 2.45 GHz, gives us a wavelength of 2.45 inches and a focal length of 0.6 inches. The parabola is described by the equation x squared = 4 pz, where *p* equals the focal length. When *p* equals .6 inches, *r* equals 2.4 z. From this, we construct Table 13-1.

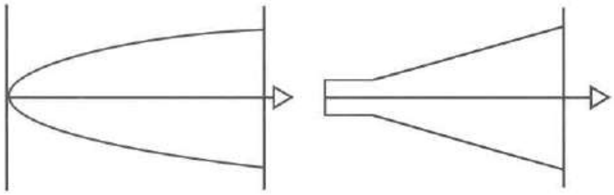

Figure 13-2 *Top and side horn templates*

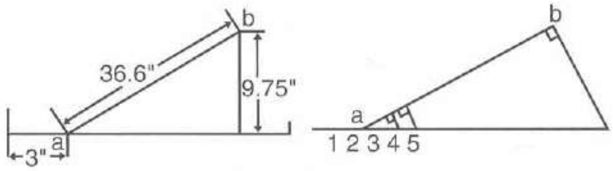

Figure 13-3 *Flair angle template*

Table 13-1 Parabolic Dimensions (measurements in inches)

z	*x*	*z*	*x*	*z*	*x*	*z*	*x*	*z*	*x*
.125	.55	3	2.68	8	4.38	16	6.2	29	8.34
.25	.77	4	3.1	9	4.65	18	6.57	32	8.76
.5	1.1	5	3.46	10	4.9	20	6.93	35	9.17
1	1.55	6	3.79	12	5.37	23	7.43	38	9.55
2	2.19	7	4.1	14	5.8	26	7.9	39.6	9.75

When regarding the side view of the horn (Figure 13-2), it can be seen that plotting the parabola, which is done by cutting the wooden sheets used for making a mold of the horn, will require an adjustment so that the sides of the parabola are not distorted where z is greater than 3 inches. The dimensions of z, if projected from the z axis to the top and bottom of the horn where z is greater than 3 inches, will have to be enlarged by dividing each inch by the direction cosine of the angle between the z axis and the top (also the bottom) of the horn. This can be done by drawing upon a sheet of 2 by 4 foot particle board, as shown in Figures 13-3.

After you have marked the board, as shown in Figure 13-3, you will begin plotting dots measured above and below the z axis at distances of x from Table 13-1 measured with a carpenter's square. You will connect these dots with a ruler, and when you cut out the parabola, the saber saw that you use will make a curved path that will very closely approximate the parabolic shape needed. After you cut out the parabola, you will make another cut along the locus of points where z equals 3 inches.

You will trace these pieces and cut duplicates from another piece of particle board, as shown in Figure 13-4. With nails, screws, glue, and two trapezoidal-shaped pieces of particle board, you will put together the form that you will use to build the copper horn (see Figure 13-5).

Stand the wooden form on its end, where z is 39.6 inches against the floor, and make a template from poster paper that has been taped together to form large enough pieces. Make a template that will wrap about the sides, and trim it to fit so that it matches the sides. Make another template for the top (also the bottom) piece the same way, as shown in Figure 13-6.

You will need 15 feet of copper roofing sheet metal to cut the pieces that you will solder together. Trace and cut the long piece from the template you made from the sides and rear of the wooden form. Add a half-inch to the length of each side so that you can bend a lip that will make the front of the sides rigid. This lip will also provide a place where you can attach the sides of a wooden frame that will make the horn more stable. Then trace the top and bottom pieces against the copper and draw a $^1/_4$-inch additional margin surrounding the parabola you just traced. Also add a half-inch to the front that you can bend into a lip to make the top and bottom more rigid. Cut out two of these pieces, one for the top and one for the bottom.

With a pair of pliers, bend a 90-degree lip from the $^1/_4$-inch margin you traced. You can cut notches in this lip to make this bending easier. This lip will be soldered to the piece that will form the sides of the horn. Bend the half-inch lip at the front edges of the pieces that you will solder together, and clean the insides of the lip that you will be soldering to the sides of the horn using steel wool or sand paper. Clean the exterior $^1/_4$-inch sides of the horn to which you will be soldering the top piece. Apply a thin coat of soldering paste and clamp the sides and top securely into place on the wooden form you built. Hammer the side lip against the sides of the horn for a tight fit. You do not need a sharp crease where the horn begins to flare out where z is 3 inches; a modest bend will do. You are now ready to begin soldering the seam, as shown in Figure 13-7.

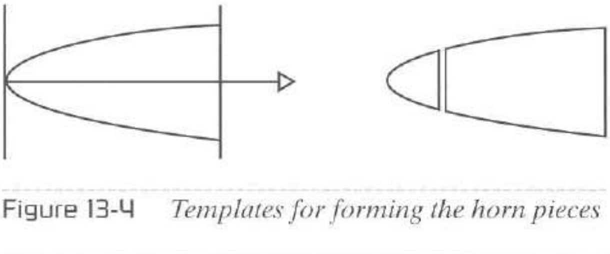

Figure 13-4 *Templates for forming the horn pieces*

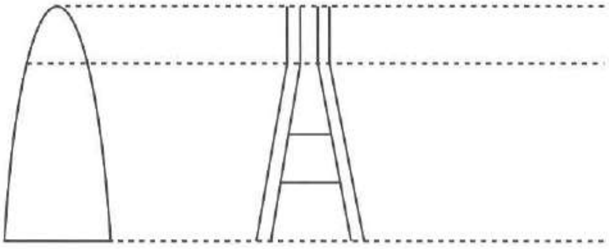

Figure 13-5 *Poster paper template*

Begin with a propane torch and plumber's solder. Apply the flame to a spot and, as soon as it is hot enough to draw in some solder, let just enough solder draw into the seam to fill the gap. Use more hammer taps to close any open spots where the lip is not tightly against the side piece. As soon as you have an inch or two soldered, remove the flame and solder another place a few inches away. Do not use enough solder to cause it to form drips or beads on the inside of the horn. Continue until the entire seam is soldered. After this is done, remove the clamps and sheet copper structure you soldered together. If you see any lumps of solder on the interior seam, you can melt it away by applying a flame.

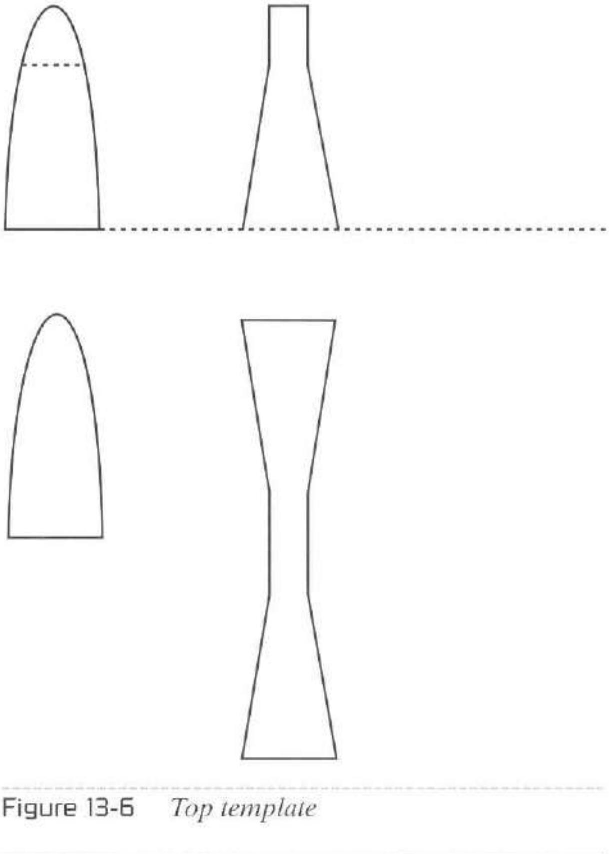

Figure 13-6 *Top template*

Now you must locate the point on the center axis at the top of the horn. This can be done by measuring the center of the front of the top section of the horn, measuring 2 feet along each side of the horn's interior, drafting a line between these points that are 2 feet back from the front, and bisecting it. You bisect the width of the front to find one point. You then measure 2 feet back on each side and draw a line between the ends. You bisect the line for the second point. This will provide two points to extend back to the vertex of the parabola's interior.

Now measure, as punctiliously as you can, 0.6 inches from the vertex. Drill a pilot hole and cut a circle that is $1^{3}/_{16}$ inches in diameter. With a hammer and a piece of pipe, tap on the interior of this hole so that its edges flare outward to make a good seal with the brass wool gasket on the magnetron you will be using (see Figure 13-8).

At this point, solder the bottom piece into place and complete the construction of the copper horn. A wooden frame must be built to make the front edges of the horn more rigid, to provide a structure for mounting the lens, and to provide a place for the front of an aluminum rail where a telescopic rifle sight will be mounted.

Here you have great artistic liberty in how you will build this cabinetry. Half-inch plywood is the most practical material to use, and cedar has the most pulchritudinous finish to stain, varnish, and paint. You can paint or carve artwork upon this frame, and you can also upholster the rear or sides with velvet and sunken buttons. In the most unlikely event the cannon becomes an exhibit in court, these characteristics will humor the jury, many of whom will be as hostile as you against the music the neighbors were forcing into your home. If it becomes newsworthy enough to attract a television crew, it will be a splendid show for the millions of people whom you will inspire to follow your example and join the revolution against cultural tyranny.

The rear of the frame will protect the operator from the deadly 4,500-volt wires between the power supply and the magnetron. It will also provide a place to mount the rear of the aluminum rail for the telescopic rifle sight.

Let us digress to a disquisition on the flair angle formed by the sides of the cannon. In order to form the best possible cancellation of standing waves, due to the impedance vicissitudes where the horn begins to flair at the mouth of the horn, the following equation must be emulated within approximately 10 percent. The dimensions of the horn must match this within a percent.

$$0.25 = [a/(2\lambda)]\tan(\theta/2)$$

where a is the dimension of the mouth in the horn in the y direction: 20 inches. λ is the wavelength in

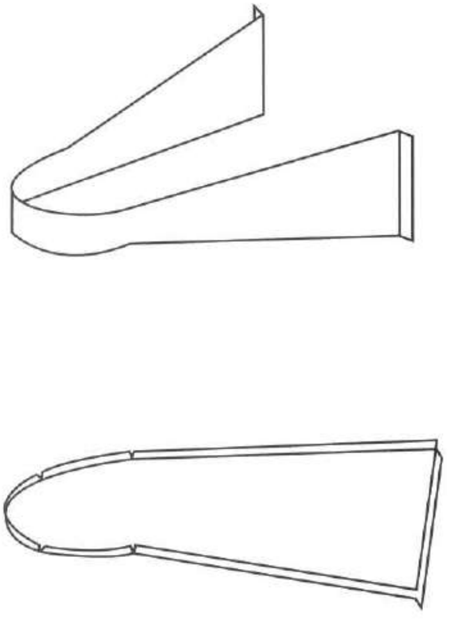

Figure 13-7 *Copper side with top and bottom pieces*

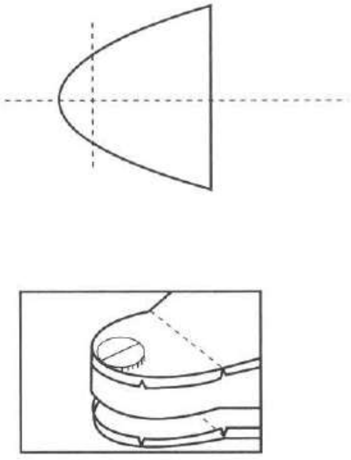

Figure 13-8 *Position of aperture hole for magnetron probe*

inches. θ is the angle between the z axis and the top (or bottom) side of the horn.

Let us now examine the efficiency of the parabolic sides of the horn for their focus of the magnetic field. The antenna is circumscribed by a magnetic field that propagates from a line source at the center of the antenna. The radiation either reflects off the sides of the parabola formed by the horn or it radiates directly from the antenna out the mouth of the horn. The latter is not useful because it dissipates rapidly with distance from the cannon. The former, the reflected radiation, is projected parallel to the zy plane, and because it originates from a line source parallel to the y axis, no diffraction occurs off the sides of the horn's mouth. The fraction of the radiation that is reflected is 92.2 percent of the total from the antenna. This is far superior to the typical 40 percent efficiency given by radar dishes that use large sizes to achieve gain.

The focusing of the electric field can now be performed by placing a lens in front of the cannon's mouth, and a description of the electric field must be provided. On the surface of a conductor there can be no voltage parallel to the surface of that conductor, because any voltage would be short-circuited out. This makes the components of the gradient of any voltage parallel to that surface zero. Since the electric field is the negative gradient of the voltage, the electric field near a conducting surface must perforce be perpendicular to that surface. As the electric field propagates forward, it forms a cylindrical-shaped wavefront in the flared part of the horn. At the mouth of the horn, this cylinder has a radius of 41 inches. A lens will convert this cylindrical wavefront to a flat wavefront, and that will complete the focusing of the radiation.

The lens will be constructed of metal plates that will be aligned parallel to the y and z axis 4 inches apart. Between these plates, any wave vector, $k = 2\mathrm{n}/\pi$, can be the vector sum of two components, kx and kz. In the x direction, the sine wave of the wave vector will be zero every 4 inches, because at the plates the electric field in the y direction must be zero. Taking recourse to the Pythagorean theorem, the wavelength in the z direction is dilated by the inverse of the refraction index, that is, the refraction index, $n = [1-(\lambda/8 \text{ inches})E2]E-.5 = 0.8$ ($\lambda = 4.8$ inches). See Figure 13-9.

The cylinder formed by the cylindrical-shaped wavefront plus the refraction index times the distance parallel to the z axis between the lens elements to a flat plane in front of the lens is to be constant so that the electric field will be in phase on the flat front of the lens. The y axis is shifted to the most concave depth of the lens elements, and the z axis is directed rearward. The algebra for calculating the concave curve to cut into the rear of the lens plates is rudimentary. From the following equation, the dimensions of the lens are shown in Figure 13-10.

[(47" - z)E2 + y]E.5 + 0.8 z = 47" from which z =
{18.8 - [(18.8)E2 - (1.44y)E2] E.5} {1/.72}

The lens plates mounted against the far right and far left sides of the lens frames should be cut from leftover sheet copper and glued to wooden sides, also cut to the shape of the lens element. The other lens elements should be made from 0.025-inch sheet aluminum. They should be 21 inches long in order to mount them to the wood lens frame you will build. Small wooden blocks can hold the ends of lens plates in place. Pieces of wood with grooves cut into them will also work.

The efficiency of the lens can also be calculated. This efficiency is due to the reflection of radiation that takes place when radiation passes from a medium of one refraction index into a medium of a different refraction index where the space impedance is changed. Two reflections occur, one when the radiation enters the rear of the lens and a second reflection when the radiation exits the front of the lens. The second reflection is in opposite phase with the first reflection, so it is 2π radians out of phase with the first reflection. The appropriate Fresnel equation for calculating the power reflected when the electric field

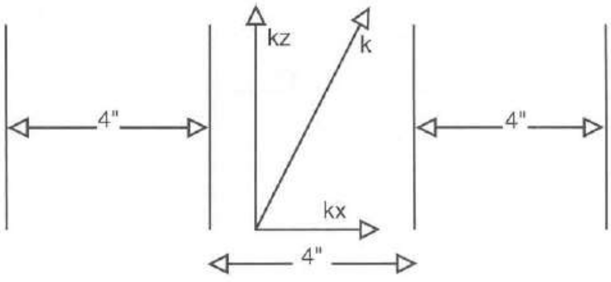

Figure 13-9 *Focus plate geometry*

is parallel to the plane containing the angle of incidence, θi, and the angle of transmission, θt, is as follows:

$$R = \{[-\cos\theta i + (n1/n2)\cos\theta t] \div [\cos\theta i + (n1/n2)\cos\theta t]\}E2$$

θt is arctan dz/dy, as can be seen in Figure 13-11. θi is $\theta t - \arctan [y/(47 - z)]$.

In the second reflection, θi and θt are both zero. n1 is 1 and nt is 0.8. The second reflection that must be subtracted from the first one is $\cos [2\pi \text{ radians } 2(z + .25'')\lambda z]T$, where T is the fraction of the power transmitted into the lens and λz is 4.8 inches/0.8. Since the total reflected power is almost everywhere less than a percent, we can make the approximation that T equals 1. If the reflected power is calculated every inch for y, the average power reflected is approximately half a percent. Thus, the gain of the horn is

$$4\pi A/\Lambda E2 = 218 \text{ or } 23.4 \text{ db}$$

With the ideal horn and lens system having been designed so that the common man can build it without having to know all the electromagnetic laws of physics that make it work, we direct our attention to the choice of a microwave source. The cardinal rule of electronic warfare is that the radiation must convey noise in order to disrupt the signals in the circuits you will be attacking. This requires modulation that will drown out enough of the unwanted music and interfere with the digital signals in disc players to make them useless. Half-wave rectification of the current through the radiation generator is subjectively the most disrupting. The lower frequencies promoted by stereo manufacturers and music producers seem to be designed to most effectively penetrate the walls of apartments and homes for the purpose of destroying the privacy and cultural autonomy of all but the wealthiest who can afford the enormous cost of solitude.

The 60-cycle voltage from a household wall socket offers a nearly ideal source of modulation to create this disruption, because in a microwave oven it is half-wave rectified by a voltage-doubling circuit. This provides a splendid spectrum of Fourier frequencies that concentrate upon the lower frequencies but distribute themselves upward through the audio range, and it gives the utmost modulation of amplitude. It also reduces the duty cycle of the magnetron while preserving the maximum power in each pulse it shoots into the neighbor's stereo for the best range.

The first inclination is to choose a magnetron that will have the highest output power. The best choice would be a radar magnetron, which might offer 30 to 100 kilowatts in pulses that are only a microsecond or two in duration. However, tests using radar against a compact disc player yield disappointing results. The Fourier series expansion on a series of microsecond-long pulses modulates less than 1 percent in the audio band. The series of ticking sounds that it introduces into a stereo amplifier is scarcely noticeable, and not enough pulses occur to put sufficient errors in the digital circuits of the compact disc player to do any good. Also, audio circuits are not as sensitive to the

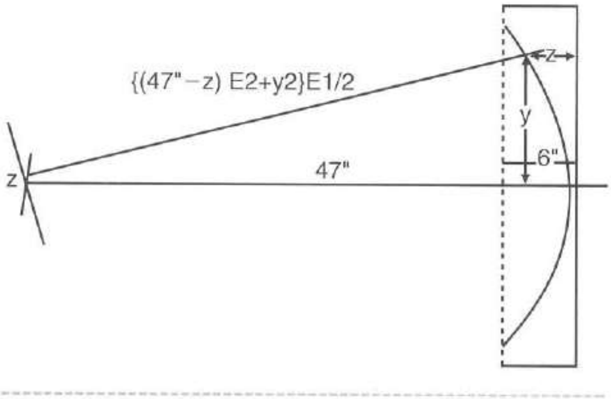

Figure 13-10 *Wavefront and electric lens geometry*

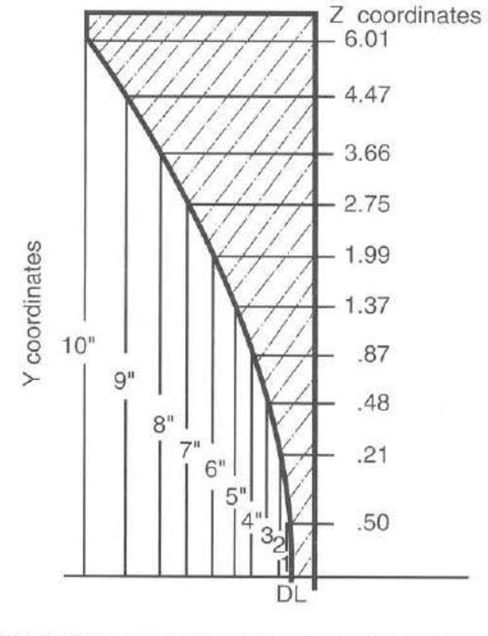

Figure 13-11 *Calculated lens dimension*

higher frequencies of radar as they are to microwave oven frequencies.

We might also be inclined to consider the higher-powered magnetrons used in industrial heaters and plasma generators, but this can be problematic. First, higher power is achieved more by increasing the duty cycle than by increasing the peak output power we desire. The 3,000-watt Panasonic 2M265 magnetron has a 7-kilowatt peak output power, but it runs on pulse-modulated direct current for a longer duty cycle than the half-wave voltage doubler of a microwave oven. Its operation is more critical; the cathode heater must have its voltage adjusted to the power output, and the cathode is more brittle in these more powerful magnetrons. It also costs as much as $1,000 and the power supply is even more expensive. The 6,000-watt Toshiba E3328 magnetron runs on filtered direct current and it has even more expensive computer controls that adjust an electromagnet to control the current. The peak output power is less than that of the 3,000-watt magnetron.

Another problem occurs with these magnetrons as well. They are so unforgiving of such things as a large metal object near the cannon in which they might be installed that reflections could mismatch the impedance enough to shorten the service life. The manufacturers recommend that they have a branch off of a waveguide to a dummy load that will absorb any energy due to such an impedance mismatch in order to protect the magnetron. This makes the parabolic horn impossible to use because it is not practical to feed with a waveguide. A pyramidal horn would be needed and it only has two-thirds of the gain of the parabolic horn. Thus, the 6,000-watt magnetron would be as good as a 4,000-watt one in a parabolic horn, and the 3,000-watt magnetron would be as good as two-thirds of 7,000 watts in a parabolic horn. Several vastly cheaper household oven magnetrons can easily outclass the more expensive and difficult-to-operate choices that appear more attractive until they are examined with greater perspicacity.

Another difficulty with purchasing industrial magnetrons also exists. Companies will not sell them to private individuals out of fear that they might be used for negative or hateful enterprises, such as in machines designed to sabotage overly loud stereos. Although almost any magnetron and power supply from any microwave oven will work, the best oven magnetron for our purposes is the 2M121A, which has its number conflated with the number 53 or 57. These last two numbers make no difference in the choice of editions of the 2M121A you procure. They are used in the following ovens: the Panasonic 1030, the Brother MF5000 and MF7000, and the Merrichef models 136M, 165M, and 206M. It has a peak output power of 5,800 watts, better than the 2,000-watt magnetrons that cost more than twice as much. It can be ordered from the Expert Appliance site (www.expertappliance.com) under the part number Z9-Panasonic NEI 030-412 if you type in Panasonic NEI 030 for the type of oven for which you are purchasing parts. The transformer number is 210 Panasonic NEI 030-5414. Global Microwave Parts has voltage-doubling capacitors that are 0.85 microfarads rated at 2,500 volts AC and diodes. You should use one rated at half an ampere. Magnetrons and parts are available and listed in Table 13-2. You will need copper and aluminum sheeting, along with mat-erials available from hardware or building suppliers.

The magnetron is attached to the antenna, as shown in Figure 13-12. Follow the instructions closely and observe that the brass washer is flush with the copper conductive surface.

The power supply is a simple voltage-doubler circuit, shown in Figure 13-13. It must be understood that the voltages in the wires will be in the thousands and can produce a spark that can jump through the air and through an insulator, conveying enough current that it would probably cause electrocution. Accordingly, the wires should be positioned where they will not be approached or touched. The magnetron and horn must be securely grounded or it will be 4,500 volts above ground potential. The frame that you must build around it will keep people's hands away from these wires. Also, after the cannon is shut off, four of the capacitors can hold a charge that would inflict a shock unless it is discharged, either through a bleeder resistor that may be built into the capacitor or it can be discharged with a safety discharge probe.

The magnetron must also have a fan to cool it during operation. The 2M121A requires at least 85 cubic feet per minute of air directed through its fins for best operation.

When your microwave cannon is completed, testing it is little more than a matter of switching on the power and aiming it at a fluorescent tube or a small 120-volt lightbulb. The lightbulb should be no more than 25 watts and should glow slightly when it is within a few feet of the mouth of the cannon. Fluorescent lamp tubes should light up as far as 10 feet away.

You should not put any body parts directly in front of the cannon at close range because, after a few seconds to a few minutes, you could burn yourself, but a distance of over 50 feet away would provide safety. However, prolonged overheating of internal organ tissue is remotely possible and the cornea of the eye can be damaged by overheating. Also, testicles can overheat, and at close enough range for the *root mean square* (RMS) microwave field to exceed 4,000 watts per square meter, a 5-minute uninterrupted exposure is supposed to build up enough internal steam pressure to cause the testicles to explode. However, the possibility of having a hot spot this intense a few inches in front of the cannon is marginal.

When testing or using your microwave cannon, you can do so with utmost safety to yourself by constructing a gown out of metal screen from a hardware store. The gown should include a hood that completely covers the head and should extend over the torso down to just below the vaudevillian parts of the body. The sex organs should be stuffed into a tin can, the mouth of which can be lined with fake fur to make it more comfortable. (Obviously, the ladies need not take recourse to this expedient.) This is more than enough to prevent any harm when performing experiments that last half an hour or more and is probably hundreds of times the protection you will need.

It can be advantageous to test the cannon in your backyard. Sooner or later the neighbors will see it, and if they see a mysterious-looking machine that lights up fluorescent tubes 10 or more feet away and they see you dressed in a wire screen gown complete with a tin can, it will frighten the emunctory indiscretions out of them. Further, when they complain to the authorities they will sound much the same way mental patients sound as they describe exotic macrunes, stereos malfunctioning, strange costumes complete with tin cans, and the like. It is even slightly possible

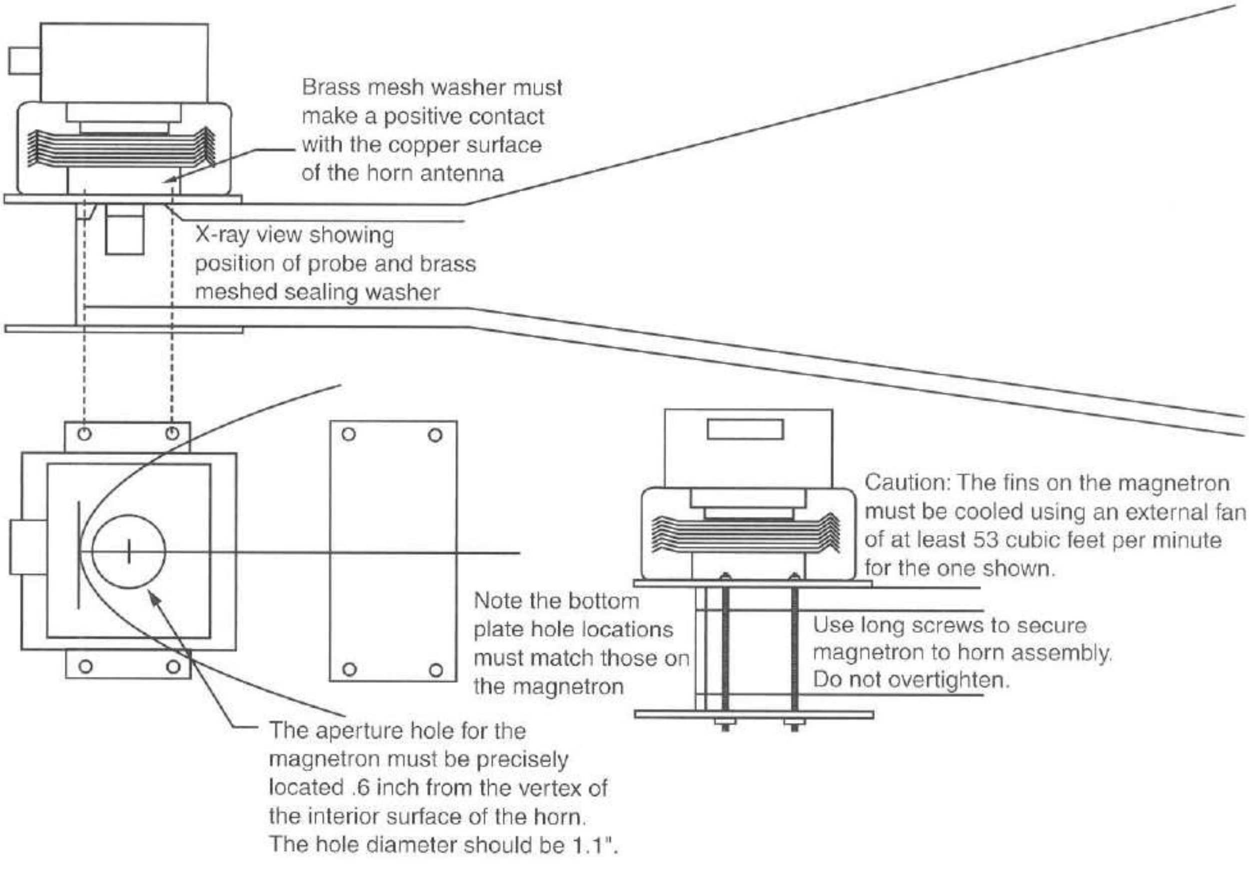
Brass mesh washer must make a positive contact with the copper surface of the horn antenna
X-ray view showing position of probe and brass meshed sealing washer
Caution: The fins on the magnetron must be cooled using an external fan of at least 53 cubic feet per minute for the one shown.
Note the bottom plate hole locations must match those on the magnetron
Use long screws to secure magnetron to horn assembly. Do not overtighten.
The aperture hole for the magnetron must be precisely located .6 inch from the vertex of the interior surface of the horn. The hole diameter should be 1.1".

Figure 13-12 *Attaching the magnetron*

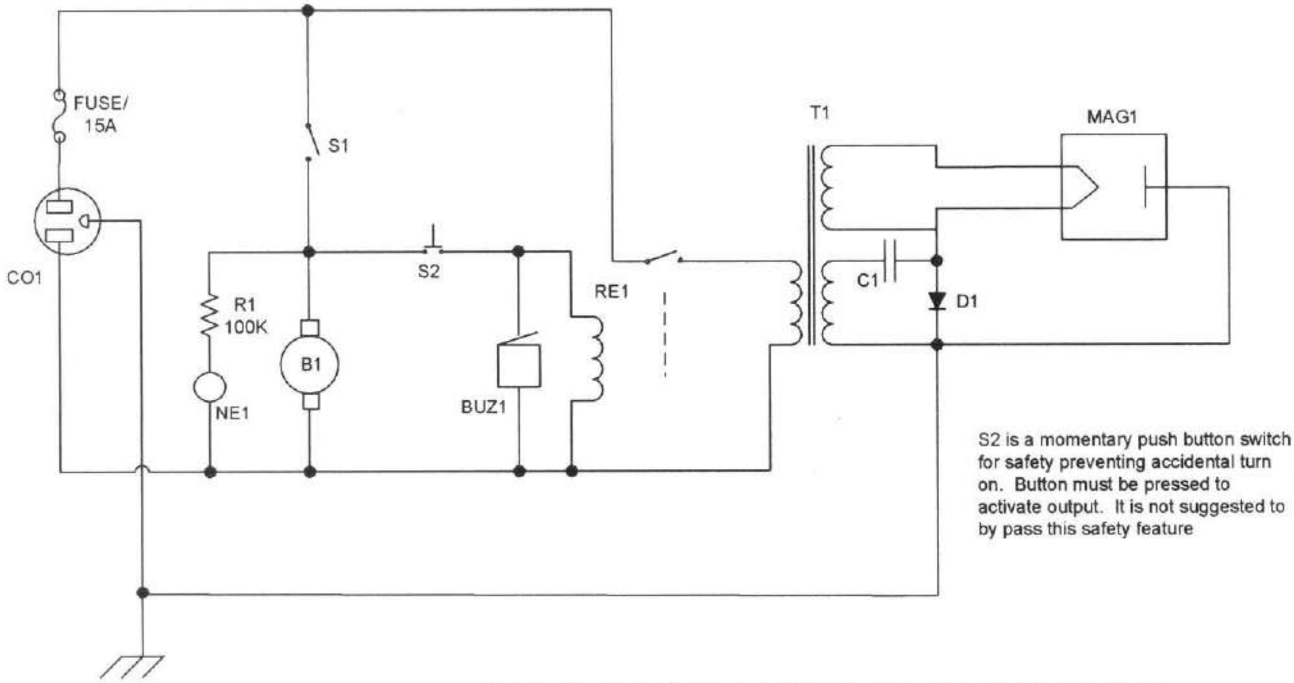

S2 is a momentary push button switch for safety preventing accidental turn on. Button must be pressed to activate output. It is not suggested to by pass this safety feature

Indicates grounding of antenna and and frame pieces

DANGER The metal antenna and all frame structure must positively be grounded to the earth via the third green lead of the power cord. Failure to do this will result is a very dangerous electrical hazard!

Connection leads indicated by the heavier lines must support current levels of 25 amps.

Operation requires switch S1 powering the neon indicator lamp NE1 and the cooling blower B1 for the magnetron. It also arms the push button switch S2.
The magnetron is powered when S2 is pressed along with a warning buzzer BUZ1 by energizing primary contactor RE1

Figure 13-13 *Microwave cannon schematic*

that they might be invited to undergo observation at the local state psychiatric hospital, where, if you are really ambitious, you could locate their room, aim your microwave cannon, and make their stereo buzz whenever the psychiatrist's back is turned and cause a diagnosis that will keep them out of the neighborhood for a longer time. But in the remote case you get caught doing this, it will be almost impossible to prosecute you because no state legislature has the time to write and process special laws against shooting microwaves into peoples' stereos. The television publicity possible if the state attempts to prosecute you will be a disaster for the system that supports cultural tyranny because it will inspire countless television viewers to join in the revolution to make the world a better place by building more microwave cannons and shooting down more obnoxious stereos. And even if they can win a battle in court, they will ultimately lose the war that they started against human dignity.

Table 13-2 Microwave Cannon Parts List

Ref. #	*Description*	*DB Part #*
R1	100K 1/4-watt resistor (br-blk-yel)	
CO1	Heavy-duty, 15-amp, three-wire power cord	
FUSE/15	15-amp fuse and holder	
B1	85 cubic feet/minute or better cooling fan for magnetron fins	
S1	*Single pole, single throw* (SPST), 5-amp, 115 VAC toggle switch	
S2	Momentary push-button switch (see note on Figure 13-13)	
NE1	Neon indicator lamp with leads or equivalent	
BUZ1	120 VAC high output safety buzzer	
RE1	10- to 20-amp contacter or relay	
T1	Magnetron power transformer NEI030-5414	DB# MAGTRAN
C1	.85-microfarad, 2500 VAC voltage-doubler capacitor	DB# MAGCAP
D1	5 kilovolt, 1/2-amp diode	DB# MAGDIODE
MAG1	2M121A magnetron	DB# MAGNETRON

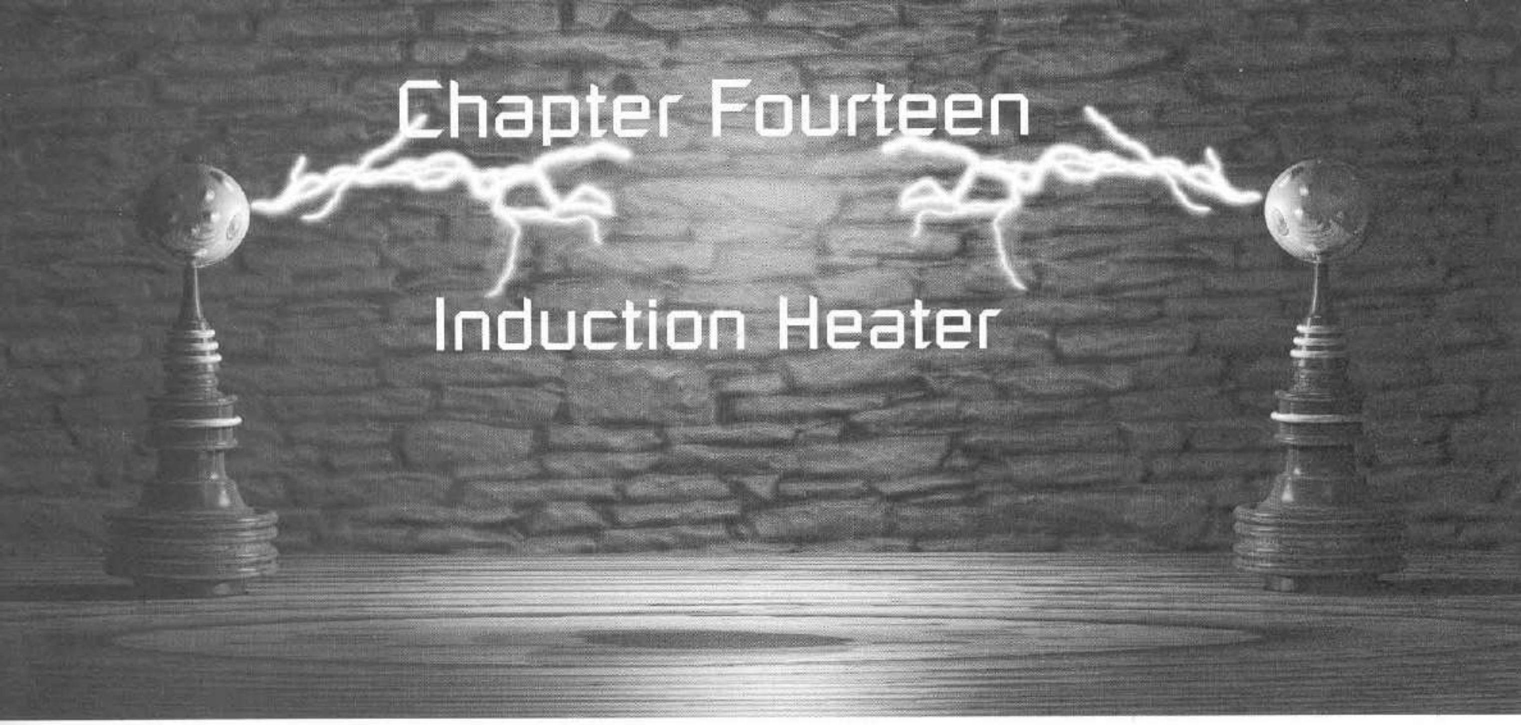

Chapter Fourteen

Induction Heater

This device shown in Figure 14-1 shows an induction heater capable of heating metals to over 1,500 degrees Fahrenheit. It can be used in the processing of gas discharge tubes often used for plasma, neon displays, laser discharge tubes, and spectrum tubes. These tubes usually require electrodes that must be heated to drive out the impurities that are now sucked out by a vacuum system. The induction heater heats these electrodes directly through the glass.

The completed system consists of a donut-shaped heating coil head connected via an umbilical cord, which is in turn connected to the power-conditioning and control box. The system is powered directly from 115 VAC and requires caution when building and using. The head is a coil of wire that surrounds the target area to be heated. As shown, it is an air-cooled system that requires a waiting time for processing between operations because the coil will get hot. The common circuit ground is above ground and requires an isolation transformer or an ungrounded scope to view the waveforms from a grounded scope.

This is an advanced-level project requiring electronic skills and the use of an oscilloscope. Expect to spend $35 to $75. All the parts are readily available, with specialized parts obtainable through Information Unlimited (www.amazing1.com), and they are listed in Table 14-1.

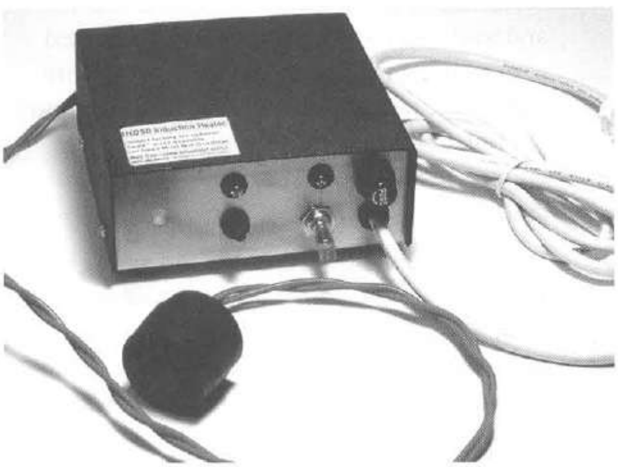

Figure 14-1 *Induction heater*

Circuit Operation

The power cord CO1 supplies 115/220 VAC from a wall receptacle where the third, green wire, grounding lead must be securely attached to the metal chassis (CHASSIS) for safety (see Figure 14-2).

Switch SW1 energizes the primary circuits, and fuse FUS1 protects the main circuit from any catastrophic faults. Indicator lamp NEON1 requires the current-limiting resistor R12 and lights up when SW1 is turned on. The peak charging current to capacitors C10 and C11 is limited by in-rush resistor Rx. The rectifiers D1, D2, D3, and D4 are shown in a full-bridge configuration when used for 220 VAC input. For 115 VAC operation, only D1 and D3 are used along with the JUMP jumper in a voltage-doubler configuration. The output voltage across C10 and C11 is 300 volts DC and is referred to as the positive and negative rails.

Dropping resistors R2 and R3 provide voltage to the oscillator driver IC1 with zener diode Z1 regulating the voltage at 15 volts. Switch SW2 is normally closed, shorting out this voltage across Z1. Capacitor C2 provides the high instant of current for the output pulses and must be physically close to IC1. Operation requires SW2 being pressed, allowing 15 volts to be applied to the oscillator driver generating the drive pulses.

These drive pulses turn the main MOSFET switches, Q1 and Q2, off and on. The switching frequency is determined by the time constant of power control pot R10 and timing capacitor C3. A limit resistor, R11, prevents driving the circuit beyond the intended limits.

The switching circuit is in a half-bridge configuration where the MOSFET connected to the positive end (+ rail) must be driven with its source pin referenced at 150 volts above the common end (− rail). This is accomplished by biasing a bootstrap capacitor (C4) through an ultrafast diode (D5), providing the correct DC level to fully control Q1. Resistors R6 and R7 eliminate the high-frequency parasitic oscillation that occurs as a result of rapidly switching the capacitive load of the MOSFET gates. Capacitor C40 limits the current flow to a second neon indicator light, NEON2, which is connected in series and is energized by the output frequency when the heating cycle is activated by SW2.

A network consisting of capacitor C7 and resistor R8 slows down the transition time of the switched pulses across Q1 and Q2. The resultant time constant limits the rate of voltage rise, dv/dt, that could cause a premature turn-on at the wrong switch, creating a catastrophic fault mode.

Capacitors C5 and C6 provide a voltage midpoint and produce the necessary storage energy to maintain the voltage level of the individual pulses. The output is taken at the junction of Q1/Q2 and C5/C6 and is a square wave fed to the induction heating coil L1. A current rise through L1 now commences as a function of Et/L where E is the voltage of the drive pulse, t is the applied time, and L is the inductance of L1. This changing current now induces eddy currents in the target piece to be heated as a function of E = L di/dt. These circulating currents quickly heat up the target piece.

Project Assembly

1. Assemble the heating head coil, as shown in Figure 14-3. This coil must be tightly and evenly wound for optimal performance. It may be difficult for the beginner and is available in three sizes, as indicated in Table 14-1.
2. Assemble the circuit board, as shown in Figure 14-4. Note that you should allow a half-inch for the Q1 and Q2 pins. This is necessary to allow these parts to be placed flat against the chassis when mounting as they exit at the midsection of the part.

   It is also strongly suggested that you use an eight-pin IC socket (SO8) because IC1 is very sensitive to static electricity and possible errors during the assembly.

   Note that if you are building from a perforated or vector circuit board, it is suggested that you use the indicated traces for the wire runs and insert components starting in the lower left-hand corner. Pay attention to the polarity of capacitors with polarity signs and all semiconductors.

   Route the leads of the components as shown and solder as you go, cutting away the unused wires. Attempt to use certain leads as the wire runs or use pieces of the #24 bus wire. The heavy foil runs should use the thicker #20 bus wire because these are the high-current discharge paths.
3. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.

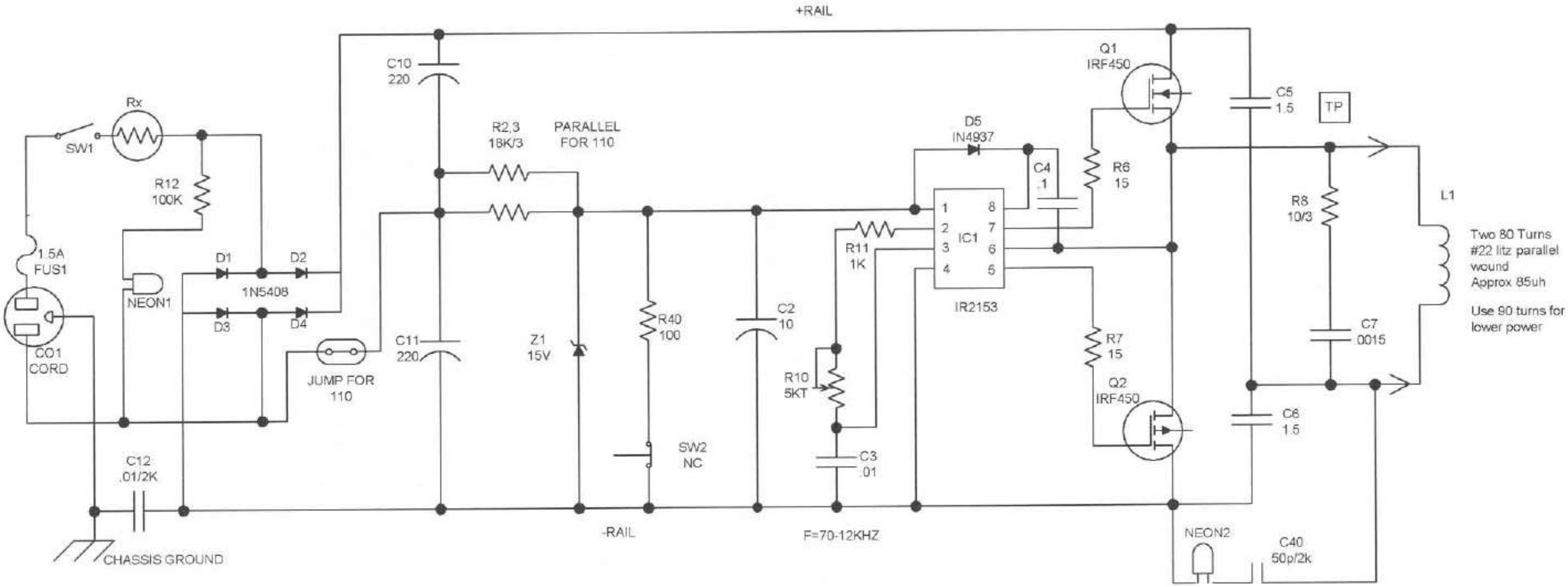
+RAIL
C10
220
SW1
Rx
R12
100K
1.5A
FUS1
CO1
CORD
NEON1
D1
D2
1N5408
D3
D4
C12
.01/2K
CHASSIS GROUND
R2,3
18K/3
PARALLEL
FOR 110
C11
220
JUMP FOR
110
Z1
15V
R40
100
SW2
NC
-RAIL
C2
10
R10
5KT
C3
.01
F=70-12KHZ
R11
1K
D5
IN4937
C4
.1
IC1
IR2153
R6
15
R7
15
Q1
IRF450
Q2
IRF450
C5
1.5
C6
1.5
TP
R8
10/3
C7
.0015
L1
Two 80 Turns
#22 litz parallel
wound
Approx 85uh
Use 90 turns for
lower power
NEON2
C40
50p/2k

Figure 14-2 *Induction heater schematic*

4. Connect the external components as shown in Figure 14-5. Use 6-inch lengths of the wire size as noted.
5. Create the chassis, as shown in Figure 14-6, from a piece of .065-inch 5052 bendable aluminum. Note to trial-fit and measure the diameters of the components before drilling the holes. Pay attention to the holes for mounting Q1 and Q2 because these must be reasonably accurate to align with the layout of the PCB. Also, there must be at least 1/8 inch of clearance between the board and the chassis for sandwiching in the PLASTPL plastic plate. This prevents grounding out of the board wiring traces to the metal chassis.
6. The final assembly is shown in Figure 14-7. Note the mounting scheme for Q1 and Q2, as they must be insulated from the chassis using the thermo pads, THERM1. These pads allow heat transfers and provide electrical insulation from the metal chassis.

   Drill the hole for grounding lug LUG1 and solder the green grounding lead from the power cord and circuit board's earth ground.

The heating head coil assembly can be built to accommodate a diameter from a 1/2" to a 1" object. The assembly with the material shown is intended for using enclosed electrodes in glass tubes such as those used in neon tubes, laser tubes, and spectrum tubes. High-temperature materials may be used if you are anticipating using for other applictions.

The coil should be wound with high-frequency wire to limit heating losses. Use #22 litz wire consisting of multiple strands of #36 magnet wire.

1. Fabricate the winding bobbin from 1.25" lengths of 1/16" wall GT fiber glass tubing of your diameter choice of 1/2, 5/8, 3/4, 7/8 or 1".
2. Wind an even 1" layer of the #22 litz wire. Glue the start wire with a fast-curing glue. Tape the finished winding using a single layer of high-temperature glass tape.
3. Continue winding layers and repeat step 2 until there are 80 turns. You may truncate the successive layers by several turns at each end to keep from slipping off or use a quick-curing glue.
4. Attach the input leads and attach by gluing and taping into place.

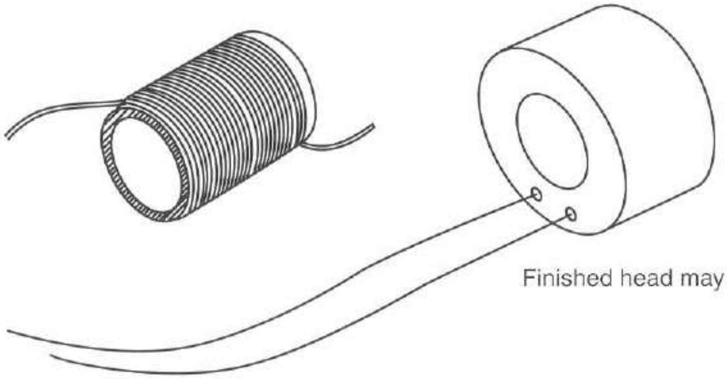

Finished head may be potted in epoxy or sealed in plastic caps

The inductances of the various heads will vary as a function of the diameter. Ideal current through the head is 20 amps peak. You should therefore adjust the frequency of R10 to maintain this value when changing the head sizes as the following:

1/2" head	50 uh	75 KHz
5/8" head	75 uh	50 KHz
3/4" head	100 uh	40 KHz
7/8" head	125 uh	35 KHz
1" head	150 uh	30 KHz

These heads are available assembled for those not wanting to attempt winding their own. They are listed in Table 14-1.

Figure 14-3 *Heater head coil assembly*

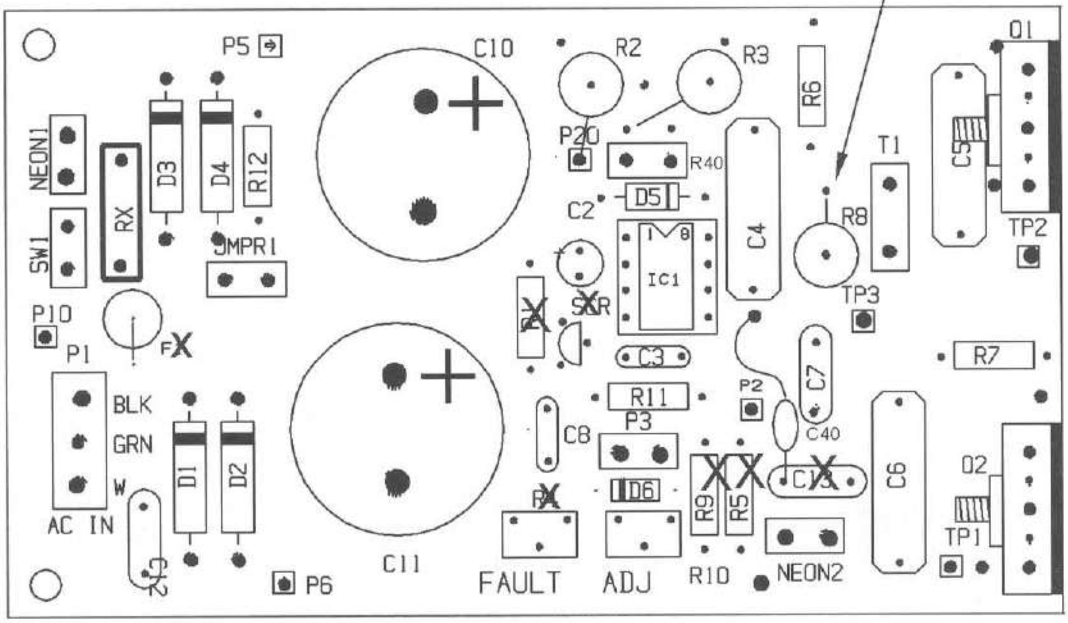
This junction is TP test point

P5
C10
R2
R3
R6
Q1
NEON1
D3
D4
R12
RX
P20
R40
T1
C5
C2
D5
C4
R8
SW1
JMPR1
IC1
TP2
P10
P1
R4
SCR
F
X
TP3
C3
R7
BLK
GRN
W
AC IN
D1
D2
R11
P2
C7
P3
C8
C40
R9
R5
C1
C6
Q2
TP1
R3
P6
C11
FAULT
ADJ
R10
NEON2

1. Black dots are connections points to external leads as shown figure 14-5

2. The PC board as shown is used in several projects and does not use all the components as indicated. Those indicated by an X are omitted for this project. R40 and C40 are added to the pads as shown and are not identified on the PC board printing

3. This circuit uses a 8 pin socket for IC1 as it is prone to failing if there are circuit faults

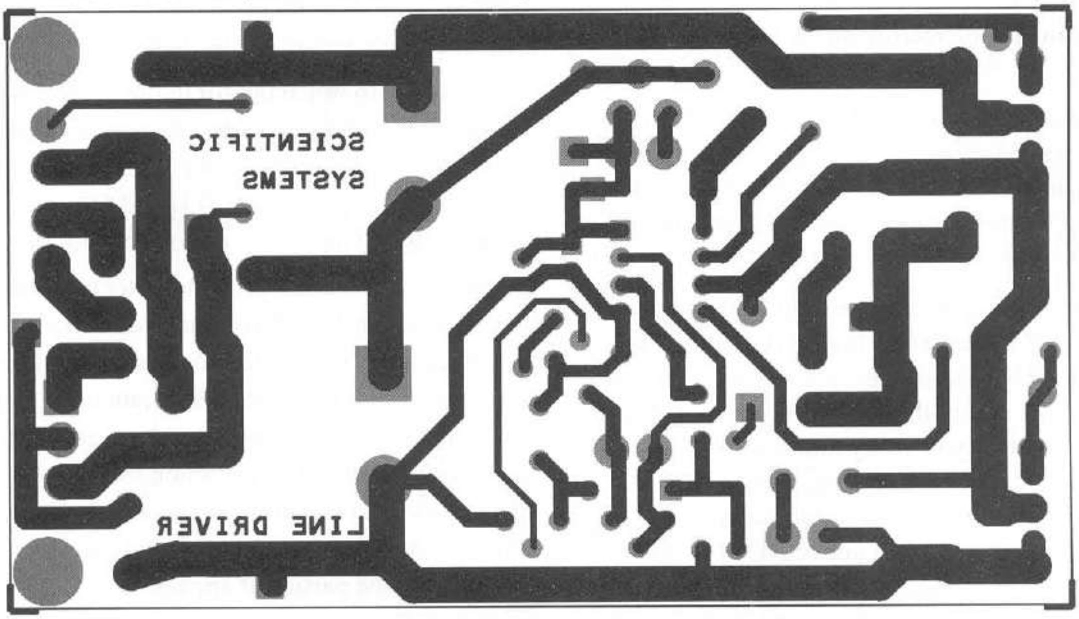
SCIENTIFIC
SYSTEMS
LINE DRIVER

JUMP

Assembly of the board assembly is shown using a printed circuit board available through www.amazing1.com Builders may use a vector board and use the component leads to duplicate the foil traces as shown.

On all IND40 assembly boards this jump lead is to be the first part on the board. Once the unit is final assembled with the heating head connected in place the jump now can now be disconnected. This step prevents static electric damage to the semiconductors.

Figure 14-4 *Assembly of the circuit board*

Note that we have shown the use of plastic bushings (BU38) for the neon indicator lamps and output leads, and also BU58CL clamp bushing for the line cord.

7. Assemble the test jig as shown in Figure 14-8. This is a very useful fixture for any 115 VAC project up to 300 watts.
8. Now fabricate the cover to fit the chassis section. Use bendable plastic or a thin, perforated aluminum sheet. Secure it to the chassis via four screws (SW1) on the folded-up lip in this section.

Project Testing

Caution: The circuit commonly noted as −RAIL is electrically above earthground by over 150 volts. It is strongly recommended that you obtain a 200-watt isolation transformer before proceeding and testing this or other similar direct line powered circuits. Experienced experimenters working on dry, wooden floors may choose to unground their test equipment. This can be a dangerous shock hazard!

1. Connect the ungrounded scope to the TP test point, as noted in the Figure 14-2 schematic or in Figure 14-4. The lead of R8 is shown connected to the junction of the drain and the source of Q1 and Q2 respectively.
2. Verify that the main power switch S1 is off and insert a 2-amp fuse into FH1. Then adjust trimpot R10 to midrange and connect CO1 to the female end of the cord on the test jig. It is important that you verify that the test jig is not shorting the ballast lamp. Plug the jig into the output of the isolation transformer. Quickly turn on S1 and notice NEON1 lighting up and the ballast lamp not lighting. If it lights, you must troubleshoot and find your error before proceeding.

Press and hold PB1, noting a picture-perfect square wave on the scope with the lamp only dimly lighting. Quickly adjust R10 for a 20-usec pulse period. This corresponds to 50 kHz. NEON2 will indicate when PB1 is activated. Experienced builders may wish to monitor the current to the head using a current probe attachment to the scope. This wave shape should be another picture-perfect sawtooth of 20 amps maximum.

Using the Heater Project

Obtain a gas discharge tube with an electrode to fit into the head size you chose in assembly.

At this point, it is a good idea to have a meter capable of measuring the AC line current, as this value is very dependent on the object that you place into the head and should not exceed 2 amps. You can drop this value by increasing the frequency, which is done by adjusting R10. When heating electrodes, you will note the ampmeter dropping in level when the electrode is red hot. This is due to the change in the magnetic characteristics of the heated material.

Something to watch out for in this project is overheating the head coil. You may implement a cooling fan stand when using the device for multiple tasks. Also, the chassis panel where Q1 and Q2 are mounted should not be uncomfortably hot to touch.

Those experienced can create other heads for different heating tasks and you can even make a "quick" heat solder pot. Just make sure you keep the switching current below the maximum of 20 amps. The inductance of the head should not go below 80 uh. Alternate head construction must always take into consideration heating fatigue and the eventual failure of the head wires, bobbin, and construction materials for the particular application.

Notes:
1. The parts on the printed circuit board with an X are not used for this project.
2. C40 and R40 are added components not indicated on the board
3. Lead lengths can be shortened in final assembly

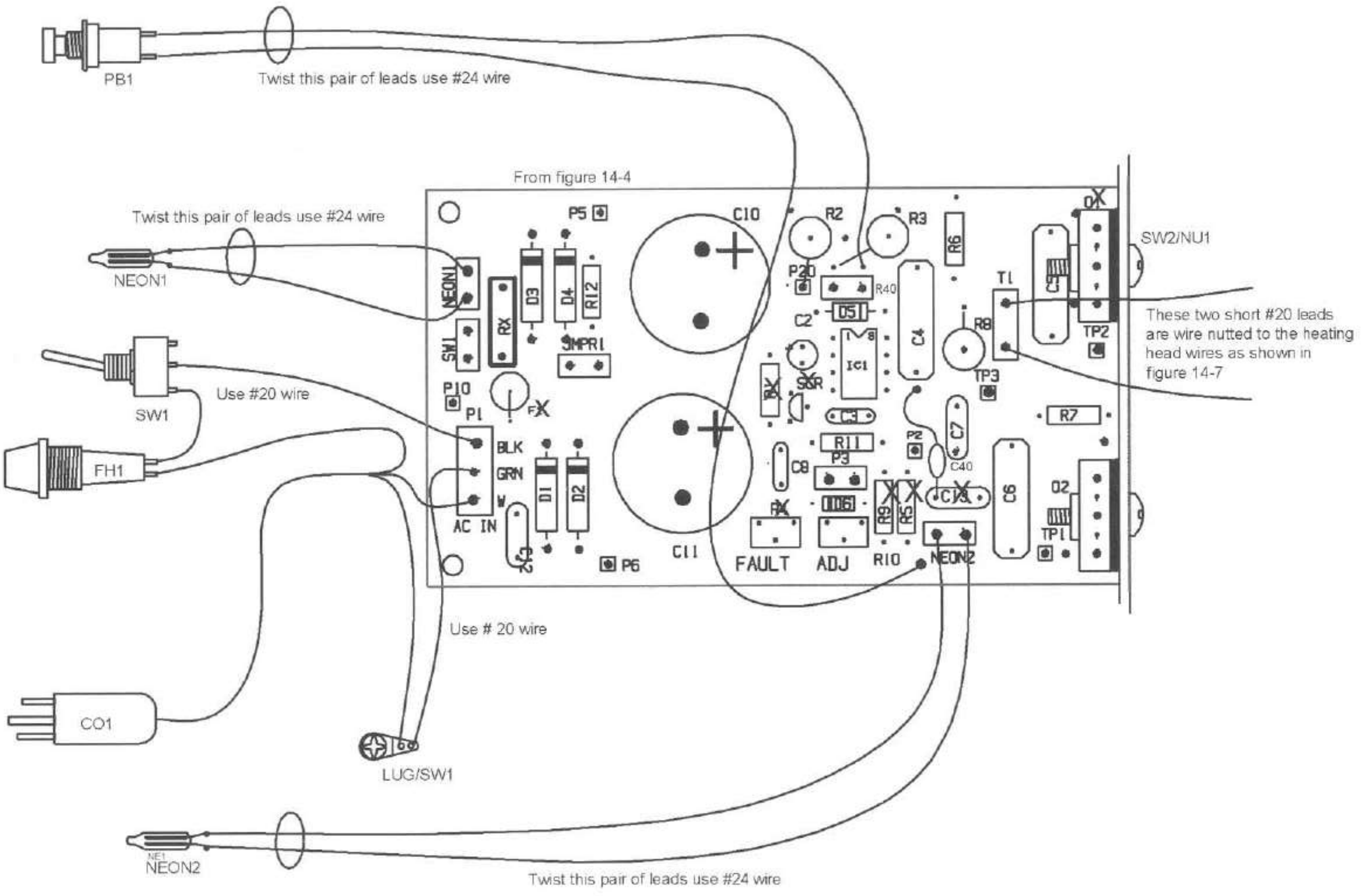
PB1
Twist this pair of leads use #24 wire
From figure 14-4
Twist this pair of leads use #24 wire
NEON1
SW1
Use #20 wire
FH1
CO1
LUG/SW1
Use # 20 wire
NE1
NEON2
Twist this pair of leads use #24 wire
SW2/NU1
These two short #20 leads are wire nutted to the heating head wires as shown in figure 14-7
P5
C10
R2
R3
R6
T1
TP2
TP3
TP1
R7
C5
C6
C4
C7
C40
R8
R40
D5
IC1
C2
P20
P2
P3
C3
R11
C8
D6
R9
R5
C1
Q1
Q2
R10
NEON2
FAULT
ADJ
C11
P6
D1
D2
C12
AC IN
BLK
GRN
W
P1
P10
SW1
NEON1
RX
D3
D4
R12
JMPR1
FX

Figure 14-5 *External wiring*

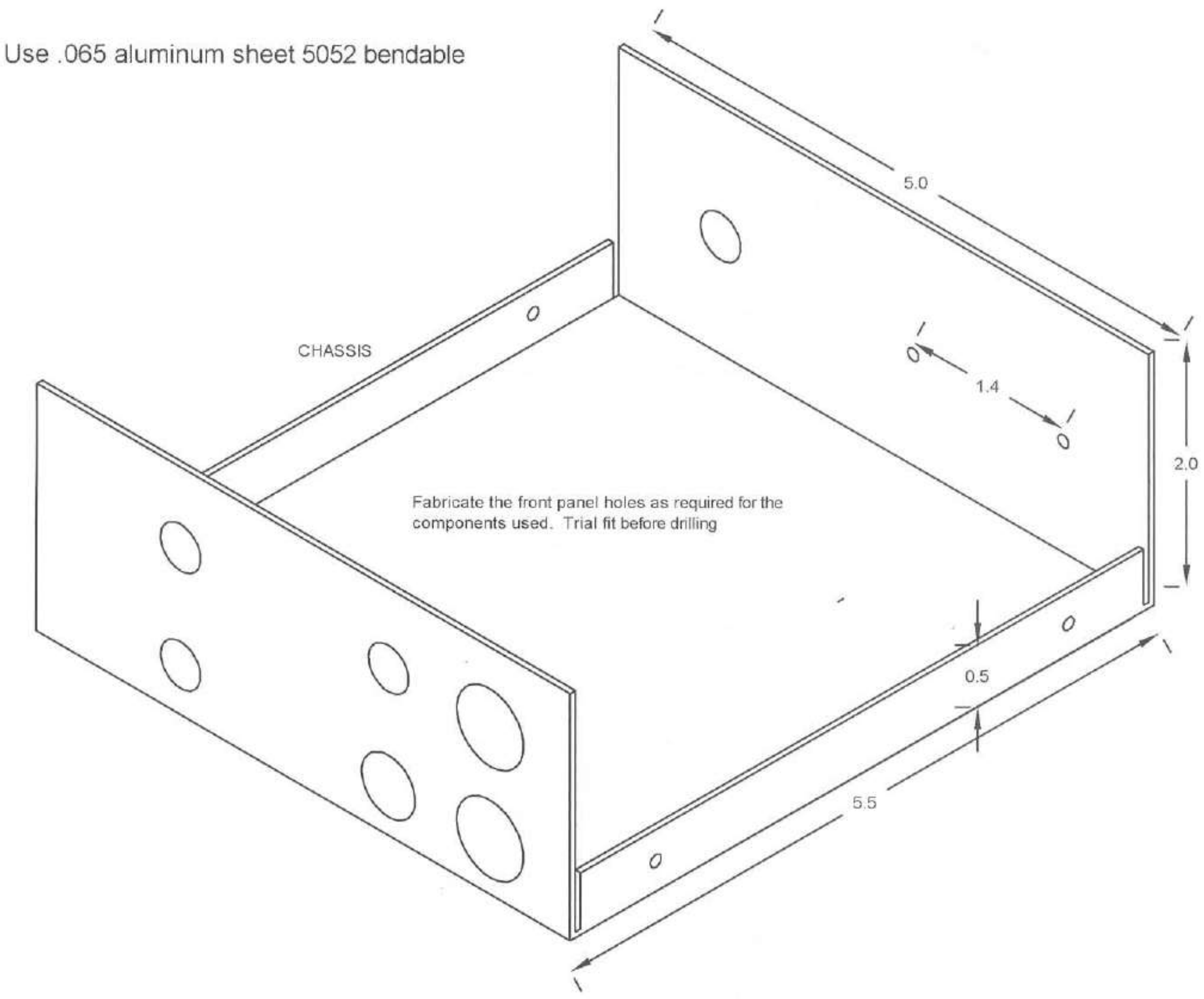
Use .065 aluminum sheet 5052 bendable
5.0
CHASSIS
1.4
2.0
Fabricate the front panel holes as required for the components used. Trial fit before drilling
0.5
5.5

Figure 14-6 *Chassis fabrication*

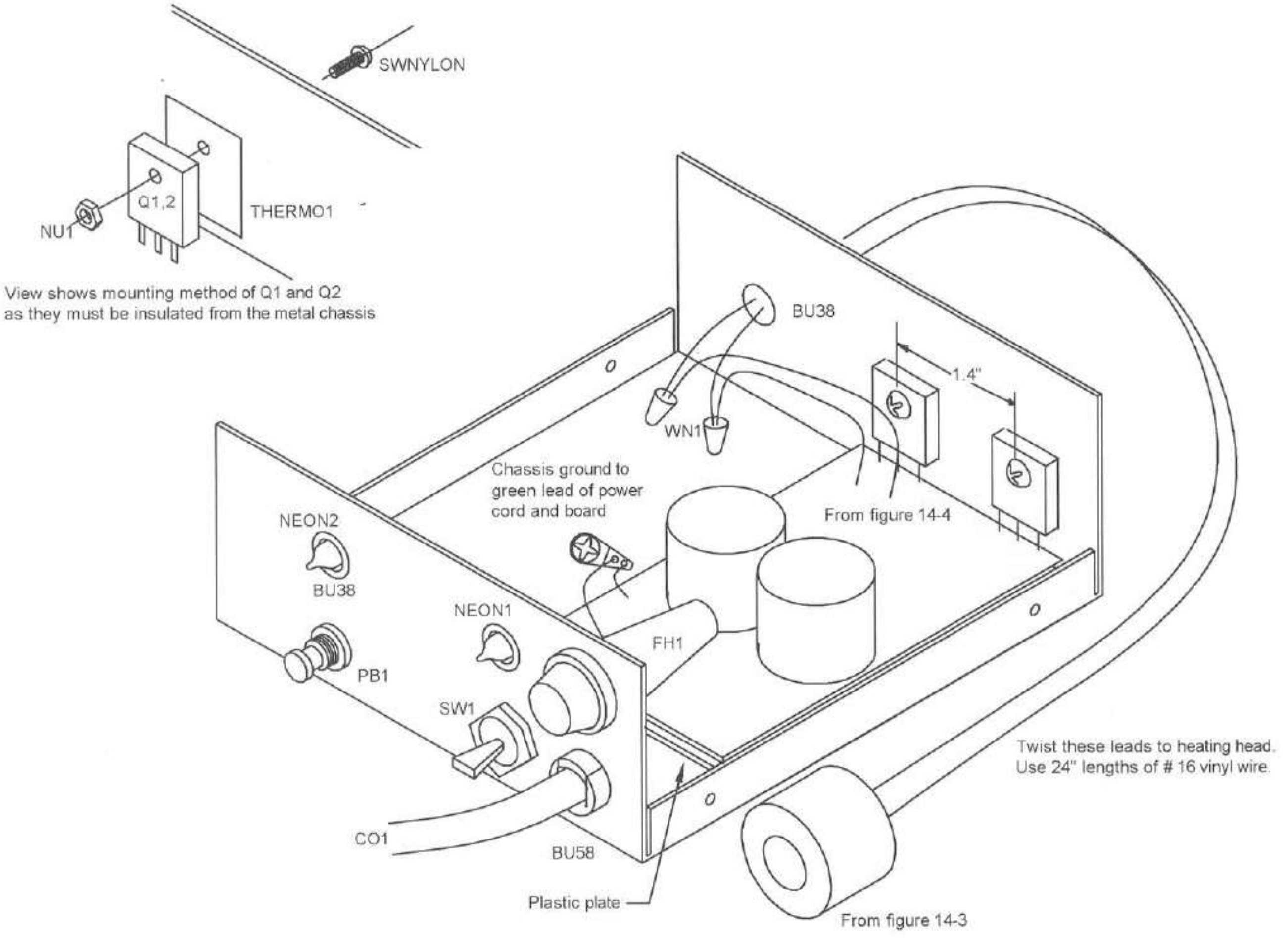
SWNYLON
NU1
Q1,2
THERMO1
View shows mounting method of Q1 and Q2 as they must be insulated from the metal chassis
BU38
1.4"
WN1
Chassis ground to green lead of power cord and board
From figure 14-4
NEON2
BU38
NEON1
FH1
PB1
SW1
Twist these leads to heating head. Use 24" lengths of # 16 vinyl wire.
CO1
BU58
Plastic plate
From figure 14-3

Figure 14-7 *Isometric of final assembly*

This test jig as shown prevents catastrophic damage to AC powered circuitry in the event of a fault when first assembled. Basically it uses a 60-watt lamp as a ballast, connected in series with the hot wire of the AC line. The lamp will usually allow ample current to flow for operational circuit verification yet limit short circuit current to that of the lamp should the circuit have a catastrophic fault. It is suggested that the serious experimenter construct this test jig and enclose it in a suitable container with receptacle for plugging in any AC circuit under test.

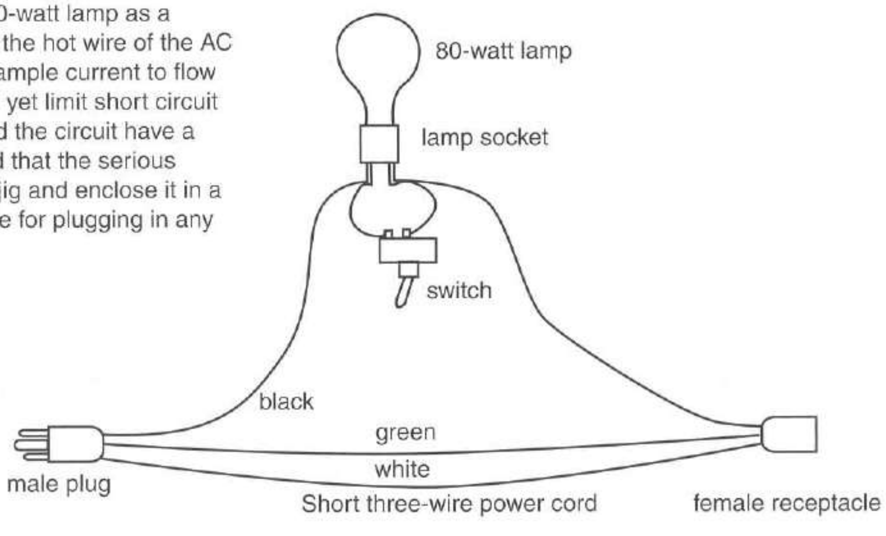
80-watt lamp
lamp socket
switch
black
green
white
male plug
Short three-wire power cord
female receptacle

Figure 14-8 *Useful test jig for line-drive circuits*

Table 14-1 Induction Heater Parts List

Ref. #	*Description*	*DB Part #*
R2, 3	Two 18K 3-watt resistor	
R4	100-ohm, ¼-watt resistor (br-blk-br)	
R6, 7	Two 15¼-watt resistor (br-grn-blk)	
R8	10-ohm, 3-watts metal oxide resistor	
R10	5K vertical trimmer	
R11	1 K ¼-watt resistor (br-blk-red)	
R12	100 K ¼-watt resistor (br-blk-yel) or 10-ohm 3-watt metal oxide resistor	
Rx	5-amp in-rush limiter (KC006L-ND)	
C2	10-microfarad, 50-volt vertical electrolytic capacitor	
C3	.01- microfarad, 50-volts plastic capacitor (103)	
C4	.1-microfarad, 600-volt plastic capacitor	
C5, 6	Two 1.5-microfarad, 400-volt metal polyester capacitors	
C7	.0015-microfarad, 600-volt polypropylene capacitor	
C10, 11	Two 200- to 800-microfarad, 200-volt vertical electrolytic capacitors	
C12	.01-microfarad, 2-kilovolt disc capacitor	
C40	50-picofarad, 6-kilovolt ceramic capacitor	
D1, 2, 3, 4	Four 1,000-volt, 3-amp rectifiers (1N5408)	
D5	1-kilovolt, 1-amp fast diode (1N4937)	
Z1	15-volt, .5-watt zener diode (1N5245)	
Q1, 2	IRFP450/460 *metal-oxide-semiconductor field effect transistors* (MOSFET)	
IC1	IR2153 half-bridge driver	
NEON1, 2	Two neon indicator bulbs with leads	
FH1	Fuse holder panel mount	
FUS1	2-amp slow blow 3AG fuse	
PB1	Normally closed push-button switch	
SW1	*Single pole, single throw* (SPST) toggle switch	
CO1	Three-wire line cord	
LUG1	#6 solder lug	
SO8	Eight-pin *integrated circuit* (IC) socket	
THERM1	Two thermo mounting pads for Q1 and Q2	
PCLINE	*Printed circuit board* (PCB)	DB# PCLINE
WN1	Two small wire nuts	
CHASSIS	Chassis fabricated from .065-inch aluminum sheet per Figure 14-6	
COVER	Cover fabricated from plastic sheet to fit chassis (see Figure 14-6)	
BU38	Three ⅜-inch plastic bushings	
BU58CL	⅝-inch clamp bushing for cord CO1	
SW1	Five #6 × ⅜-inch sheet metal screws for cover and LUG1	
SW2/NU1	Two 6-32 × ½-inch nylon screws and nuts	
WR24	24 inches of #24 vinyl stranded hookup wire	
WR20	24 inches of #20 vinyl stranded hookup wire	
WR16	48 inches of #16 vinyl stranded hookup wire	
PLASTIC	2 × 4-inch plastic insulating plate for under PC board	
HEAD12	Head with ½-inch clearance hole	DB# HEAD12
HEAD58	Head with ⅝-inch clearance hole	DB# HEAD58
HEAD34	Head with ¾-inch clearance hole	DB# HEAD34
HEAD78	Head with ⅞-inch clearance hole	DB# HEAD78
HEAD10	Head with 1-inch clearance hole and head parts for those wanting to attempt assembly	DB# HEAD10
WR22LITZ	#22 LITZ of multiple strands of #36 enamel wire	
BOBBIN	1.25 inches of ⅝-inch GT tubing	

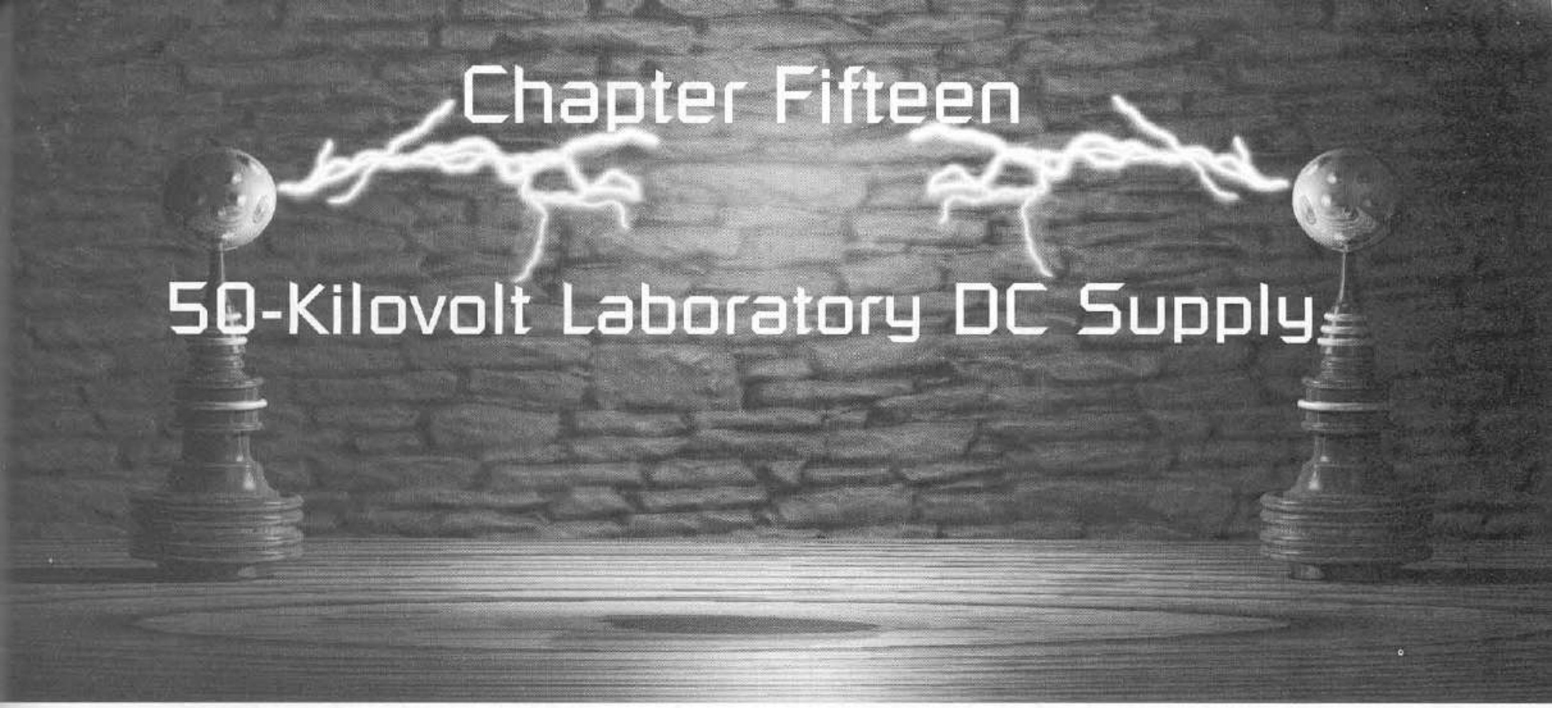

Chapter Fifteen

50-Kilovolt Laboratory DC Supply

This project, as shown in Figure 15-1, provides the high-voltage experimenter or researcher a DC source of 5 to 60 kilovolts at 5 milliamps. The system is short-circuit protected and current limited by inductive ballasting. This circuitry is excellent for high-energy capacitor charging intended for high-power applications. The charging current is fully adjustable and easily converts to a voltage source by the addition of load resistors. It is an excellent device for large gravity and lifter research, as well as normal, high-voltage laboratory functions.

This is an advanced-level project requiring electronic skills and high-voltage experience. Expect to spend $100 to $250. All parts are readily available, with specialized parts obtainable through Information Unlimited (www.amazing1.com), and they are listed in Table 15-1.

Basic Description

This project is shown constructed in two sections. The first section is the power conditioning and control circuitry. This is where the 115 VAC level is converted to a variable-frequency, high-voltage current source between 15 and 30 kHz. The second section is the high-voltage multiplier circuitry interconnected via a short umbilical cable from the power control section.

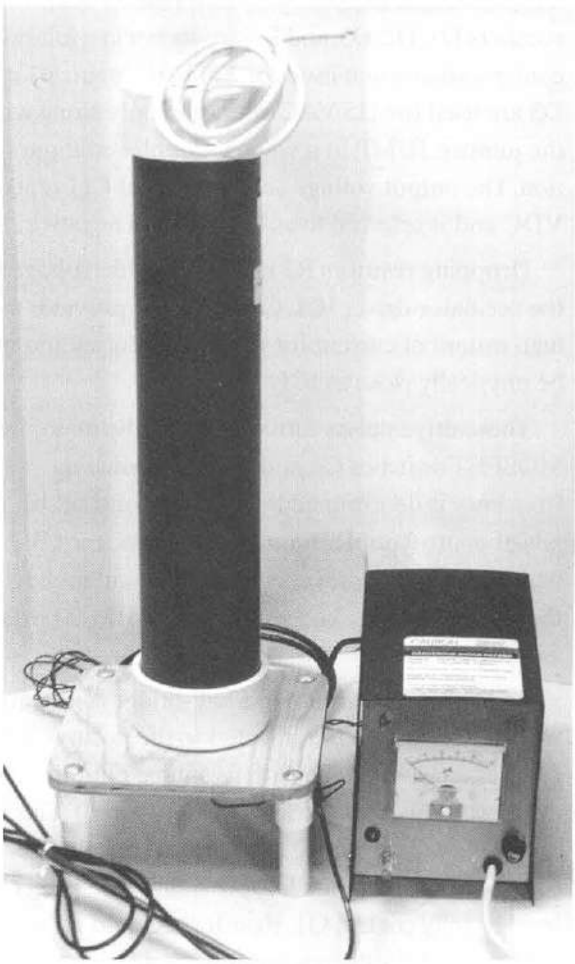

Figure 15-1 *50-Kilovolt power supply*

The multiplier section is a Cockcroft-Walton half-wave circuit consisting of 10 capacitors and diodes connected in the ladder configuration, along with a peak-current, high-voltage-limiting resistor.

Circuit Operation

The power cord (CO1) supplies 115/220 VAC from a wall receptacle where the third, green, grounding lead wire must be securely attached to the metal chassis (CHASSIS) for safety (see Figure 15-2).

Switch S1 energizes the primary circuits and fuse FUS1 protects the main circuit from any catastrophic faults. The indicator lamp NEON1 requires current-limiting resistor R12 and lights up when SW1 is turned on. The peak charging current to capacitors C10 and C11 is limited by in-rush resistor Rth. The rectifiers D1, D2, D3, and D4 are shown in a full-bridge configuration when used for 220 VAC input. D1 and D3 are used for 115 VAC operation only along with the jumper, JUMP, in a voltage-doubler configuration. The output voltage across C10 and C11 is 300 VDC and is referred to as the plus and negative rails.

Dropping resistors R2 and R3 provide voltage to the oscillator driver IC1. Capacitor C2 provides the high instant of current for the output pulses and must be physically close to IC1.

These drive pulses turn off and on the main MOSFET switches Q1 and Q2. The switching frequency is determined by the time constant of power control pot R21 and timing capacitor C3. A limit resistor R11 prevents driving the circuit beyond the intended limits. You will note that R21 is part of switch S1.

The switching circuit is in a half-bridge configuration where the MOSFET connected to the positive end (+ rail) must be driven with its source pin referenced at 150 volts above the common end (− rail). This is accomplished by biasing a bootstrap capacitor C4 through ultrafast diode D5, providing the correct DC level to fully control Q1. Resistors R6 and R7 eliminate the high-frequency parasitic oscillation that occurs as a result of rapidly switching the capacitive load of the MOSFET gates. The transition time of the switched pulses across Q1 and Q2 is slowed down by a network consisting of capacitor C7 and resistor R8. The resultant time constant limits the rate of voltage rise dv/dt rate, which could cause premature turn-on at the wrong switch, creating a catastrophic fault mode.

Capacitors C5 and C6 provide a voltage midpoint and produce the necessary storage energy to maintain the voltage level of the individual pulses. The output is taken at the junction of Q1/Q2 and C5/C6 and is a square wave fed to the primary of transformer T1, inducing the high-voltage output in the secondary winding.

A safety shutdown circuit, consisting of a spark breakover switch, fires when the voltage exceeds a certain level and is fed to isolation transformer T2. The output is now rectified by diode D6 and integrated onto capacitor C8. When this voltage exceeds a preset level controlled by trimpot R4, a trigger level now turns on the silicon switch, SCR, and crowbars the voltage to the drive chip IC1, shutting it down and disabling the high voltage until reset by power removal.

The 15,000-volt, high-frequency output is fed to the voltage multiplier stack, consisting of capacitors C14 through C21 and diodes D7 through D14. The resultant 50,000-volt DC output is in series with peak-current-limiting resistors (R17 through R20).

Assembly of the Driver Section

1. Assemble the circuit board as shown in Figure 15-3. Note to allow a length of a half-inch for the pins of Q1 and Q2. This is necessary to allow these parts to be placed flat against the chassis when mounting.

   It is also strongly suggested that you use an eight-pin integrated circuit socket (SO8) because IC1 is very sensitive to static electricity and possible errors during assembly.

   Note that if you are building from a perforated or vector circuit board, it is suggested that you use the indicated traces for the wire runs and insert components starting in the lower left-hand corner. Pay attention to the polarity of the capacitors with polarity signs and all the semiconductors.

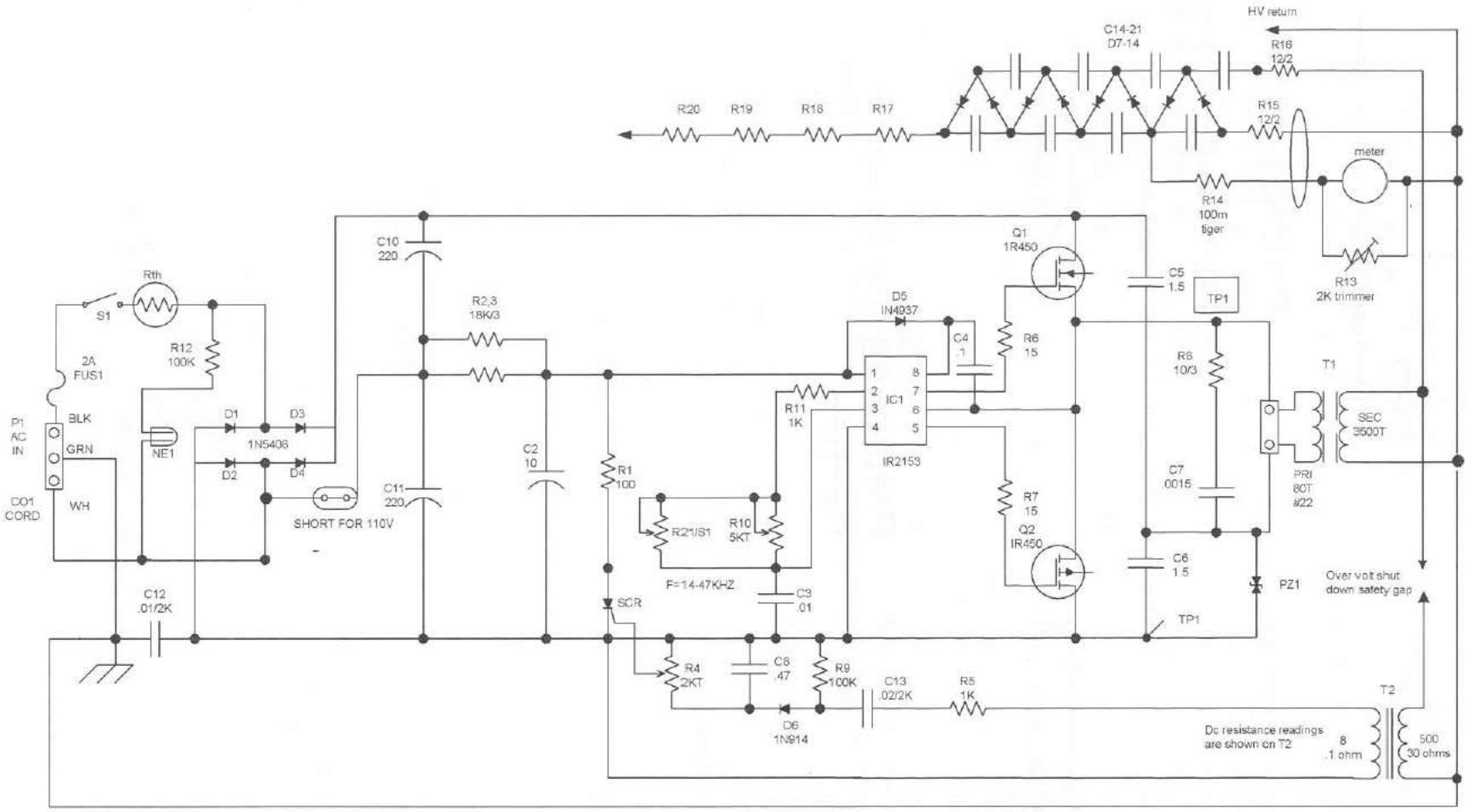
HV return
C14-21
D7-14
R16
12/2
R15
12/2
R20
R19
R18
R17
meter
R14
100m
tiger
R13
2K trimmer
Q1
1R450
C5
1.5
TP1
R8
10/3
T1
SEC
3500T
PRI
80T
#22
C7
.0015
C6
1.5
PZ1
Over volt shut
down safety gap
TP1
Q2
IR450
R7
15
R6
15
D5
IN4937
C4
.1
IC1
IR2153
R11
1K
Rth
S1
2A
FUS1
P1
AC
IN
BLK
GRN
WH
CO1
CORD
R12
100K
NE1
D1
D2
D3
D4
1N5406
SHORT FOR 110V
C10
220
C11
220
R2,3
18K/3
C2
10
R1
100
SCR
R21/S1
R10
5KT
F=14-47KHZ
C3
.01
C12
.01/2K
R4
2KT
C8
.47
R9
100K
D6
1N914
C13
.02/2K
R5
1K
T2
Dc resistance readings
are shown on T2
8
.1 ohm
500
30 ohms

TRANSFORMER				
PRIMARY	80	#22 LITZ	9MH	4.75MH
XFMR	TURNS	WIRE	Lsc	Loc
SECONDARY	3500	#39 MAG	3HY	9H
AIR GAP=	2 MILS/SIDE			

Figure 15-2 *Circuit schematic*

Route the leads of the components as shown and solder as you go, cutting away unused wires. Attempt to use certain leads as the wire runs or use pieces of the #24 bus wire. The heavy foil runs should use the thicker #20 bus wire as these are the high-current discharge paths.

2. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.
3. Assemble the transformer as shown in Figure 15-4, or you may purchase this part as noted on the parts list.
4. Connect the external components as shown in Figure 15-5.
5. Fabricate the chassis as shown in Figure 15-6 from a piece of .035-inch 5052 bendable aluminum. Note to trial-fit and measure the diameters of the components before drilling the holes. Pay attention to the holes for mounting Q1 and Q2, as these must be reasonably accurate to align with the layout of the PCB. Also, there must be at least $^{1}/_{8}$ inch of clearance between the board and chassis for sandwiching in the PLASTPL plastic plate. This prevents grounding out of the board's wiring traces.
6. Fabricate the plastic base from a 9 × 6-inch piece of $^{1}/_{16}$-inch Lexan sheet, as shown in Figure 15-7.
7. Fabricate the parts as shown in Figure 15-8 for the over-voltage shutdown switch.
8. Attach the metal chassis to the plastic base via four screws and nuts, SW12/NU1 (see Figure 15-9).
9. Attach the shutdown switch assembly from Figure 15-8 using sheet metal screws SW1.
10. Mount the assembled PCB, as in Figure 15-3. Note that Q1 and Q2 are mounted using the insulating THERMO1 pads with the nylon screws. Mount the prewired front panel components, as in Figure 15-5.
11. Mount transformer T1 via an 8-inch tie wrap. Note the cores of T1 are wrapped with a single turn of #18 bus wire and this is connected to the ground lug electrically grounding the core.
12. Mount transformer T2 using a piece of double-sided foam tape or glue it in place.
13. Mount the remaining hardware, wiring lugs, and so on.
14. Proceed to wire everything as shown in Figure 15-10. Use twist pairs where noted or tie-wrap to neaten its appearance (see Figure 15-11).
15. Double-check the wiring for any obvious errors. You should have a finished assembly resembling Figure 15-12. The unit is now ready for the basic pretest.

Pretesting the Driver Section

This step will require basic electronic laboratory equipment including a 60 MHz oscilloscope

1. Assembly the test jig as shown in Chapter 14 Figure 14-8. This fixture is a very useful for any 115 VAC project up to 300 watts.

Caution: The circuit commonly noted as −RAIL is electrically above ground by over 150 volts. It is strongly recommended that you obtain a 200-watt isolation transformer before proceeding and that you test this or similar direct-line-powered AC circuits. Experienced people working on dry, wooden floors may choose to unground their test equipment. This still can be a dangerous shock hazard!

2. Connect the scope to the test points as noted in the Figure 15-2 schematic or as shown in Figure 15-4. The lead of R8 is shown connected to the junction of the drain and the source of Q1 and Q2 respectively.
3. Short out the output leads.

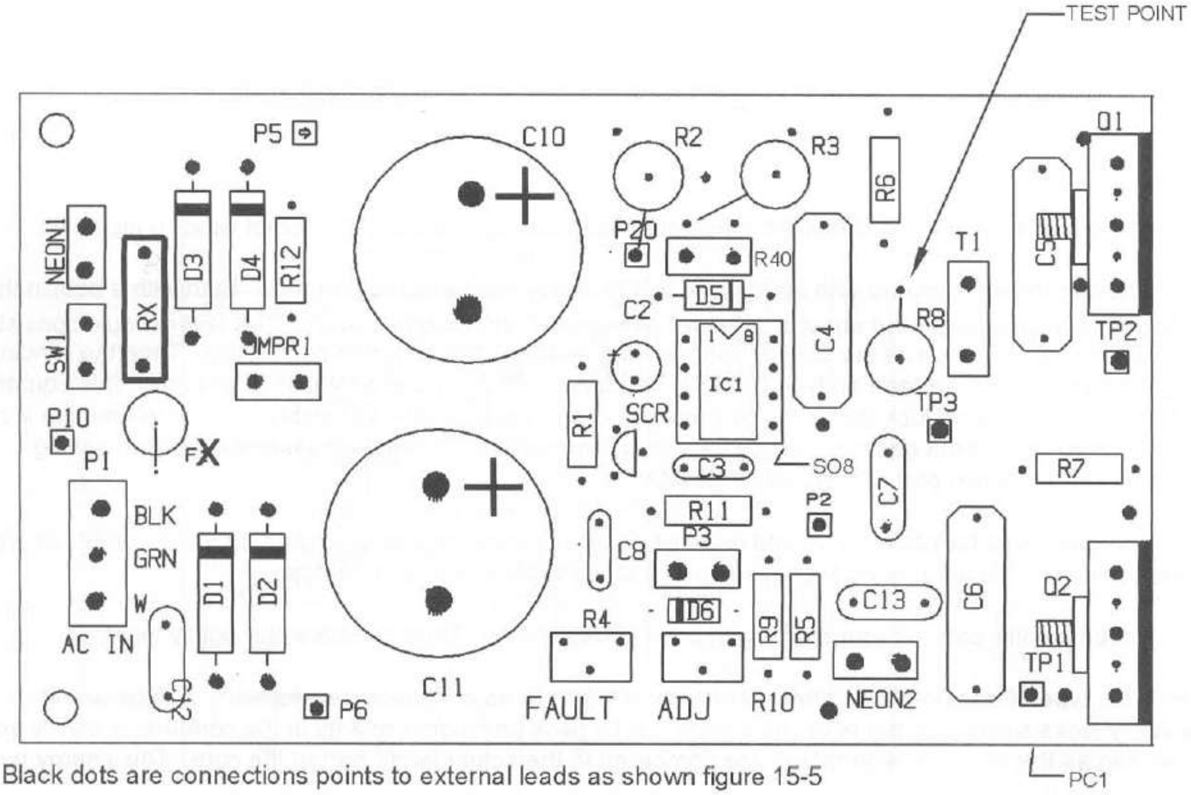
TEST POINT
P5
C10
R2
R3
R6
Q1
NEON1
RX
D3
D4
R12
P20
R40
D5
C2
C4
T1
R8
C5
TP2
SW1
JMPR1
IC1
1
8
TP3
P10
P1
FX
R1
SCR
C3
SO8
R7
C7
BLK
GRN
W
AC IN
D1
D2
R11
P2
C8
P3
C6
Q2
D6
C13
R4
R9
R5
TP1
C12
P6
C11
FAULT
ADJ
R10
NEON2
PC1

1. Black dots are connections points to external leads as shown figure 15-5

2. The PC board as shown is used in several projects and may not use all the components as indicated.

3. This circuit uses a 8 pin socket for IC1 as it is prone to failing if there are circuit faults

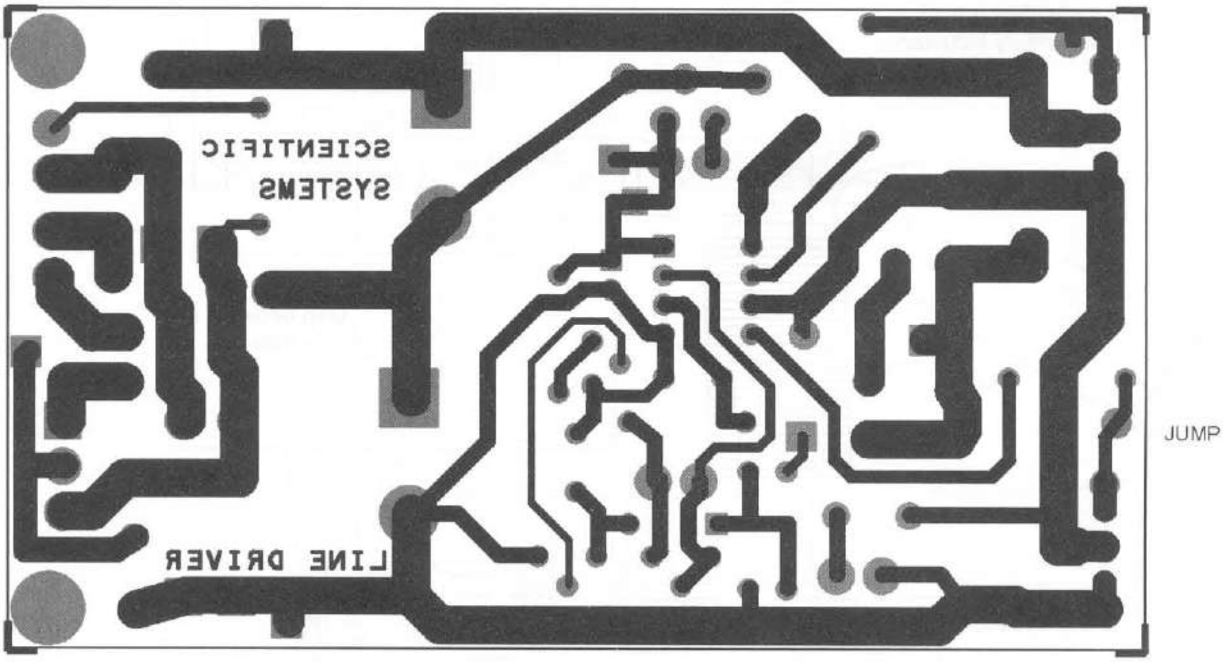
SCIENTIFIC
SYSTEMS
LINE DRIVER
JUMP

Assembly of the board is shown using a printed circuit board available through www.amazing1.com Builders may use a vector board and use the component leads to duplicate the foil traces as shown.

Figure 15-3 *Assembly of the circuit board*

The special ferrite transformer used in this project is available ready to use and is listed on table 15-1. The transformer is designed with a certain amount of leakage reactance between the primary and secondary winding to allow limited short circuit current. This approach allows charging capacitors without the use of lossy resistors. While this is an advantage in circuit protection is has the disadvantage of requiring transistor switches (FETS) that now must switch the reactive leakage current in addition to the real load current. We show our circuit using high current switches that alleviate this problem.

The design specifications are shown for those advanced builders desiring to attempt winding their own.

1. The secondary is wound with 3500 turns of #39 heavy insulated magnet wire. Start with a bobbin that will fit over the cores and wind about a ¾" layer evenly and without cross overs. The layer should consist of about 200 turns. Note to exit the start of your winding as this is the common connection. Tape the winding using 2 mil polypropylene tape and wind the second layer until you have a total of 3500 turns. This equates out to 18 layers!! (good luck doing this by hand) You now must pot the assembly and this will require a cup to hold the windings and out gassing using a vacuum pump on both the winding assembly and the potting material. We use a two part silicon rubber GE#627.

2. The primary is far easier to do and requires another bobbin now only winding 80 turns of #22 Litz high frequency wire. This will require two layers and is taped together using mylar tape.

3. Assemble the core set and space with 5 mil (.005)" shims. Tape the assembly tightly together.

Note: This type of transformer is often incorrectly referred to as a "flyback transformer". A flyback unit physically looks similar but this is where it ends. A fly back unit stores energy in the core that is mainly across the air gap as this is a very high reluctance compared to the actual ferrite part of the core. This energy builds up as long as the transistor is on and then "flies back" when the transistor switches off.

Measured parameters:

Open circuit primary inductance..........5.5 mH
short circuit primary inductance.......... 1 mH
Open circuit secondary inductance.......10 H
Short circuit secondary inductance....... 2 H

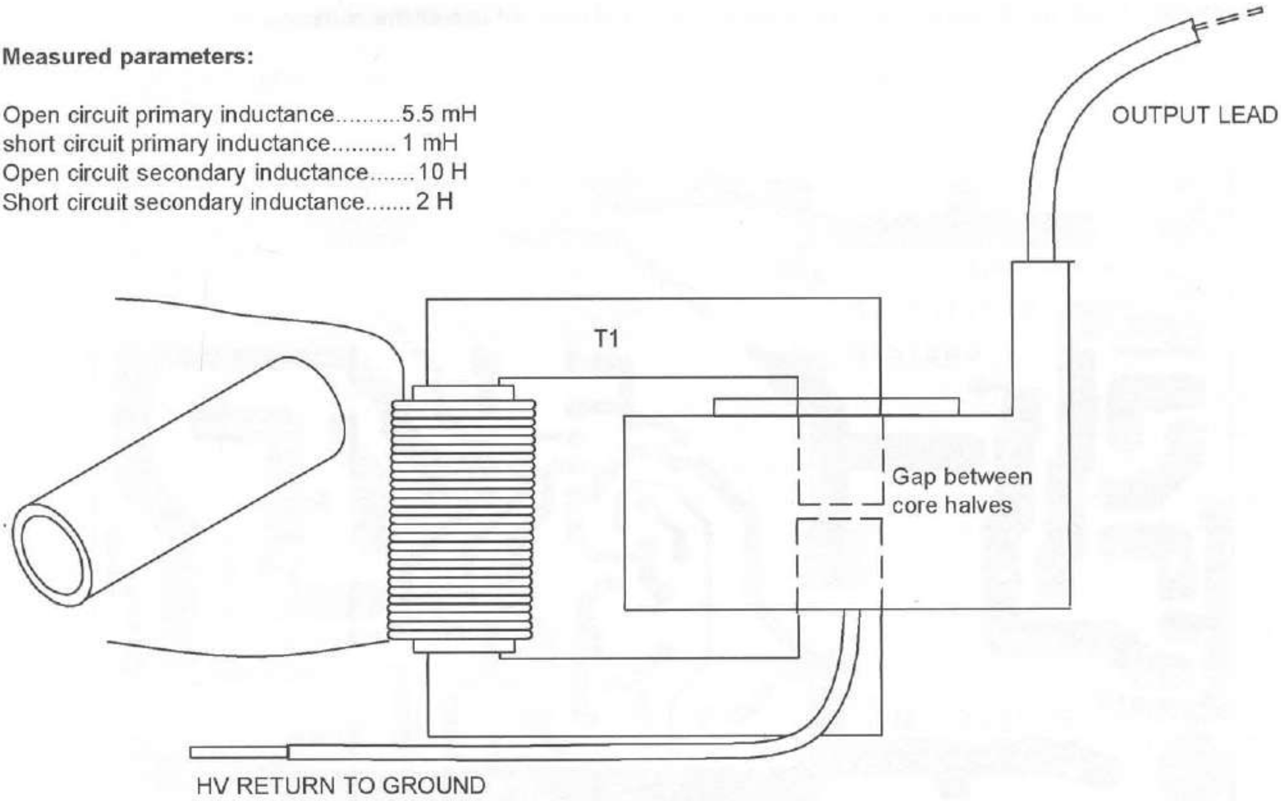

*#18 solid magnet wire can be used however high frequency LITZ wire will give a slight improvement. You can make this wire by obtaining 6 pieces of #26 magnet wire and twist together as a single wire.

Figure 15-4 *Assembly of transformer*

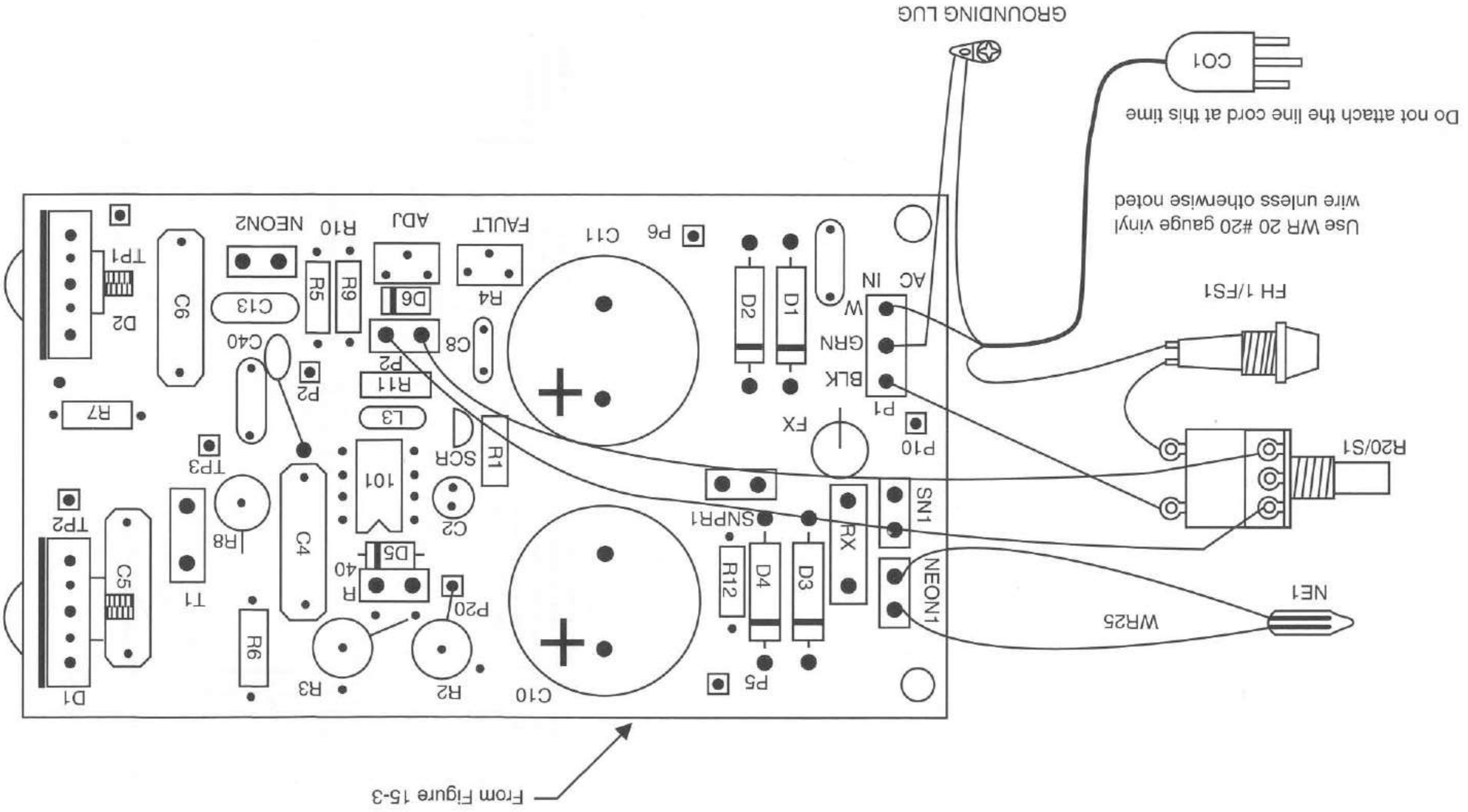
From Figure 15-3
D1
TP2
C5
D2
TP1
R7
T1
TP3
C6
R8
R6
C40
C13
NEON2
R3
C4
P2
R5
R9
R10
R
40
D5
101
L3
R11
P2
D6
ADJ
R2
P20
C2
SCR
R1
C8
R4
FAULT
C10
C11
P6
SNPR1
R12
P5
D4
D2
D3
D1
FX
RX
P1
BLK
GRN
W
AC IN
NEON1
SN1
P10
GROUNDING LUG
WR25
NE1
R20/S1
FH 1/FS1
CO1
Use WR 20 #20 gauge vinyl wire unless otherwise noted
Do not attach the line cord at this time

Figure 15-5 *External wiring*

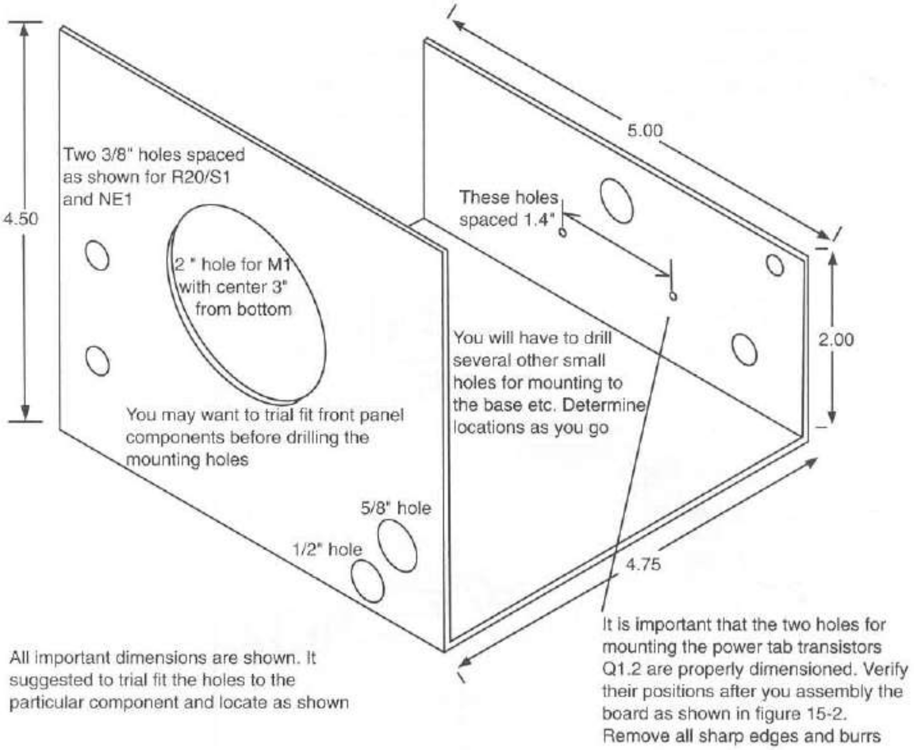
4.50
5.00
2.00
4.75
5/8" hole
1/2" hole
2 " hole for M1 with center 3" from bottom
These holes spaced 1.4"

Two 3/8" holes spaced as shown for R20/S1 and NE1

You may want to trial fit front panel components before drilling the mounting holes

You will have to drill several other small holes for mounting to the base etc. Determine locations as you go

All important dimensions are shown. It suggested to trial fit the holes to the particular component and locate as shown

It is important that the two holes for mounting the power tab transistors Q1.2 are properly dimensioned. Verify their positions after you assembly the board as shown in figure 15-2. Remove all sharp edges and burrs

Figure 15-6 *Chassis fabrication*

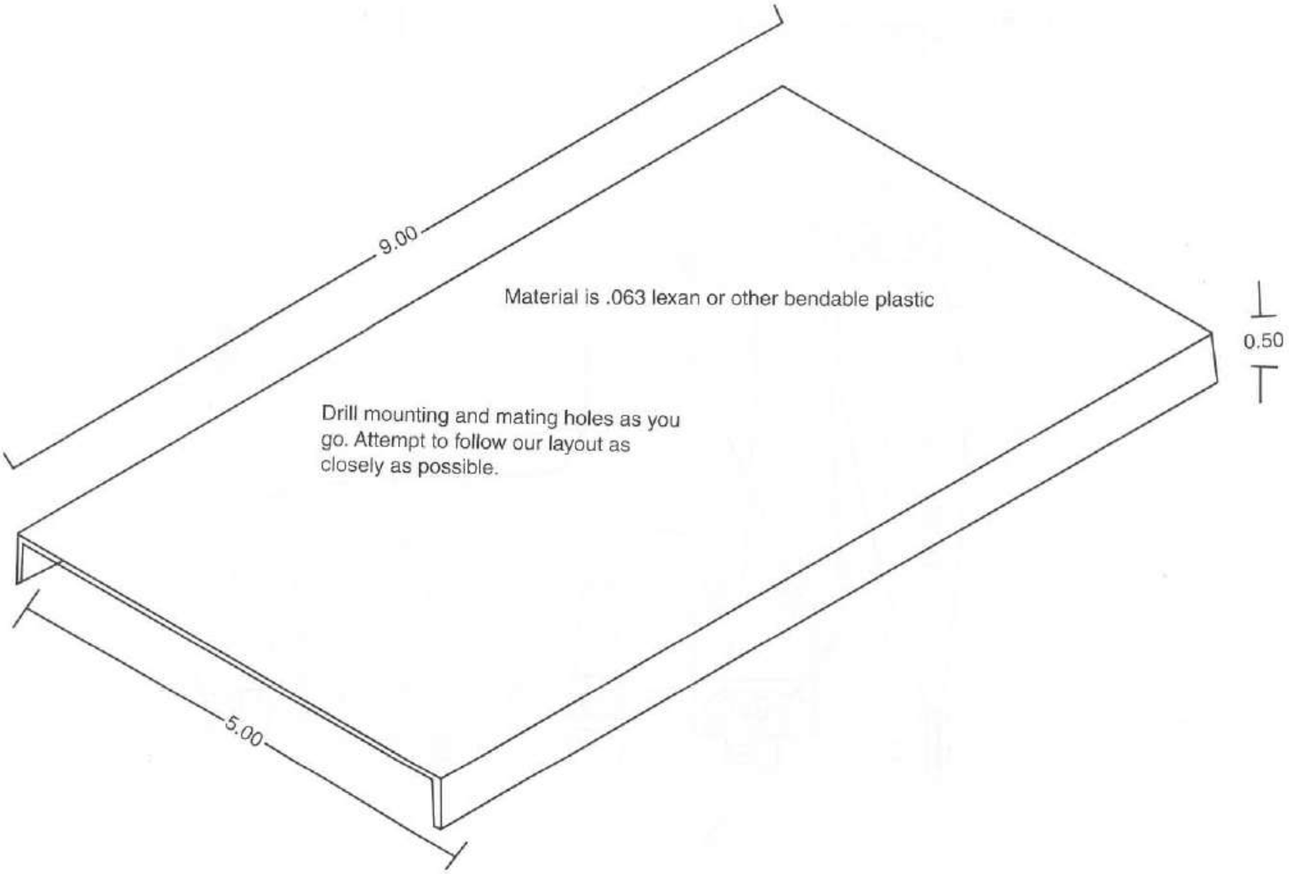

Material is .063 lexan or other bendable plastic

Drill mounting and mating holes as you go. Attempt to follow our layout as closely as possible.

Figure 15-7 *Plastic base fabrication*

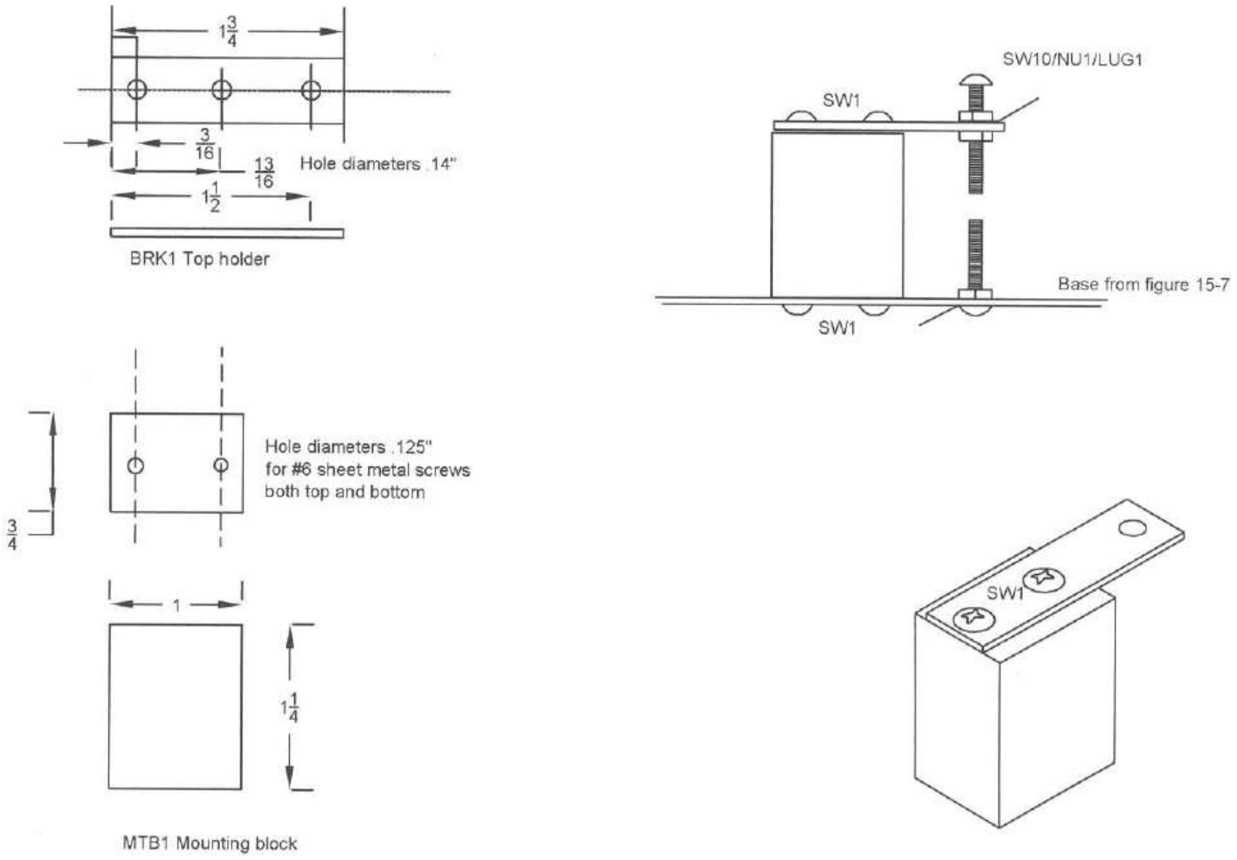

Figure 15-8 *Creating the shutdown switch*

4. Verify the main power switch S1 is off and insert a 2-amp fuse into FH1. Connect CO1 to the female end of the cord on the test jig and adjust both trimpots to midrange. Verify the test jig is not shorting the ballast lamp and plug the jig into the output of the isolation transformer. Quickly turn on S1 and note NEON1 lighting and the ballast lamp not lighting. If it brightly lights up, you must troubleshoot and find your error before proceeding. Otherwise, note a picture-perfect square wave on the scope with the lamp only dimly lighting. Adjust R21 full clockwise and adjust trimpot R10 for an 80-usec pulse period. This corresponds to 12.5 kHz. Experienced builders may wish to monitor the current to T1 using a current probe attachment to the scope. This waveshape should be another picture-perfect saw tooth of 8 amps maximum. This is a reactive current and is not reflected as an input current due to the phase angle.
5. Turn the unit off and unshort the output leads. Adjust the spark switch screws 1/8 of an inch.
6. Turn on R21/S1 and very quickly rotate clockwise, noting the unit shutting down as apparent by the absence of the waveshape at the test point. If this does not occur instantly, you will need to readjust the switch gap or trimpot R4 clockwise until the shutdown activates. If it does not, you must troubleshoot the circuit. This completes the driver section of this system.

Assembly of the Multiplier Section

Assembly of this section requires solder joints to be large, smooth, and globular. This is contrary to normal soldering practices but is necessary to reduce hv leakage.

1. Fabricate the MULTIBOARD board and two BRK2 brackets, as shown in Figure 15-13.
2. Assemble the diodes and components as shown in Figure 15-14, but do not solder at

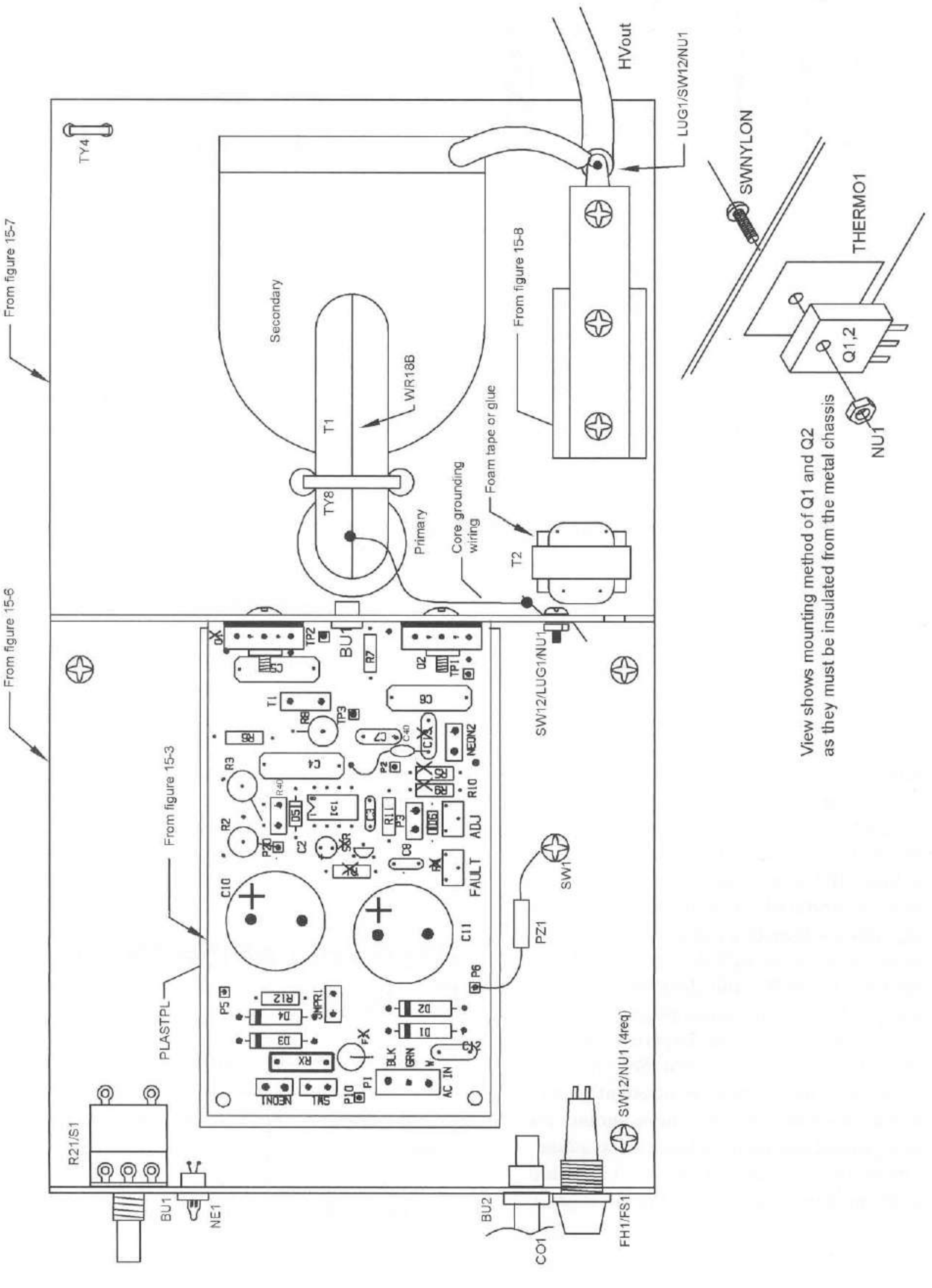
From figure 15-7
From figure 15-6
From figure 15-3
PLASTPL
R21/S1
BU1
NE1
BU2
CO1
FH1/FS1
SW12/NU1 (4req)
SW1
PZ1
P6
C10
C11
FAULT
ADJ
NEON1
NEON2
SW1
RX
FX
P1
P10
BLK
GRN
W
AC IN
D1
D2
D3
D4
R12
SNPR1
P5
C12
R2
R3
R6
R40
R8
R7
R5
R9
R10
R11
C2
C3
C4
C6
C7
C8
C1X
C40
IC1
DS1
D6
SCR
P2
P3
T1
TP1
TP2
TP3
Q1
Q2
BU1
SW12/LUG1/NU1
T2
Core grounding wiring
Foam tape or glue
Primary
TY8
T1
WR18B
Secondary
From figure 15-8
TY4
LUG1/SW12/NU1
HVout
SWNYLON
THERMO1
Q1,2
NU1

View shows mounting method of Q1 and Q2 as they must be insulated from the metal chassis

Figure 15-9 *Parts layout to driver*

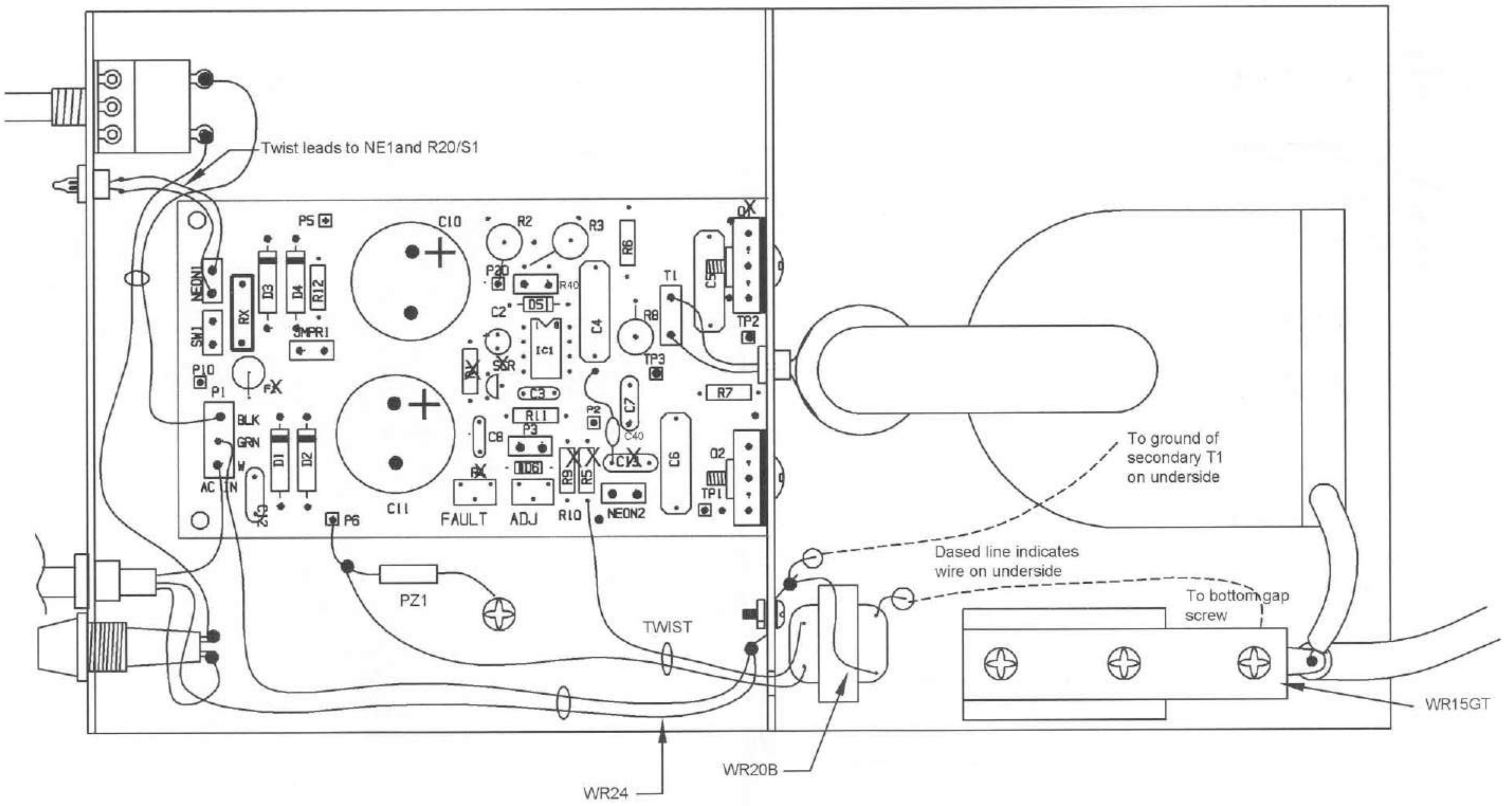
Twist leads to NE1and R20/S1
NEON1
SW1
P10
P1
RX
D3
D4
R12
JMPR1
P5
C10
FX
BLK
GRN
W
AC IN
D1
D2
C12
P6
C11
R2
R3
P20
R40
D5
C2
SGR
IC1
C3
R11
P3
C8
D6
FAULT
ADJ
R9
R5
R10
C4
R6
R8
TP3
P2
C7
C40
C13
NEON2
T1
C5
TP2
R7
C6
Q2
TP1
To ground of
secondary T1
on underside
Dased line indicates
wire on underside
To bottom gap
screw
PZ1
TWIST
WR24
WR20B
WR15GT

Figure 15-10 *Primary wiring aid*

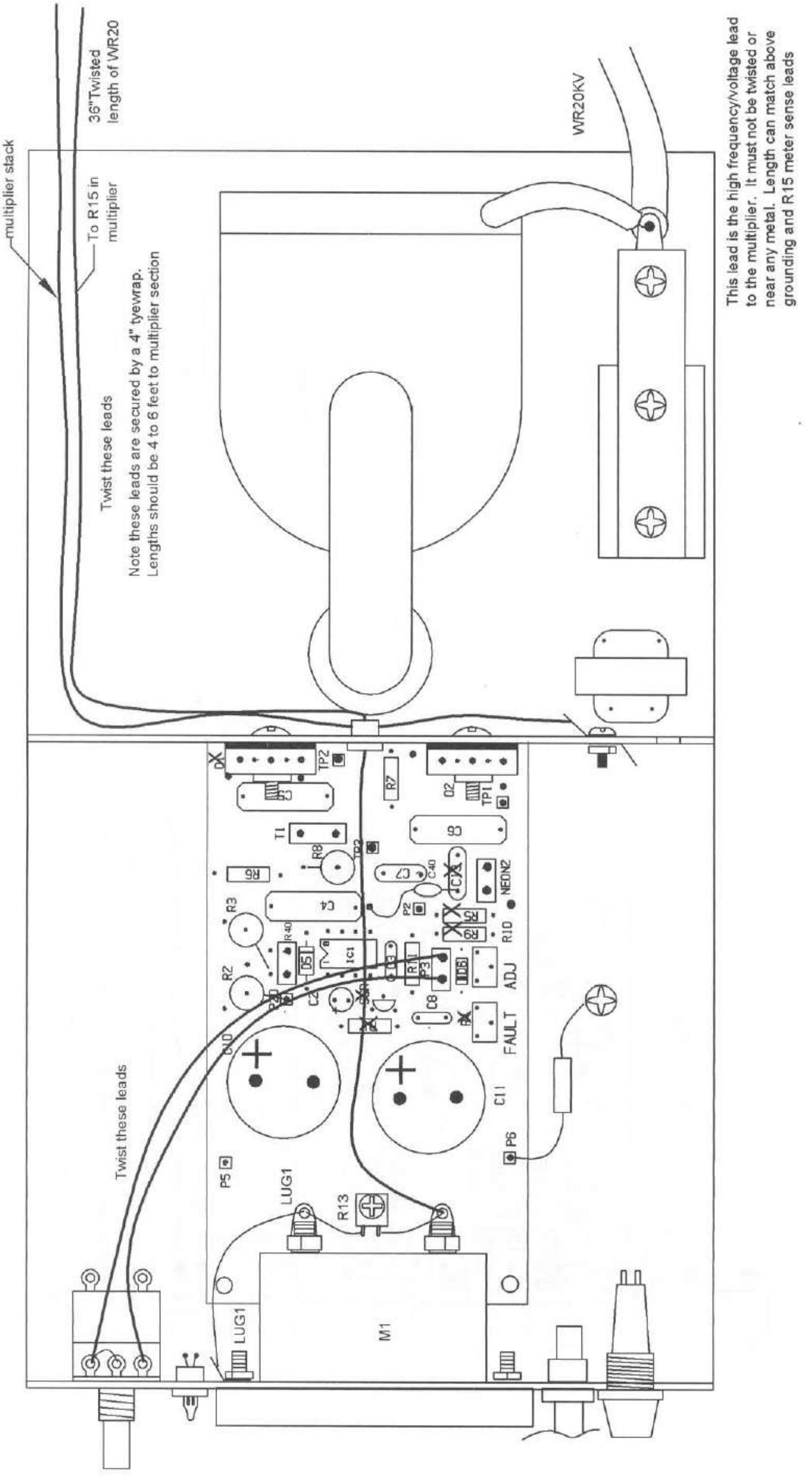

Twist these leads

Note these leads are secured by a 4" tyewrap.
Lengths should be 4 to 6 feet to multiplier section

This lead is the high frequency/voltage lead to the multiplier. It must not be twisted or near any metal. Length can match above grounding and R15 meter sense leads

HV return to multiplier stack
To R15 in multiplier
36"Twisted length of WR20
WR20KV
Twist these leads
D1
TP2
C5
R7
D2
TP1
T1
C6
R8
TP3
R6
C7
C40
CX
NEON2
R3
C4
P2
R5
R9
R10
R40
DS1
IC1
D3
R11
P3
D6
ADJ
R2
R20
C2
C3
C8
R1
FAULT
C10
C11
P6
P5
LUG1
R13
LUG1
M1

Figure 15-11 Secondary wiring aid

this time. Assemble the resistor tube for R17-20 as shown.

3. Mount the capacitors C14 through C21 and the remaining components as shown in Figure 15-15. Insert the BAF1 baffle plate to insulate and separate the left and right rows.

   Proceed to solder all the joints, noting that the capacitor and diode connections should be smooth, round balls of solder 1/8 to 3/16 of an inch wide. This may be tricky with all the leads but should be done to avoid leakage, especially the last four capacitors.

4. Sleeve resistor R14 into a plastic sleeve, as shown in Figure 15-16.

5. The final assembly is shown in Figure 15-17. Note you have a choice of an output lead or a spherical terminal.

Final Testing

This step requires basic electronic laboratory equipment including a 60 MHz oscilloscope along with a high-voltage meter with a 40 kilovolt dc rating. You will need a high-voltage load resistor to verify output power capabilities.

1. Connect the driver and multiplier sections together and separate them by at least 3 feet. Keep the high-voltage wire from the driver away from other conductive objects. Follow Figure 15-11 for the driver and Figure 15-17 for the multiplier interconnecting wiring.

2. Obtain a load resistor of around 10 megohms of at least 100 watts. This resistor must be able to support at least 50,000 volts. Our laboratory load resistor consists of 180 47k, 1-watt resistors connected in series and then inserted into

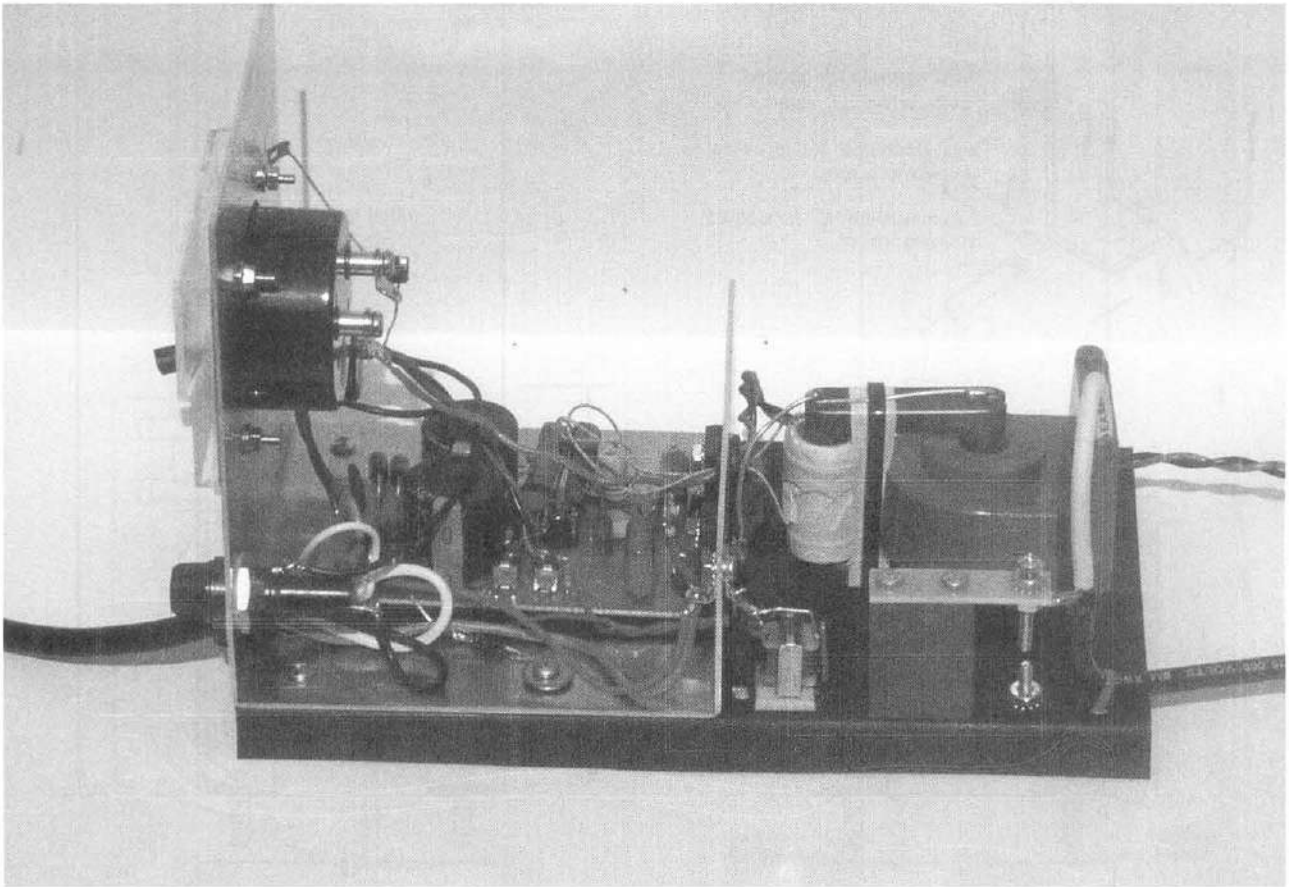

Figure 15-12 *Driver section*

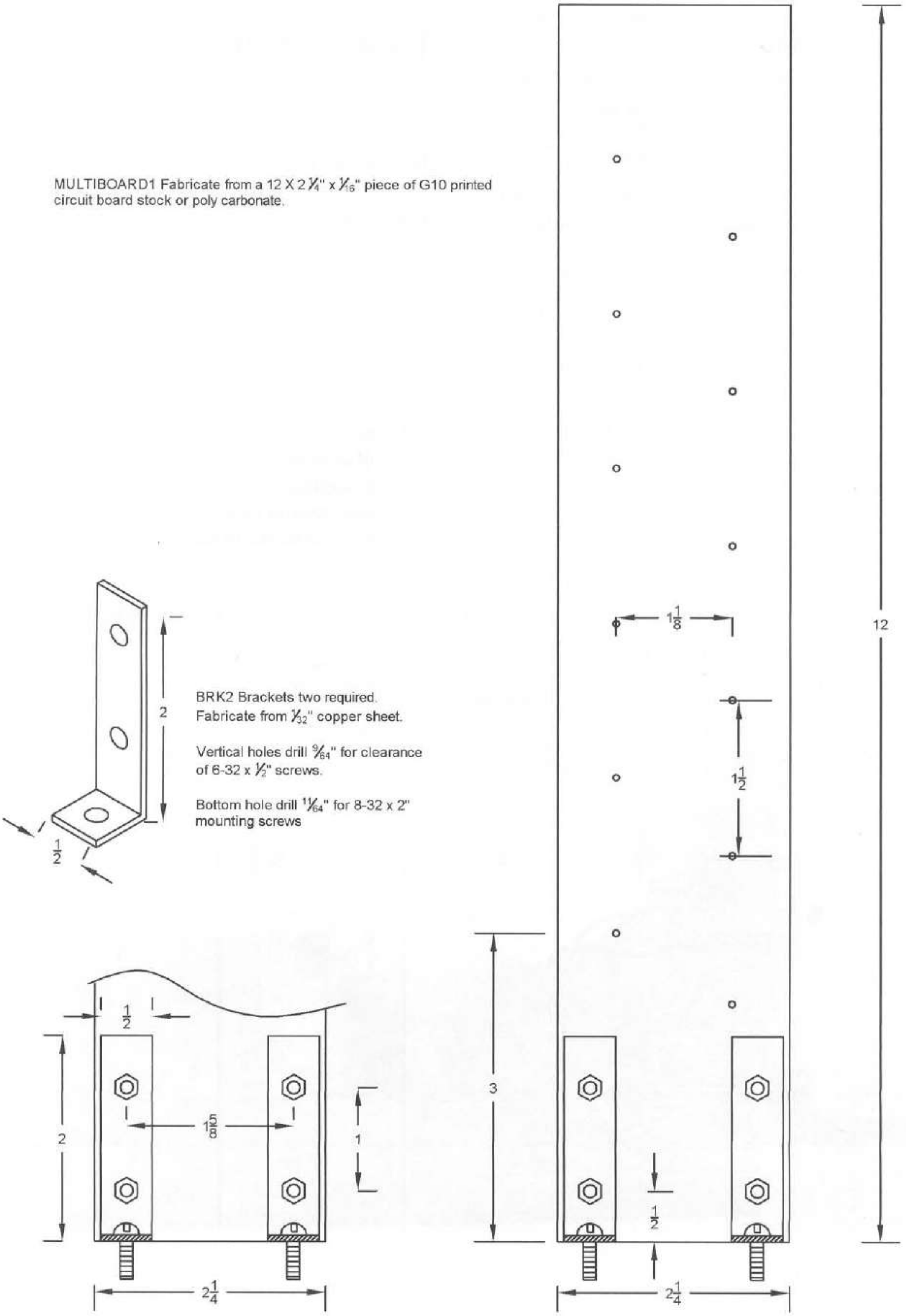

MULTIBOARD1 Fabricate from a 12 X 2 1/4" x 1/16" piece of G10 printed circuit board stock or poly carbonate.

BRK2 Brackets two required. Fabricate from 1/32" copper sheet.

Vertical holes drill 9/64" for clearance of 6-32 x 1/2" screws.

Bottom hole drill 11/64" for 8-32 x 2" mounting screws

Figure 15-13 *Creating the multiplier board*

some polyethylene tubing, along with the terminating high-voltage leads. The tubing is then coiled, maintaining adequate clearance of the terminating leads. This is a very tedious assembly but works well for powers up to 300 watts at 100 kilovolts. Connect the load to the output and common leads.

3. Assuming that the driver is properly functioning, proceed to apply power and slowly turn the power control pot R21 to full clockwise, noting a line current draw of about 2 amps. Very carefully measure around 35 kilovolts at the output using a 40-kilovolt probe such as an HV44 or similar. The load resistors should start to get warm.

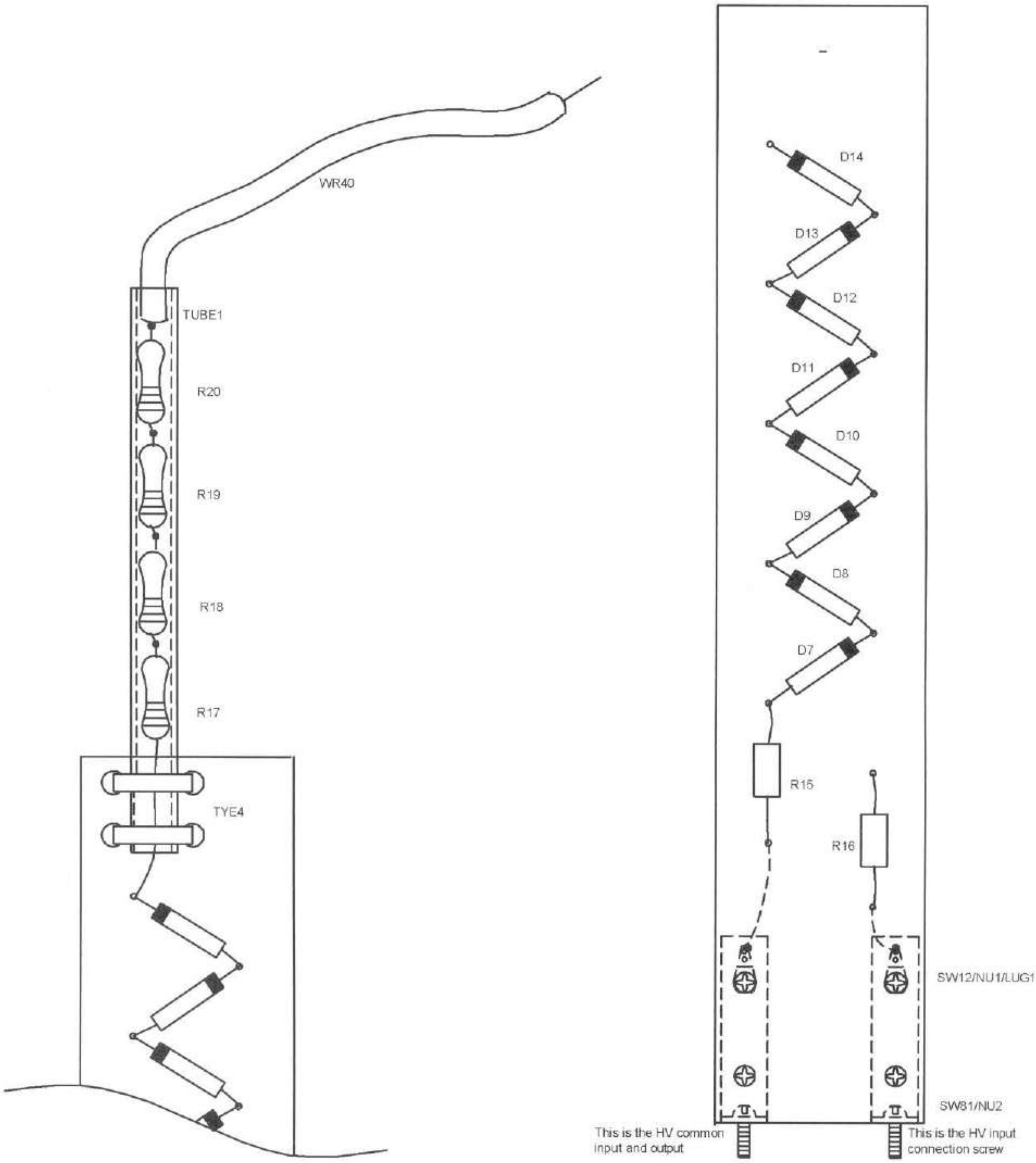
WR40
TUBE1
R20
R19
R18
R17
TYE4
-
D14
D13
D12
D11
D10
D9
D8
D7
R15
R16
SW12/NU1/LUG1
SW81/NU2
This is the HV common input and output
This is the HV input connection screw

Figure 15-14 *Diode layout and wiring*

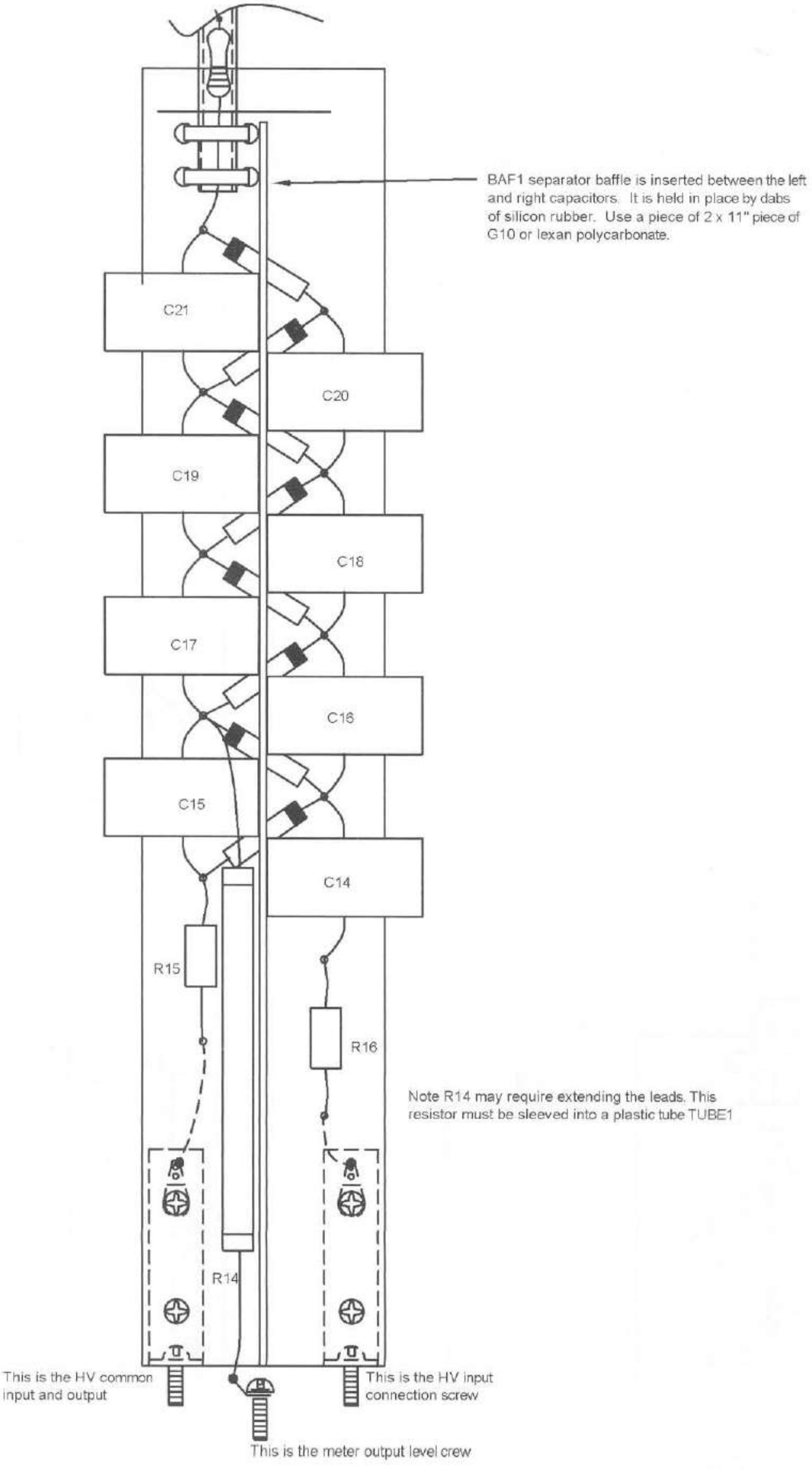
BAF1 separator baffle is inserted between the left and right capacitors. It is held in place by dabs of silicon rubber. Use a piece of 2 x 11" piece of G10 or lexan polycarbonate.
C21
C20
C19
C18
C17
C16
C15
C14
R15
R16
R14
Note R14 may require extending the leads. This resistor must be sleeved into a plastic tube TUBE1
This is the HV common input and output
This is the HV input connection screw
This is the meter output level crew

Figure 15-15 *Mount capacitors and remaining components as shown here.*

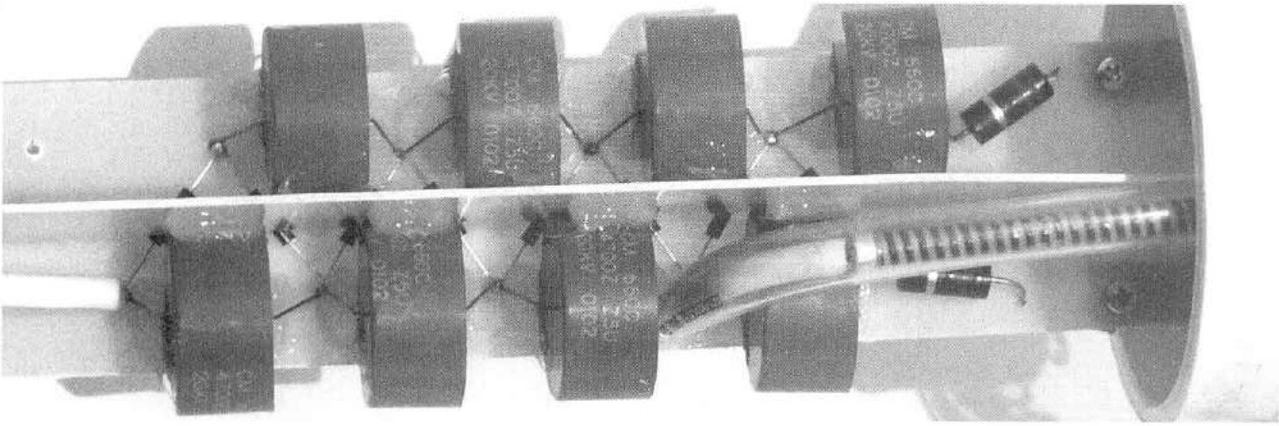

Figure 15-16 *Multiplier internal photo*

4. Allow to it run for several minutes and note that Q1 and Q2 are only slightly warm to the touch.
5. Reverify the operation of the shutdown circuit to prevent damage to the circuit components in the event of over-volting.

Applications

This unit is excellent to use for powering large antigravity lifters because it is not damaged by short-circuiting the output. It can also be used for charging capacitors up to 50,000 volts or as a voltage source with a load resistor of 20 megohms. The charging current is 4 to 5 milliamps and charges as a current source.

DANGER! Improper contact or use of high-energy capacitors can result in death by electrocution or injury due to explosive discharges.

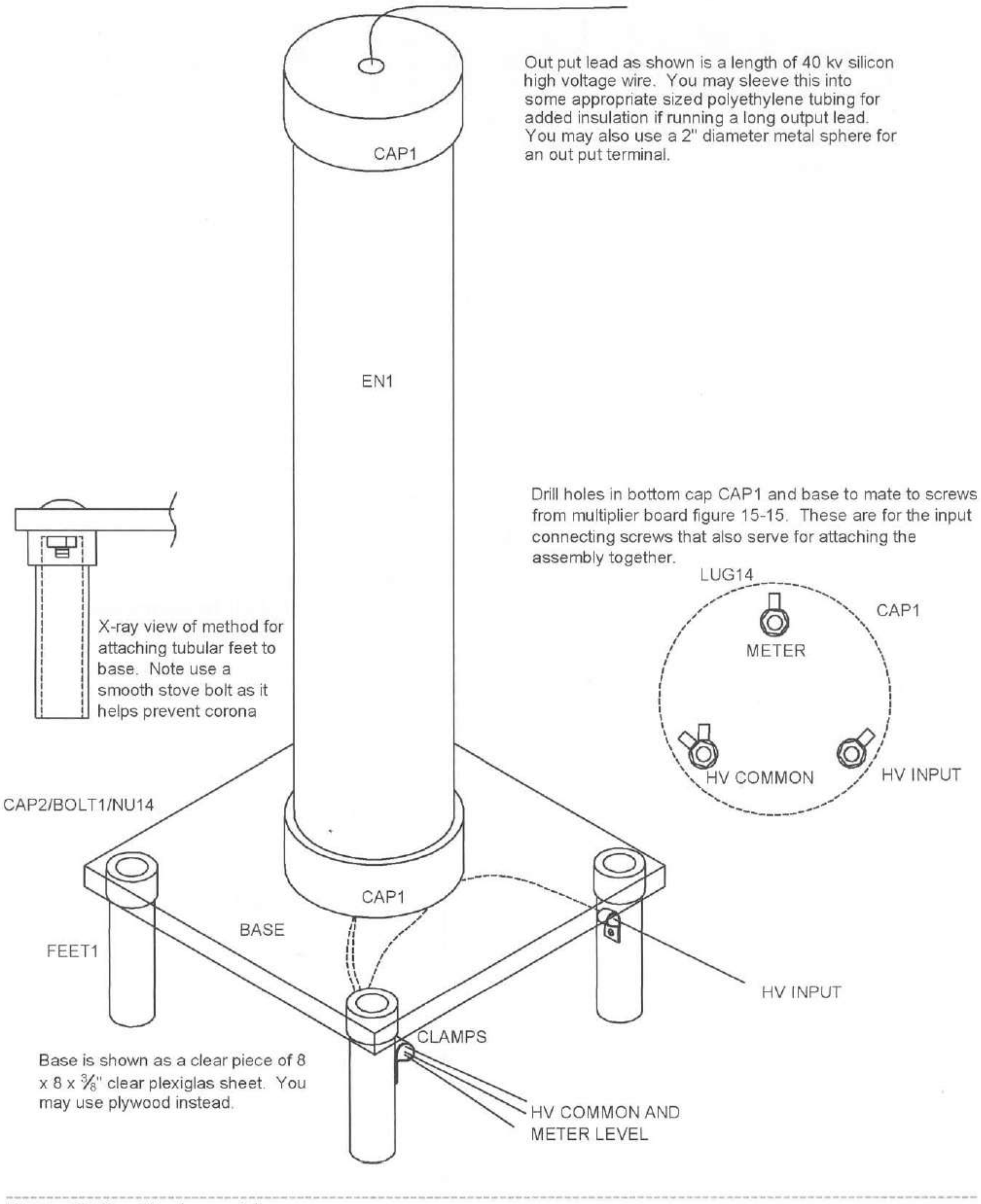

CAP1
EN1
CAP2/BOLT1/NU14
FEET1
BASE
CAP1
CLAMPS
HV INPUT
HV COMMON AND
METER LEVEL
LUG14
CAP1
METER
HV COMMON
HV INPUT

Out put lead as shown is a length of 40 kv silicon high voltage wire. You may sleeve this into some appropriate sized polyethylene tubing for added insulation if running a long output lead. You may also use a 2" diameter metal sphere for an out put terminal.

Drill holes in bottom cap CAP1 and base to mate to screws from multiplier board figure 15-15. These are for the input connecting screws that also serve for attaching the assembly together.

X-ray view of method for attaching tubular feet to base. Note use a smooth stove bolt as it helps prevent corona

Base is shown as a clear piece of 8 x 8 x $\frac{3}{8}$" clear plexiglas sheet. You may use plywood instead.

Figure 15-17 *Final assembly*

Table 15-1 50-Kilovolt Power Supply

Ref. #	*Description*	*DB Part #*
R1	100-ohm, 1/4-watt resistor (br-blk-br)	
R2, 3	Two 18K, 3-watt resistors	
R4, 13	Two 2K vertical trimmers	
R6, 7	Two 15 1/4-watt resistors (br-grn-blk)	
R8	10-ohm, 3-watt metal oxide resistor	
R10	5K vertical trimmer	
R5, 11	Two 1K 1/4-watt resistors (br-blk-red)	
R9, 12	Two 100K 1/4-watt resistors (br-blk-yel)	(KC006-LND)
Rth	limiter 5-amp in-rush	
R15, 16	Two 12-ohm, 2-watt carbon resistors (br-red-blk)	
R14	100 meg, 20-kilovolt tiger resistor	
R17–20	4 100K, 2-watt ceramic resistor	DB# 100KCER
R21/S1	10K pot and 115 VAC switch	
C2	10-microfarad, 50-volt vertical electrolytic capacitor	
C3	.01-microfarad, 50-volt plastic capacitor (103)	
C4	.1-microfarad, 600-volt plastic capacitor	
C5, 6	Two 1.5-microfarad, 400-volt metal polyester capacitors	
C7	.0015-microfarad, 600-volt polypropylene capacitor	
C8	.47-microfarad, 50-volt plastic capacitor	
C10, 11	Two 200- to 800-microfarad, 200-volt vertical electrolytic capacitors	
C12, 13	Two .01-microfarad, 2-kilovolt disc capacitor	
C14–21	Eight .0047-microfarad, 20-kilovolt ceramic capacitors	DB#.0047/20 KV
PZ1	420-volt suppressor	
D1, D2, D3, D4	Four 1,000-volt, 3-amp rectifier (1N5408)	
D5	1-kilovolt, 1-amp fast diode (1N4937)	
D6	1N914 small-signal diode	
D7–14	Eight 30-kilovolt, 5-milliamp fast-recovery diodes	DB# 3VG30
SCR	EC103 A, 12 uamp sensitive gate *silicon-controlled rectifier* (SCR)	
Q1, 2	IRFP450/460 *metal-oxide-semiconductor field effect transistor* (MOSFET)	
IC1	IR2153 half-bridge driver	
T1	High-volt transformer assembled as shown in Figure 15-4	DB# FLYPVM
T2	8/500-ohm audio transformer	
SO8	Eight-pin DIP socket	
NE1	Neon indicator bulb with leads	
FH1	Fuse holder panel mount	
FUS1	3-amp slow blow 3AG fuse	
METER	50 uamp, 3-panel meter	DB# METER50L
CO1	Three-wire line cord	
LUG1	Eight #6 solder lugs	
LUG14	Three 1/4-20 lugs	
THERM1	Two thermo mounting pads for Q1 and Q2	
PCLINE	*Printed circuit board* (PCB)	DB# PCLINE
BU38	Three 3/8-inch plastic bushings	
BU58CL	5/8-inch clamp bushing for cord CO1	
WR15GTO	4 feet of 15-kilovolt GTO flexible wire	
WR40KV	12 inches of 40-kilovolt high-voltage silicon wire	DB# WR40KV
WR20	10 feet of #20 vinyl hookup wire	
WR24	12 inches of # 24 vinyl hookup wire	
WR20B	12 inches of # 20 bus wire	
WR18B	24 inches of #18 bus wire	

Table 15-1 *continued*

Fabrication

Ref. #	*Description*	*DB Part #*
CHASSIS	Chassis created from .065-inch aluminum as per Figure 15-6	
COVER	Cover created from plastic sheet to fit chassis	
BASEPLATE	Base plate created from .065-inch Lexan per Figure 15-7	
BRK1	Bracket created from .035-inch copper per Figure 15-8	
BRK2	Two brackets created from .035-inch copper per Figure 15-13	
MTB1	Mounting block created from .75-inch PVC plastic	
PLASTPL	4.5 × 2.5 piece of insulating plastic, shown in Figure 15-9	
MULTIBOARD1	Multiplier board created from .065-inch Lexan per Figure 15-13	
BASE	Base created from 8 × 8 × 3/8-inch Plexiglas per Figure 15-17	
BAF1	Baffle separator created from .065-inch Lexan per Figure 15-15, available from a hardware store	
SW1	Five #6 1/4-inch sheet metal screws for cover and MTB1	
SWNYLON	Two 6-32 1/2-inch nylon screws	
SW12	Ten 6-32 1/2-inch screws	
SW10	Two 6-32 × 1-inch screws	
SW81	Three 8-32 × 2-inch screws	
BOLT1	Four 1/4-20 × 1 1/2-inch stove bolts	
NU1	Sixteen 6-32 keep nuts	
NU2	Six 8-32 keep nuts	
NU14	Four 14-20 hex nuts	
CAP1	Two 3-inch PVC flat caps (GENOVA# 70153)	
CAP2	Four 1/2-inch PVC flat caps (GENOVA# 30155)	
FEET1	Four pieces of 1/2 × 3-inch PVC tubing	
EN1	3 × 15-inch PVC tubing	
TUBE1	10 inches of 3/8 ID × 1/2 OD flexible plastic tubing	

Chapter Sixteen

Magnetic High-Impact Cannon

This very advanced project (shown in Figure 16-1) is intended for the experienced experimenter and researcher in the field of high-voltage and high-energy circuitry. The magnetic cannon accelerates a projectile at high velocities with considerable kinetic energy; this is accomplished strictly from a magnetic pulse. Not only is this project a dangerous electrical device, it is also kinetically hazardous, as the accelerated projectile can cause serious injury and even death. The same consideration and respect given to a firearm must be given to this project. A video demonstration of this device blasting large holes through a wall can be seen at www.amazing1.com.

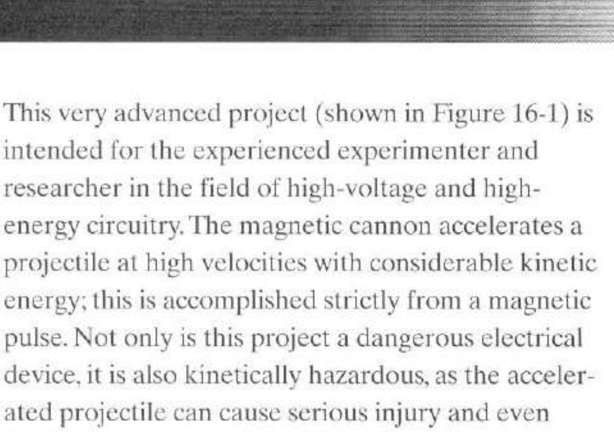

Figure 16-1 *The magnetic cannon*

Under the correct supervision, the magnetic cannon can provide a high-action science project demonstrating several important electrical laws that involve electromagnetic reactions. Lenz's law and the Lorentz JXB forces are clearly utilized in this device. The system as shown can provide the experienced hobbyist hours of fun and entertainment by experimenting with its effect (its impact) on various objects.

This is an advanced level project requiring electronic skills and high-voltage experience. Expect to spend $300 to $500 unless you have access to the surplus market. All parts are readily available and any specialized parts are available through Information Unlimited (www.amazing1.com) and are listed in the parts list at the end of the chapter.

Theory of Operation

A nonmagnetic conductor such as aluminum is placed in a time-variant magnetic field. Induced currents in the aluminum produce currents that, in turn, produce opposing magnetic fields thereby causing a moment of acceleration of the aluminum piece. In this project the aluminum piece is in the form of a large flat washer that is the projectile. The aluminum projectile has a hole in its center with a mandrel guide to keep it traveling in a straight line.

The heart of this project is the accelerating coil. It is what couples the energy (the electrical current) to the aluminum ring (the projectile). Optimum efficiency is dependent on the coupling to achieve maximum projectile kinetic energy. This requires a minimal proximity of the projectile to the coil along with minimal air gaps (such as the spacing between the wires of the coil and the geometry of the coil as related to the dimensional format of the projectile). Also the inductive value of the coil is related to the storage capacity to provide the current rise over a time period that is dependent on the physical parameters of the projectile. The associated second-order differential equations that mathematically determine these parameters are beyond the scope of this material. Although the mathematical purist will find a deviation from maximum efficiency by using different materials, practically speaking, there are materials of different sizes and materials that offer a cost-effective compromise. An example is the use of square magnetic wire in place of conventional round stock wire. Using square magnetic wire for the coil, when properly wound, will provide more kinetic energy to the projectile due to reduced air space (reluctance); however it is quite difficult to wind and usually requires purchasing a significant amount at a healthy price.

The shape and timing parameters of the coil-magnetizing pulse must be related to the projectile to achieve optimum efficiency. A current rise that is too fast will cause slippage (magnetic cavitation). Note that one definite disadvantage of this method of acceleration is that the projectile is influenced over a very short distance. Velocity, now being the square root of the acceleration times the distance, is limited. To achieve high resultant velocities requires a very high moment of acceleration. Projectiles using conventional explosives current detonators and boosters could be prone to sympathetic initiation by these high g accelerating forces at the time of launching.

Circuit Theory

This project as shown in Figure 16-2 is constructed operating from 115 vac household current. It also can be built to operate from 12 volts or built-in batteries.

A high-voltage current-limited 60 Hz transformer (T30) steps up the 115 vac to 6,500 vac and is rectified by high-voltage diodes (D31 through D34). DC current now charges the energy storage capacitors (C30) through the isolation resistor (R30) to a programmable value, as selected by the operator. It is this stored energy that is discharged into the accelerator coil (L1) as it is switched by the spark gap (GAP1). Once switched into the coil, the now rapidly rising current wave induces a current into the aluminum ring projectile (PROJ1). It is this induced current that now generates a very high magnetic moment and repels the initial field in the accelerator coil, causing a moment of intense accelerative forces. Those not familiar with this concept often ask why an aluminum ring? The answer is that a magnetic material would now be attracted and thus would neutralize the repulsion.

The initializing of the circuit commences by turning on the key switch (S2). This switch is intended to keep unauthorized personnel from powering up the system. The key switch controls 12 volts of DC power necessary to energize the relay (RE1), with normally open contacts controlling power to the high-voltage transformer (T30).

The controlling system, as shown, consists of momentary pushbutton switches that start the charging action (S4) and can stop this action via switch (S5). Triggering the momentary pushbutton switch (S3) supplies power to the trigger module (TRIG10) and firing trigger gap (GAP2), thereby initiating the main gap GAP2 and switching the energy from the storage capacitors C30 into the accelerator coil L1. Charging voltage to C30 is controlled by the potentiometer R14. Once set, the voltage will maintain its preset level until triggered or readjusted. Meter M1 indicates the charge voltage and is calibrated by the trimpot (R16).

The transformer (T1) supplies 12 volts of AC that is rectified by diodes (D1 through D4), filtered by capacitor C4, and regulated by zener diodes (Z1 and Z2) to 12 volts dc in order to power the control circuits. The indicator *light-emitting diode* (LED1) illuminates when the key switch is energized. The second indicator LED (LED2) illuminates when charging of C30 is taking place. A buzzer (BUZ1) sounds whenever there is a charge voltage on C30. This is a safety

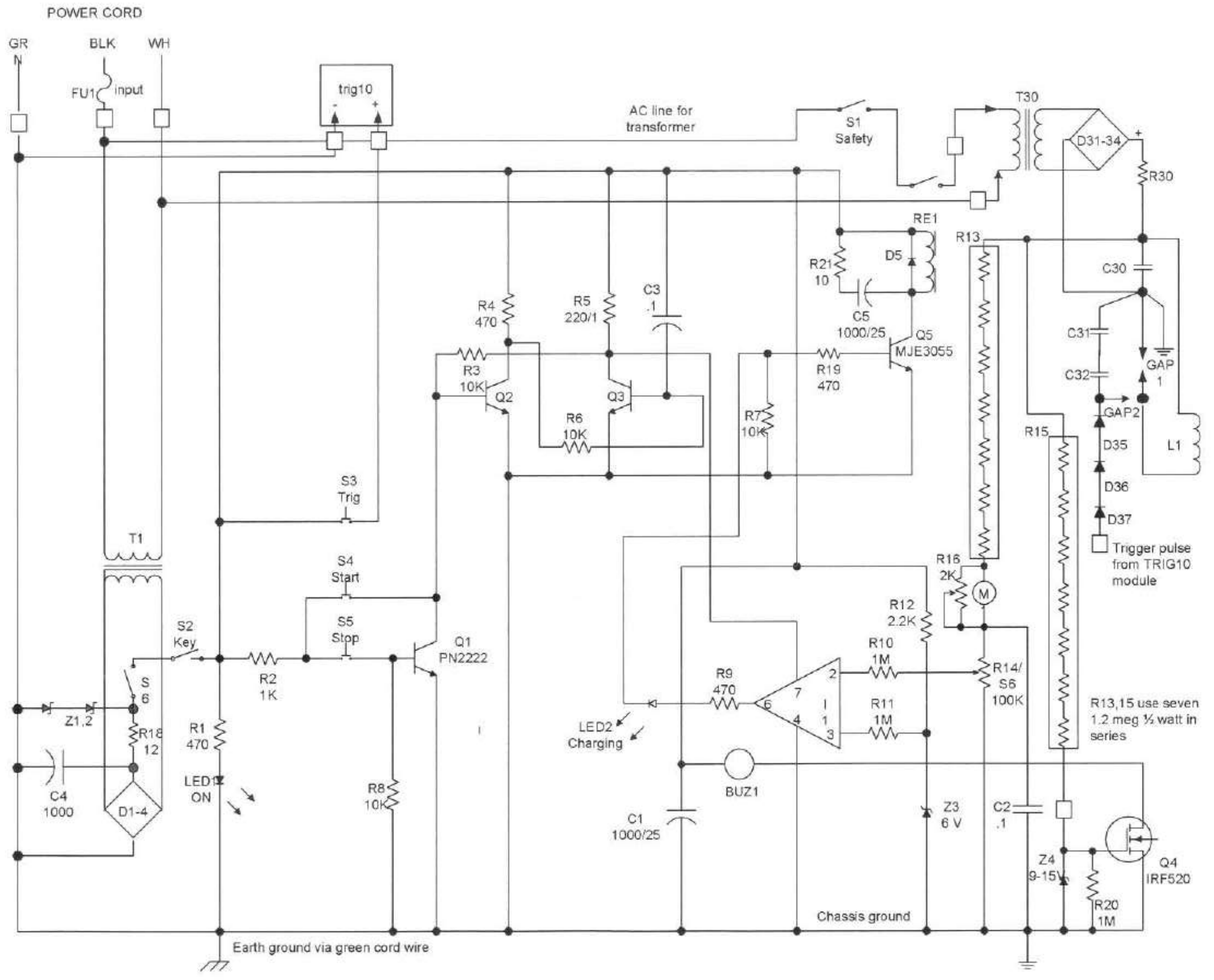
POWER CORD
GR
N
BLK
WH
FU1
input
trig10
-
+
AC line for transformer
S1
Safety
T30
D31-34
+
R30
RE1
R13
C30
R21
10
D5
C5
1000/25
Q5
MJE3055
C31
C32
GAP
1
GAP2
R4
470
R5
220/1
C3
.1
R3
10K
Q2
Q3
R6
10K
R19
470
R7
10K
R15
D35
L1
D36
D37
Trigger pulse from TRIG10 module
S3
Trig
T1
S4
Start
S5
Stop
R16
2K
M
R12
2.2K
S2
Key
Q1
PN2222
R10
1M
S
6
R2
1K
R9
470
2
7
I
1
6
4
3
R14/
S6
100K
R11
1M
R13,15 use seven 1.2 meg ½ watt in series
Z1,2
R18
12
R1
470
LED2
Charging
C4
1000
D1-4
LED1
ON
R8
10K
BUZ1
C1
1000/25
Z3
6 V
C2
.1
Q4
IRF520
Z4
9-15V
R20
1M
Chassis ground
Earth ground via green cord wire

Figure 16-2 *Schematic*

device to warn the operator that a voltage exists on the storage capacitors. Sense voltage is obtained for the meter and charge voltage control circuit via string resistor R13. The safety level voltage is obtained through the resistor string (R15).

Assembly Steps

1. Cut a piece of .1 × .1 grid perforated circuit board (PB1) 6 × 4½ inch. You will have to drill holes for transformer T1 and relay RE1.
2. Insert the components as shown in Figure 16-3, starting from the lower left-hand corner. Count the perforated holes as a guide. Note the polarity marks on capacitors and diodes. It is suggested that you use a socket (SO8) for the LM741 integrated circuit (I1).
3. Wire the components as shown in Figure 16-3 using the leads of the actual components as the connection runs. These are indicated by the dashed lines. Always avoid bare-wire bridges and globby solder joints. Check for cold or loose solder joints. Note the symbols indicating wires to external components and to solder junctions beneath board.
4. Assemble the resistor dividers (R13 and R15) as shown in Figure 16-4 from seven 1.2 meg, ½-watt resistors that are all connected in series and sleeved into a ¼-inch ID flexible plastic tubing. Place these along with the connecting leads. The leads connected to R30 must be rated for 10 kilovolts. Connect the external components using #20 vinyl-jacketed, stranded wire. Attach and solder shunt resistor R16 across meter M1. Sleeve in the connection.
5. Connect the three high-voltage diodes (D35, D36, and D37) in series as shown in Figure 16-5. Make sure there are no sharp edges on solder joints and then sleeve into a plastic tube with a short piece of high-voltage wire lead.

   Connect the remaining components, noting that some may have to be unsoldered when routing through the panel, as shown. Verify proper lengths of connecting leads with other figures in the project.
6. Fabricate the chassis from a sheet of .063 aluminum as shown in Figure 16-6. The front panel holes are shown to approximate layout and size. These should be verified with all parts for hole size and location.
7. Fabricate a 6 × 7½ inch sheet of plastic (PLATE1) to insulate the assembly board connections as shown in Figure 16-7. Assemble T30 to the chassis using screws and nuts. Include solder lugs for grounding the power cord and assembly board.
8. Fabricate a 2¼ × 7 inch piece of plastic (PLATE2) for mounting the four rectifiers (D31 through D34) and resistor R30. Use two-sided sticky tape or silicon rubber to secure. Use a long screw or piece of threaded rod with an insulating washer on top side of resistor.
9. Assemble components to the front panel, as shown in Figure 16-8, and then finalize the wiring to complete assembly. Use pieces of 10 kilovolt-rated wire for those that are shown as heavy traces. Other leads are made from #20 vinyl-jacketed, stranded wire. The grounding leads shown are from the power cord, assembly board, and D33/D34 to the lug on T30. Output leads are shown as the TRIGGER, COMMON GROUND, and HV OUTPUT. Observe all notes on this figure.

Circuit Testing

10. Obtain a ballasted 115 vac power source. You can make this simply by placing a 60-watt lightbulb in series with the hot side of the power line, which is usually designated by the black lead. Do not eliminate the green earth ground connection.
11. Verify all switches are off and then insert a 3-amp fuse into FU1. Temporarily connect the high-voltage output to the chassis ground. Output is current limited by the leakage inductance of T30.

12. Plug into the ballasted 115 vac source. Turn on S6 and the key switch. Note LED1 lighting. Check control of both these switches as they are connected in series.
13. Push S3 and note a high-voltage spark at the trigger output lead.
14. Push S4 and note the relay latching and 60-watt lamp starting to light. The unit may chatter in this mode. LED2 should light in coincidence with the relay.

This concludes the basic test of this circuit but does not verify the programming voltage function, meter reading and calibration, or the safety buzzer.

Assembly of the Completed System

15. Fabricate the upper and lower deck pieces, as shown in Figure 16-9. Fabricate four pieces of 5½-inch PVC tubing for the pillar spacers located at each corner.

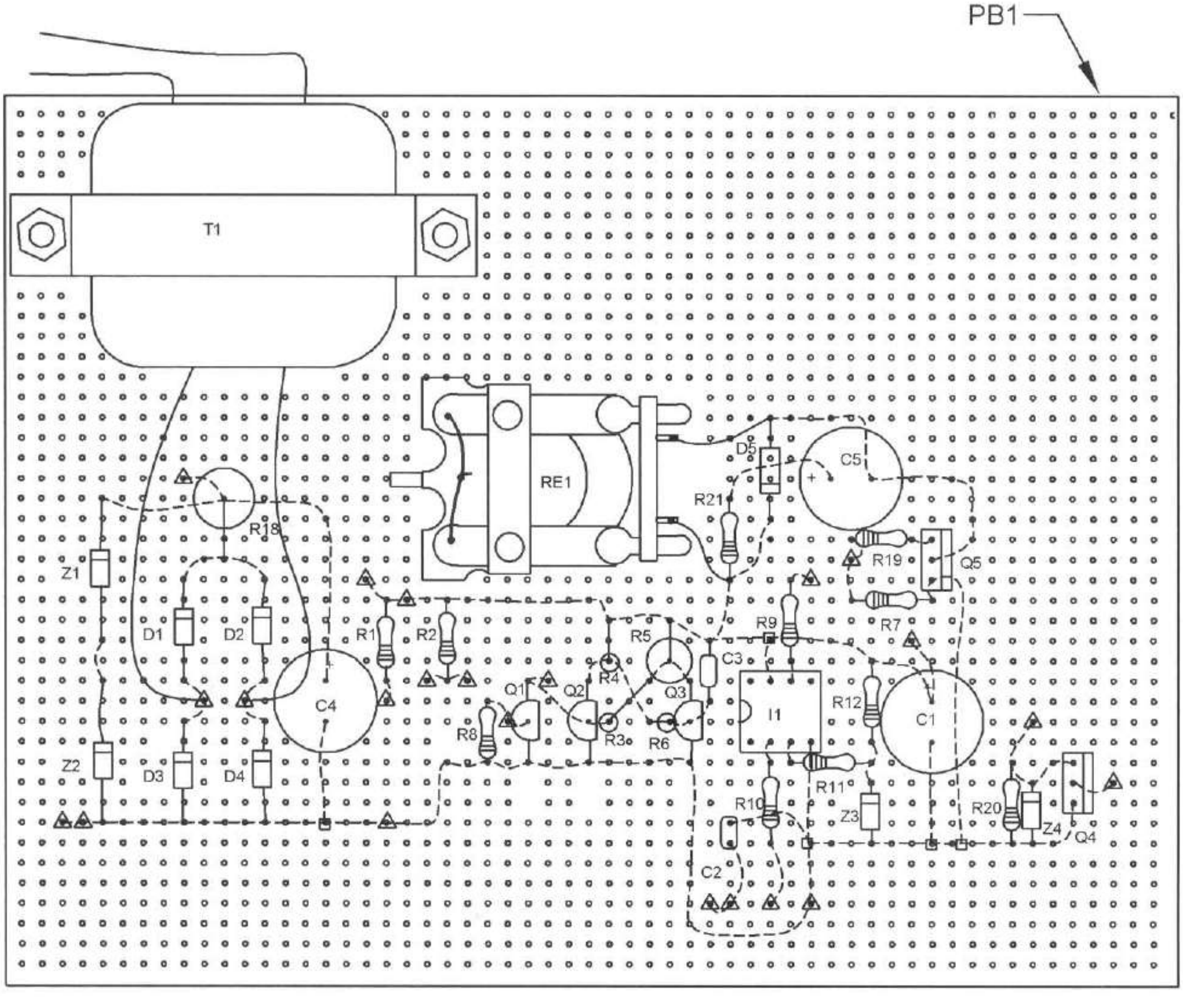

▲ This symbol indicates points for wires to external components

□ This symbol indicates solder junction of leads

- - - This symbol indicates wiring on underside of board

Figure 16-3 *Assembly board parts layout*

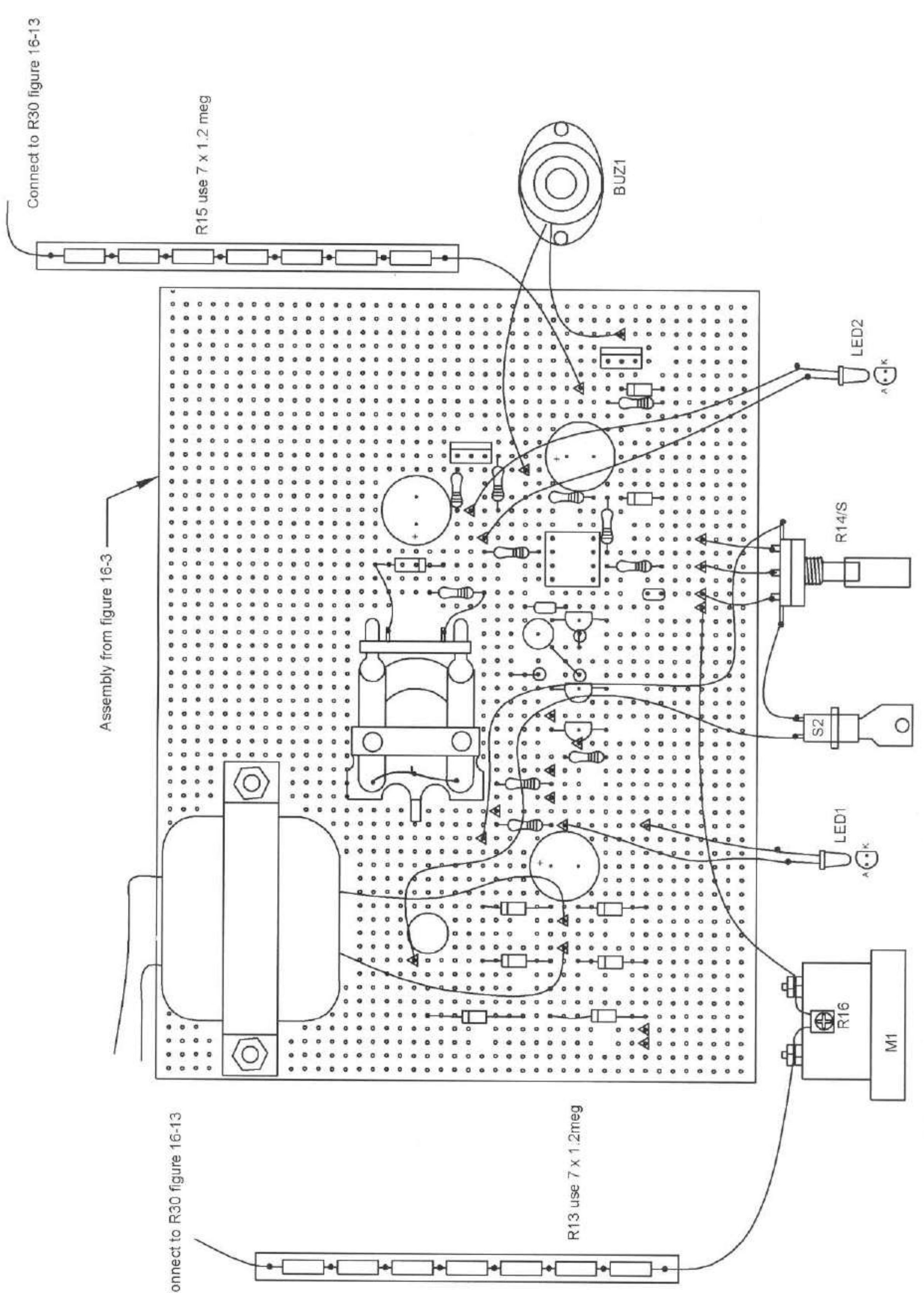
Connect to R30 figure 16-13
R15 use 7 x 1.2 meg
BUZ1
Assembly from figure 16-3
LED2
A
K
R14/S
S2
LED1
A
K
R16
M1
Connect to R30 figure 16-13
R13 use 7 x 1.2meg

Figure 16-4 *Assembly board external wiring first level*

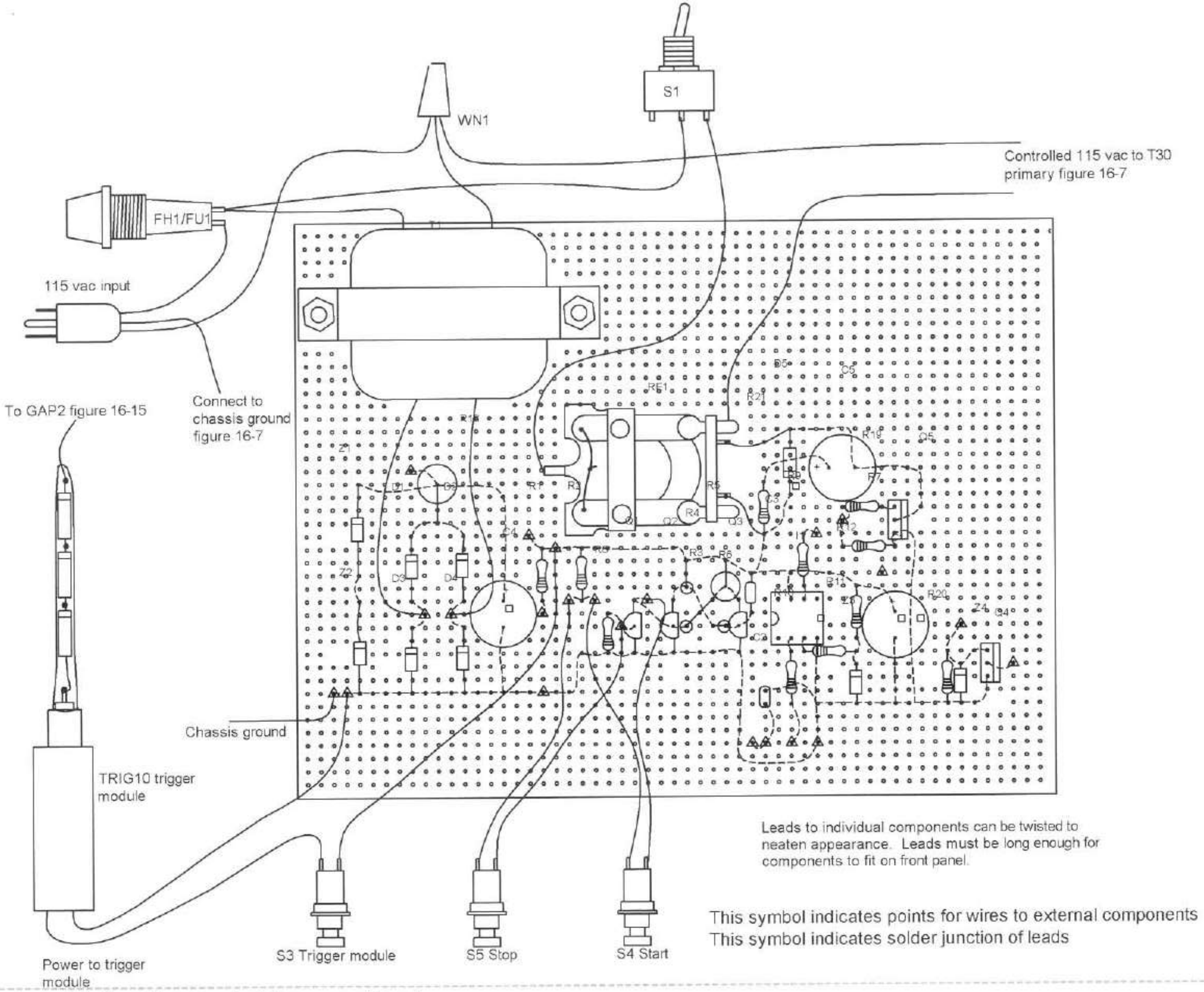
WN1
S1
FH1/FU1
Controlled 115 vac to T30 primary figure 16-7
115 vac input
Connect to chassis ground figure 16-7
To GAP2 figure 16-15
Chassis ground
TRIG10 trigger module
Power to trigger module
S3 Trigger module
S5 Stop
S4 Start
RE1
R1
R2
R3
R4
R5
R6
R7
R11
R12
R18
R19
R20
R21
Q1
Q2
Q3
Q4
Q5
C2
C3
C4
C5
D3
D4
D5
Z1
Z2
Z4
T1
Leads to individual components can be twisted to neaten appearance. Leads must be long enough for components to fit on front panel.
This symbol indicates points for wires to external components
This symbol indicates solder junction of leads

Figure 16-5 *Assembly board external wiring second level*

Fabricate from .063 aluminum sheet.

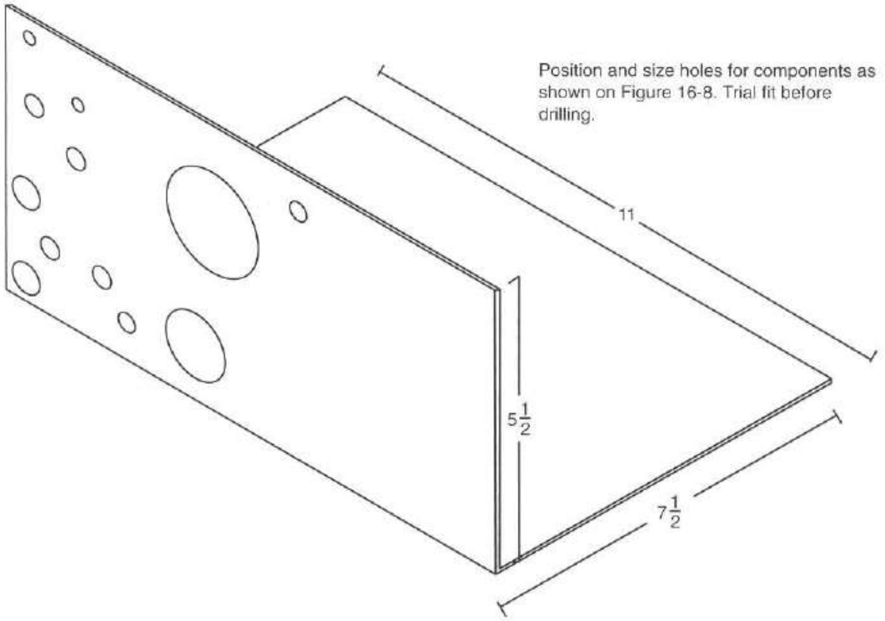

Position and size holes for components as shown on Figure 16-8. Trial fit before drilling.

11
$5\frac{1}{2}$
$7\frac{1}{2}$

Figure 16-6 *Fabrication of chassis*

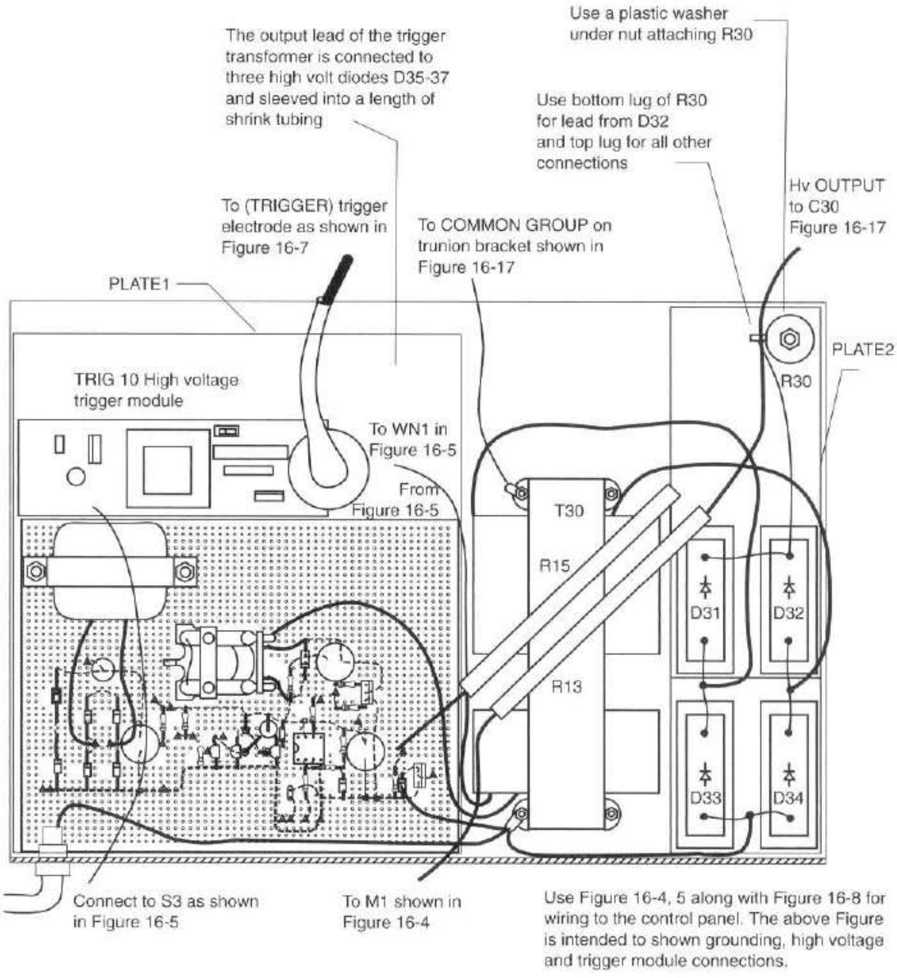

The output lead of the trigger transformer is connected to three high volt diodes D35-37 and sleeved into a length of shrink tubing

Use a plastic washer under nut attaching R30

Use bottom lug of R30 for lead from D32 and top lug for all other connections

To (TRIGGER) trigger electrode as shown in Figure 16-7

To COMMON GROUP on trunion bracket shown in Figure 16-17

Hv OUTPUT to C30 Figure 16-17

PLATE1
TRIG 10 High voltage trigger module
To WN1 in Figure 16-5
From Figure 16-5
T30
R15
R13
R30
PLATE2
D31
D32
D33
D34

Connect to S3 as shown in Figure 16-5

To M1 shown in Figure 16-4

Use Figure 16-4, 5 along with Figure 16-8 for wiring to the control panel. The above Figure is intended to shown grounding, high voltage and trigger module connections.

Figure 16-7 *Assembly and wiring of chassis*

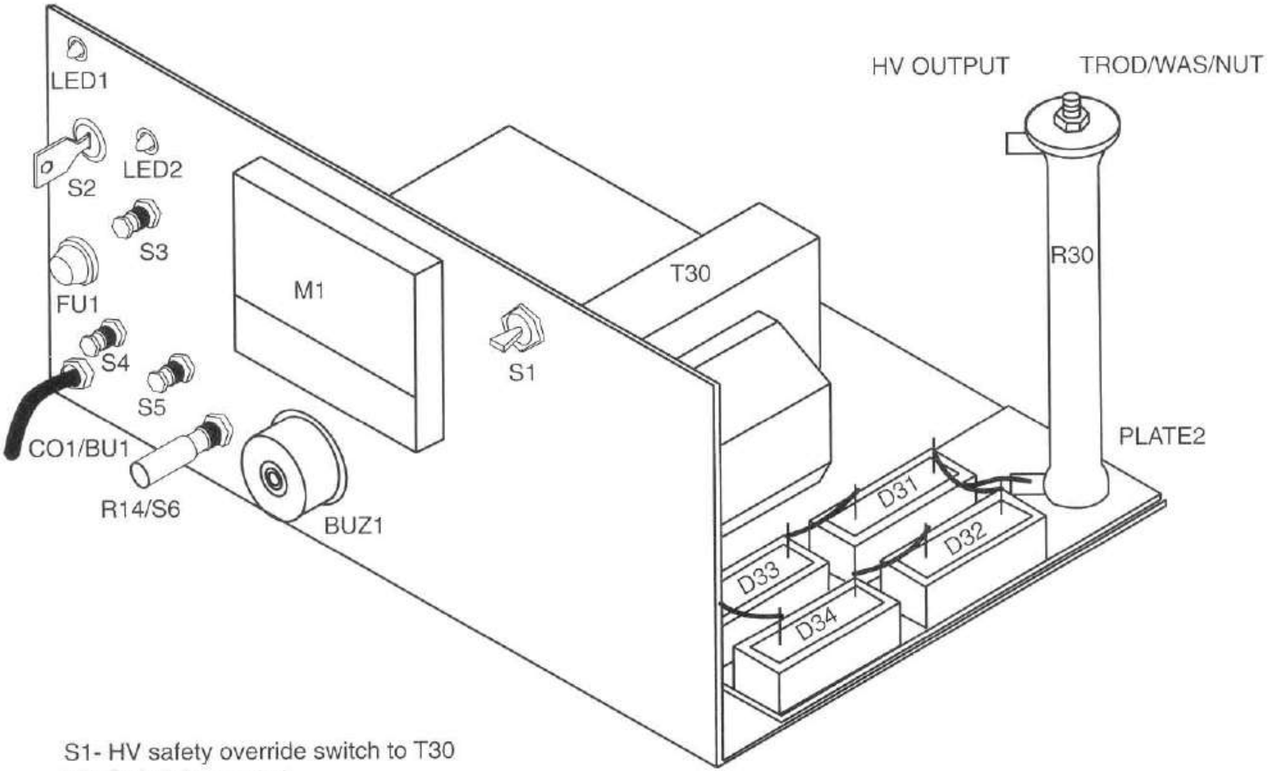

S1- HV safety override switch to T30
S2- Safety key switch
S3- Trigger "fire" switch
S4- Charge start switch
S5- Charge abort switch
S6- Low control voltage switch
LED1- Power "on" indicator
LED2- Charging cycle "on" indicator
R14- Programmed charge voltage level control
FU1- 3 amp fuse holder
BUZ1- High voltage presence audible warning
M1- Charge voltage meter.. 5 kilovolt full scale
CO1/BU1- Clamp bushing for power cord CO1

Note that it is a good idea to label the controls

Figure 16-8 *Front-panel controls and wiring*

16. Fabricate the capacitor bracket, as shown in Figure 16-10. This piece secures the three energy storage capacitors C10A, C10B, and C10C on the bottom deck.

17. Fabricate the trunion bracket, as shown in Figure 16-11. This piece secures the actual flyway and accelerating coil and therefore must be able to withstand the kinetic recoil from the accelerating reaction of the projectile. This bracket is the central electrical grounding point of the discharge circuit and must also be actively grounded to earth ground via the power cord green wire. There is also a $^3/_8$-inch hole for the electrode holding collar (COLLAR1) for the ground electrode (TUNG38) of the main spark switch (GAP1). This part must be centered and carefully soldered in place using a propane torch.

18. Fabricate the front and rear disk section, as shown in Figure 16-12. You may also fabricate the projectile (PROJ1) from a piece of aluminum, as shown.

19. Wind the coils, as shown in Figure 16-13. It is a good idea to study Figures 16-12, 16-13, and 16-14 before attempting this step. Make sure to leave connection leads that are a minimum of 10 inches long.

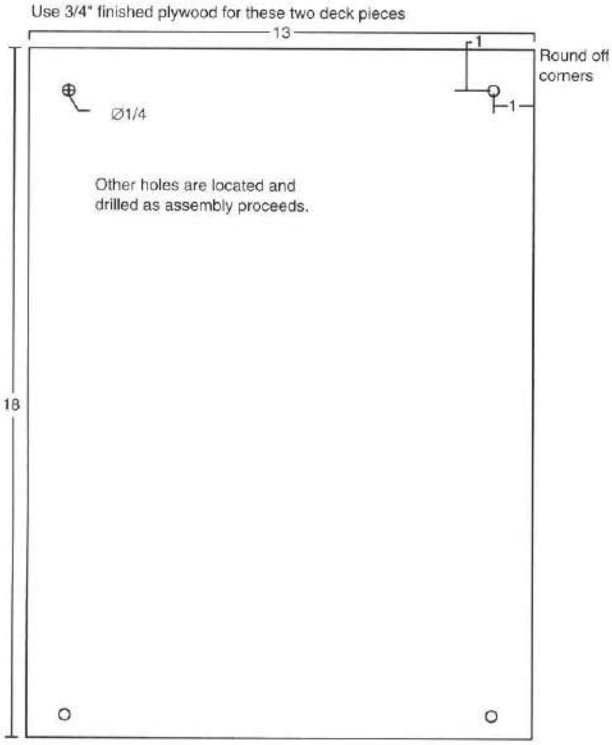

Use 3/4" finished plywood for these two deck pieces
13
1
Round off corners
1
Ø1/4
Other holes are located and drilled as assembly proceeds.
18

Figure 16-9 *Fabrication of deck sections*

20. Assemble the flyway and breech, as shown in Figure 16-14. Study the notes on this figure.

21. Fabricate the two plastic blocks (PVCBLK and TFBLK), as shown in Figure 16-15, to the suggested dimensions. Note that the heights of these pieces must allow close alignment of the electrodes as dictated by the position of the soldered collar on the trunion bracket. C31 and C32 are wired in place and must be spread apart to prevent sparking. The blocks are secured to the top deck using sheet metal screws. Use two screws for the PVC block. Do not allow these screws to penetrate more than ½ inch into the plastic material. Note the clearance hole in the deck for the high-voltage lead from the trigger module, shown in Figure 16-7.

22. Assemble the three energy-discharge capacitors (shown in Figure 16-16) along with the charger and control module.

23. Make the final assembly steps and wire, as shown in Figure 16-17. Note the direct and

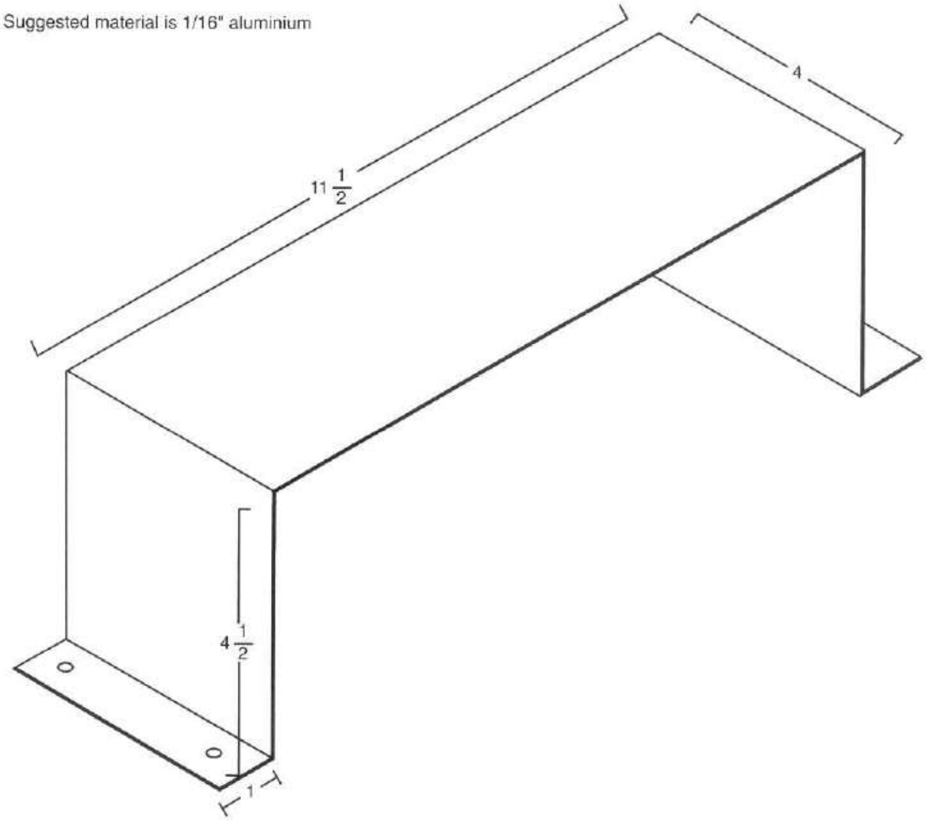

Suggested material is 1/16" aluminium
4
11 1/2
4 1/2
1

Figure 16-10 *Fabrication of the capacitor bracket*

Suggested material is 1/16" galvanized steel

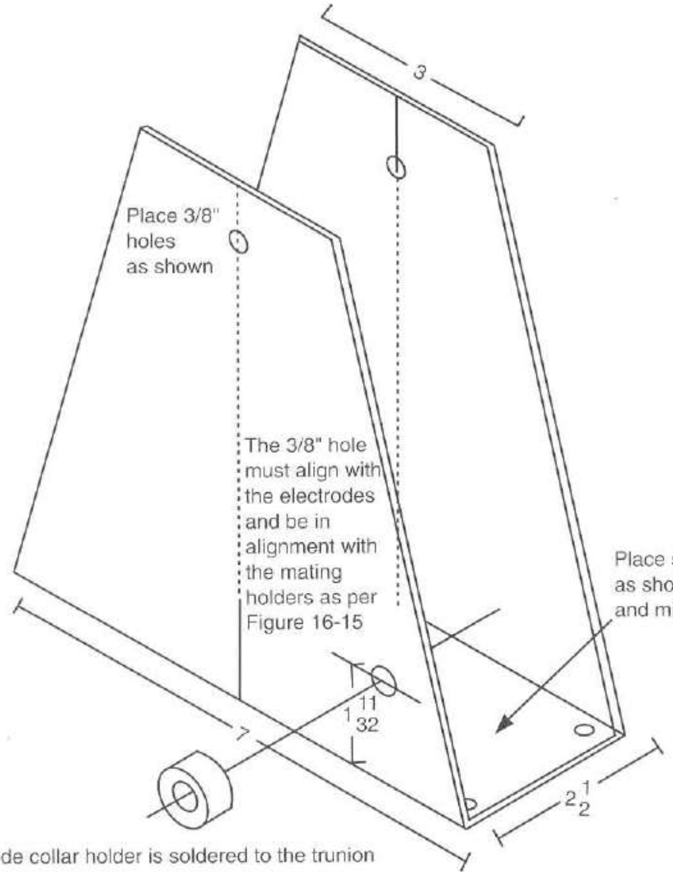
3
Place 3/8" holes as shown
The 3/8" hole must align with the electrodes and be in alignment with the mating holders as per Figure 16-15
Place six 3/16" holes as shown at ends and mid section.
$1\frac{11}{32}$
7
$2\frac{1}{2}$

The electrode collar holder is soldered to the trunion bracket using a propane torch and carefully centering with clearance hole.

Figure 16-11 *Fabrication of the trunion bracket*

heavy leads for the discharge path. You may want to sleeve some plastic tubing over these leads, as they are rated only for 600 volts.

Note: The trunion bracket is the common grounding point for the system. It is very important that the earth ground green lead of the power cord be firmly attached at this point to ensure system operating safety.

24. Verify and use the diagram shown in Figure 16-18 as a wiring aid and reconfirm grounding and general circuitry and integrity.

Testing and Operation of the Cannon

25. Locate a test site with a backstop capable of stopping the projectile. Obtain a 5,000 volt DC reading meter to set M1. It is assumed that the circuitry is correctly assembled and the basic electrical pretest as described in steps 10 through 14 was successively performed. Preset the main gap to $^1/_{16}$ inch and the trigger gap to $^1/_8$ inch.

26. Preset all switches shown back in Figure 16-8 to the off position. Familiarize yourself with these controls

Suggested material is 3/4" plywood or machined PVC

The accelerator coil construction is very important to the overall performance and efficiency of the system. These pieces as shown are fabricated from a high grade birch plywood. You may use polycarbonate for strength or PVC for cost. The objective is to provide a cavity for the flat pan cake shaped windings. The parts shown are 7 inch overall diameter of 3/4" thickness. Both pieces have all holes as shown and must be accurately aligned with one another. The front piece has the 5/16" deep cavity milled out for nesting of the coil windings. Leave a 1" shoulder allowing a 5 inch winding diameter. Both center holes should allow a reasonably tight sliding action for the 1 7/8" PVC fly way tube. Note the two feed through holes for the accelerator coil leads.

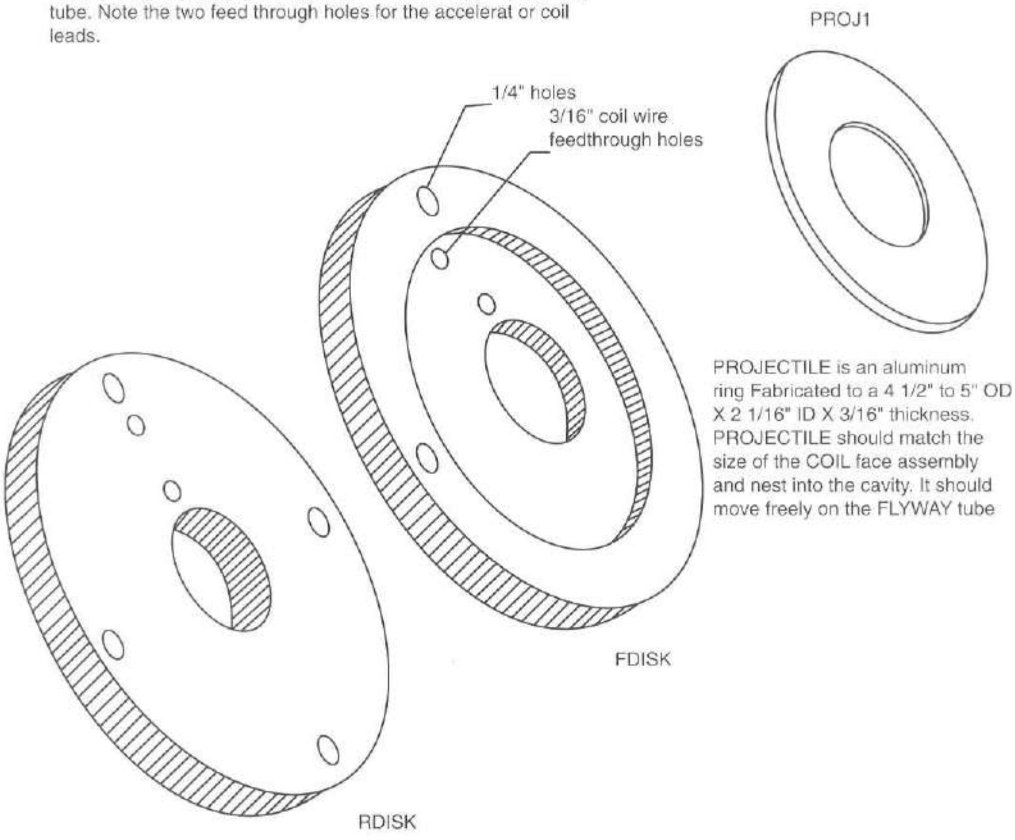

PROJECTILE is an aluminum ring Fabricated to a 4 1/2" to 5" OD X 2 1/16" ID X 3/16" thickness. PROJECTILE should match the size of the COIL face assembly and nest into the cavity. It should move freely on the FLYWAY tube

Figure 16-12 *Fabrication of the accelerator coil holder*

Danger: The next steps involve exposure to dangerous high voltages. Failure to take any of the following described steps will require a complete shutdown and power removal, along with a shorting of C30 before performing any trouble shooting.

27. Connect the test meter from chassis ground to the output end of resistor R30. Plug the unit into 115 vac and turn on S6. Turn on the key switch S2 and note LED1 lighting. Place the projectile over the accelerator coil.

28. Depress pushbutton switch S3 and note a spark occurring in both gaps.

29. Verify S1 safety switch is off. Depress pushbutton switch S4 and note the relay RE1 energizing and LED2 lighting. Depress pushbutton switch S5 and note RE1 de-energizing and LED2 turning off. Verify these on and off functions of S4 and S5 several times before the next steps.

    S5 is important as it must stop the charging action in the event that you choose to quickly abort the cycle.

30. Turn on toggle switch S1. Restart the system by again depressing R4, noting a voltage reading on both the panel meter M1 and the external test meter. This voltage will build up to a level as set by the charge-voltage level control R14 and then shut down, only to restart the charge cycle when the voltage charge level

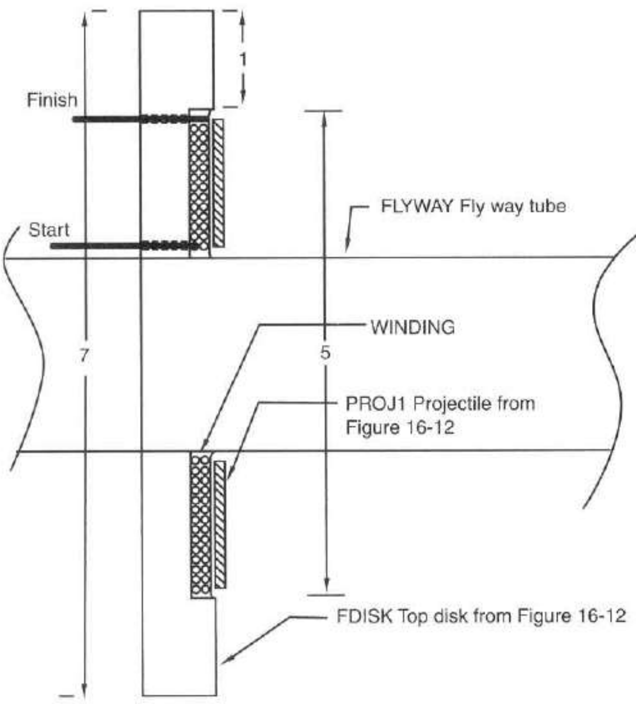
1
Finish
Start
7
5
FLYWAY Fly way tube
WINDING
PROJ1 Projectile from Figure 16-12
FDISK Top disk from Figure 16-12

1. Fabricate the FLYWAY from a 18" piece of 2" PVC tubing
2. Wind a coil of 30 turns of #12 in fifteen layers of 2 turns each. Finished coil must be in a tight pancake configuration and have 10 inch connection leads. *You may have to make up a temporary bobbin jig for this step.* Hold together with tape etc.
3. Place COIL section into cavity routing the leads as shown. COIL section should nest into cavity and be positioned flat.
4. Seal all points that can leak as epoxy must be carefully poured over the COIL forming a 1/16" layer on top.

Figure 16-13 *Side view showing winding*

Suggested material is ¾" plywood or machined PVC

You will note that the front disk housing the accelerator coil is allowed to move over the flyway tube and against the springs to absorb some of the recoil. The rear disk abuts to the shoulder of the breech tube that is non movable along with the flyway tube.

This feature in important if you are intending to repetitively fire high energy shots..

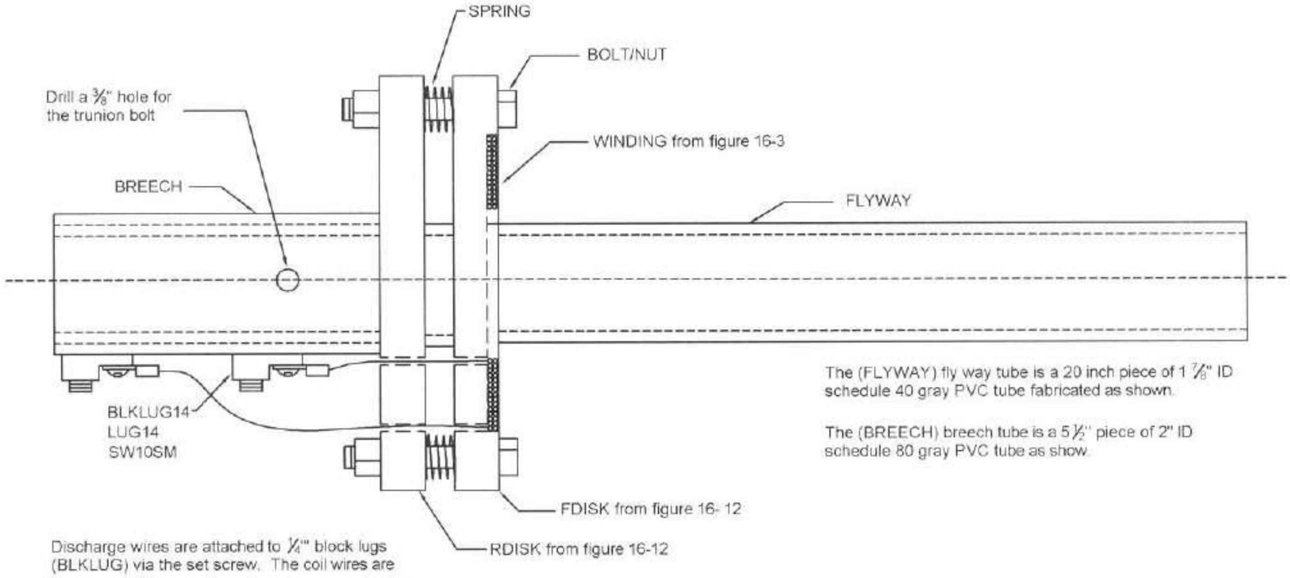
SPRING
BOLT/NUT
Drill a ⅜" hole for the trunion bolt
WINDING from figure 16-3
BREECH
FLYWAY
BLKLUG14
LUG14
SW10SM
FDISK from figure 16- 12
RDISK from figure 16-12

The (FLYWAY) fly way tube is a 20 inch piece of 1 ⅞" ID schedule 40 gray PVC tube fabricated as shown.

The (BREECH) breech tube is a 5 ½" piece of 2" ID schedule 80 gray PVC tube as show.

Discharge wires are attached to ¼" block lugs (BLKLUG) via the set screw. The coil wires are soldered to ¼" spade lugs that are secured to the block lugs via #10 x ¾" sheet metal screws, screwed into both the FLYWAY and BREECH tube.

The bare leads from the coil should be individually sleeved into plastic tubing to prevent catastrophic contact

Figure 16-14 *Assembly of the breech and flyway*

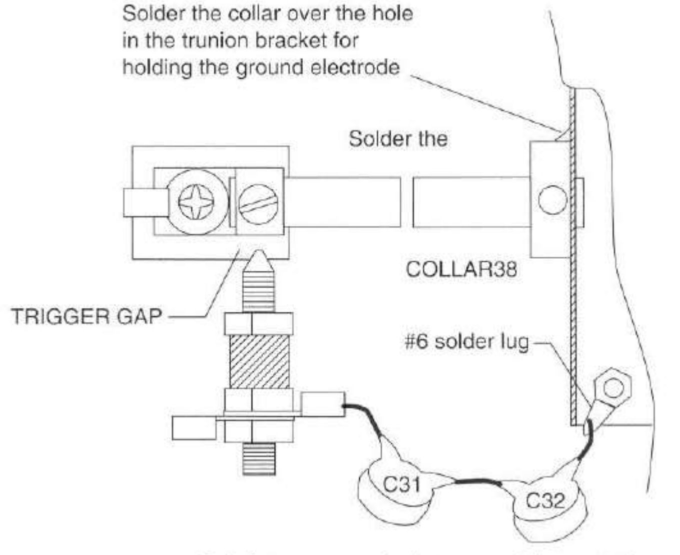
Solder the collar over the hole
in the trunion bracket for
holding the ground electrode
Solder the
COLLAR38
TRIGGER GAP
#6 solder lug
C31
C32

Solder gap switch assembly top view

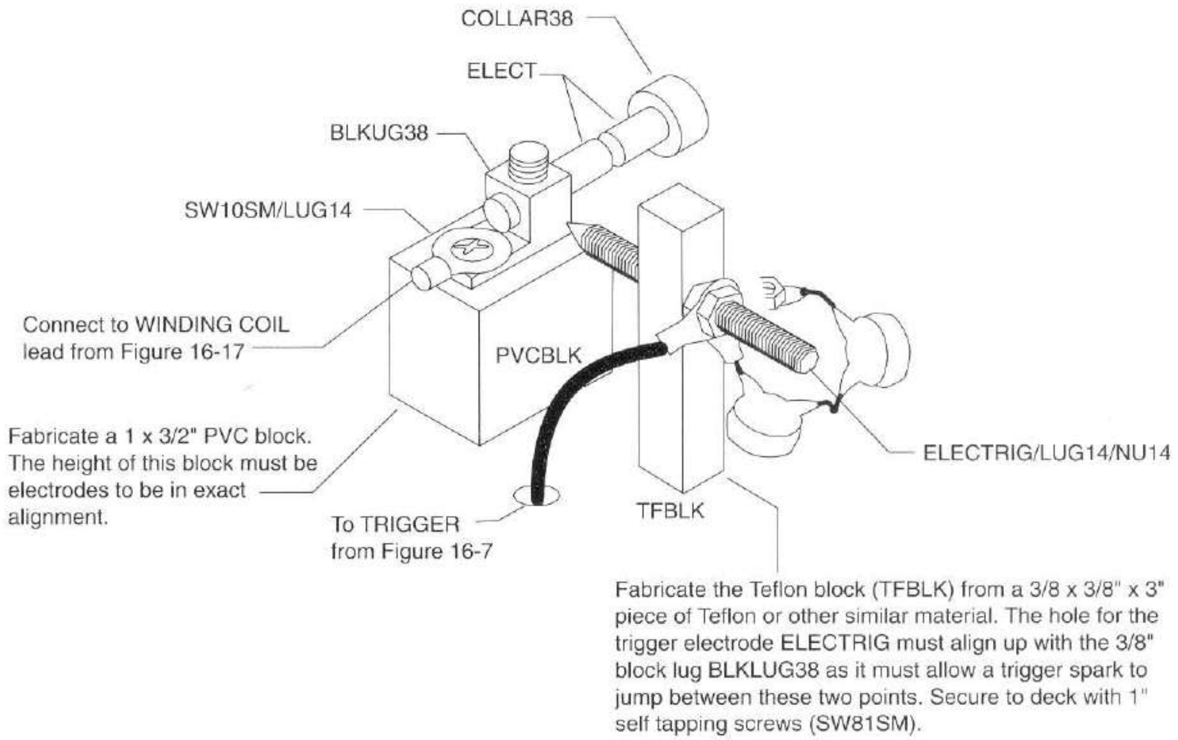
COLLAR38
ELECT
BLKUG38
SW10SM/LUG14
Connect to WINDING COIL
lead from Figure 16-17
PVCBLK
Fabricate a 1 x 3/2" PVC block.
The height of this block must be
electrodes to be in exact
alignment.
To TRIGGER
from Figure 16-7
TFBLK
ELECTRIG/LUG14/NU14
Fabricate the Teflon block (TFBLK) from a 3/8 x 3/8" x 3"
piece of Teflon or other similar material. The hole for the
trigger electrode ELECTRIG must align up with the 3/8"
block lug BLKLUG38 as it must allow a trigger spark to
jump between these two points. Secure to deck with 1"
self tapping screws (SW81SM).

Spark gap switch is o view

Figure 16-15 *Details of the spark and trigger switches*

falls below the set level. Obviously there is an amount of hysteresis that prevents off-and-on chatter. With R14 set to full counterclockwise (CCW), this voltage should reset at less than 1,000 volts. The buzzer BUZ1 should be producing an annoying sound, thus alerting you to a charge presence on the capacitors.

31. Slightly turn R14 clockwise and allow it to cycle, repeating up to 1,500 volts as read on the external test meter. Depress the trigger

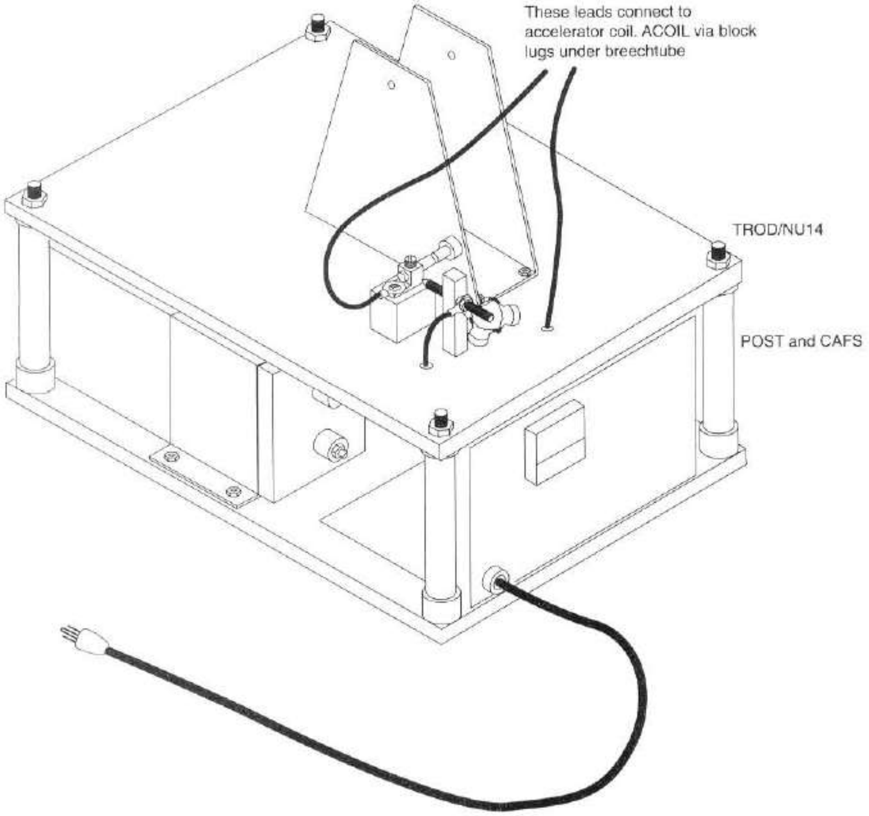
These leads connect to accelerator coil. ACOIL via block lugs under breechtube
TROD/NU14
POST and CAFS

Figure 16-16 *Final assembly isometric view*

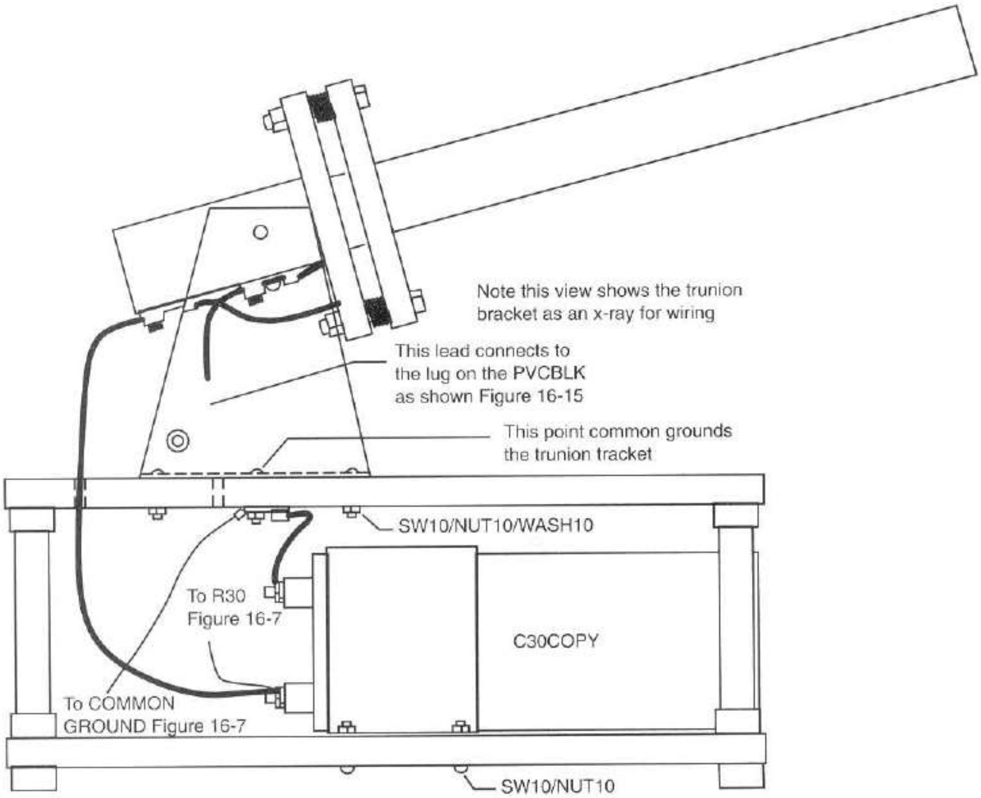
Note this view shows the trunion bracket as an x-ray for wiring
This lead connects to the lug on the PVCBLK as shown Figure 16-15
This point common grounds the trunion tracket
SW10/NUT10/WASH10
To R30 Figure 16-7
C30COPY
To COMMON GROUND Figure 16-7
SW10/NUT10

Figure 16-17 *Final assembly side view*

fire pushbutton S3 and note the projectile shooting into the backstop. Immediately press the stop switch S5 unless you want it to charge up for the next shot.

Note: The system will automatically recharge to the 1,500-volt present level unless the stop switch is activated.

32. Repeat the cycle, increasing the charge voltage up to 2,500 volts, and then adjust the trimpot R16 on the meter M1 to agree with the test meter.

33. Repeat this step allowing the charge voltage to go up to 5,000 volts, noting the impact of the projectile.

34. Verify the unit is not breaking apart as a result of recoil. Rethink the recoil springs if a problem exists.

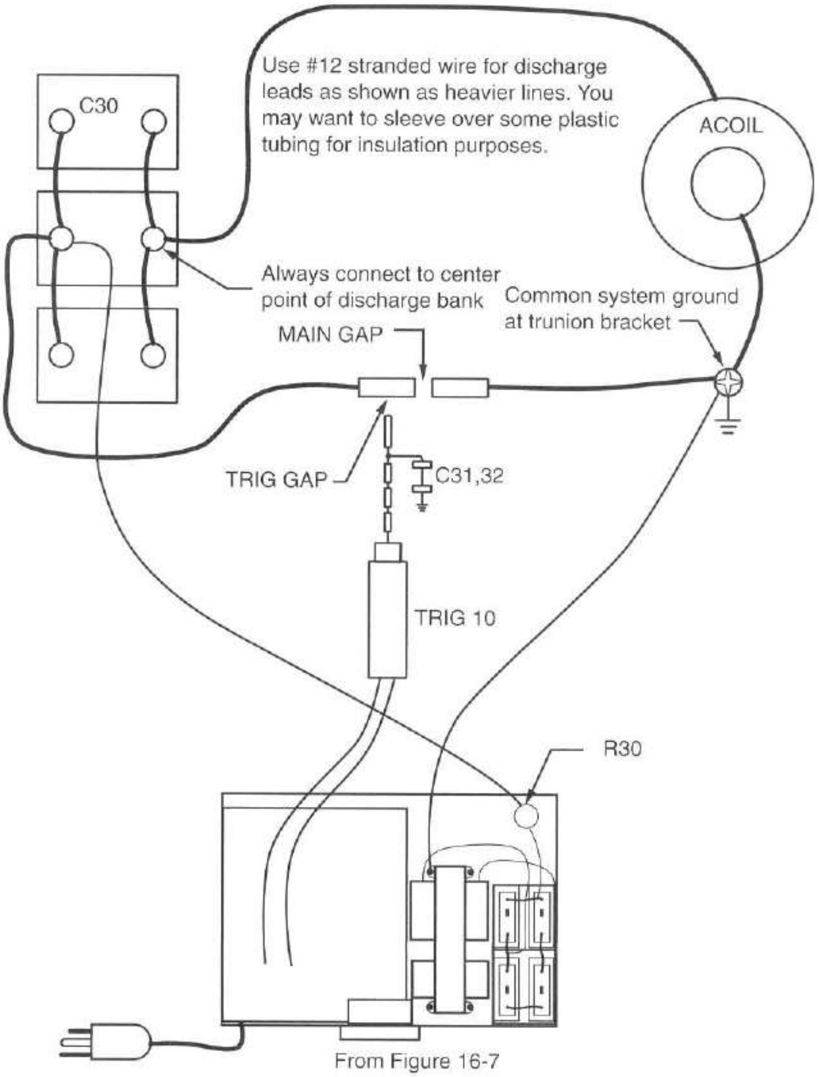
C30
Use #12 stranded wire for discharge leads as shown as heavier lines. You may want to sleeve over some plastic tubing for insulation purposes.
ACOIL
Always connect to center point of discharge bank
Common system ground at trunion bracket
MAIN GAP
TRIG GAP
C31,32
TRIG 10
R30
From Figure 16-7

Figure 16-18 *Overall system wiring diagram*

Table 16.1 Magnetic Cannon Part List

Ref. #	Description	DB Part #
R1, 4, 9, 19	Four 470, ¼-watt resistors (yel-pur-br)	
R2	1K, ¼-watt resistor (br-blk-red)	
R3, 6, 7, 8	Four 10 kilohm, ¼-watt resistors (br-blk-or)	
R5	220, ¼-watt resistor (red-red-br)	
R10, 11	Two 10M, ¼-watt resistors (br-blk-bl)	
R12	2.2K, ¼-watt resistor (red-red-red)	
R13, 15	Fourteen 1.2M, ¼-watt resistors (br-red-gr) (see schematic in Figure 16-2)	
R14/S6	100K potentiometer, 17 mm with low-volt switch	
R16	2K trimpot	
R20	1M, ¼-watt resistor (br-blk-gr)	
R21	10-ohm, ¼-watt resistors (br-blk-blk)	
R18	12-ohm, 3-watt resistor	
R30	16K, 50-watt power resistor	DB# 16K50W
C1, 4, 5	Three 1,000 mfd, 25-volt vertical electrolytic capacitors	
C2, 3	Two .1 mfd, 50-volt plastic capacitors	
*C30A	Three 32 mfd, 4,500-volt oil-filled cans in parallel for 96 mfd/4.5 kv	DB# 32m/4500
*C30B	Four 24 mfd, 5,200-volt oil-filled cans in parallel for 96 mfd/5 kv	DB# 24m/5200
I1	741 op-amp DIP IC	
Q1, 2, 3	Three PN2222 GP NPN transistors	
Q4	IRF520 Mosfet	
Q5	MJE3055 NPN power transistor	
D1-4	Four 1N4001 50-volt 1-amp diodes	
D5	1N4007 1,000-volt, 1-amp diode	
D31-34	Four 8-kilovolt, 100 ma standard recovery diodes	DB# H407
D35-37	Three 30 kilovolt, 10 ma diodes	DB# VG30
Z1, 2, 3	Three 1N4733 6-volt zener diodes	
Z4	1N4745 15-volt zener diode	
LED1	Red LED indicator	
LED2	Green LED indicator	
S1	SPST toggle switch	
S2	Key switch	
S3, 4, 5	Three pushbutton NO switches	
RE1	12-volt, 5-amp relay	
BUZ1	12-volt buzzer	DB# BUZROUND
Q1	100-volt IRF540 MOSFET	
T1	12-volt, 100 ma transformer	DB# 12/.1
T2	6,500-volt, 20 ma current-limited transformer	DB# 6kv/20
*L10	Accelerator coil assembly, as shown in Figures 16-12 and 16-13	DB# ACCOIL
*GAP1	Main gap assembly, as shown in Figure 16-15	DB# GAPMAIN
*GAP2	Trigger gap assembly, as shown in Figure 16-15	DB# GAPTRIG
CHASSIS1	Metal chassis fabricated as shown in Figure 16-6	
PLATE1	6 × 7½ × 1/16 inch plastic insulating plate	
PLATE2	2¼ × 7½ × 1/16 inch plastic insulating plate	
TAPE1	12 inches of 1-inch wide, thin sticky tape	
PLTUBE	18 inches of ¼-inch ID vinyl flexible tubing for R13, R15, and trigger diodes	
CO1	6 feet of 3-wire #18 power cord	
HVWIRE	12 inches 20 kv silicon high-voltage wire	DB# WIRE20KV
BUSSWIRE	24 inches #20 bare busswire	

WR20	10 feet #20 vinyl stranded hook-up wire	
WN1	Medium wire nut	
BU2, 3	Two bushings for LEDS	
BU1	Power cord clamp bushing	
SW6	Six 6-32 × 3/4-inch screws	
SW1	6-32 × 1/4-inch screw for mounting relay	
NU6	Six #6 hex nuts	
LUG6	#6 solder lug	
TROD	5-inch length of 1/4-20 threaded rod, 2 nuts, and plastic washers	
TRIG10	Assembled trigger pulse module	DB#TRIG10
TRIG1K	Kit of trigger pulse module	DB#TRIG1K

Parts List to Final System

DECKS	Two fabricated pieces of 7/8-inch birch plywood 13 × 17 inches
POSTS	Four fabricated pieces of 5 3/4 × 7/8 inch PVC tubing
PLATE1	Fabricate a piece of 1/16-inch plastic sheet 6 × 7 inches
PLATE2	Fabricate a piece of 1/16-inch plastic sheet 2 1/4 × 7 inches
BRKCAP1	Fabricate the capacitor bracket seen in Figure 16-10
BRKTRUN1	Fabricate the trunion bracket seen in Figure 16-11
RDISK	Fabricate the rear disk seen in Figure 16-12
FDISK	Fabricate the front disk seen in Figure 16-12
PROJ1	Fabricate the projectile seen in Figure 16-12
BLKELECT	Fabricate the electrode holding block seen in Figure 16-15
BLKTRIG	Fabricate the trigger electrode holding block seen in Figure 16-15
FLYWAY	Fabricate the flyway tube seen in Figure 16-14
BREECH	Fabricate the breech tube seen in Figure 16-14
CAPS	Twelve 7/8-inch PVC caps for posts and feet seen in Figure 16-1
SPRINGS	Four 1/2-inch diameter × 3/4-inch stiff compression springs
COLLAR	3/8-inch shaft collar
BLKLUG25	Three 1/4-inch block lugs for breech connections
BLKLUG38	3/8-inch block lug for electrode
LUG14	Five 1/4-inch wide solder lugs
ELECTRGAP	Two 3/8 × 2-inch tungsten or tool steel electrodes
ELECTRTRIG	1/4-20 × 3-inch threaded rod with point
TROD	Four 1/4-20 × 8-inch threaded rod
BOLT/NUT	Five 3 1/2 × 3/8-inch bolt and nut
NUT14	Eleven 1/4 × 20 nuts for trigger gap and posts
SW10M	Five #10 × 1/2-inch sheet metal screws
SW10	Ten #10-32 × 1.5-inch machine screw
NU10	Ten #10-32 nuts
WASH10	Twenty #10 wide shoulder washers
WIRE12	4 feet #12 stranded, vinyl-covered wire
WIRE12	30 feet #12 heavy-coated, magnetic wire for winding ACOIL

Chapter Seventeen

Vacuum Tube Tesla Coil Project

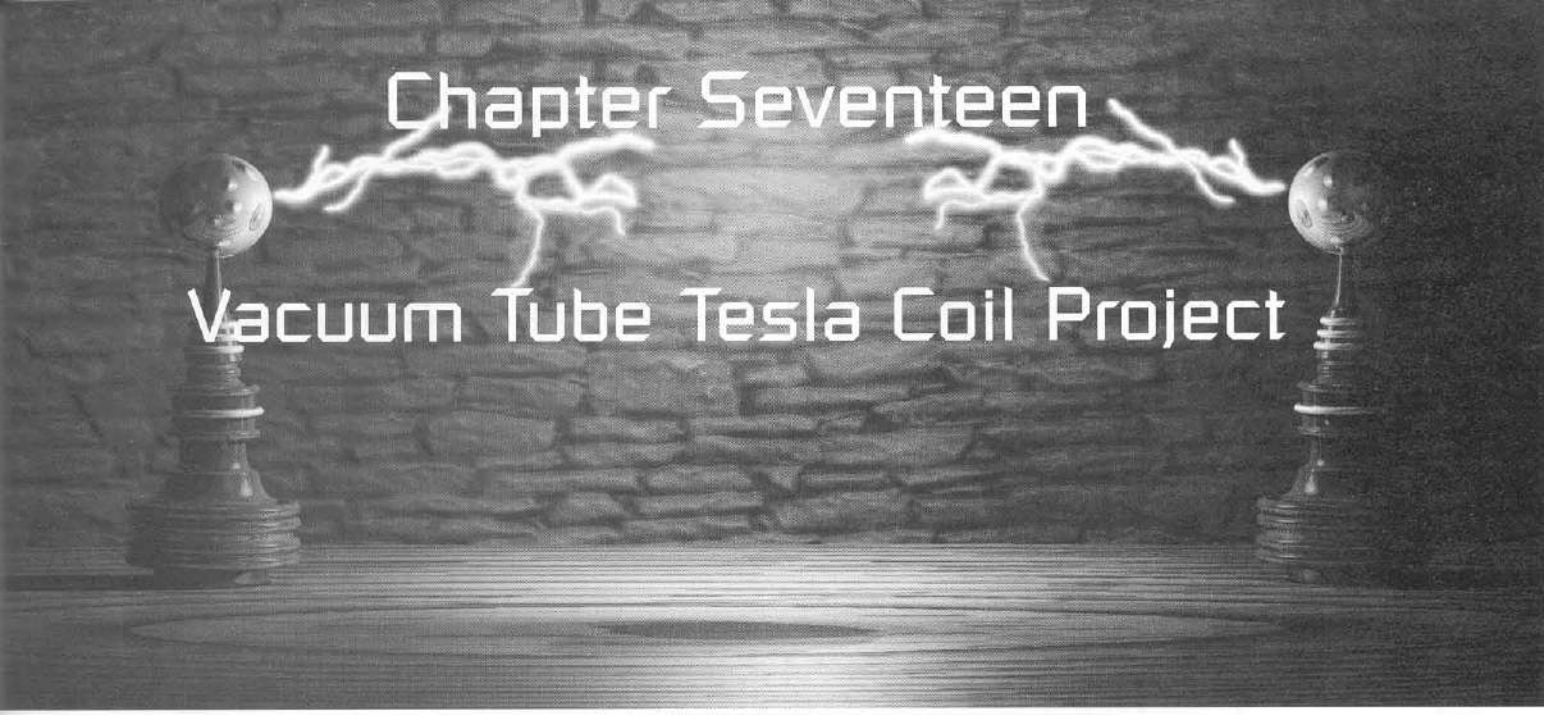

This advanced project, as shown in Figure 17-1, is similar to a coil made by John Weisner some years ago. The geometry of the secondary not being the most efficient is a good compromise for wireless energy transmission and also provides a reasonable output display. The project is intended for the experienced experimenter and researcher in the field of high-voltage and high-frequency circuitry. Not only is this project a dangerous electrical device, but it can also be a bothersome source of radio interference, especially in the standard AM broadcast band and should be used in a shielded enclosure or a Faraday cage if operated on a continuous basis. A gas discharge lamp such as a standard household fluorescent lamp will glow up to 10 feet away when this device is used. Annoying burns are experienced when contact is made with conductive objects up to similar distances.

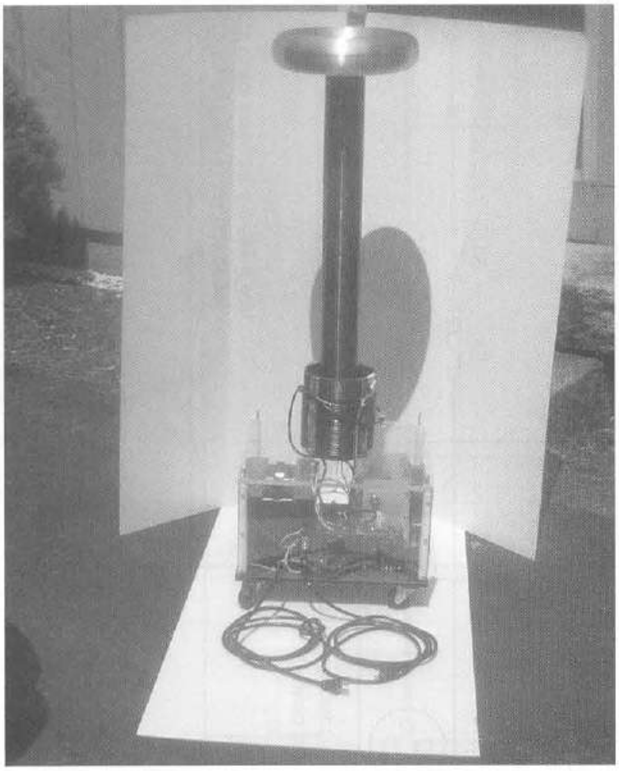

Figure 17-1 *The vacuum tube Tesla coil*

We also have built coils using 304TH, 450TH, and parallel 2500THs with a full 5,000 watts of plate dissipation. This coil produced spectacular flaming arcs up to 8 feet long with 8 kilowatts of input. A video showing the operation is on our web site. The coil shown here will easily produce 1 to 2 feet of flaming arcs.

This system can provide the experienced hobbyist with hours of fun and entertainment experimenting with terminals and settings to generate wireless energy as well as output arcs and sparks. This is an advanced-level and potentially expensive project requiring electronic skills and high-voltage experience. Expect to spend $300 to $750. All the parts are readily available, with specialized parts obtainable through Information Unlimited (www.amazing1.com), and they are listed in Table 17-1.

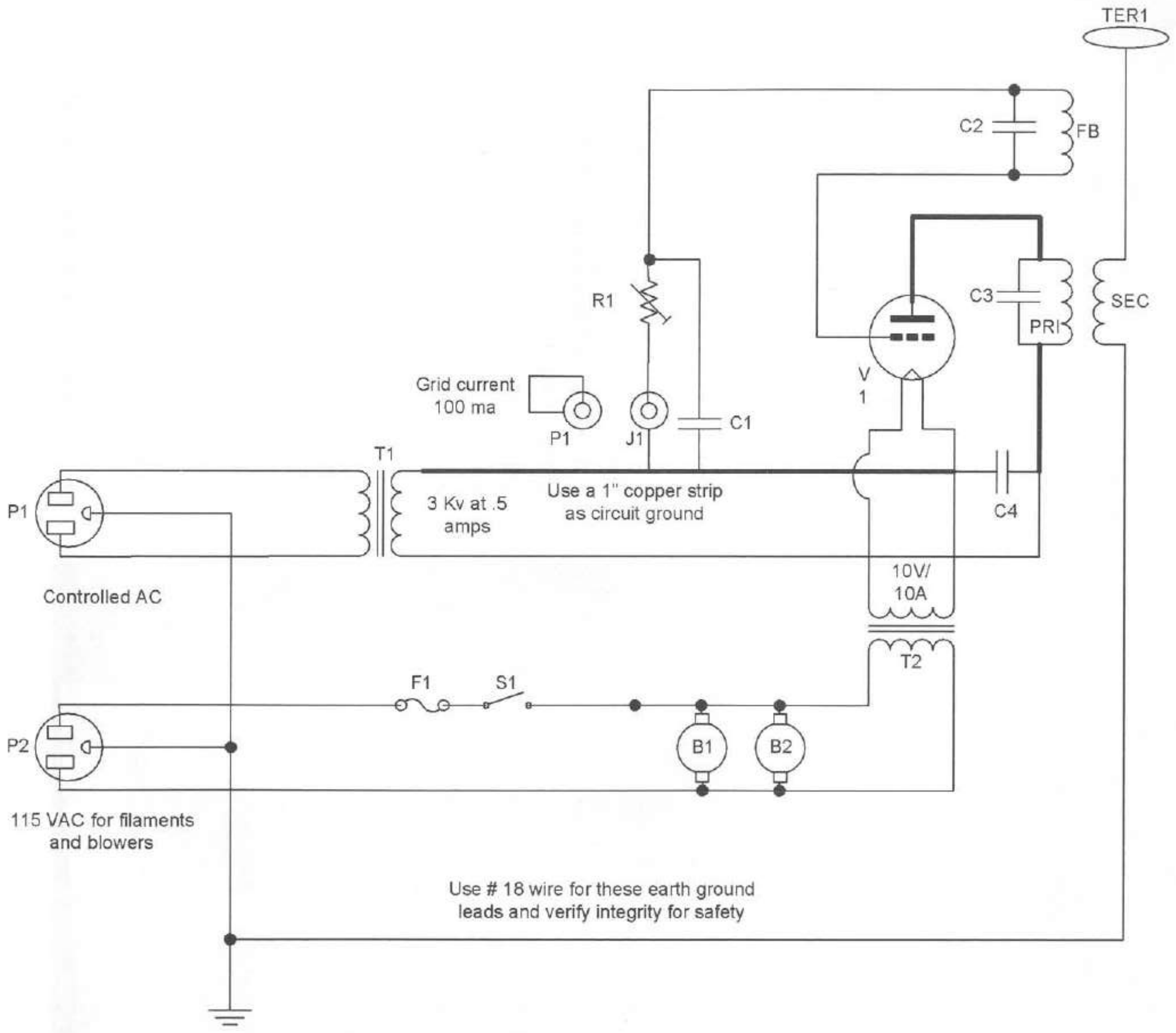
TER1
C2
FB
R1
C3
SEC
PRI
V
1
Grid current
100 ma
P1
J1
C1
T1
3 Kv at .5
amps
Use a 1" copper strip
as circuit ground
C4
P1
Controlled AC
10V/
10A
T2
F1
S1
P2
B1
B2
115 VAC for filaments
and blowers
Use # 18 wire for these earth ground
leads and verify integrity for safety

Feed back winding is 10 to 20 turns of #18 stranded vinyl wire. Open circuit inductance is approximately 75 micro-henries

Primary windings are 20 to 22 turns of #8 insulated wire. Open circuit inductance 55 micro-henries

Secondary winding is a 33" length of tightly wound #28 magnet wire for about 2000 turns. Free standing resonant frequency is approximately 200 kHz.

Resistor R1 is 5000 ohms at 100 watts

Figure 17-2 *Circuit schematic*

Basic Description

These spectacular display devices produce arcs and sparks quite unlike the undamped spark gap-driven Tesla coils. Operation does not require a noisy spark gap that produces copious amounts of *radio frequency interference* (RFI) but operates efficiently at the quarter-wave frequency of the secondary coil. Vacuum tubes, while large and requiring filament power, do offer robust and forgiving operation and are intended as a good starter approach to solid-state Tesla coils.

This approach is far less temperamental than the solid-state *metal-oxide-semiconductor field effect transistors* (MOSFET) or *insulated gate bipolar transistor* (IGBT) drivers. However, once these FET devices are debugged and tweaked, one can obtain some excellent results. Our lab FET coil generates over 4-foot sparks using only 3 kilowatts. The output frequency tracks within 5 percent of the resonant point using one of our proprietary circuits that will be on our site in the late spring of 2006 with the complete system scheduled for publishing in the third edition of this book series.

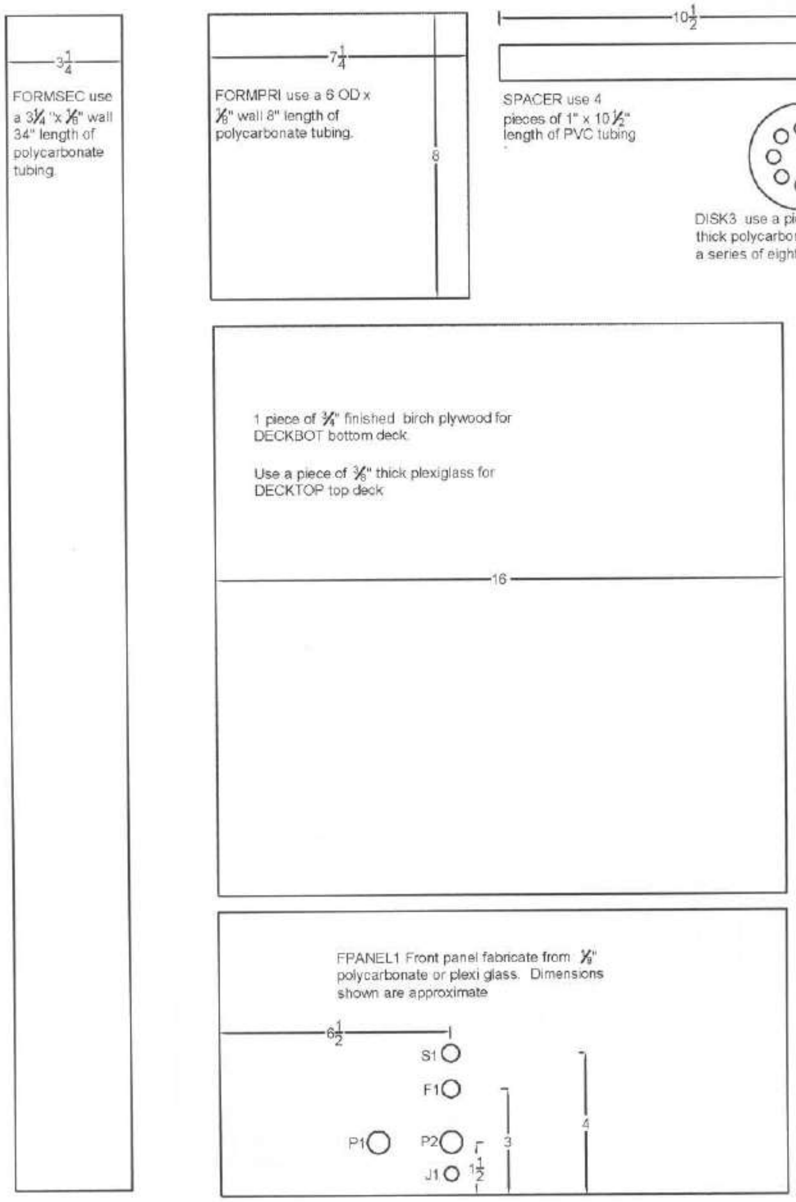
3¼
FORMSEC use a 3¼ "x ⅛" wall 34" length of polycarbonate tubing.
7¼
8
FORMPRI use a 6 OD x ⅛" wall 8" length of polycarbonate tubing.
10½
1
SPACER use 4 pieces of 1" x 10½" length of PVC tubing
DISK3 use a piece of 3" x ¼" thick polycarbonate disk. Drill a series of eight air flow holes
1 piece of ¾" finished birch plywood for DECKBOT bottom deck.
Use a piece of ⅜" thick plexiglass for DECKTOP top deck
16
FPANEL1 Front panel fabricate from ⅛" polycarbonate or plexi glass. Dimensions shown are approximate
6½
S1
F1
P1
P2
J1
1½
3
4

Figure 17-3 *Creating the major parts*

The circuit as shown on the schematic in Figure 17-2 of the vacuum tube device is nothing more than a medium-powered Hartley radio frequency oscillator tuned to the resonant frequency of the secondary coil. The circuit uses a medium-powered, readily available 833A triode transmitting tube that inherently has a high grid to plate capacitance necessary for self-oscillation. The plate load is a resonant transformer with both the primary and secondary tuned to the approximate identical frequencies. The output potential is now a function of the secondary-to-primary coil inductances times the peak primary plate voltage. This system generates voltages at frequencies sufficient to cause air breakdown, thus producing the sparks and arc jumping into open air.

The output of the oscillator is relatively closely coupled to the secondary coil designed for high Q performance and self-resonant to the quarter-wave of the oscillator frequency. The voltage distribution is now that of a quarter section with a current node at the base and voltage node at the top. The input power to the coil is raw, unrectified AC at 3,000 volts *root mean square* (rms) with a current of .5 amps. The peak voltage is over 4,000 volts and is supplied by a conventional plate transformer being fed by a voltage-adjustable variac, which allows you to adjust the output voltage of the system. A pulse signal is also shown that controls the grid of the tube, allowing a wide range of spark texture variation by changing the duty cycle and frequency.

Assembly Steps

1. Fabricate the secondary coil form (FORMSEC) from a $3^1/_4 \times {}^1/_8$-inch wall polycarbonate tubing, as shown in Figure 17-3. It is not suggested that you use PVC tubing due to moisture retention and breakdown.

   Fabricate the primary coil form (FORMPR1)as shown in Figure 17-3 from a length of $7^1/_2 \times {}^1/_4$-inch wall plastic tubing. Polycarbonate tubing is preferred.

   Fabricate two 3-inch diameter discs (DISK3) as shown in Figure 17-3 from $^1/_4$-inch polycarbonate. These pieces are for the ends of the secondary coil and are secured via three #6 × $^3/_8$-inch brass wood screws. Drill air passage holes as shown for the blower (B2).

   Fabricate four 10.5-inch spacers (SPACER) as shown in Figure 17-3 from lengths of 1-inch OD PVC tubing.

   Fabricate the 16 × 16-inch square bottom deck (DECKBOT) as shown in Figure 17-3 from a piece of $^3/_4$-inch finished plywood and paint it black.

   Fabricate the 16 × 16-inch deck top (DECKTOP) as shown in Figure 17-3 from a $^3/_8$-inch-thick Plexiglas or you may use plywood.

2. Fabricate the parts as shown in Figure 17-4. Note to use $^1/_{16}$-inch G10 or another high-temperature plastic for the actual tube

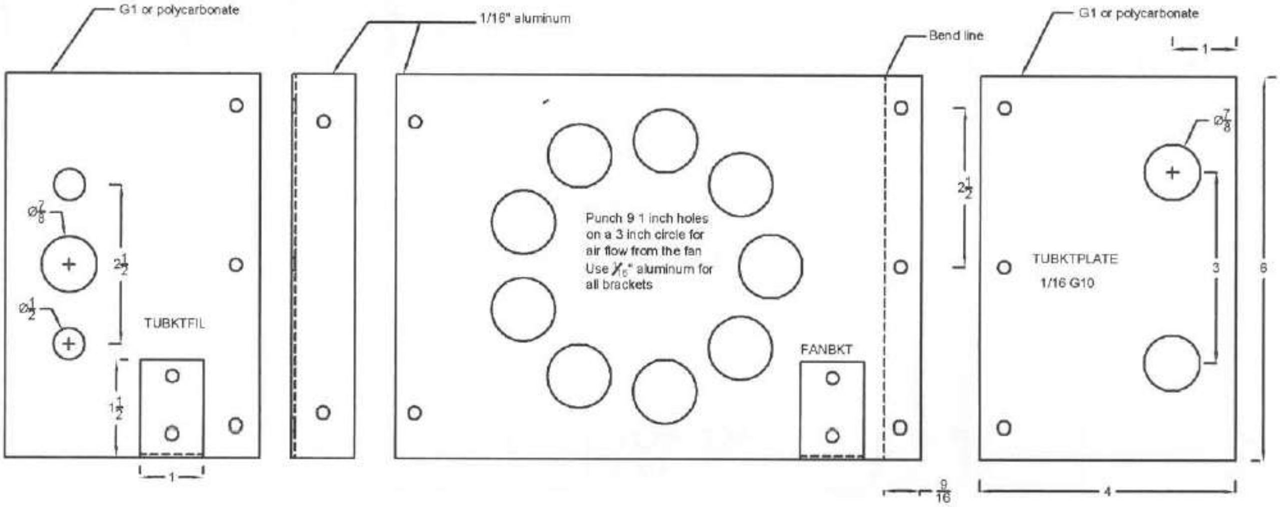

Figure 17-4 *Assembly of the tube and mounting bracket*

holders (TUBKTFIL and TUBKTPLATE). Use 1/16-inch aluminum for the other parts. Note the mating holes for assembling them together, as shown in Figure 17-4.

3. Assemble the fan and tube-mounting bracket from the parts in Figure 17-5. Use thirteen 6-32 × 1/2-inch screws and nuts. Foot brackets secure the assembly to the wooden deck via four #8 × 1/2-inch wood screws. TUBKTPLATE must be disassembled to insert the tube V1 in place. The blower fan B1 is centered to maximize air flow across the plate section of V1. It is mounted using four 8-32 × 1/2-inch screws and nuts. The entire assembly

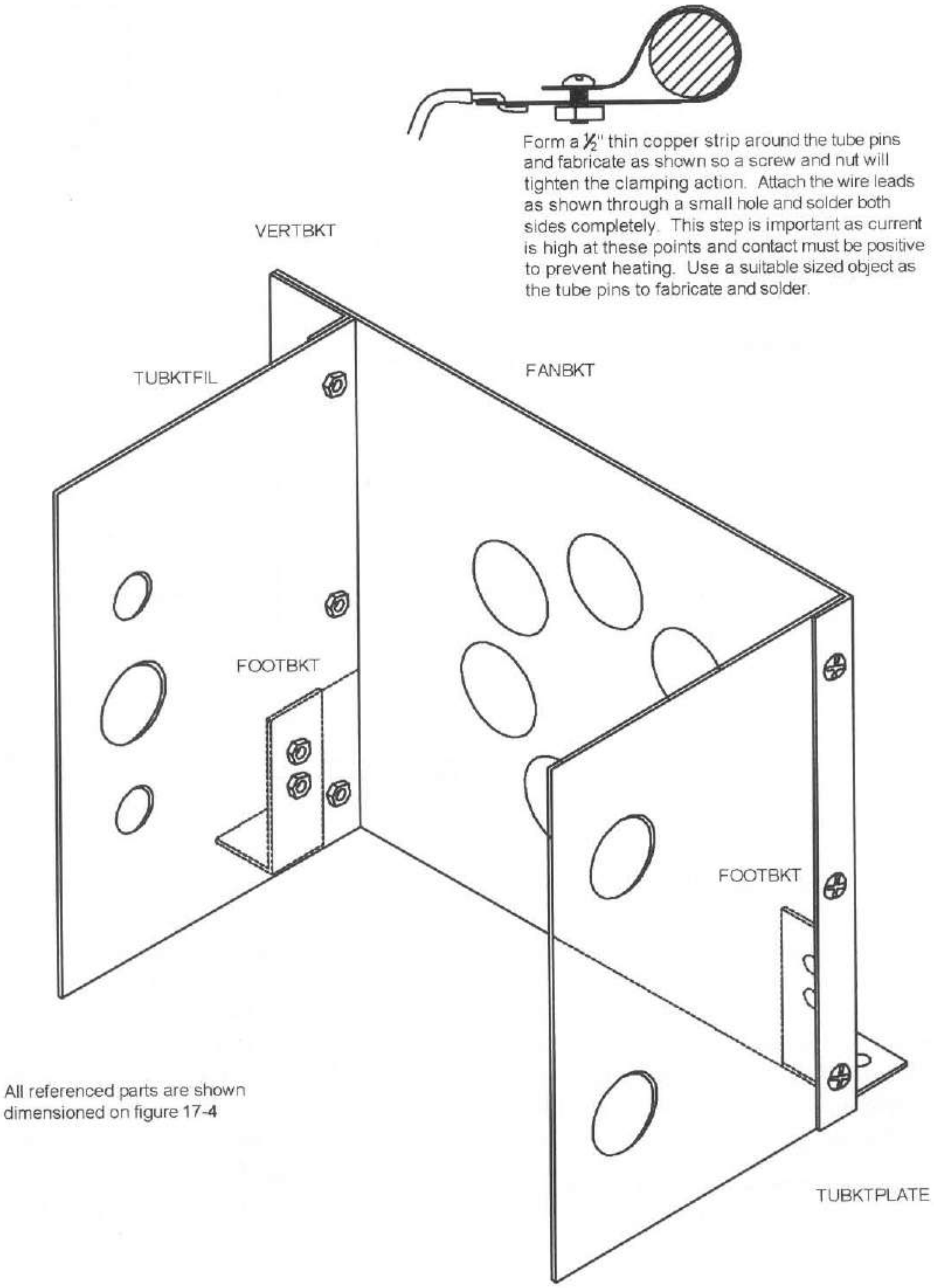
Form a ½" thin copper strip around the tube pins and fabricate as shown so a screw and nut will tighten the clamping action. Attach the wire leads as shown through a small hole and solder both sides completely. This step is important as current is high at these points and contact must be positive to prevent heating. Use a suitable sized object as the tube pins to fabricate and solder.
VERTBKT
TUBKTFIL
FANBKT
FOOTBKT
FOOTBKT
TUBKTPLATE
All referenced parts are shown dimensioned on figure 17-4

Figure 17-5 *Assembly of the tube-holding bracket*

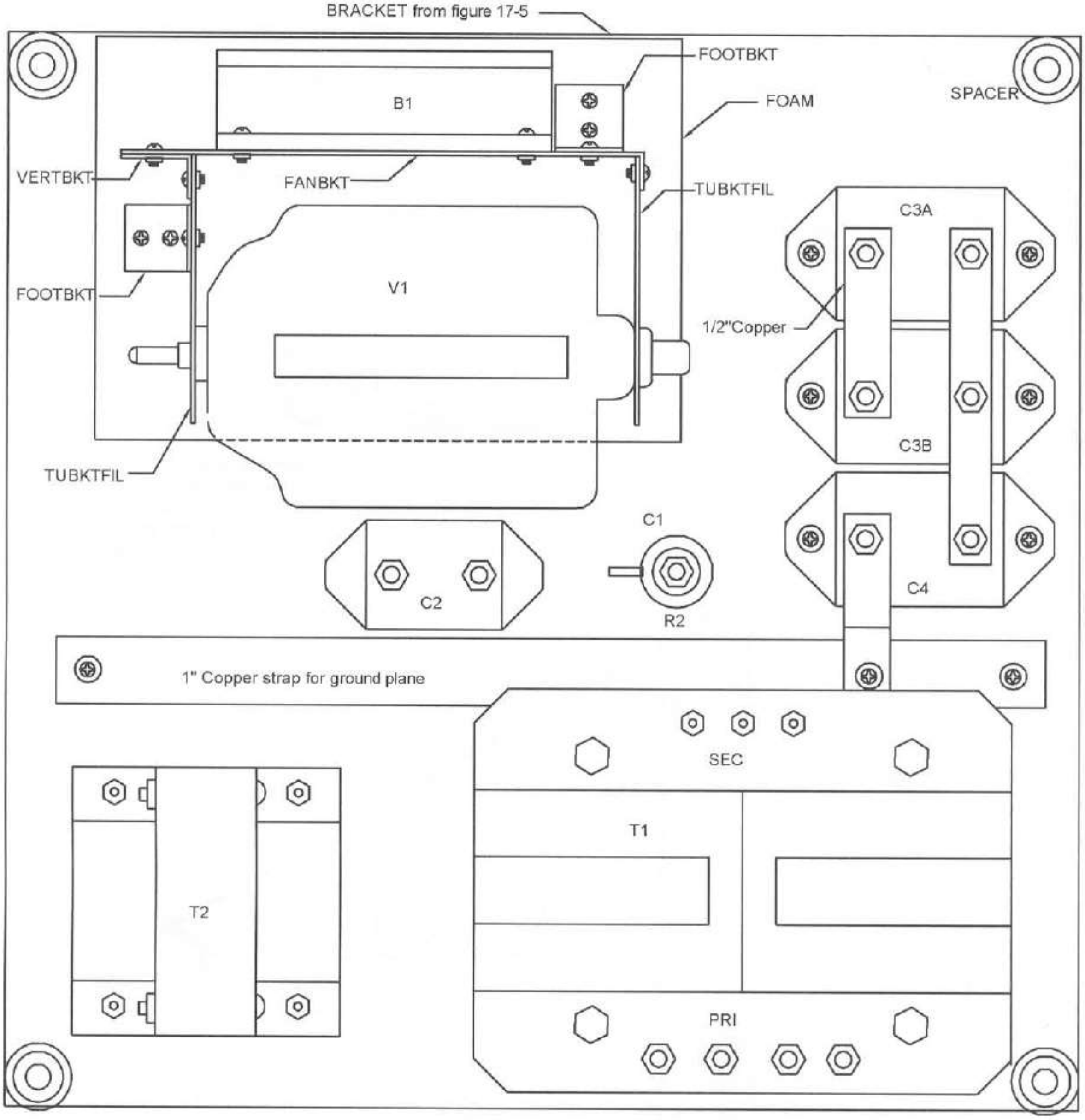
BRACKET from figure 17-5
FOOTBKT
FOAM
SPACER
B1
VERTBKT
FANBKT
TUBKTFIL
C3A
FOOTBKT
V1
1/2"Copper
TUBKTFIL
C3B
C1
C2
R2
C4
1" Copper strap for ground plane
SEC
T1
T2
PRI

Figure 17-6 *Bottom deck assembly*

must set on a 1/4-inch foam rubber pad to avoid any vibration. Fabricate the tube pin connections from 1/2-inch strips of thin copper in the form of clamping connectors. Use screws to tighten the clamps securely around the tube pins. Crimp the connecting wires in place and carefully solder the connecting leads completely. This is important as the tube pins must pass a high current.

4. Lay out the bottom deck DECKBOT and trial-fit the components, as shown in Figure 17-6. (You may want to consider castors at the corners for system mobility.) Cut a 14-inch length of a 1-inch-wide copper strip for the main ground plane. Secure it via two wood screws at the end points. Secure the plate transformer (T1) using four 1/4-20 × 1 1/2-inch bolts with flat washers on the bottom side of the wooden deck. Mount the three mica capacitors (C3A, C3B, and C4) using 6-32 × 1 1/2-inch screws, nuts, and washers. Cut 1/2-inch lengths of copper straps and attach them as shown. Note the grounding piece on C4 must be shaped and formed to have a flat surface abutting the ground plane strap and fitting as shown.

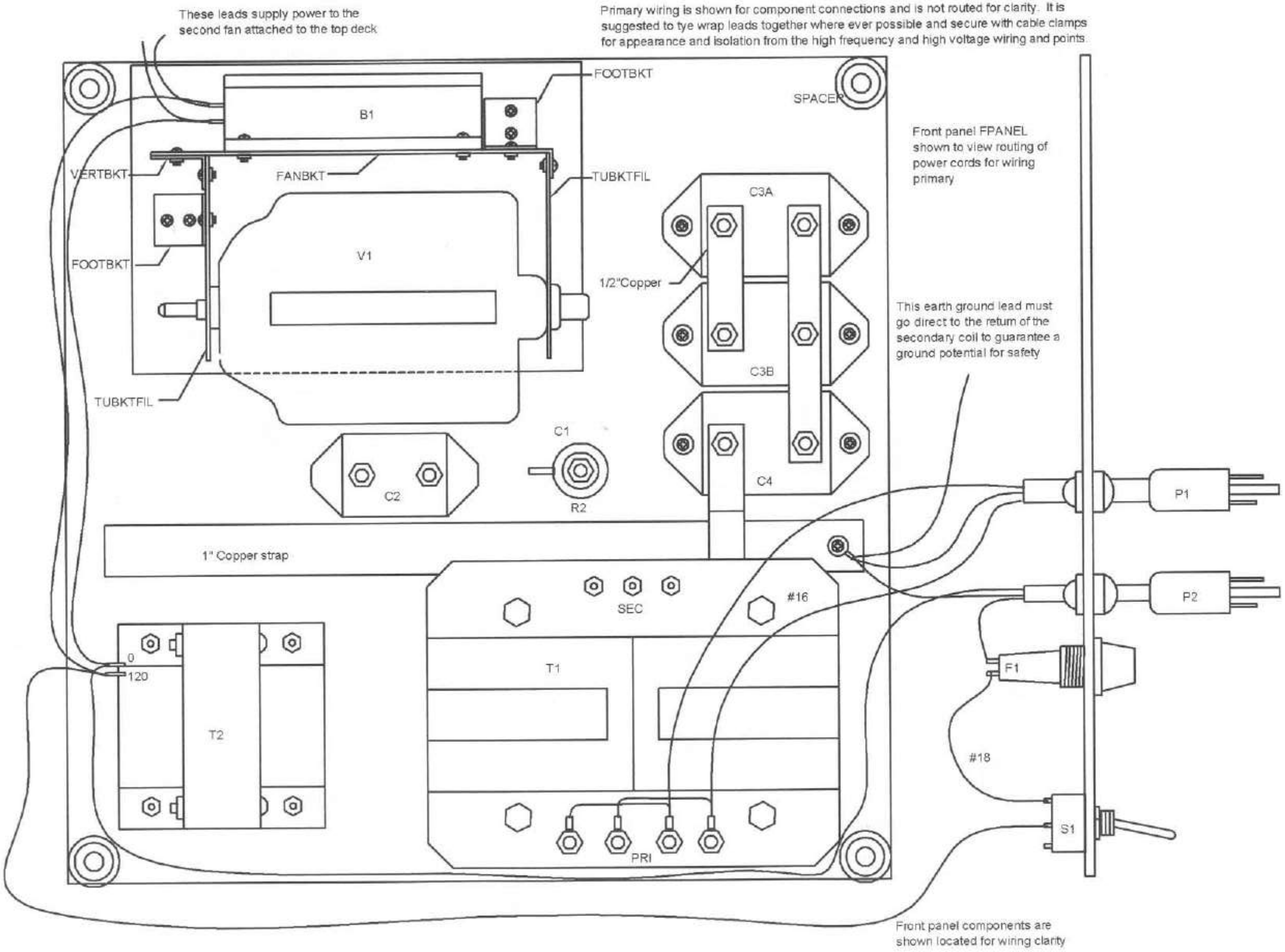
These leads supply power to the second fan attached to the top deck

Primary wiring is shown for component connections and is not routed for clarity. It is suggested to tye wrap leads together where ever possible and secure with cable clamps for appearance and isolation from the high frequency and high voltage wiring and points.

Front panel FPANEL shown to view routing of power cords for wiring primary

This earth ground lead must go direct to the return of the secondary coil to guarantee a ground potential for safety

Front panel components are shown located for wiring clarity

FOOTBKT
SPACER
B1
VERTBKT
FANBKT
TUBKTFIL
FOOTBKT
V1
TUBKTFIL
C3A
1/2"Copper
C3B
C4
C1
C2
R2
1" Copper strap
P1
P2
F1
#16
SEC
T1
T2
0
120
#18
S1
PRI

Figure 17-7 *Bottom deck primary wiring*

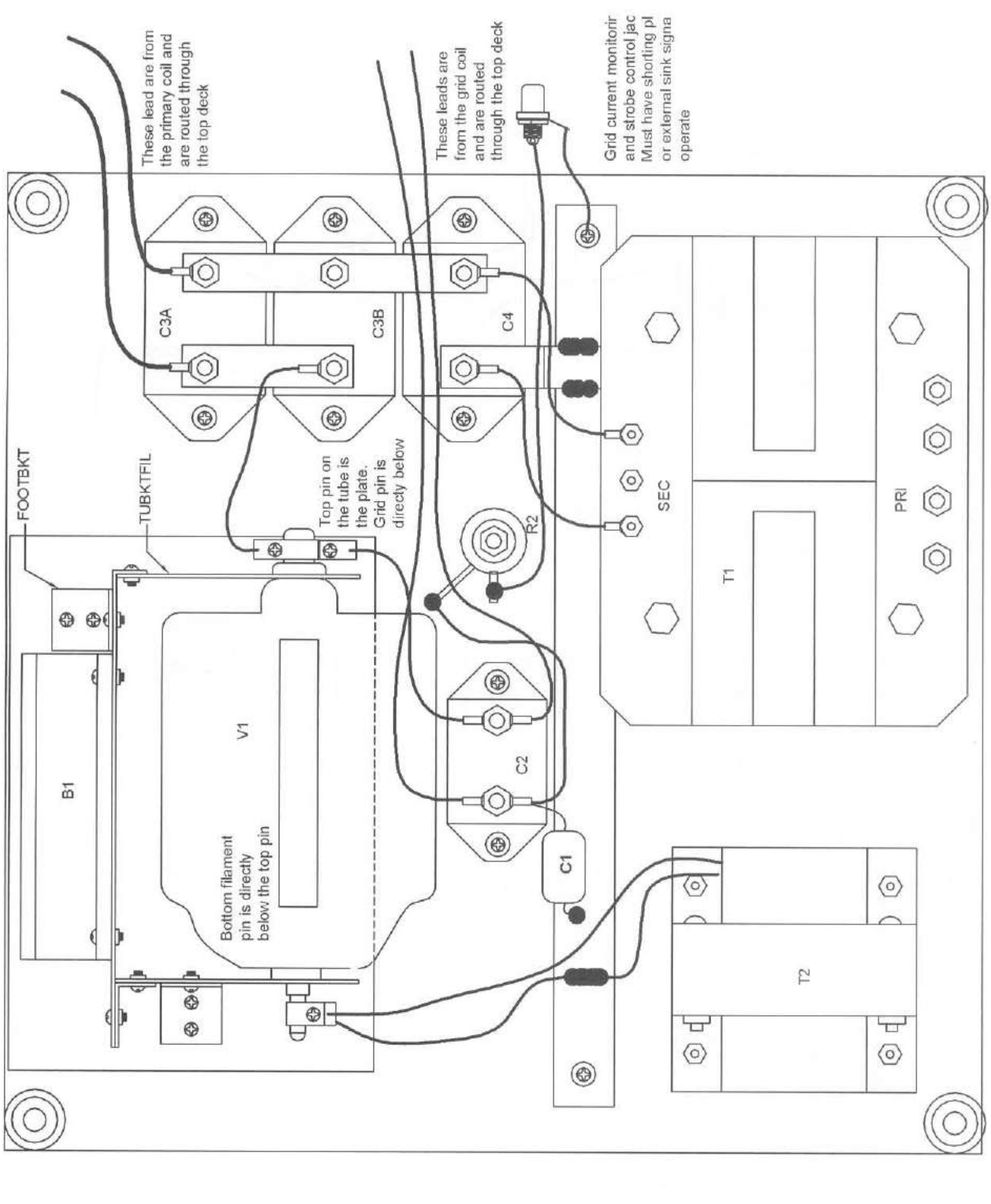
These lead are from the primary coil and are routed through the top deck
These leads are from the grid coil and are routed through the top deck
Grid current monitorir and strobe control jac Must have shorting pl or external sink signa operate
C3A
C3B
C4
FOOTBKT
TUBKTFIL
Top pin on the tube is the plate. Grid pin is directly below
R2
SEC
PRI
T1
B1
V1
C2
C1
Bottom filament pin is directly below the top pin
T2

Figure 17-8 *Bottom deck high-voltage wiring*

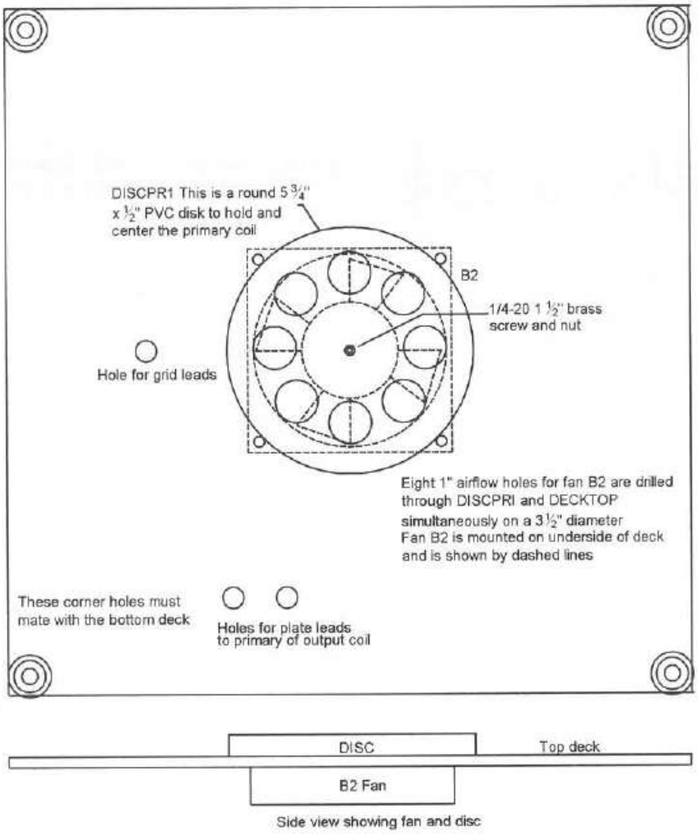
DISCPR1 This is a round 5 ¾" x ½" PVC disk to hold and center the primary coil
B2
1/4-20 1 ½" brass screw and nut
Hole for grid leads
Eight 1" airflow holes for fan B2 are drilled through DISCPRI and DECKTOP simultaneously on a 3 ½" diameter
Fan B2 is mounted on underside of deck and is shown by dashed lines
These corner holes must mate with the bottom deck
Holes for plate leads to primary of output coil
DISC
Top deck
B2 Fan
Side view showing fan and disc

Figure 17-9 *Top deck fabrication and assembly*

Do not tighten the nuts at this time. Mount mica capacitor C2 using similar hardware and grid resistor R2 vertically using a long 1/4-inch bolt or a threaded rod. Do not overtighten. Secure the filament transformer T2 using four 8-32 × 1 1/2-inch screws and nuts. Secure the tube and fan bracket via four #8 × 1/2-inch wood screws through the two foot brackets FOOBKT. Also note the foam rubber mount FOAM. Drill holes for the 1/4-inch −20 threaded rods for the deck spacer.

5. Thread the power cords (P1, P2) through the clamp bushings (BU1, BU2) as shown in Figure 17-7 on the front-panel section (FPANEL). The actual location on the panel section is shown in Figure 17-3. Figure 17-7 is shown to clarify the wiring. Power cord P1 must be a heavy-duty, #16, three-wire cord, with P2 being a #18, three-wire cord. Connect the two green ground leads to a crimp lug along with a 24-inch piece of #18 stranded wire that eventually connects to the center ground lug of the secondary coil. Solder this lug and verify the integrity as this is mandatory for electrical safety. Proceed to wire as shown using #18 stranded wires. Note that the two leads eventually connect to the second blower (B2) mounted on the top deck. Use crimp lugs and solder wherever shown. Route the wires, tie wrap, and clamp into place for appearance and isolation from other components.

6. Wire the filament clips to the 10-volt output leads of T2 as shown in Figure 17-8. These clips are shown in Figure 17-5. You will note that insulation is removed from one of these leads, and the uncovered section is soldered to the 1-inch copper strip. Solder in the bypass resistor (C1) to the copper strip and to a lug with the other leads to R2. High-voltage wire should be used from transformer T2 to capacitor C4. Proceed to complete the wiring using crimp lugs and soldering. Two pairs of wires are shown from the primary and grid coil. These are not actually connected at this time as they are part of the top deck assembly (DECKTOP). Jack (J1) is attached to the front panel shown located in Figure 17-3.

7. Fabricate the DECKTOP, as shown in Figure 17-9. This piece was originally shown as being

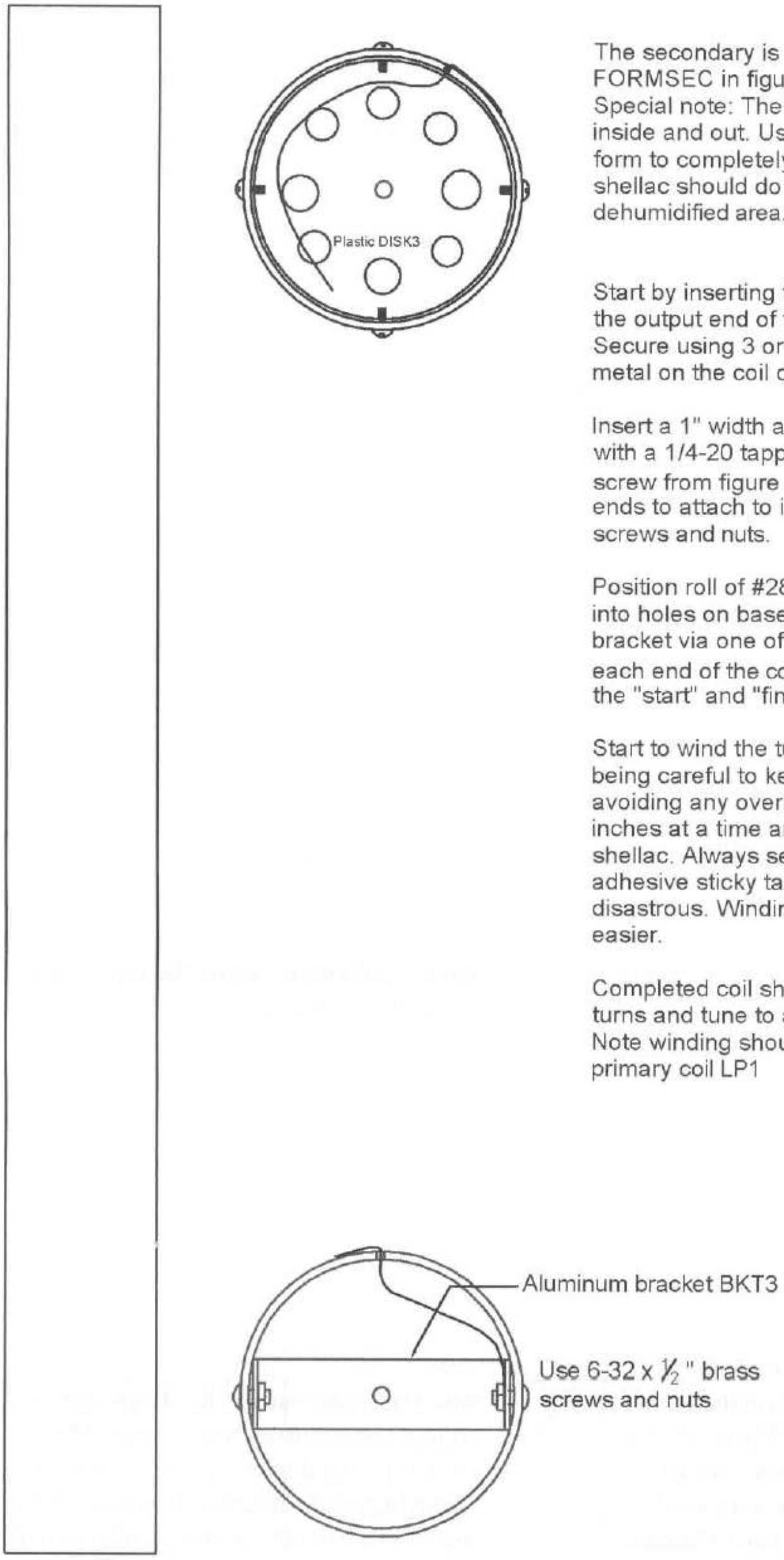

The secondary is wound on the form as shown for FORMSEC in figure 17-3.
Special note: The coil form must be totally clean both inside and out. Use isopropyl alcohol and allow the form to completely dry. Coat with a layer of orange shellac should do this on a dry day or in a dehumidified area.

Start by inserting the polycarbonate disk DISK3 into the output end of the coil form flush with the end. Secure using 3 or 4 small plastic screws. Do not use metal on the coil output end as arcing will occur.

Insert a 1" width aluminum bracket into the base end with a 1/4-20 tapped hole for screwing on to the brass screw from figure 17-9. Bend down ½" 90 degree ends to attach to inner diameter of coil via brass screws and nuts.

Position roll of #28 heavy magnet wire and thread into holes on base end of form. Attach free end to bracket via one of the screws. Drill small holes at each end of the coil form ½" from ends for threading the "start" and "finish" leads of the windings.

Start to wind the turns and continue for 33" in length being careful to keep the wire tight, free of kinks and avoiding any overlaps. Do approximately several inches at a time and shellac in place using orange shellac. Always secure lead with a piece of good adhesive sticky tape as an "unwind" will be disastrous. Winding this coil with two people is much easier.

Completed coil should contain approximately 2400 turns and tune to around 200 khz when free standing. Note winding should be in the same direction as primary coil LP1

Figure 17-10 *Assembly of the secondary coil*

a $^{3}/_{8}$-inch Plexiglas square from Figure 17-3. Fabricate a $5^{3}/_{4}$-inch PVC disc (DISCPRI) from $^{1}/_{2}$-inch-thick sheet. This piece is for securing and centering the primary coil assembly, shown later in the plans. Secure it in place on the deck via a $^{1}/_{4}$-20 × 1-inch brass screw and nut. This will be the grounding point for the bottom of the secondary coil and

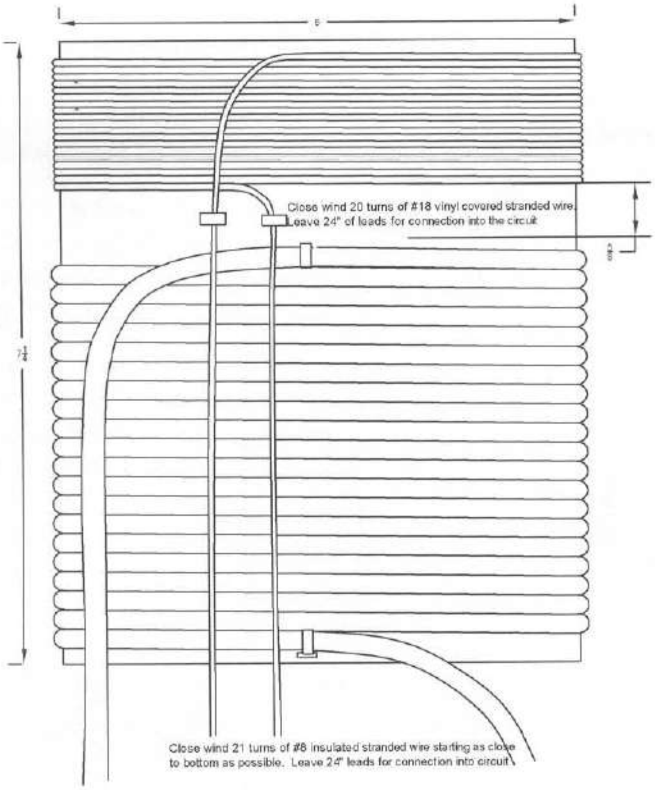
6
7¼
Close wind 20 turns of #18 vinyl covered stranded wire.
Leave 24" of leads for connection into the circuit
Close wind 21 turns of #8 insulated stranded wire starting as close
to bottom as possible. Leave 24" leads for connection into circuit

Figure 17-11 *Assembly of the primary coil*

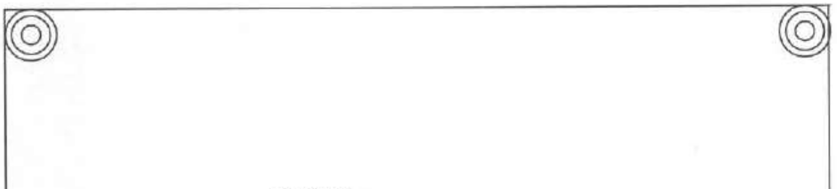
DISCPR1
FROM FIG 17-11
FROM FIG 17-10

Figure 17-12 *Final assembly of top deck*

goes to earth ground, as shown in Figure 17-7. Proceed to drill the eight 1-inch air flow holes.

At this time, the builder should be aware that if his or her coil is to be used on a continuous basis, the second blower mounted on the underside of the top deck is necessary to keep the bottom of the coils cool. They will heat up over time due to the resonant rise of current associated with the base of a quarter-wave resonator. If your coil is to be used for a short-time operation of several minutes, the part involving the drilling of the air flow holes and the second fan may be omitted.

8. Assemble and wind the secondary coil as shown in Figure 17-10. Proper winding of the coil is mandatory for achieving optimal operation. Kinks and overlaps will seriously affect output.
9. Assemble the primary coil assembly, as shown in Figure 17-11. Start at the bottom and tightly wind the turns. Use tie wraps to secure the start and finish of this heavier 21 turns of the #8 stranded wire. You may want to place dabs of silicon rubber to secure them in place. Start

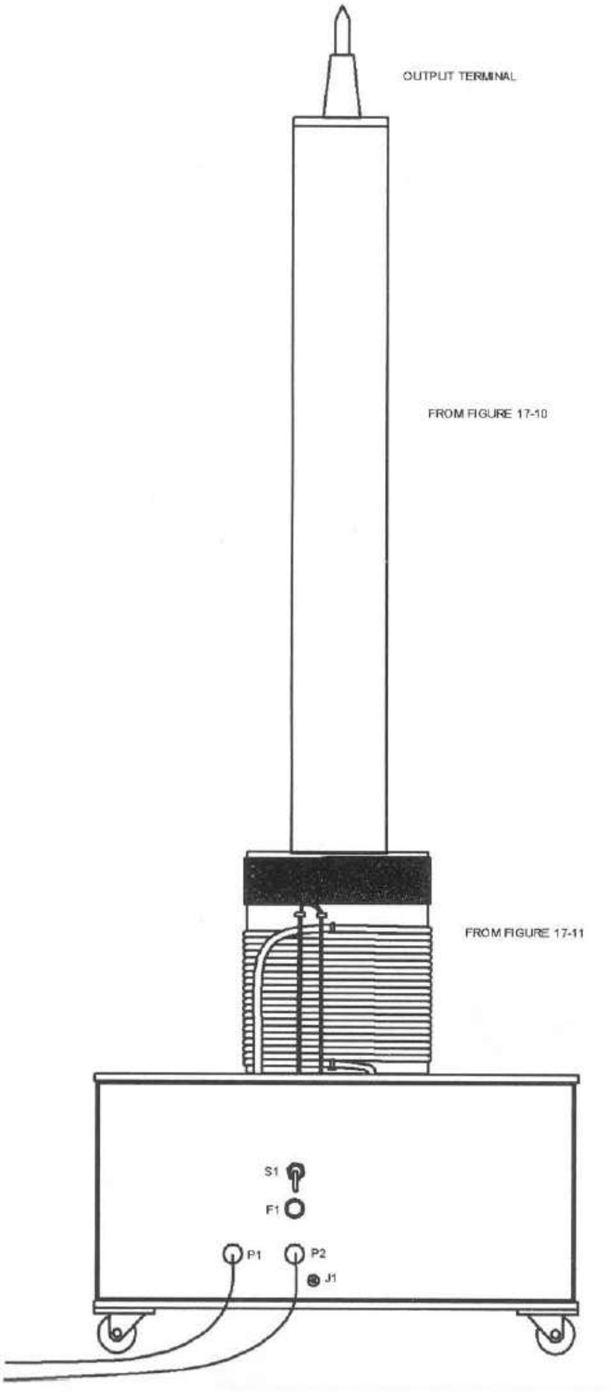

Figure 17-13 *Side view of final assembly*

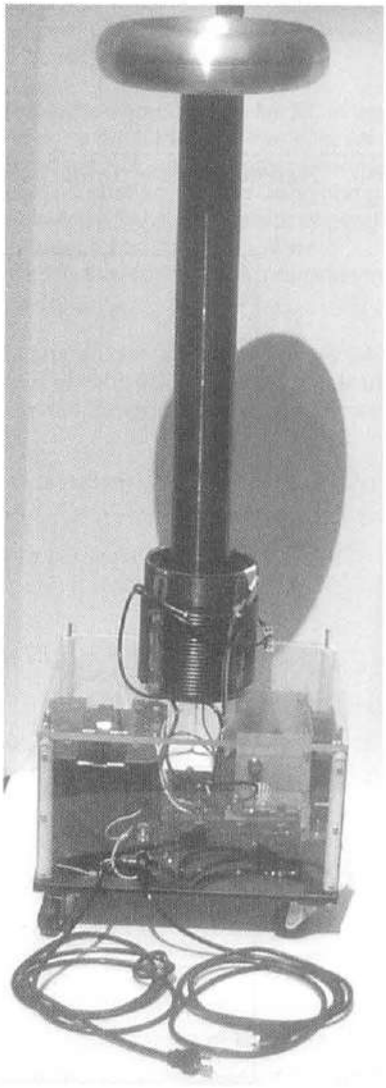

Figure 17-14 *The final assembly*

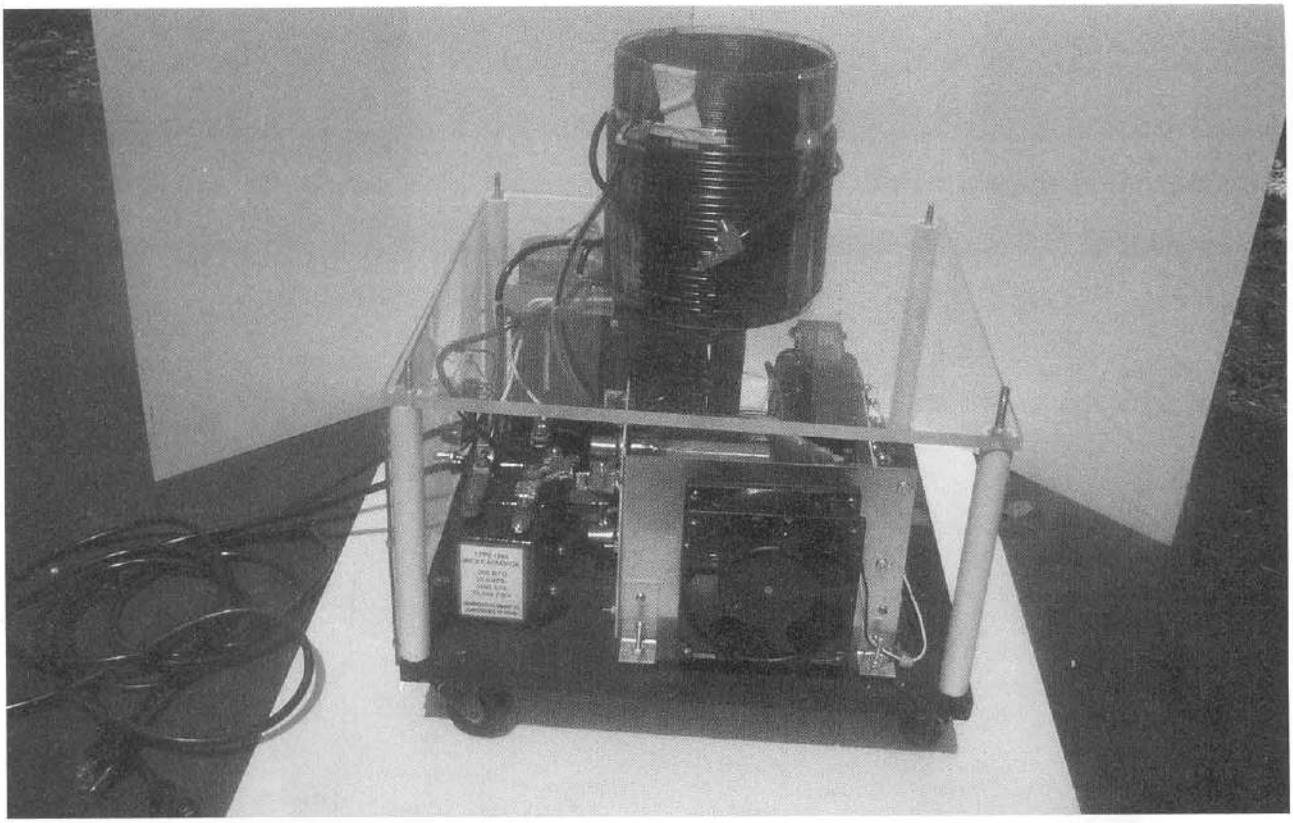

Figure 17-15 *The right-side view*

from the top and wind 20 turns of #18 vinyl-covered stranded wire and tie wrap it in place. Note that there must be at least 1/2-inch between the closest proximities of these two coils as a flashover (or spark break) could damage the vacuum tube. Leave at least 2 feet of extra leads for connections, as shown in Figure 17-8.

10. Finally, assemble the top deck, as shown in Figure 17-12. The primary coil section slips over the disc DISCPRI, as shown in Figure 17-8. The tapped hole in the bracket of the secondary coil screws onto the center brass screw shown in Figure 17-8. It is this screw that grounds and secures the secondary coil.

    Route the leads from the coils through their respective holes for connection into the circuit. You will note back in Figure 17-8 that the heavy plate leads are shown connected to C3A via heavy crimp lugs and are also soldered. The same applies for the smaller grid leads that connect across C2. These leads may require switching during the test steps. Both sets of these leads carry high-frequency energy with the heavier leads carrying high-radio-frequency voltages and currents. These heavier leads must be as short and direct as possible and not be in close proximity to other objects.

    Figure 17-13 shows the completed graphic of the coil assembly with an insulator and a pointed discharge pin. Any suitable feed through the insulator around 2 to 3 inches in height can be used. You may also use a piece of polycarbonate or Teflon rod for this part. Other plastics may carbonize and burn.

    The following figures are actual photos showing different views of the assembly apart and intact. They are intended as assembly aids.

Figure 17-14 shows a 12 × 3-inch toroid as the output terminal. You will note the two power cords exiting through the front panel.

Figure 17-15 shows the right side with a view of the blower B1 and tube-mounting bracket.

Figure 17-16 shows the left side with a view of the filament and plate transformers.

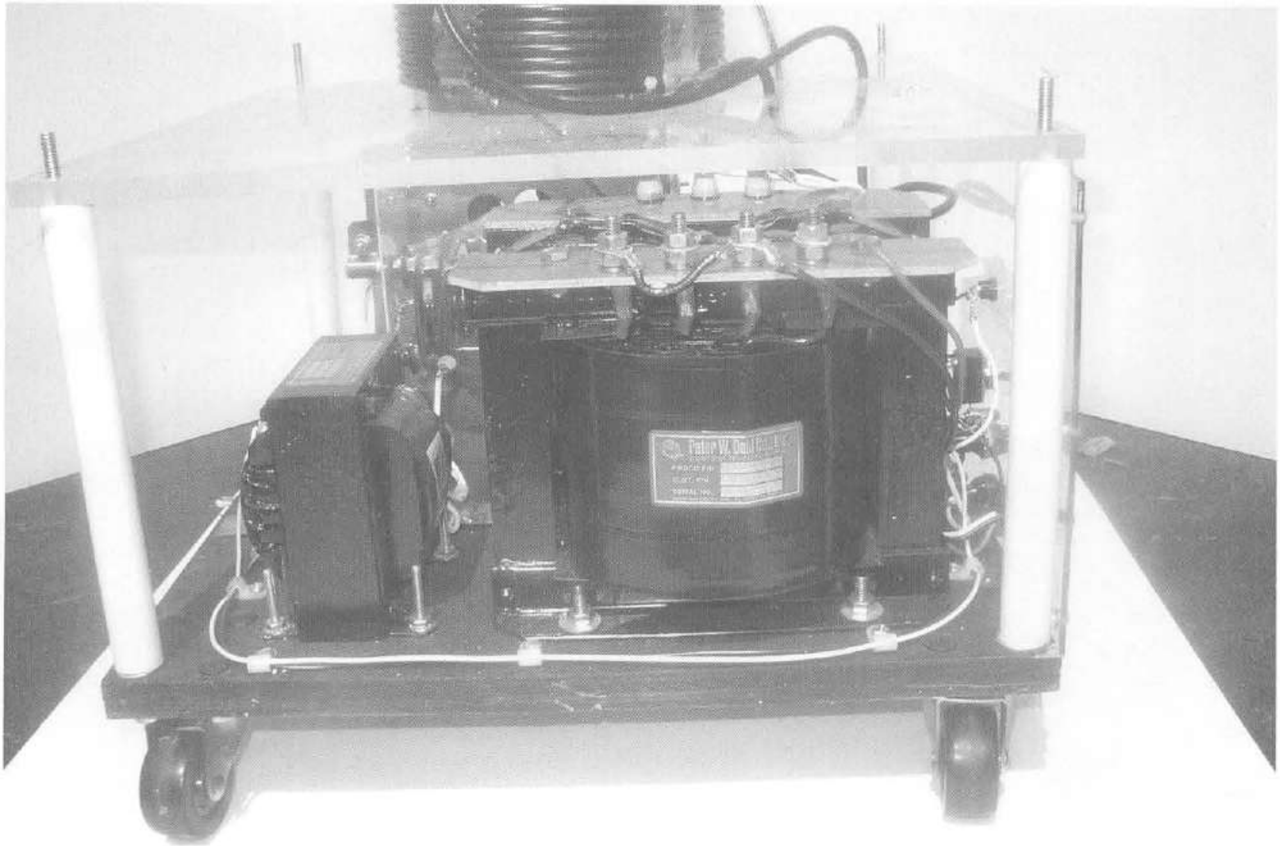

Figure 17-16 *The left-side view*

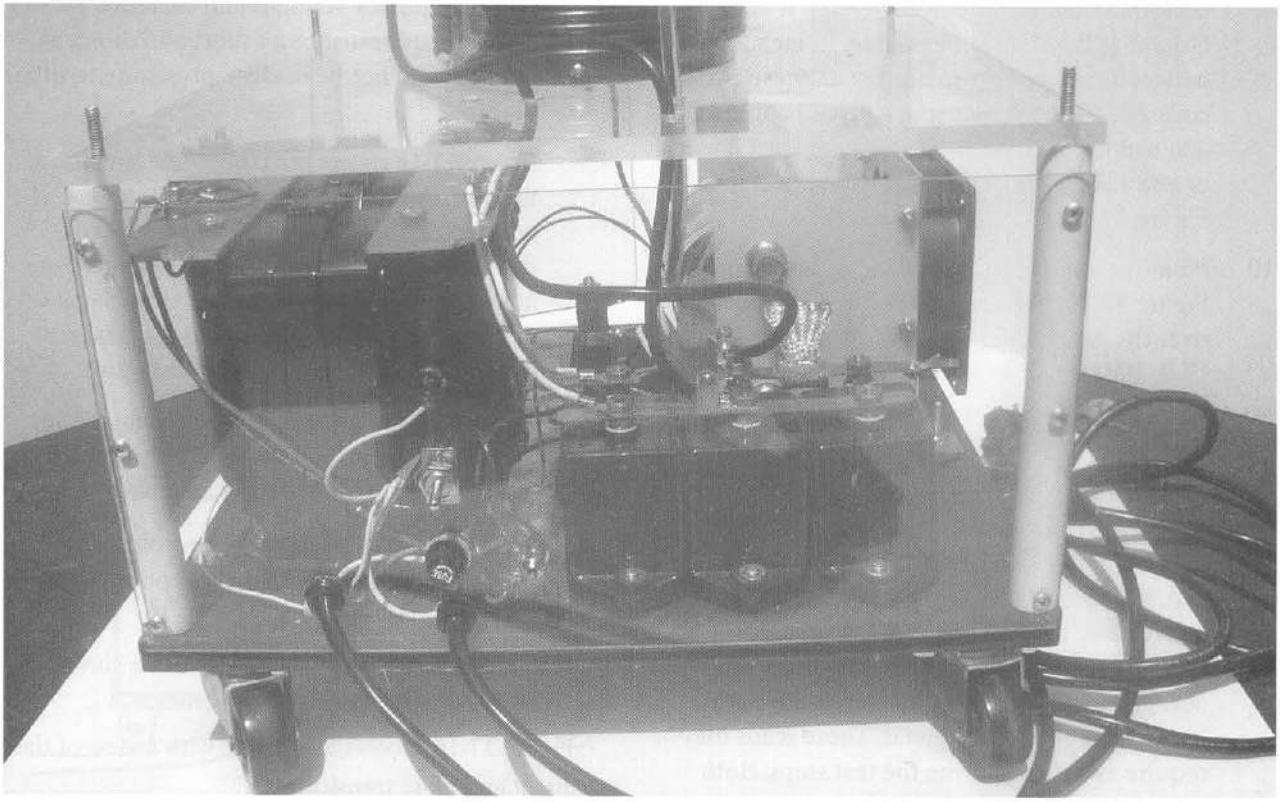

Figure 17-17 *The front view*

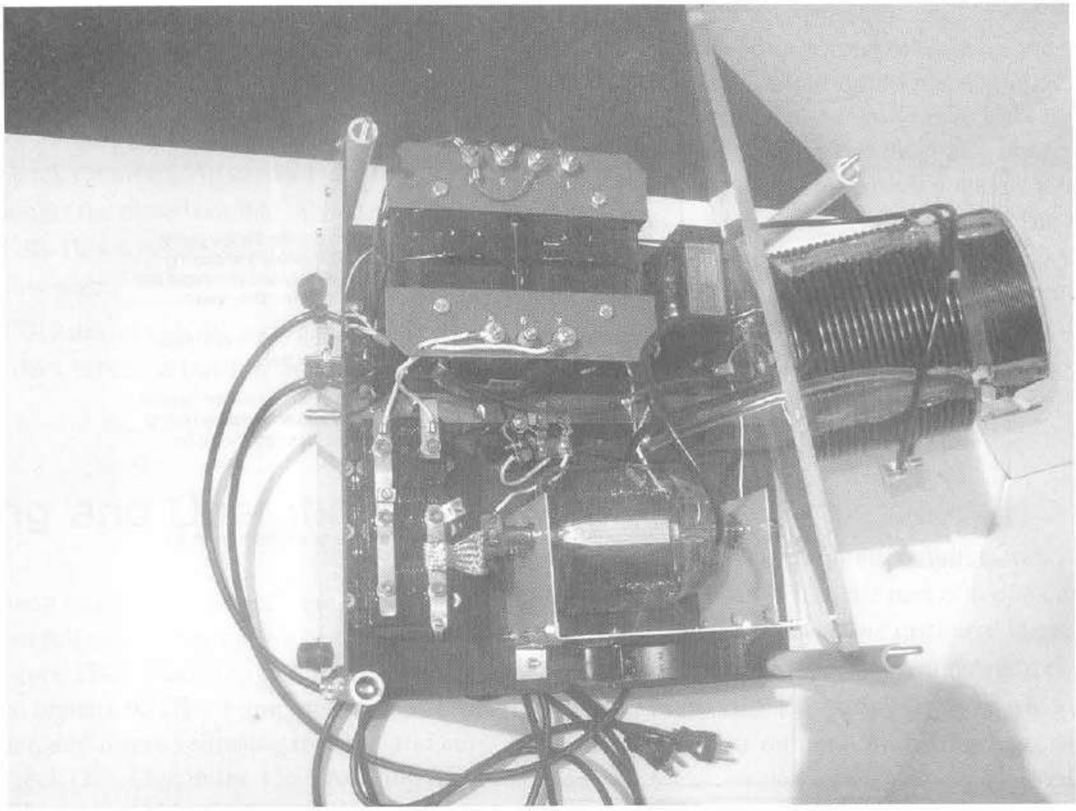

Figure 17-18 *The top view*

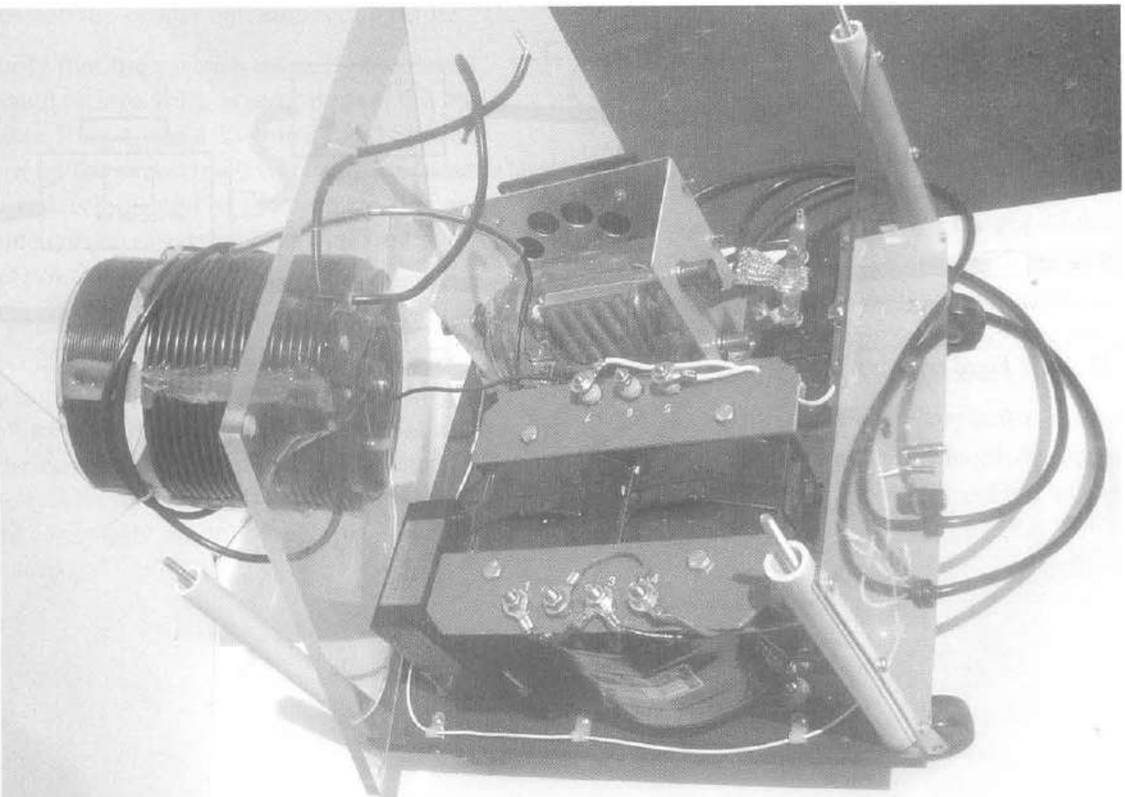

Figure 17-19 *Another top view*

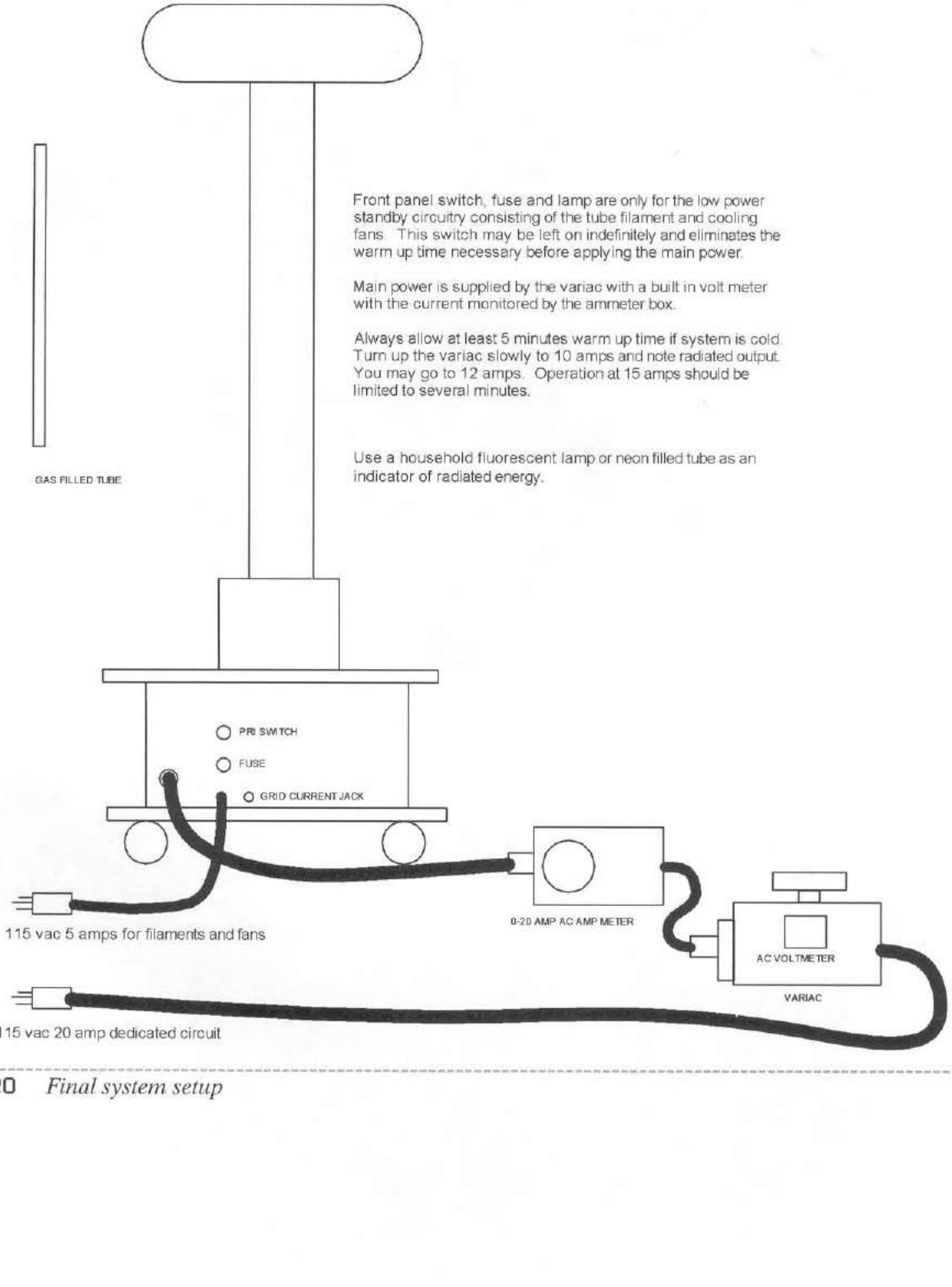

GAS FILLED TUBE
PRI SWITCH
FUSE
GRID CURRENT JACK
115 vac 5 amps for filaments and fans
0-20 AMP AC AMP METER
AC VOLTMETER
VARIAC
115 vac 20 amp dedicated circuit

Front panel switch, fuse and lamp are only for the low power standby circuitry consisting of the tube filament and cooling fans. This switch may be left on indefinitely and eliminates the warm up time necessary before applying the main power.

Main power is supplied by the variac with a built in volt meter with the current monitored by the ammeter box.

Always allow at least 5 minutes warm up time if system is cold. Turn up the variac slowly to 10 amps and note radiated output. You may go to 12 amps. Operation at 15 amps should be limited to several minutes.

Use a household fluorescent lamp or neon filled tube as an indicator of radiated energy.

Figure 17-20 *Final system setup*

Figure 17-17 shows the front view through the clear plastic panel. Note the three mica capacitors C3A, C3B, and C4.

Figure 17-18 shows the top view with the clear plastic top deck removed. We used a piece of shielded strap to connect the plate terminal of the tube to C3A and C3B. This is not necessary and may be heavy flexible wire.

Figure 17-19 also shows the top view with the clear plastic top deck removed but at a different angle.

Testing and Operating

1. You will need a dedicated 20-amp line to obtain full output. Set up the system as shown in Figure 17-20. Place the unit away from metal objects, insert a 7-amp fuse in the holder, and insert a shorting into the grid current jack (J1). Plug in the 115 VAC plug for the filament and blower fans, and wait for the tube to glow and the fans to blow. Always allow at least 3 minutes for the tube filament to reach the proper emission temperature.
2. Verify that the variac is properly wired and turned to zero volts, as indicated on the voltmeter. Plug it into a 20-amp circuit. Slowly turn up the variac to 30 volts and verify some output as indicated by the bulb glowing or some corona discharge occurring on the output pin.

   If you do not observe some output, it will be necessary to completely shut the system down by unplugging both plugs, reversing the grid coil wires that attach to C2, and repeating the previous step. This is referred to as phasing the system. You should get substantial output and easily light the tube when properly phased.
3. Once you verify that the coil is properly phased, you may start to increase the voltage, noting the output discharge, the input current, and the applied AC voltage as indicated on the variac meter. You must also observe that the plate of the tube does not start to glow, as this is an indication of mistuning. The tube may have a very soft glow at full 120 VAC where the output should be 15 inches or more. The input current will approach 20 amps.

Notes

Tuning the coil for maximum output is tricky and may take some effort on the part of the builder. The tuning values will vary as the discharge increases due to the virtual capacitance of the ionization cloud at the top of the coil. The output terminal will also greatly effect the output tuning. Try adding or subtracting a turn or two on the primary plate coil, as this will change the resonant drive signal. You will need more inductance as the output increases, and the trick is to select the optimum value at the desired input power.

A 12-inch toroid or another large capacitive terminal will cause larger circulating currents within the coil and greatly enhance the wireless transmission of energy, as can be demonstrated by observing the distance that is possible for lighting the gas discharge tube without a wire. This will eliminate the air breakdown discharge when operating in this mode as the terminal voltage drops unless you increase the input power to overcome this loading effect.

The geometry of the secondary coil as shown is not the optimum for maximum length discharges. A squatter size would be more efficient but would radiate less.

Table 17-1 Vacuum Tube Tesla Coil Parts List

Ref. #	Description	DB Part #
R2	5K, 100-watt wire-wound resistor	
C1	.004-microfarad, 1,000-volt polypropylene capacitor	DB# .0039/400
C2	.003-microfarad, 3-kilovolt mica capacitor	DB# .003/3 KVM
C3A, B, C4	Three .01-microfarad, 10-kilovolt mica capacitors	DB# .01/10 KVM
V1	833A Triode vacuum tube	DB# 833A
T1	3,000-volt, .5-amp plate transformer	DB# 3KV/.5A
T2	10-volt, 10-amp filament transformer	DB# 10V/10A
B1, 2	5 × 5-inch muffin fan and 5 × 5-inch high-output muffin fan	DB # FAN1
S1	*Single pole, single throw* (SPST) 5-amp toggle switch	
F1/FS1	Panel-mount fuse holder and 5-amp fuse	
P1	8-foot #16 three-wire power cord	
P2	8-foot #18 three-wire power cord	
AMPMETER	20-amp AC panel mount ammeter	
VARIAC	18-amp variac with built-in meter	DB# VARIAC18A
TUBE	15-inch neon tube	DB# NE15
TER1	12 × 3-inch toroidal terminal	DB# TO12

Chapter Eighteen

Universal Capacitance Discharge Ignition (CDI) Driver

Figure 18-1 shows how to construct a very useful driver circuit that can be used to pulse most induction and ignition coils in the forward direction. High-quality automotive ignition coils can produce 50- to 75,000-volt pulses capable of jumping several inches. The system allows control of the pulse repetition rate up to 100 reps/second and individual pulse energy up to .3 joules. A maximum power output of over 30 watts is possible. Our laboratory model has generated discharges of up to 10 inches from some high-performance ignition coils.

The advantage of a *capacitance discharge ignition* (CD) system is its ability to produce high peak currents and voltages simultaneously. This property is well understood and appreciated in the automotive industry where fouled spark plugs are prevented by the brisance of the discharge. Electric fences burn away vegetation growth from contacting the wire by using high current pulses. Stun guns use these high peak current pulses to thwart the neuromuscular system. All this peak power with very low resultant energy allows many useful applications where battery-operated systems can be efficiently utilized.

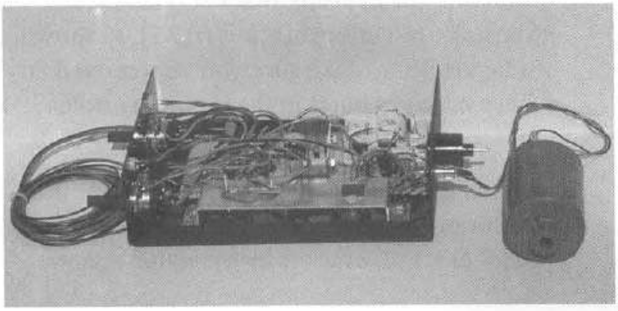

Figure 18-1 *The capacitor discharge driver*

This is an intermediate-level project requiring basic electronic skills and one should expect to spend $25 to $50. All the parts are readily available, with specialized parts obtainable through Information Unlimited (www.amazing1.com), and they are listed in Table 18-1.

Circuit Description

Figure 18-2 shows the complete circuit schematic of the adjustable energy and repetition rate capacitor discharge driver. The adjustable voltage regulator I2 controls the operating voltage to the system via control pot R9/S1 from 6 to 12 volts. Input to I2 is a 12-volt DC wall adapter transformer with a current rating of 1 to 2 amps being controlled by switch S1. This adjustable voltage is fed to the switching transistors Q1 and Q2 via the center tap of stepup transformer T1. The switching action of these transistors is controlled by a complementary pulse driver circuit I1 with its frequency controlled by trimpot R3. This circuit generates a high-frequency signal that alternates between the two primary windings, in turn inducing a higher voltage in the secondary winding.

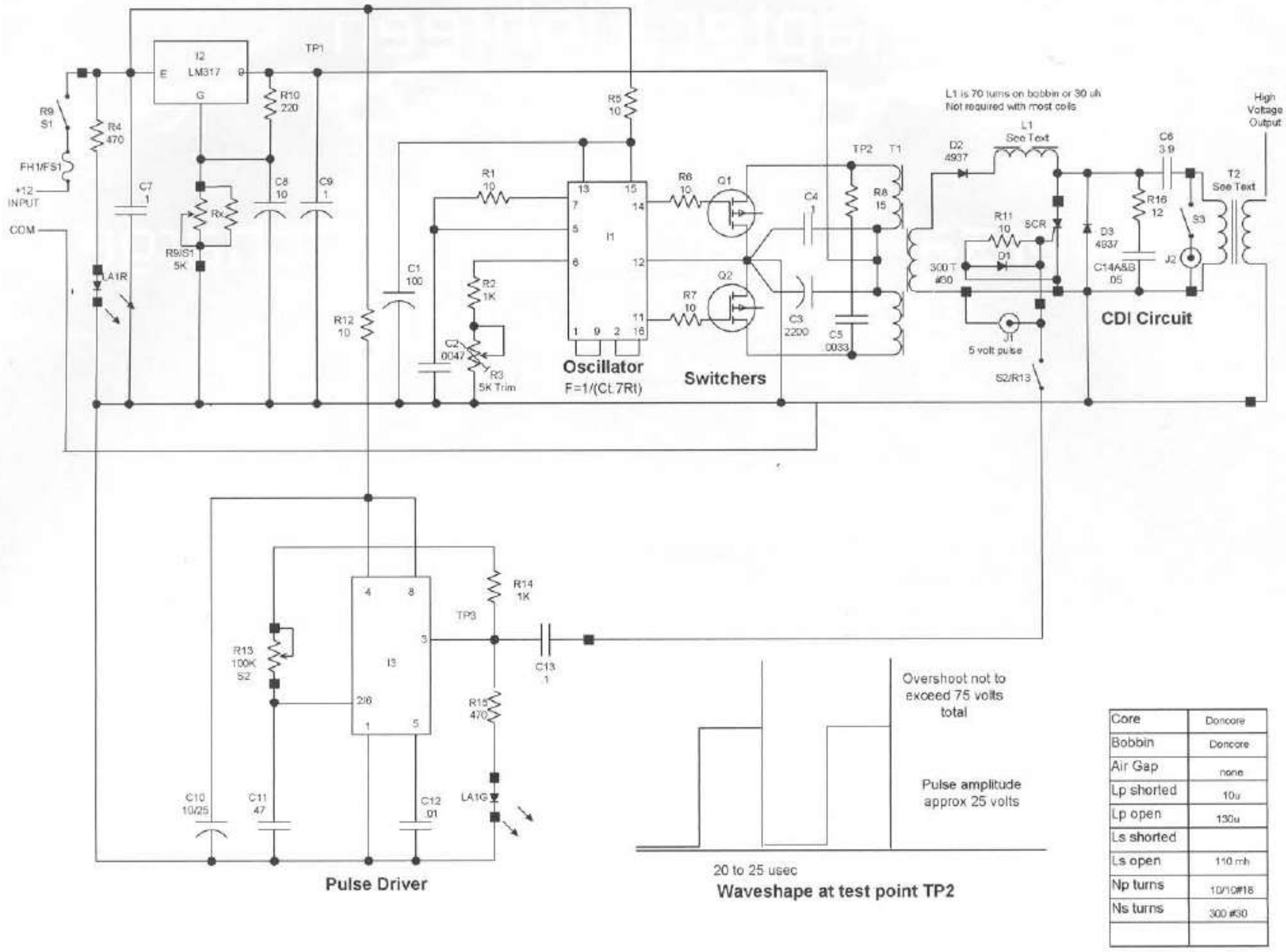
I2
LM317
E
G
TP1
R9
S1
R4
470
FH1/FS1
+12
INPUT
COM
C7
1
R10
220
Rx
R9/S1
5K
C8
10
C9
1
LA1R
R12
10
C1
100
R1
10
R2
1K
C2
.0047
R3
5K Trim
R5
10
13
15
7
14
5
I1
6
12
11
1
9
2
16
Oscillator
F=1/(Ct.7Rt)
R6
10
Q1
R7
10
Q2
Switchers
C4
1
C3
2200
C5
.0033
TP2
T1
R8
15
L1 is 70 turns on bobbin or 30 uh
Not required with most coils
L1
See Text
D2
4937
R11
10
SCR
D1
300 T
#30
J1
5 volt pulse
S2/R13
D3
4937
C14A&B
.05
R16
12
C6
3.9
S3
J2
T2
See Text
High
Voltage
Output
CDI Circuit
4
8
3
I3
2/6
1
5
R13
100K
S2
R14
1K
TP3
C13
.1
R15
470
LA1G
C10
10/25
C11
.47
C12
.01
Pulse Driver
Overshoot not to
exceed 75 volts
total
Pulse amplitude
approx 25 volts
20 to 25 usec
Waveshape at test point TP2

Core	Doncore
Bobbin	Doncore
Air Gap	none
Lp shorted	10u
Lp open	130u
Ls shorted	
Ls open	110 mh
Np turns	10/10#18
Ns turns	300 #30

Figure 18-2 *Circuit schematic*

T1 is a transformer with a preset, built-in leakage reactance that controls the short-circuit-charging current to discharge capacitor C6 through rectifying diode D2. Once C6 is charged, it can be discharged by the silicon controlled rectifier switch (SCR) through the primary high-voltage pulse transformer T2. The SCR is controlled by timer chip I3, connected as a variable pulse generator via pulse rate adjust pot R13/S2. It is these pulses that trigger the SCR to switch energy from C6 into T2, generating the high-voltage output pulses. You may also externally switch the SCR via a pulse signal at input jack J1. R13/S2 must be off for this external function. The pulse transformer is shown as a low-power unit in the parts list.

The real intention of the circuit is the ability to pulse-drive many different ignition coils at different energy levels as selected by R9 and pulse repetition rates as selected by R13. We have obtained sparks up to 4 inches with conventional automotive coils.

Assembly Steps

1. Assemble the switching transformer as shown in Figure 18-3.
2. Fabricate the heatsink bracket (HS1) and transistor-mounting bracket (BR1), as shown in Figure 18-4. Make sure you've removed any sharp edges or burrs in the mounting holes that could break through the insulating pads.
3. Attach Q1 and Q2 as shown, noting that the mounting tabs are insulated from the actual metal heatsink bracket by the nylon screws.

1. Bifilar (parallel) wind two different color #18 magnet wire for 10 turns on the primary bobbin. Tape in place leaving 8" leads. Note that different colors will help identifying the lead ends.
   #18 solid magnet wire can be used however high frequency LITZ wire will give a slight improvement. You can make this wire by obtaining 6 pieces of #26 magnet wire and twist together as a single wire.
2. Wind 300 turns of #30 magnet wire in the 10 segments of the secondary bobbin. Wind 30 turns per segment and attach the start and finish leads to the connection lugs
3. Place the wound coils onto the cores as shown and tape tightly into place.
4. You may verify the inductance as follows secondary open:
   A to B&C are around 130 micro henries.
   D to B&C are around 130 micro henries
   Secondary winding 110 millihenrys

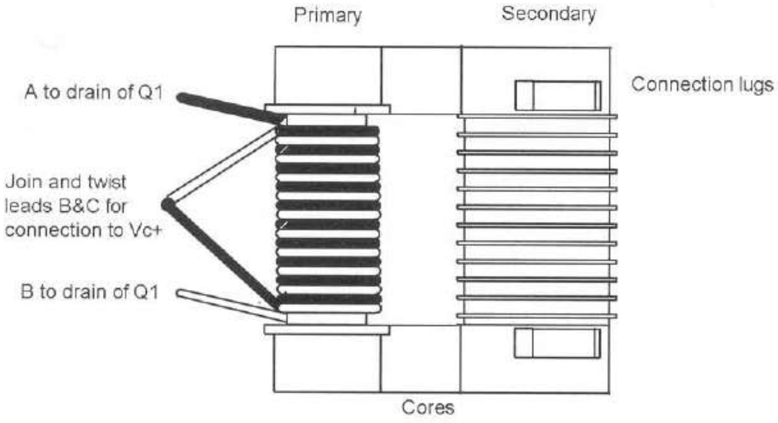
Primary
Secondary
A to drain of Q1
Connection lugs
Join and twist
leads B&C for
connection to Vc+
B to drain of Q1
Cores

*#18 solid magnet wire can be used however high frequency LITZ wire will give a slight improvement. You can make this wire by obtaining 6 pieces of #26 magnet wire and twist together as a single wire.

Figure 18-3 *T1 Transformer construction*

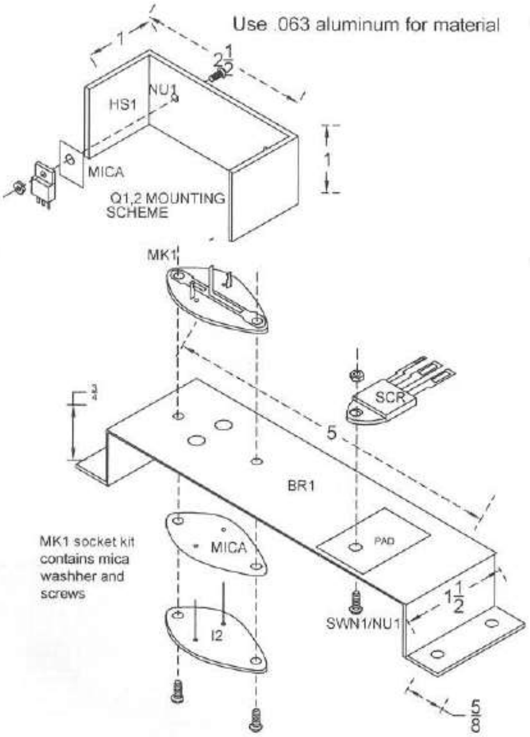
Use .063 aluminum for material
1
$2\frac{1}{2}$
HS1
NU1
MICA
Q1,2 MOUNTING SCHEME
1
MK1
SCR
$\frac{3}{4}$
5
BR1
MK1 socket kit contains mica washher and screws
MICA
PAD
I2
SWN1/NU1
$1\frac{1}{2}$
$\frac{5}{8}$

Figure 18-4 *Assembly of brackets*

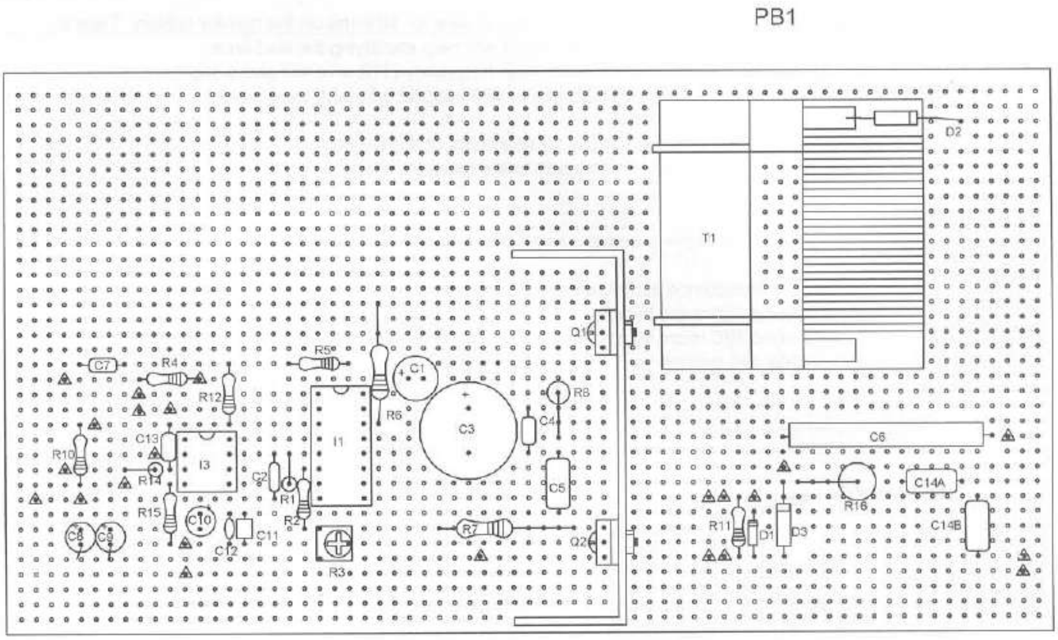
PB1
D2
T1
Q1
R5
C1
C7
R4
R12
R8
R6
C3
C4
C6
I1
C13
R10
I3
C2
R14
R1
C5
C14A
R16
R15
R11
D1
D3
C14B
C8
C9
C10
C11
R2
R7
Q2
C12
R3

▲ This symbol indicates points for wires to external components

Figure 18-5 *Assembly board and parts location*

Attach I2 using the socket assembly kit (MK1) and attach the SCR using a nylon screw and an insulating pad (PAD).

4. Cut a piece of .1 × .1 grid perforated circuit board (PB1) at 7 × 3.8 inches.

5. Insert the components as shown in Figure 18-5, starting from the lower left-hand corner. Count the perforated holes for a guide. Note the polarity marks on C1, C3, C8, C9, and C10 and diodes D1, D2, and D3. It is suggested that you use sockets SO16 for I1 and SO8 for I3.

6. Wire the components as shown in Figure 18-6 using the leads of the actual components as the connection runs, which are indicated by the dashed lines. Always avoid bare-wire bridges and messy solder joints. Check for cold or loose solder joints and secure T1 using double-sided sticky tape or plastic tie wraps.

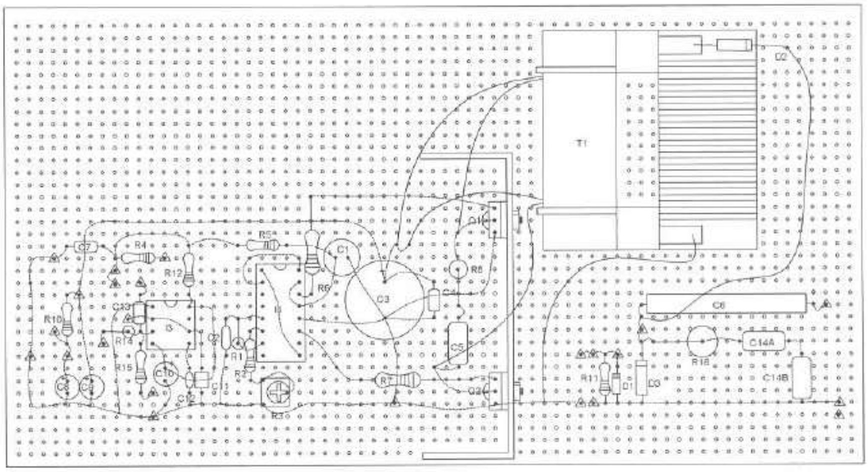

Figure 18-6 *Internal onboard wiring*

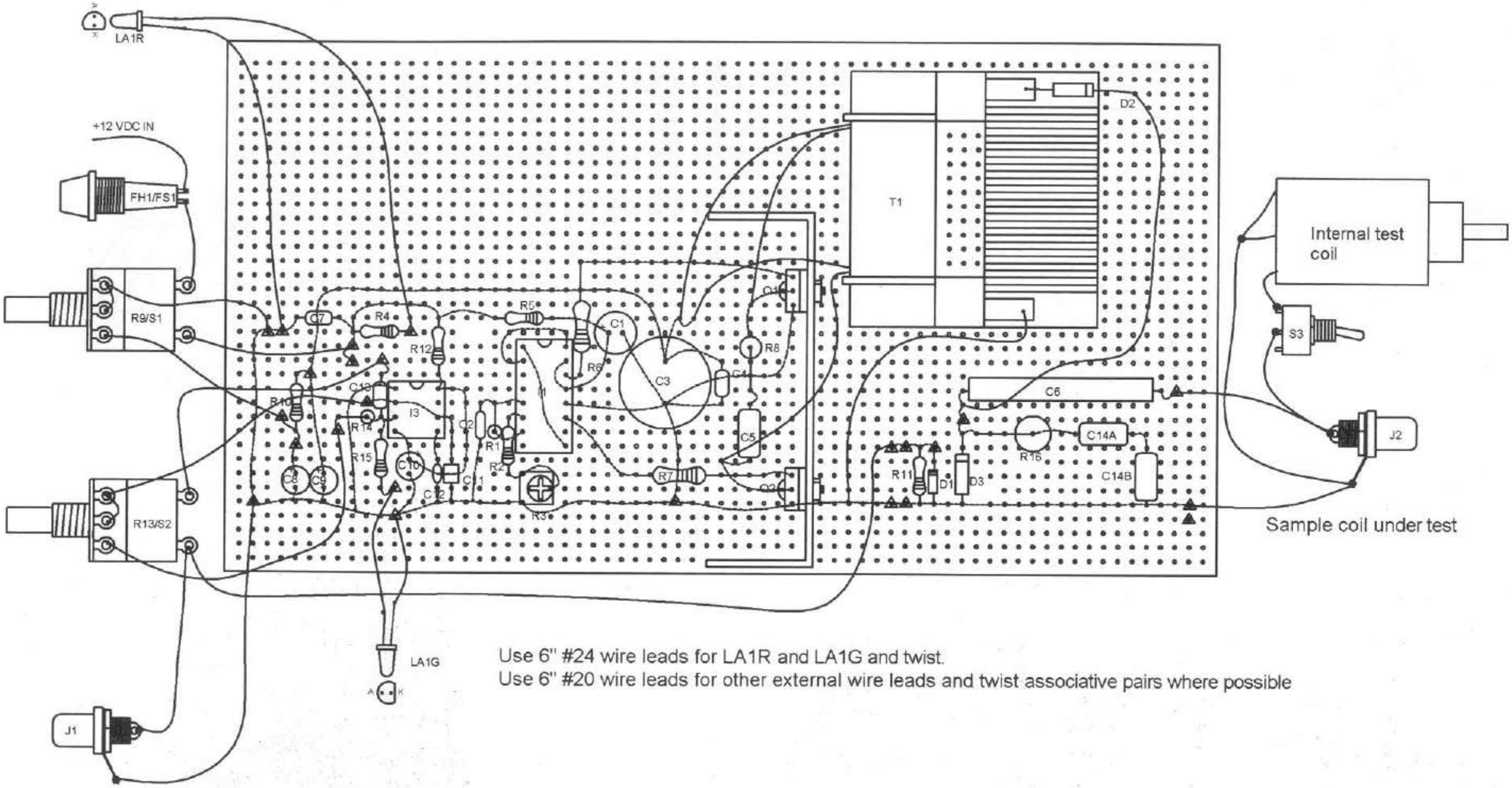
LA1R
A
K
+12 VDC IN
FH1/FS1
R9/S1
R13/S2
J1
LA1G
C7
R4
R5
R12
C13
I3
R10
R14
R15
C8
C9
C10
C12
C11
C2
R1
R2
R3
I1
R6
C1
C3
C4
R8
C5
R7
Q1
Q2
T1
D2
C6
R16
C14A
C14B
R11
D1
D3
Internal test coil
S3
J2
Sample coil under test

Use 6" #24 wire leads for LA1R and LA1G and twist.
Use 6" #20 wire leads for other external wire leads and twist associative pairs where possible

Figure 18-7 *External wiring to controls and input*

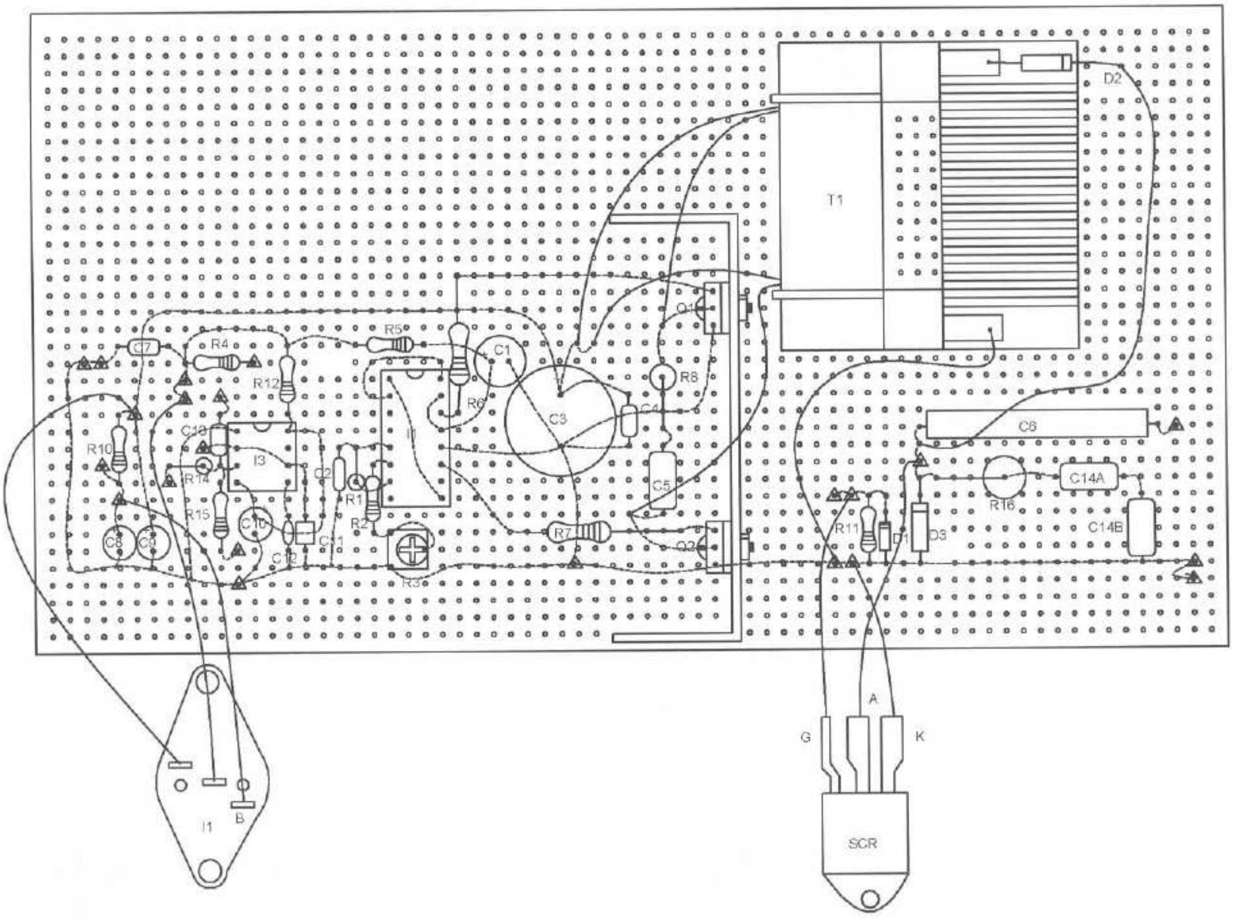
T1
D2
Q1
C7
R4
R5
C1
R12
R8
R6
C3
C4
C6
C13
R10
I3
I1
C2
R14
R1
C5
C14A
R16
C14B
R15
C10
R2
R11
D1
D3
C9
C8
C12
C11
R7
Q2
R3
I1
B
G
A
K
SCR

Figure 18-8 *External wiring to transistor modules*

The external coil of choice is connected to J3 using short lengths of #18 wire. It is a good idea to rig a ground return for the output spark to jump to. Always use one of the primary leads of the external coil as the ground return point. Other points used for ground may cause voltage loops that can damage the unit. We use a dedicated gap with connection leads containing clips. The gap is adjustable from ½ to 3".

S3 disables the internal test coil.

J1 is intended for input from an external trigger and requires R13 to be fully off.

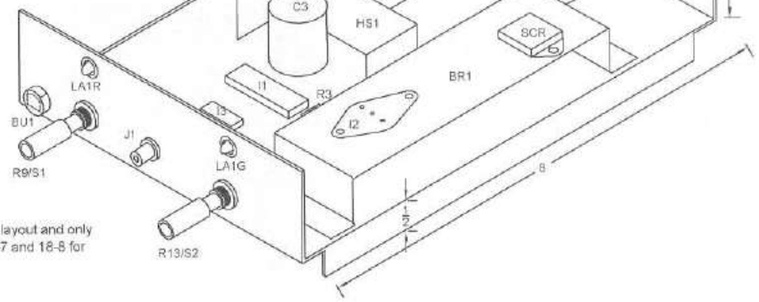
CHASSIS
TEST COIL
S3
J2
T1
C3
HS1
SCR
BR1
I1
R3
I3
I2
LA1R
BU1
J1
R9/S1
LA1G
R13/S2
8
1 9/16

Note this view is intended to show the general layout and only indicates certain components. Use figures 18-7 and 18-8 for wiring of the specific parts as shown

Figure 18-9 *Final assembly*

7. Wire the external components as shown in Figure 18-7.
8. Wire the transistors from the assembled bracket, as shown in Figure 18-8 and detailed in Figure 18-4.

Fabricate the mounting chassis at the overall dimensions shown in Figure 18-9. Trial-fit the remaining components, proceeding to drill and locate with the proper sized holes. Attempt to follow the figure and note the position of the BR1 bracket assembly and front panel controls. Route and twist all lead pair to neaten appearance. You may want to splice in some output leads to replace the connections to the test pulse transformer T2. These should be #18 wire and as short as physically possible to the external ignition coil. Also create a suitable cover if required and ventilate it with holes for air cooling.

Test Steps

This step will require basic electronic laboratory equipment, including a 60 MHz oscilloscope.

1. Turn off both controls R9 and R13. Set the R3 trimpot at midrange.
2. Connect a scope to TP2 and a voltmeter to TP1.
3. Apply 12 VDC to the input leads and turn on R9/S1. Adjust the meter for 5 volts and adjust the period of the waveshape at TP2 to a 25 us period. The input current should be less than .2 amps on a 12-volt power supply.
4. Note that the LED indicators are on with LA1G flashing at the pulse repetition rate determined by the R13 setting. This occurs even with R13/S2 fully off.
5. Turn on R13 and adjust it for 10 volts. Note the test coil sparking over its outer surface. Rotate R13 and notice the pulse rate increasing. Only operate the test coil to test circuit performance.

Notes on Operation

The unit is intended to drive larger ignition and induction coils. Jack J2 is intended as the output to the external coil under operation.

Note that the test coil must be disabled for external operation and may be disconnected by unsoldering or using a suitable high-current switch that may be mounted on the rear panel adjacent to J2.

When using unknown coils, it is advised that you monitor the DC voltage across the SCR. It should never exceed 400 volts, as the storage capacitor will be overrated. The SCR can handle up to 800 volts.

A change occurs in the charging voltage as the pulse repetition rate is changed. It decreases as the repetition rate increases. You can compensate for this by adjusting R9, but you must be careful to turn it down before lowering the repetition rate. Always start with the voltage control R9 at the lowest setting and increase as necessary. Additionally, jack J1 is intended for external triggering and must be used with R13/S2 in the off position.

Table 18-1 Parts List for CDI Driver

Ref. #	*Description*	*DB Part #*
R1, 5, 6, 7, 11, 12	Six 10-ohm ¼-watt resistors (br-blk-blk)	
R2, 14	Two 1K, ¼-watt resistors (br-blk-red)	
R3	5K horizontal trimmer	
R4, 15	Two 470 ohm, ¼-watt resistors (yel-pur- br)	
R8	15-ohm, 3-watt metal oxide resistor	
R9/S1, R13/S2	Two 10K pots and switches	
R10	220-ohm ½-watt resistor (red-red-br)	
R16	12-ohm, 2-watt metal oxide resistor	
Rx	Select a ¼-watt resistor to shunt R9 to improve range	
C1	100-microfarad, 25-volt vertical electrolytic capacitor	
C2	4,700-picofarad, 50-volt polyester capacitor	
C3	2,200-microfarad, 25-volt vertical electrolytic capacitor	
C4, 7, 13	Three .1-microfarad, 50-volt polyester capacitors	
C5	.0033-microfarad, 250-volt polypropylene capacitor	
C6	3- to 5-microfarad, 350-volt polyester capacitor	DB# 3.9M/350
C9	1-microfarad, 25-volt vertical electrolytic capacitor	
C8, 10	Two 10-microfarad, 25-volt vertical electrolytic capacitors	
C11	.47-microfarad, 50-volt polyester capacitor	
C12	.01-microfarad, 50-volt disc capacitor	
C14	.05-microfarad, 400-volt metalized polypropylene	
I1	*Integrated circuit* (IC) driver and pulse-width modulator (LM3525)	
I2	1.5-amp adjustable voltage regulator (LM317)	
I3	Timer *dual inline package* (DIP) IC (LM555)	
Q1, 2	Two 100-volt *metal-oxide-semiconductor field effect transistors* (MOSFETS) (IRF540)	
SCR1	70-amp, 700-volt *silicon-controlled rectifier* (SCR)	DB# S8070W
D1	IN914 small-signal diode	
D2, 3	Two IN4937 fast-recovery, 1,000-volt, 1-amp diodes	
LA1R	Red *light-emitting diode* (LED)	
LA1G	Green LED	
T1	Inverter high-voltage transformer, as shown in Figure 18-3	DB# TRANCDI
T2	20-kilovolt pulse transformer (see text)	DB# CD25B
FH1	Panel-mount fuse holder	
FS1	3-amp fuse	
J1, 2	RCA phono jack	
S3	*Single pole, single throw* (SPST) 5-amp or more toggle switch	
PB1	7 × 3.8 × .1-inch perforated circuit vector board	
SOCK8	Eight-pin IC socket	
SOCK16	16-pin IC socket	
MK1	TO3 transistor mounting kit	
PAD	Thermo pad insulator	
CHASSIS	#20-gauge aluminum chassis, as shown in Figure 18-9	
HS1	Heatsink, as shown in Figure 18-4	
BR1	Transistor-mounting bracket, as shown in Figure 18-4	
MICA	Two Mica insulating pads for Q1 and Q2	
SWN1	Two 6-32 × 1/2-inch nylon screws	
SW1	6-32 × 1/2-inch screws	
NU1	Seven 6-32 nuts	
WIRE20B	12 inches of #20 bus wire	
WR20	Six feet of #20 vinyl hookup wire	
BU1	Small clamp bushing	
BU2, 3	Two ⅜-inch plastic bushings for LED	

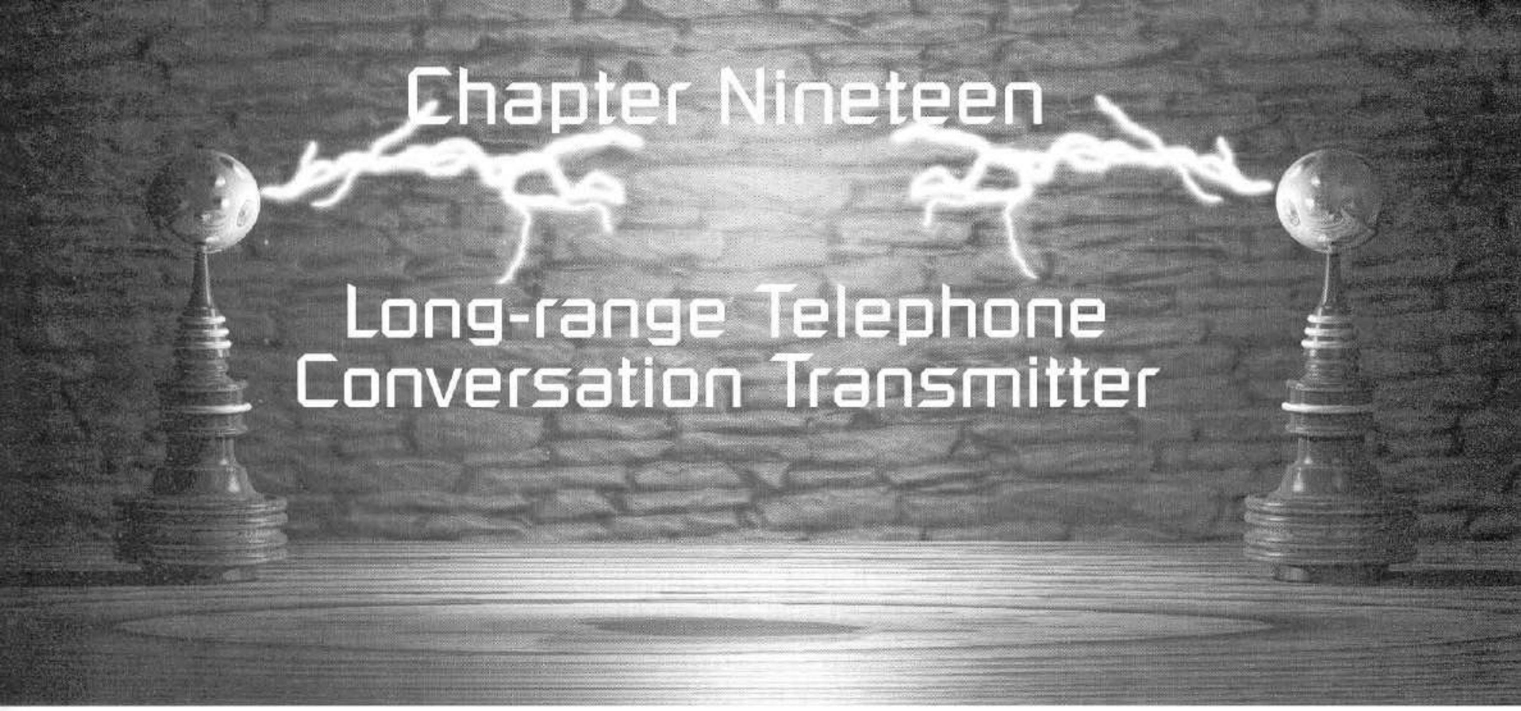

Chapter Nineteen

Long-range Telephone Conversation Transmitter

Figure 19-1 shows a device that allows the user to monitor a phone's incoming and outgoing calls without any connecting wires. The system is intended to transmit both sides of a telephone conversation to any FM radio or headset-type receiver, such as a Walkman radio. It allows the user to perform outside activities such as mowing the lawn or working outside the house while completely monitoring the telephone over the radio.

The unique feature of this device is that activation occurs only when the phone is off the hook. This eliminates transmitting a continuous signal, thus preventing potential interference and unnecessarily wearing down the batteries. It also eliminates dead-tape listening time, as now the tape only records voices when the telephone is used. You may also listen to a radio station while monitoring the telephone. When the phone is used, the radio station is replaced by the transmission of the unit.

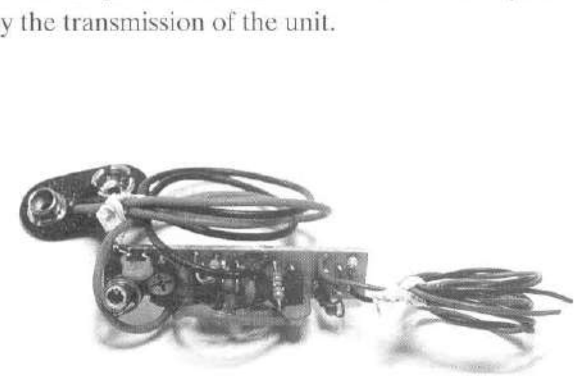

Figure 19-1 *Wireless circuit*

The device is shown assembled on a perforated circuit board and may be housed in a small plastic enclosure. Potting (or encapsulating with epoxy) and sealing is at the discretion of the builder.

Circuit Description

This chapter shows how to build a mini-powered FM transmitter that transmits both sides of a telephone conversation to a nearby FM radio. The circuit consists of a *radio-frequency* (RF) oscillator section combined with a telephone-activated switch section that controls the oscillator power (see Table 19-1). Transistor Q1 forms a relatively stable RF oscillator whose frequency is determined by the values of coil LI and tuning capacitor C4. The C4 setting determines the desired operating frequency and is in the standard FM broadcast band with tuned circuit design, favoring the high end up to 110 MHz. Capacitor C2 supplies the necessary feedback voltage developed across resistor R3 in the emitter circuit of Q1, sustaining an oscillating condition. Resistors R1 and R2 provide the necessary bias of the base emitter junction for proper operation, while capacitor C1 bypasses any RF to ground fed to the base circuit. Capacitor C3 provides an RF return path for the tank circuit of L1 and C4 while blocking the DC supply voltage fed to the collector of Q1 (see Figure 19-2).

You will note that the junction of the base bias resistors R1 and R2 is a feed point consisting of capacitor C5 and resistor R4. This point is where the speech voltage from the telephone is applied. Because of the nature of the oscillator frequency being subject to change by varying the base bias condition, a varying AC voltage superimposed at this point causes a corresponding frequency shift (FM) along with an amplitude-modulated (AM) condition. It is this property that allows the circuit to be FM modulated by the speech occurring on the telephone lines. The signal is clearly detected by any FM receiver when properly tuned.

Be aware that before the phone receiver is lifted off the hook, a DC voltage of approximately 50 volts along with some AC hum is measured across the green and red wires of the phone line. This 50 volts is used to keep transistors Q2 and Q3 "off" by preventing forward biasing of the Q2 junction. The zener diode Z1 reverse voltage is exceeded as long as the voltage on the telephone lines remains above the zener voltage. When the phone is off the hook, the voltage is now below the zener voltage, and resistor R5 now biases Q2 "on," which turns Q3 on and clamps the common return of the oscillator to the negative of the battery, commencing operation by turning on.

Note that Q2 and Q3 make a simple DC switch that produces a minimal battery current during the on-hook conditions. When the receiver is lifted, commencing normal conversation, this DC voltage drops to less than 10 volts, and an existing AC voltage corresponding to the speech conversation now modulates the oscillator as described. You will also note Q2 and Q3 being a simple DC switch that only allows a minimal battery current during the on-hook condition.

Capacitor C5 is necessary to block the DC component of the phone line when an on-hook condition exists and allows little attenuation of the varying AC speech voltage. Resistor R4 is necessary for attenuation of this speech voltage that could cause overmodulation and unnecessary serious sidebands if not properly selected.

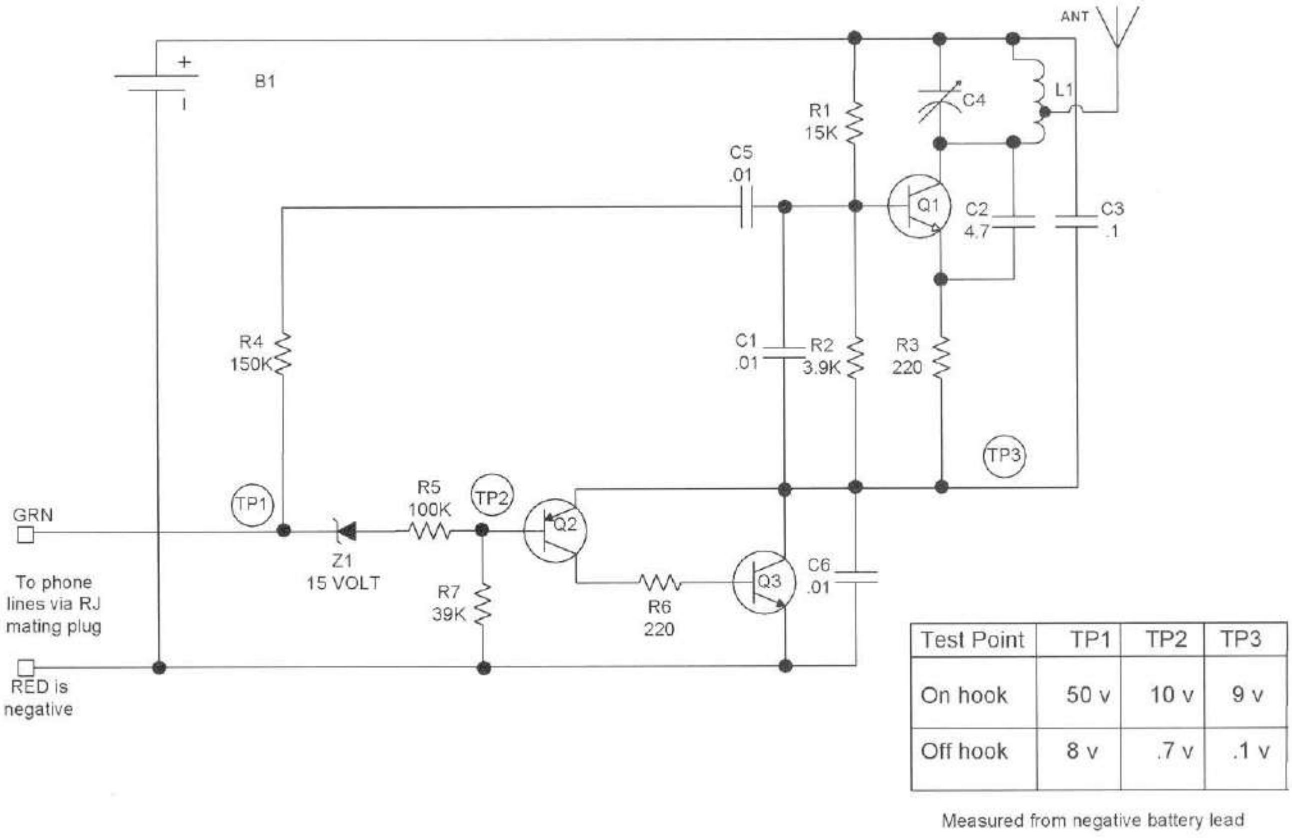

Test Point	TP1	TP2	TP3
On hook	50 v	10 v	9 v
Off hook	8 v	.7 v	.1 v

Measured from negative battery lead

Figure 19-2 *Circuit schematic*

Assembly of the Circuit Board

Assemble the circuit as shown in Figure 19-3.

1. Fabricate a piece of .1-inch grid perforated board to a size of 2 × 1 inches.
2. Form L1 by tightly wrapping eight turns of #16 bus wire on a #8 wood screw. This produces an eight-turn coil with an inner diameter of approximately .135 inches and is about .625 inches in length. Insert it in the proper hole and solder as shown.
3. Insert trimmer capacitor C4 into the holes as shown. You may put this component on either side of the PC board. This choice is up to you and should be determined by the final packaging scheme you require. Three holes are available for the part, with two of them electrically the same point. This is necessary as some trimmers may have three pins. Make sure that the common pins are connected to the same electrical points with the odd pin to the other point.
4. If you are building from a perforated board, it is suggested that you insert components starting in the lower left-hand corner. Pay attention to the polarity of the capacitors with polarity signs and all the semiconductors. Route the leads of the components as shown and solder as you go, cutting away unused wires. Attempt to use certain leads as the wire runs or use pieces of the #24 bus wire. Follow the dashed lines on the assembly drawing as these indicate connection runs on the underside of the assembly board.
5. Attach the external leads of battery clip CL1 and the RJ cable and plug them into the phone connection jack.
6. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.
7. Tune an FM receiver to a fairly strong station at the high end of the band (108 MHz or higher). Turn up the volume and position it at a distance of 25 to 50 feet.
8. If a multimeter or 50-milliamp meter is available, connect it in series with the battery lead.

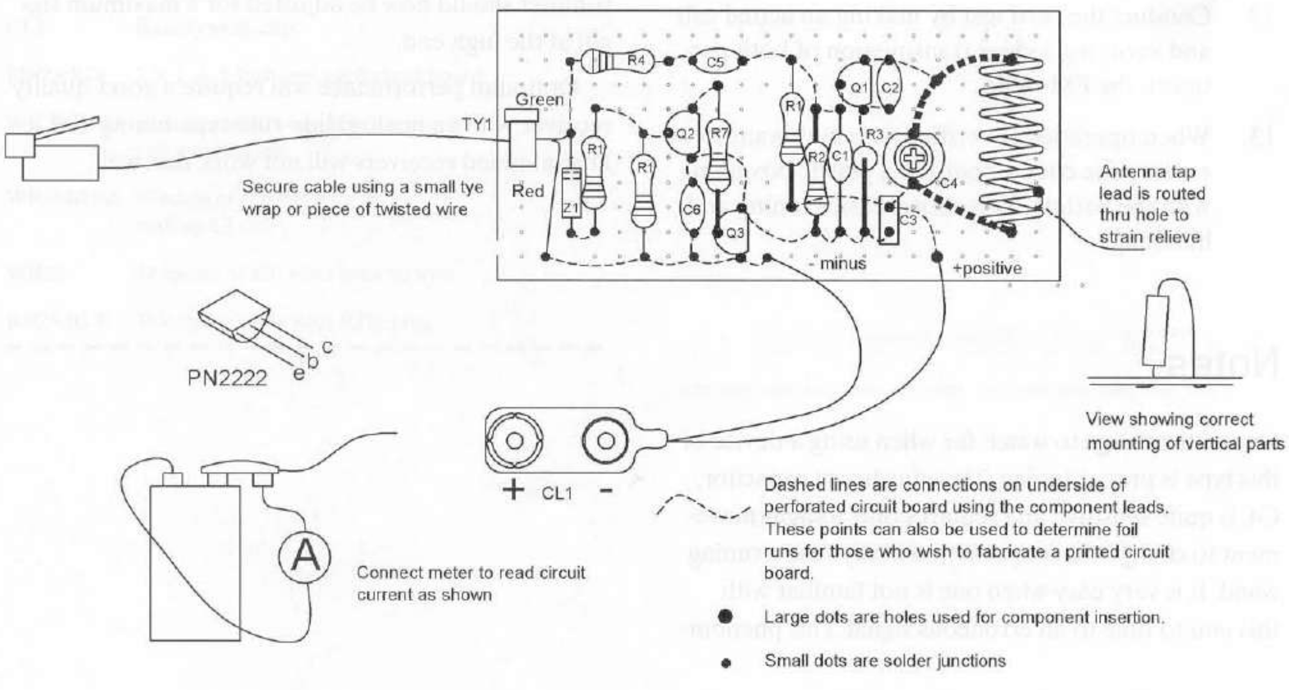

Secure cable using a small tye wrap or piece of twisted wire

Antenna tap lead is routed thru hole to strain relieve

View showing correct mounting of vertical parts

Connect meter to read circuit current as shown

Dashed lines are connections on underside of perforated circuit board using the component leads. These points can also be used to determine foil runs for those who wish to fabricate a printed circuit board.

● Large dots are holes used for component insertion.

• Small dots are solder junctions

Figure 19-3 *Assembly of circuit board*

This can be done by removing one of the clip fasteners and connecting the meter to the free contacts, as shown in Figure 19-3. The meter should read 5 to 7 milliamps. Use a short piece of bare wire to touch the coil L1 a turn at a time, starting from the C3 end. Note that as you progress turn by turn away from C3 that the indicated meter current will drop or change.

9. Connect the operating unit to the phone line with the phone on hook and notice the meter reading dropping to almost nothing. You may have to reverse the connections to the phone lines. Check the following voltage points indicated on the schematic to verify proper circuit operation, as shown in Figure 19-2.
10. Take the phone off the hook and tune C4 with a tuning wand until you hear the telephone tone. It may be difficult at first to spot the signal as this adjustment is very touchy. It is important to know that several spots in the adjustment may be erroneous and will be weak and unstable. The correct signal will be strong, stable, and clear, and it can be verified by distance-of-transmission testing.
11. Once the desired setting of C4 is found, the position should be marked on the adjustment capacitor with the frequency noted.
12. Conduct the final test by making an actual call and verifying a clear transmission of both parties to the FM radio.
13. When operation is verified, you may want to enclose the circuit board in a plastic box along with the battery. This allows easier tuning and handling.

Notes

One of the things to watch for when using a device of this type is proper tuning. The adjustment capacitor, C4, is quite sensitive and requires only a slight movement to change the frequency, so always use a tuning wand. It is very easy when one is not familiar with this unit to tune to an erroneous signal. This phenomenon is likely to occur when the unit is close to the monitoring receiver. As stated previously, an erroneous signal will be weak, distorted, and unstable (often it is mistaken for the main signal and the unit is blamed for poor performance), and the main signal will be strong, stable, and undistorted, if modulated.

Several experiments in tuning the unit should be done before attempting to use it for the desired application. Also, whenever possible, the unit should be used around 108 to 109 MHz, which is at the boundary of the aircraft band and upper FM broadcast band. When the approximate desired spot is found, final touchup tuning should be done at the receiver end for clarity. In most areas, these upper frequencies are clear and allow uninterrupted use in contrast to the lower ones in the FM band where a slight change in frequency from a clear spot results in the unit being drowned out by a strong broadcast station. Do not go above 108 MHz if near an airport or air traffic lane.

Most available FM radios can easily be detuned slightly to shift the dial readings down to where the dial reading of 108 MHz is actually 109 MHz. This is accomplished by carefully adjusting the "oscillator," the padding trimmer located on the main tuning capacitor, and "walking" a known station down the necessary megahertz or two. An antenna-peaking trimmer should now be adjusted for a maximum signal at the high end.

Optimum performance will require a good-quality receiver with an analog slide-rule-type tuning dial. Digital tuned receivers will not work that well.

Table 19-1 Long-Range Phone Conversation Transmitter Parts List

Ref. #	*Description*	*DB Part #*
R1	15K, 1/4-watt resistor (br-gr-or)	
R2	3.9K, 1/4-watt resistor (or-wh-red)	
R3, 6	Two 220 ohm, 1/4-watt resistor (red-red-br)	
R4	150K, 1/4-watt resistor (br-gr-yel)	
R5	100K, 1/4-watt resistor (br-blk-yel)	
R7	39K, 1/4-watt resistor (or-wh-or)	
C1, 5, 6	Three .01-microfarad, 50-volt disc capacitors	
C2	4- to 6-picofarad small silver mica disk (4.7)	DB# 4.7P
C3	.1-microfarad, 50-volt disc capacitor	
C4	6- to 35-picofarad trimmer capacitor	DB# 635P
Z1	15-volt zener diode (1N5245)	
Q1, 3	Two PN2222 NPN general-purpose transistors	
Q2	PN2907 PNP general-purpose transistor	
L1	Coil wound as described in assembly steps	DB# COIL8T
CL1	Battery snap clip	
PBOARD	2 × 1 × .1-inch grid perforated board	
WR24BUSS	12 inches of #24 bus wire for wiring and antenna tap	
WR16BUSS	6 inches of #16 bus wire for making L1 coil	
WR20	24 inches of #20 vinyl hookup wire	
RJCABLE	Telephone cable with RJ11 plug	

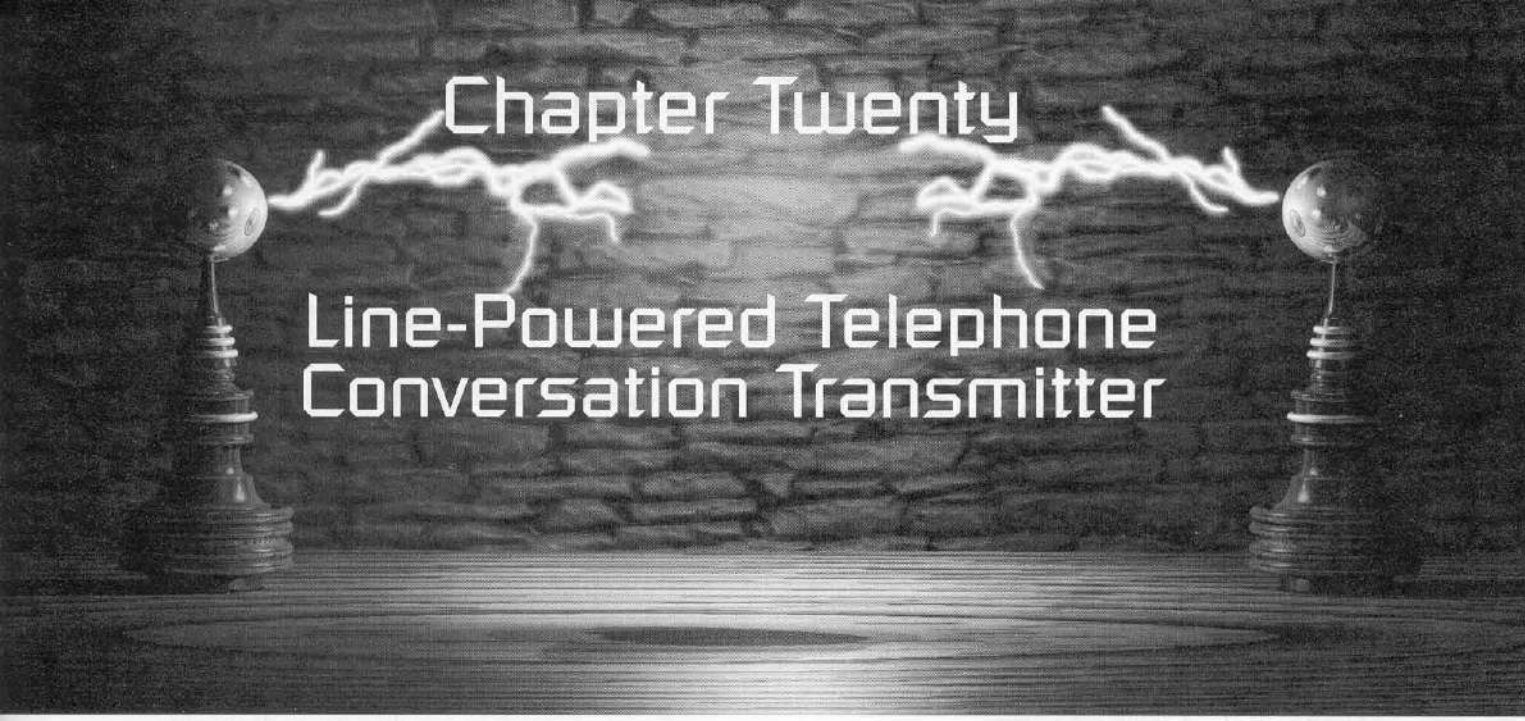

Chapter Twenty

Line-Powered Telephone Conversation Transmitter

This project shown in Figure 20-1 is similar to that described in Chapter 19, but this one can be permanently implanted in a telephone and can operate indefinitely, as it does not require batteries. The circuit as shown derives its power from the phone lines once the receiver is lifted off of the hook. Operating range is limited to several hundred feet, far less than the battery-powered version.

The unique features of this device is that it does not require a battery, and activation occurs only when the phone is off the hook. This eliminates transmitting a continuous signal, thus preventing potential interference and eliminates dead-tape listening time. You may also listen to a radio station while monitoring the telephone. When the phone is used, the radio station is replaced by the transmission of the unit. The device is shown assembled on a perforated circuit board and may be housed in a small plastic enclosure. Potting and sealing is at the discretion of the builder.

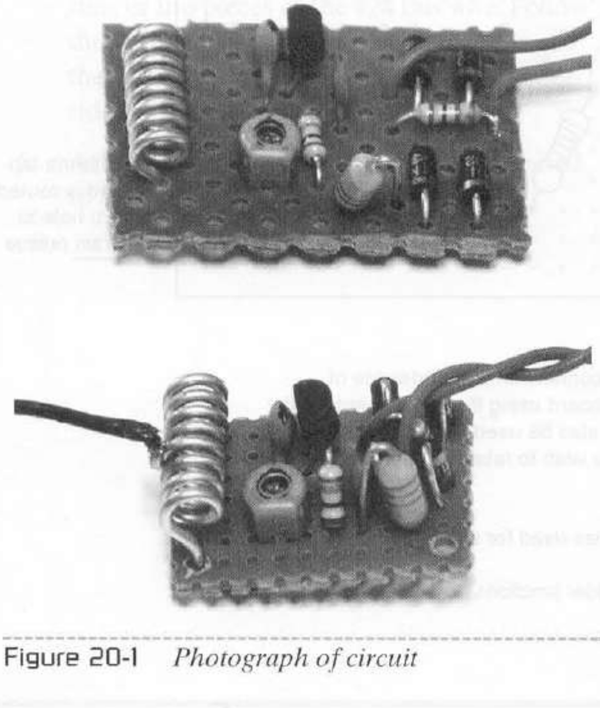

Figure 20-1 *Photograph of circuit*

Circuit Description

This chapter shows how to build a mini-powered FM transmitter that transmits both sides of a telephone conversation to a nearby FM radio. The circuit consists of a *radio-frequency* (RF) oscillator section combined with a telephone power conditioner that powers the oscillator (see Table 20-1). Transistor Q1 forms a relatively stable RF oscillator whose frequency is determined by the value of coil LI and tuning capacitor C3. The setting of C3 determines the desired operating frequency and is in the standard FM broadcast band with tuned circuit design favoring the high end up to 110 MHz. Capacitor C2 supplies the necessary feedback voltage necessary to sustain oscillation. Resistor R2 provides the necessary bias of the base emitter junction for proper operation, while

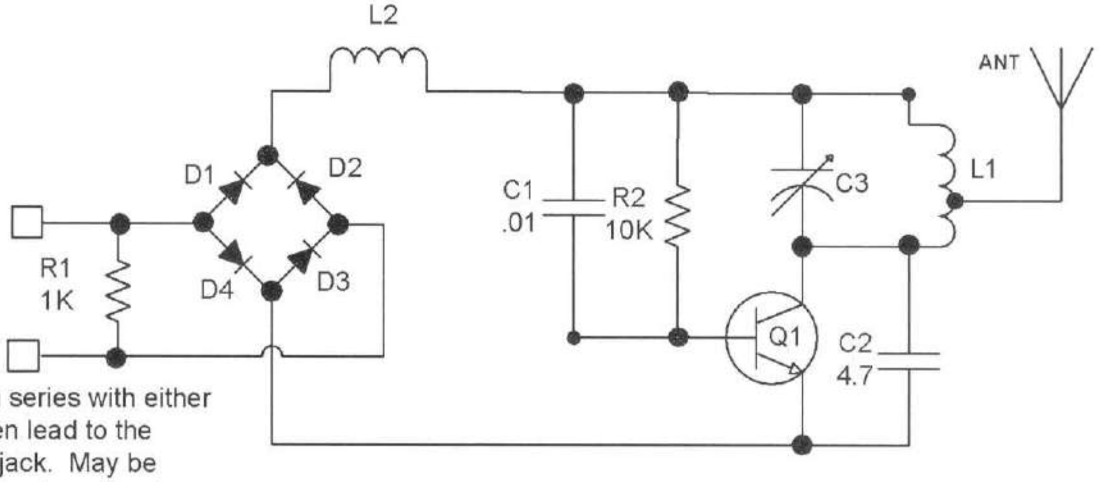

Connect in series with either red or green lead to the telephone jack. May be installed anywhere line is accessible. Connection will mean breaking one of the leads and connecting in place.

Figure 20-2 *Circuit schematic*

capacitor C1 bypasses any RF to ground. Choke L2 blocks the signal energy from going back to the phone lines yet allows DC power to Q1 (see Figure 20-2).

Connecting the circuit requires breaking one of the lines and connecting the input leads in series with the telephone. It is this method that leaches voltage from the series load resistor R1 when the telephone is in use, as current now flows across R1. The bridge diodes consisting of D1 through D4 polarize the audio voltages necessary to modulate the oscillator.

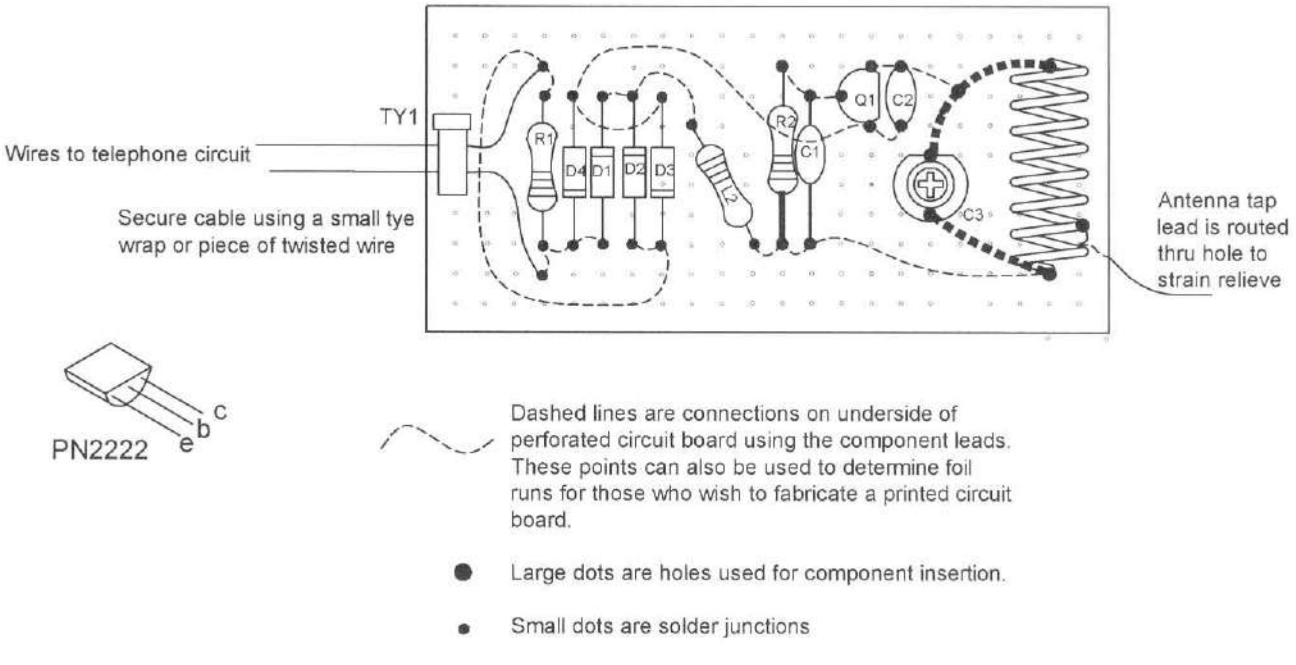

Wires to telephone circuit

Secure cable using a small tye wrap or piece of twisted wire

Antenna tap lead is routed thru hole to strain relieve

Dashed lines are connections on underside of perforated circuit board using the component leads. These points can also be used to determine foil runs for those who wish to fabricate a printed circuit board.

Large dots are holes used for component insertion.

Small dots are solder junctions

Figure 20-3 *Assembly of circuit board*

Assembly of the Circuit Board

Assemble the circuit as shown in Figure 20-3:

1. Cut a piece of .1-inch grid perforated board to 2 × 1 inches.
2. Form L1 by tightly wrapping eight turns of #16 bus wire on a #8 wood screw. This produces an eight-turn coil with an inner diameter of approximately .135 inches and is about .625 inches in length. Insert it in the proper hole and solder as shown.
3. Insert trimmer capacitor C4 into the holes as shown. You may put this component on either side of the PC board. The choice is up to you and should be determined by the final packaging scheme you require. Three holes are available for the part, with two of them electrically the same point. This is necessary, as some trimmers may have three pins. Make sure that the common pins are connected to the same electrical points, with the odd pin to the other point.
4. If you are building from a perforated board, it is suggested that you insert the components starting in the lower left-hand corner. Pay attention to the polarity of the capacitors that have polarity signs and all the semiconductors. Route the leads of the components as shown and solder as you go, cutting away unused wires. Attempt to use certain leads as the wire runs or use pieces of the #24 bus wire. Follow the dashed lines on the assembly drawing as these indicate connection runs on the underside of the assembly board.
5. Attach the external leads that are connected to the telephone lines.
6. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.
7. Tune an FM receiver to a fairly strong station at the high end of the band (108 MHz or higher). Turn up the volume and position it 25 to 50 feet away from the transmitter.
8. Connect the operating unit to the phone line with the phone on the hook. Remove the phone and adjust C3 until you hear the dial tone. Tuning the trimmer must be done with an insulating tuning wand. It is best to tune the unit by placing it on a nonconductive surface and securing it with a sponge or a piece of rubber.

   It may be difficult at first to spot the signal as this adjustment is very touchy. Also note that several spots in the adjustment may be erroneous and will be weak and unstable. The correct signal will be strong, stable, and clear and can be verified through distance-of-transmission testing.
9. Once the desired setting of C4 is found, it should be marked, with the frequency noted.
10. Conduct a final test by making an actual call and verifying the clear transmission of the FM radio to both parties.
11. When operation is verified, it may be a good idea to enclose the circuit board in a plastic box. This allows easier tuning and handling.

Notes

Please see notes in Chapter 19.

Table 20-1 Line-Powered Telephone Conversation Transmitter Parts List

Ref. #	*Description*	*DB Part #*
R1	1K ¼-watt resistor (br-blk-red)	
R2	10K ¼-watt resistor (br-blk-or)	
C1	.01-microfarad, 50-volt disc capacitor	
C2	4- to 6-picofarad small silver mica disk (4.7)	DB# 4.7P
C3	6- to 35-picofarad trimmer capacitor	DB# 635P
D1–4	100-volt diodes IN4002	
Q1	PN2222 NPN general-purpose transistor	
L1	Coil wound as described in assembly steps	DB# COIL8T
L2	3.6 microhenry inductor	DB# 3.6U
PBOARD	2 × 1 × .1-inch grid perforated board	
WR24BUSS	12 inches of #24 bus wire for wiring and antenna tap	
WR16BUSS	6 inches of #16 bus wire for making L1 coil	
WR20	24 inches of #20 vinyl hookup wire	

Chapter Twenty-One

Remote Wireless FM Repeater

Figure 21-1 shows a wireless FM transmitter designed to transmit the output of a tape recorder, VCR, DVD, or any device having composite audio and video outputs to a nearby radio or TV tuned to the unit's frequency. Assembly is built on a piece of circuit board and may be placed in any suitable plastic container along with a 9-volt battery.

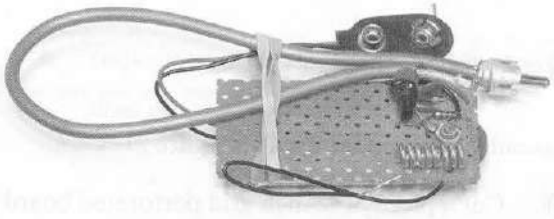

Figure 21-1 *The completed device*

Circuit Description

The parts for this project are listed in Table 21-1. Transistor Q1 forms a relatively stable *radio-frequency* (RF) oscillator whose frequency is determined by the value of coil L1 and tuning capacitor C4. The C4 setting determines the desired operating frequency and is in the standard FM broadcast band with tuned circuit design favoring the high end up to 110 MHz. Capacitor C2 supplies the necessary feedback voltage developed across emitter resistor R3 in the emitter circuit of Q1, sustaining an oscillating condition. Resistor R1 and R2 provide the necessary bias of the base emitter junction for proper operation, while capacitor C1 bypasses any RF to ground fed to the base circuit. Capacitor C9 provides an RF return path for the tank circuit of L1 and C4 while blocking the DC supply voltage fed to the collector of Q1 (see Figure 21-2). R4 is connected in series with DC-blocking capacitor C5 to the input cable and plug.

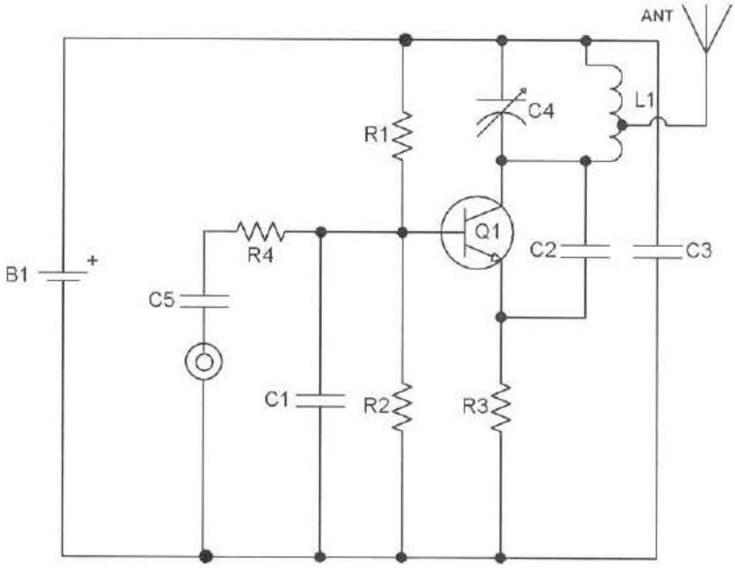

Figure 21-2 *Circuit schematic*

Assembly

Assemble the circuit as shown Figure 21-3.

1. Cut a piece of .1-inch grid perforated board to a size of 2 × 1 inches.
2. Form L1 by tightly wrapping eight turns of #16 bus wire on a #8 wood screw. This produces an eight-turn coil with an inner diameter of approximately .135 inches and a length of .625 inches. Insert it in the proper hole and solder as shown.
3. Insert trimmer capacitor C4 into the holes as shown. You may put this component on either side of the PC board. This choice is up to you and should be determined by the final packaging scheme you require. Three holes are available for the part with two of them electrically the same point. This is necessary, as some trimmers may have three pins. Make sure that the common pins are connected to the same electrical points with the odd pin to the other point.
4. If you are building from a perforated board, it is suggested that you insert components starting in the lower left-hand corner. Pay attention to the polarity of the capacitors that have polarity signs and all the semiconductors. Route the leads of the components as shown and solder as you go, cutting away unused wires. Attempt to use certain leads as the wire runs or use pieces of the #24 bus wire. Follow the dashed lines on the assembly drawing as these indicate the connection runs on the underside of the assembly board.
5. Attach the external leads of the battery clip (CL1) and shielded cable (WR10).
6. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.
7. Tune an FM receiver to a fairly strong station at the high end of the band (108 MHz or higher). Turn up the volume and position the receiver about 25 to 50 feet from the transmitter circuit.
8. If a multimeter or 50-milliamp meter is available, connect it in series with a battery lead. This can be done by removing one of the clip fasteners and connecting the meter to one of the free contacts, as shown in Figure 21-2. The meter should read 5 to 7 milliamps. Pick up a short piece of bare wire and touch coil L1. Touch it a turn at a time, starting from the C3

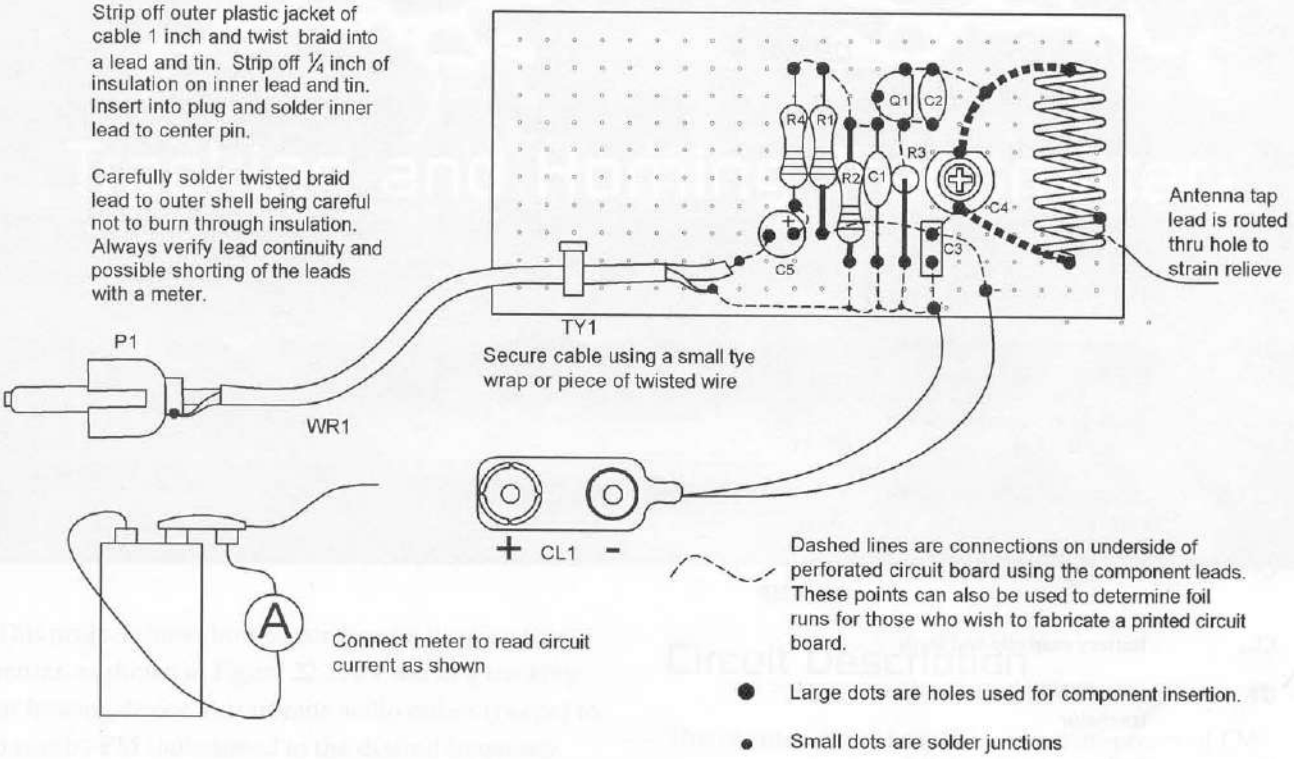

Strip off outer plastic jacket of cable 1 inch and twist braid into a lead and tin. Strip off ¼ inch of insulation on inner lead and tin. Insert into plug and solder inner lead to center pin.

Carefully solder twisted braid lead to outer shell being careful not to burn through insulation. Always verify lead continuity and possible shorting of the leads with a meter.

Secure cable using a small tye wrap or piece of twisted wire

Antenna tap lead is routed thru hole to strain relieve

Connect meter to read circuit current as shown

Dashed lines are connections on underside of perforated circuit board using the component leads. These points can also be used to determine foil runs for those who wish to fabricate a printed circuit board.

● Large dots are holes used for component insertion.

• Small dots are solder junctions

Figure 21-3 *Assembly of board*

end. Note that as you progress turn by turn away from C3 that the indicated meter current will drop or change.

9. Connect the cable into an audio source such as a recorder or another radio. Slowly rotate C4 with an insulated tuning wand until the station being received by the radio is at approximately 108 MHz. It may be difficult at first to spot the signal, as this adjustment is very touchy. Also note that several spots in the adjustment may be erroneous and will be weak and unstable. The correct signal will be strong, stable, and clear and can be verified by distance-of-transmission testing.
10. Once the desired setting of C4 *is* found, it should be marked with the frequency noted.
11. When the operation is verified, it may be desired to pot the assembly as shown. This allows easier tuning and handling.

Notes

Please refer to the notes in Chapter 19 on operation of wireless circuits.

Table 21-1 Remote Wireless FM Repeater Parts List

Ref. #	*Description*	*DB Part #*
R1	15K, 1/4-watt resistor (br-gr-or)	
R2	3.9K, 1/4-watt resistor (or-wh-red)	
R3	220-ohm, 1/4-watt resistor (red-red-br)	
R4	1K, 1/4-watt resistor (br-blk-red)	
C1, 5	Two .01-microfarad, 50-volt disc capacitors	
C2	Small 4- to 6-picofarad silver mica disk (4.7)	DB #4.7P
C3	.1-microfarad, 50-volt plastic capacitor	
C4	6- to 35-picofarad trimmer capacitor	DB #635P
CL1	Battery snap clip and leads	
Q1	PN2222 NPN general-purpose transistor	
L1	Coil wound as described in assembly steps	DB #Coil8T
PBOARD	2 × 1 × .1-inch grid perforated board	
WR24BUSS	12 inches of #24 bus wire for wiring and antenna tap	
WR16BUSS	6 inches of #16 bus wire for making L1 coil	
WR1	12 inches of shielded cable	
P1	RCA phono plug	

Chapter Twenty-Two

Tracking and Homing Transmitter

This project shows how to construct a beeping transmitter, as shown in Figure 22-1, for use as a tracking or homing device. It transmits audio pulses (beeps) to a nearby FM radio tuned to the desired frequency. The unit is intended to be used in transmitter hunt games or as a homing beacon for hunters or hikers. It is suggested that you mark a bearing heading on a small but good-quality FM radio and use it as a low-cost receiver for determining the direction of the signal pickup. The range of the device, when used as a homing device with a quarter-wave vertical placed on top of a metal roof, such as a vehicle, can be in excess of 3 miles when using a quality analog FM receiver.

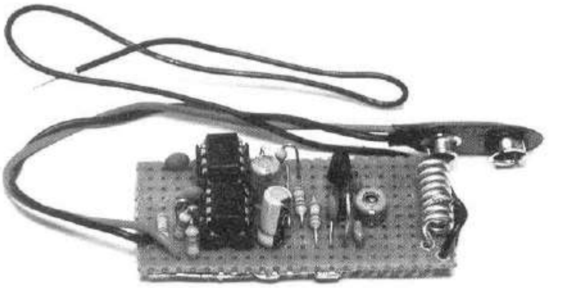

Figure 22-1 *Transmitter circuit*

Circuit Description

This chapter shows how to build a mini-powered FM transmitter that transmits audio pulses for tracking and homing applications (see Figure 22-2). The circuit consists of a *radio-frequency* (RF) oscillator section interfaced with a tone modulator and a pulse-switching section that controls the oscillator power. Transistor Q1 forms a relatively stable RF oscillator whose frequency is determined by the value of coil LI and tuning capacitor C4. The C4 setting determines the desired operating frequency and is in the standard FM broadcast band with the tuned circuit design favoring the high end up to 110 MHz. Capacitor C2 supplies the necessary feedback voltage developed across resistor R3 in the emitter circuit of Q1, sustaining an oscillating condition. Resistors R1 and R2 provide the necessary bias of the base emitter junction for proper operation, while capacitor C1 bypasses any RF to ground fed to the base circuit. Capacitor C3 provides an RF return path for the tank circuit of LI and C4 while blocking the DC supply.

You will note that the junction of the base bias resistors R1 and R2 is a feed point consisting of resistor R4. This point is where the beeping tone from the tone generator chip I2 is applied. Because of the nature of the oscillator frequency being subject to change by varying the base bias condition, a pulsed voltage superimposed at this point causes a corresponding frequency shift (FM) along with an amplitude modulated shift (AM) condition. It is this property that allows the circuit to be FM modulated, producing the tone frequency. The signal is clearly detected on any FM receiver when the receiver is properly tuned. The tone frequency is the pulse repetition rate of I1 wired as an astable pulse generator. This repetition rate is determined by timing resistor R5 and capacitor C6.

The ratio of beeping on to off time is controlled by chip I1, wired as a free-running astable device with duty cycle control. This function is determined by the ratio of resistors R7 and R8 and timing capacitor C8. The output of I1 controls the bias voltage of oscillator Q1 through resistor R1, turning the oscillator on and off (see Figure 22-2).

Assembly of the Circuit Board

Assemble the circuit as shown in Figure 22-3.

1. Cut a piece of .1-inch grid perforated board to a size of 2 × 1 inches.
2. Form L1 by tightly wrapping eight turns of #16 bus wire on a #8 wood screw. This produces an eight-turn coil with an inner diameter of .135 inches and a length of .625 inches. Insert it in the proper hole and solder as shown.
3. Insert trimmer capacitor C4 into the holes as shown. You may put this component on either side of the PC board. The choice is up to you and should be determined by the final packaging scheme you require. Three holes are available for the part, with two of them electrically the same point. This is necessary, as some trimmers may have three pins. Make sure that the common pins are connected to the same electrical points with the odd pin to the other point.
4. If you are building from a perforated board, it is suggested that you insert components, starting in the lower left-hand corner. Pay atten-

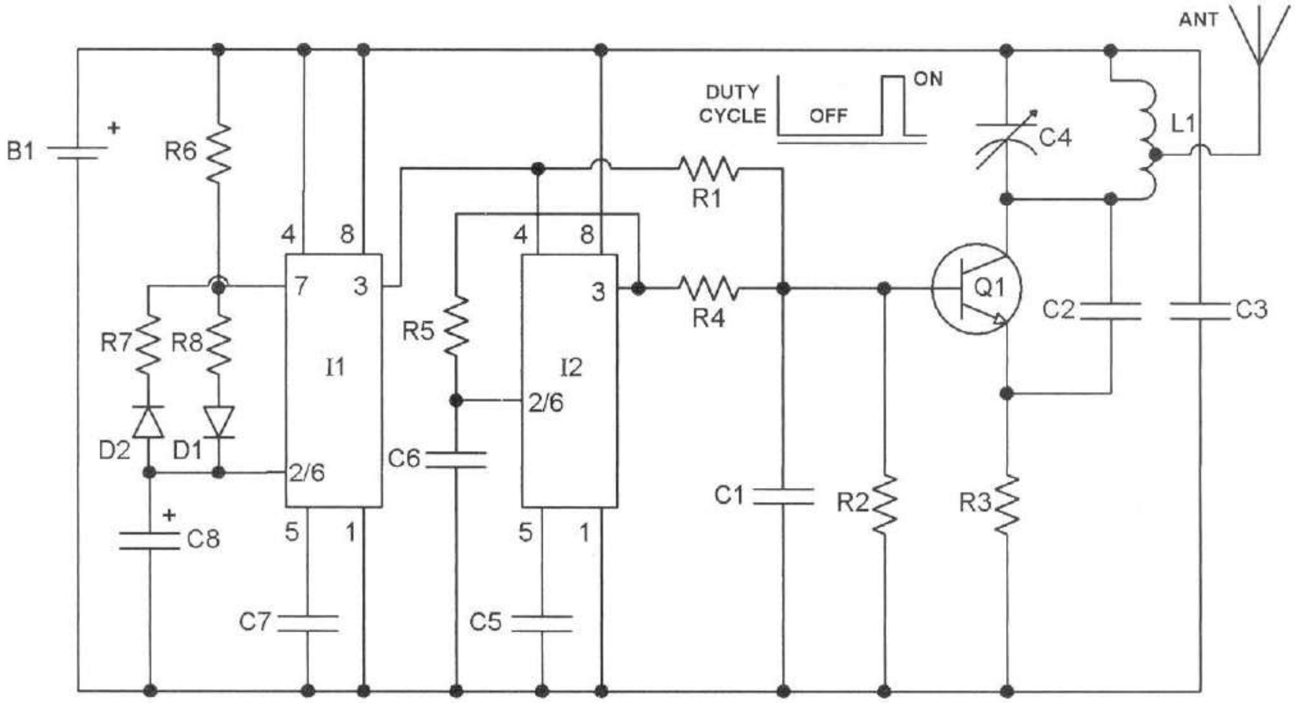

Figure 22-2 *Circuit schematic*

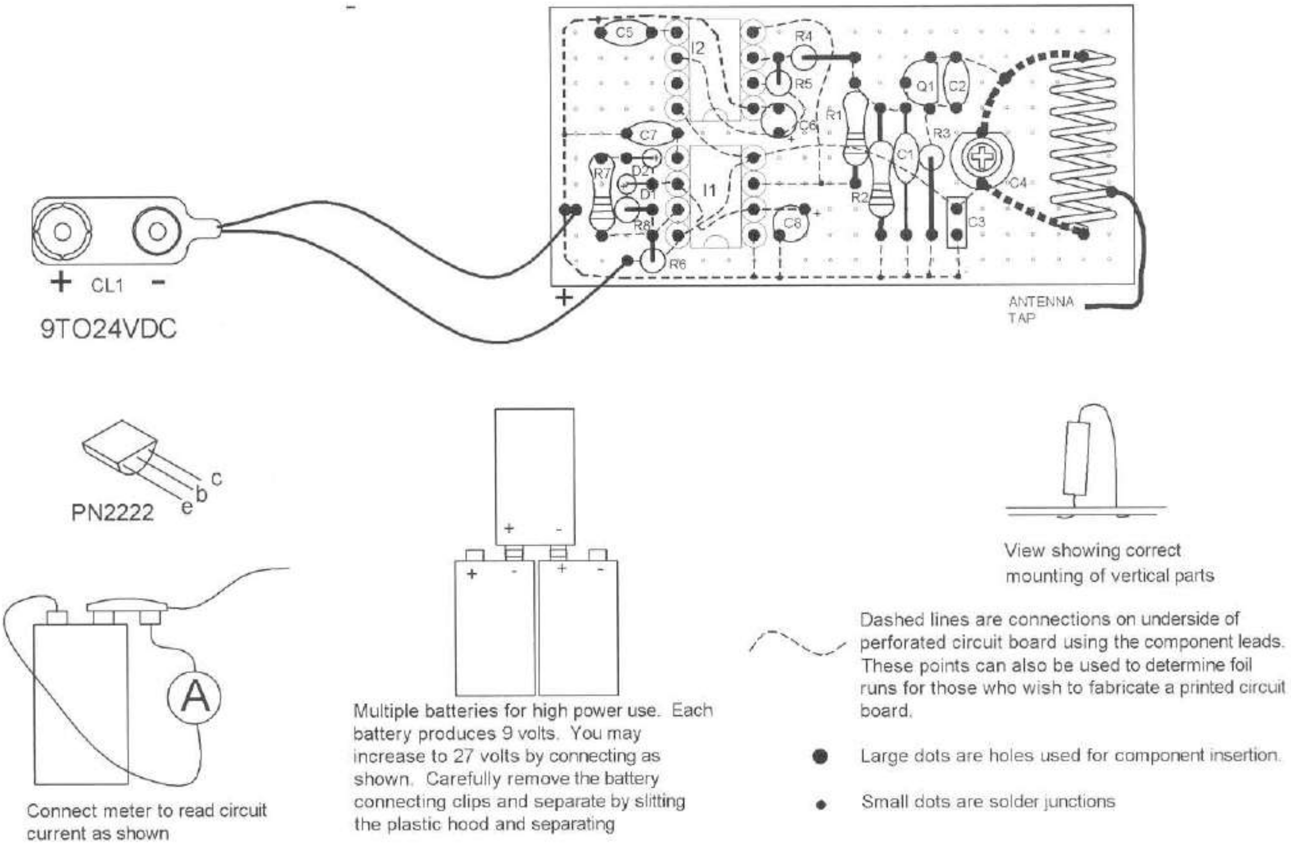

Connect meter to read circuit current as shown

Multiple batteries for high power use. Each battery produces 9 volts. You may increase to 27 volts by connecting as shown. Carefully remove the battery connecting clips and separate by slitting the plastic hood and separating

View showing correct mounting of vertical parts

Dashed lines are connections on underside of perforated circuit board using the component leads. These points can also be used to determine foil runs for those who wish to fabricate a printed circuit board.

● Large dots are holes used for component insertion.

• Small dots are solder junctions

Figure 22-3 *Assembly of the circuit board*

tion to the polarity of the capacitors with polarity signs and all the semiconductors. Route the leads of the components as shown and solder as you go, cutting away the unused wires. Attempt to use certain leads as the wire runs or use pieces of the #24 bus wire. Follow the dashed lines on the assembly drawing as these indicate the connection runs on the underside of the assembly board.

5. Attach the external leads of battery clip CL1.
6. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.
7. Tune an FM receiver to a fairly strong station at the high end of the band (108 MHz or higher). Turn up the volume and position the receiver 25 to 50 feet from the transmitter circuit.
8. If a multimeter or 50-milliamp meter is available, connect it in a series with a battery lead. This can be done by removing one of the clip fasteners and connect the meter to the free contacts, as shown in Figure 22-2. The meter should read a pulsing of 5 to 7 milliamps. Use a short piece of bare wire to touch the coil L1 a turn at a time, starting from the C3 end. Note that as you progress turn by turn away from C3 that the indicated meter current will drop or change.
9. Tune C4 with an insulated tuning tool until you hear the beeping tone. It may be difficult at first to spot the signal, as this adjustment is very touchy. Also note that several spots in the adjustment may be erroneous and will be weak and unstable. The correct signal will be strong, stable, and clear and can be verified by distance-of-transmission testing.

10. Once the desired setting of C4 is found, it should be marked with alignment marks, with the frequency noted.

11. Finally, test the unit using a good-quality FM analog receiver. Note to place an arrow on the radio once you determine the proper positioning to obtain the maximum signal strength. Do this step from a distance of at least several hundred feet.

12. When operation is verified, it may be desired to enclose the circuit board into a plastic box along with the battery. This allows easier tuning and handling.

Notes

Please see the notes in Chapter 19 on operating.

Table 22-1 Tracking and Homing Transmitter Parts List

Ref. #	*Description*	*DB Part #*
R1	Two 10K, 1/4-watt resistor (br-gr-or)	
R2	3.9K, 1/4-watt resistor (or-wh-red)	
R3	220-ohm, 1/4-watt resistor (red-red-br)	
R5	1K, 1/4-watt resistor (br-blk-red)	
R6	39K, 1/4-watt resistor (or-wh-or)	
R7	100K, 1/4-watt resistor (br-blk-yel)	
R8	22K, 1/4-watt resistor (red-red-or)	
C1, 5, 7	Three .01-microfarad, 50-volt disc capacitors	
C2	Small, 4- to 6-picofarad, silver mica disk (4.7)	DB# 4.7P
C3	.1-microfarad, 50-volt plastic capacitor	
C4	6- to 35-picofarad trimmer capacitor	DB# 635P
C6	.47-microfarad, 25-volt vertical electrolytic capacitor	
C8	4.7-microfarad, 25-volt vertical electrolytic capacitor	
D1, 2	Two 1N914 small-signal diodes	
I1, 2	Two 555 timers in a *dual inline package* (DIP)	
CL1	Battery snap clip and leads	
Q1	PN2222 NPN general-purpose transistor	
L1	Coil wound as described in assembly steps	DB# COIL8T
PBOARD	2 × 1 × .1-inch grid perforated board	
WR24BUSS	12 inches of #24 bus wire for wiring and antenna tap	
WR16BUSS	6 inches of #16 bus wire for making L1 coil	

Chapter Twenty-Three

Snooper Phone Room-Listening Device

Figure 23-1 shows a simple electronic device intended to be an intrusion detector and listening device for checking a home or office while away or on vacation. It may also be used to trigger other electrical devices. The circuitry does not require a tone encoder, so it does not provide selected activation by a particular caller. Thus, it limits the device from being used for illegal purposes and allows the sale of the kit or fully assembled unit without the requirement of special authorization.

This is an intermediate-level project requiring basic electronic skills. Expect to spend $25 to $40. All parts are readily available, with specialized parts obtainable through Information Unlimited (www.amazing1.com), and they are listed in Table 23-1.

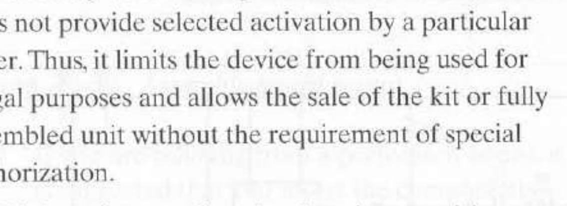

Figure 23-1 *The snooper phone*

Unlike the controversial Infinity transmitter, our system does not utilize decoder and encoder circuitry. Once connected to the phone lines, it will open a microphone to anyone who happens to dial the number. This feature obviously reduces the potential use of the device for illegal interception of oral communication. Even though the phone does not ring, it would not be long before the potential victim of an illegal surveillance would be aware of the setup. It is therefore quite obvious that the applications are strictly limited to use as a security device that checks for unauthorized intrusion or checks household appliances by placing a pickup device in certain favorable locations to overhear these desired sounds and noises. I am sure that anyone with this capability would breathe a sigh of relief to be able to dial his or her home or office phone while being away and hear the familiar sounds of appliances and other systems properly performing their duties and functions.

Circuit Operation

Your circuit consists of a high-gain amplifier (I1) fed into the telephone lines via transformer T1. The circuit is initiated by the action of a voltage transient pulse occurring across the phone line at the instant of a telephone circuit connection being made. This transient is a voltage change of about 48 to 5 VDC and usually occurs before the ring signal. It is this change that immediately forward-biases diode D1 and triggers pin 2 of timer I2, whose output pin 3 goes positive, turning on transistors Q1 and Q2. Timer I2 now remains in this state for a period depending on the values of timing resistor R12 and the selected value of the capacity determined by the activated time of switch S2. This four-position switch allows time selections of 10, 30, 100, and 1,000 seconds by the selection of the respective capacitors C12 through C15 (see Figure 23-2).

You will note that when Q2 is turned on by timer I2 that a simulated off-hook condition exists by the switching action connecting the 500-ohm winding of the transformer directly across the phone lines. Simultaneously, Q1 clamps the ground of the I1 amplifier to the negative return of battery Bl, enabling this amplifier section. System power is controlled by switch S1 being part of the pickup sensitivity control R1.

Electret microphone M1 picks up the sounds fed to operational amplifier I1 with the output coupled by capacitor C6 to the 1,500-ohm winding of Tl, and R1 controls the pickup sensitivity of the system. The system described will operate when any incoming call is received without the phone ever ringing.

Construction Steps

If you are a beginner, it is suggested to obtain our *GCAT1 General Construction Practices and Techniques*. This informative literature explains basic practices that are necessary in proper construction of electromechanical kits and is listed in Table 23-1.

1. Lay out and identify all the parts and pieces. Verify them with the parts list, and separate the resistors as they have a color code to determine their value. Colors are noted on the parts list.
2. Cut a piece of .1-inch grid perforated board to a size of 4.2 × 1.5. Locate and drill the holes as shown in Figure 23-3.

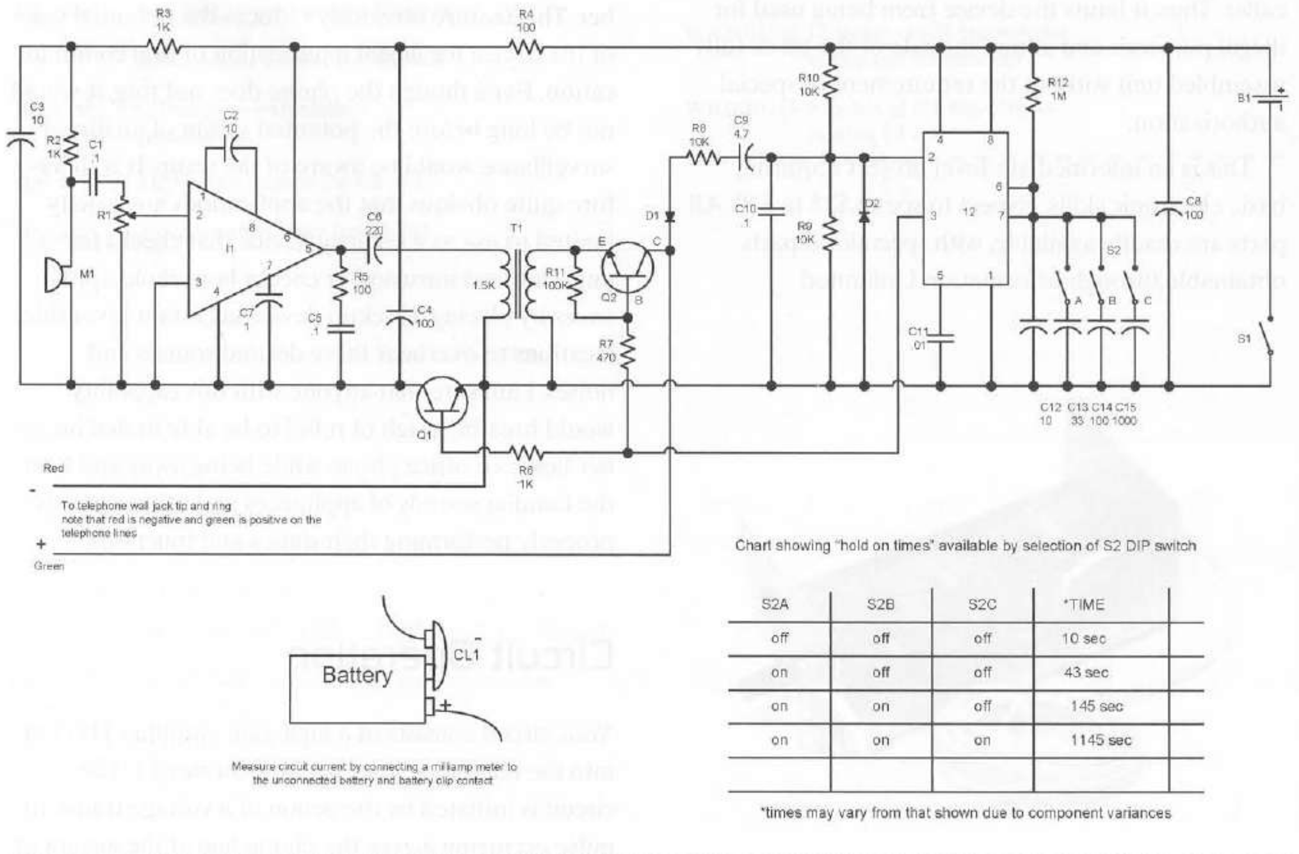

R3 1K
R4 100
C3 10
R2 1K
C1 .1
C2 10
R1
M1
I1
C6 220
T1
R5 100
1.5K
R11 100K
C7 .1
C5 .1
C4 100
E
C
B
Q2
R7 470
D1
Q1
R6 1K
R8 10K
C9 4.7
R10 10K
R12 1M
B1
R9 10K
C10 .1
D2
I2
C8 100
S2
A
B
C
C11 .01
S1
C12 C13 C14 C15
10 33 100 1000
Red
-
To telephone wall jack tip and ring note that red is negative and green is positive on the telephone lines
+
Green

Chart showing "hold on times" available by selection of S2 DIP switch

S2A	S2B	S2C	*TIME
off	off	off	10 sec
on	off	off	43 sec
on	on	off	145 sec
on	on	on	1145 sec

*times may vary from that shown due to component variances

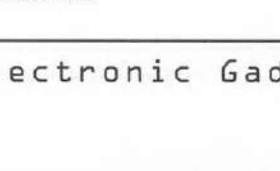

Battery
CL1
-
+

Measure circuit current by connecting a milliamp meter to the unconnected battery and battery clip contact

Figure 23-2 *Circuit schematic*

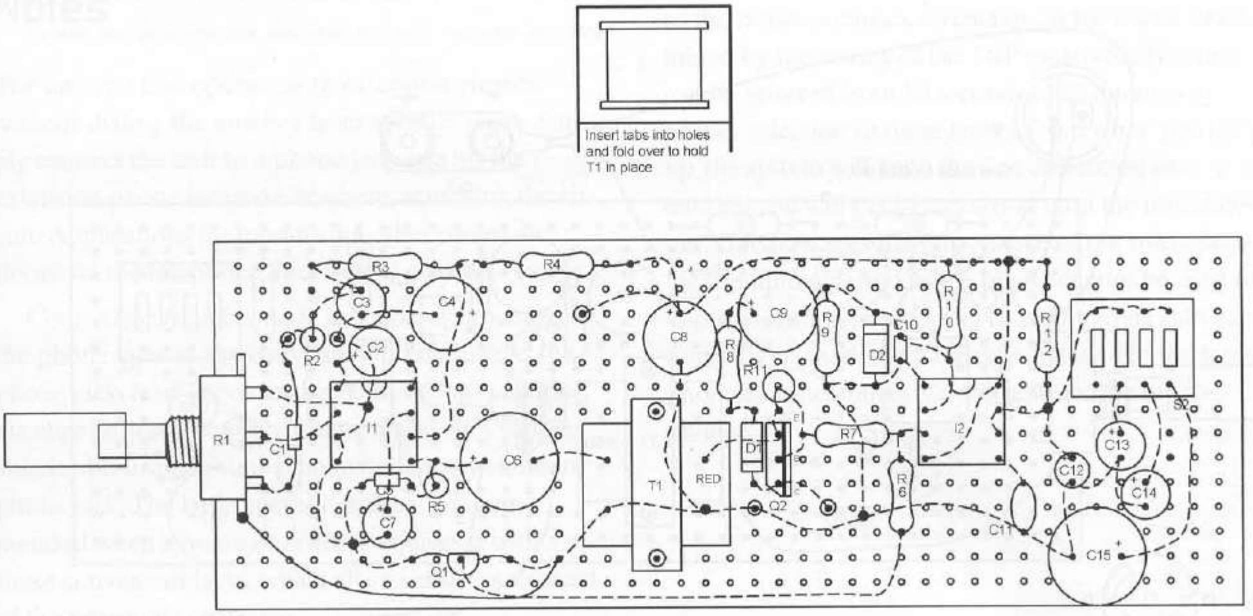

Dashed lines are connections on underside of perforated circuit board using the component leads. These points can also be used to determine foil runs for those who wish to fabricate a printed circuit board.

- Small dots are holes used for component insertion.

● Large dots are solder junctions

◉ Circles with dots are holes that may require drilling for component mounting and points for external connecting wires indicating strain relief points. It is suggested to drill clearance holes for the leads and solder to points beneath board.

Figure 23-3 *Assembly board layout*

3. If you are building from a perforated board, it is suggested that you insert the components starting in the lower left-hand corner (see Figure 23-3). Pay attention to the polarity of the capacitors with polarity signs and all the semiconductors.

   Route the leads of the components as shown and solder as you go, cutting away unused wires. Attempt to use certain leads as the wire runs or use pieces of the #24 bus wire. Follow the dashed lines on the assembly drawing as these indicate connection runs on the underside of the assembly board.

4. Attach the external leads as shown in Figure 23-4. You may relieve any strain they might have by drilling holes to allow the insertion of the entire wire. Note that the short pieces of bus wire are soldered to R1 for connection to the circuit board.

5. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.

Final Assembly

Create a suitable housing from a piece of aluminum or purchase a ready-made aluminum box approximately 5 × 3 × 2 inches. Drill the holes as required for the microphone grommet GR1, gain control R1, and strain relief bushing for the phone line cable assembly RJ11. The metal box should be grounded to the circuit common line to reduce hum pickup. Note that a piece of Velcro sticky tape can be used to secure the battery or use your own ingenuity.

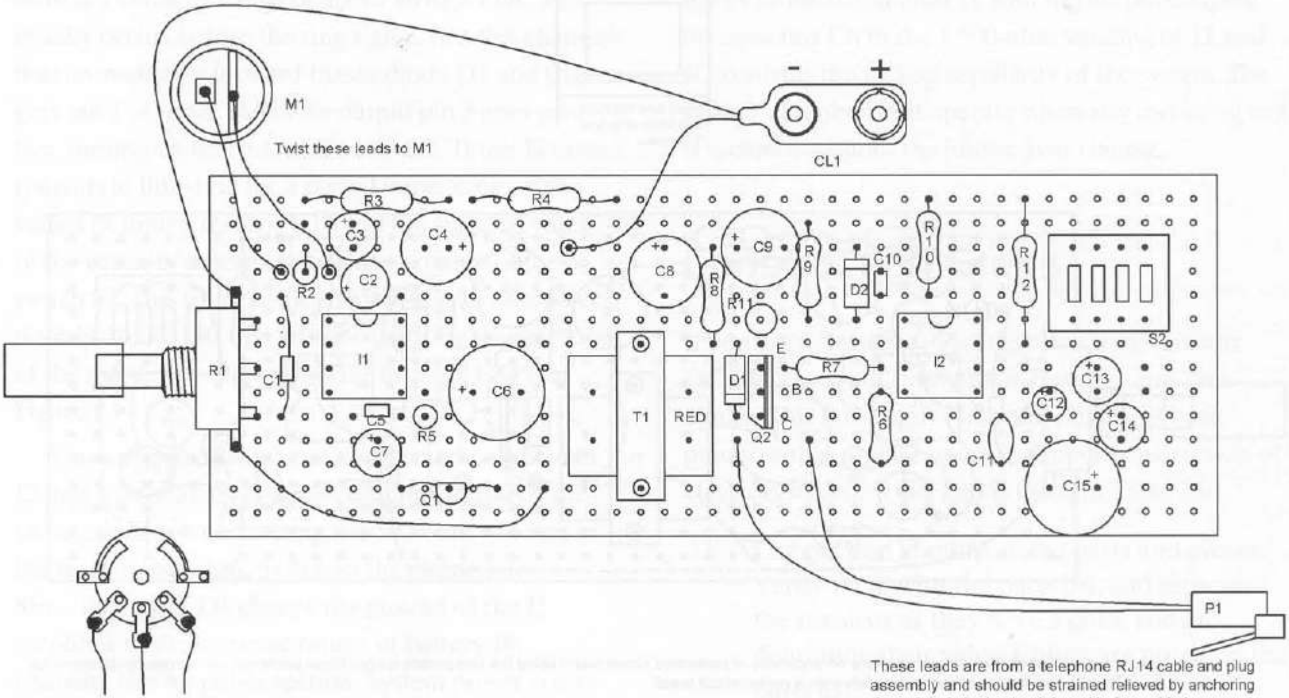

Twist these leads to M1

These leads are from a telephone RJ14 cable and plug assembly and should be strained relieved by anchoring through hole in thecircuit board before soldering

Solder short pieces of #22 buss wire for extending connecting leads

Figure 23-4 *External connection and parts identification*

Testing Steps

This step will require basic electronic laboratory equipment including a 60 MHz oscilloscope if trouble shooting the circuit is necessary.

1. Turn the unit off by rotating R1 full counter-clockwise, noting the click of the switch. Verify that all the DIP switch levers of S2 are down (off). Connect the battery as shown in series with a current meter and note zero current flow. Turn the unit on and note a current of 5 milliamps, which may momentarily jump to 30 milliamps as the unit times out.
2. Locate a phone jack so that you can monitor the DC voltage. Connect a meter across the tip and ring leads (green and red), and note the 40 to 50 volts with the phone on the hook. Also note the green lead is positive and the red is negative. This is the opposite of conventional wiring. Lift the phone and notice the 40 to 50 volts dropping down to 8 to 12 volts. This step verifies a properly working phone line.
3. Turn the unit off and connect it to the phone line. Rotate the R1 so that it just clicks on, and note the voltage remaining as or timing out to read the on-hook 40 to 50 volts. Preset R2 to midway.
4. Call the target phone using a cell or another telephone. The unit should connect at the beginning of the ring signal before the phone has a chance to ring. You should clearly hear sounds until the unit times out, at which time the device disconnects and restores the phone to normal operation. Adjust R1 for desired pickup sensitivity. The timeout time should be only about 10 seconds for the presetting of S2, as instructed in step 1. You may lengthen these times as shown in Figure 23-2.

Notes

The unit can also operate with extension phones without dialing the number from another place. Simply connect the unit to a phone jack and lift the extension phone located elsewhere, activating the circuit. Applications for monitoring other rooms and floors via the telephone lines now is possible.

Connections can be made in simple ways across the phone lines at any convenient point, such as the phone jack, lead-in box, or inside the telephone itself via clips or lugs. They can also be made via a simple, standard four-pin phone plug mated to any standard phone jack. The latter method is usually recommended when a home or office is equipped with these convenient jacks, which allow simple placement of the telephone or device.

When the phone activates, allowing you to listen to the target premises, it remains on for a time determined by the setting of the DIP switch S2. This time can be selected from 10 seconds to 20 minutes by proper selection of these levers. Even when you hang up, the system will keep the line connected so another call will get a busy signal until the unit times out. One now sees the importance of not overselecting the timing period as the phone cannot be used for either incoming or outgoing calls. Outgoing calls can normally be made with the unit installed to the target phone, but incoming calls will activate the microphone.

Table 23-1 Snooper Phone Room-Listening Device Parts List

Ref. #	*Description*	*DB Part #*
R1/S1	10K, 17-millimeter pot and switch	
R2, 3, 6	Four 1K, 1/4-watt resistors (br-blk-red)	
R4, 5	One 100-ohm, 1/4-watt resistor (br-blk-br)	
R7	470 1/4-watt resistor (yel-pur-br)	
R8, 9, 10	Three 10K, 1/4-watt resistors (br-blk-or)	
R11	100K, 1/4-watt resistor (br-blk-yel)	
R12	1M 1/4-watt resistor (br-blk-gr)	
C1, 5, 10	Three .1-microfarad, 50-volt caps	
C2, 3, 12	Three 10-microfarad, 25-volt vertical electro radial leads	
C4,8,14	Three 100-microfarad, 25-volt vertical electro radial leads	
C6	.22-microfarad, 250-volt metalized polypropylene	
C7	.1-microfarad, 50-volt cap	
C9	1-microfarad, 25-volt vertical electro cap	
C13	33-microfarad, 25-volt vertical electro radial leads	
C11	.01-microfarad, 50-volt disk (103)	
C15	1,000-microfarad, 25-volt vertical electrolytic capacitor	
C6	.22-microfarad, 250-volt metallized polypropylene	
C9	1-microfarad, 25-volt vertical electro cap	
C7	.1-microfarad, 50-volt cap (INFO#VG22)	
D1, 2	Two 1N4007 1-kilovolt diodes	
I1	LM386 operational amplifier *dual inline package* (DIP) *integrated circuit* (IC)	
I2	555 timer DIP IC	
Q1	PN2222 NPN small-signal transistor	
Q2	D40D5 high-voltage transistor	
TI	1.5K, 500-ohm audio interstage transformer	
M1	Electret microphone	
S2	4PST DIP switch	
PBOARD	4.2 × 1.5 × .1-inch grid perforated board, sized per chapter	
GR1	1/2-inch rubber grommet	
BUSCLP	1/4-inch Heyco clamp bushing	
CL1	9-volt battery clip with leads	
RJ11	RJ11 telephone plug and cable assembly	
WR20R	12 inches of #20 vinyl red wire for positive input	
WR20B	12 inches of #20 vinyl black wire for negative input	
ENC1	5 × 3 × 2-inch aluminum box and cover	

Chapter Twenty-Four

Long-Range FM Voice Transmitter

This project shows how to construct an electronic device that can transmit low-level sounds over a considerable distance (see Figure 24-1). It is intended for use as a security system when entry sounds are broadcasted over a standard FM radio, providing an early warning of an intrusion. It may be used to monitor children, swimming pools, sick and invalid people, and animals. The device can also be used for nature listening, which is made possible by the wide frequency response of the audio section, allowing good-quality reproducibility of rare bird calls, for example. An excellent application, when one is living in hostile areas, is to install the unit in one's home. This allows monitoring of the premises before possibly being surprised by a burglar.

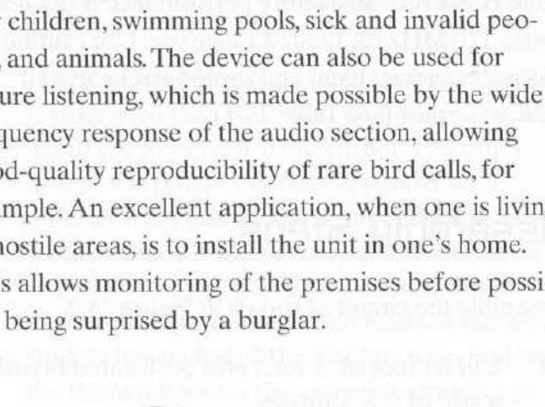

Figure 24-1 *Long-range FM voice transmitter*

The assembled transmitter is small enough to be placed at the foci of a sonic reflector, creating a highly directional wireless mike for targeting particular areas of interest. (The Audubon Society purchased 25 of these systems for monitoring bird and animal calls in mosquito-infested swamp areas.)

Circuit Description

This circuit shows how to build a super-sensitive, mini-powered FM transmitter consisting of a *radio frequency* (RF) oscillator section interfaced with a high-sensitivity, wide pass-band audio amplifier and capacitance mike with a built-in *field effect transistor* (FET) that modulates the base of the RF oscillator transistor (see Figure 24-2). Transistor Q1 forms a relatively stable RF oscillator whose frequency is determined by the value of coil L1 and tuning capacitor C4. The setting of C4 determines the desired operating frequency and is in the standard FM broadcast band with the tuned circuit design favoring the high end up to 110 MHz. Capacitor C2 supplies the necessary feedback voltage developed across resistor R3 in the emitter circuit of Q1, sustaining an oscillating condition. Resistors R1 and R2 provide the necessary bias of the base emitter junction for proper operation, while capacitor C1 bypasses any RF to ground fed to the base circuit. Capacitor C3 provides an RF return path for the tank circuit of L1 and C4 while

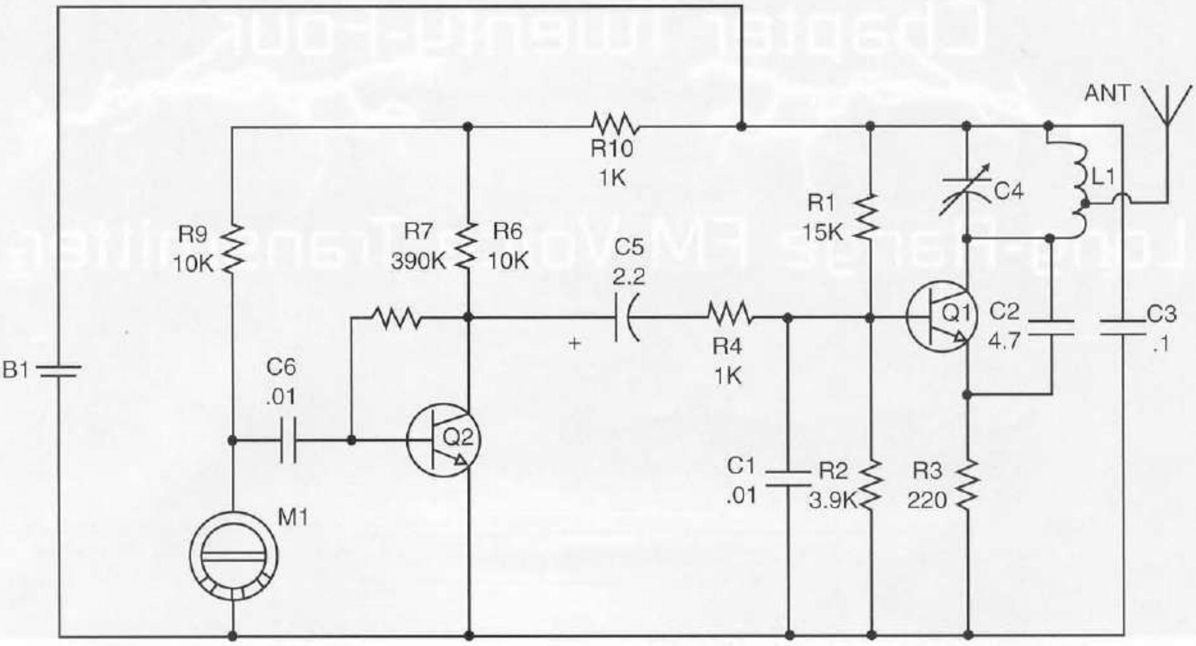

Note that the microphone M1 has polarity. You can identify the ground pad by the connection tabs to the outer shell.

Figure 24-2 *Circuit schematic*

blocking the DC supply voltage fed to the collector of Q1.

The audio section utilizes a high-sensitivity capacitance mike (M1) and built-in FET transistor (field effect) and will clearly pick up all low-level sounds in the speech audio spectrum. The speech voltage developed across resistor R9 by M1 is capacitively coupled by C6 to the base of audio amplifier transistor Q2. You will note the base of oscillator transistor Q1 is biased by resistors R1 and R2. A signal voltage now developed across resistor R6 is capacity coupled through nonpolarized capacitor C5 to the base of Q1 through resistor R4. The gain of Q2 is controlled by the ratio of R7 and R6. The operating point is set to allow full excursion of the collector to the amplified signal. The amplified speech signal now causes FM and AM modulation of the oscillator circuit by slight shifting of the operating point of the base section. Resistor R10 decouples the oscillator and audio circuits and is necessary to prevent feedback and other undesirable effects.

When properly assembled, the circuit should produce crystal-clear quality when the receiver is properly tuned to the unit. Note that a shunt capacitor may be connected across the base lead of Q1 to reduce sensitivity. The circuit with the components listed operates best in the upper FM band. This is a clear spot without interference from FM radio stations. However, satisfactory performance is obtained above 110 MHz for limited range use. Use caution as this is the aircraft band and should not be used if near an airport (see Table 24-1).

Assembly Steps

Assemble the circuit as shown in Figure 24-3.

1. Cut a piece of .1-inch grid perforated board to a size of 2 × 1 inches.

2. Form L1 by tightly wrapping eight turns of #16 bus wire on a #8 wood screw. This produces an eight-turn coil with an inner diameter of approximately .135 inches and a length of about .625 inches. Insert it in the proper hole and solder as shown.

3. Insert trimmer capacitor C4 into the holes as shown. You may put this component on either side of the PC board. This choice is up to you and should be determined by the final packaging scheme you require. Three holes are available for the part with two of them electrically the same point. This is necessary as some

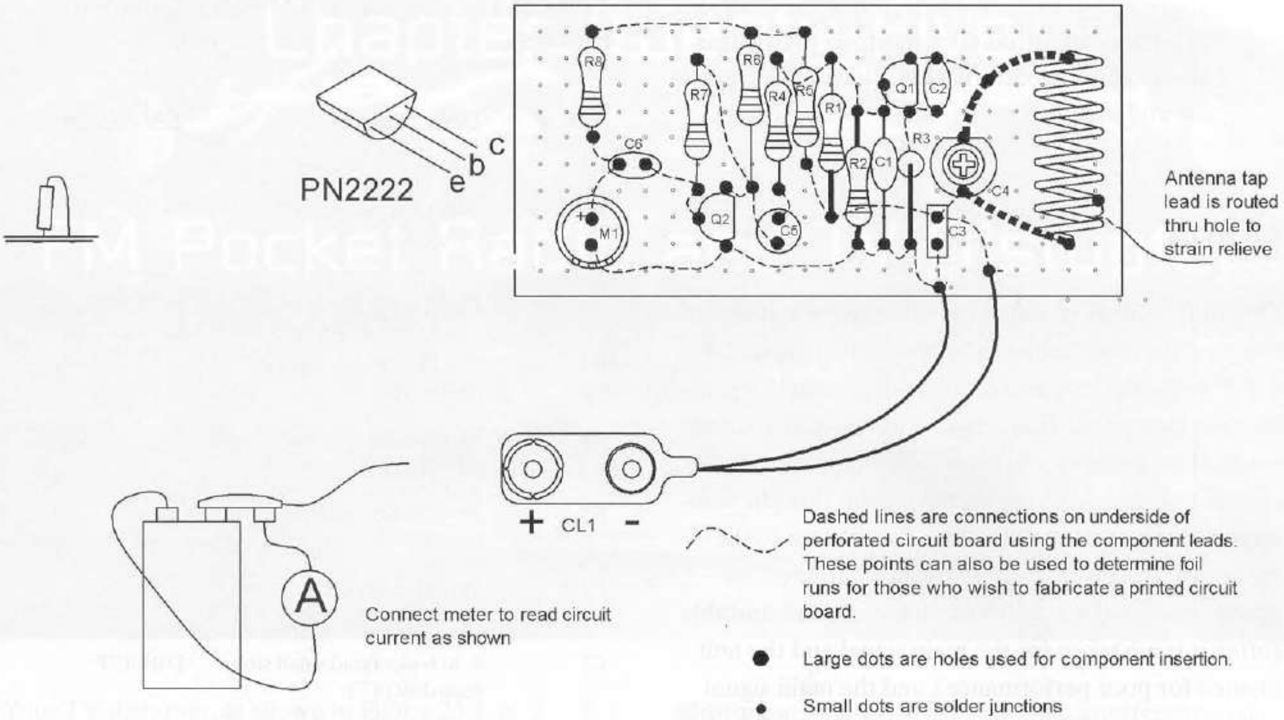
PN2222
e
b
c
R8
R7
R6
R4
R5
R1
Q1
C2
R3
C6
R2
C1
C4
M1
Q2
C5
C3
Antenna tap lead is routed thru hole to strain relieve
+ CL1 -
A
Connect meter to read circuit current as shown
Dashed lines are connections on underside of perforated circuit board using the component leads. These points can also be used to determine foil runs for those who wish to fabricate a printed circuit board.
Large dots are holes used for component insertion.
Small dots are solder junctions

Figure 24-3 *Assembly board*

trimmers may have three pins. Make sure that the common pins are connected to the same electrical points with the odd pin to the other point.

4. If you are building from a perforated board, it is suggested that you insert components starting in the lower left-hand corner. Pay attention to the polarity of the capacitors with polarity signs and all the semiconductors. Route the leads of the components as shown and solder as you go, cutting away unused wires. Attempt to use certain leads as the wire runs or use pieces of the #24 bus wire. Follow the dashed lines on the assembly drawing as these indicate the connection runs on the underside of the assembly board.
5. Attach the external leads of the battery clip (CL1).
6. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.
7. Tune an FM receiver to a fairly strong station at the high end of the band (108 MHz or higher). Turn up the volume and position the receiver 25 to 50 feet from the circuit.
8. If a multimeter or 50-milliamp meter is available, connect it in series with a battery lead. This can be done by removing one of the fasteners of the clip and connecting the meter to the free contacts, as shown in Figure 24-3. The meter should read 5 to 7 milliamps. Use a short piece of bare wire to touch the coil L1 a turn at a time, starting from the C3 end. Note that as you progress turn by turn away from C3 that the indicated meter current will drop or change.
9. If the device performs as follows with the battery connected, slowly rotate C4 with an insulated tuning wand until the station being received by the radio at approximately 108 MHz breaks out in audio feedback or is blocked out. It may be difficult at first to spot the signal as this adjustment is very touchy. Also note that several spots in the adjustment may be erroneous and will be weak and unstable. The correct signal will be strong, stable, and clear and can be verified by distance-of-transmission testing.
10. Once the desired setting of C4 is found, it should be marked with the frequency noted.

11. When operation is verified, it may be desired to enclose the circuit board into a plastic box along with the battery. This allows easier tuning and handling.

Notes

One of the things to watch for when using a device of this type is proper tuning. The adjustment capacitor, C4, is quite sensitive and requires only a slight movement to change the frequency, so always use a tuning wand. It is very easy when one is not familiar with this unit to tune to an erroneous signal. This phenomenon is likely to occur when the unit is close to the monitoring receiver. As previously mentioned, an erroneous signal will be weak, distorted, and unstable (often it is mistaken for the main signal and the unit blamed for poor performance), and the main signal will be strong, stable, and undistorted, if modulated. Several experiments in tuning the unit should be done before attempting to use it for the desired application.

Also, whenever possible, the unit should be used around 108 to 109 MHz, which is at the boundary of the aircraft band and upper FM broadcast band. When the approximate desired spot is found, final touchup tuning should be done at the receiver end for clarity. In most areas, these "upper" frequencies are clear and allow uninterrupted use in contrast to the lower ones in the FM band where a slight change in frequency from a clear spot results in the unit being drowned out by a strong broadcast station. Do not go above 108 MHz if near an airport or air traffic lane.

Most available FM radios can easily be detuned slightly to shift the dial readings down where 108 is 109. This is accomplished by carefully adjusting the "oscillator," the padding trimmer located on the main tuning capacitor, and "walking" a known station down the necessary megahertz or two. An antenna-peaking trimmer should now be adjusted for maximum signal at the high end.

Optimum performance will require a good-quality receiver with an analog slide-rule-type tuning dial. Digital tuned receivers will not work that well.

Table 24-1 Long-Range FM Voice Transmitter Part List

Ref. #	*Description*	*DB Part #*
R1	15K, 1/4-watt resistor (br-gr-or)	
R2	3.9K, 1/4-watt resistor (or-wh-red)	
R3	220-ohm, 1/4-watt resistor (red-red-br)	
R4, 5	Two 1K, 1/4-watt resistors (br-blk-red)	
R6, 9	Two 10K, 1/4-watt resistors (br-blk-or)	
R7	390K, 1/4-watt resistor (or-wh-yel)	
C1, 6	Two .01-microfarad, 50-volt disc capacitors	
C2	4- to 6-picofarad small silver mica disk (4.7)	DB# 4.7P
C3	.1-microfarad, 50-volt plastic capacitor	
C4	6- to 35-picofarad trimmer capacitor	DB# 635P
C5	2.2-microfarad, 25-volt vertical nonpolarized electrolytic capacitor	
CL1	Battery snap clip and leads	
Q1, 2	Two PN2222 NPN general-purpose transistors	
L1	Coil wound as described in assembly steps	DB# COIL8T
PBOARD	2 × 1 × .1-inch grid perforated board	
WR24BUSS	12 inches of #24 bus wire for wiring and antenna tap	
WR16BUSS	6 inches of #16 bus wire for making L1 coil	
M1	Special FET-bypassed microphone	DB# FETMIKE
PCXFM1		DB# PCXFM

Chapter Twenty-Five

FM Pocket Radio and TV Disrupter

Your TV disrupter, as shown in Figure 25-1, is designed and built on a piece of perforated circuit board that can fit inside a small plastic case. It operates using two standard, 9-volt transistor batteries in series. A unique pulsing circuit causes carrier deviation at 1,000 Hz, which creates a highly effective interfering device. The unit is tuned to a certain channel or frequency via a miniature tuning capacitor and then is controlled by a slight action of hand movement for further control.

This device is intended as a barroom joke and must not be abused. No attempts should be made to add an antenna because the range can be excessive and become a highly illegal device. When used as a harmless gag, such as in a bar, you can drive the bartender crazy because just as he is about to make an adjustment to a TV to get rid of the interference, you simply detune the device to make an adjustment to get rid of the interference and take the disrupter off the channel. When he thinks the problem is solved and continues his other duties, you hit it again. You can only begin to imagine the enjoyment (as a gag) of creating such havoc and being completely unnoticed.

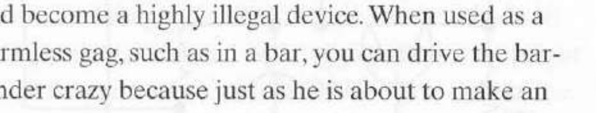

Figure 25-1 *Assembly board*

Also, any FM radio station can be interrupted with a 1,000 Hz tone that is many times louder than the normal received station. Obviously, a fun-loving person can use his or her imagination and come up with many harmless uses for the device.

It must be stated that using a device of this nature could be considered illegal, and discretion should be used at all times because a good gag can become a hateful nuisance. Also note that when cable TV is used, it may be necessary to get as close to the set as possible for maximum effect.

The following plans show how a miniature, simple, inexpensive electronic circuit can disrupt TV and radio communications in a very limited area. The unit can also be used to blank out those tiresome commercials.

Circuit Theory

The I1 timer generates positive pulses that occur across R4. The frequency of those pulses is a function of R1 and C1 and consequently establishes the pulse repetition rate of the device. This frequency modulates the base of Q1, which is a variable oscillator whose frequency is determined by C6 (see Figure 25-2). See Table 25-1 for the parts list.

Assembly Steps

Lay out and identify the parts with the bill of materials. Insert the components as shown in Figure 25-3, starting from one end of the perforated circuit board, and follow the locations using the holes as guides. Use the leads of the actual components as the connection runs, which are indicated by the dashed lines. It is a good idea to trial-fit the parts before actual soldering. Always avoid bare wire bridges and globby solder joints, and check for cold or loose solder joints. Always pay attention to the capacitors with polarity signs and all the semiconductors.

To adjust for frequency, set the TV to channel 2 and turn C6 for maximum capacitance (for a preliminary frequency, when adjusting, it is advisable to use a lower voltage due to the possibility of interfering with other than TV or radio stations such as police or aircraft bands). Adjust L1 by compressing turns until the interference is noted on channel 2. Turn to other channels, turn on C4, and note the complete coverage. Channels above 6 are covered by the harmonic output of the device.

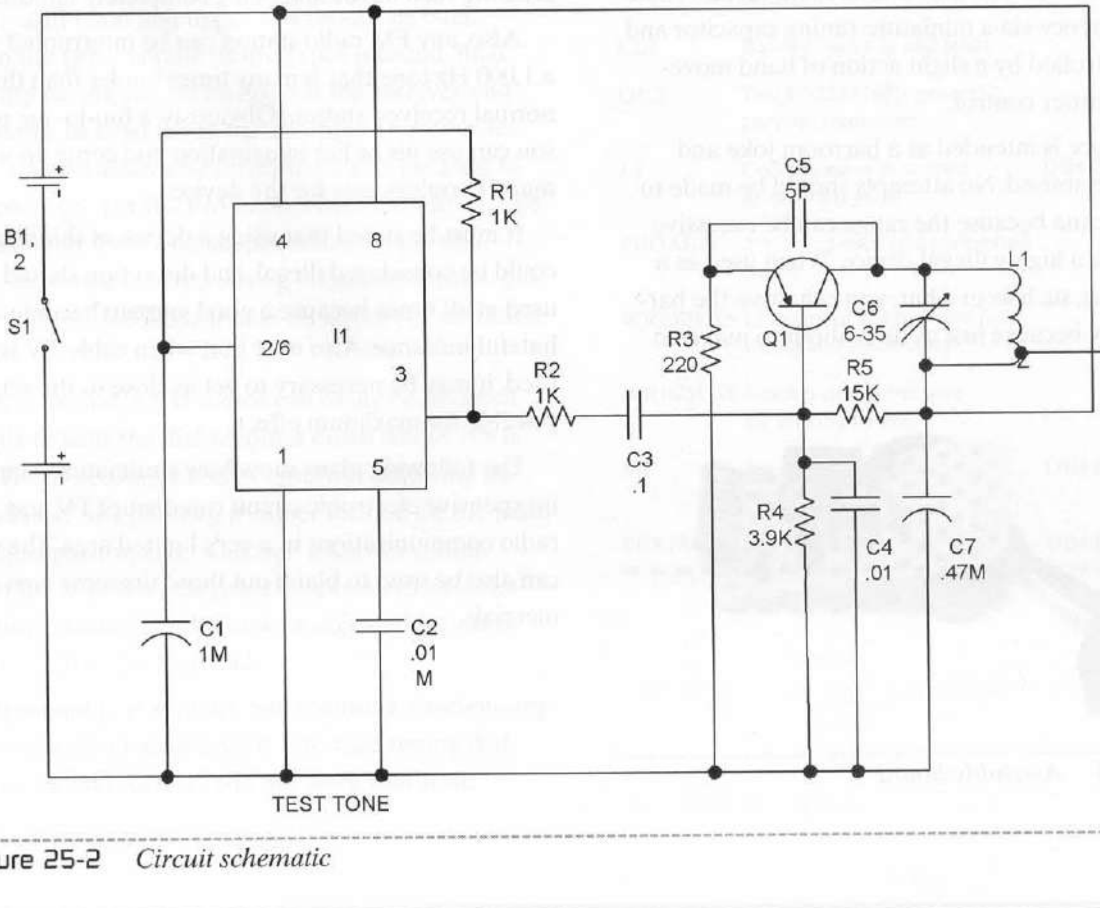

Figure 25-2 *Circuit schematic*

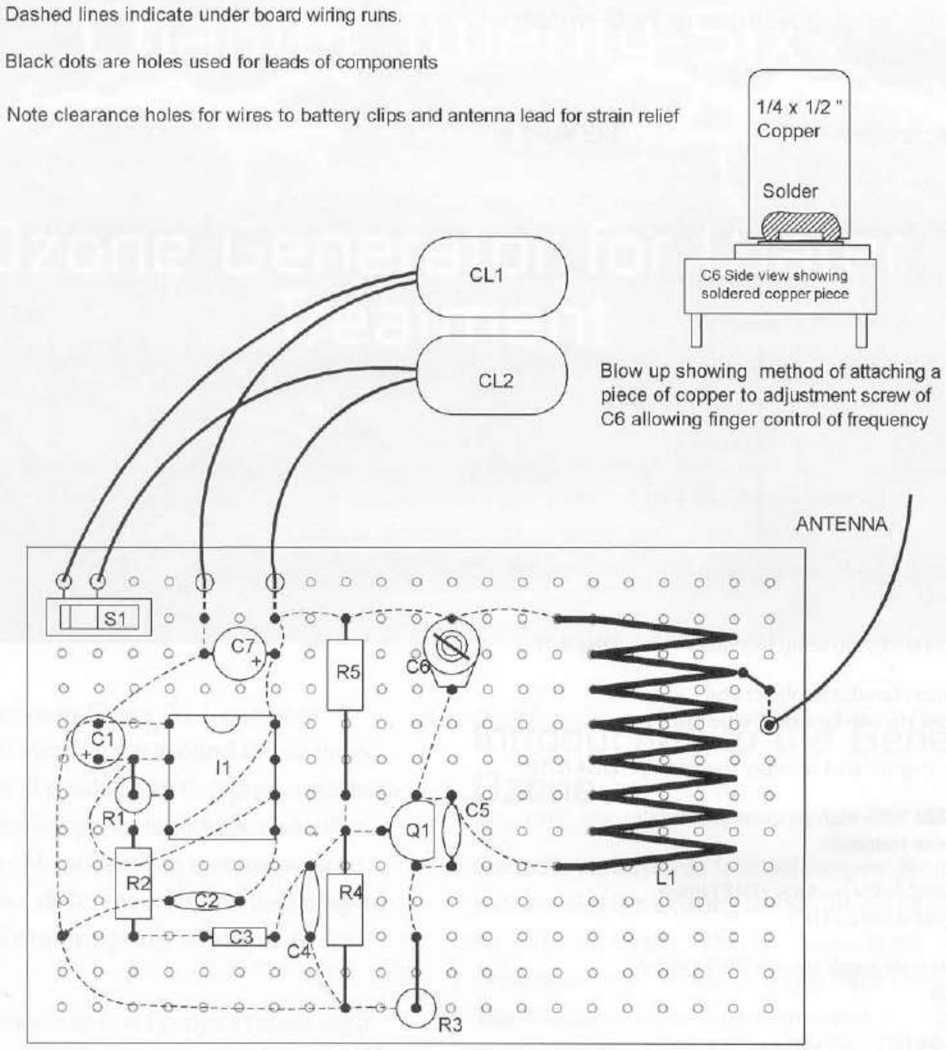
Dashed lines indicate under board wiring runs.
Black dots are holes used for leads of components
Note clearance holes for wires to battery clips and antenna lead for strain relief
1/4 x 1/2 "
Copper
Solder
C6 Side view showing soldered copper piece
Blow up showing method of attaching a piece of copper to adjustment screw of C6 allowing finger control of frequency
CL1
CL2
ANTENNA
S1
C7
+
R5
C6
C1
+
I1
R1
C5
Q1
R2
C2
R4
C3
C4
L1
R3

Figure 25-3 *Assembly board*

Several points of interest should be mentioned here, one being the ability to tune the unit on or off of a particular channel with just hand capacity (slight movement near L1); that is, the ability to detune the circuit with your hand creates a virtual capacitance. The other is the ability to also cover the FM radio band. This is accomplished by decompressing the coil (spreading the turns apart) and eventually adjusting L1 and C4 for the necessary coverage. It should be noted that all channels can be covered along with the FM band when properly adjusted. It may be necessary to remove a turn or two from L1 to a range on the FM radio stations.

Cable and satellite TV will often require the unit to be very close to cause an effect. There is no control over this, as the unit is designed to interfere with the actual frequency of the station being received, as was the case before cable and satellite services.

Table 25-1 FM Pocket Radio and TV Disrupter Parts List

Ref. #	*Description*	*DB Part #*
R1, 2	Two 1K, 1/4-watt resistors (br-blk-red)	
R3	220-ohm, 1/4-watt resistor (red-red-br)	
R3	100-ohm, 1/4-watt resistor (br-blk-br)	
R4	3.9K, 1/4-watt resistor (or-wh-red)	
R5	15K, 1/4-watt resistor (br-gr-or)	
C1	1-microfarad, 35-volt vertical electrolytic cap (green or blue can)	
C2, 4	Two .01-microfarad, 25-volt disk caps (103)	
C3	.1-meter, 50-volt ceramic cap (blue tablet .1)	
C5	6-picofarad, zero temp (4.7 disk)	DB# 4.7P
C7	.47-microfarad, 35-volt vertical electrolytic cap (green or blue can)	
C6	6- to 35-picofarad trimmer tuning Cap	DB# 635P
Q1	PN2222 NPN high-frequency, general-purpose transistor	
I1	555 *dual inline package* (DIP) timer *integrated circuit* (IC)	
S2	*Single pole, single throw* (SPST) slider switch	
CL1	Two-battery snap clip	
WR1	9 inches of #16 buss wire; wind seven turns on a pencil-sized object	
WR2	12 inches of #24 vinyl hookup wire	
PB1	2.5 × 1.5 × .1-inch grid perforated circuit board	
L1	Coil as shown	DB# COIL8T

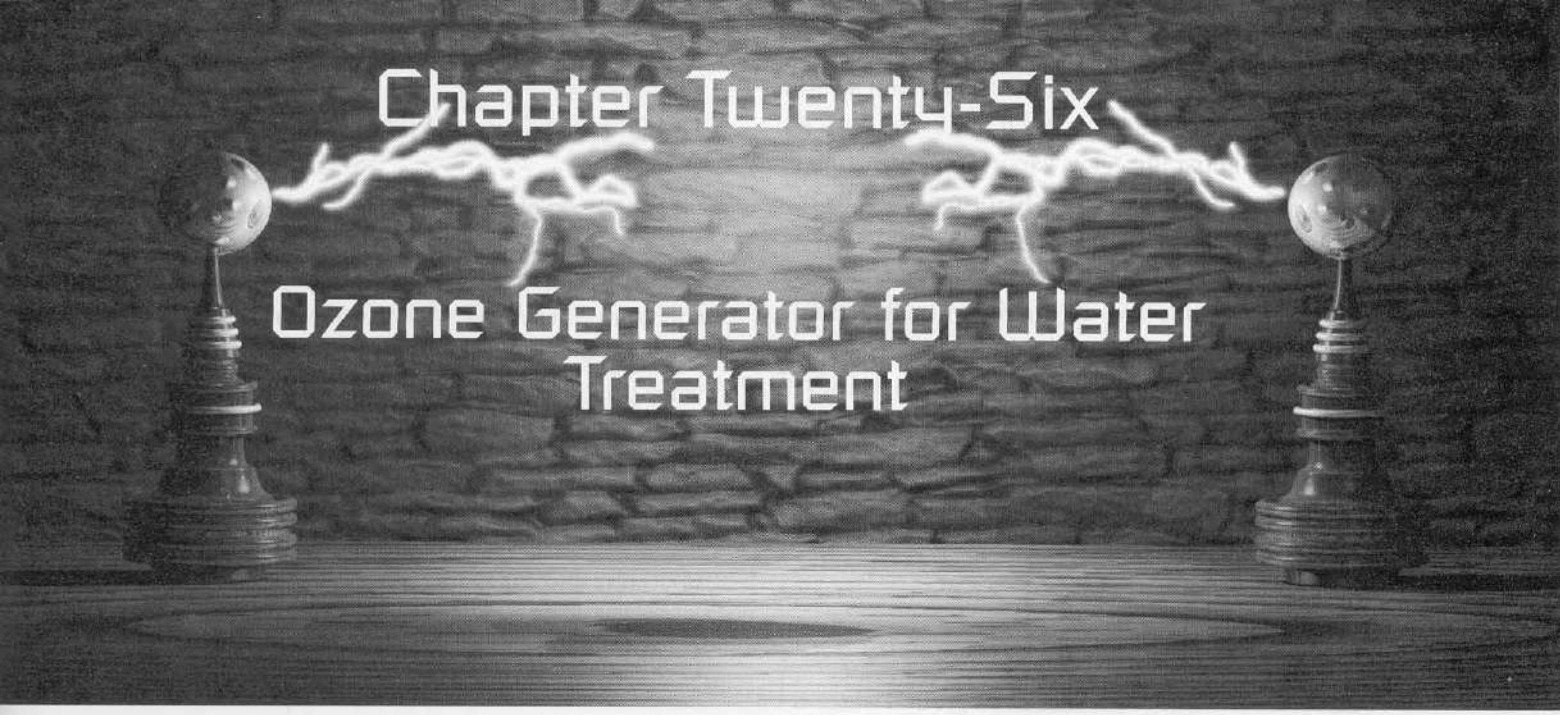

Chapter Twenty-Six

Ozone Generator for Water Treatment

This project, as shown in Figure 26-1, can very effectively purify water for use around the home or farm. The system will produce up to 5 grams per hour sufficient for swimming pools, laundries, and other medium-volume applications. The system covered here is shown in five different sizes and uses easy-to-construct modules requiring only a flow of air for operation.

This is an intermediate-level project requiring a basic electrical hookup. Expect to spend $50 to $250 depending on the size of the system chosen and the availability of the required air source. All parts are readily available, with specialized parts obtainable through Information Unlimited at www.amazing1.com, and they are listed in the parts list at the end of the chapter.

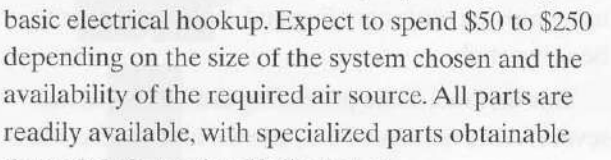

Figure 26-1 *The water treatment system*

Introduction to the Benefits of Ozone

Ozone is an unstable form of oxygen. Normal oxygen is diatomic (0_2), existing as two atoms of oxygen that form the molecule. Whereas ozone is tri-atomic (0_3), existing as three atoms of oxygen for the molecule. The tri-atomic form of oxygen is very unstable, wanting to lose the third oxygen atom and combine with whatever atom it can (oxidization). This property makes it the most active oxidizer known, with the exception of the very hazardous fluorine gas.

Ozone at normal pressures is colorless and produces a pleasing, fresh-air odor often like that after an old-fashioned thunderstorm. Under pressure, it becomes a bluish gas.

Ozone is also a very powerful bactericide. It is not affected by pH, as is chlorine, thus making it an excellent candidate for pools, spas, laundries, and general water treatment applications. It is many times more soluble in water, further enhancing its purifying effect. Ozone will combine with diatomic nitrogen N^2, forming nitrous oxide, 2NO. This oxide quickly combines with water, forming nitric acid, HNO_3. This is often a very undesirable effect when used with straight air. Pure oxygen greatly minimizes this effect and is often required in many applications. However, those requiring a supply of concentrated nitric acid for nitration may wish to consider ozone and air with a condensing apparatus to obtain this useful acid. A condensing apparatus takes vapor from the ozone gas along with water and nitrogen to form a liquid by contacting them with cooler surfaces. The liquid (now nitric acid) flows down into a collection container by gravity. In the same way the old timers made moonshine, this uses a simple and basic chemical reaction where $2O^3 + 2N^2 + 2H_2O$ yields $2H(NO^3) + H_2O + 2NO$.

Ozone for Water Applications

It is estimated that 20 percent of all ground water is contaminated by pesticides, benzene, and phenol derivatives along with other undesirable organic substances. Ozone can oxidize many of these compounds, along with deactivating many viruses and harmful bacteria. Ozone will also oxidize certain inorganic compounds, such as iron and manganese, making them more easily removable by filtration.

Chlorine and bromine are often the choice of disinfectant for swimming pools and spas. The effect of these halogens is often dictated by pH, temperature, and agitation. Extreme heat and agitation can produce chloroform, a very toxic carcinogenic. The *Environmental Protection Agency* (EPA) is already taking a look at these chemicals for this use.

Ozone-treated water will destroy fungus, mold, and many pathogens found in water when used for washing fruits and vegetables in packing lines. When discharged, ozone causes little change to the beneficial bacteria in sewage treatment facilities.

Additionally, freshly caught fish will last longer when washed with ozone-treated water, and ozone's oxidizing action can eliminate odors from stored cheese. Egg storage time is increased, and wine can be aged faster. The removal of odors produced by the bleaching of beeswax, starch, flour, straw, bones, and feathers are all aided by ozone treatment. Also, the grease and wax on cotton and wool fibers can be decomposed by ozone, and gray mold on the surface of fruits and vegetables can be controlled by ozone.

Ways to Generate Ozone

Ozone can be produced by an electrical discharge or by a high-frequency electromagnetic wave. A high frequency requires the wave to be in the ultraviolet spectrum where Planck's energy formula, $W = hc/v$, starts to become effective. This is where energy in a wave packet in ergs is equal to Planck's constant times the speed of light in centimeters divided by the wavelength in centimeters. It is this energy that causes the stable O_2 to break up and recombine with other O_2 to form unstable O_3. Germicidal lamps operating at 253.7 nanometers can produce ozone.

The method presented here utilizes the ozone-producing properties of an electrical discharge. We have all, at one time, smelled the by-products of ozone. After a thunderstorm, it can be detected, as well as on certain days where electrical activity is spawning a storm. A sparking electric discharge such as brushes on a motor will create ozone also.

Our method uses a metal tube with a conductor running down the center. The conductor and the tube are insulated by one another and must support the high frequency and high voltage necessary to create a corona without breaking down into an arc. The ozone produced in the tubular cavity must be made to flow using moving air for cooling and replenishing air to be converted.

This method, while producing usable ozone, has several disadvantages. First, normal air contains nitrogen that likes to combine with the ozone to produce nitrous oxides. Second, air contains water in the form of moisture. Without getting into basic chemistry, we know that water plus nitrous oxide forms nitric acid, which is corrosive and undesirable for air purification applications. However, a very effective way of making this acid is to allow the oxides to combine with water vapor or steam and condense in a cooling tube, producing concentrated nitric acid that can be used in the manufacture of high explosives. The 5 gram per hour system can produce enough nitric acid that, when mixed with battery acid (sulfuric acid), can produce usable amounts of high explosives by the simple nitration of many organic compounds.

Select the system you require from Figure 26-2. Note that each is complete with the mating power supply.

Wire up the system as directed in Figure 26-3. Obtain a suitable air supply and connect up the hoses as shown. Use ozone-resistant material for connections to the cell.

Turn on the air supply and note a bubbling from the output hose immersed under the water. Apply power to the high voltage and note a distinct smell coming from the hose when pulled out of the water. Reinsert and rig the hose end to stay under the water. The unit is now ozonating the water.

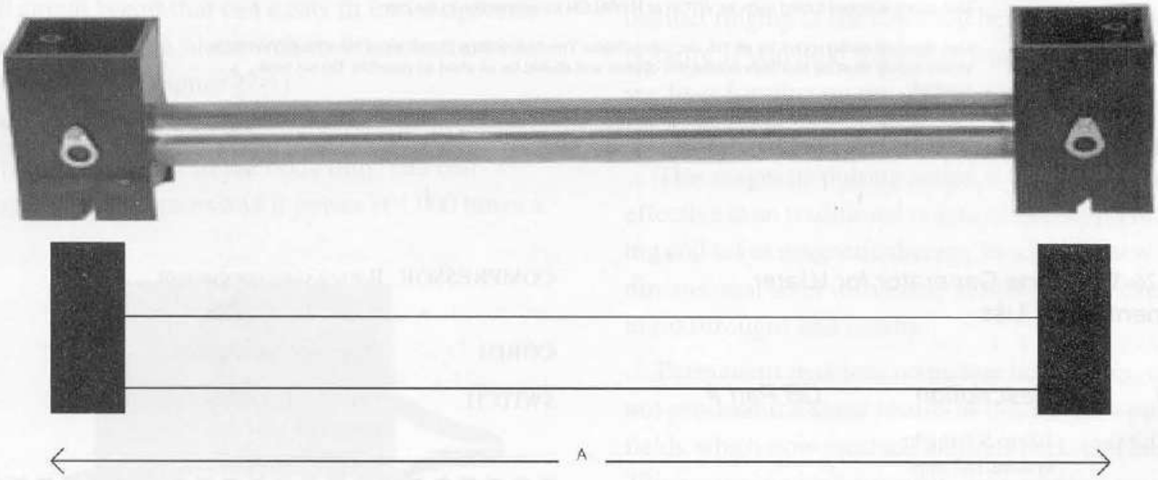

DATA BASE#	LENGTH A	OZONE OUTPUT/HR	AIR/O3 FLOW/MIN	INPUT V	INPUT I
OZONE300	182mm	100-500 mg/hr	.25-.55 CFM*	115 Vac	.2 Amps
OZONE800	282mm	500-1000 mg/hr	.35-.7 CFM*	115 Vac	.6 Amps
OZONE2000	382mm	1000-2500 mg/hr	1-1.5 CFM*	115 Vac	.9 Amps
OZONE5000	482mm	2500-5000 mg/hr	1.5-3 CFM*	115 Vac	I.2 Amps

* Multiply CFM by 28.3 to get Liters per minute

Each DATA BASE# system is complete with a matching power supply and ready to connect up as shown on figure 26-3

Figure 26-2 *System selection specifications*

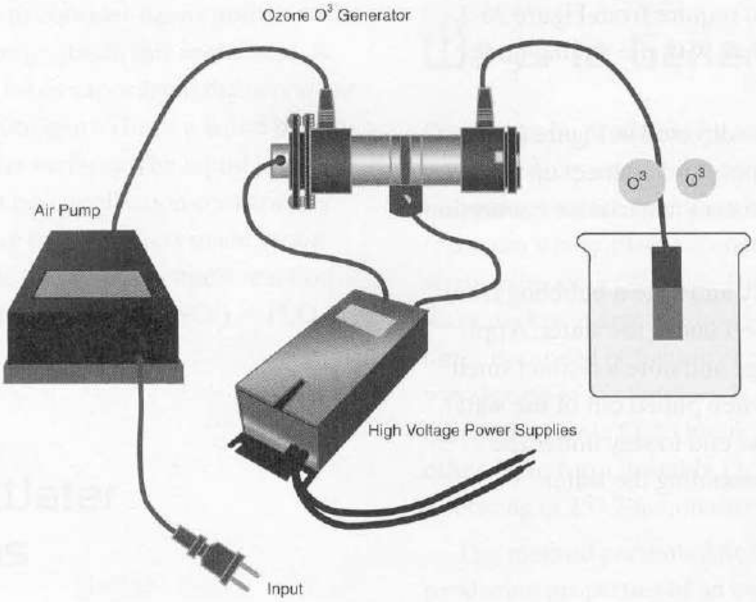

The high voltage power supplies used in these systems operate from normal 115 Vac 60 Hz. They are listed with the appropriate ozone cell on figure 26-2

The air pumps used can be that appropriate for the volume of the particular syste. A flow meter should be used when first setting up the system to verify proper air flow. Pure oxygen works the best and is available in cylinders with regulator valves and optional flow meters. Large aquarium air pumps may be used with the smaller systems while rotary vane air compressors are suggested for the larger systems. Note to multiply cubic feet per minute (CFM) by the factor 28.3 when converting to liters per minute.

Use ozone resistant tubing such as VITOR or HYPALON for connection to the cell

Use standard wiring codes for all 115 vac connections. The high voltage output wires from the high voltage power supply must be free from conductive objects and should be as short as possible. Do not twist together.

Figure 26-3 *Ozone system diagram*

Table 26-1 Ozone Generator for Water Treatment Parts List

Ref. #	*Description*	*DB Part #*
OZONE300	100 to 500 mg/hr system/182 mm/ 115 VAC @ .2 amps	DB# OZONE500
OZONE800	500 to 1,000 mg/hr system/282 mm/ 115 VAC @ .6 amps	DB# OZONE800
OZONE2000	1,000 to 2,000 mg/hr system/382 mm/ 115 VAC @ .9 amps	DB# OZONE2000
OZONE5000	2,500 to 5,000 mg/ hr system/482 mm/ 115 VAC @ .2 amps	DB# OZONE5000
HOSE	Connecting hoses, use ozone-resistant VITOR or HYPALON	
COMPRESSOR	Rotary vane compressor at required airflow	
CORD1	Three-wire power cord	
SWITCH	Two switches to control air compressor and high-voltage supply	

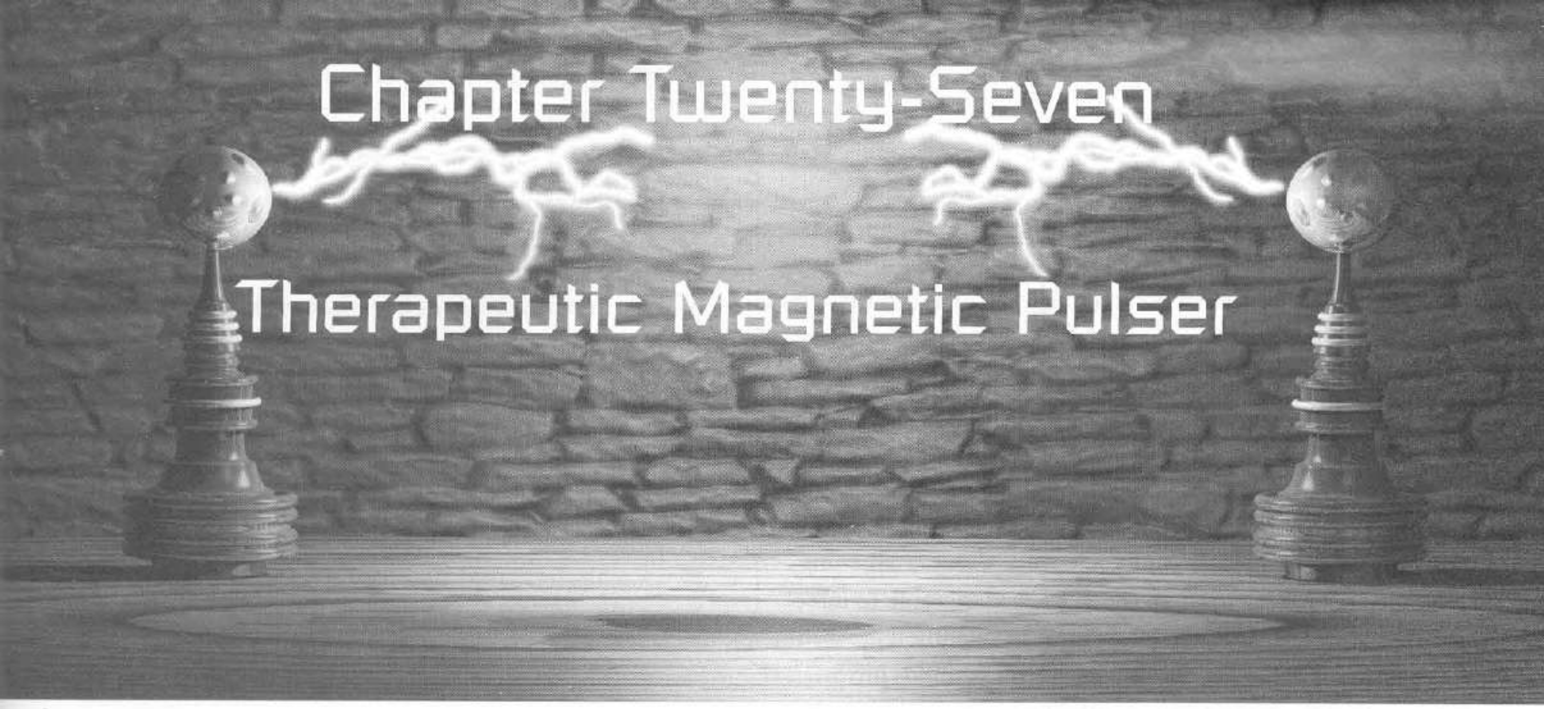

Chapter Twenty-Seven

Therapeutic Magnetic Pulser

This project shows how to build a battery-powered magnetic pulser that can provide relief from aches and pains, as well as improve upon many different health concerns and conditions. The unit is built on a small circuit board that can easily fit into a cigarette pack-size enclosure, along with its 9-volt battery or wall adapter (see Figure 27-1).

This magnetic pulser is designed so the magnetic north pole is applied to the body only. The coil strength is 10,000 gauss and it pulses at 1,000 times a second. A small switch can be used to turn the device off or you may simply unsnap the battery clip.

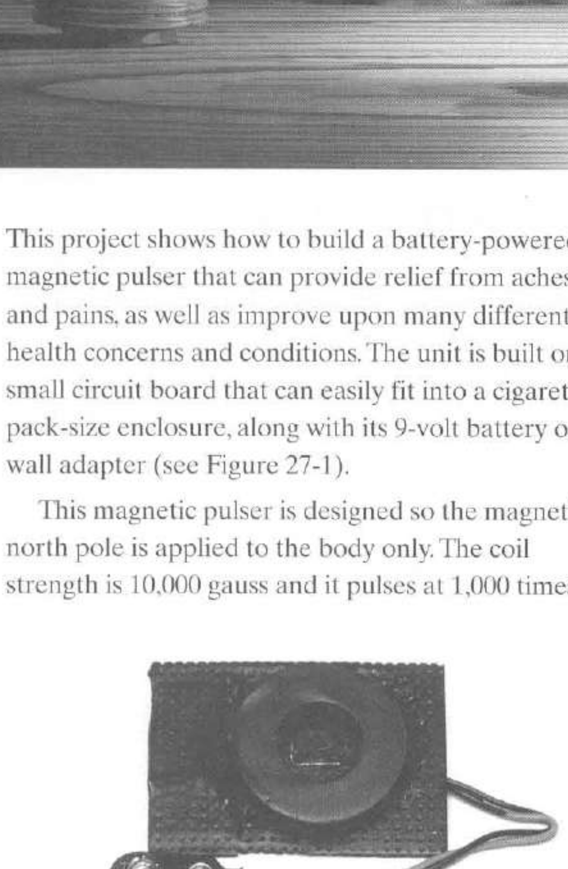

Figure 27-1 *The thermomagnetic pulser*

When an aluminum beer or pop can is held between your ear and this pulser, you hear a strong, distinct ringing as the can itself becomes the opposing polarity. If you hold the coil in front of a TV, you can see lines forming on the screen. On some TVs, the lines are visible when it's held as far away as 3 inches.

This magnetic pulsing action is more powerful and effective than traditional magnetic therapy. The pulsing coil takes magnetic therapy to a whole new dimensional level with many yet-to-explore exciting breakthroughs and results.

Permanent magnets, no matter how strong, will not produce the same results as this device's pulsed fields, which now produce induced back-emf currents. This requires a high-intensity, time-varying, magnetic impulse, not just a steady-state magnetic field.

Permanent magnets have their uses, but they work in different ways with different results. Equally important is the fact that we have "electric-powered" brains: All our thoughts and perceptions consist of complex networks of electrical signals and electromagnetic fields that pulse and sweep throughout the brain. So it then makes sense that harmonic electrical revitalization of the brain can harmonically influence your mental state and positively alter mental effectiveness.

Externally applied magnetic pulses to the lymphatic system, spleen, kidney, and liver help neutralize germinating, latent, and incubating parasites of all types, helping to block reinfection. This speeds up the elimination of disease, restores the immune system, and supports detoxification. The movement of the lymphatic system is essential in purifying, detoxifying, and regenerating the body, supporting the immune system, and maintaining health. Normally, the lymph is pumped by the movement of our body's electromagnetic field with vigorous exercise and physical activity. However, a clogged, sluggish, or weak lymphatic system prevents the body from circulating vital fluids and eliminating toxic wastes, thus weakening the immune system. It makes us vulnerable to infections and diseases. In order to be healthy, it is essential to keep the energy balanced and fluids moving so that the body's natural intelligence may operate at its full healing capacity. In addition, each cell must be enlivened with its own unique energetic frequencies and harmonic energy state and be harmonically connected to the life-force energy throughout the rest of the body.

Pulsed electromagnetic fields influence cell behavior by inducing electrical changes around and within the cell. Improved blood supply increases the oxygen pressure, activating and regenerating cells. Improved calcium transport increases the absorption of calcium in bones and improves the quality of cartilage in joints, decreasing pain dramatically. Acute and even chronic pain may disappear completely.

This device uses complex energy pulses of magnetic waves to stimulate certain body functions. It accelerates the production of vitally important hormones often providing miraculous effects. Magnetic pulsing aids in human growth hormone production and neotransmitter production. The results are remarkable, increasing vitality, sexuality, and accelerated healing. It also helps learning and the reduction of memory loss with reports of increased psychic ability for some people. The magnetic pulser also revitalizes and rejuvenates. As we get older, hormone production drops off considerably and is a reason why we age.

Users are experiencing faster healing of injuries, including bone fractures, carpal tunnel, and arthritis. The pulsed magnetic field stimulates blood flow and cellular respiration in the area applied. Reports have also documented that people's migraine headaches have ceased after magnetic therapy has been applied.

This pulser can be worn, held, or wrapped to any part of the body or held in place with a tension bandage. You can use your pulser while driving and on the go, rather than having to stop and apply it when sitting or lying down, as is necessary when using a 110-volt, standard magnetic pulser. The product is currently being sold only for experimental research and testing.

This is a simple, basic-level project requiring minimal electronic skills. Expect to spend $15 to $20. All parts are readily available, with specialized parts obtainable through Information Unlimited (www.amazing1.com), and they are listed in the parts list at the end of the chapter.

Circuit Description

Figure 27-2 shows a 555 timer (I1) connected as a free-running, astable pulse generator. The output pulses are nonsymmetrical with a ratio of low to high time of 10 to 1. Discharge resistor R3 and forward-conducting diode D2 control the low time, whereas charge resistor R2 and forward-conducting diode D1 control the high time. Timing capacitor C2 is common to both states of the pulse being alternately charged and discharged.

Magnetic energy is produced when the timer I1 turns on transistor Q1. The current now has the "on" time of the pulse to build up and then collapse in the flux coil L1 when Q1 is turned off. The stored energy is returned to the circuit by diode D3. The result is a steady train of magnetic pulses at 1,000 pulses per second.

The switch S1 allows the unit to be turned on and off and also allows switching in another range, dropping the pulse repetition to 100 per second. It is interesting to note that claims using specific pulse repetition rates of 20, 72, 95, 100, 125, 146, 440, 600, 660, and 727 have proven very beneficial. Do not attempt to extend the frequency without first determining the current rise in the coil. Current may be excessive and will require more turns as the frequency is lowered.

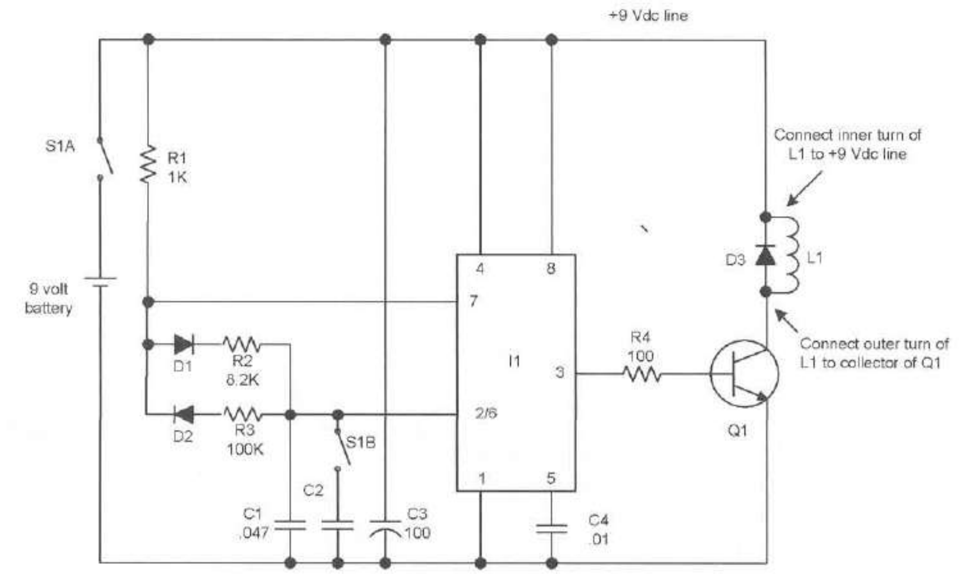

Select C2 for optional repetition rate

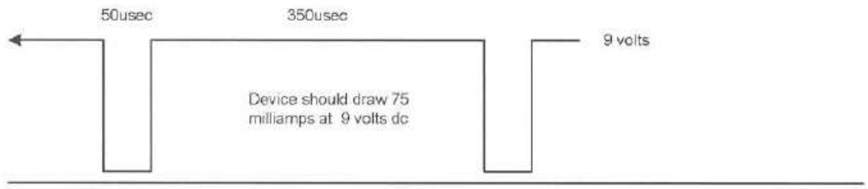

Waveform at collector of Q1

Figure 27-2 *Circuit schematic*

Assembly

1. Assemble the flux coil as shown in Figure 27-3 by first drilling a small hole adjacent to the inner diameter so that the "start" lead can exit the winding. Carefully wind 100 turns of #24 magnet wire in even windings. This is not necessary but makes for a neat-looking construction. Note to wind the wire so that its direction is the same as the figure with the leads routed as directed. This will guarantee the correct polarity as shown.
2. Cut a piece of .1-inch grid perforated board to a size of 2.6 × 1.8 inches, as shown in Figure 27-4. Locate and drill the holes as shown for the leads of battery clip CL1.
3. When building from a perforated board, it is suggested that you insert the components starting in the lower left-hand corner as shown. Pay attention to the polarity of the capacitors with polarity signs and all the semiconductors.

   Route the leads of the components as shown and solder as you go, cutting away unused wires. Attempt to use certain leads as the wire runs or use pieces of the #24 bus wire. Follow the dashed lines on the assembly drawing as these indicate the connection runs on the underside of the assembly board.
4. Attach the external leads from the battery clip and flux coil. Note the polarity of the leads on this part as indicated in Figures 27-2 and 27-4.
5. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.
6. You are now ready to test the unit. Connect it to a 9-volt battery and note a current draw of 60 to 80 milliamps. Place the coil of the unit

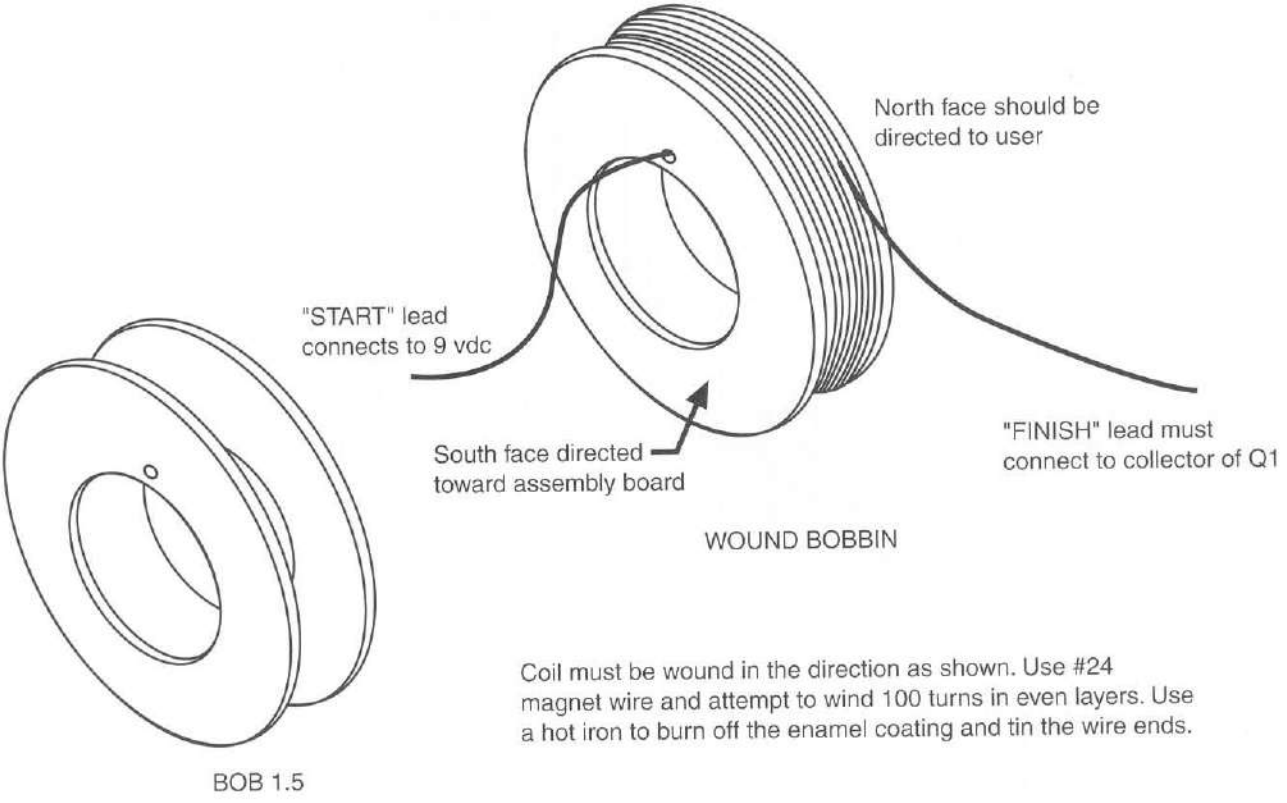
North face should be directed to user
"START" lead connects to 9 vdc
South face directed toward assembly board
"FINISH" lead must connect to collector of Q1
WOUND BOBBIN
BOB 1.5

Coil must be wound in the direction as shown. Use #24 magnet wire and attempt to wind 100 turns in even layers. Use a hot iron to burn off the enamel coating and tin the wire ends.

Figure 27-3 *Flux coil assembly*

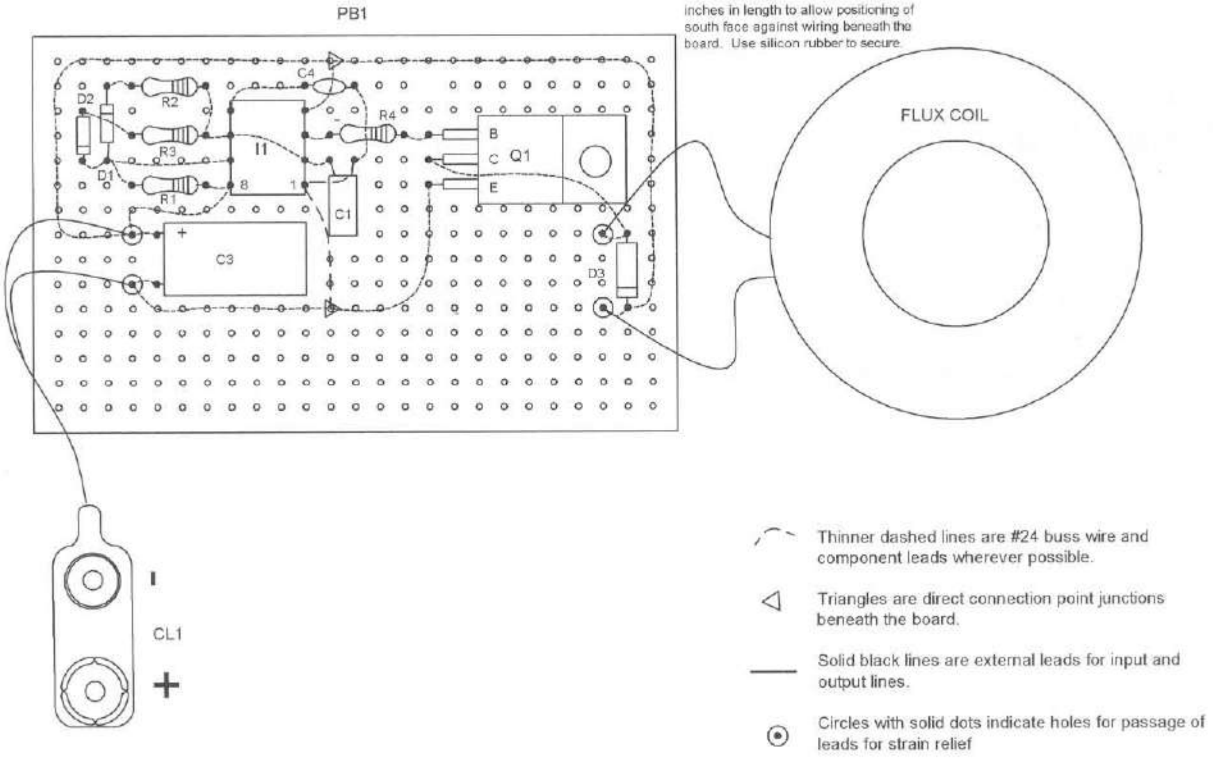
PB1
D2
R2
C4
R3
I1
R4
B
C
Q1
E
D1
R1
8
1
C1
+
C3
D3
FLUX COIL
CL1
I
+

Leads to flux coil L1 should be two inches in length to allow positioning of south face against wiring beneath the board. Use silicon rubber to secure.

Thinner dashed lines are #24 buss wire and component leads wherever possible.

◁ Triangles are direct connection point junctions beneath the board.

Solid black lines are external leads for input and output lines.

⊙ Circles with solid dots indicate holes for passage of leads for strain relief

Figure 27-4 *Board assembly*

near a television screen and note the lines appearing. You may also put a metal can onto the coil head and place your ear to the can, noting a 1 to 2 KHz tone. If you have a scope, note the waveshape shown in Figure 27-2.

7. To use the unit, simply place the north face of the coil to the target area and allow exposure for several hours.

Table 27-1 Therapeutic Magnetic Pulser Parts List

Ref. #	*Description*	*DB Part #*
R1	1K, ¼-watt resistor (br-blk-red)	
R2	8.2K, ¼-watt resistor (gray-red-red)	
R3	100K, ¼-watt resistor (br-blk-yel)	
R4	100-ohm, ¼-watt resistor (br-blk-br)	
C1	.047-microfarad, 50-volt plastic capacitor	
C2	1-microfarad, 25-volt vertical electrolytic capacitor	
C3	100-microfarad, 25-volt vertical electrolytic capacitor	
C4	.01-microfarad, 50-volt disk capacitor (103)	
D1, 2	Two IN914 silicon diodes	
D3	1-kilovolt, 1-amp diode 1N4007	
Q1	TIP31 NPN TO-220 power transistor	
I1	555 dual inline package (DIP) timer	
BOB1.5	1.5 × .3 × .75-inch ID nylon bobbin (see Figure 27-3)	DB# BOB1.5
L1	Flux coil (see Figure 27-3)	DB# FLUXCOIL
PBOARD	2.6 × 1.8 × .1-inch grid perforated board as shown in Figure 27-4	
CL1	9-volt battery snap clip	

Chapter Twenty-Eight

Noise Curtain Generator

This unique circuit shown in Figure 28-1 is designed to produce a relaxing sound like that of a breaking surf. It is technically known as pink noise, being defined as a random distributor of equal sounds in the audible spectrum, favoring the higher-frequency end.

Pink noise has a spectral intensity that is inversely proportional to frequency for a given range. Therefore equal power is dissipated into a fixed resistance in any octave bandwidth in that range. White noise is random noise, which is typically thermal and shot noise, has a constant energy per unit bandwidth, and is independent of the central frequency at the band.

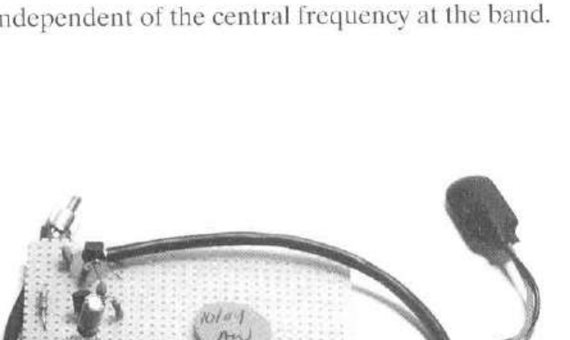

Figure 28-1 *The board-level generator*

This device is intended for use as a relaxing source of sound and also as a mask for certain types of surveillance equipment. Crashing surf is an excellent example of naturally produced pink noise that has a great relaxing effect. Due to the nearly equal amplitude of all the frequencies generated, the electronic equipment of an acoustical nature becomes easily overloaded and saturated, whereby normal voice sounds of varying amplitude and of limited bandwidth are much more finite and are unable to be processed by the equipment. In other words, the noise jams the microphone and audio preamplifier circuitry, thus rendering it unable to detect normal voices. (It should be noted that in certain instances a running shower will sometimes help simulate pink noise.)

The signal from the device is intended to be fed into the input of any amplifier, radio, or tape deck. It is easily connected to the center arm of the volume control if an input jack is not provided in a regular radio. (You will note the phono plug attached.) Also, the two batteries used are left on permanently, due to the fact that so little power is used.

The optional "speaker amplifier" reference schematic in Figure 28-2 is built on a piece of perfboard using standard audio-frequency wiring techniques. When used as a totally self-contained system, the electronics may be housed along with all batteries into a case or enclosure determined by the builder. This approach provides a convenient portable unit.

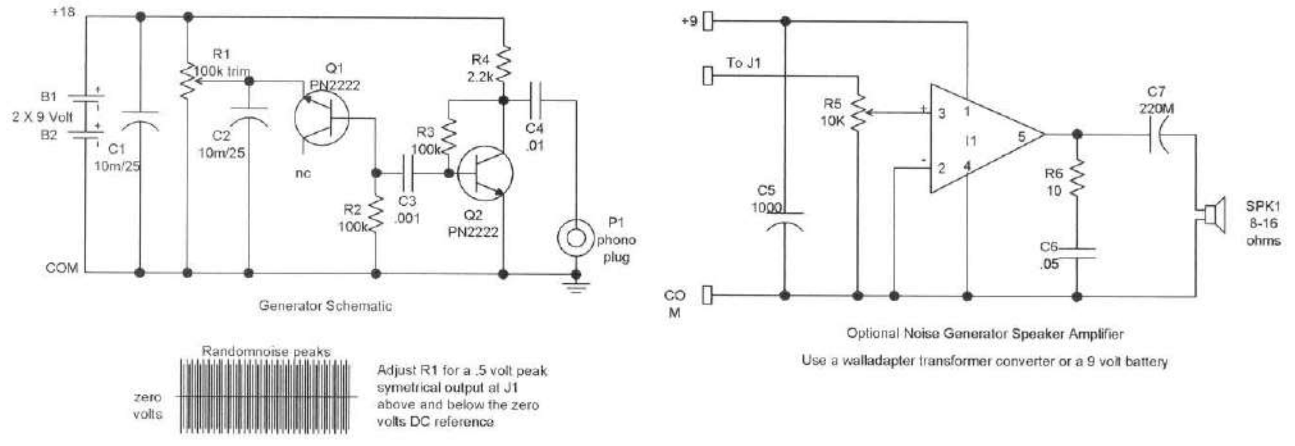

Adjust R1 for a .5 volt peak symetrical output at J1 above and below the zero volts DC reference

Use a walladapter transformer converter or a 9 volt battery

Figure 28-2 *Circuit schematic*

Circuit Description

Pink noise is a form of white noise that is the Gaussian distribution of all possible frequencies, the difference being that pink noise is more weighted to the audio spectrum. This form of noise has some very interesting properties, one being the ability to cause relaxation and a sense of well-being. Another property is that it provides a background that will completely mask an annoying device, rendering it unable to affect normal conversation. The device described is designed to work with any sound system with an audio input or is easily adapted to drive a loudspeaker by connecting it to a normal radio or building an optional speaker amp, as shown in the accompanying plans, making a complete system (see the parts list at the end of the chapter).

The circuit works in the following manner: A base emitter junction of transistor Q1 is reversed biased through a current-limiting resistor into breakdown (avalanche) condition. The random shot noise created is fed to the common emitter amplifier transistor, Q2, and to a filter, which in turn provides a low-level signal output. The unit module board and the batteries are placed in a plastic case with pieces of foam rubber or plastic. The outlet leads are fed out per the builder's requirements. The batteries are permanently installed and left on because the operating current is so low that there is no reason for a switch. Note that Jl is *not* required if using the optional speaker and amplifier circuit.

Assembly Steps

If you are a beginner it is suggested you obtain our *GCAT1 General Construction Practices and Techniques*. This informative manual explains basic practices that are necessary in the proper construction of electromechanical kits and is listed in Table 28-1.

1. Lay out and identify all the parts and pieces. Verify them with the parts list, and separate the resistors as they have a color code to determine their value. Colors are noted on the parts list.
2. Cut a piece of .1-inch grid perforated board to a size of 3 × 2.4 inches. Locate and drill the holes as shown in Figure 28-3 for the leads of the battery clips.
3. If you are building from a perforated board, it is suggested that you insert the components starting in the lower left-hand corner as shown in Figure 28-3. Pay attention to the polarity of the capacitors with polarity signs and all the semiconductors. Route the leads of the components as shown and solder as you go, cutting away unused wires. Attempt to use certain leads as the wire runs or use pieces of the #24 bus wire. Follow the dashed lines on the assembly drawing as these indicate connection runs on the underside of the assembly board.

4. Attach the external leads as shown in Figure 28-4. Notice the special note on the shielded cable.
5. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.

Testing Steps

Insert the two battery clips and plug the output cable into the microphone or auxiliary jack of the existing system. Note the rushing sound and adjust the trimpot R1 to the desired effect. If you have a scope, make the adjustment as shown in Figure 28-2. If you build the speaker amplifier section, simply connect it as shown for a self-contained system suitable for nightstand use.

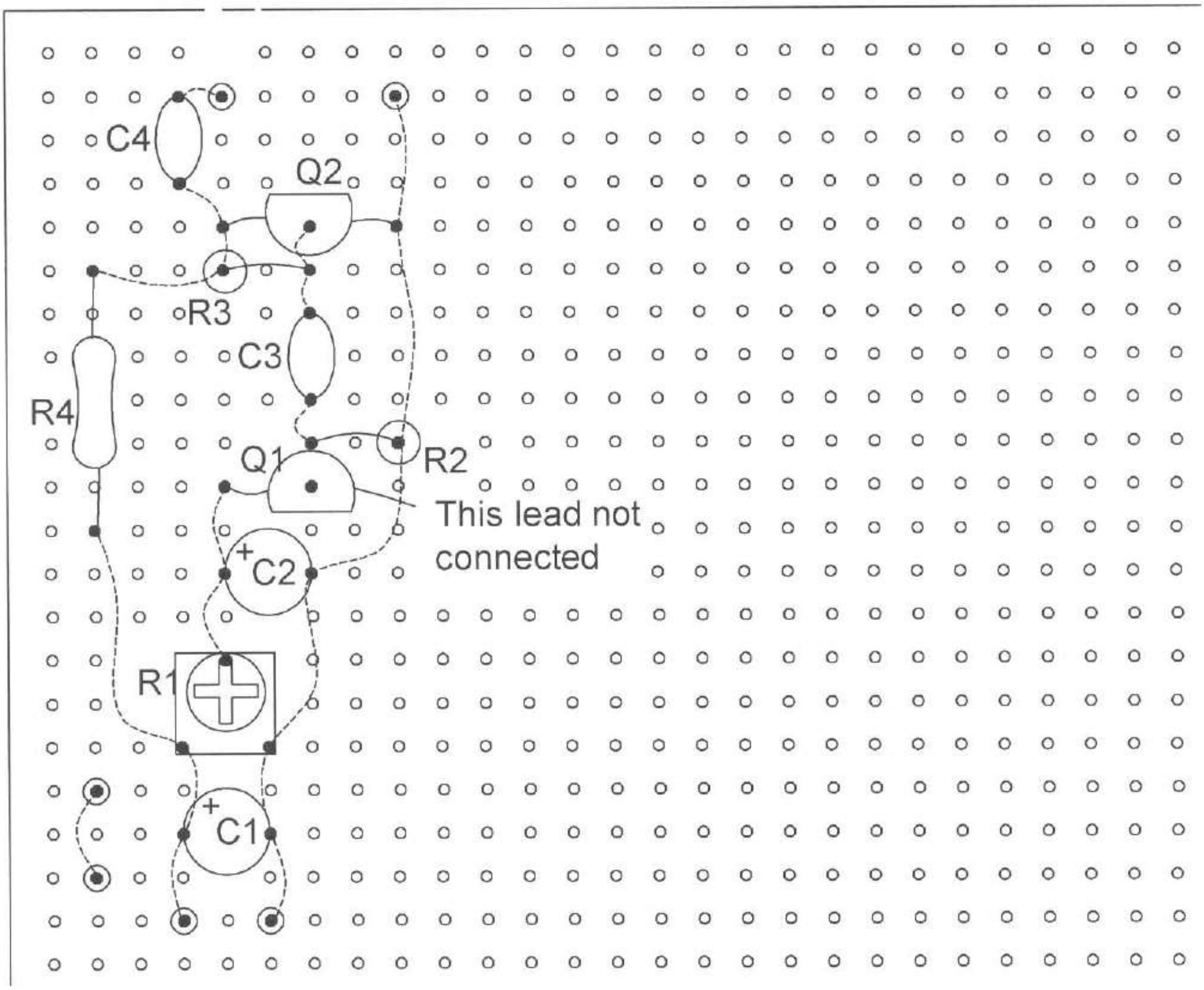

Figure 28-3 *Assembly board layout*

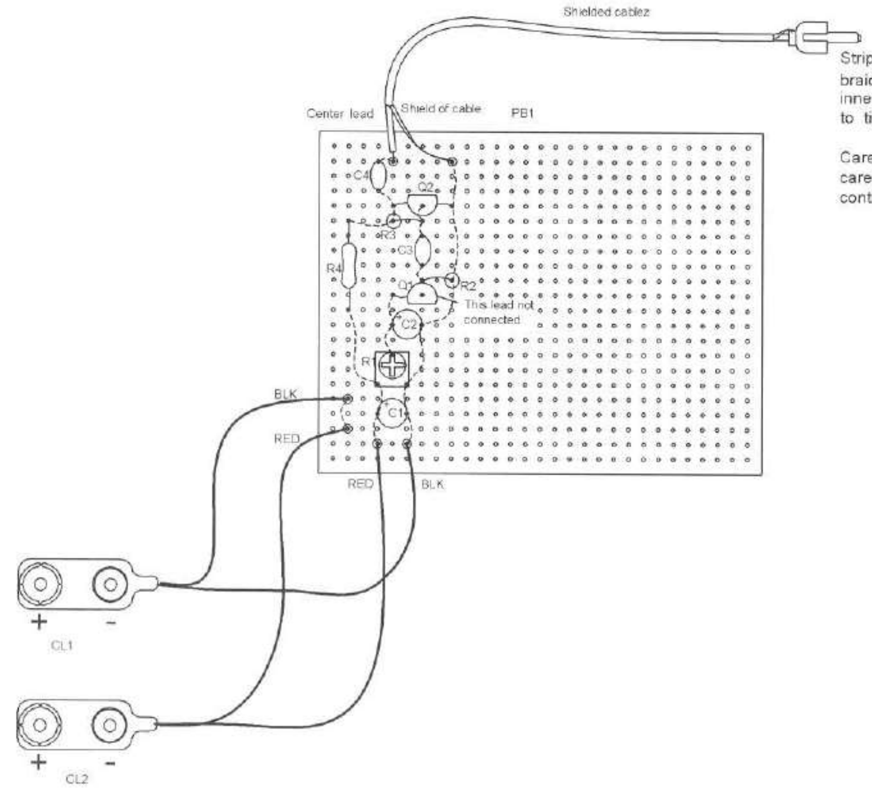

Strip off outer plastic jacket of cable 1 inch and twist braid into a lead and tin. Strip off ¼ inch of insulation on inner lead and tin. Insert into plug and solder inner lead to tip.

Carefully solder twisted braid lead to outer shell being careful not to burn through insulation. Always verify lead continuity and possible shorting of the leads with a meter.

Figure 28-4 *Assembly showing external leads*

Table 28-1 Noise Curtain Generator Parts List

Ref. #	*Description*	*DB Part #*
R1	100K trimpot vertical mount	
R2, 3	Two 100K, ¼-watt resistors (br-blk-yel)	
R4	2.2K, ¼-watt resistor (red-red-red)	
R5	10K trimpot vertical mount	
R6	10-ohm, ¼-watt resistor (br-blk-blk)	
C1, 2	10-microfarad, 50-volt vertical electrolytic capacitor	
C3	.001-microfarad, 50-volt disc capacitor	
C4	.01-microfarad, 50-volt disk capacitor	
C5	1,000-microfarad, 25-volt vertical electrolytic capacitor	
C6	.047-microfarad, 50-volt plastic capacitor	
C7	220-microfarad, 25-volt vertical electrolytic capacitor	
Q1, 2	Two PN2222 NPN general-purpose transistors	
I1	LM386 dual inline package (DIP) operational amplifier	
PB1	2.25 × 3 × .1 × .1-inch grid perforated circuit board	
SPK1	Small 8- to 16-ohm speaker	
CL1, 2	Two 9-volt battery clips and leads	
P1	RCA phono plug	
SHC1	12-inch shielded microphone cable	

Chapter Twenty-Nine

Mind-Synchronizing Generator

Caution: This machine can trigger epileptic seizures and should be avoided at all costs by anyone who suspects even slightly that they are epileptic.

Your mind synchronizer is designed to be a completely self-contained unit with built-in batteries (see Figure 29-1). The system produces a variable-rate, flashing, monochromatic, directional, "shaped" light source (a light source with a variable ratio of off to on times), and a complementary synchronized audio pulser with tone-shaping control. The unit may be used so that a group can be exposed to light and sound pulses. For individual use, headphones can be used, along with special assembled light glasses. These are shown made from an existing pair of sunglasses. Optimum performance requires a quiet and reasonably low light condition, and distractions must be at a minimal. The mind-synchronizer can produce strange and bizarre hallucinations, as well as provide a sense of relaxation and well-being.

Information on Mind Control

Today's work environment demands that we have the ability to excel, to compete, and to resolve problems. Hence, the most important tool we possess is our mind. To optimize our potential, we now offer our low-cost mind-synchronizing unit.

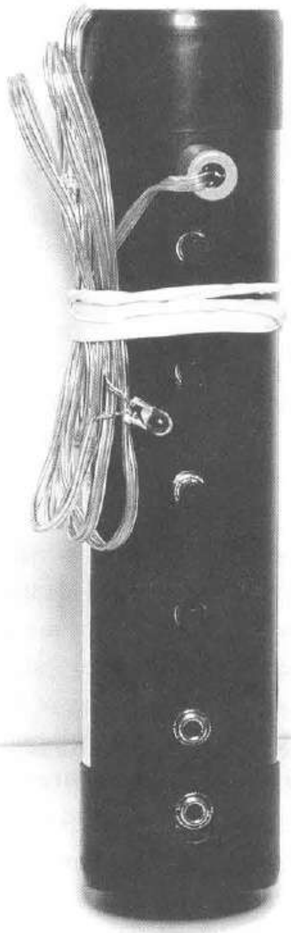

Figure 29-1 *Photograph of unit*

After 25 years of research, Dr. Axel Bruck of Berlin, Germany, created a machine that can enhance, develop, and perfect the qualities within you in order to attain the goal of mental fitness. It has been found that external stimulation evokes a mental reaction in people that is either active or passive. The mind synchronizer can determine automatically the kind of person you are. An evaluation is performed via the means of a hand sensor using a technique called biofeedback, which is sensitive to the flow of blood. This knowledge is very important for providing optimal stimulation to the users.

Having determined the kind of person you are, the mind synchronizer now uses "programmed" light pulses and "shaped" sound to help you reach a required mental condition, be it relaxation or efficiency. It has been known for hundreds of years that an external rhythm of light and sound has an effect on your mental condition. The mind synchronizer is a modern version of a drum and a flickering fireplace.

Through extensive scientific measurements, experiments, and results, the mind synchronizer can improve your performance level by using the right stimuli:

- Deep relaxation to the point of sleeping
- Relaxation
- Equilibrium
- Activation of muscles doing physical movements
- Efficiency

Whether in sports, professions, or personal lives, the development of personality and identity is vital to personal success and satisfaction. The mind synchronizer trains and perfects your mental and psychic abilities so that you have a stable way of thinking with a healthy outlook. The mind synchronizer teaches you the technique of concentration, stress handling, focusing, visualization, relaxation, and self-awareness

When using the mind synchronizer, you learn to set a goal and concentrate on this goal completely. You eliminate all distractions and thereby maintain your concentration over a long period of time. By doing so, you improve your power of concentration. In the long run, you will think clearly, make better decisions, and deal more effectively with life itself.

The program of relaxation combats stress, tension, and direct pressure. Almost everything we do causes stress. Stress means that the body is in a chaotic condition. The opposite of stress is relaxation. The right kind of relaxation is an optimization of mind, concentration, learning ability, performance, and health.

In reality, a relaxed person saves more than half a billion heartbeats over a nervous or hectic person, according to Professor Prinzinger's theory of the Quiet Pulse. Professor Prinzinger performed much research into the various relaxed states of the body and the relationship of relaxation to heart rate and blood pressure. The body consumes less energy, organs are burdened less, and life expectancy increases.

The program of efficiency is a newly developed psycho-active technique: the indication of unconscious eye movement and the employment of rhythm displacement. Certain stimulating rhythms such as that of a strobing light can produce nausea when strobing at 20 repetitions per minute. Other stimulating rhythms of lower repetitions per minute can produce relaxation and the feeling of well being. In such a case, eye movement usually follows at the lower repetition rates. The combination of visual effects (flashing *light-emitting diodes* [LEDs]) with audio input (via a headset) produces high-efficiency performance. The mind synchronizer establishes physical and mental readiness to attain this goal of high performance.

The program of activation involves more optic, visual patterns than acoustic ones to provide "imagination journeys." To experience associations with your own history and your own personality, you will be able to discover your emotional condition and personality.

The program of equilibrium allows for meditation with your eyes open. You would observe attentively, simply, and exactly each individual light source and its changes. The mind synchronizer will bring you the healthy balance you desire through light and sound.

Mind machines are available from a wide variety of sources today. Often referred to as hemispheric synchronizers or cortical-frequency entertainment devices, they all basically perform one basic function: They lead the brain to synchronize both hemispheres and then lead those synchronized hemispheric frequencies to a specific target frequency. In order to understand why anyone would want to do such a thing, let's examine what frequencies the brain produces and what people experience when the brain is at these frequencies.

Generally, brain waves are broken down into four different levels. They are referred to as beta, alpha, theta, and delta. With respect to frequency, they are as follows:

Beta: 13 cycles per second or more

Alpha: 8 to 12 cycles per second

Theta: 5 to 7 cycles per second

Delta: 1 to 4 cycles per second

Profound changes take place in a person's conscious experience at each of these levels. The beta frequency is the "normal" waking frequency, the drive-in-the-city, go-to-work, drink-coffee frequency, and it can range as high as 40 cycles per second. It is *not* a relaxing frequency.

The alpha frequency is a realm of relaxed awareness. It has been shown in a concrete fashion that this frequency is associated with an increased capacity for learning, comprehension, and retention. It is in the alpha state that the best learning occurs.

Theta is a hypnotic state often associated with meditation, a zen-like state of waking dreaminess. It can be a profoundly relaxed state where internal associations and thoughts can form ideas with clarity that could never have emerged in the normal, high-noise environment of the beta state. This is also the range that many native shamans have been found to enter when engaging in healing trance work. It is a state of profound tranquility.

The delta state is much simpler to define; you're asleep. When your brain waves are in the delta range, you're in a sound and restful sleep.

A number of studies have been done that indicate that a flashing light has a profound impact on the entire brain, more so than a pulsing sound or feeling. When the brain is exposed for a time (as short as 2 to 5 minutes) to a flashing light at a speed close to the brain's existing operating frequency, the brain will begin to synchronize with the stimulation. Moreover, both hemispheres, which are often operating at different frequencies, will harmonize. This has also been shown clinically and statistically to be an aid to mental functioning. Many books are available on the market on this subject for anyone who would like more detailed information on these studies.

Operating a Mind Machine

Now with the previous information in mind, the operation of the mind machine in all of its forms is basically the same. Start the machine off at a frequency that is roughly equivalent to the operating frequency of your brain at the time. (Use your own feelings as a gauge. If you're frazzled, start at maximum frequency; if you're pretty mellow, start at about half-frequency.) Allow yourself about 3 to 4 minutes at this frequency, and then slowly begin bringing the frequency down, very slowly. You should take about 7 to 10 minutes to bring the control all the way down. When you reach a minimum, leave it there for about 5 to 7 minutes.

After you've gotten to the bottom range, you can either lead yourself back to normal, in which case you will feel what you may consider a "good normal." If you so choose, you can exit directly from the lowest setting. If you do, have taken your time getting to it, and stayed there for several minutes, you will notice that it may be very hard to begin to move. Your limbs may feel very heavy, as if your body were asleep. If you manage to get to your feet, you may notice that things won't feel as solid; they'll seem to have a rubbery quality to them. These sensations can be quite pleasant as long as you're prepared and expect them (and don't have to drive or operate heavy machinery afterward!).

The visual effects are the most stunning and can only be likened to the experience of ingesting certain mind-altering substances. As the frequencies change, you will notice colors. Yes, two little flashing red lights in front of your closed eyes will create the illusion of fantastic, swirling colors and geometric designs ranging from ultimate simplicity to incredible complexity. One effect will be observed as myriads of little shapes that form and begin to slowly spiral around your field of vision, first clockwise, then counterclockwise, then back again. You will see dreamy, smooth, cloudlike fields of frosty blue-white, deep blues, deep reds, electric greens, yellows, whites, and purples, all in balls, stripes, stars, and triangles, dancing and moving like the most intense kaleidoscope you can imagine. The imagery is absolutely fascinating and can transfix you with the dazzling display. You will be absolutely amazed and impressed with the fact that it is your brain that produces this light show.

This unit is designed to take you from a medium-level beta down to a low theta. You can use these states as you desire. You may wish to stop the frequency decrease early and remain in the alpha state before an exam, prior to entering a class, or when trying to learn something important. You can continue on down and enter the theta level and create healing imagery in your mind to assist your body's own natural healing processes.

Most of all, play with your unit and enjoy the show. Built right, it will provide many years' worth of reliable service. Through usage, you can become familiar with the settings and be able to use the machine in a way that's exactly right for you. Always leave yourself plenty of time to use the machine and at least a half-hour after to "normalize." Use the unit to help you drift off to sleep; you can actually lead your brain down into relaxation and turn the unit off after you've gone into a deep theta, and if you're ready to sleep away, it's only a short slide from theta into the peaceful slumber of delta.

Circuit Theory

Integrated circuit I1 is an astable oscillator with a pulse rate being controlled from approximately 3 to 40 reps per second by external pot R2. The output of Ul now triggers integrated circuit I2 being connected as a monostable circuit that controls these pulses' ratios of on to off time. This duty cycle control is via external pot R5. Light-emitting diodes (LEDl and LED2) are connected to the output pin 3 of I2 through brightness control pot R7. These are the diodes you install in the sunglasses or they can be a single ultrabright LED for group use. The result is a variable flash rate from 3 to 40 pulses per second with a variable duty cycle rate (ratio of on to off time).

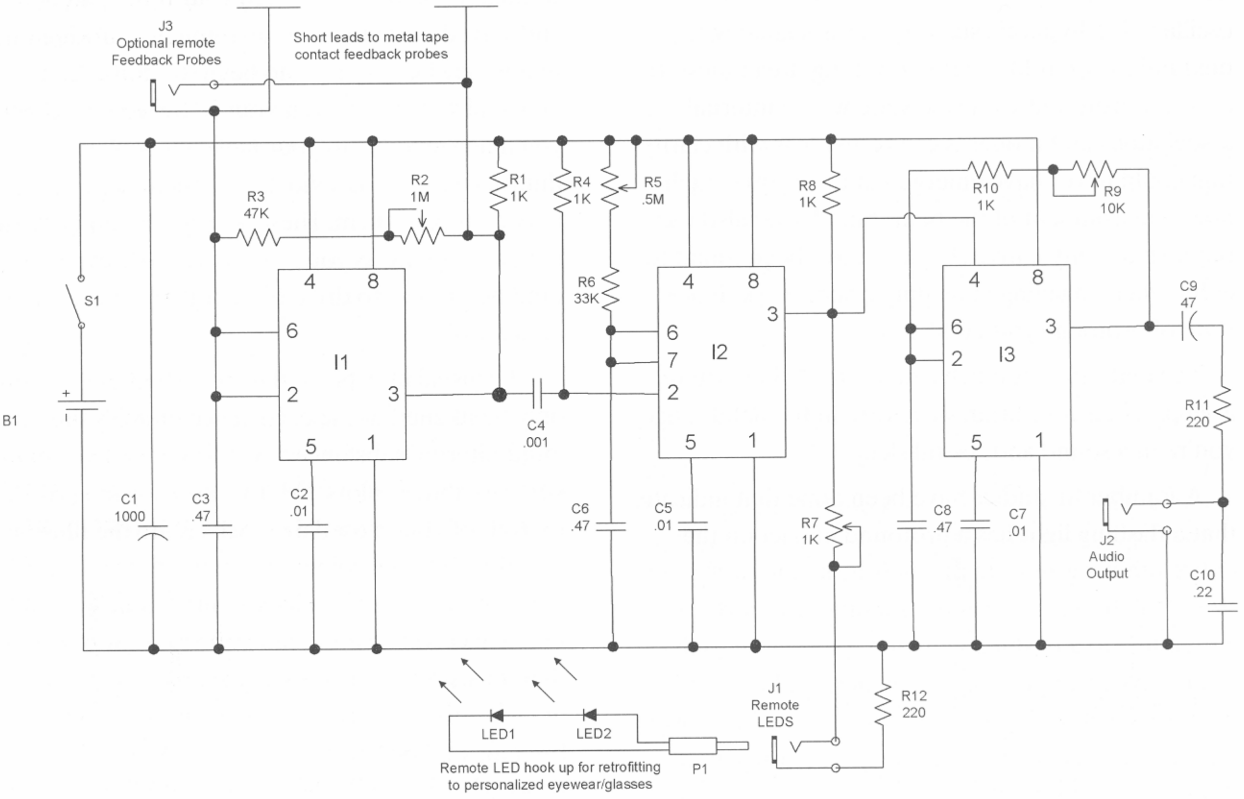
J3
Optional remote
Feedback Probes
Short leads to metal tape
contact feedback probes
R3
47K
R2
1M
R1
1K
R4
1K
R5
.5M
R8
1K
R10
1K
R9
10K
S1
B1
I1
I2
I3
R6
33K
C4
.001
C9
47
R11
220
C1
1000
C3
.47
C2
.01
C6
.47
C5
.01
R7
1K
C8
.47
C7
.01
J2
Audio
Output
C10
.22
LED1
LED2
Remote LED hook up for retrofitting
to personalized eyewear/glasses
P1
J1
Remote
LEDS
R12
220

Figure 29-2 *Circuit schematic*

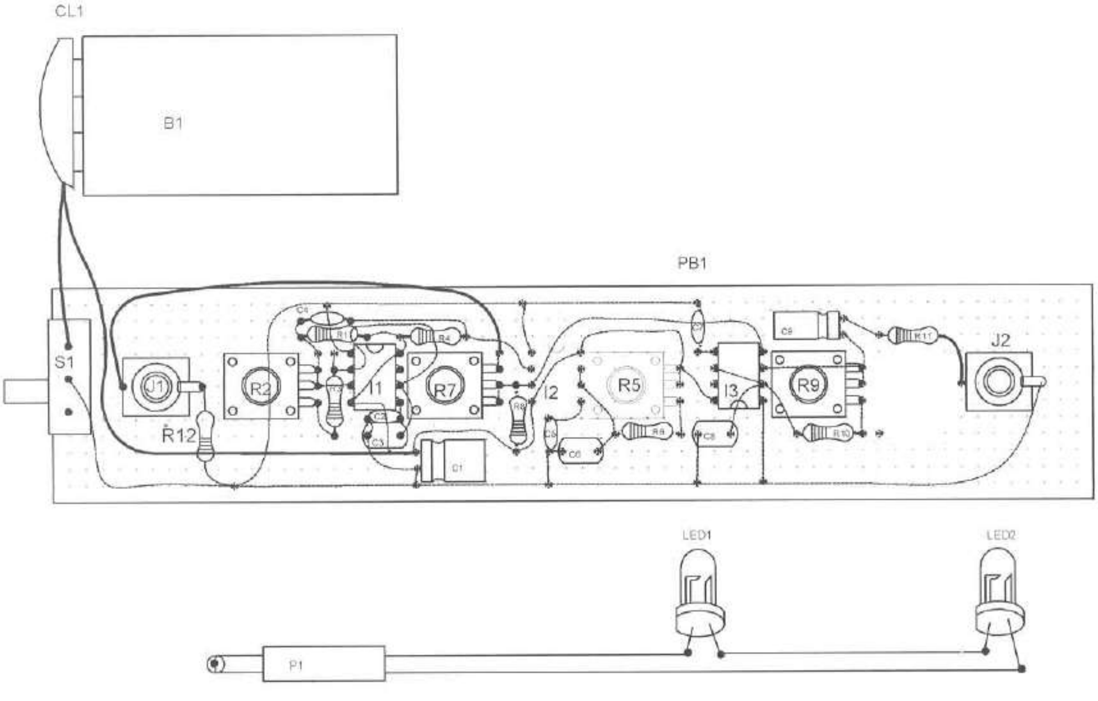

Emitter assembly for retrofitting to personalized eyewear glasses

Figure 29-3 *Assembly board*

Note it is best to experiment with and record the duty cycle settings and their effects (see Figure 29-2).

The circuit also includes integrated circuit I3 connected as a variable-rate, sonically "shaped" audio oscillator. The output is via jack J2 driving a set of 32-ohm headphones.

The feedback circuit consists of two external pieces of metallic tape providing points of contact for the user's hand. Body resistance now partially controls the pulse rate of I1. An uptight situation provides lower contact resistance and consequently raises the pulse rate. As one relaxes, this resistance becomes higher and lowers the pulse rate. An external jack (J3) provides a connection for other bodily contact probes to the feedback system. These contact probes are left to the discretion of the user. This jack is omitted when using the external tape probes. Switch S1 controls power to the circuit and should be off when not in use to conserve the internal batteries.

Assemble the Board

If you are a beginner, it is suggested to obtain our *GCAT1 General Construction Practices and Techniques*. This informative manual explains basic practices that are necessary in proper construction of electromechanical kits and is listed in Table 29-1.

1. Lay out and identify all the parts and pieces. Verify them with the parts list, and separate the resistors as they have a color code to determine their value. Colors are noted on the parts list at the end of the chapter.

2. Cut a piece of .1-inch grid perforated board to a size of $6^1/_4 \times 1^1/_4$ inches, as shown in Figure 29-3. An optional PCB is individually available.

   The circuit is shown with the more challenging perforated circuit board often required for a science fair project. A PCB is also available and requires that you only identify the particular part and insert it into the respective holes as noted. The PCB is plainly marked with the part identification. Soldering is very simple as you solder the component leads to the conductive metal traces on the underside of the board.

   The perforated board approach is more challenging as now the component leads must be routed and used as the conductive metal traces. We suggest that the builder closely follow the drawings on this section and mark the actual holes with a pen before inserting the

Obtain a pair of sunglasses preferably the ones with side shields. Measure the distance between your eye centers. This is usually about 2½". Drill the two holes just a bit smaller than the LEDS to provide a secure fit. Press into the holes and glue if necessary to further secure.

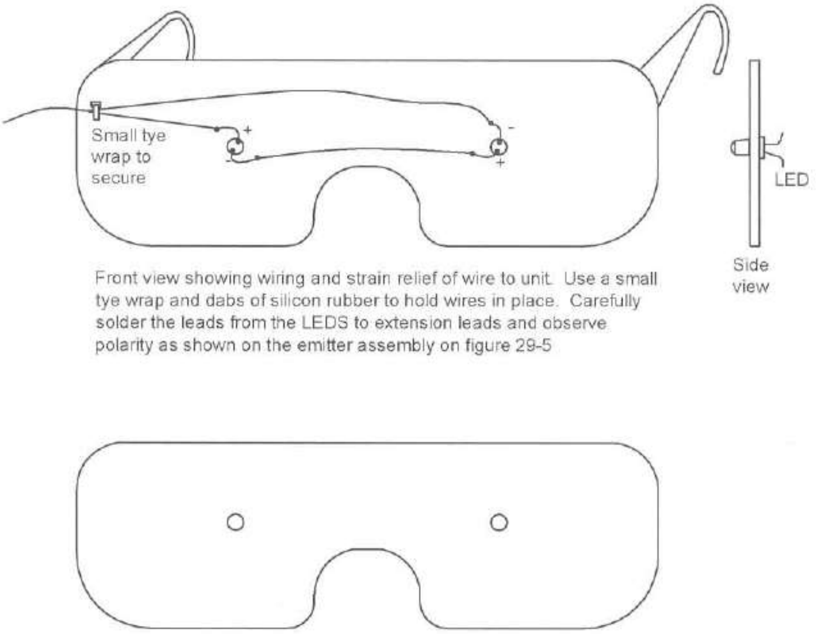

Front view showing wiring and strain relief of wire to unit. Use a small tye wrap and dabs of silicon rubber to hold wires in place. Carefully solder the leads from the LEDS to extension leads and observe polarity as shown on the emitter assembly on figure 29-5

View towards the eyes. Secure the leads to the frames of the assembly to keep out of the way. You may also use pieces of duct or electrical tape to hold all in place.

Figure 29-4 *Assembly of glasses*

parts. Start from a corner, using it as a reference, and proceed from left to right. Note that the perforated board is the preferred approach for science projects as the system looks more homemade.

3. If you are building from a perforated board, it is suggested that you insert the components, starting in the lower left-hand corner, as shown in Figure 29-3. Pay attention to the polarity of the capacitors with polarity signs and all the semiconductors. Resistor RI2 is connected from the pin of 11 to the PC board. Note the control pots R2, R5, R7, and R9 and jacks J1, J2, and J3 should be positioned to closely match the holes, as shown in Figure 29-5. Route the leads of the components as shown and solder as you go, cutting away unused wires. Attempt to use certain leads as the wire runs or use pieces of the #24 bus wire. Follow the dashed lines on the assembly drawing as these indicate the connection runs on the underside of assembly board.
4. Attach the external leads to the battery holder as shown.
5. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.
6. Assemble an LED to cable P1 as shown. Also, retrofit to the eyewear as per Figure 29-4.
7. Fabricate the enclosure tube (TUB1) as shown in Figure 29-5 and insert the assembly board and battery. Verify that the controls mounted on the assembly board will match the drilled holes with the dimensions as shown. Note the dimensions are shown for use with the optional PCB and may be changed. The unit is secured attaching the adjustment pots via nuts.
8. The final assembly is shown in Figure 29-6 and should be operated as instructed.

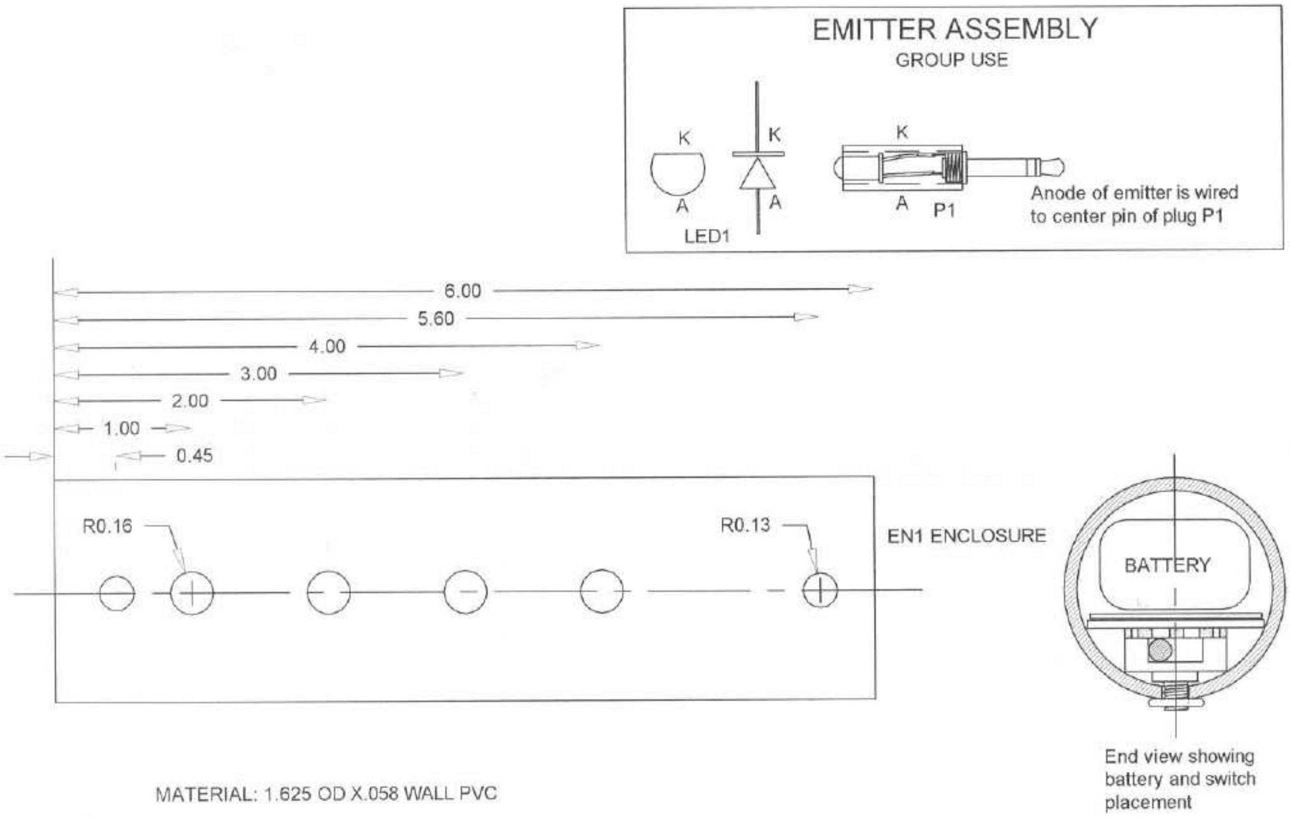
EMITTER ASSEMBLY
GROUP USE
K
K
K
A
A
A
P1
LED1
Anode of emitter is wired to center pin of plug P1
6.00
5.60
4.00
3.00
2.00
1.00
0.45
R0.16
R0.13
EN1 ENCLOSURE
BATTERY
End view showing battery and switch placement
MATERIAL: 1.625 OD X.058 WALL PVC

Figure 29-5 *Fabrication of enclosure*

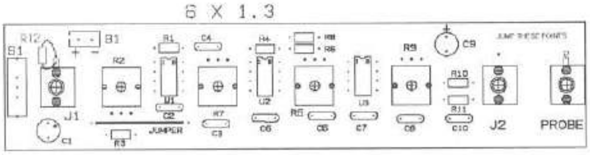
6 X 1.3
R12
B1
R1
C4
R4
R8
R6
R9
C9
R2
R5
R10
R11
J1
C1
C2
JUMPER
R3
R7
C3
C6
C5
C7
C8
C10
J2
PROBE

Figure 29-6 *Final assembly on PC Board*

Table 29-1 Mind-Synchronizing Generator Parts List

Ref. #	*Description*	*DB Part #*
R1, 4, 8, 10	Four 1K, ¼-watt resistors (br-blk-red)	
R2, 5	.5-1 meg potentiometer resistor	
R3	47K, ¼-watt resistors (yel-pur-or)	
R6	33K resistor (or-w-or)	
R7	1K pot resistor	
R9	10K pot resistor	
R11, 12	Two ¼-watt 220- to 330-ohm, resistors (red-red-br)	
C1	1,000-microfarad, 25-volt vertical electrolytic capacitor	
C2, 5, 7	Four .01-microfarad, 50-volt disc capacitors	
C4	.001 mfd, 50-volt disk capacitor	
C3, 6, 8	Three .47-microfarad, 50-volt mylar capacitors	
C9	47-microfarad, vertical electrolytic capacitor	
C10	.22-microfarad, 50-volt mylar capacitor	
I1, 2, 3	Three 555 dual inline package (DIP) timers	
*LED1, 2	Two bright-red LEDs or use single, optional 15,000 mcd assembled as shown	
S1	Single pole, single throw (SPST) PC switch	
CL1	9-volt battery snap	
11, 3	Two 3.5 mm mono jacks	

Operation and explanation of controls

1. Slide of bottom cap and insert a 9 volt battery

2. Plug in 30 ohm headsets into J2

3. Plug in cable to LEDS in glasses

4. Identify and study the functions of the four controls

   R2-Frequency control allows setting from 1-24 reps

   R7-Allows setting of LED brightness

   R5-Sets the ratio of "off to on" time. Will interact with R2 if set to long "on" time. Suggest to preset full CW and then set for desired effect

   R9- Allows tone adjustment or sound shaping.

5. Adjust the unit to produce a test tone of about 3 pulses per second. Grab feed back probes being the metal strips along the side of the unit and note pulse rate increasing. It is suggested to become familiar with the operation and resulting effects before attempting to use the feedback functions.

6. Unit is easily hand held with all controls accessible

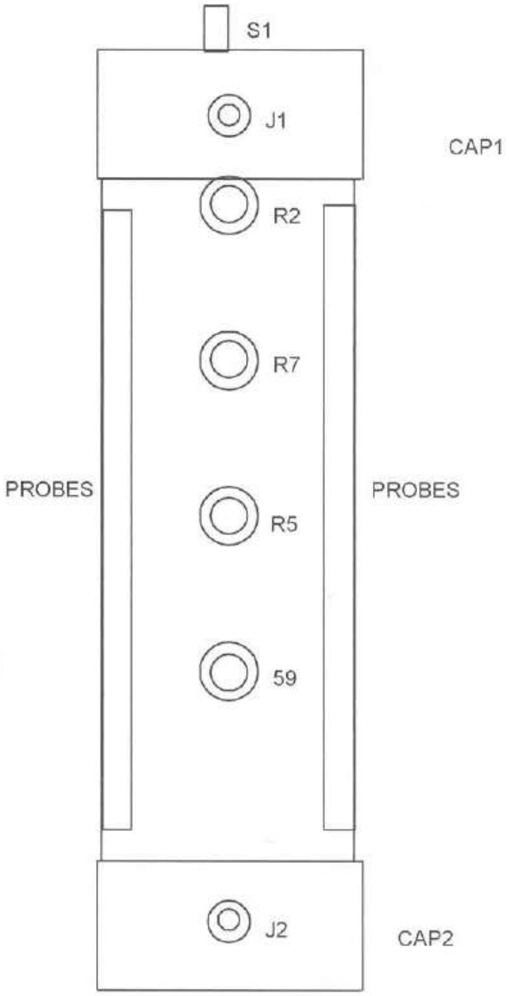

Figure 29-7 *Operation and Explanation of Controls*

J2	3.5 mm stereo jack
WR1	12 inches of #24 hookup wire
WR2	24 inches of #24 bus wire
WR3	48 inches of two-cord speaker wire
PBl	.1 × 6¼ × 1¼-inch grid perforated board
PCBOARD	Optional printed circuit board (PCB) with printing DB# PBMIND
EN1	Enclosure, sized at 1⅝ OD × 1 7/16 ID × 7 inches
CAl, 2	Two 1⅝ inch plastic caps #A 1⅝
P1	3.5 mm plug for LED cable to eyewear
HEADSETS	Stereo or mono, 32-ohm, cushioned headphones
TAPE	4 × ½-inch metallic silicon tape229

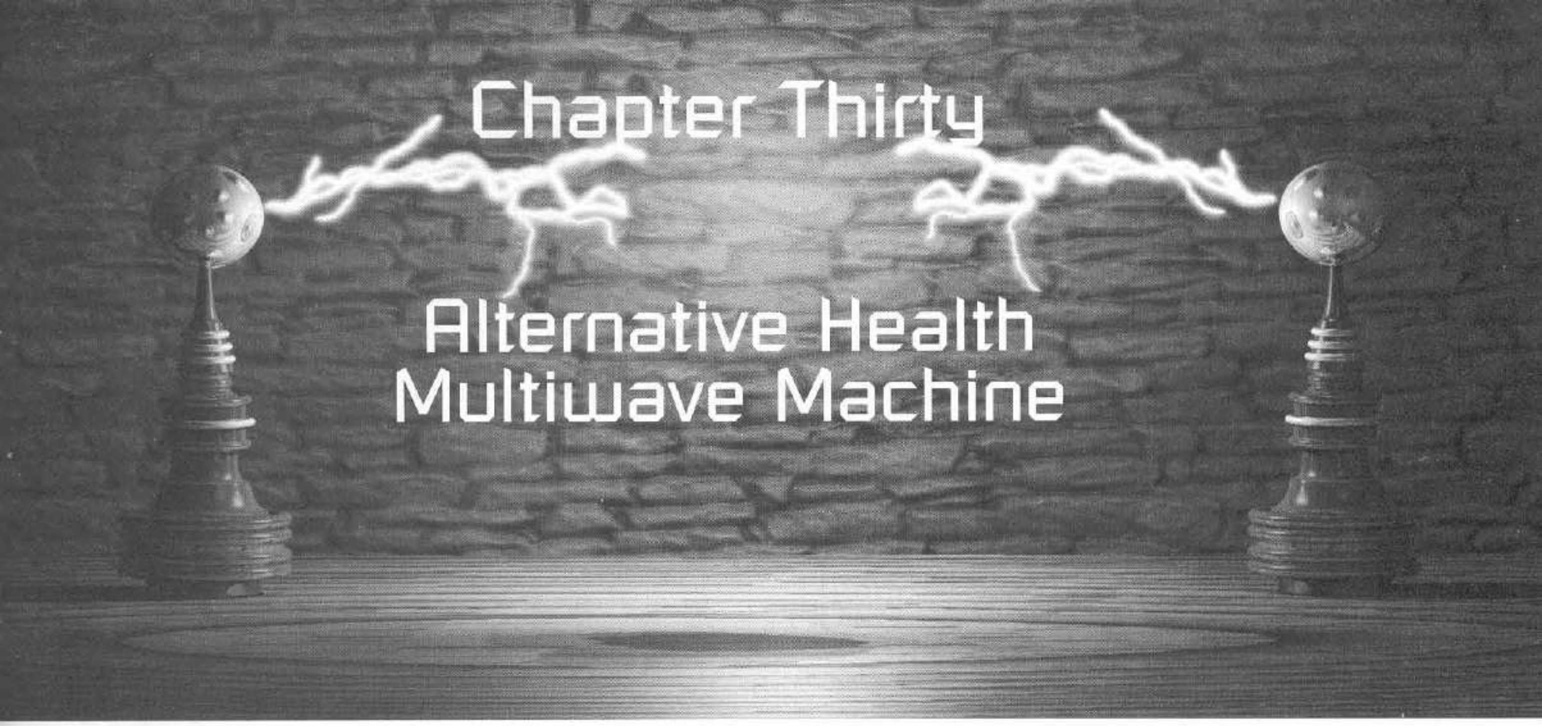

Chapter Thirty

Alternative Health Multiwave Machine

Figure 30-1 shows how to construct a Lakhovsky multiwave alternative health system. Dr. Lakhovsky was a Russian researcher who experimented with the effects of high-frequency electromagnetic radiation on various parts of the living organism. This controversial system is claimed to cure many health ills, including cancer. Its claim to fame is that all cells in a living body possess an intrinsic, vibrating resonant frequency that can be energized by an external means. Diseased cells appear to have a weaker and different frequency that can be brought back into step with the adjacent nondiseased cells. Antiaging appears to occur when cells are exposed to this full-spectrum electromagnetic vibratory energy. It is of special interest that this method of cellular regeneration has generated considerable interest in several research fields as a possible cure for cancer.

This project shows how to construct the multiwave radiating coil and antenna from readily available parts and pieces. System operation will require a working BTC30 12-inch-spark Tesla coil, as described in Chapter 14 of *Electronic Gadgets for the Evil Genius*. This Tesla coil is also available through plans, a kit, or as a complete and ready-to-use coil on the Information Unlimited Web site at amazing1.com.

It is suggested that the builder obtain the Lakhovsky handbook as listed in the parts list at the end of this chapter and study the methods of application and testimonials on the use of this controversial health machine before using it on one's self or others.

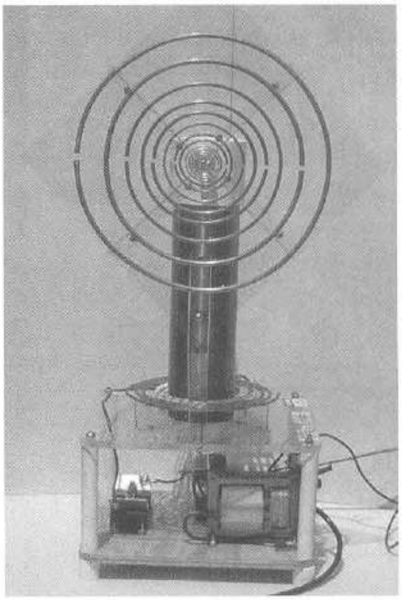

Figure 30-1 *The multiwave antenna*

Theory of Operation

Figure 30-2 shows a block diagram of the radiating coil and Tesla driver. The antenna coil contains a series of concentrically wound coil rings mounted in a flat plane. These coils all possess an inherent resonant

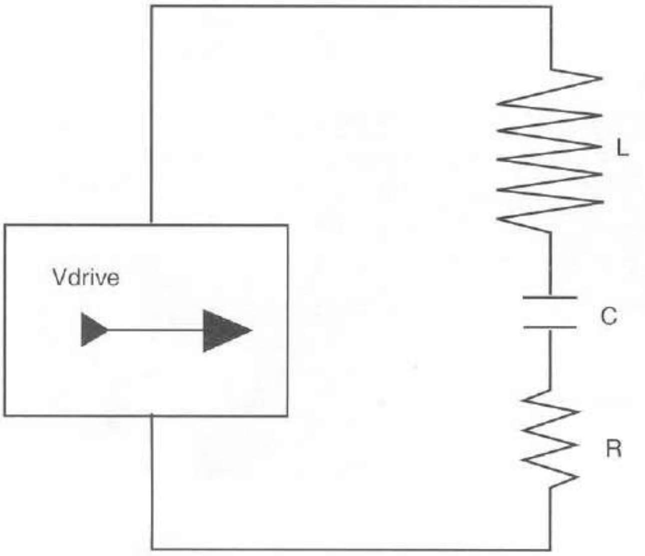

L1 to Ln represents the inductance of all of the coils.
C1 to Cn reprasents the capacity of all of the coils.
R1 to Rn is the radiation resistance of each section.
Frequency are the sum of the resonant sections ranging up to the gigahertz.

L is the sum of L1 to Ln
C is the sum of C1 to Cn
R is the sum of R1 to Rn and is mainly the radiation resistance.

The undamaged current wave from the output of the Tesla driver induces voltages at many frequencies into the multicoils of the antenna.

Figure 30-2 *Diagram of system*

1/16" clear lexan (polycarbonate)

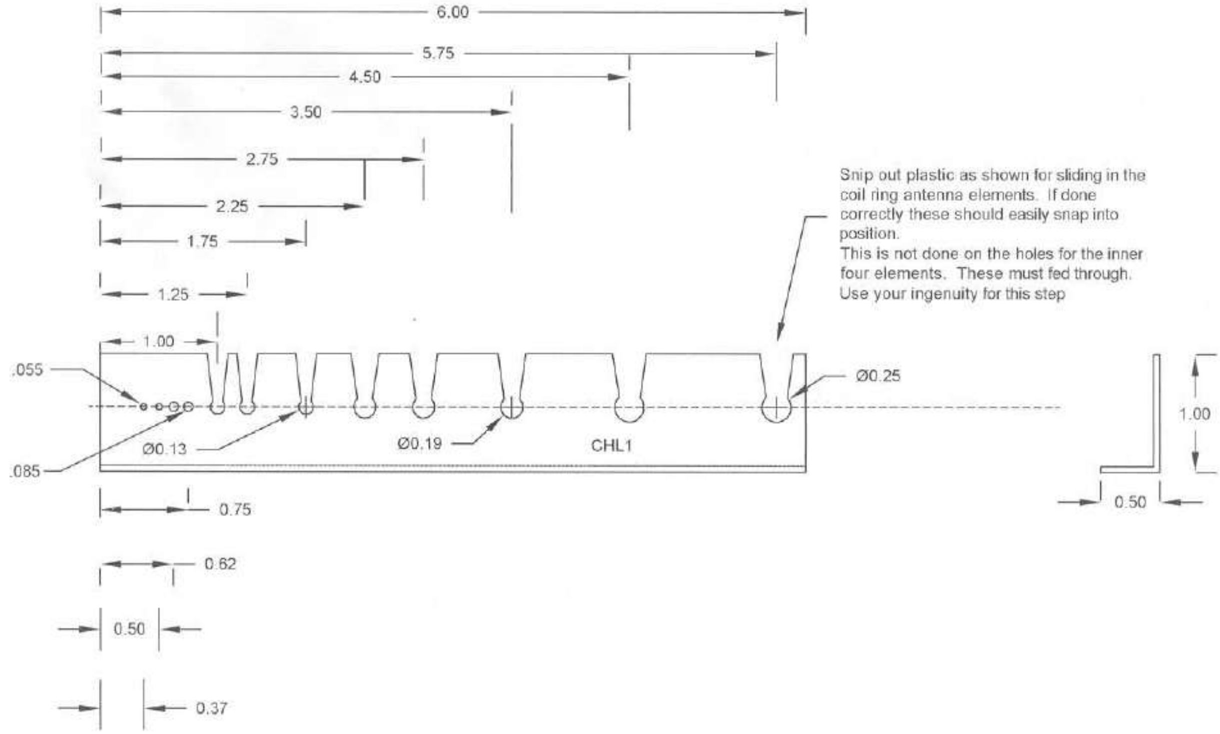

Snip out plastic as shown for sliding in the coil ring antenna elements. If done correctly these should easily snap into position.
This is not done on the holes for the inner four elements. These must fed through. Use your ingenuity for this step

Figure 30-3 *Fabricating the four coil holders*

frequency as a function of their physical parameters. Harmonics are also produced and with all the coils oscillating they create a near-continuous spectrum of radio-frequency energy. A spark-driven Tesla coil is required for optimum results, as it generates many fast steps of current rich in harmonic content that drives the antenna coil.

Construction of the Multiwave Coil and Antenna

1. Create the four coil holders (CHOLD1) as shown in Figure 30-3. Use 1/16-inch Lexan polycarbonate.
2. Create the 12 × 12-inch main plate (MP1) from a piece of 1/16-inch Lexan sheet, as shown in Figure 30-4.
3. Mount the coil holders as shown and carefully drill the mounting holes. Use nylon or brass screws and nuts for mounting the holders to the plate. It is very important to have these holders as exact as possible or the wound coils will not fit properly in their predrilled slots.
4. Form the circular coil rings from the appropriate diameter copper tube and snap the tubing into place. Again attempt to form circles as perfect as possible for proper operation and esthetics. The inner coil rings are formed pieces of bus wire and are threaded through their respective holes. Note a 1/8-inch gap at the coil ends.
5. Attach a PVC cap (CAP1) using a small metal screw, solder lug, and nut. This screw is the high-voltage feed point to the antenna and is connected to a piece of wire soldered to the innermost coil ring (see Figure 30-5).
6. The antenna may be used with any suitable high-voltage, high-frequency generator. It is shown in Figure 30-6 with our BTC30 250-kilovolt Tesla coil, as described in Chapter 14 of *Electronic Gadgets*. It is available as a kit or completed system as noted in Table 30-1.

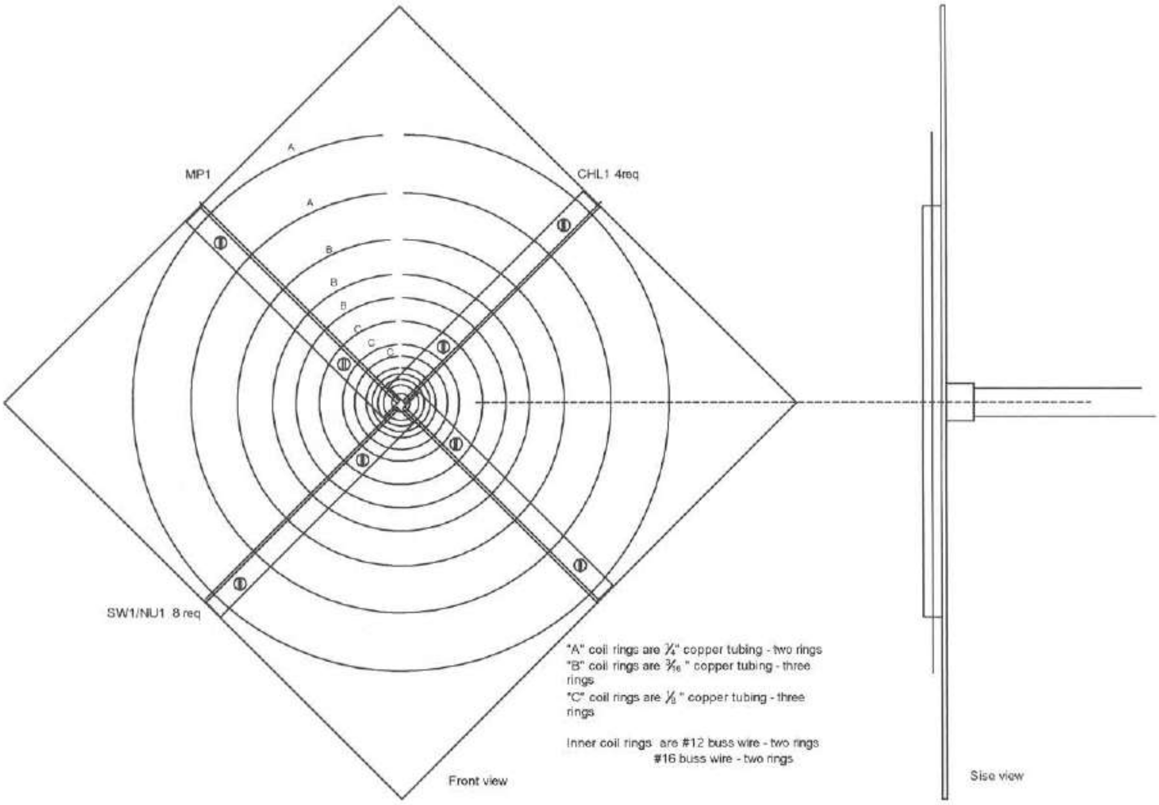

"A" coil rings are 1/4" copper tubing - two rings
"B" coil rings are 3/16" copper tubing - three rings
"C" coil rings are 1/8" copper tubing - three rings

Inner coil rings are #12 buss wire - two rings
#16 buss wire - two rings

Figure 30-4 *Creating the base*

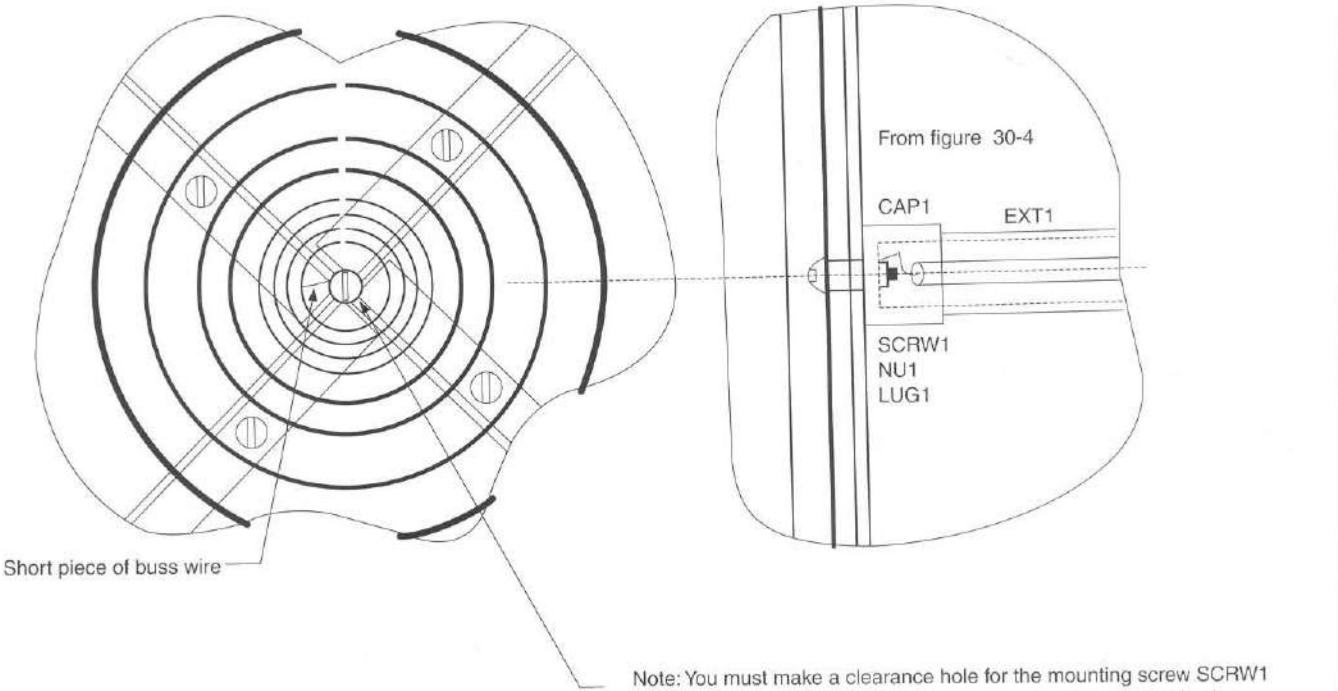

Note: You must make a clearance hole for the mounting screw SCRW1 exactly in the center. It may be necessary to heat up a suitable size diameter metal rod and carefully melt a clearance section out of the four coil holders for the head of this screw. You must now connect a short piece of buss wire from this screw to the inner most ring and solder in place.

Figure 30-5 *Mounting and feeding details*

7. If you intend to use our coil, you will need to obtain a 3-inch slip cap (SLP1) and drill a hole for the extension arm tube (EXT1). You will also need the plastic stabilizing bracket (BRK1) required to keep the coil assembly upright.
8. The connection between the output of the Tesla coil and the antenna is made via the wire sleeved through the extension tube that connects to the mounting screw and exits through the rear to make contact with the Tesla coil output lead.

Operation

The following instructions reference the use of our 250-kilovolt Tesla coil as listed in Table 30-1 or as described in Chapter 14 of *Electronic Gadgets*.

1. Preset the spark gap on the Tesla coil to no more than 1/16 of an inch. Connect the tap lead for maximum inductance.
2. Turn on the switch on the Tesla coil and note all the antenna coil rings adjacently sparking and arcing between one another. Note that the outer ring may be grounded to the chassis to increase output activity.

At this point, it is recommended that you obtain the referenced book (LSK1), the *Lakhovsky Multiple Wave Oscillator Handbook,* as referenced in the parts list and familiarize yourself with the potentials, treatments, and experiments possible with this system.

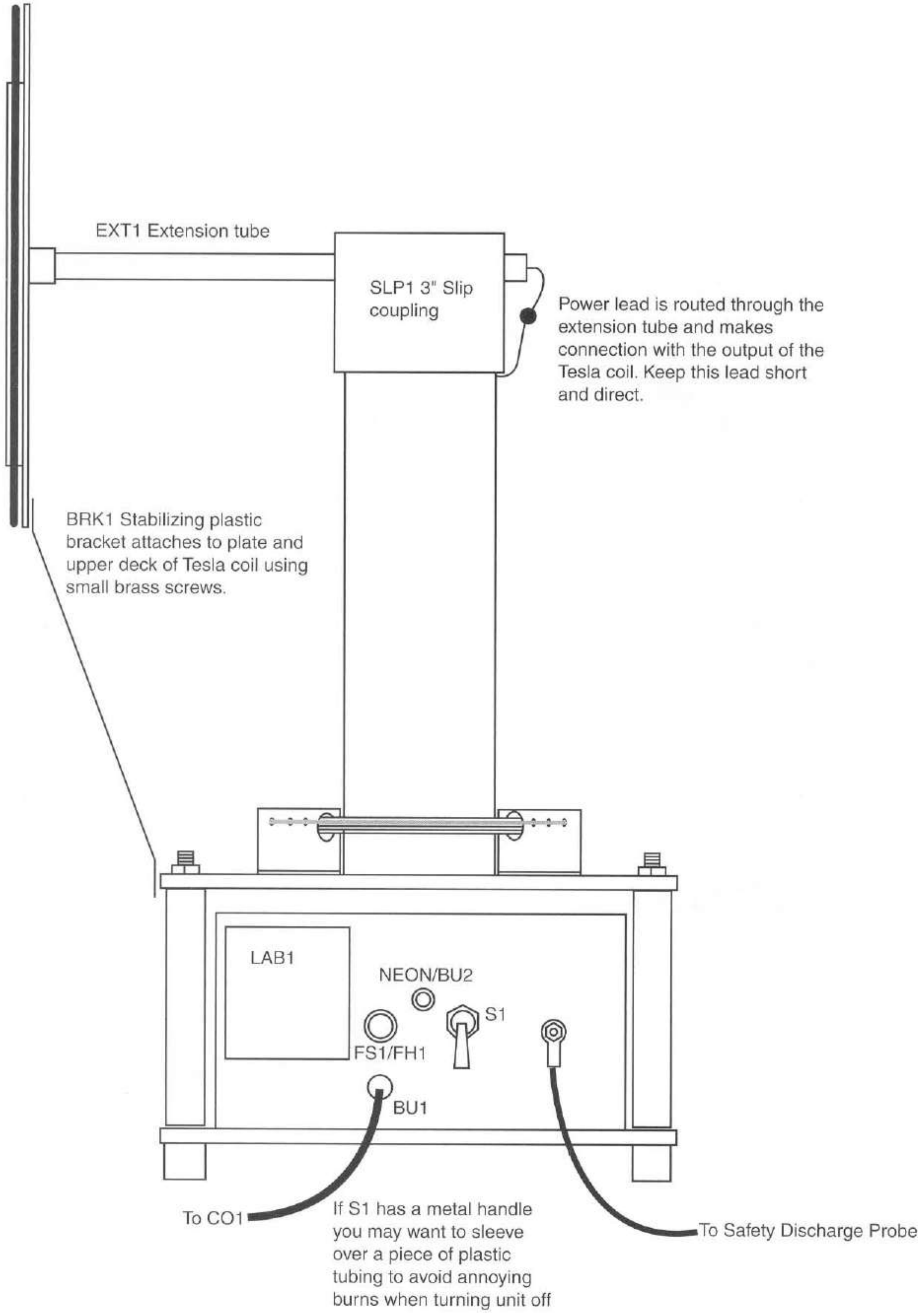
EXT1 Extension tube
SLP1 3" Slip coupling
Power lead is routed through the extension tube and makes connection with the output of the Tesla coil. Keep this lead short and direct.
BRK1 Stabilizing plastic bracket attaches to plate and upper deck of Tesla coil using small brass screws.
LAB1
NEON/BU2
S1
FS1/FH1
BU1
To CO1
If S1 has a metal handle you may want to sleeve over a piece of plastic tubing to avoid annoying burns when turning unit off
To Safety Discharge Probe

Figure 30-6 *Completed system*

Table 30-1 Multiwave Machine Parts List

Ref. #	*Description*	*DB Part #*
CHL1	Four coil-ring holders shown in Figure 30-3	DB# CHL1LAK
MP1	Mounting plate shown in Figure 30-4	
ARING	Two ¼-inch copper, soft drawn tubings	
BRING	Three 3⁄16-inch copper, soft drawn tubings	
CRING	Three ⅛-inch copper, soft drawn tubings	
INNERRING	Two #12 bus wire circular rings	
INNERRING SMALL	Two #16 bus wire circular rings	
SW1/NU1	106-32 × ½-inch brass or nylon screws and nuts	
SCRW1/NU1	6-32 × ¾-inch brass screw and nut	
LUG1	#6 solder lug	
WR20	24 inches of #20 bus wire	
BRK1	Stabilizing bracket shown in Figure 30-6	
CAP1	½-inch PVC slip cap	
SLP1	3 inches of PVC slip coupling	
EXT1	18 inches of ½-inch PVC tubing	
BTC3K	250-kilovolt Tesla coil generator kit and plans	DB# BTC3K
BTC30	250-kilovolt Tesla coil generator ready to use	DB# BTC30
LSK1	Book on Lakhovsky, 144 pages	DB# LSK1

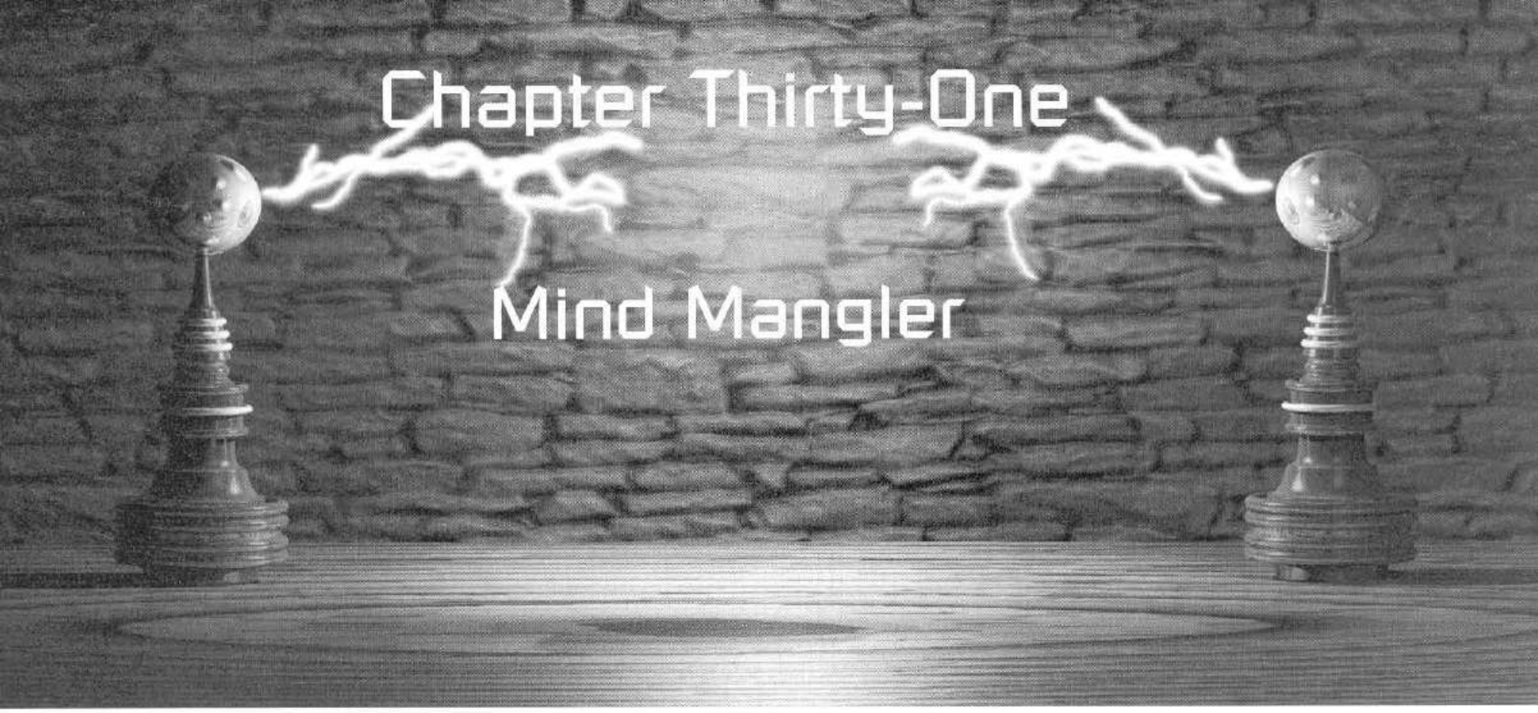

Chapter Thirty-One

Mind Mangler

This board-level device, as shown in Figure 31-1, when properly assembled, can be a great prank. When placed in one's bedroom under normal light condition, it does nothing. As soon as the lights are turned off, it comes to life, producing pulses of high-frequency sound much like an insect. These control pulses can be timed to occur about every minute. When the unsuspecting victim attempts to locate it by turning on the lights, it ceases operation. Even using a flashlight will disable it. The long time between pulses also makes it very difficult to detect. Properly hidden, it can take a long time to locate and may turn a prank into a nasty situation.

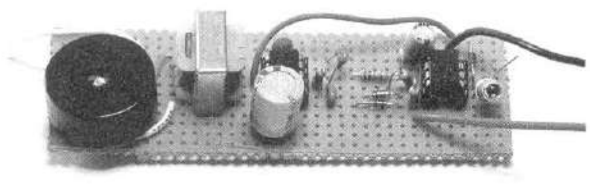

Figure 31-1 *Board-level view of the mind mangler*

Circuit Operation

Figure 31-2 shows a timer, I1, as a free-running, astable pulse generator producing a symmetrical 10-second, off and on pulse controlling the second timer, I2. Timer I1 has a phototransistor Q1 connected in series with the reservoir capacitor C1. As long as there is light on Q1, C1 can never discharge due to the conductance of Q1. Therefore, pin 3 is at a low and I2 is disabled. When the lights go out, C1 now is allowed to discharge and pin 3 goes high, turning on I2 and producing the high-frequency sound to the output transducer TD1. The high-audio frequency is determined by resistor R3 and C3, and can be changed by altering these values. The output signal from I2 is stepped up by transformer T1 (see the parts list at the end of the chapter).

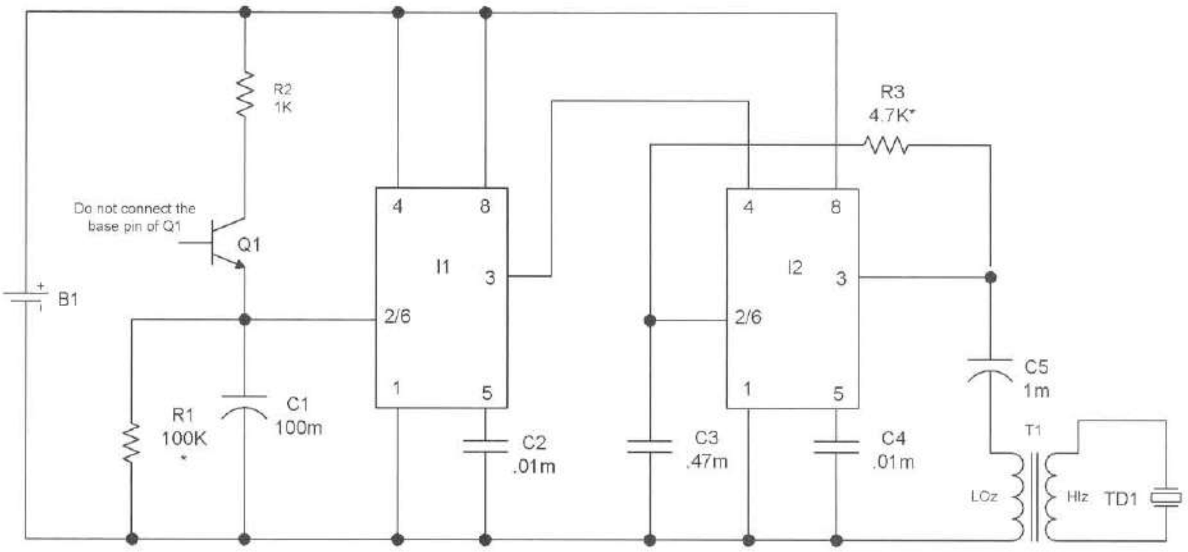
R2
1K
Do not connect the base pin of Q1
Q1
B1
R1
100K
*
C1
100m
4
8
I1
3
2/6
1
5
C2
.01m
R3
4.7K*
I2
C3
.47m
C4
.01m
C5
1m
T1
LOz
HIz
TD1

*The value of R1 May be changed to adjust for dark and light conditions. You may want to replace with a 500 K trimpot

Figure 31-2 *Circuit schematic*

Construction Steps

1. Identify all the parts and pieces and verify them with the bill of materials.
2. Insert the components starting from one end of the perforated circuit board and follow the locations shown in Figure 31-3 using the individual holes as guides. Use the leads of the actual components as the connection runs, which are indicated by the dashed lines. It is a good idea to trial-fit the larger parts before actually starting to solder. Always avoid bare wire bridges, globby solder joints, and potential solder shorts. Also check for cold or loose solder joints.

Additionally, pay attention to the polarity of the capacitors with polarity signs and all the semiconductors. The transformer position is determined using an ohmmeter.

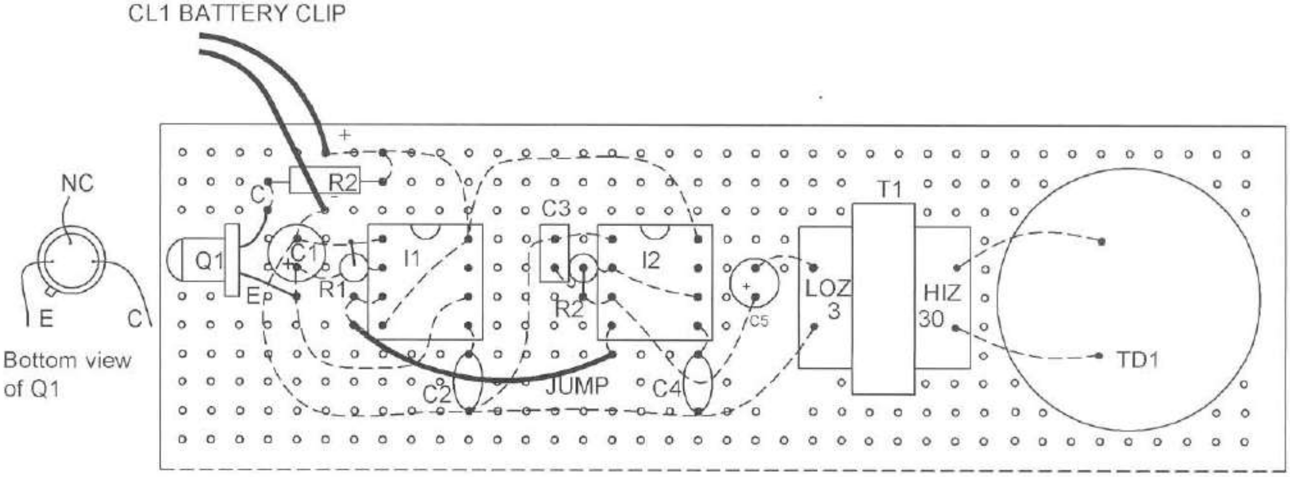
CL1 BATTERY CLIP
+
NC
E
C
Bottom view of Q1
Q1
E
C1
R2
C1
R1
I1
C3
R2
I2
C5
LOZ
3
T1
HIZ
30
TD1
C2
JUMP
C4

You can determine the low impedance side of T1 with an ohmeter. It will always read as the lower of the two

Figure 31-3 *Assembly board*

Testing

Verify the wiring accuracy and connection. Insert a 9-volt battery into the clip, cover the lens of Q1 or turn off the lights, and note the unit emitting a low-level, raspy, clicking tone like a sick cricket. Uncover the lens of Q1 and note the circuit shutting down. There will be a delay that is dependent on the level of light and dark. You may vary the response by changing the value of R1. Decreasing its value will require more light to shut down and vice versa.

Table 31-1 Mind Mangler Part List

Ref. #	*Description*	*DB Part #*
R1	100K, ¼-watt resistor (br-blk-yel)	
R2	1K, ¼-watt resistor (br-blk-yel)	
R3	4.7K, ¼-watt resistor (yel-pur-red)	
C1	100-microfarad, 25-volt vertical electrolytic capacitor	
C2, 4	Two .01-microfarad, 50-volt plastic capacitors (103)	
C3	.47-microfarad, 50-volt vertical electrolytic capacitor	
C5	1-microfarad, 50-volt vertical electrolytic capacitor	
I1, 2	Two 555 timer dual inline packages (DIP)	
Q1	L14P phototransistor	
CL1	9-volt battery clip	
TD1	Small piezo transducer	
T1	Small audio transformer, 8 to 1K	
PB1	4 × 1.5 × .1-inch grid perforated circuit board	

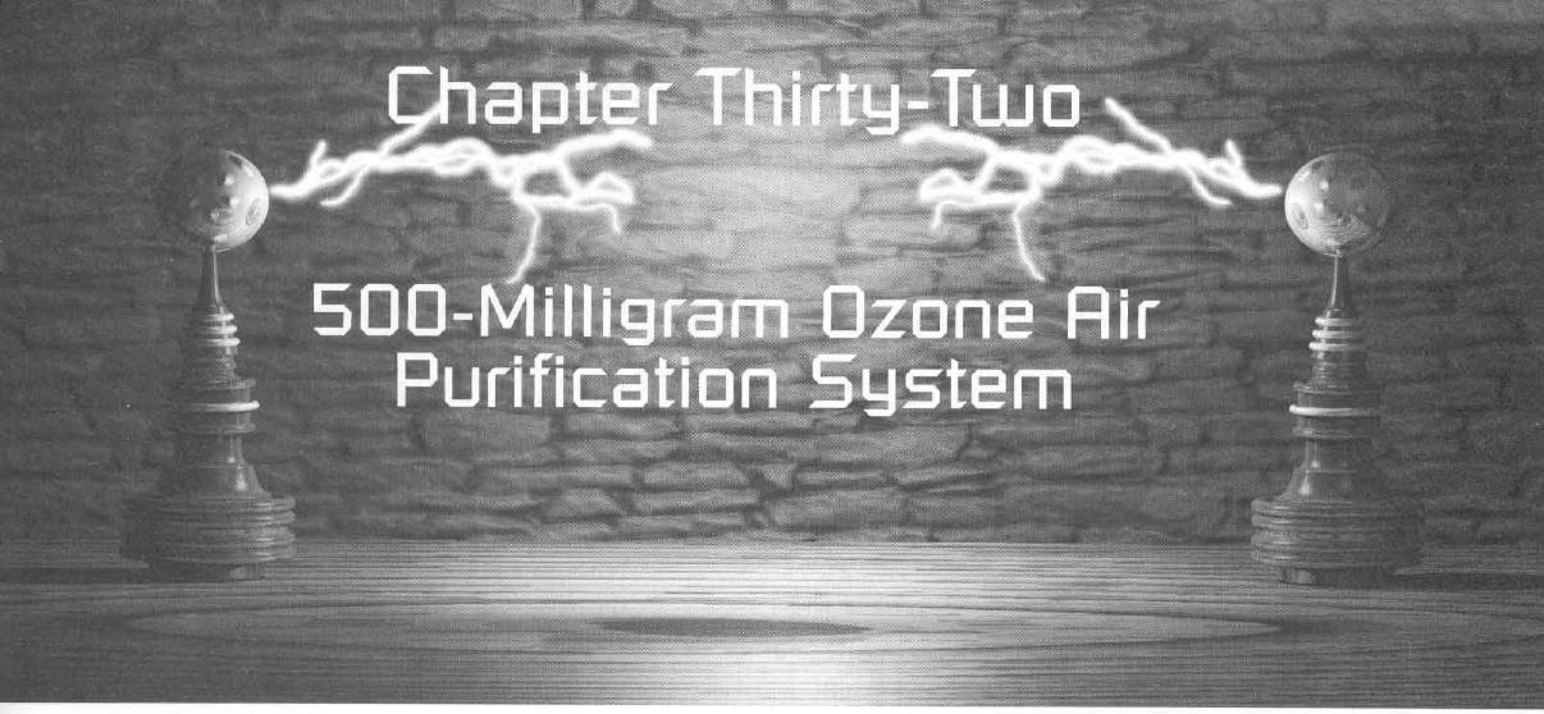

Chapter Thirty-Two

500-Milligram Ozone Air Purification System

Figure 32-1 shows a useful home device that generates ozone for eliminating odors, killing bacteria and mold, and disinfecting unsanitary areas such as pet boxes and bathrooms.

This project shows how you can build a system that could cost over $300 for less than $50. Its simple and basic design uses two 12-volt fans to blow air over a unique ozone-producing cell that is powered by a variable-output, high-voltage, high-frequency power supply. The power supply is featured in Chapter 7 of this book and is used in several other projects.

This is a beginner- to intermediate-level project requiring minimal basic electronic skills. Expect to spend $40 to $60. All the parts are readily available, with specialized parts obtainable through Information Unlimited (www.amazing1.com), and they are listed in Tables 32-1 and 7-1 at the end of this chapter and Chapter 7.

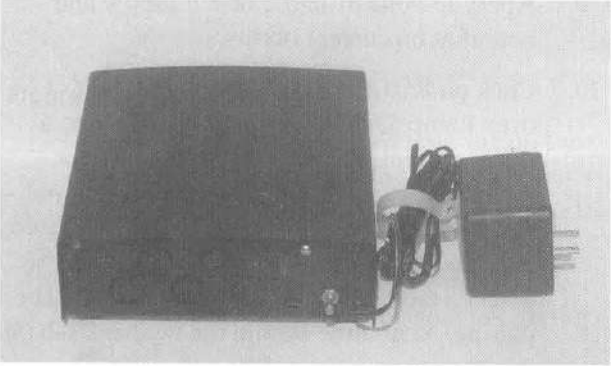

Figure 32-1 *Ozone purification system*

The Story of Ozone

Ozone is an unstable form of oxygen. Normal oxygen is diatomic (O_2), existing as two atoms of oxygen making up the molecule. Ozone is triatomic (O_3), existing as three atoms of oxygen for the molecule. The triatomic form of oxygen is very unstable, wanting to lose the third oxygen atom and combine with whatever atom it can (oxidization). This property makes it the most active oxidizer known, with the exception of the very hazardous fluorine gas. Ozone at normal pressures is colorless and odorless, yet it produces a pleasing, fresh-air odor as a result of the nitrous oxides it produces. The odor is usually noticeable after an old-fashioned thunderstorm. Under pressure, it becomes a bluish gas.

Ozone is also a very powerful bactericide. It is not affected by pH, as is chlorine, thus making it an excellent candidate for pools, spas, laundries, and general water treatment applications. It is many times more soluble in water, further enhancing its purifying effect. Ozone will combine with diatomic nitrogen, N_2, forming nitrous oxide, 2NO. This oxide quickly combines with water, forming nitric acid, HNO_3. This is often a very undesirable effect when used with straight air. Pure oxygen greatly minimizes this effect and is often required in many applications.

However, those requiring a supply of concentrated nitric acid for the nitration of many carbohydrates, such as cellulose, glycerine, hexamine, phenol, and toluene, may wish to consider combining ozone and air with a condensing apparatus to obtain this useful acid.

Ozone for Treating Air

Air pollution is one of the most serious environmental issues that we face. Visible and detectable smoke, dust, mildew, mold, and toxic odors, bacteria, pollen, and static electricity, along with the more elusive and invisible chemicals, have become a serious health threat. As buildings and homes are constructed with tighter air sealants for energy conservation, the problem will become even more severe. Often the effects from these problems include burning eyes, headaches, dizziness, depression, allergies, and general lethargy, all of which are usually attributed to colds and viruses. Air filters, such as the overpriced Sharper Image unit currently being advertised, are only partially effective, with some producing undesirable positive ions.

Ozone purifies the air from these undesirable pollutants by oxidization, as most pollutants readily combine with this highly reactive tri-atomic form of oxygen and break down into water and other non-toxic compounds. Ozone is produced artificially by electricity and is many times more antiseptically effective than oxygen.

Construction Steps

1. The system is shown using the high-voltage, high-frequency driver shown in Chapter 7. It is built on a printed circuit board or a more difficult perforated board. The PCB version is the easiest, as construction involves placing the correct components into the correct holes and soldering. The more difficult approach is also shown and uses a perforated circuit board. This is far more challenging and is intended for the experienced assembler. Figure 32-2 shows the board with the necessary changes for use with this circuit.
2. Create the base section, as shown in Figure 32-3. You can use sheet aluminum or bendable plastic sheets. Design a mating cover with a slight overhang over the front and rear panels. Use similar material and mount using #6 sheet metal screws along the lip of the base.
3. Fabricate the fan and cell mounting bracket as shown in Figure 32-4. Use clear 1/16-inch polycarbonate plastic and initially layout the piece, marking off the bend and cutting lines. You may fabricate separate pieces for the cell brackets and fans if you do not have access to the necessary equipment. Note dimensions and cuts must result in a single piece, as shown.
4. Cut out a piece of window screen large enough to cover the input air holes and secure it from the inside of the base enclosure using screws and silicon rubber cement.
5. Mechanically assemble the components, as shown in Figure 32-5. Verify all proper fittings and the clearance of components, especially the high-voltage points on the cell. Finally, wire everything as shown and verify the accuracy and any potential errors. Note: Figure 32-6 shows the wiring using the perforated board approach. Do not connect the 12 DC/1.5 adapter wall transformer at this time.
6. Obtain a 12-volt, 3-amp bench power supply with a voltage and current meter. Connect it to where the adapter's plus and minus wires go.
7. Connect a scope to the test points, as shown in Figure 32-5.
8. Preset the trimpot R1 to midrange and control pot R10 to full counterclockwise and click off.
9. Apply 12 volts from the bench supply and note that no current occurs.
10. Click on R10 and note a pulsing current of over 1 amp. Quickly adjust the trimpot to a dip in current, as noted on the meter. Our model was tuned correctly when the trimpot was set to one to two o'clock. The fans should both be in full operation and you should detect a faint, bluish glow coinciding with the pulsing. At no time should the heatsink tab on the power transistor Q1 be hot.

It will be necessary to construct the printed circuit assembly board as shown in chapter 7 with the following below list of revisions.

Wind L1 using 10 turns of #22 magnet wire. Air gap core with .007 (7mils) shims. Measured inductance should be between 15 -16 micro-henries.

Trimpot R6 is not used

Replace R5 with a wire jump

Replace C2 from a .0022 to a .0047 microfarad capacitor

Replace C6 from a .22 to a .47 microfarad 250 volt capacitor

Capacitor C4 is not used. insert a short piece of buss wire for a test point ground

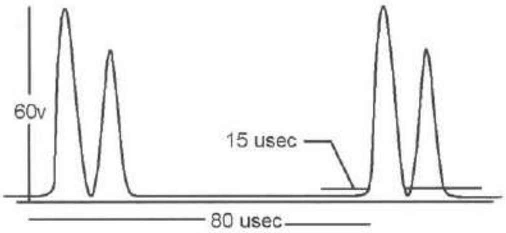

Correct wave shape measured at the drain of Q1 and test ground point

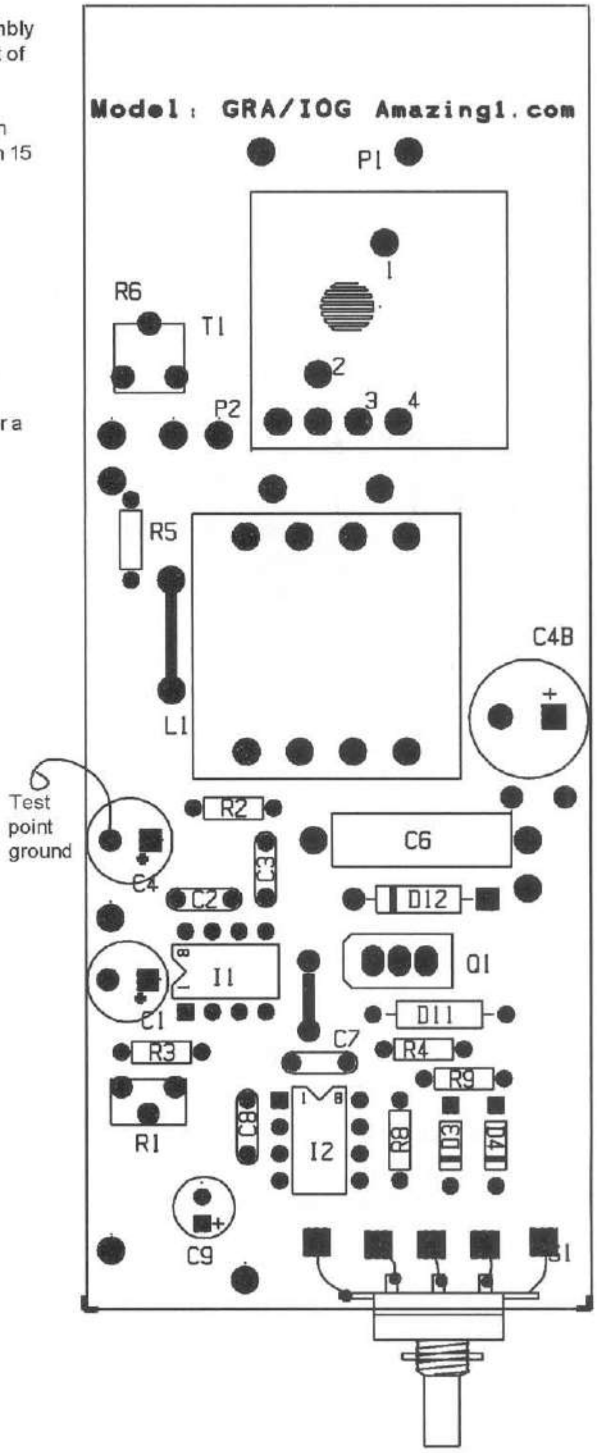
Model : GRA/IOG Amazing1.com
P1
1
2
3
4
R6
T1
P2
R5
L1
C4B
+
Test point ground
R2
C3
C6
C4
C2
D12
8
I1
Q1
C1
D11
R3
C7
R4
R9
1
8
C8
I2
R8
D13
D14
R1
+
C9
S1

Figure 32-2 *PCB rework*

11. Turn R10 full clockwise and note the on time of the pulse increasing with the current but not exceeding 1.5 amps at these pulse peaks. Allow it to operate for 30 minutes and note ozone being produced, a soft, bluish, even glow on the cell, and the heatsink not getting too hot to touch.

12. Once this is verified, you can hook in the 12 VDC/1.5-amp adapter transformer. Allow it to operate and verify that the adapter does not get hot.

13. Attach the cover and place the unit in the target area.

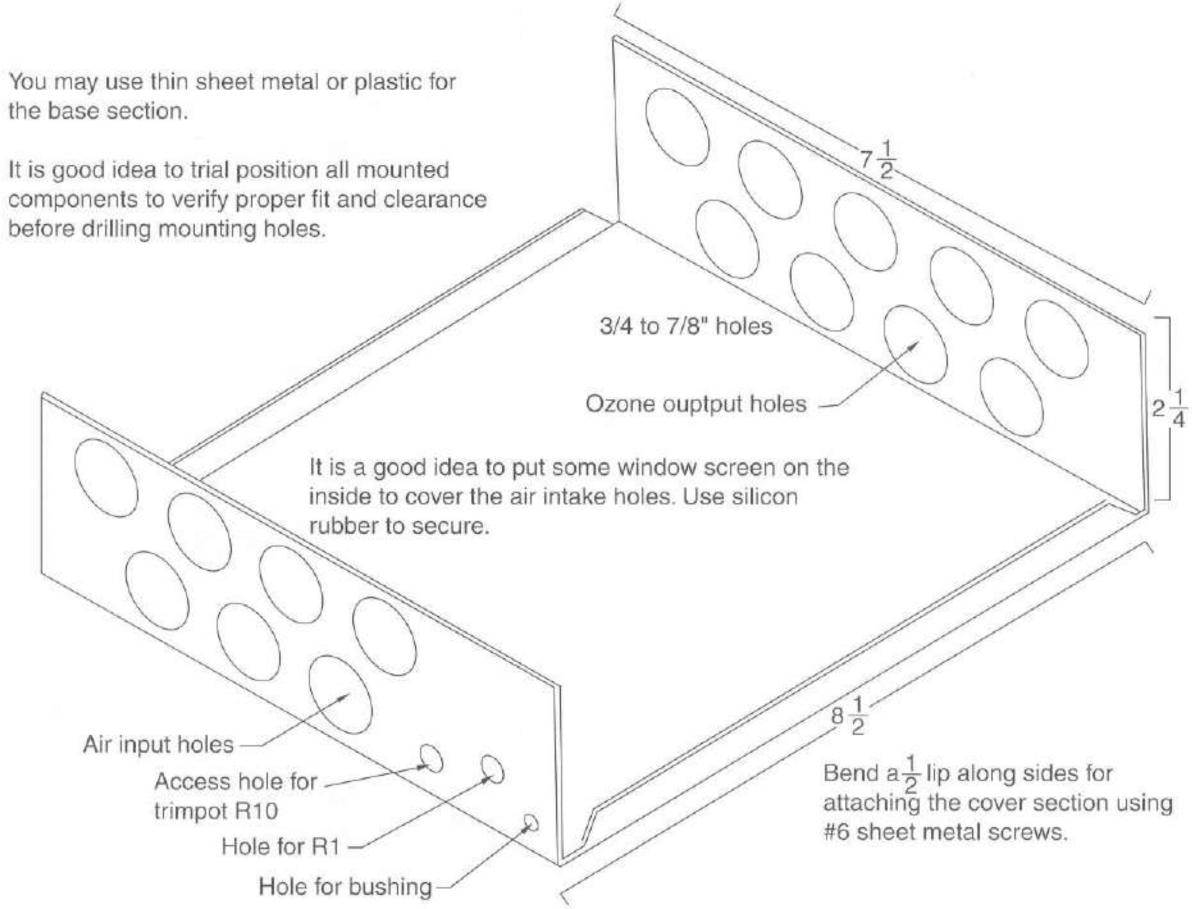
You may use thin sheet metal or plastic for the base section.
It is good idea to trial position all mounted components to verify proper fit and clearance before drilling mounting holes.
7 1/2
3/4 to 7/8" holes
Ozone ouptput holes
2 1/4
It is a good idea to put some window screen on the inside to cover the air intake holes. Use silicon rubber to secure.
8 1/2
Air input holes
Access hole for trimpot R10
Hole for R1
Hole for bushing
Bend a 1/2 lip along sides for attaching the cover section using #6 sheet metal screws.

Figure 32-3 *Base section creation*

14. Place the unit with the output directed toward the center of the area.
15. Plug in the wall adapter and rotate the control switch to on. Note the fan rotating and freely turning without obstructions. It should produce a good air flow.
16. Adjust the control a full clockwise turn and note a smell of fresh air emitting from output ports. You will observe a purplish glow around the cell when viewed in darkness.
17. Allow it to run for several hours and adjust the control to where the fresh air smell is just detectable.
18. Check both the unit and the wall adapter for excessive heating. They should only be warm to the touch.
19. This unit may need cleaning from dust. Use compressed dry air or a soft cleaning brush.

Special Notes

You should never operate the unit when the odor becomes pungent or domineering. Ozone is a colorless, odorless gas composed of unstable diatomic oxygen. The smell of fresh air is not the ozone, but the result of secondary reactions with other chemicals.

The unit can be adapted to 12 VDC operation for vehicle or battery use. Simply remove the adapter and splice in some 12-volt connections. Use caution when observing the polarity. You may request a factory modification for this step.

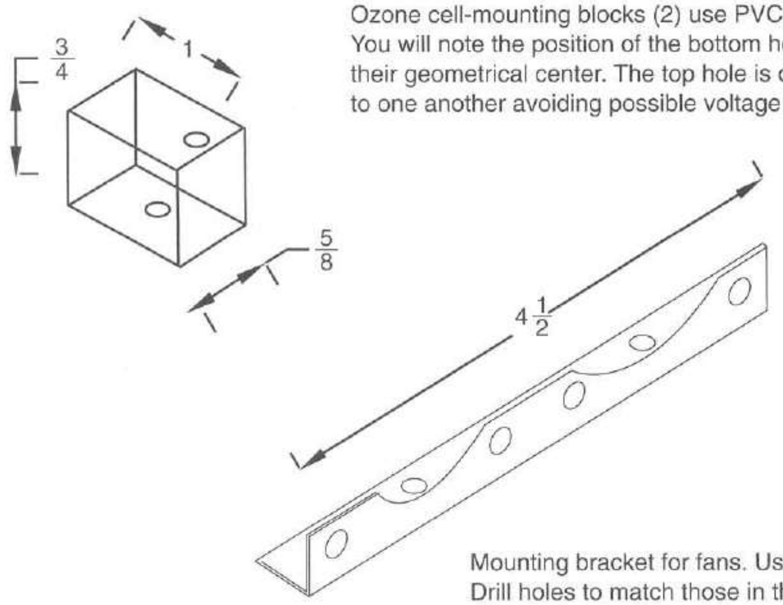

Ozone cell-mounting blocks (2) use PVC or another suitable insulting plastic. You will note the position of the bottom hole mounting the blocks of the base is at their geometrical center. The top hole is offset to this location so as not to be adjacent to one another avoiding possible voltage breakdown.

Mounting bracket for fans. Use $\frac{1}{32}$ aluminum or plastic. Drill holes to match those in the fan assemblies.

Figure 32-4 *Fabricated parts*

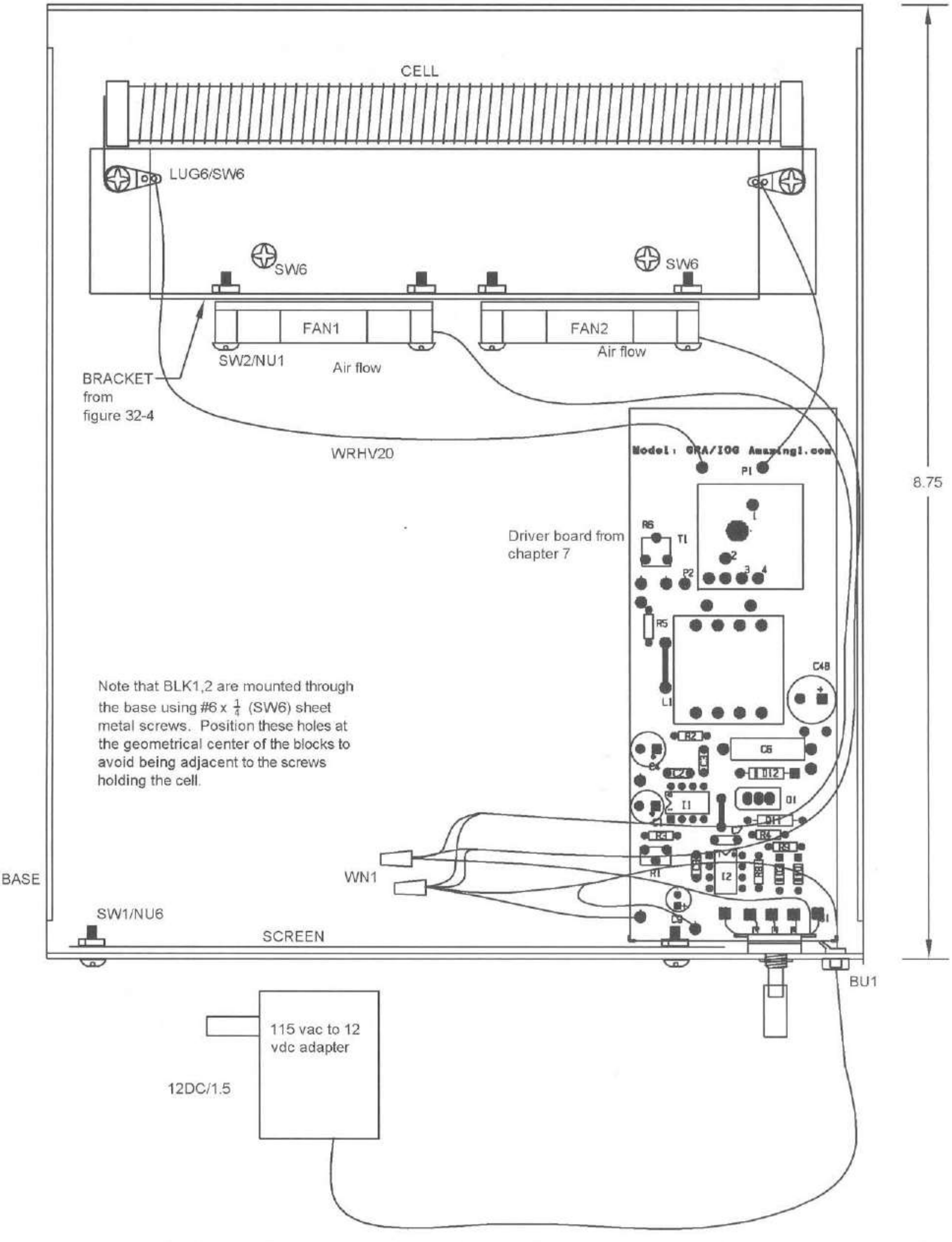
CELL
LUG6/SW6
SW6
SW6
FAN1
FAN2
Air flow
Air flow
SW2/NU1
BRACKET from figure 32-4
WRHV20
8.75
Model: GRA/IOG Amazing1.com
P1
R6
T1
1
2
3
4
P2
R5
L1
C48
Driver board from chapter 7
R2
C6
C4
IC2
D12
I1
Q1
D11
R4
R3
R9
R1
I2
C9
WN1
BASE
SW1/NU6
SCREEN
BU1
115 vac to 12 vdc adapter
12DC/1.5

Note that BLK1,2 are mounted through the base using #6 x $\frac{1}{4}$ (SW6) sheet metal screws. Position these holes at the geometrical center of the blocks to avoid being adjacent to the screws holding the cell.

Figure 32-5 *Mechanical layout showing wiring when using the PCB*

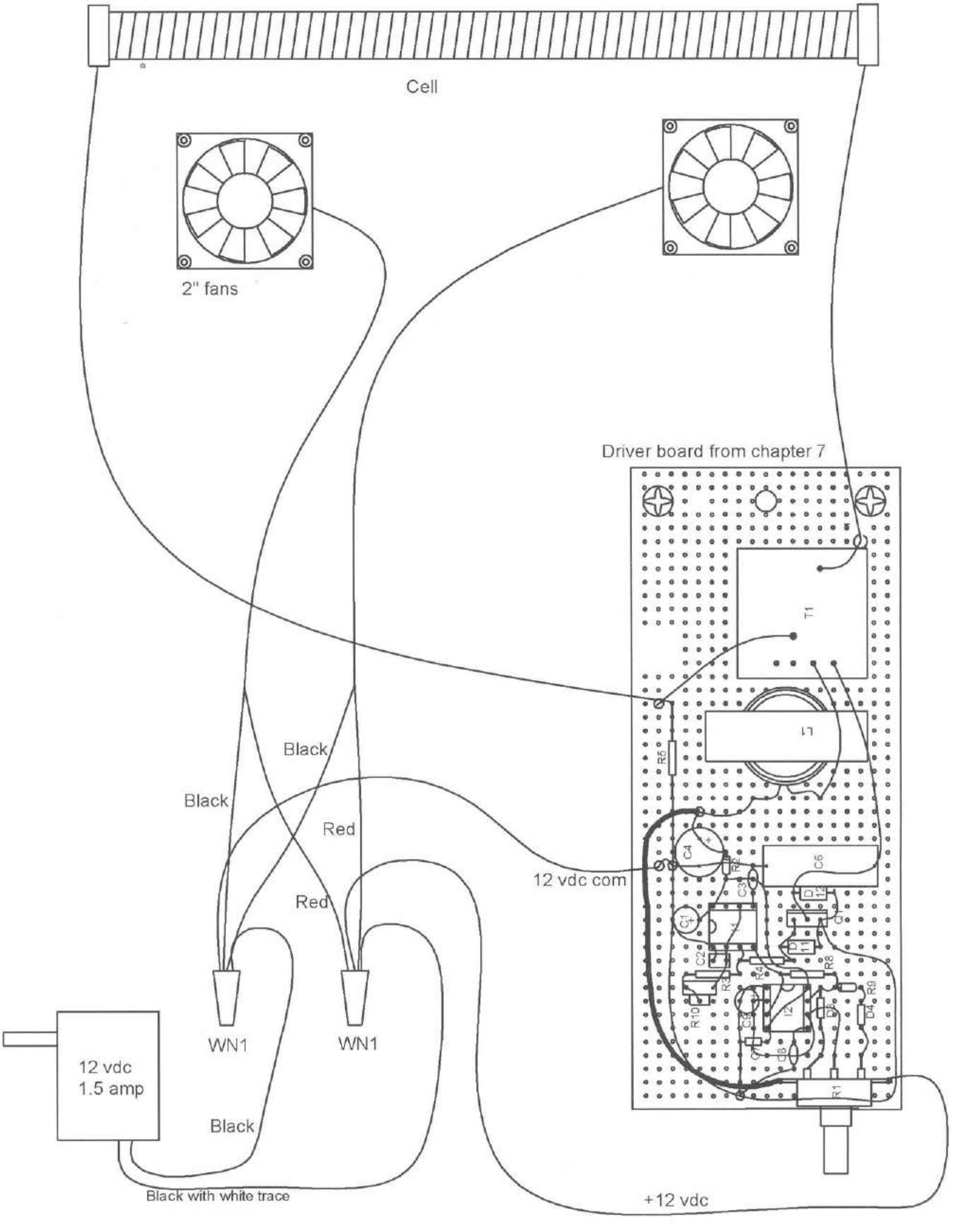
Cell
2" fans
Driver board from chapter 7
T1
L1
R6
Black
Black
Red
Red
12 vdc com
C4
R2
C6
C3
D12
Q1
C1
I1
D11
C2
R3
R4
R8
R9
R10
C9
I2
D3
D4
C7
C8
R1
WN1
WN1
12 vdc
1.5 amp
Black
Black with white trace
+12 vdc

Figure 32-6 *Final wiring using a perforated board*

Table 32-1 500-Milligram Ozone Machine Parts List

Ref. #	*Description*	*DB Part #*
DRIVER	Modified high-voltage assembly from Chapter 7	DB# GRADRIV10
BASE	Base section fabricated per Figure 32-3	
COVER	Fabricated to fit BASE section	
BLOCKS	Two blocks fabricated per Figure 32-4	
BRACKET	Bracket fabricated per Figure 32-4	
FAN1, 2	Two small, 2 × 2-inch, 12-volt fans	DB# FAN2
CELL500	500-milligram ozone corona cell	DB# CELL500
SCREEN	2 × 5-inch piece of window screen	
12DC/1.5	12-volt, 1.5-amp wall adapter transformer	DB# 12 DC/1.5
WRHV20	6-inch, 20-kilovolt, high-voltage wire	
WN1,2	Two small wire nuts	
BU1	Small Heyco bushing #2P-4	
SW6	10 #6 × ¼-inch sheet metal screws for cover and blocks	
SW2/NU2	Four #4-40 × ¾-inch machine screws and nuts for fans	
SW1/NU1	Two #6-32 × ½-inch machine screws	
LUG6	Two #6 solder lugs	
WR22	12 inches of #22 vinyl hookup wire	
FEET	Four stick-on rubber feet for base section	

Chapter Thirty-Three

Invisible Pain-Field Generator

The invisible pain-field generator device shown Figure 33-1 is a handheld, battery-operated sonic, shockwave generator that produces a *sound pressure level* (SPL) of up to 125 decibels (db) at 30 centimeters. The shock wave frequencies are user presettable at 25, 16, and 12 KHz. A sweep function is included in the circuitry where the selected center frequency varies between two set limits at an adjustable rate, providing a complex sonic signal that further enhances the effect.

Applications of the device can range from the routing out of agricultural pests in silage bins, chicken houses, grain bins, or wherever rats are a problem. The optimum frequency for this particular application is 16 KHz or just above that of human hearing. Many farmers use these devices for determining the effectiveness of pesticides, noting the reduction in infestation levels. Other applications include spooking birds and animals from unwanted areas and discouraging dogs and deer from decimating shrubs and ornamentals.

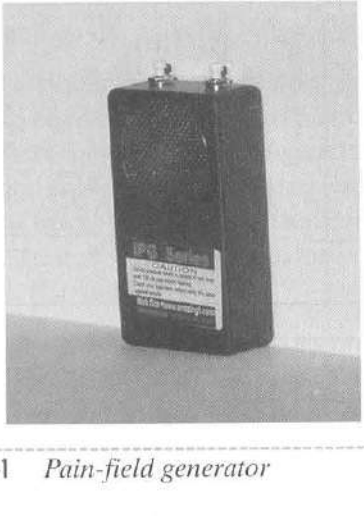

Figure 33-1 *Pain-field generator*

Adjustment of the frequency down to 12 KHz (within the range of human hearing) will produce extremely painful and annoying affects. The unit is excellent for disbursing crowds of potentially unruly people. A simple test is to point the device at an unsuspecting subject and momentarily push the button. You will notice a very positive response. Unfortunately maximum affect seems to favor younger women; older men seem to be less sensitive.

The unit is very directional when used outside, but it looses this property when used inside due to reflections from walls, ceilings, and furniture.

The unit is shown assembled into a small plastic enclosure that includes the electronics emitter transducer and the 9-volt battery. A removable plastic cover allows access in order to change the battery. Controls include a pushbutton emission control and sweep-activation pushbutton switches located on the top of the enclosure. The size of the main enclosure is 4 × 2 × 1 inches. The weight is only about 6 ounces with batteries installed

The unit can be used as a research tool for producing an effect on certain animals for their control and experimentation. The device has been successfully used for controlling certain dogs or other vicious animals by joggers or other outdoors enthusiasts. The unit has also been used in flushing out rats.

Caution: Caution must be used, as the effect on most people causes pain, headache, nausea, and extreme irritability. (Younger women are especially affected.) Do not, under any circumstances, point the unit at a person's ears or head at close range, as severe discomfort and possible ear damage may result. This also applies to dogs and other animals.

Include a caution label on the device to avoid exposure over 105 db for any continuous period of time. When using the device around people, you must be careful, as unjustified harassment is illegal and can result in prosecution.

This is an intermediate-level project requiring basic electronic skills. Expect to spend $25 to $50. All parts are readily available, and any specialized parts are available through Information Unlimited (www.amazing1.com) and are listed in Table 33-1 at the end of the chapter.

Circuit Description

A timer (IC2) is connected as an astable free-running multivibrator whose frequency is internally controlled by trimpot (R9). Resistor R8 selects the range limit of R9. The square-wave output of IC2 is via pin 3 and is directly coupled to power amplifier Q2. The drain of Q2 is DC biased through audio-frequency blocking chokes (L1A and L1B), providing a high impedance to the AC component of the signal.

DIP switches (S1,2,3) preselect the internal frequency for the required application and are shown in Figure 33-2. These switches select resonating inductors L2 and L3 connected in series with the output transducer TD1. The resonant action between the inherent capacity of TD1 and these selected inductances now produces a sinusoidal-shaped wave

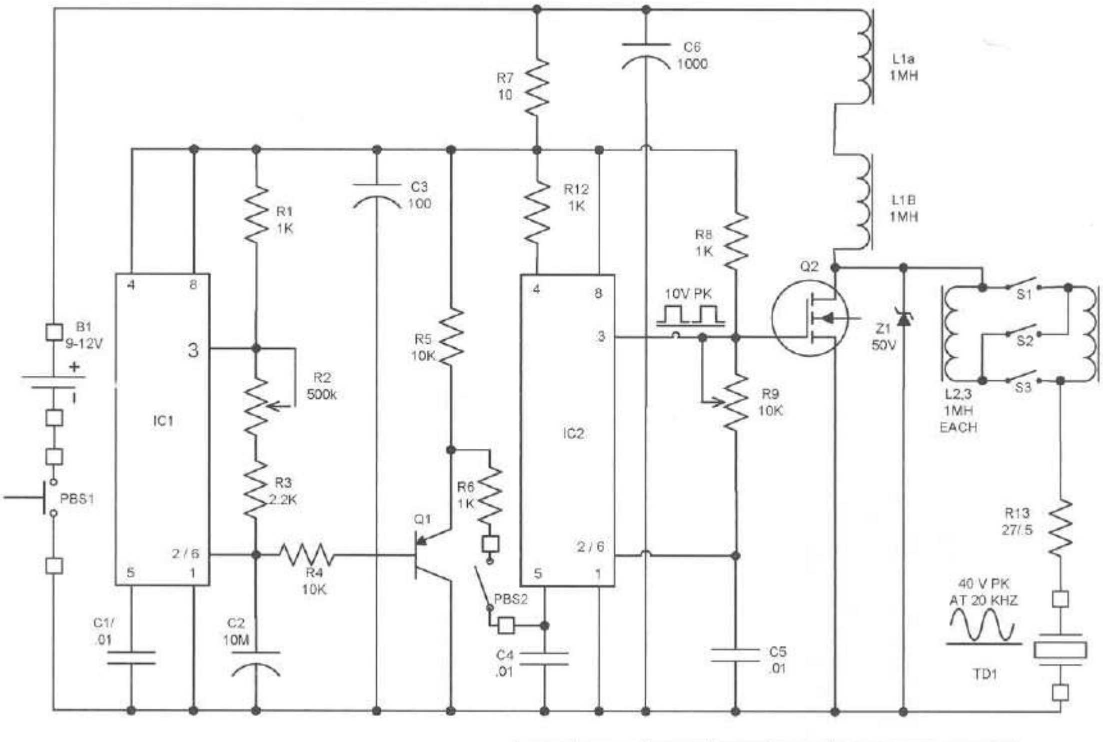

☐ These symbols represent external connection to off board components

ma	freq	1	2	3	usec
225	25khz	on	off	on	tune max 40-50 u
175	16khz	on	off	off	tune max 60-70 u
150	12khz	off	on	off	tune max 80-90u

Figure 33-2 *Pain-field generator circuit schematic*

peaking around the upper limit of the tuning range. This *signal* waveform now has a peak-to-peak voltage several times that of the *original* square wave. Transducer TD1 now can take advantage of these peak voltages to produce the high sound-pressure levels necessary without exceeding the high RMS ratings of an equivalent voltage-level square wave. Zener diode Z1 clips any excessive overshoots across Q2.

Timer IC1 is similarly connected as an astable running multivibrator and is used to produce the sweeping voltage necessary for modulating the frequency of IC2. This sweep repetition rate is controlled by trimpot R2. Resistor R3 limits the range of this repetition time. C2 sets the sweep time range. Output for IC1 is via pins 6 and 2, where the signal ramp function voltage is resistively coupled to inverter transistor Q1 via resistor R6. The output of Q1 is fed to pin 5 of IC2 and provides the output modulation voltage necessary to vary the frequency as required. Note that the modulation signal is enabled by pushbutton PBS2.

Power to the system is via battery B1 and pushbutton PBSl. Capacitor C6 guarantees an AC return path for the output signal. Power to the driver circuits IC1 and IC2 is thru a decoupling network consisting of resistor R7 and capacitor C3.

Construction

1. Layout and identify all parts and pieces. Verify with parts list. Separate resistors, as they have a color code to determine their value. Colors are noted on the parts list.
2. If you are building from a perforated board, use Figure 33-3 as a parts placement guide and insert components starting in the lower left-hand corner as shown. Pay attention to polarity of capacitors and all semiconductors. It is suggested you use sockets for integrated circuits IC1 and IC2.

   Route leads of components, as shown, and solder as you go, cutting away unused wires. Attempt to use certain leads as the wire runs or use pieces of the #24 buss wire. Follow dashed lines on assembly drawing, as these indicate connection runs on underside of assembly board. The heavy dashed lines indicate use of thicker #20 buss wire, as this is a high-current discharge path.

   Please note that this circuit is very cramped, and it is suggested you obtain the optional printed circuit board shown in the parts list.
3. Attach external leads to components as shown in Figure 33-4, noting the individual lengths and twisted pairs.
4. Double-check accuracy of wiring and quality of solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.
5. Connect an ohm meter between CLI contacts and pushbutton PBSl. Note a reading of several thousand ohms. This may vary but should not indicate a short circuit. Preset trimpots R2 and R7 to midrange.
6. Connect a 9-volt battery and note a current of 200 to 250 milliamps when R9 is set to midrange. The meter can be connected across PBSl for this step.
7. Simultaneously press both PBS2 and PBSI and note a piercing sweeping tone coming from the transducer TD1. Turn R9 fully ccw and note the decrease in signal tone.
8. Frequency range should be approximately 10 to 25 KHz, with a sweep of approximately 1 to 20 times per second selectable by trimpot R2.
9. Check the waveshapes as shown at pin 3 of the output timer IC2. This is determined by the following formula: $F = 1/(1.57 \times R9 \times C5)$.
10. Check the waveshapes as shown across transducer TD1. This waveshape is approximately 40 to 50 volts, peak to peak, and approaches a sine wave. This is the approximate resonant frequency point of the selected values of series inductances within the inherent capacitance of the transducer. A voltage-peaking effect will be noted, as the frequency is varied by trimpot R9. As the frequency is varied off of resonance, this wave will severely distort but will not damage the circuitry. Note the chart in Figure 33-2 showing the selection

It may be a good idea to use SOCK 1, 2 sockets for IC1, 2 to avoid unsoldering if you make an error.

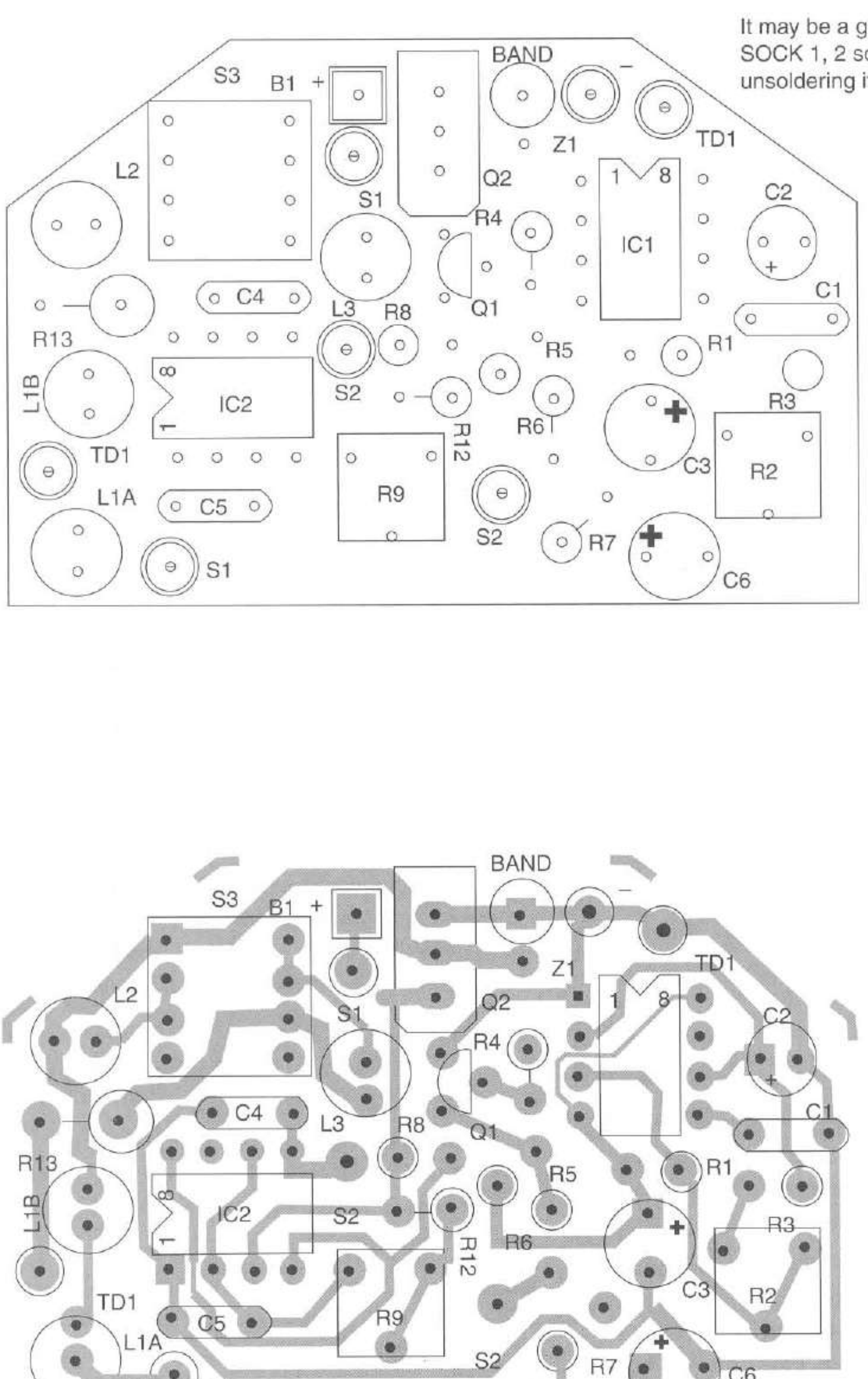
BAND
S3
B1 +
L2
Q2
S1
R4
Z1
TD1
IC1
C2
C1
C4
L3
R8
Q1
R13
R5
R1
L1B
IC2
S2
R12
R6
R3
C3
R2
TD1
R9
L1A
C5
S2
R7
C6
S1
1
8

Figure 33-3 *Printed circuit board parts and wiring*

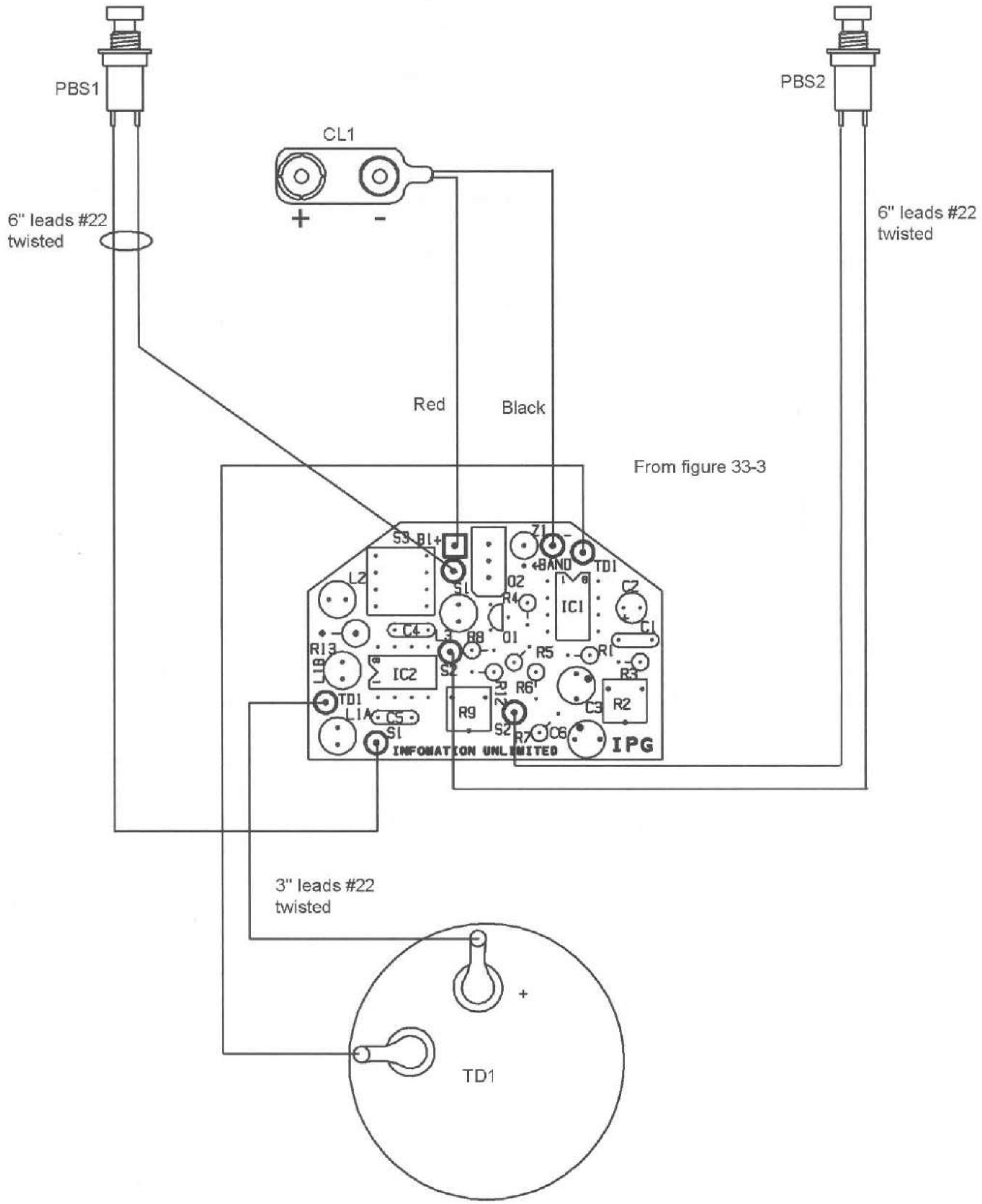
PBS1
CL1
PBS2
6" leads #22 twisted
+
-
6" leads #22 twisted
Red
Black
From figure 33-3
S3
B1+
Z1
-
+BAND
TD1
L2
S1
Q2
R4
IC1
C2
C4
L3
R8
Q1
R5
R1
C1
R13
L1B
IC2
S2
R6
R3
TD1
L1A
C5
R9
R12
S2
C3
R2
S1
R7
C6
INFOMATION UNLIMITED
IPG
3" leads #22 twisted
+
TD1

Figure 33-4 *External wiring*

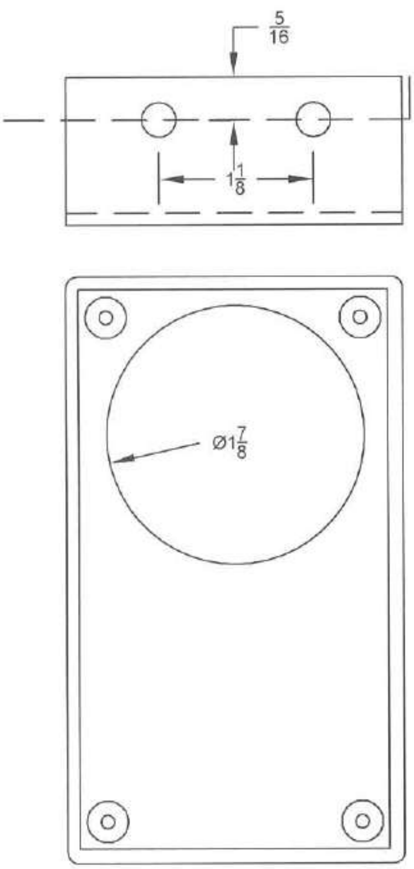

Operating instructions:

Remove rear cover and insert a 9 volt alkaline battery. Note that proper positioning is required for cover to properly fit without bowing.

Default settings of the two trimpots is midrange. To change for obtaining the maximum effect on the target you may adjust trimpots using a small plastic tuning tool or screw driver.

Simply direct unit transducer opening towards the target and push the control button. You should notice an immediate effect. The range will depend on the acoustical sensitivity of the target subject.

Use the sweep button to possible enhance the effect on the target.

This unit is designed for generating intermittent bursts of ultrasonic and high frequency acoustical energy.
Do not direct at people.
Use out side as walls and ceilings will cause the signal to loose directional characteristics and effect the user.
Use in 2 to 5 second bursts.

Figure 33-5 *Fabrication of plastic case*

of the inductances for the choice of operating frequencies that provide the maximum output.

Use a scope for these measurements. These steps are not necessary if the unit appears to function as described.

11. A sound pressure measurement of approximately 125 db was measured at 10 KHz at a distance of 16 inches on our model. Voltage across TD1 was measured at 40 volts peak.

This completes the electronic testing of the unit.

12. Fabricate the EN1 enclosure, as shown in Figure 33-5. Note $1\frac{7}{8}$-inch hole for the aperture of TDI. This hole should be made with a large punch or hole saw using a drill press. The hole may also be cut out with a sharp knife. Drill $\frac{1}{4}$-inch holes for PBSl and PBS2, as shown.
13. Cut a piece of $2\frac{1}{8} \times 2\frac{1}{8}$-inch window screen and secure it along with transducer TDI. Use RTV or suitable adhesive.
14. Make final assembly, as shown Figure 33-6, and secure assembly board with two-sided foam tape.
15. Attach the battery and position as shown. Note: The battery must be up against the enclosure with all leads properly routed or the rear cover will not fit properly.
16. See Figure 33-5 for operating instructions.

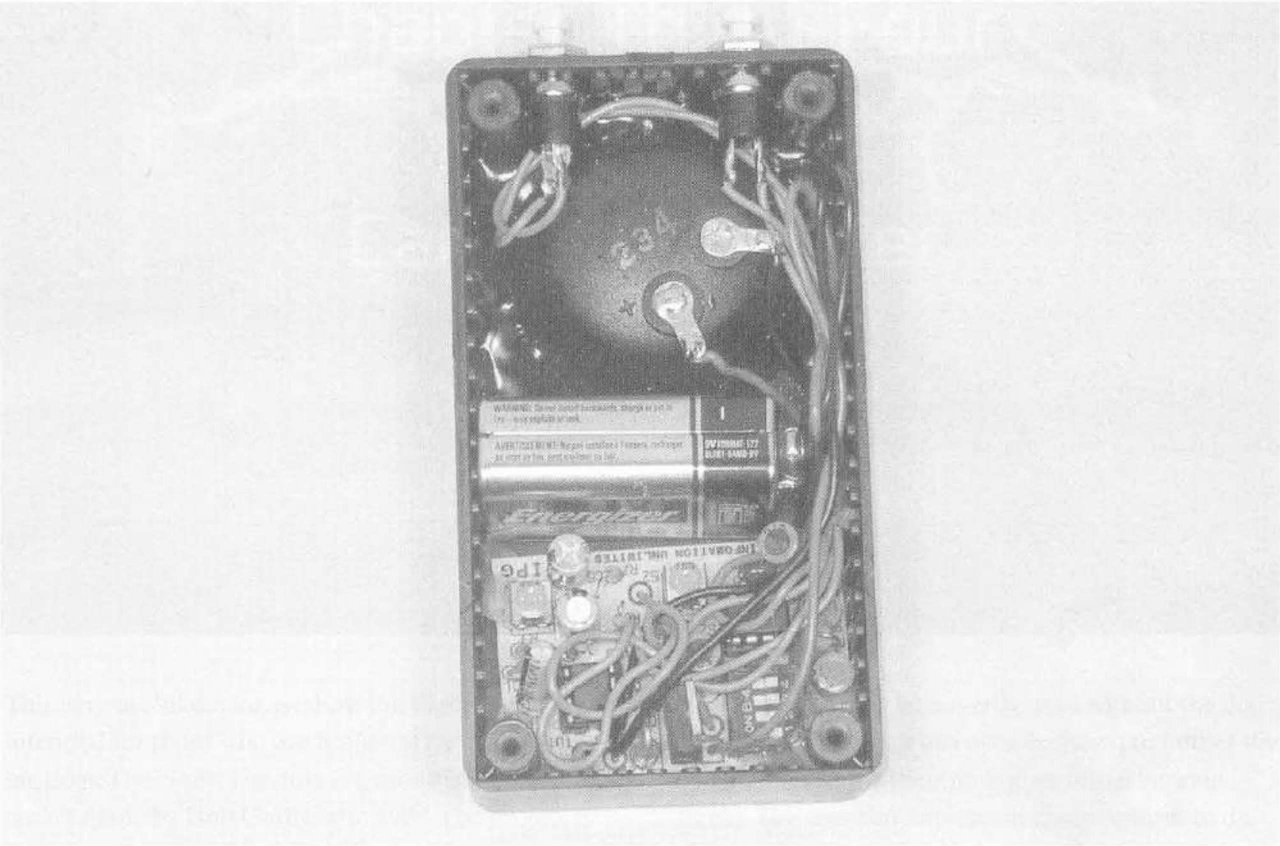

Figure 33-6 *Final assembly into EN1 enclosure*

Table 33-1 Invisible Pain-Field Generator Parts List

Ref#	*Description*	*DB Part #*
R1,6,8,12	Four 1 K, ¼-watt resistors (br-blk-red)	
R2	500 K vertical trimpot	
R3	2.2 K, ¼-watt resistor (red-red-red)	
R4,5	Two 10 K, ¼-watt resistors (br-blk-or)	
R7	10-ohm, ¼-watt (br-blk-blk)	
R9	10 K vertical trimpot	
R13	27-ohm, ½-watt resistor (red-pur-blk)	
C1,4	Two .01 mfd, 50-volt disc capacitors	
C2	10 mfd, 50-volt vertical electrolytic capacitor	
C5	.01 mfd, 50-volt plastic capacitor	
C3	100 mfd, 50-volt vertical electrolytic capacitor	
C6	1,000 mfd, 25-volt vertical electrolytic capacitor	
IC1,2	Two 555 timer DIP integrated circuits	
Q1	PN2907 PNP TO92 transistor	
SOCK1,2	Two 8-pin DIP sockets for above IC1,2 (Not shown)	
Q2	IRF540 Mosfet TO220	
L1AB,2,3	Four 1-millihenry inductors	DB# 1MH
Z1	50-volt, 1-watt Zener diode	
PBS1,2	Two pushbutton switches	
S1,2,3	3 or 4 section DIP switch	
CL1	Battery snap clip	
TD1	Special polarized transducer	DB# 1020A
PCB1	Printed circuit board	DB# PCIPG9
CASE	4 × 2⅜ × 1⅛-inch plastic enclosure	
SCREEN	2.5 × 2.5 inch piece of window screen	

Chapter Thirty-Four

Canine Controller

This very useful device, as shown in Figure 34-1, is intended for those who are bothered by nearby barking dogs. The control section is a modification of a circuit made by Bob Gaffigan in 1993. The project is intended to detect the dog's bark, which triggers a high-output, pain-field sonic generator, as described in Chapter 35, producing very uncomfortable sounds to the animal's sensitive hearing.

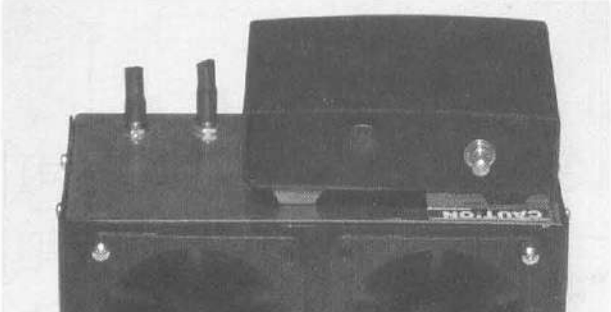

Figure 34-1 *The canine controller system with our sonic blaster*

The system may be covertly used without the dog owners knowing or it can even be tuned to bother the owners every time their mutt goes into a barking frenzy. The user can implement many options to discourage these constant barkers or give some of these inconsiderate owners a taste of their own medicine.

This is an intermediate-level project requiring basic electronic skills. Expect to spend $25 to $50. All the parts are readily available, with the specialized parts obtainable through Information Unlimited (www.amazing1.com), and they are listed in the parts list in Table 34-1 at the end of the chapter.

Device Description

A directional device picks up the dog's bark and processes it so that it turns on the sonic burst pain-field generator pointed in the offending animal's direction. The animal now experiences a very uncomfortable sound, much like a person would find the scratching of chalk on a blackboard. Eventually, the animal associates this sound with his barking, thus conditioning him to stop. The device contains adjustments that actually count the number of yelps so that it can be set not to trigger, thereby discouraging the animal from a normal barking trend that may be the warning of an intrusion or other important event.

The unit is easily mounted in any convenient location and operates from internal batteries, an external 12 volts, or 115 VAC via a wall adapter converter. The detection system contains adjustments for bark-level sensitivity, the number of barks detected before triggering, and the length of the triggered on time of the sonic burst. The pain-field generation section, as described in Chapter 35, has controls for the frequencies of the burst, sweep rate, sweep on/off, and main power control. The canine controller section is designed to easily connect to the existing jacks on the sonic generator section via a three-conductor cable. It may be mechanically attached as a single unit for a compact integrated system. The pickup microphone may even be placed at the focal point of a parabolic dish, providing very discrete selection of the target. As in any device such as this, many variables will enter into the equation that defines the overall performance. The suggested effective range will vary but can be up to 100 feet.

Special Note: In some circumstances, the owner is more to blame than the actual animal. As previously stated, the unit may be tuned to a frequency that the owner him- or herself finds very uncomfortable.

Circuit Description

Figure 34-2 shows the block diagram where the mic picks up the bark and amplifies it. A filter favors the assumed frequencies of the bark and feeds them to a bark counter. This circuit contains a selector switch to preset the necessary number of barks to trigger the bark period timer. The time of the bark period is also selectable by a switch. The bark period timer gates on the MOSFET switch transistor turns on the sonic generator, which drives the four transducers that now produce the directional, uncomfortable sound pain-field shockwaves.

Figure 34-3 shows the complete schematic of the controller section circuit, and the sonic pain-field generator section schematic is shown in Chapter 35. The bark is detected by microphone M1 and amplified by operational amplifier IC2A. The DC follower IC2D sets the midpoint bias, and resistor R2 sets the gain of this stage. The output level at this stage is controlled by potentiometer R5 and fed to amplifier/filter IC2C with a passband set by resistors R8 and R9 and capacitors C8 and C9. Schmidt discriminators

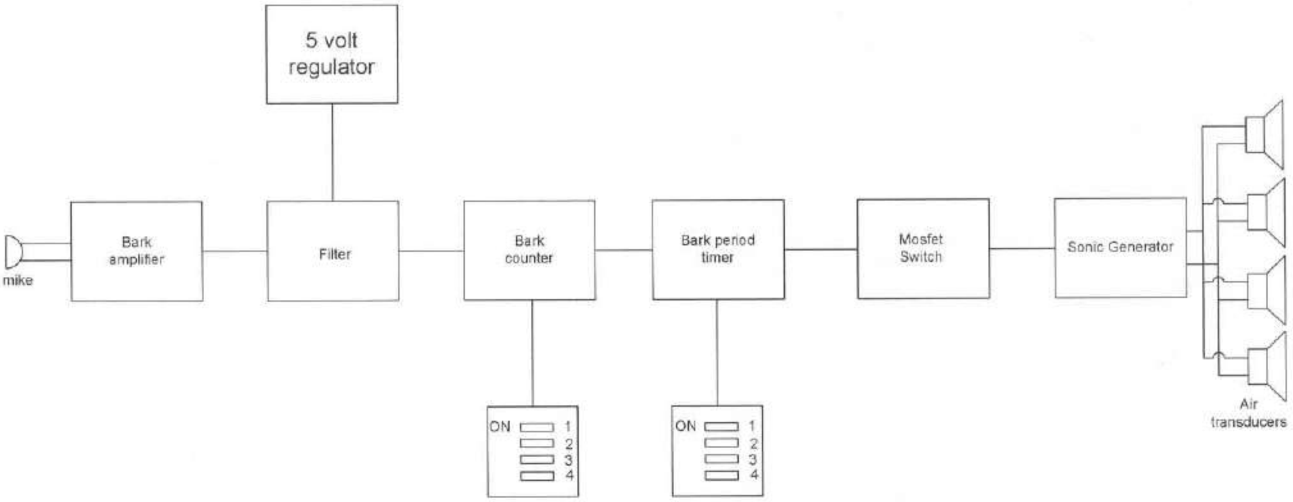

S1 determines the relative number of barks that must be detected before triggering the timer for turning on the sonic generator.
Pos1.........8
Pos2.........7
Pos3.........5
Pos4.........3

S2 determines the on time of the sonic generator once it triggers from the bark signal.
Pos1.......1sec
Pos2.......2sec
Pos3.......10sec
Pos4.......33sec
Combinations of the switch positions will be additive. Note the switch must always be on in one position.

Figure 34-2 *Circuit block diagram*

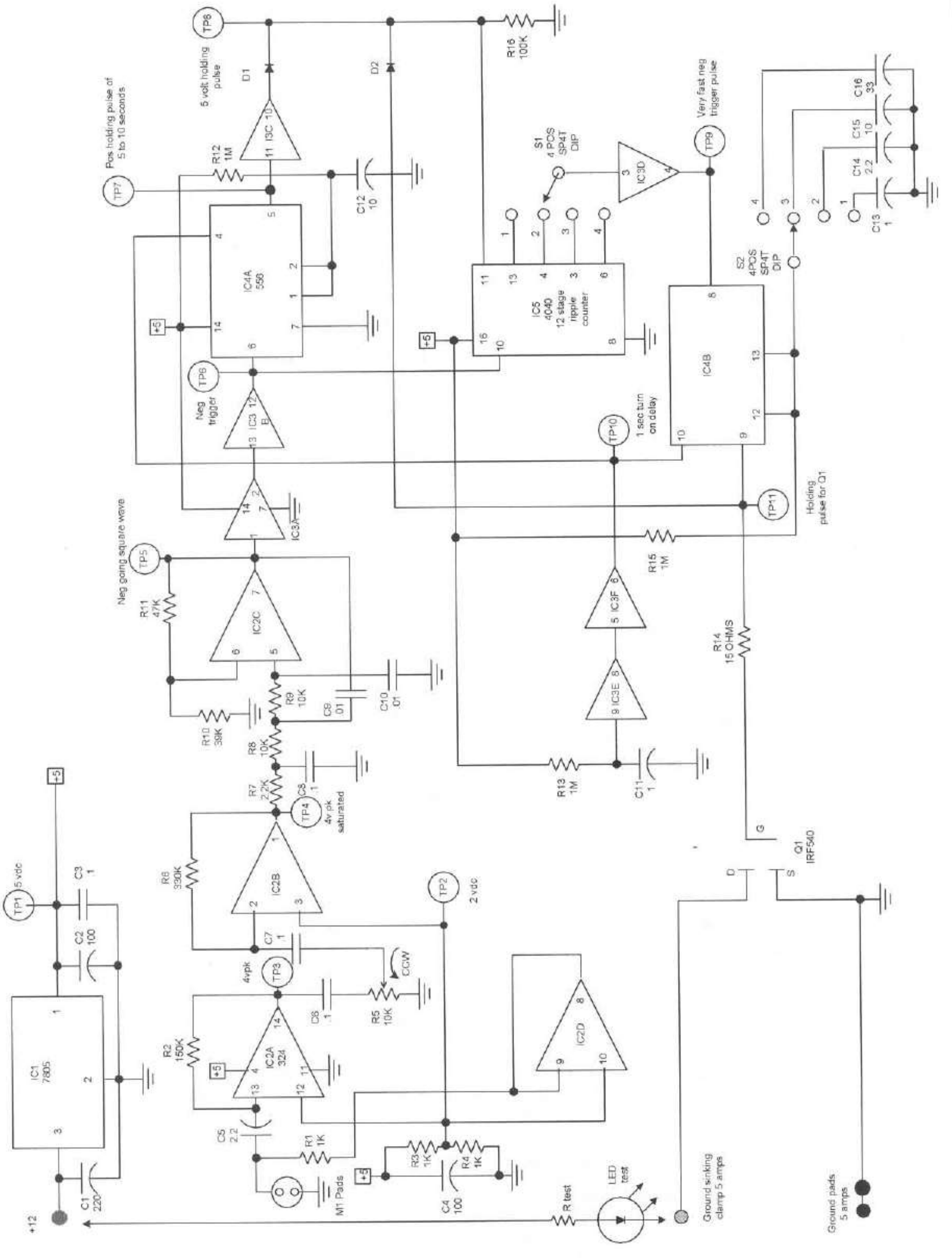
+12
C1 220
IC1 7805
TP1 5 vdc
C2 100
C3 1
+5
M1 Pads
C5 2.2
R1 1K
R2 150K
IC2A 324
TP3 4vpk
C6 .1
R5 10K
CCW
C7 .1
R6 330K
IC2B
TP2 2 vdc
R3 1K
R4 1K
C4 100
IC2D
R test
LED test
Ground sinking clamp 5 amps
Ground pads 5 amps
TP4 4v pk saturated
R7 2.2K
C8 .1
R8 10K
R9 10K
R10 39K
C9 .01
C10 .01
IC2C
R11 47K
TP5 Neg going square wave
IC3A
Neg trigger
IC3 B
TP6
IC4A 556
TP7 Pos holding pulse of 5 to 10 seconds
R12 1M
C12 10
I3C
5 volt holding pulse
D1
D2
TP8
R16 100K
IC5 4040 12 stage ripple counter
S1 4 POS SP4T DIP
IC3D
TP9 Very fast neg trigger pulse
TP10 1 sec turn on delay
IC4B
S2 4POS SP4T DIP
C13 1
C14 2.2
C15 10
C16 33
TP11 Holding pulse for Q1
R15 1M
IC3F
IC3E
R14 15 OHMS
R13 1M
C11 1
Q1 IRF540
G
D
S

Figure 34-3 Circuit schematic

IC3A and IC3B trigger the signal, provide a defined level of the correct polarity to initiate timer IC4A, and enable ripple counter IC5 to count the barks within a time period determined by IC4A. If the number of barks exceeds a limit set by bark count switch S1, timer IC4B is initiated, turning on MOSFET switch Q1 for a predetermined amount of time selected by switch S2, enabling the sonic pain-field deterrent signal to be sent toward the target and then resetting. You will note that control of the sonic pain-field generator is made by sinking the negative return lead into the ground.

Circuit Assembly

1. Assemble the circuit board as shown in Figure 34-4. Note that Figure 34-5 shows the foil traces of the printed circuit. It is not recommended that you assemble this circuit on a perforated vector board unless you are an advanced assembler.

   Note that if you are building from a perforated or vector circuit board, it is suggested that you use the indicated traces for the wire runs and insert components starting at the lower left-hand corner. Pay attention to the polarity of the capacitors with polarity signs and all the semiconductors. It is a good idea to use sockets for all the integrated circuits.

   Route the leads of the components as shown and solder as you go, cutting away unused wires. Attempt to use certain leads as the wire runs or use pieces of the #26 bus wire. The heavy foil runs should use the thicker #24 bus wire, as these are the high-current discharge paths.

2. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire

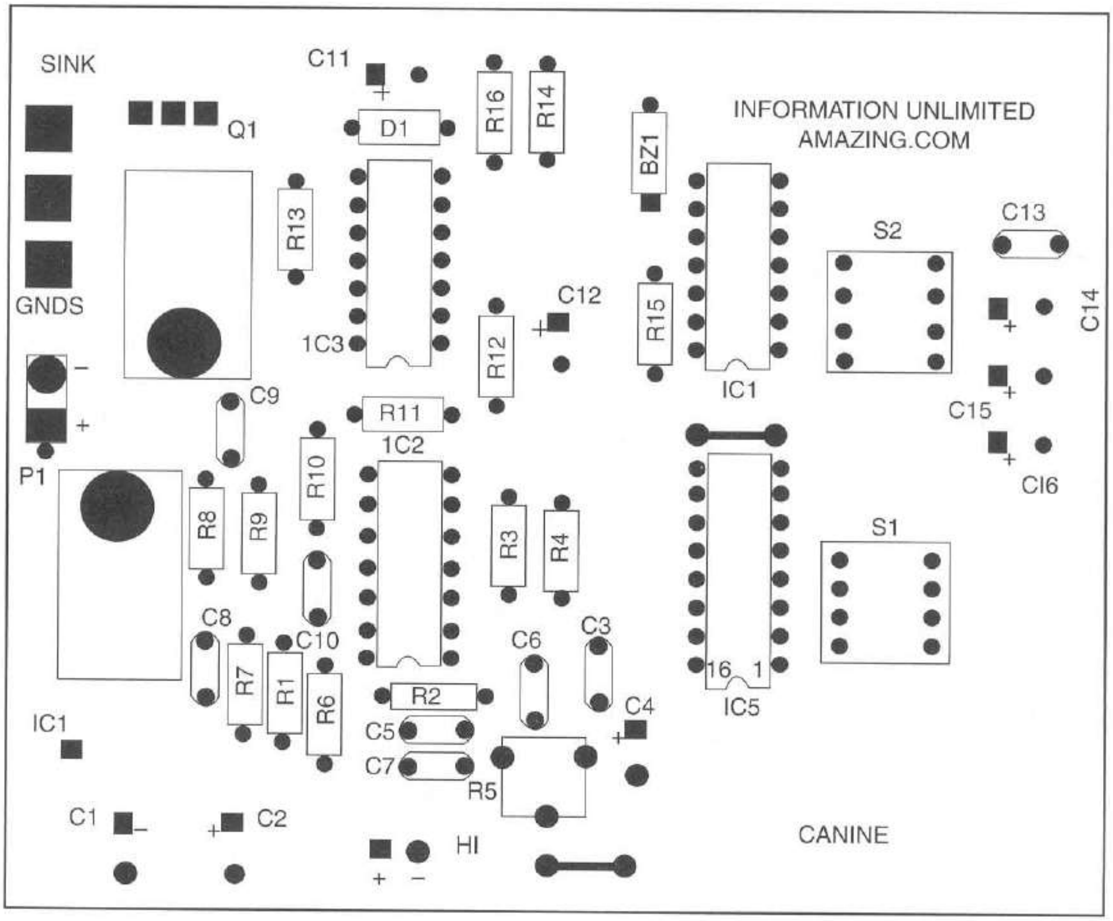
SINK
C11
Q1
D1
R16
R14
BZ1
INFORMATION UNLIMITED
AMAZING.COM
R13
S2
C13
GNDS
C12
R15
C14
1C3
R12
-
+
P1
C9
R11
IC1
C15
1C2
R10
R8
R9
R3
R4
CI6
S1
C8
C10
C6
C3
R7
R1
R6
R2
C4
16
1
IC5
IC1
C5
C7
R5
C1
C2
HI
+
-
CANINE

Figure 34-4 *Assembly of the circuit board*

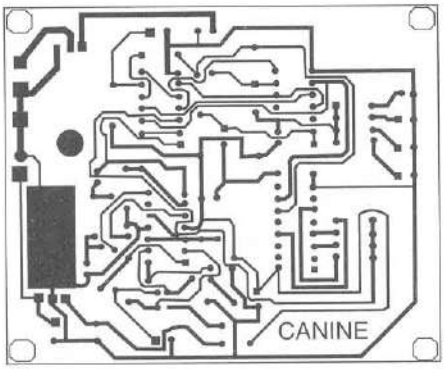

Figure 34-5 *Assembly circuit board traces*

bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.

3. Connect the external components as shown in Figure 34-6. Use shielded microphone cable if you are not installing it on the actual board.
4. Fabricate the chassis as shown in Figure 34-7 from a piece of .035-inch 5052 bendable aluminum or plastic. If you use metal, you will need a piece of plastic material under the assembly board to prevent shorting the foil traces.
5. Create a mating cover and finally assemble everything as shown in Figure 34-8. Note the hole for the microphone with a screen. Wire the P1 plug and cable for interconnecting the controller and the sonic blaster using desired lengths per your requirements. Leave the other end with leads that will eventually connect to the sonic blaster.

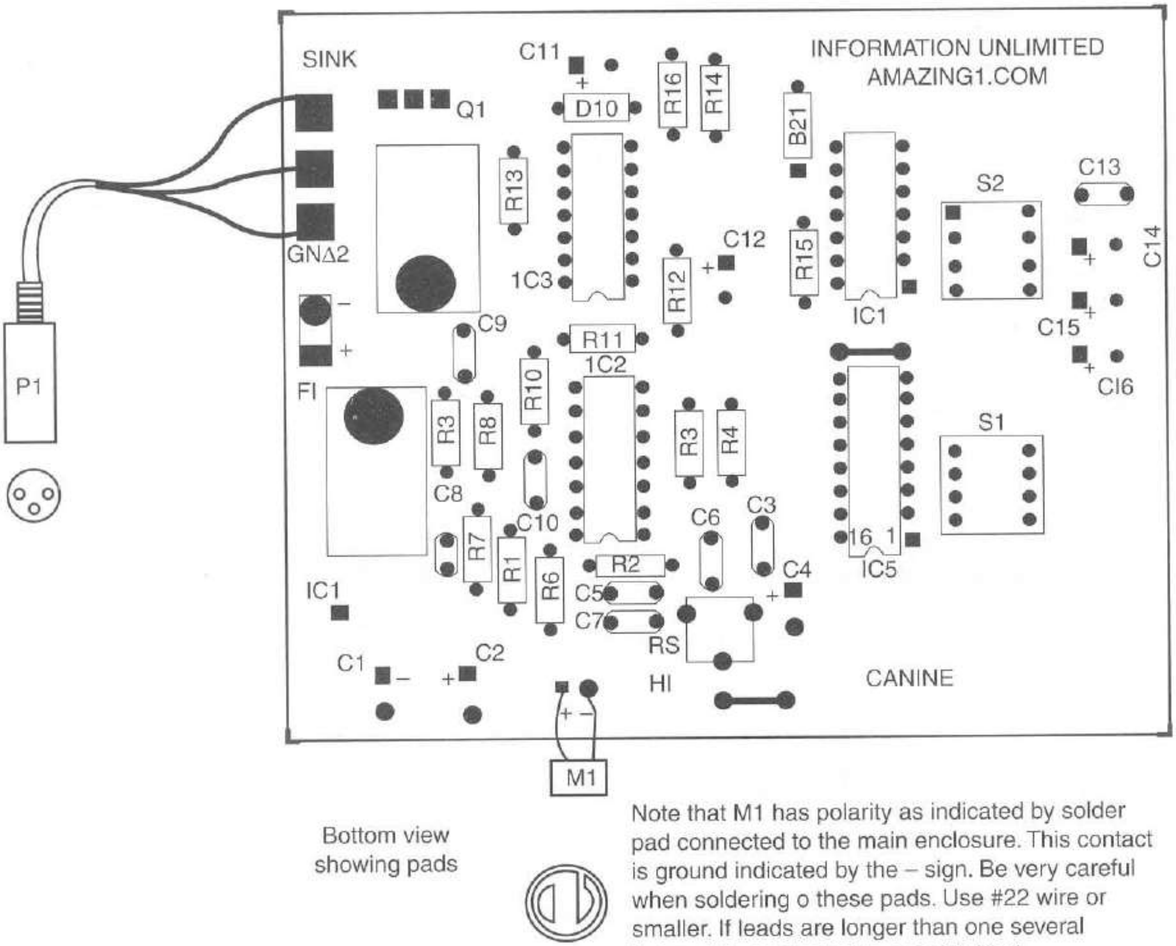
SINK
C11
Q1
D10
R16
R14
INFORMATION UNLIMITED
AMAZING1.COM
B21
R13
S2
C13
GNΔ2
C12
R15
C14
1C3
R12
IC1
C15
C9
R11
1C2
Cl6
P1
FI
R10
R3
R8
R3
R4
S1
C8
C6
C3
C10
16 1
R7
R1
R2
C4
IC5
IC1
R6
C5
C7
RS
C1
C2
HI
CANINE
M1
Bottom view
showing pads

Note that M1 has polarity as indicated by solder pad connected to the main enclosure. This contact is ground indicated by the – sign. Be very careful when soldering o these pads. Use #22 wire or smaller. If leads are longer than one several inches it is suggested to use shielded microphone cable.

Figure 34-6 *External wiring*

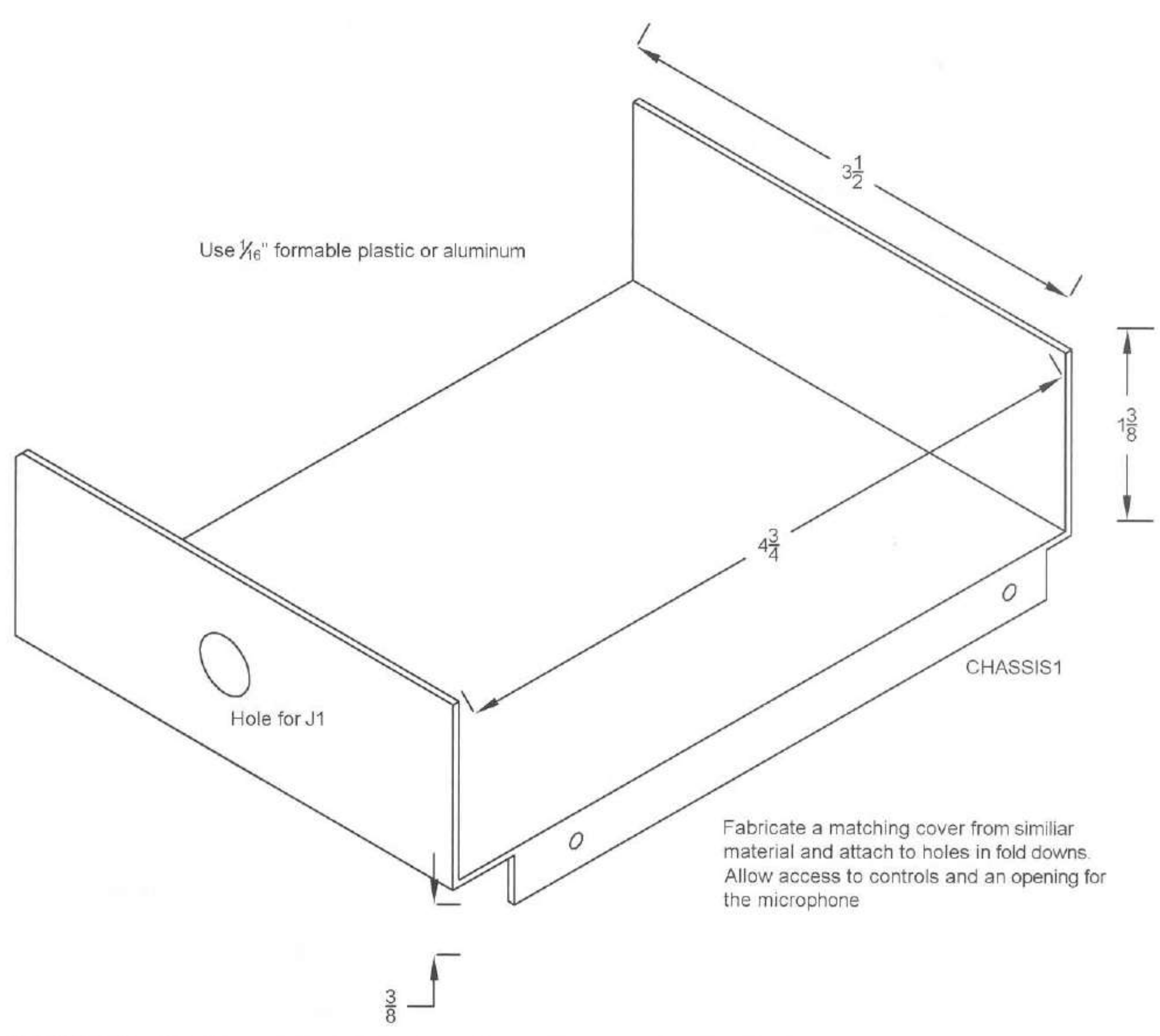

Use 1/16" formable plastic or aluminum

Fabricate a matching cover from similiar material and attach to holes in fold downs. Allow access to controls and an opening for the microphone

3 1/2
1 3/8
4 3/4
3/8
Hole for J1
CHASSIS1

Figure 34-7 *Chassis fabrication*

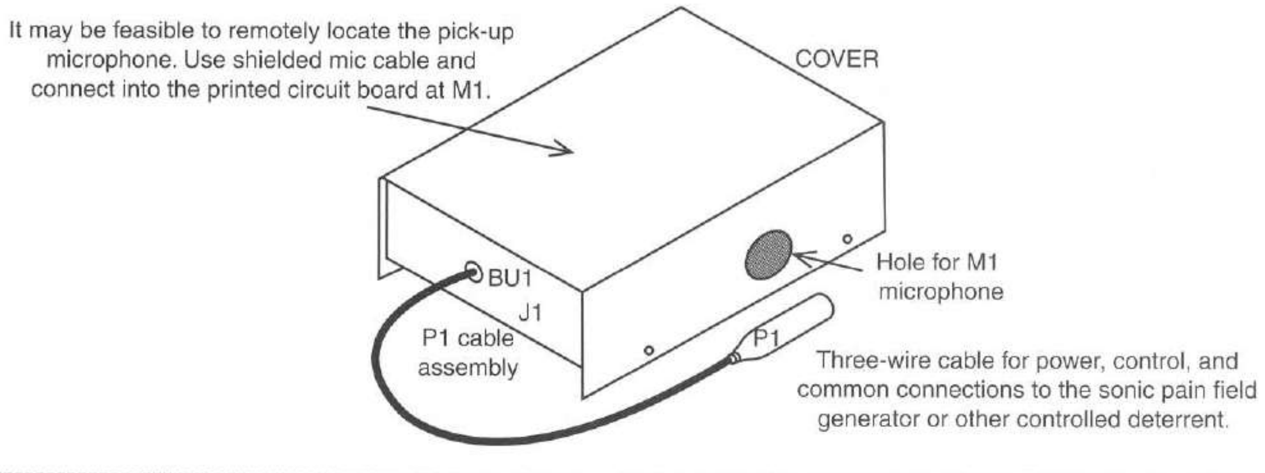

It may be feasible to remotely locate the pick-up microphone. Use shielded mic cable and connect into the printed circuit board at M1.

Three-wire cable for power, control, and common connections to the sonic pain field generator or other controlled deterrent.

COVER
BU1
J1
P1 cable assembly
P1
Hole for M1 microphone

Figure 34-8 *Isometric of final assembly*

Testing the Circuit

1. Assemble the test jig, referring back to Figure 34-3. If the circuit is correct, you may go right to the connection of the test jig as shown and perform the following steps:
    a. Preset S1 to position 4 (the lowest number of barks for activation).
    b. Preset S2 to position 1 (the shortest time for the sonic generator).
    c. Preset R5 to midrange.
2. Apply a 12 VDC input and note a low-current draw of less than 10 milliamps, which is the same value as the quiescent operating current. C cell batteries will work approximately 30 days before requiring replacement.
3. Simulate a bark and note the test LED momentarily coming on. If this occurs, you may want to test the various positions of S1 and S2. Also verify the sensitivity control, R5.
4. If the circuit does not work, you may need a scope to test the various test points as shown in Figure 34-3. They should have the following values:

    TP1: +5 volts DC

    TP2: 2 to 3 volts DC

    TP3: 4-peak-volt audio level

    TP4: Saturated audio signal

    TP5: Negative-going pulse

    TP6: Negative-going square wave

    TP7: Positive 10-second holding pulse

    TP8: Positive holding pulse

    TP9: Very fast negative trigger pulse

    TP10: 1-second turn-on delay

    TP11: Holding pulse for Q1

10. Once the controller operation is verified, you may connect the system as shown in Figure 34-9. You will need a properly working sonic pain-field generator, as described in Chapter 35.

It is suggested that you experiment with the system before actually installing it. Once familiar with it, you may position the system for the best effect. You also have the option of adjusting the sonic generator when the dog starts barking so that it can also be very annoying to the animal owners when their mutt goes into an uncontrolled barking rage. Use your own judgment as you can get tagged for harassment!

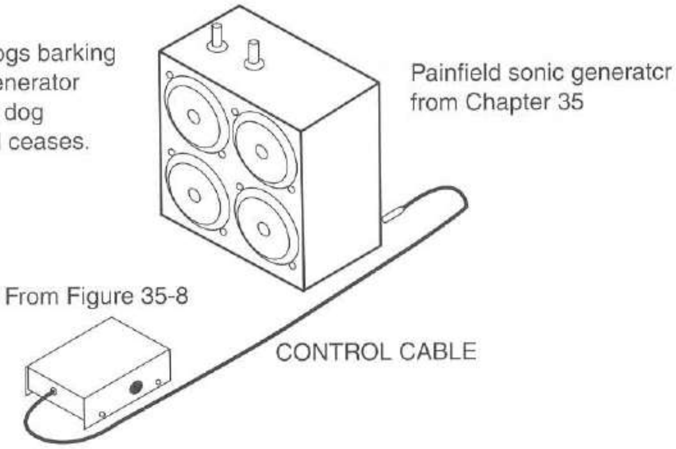

Microphone picks up the sound from the dogs barking and conditions it to turn on the pain field generator for a preselected burst time. Eventually the dog associates the sonic pain with his bark and ceases.

Figure 34-9 *Systems view*

Table 34-1 Canine Controller Parts List

Ref. #	*Description*	*DB Part #*
R1, 3, 4	Three 1K, ¼-watt resistors (br-blk-red)	
R2	150K, ¼-watt resistor (br-gr-yel)	
R5	10K pot and 12-volt switch, 17 mm	
R6	330K, ¼-watt resistor (or-or-yel)	
R7	2.2K, ¼-watt resistor (red-red-red)	
R8, 9, 10	Three 10 kilo-ohm, ¼-watt resistors (br-blk-or)	
R11	47 kilo-ohm, ¼-watt resistor (yel-pur-or)	
R12, 13, 15	Three 1M, ¼-watt resistors (br-blk-gr)	
R14	15-ohm, ¼-watt resistor (br-gr-blk)	
R16	100K, ¼-watt resistor (br-blk-yel)	
C1	220-microfarad, 25-volt vertical electrolytic capacitor	
C2, 4	Two 100-microfarad, 25-volt vertical electrolytic capacitors	
C3, 6, 7, 8	Four .1-microfarad, 50-volt plastic capacitors	
C5, 14	Two 2.2-microfarad, 50-volt vertical electrolytic capacitors	
C9, 10	Two .01-microfarad, 50-volt plastic capacitors	
C11, 13	Two 1-microfarad, 50-volt vertical electrolytic capacitors	
C12, 15	Two 10-microfarad, 50-volt vertical electrolytic capacitors	
C16	33-microfarad, 50-volt vertical electrolytic capacitor	
D1, 2	Two IN914 silicon diodes	
IC1	7805 5-volt regulator TO220	
IC2A, B, C, D	LM324 quad amp in dual inline package (DIP)	
IC3A, B, C, D, E, F	40106 hex Schmidt in DIP	
IC4	556 dual timer in DIP	
IC5	4040 complementary metal oxide semiconductor (CMOS) PLL phase lock loop in DIP	
Q1	100-volt IRF540 metal-oxide-semiconductor field effect transistor	MOSFET
M1	FET mic element	DB# FETMIK
S1, 2	Two four-position single-throw DIP switches	
PCCANINE	Printed circuit board (PCB)	DB# PCCANINE
CHASSIS1	Metal or plastic chassis as shown in Figure 34-7	
COV1	Cover fabricated as shown in Figure 34-8	
J1/P1	Three-pin chassis mount jack and mating plug	
WR3C	3-inch three-conductor cable for interconnecting	
PPF4K	Pain-field generator kit as described in Chapter 35	DB# PPF4K
PPF40	Pain-field generator assembled as described in Chapter 35	DB# PPF40

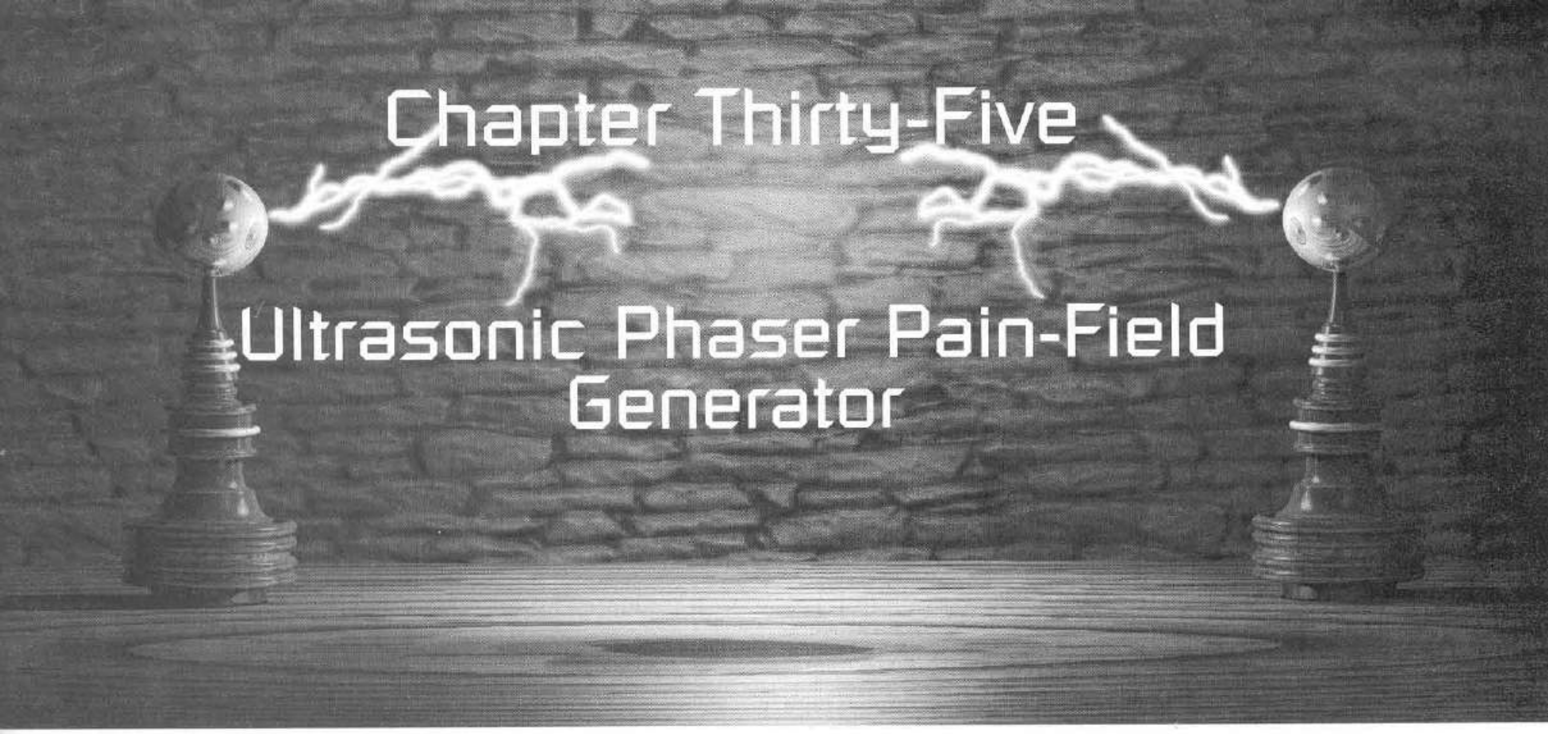

Chapter Thirty-Five

Ultrasonic Phaser Pain-Field Generator

This project, as shown in Figure 35-1, shows how to construct a moderately high powered sonic generator that can be used for tasks that range from animal control to discouraging personal encounters. It can be used as part of our laser property protection guard described in Chapter 5 or as the deterrent for use with the canine controller, as described in Chapter 34. The unit can generate a variable rate of complex waves from 5 to 25 kHz well into the ultrasonic range. These waves can be very painful or disorientating to animals and people, depending on where the controls of the unit are set.

This is an excellent device for use in animal control as well as a low-liability deterrent in anti-intrusion alarms and detection systems. This is an intermediate-level project requiring basic electronic skills. Expect to spend \$50 to \$75. All parts are readily available, with specialized parts obtainable through Information Unlimited at www.amazing1.com, and they are listed in the parts list at the end of the chapter.

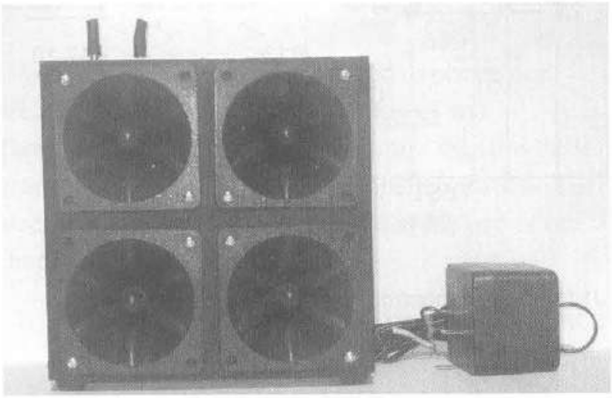

Figure 35-1 *The phaser pain-field generator*

Basic Device Description

This chapter demonstrates how to create a variable-sweep-frequency, ultrasonic, pain-field generator capable of producing the equivalent of 400 watts of resultant power such as that obtainable from a conventional dynamic transducer system. This is possible due to recently developed piezoelectric ceramic devices. The high efficiency requires very little driving power, consequently resulting in a lightweight, portable, handheld device that is battery driven and capable of producing high sound pressure. It is a directional device and can be set up in a target area such as a field or garden to discourage any sound-sensitive animals. It is also capable of being powered by a car's 12-volt system and can be mounted on the hood or roof of the vehicle. Transducers are mounted in an array for a concentration of energy in one direction.

It should be understood that certain people subjected to different degrees of exposure are affected more than others, some to a point where they may vomit or experience severe headaches and cranial pains. Some people will experience severe pain in the ear, teeth, or lower head. Statistically, women and younger children are many times more sensitive to this device than the average male adult. With this in mind, the user must exercise consideration when testing and using the device for animal control, as many people will not be aware of the source of this pain and attribute it to a headache or other physical ailment. Also, certain people are affected mentally to a point of actually losing their tempers completely or becoming extremely quick-tempered. Some will experience a state of extreme anxiety when overly exposed. Therefore, consideration must be used at all times when testing or using this or similar devices.

It should also be noted that using the transducers in an array configuration may be damaging to hearing at close range. The array approach produces high sound-pressure density occurring on or near the output axis.

Circuit Theory

Figure 35-2 shows a timer, IC2, connected as an astable free-running multivibrator whose frequency is externally controlled by pot R9. Resistor R10 selects the range limit of R9, and capacitor C5 determines the frequency range of the device along with the previous resistors.

The square wave output of IC2 is via pin 3 and is directly coupled to power amplifier Q2. The drain of Q2 is DC biased through choke L1.

The square wave output signal is then fed into transducer TD1 in series with parallel combination of resonating coils L2 and L3. The resonant action between the inherent capacity of TD1, the C7 tuning capacitor, and the inductance now produces a sinusoidal-shaped wave peaking around 25 kHz or the

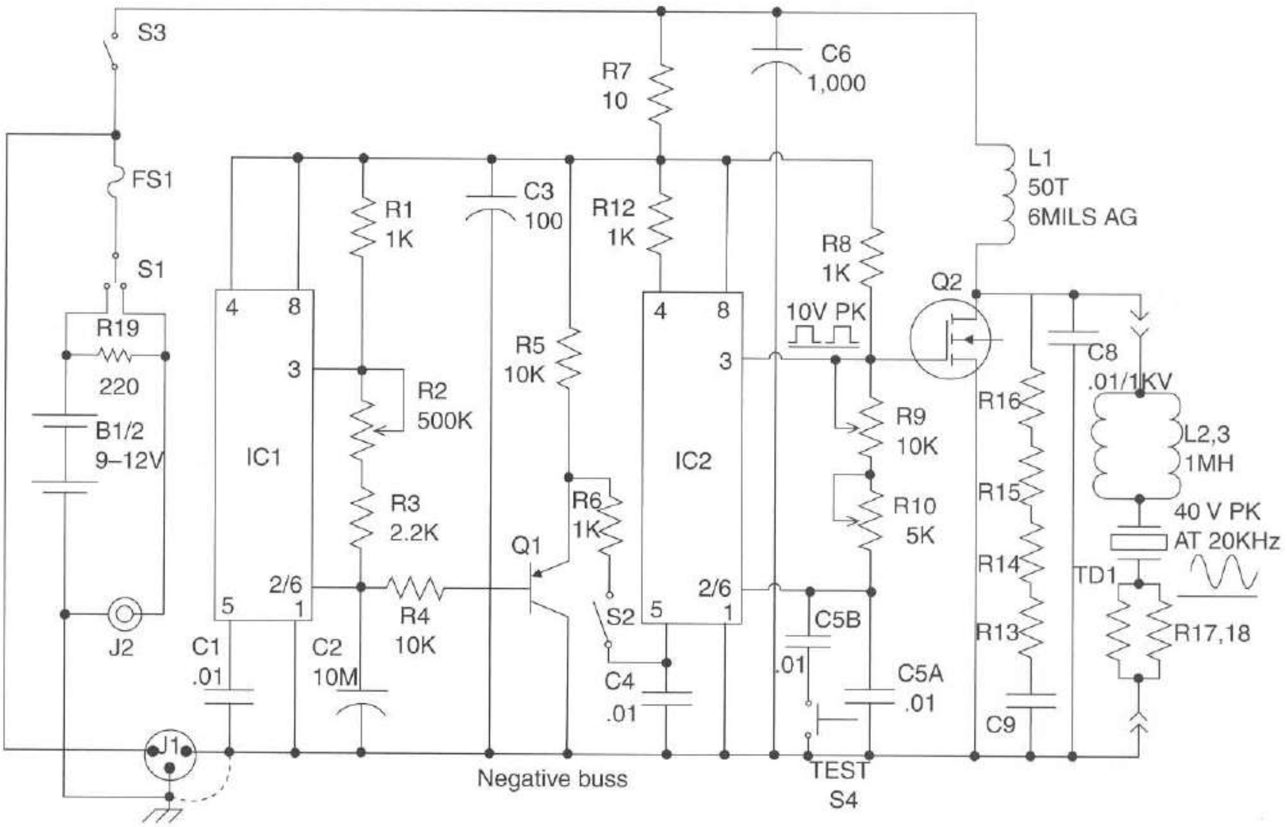
S3
R7
10
C6
1,000
FS1
L1
50T
6MILS AG
R1
1K
C3
100
R12
1K
R8
1K
S1
Q2
10V PK
R19
220
4
8
3
R2
500K
R5
10K
C8
.01/1KV
B1/2
9–12V
IC1
R9
10K
R16
L2,3
1MH
IC2
R3
2.2K
R6
1K
R10
5K
R15
40 V PK
AT 20KHz
Q1
2/6
R14
TD1
5
1
R4
10K
S2
C5B
R13
R17,18
J2
C1
.01
C2
10M
C4
.01
C5A
C9
J1
Negative buss
TEST
S4

Dashed line indicates bypass connection when not using J1 for external control. You may also short these pins using a mating plug with pins connected together.

Note four transducers and associated networks.

Figure 35-2 *Schematic of the phaser pain-field generator*

upper limit of the tuning range. This signal waveform now has a peak-to-peak voltage several times that of the original square wave. Transducer TD1 now can take advantage of these peak voltages to produce the high sound-pressure levels necessary without exceeding the high *root means square* (rms) ratings of an equivalent voltage-level square wave.

Timer IC1 is similarly connected as an astable, running multivibrator and is used to produce the sweeping voltage necessary for modulating the frequency of IC2. This sweep repetition rate controlled by pot R2 and resistor R3 limits the range of this repetition time. Resistor R1 selects the duty cycle of the pulse, while capacitor C2 sets the sweep time range. The output for IC1 is via pins 6 and 2 where the signal ramp function voltage is resistively coupled to inverter transistor Q1 via resistor R4. The output of Q1 is fed to pin 5 of IC2 and provides the output modulation voltage necessary to vary the frequency as required. Note that the modulation signal is easily disabled via R2/S2.

Capacitor C6 guarantees an AC return path for the output signal. Power to driver circuits IC1 and IC2 is through a decoupling network consisting of resistor R7 and capacitor C3.

Power to the system is via internal battery B1 or an external 12 VDC such as from a vehicle. Three-way switch S1 selects between these two power sources or shuts the unit off in the center position. Resistor R19 provides a trickle charge to the internal battery when using an external source. Jack J1 allows a connection that is similar to our canine controller, turning on the unit when it detects a barking dog.

Construction Steps

If you are a beginner it is suggested to obtain our *GCAT1 General Construction Practices and Techniques*. This informative literature explains basic practices that are necessary in proper construction of electromechanical kits and is listed at the end of the chapter.

1. Lay out and identify all the parts and pieces, and check them with the parts list. Note that certain parts may sometimes vary in value. This is acceptable as all components are 10 to 20 percent tolerant unless otherwise noted.
2. Fabricate the heatsink bracket (HS1) as shown in Figure 35-3 from a .75 × 2 × .065-inch aluminum piece. Bend it 90 degrees at its midsection and drill a hole for the SW1/NU1 screw and nut. Attach it to Q2 as shown.
3. Assemble the L1 choke coil as shown in Figure 35-4 by wrapping 50 turns of #24 magnet wire on the nylon bobbin as evenly as possible. Leave 2 inches of leads for a connection to the circuitry. Assemble the E core as shown and shim each side with pieces of yellow cardboard strips of 3 mils each for a total of 6 mils. If you have an LCR bridge, measure 1 millihenry.
4. Assemble the PCB as shown in Figure 35-5. Note the two wire jumps and component polarity. Wire in inductor L1 and secure it to the board with *room temperature vulcanizing* (RTV) silicon rubber or another suitable adhesive.

   If you are building from a perforated or vector circuit board, it is suggested that you use the indicated traces for the wire runs and insert the components starting in the lower left-hand corner. Pay attention to the polarity of the capacitors with polarity signs and all the semiconductors.

   Route the leads of the components as shown and solder as you go, cutting away unused wires. Attempt to use certain leads as the wire

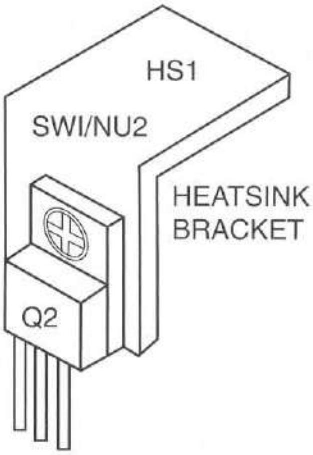

Figure 35-3 *Fabrication of heatsink*

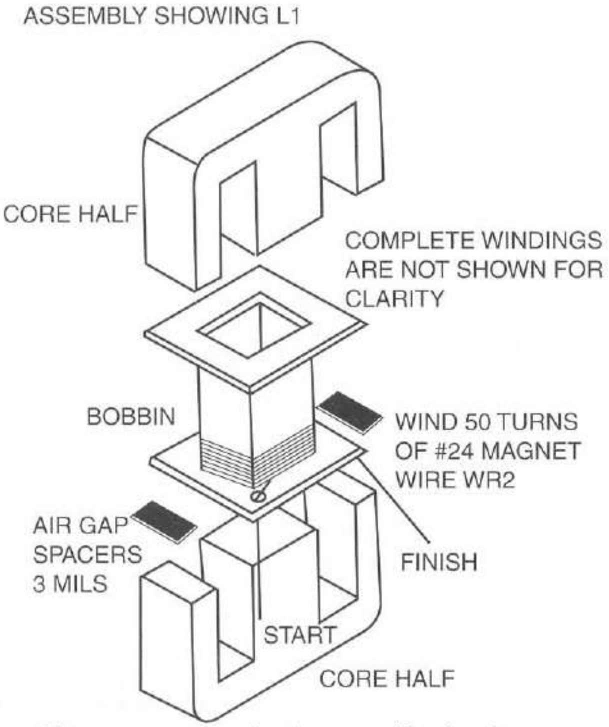
ASSEMBLY SHOWING L1
CORE HALF
COMPLETE WINDINGS ARE NOT SHOWN FOR CLARITY
BOBBIN
WIND 50 TURNS OF #24 MAGNET WIRE WR2
AIR GAP SPACERS 3 MILS
FINISH
START
CORE HALF

Note you can use a business card for the air gap shims as thickness will be close enough.

Use tape to secure core halves tightly together.

Figure 35-4 *Assembly of L1 inductor*

runs or use pieces of the #24 bus wire. The heavy foil runs should use the thicker #20 bus wire as these are the high-current discharge paths.

It is suggested that you use an 8-pin integrated circuit socket (SO8) for IC1 and IC2.

5. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.

6. Fabricate the chassis and cover, as shown in Figure 35-6 from 1/16-inch Lexan plastic or aluminum. Use a large punch or circle saw to cut out the four holes for the transducers. Drill the remaining holes as you assemble it, verifying the proper clearances and sizes with components.

7. Complete the final assembly shown in Figure 35-7 and mount the assembly board, battery holders, switch S1, jacks J1 and J2, and fuse holder FS1. It is a good idea to trial-fit all the parts in this section before actually drilling or making any mounting holes. Note a piece of plastic (SPACER) that insulates the bottom solder connections of the PCB from shorting out to the metal chassis base. Also note resistor R19 across S1. Secure the battery holders using pieces of two-sided foam sticky tape (TAPE). Wire using #20 vinyl wire as shown.

8. Mount the transducers as shown in Figure 35-8 using SW2/NU2 6-32 × 1/2-inch screws and nuts.

   Twist the leads of the two inductors L2 and L3 together, effectively paralleling these components. Repeat this for resistors R17 and R18. You should end up with four sets of these parallel components. These components should be self-supporting by being connected to stiff pieces of bus wire. Take caution to observe any potential shorts. Solder the two 18-inch connecting leads and connect them to point P6 on the assembly board.

9. The final assembly is shown in Figure 35-9. Do not attach the cover until you have tested the unit.

Testing the Assembly

10. Turn the pots counterclockwise, set the switch to mid position (off), and install eight fresh batteries into the battery holders. If you have access to a bench supply of 12 VDC at 2 amps, it may be connected to J2 via a mating plug, eliminating the need for the batteries in the preliminary test. Short out the chassis ground and negative bus pins of J1. This jack is for connection to the canine controller module or other remote control devices.

Note: A variable bench supply capable of 12 VDC at 2 amps with a voltmeter and current meter can be a great convenience for the remaining steps and the testing of other similar circuits.

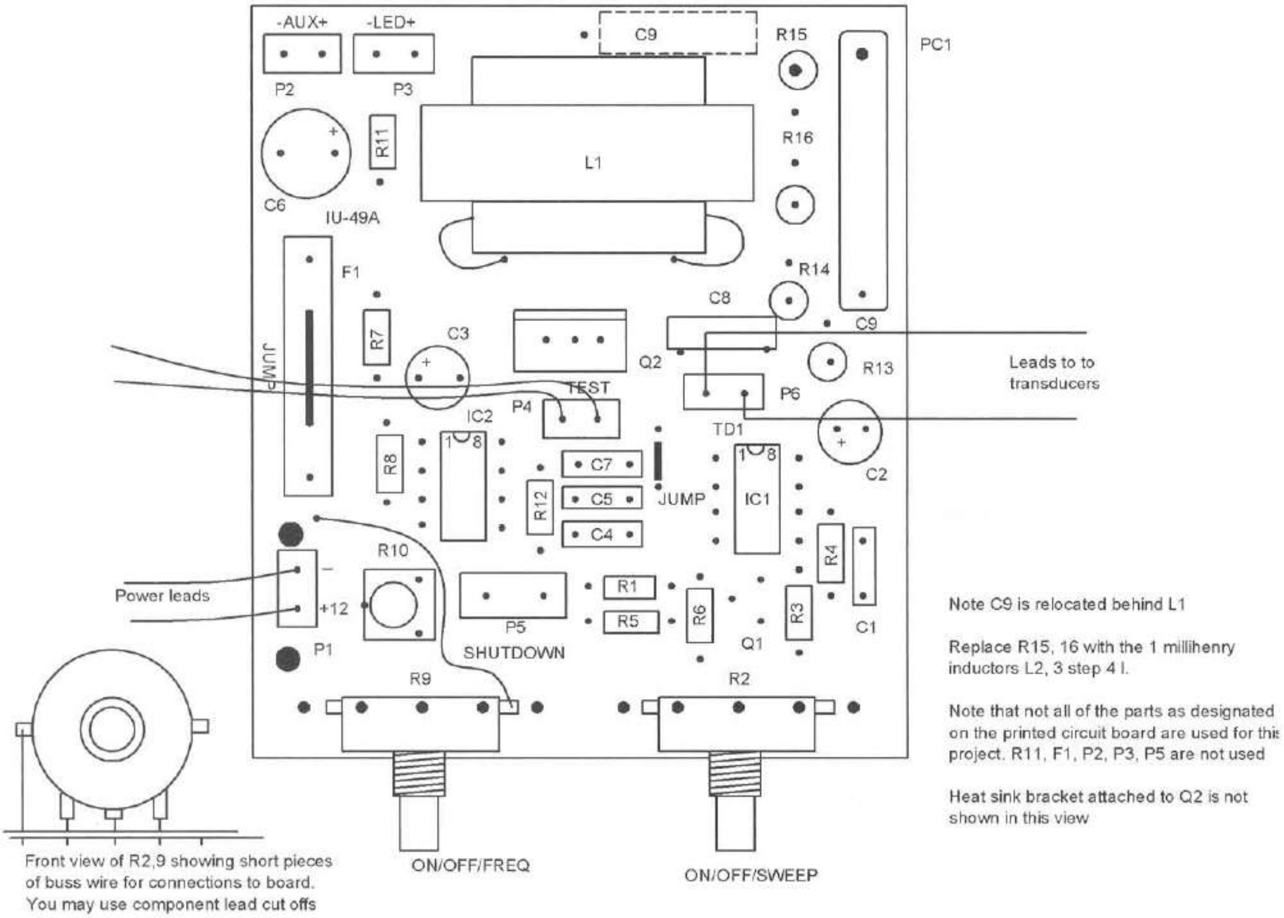

Note C9 is relocated behind L1

Replace R15, 16 with the 1 millihenry inductors L2, 3 step 4 I.

Note that not all of the parts as designated on the printed circuit board are used for this project. R11, F1, P2, P3, P5 are not used

Heat sink bracket attached to Q2 is not shown in this view

Front view of R2,9 showing short pieces of buss wire for connections to board. You may use component lead cut offs

Figure 35-5 *Assembly of PCB*

11. Press push button S1 and note a loud, piercing wave coming from TD1. Measure a current of 500 to 600 milliamps in series with a battery or bench supply. Note that the switch position when up should be power supplied by internal batteries, and the down position being supplied by the external 12 VDC supply. This function may be reversed but should be noted, as the external position will automatically trickle charge the internal batteries.

12. Rotate R9 and note the frequency increasing to above the audible range; measure a current of 400 to 800 milliamps. You may preset the limit of the high range by the R9 setting. The normal factory setting is 25 kHz maximum with R10 full clockwise. Note the wave shapes shown in Figure 35-2 for those who have a scope.

13. Turn on sweep control R2/S2 and note the frequency being modulated by a changing rate as this control is adjusted. Use caution as certain sweep rates may cause epileptic seizures and other undesirable effects. Sweep rates between 7 to 20 per sec should be used with caution.

Basic Operating Instructions

14. Connect a 12-volt source capable of supplying 2 amps of current to designated leads or use internal *nickel cadmium* (NiCad) batteries that are type C cells. These will provide approximately 3½ hours continuous operation at slightly reduced output.

15. The battery-charging function occurs when the system is connected to an external power source and will charge the NiCad batteries at a 100-milliamp rate, taking about 14 hours to fully charge.

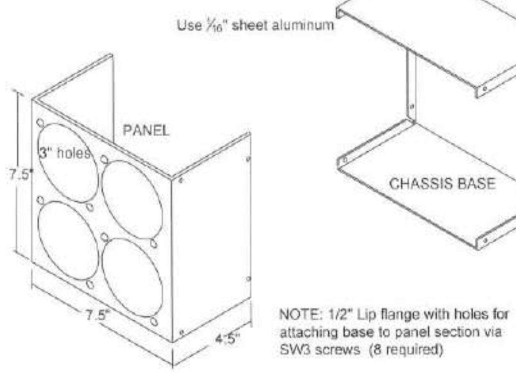
Use 1/16" sheet aluminum
PANEL
3" holes
7.5"
7.5"
4.5"
CHASSIS BASE

NOTE: 1/2" Lip flange with holes for attaching base to panel section via SW3 screws (8 required)

Figure 35-6 *Chassis fabrication*

16. If continued use is anticipated via the external 12 volts, it is a good idea to remove the internal batteries' connections as overcharge may occur that will ruin the batteries or actually cause them to explode. The recommended method of setting the controls is the following:
    a. Determine the frequency limits per the application. These types of devices have two basic applications. When used as an anti-intrusion device to discourage unauthorized entrance, adjustments are made for maximum human annoyance, usually with frequencies from 10 to 15 kHz.

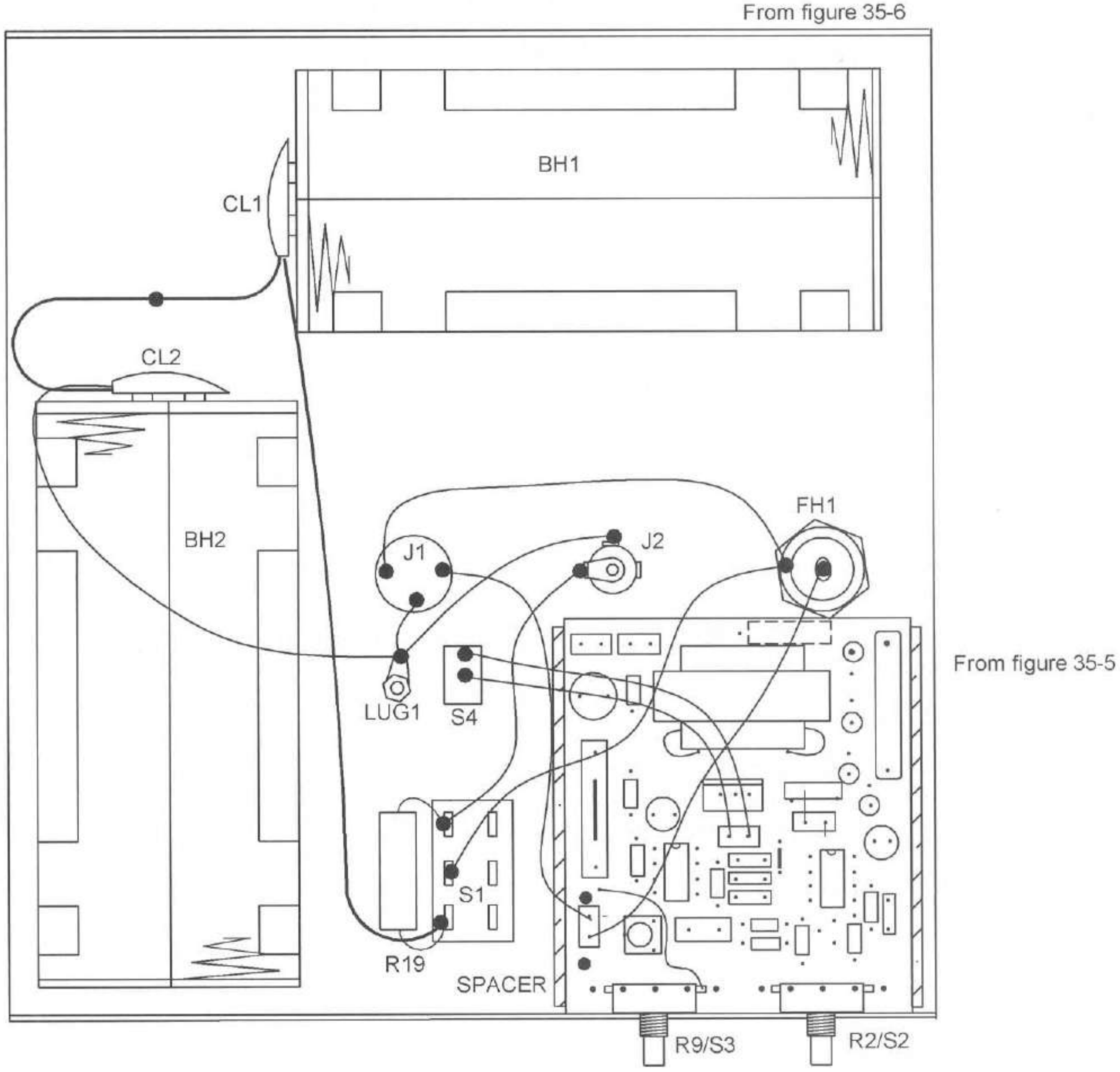
From figure 35-6
BH1
CL1
CL2
BH2
FH1
J1
J2
LUG1
S4
From figure 35-5
S1
R19
SPACER
R9/S3
R2/S2

Figure 35-7 *Chassis wiring*

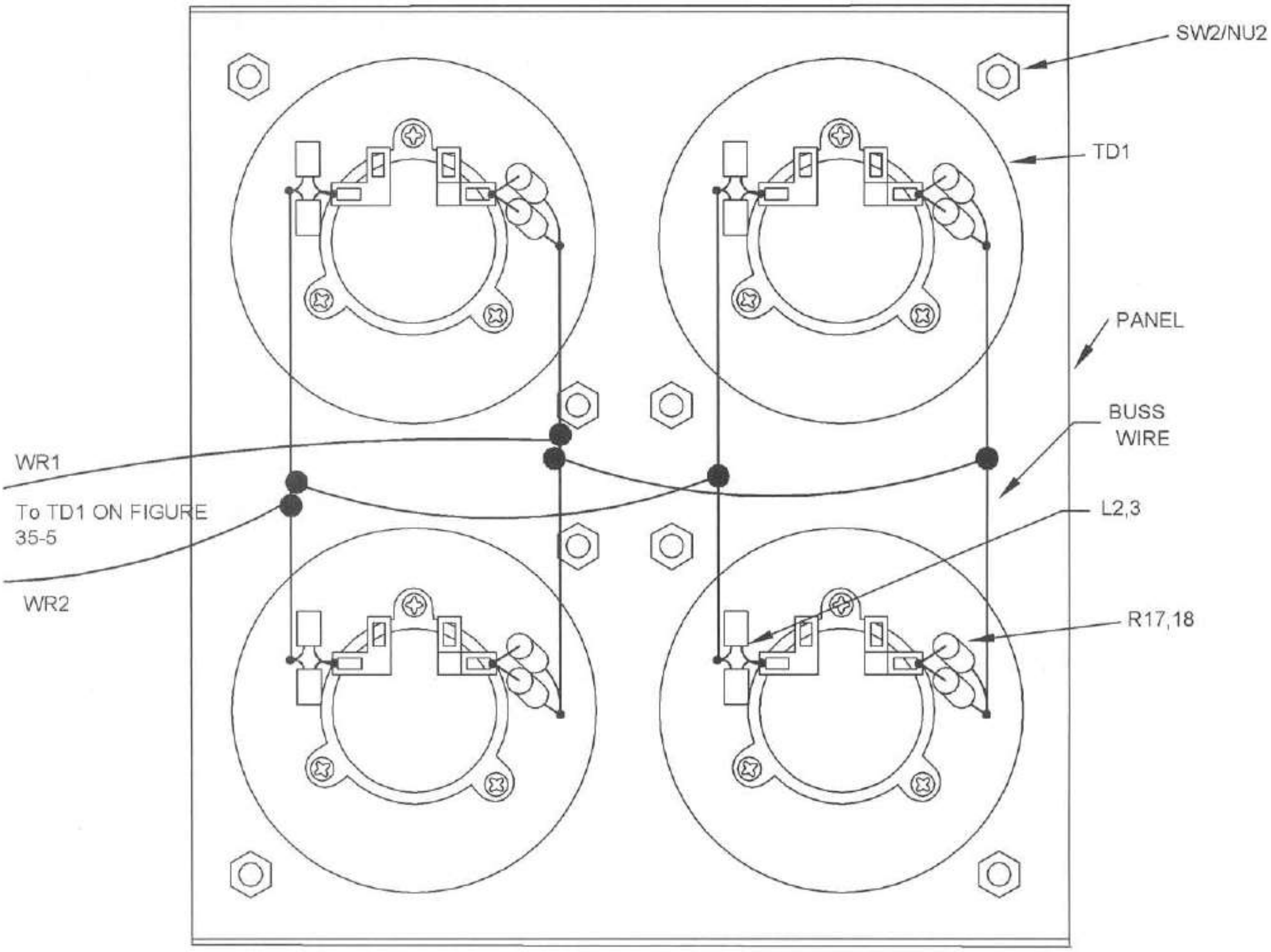

Figure 35-8 *Transducer panel wiring*

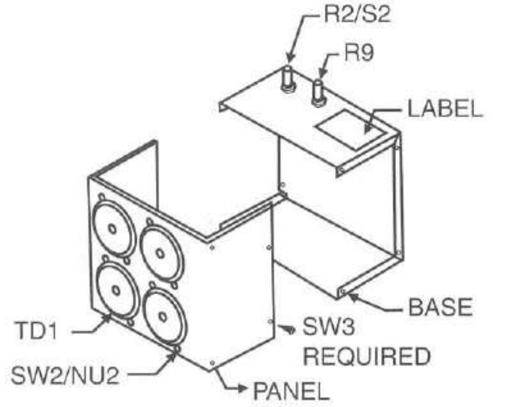

NOTE: 1/2" Lip flange with holes for attaching base to panel section via SW3 screws (8 required)

Figure 35-9 *Final assembly*

When used as a rodent device, we have found that the lowest tolerable frequency to humans usually has the greatest effect on the common species of rats. This is not always true but serves as a starting point in initially setting the adjustments. A frequency setting of 15 to 20 kHz for nonhuman living areas and 20 to 25 kHz for areas where people are present usually suffices for good rat control. Note that at no time should the unit be set higher than 25 kHz.

Dog control is usually around 18 kHz, but experimenting with the target animal may be necessary for optimum effect. Stubborn dog owners fall into the category of anti-intrusion if they will not cooperate.

b. Turn the SWP control off. Adjust the TONE control so that the output is detectable by the human ear or just above the point of annoyance.

c. Set the TEST switch to an audible verification. Note that experimentation is required for optimizing any effect on a given set of target applications.

Information on the System

Your phaser system is capable of operating in two modes. Mode one is at a frequency that is known to produce paranoia, nausea, disorientation, and many other physiological effects. Mode two allows using the system as an audible alarm to frighten off intruders or warn the user of an intrusion. Both modes may be used in combination and are easily rear-panel-controlled by the user.

A Word of Caution

Ultrasonics is a grey area in many respects when an application involves the control of animals or even a deterrent to unauthorized intrusion. It is always best to consult with local municipal and state laws before using this device to protect home or property. Remember many state laws lean more towards the right of the criminal rather than the victim.

General Information on Ultrasonics

Numerous requests have been made for information on the effect of these devices on people. First, let us make it clear that no device such as this should purposely be used on humans, and we discourage this use due to the possibility of acoustically sensitive people being highly irritated.

None of these devices has the ability to stop a person with the same effect as a gun, club, or more conventional weapon. They will, however, produce an extremely uncomfortable, irritating, sometimes painful effect in most people. Everyone will experience this effect to some degree. Unfortunately, younger women are much more affected than older men due to being more acoustically sensitive. The range of the device depends on many variables and is normally somewhere between 10 and 100 feet depending on the acoustical sensitivity of the target.

One possible use of the device (which deserves careful consideration) is the installation of all transducers in an area to protect against unauthorized intrusion. This would produce an irritating and painful feeling for the intruder, along with a condition of paranoia of not knowing what to expect next.

Application Supplement

The following describes an acoustical, ultrasonic device for dispersing potentially unruly crowds or gatherings. The reason for using the higher-frequency energy is that it is easily directed without being diffracted in all directions. Therefore, the use of directional-type transducers producing a front-to-back relative signal ratio of 20 to 30 dbs is easily achieved. The higher-frequency sound can produce nausea, headaches, and depression without the subject necessarily being able to determine the source. This effect no doubt presents an advantage to law enforcement when using these devices.

It is a known fact that the development of any type of demonstration in "psychological crowd control" follows the natural laws of growth and decay, in that a positive influx of people encourages a crowd growth, whereas a negative outflux creates a diminishing situation. One must realize that differences exist among most people regarding sound level and frequency sensitivity. When sampling a group or crowd as to the overall sensitivity of these devices, it is usually found that certain ones may not be affected to the same degree as others. However, taking into consideration that those who find the effect intolerable does generate the negative outflux that usually terminates the gathering.

This method, to this date, produces no permanent, lasting effects or symptoms and therefore becomes a more humane means of dispersing potentially unruly crowds. It should also be mentioned that younger people appear to be many times more sensitive to this effect. This usually goes hand in hand in most demonstrations, since younger people are more apt to become unruly.

This section encompasses many of the methods for applying this device via a permanent installation where transducers can be placed and adjusted to produce the necessary sound-pressure levels at distances to achieve the desired effects. This method could prove effective as an intrusion deterrent for a certain protected area.

Another method could be a handheld directional device similar to a megaphone where the amount of transducers used would be determined by the affects desired. This device would take into consideration the highly directional characteristics of this energy, being able to be directed at problem areas as a handheld portable system.

Another method for effective control would be clusters or arrays of transducers with these systems being mounted to vehicles via conventional roof mounts or on the hoods of these vehicles. Operation could take further advantage of the fair to good soundproofing found in most automobiles and vehicles could approach problem areas, taking advantage of close-proximity situations.

It has been found that certain frequencies also can be made to produce intense irritability in certain types of people. This method could justify the eventual use of more severe methods of restraint, assuming that the target subjects would eventually become more aggressive or violent, thus warranting this action. This is only to be considered as a potential extreme use of the device.

The objective here is to describe a similar method whereby the sound-pressure levels and frequencies can be obtained economically and efficiently, using state of the art methods and easily available parts, yet they can still be portable and easily handled for complete operational flexibility. Construction is based around the use of the piezoelectric tweeter transducers. These lightweight devices are easily driven and produce a conversion efficiency of over six times their electromagnetic counterpart. Being lightweight, economical, and easy to use when driving, they are an excellent candidate for this type of system. Higher driving impedance versus frequency is the inverse of the electromagnetic type.

The system, as described, contains several modes of operation. The first mode is a steady state of a certain frequency determined by the operator. This mode is termed *manual frequency control* and is adjusted by its appropriate control. The next mode is *low speed sweep*, which consists of the unit starting at a low frequency and automatically increasing to its higher limit where it again repeats itself. The remaining modes are controlled rates of this sweep, determined by the appropriate control referred to as *sweep rate control*. It is these controls that can be remotely operated from the inside of a vehicle when high-powered, outside arrays in clusters are used. As was mentioned before, certain frequencies combined with certain sweep rates can cause different degrees of effectiveness. It is not the objective of this information to analyze the potential types of effect and behavior versus control setting, but to describe the working system as a generator of these controlled sounds.

Table 35-1 Pain-Field Generator Parts List

Ref. #	*Description*	*DB Part #*
R1, 6, 8, 12	Four 1K, ¼-watt resistors (br-blk-red)	
R2/S2	500K, 1 meg pot/switch, 17 mm	
R3	2.2K, ¼-watt resistor (red-red-red)	
R4, 5	Two 10K, ¼-watt resistors (br-blk or)	
R7	10-ohm, ¼-watt resistor (br-blk-blk)	
R9/S3	10K pot/switch, 17 mm	
R10	5K horizontal trimpot	
R13–16	Four 4.7-ohm, 3-watt resistors (yel-pur-sil-gold)	
R17, 18, 19	Nine 120-ohm, 5-watt resistors (br-red-br); use two per transducer and one across S1	
C1, 4	Two .01-microfarad, 50-volt disk capacitors (103)	
C3	100-microfarad, 25-volt vertical electrolytic capacitor	
C2	10-microfarad, 25-volt vertical electrolytic capacitor	
C5A, B	Two .01-microfarad, 50-volt polyester capacitor	
C6	1,000-microfarad, 25-volt vertical electrolytic capacitor	
C8	.01-microfarad, 2-kilovolt disk capacitor	
C9	.22-microfarad, 250-volt polypropylene capacitor	
L2, 3	Eight 1 mh inductors (6080), 2 used per transducer, in place of R15 and R16 as marked on PC board	DB# 1MH
L1	Inductor: two Hitachi 30.48 E cores and mating bobbin assembled and wound L1 inductor	DB# PPPL1L1
Q1	PN2907 PNP GP transistor	
Q2	IRF530 or 540 N-channel metal-oxide-semiconductor field effect transistor (MOSFET)	
IC1, 2	Two 555 dual inline package (DIP) timers	
HS1	Heatsink bracket, created as shown	
SW1	6-32 × ¼-inch screw	
SW2	Eight 6-32 × ½-inch screw	
SW3	Eight #8 × ¼-inch sheet metal screw	
NU2	Nine 6-32 hex nuts	
FH1/FS1	Fuse holder and 2-amp fuse	
J1	Three-pin din chassis mount jack	
J2	DC 2.5 mm jack panel mounting	
CL1, 2	Two battery 9-volt clips	
BH1, 2	Two four-C-Cell holders	
S1	Single pole, single throw (SPDT) power switch	
S4	Small SPST toggle switch	
PC1	Printed circuit board (PCB) or use perforated vector board	DB# PCSONIC
TD1	Four polarized 130 db piezo transducer	DB# MOTRAN
WR1/2	36 inches of #20 vinyl hookup wire, red and black	
CHASSISBASE	Create as shown in Figure 35-6	
PANEL	Create as shown in Figure 35-6	
TAPE	8 inches of two-sided 1-inch foam tape for securing battery holders to base	
SPACER	3 × 3-inch piece of plastic sheet to insulate assembly board from metal BASE	
LABULTRA	Label, ultrasonic warning	

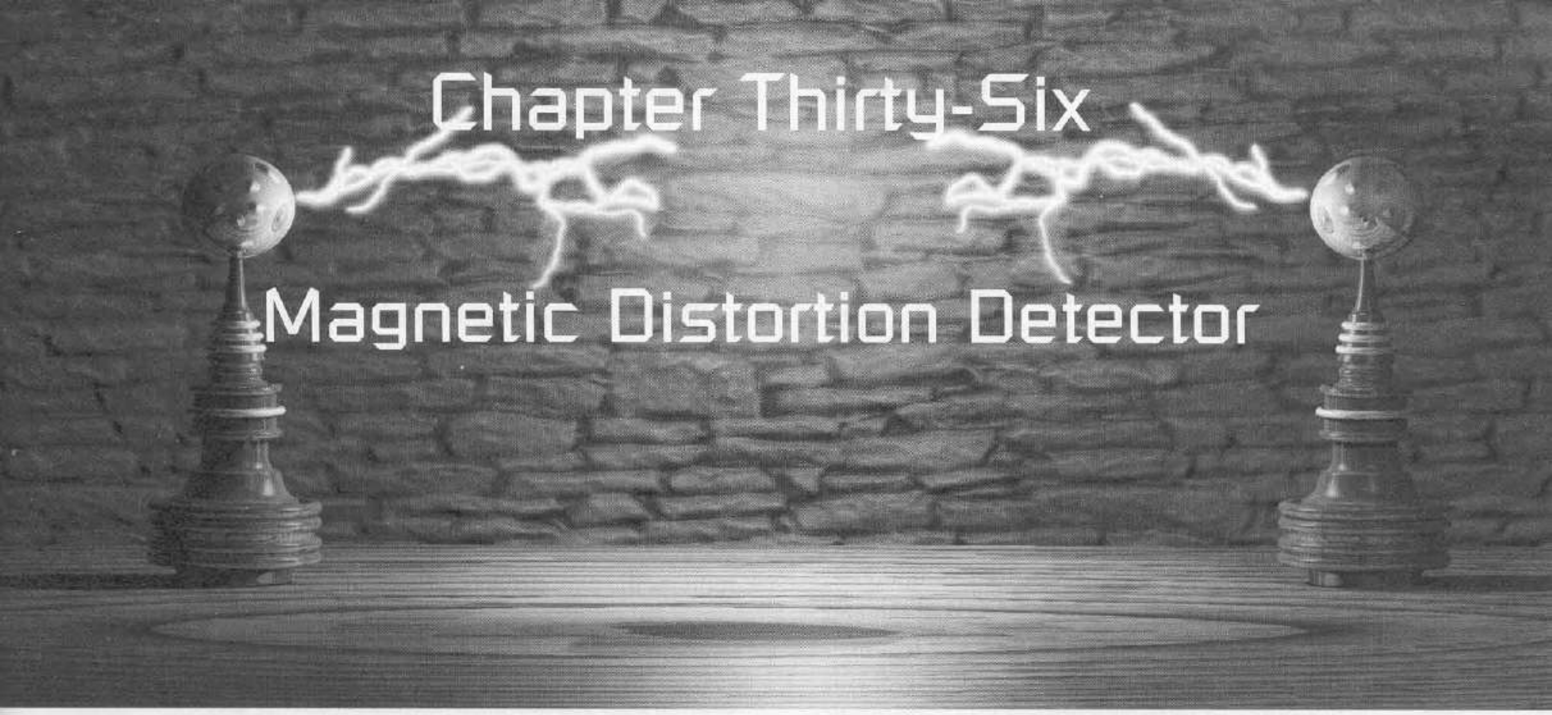

Chapter Thirty-Six

Magnetic Distortion Detector

The following project, as shown in Figure 36-1, describes an excellent, low-cost device that has amazing sensitivity in its ability to detect magnetic disturbances down into the microgauss. A microgauss is one millionth of a gauss. The earth's magnetic field is 500 milligauss where a milligauss is one thousandth of a gauss. The unit easily detects sunspot activity and auroras, and it can even be used as a vibration sensor that can detect a person's finger tapping on a rock. It can also detect moving aircraft and vehicles, as well as other changing electric and magnetic phenomena.

Actual detection is accomplished via an analog meter that responds to the relative strength of the changing magnetic fields. An audible alarm also allows the presetting of an activation level where it will alert the user of certain activity. An output port is included for connecting to a chart recorder.

The magnetic fields produced by many devices, both manmade and natural, are detectable by this system. Interesting results are produced when this unit is placed near anything electrical or magnetic such as a car, motor, or TV. The sensitivity is impressive when one sees the results of a 1-amp pulse of electrical current through a wire located a meter away from the unit. (This corresponds to a milligauss. Earth's magnetic field is .5 gauss). When the system is properly installed, it will detect passing automobiles, aircraft, or any moving, ferrous source that will cause the terrestrial magnetic field to distort. The unit is an excellent science fair project and can supply the amateur scientist with hours of knowledgeable entertainment in listening and detecting magnetic fields and their abnormalities.

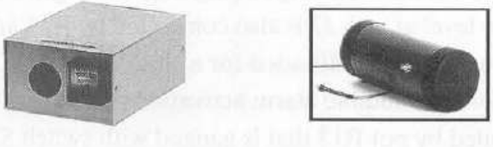

Figure 36-1 *The magnetic distortion detector*

Construction is shown to be as flexible as possible. It is built in two sections: the sensing head and the control box, which are interconnected via a length of shielded cable. The maximum usable sensitivity requires the sensing head to be placed away from power lines and other potential disturbances. The sensing head, which must be securely mounted to prevent interference from mechanical movement due to wind and vibration, can cause erroneous signals as it shifts position through the earth's magnetic field. The unit can also be made as a sensitive magnetic probe by installing the sensing head and the control box together, bearing in mind that the usable sensitivity will be decreased due to erroneous movement near magnetic materials when moving its position. The detection of field changes produces a moderately loud alarm along with a relative indication from an analog meter located on the panel. An optional jack is used for an X/Y chart recorder, and the system is fully battery powered.

This is an intermediate to advanced-level project requiring basic electronic skills. Expect to spend $45 to $65. All parts are readily available, with specialized parts obtainable through Information Unlimited at www.amazing1.com, and they are listed in the parts list at the end of the chapter.

Circuit Description

Figure 36-2 shows a coil (L1) consisting of many thousands of turns is wound on a 2-inch bobbin and placed over a soft iron core of about 1 to 2 feet in length. Any magnetic changes induce small voltages to occur at the output of this coil. This output signal is fed into a nano-amp, sensitive DC amplifier (I1), providing a current gain of 500 times the original. The iron core of L1 serves to concentrate the magnetic flux lines by offering a lower reluctance path than that of air. The input of I1 is protected by overvoltage and transients via clipping diodes Dl and D2. The output of I1 is forced to near zero via a multiturn trimpot (R4) for balancing and compensating any offset.

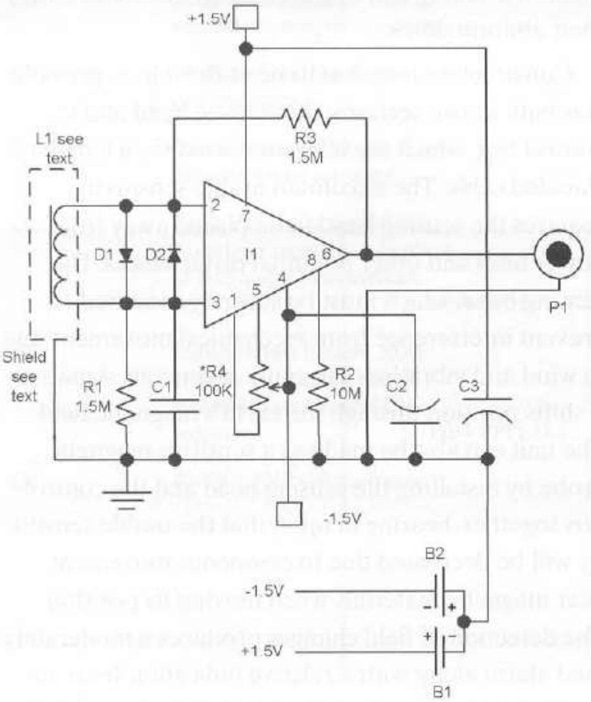

Figure 36-2 *Sense head schematic*

Power for I1 consists of batteries Bl and B2, which are left in the circuit permanently as current drain is very low and replacement can be on a yearly basis. It should be noted that if the sensing head is to be located in a hard-to-get area the builder might use D-size cells, as these will last that much longer.

The sense head output is fed to the control box via any reasonable length of shielded cable, WR10. As mentioned before, the sensing head should be located away from any potential magnetic interfering devices, such as anything electrical and large, ferrous objects.

Aluminum foil can be wrapped around the sense coil and is connected to the circuit common line. This reduces the capacitive pickup of unwanted signals.

Again, the sensing head and control box may be assembled together and used as a probe. This configuration may reduce the system's sensitivity due to erroneous movements and interaction of the circuits such as magnetic speakers and meters. This method is not encouraged due to these erroneous readings.

Figure 36-3 shows the control box providing another amplifier, I2, with a gain of 200 times the original amount, which gives the system an overall gain of 100,000. Full-scale meter deflection of the 50 ua meter (M1) occurs when a current of .5 nano-amp is obtained from L1. It is obvious that with the many thousands of turns on L1 that a detectable signal can occur with very minute flux changes, such as the result of changes in the microgausses. It should be noted that the earth's magnetic field is only .5 gauss.

I1 is connected as a conventional DC operational amplifier with its output DC coupled to I2 via a shielded cable. This cable is necessary to separate the sense head and the control box by at least 6 feet. This is necessary to prevent magnetic coupling from the speaker SPK1. Sensitivity control pot R5 controls the system gain and is usually set at maximum.

You will note that meter deflection pot R11 also is ganged with switch S1, which controls the +6-volt power to the control box section. R11 controls the deflection of meter M1 with diode D3 and capacitor C6, serving to smooth out and average the signal. The voltage level at jack J2 is also controlled by R11 and is an auxiliary jack intended for a chart or other recorder. The audible alarm activation point is controlled by pot R12 that is ganged with switch S2.

Switch S2 controls the −6 volt power to the control box.

The alarm is nothing more than a reflex-type oscillator consisting of transistors Q2 and Q3 with its feedback paths consisting of resistor R15 and capacitor C8. Trimpot resistor R16 is set for a reliable, clean-sounding signal.

Assembly Steps

1. Figure 36-4 shows coil L1 wound with thousands of turns of fine magnet wire. This may be difficult to do and it is suggested that the builder locate a high-voltage oil burner ignition, neon sign, bug killer, or other similar transformer and carefully remove the high-voltage secondary winding. The parameters of these transformers are usually around 5,000 volts at 10 to 20 milliamps. These are available from junkyards and other places. Many are rejected due to radio and TV interference and yet have intact secondary winding suitable for this project. The objective is to obtain one with as many turns as possible; however, most will work quite well. A quick test is to measure the DC resistance of between 5 and 20K across the coil. A relay coil of high impedance such as 5 to 10K may also work. The coil must have connecting leads and these should be secured with *room temperature vulcanizing* (RTV) adhesive, wax, or something similar to prevent breaking that may render the coil useless. A soft iron core of 1 to 2 feet is inserted through the coil and serves to offer a lower reluctance to the magnetic flux lines. It is these lines that, as they change in the core, induce a voltage due to magnetic induction. Note to shield the coil from electrical and capacitive pickup by wrapping it in aluminum foil and connecting the foil to the circuit ground. A special custom-made coil is available from Information Unlimited and is noted in the parts list.
2. Lay out and identify all the parts and pieces, and check them with the parts list. Note that some parts may sometimes vary in value. This is acceptable as all components are 10 to 20 percent tolerant unless otherwise noted.
3. Assemble the board by inserting components into the board holes as shown in Figure 36-5.

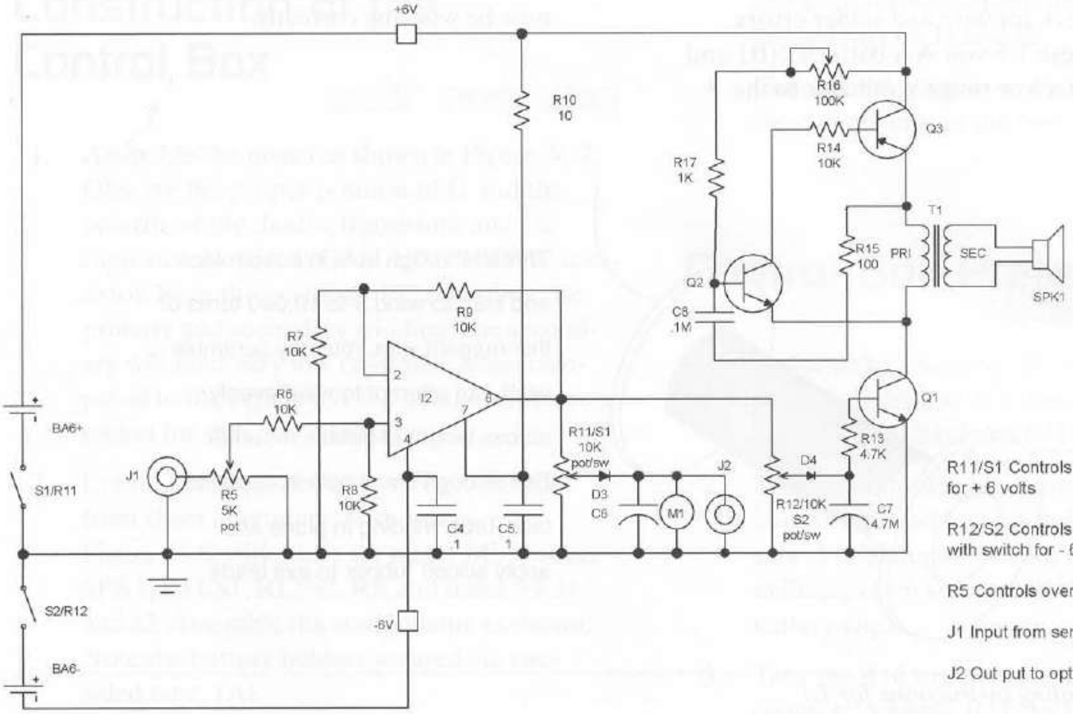

R11/S1 Controls meter sensitivity with switch for + 6 volts

R12/S2 Controls alarm threshold sensitivity with switch for - 6 volts

R5 Controls over all sensitivity

J1 Input from sense head

J2 Out put to optional chart recorder

Figure 36-3 *Control box schematic*

Proceed from left to right, attempting to obtain the layout. Note the polarity of the components as indicated. Certain leads of the actual components are to be used for connecting points and circuit runs. Do not cut or trim them at this time. It is best to temporarily fold the leads over to secure the individual parts from falling out of the board holes for now. Solder as required and clip off excess leads. Some kits contain a length of insulated wire. This must be cut, stripped, and tinned according to where it is used.

4. Wire in the 6-foot length of shielded cable, being careful not to burn through the insulation and short out the cable.
5. Attach the battery clip HOLD1 using silicon rubber.
6. Attach the L1 coil to the PB1 perforated board as shown in Figure 36-5 and secure it with RTV adhesive. You will have to drill a hole for the iron core. The core should fit snuggly and can be glued in place in the final assembly.

Pretesting the Sense Head

1. Carefully check for wire and solder errors, and install fresh 1.5-volt AA batteries (B1 and B2). Connect a low-range voltmeter to the leads of the output cable and adjust the 25-turn pot (R4) for as close to a near-zero reading as possible. Finally, adjust the lowest range of the meter, preferably the 50 microamp range. This should be done outside away from power lines.
2. With the meter still connected, observe a reading when a piece of iron is moved near sensing coil L1 and moved about. If the piece of iron contains any magnetism, the effect will be enhanced.
3. Wrap the sense coil with several layers of aluminum foil and ground the foil to the negative circuit. Diodes D1 and D2 will require shading against light sources as the junction behaves like photodiodes at these low signal levels.
4. Create enclosure EN1, as shown in Figure 36-6, from an $8 \times 3^{1}/_{2}$-inch outer-diameter piece of plastic tubing.

   Note the small, $^{3}/_{8}$-inch access holes for adjusting R4 are drilled when all is in place. Create the brackets as required for mounting the sense head.
5. Install the assembly into EN1 with caps CP1 and CP2 as shown. Recheck it by connecting the meter across P1. Readjust R4 through the access hole if needed. The sense head should now be working correctly.

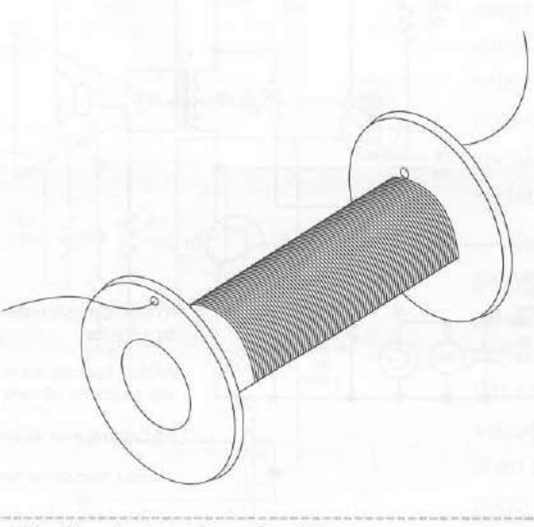

Thread through hole in bobbin face and start to wind 5 to 10,000 turns of thin magent wire.You may scramble wind, but attempt to wind evenly across length of bobbin mandrel.

Exit through small hole in opposite face.Tape winding in place and apply silicon rubber to exit leads.

Figure 36-4 *Winding instructions for L1*

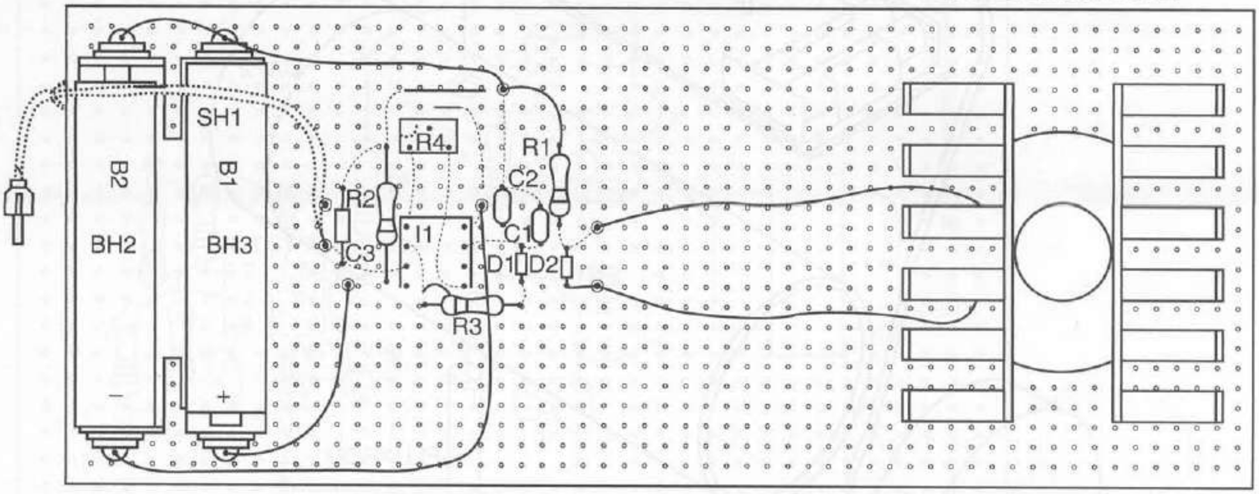

Strip off outer plastic jacket of cable 1 inch and twist braid into a lead and tin. Strip off 1/4 inch of insulation on inner lead and tin. Insert into plug and solder inner lead to lin tip.

Carefully solder twisted braid lead to outer shell being careful not to burn through insulation. Always verify lead continuity and possible shorting of the leads with a meter.

⋯ Dashed lines are connections on underside of perforated circuit board using the component leads. These points can also be used to determine foil runs for those who wish to fabricate a printed circuit board.

● Small dots are holes used for component insertion.

⊙ Circles with dots are holes that may require drilling for component mounting and points for external connecting wires, indicating strain relief points. It is suggested to drill clearance holes for the leads and solder to points beneath board.

Note: The sense head as shown in this sketch is wound from a transformer bobbin. Any plastic bobbin 2"x1" will usually suffice. See Figure 36-4.

Figure 36-5 *Assembly of the sense board section*

Construction of the Control Box

1. Assemble the board as shown in Figure 36-7. Observe the proper position of I1 and the polarity of the diodes, transistors, and C6 capacitor. Note that Q3 is a PNP PN2907 transistor. Note the position of Tl regarding the primary and secondary winding. The secondary will read very low resistance when compared to the primary. It is a good idea to use a socket for the integrated circuit.
2. Create the enclosure base section (BASE) from sheet aluminum, as shown in Figure 36-8, with holes for meter MI, speaker SPK1, R11/S1, R12/S2, R5, and holes for J1 and J2. Assemble the components as shown. Note the battery holders secured via two-sided tape, TAl.
3. Attach the interconnecting leads, as shown in Figure 36-9. Twist the leads whenever possible and keep them short and direct. Note the ground lugs on J1 and J2.
4. Connect the leads from the board assembly to the components in the box.

Control Box Pretesting

1. Position the board, but do not secure it at this point. Turn S1 and S2 off and insert four AA batteries into BH2 and BH3.
2. Turn on both switches, unsnap one side of the battery clip, and insert a milliamp meter. Measure .5 milliamps (−6 volts) at BH2 and .8 milliamps (+6 volts) at BH3. Reconnect the battery clips.
3. Turn the R16 trimpot to midrange, R11 full counterclockwise, R12 clockwise, and note the meter remaining at zero and the alarm

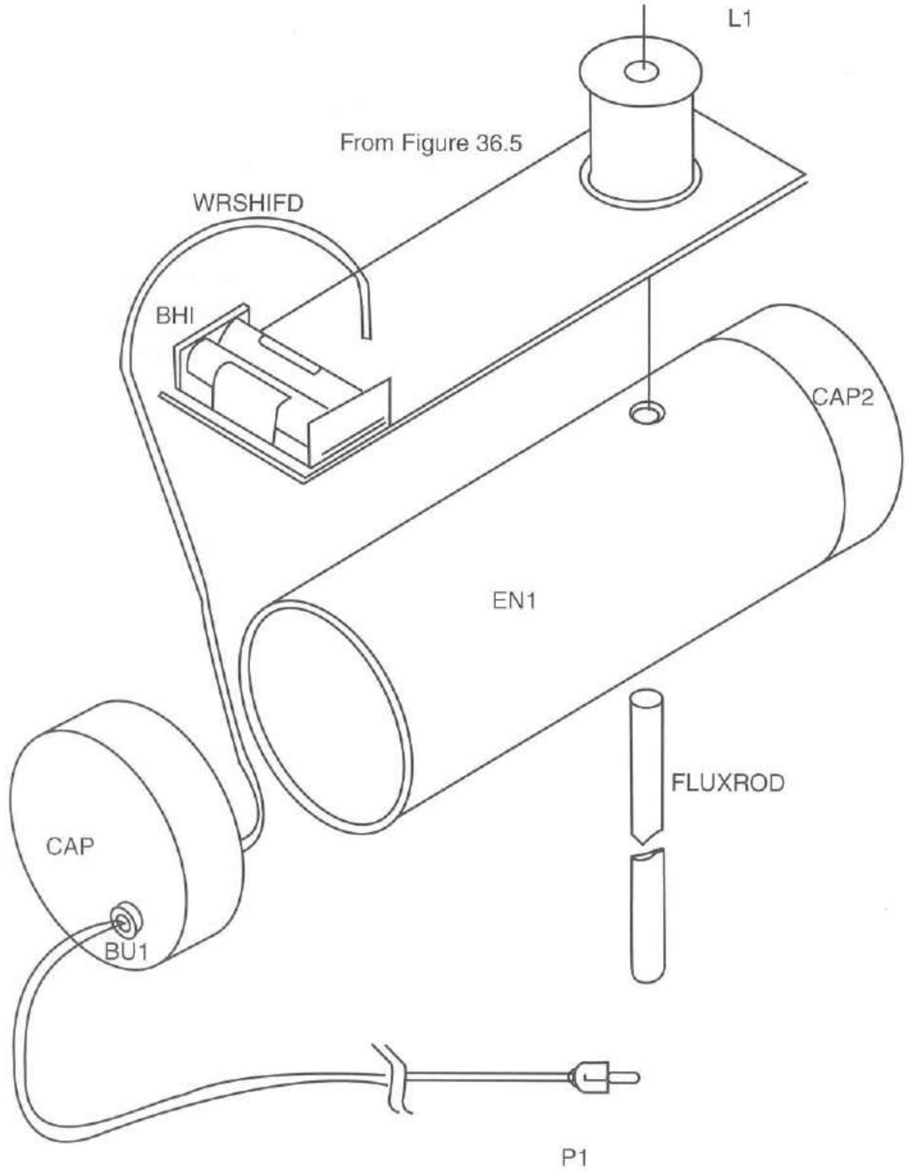

Figure 36-6 *Sense head enclosure assembly*

remaining silent. Now turn R11 clockwise and note the meter deflecting to full scale.

The following steps will require the sense head located away from electrical wiring and other potential sources of magnetic interference. The distance should be at least 5 feet from the control box to avoid feedback from the meter and speaker magnets.

4. Turn both R11 and R12 full clockwise. Plug in P1 from the sense head J2. Note the meter may pin, or jump off scale, and the alarm may sound. Have a friend slowly rotate R4 on the sense head to attempt to bring the meter into range. You may have to desensitize the circuit by turning R11 counterclockwise until the meter reads midrange and turn R12 to deactivate the alarm as it can be annoying during testing. Note that R5 is an overall sensitivity control and should be turned fully on in most applications for maximum sensitivity. The meter and alarm can be ranged by their respective controls, R11 and R12.

   When properly set, the reading will constantly fluctuate and is the result of sunspots or other phenomena distorting the earth's magnetic field.

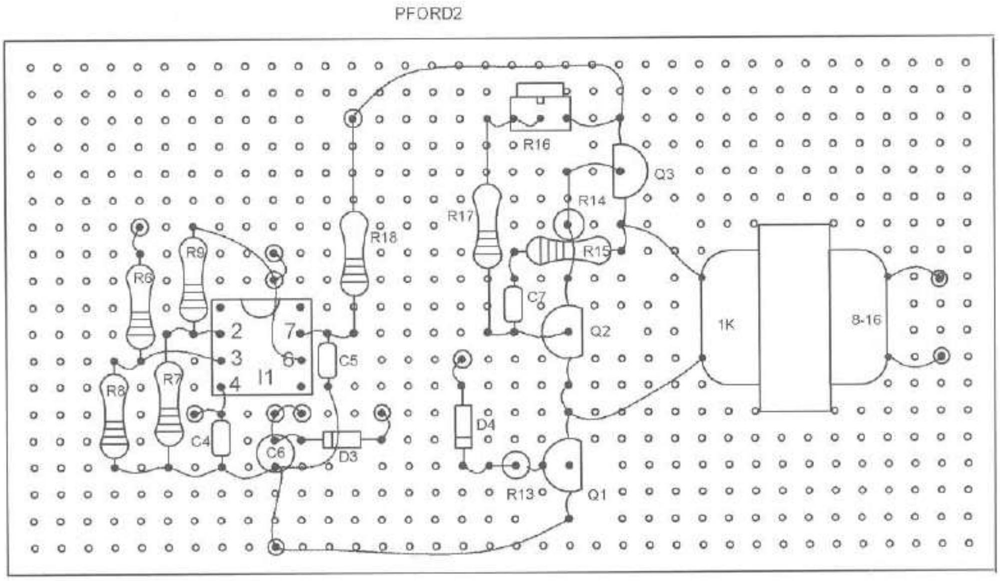

Dashed lines are connections on underside of perforated circuit board using the component leads. These points can also be used to determine foil runs for those who wish to fabricate a printed circuit board.

• Small dots are holes used for component insertion.

⊙ Circles with dots are holes that may require drilling for component mounting and points for external connecting wires indicating strain relief points. It is suggested to drill clearance holes for the leads and solder to points beneath board.

Figure 36-7 *Assembly control board*

If you choose a magnetically noisy location, you will have an added signal as the result of power lines, motors, and other magnetic disturbance. Note that if R11 cannot bring the meter into range without overly desensitizing it, it may be necessary to internally readjust R4 for proper ranging.

5. Preset R14 at midrange and adjust R12 until the alarm sounds. Readjust R16 for a clean, desirable-sounding tone if necessary.

   Appreciating the sensitivity of this unit requires operation in a magnetically quiet location where all the controls can be advanced fully. Under the conditions of a low magnetic background, the unit can detect automobiles, manmade devices, and natural magnetic disturbances. Also, simply moving a piece of iron near the sensing head should produce an indication. (Note low-level-detection sensitivity may be difficult during the peak of the sunspot cycle due to the high background signal fluctuations.)

   Jack J2 may be used to drive an XY recorder when observing sunspots or other time-related disturbances.

Note: No steps were taken to filter out 60 HZ disturbances, as it appears that many uses of this device involve the detection of power lines and other related equipment.

Applications

The uses of this device are obviously only appreciated by those who can understand its sensitivity to a minute change in magnetic flux. The unit can be made to detect the motion of large steel bodies as they distort the existing earth's magnetic field. Many applications and experiments are therefore possible with the device. Any high-impedance headphones

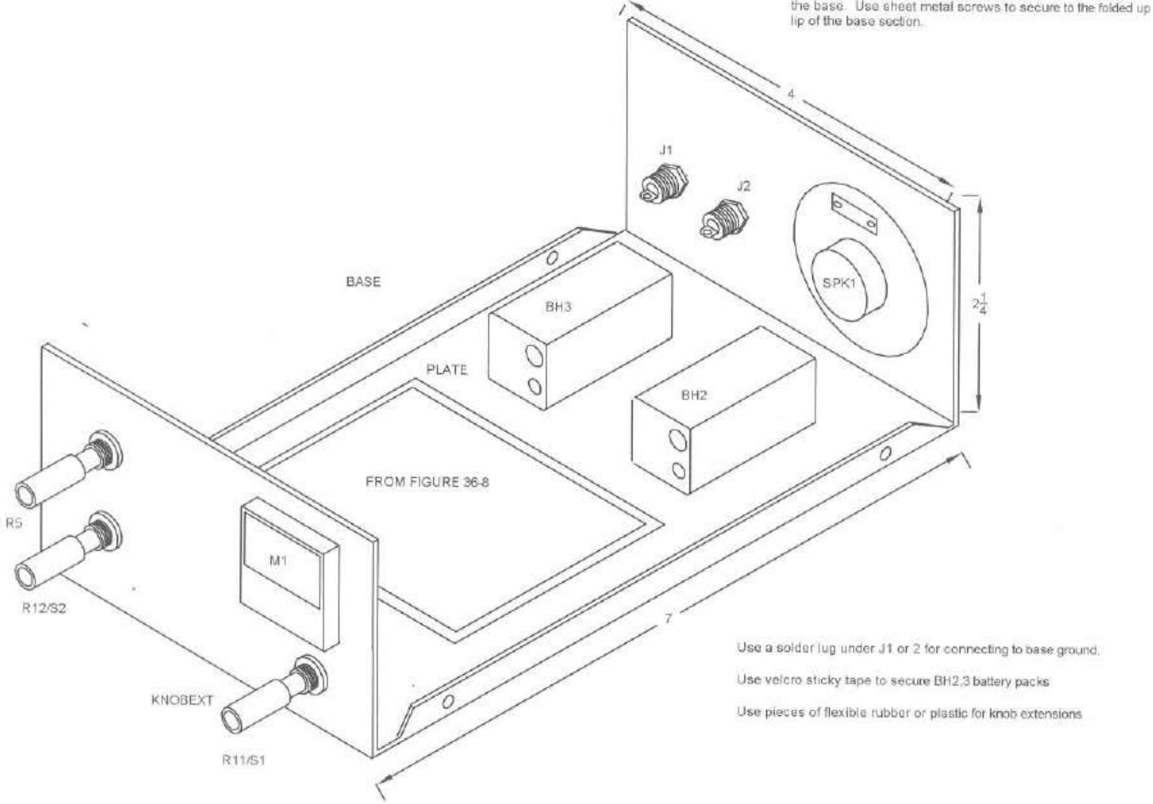

Fabricate a cove to fit the base section from any suitable material. Allow a over hang over the front and rear panels of the base. Use sheet metal screws to secure to the folded up lip of the base section.

Use a solder lug under J1 or 2 for connecting to base ground.

Use velcro sticky tape to secure BH2,3 battery packs

Use pieces of flexible rubber or plastic for knob extensions

4
J1
J2
SPK1
2¼
BASE
BH3
PLATE
BH2
FROM FIGURE 36-8
R5
M1
R12/S2
7
KNOBEXT
R11/S1

Figure 36-8 *Control box base section fabrication and layout*

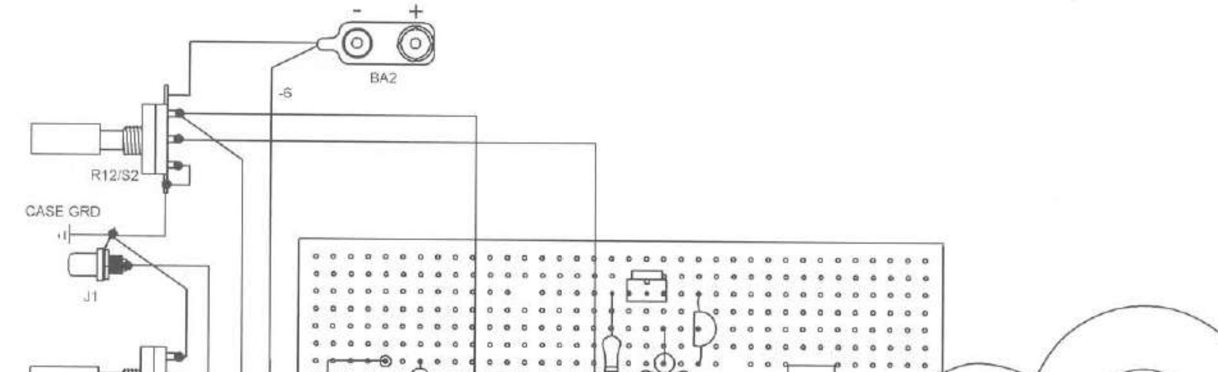
-
+
BA2
-6
R12/S2
CASE GRD
J1
R5
CASE GRD
J2
2
8
4
5
CASE GRD
SPK1 SPEAKER
R11/S1
-
+
BA1
+6
+
-
M1 METER

Figure 36-9 *Final assembly control box*

can be plugged into J2 where the experimenter can actually "listen in" and hear the field changing.

One verifying test that is easily done to determine the system's ability to detect the earth's magnetic field is the following. Rotate the sensing head with short, quick jerks in a clockwise direction. Note the meter pinning, and as rotation is continued a point will be found where the meter will respond less. This nulling point should occur with the fluxrod pointed in a north/south direction. If rotation is continued, it will be found that the short, quick jerks must occur in a counterclockwise direction to obtain a positive meter movement. This is similar to an electrical generator using the earth as its field magnet and the sensing head core as the moving armature.

Optional Ultra-Sensitive Vibration and Tremor Detector

You may easily convert this device to a vibration detector by simply modifying the sense head so it can be mounted to a rigid object such as a rock. The trick now is to place a magnet adjacent to the end of the sense coil, L1. The magnet should be held in place by elastic or other means whereby it is mechanically isolated relative to the sense coil. It should be very close to the sense coil for optimum sensitivity.

The theory is simple. Positioning a powerful magnet near a coil with many turns produces a signal. It is now the relative motion between this magnet and the coil that produces a voltage by the familiar equation $E = Nd0/dt$, where E is volts, N is the coil turns, 0 theta is the flux of the magnet, and t is the rate of relative motion in seconds. The combination of the high-resultant signal volts and the amplification sensitivity of the circuit provides amazing results.

Many variations are possible with the common objective being to obtain the highest reading for a given tremor. This unit will easily detect earthquakes occurring anywhere on the planet; however, the builder must experiment with motion dampening and other schemes to differentiate between manmade vibrations in order to isolate these signals. This is an excellent, low-cost project for any interested individual to learn by experimentation.

Table 36-1 Magnetic Distortion Detector Parts List

Ref. #	*Description*	*DB Part #*
R1, 3	Two 1.5M, ¼-watt resistors (br-gr-gr)	
R2	Two 10M, ¼-watt resistors (br-blk-bl)	
R4	100-kilo-ohm, 10-turn vertical trimmer	
R5	5K linear pot	
R6, 7, 8, 9, 14	Five 10K, ¼-watt resistor (br-blk-or)	
R10	10-ohm, ¼-watt resistor (br-blk-blk)	
R11/S1, 12/S2	Two 10K linear pots and switches	
R13	4.7K, ¼-watt resistor (yel-pur-red)	
R15	100-ohm ¼-watt resistor (br-blk-br)	
R16	100-kilo-ohm vertical trimmer	
R17	1-kilo-ohm, ¼-watt resistor (br-blk-red)	
C1, 2, 3, 4, 5, 8	Six .1-microfarad, 50-volt plastic capacitors	
C6	1-microfarad, 50-volt vertical electrolytic capacitor	
C7	4.7-microfarad, 50-volt vertical electrolytic capacitor	
I1	LM4250 nano-amp operational amplifier in a dual inline package (DIP)	
I2	LM741 operational differential amplifier in DIP	
D1, 2, 3, 4	Four IN914 small-signal diodes	
Q1, 2	Two PN2222 NPN GP silicon transistors	
Q3	PN2907 PNP GP silicon transistor	
TI	1K to 8-ohm, small audio transformer	
M1	50 uamp panel meter	DB# METER50S
SPK1	2-inch, 8- to 16-ohm speaker	
J1, 2	Two RCA panel-mount phono jacks	
P1, 2	Two RCA mating phono plugs	
CL1, 2	Two 9-volt battery clips	
SH1	3 feet of single-conductor, shielded mike cable	
WIRE24	4 feet of #24 vinyl stranded hookup wire	
BH1	Two-AA-battery holder	
BH2, 3	Two 4-AA-battery holders	
BU1	Small, plastic feed-through bushing	
L1SENSE	Sense coil per Figure 36-4	DB# SENSE
PBORD1	6¼ × 2½-inch, .1 × .1-inch grid perforated vector board	
PBORD2	4½ × 2-inch, .1 × .1-inch grid perforated vector board	
EN1	7½ × 3-inch schedule 40 PVC tubing	
CAP1	Two 3½-inch black plastic caps	
TAP1	4 inches of 4 × 1-inch strips of Velcro sticky tape	
KNOBEXT	4 inches of 3/16-inch inner-diameter plastic or rubber tubing for shafts	
FLUXROD	6 inches of ¼ × 20-inch steel-threaded rod and two mating nuts	
BASE	Created as shown in Figure 36-8 from .063 aluminum	
COVER	Cut to fit BASE section	
PLATE	5 × 2½-inch piece of insulating plastic to mount under board	

Chapter Thirty-Seven

Body Heat Detector

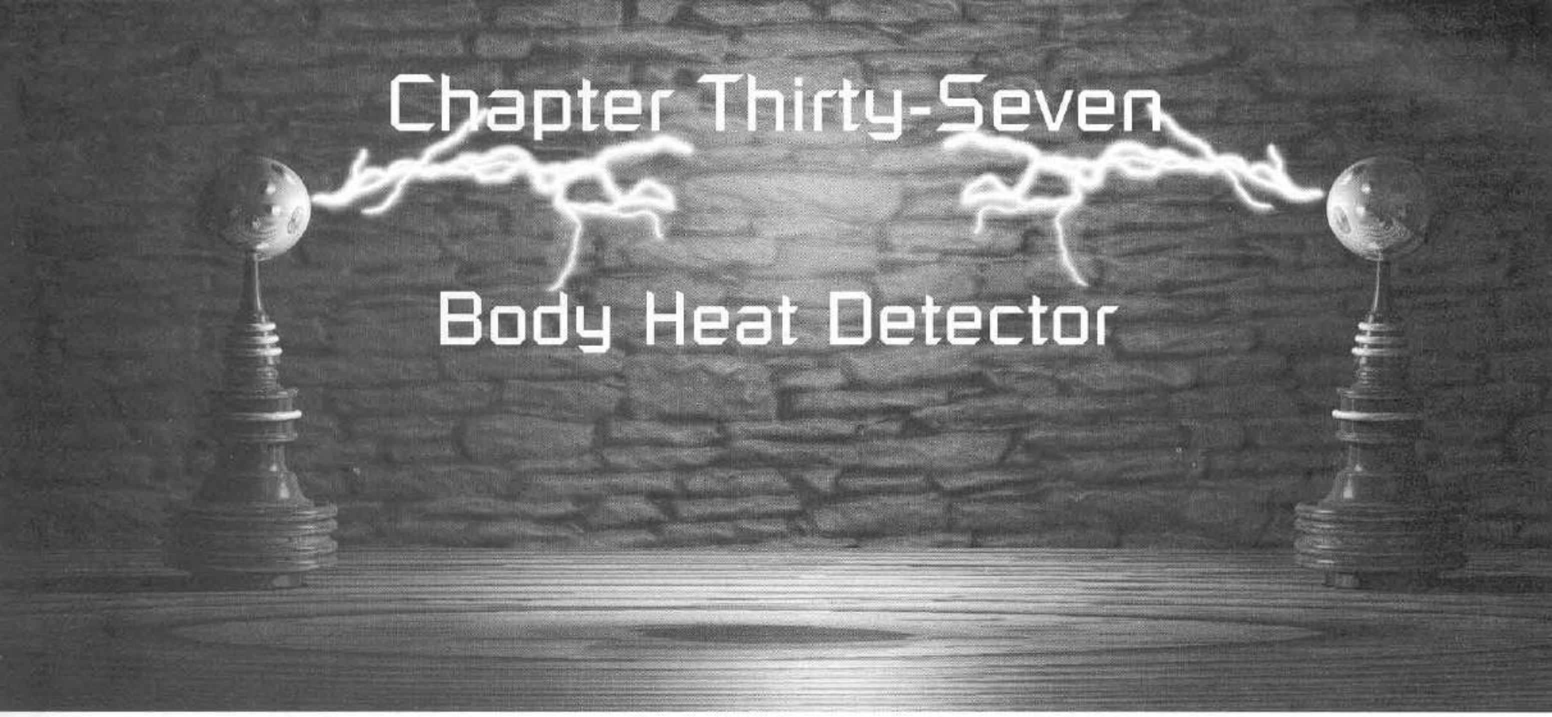

This project, as shown in Figure 37-1, shows how to construct a useful device that can detect human or animal motion up to several hundred meters. It provides a narrow field of view and can be used for paths and walkways. Farmers use these devices to detect when a predator comes into range, taking the necessary action when the alarm sounds. It may also be used as a medium-sensitive thermal detector for lab and research use. The device is an excellent science project demonstrating medium-sensitivity detection of relatively low temperatures.

This is an intermediate-level project requiring basic electronic skills. Expect to spend $30 to $45. All the parts are readily available, with specialized parts obtainable through Information Unlimited at www.amazing1.com. These are listed in the parts list at the end of the chapter.

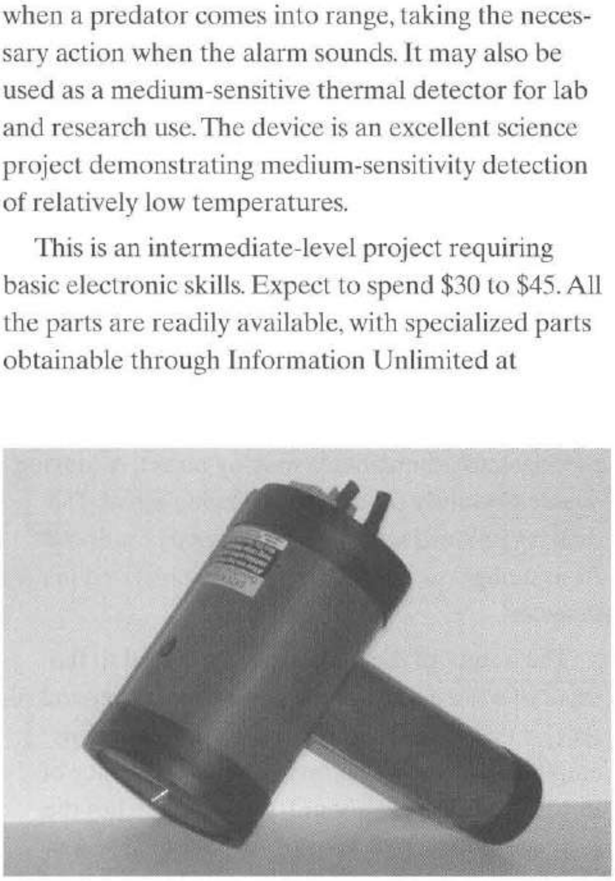

Figure 37-1 *Body heat detector*

Description

Your body heat detector is a device capable of sensing an object at a temperature increase slightly over ambient. It is this property that allows it to detect a living body at a considerable distance. The sensitivity of the device corresponds to a black body wavelength of 10 microns, corresponding to a temperature of 100° Fahrenheit. Black-body radiation defines color temperature in wavelengths below what is visible to the naked eye. All objects above absolute zero radiate energy as a function of their temperature to the fourth power. This radiation decreases in wavelength as the radiating body becomes hotter. This radiation will consist of the visible color spectrum for a given range of temperatures. Hotter bodies will show visible color such as blue being on the hotter side of red. A body hot enough to glow red is around 900 degrees F. Black bodies radiate in the infrared (invisible) and lower as the generating temperatures are near ambient being above 70 degrees F. All objects above absolute zero radiate energy as a function of the fourth power of temperature, and body heat energy is in the infrared region at around 100° degrees F.

The applications for this device are numerous and as an example you could a create motion detection system for home security. These are normally short-range, wide-angle detection systems and are limited in performance. Our system would be mounted similar to an optical scope on a rifle or a tripod. A living target now can easily be ferreted out in darkness or under moderate cover due to heat reflections. A predator, as it approaches its target victim, is easily detected, alerting the farmer or rancher to take the next step of action. Garden pests, as they emerge from their burrow, and downed game can easily be detected by the device.

The ability to detect hot spots is an extremely useful function. An example of sensitivity can easily be demonstrated by the following. Obtain a piece of cardboard or plastic approximately 6 × 12 inches. Place your hand on one end of this piece for several seconds to produce a hot spot of your hand print. Then place the piece in front of the unit, noting a detected signal increase on the surface previously touched by your hand due to the heat stored. (Don't touch the other end of the piece). Note that as time continues, the hot spot dissipates into the object thereby reducing the temperature differential and the signal pickup on the cardboard piece becomes evenly distributed. The differences in temperature detection capability is in the millidegrees. You can now realize the many useful applications of this unit.

Body Heat Detector Modes

Your body heat detector operates in two modes:

- **MODE A:** This mode is used to detect moving bodies such as animals, people, and so on. It is intended for sensing intrusions along pathways and into areas where a narrow field of view is necessary at long ranges (100 meters or so). This mode is the most stable of the two and rejects steady background sources of thermal energy. Detected bodies, however, must be moving.
- **MODE B:** Operation in this mode allows the detection of stationary hot spots and is most useful with a "cold background" such as at night. It is useful for ferreting out thermal sources such as hidden persons and animals. Hot object spots and areas are easily detected, but unfortunately this mode is so sensitive that often signals appear to be erroneous and mask the target object.

Both modes of operation require a camera tripod for reliable and simple operation. Very often a panning action can be used to reject or detect the desired targets.

Circuit Description

The heart of the detector is a pyroelectric detector (PYR1) capable of detecting very minute temperature changes. This device is responsive to temperatures of 100° Fahrenheit equivalent to approximately 10 microns. Unfortunately, its response speed is relatively slow, peaking at about .1 Hz. This is shown in Figure 37-2.

The pyroelectronic effect utilizes dual crystals of lithium tantalum oxide (LiTa 02). These crystals are polarized in an electric field below the Curie point. Electrodes attached to the crystal become electrically charged and eventually neutralize due to the surroundings. When the neutralized crystal is exposed to an infrared signal, its temperature changes, producing a charge on the electrodes and hence a detected signal. This signal is amplified by a built-in, low-noise *field effect transistor* (FET) and becomes the output. It is important to note that only a changing infrared signal can produce output; therefore, a chopper must be used for stationary sources. A chopper is nothing more than a mechanical wheel that has vanes to periodically interrupt the heat radiation signal to the pyrodetector, simulating a moving object. A moving source obviously produces a changing signal. The dual crystals will also cancel out signals produced from sunlight as they are oppositely polarized in the detector.

The output of the detector is connected to the input of a low-frequency, dual-stage amplifier and filter (U1) with a gain of approximately 2,500. This amplifier is designed to respond to a frequency of .1 to 10 Hz, peaking at about 1 Hz. This matches the response of the pyrodetector and the signal commonly generated by a living, moving body in the field of detection.

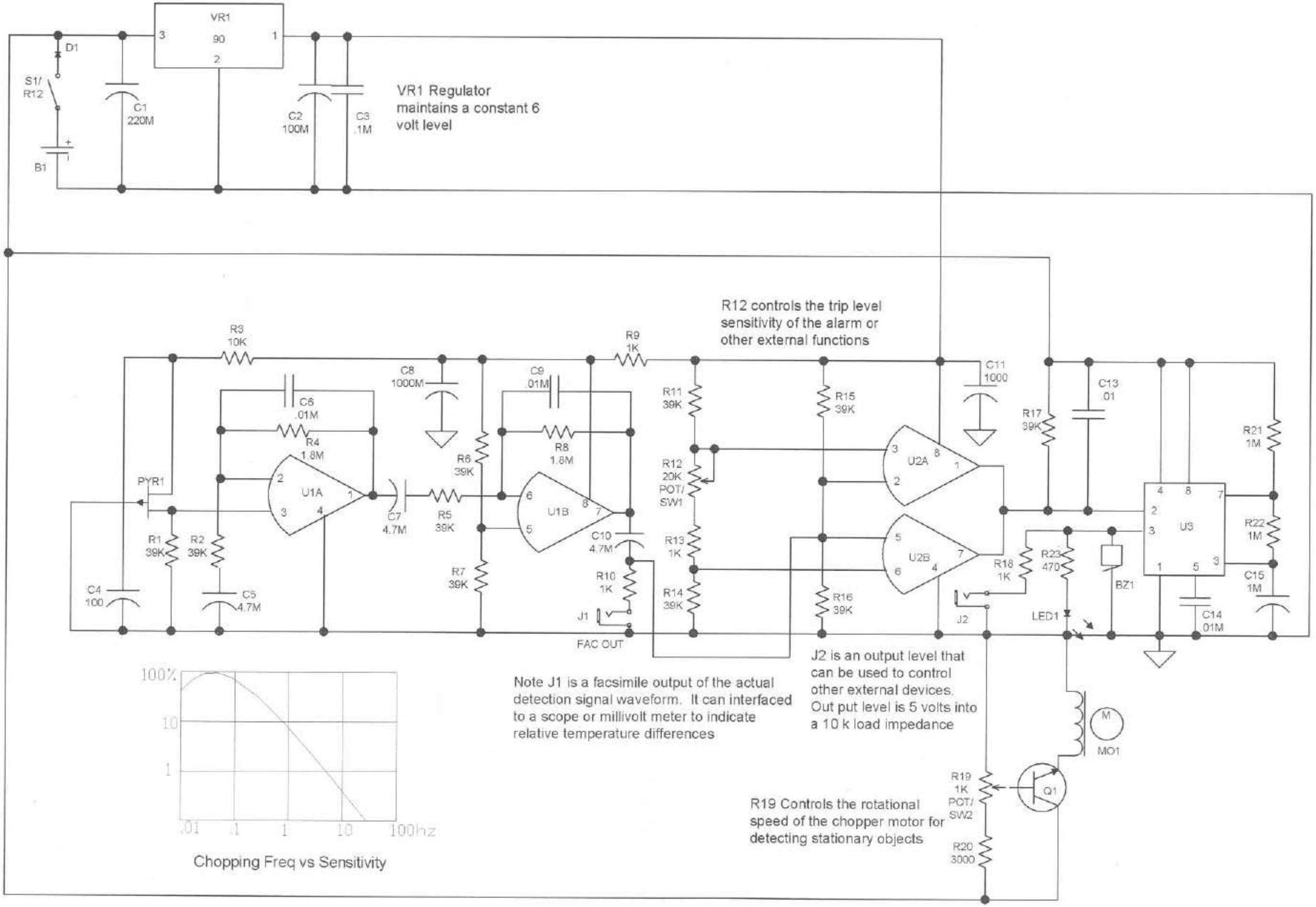
VR1
90
D1
S1/
R12
B1
C1
220M
C2
100M
C3
.1M
R3
10K
C6
.01M
R4
1.8M
PYR1
U1A
R1
39K
R2
39K
C4
100
C5
4.7M
C7
4.7M
R5
39K
C8
1000M
R6
39K
R7
39K
C9
.01M
R8
1.8M
U1B
C10
4.7M
R10
1K
J1
FAC OUT
R9
1K
R11
39K
R12
20K
POT/
SW1
R13
1K
R14
39K
R15
39K
R16
39K
U2A
U2B
C11
1000
R17
39K
C13
.01
R18
1K
R23
470
J2
LED1
BZ1
U3
R21
1M
R22
1M
C15
1M
C14
.01M
MO1
M
Q1
R19
1K
POT/
SW2
R20
3000
100%
10
1
.01
.1
1
10
100hz
Chopping Freq vs Sensitivity

VR1 Regulator maintains a constant 6 volt level

R12 controls the trip level sensitivity of the alarm or other external functions

Note J1 is a facsimile output of the actual detection signal waveform. It can interfaced to a scope or millivolt meter to indicate relative temperature differences

J2 is an output level that can be used to control other external devices. Out put level is 5 volts into a 10 k load impedance

R19 Controls the rotational speed of the chopper motor for detecting stationary objects

Figure 37-2 *Circuit schematic*

The changing AC output of the dual-signal amplifier (U1A and U1B) is capacitive coupled to output jack J1. It is this signal that is measured by an external AC millivolt meter and provides an analog indication of signal strength. This signal may be further processed for ultra-low-sensitive detection or maybe demodulated where the resulting DC level can control a VCO for audible detection via a changing tone. This AC signal is also fed to the dual plus and minus ("window") comparator (U2A and U2B). A threshold control (R12) presets the trip voltage or sensitivity of activation. The output of these two comparators is "anded" together and fed to the trigger input of the timer (U3). The output of U3 now activates audible buzzer BZ1 and visual indicator LED1. The buzzer on time is controlled by timing resistors R21 and R22 and capacitor C15.

An optional motor speed control for the chopper consists of Q1 and speed control pot R19. Jack J2 provides a 5-volt level through 1K resistor R18 and is used to trigger other systems.

Assembly Instructions

1. Lay out and identify all the parts and pieces, and verify them within the parts list at the end of the chapter.
2. Examine the PC board, as shown in Figure 37-3 and 37-4.
3. Proceed to assemble the device by inserting and soldering all the fixed resistors.
4. Insert and solder the capacitors, carefully noting the polarity on all the electrolytics: C1, C2, C4, C5, C7, C8, C10, C11, and C15. The other capacitors are not polarized.
5. Insert and solder in the semiconductors (D1 and VR1) and Q1. Note the polarity.
6. Identify and note the position marks on the integrated circuits U1, U2, and U3. Insert and carefully solder them, avoiding solder shorts between the pins. It is suggested to always use sockets for these components.
7. Identify the pots R12 and R9 and solder them in place. Note the switch section, S1, of R12. These connections are made using short pieces of cut leads to the pads directly beneath. Note these pots must be properly installed to allow final alignment with the PANEL plate.
8. Insert the wire jumps using short pieces of the cut leads of components that were used priorly.

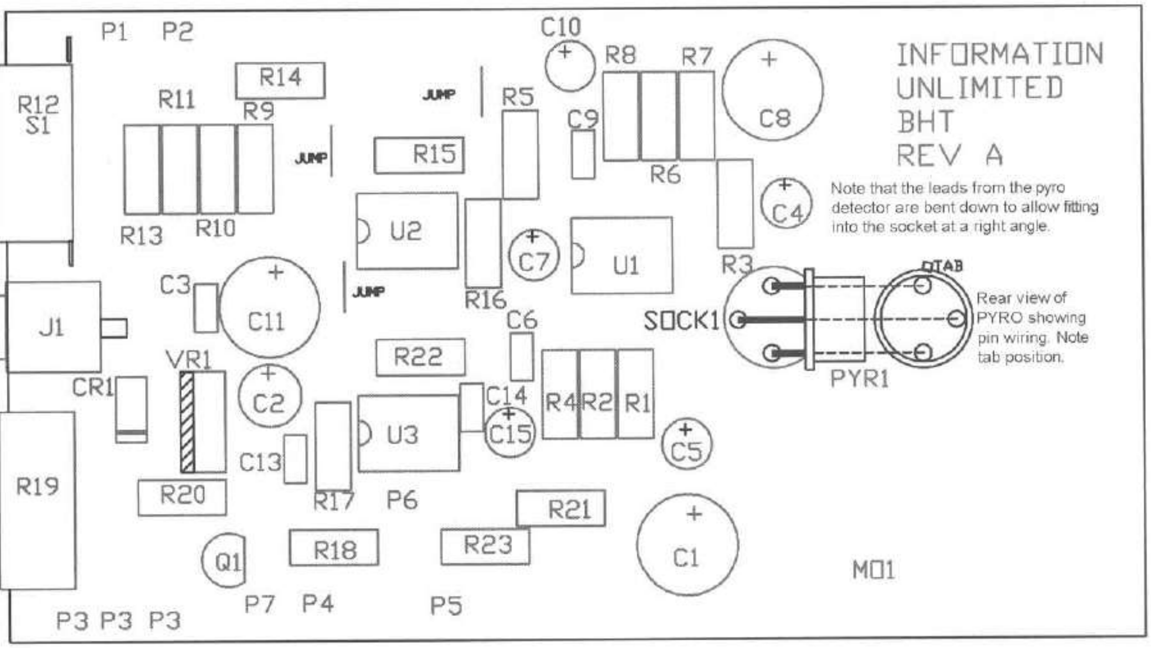

Note that the leads from the pyro detector are bent down to allow fitting into the socket at a right angle.

Rear view of PYRO showing pin wiring. Note tab position.

Figure 37-3 *PCB wiring and assembly*

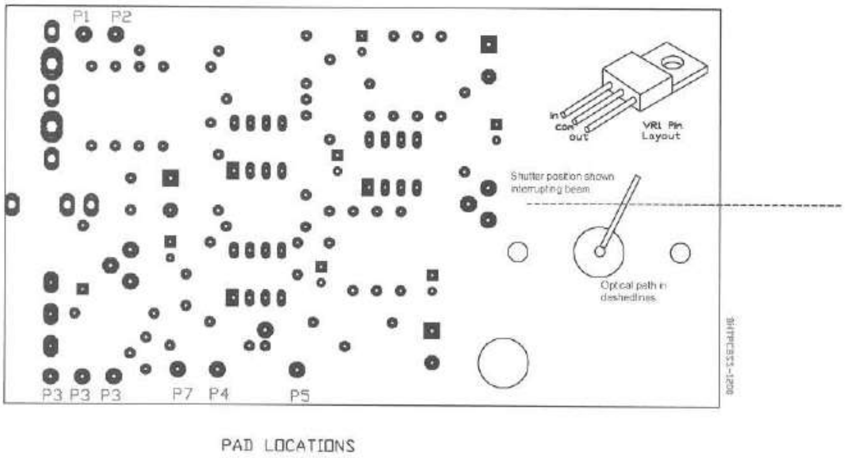
P1 P2
P3 P3 P3
P7 P4
P5
in
con
out
VR1 Pin Layout
Shutter position shown interrupting beam
Optical path in dashed lines
PAD LOCATIONS

Figure 37-4 *Pad and connections*

9. Insert J1 and note the rear connection lead requiring a short piece of bare wire to connect to the board pad directly beneath.
10. Insert the SOCK1 pyrodetector socket as shown. Carefully form the three leads of the pyrodetector at 90°. Do not sharply bend it. Note the proper leads, and position the tab as shown. This part must be at the optical center of the system for proper operation.
11. Connect the BZ1 buzzer using its leads, as shown in Figure 37-5.
12. Connect output control jack J2 and indicator light-emitting diode (LED1) using a 3-inch length of #22 vinyl wire. Note the cathode lead of LED1 directly connecting to the ground lug of J2. Note these parts eventually will be mounted to the panel.
13. Connect the leads from the chopper motor MO1 and battery clip CL1 as shown.
14. Create the panel, as shown in Figure 37-6. Attach it to the PC board via the two pots and the jack. The panel and PC board should be flush to one another. Attach the remaining J2 and LED1 to the panel. Note to use a bushing for the LED.
15. Create the shutter flap (FLAP), as directed in Figure 37-6.
16. Recheck all wiring for shorts, accuracy, and correctness of parts. Also check for solder joints. You are now ready to bench test the assembly.
17. Connect a 9-volt supply to the battery clip.
18. Rotate R12 to on, and note the shutter flap rotating with R19 full counterclockwise. It should slow down as R12 is turned in a clockwise direction. This rotation should be variable in the range of 10 *revolutions per second* (rps), down to stalling at less than 1 rps. The motor will stall if there is any obstruction.

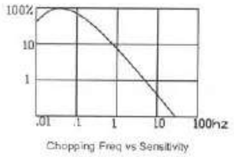

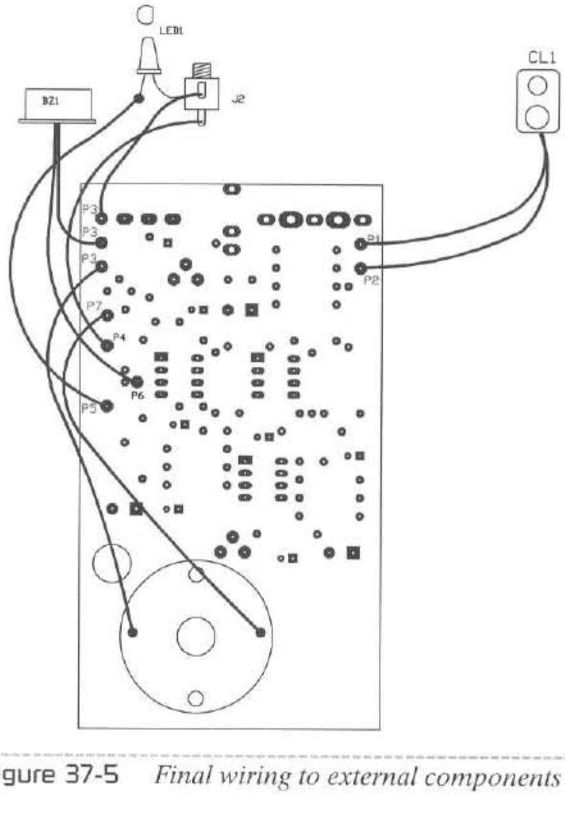

Figure 37-5 *Final wiring to external components*

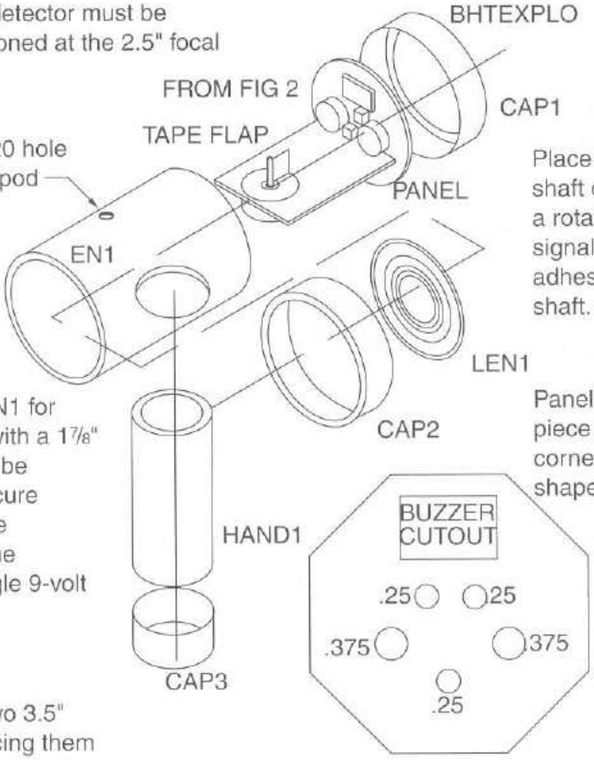

Note that the pyro detector must be centered and positioned at the 2.5" focal length of the lens.

Drill and tap a 1/4-20 hole for mounting to a tripod screw.

Place a small "flap" on to the shaft of the chopper motor to act as a rotating shutter for breaking up the signal path. Use a small piece of adhesive metal tape folded onto the shaft.

Note that hole in EN1 for handle is best cut with a 1⅞" circle saw. Fit must be tight to properly secure handle in place. The handle serves as the housing for the single 9-volt or 8 AA cells.

Panel is cut from a 3 1/4 x 3 1/4" piece of .035 aluminum. Cut corners to approach a circular shape.

Cut out center of two 3.5" plastic caps by placing them on the enclosure tube and cutting out the center section with an xacto knife using the inner wall as a guide.

Front Panel Fab

Note that holes must be accurately positioned for proper alignment to R12, R19, and J1 on the printed circuit board.

Figure 37-6 *Final housing setup*

Make sure the flap is clear and not touching the bearing bottom.

19. Connect a scope to J1 and note a changing level as your hand or another warm object is waved in view of the detector. Note the signal responding to a stationary source with the shutter flap or to some form of motion with the shutter flap stopped. Check the sensitivity control pot R12 and then verify a DC output level at J2, occurring when a body is detected. This function is for control of external alarms and alerts.
20. Check that the buzzer and light emitter activate when detecting a thermal source. Use your solder iron as a heat source and experiment with different sources.
21. Finally, assemble the device as shown in Figure 37-7 and proceed to use it as outlined in the instruction section.

Operating Instructions

1. Turn all controls fully counterclockwise to off.
2. Remove the battery compartment cover and slide out the eight-cell holder. Insert fresh batteries and reinstall.
3. Rotate the chopper control fully clockwise. View the shutter through the access hole and note how it is positioned in the figure.
4. Slowly turn the sensitivity control clockwise until the alarm sounds. If the unit does not respond, try moving your hand in front of it, noting its response. Turn the control back, just to the point where the unit does not respond.
5. Turn the chopper control until the unit starts to respond. You are now measuring the thermal background as indicated by the periodic sounding of the alarm as the shutter interrupts the heat path.

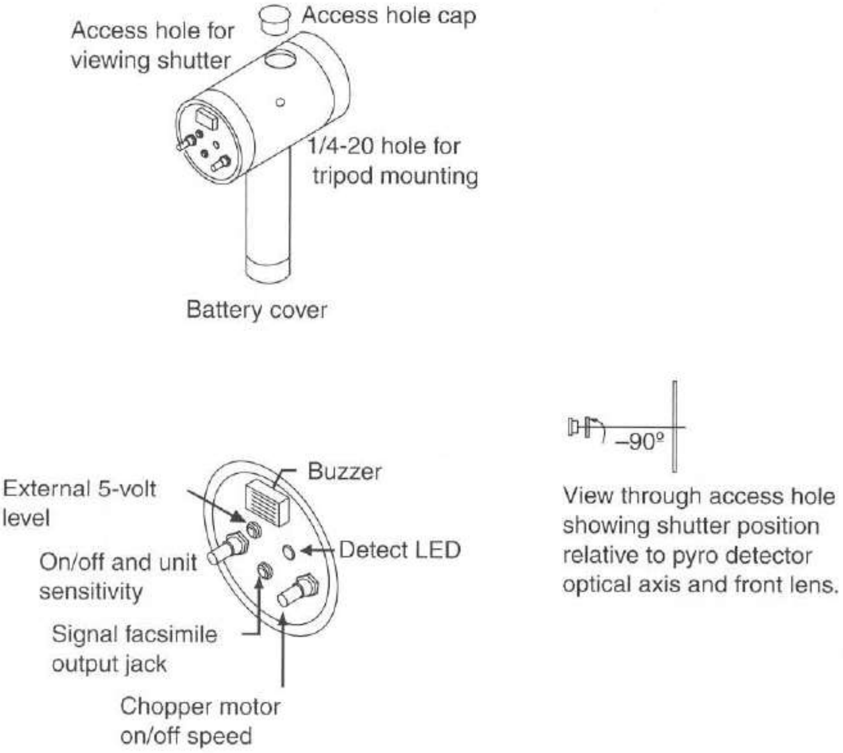
Access hole for viewing shutter
Access hole cap
1/4-20 hole for tripod mounting
Battery cover
External 5-volt level
Buzzer
On/off and unit sensitivity
Detect LED
Signal facsimile output jack
Chopper motor on/off speed
–90°
View through access hole showing shutter position relative to pyro detector optical axis and front lens.

Figure 37-7 *Final assembly view*

6. Connect a standard 3.5-millimeter plug into the 5-volt level jack and measure a positive 5-volt level every time the alarm responds. This output is through a 1K resistor and is intended for triggering other alarms or external functions.
7. Connect a standard 3.5-millimeter plug into the signal facsimile jack and connect it to a scope input set on the AC millivolt scale. Note the changing voltage level indicating the slightest amount of differential thermal energy. Check this level with and without the chopper and try it with various thermal sources, noting the voltage levels. This output port is intended for the lab technician desiring to explore low-level thermal sources. It requires a scope or a sensitive AC millivolt meter capable of measuring 10 to 100 millivolts.

The previous steps are performed without considering the thermal stabilizing of the unit or its immediate surroundings. Maximum sensitivity will not be obtainable under these conditions.

8. Select a target area and mount the unit on a tripod, using the 1/4-20 mounting hole located on the side of the unit. This scheme will place the handle in the horizontal plane (off to the side) but should not pose a problem. The access hole will be opposite the handle.
9. Point the unit at the target area and allow it several hours to thermally stabilize. Do not touch the enclosure and try not to get too close to the front of the system, as this can destabilize the system for low-level detection.
10. Point and pan the unit at some known source to verify the anticipated range of operation. Experiment with the chopper speed or no shutter to obtain the optimum results for your application.
11. Handheld operation will vary considerably due to the many variables in the thermal background. We can only suggest that you get accustomed by trial and error to determine your application.

Final Notes

The access hole for the chopper shutter is necessary in order to prevent the blockage of the thermal path when the chopper motor is off. There is no guarantee that the shutter will not randomly stop at the wrong place, blocking the detector and preventing normal operation. In such cases, the shutter must be manually moved with a pencil or some object.

The unit must be temperature stable to optimize maximum sensitivity. This usually takes about an hour if the unit has been recently handled or placed near radiant sources.

For actual body detection, the best performance will occur on a cold night. This does not mean that the unit will not work on a warm night, but that it might be less sensitive.

The detection of stationary objects may vary with chopper motor speed and background temperature. The maximum response of the pyrodetector is around .1 chops per second. This is very slow for many applications and we suggest increasing it to 1 to 3 per second.

The field of view may be narrow due to the Fresnel lens used, but you may change to other suitable lenses of polyethylene. Also, a panning action allows zeroing in on certain target subjects.

Table 37-1 Body Heat Detector Parts List

Ref.#	*Description*	*DB Part #*
R1, 2, 5, 6, 7, 11, 14, 16, 17	Ten 39K, ¼-watt resistors (or-wh-or)	
R3	10K, ¼-watt resistor (br-blk-or)	
R4, 8	Two 1.8-megohm ¼-watt resistors (br-gray-gr)	
R9, 10, 13, 18	Four 1K, ¼-watt resistors (br-blk-red)	
R20	4.7K, ¼-watt resistor (yel-pur-red)	
R21, 22	Two 1-megohm, ¼-watt resistor (br-blk-gr)	
R12/S1	20K pot and 12 VDC switch	
R19	500 to 1-kilohm pot	
R23	470-ohm, ¼-watt resistor (yel-pur-br)	
C1	220-microfarad, 25-volt vertical electrolytic capacitor	
C2, 4	Two 100-microfarad, 25-volt vertical electrolytic capacitors	
C3	.1-microfarad, 50-volt plastic capacitor	
C5, 7, 10	Three 4.7-microfarad, 25-volt vertical electrolytic capacitors	
C6, 9, 13, 14	Four .01-microfarad, 50-volt plastic capacitors (103)	
C8, 11	Two 1,000-microfarad, 25-volt vertical electrolytic capacitors	
C15	1-microfarad, 25-volt vertical electrolytic capacitor	
PC1	Printed circuit board (PCB)	DB# PCBHT
SOCK1	Three-pin TO5 socket	
Q1	PN2222 NPN transistor	
U1	LM358 dual-operational-amplifier integrated circuit	
U2	LM393 dual-comparator integrated circuit	
U3	LM555 timer integrated circuit	
VR1	7805 5-volt regulator TO220 package	
MO1	Low-voltage chopper motor	DB# CHOPMOT
PYR1	Pyroelectric detector	DB# PYR1
BZ1	Piezoelectric buzzer	
LED1	High-brightness light emitter	
J1, 2	Two 3.5-millimeter panel-mount jacks	
CL1	Battery clip and 9-inch leads	
PANEL	3.25 × 3.25-inch aluminum as shown	
EN1	6 × 3.5-inch OD schedule 40 PVC tubing as shown	
HANDLE	6 × 1.875-inch schedule 40 PVC tubing as shown	
CAP1, 2	Two 3 ½-inch plastic caps, as shown in Figure 37-6	
CAP3	2-inch plastic cap for retaining batteries in handle section	
LEN1	3-inch Fresnel polyethylene lens	DB# FRELEN3
BU1	Lens holder for LED1 or use a ⅜-inch bushing	
BHOLD1	8-AA-cell, 4 × 2 battery holder with snaps	
WR20B	24 inches #22 black vinyl hookup wire	

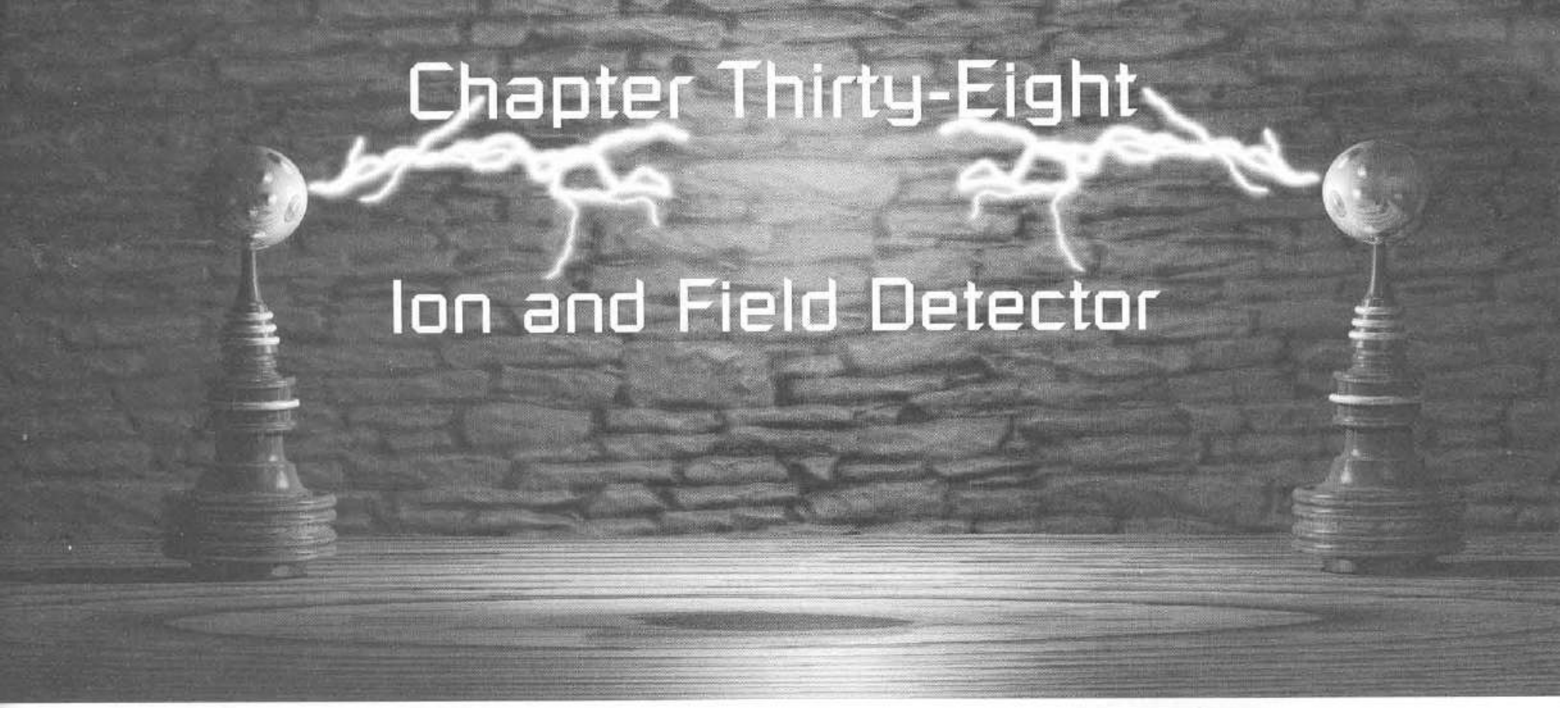

Chapter Thirty-Eight

Ion and Field Detector

This chapter focuses on a device capable of detecting and indicating the relative amount of free ions in the air (see Figure 38-1). It can determine the output of ion generators, high-voltage leakage points, static electrical conditions electric field gradients, and any other application where the presence of ions or a measurement of their relative flux density is required. It can also be used as a very sensitive lightning indicator, showing dangerous electric field levels.

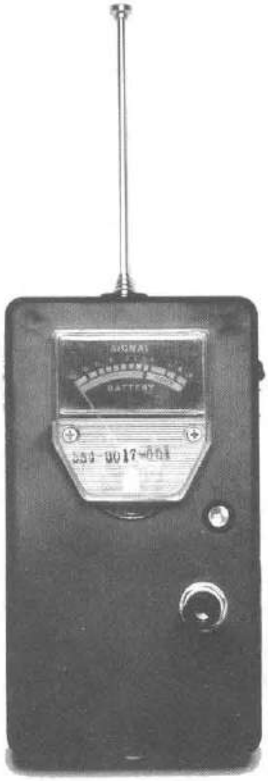

Figure 38-1 *Ion and field detector*

The unit is handheld and enclosed in a plastic enclosure about the size of a king-size pack of cigarettes. A sensitivity control with an on/off switch, a high flux indicator lamp, and a meter are located on the front panel. A telescoping antenna exits a hole in the top of the unit and is used as a moving probe or a stationary pickup point for these charge fields. Grounding the unit is a strip of metallic foil on the outside of the enclosure that contacts the user's hand or may be directly connected to earth ground via a wire. Please note that the performance of this unit can be seriously affected by high humidity conditions.

Circuit Description

As shown in Figure 38-2, ions are accumulated on the collector probe ANT1 and cause a minute, negative base current to flow in transistor Q1 connected as a PNP Darlington pair with transistor Q2. Capacitor C1 and resistor R1 form a time constant to eliminate any rapid fluctuations. The response of the time constant can be changed for various applications and is discussed in Figure 38-2. Diodes D1 and D2 prevent excessive voltage from destroying Q1 by clamping any transients to ground. The collector of Q2 is DC

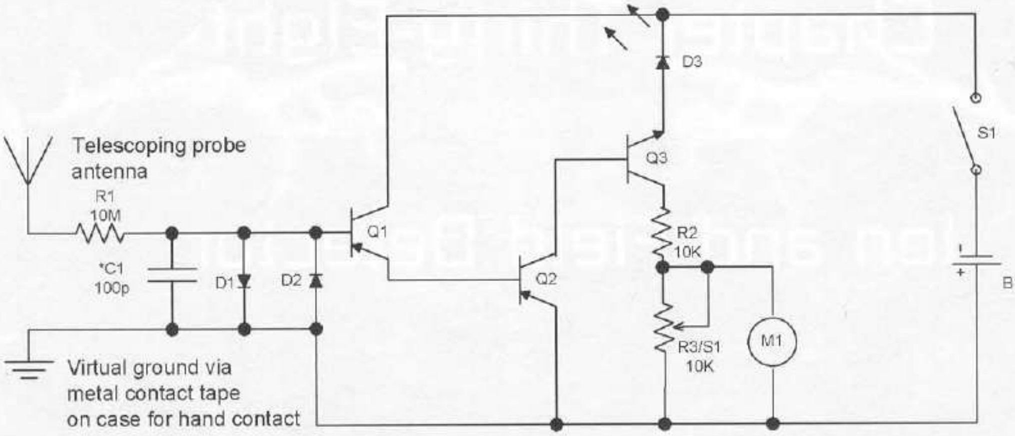
Telescoping probe antenna
R1
10M
*C1
100p
D1
D2
Q1
Q2
Q3
D3
R2
10K
R3/S1
10K
M1
S1
B1
Virtual ground via metal contact tape on case for hand contact

C1 is used to slow down the response and can be eliminated when detecting fast discharge fields such as lightning etc

Unit is shown wired for positive ion detection. You may change to negative by replacing the PN2222 transistors with PN2907 and the opposite. For a quick indication you may grab the antenna and use the body of the unit to detect the negative ions

Please note that the antenna probe on you unit is a telescoping antenna properly secured and electrically isolated. It is important to remember that any type of leakage around the input of QI can reduce the sensitivity. You may wish to coat the circuitry with a good quality varnish etc. Make sure unit is dry and clean before sealing.

Figure 38-2 *Circuit schematic*

coupled to the base of NPN transistor Q3 in series with current-limiting resistor R2 and meter sensitivity pot R3. Current flowing through Q3 now causes a voltage to develop across R3, driving the indicator meter M1. *Light-emitting diode* (LED) D3 is connected in series with the emitter of Q3 and serves as a visual indicator of strong charge fields. It should be noted that the meter serves only as a relative indicator of ion flux and is not totally that linear.

A threshold voltage, as a result of the base emitter, drops off Q1 and Q2 is only noticeable at almost non-detectable levels. It should be noted that in order for the unit to operate properly, some sort of ground is usually required. Metallic tape is used as shown and provides contact to the operator's hand, providing a partial ground. If the unit is placed remotely, it should be earth grounded to a water pipe, for example.

Measuring positive ions simply involves holding the antenna, using the body of the unit as the probe, or reversing the polarity of the transistors. A *double pole, double throw* (DPDT) switch may be optionally installed for polarity reversal, allowing reversal at the flick of a switch. Be aware that a ceramic or quality switch must be used to avoid leakage in high humidity conditions.

Construction Steps

1. Lay out and identify all the parts and pieces. Verify them with the parts list, and separate the resistors as they have a color code to determine their value. Colors are noted in the parts list at the end of the chapter.
2. Cut a piece of .1-inch grid perforated board to 2 × 1 inches. Locate and drill the holes as shown in Figure 38-3. If you are building from a perforated board, it is suggested that you insert the components starting in the lower left-hand corner as shown. Pay attention to the polarity of all the semiconductors. Route

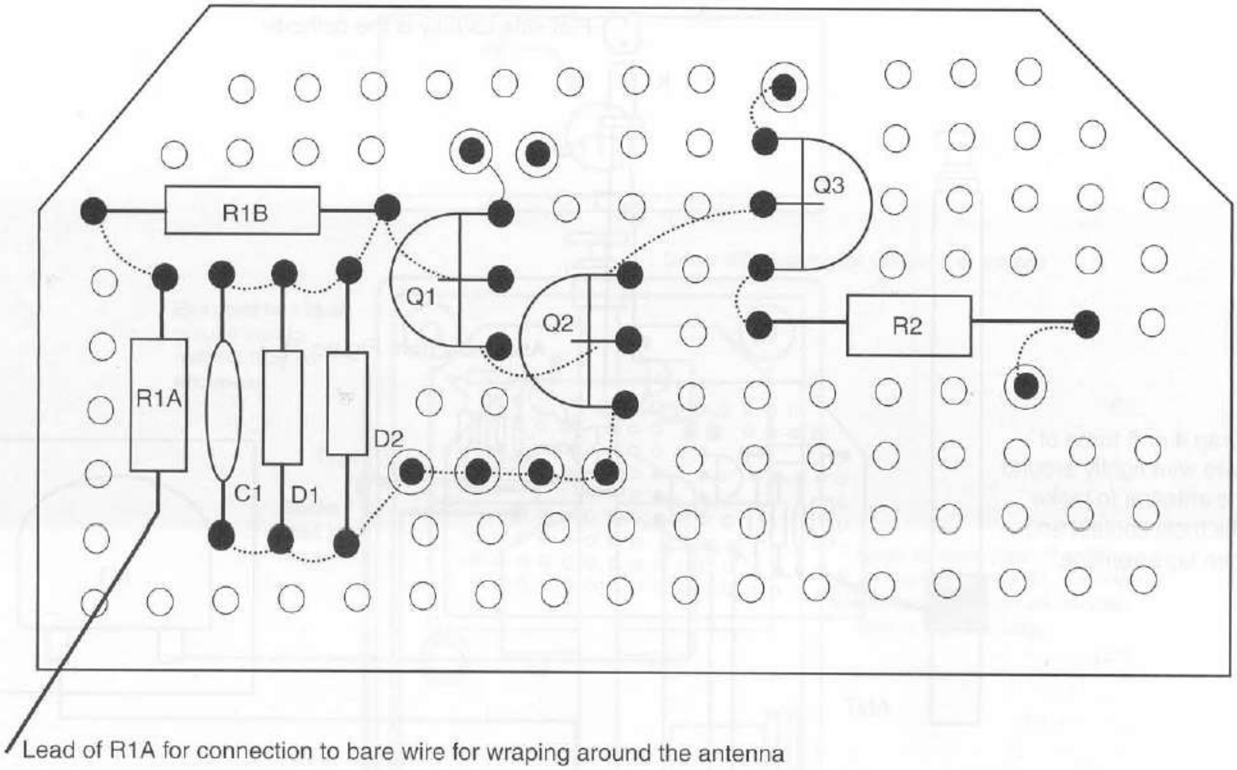

Lead of R1A for connection to bare wire for wraping around the antenna

● This symbol indicates holes used for leads of components.

◉ This symbol indicates holes used for connection points of external leads.

∿ Dashed line indicates wiring performed on underside of board. Use the leads of the components whenever possible for these wire runs.

Board size is 1.9 x 1 inch and is .1 x .1 centers.

Figure 38-3 *Assembly board wiring*

the leads of the components as shown and solder as you go, cutting away unused wires. Attempt to use certain leads as the wire runs or use pieces of the #24 bus wire. Follow the dashed lines on the assembly drawing as these indicate the connection runs on the underside of the assembly board.

3. Attach the external leads and components as shown in Figure 38-4.
4. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.
5. Create the plastic case, as shown in Figure 38-5. Note the 1-inch meter hole centered and located approximately 1½ inch from the top as shown. Holes for the switches are centered and located 1 inch below the meter line. Location is not critical, but make sure everything is clear and fits together.
6. Place a strip of foil tape as shown, noting a hole for the connection to the common circuit

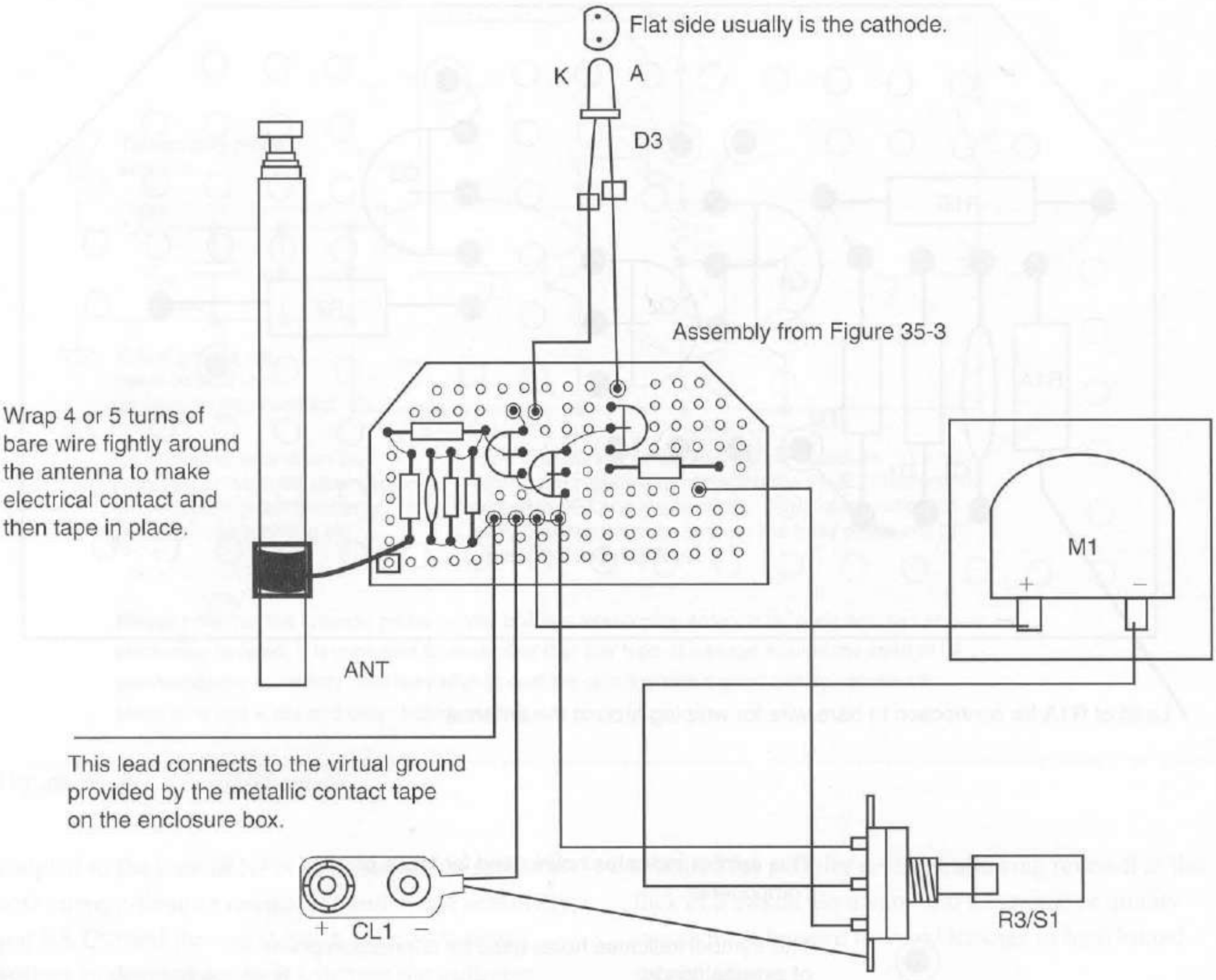

Wrap 4 or 5 turns of bare wire tightly around the antenna to make electrical contact and then tape in place.

This lead connects to the virtual ground provided by the metallic contact tape on the enclosure box.

☐ These symbols indicate solder connections of component leads to external wires.

Use #24 vinyl jacket hook-up wire for external connections.

Figure 38-4 *Final wiring and assembly*

point. Secure the assembly board with double-sided tape after circuit operation has been verified.

7. Install battery B1 as shown and turn on the control. Rotate R3 fully clockwise and note the meter showing a slight indication. This is due to transistor and circuit leakage and should not be considered an indication of the ions. If meter sensitivity responds in the opposite way, you may have to reverse the wires across the outer terminals of R3 or the meter leads.

8. Extend the antenna about 6 inches and adjust R3 fully clockwise. Obtain a plastic comb and run it through your hair or rub a plastic object with a dry cloth. Place it near the probe and note the meter indicating a strong charge and the diode lighting. This effect may be reduced if the air is damp or moist.

9. Make the final assembly, as shown in Figure 38-6.

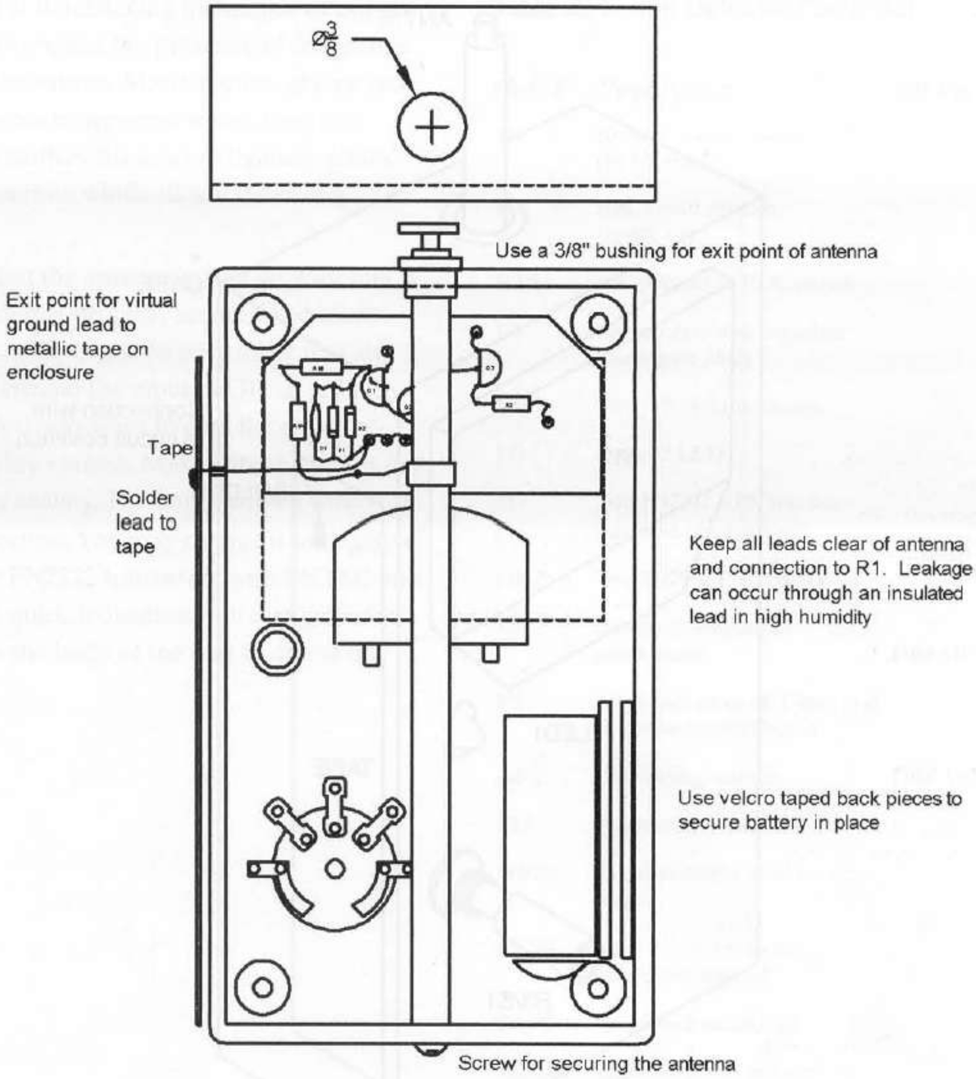

Use a 3/8" bushing for exit point of antenna

Exit point for virtual ground lead to metallic tape on enclosure

Tape

Solder lead to tape

Keep all leads clear of antenna and connection to R1. Leakage can occur through an insulated lead in high humidity

Use velcro taped back pieces to secure battery in place

Screw for securing the antenna

Rear view with cover removed. Wiring as not shown for clarity. Use figure 35-4 for wiring and this figure for component location. Keep all lead neat and short.

Figure 38-5 *Final assembly*

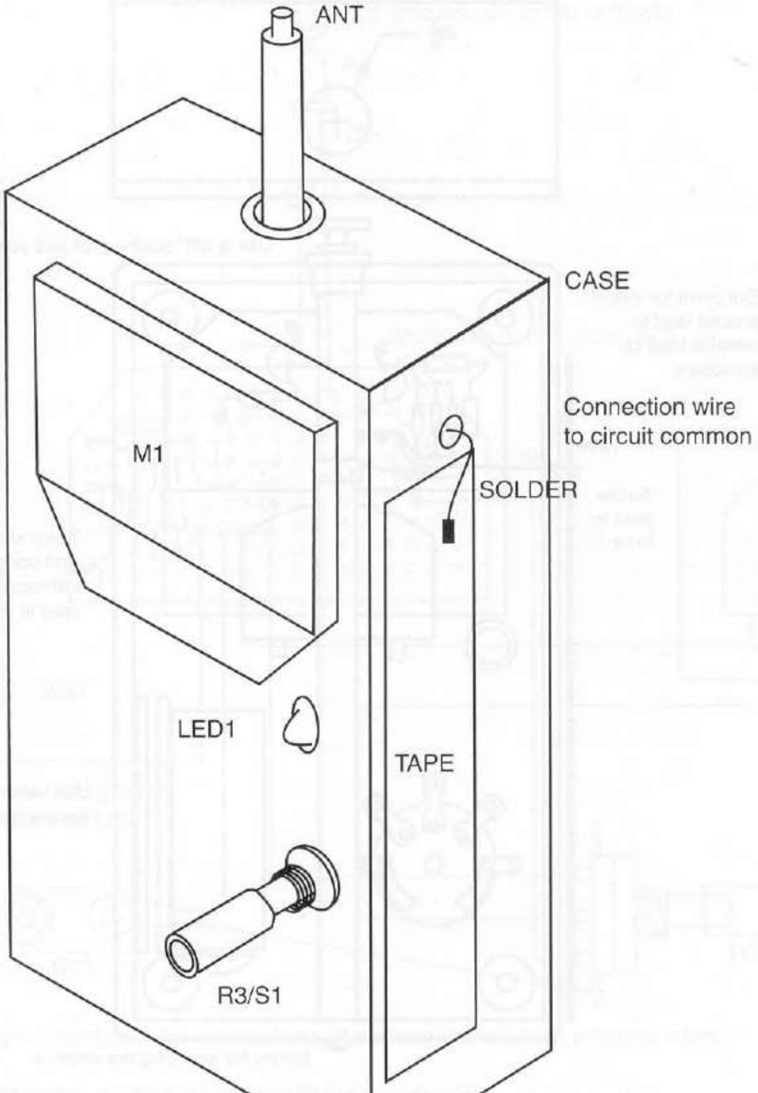

Figure 38-6 *Completed unit*

Applications

Your ion detector is a very, very sensitive ion-measuring device. It can be used for relative measurement but is not designed for absolute measurement.

For quick indications of the presence of a negative ion field, the unit may be handheld and actually used to determine where the source is. The sensitivity of this device can be realized by the simple experiment of running a plastic comb through one's hair and laying it near the probe, as described above.

Those who are familiar with the metallic leaf electroscope will soon realize the advantage of portability and sensitivity. When used for indicating or testing the relative strength of ion sources, the unit should be hard wire grounded for the best results. Adjustments to an ion source to determine the ion output may be made, noting the meter reading and then readjusting R3 to bring the meter reading to scale.

A very interesting phenomenon will be noted when using this device for detecting residual ion fields, the shielding of ions, field direction, static charges, resultant polarity, and the intensity of static charges, as well as a host of others. The unit is an invaluable tool for determining the output of ion generators, any purifiers, and the presence of dangerous static electricity situations. Many sources of charged particles soon become apparent when using the device. People's clothes, fluorescent lighting, plastic containers, and certain winds all will indicate a charge.

Please note that the antenna probe on your unit is a telescoping antenna properly secured and electrically isolated. It is important to remember that any type of leakage around the input of Q1 can reduce the sensitivity. You may want to coat the circuitry with a good-quality varnish. Make sure the unit is dry and clean before sealing. The unit is shown wired for positive ion detection. You may change it to negative by replacing the PN2222 transistors with PN2907 and vice versa. For a quick indication, you may grab the antenna and use the body of the unit to detect the negative ions.

Table 38-1 Ion Detector Parts List

Ref. #	*Description*	*DB Part #*
R1	10 meg, ¼-watt resistor (br-blk-blue)	
R2	10K, 1-watt resistor (br-blk-or)	
R3/S1	10K pot and 12 VDC switch	
C1	100-picofard disk capacitor (see Figure 38-2)	
D1, 2	Two 1N914 signal diodes	
D3	Colored LED	
Q3	Two PN2222 NPN transistor (see note on reversing)	
Q1, 2	Two PN2907 PNP transistors	
M1	100-micro-amp, small panel meter	DB# METER50S
PB1	1 × 2-inch piece of .1-inch grid perforated circuit board	
ANT	Telescoping antenna	DB# ANT1
CL1	9-volt battery clip	
WR24	24 inches of #24 vinyl hookup wire	
CASE	4⅜ × 2 7/16 × 1¼-inch plastic box with lid	
TAPE	4 × ½-inch metal tape	
WR24B	12 inches of #24 bus wire	

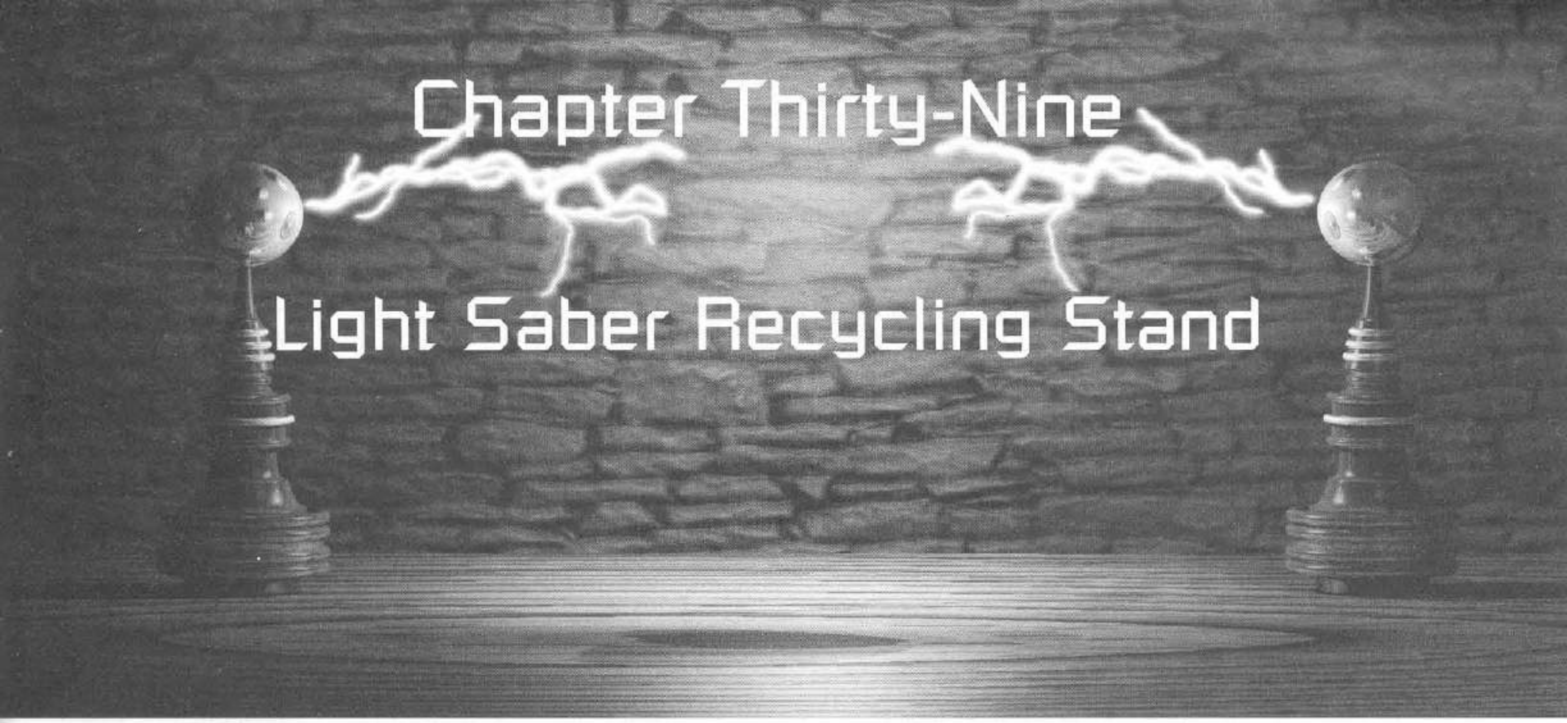

Chapter Thirty-Nine

Light Saber Recycling Stand

This project is intended for those who have built or purchased a light saber from Information Unlimited or who have built one from our first book as directed in Chapter 21 in this series of three.

The saber stand recycles your saber where the plasma display travels up the tube, appearing to evaporate into open space (see Figure 39-1). It then retracts into the hilt and continues repeating until power is removed. The effect is very relaxing and provides an excellent display piece or prop for a bar, a point-of-purchase display, an attention getter, a conversation piece, or just a nightlight for a bedroom.

Operation is powered from a 12-volt wall adapter and is completely safe for children. This is an intermediate-level project requiring basic electronic skills. Expect to spend \$20 to \$30. All the parts are readily available, with specialized parts obtainable through Information Unlimited (www.amazing1.com), and are listed in the parts list at the end of the chapter.

Circuit Description

Figure 39-2 shows a 12-volt wall adapter (T1) being fed into a voltage regulator chip (I2) where a regulated 9 volts is applied to the remaining circuitry. Capacitors C1 and C2 smooth out the voltage waveform. An astable pulse generator using a timer (I1) provides an uneven "on to off" time determined by the ratio of resistors R2 and R3. The output pulse repetition rate is a function of timing capacitor C3 with

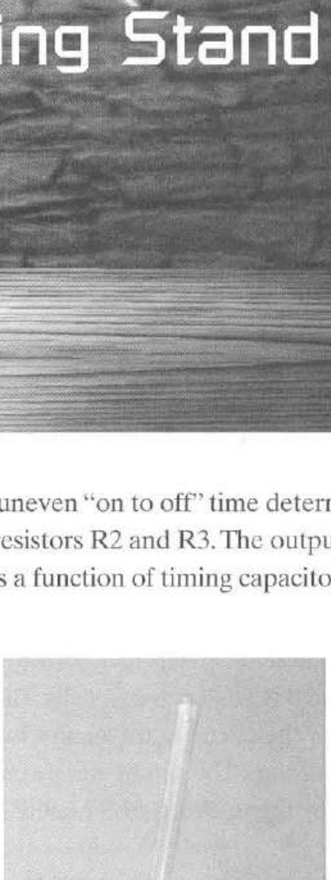

Figure 39-1 *The saber recycling stand*

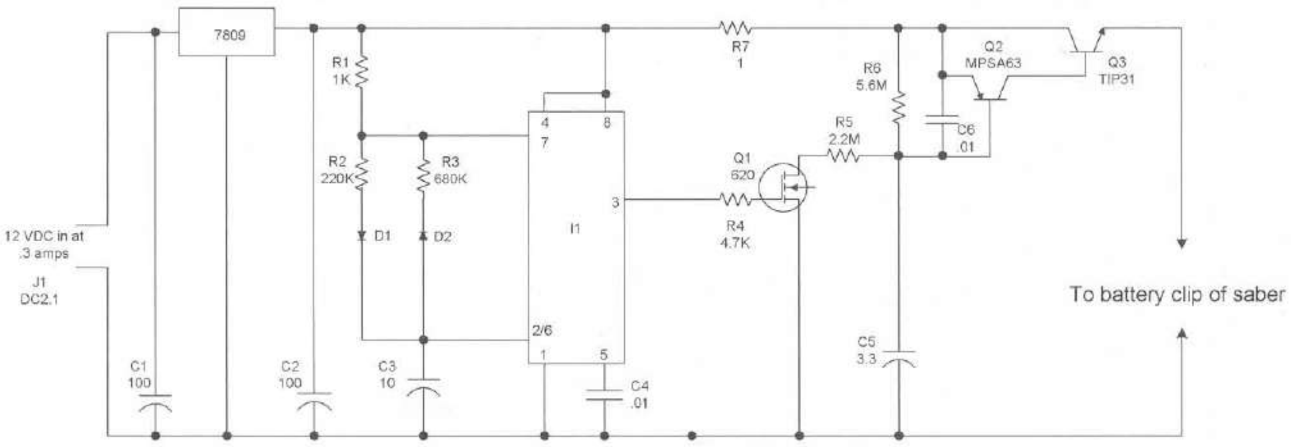

Figure 39-2 *Circuit schematic*

output at pin 3 of I1. The pulse output serves as a gate for MOSFET switch Q1, creating a charging and discharging of low-leakage capacitor C5. A quasi-linear voltage ramp now appears across C5, controlling the current through a high beta Darlington transistor (Q2) in turn controlling the main-current pass transistor Q3. It is this ramping up and down of current that now powers the saber electronics, allowing the plasma to flow up and down, all automatically. The cycling action is set by the circuit to travel up a half-second and down 3 seconds. This timing can be changed to some extent by the selection of R2 and R3.

Assembly

1. Lay out and identify all the parts and pieces. Verify them with the parts list, and separate the resistors as they have a color code to determine their value. Colors are noted on the parts list.
2. Cut a piece of .1-inch grid perforated board to a size of 3.1 × 1.8 inches, as shown in Figure 39-3. Locate and drill the holes as shown to relieve any strain on the input and output leads.
3. When you are building from a perforated board, it is suggested that you insert the components starting in the lower left-hand corner, which is used as a reference, and proceed to the opposite corner.

   Pay attention to the polarity of the capacitors with polarity signs and all the semiconductors.

   Route the leads of the components as shown and solder as you go, cutting away unused wires. Attempt to use certain leads as the wire runs or use pieces of the #24 bus wire. Follow the dashed lines on the assembly drawing as these indicate the connection runs on the underside of the assembly board.
4. Attach the external leads for battery clip CL1 and the leads to input power jack J1. *Important!* The battery clip CL1 must be wired in reverse by connecting the red lead to the circuit ground and the black lead to the emitter of Q3.
5. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.

Parts Fabrication

1. Create the chassis shown in Figure 39-4. This is attached to the BASE section using four #6 sheet metal screws.
2. Create the base section from a piece of plastic or aluminum, as shown in Figure 39-5. Note

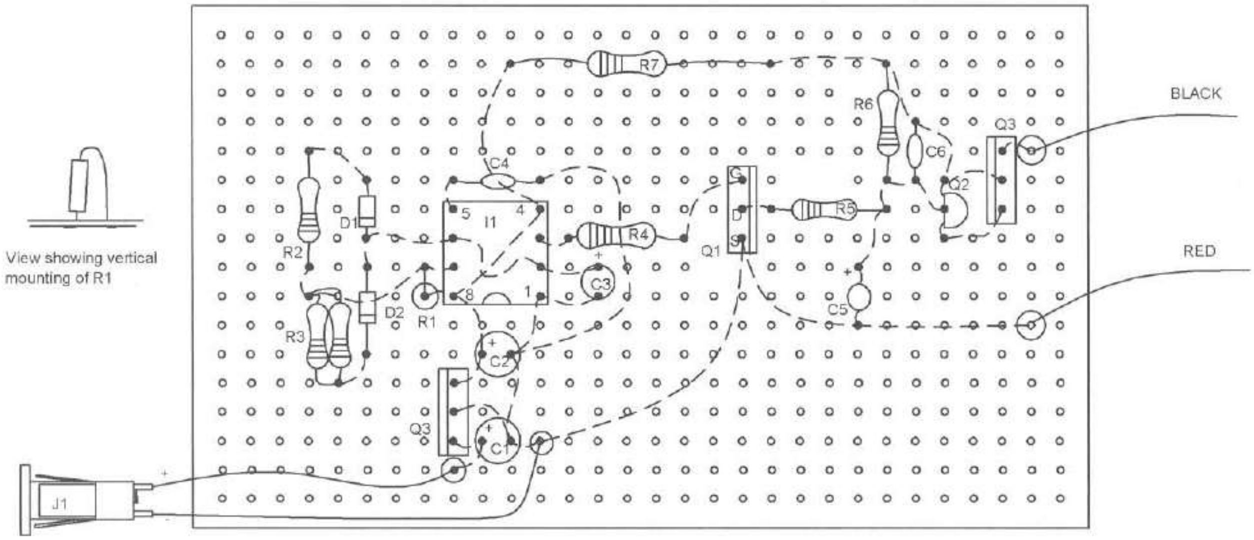

Notes

The looped dashed lind connecting pins 2 and 6 of I1 must have a sleeve of insulated tubing to prevent shorting the lead between pins 4 and 8.

Dashed lines are connections on underside of perforated circuit board using the component leads. These points can also be used to determine foil runs for those who wish to fabricate a printed circuit board.

· Small dots are holes used for component insertion.

● Large dots are solder junctions

⊙ Circles with dots are holes that may require drilling for component mounting and points for external connecting wires indicating strain relief points. It is suggested to drill clearance holes for the leads and solder to points beneath board.

Figure 39-3 *Board assembly*

the four holes for mounting the chassis section, two holes for the holder section, and two holes for the tie wraps.

3. Create the holder from a piece of 2$^{3}/_{8}$-inch, schedule 40 gray PVC tubing. Note the slight angle cut for aesthetics being congruent with the light saber when in the stand. This piece is mounted to the base using two #6 × $^{3}/_{8}$-inch sheet metal screws.

4. Form the frame as shown in Figure 39-6 from $^{3}/_{16}$-inch cold roll steel rod or an equivalent. The frame must be formed to provide a stable assembly with the saber in place in the stand. The longer 36-inch blade unit is the most critical, requiring the most attention. Attempt to follow the dimensions as shown and determine the base angle as a final adjustment with the saber in place.

   The frame is secured in place by inserting the horizontal members through the corner holes in the chassis section. The front section of the frame is secured in place by two small tie wraps, as shown in Figure 39-5.

10. Finish the assembly, as shown in Figures 39-7 and 39-8.

Operation

Remove the cap at the rear of the saber hilt and remove both batteries. Take out one of the battery clips and snap it into the mating clip of the saber stand. Position the saber in the stand and verify its mechanical stability. Plug in the wall adapter and turn on the switch. The plasma should immediately go to the end and then drop down into the hilt. Recycling should commence and continue until power is removed.

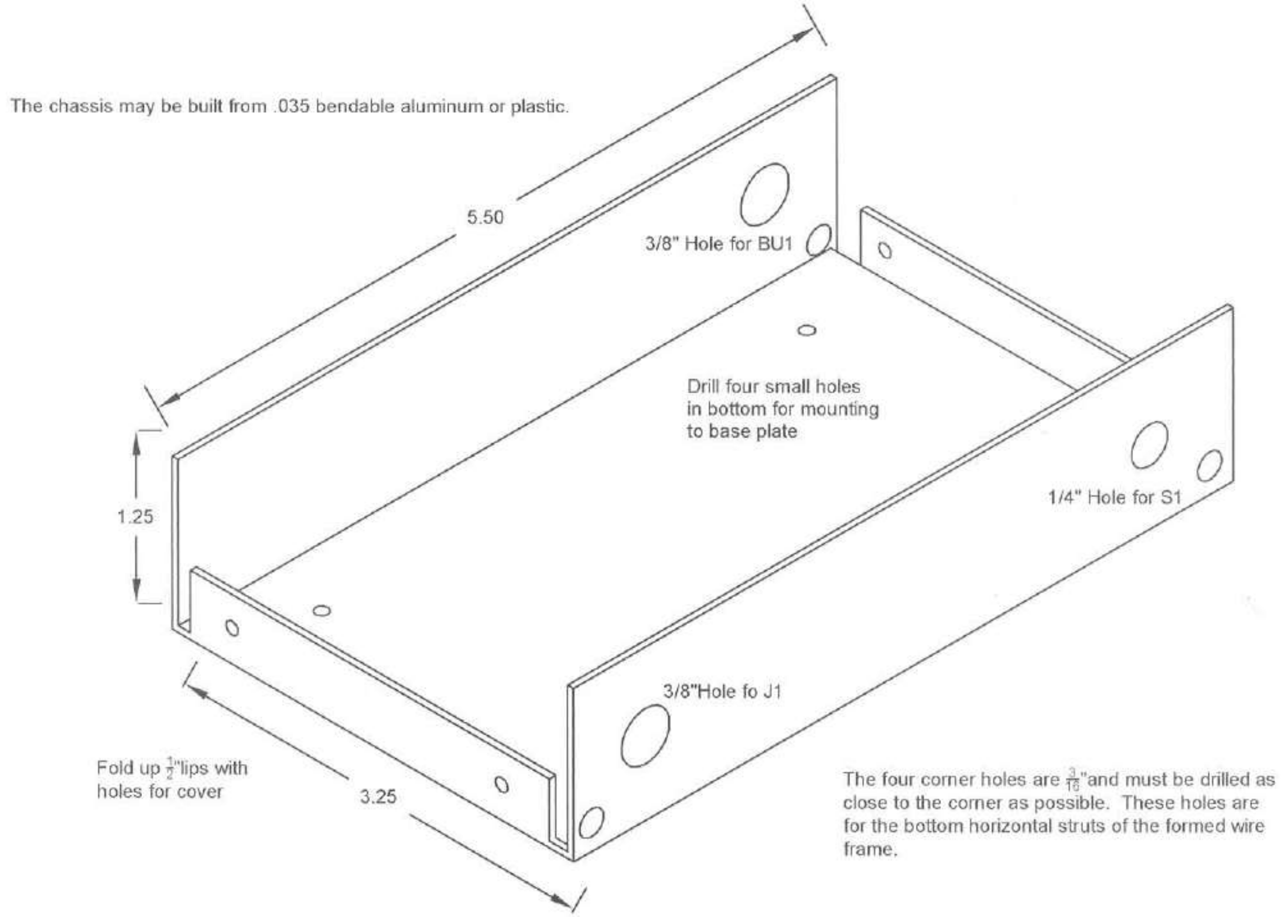

The chassis may be built from .035 bendable aluminum or plastic.

Drill four small holes in bottom for mounting to base plate

Fold up $\frac{1}{2}$"lips with holes for cover

The four corner holes are $\frac{3}{16}$"and must be drilled as close to the corner as possible. These holes are for the bottom horizontal struts of the formed wire frame.

Figure 39-4 *Chassis fabrication*

Light sabers are available from Information Unlimited at www.amazing1.com in 24- or 36-inch lengths in colors of phasor green, photon blue, starfire purple, neon red, or solar yellow.

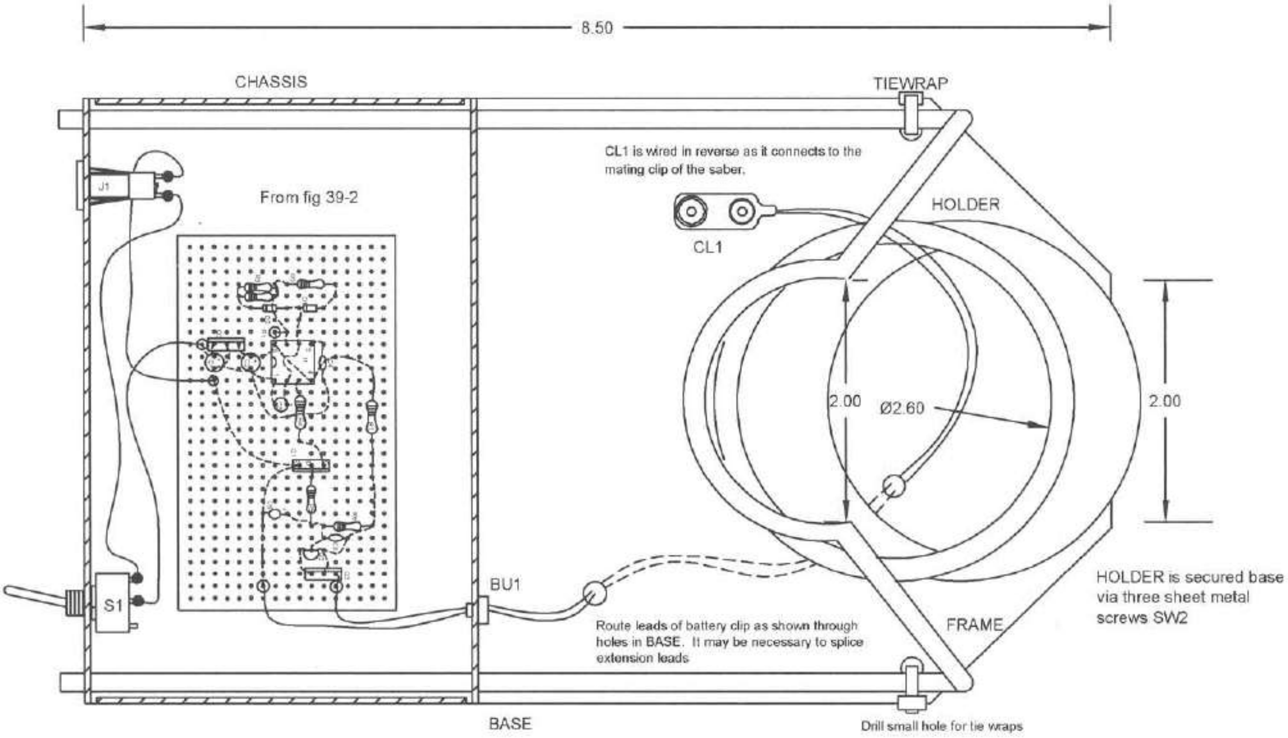
8.50
CHASSIS
TIEWRAP
J1
From fig 39-2
HOLDER
CL1
2.00
Ø2.60
2.00
S1
BU1
FRAME
BASE

CL1 is wired in reverse as it connects to the mating clip of the saber.

HOLDER is secured base via three sheet metal screws SW2

Route leads of battery clip as shown through holes in BASE. It may be necessary to splice extension leads

Drill small hole for tie wraps

Figure 39-5 *Base fabrication*

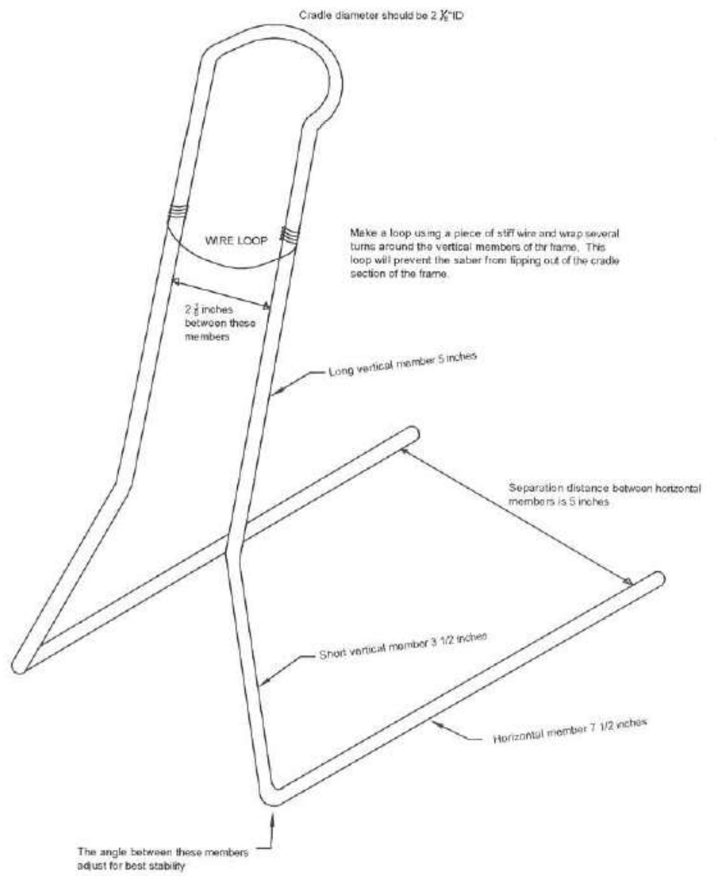
Fabricate from $\frac{3}{16}$"cold roll steel or brass
Cradle diameter should be 2 $\frac{1}{8}$"ID
WIRE LOOP
Make a loop using a piece of stiff wire and wrap several turns around the vertical members of thr frame. This loop will prevent the saber from flipping out of the cradle section of the frame.
2 $\frac{1}{8}$ inches between these members
Long vertical member 5 inches
Separation distance between horizontal members is 5 inches
Short vertical member 3 1/2 inches
Horizontal member 7 1/2 inches
The angle between these members adjust for best stability

Figure 39-6 *Frame section*

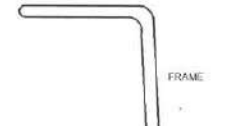
FRAME
LOOP
HOLDER
CHASSIS
BASE
FEET

Figure 39-7 *Side view of assembly*

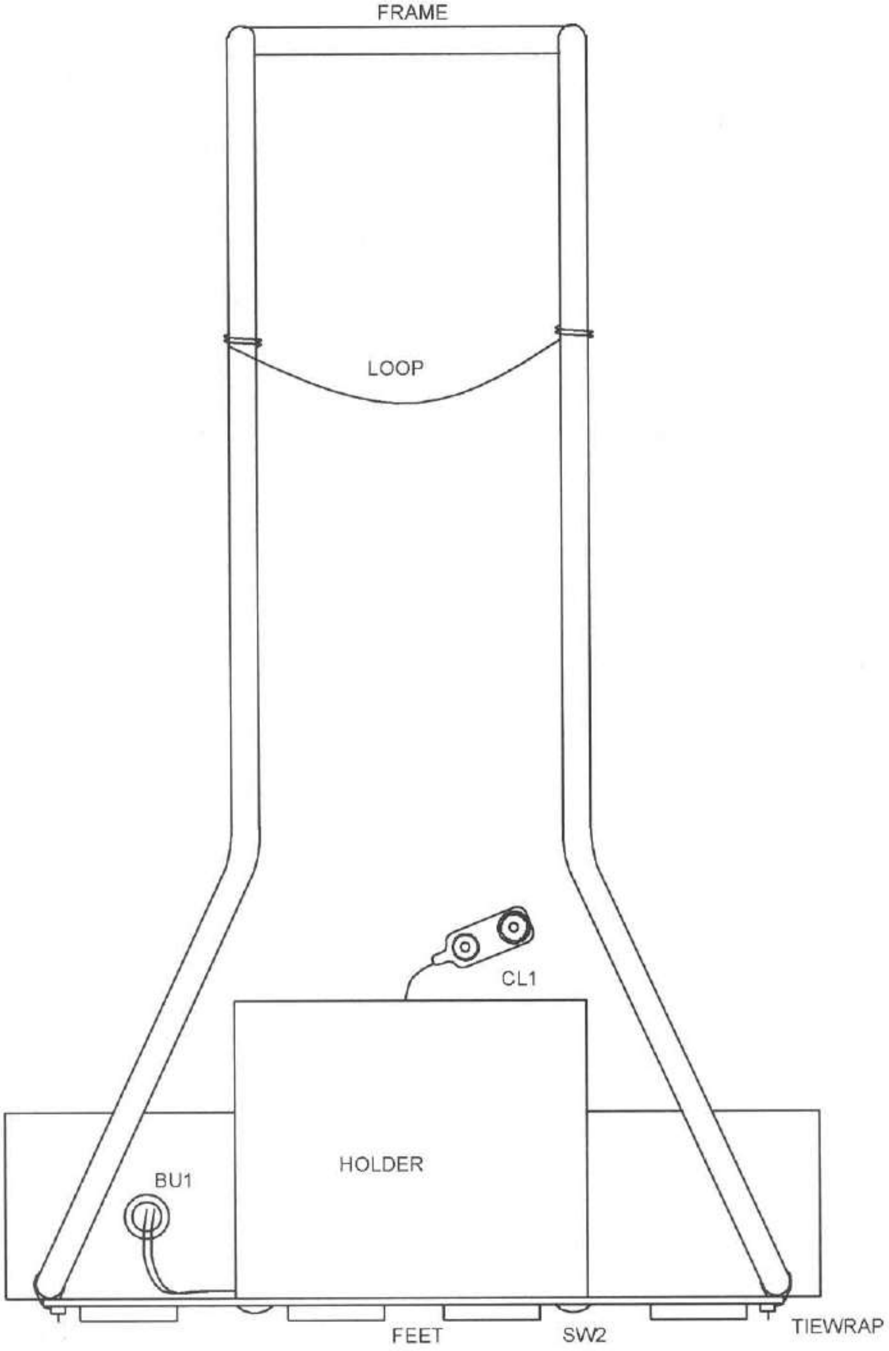
FRAME
LOOP
CL1
HOLDER
BU1
FEET
SW2
TIEWRAP

Figure 39-8 *Front view of assembly*

Table 39-1 Light Saber Stand Parts List

Ref. #	*Description*	*DB Part #*
R1	1K, ¼-watt resistor (br-blk-red)	
R2	220K, ¼-watt resistor (red-red-yel)	
R3	680K, ¼-watt resistor (blue-gray-yel)	
R4	4.7K, ¼-watt resistor (yel-pur-red)	
R5	2.2 m, ¼-watt resistor (red-red-gr)	
R6	5.6 m, ¼-watt resistor (gr-blue-gr)	
R7	1-ohm, ¼-watt resistor (br-blk-gold)	
C1, 2	Two 100-microfarad, 25-volt vertical electrolytic capacitors	
C3	10-microfarad, 25-volt vertical electrolytic capacitor	
C4, 6	Two .01-microfarad, 50-volt disc capacitors	
C5	3.3-microfarad, 35-volt tantalum capacitor	
D1, 2	Two IN914 silicon diodes	
I1	555 dual inline package (DIP) timer integrated circuit	
I2	7809 T0220 9-volt regulator	
Q1	IRF620 metal-oxide-semiconductor field effect transistor (MOSFET)	
Q2	MPSA63 PNP Darlington transistor	
Q3	TIP31 T0220 NPN transistor	
J1	Snap-in DC panel, 2.5 mm jack	
S1	Single pole, single throw (SPST) small toggle switch	
PB1	3.5 × 2 .1 × .1-inch grid perforated board; trim to size	
CL1	Battery clip with 12-inch leads or extend shorter leads	
TIEWRAP	Two small tie wraps	
CHASSIS	Create as shown in Figure 39-4 from .035-inch aluminum	
BASE	Create as shown in Figure 39-5 from .035-inch plastic	
HOLDER	Create from a length of 2⅜-inch schedule 40 gray PVC	
FRAME	Create as shown in Figure 39-6	
FEET	Six large, rubber stick-on feet	
SW2	Six #6 × ⅜-inch sheet metal screws	
BU1	Two ⅜-inch plastic bushings	
LIGHT SABER	36- or 24-inch saber in blue, green, red, or purple	

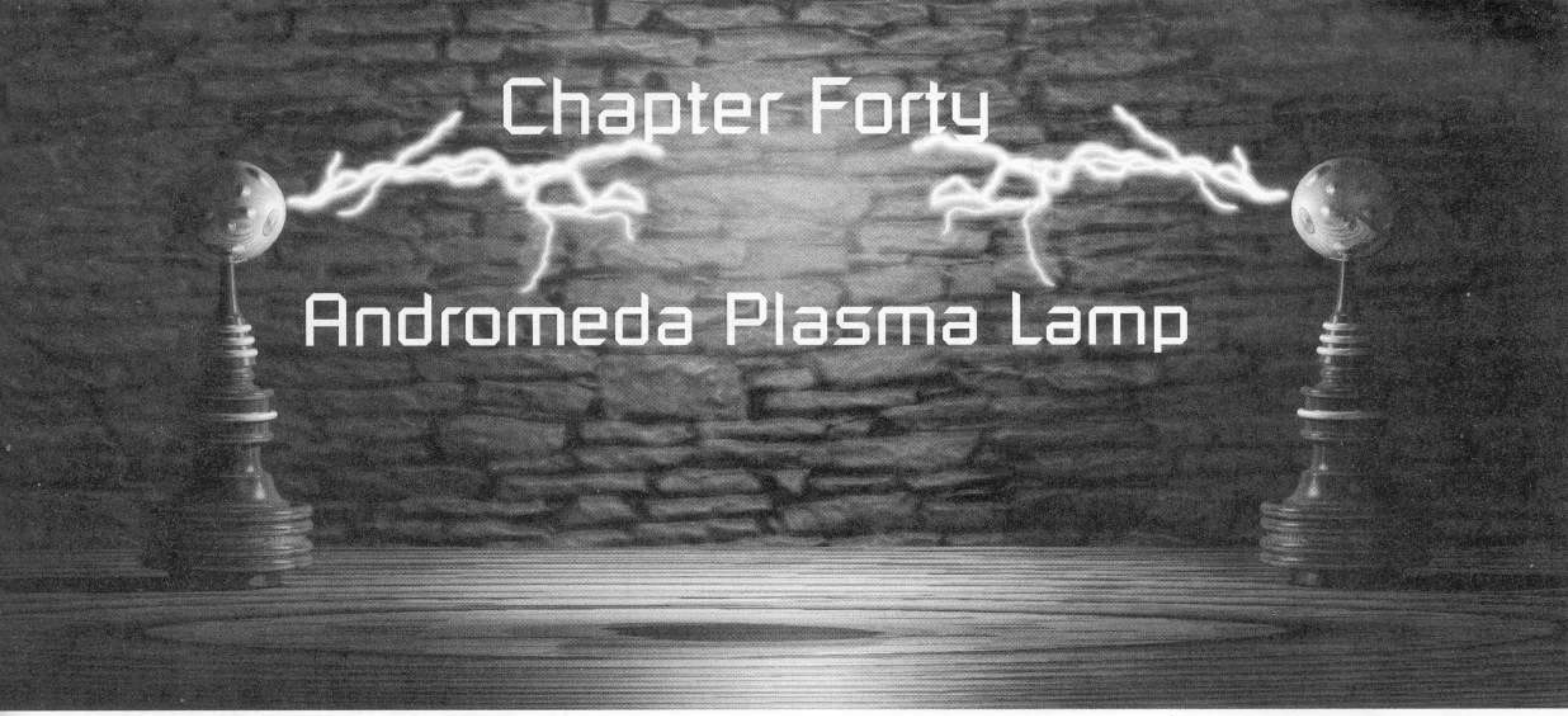

Chapter Forty

Andromeda Plasma Lamp

This project shows how to build a unique type of outside lamp suitable for illuminating driveways, pools, gardens, walkways, docks, or any place where colorful or functioning illumination is desired. The lamps are patented by the author and are currently in preproduction.

The lamp's features are high efficiency as five individual lamps can be powered by a single, conventional, 12-volt wall adapter as the lamps require only 3 watts each. The lamps provide a high amount of useful illumination and must not be compared to the low-light, solar-powered accent lamps.

Another feature of the lamp is that it never needs replacing and should last a lifetime. Other lamps such as incandescent ones are always burning out filaments, whereas conventional fluorescents darken and fail to start due to metallic sputtering.

The third feature is that the lamps are available in pure spectral colors of green, blue, purple, yellow, red, and cool and warm white. These are vivid colors and not the result of filters or paint. Their color is the result of the true spectral emission.

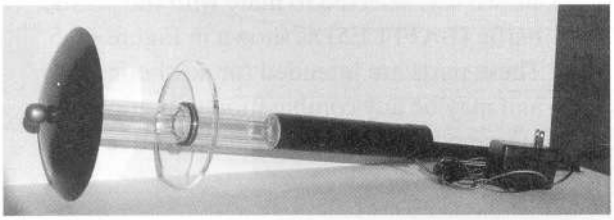

Figure 40-1 *The Andromeda plasma lamp*

This is an intermediate-level project requiring basic electronic skills. Expect to spend \$25 to \$35. All the parts are readily available, with specialized parts obtainable through Information Unlimited at www.amazing1.com, which are listed in the parts list at the end of the chapter.

Circuit Description

Transistor Q1 forms a self-oscillating circuit where current rises in the collector, COL, winding transformer T1 and inducing a voltage in the feedback (FB). This also causes a winding that keeps Q1 on until T1 saturates and the feedback voltage can no longer be induced. The collector current rapidly falls and forces the base to go negative further, turning Q1 off. When the collector current falls to zero, Q1 is turned on by bias-set resistors R1 and R2, which must charge up timing capacitor C2. The setting of trimpot R1 determines the power output by controlling the number of cycles per second, and capacitor C1 bypasses any residual signal voltage to ground (see Figure 40-2).

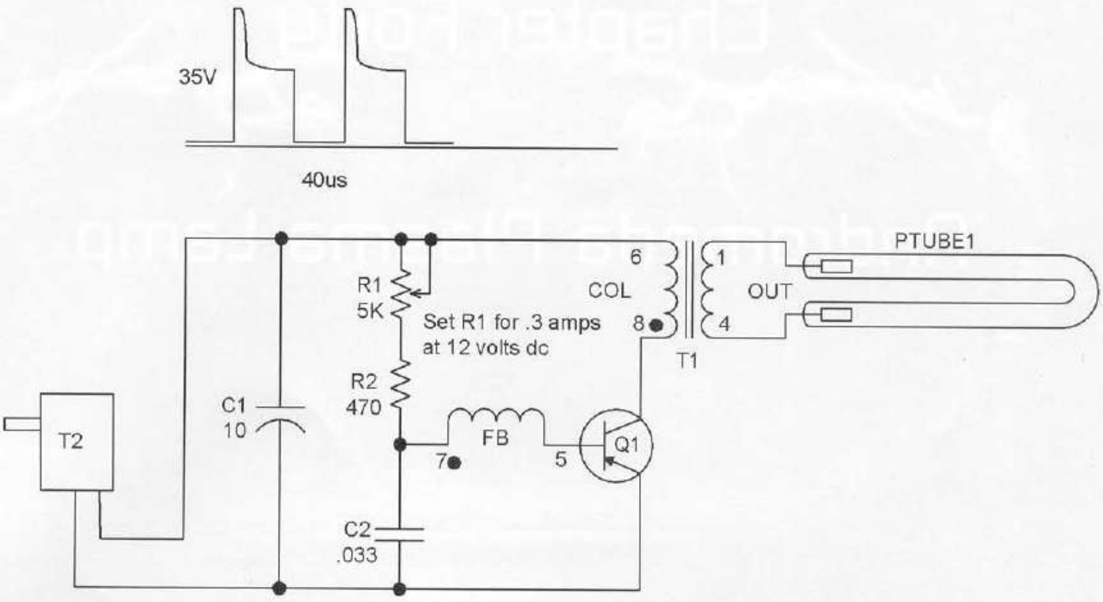

Figure 40-2 *Circuit schematic*

Assembly Board Wiring

1. Cut a piece of .1 × .1 grid vector circuit board to 1 × 4.5 inches, as shown in Figure 40-3.
2. Whenever you are using a perforated board, it is suggested that you insert components starting in the lower left-hand corner. Pay attention to the polarity of the capacitors with polarity signs and all the semiconductors. Route the leads of the components as shown and solder as you go, cutting away unused wires. Attempt to use certain leads as the wire runs or use pieces of the #24 bus wire. Follow the dashed lines on the assembly drawing as these indicate the connection runs on the underside of the assembly board.
3. Double-check the accuracy of the wiring and the quality of the solder joints. Avoid wire bridges, shorts, and close proximity to other circuit components. If a wire bridge is necessary, sleeve some insulation onto the lead to avoid any potential shorts.
4. Attach leads from the plasma tube PT1, being careful not to stress the glass enclosure. Note the small tie wraps (TY1) securing the tube to the perforated board. Again, do not cause any stress to the glass. Make sure that when tightening the wraps they do not stress the tube. When soldering, you must not allow the glass or metal seal to get too hot. Leave at least a half-inch lead between the tube and the connection. Use flux and make the solder joints quickly.
5. Connect the input power leads to a 12-volt DC source and note the lamp lighting. Quickly adjust trim pot R1 to 300 milliamps of current. The lamp should be very bright and the heatsink tab on Q1 should only be warm. Allow it to run for several hours and check adapter transformer T2 for excessive heating.
6. Create the EN1 enclosure tube and the STK1 mounting stake, as shown in Figure 40-4. Note the two mating holes for attaching the stake and the $^{5}/_{16}$-inch hole for the rubber tubular bushing for routing the input power leads.
7. Obtain a top canopy (CANOPY7). The one shown was selected to mate with the plastic baffle (BAFFLE5) as shown in Figure 40-5. These parts are intended for aesthetics only and may be any combination that appears pleasing to the builder. A good source is a

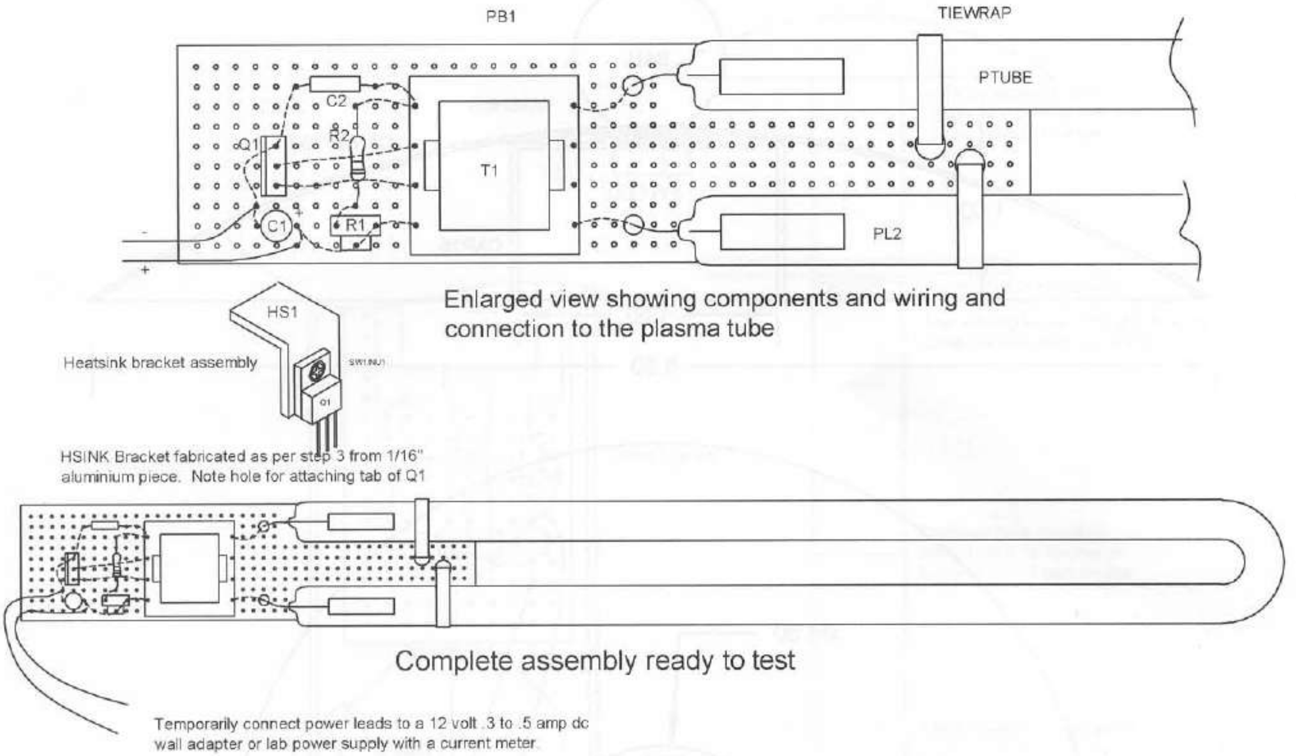
PB1
TIEWRAP
PTUBE
C2
R2
Q1
T1
C1
R1
PL2
HS1
Q1
Heatsink bracket assembly

Enlarged view showing components and wiring and connection to the plasma tube

HSINK Bracket fabricated as per step 3 from 1/16" aluminium piece. Note hole for attaching tab of Q1

Complete assembly ready to test

Temporarily connect power leads to a 12 volt .3 to .5 amp dc wall adapter or lab power supply with a current meter.

Figure 40-3 *Assembly board*

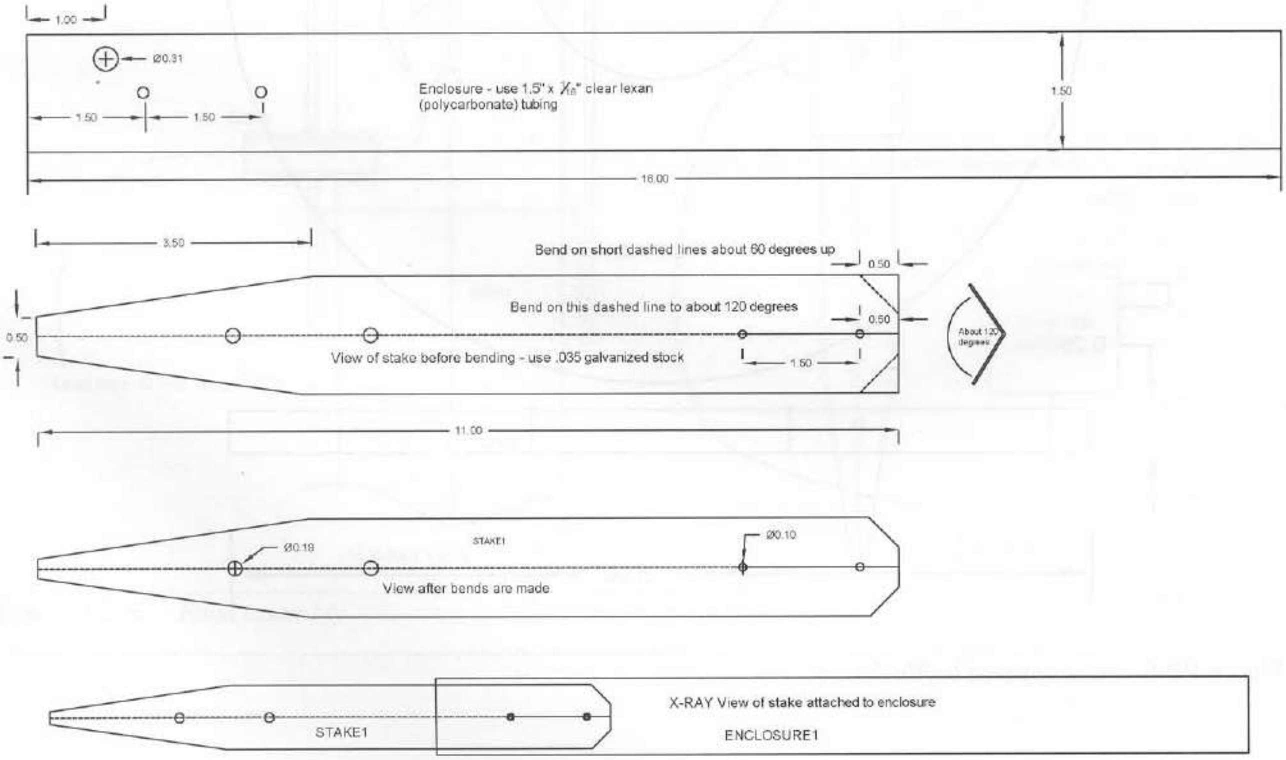
1.00
Ø0.31
1.50
1.50
Enclosure - use 1.5" x 1/16" clear lexan (polycarbonate) tubing
1.50
16.00
3.50
Bend on short dashed lines about 60 degrees up
0.50
Bend on this dashed line to about 120 degrees
0.50
0.50
View of stake before bending - use .035 galvanized stock
1.50
About 120 degrees
11.00
Ø0.18
STAKE1
Ø0.10
View after bends are made
X-RAY View of stake attached to enclosure
STAKE1
ENCLOSURE1

Figure 40-4 *Creating the mounting stake and enclosure*

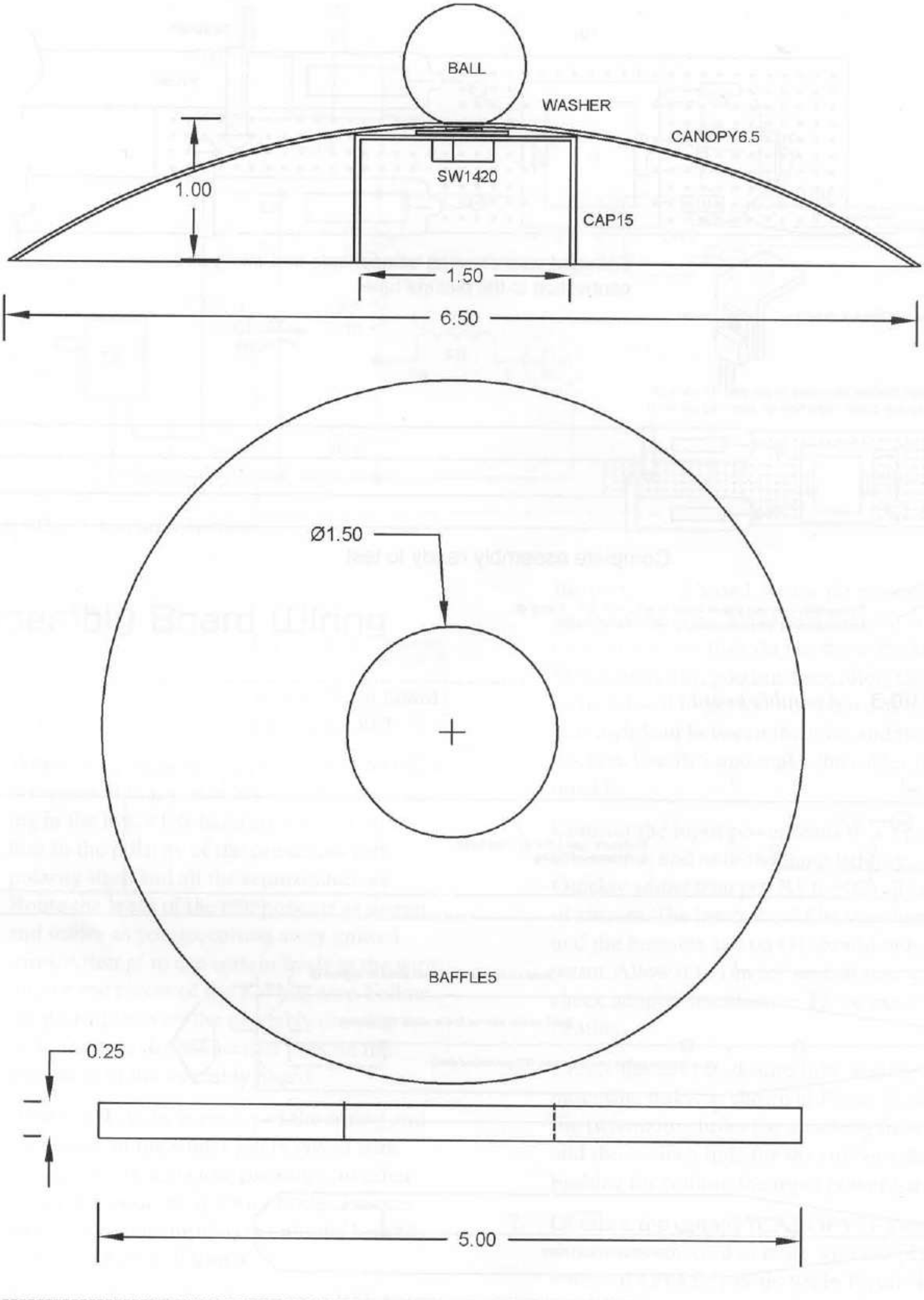
BALL
WASHER
CANOPY6.5
SW1420
1.00
CAP15
1.50
6.50
Ø1.50
BAFFLE5
0.25
5.00

Figure 40-5 *Canopy and baffle design*

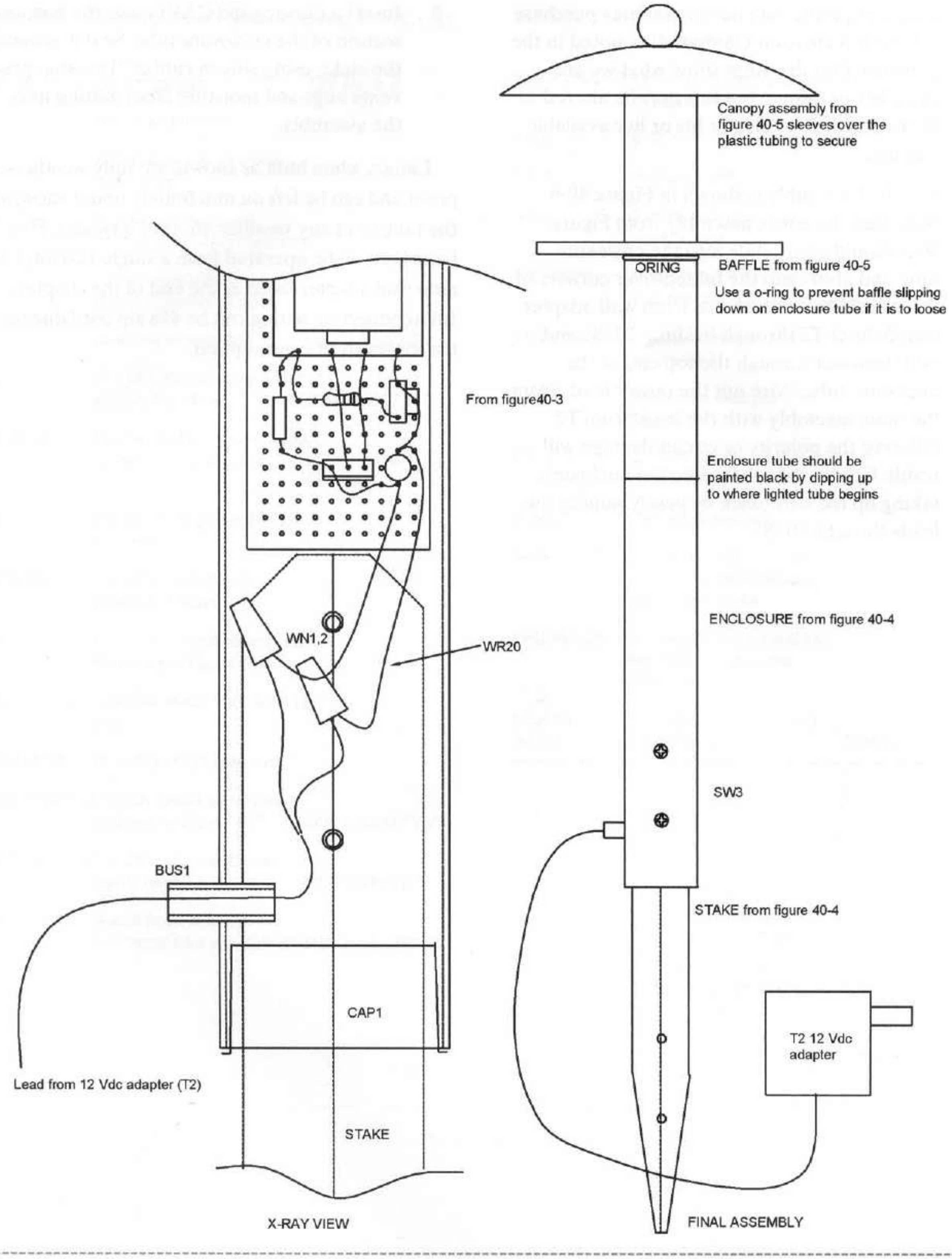
Canopy assembly from figure 40-5 sleeves over the plastic tubing to secure
ORING
BAFFLE from figure 40-5
Use a o-ring to prevent baffle slipping down on enclosure tube if it is to loose
From figure40-3
Enclosure tube should be painted black by dipping up to where lighted tube begins
ENCLOSURE from figure 40-4
WN1,2
WR20
SW3
BUS1
STAKE from figure 40-4
CAP1
T2 12 Vdc adapter
Lead from 12 Vdc adapter (T2)
STAKE
X-RAY VIEW
FINAL ASSEMBLY

Figure 40-6 *Final assembly*

lamp or lighting supplier or you may purchase it from Information Unlimited, as noted in the parts list. Our drawings show what we are using in our design, but this may be altered to fit the builder's choice or his or her available sources.

8. The final assembly is shown in Figure 40-6. Note that the entire assembly from Figure 40-3 should easily slide into the enclosure tube and abut onto the folded-over corners of the stake. Thread the leads from wall adapter transformer T2 through bushing BUS1 and pull them out through the top end of the enclosure tube. Wire nut the power leads from the main assembly with the leads from T2. Observe the polarity or circuit damage will result. Gently slide them into the enclosure, taking up the wire slack by gently pulling the leads through BUS1.

9. Insert a plastic cap (CAP1) into the bottom section of the enclosure tube. Seal it around the stake using silicon rubber. This step prevents bugs and moisture from getting into the assembly.

Lamps, when built as shown, are fully weatherproof and can be left on indefinitely under snow, in the rain, or in any weather, short of a twister. Five lamps can all be operated from a single 12-volt, 1.5-amp wall adapter listed at the end of the chapter. Interconnecting wiring can be #18 zip cord due to the low-current consumption.

Table 40-1 Andromedia Plasma Lamp Parts List

Ref. #	*Description*	*DB Part #*
R1	5K vertical trimmer	
R2	470-ohm, ¼-watt resistor (yel-pur-br)	
C1	10-microfarad, 50-volt vertical electrolytic capacitor	
C2	.033-microfarad, 50-volt plastic capacitor (332)	
T1	Inverting transformer	DB# ANQP2824
T2	12 VDC, 300-milliamp wall adapter transformer	DB# 12 DC/.3
PTUBE-C	Cool white/blue/green/purple/ red/yellow plasma tube	DB# PTUBE-COLOR
Q1	PN3055 NPN power T0-220 transistor	
PBOARD	1 × 4.5 × .1-inch grid perforated board	
HS1	1.5 × 1 × ¹⁄₁₆-inch aluminum heatsink per Figure 40-3	
WR20	12 inches of #20 vinyl hookup wire	
WR24BUSS	12 inches of #24 buss wire	
CANOPY6.5	6.5-inch, round metal canopy, as shown in Figure 40-5	DB# CANOPY65
BAFFLE	5 × .25-inch, round, clear plastic baffle	DB# BAFFLE5
BALL	⅞-inch brass ball with a hole tapped for a ¼-20 bolt	DB# BALL78
STAKE1	Metal stake created as shown in Figure 40-4	
ENCLOSURE	Plastic tube created as shown in Figure 40-4	
SW1	6-32 ¼ Phillips screw	
NUT	6-32 keep nut	
SW3	Two #6 × ¼-inch sheet metal screws	
SW1420X1	¼-20 × 1-inch hex bolt for attaching the ball (see Figure 40-5)	
TIEWRAP	Two tie wraps used per Figure 40-3	
CAP1	Plastic cap used per Figure 40-6	
CAP1.5	1.5-inch plastic cap shown in Figure 40-6	
WN1, 2	Two small wire nuts	
BUS1	⁵⁄₁₆ OD × ³⁄₁₆ ID × 1-inch length of rubber tubing (see Figure 40-6)	
ORING1.25	1¼-inch O-ring for holding the baffle in place per Figure 40-6	
5 LAMP ADAPTER	12 VDC, 1.5-amp adapter for five lamps	DB# 12/1 .5

Newnes

ELECTRONICS EXPLAINED

Fundamentals for Engineers, Technicians, and Makers

Second Edition

Louis E. Frenzel Jr.

Electronics Explained

This page intentionally left blank

Electronics Explained

Fundamentals for Engineers, Technicians, and Makers

Second Edition

Louis E. Frenzel, Jr.
Austin, TX, United States

ELSEVIER
Newnes
An imprint of Elsevier

Newnes is an imprint of Elsevier
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States

Copyright © 2018 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-12-811641-8

For information on all Newnes publications visit our website at https://www.elsevier.com/books-and-journals

ELSEVIER
Book Aid International
Working together to grow libraries in developing countries
www.elsevier.com • www.bookaid.org

Publisher: Mara E. Conner
Acquisition Editor: Tim Pitts
Editorial Project Manager: Charlotte Kent
Production Project Manager: Mohanapriyan Rajendran
Designer: Mark Rogers

Typeset by TNQ Books and Journals

To Joan Ree with love for many happy years.

This page intentionally left blank

Contents

List of Figures xv
Preface to the Second Edition xxv
Preface xxvii

1. **Learning About Electronics for Your Own Good 1**

Introduction 1
It Is Not Like It Used to Be 2
What Is in It for You? 2
Electronics: The Big Picture 3
Communications 4
Computers 4
Control 4
Instrumentation 5
Components 5
How the Electronics Industry Works? 5
Electricity Versus Electronics 6
The Painful Truth 7
Cell Phone 8
Computers 9
Robots 11

2. **Electronic Concepts: More Interesting Than You Think 15**

Introduction 15
Electricity and Electronics 15
Atoms and Electrons 16
Charge, Voltage, and Current 16
Dynamic Electricity and Current Flow 18
Direction of Current Flow 19
Conductors, Insulators, and Semiconductors 20
Magnetism 20
Magnetic Fields 20
Electromagnetism 21
Electromagnetic Induction 23

Voltage Sources 24
Direct Current Voltages 24
Alternating Current Voltages 26
Practical Direct Current and Alternating Current Voltage Sources 29
Direct Current Voltage Sources 29
Fuel Cell 32
Alternating Current Voltage Sources 33
Analog and Digital Signals 34
Fourier Theory and the Frequency Domain 34

3. The Systems Versus Components View of Electronics 41

Introduction 41
Why Systems and Not Circuits? 41
Electronic Components 45
Switches 45
Resistors 45
Capacitors 47
Inductors 50
Transformers 51
Diodes 52
Transistors 52
Integrated Circuits 55

4. Electronic Circuits: Linear/Analog 63

Introduction 63
Linear Circuits 63
Amplifiers 64
Classifying Amplifiers 65
Amplifier Specifications 71
Filters 75
Oscillators 77
Mixers 78
Phase Detector 79
Phase-Locked Loops 80
Frequency Synthesizers 82
Power Supplies 84

5. Electronic Circuits: Digital 103

Introduction 103
Binary Numbers 103
Binary-to-Decimal Conversions 104
Using Hardware to Represent Binary Quantities 105
Decimal-to-Binary Conversions 105
Maximum Decimal Value for N Bits 106
BCD and ASCII 107
Parallel and Serial Data 112

Digital Logic Elements 114
Inverter 114
AND Gate 115
OR Gate 117
NAND and NOR Gates 117
Exclusive OR Gate 118
Flip-Flops 119
Storage Registers 120
Shift Registers 121
Counters 121
Common Logic Circuits 124
Multiplexer 124
Demultiplexer 124
Decoder 124
Comparators 125
Arithmetic Circuits 125
Digital Memories 126
Random Access Memory 126
Read-Only Memory 128
Programmable Logic Devices 129
Data Conversion 131
Analog-to-Digital Conversion 132
Digital-to-Analog Conversion 132
Resolution and Sampling Interval 133
Direct Digital Synthesis 134

6. **How Microcomputers Work** **139**

Introduction 139
Concepts and Definitions 139
Computer Organization and Operation 142
Operational Details 145
Registers and the Arithmetic Logic Unit 145
Control Unit 147
Instruction-Word Formats 148
Program Execution Example 149
Microcomputer Buses 151
Popular Microcomputers 151
8- and 16-Bit Microcontrollers 152
32- and 64-Bit Processors 155
Digital Signal Processing 157

7. **Radio/Wireless** **159**

Introduction 159
Radio Communication Systems 159
Three Basic Ways of Communications 160
Simplex 161
Half Duplex 161

Full Duplex 161
What Is a Radio Wave Anyway? 162
Electric and Magnetic Fields 162
Signal Speed 164
Signal Strength 164
Electromagnetic Frequency Spectrum 164
Radio Wave Propagation 168
Low and Medium Frequencies 168
High Frequencies 168
VHF, UHF, Microwaves and Millimeter Waves 169
Modulation 170
Amplitude Modulation 171
Sidebands 171
Bandwidth 172
Single Sideband Modulation 173
Frequency Modulation 173
Phase Modulation 174
Digital Modulation 174
Spread Spectrum 175
Orthogonal Frequency Division Multiplex 176
Radio Hardware 176
Transmitters 176
Receivers 177
Transceivers 179
ISM Band Transceiver 179
Antennas 181
Ground-Plane Antenna 181
Dipole 182
Yagi 182
Other Antennas 183
Transmission Lines 183
Microwave Antennas and Waveguides 185
Two-Way Radio 186
Aircraft 187
Citizens Band 187
Amateur Radio 188
Family Radio 188
Marine 188
Land Mobile Radio 189
Satellites 189
Radio Telescopes 191

8. Cell Phones 195

Introduction 195
Can You Hear Me Now? 195
The Cellular Concept 197
What Is Inside a Cell Phone? 199

What Type of Cell Phone Technology Do You Have? 200
Digital Modulation and Wireless Transmission Methods 203
Quadrature Amplitude Modulation 203
Orthogonal Frequency Division Multiplexing 204
Multiple Input Multiple Output 206
Spectral Efficiency 208
Access Methods and Duplexing 209
A Look Inside a 4G Cell Phone 212
The Latest Cell Phone Technology 213
Long-Term Evolution 213
Fifth-Generation Wireless Overview 214
Location Technology 216

9. Networking: Wired and Wireless 217

Introduction 217
Is Everything Networked? 217
Types of Networks 218
Wide Area Networks 218
Metropolitan Area Networks 218
Local Area Networks 218
Personal Area Networks 218
Home Area Networks 219
Storage Area Networks 219
Network Relationships 219
Network Interconnection Methods 219
Wired Networking Technologies 220
Types of Cables 221
Ethernet 223
How Is Ethernet Used? 225
Synchronous Optical Network 226
Optical Transport Network 228
How the Internet Works 228
Wireless Networks 230
Wi-Fi 231
Bluetooth 233
ZigBee 233
ISM-Band Radios 235
Radio Frequency Identification and Near-Field Communications 235
Broadband Technology 236
Cable TV Connectivity 237
Digital Subscriber Lines 237
Wireless Broadband 238
The Internet of Things 238
How Internet of Things Works 239
Wireless Technologies of Internet of Things and Machine 2 Machine 240
The Issues of Internet of Things 240

10. Audio Electronics 243

Introduction 243
The Nature of Sound 243
Microphones 244
Speakers 244
Digital Audio 245
Past Recording Media 246
Digitizing Sound 247
Digital Compression 248
How an MP3 Player Works 250
The Compact Disc 251
Audio–Video Receiver 254
Radio Choices 254
Surround Sound 256
Special Sound Applications 257
Wireless Speakers and Headphones 257
Downloading and Streaming Audio 257
Musical Instruments 257
Noise-Canceling Earphones 258
Sonar 259
Hearing Aids 259

11. Video Technology 261

Introduction 261
Video Fundamentals for the Impatient 261
Scanning Fundamentals 262
Aspect Ratio 262
Persistence of Vision 263
Analog TV 264
Digital TV 266
Color Video Principles 269
Video Compression 271
TV Transmission 272
TV Receiver 274
TV Screen Technology 274
Summary of Screen Technologies 277
3D TV 278
Cable Television 278
Satellite TV 283
Over-the-Top TV 285
Cell Phone TV 285
Closed-Circuit TV 286
Digital Video Discs 286
How a DVD Player Works 287
Blu-ray DVD 288
Virtual and Augmented Reality 288

12. Industrial Control 291

Introduction 291
Open- and Closed-Loop Control 291
Open-Loop Control 292
Closed-Loop Control 292
Controllers 294
Sensors 295
Temperature Sensors 296
Pressure Sensors 297
Switches 298
Potentiometers 298
Photoelectric Sensors 299
Flow Sensors 302
Output Devices 303
Relays 303
Solenoids 309
Motors 310
DC Motors 310
AC Motors 311
Other Actuators 312
Thyristors 312
Silicon-Controlled Rectifier 312
Diac 315
Triac 315
Programmable Logic Controllers 318
Organization 318
Operation 319
Programming 320
Data Acquisition Systems 321

13. Software, Programming and Electronics 325

Introduction 325
Software Fundamentals 325
Programming Languages 327
Programming 329
Developing Software for Embedded Controllers 334
Microcontroller Unit Interfacing 336
Platforms for Learning Programming 339
Parallax BASIC Stamp 2 339
PICAXE 339
Arduino 340
LaunchPad 341
Advanced Alternatives 341
Learning Programming 342
Books 342
Learn by Example 343

Online Tutorials 343
Online College Courses 343
Apply What You Learn 343

14. Soldering 347

Introduction 347
What is Solder? 347
Soldering Tools 348
Soldering Procedures 349
Desoldering 351

Appendix A: Book References 355
Appendix B: Projects Parts List 357
Appendix C: Learning Resources 363
Index 365

List of Figures

Figure 1.1 Major sectors of the electronics industry and common applications 04
Figure 1.2 General block diagram of how the electronics industry works from raw materials to end users 06
FIgure 1.3 Model of how all electronic circuits and equipment work. Input signals are processed by circuits and equipment into new output signals 08
Figure 1.4 Model showing how a cell phone communicates with a cell site to make a call 08
Figure 1.5 Block diagram representing any digital computer, be it a mainframe, PC, or embedded controller 10
Figure 1.6 The most common form of robot emulates a human arm and hand to perform work automatically under the control of a computer 12
Figure 2.1 Physics model of what a copper atom looks like 17
Figure 2.2 Basic electrical circuit used to describe every electronic circuit 19
Figure 2.3 Permanent bar magnet illustrating the magnetic lines of flux that surround it 21
Figure 2.4 (A) Current in a wire produces a magnetic field around it. (B) Coiling the wire increases the strength of the magnetic field 22
Figure 2.5 Relative motion between a magnetic field and a conductor such as a wire causes a voltage to be induced into the wire making it a voltage source 23
Figure 2.6 DC voltages. (A) DC voltage source such as a battery or power supply. One side is usually grounded to serve as a reference. (B) Fixed, continuous positive DC voltage shown over time. (C) Negative DC voltage. (D) Varying positive DC voltage. (E) DC voltage pulses 25
Figure 2.7 AC voltages. (A) AC voltage source drawn as a circle with a wave inside. The output polarity changes periodically. (B) Sine wave, the most common AC voltage shape. (C) Rectangular wave or AC pulses. (D) Triangular wave. (E) Random AC wave that may represent voice or video 27
Figure 2.8 Basic concept of any battery 29
Figure 2.9 Battery and cell symbols used in circuit diagrams. (A) Cell. (B) Four cells in series making a battery. Cell voltage add-up. (C) Symbol for battery of any number of cells. (D) Cells in parallel to increase current capability 31
Figure 2.10 Panel comprised of solar cells in series and parallel make up DC source that charges the battery in a spacecraft 32
Figure 2.11 Frequency-domain views. (A) 1-MHz sine wave. (B) 1-MHz square wave 35
Figure 2.12 A low cost digital multimeter that measures DC and AC voltages, current, and resistance 36

Figure 2.13 A simple solar power system 38

Figure 2.14 An experimental solar power system. The solar panel feeds the charge controller below the panel. The charge controller connects to and charges the black battery on the right. The other charge controller output connects to the inverter below. The lamp plugs into the inverter. Clip leads are used to make the temporary connections 39

Figure 3.1 Hierarchy of electronics from systems to components 43

Figure 3.2 System block diagram of typical personal computer 44

Figure 3.3 System block diagram of iPod or MP3 music player 44

Figure 3.4 Switch and how it is used 46

Figure 3.5 Types of resistors and schematic symbol 47

Figure 3.6 Standard resistor color code 48

Figure 3.7 Voltage divider, the most common resistor circuit 48

Figure 3.8 Capacitor construction and schematic symbol 49

Figure 3.9 (A) Charging capacitor, (B) discharging capacitor 49

FigurE 3.10 Low-pass filter, a common capacitor use 50

Figure 3.11 How capacitor passes AC sine wave but blocks DC voltage 51

Figure 3.12 Schematic symbol for inductor 51

Figure 3.13 Schematic symbol for transformer 51

Figure 3.14 Diode, its schematic symbol, and how to bias it for conduction or cutoff 52

Figure 3.15 How a diode rectifies AC into DC 53

Figure 3.16 Schematic symbol for bipolar junction transistor (BJT) and its inputs and outputs 53

Figure 3.17 Schematic symbol for MOSFET and its input and outputs 54

Figure 3.18 Simple BJT amplifier and how it works 54

Figure 3.19 Simple MOSFET switch and how it works 55

Figure 3.20 Typical integrated circuit today. This one is a dual-satellite TV tuner for set-top boxes. The size is only 7 × 7 mm. The pins are soldered to a printed circuit board 56

Figure 3.21 Common breadboarding socket for building prototypes or just experimenting 57

Figure 3.22 Soldering wires on the battery clip so they will plug into the breadboarding socket 58

Figure 3.23 Measuring battery voltage with multimeter 58

Figure 3.24 This is the circuit used to light an LED and how to wire it on the breadboard 59

Figure 3.25 Circuits for charging and discharging capacitor and respective wiring 60

Figure 3.26 Circuit that turns an LED off and on with MOSFET and wiring 61

Figure 4.1 Linear circuit has a straight-line response. As input varies, output varies by factor of 10 larger, producing straight-line output 64

Figure 4.2 Block symbol for amplifier. Gain, A, is often given. DC power inputs are not usually shown 65

Figure 4.3 Single-ended (A) and differential input (B) amplifiers. Inputs and outputs are referenced to ground 66

Figure 4.4 Class A amplifier conducts continuously 67

Figure 4.5 Class B amplifier conducts half the time, making it more efficient 67

Figure 4.6 Class C amplifier only conducts for part of a half cycle, but the current pulse stimulates resonant tank circuit that oscillates at desired frequency 68

Figure 4.7 Class D amplifier uses pulse width modulation (PWM) at higher frequency to switch higher voltages for greater output. Class D amplifiers are more efficient than any other type, and thus do not produce as much heat 69

Figure 4.8 Most commonly used operational amplifier circuit configurations. (A) Inverting amplifier, (B) noninverting amplifier, (C) follower, and (D) differential amplifier 70
Figure 4.9 Instrumentation amplifier widely used to amplify small sensor signals in industrial applications 71
Figure 4.10 Amplifier input impedance (A) and output impedance (B) 72
Figure 4.11 When one amplifier feeds another, input and output impedances form a voltage divider that attenuates the signal and offsets some of the gain 73
Figure 4.12 Cascading amplifier provides more gain. Total gain is product of individual gains or sum of decibel (dB) gains 73
Figure 4.13 Frequency-response curves show output versus input frequency. Curve in (A) is response of operational amplifier that extends from DC to 5-MHz cutoff frequency. Curve in (B) is response of common stereo audio amplifier 75
Figure 4.14 Four common filter types and their response curves and symbols. (A) Low-pass, (B) high-pass, (C) band-pass, and (D) band-reject or notch 76
Figure 4.15 Types of oscillators and their symbols. (A) Sine wave, (B) rectangular pulse, (C) voltage-controlled oscillator, and (D) crystal clock oscillator 77
Figure 4.16 Linear mixing with resistors (A) or an operational amplifier summer (B) 79
Figure 4.17 Nonlinear mixer used to generate higher and low frequencies 79
Figure 4.18 Phase detector converts phase shift into proportional DC average 80
Figure 4.19 Phase-locked loop 81
Figure 4.20 Frequency synthesizer using phase-locked loop. Changing the divider ratio changes the frequency 83
Figure 4.21 Typical power supply configuration using bus architecture, regulators, and DC–DC converters to get desired number of outputs 85
Figure 4.22 Concept of duty cycle and how varying pulse widths can be filtered into a proportional DC average voltage 86
Figure 4.23 Common DC-to-AC solar power system 88
Figure 4.24 The Velleman PCSU1000 oscilloscope. It has dual channels for observing two signals at once and a bandwidth of 60 MHz 89
Figure 4.25 Virtual instrument units. Left: National Instruments' myDAQ, Right: Digilent Analog Discovery 2 90
Figure 4.26 An AC–AC wall adapter transformer and a pot form a variable 60 (50) Hz signal source 92
Figure 4.27 A 555 timer IC circuit that generates rectangular waves 92
Figure 4.28 One type of breadboarding socket. Not all sockets have the A and B power and ground buses in which case you can designate any one group of 5 pins to power and another for ground 93
Figure 4.29 A 60-Hz sine wave displayed on an oscilloscope 94
Figure 4.30 The output of the 555 timer circuit displayed on an oscilloscope 96
Figure 4.31 The 555 IC connected as an audio tone generator with speaker 97
Figure 4.32 The 555 IC connected to flash an LED off and on 97
Figure 4.33 A half wave rectifier and regulated 5-V power supply 98

Figure 4.34 The half sine pulses as seen on the oscilloscope 99
Figure 4.35 Operational amplifier inverter circuit showing power supply detail 100
Figure 4.36 Operational amplifier noninverting amplifier circuit 101
Figure 5.1 Binary signal that represents 0 and 1 as voltage levels 104
Figure 5.2 Off/on switches are used to enter binary data into digital circuits 105
Figure 5.3 How lights or LEDs are used to represent binary data 106
Figure 5.4 Parallel binary data transfer over bus 112
Figure 5.5 Serial data transfer where 1 bit occurs after another in time 113
Figure 5.6 Logic symbols for inverter and truth table 114
Figure 5.7 Logic symbols for AND gate and truth table 115
Figure 5.8 How AND gate is used to turn signal off or on. (A) The AND gate showing the input and output names. (B) The waveforms of the inputs and output 116
Figure 5.9 Logic symbols for OR gate and truth table 117
Figure 5.10 Logic symbols for NAND gate and truth table 118
Figure 5.11 Logic symbols for NOR gate and truth table 118
Figure 5.12 Logic symbols for XOR and XNOR gates and truth tables 119
Figure 5.13 Logic symbols for flip-flops. (A) RS. (B) D type. (C) JK 119
Figure 5.14 Toggling T input on JK flip-flop causes it to change states, producing frequency division by 2 120
Figure 5.15 Cascading JK flip-flops produce higher-frequency division ratios, which is also a binary counter that counts from 0000 to 1111 120
Figure 5.16 Storage register from one word consists of FFs. It also has clock/load and clear/reset inputs 121
Figure 5.17 (A) Shifting serial data into shift register. (B) Shifting serial data out of shift register 122
Figure 5.18 Shift registers used for (A) serial-to-parallel data conversion and (B) parallel-to-serial data conversion 123
Figure 5.19 Counter storing binary value equal to number of input pulses that occurred 123
Figure 5.20 Digital multiplexer (mux) that selects one of four inputs to appear at the single output 124
Figure 5.21 Decoders. (A) 1-AND gate decoder that identifies input 0110. (B) 4-to-16 decoder that identifies 1 of 16 different 4-bit input values 125
Figure 5.22 Basic concept for random access memory (RAM). Address selects one of eight storage locations. Then data may be read from that location or data stored in it 126
Figure 5.23 RAM storage cells. (A) RS FF for SRAM. (B) Capacitor cell for DRAM 127
Figure 5.24 Storage cell used in EEPROM and flash memories is MOSFET with a special floating gate that keeps transistor off or on 128
Figure 5.25 PLD concept. (A) Block concept of programmable AND and OR gate arrays. (B) Simple circuit of PLD showing programming by dots on AND and OR gate arrays 130
Figure 5.26 Analog-to-digital conversion is performed by sampling analog signal at equal intervals and generating proportional binary value 132
Figure 5.27 Symbol for ADC with both parallel and serial output examples 133
Figure 5.28 Converting binary data back into analog signal with DAC 133
Figure 5.29 DAC output is stepped approximation to original analog signal 134
Figure 5.30 A generic block diagram of a direct digital synthesizer (DDS) 135
Figure 5.31 Pin out details of the 74HC00 quad NAND logic IC, a NAND test circuit with LED output, and a truth table 136
Figure 5.32 The output of a 555 timer is the input to a NAND gate 137
Figure 5.33 A 74HC161 IC binary counter clocked by a 555 timer to demonstrate the binary count sequence on LEDs 138

Figure 6.1 How a microcomputer controls external devices 141
Figure 6.2 How a microcomputer monitors external devices 141
Figure 6.3 Controlling liquid level in a tank automatically with a microcomputer 142
Figure 6.4 General block diagram of any digital computer or microcontroller 143
Figure 6.5 How data is manipulated in a register. Examples of shifting data, incrementing and decrementing a register, and clearing (resetting) a register to 0 145
Figure 6.6 How ALU adds or otherwise processes two data values 146
Figure 6.7 Program counter identifies the address in memory to be accessed 147
Figure 6.8 Three common instruction-word formats with addresses or data 149
Figure 6.9 Example of program stored in RAM for execution using various CPU registers 150
Figure 6.10 Only one source may transmit at a time on a bus. Others are disconnected. All destinations can receive the transmitted data 152
Figure 6.11 General block diagram of 8051 microcontroller. It is very similar to most other controllers in terms of major circuits and their interconnections 153
Figure 6.12 ASCII letter "j" being sent over an RS-232 interface using a UART 156
Figure 6.13 Basic makeup of digital signal processor and its ADC and DAC I/O 157
Figure 7.1 Model of electronic communications system 160
Figure 7.2 Ways of communicating. (A) Simplex—one way and (B) duplex—two way: half—alternating send/receive and full—simultaneous send/receive 161
Figure 7.3 How an antenna creates a radio wave. (A) Antenna produces both electric and magnetic fields. (B) The electric and magnetic field lines are at a right angle to one another and to the direction of radiation 163
Figure 7.4 Electromagnetic frequency spectrum 165
Figure 7.5 Radio propagation of shortwaves 169
Figure 7.6 Line-of-sight (LOS) communications at VHF, UHF, and microwave 170
Figure 7.7 Amplitude modulation 171
Figure 7.8 Sidebands and bandwidth 172
Figure 7.9 Frequency modulation 173
Figure 7.10 Types of digital modulation. (A) ASK. (B) OOK. (C) FSK. (D) BPSK 175
Figure 7.11 Radio transmitter 177
Figure 7.12 Superheterodyne receiver 178
Figure 7.13 Single-chip UHF transceiver for low-speed, short-range digital data transmissions 180
Figure 7.14 Quarter-wave ground-plane antenna 182
Figure 7.15 Dipole antenna. (A) Dipole and (B) horizontal radiation pattern 183
Figure 7.16 Yagi antenna. (A) Yagi and (B) radiation pattern 184
Figure 7.17 Printed circuit board (PCB) patch antenna 184
Figure 7.18 Coaxial cable transmission line 185
Figure 7.19 Microwave horn antenna fed by waveguide transmission line 185
Figure 7.20 How a parabolic dish works. (A) Parabolic dish with horn, (B) transmission, and (C) reception 186
Figure 7.21 Concept of communications satellite 189
Figure 7.22 Satellite transponder 190
Figure 7.23 Wiring details for the ISM band short-range radio experiment 194
Figure 8.1 (A) Cell coverage of a given area. (B) Cell area covered depends on antenna pattern and environment 198
Figure 8.2 How small and large cells cover an area 198
Figure 8.3 Cell phone basic operations 200
Figure 8.4 Block diagram of basic cell phone 201

Figure 8.5 How QPSK is generated using mixers and a local oscillator shifted 90 degrees to produce in-phase (I) and quadrature (Q) signals that are summed 203
Figure 8.6 QPSK diagrams. (A) Phasor. (B) Constellation 204
Figure 8.7 How 16-QAM is generated and resulting constellation diagram 205
Figure 8.8 OFDM divides a channel into many subchannels or subcarriers, each of which is modulated by part of digital data to be transmitted 205
Figure 8.9 Generating OFDM is done by DSP and I/Q modulation 206
Figure 8.10 MIMO in a 4 × 4 configuration quadruples the data rate in an existing channel, while reducing the ill effects of multipath fading 207
Figure 8.11 The concept of MU-MIMO using beamsteering with a patch-type phased array antenna 208
Figure 8.12 Frequency division multiple access and frequency duplexing use two separate bands of frequencies divided into individual channels 210
Figure 8.13 TDMA multiplexing divides a channel into sequential time slots for the digital data 210
Figure 8.14 How CDMA is generated 211
Figure 8.15 A block diagram of 4G cell phone with multiple radios and functions 212
Figure 9.1 How LANs, MANs, and WANs connect to form a huge worldwide network of computers and other devices 220
Figure 9.2 Three main networking topologies—star, ring, and bus. Bus is the most widely used 221
Figure 9.3 Coax cable. It is now mainly used in cable TV networks. The most common is RG-6/U and uses F-type or BNC connectors 222
Figure 9.4 Twisted-pair cable. Initially used for telephones and is still used. Computer network cables use four twisted pairs. A shielded version is available where noise is a problem 222
Figure 9.5 Fiber optic cable. It is nonconductive because it is made of plastic and glass, and is used to transmit data as infrared (IR) light pulses 223
Figure 9.6 How binary data is transmitted over fiber optic cable 224
Figure 9.7 Ethernet hub or switch implements a bus that is accessed by RJ-45 connectors and twisted-pair cable 225
Figure 9.8 Ethernet frame or packet 225
Figure 9.9 A T-1 data frame commonly used in digital telephone systems as well as data interconnections in networks 227
Figure 9.10 Internet illustrated as cloud of fiber backbone connections through network access points (NAPs) that attach to WANs, MANs, and LANs 228
Figure 9.11 Protocols used in Internet. (A) Transmission control protocol (TCP). (B) Internet protocol (IP). The data in each frame are transmitted from left to right and top to bottom 230
Figure 9.12 Summary of short-range wireless networking technologies shown as data rate versus range 231
Figure 9.13 How a Wi-Fi access point or hot spot works 232
Figure 9.14 A mesh network can extend the overall range of a node and provide greater reliability through multiple possible paths. Most wireless mesh networks are ZigBee, but other wireless technologies may be used (e.g., Wi-Fi) 234
Figure 9.15 How RFID works. The tag has no power of its own but gets power from the transmitter in the reader 236
Figure 9.16 How the Internet of Things systems works 240
Figure 10.1 Cross section of speaker showing major components and how it works 244
Figure 10.2 Cross over networks keep signals of different frequencies separate for appropriate speakers 245

Figure 10.3 How audio signal is digitized. In this example, each sample is assigned a 4-bit code. The stepped approximation shown is what a DAC would deliver to the speakers 248
Figure 10.4 Voice signals up to 4 kHz for the telephone are sampled at an 8-kHz rate, producing an 8-bit sample, each 125 μs. Stringing the samples together in a serial data stream produces a digital signal at a 64-kHz rate 249
Figure 10.5 General block diagram of a digital music player like Zune MP3 or Apple iPod 250
Figure 10.6 Digital music on a CD is encoded as pits and lands on the CD surface. These are burned in by a laser. A laser light shined on the pits and lands produces a reflection that a photo-detector picks up to create the original digital data 252
Figure 10.7 Block diagram of CD player 253
Figure 10.8 Block diagram of typical audio–video receiver, the heart of the audio reproduction part of a consumer home entertainment center 255
Figure 10.9 How 5.1 surround-sound system is set up 256
Figure 11.1 How scanning converts a scene of light and color into an electronic signal 263
Figure 11.2 TV screen size and aspect ratio 264
Figure 11.3 Interlaced scanning helps reduce flicker. Sync pulses help keep the receiver and transmitter in step with one another 265
Figure 11.4 Digital TV is created by digitizing the analog video signals with an ADC, and then converting it into a serial bitstream of video data words 267
Figure 11.5 TV screens can be thought of as thousands of rows and columns of pixels, each represented by data 267
Figure 11.6 Any color of light can be reproduced by mixing red, green, and blue light in the right proportion 269
Figure 11.7 Red, green, and blue digital color signals are generated by filtering the scene into its red, green, and blue components, and then digitizing the resulting analog signals 270
Figure 11.8 HDTV packet for ATSC US digital color TV system 272
Figure 11.9 General block diagram of TV transmitter 272
Figure 11.10 US ATSC system translates the digital bitstream into an eight-level signal to transmit multiple bits per level, making higher data rates in a given bandwidth possible 273
Figure 11.11 Simplified block diagram of digital TV set 275
Figure 11.12 Inputs and outputs of modern digital TV set 276
Figure 11.13 How color pixels are formed on a TV screen. Pixel format typical of LCD or LED screens 277
Figure 11.14 Coax cable bandwidth up to 1 GHz can contain about 130 6-MHz channels 279
Figure 11.15 General block diagram of modern cable TV system 280
Figure 11.16 Simplified block diagram of modern cable box with DVR 282
Figure 11.17 TV distribution satellites in geosynchronous orbits act as wireless relay stations 283
Figure 11.18 Video from all sources is transmitted uplink to satellite that rebroadcasts it back to Earth over downlink to home receivers 284
Figure 11.19 Satellite TV home system consists of antenna with low-noise block converter, receiver, and TV set 284
Figure 11.20 Simplified block diagram of satellite TV receiver 285
Figure 11.21 Block diagram of standard DVD player 287
Figure 12.1 General block diagram of industrial control process 292
Figure 12.2 Open-loop control system for filling vat with liquid via a pump 292
Figure 12.3 Generic diagram of any closed-loop control system with feedback 293

Figure 12.4 Filling vat with liquid automatically with closed-loop control system 294
Figure 12.5 Closed-loop control system for motor speed control and regulation 295
Figure 12.6 Most resistive sensors such as an RTD or thermistor are placed in a bridge circuit to condition them to provide amplified output voltage 296
Figure 12.7 Characteristic curve of thermistor and simple voltage divider circuit producing output voltage 297
Figure 12.8 Strain gauge is a resistive element whose value changes with stress or pressure 298
Figure 12.9 Schematic symbols for photo cells. (A) Photoconductive or photoresistive, (B) photovoltaic or solar cell, and (C) photo transistor 299
Figure 12.10 Photodiode and its simple output circuit 300
Figure 12.11 Simple off/on photo control circuit. Light turns on automatically at dark and off when light is present 301
Figure 12.12 Simple tachometer circuit using phototransistor to create pulses that are averaged into a DC voltage proportional to speed 302
Figure 12.13 Cross-sectional view of typical relay showing major components and schematic diagram 303
Figure 12.14 Common relay contacts. (A) Form A, (B) form B, and (C) form C 304
Figure 12.15 Reed relay 305
Figure 12.16 Controlling a large current with a smaller current with a relay 306
Figure 12.17 Latching a relay with its own contacts 308
Figure 12.18 Operating a relay with a transistor is a common application. The diode protects the transistor when power is removed from the relay coil 308
Figure 12.19 A solenoid produces linear motion using magnetic principles 309
Figure 12.20 Using an H-bridge circuit to reverse the current flow and direction of rotation in a DC motor 311
Figure 12.21 Schematic symbol of an SCR 313
Figure 12.22 Circuit that shows how an SCR is turned on and off 313
Figure 12.23 Using an SCR to control power to a load using AC. The SCR rectifies the AC 314
Figure 12.24 Schematic symbol for triac 315
Figure 12.25 Equivalent circuit of triac shown with SCRs 316
Figure 12.26 Lamp dimmer circuit using RC phase shifter and diac to control current in the load 316
Figure 12.27 AC waveforms in a lamp load controlled by a triac. (A) Waveforms with triggering at 60 degrees, (B) waveforms with triggering at 120 degrees. The longer pulses produce higher average current in the load 317
Figure 12.28 General block diagram of PLC 318
Figure 12.29 PLC operational scan cycle 319
Figure 12.30 Sample ladder diagram and PLC application program 321
Figure 12.31 The key components and organization of a typical data acquisition system 322
Figure 12.32 A simple feedback control circuit for temperature control 323
Figure 13.1 Flowchart symbols 330
Figure 13.2 A flowchart illustrating a computer program 331
Figure 13.3 Structured programming divides a problem into smaller and smaller submodules 332
Figure 13.4 Block diagram of the complete microcontroller software development process 335
Figure 13.5 Driving an LED directly from the MCU output 336
Figure 13.6 An NPN driver transistor for higher loads 337
Figure 13.7 An N-channel enhancement mode MOSFET driving a DC motor 338
Figure 13.8 Using a pushbutton switch to generate an input logic level 338
Figure 13.9 A comparator and thermistor used for temperature sensing 338

Figure 13.10 The Parallax BASIC Stamp Activity Kit experimental board is designed to be operated from a 9-V battery. The PIC MCU is the IC above the reset button in the center 340
Figure 13.11 The Arduino Uno R3 is one of the most widely used microcontroller learning platforms 341
Figure 13.12 The Texas Instruments' LaunchPad development board uses an MPS430 microcontroller 342
Figure 14.1 A common soldering iron used in electronic work 348
Figure 14.2 Tools used in soldering. On the left is a wire stripper. Next is a pair of needle-nose pliers. On the right is a side cutter for wire 349
Figure 14.3 Soldering two wires together. Apply solder to the heated wires not to the iron tip 350
Figure 14.4 Soldering components to a printed circuit board 350
Figure 14.5 Soldering a surface mount part to a PCB 351
Figure 14.6 Tools for desoldering, a solder sucker and solder wick 352

This page intentionally left blank

Preface to the Second Edition

This revised and enhanced edition of *Electronics Explained* is still written for those of you who want to learn about electronics in the shortest amount of time. The original goals of the book as outlined in the Preface below are the same, but now the book has been expanded and updated to match the latest developments in electronics. This revised version also hopes to make you more of a hands-on electronic experimenter, hobbyist, DIYer, or maker. The highlights of the new edition are as follows:

- All chapters are revised to include any new developments, technologies, and products. Some of the new topics covered include 4K high-definition video, virtual reality, the latest wireless standards, the Internet of things, and the latest 4G and 5G cellular technologies.
- Increased number of hands-on projects with electronic components and test equipment. Virtual instruments are introduced and a wide range of analog and digital projects and experiments are provided.
- A new chapter on soldering, a core skill of any electronic person. Building a kit is a recommended project.
- A new chapter on programming and interfacing embedded controllers. It is difficult to be an electronics enthusiast today without knowledge of microcontrollers and how to program them. Modules like the Arduino are described and examples of programming using BASIC and C languages are included.

Happy learning and experimenting

Lou Frenzel
2017

This page intentionally left blank

Preface

This book is for those of you who want to learn about electronics and computers but frankly are put off by all the gory details. If you have a short attention span, a nonexistent background in science, math, and technology, and are looking for near-instant gratification of electronics knowledge, this book is for you. Electronics is a complex subject, especially if you want to be an electronics engineer. However, if you are merely curious or just need a working knowledge of electronics for your job, this book will give you a basic education on the fundamentals without boring you to death or loading you down with the excruciating details that other books lay on you. What I have attempted to do is write a book that tells you the most important concepts and fundamentals but leaves out the minutiae that only an engineer will need or appreciate. This book contains exactly what you need to know today in the 21st century.

There are lots of basic electronics books around, and you may already have read some. This one is really different. First, it does not go into detailed circuit analysis. Why? Simply, all electronic equipment today is made up of integrated circuits or chips, and we cannot see the circuits, access them, or even repair them. If you are an electronic engineer designing integrated circuits, yes, you do need to understand their operation. If you just want to learn how electronic equipment works, then you can skip this detail. This book treats circuits as simple functional building blocks whose purpose and application are easily understood.

Second, this book presents what I call the *new* electronics. Most basic electronic books dwell on the older components and circuits and ignore how electronics really is today. These books do a good job teaching you the history of electronics but little else. What you will see in this book is a fresh new way to look at and understand electronics, especially in the context of the familiar applications you use every day made with the most advanced integrated circuits. If you complete this book and still want more, one of the more traditional basic electronics books would be your next step. See the appendix for my recommendations.

The first three chapters introduce you to the electrical principles that form the basis of electronics. You will learn the core concept of electronics, which boils down to how to generate current flow and then control it to do something useful. Magnetic fields are also covered, since magnetism is just as important in electronics as current flow. But do not worry, even if you never took a physics course you can easily get through this material.

Next, you learn about the basic components of electronics such as resistors, capacitors, inductors, transformers, diodes, transistors, and especially integrated circuits. Then you will discover all the different types of electronic circuits used to make up any electronic device. The functional block diagram approach is used here so that you will not have to do the nitty-gritty circuit analysis that has you chasing electrons through components and solving simultaneous equations.

An important part of this book is the coverage of computers, and specifically of embedded controllers, that is, those tiny single-chip microcomputers or microcontrollers that are built virtually into every electronic device. Digital signal processors are also covered since they also appear in most modern electronic equipment. You cannot name an electronic product today that does not have one of these embedded computers in it. You really need to know what these are and how they work, as they are the very heart of electronic devices today. Programming and interfacing are included with C and BASIC language examples.

The rest of the book consists of descriptions of how all the major electronic products and systems work. The entire range of modern electronics is covered, including major things like cell phones (LTE and 5G), video (cable, HDTV, 4K, DVD, VR, OTT, etc.), audio (stereo, CDs, MP3, streaming, etc.), satellites (GPS, TV, radio, etc.), radio (broadcast, marine-aircraft-shortwave, ham, CB, family radio, etc.), and everything else you can think of. And it is all explained using basic building blocks rather than detailed circuits.

Finally, if you want involvement as you read the book, get involved with the projects that I list in the project sidebars in each chapter. These can lead to a deeper understanding of the concepts. Furthermore, these projects could encourage you to become a true electronic hobbyist—or is that a geek or nerd?

When you finish this book you will not be an electronic engineer nor will you be able to design or analyze circuits. But you will know about all the different kinds of circuits, how they work, and how they all fit together to create electronic equipment. The benefits to you are the ability to apply, use, select, operate, and to some extent troubleshoot most of the common electronics products with confidence, while appreciating their complexity and the value they deliver. Hope you like it.

Lou Frenzel
Austin, TX

Chapter 1

Learning About Electronics for Your Own Good

A Little Bit of Perspective Before You Get to the Details

In this Chapter:

- Why it is helpful to know electronics.
- The 10,000-foot view of the electronics industry and how it works.
- The systems view of electronics.

INTRODUCTION

You do not know how lucky you are. You live in high-tech heaven surrounded by lots of useful, entertaining, interesting, necessary, and even addicting electronic products, products you cling to and would not want to give up.

But do you have any idea how many electronic products you own? Do you really know how much electronics influences your life every day? Some of you do, but most of you probably do not because your electronic gadgets are pretty much taken for granted. Yet we spend the better part of our lives working with these devices. Most of us sit in front of a computer all day at work, slouch on the couch in front of the TV at night, listen to our iPods going to and from work, and spend a weekend downloading videos off of the Internet or taking photos with a digital camera. And let us be clear here, how many text messages did you send today or cell calls did you make? The impact on our lives of all this electronic equipment is almost overwhelming. In this chapter, I encourage you to think about how electronics impacts your life.

And that brings us to the subject of this book, learning how electronics works. If you want a first book in electronics that gives you the big picture about how electronic equipment works, this is it. I am not going to beat you to death with a lot of complex circuits and theory or math, only what you really need to know. Then I will go on to show you how each and every one of the products you use every day actually works.

Electronics Explained. http://dx.doi.org/10.1016/B978-0-12-811641-8.00001-1

Copyright © 2018 Elsevier Inc. All rights reserved.

IT IS NOT LIKE IT USED TO BE

It used to be that you learned electronics by playing around with resistors, capacitors, transistors, and other devices. You wired up simple circuits, built a one- or two-transistor radio, or even built a kit. But today electronics is so different that it is hard to do that anymore. Oh yes, you can still build little circuits and radios and other gadgets, but that is not how electronic products are made today. Electronic equipment today is made with integrated circuits or ICs that we also call chips. These chips are tightly packed on printed circuit boards with resistors and capacitors that you can barely see. It is more difficult than ever to experiment with these devices, much less troubleshoot and repair them. We are in that age of electronics where unless it is a big pricey piece of equipment, you throw a defective electronic product away and simply buy a newer and better one at a lower price.

Despite the complexity of today's electronic products, there is something just fascinating about them. Many of you actually want to know how they work. And despite their complexity, these products are actually understandable. It is possible to learn how these devices work without a whole lot of trouble. And you do not have to learn all the math and physics required by engineers to do so. This book takes the basic ideas of electronics and reduces them down to the lowest common denominator and shows you how to use those fundamentals and apply them to all of the fabulous electronic gadgets you have today.

WHAT IS IN IT FOR YOU?

There are some big benefits to being electronic literate. Take a look at some of them.

Satisfy your curiosity—Surrounded by so much electronics, you often probably wondered just how all this stuff works. For some of us, not knowing how something works nags at us until we satisfy that curiosity. If you have an interest in electronics, you tend to know more about it. This book will get you started in the right direction.

Enhance your hobby—If you dabble in a hobby that uses electronics, you obviously want and need to know more about it. Lots of hobbies involve electronics such as radio-controlled models, electric trains, amateur radio, shortwave listening, citizens or family radio, audio systems, or video. Or maybe you just like playing around with your home security system, garage door opener, or your high-speed Internet connection. A major part of any hobby is learning more about it and that is very satisfying. And do not forget, electronics itself can be your hobby. You can spend your time learning more about it and then digging deeper into those things that you want to know more about. Amateur radio is certainly one of those electronic hobbies that will take you from one extreme to the next in the radio spectrum. Anything having to do with computers is also a great hobby. With everything being controlled by an embedded microcomputer, spending more time learning about them is a great activity with nearly endless possibilities.

You are a maker—A maker is a serious do-it-yourselfer (DIY) who likes to build things small and large. Makers are hobbyists who like to create projects for home use or just for fun. Projects can be all electronic but more often involve wood or metal working and other skills. Most projects use electrical or electronic parts. That is why learning electronics is so important.

Job-related needs—You may actually be working in a job surrounded by electronic equipment of some sort, yet you do not know how it works. Often, just understanding some of the fundamentals will give you a better grasp of how to use that equipment or even troubleshoot and repair it. Understanding some of the concepts may actually help you to better select and purchase equipment that you may use.

Boost your school performance—For those of you going to a college or university to learn electronic engineering or technology, you regularly have your curiosity stimulated by the classes you take. Yet most of the classes are heavily involved with theory, physics, math, and circuit analysis. What you really do not get in school is the perspective or big picture as to how the electronic products actually work. This book will give you the big picture and show you how it all comes together so that all those detailed classes in circuits and theory will make more sense.

Just so you will not appear to be so dumb—With electronics everywhere, people often talk about it without really knowing what they are talking about. The ignorance of people who use the electronics but do not have a clue as to what it does or how it does is amazing. For instance, do you really understand what an MP3 is, or why 4K UHD TVs are better, or what a 4G or 5G cell phone really is? The answers are in this book. If you want to appear a little smarter to your coworkers, friends, and relatives, this book can help out. You will be able to deal with common questions and misconceptions, such as can you get shocked by a car battery, can you really get brain cancer or pop popcorn with your cell phone, or will sitting too close to your TV set make you sterile? While this book certainly will not make you a genius, it will indeed give you the basics so that you will not appear to be an idiot when technical questions about electronics come up.

ELECTRONICS: THE BIG PICTURE

Before we get into the nitty-gritty of how electronic equipment works, you really need to know the entire scope of electronics. It is a huge industry, and recent figures seem to show that it is one of the biggest, if not the biggest, markets in the entire world. To get a grasp of how big the industry is, we normally break it down into segments of specializations. There are five major fields of electronics: communications, computers, control, instrumentation and measurement, and components. Any electronic product or device will easily fall into one of those categories. In some cases, the product may overlap two or more or all of those categories. Take a look at Fig. 1.1, which shows each of these segments of electronics and the products and technologies that are part of them.

Communications	***Computer***	***Control***	***Instrumentation***	***Components***
Radio	Supercomputers	Automation	Test equipment	Resistors
Television	Mainframes	Robots	Test systems	Capacitors
Telephone	Servers	Manufacturing	Data acquisition	Inductors
Cable TV	Workstations	Automotive	Medical	Transformers
Satellites	Personal computers	Home control	Auto diagnosis	Diodes
Networks	Peripheral equipment	Appliances	Aircraft diagnosis	Transistors
Wireless systems	Embedded	Security		Integrated circuits
Consumer	Special purpose	Toys		Printed circuits
Cell phones				Wire and cable
				Connectors

FIGURE 1.1 Major sectors of the electronics industry and common applications.

Communications

Communications is the oldest segment of electronics and still the largest. Electronics actually started with communications—specifically telegraph, telephone, and radio—and grew from there. A lot of this book is about communications because it is such a huge segment of the industry and one that we use the most. And do not forget, communications comes in both wired and wireless forms. Wired systems include the vast telephone network, cable TV systems, and computer networks, including the Internet. Wireless systems comprise all of the radio, cell phone, television, satellite, radar, and other wireless systems.

Computers

Computers have only been around since the late 1940s and early 1950s but what an impact they have had on the way we work. Today virtually everyone uses and even owns a personal computer. And then there are all the larger and more powerful computers such as the servers that manage all of our networks, the mainframes still used by government and big business to manage huge databases, as well as the high-speed supercomputers that still do an amazing amount of scientific, engineering, and mathematical design and research. But the computer we are going to talk about most in this book is the embedded controller. These are small single-chip digital computers called *microcomputers* or *microcontrollers* or just *micros* that are literally part of every electronic product. These are miniature digital computers dedicated to a specific function inside the products in which they exist.

Control

Control is a broad general term for monitoring and control. Monitoring, of course, means sensing various physical characteristics such as temperature, humidity, physical position, motor shaft speed, or light level. Electronic devices called *sensors* or *transducers* convert these physical characteristics into electrical signals, which allow us to monitor them. Control refers to managing and exercising some degree of influence over items such as motors, lights, relays, heating elements, and other devices. Most monitoring and control functions take place in factories, chemical plants, refineries, and food processing operations. Control also occurs in the home, an example being your electronic thermostat. It also shows up in electromechanical products such as a DVD player. Electronic controls appear in appliances, home control systems, security systems, automobiles, and even toys. Robots are another major segment of the control industry.

Instrumentation

Instrumentation refers to the field involved with testing and measuring electronic equipment and other mechanical or electronic items. It refers to the test instruments such as voltmeters, oscilloscopes, and spectrum analyzers as well as the large automated test systems used for mass testing and measuring of other devices. Instrumentation refers to data acquisition systems, medical tests and measurements, and a wide range of other products.

Components

All of the electronic products that make up each of the four major segments of the electronic industry are made up of various electronic components. In the past, electronic products were made of the individual discrete components such as resistors, capacitors, diodes, and transistors. Today most electronic equipment is made with one or more ICs surrounded by a sprinkling of those other so-called *discrete* components. No longer can we access most individual components or circuits since they are sealed within ICs or chips. Just keep in mind that there is a whole industry centered on making these electronic components that are in turn made available to the manufacturers of the end electronic equipment in each of the four major categories just described.

HOW THE ELECTRONICS INDUSTRY WORKS?

As part of the big picture of learning electronics, you really need to understand how the industry itself works. This is summed up by the simplified block diagram in Fig. 1.2. It all starts with the raw materials that the component manufacturers use to create the individual components. For example, sand is refined into silicon, which, in turn, is converted into wafers that are processed into the ICs we use.

The electronic equipment manufacturers design the electronic equipment for the marketplace. It may be a TV set, a cell phone, or a military radar. These companies buy the electronic components and put them together to create the end product. Manufacturers may then sell the end product directly to customers

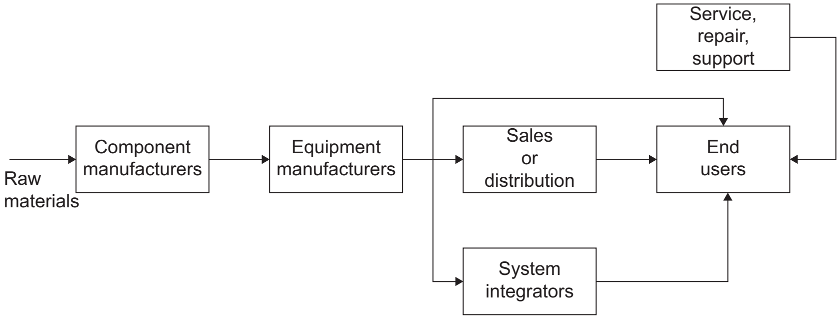

FIGURE 1.2 General block diagram of how the electronics industry works from raw materials to end users.

and end users. In some cases, the end product will go through a sales organization or distributor of some sort. Your LCD large-screen TV set does not come directly to you directly from Samsung in Korea or Sony in Japan. It is sold to an organization such as Best Buy or Amazon, which is where you buy it. Sometimes the electronic equipment is sold to so-called systems integrators. These are companies that piece together larger systems for an end user (like the military) made up of a wide range of different types of equipment.

Finally, there are the customers or end users of the equipment. These come from a wide range of backgrounds. The end users or customers may be businesses with offices, industry with factories and process control plants, or the government, military, hospitals and other medical organizations, schools, colleges and universities, and you the consumer. There is also a network of service companies designed to support customers. They may do installation or repair work that the customer is not capable of doing.

Electricity Versus Electronics

One of the questions that always comes up in the early stages of learning about electronics is: what is the difference between electricity and electronics? Let us get this straight up front. Electricity refers to the basic energy we use for power. It generally involves the generation and distribution of electrical energy to homes, businesses, and factories. Electricity is energy that is produced by the coal-fired or nuclear power plant and then sent to you over long transmission lines until it comes into your house where you access it through your AC wall outlets. You use that energy for lighting and appliances and also to power most of your electronic equipment.

Electrical power may also be generated for use in other systems. The electrical system in a car is a great example. It uses a battery plus a generator driven by the automobile engine to recharge the battery and to operate all of the electrical and electronic devices in a car. Ships, airplanes, and offshore drilling rigs also have their own electrical power–generating systems. The new electric wind generators and solar power systems are also electricity-producing systems.

All of the principles of electricity also apply to electronics. Basic electrical components such as resistors, capacitors, inductors, transformers, motors, and other similar devices are also used in electronics.

Electronics is the field of science that uses electrical principles to perform other useful functions. Electronics is the control of electricity to produce radio, television, computers, robots, and MRI machines. To control the electricity, you need more than just basic electrical components such as resistors and capacitors. What makes electronics possible are the unique components called *transistors*. Electronics began, of course, with vacuum tubes and then switched over to transistors as the main control devices. Transistors are now miniaturized and packaged in devices called *integrated circuits* that perform very specific electronic functions, such as generating signals, amplifying signals, or performing some control function. Just keep in mind that it is the control devices such as transistors and the circuits that they comprise that make electronics what it is. The transistors are used to form circuits that in turn perform all of the various functions required to produce electronic devices.

THE PAINFUL TRUTH

Okay, here comes the hard part. Electricity is moving electrons. An electron is a tiny particle that is inside the atoms that make up everything we can feel, see, and even not see. If we can get some electrons moving together in a wire, electrical or electronic component, or circuit, we can control them to do many useful things, such as amplification, for example. Moving electrons make up what we call *current flow*. And current flow is produced by electrical energy called *voltage*. Voltage comes from batteries, the AC wall outlet, and lots of other sources including power supplies, solar cells, microphones, and the like. If we can control the electrons, we can produce magic results. That is what electronics is all about.

We control the electrons with components such as resistors, capacitors, inductors, transistors, and all sorts of ICs. We create circuits that actually process the electron flow. We can call the voltages that create current flow and the resulting current *signals*. The circuit processes the signals to produce the end effect that we want, be it a radio, a computer, or a DVD player.

One of the basic models I will use in this book is shown in Fig. 1.3. Think of electronic signals as the inputs to a circuit or piece of equipment that processes the signals producing new signals that represent the output we want. Think of the small voltage that comes from a microphone when you talk into it. It is basically too small to be useful. So we apply it to a circuit called an *amplifier* that makes the signal bigger and more useful. That amplifier processes the signal to make it large enough to drive a speaker as our desired output. The result is a

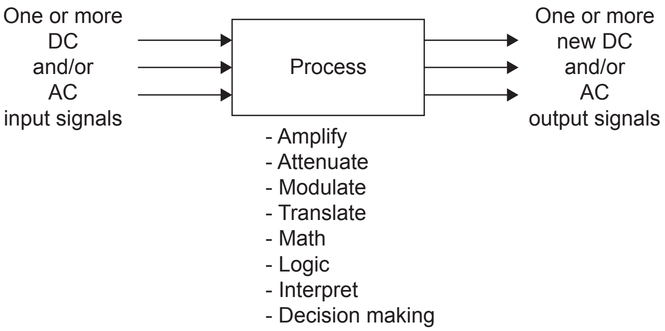
One or more
DC
and/or
AC
input signals
Process
- Amplify
- Attenuate
- Modulate
- Translate
- Math
- Logic
- Interpret
- Decision making
One or more
new DC
and/or
AC
output signals

FIGURE 1.3 Model of how all electronic circuits and equipment work. Input signals are processed by circuits and equipment into new output signals.

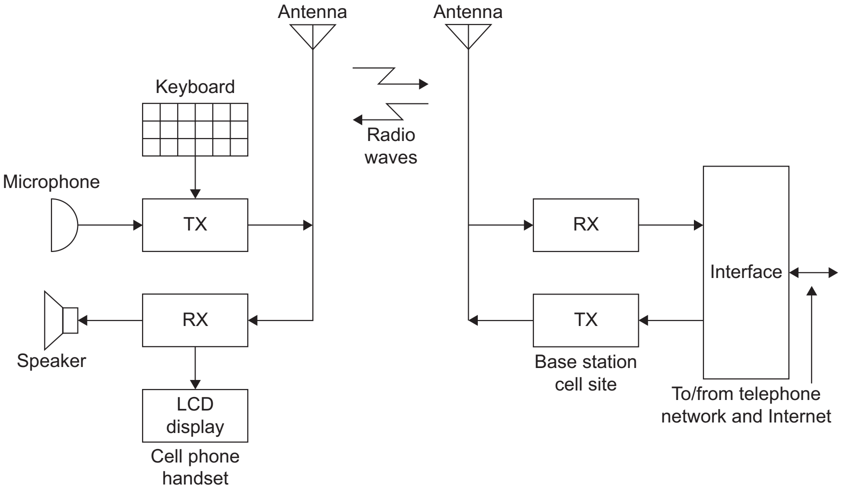
Antenna
Antenna
Keyboard
Radio
waves
Microphone
TX
RX
RX
Interface
TX
Speaker
LCD
display
Base station
cell site
To/from telephone
network and Internet
Cell phone
handset

FIGURE 1.4 Model showing how a cell phone communicates with a cell site to make a call.

public address system. Anyway, you get the picture. Just remember that *all* electronic circuits and equipment are based on this model: input–process–output. Think about it.

Here are a couple of other examples to show how most electronic things work.

Cell Phone

Take a look at Fig. 1.4. On the left is a cell phone handset. It is made up of two major pieces—a transmitter (TX) and a receiver (RX). You talk into a microphone and generate a small electrical voltage or signal representing your voice.

The transmitter takes that and processes it in such a way that the voice controls or modulates a high-frequency radio signal. That signal carrying your voice is then radiated by the antenna to a cell site or base station. The base station receiver picks it up and processes it back into the original voice signal where it then undergoes other processing before being sent over the telephone network to its destination. The dialing process or text messaging from the keyboard also generates input signals that are transmitted to the base station.

On the return side, the voice from your caller passes through the cell phone network to the base station and it is then used to control the transmitter to send that voice to you by radio. The transmitter sends the signal to the antenna that sends the signal to the receiver in your cell phone. The receiver processes the signal by amplifying, demodulating, and otherwise processing the voice that eventually is sent to the speaker that is glued to your ear. Another output may be a text message on the liquid crystal display (LCD).

Computers

The main function of a computer is to store and retrieve data, process data, and to provide automatic control. Data consists of information, facts, and figures. It may be numbers, letters, symbols, or text of any kind. It might represent names and addresses, parts inventory, accounting or banking figures, control signals, or a variety of other items. Data may also be graphics and illustrations, or voice and video that are converted to data. Such data can be stored electronically in computers and rapidly accessed.

Computers process or manipulate data in many different ways. They can perform mathematical operations on it, sort it, edit it, translate it, or otherwise operate on it to produce some useful output. Computers also make decisions based on the data.

Computers have so many applications that it is difficult to enumerate them. In general, computers are tools that help people organize and perform their work in a more efficient manner. Computers also allow people to do work that was difficult or impossible to do without computers. Computers perform accounting and banking operations; store and process mailing lists and employee records; solve complex mathematical problems in business, science, and engineering; and create graphics. The thing that makes them so valuable is that all of their operations are automatic. Humans need not be a part of the actual computation, as is necessary with a calculator. Computers carry out preprogrammed assignments to solve various problems or perform operations. Most operations are extremely fast. It is this automatic feature that makes computers so useful.

All computers, regardless of their size and features, are made up of the four basic sections shown in Fig. 1.5. The memory electronically stores the data to be processed, as well as the program that tells how the data is to be manipulated.

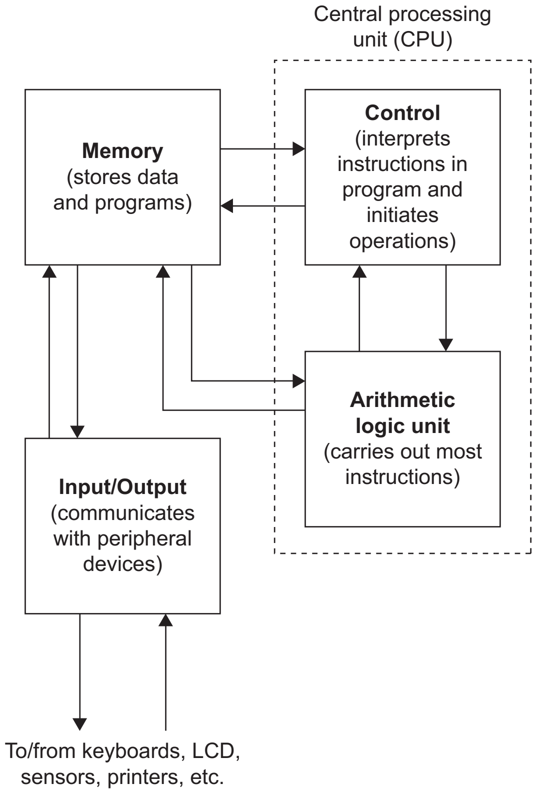
Central processing unit (CPU)
Memory (stores data and programs)
Control (interprets instructions in program and initiates operations)
Arithmetic logic unit (carries out most instructions)
Input/Output (communicates with peripheral devices)
To/from keyboards, LCD, sensors, printers, etc.

FIGURE 1.5 Block diagram representing any digital computer, be it a mainframe, PC, or embedded controller.

The program is a sequence of codes or instructions that cause the computer to carry out its preprogrammed assignment one step at a time.

Refer again to Fig. 1.5. The control section examines the program instructions one at a time in sequence, interprets them, and then sends control signals to all other parts of the computer. The desired operations, such as store, add, print, and so on, are carried out.

The arithmetic logic unit (ALU) is the section that actually carries out most of the processing. It performs data transfers, arithmetic operations, and logical decision-making. In most computers, the control and ALU sections are integral units that work closely together. They are usually referred to as the central processing unit (CPU). A microprocessor in a personal computer is a single chip or a CPU.

The input/output (I/O) section in Fig. 1.5 is the interface between the computer and the outside world. It communicates with peripheral devices that humans use to input programs and data, and receive the output results of computation. Data and instructions are usually entered via a keyboard, while outputs are displayed on a video screen or printed out. Inputs may come from sensors, and output may go to a printer or LCD readout.

There are many sizes and types of computers. Supercomputers are the largest and fastest computers. They are used primarily for difficult, complex scientific and math problems. Mainframe computers are also very large, and they are used by business and government for financial applications, data storage, and other applications. Personal computers are the smallest practical computers in use. Millions of them are used for word processing, financial spreadsheets, database management, and email and Internet access.

There are also many dedicated computers, built into other equipment to perform single-purpose operations unlike the general-purpose data processors described above. Thanks to semiconductor technology, it is now possible to put an entire computer on a single chip you can hold in your hand. These are called *embedded computers* or *microcontrollers*. These microcomputers are used like components and are built into most other kinds of equipment such as TV sets, microwave ovens, gasoline pumps, washing machines, and many others. Virtually every other electronic product contains at least one or more embedded controllers. And even very large ICs have controllers on the same chip called *cores*.

Robots

A great example of a control device is a robot. Most robots consist of an electromechanical manipulator arm used in manufacturing operations. Robots are machines designed to mimic human capabilities. Most robots simulate a human arm and hand. As a result, a robot may be able to substitute for a human in certain manufacturing jobs. A typical robot manipulator arm is shown in Fig. 1.6. Motors operate shoulder, elbow, and wrist joints, while grippers and other tools form the hand that does the work.

Robots help automate the manufacturing process to increase productivity. They lower costs, produce more goods in less time, and improve quality. Robots also replace humans in many jobs requiring repetitive, tedious, and hazardous work.

Robots perform a variety of tasks in industry. The most common are welding, spraying or painting, and “pick-and-place” operations. Various tools on the end of the manipulator arm, called *end effectors*, let the robot perform a remarkable variety of tasks. In addition, a programmable controller or computer is used to control the robot. A robot may thus be programmed to perform a wide variety of jobs and is easily reconfigured when the work changes.

In subsequent chapters you will see more of these kinds of descriptions. But in Chapters 2–4, you will learn a few more basics that make the rest of the book more understandable.

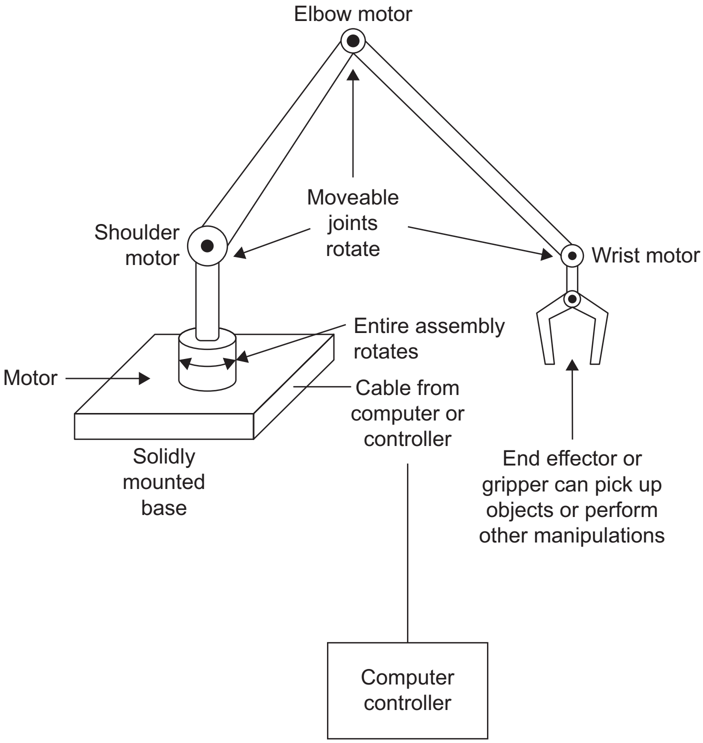
Elbow motor
Moveable joints rotate
Shoulder motor
Wrist motor
Entire assembly rotates
Motor
Cable from computer or controller
Solidly mounted base
End effector or gripper can pick up objects or perform other manipulations
Computer controller

FIGURE 1.6 The most common form of robot emulates a human arm and hand to perform work automatically under the control of a computer.

Project 1.1

How Much Electronics Do You Own?

Just to be sure that you really appreciate the extent to which electronics impacts your life, I strongly suggest that you take a few minutes and make an inventory of all the electronic devices and equipment that you use at home, at work, and away from home. Get a yellow pad and a ballpoint pen and move around your house and just write down all the electronic products that you own and use on a regular basis. TV sets, computers, and cell phones are the most obvious, but do not forget things such as telephones, your wrist watch, appliances, and other entertainment products. Make a separate listing of the electrical and electronic products you use inside your car and then another one for those electronic products that you use at work. Make the list comprehensive. Study it. Be amazed.

Project 1.2

How Much Electronics Do You Use?

Now take that same pad of paper and pen and make a quick summary of how you spend your time during the day. How much time do you spend using each of the electronic products that you listed above? Some of them are only momentarily used, but others you use over longer periods of time, particularly computers, smartphones, and TV sets. Just take a quick estimate of the time you use each one during a typical day. Again, add up the hours. Be amazed.

Project 1.3

Decisions, Decisions

Make the following decision. If you had to give up all of your electronic products but one, which one would you keep? TV, cell phone, remote control, music player? Only one. This will say a lot about who you are.

This page intentionally left blank

Chapter 2

Electronic Concepts: More Interesting Than You Think

Some Basic Stuff You Really Need to Know

In this Chapter:

- Electrons and current flow.
- DC and AC voltage.
- Voltage sources, batteries, the power line, and power supplies.
- Signal sources.
- Magnetism and induction.

INTRODUCTION

You do not have to be an engineer, physicist, or mathematical genius to learn electronics. But there are a few basic things that you need to know so you can understand what is going on in the circuits and equipment. For example, you need to know that electrons are what make up current flow, and voltage is what makes the current flow. You also need to know about magnetic fields and electric fields, where they come from, what they do, and how magnetism and current flow are related. With that background you are ready to learn about the various components and circuits. This chapter takes care of those basics.

ELECTRICITY AND ELECTRONICS

Electricity is a type of energy produced by a charge that is either fixed (static) or moving (dynamic). The main source of electricity is the electron, a subatomic particle that has a negative charge. When electrons are stored or moved, electricity is produced.

Electricity is one of our prime sources of energy. It is used for lighting, heating, and operating appliances of all sorts. It powers motors to produce mechanical energy, and it powers our electronic equipment. The uses of electricity are virtually infinite.

Electronics Explained. http://dx.doi.org/10.1016/B978-0-12-811641-8.00002-3
Copyright © 2018 Elsevier Inc. All rights reserved.

Electronics is the field of applied science that uses components such as resistors, capacitors, diodes, transistors, and integrated circuits (ICs) to control and process electricity. These components are used to create circuits that convert, modify, vary, translate, or otherwise manipulate electrical charges to perform useful functions. That is the essence of it. Now for the details.

Atoms and Electrons

This section really ought to be called Physics 101, but I do not want to scare you away. Just to make you more comfortable, I have to say that it is not all that complicated or difficult to understand. So even if you avoided physics in high school, just be aware that it is pretty easy. Trust me.

All matter, whether it is a solid, a liquid, or a gas, is made up of tiny entities called atoms. An atom is the smallest possible particle of substances called elements. An element is a chemical substance that cannot be subdivided into smaller, simpler substances. Typical elements are hydrogen, oxygen, gold, silver, copper, carbon, helium, silicon, and sodium. All matter is either an element or composed of several elements. Atoms of two or more elements often combine to form new substances called compounds. For example, water is a compound of hydrogen (H) and oxygen (O) or H_2O. Salt is a compound of sodium (Na) and chlorine (Cl) or NaCl. The smallest possible particle of a compound that exhibits all the characteristics of the substance is called a molecule.

According to the theories of physics, an atom is composed of a nucleus consisting of a core of tightly bound subatomic particles, called protons and neutrons. Protons have a positive charge. Neutrons are neutral, having neither a positive nor a negative charge.

Rotating around the nucleus are electrons. The electrons arrange themselves in multiple orbits in much the same way that the planets orbit around the Sun. The orbits are sometimes referred to as rings or shells. Electrons have a negative charge. Fig. 2.1 illustrates a copper atom. The number of electrons in orbit equals the number of protons in the nucleus. The number of positive charges equals the number of negative charges. Therefore, the atom is balanced electrically.

As far as electronics is concerned, the most important part of an atom is its electrons. Since electrons can be manipulated (i.e., stored or moved), they can be used to produce electricity. This electricity can then, in turn, be processed and controlled to perform a wide range of valuable functions. By applying an external force, the electrons in the outer shells can be stripped off to produce current flow.

Charge, Voltage, and Current

There are two basic types of electricity, static and dynamic. Static electricity is the buildup of a charge between two objects. Dynamic electricity supplies a continuous charge.

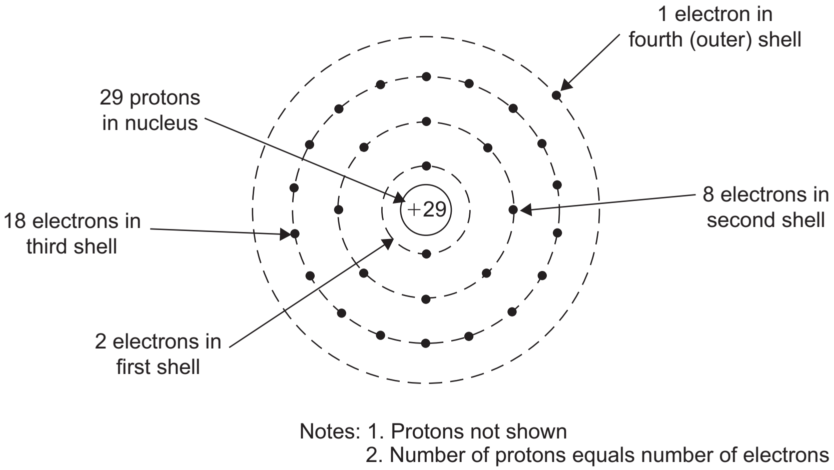

FIGURE 2.1 Physics model of what a copper atom looks like.

Static Electricity

With static electricity, one object has an excess of electrons and the other has a shortage of electrons. The object with an excess of electrons is therefore negative, while the object with a shortage of electrons is positive. A basic law of electricity states that opposite charges attract and like charges repel. The two oppositely charged bodies, one negative and one positive, will have a high physical attraction for one another. An invisible force field called an *electric field* will exist between the two charged objects. The object with a shortage of electrons attracts the object with excess electrons.

The two oppositely charged objects will naturally attract one another. The closer they are, the greater the attraction. This attraction of the charges sets up an electrical force field between the two bodies. This electric field, like a magnetic field, contains energy that can be released to perform useful work.

Whenever there are opposite charges on two objects, we say that an electrical potential or difference of potential exists between them. This difference of potential is called *voltage*. Voltage is an electrical pressure that can cause work to be accomplished.

The charge stored on the two objects is usually called static electricity because it is stationary. However, this charge is usually dissipated in some way. If the two charged objects get too close, or the charge builds up too high, the attractive force becomes so great that the electrons jump the gap between them and create a spark. That is how lightning occurs. The clouds and the earth become oppositely charged, the earth positive, the clouds negative. The charge becomes so great that the electrons leap across the free space gap, ionizing the air, and creating a path through which the electrons travel to neutralize the positive charge on the earth. The result is the massive spark or flash that we call lightning and the roar we call thunder.

In general, static electricity is not very useful. It represents energy stored as a charge, but it quickly dissipates as a spark or a brief flow of electrons between the negatively and positively charged objects when the two charged objects get close enough or when the charge gets too high. We call this an *electrostatic discharge* (ESD). Static electricity, or ESD, is regarded as more of a disadvantage or nuisance than anything else. In fact, it is one of the main causes of damage to electronic components such as transistors and ICs.

Neutralizing a charge between two objects cancels or eliminates the potential. However, if we had a constant charge that could constantly be replenished, we would have a continual source of potential (voltage) and could therefore sustain a flow of electrons. With such a source of electrical energy, useful work can be accomplished.

Dynamic Electricity and Current Flow

Electricity as we know it, whether in electrical power applications or in electronics, is actually the flow of electrons. Electrons flowing from one place to another is known as current flow. And it is voltage that causes the current to flow. When electrons flow, we have energy that can be put to practical use. Electrons flowing in a filament create light. Electrons flowing in a heating coil produce heat. Electrons flowing in a motor produce mechanical energy. The object is to convert the energy from one form to another. In electronics, we precisely control the electrons with special components and circuits to produce a variety of effects.

Fig. 2.2 provides a general block diagram model that shows how all electrical or electronic circuits work. A voltage source produces the energy that causes electrons to flow. This current flow is controlled or processed by a single electronic component, or many components that make up an electronic circuit. The processed electrons flow through a load that consumes or uses the electrical energy to produce a useful output. In an electrical application, the load may be a light, a heating element, or motor, as described above. In an electronic application, the load may be an antenna that radiates the radio signal produced by the electronic circuits. Or the load may be the magnetic head that develops and records digital data on a hard disk or a laser that records a video signal on a DVD.

A couple of key facts about the circuit in Fig. 2.2 are required here. First, there must be a voltage source to cause the current to flow. The electrons flow from the negative terminal of the voltage source around the circuit as they are attracted by the positive terminal. Second, the circuit must always form a complete loop from plus to minus.

Current is so important that we need some way to measure it. We do this by standardizing the number of electrons that move past one point in a

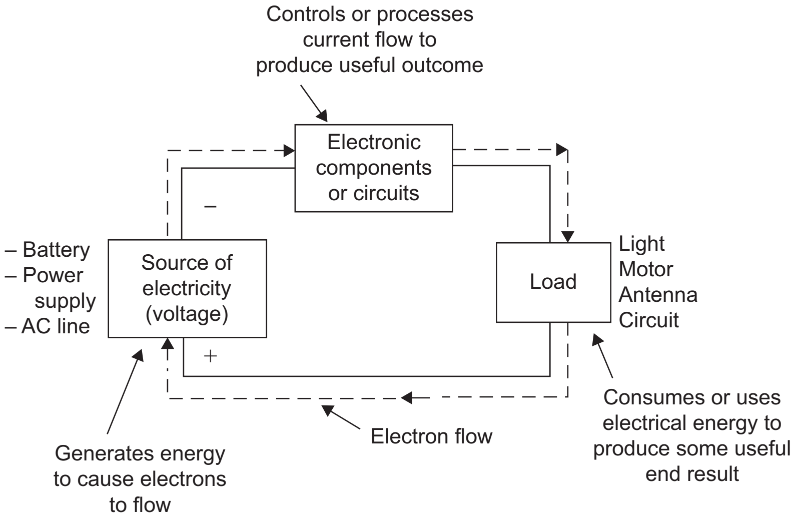
Controls or processes current flow to produce useful outcome
Electronic components or circuits
–
– Battery
– Power supply
– AC line
Source of electricity (voltage)
+
Load
Light
Motor
Antenna
Circuit
Generates energy to cause electrons to flow
Electron flow
Consumes or uses electrical energy to produce some useful end result

FIGURE 2.2 Basic electrical circuit used to describe every electronic circuit.

conductor during a specific period of time. The charge or quantity of electrons (Q) is measured in coulombs (C). One coulomb of charge is equal to 6,242,000,000,000,000,000 or 6.242×10^{18} electrons. If 1 coulomb moves past a point in 1 s, we say that the current is 1 ampere (A). One coulomb per second (1 C/s) equals 1 A.

Direction of Current Flow

The current flowing in electronic components and circuits consists of electrons that are influenced by a charge or voltage source. The early pioneers of electricity did not know that electricity was the movement of electrons. They knew that there was some kind of movement, of a charge or current, but they did not know what it was, or in which direction it flowed. To explain electrical phenomena and perform analyses of components and circuits, they simply assumed a direction of current flow that everyone agreed to and used. Their original assumption was that current flowed from positive to negative, the opposite of the actual direction. This has become known as conventional current flow.

Even after the discovery that real current was electrons moving from negative to positive, scientists and engineers continued to assume a positive-to-negative current flow. And science and engineering professors have continued to teach and support the conventional current flow concept. Most science and engineering textbooks illustrate and teach conventional current flow. Old traditions die hard or maybe never die.

As it turns out, it does not really matter what direction current flows. Any analysis, design, or explanation can be successfully carried out using either electron flow or conventional current. One way is not any better than the other in describing or understanding electrical or electronic phenomena. Why academia has perpetuated the myth of conventional current flow rather than accepting and teaching the truth is not known. It no doubt stems from long-time heritage, custom, and convention.

Conductors, Insulators, and Semiconductors

A conductor is a material that has many electrons easily freed up by an external voltage. Current flows through conductors easily. They have what we call low resistance. Most good conductors are metals such as copper, silver, or aluminum, with loosely bound electrons that can be freed by an external voltage to create current flow. Copper wire is the most commonly used conductor because of its low cost and its ability to be formed into many different shapes and sizes. Most metals are conductors but so is salt water.

Insulators are just the opposite of conductors. These are usually compounds in which the electrons are tightly bound together with the nuclei of the atoms. Even with lots of voltage applied, the electrons are hard to strip away to make current. Insulators keep current from flowing. Some common insulator materials are glass, ceramic, and plastics.

A third category of material is called *semiconductor*. It is a material that can be changed to make it a good conductor or a good insulator or anything in between. Semiconductors are used to make the transistors, diodes, and ICs. The most common semiconductor is silicon. Others are germanium and carbon. Compounds such as gallium arsenide (GaAs), indium phosphide (InP), and silicon germanium (SiGe) are examples. Most ICs or chips are made of silicon.

MAGNETISM

You are probably asking, what does magnetism have to do with anything? As it turns out, it has lots to do with electricity and electronics. As you will see, the two are tied together significantly, meaning you need to know about both to get the big picture. Magnetism is essential to all electrical and electronic phenomena. Without magnetism, there would be no electricity or electronics.

Magnetic Fields

Magnetism is an invisible force that has the power to attract pieces of iron, or alloys of iron such as steel. Iron, steel, and the metals, nickel and cobalt, are called magnetic materials because they can be magnetized to generate a magnetic field, or they support the development and passage of a magnetic field. Magnets are pieces of such materials that exhibit this force. Magnetism has no effect on nonmagnetic materials such as other metals (aluminum, brass, etc.) or on objects made of wood, plastic, or glass.

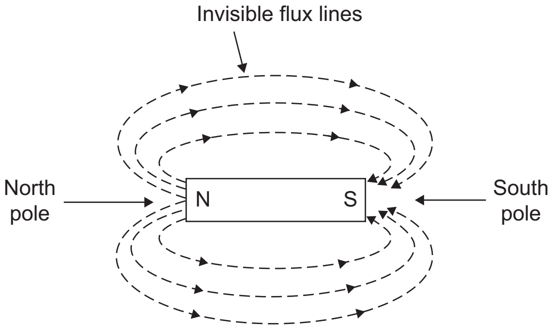

FIGURE 2.3 Permanent bar magnet illustrating the magnetic lines of flux that surround it.

You have no doubt experienced the effects of magnetism yourself. You may have played with a horseshoe magnet as a kid or used bar magnets like the one in Fig. 2.3 in a school lab. And you most likely have used flat magnets to hold paper notes on your refrigerator or file cabinet. These are called permanent magnets because they retain their magnetic force indefinitely.

A magnet is surrounded by an invisible magnetic force field made up of magnetic lines of force or flux lines. These flux lines flow out of one end of the magnet, the north (N) pole, and into the other end, the south (S) pole, as shown in Fig. 2.3. The ends of the magnet are the points of the heaviest concentration of the magnetic field.

The basic law of magnetism follows: *opposite poles attract, and like poles repel.* When you get two permanent magnets together, the north and south poles will attract one another. On the other hand, two north poles will repel one another. Two south poles will also physically push away from one another. Permanent magnets are used in a variety of electrical and electronic equipment. Their most common use is in motors, where the attraction and repulsion of magnets create rotary motion.

Electromagnetism

Magnetism is also produced by electricity. Whenever electrons flow in a conductor, they produce a magnetic field. This effect is called *electromagnetism.* Fig. 2.4A shows how the magnetic lines of force encircle a wire through which current is flowing. Note that the direction of the lines of force depends on the direction of the current flow.

The strength of the magnetic field around a wire depends on the magnitude of the current flowing. High current (composed of many electrons) produces a strong magnetic field. Despite the current amplitude, however, the magnetic field weakens because it spreads out quickly, even at a short distance from the wire.

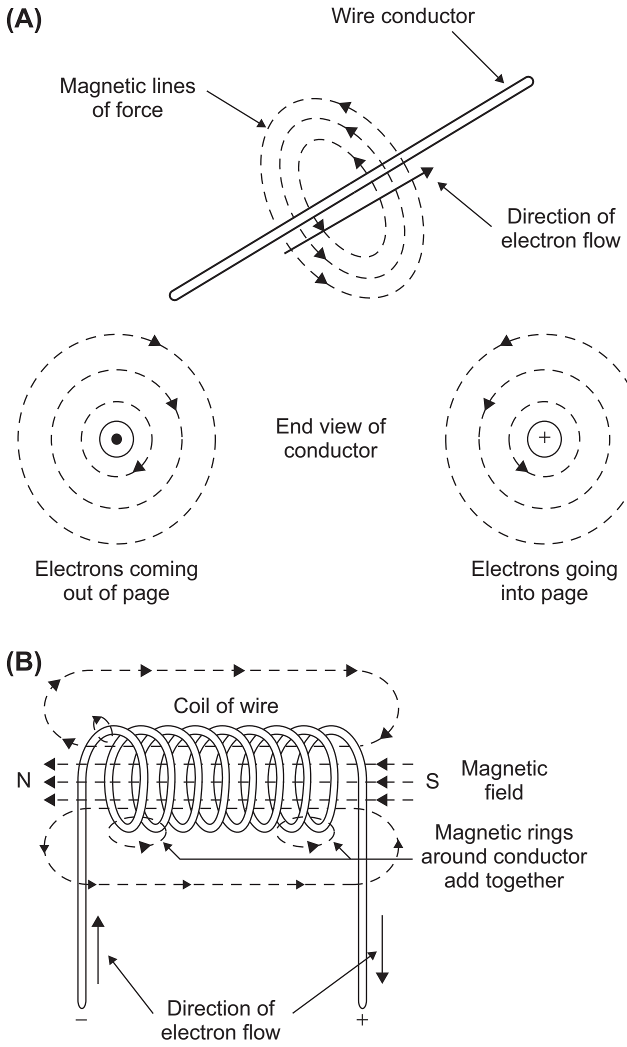
(A)
Wire conductor
Magnetic lines of force
Direction of electron flow
End view of conductor
Electrons coming out of page
Electrons going into page
(B)
Coil of wire
N
S
Magnetic field
Magnetic rings around conductor add together
Direction of electron flow
−
+

FIGURE 2.4 (A) Current in a wire produces a magnetic field around it. (B) Coiling the wire increases the strength of the magnetic field.

If you make a coil out of the wire, as shown in Fig. 2.4B, the lines of force around each turn are added together. The result is that a more powerful, highly concentrated magnetic field is produced. In fact, the coil simulates a bar magnet with north and south poles.

As indicated earlier, the strength of the magnetic field depends on the current amplitude in the wire. High current produces a strong field. The magnetic field is also increased by coiling the wire, which helps concentrate the lines of force. The closer the turns of the wire, the stronger the field. You can also increase

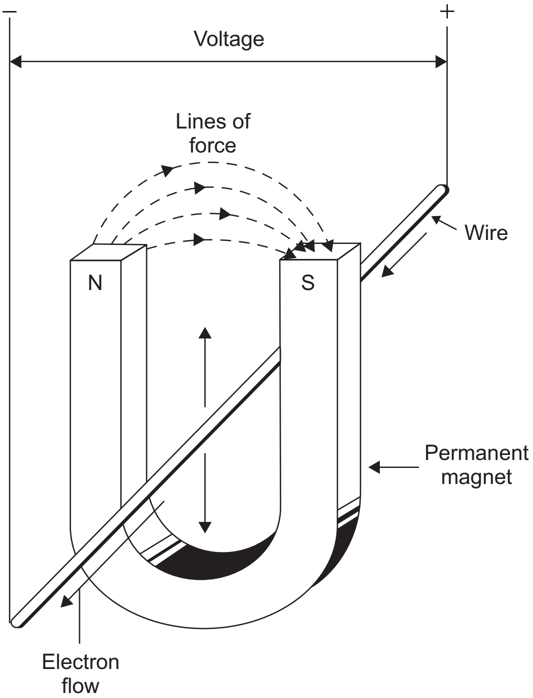

FIGURE 2.5 Relative motion between a magnetic field and a conductor such as a wire causes a voltage to be induced into the wire making it a voltage source.

field strength for a given current level by simply adding more turns to the coil. Strong electromagnets are made by tightly coiling together many turns of fine wire in several layers.

Another way to increase the strength of the magnetic field is to insert a bar or core of magnetic material inside the coil. Because flux lines flow easier in iron or steel than in air, an iron or steel bar inside the coil helps concentrate the lines of force. Most of the flux lines flow through the core because it represents a lower path of resistance than the surrounding air for the lines of force. The result is a very strong field.

Electromagnetic Induction

Another important principle of magnetism is electromagnetic induction. This principle states that whenever there is relative motion between a conductor and a magnetic field, a voltage will be induced into the conductor.

Refer to Fig. 2.5. If the wire is moved up or down so that it cuts through the magnetic lines of force between the magnet poles, voltage will be induced in the wire and cause current to flow. Voltage will also be induced into the wire if the wire is held in a fixed location and the magnetic field is moved so that the flux lines cut across the wire. The wire actually becomes a voltage source. That voltage will cause current to flow. The direction of current flow in the wire depends on both the direction of the magnetic field and the direction of the relative motion between the magnetic field and the wire.

The amount of the induced voltage depends on the number of flux lines cut and the speed of cutting. The greater the number of lines of force, the higher the induced voltage and the greater the current flow. The highest induced voltage occurs when the conductor cuts across the flux lines at a right angle. If the wire moves parallel with the lines of force, no voltage is induced.

The induced voltage can also be increased by coiling the wire. More turns passing through the magnetic field will produce greater induced voltage and current flow. The motion in the magnetic field causes a small voltage to be induced into each turn of wire. All of the voltages will add together to produce a higher voltage.

In the examples above, we have assumed that the magnetic field comes from a permanent magnet. However, it does not really matter where the field comes from. An electromagnetic field works just as well. Electromagnetic induction is the principal method of generating electricity in the world. Most electrical power is produced by giant generators containing large coils of wire that rotate in a magnetic field. The generator or alternator in your car works the same way.

VOLTAGE SOURCES

To make current flow, we must have voltage. That voltage will produce either direct current (DC) or alternating current (AC). DC is electron flow in only one direction. AC is current that flows in one direction and then reverses and flows in the opposite direction.

Direct Current Voltages

DC may be a fixed, steady current, one that varies, or one that switches off and on periodically. A DC voltage source is one that causes direct current to flow one way through the circuit, component, or conductor. It has positively and negatively charged terminals of a fixed polarity. It always causes electrons to flow from negative to positive.

Fig. 2.6A shows a fixed, positive DC voltage source, which is usually either a battery or an electronic power supply. We call it a positive source because the negative terminal of the DC source is connected to a common point called the *ground*. The other terminal is positive with respect to ground. Voltages are always looked at or measured between two points with one point (ground) being the reference. The polarity of the voltage is the polarity of the terminal not connected to ground.

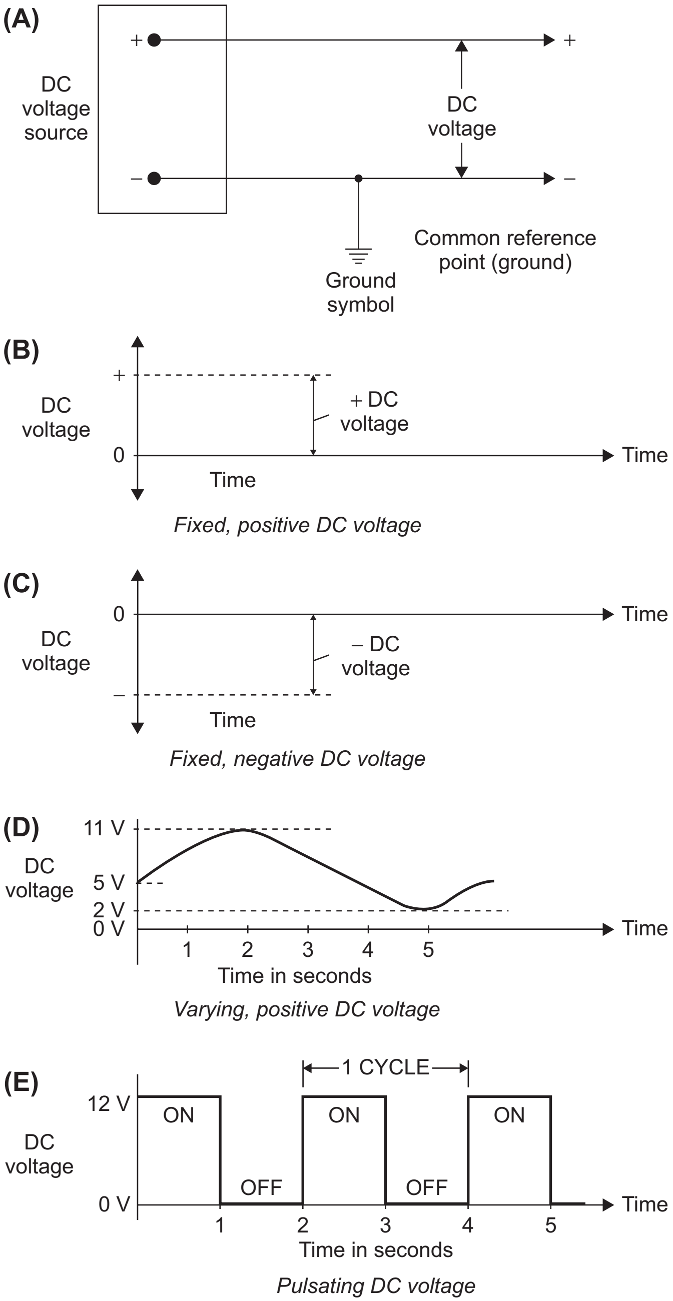
(A)
DC voltage source
+
−
+
−
DC voltage
Common reference point (ground)
Ground symbol
(B)
DC voltage
+
0
+ DC voltage
Time
Time
Fixed, positive DC voltage
(C)
DC voltage
0
−
− DC voltage
Time
Time
Fixed, negative DC voltage
(D)
DC voltage
11 V
5 V
2 V
0 V
1
2
3
4
5
Time
Time in seconds
Varying, positive DC voltage
(E)
1 CYCLE
12 V
DC voltage
0 V
ON
OFF
ON
OFF
ON
1
2
3
4
5
Time
Time in seconds
Pulsating DC voltage

FIGURE 2.6 DC voltages. (A) DC voltage source such as a battery or power supply. One side is usually grounded to serve as a reference. (B) Fixed, continuous positive DC voltage shown over time. (C) Negative DC voltage. (D) Varying positive DC voltage. (E) DC voltage pulses.

The remainder of Fig. 2.6 shows graphs of DC voltage over time. In electronics we normally view DC or AC voltages with respect to time. The positive DC voltage is shown in Fig. 2.6B as a horizontal dashed line.

A negative DC voltage is shown in Fig. 2.6C. The dashed line representing the voltage below the zero line indicates that it is negative. The voltage is negative with respect to the reference point known as ground.

Fig. 2.6D shows a varying positive DC voltage. It begins at +5V, rises upward to +11V, and then drops to 12V over a period of 5s. The left-hand scale indicates the voltage level.

Fig. 2.6E shows a pulsating DC, that is, one that turns off and on. This voltage switches on to +12V, stays on for 1 s, and then switches to zero for 1 s. Then this on–off cycle repeats. One cycle is the time of one on pulse and one off pulse or in this case 2s. Such pulse signals are usually called *binary* or *digital signals*.

Alternating Current Voltages

As its name implies, AC current flows in two directions in a conductor, component, or circuit, but not at the same time. The electrons move in one direction, then reverse and flow in the other direction, both for a short period of time. The direction of flow reverses repeatedly.

The polarity of an AC voltage reverses periodically. During one period, the terminals of the voltage source are + and −, but later reverse to − and +. This reversal continues to repeat as the voltage amplitude varies. Since it is the voltage that causes current to flow, if the polarity of the voltage changes, then naturally the direction of the electron flow will also change periodically. The electrons just flow back and forth in the circuit.

An AC voltage source is usually represented by just a circle with a varying wave inside as illustrated in Fig. 2.7A. It may be the AC outlet in your home, a signal generator, or a type of electronic circuit called an *oscillator*. AC voltages and currents can assume many shapes. The most common electrical and electronic AC voltage is a sine wave, as shown in Fig. 2.7B. The voltage rises gradually from zero to a positive peak, and then drops to zero. Then it reverses polarity, rises to a negative peak, and returns to zero. Each rise and fall is called an alternation. One positive and one negative alternation form a cycle. The cycles repeat over and over. The number of cycles that occur per second is called the frequency.

The AC power-line voltage that comes into your house and appears at the outlets is a sine wave. Its frequency is 60 cycles per second (cps). The term *hertz* (abbreviated Hz) is normally used instead of cps to express frequency. The power-line frequency, therefore, is 60Hz. All radio signals are sine waves, but have very high frequencies.

AC signals can have a variety of other shapes as well. Fig. 2.7C shows a square or rectangular wave. It is made up of pulses of alternating polarity voltages that switch from 16 to 26V. Fig. 2.7D shows a triangular wave, and Fig. 2.7E

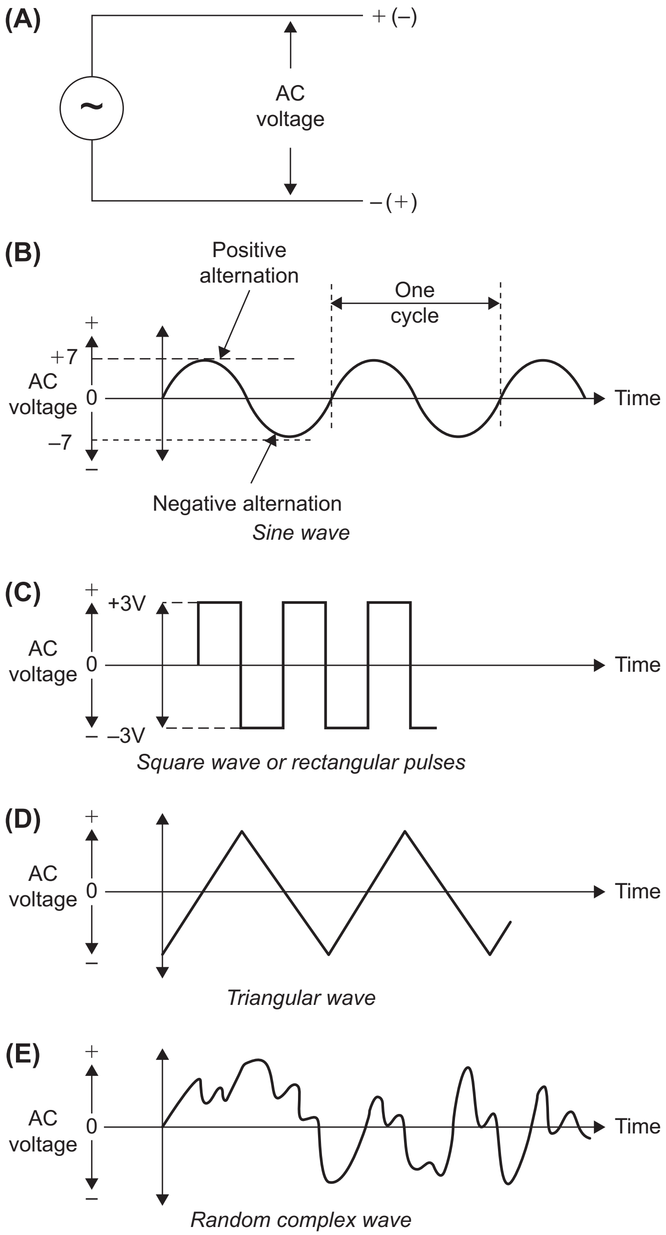
(A)
+ (–)
~
AC voltage
– (+)
(B)
Positive alternation
One cycle
+
+7
AC voltage
0
–7
–
Time
Negative alternation
Sine wave
(C)
+
+3V
AC voltage
0
–
–3V
Time
Square wave or rectangular pulses
(D)
+
AC voltage
0
–
Time
Triangular wave
(E)
+
AC voltage
0
–
Time
Random complex wave

FIGURE 2.7 AC voltages. (A) AC voltage source drawn as a circle with a wave inside. The output polarity changes periodically. (B) Sine wave, the most common AC voltage shape. (C) Rectangular wave or AC pulses. (D) Triangular wave. (E) Random AC wave that may represent voice or video.

shows a complex waveform that might represent voice or video. Regardless of their shape, AC signals have one thing in common: they reverse the polarity and direction of current flow periodically. Their frequencies can range from a fraction of a cycle per second to billions of cycles per second.

Alternating Current Frequency

Frequency is measured and expressed in units called cps or Hz. Hertz is the more common unit. To express higher frequencies, the units kilohertz (kHz), megahertz (MHz), and gigahertz (GHz) are used.

1 kHz = 1000 Hz
1 MHz = 1 million Hz or 1000 kHz
1 GHz = 1 billion Hz, 1 million kHz, or 1000 MHz

Table 2.1 presents a summary of various frequency ranges that you will encounter and the types of signals in those ranges.

Alternating Current Voltage Measurement

To measure or represent the amount of AC in a circuit, we usually state the peak, peak-to-peak, and root mean square (rms) values. The peak value of a wave is simply the voltage measured from the zero line to the maximum positive (or

TABLE 2.1 The Frequency Ranges of the Electromagnetic Spectrum and Their Applications

Frequency Range	Types of Signals
20 Hz to 20 kHz	Audio, sound, voice, and music.
300 kHz to 3 MHz (LF)	Low radio frequencies. AM broadcast.
3 MHz to 30 MHz (HF)	High-frequency radio, shortwaves. Broadcasting, amateur radio, CB.
30 MHz to 300 MHz (VHF)	Very high radio frequencies. Two-way radio, mobile, marine, aircraft, TV and FM broadcast.
300 MHz to 3 GHz (UHF) (anything above 1 GHz is called microwave)	Ultrahigh radio frequency. Two-way radio, TV broadcast, cell phones, GPS satellites, wireless broadband, wireless networks.
3 GHz to 30 GHz (EHF)	Extremely high radio frequencies. Radar, satellites, network radios.
30 GHz to 300 GHz	Superhigh radio frequencies. Radar, satellites, network radios, 5G cell phones.

negative) value. In Fig. 2.7B, the peak value of the sine wave (V_p) is 7V. The peak-to-peak value, normally abbreviated V_{pp}, is the positive peak added to the negative peak. Or, since the positive and negative peaks are identical in a sine wave, the peak-to-peak value is two times the peak value, or in this case 14V.

$$V_{pp} = 2V_p$$

Another way to measure AC voltage is to use an average value known as rms or effective value. The rms value is computed by multiplying the peak value by 0.707 or:

$$V_{rms} = 0.707V_p$$

Thus, the rms value of the sine wave is 0.707×7=4.949V. Most sine wave voltages are measured and stated in terms of rms value. The amount of AC voltage available at a standard AC outlet, for example, is 120V_{rms}. The rms value, as discussed above, is applied only to sine waves. The significance of the rms value is discussed later when we address power.

PRACTICAL DIRECT CURRENT AND ALTERNATING CURRENT VOLTAGE SOURCES

Just what is the source of DC and AC? Such voltages originate from a variety of sources.

Direct Current Voltage Sources

Batteries

The most common source of DC voltage is a battery. A battery is a collection of cells. A cell is the basic unit of voltage generation that creates electricity by chemical action. A battery is made up of two or more cells connected together. Fig. 2.8

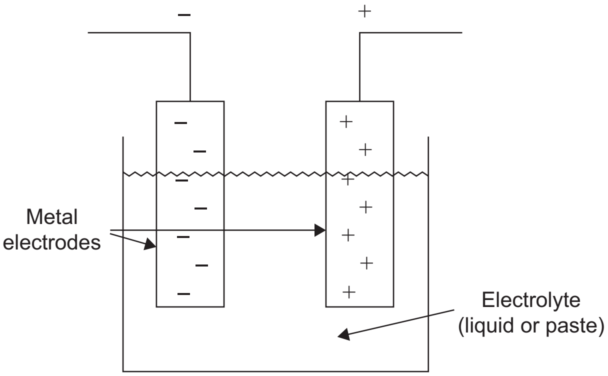

FIGURE 2.8 Basic concept of any battery.

TABLE 2.2 Most Popular Types of Cells/Batteries and Typical Applications

Materials	Voltage	Type	Applications
Alkaline	1.5	Primary	AAA, AA, C, D cells; flashlights; toys; and much more
Mercuric oxide	1.35	Primary	Watches; calculators
Silver oxide	1.6	Primary	Hearing aids; watches
Nickel–cadmium	1.2	Secondary	Tools
Nickel metal hydride	1.2	Secondary	Laptops; cell phones; power tools
Lead acid	2.1	Secondary	Cars; golf carts
Lithium ion	4.1	Secondary	Cell phones; tablets; laptops; electric cars

shows the configuration of a cell. It consists of two metal electrodes immersed in a solution called an electrolyte. The chemical interaction between the metal electrodes and the electrolyte produces a separation of charges. This separation of charges results in an excess of electrons on one electrode (the – terminal of the cell) and a shortage of electrons on the other electrode (the + terminal of the cell). If a conductor is connected externally between the terminals, electrons will flow from – to +.

The amount of voltage produced by the cell depends on the type of electrodes and the electrolyte. Various chemical combinations generate different voltages. Table 2.2 lists the most commonly used cells and the voltage they generate.

The schematic symbol of a cell is illustrated in Fig. 2.9A. To create a battery, two or more cells are connected together. The cells are usually connected in series form, that is, end to end and – to +, as shown in Fig. 2.9B. The cell voltages add up to produce the total battery voltage. For example, six small alkaline cells are connected in series to create a $1.5 \times 6 = 9$-V transistor radio battery. A car battery consists of six lead–acid cells connected in series; each produces 2.1 V for a total battery voltage of $6 \times 2.1 = 12.6$ V. If four nickel metal hydride cells are connected in series, as in Fig. 2.9B, the total battery voltage is 4.8 V.

The simplified schematic symbol for a battery is shown in Fig. 2.9C. Instead of drawing all of the individual cells, the simplified battery symbol is used.

The cells listed in Table 2.2 can be separated into two categories, primary and secondary. *Primary cells* cannot be recharged. As the chemical action

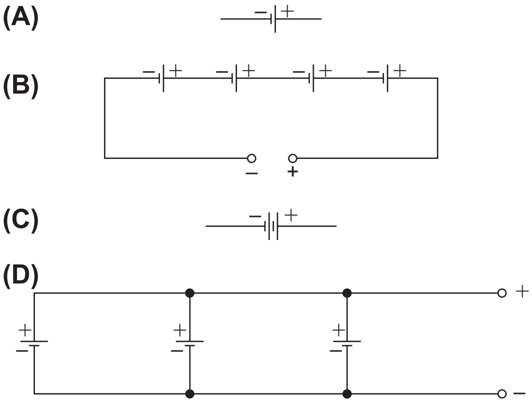

FIGURE 2.9 Battery and cell symbols used in circuit diagrams. (A) Cell. (B) Four cells in series making a battery. Cell voltage add-up. (C) Symbol for battery of any number of cells. (D) Cells in parallel to increase current capability.

produces voltage, the electrodes and electrolyte are actually used up. At some point, no further chemical action is produced, and no voltage is generated. The cell must be discarded and replaced.

Secondary cells, on the other hand, can be recharged. In these cells, the chemical action can be reversed by connecting an external DC voltage to the battery. This is called *charging*. Charging forces electrons into the battery, producing a reverse chemical action that rejuvenates the electrodes and the electrolyte. The cell can be discharged and charged repeatedly, giving it a very long life.

The amount of current that can be drawn from a cell depends on its size and the quantity of the materials used. Large electrodes and electrolytes can produce more current, but the voltage stays the same; the type of material, not its volume, determines voltage.

For example, a large D-size flashlight cell and a tiny AAA penlight cell are both made of alkaline material so their output voltages are both 1.5V. The D-cell, however, can produce more current and will have a longer life because it is larger.

A greater current capacity can be obtained by simply connecting two or more cells in parallel. In a parallel connection, the cells are wired across one another, all + leads connected together, and all – leads connected together, as shown in Fig. 2.9D. The output voltage remains the same, but the current output can be higher.

Solar Cells

A special type of voltage-producing cell is the so-called *solar cell*. Also called a *photovoltaic* cell, a solar cell converts light into electrical energy.

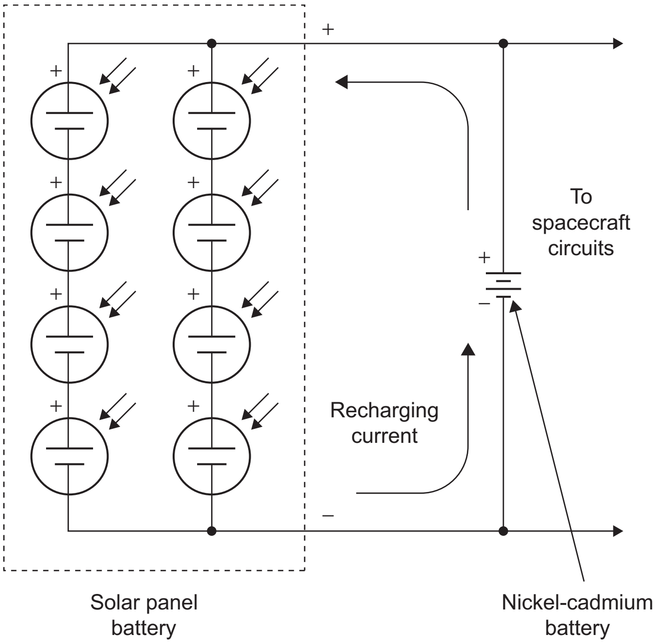

FIGURE 2.10 Panel comprised of solar cells in series and parallel make up DC source that charges the battery in a spacecraft.

When the sun or artificial light shines on the cell surface, a voltage is generated. The most commonly used material is silicon, and a typical cell generates about 0.5 V. Other materials, such as cadmium sulfide and gallium arsenide, are also used.

You have probably seen solar cells used in handheld calculators. Big arrays of solar cells are now being adopted as an energy source for homes and businesses. Another important application is spacecraft. Virtually every satellite, deep-space probe or other space vehicle, uses solar cells for power. Huge panels of solar cells connected in a complex series–parallel array produce a powerful battery that can operate all electrical and electronic equipment on board. Typically, however, such a solar battery is primarily used to recharge the main spacecraft power source, which may be a nickel–cadmium secondary battery, as shown in Fig. 2.10. The voltage produced by this array is 2 V.

Fuel Cell

A fuel cell is a special type of battery that combines oxygen and hydrogen gas to produce voltage and electron flow. That means you have to have a supply of hydrogen gas, which is usually stored in a bottle or other container. Typically, the oxygen can be extracted from the air, but if not you will need a bottle of oxygen as well. When you combine the two, charges separate and plus (+) and minus (−) terminals are formed. Most fuel cells generate between 0.5V and 0.9V, so you have to put many of them in series to be useful. Usually they are bulky and inconvenient because of the need for the gases. But as long as oxygen and hydrogen are supplied, the fuel cell produces the voltage. Heat and water are by-products of the chemical process, making them essentially impractical for many applications.

Power Supplies

Electronic circuits require a source of DC for proper operation. A great many electronic products are still powered by batteries because they are small and portable, such as cell phones, iPods, and the like. Other electronic equipment is designed to operate from the AC power line. Such equipment contains an electronic power supply that converts AC into the required DC voltages. Computers, TV sets, and other nonportable electronic equipment have an AC power supply.

Alternating Current Voltage Sources

Generators

The AC that appears at the outlets of our homes and offices is produced by huge electromechanical generators owned and operated by an electrical utility company. A source of mechanical power—such as water, or a steam turbine powered by oil, coal, or nuclear fuel—operates the generator. The turbine converts steam into mechanical energy to turn the generator. The generator rotates coils in a strong magnetic field to produce a voltage. The speed of rotation of the coils determines the frequency. That voltage is a 60-Hz sine wave.

The big wind generators popping up across the landscape also produce AC sine waves, but because the rotational speed depends on the wind speed, the AC frequency will vary, which is unacceptable. Therefore, the AC voltage they produce is converted to DC, and then back into AC of the correct frequency (60 Hz) to be used or put on the AC power grid.

Another example is the alternator in your car. “Alternator” is just another word for an AC generator. The alternator, driven by the car engine through a belt, produces AC output that is converted to DC by diodes making up a rectifier. The rectifier output is DC to charge the battery.

Inverters

Electronic circuits called inverters are used to produce AC from a DC source. This is common in solar power systems. The solar cells produce DC, which keeps batteries charged. The DC is then converted to AC by the inverter.

In electronic equipment, AC signals are generated by circuits called *oscillators* or *function generators*. The AC signals are sine waves, square (rectangular) waves, or triangular waves, among others.

Analog and Digital Signals

You have no doubt heard the terms "analog" and "digital." These terms distinguish between the two main types of voltages or signals in electronic equipment. Analog signals are voltages that vary smoothly and continuously over time. You saw some in previous figures. Analog signals may be either DC or AC. A continuous DC voltage is analog, as well as a varying DC signal. A sine wave and the random signal in Fig. 2.7E are also analog signals.

Analog signals come from microphones, video cameras, and sensors that detect light level, temperature, and other physical conditions. The most common types of analog signals are voice or music from a microphone and video from a camera.

Digital signals are pulses or on–off signals. The DC pulses in Fig. 2.6E and 2.7 are digital. Most digital signals are binary in nature, meaning they only have two voltage values. In Fig. 2.6E the two values of voltage are 112 and 0V. In Fig. 2.7C the two voltage levels are 13 and 23V.

You will see both analog and digital signals in all sectors of electronics. Today, most signals are digital. Special circuits called analog-to-digital converters (ADCs) and digital-to-analog converters (DACs) are used to translate between the two types of signals.

Fourier Theory and the Frequency Domain

We normally think of electronic signals as voltages that vary over time. Waveforms showing signals are said to be in the time domain. Time is the horizontal axis in a graph. But there is another way to look at signals in the frequency domain where frequency is on the horizontal axis. For example, a sine wave is shown as a single vertical line at its frequency as Fig. 2.11A shows. The height of the line is representative of the voltage level, either peak or rms.

The big question is: how do you show a complex signal like a square wave in the frequency domain? The only way to do this is to use the Fourier theory that says that any complex nonsinusoidal wave is made up of a group of sine waves added together. There is a fundamental sine wave that occurs at the signal frequency plus harmonic sine waves. A harmonic is a sine wave whose frequency is an integer (whole number) multiple of the fundamental sine wave. The second harmonic of a 50-MHz sine wave is a 100-MHz sine wave. The third harmonic is a 150-MHz sine wave and so on. If you add the fundamental sine wave to multiple harmonics of differing amplitudes and phases, you can create any shape of wave.

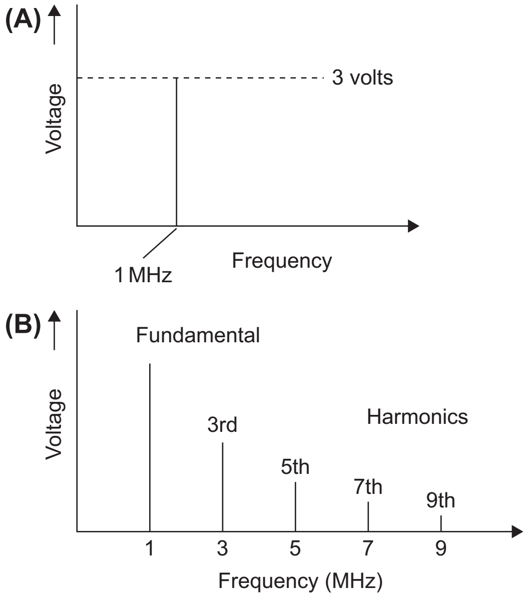

FIGURE 2.11 Frequency-domain views. (A) 1-MHz sine wave. (B) 1-MHz square wave.

For instance, a square wave is made up of a fundamental sine wave and all the odd harmonics. A 1-MHz square wave is made up of a fundamental 1-MHz sine wave and harmonics of 3 MHz, 5 MHz, 7 MHz, and so on. According to Fourier theory, it takes an infinite number of harmonics but in reality only a few (like up to the fifth or seventh) are needed to approximate the square wave. The frequency domain plot of a square wave therefore would look like the one in Fig. 2.11B.

When looking at a time domain view of a signal, try to keep a frequency-domain picture in mind at the same time. You cannot actually see the third or fifth harmonics in the time domain plot but they are buried in there just the same.

Project 2.1

Getting Familiar With a Multimeter

One of the things you are going to need as you work with electronic parts, circuits, and equipment is a multimeter. This is a test instrument that measures voltage, current, and resistance. Some also measure capacitance and other things. These multimeters are usually handheld and battery operated. There are two basic types, analog and digital. Analog meters use a dial to show the measured value. Digital multimeters (DMMs) have a liquid crystal display (LCD) to show the measurement. See Fig. 2.12. The two types work equally well; which one you use is typically a matter of preference for meters versus digital displays. I recommend a digital unit.

Continued

Project 2.1—cont'd

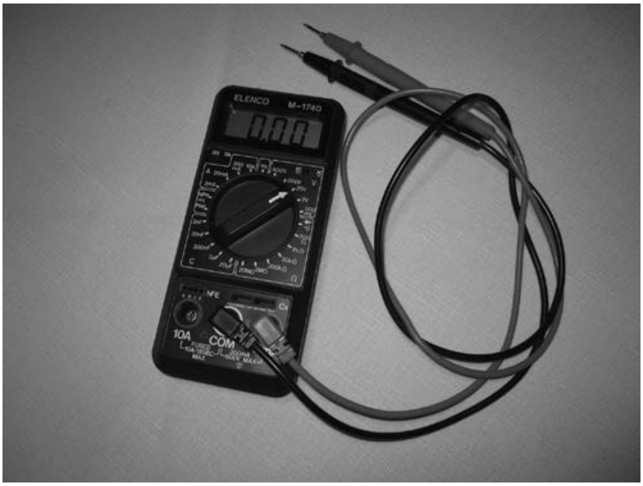

FIGURE 2.12 A low cost digital multimeter that measures DC and AC voltages, current, and resistance.

The good news is that multimeters are not expensive. You can get one for as little as $10, but a better unit may cost you several times that. The cheaper units are kits that you will need to build yourself. I do not recommend a kit if you are just starting out. Later in this book, I will discuss kits in more detail. In any case, you need to buy a multimeter and learn to use it. Plan to buy one as soon as you can.

Several good sources of multimeters are as follows:

Adafruit: www.adafruit.com
Elenco Electronics: www.elenco.com
Fry's: www.frys.com
Jameco: www.jameco.com
Kelvin Electronics: www.kelvin.com
SparkFun: www.sparkfun.com
Velleman: www.velleman.com
Ask for a printed catalog when you inquire.

Project 2.2

Hardware Sources

As an electronic experimenter, you should be familiar with buying parts and equipment. RadioShack stores still stock some parts, but today most hobbyists and experimenters buy via the mail or online. The most popular sources of parts and equipment are All Electronics, Digi-Key, Mouser, and Jameco. And you will discover others. You can go to their websites, but you really should get on their mailing list and get a catalog. Do that as soon as you can. Links to these sources are given below. You can get your multimeter from one of them as well.

Adafruit: www.adafruit.com
All Electronics: www.allelectronics.com
Digi-Key: http://www.digikey.com
Jameco: http://www.jameco.com
Mouser: www.mouser.com
SparkFun: www.sparkfun.com

Project 2.3

Magazines of Interest

As you are learning electronics, you need to be continuously aware of what is going on in the industry and with the technology. Magazines are a good source of that information. One that you should read on a regular basis is *Nuts & Volts*. It is available at some newsstands and bookstores, but if you cannot locate it, go to the website and subscribe. This magazine has lots of articles and columns, including construction projects. It is a great way to stay up to date and learn more. *Nuts & Volts* also publishes *Servo* magazine, which is aimed at people who like to build and play around with robots.

Another recommended magazine that has an embedded controller emphasis is *Circuit Cellar*. This publication contains great educational articles and projects. Get a subscription if you are serious about learning more.

Another interesting magazine is called *Make*—it is not strictly electronics but it usually contains some good electronic projects. It also features other mechanical projects such as robots, among other things. This one is definitely worth a look the next time you are at the newsstand.

Nuts & Volts and *Servo*: www.nutsvolts.com
Circuit Cellar: www.circuitcellar.com
Make: www.makezine.com

Project 2.4

Measuring Voltages

A good first step in using a multimeter is to measure voltages. Batteries are a good place to start. Look around the house and locate some batteries that you commonly use. Take them out of the equipment before you measure them. You will most likely find AA or C cells. Set the multimeter to measure DC according to the meter user's manual and connect the test leads to the battery. The red lead is + and the black lead is −. The LCD on the meter should show the voltage. It should be 1.5V. You can also check batteries in some cordless phones, cell phones, or other devices. You may be able to remove the battery from a laptop. In each case, check to determine the battery technology—lithium ion, alkaline, and so on.

Two key points to remember: first, if you get the test leads reversed, the red to the − terminal of the battery and the black to the + terminal, the meter will read a negative value. That just means that the leads are reversed, nothing more. Second, you cannot typically judge the condition of the battery by measuring its voltage. With no load on the battery, no current is drawn from it so you are just measuring the characteristic voltage. If you connect the battery to something, current will flow and its voltage will decrease slightly. If it decreases a lot, the battery is going bad. If the cell or battery voltage is several tenths of a volt lower than the characteristic value given in Table 2.2, the battery is probably bad.

Continued

Project 2.4—cont’d

You could also go out to your car and measure its battery voltage. Be careful. You cannot get shocked but you could short something. When the engine is off, the battery will probably register about 12V. With the engine running, the battery voltage will be a bit higher, as much as 15V or so. The generator is charging the battery, making its voltage a bit higher.

One last thing: You can use the multimeter to measure the AC line voltage at an outlet. Only do this if you have probes that will fit into the slots on the outlet.

And be supercareful as you can kill yourself. Be paranoid, even.

Set the multimeter for AC voltage at an appropriate level of at least 120V, according to the multimeter manual. Place the probe tips into the AC outlet and read the voltage.

Do not touch any metal with the probe tips!!

Your value will depend on things like how far you are from the power station, what appliances and other items are on, and time of day. In any case, the voltage range is about 95–130V. It is rarely lower than 110 or higher than 125V. The typical value is 120V, or what is called *nominal*. I measured mine at 117V.

Project 2.5

Build a Simple Solar Electrical System

Most electronic equipment is a system, a collection of parts, components, modules, and subassemblies. The system receives electrical or electronic inputs, processes them, and generates useful new outputs. An example is a solar power system similar to one that may be used in some homes.

Refer to Fig. 2.13. A solar panel battery receives energy from the Sun to generate a DC voltage. This voltage is used to maintain a charge on a battery. A charge controller keeps the battery from overcharging and develops the output voltage to the inverter. The battery then powers a DC to AC inverter. The inverter generates a 60 Hz AC voltage at the common line voltage of 120V. Any standard load such as a lamp or radio can be plugged in to operate from the inverter. In this project, you will build and operate such a system.

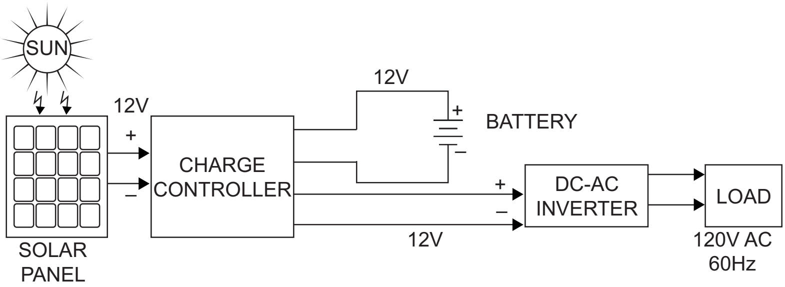

FIGURE 2.13 A simple solar power system.

Project 2.5—cont'd

- First acquire the parts. These are listed in Appendix B parts list. I got mine from All Electronics and the inverter came from a local Best Buy. There are other sources.
- Before you build the system, take the solar panel outside in the sun and measure its output voltage with your DMM. In direct sunlight a 12V panel has a no-load, open circuit output of between 16 and 20V. When you connect a load, the output will drop to its approximate nominal 12 V value.
- Assemble the parts as shown in Fig. 2.13. Follow the wiring directions given for the charge controller. This device manages the charging of the battery to prevent overcharging and damage and conditions the battery output to the inverter. Be sure to observe the polarity of each connection! The inverter usually comes with a 12V lighter plug like those used in cars. You may wish to use clip leads to make the connections faster and easier. Fig. 2.14 shows a photo of a typical system. Clip leads were used on some of the connections.

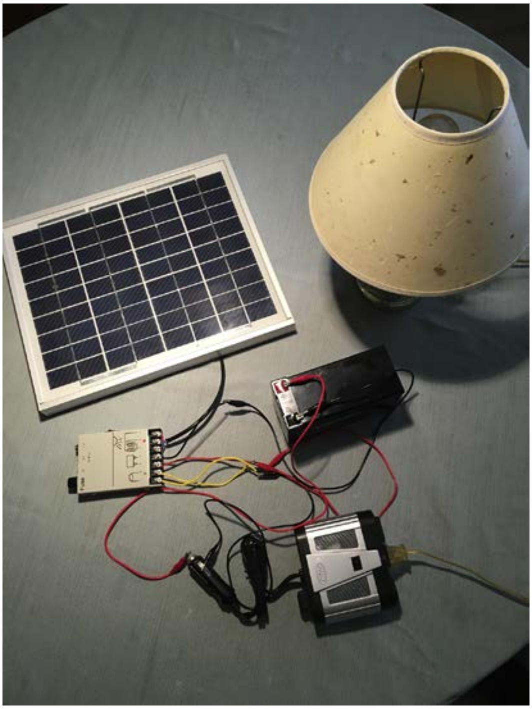

FIGURE 2.14 An experimental solar power system. The solar panel feeds the charge controller below the panel. The charge controller connects to and charges the black battery on the right. The other charge controller output connects to the inverter below. The lamp plugs into the inverter. Clip leads are used to make the temporary connections.

- Connect a load to the inverter. I used a small table lamp and plugged it into one of the inverter's normal AC outlets. Be sure the wattage of the lamp does not exceed the wattage capacity of the inverter.

Continued

Project 2.5—cont'd

- The battery should have an initial charge on it, but check it with your DMM to see that its output is about 12V or more.
- Take the system outside and let the sun charge the battery. The charge controller may have an off–on switch so turn it on if it does. Let the battery charge for about an hour.
- With all but the load turned on, use the DMM to measure the charge controller inputs and outputs to see if all are present. They should all be about 12V or so.
- Test your load. Turn on the lamp. It should work as normal.
- Use your DMM to measure the AC output voltage at the inverter outlet. Set your DMM to the AC mode and be very careful as the 120V output can severely shock you. You should measure about 120V.
- Turn off everything and disconnect the battery.

Chapter 3

The Systems Versus Components View of Electronics

A Fresh New Way to Learn About Electronics

In this Chapter:

- Systems defined.
- Examples of systems.
- Introduction to basic electronic components.

INTRODUCTION

When learning about electronics in the past, you started with basics of physics and math, and then went on to learn electrical theories. Next came instruction in how electronic components work, and then you combined them into circuits. Then you put the circuits together into specific pieces of equipment. Finally, knowing the equipment, you could understand the application, such as TV or computers, audio or home system control, or whatever. This approach to learning is called the bottom-up approach. You learned the details and added them all together to get to what you really wanted to know about—the application and its technology.

That is still the way you learn electronics in college today, and it is the approach taken by most other electronics books you may have seen. This book does not take that old legacy path. Instead, it skips most of the early background and theory, goes right for details on the circuits, and then builds to the applications. It is a much more interesting way to learn electronics. And it is also far more relevant to how electronic equipment is implemented today. We call this way of learning the "systems view" of electronics.

WHY SYSTEMS AND NOT CIRCUITS?

What is the advantage of this approach? First, thanks to modern semiconductor technology, most electronic components and circuits are inside chips or integrated circuits (ICs). You cannot see them, and you cannot access, test, or

Electronics Explained. http://dx.doi.org/10.1016/B978-0-12-811641-8.00003-5

Copyright © 2018 Elsevier Inc. All rights reserved.

replace them. About all you can do is to learn what they do and then work with the various input and output signals. Since you do not have access to the components and circuits, there is some argument as to why you need to learn all that circuit theory detail. If you are an engineer designing ICs, you definitely need to know this. Otherwise, it is just a waste of time unless you want to know just for the heck of it. Some of it is nice to know, of course, but in most cases it is not necessary. What you really need to know is what the circuit is, what it does, what its characteristics are, and what the inputs and outputs should look like.

A good example is the power supply in a PC. If it goes bad, you just replace it with a $30 unit that you can buy online. Troubleshooting the power supply would take many hours and require some expensive test equipment and access to spare parts. The whole repair could take hours and at a typical rate of, say, $50/h, or more, it could cost hundreds of dollars and take weeks to get the new parts. A quick replacement with the $30 unit takes only about a half hour at most. So for less time and money, you get your PC working faster. Not a bad deal at all.

A cell phone is the same. It takes hours of troubleshooting and very expensive test equipment to troubleshoot and repair one. A typical repair would cost hundreds of dollars. Consequently, it is rarely done. It is sometimes cheaper to get a new cell phone, and it will be a better one at that. And it may even be free if you renew your service contract. That is electronics today.

The ability to analyze circuit operation by tracing electron flow through the various components used to be necessary. That made it necessary to know about all the components in detail and how they all come together to make a circuit. But with the systems view, you rarely need this knowledge unless you are an engineer designing the circuits or the equipment. In most cases you are not, so why invest the time?

The systems view is a better approach, as you do learn about the circuits and the equipment at a higher level. Later, if you want or need to, you can delve deeper into the circuit and component details, just for fun or to satisfy your curiosity. We will not do that in this book. But you will get the system view and big picture that will serve you well in learning electronics.

The systems view is more important today because ICs have made it possible to largely ignore the internal details. The ICs and systems on a chip (SoC) have resulted in the throwaway electronics society in which we live. It is no longer possible or economical to repair most electronic equipment, thanks to ICs. The cost of labor, the cost of test instruments to troubleshoot equipment, and the difficulty of replacing ICs have made it more practical to just discard a bad piece of electronic equipment rather than repair it. It costs less and you get a newer and better replacement product as a result. Not a bad deal.

Yes, it is still possible to repair some equipment. For example, big expensive equipment such as a magnetic resonance imaging (MRI) machine is usually repairable; however, it is not done by replacing bad components, but by

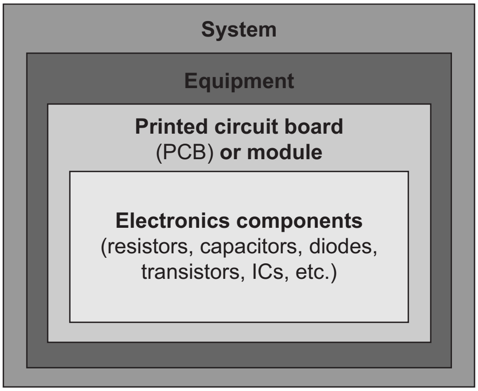
System
Equipment
Printed circuit board (PCB) or module
Electronics components (resistors, capacitors, diodes, transistors, ICs, etc.)

FIGURE 3.1 Hierarchy of electronics from systems to components.

replacing defective modules, subassemblies, or printed circuit boards (PCBs). You do not have to know how to analyze circuits to do that. You just troubleshoot by observing signal flow, measuring inputs and outputs, and replacing any module or subassembly that does not work. You get the equipment working sooner and at lower cost.

A good way to picture the systems view is shown in Fig. 3.1. It all begins with the components that make up the circuits that in turn make up the modules and subassemblies, usually on PCBs. Those are put together to form pieces of equipment. Finally, multiple pieces of equipment may end up being interconnected to form some kind of system. A few system examples will show you the idea.

A personal computer (PC) is a great example of a system (see Fig. 3.2). It is made up of modules including a motherboard with its processor, memory ICs, and input/output (I/O) interfaces. There is a power supply module, a hard disk drive, a CD/DVD drive, and perhaps a video card. Then the system is completed when it is connected to the video monitor, keyboard, mouse, printer, and Internet modem.

A consumer home entertainment system is made up of a big-screen HDTV set, a cable or satellite box, a DVD/Blu-ray player, and an audio amplifier/receiver with speakers. Other equipment may be involved such as an over-the-top Internet TV connection with a Wi-Fi unit from Apple TV or Roku. The TV itself or any of the other components are systems in their own right.

An iPod or MP3 player is a simple system. Shown in Fig. 3.3, it consists of a memory, either a solid-state flash memory that stores the music, an audio codec IC that does the digital music decompression, digital-to-analog conversion and analog-to-digital conversion for storing music, a microcontroller that controls the whole thing including the selection controls, and the LCD readout. A pair of audio amplifiers and a headset complete the system. A power management system charges the battery and doles out DC to power the various ICs.

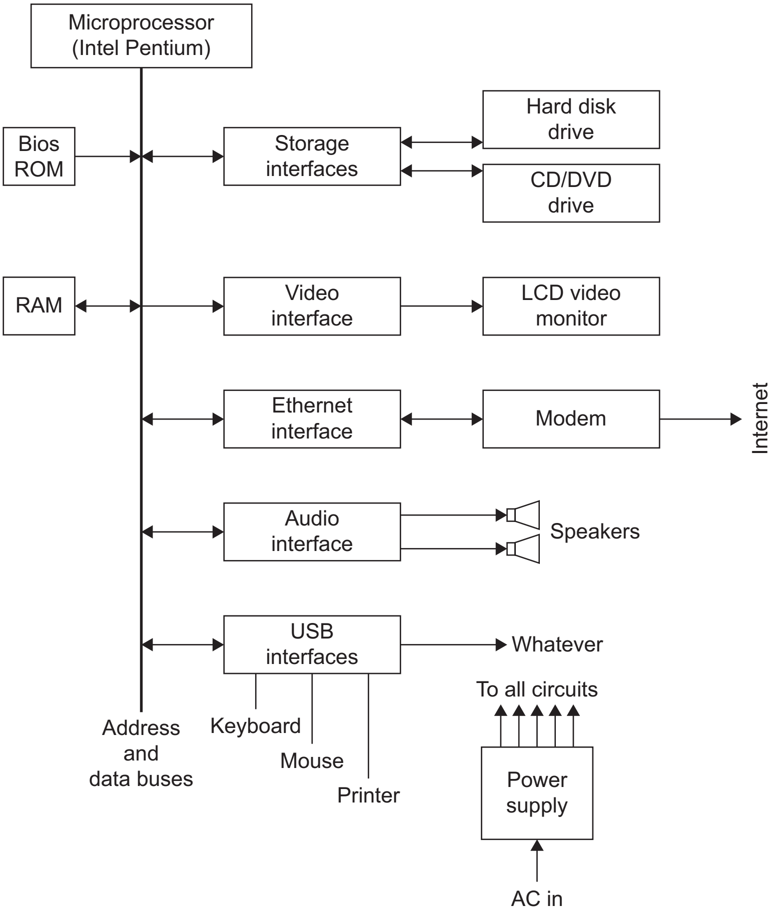
Microprocessor
(Intel Pentium)
Bios
ROM
Storage
interfaces
Hard disk
drive
CD/DVD
drive
RAM
Video
interface
LCD video
monitor
Ethernet
interface
Modem
Internet
Audio
interface
Speakers
USB
interfaces
Whatever
To all circuits
Address
and
data buses
Keyboard
Mouse
Printer
Power
supply
AC in

FIGURE 3.2 System block diagram of typical personal computer.

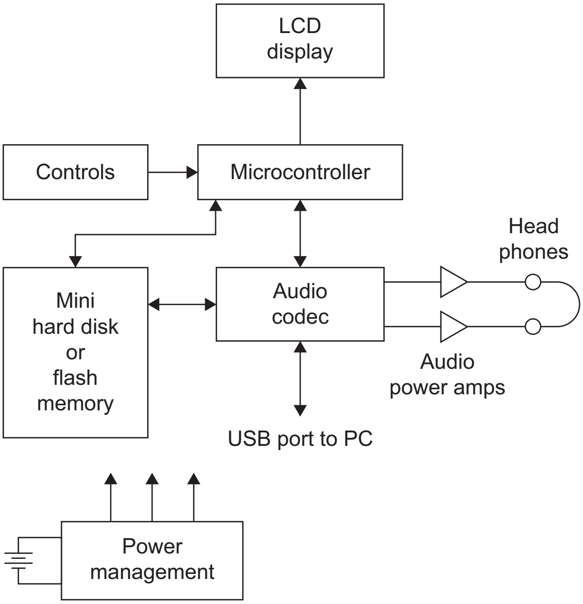
LCD
display
Controls
Microcontroller
Head
phones
Mini
hard disk
or
flash
memory
Audio
codec
Audio
power amps
USB port to PC
Power
management

FIGURE 3.3 System block diagram of iPod or MP3 music player.

An MRI machine is a whole system. It is made up of the central bed surrounded by a superconducting magnet and radio frequency–sensing coils, high-power radio frequency (RF) amplifiers and detectors, the video imaging section, and a computer that runs the whole thing.

Anyway, you get the picture. There are big systems, little systems, and lots of in-between systems. Their main characteristic today is that you cannot get at the individual components and circuits, but there are exceptions.

In this book we look mainly at the circuits, equipment, and systems, not the components. We will give you a quickie introduction to the main components but move on to the circuits and systems.

ELECTRONIC COMPONENTS

Most systems today are collections of ICs wired together on PCBs and packaged in a housing of some sort. The MP3 player/iPod in Fig. 3.3 is a great example. Each block in that diagram is an IC. In the past, systems were made with PCBs and other assemblies, but the circuits were made with discrete components, individual resistors, capacitors, diodes, transistors, and even some ICs. You will still see these here and there in modern equipment; just remember that most circuits and equipment are just a batch of ICs.

Just so you will not be ignorant of components, here is a nutshell introduction that should serve you well when learning electronics the system way. If you want to dig deeper, you can find many other books that will make your eyes glaze over with detail.

Switches

There is no simpler electrical component than a switch. It is just two (sometimes more) metal contacts and a mechanical device to open and close them. In one position the switch contacts are open and not touching so that no current will flow through them; the switch is “off.” Moving the switch lever, the contacts are made to touch making a path for current to flow; the switch is said to be “on.”

Fig. 3.4 shows a switch in a circuit diagram called a *schematic*. This shows the switch connected to a battery and a light-emitting diode (LED) in a simple flashlight. The other component is a resistor to limit the current.

Resistors

Resistors are one of the most common components in electronics. These devices offer opposition to current flow; they limit or set the level of current in a circuit. The opposition to current flow is measured in ohms. The higher the ohms value, the more resistance it offers to electron flow.

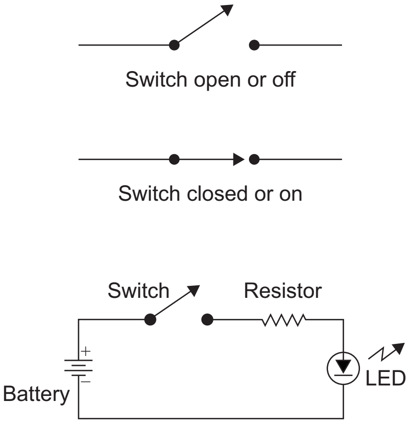

FIGURE 3.4 Switch and how it is used.

Ohm's Law

There is one key electrical relationship you should know, called Ohm's law. It is simple and easy to understand. It says the current in a circuit (I) is directly proportional to the voltage (V) and inversely to the resistance (R). It is usually expressed as a simple formula:

$$I = V/R$$

If you increase the voltage in a given resistance, the current increases. That makes sense since the voltage charge simply attracts and repels more electrons. Increasing the resistance makes the current decrease. Since resistance opposes current flow, more resistance with the same voltage will make the current decrease.

Keep this simple relationship in mind, as it is the heart of every electrical and electronic application.

Today most resistors are made along with all the other circuitries in an IC. But discrete component resistors are also still fairly widely used. The most common ones are called chip resistors and look something like the one shown in Fig. 3.5. They are small and you can just make them out when looking at a PCB. Older but still used resistors are cylindrical like the one shown in Fig. 3.5. The schematic symbol used to represent a resistor in a schematic diagram is shown in Fig. 3.5. Color-coded resistors use a set of color bands with a special code to indicate their value (see Fig. 3.6).

An example of a common resistor circuit is the voltage divider shown in Fig. 3.7. It is used to attenuate a voltage or signal or make it lower. The two resistors of the voltage divider are connected in series. When current flows, a voltage develops across the two resistors. The input voltage is divided between the two resistors according to their value. In any case, the output is taken from one of the resistors and it is lower in value than the input.

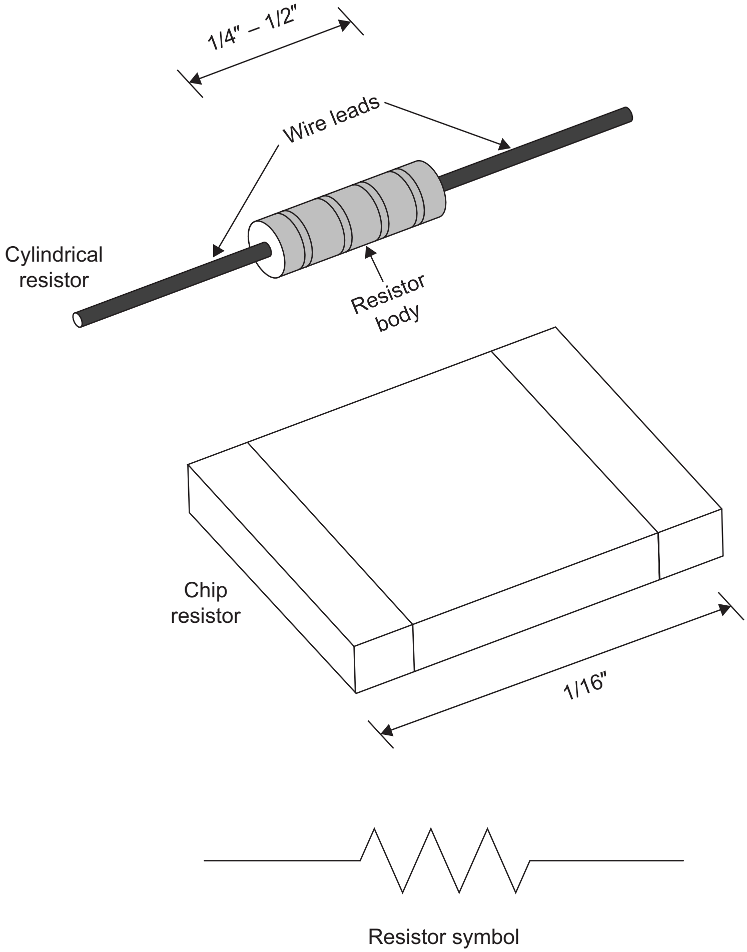

FIGURE 3.5 Types of resistors and schematic symbol.

Capacitors

A capacitor is a device that stores electrical charge. It is made up of two metallic plates separated by an insulator such as plastic or ceramic or even air (see Fig. 3.8). The symbol for a capacitor in schematics is also shown.

The basic function of a capacitor is to be charged or discharged. You charge the capacitor by applying a voltage to it as shown in Fig. 3.9A. The electrons pile up on one plate making it negative and are pulled away from the other plate making it positive. It takes a finite amount of time for the capacitor to charge as determined by the values of capacitance (C), stated in farads (F), and resistance (R). That time is called the time constant (T). In the charged state, the capacitor stores the voltage. The capacitor in its charged state is almost like a battery, as it can supply DC voltage to a circuit or load.

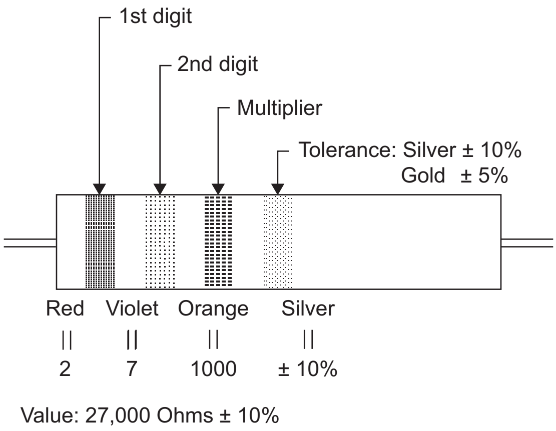
1st digit
2nd digit
Multiplier
Tolerance: Silver ± 10%
Gold ± 5%
Red
Violet
Orange
Silver
||
||
||
||
2
7
1000
± 10%
Value: 27,000 Ohms ± 10%

Resistor color code

Color of band	Value	Multiplier
Black	0	1
Brown	1	10
Red	2	100
Orange	3	1000
Yellow	4	10,000
Green	5	100,000
Blue	6	1,000,000
Violet	7	10,000,000
Gray	8	100,000,000
White	9	1,000,000,000

FIGURE 3.6 Standard resistor color code.

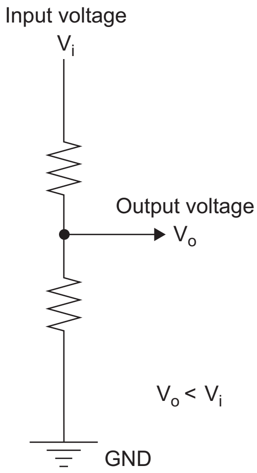
Input voltage
V_i
Output voltage
V_o
$V_o < V_i$
GND

FIGURE 3.7 Voltage divider, the most common resistor circuit.

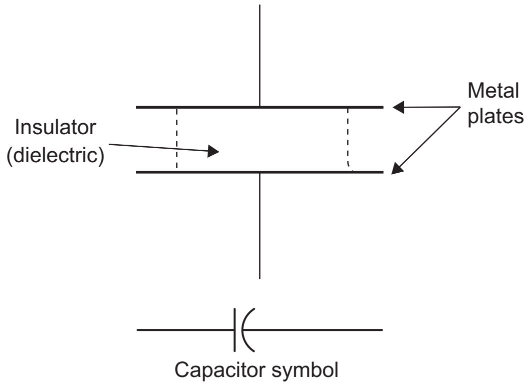
Metal plates
Insulator (dielectric)
Capacitor symbol

FIGURE 3.8 Capacitor construction and schematic symbol.

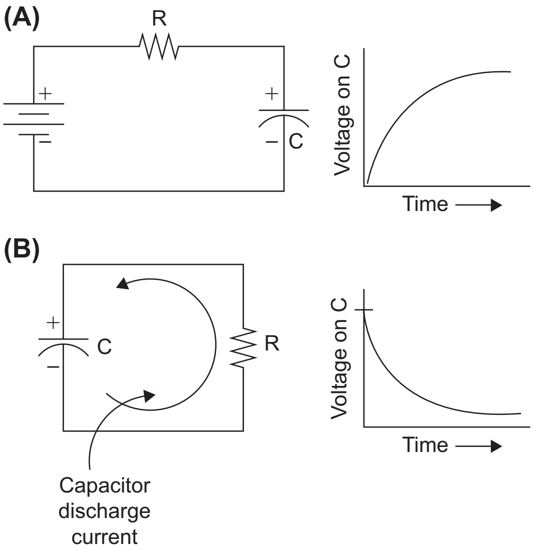
(A)
R
+
−
+
−
C
Voltage on C
Time
(B)
+
−
C
R
Capacitor discharge current
Voltage on C
Time

FIGURE 3.9 (A) Charging capacitor, (B) discharging capacitor.

The energy stored in the capacitor is used by discharging it into a load such as a resistor (see Fig. 3.9B). The electrons then flow from the negative plate through the resistor to the positive plate. Current flows until the charge is dissipated. Again, it takes time for the discharge to occur.

When capacitors are used with AC signals, they offer opposition to current flow called *capacitive reactance*. It is like resistance in that it determines current flow, but it varies with the frequency of the AC and capacitor value.

Capacitors perform all sorts of functions in electronic circuits. They are used for filtering and to pass AC but block DC. Fig. 3.10 shows a low-pass filter

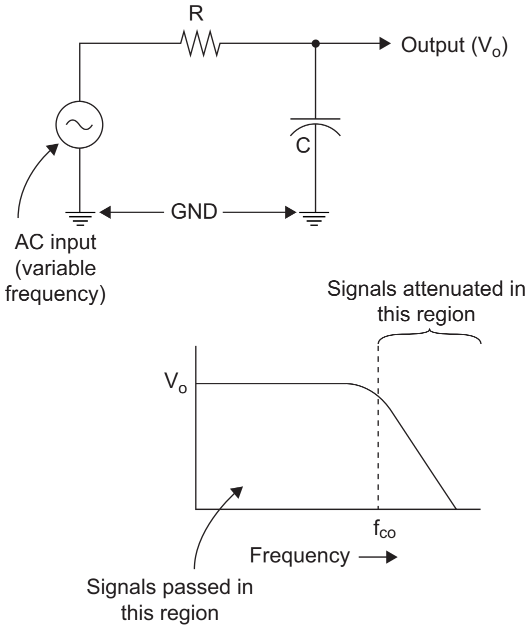
R
Output (V_o)
C
GND
AC input
(variable
frequency)
Signals attenuated in
this region
V_o
f_{co}
Frequency
Signals passed in
this region

FIGURE 3.10 Low-pass filter, a common capacitor use.

comprised of a resistor and a capacitor. This is a frequency-sensitive circuit that passes AC signals below a specific frequency called the *cutoff frequency* (f_{co}) and attenuates the frequencies above the cutoff. Fig. 3.11 shows how a capacitor passes AC but blocks DC.

A Word About Ground

In schematic diagrams like Fig. 3.10, you will see the triangular-shaped symbol called GND. This means ground. Ground is an electronic term that refers to a common connecting point. It may be a common wire, a metal chassis, a copper area on a PCB, or some other conductive part or area. In this figure, it means that the bottom of the AC generator and the bottom of the capacitor are connected together.

Ground is also the common reference point for measuring all voltages.

Inductors

Inductors are mainly just coils of wire. When current flows through them, they produce a magnetic field and induce a voltage in themselves. This is called *self-induction*. The induced voltage has an opposite polarity of the applied voltage. The result is that the induced voltage causes opposition to the current flow.

We often use that characteristic to control current flow in AC circuits. Like the capacitor, the inductor is an opposition to AC called *inductive reactance*. Fig. 3.12 shows the symbol of an inductor as used in schematic diagrams.

Transformers

A transformer is usually comprised of two coils or windings on a common core. The schematic symbol is shown in Fig. 3.13. A voltage applied to one winding called the *primary* causes current to flow and a magnetic field to develop. This magnetic field spreads out and induces voltage in the other winding called the

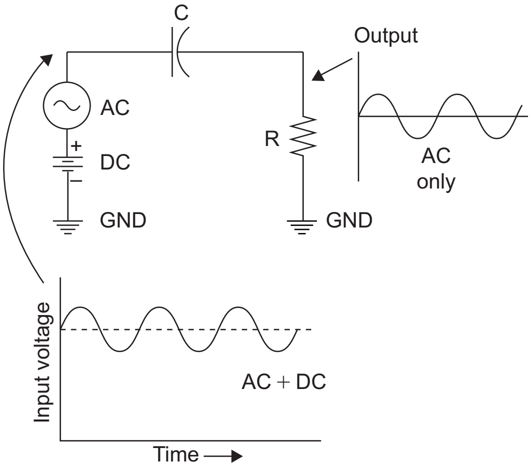
C
Output
AC
+
DC
–
GND
R
GND
AC
only
Input voltage
AC + DC
Time

FIGURE 3.11 How capacitor passes AC sine wave but blocks DC voltage.

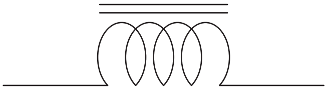

FIGURE 3.12 Schematic symbol for inductor.

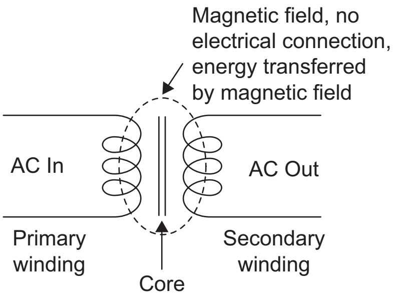
Magnetic field, no electrical connection, energy transferred by magnetic field
AC In
AC Out
Primary winding
Core
Secondary winding

FIGURE 3.13 Schematic symbol for transformer.

secondary, thereby transferring electrical energy for one coil to another by way of the magnetic field. Transformers are used to step up or step down voltage levels of AC signals. They are also used for signal isolation as well as impedance matching.

Diodes

A diode is a semiconductor component that lets current flow through it in one direction but blocks current in the other direction depending on the polarity of the voltage applied to it. It acts like a polarity-sensitive switch. Fig. 3.14 shows a typical diode, the diode schematic symbol, and how it is biased for current or no current.

Diodes are used mainly for rectification, the process of converting AC into DC. Fig. 3.15 shows how a diode converts an AC sine wave into DC pulses. If you put a capacitor across the load resistor, it will charge up to the peak sine voltage and store it. The result is that the output is a near-constant DC value. Most electronic power supplies work like this.

Special diodes are made to emit light (LED), regulate voltage (zener diode), act as a variable capacitor (varactor), or perform as a switch (PIN diode).

Transistors

A transistor is a three-terminal semiconductor device that uses a small input signal to control a much larger output signal. One type of transistor is the bipolar

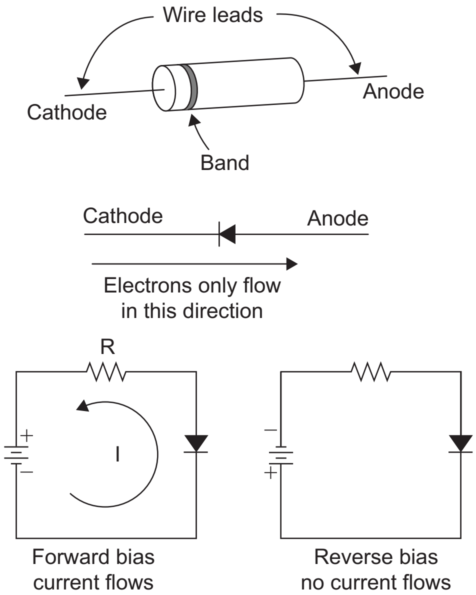

FIGURE 3.14 Diode, its schematic symbol, and how to bias it for conduction or cutoff.

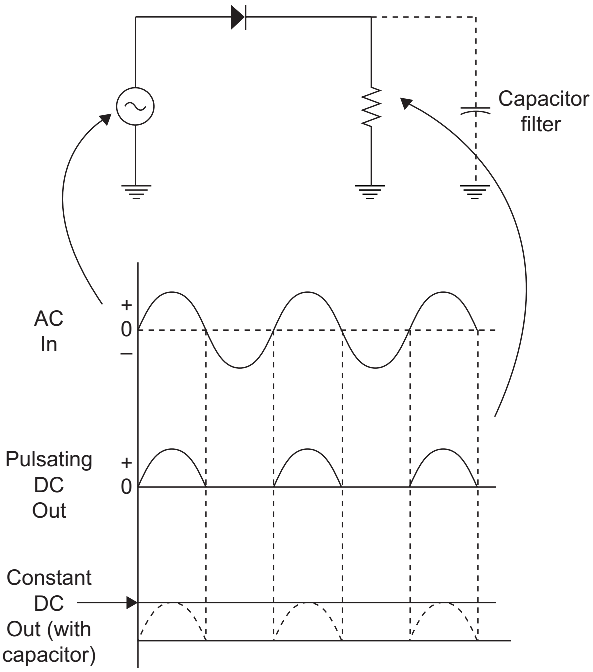
Capacitor
filter
AC
In
+
0
–
Pulsating
DC
Out
+
0
Constant
DC
Out (with
capacitor)

FIGURE 3.15 How a diode rectifies AC into DC.

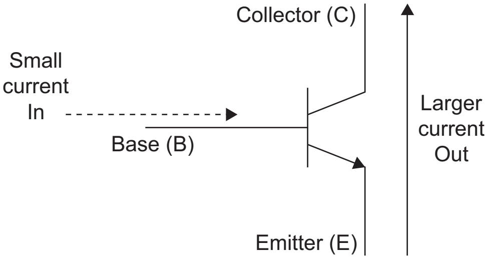
Collector (C)
Small
current
In
Base (B)
Larger
current
Out
Emitter (E)

FIGURE 3.16 Schematic symbol for bipolar junction transistor (BJT) and its inputs and outputs.

junction transistor (BJT) shown in Fig. 3.16. A small current applied to the base element is used to control a larger current flowing from emitter to collector.

Another type of transistor is the metal oxide semiconductor field-effect transistor (MOSFET). It too has three elements, as shown in Fig. 3.17. A small voltage on the gate controls a larger current flowing from source to drain.

Transistors have two basic functions: amplification and switching. In amplification, the small input variation causes larger output current variation. The small input signal is not really made bigger, but the larger output variation has the same shape and characteristics, resulting in the same effect. The transistor simply varies the larger DC current from the supply and shapes it like the input.

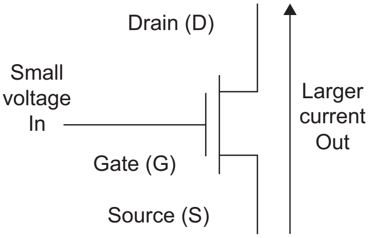
Drain (D)
Small voltage In
Gate (G)
Source (S)
Larger current Out

FIGURE 3.17 Schematic symbol for MOSFET and its input and outputs.

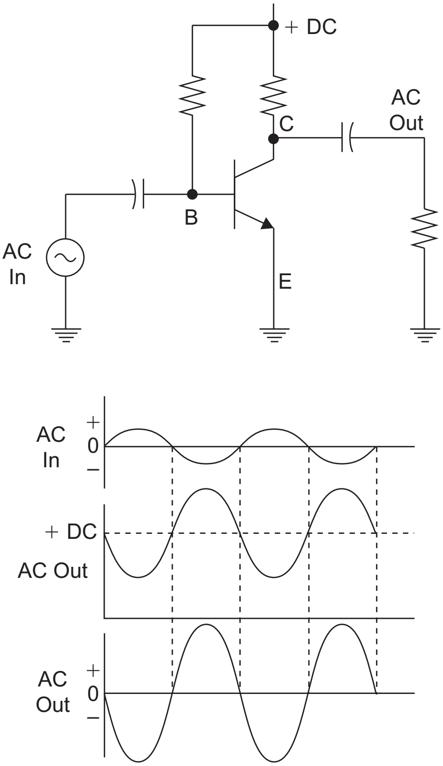
+ DC
AC Out
C
B
AC In
E
AC In
+
0
−
+ DC
AC Out
AC Out
+
0
−

FIGURE 3.18 Simple BJT amplifier and how it works.

Amplifiers

Fig. 3.18 shows a simple BJT amplifier. The base resistor sets the base current turning the transistor on so that current flows from emitter to collector. Its output voltage is a DC value at about half the supply voltage. Now, if a small AC signal is applied through the capacitor as shown, the base current will vary. This will vary the collector current and the voltage across the collector resistor. A small variation in the base current produces a larger variation in the collector current. The output voltage is an enlarged version of the input, except that it is reversed in phase. We say that the output is 180° out of phase with the input, called *phase inversion*. The output is a DC signal, but to convert it to AC we pass it through a capacitor to the load. All amplifiers generally work like this.

Switches

As a switch, the transistor output current is turned off or on by the smaller input current or voltage. A small input signal can switch a larger output current off and on. Switching transistors can switch very fast (in less than nanoseconds). Fig. 3.19 shows a MOSFET switch. If the input voltage on the gate is zero, the MOSFET does not conduct. It acts like an open or off switch. The output then is just the supply voltage (1 V) as seen through the resistor.

Now, if we apply voltage to the gate above a certain threshold level, the MOSFET will conduct. It acts like a closed or on switch so current flows in the resistor. The output is a very small value near zero. Fig. 3.19 also shows what the inputs and outputs look like with a square wave applied.

Integrated Circuits

An IC is also called a chip. It is a complete circuit made on a single chip of silicon. All the components are made at the same time and interconnected to form an amplifier, microcomputer, memory, oscillator, radio transceiver, or whatever the function. The IC is then connected to DC power and it works. All you do is to apply the inputs, and new outputs are produced according to what the IC function is. Most circuits today are ICs.

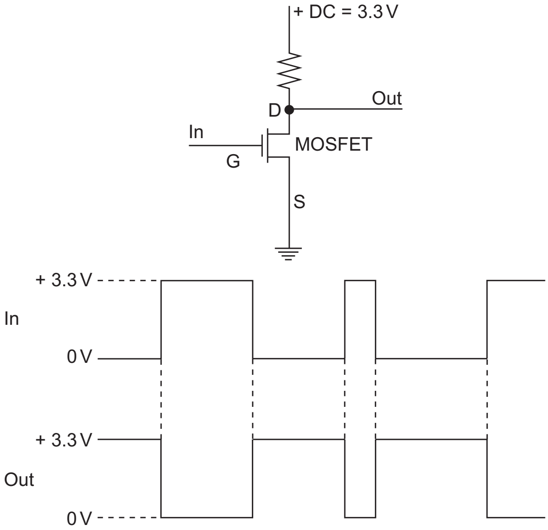

FIGURE 3.19 Simple MOSFET switch and how it works.

MDM9600
QUALCOMM

FIGURE 3.20 Typical integrated circuit today. This one is a dual-satellite TV tuner for set-top boxes. The size is only 7 × 7 mm. The pins are soldered to a printed circuit board. *Courtesy NXP Semiconductor.*

There are all sizes and types of ICs. There are linear circuits for analog signals, digital circuits for binary signals, and mixed signal IC for both analog and digital signals. There are small ICs that are just an amplifier or logic circuit. Then there are even larger circuits like a computer memory or a microprocessor that is the heart of every computer. Finally, there are complete SoC. Each year, it is possible to make the components on a silicon chip smaller so you can put more on a chip economically. That is why we have iPods, cell phones, and other really small devices that in the past would have been huge pieces of equipment. Fig. 3.20 is representative of ICs today.

There are lots of other electronic components but these are the main ones. A few others are introduced in later chapters.

While this book attempts to get you thinking about the big picture—the systems rather than the individual components—we cannot ignore discrete components completely. Well, actually we can, but you may be curious and want to have a bit more involvement with them. If so, the projects to follow will show you how individual parts are used. The circuits presented here are trivial compared with the sophisticated technology you use every day, such as big-screen LCD HDTV sets, laptops, cell phones, and your wireless Internet broadband router. Feel free to skip these if you wish.

Project 3.1

Setting Up a Simple Lab

If you want to play around with some components just to get the feel for how it used to be, you need to set up a simple lab. If you performed the previous projects, you already have a multimeter. Now you will need a way to interconnect parts in circuits and a power source. Appendix B is a complete parts list for the projects in each chapter.

Project 3.1—cont'd

The power source can be four AA cells in a holder. This will provide DC voltage of 6 V to power the circuits. Or you can use a common 9-V battery. A battery clip with leads will allow you to connect the 9-V battery to the circuit.

Next you will need a breadboard. This is a connector socket with multiple holes in it like that shown in Fig. 3.21. You plug the wire leads of the components into the holes to connect them to one another without soldering them.

Then you will need some components such as resistors, capacitors, and transistors. In this project, the goal is to get the battery and the breadboarding socket. The remaining projects for this chapter are circuit experiments. See Appendix B.

You can get the breadboard from a local RadioShack or from a mail order house such as Jameco. There are usually two sizes, large and small. The larger one is recommended. Fig. 3.21 shows a larger one.

Get a battery holder that will take four AA batteries. All Electronics or Jameco are good sources. Some battery holders also have an on/off switch that is useful. The battery holder may have two leads for connecting the battery to the circuit. The 9-V battery clip is shown in Fig. 3.21. You will need to solder some #22 solid copper wires to the battery clip leads as these leads use stranded wire that does not plug into breadboarding socket. Fig. 3.22 shows how. You may need to do the same thing with the AA battery holder wires. See Chapter 14 on how to solder if you need to.

Once you get the parts, put the AA cells into the holder so that you connect the negative end (–) of one battery to the positive end (+) of the next so that the cells will be in series. That way their voltages will add together. This wiring may already be present in the battery holder. The circuit should look like that in Fig. 3.23.

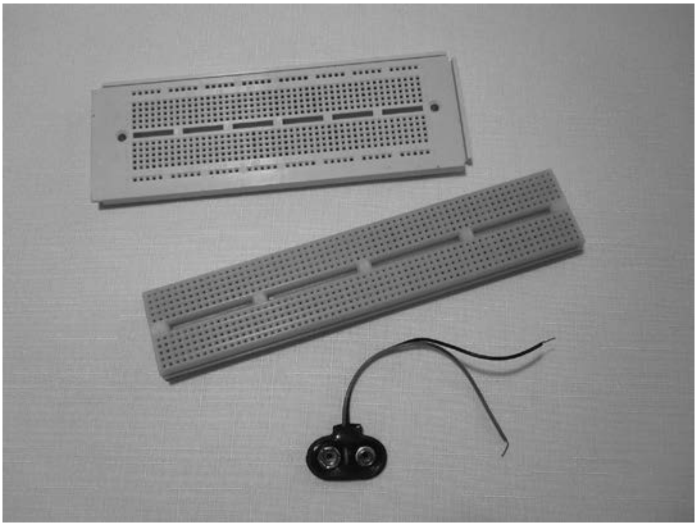

FIGURE 3.21 Common breadboarding socket for building prototypes or just experimenting. *Courtesy Global Specialties.*

Continued

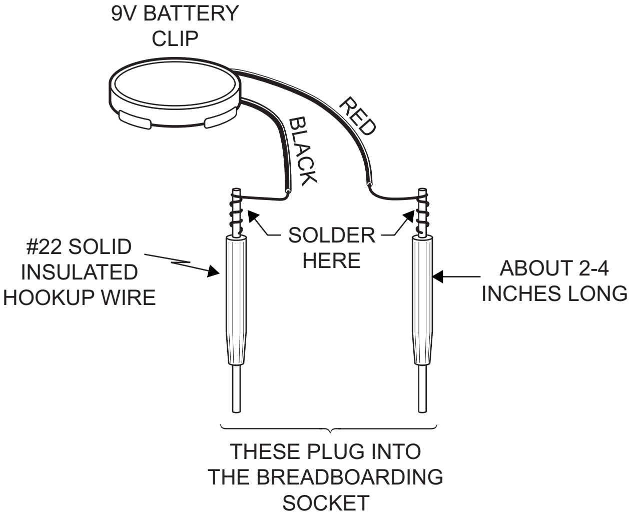
9V BATTERY
CLIP
RED
BLACK
SOLDER
HERE
#22 SOLID
INSULATED
HOOKUP WIRE
ABOUT 2-4
INCHES LONG
THESE PLUG INTO
THE BREADBOARDING
SOCKET

FIGURE 3.22 Soldering wires on the battery clip so they will plug into the breadboarding socket.

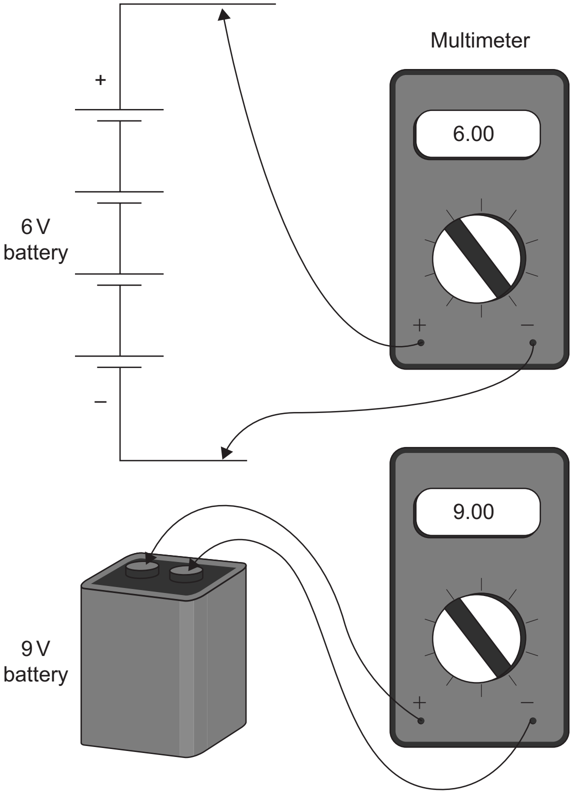
Multimeter
+
6.00
6 V
battery
+
−
−
9.00
9 V
battery
+
−

FIGURE 3.23 Measuring battery voltage with multimeter.

Project 3.1—cont’d

Next, test the battery. Using your multimeter on the DC ranges, measure the voltage across the battery holder wires. What do you get? You should read 6 V or a bit more. The 9-V battery should, of course, measure 9 V.

Now you are ready for a few additional component projects.

Project 3.2

Light-Emitting Diode

When you are buying your parts, get an LED and a set of resistors. Any old common red or green LED will work. The resistor set will be a package of all the most common values of resistors.

Now, wire up the circuit in Fig. 3.24. A pictorial is shown in the figure as well. Be sure to get the LED leads connected correctly, for if not, you will zap it. See the wiring details in the figure.

The resistor is needed to control or limit the current so that you do not blow out the LED. A value of 220 Ω is a good one if you use the 6-V battery. The color code for that resistor is red–red–brown–gold. If you are using the 9-V battery, use a 330-Ω resistor (orange–orange–brown–gold).

Wire the circuit and connect the battery. The LED should light. Now you are an electronic circuit genius.

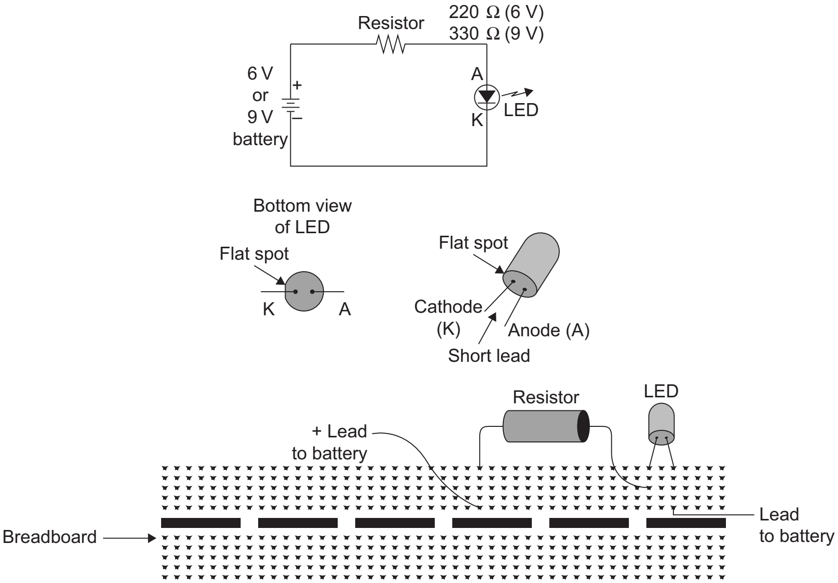
Resistor
220 Ω (6 V)
330 Ω (9 V)
6 V
or
9 V
battery
+
−
A
K
LED
Bottom view
of LED
Flat spot
K
A
Flat spot
Cathode
(K)
Anode (A)
Short lead
Resistor
LED
+ Lead
to battery
Lead
to battery
Breadboard

FIGURE 3.24 This is the circuit used to light an LED and how to wire it on the breadboard.

Project 3.3

Capacitors

Capacitors store a charge and then release it. The charge is stored energy. You charge the capacitor by applying a voltage to it. Then you discharge the capacitor by connecting it to a load such as a resistor. This brief project will illustrate that idea.

You will need the following parts from a parts source. Get a 100K-Ω resistor (color code brown–black–yellow–gold) and a 100-μF (microfarad) electrolytic capacitor. This capacitor is polarized, meaning that you have to observe the polarity of the voltage applied to it. The wire leads are usually marked with either a + sign, or sometimes with a – sign.

Wire the circuit shown in Fig. 3.25 on the breadboarding socket. Check the capacitor lead polarity. Do not initially connect the positive lead of the battery to the resistor. Connect your multimeter across the capacitor. Now connect the 6 V (or 9 V) from the battery supply as indicated when you are ready to start charging the capacitor. You should see the voltage begin to rise slowly. It should take about 50 s for the capacitor to fully charge to 6 (or 9) volts.

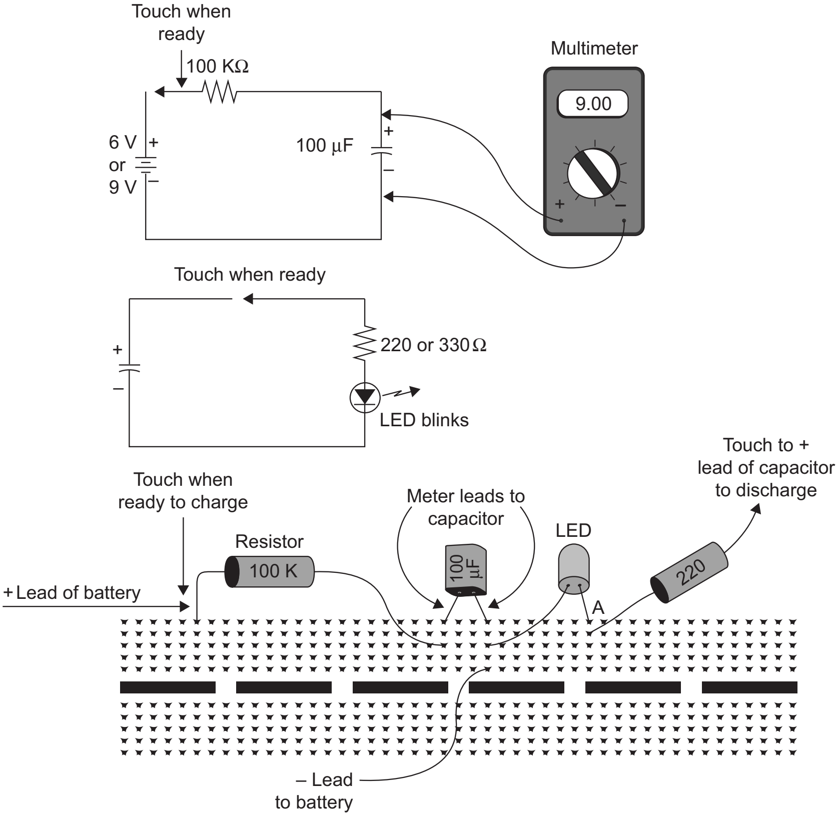

Touch when ready
100 KΩ
Multimeter
9.00
6 V or 9 V
+
–
100 μF
+
–
+
–
Touch when ready
+
–
220 or 330 Ω
LED blinks
Touch when ready to charge
Touch to + lead of capacitor to discharge
Meter leads to capacitor
LED
Resistor
100 K
100 μF
A
220
+ Lead of battery
– Lead to battery

FIGURE 3.25 Circuits for charging and discharging capacitor and respective wiring.

Project 3.3—cont'd

Now disconnect the battery's positive lead. The capacitor retains the charge, as you should see on the meter. It will probably begin to drop slowly as it discharges into the meter.

Now use the capacitor as a voltage source. Connect the capacitor to the resistor and LED you used earlier as shown in Fig. 3.25. The capacitor will discharge quickly, and you should only see the LED blink briefly. To see it again, recharge the capacitor and repeat.

Project 3.4

Transistor Switch

A transistor makes a good switch, especially a MOSFET. And it is easy to use. This project demonstrates how the MOSFET can be used to turn the LED off and on.

Acquire a MOSFET, such as the IRF510, or an equivalent such as the 2N7000, etc. Then build the circuit shown in Fig. 3.26. Be sure to follow the guidelines for wiring the transistor. Use the LED and 220 (or 330)-Ω resistor that you used before.

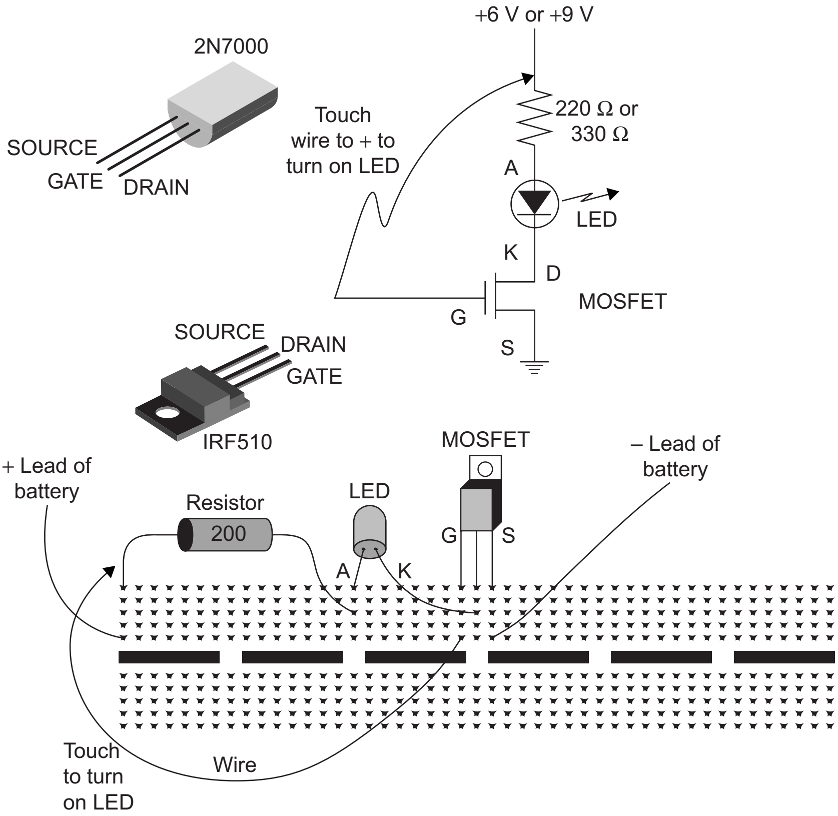
+6 V or +9 V
2N7000
SOURCE
GATE
DRAIN
Touch wire to + to turn on LED
220 Ω or 330 Ω
A
LED
K
D
MOSFET
G
S
SOURCE
DRAIN
GATE
IRF510
MOSFET
– Lead of battery
+ Lead of battery
Resistor
200
LED
G
S
A
K
Touch to turn on LED
Wire

FIGURE 3.26 Circuit that turns an LED off and on with MOSFET and wiring.

Now, connect the battery. Touch the wire from the gate on the MOSFET to the +6 V or +9 V connection like on the resistor lead. This will turn the transistor on, and it will act as a switch to turn on the LED. Disconnect the gate from +6 V and the transistor switch goes off, turning off the LED.

Project 3.5

Measuring Resistance and Continuity

In some instances, you may need to measure a resistance. Your multimeter can do this. Here is a short project to give you some practice in measuring a resistor value. In addition, you will see how to use your multimeter to check for continuity, that is, to verify that there is a direct connection between two different points.

Get some of the resistors you used in the previous projects and check their color code. For example, one resistor may have the color code yellow–purple–red–gold. The color code says that this is a 4700-Ω 5% resistor. The 5% means that the actual value of the resistor will be between +5% and −5% of the color-coded value. 5% of 4700 is 235 Ω. So the resistor value will be between $4700 - 235 = 4465$ and $4700 + 235 = 4935$ Ω.

Now, choose a resistor and calculate the possible range. Then, use your multimeter to measure the actual value. Set the multimeter to the resistance (R) range. Try the different ranges to get the right reading. The resistor value should be within the range. You may want to repeat this process with another resistor to be sure you know how to do it.

Now check the continuity of something. This is the process of measuring the resistance between two points that should be connected together with a wire, connector, or other conductor. The resistance should be zero or very low. A good test is to measure the resistance of piece of wire. Wire does have resistance but it is very low and your meter may not record it.

If you have a roll of #22 hookup wire, measure its resistance. Set the multimeter to its lowest resistance range. The meter digital readout should initially be 1 meaning infinite resistance or an open circuit. Now touch the meter probes to the two ends of the wire. You should read 0 Ω or close to it. This means that the wire has continuity.

Note: Just remember that when you are measuring resistance or checking for continuity, there should be no current flowing or voltage applied. Remove the resistor completely from any circuit and disconnect any wire or cable from power before you use a multimeter on it. Otherwise, you will damage your meter.

Chapter 4

Electronic Circuits: Linear/Analog

The Building Blocks of Electronic Equipment

In this Chapter:

- Linear circuits defined.
- Types of linear circuits.
- Operation of the most common linear circuits such as amplifiers and power supplies.

INTRODUCTION

The basic building blocks of electronic equipment are individual circuits made up of transistors, resistors, capacitors, and sometimes transformers, inductors, or other components. It used to be that circuits were made up of individual discrete components that were wired together on a printed circuit board (PCB). But today, most electronic circuits are in integrated circuit (IC) form, or more likely multiple circuits are already combined on a single silicon chip to create a larger circuit, subsystem, or system on a chip.

While most circuits are ICs there are still some cases where discrete components are necessary or desirable. The most common examples are high voltage or high power circuits that would burn up a tiny silicon chip. Discrete components have not gone away entirely, and you will find some of them external to an IC such as larger capacitors or inductors that just cannot be easily integrated onto a chip or transistors that can handle higher voltage or power. As mentioned before, we are not going to talk about the makeup or operation of individual circuits. Instead, this book treats circuits as just building blocks with an input and an output that process the input in some way to create a new, different output.

LINEAR CIRCUITS

There are two basic types of circuits, linear and digital. Digital circuits are covered in the next chapter. Linear circuits are those circuits that process analog

Electronics Explained. http://dx.doi.org/10.1016/B978-0-12-811641-8.00004-7
Copyright © 2018 Elsevier Inc. All rights reserved.

signals. Recall that analog signals are those that vary smoothly and continuously over time in contrast to pulse and digital signals that are voltages that switch off and on very fast. Analog signals are like sine waves, voice signals, video signals, radio waves, and voltages from sensors like a temperature transducer. DC, either a constant value of voltage or one that is varying, is also an analog signal.

A linear circuit is one whose output is proportional to its input. See Fig. 4.1. Its so-called response is a straight line or what a mathematician would call a linear function. The best example of this is the amplifier where its output is proportional to its input so that it accurately reproduces the input faithfully but at a higher voltage or power level. In the figure, with 0.4-V input you get 4-V output. That means the amplifier has a gain of 10. Linear circuits do not distort the signal. This chapter covers linear circuits with most of the emphasis on amplifiers since that is the largest category of linear circuits. And as it turns out, most other types of linear circuits are just amplifiers connected in different ways to give different processing results.

Amplifiers

The basic block diagram symbol for an amplifier is a triangle. See Fig. 4.2. It has an input and an output. Amplifiers also have one or more DC inputs as well. This DC from a battery or other power supply is what powers the circuit, and it is this DC that is ultimately converted into the new, larger output signal. In most cases the DC inputs are not shown in a block diagram. The amplifier processes the input to create an output of the same shape but at a larger amplitude.

The primary characteristic of any amplifier is its *gain*. The gain (usually represented by A) is simply a number equal to the ratio of the output to the input.

$$A = \frac{\text{Output}}{\text{Input}}$$

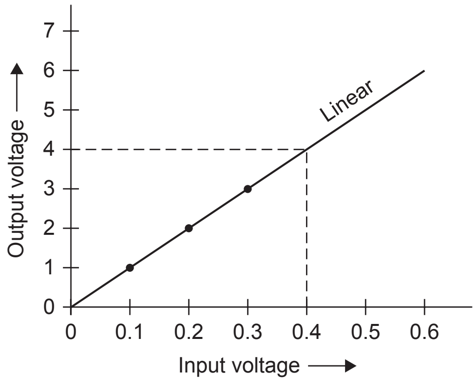

FIGURE 4.1 Linear circuit has a straight-line response. As input varies, output varies by factor of 10 larger, producing straight-line output.

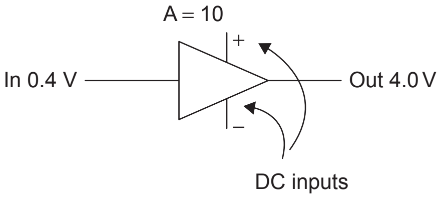

FIGURE 4.2 Block symbol for amplifier. Gain, A, is often given. DC power inputs are not usually shown.

If the output is 4 V and the input is 0.4 V, the gain is simply:

$$A = \frac{4}{0.4} = 10$$

The amplifier multiplies the input by 10 to get the output.

$$\text{Output} = 10 \times \text{Input}$$

Classifying Amplifiers

The amplifier is the most common linear circuit. There are so many different types of amplifiers that we divide them up into different types to distinguish one from another. And there are multiple ways of doing this. Here are the most common.

Small Signal Versus Large Signal

Small signal amplifiers are those that, as the name implies, only amplify small signals that are those roughly below about 1 V. Small signal amplifiers amplify millivolt, microvolt, or nanovolt signals. Large signal amplifiers amplify larger signals and typically amplify the power of a signal rather than the voltage. These are called power amplifiers (PAs).

Frequency Classification

Amplifiers are also categorized by the frequency of the signals they amplify. Some examples are DC amplifiers, audio frequency amplifiers, intermediate frequency (IF) amplifiers, video amplifiers, radio frequency (RF) amplifiers, and microwave amplifiers.

Configuration Classification

The two main configurations are single ended and differential. A single-ended amplifier is one in which the input and output are referenced to a common ground. These amplifiers have two inputs, one is the signal and the other is ground. See Fig. 4.3A. Normally only the input is shown as the ground is just assumed to be there.

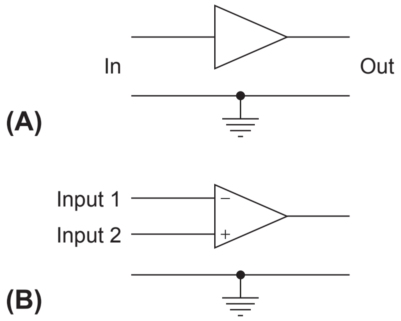

FIGURE 4.3 Single-ended (A) and differential input (B) amplifiers. Inputs and outputs are referenced to ground.

A differential amplifier has two inputs as Fig. 4.3B shows. These are usually called the inverting (2) and the noninverting (1) inputs. Both inputs are referenced to ground. The amplifier is called a differential amplifier because it actually takes the mathematical difference between the two inputs and amplifies the result. One input is subtracted from another, then the difference is amplified by the gain A. The formula for this operation is:

$$\text{Output} = (\text{Input } 2 - \text{Input } 1) \times \text{A}$$

Differential amplifiers have one neat feature in that any signal common to both inputs like noise is automatically canceled out. This feature is widely taken advantage of in very small signal amplifiers used in memory chips and high gain amplifiers that amplify small signals from sensors. Noise that is picked up simultaneously on both inputs simply subtracts itself out without bothering the desired small difference signal.

Classification by Class of Operation

One way to categorize amplifiers is to designate the conduction of the transistors in the amplifiers. The primary classes are A, B, AB, C, D, E, and F. There are a few others but these are the most common. Class A amplifiers conduct continuously. We say that the amplification occurs over the full 360 degrees of a sine wave cycle. See Fig. 4.4. Note: The "t" on the waveforms means time as the voltage is varying over time.

Class B amplifier transistors only conduct for half a cycle or 180 degrees of a sine wave. See Fig. 4.5. Such an amplifier seems useless since it distorts the signal. But if you combine two transistors, one to amplify positive cycles and the other to amplify negative cycles, then combine the two, the complete wave is amplified. Such an amplifier is called a push–pull amplifier. The benefit of this arrangement is that the circuit is more efficient.

Class AB amplifiers are a variation of the class B. Class B amplifiers do add a little distortion to the signal around the point where the sine wave crosses

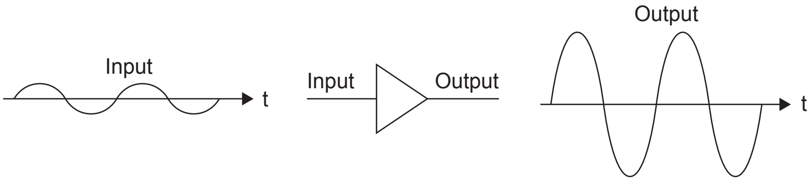
Input
t
Input
Output
Output
t

FIGURE 4.4 Class A amplifier conducts continuously.

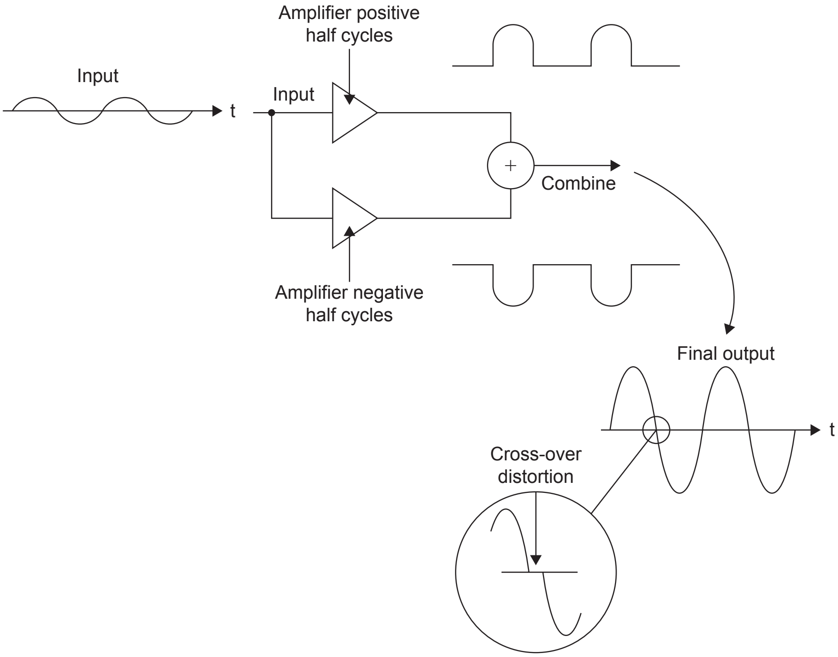
Amplifier positive
half cycles
Input
t
Input
+
Combine
Amplifier negative
half cycles
Final output
t
Cross-over
distortion

FIGURE 4.5 Class B amplifier conducts half the time, making it more efficient.

zero. This is caused by the transistors not turning on and off precisely at the zero point. This glitch shown in Fig. 4.5 is called crossover distortion. To get rid of that, class AB amplifiers allow some small current to flow continuously. The result is less distortion but a little less efficiency.

Class C amplifiers are for RF signals and are usually PAs. They only allow the transistor to conduct for less than 180 degrees of a sine wave input, usually 90–150 degrees. It produces massive distortion. However, the amplifier output is made up of an inductor and capacitor (LC) tuned circuit that oscillates at the operating frequency and removes the distortion. These amplifiers are very efficient.

In a class C amplifier the MOSFET acts like a switch that is turned off and on by the input signal, usually a sine wave. See Fig. 4.6. When the MOSFET is on, the capacitor charges up to the DC supply voltage and current flows in the

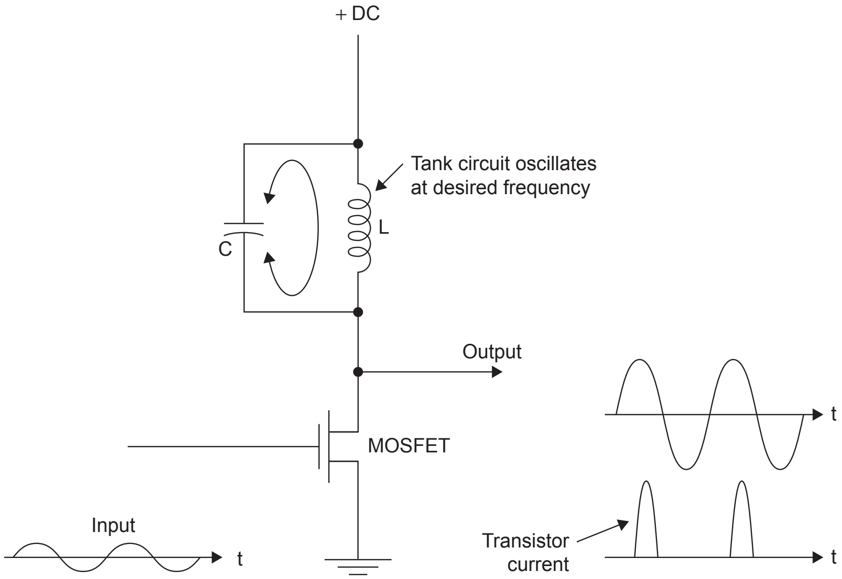

FIGURE 4.6 Class C amplifier only conducts for part of a half cycle, but the current pulse stimulates resonant tank circuit that oscillates at desired frequency.

inductor creating a surrounding magnetic field. When the transistor turns off, the inductor and capacitor begin to exchange energy and set up an oscillation at the resonant frequency of the LC circuit. This is called a tank circuit. The result is that the energy stored in the tank circuit produces the amplified sine wave output.

Another way of looking at this is that the transistor switch distorts the input creating a pulse waveform that contains lots of harmonics. The tank circuit acts like a selective band-pass filter that lets the fundamental sine wave pass while the harmonics are filtered out.

Don't forget the Fourier theory and harmonics. Remember that a harmonic is an integer multiple of a fundamental sine wave. If the fundamental sine wave is 20 MHz, the second harmonic is 40 MHz, the third harmonic is 60 MHz, the fourth harmonic is 80 MHz, and so on. Any nonsine wave signal, such as a half sine wave, square wave, or pulse is said to be made up of the fundamental sine wave added to multiple harmonics. A distorted sine wave has the fundamental sine wave in it but also harmonics. This concept is called the Fourier theory after the French mathematician.

Class D amplifiers are a special class called a switching amplifier. It is made up of transistor switches rather than a real linear amplifier. This amplifier chops the input analog signal into high-frequency pulses with a varying width. See Fig. 4.7. This process is called pulse width modulation (PWM). The sine wave audio signal to be amplified is sent to one input of a special amplifier called a comparator

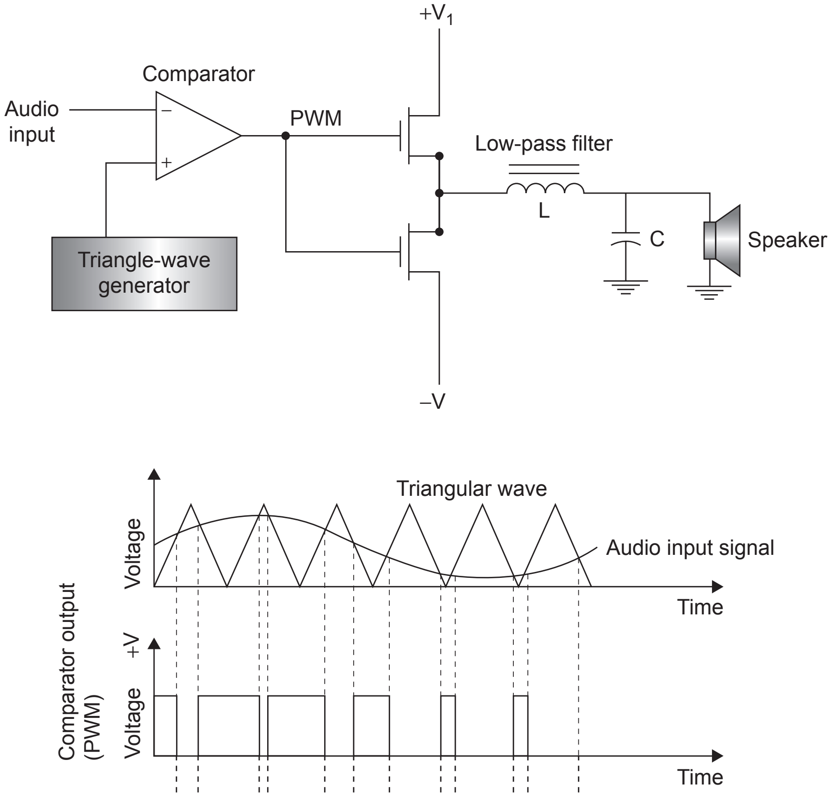
+V$_1$
Comparator
Audio input
−
+
PWM
Low-pass filter
L
C
Speaker
Triangle-wave generator
−V
Triangular wave
Audio input signal
Voltage
Time
Comparator output (PWM)
+V
Voltage
Time

FIGURE 4.7 Class D amplifier uses pulse width modulation (PWM) at higher frequency to switch higher voltages for greater output. Class D amplifiers are more efficient than any other type, and thus do not produce as much heat.

along with a higher-frequency triangle wave. The comparator output switches when the triangle wave and sine wave values are equal. The resulting PWM signal is then fed to MOSFET switches to make the signal bigger. The higher amplitude output it then filtered back into an analog signal by a low-pass filter (LPF) made up of a capacitor and inductor. Most class D amplifiers are audio PAs whose load is a speaker or headphone. For low power less than a few watts, all the circuitry is in a single IC. For higher power, the MOSFETs are larger external devices.

The big benefit of switching amplifiers is that for a given output power they are very efficient. Whereas a class AB amplifier may only be 20%–30% efficient, the class D is over 90% efficient. This means that the amplifier can be smaller and use less power and dissipates less heat. Class D amplifiers are great for portable battery-operated devices such as MP3 players and cell phones.

Classes E and F are special switching amplifier variations used at RF frequencies only. Like the class C amplifier, they use LC circuits to filter out the harmonics they produce. They are very efficient.

Operational Amplifiers

There are many special types of amplifiers. The most widely used is the operational amplifier, or op amp. This is a versatile high-gain DC differential amplifier that can be configured with external resistors, capacitors, diodes, etc., to perform a wide variety of amplification functions. The op amp may be the most widely used type of amplifier because of its versatility.

Fig. 4.8 shows the most common op amp configurations. In most cases, the resistors are external to the IC amplifier. Note that the gain is determined by the values of the external resistors. Note formulas for gain (A).

A variation of the op amp is the instrumentation amplifier. See Fig. 4.9. It is a combination of multiple op amps that make it very useful in industrial

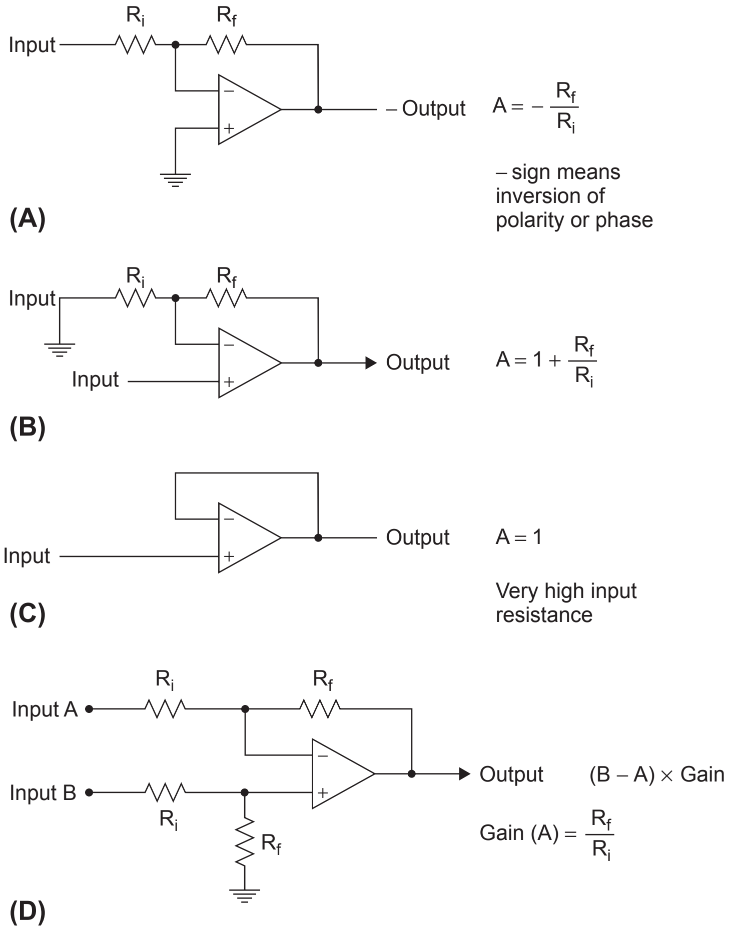
(A)
Input
R_i
R_f
–
+
– Output
$A = -\frac{R_f}{R_i}$
– sign means inversion of polarity or phase
(B)
Input
Input
R_i
R_f
–
+
Output
$A = 1 + \frac{R_f}{R_i}$
(C)
Input
–
+
Output
A = 1
Very high input resistance
(D)
Input A
Input B
R_i
R_f
R_i
R_f
–
+
Output
(B – A) × Gain
Gain (A) = $\frac{R_f}{R_i}$

FIGURE 4.8 Most commonly used operational amplifier circuit configurations. (A) Inverting amplifier, (B) noninverting amplifier, (C) follower, and (D) differential amplifier.

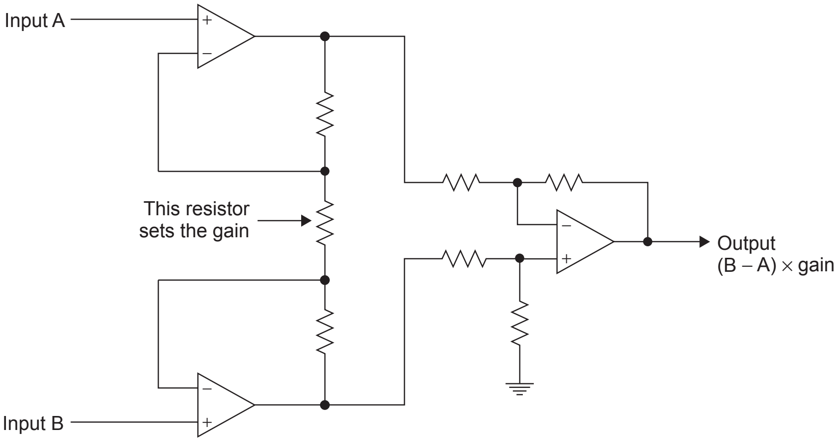

FIGURE 4.9 Instrumentation amplifier widely used to amplify small sensor signals in industrial applications.

applications such as amplifying small signals from sensors. Again the gain is determined by the resistor values. You can make this amplifier from individual op amps but usually you just buy it as a single IC. The external resistor sets the gain.

A variable or programmable gain amplifier is one whose gain can be controlled from an external DC input.

Amplifier Specifications

While you really do not need to know much about what the circuitry in an amplifier IC is, you do need to understand the specifications. Here are the most common specifications to look for.

As mentioned before, gain is the primary amplifier characteristic. It is usually a voltage or power ratio, depending on the type of amplifier.

$$\text{Voltage gain is output voltage over input voltage}: A_v = \frac{V_o}{V_i}$$

$$\text{Power gain is output power over input power}: A_p = \frac{P_o}{P_i}$$

Many amplifiers have their gain expressed in decibels (dBs). A dB is a gain measurement based on a logarithmic scale rather than a linear scale. This has the advantage of being able to express large and small gains in smaller numbers. And gains of multiple stages can be added rather than multiplied. To calculate dB gains you used these formulas:

$$\text{Voltage gain}: \text{dB} = 20 \log (A_v)$$

$$\text{Power gain}: \text{dB} = 10 \log (A_p)$$

In these formulas, log means the common (base 10) logarithm of the voltage or power gain. The logarithm (log) is easy to find using a scientific calculator.

For example, a voltage gain of 50 expressed in dB is:

$$\text{dB} = 20 \log 50 = 20\,(1.7) = 34\ \text{dB}$$

A power gain 300 expressed in dB is:

$$\text{dB} = 10 \log 300 = 10\,(2.5) = 25\ \text{dB}$$

Input Impedance

Input impedance is the load seen by the signal source to the amplifier. It is usually resistive but typically also has some capacitance across it. Fig. 4.10A shows an equivalent circuit of an amplifier. Note the input resistance. That is the input impedance. It may be as high as several megohms or could be as low as, say, 50 Ω. Usually the higher the better for voltage amplifiers but not for PAs.

Impedance

The term *impedance* means opposition to current flow. It is actually the combined opposition offered by resistance plus any capacitive or inductive reactance. If there is no reactance, the impedance is just the resistance so the two terms can be used interchangeably. Impedance is represented by the letter Z.

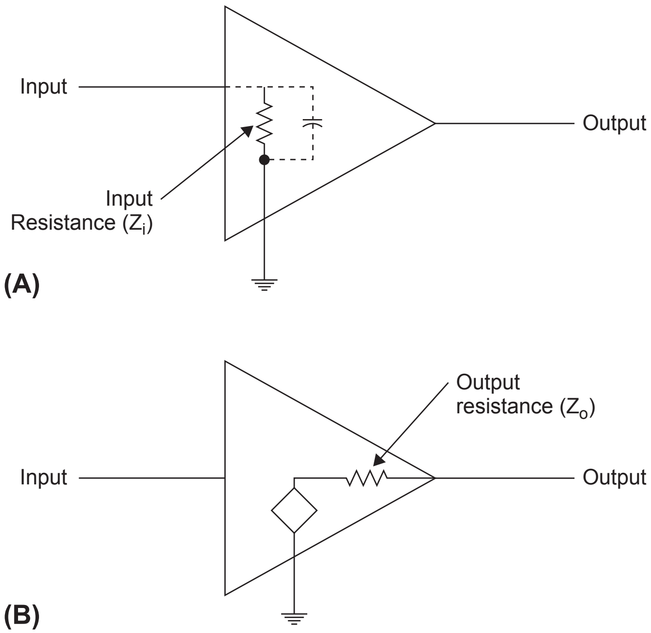

FIGURE 4.10 Amplifier input impedance (A) and output impedance (B).

Output Impedance

Amplifiers act as signal generating sources because they essentially just create a new higher voltage or higher power signal with the same shape as the input. They appear as a signal generator to the next circuit or load. Fig. 4.10B shows the amplifier as a generator shaped as a diamond. It has an internal impedance or series resistance called the output impedance. It is not a real resistor but it is there and it appears in series with whatever load is connected, which may be another amplifier or something else.

Fig. 4.11 shows the amplifier connected to another amplifier. Note that the output impedance for amplifier A_2 and the load form a voltage divider so some of the amplification is lost to the voltage divider effect. The input impedance of A_2 is the load for the output impedance of A_1.

Cascading Amplifiers

If an amplifier does not have enough gain, then multiple amplifiers can be cascaded one after the other to give the needed gain. One amplifier amplifies the output of the next and so on. See Fig. 4.12. To calculate the overall gain of an amplifier you just multiply the gains as shown in the figure. This procedure assumes that one amplifier does not load the next. That is, the output impedance

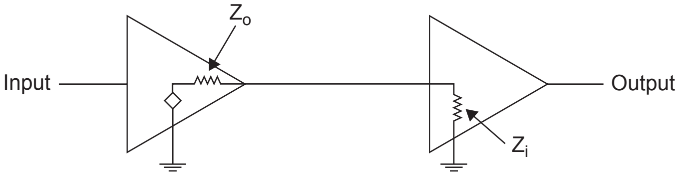

Z_o and Z_i form a voltage divider that lowers or cancels gain

FIGURE 4.11 When one amplifier feeds another, input and output impedances form a voltage divider that attenuates the signal and offsets some of the gain.

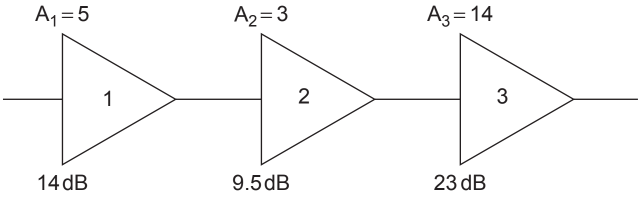

$A_T = A_1 \times A_2 \times A_3$
$A_T = 5 \times 3 \times 14 = 210$
$A_T = 14 + 9.5 + 23 = 46.5\,\mathrm{dB}$
$210 = 46.5\ \mathrm{dB}$

FIGURE 4.12 Cascading amplifier provides more gain. Total gain is product of individual gains or sum of decibel (dB) gains.

of each amplifier is assumed to be zero and the input impedance of the next amplifier is assumed to be infinite. That way, there is no voltage divider effect that lowers the gain.

If the amplifier gains are given in dB, you just add the dB figures to get the total gain.

Efficiency

Efficiency is the ratio of the output power to the input power expressed as a percentage.

$$\%\ \mathrm{Eff} = \frac{P_o}{P_i} \times 100$$

An amplifier takes the DC supply voltage and converts it into the AC output. The efficiency expresses how well the amplifier does this. Any lost power usually gets converted to heat which is not good. Nevertheless, some heat will always be generated just because electronic circuits are not perfect. High efficiency is a good thing of course, but often efficiency is traded off for better linearity, fidelity, and lower distortion of the signal.

Class A amplifiers have the poorest efficiency commonly in the 10%–30% range. Classes B and AB are more efficient in the range of 50%–70%. Class C amplifiers have an efficiency of 70%–85%. Class D amplifiers are the best with efficiencies in the 80%–95% range.

Frequency Response

This is a measure of the range of frequencies that amplifiers cover. DC amplifiers may only cover DC up to about a few hundred Hertz. Audio amplifiers may cover voice frequencies from 300 to 3000 Hz or stereo amplifiers that cover 20 Hz–20 kHz. IF amplifiers may only cover a specific band, say, 455 kHz, 9 MHz, 45 MHz, 70 MHz, or some other popular IF value. Op amps amplify signals from DC up to some very high frequency as high as several hundred MHz in some cases.

RF amplifiers cover high radio frequencies from a few MHz up to microwaves from 1 to 100 GHz. Their frequency and range of coverage is usually set by a filter or LC tuned circuit.

Fig. 4.13A shows the response curve for an op amp with an output down to DC and up to the cutoff frequency (f_{co}) or 5 MHz. Fig. 4.13B is the response curve of a stereo amplifier with a range of 20 Hz–20 kHz.

Output Capability

This specification indicates the range of output signals the amplifier is capable of. It is usually stated as a maximum voltage swing or as a maximum amount of output power. In both cases, the maximum output is usually given for a given value of load resistance. For example, an op amp may be capable of producing

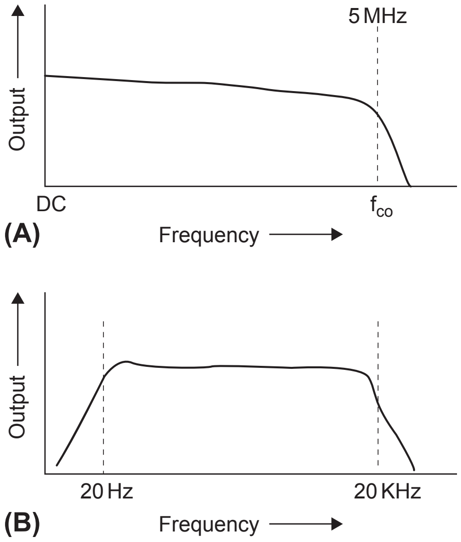

FIGURE 4.13 Frequency-response curves show output versus input frequency. Curve in (A) is response of operational amplifier that extends from DC to 5-MHz cutoff frequency. Curve in (B) is response of common stereo audio amplifier.

an output of ±12 V across a 2-k load. A PA may produce a maximum output of 10 W across an 8-Ω speaker load or, say, 50 W across a 50-Ω antenna load.

Filters

A filter is a frequency sensitive circuit that is designed to pass some frequencies, but reject others. There are various ways of making filter circuits that will accept some frequencies, but greatly attenuate others. There are literally thousands of applications for such circuits. But the main use of filters is to retain a desired signal, but eliminate interference and noise.

There are four basic types of filters: LPF, high-pass filter (HPF), band-pass filter (BPF), and band-reject filter. The names, of course, tell their function.

A LPF is a circuit that passes all frequencies between DC and some upper cutoff frequency (f_{co}). All frequencies above the cutoff frequency are rejected. Fig. 4.14A shows the response curve of an LPF. This is a plot of the filter output amplitude versus frequency. Note the gradual slope as signals above the cutoff frequency (f_c) are reduced in amplitude. The block symbol is how LPF is represented in some diagrams.

A high-pass filter passes frequencies above the cutoff and greatly attenuates those below the cutoff. Fig. 4.14B shows the ideal response curve of an HPF. Note the HPF symbol.

BPF pass a certain range of frequencies, but reject all those above and below that range. Fig. 4.14C shows the practical response curve of a BPF. Note that

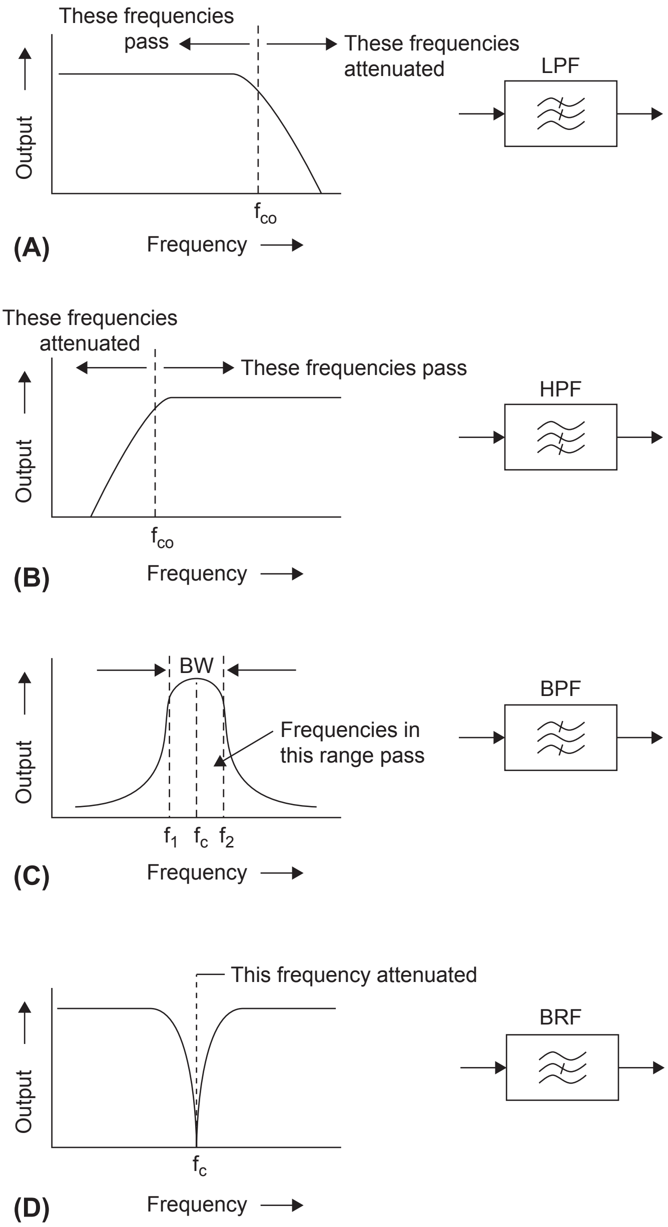
These frequencies pass
These frequencies attenuated
LPF
Output
f_{co}
(A)
Frequency
These frequencies attenuated
These frequencies pass
HPF
Output
f_{co}
(B)
Frequency
BW
BPF
Frequencies in this range pass
Output
f_1 f_c f_2
(C)
Frequency
This frequency attenuated
BRF
Output
f_c
(D)
Frequency

FIGURE 4.14 Four common filter types and their response curves and symbols. (A) Low-pass, (B) high-pass, (C) band-pass, and (D) band-reject or notch.

there are two cutoff frequencies, one above and one below the center frequency f_c. The difference between the upper and lower cutoffs is called the filter bandwidth (BW):

$$BW = f_2 - f_1$$

This is the range of frequencies passed. Again note the schematic symbol for a BPF.

Fig. 4.14D illustrates the curve for a band-reject or notch filter. It is used to eliminate one single frequency and those close around it.

Note in the block symbols in Fig. 4.14 that each has three waves representing high, middle, and low frequencies. The wave with the slash through it means that these frequencies are cut out or greatly attenuated.

Filters can be implemented in many ways. Resistor and capacitor (RC) filters are common as are LC filters. You can make filters with op amps called active filters. Other filters are made with mechanical vibrating elements, a ceramic or quartz crystal, or a special surface acoustic wave filter. Many filters today are made with digital signal processing. These will be discussed in Chapter 6.

Oscillators

An oscillator is a circuit that generates a signal. It is usually a sine wave or a rectangular wave. The electronic symbols used to represent an oscillator are given in Fig. 4.15.

The main specification of an oscillator is its output frequency. Most oscillators have a fixed frequency but there are oscillators whose frequency can be varied. A tunable frequency oscillator is called a variable frequency oscillator.

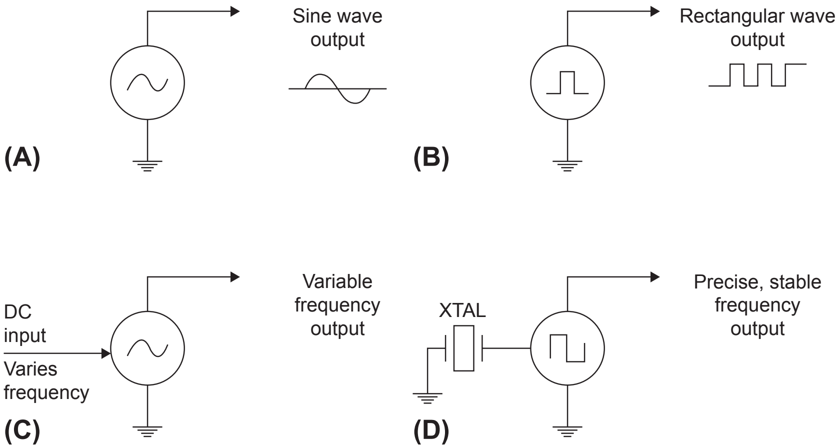

FIGURE 4.15 Types of oscillators and their symbols. (A) Sine wave, (B) rectangular pulse, (C) voltage-controlled oscillator, and (D) crystal clock oscillator.

You can often tune it by turning a knob that varies a capacitor or inductor to change the frequency. Some oscillators are tuned by an external voltage. These are called voltage-controlled oscillators (VCOs). By varying the value of a DC input voltage, the frequency is changed.

The frequency of an oscillator is set by some frequency determining element or circuit. That circuit may be made up of RC, LC, or a crystal. A crystal is a thin piece of quartz that can be made to vibrate at a very precise frequency. It maintains its frequency very closely unlike RC or LC networks that can drift with temperature or change from vibration or other conditions. Crystals come in a small metal package and are used to set the frequency in most oscillators today.

A special type of oscillator is called a clock oscillator. It is used as an accurate time and frequency source. The frequency determining element is usually a quartz crystal that vibrates at a precise frequency and is very stable. The clock signal is used to operate all the circuits in synchronism. An example is the clock oscillator in a PC that sets the speed of calculations. Sometimes you will see a quartz crystal (XTAL) symbol drawn with the oscillator symbol to illustrate this. See Fig. 4.14D.

Mixers

A mixer is a circuit that takes two input signals and mixes them in a special way to produce new output signals. There are two types of mixers, linear and nonlinear.

A linear mixer is one that takes the input signals and just adds them together. You can do this with some resistors as shown in Fig. 4.16A. The sum of both input signals appears across the output resistor.

You can also make a linear mixer with an op amp as shown in Fig. 4.16B. This op amp circuit is called a summer. Setting the resistor values allows you to mix and add gain at the same time. A good example of a linear mixer is those used by musicians so they can mix the outputs from microphones and musical instruments so they can all be amplified as one by the PAs that drive the speakers. Note the variable resistors called pots that are used as amplitude or volume controls.

A nonlinear mixer takes two input signals and effectively multiplies them together. The result is that the mixer produces four output signals. Fig. 4.17 shows the symbol for a mixer. The inputs are frequencies 1 and 2, or f_1 and f_2. The mixer outputs are f_1, f_2, and $(f_1 + f_2)$ or the sum, and $(f_1 - f_2)$ or the difference. The sum and difference frequencies are the ones of most interest. Usually a filter is used at the output of the mixer to select the desired frequency and eliminate the others. If the lower difference frequency is selected, the mixer is often called a down converter. If the sum frequency is selected, the mixer is called an up converter. The mixer is a way to translate a frequency to another

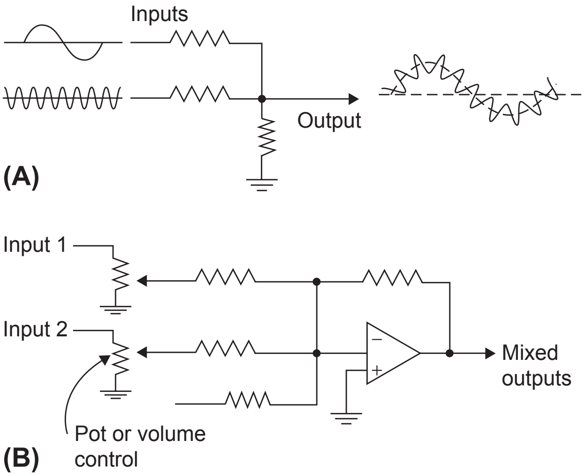
Inputs
Output
(A)
Input 1
Input 2
Mixed outputs
Pot or volume control
(B)

FIGURE 4.16 Linear mixing with resistors (A) or an operational amplifier summer (B).

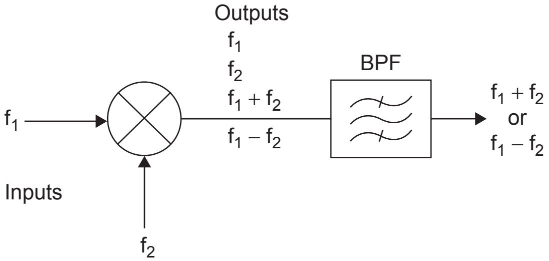
Outputs
f1
f2
f1 + f2
f1 − f2
BPF
f1 + f2
or
f1 − f2
f1
Inputs
f2

FIGURE 4.17 Nonlinear mixer used to generate higher and low frequencies.

higher or lower frequency that is better for processing. Any signal amplitude or frequency variations on the inputs are accurately carried through to the new output frequency.

Phase Detector

A phase detector is a mixer-like circuit that puts out a signal that is proportional to the phase difference between two input signals of the same frequency. See Fig. 4.18. A phase shift is a time difference between two signals of the same frequency. We sometimes need to know how much phase shift is present. The phase detector produces a series of output pulses whose width is proportional to the phase difference. Passing the pulses through a LPF smoothes them into a proportional DC voltage.

Filtering Pulses Into DC

A common occurrence in electronic circuits is the filtering of pulses into a DC voltage. You saw this back in Fig. 3.15 where positive pulses from a sine wave rectifier are smoothed into a near constant DC voltage. It is the capacitor that charges up and stores the pulse voltage. Then between pulses the capacitor discharges into the load maintaining a near constant DC voltage. The LPF usually includes a capacitor so performs the same function in the phase detector of Fig. 4.18. The output is a proportional average DC output.

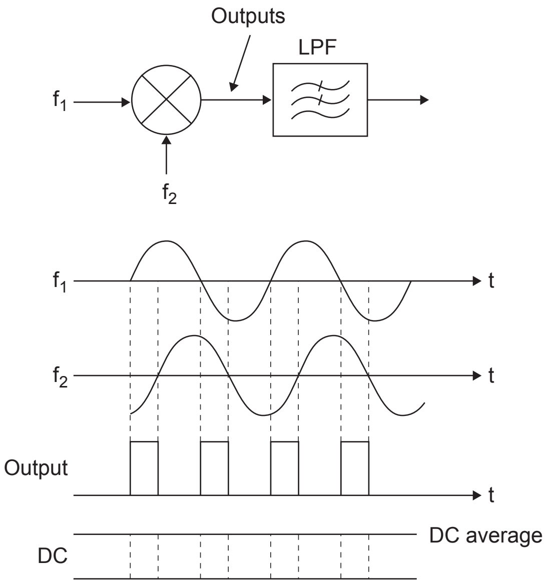

FIGURE 4.18 Phase detector converts phase shift into proportional DC average.

Phase-Locked Loops

The phase-locked loop (PLL) is a frequency- and phase-sensitive feedback control circuit. It consists of three major parts: a mixer or phase detector, a LPF, and a VCO, as shown in Fig. 4.19. The signal from the VCO is compared to the input signal. If there is a frequency (or phase) difference between the two, an error signal is generated. This error signal is filtered by the LPF into a varying DC level and is used to control the VCO frequency. This is the feedback signal.

If the VCO frequency, f_o, differs from the input reference signal frequency, the phase detector sees this as a phase shift. This causes the phase detector to produce an error signal. This error signal is filtered into a varying DC signal that is used to control the VCO.

The VCO can be either a sine wave oscillator or a rectangular wave oscillator depending upon the desired waveshape of the output. In either case, the VCO

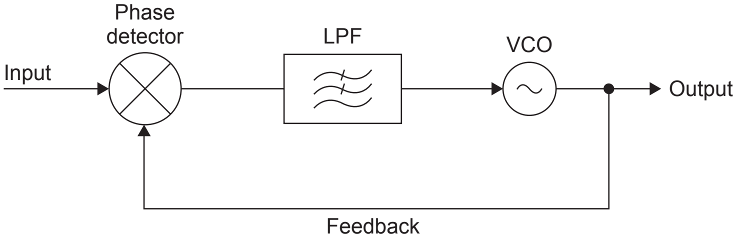

FIGURE 4.19 Phase-locked loop.

output frequency is made directly proportional to the DC control voltage. There is a linear voltage-to-frequency relationship in the VCO. This means that changing the control voltage produces a proportional change in output frequency. If the voltage goes up, so does the frequency.

Now if an input reference signal whose frequency is near that of the VCO is applied to the PLL, the phase detector will produce an output voltage proportional to the frequency difference. This signal is filtered and the resulting DC control voltage is applied to the VCO. The control voltage is such that it forces the VCO frequency to move in a direction that reduces the error signal. This means that the VCO frequency will change until it is equal to the input reference signal frequency. When this happens, the two signals are synchronized or "locked." The phase difference causes the phase detector to produce the DC voltage at the VCO input to keep the PLL locked to the input signal.

If the input reference signal changes, then the phase detector will recognize a frequency (or phase) difference between the input and the VCO output. As a result, the LPF will produce a different DC control signal that will force the VCO to change such that it is equal to the new input frequency. As you can see then, the PLL will "track" an input signal frequency as it changes.

The range of frequencies over which the PLL will track an input signal and remain locked is known as the lock range. This is a range of frequencies above and below the VCO free running frequency. The PLL can track and "lock" to any input frequency in this range. If an input signal out of the lock range is applied, the PLL will not synchronize.

If the input signal is initially outside of the lock range, the PLL will not lock. But, the PLL will jump into a locked condition as soon as the input frequency gets close to the VCO frequency. In other words, the PLL will "capture" the incoming signal if it is close enough to the VCO frequency. Once the input signal is captured, the PLL is locked and will track further changes in the input signal frequency.

The range of frequencies over which a PLL can capture a signal is known as the capture range. Like the lock range, it too is centered on the free running frequency. But the capture range is narrower than the lock range. The PLL acts as a frequency sensitive circuit over a narrow range of frequencies.

Since the PLL will only capture and lock on to input signals within a certain narrow band, the PLL acts like a band-pass filter. For that reason, the PLL is an excellent signal conditioner. You can take a noisy input signal or one with undesirable interference on it and filter it with a PLL. The PLL will lock on to only the desired frequency component of the input. The VCO reproduces the input signal at the same frequency but with the noise and interference removed. The PLL not only cleans up a signal but also can track it if its frequency changes.

The PLL is widely used for a variety of purposes. It is used to recover the clock signal in some wireless applications. It is used to recover the original signal in frequency modulation radio. It is used to multiply a frequency by a fixed factor. It can be used for motor speed control. Almost all electronic products of some kind contain a PLL.

Frequency Synthesizers

One of the most popular uses of the PLL is in frequency synthesizers. A frequency synthesizer is a very stable signal source that can be varied over a specific range of frequencies in finite increments or steps. Frequency synthesizers are used as precision signal generators for test and measurement purposes and replace more conventional oscillator circuits as the primary signal sources in radio transmitters or receivers.

The big advantage of a PLL synthesizer is its ability to generate a wide range of frequencies with great accuracy and stability by using only a single stable signal source. For example, a PLL synthesizer with a single crystal controlled reference input can readily generate hundreds or even thousands of discrete frequencies that are as stable as the crystal reference. By using a synthesizer as the initial signal source in a transmitter, a wide range of channels can be obtained with a single crystal. When many channels must be used, a synthesizer greatly reduces the cost since crystals are expensive.

The synthesizer circuit in Fig. 4.20 shows that it is a PLL with a frequency divider connected between the VCO output and the phase detector. The frequency divider is a circuit that divides its input frequency by some whole number. For example, the output of a divider with a division factor of 12 and an input of 48 MHz is 48/12 = 4 MHz.

The frequency synthesizer PLL also uses a stable, crystal oscillator input to the phase detector. This is usually derived from a crystal oscillator. When the PLL is locked, the stability and accuracy of the VCO output will be the same as that of the reference input. The frequency divider is usually made so that its divide ratio can be changed by setting a switch or changing the control inputs. Changing the division ratio changes the output frequency at which the loop will lock.

The reason for this is that the two inputs to the phase detector must always be equal in frequency if a locked condition is to be achieved. The reference frequency is fixed so the only thing that can change is the VCO output if the divider

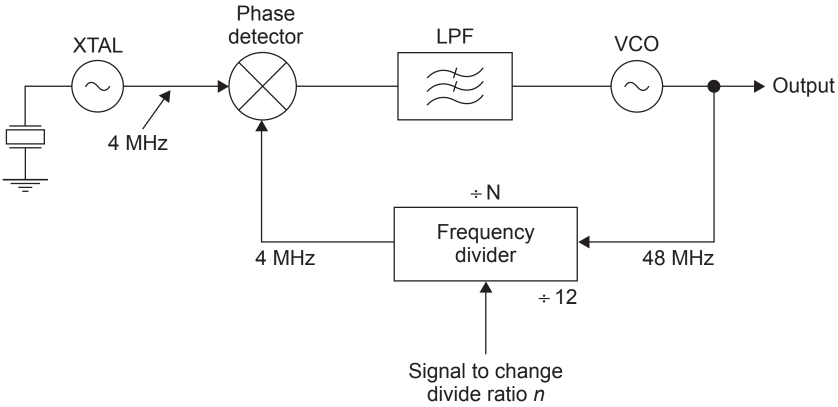

FIGURE 4.20 Frequency synthesizer using phase-locked loop. Changing the divider ratio changes the frequency. *LPF*, Low-pass filter; *VCO*, Voltage-controlled oscillator.

ratio is altered. If the loop is locked, the output of the divider will be 24 MHz no matter what the divide ratio is. Therefore, with an output of 4 MHz, a divide by six circuits must have a 24 MHz input. Since the VCO output is the divider input, the VCO and synthesizer output will be 24 MHz too.

As you can see, changing the divide ratio changes the output frequency. In this circuit, if the divide ratio is changed in integer values between 6 and 12, the output frequencies will be as shown in Table 4.1. With this setup, the output is switched in 4-MHz increments.

If the input reference is made 1 MHz, then when the divide ratio is changed, the output frequency will step in 1-MHz increments. By carefully selecting the reference frequency and the divide ratio, it is possible to synthesize practically any range of frequencies in many increments from subaudio to microwave.

Note one final fact about the PLL circuit in Fig. 4.20. If the reference is considered the input, and the VCO is considered the output, then the PLL is acting as a frequency multiplier. Here the output is 12 times the input. You will occasionally see the PLL used as a frequency multiplier in some applications.

The PLL synthesizer is a very versatile circuit. It is commonly available in IC form. PLL frequency synthesizers are used in TV channel selectors, stereo receivers, CB radios, cellular telephones, and other two-way radios. In such equipment the divider ratio is supplied by an input keyboard or a built-in microcomputer programmed to supply the desired divide factor for a selected output frequency.

A newer method of frequency synthesis is called direct digital synthesis (DDS). It uses digital circuits such as registers, adders, an accurate clock, a memory, a digital-to-analog converter to produce highly accurate sine waves and other waves whose frequency can be changed in small increments. The DDS is covered at the end of Chapter 5.

TABLE 4.1 Phase-Locked Loop Divide Ratio and Voltage-Controlled Oscillator (VCO) Output

Divide Ratio	VCO Output (MHz)
6	24
7	28
8	32
9	36
10	40
11	44
12	48

Power Supplies

A power supply is, as its name implies, a source of power to an electronic circuit. Most electronic circuits operate from DC as they process AC or DC signals. The most common power supply is the one that gets its input from the standard AC wall outlet that supplies 120 V at 60 Hz. This common voltage is then converted by the power supply into one or more DC voltages that go to operate the TV set, computer, or other equipment. This section discusses the main components and circuits of a power supply.

Batteries

One of the most widely used power supplies is a battery. A battery is a great source of DC by itself and no AC source is needed. Batteries were the very first form of voltage sources for electrical circuits, and they are still widely used today. What would we do without the batteries for our cell phones, iPods, cordless telephones, laptop computers, and remote controls? Just remember that a special form of power supply is the battery charger that normally uses AC to recharge lead–acid, nickel cadmium, nickel metal hydride, and lithium-ion batteries.

Standard Power Supply

A very common AC to DC power supply is shown in Fig. 4.21. It normally consists of a transformer that translates the 120-V AC line voltage into a higher or lower voltage as needed by its step up or step down characteristics. Today, since most equipment uses ICs, a lower AC voltage is needed. The lower AC voltage is then converted into a pulsating DC voltage by a rectifier. A rectifier is one or more diodes that act like polarity sensitive switches to change the positive and/or negative sine half cycles into pulses of DC. A filter is then used to smooth the pulses into a more constant DC voltage. The filter is usually a large capacitor

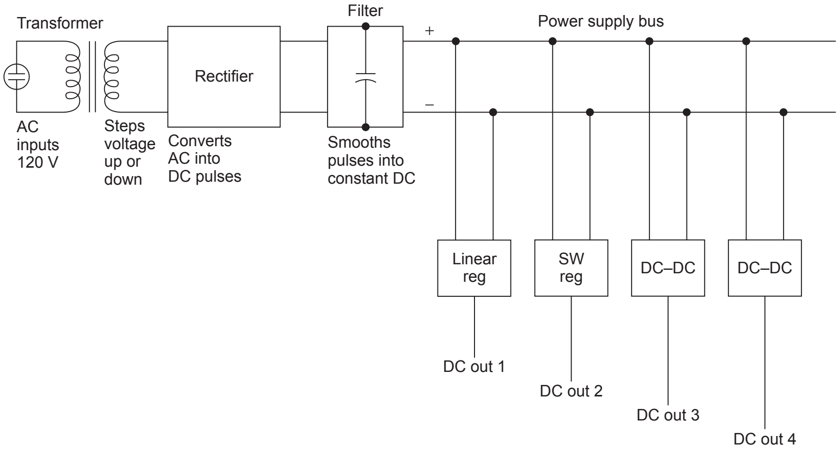

FIGURE 4.21 Typical power supply configuration using bus architecture, regulators, and DC–DC converters to get desired number of outputs. *reg*, Regulator.

that charges up to the peak of the DC pulses then only discharges very slowly into the load.

Regulators

A regulator is a circuit that maintains a fixed DC output voltage of a desired value. The DC produced by the power supply at the filter output will vary over a wide range as the AC input line voltage changes. Any changes in the load meaning variations in current taken from the supply will also cause the output voltage to vary. A regulator absorbs those variations and creates a constant DC output suitable for powering ICs and other electronic circuits. Common DC output voltages are 3.3, 5, 12, 15, and 24V. Other typical DC supply voltages are 1.2, 1.8, 6, and 48V.

There are two types of regulators, linear and switching. A linear regulator inserts a transistor in series between the power supply output and the load. If the load varies, circuitry in the regulator adjusts the conduction of the transistor so that a constant voltage is maintained. It is as if a resistor in series with the load forms an adjustable voltage divider. If the output voltage drops, the series transistor conducts more to keep the output voltage the same. If the output increases, the transistor conducts less to decrease the voltage and maintain the same value.

The other type of regulator is the switching regulator. It puts a transistor in series with the load as well but the internal circuitry turns the transistor off and on at a high rate of speed. By varying the duty cycle (ratio of on to time for one period) of the on–off pulses, the average DC output may be varied. Then if any output voltage changes are sensed, the duty cycle is adjusted to ensure that the average output stays the same.

Duty Cycle and Pulse Width Modulation

The duty cycle is the ratio of the pulse width to the period of a rectangular wave as shown in Fig. 4.22. The period is the reciprocal of the frequency of the signal or T = 1/f. It is constant. The on-time is the duration of the pulse voltage (t). The pulse duration can vary depending upon what the application is. If the period is 50 ms and the pulse on-time is 10 ms, the duty cycle (D.C.) is:

$$\text{D.C.} = 10/50 = 0.2 \times 100 = 20\,\%$$

Circuits can be made to change the on-time so that the duty cycle can be varied from a few percent to nearly 100%. This is called PWM. It was introduced earlier in the class D amplifier discussion. PWM is widely used in switching power supplies such as regulators and DC–DC converters.

Keep in mind that if you filter the pulses into an average proportional DC with a capacitor or LPF, you can vary the DC from some low value with a low duty cycle or some higher value with a higher duty cycle.

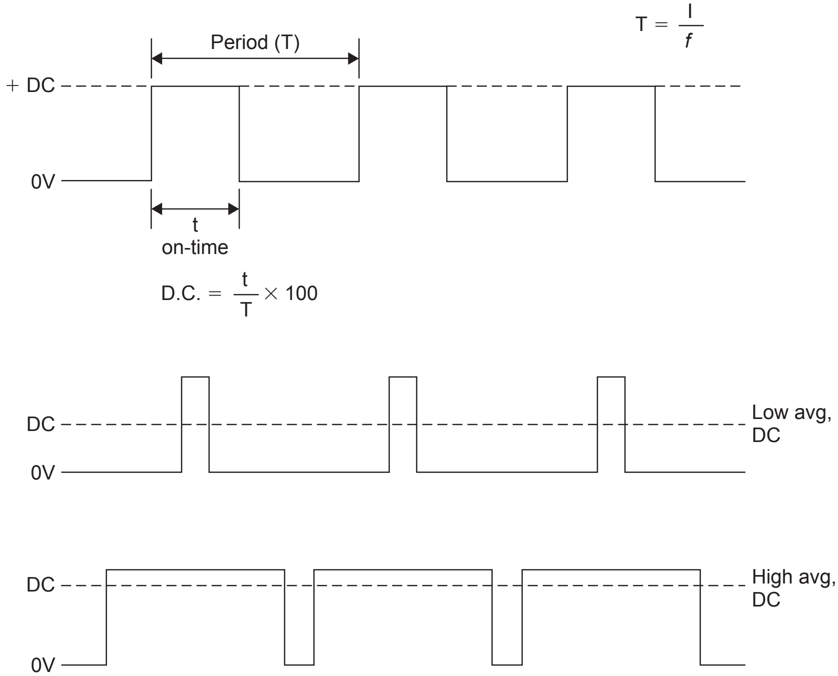

FIGURE 4.22 Concept of duty cycle and how varying pulse widths can be filtered into a proportional DC average voltage. *avg*, Average.

Linear regulators work very well but are not very efficient as they waste power in the series transistor. Typical efficiency is 10%–40%. A switching regulator is far more efficient as it is off for a large part of the time. Efficiencies in the 70%–95% range are common.

DC–DC Converters

A DC–DC converter is a circuit that does what its name implies. It translates one DC voltage to another. It can take the voltage from a 12-V battery and generate several hundred or thousand volts. Or it can take a 4.2-V battery and produce voltages of 3.3 or 2.5V and 15V. By building one DC power supply, several other output voltages can be obtained by using different DC–DC converters.

Bus-Oriented Architecture

Many power supplies have what we call a bus architecture. See Fig. 4.21. The DC power supply produces one main output voltage. That voltage is applied to a common DC bus. A bus is just the name for the two + and − DC lines that carry the voltage. Then, regulators and DC–DC converters are used to generate the voltages needed by the equipment. Again see Fig. 4.21.

Power Management

Many power supplies today include a circuit that manages the power supply voltages. It is usually a single IC that does things like turn on the various DC voltages in a power supply in a specific sequence. Some ICs require multiple DC voltages and they must be turned on one at a time in a specific sequence. In addition, the power management chip does things like monitor current usage. If one power supply is not being used, meaning current not being drawn from it, that part of the power supply is shut down to conserve power. A good example of this is the power management in a laptop computer or cell phone that monitors and controls all DC power to preserve the battery charge for longer life. Some other functions performed by power management chips are voltage monitoring with voltage alarm if a voltage fails, temperature monitoring, fault detection, over voltage or under voltage detection, and power control. Some power management chips also have the battery charging circuits built in. Every cell phone and laptop has one of these.

Inverters

An inverter is a power supply that converts DC into AC. A common unit converts 12V from a car battery into standard 120V 60Hz AC. With an inverter you can operate common household lights and appliances from a battery.

A special variation of an inverter is the uninterruptible power supply (UPS). These are widely used with computers and networks to provide power even during loss of AC input from the power line. A battery operates an inverter that is used to power the computers and other equipment. If the AC power fails, the equipment continues to operate. The battery is continuously charged by a built-in battery charger that operates from the AC power line.

Solar Power System

Any home or business solar power system is similar to a large UPS (see Fig. 4.23). It uses the solar panels to charge a bank of batteries that store the energy for use later when the sun goes down. The inverters convert the DC into standard 120/240V AC to power the house. In some cases, unused power is sent back to the power grid.

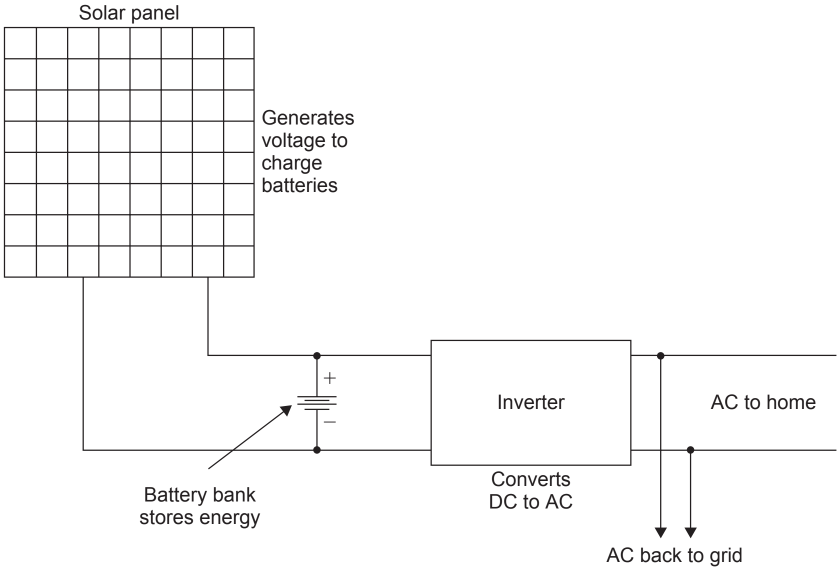
Solar panel
Generates voltage to charge batteries
+
−
Battery bank stores energy
Inverter
Converts DC to AC
AC to home
AC back to grid

FIGURE 4.23 Common DC-to-AC solar power system.

Project 4.1

Build a Kit

If you are really serious about electronics, then one of the first things you should do is build a kit. A kit, in case you really do not know, is a complete package of electronic parts to make a specific product. You build it yourself by placing components on a printed circuit board and soldering them in place. There is no better way to see how electronic products get put together than to handle the wide variety of parts. It is a great learning experience.

There are two main questions to answer about this project. First, what kind of kit do you want? Second, can you learn to solder? You will have to answer the first question, but as for the second, yes, you can learn to solder. It's not that hard. You can learn in Chapter 14.

One way to get an overview of the kits available is to obtain print catalogs (if available) from the sources listed below. You can order a print copy online and see the whole range of kits available. Go to:

Adafruit: www.adafruit.com
All Electronics: www.allelectronics.com
Digi-Key: www.digikey.com
Jameco: www.jameco.com
Mouser: www.mouser.com
SparkFun: www.sparkfun.com
Velleman: www.velleman.com

What I recommend is to get a kit related to one of the topics in this chapter. A power supply is a great choice, and it will give you a variable DC voltage source that you can use in future projects. I recommend the Jameco dual variable supply

Project 4.1—cont'd

kit 20626. Most of the companies have lots of radio- and wireless-related kits, but save these for the later chapters on wireless products.

Don't forget to get a small soldering iron (25 W or so) with a small tip and some solder. You will also need a few hand tools such as a wire stripper, needle-nose pliers, side cutters, and a small screwdriver, if you do not already own these. See Chapter 14.

Project 4.2

Getting Familiar With Virtual Instruments and the Oscilloscope

To really appreciate how electronic circuits work, you need a way to look at the voltages inside and outside circuits. To do that, you need an oscilloscope (scope). This is a test instrument that displays signal voltages, analog or digital, versus time on a screen. It allows you to measure voltage, frequency and time on the display. Using the probes supplied you can observe input and output signals and unique waveforms going on inside the circuit. Scopes are ideal for design and testing or for troubleshooting and repair.

Oscilloscopes are relatively expensive bench instruments. However, today you can have an oscilloscope for a few hundreds of dollars. It is part of a product known as a virtual instrument (VI). A VI is a small package of hardware that connects to the signals to be observed plus a PC or laptop computer. The laptop runs software that implements the oscilloscope functions and the LCD video monitor is the display.

Some VI products are just an oscilloscope such as the one shown in Fig. 4.24. Another version of this product also contains a signal generator that produces

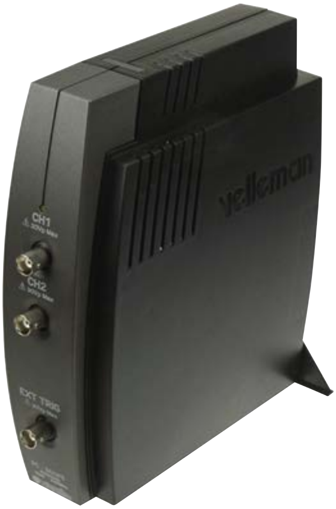

FIGURE 4.24 The Velleman PCSU1000 oscilloscope. It has dual channels for observing two signals at once and a bandwidth of 60 MHz. *Photo courtesy of Velleman.*

Continued

Project 4.2—cont'd

variable frequency sine, square, and other waves you can use as inputs to test circuits and equipment.

Another type of VI is more of a complete instrument package. In addition to the scope and generator, these units also contain DC power supplies and other instruments such as a digital multimeter (DMM). Software implements the instruments, and the laptop displays the output and controls. Two such units are shown in Fig. 4.25. On the left (A) is National Instruments' myDAQ. On the right (B) is the Digilent Analog Discovery 2. Both contain a scope, generators and power supplies. For more details on these and the Velleman units, go to the manufacturers' websites. Other similar products are available.

Digilent: www.diligent.com
National Instruments; www.ni.com
PicoScope: www.picotech.com
Velleman: www.velleman.com

If you are going to be a serious hobbyist or DIY maker, it is strongly recommended that you acquire one of these units. The projects that follow assume that you have one of these units or the equivalent.

To wrap up this project, do the following:

1. Buy one of the units suggested above or an equivalent.
2. Install the unit and software.
3. Read the instruction manual.
4. Demonstrate the scope function by observing the sine output of the internal signal generator.

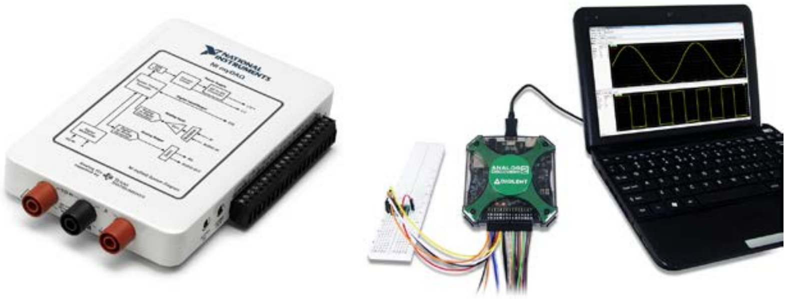

FIGURE 4.25 Virtual instrument units. Left: National Instruments' myDAQ, Right: Digilent Analog Discovery 2. *Photos courtesy of National Instruments and Digilent.*

Project 4.3

Build a Power Supply

You will need two power supplies for the projects that follow in this and later chapters. If you built a power supply kit as recommended in Project 4.1, you are all set. Or you can use batteries. The four AA cell source you built earlier is excellent but you will need a second one for some projects. You can also use the 9-V battery. Get a second 9-V battery and clip for some projects. Another choice is to get one of the wall adapter supplies. Choose the AC to DC type with 5- or 9-V output. The regulated versions are best and have a cleaner output. The current rating should be 0.2 A or more. See the parts list in Appendix B.

If you acquired one of the VI products such as the myDAQ or Analog Discovery 2, you can probably use the internal supplies. The circuits to be demonstrated will indicate the voltage options to use.

For this project, do the following:

1. Acquire what you need.
2. Test the supplies with your DMM (measure output voltages).

Project 4.4

Build a Signal Source

For some projects you will need a signal source to use as an input signal. A sine wave is best. Some projects such as the digital circuits in Chapter 5 need a rectangular/square wave signal. If you bought one of the VI units with a signal generator, you are ready to go. If not, here are a couple of options.

An easy choice is to use the 60-Hz or 50-Hz power line AC as a signal. Use a transformer to step the voltage down to a safe lower level then use a potentiometer to adjust the output to a suitable value. A safe choice is to use an AC to AC wall adapter transformer with a 5-V output. Cut off the connector on the cable then strip the wires and tin them with solder so they plug into the breadboard. (Note: You may need to solder the transformer wires to solid #22 hookup wire if they do not plug into the breadboarding socket holes.) Connect the wires to a 1-kΩ potentiometer as a voltage divider. See Fig. 4.26. See the parts list in Appendix B.

A more complex generator is the one shown in Fig. 4.27. It uses the 555 timer IC. This popular chip generates rectangular pulses at a frequency set by resistors R_1 and R_2 and C_1. You can build this circuit on your breadboarding socket. The parts list in Appendix B.

Before you tackle this project, look at Fig. 4.28 which is a review of breadboard specifics. Look this over carefully before you build the circuit. Keep in mind that there is no specific layout for a circuit. You build it up as you go along. For this circuit, build it on one end of the breadboarding socket so that you will have room for other circuits to be built later. Here are some other guidelines:

- Use #22 solid insulated hookup wire for direct point to point connections.
- Integrated circuits are to be mounted so that they straddle the center area between the two groups of holes. See Fig. 4.28.

Continued

Project 4.4—cont'd

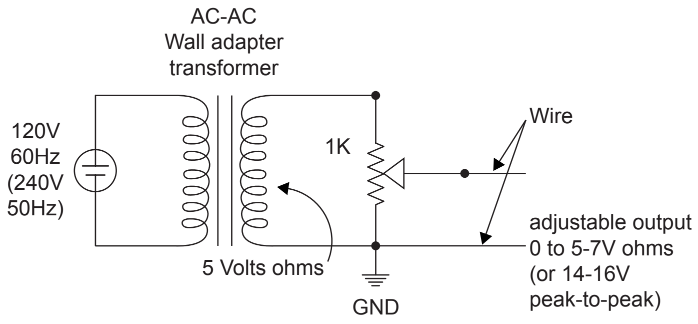
AC-AC
Wall adapter
transformer
120V
60Hz
(240V
50Hz)
1K
Wire
adjustable output
0 to 5-7V ohms
(or 14-16V
peak-to-peak)
5 Volts ohms
GND

FIGURE 4.26 An AC–AC wall adapter transformer and a pot form a variable 60 (50) Hz signal source.

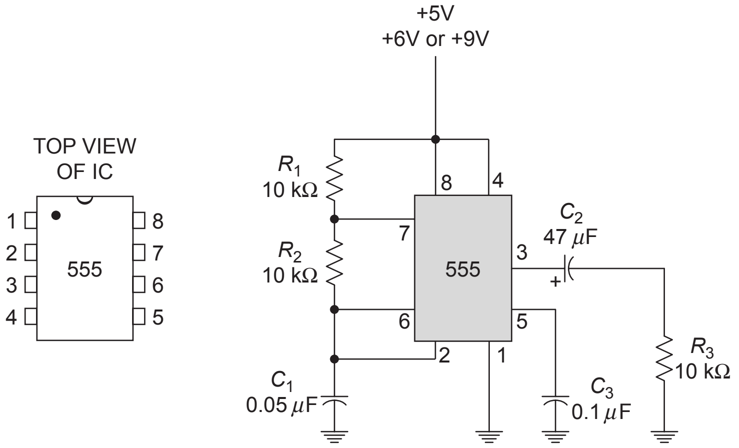
+5V
+6V or +9V
TOP VIEW
OF IC
1
2
3
4
8
7
6
5
555
R1
10 kΩ
R2
10 kΩ
C1
0.05 μF
8
4
7
3
6
5
2
1
555
C2
47 μF
+
C3
0.1 μF
R3
10 kΩ

FIGURE 4.27 A 555 timer IC circuit that generates rectangular waves.

- Be sure to pay attention to the IC orientation and pin numbers. See Fig. 4.27 for details.
- Some capacitors such as the electrolytics and tantalums are polarity sensitive. Make sure the + or – leads are wired as the schematic shows or damage will result.
- If you are still unsure about how to build the circuit on the breadboard, do an internet search on "how to use a breadboard." You will find YouTube videos and many other tutorials to help you. However, the best way to do this is just dive in and give it a try correcting as you go along.

Once you have built one or more of the circuits, you will test them in Project 4.5.

Project 4.4—cont'd

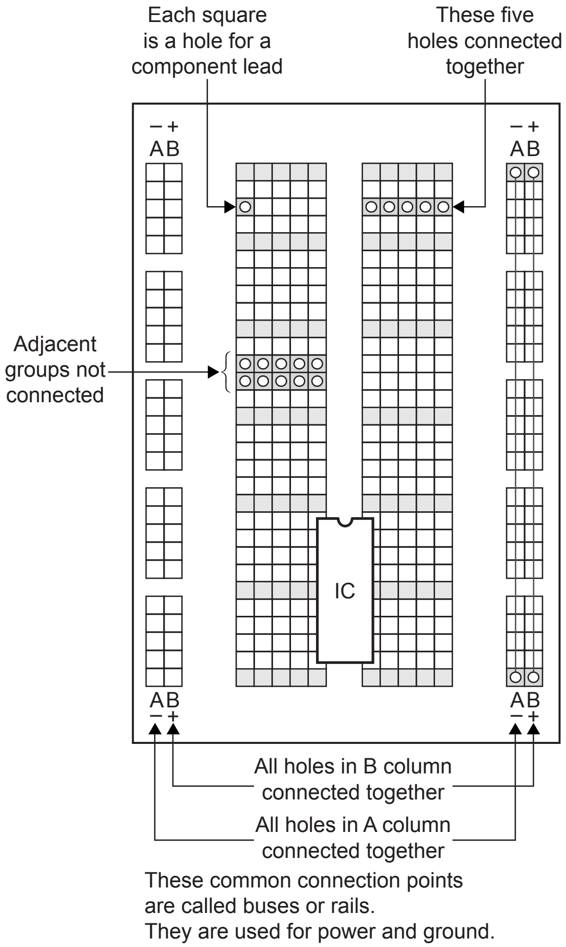
Each square is a hole for a component lead
These five holes connected together
− +
A B
− +
A B
Adjacent groups not connected
IC
A B
− +
A B
− +
All holes in B column connected together
All holes in A column connected together
These common connection points are called buses or rails.
They are used for power and ground.

FIGURE 4.28 One type of breadboarding socket. Not all sockets have the A and B power and ground buses in which case you can designate any one group of 5 pins to power and another for ground.

Project 4.5

Testing Circuits With an Oscilloscope

Before you test the circuits you built in the previous project, you need to familiarize yourself with the oscilloscope itself. Read the manual first. Then power up the scope and be sure the screen is displayed on the video monitor. The scope screen should look something like that shown in Fig. 4.29. The screen is a grid of horizontal and vertical lines. The horizontal scale is time and the vertical scale is volts. Each of the grid's major segments represents a calibrated value. Horizontal grid increments are usually in milliseconds or microseconds per division such as 5 milliseconds (ms) per division. The vertical grid is calibrated in millivolts or volts. For example, one setting may be 5V per division. See Fig. 4.29 for clarification. The scope lets you change these settings so you can adjust it to view a wide range of signals. Check the scope instructions to be sure you know how to do this. Because the grid is calibrated you can measure the time period (t) of a

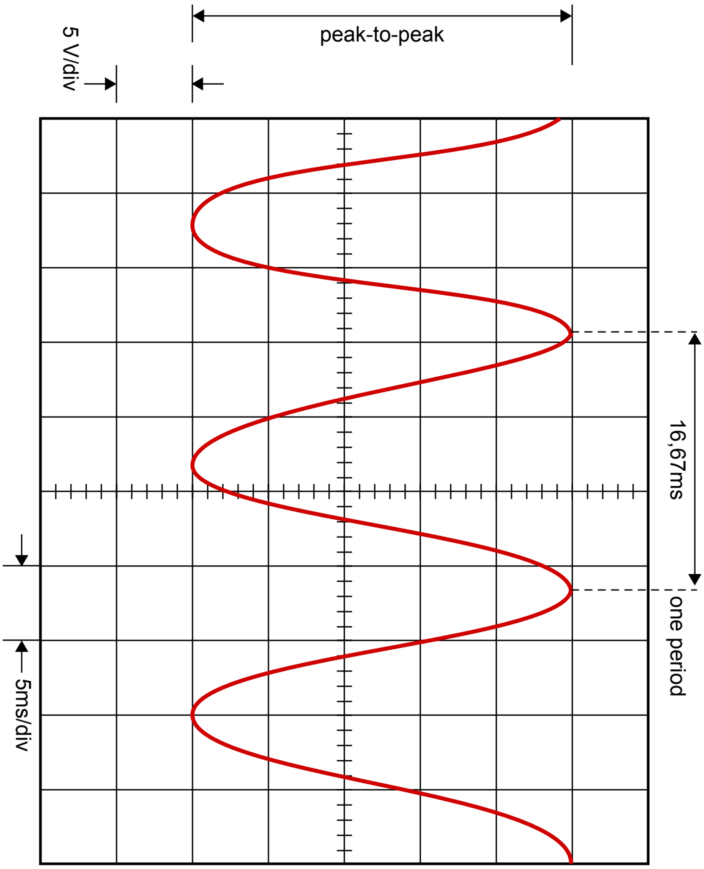

FIGURE 4.29 A 60-Hz sine wave displayed on an oscilloscope.

Project 4.5—cont'd

signal and its voltage amplitude. Remember if you know the period of a signal (t), you can calculate the frequency (f) with the formula $f=1/t$. Now on to the testing.

If you built the simple signal source as shown in Fig. 4.26, get it ready and turn on your oscilloscope, and set the horizontal and vertical scales to those shown in Fig. 4.29. Follow these steps:

- Plug the wall adapter transformer into an AC outlet.
- Connect short hookup wires (2 or 3 inches) to the pot on the breadboarding socket. One wire is connected to ground, and the other to the slider arm of the pot (usually the center pin). These wires will let you connect the scope probe inputs to the circuit. See Fig. 4.26. Some VIs may have other arrangements so follow the guidelines in the VI manual.
- Rotate the pot knob to the maximum clockwise (CW) position.
- Connect the scope to the pot output. You should see a sine wave as shown in Fig. 4.29. Make adjustments as needed to see all of the signal.
- If you do not see a signal, rotate the pot to the maximum counter clockwise position (CCW).
- Now measure the period of the sine wave between two successive peaks. Each major division is 5 ms so each smaller increment is 1 ms. For a 60-Hz sine wave the period is 16.67 ms. You should be able to estimate that from the display.
- Next measure the peak-to-peak amplitude of the signal. This is the voltage between the maximum positive peak and maximum negative peak. If you bought a 5-V transformer, the peak should be 14–16V. The 5-V rating is an rms value and the peak-to-peak value 2.828 or more times the rms value.
- Now vary the pot knob from maximum CW to CCW positions and note how the sine wave amplitude varies from zero to its maximum value. This lets you set the voltage as you need.
- Unplug the transformer.

If you built the circuit in Fig. 4.27, check your wiring and get ready for the test. Do this:

- Apply DC power. The circuit is flexible so you can use what DC voltage you have from about 4.5–15V. Your 5-, 6-, or 9-V supplies will work ok.
- Set up the scope. Some recommended settings are 0.5 or 1V per division vertically and 200 or 500 ms per division horizontally.
- Connect the scope input to the top of the 10-k resistor R3 and ground. You should see a signal like that in Fig. 4.30. This figure assumes a power supply voltage of 6V. Adjust the scope setting to display the signal as in the figure.
- Measure the period and peak-to-peak values. The p–p value should be a bit less than the DC supply voltage. With resistor and capacitor values shown, the frequency computes to 962 Hz. That is close to 1000 Hz or 1 kHz. The period should be little bit more than 1 ms between consecutive positive or negative going edges of the pulses.
- Disconnect the power.

Continued

Project 4.5—cont'd

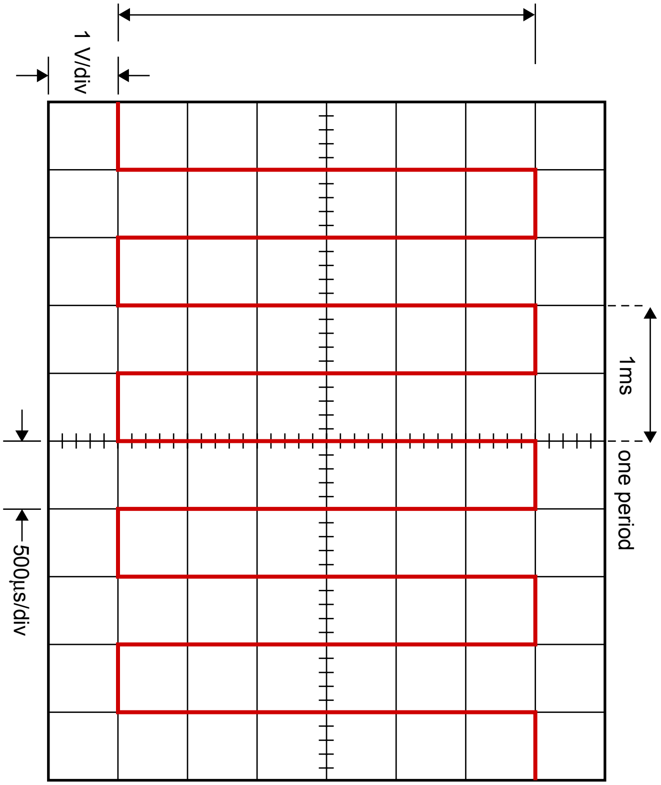

FIGURE 4.30 The output of the 555 timer circuit displayed on an oscilloscope.

A Word About Troubleshooting

If your circuit does not work, you will need to troubleshoot it. Here are a few suggestions.

- The problem is most likely a wiring error. Check all of your connections carefully. Rewire as needed. A common mistake is plugging a component leads into the wrong adjacent hole. Refer back to Fig. 4.28 for a review.
- Be sure that the right part is in the right position. Double check resistor color code values. If polarized capacitors are used, be sure the + and – leads are in the designated position.
- Use your DMM to see that the circuit is getting DC power. Measure between pins 1 and 8.
- If the 555 timer circuit of Fig. 4.27 does not work again, check the wiring. Try rewiring according to that in Fig. 4.28 and try testing again.
- If available, a knowledgeable friend or acquaintance may be able to help.

Project 4.6

Experimenting With the 555 Timer

The 555 timer IC is a very versatile device. Here are several different uses to try.

- Remove the 10-kΩ resistor R3 from the circuit in Fig. 4.27 and replace it with an 8-Ω speaker and 100-Ω resistor. See Fig. 4.31.
- Apply DC power. You will hear the 962-Hz audio tone in the speaker.
- Turn off the power.
- Rewire the circuit so that it is like that shown in Fig. 4.32. You will replace the 0.05-µF capacitor with the 47-µF capacitor. Be sure to observe the correct polarity. Then connect the 220-Ω resistor and LED as shown to replace the speaker. Again be sure to orient the LED correctly as shown.
- Apply power. The LED should blink at an approximate rate on once per second.
- Turn off the power.
- Turn off the power but do not disassemble the 555 circuit as you will use it in Chapter 5.

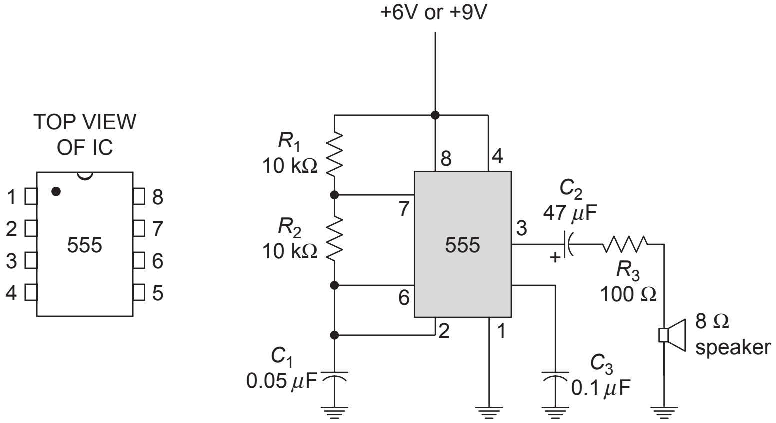
+6V or +9V
TOP VIEW
OF IC
1
2
3
4
555
8
7
6
5
R_1
10 kΩ
R_2
10 kΩ
C_1
0.05 μF
8
4
7
555
6
2
1
3
C_2
47 μF
+
R_3
100 Ω
8 Ω
speaker
C_3
0.1 μF

FIGURE 4.31 The 555 IC connected as an audio tone generator with speaker.

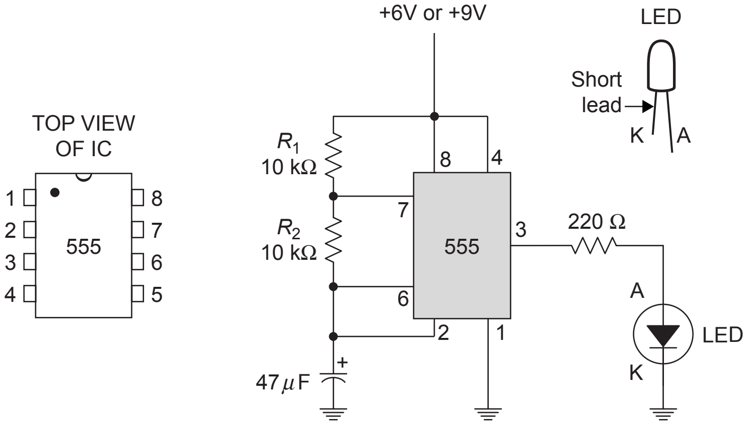
+6V or +9V
LED
Short
lead
K
A
TOP VIEW
OF IC
1
2
3
4
555
8
7
6
5
R_1
10 kΩ
R_2
10 kΩ
+
47 μF
8
4
7
555
6
2
1
3
220 Ω
A
LED
K

FIGURE 4.32 The 555 IC connected to flash an LED off and on.

Project 4.7

Building a Half Wave Rectifier and Power Supply

A rectifier is a diode that converts AC into pulsating DC. Rectifiers are a part of most power supplies. Here you will demonstrate how a diode produces half sine pulses that can then be filtered into a steady DC. By adding an IC voltage regulator you will create a useful power supply you can use in later projects.

- Remove the 1-k pot from the wall adapter transformer in Fig. 4.26 and wire the circuit in Fig. 4.33. Add only the 1N4001 diode and 1-k-Ω resistor at this time.
- Plug the transformer into an AC outlet. Connect the oscilloscope to the 1-k resistor output. You should see half sine pulses like that in Fig. 4.34. The peak pulse value should be about 7–9V with a 5V transformer.
- Measure the voltage across the 1-k resistor with your DMM. It is the DC average voltage of about 2–4V with the 5V transformer.
- Now connect a 1000 μF capacitor across the 1 k resistor as shown in Fig. 4.33. Be sure to get the polarity right as shown.
- Measure the voltage across the capacitor with your DMM. This is the DC average voltage with peaks of about 7–9V with the 5-V transformer. The capacitor filters out the pulses into a steady DC.
- When you connected the 1000 μF, you noticed that the pulses went away on the scope screen. There will still be some minor variations on the output called ripple. It will be only a few millivolts or less. You can see it on the scope if you change the vertical sensitivity.

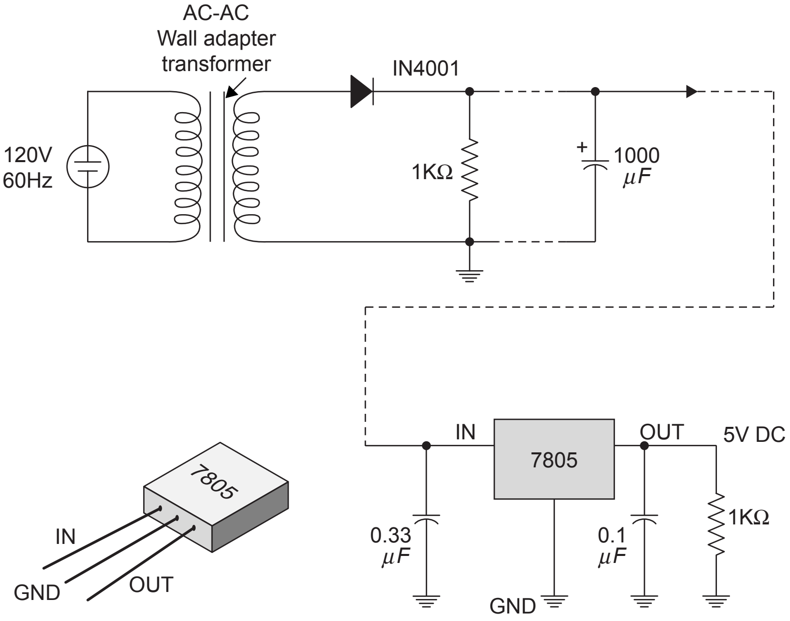

FIGURE 4.33 A half wave rectifier and regulated 5-V power supply.

Project 4.7—cont'd

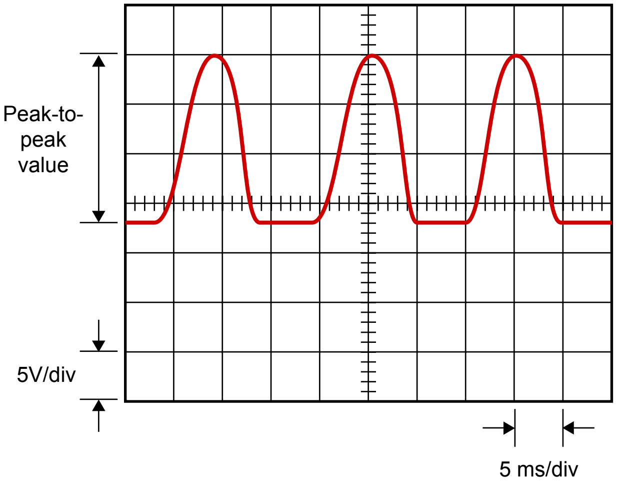

FIGURE 4.34 The half sine pulses as seen on the oscilloscope.

- Next, connect a 7805 IC regulator to the output as shown in Fig. 4.33. This IC converts the voltage across the 1000-μF capacitor into a stable and accurate 5V.
- Measure the output across the 1-k-Ω resistor with your DMM.
- Unplug the transformer.

Project 4.8

Experimenting With Operational Amplifiers

An op amp is an all-purpose amplifier that can be connected in a variety of ways to meet many amplification needs. Most op amps require two power supplies, one positive and the other negative. With two supplies the output voltage swing can be positive and negative. Fig. 4.35 shows the most common op amp configuration. This is an inverting amplifier. See Fig. 4.8 and the related text for a review. Note how the power supplies are connected to pins 7 and 4. You can use two 5-volt supplies, the 6-V AA cell supplies, or two 9-V batteries or the supplies in your VI device. The junction of the two supplies is the circuit ground.

Wire the circuit of Fig. 4.35 on the breadboard. The 1-k pot is connected to the positive supply to provide a variable DC voltage to use as an amplifier input. Double check your wiring before you go to the procedure below.

- Apply voltage to the circuit. Use your DMM to check for voltages on pins 7 and 4 on the IC.
- Measure the pot output voltage and vary the pot to set its output to 1V. Connect the +1V to the op amp as shown in the figure.

Continued

Project 4.8—cont'd

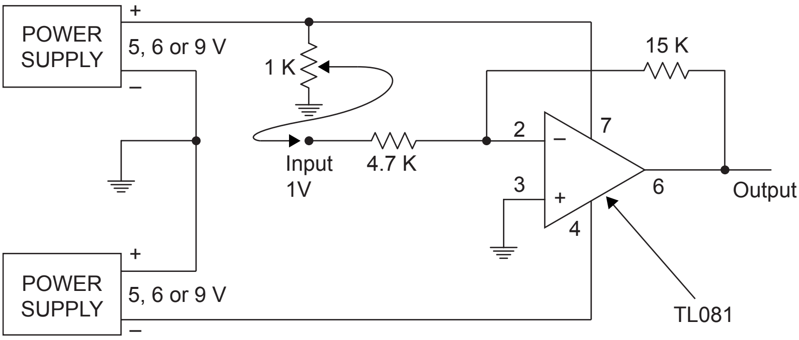

FIGURE 4.35 Operational amplifier inverter circuit showing power supply detail.

- Using the resistor values given in the figure, calculate the circuit gain.
- Calculate the value of the output voltage given the 1-V input.
- Now measure the op amp output at pin 6 with your DMM. You should get −3.2V. The circuit gain is 15 k/4.7 k = 3.2 so with 1V in you will get −3.2V out. The voltage is negative because the op amp inverts the + input. Op amps can amplify DC and well as AC signals.
- Turn off the power and rewire the circuit as shown in Fig. 4.36. This is a non-inverting amplifier.
- Use the formula in Fig. 4.8 to calculate the gain using the resistor values given (see Fig. 4.8).
- Connect an input signal to the op amp at pin 3. Use the 60-Hz output from the circuit in Fig. 4.26. Disconnect the wall transformer from the power supply you built in Project 4.7 and connected to the 1-k ohm pot as shown in Figs. 4.26 and 4.36.
- Using the scope, set the output of the 1-k pot to 1V peak-to-peak. Connect the pot output to pin 3 of the op amp. Now look at the output at pin 6 of the op amp with the scope. Since the calculated gain is 6.8 k/1 k + 1 = 7.8, the output should be 7.8V peak-to-peak.
- While observing the output on pin 6 of the op amp, rotate the 1-k pot to increase the input voltage. At some point, you will see the output sine wave increase until the output saturates and clips creating a square wave. The output voltage cannot be greater than the power supply voltages.
- Disconnect power.

Project 4.8—cont'd

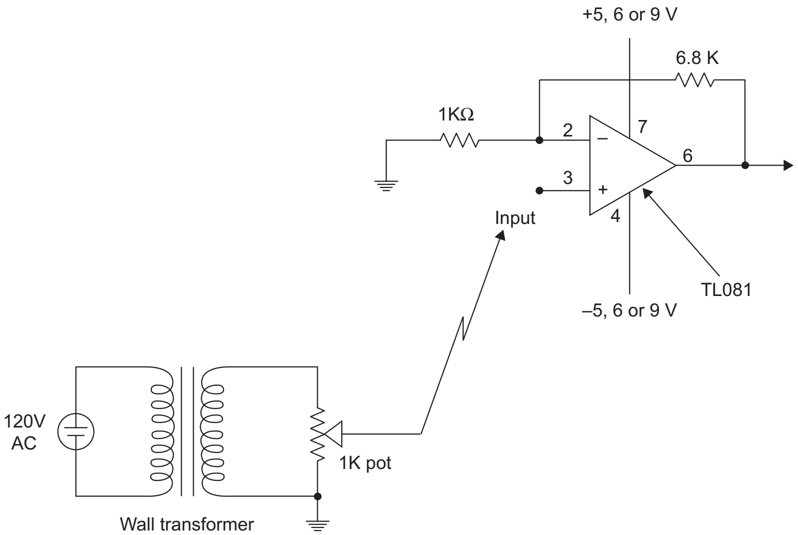

FIGURE 4.36 Operational amplifier noninverting amplifier circuit.

Project 4.9

Try Out a Circuit Simulator

A circuit simulator is a software program that lets you build an electronic circuit and test it on your computer. A schematic capture program helps you construct the circuit on screen using built-in component symbols and wiring. The software then converts that circuit into a sequence of operations that simulate the various components and the operation of the circuit. Built-in waveform viewers and read out capability lets you see how the circuit works and performs. Electronic engineers regularly use simulators to help design circuits and evaluate them before constructing physical prototypes and before committing to an expensive IC design. Some simulators are accompanied a PCB layout program that helps you design a PCB for your circuit.

Circuit simulators require some learning time but they are worth the effort. You can also substitute a circuit simulator for the actual hands-on projects in this book, although that is not recommended. Use the simulators to complement the real lab work projects.

Listed below are several circuit simulators to try. Some are free while others require payment to access.

https://circuits.io/
https://www.circuitlab.com/editor/#?id=7pq5wm

Continued

Project 4.9—cont'd

http://dcaclab.com/en/home
http://www.docircuits.com/circuit-editor
http://everycircuit.com/
http://www.partsim.com/simulator
http://www.schematics.com/

Recommended action:

- Try out one or more circuit simulators listed. Build and test at least one circuit.
- Implement one or more of the hardware projects you experimented within this chapter. The 555 timer, half wave rectifier and op amp projects are suggested.
- Compare your hands-on results with the simulated results.

Chapter 5

Electronic Circuits: Digital

Practically Everything Is Digital These Days

In this Chapter:

- Binary number system.
- Logic gates and flip-flops.
- Common combinational and sequential circuits.
- Semiconductor memory.
- Programmable logic devices.
- Analog-to-digital and digital-to-analog conversion.

INTRODUCTION

Recall that there are two basic types of electronic signals—analog and digital. A digital signal is one that varies in discrete steps. Unlike an analog signal, which varies continuously, a digital signal has two levels or states. The signal switches or changes abruptly from one state to the other.

Fig. 5.1 shows a DC digital signal that switches between two known levels such as 0V and close to it (0.1V) or 0 and 3.3V. The positive voltage can be anything between about 1 and 12V with 3.3 and 5V being the most common.

Digital signals with two discrete levels are also referred to as *binary signals. Binary* means two—two states or two discrete levels of voltage.

Humans use the decimal number system that represents quantities with digits 0 through 9. However, digital equipment and computers do not. Internally, digital equipment processes binary data.

BINARY NUMBERS

The binary number system is a set of rules and procedures for representing and processing numerical quantities in digital equipment. Because the base of the binary number system is 2, only two symbols (0 and 1) are used to represent any quantity. The symbols 0 and 1 are called *binary digits* or *bits*. For example, the 6-bit number 101101 represents the decimal number 45. Your understanding of

Electronics Explained. http://dx.doi.org/10.1016/B978-0-12-811641-8.00005-9
Copyright © 2018 Elsevier Inc. All rights reserved.

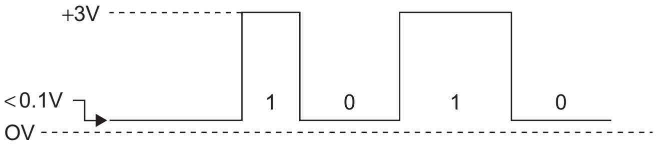

FIGURE 5.1 Binary signal that represents 0 and 1 as voltage levels.

how digital circuits, microprocessors, and related equipment process data is tied directly to an understanding of the binary number system.

The reason for using binary numbers in digital equipment is the ease with which they can be implemented. The electronic components and circuits used to represent and process binary data must be capable of assuming two discrete states to represent 0 and 1. Examples of two-state components are switches and transistors. When a switch is closed or on, it can represent a binary 1. When the switch is open or off, it can represent a binary 0. A conducting transistor may represent a 1, whereas a cutoff transistor may represent a 0. The representation may also be voltage levels. For example, a binary 1 may be represented by +3.3 V and a binary 0 by 0 V as previously shown in Fig. 5.1.

Binary-to-Decimal Conversions

The binary system is similar to the decimal system in that the position of a digit in a number determines its weight. Recall that in the decimal system the weights are powers of 10. The rightmost digit is units or 1s and then 10s, 100s, 1000s, and so on, moving from right to left from one digit to the next. The position weights of a binary number are powers of 2:

$2^0 = 1$
$2^1 = 2$
$2^2 = 4$
$2^3 = 8$
$2^4 = 16$
$2^5 = 32$
$2^6 = 64$
$2^7 = 128$
$2^8 = 256$

The position weights of an 8-bit binary number are as follows:

Bit position	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
Weight	128 (MSB)	64	32	16	8	4	2	1 (LSB)

The most significant digit or bit (MSB) and the least significant bit (LSB) are identified.

Now let us evaluate the decimal quantity associated with a given binary number, 101101. You simply multiply each bit by its position weight and add the values to get the decimal equivalent, 45.

Position weight	32	16	8	4	2	1
Binary number	1	0	1	1	0	1

1 × 1 = 1
0 × 2 = 0
1 × 4 = 4
1 × 8 = 8
0 × 16 = 0
1 × 32 = 32
Decimal equivalent 45

You can see that positions with a 0 bit have no effect on the value. Therefore, they can be ignored. To quickly determine the decimal equivalent of a binary number, simply sum the weights of the positions containing a 1 bit. For example, in the number 11101, the weights of those positions with a 1 bit from right to left are 1 + 4 + 8 + 16 = 29.

Using Hardware to Represent Binary Quantities

Switches are widely used to enter binary data into computers and digital equipment. If the switch is set to the up position, a binary 1 is represented. If the switch is down, a binary 0 is represented. The switch contacts can be opened or closed depending on the circuit in which the switch is connected. Fig. 5.2 shows a group of slide switches set to represent a binary number, 11000101 or 197.

Indicator lights such as light-emitting diodes (LEDs) are sometimes used to read out binary data in digital equipment. An “on” light is a binary 1 and an “off” light is a binary 0. See Fig. 5.3. The decimal value being represented by the display is 178.

Decimal-to-Binary Conversions

You may find it necessary to convert decimal numbers into their binary equivalents. This can be done by dividing the decimal number by 2, dividing the quotient by 2, dividing that quotient by 2, and so on, recording the remainders until the quotient is zero. The remainders form the binary equivalent.

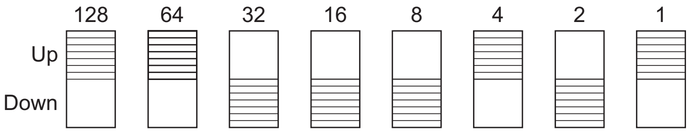

FIGURE 5.2 Off/on switches are used to enter binary data into digital circuits.

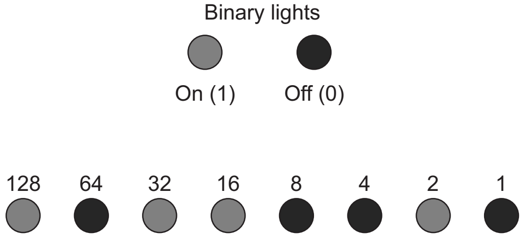

FIGURE 5.3 How lights or LEDs are used to represent binary data.

In the following example, we convert the decimal number 57 to its binary equivalent.

Quotient	**Remainder**
57 ÷ 2 = 28	1 (LSB)
28 ÷ 2 = 14	0
14 ÷ 2 = 7	0
7 ÷ 2 = 3	1
3 ÷ 2 = 1	1
1 ÷ 2 = 0	1 (MSB)

You can always check your work by reconverting the binary number back to decimal using the procedure described earlier. Also, keep in mind that some scientific calculators can do binary-to-decimal and decimal-to-binary conversions.

Another name for a binary number is *binary word*. The term *word* is more general. It refers to a fixed group of bits that can mean numbers, letters, or special characters and codes. We say that digital equipment processes *binary data words*.

All digital circuits and microcomputers work with a fixed-length binary word. A common binary word length in microcomputers and other digital equipment is 8 bits. All data storage, processing, manipulation, and transmission are carried out in 8-bit groups. Word lengths of 4, 8, 12, 16, 32, and 64 bits are common in digital equipment.

An 8-bit word is usually called a *byte*. In data communication applications, an 8-bit word is sometimes called an *octet*. A 4-bit word is sometimes referred to as a *nibble*.

Maximum Decimal Value for N Bits

The number of bits in a binary word determines the maximum decimal value that can be represented by that word. This maximum value is determined with the formula:

$$M = 2^N - 1$$

where M is the maximum decimal value and N is the number of bits in the word.

TABLE 5.1 Number of Bits in Binary Number and Maximum Number of States

Number of Bits	Maximum States
8	256
12	4096 (4 K)
16	65,536 (64 K)
20	1,048,576 (1 M)
24	16,777,216 (16 M)
32	4,294,967,296 (4 G)

G, gigabits = 1,0737,41,824; *K*, kilobits = 1024; *M*, megabits = 1,048,576.

For example, what is the largest decimal number that can be represented by 4 bits?

$$M = 2^N - 1 = 16 - 1 = 15$$

With 4 bits, the maximum possible number is binary 1111 or decimal 15.

The maximum decimal number that can be represented with 1 byte is 255 or 11111111. An 8-bit word greatly restricts the range of numbers that can be accommodated. But this is usually overcome by using larger words.

There is one important point to know before you leave this subject. The formula $M = 2^N - 1$ determines the maximum decimal quantity (M) that can be represented with a binary word of N bits. This value is 1 less than the maximum number of values that can be represented. The maximum number of values that can be represented (Q) is determined with the formula:

$$Q = 2^N$$

With 8 bits, the maximum number of values is 256 or 0 through 255.

Table 5.1 gives the number of bits in a binary number and the maximum number of states that can be represented.

BCD and ASCII

The binary numbers we have been discussing are usually referred to as pure binary codes. But there are other types of binary codes. For example, the binary-coded decimal (BCD) system is popular. BCD is a hybrid code between the binary and decimal systems. It was developed in an attempt to simplify the conversion processes between the two systems and to improve human–machine communication.

TABLE 5.2 Binary-Coded Decimal

Decimal Value	BCD Code
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001

To represent a decimal number in the BCD system, each digit is replaced by its 4-bit binary equivalent. Thus, the number 729 in BCD is:

7 2 9

0111 0010 1001

The BCD code is given in Table 5.2.

It is important to note that the 4-bit binary numbers 1010 through 1111 that represent the decimal values 10 through 15 are invalid in BCD.

To convert a BCD number to decimal, simply substitute the decimal equivalent of each 4-bit group. The BCD number 1001 0100 0110 in decimal is 946.

A special form of BCD code is used in computers and data communications systems. It is a 7- or 8-bit code that is used to represent not only numbers but also letters (both uppercase and lowercase), special symbols, and control functions. This code is called the American Standard Code for Information Interchange, or ASCII (pronounced “ask-key”).

The following are examples of ASCII:

Number, letter, or symbol	ASCII
F	01000110
8	00111000
J	01101010
+	00101011

The complete ASCII is given in Table 5.3. It provides the decimal and hexadecimal (hex) values of the codes. The code contains both uppercase and lowercase letters, the digits 0 through 9, punctuation marks, common mathematical

TABLE 5.3 The ASCII Character Code

Char	Dec	Oct	Hex	Char	Dec	Oct	Hex	Char	Dec	Oct	Hex	Char	Dec	Oct	Hex
(nul)	0	0000	0300	(sp)	32	0040	0320	@	64	0100	0340	′	96	0140	0360
(soh)	1	0001	0301	!	33	0041	0321	A	65	0101	0341	a	97	0141	0361
(stx)	2	0002	0302	“	34	0042	0322	B	66	0102	0342	b	98	0142	0362
(etx)	3	0003	0303	#	35	0043	0323	C	67	0103	0343	c	99	0143	0363
(eot)	4	0004	0304	$	36	0044	0324	D	68	0104	0344	d	100	0144	0364
(enq)	5	0005	0305	%	37	0045	0325	E	69	0105	0345	e	101	0145	0365
(ack)	6	0006	0306	&	38	0046	0326	F	70	0106	0346	f	102	0146	0366
(bel)	7	0007	0307	′	39	0047	0327	G	71	0107	0347	g	103	0147	0367
(bs)	8	0010	0308	(	40	0050	0328	H	72	0110	0348	h	104	0150	0368
(ht)	9	0011	0309	)	41	0051	0329	I	73	0111	0349	i	105	0151	0369
(nl)	10	0012	030a	*	42	0052	032a	J	74	0112	034a	j	106	0152	036a
(vt)	11	0013	030b	1	43	0053	032b	K	75	0113	034b	k	107	0153	036b
(np)	12	0014	030c	,	44	0054	032c	L	76	0114	034c	l	108	0154	036c
(cr)	13	0015	030d	2	45	0055	032d	M	77	0115	034d	m	109	0155	036d
(so)	14	0016	030e	.	46	0056	032e	N	78	0116	034e	n	110	0156	036e
(si)	15	0017	030f	/	47	0057	032f	O	79	0117	034f	o	111	0157	036f
(dle)	16	0020	0310	0	48	0060	0330	P	80	0120	0350	p	112	0160	0370
(dc1)	17	0021	0311	1	49	0061	0331	Q	81	0121	0351	q	113	0161	0371
(dc2)	18	0022	0312	2	50	0062	0332	R	82	0122	0352	r	114	0162	0372
(dc3)	19	0023	0313	3	51	0063	0333	S	83	0123	0353	s	115	0163	0373

Continued

TABLE 5.3 The ASCII Character Code—cont'd

Char	Dec	Oct	Hex	Char	Dec	Oct	Hex	Char	Dec	Oct	Hex	Char	Dec	Oct	Hex
(dc4)	20	0024	0314	4	52	0064	0334	T	84	0124	0354	t	116	0164	0374
(nak)	21	0025	0315	5	53	0065	0335	U	85	0125	0355	u	117	0165	0375
(syn)	22	0026	0316	6	54	0066	0336	V	86	0126	0356	v	118	0166	0376
(etb)	23	0027	0317	7	55	0067	0337	W	87	0127	0357	w	119	0167	0377
(can)	24	0030	0318	8	56	0070	0338	X	88	0130	0358	x	120	0170	0378
(em)	25	0031	0319	9	57	0071	0339	Y	89	0131	0359	y	121	0171	0379
(sub)	26	0032	031a	:	58	0072	033a	Z	90	0132	035a	z	122	0172	037a
(esc)	27	0033	031b	;	59	0073	033b	[	91	0133	035b	{	123	0173	037b
(fs)	28	0034	031c	,	60	0074	033c	\	92	0134	035c	|	124	0174	037c
(gs)	29	0035	031d	5	61	0075	033d	]	93	0135	035d	}	125	0175	037d
(rs)	30	0036	031e	.	62	0076	033e	^	94	0136	035e	~	126	0176	037e
(us)	31	0037	031f	?	63	0077	033f	_	95	0137	035f	(del)	127	0177	037f

Char, character; *Dec*, Decimal; *Hex*, Hexadecimal (4-bit groups); *Oct*, Octal (3-bit groups).
Characters in parentheses are control codes for specific operations.
X, "Don't care" character
Example: ASCII capital letter A
Decimal 65
Binary 0000000001000001
Octal 0/000/000/001/000/001
0 0 1 0 1
Hex 0000/0000/0100/0001
0 x 4 1

symbols, and many special two- and three-letter codes. The special codes are known as control codes. They are used to indicate the start and end of a string of characters or to initiate special operations. Some examples are as follows:

STX—Start of text sequence
EOT—End of transmission of a string of characters
BEL—Ring a bell
DEL—Delete
LF—Line feed
CR—Carriage return

These codes not only represent alphanumeric data but can also make things happen in a printer or computer.

The main use of ASCII is in data communication. Computers use ASCII to "talk" to their peripheral units, such as printers, or to one another. A common example is a personal computer that receives data via a modem from the Internet. Another example is the data sent from a computer to a laser printer. ASCII is the communication code for text.

Hexadecimal Notation

Hexadecimal notation is a way to represent binary values with numbers and letters in an effort to make data easier to remember and to reduce the errors normally associated with writing or reading long strings of 1s and 0s. The hex code is just an extension of the BCD code, with the decimal values from 11 to 15 represented by the letters A through F. See the following table.

Decimal Value	Hex Code
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001
A (10)	1010
B (11)	1011
C (12)	1100
D (13)	1101
E (14)	1110
F (15)	1111

As an example to convert the binary number 101000110101100 to hex, you divide the word into 4-bit groups starting on the right and then replacing each with the hex codes. In this case, the hex code is A35C.

Parallel and Serial Data

It is frequently necessary for digital devices to communicate. Someone using a computer may have to send data to a printer. Or two people using personal computers may have to communicate over a local area network (LAN). To accomplish either of these objectives, one must transmit binary data from one circuit to another. This is done in one of two ways: parallel transmission or serial transmission.

Parallel Data

When the parallel method is used, all bits of a binary word or number are transmitted simultaneously. Fig. 5.4 shows an 8-bit word being transferred from logic circuit 1 to logic circuit 2.

A *bus* is a set of parallel data lines over which digital data is transferred. A data bus can have any number of data lines as required by the application. However, the main characteristic of a bus is parallel data transfer. In physical form, a bus is usually implemented on a connector, with lines of copper on a printed circuit board or with a ribbon cable with one wire per bit.

Because all bits are transmitted at the same time over a parallel bus, data movement is extremely fast. The main benefit of parallel data transmission is its high speed. With high-speed logic circuits, a binary word can be transferred from one point to another in as little as several nanoseconds. Parallel data transmission over buses in computers takes place at rates up to several hundred megahertz. It is not uncommon to see a rate of 100 million bits per second (Mbps) on a personal computer data bus. On long buses, data rates are slower because the stray inductance and capacitance of the cable typically distorts the data and slows the transmission speed.

Although parallel transmission is fast, it is expensive, because there must be circuitry for each bit on both the sending and receiving ends. There must be one wire per bit in a cable plus a separate ground wire. This increases the complexity and thereby the cost of the circuitry.

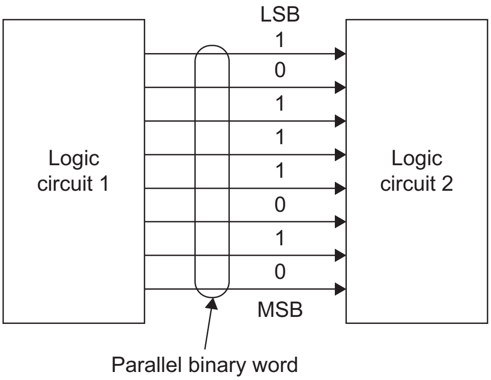

FIGURE 5.4 Parallel binary data transfer over bus.

Serial Data

The other method of moving binary data from one point to another is serial transmission. In the serial method, the bits of a word are transmitted one at a time, or sequentially. Fig. 5.5 shows the waveform of an 8-bit serial binary word. This is the voltage that you would monitor at the single output of the circuit generating the word. It is the waveform that an oscilloscope would display. Each bit occurs for a fixed time interval such as 1.5 ms. Thus, it takes 1.5 × 8 = 12 ms to transmit an 8-bit word.

The speed of data transmission is measured in bits per second (bps). That figure can be calculated from the expression:

$$\text{bps} = \frac{1}{t}$$

Here, t is the time for 1 bit. In Fig. 5.5, 1 bit takes 1.5 μs. The speed of the data, then, is

$$\text{bps} = 1 \div \left(1.5 \times 10^{-6}\right) = 666,666.7 \text{ bps or } 666.7 \text{ kbps}$$

If you know the speed in bps, you can calculate the bit time (t) by rearranging the formula:

$$T = \frac{1}{\text{bps}}$$

Assume a data rate of 14,400 bps. The bit time is 69.4 μs.

As seen in Fig. 5.5, the MSB is transmitted first. Depending on the system used, however, the LSB can be sent first. In any case, the speed of transmission depends on the number of bits in the word and the duration of each bit. It is this longer transmission time that is the primary disadvantage of the serial method. However, although serial transmission is slower than parallel transmission, it is perfectly suitable for many applications. Common data rates are in excess of millions of bits per second. Data rates of up to 100 Gbps are common.

The main benefits of the serial method are its simplicity and low cost. Only a single line rather than multiple lines is needed to interconnect the

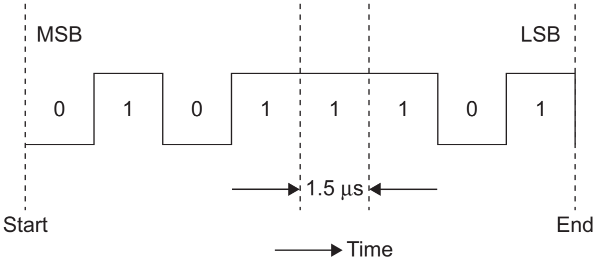

FIGURE 5.5 Serial data transfer where 1 bit occurs after another in time.

equipment. In addition, only one set of sending and receiving circuits is needed. All data transmissions by copper cable or fiber optical cable exceeding about 100 Mbps are made by the serial method. All wireless data transmissions are serial.

DIGITAL LOGIC ELEMENTS

Digital logic elements are the basic circuits that are used to process the binary data. The logic element has one or more binary data inputs to be processed. The logic element processes or manipulates the binary input signals in a fixed way and generates an appropriate output signal. The output is a function of the binary states of the inputs and the unique processing capability of the logic element. The logic element "looks" at the binary input signals and then makes a decision and generates an appropriate output. There are five elementary logic circuits: inverter or NOT, AND, OR, NAND, and NOR. These functions are performed by logic elements known as inverters and gates.

You need not be concerned with the inner workings of these logic elements. Instead, your primary concern is with the basic logic functions. Needless to say, the gates and inverters are inside every microcomputer and digital integrated circuit (IC).

Inverter

The inverter has a single input and a single output. The logic function performed by an inverter is *inversion*. The output of the inverter is simply the inverse, or opposite, of the input. Because binary signals can assume only one of two different states, either 0 or 1, the inverter generates a 0 when the input is a 1 and a 1 when the input is a 0.

Two common symbols used to represent an inverter are shown in Fig. 5.6. The input and output are given names, which are usually letters of the alphabet.

There are a variety of ways to express the operation of various logic functions. One way is to use Boolean algebraic expressions. A Boolean expression

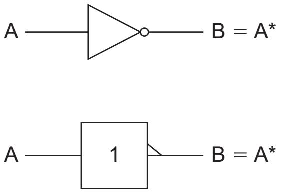

A	B
0	1
1	0

FIGURE 5.6 Logic symbols for inverter and truth table.

is a simple mathematical or algebraic formula that expresses the output in terms of the input. The Boolean expression for the inverter is:

$$B = A^*$$

This equation is read, "B equals not A." The asterisk means NOT or inversion. What this expression is telling you is that if the input is A, the output, B, is not A.

Another method of expressing the function of a logic element is a truth table. This is simply a table listing all possible combinations of the inputs and outputs of a logic element. Fig. 5.6 also shows the truth table for an inverter. There are only two possible inputs, 0 and 1, and the corresponding outputs. A truth table completely defines the operation of a logic element.

AND Gate

A gate is a circuit that has two or more inputs and a single output. The gate generates a binary output that is a function of the states of the inputs and the nature of the gate itself.

The type of gate determines how the binary inputs are processed. The two basic logic gates are the AND gate and the OR gate. An AND gate is a logic circuit that generates a binary 1 output if all of its inputs are binary 1s. Otherwise, the circuit generates a binary 0. All of the inputs have to be binary 1 for the AND gated to produce a binary 1 output. If any one of the inputs is a binary 0, the output is binary 0.

Fig. 5.7 shows the logic symbols used to represent an AND gate. Letters of the alphabet are used to represent inputs and outputs. There may be more than two inputs, but there is only one output.

As with the inverter, the output of the AND gate can be expressed in terms of the inputs with a Boolean expression. The Boolean expression for an AND gate is:

$$C = AB$$

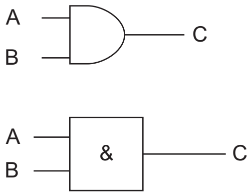

A	B	C
0	0	0
0	1	0
1	0	0
1	1	1

FIGURE 5.7 Logic symbols for AND gate and truth table.

This equation is read, “C equals A and B.” Parentheses also can be used to separate the inputs; for example, C=(A) (B).

The truth table for a simple two-input AND gate is also shown in Fig. 5.7. Note that the output C is a binary 1 when both inputs are binary 1s. At all other times, the output is binary 0. With two inputs, there are a total of four different input combinations ($2^2=4$). With three inputs, there are $2^3=8$ possible input combinations for an AND gate. These inputs are the binary numbers 000 through 111.

One of the most common applications of an AND gate in digital circuits is gating. Gating simply refers to the use of one binary signal to control another. A two-input AND gate is most often used as a control gate. One input signal is the control that either keeps the other input signal from passing through to the output or allows it to pass.

Fig. 5.8 shows the inputs and output of a typical control gate. Note that multiletter names called *mnemonics* are used instead of single letters to designate the input and output signals. The ENB or enable input represents the control signal. As long as the ENB input is a binary 0, the output is a binary 0. The gate is said to be inhibited because nothing more than a binary 0 output occurs.

When the ENB control input is a binary 1, the gate is enabled. At this time, the main input signal, CLK for clock, is allowed to pass through to the output BST. CLK is a periodic clock signal that switches repetitively between the binary 0 and the binary 1 levels at a fixed frequency. The output of the AND gate follows the CLK input as long as the ENB control input is a binary 1.

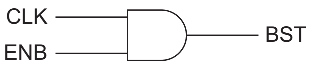

(A) Gate symbol

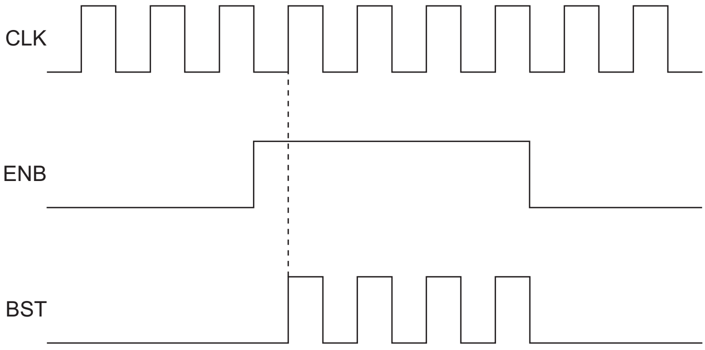

(B) Waveforms

FIGURE 5.8 How AND gate is used to turn signal off or on. (A) The AND gate showing the input and output names. (B) The waveforms of the inputs and output.

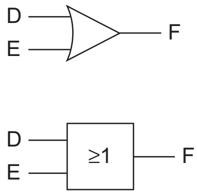

D	E	F
0	0	0
0	1	1
1	0	1
1	1	1

FIGURE 5.9 Logic symbols for OR gate and truth table.

OR Gate

Another commonly used logic gate is the OR circuit. Like the AND gate, the OR gate has two or more inputs and a single output. The OR gate generates a binary 1 if any one or more of its inputs are binary 1. The only time the output of an OR gate is binary 0 is when all its inputs are binary 0.

The symbols used to represent an OR gate are shown in Fig. 5.9. Letters of the alphabet are used to designate the inputs and outputs. Using these designations, the Boolean output expression for the OR gate is:

$$D = E + F$$

This equation is read, "D equals E or F." In this Boolean expression, the plus sign designates the OR function. In a Boolean logic expression, the plus sign does not mean addition. In Fig. 5.9, the mathematical symbol (greater than or equal to) is used to designate the OR logic function. The truth table for an OR gate further defines its operation. The output D is a binary 1 when input E or input F or both are a binary 1. Like an AND gate, an OR gate can have any number of inputs.

NAND and NOR Gates

The AND and OR logical functions are basic to all digital systems. Although these functions are implemented with AND and OR gates, a variation of these gates is even more widely used. These are NAND and NOR gates, which are a combination of either an AND gate or an OR gate and an inverter. These gates are more flexible in their application when they are used together.

A NAND gate can be made with an AND gate followed by an inverter. Fig. 5.10 shows this circuit configuration. Note the two special symbols in Fig. 5.10 used to represent the logical NAND function. The truth table is also given.

A NAND gate can perform all of the gating and detection functions mentioned earlier for an AND gate. The only difference is that the output is inverted.

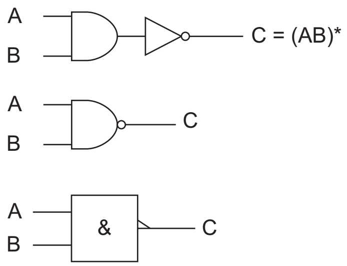

A	B	C
0	0	1
0	1	1
1	0	1
1	1	0

FIGURE 5.10 Logic symbols for NAND gate and truth table.

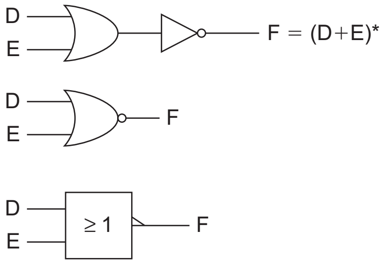

D	E	F
0	0	1
0	1	0
1	0	0
1	1	0

FIGURE 5.11 Logic symbols for NOR gate and truth table.

The other widely used digital logic element is a NOR gate. A NOR gate is a combination of an OR gate followed by an inverter (see Fig. 5.11). The symbols used to represent the NOR logic function are also shown. The output of a NOR gate is simply the complement of the output of an OR gate, as shown by the truth table. An XOR actually is a 1-bit binary adder.

Exclusive OR Gate

A variation of the OR gate is the exclusive OR, usually designated XOR. Its symbols are given in Fig. 5.12. The truth table is also shown. Note that the circuit only produces a 1 output if the two inputs are opposite. The output is 0 if the two inputs are the same. An XOR gate is used to make an adder.

An inverted XOR or XNOR has the symbols and truth table shown in Fig. 5.12. It is useful as a 1-bit comparator since the output is 1 if both inputs are the same.

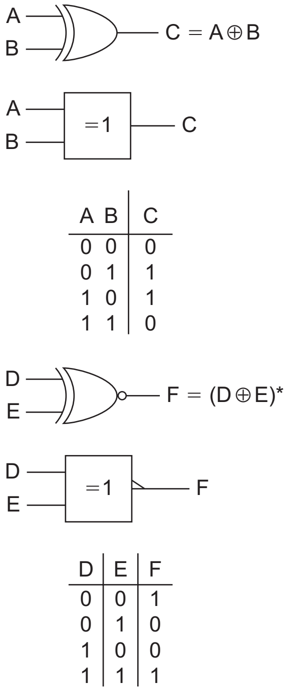

A	B	C
0	0	0
0	1	1
1	0	1
1	1	0

D	E	F
0	0	1
0	1	0
1	0	0
1	1	1

FIGURE 5.12 Logic symbols for XOR and XNOR gates and truth tables.

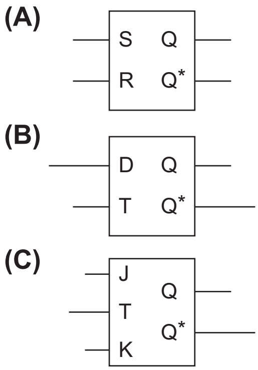

FIGURE 5.13 Logic symbols for flip-flops. (A) RS. (B) D type. (C) JK.

Flip-Flops

A flip-flop (FF) is another basic building block of electronics. It is essentially a circuit that stores 1 bit of data. The circuit has two states—set and reset. If the FF is storing a binary 1, it is set. If it is storing a binary 0, it is reset. Fig. 5.13A shows the basic block diagram of a reset–set (R–S) FF. The inputs are set (S) and reset (R). The two outputs Q and Q* are complementary. You apply an input to the appropriate input to set or reset the FF.

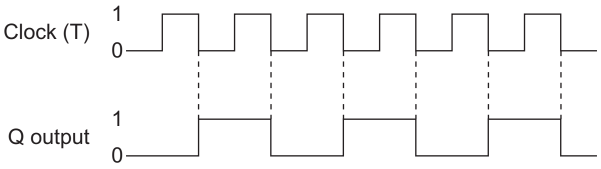

FIGURE 5.14 Toggling T input on JK flip-flop causes it to change states, producing frequency division by 2.

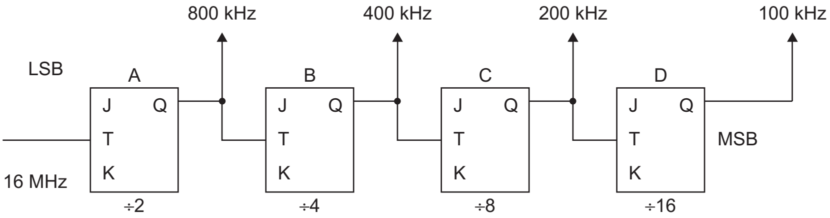

FIGURE 5.15 Cascading JK flip-flops produce higher-frequency division ratios, which is also a binary counter that counts from 0000 to 1111.

A D-type FF is shown in Fig. 5.13B. It has a D or data input and a clock (T) input. When the clock occurs, the FF is either set or reset depending on the value of the D input. The D FF is used to store 1 bit of data.

Another type of FF is the JK, shown in Fig. 5.13C. It has J and K inputs that correspond to S and R inputs, respectively. It also has a clock or T input that performs the set or reset operations. To store a binary 1, you make $J = 1$ and $K = 0$. When a clock input occurs, the FF is then set or reset.

The T input also causes the state of the FF to toggle or change state when a clock transition occurs (see Fig. 5.14). This feature is used to produce frequency division by 2 as shown. The output is a square wave half the frequency of the T input. Additional JK FFs can be cascaded to produce frequency division by the power of 2 or 2, 4, 8, 16, 32, 64, and so on, as shown in Fig. 5.15.

Storage Registers

A storage register is a place to store a binary word or number. It is made up of 1 FF per bit. An 8-bit register stores bytes and has 8 FFs. Fig. 5.16 shows a register with eight inputs and eight outputs. Most registers also have a clock input also called LOAD or STORE that determines when the input data is stored. Some also have a CLEAR or RESET input that puts all the FFs in the binary 0 state.

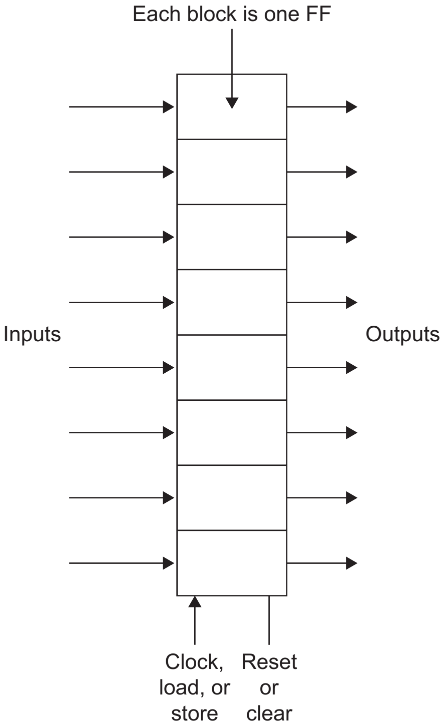

FIGURE 5.16 Storage register from one word consists of FFs. It also has clock/load and clear/reset inputs.

Shift Registers

A shift register (SR) is like a storage register, but the bits stored there may be shifted from one FF to the next as each clock pulse occurs. Fig. 5.17A shows how serial binary data is shifted into the SR for storage. As new data is shifted in, the existing data will be shifted out and sent to another circuit. Fig. 5.17B shows data being shifted out as binary 0s are shifted in.

SRs are often used for parallel-to-serial and serial-to-parallel conversions. A serial-to-parallel conversion is shown in Fig. 5.18A. The SR initially contains all binary 0s. Then serial data is shifted in, and the output is taken in parallel from the FF outputs. For parallel-to-serial conversion, a binary word is initially stored in the register. Then the word is shifted out as the clock pulses occur (see Fig. 5.18B).

Counters

As its name implies, a counter is a circuit that counts input pulses. Counters are made of multiple cascaded FFs, and they are sometimes accompanied by

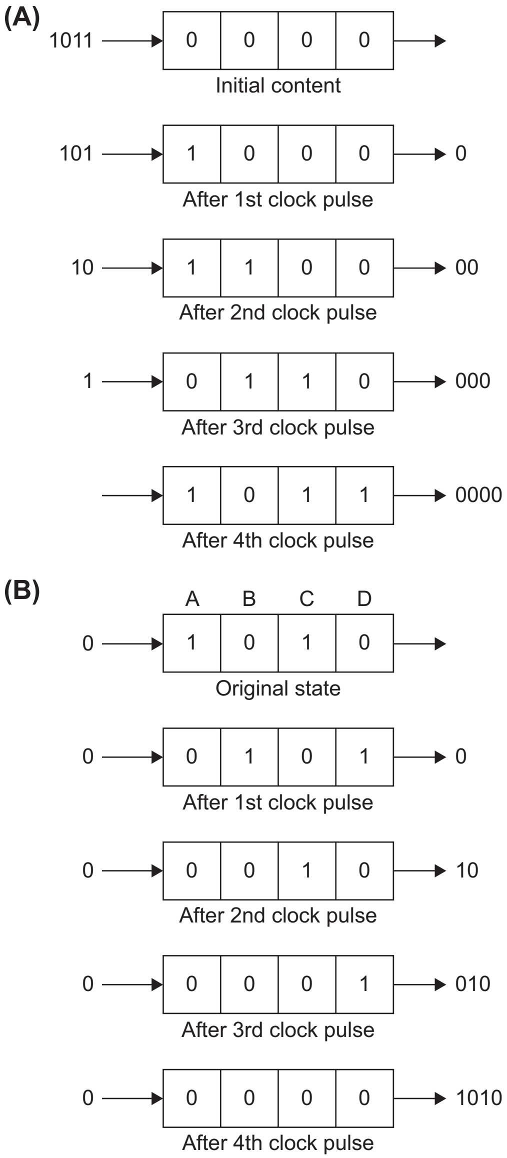
(A)
1011
0 0 0 0
Initial content
101
1 0 0 0
0
After 1st clock pulse
10
1 1 0 0
00
After 2nd clock pulse
1
0 1 1 0
000
After 3rd clock pulse
1 0 1 1
0000
After 4th clock pulse
(B)
A B C D
0
1 0 1 0
Original state
0
0 1 0 1
0
After 1st clock pulse
0
0 0 1 0
10
After 2nd clock pulse
0
0 0 0 1
010
After 3rd clock pulse
0
0 0 0 0
1010
After 4th clock pulse

FIGURE 5.17 (A) Shifting serial data into shift register. (B) Shifting serial data out of shift register.

logic gates. As the input pulses occur, the binary number stored in the counter FFs tells how many pulses have occurred. The frequency divider shown in Fig. 5.15 is also a 4-bit counter. It can count from 0 (0000) to 15 (1111). If all FFs are initially reset to 0, then 10 pulses occur, and the FFs from A to D would be storing 0101 from left to right. The rightmost FF (D) is the MSB, so you read the counter as 1010 or the binary equivalent of decimal 10.

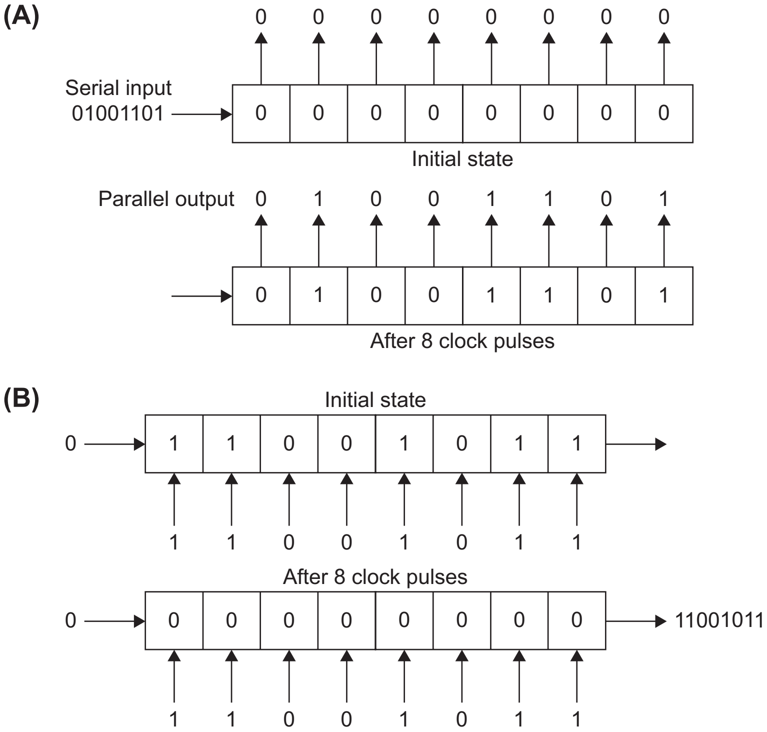
(A)
Serial input
01001101
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
Initial state
Parallel output 0 1 0 0 1 1 0 1
0 1 0 0 1 1 0 1
After 8 clock pulses
(B)
Initial state
0
1 1 0 0 1 0 1 1
1 1 0 0 1 0 1 1
After 8 clock pulses
0
0 0 0 0 0 0 0 0
11001011
1 1 0 0 1 0 1 1

FIGURE 5.18 Shift registers used for (A) serial-to-parallel data conversion and (B) parallel-to-serial data conversion.

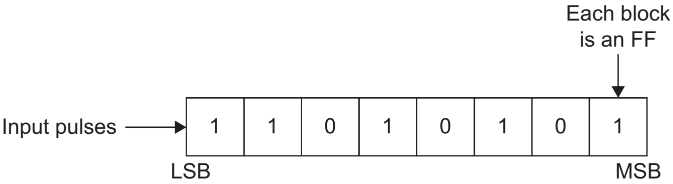
Each block
is an FF
Input pulses
1 1 0 1 0 1 0 1
LSB
MSB

FIGURE 5.19 Counter storing binary value equal to number of input pulses that occurred.

Binary counters can store numbers up to $2^N - 1$, where N is the number of FFs. For example, an 8-bit counter would count from 00000000 to 11111111 or $2^8 - 1 = 256 - 1 = 255$.

Fig. 5.19 shows a counter where 178 pulses have occurred, so the counter stores the number 10101011 that is the binary equivalent of 179.

Another popular form of counter is the BCD version that counts from 0000 to 1001 or 0 through 9. BCD counters are used to keep track of decimal quantities in binary form. BCD counters can serve as frequency dividers as well. Each divides by 10.

COMMON LOGIC CIRCUITS

There are several types of logic circuits that occur again and again in digital equipment. The counters and registers just described are good examples. A few others that you may encounter are described in this section. Remember that these circuits can be made from individual gates and FFs in IC form. You can also buy them in IC form already wired for their specific function. But mostly these circuits appear as a collection of circuits inside some larger IC.

Multiplexer

A multiplexer or mux or mpx is a circuit with two or more inputs and one output (see Fig. 5.20). Here there are four inputs and one output. Two other inputs are used to select one of the inputs to appear as the output. The four selected inputs are 00, 01, 10, and 11. In the figure, note which input is routed to the output with each input select code. Multiplexers can have as many inputs as needed.

Demultiplexer

A demultiplexer or demux is the opposite of a mux. It has one input and multiple outputs. Like the mux, select inputs specify which of the output lines the input is connected to.

Decoder

A decoder is a circuit that can detect a specific binary code. In its simplest form, it is just an AND gate, as shown in Fig. 5.21A. If the input 0110 appears at the inputs, the AND gate will have all its inputs at binary 1, so the output will

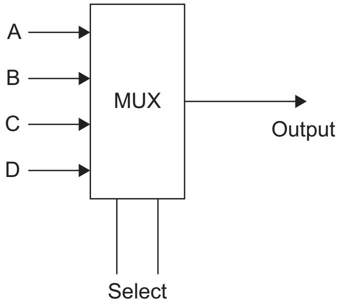

Select	Input at output
00	A
01	B
10	C
11	D

FIGURE 5.20 Digital multiplexer (mux) that selects one of four inputs to appear at the single output.

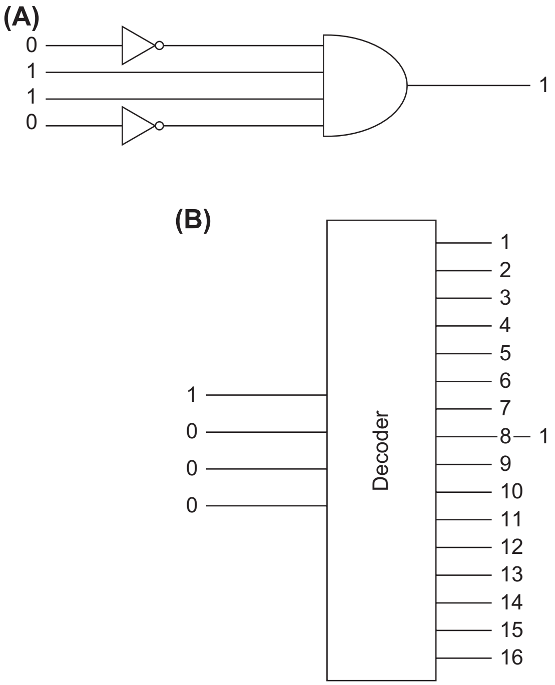

FIGURE 5.21 Decoders. (A) 1-AND gate decoder that identifies input 0110. (B) 4-to-16 decoder that identifies 1 of 16 different 4-bit input values.

be binary 1 indicating that it is detecting the desired input. For any other 4-bit input, the output will be binary 0.

Decoders can also be made to detect all states of a given binary input. For example, a 4-bit decoder like that in Fig. 5.21B looks at the 4-bit inputs and decodes all states from 0000 to 1111. If the input is 1000, the 8 output is binary 1, and all other outputs are 0.

Comparators

A comparator takes two binary words as inputs and compares them. If the two inputs are equal, a binary 1 output occurs. Otherwise, if the two inputs are different, the output is 0. Some comparators have outputs that can tell if one input is larger than or smaller than the other.

Arithmetic Circuits

Digital logic does a good job of basic arithmetic functions such as add and subtract. The basic computing element is called an arithmetic logic unit (ALU). It does binary addition and subtraction. It can also perform basic logic operations such as AND, OR, or XOR on two binary input words. Circuits for multiplication and division are more complex, but are common in digital computers.

DIGITAL MEMORIES

Digital memories are semiconductor circuits that store binary data. You have probably heard of the terms RAM and ROM as they apply to personal computers. These are the two main types of memory, random access memory (RAM) and read-only memory (ROM).

Random Access Memory

The RAM is one in which you can access hundreds, thousands, or millions of binary storage locations. You can store one binary number or word in each location or read the data out of the location. Each location is referred to by a unique binary number called the address. You send the memory an address and activate the selected location. Then you can read data from the location or write data into that location. RAMs therefore are what we call read-write memories.

Fig. 5.22 shows the concept of a RAM. It has eight locations for 4-bit words. To address eight locations you need a 3-bit address ($2^3=8$). When you apply the address to the address decoder, the decoder activates the desired location. In the figure you see that the binary number 1001 stored in location 011 is being accessed and read out.

Most memories are large today. We talk about megabytes (MB) or millions of bytes of data or gigabytes (GB) or billions of bytes of data in a modern PC. Embedded controllers will have fewer locations, usually in the thousands of bytes (KB) or perhaps more in the larger systems. Storage word sizes are usually 8 (bytes), 16, 32, or 64 bits.

There are two ways to make a RAM. The first way is just to build one FF for each bit to be stored. The basic set–reset FF described earlier is the basic

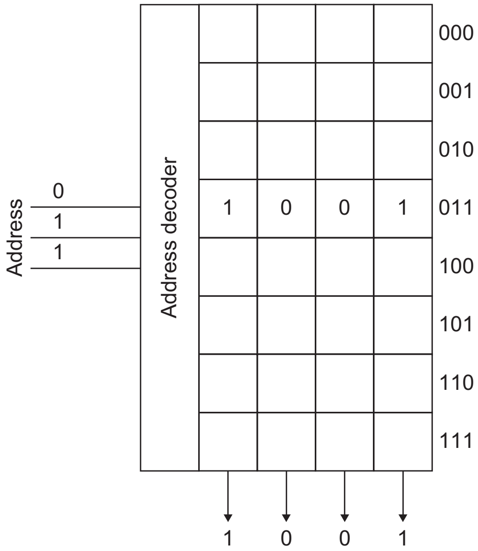

FIGURE 5.22 Basic concept for random access memory (RAM). Address selects one of eight storage locations. Then data may be read from that location or data stored in it.

storage circuit (see Fig. 5.23A). Memories using this type of circuit are called static RAM or SRAM. They are called static because once you store the bit in the FF, it stays there as long as power is applied to the circuit.

The other form of RAM is called *dynamic RAM* or DRAM. The basic storage circuit is a capacitor as shown in Fig. 5.23B. If you charge the capacitor, it is storing a binary 1. If the capacitor is discharged, a binary 0 is being stored. A MOSFET is used as a switch to access the capacitor. If the MOSFET is turned on by an input address on the gate, it connects the capacitor to the data line, where it may be read out or where a new bit may be stored. When the MOSFET is turned off, the capacitor is isolated and just stores the bit.

The reason why the term *dynamic* is used for this type of RAM is that the charge on the capacitor will leak off very quickly. The capacitor is very small, and millions of them are made on a tiny chip of silicon. For this reason, the capacitor has to be recharged every so often so that the data is retained. This is called a refresh operation, and it usually occurs every 2 ms or so.

Comparing SRAM and DRAM, you should know that SRAM is the faster of the two in that you can store data or read data out faster than you can with a DRAM. The downside is that the SRAM is more complex and takes up more space on the chip, and thus costs more. The DRAM is slower but not much. Its advantage is that you can put more storage cells on a chip, making them cheaper. The need to refresh is a downside but easy to live with, as all the refresh circuitry is on the chip, and the process is transparent to the user.

Just keep in mind a key fact. RAM is fast and reliable, but it loses all data if the DC power to the memory circuit is removed. All data is erased. RAMs are therefore said to be volatile.

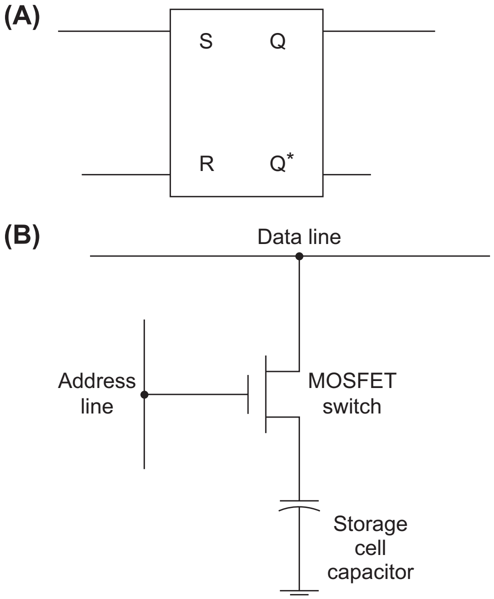

FIGURE 5.23 RAM storage cells. (A) RS FF for SRAM. (B) Capacitor cell for DRAM.

A newer memory type is the ferroelectric RAM or FRAM. Its basic structure is similar to that of the capacitive DRAM cell of Fig. 5.23B. The difference is that the dielectric material in the capacitor has magnetic capabilities. When data is written into or data is extracted from the cell, electric field state is stored magnetically. That means that the cell remembers its state after an operation even if power is removed. This makes FRAM nonvolatile like ROM. FRAM is fast and uses very low power. This type of memory is used in some newer microcontrollers.

Read-Only Memory

ROMs are nonvolatile memories. Once you store data in them, they retain the data when power is removed. In some ROMs, you can change the data stored there. Incidentally, all ROMs are random access in that you can go to any desired location to read or maybe write data.

There are several categories of ROM. The older types are called masked ROM, where the mask refers to the actual wiring of the ROM to permanently fix the data in the circuitry when it is made. The storage cells are essentially a MOSFET matrix where each MOSFET is permanently turned on or off to store 1 or 0. No changes are possible. Later a programmable ROM or PROM was developed. It used a special MOSFET that could be programmed on or off. You could then erase the programming by shining ultraviolet (UV) light on the chip for a short time. Then the PROM could be reprogrammed. These are no longer used.

Most ROMs today are what we call electrically erasable programmable ROMs or EEPROMs or E^{2}PROMs. The storage element is a MOSFET that you turn off or on with a programmable voltage (see Fig. 5.24). The MOSFET retains that state until it is electrically changed.

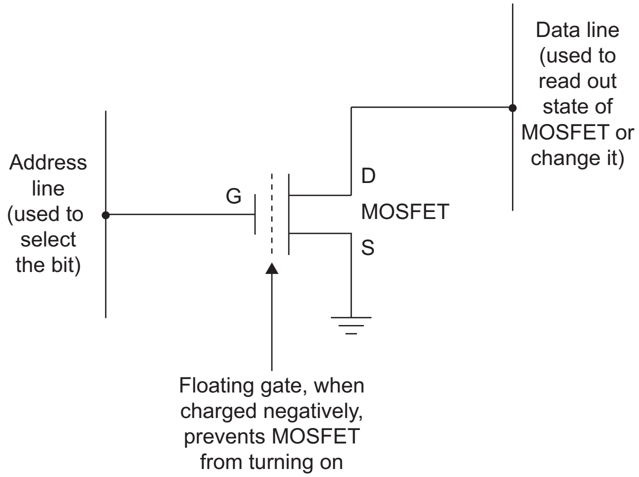
Data line (used to read out state of MOSFET or change it)
Address line (used to select the bit)
G
D
MOSFET
S
Floating gate, when charged negatively, prevents MOSFET from turning on

FIGURE 5.24 Storage cell used in EEPROM and flash memories is MOSFET with a special floating gate that keeps transistor off or on.

The main problem with EEPROMs is that they are slow to erase and reprogram. But today we have a special version of EEPROMs called flash memories. These are also EEPROMs, but they actually are almost as fast as RAMs for reading and writing data. Yet they are nonvolatile.

You will hear of two types of flash memory—NOR and NAND. NOR flash is used to store programs for microcontrollers. It can only be erased on large block sizes and not individual cells. NAND flash is faster, and you can erase individual memory locations. The NAND flash is gradually replacing the NOR flash in many applications.

Flash memory has become so popular that it is used in almost every electronic product today. MP3 players and iPods store music in flash memory. Cell phones store phone numbers, digital cameras store photos, and in embedded controllers, the flash usually stores the program that tells the processor what to do. And, let us not forget the flash drive or "thumb" drive as it is called. These devices with a USB connector can store many gigabytes of data and are very inexpensive. Flash memory is also used to make solid-state drives (SSDs) that are the equivalent of the hard disk drives used in PCs and laptops. They are still expensive, but as prices continue to decline, you will see more SSDs replacing conventional mechanical/magnetic hard drives.

PROGRAMMABLE LOGIC DEVICES

There are two ways that digital devices are built today. The most common is to use an embedded controller or microcomputer. This is a single-chip computer that you can program to do whatever digital function you want. Most electronic products have at least one embedded controller at the heart of their design.

In some cases, the embedded controller is not fast enough to do the job. One way to make the circuitry is to use individual logic circuits such as gates, FFs, counters, and so on. That is the way it used to be done. Now, if you need fast custom digital logic, you build the circuit with a programmable logic device (PLD). A PLD is an IC that contains a mix of gates, FFs, and other circuits that can be programmed to implement any digital function. The programming is usually done on a computer, and a file is created. The file is then downloaded to the PLD, where it is stored in a RAM or ROM. The file in the memory determines how the unique circuit is formed.

There are three basic types of PLDs: simple PLDs (SPLDs), complex PLDs, and field programmable gate arrays (FPGAs). The SPLDs are small devices for simpler digital circuits. The most common types are called programmable array logic (PAL), programmable logic arrays (PLAs), and generic array logic (GAL). PROMs are also used as SPLDs. For example, the address input to a ROM is some binary number or a collection of binary inputs. These form an address to locate a specific binary word stored in that location. When that particular selection of input bits occurs, it forms the address that identifies the memory location where the desired set of output responses is stored. This form of SPLD is also called a lookup table or LUT.

The other SPLDs are essentially a set of AND and OR gates that can be programmed. They all take the form shown in Fig. 5.25. You can select which inputs go to the AND gates, and/or select which AND outputs go to which OR gate inputs. A PROM has the AND array fixed, and you can program the OR gates. A PAL is an SPLD in which you can program the AND inputs, but the OR connections are fixed. A PLA is a PLD that lets you program both the AND

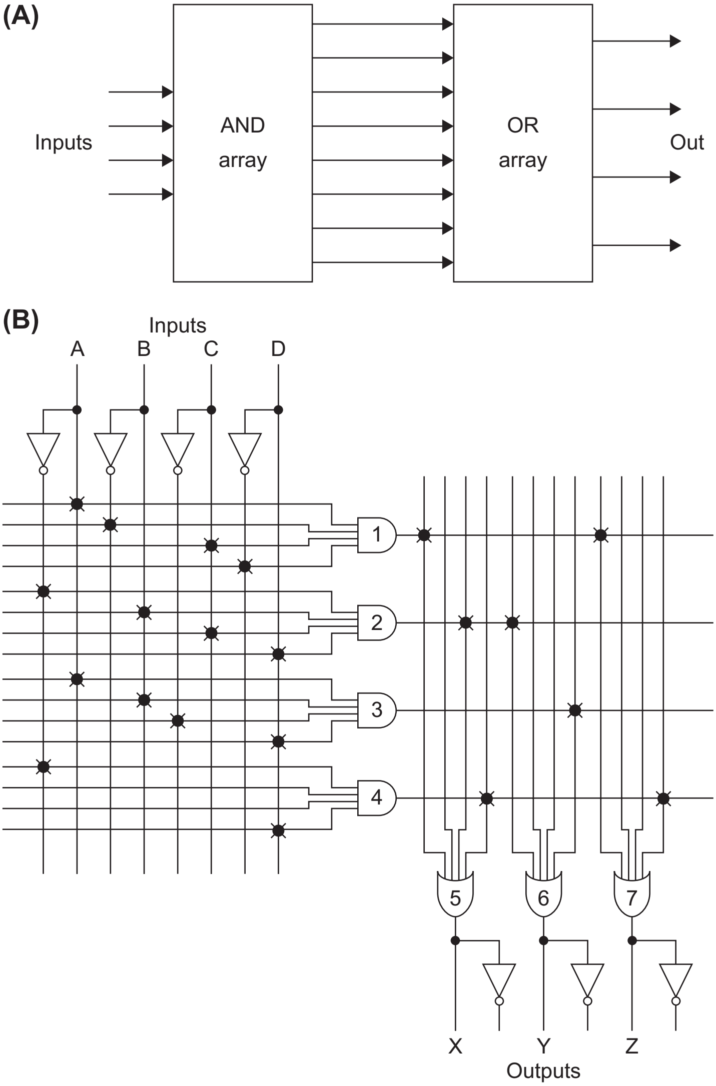

FIGURE 5.25 PLD concept. (A) Block concept of programmable AND and OR gate arrays. (B) Simple circuit of PLD showing programming by dots on AND and OR gate arrays.

inputs and the OR connections. In both PLAs and PALs, the programming is usually permanent.

A GAL is a variation of the PAL with its AND inputs programmable. The difference is that the programming can be changed. The program you develop is actually stored in flash memory. The flash memory cells send the programming bits to the various AND and OR gates telling them what logic to implement. You can then erase it and reprogram if you want or need to. GALs also have one or more FFs associated with the logic gates, so you can store bits or actually build counters and registers by programming.

CPLDs are complex PLDs similar to SPLDs, only larger. They contain more gates and FFs so that very large logic projects can be built on a single chip. A typical CPLD is really just a large collection of PALs or GALs that can be individually programmed and then connected together with a large interconnect matrix. Input/output (I/O) circuits also provide inversion if needed or extra power to drive external circuits. Again, flash memory is used to store the programming.

The largest and most flexible PLD is the field programmable logic array or FPGA. It has become the most popular and common way to implement digital equipment other than a microcontroller. FPGAs are large chips that contain thousands of logic circuits. The largest devices have millions of gates and other circuits. Most FPGAs are programmed by storing a bit pattern in the SRAM cells inside. These cells tell the logic circuits what configuration to take and what functions to perform. If you turn the power off, the programming goes away, and all you have is a dumb FPGA with no logic interconnections or functions. When you power up an FPGA, the desired program, usually from a flash memory, or a computer loads the RAM cells with the program, thereby enabling the FPGA to perform its desired function.

The logic inside an FPGA is essentially thousands of LUTs that can be individually programmed. Each LUT has its own program memory block as well as one or more FFs and I/O circuits. The LUTs may also be interconnected to one another by programming. With this arrangement virtually any digital logic operation can be programmed. Some of the newer FPGAs also include microprocessor cores and common I/O interfaces.

There are about a half-dozen companies that make PLDs. The two oldest and largest are Xilinx and Altera (Intel). Both have extensive lines of CPLDs and FPGAs.

DATA CONVERSION

The term *data conversion* refers to the translation of analog signals into digital form and vice versa. While it used to be an all-analog electronics world, today it is virtually all digital. We still have analog signals such as voice, video, radio waves, and sensor voltages, but mostly we process, store, and transmit data by digital means. For that reason, circuits that perform these conversions are critical to virtually all electronic applications. All cell phone voice is digital, all

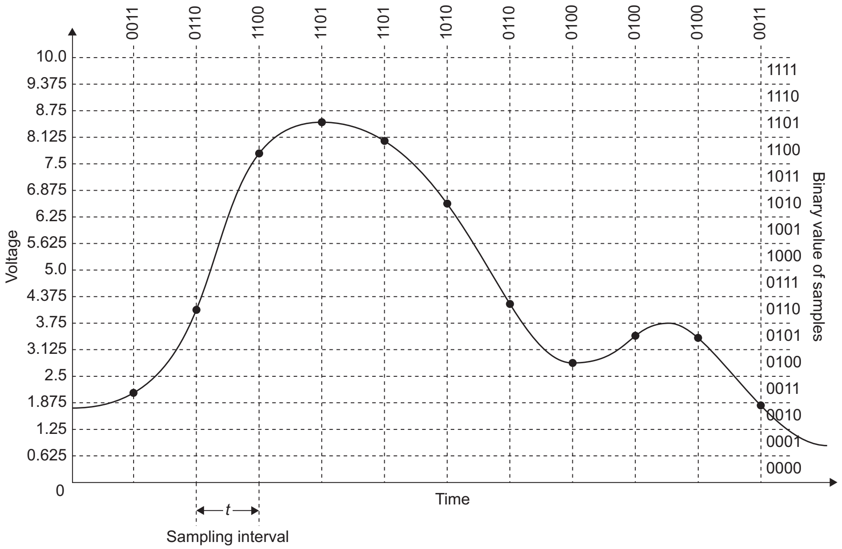

FIGURE 5.26 Analog-to-digital conversion is performed by sampling analog signal at equal intervals and generating proportional binary value.

modern TV is digital, and sensor outputs are captured and processed in digital form. It is difficult to name a product today that does not use some form of data conversion.

Analog-to-Digital Conversion

The process of converting an analog signal into a digital one is called analog-to-digital conversion, and it is performed by an analog-to-digital converter (ADC). The process, also referred to as sampling, is illustrated in Fig. 5.26. The ADC looks at the analog input and periodically takes a sample of the voltage at that instant, captures it, and then converts it into a proportional binary number. We say that we are digitizing the signal. The sample points are shown by the dots on the analog curve. The binary value of the sample is shown to the right of the curve. The conversion process actually results in a sequence of binary numbers that represent the analog waveform. These values are usually stored in a RAM or transmitted to other circuits as shown in Fig. 5.27. Note the symbol for an ADC.

Digital-to-Analog Conversion

To recover the original signal, we put the data sequence previously captured by the ADC into a digital-to-analog-converter (DAC) (see Fig. 5.28). The output is a version of the analog signal. The DAC output is not a perfect reproduction, but just an approximation. This is shown in more detail in Fig. 5.29. Each

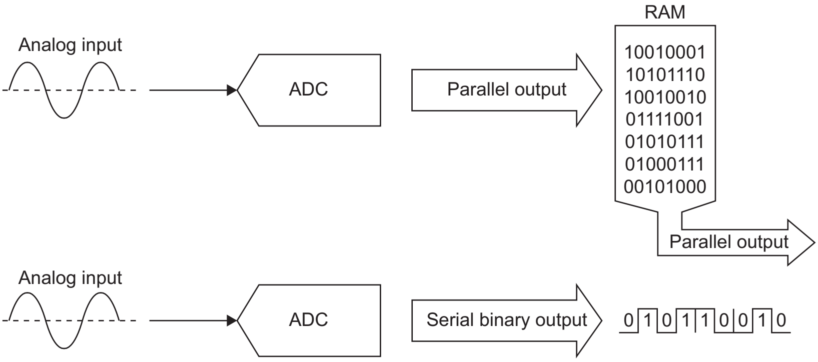
RAM
Analog input
ADC
Parallel output
10010001
10101110
10010010
01111001
01010111
01000111
00101000
Parallel output
Analog input
ADC
Serial binary output
0 1 0 1 1 0 0 1 0

FIGURE 5.27 Symbol for ADC with both parallel and serial output examples.

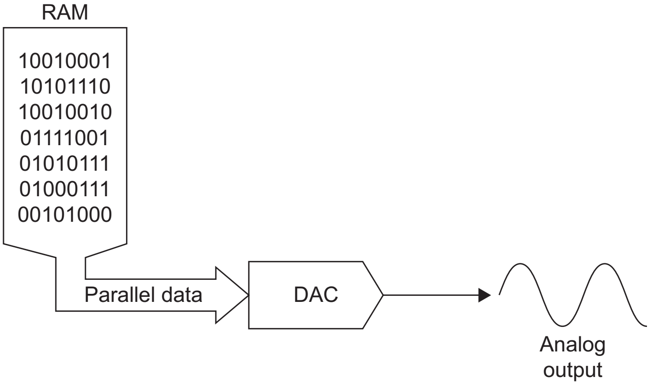
RAM
10010001
10101110
10010010
01111001
01010111
01000111
00101000
Parallel data
DAC
Analog output

FIGURE 5.28 Converting binary data back into analog signal with DAC.

binary input results in a constant voltage output from the DAC during the sample period. The result is a stepped approximation to the original signal. The rate at which the binary data is sent to the DAC must be the same as the sampling interval to recover the original frequency information in the signal.

Resolution and Sampling Interval

The key to good data conversion is to use greater resolution and faster sampling rates. Resolution refers to the number of bits used in the data conversion. In Fig. 5.27, only 4 bits are used, so the resolution is poor. The voltage range is only divided into 16 intervals, meaning that amplitude variations at less than 0.625 V are missed. This problem can be corrected by using more bits. ADCs are available in many bit sizes. The most common are 8, 10, and 12 bits, but 14 and 16 bits are available. Some methods of ADC produce resolutions of 20–26 bits. The result is a finer conversion of amplitude detail. As an example, if the 0- to 10-V range in Fig. 5.27 was sampled with a 12-bit ADC, the individual smallest

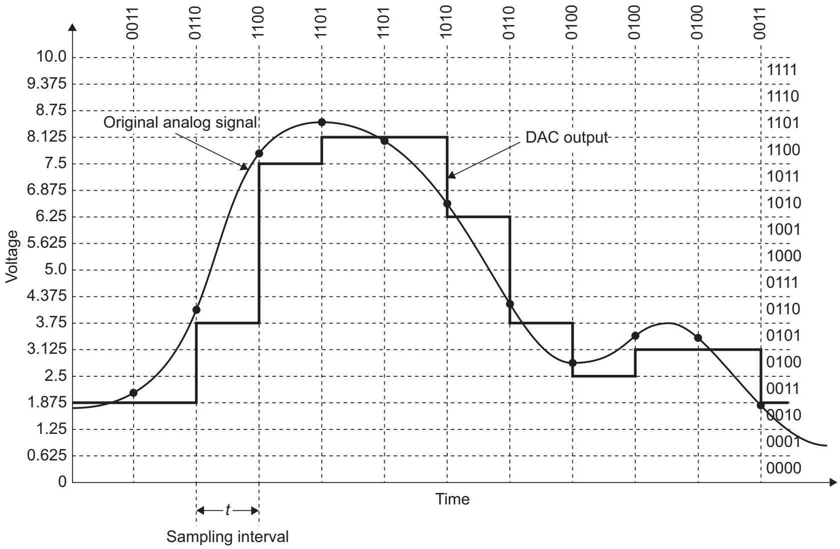

FIGURE 5.29 DAC output is stepped approximation to original analog signal.

voltage increment that can be detected is $10/2^{12} = 10/4096 = 2.44$ mV instead of the 0.625V in the figure.

Another critical specification is sampling rate. To retain all the frequency detail in a signal, the sampling rate must be at least twice the highest frequency in the signal. This is called the Nyquist criterion. For example, when digitizing music with a frequency range of 20Hz–20kHz, the sampling rate must be at least double the 20-kHz frequency. In most systems, a rate of 44.1 or 48kHz is used.

ADCs with sampling rates to tens of gigahertz are commonly available, meaning that even radio signals can be digitized.

Direct Digital Synthesis

In Chapter 4 you learned how a phase-locked loop (PLL) synthesizer works to generate a variable frequency output. A newer way to generate an output signal is direct digital synthesis (DDS). It uses the sampling principles described here along with a DAC to produce a variable frequency output of any desired waveshape.

Fig. 5.30 shows a basic DDS. Stored in the memory are samples of a sine wave as they may have been produced by an ADC. A memory address register sends sequential addresses to the memory, and the samples are read out and sent to a DAC where a stepped approximation of the sine wave is produced. A low-pass filter (LPF) smoothes the output into a near perfect sine wave.

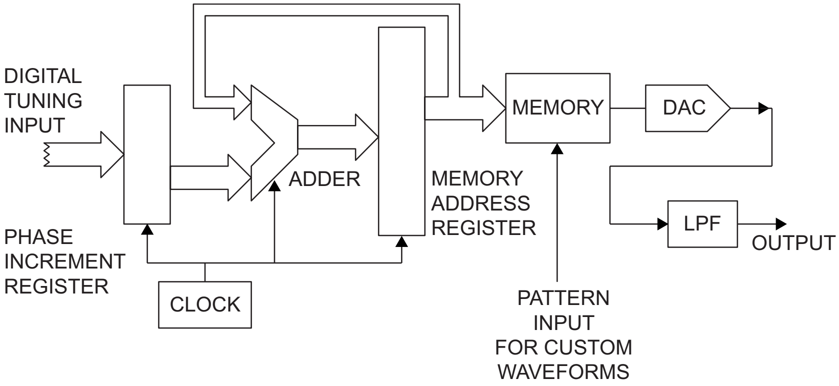

FIGURE 5.30 A generic block diagram of a direct digital synthesizer (DDS).

Changing frequency is a bit tricky since the clock is a fixed frequency. To change frequency, the circuit selects the specific samples to be used for the output. This is done by adding a phase increment constant value to the memory address register value in an adder. The effect is to use more or less of the samples in the output. Increasing the phase increment value increases the output frequency and vice versa.

The DDS has become quite popular because it is available as a completely easy-to-use IC. Furthermore, almost any unique waveform pattern can be stored in the memory making the DDS an arbitrary waveform generator (AWG). DDS chips are available to produce outputs up to about 100 MHz.

Project 5.1

Building a 5 V Power Supply

The digital projects to follow use CMOS logic circuits that operate from +5 V. If you apply more than 5 V, you will destroy the IC. Any of the following power sources is suitable:

- the power supply kit you may have built in Project 4.1 set to 5 V;
- a regulated AC to DC 5 V wall adapter;
- the power supply you may have built in Project 4.7 (Fig. 4.33);
- the 6 or 9 V battery supply plus the 7805 regulator circuit in Fig. 4.33.

Project 5.2

Examining a NAND Gate

A NAND gate is a logic gate that implements the NOT AND function (Fig. 5.10). In this project you will demonstrate the operation of a NAND gate and a common application.

Continued

Project 5.2—cont'd

The IC you will use is the 74HC00, a quad two-input NAND. Four NAND gates in a 14-pin dual inline package (DIP). Fig. 5.31 shows the logic diagram and pin connections. Do the following:

- Plug the 74HC00 into your breadboard.
- Connect pin 14 to +5 V and pin 7 to ground or minus.
- Connect two short #22 wires into pins 1 and 2. These are the inputs. You will connect the wires to +5 V for a binary 1 and to ground for binary 0.
- Connect the 330-Ω resistor and LED as shown in the figure.
- Apply power to the circuit.
- Apply the inputs shown in the truth table of Fig. 5.31, and for each pair on inputs, note the output and record a 0 or 1 in the truth table. Your truth table should be like that in Fig. 5.10.

One common application of the NAND gate is to gate another signal. That is the circuit serves as an open or shut gate. Gating a clock signal is one widely used case. You will use the 555 timer circuit you built earlier as the clock. Do this:

- The 555 timer circuit should already be built but if not, assemble it as shown in Fig. 5.32. Use 5 V for the supply.
- Connect the 555 timer output from pin 3 to pin 1 of the NAND gate. The wire input to pin 2 of the gate is the enable or disable signal.
- Connect the wire from pin 2 to ground or binary 0. Does the LED flash?
- Now connect the wire from pin 2 to +5 V or binary 1. Does the LED flash? It should indicate that the gate is allowing the signal through. The voltage on pin 2 is the enable or gating signal.
- Turn off the power.

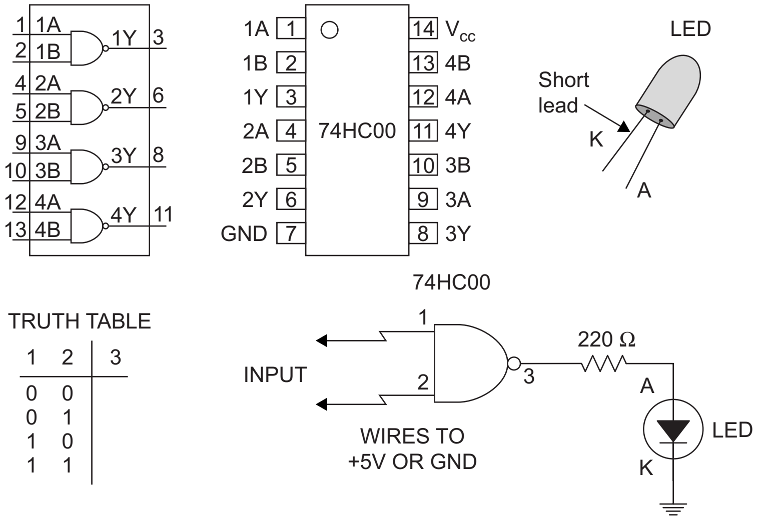
1 1A
2 1B
1Y 3
4 2A
5 2B
2Y 6
9 3A
10 3B
3Y 8
12 4A
13 4B
4Y 11
1A 1
1B 2
1Y 3
2A 4
2B 5
2Y 6
GND 7
74HC00
14 V$_{CC}$
13 4B
12 4A
11 4Y
10 3B
9 3A
8 3Y
LED
Short lead
K
A
TRUTH TABLE
74HC00
1
2
3
INPUT
WIRES TO +5V OR GND
220 Ω
A
LED
K

1	2	3
0	0	
0	1	
1	0	
1	1	

FIGURE 5.31 Pin out details of the 74HC00 quad NAND logic IC, a NAND test circuit with LED output, and a truth table.

Project 5.2—cont'd

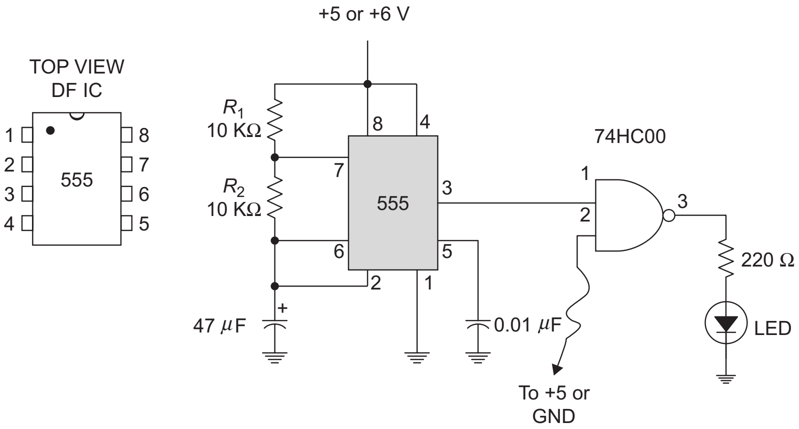

FIGURE 5.32 The output of a 555 timer is the input to a NAND gate.

Project 5.3

Demonstrating a Binary Counter

A binary counter is made up of FFs connected in a sequential arrangement. When an input is applied to the counter, it will toggle the FFs in a binary manner. The counter actually counts the number of input pulses and stores a binary value equal to the number of pulses received. As an example, a 4-bit counter counts from zero 0000 binary to 15 or 1111 binary. If the counter is initially reset or set to 0000 and 6 input pulses occur, the binary number stored in the counter will be 0110 or decimal 6. The binary state of each FF can be observed at each FF output.

The binary counter is also a frequency divider. Since each FF divides its input frequency by 2, an input of 50 kHz will produce a 25 kHz output. Cascading FF causes each FF to divide by 2. With a 4-bit binary counter a 50 kHz input will be divided down by a factor of 2^4 or 16–3.125 kHz observed at the fourth FF output.

In this project you will build a binary counter and observe its outputs as it counts. You will also demonstrate the counter as a frequency divider by looking at the FF outputs on an oscilloscope.

- Construct the circuit shown in Fig. 5.33. The 555 timer circuit of Fig. 5.32 that should still be working will be used as the clock signal to drive the counter.
- The counter is a 74HC161 4-bit counter IC. Its four outputs operate the four LED outputs. Be sure to connect pin 16 to +5 V and pin 8 to ground. Pins 1, 7, 9, and 10 are not used but need to be at +5 V and are connected by the 1k-Ω resistors. Double check the wiring of the circuit before you apply power.

Continued

Project 5.3—cont'd

- Apply 5 V. You should immediately see the LEDs flashing. The clock frequency is about 1 Hz. Observe for a while until you can identify the binary code sequence. It should correspond to the binary sequence shown in Box Hexadecimal Notation. Note in Fig. 5.33 which LED represents the LSB and the MSB. The pattern will repeat after each 16 input pulses.
- Turn off the power.
- Replace the 47-μF capacitor in the 555 circuit to a 0.01 μF ceramic. This will increase the clock frequency to about 4.8 kHz with a period of just over 200 μs.
- Turn on the power. The LEDs are flashing too fast for you to follow so they all just glow dimly.
- Use your oscilloscope to view the FF outputs. If you have a dual trace scope, you can look at two waveforms at the same time. Look at the 555 timer output at pin 3 or at pin 2 of the 74HC161. Then look at each FF output at pins 14, 13, 12, and 11 and note how each output frequency is half the previous output. The output at the fourth FF (pin 11) should be about 300 Hz with a period of about 3.33 ms.
- Turn off the power.

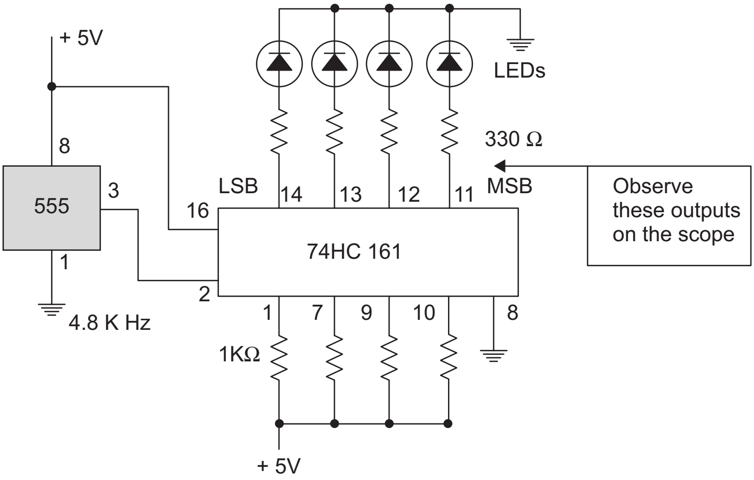

FIGURE 5.33 A 74HC161 IC binary counter clocked by a 555 timer to demonstrate the binary count sequence on LEDs.

Chapter 6

How Microcomputers Work

The Brains of Every Electronic Product Today

In this Chapter:

- Microprocessors, microcomputers, and embedded controllers defined.
- The stored program concept.
- Central processors and how they work.
- Types of microcomputers.
- Modern examples of embedded controllers.
- Digital signal processing.

INTRODUCTION

Whenever you hear or see the word *microcomputer*, you probably think of a personal computer. A personal computer is, of course, one type of microcomputer. But you may not be aware of other forms of microcomputers. For example, do you know that almost every piece of electronic equipment you own or use has a microcomputer in it? This type of microcomputer is called an embedded controller, microcontroller, or microcontroller unit (MCU). It is usually a single integrated circuit that performs all of the basic functions of a computer but is dedicated to a specific task. Such micros (*micro* is short for microcomputer) are in TV sets, stereo receivers, microwave ovens, CD players, DVD players, cell phones, blenders, copiers, and most other types of electronic devices. You will find them in your car (many of them), your iPod, your bathroom scale, gasoline pumps, and dozens of other things you use every day. Microcomputers do indeed make your day—you cannot live without them.

CONCEPTS AND DEFINITIONS

A microcomputer is a small digital computer that can take several different forms. It can be a single integrated circuit, or it can be a module made up of several integrated circuits on a printed-circuit board.

The two most common types of microcomputers are the embedded microcontroller and the personal computer. One special type of microcomputer is the programmable logic controller (PLC) used for industrial control (Chapter 12).

Electronics Explained. http://dx.doi.org/10.1016/B978-0-12-811641-8.00006-0
Copyright © 2018 Elsevier Inc. All rights reserved.

The larger microcomputers, such as personal computers and laptops, are made with a microprocessor. A microprocessor is a large-scale integrated circuit that contains most of the digital logic circuitry usually associated with a digital computer. This logic circuitry is referred to as the central processing unit (CPU). A microprocessor is a single-chip CPU.

Another common micro is called a *core*. A core is a microprocessor or microcomputer integrated with some other circuits. Today multicore micros are common. Two or more cores are used to get more processing power. Each core can run its own programs in parallel giving faster processing.

Embedded microcontrollers are complete micros on a single chip, including the CPU, memory, and input/output circuits. They are also called microcontrollers or MCUs. But in all cases, a microcomputer is an assembly of digital logic circuits, such as gates and flip-flops, that is used to process data. It is sometimes referred to as a *data processor* or simply *processor*.

Data, of course, refers to the binary numbers and words the processor works with. Processing refers to the way the data is manipulated or handled. Types of processing include arithmetic, logic, sorting, translating, editing, counting, and searching. Any action taken on the data is called processing. Processing normally implies that the data is changed in some way or is used to create new data. Data that is not processed as suggested above is dealt with in other ways. Four common ways are storing, retrieving, input, or output. The data is not changed by any of these techniques.

Storing data means putting it in a safe place, like a semiconductor memory or disk drive, where it can be accessed later. *Retrieving*, of course, is the opposite of storing, or going to get the data for reuse.

Input means taking data into the computer to be stored or processed. *Output* means sending the data from the computer to some external device. Input and output (I/O) are ways to transmit data from one place to another.

Another feature of a microcomputer is its *decision-making* capability. During processing, the computer can make decisions and alter its sequence of operation. In other words, the computer can "make up its mind" based on the state of the data or outside conditions.

For example, the computer can tell if a number is less than, greater than, or equal to another. It can choose among alternative courses of actions or say yes or no, true or false, if given enough input facts.

Another major application of computers is *control*. Computers can be used to actuate relays and solenoids or turn lights and motors off and on (see Fig. 6.1). In control applications, the computer actually determines when external devices are turned on or off. The computer serves as an electronic clock to time various operations. One example is the embedded single-chip microcontroller inside a microwave oven or a washing machine. Another example is a PLC, a special type of microcomputer used in industrial applications, that is commonly used to sense, sequence, and do time operations in a factory.

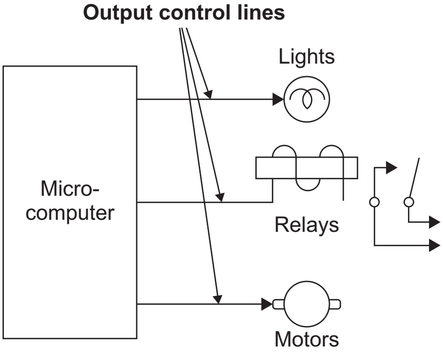
Output control lines
Lights
Micro-
computer
Relays
Motors

FIGURE 6.1 How a microcomputer controls external devices.

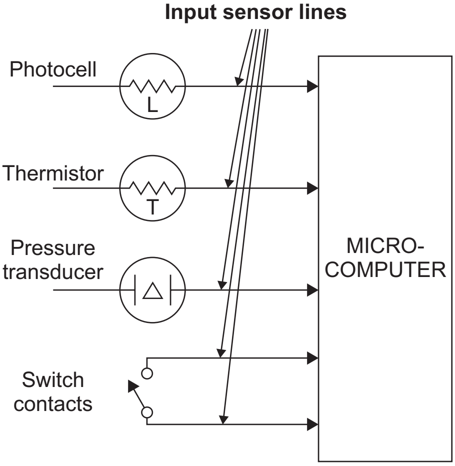
Input sensor lines
Photocell
L
Thermistor
T
Pressure
transducer
Switch
contacts
MICRO-
COMPUTER

FIGURE 6.2 How a microcomputer monitors external devices.

The most important thing to remember is that the control is automatic. The micro knows when to perform the various operations that it has been assigned, because it has been programmed to do so.

To implement certain control applications, the micro must also perform a monitoring function. The micro "looks at" the process or devices being controlled to see what is happening. For example, the computer can monitor switch closures to determine physical state or position, pressure, and many other parameters. Transducers (sensors) convert the monitored physical characteristics, such as temperature or light level, into electrical signals that the computer can understand and respond to (see Fig. 6.2). In most microcontrollers or PLCs, the controlled output change takes place only if a specific input condition is sensed or not sensed. For example, if a temperature sensor indicates that the temperature has risen to a specific level, the computer will turn on a cooling fan.

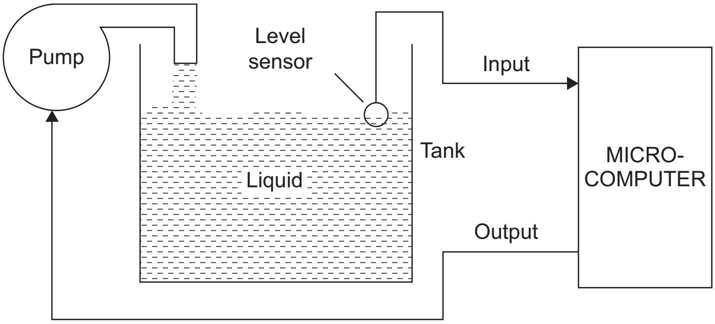

FIGURE 6.3 Controlling liquid level in a tank automatically with a microcomputer.

Micros sometimes change the control function in response to one or more of the inputs they are monitoring. For example, if the micro senses that the liquid in a tank exceeds a given level, it can automatically turn off the pump that is filling the tank. If the liquid level goes below a predetermined level, the micro detects this and automatically starts the pump (see Fig. 6.3). The key point here is that the computer makes its own decisions based on the input data that it receives. The result is full automation of some process or device.

The micro's ability to monitor and control operation may be used to automate simple processes in a toy or to operate entire factories.

COMPUTER ORGANIZATION AND OPERATION

All microcomputers are made up of four basic sections: memory, control unit, arithmetic logic unit (ALU), and I/O unit. A general block diagram of a microcomputer showing these four sections is given in Fig. 6.4. These four sections communicate with one another over multiple parallel electrical conductor data paths called a bus, as shown, to process the data or perform a control function.

Note that the control and ALU are shown together in a common structure. This is the CPU. The CPU is, of course, usually a microprocessor.

The memory is that part of the computer where data and programs are stored. The memory in any computer may contain thousands, millions, or even billions of locations used for storing numbers, words, or other forms of information.

Two primary types of information are stored in computer memory. The first type is the data to be processed. These are the codes, numbers, letters, and other forms of data to be manipulated.

The other type of data stored in memory is an instruction. Instructions are special binary numbers or codes that tell the computer what to do. Instructions specify the ways in which the data is to be processed. For instance, there are instructions that cause arithmetic operations to take place or data to be transferred from one place to another.

The instructions listed in a special sequence form a program. A program is a list of instructions that causes data to be processed in a unique way. A program

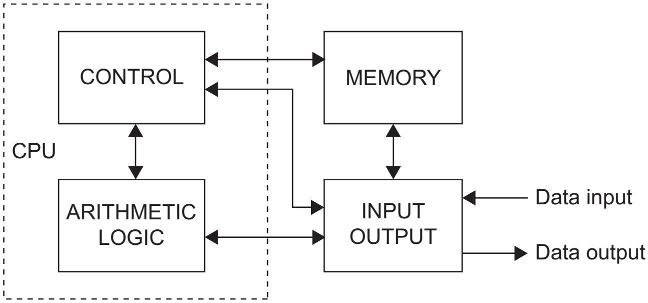

FIGURE 6.4 General block diagram of any digital computer or microcontroller.

is a step-by-step procedure that solves a problem, performs a control operation, or otherwise manipulates data according to some recipe. Programs are stored in memory along with the data that they process. In some micros there are separate memories for data and instructions. The programs that a computer uses are called software.

Data and instructions can be stored in or retrieved from memory during processing. When data is stored, we say that it is written into memory. When data is retrieved, we say that it is read from memory. A typical write operation transfers data from an analog-to-digital converter and stores it in memory. A common read operation accesses data in memory to be transferred to a liquid crystal display (LCD) screen.

Now let us look at the operation of the CPU. The control unit is that portion of the digital computer responsible for the automatic operation. The control unit sequentially examines the instructions in a program and issues signals to the other sections of the computer that carry out designated operations.

Each instruction is fetched (read) from memory by the control unit, interpreted, and then executed one at a time until the program is completed. This is called the fetch–execute cycle, which is repeated on each instruction until the program runs to completion. The execution of each instruction may call for accessing one or more data words in the memory or storing a data word in memory.

The ALU is the section of the computer that carries out many of the functions that are specified by the instructions. In other words, the ALU actually processes the data. Specifically, the ALU carries out two main types of processing: arithmetic operations (such as addition, subtraction, multiplication, and division) or logic operations (such as AND, OR, complement, or exclusive OR). For example, if an add instruction is stored in memory, the control section will fetch it, interpret it, and send signals to the ALU that cause two numbers to be added.

The ALU also performs data movement operations. It can move data or instruction words from one place to another inside the CPU, or it can carry out memory read/write or input/output operations. These are called load and store instructions.

A key part of the CPU associated with the ALU is the registers. Most ALUs have two or more registers. High-powered CPUs may have a group of 16 or more registers, called *general purpose registers* (GPRs). The registers are used to temporarily store the data being processed by the ALU and the results of the computations. Other registers in the CPU store a number called an *address* where the data or instructions are stored, store the instruction being executed, or act as a stop-off place for data into or out of the CPU.

The control section, ALU, and the registers in all micros are very closely related. They operate together and are always considered as a single unit. As indicated earlier, the combination of the control and the ALU sections is called the CPU (see Fig. 6.4). In addition, microprocessors are single-chip CPUs. Besides being called CPUs, microprocessors are sometimes called MPUs or *microprocessing units*. You will also see the expression μP used to refer to a microprocessor. The μ is the Greek letter mu, which means micro; the *P* means processor.

A microcomputer consists of a microprocessor (the CPU) plus external memory and I/O circuits. Or it could be an embedded controller with everything on one chip. The memory is either SRAM or DRAM for data and programs, and EEPROM or flash for the program.

The I/O unit of the computer is the set of logic circuits that permits the CPU and memory to communicate with the outside world. The I/O circuits are referred to as *interfaces*. All data transfers into and out of the computer pass through the I/O interfaces.

The external peripheral devices, or *peripherals*, connected to the I/O unit are electronic or electromechanical units that are used for data entry or data display. Data is most commonly entered into the computer through an input keyboard or a disk drive.

The output data is usually displayed on a video monitor or small LCD screen. “Hard copy” on paper is created by a printer. There are a wide variety of other external input/output devices such as scanners, voice synthesizers, and barcode readers.

In an embedded microcontroller, the peripheral devices include keyboard, ADC, DAC, sensors, magnetic stripe readers, barcode scanner, and so on. Outputs go to relays, motors, lights, and LCDs.

There are two basic ways that microcomputers are organized. One is called the Von Neumann architecture, and the other is called the Harvard architecture. The Von Neumann architecture is the older basic design we have discussed so far. Its main characteristic is that both the program instructions and the data to be processed are stored in a common memory. The Harvard architecture features separate memories for data and programs. Older micros use the original Von Neumann architecture, but the newer microcontroller chips use the Harvard architecture. Data memory is usually volatile RAM while the program or instruction memory is larger and usually nonvolatile EEPROM or flash.

OPERATIONAL DETAILS

Registers and the Arithmetic Logic Unit

The main circuit element in a CPU is the register. Usually the data to be processed by the CPU is taken from memory and stored temporarily in a register. Like a memory location, a register usually has a fixed length such as 8, 16, or 32 bits. But unlike a memory location, registers are often used to manipulate data as well as store it. That is, the register can alter or process the data in some way.

Examples of how a register can process data are shown in Fig. 6.5. A binary word in a register can be shifted one or more bit positions to the right or left, or the register may be connected as an up/down counter, so that it can be incremented (add one to the content) or decremented (subtract one from the content). A register can also be reset or cleared, thereby erasing any data in it and leaving the content zero.

Data transfers and manipulations performed on a register are initiated by individual computer instructions. For example, one instruction may cause the register to be loaded from a memory location. Another instruction may cause the word in the register to be transferred to a memory location. All data transfers are parallel, meaning that all bits are moved simultaneously from the source to the destination.

The main working register in most digital computers is called the *accumulator*. It may also be called the *A register*, *W register*, or something else. The accumulator can hold one word whose bit length is equal to that of a memory location. An 8-bit microprocessor has an 8-bit accumulator. The accumulator is part of the ALU.

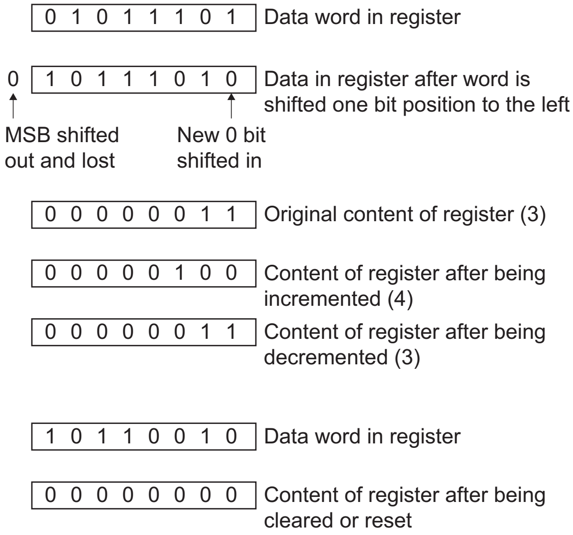
0 1 0 1 1 1 0 1
Data word in register
0
1 0 1 1 1 0 1 0
Data in register after word is shifted one bit position to the left
MSB shifted out and lost
New 0 bit shifted in
0 0 0 0 0 0 1 1
Original content of register (3)
0 0 0 0 0 1 0 0
Content of register after being incremented (4)
0 0 0 0 0 0 1 1
Content of register after being decremented (3)
1 0 1 1 0 0 1 0
Data word in register
0 0 0 0 0 0 0 0
Content of register after being cleared or reset

FIGURE 6.5 How data is manipulated in a register. Examples of shifting data, incrementing and decrementing a register, and clearing (resetting) a register to 0.

The accumulator can be loaded from memory or the accumulator content can be stored in any memory location. I/O operations with peripheral units sometimes also take place through the accumulator.

Data to be processed by the arithmetic/logic section is usually held in the accumulator. This data is fed to the ALU. The ALU is a digital logic circuit that can add, subtract, and perform a wide variety of logic operations.

The ALU is capable of processing two inputs, one from the accumulator and one from the memory of another register (see Fig. 6.6). Let us assume that you wish to add two binary numbers. To do this, you first load one of the numbers (the *augend*) into the accumulator. This is done with a load accumulator instruction that takes the number from a memory location and puts it into the accumulator. Next, an ADD instruction is executed. This causes another word (the *addend*) to be taken from memory and placed in register B. It is then added to the content of the accumulator. The sum is usually stored back in the accumulator, replacing the augend originally there.

Some microprocessors have two or more accumulator registers that share a single ALU. Two accumulators provide greater flexibility in the manipulation of data than a single accumulator. Such multiple registers simplify, speed up, and shorten a program to perform a given operation. Some sophisticated CPUs have 4, 8, 16, or even more accumulators. Usually they are referred to as GPRs or a register file. These registers can each use the ALU and act as temporary storage locations for data and the intermediate results of calculations. Instructions are provided to move data from one register to another.

In most micros, the GPRs share a single ALU. The ALU accepts only two input words and generates a single output word. The two words to be processed

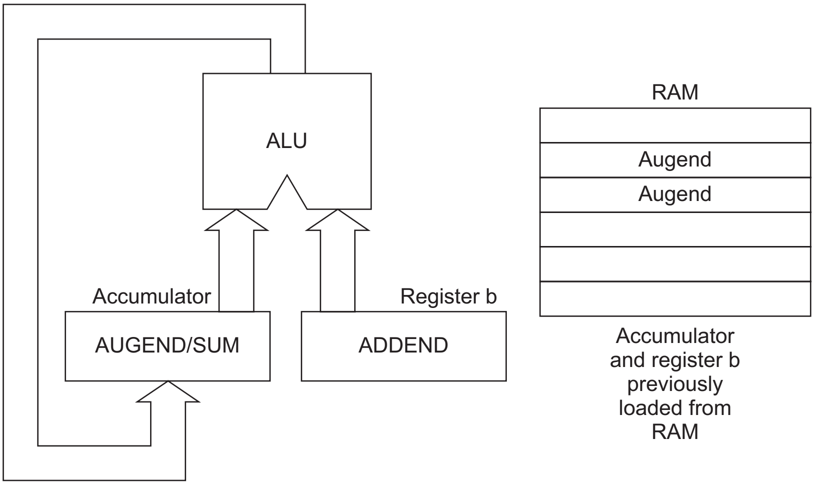

FIGURE 6.6 How ALU adds or otherwise processes two data values.

by the ALU usually come from any two GPRs or any GPR and a designated memory location. The destination of the ALU output can also be one of the GPRs in some CPUs.

Control Unit

CPUs also have several other registers, including the instruction register (IR); the program counter (PC), also called the instruction counter; and the memory address register (MAR), also called the address buffer.

The IR is used to store the instruction word. When the CPU fetches an instruction from memory, it is temporarily stored in the IR. The instruction is a binary word or code that defines a specific operation to be performed. The instruction word is also called the *op code* or *operation code*. The CPU decodes the instruction, and then executes it.

The PC is really a counter and a register. It stores a binary word that is used as the address for accessing the instructions in a program. If a program begins with an instruction stored in memory location 43, the PC is first loaded with the address 43. The address in the PC is applied to the memory, causing the instruction in location 43 to be fetched and executed. After the instruction is executed, the PC is incremented (add 1) to the next address in sequence, or 44. The instructions in a program are stored in sequential memory locations.

The MAR or *address buffer* also stores the address that references memory. This register directly drives the address bus and the memory address decoder in RAM or ROM. The MAR gets input from the PC when an instruction is to be accessed (see Fig. 6.7). The MAR can also be loaded with an address that

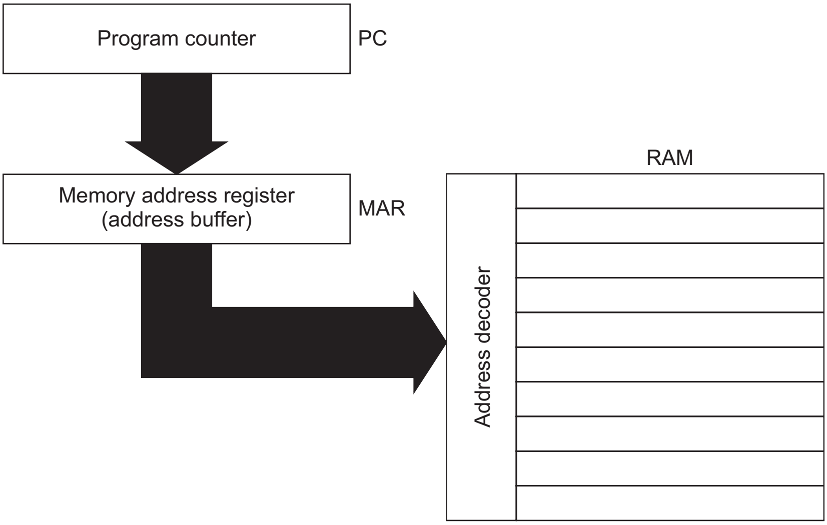

FIGURE 6.7 Program counter identifies the address in memory to be accessed.

is used to access data words stored in memory. To retrieve a data word used in an arithmetic operation, the MAR is loaded with the binary word that points to the location of that word in RAM. This address is often a part of the instruction.

It is important to note that the PC and the MAR (address buffer) have a fixed length of so many bits. And that limits the amount of memory that can be accessed. For example, with a 16-bit address register, the address bus has 16 bits to address RAM and ROM. With 16 bits, a maximum of $2^{16}=65{,}536$ words can be addressed.

There are usually two other registers, the *flag* and *stack pointer registers*. The flag or *F register* is an 8-bit register whose individual flip-flops are set and reset by the ALU as the various arithmetic and logic operations are carried out. Each flip-flop is called a flag. As an example, there are *zero* (Z) and *carry* (C) *flags*. If the accumulator content is zero after an operation is performed, the Z flag is set indicating this condition. If an arithmetic operation (addition) results in a carry from the most significant bit (MSB) of the accumulator, the C flag is set indicating this condition. These flags can be monitored or tested by the control circuitry to change the sequence of processing.

The stack register is a 16-bit or larger register used to address a selected area of RAM known as the *stack*. This memory is used to store register contents and status information when subroutines and interrupts are used.

Instruction-Word Formats

There are three types of instruction formats used in typical 8-bit microprocessors, illustrated in Fig. 6.8. In the single-word format, the instruction is a single 8-bit word. This word is called the *op code*. The op code tells the ALU, the registers, and other elements of the system what to do. In this format, no address is used. The focus of the instruction is implied in the instruction. That is, the data to be processed is already in a location designated by the instruction. Usually the data is in a register. Typical instructions using this format are register-to-register transfer, shift data left (or right), or clear to zero.

The two-word instruction format in Fig. 6.8 requires two 8-bit words to define the operation. These two words are stored in sequential memory locations. The first word is the op code. The second word is usually an address that specifies a memory location where the data word to be processed is stored. For example, if the op code calls for an add operation, the address word designates the location in RAM of the number to be added to the contents of the accumulator. The 256 bytes of RAM can be addressed with 1 byte of address.

In some 2-byte instructions, the second byte is not the address. Instead, it is the data itself. This is called an immediate instruction since it is not necessary to address the data that is available immediately within the instruction itself.

The three-word instruction format in Fig. 6.8 is comprised of an 8-bit op code and two 8-bit address words stored in sequential memory locations. The second and third bytes together form a 16-bit address word that designates the location in RAM of the data to be processed. In the 3-byte instruction format,

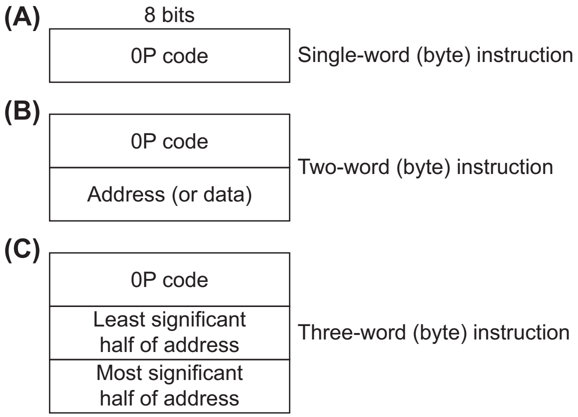
(A)
8 bits
0P code
Single-word (byte) instruction
(B)
0P code
Address (or data)
Two-word (byte) instruction
(C)
0P code
Least significant half of address
Most significant half of address
Three-word (byte) instruction

FIGURE 6.8 Three common instruction-word formats with addresses or data. *Op code*, operation code.

the first byte is the op code, the second byte is the least significant half of the address, while the third byte is the most significant part of the address. Different formats are used by other micros.

To access a word in RAM, the instruction address word must be stored in the MAR. This happens during the instruction fetch operation. When an instruction is fetched from memory, the op code is stored in the IR while the address is stored in the MAR. The instruction is then executed. The MAR usually gets its input from the PC. Once an instruction is fetched and executed, the PC is incremented.

The PC may be incremented once, twice, or three times, depending on the length of the instruction just executed. If a 2-byte instruction is executed, the PC is incremented twice so that the PC points to the address of the next instruction op code.

Program Execution Example

Now let us illustrate how a typical CPU executes a simple program. Here are a few common instructions similar to those found in any micro.

Move A to B (MOV B,A): This is a 1-byte instruction that causes the content of the accumulator (A) to be moved to register B. The content of the accumulator is not erased.
Load Accumulator (LDA): This is a 3-byte instruction that takes the content of the address specified by the second and third bytes and loads it into the accumulator.
Add Immediate (ADI): This is a 2-byte instruction that takes the content of the second byte of the instruction and adds it to the content of the accumulator. The sum is stored in the accumulator, replacing the number that was there previously.
Halt (HLT): Halt is a 1-byte instruction that stops processing.

Note that we refer to each instruction by a three-letter abbreviation called a *mnemonic*.

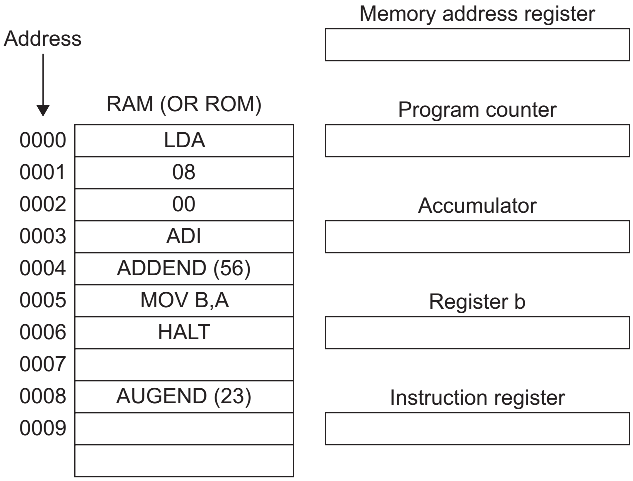

FIGURE 6.9 Example of program stored in RAM for execution using various CPU registers.

Fig. 6.9 shows a complete CPU and the RAM containing a simple program. The first instruction in the program is the LDA stored in location 0000.

To begin the execution, the address of the first instruction is loaded into the PC that points to the location of the instructions in RAM.

Next, the content of the PC is transferred to the MAR that addresses RAM. The LDA instruction op code is fetched and stored in the IR. The address bytes of the LDA instruction in locations 0001 and 0002 are then transferred to the MAR. The address of the data to be loaded into the accumulator in bytes 2 and 3 is 0008. The desired data word in location 0008 is retrieved and loaded into the accumulator. This is the augend (23).

After the LDA instruction is executed, the number 23 is stored in the accumulator. Next, the PC is incremented three times. The ADI instruction is then fetched and stored in the IR. The number to be added is stored in location 0004.

This is an Add Immediate instruction where the second byte contains the data, in this case the addend (56). The number in location 0004H is 56, which is added to 23 in the accumulator, creating the sum 79. The mnemonic of the next instruction to be executed is MOV B,A in location 0005.

The sum in the accumulator is then moved to the B register by the MOV B,A instruction. The content of the accumulator is not changed when data is transferred to the B register.

The PC is incremented once since the HLT is a 1-byte instruction.

As you can see by this step-by-step analysis of the execution of a computer program, there is nothing mysterious about its operation. The CPU simply fetches and then executes the sequentially stored instructions at very high speed until the operation is complete. This process is sometimes referred to as the *stored program concept*.

Microcomputer Buses

Data transfers inside a microprocessor take place in parallel. This means that all bits in a word are transferred simultaneously from one place to another. It takes only a few nanoseconds (one billionth of a second) for all data bits in one register to be moved to another register.

The parallel data transfers take place over a data bus. A bus is simply multiple parallel electrical connections from a source to a destination. A number of these buses are contained within the CPU and are known as internal buses.

Microprocessors usually have three major buses—a *data bus*, an *address bus*, and a *control bus*. These are made available to external circuits in some micros. A typical 8-bit CPU has an 8-bit data bus and a 16-bit address bus. The data bus sends data to and from the CPU, RAM, ROM, and I/O sections. All data transfers between the CPU and memory or I/O sections take place over the data bus. The address bus drives all of the memory and I/O devices.

When an instruction is fetched from RAM or ROM, it is transferred over the data bus from the memory into the IR. Any data word retrieved from memory or a peripheral device via the input section also passes over the data bus into the accumulator or GPR.

When a store instruction is executed, the word in the accumulator is transmitted over the data bus into RAM. Data can also be transferred from the accumulator over the data bus to an external device such as an LCD. Or, data from an input device such as a keyboard passes over the data bus and is placed into the accumulator. The important point here is that data can move in either direction over the data bus. We say that it is a *bidirectional bus*.

Another key point is that the data bus can be connected to only one data source at a time. Data can originate only at a single source, but it can be sent to one or more destinations. To accomplish this, circuits called *bus multiplexers*, or *three-state line-driver circuits*, are used to connect or disconnect the various data sources to or from the bus. Fig. 6.10 illustrates this concept.

The address bus is a unidirectional bus. It transfers an address from the CPU to all external circuits (memory and I/O). Address words are produced in the CPU. The PC generates the address that points to the instruction to be fetched. The content of the PC is transferred over a parallel 16-bit internal address bus to the MAR or address buffer. The output of the MAR or address buffer is the address bus.

The control bus consists of numerous signals generated by the microprocessor and used to initiate various memory or I/O operations. The control bus also contains input lines from external circuits that tell the CPU what to do and when. The number of control signals varies from CPU to CPU.

POPULAR MICROCOMPUTERS

The first practical microprocessors and microcomputers became available in the mid-1970s. Over the years lots of different models have been developed. As semiconductor processing technology has improved during the past 40 years, it

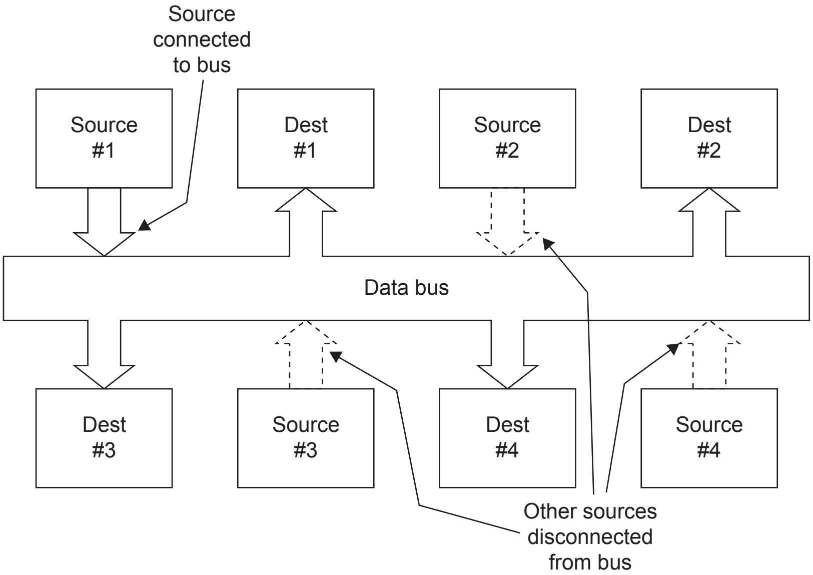

FIGURE 6.10 Only one source may transmit at a time on a bus. Others are disconnected. All destinations can receive the transmitted data. *Dest*, destination.

has become easier and cheaper to put more circuits on a chip plus the memory that earlier was always external. And processing speeds have multiplied many times over. Today's embedded controllers are indeed complete computers on a chip and they run at very high clock rates. And their cost is low. In high-volume manufacturing, an embedded controller can cost less than $1.

Most of the initial processors were 8-bit devices. These are still popular today. Larger, faster processors are available for not much additional cost. 16- and 32-bit processors are commonplace. Even 64-bit devices are available if needed. Many are individual chips, while others are available only as cores on a larger systems chip. Following is a summary of a few of the most widely used processors.

8- and 16-Bit Microcontrollers

8051

The 8051 is one of the oldest but still widely used. It was originally developed by Intel in the late 1970s and 1980s. It is still around and available from a half dozen or more companies. The more recent versions are faster and have more memory and I/O features than the original. Original clock speeds were only 8–12 MHz, but today you can get models with a clock rate up to 100 MHz.

A general block diagram of the 8051 is given in Fig. 6.11. The 8051 features an internal RAM of 128 or 256 bytes but can handle 64 or 128 KB

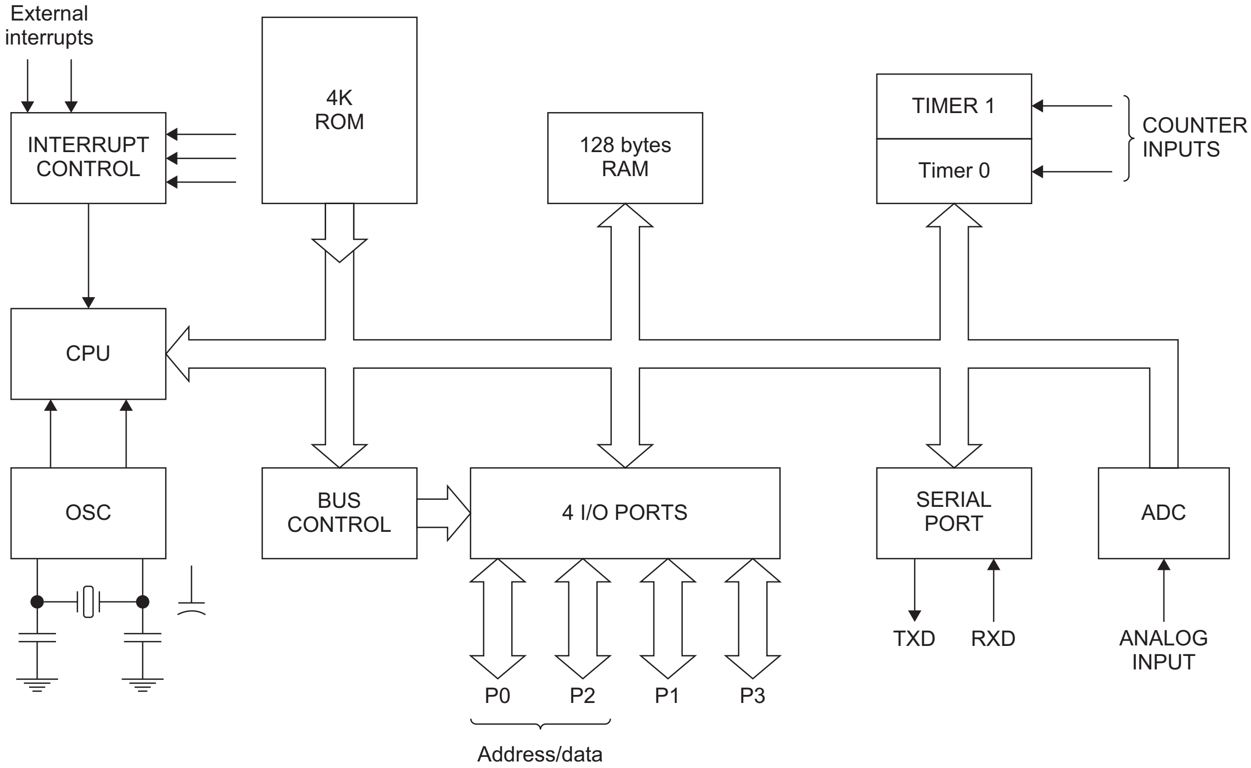
External interrupts
INTERRUPT CONTROL
4K ROM
128 bytes RAM
TIMER 1
Timer 0
COUNTER INPUTS
CPU
OSC
BUS CONTROL
4 I/O PORTS
SERIAL PORT
ADC
P0
P2
P1
P3
Address/data
TXD
RXD
ANALOG INPUT

FIGURE 6.11 General block diagram of 8051 microcontroller. It is very similar to most other controllers in terms of major circuits and their interconnections.

of RAM externally. The internal ROM can be had from 512 bytes to 128 KB. Some models also feature flash memory. As for I/O ports, the figure shows four 8-bit parallel I/O ports. You can get a standard serial port called a UART, and more specialized I/O ports such as SPI and I^2C. See the Interfaces sidebar in this chapter. Many models also come with an internal 10- or 12-bit ADC.

A few other items in this diagram that were not previously discussed are the clock, interrupts, and timers. The *clock* is an oscillator that controls all operations in the micro. Its frequency is usually set by an external crystal that maintains a precise and stable frequency. The higher the frequency, the faster the operations. Fast is good.

An *interrupt* is, as its name implies, an external signal that can interrupt the CPU when an important event occurs. It can signal the CPU of the event and even cause the CPU to stop executing the current program and jump off to another program to deal with the interrupt.

Timers are usually just one or more up/down counters that can be programmed with various count values. These are used to implement count functions in programs and to provide timing operations for I/O operations. Most controllers have one or more timers.

PIC Processors

Microchip Technology makes one of the most widely used 8-bit controllers. Called PIC processors, they are extremely simple and cheap. There are hundreds of versions with different memory sizes and configurations and various mixes of interfaces. One of the simpler versions is the 16C54. It has a single working register, W, that is the accumulator. It uses a bank of 24–73 registers for RAM. ROM for program storage ranges from 512 bytes to 2 KB. I/O occurs via 4- or 8-bit parallel ports. Larger versions have more RAM and more ROM, higher clock speeds, a built-in ADC, and special network interfaces like CAN. Several 16-bit versions are also available.

MPS430

The MPS430 is a line of 16-bit microcontrollers from Texas Instruments. It has 16 GPRs and a maximum clock speed of 25 MHz. There are multiple versions each with a different mix of memory types including up to 512 KB of flash and 64K of SRAM. Some models include a new type of nonvolatile memory called ferroelectric random access memory (FRAM). FRAM uses a cell similar to a standard DRAM cell that stores bits in a capacitor. Using a ferroelectric dielectric in the capacitor allows data to be retained when power is removed. The MPS430 has an ADC on board, a DAC, timers, PWM outputs and all the common serial interfaces. It is a popular chip because its ultra low power consumption makes it ideal for portable and battery-powered devices.

32- and 64-Bit Processors

As it turns out, it is almost as cheap and easy to make a 32-bit processor as it is to make an 8-bit device. For that reason, for a few cents more, designers can make their products faster and more powerful by selecting a 32-bit device. With this many bits being processed at the same time, it is possible to move more data faster, compute with increased resolution, and to perform much more complex operations faster and easier. 32-bit processors can also address much more memory. A few years ago, 32-bit processors were reserved for only the largest or most critical of needs. Today, the 32-bit processor is almost as widely used as the 8-bit processor.

Three devices account for most applications. They are made by ARM Holdings, NXP, and MIPS. ARM is the name of a British company that invented a supersimple, fast, and flexible 32-bit processor that does everything well at very low cost. It is referred to as a reduced instruction set computer (RISC) processor. It has a few simple instructions but is extremely fast. ARM does not actually make chips, but instead licenses the design to others. There are several vendors of ARM processors but most of them are actually cores that are integrated on chip with other circuits to form a system on a chip. With over 75% of the market share, you are sure to encounter an ARM device sooner or later. Chances are you own at least one inside a product such as a cell phone, which is where most are used.

NXP's PowerPC is another popular 32-bit processor. Many embedded versions have since been made. Its primary usage is in communications, networking, automotive, and industrial applications.

For really super processing power, the MIPS processors are a popular choice for certain embedded applications. These devices come in 32-bit as well as 64-bit models, in case you really have a hot application for which nothing else will do. The MIPS processors are great at advanced math and have excellent resolution.

Microcontroller Interfaces

An *interface* is the circuitry that connects an embedded controller to the outside world. Initially, all I/O in early micros was just the parallel data and address buses extended to accommodate additional I/O circuitry. Today, much of the I/O interface circuits are put right on the same chip with the rest of the micro. Following are the most widely used interfaces you are likely to encounter. All of these are serial interfaces. Parallel I/O is still used, but today the preference is serial I/O because of fewer interconnecting lines.

RS-232/UART—This universal asynchronous receiver–transmitter is actually the physical circuitry implementing a popular serial interface standard referred to as RS-232. It was originally developed for connecting computers to teletype machines and early video terminals. It is still used today, especially for connecting to industrial equipment for monitoring or control. You will hear this interface called a serial port.

Continued

Microcontroller Interfaces—cont'd

The RS-232 interface sends and receives data in bytes (8 bits) plus two extra bits, called Start and Stop bits, that tell the receiving device when the data begins and ends (see Fig. 6.12). The logic levels are −3 to −25 V for a binary 1 and +3 to +25 V for a binary 0. Typical voltage levels are in the 5- to 12-V range. The data rate is slow, typically from 300 bps to 115.2 kbps. Standard data rates are used, so no clock signal is needed. Common speeds are 2400, 9600, and 19,200 bps. The physical communications medium is a two-wire (signal plus ground) cable that can be up to 50 ft long. Special 9- and 24-pin connectors are part of the standard. Many PCs and laptops still have a 9-pin RS-232 port.

I^{2}C—Shorthand for interintegrated circuit or I squared C, this is a low-speed serial bus that is used for sending data from one chip to another on the same PC board or over short cables between two pieces of equipment. The data rate is typically in the 10–400 kbps range, but a faster version (3.4 Mbps) is available. Speeds rarely exceed 100 kbps. Only three wires are needed, data, a clock signal that controls the data transfer, and a ground.

SPI—This is the serial peripheral interface. It uses four wires, data in, data out, clock, and a select signal that determines whether a device is a master or slave and either sending or receiving data. A fifth connection is ground. It is much faster than the I^{2}C interface, with speeds up to 20 Mbps.

CAN—This is controller area network serial bus in an interface used to link micros together in small networks. It is used in automotive and industrial applications. The usual cable is twisted pair and a ground. Speeds range from 10 kbps to 1 Mbps with 20 kbps being typical. The data is sent in specifically formatted frames as determined by a standard protocol. It uses start and stop bits like the RS-232 and some similar control codes defined in ASCII. It also includes error detection and correction capability, so it is very reliable.

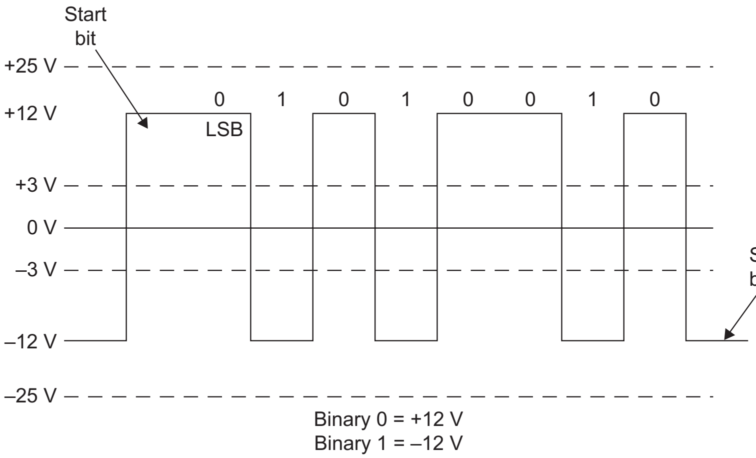

FIGURE 6.12 ASCII letter "j" being sent over an RS-232 interface using a UART.

Continued

Microcontroller Interfaces—cont'd

USB—The universal serial bus is the most common PC interface for peripherals including keyboard, mouse, printer, flash drive, external disk, and so on. It is also used in some embedded micros. The interface uses two wires but also includes 15 V DC and ground. Version 2.0 provides speeds to 480 Mbps over cables no longer than 16 ft. A new higher-speed version 3.0 is also available with speeds up to 5 Gbps. Data is sent in bytes in packets according to a unique protocol.

RS-485—This is an interface widely used in industrial applications. It uses twisted pair cable that can be up to 4000 ft long. The cable is actually a bus to which many devices can be attached. Data rates can be up to 10 Mbps for shorter lengths. Various protocols may be used.

DIGITAL SIGNAL PROCESSING

Digital signal processing (DSP) is the technique of programming a microcomputer to perform operations normally carried out by analog or linear circuits. The method is illustrated in Fig. 6.13. Assume you want to process an analog signal by filtering. You could use a hardware filter made up of inductors and capacitors or resistors and capacitors. Another way to do it is with DSP. The analog signal to be filtered is first digitized with an ADC and the resulting data stored in an RAM. This data is then acted upon by a processor programmed to perform a filter function. Special mathematical algorithms have been developed to do this. The resulting calculations generate new data that is also stored in the RAM. This new data is then sent to a DAC where it is converted back to an analog signal. The result is that the signal is filtered

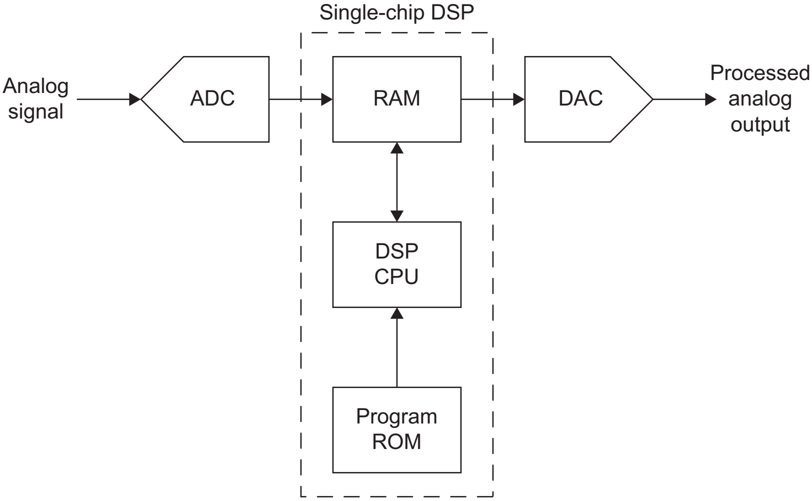

FIGURE 6.13 Basic makeup of digital signal processor and its ADC and DAC I/O.

as if were processed by an electronic circuit of capacitors and inductors. The processing was mathematical.

There are many other operations that can be performed by DSP. Filtering is the main application, but you can also do things like modulation and demodulation as needed in wireless applications. It will perform other things such as data compression, equalization, tone control, and even frequency analysis. Using a program called the fast Fourier transform, the DSP can analyze a signal and tell you exactly what frequency components are in it. It identifies the harmonic frequencies and their various amplitudes and phases just as in a frequency–domain plot.

The big question is why use a complicated technique such as DSP when you can do it with simple analog/linear hardware? Probably the best answer is “just because we can.” With cheap single-chip processors, DSP is often no more expensive than an equivalent analog circuit. Since everything else is digital these days, it makes some sense to do as many analog functions digitally as well. But the best reason is that DSP provides superior results over analog methods. Filtering is substantially improved with better selectivity that was not possible with analog filters.

While DSP can be programmed on any processor or controller, it is best done on special DSP chips designed for the purpose. They use special architectures and processing schemes to speed up and improve the processing over standard conventional processors. A good example is that DSP chips implement a multiply and add or accumulate function, called MAC, that is common to most DSP processes. The MAC has to be programmed in a standard processor, but in a DSP chip it is implemented with special hardware that speeds up the process and simplifies programming. Some more conventional processors also incorporate DSP instructions to facilitate DSP operations. It is also common to implement DSP in an FPGA or in custom digital logic inside a larger IC.

Today, DSP is found in most electronic products today, such as TV sets, cell phones, MP3 players, cable modems, and many other common products.

Chapter 7

Radio/Wireless

The Invisible Cables of Modern Electronics

In this Chapter:

- The oldest electronic application.
- What are radio waves?
- The electromagnetic frequency spectrum.
- The hardware of radio: transmitters, receivers, and antennas.
- Two-way radio applications.
- Satellites.
- Radio telescopes.

INTRODUCTION

First they called it wireless. Then they called it radio. Today, the term *wireless* is trendy again. Whatever you call it, radio is communication over a distance without any physical connection. Radio is cool because it is completely silent and invisible. It can occur over a distance of only a few feet, all the way around the world, or between planets and spacecraft. Radio, despite the fact that it has been around for over a century, and the fact that we take it for granted on a grand scale, is truly black magic. When you finish reading this chapter, you will know how this magic works.

RADIO COMMUNICATION SYSTEMS

Radio is an electronic communications system. All electronic communications systems have the basic form shown in Fig. 7.1, whether it is a wired or wireless system. The information to be communicated, usually voice, is first converted into an electrical signal by a microphone. That signal is then processed by the circuits in a transmitter (TX) and then sent via a communications medium to a receiver (RX) that processes the signal and converts it back into voice by a speaker or headphones. In the telephone system, the communications medium is a cable made of copper wire. In a cable TV system, the medium is a coaxial cable, twisted-pair cable, or fiber optic cable. But in radio, the communications

Electronics Explained. http://dx.doi.org/10.1016/B978-0-12-811641-8.00007-2
Copyright © 2018 Elsevier Inc. All rights reserved.

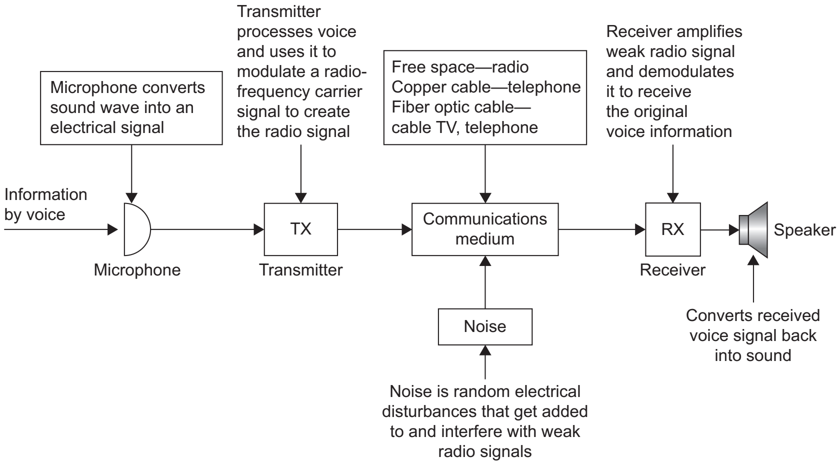
Microphone converts sound wave into an electrical signal
Transmitter processes voice and uses it to modulate a radio-frequency carrier signal to create the radio signal
Free space—radio
Copper cable—telephone
Fiber optic cable—cable TV, telephone
Receiver amplifies weak radio signal and demodulates it to receive the original voice information
Information by voice
TX
Communications medium
RX
Speaker
Microphone
Transmitter
Receiver
Noise
Converts received voice signal back into sound
Noise is random electrical disturbances that get added to and interfere with weak radio signals

FIGURE 7.1 Model of electronic communications system.

channel is free space. The radio signal travels from one place to another almost instantaneously, and it passes right through most obstructions such as walls, buildings, smog, trees, and even people. Note that noise is added to the signal as it passes through the channel or medium.

Noise

Noise is any random electrical voltage that interferes with the weak radio signal. Noise comes from the atmosphere, space, human-made objects, and electronic components. Atmospheric noise is lightning. Space noise comprises random signals from the Sun and remote stars. Human-made noise comes from electrical power generation, distribution, and switching. It also comes from fluorescent lights, motors, auto ignitions, and many other sources. The worst noise comes from the electronic components that actually process the weak radio signal. Resistors, transistors, and integrated circuits generate random noise because of thermal agitation. Heat produces random movement of the electrons in the components, and that produces a noise voltage which is added to the radio signal. Even though noise voltages are very small (microvolts or nanovolts), they are still large enough to interfere with and in some cases obliterate the very small received radio signal. In most radio systems, it is noise that limits intelligibility, transmitted data speed, and transmitted distances.

THREE BASIC WAYS OF COMMUNICATIONS

Radio communications systems work in one of three basic ways: simplex, half duplex, and full duplex (see Fig. 7.2).

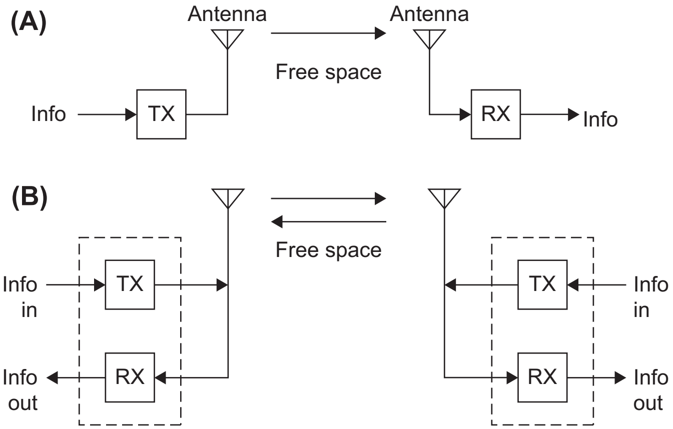

FIGURE 7.2 Ways of communicating. (A) Simplex—one way and (B) duplex—two way: half—alternating send/receive and full—simultaneous send/receive.

Simplex

Simplex is one-way communications. Examples are radio broadcasting, paging, and remote control. A radio telescope is also one-way: there is a transmitter at one end and a receiver at the other.

Half Duplex

Duplex means two-way, and most radio communications are really two-way affairs. There is a transmitter and a receiver at both ends of the communications system. Both parties can send and receive. And there are two types of duplex systems: half duplex and full duplex.

In half-duplex systems, only one party can transmit at a time. Those communicating with one another take turns talking. This is the kind of radio where you have to say “over” or “come back” after you finish speaking to let the other party know that it is time for her or him to talk. Citizens band (CB) radio is like this. Most two-way communications services (police, fire, taxi, aircraft, marine, etc.) use this method because it is cheap and simple. Fax machines are half duplex. Most computer modems are half duplex.

Full Duplex

Full duplex is more complex and expensive. This is two-way radio where both parties can both send and receive simultaneously. Probably the best example of this is a cordless or cellular telephone. It is great to be able to talk and listen at the same time.

WHAT IS A RADIO WAVE ANYWAY?

Any discussion of radio always gets around to this basic question. How do you describe something that you cannot see, feel, or hear? You imagine it.

Electric and Magnetic Fields

A radio wave is an invisible force field that is made up of both a magnetic field and an electric field. In combination, we call them an *electromagnetic field*. We cannot see or feel them, but we can experience their effects. In Chapter 2, electric and magnetic fields are described briefly. To review, a magnetic field is a force generated by a magnet, either a permanent magnet or electromagnet. Current flow in a wire or other conductor produces the magnetic field. Invisible magnetic flux lines emanate from the magnet and may influence objects in or near them.

An electric field derives from a voltage, or what we call a *potential*, across two conductors. The attraction of the positive and negative charges across an open space introduces another kind of invisible field that can also influence external objects.

Now if you combine these two types of fields into a self-supporting entity, you would have an electromagnetic field or *radio wave*. We create a radio signal with the circuits in a transmitter, but it is the magic of a device called an *antenna* that produces the combined electric and magnetic field that we refer to as a radio wave.

In the 19th century, a British scientist named James Clerk Maxwell figured out that if you generate a changing electric field, it would, in turn, generate a magnetic field. If the magnetic field is changing and moving forward, it will generate an electric field. The two fields support or regenerate one another as they change and move outward from the antenna that produces them. Maxwell expressed his theories in his famous mathematical equations that form the foundation of radio and light transmission.

Trying to draw a picture of a radio wave is somewhat like trying to draw a picture of the wind. Nevertheless, Fig. 7.3A is an attempt. Try to think in three dimensions. Current flow in the antenna wire creates the magnetic field lines. The voltage across the two segments of the antenna produces the electric field lines. The picture is somewhat simplified if we just show the field lines as in Fig. 7.3B. The electric and magnetic field lines are always at right angles (90 degrees) to one another. The radio wave moves in a direction at a right angle to both field lines, in this case either into or out of the page. As it gets some distance from the antenna, it becomes self-supporting. That is, the magnetic field sustains the generation of the electric field and vice versa.

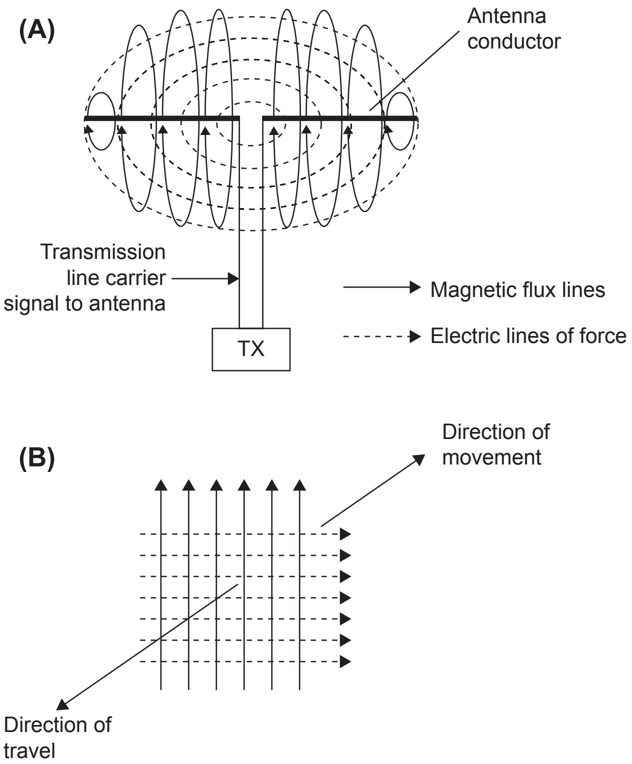
(A)
Antenna conductor
Transmission line carrier signal to antenna
TX
Magnetic flux lines
Electric lines of force
(B)
Direction of movement
Direction of travel

FIGURE 7.3 How an antenna creates a radio wave. (A) Antenna produces both electric and magnetic fields. (B) The electric and magnetic field lines are at a right angle to one another and to the direction of radiation.

Polarization

Polarization of a radio wave means how it is oriented with respect to the surface of the Earth. If the electric field is vertical to the surface of the Earth, the wave is vertically polarized. If the electric field is horizontal to the Earth, the wave is horizontally polarized. It is the orientation of the antenna generating the radio signal that determines the polarization. A vertical antenna produces vertical polarization, and a horizontal antenna produces horizontal polarization. For the best reception, the transmitting and receiving antennas should both be either vertical or horizontal.

There is also circular polarization. Some antennas make the signal rotate clockwise or counterclockwise as the signal travels from transmitter to receiver. Clockwise rotation, as viewed from the transmitter, is called right-hand circular polarization. Counterclockwise rotation is called left-hand circular polarization.

Signal Speed

Radio waves travel at the speed of light in space. That is 300,000,000 m/s or 186,410 miles/s. So while radio transmission is not instantaneous, it is pretty fast. Just as an example, to transmit a signal to the moon that is 250,000 miles away takes 250,000/186,400 = 1.34 s.

Signal Strength

As the radio wave gets farther and farther from the antenna, it gets weaker and weaker. The power in a radio wave, in fact, decreases by the square of the distance from the antenna. For example, doubling the distance decreases the received power by two squared (2^2) or four. Even when the transmitter puts many hundreds or thousands of watts of power into the antenna, the signal received at a distance is many times smaller. Typical radio signals are usually expressed in terms of microwatts (1-millionth of a watt) or nanowatts (1-billionth of a watt).

As the radio wave encounters a receiving antenna, the magnetic and electric fields induce a voltage into the antenna. This tiny voltage is then amplified and translated back into the original information signal by the receiver.

ELECTROMAGNETIC FREQUENCY SPECTRUM

The radio signal applied to the antenna by the transmitter is a sine wave. Remember that all sine waves have both an amplitude and a frequency. Frequency is stated in terms of cycles per second or hertz (Hz). Radio transmissions can take place over a very wide frequency range. We call this range the *electromagnetic frequency spectrum*. It extends from approximately 10 kHz–300 MHz. Fig. 7.4 shows a diagram designating the major divisions of the electromagnetic spectrum.

The electromagnetic spectrum is in free space, not only here on Earth but also throughout the universe. The spectrum space may be used by anyone and, in fact, on Earth we all share it. The spectrum is divided up and used for many different radio services. It is also partitioned such that various frequency ranges are assigned to or used by different countries for their various radio services. In many cases, portions of the frequency spectrum are shared concurrently among several countries. This allocation of the frequency spectrum is necessary to prevent radio stations from interfering with one another.

In the United States, the frequency spectrum is regulated by the Federal Communications Commission (FCC). Congress established the Communications Act of 1934, which has subsequently been amended to permit the FCC to regulate all forms of radio and wired (telephone, cable TV, etc.) communications services. The National Telecommunications Information Administration (NTIA) manages the spectrum for the government and the military. Throughout the world, the frequency spectrum is allocated and managed by a worldwide organization known as the International Telecommunications Union (ITU).

The various areas of the frequency spectrum are assigned to different types of communications services. For example, in the LF range, most communications

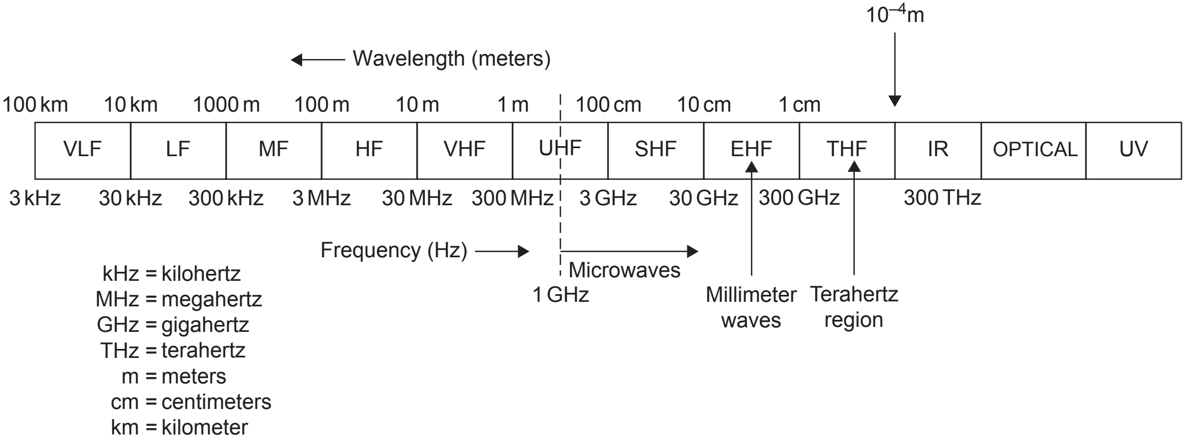
10^{-4}m
Wavelength (meters)
100 km
10 km
1000 m
100 m
10 m
1 m
100 cm
10 cm
1 cm
VLF
LF
MF
HF
VHF
UHF
SHF
EHF
THF
IR
OPTICAL
UV
3 kHz
30 kHz
300 kHz
3 MHz
30 MHz
300 MHz
3 GHz
30 GHz
300 GHz
300 THz
Frequency (Hz)
Microwaves
1 GHz
Millimeter waves
Terahertz region
kHz = kilohertz
MHz = megahertz
GHz = gigahertz
THz = terahertz
m = meters
cm = centimeters
km = kilometer

FIGURE 7.4 Electromagnetic frequency spectrum.

Major Segments of the Frequency Spectrum

30–300 kHz—Low frequencies (LFs)
300 kHz–3 MHz—Medium frequencies (MFs)
3–30 MHz—High frequencies (HFs) or shortwaves (SWs)
30–300 MHz—Very-high frequencies (VHFs)
300 MHz–3 GHz—Ultrahigh frequencies (UHFs)
3–30 GHz—Superhigh frequencies (SHFs)
30–300 GHz—Extremely high frequencies (EHFs)
300 GHz to IR—Tremendously high frequencies (THFs) or the terahertz (THz) region
Note: One terahertz is 1000 GHz.

services are navigational systems used by the Coast Guard and military. The AM broadcast band (530–1710 kHz) is in the MF range. SW radio broadcasts, CB radio, amateur radio, and a variety of other communications services are in the 3–30 MHz or HF range.

One of the most widely used areas of the spectrum is the VHF range from 30 to 300 MHz. Virtually all two-way radio communications for marine, aircraft, mobile, and other services are in this range. FM broadcasting (88–108 MHz) and TV channels 2 through 7 are also in this region.

The UHF frequency range, from 300 to 3000 MHz (3 GHz), also contains a wide variety of communications services. It includes two-way mobile radio, TV, and cellular telephone.

Frequencies above 1 GHz are designated as microwaves. In the microwave region between 1 and 3 GHz are satellite and navigational systems. For example, the widely used global positioning system (GPS) operates in this range. This area is also occupied by the newer cellular telephone service and wireless networks such as Wi-Fi and Bluetooth.

The frequencies from 30 to 300 GHz are called millimeter waves because their wavelength is in millimeters. These frequencies are typically used only by satellite and radar systems. Other services are point-to-point wireless data services and cellular backhaul. The forthcoming fifth generation (5G) cell phone system will the use millimeter wave bands. Some short-range wireless devices have also been defined for the 60-GHz range. These are used in consumer applications as well as in faster wireless networks. Automotive radars use frequencies in the 76–81 GHz range.

Critical Issue: Spectrum Space

There is only so much radio spectrum space and we must all share it. It is a true natural resource. That is why spectrum space is regulated by governments. But we have just about used it all up. Virtually every hertz of space is spoken for. Diverse radio services compete for space. In some cases, an old service must die before a new one can begin. Sometimes space is stolen from one service to give to another,

Critical Issue: Spectrum Space—cont'd

such as in the mid-1980s when the upper UHF TV channels were sacrificed to make room for cellular telephones. Lately, the US government has been selling space to the highest bidders for new cell phone and other wireless services. During the recent switchover to digital TV in the United States, a huge chunk of the spectrum was freed up (108 MHz, from 698 to 806 MHz) as TV stations abandoned this region. Known as the 700-MHz band, it was auctioned off by the government for billions for more cell phone spectrum as well as broadband wireless services.

The trend has been to push higher and higher in the frequency range to find more space. During the past several decades, the microwave region (above 1 GHz) has been developed. This region, in the UHF, SHF, and EHF ranges, was previously used only for radar and satellites. But now it is being used and developed for digital cell phones, navigation, high-speed networking, broadband (Internet) wireless, and digital TV.

Beyond the millimeter wave range above 300 MHz is the terahertz region, infrared (IR) and optical spectrum. That is, above 300 GHz, the electromagnetic radiation becomes light, namely IR, visible light, and ultraviolet light. But that 'is another story altogether. IR is widely used for wireless connectivity as most TV and other remote controls use it.

Wavelength

While we normally use frequency to refer to location of a radio signal in the spectrum, sometimes you will hear about frequencies or ranges of frequencies designated by wavelength.

Wavelength is the actual physical length of a radio wave. It is the amount of space the wave occupies. Specifically, wavelength is the length of one cycle of the sinusoidal variation of the electromagnetic fields. If we could see the electric and magnetic fields, we could measure the distance between the maximum intensity points in meters (1 m = 3.28 ft or 39.37 in.). This is the wavelength. Wavelength is also the distance that the radio wave travels in the time of one cycle.

You can determine the wavelength of a radio signal with the following simple formula:

$$\lambda = 300,000,000/f$$

Wavelength is represented by the Greek letter lambda (λ), and frequency in hertz is designated by the letter "f." The value 300,000,000 is the speed of light of a radio wave in meters per second in free space. (This translates to a speed of 186,410 miles/s.)

A radio signal with a frequency of 150 MHz is expressed in hertz as 150,000,000 Hz. Therefore,

$$\lambda = 300,000,000/150,000,000 = 2\text{ m}$$

If we give the frequency in megahertz, the formula can be shortened to:

$$\lambda = 300/f$$

Continued

Wavelength—cont'd

For example, we can calculate the wavelength of the SW or HF part of the spectrum (3–30 MHz) as follows:

$$\lambda = 300/3 = 100 \text{ m}$$

$$\lambda = 300/30 = 10 \text{ m}$$

If you rearrange the formula, you can also find the frequency if you know the wavelength:

$$f = 300/\lambda$$

What is the frequency of the 25-m band?

$$f = 300/25 = 12 \text{ MHz}$$

RADIO WAVE PROPAGATION

Propagation refers to the way in which a radio signal travels from the transmitting antenna to the receiving antenna. The way the signal behaves is directly related to its frequency.

Low and Medium Frequencies

For example, in the low-frequency and medium-frequency ranges, the radio signals tend to hug the Earth as they move from transmitter to receiver. These ground wave signals can only travel relatively short distances because they are absorbed and attenuated by the Earth, trees, buildings, and other obstructions. Depending on the frequency of operation, transmitter power, and the nature of the terrain, distances are limited to several hundred miles. Distances can be extended with greater power or if the signal travels over water.

High Frequencies

In the HF or SW frequency range, the ground waves are extremely weak and travel only short distances, usually several miles. Instead, propagation is by way of refraction of the ionosphere. The ionosphere is an electrified area above the Earth located from approximately 30–250 miles above the Earth. The Sun ionizes the gases in the air, thereby giving them electrical characteristics. The closer the ionosphere to the Sun, the greater the degree the ionization. The ionosphere is actually several different layers or ionization levels as shown in Fig. 7.5. The upper layers have greater ionization than the lower layers. Obviously, the degree of ionization is determined by whether it is day or night. During the daylight hours, the ionosphere is highly charged electrically. At night, the ion layers essentially disappear.

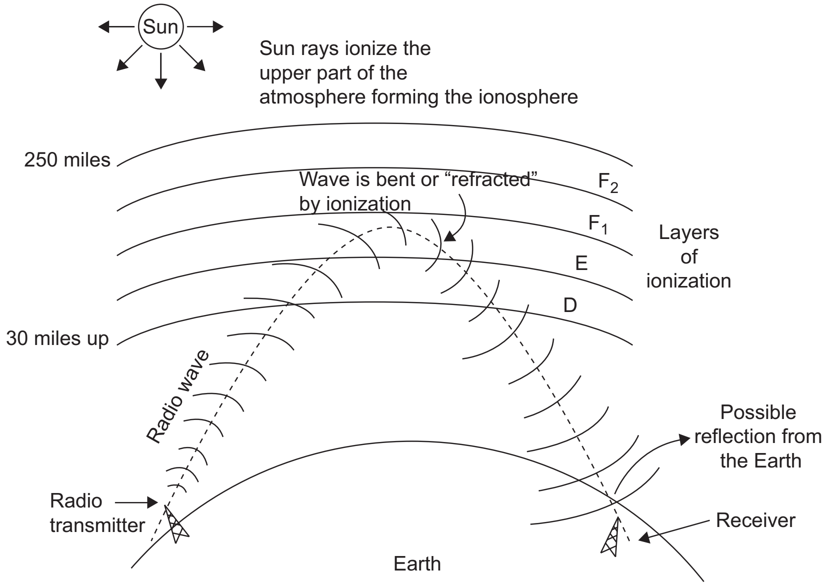
Sun
Sun rays ionize the upper part of the atmosphere forming the ionosphere
250 miles
Wave is bent or "refracted" by ionization
F_2
F_1
E
D
Layers of ionization
30 miles up
Radio wave
Possible reflection from the Earth
Radio transmitter
Receiver
Earth

FIGURE 7.5 Radio propagation of shortwaves.

A radio wave in the 3- to 30-MHz range is transmitted up to the ionosphere where it is refracted or bent by ionization. The degree of bending depends on the frequency of operation and how high in the ionosphere the radio wave penetrates. In most cases, the radio wave is bent so much that it actually returns to Earth. The effect appears to be similar to that of a light beam being reflected from a mirror.

As you can see, the radio wave actually returns to the Earth at some distance from the transmitter. Receiving stations in the area of the refracted signal can receive it. In some cases, if the signal is strong enough and the characteristics of the Earth are right, the radio signal may actually be reflected by the Earth back up to the ionosphere where it is again bent and returned to Earth. Under the right conditions, the radio signals can make several hops, thereby greatly extending the distance of communications. With the optimum frequency, power level, and time of day, communications halfway around the world can easily be achieved. That is why it is easy for you to listen to worldwide radio broadcasts that occur in the SW frequency bands.

VHF, UHF, Microwaves and Millimeter Waves

Radio wave propagation in the frequency range above 30 MHz occurs essentially by direct antenna-to-antenna space waves. This is usually called *line-of-sigh (LOS) communications*. Further, the higher the frequency of the radio wave, the less it is affected by the ionosphere. UHF and microwave signals penetrate the ionosphere, making satellite and long-distance spacecraft communication possible.

$$D = \sqrt{2H_t} + \sqrt{2H_r}$$

D = Line-of-sight distance (miles)
H_t = Height of transmitting antenna
H_r = Height of receiving antenna

If $H_t = 200$ ft and $H_r = 40$ ft,

$$D = \sqrt{2(200)} + \sqrt{2(40)} = \sqrt{400} + \sqrt{80}$$

$$D = 20 + 8.94 = 28.94 \text{ miles}$$

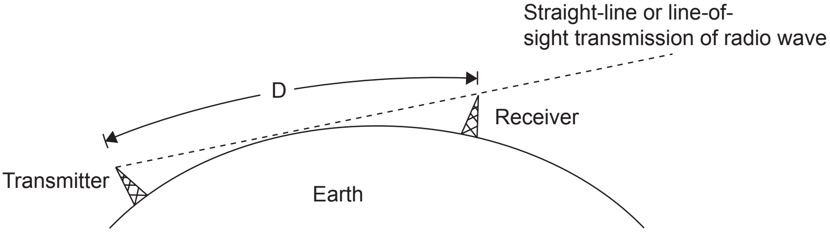

FIGURE 7.6 Line-of-sight (LOS) communications at VHF, UHF, and microwave.

LOS communications means that the transmitting antenna must be able to "see" the receiving antenna. If the distance of the receiving antenna is beyond the curvature of the Earth, the transmitted signal will simply pass over the antenna. To receive the signal, the receiving antenna or transmitting antenna can be increased in height, thereby extending the total range of communications. Fig. 7.6 shows this idea. The simple formula along with it allows you to calculate the maximum LOS communications if you know the height of the receiving and transmitting antennas. That 'is why at VHF, UHF, and microwave frequencies, the height of the antenna is the most critical determining factor in overall communications distance.

MODULATION

A radio wave is a sine wave of voltage that has been converted into electric and magnetic fields by the antenna. The transmitter generates the sine wave and amplifies it to a high power level. The question is: how does the information to be transmitted fit into this?

The information or intelligence to be conveyed is usually a voice signal produced by a microphone. Other forms of intelligence signals include pictures or video and digital (binary) data from a computer. The purpose of the HF sine wave radio signal is to give the intelligence signal a ride. That 'is why the radio signal generated by the transmitter is called the carrier. Modulation is the process by which the information is impressed on or embedded in the carrier. Modulation causes the carrier to be modified by the information signal.

There are three basic forms of modulation used to put information on a carrier: *amplitude modulation* (AM), *frequency modulation* (FM), and *phase modulation* (PM).

Amplitude Modulation

In AM, the voltage or power level of the information signal changes the amplitude of the carrier in proportion (see Fig. 7.7). With no modulation, the AM carrier is transmitted by itself. When the modulating information signal (a sine wave) is applied, the carrier amplitude rises and falls in accordance. The carrier frequency remains constant during AM.

AM is widely used in radio. AM broadcast stations are, of course, amplitude modulated. So are CB radios, aircraft radios, and the video modulation of a TV broadcast transmitter. A special form of AM, known as quadrature amplitude modulation (QAM), is also widely used in modems to transmit digital data over cable or wireless.

Sidebands

The modulation process causes new signals to be generated. These new sine wave signals are called sidebands. Their frequencies are the sum and difference

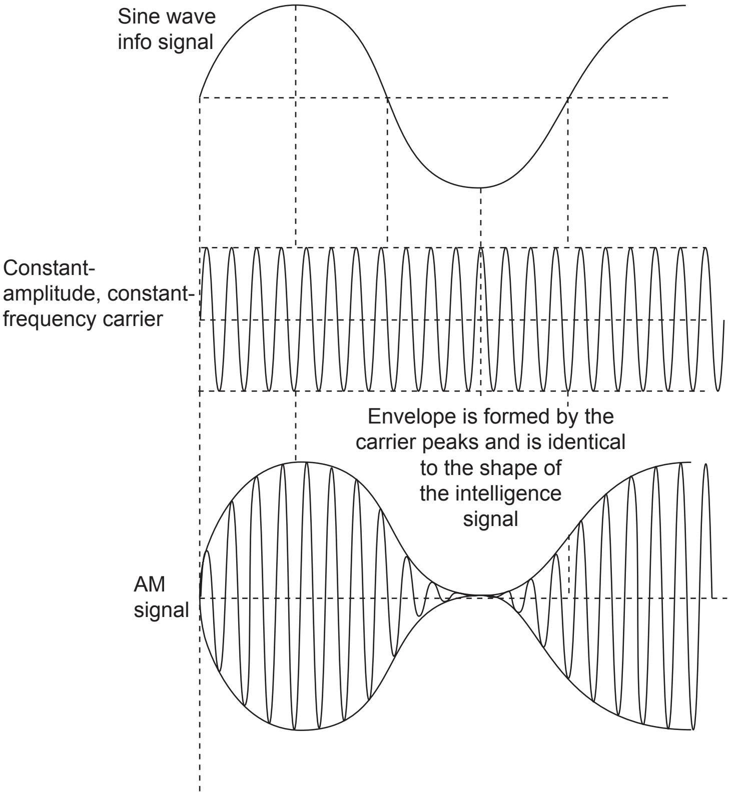
Sine wave
info signal
Constant-
amplitude, constant-
frequency carrier
Envelope is formed by the
carrier peaks and is identical
to the shape of
the intelligence
signal
AM
signal

FIGURE 7.7 Amplitude modulation.

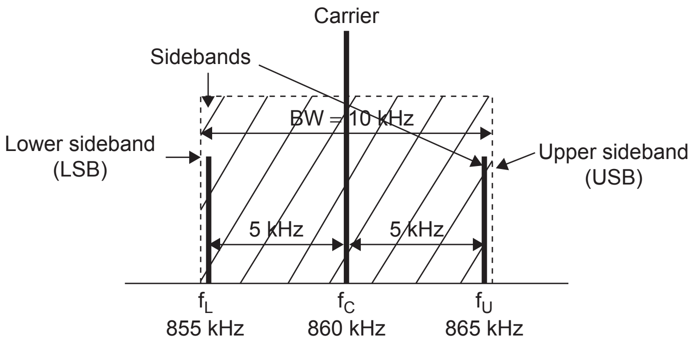

f_M = Frequency of modulating signal (5 kHz)
f_C = Carrier frequency (860 kHz)
f_L = Lower sideband frequency
f_U = Upper sideband frequency

$$f_L = f_C - f_M = 860 - 5 = 855 \text{ kHz}$$
$$f_U = f_C + f_M = 860 + 5 = 865 \text{ kHz}$$

$$BW = f_U - f_L = 865 - 855 = 10 \text{ kHz}$$

FIGURE 7.8 Sidebands and bandwidth.

of the carrier and modulating signal frequencies. For example, in an AM radio station, audio frequencies as high as 5 kHz can be transmitted. If a 5-kHz sine wave tone is to be transmitted, the modulation process causes sidebands 5 kHz below and 5 kHz above the carrier to be produced. This is illustrated in Fig. 7.8. For an AM radio station with a carrier frequency of 860 kHz, the lower sideband (LSB) would occur at 860 − 5 = 855 kHz, while the upper sideband (USB) occurs at 860 + 5 = 865 kHz. The carrier and the sidebands combined produce the composite waveform shown at the bottom of Fig. 7.7.

Bandwidth

This brings up the very important concept of bandwidth. Bandwidth refers to a range of frequencies over which a radio signal operates. This is also referred to as a radio channel. As you can see from Fig. 7.8, the AM signal consists of the carrier and the sidebands. Together these signals occupy a bandwidth of 10 kHz. You can figure out the bandwidth of any signal by simply subtracting the LSB frequency from the USB frequency.

$$BW = USB - LSB = 865 - 855 = 10 \text{ kHz}$$

All of the receiver circuits must be set to pass signals in this 10-kHz range to avoid distortion of the signal or lost information.

Single Sideband Modulation

Single sideband (SSB) modulation is a special form of AM in which the carrier and one sideband are suppressed. All of the information to be transmitted such as voice is fully contained in one sideband. This does two things. First, it reduces the required spectrum to one half of that of a normal AM signal thereby saving spectrum space. Second it eliminates the wasteful carrier power and allows more power to be put into the sideband providing more reliable communications. SSB is used in some military applications and is widely used in amateur radio and citizen's band (CB) radios.

Frequency Modulation

In frequency modulation, the carrier amplitude remains constant, but its frequency is changed in accordance with the modulating signal. Specifically, the higher the amplitude of the information signal, the greater the frequency change. The actual carrier frequency deviates above and below the center carrier frequency as the information signal amplitude varies. Fig. 7.9 shows frequency modulation with a sine wave information signal. Note that the carrier frequency gets higher on the positive peaks and lower on the negative peaks of the information signal.

Like AM, FM also produces sidebands. But unlike AM, which produces a single pair of sidebands for each frequency in the modulating signal, the FM process produces an infinite number of pairs of sidebands for each frequency in the information signal. As a result, the bandwidth occupied by an FM signal is enormous. Luckily, the number of sidebands produced can be controlled by

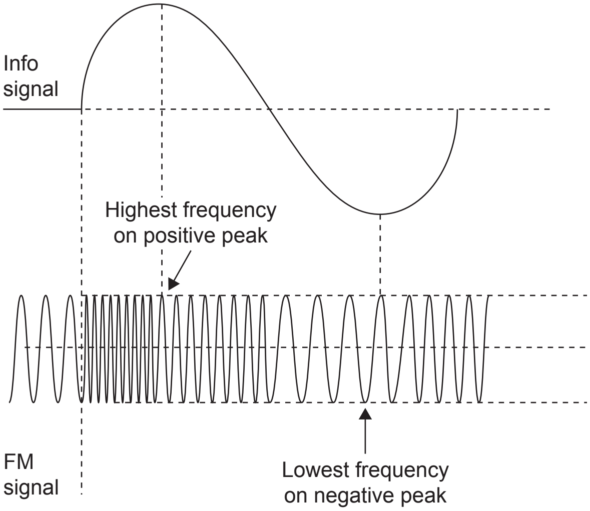

FIGURE 7.9 Frequency modulation.

properly selecting the amount of deviation permitted in the carrier. Small deviations result in fewer sidebands. Further, some of the higher-order sidebands are extremely low in amplitude and, therefore, contribute little to the FM signal. But while the bandwidth of an FM signal can be controlled and established to fit a desired frequency range, it does nevertheless usually take more room in the spectrum than an AM signal.

The primary benefit of FM is that it is less sensitive to noise, which consists of undesirable amplitude variations that get involuntarily added to a signal. Noise is easily eliminated in an FM system where a constant carrier amplitude is used. Some of the most common applications of FM include FM radio broadcasting, and two-way mobile and marine radios.

Phase Modulation

The third type of modulation is PM. Remember that phase shift is a time shift between two sine waves of the same frequency. We can use the information signal to shift the phase of the carrier with respect to the carrier reference. The result of this, however, is a signal that looks virtually the same as an FM signal (Fig. 7.9). PM is usually a little easier to implement electronically than FM, so most so-called FM systems typically use PM instead. Both FM and PM are often referred as types of *angle modulation*.

Digital Modulation

All of the three basic types of modulation, AM, FM, and PM, can also be used to transmit digital or binary data. However, they are usually given different names. AM of a carrier by a binary signal is usually referred to as amplitude shift keying (ASK). This is illustrated in Fig. 7.10A. The binary signal simply shifts the carrier amplitude between two specific levels. A special form of ASK is called on–off keying (OOK). This is illustrated in Fig. 7.10B. Here the binary signal simply turns the carrier on for a binary 1 and off for a binary 0.

Frequency modulation of a carrier by a binary signal is called frequency shift keying (FSK). Here the binary signal shifts the carrier between two discrete frequencies (see Fig. 7.10C).

PM of a carrier by a binary signal is referred to as phase shift keying (PSK). The term binary PSK (BPSK) is also used. The phase of the carrier is changed as the binary signal switches from 0 to 1 or 1 to 0 (see Fig. 7.10D). Phase shift is 180 degrees, which is easily detected at the receiver.

One of the most widely used forms of digital modulation is QAM, which is really a composite of both AM and PSK. The binary signal is converted into a multilevel digital signal that simultaneously modifies both the amplitude and the phase of the signal. QAM is used in digital data transmission in satellites and in modems sending digital data in audio form over telephone lines or cable TV systems. It is also widely used in the newer digital wireless networks. See Chapter 8 for more detail.

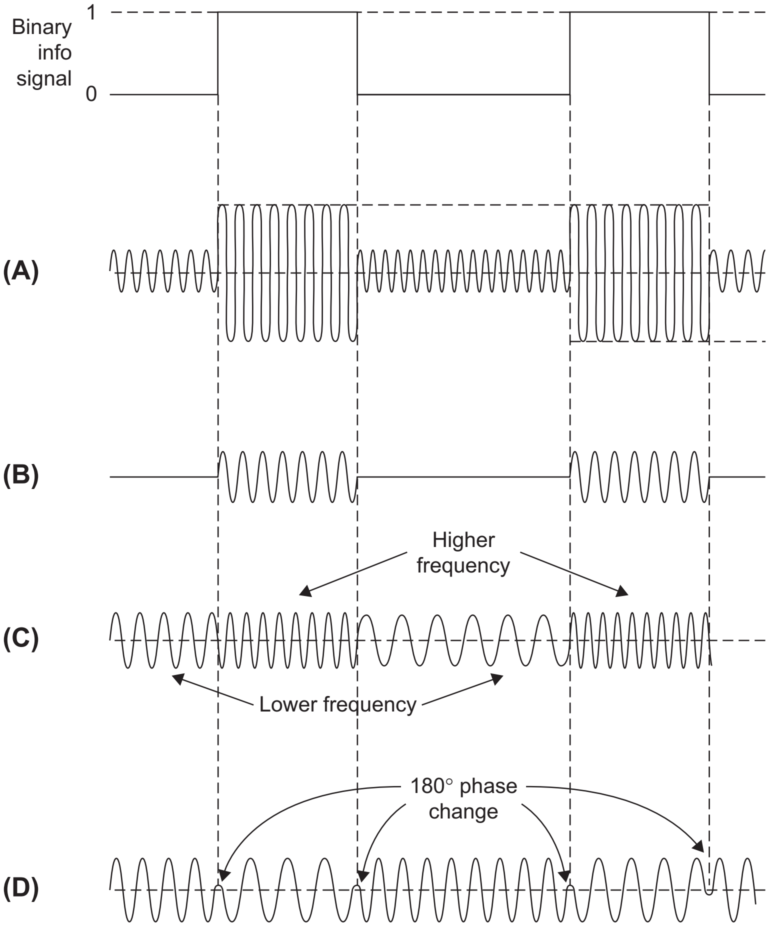

FIGURE 7.10 Types of digital modulation. (A) ASK. (B) OOK. (C) FSK. (D) BPSK.

Because binary signals contain many higher-frequency harmonics, the resulting signals produced by modulation have an enormous bandwidth. Many schemes have been developed to limit the spectrum of the modulating signal and, as a result, to reduce the amount of bandwidth occupied by a binary modulated radio signal.

Spread Spectrum

Spread spectrum (SS) is unusual in that it is not only a form of modulation but also a system for encrypting the information signal for security purposes and for multiplexing multiple signals over a range of frequencies.

SS was originally developed during World War II as a way to prevent the enemy from interpreting radio transmissions. For years the government and military protected the concepts of SS modulation. However, today it is widely used in radio and telephone communications. Most of the newer cellular telephones are based on this concept, specifically code division multiple access (CDMA) cell phones.

When using SS, the analog information signal is converted into a digital signal before transmission. The digital voice signal is then used to modify the carrier in a special way. The two most common ways of modifying the carrier are referred to as *direct-sequence* (DS) and *frequency-hopping* (FH) *spread spectrum*. In DSSS, the serial digital signal representing the voice is chopped up into random sequences and then used to PSK the carrier. In FHSS, the carrier is rapidly switched from one frequency to another for a short dwell time. The result of SS modulation is that the signal itself is broken up at random, and spectral energy is spread over a very wide frequency range. Because of the random sequences of the coding and FH, such a radio signal is nearly impossible to put back together at a receiver without knowing the code. SS receivers designed to pick up the SS signal know the code and can unscramble the signal and reassemble it into the original modulating information signal.

Orthogonal Frequency Division Multiplex

Orthogonal frequency division multiplex (OFDM) is another widely used modulation technique. OFDM and its usual companion modulation QAM are covered in detail in Chapter 8 on cell phones.

RADIO HARDWARE

All radio communications systems consist of three basic pieces of hardware: *transmitter*, *receiver*, and *antenna*. In some systems, the transmitter and receiver are packaged and used separately, while in others they are combined into a single package and called a *transceiver*.

Transmitters

The radio signal is produced by the transmitter. A general block diagram that applies to virtually every transmitter is shown in Fig. 7.11. The radio signal is generated at a low power level by a frequency synthesizer. This uses a phase-locked loop (PLL) with a crystal (XTAL) oscillator reference. A variable-frequency divider is used to set the frequency of operation. The FCC is really fussy about the frequency at which transmitters work and, therefore, the frequency must be precisely set and extremely stable. A quartz crystal reference oscillator is normally used to achieve this objective.

The low-level signal produced by the voltage-controlled oscillator (VCO) is then amplified by one or more driver amplifiers to boost the power level. Next, the microphone signal is fed to the modulator. The modulator changes the

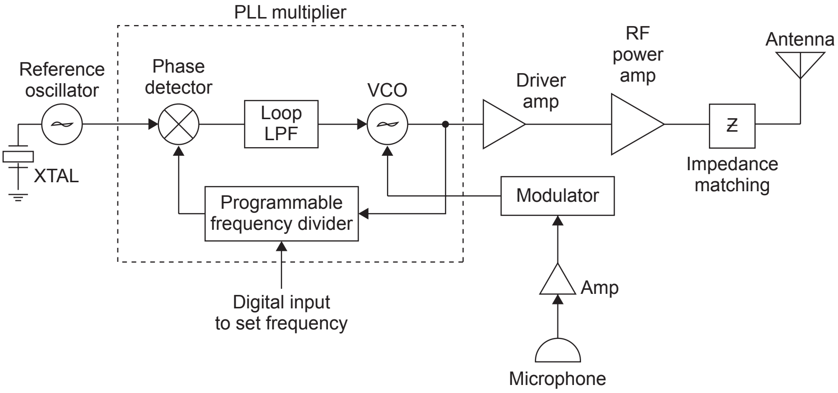

FIGURE 7.11 Radio transmitter.

frequency or phase of the RF signal in accordance with the voice, digital data, or other information to be transmitted. That signal is applied to the PLL VCO.

The modulated signal is sent to additional RF power amplifiers, usually referred to as *drivers*. Each succeeding stage boosts the signal power by a specific amount. A final RF power amplifier produces the desired amount of output power. From there, impedance (Z) matching circuits are used to match the high-power amplifier (HPA) to the antenna for maximum power radiation.

The physical realization of the transmitter can vary all the way from a single integrated circuit chip, the size of a postage stamp to something that can fill up a moderately large building. The size depends on the operating frequency and the amount of output power desired. Very-low-power, HF radio transmitters for garage door openers and remote keyless entry systems on cars are fully self-contained within a single integrated circuit. High-power radio broadcast transmitters at LFs may fill a small room. The higher the power and the lower the frequency, the larger the transmitter.

Receivers

The job of the receiver is to translate the weak radio signal picked up by the antenna into an output that fully recovers the originally transmitted information. By the time the radio signal reaches the receiving antenna, it is extremely weak and, therefore, the primary job of the receiver is to boost the signal level. Therefore, the receiver consists mainly of a series of amplifiers that increase the level of the small signal to a point where the original information can be recovered.

The amount of amplification provided by a receiver determines its *sensitivity*. Another job of the receiver is to provide selectivity so that the desired signal can be picked out of the thousands of signals being transmitted simultaneously by other transmitters. The receiver's job is to select just the desired one and to reject all of the others. This is what we refer to as the *selectivity* of the receiver.

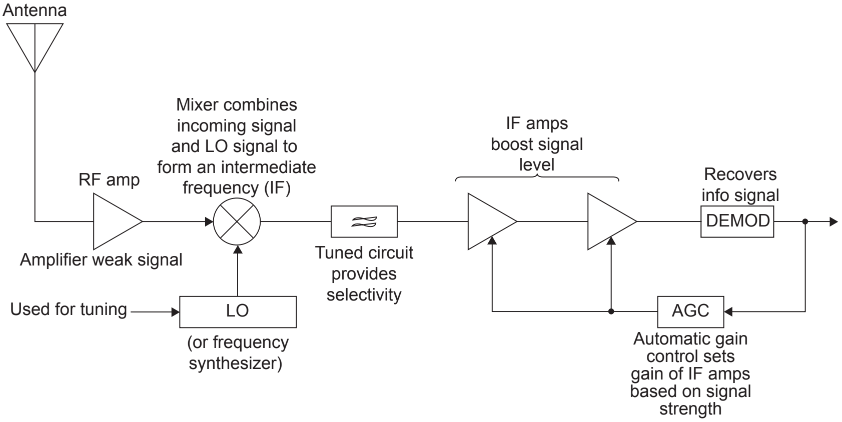
Antenna
RF amp
Amplifier weak signal
Mixer combines incoming signal and LO signal to form an intermediate frequency (IF)
Used for tuning
LO
(or frequency synthesizer)
Tuned circuit provides selectivity
IF amps boost signal level
Recovers info signal
DEMOD
AGC
Automatic gain control sets gain of IF amps based on signal strength

FIGURE 7.12 Superheterodyne receiver. *LO*, local oscillator.

A third function of the receiver is to recover the original information. A circuit called a *demodulator* removes the voice, video, or digital data from the carrier. Finally, output amplifiers reproduce the signal as desired, such as a voice signal in a speaker.

Fig. 7.12 is a block diagram of a superheterodyne receiver. The incoming signal is usually amplified by an *RF input amplifier* often called a *low-noise amplifier* (LNA) and then fed to a circuit called a *mixer*. The mixer is a frequency-translation circuit that converts the incoming signal, regardless of its frequency, along with its modulation to a lower intermediate frequency (IF). This allows the receiver to use a fixed-frequency amplifier system where most of the gain and selectivity of the receiver is achieved. The mixer takes the input signal and mixes it with a HF sine wave signal generated internally in the receiver by what is called a local oscillator (LO). The LO is usually a frequency synthesizer. The result of the mixing is output signals that are the sum and difference frequencies of the input and oscillator signals. The IF amplifier is normally tuned to select only the difference frequency. Tuning is accomplished by changing the LO frequency. The LO in most channelized receivers is some type of PLL frequency synthesizer.

The IF amplifier boosts the signal level considerably. Its selectivity ensures that signals whose frequencies are above or below the desired signal are effectively eliminated. The IF amplifier also has a built-in feedback control circuit that automatically adjusts its gain to the level of the incoming signal. This feature is called *automatic gain control* (AGC). AGC permits the receiver to pick up both strong local signals and very weak distant signals. When very strong signals are received, the AGC automatically reduces the gain of the IF amplifier to minimize circuit overload and signal distortion. When very weak signals are received, the AGC adjusts the gain of the IF amplifier for higher amplification.

The *demodulator circuit*, sometimes called the *detector*, follows the IF amplifier. It removes the carrier and recovers the desired information signal. From there, additional processing takes place to either boost the signal level or to shape it before it appears at the output.

Most receivers that we are familiar with have a speaker as the primary output. But there are exceptions, such as the picture tube in a TV set or the motor in a garage door opener. The receiver may also put out digital data that goes straight to a computer.

Transceivers

Most communications systems use both a transmitter and a receiver. In two-way radio, both send and receive capability is needed at both ends of the communications channel. Therefore, the receiver and transmitter are packaged together in the same housing to create a transceiver. Small portable transceivers are called handhelds, walkie-talkies, or cellular telephones.

The transmitters and receivers are typically the same as those described earlier. However, in some cases, the transmitter and receiver may share circuits. They virtually always share the antenna and a common power supply like a battery. In some cases, a single-frequency synthesizer may generate both the receiver LO signal as well as the transmitter carrier.

ISM Band Transceiver

The ISM band refers to specific frequencies set aside for wireless industrial, scientific, and medical (ISM) applications. The most common frequencies in the United States are 315, 433, and 915 MHz. The frequency 868 MHz is widely used in Europe. These unlicensed frequencies are used for simple monitor and control applications using low-speed digital data over short ranges. Some common examples are garage door openers, remote keyless entry on cars, remote thermometers, tire pressure gauges, security and fire alarms, home automation, and any telemetry or remote control application. The radios used for this are typically single ICs to which an antenna, a battery, a housing, and the input data are added. The devices are available as a receiver, a transmitter, or a complete transceiver. Fig. 7.13 is a block diagram of a typical transceiver.

The transmitter section uses an external crystal to serve as the reference to an internal PLL frequency synthesizer to set the frequency. The PLL actually multiplies the crystal frequency by some N value to get the desired output. If the crystal frequency is 6.77 MHz and the divide ratio is 64, the output would be 433.28 MHz. The serial digital data input coming from a sensor, an analog-to-digital converter (ADC), or a computer frequency modulates the crystal oscillator to create the FSK signal. A power amplifier after the transmit VCO boosts the power up into the 1- to 100-mW range. The output goes to the antenna. A transmit/receive switch allows the antenna to be shared in a half-duplex system.

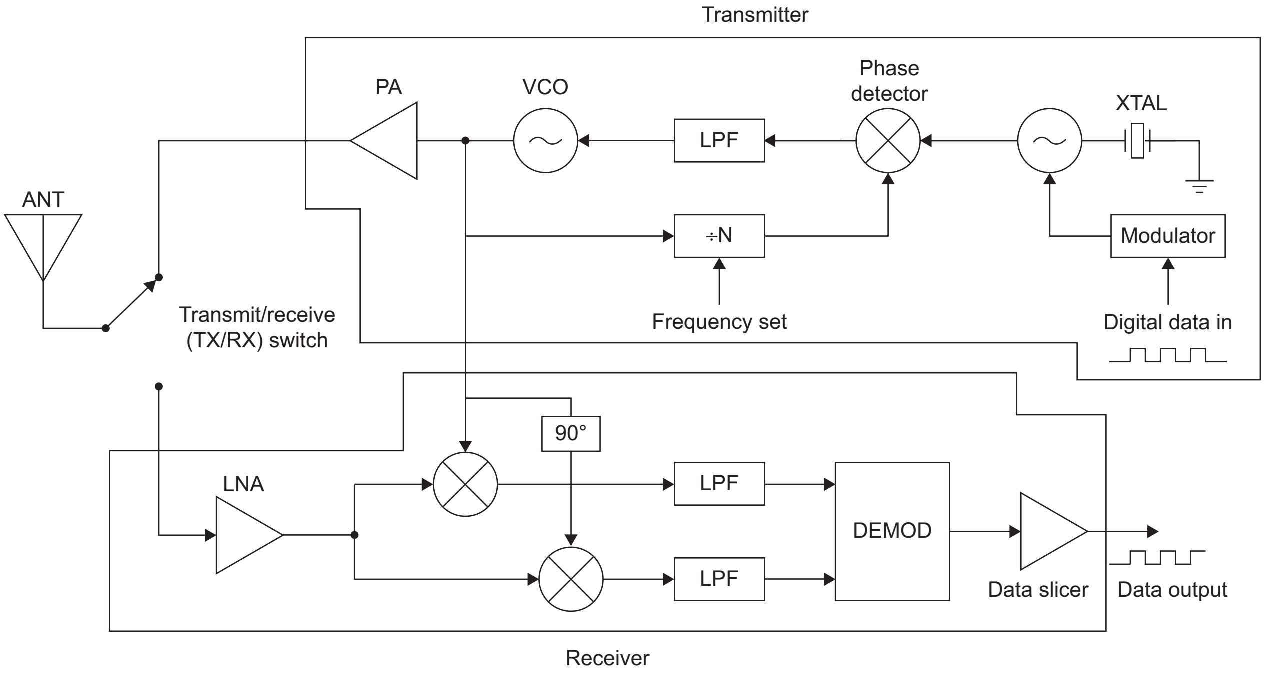
Transmitter
Phase
detector
PA
VCO
XTAL
LPF
ANT
÷N
Modulator
Transmit/receive
(TX/RX) switch
Frequency set
Digital data in
90°
LNA
LPF
DEMOD
LPF
Data slicer
Data output
Receiver

FIGURE 7.13 Single-chip UHF transceiver for low-speed, short-range digital data transmissions.

The receiver is a special version of a superheterodyne. The input signal from the antenna is boosted in amplitude by the LNA and applied to the mixers. The LO is set to the incoming signal frequency (also the transmit frequency), so the mixers generate a difference of zero. This is called *direct conversion* or *zero IF*. Two mixers are used and the LO is shifted 90 degrees in phase between the two. This gives two zero IF outputs that are filtered. Zero IF does not mean a nonexistent output signal. It just means that the FSK signal is recovered directly at its data rate. The I and Q signals are 90 degrees out of phase, and this allows the digital data to be more easily demodulated. A data slicer, a circuit like a comparator, is used to shape the signal into a clean binary bitstream. With this kind of radio, the range is limited. It may be only a few feet or up to a mile or so depending on the antenna, environment, or application.

ANTENNAS

The antenna takes the voltage and current produced by the transmitter and converts it into the electromagnetic wave we know as a radio wave. At the receiver, the antenna converts the electromagnetic field back into a voltage we know as the radio signal. The radio waves pass over the receiving antenna, inducing a voltage that causes current to flow in the receiver.

Antennas can have an amazing variety of shapes. Essentially an antenna or aerial is some arrangement of electrical conductors whose size and length are critical. For optimum operation, the antenna length must be related to the wavelength (λ) of the signal. Common antenna sizes are one-quarter ($\lambda/4$) and one-half ($\lambda/2$) wavelength at the operating frequency. The higher the frequency, the shorter the wavelength and thus the smaller the antenna. For example, one wavelength at 150 MHz, a popular two-way radio frequency range measures 2 m. A half wavelength is 1 m and a quarter wavelength is 0.5 m. Since 1 m is 39.37 in., a half wavelength is 19.68 in. At 900 MHz, the frequency range of cell phones, one wavelength is 0.3333 m. A quarter wavelength antenna would be $0.3333/4 = 0.08333$ m or 3.28 in.

Ground-Plane Antenna

Antennas can be anywhere from a long horizontal copper wire to a vertical steel or aluminum tower several hundred feet tall. Or it may be an array of short conductors like those in a TV antenna. One of the most common antennas is a short quarter wavelength vertical that is widely used in mobile radios, cell phones, and other handheld units (see Fig. 7.14). The quarter-wave or ground-plane antenna, as it is sometimes called, radiates equally well in all directions, making it omnidirectional. We say that its radiation or reception pattern is a circle.

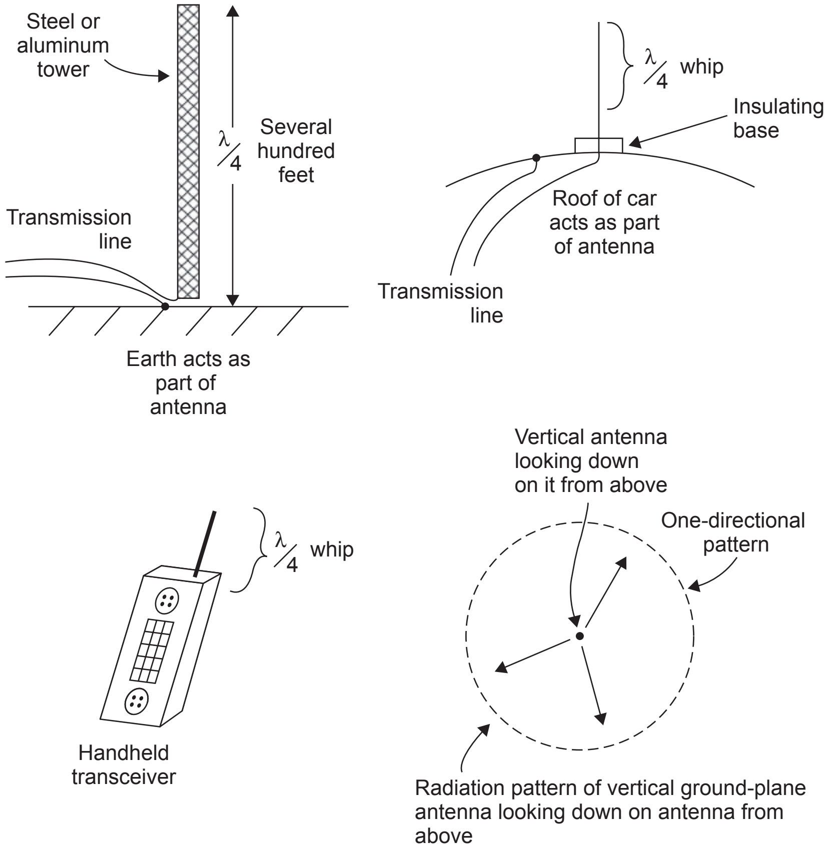
Steel or aluminum tower
λ/4
Several hundred feet
Transmission line
Earth acts as part of antenna
λ/4 whip
Insulating base
Roof of car acts as part of antenna
Transmission line
Vertical antenna looking down on it from above
One-directional pattern
λ/4 whip
Handheld transceiver
Radiation pattern of vertical ground-plane antenna looking down on antenna from above

FIGURE 7.14 Quarter-wave ground-plane antenna.

Dipole

Another common antenna type is the dipole, which is one-half wavelength long and usually mounted horizontally. Fig. 7.15A shows an example. Its radiation or reception pattern in Fig. 7.15B is a figure eight. The greatest radiation or strongest reception occurs at a right angle to the dipole element itself.

Yagi

A variation of the dipole is the Yagi, a multielement antenna based on the dipole shown in Fig. 7.16. To the dipole is added a reflector element that is slightly longer than one-half wavelength, and one or more director elements that are progressively shorter than the dipole. The directors and reflector make the Yagi highly directional. The extra elements, called *parasitic elements*, concentrate and aim the beam of radiation over a narrow range. At the transmitter, this has the effect of increasing the transmitted power,

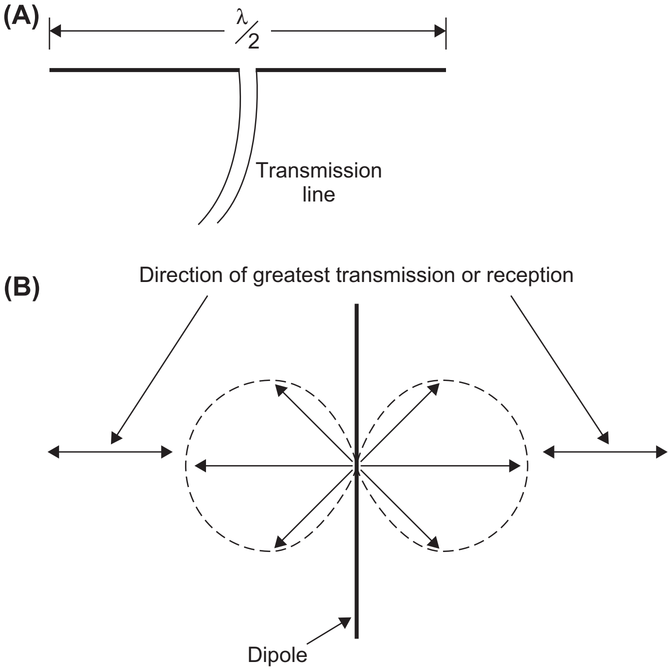

FIGURE 7.15 Dipole antenna. (A) Dipole and (B) horizontal radiation pattern.

making the signal stronger at the receiver as if the transmitter power had been increased. A similar gain effect occurs at the receiver.

Other Antennas

There are many other antenna variations. Antennas formed with copper patterns on a printed circuit board (PCB) are widely used. One example is the *patch antenna* (see Fig. 7.17). It is simply a square area of copper about one-half wavelength on a side made on a PCB. Another is the *loop antenna* that is just a loop of copper on the PCB. A popular cell phone antenna is the inverted F, which is an F-shaped pattern on the PCB. Antennas also come as components. For microwave frequencies, there are special antennas made on a ceramic base that are small and can be soldered to a PCB.

Transmission Lines

The antenna is usually connected to the transmitter or receiver by way of a coaxial cable (see Fig. 7.18). It consists of a copper center conductor surrounded by an insulator, such as Teflon. Around that is a shielded copper braid or foil. The whole thing is covered with a plastic insulator. Coaxial

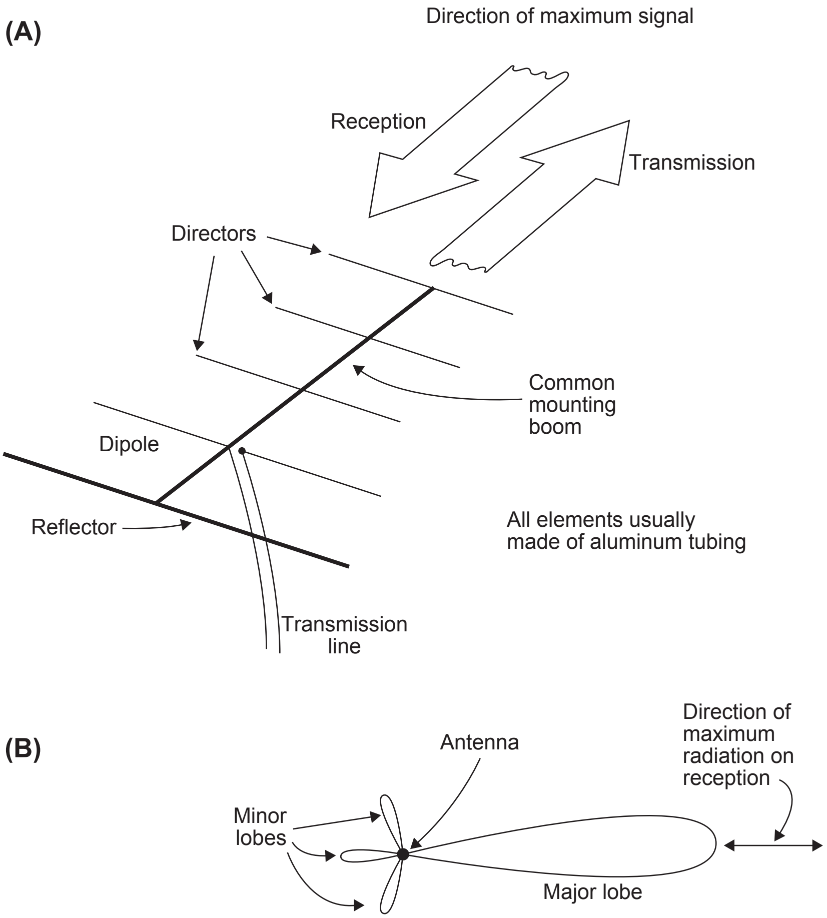
(A)
Direction of maximum signal
Reception
Transmission
Directors
Common mounting boom
Dipole
Reflector
All elements usually made of aluminum tubing
Transmission line
(B)
Antenna
Direction of maximum radiation on reception
Minor lobes
Major lobe

FIGURE 7.16 Yagi antenna. (A) Yagi and (B) radiation pattern.

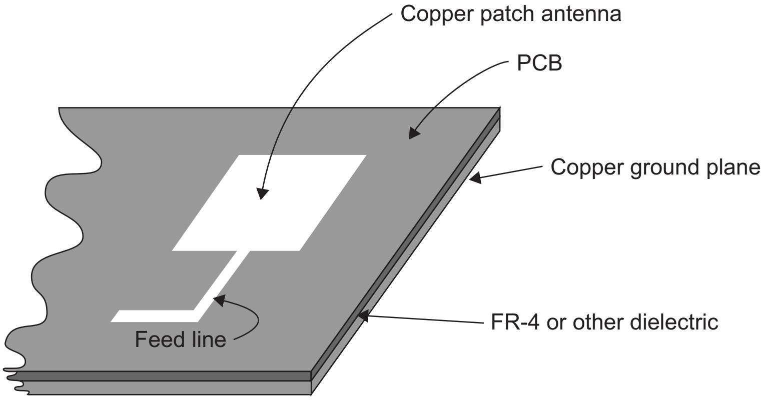
Copper patch antenna
PCB
Copper ground plane
Feed line
FR-4 or other dielectric

FIGURE 7.17 Printed circuit board (PCB) patch antenna.

cable is the most commonly used transmission line to connect the equipment to the antenna. It provides an efficient method of transferring the voltages and currents to and from the antenna while providing a certain amount of shielding from noise. For PCB antennas, the transmission line is just a short pair of copper lines from the transceiver IC to the antenna. The transmission line impedance most commonly used in RF systems is 50 Ω. Some systems such as cable TV use 75-Ω cable.

Microwave Antennas and Waveguides

Antennas used at microwave frequencies are quite different from those used at the lower frequencies. Perhaps the most common is the *horn antenna*. It looks like the flared end of a musical horn, only most of them have an opening or aperture that is rectangular rather than round.

As seen in Fig. 7.19, the horn is directional and has gain. The microwave energy is fed to the horn by a transmission line. Coax can be used for the lower microwave frequencies, but at the higher microwave frequencies, coaxial cable loss is too high for long runs. A high percentage of the signal power is lost in the coax cable itself. This problem is overcome by using a waveguide. A waveguide is nothing more than a hollow copper pipe, usually with a rectangular cross section through which the microwave energy passes from transmitter to antenna or from antenna to receiver.

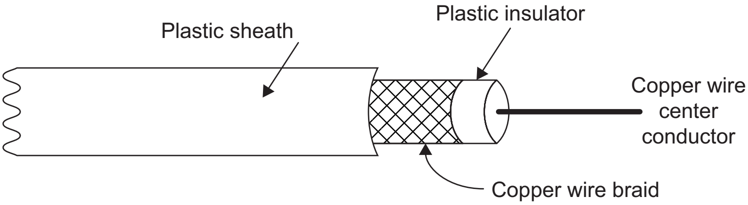
Plastic insulator
Plastic sheath
Copper wire center conductor
Copper wire braid

FIGURE 7.18 Coaxial cable transmission line.

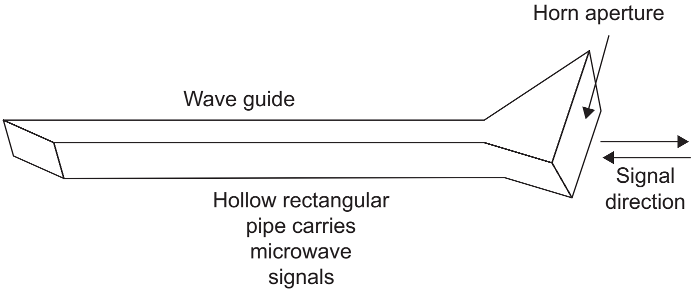
Horn aperture
Wave guide
Signal direction
Hollow rectangular pipe carries microwave signals

FIGURE 7.19 Microwave horn antenna fed by waveguide transmission line.

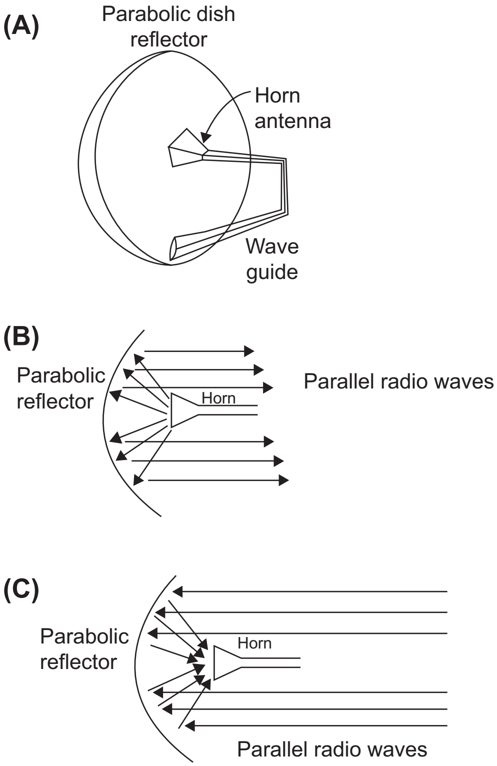
(A)
Parabolic dish reflector
Horn antenna
Wave guide
(B)
Parabolic reflector
Horn
Parallel radio waves
(C)
Parabolic reflector
Horn
Parallel radio waves

FIGURE 7.20 How a parabolic dish works. (A) Parabolic dish with horn, (B) transmission, and (C) reception.

The most effective and widely used microwave antenna is the parabolic dish (see Fig. 7.20). This is usually a horn antenna combined with a reflector shaped like a dish. Its shape is a mathematical parabola. Signals transmitted by the horn are reflected forward by the parabolic dish shaping the radio waves into a very narrow beam. Received energy is reflected by the dish into the horn antenna. The highly directional nature of the dish gives it a very narrow beamwidth and enormous gain. These are widely used in satellite TV and radar.

TWO-WAY RADIO

Two-way radio communications between individuals are by far the most widespread use of radio. And there are many forms of it. You have seen such radio communications as they take place and you have no doubt engaged in some form of it yourself. In any case, this section takes a look at some of the diverse uses of such two-way communications.

Most two-way radios operate in the VHF and UHF bands. The simplest two-way radios operate on multiple frequencies usually selected by a frequency synthesizer. Virtually all two-way radio communications are half duplex, meaning that one party transmits while the other receives. Transmit and receive operations are controlled with a push-to-talk switch on the side of the microphone or handheld unit.

Aircraft

All airplanes from the smallest single-prop planes to the largest jumbo superjets use two-way radio to communicate with airport towers and the major air traffic management centers across the country. Two-way radio is absolutely essential to the safety of any flight.

Most aircraft radios are transmitters installed in front of the pilot or copilot. They usually operate in the 108- to 135-MHz band and use AM. Specific channels are assigned for different regions and for different communications functions. The pilot or copilot uses front panel switches to select the desired communications frequency. Handheld microphones with a push-to-talk switch are the most commonly used input devices, but other forms of microphones such as throat mikes or microphones attached to headsets are also widely used. Most transceivers operate at very low power, and transmission distances are considerable. The height of the airplane gives the radio transceiver an enormous transmission range even with low power at VHF frequencies.

Citizens Band

You have no doubt heard of CB radio and perhaps even used one yourself. Anyone can buy and use a CB radio for two-way communications. The FCC has allocated 40 frequencies or channels in the 27-MHz frequency range for CB operation. The maximum transmission power is restricted to 5 W, although some handheld units use less. Modulation is AM but SSB is used in some units.

Citizens band radio has enjoyed wide popularity with individuals and sportsmen as well as truckers. While the advent of the cellular telephone has reduced the number of individuals using CB radio, it is still widely used. The primary limitation of most CB radios is the antenna. Since CB radios operate in the 27-MHz range, the required minimum antenna length is approximately 9 ft. This can be shrunk somewhat by the use of coils attached to the antenna, but effectiveness is reduced. A CB radio antenna of 5 or 6 ft is common and certainly manageable on a car, truck, or boat. In handheld transceivers, any antenna longer than about 18 in. is a nuisance.

A 27-MHz frequency allocation of CB radio also gives it some peculiar transmission characteristics. While communications is reliable over short distances of up to several miles, at certain times of the day or evening, worldwide communication is also possible. This is probably more a disadvantage than an advantage, primarily because interference can be received from stations literally anywhere in the world.

Amateur Radio

Amateur or ham radio is a hobby. Technically knowledgeable individuals are licensed by the FCC to use selected frequency bands across the spectrum for personal communications. Hams use FM or SSB voice as well as CW (continuous wave) or Morse code transmissions. Hams are also beginning to use digital modes such as PSK. Hams talk to one another about their equipment, the weather, and other common matters. Communication is worldwide and helps to promote good international relationships. Hams have proven their worth with emergency communications during hurricanes and other natural disasters. Hams frequently design, build, and experiment with their own equipment, although many hams prefer commercially manufactured radios. Hams also experiment with digital radios, satellites, and TV.

Family Radio

Family radio is another radio service recently instituted by the FCC for the purpose of short-range personal communications. It supplements CB radio in a wide range of personal communications needs. Because of a different frequency allocation, family radios are smaller, more portable, and have more reliable communications over distances of several miles.

All family radios are handhelds using FM and operating in the 462 and 467 MHz range. Only several frequencies in the 462/467 MHz band are allocated to the Family Radio Service (FRS). While many users will share these frequencies, the short communication distances minimize interference between stations using the same frequency. The transceivers are restricted to 0.5 W in power. At a frequency of 462 MHz, a quarter-wave whip antenna is only about 6 in. long. Communications are extremely reliable up to 1 mile and even farther if transmission takes place outdoors.

You can use family radios in a variety of ways. They are particularly handy when you are camping, hiking, or doing other outdoor activities. Many families use them in shopping malls and even around the house to keep in touch. They are great on farms and ranches and for communications between two nearby cars.

Marine

The very first application of two-way radio in the world was on ships at sea. Ships were always in danger without any form of communication. If a ship was disabled or began sinking, no one would know for a long period of time or maybe never. With a two-way radio or even a one-way radio on a ship, the operator could communicate with a shore-based radio station to signal danger with the familiar SOS.

Today, a radio is a necessity on virtually every boat and ship. While small boats are not required to carry radios, most larger boats are. And what seaman would be without a radio for safety reasons?

Marine radios have operated on a wide range of frequencies over the years. Ocean-going liners have operated on SW frequencies so that long-distance communications are possible. Ships also operate on lower frequencies for exchanging messages and at the higher frequency for reliable communications near land and in port.

The typical marine radio for communicating near ports is a transceiver of 25–40 W of power operating with frequency modulation in the 150- to 170-MHz range. Most of these units transmit on multiple frequencies, making it always possible to find a clear channel. Channel 16, operating on a frequency of 156.8 MHz, is used for emergency calls. This frequency is typically monitored by other boaters and always by the Coast Guard.

Land Mobile Radio

Land mobile radio refers to two-way radios mounted in cars and trucks. These are widely used by police, fire, and other public safety and civil services. They are also widely used by taxis, buses, trains, and by companies using fleets of cars or trucks for convenient communications. The typical unit is mounted in the vehicle and provides relatively high power in the 25- to 50-W range. Most communications take place in the 150- to 170-MHz range. Other mobile radio frequency ranges are 450–470 MHz. Many of the radios today are handheld types.

SATELLITES

A satellite is a space station that is designed to orbit the Earth and provide a remote radio signal relaying station for the purpose of extending the range of communications worldwide. See Fig. 7.21. Such communications satellites contain both a receiver and transmitter that operate simultaneously. The receiver

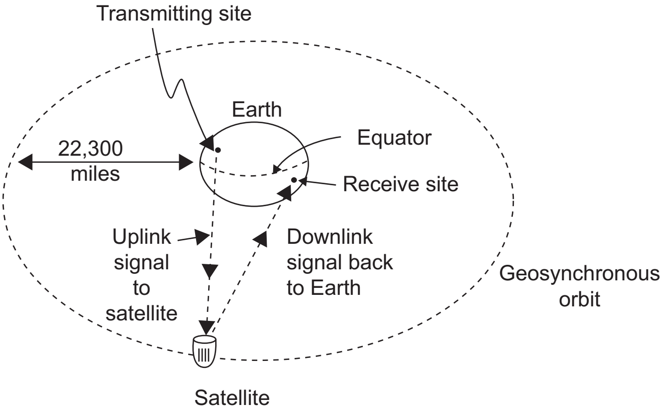

FIGURE 7.21 Concept of communications satellite.

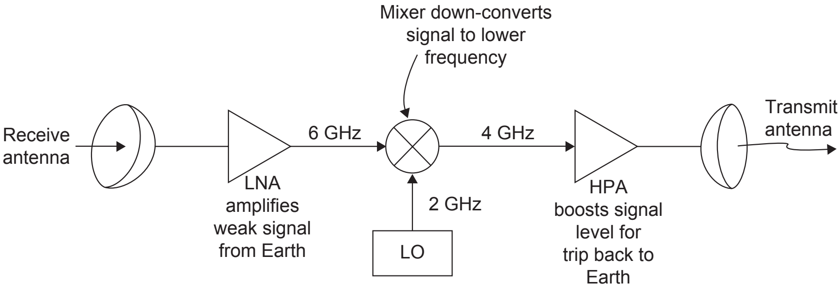
Mixer down-converts signal to lower frequency
Receive antenna
6 GHz
4 GHz
Transmit antenna
LNA amplifies weak signal from Earth
2 GHz
LO
HPA boosts signal level for trip back to Earth

FIGURE 7.22 Satellite transponder. *LO*, local oscillator.

picks up signals from the Earth on one frequency and retransmits them simultaneously on another frequency. In this way, the transmitter and receiver in the satellite do not interfere with one another.

Most communications satellites are geosynchronous. These are satellites positioned to orbit directly around the Earth's equator at a distance of 22,300 miles from Earth. When located in this position, the satellite rotates in exact synchronization with the Earth's rotation; thus, the term *geosynchronous*. Because the satellite and the Earth are rotating together, it appears as though the satellite is stationary overhead to an observer on Earth. With the satellite in a fixed position with respect to the Earth, the satellite becomes a near-perfect radio repeater station.

Fig. 7.22 shows a general block diagram of a satellite. The communications part of a satellite consists of a combined receiver and transmitter called a transponder. Satellites operate in the microwave region. The C band from 4 to 6 GHz and the K_u band from 11 to 18 GHz are the most popular operating ranges. A LNA picks up the very small 6 GHz uplink signal from Earth. It is mixed with a 2-GHz LO signal and downconverted to a 4-GHz downlink signal, which is then amplified by an HPA and retransmitted back to Earth. Most satellites contain many transponders with wide bandwidth capable of handling many signals simultaneously.

The satellite power source is batteries combined with a system of solar cells. The solar cells convert light from the Sun into electrical energy that is used to recharge the batteries. This system gives the satellite long life in space.

Communications satellites are used primarily in long-distance telephone communications and for distribution of TV signals. Some satellites are used for navigation, a good example being the GPS widely used by the military and commercial organizations. It is also widely used by individuals as they find their way with personal navigation devices like those from Garmin and Tom Tom. Many of the newer cars and trucks have built-in GPS navigation systems with maps using an LCD screen. The GPS system uses 26–31 satellites about 10,800 miles above the Earth orbiting in overlapping patterns so that at least three of four are "visible" by wireless at any point on the Earth.

Other satellites are used for surveillance. Spy satellites only 102 miles high pick up radio signals as well as IR images and photographs. Weather satellites monitor cloud formations and movements. There are also several satellite telephone networks with low-Earth-orbit satellites to permit communications to/from any point on the planet.

RADIO TELESCOPES

Observation of the planets and stars is usually done with an optical telescope. However, in 1932, researchers discovered that distant stars or suns emitted radio waves. All stars, like our Sun, emit radio frequencies. Most of this energy occurs in the microwave region at 1.4218 GHz. This is the frequency of ionized hydrogen. By building a highly sensitive radio receiver and coupling it to a large highly directional radio antenna, it is possible to plot the position of distant stars just as effectively as an optical telescope. Radio telescopes are, in fact, even more effective in that they can more easily detect distant stars that cannot be seen even by the best optical telescopes.

Radio astronomy observatories consist of an enormous parabolic dish antenna usually 60–250 ft in diameter, which can be rotated and tilted so that it scans the skies. An array of multiple smaller dishes can also be used as in the SETI (search for extraterrestrial intelligence) project in New Mexico. The antennas are coupled to one or more very sensitive radio receivers. These receivers have very high gain so that they can sufficiently amplify very small signals received from stars that are many light years away. Often special cooling systems are used to reduce the temperature of the receiver front end to minimize noise, thereby permitting even weaker signals to be received. One of the largest radio telescopes is a curved dish reflector built of screen mesh 1000 ft in diameter. The dish is built into the valley between adjacent mountains near Arecibo, Puerto Rico. The horn antenna is moveable on a long cable so it can be positioned to see any part of the sky. The world's largest radio telescope is in China.

Wireless Everything

It seems like everything is going wireless these days, such as cordless telephones, wireless local area networks, and your TV remote control. But wireless has always been around just not by that name. Most radio applications are old and well known. A few examples you may have encountered follow.

Garage door opener. A tiny transmitter in your car transmits a binary coded message to the receiver in your garage hooked up to the door motor.

Remote keyless entry. A miniature transmitter on your key ring transmits a signal to a receiver in your car that opens the door locks. Some even turn on lights and start your car from 10 to 30 ft away.

Wireless data acquisition. Remote transmitters are set up with solar batteries at remote sites to gather data from sensors and transmit it back to a collection

Continued

Wireless Everything—cont'd

center. A good example is electric, gas, and water utilities that use such systems to monitor critical conditions at remote sites and transmit the data back to a central monitoring facility where decisions can be made. Home wireless thermometers are very common.

Radio control of models. This is very common in toy cars, model airplanes, and boats.

Drones. Military, commercial, and recreational drones or unmanned aerial vehicles are all wirelessly controlled and that includes a wireless link for the video usually deployed.

RFID. Ultraminiature radios are being embedded in shipping labels, cars, and other objects for identification purposes. The units transmit their unique ID code when scanned by a nearby receiver.

HD radio. This is digital radio broadcasting in the United States. Most US AM and FM stations now transmit a digital version of their signals on the same frequencies as their analog transmissions. They use a form of digital modulation called OFDM (see Chapter 8). It produces greater frequency response than the regular analog broadcasts, making AM sound more like FM and FM sound more like a CD. A special radio is needed to receive it. The digital technology makes noise and fading less of a problem. And it permits each station to add several additional channels thanks to the multiplexing capabilities of the HD technology.

Security tags. The plastic tags clipped on new clothes to prevent their being stolen contain a tiny antenna and electronic circuit that make their presence known at a nearby transmitter/receiver unit.

Surveillance and bugging, or secret monitoring or electronic spying. This is not just something you see in the movies or on TV. It is more widely used in private investigation, police work, and industrial espionage than you may think. Scary!

Face it, wireless is truly everywhere.

Project 7.1

Investigate HD Radio

HD radio is digital radio in the United States that shares the same frequencies as the regular AM and FM bands. You can buy stand-alone HD receivers, and some stereo receivers include it. HD radio is also available in some cars. To learn more about this technology, go to the website of the developer iBiquity or www.ibiquity.com. Also try a Google/Bing/Yahoo search on HD radio.

Project 7.2

Become Familiar With ISM Radio

ISM radio is the term used to refer to short-range data radios used in the FCC's unlicensed spectrum designated for industrial–scientific–medical (ISM) applications. Got to the FCC website www.fcc.gov and find the most commonly used ISM frequencies of operation in Parts 15 and 18 of the rules and regulations. Do a general search of ISM to determine some common applications.

Project 7.3

Radar

Radar (radio detection and ranging) is a type of wireless. It has many uses, mainly military but also commercial and civilian. Marine and aircraft navigation are key applications. Police radar is widely used. Radar is also showing up in automobiles as detectors of distance between vehicles for automatic braking. Radar plays a major role in advanced driver assistance systems (ADAS) now being included in most new vehicles and in the forthcoming self-driving cars and trucks. You may wish to learn more about this technology. Do a search on radar.

Project 7.4

Shortwave Radio

Buy a low-cost SW radio and experience listening to international broadcasts and amateur radio operation.

Project 7.5

Experience Citizens Band or Family Radio Service

Go to your local electronic store and buy a pair of FRS walkie-talkies. Then test them to determine maximum range. As an alternative, buy a CB radio and listen to what is going on with the truckers or other locals.

Project 7.6

Experiment with a Short-Range Radio

This project lets you build a short-range radio system using ISM band transmitter (TX) and receiver (RX) modules. The modules are from Radiometrix and available from distributor Lemos. The US modules operate in the ISM band on a frequency of 914.5 MHz and the European modules operate on 869.85 MHz. The modulation is FSK, and data rates to 64 kbps are possible. The project will show how a digital signal can remotely flash an LED or control a relay. Use the following procedure.

- Go to the Radiometrix website and print out the data sheets and any application notes for the TX3A transmitter and the RX3A receiver: www.radiometrix.com. The modules are designed for PCB mounting but also plug easily into a standard breadboarding socket. Read the data sheets.
- Mount the transmitter and receiver on separate breadboarding sockets so that you can space them from one another.
- Provide power to each module separately. The modules will operate from any DC supply in the 2.2–16 V range. Use battery power for portability. You can use your 6 V AA supply or a 9 V battery.
- Wire the modules according to the circuit in Fig. 7.23. The 555 timer oscillator will supply the repeating square wave of data to the TX3A. The RX3A receiver output will drive a transistor switch to turn the LED on and off.

Continued

Project 7.6—cont'd

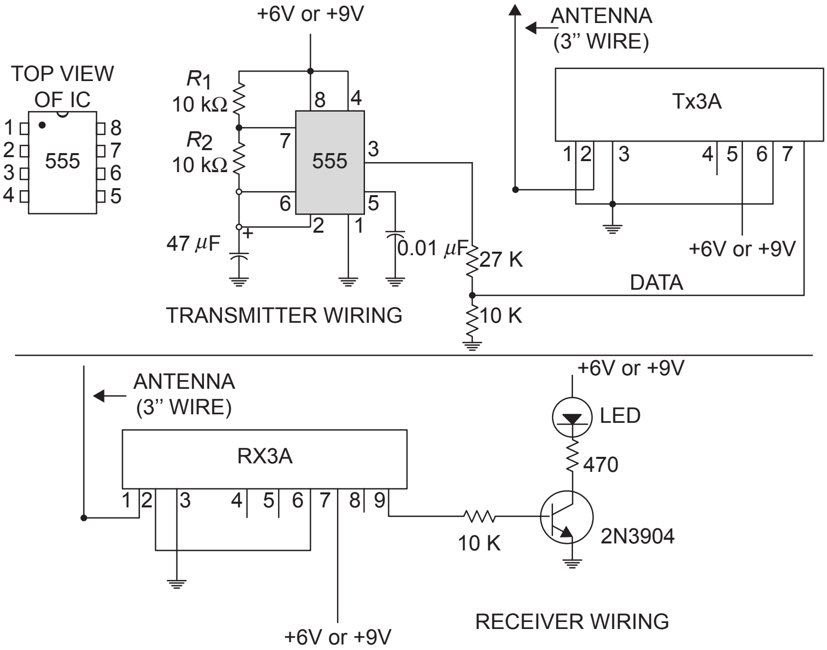
TOP VIEW OF IC
555
1
2
3
4
8
7
6
5
+6V or +9V
R1
10 kΩ
R2
10 kΩ
47 μF
+
8
4
7
3
6
5
2
1
0.01 μF
27 K
10 K
TRANSMITTER WIRING
ANTENNA (3" WIRE)
Tx3A
1 2 3 4 5 6 7
+6V or +9V
DATA
ANTENNA (3" WIRE)
RX3A
1 2 3 4 5 6 7 8 9
+6V or +9V
LED
470
10 K
2N3904
+6V or +9V
RECEIVER WIRING

FIGURE 7.23 Wiring details for the ISM band short-range radio experiment.

- For antennas use a 3 in. piece of #22 hook up wire. Just plug one end of the antenna wire into the transmitter RF OUT output on pin 2. The receiver wire antenna should be connected to the RF IN connection on pin 1. Orient the antenna wires vertically.
- Double-check your wiring before applying power to both units.
- Apply power. You should immediately see the LED flashing. This confirms that the wireless link has been established.
- Experiment with transmission range. Move the units as far apart as possible and determine where the connection fails.
- Disconnect power.
- Examine the circuit in Fig. 7.23. You could replace the 555 timer with a push-button and the LED with a relay to perform some remote control function. Other scenarios are possible.

Chapter 8

Cell Phones

It Is Now Possible to Do Anything Wirelessly: Talk, Text, Email, Web Browse, Games, Whatever

In this Chapter:

- How the cellular system works.
- Cell phone radio technologies.
- Digital modulation and access schemes.
- Cell phone data capabilities.
- Smartphones, 4G, LTE, 5G, MIMO, and location technology.

INTRODUCTION

The cell phone is one of the coolest, most useful, and convenient electronic devices of all time. It also happens to be one of the most complex technically, not only internally but also because it is part of the largest communications network in the universe, the telephone system. The reason for a separate chapter on this topic is not only because of the advanced wireless technology involved, but also its importance to all of us personally and to the national economy. Since practically everyone has a cell phone today, it plays a huge role in our lives. And for you techies out there, it is worthwhile knowing a bit of the details of how phone calls work, how text messages get sent, and how you can now access the Internet via your cell phone.

CAN YOU HEAR ME NOW?

We have all said that before. Why? For one simple reason: cell phones are two-way radios with all the usual faults. The cell phone industry usually refers to itself as the wireless industry, but it is all the same thing—radio. Your cell phone contains a *transmitter* and a *receiver* that collectively are called a *transceiver*. You will also hear them called *handsets*. People in the industry also call them *mobile terminals*. The most common designation of a cell phone is *smartphone*. It is much more than just a two-way radio. In fact the smartphones of today

Electronics Explained. http://dx.doi.org/10.1016/B978-0-12-811641-8.00008-4

Copyright © 2018 Elsevier Inc. All rights reserved.

contain another three to five radios. In addition, smartphones have been characterized as powerful handheld computers with multiple input/output radios. This later designation may be the most accurate.

Because cell phones have to be so small to be practical, they are limited in the amount of power they can transmit and the type of antenna that is practical. Battery drain is a major factor in cell phone design, as you want to get as much talk time as you can on a battery charge. But by limiting the transmit power, you automatically minimize the distance over which the unit will transmit. In any case, communications range is related to power as well as antenna type and orientation. And the typical range of a cell phone is a mile or so, depending on conditions. Take a look at the sidebar to review some of the facts about wireless and radio covered in Chapter 7.

Factors Affecting Range and Reliability to Radio Transmission

Range—The greater the distance between any transmitter and receiver, the smaller the signal. Signal strength varies inversely by the square of the distance between transmitter and receiver. Range is also affected by the frequency of operation. The higher the frequency, the shorter the range for a given transmitter power output.

Power—The higher the power, the greater the transmission range. More power translates into increased communications distance up to a point. For portable units, higher power means shorter battery life or faster discharge.

Antenna gain—Some antennas focus the signal in a specific direction. Doing so gives the antenna *gain*, meaning it multiplies the power of the transmitter and the signal strength at the receiver.

Environment—Cell phones and any other microwave wireless device always transmit farther in clear open space. The ideal is direct line of sight (LOS), meaning that the two antennas can "see" one another. While cell phones don't have to have such a clear path to work, they always work better if they do. Environment also means transmitting from inside a building. Yes, radio waves do pass through walls and other obstructions but as they do they lose considerable strength as they encounter obstacles. Walls, ceilings, floors, and any other obstructions greatly attenuate (make it smaller) the radio signal. So if you are inside a car, shopping mall, office, or airplane, your signal loses over half its strength and that severely restricts your range.

Height—Getting your cell phone to the highest possible point will greatly increase transmission distance. This helps the LOS effect so important to reliable transmission. Getting higher typically is not practical, but almost nothing helps more than height, especially outside.

Receiver sensitivity—This means how much gain the receiver has. High sensitivity translates into greater range.

Multipath—Whenever you are transmitting in an environment with lots of obstructions, your signal will bounce off many objects and be reflected. Some reflected signals will get lost and never get to the receiver. Others will bounce off objects that will ultimately direct them to their intended destination. These reflected

Factors Affecting Range and Reliability to Radio Transmission—cont'd

signals will be delayed in time. Depending on factors such as operating frequency and distance traveled, these signals could arrive in phase and actually boost reception level. Or they could arrive out of phase and literally cancel the more direct signal that arrived earlier.

Noise—Do not forget that noise also affects communications distance and reliability. See discussion in Chapter 7.

THE CELLULAR CONCEPT

The original designers of the cell phone created a system that helps overcome the range and power problem described in the sidebar. If distances are kept short, then other limitations in power and environment are offset. The working word here is "cell." A cell or cell site is one of many transmit/receive stations set up to communicate with individual cell phones. Also known as base stations, these facilities are easily recognized by their tall antenna towers. Modern fourth-generation (4G) cell sites are referred to as eNodeB. The small building at the base of the tower houses racks of transceivers that share the big antennas at the top of the tower. Long coax cables carry the signals to and from the antenna. They actually look like pipes going up the side of the tower.

Cell phone systems have dozens to hundreds of cell sites. The overall network is called the radio access network (RAN). An ideal design is shown in Fig. 8.1A. Each cell represents an area covered by the base station antennas. Cell range and overall coverage are deliberately restricted by antenna height and radiation pattern as well as transmitting power. The hexagon-shaped cells are only theoretical. In reality, the shape of the cell is more circular or rather an irregular circular shape because of antenna characteristics and environmental characteristics that affect the cell shape (nearby trees, buildings, etc.) (see Fig. 8.1B).

While equal-size cells are desirable and operation is more predictable, in the real world, the coverage of each cell depends a great deal on the number of cell phone subscribers as well as the terrain and environment. In large cities with lots of people and buildings, smaller cell sizes are used to increase subscriber capacity and provide the most reliable coverage possible in a given area. Many cities have what are called microcells that may only cover one block or picocells that cover one area of a building. On the other hand, in rural areas and along major highways, the cell sizes are larger and cover fewer users. The cell coverage may look more like that in Fig. 8.2.

A key feature of the cellular concept is that by keeping the cell sizes small, the frequency channels assigned to cellular service can be reused. One frequency may carry different calls if the cell sites are far enough apart and do not interfere with one another. This frequency reuse concept multiplies the overall number of channels available to handle calls.

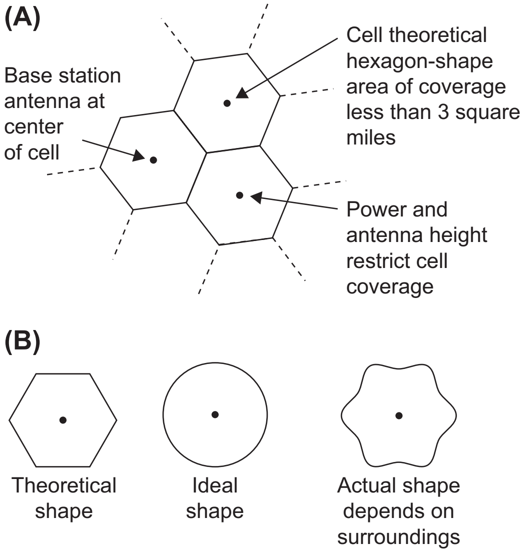
(A)
Base station antenna at center of cell
Cell theoretical hexagon-shape area of coverage less than 3 square miles
Power and antenna height restrict cell coverage
(B)
Theoretical shape
Ideal shape
Actual shape depends on surroundings

FIGURE 8.1 (A) Cell coverage of a given area. (B) Cell area covered depends on antenna pattern and environment.

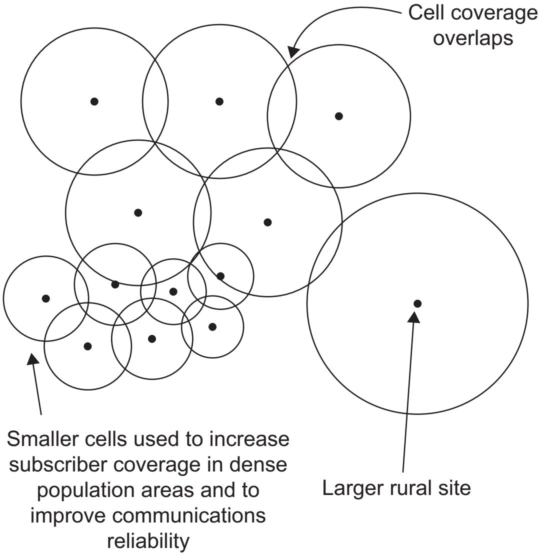
Cell coverage overlaps
Smaller cells used to increase subscriber coverage in dense population areas and to improve communications reliability
Larger rural site

FIGURE 8.2 How small and large cells cover an area.

One other factor to keep in mind is that in most cities, there are two or more cellular companies vying for your business. Each company has its own cellular system with base stations with coverage that overlays the cell coverage of other companies. The competing systems do not interfere with one another because they usually operate on different frequencies with and sometimes different radio technology, and the cell phones they support know which base stations to talk to.

One other thing: when you are talking on your cell phone, you are talking to only one base station—typically the one closest to you since both the power and range of your handset and that of the base station are limited. If you are moving (such as in a car), you will eventually leave the coverage area of one cell and enter the coverage area of an adjacent cell. You never really know when this happens, but it is a pretty neat technical trick to make this happen. The process of having your call transferred from one cell site to another automatically is called the *handoff*.

All of the base stations are tied to a central office known as the *mobile telephone switching office* (MTSO) or *mobile switching center* (MSC). Many companies simply refer to it as the *switch*, as its main job is switching between base stations as well as linking to the main telephone system. The connection between the base stations and the switching station can be by copper cable or fiber optic cable or in many cases by a separate microwave wireless link. This is known as *backhaul*. If a cell site tower has one or more small "dish" antennas on it, you can bet that your signal eventually is communicated back to the switch via another wireless link.

The big job of the MSC is to control and keep track of everything. It validates you as a subscriber when you turn on your phone and make a call. It keeps track of how many minutes of time you use. And it will get messages from remote switches when you are in the roam mode outside of your usual home subscriber area. The cell phone companies have the ability to use other systems in most major cities and in many foreign countries, so you can use the phone virtually anywhere. Fig. 8.3 shows just what goes on when you turn on your phone and make or receive a call.

WHAT IS INSIDE A CELL PHONE?

A cell phone can legitimately be called the world's most complicated two-way radio. Sure, it is just a transceiver, but it uses practically every radio trick in the book to ensure that you can make your calls reliably anytime with minimal hassle. It is a two-way radio that is used just like any other phone you are familiar with. In other words, you do not have to say "over," "come back," or any other phrase when you are finished talking and want to listen, as with some two-way radios (citizens band, family radio, ham radio, aircraft, marine, etc.).

Simultaneous send and receive is referred to as *full duplex*. All telephones are full duplex. It is easy to do in the wired telephone system but complex and costly to do by radio. Yet, full duplex is what makes cell phones so comfortable to use. There is none of this "roger and over" stuff. What this means is that your

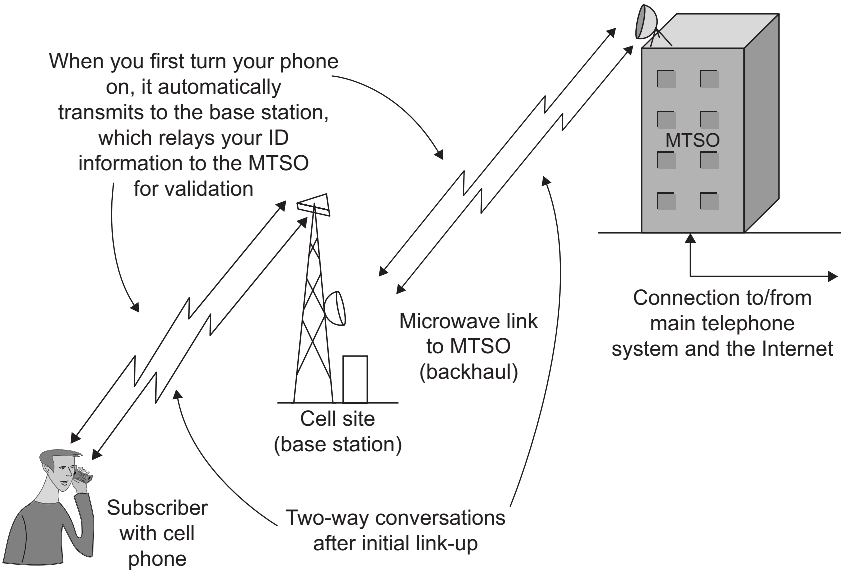
When you first turn your phone on, it automatically transmits to the base station, which relays your ID information to the MTSO for validation
MTSO
Microwave link to MTSO (backhaul)
Connection to/from main telephone system and the Internet
Cell site (base station)
Subscriber with cell phone
Two-way conversations after initial link-up

FIGURE 8.3 Cell phone basic operations.

transmitter and receiver are working at the same time. There is more about how duplexing works later.

Fig. 8.4 shows a very simplified general block diagram of what is in a cell phone. The receiver and transmitter making up the transceiver are obvious, and they share the single antenna. The transceiver is usually a single chip. What is not so obvious is what we call the baseband part of the cell phone. Baseband refers to the voice and data to be transmitted or received. In most modern digital cell phones, this is a very complex integrated circuit that handles the translation of voice between analog and digital, modulation and demodulation, and voice compression. All these functions are done digitally by a digital signal processor (DSP) or special DSP circuits in combination with a powerful embedded processor. This processor or a separate processor handles all the transparent housekeeping jobs of managing the touch keyboard and the LCD screen as well as the basic control features of the cell system, which also include automatic operating frequency selection and automatic power control under the direction of the cell site. The DSP also implements the messaging functions.

What Type of Cell Phone Technology Do You Have?

The cell phone radio and networking technologies have changed over the years. Each technology is known by its generation. We are currently in the fourth generation, and most new phones use a 4G technology known as Long-Term Evolution (LTE). But some older generation are still around.

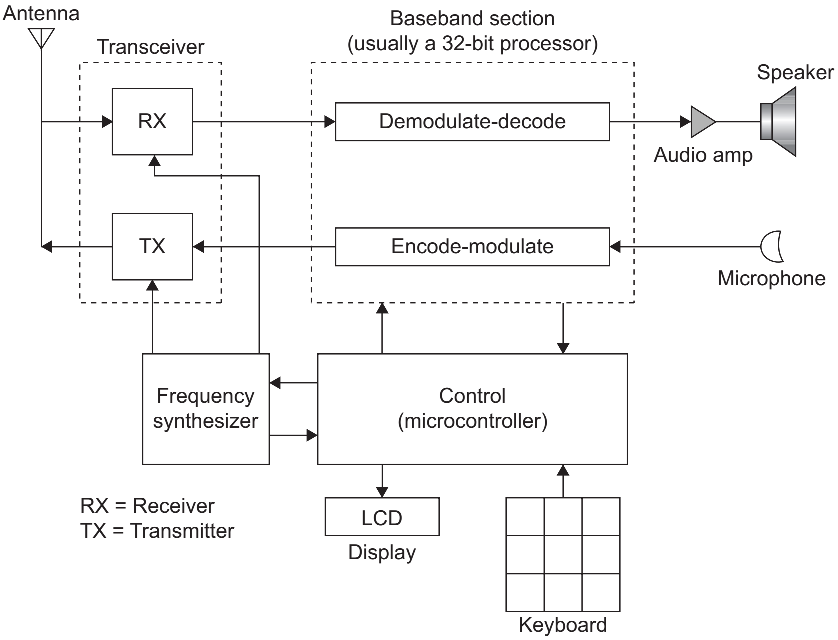
Antenna
Transceiver
RX
TX
Baseband section
(usually a 32-bit processor)
Demodulate-decode
Encode-modulate
Speaker
Audio amp
Microphone
Frequency synthesizer
Control (microcontroller)
RX = Receiver
TX = Transmitter
LCD
Display
Keyboard

FIGURE 8.4 Block diagram of basic cell phone.

The first characteristic of a cell phone is the frequencies it uses. Most cell phones operate in the 800- to 900-MHz band where there are hundreds of channels for cell phone calls. Some 700-MHz spectrum is also now in use. Another range of frequencies is the so-called personal communications system (PCS) band from 1850 to 1990 MHz (1.85–1.99 GHz), which has hundreds of more channels. Some of the newer phones use the advanced wireless services (AWS), 1700- and 2100-MHz spectrum assignments. Some cellular companies also use spectrum in the 2300–2500 MHz range. Because of the higher frequency, PCS and AWS systems have a shorter range, and thus use smaller area cells, which means many more cell sites to cover a given area. Most cell phones actually operate on two or more bands that are assigned to your carrier.

Next, a cell phone is known by the technology and access methods it uses. The early original analog phones used first-generation (1G) technology, or FM, but have now been phased out. Most second-generation (2G) digital phones have also been phased out although there probably are some still in use in third world countries. Some third-generation (3G) cell phones are still with us to some degree; however, most phones today are of the 4G variety and fifth-generation (5G) is almost with us. A summary of these technologies follows.

GSM—The Global System for Mobile Communications (originally called Groupe Spécial Mobile) is the 2G digital system developed in Europe. It uses time division multiple access (TDMA), a digital system using digital modulation and a method of multiplying the number of channels in a given amount of frequency spectrum. Your voice is converted to binary 1's and 0's before being transmitted. This allows the carrier to put eight subscribers in one 200-kHz-wide frequency channel, thereby greatly multiplying their subscriber capacity and income. GSM uses a form of FSK called Gaussian minimum shift keying (GMSK). GSM uses several enhancements that facilitate high-speed data transmissions such as the General Packet Radio System (GPRS) and Enhanced Data Rate for GSM Evolution (EDGE). These are no longer widely used.

CDMA—Code division multiple access is another 2G digital cell phone technology developed by San Diego–based Qualcomm Inc. The basic technology is known as *spread spectrum* where many signals are transmitted simultaneously over a very wide 1.25-MHz frequency band without interfering with one another. The original standard is called IS-95, and CDMA signals coexist in the same spectrum with TDMA systems in many areas.

cdma2000—This is the 2.5G/3G version of CDMA. It adds high-speed packet data transmission called 1xRTT with a data rate up to 144 kbps. A more recent upgrade called 1xRTT EV-DO has even higher data rates up to 2 Mbps, making Internet access a cinch. Many cdma2000 systems are still in operation around the United States and other countries.

WCDMA—This is wideband CDMA, which is a 3G technology. The access technology is CDMA, but it uses wider 5-MHz bands to provide greater user capacity and very high potential data rates from 384 kbps to 2 Mbps. WCDMA offers greater data speeds and services with higher subscriber capacity.

HSDPA/HSUPA—High-speed downlink packet access (HSDPA) and high-speed uplink packet access (HSUPA) are upgrades from the WCDMA system. Both use QAM to boost data speeds in the same 5-MHz channels. Data rates as high as 14 Mbps downlink and 5 Mbps uplink are possible. Most cellular companies have now implemented this technology. More advanced versions called HSPA or HSPA+ can give even higher data rates. Such systems are no longer widely used as it has been replaced by 4G technology.

cdma2000 Rev A and **Rev B**—These are more advanced CDMA versions that are used to boost data speeds in cdma2000 handsets. Rev A boosts download data rates to 3.1 Mbps using QPSK and 16-QAM in the standard 1.25-MHz channel. Rev B is not widely implemented, but can use three 1.25-MHz channels and boost download speeds to 14.7 Mbps and higher. Like the HPSA 3G technology these systems have been mostly replaced by 4G systems.

Long-Term Evolution—LTE is the current 4G cell phone technology. It uses an entirely different radio technology called orthogonal frequency division multiplexing (OFDM) to give even higher data rates. More on that later.

Today most phones are 4G LTE and will fall back to 3G if no LTE service is available. As this is written the 5G standards are still being developed and finalized.

DIGITAL MODULATION AND WIRELESS TRANSMISSION METHODS

In Chapter 7, basic digital modulation methods ASK/OOK and FSK were covered. These are not used except for an advanced FSK as used in GSM. Instead, all cellular systems used advanced digital modulation methods. This section provides a look at these techniques.

Quadrature Amplitude Modulation

Quadrature amplitude modulation (QAM) is a technique that mixes both amplitude and phase variations in a carrier at the same time. What this technique does is to allow higher data rates within the same bandwidth. See the sidebar on spectral efficiency. The technique consists of transmitting multiple bits for each time interval of the carrier *symbol*. The term "symbol" means some unique combination of phase and amplitude. The term M as in M-QAM indicates how many bits are transmitted per time interval or symbol for each unique amplitude/phase combination.

The simplest form of QAM is 2-QAM, more commonly called QPSK or quadrature phase shift keying. It is produced by the circuit shown in Fig. 8.5. It takes the serial bitstream and passes it through a 2-bit shift register producing two parallel bitstreams at half the rate. One is called the in-phase (I) bitstream and the other the quadrature (Q) bitstream. The I and Q signals are each applied to a mixer along with a local oscillator (LO) signal. The mixer output becomes part of the RF carrier. Note that the LO signal is applied directly to one mixer, but is shifted by 90 degrees and applied to the other mixer. The term "quadrature" means shifted by 90 degrees. The two mixer outputs are AM signals but at different phases. But when they are added together, the QPSK signal is generated. With this arrangement, you can transmit data at twice the rate of BPSK in the same bandwidth.

Fig. 8.6 shows how we illustrate this with what we call the phasor and constellation diagrams. Fig. 8.6A shows the amplitude and phase of each possible

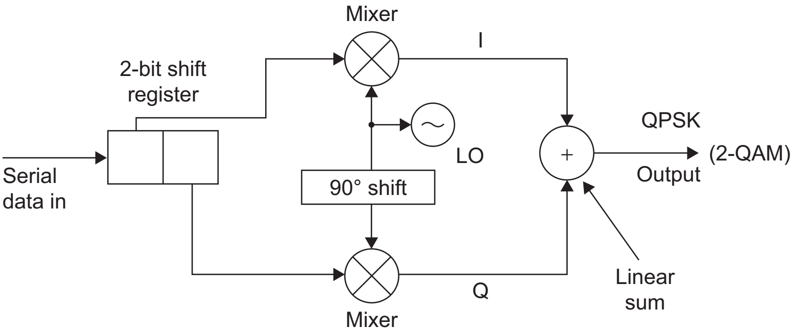

FIGURE 8.5 How QPSK is generated using mixers and a local oscillator shifted 90 degrees to produce in-phase (I) and quadrature (Q) signals that are summed.

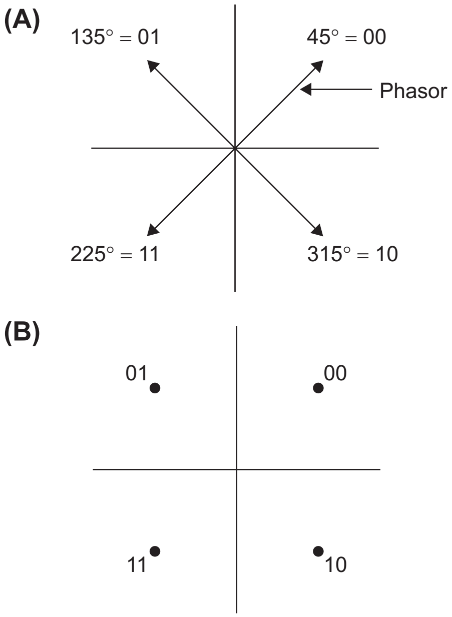

FIGURE 8.6 QPSK diagrams. (A) Phasor. (B) Constellation.

combination of 2 bits. The length of the arrow called a *phasor* is the amplitude, and its direction or angle is the phase. If the receiver gets a signal that is at a phase angle of 135 degrees, it means that the 2-bit combination 01 is being transmitted. The constellation diagram in Fig. 8.6B is usually shown without the arrows or phasors.

This technique can be modified to produce 16-QAM, which transmits 4 bits per time interval or symbol causing the data rate to quadruple in the same bandwidth. The circuit for producing 16-QAM is shown in Fig. 8.7. The 4-bit shift register produces two groups of I and Q bits. These are converted into four levels by a two- to four-level converter that works like a digital-to-analog converter (DAC). The resulting constellation diagram for 16-QAM is shown in Fig. 8.6 as well. Other common versions are 64-QAM and 256-QAM, which transmit 6 and 8 bits per symbol, respectively, in the same bandwidth. A 1024QAM is also an option in some systems.

QPSK and QAM modulation methods are common in cable TV for digital TV as well as high-speed Internet service. They are also used with satellites and any broadband wireless application. All modern cell phones and most other digital wireless schemes use this type of modulation because of its spectral efficiency.

Orthogonal Frequency Division Multiplexing

OFDM is a broadband modulation method like spread spectrum/CDMA. It takes a high-speed serial binary signal and spreads it over a wide bandwidth. The

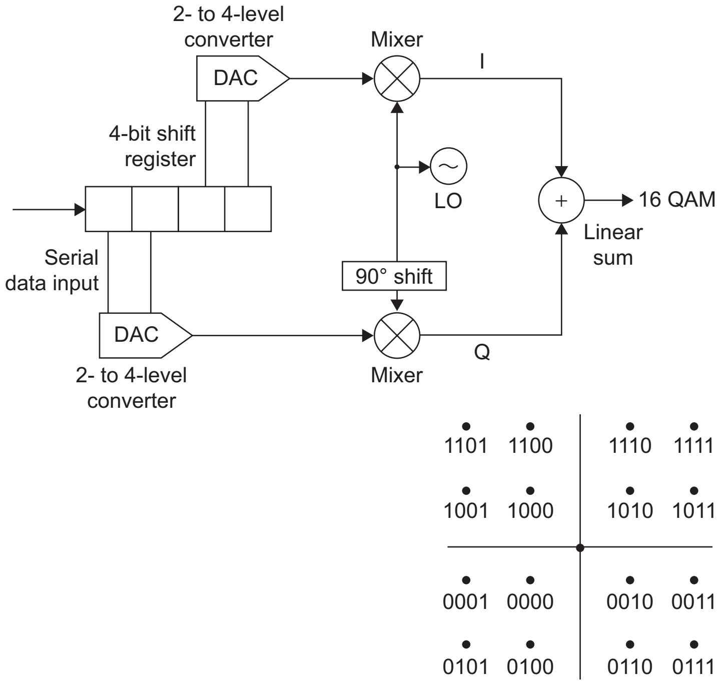
2- to 4-level
converter
DAC
Mixer
I
4-bit shift
register
LO
+
16 QAM
Linear
sum
Serial
data input
90° shift
DAC
Q
2- to 4-level
converter
Mixer
1101
1100
1110
1111
1001
1000
1010
1011
0001
0000
0010
0011
0101
0100
0110
0111

FIGURE 8.7 How 16-QAM is generated and resulting constellation diagram.

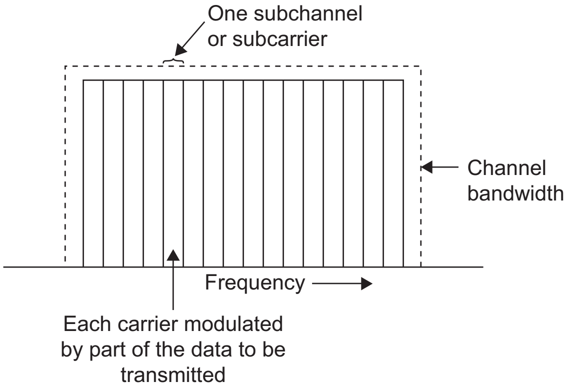
One subchannel
or subcarrier
Channel
bandwidth
Frequency
Each carrier modulated
by part of the data to be
transmitted

FIGURE 8.8 OFDM divides a channel into many subchannels or subcarriers, each of which is modulated by part of digital data to be transmitted.

serial data is passed through a circuit that maps out the constellation diagram for the modulation to be used. It then divides it into many slower-speed serial bitstreams. Each bitstream modulates a carrier on one of many adjacent carriers in the available bandwidth. This technique effectively divides a wide bandwidth into many narrower subchannels or subcarriers as shown in Fig. 8.8. Sometimes dozens of channels are used, and in other cases hundreds or even thousands of carriers are used. The carrier frequencies are selected so they are orthogonal and as a result they will not interfere with one another even though they are directly adjacent. All the modulated channels are then added together and the combination transmitted in the available bandwidth. The type of modulation varies with the application but it is usually BPSK, QPSK, 16-QAM, 64-QAM, or 256-QAM.

The basic technique is shown in Fig. 8.9. The only practical way to create an OFDM signal is by using DSP. The DSP executes an algorithm called the inverse fast Fourier transform (IFFT) that creates the multiple adjacent modulated carriers. The DSP data is applied to DACs to create I and Q analog signals that are sent to mixers for creation of the final signal. That signal is then amplified by a broadband linear amplifier before being transmitted. At the receiver, the data is recovered by another DSP executing the FFT. The output is the original fast serial data.

OFDM seems impossibly complex while seemingly hogging spectrum space. Yet, it can transmit higher speeds in smaller bandwidths than most other types of digital modulation. It is very spectrally efficient. Furthermore, it is more resistant to multipath interference that can cause microwave links to lose data due to a reflected signal interfering with another or a direct signal causing cancellation and fading. Most of the new wireless technologies today use OFDM, including the LTE cellular systems. Some examples of other OFDM uses include wireless local area networks (WLANs) such as Wi-Fi, wireless digital subscriber line (DSL) Internet access, some (European) digital TV, and AC power line networking.

Multiple Input Multiple Output

Another technique for boosting speed within a given bandwidth is multiple input multiple output (MIMO). Also known as spatial multiplexing, it is the use of multiple transmitters transmitting to multiple receivers, each sending different bitstreams on the same channel. The high-speed serial data is

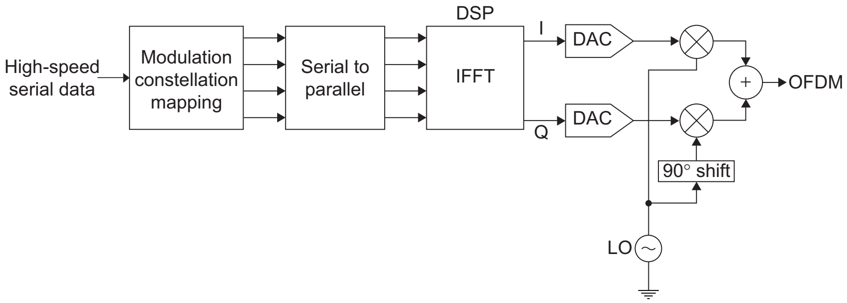

FIGURE 8.9 Generating OFDM is done by DSP and I/Q modulation.

divided into two, three, four, or even more bitstreams, and each is used to modulate a different transmitter but operating on the same frequency. The signals from each transmitter reach the antennas of each receiver along with any reflected signals. See Fig. 8.10, which shows four transmitters (Tx) and four receivers (Rx) or 4×4 MIMO. What allows the receivers to sort out each bitstream is the fact that each signal travels a different path and will have different characteristics or spatial signature that can be sorted out by DSP in the receiver.

There are several things that make MIMO practical. First, MIMO is used at very high frequencies, especially microwave (0.1 GHz). This means that antennas are short, and interantenna space may be one wavelength or more, which is required for MIMO to work. The distance between the antennas ensures different spatial paths with different characteristics that let the receivers recover each bitstream. Second, thanks to semiconductor technology, it is possible to make the multiple transmitters and receivers very inexpensively, typically all on a single silicon chip.

MIMO has several benefits. First, it allows a higher data rate in the same channel bandwidth. Using a 4×4 MIMO, for example, would boost speed by a factor of 4. Configurations of MIMO can be any combination of transmitters and receivers. The 2×2 mode is popular but 4×4 is also used. Second, MIMO provides a more reliable link. It actually takes advantage of the multipath problems that usually plague microwave wireless and makes it better. Finally, MIMO is usually combined with OFDM, which provides the ultimate data rate boost, reliability, and spatial efficiency. MIMO is already used in WLANs such as Wi-Fi. MIMO is also used with LTE Handsets typically have only two antennas, but the base stations have four or more.

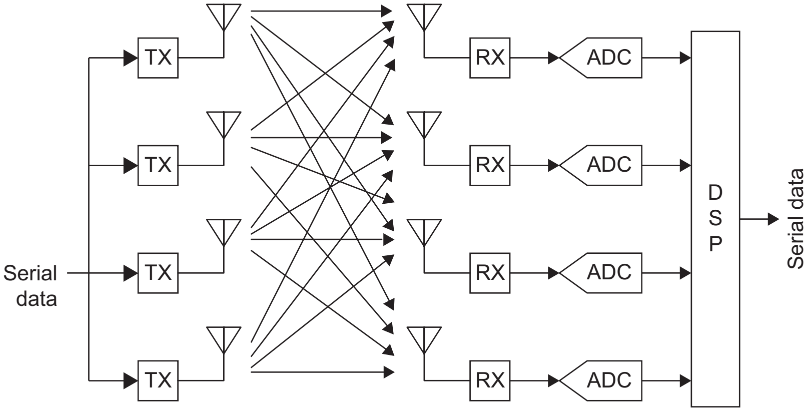

FIGURE 8.10 MIMO in a 4×4 configuration quadruples the data rate in an existing channel, while reducing the ill effects of multipath fading.

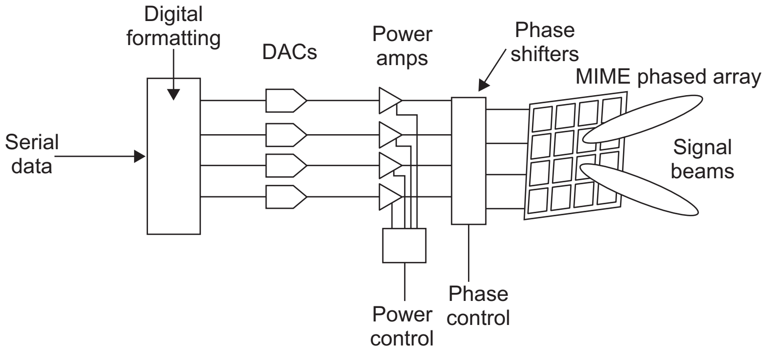

FIGURE 8.11 The concept of MU-MIMO using beamsteering with a patch-type phased array antenna.

An advanced version of MIMO called multiuser or MU-MIMO is also being widely adopted. Also called massive MIMO because it uses a large (64–1024) number of antennas combined to form what is called a phased array. Each antenna has a transmitter and receiver associated with it. By controlling the phase and amplitude of the signals sent to the antennas, the signal can be focused into a narrow beam creating gain in the antenna. Fig. 8.11 shows the concept with a 16-patch antenna array. In addition to beam forming, the signal beam can also be steered in different directions horizontally or vertically. Another feature is that separate multiple beams can be formed so that several different independent signals can be sent or received at the same time. By using massive MIMO, multiple users may be served by a single base station and higher data rates can be achieved.

Spectral Efficiency

Spectral efficiency refers to how much data you can transmit in a given bandwidth. It is usually measured in bits per second per hertz (bps/Hz). ASK/OOK and FSK have efficiencies of <1. Special forms of FSK, such as GMSK used with older GSM cell phones, are a bit better. BPSK has a spectral efficiency of 1. Table 8.1 gives the efficiencies of most of the popular digital wireless modulation schemes.

One last thing: the type of modulation is not the only thing that affects spectral efficiency. The level of noise in a channel also determines the number of bit errors that occur in transmission. The higher levels of QAM have more errors, so this limits their efficiency to some extent. Plain FSK, BPSK, and QPSK are very resistant to noise, so are more suitable in applications where noise is a problem. Noise and bit rates have to be traded off for good reliable transmission. And by the way, the spectral efficiency issue applies to wired as well as wireless communications channels.

TABLE 8.1 Spectral Efficiency of Selected Digital Wireless Modulation Schemes

Modulation Method	Spectral Efficiency (bps/Hz)
FSK	1
GMSK	1.35
BPSK	1
QPSK	2
16-QAM	4
64-QAM	6
256-QAM	8
OFDM	Depends on modulation of subcarrier and number of subcarriers
MIMO	Depends on configuration

ACCESS METHODS AND DUPLEXING

Access methods are wireless techniques that allow multiple users to share a given part of the frequency spectrum. Access methods are also called *multiplexing*. Multiplexing means allowing two or more signals to be transmitted *concurrently* over the same communications medium such as a cable or within a given frequency band. Access methods allow many subscribers to share precious spectrum space. And they provide the means to implement full-duplex simultaneous transmit/receive that most phone users expect. There are four types: frequency division multiple access (FDMA), TDMA, CDMA, and orthogonal frequency division multiplex access (OFDMA). For full-duplex operation, there are two methods—frequency division duplexing (FDD) and time division duplexing (TDD).

FDMA—A given bandwidth is divided into many narrower channels as shown in Fig. 8.12. A given radio will cover all of the bands but has the ability to select a specific operating frequency. GSM divides the band into 200-kHz channels. cdma2000 uses 1.25-MHz channels and WCDMA uses 5-MHz channels. The newer OFDM uses 5-, 10-, 15-, or 20-MHz channels.

FDD—Simultaneous transmit and receive is accomplished by using two different bands where half are used for transmit and the other half for receive. The transmitter and receiver operate at the same time but on different frequencies so they do not interfere with one another (see Fig. 8.12).

TDMA—With this method, a single-frequency channel is used, but the signal is divided into fixed-duration time slots into which digitized segments

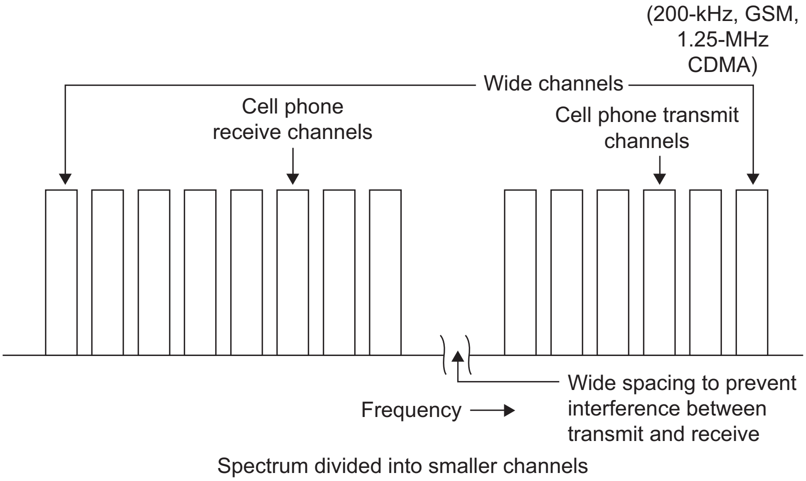
(200-kHz, GSM,
1.25-MHz
CDMA)
Wide channels
Cell phone
receive channels
Cell phone transmit
channels
Wide spacing to prevent
interference between
transmit and receive
Frequency
Spectrum divided into smaller channels

FIGURE 8.12 Frequency division multiple access and frequency duplexing use two separate bands of frequencies divided into individual channels.

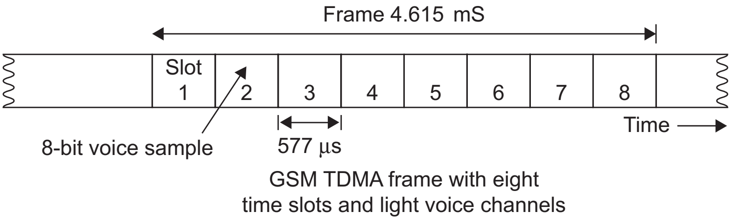
Frame 4.615 mS
Slot
1
2
3
4
5
6
7
8
8-bit voice sample
577 µs
Time
GSM TDMA frame with eight
time slots and light voice channels

FIGURE 8.13 TDMA multiplexing divides a channel into sequential time slots for the digital data.

of voice or data are transmitted. In digital voice transmission, the analog voice signal is periodically sampled and digitized into a stream of binary numbers. These binary values are then compressed into a smaller number of bits by a vocoder circuit. The resulting bits are then transmitted to the receiver. With TDMA, binary values from two or more voice sources can be interleaved and sequenced, and then sent over the single channel. For example, in GSM systems, there are eight time slots, as shown in Fig. 8.13. The data is transmitted at a 270-kbps rate in a 200-kHz-wide channel. Each time slot contains 1 byte of the voice data for one phone call. The frame of eight signals is then repeated.

TDD—TDD facilitates concurrent send and receive by assigning transmitted signals in one time slot and received signals in another time slot. They share the same frequency channel. Because all this is happening at a high rate of speed, you cannot determine that transmit and receive are not simultaneous, although in reality they are not.

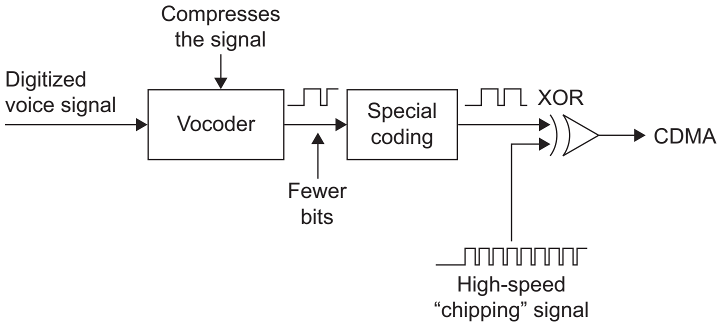

FIGURE 8.14 How CDMA is generated.

CDMA—This method uses spread spectrum technology. It takes the digitized voice and further processes it to give it special coding, and then spreads it over a wide bandwidth. In cdma2000 phones, the channel bandwidth is 1.25 MHz wide, and all cell phone signals are coded and spread over this bandwidth. The process goes like this: the analog voice signal produced by the microphone is digitized in an ADC, and then compressed by a special digital circuit called a *vocoder* or *voice coder*. This reduces the total number of bits and the speed of the digital signal. This vocoded signal is then mixed with a high-speed serial code signal, making it unique so that the receiver will recognize it. This is done with an XOR gate as shown in Fig. 8.14. The higher-frequency signal, at 1.23 Mbps, is called the *chipping data*. The resulting signal fills the 1.25-MHz channel. In WCDMA systems, the process is the same, but the chipping signal is 3.84 Mbps and the channel width is 5 MHz.

Using CDMA many cell phone calls can be transmitted on the same channel. They actually all mix together, but because they occur randomly with one another, they do not interfere with one another. Typically, up to about 64 users can share the same bandwidth. One signal simply appears as a low-level noise to another. The desired signal is picked up at the receiver by recognizing the unique code given to the transmitted signal. The two most common types of spread spectrum are direct sequence and frequency hopping. Direct-sequence spread spectrum is used in cell phones as illustrated in Fig. 8.14. Frequency hopping is used in Bluetooth wireless systems (see Chapter 9). Most duplexing is by FDD, but some of the newer versions of LTE also use TDD.

OFDMA—Access with OFDM is done by just assigning one or more of the subchannels to each transaction. With each subchannel 15 kHz wide, each is capable of carrying a compressed (vocoded) voice signal so that many calls or data transactions can be handled. In practice, more than one channel may be used, depending on the specific frequency assignments, bandwidths, and other features of the LTE system being used.

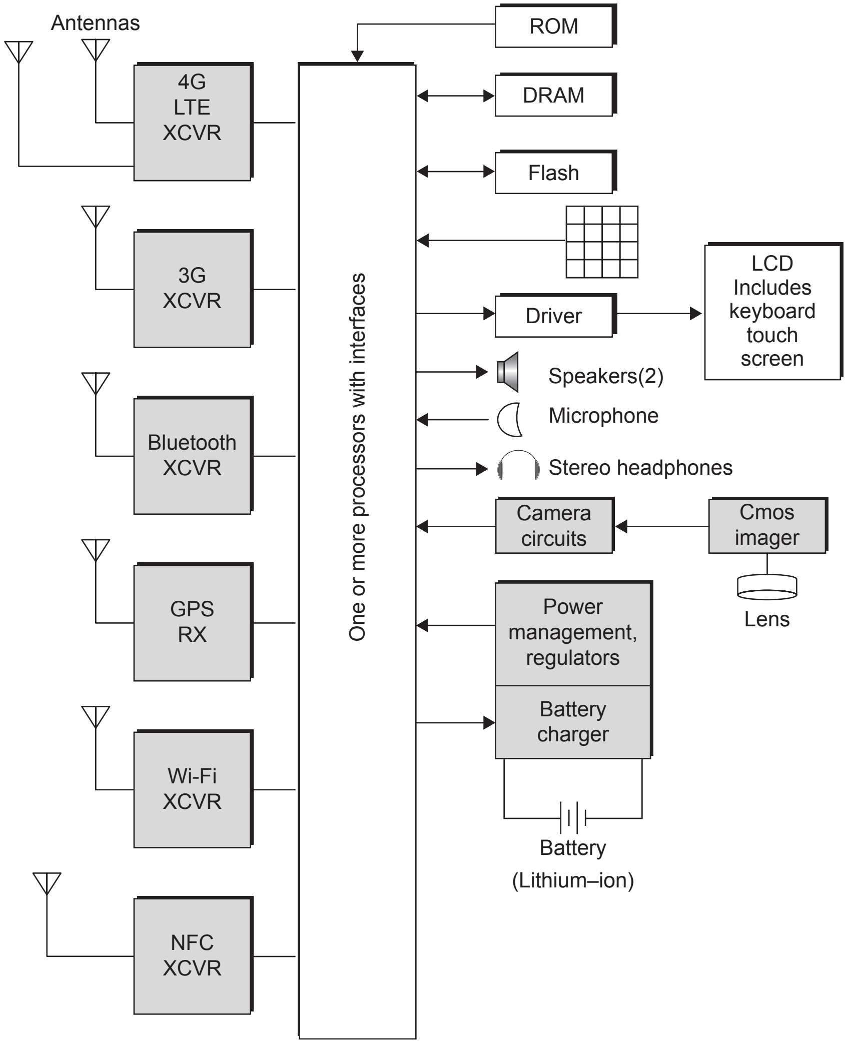
Antennas
4G
LTE
XCVR
3G
XCVR
Bluetooth
XCVR
GPS
RX
Wi-Fi
XCVR
NFC
XCVR
One or more processors with interfaces
ROM
DRAM
Flash
Driver
LCD
Includes
keyboard
touch
screen
Speakers(2)
Microphone
Stereo headphones
Camera
circuits
Cmos
imager
Lens
Power
management,
regulators
Battery
charger
Battery
(Lithium–ion)

FIGURE 8.15 A block diagram of 4G cell phone with multiple radios and functions.

A LOOK INSIDE A 4G CELL PHONE

Fig. 8.15 shows the insides of a modern 4G smartphone. This one uses the LTE 4G technology, but is also backward compatible with 3G. There are separate radio transceivers (XCVR) for each. All of the baseband functions such as modulation and demodulation and handset management are handled by one or more processors. Dual core or quad core processors are used. One processor implements DSP. This handset also has a Bluetooth radio for the headset (not shown) and a Wi-Fi transceiver for Internet access via a WLAN access point or hot spot.

Most of the newer smartphones also contain a near-field communications transceiver for payment and access applications. These technologies will be covered in the next chapter. A global positioning system (GPS) radio provides navigation as well as fulfills the E911 location requirement. Note that most radios have separate antennas. In some cases if the frequency range is the same, an antenna can be shared. A camera for digital stills and video is included. The CMOS imager is the device that converts the light and color variations in the scene into electrical signals. Other light sensor types are used. The stereo headphones are for the music player. A recent addition to many smartphones is wireless charging using inductive coils. Finally, all smartphones have a massive amount of memory, both DRAM and flash. The power supply is a lithium-ion battery along with the charging circuits, regulators, and power management circuits.

THE LATEST CELL PHONE TECHNOLOGY

The cell phone industry is one of the fastest changing in electronics. New products come out almost daily. And new features and capabilities are invented regularly to keep the technology fresh and the sales continuous. Here is a snapshot in time of the latest cell phone developments.

Long-Term Evolution

LTE is the current 4G cell phone technology. Most modern smartphones use LTE with a fallback to 3G systems where LTE is not available. LTE uses OFDM with QAM modulation in 10 or 20 MHz channels along with MIMO to provide downlink data rates of up to 150 Mb/s. More recent versions such as LTE-Advanced and LTE-Advanced Pro use carrier aggregation (CA) to widen the bandwidth by combining up to 32 channels to achieve 100 MHz or more of bandwidth to get even higher speeds. With the right combination of CA, 256-QAM and 4×4 MIMO, LTE can achieve data rates up to 1 Gb/s.

The advanced versions of LTE-A also use unlicensed spectrum to further increase bandwidth. Features such as Licensed Assisted Access (LAA), LTE-WLAN (LWA), and LTE-U for unlicensed combine LTE spectrum and the unlicensed 5 GHz spectrum used by Wi-Fi 802.11ac/ax to boost downlink speed.

LTE-A Pro adds IoT/M2M connectivity (see Chapter 9). LTE-M uses only 1.4 MHz of bandwidth to achieve speeds up to 1 Mb/s. An ever more frugal version called NB-IoT uses 180 kHz to achieve a data rate to 250 kb/s. These new versions use less power and will find applications in IoT and in vehicle-to-everything (V2X) applications such as the advanced driver assistance systems (ADAS) and self-driving cars.

LTE Advanced also provides for a feature called device-to-device (D2D) communications. It lets LTE handset users talk directly to one another rather than through a cell site. Public safety officials are expected to adopt this version of LTE in lieu of more traditional two-way radio technologies.

LTE is the dominant worldwide cell phone technology. Many cellular operators are still up grading to the advanced versions. LTE works well and satisfies most of the demands of subscribers. However, the future is expected to require even faster networks. For that reason, the 5G technology is under development.

Fifth-Generation Wireless Overview

Several key trends are driving the development of a newer and better cellular system. One trend is the increased use of mobile video and the need to be able to handle 4K resolution video and virtual reality. Higher data rates (>1 Gb/s) are needed for some applications.

Another trend is the increase in connected devices using IoT or M2M technologies. Billions of devices may be connected, and many will use the cellular system that must be able to handle the increased number of subscribers.

In addition to a higher data rate, another major need is lower latency as it is critical to driving safety as well as many IoT/M2M uses in factory automation. Latency is the delay time between the initial triggering of some action and the instant the action occurs. Many robotic and other machine operations demand millisecond (ms) range latency. 4G radios can usually have a latency of 25 ms or more. 5G is expected to provide for a latency of 5 ms or less.

The 5G standards are being developed by the Third Generation Partnership Project (3GPP). The new standard is called New Radio (NR). Here is a summary of 5G NR features.

- Maximum data rates should be in the 1 to 10 Gb/s range depending on the bandwidth allotted and signal path environment. Typical mobile data rates should be hundreds of megabits per second.
- To achieve these data rates, more bandwidth is needed and in many instances new spectrum will be needed. Some systems will use available spectrum below 6 GHz. However, most 5G systems will use the millimeter wave (mmwave) bands. In the United States the 27–28 and 37–40 GHz bands have been allocated to 5G with most operators are building systems in the 28 GHz band.
- CA (carrier aggregation) will be used to further increase bandwidth to support the desired higher data rate. Contiguous or noncontiguous bands as well as unlicensed spectrum can be aggregated to achieve the desired bandwidth.
- Modulation will be some variation of OFDM. Multiple schemes have been considered. The final standard has not been announced. QAM up to 256-QAM will be incorporated.
- TDD and traditional FDD will be used. By using TDD only half the spectrum is needed.
- The 5G network is expected to be made up of small cells instead of the larger macrocells common today. Small cells including their antennas, the size of a loaf of bread, can be attached to light poles and the sides of buildings rather than require separate towers. Indoor units will also be common. Multiple small cells are needed to cover a given area because of the short range of mmwave signals.

- Massive MIMO is key feature of 5G. Using more than 64 antenna systems, the new cell sites can handle higher data rates and multiple users can be accommodated with beamforming. Beamforming implements spatial multiplexing that enables frequency reuse. Beamforming also increases signal power and permits small cells to avoid interference and establish more reliable links.
- A key feature of the 5G network is the use of software-defined network (SDN) and network function virtualization (NFV) software. These software technologies will help the carriers to manage a large number of small cells as well as to rapidly reconfigure the network and provide faster easier addition of new cells or subscribers.

Some 5G pilot evaluation systems are being tested as this is written and some prestandard networks could be in limited use in 2018. Final standards are not expected until 2019. Full build-out may not occur until 2020.

Most cell phones today are smartphones with advanced features that expand its use. There are about a dozen smartphone manufacturers, but the market leaders are Apple's iPhone and Samsung's Galaxy series. Over a billion smartphones are sold annually. There are too numerous to mention here, and they go out of date monthly or yearly. Here is a general summary of their capability.

High-speed data capability. Most smartphones use 4G LTE technology, so they can deliver as fast a data rate as your home Internet connections, usually 10–20 Mbps or better on average. Peak rates are over 100 Mbps under some favorable conditions.

Internet access. A built-in browser lets you access the Internet. The main limitation is the small screen size, which prevents you from doing as much as you would like, but you can still do it.

Email. Email is one of the key features of a smartphone. It lets you stay in touch without a PC or laptop.

Messaging. It goes without saying that text messaging is a main data feature.

Digital camera. Most smartphones have an excellent high-resolution camera built in. The camera does have about 12 megapixels of resolution or more. All high-end smartphones contain both front and back cameras.

Camcorder. All of the digital cameras have video capability as well. The main limitation is the amount of memory available to store videos.

MP3/iPod. A music player is a standard feature of many smartphones. The key here is to have sufficient flash memory to store the songs you want.

WLAN access. You will learn more about this in the next chapter, but a smartphone has a built-in separate transceiver to connect to a Wi-Fi hot spot or other access point in airports, hotels, coffee shops, and other locations just like a laptop. You can access the Internet or email as an alternative to doing it through the cellular network.

Bluetooth. Bluetooth is another short-range wireless technology covered in Chapter 9. It is used in smartphones for wireless headsets. A headset contains a headphone and a microphone that fits in the ear. It talks to the cell phone wirelessly via a Bluetooth wireless link.

Near Field Communications. NFC is a short-range wireless technology used for automatic payment, transportation, or facility access.

GPS and maps. The GPS is the worldwide satellite navigation system. All phones have a built-in GPS receiver so that you can find your way. All phones include full navigation software, display, and maps such as those in personal navigation devices (PNDs) like those from Garmin and Tom Tom.

Apps. Apps or applications are software programs that implement special functions or services. Games are a popular type but others include banking, investment, special calculators, and social media such as Facebook, Twitter, LinkedIn, Instagram, and others. Millions of apps are available and convert your smartphone into a useful computer.

Location Technology

Location technology refers to systems and circuits that are designed to pinpoint the location of a cell phone in use. This can be done now to a certain extent, as a carrier can always determine which cell site a phone is currently using. That only locates the phone within an area of a few miles. The Federal Communications Commission (FCC) has mandated that all cell phones can be located within 50–100 ft in case of an emergency. This is part of the E911 system for all cell phones. E911 means that your cell phone company has a system that will send your physical location to emergency services if you call 911 from your cell phone. Most systems use GPS, but others may have another type of location technology. The whole thing is transparent to a user, and most users actually do not know that their phone has this capability.

The location capability has many concerned for privacy reasons. But if you do not want to be located, turn off your phone. Otherwise, learn to live with it, as it may save your life 1 day. Location technology is also supposed to lead to services that are based on location, such as advertisements for restaurants and other businesses near you.

When GPS is used, the GPS receiver computes your coordinates from the satellite signals and sends them to the carrier for use if called for. Most systems can usually pinpoint you within 50 ft or so depending on terrain and environment.

Project 8.1

Examine a Base Station

Go look at a cell site. The wireless phone company will not let you inside, but you can drive close to many and examine what you can see. Especially note the antenna arrangement at the top of the tower. Most have a triangular arrangement of three antenna arrays that divide the coverage into three 120-degree wide areas. You will see multiple coax cables going up the tower to carry the received and transmitted signals. Check for one or more backhaul dish antennas that connect the site to the main cellular system.

Chapter 9

Networking: Wired and Wireless

All Devices Talking to One Another

In this Chapter:

- Types of networks defined: WAN, MAN, LAN, and PAN.
- Ethernet, Sonet, and other wired networking technologies.
- Internet transmission fundamentals.
- Wireless networking technologies, including Wi-Fi, Bluetooth, ZigBee, and others.
- Broadband technologies.
- Introduction to the Internet of Things.

INTRODUCTION

You already know from reading this book that virtually every electronic product has a digital computer built into it in the form of an embedded controller or microcomputer. But one other fact about electronic equipment that you may not know is that a large percentage of these devices are now being interconnected to one another to form networks. This is particularly true of computers. After all, what computer today is not connected to some form of network? And while we may be nearing the point where all computers are networked, the trend continues toward networking many other noncomputer devices. This chapter takes a look at networking concepts and the most popular forms of networks being used today, both wired and wireless.

IS EVERYTHING NETWORKED?

A network is just a system of people or things that are connected to one another. If someone was to ask you to give an example of a network, you probably would use the computer network definition. But there are all sorts of other networks in which we all participate in some form or another. One of the biggest networks is our interconnecting road system. Our highways and byways interconnect cities with one another and provide transportation routes for people and goods. The railroad is a massive network. You could also call the airline system a massive network of cities connected by airplane flights. Our whole electrical grid is a massive network.

Electronics Explained. http://dx.doi.org/10.1016/B978-0-12-811641-8.00009-6
Copyright © 2018 Elsevier Inc. All rights reserved.

And one of the largest networks in the world is the telephone system, both wired and wireless. And then again, let us not forget the Internet itself that connects us all to everything else. For our purposes here, our network definition is a system designed for communications via computers talking to one another or individuals talking to one another over cell phones. In all cases, a network is involved.

TYPES OF NETWORKS

There are lots of different ways to interconnect individual computers or other electronic products. You can do it by electrical cable or fiber optic cable, or you can do it in a wireless manner. In any case, each computer or user in the network is called a *node*. The goal of a network is to connect each node to all other nodes in some way so that they can communicate. Over the years, several different levels of networks have evolved.

Wide Area Networks

A wide area network (WAN) is the largest network of all, and we could consider the entire telephone network or the Internet as a large WAN. Sometimes WANs are smaller and may be localized to a country, a state, a city, or other large geographical area. The Internet is certainly a WAN. Most WANs are made up of fiber optic cable so that they can carry lots of data at very high speeds.

Metropolitan Area Networks

A metropolitan area network (MAN) is a network that covers a smaller geographical area such as a city or a large college campus system. Good examples of MANs are the local telephone company and your local cable TV company. MANs are also widely used in large companies and in governments to interconnect their computers. Most MANs are also implemented with fiber optic cable to maintain speed and data capacity.

Local Area Networks

A local area network (LAN) covers a small group of computers, typically a thousand computers or less. LANs are implemented in companies, small businesses, and even in homes. A LAN may only have a couple of users but could have several hundred, depending on the circumstances. Even two connected PCs at home comprise a LAN. In most cases, LANs are connected externally to a MAN or a WAN for external communications.

Personal Area Networks

Most personal area networks (PANs) are short-range wireless networks. A personal area network (PAN) is formed when two or more computers or cell phones interconnect to one another wirelessly over a short range, typically less than about 30 feet. Most PANs are what we call ad hoc networks that are only set up temporarily for a specific purpose. The Bluetooth headset option on a cell phone is an example of a PAN.

Home Area Networks

A home area network (HAN) is a network inside the home used to provide monitoring and control over energy usage. It connects to the utility's electric meter and monitors energy usage so the home owner can see what energy is being used and where. It is also used to give the utility a way to control the heating and air conditioning to save energy. The HAN also provides a way to conveniently turn appliances off and on. The HAN may use wireless or communications over the AC power line. The HAN is the home portion of the national Smart Grid effort to conserve energy.

Storage Area Networks

A storage area network (SAN), as its name implies, is a network that interconnects large data storage devices to computers and to one another. Large arrays of hard disks are usually set up to store massive amounts of data needed by business, the government, and other organizations. The SAN provides a way to access all this information in a fast and easy way.

Network Relationships

As it turns out, the different types of networks are connected to one another, the result being one massive network. Fig. 9.1 is an example. For example, local LANs are connected to MANs, while the MANs themselves are connected to WANs, and so on. In this figure, the circles represent PCs while the squares represent network connecting points. These include servers, the computers that manage LANs and MANs; routers that determine connections in the network; and add/drop multiplexers that put data on the network and take it off. All the interconnecting links are fiber optic cable except for the copper cable in the LANs. This complex hierarchy essentially lets any computer talk to any other computer under the right conditions. The Internet implements such a hierarchy of networks.

NETWORK INTERCONNECTION METHODS

As you can imagine, there are lots of different ways to connect computers or other electronic devices to one another. For example, you could run a cable between each of the computers you want to interconnect. The result would be a large, complex, and very messy and expensive interconnection system. It is called a *mesh*. Except for a few computers, such systems are not used. Instead,

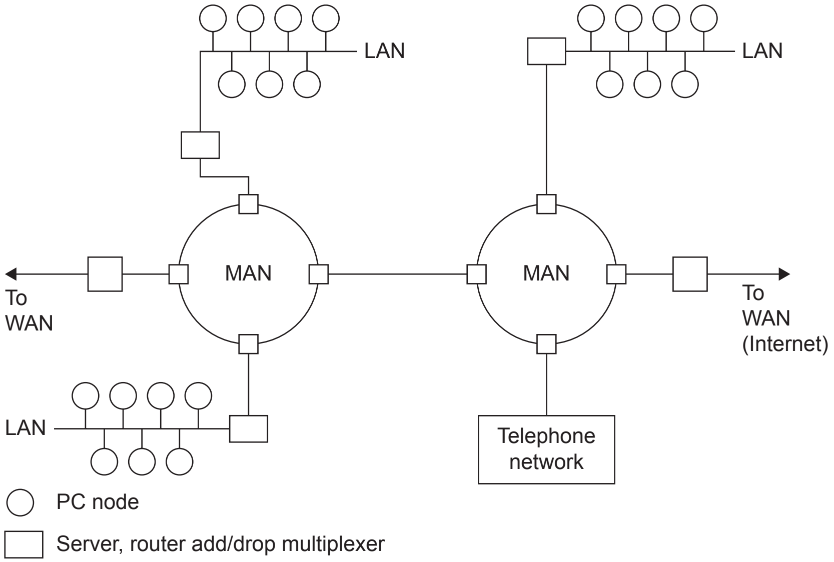

FIGURE 9.1 How LANs, MANs, and WANs connect to form a huge worldwide network of computers and other devices.

over the years, several different interconnection methods called network topologies have emerged. These are the star, the ring, and the bus, all simply illustrated in Fig. 9.2.

In the *star* method, each node is connected to a central control computer or server that manages the interconnections from any PC to any other. In the *ring* interconnection, the computers are simply connected into a single closed loop. Then data is transferred from one computer to the next by sending it around the ring. The computer destined for the information will recognize it and grab it off the ring as it comes by. The *bus* method of interconnection is one of the most popular, as it is simply a common cable or connecting point to which all computers are attached. The problem with the bus is that there can only be one sender or transmitting party of the information on the bus at a time. This requires some method of managing who transmits to whom and when. All nodes can receive.

The mesh network mentioned earlier is a messy thing to implement with cables. However, it can be implemented wirelessly. In a mesh, all nodes can talk to all other nodes within their range. This is advantageous in that there are multiple paths for data to take if one path should be blocked for some reason. More on that later.

WIRED NETWORKING TECHNOLOGIES

Wired networks use three types of cable: twisted pair, coax, or fiber optic. The telephone system has used twisted-pair cable from the very beginning, and it is

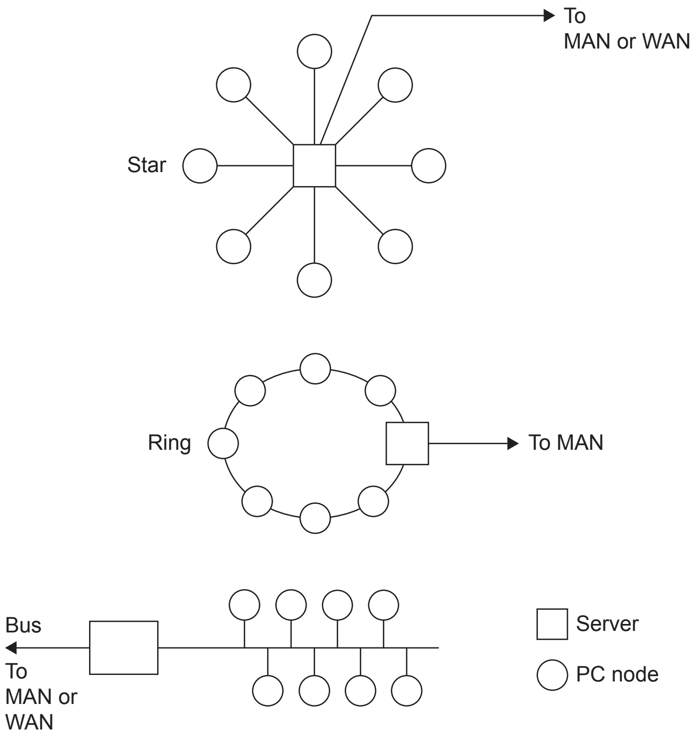

FIGURE 9.2 Three main networking topologies—star, ring, and bus. Bus is the most widely used.

still in use today. It is also the predominant way that LANs are interconnected. Coax cable was once popular but is not widely used today except in cable TV systems. Fiber optic cables were not practical until the 1970s, but since then have been developed into the best way to wire computer networks because of their high speed and low loss.

Types of Cables

Coax cable was the original networking medium (see Fig. 9.3). It has a copper wire center conductor surrounded by an insulator such as Teflon or another type of plastic that in turn is usually covered with a wire mesh braid. A plastic outer jacket covers the whole thing. The center wire and the mesh are the two connection points. Coax can handle frequencies up to many gigahertz. But it has high losses over long distances. Coax is mainly used in cable TV networks.

Twisted pair is the widely used cable for telephones and LANs. It consists of two solid insulated wires twisted together as shown in Fig. 9.4. Most telephone or LAN cables contain multiple pairs. Typical LAN cables have four pairs in a single cable. These cables are standardized and categorized as shown in

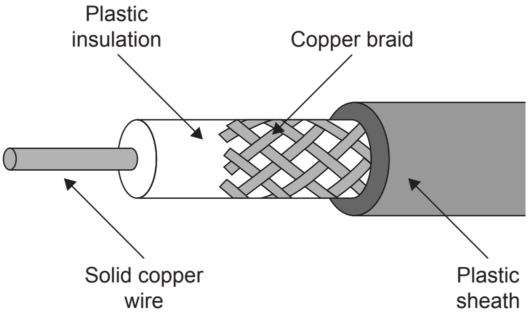
Plastic insulation
Copper braid
Solid copper wire
Plastic sheath

FIGURE 9.3 Coax cable. It is now mainly used in cable TV networks. The most common is RG-6/U and uses F-type or BNC connectors.

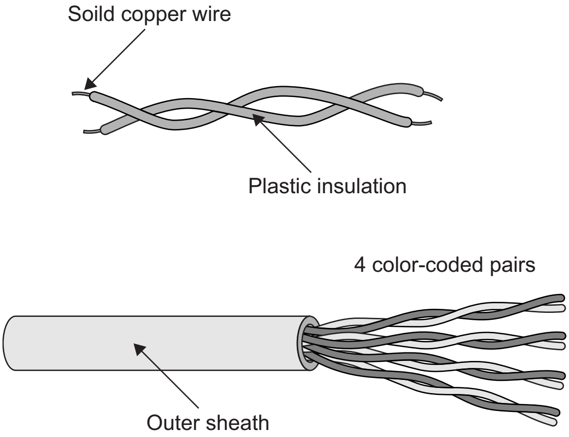
Soild copper wire
Plastic insulation
4 color-coded pairs
Outer sheath

FIGURE 9.4 Twisted-pair cable. Initially used for telephones and is still used. Computer network cables use four twisted pairs. A shielded version is available where noise is a problem.

Table 9.1. CAT5 and CAT5e are the most common. The eight wires are usually terminated in a connector called an RJ-45.

Fiber optic cable consists of a thin glass or plastic center piece surrounded by a plastic outer cover. Then that is further wrapped by an outer plastic covering (see Fig. 9.5). Light waves are passed down the center glass or plastic cable. Most light comes from an infrared (IR) light laser. You cannot see IR but it is light nonetheless. IR light is what your TV remote control uses. The IR light is pulsed off and on to create binary signals. These are picked up at the receiving end by a light detector and converted back into binary 0 and 1 voltages (see Fig. 9.6).

Over the years, dozens of different wired networks have been developed. But in the past 2 decades, only two types of wired networks are dominant: Ethernet and synchronous optical network (Sonet).

TABLE 9.1 Types of Twisted-Pair Cable

Category	Maximum Data Rate	Application/Comment
CAT5	100 Mbps	Ethernet LAN, telephones
CAT5e	1 Gbps	Ethernet LAN
CAT6	10 Gbps	Ethernet LAN, video
CAT7	10 Gbps	LAN, point to point

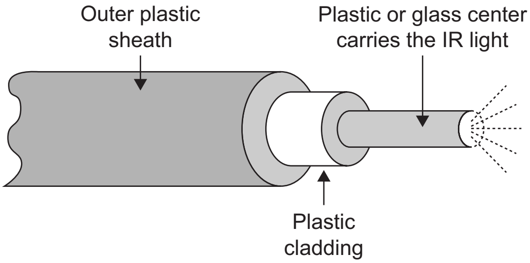

FIGURE 9.5 Fiber optic cable. It is nonconductive because it is made of plastic and glass, and is used to transmit data as infrared (IR) light pulses.

Ethernet

Ethernet was originally developed as a LAN in the 1970s. It used coax cable in a bus topology to interconnect multiple computers. Ethernet is also known by its Institute of Electrical and Electronic Engineers (IEEE) standard designation 802.3. Today, Ethernet is the dominant networking technology for LANs and is often used in MANs and WANs.

There are many variations of this technology, but the most common one uses unshielded twisted-pair cable. The bus topology is generally implemented inside a piece of equipment called a *hub* or *switch* (see Fig. 9.7). The cable from each computer or other device is connected to one of the jacks on the hub or switch. The hub or switch connects to a master computer called a *server* that manages the network. Generally speaking, connections to the hub or switch are limited in length to about 100 m or 300 feet.

The initial Ethernet LAN had a maximum data rate of 2.93 Mbps over coax. When Ethernet was standardized, the basic data rate was 10 Mbps. Over the years, Ethernet has been improved and developed to create versions with data speeds over twisted-pair cable of 10 Mbps, 100 Mbps, 1 Gbps, and 10 Gbps. Two new versions operate at 2 Mbps or 5 Mbps. While 10 Mbps can be implemented over a single twisted pair, to achieve speeds of 100 Mbps, 1 Gbps, and 10 Gbps, all four twisted pairs in the CAT5/6 cable are used. The data is divided between and transmitted in parallel at the same time using special multilevel encoding that permits such higher data rates to be achieved.

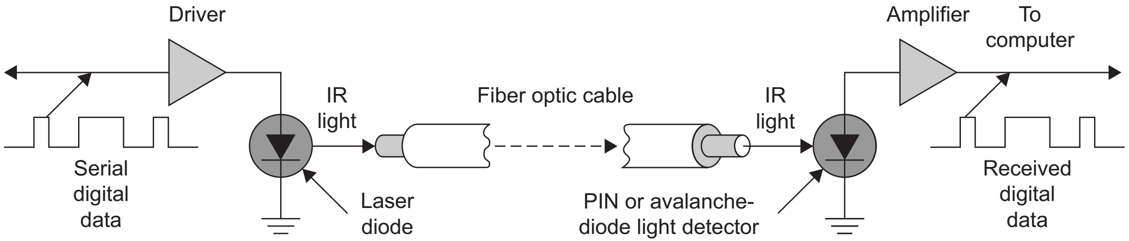
Driver
Amplifier
To computer
IR light
Fiber optic cable
IR light
Serial digital data
Laser diode
PIN or avalanche-diode light detector
Received digital data

FIGURE 9.6 How binary data is transmitted over fiber optic cable.

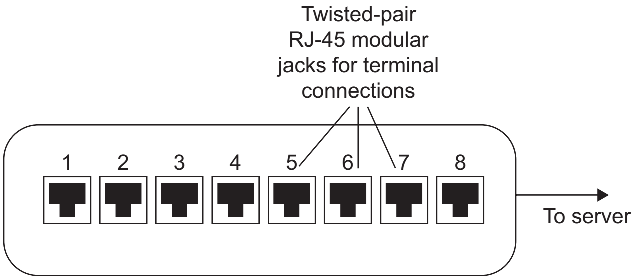

FIGURE 9.7 Ethernet hub or switch implements a bus that is accessed by RJ-45 connectors and twisted-pair cable.

Preamble	Start of frame delimiter	Destination address	Source address	Length	Data	Frame check sequence
7 bytes	1 byte	6 bytes	6 bytes	2 bytes	46–1500 bytes	4 bytes

FIGURE 9.8 Ethernet frame or packet.

There are also versions of Ethernet that use fiber optic cable. Fiber optic cable is normally used on the 1-Gbps and 10-Gbps versions. The latest versions of Ethernet provide 40-Gbps, 100-Gbps, and 400 Gbps data rates. A one terabit per second (1 Tbps) version is being developed.

How Is Ethernet Used?

Ethernet uses a bus topology connection where only one node can transmit at a time. This requires the use of an access method that permits the nodes on the bus to share it. Ethernet uses a technique known as *carrier sense multiple access with collision detection* (CSMA/CD). This is part of the IEEE 802.3 standard. Essentially what happens is that if a computer wants to transmit, it monitors the bus, and if no one is transmitting, it will begin transmitting its signal. If someone else tries to transmit at the same time, a collision occurs and both computers stop transmitting. Each will wait a random period of time and then try again. Whoever captures the bus first will be allowed to transmit. The data is transmitted over the bus, and any computer connected to it can receive it. However, typically the transmission is only intended for one other computer. To identify that computer, each is given an Ethernet address that is transmitted along with the data to be sent.

The data to be transmitted are packaged together in a group of bits called a packet. Fig. 9.8 shows what a typical Ethernet packet looks like. The preamble consists of 7 bytes of alternating 1's and 0's that help the receiving computer establish clock synchronization with the transmitting computer. The start frame delimiter presents a unique 1-byte code indicating that a frame of Ethernet data is to be transmitted. Next, a 6-byte destination address identifies the desired receiving node. The next 6 bytes identify the sending node. This is followed by another 2 bytes of data that indicate how long the data message is. The block of data is then transmitted. It can be any length from about 46 to 1500 bytes. The packet ends with a frame check sequence that is essentially a unique code designed for error detection and correction. If bit errors occur during transmission, the receiving computer can perform specific logic functions that allow the error to be detected and then corrected.

While Ethernet started out as a LAN, it has morphed into other larger forms of networks. Because of its high-speed capability with fiber optic cable, it is also used to form MANs and even WANs. A special version for MANs and WANs is called *Carrier Ethernet*. There is also a version of Ethernet used in SANs, called *Internet Small Systems Computer Interface* or *iSCSI*.

More recently Ethernet is now the networking technology of choice in industry. Factories and plants now use Ethernet to connect computers, machine tools, robots, motor drives, and other manufacturing equipment. Ethernet is also being adopted for use in cars and trucks. It interconnects all of the many embedded controllers that operate the various subsystems such as emissions control, braking and stability, and safety. Another automotive use of Ethernet is in connecting infotainment equipment such as the radio, audio, video, and navigation.

Synchronous Optical Network

Sonet means *synchronous optical network*. Also known by its international standard name, *synchronous digital hierarchy*, this fiber optic network system was designed specifically for high-speed data transmission in telephone networks and computer systems. Basically, Sonet defines how data is transmitted over the fiber optic cables. This includes the data format as well as data speeds. Table 9.2 shows the various standardized transmission speeds from 51.84 to 39.812 Gbps. The bulk of Sonet installations, which are used primarily in the telephone networks and the Internet, use 2.488 Gbps (also generally called 2.5 Gbps) and 9.953 Gbps (generally referred to as 10-Gbps Sonet). More

TABLE 9.2 Sonet/SDH Designations and Data Rates

Sonet Level	Data Rate
OC-1	51.84 Mbps
OC-3	155.52 Mbps
OC-12	622.08 Mbps
OC-48	2.488 Gbps
OC-192	9.953 Gbps
OC-768	39.812 Gbps

recently, new 39.812 (40) Gbps networks are being built for the Internet. Data is transmitted in fixed-length frames made up of 810 bytes, most of which is data. The network topology is either point to point or in a ring.

Most big WANs and MANs use the Sonet system. However, Sonet is often used to transmit other data, including the telephone company T1 and T3 line data, asynchronous transfer mode (ATM), frame relay, and other telephone company formats. It is also widely used to carry Ethernet packets as well as the unique packets that are part of the Internet protocol. Today, Sonet equipment is gradually being phased out in favor of the optical transport network (OTN).

Telephone Data Systems

When the wired telephone system converted to digital many years ago, several key data transmission standards were developed. The most common are the T carrier systems, as well as frame relay and ATM. The most widely used digital telephone networking standard is the T-1 system. It was developed to transmit 24 digitized telephone lines over a single cable. The voice signals were digitized into 8-bit chunks and then transmitted sequentially over a T-1 line from one central office to another. The basic T-1 frame format is shown in Fig. 9.9. It consists of 192 data bits made up of 24 bytes representing individual chunks of telephone calls. An additional synchronizing bit is added. The result is a T-1 signal that transmits at a rate of 1.544 Mbps. And while T-1 is still widely used for digital telephone calls, it is also used for broadband computer data connections and cell phone backhaul.

There are other T carrier systems that are used to multiplex T-1 lines onto larger cables. The most common is the T-3, which allows dozens of T-1 lines to be multiplexed on a single cable at a 44.736-Mbps data rate.

Frame relay transmits data mostly in telephone systems. Both the T carrier and frame relay are synchronous systems, meaning that they are synchronized to a master clock signal.

Asynchronous transfer mode, as its name implies, is an asynchronous system that transmits data in packets. Each packet is made up of 53 bytes—5 bytes for a header that designates the destination and a 48-byte payload segment. Packets may be digital voice or computer data.

All these systems still exist to some degree but they are gradually being replaced by Ethernet and the OTN.

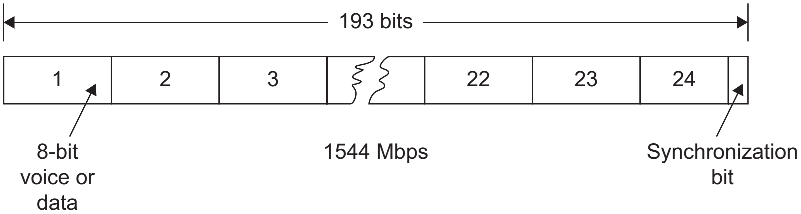

FIGURE 9.9 A T-1 data frame commonly used in digital telephone systems as well as data interconnections in networks.

Optical Transport Network

The OTN is a technology used to implement the Internet backbone network. This is the core long haul fiber optical network that connects the world together. OTN uses special polarization and modulation techniques to transmit data rates of 40, 100, and 400 Gbps.

HOW THE INTERNET WORKS

The Internet is an enormous WAN covering the entire world. It consists of many superfast, fiber optic cable links that are generally referred to as the backbone of the network. These high-speed links are owned by companies and governments and provide connection points for companies, governments, WANs, Internet service providers, and hosts for websites. These connection points are known as network access points (NAPs). See Fig. 9.10. There are dozens of NAPs in the United States, and they provide the link to the backbone connections. These backbone connections are worldwide. Because they are complex and interlinked, they are often referred to as the Internet cloud. Since data being transmitted from one point to the other can go by multiple paths in the backbone, there is no way to know the exact path, so the cloud designation is appropriate. You put the data transmission into one NAP entry point to the cloud, and it

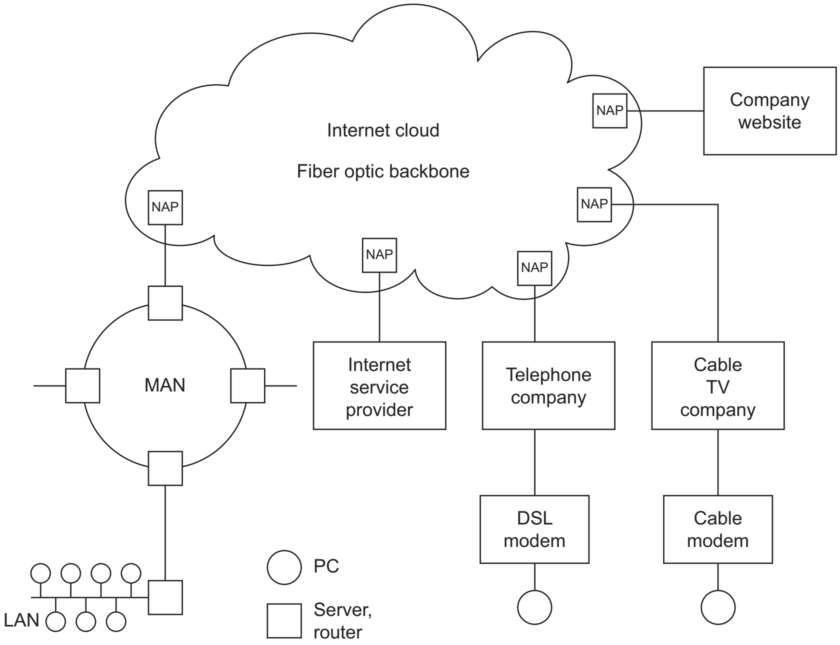
Internet cloud
Fiber optic backbone
NAP
NAP
NAP
NAP
NAP
Company website
MAN
Internet service provider
Telephone company
Cable TV company
DSL modem
Cable modem
LAN
PC
Server, router

FIGURE 9.10 Internet illustrated as cloud of fiber backbone connections through network access points (NAPs) that attach to WANs, MANs, and LANs.

comes out in another. From there it goes to other MANs and LANs for connections to other networks and to individual users.

Data is transmitted over the Internet using a standardized method referred to as TCP/IP. This means transmission control protocol/Internet protocol. Both are protocols for building packets that are transmitted over the Internet.

Assume that a LAN and MAN using Ethernet want to transmit information over the Internet. The Ethernet packets are then sent to a NAP where they are assembled into the packets that actually get transmitted over the fiber optic cable. The TCP assembles the data to be transmitted into unique packets and attaches a header to it, giving the source and destination addresses, a sequence number, and some error detection and correction information. The general format is shown in Fig. 9.11A.

The TCP packets then become the data that is assembled further into Internet packets (see Fig. 9.11B). The Internet protocol attaches a header giving source and destination addresses as well as other information about where and how the data is to be sent.

The IP header contains a 32-bit destination address identifying the location of the computer receiving the data. This 32-bit address is divided into four 8-bit segments. Each 8-bit segment is usually referred to by its binary equivalent. The result is the IP address. Every computer and most other devices have an IP address assigned by an organization set up for assigning and managing Internet addresses. You may have seen the address referred to in the format 51.78.23.189. This is called the *dotted decimal format*.

With a 32-bit address, over 4 billion individual computers and other devices can be assigned a unique address. However, we are already beginning to run out of addresses, and a new updated form of Internet protocol called IPv6 (Internet Protocol version 6) is gradually replacing it. It has a 128-bit address that provides billions and trillions of additional addresses so that even the smallest inconsequential device can have its own Internet address.

As the data is transmitted over the Internet, it passes through many devices referred to as *routers*. The router is the basic transmission control device of the Internet. What the router does is to examine all of the packets transmitted to it and to specifically look at their IP destination addresses. The router temporarily will store data being transmitted and then decide which part of the network it should be connected next to reach its desired destination. Routers store sets of information called *routing tables* that help the router decide the best transmission path through the Internet.

The data being transmitted ultimately reaches the NAP connected to the network where the destination exists. This may be a company, MAN, or LAN, or it could be an Internet service provider, such as the telephone company or cable TV company that will ultimately connect to the source computer. The source computer then strips away the TCP/IP headers to recover the original Sonet or Ethernet data.

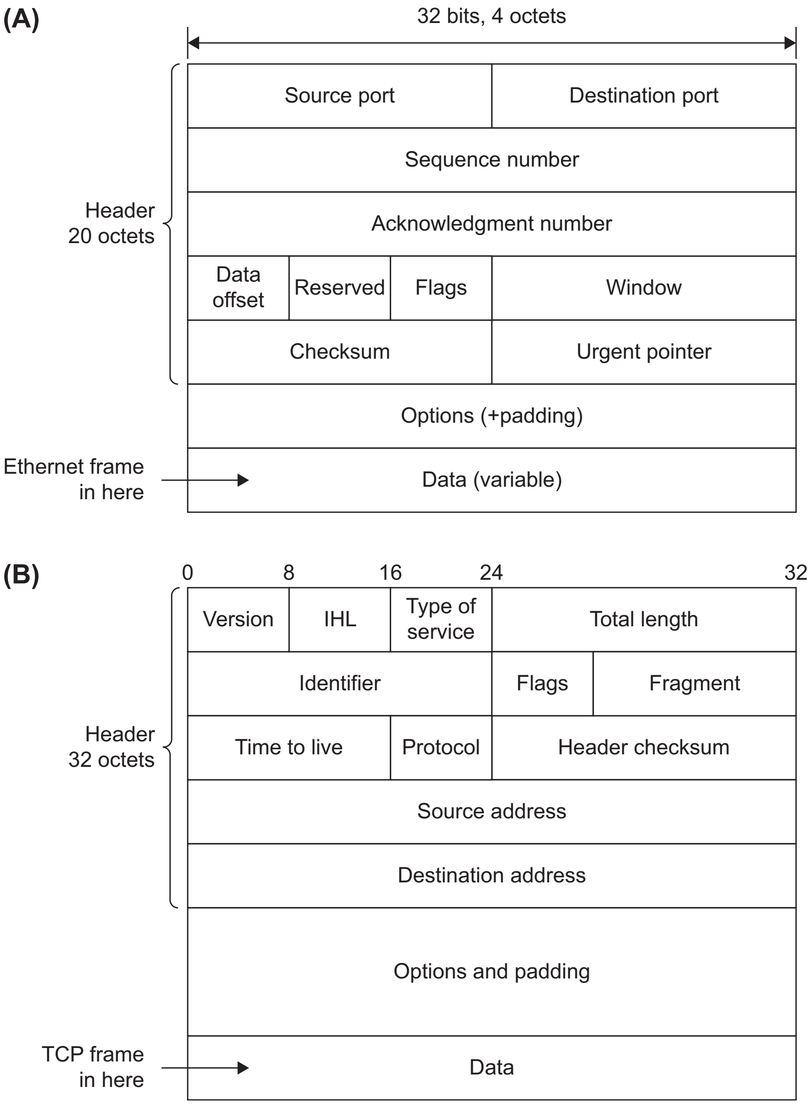
(A)
32 bits, 4 octets
Source port
Destination port
Sequence number
Acknowledgment number
Data offset
Reserved
Flags
Window
Checksum
Urgent pointer
Options (+padding)
Data (variable)
Header 20 octets
Ethernet frame in here
(B)
0
8
16
24
32
Version
IHL
Type of service
Total length
Identifier
Flags
Fragment
Time to live
Protocol
Header checksum
Source address
Destination address
Options and padding
Data
Header 32 octets
TCP frame in here

FIGURE 9.11 Protocols used in Internet. (A) Transmission control protocol (TCP). (B) Internet protocol (IP). The data in each frame are transmitted from left to right and top to bottom.

WIRELESS NETWORKS

Wireless networks are usually LANs or PANs, but may also be MANs. Most wireless networks are short-range wireless connections that bring greater mobility as well as freedom from interconnecting cables to networking. Fig. 9.12

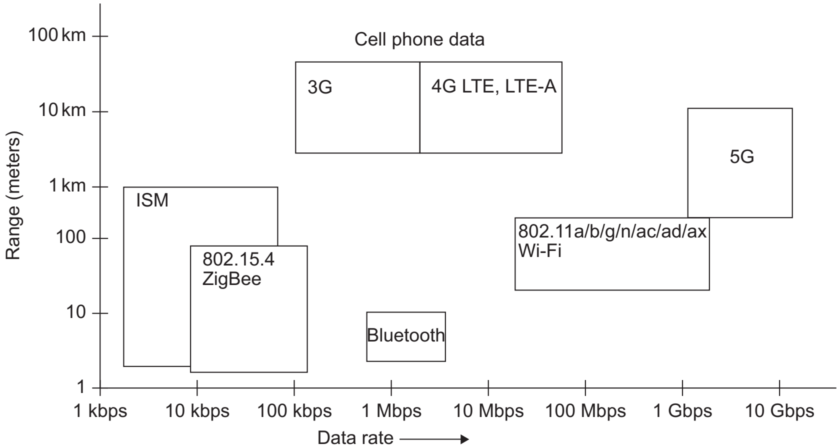

FIGURE 9.12 Summary of short-range wireless networking technologies shown as data rate versus range. See Fig. 9.12.

shows a graph designating the various types of wireless networking schemes in use today. In this graph, each networking technology is defined in terms of its data rate as well as its range. In this section, we take a look at all of these technologies, most of which you have heard of or already in use on a regular basis. The concepts of the Internet of Things (IoT) and machine-to-machine (M2M) communications are also introduced here.

Wi-Fi

Wi-Fi is one of the most widely used wireless networking technologies. Its trade name Wi-Fi, means *wireless fidelity*. This the name given to the IEEE's wireless Ethernet standard designated as 802.11. Wi-Fi was originally developed as a wireless extension to normal Ethernet LANs. It allowed laptops and personal computers and even cell phones to connect to a company or organizational LAN wirelessly instead of the normal CAT5 connection. This not only simplified and in many case eliminated new wiring but also gave employees greater freedom and mobility, particularly since most employees use laptop computers rather than fixed PCs. Today, Wi-Fi is the primary wireless link for tablets and smartphones to connect to the Internet.

Fig. 9.13 shows how Wi-Fi works. Each PC, laptop, or smartphone has a built-in data transceiver and antenna that communicate with a wireless access point (AP). This wireless AP is a transceiver that connects directly to an Ethernet switch, which, in turn, connects to the LAN wiring and its server. APs are placed at strategic locations within buildings to allow as many people as possible to connect wirelessly.

There are also public APs known as *hot spots*. These are present in coffee shops, airports, hotels, and in many other locations, allowing anyone with a

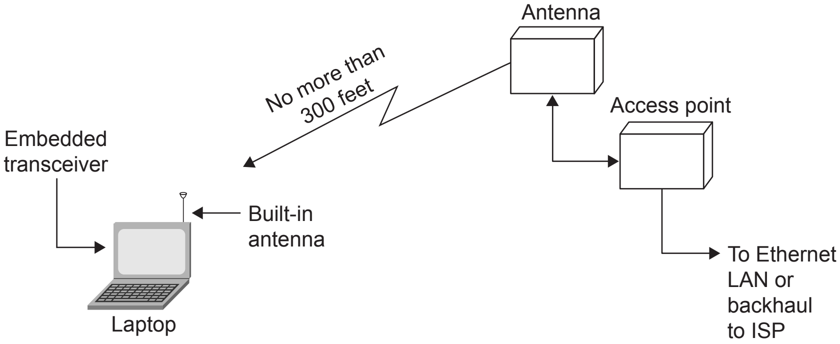

FIGURE 9.13 How a Wi-Fi access point or hot spot works.

laptop or cell phone to connect. These hot spots talk back to a central point by a backhaul method, such as a T1 telephone line or microwave relay link that connects to an Internet service provider that, in turn, is connected to the Internet.

The 802.11 standard exists in several basic forms. The original version of this standard is 802.11b. It operates in the 2.4- to 2.483-GHz unlicensed band. It uses the direct-sequence spread spectrum (DSSS) and is capable of data rates up to 11 Mbps in a range of up to 100 m. This frequency band is divided into 11 channels, each 22 MHz wide. The computer and AP decide on which channel to use to minimize interference with other transmissions. If you are close enough to the AP, you can actually get the 11-Mbps data rate, but if you are not, the system automatically ratchets down to 5.5, 2, or even 1 bps depending on the range, interference, and noise or other environmental conditions.

Another popular version of the standard is 802.11g. It is a faster version of Wi-Fi that also operates in the 2.4-GHz band. It can achieve data rates as high as 54 Mbps up to 100 m. It uses OFDM with 52 subcarriers.

Another variant of the standard is 802.11n, which uses OFDM and MIMO to achieve even higher data rates up to approximately 600 Mbps. Another less popular version is 802.11a, which uses the 5-GHz unlicensed band. It too uses OFDM and can achieve data rates to 54 Mbps. Even newer and faster versions are available. 802.11ac uses the 5 GHz band and can produce data speeds up to several Gbps under the most favorable conditions. It uses wider 80 and 160 MHz bands and higher-level 256-QAM modulation as well as increased MIMO configurations.

The 802.11ax version works in the 2.4 and 5 GHz bands and is even faster. It uses the wider bands of 802.11ac and 1024-QAM to produce even high rates to 7 Gbps. 802.11ax also uses multiuser MIMO to boost coverage range and data rate.

Faster data speed but over a shorter distance can be achieved with the 802.11ad, also known as WiGig. It operates in the unlicensed millimeter wave band from 57 to 64 GHz. The range is limited to about 10 m but a data rate to

7 Gbps is possible. Its most common use is video transport in consumer electronics gear.

Wi-Fi is also built into most of the new smartphones. While a smartphone has data capability because of its 4G LTE digital technology, Wi-Fi gives you one more way to access the Internet and email.

Wi-Fi is widely used throughout business and industry. It is also the most popular home networking technology. With no wires to install, it is easy to set up and use with a broadband DSL connection or cable TV access. Chances are you already have Wi-Fi in multiple forms.

Bluetooth

Bluetooth was developed as a PAN for connecting computers, cell phones, and other devices up to a range of about 30 feet. It is a low-speed data transmission method. Bluetooth operates in the same 2.4- to 2.483-GHz unlicensed spectrum as Wi-Fi. It uses a technique known as frequency-hopping spread spectrum (FHSS), where the data is divided into chunks and transmitted via a carrier that hops from one random frequency to another. Data is transmitted at a 1-Mbps rate using FSK. An enhanced data rate form of Bluetooth is also available to transmit at higher speeds up to 3 Mbps. The range is 10–50 m depending on the environment. The standard is managed by the Bluetooth Special Interest Group.

More recent versions of Bluetooth use a different form of FHSS and are designed to operate on less power. Called Bluetooth Low Energy (BLE) it is available in several forms for data rates of 1 or 2 Mbps. BLE nodes can operate for years from a single button cell because of the very low current drain. A newer version of BLE called Bluetooth 5 uses different modulation and coding schemes to achieve data rates to 2 Mbps over a longer range up to 50 m or more.

One of the basic features of Bluetooth is that it is capable of forming small networks called piconets. It does this by linking two Bluetooth devices together. One serves as a master controller, and it can connect to seven other Bluetooth slave devices. Once the PAN has been set up, the various connected devices can exchange information with one another through the master.

By far the most common application for Bluetooth is cordless headsets for smartphones. Other common uses are Bluetooth speakers for smartphones or tablets as well as hands-free systems in automobiles. But you will also encounter it in some wireless connections between a PC and a mouse or keyboard. Bluetooth may be the most widely used wireless standard in the world because it is used in so many smartphones, laptops, and other consumer equipment. Billions of Bluetooth chips have been sold.

ZigBee

ZigBee is another short-range PAN network technology with the IEEE designation 802.15.4. It uses low power, so the range is typically 100 m or less, basically depending on the antennas and the physical environment where it is used. ZigBee also operates in that unlicensed spectrum from 2.4 to 2.483 GHz.

The IEEE 802.15.4 standard defines the basic radio technology. It usesDSSS and a version of QPSK that gives a data rate of 250 Kbps. The frequency spectrum is divided into 16 1-MHz channels.

The basic function for ZigBee is monitoring and control. Monitoring refers to telemetry that allows sensors to collect information, which is then transmitted by radio to some central collection point. It also allows for remote control of the devices such as lights, motors, and other items. The ZigBee Alliance, the organization that promotes and enhances this wireless technology, has developed a series of applications for industrial monitoring and control, home monitoring and control, energy monitoring (automatic meter reading) and control, and several others.

A major feature of the ZigBee technology is its ability to perform mesh networks. While a typical ZigBee radio node may only have a range of 30 m or less, the range can be extended by simply transmitting data from one node to another in a mesh network that may cover hundreds of meters, or in some cases even miles. A mesh is shown in Fig. 9.14. Each circuit is a ZigBee transceiver node, and each line is a wireless path. Mesh networks have many different paths for data and, therefore, are extremely reliable. If one path is blocked or disabled, data can usually find its way to the desired destination by way of another path. For example, in Fig. 9.14, a direct path may be from A-D-F-I-K, but if the D-to-F link is blocked or the I node is disabled, the mesh would automatically reroute the data through another path such as A-D-G-J-K.

Another application of the 802.15.4 standard is the replacement of the some IR TV remote controls. A radio frequency (RF) version called RF4CE is gradually replacing IR remotes in consumer electronic equipment providing longer range and no line of sight limitations.

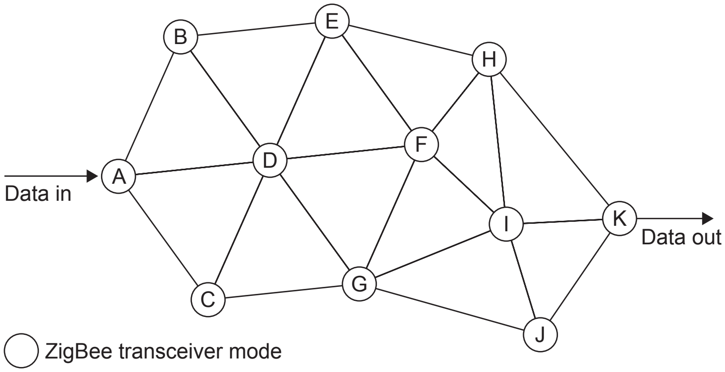

FIGURE 9.14 A mesh network can extend the overall range of a node and provide greater reliability through multiple possible paths. Most wireless mesh networks are ZigBee, but other wireless technologies may be used (e.g., Wi-Fi).

ISM-Band Radios

The industrial–scientific–medical (ISM) band has been mentioned before. This comprises a group of frequencies or bands set aside for unlicensed data transmission over short ranges. The main application is telemetry monitoring and remote control. The US frequencies are 13.56, 315, 433, and 902–928 MHz, with 915 MHz the most common. In Europe, 868 MHz is an unlicensed frequency. Transmitters and receivers are tiny single ICs or modules with power levels from about 1 mW to 1 W. Modulation is usually ASK/OOK or FSK. Data rates are low from a few kbps to usually no more than 100 kbps.

Some typical applications are garage door openers, remote keyless entry for vehicles, remote thermometers, medical telemetry, remote-controlled toy cars and boats, and sensor reading in industrial applications. Range is no more than a hundred feet or so, but longer ranges to several miles can be achieved with higher antennas.

Radio Frequency Identification and Near-Field Communications

Radio frequency identification (RFID) is a very short-range technology that is intended to substitute for standard bar codes. Practically everything has a printed bar code on it today so that with optical scanners the device can be recognized by a particular part number or other information that identifies it. The bar code readers are normally attached to a computer and, in turn, a network.

RFID devices are small thin plastic tags containing a radio chip that can be read by a wireless reader. The RFID chip contains an EEPROM memory in which is stored the device's unique identification number and other related information. A key characteristic of the RFID tag is that it contains no power source such as a battery. Yet it is designed to transmit its digital code to some external reading device whenever it is interrogated. Fig. 9.15 shows how it works.

The interrogation takes place when an external reader made up of a transmitter (TX) sends an RF signal to the nearby tag. The RF energy received by the tag's antenna is then rectified and filtered into a DC voltage that is used to power up the memory and the modulator circuit that sends data back to the reader. The reader also contains a receiver (RX) that picks up the small signal from the tag and sends it on to a computer where identification takes place.

RFID tags are generally used for what is known as asset tracking. RFID tags are attached to any relatively expensive item whose position is to be monitored, tracked, or otherwise controlled. It is good for keeping track of capital equipment, baggage, manufacturing tools, and other items. It is also used for personnel security on badges that employees wear and is good for animal tracking. One of the most widely used applications is automatic toll collection on highways.

RFID tags come in a variety of sizes, shapes, and technologies. Some of the older tags work at a frequency of 125 kHz or 13.56 MHz. The newer tags work

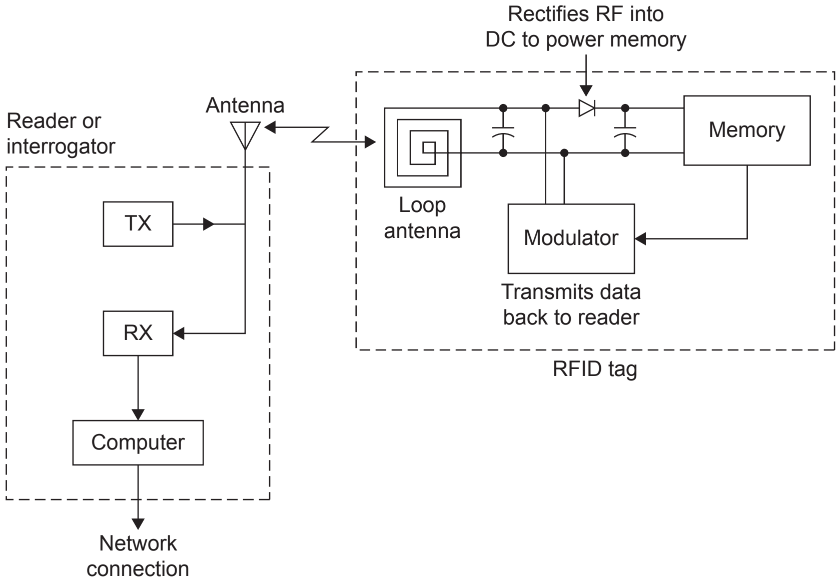
Rectifies RF into
DC to power memory
Antenna
Reader or
interrogator
TX
RX
Computer
Network
connection
Loop
antenna
Memory
Modulator
Transmits data
back to reader
RFID tag

FIGURE 9.15 How RFID works. The tag has no power of its own but gets power from the transmitter in the reader.

at UHF frequencies in the 902- to 928-MHz range. All of these tags are passive devices with no power other than that received from the reader. Some longer-range RFID tags with small batteries have also been developed.

The unique code stored in the EEPROM is stored during the tag's manufacture. It is then clocked out at a relatively slow rate of speed around 70 kbps. The serial binary data is used in a form of amplitude modulation called backscatter modulation that produces minor variations in the reader's signal amplitude that can be converted back into a binary code.

Another similar RFID-like technology is known as near-field communications (NFC). It has an even shorter range of only about 1 foot. It operates on 13.56 MHz. It is mainly designed to be used in credit cards and in smartphones to produce automatic transactions such as credit card purchases, payment for transportation (e.g., train, subway, bus), or access to some facility.

BROADBAND TECHNOLOGY

Whenever you hear the term "broadband," it generally refers to a high-speed interconnection or Internet access. Most broadband systems provide data rates from roughly 1 to 100 Mbps, depending on the level of service purchased. Such broadband services are usually provided by cable TV companies or telephone carriers. Other organizations also offer various other types of Internet connectivity.

Cable TV Connectivity

Most cable TV companies offer high-speed Internet connectivity services. What they do is allocate one or more of the 6-MHz wide channels on the cable to this service (see Chapter 11 on TV). By using QPSK, 16-QAM, or 64-QAM, a wide range of customers can be offered multiple levels of Internet connectivity service. The customer gets a cable TV modem that connects to the cable along with the TV set. This modem provides the connectivity to a computer via Ethernet. A Wi-Fi AP is often connected to the cable modem to provide wireless connections to the Internet for PCs and laptops.

Digital Subscriber Lines

Digital subscriber lines (DSLs) are high-speed Internet connections that use standard telephone lines. DSL uses the standard twisted-pair telephone lines that come into every home for normal telephone service. These lines, because they were developed only for voice signals, are restricted in bandwidth and data rate. However, special techniques have been developed to allow very high-speed data transmissions on them. DSLs, also referred to as *asymmetric digital subscriber lines* (ADSLs), use a variation of OFDM called *discrete multitone*. What it does is to divide the restricted bandwidth of the twisted-pair cable into multiple OFDM channels, each 4 kHz wide. Then the data to be transmitted is divided into parallel paths and modulated using some form of QAM. All of this is handled by a DSL modem connected to subscribers' computers and the home telephone lines.

The speed potential of an ADSL line depends on how far away the subscriber is from the central office. The greater the distance, the lower the data rate. For even the longest runs from 12,000 to 18,000 feet, data rates of up to about 2 Mbps are possible.

Newer versions of ADSL have also been developed to permit data rates of up to about 12 Mbps at a range of 8000 feet and 20 Mbps at a range of about 4000 feet. Another version referred to as video digital subscriber lines or VDSLs extends the bandwidth further and uses higher-level versions of QAM to get data rates of up to 52 Mbps. A later version called VDSL2 can achieve up to 200 Mb/s over less than 2000 feet. The overall deliverable rate depends on the distance from the subscriber's modem to the neighborhood AP called a digital subscriber line access multiplexer (DSLAM). The Internet service provider (ISP) usually runs a fiber optic cable to the DSLAM and then connects via the existing telephone cable to the homes.

The newest form of DSL is called G.fast. Its main feature is that it can deliver combined download/upload speeds of up to 1 Gb/s over existing telephone wiring. G.fast achieves higher speeds by locating the distribution point unit closer to the home. The 1 Gb/s rate is usually achieved at a range less than about 70 m. Cable length is generally restricted to a maximum of 250 m. G.fast uses time division duplexing rather than frequency division duplexing of most DSL versions. For that reason, the total data rate of 1 Gbps is the combined rates of download and upload speeds. An example is a mix of 800 Mb/s down and 200 Mb/s up.

Wireless Broadband

Wireless broadband services provide Internet connectivity via some wireless technology. Such services are provided by operating companies that service small towns and rural areas not served by standard telecom or cable TV companies. Such services usually provide fixed wireless data services with data rates in the 1–5 Mbps range. Wireless modems are connected to consumer computers that access nearby base stations within several miles.

The beauty of wireless broadband is that no cables or other hard connectivity is required. These are easy to set up, although at microwave frequencies, range is a problem as are multipath problems with buildings, trees, and other obstructions. However, with high enough antennas and sufficient power, wireless broadband has proved to be practical. Wireless broadband is growing rapidly, particularly in those areas where DSL and cable TV services are not currently available. An early application of the forthcoming fifth-generation (5G) cellular service is expected to be broadband access. Do not forget that wireless Internet connectivity can also be achieved by using the regular cell phone networks. You can purchase a data modem that plugs into your laptop or PC USB connector. This modem connects you just as it would a cell phone to your cellular network. You can also use your smartphone as a Wi-Fi hot spot. From there you can use your normal 4G LTE service to connect to the Internet, access email, or perform other networking operations. One of the first applications of the 5G NR technology is fast wireless broadband service.

THE INTERNET OF THINGS

The IoT is a movement that proposes to interconnect electrical or mechanical “things” to the Internet for the purpose of monitoring and/or controlling them. IoT, sometimes called the Internet of Everything, indicates how devices, machines, appliances, and other apparatus can communicate with humans or with one another to achieve some benefit or convenience. The IoT movement is well under way with billions of things already connected. Some sources predict that as many as 50 billion or more devices will be connected in the coming years.

There are two major segments of IoT, industrial and consumer. Industrial applications are particularly useful as it provides a way to monitor machines in a factory or systems in a process control plant. Sensors are used to keep track of environmental and physical conditions. Sensor data such as temperature, light level, pressure, mechanical position, and other factors are collected and sent via the Internet to a server that collects the data. Software then analyzes the data to extract meaningful knowledge so that sound decisions can be made about machine maintenance or production efficiency. In addition to monitoring conditions, the IoT can be used to control things such as motors, lighting, or machine tools.

Other uses of industrial IoT include the smart grid that monitors home and business electrical usage by way of wireless electric meters. Monitoring and control of city street lighting is another. An older variation of IoT called M2M (machine-to-machine) uses wireless nodes to monitor fleet trucks or cars and their location by Wi-Fi or cellular.

Consumer applications of IoT also implement monitoring and control functions but in the home or automobile. Today, applications are everywhere. Smart fitness watches monitor things such as heart rate, steps taken, or other physical attributes of the wearer. These are usually transmitted by Bluetooth to a smartphone and perhaps then on to the Internet for storage and analysis by some computer. The result is a record of a person's training and physical conditioning.

Another consumer application is a video camera monitoring a home for security purposes. The camera transmits its observations by Wi-Fi to the Internet and then to a server that connects to a smartphone with the appropriate app where the video is viewed. Other examples are the monitoring and setting of a home thermostat, lighting, or security system over the Internet via a smartphone app. Voice control of products is one more use. Voice recognition software packages such as Apple's Siri, Amazon's Alexa, Google's Assistant, or Microsoft's Cortana are being incorporated into a wide range of Internet-connected products that seek information, answer questions, play music, or initiate some action. The possibilities are endless, it seems, and there are many more to come.

How Internet of Things Works

Fig. 9.16 shows the basic IoT system. In industry, sensors are used to monitor all manner of physical conditions. The sensor is part of a wireless network node that wirelessly transmits its data to a nearby gateway. ZigBee is one widely used wireless technology in industry. The gateway receives the ZigBee data and translates into packets suitable for transmission over the Internet. The gateway connects to the Internet cloud by way of an ISP. The collected data goes to a remote server where it is then routed to a storage system and/or a computer where it is analyzed and consumed by some user.

The approach in consumer IoT is similar. Items to be monitored or controlled contain a wireless sensor or actuator node. Wi-Fi is often used although Bluetooth is also popular. The nodes connect to a consumer gateway that interfaces to the Internet by way of a cable, DSL, or fiber ISP. In many homes the ISP Wi-Fi router serves as the gateway. As with industrial IoT, the data eventually goes to a server where it is the routed to some appropriate destination where data storage or analysis takes place.

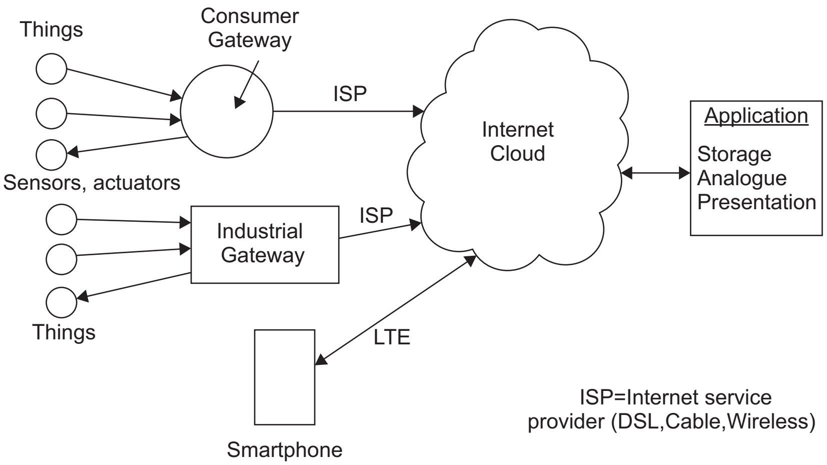
Things
Consumer Gateway
ISP
Internet Cloud
Application
Storage
Analogue
Presentation
Sensors, actuators
Industrial Gateway
ISP
Things
LTE
Smartphone
ISP=Internet service provider (DSL,Cable,Wireless)

FIGURE 9.16 How the Internet of Things systems works.

In an M2M application, the sensor or actuator node uses cellular wireless to connect to the Internet and then the storage and analysis center by way of the existing cellular network. Smartphones are also widely used at both ends of IoT or M2M connection.

Wireless Technologies of Internet of Things and Machine 2 Machine

Wi-Fi and Bluetooth are by far the most often used short-range wireless links for IoT. However, many others are also used. Table 9.3 summarizes the most popular of these standards. The IEEE 802.15.4 standard is the base for ZigBee and Thread and several others not listed. ZigBee is used in both consumer and industrial spaces. Z-Wave and Thread are typically consumer-only technologies. Sigfox and LoRa are mainly industrial standards. A growing trend is the increased use of cellular for IoT connectivity especially in business and industrial applications. Standard 2G and 3G systems have been used, but the movement is to use simpler, less expensive, and more constrained versions of LTE such as LTE M1 and NB-IoT. Most of these standards use the unlicensed 2.4 GHz band. Others use the below 1 GHz bands, either cellular or ISM bands. There are lots of choices and no interoperability or compatibility between them. That may be a problem as it creates the need for multitechnology gateways if nodes are to exchange information.

The Issues of Internet of Things

The lack of a single wireless standard may be a challenge in both consumer and industrial applications. If possible, users should select one desired standard and

TABLE 9.3 A Summary of the Most Commonly Used IoT Short-Range Wireless Technologies

Standard	Frequency (GHz)	Maximum Rate	Modulation	Maximum Range	Application
IEEE 802.15.4	2.4	250 kbps	DSSS, O-QPSK	100 m	Consumer, industrial
Wi-Fi 802.11a/b/g/n/ac/ax	2.4 and 5.8	11 Mbps–7 Gbps	DSSS, OFDM	200 m	Consumer. Commercial, industrial
Bluetooth 4.x, 5	2.4	2 Mbps	FHSS, GFSK	50 m	Consumer
ZigBee	2.4	250 kbps	DSSS, O-QPSK	100 m	Consumer, industrial
Z-Wave	Sub-1	100 kbps	FSK	30 m	Consumer
Sigfox	Sub-1	600 bps	GFSK, DBPSK	1 km+	Industrial
LoRa	Sub-1	50 kbps	SS chirp	1 km+	Industrial
LTE-M1	Sub-1	1 Mbps	OFDM	1 km+	Industrial
NB-IoT	Sub-1	250 kbps	OFDM	1 km+	Consumer, industrial

use it for everything. Next, be aware of privacy and security problems. Hackers can disrupt a network or steal data or do other more nefarious mischief. While most standards now include encryption and authentication capability, be aware that more advanced methods may be necessary to prevent hacking. Finally, realize that billions of sensors will generate a massive amount of data that has to be stored somewhere. Furthermore, the data is of little use until it is processed, analyzed, and presented. This will require significant storage and computing capability if IoT is to be useful.

Project 9.1

Become More Familiar With Wireless Data Protocols

You may want to find more information on the most popular short-range wireless data methods. The following websites are recommended. And you may want to do a general Internet search with Bing, Google, or Yahoo!

Wi-Fi: www.wi-fi.org
Bluetooth: www.bluetooth.com and www.bluetooth.org
ZigBee: www.zigbee.org

The Wikipedia entries are also quite good.

Chapter 10

Audio Electronics

Digital Voice and Music Dominate

In this Chapter:

- Sound and the electronics that deliver it.
- Digital audio.
- The compact disc (CD).
- Voice and music electronics.
- Wireless audio
- Streaming audio

INTRODUCTION

While the telephone, telegraph, and radio are generally considered the first real electronic applications, audio was certainly next. By "audio," I mean sound: voice and music, things we can hear. Audio signals have frequencies within the range of the human ear. Over the years many electronic devices have been developed to amplify, capture, and reproduce sounds of all types. Today, like everything else in electronics, audio electronics is predominantly digital, although many analog devices and systems are still in use. This chapter is a summary of audio devices you know and use every day including stereo systems, iPods, wireless speakers, and streaming audio.

THE NATURE OF SOUND

Sound is just pressure waves in the air that your ear can hear. Anything that produces a noise or other disturbance produces pressure waves that travel to your ears where the sound waves are converted into signals that your brain recognizes as sound. The frequency range of these air vibrations is 20 Hz–20 kHz. Our ears do not respond to lower or higher frequencies even though they may be present. Human hearing also varies widely due to age and health status. Age brings on hearing loss, especially of the higher frequencies.

Electronics Explained. http://dx.doi.org/10.1016/B978-0-12-811641-8.00010-2
Copyright © 2018 Elsevier Inc. All rights reserved.

There are four key things needed in audio electronics:

- The ability to convert sound waves into electrical signals.
- The means to convert the electrical signals into sound.
- A way to amplify and reproduce the signals.
- Media to store or record the signals.

This is what audio electronics is all about.

Microphones

The microphone is the transducer that converts sound waves into an analog signal. The sound waves impinge on the microphone, and it causes vibrations that are translated into a signal by either capacitive element or an inductive coil. A technical discussion about microphones is way beyond the scope of this book. In any case, the resulting signal is usually very small, in the millivolt range or even microvolt range. Therefore, it must usually be amplified before it is useful. This is the function of a small-signal amplifier usually called a *preamplifier*. What you do with the signal after that is up to you. Typically it will be amplified further for use in a public address or other sound system. Or alternately, the voice may be recorded. Many voice signals will be used in a telephone, cordless, or cellular.

Speakers

A speaker is a transducer that converts the audio signal into sound waves that you can hear. You have obviously seen a speaker, so little explanation is needed. Its main element is a paper or plastic cone that vibrates at the audio frequencies applied to it. Then it produces the sound pressure waves that our ears will hear.

Fig. 10.1 shows how a speaker works. The cone is attached to a form around which a coil of wire is wound. This is the so-called *voice coil*. That voice coil is

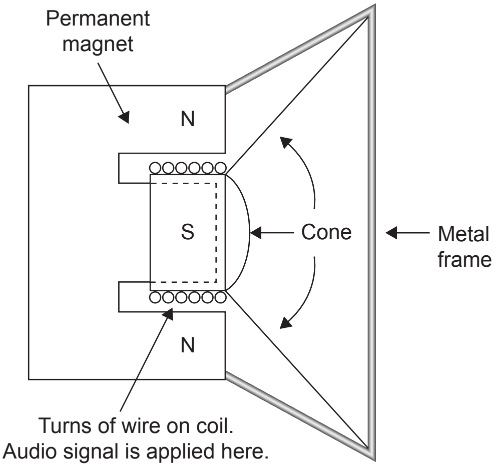

FIGURE 10.1 Cross section of speaker showing major components and how it works.

positioned inside a strong permanent magnet with north and south poles. Now when you apply the audio signal to the coil, current flows and causes the coil to become an electromagnet. The electromagnet generates alternating north and south poles as the voice or music signal occurs. These magnetic fields interact with the permanent magnet, causing attraction and repulsion of the voice coil, and the cone moves. The motion of the cone accurately reproduces the original sound waves.

Speakers come in a wide range of sizes and types. Cone speakers are usually called midrange speakers in which they can reproduce most of the frequency range except for the very low and very high frequencies. Special speakers are used for those frequencies. A woofer or subwoofer is a larger speaker designed for frequencies of about 600 Hz and below. A tweeter is a special speaker optimized for the higher frequencies above 10 kHz. Special filters called *crossover networks* separate the signal into low, mid, and high frequencies as shown in Fig. 10.2.

The remainder of this chapter addresses the amplification and recording or storage of audio signals. Along the way, most analog audio signals are digitized. And many are also compressed to save storage space or reduce the binary data rate for transmission.

DIGITAL AUDIO

The sounds we hear and the sounds we generate are analog signals. A microphone picks up the voice and music and produces an analog signal to be amplified or stored. We have amplifiers and speakers that will accurately reproduce the sounds. But the real challenge with audio is capturing, storing, or recording

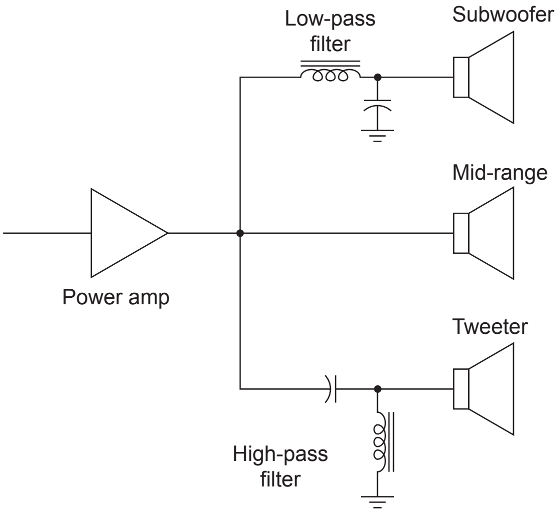

FIGURE 10.2 Cross over networks keep signals of different frequencies separate for appropriate speakers.

the sounds for later reproduction. With today's digital techniques, we can store not only voice and music more accurately than before, but we can also transmit it more easily by radio, TV, or the Internet.

Past Recording Media

Edison invented the phonograph in 1877. This was the first device ever conceived to record voice. A microphone converted the sound into vibrations that moved a sharp needle in direct accordance with the sound. The needle cut a varying groove in the wax coating on a rotating cylinder. Later, a needle placed in the groove moved the diaphragm in a horn speaker that translated the movement into sound. The quality was poor but it worked, and the concept was soon translated into what we know as a *phonograph record*. The record disc is rotated while a needle cuts a spiral groove into the plastic that accurately captures the sound. To reproduce the sound the disc is rotated at the same speed and a needle is placed in the groove to reproduce the sound in a speaker. A transducer converts the movement of the needle in the groove into an analog signal that is amplified.

Early records spun at a speed of 78 rpm and could only hold a few minutes of sound. Later a smaller 45-rpm record was invented to play minutes of music at a lower cost. Then a larger long-playing record spinning at only $33^1/_3$ rpm was produced that could hold an hour or so of sound. Stereo or two channel sound was also effectively recorded into a V-shaped groove in the record.

Records were soon upstaged by magnetic tape. A plastic such as Mylar was coated with a powdered iron or ferrite magnetic material. The music or sound was then applied to a magnetic coil that put a magnetic pattern on the tape representing the sound. A coil passed near the tape would convert the magnetic pattern variations into a small voltage to recover the sound. Reel-to-reel tape was popular for a while, but magnetic tape became the recording media of choice for years when the Philips cassette tape was invented. Other formats like the ever-popular 8-track unit were also used.

All of these formats worked well but had their limitations. Record grooves wore out with usage causing the high-frequency response to be eroded. Scratches and dirt on the record also produced noise. As for tape, it had good frequency response but had its own noise that could not be reduced below a certain level. The dynamic range of both was poor.

Digital technology has solved all these problems. Now the most popular medium is the compact disc (CD), which records sound in full digital format. It has very low, practically undetectable noise, wide frequency response, and a very long life. The cost is also low.

Digital sound is also routinely stored in computer memories like flash EEPROMs and transmitted over the Internet. With digital techniques, sound can also be processed with digital signal processing (DSP) such as compression, filtering, and equalization.

Dynamic Range

Dynamic range refers to the difference between the highest and lowest volumes of sound that a system can handle. It is a measure of how wide an amplitude range a particular storage media can record or an audio system can accommodate. It is also a way to state the maximum sound pressures that our ears can withstand and the lowest levels that we can hear.

The upper level of amplitude is set by our ears and the level at which distortion occurs. The lowest levels of the dynamic range are determined by the noise in the system. Noise is the random voltage variation caused by thermal agitation of electronics in electronic components and random variations in the magnetism on a magnetic tape or disc. As the volume of a sound gets lower at some point, it will be smaller than any noise present and not distinguishable from the noise background.

In general, the wider the dynamic range the better. It usually means lower noise and greater volume (loudness) to better match the range of human hearing.

Dynamic range can be stated in terms of power or voltage. It is usually a very wide range, so we resort to a unique mathematical method to express it. The term decibel (dB) refers to how dynamic range is presented. In math form you take the ratio of the maximum and minimum voltage, take the logarithm of that, and multiply by 20:

$$\mathrm{dB} = 20 \log (V_{max}/V_{min})$$

For power,

$$\mathrm{dB} = 10 \log (P_{max}/P_{min})$$

The dynamic range of a typical phonograph record is probably about 50 dB, and magnetic tape, about 55 dB. But a CD is approximately 90 dB, the best we have.

Dynamic range can be improved by lowering the noise in a system. This can be done with special signal-processing techniques. The most notable are methods produced by Dolby.

Digitizing Sound

Most digital sound and music is digitized with a sampling rate of 44.1 or 48 kHz. Recall the Nyquist requirement that to capture and retain the frequency content accurately, the analog signal has to be sampled at least twice the highest frequency of the sound. Since most music and the human ear cutoff at about 20 kHz, then any sampling rate greater than 2 kHz × 20 kHz = 40 kHz will satisfy that requirement. The rate of 44.1 kHz was standardized for CDs and 48 kHz for professional recordings. At 44.1 kHz, a sample is taken every 1/44,100 = 22.676 μs.

Fig. 10.3 shows the sampling of an analog signal. The smooth curve is the analog music or voice. At each vertical interval a sample is taken. In this example, a 4-bit ADC is used so there are $2^4 = 16$ possible levels. Note the 4-bit code associated with each sample amplitude. The binary samples are stored

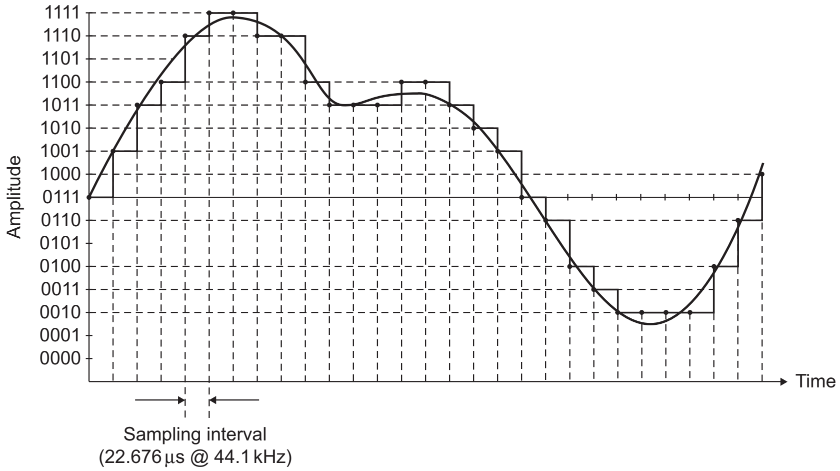

FIGURE 10.3 How audio signal is digitized. In this example, each sample is assigned a 4-bit code. The stepped approximation shown is what a DAC would deliver to the speakers.

for later playback. When sent to a DAC, the samples will reproduce the audio in a stepped approximation shown overlaid with the analog signal. Some low-pass filtering of that signal will produce a sound which almost duplicates the original.

Digitizing audio usually results in each sample producing a 16-bit word. With 16 bits you can represent $2^{16} = 65{,}536$ levels. That produces a dynamic range from the highest to lowest amplitude levels of 65,536 to 1. On the decibel (dB) scale, that is a range of 96 dB—far greater than any dynamic range of previous analog recording methods. And noise is virtually nonexistent. It is so low no one can hear it. The overall result is a highly accurate way to capture sound and music. The 16-bit words can then be conveniently stored in a memory or captured on a compact disc.

Digital Compression

Digital compression is a mathematical technique that greatly reduces the size of a digital word or bitstream so that it may be transmitted faster or stored in a smaller memory. Digitizing sound creates a huge number of bits. Assume stereo music that sampled at a rate of 44.1 kHz to create 16-bit words for each sample. One second of stereo music, then, produces $41{,}000 \times 16 \times 2 = 1{,}411{,}200$ bits. A 3-min song is $60 \times 3 = 180$ s long. The result is $1{,}411{,}200 \times 180 = 254{,}016{,}000$ bits. Since there are 8 bits per byte, the result is 31,742,000 bytes or nearly 32 MB or megabytes. That is an enormous amount of memory for just one song. With a recording medium like the CD with a storage capacity of about 700 MB, that is okay. But for computers or portable music devices, it is impractical, not to

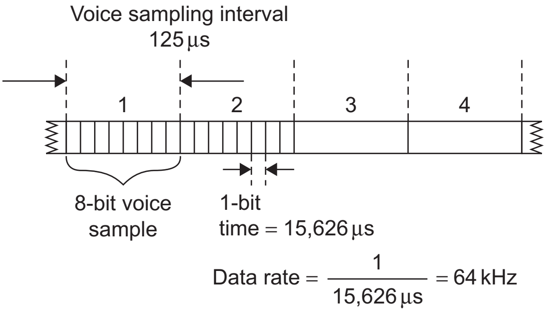

FIGURE 10.4 Voice signals up to 4 kHz for the telephone are sampled at an 8-kHz rate, producing an 8-bit sample, each 125 µs. Stringing the samples together in a serial data stream produces a digital signal at a 64-kHz rate.

mention expensive. And to transmit that over the Internet would take about 4 min at a 1 Mb/s rate. Pretty slow by today's standards.

The solution to this storage and transmission problem is to compress the bitstream into fewer bits. This is done by a variety of mathematical algorithms that greatly reduce the number of bits without materially affecting the quality of the sound. The process is called *digital compression*. The music is compressed before it is stored or transmitted. Then it has to be decompressed to hear the original sound.

The two most commonly used music compression algorithms are MP3 and AAC. MP3 is short for MPEG-1 Audio Layer 3, the algorithm developed by the Motion Picture Experts Group as part of a system that compressed video as well as audio. AAC means advanced audio coding. MP3 is by far the most widely used for storing music in MP3 music players and sending music over the Internet. AAC is used in the Apple iPod and iPhone and used on the iTunes site to send music. It is also part of later MPEG-2 and MPEG-4 video compression formats. Both methods significantly reduce the number of bits to roughly a 10th of their original size, greatly speeding up transmission and easing storage requirements. There are many more compression standards out there, but these are by far the most used and the ones you will most likely encounter.

To perform the compression process you actually need a special CPU or processor. It is typically a special DSP device programmed with the algorithm for either compressing or decompressing the audio.

There are also a number of compression methods used just for voice. Voice compression was created to produce signals for telephony. Most phone systems assume a maximum voice frequency of 4 kHz. The most common digitizing rate is twice that or 8 kHz. 8-bit samples are typical. If you digitize voice creating a stream of samples in serial format, the signal would look like that in Fig. 10.4. Each sample produces an 8-bit word where each bit is 125/8 = 15.625 µs long. That translates to a serial data rate of 1/15.625 µs = 64 kbps. This takes up too much bandwidth in a telephone system, so compression is used. The

International Telecommunications Union (ITU), an international standards organization, has created a whole family of compression standards. These are designated as G.711, G.723, G.729, and others. These mathematical algorithms reduce the bit rate for transmission to about 8 kbps. You will see them used in voice over Internet phone (VoIP) digital phones, which are gradually replacing regular old-style analog phones.

There are many other forms of audio compression. Another common one is Dolby Digital or AC-3 that is used in digital movie theater presentations and some DVD players.

How an MP3 Player Works

MP3 players such as the Apple iPod are amazing electronic systems in a hand-held package. They all take the general form shown in Fig. 10.5. At the heart of the player is a processor. It may be a general processor or a DSP. In some cases, two processors are used, one to control the player buttons, controls, and display, and other general housekeeping functions; a separate DSP handles the compression/decompression. A ROM holds the general control program for the device. A large flash EEPROM is used to store the music in compressed form. Some devices also provide for an external flash memory plug-in device to hold more music. A USB port is the main I/O interface to connect to a PC for the music downloads.

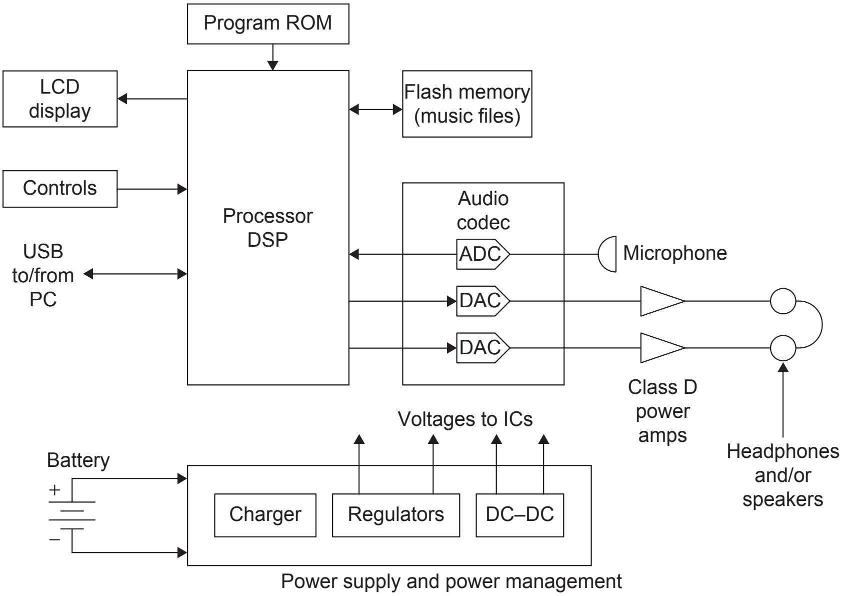
Program ROM
LCD display
Flash memory (music files)
Controls
Processor DSP
Audio codec
USB to/from PC
ADC
Microphone
DAC
DAC
Class D power amps
Headphones and/or speakers
Voltages to ICs
Battery
+
−
Charger
Regulators
DC–DC
Power supply and power management

FIGURE 10.5 General block diagram of a digital music player like Zune MP3 or Apple iPod.

The codec in the device is the coder–decoder. This chip holds the ADC to digitize music from an external source such as a microphone (coder) and the DACs that decode or translate the decompressed audio back into the original analog music for amplification. Class-D switching power amplifiers are used for the headphones to reduce power consumption. One or more internal speakers may also be used. Other circuits include the LCD display and its drivers and LED backlights if used. A power management chip is normally used to provide multiple DC voltages to the different circuits and to manage power to conserve it. This section also contains the battery charger.

The Compact Disc

CD is by far the most popular digital audio medium in use today. It has been around since the 1980s and is still going strong. It has virtually totally replaced the Philips cassette magnetic tape cartridges and vinyl phonograph discs. About the only competition the CD has is from flash EEPROM devices such as those used to store MP3/AAC music for portable music players.

You have probably seen a CD. Let us add a little background so you will appreciate the CD. It is a 4.72-in. (120-mm) diameter disc made of plastic. It is about 0.05-in. (1.2-mm) thick and has a hole in the center for mounting on the motor that spins the disc. It is actually a sandwich of clear plastic and a thin metal layer for reflecting light.

The music is recorded on the disc in digital format. A laser actually burns *pits* into the plastic representing the 1's and 0's of the digital data (see Fig. 10.6). These pits are extremely small, about 0.5-μm wide (a micron designated μm is 1-millionth of a meter). The flat areas between the pits are called *lands*. A binary 1 occurs at a transition from a pit to a land or vice versa while the land is a string of binary 0's. The binary data is recorded as a continuous spiral track about 1.6-μm wide, starting at the center and working outward. That is a very efficient way to store the data, but it means that the motor speed must vary to ensure a steady bit rate when the music is played back. The pickup mechanism will experience higher speeds at the center of the disc and lower speeds at the outer edges of the disc, so that a speed control mechanism keeps the recovered data rate constant.

The music on a CD is derived from two stereo channels of music with a frequency response from 20 Hz to 20 kHz. Digitization occurs at the standard 44.1-kHz rate. Each sample is 16-bits long. The 16-bit words from the left and right channels are alternated and formatted into a serial data stream that occurs at a rate of 44.1 kHz × 16 × 2 = 1.4112 MHz. The 16-bit words are then encoded in a special way. First, they undergo an error detection and correction encoding scheme using what is called *cross-interleaved Reed–Solomon code* (CIRC). This coding helps detect errors in reading the disc caused by dirt, scratches, or other distortion. The CIRC adds extra bits that are used to find the errors and fix them prior to playback.

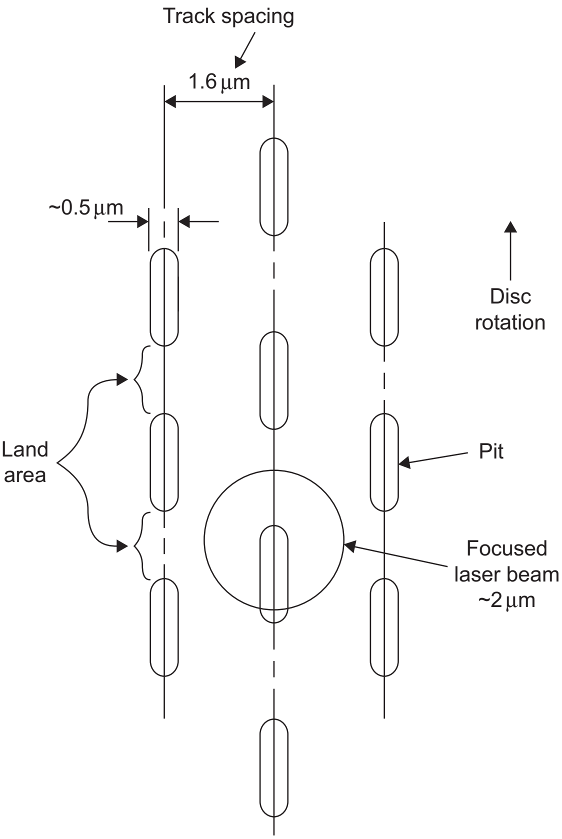

FIGURE 10.6 Digital music on a CD is encoded as pits and lands on the CD surface. These are burned in by a laser. A laser light shined on the pits and lands produces a reflection that a photodetector picks up to create the original digital data.

Next, the serial data string is then processed using what is called eight-to-fourteen modulation (EFM). Each 8-bit piece of data is translated into a 14-bit word by a lookup table. EFM formats the data for the pit-and-land encoding scheme. Finally, the completed data is formatted into frames 588-bits long and occurring at a rate of 4.32 Mbps. That is the speed of the serial data coming from the pickup assembly in a CD player before processing.

To recover the audio from the CD, the disc is put into the player and a motor rotates it. Refer to Fig. 10.7. A laser beam is shined on the disc as indicated in Fig. 10.6. The reflections from the pits and lands produce an optical light pattern that is picked up by a photodetector that converts the light variations into the 4.32-Mbps bitstream. A motor control system varies the speed of the motor to keep the data rate steady. The data stream goes to a batch of processing circuits. First, the EFM is removed, and then a CIRC decoder identifies and repairs

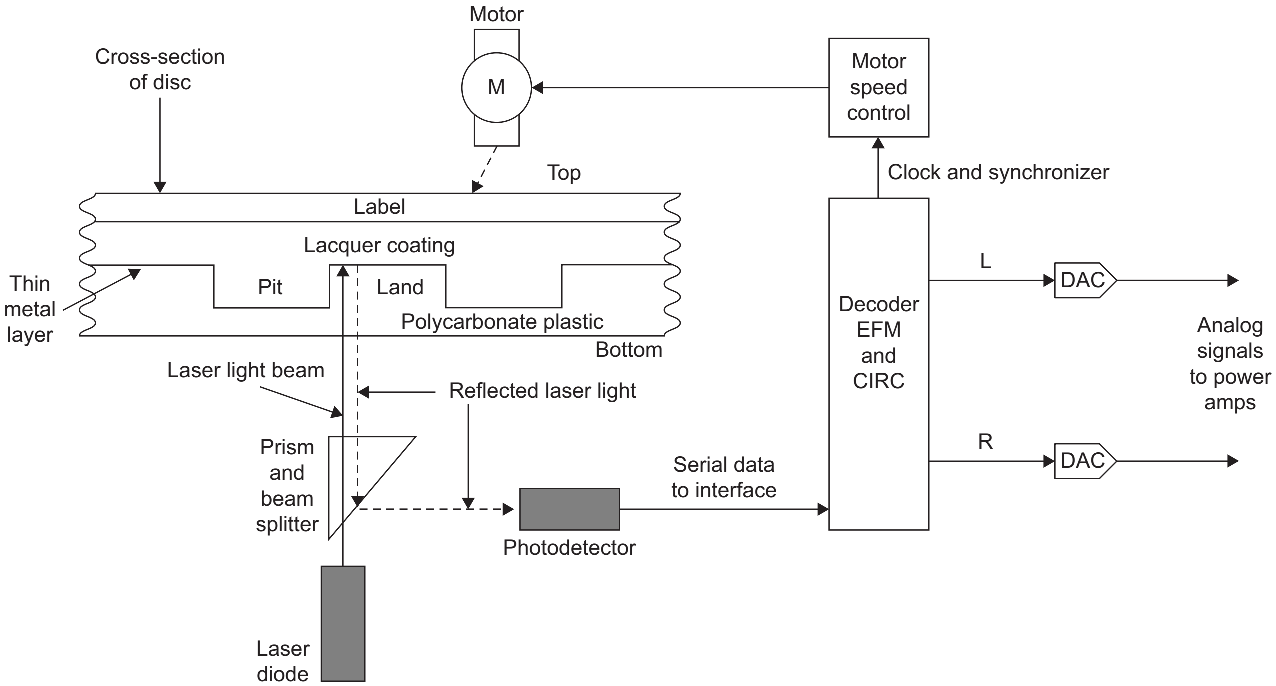
Motor
M
Motor speed control
Cross-section of disc
Top
Clock and synchronizer
Label
Lacquer coating
Thin metal layer
Pit
Land
Polycarbonate plastic
Bottom
L
DAC
Decoder EFM and CIRC
Analog signals to power amps
R
DAC
Laser light beam
Reflected laser light
Prism and beam splitter
Serial data to interface
Photodetector
Laser diode

FIGURE 10.7 Block diagram of CD player.

any errors and recovers the original 1.41-Mbps data stream. A demultiplexer separates the left- and right-channel 16-bit words and sends them to the DACs, where the original analog music is recovered and sent to the power amplifiers and speakers.

The CD can store lots of data. Its capacity is in the 650- to 700-MB range. That translates into a maximum of about 74 min of audio. And this is not compressed.

AUDIO–VIDEO RECEIVER

An HDTV receiver is usually at the heart of most consumer electronics home entertainment systems. A secondary piece is the audio–video (AV) receiver, which provides the audio component of the entertainment. The systems usually include the CD player, several radio options, and all the audio power amplifiers that operate the multiple speakers. The AV receiver also accepts inputs from the TV set, DVD player, and other external devices to provide higher quality sound. This section provides a look at this piece of equipment and how it works.

Fig. 10.8 shows a general block diagram of the AV receiver. Note the multiple power amplifiers on the right. These drive the speakers that are part of a surround-sound system. Most of these are class AB linear amplifiers that deliver power levels from about 20 W per channel to over 100 W in the larger units. The frequency response is 20 Hz–20 kHz or more with a distortion level of 0.5%–0.9% or less.

The power amplifiers get their inputs from multiple sources. These sources include audio from the internal radio receivers and the CD player as well as many other external sources. These sources include the TV receiver, DVD player, VCR, and a cable or satellite box, as well as audio from older devices such as tape players and phonographs. Many receivers also accept inputs from an iPod or other MP3 music player. A large switching matrix is used to select the desired input. That selection can be made from the front panel controls or via the remote control.

Radio Choices

All AV receivers contain the traditional AM and FM analog radios. These are still popular sources of music. But more and more the most recent receivers contain one or more digital radio sources such as HD radio or satellite radio.

HD radio is becoming more popular each year. This is a digital radio broadcast option available at most US radio stations today. The AM and FM sources are digitized and broadcast on the same frequencies but with digital modulation techniques like orthogonal frequency division multiplex a wide separate channel is needed. The result is an improvement in the signal. Digital signals offer slightly better frequency response on both AM and FM. AM signals will sound

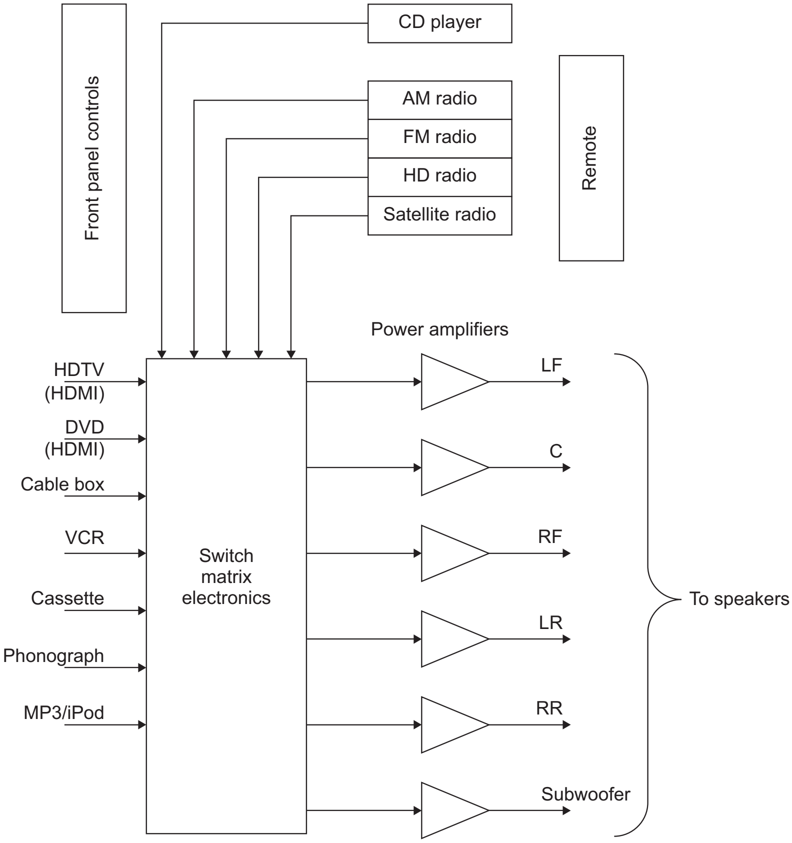
Front panel controls
CD player
AM radio
FM radio
HD radio
Satellite radio
Remote
Power amplifiers
HDTV (HDMI)
DVD (HDMI)
Cable box
VCR
Cassette
Phonograph
MP3/iPod
Switch matrix electronics
LF
C
RF
LR
RR
Subwoofer
To speakers

FIGURE 10.8 Block diagram of typical audio–video receiver, the heart of the audio reproduction part of a consumer home entertainment center.

more like FM, and FM signals will be nearly CD quality. Furthermore, the digital signals are more noise and fade free than the standard AM and FM signals. While HD radio was first made available in automobiles, it is now widely available for home use. Separate HD radios can be purchased at most consumer electronics stores. And more HD radios are being built into AV receivers.

Satellite radio was also first made popular in automobiles. Yet separate home receivers are available. Some AV receivers now include satellite radio capability. Satellite radio is available from one source, SiriusXM Radio. This is a subscription service with hundreds of radio, music, news, sports, and other channels. The signals are digital, delivering near CD-quality sound. The satellites use the 2.3-GHz microwave band. The main difficulty when using satellite radio at home is that an outside antenna is needed. Some radios supply a small antenna for window mounting so that the antenna can "see" the satellites in the sky. Otherwise, poor or no reception may occur.

Surround Sound

Virtually all audio sound today is stereo. That is, the music comes from at least two speakers, one on the left and one on the right. The sound is recorded as two separate channels with different microphones to provide a stereophonic experience to the listener. A minimal system uses two speakers but some add a subwoofer for good bass reproduction. This is sometimes referred to as a 2.1 stereo system, that is, two speakers and one subwoofer.

New audio systems offer surround sound. These systems feature six, seven, or eight speakers. The idea is to totally envelop the listener with sound as it might be heard in a concert hall or theater. There are speakers in front and at the rear. The first surround-sound implementations were analog but today most surround sound is fully digital. The most common surround-sound method is Dolby Digital 5.1, which uses five speakers and a subwoofer as shown in Fig. 10.9. These are left front (LF), right front (RF), center (C), left rear (LR), and right rear (RR). Then there is one subwoofer. The woofer can be positioned anywhere but is usually at the front. Each speaker is driven by its own amplifier in the AV receiver.

The big question that you may have is where do the six channels of sound come from? Not from the radio or a CD for sure, both of which are stereo only. In most cases, the sound signals will come from a DVD player. Right now it is about the only common source of 5.1 surround sound, although some HDTV programs support it. During recording of the original material, one microphone is used for each channel. The digitized audio for each of the channels is stored on the DVD.

Incidentally, there are more elaborate versions of surround sound. For example, 7.1 adds two more speakers on the left and right. As of now there are few, if any, 7.1 music sources available.

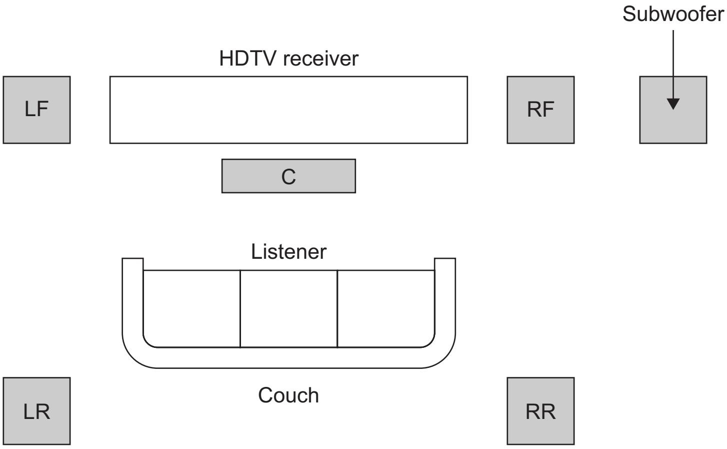

FIGURE 10.9 How 5.1 surround-sound system is set up.

When you play the DVD, the digital signals from each channel are sent to a special chip that decodes them and otherwise performs necessary noise reduction, equalization, and other operations. The recovered signals are then sent to DACs for conversion to their original analog format for amplification.

SPECIAL SOUND APPLICATIONS

When you think of audio or sound, you think mainly of voice or music reproduction. Yet, electronic sound is used in lots of additional places. Some interesting examples are in wireless audio, streaming music, musical instruments, sonar, ultrasound, and hearing aids.

Wireless Speakers and Headphones

A popular audio feature today is wireless speakers. Such speakers are powered by batteries or may be plugged into the AC mains. The audio connection is wireless commonly using the Bluetooth radio technology. Some newer systems use Wi-Fi for the wireless connection. Such speakers help to minimize the wiring mess in a home entertainment system. They also are popular for providing louder and better quality audio for smartphones and tablet/laptop computers. Bluetooth headphones are also a popular accessory for smartphones.

Downloading and Streaming Audio

While CDs are still as popular music medium, even more popular today is downloading music. Songs are selected from a music source such as Apple's iTunes and the compressed digital music file is downloaded via the Internet to an MP3 player or a smartphone. Music is usually purchased one song at a time and once downloaded it can be played.

Streaming audio is the term given to receiving audio mostly music over the Internet. The compressed music files are delivered in real time to a computer or smartphone and played directly. The streaming requires specific software that performs the transfer, memory buffering, decompression, and playing. The computer or smartphone and its software takes the music file into a buffer memory then parses it out continuously to the player. A wide range of streaming services are available such as iTunes Radio, Pandora, Spotify, Google Play, Beats, and others.

Musical Instruments

Musical instruments produce their own sounds, of course. Each has its own unique signature made up of the basic notes or tones to which are added the mix of harmonics or overtones unique to the physical structure of the instrument. Some instruments like horns and drums rarely need amplification. Other instruments like guitars and violins commonly need some amplification, especially in large venues such as on stage addressing hundreds or thousands of people. Special microphone-like devices called pickups are used with string instruments. The pickup is usually one that converts the string vibrations into a small signal that is then amplified.

One of the most popular electronic instruments is the *synthesizer*. It has a keyboard like a piano. The keys are equivalent to those on a piano. Each key then triggers the electronics to generate a tone of the desired frequency. That tone is especially modified by DSP circuits to sound like that note played on a piano, an organ, or any other musical instrument. The DSP synthesizes or constructs the tones digitally and sends them to a DAC that converts them into the analog sound to be amplified.

Are Vacuum-Tube Amplifiers the Best Audio Amplifiers?

Vacuum tubes? I'm kidding, right? Actually, no. There are many audio experts, especially guitar players and some recording studio engineers, who still believe that vacuum-tube amplifiers are superior to solid-state amplifiers. And as it turns out, the largest market for general vacuum tubes today is guitar, high-end stereo, and recording studio amplifiers.

Contrary to popular belief, vacuum tubes did not really go away when transistors and integrated circuits came along. They are still widely used in China, Russia, and a few other countries. And that is where most tubes are made these days. Vacuum tubes are still widely used. Cathode ray tubes or picture tubes are still at the heart of many TV sets and computer video monitors. Magnetrons are vacuum tubes that are in every microwave oven. And microwave vacuum tubes, such as traveling-wave tubes and klystrons, are still used in high-power microwave transmitters. And vacuum-tube audio amplifiers.

Why are vacuum-tube amplifiers so popular still? It is the unique sound they produce. It is a sound that musicians appreciate and want in their sound mix. And the sound is difficult and expensive to reproduce in a solid-state amplifier. The sound is smoother and less harsh. The sound is "warmer," and the compression that such amplifiers produce when overloaded is exactly what makes rock and country music sound like it does. It looks like vacuum tubes will be around much longer, but prices are sky high and some tube models are hard to find.

Noise-Canceling Earphones

You have probably heard of these or maybe seen or used them. They make clever use of analog signal processing to greatly minimize surrounding noise when you are trying to listen to music with headphones.

First, most of these headphones try to isolate the actual headphone from the surroundings with a tight-fitting ear bud or soft foam pads around the earphone that limit the amount of external noise getting to your eardrum. This is called *passive noise control.*

Second, each earphone contains a built-in microphone that picks up the surrounding noise. It amplifies it then inverts it in phase. The original noise from the microphone is then added to the inverted noise. Being the same but exactly out of phase with one another, the two noise signals cancel one another. The cancellation is not perfect, but the noise reduction is significant. It can greatly minimize noise on a plane or in a car or other types of noisy environments.

Sonar

Sonar means sound navigation and ranging. It is basically the underwater equivalent of radar. It is used on ships and submarines for making depth measurements, detecting underwater objects, and for warfare. The two types of sonar are passive and active.

Passive sonar is simply the idea of putting underwater microphones called *hydrophones* on a long wire or arrays of such microphones in the water to listen for any sounds that occur. Amplifiers boost the signal levels and various filters help sift through all the complex signals that can be heard. In military applications, the sounds are digitized and various DSP programs help filter and identify specific sounds. Many ships create unique sounds called signatures that can be digitized and stored for comparison to any sounds picked up for positive identification. The sounds include fish and other sea life, boat and ship engines, motion through the water by a ship or boat, or any noise that occurs inside a submarine. Sound travels very easily through water so is easy to pick up. With practice and experience, a sonar operator can identify almost everything occurring over a wide range around the microphone array.

The other form of sonar is *active*. This is like radar where a high-frequency pulse of sound is applied to a transducer to convert the pulse into sound waves for transmission through the water. The frequency of the pulse is usually ultrasonic, that is, above most human hearing. Frequencies from about 15 kHz to 1 MHz are used, depending on the application. The pulse of sound travels out from the transducer and is then reflected off distant objects. The reflections go back to the source where they are picked up by a microphone. Then, knowing the speed of sound transmission in water, the distance to the object can be computed. Sonar depth sounders point the transducer downward to get a measure of the distance to the bottom. Sonars can easily pick up reflections from ships miles away for navigation safety purposes or for military actions.

Hearing Aids

Electronic hearing aids have been around a long time. The older analog types were large and cumbersome but worked fine. Integrated circuits made them much smaller and more acceptable to the hearing impaired. Today, the newer hearing aids use digital technology. The microphone picks up the sound and amplifies it, then digitizes the sound in an ADC. A DAC can then translate the sound back into audio for amplification and application to the earphone. Most digital hearing aids now incorporate DSP filters. These filters may be customized to the frequency response of the ear, correcting for different frequencies. Mostly the high-frequency response is lost due to aging. Using an external computer, each hearing aid can be adjusted from volume and frequency response to fit the exact deficiency.

Project 10.1

Experience Digital Radio

Digital radio in the United States is available from satellites and via the existing AM and FM analog stations through HD radio. Both offer commercial free music and other content in multiple channels. HD radio is free but satellite radio is a pay-for-listen service.

In the United States, two satellite radio services were established in the early 2000s, XM Radio and Sirius Radio. The two services merged to form Sirius XM Satellite Radio. Both offer subscription services to music, news, sports, and other information. Both services operate in the 2.3-GHz band, and get their signals from satellites. Special receivers and antennas are required.

Satellite radio is a digital service and both systems used compressed audio for signal transmission. HD radio is simultaneously broadcast with existing AM and FM analog signals. A special receiver is needed to receive it. The digital techniques greatly improve sound quality.

A good project if you are a music lover is to try one or both of these services. You can get an HD or satellite radio for your vehicle or you can buy a home unit. Some high-end AV receivers have one or the other in addition to the traditional AM and FM radios. You can purchase home desktop units from Best Buy and other electronics dealers for a reasonable price.

HD radio is free and you will find that some stations transmit additional digital channels not available in analog form. These give you more music genres. No special antenna is needed.

To learn more about HD radio, do a Bing, Google, or Yahoo! search. Also go directly to the website of the company that invented HD radio, iBiquity (www.ibiquity.com), for more details.

You can also buy a satellite radio from several local sources. But you will need to subscribe. Go to www.siriusxm.com.

Project 10.2

Test Your Hearing Response

To check the frequency response of your ears, go to www.audiocheck.net/audiotests_frequencycheckhigh.php. There are a few online audio tests for your hearing. This website has a whole batch of other audio response tests that you may be interested in.

Chapter 11

Video Technology

TV and Video Is All Digital Now

In this Chapter:

- How scanning works to create video.
- Basic color TV principles.
- US and other standards for digital and high-definition TV.
- How a digital TV set works.
- TV screen technologies.
- Cable TV.
- Satellite TV.
- Over-the-Top TV
- Cell phone TV.
- DVD players.
- Virtual reality.

INTRODUCTION

Nothing in electronics has affected us more than the development of television. Okay, I have not forgotten the PC and the Internet, which have also seriously impacted our lives. But TV came first and has been working its magic on us for far longer. The impact of television or video on our lives is significant in that we spend more than half of our waking hours in front of a video screen, either watching TV or working with a video monitor on a computer. And, of course, you also interact with the liquid crystal display (LCD) screens on smartphones, laptops, smart watches, and personal navigation devices. The social and political impact has been enormous. For this reason it helps to understand a little bit about how video works. Video is perhaps one of the most complex segments of electronics, but this chapter helps make those fundamentals easy to understand.

VIDEO FUNDAMENTALS FOR THE IMPATIENT

The main feature of any video device is the screen. The first video or TV screens were cathode ray tubes (CRTs), and those are still around in some older TV sets

Electronics Explained. http://dx.doi.org/10.1016/B978-0-12-811641-8.00011-4
Copyright © 2018 Elsevier Inc. All rights reserved.

and computer monitors. The most commonly used TV and video screen today is the LCD. LCD screens are used in the newer TV sets and in most personal computers. They are also used in laptops, smartphones, personal navigation devices, and anything else that has a screen on it. Some older TV sets also use plasma, digital light processing (DLP) projection technology, and other methods to display very large pictures. Now newer screens use light-emitting diodes (LEDs). There is more detail on TV screens later in this chapter, but for now, the most important part of understanding video is to learn the methodology used to create and present video regardless of the screen type.

A video system is a method for converting a picture or scene into an electronic signal that can be transmitted by radio or cable or stored electronically. Some type of device is needed to take in the light and color information from a picture or scene and convert it into an electrical signal. Once in electrical form, the video can be modulated onto a radio carrier for transmission or electronically stored. The process of converting the scene or picture into an electrical signal is known as scanning.

Scanning Fundamentals

A video signal is usually generated by a video camera. The camera takes the scene and scans it electronically, dividing it into hundreds or even thousands of very thin horizontal lines (see Fig. 11.1).

The horizontal scan lines comprise a sequence of light and color variations. The camera produces an analog signal whose frequency and amplitude vary with the brightness and color changes along the scan line. The whole idea in video is to transmit the video signal one scan line at a time at very high speeds. If you are not transmitting the signal by radio or cable, you can store it, such as recording it on a computer hard drive or magnetic tape. In any case, the received signals are then converted back into scan lines and applied to the screen by scanning the lines across the receiving screen very fast from left to right and top to bottom.

The secret to producing a high-quality video picture is to use enough scan lines with enough detail in each, so that the color, brightness, and fine detail in the picture are retained. Note in Fig. 11.1 that the analog signal for one scanned line is such that the higher-amplitude voltages are closer to black, while the lower-amplitude signals are closer to the white level. This is what a scan line from a monochrome or black-and-white (B&W) video system produces. It is called the *brightness* or *Y signal*. Color will be covered later.

Aspect Ratio

When talking about TV screens, the key factors are size and aspect ratio. The *size* is usually the diagonal measure of the screen in inches. The *aspect ratio* is the ratio of the picture width to the picture height. Standard TV sets and video

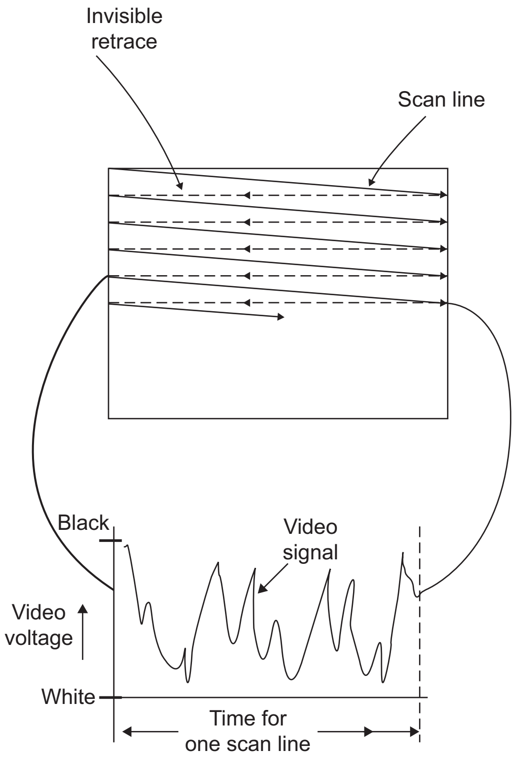

FIGURE 11.1 How scanning converts a scene of light and color into an electronic signal.

monitors previously used an aspect ratio of 4:3 as shown in Fig. 11.2. Most of the newer TV screens and video monitors have an aspect ratio of 16:9. This wider format was chosen because it is a better fit with the wide-screen movie formats.

Persistence of Vision

When you think about it, it just does not seem possible that a complete picture or scene can be reproduced in electronic format by scanning. The scanning process just breaks the picture or scene down into hundreds or perhaps thousands of scan line signals, and the only way that a picture can be created is for those scan lines to occur at a very high rate of speed. Luckily, this technique works because we all have a visual characteristic known as persistence of vision. Persistence is that characteristic of the eye that prevents it from following very high-speed changes in light variations. In other words, if the scan lines are traced across the screen at a very high rate of speed, your eye simply cannot follow them one at a time but instead puts the whole thing together and sees a single steady picture. If you scan at too low rate, you will actually begin to see the scan lines, and your eye will interpret this as a kind of flicker.

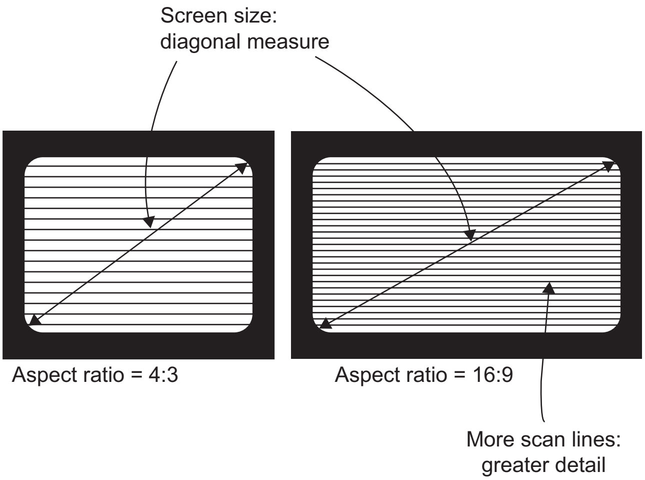

FIGURE 11.2 TV screen size and aspect ratio.

The concept of persistence of vision has been known for decades. Motion picture films take advantage of this characteristic of our eyes as well. Remember that movies are just a high-speed sequence of individual still pictures recorded on film. If you can get the frame rate up to a level of approximately 24 frames per second or faster, your eye will translate the changes from one still frame to the next as motion.

Analog TV

As you may know, the Federal Communications Commission eliminated analog TV broadcasts in the United States as of June 12, 2009. But while analog TV has been replaced by digital TV (DTV), that does not mean that it is not still around. You will still find analog TV in many other countries of the world, and it is also still used in closed-circuit television. In any case, it is such a widely used standard that understanding it is useful.

Back in the late 1940s and early 1950s, the United States created what is called the National Television Standards Committee (NTSC) TV standard. It basically specifies that a picture is made up of 525 horizontal scan lines, known as a video *frame*. As it turns out, one frame is actually made up of two fields of 262.5 scan lines. The scene or picture is first scanned 262.5 times from top to bottom, creating one field. This field is then transmitted at a rate of 60 fields per second. Then the scene is scanned again with another 262.5 lines, and these are transmitted next. The second field is interlaced with the first field to form the complete 525-line picture. See Fig. 11.3. The result is that you will see 30 complete 525-line frames per second. This is fast enough for persistence of vision to work nicely. The interlacing helps significantly in reducing any perceived flicker.

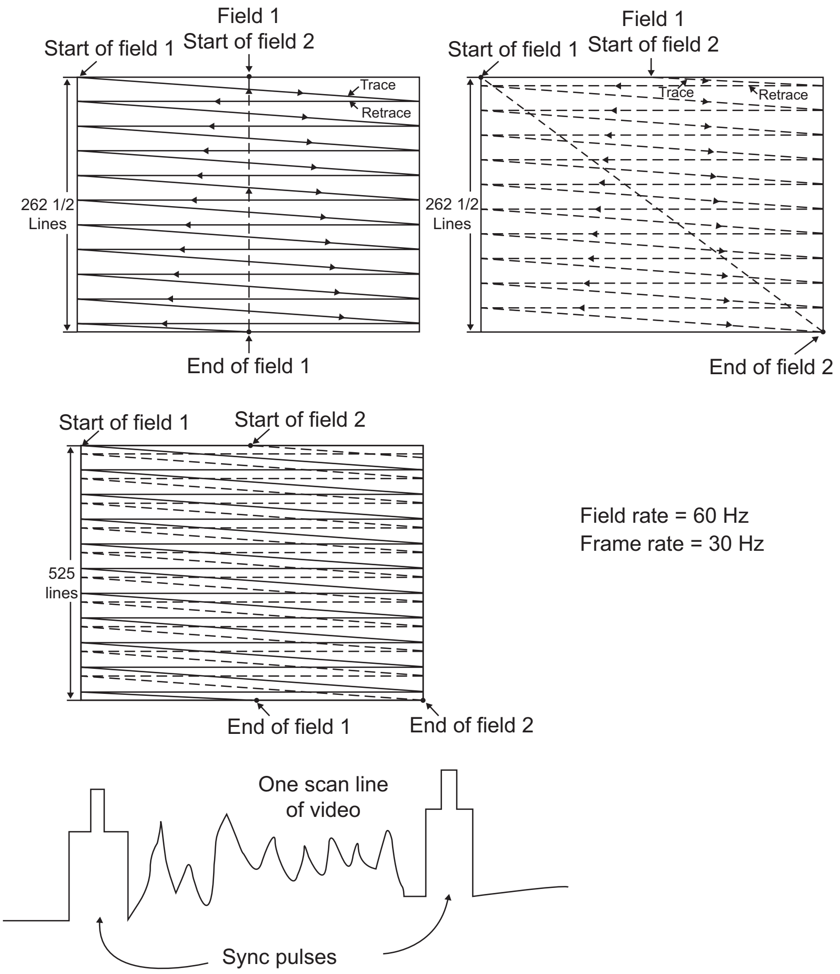
Field 1
Start of field 1
Start of field 2
Trace
Retrace
262 1/2 Lines
End of field 1
Field 1
Start of field 1
Start of field 2
Trace
Retrace
262 1/2 Lines
End of field 2
Start of field 1
Start of field 2
525 lines
End of field 1
End of field 2
Field rate = 60 Hz
Frame rate = 30 Hz
One scan line of video
Sync pulses

FIGURE 11.3 Interlaced scanning helps reduce flicker. Sync pulses help keep the receiver and transmitter in step with one another.

One thing that you may be wondering is how do you keep the TV receiver screen in step with the video camera or other video source? This is solved by creating synchronizing pulses that are transmitted before and after each scan line (see Fig. 11.3). After one field is scanned, a larger group of synchronizing pulses occurs between fields to cause the receiver to start the scanning back at the top again. The synchronizing pulses are transmitted or stored along with the analog video signal for each line. These synchronizing pulses are then used to trigger the receiver's circuitry to make sure that the scene is sequenced properly on the screen.

Other World TV Systems

The US NTSC analog system is used in Canada, Mexico, and Japan. But other analog systems are used throughout the world. In Europe, specifically the UK, Germany, Italy, and Spain, the PAL (phase alternation line) standard is used. It uses 625 scan lines with a frame rate of 25 frames per second. In France and in some other countries, a system called SECAM for Sequenteil Couleur a Memorie (or sequential chrominance and memory) is used. It too uses 625 scan lines and a 25-frame-per-second rate. PAL and SECAM systems have a better vertical resolution for picture detail because of the greater number of scan lines, giving a perceptibly improved picture over the NTSC system. Most of the other countries of the world have adopted one of these three systems. However, now the United States and most of Europe have digital systems.

Digital TV

Most TV and video now is transmitted and stored in digital or binary format. Scanning is still the process used to convert the scene into an electrical signal, in this case a stream of binary numbers representing the brightness and color detail. Each scan line is still essentially an analog signal but that is converted into a sequence of binary numbers by an analog-to-digital converter (ADC). The basic process is illustrated in Fig. 11.4. The ADC samples are measures of the analog signal at a constant rate, as shown by the dots along the curve in Fig. 11.4. Remember that to retain the information in an analog signal when it is converted into digital, the signal must be sampled at a rate at least twice as fast as the highest-frequency signal contained in the analog signal. Since video signals contain extremely high frequencies, the sampling rate is very fast. Most video contains scene details with frequencies beyond 4 MHz. That means a sampling rate of at least 8 MHz to capture the detail. In any case, each sample results in an 8-bit binary number. These bytes, one representing each of the sequential samples, are sent one after another in a serial binary data stream as shown. The video data can then be transmitted or stored.

Each of the binary numbers produced by sampling produces what we call a *pixel* or picture element. This particular binary number will ultimately translate to a specific analog voltage value and produce a specific brightness level along the scan line.

One way to visualize a digital video display is to think of it as an array, or horizontal scan lines, each made up of a sequence of pixels (see Fig. 11.5). Each pixel is a dot of light or, in this case, a square. That pixel is derived from a specific binary number developed during the scanning and sampling process. Each of those pixels represents a specific brightness level from white to black. A digital-to-analog converter (DAC) converts the binary brightness value back into a specific shade of gray.

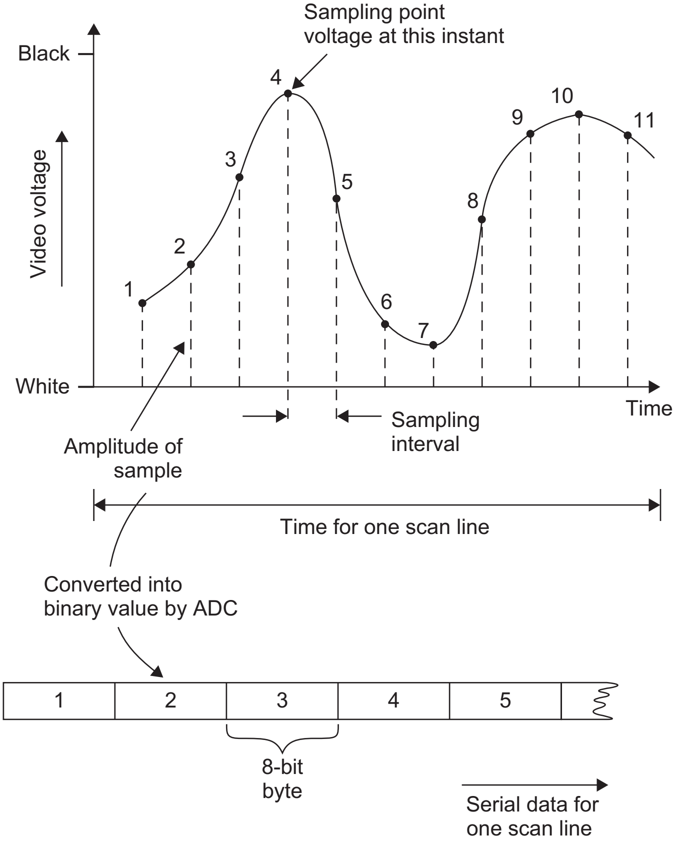
Sampling point
voltage at this instant
Black
Video voltage
White
1
2
3
4
5
6
7
8
9
10
11
Time
Sampling
interval
Amplitude of
sample
Time for one scan line
Converted into
binary value by ADC
1
2
3
4
5
8-bit
byte
Serial data for
one scan line

FIGURE 11.4 Digital TV is created by digitizing the analog video signals with an ADC, and then converting it into a serial bitstream of video data words.

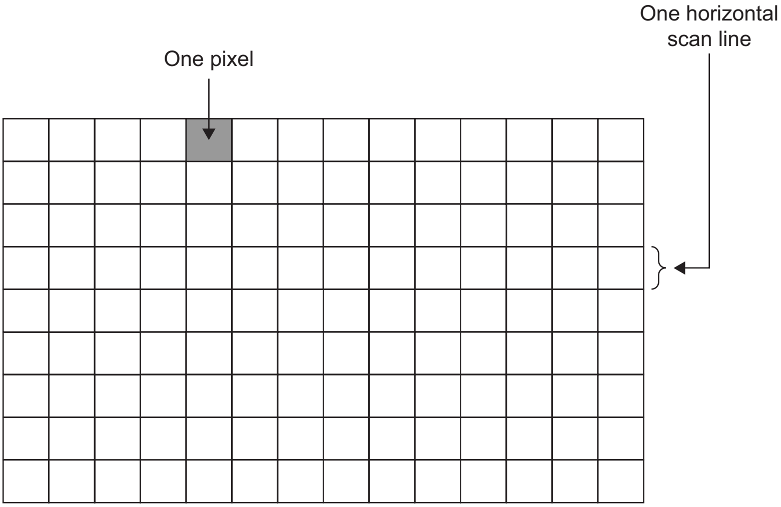
One pixel
One horizontal
scan line

FIGURE 11.5 TV screens can be thought of as thousands of rows and columns of pixels, each represented by data.

DTV uses the interlaced scanning principles described earlier to help minimize flicker. But it also uses a faster progressive scan. This is where each scan line making up the screen is scanned sequentially one after the other from top to bottom. To minimize flicker, *progressive scanning* must be done at a higher rate of speed than interlace scanning. Progressive scanning is used simply because it creates a sequence of digital numbers that is more conveniently stored and processed one after the other than in a two-step interlaced process.

DTV uses a variety of different formats. In the basic format, it produces 480 scan lines with 640 pixels per line. This is roughly the equivalent of what you would see with an analog TV signal. This is what we call the DTV standard resolution format. Both interlaced and progressive scanning formats can be used with rates of 24, 30, and 60 frames per second.

Picture Resolution

One thing that you should understand about video is the concept of resolution. This refers to how much fine detail you can see on the screen. In analog TV, brightness and color detail results in very-high-frequency analog signals. The older original analog TV was limited to frequencies of approximately 4.2 MHz. Modern HDTV uses much higher-frequency responses up to about 10 MHz. To see that picture detail, you must transmit the analog signal over a channel that has sufficient bandwidth. Most radio channels and cable systems have bandwidth limitations and essentially act as a low-pass filter that will remove some of the higher frequencies. While you will still see a picture, it will lack some of the fine detail that your eye may or may not notice.

Resolution as it relates to DTV is a combination of how many scan lines plus how many pixels per line are used. Obviously, the greater the number of scan lines and the greater the number of pixels, the more detail you are going to be able to transmit and ultimately see.

A higher-resolution format to produce what we call high-definition (HD) TV uses 720 scan lines with 1280 pixels per line. The HD format uses 1080 horizontal lines with 1920 pixels per line. Both of those HD formats can use either interlace or progressive scanning at rates of 30, 60, 120, and 240 frames per second. Just remember that the frame rate simply refers to the number of times per second that each frame is presented on the screen. The earlier slow frame rates of 30 and 60 frames per second often produce a kind of picture smearing when the scenes are changing at a high rate of speed. This is noticeable in action movies or in sports events. The newer 120- and 240-frames-per-second scan rates essentially eliminate this problem.

The newest form of DTV is called ultrahigh definition (UHD). Other designations are UHD-1, Quad TV, or 4K. The 4K is derived from the pixel format of 3840 × 2160 where there are almost 4000 (4K) lines with 2160 pixels each. The resolution is four times the standard HD resolution. The scan is progressive at 50 or 60 frames per second. The resolution is astonishing and fine detail can easily be seen. 4K TV sets have been available for a few years, but prices are still very high. 4K resolution content is not yet widely available but there is some 4K over-the-air (OTA) broadcasts and some 4K movies on DVD.

Another term you will hear associated with 4K is HDR or high dynamic range. HDR is a feature that offers better picture contrast meaning whiter whites, darker blacks, and a wider range of colors. Not all 4K TV sets have HDR but those that do deliver amazing picture quality.

Also now available are 8K screens with a resolution of 7680 lines with 4320 pixels per line. It is meant primarily for the very large screens (>85 in.). Frame rate is 60 Hz and 8K is usually accompanied by HDR.

Color Video Principles

Thus far, the signals we have described are what we call monochrome or B&W video. In *monochrome video*, all we capture and transmit and reproduce is the brightness variations along a scan line that range from black to white and an infinite number of shades of gray in between. But, of course, today most video is color. Capturing, transmitting, processing, and displaying color is a far more complex process than monochrome.

One physics principle that makes color video possible is the concept that any color of light can be reproduced by simply mixing the right combination of red, green, and blue light. Take a look at Fig. 11.6. Mixing red and green creates yellow. Mixing red and blue produces violet (magenta), while mixing green and blue produces the color cyan. In any case, as it turns out virtually, any color can be created using just the right mix and intensities of red, green, and blue.

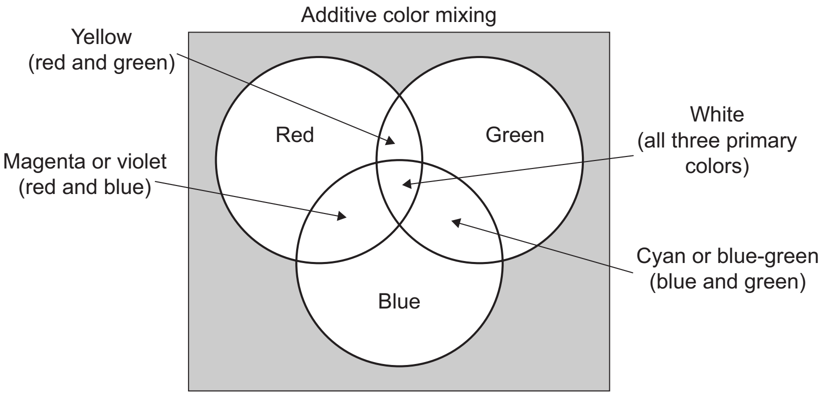

FIGURE 11.6 Any color of light can be reproduced by mixing red, green, and blue light in the right proportion.

Strangely enough, there is one unique combination of the three colors that creates white. Black is produced, of course, by just turning off all the light sources. The color TV problem is how do you capture the individual colors and then how do you reproduce them?

Video cameras use a light-sensitive imaging chip called a CMOS imager or charge-coupled device (CCD). An optical lens focuses the scene on the chip. The chip itself is made up of millions of tiny light-sensitive elements that are charged like a capacitor with a voltage depending on the brightness level reaching them. The analog video signal for a scan line is then created by sampling those light-sensitive elements one at a time horizontally. The result is an analog signal whose amplitude is proportional to the light level. The result is a B&W brightness or Y signal.

Fig. 11.7 shows how the color signals are produced. The scene is actually passed through a lens and a beam splitter that divides the scene into three identical segments. Each is passed through individual red, green, and blue filters. Three light-sensitive imaging chips then produce three separate analog signals for the red, green, and blue elements in a scene. The red, green, and blue analog signals are then individually amplified and sampled by an ADC to translate them into a binary sequence. The three binary streams are then transmitted serially as a sequence of red, green, and blue for each picture element on the imaging chip. Each of these RGB colors will then later form a pixel on the viewing screen.

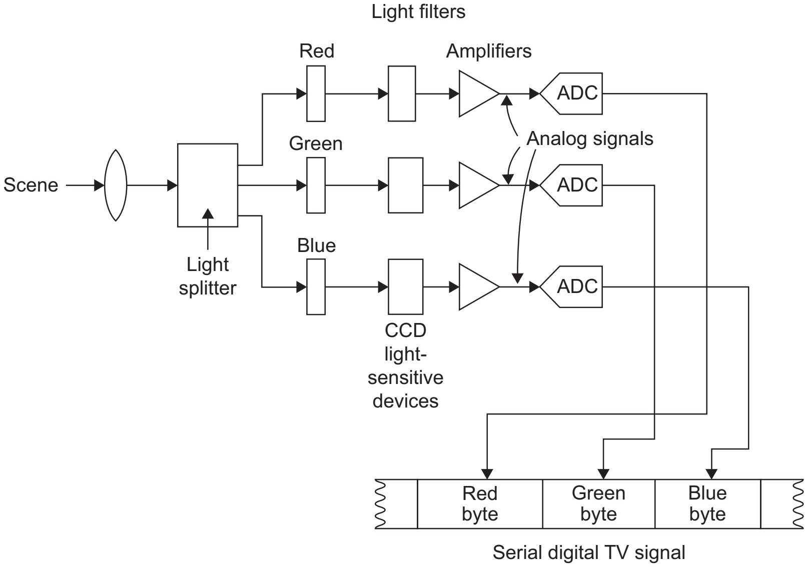

FIGURE 11.7 Red, green, and blue digital color signals are generated by filtering the scene into its red, green, and blue components, and then digitizing the resulting analog signals.

Video Compression

Generating a color video signal produces a serial bitstream that varies at a rate of hundreds of millions of bits per second. Consider these calculations:

- For each pixel, there are three color signals of 8 bits each making a total of 8 × 3 = 24 bits per pixel.
- For a standard DTV screen, there are 480 lines of 640 pixels for a total of 307,200 pixels per frame or screen.
- For one full screen or frame, there are 307,200 × 24 = 7,372,800 bits.
- If we want to display 60 frames per second, then the transmission rate is 7,372,800 × 60 = 442,368,000 bits per second, or 442.368 Mbps. That is extremely high frequency.
- To store one frame of video of this format, you would need a memory with 55,296,000 bytes or 55.296 MB. And that is just one frame.
- To store 1 min of video you would need a memory of 3,317,760,000 bytes or just over 3.3 gigabytes (GB). And that is just 1 min. Multiply by 60 to get the amount of memory for an hour.

Imagine the speeds and memory requirements for 1080p or a 120 frames-per-second format. And we have not even factored in the stereo audio digital signals that go along with this. Anyway, you are probably getting the picture here. First, data rates of 442.368 Mbps are not impossible, but they are impractical as they require too much bandwidth over the air and on a cable. And they are expensive. Furthermore, while memory devices such as CDs and DVDs are available to store huge amounts of data, they are still not capable of those figures. Therefore, when it comes to transmitting and storing digital video information, some technique must be used to reduce the speed of that digital signal and the amount of data that it produces.

This problem is handled by a digital technique known as *compression*. Digital compression is essentially a mathematical algorithm that takes the individual color pixel binary numbers and processes them in such a way as to reduce the total number of bits representing the color information. The whole compression process is way beyond the scope of this book, but suffice it to say that it is a technique that works well and produces a bitstream at a much lower rate. And, the compressed video will take up less storage in a computer memory chip.

The digital compression technique used in DTV is known as MPEG-2. MPEG refers to the Motion Picture Experts Group, an organization that develops video compression and other video standards. Another standard is MPEG-4 AVC. For 4K and 8K a compression method called high efficiency video coding is used.

What you have to think about when you consider the transmission of digital video from one place to another is that it is typically accompanied by audio. The audio is also in serial digital format. Those digital words will be transmitted along with the video digital words to create a complete TV signal.

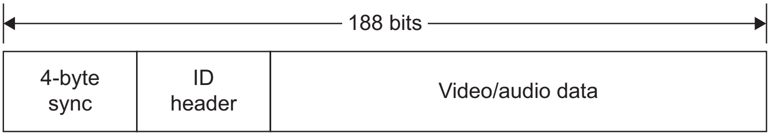

FIGURE 11.8 HDTV packet for ATSC US digital color TV system.

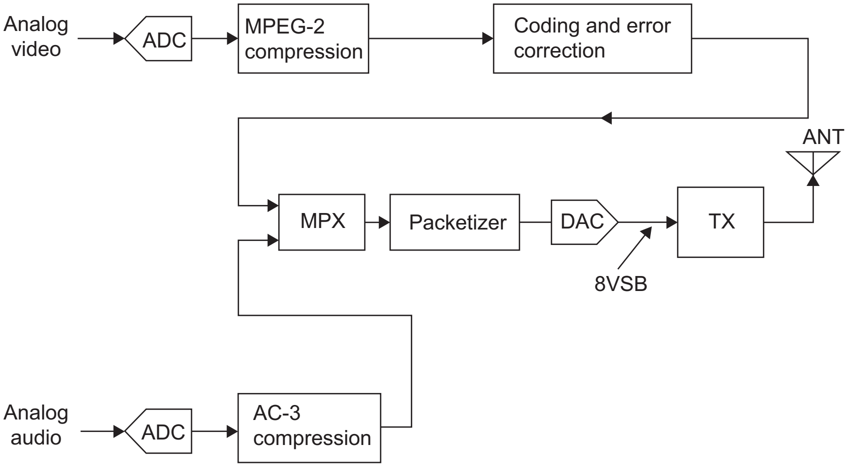

FIGURE 11.9 General block diagram of TV transmitter.

TV Transmission

The compressed audio and video of a TV signal are assembled into packets or frames that look like that in Fig. 11.8. The 4 bytes up front provide synchronizing bits for the receiver. The ID header tells the packet number, sequence, and video format used. The remaining bits are the serial compressed video and audio.

Fig. 11.9 shows a simplified TV transmitter. The serial digital video is put through the MPEG-2 compression process. This is followed by additional bit manipulation and coding for error correction and improved signal recovery by the receiver. This serial data is then multiplexed with the serial compressed audio data. Then the packets are formed. This is what will modulate the transmitter.

The modulation begins by translating the fast serial bitstream into an eight-level signal with a DAC. Each 3-bit segment of the serial data is assigned a specific voltage level as shown in Fig. 11.10. Since each level represents 3 bits, you can transmit more bits per second in the same time period. It is this eight-level signal that modulates the transmitter carrier. The modulation is AM, where all of the upper sidebands are transmitted but only part of the lower sidebands. This saves spectrum. The modulation is referred to as 8VSB for *vestigial sideband*, where vestigial refers to part of the lower sideband. The signal is then upconverted to the final frequency by a mixer. A high-power amplifier in the transmitter sends the signal via coax cable to the antenna. Typical TV station

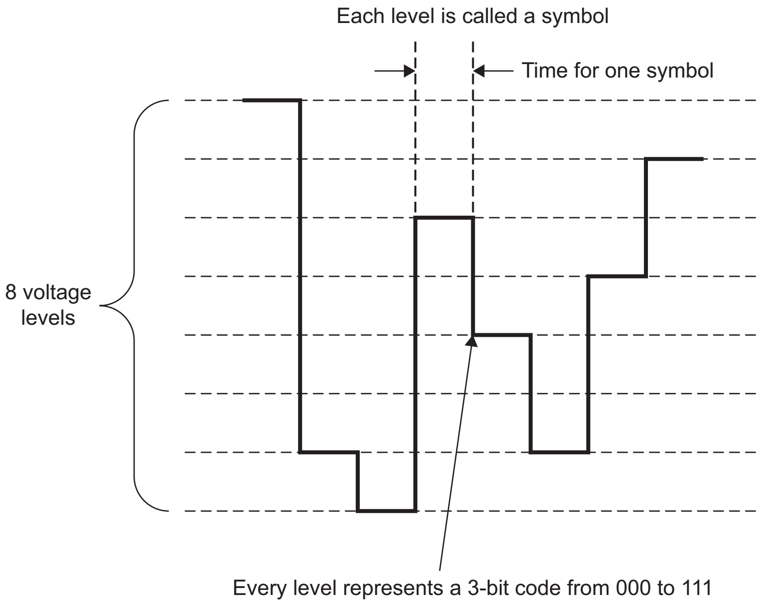
Each level is called a symbol
Time for one symbol
8 voltage levels
Every level represents a 3-bit code from 000 to 111

FIGURE 11.10 US ATSC system translates the digital bitstream into an eight-level signal to transmit multiple bits per level, making higher data rates in a given bandwidth possible.

range is about 70 miles maximum, and that is determined by tower height and terrain. The standard for OTA DTV in the United States was developed by the Advanced Television Standards Committee (ATSC). A new standard for 4K UHD TV called ATSC 3.0 using orthogonal frequency division multiplex (OFDM) instead of 8VSB is forthcoming.

In the United States, each TV channel is 6 MHz wide. There are 50 channels starting at channel 2 from 54 to 60 MHz. Channels 3 through 6 range from 60 to 88 MHz. Channels 7 through 13 run from 174 to 216 MHz. Channels 14 through 51 run from 470 MHz and ending with channel 51 from 692 to 698 MHz. In the United States the 600–700 MHz channels are no longer used for TV and are being repurposed for cell phone use and wireless broadband services.

World Digital TV Standards

The US DTV standard from ATSC has also been adopted by Canada, Mexico, South Korea, and a few others. However, the rest of the world uses different digital formats. The UK, Europe, Australia, and New Zealand use a standard called Digital Video Broadcasting (DVB). Another is the Integrated Services Digital Broadcasting (ISDB) standard used in Japan, South America, and some African countries. The Digital Terrestrial Multimedia Broadcasting (DTMB) standard is used in China. What all these non-US standards have in common is the use of OFDM modulation that has proven superior in OTA transmissions. The United States plans to adopt OFDM in its forthcoming ATSC 3.0 standard for 4K.

TV Receiver

Fig. 11.11 shows a general block diagram of a modern DTV receiver. The signal from the antenna or cable TV box is sent to the tuner where the signal is amplified and sent to a mixer along with a local oscillator signal to downconvert the signal to a lower frequency called the *intermediate frequency* (IF). This is commonly 44 MHz. The local oscillator is a frequency synthesizer that selects the desired channel, usually with a remote control. The IF signal is filtered to select only the desired 6-MHz channel.

The IF is then sent to an ADC for digitizing. The 8VSB signal is then recovered in the demodulator. It is then translated into the original bitstream, and any coding or error correction is removed. The signal is then demultiplexed into the video and audio streams. The video goes to an MPEG-2 decompression circuit for recovery of the original video. The individual RGB components are sent to DACs to recover the picture detail. The RGB signals are then amplified and sent to the screen for display. The audio is also decompressed and sent to the DACs and amplifiers and speakers.

The whole TV set looks something like Fig. 11.12. Note all the possible inputs. The set gets its RF inputs from the antenna or cable box. The cable box may also have a digital signal output, meaning the cable box does all the decoding, demultiplexing, and decompression. The output comprises a serial data stream of the video and audio. A special high-speed digital interface called the *high-definition media interface* (HDMI) is used. It uses a special cable no more than 15 ft long and with a special 19-pin connector. The HDMI carries video and audio from the TV to cable box or DVD player or vice versa.

Some TV sets also accept analog video inputs from other sources such as VCRs, video cameras, and digital cameras. A special fiber optic cable input called TOSLINK serves as the interface for TV sets, cable boxes, and DVD players.

Finally, some TV sets include standard computer interfaces such as USB or Ethernet ports. The so-called Smart TV sets are configured for Internet access from a modem by way of the Ethernet port or a built-in Wi-Fi wireless capability.

TV set outputs, of course, are sent to the screen and the speakers. A remote control sends infrared (IR) or RF wireless signals to an embedded controller that tells the set what to do.

TV Screen Technology

As indicated earlier, there are several ways that the video signal is translated back into a viewable picture. The CRT used in older sets is gone. LCD screens are by far the most popular and most widely used today. Plasma screens and DLP projection screens have mostly disappeared. There are other emerging technologies that we may see in the future. In any case, what we need is a light source to reproduce the varying intensities of red, green, and blue light.

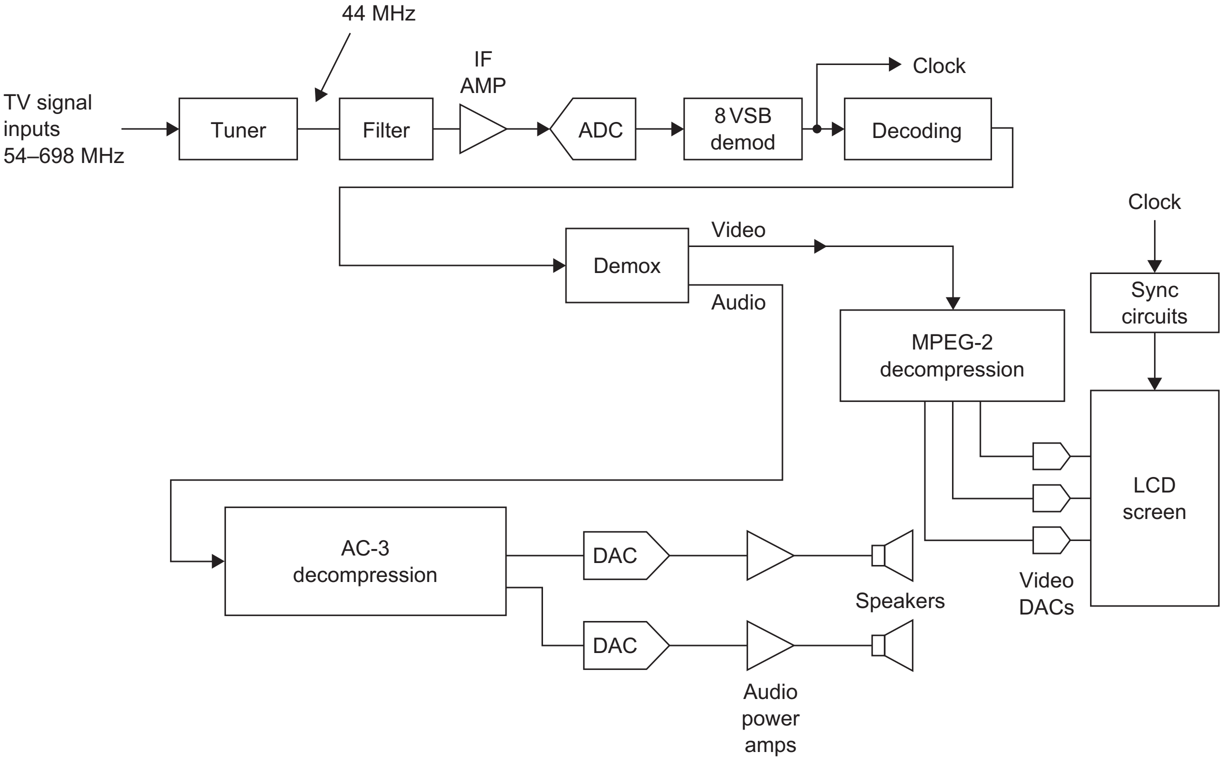
44 MHz
IF
AMP
Clock
TV signal
inputs
54–698 MHz
Tuner
Filter
ADC
8 VSB
demod
Decoding
Clock
Video
Demox
Audio
Sync
circuits
MPEG-2
decompression
LCD
screen
AC-3
decompression
DAC
DAC
Speakers
Video
DACs
Audio
power
amps

FIGURE 11.11 Simplified block diagram of digital TV set.

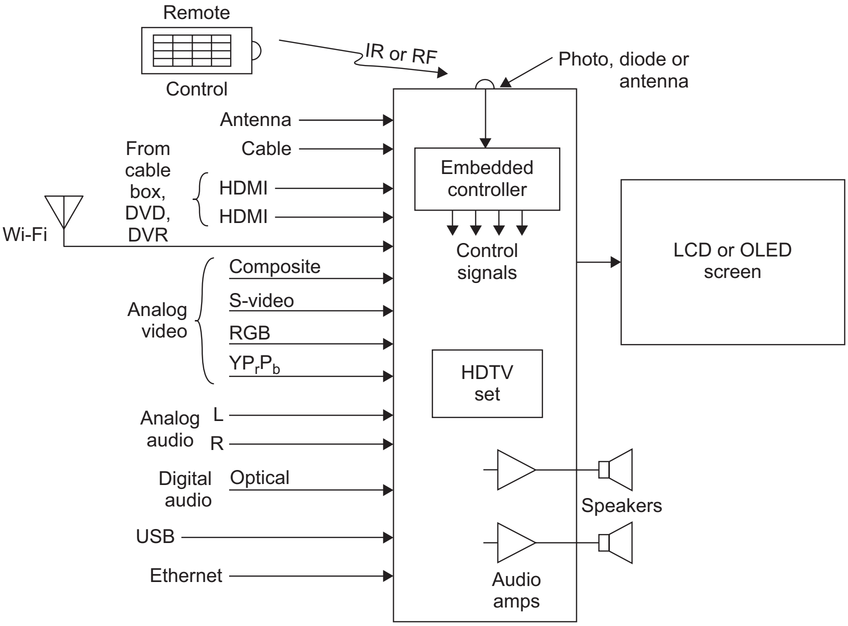
Remote
Control
IR or RF
Photo, diode or antenna
Antenna
Cable
From cable box, DVD, DVR
HDMI
HDMI
Wi-Fi
Embedded controller
Control signals
LCD or OLED screen
Analog video
Composite
S-video
RGB
YP_rP_b
HDTV set
Analog audio
L
R
Digital audio
Optical
Speakers
USB
Ethernet
Audio amps

FIGURE 11.12 Inputs and outputs of modern digital TV set.

Color video screens are made up of a very large number of tiny red, green, and blue light sources. These are used to make up the pixels or dots of light that recreate the picture. These pixels are very small, so that your eye cannot actually see them even when you are relatively close to the screen.

Fig. 11.13 shows one example of a screen pixel format in use today. LCD and LED screens use a pixel format somewhat like that shown in Fig. 11.13. These are usually rectangular red, green, and blue areas. They are LEDs that produce their own light or are LCDs backlighted with the appropriate degree of intensity.

One pixel as displayed on the screen might actually be made up of a collection of these three color groups. The actual pixel size may be something like Fig. 11.13, made up of multiple groupings of red, blue, and green elements as shown.

The actual format displayed on the screen is determined by the TV display standards. As indicated earlier, a standard DTV screen uses 480 horizontal lines with 640 pixels per line. That creates a screen with a total of 307,200 pixels. The 720 × 1280 screen puts 1280 pixels on 720 lines for a total of 921,600 pixels. HD screens use 1080 lines with 1920 pixels per line for a total of 2,073,600 pixels. Just remember that the pixels are turned on one at a time from left to right and top to bottom on the screen at a very high rate of speed.

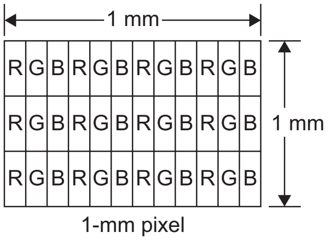

FIGURE 11.13 How color pixels are formed on a TV screen. Pixel format typical of LCD or LED screens.

Summary of Screen Technologies

The actual technical details of each of the different types of screens are way beyond the scope of this book, but a summary might be useful. A quickie review of each of these screen types follows.

CRT—No longer used in modern TV sets.

LCD—Liquid crystals are, as their name implies, a liquid made up of molecules whose physical orientation can be changed by the application of an electric signal. Different voltages change the polarization of the light through the liquid. By applying one level of voltage to the liquid crystal, it can be made almost totally transparent where maximum light passes through it. Another level of voltage will twist the molecules and practically block all light from being passed. In this way, the liquid crystal acts almost like a camera shutter but in this case with varying degrees of transparency being possible. The liquid crystal material is contained within glass plates. A bright light source, such as a fluorescent lamp or high-brightness LEDs, is placed behind the liquid crystal. In front of the liquid crystal is another transparent sheet containing the red, blue, and green pixels described earlier. The whole thing is sandwiched together in a very thin screen. The electrical signals applied to a matrix of wires cause the liquid crystals to be properly oriented for each pixel and to allow the correct proportion of light through to each of the red, blue, and green light filters.

Plasma—Plasma screens are no longer popular.

DLP—DLP TV sets are no longer available.

OLED—The newest TV screens are made of tiny organic red, green, and blue LEDs arranged in a pixel matrix as described earlier. The base is an electroluminescent layer of film made of an organic compound. The organic or OLEDs omit their own light so no backlight is needed. They typically have a wider contrast for an improved picture. A newer enhanced version is called quantum dot technology. QLEDs as they are called give brighter colors. The OLED screens are the brightest available and respond fast to action scenes. However, they are the most expensive screens.

3D TV

Yes, you actually can buy a 3D TV set that gives you that same 3D "coming-at-you" feeling of theater 3D. These sets are very expensive, and there is very little 3D material to see. Right now it is only available on DVD. And, yes, you still have to wear the special glasses. These glasses are not the red and green lens type but a special electronic version. The lenses use electrical polarization like a shutter that turns the left and right lenses off and on in synchronization with the two alternating sets of frames on the screen. A wireless signal (IR or radio) signals the glasses when to switch. The switching occurs at very high speed so the effect is to see the 3D scene. 3D TV has not been widely accepted and for now will remain a gimmick or niche TV product.

Which Screen Type Is Best?

LCD screens are available in a wide range of sizes, but the minimum today for a TV set is about 19 in. Remember that these dimensions are the diagonal dimension from an upper corner of the screen to the lower diagonal corner. A popular LCD screen size is 32 in., but they are available in a wide range of sizes such as 42, 55, 65, 85, and 105 in.

The ability of a screen to produce high-speed motion is generally good on LCD screens. However, some early LCDs did produce some smearing. Recently the higher scan rates of 120 and 240 frames per second have essentially eliminated this problem.

LCD screens dominate TV today as their price continues to drop with better manufacturing methods. LED and organic LED (OLED) screens are growing in popularity and will become more widely used as prices come down in the future. Even curved screens are now available but expensive and not popular as they offer no improvement in picture quality and suffer from a narrower angle of view.

CABLE TELEVISION

It is estimated that about 15% of all US citizens get their TV over the air with an antenna. But over 80% get TV via some pay TV service. Roughly 60% get TV from cable. While the remaining 20% get satellite TV. Why? Mainly because generally cable offers not only better, more reliable signals than by antenna, but also offers hundreds of channels of programming. But that is not all. Most of those who have cable also use the cable connection for high-speed Internet access service. Of course, you pay for all that, and it has become like a monthly utility bill. With the cable industry being so big, it pays to know a little more about it.

It all starts with a facility called the *cable head end*. Another name you might hear is cable modem termination system, which specifically is the

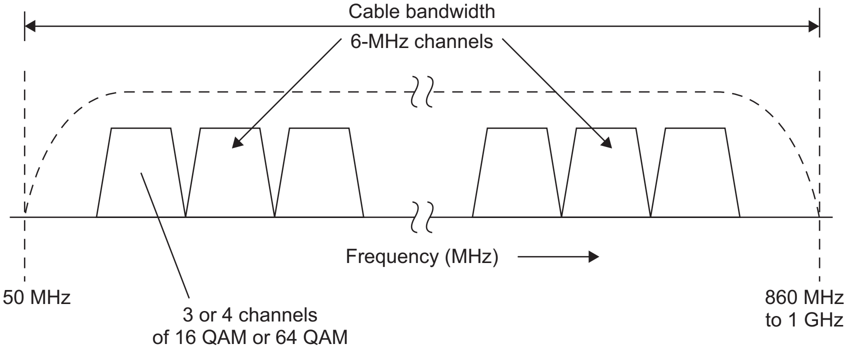

FIGURE 11.14 Coax cable bandwidth up to 1 GHz can contain about 130 6-MHz channels.

equipment at the head end. This is a place where the company collects video and programming from many different sources. It snatches signals off the air, it receives premium programming such as HBO, ESPN, Showtime, and Fox via satellites, stores movies on its large computers called video servers, and may even do some of its own local programming. The result is a hundred channels or more.

Now each of those TV signals—video plus audio—is modulated onto carriers that are spaced 6 MHz apart just like the regular OTA TV channels. Then all of those modulated carriers are added together to form a very complex signal. That signal is then amplified and put on the outgoing cable. Fig. 11.14 shows what the spectrum looks like. Most cables have a bandwidth from about 50 MHz up to about 860 MHz–1 GHz, which will accommodate up to about 130 6-MHz channels. We call this process frequency-division multiplexing, where a single cable can carry many separate channels of data, each on a different frequency and bandwidth. And since the cable is shielded, it does not radiate, so it, in effect, duplicates the free-space electromagnetic spectrum.

Since TV today is digital, digital modulation methods are used to put multiple TV signals within each 6-MHz channel. Using 16-QAM or 64-QAM, a cable company can squeeze several channels of video/audio into each 6-MHz chunk of spectrum, providing hundreds of additional slots for programming. That is why cable companies can advertise up to 500 channels of entertainment, news, and so on.

All of the digital video/audio signals are then remodulated onto different channel assignments. All of these 6-MHz channels are added together or linearly mixed in a multiplexer (MUX) to create a massively complex composite signal. This is what goes out on the cable (see Fig. 11.15).

The signal distribution network is called a *hybrid fiber cable* system. This multiplexed signal is applied to a fiber optic cable, called a *trunk cable*. It is routed around the city, and it terminates at selected points where the signal is divided into multiple paths of coax cable. These so-called *feeder cables* are

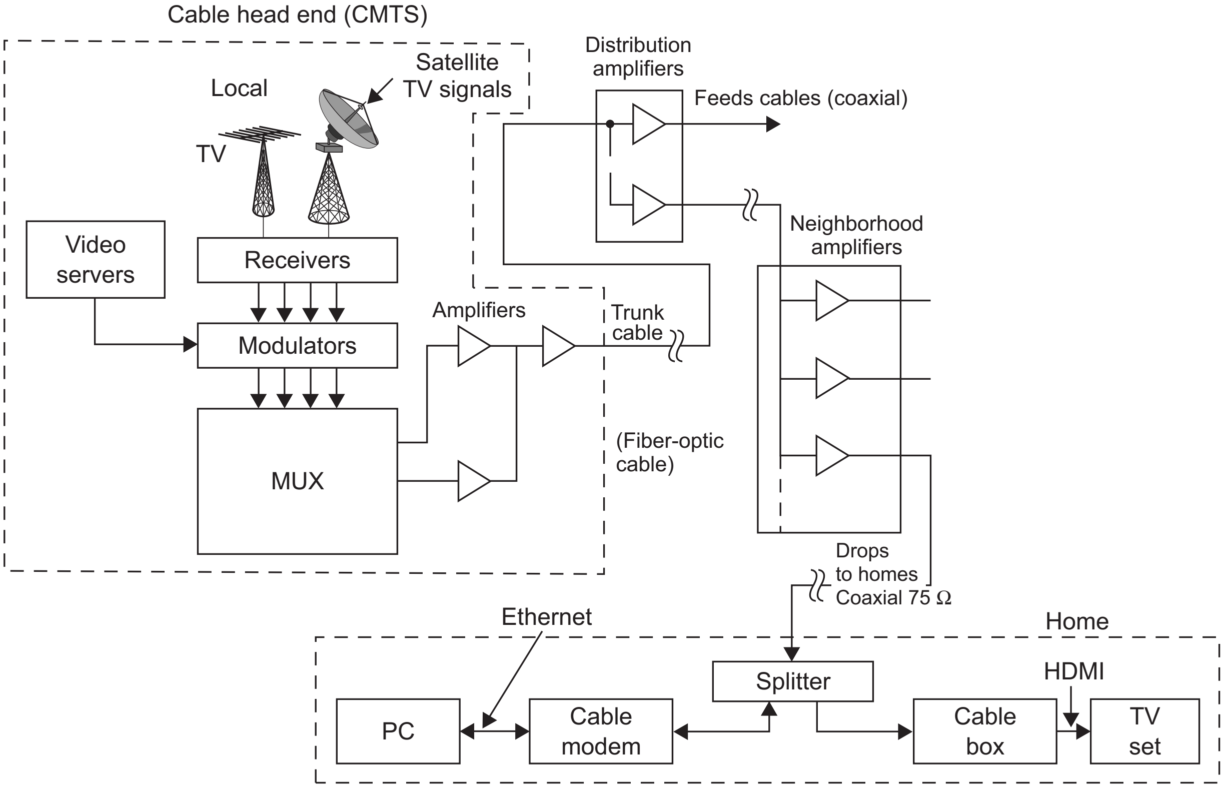
Cable head end (CMTS)
Local
TV
Satellite
TV signals
Video servers
Receivers
Modulators
MUX
Amplifiers
Trunk cable
(Fiber-optic cable)
Distribution amplifiers
Feeds cables (coaxial)
Neighborhood amplifiers
Drops to homes
Coaxial 75 Ω
Ethernet
Home
HDMI
PC
Cable modem
Splitter
Cable box
TV set

FIGURE 11.15 General block diagram of modern cable TV system.

connected to amplifiers along the way to overcome the large attenuation of the cables. At some point, the feeder cables come into your neighborhood and are further amplified and split into multiple signals. These signals are then sent to your house by a coax drop cable. The most common type is 75-Ω RG-6/U coax with F connectors. The cable comes into your home and is usually split two or more times and sent to wall connectors in the various rooms. These connectors connect the signal to a cable box and/or a cable modem for Internet access to a PC or a Wi-Fi router.

The cable box is like a tuner in that its main job is to select one of the many channels and send it to your TV set. Fig. 11.16 shows a simplified block diagram of a typical cable box, usually called a set-top box (STB). The input is a TV tuner with channel selection via a frequency synthesizer local oscillator by way of your remote control. New STBs feature two or more tuners so that you can watch one channel and record another. The selected signal is then demodulated and decompressed by a digital media processor chip and then sent to the TV set via one of the many possible interfaces, usually the HDMI or *high-definition multimedia interface*. This is a very fast digital interface with special cable and connector. Most TV sets have this as the most common input. Other video format interfaces, such as RGB, PrPbPy, or S-video, are usually provided to accommodate older analog TV accessories. An additional feature of many of the newer STBs is a built-in digital video recorder (DVR). This is essentially a computer hard disc drive that records the program you want to save for viewing later.

Other Wired TV Distribution Systems

There are a couple of other ways that digital video is distributed to homes. One is a fully fiber optic system. The fiber cables go all the way from the head end to the homes. This was once too expensive, but today, thanks to improved fiber and other technologies, it is cost effective. Such systems are generally referred to as *passive optical networks* (PONs). "Passive" simply means that the signals are strong enough and the cable runs short enough so that no amplification is needed along the way. The digital video and audio are modulated onto IR light beams and sent down the cable. Because fiber has such wide bandwidth, data rates are extremely high, over 1 GB/s in the newer systems. This allows not only digital video to be transmitted but also very-high-speed Internet service. Such systems use standards such as Ethernet PON or GPON for gigabit PON. Typical US systems are Google Fiber and Verizon Fios.

TV is also distributed over digital subscriber lines (DSLs) (Chapter 9). The signals travel via fiber to the neighborhoods. From there they are split off, and the signals travel via in-place twisted-pair telephone lines. New forms of modems using a form of OFDM called *discrete multitone* transmit the signal over the twisted pair at high speeds. These lines are now used just for high-speed Internet service and are referred to as *asynchronous digital subscriber lines* (ADSLs or just DSLs). Newer versions called ADSL2 and ADSL2+ or VDSL for video DSL provide data rates from 26 to 52 Mbps on standard phone lines. Newer systems use an advanced form of DSL called G-fast that can achieve an aggregate of 1 Gbps on standard installed lines.

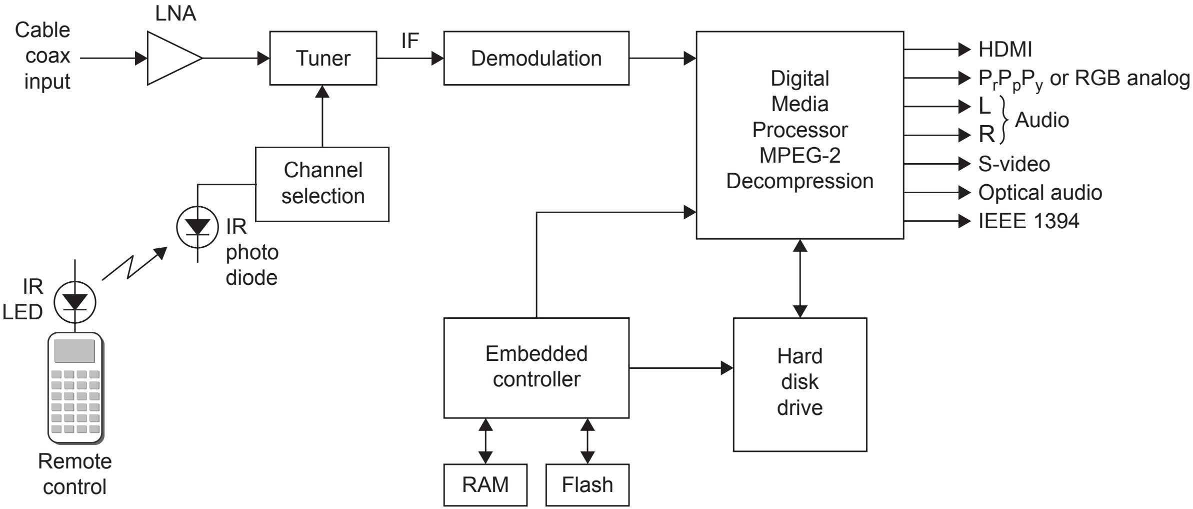
Cable coax input
LNA
Tuner
IF
Demodulation
Digital Media Processor MPEG-2 Decompression
HDMI
$P_rP_pP_y$ or RGB analog
L
R
Audio
S-video
Optical audio
IEEE 1394
Channel selection
IR photo diode
IR LED
Remote control
Embedded controller
Hard disk drive
RAM
Flash

FIGURE 11.16 Simplified block diagram of modern cable box with DVR.

For Internet access, you connect a cable modem to the coax cable, and it automatically selects the channel where the data services are assigned. The modem goes to your computer or a Wi-Fi router via an Ethernet cable.

SATELLITE TV

Satellite TV technically is also OTA TV or radio, but instead of the signals coming from one or more local TV stations over broadcast towers, the signals emanate from a geosynchronous satellite 22,300 miles away orbiting around the equator (see Fig. 11.17). The satellite transponders operate in the Ku band with frequencies of 10.95–14.5 GHz. This system is referred to as *direct broadcast satellite*.

The satellite TV company gets its TV like the cable companies and brings it all together at a single Earth station head end where it is all collected and transmitted on an uplink to the satellite in the 14–14.5 GHz range (see Fig. 11.18). The signals are then amplified in the satellite and rebroadcast on the downlink frequencies at 10.95–12.75 GHz. A high-gain antenna focuses the signals on the United States or other country of interest.

Down on Earth, satellite TV subscribers have a special receiver and antenna. These are illustrated in Fig. 11.19. The antenna is usually a small dish or parabolic reflector antenna about 18 in. in diameter. The antenna is usually a small horn. Also inside the antenna assembly is a low-noise amplifier (LNA) to boost the small signal from the satellite. This is then sent to a mixer along with a local oscillator signal to downconvert the signals into the 950–2100 MHz range. This signal is then routed by coax to the satellite receiver. The LNA/mixer assembly at the antenna is called a low-noise block converter. This receiver signal processing at the antenna is necessary because the Ku band frequencies of 10.95–12.75 GHz are so high they would be far too greatly attenuated by the antenna

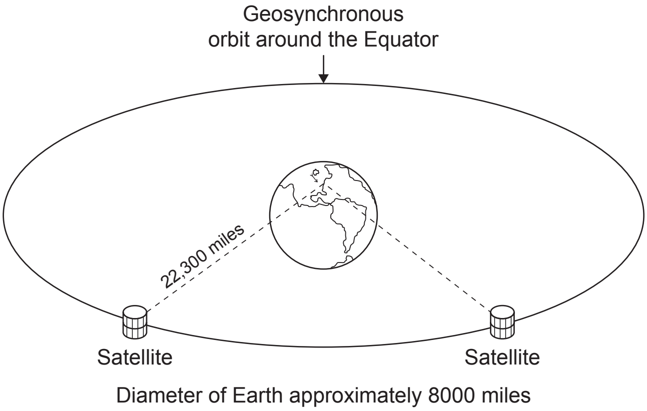

FIGURE 11.17 TV distribution satellites in geosynchronous orbits act as wireless relay stations.

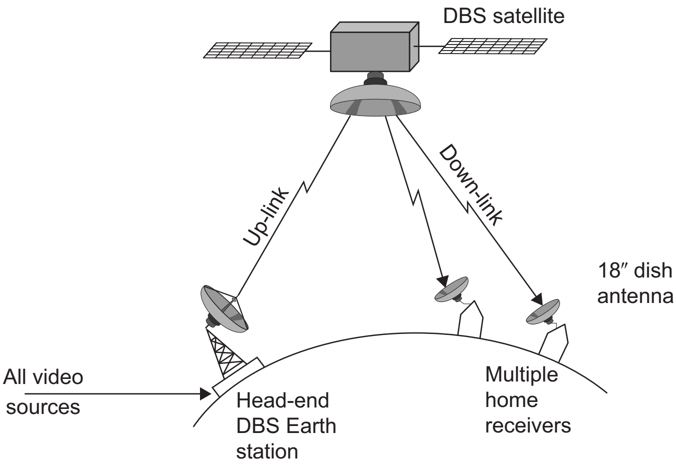
DBS satellite
Up-link
Down-link
18″ dish
antenna
All video
sources
Head-end
DBS Earth
station
Multiple
home
receivers

FIGURE 11.18 Video from all sources is transmitted uplink to satellite that rebroadcasts it back to Earth over downlink to home receivers.

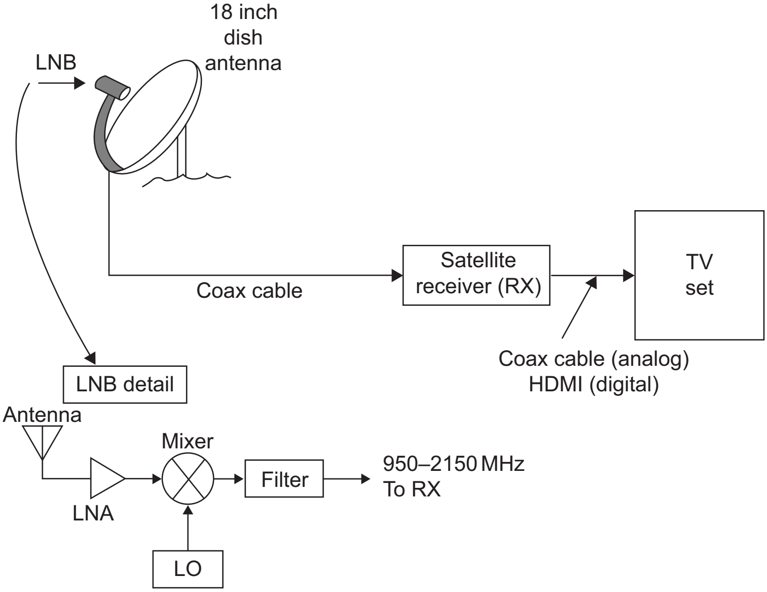
18 inch
dish
antenna
LNB
Coax cable
Satellite
receiver (RX)
TV
set
Coax cable (analog)
HDMI (digital)
LNB detail
Antenna
Mixer
LNA
Filter
LO
950–2150 MHz
To RX

FIGURE 11.19 Satellite TV home system consists of antenna with low-noise block converter, receiver, and TV set.

coax cable to the receiver. So downconverting to the lower frequency allows coax to be used to the receiver.

The satellite receiver is like the cable box in that it is used to select the desired channel and demodulate it for the TV set. The satellite signals are all digital and transmitted in packets 147 bytes long. A satellite receiver is shown

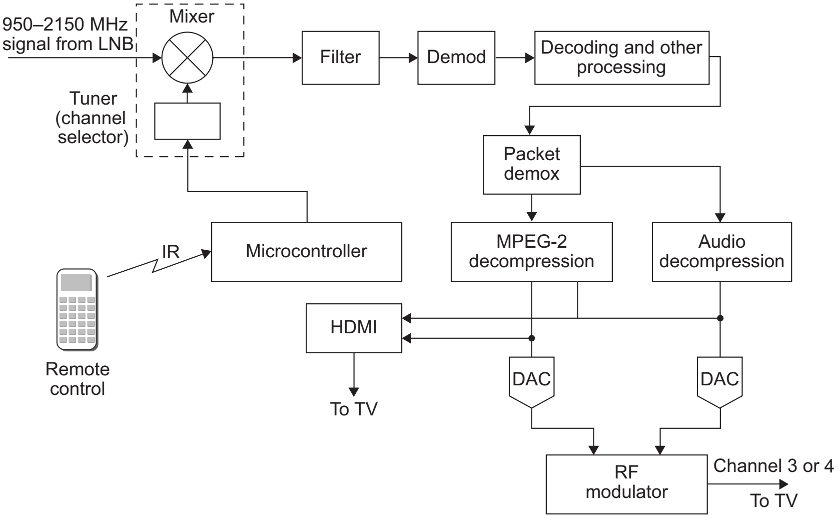
950–2150 MHz signal from LNB
Mixer
Tuner (channel selector)
Filter
Demod
Decoding and other processing
Packet demox
IR
Microcontroller
MPEG-2 decompression
Audio decompression
Remote control
HDMI
To TV
DAC
DAC
RF modulator
Channel 3 or 4
To TV

FIGURE 11.20 Simplified block diagram of satellite TV receiver.

in Fig. 11.20. MPEG-2 video decompression is used. The recovered signals are then sent to a DTV receiver, typically via an HDMI port.

OVER-THE-TOP TV

Over-the-top (OTT) TV is a video that streams the DTV signal to you over the Internet. In the United States some of the popular sources are Amazon, Hulu, and Netflix. The signal comes to you by your cable or DSL service usually to a Wi-Fi router. The router delivers the signal wirelessly to an OTT box or converter. Some common boxes are from Apple TV, Roku, or Amazon. The box is then connected to your TV set by and HDMI cable. Smart TVs have the Wi-Fi built in.

Streaming means that the compressed video comes to you in real time assuming your Internet service is fast enough. Streaming is video as it is occurring in contrast to downloading a video and storing it for play later. Your OTT box usually buffers or stores the video stream and feeds it continuously to the display.

Movies are the most common content but also available are older TV series, documentaries, and travel videos. Audio can also be streamed this way from services such as Spotify.

CELL PHONE TV

You can watch TV on your smartphone and you probably already do. Short video clips come via text or email. With Internet access you can also see TV from a mix of sources such as YouTube. If you have an OTT service such as

Amazon, Hulu, or Netflix, you can watch TV series and movies by way of a Wi-Fi connection to your home router and OTT provider box (Roku, Apple TV, etc.) The picture is small but of high quality.

CLOSED-CIRCUIT TV

Closed-circuit TV (CCTV) is used primarily for surveillance. You have seen the small wall- or ceiling-mounted TV cameras in stores, airports, companies, and today on the streets. The video signals are routed to a central point where multiple TV sets or monitors are set up for observation.

In many installations, the video is transmitted directly over coax cable to the monitors. There is no modulation involved. However, in some cases, when remote cameras are used, a wireless link via 2.4 GHz is typical. Of course, modulation is then used in the transmission so that a receiver is needed at the viewing end to recover the video.

Many installations also have video recorders (VHS tape or DVD) to capture several minutes or hours of video. In the past, most cameras were black and white, but today more and more use color and high definition.

DIGITAL VIDEO DISCS

The formal meaning of DVD is *digital versatile disc*. Its main use is video but it can also store computer data or audio. It is the storage medium of choice for movies, music videos, and other video programming. It has virtually replaced the VCR.

The basic DVD storage medium is a disc that is almost identical to the CD discussed in Chapter 10. The disc is the same size (120-mm diameter), but it is formatted differently to store many more bits and bytes of data. While the CD can store up to about 700 MB, the DVD can store about 4.7 GB, that is, gigabytes or billions of bytes. And that is just the basic one-sided disc. There are also single-side/double-layer discs, double-sided/single-layer, and double-sided/double-layer versions that can store even more, up to 15.9 GB.

The storage method is the same. A laser beam burns pits into a spiral track. The data is represented as pits and lands as described earlier in the CD discussion (see Figure 10.6). The big difference is that the size of the pits is significantly smaller and the track spacing is less than half the CD spacing, that is, 1.6 μm (1600 nm) versus 740 nm. A nanometer is 1-billionth of a meter. With smaller pits and tighter spacing, the disc can hold significantly more data.

But that is not the whole story. The video is also compressed. Unlike the audio on a CD, which is not compressed, digital video is compressed with the MPEG-2 standard described earlier. This gives roughly a 40-to-1 compression. In that way, a full 2 h of video can be stored.

As for audio, the format is as described in Chapter 10 for CDs, but DVD audio is Digital Dolby Surround Sound 5.1 using six channels of 16-bit, 44.1-kHz digital sound. Audio DVDs are also available with 192-kHz sampling of the audio with 24-bit ADCs, giving even greater fidelity and low-noise performance.

How a DVD Player Works

Fig. 11.21 depicts the block diagram of a typical DVD player. First, note the motor that drives the disc. It is a variable-speed device that changes speeds as the pickup mechanism moves from inside to outside as it reads the spiral track. The pickup head uses a laser that shines on the disc to read the pits and lands. Because of the closer spacing, a higher-frequency (shorter wavelength) laser is used. CD players use a 780-nm laser, while DVD players use a 640-nm laser that has a smaller focused beam.

The photodetector in the head picks up the laser reflections and sends the 10 Mb/s data stream to the processing chips. In some players, a single-chip application-specific device handles all of the digital processing from demultiplexing, to error detection and correction, decoding, and MPEG-2 decompression. All this may be in separate chips in some players.

In any case, the MPEG-2 data goes to a video decoder and to multiple DACs that recreate the original analog RGB signals. Various video formats are used, such as RGB, YCrCb, and YPrPb, each of which needs three DACs each. Most DVD players also decode regular composite video and S-video.

The separated audio signals go to an audio codec that contains the DACs that recover the audio content either in basic stereo or 5.1 surround sound. As

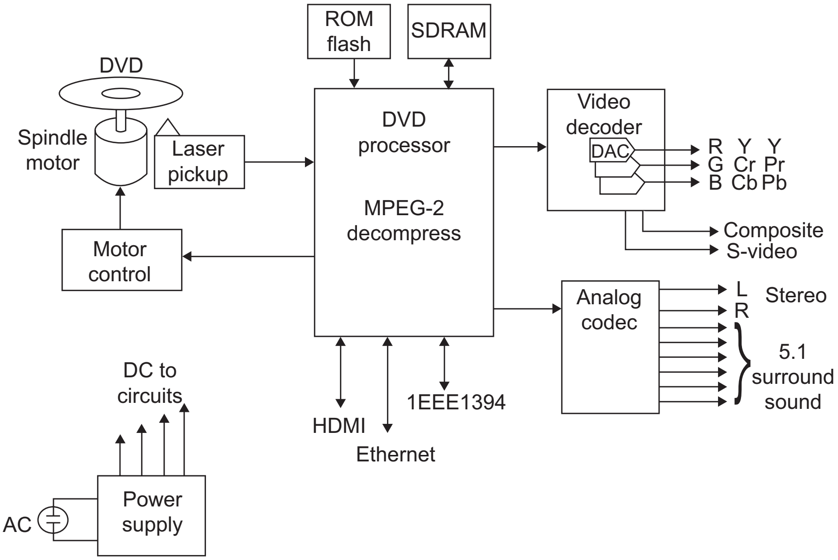
ROM flash
SDRAM
DVD
Spindle motor
Laser pickup
DVD processor
MPEG-2 decompress
Video decoder
DAC
R Y Y
G Cr Pr
B Cb Pb
Composite
S-video
Motor control
Analog codec
L
R
Stereo
5.1 surround sound
DC to circuits
HDMI
Ethernet
1EEE1394
AC
Power supply

FIGURE 11.21 Block diagram of standard DVD player.

for interfaces, an HDMI output is provided. Most players also have an Ethernet port and USB connections.

Do not forget that some of the later models are also DVD recorders. They play DVDs but also burn new ones with video from the TV set, cable box, satellite box, or a built-in tuner.

Blu-ray DVD

Blu-ray is the most recent generation of DVD medium and players. The discs are the same size but thanks to even smaller pits, lands, and track spacing, even greater storage capacity is available. With 320-nm track spacing and a 405-nm (blue light wavelength) laser, the single-sided disc can hold up to 25 GB of data. That is enough to store two full-length HD movies or 13 h of standard-definition video. A double-layer version is also available that doubles those storage figures. Blu-ray uses MPEG-2 compression but also supports other video compression standards such as MPEG-4/H.264 and VC-1.

Virtual and Augmented Reality

Virtual reality (VR) is an emerging video specialty that is used to reproduce or simulate an alternative reality. Using computer generated images and related sound, the user can experience an almost real or even an imaginary environment. 3D images can be produced.

VR is experienced by way of special goggles called a VR headset. This binocular head-mounted display completely covers the eyes and is often used with speakers or headphones in addition to motion sensors attached to the hands to duplicate motion and generate interaction and other involvement in the scenes presented. The computer generated scenes and situations that make the viewer feel as though they actually exist in this artificial setting. Various scenarios are produced, presented, and interacted with by the user.

Commercial VR products are now available from several sources. A powerful PC is required to meet the software requirements. Most of the software is related to computer games although a developing application is simulating real environments for training and education. Other applications are being explored.

VR produces a very real, immersive, and sometime intense experience. Any simulated motion has produced nausea in some users.

Augmented reality (AR) is a variation of VR that also uses goggles to enhance an existing scene or environment. The user views a current setting directly or by way of a video camera and that scene is overlaid with a computer-generated augmentation for some application. A heads-up display in a car windshield is a type of example. A current scene or situation is artificially modified by the computer overlay in real time.

AR applications are still developing, but current uses include games, military, training and education, manufacturing, and medical. Microsoft HoloLens is an AR product.

Project 11.1

Look at a Pixel

Get a magnifying glass and hold it up to your TV screen to see the individual color pixels or subpixels. What is their shape?

Project 11.2

Try Over-the-Air TV

If you get your TV by cable or satellite, try receiving local channels on an antenna. Use an indoor antenna for simplicity. You can buy a set of “rabbit ears” or other indoor antenna from a local electronic store such as Best Buy, Fry’s, or Walmart. Or try an outdoor antenna. You can get some channels on a simple loop of wire 1 yard long.

This page intentionally left blank

Chapter 12

Industrial Control

How to Automate: Monitor, Process, and Control

In this Chapter:

- Control defined.
- Open and closed control systems.
- Sensors and transducers.
- Actuators like relays, solenoids, and motors.
- Thyristors.
- Programmable logic controllers.
- Data acquisition systems.

INTRODUCTION

Industrial control is one of the four major applications of electronics. In industrial control, electronic components, circuits, and equipment are used to operate various types of machines in manufacturing plants. Most industrial machines such as robots are mechanical in nature but electrically operated. Most are powered by electric motors or other actuators. To operate the machine, electrical power must be properly controlled. Industrial control electronics is used to turn machines off and on at the appropriate time, control their speed of operation, and otherwise produce the desired manipulations.

Electronic control is not confined to industry. There are lots going on in the home and car as well. The principles given here apply to those applications as well.

OPEN- AND CLOSED-LOOP CONTROL

The basic industrial control process is illustrated in Fig. 12.1. Electrical inputs, such as switches, sensors, or other devices, are used to initiate an electronic control process. The signal-processing circuits generate control output signals that are used to operate the industrial machines. This basic process is known as open-loop control. The typical device being controlled by the output signals is an electric motor that operates a machine. Other outputs could operate lights, relays, solenoids, or a variety of other devices.

Electronics Explained. http://dx.doi.org/10.1016/B978-0-12-811641-8.00012-6
Copyright © 2018 Elsevier Inc. All rights reserved.

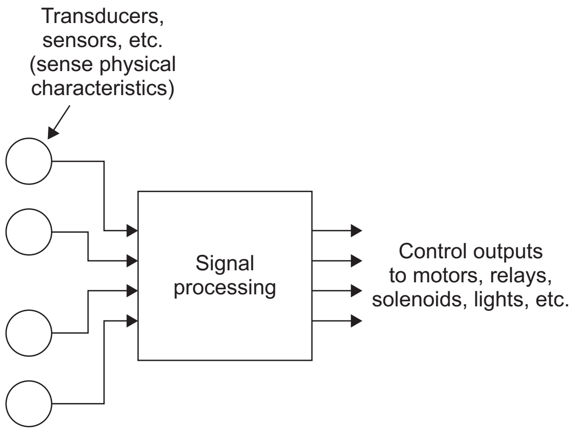
Transducers, sensors, etc. (sense physical characteristics)
Signal processing
Control outputs to motors, relays, solenoids, lights, etc.

FIGURE 12.1 General block diagram of industrial control process.

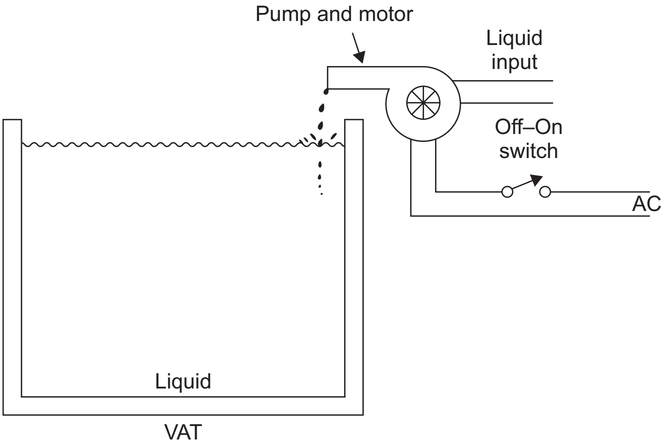
Pump and motor
Liquid input
Off–On switch
AC
Liquid
VAT

FIGURE 12.2 Open-loop control system for filling vat with liquid via a pump.

Open-Loop Control

A simple example of an open-loop control system is shown in Fig. 12.2. The input is a switch that applies electrical power to a motor. The motor, in turn, operates a pump that causes liquid to be put into a vat or tank. As soon as the vat is full, the operator turns the switch off. The basic control process in this system is simply turning power off and on to control the pump. The operator visually monitors the level in the vat until the correct level is reached.

Closed-Loop Control

An improved form of industrial control system is illustrated in Fig. 12.3. It is known as a closed-loop control system. The system is given an initial input

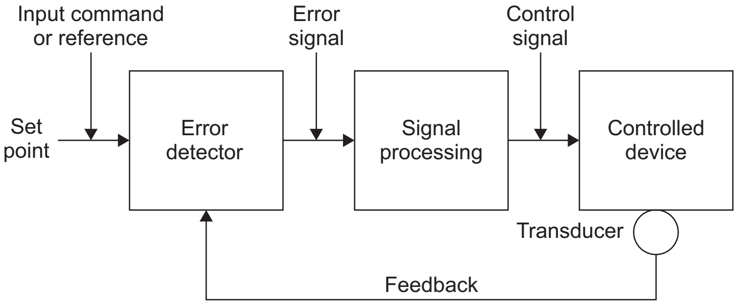

FIGURE 12.3 Generic diagram of any closed-loop control system with feedback.

reference called the set point. The set point is usually a voltage that represents some physical value that is to be achieved. It is applied to an error detector that compares this input to a signal from the controlled devices. If the two signals are different, an error output is produced. This error signal is processed (usually amplified), and an output control signal is derived from it. It is this output signal that operates the controlled device. A transducer monitors the controlled device to see that the desired outcome is obtained. The signal from this transducer is called *feedback*. The feedback signal is compared to the input to produce the error signal. The control signal operates the controlled device until the error is reduced to zero. At that time, the desired output condition, where the feedback equals the set point, is obtained.

The key to closed-loop control is the feedback. Feedback signals tell the system whether the machine is performing correctly. The feedback signal allows the system to adjust itself in such a way that the desired output is continuously accomplished. Control is automatic.

A closed-loop version of our system to fill a vat with liquid is shown in Fig. 12.4. An initial input signal is applied to start the process. If the vat is not full, the transducer will sense it and send a signal to the error detector that says the liquid level is low. The error detector generates an output signal that is processed and used to start the pump motor. The motor rotates the pump, which puts liquid in the tank. When the vat is full, the sensor will generate a signal that is sent to the error detector. The error is now zero, so the pump motor is turned off.

The primary benefit of a closed-loop control system is that its operation is automatic. In the open-loop system, an operator has to turn the system off and on manually to fill the tank. In the closed-loop control system, the operator gives an initial input, called the set point, to which the error detector will compare the feedback signal from the sensor. He or she turns the system on. From that point on, operation is automatic. The pump turns itself off when the vat is full. If the liquid level drops below the desired level due to usage, the sensor indicates a low level and automatically turns on the motor so that the pump will again fill the tank. When the tank is full, the unit shuts off by itself.

Most industrial control systems are closed loop to make them automatic. And your home has some as well. Your toilet flusher is a closed loop as is your heating

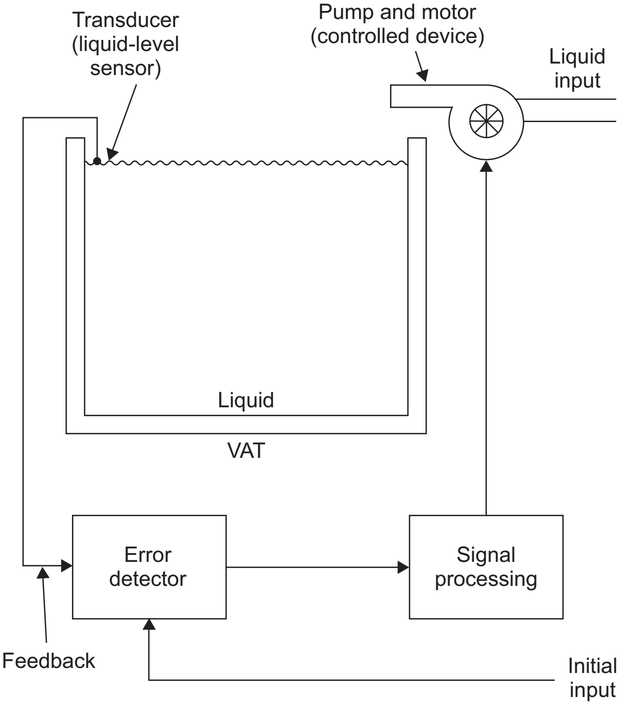

FIGURE 12.4 Filling vat with liquid automatically with closed-loop control system.

and air conditioning (HVAC) thermostat. Appliances such as your washer, dryer, and dishwasher are other examples of closed-loop control. All of them contain a microcontroller with sensors, timers, and feedback to make them totally automatic.

Controllers

While many industrial control applications are implemented with simple control circuits, most industrial control is carried out by specialized instruments called controllers. These units accept inputs from sensors, condition them, process them, and then generate output control signals.

A good example of an analog controller is one used to control the temperature in a system. A resistive temperature sensor is used to monitor the temperature of the liquid in a pipe. If the temperature goes up, the sensor output goes up. This is sensed by the controller circuitry. Inside the controller, a comparator is set to a desired temperature threshold. Should the temperature exceed a specified set point, the controller will turn off the device heating the pipe and liquid. Should the temperature drop below another set point temperature, the controller will turn on the heating device in an effort to maintain the set point temperature. While the heating element will be turning off and on, it will produce the desired average overall temperature.

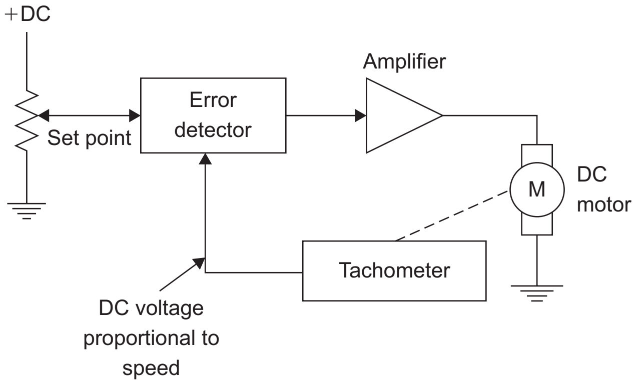

FIGURE 12.5 Closed-loop control system for motor speed control and regulation.

A continuous analog controller will provide proportional control rather than an on/off response. For instance, suppose that you want to control the speed of a motor. The system depicted in Fig. 12.5 does this. You want the motor to stay set to the speed you want, say 1200 rpm. A pot connected to a voltage represents that speed value. The voltage is applied to an error detector along with the feedback input from a tachometer. A tachometer is a device that senses rotational speed and develops a proportional DC voltage. That voltage is sent to the error detector and compared to the set point. If they are different, an error voltage is developed. This voltage is then amplified and applied to the motor. The speed of a DC motor is proportional to the voltage applied to it. So if the motor slows down, the error voltage is increased and that is amplified and applied to the motor speeding it up to compensate for the speed drop. A speed up in the motor produces an increased tachometer output. The resulting error signal causes the motor speed to drop until the error is zero. With such a circuit, the motor speed remains constant.

An analog controller is a product made up of the set point pot, error detector, and a selection of amplifiers whose gain can be varied and an output amplifier that can drive the motor or other device. Many different types of commercial electronic controllers have been developed to monitor and control physical characteristics such as pressure, physical strain, weight, liquid flow rate, physical position, or liquid level in a tank. Digital controllers that contain a microcomputer are also available.

SENSORS

An essential element of most control systems is the inputs from sensors. Sensors are the components that detect physical changes or events and convert them into electrical signals to be processed. Sometimes the term transducer is used to refer to the sensor. A transducer is a device that converts one type of energy into another such as mechanical energy to electrical. There are literally hundreds of different types of sensors. A few examples of the most common types follow.

Temperature Sensors

Temperature is probably the most often sensed physical characteristic. Temperature sensors convert temperature to a resistance change or voltage change. The most commonly used temperature sensors are RTDs, thermistors, and thermocouples.

RTD—An RTD is a resistive temperature device. It is essentially just a platinum wire whose resistance varies with temperature. Most of them have a resistance of 100 Ω at zero degrees Celsius (0°C). As the temperature changes, the resistance changes to reflect the change. RTDs have what we call a linear positive temperature coefficient. That is, as the temperature goes up, the resistance goes up and vice versa.

To convert the resistance variation into a voltage, the RTD is usually put into a bridge circuit as shown in Fig. 12.6. The bridge is initially balanced with the variable calibrate control (CAL) to produce zero output at a given temperature. Then if the temperature varies, the bridge becomes unbalanced and a voltage is produced. This is then amplified into a larger voltage that is a measure of the temperature. The output can be measured with a meter or it can be digitized for use as input to a digital system for display or control.

Thermistor—A thermistor is a resistor whose resistance has a negative temperature coefficient. As temperature goes up, the resistance goes down and vice versa (see Fig. 12.7A). The resistance change is much greater than that of an RTD. The thermistor can be used in a simple voltage divider to develop an output voltage or in a bridge circuit like the RTD (see Fig. 12.7B).

Thermocouple—A thermocouple is a unique type of temperature sensor, as it develops a voltage rather than a resistance change. It is formed with two dissimilar metals. If the junction of the metals is heated, a voltage is developed. The voltage is usually in the millivolt range so that it usually has to be amplified before it becomes useful. The thermocouple's main advantage is its accuracy at very high temperatures.

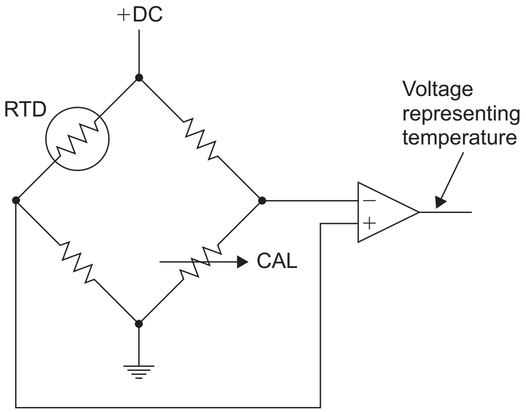

FIGURE 12.6 Most resistive sensors such as an RTD or thermistor are placed in a bridge circuit to condition them to provide amplified output voltage.

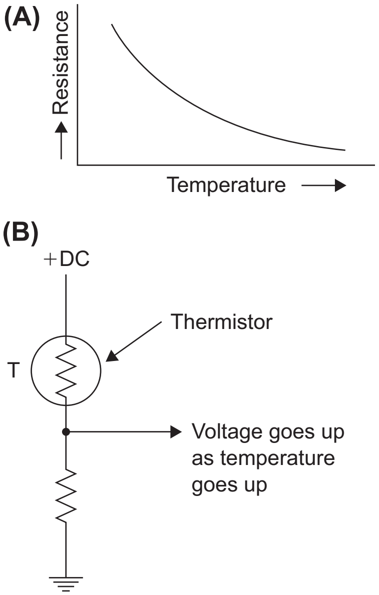

FIGURE 12.7 Characteristic curve of thermistor and simple voltage divider circuit producing output voltage.

Solid state—Diodes and transistors make good temperature sensors. A silicon diode has a very linear voltage drop variation that decreases with an increase in temperature. This is true of the emitter–base junction of a bipolar transistor as well. Zener regulator diodes exhibit a similar response. The sensor is usually packaged into a complete integrated circuit with an amplifier that produces a voltage output proportional to the temperature variation.

Pressure Sensors

A pressure sensor responds to force or pressure. There are many different types. A widely used kind is called a *strain gauge*. A strain gauge is essentially a thin pattern of metal deposited on a plastic base like that shown in Fig. 12.8. It has a specific value of resistance. A value of 120 Ω is common. The strain gauge is then glued or cemented with an adhesive to the object to which the pressure is to be applied. An example is a steel beam. When pressure is applied to the beam, it will bend. As it bends, it stretches or compresses the strain gauge resistance element. If the resistive element is stretched, its resistance increases. If it is compressed, the resistance decreases. Knowing the specific characteristics of the resistance change per pound of pressure, an exact measure of the pressure can be determined. Mostly strain gauges are used in a bridge circuit like that shown in Fig. 12.6. Since the resistance variation is small and the resulting bridge circuit output is small, a special amplifier called an *instrumentation amplifier* is used. It has differential inputs and very high gain.

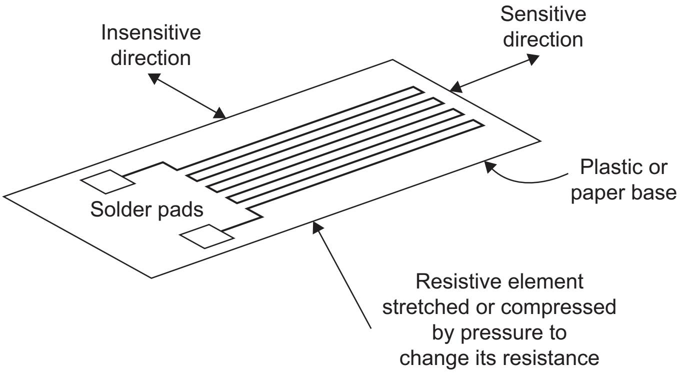
Insensitive direction
Sensitive direction
Plastic or paper base
Solder pads
Resistive element stretched or compressed by pressure to change its resistance

FIGURE 12.8 Strain gauge is a resistive element whose value changes with stress or pressure.

A special pressure sensor called a *load cell* is typically made up of strain gauges. It is fully calibrated to provide a specific voltage output with a specific pressure. Load cells are widely used for weight measurement. An example application is the use of load cells in truck weighing scales.

There are other forms of pressure transducers. Many are solid-state devices with an output voltage proportional to pressure. *Piezoelectric pressure transducers* use the voltage produced by certain types of crystals or ceramics when pressure is applied to them.

Switches

Switches are widely used as sensors. Switches are mechanical devices with electrical contacts that open or close with a particular physical movement. Switches are used to detect mechanical movement and specifically to determine when the limit of some physical device has been reached. Called *limit switches*, these devices have various arms and mechanisms that permit ease of use in a variety of situations. Examples of such switches are those that can detect when a cabinet door is open or closed or a switch that senses when a mechanical device reaches a specific point where motion must be stopped. And of course, a human-operated switch may be the most common. It is usually a button that is pushed to initiate or stop some action.

Potentiometers

A potentiometer is a variable resistor or variable voltage divider. While most electronic pots are rotary devices, they also come in a linear variety. The variable arm of the pot is controlled by some external mechanical device. Then by using the resistance value of the pot or the voltage it produces as a variable voltage divider, it can be used to sense mechanical position.

Photoelectric Sensors

One of the most widely used industrial transducers is a photoelectric sensor. A photoelectric sensor is a component that responds to light and produces an electrical signal that can initiate some operation.

There are three basic types of photoelectric sensors. These are the *photoresistive* or *photoconductive cell*, the *photovoltaic cell*, and the *phototransistor*. The photoconductive sensor is a light-sensitive resistor. Its schematic symbol is shown in Fig. 12.9A. The resistance of the device varies with the amount of light falling on it. With no light falling on the sensor, its resistance will be very high, 100,000 Ω or more. When bright light shines on the sensor, however, its resistance will drop to a very low value, usually several hundred ohms or less. Of course, at a light level between bright and dark, the resistance of the device will be somewhere between 100 Ω and several hundred thousand ohms. The resistance is inversely proportional to the light level.

Another type of photoelectric device is the photovoltaic cell, whose schematic symbol is illustrated in Fig. 12.9B. This is the solar cell described in an earlier chapter. Whenever bright light falls on the cell, it will generate a small DC voltage. The maximum output of a typical photovoltaic cell is approximately 0.45–0.5 V. At lower light levels, the output voltage will be less. Photovoltaic cells are used primarily in power generation systems for producing charging voltage for batteries in satellites and other remote systems. Multiple cells are connected in various series and parallel combinations to get the desired voltage level and current capacity. In some applications, the photovoltaic cell can be used for industrial control. These are also found in portable calculators. Two AA cells at 1.5 V each produce 3 V in series. To generate 3 V with 0.5-V solar cells, you would need to connect six in series (6 × 0.5 = 3 V).

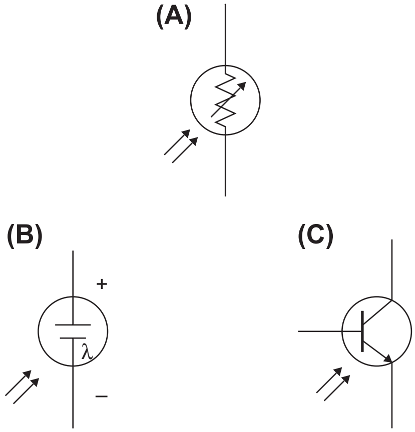

FIGURE 12.9 Schematic symbols for photo cells. (A) Photoconductive or photoresistive, (B) photovoltaic or solar cell, and (C) photo transistor.

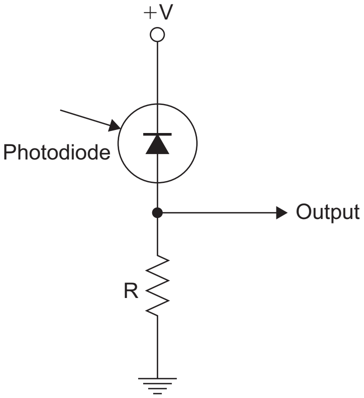

FIGURE 12.10 Photodiode and its simple output circuit.

A silicon diode can also be used as a light detector. Special diodes in clear housings are often used to detect light. These are called photodiodes. Photodiodes are operated in the reverse-biased direction, as shown in Fig. 12.10. With no light falling on the diode, it is cut off, so there is no voltage across R_1. When light strikes the diode, its reverse leakage current increases dramatically, causing a voltage drop to occur across R_1.

A phototransistor is a light-sensitive transistor. Its symbol is given in Fig. 12.9C. It only has two terminals, the emitter and the collector. The device has a base, but typically it has no lead. Instead, the device is built so that when the base is exposed to light, it will cause the emitter–base junction to conduct. When light strikes the base, it causes ionization, which simulates base current flow. Therefore, the transistor conducts between emitter and collector. The primary advantage of a phototransistor is its very high sensitivity over a photodiode. It will conduct even with a small amount of light applied to the base. The transistor essentially provides amplification so that small light levels can control a large current.

A typical closed-loop control circuit using a photoelectric cell is illustrated in Fig. 12.11. This circuit will automatically turn a light on when it gets dark and turn the light off when it is light again. Note that a photoresistive cell is connected as part of a voltage divider with resistor R_1. The DC voltage from this voltage divider is applied to one input of an op amp comparator. The DC reference input to the comparator comes from a voltage divider made up of R_2 and R_3. The potentiometer can be used to adjust the DC reference voltage or set point. The output of the comparator drives a switching transistor Q_1, which will turn the relay K_1 off or on. This, in turn, operates the light.

When the photoresistive cell is exposed to light, its resistance will be very low. Therefore, the voltage from the voltage divider will be low. It will be less than the DC set point voltage and the comparator will not be triggered. The output of the comparator will be a negative voltage, which will reverse bias the

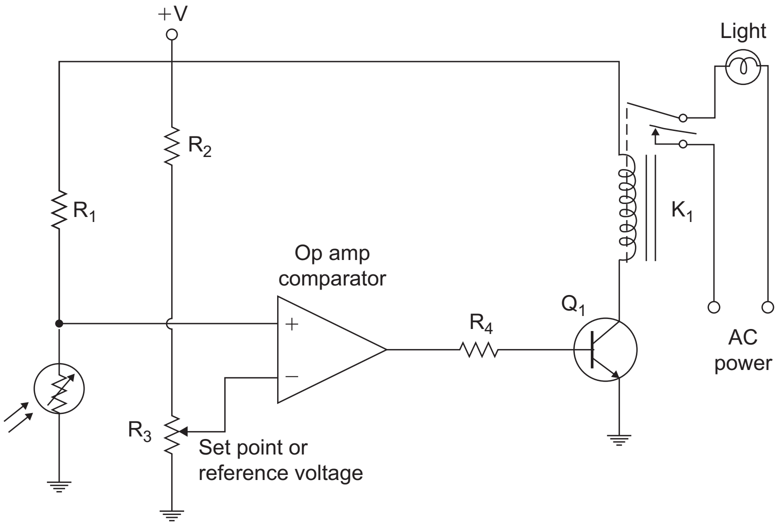

FIGURE 12.11 Simple off/on photo control circuit. Light turns on automatically at dark and off when light is present.

emitter–base junction of the transistor and keep it turned off. As soon as it gets dark outside, the resistance of the photocell will increase. This will cause the voltage divider output to increase. At some point, the voltage will be higher than the set point voltage, causing the comparator to switch. This applies a positive voltage to the transistor base, turning it on, and then turning on the relay and the light. The light will remain on as long as the photoconductive cell is in the dark. But as soon as light shines on the photocell, the voltage divider output voltage will drop, thereby triggering the comparator and turning off the transistor, the relay, and the light bulb. By adjusting the reference or set point voltage with potentiometer R_3, the triggering point can be varied, so that the light will be turned off or on at the desired level of ambient lighting.

This same basic concept can be used to produce a variety of industrial control operations. For example, you could use the circuit as an entry or intrusion detector. Shine a light on the photocell. If someone or something comes between the light and photocell, the circuit will trigger and the relay could then be used to operate a bell or buzzer or light.

Another example of a photoelectric controller is a tachometer circuit used for measuring the speed of a rotating shaft. Fig. 12.12 shows a small, plastic wheel connected to the shaft of a motor. There are alternate clear and opaque areas on the disk. The disk is positioned so that it is between the light-emitting diode (LED) and the phototransistor. As the disk rotates, the dark opaque areas will block the light, while the clear areas pass the light from the LED to the phototransistor. When a clear area on the disk passes, the LED light will shine through it, hitting the phototransistor, and turning it on. When a dark area

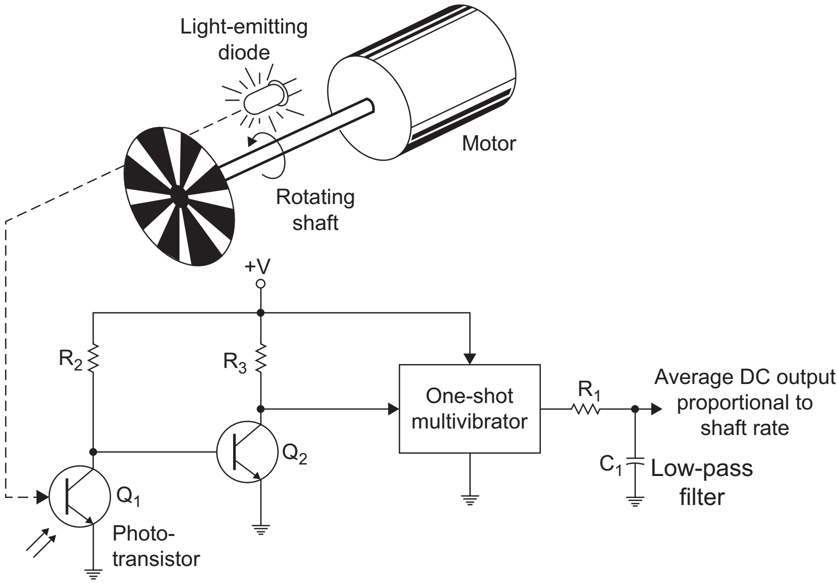

FIGURE 12.12 Simple tachometer circuit using phototransistor to create pulses that are averaged into a DC voltage proportional to speed.

passes, the light will be blocked and the phototransistor will turn off. The rapidly rotating disk will cause the transistors to turn off and on very rapidly. This rapid conduction and nonconduction of the phototransistor causes a series of pulses to occur at the output of Q_2. These pulses are used to trigger the one-shot multivibrator.

The one-shot multivibrator produces a series of fixed, amplitude-width pulses. These are applied to an RC low-pass filter made up of C_1 and R_1. This filter will average the pulses into a DC voltage. This output voltage is proportional to the motor speed. If the speed of the motor should increase, the pulses occur at a more rapid rate. This means that the pulses occur closer together, and thus the average voltage across C_1 goes up. A DC voltmeter connected to the output can be used to indicate the speed. At low speeds, the meter indication will be low, while at high speeds it will be high. The meter dial itself can be calibrated in terms of revolutions per minute (rpm), thereby giving an accurate speed indication.

Flow Sensors

These devices have liquid or gas pass through them and measure the amount of flow in gallons per minute or other measure. They develop an output voltage proportional to flow rate.

There are too many other sensors to catalog here but in most cases there is a sensor to convert any physical variable into a voltage to be measured.

OUTPUT DEVICES

An output device or an *actuator* is the thing that is most often controlled in a control system. The two most common types of actuators are relays and motors. There are many types of both.

Relays

A relay is an electromagnetically operated switch. A typical relay, and its schematic symbol, is illustrated in Fig. 12.13. Its main components are a magnetic coil and a set of switching contacts. With no current applied to the coil, the spring keeps the armature pulled down, so that the contact arm makes connection with

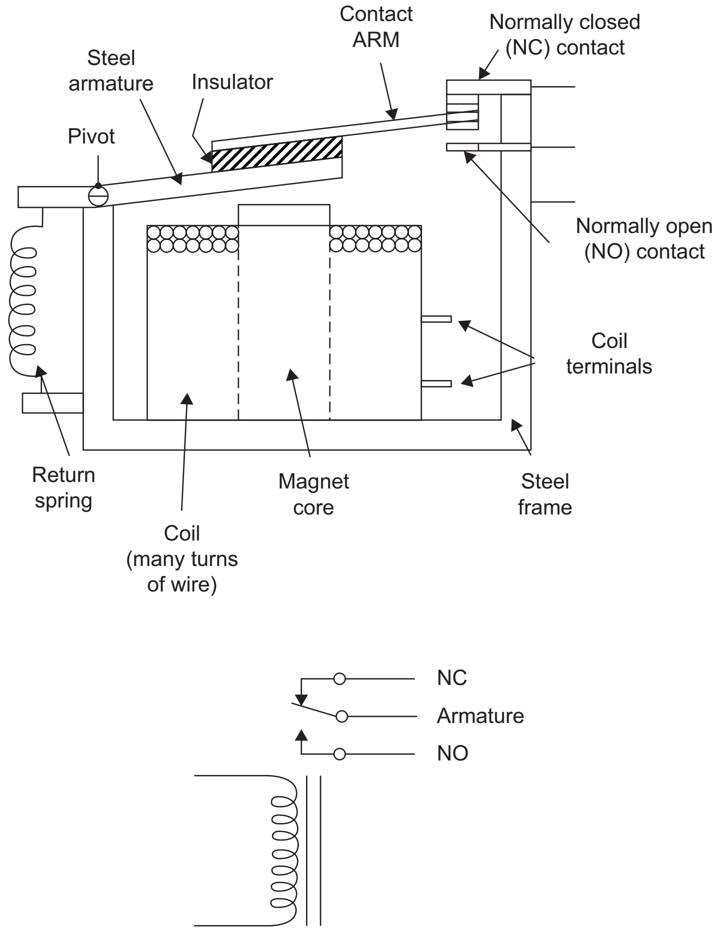
Contact
ARM
Normally closed
(NC) contact
Steel
armature
Insulator
Pivot
Normally open
(NO) contact
Coil
terminals
Return
spring
Magnet
core
Steel
frame
Coil
(many turns
of wire)
NC
Armature
NO

FIGURE 12.13 Cross-sectional view of typical relay showing major components and schematic diagram.

the upper, normally closed (NC) contact. When current is applied to the coil, a magnetic field is produced. This magnetic field attracts the steel armature, opening the upper contact, and closing the lower, normally open (NO) contact. The contacts form a single-pole, double-throw (SPDT) switch. When power is disconnected from the coil, the magnetic field ceases, and the contacts return to their initial state because of the spring.

Relay coils are designed to be operated from either AC or DC. The contacts are small circular disks of silver or tungsten. These materials are chosen because they will handle a large amount of current with minimum burning, arcing, and pitting.

The various kinds of switching contacts used in relays are illustrated in Fig. 12.14. NO contacts are referred to as *form A*. Applying power to the coil closes the contacts. NC contacts are referred to as *form B*. When power is applied to the relay coil, the contacts open. SPDT switching contacts are referred to as *form C*. When power is applied to the relay coil, the contact arm moves from contact A to contact B. Form C contacts are usually of the break-before-make variety. That is, contact A opens before the arm touches or "makes" contact B.

Sometimes you will hear the expression that a relay has been "picked" or "picked up." This means that power has been applied to the relay coil, and its contacts have been moved according to their function, and switching of a circuit has occurred.

Most relays usually have more than one set of contacts, so that multiple switching operations can be performed simultaneously. The various contact arms are ganged together so that they operate together when power is applied to the coil. A common configuration is two form-A contacts ganged to form a double-pole, single-throw (DPST) relay. Putting two form-C contacts together on the same relay produces a double-pole, double-throw (DPDT) relay.

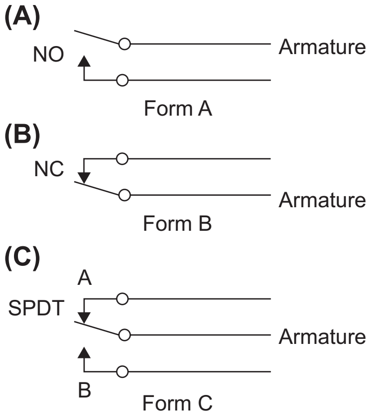

FIGURE 12.14 Common relay contacts. (A) Form A, (B) form B, and (C) form C.

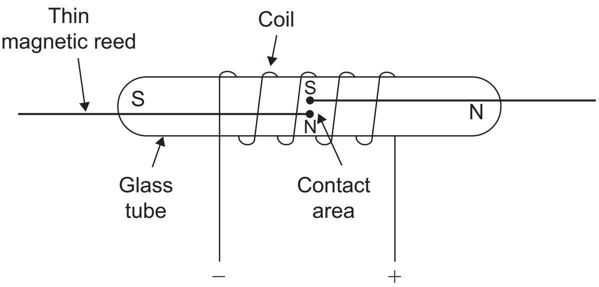

FIGURE 12.15 Reed relay.

Another widely used type of relay is the *reed relay*, illustrated in Fig. 12.15. The heart of the reed relay is a reed switch that consists of two thin, metallic, reed contacts sealed in a glass tube. The reed contacts are made of a magnetic material, so that they will be magnetized when a magnetic field is applied to them. A coil consisting of many turns of fine wire is wound on a form and placed over the reed switch. When power is applied to the coil, a magnetic field is produced. This magnetic field magnetizes the two reeds as if they were bar magnets. One end of each reed will have a north pole and a south pole. The reeds are positioned so that the north pole of one will attract the south pole of the other. Remember, magnetic theory says that unlike poles attract, whereas like poles repel. The result is that the contacts will move toward one another and touch, making a good electrical connection. When power is removed from the coil, the contacts become demagnetized and spring apart. The reed relay will also operate by moving a permanent magnet near it. When the magnet is near, the contacts will close. This is a common way to sense open windows or doors in a home security system.

The contacts on reed relays cannot handle as much power as the contacts on conventional relays because they are smaller. As a result, they are used only in lower current power applications. However, the contacts on a reed relay open and close faster than those of a larger standard relay. Reed relay contacts can open and close in a few hundred microseconds, while standard relays require many milliseconds to open or close because of the larger contacts and the related structures that must be moved.

Relays are used in two basic ways in industrial control:

1. To control a larger current with a smaller current.
2. For remote control.

Let us examine the use of a relay for controlling a larger current with a smaller current. A good example of this is the starting system in your car. A large, heavy-duty DC motor is used to spin the flywheel on the car's engine to start it. This motor draws a current of hundreds of amperes. To turn the engine and start the car, the battery must be connected to the starter motor. This could be done directly by having the ignition switch connect the battery to the motor. However, this switch would have to have very large contacts to handle the high current. Such a switch would be large and expensive to build. The large heavy wires from the battery to the motor would have to be lengthened and brought up into the car.

A better solution is to use a smaller ignition switch to control the current applied to a relay coil (see Fig. 12.16). The relay coil draws considerably less current than the motor; therefore, the switch contacts need not be large or expensive. The switching contacts on the relay are much larger and will carry the high motor current. When the ignition switch is turned on, its contacts close, applying current to the relay coil. The form-A relay contacts close, connecting the battery to the starting motor. When the car starts, the ignition switch is released, the relay drops out, and the starter motor shuts off.

Another application of a relay is remote control. Again, the starting system of a car is a good example. The load to be controlled, in this case the starter motor, is located remotely from the ignition switch. If the ignition switch were used to apply current to the motor, long wires would be required to connect the battery, the motor, and the switch. Heavy wire conductors would have to be used to carry the high current. The longer the interconnecting wires, the higher the resistance of the circuit. While large conductors have minimum resistance, when high current flows through them, voltage drops still occur. The voltage dropped across the interconnecting lines subtract from the battery voltage available for the motor. The voltage drops might be excessive, unless very large, expensive, and inconvenient conductors are used.

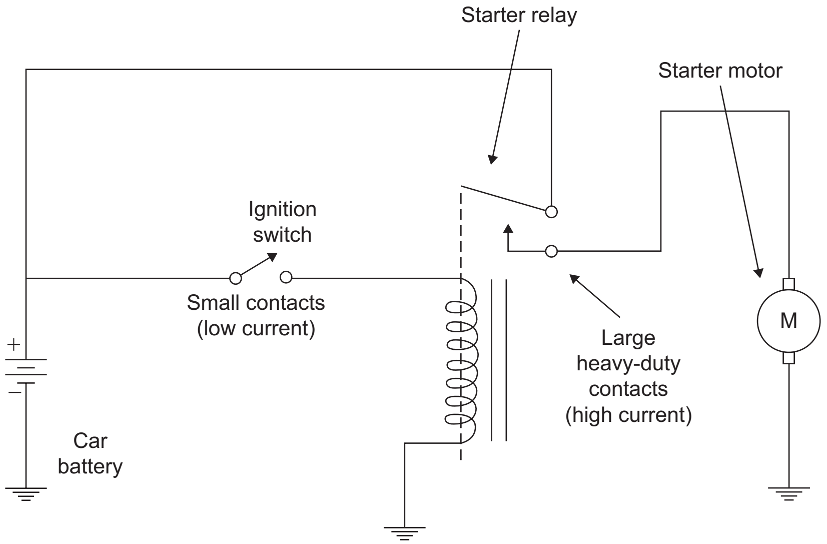
Starter relay
Starter motor
Ignition switch
Small contacts (low current)
+
−
Car battery
M
Large heavy-duty contacts (high current)

FIGURE 12.16 Controlling a large current with a smaller current with a relay.

To eliminate this problem, a remote relay is used. The relay itself is located near the starter motor and the battery. In this way, the interconnecting leads can be kept short, thereby minimizing voltage drops. Since the relay operates from very low current, longer interconnecting wires can be used to attach it to the ignition switch without noticeable voltage drops.

In some industrial control applications, the load to be operated is located at extreme distances from the switch used to control it. The current controlling the device does not have to be passed through very long conductors if a relay is used. Small conductors running to the switch remotely from the load control the relay while the relay turns the load off and on.

Relay contacts stay closed (or open) as long as power is applied to the coil. This means that the switch operating the relay coil must be kept closed if the relay is to remain actuated. There are applications, however, where it is desirable to use a momentary contact switch to operate the relay. An example is a pushbutton that, when depressed, causes contacts to be closed and power to be applied to the relay. When the button is released, a spring causes the contacts to open. With such an arrangement, the relay will close momentarily while the button is depressed, but will drop out when the button is released. The objective is to cause the relay to latch into its actuated position when the coil is momentarily pulsed. Such a relay is called a *latching relay*. Some latching relays use a mechanical arrangement on the armature to cause the contacts to latch closed when the coil is momentarily pulsed.

A relay can be made to latch electrically if it has an extra set of contacts. A typical latching circuit is shown in Fig. 12.17. This relay has two sets of form-C contacts that operate simultaneously. The lower set of contacts is used to operate the load. The upper set of contacts is used to automatically latch the relay once it is pulsed. Switch S_1 is a NO momentary contact pushbutton that is used to apply power to the relay coil. Depressing the button causes the relay coil to operate, closing the contacts. The lower set of contacts operates the load as desired. The upper set of contacts applies the power to the coil. Thus, when pushbutton S_1 is released, the circuit to the relay coil remains closed through the upper-latching contacts. To turn off the circuit, NC pushbutton S_2 must be depressed. This breaks the relay coil circuit, removing power. The relay drops out, opening the load contacts and the latching contacts.

While many relay coils are operated by mechanical switches, others are operated by a transistor. A typical transistor-controlled relay circuit is shown in Fig. 12.18. The transistor is operated as a switch. When the base is grounded through contact B on S_1, the transistor is cut off and acts as an open switch. Therefore, no current flows in the relay coil. When a voltage is applied to the base resistor, through contact A on S_1, the transistor saturates, acting like a very-low-resistance or "on" switch. Current flows through the relay coil, operating the contacts. Removing the base current from the transistor turns the transistor off, terminating current in the relay. With this arrangement, an even smaller current can be used to control the already small relay current. A base current, of only microamperes, is sufficient to cause the transistor to operate.

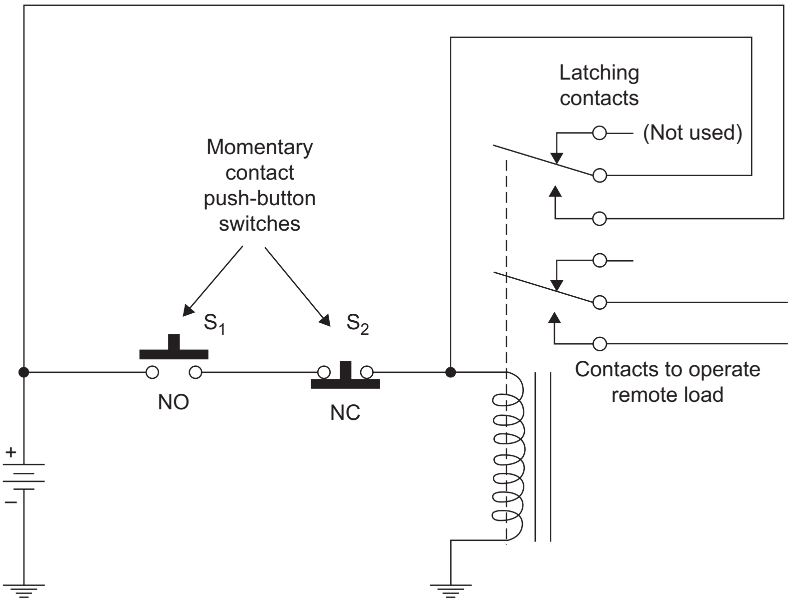
Latching
contacts
(Not used)
Momentary
contact
push-button
switches
S_1
S_2
NO
NC
+
–
Contacts to operate
remote load

FIGURE 12.17 Latching a relay with its own contacts.

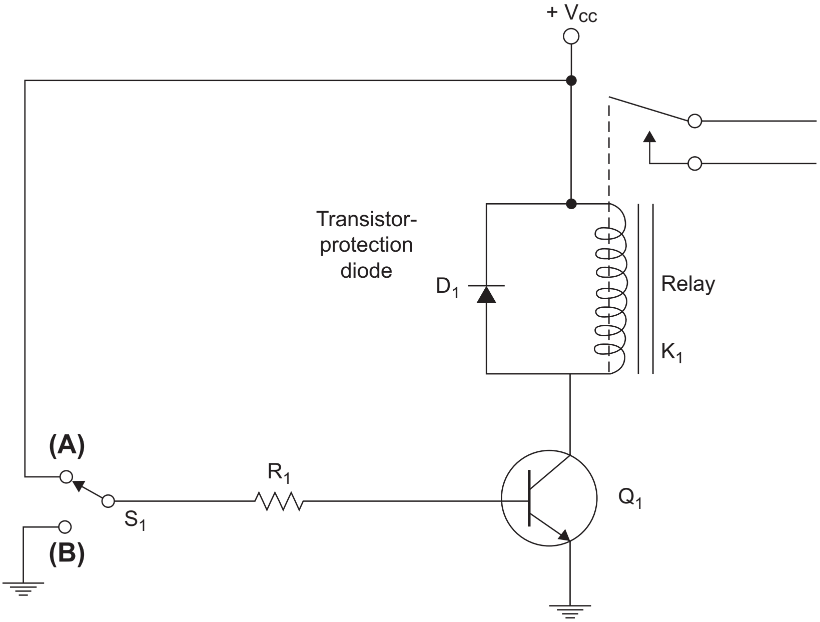
+ V_{CC}
Transistor-
protection
diode
D_1
Relay
K_1
(A)
(B)
S_1
R_1
Q_1

FIGURE 12.18 Operating a relay with a transistor is a common application. The diode protects the transistor when power is removed from the relay coil.

In the circuit of Fig. 12.18, when base current is removed from the transistor, the transistor will cut off. When current through the relay coil ceases, the magnetic field around it collapses. In doing so, the magnetic field induces a very high voltage in the relay coil. This very high voltage will be applied to the collector of the transistor. This voltage could be as high as several hundred, or even thousands, of volts, depending on the relay coil. Such a high voltage will damage the transistor. To protect the transistor against such occurrences, some kind of voltage spike eliminator or surge suppressor must be used across the relay coil. A commonly used protection device is a silicon diode D_1, as shown in Fig. 12.18. When current is applied to the relay coil through the transistor, the diode is reverse biased and has absolutely no effect on the circuit. When the transistor turns off, the high voltage will be inducted into the coil with the polarity reversed from the supply voltage V_{cc}. This causes the diode to become forward biased and conduct. Thus, the diode effectively shorts out the high voltage spike, protecting the transistor.

Solenoids

A solenoid is a device that produces linear motion with an electromagnetic coil (see Fig. 12.19). A magnetized steel rod called a plunger is inserted into a coil of wire. When a voltage is applied to the coil, a magnetic field is produced. It interacts with the magnetic rod producing attraction and/or repulsion that causes the rod to move into or out of the coil. The rod is usually physically constrained in one or both directions, and is often spring loaded so that it always returns to the same place when power is removed.

Solenoids produce linear motion. They are usually connected to a lever or ratchet mechanism that produces other forms of motion including rotation. Solenoids are also used in valves to turn off the flow of a liquid or gas. By continuously controlling the current in the coil, the rod may also be precisely set to some position other than off or on. This effect is used in valves to continuously control the opening.

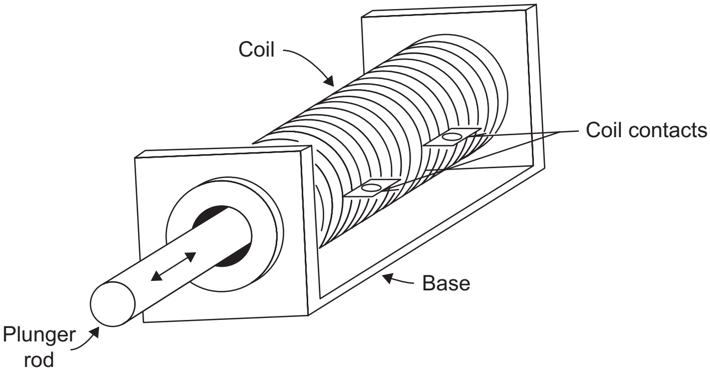

FIGURE 12.19 A solenoid produces linear motion using magnetic principles.

Motors

A motor is a device that converts electrical energy to mechanical energy. The mechanical output is typically the rotation of a shaft. That rotation can then be used directly or converted into other mechanical forms with levers, gears, ratchets, or cams. There are dozens of different motor types but most are categorized as either DC or AC.

Motors: Billions and Billions Sold and Used

Motors are one of those devices we take for granted. We use them every day but do not realize it. Most motors are hidden but do an enormous amount of physical manipulation. In industry, motors drive machine tools, conveyer belts, pumps, and other mechanical systems. Elevators use big motors.

At home, motors are everywhere. Just think a minute and try to identify all the motors you have in your home and in your car. All of your appliances have motors, such as the washer, dryer, dishwasher, and HVAC units. There are motors in your mixers, blenders, and can openers. Shavers and hair dryers have motors. An analog clock is a motor. Then there are the motors in your electronics. In a PC or laptop, there are the fan motor and the ones in your hard disk and CD drives. Your printer, scanner, and copier and fax machine have motors. Your DVD player and CD player have multiple motors.

Then in your car is the starter motor, as well as the motors for windows, seats, mirrors, and windshield wipers.

Can you name some others?

DC Motors

A DC motor operates from a DC voltage source like a battery or power supply. It is made up of a magnet and a rotating coil. When a current is applied to the coil, it forms an electromagnet. The magnetism of the coil interacts with the magnet field, producing rotary motion. The magnet is commonly a permanent magnet in smaller DC motors but an electromagnet in larger motors. In the larger motors, a special field coil is energized with DC voltage as well to create the magnetic field.

The main specifications of a motor are its horsepower rating, speed, and torque. Speed and torque are controlled by the amount of current flowing in the coil and/or the strength of the magnetic field. Increasing the voltage applied to the motor usually causes speed and torque to increase. Some kind of electronic control circuit is usually involved to set the speed to the desired value. A variable resistor in series with the motor will control its speed, but usually more complex electronic circuits are involved.

Incidentally, the direction of rotation is determined by the polarity of the DC voltage. Reversing the polarity reverses the magnetic field, thus changing the direction of rotation. Fig. 12.20 shows a neat circuit used to change the direction of

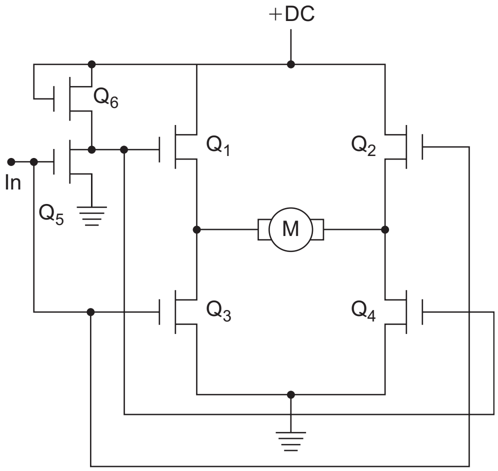

FIGURE 12.20 Using an H-bridge circuit to reverse the current flow and direction of rotation in a DC motor.

rotation of a DC motor. This is called an H-bridge and is made up of MOSFET switches. Q_2 and Q_3 are turned on by a positive input voltage IN. Q_5 inverts this voltage to zero, which keeps Q_1 and Q_4 off. Therefore, current (electrons) flows from ground through Q_3 through the motor (M) from left to right through Q_2. Setting the input to zero turns off Q_2 and Q_3 and turns on Q_1 and Q_4. Now current flows from ground through Q_4 and the motor from right to left and then through Q_1. This changes the direction of the motor with just a simple on/off input control voltage.

Most DC motors rotate too fast for many applications. A typical speed range is 1000–10,000 rpm. To achieve lower speeds, the motor is used in conjunction with a gear box or pulleys.

A special type of DC motor is the *stepper motor*. This is a DC motor driven by two DC pulse signals. The pulses operate coils or windings that produce a rotating magnetic field. A permanent magnet rotor interacts with the rotating magnetic field to produce motion. The interesting characteristic is that the rotation occurs in increments or steps instead of a continuous rotation as with a standard DC motor. Depending on the number of windings and coils and elements on the rotor, the increments usually vary in increments of 22.5, 18, 15, 7.5, or 3 degrees or smaller. The stepper motor therefore permits very precise positioning. Yet it can be rotated at several thousand rpm if needed.

AC Motors

AC motors usually operate from a 50- or 60-Hz power line. Again the rotation is produced by interacting magnetic fields. The speed of rotation is fixed by the number of coils and magnet poles inside the motor. That speed is also often higher than needed, so gears and/or pulleys are used to drop the speed to the desired value.

To change the speed of an AC motor, you must vary the frequency of the AC voltage applied to it. This means that you cannot operate the motor directly from the AC power line if the speed must be changed. Instead, you must use a so-called AC drive. This is a box of electronics that generates a variable frequency AC voltage to apply to the motor. An AC drive takes the AC line input and uses semiconductor switches to produce an AC signal at a different frequency.

Other Actuators

Besides motors, solenoids, and relays, there are a variety of different controlled devices. Valves are common device. Valves are used to turn the flow of a liquid or gas on or off. Proportional valves very carefully regulate the amount of flow.

Other devices are hydraulic or pneumatic devices. Hydraulic devices use oil in cylinders to provide mechanical force where needed or position something mechanically. Pneumatic cylinders work with compressed air to apply force or control the position of a device.

THYRISTORS

Many industrial switching circuits are implemented with MOSFETs designed for high voltage and high power. You will sometimes hear these devices called *insulated gate field-effect transistors* (IGFETs).

Other power-switching devices are also widely used, especially thyristors. A thyristor is a semiconductor device used for switching purposes in industrial control. Like the relay, it is used to apply and control electrical power to motors, heating elements, lights, and other loads in industrial applications. In many applications, thyristors replace relays. A thyristor acts as a switch, but because of its solid-state nature, is far faster in switching than a relay. There are three basic types of thyristors: silicon-controlled rectifiers (SCRs), triacs, and diacs.

Silicon-Controlled Rectifier

The SCR is a three-terminal thyristor that acts like a silicon rectifier diode whose conductor is controlled by an input current. The schematic symbol for an SCR is shown in Fig. 12.21. The symbol is similar to that of a diode with a cathode and an anode. Note that the third element of the SCR is known as the gate. The SCR will conduct current between cathode and anode, but only if the proper control current is applied to the gate. The gate must be made positive with respect to the cathode for the SCR to conduct. When conducting, the SCR acts like a closed switch. The voltage drop across the cathode and anode will be approximately 0.7–1.8V, depending on the size of the SCR and how much current is flowing through it. When the cathode and anode are reverse biased, current will not flow through the device.

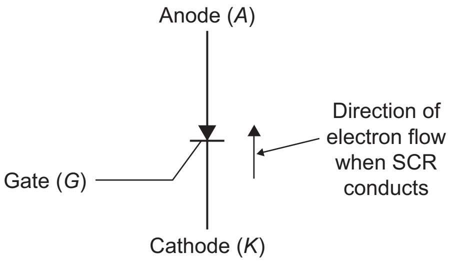

FIGURE 12.21 Schematic symbol of an SCR.

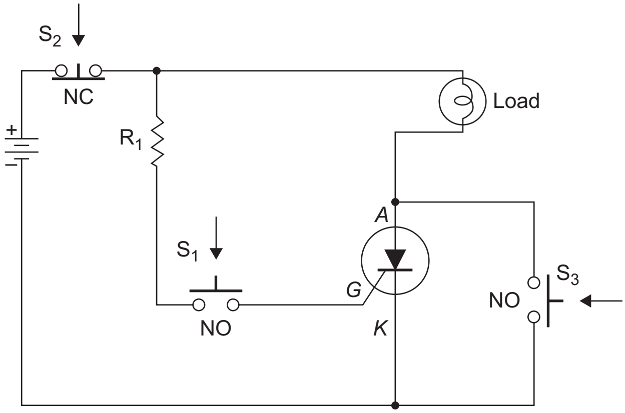

FIGURE 12.22 Circuit that shows how an SCR is turned on and off.

Thyristors, like relays, are generally used to control a larger current with a smaller current. Fig. 12.22 shows how an SCR is used as a switch to apply DC power to a light bulb. The load could also be a motor or heating element. The DC voltage is connected so that the cathode and anode of the SCR are forward biased, but no current will flow through the device until a current is applied to the gate. This is done with the switch S_1. With the switch open, no current flows in the gate. Closing the switch, however, applies a positive voltage to the gate through resistor R_1. This causes the SCR to turn on. When it conducts, it acts as a low-resistance "on" switch and the light bulb goes on. At this point, switch S_1 may be opened. It is not necessary to maintain current in the gate for the device to continue to conduct. The gate current only needs to be momentary, for it is only required to turn the device on. The device remains on like a latching relay that has been pulsed.

The gate element is only used to turn the SCR on. Removing gate current will not turn the device off. To cause the SCR to stop conducting, the circuit must be broken. This can be done with a momentary contact, NC pushbutton, in series with the circuit, like S_2 in Fig. 12.22. Depressing this pushbutton will break the circuit, stopping the current in the SCR, and the light bulb will go off. To turn the bulb back on, S_1 must again be depressed to apply gate current momentarily.

Another way to stop conduction is to momentarily short around the SCR, as shown in Fig. 12.15. When switch S_3 is closed momentarily, current will flow through it and the bulb, bypassing the SCR. The current in the SCR will drop to zero. When switch S_3 is opened, the circuit will be off.

While SCRs are sometimes used to control DC power, in most applications they are used to control AC. Fig. 12.23A shows the SCR used to apply AC to a light bulb. If switch S_1 is closed, gate current will be applied to the device. The device, however, will only conduct when the anode is made positive with respect to the cathode. This, of course, occurs when the applied AC voltage has the correct polarity. Since the SCR operates as a rectifier diode, current will only flow through the device on the positive half-cycles of the sine wave. The current through the light bulb will be a pulsating DC, as illustrated in Fig. 12.23B. The brightness of the lamp will depend on the average amount of current flowing.

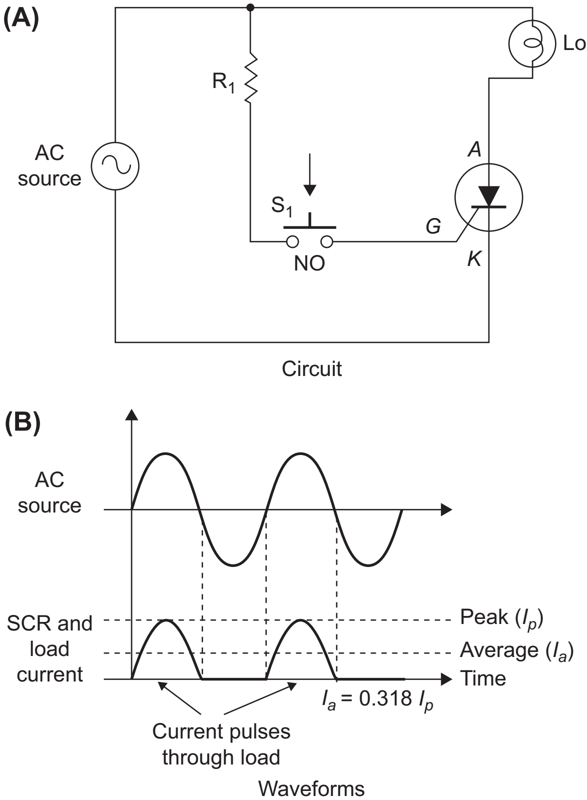
(A)
Load
R1
AC source
A
S1
G
NO
K
Circuit
(B)
AC source
SCR and load current
Peak (Ip)
Average (Ia)
Time
Ia = 0.318 Ip
Current pulses through load
Waveforms

FIGURE 12.23 Using an SCR to control power to a load using AC. The SCR rectifies the AC.

The average amount of current through the light bulb, or other load, can be controlled by using the gate element. Various electronic circuits are used to adjust the time when the SCR turns on. By turning on the SCR at the appropriate time in the positive half-cycle, the duration of the current flow can be controlled. The longer the current is allowed to flow, the higher the average current in the load. By changing the control point, the brightness of the bulb can be varied.

Diac

A diac is a two-terminal semiconductor device that conducts in either direction when a voltage of a specific level is exceeded. This voltage is called the *triggering voltage*. Typical diac triggering voltages are in the 20- to 45-V range. The diac is used as a triggering device for SCR and triacs as it determines when the device will conduct or cut off.

Triac

A triac is another three-terminal thyristor. Its operation is similar to that of an SCR, but the triac will conduct in either direction. The schematic symbol for the triac is shown in Fig. 12.24. Current will flow from main terminal 1 (MT1) to main terminal 2 (MT2), or from main terminal 2 (MT2) to main terminal 1 (MT1), depending on the polarity of the voltage applied, and if the gate triggers the device on.

An approximate, equivalent circuit of the triac is shown in Fig. 12.25. Here, two SCRs are connected back to back with their gates connected together. Current can flow in either direction, depending on whether D_1 or D_2 conducts. Of course, SCR conduction depends on whether a gate-triggering current is applied. Triacs, like SCRs, are used primarily to switch AC voltage to a load. If gate current is applied continuously, the triac will conduct in both forward and backward directions, causing both positive and negative half-cycles of the AC to be applied to a load.

Fig. 12.26 shows a circuit used for light-dimming purposes. The load is a light bulb. An RC phase shifter and diac-triggering device are used on the gate of the triac. Assume that the diac has a triggering voltage of 30 V. If the voltage

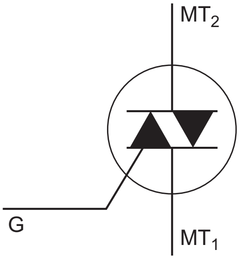

FIGURE 12.24 Schematic symbol for triac.

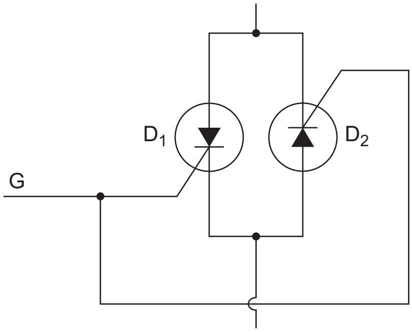

FIGURE 12.25 Equivalent circuit of triac shown with SCRs.

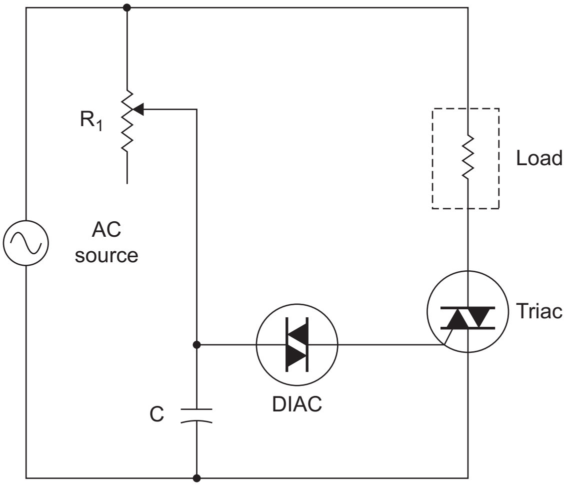

FIGURE 12.26 Lamp dimmer circuit using RC phase shifter and diac to control current in the load.

across the diac is less than 30V, it will not conduct. As soon as the 30-V point is reached or exceeded, the diac rapidly conducts and fires the SCR. Keep in mind that the diac is a bidirectional device and that current will flow through it in either direction if the bias voltage in either direction exceeds the triggering voltage level.

The triac will conduct on both positive and negative half-cycles of the applied AC. The phase shifter made up of R_1 and C will set the triggering delay. It will determine at what point in the AC cycle the diac will fire and apply gate current to the triac. Fig. 12.27 shows the waveforms at two different phase-shift points in the circuit. Note that current flows through the load in both the positive and negative half-cycles, but the duration of current flow depends on the point where the gate triggers. In Fig. 12.27, the average current through the load will be relatively high at a 60-degree phase shift, and the bulb will be

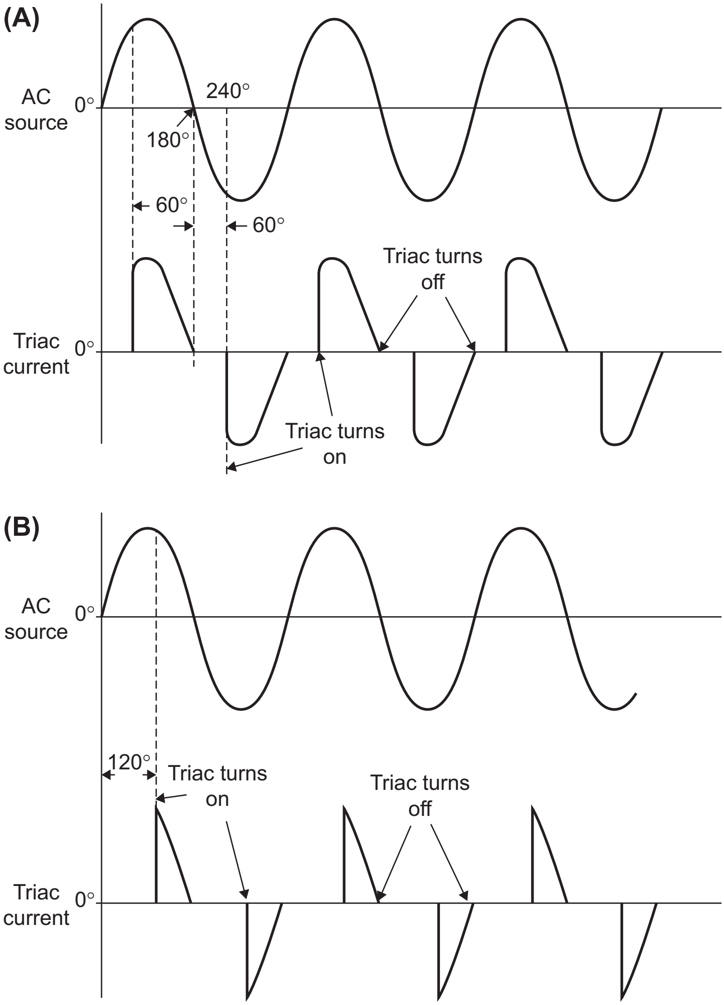

FIGURE 12.27 AC waveforms in a lamp load controlled by a triac. (A) Waveforms with triggering at 60 degrees, (B) waveforms with triggering at 120 degrees. The longer pulses produce higher average current in the load.

bright. At 120 degrees, the triggering point is late in each half-cycle, allowing only a small amount of current to flow through the bulb. The bulb, therefore, will glow dimly. Adjusting R_1 will allow the brightness of the bulb to be varied from nearly full on to full off. This is the circuit normally used in household and industrial light dimmers.

This same circuit can be used to control the speed of AC motors. In connecting a hand drill as the load, for example, R_1 will vary the speed of rotation.

Triacs are also widely used as solid-state relays.

PROGRAMMABLE LOGIC CONTROLLERS

A very popular type of controller used in industry and process control is a specialized computer called a programmable logic controller (PLC), or programmable controller. It uses sensors to monitor physical variables and generates output control signals to operate heating elements, motors, pumps, solenoids, and other devices. The PLC, however, contains a microprocessor that can be programmed in a special language that permits the PLC to be customized to the particular control function. In this way, a standard PLC can be programmed and configured for virtually any application. The PLC is by far one of the most popular and versatile control devices in use in industry.

The main function of a PLC is factory automation. It is a versatile special-purpose computer that can be quickly and easily reconfigured to meet changes and additions in factories, plants, and other operations that routinely need to control and sequence various machines and processes. While an embedded controller or personal computer could be used in some operations, they are harder to program and apply and are less flexible when rapid changes are needed. Furthermore, these common microcomputers are not designed for the factory environment. But PLCs are rugged and hardened against the harsh environment of most factories and plants. They can withstand the temperature ranges, vibration, and dirty conditions usually existing in a manufacturing or process control setting.

Organization

A general block diagram of a PLC is shown in Fig. 12.28. It is like most other microcomputer diagrams but there are some differences. A microprocessor is at the heart of the design. This CPU is coupled to a ROM where a special control program is stored. This program implements the basic scan operational mode of

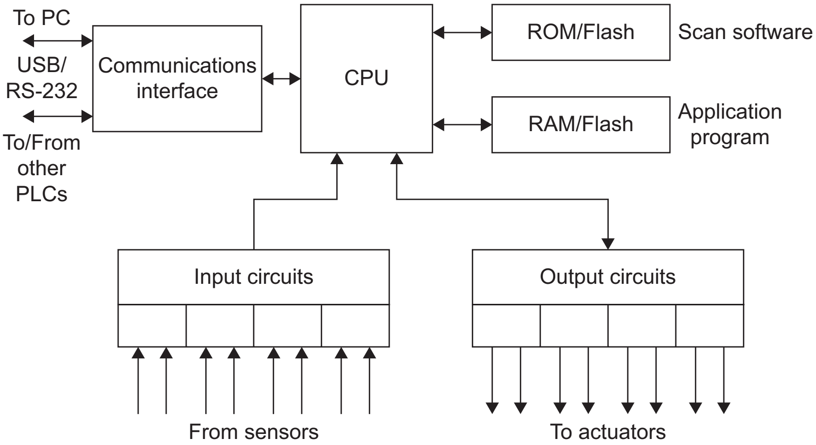

FIGURE 12.28 General block diagram of PLC.

the PLC. RAM is also available to store the application program. Flash memory may also be used in place of RAM.

The key to the PLC's versatility is its extensive I/O section. It has a variety of input and output modules. These are fixed in small, single-function PLCs, but in the larger more flexible units, the PLC is designed to accept a wide variety of separate I/O modules. These usually plug into a rack containing the main PLC unit and power supply. The different I/O modules let the user customize the PLC to the specific application.

Some examples of I/O modules are digital inputs and outputs with common logic levels or conditioned levels, analog input modules that accept analog signals from sensors, and analog output modules that generate analog outputs for controlling actuators of almost any kind. A huge variety is available. A communications interface is also available so that the PLC can talk to external devices such as a PC or other PLCs that are connected to form a larger control system. RS-232 and RS-485 serial interfaces are common as is USB. The program describing the operations to be implemented is developed on an external PC, and then downloaded to the PLC RAM or flash memory via the USB or RS-232 interfaces.

Operation

The general operation of a PLC is shown in Fig. 12.29. When the PLC is started, the internal operational program looks at the application program and interprets what is to be done. The sequence of operations is to first scan the inputs. The PLC looks at what the various sensors are doing. It gets a status

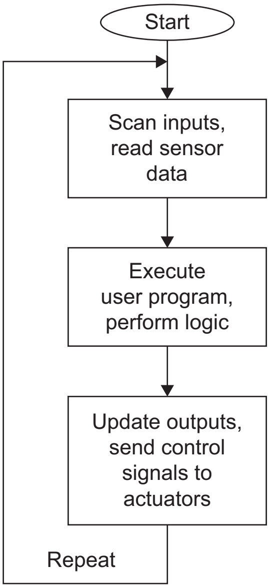
Start
Scan inputs, read sensor data
Execute user program, perform logic
Update outputs, send control signals to actuators
Repeat

FIGURE 12.29 PLC operational scan cycle.

update from the application itself. The program then interprets what to do next based on the application program. It processes the input to determine what the outputs should be. Then it updates the outputs to carry out the application. Then the cycle repeats itself at a high rate of speed. The inputs are scanned and outputs are updated every 1 ms or so, slower or faster as the application requires.

A simple example is using a switch to turn on a motor. The switch contacts are connected to an input module by wires. Output wires connect to a relay that in turn is used to operate the motor. As the PLC is scanning, it sees that the switch is off, so it does nothing. But as soon as the switch is turned on, the next scan detects this new condition. The program then tells the CPU that if the switch is on, turn on the relay. An output signal is sent to the output module, which then actuates an external relay. The relay contacts then turn the motor on. If the switch is then turned off, the next scan detects this condition and causes the output to turn off the relay and motor.

Programming

As for programming, PLCs can be programmed in various ways. Most PLC vendors have special languages to do the programming. In some cases, standard BASIC or C languages may be used. One widely used approach is a graphical language called *ladder logic*. Ladder logic is the technology that the PLC replaced decades ago. Control systems were made with relays, and they were wired so the diagram looked like a ladder (see Fig. 12.30). The two vertical lines represent a DC or AC voltage source. Circuits are then connected between these two "rails," forming rungs on the ladder. The circuits are connections of switches, relays, and other components. The software allows the user to build the ladder logic program on the screen, while a translation program in the PC converts it into the binary code that the PLC can execute.

This particular circuit in Fig. 12.30 does the following: when the ON pushbutton is pressed, current is applied to relay coil 1CR through NC contacts of SW1. SW1 is a thermal switch whose contacts stay closed as long as the temperature is less than 300 degrees Fahrenheit (300°F). If the temperature rises above that level, the contacts open, thereby breaking the circuits. Relay contacts 1CR around the ON pushbutton latch the circuit on when the ON pushbutton is released.

In the next "rung" of the ladder contacts 1CR are on the relay. They close when the relay turns on. This applies voltage to a heating element. This current flows through the NC contacts of thermal switch SW2. A green light in parallel with the heating elements turns on to indicate the heater is on.

When the heater temperature reaches 200°F, SW2 opens, turning off the heater and the green light. As the heater cools, SW2 will then close again, applying voltage to the heater. The heater will then cycle off and on keeping its temperature near the 200°F value.

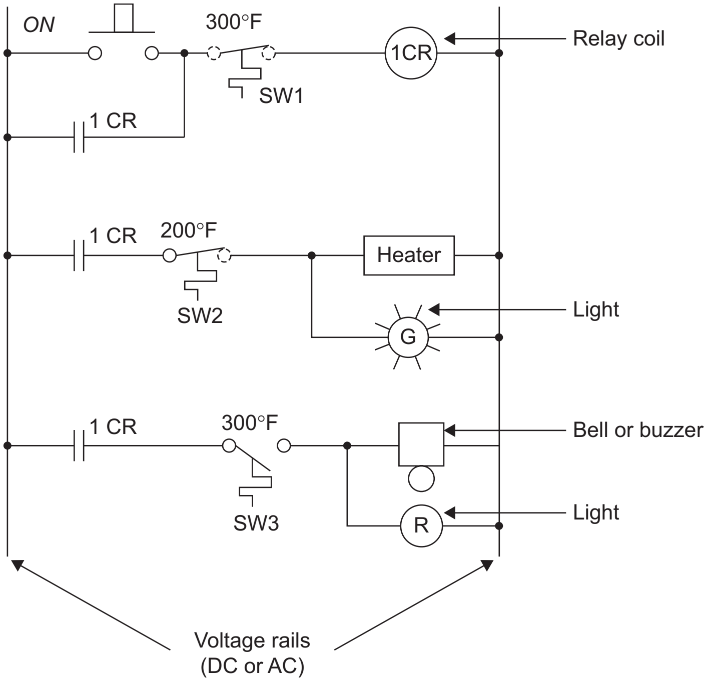

FIGURE 12.30 Sample ladder diagram and PLC application program.

If the temperature should rise above 300°F due to a failure of SW2 F, then SW1 opens. Relay 1CR is turned off, and its contacts open, releasing the latch on rung 1 and opening the circuit on rung 2. Thermal switch SW3 closes at 300°F, turning on a red light and ringing an alarm bell.

This example could be programmed with a graphical ladder logic program or with special commands and syntax of a traditional programming language.

In addition to being able do simple on/off tasks common to industrial control, special hardware in the PLC can also implement counters and timers so that external events can be counted and the sequencing of events can be programmed.

Data Acquisition Systems

A data acquisition system, often abbreviated DAQ, is a collection of circuits and equipment used for capturing data from multiple sensors when testing or monitoring conditions in a specific product or system. While testing automobiles or aircraft, it is common practice to measure multiple points in the target product to see how it performs to specifications or to detect problems. Common physical factors monitored include temperature, stress, pressure, humidity, mechanical position, liquid flow rate, and others related to the system. Industrial factories and process control plants such as refineries use DAQ to keep an eye on the state of the systems and its performance.

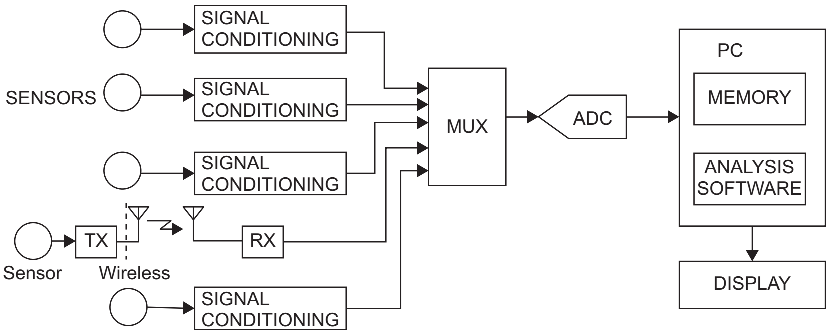

FIGURE 12.31 The key components and organization of a typical data acquisition system.

The main components of a DAQ are shown in Fig. 12.31. The inputs are from sensors that are usually connected by cables to signal conditioning circuits to prepare their outputs to be digitized. Signal conditioning includes amplification, filtering for noise, level shifting, or other corrections. In some applications the sensor may be far from the DAQ system in which case a wireless link can be established using Wi-Fi or other wireless technology.

The processed sensor signals then go to a multiplexer (MUX) that selects the sensor to be digitized and passes the signal along to an analog-to-digital converter (ADC) that samples the analog signal from the sensor and converts it to a stream of binary values, usually 8 to 16 bits long. Sampling rates are usually low, from several samples per hour to 1 MS/s. The resulting data words are then sent to a PC and stored in a file. The data can then be analyzed by software, used in analysis or converted into appropriate displays or graphs.

Project 12.1

Build a Robot

Robots are great examples of basic industrial control. While most industrial robots are usually manipulator arms, you can actually buy a mobile robot with wheels that can be programmed for motion and other functions. A wide range of kits are available. It is a great way to get familiar with sensors, a microcomputer controller, and various actuators. Typical inputs are switches, photo cells, temperature sensors, and common outputs are relays, lights, and motors.

Some recommended kits are those supplied by Lego and Parallax. Suggested websites to learn more follow:

Jameco: www.jameco.com
Lego Mindstorms NXT: www.mindstorms.lego.com
Parallax: www.parallax.com

Two magazines regularly cover robots from a hobbyist/experimenter viewpoint. These are *Nuts & Volts* and *Servo*. You may be able to find these at a newsstand; otherwise, go to their websites:

Nuts & Volts Magazine: www.nutsvolts.com
Servo Magazine: www.servomagazine.com

Project 12.2

Build a Feedback Control System

Feedback control systems as described in this chapter are very common in industrial electronics. This project lets you build one and demonstrate its operation. The basic circuit is shown in Fig. 12.32. A heating element, the 50 Ω resistor, is monitored by a thermistor. If the resistor gets too hot, the thermistor value will decrease and change the input voltage to a comparator. Voltage divider R_2-R_3 provides a 3.75 V set point to the comparator. The comparator will switch its output high turning on the MOSFET that, in turn, turns on a fan. The fan then cools the heating element. This causes the temperature to decrease and the thermistor value to go back up. The comparator switches again. Its output goes low turning off the MOSFET and the fan. The process then repeats automatically.

Build the circuit on your breadboarding socket. The parts list is given in Appendix B. You will need a 12 V supply for this project. An electronic supply is preferred but you can use two of your AA cell 6 V supplies in series. Use the long leads of the 50-Ω resistor to keep the body of the resistor away from the breadboarding socket as the resistor will get very hot and melt the plastic. Be careful as the resistor will burn badly. Physically place the fan several inches away from the resistor. The resistor should also be positioned within inches of the thermistor. Check your wiring to be sure there are no errors.

Apply power to the circuit. The 50 Ω resistor will draw 240 mA of current that will quickly deplete the charge on the batteries. Check the supply voltage with

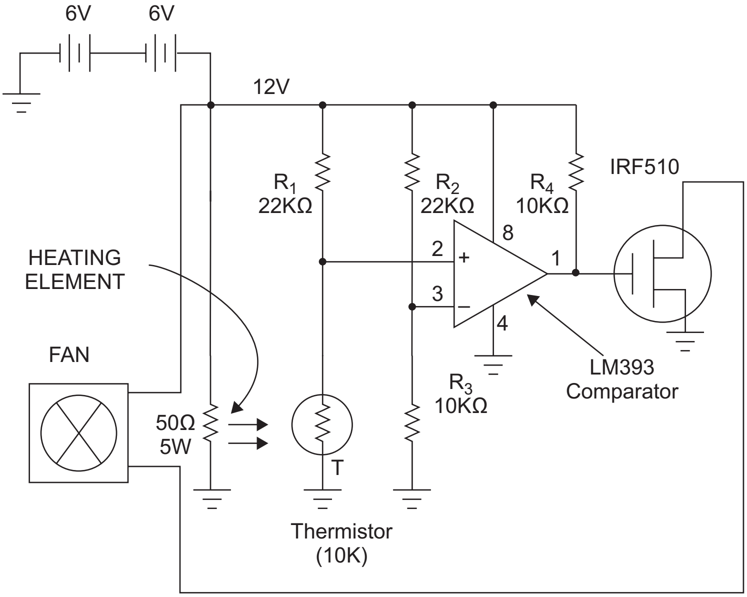

FIGURE 12.32 A simple feedback control circuit for temperature control.

Continued

Project 12.2—cont'd

your multimeter to be sure it is 12 V. Within a minute or so, the fan should come on. If it does not, reduce the distance between the resistor and the thermistor. Once the fan turns on, it will cool the resistor within a minute or so. The fan will then turn off. Experiment with the distances between the resistor and the fan and thermistor until the circuit works as described. Again, be careful as the resistor gets seriously hot. Turn off the power.

Chapter 13

Software, Programming and Electronics

Today, Most Electronics Is a Combination of Hardware and Software

In this Chapter:

- Micros in every electronic product.
- What is software?
- Types of software.
- Programming languages for embedded controllers.
- Learning to program.
- Micro projects.

INTRODUCTION

Electronics equipment used to be strictly hardware. However, over the years, vacuum tube circuits on a metal chassis gave way to transistors on printed circuit boards (PCBs) that, in turn, gave way to large-scale integrated circuits on smaller multilayer PCBs today. And that brings us to the key point of this chapter. That is, virtually all modern electronic products contain at least one microcontroller also known as an embedded controller.

The secret to making that microcontroller implement the product's functions is software. Those microcontrollers are simply pieces of hardware that have no specific application until they are given a program that dedicates the chip to the product's characteristics. This chapter picks up where Chapter 6 ends and then expands on the subject of software and introduces programming languages of interest. Projects using available microcontroller hardware and software are included.

SOFTWARE FUNDAMENTALS

Software is the general term that describes any program that tells a computer or microcontroller chip what to do. A program is a sequence of binary codes stored in memory that represent instructions that define specific operations that the

Electronics Explained. http://dx.doi.org/10.1016/B978-0-12-811641-8.00013-8
Copyright © 2018 Elsevier Inc. All rights reserved.

computer is capable of performing. The computer or microcontroller unit (MCU) is the visible hardware while the programs are the invisible magic ingredient buried in memory. The combination defines a specific product. In addition to the microcontroller and its programs are the external components and devices that serve input and output operations specific to the product. Our discussion here focuses on microcontrollers rather than general-purpose computers. Refresh your knowledge of microcontroller organization and operation in Chapter 6.

The microcontroller can perform almost any digital function from the simplest logic operation to a complex mathematical calculation. For many products, dozens of digital logic chips can be replaced by a single microcontroller. The only problem is that you have to tell the MCU to do what you want. And you do that by programming. A program is the sequence of instructions that detail the order in which you do the various processing steps that lead to your application. The microcontroller chip itself is a dumb piece of hardware waiting for its program. What this means is that today, designing an electronic product is as much about writing programs as it is building electronic hardware.

Software is the name given to all the programs that a processor uses for developing and executing a program. The program that implements your product's functions is the applications program. Other types of software used with the MCU are as follows:

Language—A programming language is a set of rules or syntax that defines how you tell the processor what to do. A language is the tool that you actually use to create your application program. It is a kind of shorthand code that lets you express the operations in simple terms.

Machine code—This is the binary words that represent the instructions of a program. It is what tells the MCU hardware what to do and when. Hex characters are commonly used instead of binary for programming clarity and simplicity.

Assembler—An assembler is a programming language based on the instruction set native to the MCU. It is a collection of terms called mnemonics that define specific hardware functions. A programmer uses the instruction terms to create a step-by-step program. It is the assembler that translates the assembler code into machine code.

Compiler—A compiler is the program that translates your program code into binary code that can be stored in the memory and used by the processor. Processors only understand binary code, but humans are not good with such code. So, you write your program in a language you can understand, and the compiler converts it to the code that the processor understands. The compiler is not used to execute the program.

Interpreter—An interpreter is a program that resides along with the application program in memory and executes the application program one line of code at a time. The interpreter translates the instructions of the language into machine code to be executed by the MCU.

Operating system—For large complex programs or for applications that require the processor to execute multiple programs at different times, you need a piece of software that manages the multiple programs and operations. The operating system (OS) oversees things such as memory usage and access, timing, and I/O operations. Not all embedded controllers need an OS, just the more powerful ones with multiple functions. For example, most cell phones have an OS.

Editor—An editor is like a word processing program that lets you type in your program using the language syntax. You then save it as a file that the compiler or assembler uses to create machine code.

Integrated development environment (IDE)—This is a complete package of software tools that facilitates software development. It consists of an editor, languages, and compilers, as well as programs for testing and debugging your code.

Debugger—This is a feature of most IDEs that makes it easy to find problems in a program such as syntax errors, misused instructions, address conflicts, timing problems, and others.

Subroutines—These are short sections of code that perform a specific operation. It may be a math operation, some type of I/O process, or other function that is performed multiple times. You only write the subroutine once, and then access it when you need to do that operation.

Libraries—Many IDEs have libraries of subroutines that you can tap so you do not have to program common functions yourself. Subroutines shorten your development time and speed execution.

Source code—This is the program that you write with the editor and store in a file. It is then given to the compiler or interpreter.

Object code—This is the machine code that is the output of the compiler.

Programming Languages

There are dozens of languages used for programming embedded controllers. The more popular ones are assembly language and C. An older language BASIC is not widely used in industry, but it is popular for learning programming for microcontrollers. Here is an overview of each.

Assembly Language

Assembly language uses shorthand mnemonics such as three- or four-letter designations for the instructions that the processor can execute. Instead of programming with binary code, you use the mnemonics. In addition, you use names that you make up to identify variables, constants, memory locations, I/O ports, and other things normally referenced by a binary code. Following is what a simple assembly program looks like:

```
ORG $0000
LDAA FIRST
MUL SECOND
STA PRODUCT
OUT PORT 1
END
```

This program loads one number from a memory location given the name FIRST, multiplies it by another number in memory location named SECOND. It then stores the result in another memory location called PRODUCT. It next outputs the value to an I/O port called PORT 1. The ORG $0000 simply says to start the program whose first instruction is in memory location 0000, a hex address.

The software that converts the assembly language program to binary code is the *assembler*. The assembler gives actual binary assignments to the memory locations named as well as the instruction op codes.

BASIC Language

Some embedded controllers use a simplified version of BASIC. BASIC was invented before personal computers but then initially became the main language of personal computer programming. Microsoft still sells its Visual BASIC software for program development. BASIC is a very simple higher-level language that is easy to learn and use. The program created is then translated into binary code by a compiler. Alternately, the microcontroller runs a program called an *interpreter* that resides in memory and executes the program, a command at a time.

The following program is written in PBASIC, a simple version of BASIC used with an embedded controller called the BASIC Stamp made by Parallax. It uses a PIC microprocessor and stores a program in flash memory. The program causes an LED to flash off and on.

```
Do
    High 7
    Pause 500
    Low 7
    Pause 500
Loop
```

The first line tells the controller to make the output pin 7 on the microcontroller high or binary 1. The next command says pause for 500 ms or one half second. Next, pin 7 is made to go low or to binary 0 via the Low command. Pause makes it stay there for another 500 ms. Finally, the Loop command causes the program to loop back and repeat itself again and again. This is what is called a Do…Loop for repeating a program over and over again. The result is an LED that turns off and on at a 1000-ms (1-s) rate.

C Language

The C language is referred to as a higher-level language in that the various syntax and terminology of the language usually translates into multiple binary machine instructions. The C language lets you write the program in a more conversational or casual format using English instead of cryptic mnemonics and other terms. The language is a bit of a challenge to learn, but after that programming goes much faster. A compiler converts the C language program code into binary code.

A simple C program may appear like this:

```
#include <stdio.h>
#include “stdio_setup.h”

int main (void)
{
UartInit();
int x, y, z;
x = 38;
y = 57;
z = y – x;
printf(“The difference is %d”, z);
}
```

The #include statements tell the compiler to include instructions to accomplish some of the actions specified by the program. The int main(void) is something that must be included in all C programs. The UartInit() tells the compiler to set up an output port for the result of the program. An open bracket symbol indicates the beginning of the program. The next line identifies three integer values x, y, and z; 38 is assigned to x, and 57 to y. Then the difference between the two is designated z. Finally, the printf line accesses a subroutine that causes the expression “The difference to be printed is 19.” The z value is substituted for %d indicator. Note that all lines of the program end with a semicolon. The end of the program is indicated by the close bracket.

PROGRAMMING

Programming is the process of solving a problem with a computer. Sometimes called coding, programming is a multistep process that involves identifying and defining the problem, developing a solution to it, and then preparing a sequence of instructions that solve the problem.

There are five steps in the programming process. These are as follows:

1. Problem definition
2. Solution development
3. Program coding
4. Program testing and debugging
5. Documentation

In problem definition, you state as clearly as possible what you want to accomplish. Typically it is best to make a written statement of the problem. By actually writing out what is to be accomplished, you will think through the problem more thoroughly. Making charts, tables, graphs, drawings, or the like is also useful. Define your inputs, define the desired outputs, and then outline the processing steps required. Thorough problem definition produces a solution faster and easier. This is the most important step in programming.

Once the problem has been defined, you are ready to develop a solution. It is virtually impossible to give you guidelines on how to do this. Developing the solution is really the creative part of programming. The best way to learn it is by actually doing some programming. The more problems you solve, the more creative and proficient you will become. In any case, developing the solution is best done when the problem has been clearly defined and when you have a good working knowledge of the MCU, its instruction set, and the programming language you are going to use.

The solution you develop is typically a step-by-step procedure. Each step in the solution will cause some part of the problem to be solved. A step-by-step sequence of procedures for solving a specific problem is called an *algorithm*. An algorithm may be a mathematical formula or just a recipe-like list of steps.

A useful tool in developing a solution to your programming problem and in preparing your algorithm is the *flowchart*. A flowchart is widely used by programmers to provide a visual or graphical presentation of the solution. Sometimes the solution is more easily seen or understood by blocking out a problem in graphical or visual form.

Flowcharts are made up of a number of simple symbols that depict typical computer operations. Fig. 13.1 shows the most commonly used symbols. The two most important symbols are the rectangle and the diamond. The rectangle simply specifies some particular operation. The diamond is a decision block. It usually asks a question that can be answered with a yes/no or true/false decision. This is the way most computers "think" or make decisions.

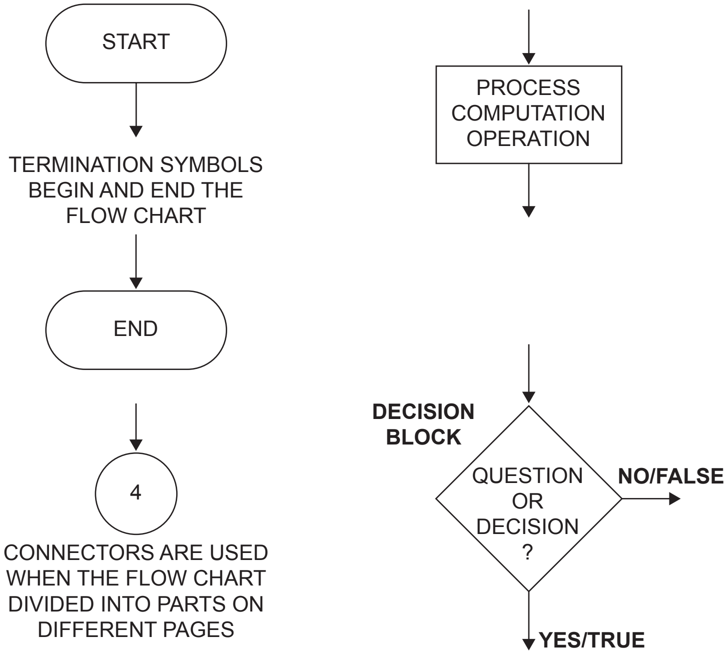
START
TERMINATION SYMBOLS BEGIN AND END THE FLOW CHART
END
4
CONNECTORS ARE USED WHEN THE FLOW CHART DIVIDED INTO PARTS ON DIFFERENT PAGES
PROCESS COMPUTATION OPERATION
DECISION BLOCK
QUESTION OR DECISION ?
NO/FALSE
YES/TRUE

FIGURE 13.1 Flowchart symbols.

An example of a flowchart is given in Fig. 13.2. This flowchart represents the solution to a typical MCU problem. The problem in this example is to determine which of two numbers, A or B, is the larger. The algorithm is to subtract B from A. If the difference is positive, A is obviously greater than B. On the other hand, if the subtraction results in a negative difference, A is less than B, which means B is the larger. The flowchart illustrates the entire problem and solution.

One of the best approaches to programming a solution is to use *structured programming*. Also known as *top-down design*, structured programming is the process of taking the problem and partitioning it into modules that perform the major functions of the solution. In structured programming, you break down a

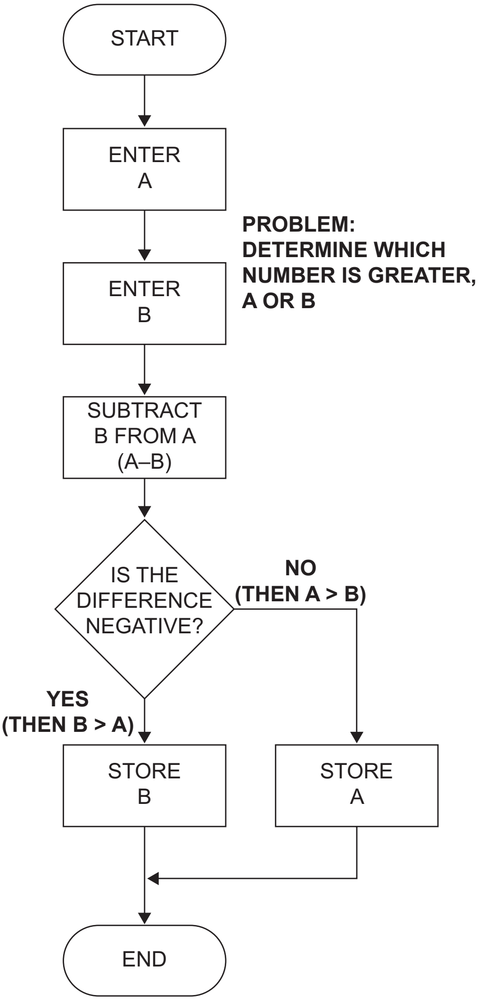
START
ENTER
A
PROBLEM:
DETERMINE WHICH
NUMBER IS GREATER,
A OR B
ENTER
B
SUBTRACT
B FROM A
(A–B)
IS THE
DIFFERENCE
NEGATIVE?
NO
(THEN A > B)
YES
(THEN B > A)
STORE
B
STORE
A
END

FIGURE 13.2 A flowchart illustrating a computer program.

problem into smaller and smaller logical pieces. Each piece, or module, performs some distinct segment of the solution.

All of the modules are then linked by a hierarchical structure that shows their relationship. This structure looks like an organization chart for a company. An example is given in Fig. 13.3. The main solution is divided up into three major modules. These modules are then divided into submodules that implement the primary module with smaller stand-alone programs. These submodules may then be further subdivided into even smaller logical segments. Some modules may be subroutines that may be used several times or called by any of the other modules or submodules.

Another feature of structured programming is to minimize or, if possible, completely eliminate unconditional branch operations. An *unconditional branch* or *jump* is a command or instruction that changes the order of a program from its normal, logical step-by-step sequence to something that might not be obvious. A common unconditional branch is the GOTO instruction. The tendency is to end one module with a GOTO so that the program branches to a submodule or some other logical segment of code. If too many GOTOs are used, the result is what is normally called spaghetti code. The result is a tangle sequence that is hard to follow in debugging or in maintaining the program.

The benefit of structured programming is that by dividing the program into smaller logical segments, it is easier to code, test, debug, and maintain. Also,

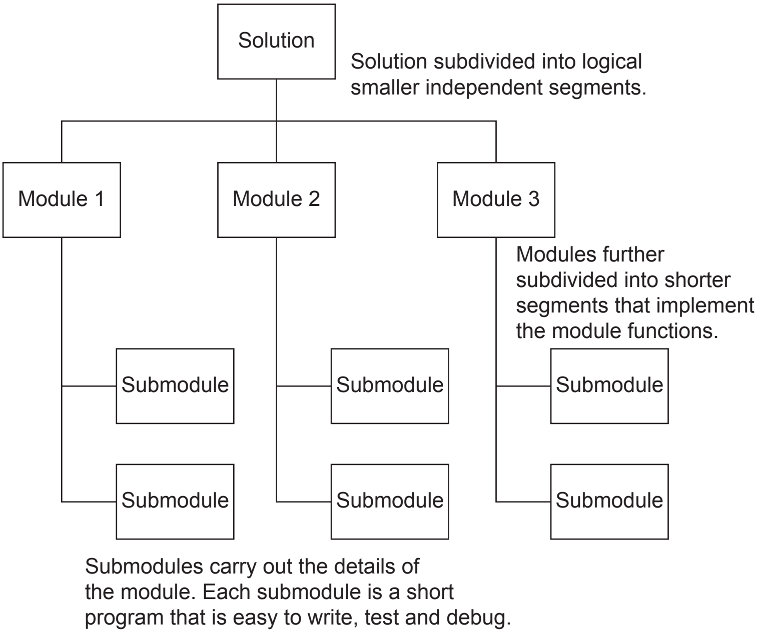
Solution
Solution subdivided into logical smaller independent segments.
Module 1
Module 2
Module 3
Modules further subdivided into shorter segments that implement the module functions.
Submodule
Submodule
Submodule
Submodule
Submodule
Submodule
Submodules carry out the details of the module. Each submodule is a short program that is easy to write, test and debug.

FIGURE 13.3 Structured programming divides a problem into smaller and smaller submodules.

it is easier to work with a smaller program when devising a solution. So if you take the time to do the partitioning up front when designing the program, the actual development time will be shorter, and chances are that the program will work right the first time.

The next step in the programming process is *coding*. This is the process of converting your solution, algorithm, or flowchart into an actual computer program. Here you take the instructions, expressions, or statements of the language you are using and write them in a sequence that implements your algorithm.

A common procedure with many programmers is to first code the solution in pseudocode. *Pseudocode* is a hypothetical or made-up language that is some hybrid of an English language statement of the operation and the actual instruction, statement, or command of the programming language. It enables the programmer to write the sequence of operations on paper first before actually using the language itself. Often the pseudocode can be written from the flowchart. This procedure helps you better see the logic of the procedure before committing to the final program code.

Programming involves much more than just the coding process, which most people associate with the term. Actually, coding usually takes roughly 20% or less of the total time involved in programming. Studies have shown that 40% of a programmer's time is usually devoted to problem definition, structure design, and the development or identification of an algorithm.

Most programmers agree that coding is actually the easiest part of programming. Once you become familiar with the characteristics and syntax of the language of use, and some basic programming tricks, coding becomes fast and easy. The use of subroutines greatly shortens and simplifies the coding process. On the other hand, defining the problem and developing a workable solution is much more difficult.

Once the coding is complete, you will enter the program into the MCU and test it. You enter the program with the editor to create your source code. The computer will then assemble or compile your program into machine code. In either case, the result is the binary object code stored in memory. At this point you are ready for the next programming step, testing and debugging.

Testing involves verifying that your program works. You will want to thoroughly test it to be sure that it does what it is supposed to do. Invariably, most programs do not work properly the first time. You may have made logical errors in sequencing, or perhaps you made errors in entering the code. If you used structured programming, you can test each submodule or subroutine by itself before linking together all of the various modules to perform the total operation.

Whatever the problems are, testing leads to *debugging*. Debugging is the process of working with the computer program and attempting to find errors and glitches. Most languages and program development software have built-in features that facilitate the debugging process. Once all the errors have been found and corrected, the program is ready to use.

The final step in the programming process is called *documentation*. This is an often-neglected but extremely important part of programming. Once the program is written and debugged, it is ready to use. The temptation then is to put the program into use and drop the project. However, it is extremely important that you document the program. Your problem definitions and solution work should be written up and filed away. Any charts, graphs, drawings, flowcharts, and other materials that you developed should be saved. All program listings should also be retained. In fact, it is desirable to prepare a detailed, written statement explaining the algorithm and the procedures involved. This is important for program maintenance, because it may be necessary to change the program at a later date. You may wish to add a feature or correct some part of the program to accommodate new conditions. It is easy to forget what you did months or years earlier. Also the original programmer may no longer be involved with the program. If a new programmer becomes responsible for changing a program, complete documentation is invaluable. Changing a program written by someone else is extremely difficult and, in some cases, virtually impossible without documentation.

The major investment in any programming project is programmer's time. A significant amount of time is usually required to develop a useful piece of software, which is why software is so expensive. In fact, the hardware is the least expensive part of the system in most embedded controllers.

One final word: Some programs may have to be changed over time. As the problem evolves, grows, or changes, the solution will have to be altered to meet current needs. New issues must be faced, or features may need to be improved to save time. In the case of a software product, improvements may have to be made to fix user-discovered bugs, improve performance, or keep the software competitive in the marketplace. It may be necessary to port the software to another MCU or OS.

All of these conditions fall into the category of maintenance. In fact, more time is spent on maintenance than in creating the original software. And, as indicated earlier, the better the documentation the faster and easier the maintenance can be carried out.

Developing Software for Embedded Controllers

The programming process just described applies to any microcomputer, be it a general-purpose personal computer or an embedded microcontroller. However, the procedure is somewhat different if you are developing programs for MCUs that are to be built into other equipment. Fig. 13.4 shows a block diagram of the complete microcontroller software development process.

A typical arrangement is to install the appropriate IDE on a PC. The PC then connects via a USB port to a development or evaluation board containing the target MCU. Next, once you have written your program in the programming language with the text editor, the assembler or compiler is called on to translate

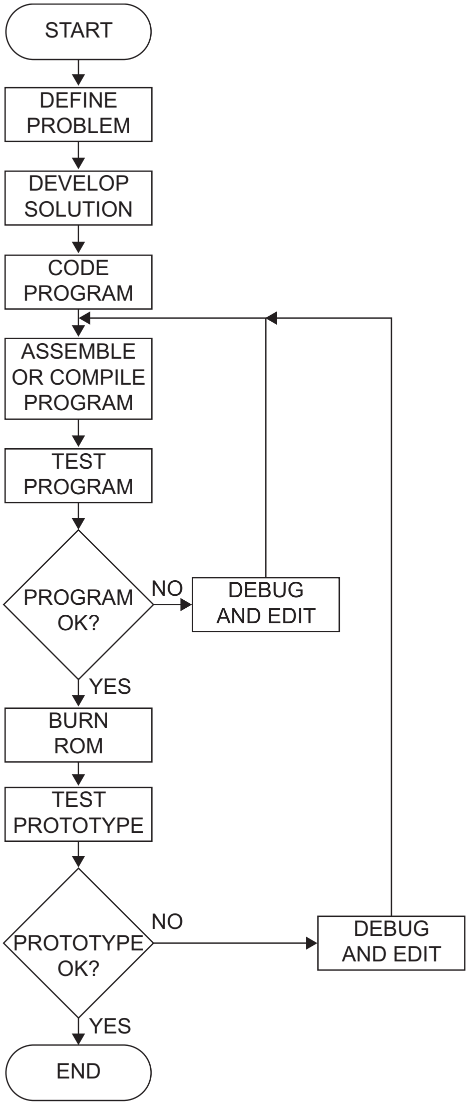
START
DEFINE PROBLEM
DEVELOP SOLUTION
CODE PROGRAM
ASSEMBLE OR COMPILE PROGRAM
TEST PROGRAM
PROGRAM OK?
NO
DEBUG AND EDIT
YES
BURN ROM
TEST PROTOTYPE
PROTOTYPE OK?
NO
DEBUG AND EDIT
YES
END

FIGURE 13.4 Block diagram of the complete microcontroller software development process.

that program into the machine code for the target microprocessor. All this is done on the PC.

Once the program has been written and the machine code generated, you test the program. The personal computer is not capable of executing the microprocessor instruction codes directly. However, an emulator program is usually available in the IDE that will allow the personal computer to test the program. An *emulator program* is a system-level program that duplicates the execution of the MCU instructions on the personal computer. The emulator mimics the microprocessor so that the object code developed under the assembler or compiler can be tested. This allows the programmer to do any initial debugging. The problem with this procedure, however, is that the emulator sometimes does not operate in real time. It will not precisely duplicate the machine-cycle instruction execution rates of the microprocessor. The program will not be executed in exactly the same time frame as with the microprocessor. If time-related functions are important, the program should be thoroughly tested on a prototype of the MCU before being finalized.

Most development software includes a debugger, a program that helps to identify and isolate coding errors, logical errors, and other problems. The next step is to download the program to the MCU. This is usually done by way of a USB cable that connects the PC to the MCU development board or a prototype of the product being developed. Functional testing can then take place. The program is downloaded to the development board and stored in a ROM or flash memory. Thus the microprocessor becomes dedicated to a specific application. Then additional testing can be performed. The program in a ROM is typically referred to as *firmware*, that is, software that exists in an electronic component.

Microcontroller Unit Interfacing

In addition to programming an MCU for a specific task, you also need to interface the MCU to the external devices that make up the device you are designing. Such peripherals are sensors such as light or temperature sensors or controlled devices such as LEDs, relays, or motors. Usually some form of circuit or device is needed between the MCU input/output lines and the peripheral devices.

Outputs. The I/O lines of the MCU input or output standard binary voltage levels. These are typically in the 1.8–5V range for a binary 1 and 0 to a few tenths of a volt for a binary 0. Depending on the MCU and supply voltage, the output lines can usually sink or source current in the 15–40mA range. As a result, the output line can operate low-power devices such as LEDs directly

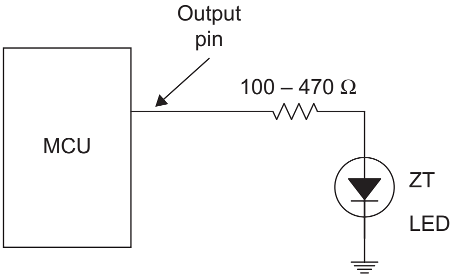

FIGURE 13.5 Driving an LED directly from the MCU output.

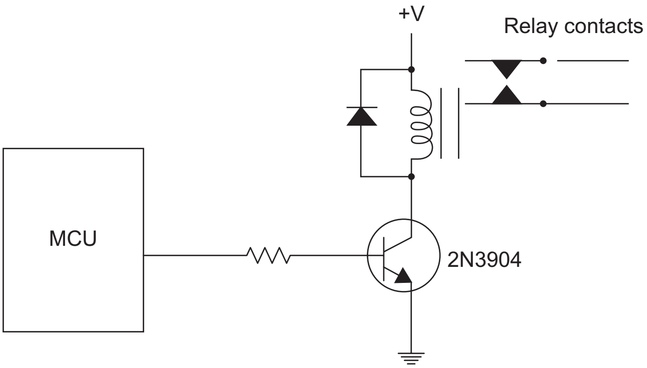

FIGURE 13.6 An NPN driver transistor for higher loads.

without an interface. Fig. 13.5 is a simple example. The resistor limits and sets the current level.

For higher current/voltage devices an interface transistor is usually needed. Fig. 13.6 shows an NPN transistor driving a relay coil. The coil may be operated by the same power supply as the MCU or a higher voltage. The diode protects the transistor from the inductive spike caused by the inductive coil being deenergized. This simple interface can also control a solenoid.

Fig. 13.7 shows a MOSFET operating a motor. A protective diode is usually built into the transistor. If bidirectional control of the motor is needed, the H-bridge circuit discussed in Chapter 12 (Figure 12.20) can be used.

A useful output is pulse width modulation (PWM). A PWM output from an MCU is a continuous DC rectangular wave whose duty cycle can be varied by the program. Some devices or circuits take this signal and average it into a continuous DC voltage. Varying the duty cycle causes the DC average to vary. (See Duty Cycle and PWM and Figure 4.22 in Chapter 4). The duty cycle is selected in the program usually by setting a numerical value. The duty cycle can also be changed by the program. Two example applications are varying the brightness of an LED and controlling the speed of a DC motor.

Inputs. A common input is a switch or pushbutton. The idea is to connect the switch into a circuit that generates a binary 0 or 1. A typical circuit is shown in Fig. 13.8A. The single-pole single-throw (SPST) switch is normally open. The pull-up resistor keeps the MCU input at +5 V. When the button is pressed, the switch contacts close grounding the resistor and producing a zero input. Instructions in the program read the status of the input line. Some action is initiated if the input goes low.

An alternate arrangement is shown in Fig. 13.8B. The resistor keeps the MCU input low or ground and binary 0 level. When the pushbutton is depressed, +5 V is applied to the input. Again the MCU tests the input with an instruction to see if it is binary 1. If it is, some other action will occur.

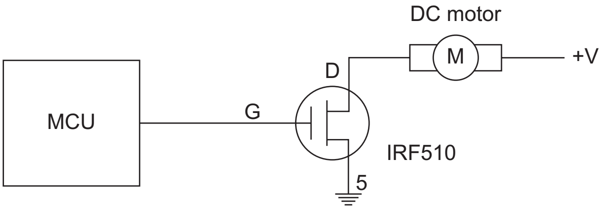

FIGURE 13.7 An N-channel enhancement mode MOSFET driving a DC motor.

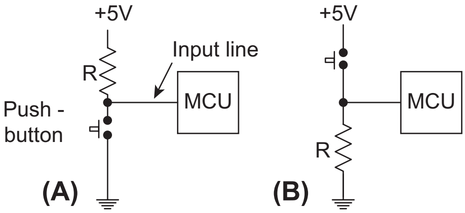

FIGURE 13.8 Using a pushbutton switch to generate an input logic level.

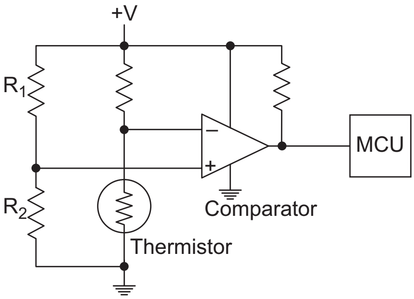

FIGURE 13.9 A comparator and thermistor used for temperature sensing.

A comparator can also be used for an interface. Fig. 13.9 shows a thermistor temperature sensor as the input to the comparator. Voltage divider R_1-R_2 sets the temperature level. As the temperature increases, the thermistor resistance decreases when the voltage from the thermistor divider changes the comparator input and switching the output from a 0 to 1 or vice versa. This signal is applied to an MCU input that is read by the program to determine if the temperature is above or below the level set by the pot. Other resistive sensors such as a photoresistive light sensor can be used for a similar function.

Some devices may require the use of an analog-to-digital converter (ADC). Some sensors or sensor circuits such as a voltage divider or bridge output a DC voltage proportional to the analog quantity being sensed. The ADC translates this voltage into an equivalent binary value that is sent to the

MCU. As it turns out, most MCUs have one or more internal ADCs so no interfacing is needed other than conditioning the analog signal for application to the MCU input.

PLATFORMS FOR LEARNING PROGRAMMING

The best way to learn programming is to start doing it. You will probably need to read some books, participate in online tutorials, or take a formal course. Whatever your approach to teach yourself programming, you will need a platform for learning. The platform, for lack of a better term, is a mix of hardware and software that is organized for learning. Many of these platforms are available for different languages and MCUs. I have reviewed several of these, and described here are the four that I can recommend. They teach either BASIC or C. For complete beginners I recommend those teaching BASIC. If you plan to use C in your work or just want to challenge yourself, select a C platform.

Parallax BASIC Stamp 2

The BASIC Stamp 2 is a complete microcontroller module that uses a Microchip Technology PIC MCU with RAM and flash memory on-chip in addition to inputs and outputs including a USB port for connecting to a PC for programming. It is designed to be built into other equipment or projects. The Stamp 2 is programmed in the PBASIC language. The Parallax Stamp has been around for about 17 years, so it is highly developed and widely used in education to teach programming.

The Stamp 2 is available in a wide range of forms from simple plug-in modules to complete learning kits that include a breadboarding socket, multiple components, and comprehensive project manuals. A free IDE is available for download from the Parallax site. I reviewed the BASIC Stamp Activity Kit that comes with the breadboard and parts. See Fig. 13.10. The manual is a self-teaching guide that instructs you in the BASIC language and electronic interfacing. The kit is a complete package of all the things you need to start learning programming. I recommend it.

PICAXE

PICAXE is another platform that uses a Microchip Technology PIC MCU and is programmed in BASIC. The package is essentially a version of the PIC MCU family in a DIP IC that can be plugged into a breadboarding socket. It has been preprogrammed with firmware that allows the MCU to be programmed by way of IDE software on a PC. Connection is by way of a USB port on the PC. You can then add other interfaces such as sensors or controlled devices as your project dictates. The BASIC language is easy to learn and use. The IDE is a

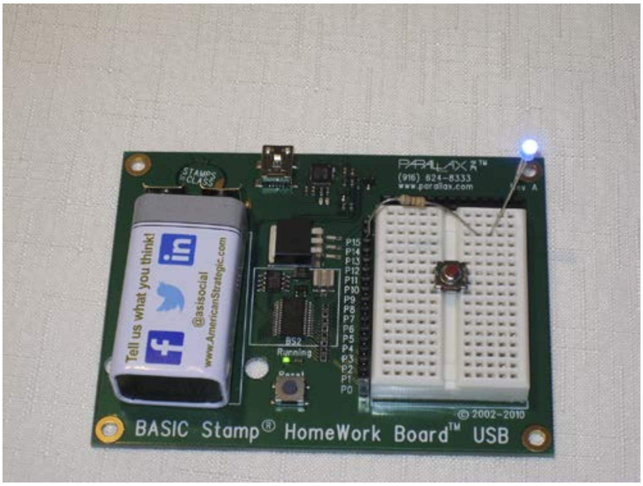
PARALLAX
(916) 624-8333
www.parallax.com
Rev A
STAMPS CLASS
Tell us what you think!
@asisocial
www.AmericanStrategic.com
P15
P14
P13
P12
P11
P10
P9
P8
P7
P6
P5
P4
P3
P2
P1
P0
BS2
Running
Reset
© 2002-2010
BASIC Stamp® HomeWork Board™ USB

FIGURE 13.10 The Parallax BASIC Stamp Activity Kit experimental board is designed to be operated from a 9-V battery. The PIC MCU is the IC above the reset button in the center.

free download, and there is plenty of support, help, and instruction on the companion website. This is a low cost way to build projects with a microcontroller.

Arduino

The Arduino is a wildly popular microcontroller module that uses an Atmel (now part of Microchip Technology) MCU. The module provides a USB port for connection to a host PC for programming and access to multiple input and output lines. It is programmed in a version of the C language. A free IDE is available for download.

The Arduino has been around for several years now, and multiple versions are available. Dozens of different products are offered by multiple companies. You choose the MCU type depending on your need for memory sizes, interfaces, and I/O ports. My recommendation is to select the simplest form called the Arduino Uno R3. It uses an 8-bit MCU called the Atmel ATmega328P. The Uno is powered by the USB port or an external 5V supply. A free online IDE is available. Using the C language variant provided, you write code to develop a program called a "sketch."

The C language is more daunting to learn than BASIC, but it offers far more power and capability than BASIC. If you want to create more complex projects as a maker or begin to use C in your career, then C is the better choice. The Arduino is probably the best learning platform as there is a large quantity of information available to help you including books, tutorials, and products. One popular type of accessory is the "shield," a small PCB that plugs into the connectors on the Arduino board. The shields are various forms of interfaces and I/O devices such as wireless, motor control, GPS, and many others.

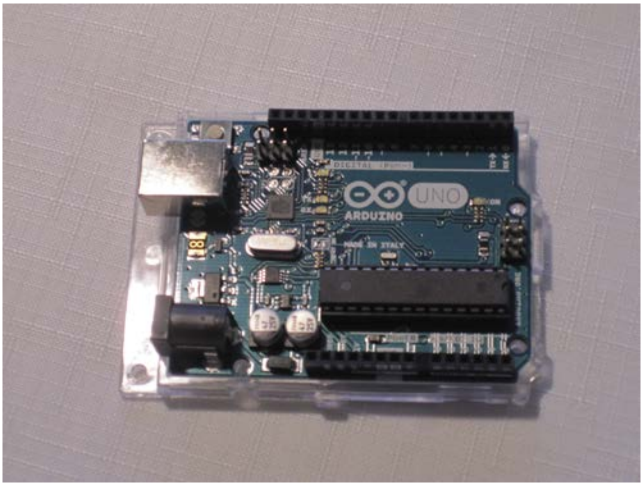

FIGURE 13.11 The Arduino Uno R3 is one of the most widely used microcontroller learning platforms.

I evaluated the Arduino by using supplier SparkFun's Inventors's Kit. See Fig. 13.11. It contains the Uno, a breadboarding socket and a batch of parts to use in interfacing and building external circuits. An instruction manual leads you through projects that demonstrate interfacing and C coding, a good choice for the beginning programmer.

LaunchPad

The LaunchPad is one of the Texas Instruments' development kits for the MPS430 series of MCUs. It comes as a PCB module with the MPS430FR6989 16-bit MCU that features 128 KB of FRAM. Refer to Fig. 13.12. The board also includes some LEDs, pushbuttons, an LCD readout, and other components to use in demonstration projects. A USB port connects the LaunchPad to the PC. Free IDEs are available for download. This is a very serious professional platform for those committed to learning C. No companion instructions are provided but many manuals and tutorials are available online. I acquired the LaunchPad as part of an online college course from Udemy that teaches C with lab projects.

Advanced Alternatives

If you are really adventuresome, you may want to select a more advanced platform. One of the most popular is the Raspberry Pi 2 or the newer more powerful Raspberry Pi 3. Both use a version of the popular ARM type processor, a powerful 32-bit microprocessor. It comes in the form of a development board with memory, I/O, and other resources. The Raspberry Pi uses variations of the Linux OS. IDEs

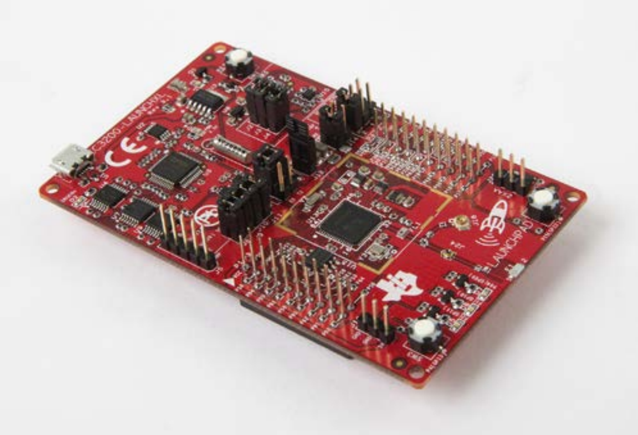

FIGURE 13.12 The Texas Instruments' LaunchPad development board uses an MPS430 microcontroller.

supporting different programming languages are available. Prices are very reasonable. I do not recommend these for beginners, but they can be the next step up after you have mastered programming.

Another interesting advanced option is Parallax's Propeller. This platform uses Parallax's own MCU chip called the P8X32A. It features eight (yes, eight) MPU cores along with memory and I/O. You program it in C, assembler, or a companion language called Spin. Once you have mastered BASIC on your Stamp platform and need more compute power, this may be a good next step for you.

LEARNING PROGRAMMING

There is not enough space in this book to teach a programming language. But you can learn it on your own. If you are not sure about self-learning, you could find a local college or university that offers continuing education courses for adults. Otherwise, give the self-learning idea a try. Here are some suggestions.

Books

There are literally dozens of books covering C programming. Not so many on BASIC. Some books are more specific to teaching the Arduino version of C. I reviewed several, but I can recommend two that may be helpful to you.

- *C Programming With Arduino*, by Warwick A. Smith, Elektor, 2016.
- *Arduino for Dummies,* by John Nussey, John Wiley & Sons, 2013.

I suggest that you go to your bookstore and see what is on hand. Your local library may have some. Alternately, do an Internet search on the Amazon and Barnes & Noble websites to see what is available. I recommend that you invest in at least two or three books as when one is not clear and helpful the others may be.

Learn by Example

Most of the books provide example programs. Kits such as the SparkFun (Arduino) Inventor's Kit and the Parallax BASIC Stamp Activity Kit provide hands-on experiments and projects with typical programs. Initially no programming is necessary. You simply enter the programs and run them. By implementing enough of the programs and analyzing them in detail, you can learn how to program on your own. In fact, with enough examples, you can put together several programs and modify them to fit your needs. If you are just learning programming, I highly recommend one of the experimenter's kits as a great starting point.

Online Tutorials

If you do an Internet search for programming tutorials, you will be surprised at how many there are for both BASIC and C. Some are free and others have a fee associated with them. Some are video based (such as YouTube), others are PDF documents, and some have presentations. Take your pick. Take one to start and then do another just to firm up your knowledge.

Online College Courses

Many colleges and universities now offer online college-level courses in a wide range of subjects. Such programs are referred to a Massive Open Online Courses or MOOCs. Programming is one popular topic. Many are free and others have small fees. If you are interested in a more formal course at the college level, from major institutions, take a look. Go to the websites of Coursera, edX, and Udemy, and search for a C or BASIC course. I selected a course called *Microcontrollers and the C Programming Language* from Udemy. It is taught online by Professor Mark Budnik of Valparaiso University in Indiana. It includes lab programming and interfacing projects using the Texas Instruments' LaunchPad described earlier. The course was offered at very low cost, but the LanuchPad was a minor added expense. A very challenging and time-consuming course but it does the job. There are others you can identify online.

Apply What You Learn

Reading relevant books and taking courses are good ways to get started, but they will not make you a programmer. To learn programming, you must do programming. You must write code, screw up, and retry. You must test the process yourself, failing along the way, and adjusting. This is still the best way to learn programming. Start with small programs and then progress to more complex applications as you improve your knowledge and skill. The projects outlined next will get you started.

Project 13.1

Select a Learning Platform and Language

If you have not already done so, choose your language and acquire one of the platforms or its equivalent described earlier. The complete kits cost a bit more but will get you started faster and easier than just investing in a development board. See Appendix B for sources. You should be able to get started for about $100. It is a good investment.

An interesting exercise is to investigate the programming languages to find out which ones are the most widely used. As a challenge in this project, perform an Internet search to discover the top three most popular programming languages.

A Word About the Following Projects

The projects that follow are intended as demonstrations that show what is possible and how to do it. If you invest in a microcontroller, you can try out these projects. You can just copy the programs into the IDE editor and test them out. The programs are instructive in showing you the programming process. However, you will need further learning before you become a programmer. Your goal should be further study as outlined earlier. Practice makes you more knowledgeable as you try, fail, and then try again and succeed.

Project 13.2

Blink an LED

One of the simplest applications is to flash an LED off and on. You first connect a resistor and LED to one of the output pins on your development board as shown in Fig. 13.5. Some boards such as the Arduino Uno R3 and the TI LaunchPad have this LED already connected on the PCB. As for the code, it is simple. Earlier in this chapter you saw the BASIC code to do this. Here is the code for the BASIC Stamp 2. Wire the LED as shown in Fig. 13.5 using a 470-Ω resistor on pin 7.

```
' {$STAMP BS2}
' {$PBASIC 2.5}
DO
     HIGH 7
     PAUSE 500
     LOW 7
     PAUSE 500
LOOP
```

The first two lines of code that begin with an apostrophe set up the programming software in the Stamp IDE. The remaining code flashes the LED as explained earlier in the chapter.

Project 13.2—cont'd

Below is a listing for the C code to flash the LED. Wire the LED as in Fig. 13.5 to pin 9. Use a 470-Ω resistor. If you are using the Arduino Uno, you can use the onboard LED already connected to output pin 13. Just substitute 13 for 9 in the code below.

```
int ledPin = 9;
void setup()   {
pinMode(ledPin, OUTPUT); }
void loop()   {
        digitalwrite(ledPin, HIGH);
        delay(500);
        digitalwrite(ledPin, LOW);
        delay(500);
}
```

The first line tells the Arduino to assign the variable ledPin to the integer value of 9. The LED and resistor would be connected to pin 9. Next the void setup() command tells the processor to execute the code between the two following curly brackets. The pinMode command tells the processor that pin 9 is an output port. Now the void loop() tells the processor to set up a loop to run the code between the following curly brackets. The digitalwrite command tells the processor to set pin 9 to the high or binary 1 state turning on the LED. The delay command then causes a pause of 500 ms or one half second. Next the digitalwrite command tells the processor to set pin 9 to low or binary 0 turning off the LED. Another delay or pause of 500 ms occurs. Note all command lines have to end in a semicolon. The program then loops back to repeat.

Project 13.3

Use a Pushbutton Switch for Control

This project will illustrate how to apply an input to the microcontroller. The pushbutton is a common input device. The basic interface is like that shown in Fig. 13.8. The micro looks at the input pin and senses the state. If it reads the right condition, it will cause something to happen. In the example to follow, the goal is to flash the LED five times when the pushbutton in depressed. The code in both BASIC and C are outlined below.

Here is the code for BASIC Stamp 2. It uses the same flash code in the previous BASIC example but is modified by to flash five times and then stop. The pushbutton should be wired like that in Fig. 13.8B on pin 5.

```
‘ {$STAMP BS2}
‘ {$PBASIC 2.5}
counter VAR Byte
FOR counter = 1 TO 5
DO
        IF (IN5 = 1) THEN
        HIGH 7
```

Continued

Project 13.3—cont’d

```
        PAUSE 500
        LOW 7
        PAUSE 500
NEXT
LOOP
```

The first two lines beginning with the apostrophe sets up the programming software and must be included. The next line sets up a variable called counter that can be up to a byte long. The DO instruction sets up loop to repeat the program. The IF…THEN line sets up another inside loop that reads input pin 5 to determine its state. It will look at the pin ten times per second. If input pin 5 goes high or 1, it signals the program to execute the code flashing the LED. The program then flashes the LED five times as the loop NEXT command counts the number of flashes and then stops. The LOOP instruction sets up the loop again as the program waits until the button is pressed again. The program is a loop within a loop where the inner loop is conditioned by the number of repeated flashes. The outer loop is conditioned by the state of the pushbutton.

The C code for the Arduino to initiate the five LED flashes is shown below. The pushbutton should be wired as shown in Fig. 13.8A using a 10k-Ω resistor. The code is set up to use pin 2 as the input. Pin 9 is used as an output for the LED as in the previous code.

```
int ledPin = 9;
int inputP = 2;
int val = 0;
void setup()   {
pinMode(ledPin, OUTPUT);
pinMode(inputPin, INPUT);
}
void loop()   {
        val= digitalRead(inputPin);
        if (val == HIGH) {
        digitalwrite(ledPin, LOW);
        } else {for (int i = 0; i < 5; 1++) }
        digitalwrite(ledPin, HIGH);
        delay(500);
        digitalwrite(ledPin, LOW);
        delay(500);
}
```

The program sets up pin 9 as the output for the LED and pin 2 as the input. The void loop() sets up a loop to repeat the flashing five times. The program reads the state of the input pin with the digitalRead statement. If it is HIGH, the button has not been pressed to close the circuit. The “if” statement evaluates the input, and if it is HIGH, it turns off the LED with a digitalwrite command. However, if the input goes LOW, the “else” statement initiates the flash sequence. The “for” statement counts the number or flashes and then terminates the program.

Chapter 14

Soldering

An Essential Skill for Electronics

In this Chapter:

- The importance of soldering.
- What is solder?
- Soldering tools.
- How to solder.
- How to desolder.
- Projects.

INTRODUCTION

Soldering is the process of melting a metal alloy called solder for the purpose of joining together two or more other metals. The solder, when melted, becomes a liquid and flows over the other metals to be joined. When the solder cools, it hardens and adheres to the other metals creating a strong mechanical bond as well as a highly conductive connection. Most interconnections of electronic parts and printed circuit boards (PCBs) are made with solder. For that reason, soldering is an important skill for any electronic engineer, technician, hobbyist, experimenter, maker, or other do-it-yourselfer. This chapter shows fundamentals of soldering and some projects to practice with and learn from.

What is Solder?

Solder is a soft metal alloy made up of tin and lead. The usual mix is 60% tin and 40% lead. Other mixes are available, but the 60/40 category is the most common in electronics work. Solder has a low melting point of about 360–370°F or in the 180–190°C range. It is usually melted with a handheld tool called a soldering iron. The solder is usually made in the form of a thin flexible wire. When melted, the solder becomes a liquid and easily flows over the other metals to be joined. The solder cools and hardens and adheres to the other metals making a solid joint.

Electronics Explained. http://dx.doi.org/10.1016/B978-0-12-811641-8.00014-X
Copyright © 2018 Elsevier Inc. All rights reserved.

Most wire solder contains a continuous core of a material called rosin flux. When heated, the flux melts with the solder and is used as a cleaning agent. The metals to be joined often oxidize and resist the solder. The flux clears away the oxidation and other surface obstacles to allow the solder to adhere to the metals. Some types of solder used in mechanical work use an acid core flux. This is deadly to electronic parts. Never use it. Use rosin core solder only.

Solder containing lead has been ruled undesirable in some applications. Yet it is still the best solder. It will not harm you. You can also get solder without lead. Such solder is mostly tin with a small amount of silver or copper or both. The melting point is typically 5%–10% higher so extreme care should be taken when soldering delicate temperature-sensitive electronic parts.

Soldering Tools

The main tool of soldering is the soldering iron. See Fig. 14.1. It has an insulated handle, a heating element with an AC power line cord and plug. The heating element heats a metal tip that actually does the soldering. The tips come in many sizes but a small tip is best for use on modern components.

Soldering irons are rated by the power they consume and heat they dispense. Ratings run from about 15W to over 100W. For electronics work an iron with a rating of 25–35W is best. It is hot enough to do the soldering but will not do heat damage to sensitive electronic components. Be sure that the soldering iron has a small narrow or pointed tip to accommodate most small electronic component leads.

A helpful accessory is a soldering iron stand or holder to keep the iron from burning the bench or table you are using. Also useful is a wet sponge. It is used

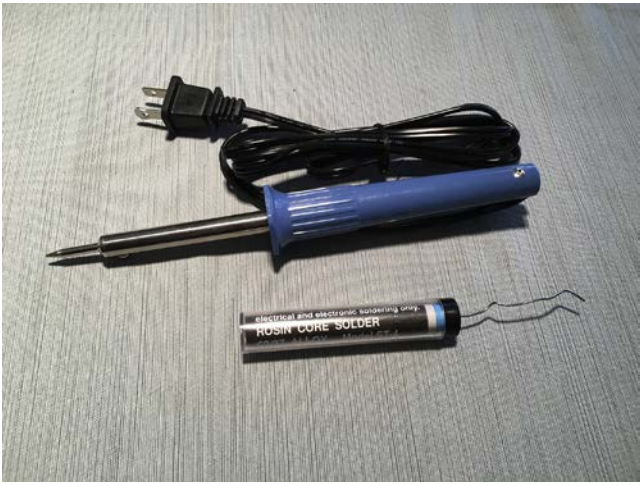

FIGURE 14.1 A common soldering iron used in electronic work.

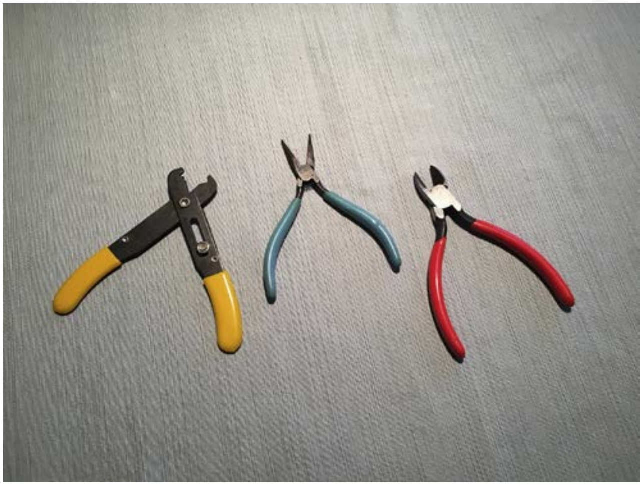

FIGURE 14.2 Tools used in soldering. On the left is a wire stripper. Next is a pair of needle-nose pliers. On the right is a side cutter for wire.

to keep the soldering tip clean. A sponge holder is often part of the soldering stand.

Other useful soldering tools are illustrated in Fig. 14.2. The wire stripper helps remove the insulating cover from wires to be soldered. It cleanly cuts off the insulation without damaging the copper wire. This is a must tool and a mandatory substitution for a knife that can nick the wire while stripping the insulation.

A pair of needle-nose pliers is sometimes helpful in bending or holding component leads or wires. The side cutters are used to cut wires or trim away component leads or wires after soldering. All of these tools speed up and simplify your soldering work and are a good investment.

If you are going to work with tiny surface mount parts, you should have a pair of tweezers to handle them or to hold them in place as you solder. A magnifying glass is also a useful accessory when using tiny parts.

Soldering Procedures

The most common soldering scenarios are soldering wires together or soldering electronic parts to a PCB. Fig. 14.3 shows the first case. When connecting two wires, you first remove any insulation first. You will need an inch or more of bare wire to make the connection. Then make a good physical connection by wrapping one wire around the other as shown. Then touch the clean tip of the iron to the connection. Let the connection heat for a few seconds, then apply the solder. The secret to a good connection is to apply the solder to the heated wires and not to the iron tip. Let the solder flow over the wires until coated. Do not use

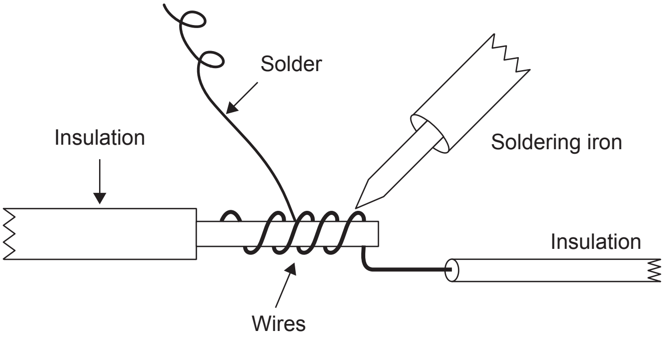
Solder
Insulation
Soldering iron
Insulation
Wires

FIGURE 14.3 Soldering two wires together. Apply solder to the heated wires not to the iron tip.

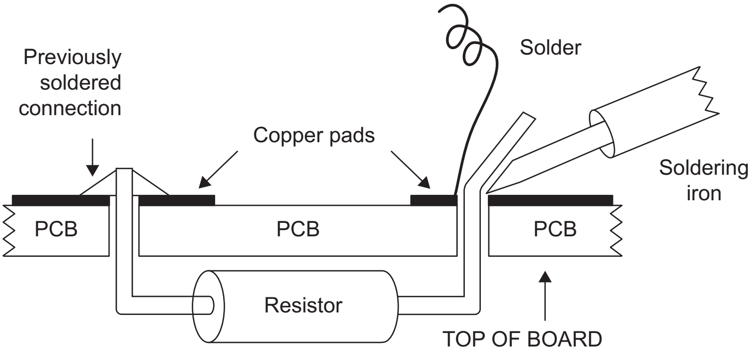
Previously soldered connection
Copper pads
Solder
Soldering iron
PCB
PCB
PCB
Resistor
TOP OF BOARD

FIGURE 14.4 Soldering components to a printed circuit board.

too much solder. Once the solder flows over the wires, remove the iron and let the connection cool. It should look shiny and bright, not dull gray.

Another common soldering situation is connecting components to a PCB. See Fig. 14.4. This shows a resistor lead being soldered to a copper pad on the PCB. The resistor lead is usually inserted into a hole on the board and is then soldered to a copper pad or area on the bottom of the board.

The procedure is to bend the lead and insert it into the hole. The resistor will lay flat on the board as shown. Bend the lead over to hold it in place, then solder. As with soldering wires, or anything for that matter, apply the heat to the copper pad and resistor lead for a few seconds. Then flow the solder over the connection. Do not apply the solder to the iron tip. And do not let the lead or the board move as you solder. Use minimum amount of solder to cover the connection and not more.

Remove the iron and let the connection cool. No movement should occur during the cooling process, or it will produce what is called a "cold" solder joint. Instead of looking bright and shiny the connection will look crumbly and dull gray. Movement during cooling is the most common cause of cold

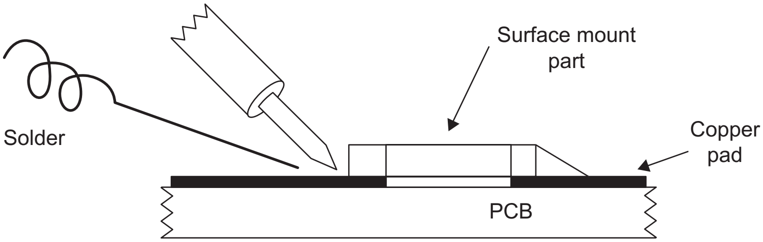

FIGURE 14.5 Soldering a surface mount part to a PCB.

joints. A cold joint does not make a good electrical connection. If this occurs, simply reheat the joint and then let it cool without movement. Do not add more solder.

Fig. 14.5 shows another soldering situation. Here a surface mount part is placed on the top of the PCB and is soldered to a copper pad. No hole is used or needed. Surface-mount parts are tiny and hard to work with. Hold the part carefully in place, then heat the junction of the part and the copper pad. Use a small iron tip to avoid accidentally touching adjacent pads or parts. Then flow the solder over the joint. Use as little solder as possible. Be very careful not to move anything as the solder cools.

There are a couple of precautions to observe. First, use as little heat as possible. Most PCBs are sensitive to heat and with too much heat you can cause the copper pad on the PCB to peel free and possibly break off. This is a serious problem that is often catastrophic for an expensive electronic PCB. Too much heat can sometimes damage a component such as an integrated circuit or resistor. The correct balance is to use just enough heat to melt the solder but not enough to damage anything. A few seconds is all you usually need.

Second, be careful of cold solder joints. Do not move anything as the solder cools. A cold joint will also result with not enough heat. The giveaway is a dull or rough-looking joint rather one that is smooth and shiny. Cold solder joints usually do not create a conductive path between the copper pad and the component. An open circuit is the result.

Third, most connections on a PCB are very close together. Spacings between components and IC pins of 0.1 in. (2.54 mm) or less are common. While you are soldering one connection, be careful not to disturb an adjacent connection. A common problem is using too much solder on a connection that slops over to a nearby connection causing a short or solder "bridge." To fix this, remove any excess solder from the connections to remove the bridge.

Desoldering

A related skill is desoldering. Desoldering is the process of disconnecting things that were previously soldered. It involves removing the solder first, then physically breaking the parts from one another. Desoldering is usually common when

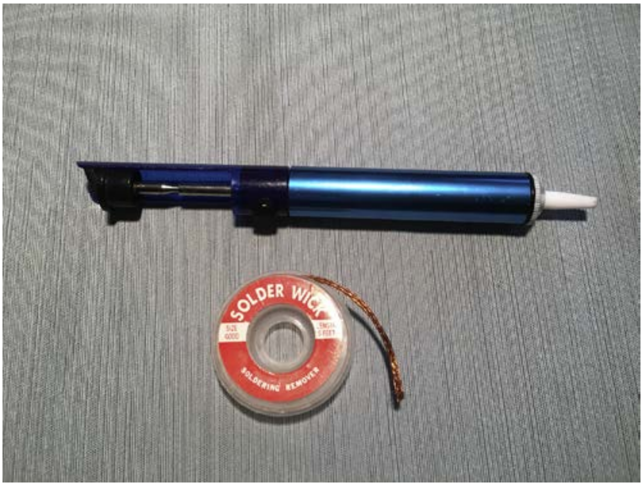

FIGURE 14.6 Tools for desoldering, a solder sucker and solder wick.

repairing electronic equipment by replacing a defective part or when making and correcting a mistake.

The tools of desoldering are shown in Fig. 14.6. The first is a solder sucker. It is a simple vacuum pump that literally sucks the molten solder from a joint. You heat the joint to be disconnected, then simultaneously apply the solder sucker to remove the solder. You usually have to repeat the procedure several times to get the solder off. Once you remove as much solder as possible, you physically pry the parts apart with the tip of a knife or a pair of needle-nose pliers. Be careful not to pull the copper foil pad off the PCB.

An alternative or a companion to the solder sucker is solder braid or wick. This is a narrow flat braid of fine copper wires coated with flux or other chemical that attracts solder. You place the braid on the connection to be desoldered, then heat the braid with the iron. Once the solder melts, it wicks up into the braid. The braid absorbs the solder. As with the solder sucker, you usually have to apply it several times. Be sure to cut off the braid that has absorbed the solder. Always use fresh piece of braid. Again, do not use too much heat and damage the PCB or part.

Projects

The best way to learn soldering is to solder. It is a manual skill that takes a bit of practice. It is not difficult but you do need to try it out first, then practice it several times. The following projects will help.

Project 14.1

Acquire the Tools

If you are planning to do serious electronic work, you need the right tools. The first step is to go get the tools described above. You may be able to get them individually at a DIY store such as Lowe's or Home Depot. RadioShack or Fry's should have them. Online catalog sources, such as All Electronics, Jameco, or others, have them. Some vendors offer complete packages of all the tools in one kit. This complete kit is the best approach if you have no tools now.

Project 14.2

Solder Two Wires

A common need is to connect two wires. An example you may have experienced is the problem of not being able to plug stranded wire into a breadboarding socket. The 9-V battery connector you may have used has stranded wires that are not stiff enough to plug into the breadboard. One way to fix this problem is to attach some short solid hookup wire leads to the battery clip wires. The hookup wires will plug right into the socket.

Here is the procedure:

- Cut two lengths of #22 gauge hookup wire several inches long.
- Use the wire strippers to remove about one half inch (1.25 cm) of insulation from each end of the wires.
- Also strip about one half inch of insulation from each battery clip wire.
- Then wrap the stranded battery clip wire around the solid hookup wire as previously shown in Fig. 14.3.
- Solder the connection by heating the wires first, then flowing solder over them. Do not apply the solder to the soldering iron tip. Also use a minimal amount of solder. Let the connection cool.
- Repeat the procedure for the other battery clip wire.

Project 14.3

Build a Kit

The best way to practice soldering is to build a kit. This was recommended earlier in Project 4.1. If you have not built a kit yet, now is a good time to do so. I recommend the solder practice kit (SP-1A) where you build a siren with flashing LEDs. The kit also provides an area on the PCB to practice soldering skills. You can get this kit from Elenco (www.elenco.com) or Jameco (www.jameco.com).

Any other small low-cost kit such as a power supply will also provide plenty of soldering experience. Refer back to Project 4.1.

This page intentionally left blank

Appendix A

Book References

Here is a list of books for those who wish to dig deeper. You will find lots of basic electronics books out there, but these are my favorites as well as some of my own.

American Radio Relay League. *The ARRL Handbook for Radio Communications*; 2017. The focus is on wireless but the book also has excellent coverage of electronic fundamentals.

Ashby D. *Electrical engineering 101*. Newnes/Elsevier; 2008. A good basic book that delves more into the circuits than this book. A good next step from here.

Frenzel LE. *Crash course in electronics technology*. 2nd ed. Newnes/Elsevier; 1996. Out of print now but still available in the used market. A programmed introduction to electronics.

Frenzel LE. *Crash course in digital technology*. 2nd ed. Newnes/Elsevier; 1998. Out of print now but still available in the used market. A programmed introduction to digital circuits. Getting a bit dated now but still valid.

Frenzel LE. *Crash course in PC and microcontroller technology*. Newnes/Elsevier; 1999. Out of print now but still available in the used market. A programmed introduction to micros. Getting dated but the fundamentals are still good.

Frenzel LE. *Principles of electronic communications systems*. 4th ed. McGraw-Hill; 2016. Recently updated and it takes the systems approach to the subject. A college text but very readable.

Frenzel LE. *Contemporary Electronics: Fundamentals, Devices and Circuits, and Systems*, McGraw Hill, 2014. A college textbook covering DC, AC, semiconductors, and linear circuits with a systems emphasis.

Horowitz P, Hill W. *The art of electronics*. 3rd ed. Cambridge University Press; 2015. Recently updated, a good example of how to present electronics. A great reference. A lab manual is also available.

Kuphaldt, TR. *Lessons in electric circuits*. Self-published; 2000–2009. A complete electronics book online and free for download. It covers DC and AC circuits, semiconductors and basic circuits, and digital. A lab manual is available. Go to www.openbookproject.new/electricCircuits.

Maxfield C. *Bebop to the Boolean Boogie*. 3rd ed. Newnes/Elsevier; 2009. First published in 1995, it was a hit. A great introduction to digital practice and circuits.

This page intentionally left blank

Appendix B

Projects Parts List

PARTS SOURCES

These are recommended but there are others.

Adafruit: www.adafruit.com
All Electronics: www.allelectronics.com
Digi-Key: www.digikey.com
Jameco: www.jameco.com
Marlin P. Jones & Associates: www.mpja.com
Mouser: www.mouser.com
SparkFun: www.sparkfun.com

CHAPTER 2

Digital multimeter (DMM). Must measure DC and AC voltages, DC current, and resistance. More expensive models may measure capacitance and test transistors, but these are not necessary for the projects in this book.

CHAPTER 3

- Batteries, four AA cells or a 9V battery
- Battery holder for AA cells (Get one with preconnected cells and an off/on switch. Jameco 216187 is recommended. In fact, buy two as you will need another supply for later projects.)
- Battery clip for 9-V battery (Buy two as you will need another supply later.)
- Breadboarding socket (Buy the larger one for more complex circuits. Two is recommended.)
- #22 solid copper hook up wire, insulated (At least several feet for future uses.)
- LED, any generic type or color
- 220-Ω ¼-W resistor
- 330-Ω ¼-W resistor
- 100k-Ω ¼-W resistor
- 100-µF electrolytic capacitor, voltage rating >10V

- MOSFET transistor IRF510, 2N7000, or equivalent
- Tools: soldering iron (25–35 W), rosin core solder, side cutters, wire strippers

Note: If you plan to do the projects in this book, you will need multiple values of resistors. Most parts vendors offer assortments of the most common value resistors with 5 or 10 of each value. This is a highly recommended purchase as it will save time and be more convenient.

CHAPTER 4

Project 4.1

Recommended kit is the dual adjustable regulated power supply from Jameco: Model JE215 (20626).

Project 4.2

Recommended VI packages:

- Digilent Analog Discovery or Analog Discovery 2 with Waveforms software
- National Instruments' my DAQ with LabVIEW software
- Velleman PCSU1000 oscilloscope or PCSGU250 oscilloscope with signal generator

Project 4.3

Possible power supply choices:

- Dual adjustable regulated power supply kit from Jameco: Model JE215 (20626)
- Batteries: Two each 4-AA cells in a holder (6V) or two 9-V batteries with clips
- Power supplies in VI units
- Wall adapter AC to DC power supplies, regulated preferred, 5V

Project 4.4

Parts for Fig. 4.26:

- AC to AC wall adapter transformer: 5V (Cut the connector off the cable and tin the wires with solder to plug into the breadboarding socket.)
- 1k-Ω potentiometer (Get the kind that plugs into the breadboarding socket.)

Parts for Fig. 4.27:

- 555 timer IC 8 pin dual inline package (DIP)
- 10k-Ω ¼-W resistors (3)
- 0.05-μF ceramic capacitor
- 0.1-μF ceramic capacitor
- 47-μF electrolytic or tantalum capacitor, voltage > 10V

Project 4.5

Parts for Fig. 4.31:

Same as for Fig. 4.27 but add:

- 100-Ω ¼-W resistor
- 2-in., 8-Ω speaker

Parts for Fig. 4.32:

Same as for Fig. 4.27 but add:

- 220-Ω ¼-W resistor
- LED

Project 4.7

Parts for Fig. 4.33:

Same as for Fig. 4.26 plus:

- 1N4001 silicon diode
- 1000-μF electrolytic capacitor, 16V or higher
- 0.33-μF capacitor
- 0.1-μF capacitor
- 1k-Ω, ¼-W resistor
- 7805 linear regulator, TO-220 package

Project 4.8

Parts for Fig. 4.35:

- TL081 op amp or equivalent (741, etc.)
- 15k-Ω, ¼-W resistor
- 4.7k-Ω, ¼-W resistor
- 1k-Ω pot

Parts for Fig. 4.36:

- TL081 op amp or equivalent (741, etc.)
- 6.8k-Ω, ¼-W resistor
- 1k-Ω, ¼-W resistor
- 1k-Ω pot
- Wall transformer

CHAPTER 5

Project 5.2

Parts for Fig. 5.31:

- 5V power supply as suggested in Project 5.11
- 74HC00 quad NAND gate IC
- 220-Ω, ¼-W resistor

Continued

Project 5.2—cont’d

- LED

Parts for Fig. 5.32:

- Same as parts for Fig. 4.32.
- Same as parts for Fig. 5.31.

Project 5.3

Parts for Fig. 5.33:

Same as parts for Fig. 4.32, plus:

- 74HC161 binary counter IC
- LEDs (4)
- 330-Ω, ¼-W resistors (4)
- 1k-Ω ¼-W resistors (4)
- 0.01-μF capacitor

CHAPTER 7

Project 7.6

Parts for Fig. 7.23:

- Two 6 or 9V supplies
- Part for 555 timer circuit same as for Fig. 4.32
- Radiometrix TX3A transmitter
- Radiometrix RX3A receiver
- 2N3904 NPN transistor
- 470-Ω, ¼-W resistor
- 10k-Ω, ¼-W resistor (2)
- 27k-Ω, ¼-W resistor
- LED

CHAPTER 12

Project 12.2

Parts for Fig. 12.31:

- LM393 quad comparator IC
- IRF510 MOSFET or equivalent
- 10k-Ω thermistor
- 12-V DC fan, 2 or 3 in. in diameter
- 10k-Ω, ¼-W resistors (2)
- 22k-Ω, ¼-W resistors (2)
- 50-Ω, 5-W resistor

CHAPTER 13

Project 13.1

Sources of microcontrollers:

Adafruit: www.adafruit.com
Arduino: www.arduino.c
Jameco: www.jameco.com
Marlin P. Jones & Associates, www.mpja.com
Mouser: www.mouser.com
Parallax: www.parallax.com
PICAXE: www.pickaxe.com
SparkFun: www.sparkfun.com

Project 13.2

Parts to flash an LED:

- Standard LED
- 470-Ω, ¼-W resistor

Project 13.3

Parts for pushbutton control of an LED:

- Standard LED
- Pushbutton switch with leads to plug into a breadboarding socket
- 470-Ω, ¼-W resistor
- 10k-Ω, ¼-W resistor

CHAPTER 14

Project 14.1

Acquire a soldering iron and some solder. A model with a 20–30W rating with a narrow tip is best. The solder should be 60/40 rosin core. A desoldering tool or braid is an option. Also get basic tools such as wire strippers, needle nose pliers, and side cutters. A knife is also handy.

Project 14.2

- 9-V battery clip and/or four AA cell holder
- #22 solid insulated hook up wire

Project 14.3

Acquire the solder practice kit (SP-1A) or some other kit to practice with.

This page intentionally left blank

Appendix C

Learning Resources

This book provides a good introduction to electronics. But as you will find out, the more you learn the more you realize what you do not know. Most people want to learn more.

Today, most people learn from the Internet. With searches you can find information on almost anything you want to learn. Wikipedia is an example. Listed here are some additional sources that I recommend for your continued learning.

- *All About Circuits* is an online resource for electronic engineers and enthusiasts that offers a wide range of materials including design calculators, articles, textbooks, and projects (http://www.allaboutcircuits.com/tools/).
- *MOOCs* or *Massive Open Online Courses* are online college-level courses covering an extensive array of topics. Many are free; others are low cost. The offerings are extensive so you need to do a search on each website to find what you want. The following are leading providers of MOOCs:
    - Coursera: www.coursera.com
    - EdX: www.edx.com
    - Udemy: www.udemy.com
    - MOOCs listing: www.mooc-list.com
    - MOOCs directory: www.moocs.co
- The *Great Courses* is a source of excellent video courses on DVD, CD, or online. They offer a general electronics course called Understanding Modern Electronics by Professor Richard Wolfson of Middlebury College. Its price varies with discounts offered. Other courses of potential interest are those on science and math (www.thegreatcourses.com).
- EEWeb is a website offering a variety of free electronic tools and calculators. The tools include schematic capture, PCB design and calculations, RF design tools, and passive components calculators. Other features are an online scientific calculator, graph paper and math tools for algebra, trig and calculus (https://www.eeweb.com/toolbox).
- Radio Electronics is a website with an extensive collection of articles and tutorials about all aspects of electronics (www.radio-electronics.com).
- Kahn Academy is a free online learning site offering courses in math, electrical engineering, and programming (www.kahnacademy.com).

This page intentionally left blank

Index

'*Note*: Page numbers followed by "f" indicate figures, "t" indicate tables, and "b" indicate boxes.'

#

3D TV, 278
4G, 3, 197, 212–213
4K, 268–269
5G, 3, 166, 201, 214, 238
555 Timer, 91b–92b, 92f, 96f, 97b, 137f–138f
8K, 269

A

AC. *See* Alternating current
Accumulator
  in ALU, 146
  register, 145–147
Actuator, 303–312. *See also* Motor; Relay; Solenoid
ADC. *See* Analog-to-digital converter
Address buffer, 147–148, 147f
Address bus, 151
Address register, memory, 147–148, 147f
Addressing, in RAM, 126f
Aircraft, radio in, 187
Alkaline battery, 30t
Alternating current (AC), 26–29, 27f
  capacitors' passing of, 51f
  frequency of, 28
  generators for, 33
  inverters for, 33–34
  motor, 311–312
  oscillator in, 26, 34
  polarity of, 26
  signal shapes of, 26–28, 27f
  in solar power system, 88f
  sources for, 33–34
  voltage measurement for, 28–29
Alternator, 33
ALU. *See* Arithmetic logic unit
AM. *See* Amplitude modulation
American Standard Code for Information Interchange (ASCII), 107–112
  character code of, 109t–110t
  control codes of, 108–111
Ampere, 18–19
Amplifier, 7–8, 54–55, 64–65, 71–75
  audio frequency, 65
  BJT, 54f
  cascading, 73–74, 73f
  class A, 66, 67f
  class AB, 66–67
  class B, 66, 67f
  class C, 67–68, 68f
  class D, 68–69, 69f
  class E, 69
  class F, 69
  classification of, 65–71
  configuration of, 65–66
  DC, 64
  differential, 66, 66f, 70f
  efficiency of, 74
  follower, 70f
  frequency response of, 74, 75f
  gain of, 64, 73f
  impedance in, 72–73, 72f
  instrumentation, 70–71, 71f, 297
  intermediate frequency, 65
  inverting and noninverting, 70f
  as linear circuit, 65
  microwave, 65
  operational, 70–71, 70f
  output capability of, 74–75
  power, 65
  pre-, 244
  radio frequency, 65
  in radio receiver, 177
  signal size of, 65
  single-ended, 65, 66f
  switching, 68–69, 69f
  symbol for, 65f
  vacuum-tube, 258b
Amplitude modulation (AM), 170, 171f. *See also* Quadrature amplitude modulation
Amplitude shift key (ASK), 174, 175f
Analog-to-digital converter (ADC)
  for data conversion, 132
  for digital TV, 267f

Analog-to-digital converter (ADC) (*Continued*)
  parallel output of, 133f
  serial output of, 133f
Antenna, 162, 181–186
  dipole, 182, 183f
  electromagnetic fields from, 163f
  ground-plane, 181, 182f
  horn, 185
  loop, 183
  microwave, 185–186, 185f
  patch, 183, 184f
  transmission lines to, 183–185, 185f
  Yagi, 182–183, 184f
Arduino, 340–341, 341f, 343
Arithmetic circuit, 125
Arithmetic logic unit (ALU), 10, 10f
  accumulator in, 146
  in computer, 10, 10f, 142
  data processing by, 146f
  as logic circuit, 125
  operation of, 145–147
ASCII. *See* American Standard Code for Information Interchange
ASK. *See* Amplitude shift key
Aspect ratio, 262–263, 264f
Assembly (programming language), 327–328
Astronomy, radio, 191
Asynchronous transfer mode (ATM), 227, 227b
ATM. *See* Asynchronous transfer mode
Atom, 16, 17f
Audio
  amplifier, 65
  digital, 245–254
  dynamic range of, 247b
  past recording media, 246–247
  electronics, 243
Audio-video receiver (AV receiver), 254–257
  diagram of, 255f
  radio in, 254–255
Augmented reality (AR), 288
Automatic gain control, 178
AV receiver. *See* Audio-video receiver

B

Band-pass filter (BPF)
  bandwidth of, 75–77
  response curve of, 76f
  symbol for, 76f
Band-reject filter (BRF), 76f
Bandwidth
  of BPF, 75–77
  modulation and, 172f
  of radio, 172, 172f
BASIC (programming language), 328
BASIC Stamp 2, 339
Battery, 29f
  cell voltages in, 30t
  charging of, 31
  as DC source, 29–31
  diagram of, 31f
  multimeter measurement of, 58f
  as power supply, 84
  primary cells in, 30–31
  secondary cells in, 31
  types of, 30t
BCD. *See* Binary-coded decimal
Binary data word, 106
Binary digits, 103–104
Binary numbers, 103–114
  decimal conversion from and to, 104–105
  hardware representation of, 105
  maximum number of states in, 107t
Binary phase shift keying (BPSK), 174, 175f
  spectral efficiency of, 209t
Binary-coded decimal (BCD), 107–112
  code of, 108t
  hexadecimal notation in, 111b
Bipolar junction transistor (BJT), 52–53, 53f–54f
Bits. *See also* Microcontroller; Microprocessor
  binary numbers as, 103–114
  maximum decimal value calculation for, 106–107
  significant, most and least, 104
BJT. *See* Bipolar junction transistor
Bluetooth, 233
Blu-ray DVD, 288
Boat, radio on, 188
Boolean algebraic expressions, 114–115
BPF. *See* Band-pass filter
BPSK. *See* Binary phase shift keying
Breadboard
  in home lab, 56b–59b
  LED wired on, 59b
  socket, 58f
BRF. *See* Band-reject filter
Broadband technology
  network with, 237–238
  wireless, 238
Bus, 112. *See also* Universal serial bus
  address, 151
  bidirectional, 151
  computer, 151
  control, 151
  data, 151, 152f
  multiplexer, 151
  network topology, 220, 221f
  parallel data transfer over, 112f

power supply and, 87
  sources connected to, 152f
Bus-oriented architecture, 87
Byte, 106

C

C (programming language), 328–329
Cable, 221–222
  coax, 221, 222f, 234f
  fiber optic, 222, 223f–224f
  twisted pair, 221–222, 222f
Cable TV, 278–283
  diagram of, 280f, 282f
  Internet connectivity with, 237
CAN. *See* Controller area network
Capacitive resistance, 49
Capacitors, 47–50
  AC vs. DC with, 51f
  charging of, 49f, 60b–61b
  construction of, 49f
  discharging from, 49f
  in LPF, 50f
  symbol for, 49f
Carbon, as semiconductor, 20
Cascading amplifier, 73–74, 73f
Cathode ray tube (CRT), 258b
CB. *See* Citizens band
CCD. *See* Charge-coupled device
CD. *See* Compact disc
CDMA. *See* Code division multiple access
Cell. *See also* Battery; Solar cells
  fuel, 32–33
  load of, 298
  photo, 299–302, 299f
  symbols for, 299f
Cell phone, 195
  4G, 212–213, 212f
  CDMA, 202, 211
  cellular concept of, 197–199, 198f
  communication model of, 8–9, 8f
  components of, 199–202
  diagram of, 201f
  digital modulation in, 203–208
  duplex capability of, 199–200
  EDGE, 201–202
  GPRS, 201–202
  GPS in, 216
Cell phone (*Continued*)
  GSM, 201–202
  hand-off of, 199
  HSPA, 202
  location technology in, 216
  LTE, 202
  operation of, 200f
  as radio, 195–196
  range and reliability of, 196b–197b
  smartphone as, 213
  spread spectrum in, 175–176
  transmission frequency of, 28t
  TV on, 285–286
  types of, 200–202
Central processing unit (CPU)
  components of, 142
  control unit of, 142, 147–148
  instruction-word formats in, 148–149
  program execution in, 149–150
  registers in, 144–145
Charge, 16–18
Charge-coupled device (CCD), 270
Chip. *See* Integrated circuit
Circuit, 63. *See also* Digital circuit; Linear circuit; Logic circuit
  arithmetic, 125
  filter as, 75–77
  mixer as, 78–79
  model of, 7, 8f, 19f
  photo control, 300, 301f
  theory of, 41–45
  three-state line-driver, 151
Citizens band (CB), 187, 193b
Clock
  crystal, 77, 77f
  in microprocessor, 154
Coax cable, 221, 222f, 279f
Code. *See also* American Standard Code for Information Interchange; Binary-coded decimal; Decoder
  operation, 147–149
  for resistors, 48f
Code division multiple access (CDMA)
  cell phone use of, 202, 211
  generation of, 211f
Codec, 251
Communication. *See also* Duplex communication
  cell phone model of, 8–9, 8f
  as industry sector, 4, 4f
  LOS, 170, 170f
  radio system for, 159–160, 160f
  with satellite, 189–191, 189f
  simplex, 161, 161f
Compact disc (CD), 251–254
  dynamic range of, 247b
  encoding on, 252f
  past recording media vs., 246–247
  player, diagram of, 253f

Comparator, 125
Compiler, 326
Components
  capacitors as, 47–50
  of cell phone, 199–202
  of CPU, 142
  diodes as, 52
  in electronics hierarchy, 43f
  IC as, 55–56
  inductors as, 50–51
  as industry sector, 4f, 5
  of microprocessor, 142
  resistors as, 45–46
  switches as, 45, 55, 104
  transformers as, 51–52
  transistors as, 7, 52–55, 104
  types of, 45–56
Compounds, atomic elements in, 16
Compression
  of sound, 248–249
  of video, 271
Computer, 139–142, 143f. *See also* Central processing unit; Memory; Microprocessor
  ALU in, 10, 10f, 142
  applications of, 9
  bus, 151
  control with, 140
  diagram of, 44f
  embedded, 11
  external devices with, 141f
  FFs in, 140
  function of, 9–11
  gates in, 140
  as industry sector, 4, 4f
  input-output system of, 10–11, 10f, 140, 142
  memory of, 142
  as microcontroller, 140
  monitoring by, 141f
  personal, 140
  repairing, 42
  RISC, 155
  sections of, 9–10, 10f
  as system, 43
  types of, 139
Conductor, 20. *See also* Semiconductor
Control. *See also* Industrial control; Microcontroller; Programmable logic controller; Relay
  of actuator, 303–312
  bus, 151
  closed-loop, 291–295, 293f–294f
  computers for, 140
  gain, 178
  as industry sector, 4–5, 4f
  open-loop, 291–295, 292f
  of output device, 303–312
  photo, 300, 301f
  robots as device for, 11
  with thyristor, 312–317
Controller, 294–295. *See also* Microcontroller; Programmable logic controller
Controller area network (CAN), 155b–157b
Copper, atomic composition of, 17f
Core, 140
Coulomb, 18–19
Counter, 121–123
  data conversion with, 123f
  FFs in, 121–122
  program, 147, 147f
CPU. *See* Central processing unit
Crossover, 245, 245f
CRT. *See* Cathode ray tube
Current flow, 7, 16–20. *See also* Alternating current; Direct current
  conventional, 19–20
  direction of, 19–20
  electrons in, 15
  in magnetic fields, 21, 22f
  measurement of, 18–19
  voltage in, 15
Cut-off frequency, 49–50

D

DAC. *See* Digital-to-analog converter
Data
  ALU's processing of, 146f
  bus, 151, 152f
  on fiber optic cable, 224f
  Internet's protocols for, 230f
  LED representation of, 106f
  parallel transmission of, 112–114, 112f
  register's manipulation of, 145f
  retrieval of, 140
  serial transmission of, 112–114
  storage of, 140
  word, binary, 106
Data acquisition system, 321–323
Data conversion, 131–135
  ADC for, 34, 132
  with counter, 123f
  DAC for, 34, 132–133, 134f

  resolution in, 133–134
  sampling in, 132–134, 132f, 134f
Data processor. *See* Processor
dB. *See* Decibels
DC. *See* Direct current
Decibels (dB), 71
Decimal number system. *See also* Binary-coded decimal; Hexadecimal notation
  binary conversion from and to, 104–106
  bits' maximum value calculation in, 106–107
Decoder, 124–125, 125f
Demodulator, 179
Demultiplexer, 124
Demux. *See* Demultiplexer
Desoldering, 351–352
Detector. *See* Demodulator
Integrated Development environment (IDE), 327
Diac, 315
Differential amplifier, 66, 66f, 70f
Digital audio, 245–254
Digital circuit, 103–104. *See also* Binary numbers; Data conversion; Gate; Logic circuit; Memory; Programmable logic device
  logic elements in, 114–123
Digital compression, 248–250
Digital light processing (DLP), 261–262, 274, 277
Digital modulation
  advanced, 203
  basic, 174–175, 175f
  in cell phone, 203–208
  MIMO in, 206–208, 207f
  OFDM in, 204–206, 205f–206f, 211
  QAM, 203–204, 203f–205f
  spectral efficiency of, 208, 209t
Digital radio, 192b, 260b
Digital signal processing (DSP), 157–158, 157f
Digital subscriber line (DSL), 237–238
Digital video disc (DVD), 286–288
  Blu-ray, 288
  player, 287–288, 287f
Digital video recorder (DVR), 282f
Digital-to-analog converter (DAC), 34, 132–133, 134f
Diode. *See also* Light emitting diode
  bias of, 52f
  as component, 52
  photo, 300f
  rectification by, 53f
  silicon, 300
  symbol for, 52f
  temperature sensing with, 297
  zener, 52
Dipole antenna, 182, 183f
Direct current (DC), 24–26, 25f, 29–32
  amplifier, 65
  from batteries, 29–31
  capacitors' blocking of, 51f
  conversion of, 87, 88f
  ground in, 24
  motor, 310–311
  negative, 26
  positive, 24
  power supplies for, 33
  pulsating, 26
  pulse filtering to, 80b
  from solar cells, 31–32
  in solar power system, 88f
Direct digital synthesis (DDS), 134–135
DIY, do it yourselfer, 3, 89b–90b
Dish, parabolic, 186, 186f
DLP. *See* Digital light processing
Driver
  radio amplification with, 177
  three-state line, 151
DSL. *See* Digital subscriber line
DSP. *See* Digital signal processing
Duplex communication, 161, 161f
  access methods and, 209–211
  of cell phone, 199–200
  FDD, 209
  TDD, 210
Duplexing, 209–211
Duty cycle, 86f, 86b
DVD. *See* Digital video disc
DVR. *See* Digital video recorder
Dynamic range, audio's, 247b

E

Earphones. *See* Headphones, noise-cancelling
EDGE, 201–202
Editor, 327
Effector, end, 11
Efficiency, 74
  of amplifiers, 74
  spectral, 208, 209t
8051 microcontroller (Intel), 152–154, 153f
Electricity
  dynamic, 18–19
  electronics vs., 6–7
  electrons in, 7, 15–20

Electricity (*Continued*)
  as energy source, 15
  production of, 15
  static, 16–18
Electromagnetic induction, 23–24, 23f
Electromagnetism, 21–23. *See also* Magnetic field; Magnetism
  from antenna, 163f
  frequency spectrum of, 28t, 164–168, 165f
  in radio, 162–164
  spectrum space in, 166b–167b
  wavelength in, 167b–168b
Electronics
  as applied science, 16
  audio, 243
  concepts in, 15
  in daily life, 1
  electricity vs., 6–7
  hierarchy of, 43f
  input-process-output model of, 7–8, 8f
  learning about, 2–3
Electronics industry
  communications as sector of, 4, 4f
  components as sector of, 4f, 5
  computers as sector of, 4, 4f
  control as sector of, 4–5, 4f
  instrumentation as sector of, 4f, 5
  manufacturing in, 5–6, 6f
  raw materials in, 5, 6f
  workings of, 5–7, 6f
Electrons
  in atoms, 16
  in current flow, 15
  in electricity, 7, 15–20
Electrostatic discharge (ESD), 18
Elements, atomic, 16
End effector, 11
ESD. *See* Electrostatic discharge
Ethernet
  frame of, 225f
  hub, 225f
  networking with, 223–225
  operation of, 225–226

F

Fast Fourier transform (FFT), 158
FCC. *See* Federal Communications Commission
FDD. *See* Frequency-division duplexing
FDMA. *See* Frequency division multiple access
Federal Communications Commission (FCC), 164
Feedback, in closed-loop control, 293f
FF. *See* Flip-flop
FFT. *See* Fast Fourier transform
Fiber optic cable, 222, 223f
  data transmission over, 224f
  TV over, 281b
Filter. *See also* Band-pass filter; Band-reject filter; High-pass filter; Low-pass filter
  as circuit, 75–77
  pulse to DC, 80b
  surface acoustic wave, 77
Flag, 148
Flag register, 148
Flip-flop (FF), 119–120
  cascading, 120f
  in computer, 140
  in counters, 121–122
  in storage register, 121f
  symbols for, 119f
FM. *See* Frequency modulation
Fourier theory, 34–35. *See also* Fast Fourier transform
  harmonics and, 68
Frequency. *See also* Amplifier
  of AC, 28
  of cell phone transmission, 28t
  cut-off, 49–50
  domain, 34–35, 35f
  of electromagnetic spectrum, 28t, 164–168, 165f
  Fourier theory and, 34–35
  intermediate, 274
  of oscillators, 77–78
  of radar transmission, 28t
  of satellite transmission, 28t
  synthesizers, 82–83, 83f
  of TV transmission, 28t
Frequency division multiple access (FDMA), 209
Frequency modulation (FM), 173–174, 173f
  sidebands on, 173–174
Frequency shift keying (FSK), 174, 175f
  spectral efficiency of, 209t
Frequency-division duplexing (FDD), 209
FSK. *See* Frequency shift keying
Fuel cell, 32–33
Function generators, 34

G

Gain
  of amplifier, 65, 73f
  control, 178
  dB for measurement of, 71

  power, 71
  voltage, 71
GAL. *See* Generic array logic
Gate
  AND, 115–116, 115f–116f
  OR, 117, 117f
  in computer, 140
  NAND, 117–118, 118f
  NOR, 117–118, 118f
  XNOR, 119f
  XOR, 118, 119f
Gateway, 239–240
Gauge, strain, 297, 298f
Gaussian minimum shift keying, 201–202
General Packet Radio System (GPRS), 201–202
General-purpose register (GPR), 144
Generator
  AC, 33
  function, 34
Generic array logic (GAL), 129
Germanium, 20
Global positioning system (GPS), 190
  in cell phones, 216
Global System for Mobile Communications (GSM), 201–202
GPR. *See* General-purpose register
GPRS. *See* General Packet Radio System
GPS. *See* Global positioning system
Ground, 50b
  in DC, 24
  symbol for, 51f
Ground-plane antenna, 181, 182f
Groupe Spécial Mobile. *See* Global System for Mobile Communications
GSM. *See* Global System for Mobile Communications

H

Ham radio, 188
HAN. *See* Home area network
Hard copy, 144
Hardware
  binary numbers represented by, 105
  purchasing, 36b
  of radio, 176–181
Harmonics, 68
HDMI. *See* High-definition media interface
Headphones, noise-cancelling, 258–259
Hearing
  aids, 259–260
  testing of, 260b
Hertz (Hz), 26
Hexadecimal notation, 111b
High frequency radio, 169
High-definition media interface (HDMI), 274
High-definition TV (HDTV), 43, 254
  packet in, 272f
  resolution of, 268
High-pass filter (HPF), 76f
High-speed packet access (HSPA), 202
Home area network, 219
Home entertainment system
  satellite TV in, 284f
  as system, 43
Horn antenna, 185
HPF. *See* High-pass filter
HSPA. *See* High-speed packet access
Hydraulic device, 312
Hydrophone, 259
Hz. *See* Hertz

I

IC. *See* Integrated circuit
IDE. *See* Integrated Development Environment
Impedance, 72b
  in amplifier, 72f, 73–74
  input, 72–73
  output, 73
Induction
  electromagnetic, 23–24, 23f
  self-, 50–51
Inductive reactance, 50–51
Inductors, 50–51, 51f
Industrial, scientific, and medical band (ISM band)
  applications of, 235
  project on, 192b
  transceiver, 179–181
Industrial control, 291
  actuator in, 303–312
  closed-loop, 291–295, 293f–294f
  controllers in, 294–295
  diagram of, 292f
  motor in, 310
  open-loop, 291–295, 292f
  output device in, 303–312
  relay in, 305
  sensors in, 295–302
  solenoid in, 309, 309f
  with thyristor, 312–317
Instruction register (IR), 147
Instruction-word formats, 148–149, 149f
Instrumentation, as industry sector, 4f, 5

Instrumentation amplifier, 70–71, 71f, 297
Instruments, musical, 257–258
Insulators, 20
Integrated circuit (IC), 2, 55–56
  transistors in, 7
Integrated Development Environment
Interface, 144
  high-definition media, 274
  to microcontroller, 155b–157b
  serial peripheral, 155b–157b
International Telecommunications Union (ITU), 164
Internet, 228–229
  on cable TV, 237
  diagram of, 228f
  protocols used on, 230f
Internet of Things (IoT), 230–231, 238–242, 240f
Interrupt, 154
Inversion, 114
  in amplifiers, 70f
  phase, 54–55
Inverters
  in AC, 33–34
  logic of, 114–115
  as power supply, 87
  symbols for, 114f
IPTV. *See* Internet protocol TV
IR. *See* Instruction register
ISM band. *See* Industrial, scientific, and medical band
ITU. *See* International Telecommunications Union

K

Kit building, 88b–89b

L

Lab, set up of, 56b–59b
LAN. *See* Local area network
Language. *See* Programming language
LaunchPad, 341, 342f, 343
LCD. *See* Liquid crystal display
Lead acid battery, 30t
LED. *See* Light emitting diode
Light
  photo control circuit for, 300, 301f
  speed of, 164
Light emitting diode (LED), 52
  binary data represented by, 106f
  on breadboard, 59b
  MOSFET switching of, 61b
  OLED, 277, 278b
Lightning, 17–18
Linear circuit, 63–100
  amplifier as, 65
  filter as, 75–77
  mixer as, 78
  phase detector as, 79
  straight-line response of, 64f
Line-of-sight communication (LOS communication), 170, 170f
Liquid crystal display (LCD)
  operation of, 143
  TVs made with, 261–262
Lithium ion battery, 30t
Local area network (LAN), 218
Logic. *See also* Gate; Programmable logic controller
  Boolean, 114–115
  digital, 114–123
  generic array, 129
  of inverter, 114–115
  ladder, 320, 321f
  in PAL, 129
  in PLA, 129
  in PLD, 129–131, 130f
Logic circuit, 124–125
  ALU as, 125
  comparator as, 125
  decoder as, 124–125, 125f
  demultiplexer as, 124
  multiplexer as, 124, 124f
Logic gate. *See* Gate
Long-Term Evolution (LTE), 202
Loop antenna, 183
LOS communication. *See* Line-of-sight communication
Low frequency radio, 168
Low-pass filter (LPF)
  capacitors in, 50f
  function of, 75
  in PLL, 80
  response curve of, 76f
  symbol for, 76f
LPF. *See* Low-pass filter
LTE. *See* Long-Term Evolution

M

Machine-to-Machine (M2M), 214, 230–231, 239–240
Magazines, 37b
Magnetic field, 20–21, 21f
  current flow in, 21, 22f
  voltage induction with, 23f

Magnetic tape, 246
Magnetism, 20–24
  attraction and repulsion in, 21
  flux lines in, 21f
  magnetic fields in, 20–21, 21f
  materials exhibiting, 20
Maker, 3, 347
MAN. *See* Metropolitan area network
Manipulator arm, robot's, 11, 12f
Manufacturing, electronics industry, 5–7, 6f
Marine radio, 188–189
Memory. *See also* Random access memory; Read-only memory
  in computer, 142
Memory address register, 147–148, 147f
Mercuric oxide battery, 30t
Mesh network topology, 219–220, 234f
Metal oxide semiconductor-field-effect transistor (MOSFET), 53, 54f
  in class C amplifier, 67–68
  LED switching by, 61b
  switch, 55f
Metropolitan area network (MAN), 218
Microcomputer. *See* Computer
Microcontroller, 11
  8-bit, 152–154
  16-bit, 152–154
  32-bit, 155–157
  64-bit, 155–157
  computer as, 140
  8051 (Intel), 152–154, 153f
  interfaces to, 155b–157b
  MPS430 (Texas Instruments), 154
Microphone, 244
Microprocessing unit. *See* Central processing unit; Microprocessor
Microprocessor, 140. *See also* Central processing unit
  8-bit, 152–154
  16-bit, 152–154
  32-bit, 155–157
  64-bit, 155–157
  clock in, 154
  components of, 142
  instruction-word formats in, 148–149
  popular, 151–157
  RISC, 155
Microwave
  amplifier, 65
  antenna, 185–186, 185f
  propagation of, 169
  radio, 169–170
MIMO. *See* Multiple input multiple output
Mixer, 78–79
  as circuit, 78
  linear, 78, 79f
  nonlinear, 78–79, 79f
  outputs, 78–79
  in radio receiver, 178
Mobile telephone switching office, 199
Modulation, 170–176. *See also* Amplitude modulation; Digital modulation; Frequency modulation; Quadrature amplitude modulation
  angle, 174
  ASK in, 174
  bandwidth and, 172f
  demodulator and, 178
  OOK in, 174
  phase, 174
  pulse width, 68–69, 86b
  sidebands from, 171–172
MOSFET. *See* Metal oxide semiconductor-field-effect transistor
Motor
  AC, 311–312
  applications of, 310
  DC, 310–311
  industrial control with, 310b
  speed regulation of, 295f
  stepper, 311
MP3 player
  diagram of, 44f, 250f
  operation of, 250–251
  as system, 43
MPS430 microcontroller (Texas Instruments), 154
Multimeter, 35b–36b
  battery measurement with, 58f
Multiple input multiple output (MIMO)
  in digital modulation, 206–208, 207f
  spectral efficiency of, 209t
Multiplexer. *See also* Demultiplexer
  bus, 151
  as logic circuit, 124, 124f
Multiplexing, 209–211. *See also* Orthogonal frequency division multiplexing
Musical instruments, 257–258
Mux. *See* Multiplexer

N

Network, 217–218
  Bluetooth, 233
  with broadband technology, 236–238
  CAN as, 155b–157b

Network (*Continued*)
  DSL for, 237–238
  with Ethernet, 223–225
  HAN as, 219
  interconnection methods for, 219–220, 220f
  LAN as, 218
  MAN as, 218
  near-field, 235–236
  PAN as, 218–219
  relationships of, 219
  RFID as, 235–236, 236f
  SAN as, 219
  with Sonet, 226–228, 226t
  topologies of, 219–220, 220f, 234f
  types of, 218–219
  WAN as, 218
  Wi-Fi, 231–233
  wired, 220–228
  wireless, 230–236, 231f
  ZigBee, 233–234
Nibble, 106
Nickel metal hydride battery, 30t
Nickel-cadmium battery, 30t
Noise, in radio, 160b
Noise-cancelling headphones, 258–259

O

Octet, 106
OFDM. *See* Orthogonal frequency division multiplexing
Ohm's law, 46b
On-off keying (OOK), 174, 175f
OOK. *See* On-off keying
Operating system (OS), 326–327
Operation code, 147, 149f
Orthogonal frequency division multiplexing (OFDM)
  digital modulation with, 204–206, 205f–206f, 211
  spectral efficiency of, 209t
OS. *See* Operating system
Oscillator, 77–78. *See also* Clock; Voltage-controlled oscillator
  in AC, 26, 34
  frequency of, 77–78
  types of, 77f
OTN. *See* Optical transport network
OTT. *See* Over the Top TV
Optical transport network, 228
Over the Top TV (OTT), 285

P

PAL. *See* Programmable array logic
PAN. *See* Personal area network
Parabolic dish, 186, 186f
Patch antenna, 183, 184f
PCB. *See* Printed circuit board
Peripherals, 144. *See also* Serial peripheral interface
Personal area network (PAN), 218–219
Phase detector, 79–80, 80f
  as circuit, 79–80
Phase inversion, 54–55
Phase modulation, 174
Phase shift keying (PSK), 174, 175f
Phase-locked loop (PLL), 80–82
  divide ratio of, 84t
  in frequency synthesizers, 82–83, 83f
  LPF in, 80
  VCO in, 80–82
Phonograph record, 246, 247b
Photo cell, 299–302, 299f
Photodiode, 300f
Photoelectric sensor, 299–302
Phototransistor, 299f, 300, 302f
Photovoltaic (PV), 31–32, 299
PIC processor (Microchip Technology), 154
PICAXE, 339–340
Piezoelectric pressure transducer, 298
Pixel, 266, 267f, 277f, 289b
PLA. *See* Programmable logic array
Plasma screen
  operation of, 143
  TVs made with, 261–262
PLC. *See* Programmable logic controller
PLD. *See* Programmable logic device
PLL. *See* Phase-locked loop
Pneumatic device, 312
Polarization, radio, 163b
Potential, 162
Potentiometer, 298
Power amplifier, 65
Power gain, 71
Power supply, 84–100
  battery as, 84
  bus-oriented architecture in, 87
  for DC, 33
  DC-DC converters in, 87
  inverters as, 87
  power management in, 87
  regulators in, 85–86
  solar, DC-to-AC, 88f
  standard, 84–85, 85f

PowerPC (Freescale), 155
Preamplifier, 244
Printed circuit board (PCB)
  in kit, 88b–89b
  patch antenna on, 184f
  soldering, 347, 351, 351f
Processor, 140. *See also* Central processing unit; Microprocessor
Program
  counter, 147, 147f
  CPU's execution of, 149–150
  in RAM, 150f
  storage concept for, 150
Programmable array logic (PAL), 129
Programmable logic array (PLA), 129
Programmable logic controller (PLC), 318–323
  diagram of, 318f
  operation of, 319–320
  organization of, 318–319
  programming of, 320–321
  scan cycle of, 319f
Programmable logic device (PLD), 129–131, 130f
Programming language, 326–329
  assembly, 327–328
  BASIC, 328
  C, 328–329
  with PLC, 320–321
PSK. *See* Phase shift keying
Pulse filter into DC, 80b
Pulse width modulation (PWM), 68–69
  duty cycle and, 85
PV. *See* Photovoltaic
PWM. *See* Pulse width modulation

R

Radar
  project, 193b
  transmission frequency of, 28t
Radio, 162. *See also* Duplex communication; Modulation; Sidebands
  in aircraft, 187
  AM, 171, 171f
  AV receiver with, 254
  bandwidth of, 172, 172f
  CB, 187, 193b
  cell phone as, 196
  communication system, 159–160, 160f
  digital, 192b, 260b
  driver for, 177
  electromagnetic field as, 162–164
  family, 188
  frequency amplifier, 65
  frequency types, 28t, 168–170
  GPRS, 201–202
  ham, 188
  hardware of, 176–181
  LOS communication with, 170, 170f
  marine, 188–189
  microwave, 169–170
  mobile, 189
  modes of, 160–161
  multipath transmission of, 196b–197b
  noise in, 160b
  polarization of, 163b
  propagation of, 168–170
  range and reliability of, 196b–197b
  receiver, 177–179, 178f
  shortwave, 169f
  signal speed and strength of, 164
  simplex, 161, 161f
  spread spectrum in, 175–176
  transceiver, 179
  transmission lines in, 183–185, 185f
  transmitter, 176–177, 177f
  two-way, 186–189
Radio astronomy, 191
Radio frequency identification (RFID), 235–236, 236f
Radio telescope, 191
RAM. *See* Random access memory
Random access memory (RAM), 126–128
  addressing in, 126f
  dynamic, 127, 127f
  program stored in, 150f
  static, 126–127, 127f
Raspberry Pi, 341–342
Reactance, inductive, 50–51
Read-only memory (ROM), 128–129
Receiver. *See also* Audio-video receiver
  radio, 177–179, 178f
  satellite TV, 285f
  superheterodyne, 178f
  TV, 274
Recommended Standard 232 (RS-232), 155b–157b, 156f
Recommended Standard 485 (RS-485), 155b–157b
Record, phonograph, 246, 247b
Reduced instruction set computer (RISC), 155
Reed relay, 305, 305f
Register
  accumulator, 145
  in CPU, 143, 145
  data manipulation in, 145f
  flag, 148

Register (*Continued*)
  general-purpose, 144
  instruction, 147
  memory address, 147–148, 147f
  operation of, 145–151
  shift, 121, 122f
  stack pointer, 148
  storage, 120, 121f
Regulator
  linear, 85
  in power supplies, 86b
  switching, 85
Relay, 303–309
  common contacts of, 304f
  current control with, 305–306
  diagram of, 303f
  industrial control with, 305
  latching of, 307, 308f
  reed, 305, 305f
  remote control with, 306
  as switch, 303–304
  transistor with, 308f
Resistance, capacitive, 49
Resistive sensor, 296f
Resistive temperature device (RTD), 296, 296f
Resistors, 45–46
  color codes for, 48f
  symbol for, 47f
  types of, 47f
  in voltage divider, 48f
Resolution
  in data conversion, 133–134
  of HDTV, 268
  of video, 269
RFID. *See* Radio frequency identification
Ring network topology, 220, 221f
RISC. *See* Reduced instruction set computer
Robots, 11
  building of, 322b
  as control device, 11
  end effectors of, 11
  manipulator arm of, 11, 12f
ROM. *See* Read-only memory
Routers, 229
RS-232. *See* Recommended Standard 232
RS-485. *See* Recommended Standard 485
RTD. *See* Resistive temperature device

S

Sampling
  in data conversion, 132–134, 132f, 134f
  sound, 249f
SAN. *See* Storage area network
Satellite
  communication, 189f, 190
  geosynchronous orbit of, 190
  transmission frequency of, 28t
  transponder of, 190f
Satellite TV, 283–285
  distribution system for, 283f
  in home entertainment system, 284f
  receiver diagram with, 285f
SAW. *See* Surface acoustic wave
Scanning, 262, 263f
  interlaced, 265f
  progressive, 268
SCR. *See* Silicon-controlled rectifier
Self-induction, 50–51
Semiconductor, 20. *See also* Metal oxide semiconductor-field-effect transistor; Thyristor
  carbon as, 20
  germanium as, 20
  silicon as, 20
Sensor
  of flow, 302
  in industrial control, 295–302
  photoelectric, 299–302
  potentiometer as, 298
  pressure, 297–298
  resistive, 296f
  switch as, 298
  of temperature, 296–297
Serial peripheral interface (SPI), 155b–157b
Shift key, 174, 175f. *See also* Quadrature amplitude modulation
  with ASK, 174, 175f
  with BPSK, 174, 175f, 209t
  with FSK, 174, 175f, 209t
  with Gaussian minimum, 201–202
  with OOK, 174, 175f
  with PSK, 174, 175f
Shift register, 121, 122f
Sidebands
  on FM, 173–174
  from modulation, 171–172
  vestigial, 272–273
Signal, 7. *See also* Digital signal processing
  in AC, 26, 27f
  from amplifier, 65
  analog, 34, 63–64
  binary, 26, 103, 104f
  digital, 26, 34
  distortion of, 66–67
  radio, 164

Silicon
  diode made of, 300
  as semiconductor, 20
Silicon-controlled rectifier (SCR), 312–315, 313f–314f
Silver oxide battery, 30t
Simplex communication, 161, 161f
Smartphone, 195–196, 211–213, 215–216
SoC. *See* System on a chip
Software, 325–329
Solar cells
  as DC source, 31–32
  in panel, 32f
  symbol for, 299f
Solar power system, DC-to-AC, 88f
Solder, 347–348, 353b
Soldering, 347, 349–351, 350f–351f, 352b
Soldering iron, 347–348, 348f
Solenoid, 309, 309f
Sonar, 259
Sonet. *See* Synchronous optical network
Sound, 243–245
  compression of, 248–249
  digitization of, 247–248, 248f
  frequency of, 28t
  sampling of, 249f
  in special applications, 257–260
  surround, 256–257, 256f
Speakers, 244–245
  diagram of, 244f
  surround sound with, 256–257, 256f
  as transducer, 244
  types of, 245
Spectral efficiency, 208, 209t
Spectrum
  electromagnetic frequency, 28t, 164–168, 165f
  space, 166b–167b
  spread, 175–176
SPI. *See* Serial peripheral interface
Spread spectrum, 175–176
Stack, 148
Stack pointer register, 148
Star network topology, 17, 19f
Static electricity, 16–18
Stepper motor, 311
Storage area network (SAN), 219
Storage register, 120
  FFs in, 121f
Stored program concept, 150
Strain gauge, 297, 298f
Streaming audio, 257
Subroutines, 327
Surface acoustic wave (SAW), 77
Surround sound, 256–257, 256f
Switch, 45
  as component, 42–43, 45, 55
  limit, 298
  MOSFET, 55f
  relay as, 303–309
  schematic of, 45
  as sensor, 298
Switching
  amplifier, 68–69, 68f
  of LED, 61b
  mobile telephone, 199
  regulator, 85–86
  transistor, 55
Synchronous digital hierarchy. *See* Synchronous optical network
Synchronous optical network (Sonet), 226–228
Synthesizer, frequency, 82–83, 83f
System on a chip (SoC), 63
Systems approach, 1, 41–45

T

Tachometer, phototransistor in, 302f
TDD. *See* Time-division duplexing
TDMA. *See* Time division multiple access
Telephone, mobile switching of, 199
Telephone data system, 227b
  ATM in, 227b
Television (TV). *See also* Cable TV; High-definition TV; Internet protocol TV; Satellite TV
  ADC for, 267f
  analog, 264–266
  aspect ratio of, 262–263, 264f
  on cell phone, 285–286
  diagram of, set, 119f
  digital, 266–269
  frequency of, 28t
  inputs and outputs of, set, 119f
  international systems for, 266b
  over fiber optic cable, 281b
  with plasma screen, 274
  receiver, 274
  screen technology of, 274–276, 277f
  transmission of, 272–274, 273f
Temperature sensor, 296–297
Thermistor, 296
Thermocouple, 296
Thyristor
  control with, 312–317
  diac, 315
  SCR, 312–315, 313f–314f
  triac, 315–317, 315f

Time division multiple access (TDMA), 209–210, 210f
Time-division duplexing (TDD), 210
Timer, 154
Transceiver
  ISM band, 179–181
  radio, 179
  single-chip, 180f
Transducer
  piezoelectric pressure, 298
  speaker as, 244–245
Transformer, 51–52, 51f
Transistor, 52–55. *See also* Metal oxide semiconductor-field-effect transistor
  bipolar junction, 52–53, 53f–54f
  as component, 7, 52–55, 104
  IC in, 7
  photo-, 299f, 300, 302f
  relay with, 308f
  switching, 55
  symbols for, 53f
  temperature sensing with, 297
  types of, 52–53
Transmitter
  radio, 176–177, 177f
  TV, 272f
  UART as, 155b–157b, 156f
Transponder, satellite, 190f
Triac, 315–317, 317f
Truth table, 115
  for AND gate, 115f
  for OR gate, 117f
  inverter for, 114f
  for NAND gate, 118f
  for NOR gate, 118f
  for XNOR gate, 119f
  for XOR gate, 119f
TV. *See* Television
Twisted pair cable, 221–222, 222f

U
UART. *See* Universal asynchronous receiver-transmitter
Ultra high frequency radio, 169–170
Universal asynchronous receiver-transmitter (UART), 155b–157b, 156f
Universal serial bus (USB), 155b–157b
USB. *See* Universal serial bus

V
Vacuum tubes, 258b
Valve, 312
Varactor, 52
VCO. *See* Voltage-controlled oscillator
Very high frequency radio, 169–170
Video, 261–277
  aspect ratio of, 262–263, 264f
  CCD for, 270
  color, 269–270, 269f–270f
  compression of, 271
  monochrome, 269
  persistence of vision in, 263–264
  resolution of, 268b
  scanning in, 262, 263f, 265f, 266, 268
Virtual reality (VR), 288
Vision, persistence of, 263–264
Voice
  coil, 244–245
  sampling of, 249f
Voltage, 3–5, 7
  of AC, 28–29
  of battery cell, 30t
  in current flow, 15
  divider, 48f
  as electrical potential difference, 17
  gain, 71
  induction of, 23f
  measurement of, 37b–38b
  sources of, 24–29
Voltage-controlled oscillator (VCO), 77–78, 77f
  in PLL, 80–82

W
WAN. *See* Wide area networks
Wave. *See also* Microwave
  short, 169f
  surface acoustic, 77
Waveguide, 185–186
Wavelength, in electromagnetism, 167b–168b
Wide area networks (WAN), 218
Wi-Fi, 231–233
Wireless, 147. *See also* Radio
  broadband, 238
  networks, 230–236, 231f
  trend in, 227b
Word
  binary data, 106
  instruction, 148–149, 149f

Y
Yagi antenna, 182–183, 184f

Z
ZigBee, 233–234

Electronic Engineering

ELECTRONICS EXPLAINED

Fundamentals for Engineers, Technicians, and Makers

Second Edition

Louis E. Frenzel Jr.

Take a fresh, systems based approach to understanding how electronics work!

Electronics Explained takes a systems based approach to the fundamentals of electronics, covering the different types of electronic circuits, how they work, and how they fit together to create modern electronic equipment, enabling you to apply, use, select, operate and discuss common electronic products and systems.

This new edition has been updated to show the latest technological trends with added coverage of:

- Internet of Things (IoT)
- Machine-to-Machine (M2M) technology
- Ethernet to 100 Gb/s
- Wi-Fi, Bluetooth, and other wireless technologies
- 5G New Radio cellular standards
- Microcontrollers and programming with the Arduino, BASIC Stamp, and others

With this book you will:

- Learn about the basic components of electronics such as resistors, capacitors, inductors, transformers, diodes, transistors, and integrated circuits.
- Discover different types of circuits, using the functional block diagram approach which makes it easy to understand their purpose and application.
- Get involved with Hands-On projects in each chapter, using components and ICs with the breadboarding socket.

About the author

Louis E. Frenzel Jr. has been a Technology Editor with Electronic Design magazine since 2000 and covers the wireless, communications, and networking and test/measurement fields Lou has over 30 years of electronic industry experience as an engineer in the geophysical, aerospace and computer fields. Prior to joining ED he was a college professor teaching electronics and semiconductor manufacturing and working on several NSF grants. He holds bachelors and masters degrees and is author of 21 books on communications, computers and electronics.

Cover image © iStockphoto

ELSEVIER

Newnes
An imprint of Elsevier
elsevier.com/books-and-journals

If you enjoyed this book please post a review to your favorite online bookstore today.

ISBN 978-0-12-811641-8

9 780128 116418

US008508907B2

(12) **United States Patent**
Mayer

(10) **Patent No.:** **US 8,508,907 B2**
(45) **Date of Patent:** **Aug. 13, 2013**

(54) **METHOD AND DEVICE FOR COLLECTING ELECTROSTATIC CHARGE FROM THE ATMOSPHERE**

(75) Inventor: **Robert Mayer**, Manassas, VA (US)

(73) Assignee: **eFluxor LLC**, Manassas, VA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 299 days.

(21) Appl. No.: **12/976,015**

(22) Filed: **Dec. 22, 2010**

(65) **Prior Publication Data**

US 2012/0106023 A1 May 3, 2012

Related U.S. Application Data

(60) Provisional application No. 61/408,973, filed on Nov. 1, 2010.

(51) **Int. Cl.**

H05F 3/00	(2006.01)
H02N 1/00	(2006.01)
H02N 1/04	(2006.01)

(52) **U.S. Cl.**
USPC **361/221**; 361/231; 310/309; 310/310

(58) **Field of Classification Search**
USPC 361/221, 231; 310/309, 310
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2,644,903	A *	7/1953	Lehrer	310/309
2,649,566	A *	8/1953	Morel	322/2 A
3,514,644	A	5/1970	Fonberg	
3,600,632	A	8/1971	Bright	
3,614,481	A	10/1971	Halliday	
3,857,066	A	12/1974	Cline	
3,870,933	A	3/1975	Huber	
4,048,667	A	9/1977	Brennecke	
5,248,930	A	9/1993	Taylor	
5,506,491	A	4/1996	Ford	
5,592,357	A	1/1997	Rader	
5,923,130	A	7/1999	Roman	
6,498,291	B2	12/2002	Brammer	
2005/0040731	A1	2/2005	Schmidt	
2006/0284004	A1	12/2006	Jones	
2008/0042862	A1	2/2008	Chen	
2009/0158537	A1	6/2009	Ishijima	
2010/0007218	A1	1/2010	Ogram	
2010/0008011	A1	1/2010	Ogram	

* cited by examiner

Primary Examiner — Scott Bauer
(74) *Attorney, Agent, or Firm* — eFluxor LLC; Robert A. Mayer

(57) **ABSTRACT**

A device collects electrostatic charge from the atmosphere and stores the electrostatic charge for further use. The device includes a primary array including a plurality of electrically-conducting collectors, a plurality of electrically-conducting inductors, a charge regulator, and a charge storage device. A secondary array can be added to improve the efficiency of the primary array.

6 Claims, 2 Drawing Sheets

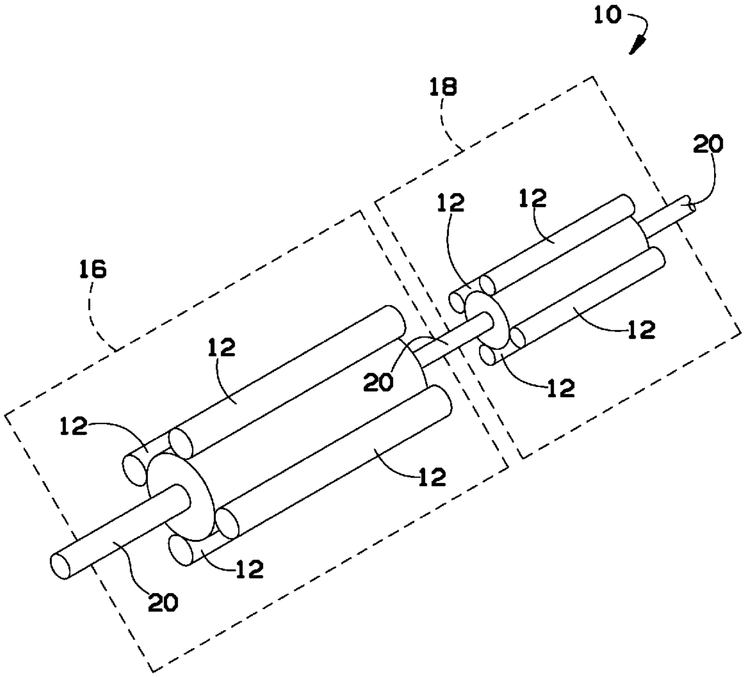

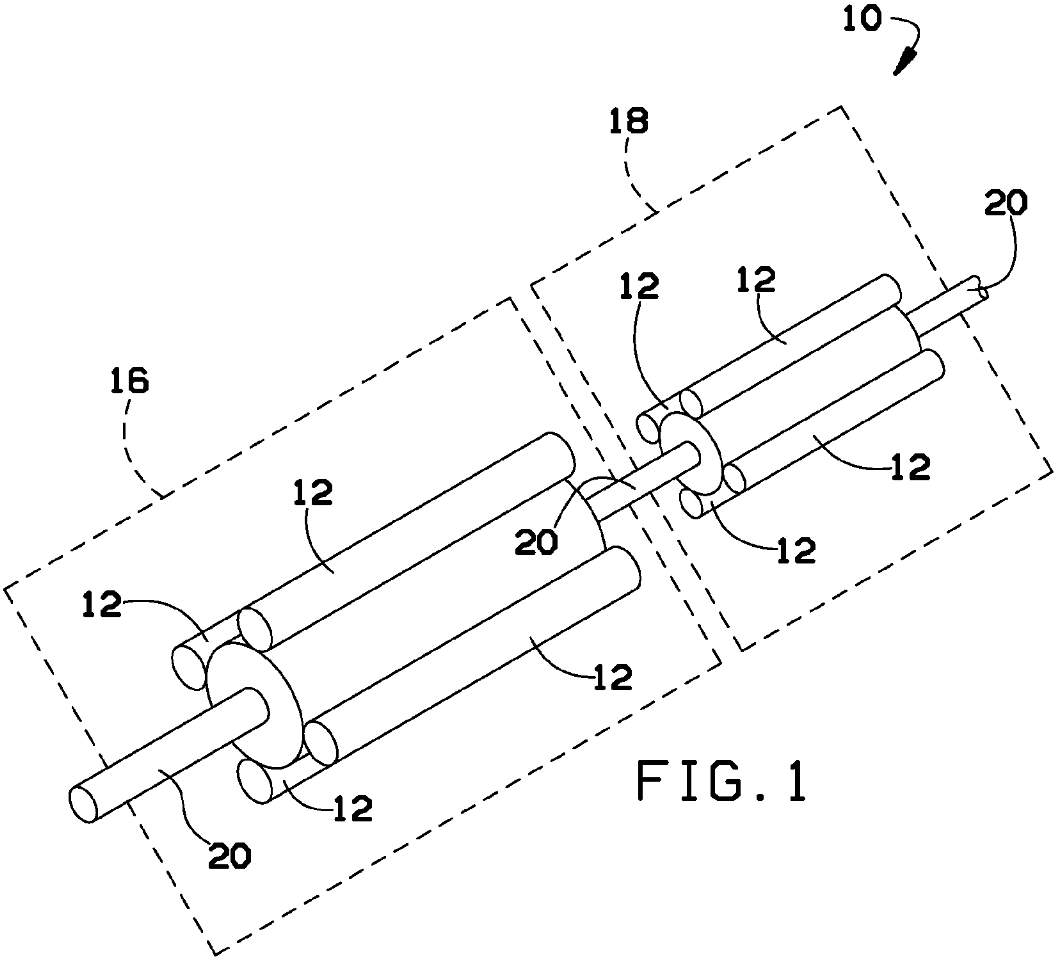

FIG. 1

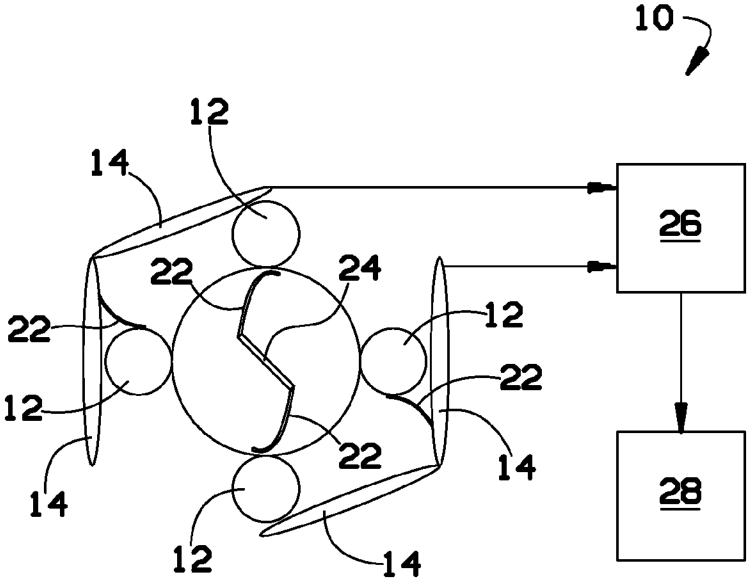

FIG. 2

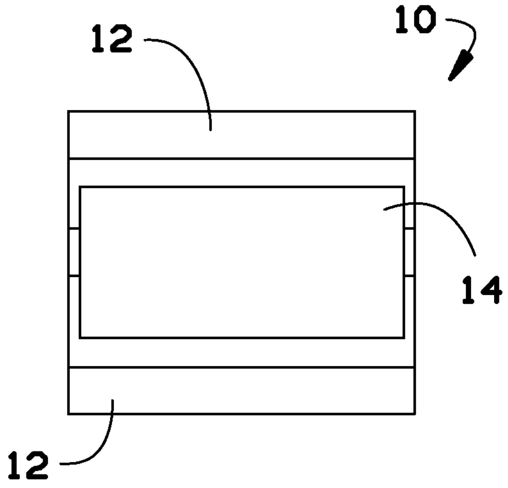

FIG. 3

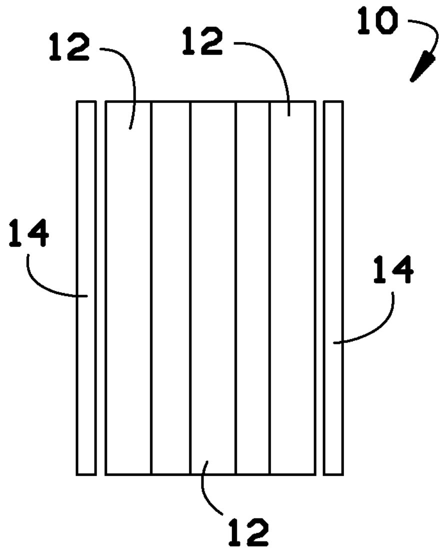
10
12
12
14
14
12

FIG. 4

METHOD AND DEVICE FOR COLLECTING ELECTROSTATIC CHARGE FROM THE ATMOSPHERE

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of Provisional Application No. 61/408,973 filed Nov. 1, 2010, the contents of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

The present invention generally relates to a method and apparatus for collecting excess electrostatic charge from the environment. More specifically, the present invention relates to a method and apparatus for collecting excess electrostatic charge from the environment and storing the electrostatic charge for dispersal or further use.

The buildup of static electricity frequently occurs in many different natural and commercial environments. People and objects may build up a high voltage charge. In addition, the air may acquire an undesirable charge. Unfortunately, when a person gets close to another person and/or an object, the difference in the voltage charge produces an undesirable discharge in the form of a spark. Sparks may be lethal to individuals and sensitive machines in some industrial environments.

Current devices for electrostatic charge collection typically rely on electrically-powered systems which operate at high voltages and are ineffective in actually collecting ion charges for controlled dispersal.

As can be seen, there is a need to provide a reliable device for electrostatic charge collection that operates at low voltage.

SUMMARY OF THE INVENTION

In one aspect of the present invention, a device for collecting electrostatic charge includes an array having a plurality of electrically-conducting collectors to collect the electrostatic charge, a plurality of electrically-conducting inductors electrically coupled to the electrically-conducting collectors, a charge regulator coupled to the electrically-conducting inductors, a charge storage device coupled to the charge regulator, a plurality of brushes for transferring collected electrostatic charge from the electrically-conducting collectors to the charge storage device; and a plurality of charge equalization wires coupled to the plurality of brushes.

In another aspect of the present invention, a second collector/inductor assembly can be connected to the first collector/inductor assembly, where each assembly includes an array having a plurality of electrically-conducting collectors to collect the electrostatic charge, a plurality of electrically-conducting inductors electrically coupled to the electrically-conducting collectors, a plurality of brushes for transferring collected electrostatic charge from the electrically-conducting collectors to the charge storage device; and a plurality of charge equalization wires coupled to the plurality of brushes.

These and other features, aspects and advantages of the present invention will become better understood with reference to the following drawings, description and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** illustrates a perspective front view of a device for collecting electrostatic charge from the atmosphere according to an exemplary embodiment of the present invention showing a primary assembly and a secondary assembly;

FIG. **2** illustrates a perspective top view of a device for collecting electrostatic charge from the atmosphere according to another exemplary embodiment of the present invention;

FIG. **3** illustrates a side view of the device of FIG. **2**; and

FIG. **4** illustrates a top view of the device of FIG. **2**.

DETAILED DESCRIPTION OF THE INVENTION

The following detailed description is of the best currently contemplated modes of carrying out exemplary embodiments of the invention. The description is not to be taken in a limiting sense, but is made merely for the purpose of illustrating the general principles of the invention, since the scope of the invention is best defined by the appended claims.

Various inventive features are described below that can each be used independently of one another or in combination with other features.

Broadly, embodiments of the present invention generally provide a rotating device for collecting electrostatic charge from the atmosphere and storing the electrostatic charge for dispersal or further use.

FIGS. **1** and **2** illustrates the device **10** for collecting electrostatic charge including an array having a plurality of electrically-conducting collectors **12** to collect the electrostatic charge, a plurality of electrically-conducting inductors **14** electrically coupled to the electrically-conducting collectors **12**, a charge regulator **26** coupled to the electrically-conducting inductors **14**, a charge storage device **28** coupled to the charge regulator **26**, a plurality of brushes **22** for transferring collected electrostatic charge from the electrically-conducting collectors **12** to the charge storage device **28**; and a plurality of charge equalization wires **24** coupled to the plurality of brushes **22**.

FIG. **1** illustrates an alternative embodiment of the present invention in which the device **10** may include a primary collector/inductor assembly **16** and a secondary collector/inductor assembly **18** mounted on a common shaft **20**. In one embodiment, the primary collector/inductor assembly **16** and secondary collector/inductor assemblies **18** may be directly connected by using rotational bearing assemblies attached to each end of the primary collector/inductor assembly **16** and the secondary collector/inductor assembly **18**.

The collectors **12** may have different shapes. In one embodiment, the collectors **12** may have a cylindrical, rectangular, pyramidal, planar, or elongated shape. The collectors **12** may be placed parallel to each other and arranged at approximately equal distances from an axis of rotation of their aggregate collector/inductor assemblies **16** and **18**. In one embodiment, the collectors **12** may be made of metal, composites, alloys, graphite, or any other conductive materials. The collectors **12** may have a solid body or a hollow body.

The exemplary invention embodiment shown as device **10** may include a plurality of inductors **14**. The inductors **14** may be operatively connected to the collectors **12** to extract charged ions from the ambient environment. In one embodiment, the inductors **14** may have a length greater than twice the radius of rotation of the rotating array with the plurality of conductors **12**.

The charge regulator **26** may be a standard commercially-available charge regulator. In one embodiment, the charge regulator **26** may be a spark gap. The charge regulator **26** may include electrical components that limit the voltage build-up in the rotating device **10** and the charge storage device **28**.

A plurality of brushes or brush equivalents **22** may be used to transfer accumulated charges between the collectors **12** and inductors **14** during the operation of the device **10**. The number of brushes **22** may be determined by the implemented configuration. In one embodiment, the device **10** may include four brushes **22**.

The charge equalization wire **24** may momentarily connect two of the collectors **12** under the influence of two inductors **14**. This wire may allow the transfer of charge between oppositely-charged collectors **12** at appropriate times in the collector assembly **16** during the operation of the device **10**.

When the collector **12** and inductor **14** assemblies of the device **10** are counter-rotated relative to each other by any appropriate means, the ion extraction process may be a normal result of the rotation of the device **10**. The collection and removal of the charge may be achieved by the unique combination of accepted electrostatic engineering principles and the new component configuration to create an electrostatic collection system which may be capable of low-voltage electrostatic ion collection capabilities.

The device **10** may be placed in a selected environment which contains excess electrostatic charge. The operation of the device **10** may be initiated by applying mechanically rotational power to the collector and/or inductor assemblies **12** to initiate the charge collection process, which continues until the rotation stops. The resulting operation may cause a transfer of the excess charge to the charge storage device **28** where it can be immediately dispersed and/or collected.

The elements of the device **10** may be in a number of different configurations which may operate at different efficiencies. The size, shape, physical relationships, and materials employed may be configured to have an apparently different appearance when actually they are alternative implementations of this invention.

The device **10** adapted to rotate for collecting electrostatic charge may be considered to be based on classic electrostatic accumulation principles, but is configured to operate at significantly lower voltages, thus requiring fewer operating components. These improvements may allow major simplifications in the construction and maintenance of the device **10** and its associated systems, reducing initial production costs, operational power requirements, and ongoing maintenance costs.

The present invention has been successfully demonstrated at the bench-scale level.

It should be understood, of course, that the foregoing relates to exemplary embodiments of the invention and that modifications may be made without departing from the spirit and scope of the invention as set forth in the following claims.

I claim:

1. A device for collecting electrostatic charge comprising:

a primary assembly, wherein the primary assembly includes;

an array having a plurality of electrically-conducting collectors to collect the electrostatic charge;

a plurality of electrically-conducting inductors electrically coupled to the electrically-conducting collectors;

a primary charge regulator coupled to the primary assembly electrically-conducting collectors;

a charge storage device coupled to the charge regulator;

a plurality of brushes for transferring collected electrostatic charge from the electrically-conducting collectors to a charge storage device; and

a plurality of primary assembly charge equalization wires coupled to the plurality of primary assembly brushes; and

a secondary assembly electrically connected to the primary assembly, wherein the secondary assembly includes:

an array having a plurality of secondary assembly electrically-conducting collectors to collect electrostatic charge;

a plurality of secondary assembly electrically-conducting inductors electrically coupled to the secondary assembly electrically-conducting collectors;

a secondary charge regulator electrically coupled to the secondary assembly electrically-conducting collectors and the primary assembly electrically-conducting collectors;

a plurality of brushes for transferring collected electrostatic charge from the secondary assembly electrically-conducting inductors to the secondary assembly electrically-conducting collectors; and

a plurality of secondary assembly charge equalization wires coupled to the plurality of secondary assembly brushes.

2. The device according to claim **1**, wherein the sizes of the primary and secondary assemblies are different from each other.

3. The device according to claim **1**, wherein the primary and secondary assemblies are mounted and rotated on the same shaft.

4. The device according to claim **1**, wherein the primary and secondary assemblies are mounted and rotated on separate shafts.

5. The device according to claim **1**, wherein the primary assembly collector array and the secondary assembly collector array do not rotate.

6. The device according to claim **1**, wherein the primary assembly inductor array and the secondary assembly inductor array rotate.

* * * * *

United States Patent [19]

Breaux

[11] **4,127,804**

[45] **Nov. 28, 1978**

[54] ELECTROSTATIC ENERGY CONVERSION SYSTEM

[75] Inventor: **Onezime P. Breaux**, Trotwood, Ohio

[73] Assignee: **The United States of America as represented by the Secretary of the Air Force**, Washington, D.C.

[21] Appl. No.: **689,272**

[22] Filed: **May 24, 1976**

[51] **Int. Cl.**² **H02N 1/00**

[52] **U.S. Cl.** .... **322/2 A**; 310/308; 320/1

[58] **Field of Search** .... 322/2 A; 310/5, 10, 310/308, 309; 320/1

[56] **References Cited**

U.S. PATENT DOCUMENTS

2,540,327	2/1951	Felici	322/2 A X
3,094,653	6/1963	LeMay et al.	322/2 A
3,971,938	7/1976	O'Hare	310/5 X
4,054,826	10/1977	Wahlstrom	322/2 A

Primary Examiner—Robert J. Hickey

Attorney, Agent, or Firm—Joseph E. Rusz; Robert Kern Duncan

[57] ABSTRACT

Two inversely ganged variable capacitors having a common movable element varying the capacities such that when one capacitor is at maximum capacitance, the other is at minimum capacitance, have, after initially charging, a substantially constant contained or trapped charge distributed between the two capacitors. As the capacities of the capacitors are varied, a potential difference is developed between the fixed plate of one capacitor and the fixed plate of the other. When these two potential points are connected to a load, charge is transferred from one capacitor to the other as a current flow through the load. The total charge is not diminished, the energy supplied the load being the energy expanded on moving the movable charged plates. As the movable plates are moved in a cyclic manner an alternating current is provided to the load.

8 Claims, 19 Drawing Figures

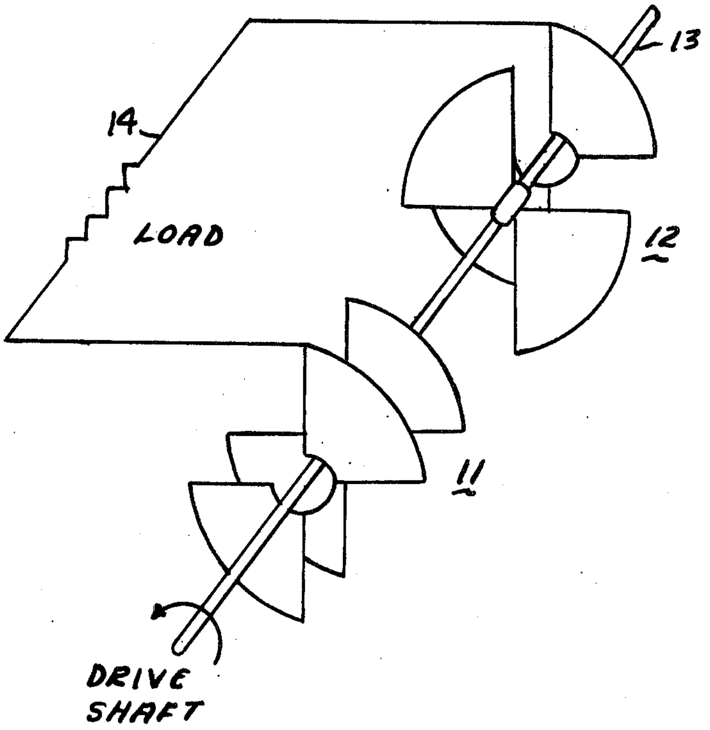

Fig-1

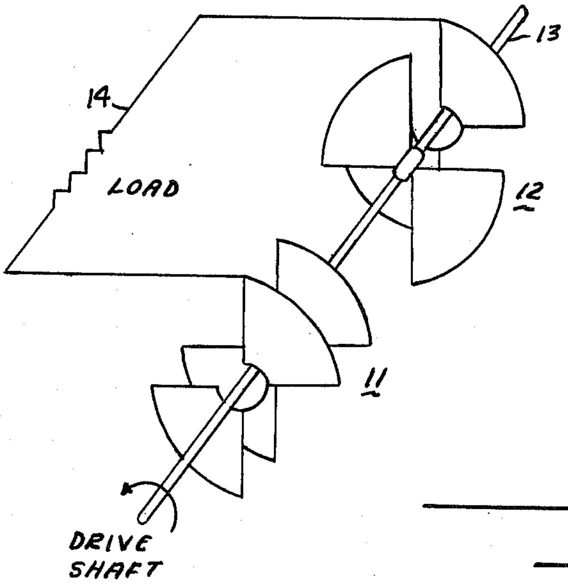
13
14
LOAD
12
11
DRIVE SHAFT

Fig-2

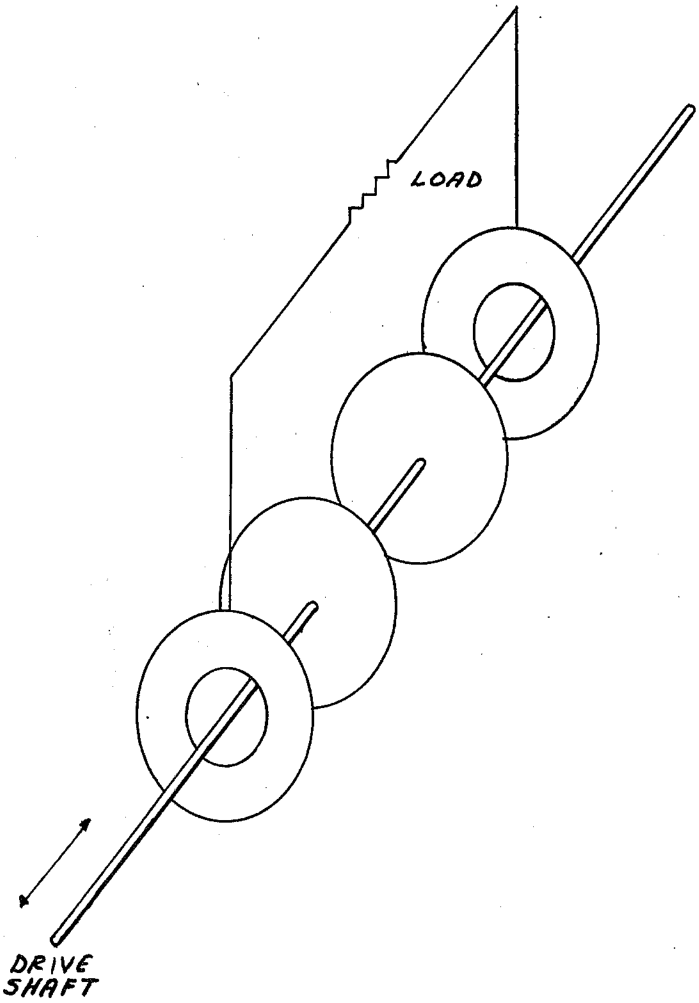
LOAD
DRIVE SHAFT

Fig-3

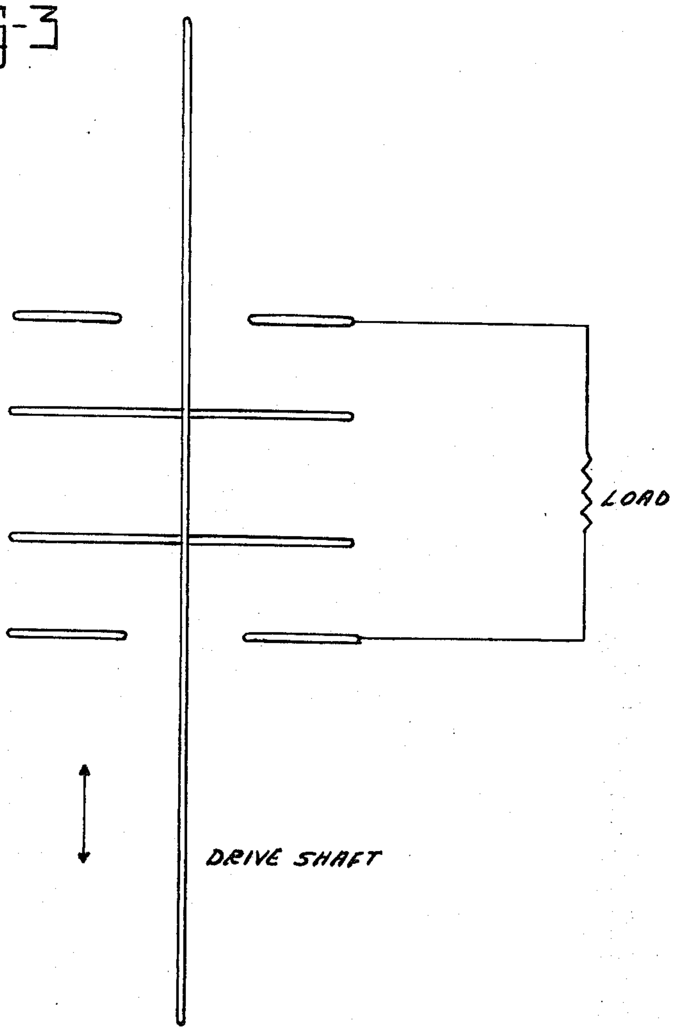
LOAD
DRIVE SHAFT

Fig-5

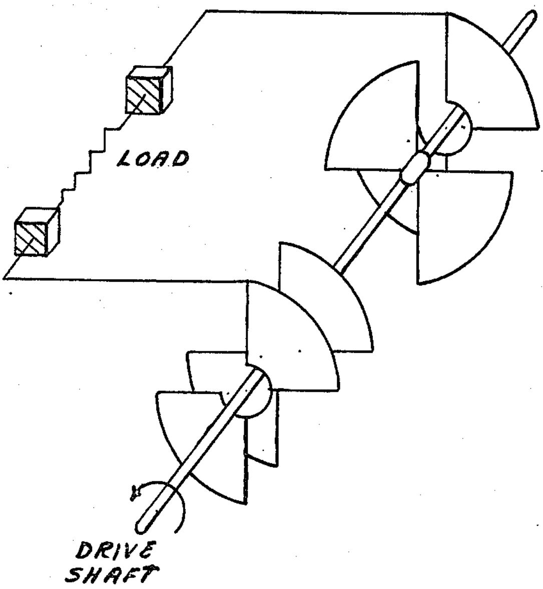
LOAD
DRIVE SHAFT

Fig-4

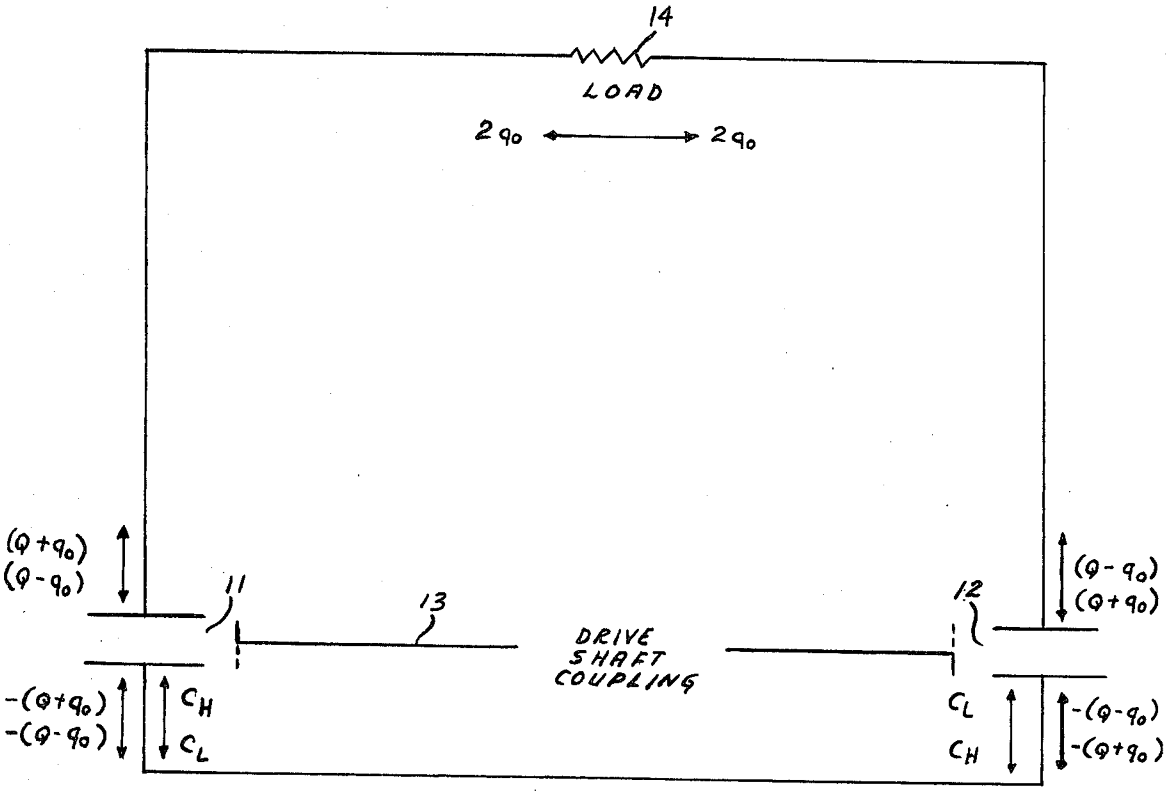
14
LOAD
$2q_0$
$2q_0$
$(Q+q_0)$
$(Q-q_0)$
11
13
DRIVE SHAFT COUPLING
12
$(Q-q_0)$
$(Q+q_0)$
$-(Q+q_0)$
$-(Q-q_0)$
C_H
C_L
C_L
C_H
$-(Q-q_0)$
$-(Q+q_0)$

Fig-6

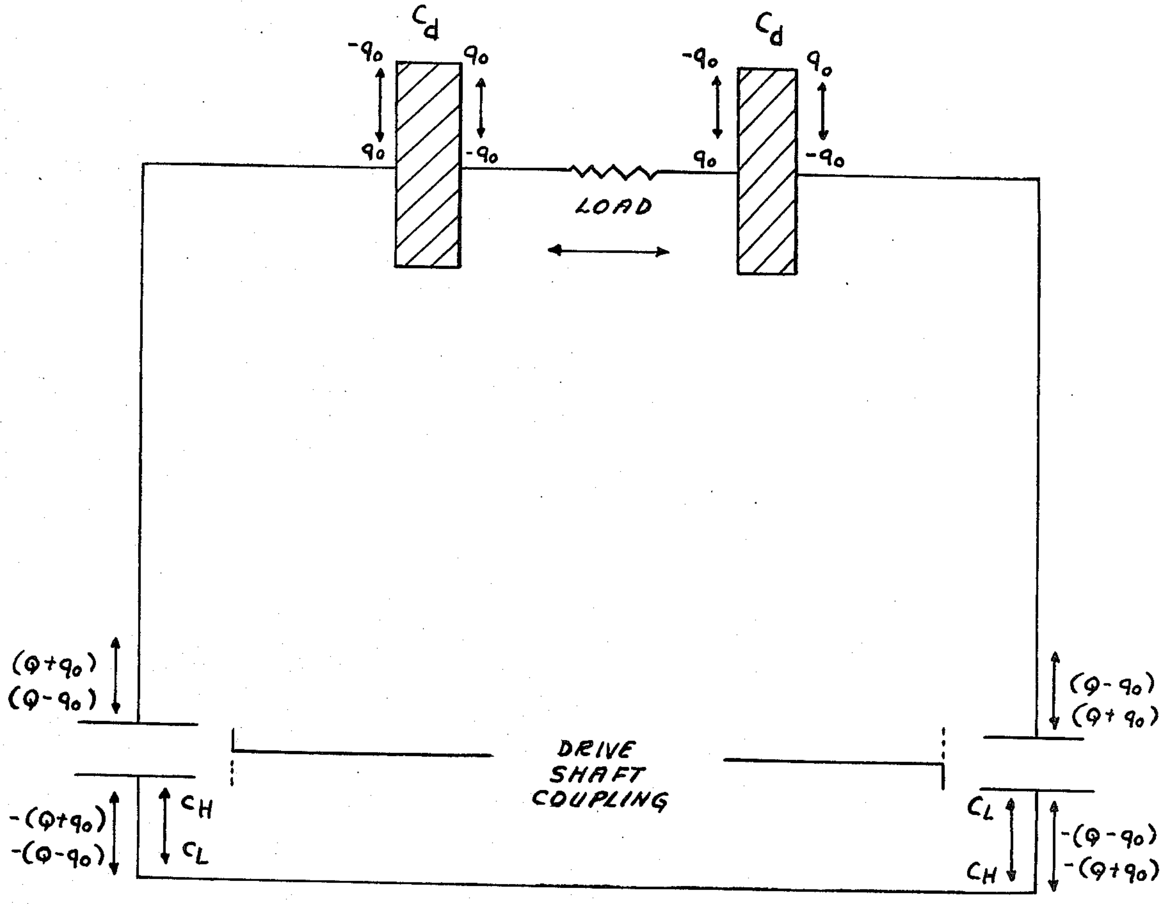
C_d
$-q_o$
q_o
q_o
$-q_o$
C_d
$-q_o$
q_o
q_o
$-q_o$
LOAD
$(Q+q_o)$
$(Q-q_o)$
$(Q-q_o)$
$(Q+q_o)$
DRIVE SHAFT COUPLING
$-(Q+q_o)$
$-(Q-q_o)$
C_H
C_L
C_L
C_H
$-(Q-q_o)$
$-(Q+q_o)$

Fig-7

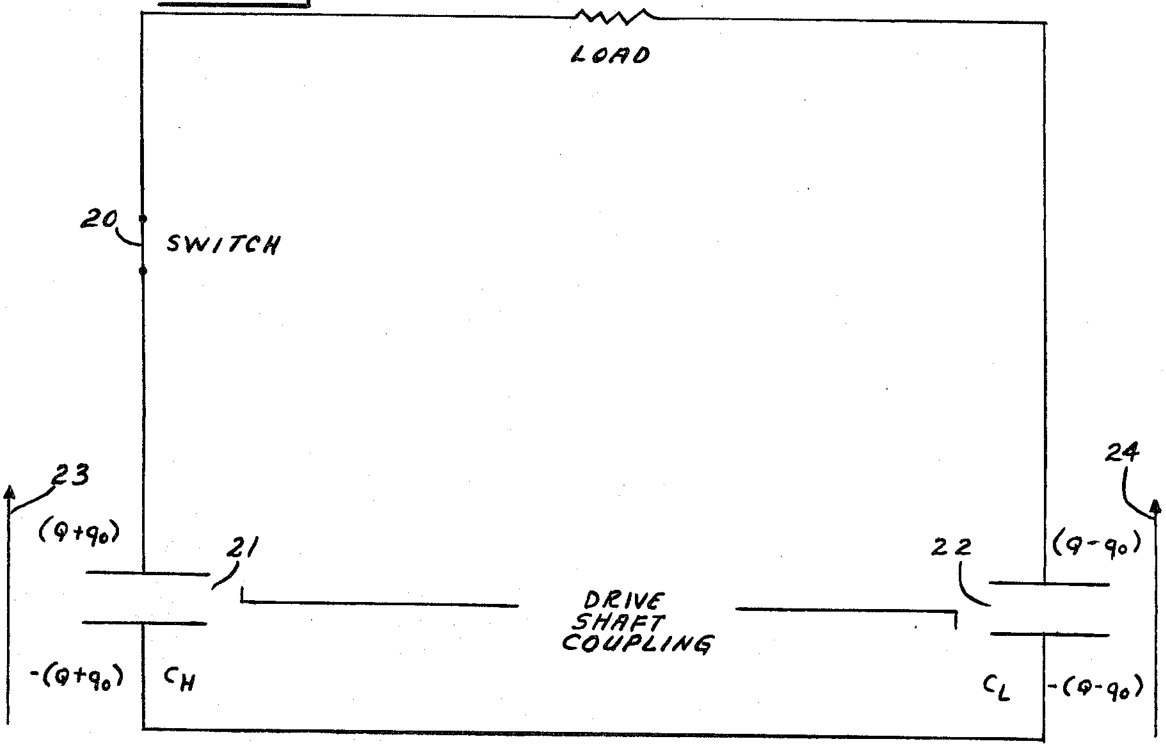
LOAD
20
SWITCH
23
$(Q+q_0)$
21
$-(Q+q_0)$
C_H
DRIVE SHAFT COUPLING
24
22
$(Q-q_0)$
C_L
$-(Q-q_0)$

Fig-8

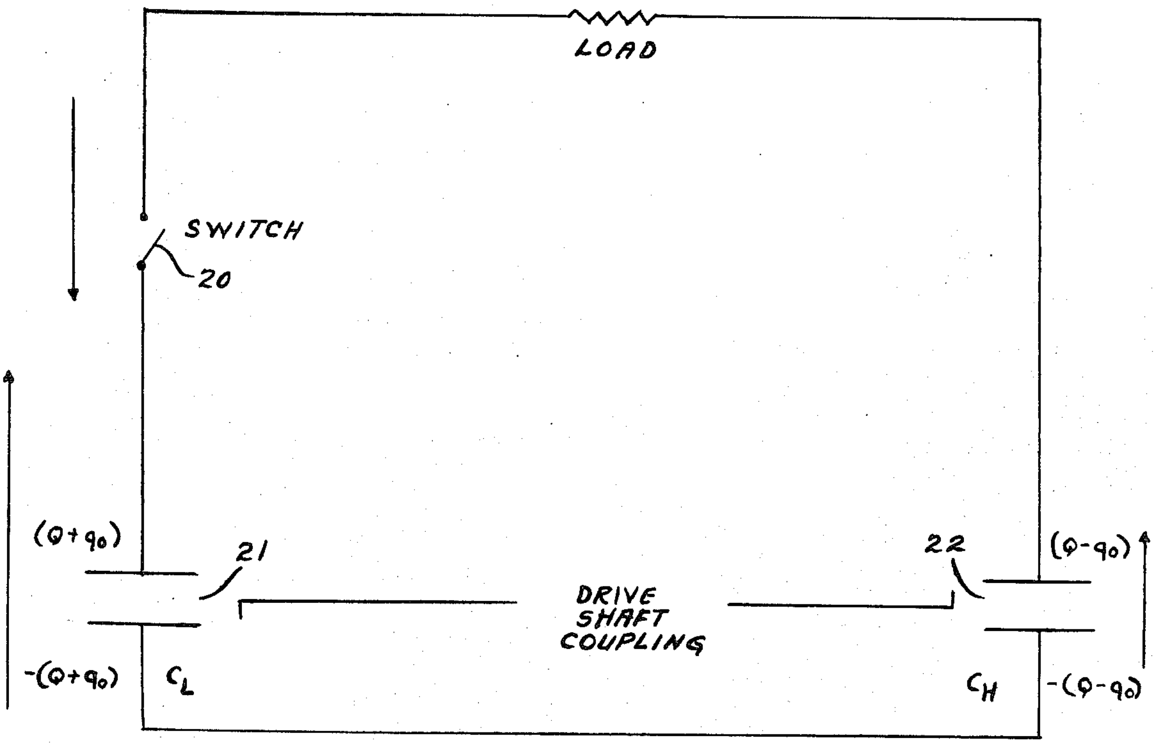
LOAD
SWITCH
20
$(Q+q_0)$
21
$-(Q+q_0)$
C_L
DRIVE SHAFT COUPLING
22
$(Q-q_0)$
C_H
$-(Q-q_0)$

Fig-9

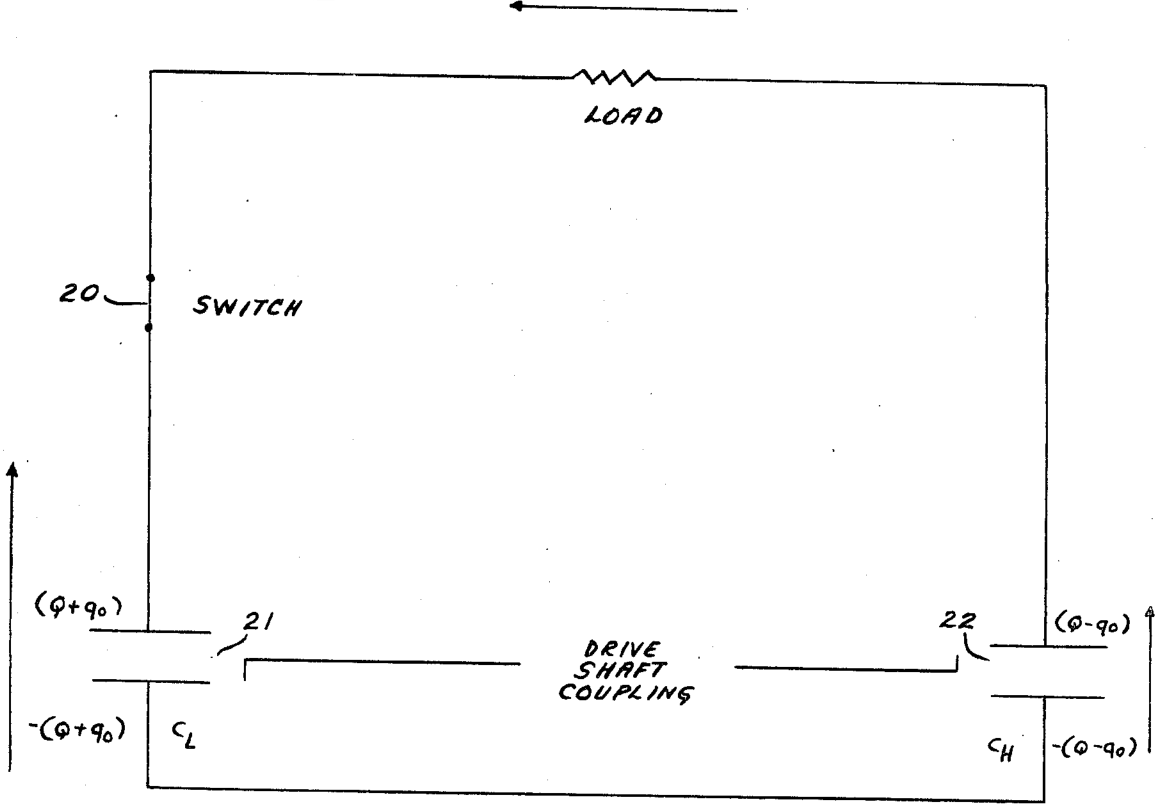
LOAD
20
SWITCH
$(\varphi+q_0)$
21
DRIVE SHAFT COUPLING
22
$(\varphi-q_0)$
$-(\varphi+q_0)$
C_L
C_H
$-(\varphi-q_0)$

Fig-10

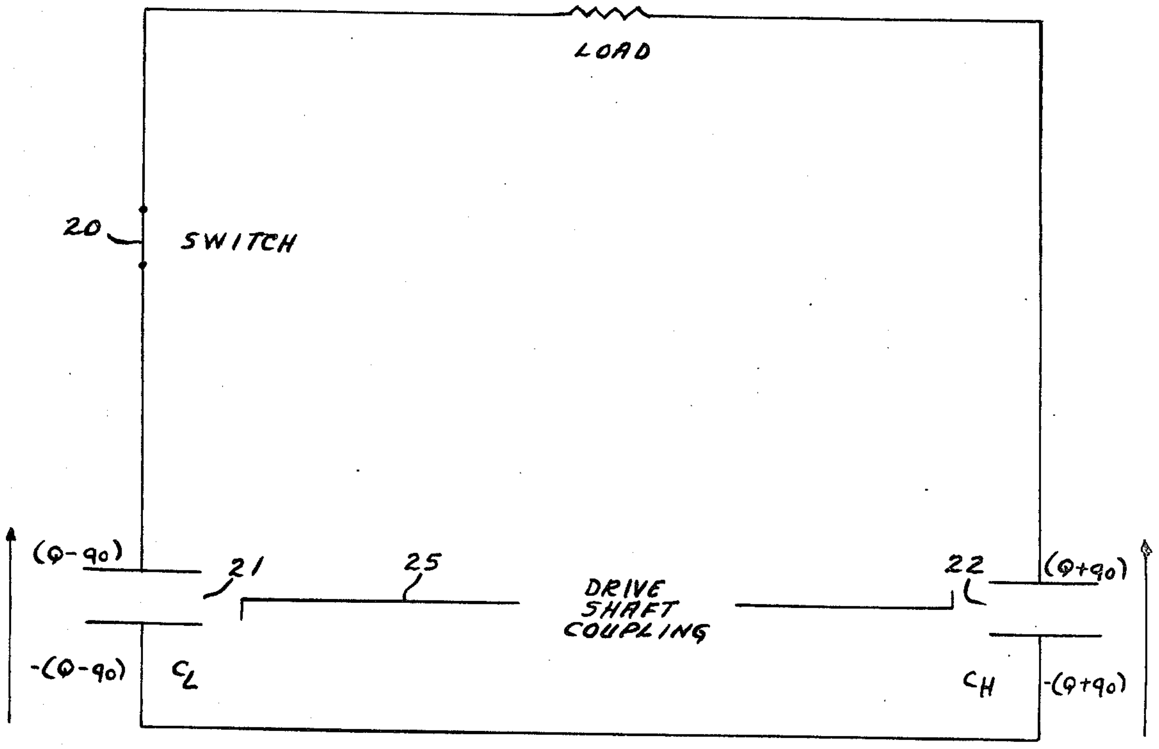
LOAD
20
SWITCH
$(\varphi-q_0)$
21
25
DRIVE SHAFT COUPLING
22
$(\varphi+q_0)$
$-(\varphi-q_0)$
C_L
C_H
$-(\varphi+q_0)$

Fig 11

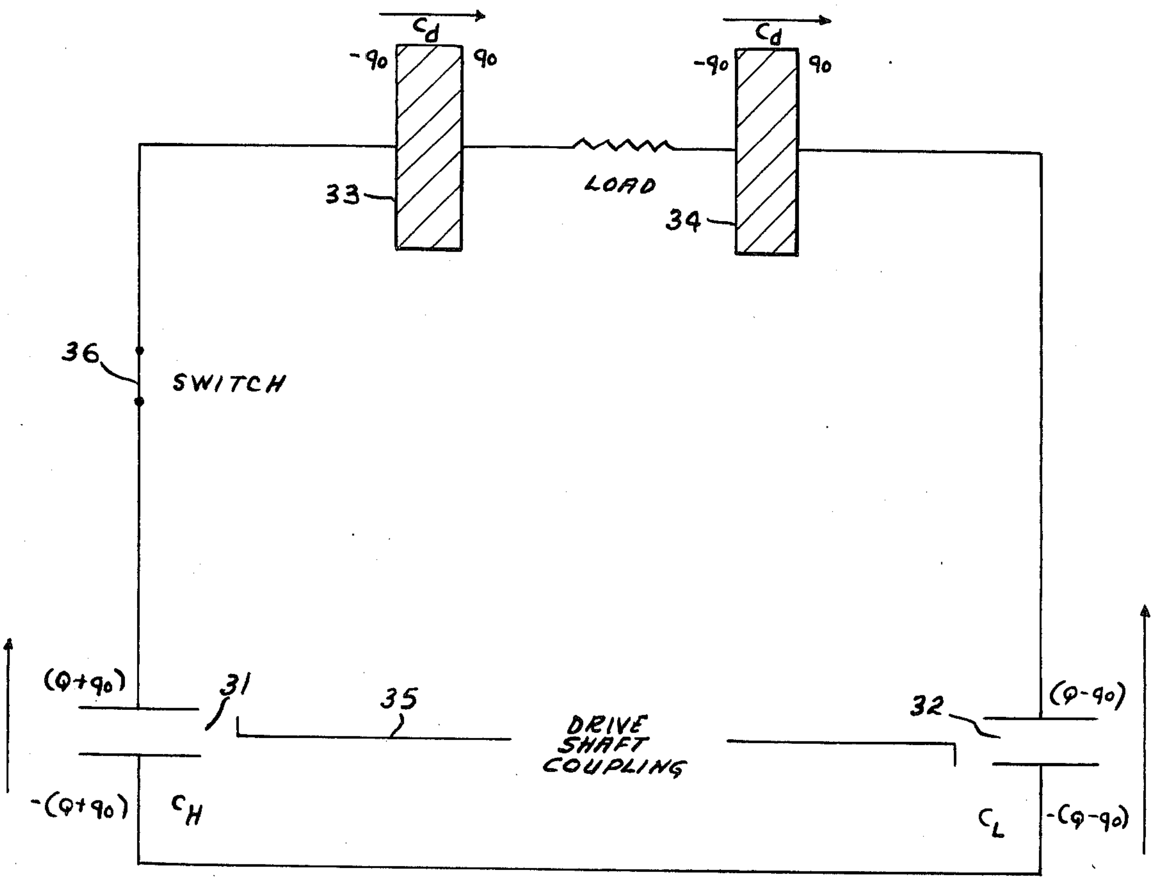
c_d
$-q_0$
q_0
c_d
$-q_0$
q_0
33
LOAD
34
36
SWITCH
$(Q+q_0)$
31
35
DRIVE SHAFT COUPLING
32
$(Q-q_0)$
$-(Q+q_0)$
C_H
C_L
$-(Q-q_0)$

Fig-12

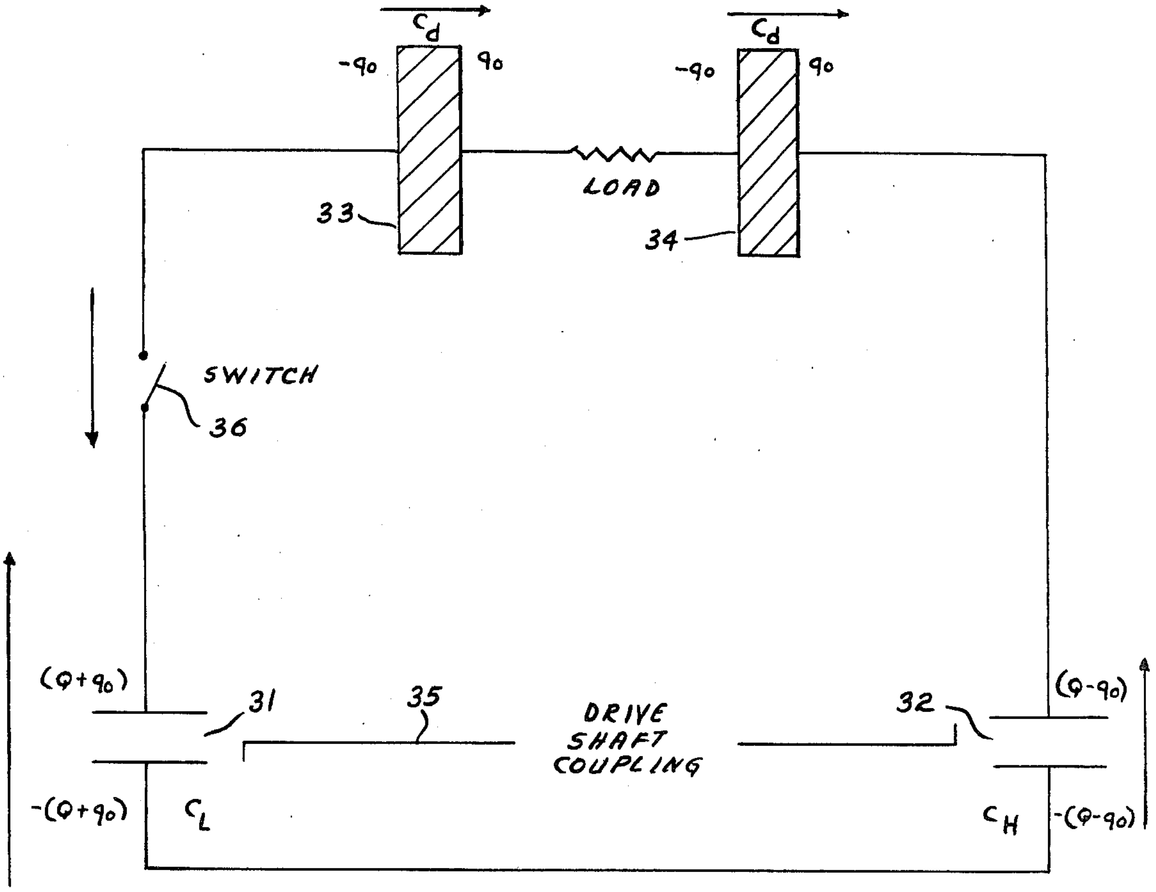
C_d
$-q_0$
q_0
C_d
$-q_0$
q_0
LOAD
33
34
SWITCH
36
$(Q+q_0)$
31
35
DRIVE SHAFT COUPLING
32
$(Q-q_0)$
$-(Q+q_0)$
C_L
C_H
$-(Q-q_0)$

Fig-13

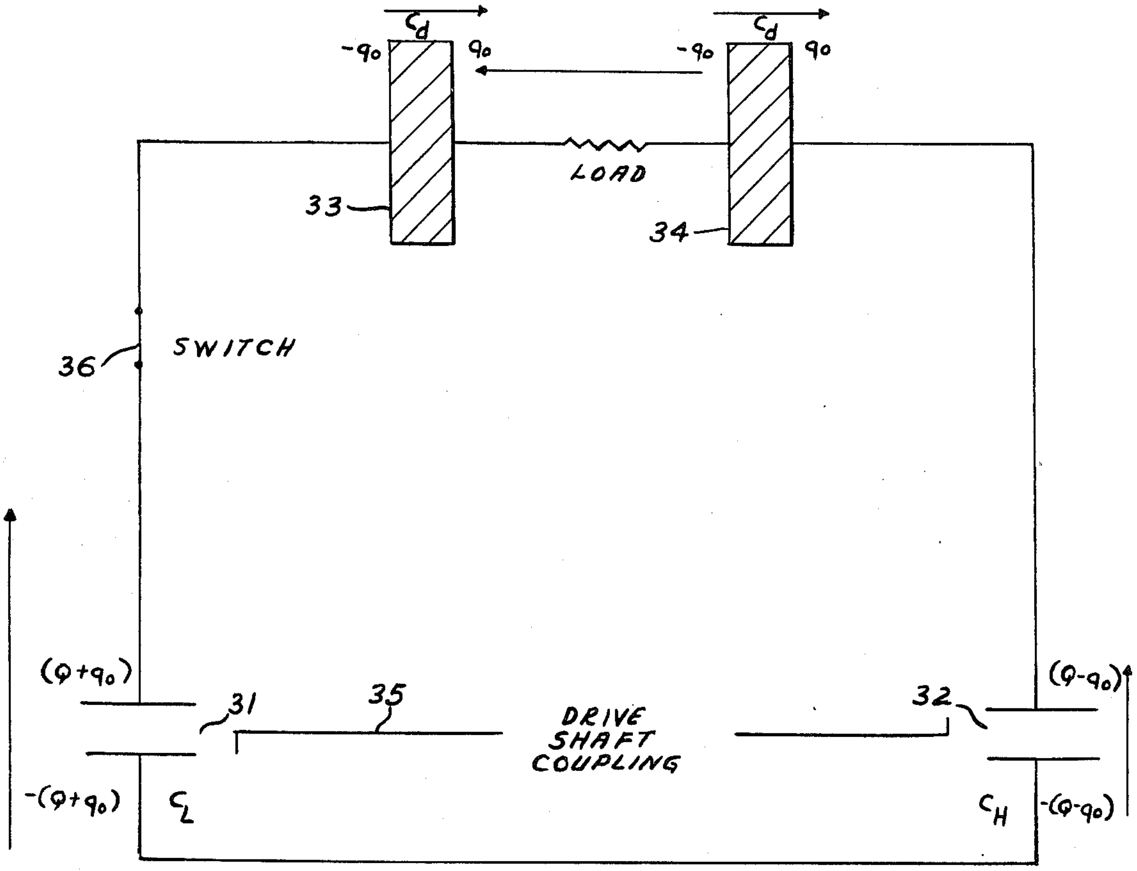
C_d
-q_o
q_o
C_d
-q_o
q_o
LOAD
33
34
36
SWITCH
$(q+q_o)$
31
35
DRIVE SHAFT COUPLING
32
$(q-q_o)$
$-(q+q_o)$
C_L
C_H
$-(q-q_o)$

Fig-14

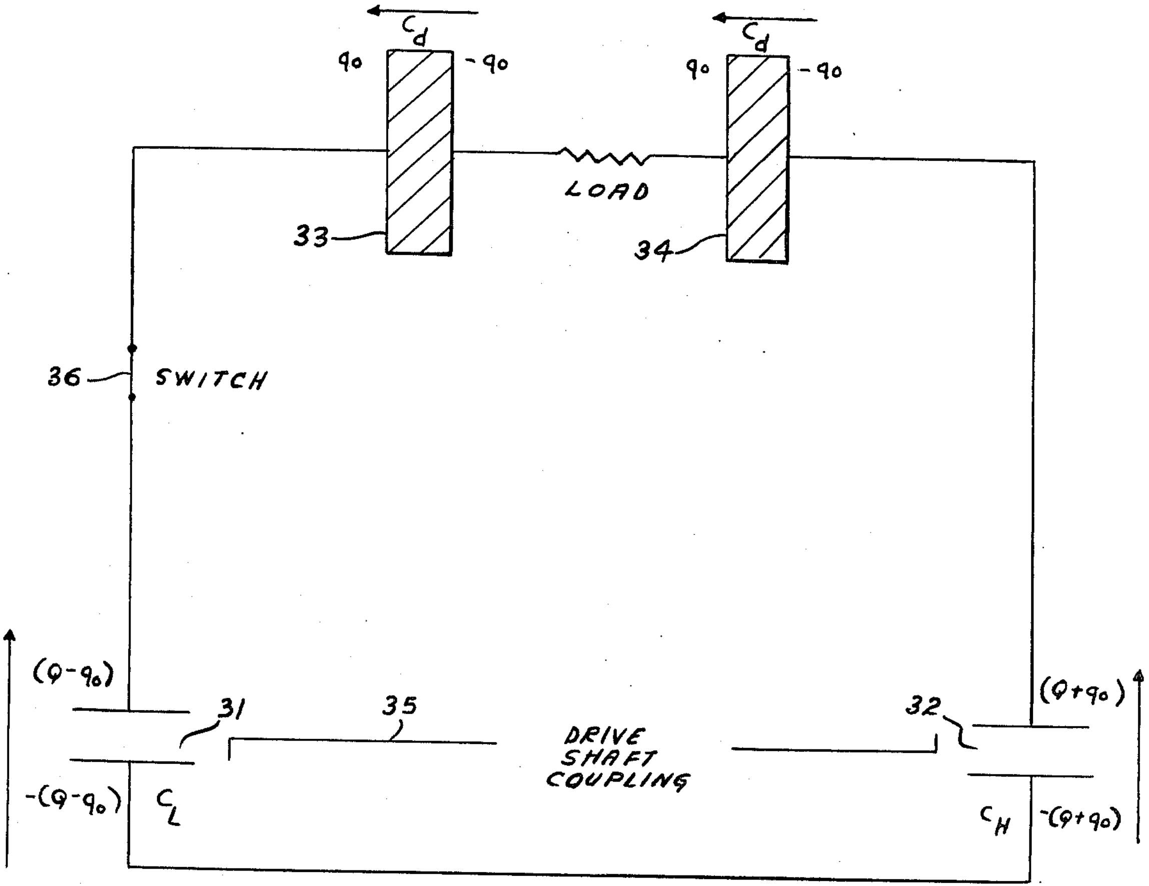
C_d
q_0
$-q_0$
C_d
q_0
$-q_0$
LOAD
33
34
36
SWITCH
$(Q-q_0)$
$-(Q-q_0)$
C_L
31
35
DRIVE SHAFT COUPLING
32
$(Q+q_0)$
$-(Q+q_0)$
C_H

Fig-15

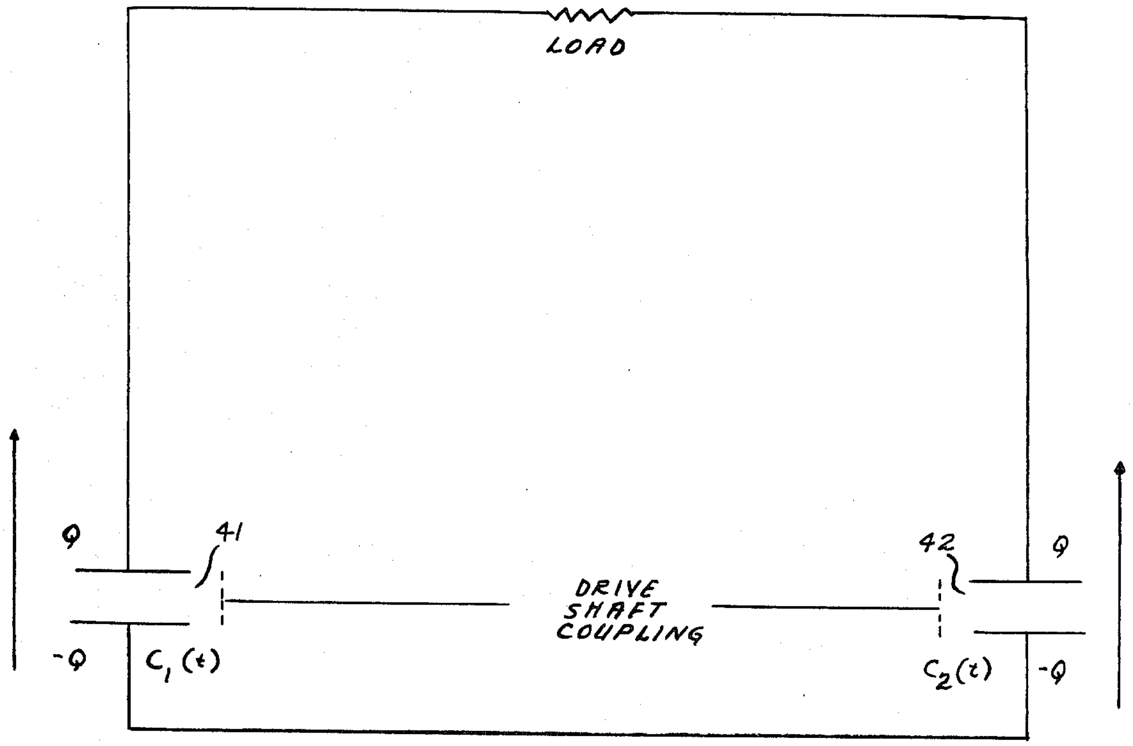
LOAD
Q
41
42
Q
DRIVE SHAFT COUPLING
-Q
C1(t)
C2(t)
-Q

Fig-16

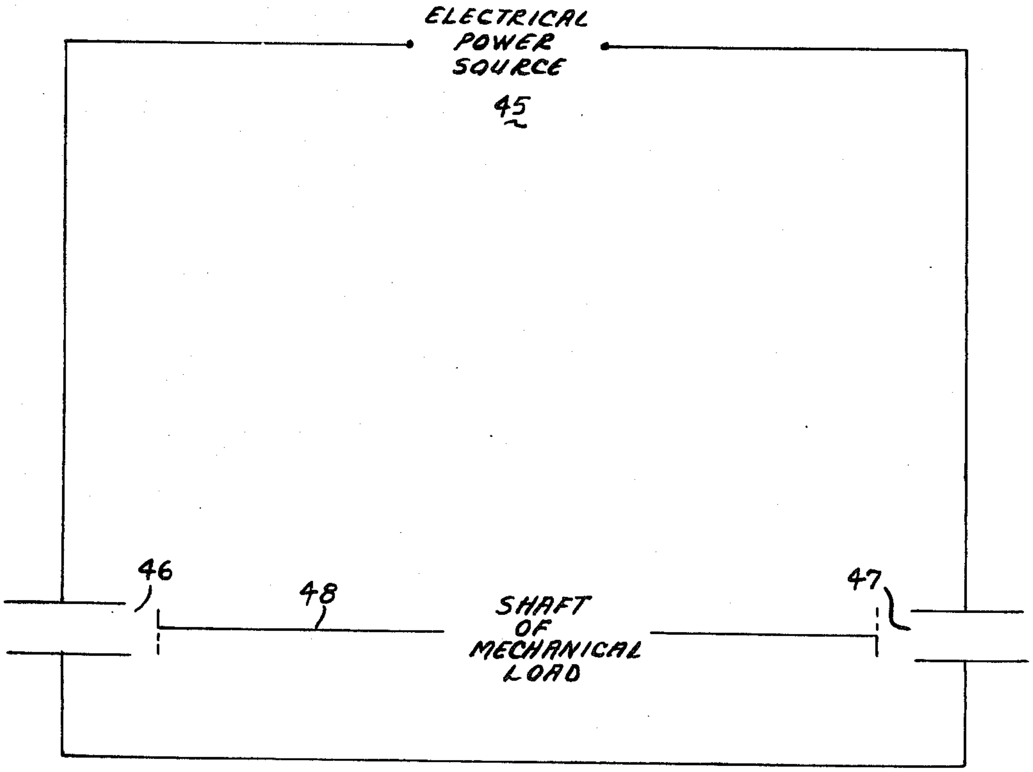
ELECTRICAL POWER SOURCE
SHAFT OF MECHANICAL LOAD

Fig 17

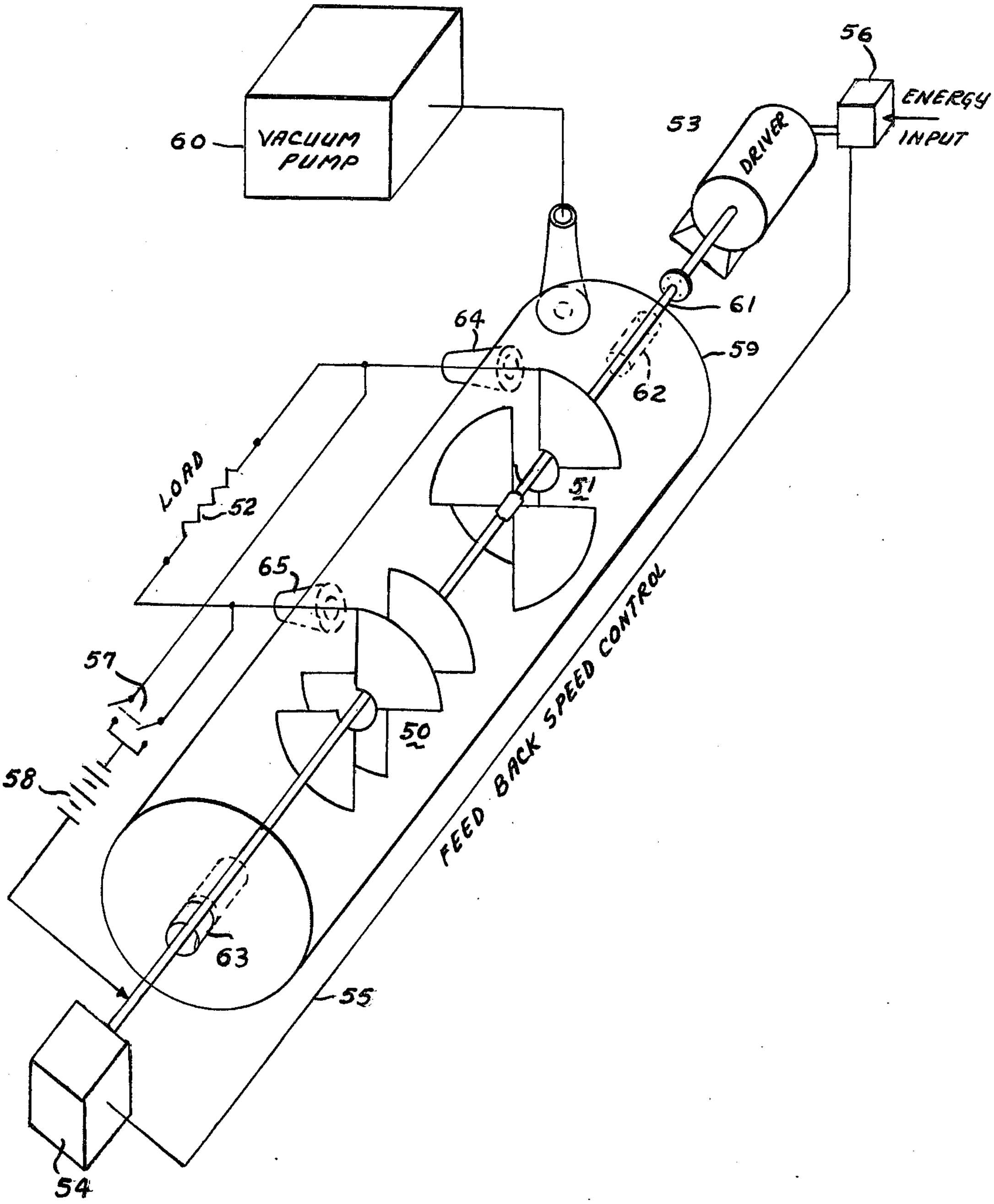
ENERGY
INPUT
DRIVER
VACUUM PUMP
LOAD
FEED BACK SPEED CONTROL

Fig-18

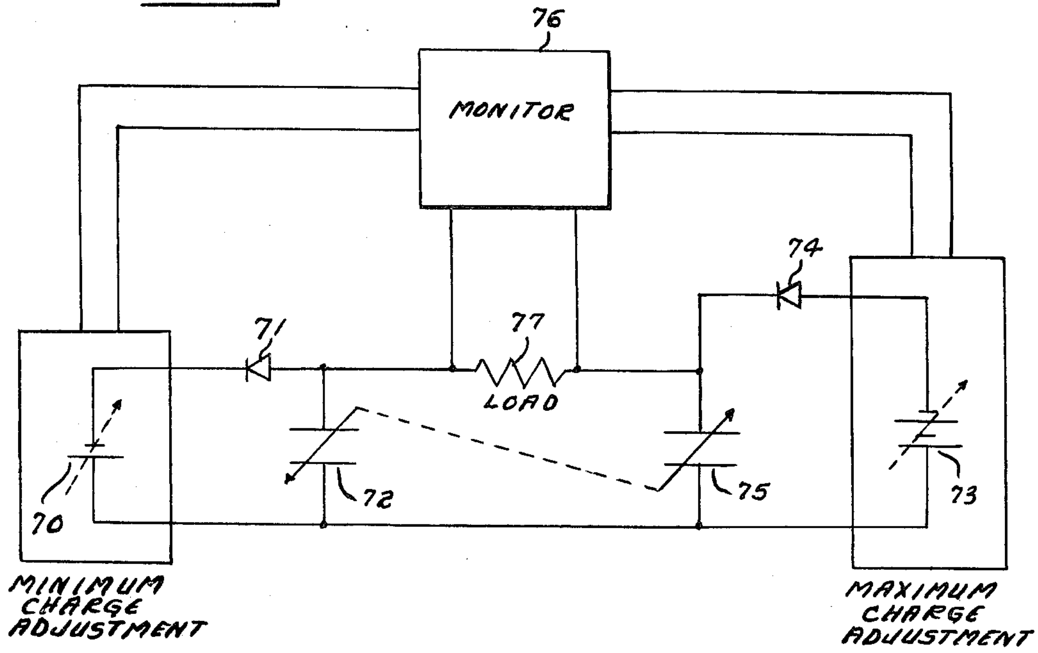
MONITOR
LOAD
MINIMUM CHARGE ADJUSTMENT
MAXIMUM CHARGE ADJUSTMENT

Fig-19

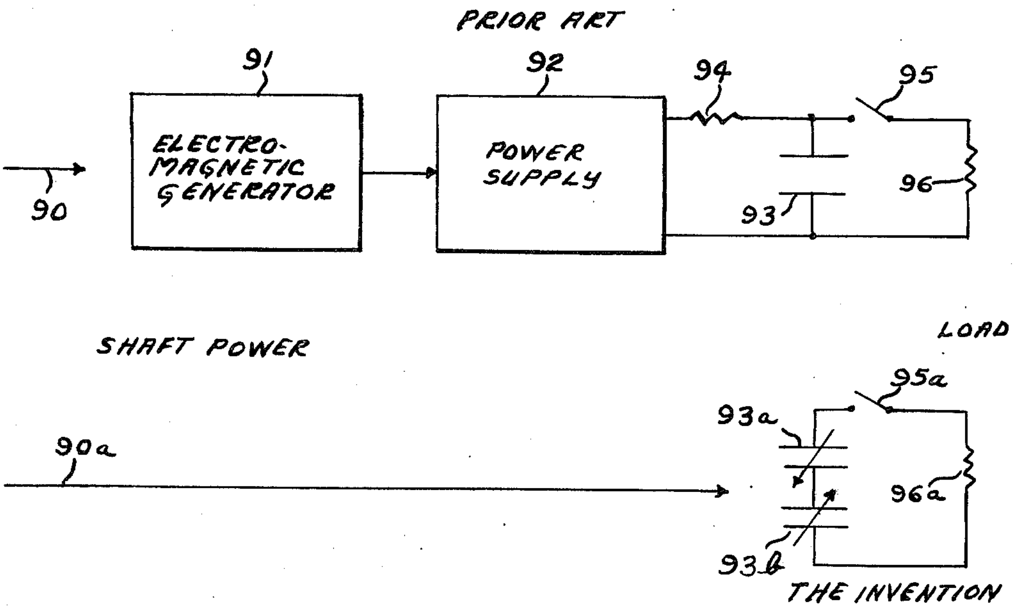
PRIOR ART
ELECTRO-MAGNETIC GENERATOR
POWER SUPPLY
SHAFT POWER
LOAD
THE INVENTION

ELECTROSTATIC ENERGY CONVERSION SYSTEM

RIGHTS OF THE GOVERNMENT

The invention described herein may be manufactured and used by or for the Government of the United States for all governmental purposes without the payment of any royalty.

BACKGROUND OF THE INVENTION

The field of the invention is in electrostatic machines.

Electrostatic machines have been well known for many years. The Toepler-Holtz and the Wimshurst machines were early electrostatic generators. Now their usage is mainly for demonstrations in the Physics Laboratory. The only remaining electrostatic machine to remain in extensive usage is the Van de Graaff type electrostatic generator. It is presently used to generate extremely high voltage potentials for atomic research. Substantially all of the prior art devices are direct current type machines.

The following publications may be helpful in further understanding the art. Electrostatic Sources of Electric Power, John G. Trump, Electrical Engineering, Vol. 66, No. 6, June 1947, pages 525-534; U.S. Pat. No. 2,810,878 to patentee N. Felici, and U.S. Pat. No. 3,094,653 to patentees D. B. LeMay et al.

SUMMARY OF THE INVENTION

An ac electrostatic variable capacity energy conversion system utilizing a fixed contained charge functions as an electrical generator when mechanically driven or as a motor providing mechanical shaft power when electrically driven.

BRIEF DESCRIPTION OF THE DRAWING

FIG. 1 is a simplified pictorial-schematic diagram representative of a rotational generator embodiment of the invention;

FIG. 2 is a simplified pictorial schematic diagram representative of a reciprocally driven generator embodiment of the invention;

FIG. 3 is an electrical schematic diagram of the embodiment illustrated in FIG. 2;

FIG. 4 is a schematic diagram illustrative of the electrical charges and capacitances of the embodiment physically illustrated in FIG. 1;

FIG. 5 is a pictorial-schematic representation of an embodiment of the invention wherein the load is between dielectric materials;

FIG. 6 is a schematic diagram of the system shown in FIG. 5 illustrative of the charges and capacitances;

FIG. 7 is a schematic diagram of a pulsed energy input embodiment of the invention in State A;

FIG. 8 is the same embodiment except in State B;

FIG. 9 is the same embodiment except in State C;

FIG. 10 is the same embodiment except in State D;

FIG. 11 is a schematic representation of an embodiment of the invention having pulsed energy input with the generators indirectly connected to the load through dielectric material, in State A;

FIG. 12 is the same embodiment of FIG. 11 except in State B;

FIG. 13 is the same embodiment of FIG. 11 except in State C;

FIG. 14 is the same embodiment of FIG. 11 except in State D;

FIG. 15 is a schematic diagram of an embodiment of the invention having continuous mechanical input providing a continuous wave (cw) electrical output with the generators directly connected to the load;

FIG. 16 is a schematic diagram of an energy conversion system of the invention wherein the system functions as a motor receiving electrical energy and delivering mechanical energy;

FIG. 17 is a pictorial-block-schematic diagram of an embodiment of the invention for receiving rotational energy from a mechanical source and delivering electrical energy to a load;

FIG. 18 is a schematic drawing shown a control system for maintaining a given maximum and minimum charge on the generating capacitors; and

FIG. 19 is a block-schematic diagram showing the improvement of the disclosed invention over a typical prior art system.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

In embodiments of the invention, energy conversion is obtained for the Variable Capacitance Electrostatic Energy Conversion System by operation of variable capacitance electrostatic generators, involving contained charge, in paired opposition to one another across the load (or electrical source of power), either directly or indirectly through dielectric material, to provide for conversion of mechanical energy to electrical energy (or electrical energy to mechanical energy).

For a variable capacitance electrostatic generator, the electrical power input to the capacitor, P_e (watts), is equal to the increase in capacitor electrostatic field energy, W_{ef} (joules), per unit time, t (seconds), and the mechanical power output, P_M (watts):

$$P_e = (dW_{ef}/dt) + P_M$$

with

$P_e = Vdq/dt$

V = capacitor potential (volts)

q = charge on capacitor (coulombs)

and $W_{ef} = \frac{1}{2} CV^2$

C = capacitor capacitance (farads)

Also, $q = CV$

Therefore, $P_M = \frac{1}{2} V^2 (dC/dt)$

and the time rate of conversion of mechanical to electrical energy, dW_{ME}/dt, equal to $-P_M$, is given as

$$dW_{ME}/dt = -\tfrac{1}{2} V^2 (dC/dt)$$

Conversion of mechanical shaftpower to electrical power to a load by variable capacitance electrostatic generators operating in paired opposition to one another across the load (and involving contained charge) shall be considered first.

Examples of two specific physical arrangements of electrostatic generators operating in paired opposition to one another connected directly to a load R (ohms) are shown in FIGS. 1 and 2. Capacitance change is by rotary motion in the embodiment of FIG. 1 and reciprocating motion for the embodiment represented by FIG. 2. FIG. 2 is shown in partial schematic form in FIG. 3 and the electrical schematic diagram for the device of FIG. 1 is shown in FIG. 4. Referring to FIGS. 1 and 4, the capacitors 11 and 12 each range in value from high capacitance C_H to low capacitance C_L and are coupled by the drive shaft 13 so that they operate in opposition to each other; that is, when one capacitor has value C_H, the other capacitor has value C_L, and vice versa. The capacitors each are initially charged to possess a charge Q, and hence there exists a contained (or "trapped") charge of 2Q for the system. As variable capacitor **11** varies from C_H to C_L, its charge varies from $(Q + q_o)$ to $(Q - q_o)$; as variable capacitor **12** varies from C_L to C_H, its charge varies from $(Q - q_o)$ to $(Q + q_o)$. A charge $2q_o$ is directly transferred across the load **14**. As capacitor **11** is decreasing in capacitance mechanical energy is converted by it to electrical energy and electrical energy is converted to mechanical energy by variable capacitor **12** which is increasing in capacitance. However, since capacitor **12** is drive-shaft coupled to capacitor **11**, this mechanical energy may be considered as simultaneously reconverted to electrical energy. With rotation, capacitor **11** reaches the state initially occupied by capacitor **12** and capacitor **12** is now in the state initially occupied by capacitor **11**. As the drive shaft continues to rotate, the capacitors return to their respective initial states with another transfer of charge $2q_o$ across the load, and a cycle is completed.

A schematic-pictorial physical arrangement of the electrostatic generators operating in paired opposition to one another, indirectly connected to the load through dielectric (insulating) material, is shown in FIG. **5**; an electrical schematic diagram for this embodiment is shown in FIG. **6**. The advantages of using a dielectric as electrode material is well known as exemplified by U.S. Pat. No. 3,745,481 to patentees G. F. Erickson et al. In this embodiment for each cycle a charge of $4q_o$ is transferred by displacement across the load; there is charge q_o associated with the dielectric material capacitors of capacitance c_d.

The simplified embodiments presented above were for initial explicit exposition of the concept involved for the Energy Conversion System. The following embodiments, presented in more detail, will explain the operation of the system and aid in the comprehension of the invention.

The pulsed energy input systems with the generators directly connected to the load are perhaps the easiest to comprehend. The electrical schematic diagram of these embodiments is shown in FIG. **7**. It is similar to the electrical schematic diagram of previous FIG. **4** except for the added switch **20**. The pictorial diagram of FIG. **1** is also a suitable physical representation if a switch is placed in the line going to the load. The operation will be explained considering four states or positions of the capacitors after a movement from their prior state. It is assumed that the capacitors have been initially charged and that operation through previous cycles has taken place. The switch is triggered cyclicly as a function of drive shaft position.

State A (FIG. **7**): Capacitor **21** has capacitance C_H and a charge $(Q + q_o)$; capacitor **22** has capacitance C_L and a charge $(Q - q_o)$; the potentials (volts) across the capacitors are equal. (NOTE: The magnitude of the potentials associated with the various circuit elements are indicated by the relative length of the arrows in the schematics as presented in the figures, e.g., arrows **23** and **24**, with the tip of the arrow indicating the direction of increasing potential for the charge; the arrows, representing the potential, are not necessarily drawn to scale.) The switch is in closed position; the switch, for the purpose of this exposition, shall be considered as an ideal switch (that is, it has negligible inductance, infinitely small capacitance relative to other circuit elements, infinite resistance when open, and zero resistance when closed). The potential across a capacitor is equal to the charge on the capacitor divided by its capacitance, therefore;

$$(Q + q_o)/C_H = (Q - q_o)/C_L$$

The energy associated with the electrostatic field of a capacitor is given by one-half the value of its charge squared, divided by its capacitance. Therefore, the total electrical energy of the electric circuit in State A, W_A, is given as

$$W_A = \tfrac{1}{2}(Q + q_o)^2/C_H + \tfrac{1}{2}(Q - q_o)^2/C_L$$

The switch is opened and, as the drive shaft rotates, the capacitances of capacitors **1** and **2** change to State B.

State B (FIG. **8**): Capacitor **21** now has capacitance C_L and capacitor **22** has capacitance C_H. The total electrical energy for the electric circuit in State B, W_B, is given as

$$W_B = \tfrac{1}{2}(Q + q_o)^2/C_L + \tfrac{1}{2}(Q - q_o)^2/C_H$$

The potential rise across capacitor **21** is equal to the potential drop across the switch **20** and across capacitor **22**.

State C (FIG. **9**): The switch **20** is now closed. The potential rise across capacitor **21** is now equal to the potential drop across the load and the potential drop across capacitor **22**. It is assumed, for the purpose of this exposition, that the transient response time of the electric circuit is much faster than the time required for one complete cycle of the variable capacitors (that is, the time required for the variable capacitors, starting at respective initial states, to return to those respective states, as the drive shaft rotates). With q now representing the charge being transferred, the equation for the electrical circuit response is given as

$$(Q + q_o - q)/C_L = R\,(dq/dt) + (Q - q_o + q)/C_H$$

When $t = 0$, $q = 0$; when $t = \infty$, $q = 2q_o$.
The solution of the previous equation is as follows:

$$q = 2q_o[1 - \exp[-(1/R)\,(1/C_L + 1/C_H)t]]$$

And from the earlier equation $(Q + q_o)/C_H = (Q - q_o)/C_L$

$$q_o = \frac{(1/C_L - 1/C_H)}{(1/C_L + 1/C_H)}\,Q$$

State D (FIG. **10**): After charge rearrangement by a flow of current through the load, capacitor **22** is in the same state as capacitor **21** in State A; capacitor **21** is in the same state as capacitor **22** in State A; and the total electrical energy for the electrical circuit in State D, W_D, is given as

$$W_D = \tfrac{1}{2}(Q + q_o)^2/C_H + \tfrac{1}{2}(Q - q_o)^2/C_L$$

By subtracting this equation from the one in State B, it is seen that:

$$W_B - W_D = 2q_o^2\,(1/C_L + 1/C_H) = 4[\tfrac{1}{2}\,q_o^2(1/C_L + 1/C_H)]$$

-continued

$$\int_{o}^{\infty} R\left(\frac{dq}{dt}\right)^2 dt = 2q_o^2(1/C_L + 1/C_H) = W_B - W_D$$

Electrostatic energy is delivered to the load. As the drive shaft **25** rotates, capacitors **21** and **22** will again go through States B and C, with their roles interchanged, until the capacitors return to their respective states as given in State A. For this second half of the cycle, the same amount of energy is delivered to the load as was delivered in the first half of the cycle; hence, the energy delivered to the load per cycle, W, is given as follows:

$$W = 2(W_B - W_D) = 8[\tfrac{1}{2}\, q_o{}^2(1/C_L + 1/C_H)]$$

The limit on energy transfer may be determined as follows. The breakdown potential for the variable capacitors, V_{CB} (volts), may be related to the charge on the variable capacitors:

$$(Q + q_o)/C_L = V_{CB}$$

By use of this equation and the one given at the end of State C, the limiting value of q_o may be given as follows:

$$q_o = (C_L/2)\,(1 - C_L/C_H)V_{CB}$$

By substitution of the above value of q_o into the foregoing equation for W, the limit on W, W_{CB}, is given as

$$W_{CB} = (1 + C_L/C_H)\,(1 - C_L/C_H)^2 C_L V_{CB}{}^2$$

The range of W is then given as follows: $0 \leqq W < W_{CB}$

For non-negligible circuit inductance, such as, for example, a series inductance L (henries), the equation for the electrical circuit response is given as

$$(Q + q_o - q)/C_L = R(dq/dt) + L(d^2q/dt^2) + (Q - q_o + q)/C_H$$

When $t = 0$, $q = 0$; when $t = \infty$, $q = 2q_o$; the solutions are given as follows.

For over damped response, $(R/2L)^2 > (1/L)\,(1/C_L + 1/C_H)$,

$$q = 2Q\left[\frac{C_H - C_L}{C_H + C_L}\right]\left[1 - \exp\left(-\frac{Rt}{2L}\right)\left[\frac{R}{2La}\sinh(at) + \cosh(at)\right]\right]$$

With

$$a = \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{L}\left(\frac{1}{C_L} + \frac{1}{C_H}\right)}$$

For critically damped response, $(R/2L)^2 = (1/L)\,(1/C_L + 1/C_H)$,

$$q = 2Q\left[\frac{C_H - C_L}{C_H + C_L}\right]\left[1 - \exp\left(-\frac{Rt}{2L}\right) - \frac{Rt}{2L}\exp\left(-\frac{Rt}{2L}\right)\right]$$

For under damped response, $(R/2L)^2 < (1/L)\,(1/C_L + 1/C_H)$,

$$q = 2Q\left[\frac{C_H - C_L}{C_H + C_L}\right]\left[1 - \exp\left(-\frac{Rt}{2L}\right)\left[\frac{R}{2Lb}\sin(bt) + \cos(bt)\right]\right]$$

With b being given by a, as defined above, divided by the square root of minus one. The relation between Q and q_o is

$$q_o = \frac{(1/C_L - 1/C_H)}{(1/C_L + 1/C_H)}\,Q.$$

as previously stated.

An embodiment of the invention for operation with pulsed energy input and the electrical load connected through dielectric material, such as barium titanate, is schematically illustrated in FIG. **11**. This type of system is ideally suited for the production of molecular excitation and volume ionization of high-pressure lasing media. As in the previous embodiment, four states or end positions of arcs of rotational movement will be illustrated.

State A (FIG. **11**): Capacitor **31** has capacitance C_H and a charge $(Q + q_o)$; capacitor **32** has capacitance C_L and a charge $(Q - q_o)$. The potential rise across capacitor **32** is equal to the sum of the potential drops across each dielectric electrode (**33** and **34**, each of capacitance c_d) and capacitor **31**:

$$(Q - q_o)/C_L = 2q_o/c_d + (Q + q_o)/C_H$$

The total electrical energy of the electric circuit in State A is given as follows:

$$W_A = \tfrac{1}{2}\,(Q + q_o)^2/C_H + q_o{}^2/c_d + \tfrac{1}{2}\,(Q - q_o)^2/C_L$$

Switch **36** is opened and, as drive shaft **35** rotates, the capacitance of capacitors **31** and **32** changes, and state B is reached.

State B (FIG. **12**): Capacitor **31** now has capacitance C_L and capacitor **32** has capacitance C_H. The total electrical energy for the electric circuit in State B is given as follows:

$$W_B = \tfrac{1}{2}\,(Q + q_o)^2/C_L + q_o{}^2/c_d + \tfrac{1}{2}\,(Q - q_o)^2/C_H$$

The potential rise across capacitor **31** and the two capacitances c_d equals the potential drop across the switch and capacitor **32**.

State C (FIG. **13**): Switch **36** is now closed. The potential rise across capacitor **31** and the two capacitances c_d equals the potential drop across the load and capacitor **32**. The equation for the electrical circuit response is given as follows:

$$(Q + q_o - q)/C_L + 2(q_o - q)/c_d = R(dq/dt) + (Q - q_o + q)/C_H$$

When $t = 0$, $q = 0$; when $t = \infty$, $q = 2q_o$.

Solution of the equation is as follows:

$$q = 2q_o[1 - \exp[-(1/R)\,(1/C_L + 2/c_d + 1/C_H)t]]$$

With, as given previously,

$$q_o = \frac{(1/C_L - 1/C_H)Q}{(1/C_L + 2/c_d + 1/C_H)}$$

State D (FIG. **14**): After charge rearrangement, capacitor **32** is in the same state as capacitor **31** in State A; capacitor **31** is in the same state as capacitor **32** in State A; and the polarity of the charge on the dielectric capacitances has been reversed. The total electrical energy for the electric circuit in State D is given as follows:

$$W_D = \tfrac{1}{2}(Q + q_o)^2/C_H + q_o^2/c_d + \tfrac{1}{2}(Q - q_o)^2/C_L$$

$$(Q - q)/C_1 = R(dq/dt) + (Q + q)/C_2$$

with $q = 0$, for $t = 0$.

With sinusoidal input to the load, for the following assumed capacitance:

$$C_1 = \frac{1}{\frac{1}{C_H} + (\frac{1}{C_L} - \frac{1}{C_H})[\frac{1 - \cos\frac{2\pi}{T}(t + \frac{T}{4})}{2}]}$$

$$C_2 = \frac{1}{\frac{1}{C_H} + (\frac{1}{C_L} - \frac{1}{C_H})[\frac{1 + \cos\frac{2\pi}{T}(t + \frac{T}{4})}{2}]}$$

$$q = \frac{\frac{Q}{R}[\frac{1}{C_L} - \frac{1}{C_H}]}{\sqrt{(\frac{2\pi}{T})^2 + [\frac{1}{R}(\frac{1}{C_L} + \frac{1}{C_H})]^2}} X$$

$$\left[\begin{array}{l} \cos[\frac{\pi}{2} - \arctan([\frac{2\pi}{T}]/[\frac{1}{R}(\frac{1}{C_L} + \frac{1}{C_H})])]\exp[-\frac{1}{R}(\frac{1}{C_L} + \frac{1}{C_H})t] \\ -\cos[\frac{2\pi}{T}(t + \frac{T}{4}) - \arctan([\frac{2\pi}{T}]/[\frac{1}{R}(\frac{1}{C_L} + \frac{1}{C_H})])] \end{array}\right]$$

By combining equations, it is seen that:

$$W_B - W_D = 2q_o^2(1/C_L + 2/c_d + 1/C_H) = 4[\tfrac{1}{2} q_o^2 (1/C_L + 2/c_d + 1/C_H)]$$

And that:

$$\int_0^\infty R(\frac{dq}{dt})^2 dt = 2q_o^2(1/C_L + 2/c_d + 1/C_H) = W_B - W_D$$

Electrostatic energy is delivered to the load. As drive shaft **35** rotates, capacitors **31** and **32** will again go through States B and C, with their roles interchanged, until the capacitors return to their respective states as given in State A. For this second half of the cycle, the same amount of energy is delivered to the load as was delivered in the first half of the cycle; hence, the energy delivered to the load per cycle, W, is given as follows:

$$W = 2(W_B - W_D) = 8[\tfrac{1}{2} q_o^2](1/C_L + 2/c_d + 1/C_H)$$

Assuming q_o limited by the breakdown potential of the dielectric material, V_{dB}, $q_o = c_d V_{dB}$. Then, the limit on W, W_{dB}, is as follows:

$$W_{dB} = 4(c_d/C_L + 2 + c_d/C_H)c_d V_{dB}^2$$

The range of W is then given as follows: $0 \leqq W < W_{dB}$

FIG. **15** schematically illustrates an embodiment of the invention having continuous mechanical energy input and providing a continuous wave output to the loads. It is similar to previous FIG. **4** except for designations of charge and capacitance.

Consider the system schematic as shown in FIG. **15**; and, at time $t = 0$, the variable capacitors **41** and **42** have the same capacitance, the same charge Q, and hence are at the same potential. Capacitor **42** is out of phase with capacitor **41** by half a period; the capacitors vary periodically with period T (seconds).

The circuit equation for the transported charge q is as follows: assuming negligible circuit inductance:

with average power, P (watts), given, for the electrical circuit time constant,

$$1/[\frac{1}{R}(\frac{1}{C_L} + \frac{1}{C_H})],$$

much less than T, and after transient decay, as follows:

$$P = \frac{1}{T}\int_t^{t+T} R(\frac{dq}{dt})^2 dt = \frac{2\pi^2 R Q^2 (C_H - C_L)^2}{T^2 (C_H + C_L)^2}$$

and with Q limited by the breakdown potential V_{CB} of the capacitors as follows:

$$[\frac{(Q - q)}{C_1}]_{maximum} = \frac{Q(C_H + C_L)}{2C_H C_L} = V_{CB}$$

By use of these last two equations, the limit on P, P_{CB}, is:

$$P_{CB} = \frac{8\pi^2 C_H C_L (C_H - C_L)^2 V_{CB}^2}{T^2 (C_H + C_L)^3 (\frac{1}{R})(\frac{1}{C_L} + \frac{1}{C_H})}$$

The range of P is then given as follows: $0 \leqq P < P_{CB}$

As with the pulsed embodiments the load on the *cw* embodiments may be an electrostatic field between dielectric electrodes. It is also to be understood that the electrical outputs of generating embodiments of the invention, while generally of quite high electrical potential, may be rectified (and filtered), or run through a transformer effecting an impedance change, or any of the other conventional electrical changes and usages that are common practice with continuous wave or pulsed alternating current voltages.

It is also to be understood that for simplicity, only one movable and one fixed plate has been shown for a variable capacitor, and that embodiments may be constructed that have capacitors with many fixed and movable plates. A plurality of embodiments may be connected in series for providing higher output voltages and better voltage breakdown characteristics. The polarity of the charge placed on the capacitors, i.e., the common polarity of all the fixed plates of an embodiment with respect to the movable plates is immaterial.

The electrostatic energy conversion system of the invention may also be operated as a motor. FIG. **16** is a simplified schematic representation of an embodiment of the invention for receiving electrical energy from a power source **45** and providing torque to rotate shaft **48**. Variable capacitors **46** and **47** are similar to those in the generating device, with the capacitance of capacitor **46** being a maximum when the capacitance of capacitor **47** is a minimum. Shaft **48** is driven by the forces of the capacitor plates to rotate in synchronism with the frequency of source **45**. Thus, unless the frequency of source **45** is relatively slowly increased from dc to the desired operating frequency, other conventional means such as are normally used to get synchronous motors up to synchronous speed may be used to bring shaft **48** up to substantially synchronous speed with the line frequency of the source of electrical power before connection is made.

It is to be understood that the devices of the invention, generally like other electrostatic devices from a practical consideration, generally function with relatively high voltages compared to conventional electromagnetic equipment. As pointed out by Trump in the previously referenced AIEE (now IEEE) paper, two metallic plates of 100 square inches area which are parallel and insulated from each other, with an electric field of 300 volts per centimeter between them have an electrostatic force of attraction of about 0.0005 pound. If the active gradient is increased to 30,000 volts per centimeter, which is the highest voltage that can be insulated in atmospheric air, the electrostatic force between the plates is increased to little more than 0.5 pound. When the plates are immersed in a high vacuum and the gradient increased to 3 million volts per centimeter, the electrostatic force then becomes approximately 5,700 pounds.

From the foregoing, it is readily apparent that for embodiments of the invention to be practical, the variable capacitors should be operated in a high vacuum. For embodiments used in outer space, this is generally not a problem, but for terrestial embodiments, it is generally preferred that a sealed chamber having a high vacuum be used. Insulative gases such as sulphur hexafluoride or air at high pressures may be used instead of the high vacuum to provide the necessary insulation at the high voltages required for practical operation. Another reason for generally preferring the high vacuum environment for embodiments of the invention is the elimination of drag on the moving members. A complete system for the cw generation of electrical power under normal terrestial conditions is shown schematically in FIG. **17**. Variable capacitors **50** and **51**, with rotation, pass a contained charge back and forth between them causing a current to flow through load **52** delivering power to it. Driver **53** may be any suitable device that can supply the torque necessary to rotate the movable plates of the capacitors such as an electric motor, a turbine, or even a hand crank. Generally, it is desirable to provide conventional speed sensor **54** with conventional feedback control **55**, actuating energy input controller **56**, so that the frequency of the generated electrical energy supplied to load **52** is relatively constant. An initial charge is placed on variable capacitors **50** and **51** by momentarily closing switch **57** causing conventional high potential source **58** to place an equal potential charge of the same polarity on each capacitor. It is generally desirable to position the movable plates of the capacitors such that they have equal capacity when initially charged so that the initial charge magnitudes will also be equal. The capacitors are operated in a high vacuum in conventional sealed chamber **59**. Conventional vacuum pump **60** provides the high vacuum in the chamber (approximately 10^{-8} Torr). Mechanical drive shaft **61** extends through the chamber walls in magnetically sealed rotary motion feedthroughs **62** and **63**. These conventional, commercially available, sealed bearings having ferrofluid seals provide substantially zero leakage (at speeds up to 10,000 rpm) to a high vacuum on one side of the bearing seal and atmospheric pressure on the other side of the bearing. Conventional capacitor plate material such as stainless steel is suitable and generally preferable. Conventional high voltage feedthroughs **64** and **65** provide for transfer of charge from the stator capacitor plates to the load.

As previously stated, once the charge is placed on the capacitors it is not diminished or dissipated in supplying power to the load. It is a contained or trapped charge, and is not run down by the load. If the insulation within the chamber were perfect, the charge would never have to be replenished. Since nothing is perfect the charge will, over a long period of time, decrease. It may be replenished by momentarily actuating switch **57** as in the initial charging of the capacitors or the charge may be automatically held within predetermined maximum and minimum values on the capacitor plates by an electrical circuit as shown schematically in FIG. **18**. For simplicity of explanation, maximum voltage monitoring is shown on one side of the capacitor pair and minimum voltage monitoring on the other. It is to be understood that both a maximum and a minimum monitoring may be on the same side, or simultaneously on both sides, if desired. Potential source **70** and diode **71** add charge when the voltage potential of capacitor **72** is below that of source **70** when capacitor **72** is at maximum capacitance. Potential source **73** in cooperation with diode **74** removes charge when the voltage of capacitor **75** (when going through its minimum capacitance) is greater than the potential of **73**. By having potential sources **70** and **73** adjustable and controllable by conventional voltage monitor **76**, the correct value of trapped charge for the system is maintained for desired voltage (or power) for load **77**. The regulating system shown in FIG. **18** may also be used to compensate for leakage losses.

FIG. **19** illustrates in block schematic form the advantages of the disclosed invention in supplying a high potential to a load such as a laser. In the conventional prior art systems, shaft power **90** turns conventional electrical generator **91** generating relatively low voltage, high current energy, to operate conventional high voltage power supply **92** which changes capacitor **93** through resistor **94**. Capacitor **93** discharges when switch **95** is closed to supply the load **96** with energy. In the present invention, shaft power **90*a*** directly actuates inversely ganged capacitors **93*a*** and **93*b*** providing an electrostatic energy charge equivalent to that placed on capacitor **93** by power supply **92**. Thus, the closing of switch **95*a*** provides the same energy to the identical load **96*a*** as in the prior art system. In the illustration bipolar pulses flow to load **96*a***. If unipolar pulses are desired for the load **96*a*** as for load **96** in the prior art circuit, conventional rectification of the bipolar pulse may be used.

I claim:

1. An alternating current electrostatic energy converter comprising:

a. a first and a second variable capacitor, each having a maximum and a minimum capacitance, connected in paired opposition and inversely ganged such that one capacitor passes through a maximum capacitance as the other capacitor is going through a minimum capacitance;

b. means for placing an initial and contained electrical charge on the said first and second capacitors;

c. means for actuating the said capacitors through their maximum and minimum capacitance passing contained charge between them; and

d. means for extracting power from the said capacitors.

2. An electrostatic alternating current generating system comprising:

a. a first and a second variable capacitor, each having a maximum and a minimum capacitance, inversely ganged on a common shaft such that one capacitor passes through a maximum capacitance as the other capacitor is going through a minimum capacitance;

b. means for placing an initial and contained electrical charge on the said first and second capacitors;

c. means for actuating the said shaft moving capacitors through their maximum and minimum capacitance;

d. an electrical load; and

e. means for connecting the said first and second variable capacitors in paired opposition to one another across the said load whereby contained charge is passed back and forth between the said capacitors through the said load.

3. An electrostatic alternating current generator comprising:

a. a first variable capacitor having a fixed plate and a movable plate providing a maximum and a minimum capacitance between the said fixed plate and the said movable plate;

b. a second variable capacitor having a fixed plate and a movable plate providing a maximum and a minimum capacitance between the said fixed plate and the said movable plate;

c. means for mechanically coupling the said movable plate of the said first capacitor to the movable plate of the said second capacitor such that the capacity of one variable capacitor is at a maximum capacitance when the other variable capacitor is at a minimum;

d. means for electrically connecting the said movable plates;

e. means for placing a contained electrical charge on the said variable capacitors such that the said fixed plates have like polarity charge;

f. means for actuating the said variable capacitors through a change of capacitance; and

g. means for connecting an electrical load between the said fixed plates of the variable capacitors, whereby with change of capacitance of the said capacitors, contained charge is passed between the said capacitors through the said load.

4. The apparatus as claimed in claim 3 wherein the said means for placing a charge on the said variable capacitors includes means for placing equal initial charges on the said variable capacitors.

5. The apparatus as claimed in claim 4 wherein a means for providing an evacuated enclosure and a means for positioning the said first and second variable capacitors within the said evacuated enclosure are provided.

6. The apparatus as claimed in claim 5 wherein the said means for actuating the said variable capacitors is rotational.

7. The apparatus as claimed in claim 5 wherein the said means for actuating the said variable capacitors is reciprocating.

8. A synchronous alternating current electrostatic motor comprising:

a. a first and a second variable capacitor each having a maximum and a minimum capacitance, connected in paired opposition inversely ganged on a common shaft such that one capacitor passes through a maximum capacitance as the other capacitor is going through a minimum capacitance;

b. means for placing an initial, contained, electrical charge on the said first and second capacitors;

c. means for placing an alternating potential voltage on the said first and second capacitors; and

d. means for connecting a mechanical load to the said shaft.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 4,127,804
DATED : Nov. 28, 1973
INVENTOR(S) : Onezime P. Breaux

It is certified that error appears in the above-identified patent and that said Letters Patent are hereby corrected as shown below:

On the front page delete:

--/73/ Assignee: The United States of America as represented by the Secretary of the Air Force, Washington, D.C.

Signed and Sealed this

Twenty-fourth Day of *April 1979*

[SEAL]

Attest:

RUTH C. MASON
Attesting Officer

DONALD W. BANNER
Commissioner of Patents and Trademarks

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 4,127,804
DATED : Nov. 28, 1973
INVENTOR(S) : Onezime P. Breaux

It is certified that error appears in the above-identified patent and that said Letters Patent are hereby corrected as shown below:

On the front page delete:

--/73/ Assignee: The United States of America as represented by the Secretary of the Air Force, Washington, D.C.

Signed and Sealed this

Twenty-fourth **Day of** *April 1979*

[SEAL]

Attest:

RUTH C. MASON
Attesting Officer

DONALD W. BANNER
Commissioner of Patents and Trademarks

United States Patent [19]

Paszyc et al.

[11] **4,223,241**

[45] **Sep. 16, 1980**

[54] **ELECTROSTATIC CHARGE GENERATOR**

[75] Inventors: **Aleksy J. Paszyc; Kwang T. Huang,** both of Ventura; **Dallas M. Shiroma,** Oxnard, all of Calif.

[73] Assignee: **The United States of America as represented by the Secretary of the** Navy, Washington, D.C.

[21] Appl. No.: **937,646**

[22] Filed: **Aug. 28, 1978**

[51] **Int. Cl.³** **H02N 1/10**

[52] **U.S. Cl.** .... **310/308;** 310/309; 322/2 A

[58] **Field of Search** .... 310/300, 308, 309; 322/2 R, 2 A

[56] **References Cited**

U.S. PATENT DOCUMENTS

2,004,352	6/1935	Simon	310/308
4,123,697	10/1978	Paszyc et al.	322/2 A

Primary Examiner—Donovan F. Duggan

Attorney, Agent, or Firm—Richard S. Sciascia; Joseph M. St. Amand

[57] **ABSTRACT**

An electrostatic charge generator for continually providing charge carriers of a given polarity in a flow of dielectric fluid. The charge generator comprises a coaxial tube with an inner conductor wire coaxial with a hollow, cylindrical outer conductor. A D.C. potential is applied between the inner and outer conductors of the coaxial tube and a dielectric liquid, such as oil, is caused to flow therethrough. Since the intensity of the non-uniform electric field in the conductor tube is highest near the inner conductor, the electrostatic charges of the opposite polarity as the inner conductor are removed from the dielectric fluid faster than charges of the same polarity are removed at the outer conductor. This results in the dielectric fluid exiting the conductor tube carrying a charge of the remaining polarity.

10 Claims, 3 Drawing Figures

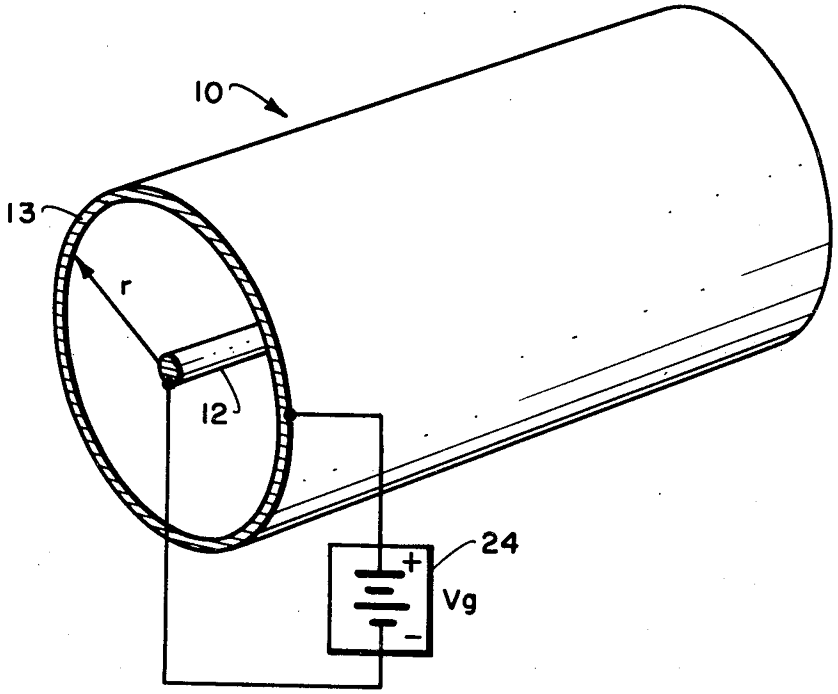

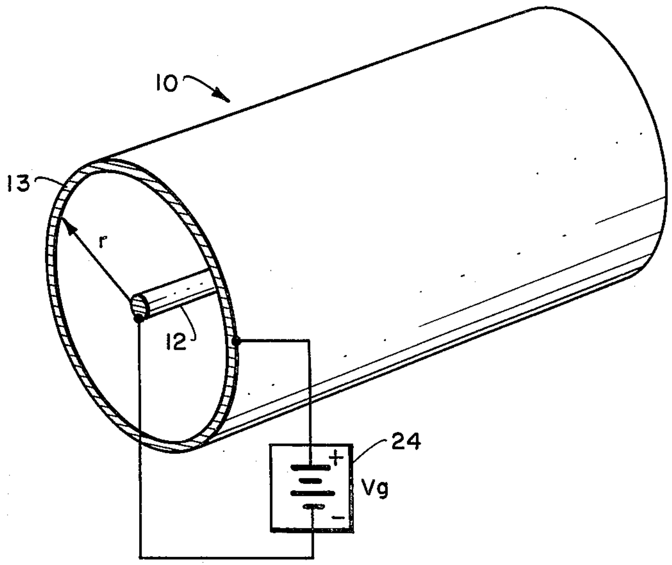
10
13
r
12
24
+
Vg
−

Fig. 1.

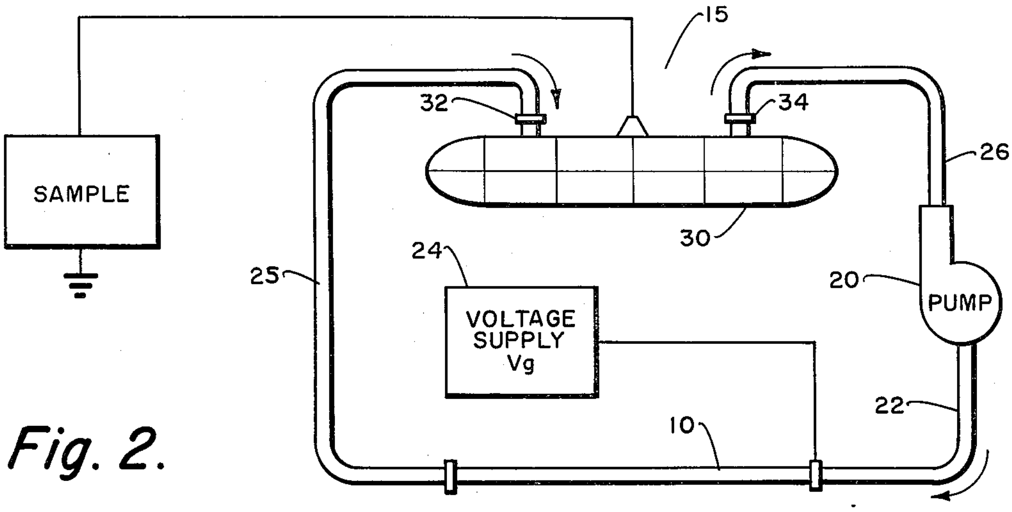
15
32
34
26
SAMPLE
30
24
VOLTAGE SUPPLY Vg
20
PUMP
25
22
10

Fig. 2.

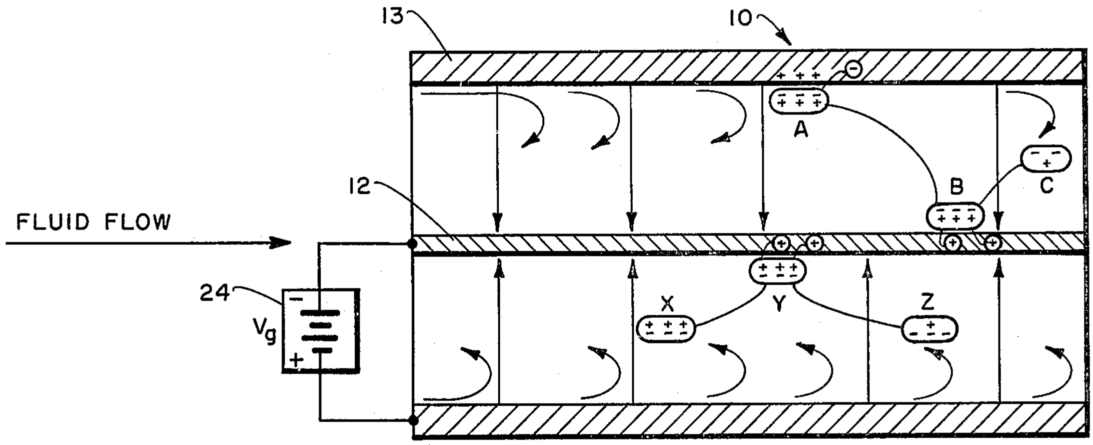
13
10
A
B
C
12
FLUID FLOW
24
Vg
−
+
X
Y
Z

Fig. 3.

ELECTROSTATIC CHARGE GENERATOR

CROSS REFERENCE TO A RELATED APPLICATION

This application is related to patent application Ser. No. 812,883 filed July 5, 1977, now U.S. Pat. No. 4,123,697.

BACKGROUND OF THE INVENTION

The present invention relates to the field of electrostatic charge generators and more particularly to electrostatic charge generators for providing electrostatic charge carriers of a given polarity in a dielectric fluid. In even greater particularity, the present invention relates to such electrostatic charge generators for use in electrostatic high potential devices.

High potential devices capable of providing voltages in the tens-of-kilovolts range are widely used in the testing of insulation for electric machinery and cables. Standard high potential tests currently used include the AC high potential test and the DC high potential test.

AC testing is essentially a go/no-go test. The voltage is raised to a specified value and held for a specified length of time. The insulation is considered good if it can withstand the voltage without actually flashing over or breaking down or if the leakage current is not excessive. The current supply requirements in AC testing are quite high, and AC test sets tend to be very large. Since equivalent DC test sets are significantly smaller, DC high potential tests are widely employed to test insulation operation in an alternating current environment.

A standard DC high potential test is the leakage current versus voltage test. In this test, a voltage is applied across the insulation and raised in incremental steps up to and somewhat above the operating voltage. The steady state value of leakage current (after high initial capacitance and absorption currents) is noted at each step of voltage; as the test progresses, a curve of leakage current versus test voltage is determined. As long as this curve is relatively flat (equal increments of voltage giving equal increments of current), the item under test is considered to be in good condition. If the voltage is raised high enough, at some point, the current will start rising at a more rapid rate, appearing as a "knee" on the leakage current versus voltage curve. If the test is carried on beyond this point, the current increases at a much more rapid rate and breakdown soon occurs. Generally, the test procedure is to stop the voltage increments at this point, extrapolating the curve to estimate the breakdown voltage.

The DC high potential test, while versatile and widely used, has a number of disadvantages. The test set is still bulky and relatively immobile. In addition, for some types of insulation the knee occurs so close to the breakdown voltage that breakdown cannot be avoided. In this case, the DC high potential test exhibits the go/no-go characteristic of AC high potential testing which is undesirable in certain situations. This breakdown problem is a result of the current levels in present DC high potential test procedures, which, although in the milliamp range, result in large power dissipation requirements at tens-of-kilovolts.

The above limitations attendant to present DC high potential testing devices evidences a need for a high

potential test device capable of providing the required voltage level at significantly lower current levels.

One such high potential test device is an electrostatic high potential system as disclosed in our copending patent application Ser. No. 812,883 filed July 5, 1977, now U.S. Pat. No. 4,123,697. The system includes a dielectric fluid circulating through an electrostatic charge generator comprised of a dome shaped ionizing chamber enclosing a series of spiral chambers and baffles. This electrostatic charge generator causes a circulating dielectric liquid (oil) to carry an electrostatic charge of a given polarity, making use of the excellent electrostatic charge carrying (insulating) properties of these fluids. These charges are removed from the dielectric fluid in an electrostatic charge collection apparatus which steadily accumulates electrostatic charge to obtain the extremely high (tens of kilovolts) potentials associated with DC high potential test equipment. The disclosed electrostatic high potential system has the capacity to generate these high potentials with a relatively small amount of charge (corresponding to the microamp range).

SUMMARY OF THE INVENTION

In the above context, the present invention is directed to an improved electrostatic charge generator capable of generating electrostatic charge carriers of a given one of either polarity in a circulating dielectric fluid.

Accordingly, the present invention provides an electrostatic charge generator which utilizes the phenomenon of electrostatic charge generation in a dielectric fluid flow. The electrostatic charge generator comprises a coaxial tube with an inner length of conductor wire coaxial with a hollow cylindrical outer conductor. A voltage supply provides a potential difference between the inner and outer conductor of the charge generator, and a dielectric fluid is circulated therethrough.

Since the intensity of the non-uniform electric field in the coaxial charge generator is highest near the inner conductor, electrostatic charges of the opposite polarity of this inner conductor are removed from the dielectric fluid more rapidly than the complementary charges are removed at the outer conductor. Thus, the dielectric fluid exiting the charge generator carries an amount of charge of the remaining polarity.

The novel features believed to be characteristic of the present invention are set forth in particularity in the appended claims. However, a more complete appreciation of the present invention, both as to organization and mode of operation, and many of the attendant advantages thereof will result as the same becomes better understood by reference to the following detailed description when considered in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a cross-section, isometric view of a portion of the coaxial charge generator according to the present invention showing the inner and outer conductors.

FIG. 2 is a schematic diagram showing an electrostatic high potential device incorporating the electrostatic charge generator.

FIG. 3 is a longitudinal cross-section of the coaxial charge generator illustrating the charge generation mechanisms which provide the electrostatic charge carriers.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Electrostatic charges are generated in a dielectric fluid whenever the fluid comes in contact with another material. As used in the discussion herein, dielectric fluid refers to a liquid such as oil or a mixture of oil and water as hereinafter described. The generation and accumulation of electrostatic charges are dependent upon certain properties of the dielectric fluid, such as electrical conductivity and electrostatic charging tendency, as well as system design and fluid flow conditions. With the improved filtering and refining today, present dielectric fluids exhibit greater electrical resistance (lower conductivity) inhibiting charge recombination and thus allowing greater charge accumulation. This is to be contrasted with fluids such as water which generate considerable electrostatic charges which, due to high conductivity, recombine very quickly, precluding the charge accumulation.

The principal mechanism responsible for the electrostatic charge generation phenomenon in a dielectric fluid is the electric "double layer" at the boundary between a solid and a liquid or between two different liquids. (It is a misconception that the charges in a dielectric fluid are generated by the friction between the moving fluid and a conduit wall.) In the hydrocarbon fluid (oil), some impurities or foreign matters are dissolved as positive and negative charged ions. As a result of the absorption of ions of one polarity on the surface of the conduit carrying the flow, the dielectric fluid generally carries ions with the opposite charge to those on the conduit surface. An electric double layer is thus formed at the interface between the fluid and the solid conduit surface. The electrostatic charges remaining in the liquid are carried by the flow of the dielectric fluid resulting in a flow of electricity known as the streaming current. Continuous flow of the dielectric fluid can accumulate enough electrostatic charge to provide high potential levels (voltages in excess of 24 Kv have been observed in fueling operations).

The electrostatic charging tendency of a fluid varies from fluid to fluid but is greatly affected by the presence of impurities. For instance, it has been noted that a two-phase mixture (such as oil and water) is a very good electrostatic charge generator, with the types of impurities permitting higher charge generation appearing to be those that increase the charging tendency of the fluid without significantly affecting its electrical conductivity. As the conductivity of the two-phase mixture reaches a certain level, the electric double layer is destroyed (due to the rapid recombination of charges) and the charge generation effect ceases.

Referring now to the drawings wherein like reference characters designate like or corresponding elements throughout the several views, FIG. **2** shows in schematic form an electrostatic charge generator **10** according to the present invention incorporated in an electrostatic high potential system **15**. The electrostatic high potential system additionally comprises a pump **20** and an electrostatic charge collector **20**. An electrostatic high potential system similar to that indicated at **15** is the subject of the above referenced copending application. As noted above, electrostatic charge generation therein is accomplished by means of a dome shaped ionizing shell enclosing a system of baffles and spiral passageways. Electrostatic charge generator **10** is an improved device for generating electrostatic charges of a given polarity in a circulating dielectric fluid, for example, oil or a mixture of oil and distilled water.

Referring to FIG. **1**, electrostatic charge generator **10** comprises a coaxial conductor having an inner conductor wire **12** coaxial with a hollow, cylindrical outer conductor **13**. A voltage supply **24** applies a potential V_g between outer conductor **13** and inner conductor **12** (inner conductor **12** being indicated to be at a negative potential).

Referring again to FIG. **1** and electrostatic high potential device **15**, pump **20** causes a dielectric fluid to flow, via a conduit **22**, through coaxial charge generation tub **10** at a predetermined rate (as explained below, this rate of flow should at a minimum be sufficient to cause the dielectric fluid to turbulently flow through the charge generator). As discussed below, the dielectric fluid exiting charge generation tube **10** carries a negative electrostatic charge. The negatively charged dielectric fluid flows through a conduit **25** into charge collection apparatus **30** via an inlet port **32**. The dielectric fluid in charge collection apparatus **30** then flows back to pump **20** via an outlet port **34** and a conduit **26** to be recirculated.

Charge collection apparatus **30**, as disclosed in the above referenced copending patent application, comprises an electrically conductive dome shaped shell enclosing spiral passageways and tree electrodes. The points of the tree electrodes intensify the electric field and act to discharge the negative electrostatic charge carried by the dielectric fluid, i.e., draw off the negative charges and accumulate them along the equipotential surface of the conductive shell of charge collection apparatus **30**. In this manner, the large (tens-of-kilovolts) potentials associated with high potential devices can be accumulated on charge collection apparatus **30**.

The electrostatic charge generation process embodied in coaxial charge generator **10** actually comprises two separate mechanisms: the electric double layer effect, described above with respect to ionized impurities in dielectric fluid flow, and a dielectrophoresis effect arising from the non-uniform electric field within the charge generator. As discussed above, a certain degree of electrostatic charge generation (the streaming current) occurs whenever a dielectric fluid flows through a solid tube, as a result of the transfer of charge from ionized impurities of a given polarity to the tube. This electric double layer effect is enhanced in the present invention by reason of the applied potential difference V_g between inner conductor **12** and outer conductor **13** of coaxial charge generator **10**. Ionized impurities in the dielectric fluid are drawn toward the complementarily charged conductor. However, since the intensity of the non-uniform electric field is highest near inner conductor **12** (which is at a negative potential with respect to outer conductor **13**), positive charges are drawn out of the dielectric fluid relatively more rapidly than are negative charges. This results in a gradual accumulation of negative charge carriers as the dielectric fluid flows through coaxial charge generator **10**.

Further enhancing the electric double layer mechanism is the ionizing effect on the electrically neutral hydrocarbon molecules of the dielectric fluid within coaxial charge generator **10**. As best explained with reference to FIG. **3**, electrically neutral hydrocarbon molecules contacting the relatively large surface area of outer conductor **13** (A) tend to be positively ionized through the loss of negative charge to the positively charged outer conductor. These positively charged ions than behave in the same manner as positively ionized impurities, migrating under the influence of the non-uniform electric field toward inner conductor **12** (B). As in the case of ionized impurities, the negatively charged inner conductor **12** draws off positive charge leaving a negatively charged hydrocarbon molecule (C). And, if the dielectric fluid flows through coaxial charge generation tube **10** at a rate sufficient to produce turbulence within the tube, the turbulent flow adjacent to outer conductor **13** (indicated by the curved arrows in FIG. **3**) continuously sweeps ionized hydrocarbon molecules away from the outer conductor toward inner conductor **12**, enhancing the flow of positively charged ions toward this inner conductor. Thus the importance of circulating the dielectric fluid through charge generator **10** at a rate sufficient to insure a turbulent flow.

The second charge generation mechanism results from the "dielectroporesis" effect of a non-uniform electric field on an electrically neutral molecule. As best explained with reference to FIG. **3**, electrically neutral hydrocarbon molecules become polarized in the non-uniform electric field which exists between outer conductor **13** and inner conductor **12** of coaxial charge generator **10** (X). This polarization effect of the electric field within charge generator **10** can be seen as producing induced electric dipoles of the electrically neutral hydrocarbon molecules within the dielectric fluid. The non-uniform electric field causes a measure of corresponding alignment of these induced dipoles. Because the field is non-uniform, one end of the dipole is in a weaker field than the other. Thus, a net force results and the hydrocarbon dipole is pulled towards the area of greatest field intensity, i.e., the area of inner conductor **12** (Y). As these electric dipoles are drawn into contact with the negative polarity of inner conductor **12**, positive charges are removed from the molecules resulting in additional negative charge carriers (Z).

The above described electrostatic charge generation mechanisms combine to produce an abundance of negative charge carriers in the dielectric fluid as the fluid traverses coaxial charge generator **10**. Also, it is apparent that if the polarity of V_g (and, consequently, of the inner conductor of electrostatic charge generator **10**) is reversed, the dielectric fluid would carry a predominance of positive rather than negative charges. Thus, the polarity of the accumulated electrostatic potential on electrostatic charge collection apparatus **30** can be controlled.

In a preferred embodiment of the electrostatic charge generator according to the present invention, coaxial charge generator tube **10** is approximately three feet long and comprised of a one-half inch diameter copper tube as outer conductor **13**, and an uninsulated copper wire as inner conductor **12**. When this electrostatic charge generator is disposed in electrostatic high potential device **15** and a two-phase dielectric fluid mixture (a liquid consisting of ten parts transformer oil to one part distilled water) is caused to circulate therethrough at a rate of approximately three gallons per minute, electrostatic potentials in the neighborhood of 50 kilovolts can be obtained for a value of V_g of forty volts DC. (For most high potential testing, a DC voltage of 30 kilovolts is sufficient.)

The value of V_g can be increased to obtain a more rapid generation of charge carriers and a correspondingly more rapid electrostatic potential build-up in electrostatic charge collection apparatus **30**. However, increases in V_g beyond a certain point (depending upon the type of oil and the type and amount of additives) results in an abrupt cessation of generated voltage, due to the breakdown of the oil causing an increase in conductivity of the dielectric fluid. This prevents any substantial charge from building up.

Obviously, other embodiments and modifications of the present invention will readily come to those of ordinary skill in the art having the benefit of the teachings presented in the foregoing description and the drawings. It is therefore to be understood that this invention is not to be limited thereto and that said modifications and embodiments are intended to be included within the scope of the appended claims.

What is claimed is:

1. An electrostatic charge generator for continuously providing charge carriers of a given polarity in a flow of dielectric liquid, comprising:

a. a coaxial tube comprising:
  1. an inner conductor wire; and
  2. a hollow, cylindrical outer conductor coaxial with said inner conductor;

b. voltage means for applying a DC potential between said inner and said outer conductor; said DC potential creating a non-uniform field the intensity of which is highest near said inner conductor;

c. a circulating dielectric liquid;

d. flow means for causing said dielectric liquid to continuously circulate through said coaxial tube at a rate to insure a turbulent flow;

e. said dielectric liquid being depleted of charges of the opposite polarity of said inner conductor more rapidly then complementary charges are removed at the outer conductor as the dielectric liquid is caused to flow through said coaxial tube, such that the dielectric liquid exiting said coaxial tube carries charges of the complementary polarity;

f. the polarity of said voltage means operable to be reversed for changing the polarity of the charges carried by said dielectric liquid and thus control the polarity of the accumulated charges; and

g. said dielectric liquid comprising a two-phase mixture consisting of from 90 to 100 percent oil and from 0 to 10 percent distilled water.

2. The electrostatic charge generator as recited in claim **1** wherein said dielectric liquid comprises 100 percent oil and 0 percent distilled water.

3. The electrostatic charge generator as recited in claim **2** wherein said coaxial tube comprises:

a. an inner conductor of uninsulated copper wire; and

b. an outer, cylindrical conductor of copper tubing.

4. An electrostatic charge generator as in claim **1** wherein the oil component of said dielectric liquid is transformer oil.

5. In an electrostatic high potential system which includes an electrostatic charge generation chamber and an electrostatic charge accumulation chamber together with means for continuously circulating a dielectric liquid therethrough, an improved electrostatic charge generation chamber comprising:

a. a coaxial tube comprising:
  1. an inner conductor wire;
  2. a hollow, cylindrical outer conductor coaxial with said inner conductor; and

b. voltage means for applying a DC potential between said inner and said outer conductor; said DC potential creating a non-uniform field the intensity of which is highest near said inner conductor;

c. said dielectric liquid being circulated through said coaxial tube at a rate to insure a turbulent flow and being depleted of charges of the opposite polarity of said inner conductor more rapidly than complementary charges are removed at the outer conductor such that the dielectric liquid exiting the coaxial tube carries electrostatic charges of the complementary polarity; and

d. the polarity of said voltage means operable to be reversed for changing the polarity of the charges carried by said dielectric liquid and thus control the polarity of the accumulated charges.

6. The electrostatic charge generator as recited in claim **5** wherein said coaxial tube comprises:

a. an inner conductor of uninsulated copper wire; and

b. an outer, cylindrical conductor of copper tubing.

7. The electrostatic charge generator as recited in claim **6** wherein said coaxial tube is approximately three feet in length and one-half inch in diameter.

8. The electrostatic charge generator as recited in claim **6** wherein said voltage means supplies a DC voltage of approximately 40 volts.

9. An electrostatic charge generator as in claim **6** wherein said dielectric liquid comprises a two-phase mixture consisting of from 90 to 100 percent oil and from 0 to 10 percent distilled water.

10. An electrostatic charge generator as in claim **9** wherein said oil is transformer oil.

* * * * *

F. GHILARDUCCI.
ELECTROSTATIC MACHINE.
APPLICATION FILED JULY 2, 1908.

1,011,939. Patented Dec. 19, 1911.

Fig. 1.

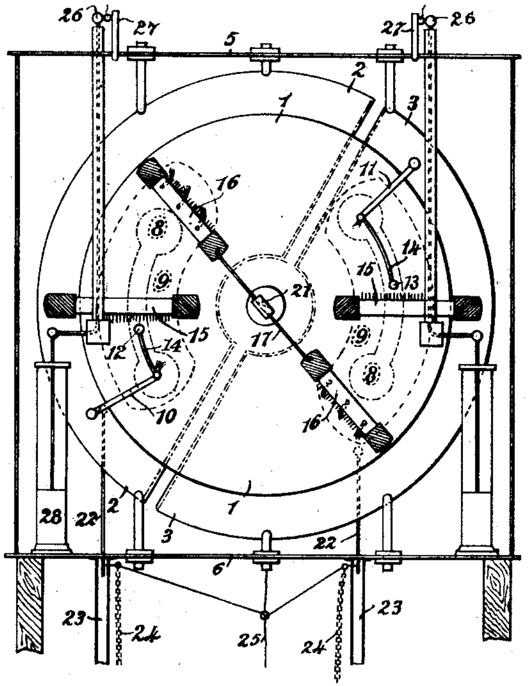
26
27
5
2
1
3
16
11
8
14
15
13
9
21
17
12
10
28
22
6
23
24
25

Fig. 2

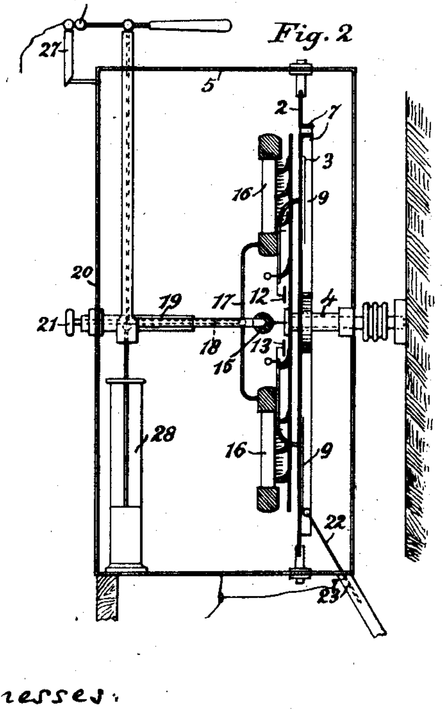
26
27
5
2
7
3
16
9
20
21
19
17
12
4
18
13
15
28
22
23

Fig. 3

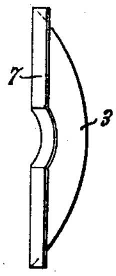

Witnesses:
Louise H. Staaden.
J. Müller

Inventor:
Francesco Ghilarducci
By Alfred Müller
Atty.

UNITED STATES PATENT OFFICE.

FRANCESCO GHILARDUCCI, OF ROME, ITALY.

ELECTROSTATIC MACHINE.

1,011,939. Specification of Letters Patent. Patented Dec. 19, 1911.

Application filed July 2, 1908. Serial No. 441,697.

To all whom it may concern:

Be it known that I, FRANCESCO GHILARDUCCI, a subject of the King of Italy, residing at Rome, Italy, professor at the Roman University, have invented certain new and useful Improvements in Electrostatic Machines, of which the following is a specification.

My invention relates to electrostatic machines, as they are used for producing Röntgen rays, for space telegraphy and in electro-therapeutics, and has for its object to make certain improvements in electrostatic machines of a well-known type, such as the Holtz machine.

My invention more particularly consists in making the stationary disk in two parts, supported by the insulating casing, and provided with flanges at their opposite straight edges, and mounting the rotatable disk on a shaft, supported independently of the casing upon a wall or other rigid structure. The most important advantage obtained by this improvement is the prompt and certain self-excitability of the machine under any atmospheric conditions, and in addition an enormously superior output is obtained from my machine as compared with other machines of equal dimensions, because the rotatable disk can be operated at an extraordinarily high speed, whereas the support for the stationary disk can be made of glass, which is a good insulating material. On account of the gap between the two parts of the stationary disk and on account of the flanges provided at their opposed straight edges, a most perfect insulation of the two parts of the stationary disk is effected. My machine, moreover, works perfectly even in a humid atmosphere and is therefore always ready to perform any radio-telegraphic and electro-therapeutic operations.

In the accompanying drawing, Figure 1 is a front view of an electrostatic machine of the well-known Holtz type with my new improvements; Fig. 2 is a side view of the machine as shown in Fig. 1; Fig. 3 a perspective view of a section of the stationary disk.

1 designates the rotatable disk, which is made of hard rubber, or glass, or the like, and is mounted on a shaft 4, which shaft has no support at its front end, but is journaled in a bearing provided in a socket of a wall or other rigid body, so that a very rapid rotation of about 3,000 revolutions per minute may be given to the disk without setting up oscillation therein. By thus journaling the shaft 4, carrying the rotatable disk 1 in a bearing independent of the casing, the bottom plate of the latter may be made of glass, since it is not subjected to any strains, which would occur, if the shaft 4 were journaled in the casing. This arrangement results, at the same time, in a great gain in insulation. The casing is therefore made entirely of glass, and the sections 2 and 3 of the stationary disk are supported each from the top 5 and the bottom plate 6 of the casing, so as to leave a gap of at least one centimeter between their straight edges, as clearly shown in Fig. 1. In the center of their straight edges, the sections 2 and 3 are recessed, so as to form a circular opening, through which the shaft 4 supporting the disk 1 passes. The sections 2 and 3 of the stationary disk are also made of hard rubber, or preferably of glass, and are each provided at their straight edges with an upright flange 7 at right angles to the planes of the sections and preferably about eight centimeters high. The distance between the two flanges is about 1 centimeter. Each section is provided with a tinfoil armature 8, which is covered with paraffin paper 9. The armatures 8 are connected with the brush-carriers 10 and 11, which are adapted to slide on the front face of the disk 1, while the latter rotates. By a metal rod 14 there is connected to each brush-carrier a small capacity 12 and 13 respectively. By the arrangement of these capacities, the inductive effect of the machine is increased and a corresponding potential increase is effected. The capacities are preferably so located that their rims are about three centimeters away from the collecting combs 15. The collecting combs 15 are supported from the top 5 of the casing and terminate in small knobs 26, which are adjustable by means of brass-rods provided with ebonite handles, so that the knobs 26 can be moved nearer to and farther away from other knobs 27, whereby a variable spark gap may be formed. The two sections of the stationary disk can be brought into contact with the ground by means of two metal rods 22, adapted to slide in insulating tubes 23, secured to the under side of the bottom plate 6 of the casing. Contact is made between the said metal rods and

the armature by the pulling of a cord 25, whereby the armatures of the two sections of the stationary disk are discharged, and the polarity of the machine is reversed. Thus the polarity of the machine may be changed even at a distance. The grounding of the discharges is effected by means of metal chains 24, which, as shown in Fig. 1, are connected to the metal rods 22.

As seen from Figs. 1 and 2, a neutralizing rod is provided, provided at each of its metallic terminals 16 with three brushes. The said terminals are connected by a bent wire 17, which is bow-shaped and fixed at its center to a metal rod 18, adapted to slide in an insulating tube 19, secured to the front plate of the casing. By means of a handle 21 attached to the rod 18, the latter can be turned to give a right or left hand inclination to the neutralizing rod, the effect of this being to vary the tension, or to reverse the polarity of the machine. The polarity is reversed by giving the neutralizing rod a maximum right hand inclination and then returning it to its normal position, shown in Fig. 1. By moving the neutralizing rod nearer to the disk 1, or farther away therefrom, also the electric tension of the machine can be varied.

By making the stationary disk of two sections, as described and shown, and providing a neutralizing rod, the machine can be easily and promptly deënergized by giving the neutralizing rod a turn to move it from its normal position shown in Fig. 1 into a position in which it will lie above the gap formed between the straight edges of the two sections 2 and 3. The deënergizing of the machine can also be obtained by establishing a metallic connection between the two armatures 8 of the sections 2 and 3 of the stationary disk. As soon as the neutralizing rod is returned to its normal position, or after the metallic connection between the armatures has been interrupted, the machine immediately becomes again energized.

In conjunction with the knobs 26 above referred to, two condensers 28 of adjustable capacity may be provided to increase the strength of the electric discharges.

I claim:

1. In an electrostatic machine of the Holtz or a similar type, the combination with an insulating casing of a rotatable ebonite disk supported independently of the said casing, a stationary disk made of insulating material, and comprising two independent sections, forming a gap between them and supported each by the top and bottom of said casing, an armature on each of said sections and brushes connected to said armature and adapted to slide on the front faces of said rotatable disk.

2. In an electrostatic machine the combination with an insulating casing, a rotatable shaft journaled independently of said casing, an ebonite disk mounted on said shaft, a stationary disk made of insulating material, and comprising two sections, forming a gap between them and supported each by the top and bottom of the said casing, an armature provided on each section, brushes connected to said armatures and adapted to slide on the front face of said disk during its rotation, collecting conductors, a neutralizing rod, a plurality of brushes carried at its terminals, an insulating tube carried by the front plate of the casing, a rod carrying said neutralizing rod and capable of adjustment in said insulating tube, an operating handle for moving said metal rod in said tube both longitudinally and axially.

3. In an electrostatic machine the combination with an insulating casing, a rotatable shaft journaled independently of said casing, an ebonite disk mounted on said shaft, a stationary disk made of insulating material, and comprising two sections, forming a gap between them and supported each by to the top and bottom of the said casing, an armature provided on each section, brushes connected to said armatures and adapted to slide on the front face of said disk during its rotation, collecting conductors, a neutralizing rod, a plurality of brushes carried at its terminals, an insulating tube carried by the front plate of the casing, a rod carrying said neutralizing rod and capable of adjustment in said insulating tube, an operating handle for moving said metal rod in said tube both longitudinally and axially, and insulating tubes attached to the bottom of the casing, and metal rods adapted to slide in said insulating tubes and to connect the armatures of the two sections of the said stationary disk.

In testimony whereof I have affixed my signature in presence of two witnesses.

FRANCESCO GHILARDUCCI.

Witnesses:

G. B. Zanardo,

Zlamelle Person.

Copies of this patent may be obtained for five cents each, by addressing the "Commissioner of Patents, Washington, D. C."

United States Patent

[11] **3,614,481**

[72] Inventor **Robert B. Halliday**
R.D. #1, Chenango Forks, N.Y. 13746

[21] Appl. No. **833,289**
[22] Filed **June 16, 1969**
[45] Patented **Oct. 19, 1971**

[54] **ELECTROSTATIC GENERATOR**
10 Claims, 6 Drawing Figs.

[52] **U.S. Cl.** **310/6**
[51] **Int. Cl.** **H02n 1/00**
[50] **Field of Search** 310/5, 6, 7

[56] **References Cited**

UNITED STATES PATENTS

410,053	8/1889	Voss	310/7
1,011,939	12/1911	Ghilarducci	310/6

Primary Examiner—D. X. Sliney
Attorney—Charles S. McGuire

ABSTRACT: A device for collecting a charge of static electricity by rotating a disc or wheel carrying a plurality of individual conductors in proximity to a pair of stationary conductors. Four stationary brushes are arranged in two pairs to wipe the charge from the individual conductors which is initially induced by the natural imbalance in static charge between the stationary and moving conductors. The charges from each pair of brushes are returned to the stationary conductors and stored on an appropriate capacitor.

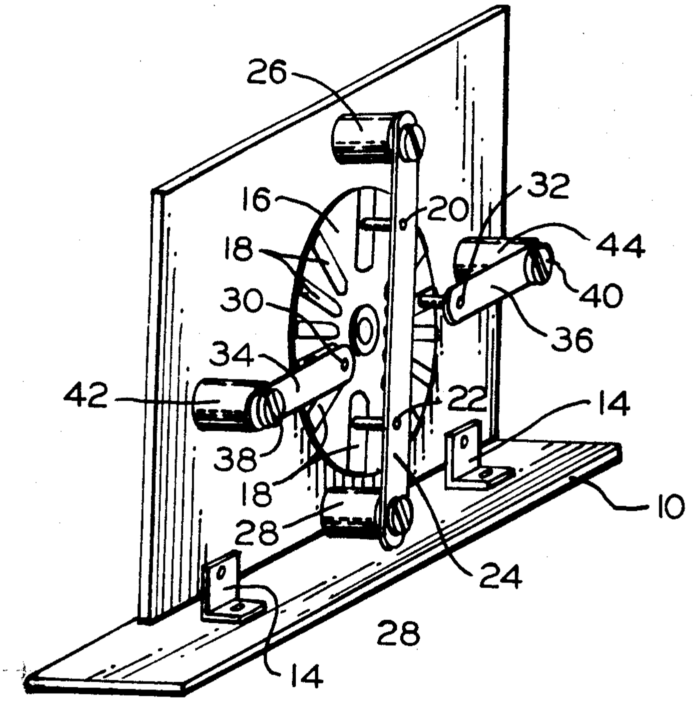

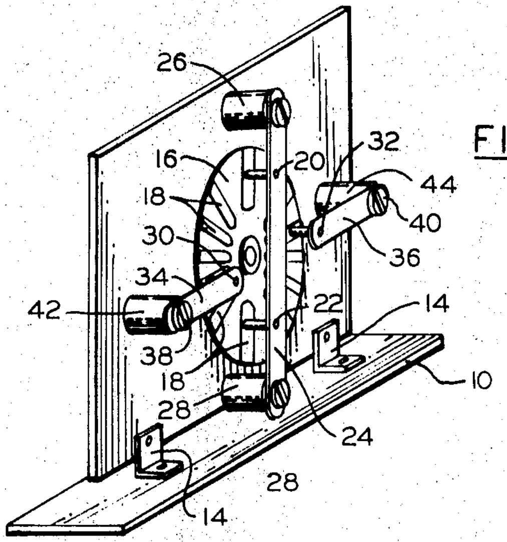
26
16
20
32
44
18
40
30
36
34
22
14
42
38
18
10
28
24
28
14

FIG. 1

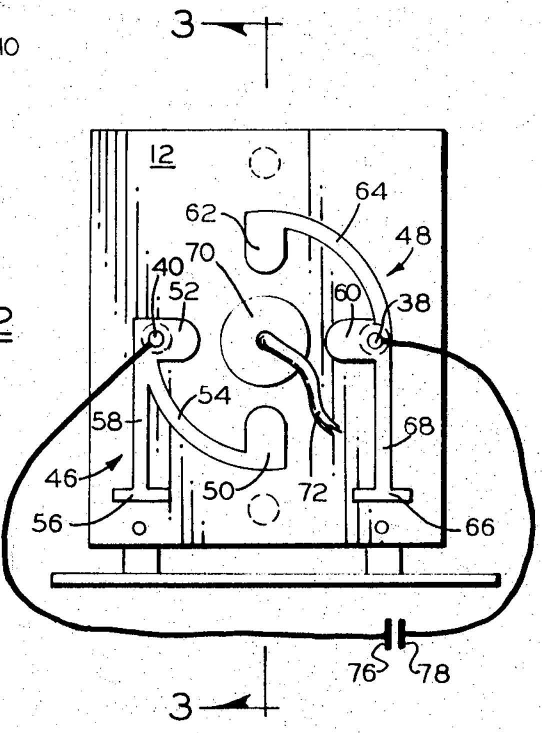
3
12
64
62
48
70
40
52
60
38
54
58
68
46
50
72
56
66
76
78
3

FIG. 2

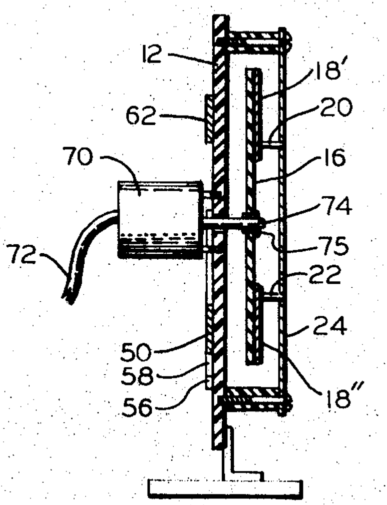
12
18'
62
20
70
16
74
72
75
22
24
50
58
56
18"

FIG. 3

Robert B. Halliday
INVENTOR.

BY Charles L. McGuire,
Attorney

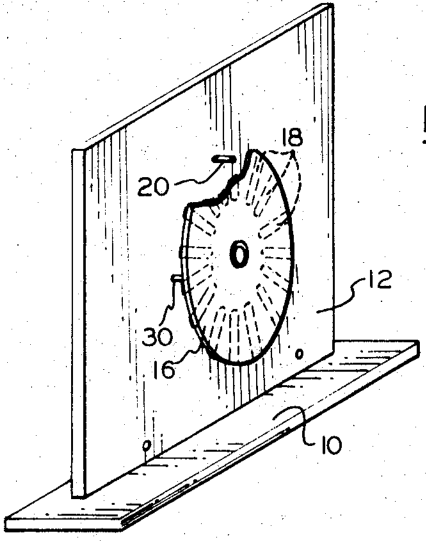
18
20
12
30
16
10

FIG. 4

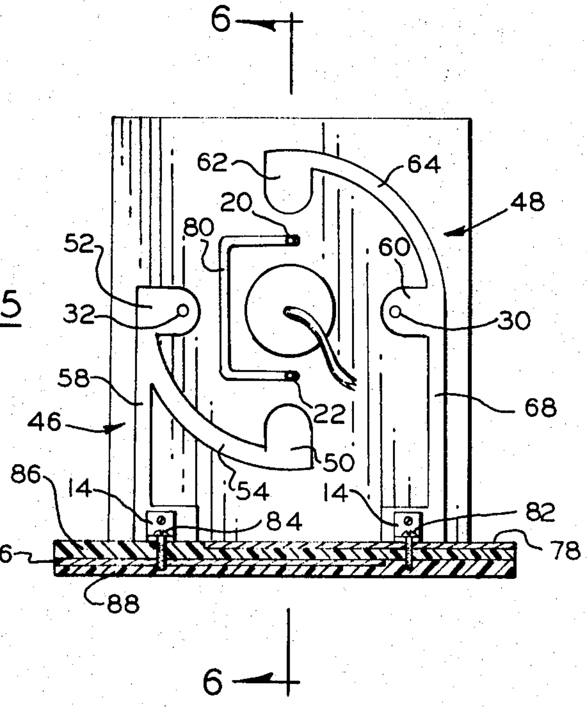
6
62
64
48
20
80
52
60
32
30
58
22
68
46
50
54
86
14
14
82
84
76
78
88
6

FIG. 5

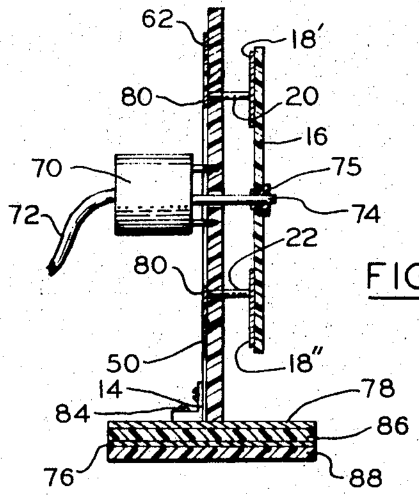
62
18′
80
20
16
70
75
72
74
22
80
50
18″
14
78
84
86
76
88

FIG. 6

Robert B. Halliday
INVENTOR

BY Charles S. McGuire
Attorney

ELECTROSTATIC GENERATOR

This invention relates to electrostatic generators or static electric machines.

Although static electricity has relatively few practical uses, as compared to current electricity, a large number of mechanical arrangements have been devised for concentrating electrical charges produced by attraction and repulsion between particles carrying like or unlike electrical charges. Certain types of electrostatic generators have been successfully used to boost protons and other nuclear particles to energy levels of 10 million electron volts, producing narrow beams of known energy which may be used to study nuclear forces. Other such devices are intended to accumulate much lower charges and may be used merely as instructional aids in demonstrating the physical phenomena which they represent.

The present invention comprises an electrostatic generator of the rotary type wherein electrons are moved through brushes from moving conductors to stationary conductors to cause an initial, low imbalance in electrical charge to become greater and greater. The brushes are preferably provided at intervals of 90° around the path of the rotating conductors, the brushes of one opposing pair being electrically connected to one another and the other pair being insulated from one another and individually connected to the two stationary conductors. The latter may conveniently be carried on a rigid backing which also supports the rotatable disc with the movable conductors in a closely spaced, parallel plane. The backing plate and disc may be carried by a base support which also holds the plates of a capacitor which stores the charge conducted from the stationary conductors. The brushes may be positioned between the backing plate and rotatable disc, or may be (preferably in smaller dimensioned units) carried by arms attached to the backing plate and extending on the opposite side therefrom of the disc.

It is a principal object of the invention to provide an electrostatic generator adapted to accumulate an electrical charge of relatively high voltage in relation to its size.

A further object is to provide a simple and economical device of instructional value in demonstrating the phenomenon of static electricity and accumulation of high-voltage charges therefrom.

Still another object is to provide a device extremely inexpensive to fabricate and assemble which will cyclically charge a capacitor and allow discharge or arcing thereof each time the charge reaches a predetermined voltage level.

Other objects of the invention will in part be obvious and will in part appear hereinafter.

The invention accordingly comprises the features of construction, combination of elements, and arrangement of parts, which will be exemplified in the construction hereinafter set forth, and the scope of the invention will be indicated in the claims.

For a fuller understanding of the nature and objects of the invention reference should be had to the following detailed description taken in connection with the accompanying drawings, in which:

FIG. 1 is a front perspective view of one embodiment of apparatus embodying the invention;

FIG. 2 is a rear elevational view of the apparatus of FIG. 1;

FIG. 3 is a side view in section on the line 3—3 of FIG. 2;

FIG. 4 is a front perspective view of a second embodiment of the invention;

FIG. 5 is a rear view of the FIG. 4 embodiment; and

FIG. 6 is a side view in section on the line 6—6 of FIG. 5.

The invention is shown in a first embodiment in FIGS. 1–3 comprising base 10, adapted to rest on a flat surface and support the other elements, backing plate 12, attached to base 10 by brackets 14, and disc 16 rotatably supported with respect to the backing plate. Disc 16 is made of material which is a good electrical insulator and carries a plurality of individual electric conductors 18, each insulated from the others. A first pair of brushes 20 and 22 is carried by arm 24, which is made of electrically conducting material at least in the portion between brushes 20 and 22 so that the latter are in electrical contact with one another.

Arm 24 is attached to backing plate 12 by screws, rivets or other appropriate fasteners and insulated therefrom by spacers 26 and 28 which hold the arm in slightly spaced relation to the front surface of disc 16. A second pair of brushes 30 and 32 is held by conducting arms 34 and 36, respectively, which are attached to backing plate 12 by conducting screws 38 and 40 and held in properly spaced relation by spacers 42 and 44. Arms 34 and 36 may be joined, of course, to form a single arm, crossing arm 24, if the portion between brushes 30 and 32 is made of an insulator so that the brushes are not in electrical contact with one another.

Turning now to FIG. 2, backing plate 12 carries on its rear surface a pair of stationary conductors designated generally by the reference numerals 46 and 48. Conductor 46 includes a pair of lobes 50 and 52, spaced at 90° on the path of a circle rearwardly adjacent conductors 18 on disc 16, and joined by conducting path 54. Lower conducting portion 56 is joined by path 58 to lobe 52. Conductor 48 includes lobes 60 and 62, also spaced 90° from one another and from lobes 50 and 52, electrically connected by conducting path 64 and connected to lower portion 66 by conducting path 68. Lower portions 56 and 66 of the stationary conductors are connected to opposite plates of an appropriate capacitor, shown schematically in FIG. 2 and indicated by the reference numerals 76 and 78.

Also carried on the rear surface of backing plate 12 is electric motor 70 having cord 72 for connection to an appropriate power source for driving the motor. Shaft 74 (FIG. 3) of motor 70 extends through an opening somewhat larger than the shaft in backing plate 12 and disc 16 is mounted on the end of the shaft with rubber grommet 75 providing a tight frictional engagement.

The space between adjacent surfaces of backing plate 12 and disc 16 should be minimal since the initial flow of electrons from conductors 18 through the brushes upon movement of disc 16 is induced by the natural imbalance in charge between the stationary conductors on the backing plate and the adjacent movable conductors on the disc. For example, lobe 62 of stationary conductor 48 will naturally have an initial concentration of electrons either higher or lower than that of the adjacent conductor 18, shown in the uppermost position in FIG. 3 and designated with the notation 18′. Assuming lobe 62 is more negative it will tend to repel any negative charge in conductor 18′, driving electrons through conducting brush 20. The loss of electrons results in conductor 18′ becoming positively charged as electrons flow through brush 20, conducting support arm 24 and brush 22 to collect on the movable conductor shown in the lowermost position in FIG. 3 and designated with the notation 18″. Thus, conductors 18′ and 18″ carry positive and negative charges, respectively, as they move in a clockwise direction out of contact with brushes 20 and 22. When these two conductors reach brushes 32 and 30, respectively, the charges are wiped off, the positive charge from conductor 18′ flowing through brush 32, conducting arm 36, conducting screw 40 and is transferred to stationary conductor 46. The negative charge from conductor 18″ flows through brush 30, conducting arm 34, conducting screw 38 and is transferred to stationary conductor 48. These two movable conductors (18′ and 18″) are then neutral until returning to their respective original positions at the top and bottom of disc 16 as shown in FIG. 3 and the process is repeated. The same cycle applies, of course, with all the movable conductors as the disc continues to rotate and the positive and negative charges transferred to stationary conductors 46 and 48, and thus to capacitor plates 76 and 78, respectively, continue to increase.

Turning now to the embodiment of FIGS. 4–6, wherein like reference numerals are used to denote elements which may be the same as those shown in the first embodiment, base 10 again supports backing plate 12, attached thereto by brackets 14. Disc 16 carries a plurality of radially spaced and individually insulated conductors 18, this time facing backing plate **12**. Disc **16** is centrally supported on shaft **74** of motor **70**, with rubber grommet **75** providing a tight frictional engagement. Again stationary conductor **46**, having lobes **50** and **52** and conducting paths **54** and **58**, and conductor **48**, having lobes **60** and **62** and conducting paths **64** and **68**, are provided on the side of backing plate **12** opposite disc **16**.

In the present embodiment, however, the arms supporting the brushes are omitted since the latter are mounted directly on the backing plate. Brushes **30** and **32** are connected to stationary conductors **46** and **48**; brushes **20** and **22** are again in electrical contact with one another, but in the present embodiment this is accomplished by an additional conducting path **80** on backing plate **12**, as seen in FIG. **5**. Thus, the basic operation of the apparatus to cause steadily increasing positive and negative charges on stationary conductors **46** and **48** as disc **16** is rotated is the same as in the first embodiment. The structure is somewhat simplified, however, by elimination of the brush support arms. The linear distance between the closest points on conducting path **80** and stationary conductors **46** and **48** must, of course, be great enough that arcing does not occur as the electrostatic charge builds up.

Also shown in the FIGS. **4–6** embodiment is a means for incorporating the capacitor plates directly in base **10**. Brackets **14** are made of a conducting material and attached directly to stationary conductors **46** and **48**. The bracket attached to stationary conductor **48** rests on plate **78**, which is exposed on a portion of the upper surface of base **10**, being secured thereto by screw **82**. The bracket attached to stationary conductor **46** is secured to base **10** by conducting screw **84**, which passes through insulating layer **86**, plate **76** and into insulating layer **88**. Thus, conductor **46** is placed in electrical contact with capacitor plate **76** by the bracket and screw **84**. In this embodiment base **10** may most conveniently be made of two pieces, namely, insulating sheets **86** and **88** having conducting plates **78** and **76**, respectively, integral therewith. Also, it may be desirable to have a portion of the upper surface of plate **76** exposed, by cutting away some of insulating layer **86**, to facilitate shorting the capacitor plates as desired. The plates may be spaced in either embodiment, of course, so that arcing between the plates occurs each time the charge reaches a predetermined level.

Backing plate **12**, carrying stationary conductors **46** and **48**, and disc **16**, carrying movable conductors **18**, may be conveniently fabricated in the manner of circuit boards, i.e., by providing sheets of insulating material covered with a thin conducting layer of copper, or the like, and removing portions of the latter by etching to leave the desired conducting areas on the sheet. Brushes **20, 22, 30** and **32** may be economically and conveniently formed as loops of conducting tape, such as rubber tape containing carbon.

I claim:

1. An electrostatic generator comprising, in combination:

a. a first pair of conducting brushes electrically connected to one another;

b. A second pair of conducting brushes electrically insulated from one another and from the brushes of said first pair;

c. first and second electrical conductors mounted on a common first support and electrically insulated from one another,

d. means electrically connecting one of the brushes of said second pair to said first conductor, and the other to said second conductor;

e. a plurality of individually insulated additional conductors mounted on a common second support;

f. means for causing movement of said additional conductors relative to said brushes, in successive winding contact with brushes of opposite ones of said pairs, and to said first and second conductors, in physical proximity thereto close enough to allow significant mutual influence of the respective electrical charges of said additional conductors on said first and second conductors, but electrically insulated therefrom; and

g. means forming an electrical capacitor with the plates connected respectively to said first and second conductors.

2. The invention according to claim **1** wherein said additional conductors are arranged in a circular configuration and said relative movement is rotary.

3. The invention according to claim **2** wherein said additional conductors are arranged on an insulating disc which is rotated while said brushes and said first and second conductors remain stationary.

4. The invention according to claim **3** wherein said brushes and said first and second conductors are carried on an electrically insulating backing plate which also supports motive means for rotating said disc.

5. The invention according to claim **4** wherein said disc and said backing plate are arranged in closely adjacent, parallel planes and said first and second conductors are on the opposite side of said backing plate from said disc.

6. The invention according to claim **5** wherein said first pair of brushes are supported on said backing plate and electrically connected to one another by a further conductor arranged on said backing plate.

7. The invention according to claim **5** wherein said first pair of brushes are supported on and electrically connected to one another by an arm supported on said backing plate in spaced relation to said disc.

8. The invention according to claim **6** wherein said additional conductors are carried on the side of said disc which faces said backing plate and said brushes are spaced 180° from the other brush of the same pair and 90° from the brushes of the opposite pair.

9. The invention according to claim **8** wherein said first conductor includes two enlarged portions, electrically connected and spaced 90° from one another on the path of a circle in a plane adjacent the plane of rotation of said additional conductors, and said second conductor includes two enlarged portions, electrically connected and spaced 90° from one another and from the two portions of said first conductor on the path of said circle.

10. The invention according to claim **9** wherein said backing plate is supported on a base whereon the plates of said capacitor are arranged.

United States Patent [19]

Wahlstrom

[11] **4,126,822**

[45] **Nov. 21, 1978**

[54] **ELECTROSTATIC GENERATOR AND MOTOR**

[76] Inventor: **Sven E. Wahlstrom,** 570 Jackson Dr., Palo Alto, Calif. 94303

[21] Appl. No.: **801,242**

[22] Filed: **May 27, 1977**

[51] **Int. Cl.**^{2} **H02N 1/00**

[52] **U.S. Cl.** .... **322/2 A;** 310/309; 320/1

[58] **Field of Search** .... 322/2 A; 310/300, 308, 310/309, 310; 320/1

[56] **References Cited**

U.S. PATENT DOCUMENTS

2,413,391	12/1946	Usselman	171/97
2,449,077	9/1948	Lindenblad	175/363
2,628,330	2/1953	Williams	315/200
2,700,901	2/1955	Rickner	73/304
3,005,305	10/1961	Thoma	58/28
3,008,334	11/1961	Lees	73/503
3,013,201	12/1961	Goldie	322/2
3,094,653	6/1963	Le May et al.	322/2
3,147,390	9/1964	Beam	310/4
3,154,699	10/1964	Courtney-Pratt	307/108
3,405,334	10/1968	Jewett et al.	317/250
3,412,318	11/1968	Robinson et al.	322/2
3,582,693	6/1971	O'Hare	310/10
3,610,970	10/1971	Skinner	310/10
3,737,688	6/1973	O'Hare	310/2

Primary Examiner—Robert J. Hickey
Attorney, Agent, or Firm—Flehr, Hohbach, Test, Albritton & Herbert

[57] **ABSTRACT**

Method and apparatus for electrostatically generating power by converting random mechanical motion into electrical energy. The apparatus includes two variable capacitors that have their capacitances varied in an alternating manner by the physical displacement of a common conductor that forms one plate in each of the capacitors. The variable capacitors are formed in a container which holds a fluid for increasing the output power of the apparatus. The fluid in combination with movable conductor also generates sufficient charge to bias the variable capacitors.

29 Claims, 6 Drawing Figures

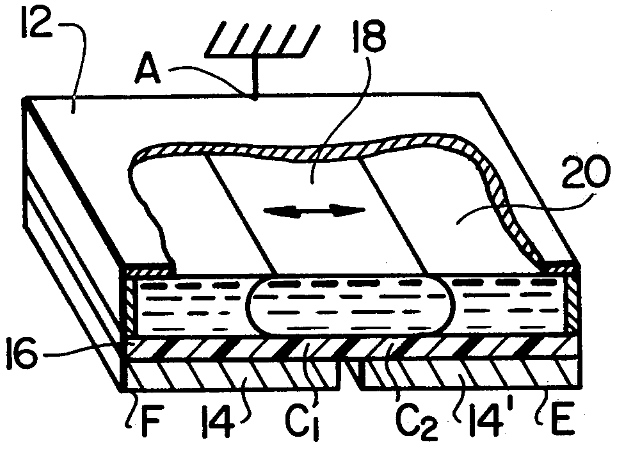

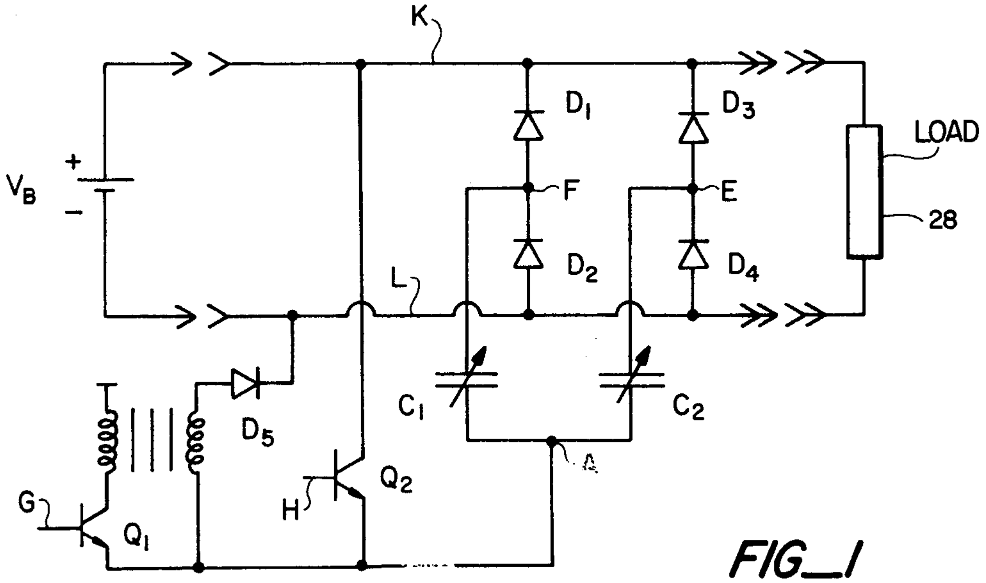
K
D1
D3
LOAD
28
VB
+
−
F
E
D2
D4
L
D5
C1
C2
A
G
Q1
H
Q2

FIG_1

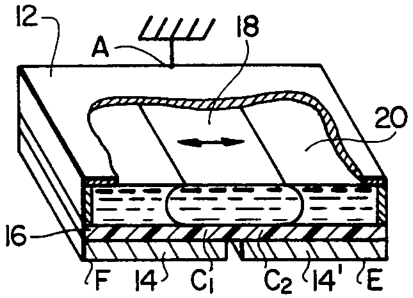
12
A
18
20
16
F
14
C1
C2
14'
E

FIG_2

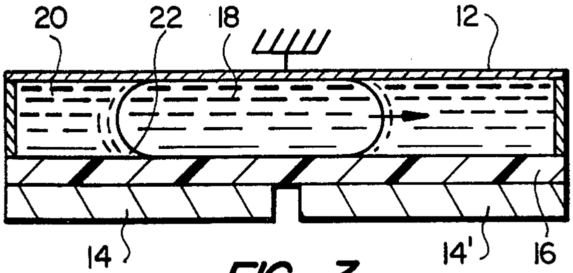
20
22
18
12
14
14'
16

FIG_3

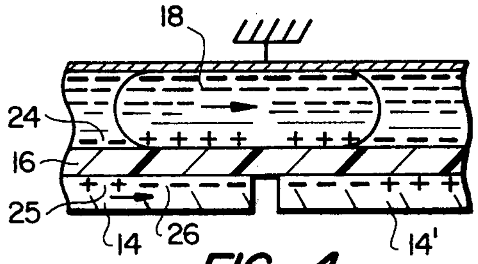
18
24
16
25
14
26
14'

FIG_4

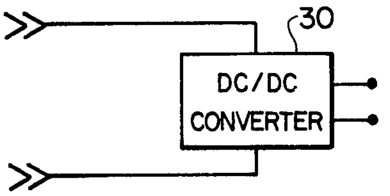
30
DC/DC
CONVERTER

FIG_5

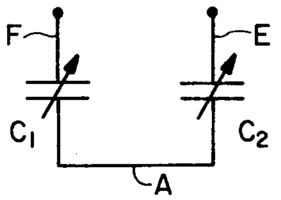

FIG_6

ELECTROSTATIC GENERATOR AND MOTOR

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention generally relates to electrical power-generating systems, and more particularly, to electrostatic generators.

2. Description of the Prior Art

In the past the need for replacing storage batteries has presented problems in convenience, planning and cost. These problems were especially important for the batteries used in portable electronic and electrical equipment. For example, the batteries used for electrical watches, portable radios, calculators, and hearing aids heretofore required frequent replacement. This replacement often was annoying to the owner because of the length of time the equipment was out of service and the high replacement cost of the batteries.

In the watch industry there has been a continuing search for a power source that can make electronic watches as maintenance free as self-winding watches. The industry has been looking for a system that will permit electronic watches to operate indefinitely.

Recently, there has developed a particular problem in the electronic watch industry. Digital watches have become fashionable and some are equipped with light emitting diodes. These diodes draw more current than conventional electronic watch circuits, and often the batteries in these watches require replacement every three or four months. At the present time none of the commercially available electronic watch circuits are equipped with a battery charging system, and the watch battery must be replaced each time the battery is discharged.

The present invention incorporates the well known physical principle that electrical energy can be generated from mechanical motion by mechanically alternating the capacitances of a plurality of capacitors. This principle, however, has never been successfully incorporated into battery charging systems and power generators because the output of these previous systems was too unsymmetrical. Although current was generated in these older systems, the systems could not be adapted to efficiently generate usable AC power.

Another problem encountered with prior electrostatic generators has been the electrical breakdown of the medium between the plates of the capacitors. These generators most efficiently generate power when operated at as high a voltage as possible. Heretofore the upper operating voltage limit was determined by the electrical breakdown of the medium between the plates. For example, if air surrounds the plates, then at high voltage, the electric field causes the air to ionize. The conductive ions thus produced deposit on the surface of the dielectric and make its surface electrically neutral. This type of electrical breakdown of the medium greatly reduces the efficiency of the system.

The present invention also makes use of the triboelectric effect as an alternative principle of operation. The triboelectric effect is the generation of electric charge as a result of the friction between two bodies. The bodies move with respect to each other and electric charge builds up. The moving bodies are fabricated from different materials and the amount of charge produced is proportional to the amount and the frequency of the frictional contact.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide a novel method and apparatus that overcomes the limitations and disadvantages of the prior art.

A further object of the present invention is to convert random mechanical motion into electrical energy by utilizing both gravitational and acceleration forces.

An additional object of the present invention is to develop an electrical energy generator for use in applications where space is at a minimum. This objective includes developing a circuit that operates in the simplest possible manner, utilizes a minimum number of components, and contains the simplest possible circuit. One application of the present invention is for charging electronic watch batteries where space and weight are both limited.

Another object of the present invention is to increase the power-generating efficiency of an electrostatic generator by raising the capacitor breakdown voltage. This permits the generator to operate at a higher voltage and gives the capacitors higher field strength.

Still another object of the present invention is to utilize the triboelectric effect and to eliminate the need for a bias supply for the capacitors. The elimination of the bias supply in the present invention both simplifies the electrical circuit and reduces the cost of producing the apparatus.

These and other features are embodied in a method and apparatus for electrostatically generating power by converting random mechanical motion into electrical energy. The apparatus includes two variable capacitors formed in a container for holding a fluid. The capacitors share a common dielectric member and a common plate. Mechanical motion is used to inversely alternate the capacitance of the two capacitors. The container is filled with a fluid for reducing charge leakage and for increasing the electrical breakdown voltage of the capacitors. This fluid with properly selected properties in combination with the movable conductive plate also generates sufficient charge via the triboelectric effect to bias the two variable capacitors during operation.

A feature of the present invention is that the threshold voltage at which power generation is commenced is easily reached for any small, relative change in the values of capacitance of the variable capacitors. In addition, although the cycle of voltage oscillation and the amount of electrical charge transferred between the capacitors are both dependent on the amount of mechanical motion, these two quantities are completely independent of the rate at which the mechanical motion occurs. Although a slow rate of mechanical motion generates a lower current, power generation in the present invention never ceases as long as capacitor modulation exceeds a predetermined minimum value.

Additional objects and features of the invention will appear from the following description in which the preferred embodiment has been set forth in detail in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a schematic diagram of one embodiment of the circuit of the present invention.

FIG. 2 is a perspective view, partially cut away, of a variable capacitance, electrical energy generator according to the present invention.

FIG. 3 is a side elevation, in section, of the electrical energy generator of FIG. 2.

FIG. **4** is a side elevation, in section and partially cut away, of the electrical energy generator of FIG. **2**, illustrating the triboelectric generation of charges.

FIG. **5** is a block diagram of one output for the circuit of FIG. **1**.

FIG. **6** is a schematic diagram of an alternative embodiment of the present invention illustrating the apparatus as an AC generator.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

FIG. **1** is a schematic diagram of the basic circuit of the present invention. The apparatus includes two variable capacitors **C1** and **C2** that are mechanically connected together so that the capacitance of each varies in an alternating manner. When the capacitance of **C1** is at a maximum, the capacitance of **C2** is at a minimum, and when the capacitance of **C2** is at a maximum, the capacitance of **C1** is at a minimum. The capacitance of these capacitors is on the order of 10,000 picofarads.

FIG. **2** illustrates the construction of the two capacitors **C1**, **C2**. The capacitors are formed in a wall of a container **12** for holding a fluid **20**. The other walls of the container are electrically conductive and the container is fluid tight. Each capacitor includes a separate stationary conductor **14, 14'**. These conductors are two metal plates that are positioned at the extreme ends of the motion of a moving conductor **18**. The moving conductor serves as a common capacitor plate to both capacitors. In the preferred embodiment the moving conductor is a drop of mercury. The two variable capacitors also share a common dielectric member **16**. The two stationary conductors are rigidly attached to the bottom surface of the dielectric member **16**. The moving conductor **18** slides back and forth within the container **12** and on top of the dielectric member **16**. The capacitors **C1** and **C2** are formed by the dielectric member **16**, the common conductor **18**, and the separate stationary conductors **14, 14'**. The circuit illustrated in FIG. **1** is connected to the apparatus of FIG. **2** at points A, E, and F, as indicated in FIG. **2**.

The movable conductor **18** is surrounded by a fluid **20** which raises the breakdown voltage of the two capacitors. This fluid is retained in the container **12** and is selected to have very low conductivity and high breakdown voltage.

It has been observed that when atmospheric air is used between the capacitor plates of an electrostatic generator, the air ionizes at high voltages and reduces the efficiency of the generator. This ionization can be explained with the help of FIG. **3**. When the drop of mercury **18** moves from left to right, the trailing edge of the mercury continuously uncovers a small wedge-shaped gap **22**. In this gap, the force of the electric field is intensified because of the narrow dimensions. The field causes the thin film of air in the gap to ionize. When the air ionizes, the negative ions are attracted to the positively charged mercury drop **18** and the positive ions are attracted to the upper surface of the dielectric member **16** by the negatively charged plates **14, 14'**. These ionized particles tend to maintain a potential on the dielectric surface approaching that of the mercury, thus significantly reducing the transfer of charge between the plates of the capacitors.

The wedge-shaped gap **22** at the edge of the mercury drop can also be visualized as one of two dielectrics connected in series. One dielectric is the film of air in the gap and the other is the dielectric member **16**. In the

preferred embodiment the dielectric member is fabricated from Mylar. Since the breakdown voltage of Mylar is substantially more than the breakdown voltage for atmospheric air, the air ionizes long before the maximum electric field corresponding to the Mylar can be developed across the capacitor plates.

To avoid this problem, a fluid **20** having a high breakdown voltage is used to surround the mercury drop and to eliminate the narrow gaps between the capacitor plates. The fluid also must have very low electrical conductivity in order to prevent any of the charge on the mercury **18** from leaking through the fluid **20** and reaching the surface of the dielectric member **16**. This leakage charge, like the ionized air particles, tends to cancel the charge transfer between the capacitors.

Mylar has a breakdown voltage of about 5,000 volts/-mil. It is believed that a fluid having a breakdown voltage of at least 1,000 v/mil and a resistivity of at least 10^{12} ohm-cm would be satisfactory in a biased system. One fluid having these characteristics is the oil used in commercial high power electrical transformers. For example, the Diola Ax oil obtained from the Shell Oil Company has been used in one embodiment actually constructed. Gases such as Freon and helium can also be used to increase the breakdown voltage. In addition, since the breakdown voltage of a gas is typically directly proportional to its pressure, filling the fluid container **12** with a low conductivity gas at high pressure would also prevent ionization.

In the preferred embodiment, the fluid **20** in the container **12** is also selected to enhance the triboelectric generation of charge. The triboelectric effect described above induces the deposition of negative charges on the upper surface of the dielectric member **16**. In FIG. **4** these charges are indicated by reference numeral **24**. When one negative triboelectric induced charge is positioned on the upper surface of the dielectric member **16**, a corresponding positive charge **25** is bound opposite to it on the surface of the stationary conductor **14**. Since the conductor is electrically neutral across its surface, there is an additional negative charge **26** which is bound opposite to the positive charge left on the mercury, when charge **24** was deposited. The net result of triboelectric charge generation is the addition of negative charges to the capacitors **C1** and **C2**.

With each increment of motion of the mercury **18** from left to right, as illustrated in FIG. **4**, an equal amount of charge is deposited on the surface of the dielectric member **16**. This deposition of charge causes an increase in the overall charge on capacitor **C1** as described above. Concurrently with this increase in charge, the capacitance of the capacitor is decreasing because of the motion of the mercury conductor. The voltage V1 across a capacitor C1 is a function of the following expression:

$$V1 = \frac{K(C_{max} - C_1)}{C_1}$$

Where K is a constant representing the amount of charge triboelectrically generated for each increment of capacitance reduction C_1 is the instantaneous capacitance of the capacitor, and C_{max} is the maximum value of C_1 when the motion of the mercury starts.

It has been observed that a fluid having low electrical conductivity and low viscosity can substantially increase the triboelectric generation of charge. A fluid of low conductivity is preferred in order to prevent the leakage of charge from the mercury drop **18** to the surface of the dielectric member **16** as described above. A fluid having low viscosity is preferred because the frictional contact between the mercury drop and the dielectric member is increased. It is believed that the rubbing contact between the movable conductor and the dielectric member causes the triboelectric effect to occur. A low viscosity fluid is also preferred in order to prevent the fluid from lubricating the mercury. It is also believed that a fluid having a resistivity of at least 10^{15} ohm-cm and a viscosity of less than 600 centistokes will satisfy these requirements.

One fluid having low viscosity, a high breakdown voltage and low electrical conductivity is Dimethylpolysiloxane. This fluid is obtainable from the Dow Corning Corporation of Midland, Michigan 48640. This fluid has a resistivity of 10^{15} ohms-cms, a dielectric strength of 35kv/100 mil, and a viscosity of 0.65 centistokes. When used in one embodiment of the present invention that was actually constructed, this fluid generated a bias of approximately 1,000 volts across a 6 um layer of Mylar. This voltage was sufficient to bias the variable capacitors C1 and C2 and to eliminate the need for a separate bias supply.

The present invention contemplates an electrostatic generator for producing electrical power. One application of this generation of power is for charging storage batteries. Referring to FIG. **1**, the circuit is connected to a rechargeable battery indicated by reference letters V_B. The rechargeable battery has an operating voltage, depending on its construction, of approximately 1.3 to 1.7 volts. Although any suitable rechargeable battery can be used, a nickel cadmium battery is preferred for electronic watch circuits. In FIG. **1** the rechargeable battery is connected in series with a load **28** representing the operating mechanism of the watch.

Connected across the battery V_B in parallel with the load **28** are two diodes, D1 and D2. The cathode of diode D1 is connected to the positive side of the battery, and the cathode of diode D2 is connected to the anode of diode D1. The anode of diode D2 is connected across the battery and in parallel with both the load **28** and the diodes, D1 and D2, are two other diodes, D3 and D4. The cathode of diode D3 is connected to the positive side of the battery, and the cathode of diode D4 is connected to the negative side of the battery. The diodes are selected to have low leakage in order to increase the efficiency of the circuit. The four diodes D1–D4 form a rectifier circuit that converts the AC output from the two variable capacitors C1, C2 into a DC output for charging the battery V_B.

In the preferred embodiment, a liquid mercury moving conductor **18** is preferred for several reasons. Because of its high density, the mercury maximizes the conversion of available work into electrical energy. When made of mercury, the moving conductor **18** performs more work in a limited space with minimum motion. Mercury is also preferred because it has high surface tension. During operation the moving conductor is subjected to large acceleration forces and a large electrostatic field. The high surface tension of mercury prevents the liquid drop from breaking up into a plurality of small droplets. These small droplets would retain a charge of their own, and the capacitor, in effect, would lose its power generating capability.

To maximize the generation of power an electrostatic generator must maximize the stored electrical energy per unit of surface area of each capacitor. The stored electrical energy in a capacitor is proportional to CV^2, where C is the capacitance of the capacitor and V is the voltage induced across the plates. The capacitance C is proportional to the dielectric constant of the dielectric material and is inversely proportional to its thickness. The voltage V is proportional to the breakdown voltage of the dielectric material. It has been observed that the breakdown voltage for most dielectric materials is proportional to the square root of the thickness of the dielectric. Thus, the electrical energy stored in a capacitor tends to be constant regardless of the thickness of the dielectric material. In the present invention the preferred dielectric material is one which has the highest product of dielectric constant and the square of the breakdown voltage for a given thickness.

A practical means to maximize the generation of power and the amount of energy stored in the capacitors without operating at bias voltages that are too high is to use very thin dielectric materials. For example, the dielectric member **16** can be fabricated from Mylar. In its thinnest presently available commercial form, Mylar has a breakdown voltage of approximately 1,000 volts. An additional, perhaps more suitable dielectric material, is Parylene, obtainable from the Union Carbide Company. A one micron thick layer of Parylene has the breakdown voltage of approximately 500 volts. Thus, Parylene permits the storage of very high energy without requiring a high bias voltage.

The electrostatic generator is operated by subjecting the container **12** and the moving conductor **18** to either random or cyclical forces. These forces can be acceleration forces that either operate directly on the mercury to cause it to slide back and forth or gravitational forces that generate potential energy that is expended in moving the mercury. In the embodiment designed for electronic watches, these forces are provided by the random motion of the watch wearer's wrist. At all times the mercury is free to slide within the container and is only constrained by the dimensions of the container.

In order to describe the operation of the circuit illustrated in FIG. **1**, assume as initial conditions that there is a bias voltage of −500 volts on point A. This bias voltage is obtained from either the triboelectric effect or from a separate bias supply as described below. The voltage across the battery V_B is approximately 1.7 volts. The capacitance of capacitor C1 is at its maximum value and the capacitance of capacitor C2 is nearly zero. The electrical charge Q on the capacitor C1 is given according to the following equation $Q_1 = C_1V$, and the energy stored therein is given by the following equation $E_1 = C_1V^2$.

The energy stored in the capacitors C1 and C2 is transferred by the motion of the mercury **18** in a manner to inversely alternate the capacitance of capacitors C1 and C2. When the capacitance of capacitor C1 is decreased and the capacitance of C2 is increased, the voltage at point F (FIG. **1**) becomes positive relative to the voltage at point A. In addition, the voltage at point E becomes negative with respect to the voltage at point A. This change in voltage occurs because both capacitors attempt to maintain the quantity of charge on their plates unchanged. Since the capacitance of C1 has decreased, the voltage across C1 must increase in order to maintain the quantity of charge of Q1 at a constant value. When the voltage across C1 has increased sufficiently, the diode D1 becomes forward biased and current flows through D1. The current path is through diode D1, into the positive side of the battery V_B, and through diode D4 to point E. The common terminal point A has no net current flowing through it after it is initially charged. Simultaneously, the capacitance of capacitor C2 is increasing in value and tends to forward bias diode D4. Capacitor C2 tends to drive point F more positive and point E more negative. In summary, as the capacitance of capacitor C1 decreases from its maximum value to zero, the electric charge Q on capacitor C1 is actually transferred from point F through the battery V_B to point E and, thus, the battery is charged by this current.

The energy stored in the capacitors C1 and C2 is transferred back in a cycle that is the reverse of the hereinbefore described sequence. As initial conditions assume that there is a bias voltage at point A of −500 volts and the voltage across the battery is approximately 1.7 volts. The capacitance of capacitor C1 is at a minimum value and the capacitance of capacitor C2 is at a maximum value. Capacitor C2 contains an electronic charge given by the equation $Q2 = C_2V$.

When the mechanical motion, hereinafter described, causes the capacitance of capacitor C2 to decrease, the voltage at point E becomes positive relative to the voltage at point A. Simultaneously, the capacitance of capacitor C1 tends to increase and the voltage at point F tends to go negative. As before both capacitors initially try to maintain the quantity of charge on their plates unchanged. When the voltage across C2 increases sufficiently, diode D3 becomes forward biased. The electric charge flows from point E through diode D3, into the positive side of the battery, and through the diode D2 to point F. The common terminal point A has no net current flowing through it.

It should be noted that in each part of the cycle there is a net current flow through the battery and this flow is always from the positive side to the negative side of the battery. Thus, battery charging current is produced regardless of which direction the values of capacitance are changing.

It should be further noted that the sum of the electric charges on points E and F remains unchanged throughout the cycle of capacitance change. Although there is a net production of energy which charges the battery, there is no loss of electric charge. The energy for charging the battery is generated by a moving conductor **18** that moves against the electrostatic forces caused by the voltage difference between points E and F.

Although the preferred embodiment incorporates a battery V_B, it should be appreciated that any suitable electrical energy storage device can be substituted for the battery V_B. Suitable energy storage devices are those that can store between cycles the electrical energy generated by mechanical motion and then can be discharged into the load. The construction and operation of the remainder of the circuit remains exactly the same. These storage devices include, for example, a large capacitor that is charged up by the current flow and then discharged into the load. The large capacitor replaces the battery V_B and is connected at the same point in the circuit, FIG. **1**. In like manner, the battery V_B can also be replaced by a DC motor connected to a fly wheel. The electrical energy generated by the circuit drives the DC motor and attached fly wheel thereby storing the generated electrical energy as kinetic energy. The stored energy can be recovered by a DC or AC generator also connected and driven by the fly wheel.

Another suitable electrical energy storage device that can be substituted for the battery V_B, FIG. **1** is a power distribution system. These distribution systems include the domestic power systems operated by public utilities and the private systems operated on ships and in independent commercial installations. When operating, a power distribution system is a short term, energy storage device because of the transmission time through the system between the energy generator and the load. In effect, the energy generator charges up the transmission lines and after a short period of time, the lines discharge into the load. In some cases there can be hundreds of miles of transmission lines between the generator and the load and the transmission time can be quite substantial. In addition, a power distribution system acts as a buffer between the energy generator and the load. In most applications the load varies continuously and correspondingly changes the power demanded from the generator. The distribution system smooths out the fluctuations in power demand. This operation of the distribution system as a buffer between the energy generator and the load is comparable to the operation of the battery V_B described hereinbefore.

To provide sufficient charging current for a power distribution system, a plurality of simultaneously operating energy generators are used and their outputs are connected to a conventional DC to AC converter. For example, a large number of energy generators can be floated on the surface of the ocean and accelerated by the motion of the ocean waves. The generators are disbursed over a wide area so that at any instant in time each energy generator is undergoing a different acceleration force. The energy generators are all connected in common to a DC/AC converter so that a nearly continuous AC output is obtained. This AC output is connected to the power distribution system and the electrical energy is thereby stored and transmitted to the load.

Although the preferred embodiment incorporates a battery V_B, it also should be appreciated that the circuit of FIG. **1** can be operated without a battery or an energy storage device. The two variable capacitors C1 and C2 and the four diodes D1–D4 form an electrostatic power generator and can be connected directly to a load as illustrated in FIG. **1**.

Although the triboelectric effect described above can develop sufficient voltage to bias the circuit of FIG. **1**, less severe requirements on the fluid surrounding the moving conductor can be applied by pre-charging the two variable capacitors C1 and C2 to a high voltage. To pre-charge the variable capacitors, the pulse transformer circuit illustrated in FIG. **1** is used to apply a high voltage across both variable capacitors. When point G is given a positive pulse, the transistor Q1 and the transformer generate a high voltage which is rectified through diode D5. A conventional blocking oscillator can also be used to charge the variable capacitors.

It has also been observed that when some fluids are used in the container, there is a certain amount of charge leakage between the moving conductor **18** and the dielectric member **16**. One example of such a fluid is transformer oil that has been contaminated by water. This leakage causes electrical charges to accumulate on the surface of the dielectric member **16** and to neutralize the electrostatic field between the plates of the two capacitors. This leakage is a problem because it decreases the efficiency of power generation.

Referring to FIG. **1**, the circuit including transistor Q2 is used to discharge the two capacitors C1, C2 and to remove the build up of charge on the capacitor plates due to charge leakage through the fluid **20**. When transistor Q2 is pulsed at point H, capacitors C1 and C2 are discharged through diodes D1 and D3.

To increase the energy exchange between capacitors C1 and C2 in FIG. **1** and thereby the charging energy to the load, an impedance matching arrangement is used. This may take the form of a DC/DC converter **30**, FIG. **5** for converting the high voltage differentials between points K and L, FIG. **1**, to a low voltage, high current output applied to the battery and the load.

It should also be noted that if the load or the electrical energy storage device can be charged using AC power, then the diodes D1–D4 are not required. Thus, the circuit of FIG. **6** is used for an AC load. This circuit merely comprises the two variable capacitors C1 and C2 connected in series, and the AC load is placed across points E, F of FIG. **6**. The construction and operation of the two capacitors is the same as described above.

In addition, although mercury is disclosed in the preferred embodiment, any conductive liquid or solid can be used for the moving conductor **18**.

It is also contemplated that the present invention is operable in all battery power equipment subjected to mechanical motion. This type of equipment includes electronic wrist watches, portable radios, hearing aids, cardiac pacemakers, small portable calculators, navigational buoys, other water borne equipment, and vehicular mounted electronic equipment.

Thus, although the best mode contemplated for carrying out the present invention has been herein shown and described, it will be apparent that modification and variation may be made without departing from what is regarded to be the subject matter of the invention.

What is claimed is:

1. Apparatus for electrostatically generating power by converting mechanical motion into electrical energy, comprising:

(a) a container for fluid;

(b) two variable capacitors formed in the container, said capacitors each having a separate conductor and sharing a common dielectric member and a common movable conductor, said common conductor is separated from the separate conductors by the dielectric member, moves between the separate conductors in response to mechanical motion, and thereby inversely alternates the capacitance between the two variable capacitors;

(c) an output power increasing, electrically resistive fluid held within the container and in contact with the common conductor and the common dielectric member; said fluid having a breakdown voltage substantially greater than atmospheric air;

(d) circuit means connected to the common conductor and the separate conductors for transferring charge between the variable capacitors and having two output junctions for transmitting electrostatically generated power.

2. An apparatus as in claim **1** in which the fluid has a breakdown voltage of at least 1,000 volts per mil.

3. Apparatus for electrostatically generating power by converting mechanical motion into electrical energy, comprising:

(a) a container for fluid;

(b) two variable capacitors formed in the container, said capacitors each having a separate conductor and sharing a common dielectric member and a common movable conductor, said common conductor is separated from the separate conductors by the dielectric member, moves between the separate conductors in response to mechanical motion, and thereby inversely alternates the capacitance between the two variable capacitors;

(c) a capacitor breakdown voltage increasing fluid held within the container and in contact with the common conductor and the common dielectric member, said fluid having a breakdown voltage substantially greater than atmospheric air and resistivity of greater than 10^{12} ohm- centimeters; and

(d) circuit means connected to the common conductor and the separate conductors for transferring charge between the variable capacitors and having two output junctions for transmitting electrostatically generated power.

4. An apparatus as in claim **3** in which the breakdown voltage increasing fluid has a breakdown voltage of at least 1,000 V/mil.

5. An apparatus as in claim **3** in which the breakdown voltage increasing fluid is transformer oil.

6. An apparatus as in claim **3** in which the breakdown voltage increasing fluid is Dimethypolysiloxane.

7. An apparatus as in claim **3** in which the breakdown voltage increasing fluid is Freon.

8. An apparatus as in claim **3** in which the breakdown voltage increasing fluid is helium.

9. An apparatus as in claim **3** in which the common conductor is an electrically conductive liquid connected to the circuit means.

10. An apparatus as in claim **3** in which the common conductor is a drop of mercury.

11. An apparatus as in claim **3** in which means for pre-charging the variable capacitors is connected to the circuit means.

12. An apparatus as in claim **3** in which means for discharging the variable capacitors is connected to the circuit means.

13. Apparatus for electrostatically generating power by converting mechanical motion into electrical energy, comprising:

(a) a container for fluid;

(b) two variable capacitors formed in the container, said capacitors each having a separate conductor and sharing a common dielectric member and a common movable conductor, said common conductor is separated from the separate conductors by the dielectric member, moves between the separate conductors in response to mechanical motion, and thereby inversely alternates the capacitance between the two variable capacitors;

(c) a triboelectric charge generating fluid held within the container and in contact with the common conductor and the common dielectric member, said fluid in combination with the movable conductor generates sufficient charge to bias the variable capacitors, said fluid also has a resistivity substantially greater than atmospheric air; and p1 (d) circuit means connected to the common conductor and the separate conductors for transferring charge between the variable capacitors and having two output junctions for transmitting electrostatically generated power.

14. An apparatus as in claim **13** in which the charge generating fluid has a resistivity of at least 10^{15} ohm-cms and a viscosity of less than 600 centistokes.

15. An apparatus as in claim **13** in which the charge generating fluid is Dimethylpolysiloxane.

16. An apparatus as in claim 13 in which the common movable conductor is an electrically conductive liquid connected to the circuit means.

17. An apparatus as in claim 13 in which the common conductor is a drop of mercury.

18. An apparatus as in claim 13 in which means for pre-charging the variable capacitors is connected to the circuit means.

19. An apparatus as in claim 13 in which means for discharging the variable capacitors is connected to the circuit means.

20. Apparatus for charging an electrical energy storage means by converting mechanical motion into electrical energy, comprising:

(a) an electrical energy storage means chargeable by passing a charging current therethrough;

(b) first and second diode means connected across said energy storage means;

(c) third and fourth diode means connected across an energy source, said first and second diode means being connected in parallel with said third and fourth diode means and all of said beforementioned diode means being forward biased with respect to the charging current;

(d) two variable capacitors each having a separate conductor and sharing a common dielectric member and a common movable conductor, one separate conductor being connected between the first and second diode means and the other separate conductor being connected between the third and fourth diode means, said common conductor is separated from the separate conductors by the dielectric member, moves between the separate conductors in response to mechanical motion, and thereby inversely alternates the capacitance between the two variable capacitors; and

(e) an output power increasing, electrically resistive fluid held within the two variable capacitors and in contact with the common conductor and the common dielectric member, said fluid having a resistivity substantially greater than atmospheric air.

21. An apparatus as in claim 20 in which the fluid is a capacitor breakdown voltage increasing fluid.

22. An apparatus as in claim 20 in which the fluid is a triboelectric charge generating fluid.

23. An apparatus as in claim 20 in which the electrical energy storage means is a rechargeable DC energy source.

24. An apparatus as in claim 20 in which the electrical energy storage means is a battery.

25. An apparatus as in claim 20 in which the electrical energy storage means is a capacitor that can be charged-up by the variable capacitors and can be discharged into a load.

26. An apparatus as in claim 20 in which the electrical energy storage means is a fly-wheel rotated by a DC motor.

27. An apparatus as in claim 20 in which the electrical energy storage means is a power distribution system.

28. Method for electrostatically generating power by converting mechanical motion into electrical energy, comprising the steps of:

(a) triboelectrically charging first and second variable capacitors with a quantity of electrical charge;

(b) generating an electrostatic force field in the first capacitor;

(c) moving a body against the electrostatic force field in the first capacitor so that charge is transferred through a circuit from the first capacitor to the second capacitor;

(d) generating an electrostatic force field in the second capacitor; and

(e) moving the body against the electrostatic force field in the second capacitor so that charge is transferred from the second capacitor back to the first capacitor through the circuit, whereby said transfer of charge between the variable capacitors is the electrostatically generated power.

29. A method as in claim 28 in which the first and second variable capacitors are triboelectrically charged by relative motion between a dielectric material, the moving body, and a fluid in contact with both.

* * * * *

US 20170237466A1

(19) **United States**

(12) **Patent Application Publication**

Carr

(10) **Pub. No.: US 2017/0237466 A1**

(43) **Pub. Date: Aug. 17, 2017**

(54) **SENSOR SYSTEM WITH ENERGY HARVESTING**

(71) Applicant: **William N. Carr**, Raleigh, NC (US)

(72) Inventor: **William N. Carr**, Raleigh, NC (US)

(21) Appl. No.: **15/434,939**

(22) Filed: **Feb. 16, 2017**

Related U.S. Application Data

(60) Provisional application No. 62/388,989, filed on Feb. 16, 2016.

Publication Classification

(51) **Int. Cl.**

H04B 5/00	(2006.01)
H02J 50/12	(2006.01)

(52) **U.S. Cl.**

CPC ***H04B 5/0031*** (2013.01); ***H02J 50/12*** (2016.02); ***H04B 5/0037*** (2013.01)

(57) **ABSTRACT**

A sensor system with energy harvesting has an energy-harvesting mode of operation and a load-enabling mode of operation, enabled at separate times. The system includes an antenna, an impedance-matching network, an RF switch, inductive and diode components (for forming a resonant RF-to-DC voltage multiplication circuit in conjunction with other elements of the system), a DC switch, an energy-storage device, and a primary load circuit.

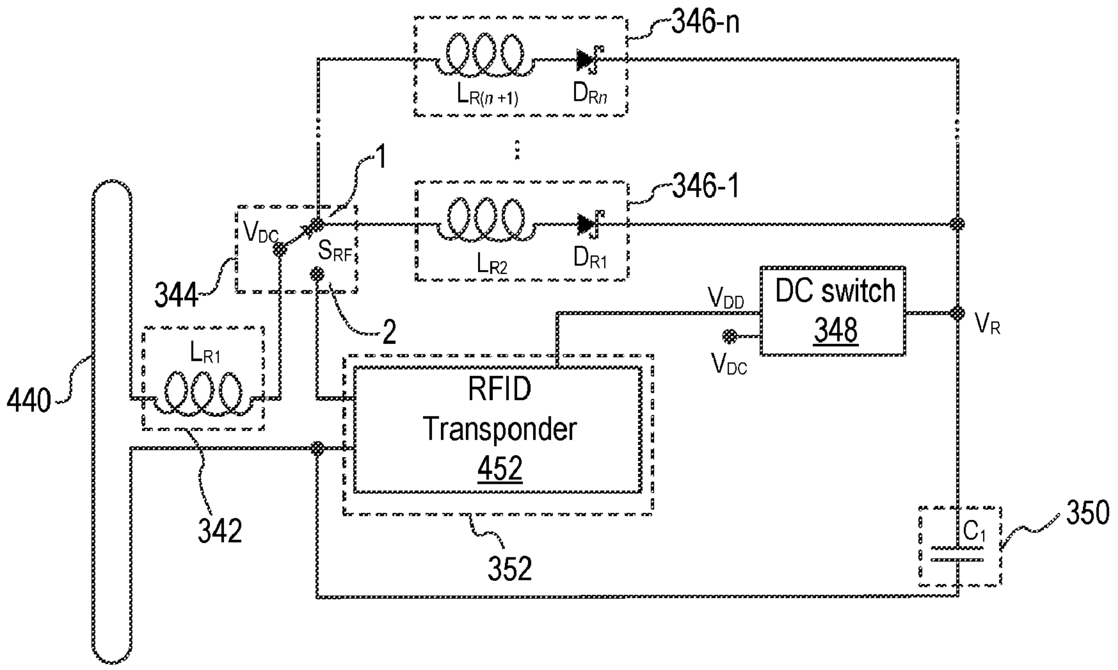

FIG.1
Prior Art

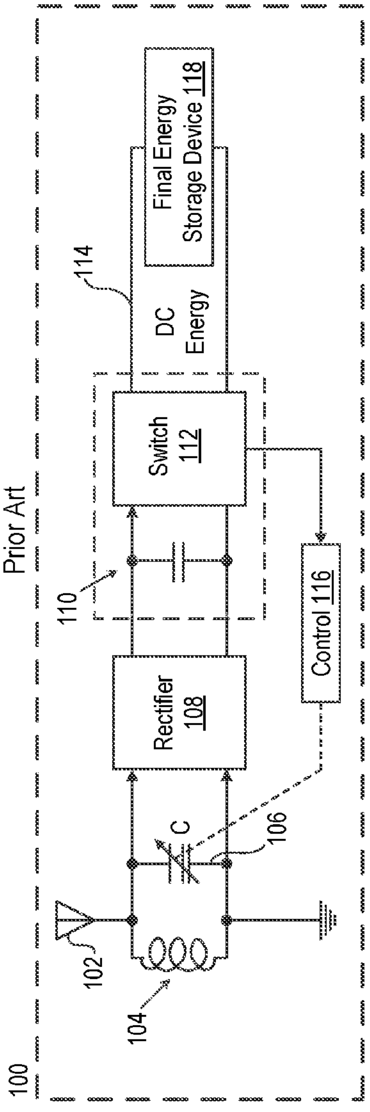
100
102
104
106
C
Rectifier 108
110
Switch 112
Control 116
114
DC Energy
Final Energy Storage Device 118

FIG.2
Prior Art

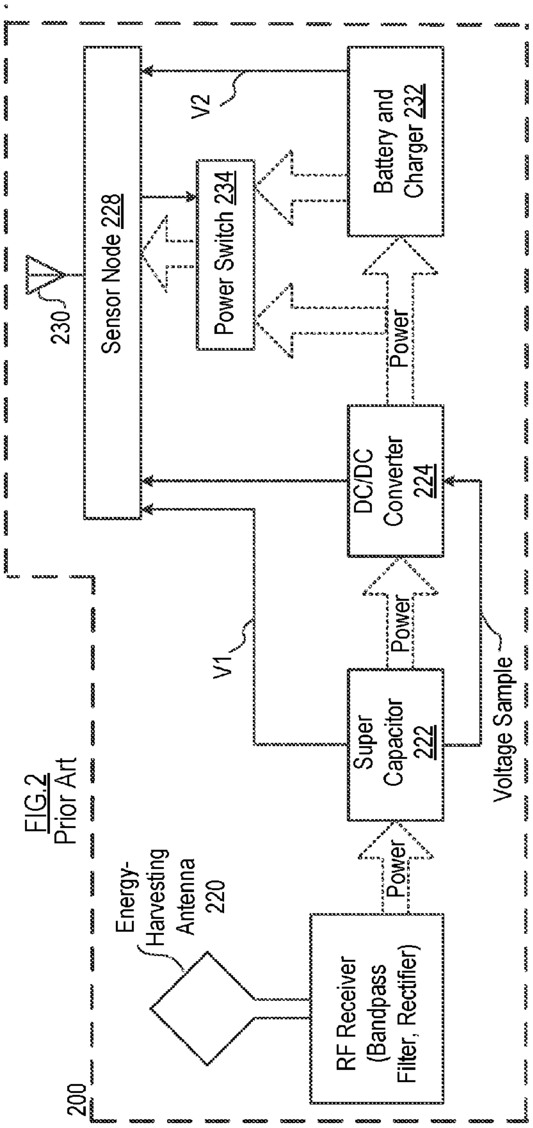
200
Energy-Harvesting Antenna 220
RF Receiver (Bandpass Filter, Rectifier)
Power
Super Capacitor 222
Power
Voltage Sample
V1
DC/DC Converter 224
Power
Sensor Node 228
230
Power Switch 234
Battery and Charger 232
V2

FIG.3A

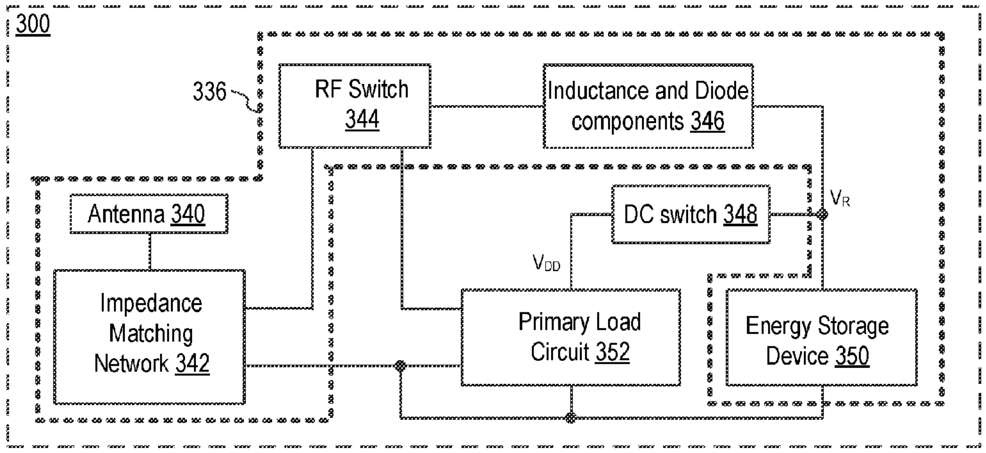
300
336
RF Switch 344
Inductance and Diode components 346
Antenna 340
DC switch 348
V_R
V_{DD}
Impedance Matching Network 342
Primary Load Circuit 352
Energy Storage Device 350

FIG.3B

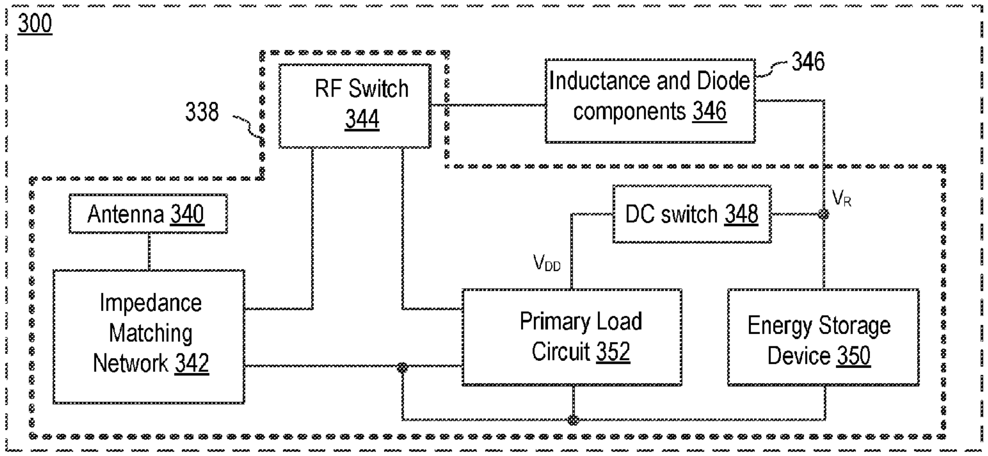
300
338
RF Switch 344
Inductance and Diode components 346
346
Antenna 340
DC switch 348
V_R
V_{DD}
Impedance Matching Network 342
Primary Load Circuit 352
Energy Storage Device 350

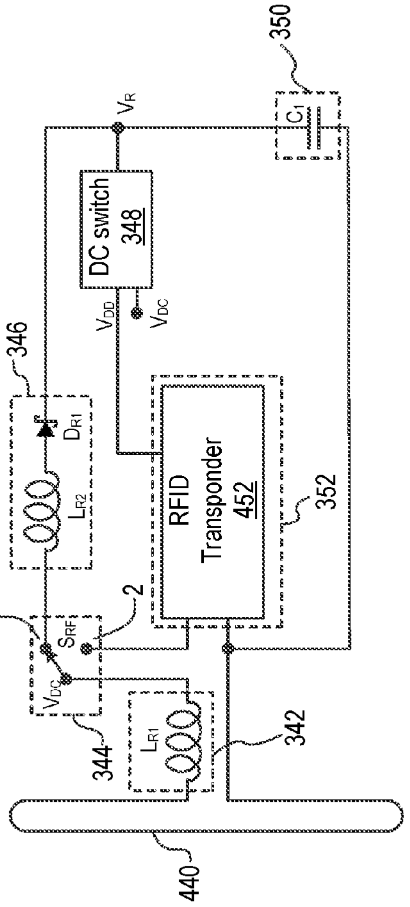
350
C_1
V_R
DC switch
348
V_{DD}
V_{DC}
346
D_{R1}
L_{R2}
RFID
Transponder
452
352
1
2
S_{RF}
V_{DC}
344
L_{R1}
342
440

FIG.4A

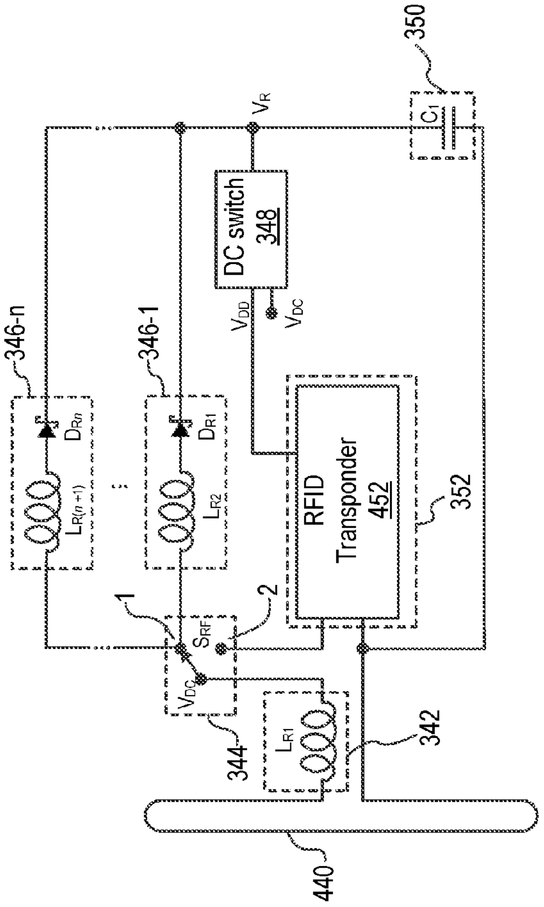
350
C_1
V_R
DC switch
348
V_{DD}
V_{DC}
346-n
346-1
D_{Rn}
$L_{R(n+1)}$
D_{R1}
L_{R2}
RFID
Transponder
452
352
1
2
S_{RF}
V_{DC}
344
L_{R1}
342
440

FIG.4B

FIG.5

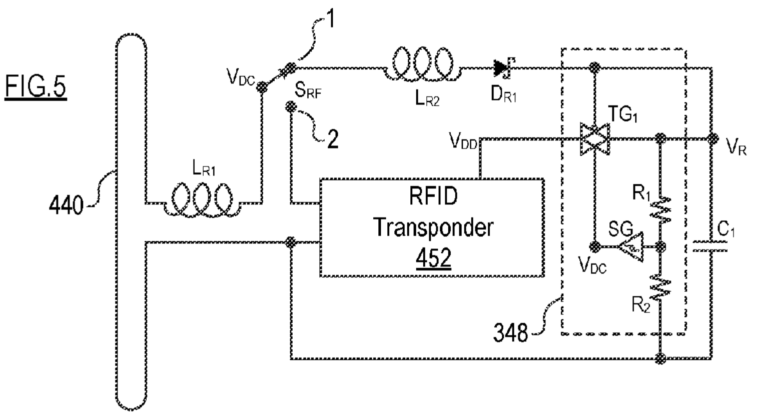
440
LR1
VDC
1
SRF
2
LR2
DR1
VDD
RFID
Transponder
452
TG1
VR
R1
SG
VDC
C1
R2
348

FIG.7

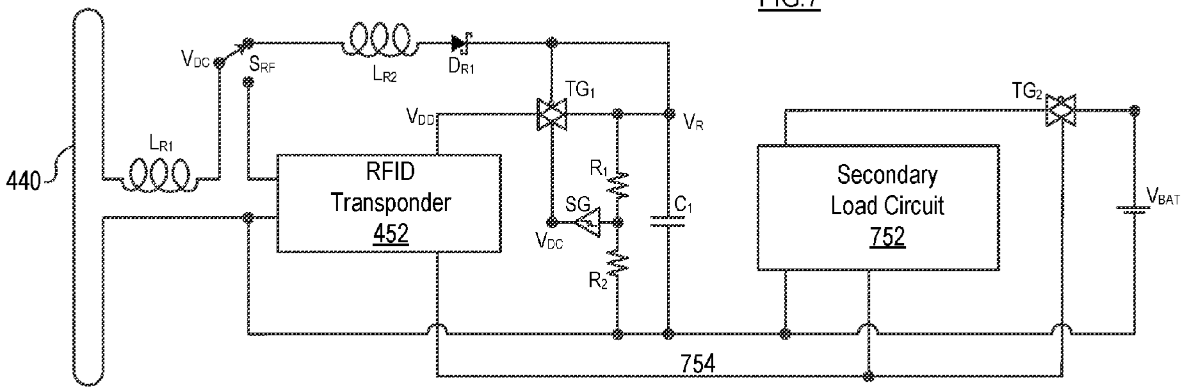
440
LR1
VDC
SRF
LR2
DR1
VDD
RFID
Transponder
452
TG1
VR
R1
SG
VDC
C1
R2
754
Secondary
Load Circuit
752
TG2
VBAT

FIG.6

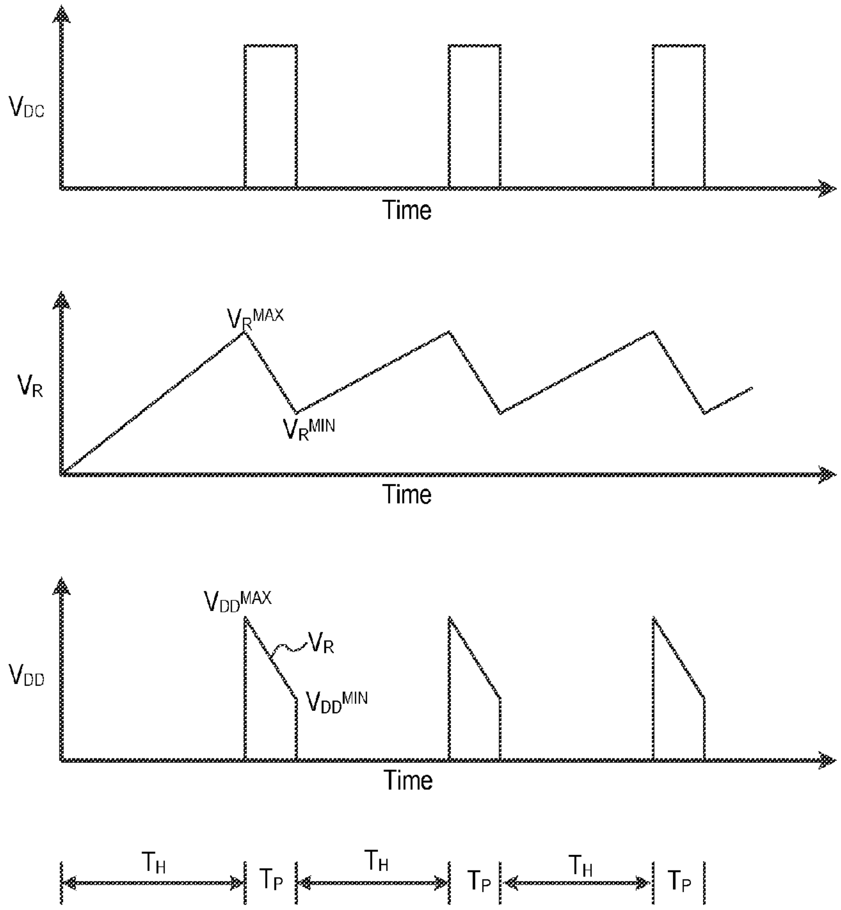
V_{DC}
Time
V_R
V_R^{MAX}
V_R^{MIN}
Time
V_{DD}
V_{DD}^{MAX}
V_R
V_{DD}^{MIN}
Time
T_H
T_P
T_H
T_P
T_H
T_P

FIG.8

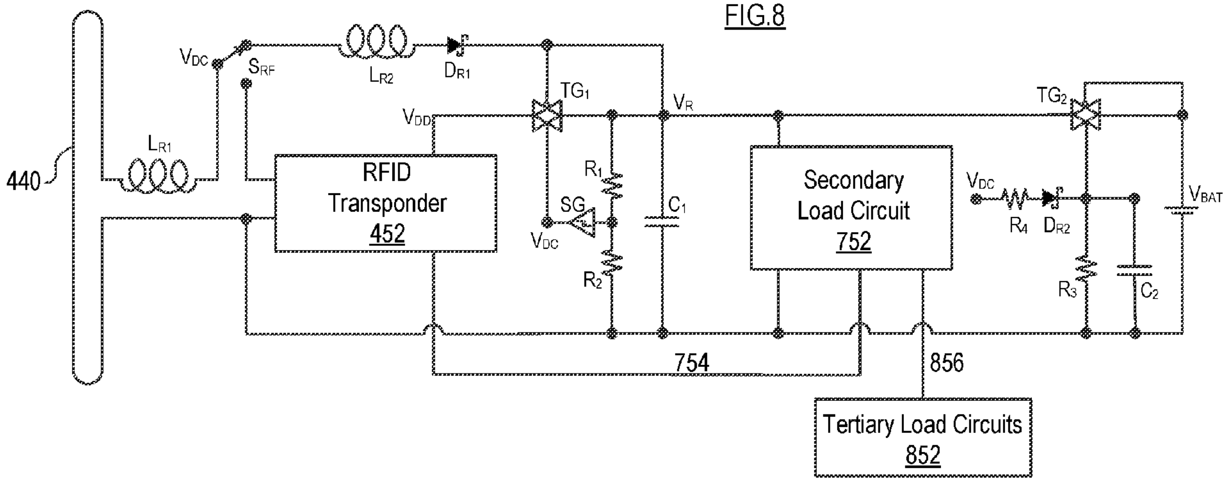
440
L_{R1}
V_{DC}
S_{RF}
L_{R2}
D_{R1}
V_{DD}
RFID Transponder 452
TG_1
R_1
SG
V_{DC}
R_2
C_1
V_R
Secondary Load Circuit 752
754
856
Tertiary Load Circuits 852
V_{DC}
R_4
D_{R2}
TG_2
R_3
C_2
V_{BAT}

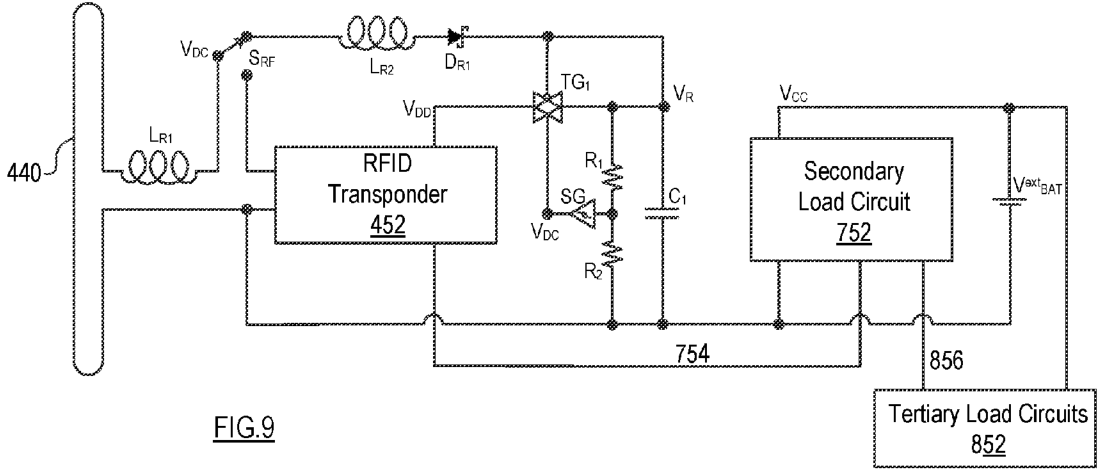
440
L_{R1}
V_{DC}
S_{RF}
L_{R2}
D_{R1}
V_{DD}
RFID Transponder 452
TG_1
R_1
SG
V_{DC}
R_2
C_1
V_R
V_{CC}
Secondary Load Circuit 752
V^{ext}_{BAT}
754
856
Tertiary Load Circuits 852

FIG.9

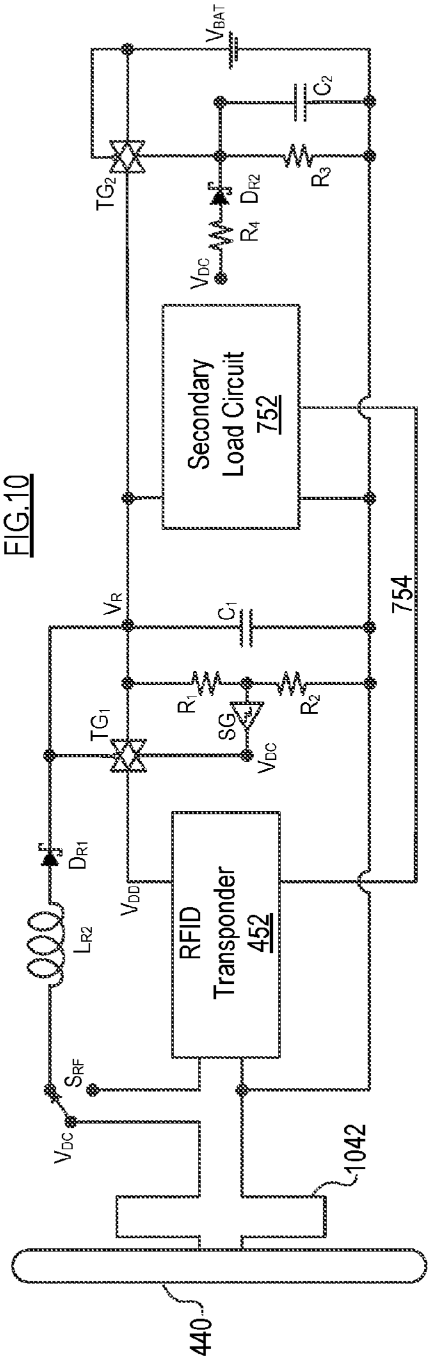
FIG.10
V_{BAT}
C_2
TG_2
D_{R2}
R_3
R_4
V_{DC}
Secondary Load Circuit 752
V_R
C_1
754
R_1
R_2
SG
TG_1
V_{DC}
D_{R1}
V_{DD}
RFID Transponder 452
L_{R2}
S_{RF}
V_{DC}
1042
440

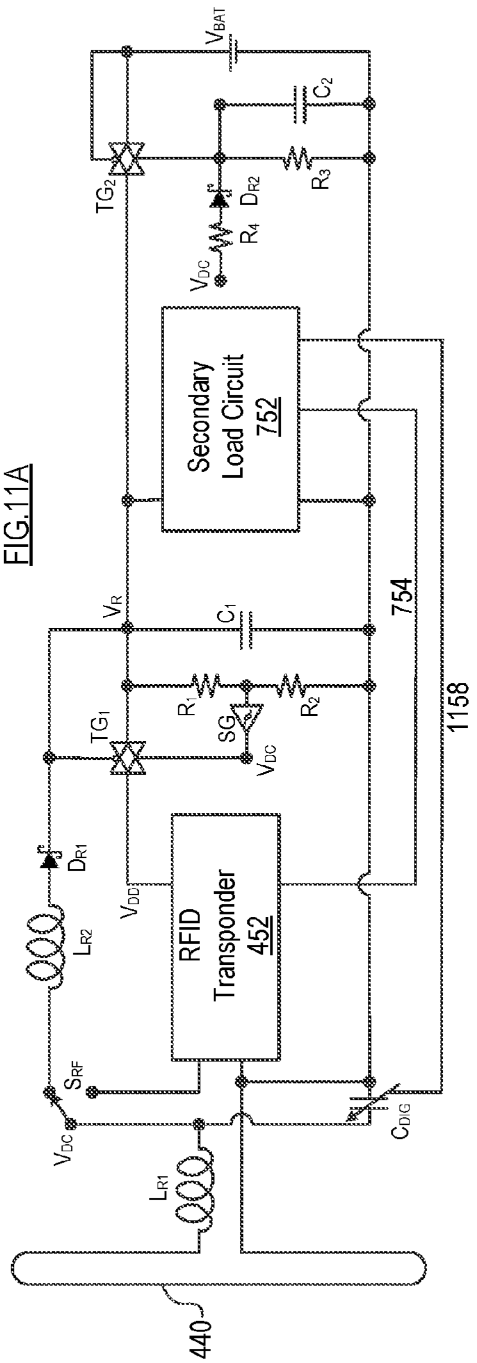
FIG.11A
V_{BAT}
C_2
TG_2
D_{R2}
R_3
R_4
V_{DC}
Secondary Load Circuit 752
V_R
C_1
754
R_1
R_2
SG
TG_1
V_{DC}
1158
D_{R1}
V_{DD}
RFID Transponder 452
L_{R2}
S_{RF}
V_{DC}
C_{DIG}
L_{R1}
440

FIG.11B

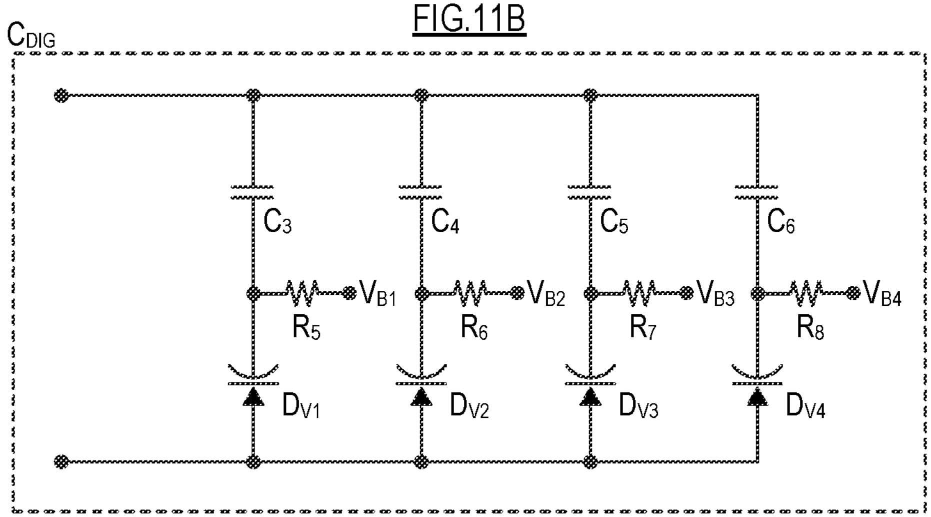
C_{DIG}
C_3
C_4
C_5
C_6
V_{B1}
V_{B2}
V_{B3}
V_{B4}
R_5
R_6
R_7
R_8
D_{V1}
D_{V2}
D_{V3}
D_{V4}

SENSOR SYSTEM WITH ENERGY HARVESTING

STATEMENT OF RELATED CASES

[0001] This case claims priority of U.S. Pat. App. Ser. No. 62/388,989 filed Feb. 16, 2016 and which is incorporated herein by reference.

FIELD OF THE INVENTION

[0002] This invention relates generally to harvesting environmental energy, and more particularly to harvesting energy from RF signals for powering very low power sensors.

BACKGROUND OF THE INVENTION

[0003] In many sensor applications, the sensors operate at very low power levels—as low as microwatts and even nanowatts. Such applications often require that the sensors be incorporated into wireless networks and operate with minimal power supplies. Power supplies that include energy-harvesting technology compatible with the aforementioned ultra-low power levels are, in some cases, used to provide power in these applications.

[0004] FIG. **1** depicts an energy-harvesting system disclosed in U.S. Pat. No. 7,084,605. Energy-harvesting system **100** includes energy-harvesting antenna **102**, inductive element **104**, variable capacitive element **106**, rectifier **108**, intermediate storage capacitor **110**, switch **112**, control circuitry **116**, DC energy output **114**, final energy-storage device **118**.

[0005] Energy-harvesting system **100** operates to charge final energy-storage device **118**. Inductive element **104** and variable capacitive element **106** form a tuning circuit to maximize the amount of energy harvested from the environment. Rectifier **108** converts the RF or AC energy received from antenna **102** into a DC energy, which is stored in capacitor **110**. Switch **112** controls the flow of DC energy **114** from storage capacitor **110** to final energy storage device **118**. The switch provides energy to operate electronic devices at fixed time intervals while also storing any additional unused energy on final energy storage device **118**. Switch **112** also functions to provide an input to control circuitry **116** to tune circuit elements **104** and **106** for best energy harvesting.

[0006] FIG. **2** depicts an energy-harvesting system for use with a wireless sensor, as disclosed in U.S. Pat. No. 8,552,597. Energy-harvesting system **200** includes energy-harvesting antenna **220**, super capacitor **222**, DC/DC converter **224**, battery and charger **226**, and power switch **228**. System **200** is directly connected to sensor node **230** and communications antenna **232**. The system can be bandpass or all-pass filtered to match all AM frequencies to maximize the capture of harvestable energy to the extent possible. Supercapacitor **222** provides a buffer for rechargeable battery **226**. Harvested energy can be used to power sensor node **230** or charge the sensor node's local power source **226**. The system is disclosed to provide power of over 10 mW to load devices.

SUMMARY

[0007] The present invention provides a source of switched DC power for a variety of sensor and control systems, including, without limitation, biomedical devices and remote and/or portable environmental sensors.

[0008] The illustrative embodiment of the invention is a sensor system including a primary load circuit with energy harvesting. In the illustrative embodiment, the primary load circuit is an RFID transponder, which can be structured for operation over a frequency range from about 100 kHz up to about 40 GHz and in accordance with any number of protocols including IEEE 802.15.4, ISO 18000, IEEE 802.15.1 (Bluetooth), and IEEE 802.11 (Wi-Fi). In one embodiment, the system is used to wake-up and/or synchronize a battery-powered UHF communication node within a network having a plurality of such nodes. By virtue of the invention and, in particularly, the manner in which energy harvesting is implemented, the distance between a transponder (included in the sensor system) and a remote interrogator is greatly increased compared to the prior art.

[0009] In accordance with an illustrative embodiment, the system includes an antenna, an impedance-matching network, an RF switch, inductive and diode components (for forming a resonant RF-to-DC voltage multiplication circuit in conjunction with other elements of the system), a DC switch, an energy-storage device, and a primary load circuit.

[0010] In accordance with the illustrative embodiment and unlike the prior art, the same antenna is used for both harvesting RF energy and for communications (e.g., RFID transponder communications, etc.). This reduces the size of the sensor system. In some embodiments, the antenna is a multi-band antenna.

[0011] In operation of some embodiments of the invention, incident RF energy that excites the antenna is coupled, at different times, into one of: (1) the energy storage device (e.g., capacitor, rechargeable battery, etc.) or (2) the primary load circuit. The destination for the harvested energy depends upon the state of the system. More particularly, the RF switch operates to direct the harvested energy to the energy storage device until a pre-determined "high" threshold is exceeded. In the illustrative embodiment, this high threshold is a specified voltage measured across the energy storage device.

[0012] Energy harvesting is implemented via a resonant voltage multiplication circuit. This circuit enables energy to be stored at a voltage level higher than is possible without the multiplier. The resonant voltage multiplication circuit permits harvesting charge in nano and microJoule increments. Typical circuits providing an RF-to-DC voltage multiplier include a diode rectifier and a DC-to-DC up-converter. These circuits do not operate at nanoWatt power levels. The voltage multiplication circuit is a resonant LC loop, wherein the capacitance of a semiconductor rectifying diode determines the resonant frequency of the voltage multiplier. This same diode, which in the illustrative embodiment is a Schottky diode, provides the unidirectional charging current to the energy storage element.

[0013] When the high voltage threshold is exceeded, in some embodiments, the DC switch enables harvested energy from the energy storage device to flow to the primary load circuit. Also, the state of the RF switch changes so that voltage multiplication is disconnected and the primary load circuit is coupled to the antenna for communications. The DC switch continues to electrically couple the energy storage device to the primary load circuit until the energy available from the energy storage device drops below a "low" voltage threshold. When the voltage drops below this low threshold, the RF switch and DC switch change state to enable energy harvesting and disconnect the primary load circuit from the antenna. Thus, in the illustrative embodiment, the system cycles between these two modes of operation as a function of the voltage across the energy storage device.

[0014] The impedance-matching circuit ensures that the RF impedance presented to the antenna by the voltage multiplier and the impedance presented to the antenna by the primary load circuit is the same. Impedance matching provides the maximum RF power transfer between the antenna and the primary load circuit, therefore providing a maximum range for the separation distance between, for example, an RFID transponder (as an example of the primary load circuit) and an RFID interrogator.

[0015] In some other embodiments, the sensor system includes one or more of the following:

- **[0016]** secondary load circuit(s) and tertiary load circuit(s), such as and without limitation, additional sensors (such as sensors for temperature, humidity, electrical conductivity, corrosion, media permittivity, inertial motion, fluid acidity, heartbeat rate, breath rate, magnetic field, gravitational vector force, and localized imaging), microprocessors, Internet connections, optical communications links, and actuatable control devices;
- **[0017]** additional control circuits;
- **[0018]** one or more batteries;
- **[0019]** devices/circuits that are operable to tune the operation at different frequencies; and
- **[0020]** devices/circuits that harvest RF energy at frequencies in addition to the operational RFID communications frequency.

[0021] In addition to its dual-purpose antenna, embodiments of the inventions include one or more of the following innovations, among any others:

- **[0022]** Power is supplied to the primary load circuit only when the energy that is harvested reaches a threshold level. In other words, the energy harvesting circuit and the primary load circuit are electrically coupled to the antenna at different times. This eliminates the undesirable loading effect of the energy harvesting circuit on the antenna that would otherwise occur if these two circuits were connected to the antenna at the same time. This enables the primary load circuit—the RFID transponder in the illustrative embodiment—to operate with lower ambient RF power levels.
- **[0023]** The energy-harvesting loop provides resonant, rectifying voltage multiplication. The resonant characteristic of the harvesting loop permits direct charging of the energy storage device to a higher voltage than is possible with a non-resonant loop. Not used in prior-art RFID systems, this permits charging of the energy storage device with incident RF power levels that are an order of magnitude smaller than required to operate an RFID transponder.
- **[0024]** In this regard, the prior-art system depicted in FIG. **1** includes a resonant circuit, but this is a parallel, resonant-circuit connection that does not enable directly charging the final energy storage device **118** to a voltage greater than the RF voltage provided by antenna **102**. Again, in embodiments of the present invention, the series resonant loop circuit including the resonant voltage multiplier is capable of charging the energy storage device to a voltage level that is greater than the RF voltage received from the antenna.
- **[0025]** The prior-art system depicted in FIG. **2** depicts a DC/DC converter **224**, which increases the DC voltage available from capacitor **222**. This converter is generally used to step-up the harvested DC voltage. However, such converters do not operate at the power levels of interest for embodiments of the present invention. In particular, typical RFID transponders operate at microwatt power levels. The inventor's use of the resonant voltage multiplier enables embodiments of the present invention to harvest RF power at nanoWatt levels, as enabled by the increased RF-to-DC voltage conversion.

[0026] In some embodiments, the sensor system harvests RF power from a single RF source, such as an RFID interrogator, at a single frequency. In some other embodiments, the sensor system is powered from multiple RF sources at multiple frequencies. In embodiments in which power is obtained from multiple RF sources, the sensor system includes a multi-frequency antenna, such as is disclosed in U.S. Pat. Nos. 8,581,793, 9,160,079, and 9,160,070, and matched resonant voltage multiplication circuits, that incrementally charge a single energy storage device (e.g., capacitor, etc.). For example, in one embodiment, the sensor system is configured with resonant harvesting circuits for multiple frequencies such as 1 MHz, 98 MHz and 915 MHz, whereas the primary load circuit, embodied as an RFID transponder, operates only within the 915 MHz band. The sensor system, however, is capable of harvesting power from all three wavelength bands to power the RFID transponder.

[0027] It will often be the case for such embodiments that energy is harvested from an RF energy source that, although readily available as a harvesting source, is not at the preferred frequency of the primary load circuit (e.g., not at the frequency preferred for RFID interrogation). In some of such embodiments, one or more harvesting circuits are provided to charge the energy storage device more or less continuously from ambient RF energy in the environment, such as a commercial AM, FM or TV broadcast station. In some embodiments, the energy storage device is maintained at a substantially "full" level of charge as a result of accumulated harvesting. The primary load circuit (e.g., RFID transponder) may operate in any of several available unlicensed frequency bands such as 125 kHz, 13.56 MHz, 860-960 MHz, or 2.45 GHz while the energy harvester is powered from other RF sources such as AM, FM or TV broadcast stations. In embodiments of the present invention, the energy harvesting circuits do not load the micro-power RFID transponder when it is enabled.

BRIEF DESCRIPTION OF THE DRAWINGS

[0028] FIG. **1** depicts a first prior-art energy harvesting system.

[0029] FIG. **2** depicts a second prior-art energy harvesting system.

[0030] FIGS. **3**A and **3**B depict a block diagram of a sensor system in accordance with an illustrative embodiment of the present invention.

[0031] FIG. **4**A depicts an embodiment of the sensor system of FIGS. **3**A/**3**B.

[0032] FIG. **4**B depicts an embodiment of the sensor system of FIGS. **3**A/**3**B wherein plural harvester circuits tuned for different frequencies and all charging the same energy storage device simultaneously are included.

[0033] FIG. **5** depicts an embodiment of the sensor system of FIGS. **3**A/**3**B wherein RF switching is controlled by a Schmitt gate and a transfer gate.

[0034] FIG. **6** depicts the control voltage and functions of sensor system of FIG. **5**.

[0035] FIG. **7** depicts an embodiment of the sensor system of FIGS. **3**A/**3**B with a secondary load circuit and programmed control connection to a battery power source.

[0036] FIG. **8** depicts an embodiment of the sensor system of FIGS. **3**A/**3**B with secondary load circuits and a delay connection to a battery power source for a timed interval.

[0037] FIG. **9** depicts an embodiment of the sensor system of FIGS. **3**A/**3**B with secondary load circuits and external devices wherein a battery power source is connected to the secondary load circuits under programmed control from the first load circuit.

[0038] FIG. **10** depicts an embodiment of the sensor system of FIGS. **3**A/**3**B wherein the impedance matching network is a T-match filter.

[0039] FIG. **11**A depicts and embodiment of the sensor system of FIGS. **3**A/**3**B wherein the RF frequency for operation is tuned by a variable capacitance device under programmed control.

[0040] FIG. **11**B depicts an embodiment of the variable capacitance device used in the embodiment of FIG. **11**A.

DETAILED DESCRIPTION

[0041] The following terms and their inflected forms are explicitly defined for use in this disclosure and the appended claims:

[0042] "RFID device" means a communication sensor, operated in a passive, semi-passive, or active mode. The sensing function includes a means of self-identification and, in some embodiments, one or more environmental sensors, actuators, controllers, etc.

[0043] "Load circuit" means the portion of the sensor system that consumes power, such as, without limitation, integrated circuits, sensors, actuators, imagers, LEDs, and lasers.

[0044] "Semi-passive RFID" means an RFID device powered with a local DC source including harvested energy stored in an energy storage device and communicating with an external interrogator via means of reflected signal carrier.

[0045] FIGS. **3**A and **3**B depict a block diagram of sensor system **300** in accordance with the illustrative embodiment of the present invention. Sensor system **300** includes antenna **340**, impedance matching network **342**, RF switch **344**, Inductance and Diode components (partial RF-to-DC voltage multiplier circuit) **346**, DC switch **348**, energy storage device **350**, and primary load circuit **352**.

[0046] Sensor system **300** is characterized as having energy-harvesting circuit **336** and load-enabling circuit **338**, with some elements in common. Energy-harvesting circuit (or RF-to-DC voltage multiplier) **336** includes antenna **340**, impedance matching network **342**, RF switch **344**, inductance and diode components **346**, and energy storage device **350**. Load-enabling circuit **338** includes antenna **340**, impedance matching network **342**, RF switch **344**, DC switch **348**, energy storage device **350** and primary load circuit **352**.

[0047] Sensor system **300** receives RF power into impedance matching network **342** through antenna **340**. The RF voltage, which is output from impedance matching network **342**, is directed to either (the rest of) energy-harvesting circuit **336** or (the rest of) load-enabling circuit **338**, depending on the position of RF switch **344**. The position of the RF switch is a function of the state of sensor system **300**; in particular, the charge level of energy storage device **350**.

[0048] When RF switch **344** is in a first position, it enables energy harvesting. In this mode of operation, RF energy received through the antenna from a remote source charges energy storage device **350** until voltage V_R across its terminals exceeds a "high" threshold voltage level, as monitored by DC switch **348**. When the high voltage threshold is exceeded, the mode of operation changes; primary load circuit **352** is enabled via DC switch **348**. Specifically, the DC switch electrically couples energy storage device **350** to primary load circuit **352**, thereby supplying DC power thereto. This enables primary load circuit **352** for operation with the voltage $V_{DD}=V_R$. Also, during this same switching interval, RF switch **344** changes to a second position disconnecting the antenna from the energy harvesting circuit and connecting it to the primary load circuit.

[0049] FIGS. **4**A and **4**B depict an embodiment of sensor system **300** wherein antenna **340** is implemented by dipole antenna **440**, impedance matching network **342** is implemented via inductor L_{R1}, RF switch **340** is implemented by single-pole double-throw (SPDT) switch S_{RF}, groupings of inductive and diode components **346-1** to **346**-*n* are each implemented via an inductor and a Shottky diode (i.e., L_{R2} and D_{R1} through L_{Rn+1} and D_{Rn}, respectively), energy storage device **350** is implemented by capacitor C_1, and primary load circuit **352** is implemented as RFID transponder **452**.

[0050] The inductive and diode components **346-1** through **346**-*n* tune the respective harvesting circuits providing RF harvesting at n different frequencies such as a UHF waveband and a lower HF or LF band. The selected frequency bands should be adequately separated to avoid cross talk interference. It is within the capabilities of those skilled in the art to select appropriately separated frequencies to avoid cross talk.

[0051] In some embodiments, RF switch S_{RF} comprises a GaAs JFET transistor and a silicon NPN transistor each controlling the two arms of the SPDT switch. The JFET provides a low resistive impedance source-drain circuit path with zero gate-voltage bias for RF energy harvesting.

[0052] RF switch S_{RF}, at position "1" as depicted in FIGS. **4**A and **4**B, enables energy harvesting. The energy harvesting loop includes inductive RF-to-DC voltage multiplication, which provides significant benefits over the prior art. RF power received by antenna **440** is partially rectified by the Schottky diode(s). The energy-harvesting circuit is an RLC resonant loop; the capacitance of Schottky diode(s) resonates with the series loop inductance provided by (in FIG. **4***a*) inductors L_{R1}, L_{R2}, and antenna **440**. The harvesting loop resonates at the frequency of the external RF power source (e.g., RFID interrogator, etc.). The resonant characteristic of the harvesting loop enables charging the energy storage device to a higher voltage than is possible with a non-resonant circuit.

[0053] FIG. **5** depicts an embodiment of sensor system **300** wherein DC switch **348** is embodied as transfer gate TG_1, Schmitt gate SG, and resistors R_1 and R_2. In this embodiment, and all other embodiments depicted herein, RF switch S_{RF} is controlled by the DC switch. It is important to note that, due to this arrangement, primary load circuit **352**—RFID transponder **452** in the illustrative embodiments—does not receive DC power during those times when RF energy harvesting is enabled. Harvesting and transponder functions are enabled at different times through RF switch S_{RF}, which is itself controlled by the control voltage V_{DC}. In particular, when (harvested) voltage V_R (the voltage across the energy storage device) exceeds the high threshold level, control voltage V_{DC} enables the following:

[0054] (1) DC switch **348** connects the supply voltage V_R into RFID transponder **452** and actuates the RF switch S_{RF} from position **1** to position **2**.

[0055] (2) Due to the move from position **1** to position **2**, the energy harvesting circuit(s) are disabled.

[0056] (3) Due to the move from position **1** to position **2**, antenna **440** and its impedance matching network are connected to RFID transponder **452**.

RFID transponder **452** is made operational through this switching action and a wireless communication link is established with a remote RFID interrogator (not depicted), through antenna **440**.

[0057] The DC threshold sensing function is provided by DC switch **348**. With RF switch S_{RF} in position "**1**" for energy harvesting, the resonant voltage multiplier is enabled and capacitor C_1 charges. Energy harvesting is disabled when the voltage V_R across capacitor C_1 exceeds the pre-determined high threshold level, as determined by the voltage divider connection R_1 and R_2. When the voltage exceeds the threshold, Schmitt gate SG goes to a high state, supplying control voltage V_{DC} to transfer gate TG_1 and RF switch S_{RF}. The enabled transfer gate TG**1** causes the voltage level V_{DD} across RFID transponder to increase from a minimal value up to approximately V_R (the voltage level of "fully" charged capacitor C_1). Control voltage V_{DC} causes RF switch S_{RF} to move to position "**2**". There is a hysteresis in voltage V_{DC} provided from Schmitt gate SG that enables the direction connection $V_R=V_{DD}$ throughout the hysteresis range of the Schmitt gate.

[0058] Thus enabled, RFID transponder **452** continues to operate until the voltage V_R falls below a pre-determined lower threshold voltage, indicating that the energy storage device—capacitor C_1 in the illustrative embodiment—has drained to the point that it cannot sustain the operation of the RFID transponder. Schmitt gate SG responds when the voltage V_R falls below the lower voltage threshold and the control voltage V_{DC} goes low. The transfer gate TG_1 then stops the flow of energy to the primary load circuit—RFID transponder **452** in the illustrative embodiment—and causes RF switch S_{RF} to move to position **1**, thereby disconnecting the RFID transponder and re-establishing the energy harvesting circuit. The charging cycle then repeats, etc. In some other embodiments, voltage V_{DC} is supplied directly as V_{DD} into the primary load circuit **352** and transfer gate TG_1 is not used.

[0059] The high and low threshold voltages are determined from calibrations and designed into the DC switch circuit. A typical value for the high threshold is about 3.5 volts and a typical value for the low threshold is about 1.5 volts.

[0060] FIG. **6** depicts the sequence of harvesting operation T_H and RFID transponder operation T_P. The plot depicting voltage V_R across the energy storage device shows voltage V_R rising to a maximum V_R^{MAX} during the energy harvesting operation. Once the "high" threshold voltage is exceeded, control voltage V_{DC} goes high and stored energy from the energy storage device is directed to the RFID transponder. The plot depicting voltage V_{DD} across the transponder shows voltage V_{DD} at its peak value V_{DD}^{MAX} when the transponder is first switched in. As the energy storage device drains, supplying energy to the primary load circuit transponder, voltage V_{DD} falls along with voltage V_R. When the voltage across the energy storage device (i.e., capacitor C_1) falls past the low threshold voltage value V_R^{MIN}, at which point the voltage across the transponder reaches its minimum value V_{DD}^{MIN}, control voltage VDC goes low and energy harvesting operation T_H resumes.

[0061] FIG. **7** depicts an embodiment of sensor system **300** that includes secondary load circuit **752** and a data bus or control line **754** from RFID transponder **452** to the secondary load circuit. In this embodiment, secondary load circuit **752** is enabled via second transfer gate TG_2 and controlled from the RFID transponder. Second load circuit is enabled typically only during operation of the RFID transponder and thus energy drain from battery power source V_{BAT} is minimal. This embodiment is particularly useful for applications in which the available battery is very small and it is desirable to conserve its stored energy.

[0062] FIG. **8** depicts an embodiment of sensor system **300** that provides a delayed connection to battery power source V_{BAT} for a timed interval. In this embodiment, Schmitt gate SG provides a control voltage V_{DC} to control second transfer gate TG_2. The control circuit electrically couples battery V_{BAT} for powering both RFID transponder **452** and second load circuit **752** for a fixed interval of time. The circuit comprised of resistors R_3, R_4, capacitor C_2 and diode D_{R2} provide the voltage which enables the transfer gate TG_2 supplying the battery voltage V_{BAT} to both load circuits for a fixed time interval. This embodiment is particularly useful when RFID transponder **452** is used to control loads, such as secondary load circuit **752** (via data bus **754**) and tertiary load circuits **852** (via data and power bus **856**) requiring additional power and wherein the operational interval of the RFID transponder is extended in time.

[0063] FIG. **9** depicts an embodiment of sensor system **300** wherein external power source V^{ext}_{BAT} is continually supplying power to secondary load circuit **752**. The secondary load circuit is controlled via RFID transponder **452** over data bus **754**. In this embodiment, tertiary load circuits **852**, which are electrically connected to secondary load circuit **752** via data and power bus **856** can draw watts or even kilowatts from power source V^{ext}_{BAT}.

[0064] FIG. **10** depicts an embodiment of sensor system **300** in which impedance matching network **342** is implemented as T-match filter **1042**. This embodiment is suitable for UHF sensors (operating at frequencies in the range of 300 MHz to 3 GHz).

[0065] T-match filter **1042** can be made by patterning of metallization on printed circuit boards and flexible substrates. A typical T-match filter has an equivalent inductive impedance component that is compatible with the use of the impedance element (e.g., L_{R2}, etc.) in the RF-to-DC voltage multiplication circuit.

[0066] FIG. **11**A depicts an embodiment of sensor system **300** in which the RF frequency for operation is tuned by variable capacitance device C_{DIG} under the control of secondary load circuit **752** over data bus **1158**. In this embodiment, a controlled capacitance is connected across the output of impedance matching network L_{R1} to provide a desired impedance match to antenna **440**/impedance matching network L_{R1}. Small incremental changes in the capacitance of variable capacitance device C_{DIG} shift the resonant frequency of sensor system **300**. In some other embodiments, variable capacitance device C_{DIG} is under the autonomous control of RFID transponder **452**. In yet some additional embodiments, a remote RFID interrogator controls the capacitance through RFID transponder **452**.

[0067] FIG. **11**B depicts an embodiment of C_{DIG} for providing a variable capacitance. In this embodiment, capacitance is controlled with 4-bit digital data bus **1158**. In this circuit, each control bit changes the capacitance of respective varactor diodes D_{V1}, D_{V2}, D_{V3}, D_{V4}. The capacitances of the varactor diodes are much smaller than that of respective coupling capacitors C_3, C_4, C_5 and C_6. Respective resistors R_5, R_6, R_7, and R_8 have impedances that are much higher than the reactance of capacitors C_3 through C_6 and do not load the variable capacitor circuit. The sum of the parallel connection of varactor-diode capacitors biased under programmed bus control provides variable capacitor C_{DIG} that tunes the wireless sensor.

[0068] It is to be understood that although this disclosure teaches many examples of embodiments in accordance with the present teachings, many additional variations of the invention can easily be devised by those skilled in the art after reading this disclosure. As a consequence, the scope of the present invention is to be determined by the following claims.

What is claimed:

1. A sensor system comprising:

an RF front end that receives and switches RF energy;

an energy-harvesting circuit that provides resonant RF-to-DC voltage multiplication, the energy-harvesting circuit including:

(a) the RF front end,

(b) inductive and capacitive elements to provide a partially rectifying, resonant LC loop, and

(c) an energy storage device; and

a load-enabling circuit that powers a primary load circuit from the energy storage device, the load-enabling circuit including:

(a) the RF front end,

(b) the primary load circuit, and

(c) the energy storage device.

2. The sensor system of claim **1** wherein the RF front end comprises:

an antenna that both receives RF energy from ambient sources thereof and transmits signals originating from the primary load circuit;

an impedance matching network that is coupled to the antenna; and

an RF switch that switches the received RF energy between the energy-harvesting circuit and the load-enabling circuit.

3. The sensor system of claim **1** wherein the inductive and capacitive elements comprise an inductor and a Schottky diode connected in series.

4. The sensor system of claim **1** wherein the energy storage device is a capacitor or a rechargeable battery.

5. The sensor system of claim **1** wherein the load-enabling circuit further comprises a DC switch, wherein the DC switch is electrically coupled to the energy storage device and the primary load circuit.

6. The sensor system of claim **5** wherein the DC switch comprises a transfer gate and a Schmitt gate.

7. The sensor system of claim **5** wherein the DC switch comprises a Schmitt gate and is configured so that when a voltage across the energy storage device exceeds a first threshold, the Schmitt gate goes to a high state generating a DC control voltage.

8. The sensor system of claim **7** wherein the RF front end comprises an RF switch that switches received RF energy between the energy-harvesting circuit and the load-enabling circuit, and further wherein, when the control voltage is generated by the Schmitt gate:

(a) a DC supply voltage from the energy storage device is coupled into the primary load circuit; and

(b) the RF switch changes state, disabling energy harvesting and electrically coupling the primary load circuit to the RF front end.

9. The sensor system of claim **1** wherein the load-enabling circuit further includes a secondary load circuit, wherein the primary load circuit controls the secondary load circuit.

10. The sensor system of claim **2** further comprising a variable capacitive device electrically connected to the impedance matching network, wherein the variable capacitive device controls a resonance frequency of the sensor system, thereby dictating which frequencies of RF energy are harvested by the sensor system.

11. The sensor system of claim **2** wherein the antenna is a multi-band antenna, the RF front end further comprising plural groups of resonant energy-harvesting inductive and capacitive elements, each group tuned for a frequency band of the multi-band antenna, said respective energy-harvesting circuits all electrically coupled to a common energy storage device.

12. The sensor system of claim **1** wherein the primary load circuit is a RFID transponder.

13. A method for operating a sensor system, the method comprising:

receiving, at an antenna, RF energy from an ambient source of RF energy;

harvesting the received RF energy in an energy storage device via a circuit that provides resonant voltage multiplication;

when a voltage across the energy storage device exceeds a first threshold:

(a) electrically coupling the energy storage device to a primary load circuit, enabling energy to flow from the energy storage device to the primary load circuit, and

(b) changing the state of an RF switch, wherein the change in state:

  (i) decouples the circuit that provides resonant voltage multiplication from the antenna, and

  (ii) electrically couples the primary load circuit to the antenna so that the primary load circuit can transmit a signal;

when the voltage across the energy storage device drops below a second threshold:

(a) electrically decoupling the primary load circuit from the energy storage device,

(b) changing the state of the RF switch, wherein the change in state:

  (i) decouples the primary load circuit from the antenna; and

(ii) electrically couples the circuit that provides resonant voltage multiplication to the antenna so that RF energy is harvested.

14. The method of claim **13** wherein the source of RF energy is an RFID transponder.

15. A method for operating a sensor system, the method comprising switching between an energy harvesting mode and a load-enabling mode as follows:

enabling the energy harvesting mode when a voltage measured across an energy storage device falls below a first threshold voltage, the energy harvesting mode including:

(a) processing RF energy received at an antenna via resonant voltage multiplication, and

(b) storing the processed RF energy in the energy storage device; and

enabling the load-enabling mode when the voltage measured across the energy storage device exceeds a second threshold voltage, the load-enabling mode including:

(a) electrically couple the energy storage device to a primary load circuit, and

(b) transmitting a signal from the primary load circuit via the antenna.

16. The method of claim **15** and further wherein enabling the load-enabling mode further comprises:

(c) disabling the processing of the received RF energy, and

(d) electrically coupling the antenna to the primary load circuit.

17. The method of claim **15** and further wherein enabling the energy-harvesting mode further comprises:

(c) decoupling the primary load circuit from the energy storage device,

(d) decoupling the primary load circuit from the antenna, and

(e) coupling the antenna to inductive and capacitive elements that collectively provide resonant voltage multiplication.

18. The method of claim **15** wherein processing RF energy further comprises receiving RF energy sourced from plural sources, each source radiating RF energy at a different frequency, wherein the antenna is a multi-frequency antenna and the RF energy at each different frequency is processed by frequency matched resonant voltage multiplication circuits.

19. The method of claim **15** wherein enabling the load-enabling mode further comprises generating, from a Schottky diode, a control voltage that causes a transfer gate to electrically couple the energy storage device to the primary load circuit.

20. The method of claim **15** wherein enabling the load-enabling mode further comprises controlling, from the primary load circuit, a secondary load circuit.

* * * * *

US 20170308783A1

(19) **United States**
(12) **Patent Application Publication**
Rokhsaz et al.
(10) **Pub. No.: US 2017/0308783 A1**
(43) **Pub. Date: Oct. 26, 2017**

(54) **METHOD AND APPARATUS FOR DETECTING RF FIELD STRENGTH**

(71) Applicant: **RFMicron, Inc.**, Austin, TX (US)

(72) Inventors: **Shahriar Rokhsaz**, Austin, TX (US); **Edwin de Angel**, Austin, TX (US)

(73) Assignee: **RFMicron, Inc.**, Austin, TX (US)

(21) Appl. No.: **15/643,962**

(22) Filed: **Jul. 7, 2017**

Related U.S. Application Data

(60) Continuation of application No. 14/644,471, filed on Mar. 11, 2015, now Pat. No. 9,704,085, which is a continuation of application No. 13/209,425, filed on Aug. 14, 2011, now Pat. No. 9,048,819, which is a continuation-in-part of application No. 12/462,331, filed on Aug. 1, 2009, now Pat. No. 8,081,043, which is a division of application No. 11/601,085, filed on Nov. 18, 2006, now Pat. No. 7,586,385.

(60) Provisional application No. 61/428,170, filed on Dec. 29, 2010, provisional application No. 61/485,732, filed on May 13, 2011.

Publication Classification

(51) **Int. Cl.**

G06K 19/07	(2006.01)
H04B 5/00	(2006.01)
H04B 17/318	(2006.01)

(52) **U.S. Cl.**

CPC ***G06K 19/0726*** (2013.01); ***H04B 17/318*** (2015.01); ***H04B 5/0062*** (2013.01)

(57) **ABSTRACT**

A radio frequency (RF) circuit comprises an antenna operably coupled to receive an RF signal, a power harvesting circuit operable to generate a first power source from the RF signal, a tank circuit coupled to the antenna and a tuning circuit. The tank circuit includes a selectively variable impedance and the tuning circuit is adapted to dynamically vary the selectively variable impedance of the tank circuit based on resonance of the RF circuit and frequency of the RF signal to substantially align the resonance of the RF circuit with the frequency of the RF signal. When the resonance of the RF circuit is substantially aligned with the frequency of the RF signal, the power harvesting circuit generates a second power source from the RF signal such that the second power source is greater than the first power source.

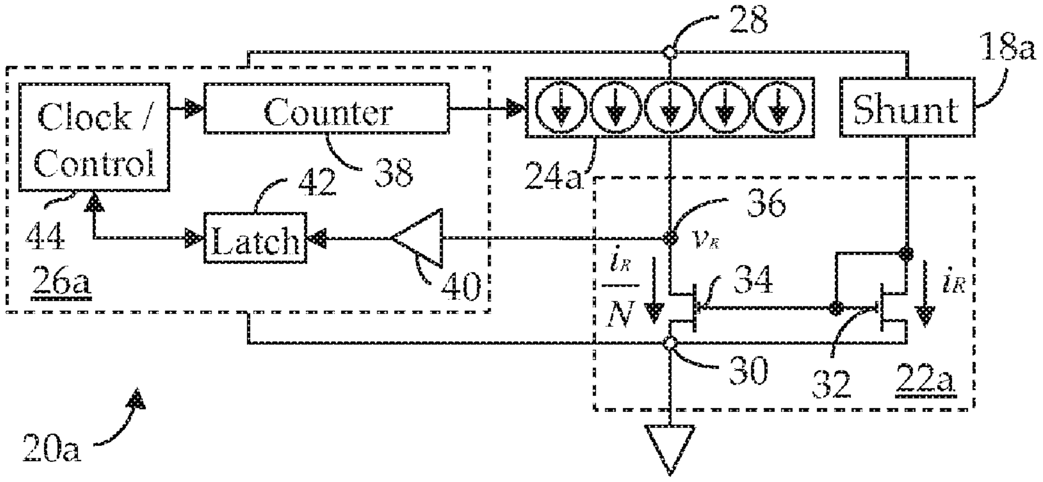

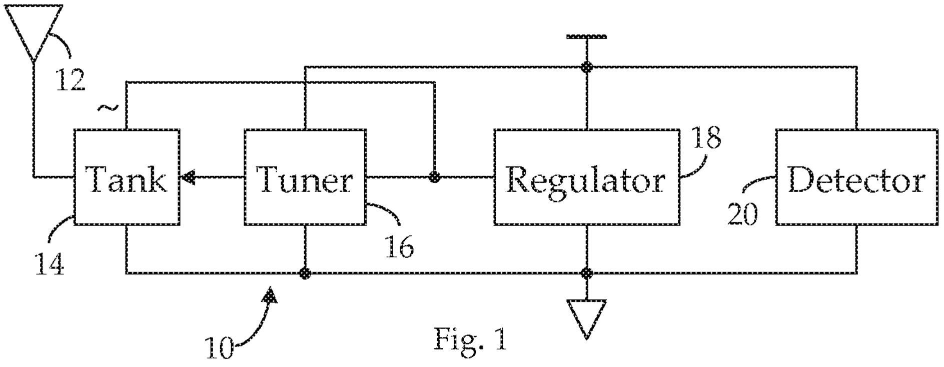
12
~
Tank
Tuner
Regulator
18
Detector
14
16
20
10

Fig. 1

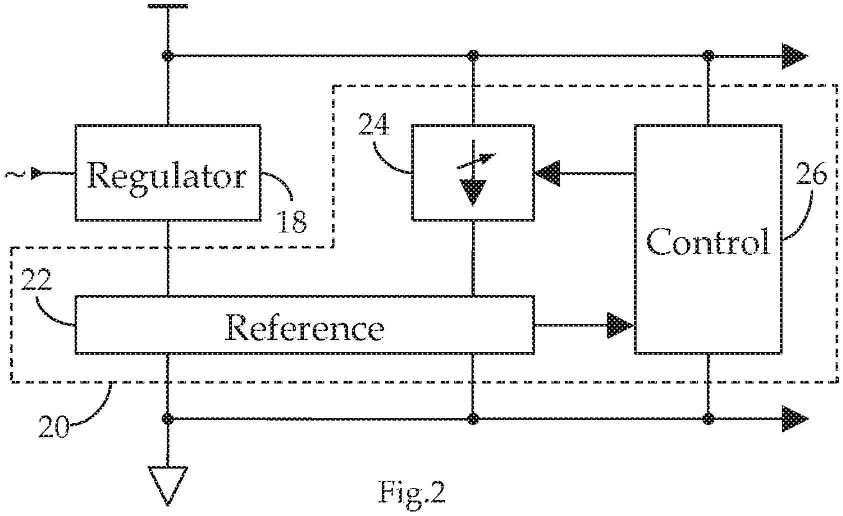
~
Regulator
18
24
Control
26
22
Reference
20

Fig.2

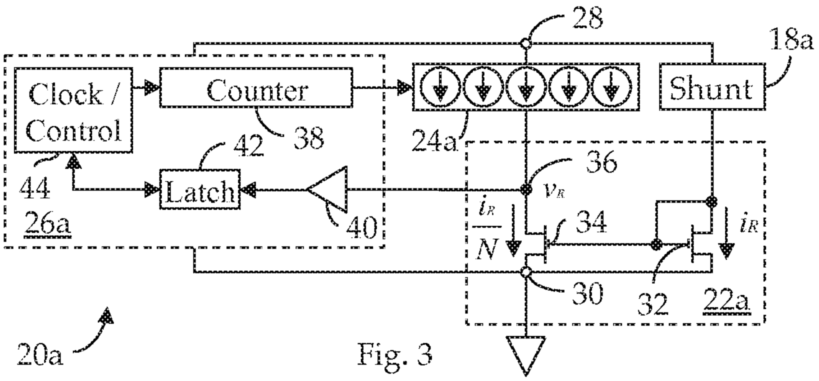
28
18a
Clock /
Control
Counter
Shunt
44
42
38
24a
36
26a
Latch
40
v_R
$\frac{i_R}{N}$
34
i_R
30
32
22a
20a

Fig. 3

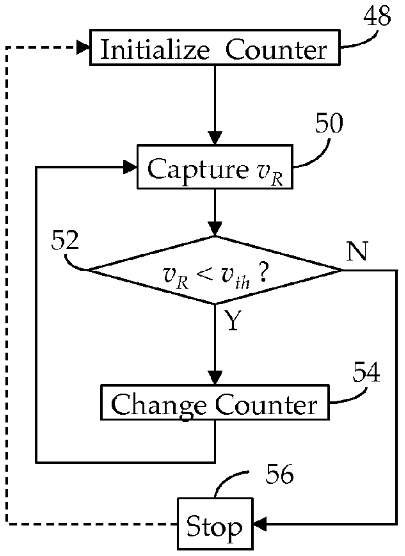
Initialize Counter
48
Capture v_R
50
52
$v_R < v_{th}$?
N
Y
Change Counter
54
56
Stop

Fig. 4

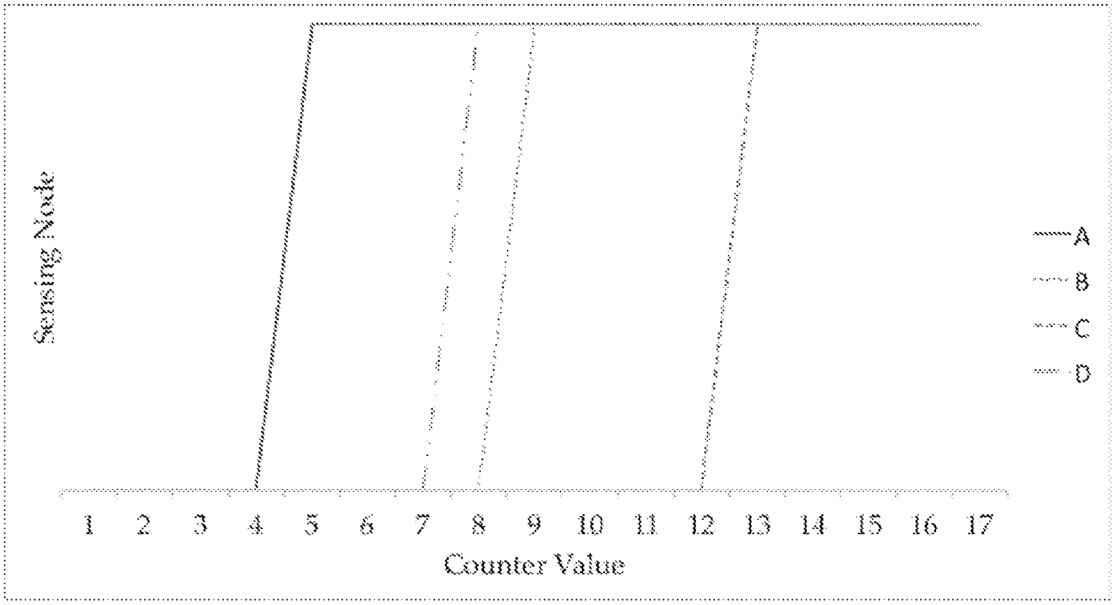
Sensing Node
A
B
C
D
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
Counter Value

Fig. 5

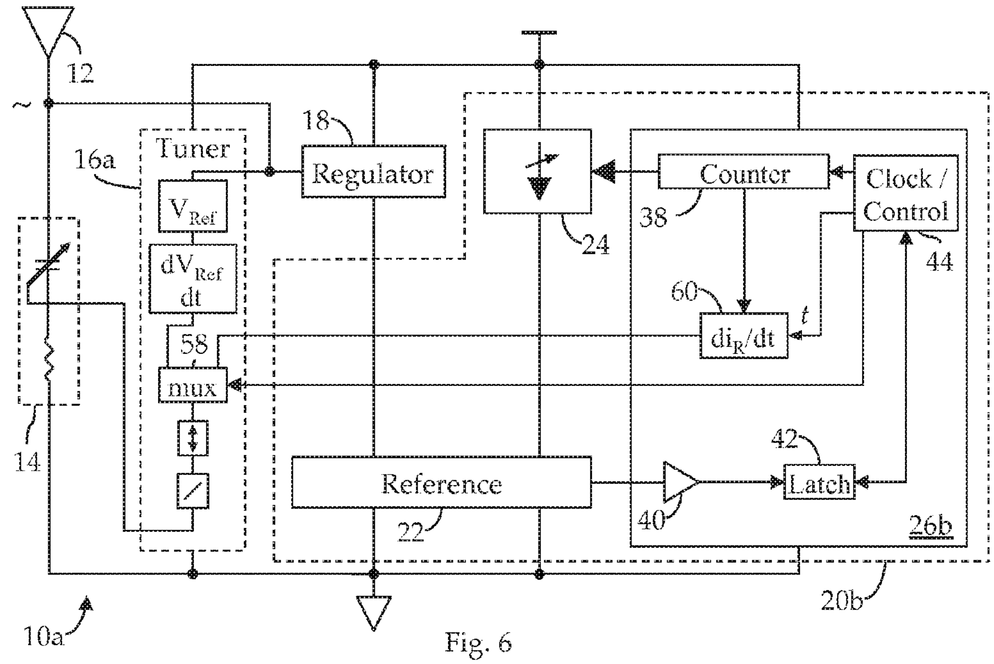
12
~
18
Tuner
16a
Regulator
Counter
Clock /
Control
V_{Ref}
38
24
44
dV_{Ref}
dt
60
t
58
di_R/dt
mux
14
42
Reference
Latch
40
22
26b
20b
10a

Fig. 6

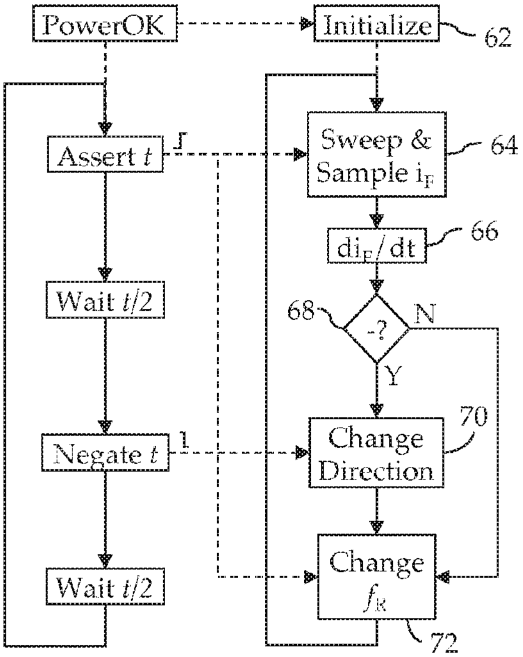
PowerOK
Initialize
62
Assert t
Sweep &
Sample i_F
64
di_F/dt
66
Wait t/2
68
-?
N
Y
Negate t
Change
Direction
70
Wait t/2
Change
f_R
72

Fig. 7

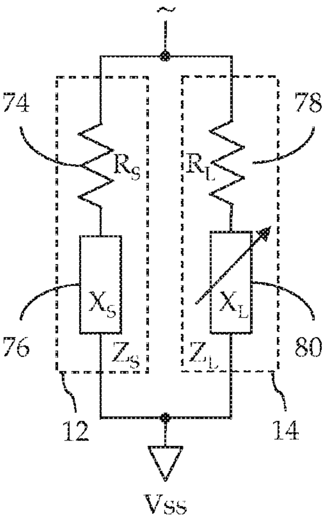
~
74
78
R_S
R_L
X_S
X_L
76
80
Z_S
Z_L
12
14
Vss

Fig. 8

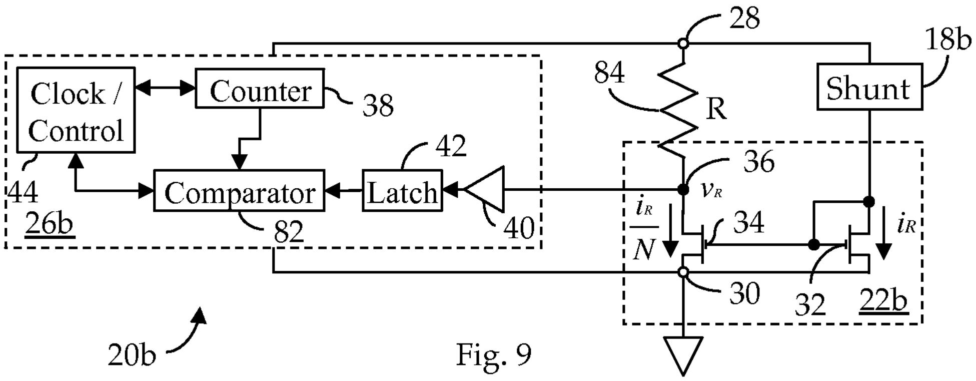
28
18b
Shunt
84
R
Clock / Control
Counter
38
42
Comparator
Latch
44
26b
82
40
36
v_R
$\frac{i_R}{N}$
34
i_R
30
32
22b
20b

Fig. 9

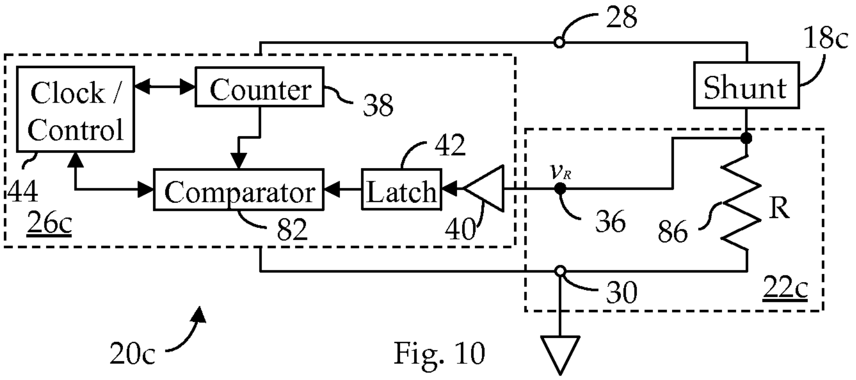
28
18c
Shunt
Clock / Control
Counter
38
42
Comparator
Latch
v_R
44
26c
82
40
36
86
R
30
22c
20c

Fig. 10

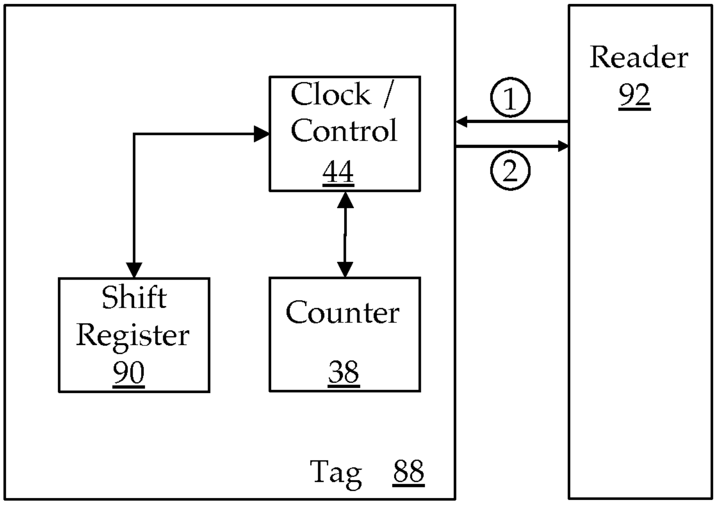
Clock / Control
44
①
②
Reader
92
Shift Register
90
Counter
38
Tag 88

Fig. 11

METHOD AND APPARATUS FOR DETECTING RF FIELD STRENGTH

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present U.S. Utility patent application claims priority pursuant to 35 U.S.C. §120 as a Continuation of U.S. Utility application Ser. No. 14/644,471, entitled "Method and Apparatus for Detecting RF Field Strength", filed 11 Mar. 2015, issuing as U.S. Pat. No. 9,704,085 on 11 Jul. 2017, which is a Continuation of U.S. Utility application Ser. No. 13/209,425, entitled "Method and Apparatus for Detecting RF Field Strength", filed 14 Aug. 2011, now U.S. Pat. No. 9,048,819, issued 2 Jun. 2015, which is a Continuation-In-Part of application Ser. No. 12/462,331, filed 1 Aug. 2009, now U.S. Pat. No. 8,081,043, issued 20 Dec. 2011 ("Related application"), which is in turn a Division of application Ser. No. 11/601,085, filed 18 Nov. 2006, now U.S. Pat. No. 7,586,385, issued 8 Sep. 2009 ("Related patent") (collectively, "Related References"). application Ser. No. 13/209,425 also claims priority pursuant to 35 U.S.C. §119(e) to U.S. Provisional Application No. 61/428,170, filed 29 Dec. 2010, and U.S. Provisional Application No. 61/485,732, filed 13 May 2011. The subject matter of the Related References, each in its entirety, is expressly incorporated herein by reference.

[0002] This application is related to application Ser. No. 13/209,420, filed on 14 Aug. 2011, now U.S. Pat. No. 8,749,319, issued 10 Jun. 2014 ("Related Co-application").

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0003] The present invention relates generally to detecting RF field strength, and, in particular, to detecting RF field strength in a passive RFID system.

2. Description of the Related Art

[0004] In general, in the descriptions that follow, we will italicize the first occurrence of each special term of art that should be familiar to those skilled in the art of radio frequency ("RF") communication systems. In addition, when we first introduce a term that we believe to be new or that we will use in a context that we believe to be new, we will bold the term and provide the definition that we intend to apply to that term. In addition, throughout this description, we will sometimes use the terms assert and negate when referring to the rendering of a signal, signal flag, status bit, or similar apparatus into its logically true or logically false state, respectively, and the term toggle to indicate the logical inversion of a signal from one logical state to the other. Alternatively, we may refer to the mutually exclusive Boolean states as logic_0 and logic_1. Of course, as is well known, consistent system operation can be obtained by reversing the logic sense of all such signals, such that signals described herein as logically true become logically false and vice versa. Furthermore, it is of no relevance in such systems which specific voltage levels are selected to represent each of the logic states.

[0005] In accordance with our prior invention previously disclosed in the Related References, the amplitude modulated ("AM") signal broadcast by the reader in an RFID system will be electromagnetically coupled to a conventional antenna, and a portion of the current induced in a tank circuit is extracted by a regulator to provide operating power for all other circuits. Once sufficient stable power is available, the regulator will produce, e.g., a power-on-reset signal to initiate system operation. Thereafter, the method disclosed in the Related References, and the associated apparatus, dynamically varies the capacitance of a variable capacitor component of the tank circuit so as to dynamically shift the f_R of the tank circuit to better match the f_C of the received RF signal, thus obtaining maximum power transfer in the system.

[0006] In general, the invention disclosed in the Related References focused primarily on quantizing the voltage developed by the tank circuit as the primary means of matching the f_R of the tank circuit to the transmission frequency, f_C, of the received signal. However, this voltage quantization is, at best, indirectly related to received signal field strength. We submit that what is needed now is an effective and efficient method and apparatus for quantizing the received field strength as a function of induced current. It is further desirable to develop this field quantization in a form and manner that is suitable for selectively varying the input impedance of the receiver circuit to maximize received power, especially during normal system operation. Additionally, in light of the power sensitive nature of RFID systems, it is desirable to vary the input impedance with a minimum power loss.

BRIEF SUMMARY OF THE INVENTION

[0007] In accordance with the preferred embodiment of our invention, we provide a sensing system for use in an RFID system. In general, the sensing system comprises an RFID tag and an RFID reader. In one embodiment, the RFID tag comprises a tank circuit having a selectively variable impedance; and a tuning circuit adapted to dynamically vary the impedance of the tank circuit, and to develop a first quantized value representative of the impedance of said tank circuit. In one alternate embodiment, the RFID tag comprises a detector circuit adapted to develop a second quantized value as a function of a field strength of a received RF signal. In yet another embodiment, the RFID tag comprises both the tank and tuning circuit, and the detector circuit. In these embodiments, RFID reader is adapted selectively to retrieve one or both of the first and second values, and, preferably, to use the retrieved values to sense changes to an environment to which the RFID tag is exposed

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[0008] My invention may be more fully understood by a description of certain preferred embodiments in conjunction with the attached drawings in which:

[0009] FIG. **1** illustrates, in block diagram form, an RF receiver circuit having a field strength detector constructed in accordance with an embodiment of our invention;

[0010] FIG. **2** illustrates, in block diagram form, a field strength detector circuit constructed in accordance with an embodiment of our invention;

[0011] FIG. **3** illustrates, in block schematic form, a more detailed embodiment of the field strength detector circuit shown in FIG. **2**;

[0012] FIG. **4** illustrates, in flow diagram form, the sequencing of operations in the field strength detector circuit shown in FIG. **3**;

[0013] FIG. **5** illustrates, in graph form, the response of the field strength detector circuit shown in FIG. **3** to various conditions;

[0014] FIG. **6** illustrates, in block schematic form, an RF receiver circuit constructed in accordance with another embodiment of our invention;

[0015] FIG. **7** illustrates, in flow diagram form, the sequencing of the operations in the RF receiver circuit shown in FIG. **6**;

[0016] FIG. **8** illustrates, in block schematic form, an alternative representation of the impedance represented by the antenna and the tank circuit of the exemplary RFID receiver circuit.

[0017] FIG. **9** illustrates, in block schematic form, an alternative exemplary embodiment of the field strength detector circuit shown in FIG. **3**.

[0018] FIG. **10** illustrates, in block schematic form, an alternative exemplary embodiment of the field strength detector circuit shown in FIG. **3**.

[0019] FIG. **11** illustrates, in block schematic form, an exemplary RFID sub-system containing tag and reader.

[0020] In the drawings, similar elements will be similarly numbered whenever possible. However, this practice is simply for convenience of reference and to avoid unnecessary proliferation of numbers, and is not intended to imply or suggest that our invention requires identity in either function or structure in the several embodiments.

DETAILED DESCRIPTION OF THE INVENTION

[0021] Shown in FIG. **1** is an RF receiver circuit **10** suitable for use in an RFID application. As we have described in our Related References, an RF signal electromagnetically coupled to an antenna **12** is received via a tank circuit **14**, the response frequency, f_R, of which is dynamically varied by a tuner **16** to better match the transmission frequency, f_C, of the received RF signal, thus obtaining a maximum power transfer. In particular, as further noted in the Related Applications, the RMS voltage induced across the tank circuit **14** by the received RF signal is quantized by tuner **16** and the developed quantization employed to control the impedance of the tank circuit **14**. As also described in the Related References, the unregulated, AC current induced in the tank circuit **14** by the received RF signal is conditioned by a regulator **18** to provide regulated DC operating power to the receiver circuit **10**. In accordance with our present invention, we now provide a field strength detector **20**, also known as a power detector, adapted to develop a field-strength value as a function of the field strength of the received RF signal. As we have indicated in FIG. **1**, our field strength detector **20** is adapted to cooperate with the regulator **18** in the development of the field-strength value. As we shall disclose below, if desired, our field strength detector **20** can be adapted to cooperate with the tuner **16** in controlling the operating characteristics of the tank circuit **14**.

[0022] Shown by way of example in FIG. **2** is one possible embodiment of our field strength or power detector **20**. In this embodiment, we have chosen to employ a shunt-type regulator **18** so that, during normal operation, we can use the shunted 'excess' current as a reference against which we develop the field-strength value. In this regard, we use a reference **22** first to develop a shunt current reference value proportional to the shunted current, and then to develop a mirrored current reference value as a function of both the shunted current and a field strength reference current provided by a digitally-controlled current source **24**. Preferably, once the tuner **16** has completed its initial operating sequence, whereby the f_R of the tank circuit **14** has been substantially matched to the f_C of the received signal, we then enable a digital control **26** to initiate operation of the current source **24** at a predetermined, digitally-established minimum field strength reference current. After a predetermined period of time, control **26** captures the mirrored current reference value provided by the current reference **22**, compares the captured signal against a predetermined threshold value, and, if the comparison indicates that the field strength reference current is insufficient, increases, in accordance with a predetermined sequence of digital-controlled increments, the field strength reference current; upon the comparison indicating that the field strength reference current is sufficient, control **26** will, at least temporarily, cease operation.

[0023] In accordance with our invention, the digital field-strength value developed by control **26** to control the field strength current source **24** is a function of the current induced in the tank circuit **14** by the received RF signal. Once developed, this digital field-strength value can be employed in various ways. For example, it can be selectively transmitted by the RFID device (using conventional means) back to the reader (not shown) for reference purposes. Such a transaction can be either on-demand or periodic depending on system requirements. Imagine for a moment an application wherein a plurality of RFID tag devices are distributed, perhaps randomly, throughout a restricted, 3-dimensional space, e.g., a loaded pallet. Imagine also that the reader is programmed to query, at an initial field strength, all tags "in bulk" and to command all tags that have developed a field-strength value greater than a respective field-strength value to remain silent'. By performing a sequence of such operations, each at an increasing field strength, the reader will, ultimately, be able to isolate and distinguish those tags most deeply embedded within the space; once these 'core' tags have been read, a reverse sequence can be performed to isolate and distinguish all tags within respective, concentric 'shells' comprising the space of interest. Although, in all likelihood, these shells will not be regular in either shape or relative volume, the analogy should still be apt.

[0024] In FIG. **3**, we have illustrated one possible embodiment of our field strength detector **20***a*. In general, we have chosen to use a shunt circuit **18***a* to develop a substantially constant operating voltage level across supply node **28** and ground node **30**. Shunt regulators of this type are well known in the art, and typically use zener diodes, avalanche breakdown diodes, diode-connected MOS devices, and the like.

[0025] As can be seen, we have chosen to implement current reference **22** in the form of a current mirror circuit **22***a*, connected in series with shunt circuit **18***a* between nodes **28** and **30**. As is typical, current mirror circuit **22***a* comprises a diode-connected reference transistor **32** and a mirror transistor **34**. If desired, a more sophisticated circuit such as a Widlar current source may be used rather than this basic two-transistor configuration. For convenience of reference, we have designated the current shunted by shunt circuit **18***a* via reference transistor **32** as i_R; similarly, we have designated the current flowing through mirror transistor **34** as i_R/N, wherein, as is known, N is the ratio of the widths of reference transistor **32** and mirror transistor **34**.

[0026] We have chosen to implement the field strength current source **24** as a set of n individual current sources **24***a*, each connected in parallel between the supply node **28** and the mirror transistor **34**. In general, field strength current source **24***a* is adapted to source current at a level corresponding to an n-bit digital control value developed by a counter **38**. In the illustrated embodiment wherein n=5, field strength current source **24***a* is potentially capable of sourcing thirty-two distinct reference current levels. We propose that the initial, minimum reference current level be selected so as to be less than the current carrying capacity of the mirror transistor **34** when the shunt circuit **18***a* first begins to shunt excess induced current through reference transistor **32**; that the maximum reference current level be selected so as to be greater than the current carrying capacity of the mirror transistor **34** when the shunt circuit **18***a* is shunting a maximum anticipated amount of excess induced current; and that the intermediate reference current levels be distributed relatively evenly between the minimum and maximum levels. Of course, alternate schemes may be practicable, and, perhaps, desirable depending on system requirements.

[0027] Within control **26***a*, a conventional analog-to-digital converter ("ADC") **40**, having its input connected to a sensing node **36**, provides a digital output indicative of the field strength reference voltage, v_R, developed on sensing node **36**. In one embodiment, ADC **40** may comprise a comparator circuit adapted to switch from a logic_0 state to a logic_1 when sufficient current is sourced by field strength current source **24***a* to raise the voltage on sensing node **36** above a predetermined reference voltage threshold, v_{th}. Alternatively, ADC **40** may be implemented as a multi-bit ADC capable of providing higher precision regarding the specific voltage developed on sensing node **36**, depending on the requirements of the system. Sufficient current may be characterized as that current sourced by the field strength current source **24***a* or sunk by mirror transistor **34** such that the voltage on sensing node **36** is altered substantially above or below a predetermined reference voltage threshold, v_{th}. In the exemplary case of a simple CMOS inverter, v_{th} is, in its simplest form, one-half of the supply voltage (VDD/2). Those skilled in the art will appreciate that v_{th} may by appropriately modified by altering the widths and lengths of the devices of which the inverter is comprised. In the exemplary case a multi-bit ADC, v_{th} may be established by design depending on the system requirements and furthermore, may be programmable by the system.

[0028] In the illustrated embodiment, a latch **42** captures the output state of ADC **40** in response to control signals provided by a clock/control circuit **44**. If the captured state is logic_0, the clock/control circuit **44** will change counter **38** to change the reference current being sourced by field strength current source **24***a*; otherwise clock/control circuit **44** will, at least temporarily, cease operation. However, notwithstanding, the digital field-strength value developed by counter **38** is available for any appropriate use, as discussed above.

[0029] By way of example, we have illustrated in FIG. **4** one possible general operational flow of our field strength detector **20***a*. Upon activation, counter **38** is set to its initial digital field-strength value (step **48**), thereby enabling field strength current source **24***a* to initiate reference current sourcing at the selected level. After an appropriate settling time, the field strength reference voltage, v_R, developed on sensing node **36** and digitized by ADC **40** is captured in latch **42** (step **50**). If the captured field strength reference voltage, v_R, is less than (or equal to) the predetermined reference threshold voltage, v_{th}, clock/control **44** will change counter **38** (step **54**). This process will repeat, changing the reference current sourced by field strength current source **24***a* until the captured field strength reference voltage, v_R, is greater than the predetermined reference threshold voltage, v_{th}, (at step **52**), at which time the process will stop (step **56**). As illustrated, this sweep process can be selectively reactivated as required, beginning each time at either the initial field-strength value or some other selected value within the possible range of values as desired.

[0030] The graph illustrated in FIG. **5** depicts several plots of the voltage developed on sensing node **36** as the field strength detector circuit **20***a* sweeps the value of counter **38** according to the flow illustrated in FIG. **4**. As an example, note that the curve labeled "A" in FIG. **5** begins at a logic_0 value when the value of counter **38** is at a minimum value such as "1" as an exemplary value. Subsequent loops though the sweep loop gradually increase the field strength reference voltage on sensing node **36** until counter **38** reaches a value of "4" as an example. At this point, the "A" plot in FIG. **5** switches from a logic_0 value to a logic_1 value, indicating that the field strength reference voltage, v_R, on sensing node **36** has exceeded the predetermined reference threshold voltage, v_{th}. Other curves labeled "B" through "D" depict incremental increases of reference currents, i_R, flowing through reference device **32**, resulting in correspondingly higher mirrored currents flowing through mirror device **34**. This incrementally higher mirror current requires field strength current source **24** to source a higher current level which in turn corresponds to higher values in counter **38**. Thus, it is clear that our invention is adapted to effectively and efficiently develop a digital representation of the current flowing through sensing node **36** that is suitable for any appropriate use.

[0031] One such use, as discussed earlier, of our field strength detector **20** is to cooperate with tuner **16** in controlling the operating characteristics of the tank circuit **14**. FIG. **6** illustrates one possible embodiment where receiver circuit **10***a* uses a field strength detector **20***b* specially adapted to share with tuner **16***a* the control of the tank circuit **14**. In our Related References we have disclosed methods, and related apparatus, for dynamically tuning, via tuner **16***a*, the tank circuit **14** so as to dynamically shift the f_R of the tank circuit **14** to better match the f_C of the received RF signal at antenna **12**. By way of example, we have shown in FIG. **6** how the embodiment shown in FIG. 3 of our Related patent may be easily modified by adding to tuner **16***a* a multiplexer **58** to facilitate shared access to the tuner control apparatus. Shown in FIG. **7** is the operational flow (similar to that illustrated in FIG. 4 in our Related patent) of our new field strength detector **20***b* upon assuming control of tank circuit **14**.

[0032] In context of this particular use, once tuner **16***a* has completed its initial operating sequences as fully described in our Related patent, and our field strength detector **20***b* has performed an initial sweep (as described above and illustrated in FIG. **4**) and saved in a differentiator **60** a base-line field-strength value developed in counter **38**, clock/control **44** commands multiplexer **58** to transfer control of the tank circuit **16***a* to field strength detector **20***b* (all comprising step **62** in FIG. **7**). Upon completing a second current sweep, differentiator **60** will save the then-current field-strength value developed in the counter **38** (step **64**). Thereafter, differentiator **60** will determine the polarity of the change of the previously saved field-strength value with respect to the then-current field-strength value developed in counter **38** (step **66**). If the polarity is negative (step **68**), indicating that the current field-strength value is lower than the previously-saved field-strength value, differentiator **60** will assert a change direction signal; otherwise, differentiator **60** will negate the change direction signal (step **70**). In response, the shared components in tuner **16***a* downstream of the multiplexer **58** will change the tuning characteristics of tank circuit **14** (step **72**) (as fully described in our Related References). Now, looping back (to step **64**), the resulting change of field strength, as quantized is the digital field-strength value developed in counter **38** during the next sweep (step **64**), will be detected and, if higher, will result in a further shift in the f_R of the tank circuit **14** in the selected direction or, if lower, will result in a change of direction (step **70**). Accordingly, over a number of such 'seek' cycles, our invention will selectively allow the receiver **10***a* to maximize received field strength even if, as a result of unusual factors, the f_R of the tank circuit **14** may not be precisely matched to the f_C of the received RF signal, i.e., the reactance of the antenna is closely matched with the reactance of the tank circuit, thus achieving maximum power transfer. In an alternative embodiment, it would be unnecessary for tuner **16***a* to perform an initial operating sequence as fully described in our Related patent. Rather, field strength detector **20***b* may be used exclusively to perform both the initial tuning of the receiver circuit **10***a* as well as the subsequent field strength detection. Note that the source impedance of antenna **12** and load impedance of tank circuit **14** may be represented alternatively in schematic form as in FIG. **8**, wherein antenna **12** is represented as equivalent source resistance R_S **74** and equivalent source reactance X_S **76**, and tank circuit **14** is represented as equivalent load resistance R_L **78** and equivalent, variable load reactance X_L **80**.

[0033] In FIG. **9**, we have illustrated an alternate embodiment of our field strength detector illustrated in FIG. **3**. Here, as before, shunt circuit **18***b* is used to develop a substantially constant operating voltage level across supply node **28** and ground node **30**. Also, as before, the current reference **22** is implemented as a current mirror circuit **22***b* connected in series with shunt circuit **18***b* between nodes **28** and **30**. However, in this embodiment, the field strength current source comprises a resistive component **84** adapted to function as a static resistive pull-up device. Many possible implementations exist besides a basic resistor, such as a long channel length transistor, and those skilled in the art will appreciate the various implementations that are available to accomplish analogous functionality. The field strength voltage reference v_R developed on sensing node **36** will be drawn to a state near the supply voltage when the mirrored current flowing though transistor **34** is relatively small, e.g. close to zero amps, indicating a weak field strength. As the field strength increases, the current flowing through mirror transistor **34** will increase, and the field strength voltage reference v_R developed on sensing node **36** will drop proportionally to the mirrored current flowing through mirror transistor **34** as i_R/N. ADC **40**, having its input connected to sensing node **36**, provides a digital output indicative of the field strength reference voltage, v_R, developed on sensing node **36**, as described previously.

[0034] In this alternate embodiment, latch **42** captures the output state of ADC **40** in response to control signals provided by a clock/control circuit **44**. As disclosed earlier, the ADC **40** may comprise a comparator circuit. In this instance, ADC **40** is adapted to switch from a logic_1 state to a logic_0 when sufficient current is sunk by mirror transistor **34** to lower the voltage on sensing node **36** below a predetermined reference voltage threshold, v_{th}. Alternatively, ADC **40** may be implemented as a multi-bit ADC capable of providing higher precision regarding the specific voltage developed on sensing node **36**, depending on the requirements of the system.

[0035] Comparator **82** subsequently compares the captured output state held in latch **42** with a value held in counter **38** that is selectively controlled by clock/control circuit **44**. In response to the output generated by comparator **82**, clock/control circuit **44** may selectively change the value held in counter **38** to be one of a higher value or a lower value, depending on the algorithm employed. Depending upon the implementation of counter **38** and comparator **82**, clock/control circuit **44** may also selectively reset the value of counter **38** or comparator **82** or both. The digital field-strength value developed by counter **38** is available for any appropriate use, as discussed above.

[0036] In FIG. **10**, we have illustrated another alternate embodiment of our field strength detector illustrated in FIG. **3**. Here, as before, shunt circuit **18***c* is used to develop a substantially constant operating voltage level across supply node **28** and ground node **30**. In this embodiment, the current reference **22** is implemented as a resistive component **86** that functions as a static pull-down device. Many possible implementations exist besides a basic resistor, such as a long channel length transistor and those skilled in the art will appreciate the various implementations that are available to accomplish analogous functionality. The field strength voltage reference v_R developed on sensing node **36** will be drawn to a state near the ground node when the current flowing though shunt circuit **18***c* is relatively small, e.g., close to zero amps, indicating a weak field strength. As the field strength increase, the current flowing through shunt circuit **18***c* will increase, and the field strength voltage reference v_R developed on sensing node **36** will rise proportionally to the current flowing through shunt circuit **18***c*. ADC **40**, having its input connected to a sensing node **36**, provides a digital output indicative of the field strength reference voltage, v_R, developed on sensing node **36**, as described previously.

[0037] In this alternate embodiment, latch **42** captures the output state of ADC **40** in response to control signals provided by a clock/control circuit **44**. As disclosed earlier, the ADC **40** may comprise a comparator circuit. In this instance, ADC **40** is adapted to switch from a logic_0 state to a logic_1 when sufficient current is sourced by shunt circuit **18***c* to raise the voltage on sensing node **36** above a predetermined reference voltage threshold, v_{th}. Alternatively, ADC **40** may be implemented as a multi-bit ADC capable of providing higher precision regarding the specific voltage developed on sensing node **36**, depending on the requirements of the system.

[0038] Comparator **82** subsequently compares the captured output state held in latch **42** with a value held in counter **38** that is selectively controlled by clock/control circuit **44**. In response to the output generated by comparator **82**, clock/control circuit **44** may selectively change the value held in counter **38** to be one of a higher value or a lower value, depending on the algorithm employed. Depending upon the implementation of counter **38** and comparator **82**, clock/control circuit **44** may also selectively reset the value of counter **38** or comparator **82** or both. The digital field-strength value developed by counter **38** is available for any appropriate use, as discussed above.

[0039] In another embodiment, our invention may be adapted to sense the environment to which a tag is exposed, as well as sensing changes to that same environment. As disclosed in our Related References, the auto-tuning capability of tuner **16** acting in conjunction with tank circuit **14** detects antenna impedance changes. These impedance changes may be a function of environmental factors such as proximity to interfering substances, e.g., metals or liquids, as well as a function of a reader or receiver antenna orientation. Likewise, as disclosed herein, our field strength (i.e., received power) detector **20** may be used to detect changes in received power (i.e., field strength) as a function of, for example, power emitted by the reader, distance between tag and reader, physical characteristics of materials or elements in the immediate vicinity of the tag and reader, or the like. Sensing the environment or, at least, changes to the environment is accomplished using one or both of these capabilities.

[0040] As an example, the tag **88** of FIG. **11**, contains both a source tag antenna **12** (not shown, but see, e.g., FIG. **6**) and a corresponding load chip tank circuit **14** (not shown, but see, e.g., FIG. **6**). Each contains both resistive and reactive elements as discussed previously (see, e.g., FIG. **8**). A tag **88** containing such a tank circuit **14** mounted on a metallic surface will exhibit antenna impedance that is dramatically different than the same tag **88** in free space or mounted on a container of liquid. Table 1 displays exemplary values for impedance variations in both antenna source resistance **74** as well as antenna source reactance **76** as a function of frequency as well as environmental effects at an exemplary frequency:

TABLE 1

Antenna Impedance Variations			
In Free Air	860 MHz	910 MHz	960 MHz
R_s	1.9	2.5	3.7
X_s	124	136	149
@910 MHz	Free Air	On Water	On Metal
R_s	2.5	26	1.9
X_s	136	136	27

[0041] The tuner circuit **16** of our invention as disclosed in the Related References automatically adjusts the load impendence by adjusting load reactance **80** (see, e.g., FIG. **8**) to match source antenna impedance represented by source resistance **74** (see, e.g., FIG. **8**) and source reactance **76** (see, e.g., FIG. **8**). As previously disclosed, matching of the chip load impedance and antenna source impedance can be performed automatically in order to achieve maximum power transfer between the antenna and the chip. My invention as disclosed in the Related References contained a digital shift register **90** for selectively changing the value of the load reactive component, in the present case a variable capacitor, until power transfer is maximized. This digital value of the matched impendence may be used either internally by the tag **88**, or read and used by the reader **92**, to discern relative environmental information to which the tag **88** is exposed. For example, tag **88** may contain a calibrated look-up-table within the clock/control circuit **44** which may be accessed to determine the relevant environmental information. Likewise, a RFID reader **92** may issue commands (see transaction **1** in FIG. **11**) to retrieve (see transaction **2** in FIG. **11**) the values contained in digital shift register **90** via conventional means, and use that retrieved information to evaluate the environment to which tag **88** is exposed. The evaluation could be as simple as referencing fixed data in memory that has already been stored and calibrated, or as complex as a software application running on the reader or its connected systems for performing interpretive evaluations.

[0042] Likewise, consider a tag **88** containing our field strength (i.e., received power) detector **20** (not shown, but, e.g., see FIG. **6**) wherein the method of operation of the system containing the tag **88** calls for our field strength detector **20** to selectively perform its sweep function and developing the quantized digital representation of the current via the method discussed earlier. As illustrated in FIG. **11**, counter **38** will contain the digital representation developed by our field strength detector **20** of the RF signal induced current, and may be used either internally by the tag **88**, or read and used by the reader **92**, to discern relative environmental information to which the tag **88** is exposed. For example, reader **92** may issue a command to the tag **88** (see transaction **1** in FIG. **11**) to activate tuner **16** and/or detector **20** and, subsequent to the respective operations of tuner **16** and/or detector **20**, receive (see transaction **2** in FIG. **11**) the digital representations of either the matched impedance or the maximum current developed during those operations. Once again, this digital value of the field strength stored in the counter **38** may be used either internally by the tag **88**, or read and used by the reader **92**, to discern relative environmental information to which the tag **88** is exposed. For example, tag **88** may contain a calibrated look-up-table within the clock and control block **44** which may be accessed to determine the relevant environmental information. Likewise, a RFID reader may issue commands to retrieve the values contained in digital shift register **90**, and use that retrieved information to evaluate the environment to which tag **88** is exposed. The evaluation could be as simple as referencing fixed data in memory that has already been stored and calibrated, or as complex as a software application running on the reader or its connected systems for performing interpretive evaluations. Thus, the combining of the technologies enables a user to sense the environment to which a tag **88** is exposed as well as sense changes to that same environment.

[0043] Thus, it is apparent that we have provided an effective and efficient method and apparatus for quantizing the received RF field strength as a function of induced current. We have developed this field quantization in a form and manner that is suitable for selectively varying the impedance of the tank circuit to maximize received power, especially during normal system operation. Those skilled in the art will recognize that modifications and variations can be made without departing from the spirit of our invention.

Therefore, we intend that our invention encompass all such variations and modifications as fall within the scope of the appended claims.

What is claimed is:

1. A radio frequency (RF) circuit comprises:

an antenna operably coupled to receive an RF signal;

a power harvesting circuit operable to generate a first power source from the RF signal;

a tank circuit coupled to the antenna, wherein the tank circuit includes a selectively variable impedance; and

a tuning circuit operable to dynamically vary the selectively variable impedance of the tank circuit based on resonance of the RF circuit and frequency of the RF signal to substantially align the resonance of the RF circuit with the frequency of the RF signal, wherein, when the resonance of the RF circuit is substantially aligned with the frequency of the RF signal, the power harvesting circuit generates a second power source from the RF signal, wherein the second power source is greater than the first power source.

2. The RF circuit of claim **1** further comprises:

a processing module operable to generate a digital value representative of dynamic variance of the selectively variable impedance of the tank circuit;

memory operable coupled to the processing module; and

a transmitter section operable to transmit an outbound RF signal to a reader, wherein the outbound RF signal includes the digital value.

3. The RF circuit of claim **2** further comprises:

the antenna operable to receive a second RF signal, wherein the second RF signal has an adjusted transmit power with respect to the RF signal based on an interpretation of the digital value such that the second power source is at least one of a desired voltage level and a desired current level.

4. The RF circuit of claim **1**, wherein the tank circuit comprises:

an inductor; and

a variable capacitor coupled to the inductor, wherein the tuning circuit is operable to adjust capacitance of the variable capacitor such that a resonant frequency of the tank circuit substantially matches a frequency of the RF signal.

5. The RF circuit of claim **1**, wherein the power harvesting circuit comprises:

a regulator coupled to the tank circuit, wherein a current is induced in the tank circuit from the received RF signal, wherein the regulator is operable to extract a portion of the induced current to provide direct current (DC) operating power to the RF circuit.

6. The RF circuit of claim **1** further comprises:

a sensing circuit that includes:

the tank circuit having the selectively variable impedance;

the tuning circuit operable to dynamically vary the impedance of the tank circuit, and to develop a first quantized value representative of the impedance of the tank circuit; and

a detector circuit operable to develop a second quantized value as a function of a field strength of a received RF signal; and

memory, wherein the memory stores the first and second quantized values; and

a control circuit operable selectively to retrieve the first and second quantized values from the memory, wherein the retrieved first and second quantized values are used to sense changes to an environment to which the RF circuit is exposed.

7. The RF circuit of claim **6**, wherein the environment includes one or more of:

free space;

air;

liquids; and

metals.

8. The RF circuit of claim **1**, wherein the selectively variable impedance is set to an initial value.

* * * * *

US 20170358957A1

(19) **United States**

(12) **Patent Application Publication**

Mitcheson et al.

(10) **Pub. No.: US 2017/0358957 A1**

(43) **Pub. Date: Dec. 14, 2017**

(54) **RF ENERGY HARVESTER**

(71) Applicant: **Drayson Technologies (Europe) Limited**, London (GB)

(72) Inventors: **Paul David Mitcheson**, London (GB); **Stepan Lucyszyn**, London (GB); **Manuel Pinuela Rangel**, London (GB); **David Christopher Yates**, London (GB)

(21) Appl. No.: **15/669,023**

(22) Filed: **Aug. 4, 2017**

Related U.S. Application Data

(63) Continuation of application No. 14/909,862, filed on Feb. 3, 2016, filed as application No. PCT/GB2014/052431 on Aug. 8, 2014.

(30) **Foreign Application Priority Data**

Aug. 9, 2013 (GB) 1314307.8

Publication Classification

(51) **Int. Cl.**

H02J 50/20	(2006.01)
H01Q 5/364	(2006.01)
H02J 50/80	(2006.01)
H01Q 9/04	(2006.01)
H01Q 1/24	(2006.01)
H02J 7/02	(2006.01)

(52) **U.S. Cl.**

CPC ***H02J 50/20*** (2016.02); ***H01Q 9/0464*** (2013.01); ***H01Q 1/248*** (2013.01); ***H01Q 5/364*** (2015.01); ***H02J 50/80*** (2016.02); *H02J 7/025* (2013.01)

(57) **ABSTRACT**

An antenna apparatus for harvesting ambient radio frequency, RF, energy, comprises one or more RF antenna components arranged to receive RF energy. The antenna components comprise a plurality of frequency filtering components, each filtering a respective frequency band of the received RF energy. A rectifying circuit is arranged to convert a variable electrical signal received at an input from an associated antenna, into a direct current electrical signal for supplying to an electrical energy storage unit. A power management module has an input arranged to receive the direct current and control supply of the direct current to the electrical energy storage unit. The rectifying circuit comprises a plurality of transmission lines, and the input of the rectifying circuit and the input of the power management module are connected via the transmission lines. The power management module is arranged at least partially within a boundary defined by the transmission lines.

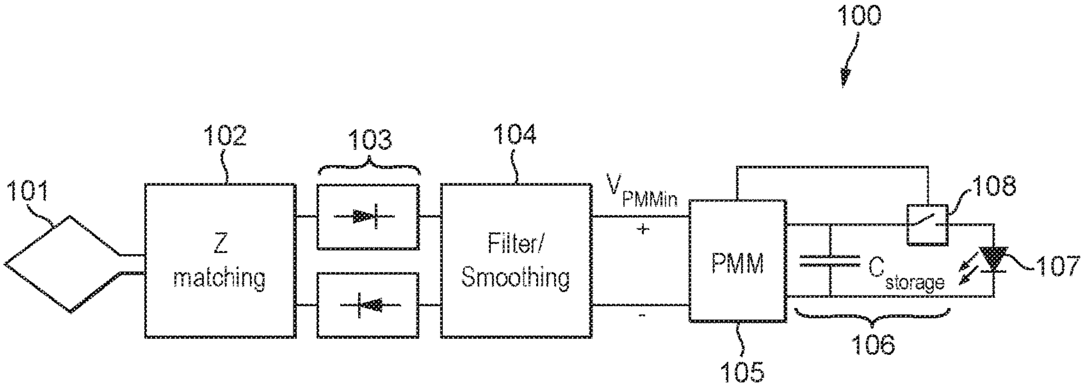

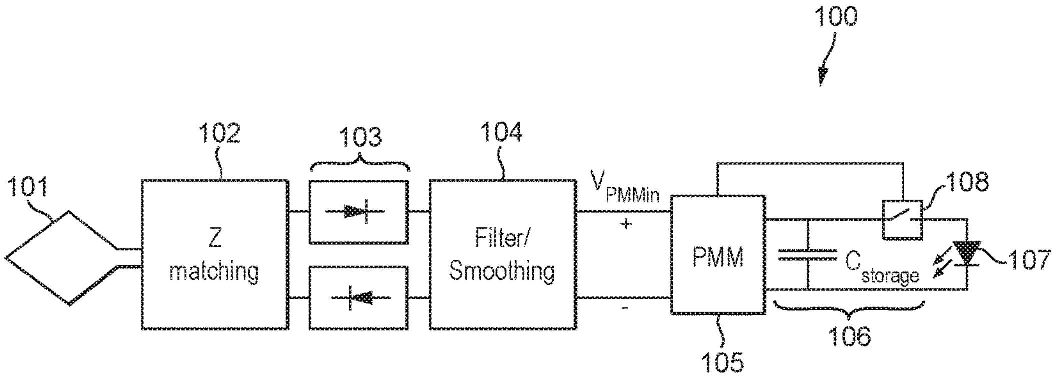
100
101
102
Z
matching
103
104
Filter/
Smoothing
V_{PMMin}
+
-
PMM
105
108
$C_{storage}$
106
107

FIG. 1

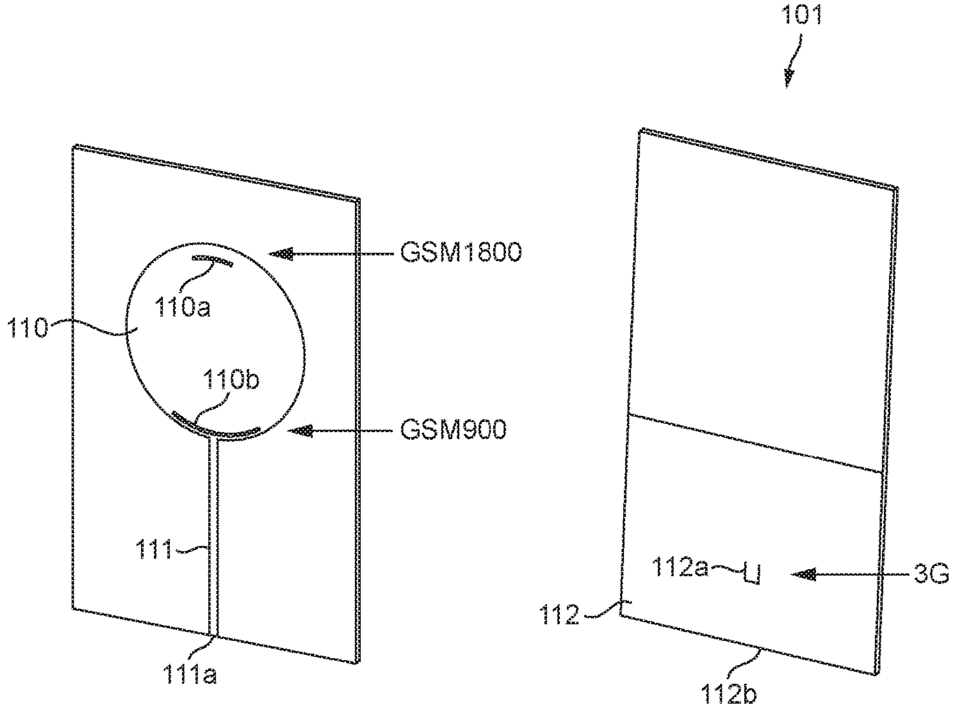
GSM1800
GSM900
3G

FIG. 2

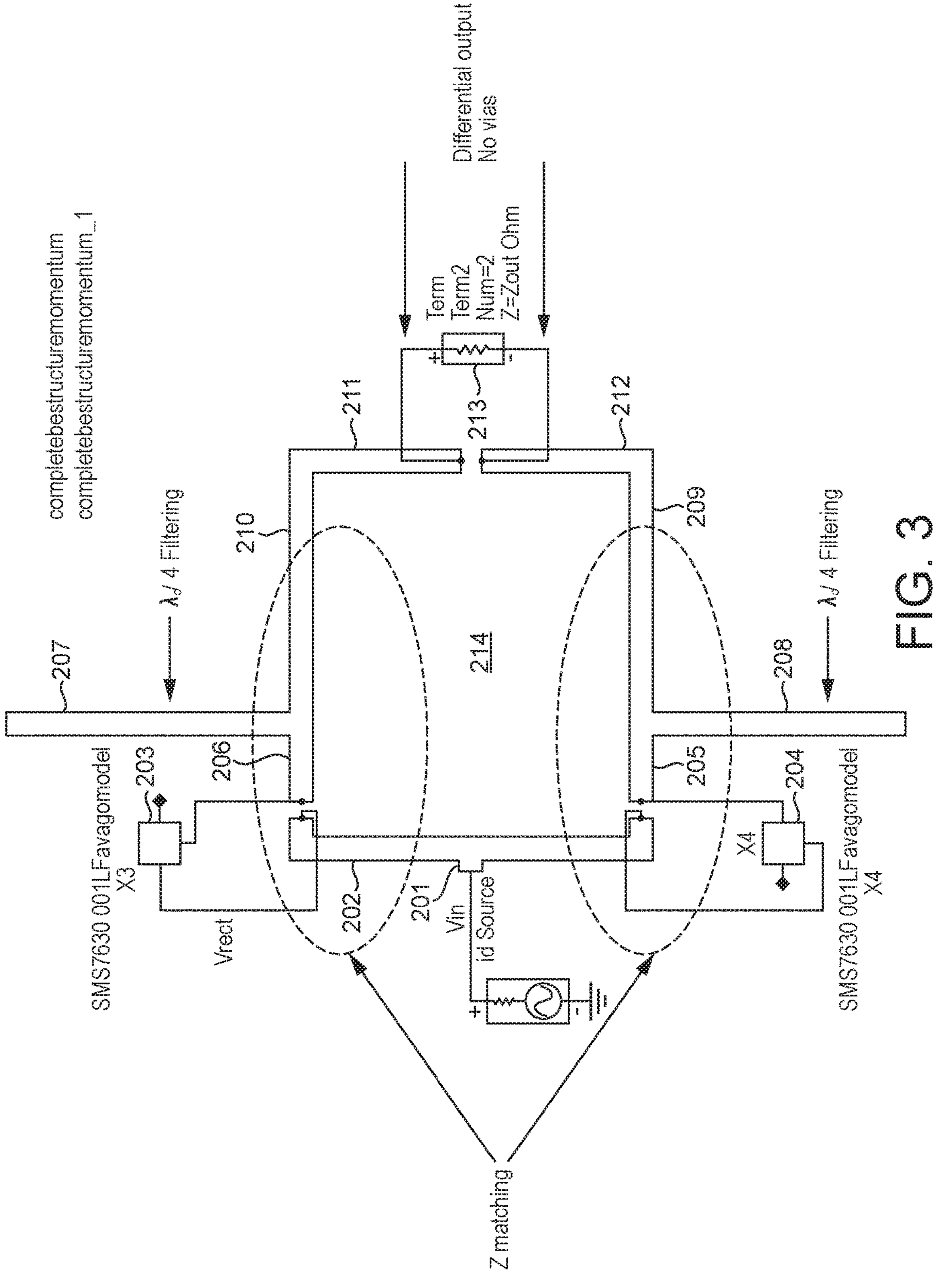
completebestructuremomentum
completebestructuremomentum_1
λ/4 Filtering
207
210
211
213
212
209
208
λ/4 Filtering
214
Term
Term2
Num=2
Z=Zout Ohm
+
-
Differential output
No vias
SMS7630 001LFavagomodel
X3
203
206
205
204
X4
SMS7630 001LFavagomodel
X4
Vrect
202
201
Vin
id Source
+
-
Z matching

FIG. 3

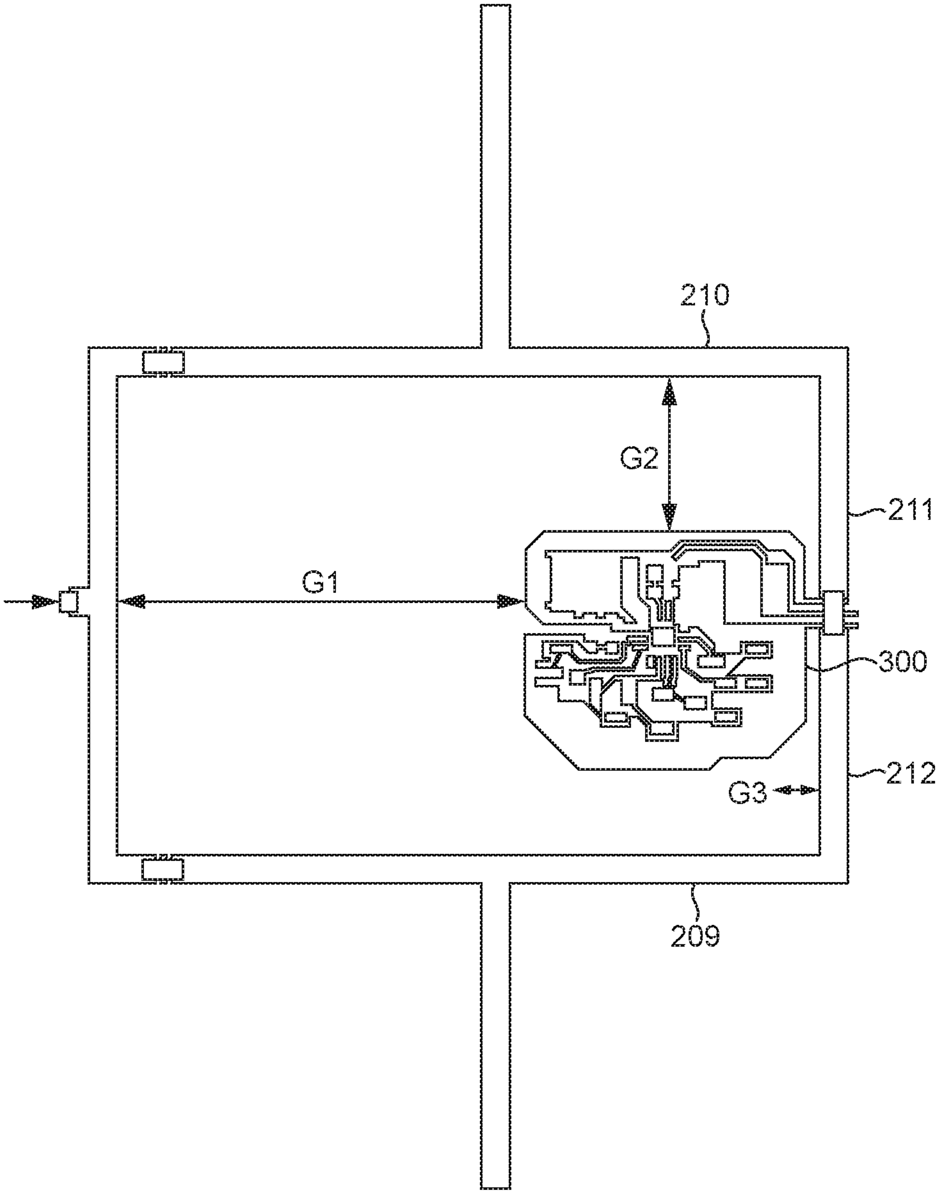
210
G2
211
G1
300
212
G3
209

FIG. 4

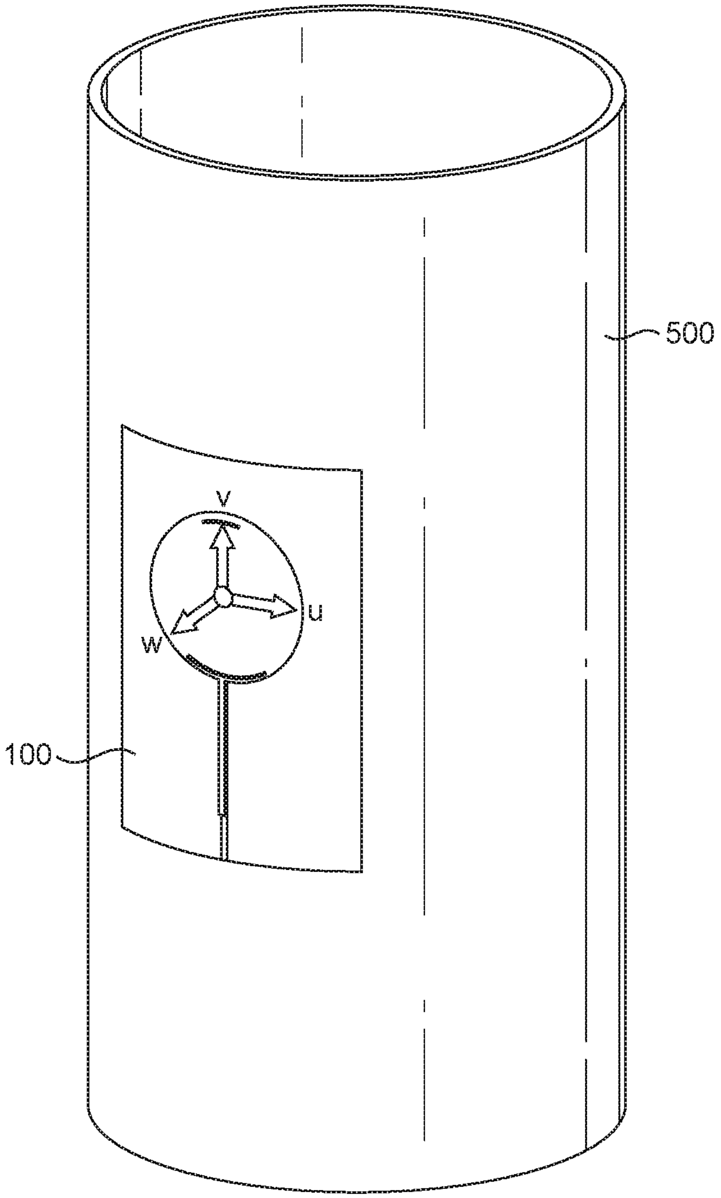

FIG. 5

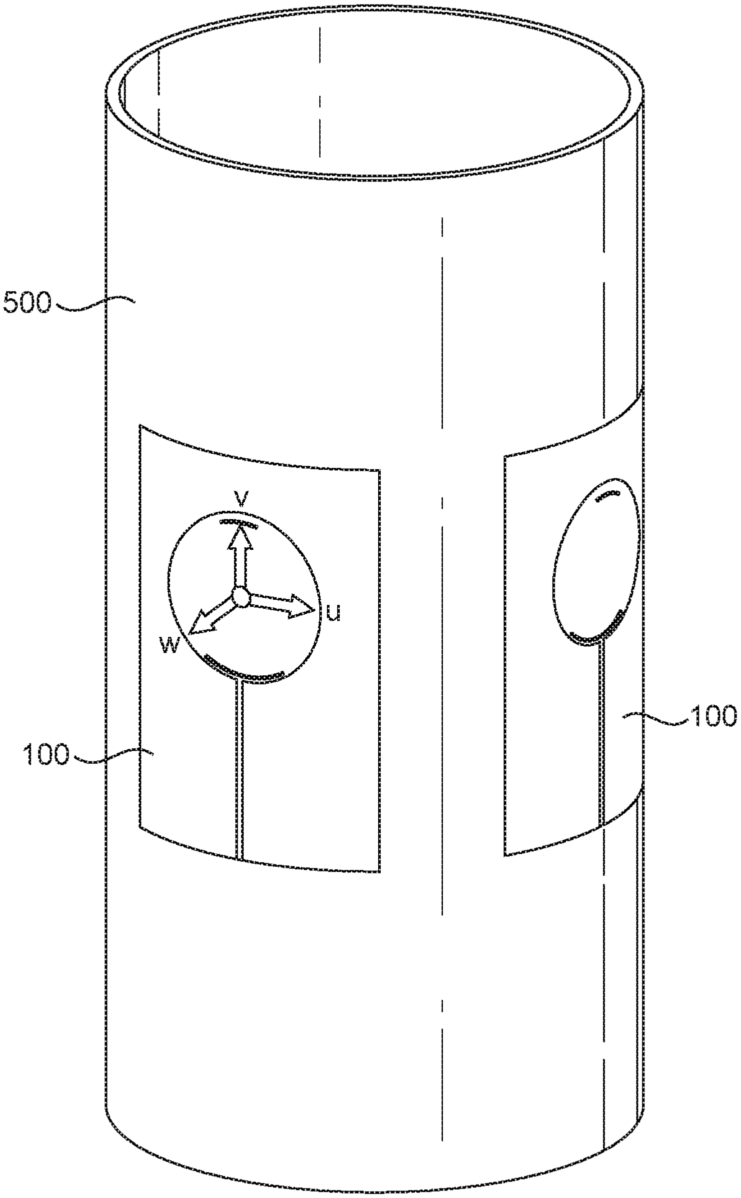
500
v
u
w
100
100

FIG. 6

RF ENERGY HARVESTER

RELATED APPLICATIONS

[0001] This application is a continuation of U.S. patent application Ser. No. 14/909,862, entitled "RF Energy Harvester" and filed Feb. 3, 2016. U.S. application Ser. No. 14/909,862 claims priority to PCT/GB2014/052431, entitled "RF Energy Harvester" and filed Aug. 8, 2014, which in turn claims priority to UK Patent Application No. 1314307.8, entitled "RF Energy Harvester" and filed Aug. 9, 2013. The contents of all of these applications are hereby incorporated by reference in their entireties.

FIELD OF INVENTION

[0002] This disclosure relates to an apparatus for harvesting RF energy. More specifically, but not exclusively, an antenna for receiving ambient RF energy is disclosed. In addition, a rectifier forming part of an RF energy harvesting system and the integration of the rectifier with a power management module is disclosed.

BACKGROUND TO THE INVENTION

[0003] Many electrical devices, particularly portable or remote electrical devices are powered by batteries. Batteries usually require recharging or replacing on a regular basis for continued operation of the electrical device. However, it is not always desirable or possible to connect a device to a mains power supply for recharging, or easy to gain access to the device for replacing the batteries. Moreover, it may not be possible to rely on a battery to perform with 100% reliability over extended periods and/or extreme operational/environmental conditions. As such, other means for powering/charging devices have been explored.

[0004] One proposal for obtaining electrical energy for electrical devices is by means of energy harvesting at the electrical device. Energy harvesting involves converting energy from clean, freely available sources into electrical energy and represents an environmentally friendly solution. Commonly used free energy solutions for powering electrical devices include solar, thermal and vibrational energy sources. For example, practical solar harvesters, such as solar panels and solar rectennas, utilize photovoltaic cells to convert photons into electrical energy via the photovoltaic effect. However, solutions such as solar harvesters are not appropriate for many types of remote electrical devices and are also not ideal for use in many geographical locations.

[0005] An alternative proposal is the wireless delivery of electrical power. For example, high power RF sources can be used to deliver RF power in a specific frequency band to recipient devices. The recipient devices then receive the RF energy and convert it into an electrical form, which can be used to power the device. Such systems can work efficiently in certain environments. However, such systems rely upon the presence of high energy RF transmitters, which reduces the overall attractiveness of such systems due to the need to install infrastructure, and therefore limits the number of possible applications that can use such technology.

[0006] Due to the expanse in the mobile telecommunications industry in recent decades and the corresponding increase in ambient RF energy, some research has looked at whether such ambient RF energy could be harvested. Presently, few practical energy collection solution for harvesting ambient RF energy have been reported, despite the abundance of such energy. One of the key challenges in respect of harvesting RF energy includes the collection of very low RF power, which can be as low as 10 nW. At present, RF energy harvesting does not look to be a viable option.

[0007] There is therefore still a requirement for alternative sources of power for remote devices.

SUMMARY OF INVENTION

[0008] In accordance with an aspect of the invention there is provided an antenna apparatus for use in harvesting ambient radio frequency, RF, energy. The apparatus comprises one or more RF antenna components arranged to receive RF energy for producing electricity. The one or more RF antenna components comprise a plurality of frequency filtering components, each frequency filtering component being arranged to filter a respective frequency band of the received RF energy.

[0009] The frequency filtering components may form part of the one or more RF antenna components. The frequency filtering components may be features formed in or on the one or more RF antenna components. Filtering characteristics of the features may be determined according to the dimensions of the features. The features may be formed in or on the one or more RF antenna components at positions associated with one or more resonant frequencies of the one or more RF antennas within which they are formed.

[0010] The respective frequency bands that the plurality of frequency filtering components are arranged to filter may be one or more known electromagnetic radiation frequency bands. The one or more known electromagnetic radiation frequency bands may include one or more of GSM900, GSM1800, 3G, and 4G.

[0011] One of the one or more RF antenna components may be a circular radiating patch arranged to receive a first frequency range. A first of the plurality of frequency filtering components may be a curved feature formed concentrically in or on the circular radiating patch. The first filtering component may be arranged for filtering one or more frequencies within the first frequency range.

[0012] The first frequency filtering component may be positioned at a distance from the centre of the circular radiating patch of between $\lambda_g/12$ and $\lambda_g/16$, between $\lambda_g/6$ and $\lambda_g/15$ in length, and between $\lambda_g/18.5$ to $\lambda_g/24.2$ in width, wherein λ_g is a wavelength at a centre frequency of the respective frequency band associated with the respective frequency filtering component.

[0013] A second of the plurality of frequency filtering components may be a curved feature formed concentrically in or on the circular radiating patch arranged for filtering one or more frequencies within the first frequency range. The one or more frequencies filtered by the second frequency filtering component may differ to the one or more frequencies filtered by the first frequency filtering component.

[0014] The second frequency filtering component may be positioned at a distance from the centre of the circular radiating patch of between $\lambda_g/12$ to $\lambda_g/24$, between $\lambda_g/4.5$ to $\lambda_g/5.7$ in length, and between $\lambda_g/18.5$ to $\lambda_g/24.2$ in width.

[0015] One of the one or more RF antenna components may be a ground plane arranged to define the radiation pattern of the antenna.

[0016] One of the plurality of frequency filtering components may be a feature formed in or on the ground plane for filtering one or more frequencies within the second frequency range. The feature formed in or on the ground plane may be U-shaped.

[0017] The feature formed in or on the ground plane may be at a distance from a nearest edge of the circular radiating patch of between $\lambda_g/5.4$ to $\lambda_g/8.1$, between $\lambda_g/7.4$ to $\lambda_g/10.5$ in length, and between $\lambda_g/15.2$ to $\lambda_g/24.2$ in width.

[0018] The RF antenna arrangement may comprise a substrate on which the circular radiating patch and the ground plane are formed. The RF antenna arrangement may comprise a connection portion at one edge of the substrate for connecting the circular radiating patch and the ground plane to a harvesting circuit. The circular radiating patch may join the connection portion via an elongate feed-line. A length of the elongate feed-line may be set so that a distance between the circular radiating patch and the ground plane defines an output impedance of the antenna. The distance between the circular radiating patch and the ground plane may be $\lambda_g/4.8$. The substrate may be flexible.

[0019] The plurality of frequency filtering components may be one or more of notches, grooves, or protrusions.

[0020] In accordance with another aspect of the invention there is provided apparatus comprising a rectifying circuit arranged to convert a variable electrical signal received at an input from an associated antenna into a direct current electrical signal for supplying to an electrical energy storage unit, the antenna for use in harvesting ambient radio frequency, RF, energy. The apparatus also comprises a power management module having an input arranged to receive the direct current and control supply of the direct current to the electrical energy storage unit. The rectifying circuit comprises a plurality of transmission lines, wherein the input of the rectifying circuit and the input of the power management module are connected via the plurality of transmission lines. The power management module is arranged at least partially within a boundary defined by the plurality of transmission lines.

[0021] The plurality of transmission lines may form first and second arms. The first arm may connect the input of the rectifying circuit to the input of the power management module via a forward biased rectifying device. The second arm may connect the input of the rectifying circuit to the input of the power management module via a reverse biased rectifying device. The first arm may comprise a first transmission line that connects the input of the rectifying circuit to the forward biased rectifying device, and a second transmission line that connects the forward biased rectifying device to the input of the power management module. The first transmission line may be arranged for impedance matching the rectifier to the associated antenna, and the second transmission line may be arranged for impedance matching the rectifying circuit with the power management module. The first transmission line may have a length of between $\lambda_g/6.8$ and $\lambda_g/8.1$, wherein λ_g is a wavelength at a centre frequency of a respective frequency band arranged to be received by the associated antenna. A filtering circuit may be connected to the second transmission line. The filtering circuit may be a quarter wavelength stub.

[0022] The second transmission line may comprise a linking transmission line and an impedance matching transmission line. The linking transmission line may be arranged between the rectifying device and the filtering circuit. The linking transmission line is arranged to absorb a capacitive reactance associated with the rectifying device. The impedance matching transmission line may be arranged between the filtering circuit and the power management module. The linking transmission line may have a length of between $\lambda_g/4.3$ and $\lambda_g/5.6$. The impedance matching transmission line may have a length of between $\lambda_g/3.8$ and $\lambda_g/4.4$.

[0023] The power management module may be arranged within the rectifying apparatus with the following characteristics. The power management module may be connected at a first end to the output of the rectifying circuit. A distance between a second end of the power management module and the input of the rectifying circuit may be set to be between $\lambda g/2.5$ and $\lambda g/5$. A distance between a first side of the power management module may be set at between $\lambda g/6.5$ and $\lambda g/12$ from an edge of the first arm. A distance between a second side of the power management module may be set at between $\lambda g/9$ and $\lambda g/16.4$ from an edge of the second arm.

[0024] In accordance with a further aspect of the invention there is provided an RF energy harvesting system, comprising an antenna as described herein and an apparatus as described herein. The RF energy harvesting system may further comprise an electrical energy storage unit arranged to receive and store electrical energy from the power management module of the apparatus.

[0025] A rectenna for harvesting ambient RF energy at power levels below 3 μW is disclosed. The rectenna may comprise an ultra-wide band planer monopole antenna operational over the GSM900, GSM1800 and 3G RF frequency bands. The rectenna may also comprise a full-wave rectifier having a differential voltage output and operational input powers below 3 μW.

[0026] An ambient RF harvester is presented where efficiencies above 45% are achievable at input RF power levels below 3 μW. To achieve this efficiency, a single multiband antenna and a full-wave detector may be provided. Such a harvester not only charges an energy storage unit faster, but is capable of operating in more locations and/or with more flexibility in spatial orientation.

[0027] A full-wave rectifier is disclosed with differential output and a modified omnidirectional disc monopole antenna with harmonic rejection filters. The new harvester design may be capable of operating with an efficiency of 45% in locations with lower available RF power.

[0028] A tri-band RF energy harvester is disclosed with an UWB antenna and integrated harmonic band rejection. A tetra or N-element harvester array may be provided.

[0029] Advantageously, a harvester is provided that is capable of providing duty cycles higher than 11% for applications requiring up to 1 mW.

[0030] A harvester that avoids harmonic re-radiation thereby improving efficiency and minimizing detection is disclosed. The harvester may achieve this by filtering the higher order harmonics such that they are not radiated by the antenna. Consequently, it is difficult for a dedicated receiver to detect the presence of the harvester when covert operation is required.

[0031] In addition, a harvester that provides DC-load isolation from higher-order harmonics is disclosed. The harvester may achieve this by use of a high frequency filtering circuit. For example, quarter wavelength stubs may form part of the rectifier for performing such filtering. Such stubs act as short circuits at the fundamental frequency, thus filtering RF signal from the output. Alternatively, resonant structures such as LC tuned circuits can be used to perform such filtering functionality.

[0032] A harvester that is capable of 3rd harmonic rejection for all received bands is disclosed. Since half-wave rectification produces odd harmonics and full wave rectification produces even harmonics in the current waveform the high order harmonics may be filtered at the output by quarter wavelength stubs and by notches formed in the antenna.

[0033] The antenna may be formed on a substrate having a bend radius 10× the thickness of the antenna substrate. The minimum bend radius may be 8.6 mm. The substrate thickness may be 790 μm. The metal antenna formed on the substrate may be 35 μm thick. The metal may be formed of any suitable conductive material. For example, the metal antenna may be formed of copper. The dielectric properties of the substrate of dielectric constant ϵr may be 2.33 and loss tangent tan δ may be 0.0012.

BRIEF DESCRIPTION OF THE DRAWINGS

[0034] Exemplary embodiments of the invention shall now be described with reference to the drawings in which:

[0035] FIG. **1** illustrates an RF energy harvesting system;

[0036] FIG. **2** illustrates a single multi-band antenna used in the system of FIG. **1**;

[0037] FIG. **3** illustrates a microstrip full-wave rectifier with a differential output as used in the system of FIG. **1**;

[0038] FIG. **4** illustrates the rectifier of FIG. **4** with the PMM circuit placed inside the rectifier;

[0039] FIG. **5** illustrates an antenna conformed to the surface of a pipe; and

[0040] FIG. **6** shows an array of antennas conformed to the surface of a pipe.

[0041] Throughout the description and the drawings, like reference numerals refer to like parts.

SPECIFIC DESCRIPTION

[0042] A general overview of an RF energy harvesting system shall now be described with reference to FIG. **1**.

[0043] A rectenna **100** is shown that includes an antenna **101** arranged to receive ambient RF energy. The antenna is designed to receive ambient RF energy across a large frequency range and therefore the electrical output of the antenna **101** has a power profile that fluctuates considerably across its reception frequency range. The output of the antenna **101**, which provides a variable electrical signal, is connected to a rectifier **103** via an impedance matching network **102**. The impedance matching network **102** is provided to minimize losses due to impedance mismatch between components of the rectenna. The rectifier then converts the received time varying electrical signal into a DC electrical signal. The output of the rectifier is then smoothed and filtered in a smoothing/filtering circuit **104**. This smoothed DC energy is then fed into a power management module (PMM) **105** to provide useful regulated voltages and charge a storage unit **106**, which in this case is a storage capacitor $C_{storage}$. The transmission of current to the storage unit **106** and corresponding electrical storage in the storage unit **106** is controlled by a power management module (PMM) **105**. The PMM provides maximum power point tracking in addition to working as a boost converter. The maximum power tracking capabilities of the PMM allow the system to maximize the power transfer to the load by ensuring that the ideal load is present across a wide range of input RF power levels. Therefore, the PMM also helps to provide a good impedance match with the antenna. The PMM is also arranged to control a switch **108** that connects the electrical energy stored in the storage unit **106** to a load, which in this case is an LED for simplicity of explanation. However, in practice the PMM controls the supply of power from the storage unit **106** to the device, i.e. the load being powered.

[0044] The rectenna **100** is designed to maximize the ambient RF energy received by the antenna **101**, and minimize the losses in the conversion of the received energy to energy suitable for storing in the storage unit **106**. The various components of the rectenna **100** shall therefore now be described in detail.

[0045] In order to maximize the power harvested by the rectenna and reduce the size of the harvester, the antenna **101** is a single multi-band antenna, as can be seen in FIG. **2**.

[0046] The antenna **101** is a planar antenna structure comprising a substrate upon which the receiving components of the antenna are formed. The antenna **101** is a modified microstrip disc monopole antenna comprising receiving components including a circular-shaped radiating patch **110**, a feed line **111** and a rectangular shaped ground plane **112**. The circular-shaped radiating patch **110** and feed line **111** are arranged on one side of the substrate, and the ground plane **112** is arranged on the opposing side of the substrate. While the antenna is shown as having a microstrip-based feedline it will be appreciated that a coplanar waveguide feed is provided in alternative arrangements, or a combination of both a microstrip-based feedline and a coplanar waveguide feed. The substrate may be made of any low loss material, for example, glass, ceramic, FR4. The planar characteristics of the antenna allow it to be formed on the same substrate as the rest of the harvester system.

[0047] The antenna **101** is arranged to be substantially omnidirectional in order to maximize the received RF radiation. Furthermore, the structural components of the antenna are arranged to maximize the receipt of RF energy from different possible sources. Since most ambient RF energy comes from mobile telecommunications, the antenna **101** is arranged to operate across an ultra-wide band covering a plurality of the following bands: GSM900, GSM1800, 3G, WiFi and 4G. It will be appreciated that an alternative antenna arrangement could be arranged to receive other frequency bands, higher and lower than those mentioned above to cover radio, TV and other ISM bands, if desired. The circular-shaped radiating patch **110** is primarily arranged for receiving all bands, while the ground plane **112** is modified to enhance the 3^{rd} harmonic rejection of specific bands 3G and 4G, which is generated by the rectification process. The arrangement of the radiating patch **110** and ground plane **112** for receiving these different frequency bands shall now be discussed in detail.

[0048] The radiating patch **110** is circular in shape and positioned on an upper portion of the substrate. An elongated electrical feed line **111** in the form of a microstrip transmission line provides a connection from the circular-shaped radiating patch **110** to an end surface of the substrate at point **111***a*. It is at this point that the antenna **100** is connectable to the next stage of the energy harvesting system. The elongate track **111** not only acts to enable the remainder of the rectenna to be connected to the radiating patch **110**, but also provides impedance matching, as will be discussed. The radius of the disc is a quarter wavelength of the lowest frequency that the antenna is arranged to receive, e.g. 945 MHz for GSM900.

[0049] The ground plane **112** is rectangular in shape and extends from a bottom of the antenna substrate to a point just under half way along the length of the antenna substrate. In alternative arrangements the ground plane may take a different form. The shape of the ground plane defines the impedance and the radiation pattern. An electrical connection on an end surface of the substrate allows for the elongated track to be electrically connected to the rest of the energy harvesting system. The circular-shaped radiating patch, feedline and ground plane are formed of a conductive material such as copper.

[0050] The antenna's impedance can be modified to conjugately match that of the rectifier by varying the width of the antenna feed and the distance between the circular-shaped radiating patch and the ground plane. In particular, the distance between the bottom of the circular-shaped radiating patch **110** and the top of the ground plane **112** can be used to tune the impedance to 50Ω (or any chosen reference impedance) and optimize the desired omnidirectional pattern. The shape and overlap of the ground plane with the disc defines the radiation pattern by reflecting the field lines coming from the disc. In consequence, this also defines the gain of the antenna which can be between 2.3 and 4.8 dBi. This distance is also dependent upon the wavelengths that the antenna is arranged to receive. In practice the distance between the bottom of the circular-shaped radiating patch and the top of the ground plane is set between: –λg/20 and λg/12.5, where λg is the guided wavelength of the signal at the centre frequency of the selected frequency band. A distinction of λg and λo is used to highlight that this is the wavelength of the guided-wave structure, which is different than that in free space.

[0051] The antenna **101** has a number of frequency filtering features that are formed on the antenna and change the frequency response characteristics of the antenna. In particular, these filters are arranged to filter out the third harmonic generated by the full-wave rectifier at each one of the selected bands that the antenna is arranged to receive, e.g. GSM900, GSM1800, and 3G. When the antenna is used with a half-wave rectifier, a second harmonic filter is required, rather than a third harmonic filter.

[0052] First and second filtering notches **110*a*** and **110*b*** are provided on the circular-shaped radiating patch, while a further notch **112*a*** is provided on the ground plane **112**. The first and second notches **110*a***, **110*b*** are curved and concentrically arranged on the circular-shaped radiating patch. The first filtering notch **110*a*** has a length of λg/6 to λg/7.5. The first filtering notch is radially positioned at between λg/12 and λg/16 from the centre of the circular-shaped radiating patch. The width of the first filtering notch is between λg/18.5 and λg/24.2, which optimizes the return loss and the bandwidth of the notches. The second filtering notch **110*b*** has a length of between λg/4.5 and λg/5.7, a width of between λg/18.5 and λg/24.2, and the notch is radially positioned λg/12 to λg/24 from the centre of the circular-shaped radiating patch. By modifying the position and the width of the notches, the return loss can be increased and the bandwidth optimized to cover all of the targeted frequencies. It is the resonances generated by the position and geometry of the notches that defines the associated return loss and bandwidth of the filtering provided by the notches.

[0053] The notch **112*a*** formed in the ground plane **112** is U-shaped with the base of the U facing downwards. The centre of the notch in the ground plane **112** is between λg/5.4 to λg/8.1 from the lower edge of the circular radiating patch, and the notch **112*a*** has a length of between λg/7.4 and λg/10.5 and a width of between λg/15.2 and λg/24.2.

[0054] In summary, the length of the first and second notches is approximately a third of the wavelength of the third harmonic and the width is proportional to the bandwidth characteristics of the filter. The U-shaped **112*a*** notch is included to filter the harmonics created by the 3G rectifier Since there is a different rectifying circuit for each band, and each rectifier generates third order harmonics corresponding to their frequency of operation, different filters or notches are required for each respective component of the antenna.

[0055] In alternative arrangements additional notches can be added to the structure, especially in the circular-shaped radiating patch. Additional notches may be used if harvesting is also desired at other bandwidths, such as 4G or WiFi. However, harmonic rejection will be lower than –2 dB. In other words, the additional filters will not be able to achieve the desired filtering properties and it is possible than around 50% of the power in the harmonics will not be filtered.

[0056] FIG. **3** illustrates the rectifying circuit **103** of FIG. **1**. The rectifying circuit **103** is a microstrip full-wave rectifier with a differential output. The rectifying circuit shown in FIG. **3** includes the impedance matching **202**, rectifier **203**, and filtering **104** components shown in the schematic diagram of FIG. **1**. The rectifying circuit is designed to be as lossless as possible to thereby reduce diode voltage drops resulting of a conventional full-wave (four diode) rectifier. This is primarily achieved by ensuring that accurate impedance matching is provided throughout the circuit. The rectifier **203** is tuned to the frequency of the frequency range that it is receiving form the antenna **101**. In the case of the antenna **101**, which is arranged to receive multiple bands, multiple rectifiers **203** are provided for each respective band. However, in the Figures only a single rectifier is shown and explained for ease of explanation.

[0057] The output from the antenna **101** is received at the feed **201** of a microstrip transmission line. At this point the microstrip transmission line splits into two arms, one connecting the input of the rectifier to the output via a forward biased rectifying element **203**, and a second arm connecting the input of the rectifier to the output via a second reverse biased rectifying element **204**.

[0058] The arms of the microstrip transmission line **202** therefore extend perpendicularly left and right from the input **201**. The track width of the microstrip transmission line is set according to the desired characteristic impedance of the circuit. The track width can also be set in order to simplify the impedance matching of the rectifying circuit with the antenna. At each end of the impedance matching microstrip transmission line **202** further parallel tracks run for a short distance perpendicular to the impedance matching microstrip transmission line **202** and then each join a first terminal of a respective rectifying element **203**, **204**. The length of impedance matching microstrip transmission line **202** is set so as to match the impedance of the antenna **101**, which in this case is 50Ω. This length is preferably between λg/6.8 and λg/8.1. Hence, the impedance matching microstrip transmission line is performing the functionality of the impedance transformer **102** of FIG. **1**. The width and length of the microstrip transmission lines between the input and the rectifying elements are used to tune the input impedance of the rectifier.

[0059] The rectifying elements are Schottky diodes **203**, **204**, with the first diode **203** being forward biased and the second diode **204** being reverse biased. It will be appreciated that in alternative arrangement a transistor based rectifying arrangement may be provided, for example FETs may be modified to perform the rectification functionality.

[0060] The further microstrip transmission line then extends for another short portion **205**, **206** from a second terminal of the respective diode **203**, **204**. The length and width of this part of microstrip line is arranged to absorb the capacitive reactance of the diode because they define an inductive reactance. In addition, the length and width of this part of the microstrip line can be used in tuning the input impedance of the rectifying circuit. The length of the further microstrip transmission line **205**, **206** is preferably between λg/4.3 and λg/5.6. At this point, λ/4 stubs **207**, **208** extend perpendicularly from the ends of the microstrip transmission lines **205**, **206**. The λ/4 stubs **207**, **208** perform the filtering functionality **104** of FIG. **1**. The stubs are used to filter the fundamental harmonic from the output, providing DC at the output. The λg/4 stub ensures that the fundamental and the higher order harmonics are below −45 dBm.

[0061] The microstrip transmission line then continues past the stubs for a longer distance **209**, **210**, until the two arms of the microstrip transmission line extend perpendicularly such that they are again aligned and join at the two output terminals. This section is preferably between λg/3.8 and λg/4.4 in length. This section works as an impedance transformer so that the load and the rest of the rectifier have the conjugate match of the antennas' input impedance throughout a range of frequencies and input power levels.

[0062] In use, diode **203** rectifies the positive part of the received signal and diode **204** rectifies the negative part of the received signal. The quarter wavelength stubs filter the fundamental harmonic from the output ensuring that DC is available. Since the output connections are not grounded, the negative voltage coming from the diode **204** is the reference to the rest of the system blocks, providing a higher voltage to reach the cold-start voltage (330 mV) of the PMM.

[0063] The output is a differential output, which allows the system to provide a higher output voltage VPMM and reach the minimum voltage required by the Power Management Module (PMM), 330 mV, with input power levels below 3 μW (−25 dBm). Furthermore, by using a differential output ground vias are avoided, reducing losses, helping to achieve a better match between simulations and measurements, as well as simplifying fabrication.

[0064] A single full-wave rectifier can achieve 12% efficiency at −26 dBm of input power and up to 46% at −10 dBm. The return loss is below the −10 dB limit ensuring an almost complete cover of the selected operating band. The fractional bandwidth varies between 2 to 6.2% depending on the operating frequency. The output of the harvester for the input power range mentioned previously is between, 75 to 550 mV.

[0065] In order to minimize the footprint of the complete RF energy harvester, the PMM circuit **300** is placed inside a boundary defined by the transmission lines of the full-wave rectifier as shown in FIG. **4**. This is in contrast to known systems which place the PMM **300** in series with the rectifier outside the perimeter of the rectifier. However, careful considerations are required to avoid coupling between the rectifying structure and the miniaturized PMM circuit **300**. To minimize the electromagnetic interaction between the PMM circuit **300** and the full-wave rectifier the distance between the edge of the PMM circuit **300** and the microstrip transmission line of the rectifier are set as follows. The PMM circuit **300** joins the differential output directly. Hence, the distance between the PMM circuit **300** and the microstrip transmission line **211**, **212** joining the differential output is negligible. Then, the distance between the opposite end of the PMM circuit **300** to the one joining the differential output and the input to the rectifying circuit is set to be between λg/2.5 and λg/5. A first side of the PMM circuit **300** is set at between λg/6.5 and λg/12 from a first upper edge **210**, as well as from the second lower edge of the microstrip transmission line and the distance between the PMM **300** and the differential output is set at between λg/9 and λg/16.4.

[0066] In the arrangement shown in FIG. **4** the PMM **300** is provided substantially within the boundary defined by the transmission lines. It will be appreciated that in alternative arrangements the PMM may only be provided within some of the boundary of the transmission line, which would still provide some of the space savings the energy harvester system as a whole.

[0067] In addition to the location of the PMM circuit, the layout shown in FIG. **4** avoids the use of vias, while providing connections for the rectifier's differential output and the "grounded" PMM output.

[0068] FIG. **5** shows how the antenna, and in consequence the rectenna, can conform to surfaces. In this case, the substrate on which the antenna **100** is formed is made from a flexible material. Consequently, the antenna **100** is bent around a pipe **500**. In this conformal antenna design the pipe **500** is a PVC water pipe with the following characteristics: radius=5-10 cm, ϵr=2.88, and tan δ=0.089. There is no return loss degradation from bending the antenna around the pipe.

[0069] It will also be appreciated that in an alternative arrangement multiple antennas may be provided in an array. For example, FIG. **6** shows an array of antennas **100** conformed to a pipe **500**. Use of several rectennas increases the total harvested power. The distance between each harvester can be as close as λo/4.8, where λo is the wave-length of the antenna's first resonance.

[0070] It will be appreciated that these examples of conforming the antenna to a pipe are applicable to any cylindrical object.

[0071] It will be appreciated that alternative arrangements could be provided wherein similar notches are placed in other types of monopoles and planar antennas such as a bow-tie and a square patch. The ground plane could be extended to modify the radiation pattern of the antenna. This would also modify the gain of the antenna.

[0072] In alternative arrangements multiple filtering components are formed together as a filtering arrangement. In other words, as single filtering arrangement is provided having multiple filtering components, each filtering component arranged to perform a specific filtering functionality such as performing filtering for a specific frequency band.

[0073] It will be appreciated that the harvester described herein may be used in various applications such as for electronic paper displays, low power displays, industrial sensors (e.g. strain-gauge, temperature, pressure, inclination, flow, proximity, smoke, gases, COx, NOx), powering LEDs, and charging storage units.

[0074] Furthermore, it will be appreciated that the impedance matching network can be omitted if the port impedance of the antenna is designed to directly match that of the rectifier.

[0075] It will be appreciated that the rectifying circuit need not be a microstrip-based circuit. For example, in alternative arrangements stripline, coplanar strip or CPW guided-wave structures are used. In yet further arrangements lumped-element or lumped-distributed components are used.

[0076] In alternative arrangements the full-wave rectifier can be scaled up or down in frequency to any desired frequency band.

[0077] In alternative arrangements the antenna is at least one of a bow-tie antenna, a modified square patch antenna, dual-monopole antenna, a log-periodic antenna, a 3D antenna, or a non-planar antenna such as a helical or double-helix antenna.

[0078] It will be appreciated that other filtering features such as protrusions could achieve the same band filtering response as the notch filter described herein. Furthermore, while the notches formed in the circular radiating patch are shown as being concentric, it will be appreciated that the notches need not be concentric. However, improved performance can be achieved by use of concentric notches.

[0079] It will be appreciated that the filtering features for different bands may be provided on either side of the antenna structure. However, use of a U-notch in the ground plane achieves an improved performance for higher frequencies than for lower frequencies.

[0080] The rectenna described above is primarily described for use in harvesting ambient RF energy. However, it will be appreciated that the rectenna can also be used for receiving RF energy from a dedicated source, or for receiving some RF energy from a dedicated source as well as harvesting ambient RF energy.

[0081] The rectifier described herein is arranged for use with a single frequency band. When an antenna is arranged to receive multiple frequency bands a respective rectifier is utilized for each band.

[0082] Where components of the system are shown or described in a way that defines a particular orientation such as upper or lower, it will be appreciated that the orientation is not important but only the relative position(s) of system components.

What is claimed is:

1. An antenna apparatus (**101**) for use in harvesting ambient radio frequency, RF, energy and connectable to a rectifier (**103**), the antenna apparatus (**101**) comprising:

a substrate comprising a first side and a second side opposing the first side;

a radiating patch (**110**) formed on the first side of the substrate; and

a ground plane (**112**) formed on the second side of the substrate,

wherein the radiating patch (**110**) is arranged to receive RF energy for producing electricity,

wherein the radiating patch (**110**) comprises first and second filtering notches (**110***a*, **110***b*) each being arranged to modify a return loss and bandwidth of the antenna apparatus (**101**) with respect to a targeted frequency of the received RF energy,

wherein the ground plane (**112**) comprises a ground plane notch (**112***a*), and

wherein the ground plane notch (**112***a*) is arranged to filter harmonics created by the rectifier (**103**).

2. The antenna apparatus according to claim **1**, wherein the return loss and bandwidth of the antenna apparatus (**101**) modified by the first and second filtering notches (**110***a*, **110***b*) are determined according to the dimensions of the first and second filtering notches (**110***a*, **110***b*), and wherein the first and second filtering notches (**110***a*, **110***b*) are formed in or on the one or more RF antenna components at positions associated with one or more resonant frequencies of the radiating patch (**110**) within which they are formed.

3. The antenna apparatus according to claim **1**, wherein the targeted frequency of the received RF energy lies in one or more known electromagnetic radiation frequency bands, and wherein the one or more known electromagnetic radiation frequency bands include one or more of GSM900, GSM1800, 3G, and 4G.

4. The antenna apparatus according to claim **1**, wherein the radiating patch (**110**) is arranged to receive a first frequency range and the radiating patch (**110**) is circular-shaped.

5. The antenna apparatus according to claim **4**, wherein the first filtering notch (**110***a*) is a curved feature formed concentrically in or on the circular-shaped radiating patch (**110**), the first filtering notch (**110***a*) being arranged for filtering one or more frequencies within the first frequency range.

6. The antenna apparatus according to claim **5**, wherein the first filtering notch (**110***a*) is:

positioned at a distance from the centre of the circular-shaped radiating patch of between λg/12 and λg/16;

between λg/6 and λg/15 in length; and

between λg/18.5 to λg/24.2 in width, wherein

λg is a wavelength at a centre frequency of the respective frequency band associated with the respective filtering notch (**110***a*).

7. The antenna apparatus according to claim **5**, wherein the second filtering notch (**110***b*) is a curved feature formed concentrically in or on the circular-shaped radiating patch (**110**) arranged for filtering one or more frequencies within the first frequency range, wherein the one or more frequencies filtered by the second filtering notch (**110***b*) differ to the one or more frequencies filtered by the first filtering notch (**110***a*).

8. The antenna apparatus according to claim **6**, wherein the second filtering notch (**110***b*) is:

positioned at a distance from the centre of the circular-shaped radiating patch (**110***a*) of between λg/12 to λg/24;

between λg/4.5 to λg/5.7 in length; and

between λg/18.5 to λg/24.2 in width, wherein

λg is a wavelength at a centre frequency of the respective frequency band associated with the respective filtering notch (**110***b*).

9. The antenna apparatus according to claim **4**, wherein the ground plane is arranged to define the radiation pattern of the antenna apparatus (**101**).

10. The antenna apparatus according to claim **9**, wherein the ground plane notch (**112***a*) is for filtering one or more frequencies within a second frequency range.

11. The antenna apparatus according to claim **10**, wherein the ground plane notch (**112***a*) is U-shaped.

12. The antenna apparatus according to claim **10**, wherein the ground plane notch (**112***a*) is formed in or on the ground plane:

at a distance from a nearest edge of the circular-shaped radiating patch of between λg/5.4 to λg/8.1;

between λg/7.4 to λg/10.5 in length; and

between λg/15.2 to λg/24.2 in width, wherein λg is a wavelength at a centre frequency of the respective frequency band associated with the respective frequency filtering component.

13. The antenna apparatus according to any of claim **9**, wherein the substrate is flexible.

* * * * *

US 20220069588A1

(19) **United States**
(12) **Patent Application Publication**
Dinwiddie et al.
(10) **Pub. No.: US 2022/0069588 A1**
(43) **Pub. Date: Mar. 3, 2022**

(54) **POWER RECEIVER FOR EXTRACTING POWER FROM ELECTRIC FIELD ENERGY IN THE EARTH**

(71) Applicant: **Earth Energies, Inc.**, Alpharetta, GA (US)

(72) Inventors: **John Dinwiddie**, Pittsboro, NC (US); **Terry L. Wright**, Suwanee, GA (US); **David R. Ames**, Johns Creek, GA (US)

(21) Appl. No.: **17/501,457**

(22) Filed: **Oct. 14, 2021**

Related U.S. Application Data

(63) Continuation of application No. 16/521,764, filed on Jul. 25, 2019, now Pat. No. 11,152,793, which is a continuation of application No. 15/397,281, filed on Jan. 3, 2017, now Pat. No. 10,389,138, which is a continuation-in-part of application No. 14/509,772, filed on Oct. 8, 2014, now Pat. No. 9,564,268.

(60) Provisional application No. 61/889,894, filed on Oct. 11, 2013.

Publication Classification

(51) **Int. Cl.**

H02J 4/00	(2006.01)
H05F 7/00	(2006.01)
H01F 38/14	(2006.01)

(52) **U.S. Cl.**
CPC ***H02J 4/00*** (2013.01); *H01F 38/14* (2013.01); *H05F 7/00* (2013.01)

(57) **ABSTRACT**

A resonant transformer connected between a ground terminal and elevated terminal draws current from the earth's electric field through a primary winding of the transformer. An impulse generator applies a high voltage impulse to the primary winding of the resonant transformer to cause current to flow from the ground terminal through the primary winding. The flow of current through the primary winding of the resonant transformer induces a current in the secondary winding, which may be converted and filtered to a usable form, e.g. 60 Hz AC or DC.

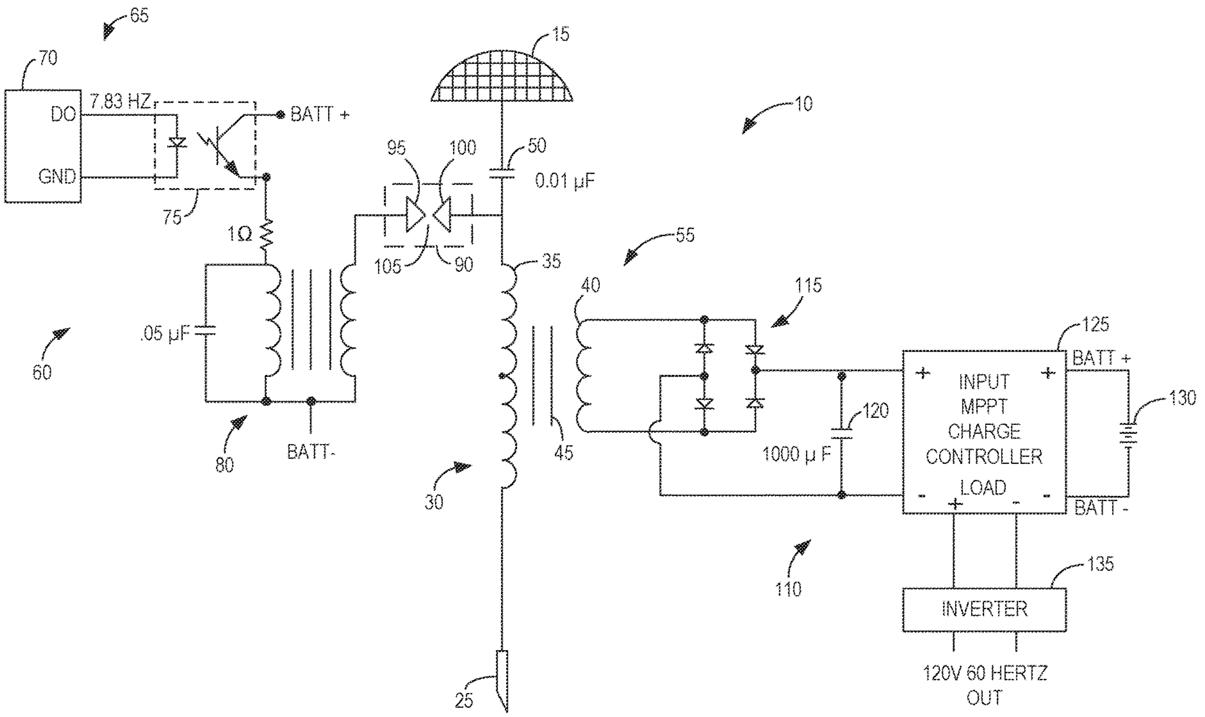
65
70
DO
GND
7.83 HZ
BATT +
75
1Ω
.05 µF
60
80
BATT-
95
100
50
0.01 µF
105
90
15
10
35
55
40
115
30
45
120
1000 µ F
125
BATT +
INPUT MPPT CHARGE CONTROLLER
LOAD
BATT -
130
110
135
INVERTER
120V 60 HERTZ OUT
25

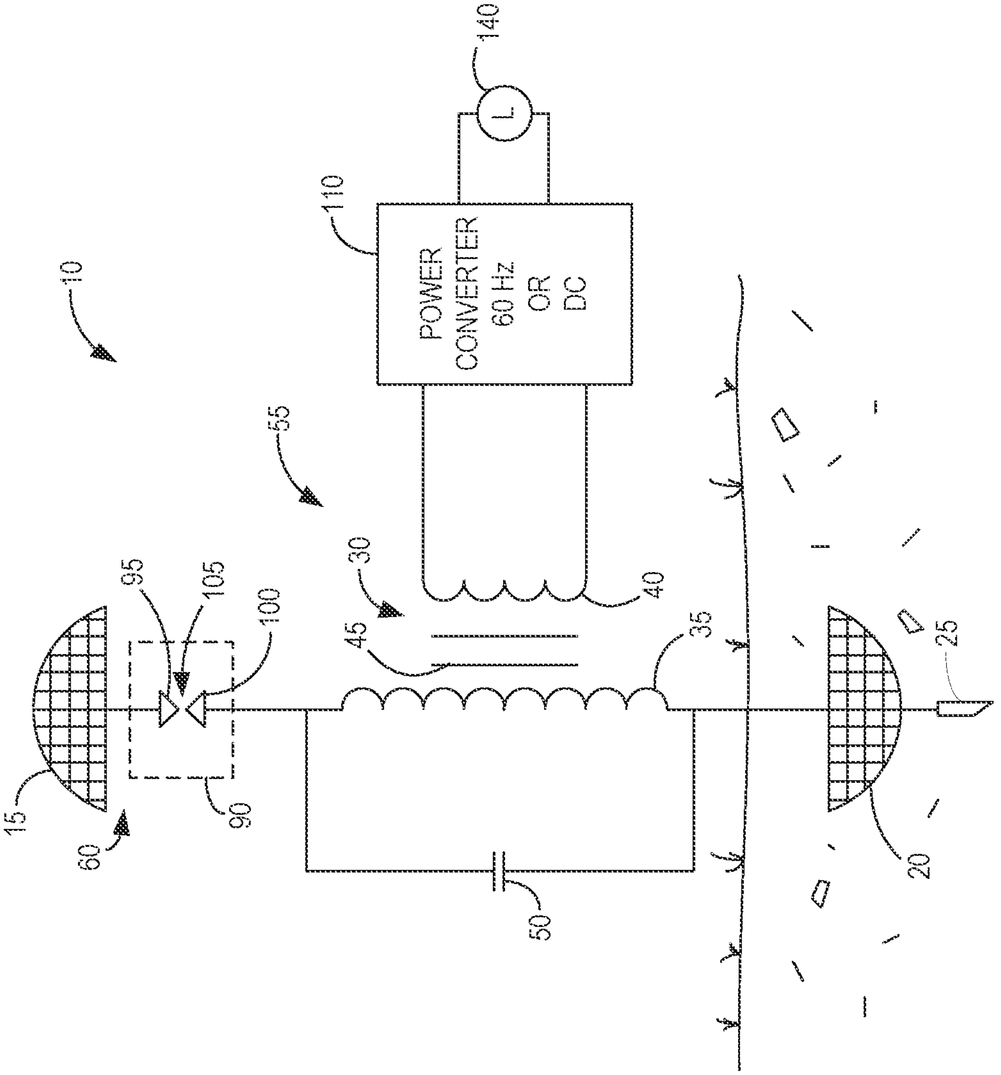
POWER CONVERTER 60 Hz OR DC
L

FIG 1

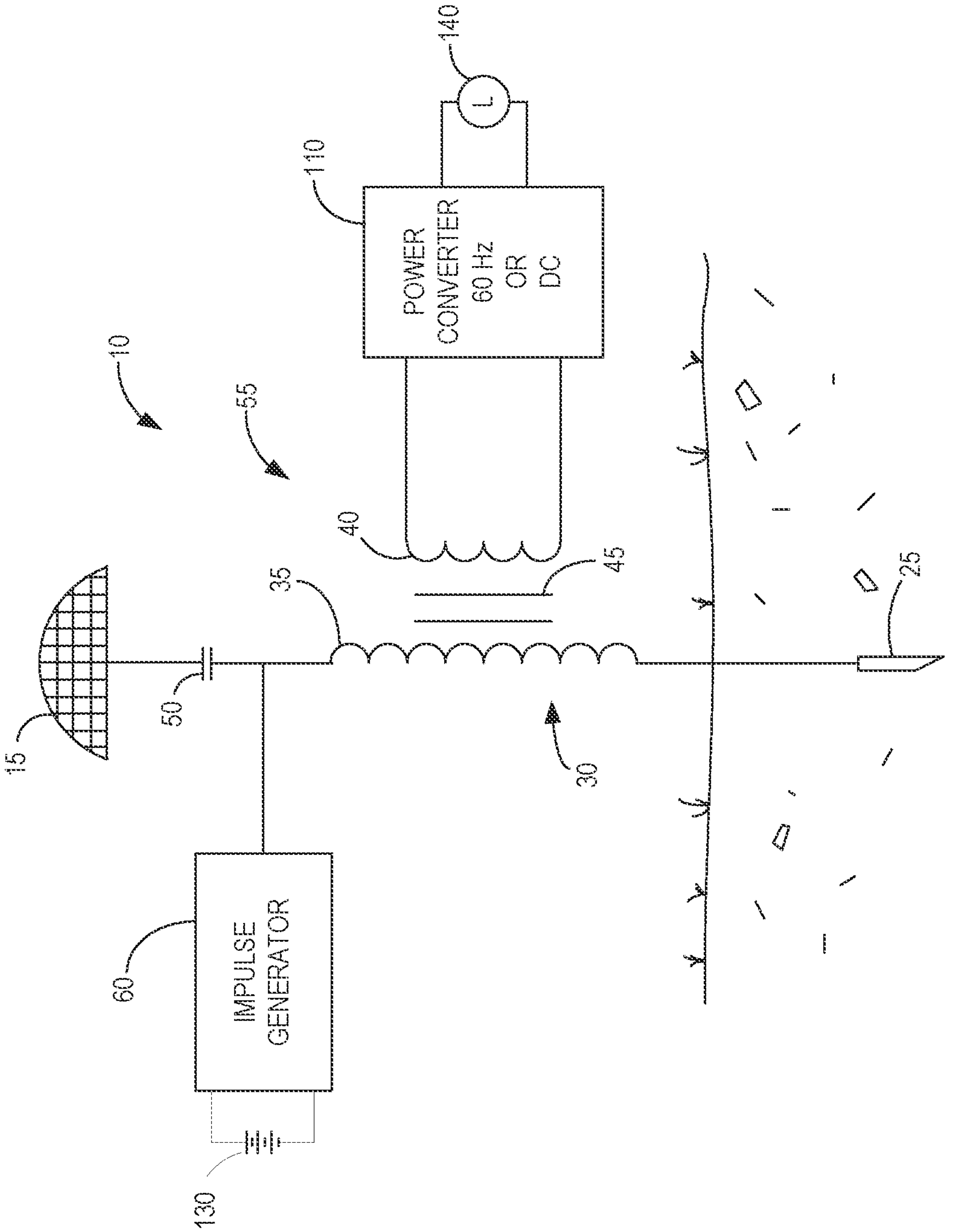
L
POWER
CONVERTER
60 Hz
OR
DC
IMPULSE
GENERATOR

FIG 2

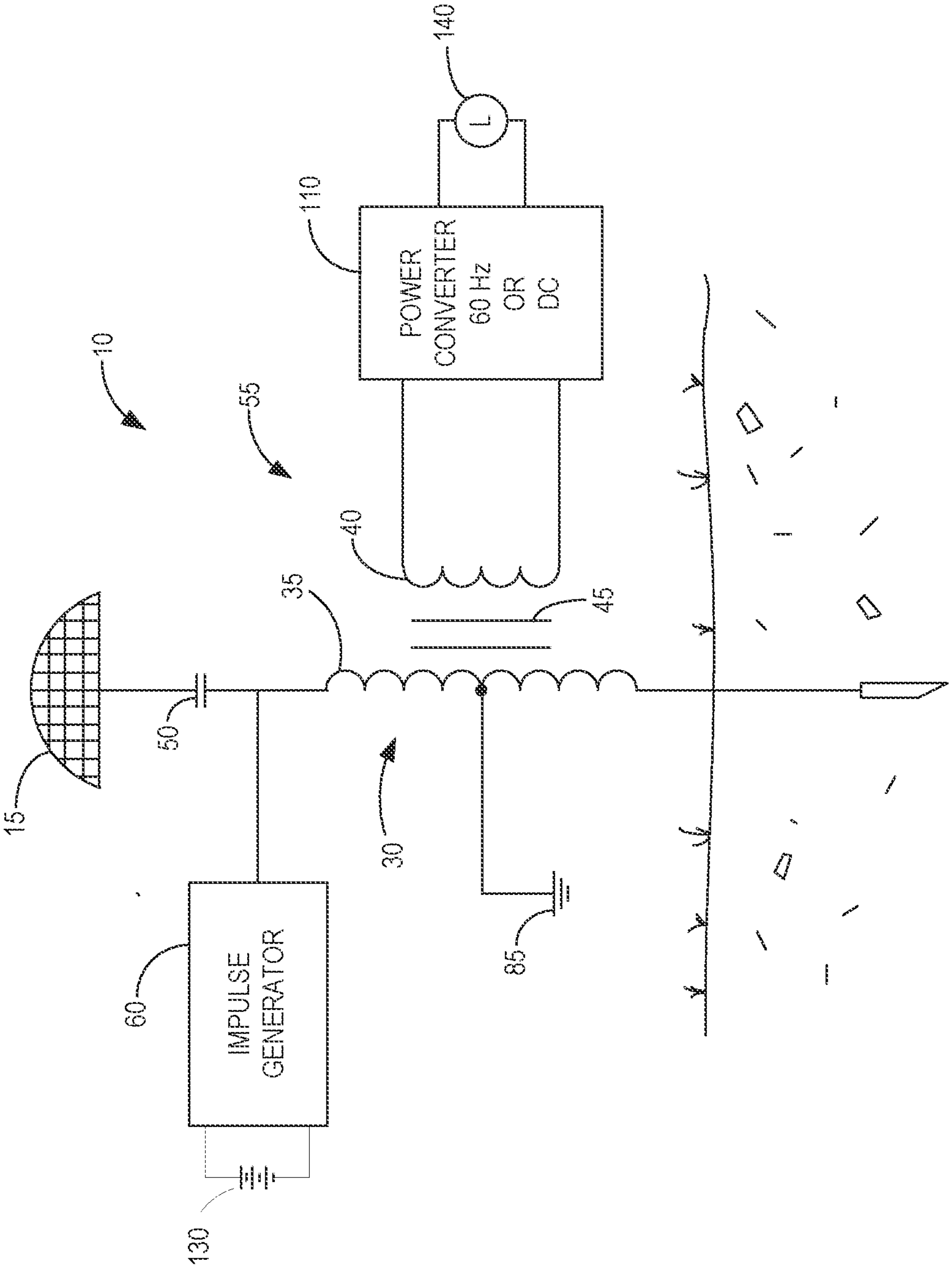
L
IMPULSE
GENERATOR
POWER
CONVERTER
60 Hz
OR
DC

FIG 3

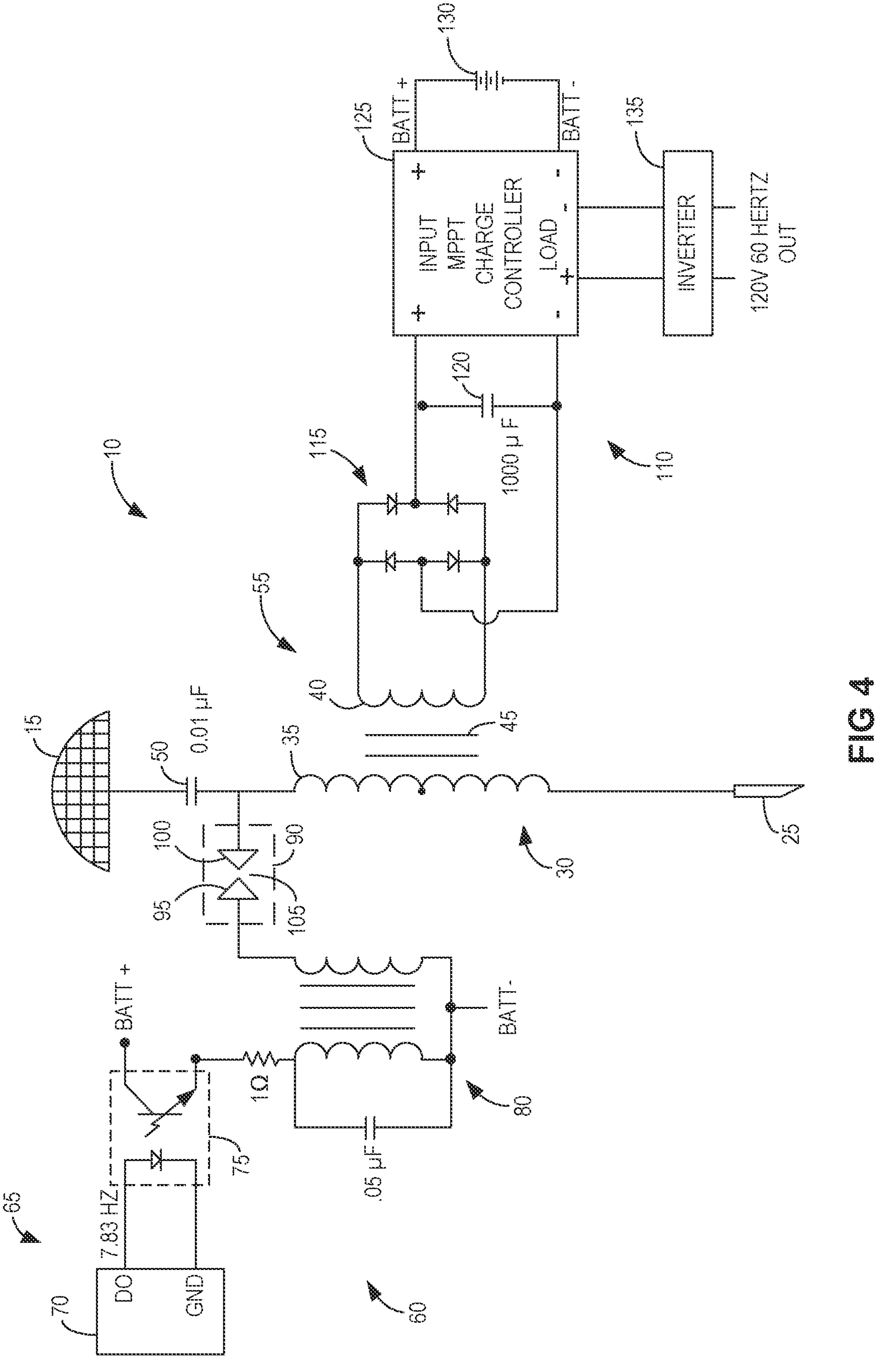
10
15
50
0.01 µF
35
45
40
55
115
120
1000 µ F
110
125
BATT +
130
BATT -
135
INPUT
MPPT
CHARGE
CONTROLLER
LOAD
+
-
+
-
+
-
INVERTER
120V 60 HERTZ
OUT
100
90
95
105
30
25
BATT +
BATT-
1Ω
.05 µF
80
75
7.83 HZ
DO
GND
70
60
65

FIG 4

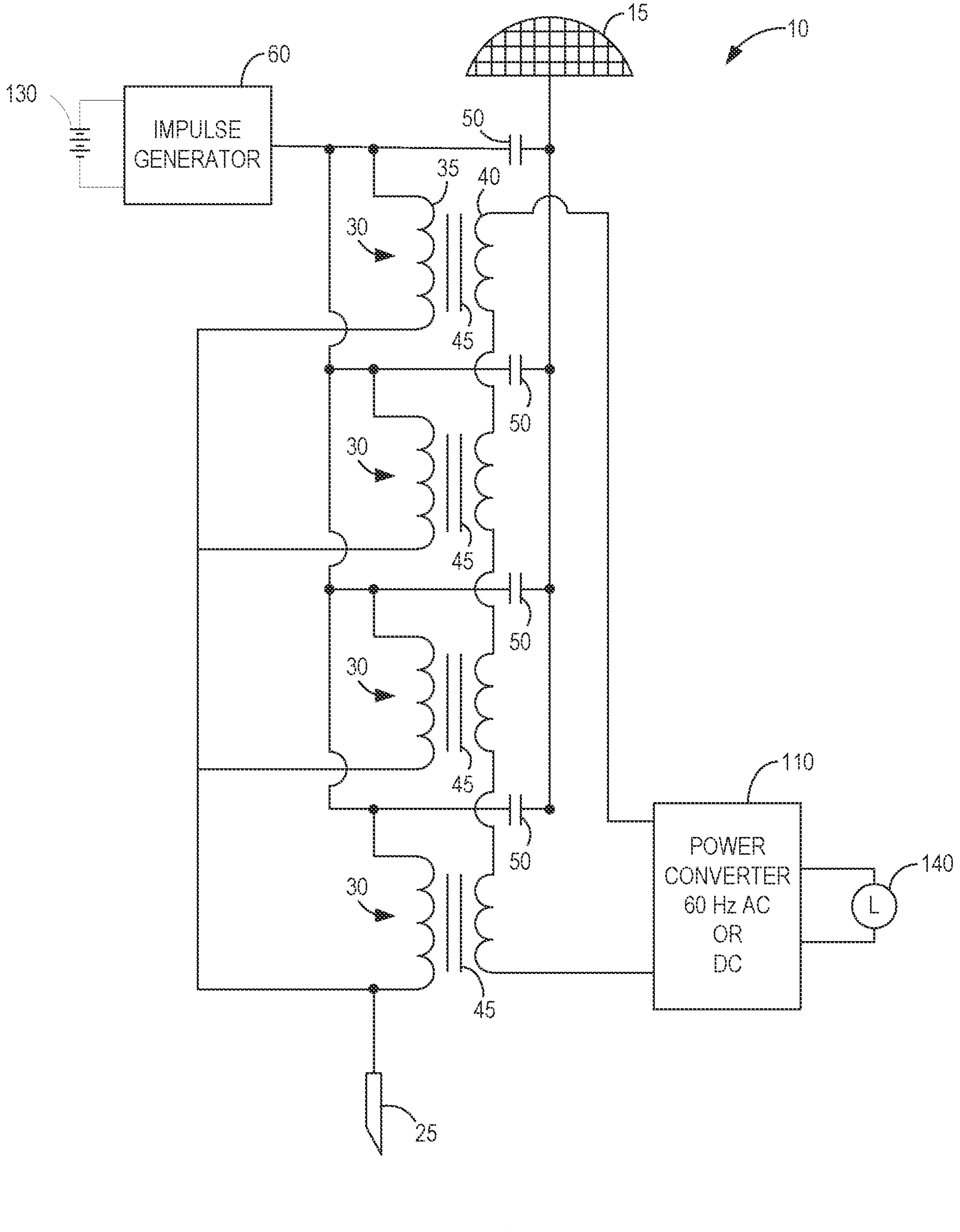
IMPULSE GENERATOR
POWER CONVERTER 60 Hz AC OR DC
L

FIG 5

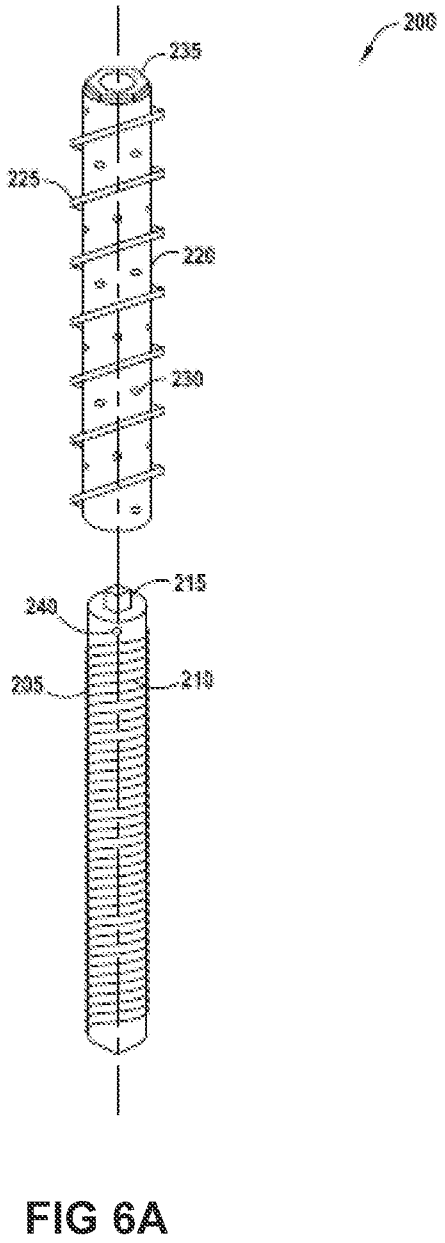
200
235
225
220
230
215
240
205
210

FIG 6A

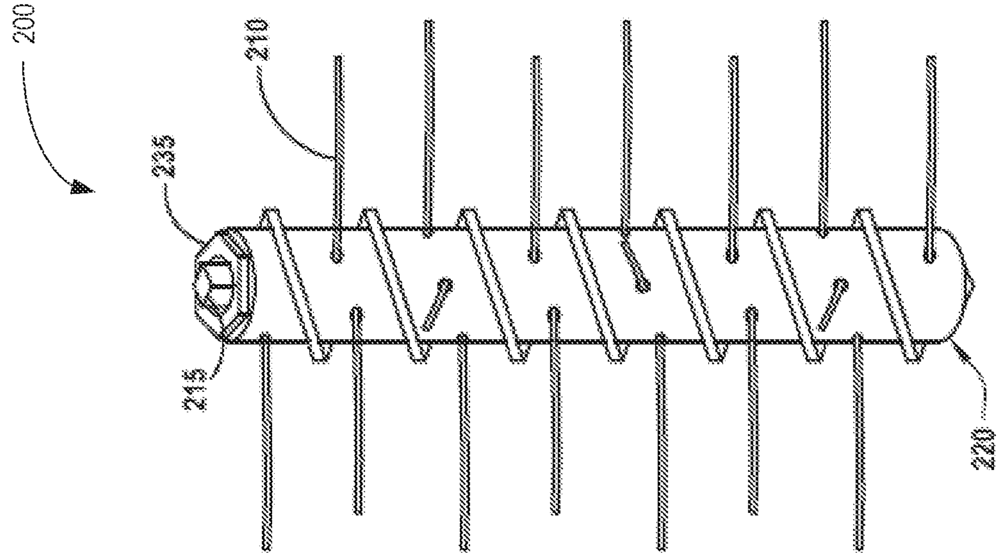
200
235
210
215
220

FIG 6C

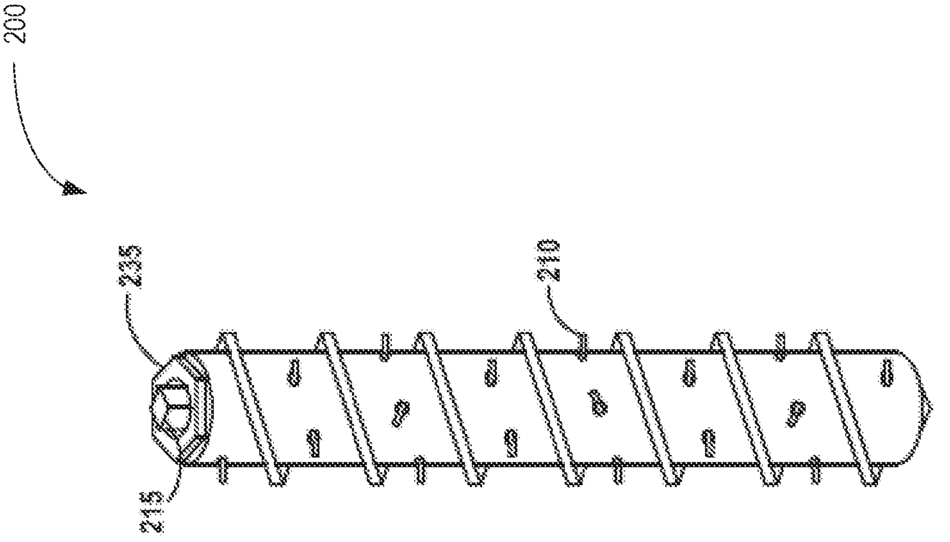
200
235
210
215

FIG 6B

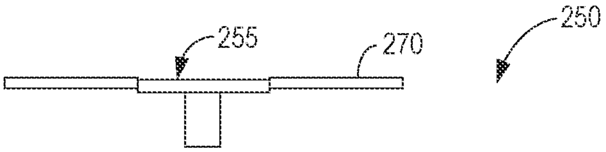

FIG 7A

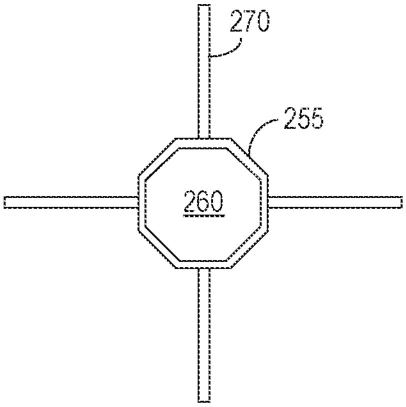

FIG 7B

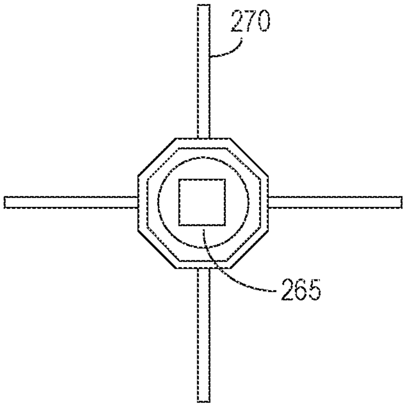

FIG 7C

POWER RECEIVER FOR EXTRACTING POWER FROM ELECTRIC FIELD ENERGY IN THE EARTH

[0001] This application is a continuation of U.S. application Ser. No. 16/521,764, filed Jul. 25, 2019, which is a continuation of U.S. application Ser. No. 15/397,281, filed Jan. 3, 2017, now U.S. Pat. No. 10,389,138, which is a continuation-in-part of U.S. application Ser. No. 14/509,772, filed Oct. 8, 2014, U.S. Pat. No. 9,564,268, which claims the benefit of U.S. Provisional Application No. 61/889,894, filed Oct. 11, 2013, the disclosures of each of which are incorporated by reference herein in their entirety.

TECHNICAL FIELD

[0002] The present invention relates generally to renewable energy, and more particularly to methods and apparatus for extracting energy from subsurface electrical fields beneath the earth's surface.

BACKGROUND

[0003] The earth and the ionosphere cavity may be viewed as a global electric circuit. Electrical currents are constantly flowing within the earth and its atmosphere. Within the earth, the majority of the earth's energy is carried by extremely low frequency (ELF) and ultralow frequency (ULF) waves in the 0-200 Hz frequency range. The earth's rotating magnetic field and positive lightning are two energy sources that sustain the ELF/ULF waves within the earth and the atmosphere.

[0004] A great deal of research has been devoted to studying the electric field present in the earth's ionosphere cavity. Joseph M. Crawley, the "Fair Weather Atmosphere as a Power Source", Proceedings ESA Annual Meeting on Electrostatics 2011; O. Jefimenko, "Operation of Electric Motors from Atmospheric Electric Field," American Journal of Physics, Vol. 39, Pgs. 776-779, 1971; M. L. Breuer, "Usability of Tapping Atmospheric Charge as a Power Source," Renewable Energy, Vol. 28, Pgs. 1121-1127, 2003. Numerous attempts have been made in the past to extract electrical energy from the earth's atmosphere. For example, U.S. Pat. No. 1,540,998 to Plauson describes a system for converting atmospheric electrical energy into usable power. These past attempts have been successful in producing only small amounts of power from the electrical field in the earth's ionosphere cavity. The modest success of these experiments compared to results from other renewable energy sources, such as solar and wind, has tempered further research and prevented widespread use of the electric field in the ionosphere cavity as an energy source.

SUMMARY

[0005] The present invention relates to a power receiver for extracting power from electric fields beneath the earth's surface. In embodiments of the present disclosure, a resonant transformer connected to a ground terminal draws current from the earth's electric field through the primary winding of the transformer. Current flow through the resonant transformer is induced by applying a high voltage impulse to the primary winding. The flow of current through the primary winding of the resonant transformer induces a current in the secondary winding, which may be converted and filtered to a usable form, e.g. 60 Hz AC or DC.

[0006] In some embodiments of the power receiver, the resonant frequency of the resonant transformers is below 200 Hz.

[0007] In some embodiments of the power receiver, the resonant transformer comprises a ferro-resonant transformer.

[0008] In some embodiments, the power receiver further comprises an elevated terminal.

[0009] In some embodiments of the power receiver, the primary winding of the resonant transformer is connected between the ground terminal and elevated terminal.

[0010] In some embodiments of the power receiver, the elevated terminal comprises an upper capacitive plate coupled to the earth's ionosphere cavity.

[0011] In some embodiments of the power receiver, the impulse generator comprises the upper capacitive plate and a spark gap connected between the upper capacitive plate and the primary winding of the resonant transformer. The spark gap comprises a pair of electrodes separated by a gap and configured to generate a spark when a voltage difference between the electrodes reaches a predetermined level.

[0012] In some embodiments of the power receiver, the impulse generator comprises a pulse generator for generating low voltage pulses, a step-up transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses, and a spark gap connected between the step-up transformer and the primary winding of the resonant transformer to generate a spark responsive to the high voltage impulses from the step-up transformer.

[0013] In some embodiments of the power receiver, the impulse generator comprises a pulse generator for generating low voltage pulses, and a step-up transformer connected to the primary winding of the resonant transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses.

[0014] In some embodiments of the power receiver, the impulse generator comprises a solid state spark generator.

[0015] In some embodiments of the power receiver, the resonant transformer includes a capacitor connected in parallel with the primary winding.

[0016] In some embodiments of the power receiver, the resonant transformer includes a capacitor connected in series with the primary winding between the impulse generator and the elevated terminal.

[0017] In some embodiments, the power receiver comprises multiple resonant transformers having primary windings connected in parallel between the ground terminal and the elevated terminal.

[0018] In some embodiments of the power receiver, the resonant transformers have different resonant frequencies.

[0019] In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are all below 200 Hz.

[0020] In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

[0021] Another embodiment of the power receiver comprises a resonant circuit connected to a ground terminal disposed below the surface of the earth, an impulse generator for generating and applying a high voltage electrical impulse to the resonant circuit to induce current flow from the ground terminal through the resonant circuit, and a power conversion circuit connected to the resonant circuit to convert electrical current flowing through the resonant circuit to a desired form. The resonant circuit has a resonant frequency below 200 Hertz.

[0022] In some embodiments of the power receiver, the resonant circuit comprises a resonant transformer having a primary winding, a secondary winding, and resonant capacitor connected in series with the primary winding.

[0023] In some embodiments of the power receiver, the resonant circuit comprises multiple resonant transformers having primary windings connected in parallel to the ground terminal.

[0024] In some embodiments of the power receiver, the resonant transformers have different resonant frequencies.

[0025] In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are all below 200 Hz.

[0026] In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

[0027] Other embodiments of the disclosure comprise a ground terminal for a power receiver. In one embodiment, the ground terminal comprises a ground shaft configured for insertion beneath the surface of the earth, a hollow cylinder surrounding the ground shaft and having a plurality of openings, and a plurality of ground wires connected at one end to the ground shaft. The ground wires are wound around the ground shaft and have free ends protruding through respective openings in the hollow shaft so that rotation of the ground shaft relative to the hollow cylinder causes the ground wires to extend radially into the earth.

[0028] Other embodiments of the disclosure comprise methods of extracting power from the earth. In one embodiment, the method comprises applying a high voltage impulse to resonant circuit coupled to a ground terminal disposed beneath the surface of the earth to initiate resonance in the resonant circuit and induce the flow of current from the ground terminal to the resonant circuit, and converting the current flowing from the ground terminal into the resonant circuit into a useful form.

[0029] In some embodiments of the method, the resonant circuit comprises a resonant transformer including a primary winding coupled to the ground terminal and a second winding coupled to a power converter, and applying a high voltage impulse to resonant circuit comprises applying a high voltage impulse to the primary winding of the resonant transformer.

[0030] In some embodiments of the method, applying a high voltage impulse to the primary winding of the resonant transformer comprises applying an impulse in the range to 10,000 to 40,000 volts to primary winding of the transformer.

BRIEF DESCRIPTION OF THE DRAWINGS

[0031] FIG. **1** illustrates a first exemplary embodiment of a power receiver.

[0032] FIG. **2** illustrates a second exemplary embodiment of a power receiver.

[0033] FIG. **3** illustrates a third exemplary embodiment of a power receiver.

[0034] FIG. **4** illustrates a fourth exemplary embodiment of a power receiver.

[0035] FIG. **5** illustrates a fifth exemplary embodiment of a power receiver.

[0036] FIG. **6**A is an exploded perspective view of an exemplary ground antenna array for the power receiver.

[0037] FIG. **6**B is a perspective view of an assembled ground antenna array before being deployed.

[0038] FIG. **6**C is a perspective view of an assembled ground antenna array after being deployed.

[0039] FIG. **7**A is a side view of an insertion tool for installing the ground antenna array.

[0040] FIG. **7**B is a top view of the insertion tool for installing the ground antenna array.

[0041] FIG. **7**C is a bottom view of the insertion tool for installing the ground antenna array.

DETAILED DESCRIPTION

[0042] Referring now to the drawings, a power receiver for extracting energy from the earth's electric field are illustrated and indicated generally by the numeral **10**. Various embodiments of the power receiver **10** are described and similar reference numbers are used throughout the description to indicate similar components.

[0043] The power receiver **10** converts energy in the ELF/ULF waves to useful form, e.g. 60 Hz AC or DC. The power receiver **10** is essentially a resonance circuit that resonates at the natural resonance frequencies in the earth's electric field. These resonance frequencies, known as Schumann resonance frequencies, occur at 7.83 Hz, 14.3 Hz, 20.8 Hz, 27.3 Hz, and 33.8 Hz. A high voltage impulse initiates resonance within the power receiver **10**. In the resonant mode, the impedance of the power receiver **10** is reduced to near zero thus inducing ground currents to flow into the power receiver **10** where the ground currents are converted to useful form.

[0044] FIG. **1** illustrates a first embodiment of the power receiver **10**. The power receiver **10** comprises a resonant transformer **30** connected between an elevated terminal **15** and ground terminal **25**. In this embodiment, the elevated terminal **15** is capacitively coupled to electric fields within the earth's ionosphere cavity and functions as an upper capacitive plate. A lower capacitive plate **20** is connected to the ground terminal **25** beneath the surface of the earth.

[0045] The resonant transformer **30** comprises a primary winding **35**, secondary winding **40**, ferromagnetic core **45**, and capacitor **50**. One end of the primary winding **35** is connected to the lower capacitive plate **20** and ground terminal **25**. The opposite end of the primary winding **35** is connected via a spark gap **90** to the elevated terminal **15**. The capacitor **50** is connected in parallel with the primary winding **35** of the resonant transformer **30** to form an LC circuit **55** with a resonance frequency range of between about 0.1 and 200 Hz. In a preferred embodiment, the resonant transformer has a Q of about 10 or greater and resonance frequency in the range of about 0.1-200 Hertz. For example, the resonant transformer 30 may have a resonance frequency of about 7.83 Hz, the fundamental Schumann resonance frequency. The secondary winding **40** of the resonant transformer **30** is connected to a power converter **110** as will be hereinafter described in greater detail. The power converter **110** converts the energy extracted from the earth's electric field by the power receiver **10** into a usable form for driving a load **140**.

[0046] The elevated terminal/upper capacitive plate **15** comprises an insulated, dish-shaped plate with a large radius of curvature. The capacitance and resistance of the elevated terminal is chosen for receiving broadband electric field frequencies in the 0-200 Hz range. The upper capacitive plate **15** is sized to maximize to the extent practical coupling with the electric field in the earth's ionosphere cavity.

[0047] The lower capacitive plate **20** is also a dish-shaped plate with a large radius of curvature. One function of the lower capacitive plate **20** is to collect charge from the earth's ground currents and provide an instantaneous source of current as hereinafter described. The capacitance and resistance of the lower capacitive plate **20** is selected to promote the flow of current from the ground with minimal losses.

[0048] The spark gap **90** connected between the elevated terminal **15** and resonant transformer **30** comprises a pair of electrodes **95**, **100** separated by an evacuated air gap **105**. Electrode **95** is connected to the upper capacitive plate **15**. Electrode **100** is connected to the resonant transformer **30**. The spark gap **105** prevents electrical discharge from the upper capacitive plate **15** to the earth's ionosphere cavity. The spark gap **90** in combination with the elevated terminal **15** function as an impulse generator that applies a high voltage impulse of about 10,000-40,000 volts to the primary winding **35** to initiate resonance in the transformer **30**.

[0049] In operation, the capacitive coupling of the upper capacitive plate **15** induces a high voltage operating current in the upper capacitive plate **15**. The upper capacitive plate is connected to a first electrode **95** to the spark gap **90**. When the voltage difference between the electrodes **95** and **100** reaches a threshold, a spark forms across the electrodes **95**,**100** and a high voltage impulse is applied to the primary winding **35** of the resonant transformer **30**. This high voltage impulse initiates resonance within the transformer **30**.

[0050] In resonant mode, the impedance of the resonance transformer is reduced to nearly zero allowing current to flow from the capacitive plate **20** and ground terminal **25** through the primary winding **35** of the transformer **30**, which in turn induces current in the secondary winding **40**. Power converter **110** converts the current flowing through the secondary winding **40** into a usable form for driving a load **140**. The transformer **30** will continue to resonate for a short period of time. By providing high voltage impulses to the primary winding **35** of the resonant transformer **30** at periodic intervals, it is possible to maintain a continuous flow of current from the earth into the resonant transformer **30**, thus producing a continuous supply of power.

[0051] FIG. **2** discloses a second embodiment of the primary receiver **10**. This embodiment includes a resonant transformer **30** connected between an elevated terminal **15** and ground terminal **25**. The resonant transformer **30** comprises a primary winding **35**, secondary winding **40**, ferromagnetic core **45** and a high voltage capacitor **50**. One end of the primary winding **35** is connected to the ground terminal **25**. The opposite end of the primary winding **35** is connected to the elevated terminal **15**. The capacitor **50** has a capacitance of about 0.01 micro-farads. In contrast to the previous embodiment, capacitor **50** is connected in series with the primary winding **35** and elevated terminal **15** and forms a LC circuit **55** with a Q of about 10 or greater and a resonance frequency in the range of about 0.1 to 200 Hz. In a preferred embodiment, the resonance frequency of the transformer **30** is 7.83 Hz, the fundamental Schumann resonance frequency. An impulse generator **60** is connected between the primary winding **35** of the resonant transformer **30** and the series capacitor **50** and applies a high voltage impulse to the primary winding **35** of the resonant transformer **30**. A battery **130** or other external power source supplies power to the impulse generator **60**. As previously described, the high voltage impulse applied by the impulse generator **60** initiates resonance within the resonant transformer **30** inducing current flow from the ground terminal **25** into the primary winding **35** of the resonant transformer **30**. The flow of current from the ground terminal **25** through the primary winding **35** induces current in the secondary winding **40**. Power converter **110** converts the electrical energy in the current flowing through the primary winding **40** into a usable form.

[0052] In contrast to the first embodiment, it is not required to capacitively couple the elevated terminal **15** in the second embodiment to the earth's ionosphere cavity. Rather, the elevated terminal **15** in this embodiment provides lightning protection and dissipates some of the energy flowing into the power receiver **10** to the earth's ionosphere cavity. Also, in contrast to the first embodiment, the capacitor **50** is connected in series between the primary winding **35** of the transformer **30** and the elevated terminal **15**. Those skilled in the art will appreciate that the capacitor **50** could also be connected in parallel rather than series with the primary winding **35** as shown in FIG. **1**. Another difference is that the impulse generator **60** has an external power source. The amount of energy generated by the power receiver **10**, however, is far greater than the energy needed to generate high voltage impulses. The first embodiment does not require an external power source to generate high voltage impulses.

[0053] FIG. **3** illustrates a third embodiment of the power receiver **10**. This embodiment is essentially the same as the embodiment shown in FIG. **2**. The main difference is that a center tap of the primary winding **35** in the resonant transformer **30** is connected to an electrical ground **85**. It should be appreciated that the electrical ground **85** may be different than the earth ground. When the center tap of the resonant transformer **30** is grounded at a distance away from the ground terminal **25** (e.g. 50 ft to 100 ft), the power receiver **10** becomes a transmitter via the ground loop formed.

[0054] FIG. **4** illustrates the power receiver **10** of FIG. **2** in greater detail. The power receiver includes a resonant transformer **30** connected between a ground terminal **25** and an elevated terminal **15**. The ground terminal **25** may comprise a ⅝-inch×8-foot copper ground rod, such as the ERICO 615880UPC. The elevated terminal **15** may comprise a 90% copper mesh formed into a hemisphere with a radius of about 9 inches. The elevated terminal **15** may be elevated at a height of approximately 6 feet above the ground.

[0055] The resonant transformer **30** includes a primary winding **35**, secondary winding **40**, ferromagnetic core **45** and series capacitor **50** configured as previously described. The resonant transformer **30** may have a Q of about **10** and a resonance frequency in the range of about 0.1 to 200 Hz. The resonant transformer **30** may be made using an Allanson transformer (part #1530BP120R) connected in series with a 0.01 micro-farad capacitor, such as the Condensor Products high voltage capacitor (part 190 TC 103-17-125). The resonant transformer **30** is used in a step-down configuration. The center tap of the resonant transformer **30** may optionally be connected to a ground.

[0056] An impulse generator **60** is connected between the primary winding **35** of the resonant transformer **30** and the series capacitor **50** and applies a high voltage impulse in the range of about 10,000 to 40,000 volts to the primary winding of the transformer **30**. A battery **130** or other external power source supplies power to the impulse generator **60**. The power converter **110** connects to the secondary winding **40** of the resonant transformer **30** for converting current in the secondary winding of the transformer to a useful form.

[0057] The impulse generator **60** comprises a pulse generator **65** for generating low voltage pulses, a step-up transformer **80** for converting the low voltage pulses from the pulse generator **65** to high voltage pulses, and a spark gap **90** for generating sparks responsive to the high voltage pulses from the step-up transformer **80**.

[0058] The pulse generator **65** comprises a square wave generator **70**, such as a Sinometer VC2002 function signal generator, and solid state relay **75**. The square wave generator **70** generates a digital pulse stream. In one embodiment, the digital pulse stream generates a square waveform with a frequency of about 7.83 Hz. The frequency of the digital pulse stream is selected to match the resonance frequency of the transformer **30**, though such is not necessarily required. The pulse stream output from the square wave generator **70** is applied to the solid state relay **75**. The solid state relay **75** is connected between a battery or other power source and a first winding of the step-up transformer **80**. The battery may comprise a 12 V, 7.0 A/H sealed lead acid battery, such as the ELB 1270A by Lithonia Lighting. The solid state relay **75** functions as a switch that is activated responsive to the waveform from the square wave generator **70** to provide a continuous stream of low voltage pulses from the battery to the first winding of the step-up transformer **80**. A 1 ohm resistor is connected between the solid state relay **75** and step-up transformer **80**.

[0059] The step-up transformer **80** may comprise a Transco 15 kV, 30 mA neon sign transformer (part #S15612). The step-up transformer **80** converts the low voltage pulses from the pulse generator **65** to high voltage pulses that are applied to the spark gap **90**. The step-up transformer has a 0.5 micro-farad capacitor connected in parallel with the primary winding of the step-up transformer **80**. The step-up transformer produces pulses at the output of about 30,000 to 40,000 volts.

[0060] The spark gap **90** comprises a pair of electrodes **95**, **100** separated by an air gap **105**. A suitable spark gap electrode pair is the Information Unlimited SPARK05 ¼-inch×1-inch tungsten electrodes. As previously described, when the voltage potential between the electrodes **95**, **100** reaches a threshold, a spark forms between the electrodes **95**, **100** and supplies a nearly instantaneous, high voltage impulse to the primary winding **35** of the resonant transformer **30**. This high voltage impulse initiates resonance in the resonant transformer **30** inducing current flow from the ground terminal **25** through the primary winding **35** of the resonant transformer **30**.

[0061] The power converter **110** comprises a bridge rectifier **115**, filter capacitor **120**, charge controller **125**, and inverter **135**. A suitable rectifier is the Micro Commercial Components **10** amp, 1000 volt bridge rectifier (Part #GBJL 1010). The bridge rectifier **115** converts the AC current flowing through the secondary winding **40** of the resonant transformer to a DC current. A filter capacitor **120** removes unwanted frequencies from the DC current. A suitable capacitor **120** is Cornell Dubilier 1000uF 450VDC capacitor (part #383LX102M450N082). The filter capacitor **120** has a capacitance of about 1000 micro-farads. The DC current is input to the charge controller **125**. The charge controller **125** may, for example, comprise a maximum power point tracking (MPPT) charge controller, such as a Tracer **4215** BN MPPT Solar Charge Controller, which is commonly used in solar power generating systems. The charge controller **125** applies a small amount of energy to a battery **30** to charge the battery **130**. As previously noted, the battery **130** serves as a power source for the impulse generator **60**. The remaining current is supplied to an inverter **135**, which converts the DC current to an AC current with a desired voltage and frequency, e.g., 120 volts/60 Hz AC. A suitable inverter **135** is the 1500 W Pure Sine power inverter (AIMS)(part #PWR11500125).The power converter **110** as shown in FIG. **4** may be utilized in the embodiment shown in FIGS. **1**, **2** and **3**.

[0062] FIG. **5** illustrates a power receiver **10** according to another embodiment. The power receiver **10** comprises a plurality of resonant transformers **30** connected between a ground terminal **25** and elevated terminal **15**. Each of the resonant transformers **30** comprises a primary winding **35**, secondary winding **40**, ferromagnetic core **45** and series capacitor **50**. The primary windings **35** of the resonant transformers **30** are connected in parallel. The secondary windings **40** are connected in series with the power converter **110**. An impulse generator **60** applies a high voltage impulse to the primary windings **35** of the resonant transformers **30**. A battery **130** or other external power source supplies power to the impulse generator **60**. The power converter **110** converts the current in the power converter circuit to a usable form for driving a load **140**.

[0063] In one embodiment, each of the resonant transformers **30** shown in FIG. **5** is configured to have a different resonant frequency. In one embodiment, the resonant transformers **30** are configured to resonate at frequencies of 7.83 Hz, 14.8 Hz, 20.3 Hz and 26.8 Hz respectively. Additional resonant transformers **30** could be added to operate at other resonance frequencies.

[0064] FIGS. **6A-6C** illustrate a high quality ground antenna array **200** which may be used as a ground terminal **25**. The ground antenna array **200** comprises a generally cylindrical ground shaft **205** disposed with a hollow cylinder **220** and a plurality of reinforced, heavy gauge ground wires **210** attached at one end to the ground shaft **205**. The ground shaft **205** and ground wires **210** should be highly conductive and have low resistance to supply current from the ground to the power receiver **10**. In one embodiment, the ground wires **210** may be copper or other highly-conductive metal. The end of the ground shaft may be pointed to facilitate insertion into the earth. A connection port on the ground shaft **220** is provided to electrically connect the ground antenna array **220** to the resonant transformer **30**.

[0065] The hollow cylinder **220** has external threads **25** to facilitate insertion into the ground. A rotator nut **235** is fixedly secured to the top end of the hollow shaft **220**. A square shaft **215** protrudes from the top end of the ground shaft **205** into the opening in the rotator nut **235**. FIG. **6B**. A tool **250**, shown in FIG. **7**, engages with the rotator nut **235** and square shaft **215** during insertion of the ground antennas array **200** into the ground as will be hereinafter described.

[0066] The insertion tool **250** is shown in FIG. **7**. The insertion tool **250** includes a tool body **255** having a first socket **260** on one side to fit the rotator nut **235** on the hollow cylinder **220** and a second socket **265** on the other side to fit the square shaft **215** on the ground shaft **205**. Arms **270**

extend from the outer periphery of the tool body **255** for manually or mechanically turning the insertion tool **250**.

[0067] Before the antenna array **200** is deployed, the ground wires **210** are wound around the ground shaft **205** with the free ends protruding slightly from respective openings **230** in the hollow cylinder **220** to a distance not to exceed one half (½) the depth of the external threads **225** on the hollow cylinder **220**. FIG. **6**B illustrates the ground antenna array **200** before deployment. FIG. **6**C illustrates the ground antenna array in a deployed configuration.

[0068] Installation of the ground antenna array **200** is performed in two stages. In the first stage, a hole slightly smaller in diameter than the threads **235** of the hollow cylinder **220** is drilled into the Earth to a depth matching the length of the hollow cylinder **220** or slightly longer. The hole is filled with water and the water is allowed to soak into the soil. After the ground is softened, the hollow cylinder **220** is rotated using the insertion tool **250** to insert the ground antenna array **200** into the ground. The first socket **260** of the insertion tool **250** is engaged with the rotator nut **230** and the insertion tool **250** is turned by hand or a mechanized rotating shaft fitted and attached to the tool arms **270** to thread the ground assembly into the hole. During the initial insertion of the ground antenna array **200**, the ground shaft **205** is fixed to the hollow shaft **220** and rotates with the hollow shaft. The hollow cylinder **220** is rotated until it reaches the full depth of the hole.

[0069] Once the ground antenna array **200** has been fully inserted into the earth, the insertion tool **250** is flipped over and the second socket **265** of the insertion tool **250** is engaged with the square shaft **215**. The insertion tool **250** is turned by hand or a mechanized rotating shaft fitted and attached to the tool arms **270** to rotate the ground shaft **205**. During the second phase, the ground shaft **205** rotates freely inside the hollow cylinder **220**. Rotation of the ground shaft **205** causes the reinforced ground wires **210** to extend radially into the earth. The ground shaft **220** is rotated until the ground wires are fully extended. The ends of the ground wires may be sharpened to aid in the extension of the ground wires during the second phase.

[0070] After the ground antenna array **200** is deployed, a connection cable **280** is attached to a connection port **240** on the ground shaft **220** to electrically connect the ground antenna array **220** to the resonant transformer **30** in the power receiver **10**.

1. A power receiver for extracting electrical energy from the earth's electric field, said power receiver comprising:
- a resonant transformer connected to a ground terminal disposed below the surface of the earth;
- an impulse generator for generating and applying a high voltage electrical impulse to a primary winding of the resonant transformer to induce an alternating current flow from the ground terminal through the transformer; and
- a power converter connected to a secondary winding of the resonant transformer to convert an alternating current flowing through the secondary winding to a direct current.

2. The power receiver of claim **1** wherein a resonant frequency of the resonant transformer is below 200 Hz.

3. The power receiver of claim **1** wherein the resonant transformer comprises a ferro- resonant transformer.

4. The power receiver of claim **1** wherein the impulse generator comprises:
- a pulse generator for generating low voltage pulses;
- a step-up transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses;
- a spark gap connected between the step-up transformer and the primary winding of the resonant transformer to generate a spark responsive to the high voltage impulses from the step-up transformer.

5. The power receiver of claim **1** wherein the impulse generator comprises a solid state spark generator.

6. The power receiver of claim **1** wherein said resonant transformer includes a capacitor connected in parallel with the primary winding.

7. The power receiver of claim **1** wherein said resonant transformer includes a capacitor connected in series with the primary winding.

8. A method of extracting electrical energy from the earth's electric field, the method comprising:
- applying a high voltage electrical impulse to a primary winding of a resonant transformer connected to a ground terminal disposed below the surface of the earth to induce an alternating current flow from the ground terminal through the transformer; and
- converting an alternating current flowing through a secondary winding of the resonant transformer to a direct current.

9. The method of claim **8** wherein a resonant frequency of the resonant transformers is below 200 Hz.

10. The method of claim **8** wherein the resonant transformer comprises a ferro-resonant transformer.

11. The method of claim **8** wherein the impulse generator comprises:
- a pulse generator for generating low voltage pulses;
- a step-up transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses;
- a spark gap connected between the step-up transformer and the primary winding of the resonant transformer to generate a spark responsive to the high voltage impulses from the step-up transformer.

12. The method of claim **8** wherein the impulse generator comprises a solid state spark generator.

13. The method of claim **8** wherein said resonant transformer includes a capacitor connected in parallel with the primary winding.

14. The method of claim **8** wherein said resonant transformer includes a capacitor connected in series with the primary winding.

* * * * *

rimstar.org
... an addiction to science, renewable energy and building stuff

Using earbuds/iphone earpods with a crystal radio

Related links:
[Crystal earphone], [Radio earphone amplifier]

Home
Renewable Energy
Science/Electronics
Projects
Build Van de Graaff
Wimshurst machine
Kelvin water dropper
Electroscope
Corona motors
Ball bearing motor
Crystal radios
Tips and tricks
Crystal earphones
Using earbuds
Razor blade diode
Amplifier for speaker
Amplifier for earphone
Loop antenna radio
AM Radio Transmitter
Fresnel lens
Electrets
Piezoelectric
Electrolytic capacitor
Joule thief
Peltier/Seebeck effect
Franklins bell
Can Stirling
Big Stirling
555 timer music
TEA laser
Laser diode fun
How Ion Prop. Works
Lifter/ionocraft
Star Trek Enterprise
Smoke precipitator
Moving with sound
Straw music instruments
Gravity light - DIY
Nitinol wire
Pyramid hologram
Arduino speech/Talkie
Arduino controlled Skull
Pinhole camera
Wireless electricity
Fly swatter powered CFL
Chlorophyll/fluorescence
Small SG Tesla coil
Photophone
Laser communicator
Hero's steam engine
Alcohol stove
Sunflower heat engine
BB-8 Droid
Ball cyclotron
Atmos. electricity
Photoelectric effect
Bottle rocket
Neural networks
Ornithopter
Curie temp. experiment
Lightning in bottle
Gyroscope - vinyl records
Raspberry Pi
Pepper's ghost phone
Potato chip circuit
Ion wind rotor
LED flashlight
Resources
Power supplies
HV Probes/measuring
Miscellaneous
Efficiency & Conservation
Composting
Food
Non-conv. Energy
Non-conv. Propulsion
Space
Stories
Donate

icims
iCIMS Employee Onboarding Software
Simplify Onboarding and Decrease Turnover
VIEW DEMO

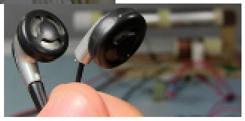

To use earbuds/iPhone earpods with a crystal radio you need to insert a transformer where the crystal earpieces would normally connect to a crystal radio. The jack for the earbuds would be connected to the transformer. Technically speaking, this is for purposes of impedance matching. Below is a diagram for one such crystal radio and some sample photos.

Crystal radio with crystal earpiece.

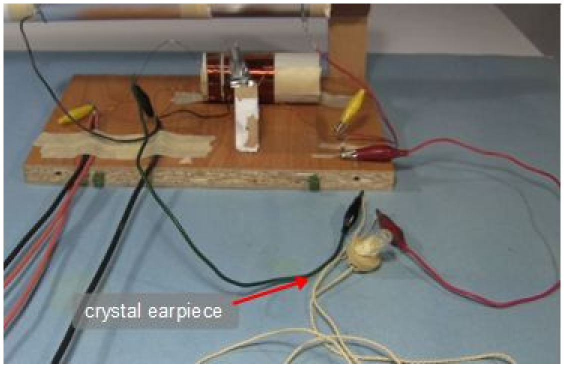
crystal earpiece

Crystal radio with transformer and jack going to earbuds.

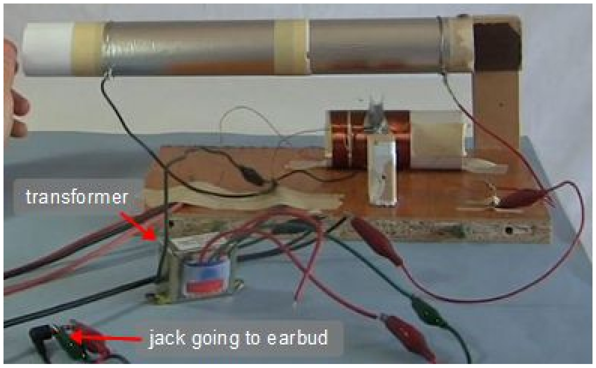
transformer
jack going to earbud

Diagram with crystal earpiece.

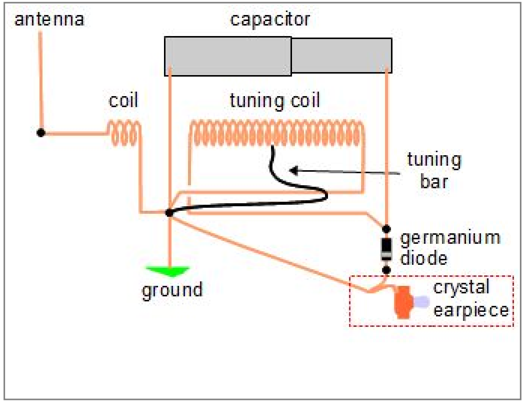
antenna
capacitor
coil
tuning coil
tuning bar
germanium diode
ground
crystal earpiece

Diagram with transformer and jack going to earbuds.

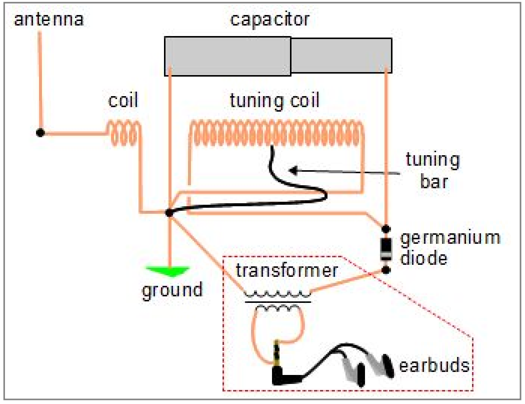
antenna
capacitor
coil
tuning coil
tuning bar
germanium diode
ground
transformer
earbuds

In the photo below are shown the transformers I've tried. Note that the transformer needs to be a *step-up* or *step-down* transformer. I'm not certain that a one-to-one transformer will work. The three smaller ones all gave around the same results. The large microwave oven transformer didn't work as well; the sound was too quiet.

Transformers I've tried with earbuds.

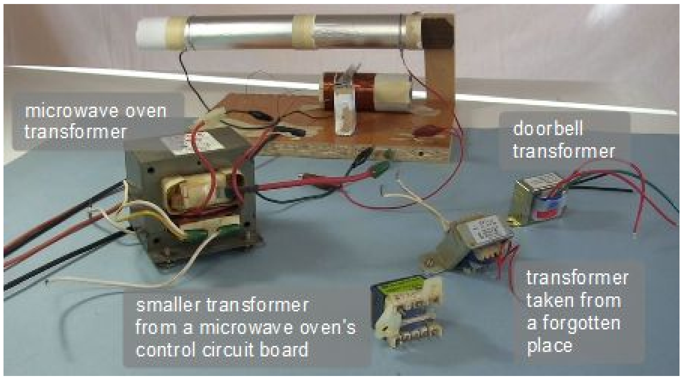
microwave oven transformer
doorbell transformer
smaller transformer from a microwave oven's control circuit board
transformer taken from a forgotten place

Connecting to the earbud/earpod

Earbud/earpod connectors usually have three places to connect on them. As shown in the following diagram, the tip is for one earbud, the middle is for the other earbud and the place nearest the plastic molded casing is common to the two.

Where to connect to the earbud jack.

AdChoices
Headphones
Model Kits
PRECISION DESIGN PREMIER SOLUTIONS
RC Networks & RC Filters
Custom designed to prevent or minimize the occurrence of arcing & noise generations

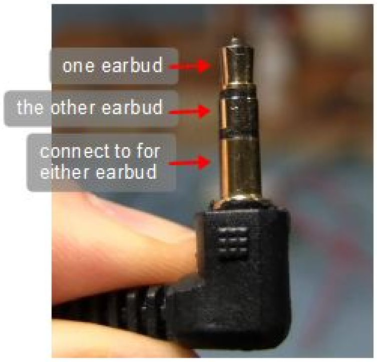
one earbud
the other earbud
connect to for either earbud

For connecting to one earbud only.

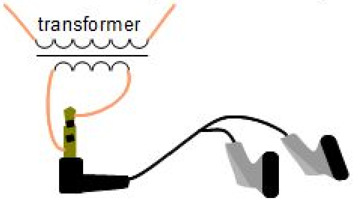
transformer

For connecting to both earbuds.

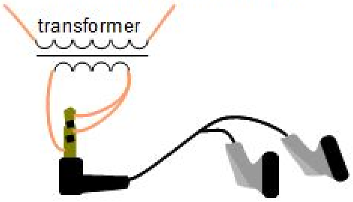
transformer

How to determine which side of the transformer to use where

A transformer has two coils/sides and each coil has two wires coming from it. You need to connect the two wires from one side to the crystal radio and the two wires from the other side to the earbuds.

Here are some ways to determine which side goes where. See the corresponding diagrams and photos below. The side that connects to the crystal radio is the side:

- that's labelled with the higher voltage,
- that has the coil with the most turns, or
- that has the highest resistance.

Determining using labelled input/output voltages.

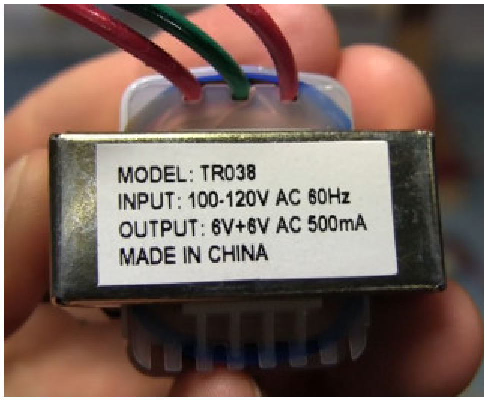
MODEL: TR038
INPUT: 100-120V AC 60Hz
OUTPUT: 6V+6V AC 500mA
MADE IN CHINA

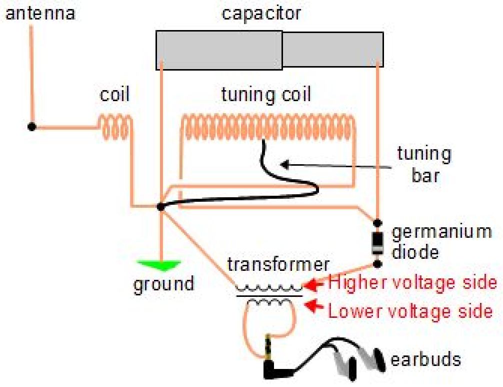
antenna
capacitor
coil
tuning coil
tuning bar
germanium diode
ground
transformer
Higher voltage side
Lower voltage side
earbuds

Determining using number of turns for the coils.

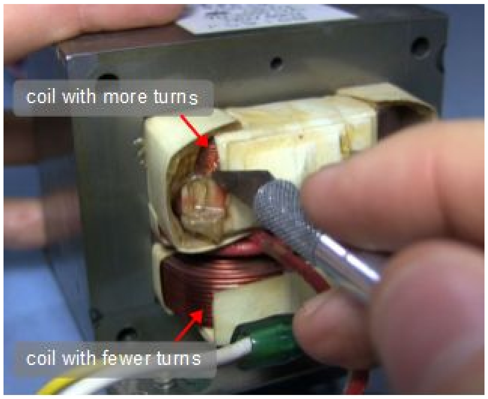
coil with more turns
coil with fewer turns

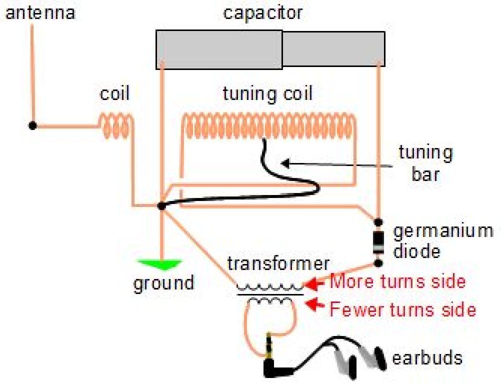
antenna
capacitor
coil
tuning coil
tuning bar
germanium diode
ground
transformer
More turns side
Fewer turns side
earbuds

Determining using resistance of the coils.

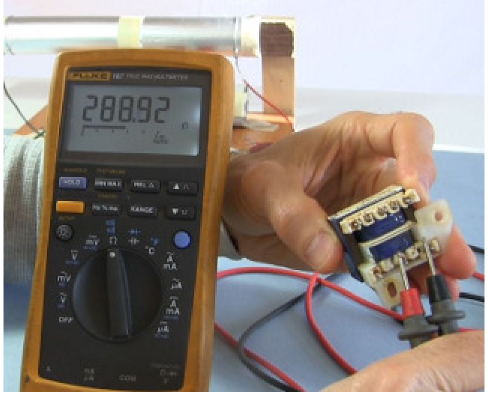

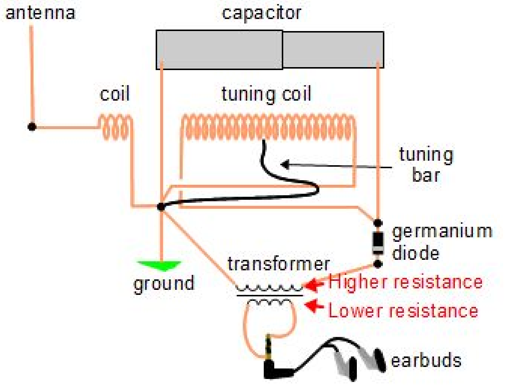
antenna
capacitor
coil
tuning coil
tuning bar
germanium diode
ground
transformer
Higher resistance
Lower resistance
earbuds

Video - Use earbuds/earphones with a crystal radio

The following is a video showing all of the above - basically how to make earbuds/earpods work with a crystal radio using a transformer.

iCIMS
iCIMS Employee Onboarding Software
Simplify Onboarding and Decrease Turnover
VIEW DEMO

Liked this? Share it with: Digg del.icio.us StumbleUpon Google Donate
Related links: [Crystal earphone], [Radio earphone amplifier]
rimstar.org - Share your project on rimstar.org - About - Privacy policy - © 2011 Steven Dufresne -
Contact: stevend@rimstar.org

Ancient Flying Machines

Vimanas - Vaimanika Shastra

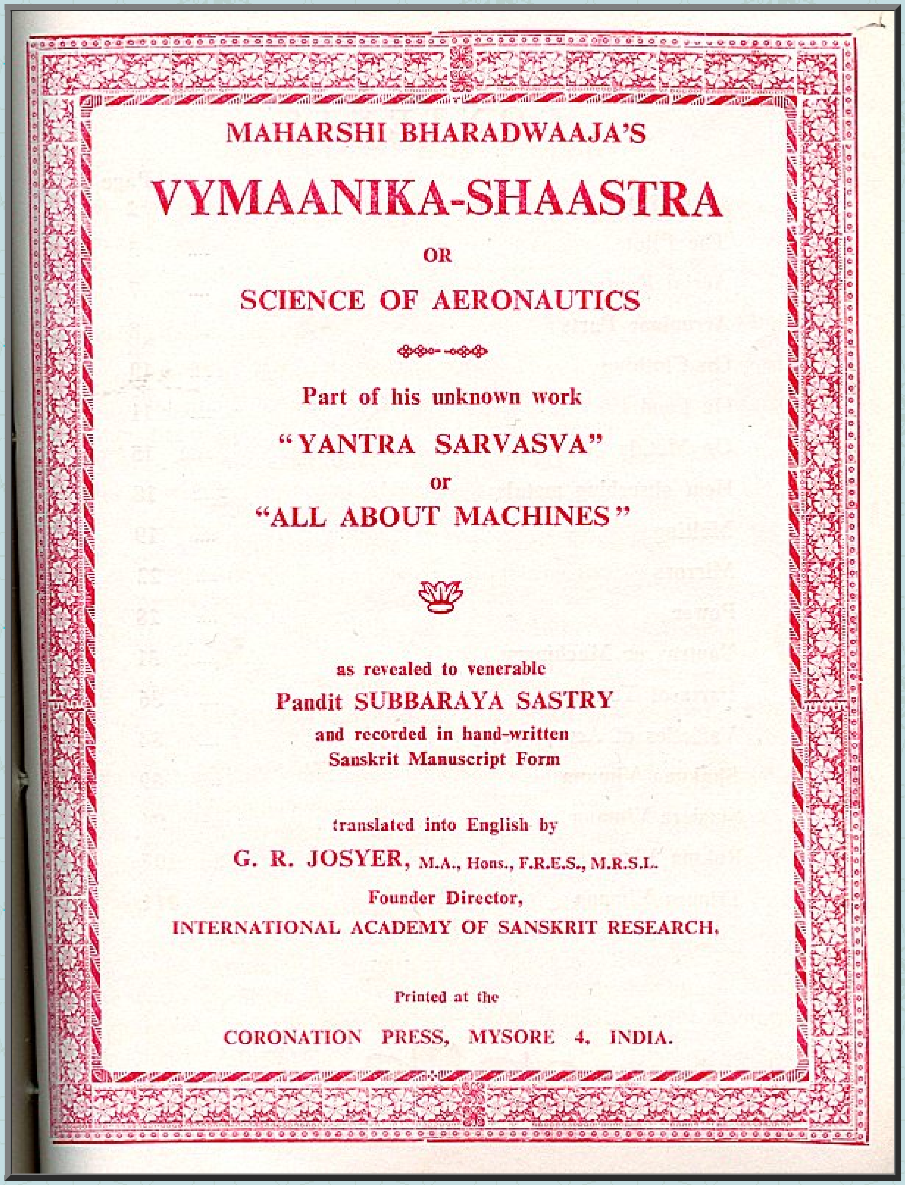

MAHARSHI BHARADWAAJA'S

VYMAANIKA-SHAASTRA

OR

SCIENCE OF AERONAUTICS

Part of his unknown work

"YANTRA SARVASVA"

or

"ALL ABOUT MACHINES"

as revealed to venerable

Pandit SUBBARAYA SASTRY

and recorded in hand-written

Sanskrit Manuscript Form

translated into English by

G. R. JOSYER, M.A., Hons., F.R.E.S., M.R.S.L.

Founder Director,

INTERNATIONAL ACADEMY OF SANSKRIT RESEARCH.

Printed at the

CORONATION PRESS, MYSORE 4. INDIA.

Title page of the English translation of Vyamanika Shastra published in 1973

Vaimanika Shastra

The Vaimānika Shāstra वैमानिक शास्त्र ("Science of Aeronautics" [1]; also *Vimanika, Vymanika*) is an early 20th century Sanskrit text on aeronautics, claimed to be obtained by mental channeling, about construction of *vimāna*s, the "chariots of the Gods", self-moving aerial cars mentioned in the Sanskrit epics.

The existence of the text was revealed in 1952 by G. R. Josyer, according to whom it is due to one Pandit Subbaraya Shastry, who dictated it in 1918-1923. A Hindi translation was published in 1959, the Sanskrit text with an English translation in 1973. It has 3000 shlokas in 8 chapters and was attributed by Shastry to Maharishi Bharadvaja, [2] which makes it of purportedly "ancient" origin, and hence it has a certain notability in ancient astronaut theories.

A study by aeronautical and mechanical engineering at Indian Institute of Science, Bangalore in 1974 concluded that the aircraft described in the text were "poor concoctions" and that the author showed a complete lack of understanding of aeronautics. [3]

Origin and publication

Subbaraya Shastry was a mystic from Anekal, who was reputed to speak out verses (*slokas*) whenever he got inspiration, described by Josyer as "a walking lexicon gifted with occult perception". According to Josyer, he dictated the text to G. Venkatachala Sharma in the early 1900s (completing it in 1923).

Subbaraya Shastry died in 1941, and Venkatachala took his manuscripts into keeping. The Vaimanika Shastra manuscript appeared at Rajakiya Sanskrit Library, Baroda by 1944.[4] The text was published in Hindi in 1959 [5] and later in English by G.R. Josyer, titled *Vymanika Shastra*. Josyer's edition, also added illustrations drawn by T. K. Ellappa, a draughtsman at a local engineering college in Bangalore, under the direction of Shastry, which had been missed in the 1959 edition. [6]

Its existence was first announced publicly in a 1952 press release by G.R. Josyer, who had founded his "International Academy of Sanskrit Research" in Mysore the year before. In the foreword to the 1973 publication that contained the full Sanskrit text with English translation, Josyer quotes a 1952 press release of his which was "published in all the leading dailies of India, and was taken up by Reuter and other World Press News Services":[7]

> **Mr. G. R. Josyer, Director of the International Academy of Sanskrit Research in Mysore, in the course of an interview recently, showed some very ancient manuscripts which the Academy had collected. He claimed that the manuscripts were several thousands of years old, compiled by ancient rishis, Bharadwaja, Narada and others, dealing, not with the mysticism of ancient Hindu philosophy of Atman or Brahman, but with more mundane things vital for the existence of man and progress of nations both in times of peace and war. [...] One manuscript dealt with Aeronautics, construction of various types of aircraft for civil aviation and for warfare. [...] Mr. Josyer showed some types of designs and drawing of a helicopter-type cargo-loading plane, specially meant for carrying combustibles and ammunition, passenger aircraft carrying 400 to 500 persons, double and treble-decked aircraft. Each of these types had been fully described.**

Josyer then tells how he was visited by "Miss Jean Lyon, journalist of Toronto and New York" for an interview, and how Lyon in her *Just Half a World Away* (1954) concluded that he was "guilty of a rabid nationalism, seeking to wipe out everything since the Vedas".

A critical review pronounced Josyer's introduction to be "least scholarly by any standards." and said that "the people connected with publication – directly or indirectly – are solely to blame either for distorting or hiding the history of the manuscripts." perhaps in an attempt to "eulogise and glorify whatever they can find about our past, even without valid evidence". By tracing the provenance of the manuscript, interviewing associates of S. Shastry (including G. V. Sharma to whom the text was originally dictated), and based on the linguistic analysis of the text, the review concluded that it came into existence sometime between 1900 and 1922.[6]

Structure and content

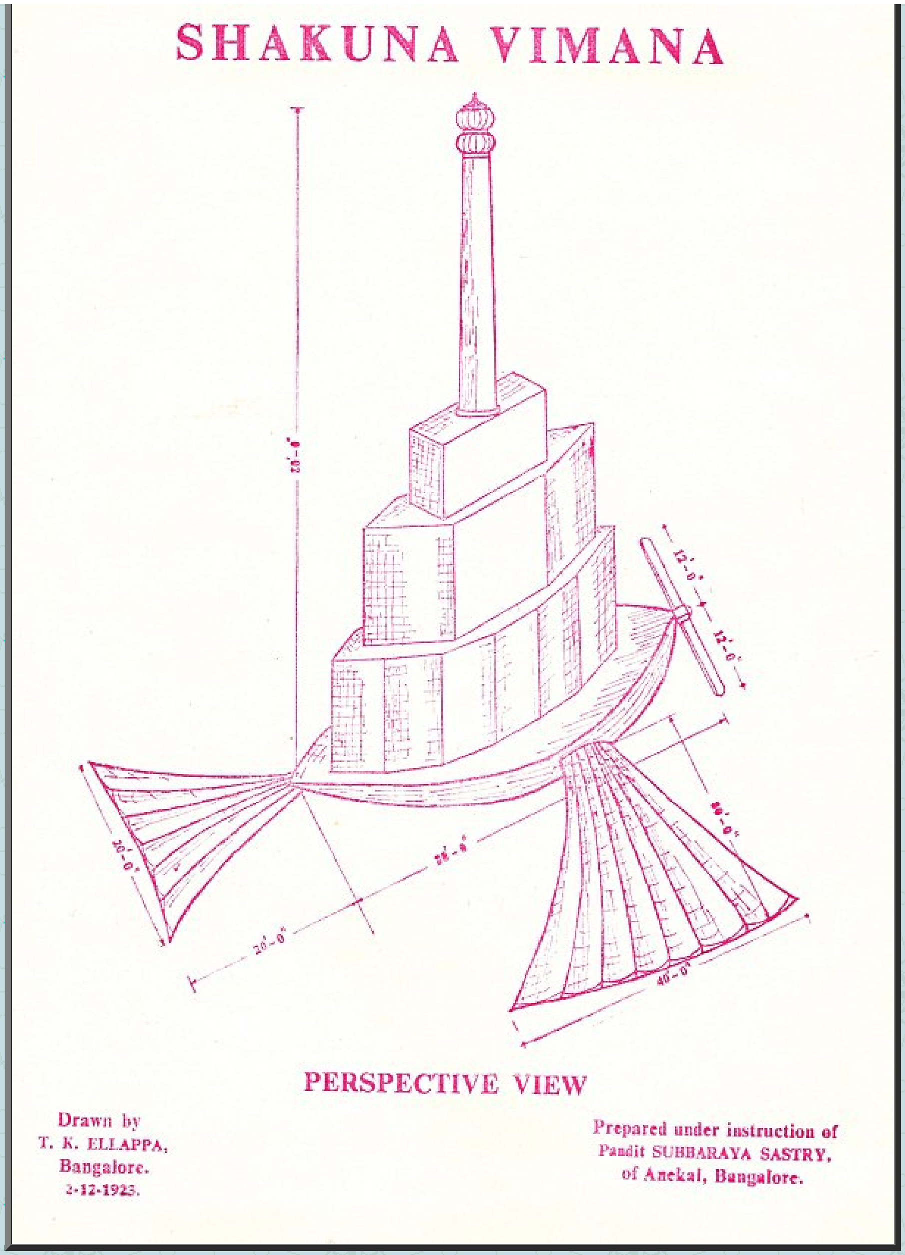

An illustration of the Shakuna Vimana that is supposed to fly like a bird with hinged wings and tail.[6]

Unlike modern treatises on aeronautics that begin by discussing the general principles of flight before detailing concepts of aircraft design, the *Vaimanika Shastra* straightaway gets into quantitative description, as though a particular aircraft is being described. The topics covered include, "definition of an airplane, a pilot, aerial routes, food, clothing, metals, metal production, mirrors and their uses in wars, varieties of machinery and yantras, planes like 'mantrik', 'tantrik', and 'kritak'" and four planes called *Shakuna*, *Sundara*, *Rukma*, and *Tripura* are described in greater detail. The extant text is claimed to be only a small (one-fortieth) part of a larger work *Yantra Sarvaswa* ("All about machines[7]) composed by Maharishi Bharadwaj and other sages for the "benefit of all mankind".[6]

In 1991, the English portion and the illustrations from the Josyer book were reprinted by David Hatcher Childress in *Vimana Aircraft of Ancient India & Atlantis* as part of the *Lost Science Series*. According to Childress, the 8 chapters treat the following:

- **The secrets of constructing aeroplanes, which will not break, which cannot be cut, will not catch fire, and cannot be destroyed.**
- **The secret of making planes motionless.**
- **The secret of making planes invisible.**
- **The secret of hearing conversations and other sounds in enemy places.**
- **The secret of retrieving photographs of the interior of enemy planes**
- **The secret of ascertaining the direction of enemy planes approach.**
- **The secret of making persons in enemy planes lose consciousness.**
- **The secret of destroying enemy planes.**

The propulsion of the Vimanas according to Kanjilal (1985) is by a "Mercury Vortex Engines"[8], apparently a concept similar to electric propulsion. Childress finds evidence for this "mercury vortex engine" in the *Samarangana Sutradhara*, an 11th century treatise on architecture.

J. B. Hare of the Internet Sacred Text Archive in 2005 compiled an online edition of Josyer's 1973 book, in the site's "UFOs" section. In his introduction, Hare writes

> **The *Vymanika Shastra* was first committed to writing between 1918 and 1923, and nobody is claiming that it came from some mysterious antique manuscript. The fact is, there are no manuscripts of this text prior to 1918, and *nobody is claiming that there are*. So on one level, this is not a hoax. You just have to buy into the assumption that 'channeling' works. ... there is no exposition of the theory of aviation (let alone antigravity). In plain terms, the VS never directly explains how Vimanas get up in the air. The text is top-heavy with long lists of often bizarre ingredients used to construct various subsystems. ... There is nothing here which Jules Verne couldn't have dreamed up, no mention of exotic elements or advanced construction techniques. The 1923 technical illustration based on the text ... are absurdly un-aerodynamic. They look like brutalist wedding cakes, with minarets, huge ornithopter wings and dinky propellers. In other words, they look like typical early 20th century fantasy flying machines with an Indian twist.**

A 1974 study by researchers at the Indian Institute of Science, Bangalore found that the heavier-than-air aircraft that the *Vaimanika Shastra* described were aeronautically unfeasible. The authors remarked that the discussion of the principles of flight in the text were largely perfunctory and incorrect, in some cases violating Newton's laws of motion. The study concluded:[9]

> **Any reader by now would have concluded the obvious – that the planes described above are the best poor concoctions, rather than expressions of something real. None of the planes has properties or capabilities of being flown; the geometries are unimaginably horrendous from the point of view of flying; and the principles of propulsion make then resist rather than assist flying. The text and the drawings do not correlate with each other even thematically. The drawings definitely point to a knowledge of modern machinery. This can be explained on the basis of the fact that Shri Ellappa who made the drawings was in a local engineering college and was thus familiar with names and details of some machinery. Of course the text retains a structure in language and content from which its 'recent nature' cannot be asserted. We must hasten to point out that this does not imply an oriental nature of the text at all. All that may be said is that thematically the drawings ought to be ruled out of discussion. And the text, as it stands, is incomplete and ambiguous by itself and incorrect at many places.**

The authors expressed puzzlement at the contradiction and errors in the *Vaimanika Shastra* text, especially since its compilers supposedly had access to publications that did not make such errors (such as, Dayanand Saraswati's commentary on the Rigveda published in 1878 or earlier).[10]

Related Links:

- **A critical study of the work "Vyamanika Shastra""** - ***Scientific Opinion - [PDF][Archived]***
- **Ancient Flying Machines** - ***World Mysteries***
- **Ancient Indian Vimanas and Modern Mercury Vortex Technology**
- **Chariots of the Gods** - ***Erik von Danekin***
- **Flights of fancy?**, ***The Week, 24 June 2001.***
- **Shivkar Bapuji Talpade**
- **Pranava-Vada of Gargyayana** - ***another Sanskrit text channeled in the early 1900s***
- **UFOs and Vimanas** - ***by Stephen Knapp***
- **Vaimanika Shastra - Wikipedia** - ***The propulsion of the Vimanas according to Kanjilal (1985) is by a "Mercury Vortex Engines", apparently a concept similar to electric propulsion***
- **Vaimanika Shastra Rediscovered** - ***Bibliotecapleyades***
  - **Vimanika Shastra / Vymaanika Shaastra** - ***An incomplete version, containing the first six chapters***
  - **The Vimanas - The Ancient Flying Machines**
  - **Ancient Vimana Aircraft**
- **Vedic Vimanas**--***8 Years Before Wright Brothers***
- **Vimana** - ***Wikipedia***
- **Vimanas** - ***Hindu Wisdom***
- **Vimana Aircraft of Ancient India and Atlantis** - ***Author: David Hatcher Childress reviewed by Theresa Welsh***
  - **Vimanas - Ancient Flying Machines** - ***[PDF][Archived]***
- **Vyamanika Shastra: A discussion** - ***ATS Forum***

Notes:

1. **lit. "shastra on the topic of Vimanas"; Shastry & Josyer 1973**
2. **Childress (1991), p. 109**

3. "Flights of fancy? (Part X of XII)". The Week. 2001-06-24
4. Mukunda 1974, p. 2.
5. Swami Brahmamuni Parivrajaka, *Brihad Vimana Shastra*, Sarvadeshik Arya Pratinidhi Sabha. Dayanand Bhavan, New Delhi, 1959.
6. Mukunda 1974
7. Shastry & Josyer 1973
8. Childress (1991), p. 249
9. Mukunda 1974, p. 11
10. Mukunda 1974, p. 12

References:

- Mukunda, H.S.; Deshpande, S.M., Nagendra, H.R., Prabhu, A. and Govindraju, S.P. (1974). "A critical study of the work "Vyamanika Shastra"". *Scientific Opinion*: 5–12
- A.S. Shastry, G.R. Josyer, *Vymanika Shastra - Pronouncements of Maharshi Bhradwaja* (1973) [1][2]
- Dileep Kumar Kanjilal, *Vimana in Ancient India : Aeroplanes Or Flying Machines in Ancient India*, Sanskrit Pustak Bhandar (1985).
- David Hatcher Childress, *Vimana Aircraft of Ancient India*, Adventures Unlimited Press (1991), ISBN 0932813127

SOURCE: Wikipedia Vaimanika_Shastra

Ancient Indian Vimanas and Modern Mercury Vortex Technology

Vimanas

This article deals with vimanas that are running on a mercury vortex propulsion system. This system is developed by, created by Bill Clendenon, an inventor of the mercury vortex propulsion system. Clendenon explains that the vimana was an ancient Indian aircraft, mention in variuous Vedic scriptures. This was a machine driven by a now unconventional manner. The mercury vortex propulsion engine.

The workings of the mercury vortex technology

David hatcher Childress has explained a bit further, the theory of the mercury vortex technology. I will quote it here in full:

- The electromagnetic field coil, which consists of the closed circuit exchanger / condenser coil circuit containing the liquid metal mercury and / or its hot vapor, is placed with its core axis vertical to the craft.
- A ring conductor (directional gyro-armature) is placed around the field coil (heat exchanger) windings so that the core of the vertical heat exchanger coils protrudes through the center of the ring conductor.
- When the electromagnet (heat exchanger coils) is energized, the ring conductor is instantly shot into the air, taking the craft as a complete unit along with it.
- If the current is controlled by a computerized resistance, (rheostat), the ring conductor armature and craft can be made to hover or float in the Earth's atmosphere.
- The electromagnet hums and the armature ring (or torus) becomes quite hot. In fact, if the electrical current is high enough, the ring will glow dull red or rust orange with heat.
- The phenomenon (outward sign of a working law of nature) is brought about by an induced current effect identical with an ordinary transformer.
- As the repulsion between the electromagnet and the ring conductor is mutual, one can imagine the craft being affected and responding to the repulsion phenomenon as a complete unit.
- Lift or repulsion is generated because of close proximity of the field magnet to the ring conductor. Clendenon says that lift would always be opposed to the gravitational pull of the planet Earth, but repulsion can also be employed to cause fore and aft propulsion.'(David hatcher Childress (2000), p.180)

The basic turbo-pump engine has four main sections: compressor, combustion, or heating chambers, turbo-pump and exhaust. Burning gases are exhausted through the turbo-pump wheel to generate power to turn the electric generator:

1. Propellant tanks will be filled with liquid air (obtained directly from the atmosphere by on-board reduction equipment).
2. Liquid air may be injected into expansion chambers and heated by the metal working-fluid mercury confined in a boiler coupled to a heat exchanger.
3. The super heated M.H.D. plasma (or air) will expand through propellant cooled nozzles.
4. The ship may recharge its propellant tanks with liquid air and condensate water collected directly from the upper atmosphere by the on-board reducing plant.

Important is that this mercury vortex propulsion model is intented for terrestrial flight only. The strange ball of light that is often seen by UFO like craft, is the ball of light that surrounds a craft is: the magneto-hydrodynamic plasma, a hot continuously recirculating air flow through the ship's gas turbine which is ionized (electrically conducting). Magnetohydrodynamics (MHD) is described by Childress as an ionized gas that is passed through a magnetic field to generate electricity. (Childress (2000), p.182)

This effect of the ball of light makes the craft to appear alive and breathing. The reason why ships disappear from view is according to Clendenon: The ionized bubble of air surrounding the UFO may be controlled by a computerized rheostat so the ionization of the air may shift through every color of the spectrum, obscuring the aircraft from view. (Childress (2000), p. 181)

Here is a scientific view on the socalled invisiblility of aerial craft. All these ideas are taken from an ancient manuscript the Vimaana Shaastra. In this manuscript there are many detailed characteristists about what could very possible have been mercury vortex generators. These mercury vortex generators could have been used in various types of vimanas.

Dangers of mercury vortex propulsion

Clendenon points out the danger of this mercury vortex propulsion system. When the liquid metal mercury is heated, it gives forth a hot vapor. This hot vapour is deadly poisonous, because; '... if the liquid metal mercury is made radioactive and heated sufficiently to emit radiation, any leaks in the mercury would, therefore, be a double danger to the crew and maintenance personnel of any vehicle powered by a mercury vapor turbine.'(Quoted from the book by Bill Clendenon and David Hatcher Childress, Mercury: UFO Messengers of the Gods, 1990, p. 85)

SOURCE: http://illuminanet.tripod.com/id4.html

UFOs and Vimanas
By Stephen Knapp

In supplying information about the advancements of Vedic science, the subject of Vedic airplanes, vimanas, is almost in a classification of its own. Some of this information is so amazing that for some people it may border science fiction. Nonetheless, as we uncover and explain it, it provides serious food for thought.

First of all we need to understand that the Vedic conception of universal time is divided into different periods. For example, a period called one day of Brahma is equivalent to 4,320,000,000 of our years on earth. Brahma's night is equally as long and there are 360 of such days and nights in one year of Brahma. Each day of Brahma is divided into one thousand cycles of four yugas, namely Satya-yuga, Treta-yuga, Dvapara-yuga, and finally the Kali-yuga, which is the yuga we are presently experiencing. Satya-yuga lasts 1,728,000 years, and is an age of purity when all residents live very long lives and can be fully developed in spiritual understanding and mystical abilities and remarkable powers. Some of these abilities, or mystic siddhis, include changing one's shape, becoming very large or microscopically small, becoming very heavy or even weightless, securing any desirable thing, becoming free of all desires, or even flying through the sky to wherever one wanted to go on one's own volition. So at that time, the need for mechanical flying machines was not necessary.

As the yugas continued, the purity of the people, along with their mystical abilities, decreased by 25% in each age. The age of Treta-yuga lasts 1,296,000 years. During that age, the minds of humanity became more dense, and the ability for understanding the higher spiritual principles of the Vedic path was also more difficult. Naturally, the ability to fly through the sky by one's own power was lost. After Treta-yuga, Dvapara-yuga lasts 864,000 years, and Kali-yuga lasts 432,000 years, of which 5,000 have now already passed. At the end of Kali-yuga, the age of Satya-yuga starts again and the yugas continue through another cycle. One thousand such cycles is one day of Brahma. Now that we are in Kali-yuga, almost all spiritual understanding disappears, and whatever mystical abilities that remain are almost insignificant.

It is explained that it was not until the beginning of Treta-yuga that the development of vimanas took place. In fact, Lord Brahma, the chief demigod and engineer of the universe, is said to have developed several vimanas for some of the other demigods. These were in various natural shapes that incorporated the use of wings, such as peacocks, eagles, swans, etc. Other vimanas were developed for the wiser human beings by great seers of Vedic knowledge.

In the course of time, there were three basic types of vimanas. In Treta-yuga, men were adept in mantras or potent hymns. Thus, the vimanas of that age were powered by means of knowledge of mantras. In Dvapara-yuga, men had developed considerable knowledge of tantra, or ritual. Thus, the vimanas of Dvapara-yuga were powered by the use of tantric knowledge. In Kali-yuga, knowledge of both mantra and tantra are deficient. Thus, the vimanas of this age are known as kritaka, artificial or mechanical. In this way, there are three main types of vimanas, Vedic airplanes, according to the characteristics of each yuga.

Of these three types, there is listed 25 variations of the mantrika vimanas, 56 variations of the tantrica vimanas, and 25 varieties of the kritakaah vimanas as we find today in Kali-yuga. However, in regard to the shape and construction, there is no difference between any of these vimanas, but only in how they were powered or propelled, which would be by mantras, tantras, or mechanical engines.

The controversial text known as Vimaanika Shastra, said to be by Maharshi Bharadwaja, also describes in detail the construction of what is called the mercury vortex engine. This is no doubt of the same nature as the Vedic Ion engine that is propelled by the use of mercury. Such an engine was built by Shivkar Bapuji Talpade, based on descriptions in the Rig-veda, which he demonstrated in Mumbai (Bombay), India in 1895. I more fully explained this in Chapter Three of this volume. Additional information on the mercury engines used in the vimanas can be found in the ancient Vedic text called the Samarangana Sutradhara. This text also devotes 230 verses to the use of these machines in peace and war. We will not provide the whole description of the mercury vortex engine here, but we will include a short part of William Clendenon's translation of the Samarangana Sutradhara from his 1990 book, Mercury, UFO Messenger of the Gods:

"Inside the circular air frame, place the mercury-engine with its electric/ultrasonic mercury boiler at the bottom center. By means of the power latent in the mercury which sets the driving whirlwind in motion, a man sitting inside may travel a great distance in the sky in a most marvelous manner. Four strong mercury containers must be built into the interior structure. When those have been heated by controlled fire from iron containers, the vimana develops thunder-power through the mercury. At once it becomes like a pearl in the sky."

This provides a most simplistic idea of the potential of the mercury engines. This is one kind of a propulsion mechanism that the vimanas of Kali-yuga may use. Other variations are also described. Not only do these texts contain directions on how to make such engines, but they also have been found to contain flight manuals, aerial routes, procedures for normal and forced landings, instructions regarding the condition of the pilots, clothes to wear while flying, the food to bring and eat, spare parts to have, metals of which the craft needs to be made, power supplies, and so on. Other texts also provide instructions on avoiding enemy craft, how to see and hear what occupants are saying in enemy craft, how to become invisible, and even what tactics to use in case of collisions with birds. Some of these vimanas not only fly in the sky, but can also maneuver on land and fly into the sea and travel under water.

There are many ancient Vedic texts that describe or contain references to these vimanas, including the Ramayana, Mahabharata, Rig-veda, Yajur-veda, Atharva-veda, the Yuktilkalpataru of Bhoja (12th century A.D.), the Mayamatam (attributed to the architect Maya), plus other classic Vedic texts like the Satapathya Brahmana, Markandeya Purana, Vishnu Purana, Bhagavata Purana, the Harivamsa, the Uttararamcarita, the Harsacarita, the Tamil text Jivakacintamani, and others. From the various descriptions in these writings, we find vimanas in many different shapes, including that of long cigars, blimp-like, saucer-shapes, triangular, and even double-decked with portholes and a dome on top of a circular craft. Some are silent, some belch fire and make noise, some have a humming noise, and some disappear completely.

These various descriptions are not unlike the reports of UFOs that are seen today. In fact, David Childress, in his book Vimana Aircraft of Ancient India & Atlantis, provides many reports, both recent and from the last few hundred years, that describe eye witness accounts of encounters with UFOs that are no different in size and shape than those described in these ancient Vedic texts. Plus, when the pilots are seen close up, either fixing their craft or stepping outside to look around, they are human-like, sometimes with a Oriental appearance, in clothes that are relatively modern in style. In other reports, we have read where the craft may have alien type beings on board along with ordinary humans navigating the craft.

Does this mean that these are ancient vimanas that still exist today? Are they stored in some underground caverns somewhere? Or are they simply modern-built, using the ancient designs as described in the Vedic texts? The UFOs that have been seen around the world may not be from some distant galaxy, but may be from a secret human society, or even military installation. However, many of the Vedic texts do describe interplanetary travel. So even if these space machines are from some other planet, they may be using the same principles of propulsion that have already been described in the universal Vedic literature. The answer awaits us.

SOURCE: http://www.stephen-knapp.com/ufos_and_vimanas.htm

Vedic Vimanas - 8 Years Before the Wright Brothers
Shivkar Bapuji Talpade

Orville Wright demonstrated on December 17th 1903 that it was possible for a ‘manned heavier than air machine to fly’. But, in 1895, eight years earlier, the Sanskrit scholar Shivkar Bapuji Talpade had designed a basic aircraft called Marutsakthi (meaning Power of Air) based on Vedic technology and had it take off unmanned before a large audience in the Chowpathy beach of Bombay. The importance of the Wright brothers lies in the fact, that it was a manned flight for a distance of 120 feet and Orville Wright became the first man to fly. But Talpade’s unmanned aircraft flew to a height of 1500 feet before crashing down and the historian Evan Koshtka, has described Talpade as the ‘first creator of an aircraft’.

As the world observes the one hundredth anniversary of the first manned flight, it is interesting to consider the saga of India’s 19th century first aircraft inventor for his design was entirely based on the rich treasury of India’s Vedas. Shivkar Bapuji

Talpade was born in 1864 in the locality of Chirabazar at Dukkarwadi in Bombay.

He was a scholar of Sanskrit and from his young age was attracted by the Vaimanika Sastra (Aeronautical Science) expounded by the great Indian sage Maharishi Bhardwaja. One western scholar of Indology Stephen-Knapp has put in simple words or rather has tried to explain what Talpade did and succeeded!

According to Knapp, the Vaimanika Shastra describes in detail, the construction of what is called, the mercury vortex engine the forerunner of the ion engines being made today by NASA. Knapp adds that additional information on the mercury engines can be found in the ancient Vedic text called Samaranga Sutradhara. This text also devotes 230 verses, to the use of these machines in peace and war. The Indologist William Clarendon, who has written down a detailed description of the mercury vortex engine in his translation of Samaranga Sutradhara quotes thus 'Inside the circular air frame, place the mercury-engine with its solar mercury boiler at the aircraft center. By means of the power latent in the heated mercury which sets the driving whirlwind in motion a man sitting inside may travel a great distance in a most marvellous manner. Four strong mercury containers must be built into the interior structure. When these have been heated by fire through solar or other sources the vimana (aircraft) develops thunder-power through the mercury.

NASA (National Aeronau-tical and Space Administra-tion) world's richest/ most powerful scientific organisation is trying to create an ion engine that is a device that uses a stream of high velocity electrified particles instead of a blast of hot gases like in present day modern jet engines. Surprisingly according to the bi-monthly Ancient Skies published in USA, the aircraft engines being developed for future use by NASA by some strange coincidence also uses mercury bombardment units powered by Solar cells! Interestingly, the impulse is generated in seven stages. The mercury propellant is first vapourised fed into the thruster discharge chamber ionised converted into plasma by a combination with electrons broke down electrically and then accelerated through small openings in a screen to pass out of the engine at velocities between 1200 to 3000 kilometres per minute! But so far NASA has been able to produce an experimental basis only a one pound of thrust by its scientists a power derivation virtually useless. But 108 years ago Talpade was able to use his knowledge of Vaimanika Shastra to produce sufficient thrust to lift his aircraft 1500 feet into the air!

According to Indian scholar Acharya, 'Vaimanika Shastra deals about aeronautics including the design of aircraft the way they can be used for transportation and other applications in detail. The knowledge of aeronautics is described in Sanskrit in 100 sections, eight chapters, 500 principles and 3000 slokas including 32 techniques to fly an aircraft. In fact, depending on the classifications of eras or Yugas in modern Kaliyuga aircraft used are called Krithakavimana flown by the power of engines by absorbing solar energies!' It is feared that only portions of Bharadwaja's masterpiece Vaimanika Shas-tra survive today.

The question that comes to one's mind is, what happened to this wonderful encyclopaedia of aeronautical knowledge accumulated by the Indian savants of yore, and why was it not used? But in those days, such knowledge was the preserve of sages, who would not allow it to be misused, just like the knowledge of atomic bombs is being used by terrorists today!

According to scholar Ratnakar Mahajan who wrote a brochure on Talpade. 'Being a Sanskrit scholar interested in aeronautics, Talpade studied and consulted a number of Vedic treatises like Brihad Vaimanika Shastra of Maharishi Bharadwaja Vimanachandrika of Acharya Narayan Muni Viman yantra of Maharish Shownik Yantra Kalp by Maharishi Garg Muni Viman Bindu of Acharya Vachaspati and Vimana Gyanarka Prakashika of Maharishi Dhundiraj'. This gave him confidence that he can build an aircraft with mercury engines. One essential factor in the creation of these Vedic aircraft was the timing of the Suns Rays or Solar energy (as being now utilised by NASA) when they were most effective to activate the mercury ions of the engine. Happily for Talpade Maharaja Sayaji Rao Gaekwad of Baroda a great supporter of the Sciences in India, was willing to help him and Talpade went ahead with his aircraft construction with mercury engines. One day in 1895 (unfortunately the actual date is not mentioned in the Kesari newspaper of Pune which covered the event) before an curious scholarly audience headed by the famous Indian judge/ nationalist/ Mahadeva Govin-da Ranade and H H Sayaji Rao Gaekwad Talpade had the good fortune to see his un manned aircraft named as 'Marutsakthi' take off, fly to a height of 1500 feet and then fall down to earth.

But this success of an Indian scientist was not liked by the Imperial rulers. Warned by the British Government the Maharaja of Baroda stopped helping Talpade. It is said that the remains of the Marutsakthi were sold to 'foreign parties' by the relatives of Talpade in order to salvage whatever they can out of their loans to him. Talpade's wife died at this critical juncture and he was not in a mental frame to continue with his researches. But his efforts to make known the greatness of Vedic Shastras was recognised by Indian scholars, who gave him the title of Vidya Prakash Pra-deep.

Talpade passed away in 1916 un-honoured, in his own country.

As the world rightly honours the Wright Brothers for their achievements, we should think of Talpade, who utilised the ancient knowledge of Sanskrit texts, to fly an aircraft, eight years before his foreign counterparts.

Original from: http://www.deccanherald.com/deccanherald/dec16/snt2.asp

SOURCE: Hindu Answers

FAIR USE NOTICE: This page contains copyrighted material the use of which has not been specifically authorized by the copyright owner. Pegasus Research Consortium distributes this material without profit to those who have expressed a prior interest in receiving the included information for research and educational purposes. We believe this constitutes a fair use of any such copyrighted material as provided for in 17 U.S.C § 107. If you wish to use copyrighted material from this site for purposes of your own that go beyond fair use, you must obtain permission from the copyright owner.

~ MENU ~

Webpages © 2001-2015
Blue Knight Productions

portable
DX
pedestrian mo
contests
vk3ye dot com
cw
antennas
shortwave
CB
videos
HF

[Return to VK3YE amateur radio projects]

A simple crystal set for free power radio

Of any electronic project, the crystal set would have to rate as one of the most popular. Many amateurs are on the air today because of their early construction of a crystal set. Most practical electronic books for beginners include at least one crystal set project. Unfortunately, some of these circuits take simplicity too far and deliver mediocre performance, often by omitting key components such as the tuning capacitor, or failing to provide coil taps.

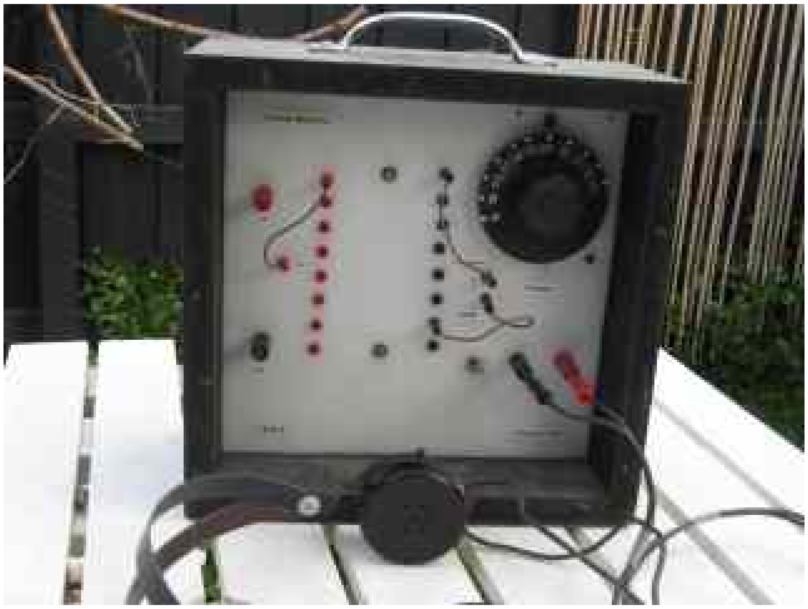

This article describes a crystal set of medium complexity. It features coil taps for the antenna and diode to make it useful for both country and metropolitan listeners. The taps allow the set to cover 160 metres if desired. All parts are easily obtainable, making it a good choice for the beginner. The endless possibilities for experimentation also make crystal sets interesting novelty projects for experienced constructors. The schematic is shown here:

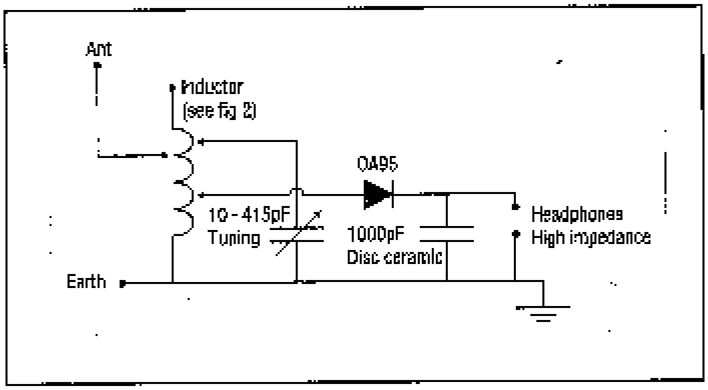
Ant
Inductor
(see fig 2)
OA95
10 - 415pF
Tuning
1000pF
Disc ceramic
Headphones
High impedance
Earth

Figure One: Schematic diagram of crystal set

VK3YE Crystal Set

Schematic & Layout diagrams

Press Back button to return to main article

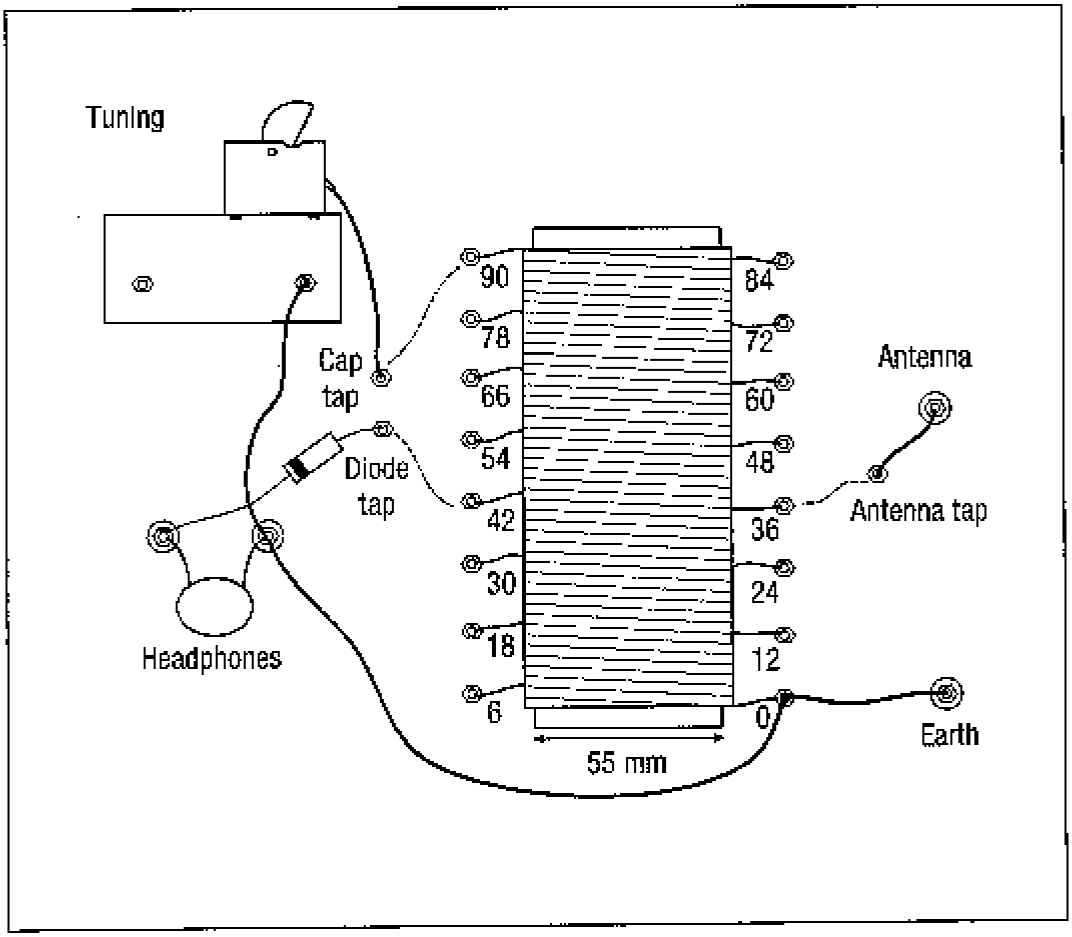
Tuning
Cap
tap
Diode
tap
Headphones
90
84
78
72
66
60
54
48
42
36
30
24
18
12
6
0
55 mm
Antenna
Antenna tap
Earth

Figure Two: Rear view of front panel, showing coil details

(from Amateur Radio February 2001, page 20)

Obtaining the parts

Tuning capacitor

A large air-spaced type, covering about 10 to 415 picofarads is preferred. These capacitors were common in valve and early transistor radios and often appear at hamfests. Their long shafts make it easy to attach large tuning knobs. When purchasing one see that the shaft turns freely, but is not loose. Ensure that the plates are straight and do not touch when meshed - use a multimeter (preferably with audible continuity function) to test this. Avoid capacitors with 3/8 inch shafts unless nothing else is available â€“ knobs for these are not obtainable, and an old valve radio dial drum will need to be used instead.

If a large capacitor is unavailable, a small plastic dielectric unit is suitable (such as stocked by Jaycar and online). The lower maximum capacitance (160 pF) means that more coil turns are required to provide coverage of the lower end of the broadcast band. This can be partially overcome by connecting the 60 and 160 pF sections in parallel (link the 'A' and 'O' tabs). The main disadvantage of these capacitors is their short shafts, which makes it harder to attach most types of knobs.

Vernier Drive and dial

The use of a vernier reduction drive is not necessary. However, its inclusion makes tuning easier, particularly on the higher frequencies. These are rare new but sometimes come up at hamfests. If your drive lacks a dial, one can be fashioned from a plastic or metal disc, such as a jar lid or salvaged computer hard disc. Glue the dial directly to the skirt of the tuning knob if you lack a vernier drive.

Diode

This is the most easily obtainable and cheapest component in the project. A germanium diode, such as a 1N60, 0A90, 0A91, 0A95 or 1N34A will be suitable. The purists still make their own diode detector with a 'cats whisker' and lump of galena, but modern diodes provide more stable and repeatable performance.

Headphones

The very old high impedance headphones are required for this circuit. A minimum of two kilohms is suggested. Medium impedance headphones (approx 600 ohms) will also work, but are less sensitive.

High impedance headphones have become difficult to obtain. Alternatives include:

1. Old-style telephone earpiece. Quite sensitive. Found in the old Telecom corded phones.

2. Crystal earpiece. These are sensitive, easy to obtain and inexpensive. You may need to connect a 100k - 470k resistor in parallel for it to work properly.

3. Piezo transducer. Believe it or not, these actually will work as an earphone. Some sizes even fit snugly in the outer ear in a similar manner to modern earpieces, such as used with mobile phones. The main drawback with transducers is their peaky audio response. Â In some cases it may be necessary to wire a 100k - 470k ohm resistor across the earphone connections for correct operation.

4. 1k to 8 ohm audio transformer and standard low-impedance headphones. Works well, but not as sensitive as a crystal earpiece.

5. Cheating! Use a transistor or IC amplifier kit to run a speaker. This approach eliminates the 'free radio' advantage of the crystal set, but provides louder reception in weak signal areas and allows speaker listening.

Coil and Coil former

This needs to be a cylinder about 55 millimetres in diameter and 150 mm long. The length needs to be long enough to accommodate all ninety coil turns used, with enough left over for mounting to the front panel. Plastic pipe, shampoo container or similar will suffice. Though enamelled copper wire can be used for the winding, the prototype used thin plastic-covered stranded insulated wire.

Front panel

All parts are mounted on a 6mm-thick polyethylene chopping board, which forms the front panel. A hacksaw was used to cut the panel to fit inside the wooden case. Use the thinnest chopping board available so that the many screw-mounted sockets used can be fastened properly. The front panel pictured was cut to 240 mm square.

Case and handle

Use non-metallic material for the enclosure. The box used in the prototype was originally a speaker bought cheaply at a school fete. The lid (which held the speaker) was removed, and the rest of the box painted. The top carry handle is optional and came from a hardware store.

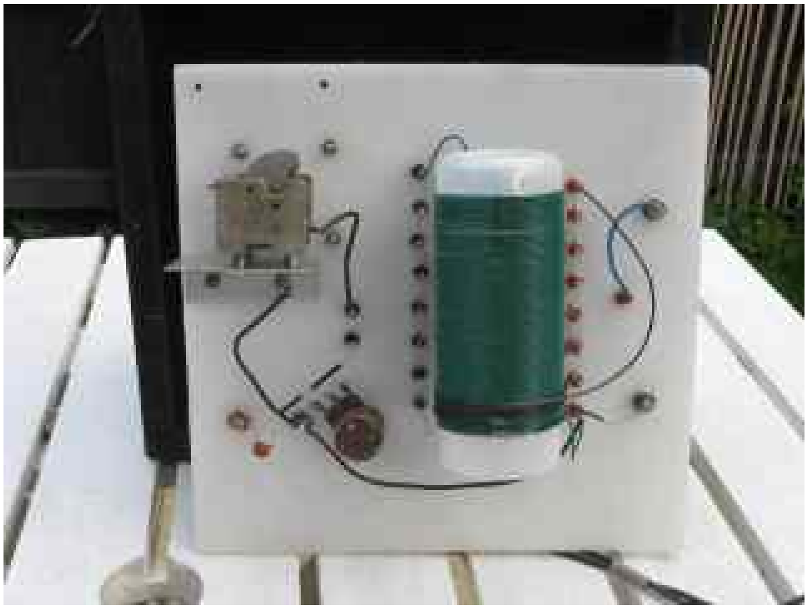

Construction

Commence construction once all components have been obtained. Plan how the parts will fit behind the front panel. The diagram and pictures above show the arrangement used in the prototype. The coil is fastened with stand-offs and the variable capacitor is screwed to an aluminium L-shaped bracket. 4mm binding posts with banana sockets are used for the antenna and headphone connections, and 2mm micro sockets for the coil tapping points.

Start by winding the coil. This consists of ninety turns of thin stranded insulated wire close wound on a plastic tube approximately 55 millimetres in diameter. A large number of tapping points are provided so that the user can vary the set's frequency coverage, and antenna and diode coupling. This makes it possible to obtain the best compromise between volume and selectivity for a particular station.

Figure Two shows the coil construction. Start from the earth end (identified as '0' in the diagram). Make two holes in the former to anchor the end of the wire. Wind six turns and then an extra half-turn. With a knife remove about 1cm of insulation, taking care not to cut the wire. Form the bare wire into a loop and lightly coat with solder. Do not apply excessive heat - the wire insulation easily melts. Wind another five and a half turns and make another tap. Repeat for the remainder of the coil until approximately ninety turns have been wound. Add more turns and taps if using a smaller variable capacitor than specified. Again make two small holes in the former to anchor the wire.

Place the completed coil aside and start work on the front panel. Mount the 4mm banana binding post terminals for the antenna, earth and headphones, as shown in Figure Two. Drill holes and mount the 2mm terminals for the coil taps and the antenna, diode, variable capacitor taps. The tuning capacitor can also be fastened at this time.

Two sets of screws and spacers can be used to mount the coil to the rear of the front panel. A 10mm separation between the coil and the panel is adequate. Solder in the various components and connecting wires as per Fig 2. Use insulated wire for the connections between the sockets and to the variable capacitor. Tinned copper wire can be used for the short links between the coil taps and the 2mm sockets. Use insulated wire for the three jumper leads. The jumpers should be sufficiently long to be able to make connections with all taps along the coil.

This completes the construction. The panel can now be inserted into the box. In the unit pictured, the front panel is recessed â€“ this protects the banana sockets and dial and makes the set more rugged. It also allows attachment of a hinged lid if required.

Parts List

* 10 â€“ 415 pF variable capacitor x1 (see text)

* 0.001 uF disc ceramic capacitor x1

* 1N60 germanium diode x1

* Vernier dial or drive x1 (optional)

* 2mm micro socket x19

* 2mm micro plug x6

* Banana socket (red) x2

* Banana socket (black) x2

* Insulated wire 20m

* Tinned copper or bell wire 1m

Other items: case and handle; polyethylene chopping board; Coil former â€“ 55mm dia, 150 mm long; screws, nuts, washers and spacers; mounting bracket for variable capacitor.

How it works

To receive signals, a radio circuit must perform three functions: selection, detection and reproduction.

The inductor and variable capacitor form a tuned circuit. The role of the tuned circuit is to select one of the many signals present at the antenna. The size of the inductor and capacitor determines the frequencies that can be tuned. The capacitor is made variable to allow the full range of AM broadcast band frequencies (531 to 1602 kHz) to be received.

The diode detector converts the selected radio frequency signal to an electrical current varying at audio frequencies.

The headphones convert this audio frequency energy to sound. The principle is similar to a relay â€“ the signal cause current to flow in a winding that forms an electromagnet. The magnetism generated vibrates the metal diaphragm, thus creating sound. Crystal earpieces perform the same function, but rely on the piezo-electric effect.

Unlike in a conventional radio, which uses amplifying devices such as transistors and integrated circuits, crystal sets are powered by the signal from the incoming station, so no batteries is required. If provided with an efficient antenna and earth, crystal sets can receive signals thousands of kilometres away.

Antenna and earth

A crystal set requires a wire antenna to operate properly. The longer and higher it is the better. A length of at least 10 metres in urban areas, and 20 â€“ 30 metres elsewhere should provide reception in most cases. The antenna should always be installed away from power lines for safety reasons and to reduce interference pick-up. An existing amateur or TV antenna can also be effective, especially if the coaxial feedline is used as part of the antenna. This is achieved by connecting both the outside and the inside of the coaxial connector to the receiver's antenna terminal.

An earth provides stronger signals, and is essential in remote areas. In homes with copper water pipes, this can simply be a lead to the nearest cold water tap. In newer homes, where plastic pipes are used, an outside ground stake can be used instead.

For long distance reception (hundreds or thousands of kilometres) more than usual effort needs to be taken when installing the aerial and earth. Reference One suggests a length of about 100 metres and a height of at least 12 metres. A series of buried radials is suggested for the earth, rather than the water pipe suggested above.

Operation

Connect antenna, earth and earphones. Install the three jumper leads. Set the capacitor tap to near the top of the coil (either the 78th or 90th turn) and the diode and antenna taps to approximately midway along the coil.

In a quiet room, adjust the tuning control and listen for a station. If several stations are audible, move the diode or antenna taps nearer the earth end (lower numbered turns) of the coil. This increases the set's selectivity and makes it possible to separate stations. In a capital city it should be possible to separate at least nine or ten stations. Optimum tap settings vary across the broadcast band â€“ lower frequency stations are often best received with higher tap settings. In rural areas volume is normally more important than selectivity, so the taps can be moved near the top of the coil.

Reception of AM operators on the 1.8 MHz (160 metre) amateur band is possible by moving the capacitor tap lower down the coil, to the 54th or 66th turn. Performance will be well down on a superhet or regenerative receiver, and SSB signals cannot be resolved. Whether you hear amateurs or not depends on your antenna system and the extent of activity from nearby operators. Here in Melbourne 160 metre AM activity includes the Monday to Saturday AM morning net starting at 11am.

In many areas there are narrowcast stations between the top of the official AM broadcast band and 160 metres. Because of their low power these stations will be weaker than the mainstream broadcasters. However these stations are excellent tests of your receiver and antenna system.

Video demonstrations of this project

Conclusion

A crystal set of moderate complexity has been described. It is the minimum required to provide good reception of local stations in urban and rural areas. However numerous refinements to increase sensitivity, selectivity or audio output can be made. These include:

1. Double tuned circuits (with variable coupling between them) to improve selectivity

2. Use of a tuned trap to null out interfering signals

3. Attention to the construction of coils to provide the highest possible Q

4. Addition of an impedance matching network to provide efficient power transfer between the antenna and the tuned circuit

5. Use of a large loop antenna for the coil to allow reception of signals without an external antenna and nulling of unwanted signals

6. Voltage doubler diode detector circuit using two diodes to increase volume

7. Use of DC bias (from a DC voltage applied to the diode) or RF bias (from a locally generated RF signal on the receiving frequency) to improve sensitivity, or, in the case of the latter, to provide CW and SSB reception.

8. Use of a Q multiplier to increase sensitivity and allow CW and SSB reception.

Should you decide to experiment with these changes, it would be desirable to keep this set as a reference and build a second receiver as a test bed for the experiments.

Obtaining the parts

Suitable parts were discussed in detail above. Many if not all can be bought online, with examples presented below.

Variable capacitor and knob

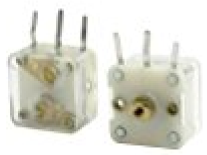

amazon

Variable Capacitor -...

$2.26

Shop now

amazon

Capacitor - Variable,...

$13.95

Shop now

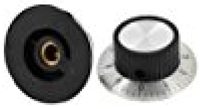

amazon

2 Pcs 24mm Top Potentiometer...

$5.24 Prime

Shop now

Inductor (insulated wire and cylinder former only required only for this project - other crystal sets use ferrite rods)

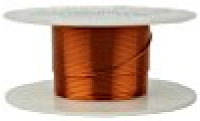

amazon

TEMCo 24 AWG Copper...

$6.09

Shop now

amazon

Remington Industries...

$10.35 Prime

Shop now

amazon

Ferrite Rod, 33 Material,...

$8.95

Shop now

amazon

Creative one 25 Pcs Silver...

$6.99 Prime

Shop now

Other parts and further reading

amazon
Crystal Radio Earphone,...
$8.88 Prime
Shop now

amazon
uxcell 1N60P Germanium...
Shop now

amazon
PIXNOR 1/4W Resistor Kit 56...
$9.99 Prime
Shop now

amazon
SMAKN® LM386 Audio Amplifier...
$6.90 Prime
Shop now

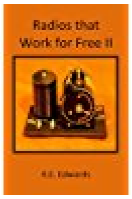

amazon
Radios That Work For Free II
$9.99
Shop now

amazon
Build Your Own Transistor...
$21.19
Shop now

Disclosure: I receive a small commission from items purchased through links on this site.
Items were chosen for likely usefulness and a satisfaction rating of 4/5 or better.

Note: This item is an updated version of an article that first appeared in *Amateur Radio*, December 2000.

Books by VK3YE

amazon

99 things you can do with...
$4.99
Shop now

amazon

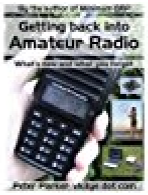

Getting back into Amateur...
$4.99
Shop now

amazon

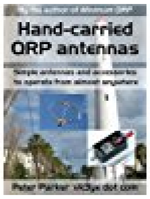

Hand-carried QRP antennas:...
$4.99
Shop now

amazon

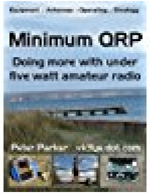

Minimum QRP: Doing more...

$4.99

Shop now

All material on this site (c) Peter Parker VK3YE 1997 - 2017.

Material may not be reproduced without permission.

freshdesk
Take Your Support
Everywhere Your Customers Go.
FREE 30-DAY TRIAL

US 20030114313A1

(19) **United States**

(12) **Patent Application Publication**

(10) **Pub. No.: US 2003/0114313 A1**

(43) **Pub. Date: Jun. 19, 2003**

Worsley et al.

(54) TECHNICAL AND THEORETICAL SPECIFICATIONS FOR WARP DRIVE TECHNOLOGY

(76) Inventors: **Andrew Peter Worsley**, Kent (GB); **Peter John Twist**, Tortola (GB)

Correspondence Address:
LARSON & TAYLOR, PLC
1199 NORTH FAIRFAX STREET
SUITE 900
ALEXANDRIA, VA 22314 (US)

(21) Appl. No.: **10/182,373**

(22) PCT Filed: **Jan. 30, 2001**

(86) PCT No.: **PCT/GB01/00381**

(30) **Foreign Application Priority Data**

Jan. 31, 2000 (GB) 0002221.0
Nov. 24, 2000 (GB) 0028721.9

Publication Classification

(51) **Int. Cl.^{7}** **H01S 1/00**
(52) **U.S. Cl.** **505/180**

(57) ABSTRACT

The present invention relates to the use of technical drive systems, which operate by the modification of gravitational fields. These drive systems do not depend on the emission of matter to create thrust but create a change in the curvature of space-time, in accordance with general relativity. This allows travel by warping space-time to produce an independent warp drive system. Differentials electron flow through a body in rotation is directed so as to simultaneously pass through a said body In its direction of

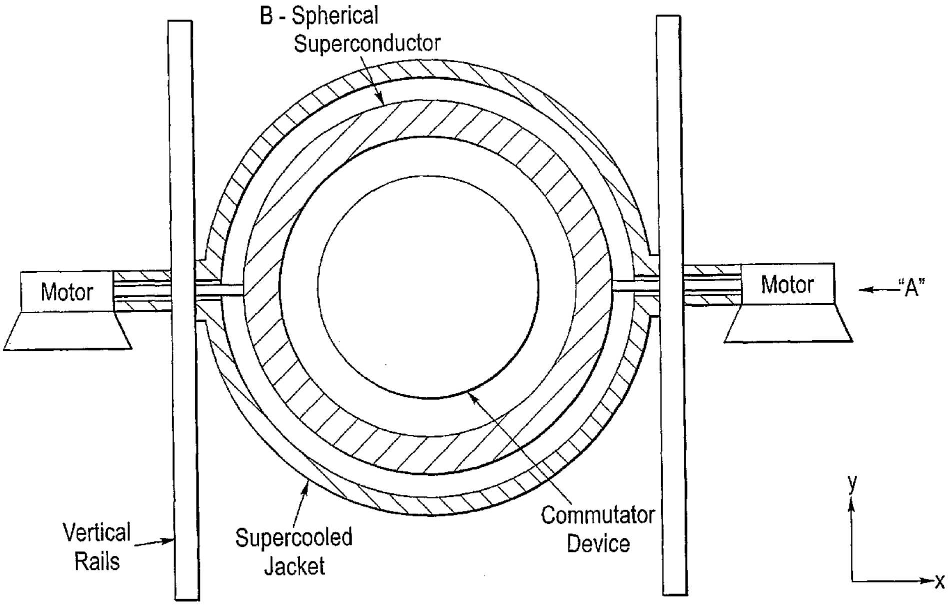

Spherical Superconductor
Sphere = 0.25m external radius
Sphere = 0.20m internal radius

Figure 1 & 2
B = Superconductor

Direction of Motion
y vector (vertical rails)

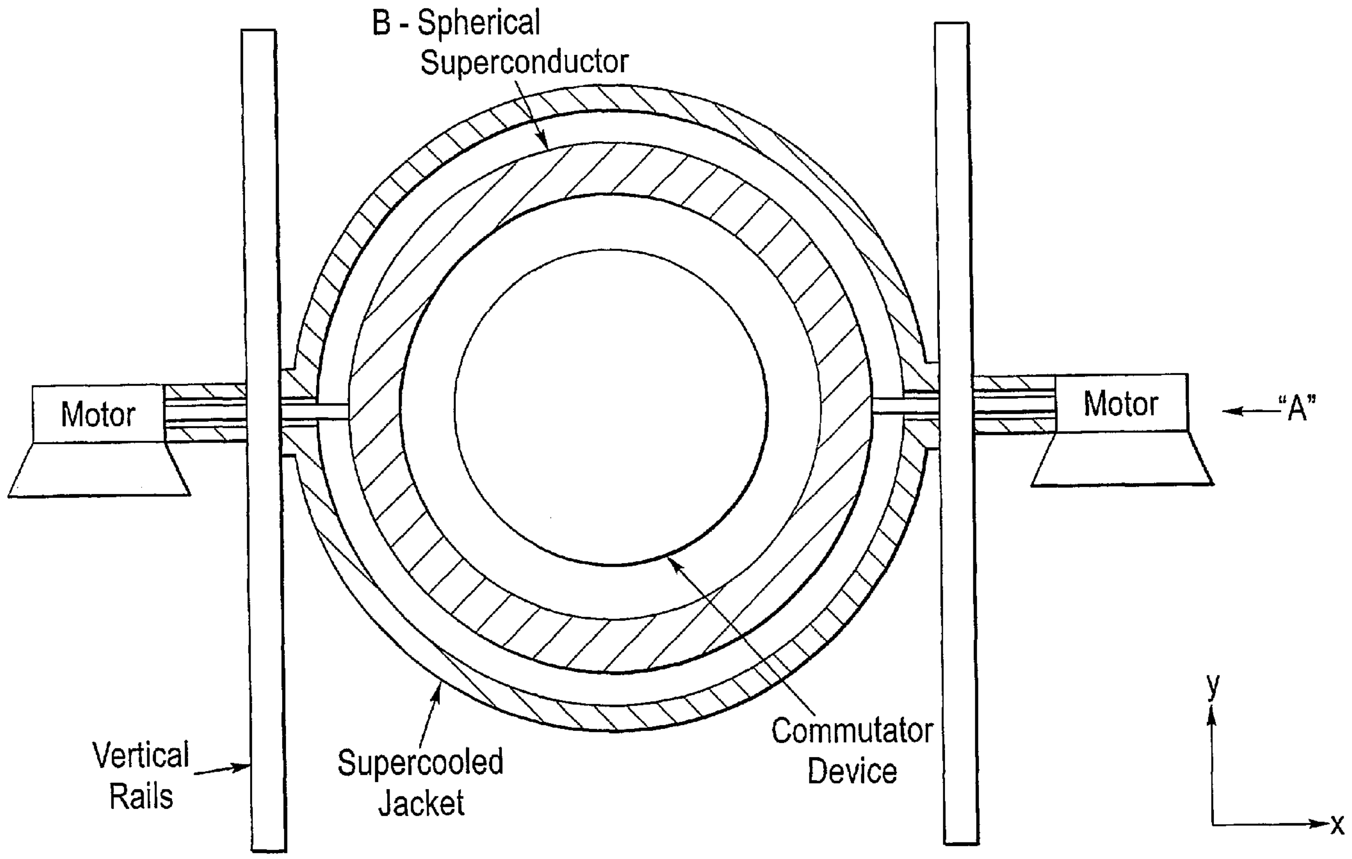
B - Spherical Superconductor
Motor
Motor
"A"
Vertical Rails
Supercooled Jacket
Commutator Device
y
x

Spherical Superconductor
Sphere = 0.25m external radius
Sphere = 0.20m internal radius

Figure 1 & 2
B = Superconductor

Fig. 1

Direction of Motion
y vector (vertical rails)

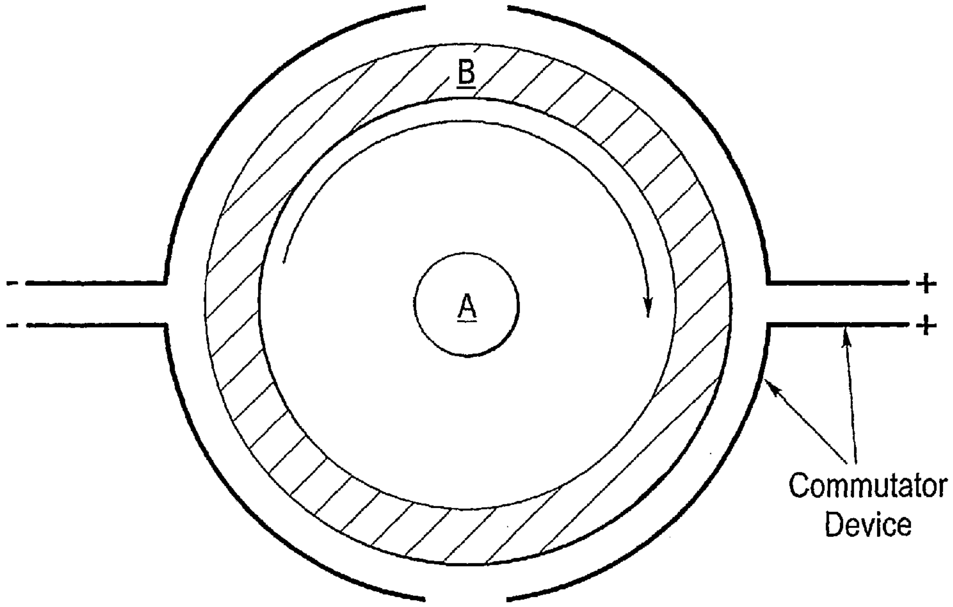
B
A
-
-
+
+
Commutator
Device

A = Ultracentrifugational Axle
(rps 10,000 revs/sec)
B = Spherical Superconductor

Fig.2

TECHNICAL AND THEORETICAL SPECIFICATIONS FOR WARP DRIVE TECHNOLOGY

FIELD OF THE INVENTION

[0001] The present invention relates to the use of technical drive systems which operate by the modification of gravitational fields. These drive systems do not depend on the emission of matter to create thrust, but create a change in the curvature of space-time, in accordance with general relativity. This allows travel by warping space-time, to produce an independent warp drive system.

THEORETICAL BACKGROUND OF THE INVENTION

[0002] The concept of gravity should be examined in the light of quantum gravity and in turn as a component of quantum physics itself. The fundamental minimal quantum of energy in quantum physics is Planck's constant; h. Thus in accordance with the energy equivalence formula $E=mc^2$, the fundamental minimum quantity of mass (m_q) can therefore be derived, from known constants by; $m_q=h/c^2$ (1). Taking this minimal mass, it is possible to show that the formation of all matter, the forces of nature and indeed space-time itself derive from this single quintessential quantity.

[0003] Thus if the number of quintessences in a system is; $n_q=m/m_q$: then the total Energy of the system is more logically given by, the energy of a single quintessence (h); directly multiplied by the number of quintessences (n_q) in that system, thus $E=hn_q=mc^2$ (1 a).

[0004] Furthermore, this minimal mass, termed quintessence, can form the basis of the existence of a quantum gravitational field in the form of a space-time lattice, from which quantum gravity may be derived from first principles. Furthermore, the conglomeration of these quintessences also accounts for the formation of the elementary particles and the forces acting between them, as in superstring theory. This concept explains the formation of matter and the forces of nature on a quantum mechanical basis and directly explains the existence of wave particle duality. Thus as $n_q=m/m_q$; the frequency of light and matter (f) is determined, directly, from the number of constituent quintessences. This leads automatically to the fundamental equation, derived from (1), $f=n_q=E/h$, where n_q is the number of quintessences, which leads directly to the frequency of both light and matter. This in turn leads directly to a Universal wave equation for matter and light $\lambda=c/\beta n_q=hc/\beta E$ (2), where β is the relative directional velocity, v/c. As the momentum, $p=\beta E/c$, then this equation also gives the standard de-Broglie wave equation, $\lambda=h/p$ in agreement with current theory and experiments.[1]

[0005] Using the Universal wave equation, the standard equation for special relativity, $m'=m_o/(1-\beta^2)^{1/2}$, derives from first principles. Also from these observations, a modified Dirac wave equation may be derived, $E\psi=(-j\beta\cdot\nabla+\beta m)\psi$ (2a), the results of which have been recently verified by a paper in which the orbitals of electrons were experimentally directly visualised.[2] Moreover, a fundamental equation for general relativity can be formulated, where G is the gravitational constant and r_q is the given radius of quintessence; $G=9r_q^2c^4/\lambda\beta E$ (3), such that the Universal wave equation is in direct agreement with general relativity.[3] Thus special and general relativity and quantum mechanics can be unified.

[0006] From here it is possible to proceed in a number of ways; the geometric structure of the electron and the forces of Nature may be derived from first principles and in turn the structure of the quarks, including the top and bottom, otherwise known as truth and beauty can be seen. Moreover, the presence of a space-time lattice results in an understanding of quantum EPR effects. By allowing a theoretical flow of energy through the space-time lattice it can be shown that:

[0007] Energy is not bound by space-time

[0008] Thus logically accounting for phenomena such as entanglement and quantum tunnelling. Quintessence can also be used to explain, logically, the inner physics of a black hole, the missing mass of the Galaxy, the continuing expansion of the Universe, Guth's inflationary theory and the Big Bang. Hence, it is now possible to understand the Universe, including space-time, matter and the forces of nature from the radius, mass and vibration of a single quantity, quintessence.

[0009] With this understanding of space-time, matter and the forces of Nature, and in particular gravity, it is possible to demonstrate that the modification of gravitational fields, and in turn the warping of space-time, can be technically readily achieved.

[0010] Using standard equations for special relativity, $m'=m_0/(1-\beta^2)^{1/2}$, it can be demonstrated that by differentially increasing the velocity of electrons, by applying a differential current, their mass can be increased in a specific way. In turn by increasing the mass of electrons, by general relativity, the number of gravitons emitted from these electrons can be modulated. By multiplying this effect using an ultracentrifugational device the differential graviton emission can be manifestly amplified. This in turn, in accordance with general relativity, will cause a change in the curvature of space-time.

[0011] This effective warping of space-time does not, of necessity, imply superluminal velocities, but does allow the creation of warp drive systems, which do not depend on the creation of thrust by the ejection of material as used in current space technologies.

Part I

Fundamental Laws of Physics

[0012] Quintessential Mass

[0013] The quantum physical, minimum component of energy is Planck's constant; h. To define the minimal component of mass, using the standard energy equivalence formula; $E=mc^2$, such a minimal mass (m_q) would be required to have the value equivalent to; $m_q=h/c^2$ (1). The total mass of a system (m) would then be; $m=m_qn_q$, where n_q is the number of these minimal units. Thence, the total energy of a system can be derived from the minimal energy; h, multiplied by the number of these energy units (n_q). Thus as, $E=mc^2$, then also $E=m_qn_qc^2$ and substituting $m_q=h/c^2$, the energy equivalence formula has the more logical formulation; $E=hn_q$ (1a). Thus the energy of a system is equivalent to the minimal energy unit; h, multiplied by the number of those minimal energy units (n_q).

[0014] Taking this minimal mass/energy, it is possible to show that all matter, the forces of nature and space time can be constructed from this single quintessential quantity. Moreover, using this quantity the laws of physics can be derived from first principles. Thus, a priori, all components of the physical universe, including space-time, can be constructed from this minimal mass component, termed quintessence.

[0015] Wave Particle Duality

[0016] If the presence of quintessence accounts for the structure of matter and if matter itself forms from the number of quintessences, then the frequency of matter and thus wave particle duality directly arises from first principles. Specifically the wavelength of matter derives from the vibration of quintessence from which it is constituted. Thus the frequency (f) and in turn the wavelength of light and matter is directly equivalent to the number of quintessences contained within it. We find that the actual frequency of light can be directly derived from first principles from the effective mass of the photon (m,) and thus by the number of quintessences (n_q) it contains.

[0017] Thus for light conventionally:

$$f=E/h$$

[0018] and if $E=mc^2$, and $h=m_q.c^2$, then

$$f=m_\gamma c^2/m_q.c^2$$

[0019] and

$$f=m_\gamma/m_q=n_q$$

[0020] Thus

$$f=n_q \tag{4}$$

[0021] Thus the formula for the frequency of light (E=hf) is now readily explained by the observation that the frequency is determined quite directly from the number of quintessences (n_q) within the photon.

[0022] The wavelength is thus also given by:

$$\lambda=c/f=m_q c/m_\gamma=h/p$$

[0023] We can now show that the frequency of matter also has the same derivation from quintessence, as has the frequency of light. The frequency of matter is again equivalent to the number of quintessences it contains. Thus the wave particle duality of matter itself can be explained by its composition from quintessence. The amount of quintessences contained within a electron sphere will depend on the number of quintessences constituting the electron and those passing through it as a result of its relative velocity β^2 (where $\beta=v/c$); effectively its relativistic momentum (p). The frequency will then be related to the total number of quintessences. Thus for matter,

$$f=\beta^2 n_q \tag{4a}$$

[0024] Thus it is possible to derive the conventional de Broglie wave equation for matter from first principles. Thus, as $\lambda=v/f$, we have:

$$\lambda=v/\beta^2 n_q \tag{5}$$

[0025] thus as $n_q=E/h$

$$\lambda=hc/\beta E \tag{2}$$

[0026] and as conventionally $\beta E/c=p$, then for matter:

$$\lambda=h/p$$

[0027] Provided that in the de Broglie equation, the momentum of the object is calculated using the relativistic mass, thus accounting for the total number of quintessences (n_q) in an object, this gives an accurate value for the wavelength of matter.[1]

[0028] Thus the wavelength of matter follows directly from its constitutents, quintessence. As matter is made of quintessence, similarly to light, its frequency depends on the number of quintessences (n_q) within it, travelling relative to the speed of light. Moreover, $\lambda=hc/\beta E$, underpins a fundamental relationship between wavelength and energy. Furthermore, this is mathematically the same as the term $\lambda=hv/\beta^2E$, giving a relativistic expression for the the wavelength of matter, from which the relativistic equations may be directly derived (see Wave Particle Duality and Relativity).

[0029] Wave Equations

[0030] The derivation of wave particle duality from first principles also now allows the derivation of a modified wave equation for matter.

[0031] To derive his wave equation Shrödinger commenced with the de Broglie equation using momentum (p). For lower energies the momentum of an electron is conventionally derived from the kinetic energy of the electron and the mass of the electron mo. Thus conventionally:

$$E_K=\tfrac{1}{2}mv^{1/2}\text{and } p=m_0v$$

[0032] Thus

$$E_K=p^2/2m_0$$

[0033] then

$$p=(E_K.2m_0)^{1/2}$$

[0034] and conventionally, the de Broglie equation can also be written as:

$$\lambda=h/p=h/(E_K.2m_0)^{1/2}$$

[0035] In turn the Shrödinger wave equation directly derives from the square of the above classical non relativistic term for kinetic energy:

$$\lambda^2=h^2/E_K.2m_0$$

[0036] thus

$$E_k=\frac{h^2}{2m}\cdot\frac{1}{\lambda^2}$$

[0037] As $E=E_k+V$

[0038] then

$$E\psi=-\frac{h^2}{2m}\cdot\frac{d^2\psi}{dx^2}+V\psi=jh\cdot\frac{d\psi}{dt}$$

[0039] However, the Shrödinger equation, may be refined by taking into account relativity. Thus the true values for the energy are given by the relativistic momentum (p).

[0040] A fundamental relativistic wave equation for ψ, and its logical derivation may now be developed through the concept of quintessence as a fundamental constituent of matter.

[0041] The amount of quintessences in the electron is determined by the number of quintessences forming the electron at rest, plus the amount of quintessences passing through it due to its relativistic velocity, which will determine the relativistic momentum (p) of a particle.

[0042] The frequency of matter can now be readily calculated from first principles to give a more accurate result. Thus as matter is made of quintessence, similarly to light, its frequency is equal to the number of quintessences (n_q) within it. The wavelength will depend on its velocity travelling relative to the speed of light and thus multiplied by the relative velocity compared to c ($\beta=v/c$);

[0043] Hence for matter as previously shown:

$$\lambda=v/\beta^2 n_q=hc/\beta E \tag{2}$$

[0044] And conventionally

$$E=(p^2c^2+m_0{}^2c^4)^{1/2}$$

[0045] Using these equations, we can now, also, reformulate the Shrödinger wave equation, which has the advantage that relativity can be treated in a quantum mechanical way. Thus if the wave energy of matter is defined as:

$$E_\lambda=\beta(p^2c^2+m_0{}^2c^4)^{1/2}/c^2$$

[0046] thus

$$E_\lambda=(\beta^2p^2/c^2+\beta^2m_0{}^2)^{1/2}$$

[0047] which in complex space generalises to

$$E_\psi=(-j\beta\cdot\nabla+\beta m)_\psi \tag{2a}$$

[0048] As the term $\alpha=e^2/hc.4\pi\epsilon_0$; also represents the ground state ratio of the velocity of the electron to c. Thus $\alpha=\beta=v/c=1/137$.

[0049] Thus, also

$$E_\psi=(-j\alpha\cdot\nabla+\beta m)_\psi$$

[0050] This is thus the standard relativistic equation that Dirac was able to construct from the Shrödinger wave equation. This relativistic equation can be derived from the modified wave equation (2). This takes into account the relative mass energy which the quintessential wave equation (2) contains.

[0051] Where importantly the term βm is the mass m, multiplied by the ratio of the relative velocity to light $\beta=v/c$, and the term a is also essentially the relative velocity of the electron.

[0052] The Dirac equation was an empirical formula which worked mathematically, nevertheless even Dirac admitted it was not logically understood. The importance of these equations (eq. 2, 2a) is that they show that the existence of quintessence allows the waveparticle duality of matter to be explained and mathematically derived from first principles, Thus the frequency of matter or even light is simply determined by the number of quintessences it contains.

[0053] Indeed, a recent publication in Nature has suggested that the direct visualisation of the orbitals of electrons shows that these are in very close agreement with theory. However, there is a significant departure from theory, in the interstitial molecular regions, suggesting that the higher velocities of the electrons obey the modified Dirac equation (2a).Thus these orbitals were in keeping with the modified Dirac equation, which itself may be derived from the wave equation above, $\lambda=hc/\beta E$ (eq. 2) (ref 2)

[0054] The Shrödinger wave equation will approximate to the correct values until v approaches c. Indeed the Shrödinger equation will give similar answers as that derived from equation (2a), under most experimental conditions.

[0055] However, equation 2 and its derivative (2a) may have advantages over standard Shrödinger theory with relativistic speeds. Furthermore, equation 2, conceptually shows that the wave particle duality of matter derives from the principle that the frequency of matter is directly equal to the number of quintessences it contains. Importantly it also mathematically allows relativity and quantum mechanics to be united.

[0056] With $v \cong c$, the modified Dirac equation (2a) will yield more accurate results, particularly compared with the Schrödinger equation. We also find that the equation $\lambda=hc/\beta E$ (2) is equivalent to the de Broglie wave equation, $\lambda=h/p$, provided we use the relativistic mass in the de Broglie equation. Given this, these equations yield accurate experimental results (ref 1).

[0057] Thus we find that the modified formulation of de Broglie wave equation $\lambda=hc/\beta E$ (eq. 2a) leads directly to a modified Dirac relativistic wave equation and is supported by recent experiments which measure the wavelength of matter and demonstrate the electron orbitals experimentally from these wave equations for matter (ref 1,2).

[0058] Wave Particle Duality and Relativity

[0059] From here it is possible to proceed in several ways using the relativistic wave equation (2). It is apparent that the reintroduction of the term for relative velocity into the wave equations will enable the reintroduction of special relativity into quantum mechanics. In particular we should now be able to derive the term $(1-v^2/c^2)^{1/2}$ as a special case of quantum mechanics.

[0060] Thus if:

$$\lambda=hc/\beta E \tag{2}$$

[0061] As $E=(p^2c^2+m_0{}^2c^4)^{1/2}$, squaring

$$\lambda^2=h^2c^2/\beta^2(p^2c^2+m_0{}^2c4)$$

[0062] Conventionally $p^2c^2=E^2v^2/c^2$

[0063] then

$$\lambda^2=h^2c^2/\beta^2(E^2v^2/c^2+m_0{}^2c^4)$$

[0064] Thus as $\beta^2=v^2/c^2$ and $m_0{}^2c^4{}^{=E}{}_0{}^2$, then:

$$\beta^4E^2+\beta^2E_0{}^2=h^2c^2/\lambda^2$$

[0065] hence

$$\beta^4E^2=h^2c^2\cdot\frac{1}{\lambda^2}-\beta^2m_0^2c^4$$

thus

$$\beta^2=\frac{h^2c^2}{\beta^2E^2}\cdot\frac{1}{\lambda^2}-\frac{\beta^2m_0^2c^4}{\beta^2E^2}$$

-continued

As $E^2 =' m^2c^4$

$$\beta^2 = \frac{h^2c^2}{\beta^{2\prime}m^2c^4}\cdot\frac{1}{\lambda^2} - \frac{\beta^2m_0^2c^4}{\beta^2E^2}$$

Substituting $h = m_qc^2$

$$\beta^2 = \frac{m_q^2c^6}{\beta^{2\prime}m^2c^4}\cdot\frac{1}{\lambda^2} - \frac{\beta^2m_0^2c^4}{\beta^2E^2}$$

As $\mathrm{m}_q/'m = 1/n_q$(eq. 2)

$$\beta^2 = \frac{c^2}{\beta^2n_q^2}\cdot\frac{1}{\lambda^2} - \frac{\beta^2m_0^2c^4}{\beta^2E^2}$$

[0066] Thus if $f=\beta^2 n_q$ (eq. 7a);

$$\beta^2 = \frac{v^2}{f^2}\cdot\frac{1}{\lambda^2} - \frac{\beta^2m_0^2c^4f^2}{\beta^2E^2}$$

As $1/\lambda^2 = f^2/v^2$

$$\beta^2 = \frac{v^2}{f^2}\cdot\frac{f^2}{v^2} - \frac{\beta^2m_0^2c^4f^2}{\beta^2E^2}$$

Thus:

$$\beta^2 = 1 - \frac{\beta^2m_0^2c^4}{\beta^2E^2}$$

As $E^2 =' m^2c^4$

$$\beta^2 = 1 - \frac{m_0^2}{'m^2}$$

Hence

$$m_0/'m = (1-\beta^2)^{1/2}$$

Thus

$$'m = \frac{m_0}{\left[1 - \frac{v^2}{c^2}\right]^{1/2}}$$

[0067] Thus this derivation now allows relativity as a universal case of the quintessential wave nature of matter.

[0068] The original premises on which special relativity was based were: that the speed of light is a constant and that all observers are equal. As the speed of light has dimensions of length and time but not apparently of mass, the relativistic change in mass is not accounted for. Using quintessence logically and directly accounts for the relativistic mass changes.

[0069] Moreover, relativity can be derived from the de Broglie equation, and visa versa, directly, thus linking relativity and quantum mechanics by taking into account the existence of quintessence mass.

[0070] Hence, it is now possible to derive the relativistic equations for mass and in turn for space and time from the quintessential wave equation, thus deriving special relativity as a universal case of quantum mechanics and thus uniting special relativity and quantum mechanics. This now allows a further understanding of the nature of space-time.

[0071] The Space-Time Lattice

[0072] The understanding of the true nature of space-time and how it is formulated in three dimensions of real space is crucial. To simply assume that space-time exists, and thence not to question the nature of that existence, denies a deeper understanding of the universe.

[0073] In order to understand the nature of space-time itself, at the quantum level a further look at the nature light and the photon is necessary. Since Einstein's description of light as a particle (the photon) and the description of the photoelectric effect, the standard picture of light as simply a wave can, no longer be applied. If light was to exist as a photon, it could not exist in one dimension, as ordinary waves do, it would need to be three dimensional, with the addition of time. Let us suppose, in this case, that a photon is a three dimensional helical ringlet of light, travelling in the x vector, and spinning around the x-axis. Conventionally this ringlet has a radius; $r=\lambda/2\pi$. The ringlet itself would be vibrating in the y and z vectors. The vectors x, y and z would represent the photon, the substance of which, would be travelling in the x direction and oscillating in the y and z vectors, which would represent oscillatory energy. This in turn would allow it to act as a wave, and create oscillatory electromagnetic fields.

[0074] It is important to re-examine space-time itself in this light, this would have one directional vector with two vector dimensions of energy, one of capacitance and one of electrical permeability, thus accounting for the well known constants of free space; the permittivity of free space (ϵ_0) and the permeability of free space (μ_0) respectively. The vector dimension of direction x, would be the direction of travel and those "quintessences" travelling in an outwardly direction would account for none other than the expansion of the universe. Three of these quintessences would naturally constitute three dimensional visible space-time. These constituents of space-time would interact with the generations of the other vector dimensions reciprocally. Thus one quintessence would sweep out one vector of permeability and one vector of permittivity, through which the other two quintessences could travel, and vica versa, creating a three dimensional space-time lattice.

[0075] The permittivity of free space, (ϵ_0) which is equivalent to capacitance, would as with capacitance plates, be determined by the effective separation between quintessences. The permeability of free space (μ_0) is in fact a force, measured as $4\pi\times10^7$ N A^{-2}, would result from the force produced by the vibration of quintessence and would be dependent on the density of quintessence. Hence these two parameters would be reciprocal and thus the product of these two would therefore be a constant, which is recognised as none other than the speed of light.

$$c=(\mu_0.\epsilon_0)^{-1/2}$$

[0076] This space time lattice would in effect be created by quintessences travelling in all directions with a speed of c within the lattice. The quintessences of the space time latice would in effect produce a non-static ether. A non-static ether is fully compatible with special and general relativity. Indeed such an ether explains how space time can be curved as in general relativity. Furthermore, the existence of a non-static ether, was espoused by Einstein in his University of Leyden lecture on general relativity of May 5, 1920. In Einstein's own words;

[0077] "According to the general theory of relativity space without ether is unthinkable."(ref 4)

[0078] Recent evidence from a number of sources now strongly support the presence of this non-static ether, in the form of quintessence. An editorial from a major journal states "combined with other observations such as those of distant Supernova, the QMAP results corroborate the prevailing theory of inflation with the twist that the Universe is only one third matter (both ordinary and dark) and two thirds quintessence, a form of energy possibly inherent in empty space".(refs 5, 6, 7)

[0079] If we take into account the existence of quintessence and as such a three dimensional space-time lattice, matter which is intrinsically made of contituents of charge would interact with this lattice to produce the effects of mass. Mass would be perceived as a result of matter (whose constituent particles appear to contain charge) interacting with this lattice directly due to the inhibition of motion by the lattice's electrical permeability and permittivity vectors, which would form the existence of complex space. These quintessences would in the direction in the y and z vectors produce small vibrations of the order of the Planck length (10^{-35} m), whilst passing through the vectors of permeability and permittivity, thus producing the effects of mass.

[0080] The vibration would endow quintessence itself a (non rest) mass (m_q) equivalent, to the minimal mass of:

$$m_q = h/c^2 = 7.373\times10^{-51}\ \text{kg.sec} \tag{1}$$

[0081] The presence and magnitude of Planck's constant (h) and especially the speed of light (c) is thus explained. Indeed, the speed of light $c=(\mu_{0.}\epsilon_0)^{-1/2}$ is not in itself a fundamental quantity.

[0082] As the energy equivalence formula is $E=mc^2$, the minimal mass of a single quintessence, would thus be the minimal mass, h/c^2, hence again:

$$m_q = h[\mu_0.\epsilon_0] = h/c^2 = 7.373\times10^{-51}\ \text{kg.sec} \tag{1b}$$

[0083] or

$$m_q c^2 = h \tag{1c}$$

[0084] It is postulated by general relativity that the shape of space time itself can be altered, indeed the presence of the space time lattice now allows this to be altered by altering the density of quintessence. It is further clear that if quintessences underly the structure of the space-time lattice, they may also underly the structure of matter itself.

[0085] With regards a single quintessence, this passing through an energy vector of the space-time lattice would appear as a vibrating string. In a similar way to string theory, the conglomeration of these quintessences would produce the constituents of ordinary matter. Thus the general equation for the number of quintessences (n_q) in an object of mass (m) would be

$$m/m_q = n_q$$

[0086] The mass of the electron (m_e) for example, would be directly determined by the number of quintessences in the electron, multiplied by the mass of quintessence.

[0087] Quintessence and Complex Space

[0088] Quintessence is postulated to constitute the fundamental nature of space-time. Three quintessences each travelling in their respective x vectors at 90° to each other would create three dimensional real space-time. These quintessences would in the direction in their respective y and z vectors produce small vibrations of the order of the Planck length (10^{-35} m), this would create the vector dimensions of permeability and permittivity. The result would give space-time 9 dimensions of space as in superstring theory. However, unlike superstring theory the six hidden dimensions would not be "curled up so as to be so small as to be invisible" these six dimensions would be present in complex space. Thus, only three of these dimensions would represent ordinary three dimensional particulate space time i.e. three dimensional objects. The other six dimensions produced by the vibrations of quintessence would form complex space.

[0089] The mathematics of complex space, using imaginary √−1 or (j) numbers, is assumed in the standard formulation of the Shrödinger wave equation. Thus the presence of complex space is an integral part of quantum mechanics.(ref 8)

$$-\frac{h^2}{2m}\cdot\frac{d^2\psi}{dx^2} + V\psi =_j h\cdot\frac{d\psi}{dt}$$

[0090] The mathematics of complex space is also an essential and integral part of the principles and application of modern electronic and control engineering. Indeed it has been well recognised for some time that each direction vector in electronic engineering can , be associated with complex vectors.(ref 9)

[0091] As this complex space consists of the vectors of permittivity and permeability it would only be "felt" by charged particles as in the electron. Nevertheless, as all particles are fundamentally composed of charged particles the effects of complex space would be felt by endowing these particles with mass and in turn kinetic energy.

[0092] In conventional complex space, a 2 dimensional Cartesian Argand diagram is mathematically used. However, in order to formulate the equations for particles a three dimensional Argand diagram is essential. This will have three dimensional vectors, one real vector and two imaginary vectors. Three of these diagrams will be required to fully describe the nature of particles, each with a real vector in the x, y and z vectors, respectively. Nevertheless, in the instance below the real vector is the x vector and the two imaginary vectors are given by ($_j$y, $_j$z)

The Three Dimensional Argand Diagram

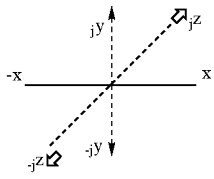

[0093] The Three Dimensional Argand Diagram

[0094] The beauty of a three dimensional Argand diagram is that the complex conjugate (i.e. the mirror image which confers mathematical reality on the coordinates) is formed by the value of the minus coordinate in the other complex vector dimension. Thus the complex conjugate of $(c_x^{1/2}+{}_jc_y^{1/2}+{}_{-j}c_z^{1/2})$ is $(c_x^{1/2}+{}_jc_y^{1/2}+{}_jc_z^{1/2})$. These two sums when multiplied thus give a real number solution.

[0095] Furthermore it is clear that nine dimensions of space time are necessary in the general relativistic equations. By including complex space we thereby create the nine dimensional spacial metrictensor and the metric energy tensor of matter necessary for computations for general relativity (see quantised general relativity pp 34-35). From here we can begin to understand the true structure of matter.

[0096] Energy and the Space-Time Lattice

[0097] The presence of numerous experimental data for quantum tunnelling(refs 10, 11, 12, 13) and indeed the recent observations by Nicholas Gisin, on the entanglement of distant photons now returns us to EPR experiments.

[0098] Using the quintessential modification of the de Broglie wave equation, gives us an insight into these teleportation and EPR effects.

[0099] As

$$\lambda=hc/\beta E \tag{2}$$

[0100] and

$$E=hn_q \tag{1a}$$

[0101] then

$$\lambda=c/\beta n_q \tag{2b}$$

[0102] Importantly, as indicated by equation (2b), energy having no quintessence; would have a wavelength of infinity. Specifically pure energy containing no quintessences, would have a lambda of infinity. According to quantum mechanics an infinite wavelength would result in the probability of that energy being anywhere. As energy itself has no electrical charge it would not be impeded by the permittivity and permeability of the three dimensional space-time lattice. Moreover, energy would not be detectable in three dimensional space-time, unless it interacted with matter, as in the EPR experiments. Indeed, energy is not observed when not bound to any form of mass or particle. Thus equation 9d, takes us to our original assertion

[0103] Energy is Not Bound by the Space-Time Lattice

[0104] Thus, as the EPR experiments suggest the existence of energy separate from matter and thus separate from the three dimensional space-time lattice, it is interesting to find that experiment suggests the existence of free energy in a continuum separate from space time to produce the effects of quantum teleportation.(refs 10, 11, 12, 13)

[0105] This is not, however, teleportation across an additional dimension, this is a term to describe in partially familiar terms the dissociation of energy from the three dimensional space-time lattice. As time is inextricably linked to each dimension of space, the effects of energy would be inextricably linked to the events, such as the creation of virtual particles, we see interacting within space-time.

[0106] It is unlikely that observers have any direct day to day experience to explain quantum events. Nevertheless, quantum physics may have given us a window into the hitherto hidden workings of the Universe. Thereby, the mystery of the uniformity of the Universe, across distances which the speed of light could not apparently traverse, is readily explained by the fact that the free energy contained in the Universe is not bound by the space-time lattice.

[0107] In the case of light, due to the exceedingly small masses involved, there would be relatively easy exchange of matter with free energy within a photon. This would make the photon the ideal experimental tool to look for energy which is not bound by matter and in turn energy which is not bound in space-time. Indeed, very recently Furusawa et at. have reported to have observed the transference of energy as photons from A to B, without those photons traversing space-time(ref 10). This finding which has been supported using other experimental techniques(refs 11, 12, 13,), is very important as it suggests the existence of such a quantum continuum.

[0108] We have already seen strong experimental data using photons(ref 10), atomic spins and other data for quantum teleportation which have recently been published (refs,11, 12, 13), which support these findings. According to the above equations the teleportation would vary in a predictable fashion, as with photons, in line with the wavelength of the light used, relative to the size of vibration of quintessence. As regards matter, the results do confirm that the effect of quantum tunnelling is indeed dependant on the wavelength of matter and the size of that matter(ref 10).

Part II

Particle Physics

[0109] Electron Structure

[0110] Understanding the electron is fundamental to the understanding of the elementary particles. The hidden nature of the electron may recently have been revealed through observations by Horst Stormer, Daniel Tsui and Robert Laughlin for which a Nobel prize has recently been awarded. They describe a quasi electron particle of charge ⅓e. This has been described on a quantum basis as a vortex of energy, bound as a quasi particle in one dimesion x, but not bound in the other two dimensions y and z, allowing dispersion in space-time as a vortex. What is more intriguing are the experimental conditions in which this occurs. First of all a two dimensional electron gas is created and held between two capacitance plates. A magnetic force is then applied in the remaining dimension, virtually creating a one dimensional passage through which only a quasi electron appears to be able to pass.(refs 15, 16)

[0111] Given the presence of charge of ⅓e, then three of these quasi electrons could form an entire electron in three dimensional visible space time. Nevertheless, each would have energy and hence a wave function which would be present in the other vectors. This electron could thus follow the probability functions as described by the Shrödinger wave equation for ψ (otherwise termed as "essence" by Shrödinger)

[0112] If the mass of the electron (m_e) is constituted from quintessence, using the formula:

$$m_e/m_q=n_q$$

[0113] Then an electron would be constituted from:

$$9.11\times10^{-31}\ \text{kg}\div7.373\times10^{-51}\ \text{kg.sec}=1.236\times10^{20}\ \text{quintessences/sec.}$$

[0114] Thus taking into account the mass-energy content of quintessence (m_q) it is independantly possible to derive the magnitude of the charge of an electron (e) using the following equation.

$$e=[m_q\epsilon_0/\tfrac{4}{3}\pi hc]^{1/2}=1.61\times10^{-19}C$$

[0115] This is in close agreement with the experimentally observed charge on the electron of 1.602×10^{-19} C.

[0116] Interestingly substituting $m_q=h/3c^2$ in the above equation we have:

$$e=[\epsilon_0/3(\tfrac{4}{3}\pi c^3)]^{1/2} \tag{6}$$

[0117] This can also be written as

$$e=\left[\frac{\varepsilon o}{3(4/3\pi c^3)}\right]^{1/2} \tag{6a}$$

[0118] Equation (6) has a number of very special implications, if re-examined, firstly three of these quasi electron spheres appear to to be required to constitute the charge of the electron . More intriguingly, it indicates that the charge is related to the volume of a sphere with an apparent radius of c. Thirdly it indicates that the square of the charge of an electron (e) is proportional to the permittivity of free space (so). The charge given from equation (3) is in close agreement with the measured charge of the electron. Furthermore a more exact value for the charge of the electron (to seven decimal places) can be deduced by taking into account the gravitational field of the Earth (see Gravity and the Charge of the Electron). Furthermore the charge of the electron (e) can now be derived from first principles. Thus, equation (3) corroborates the evidence that the electron is indeed composed of three quasi electrons in keeping with recent experimental findings.(ref 16)

[0119] The significance of the electron, composed of three spheres each with a radius of c, is not immediately clear, but can be understood if the frequency of rotation of the electron is taken into account. Thus if the diameter of the electron was approximately 10^{-19} m, then its spin would need to be $1/c.10^{-15}$ m$\approx10^{6}$ cycles/sec. Thus given a very high rotation rate an electron could have an effective radius of 1/c and still occupy subatomic sizes. Indeed these observations might be used to estimate the rate of rotation of the quasi electron and its size (see Appendix 1).

[0120] With regards a single quintessence, this passing through an energy vector of the space-time lattice would appear as a vibrating string. In a similar way to string theory, the conglomeration of these quintessences would produce the constituents of ordinary matter. The electron, for example, would be constituted from approximately 1.236×10^{20} quintessences.

[0121] The dimensions of the equation for the electron can be readily resolved by considering each of the three vector dimensions. The exact dimensions of the equation need to be considered in the light of the nature of space-time itself. These dimensional equations help explain the nature of matter. Indeed the equation for the electron may be necessary for the full understanding of gravity, this is also fully addressed (Appendix 1, Dimensional Equations)

[0122] Complex Space and Electron Structure

[0123] The presence of complex space also now further explains the conformation of the the electron, and its formulation at the quantum level, and the presence of particles, anti-particles and their spin up and spin down characteristics.

[0124] Indeed the short form equation for the charge of the electron (−e) can now be rewritten as a metric tensor with three dimensions in real space and six in complex space.

[0125] Thus if three of the x, y and z vectors are in real space and six vectors in complex space, where c is the speed of light in the real space vector, ${}_jc$ is the speed of light in the complex vector and ${}_{-j}c$ is the complex conjugate of ${}_jc$, thus the electron can be mathematically represented by the equation:

$$-e=\varepsilon_{qe}/(4/3\pi)^{1/2}\cdot\begin{matrix}(c_x)^{1/2} & \cdot(jc_y)^{1/2} & \cdot(-jc_z)^{1/2}\\ + & + & +\\ (-jc_x)^{1/2} & \cdot(c_y)^{1/2} & \cdot(jc_z)^{1/2}\\ + & + & +\\ (jc_x)^{1/2} & \cdot(-jc_y)^{1/2} & \cdot(c_z)^{1/2}\end{matrix} \tag{7}$$

[0126] Which now elegantly gives the real number solution

$$e=\epsilon_{qe}/3(\tfrac{4}{3}\pi c^3)^{1/2}$$

[0127] Where ϵ_{qe} is given as the permittivity of free space for a single quasi electron (see appendix 1). Equation 4 represents a "complex" tensor

[0128] Whilst the two dimensional Argand diagram has four quadrants, the three dimensional Argand diagram has eight cubic sectors. Two of these cubic sectors are diametric opposites and can represent "real" particulate objects. These have the primary coordinates x, y, −z; as in the electron described above, and the −x, −y, z, with the real vector x now having a minus sign. These two "real" cubic sectors, therefore, mathematically represent particles and their anti-particles.

[0129] The mathematical presence of the two primary diagonal mirror images (x, y, −z and −x, −y, z) now allow the introduction of the concept of antiparticles. This extension of the maths into a three dimensional Argand diagram thus results in the automatic formulation of the maths of anti-particles. Thus the charge of the positron (${}^+$e) is formulated by the shortened form equation, where the real vectors now each have the minus sign, and therefore exist in the −x, −y, z sector of the three dimensional Argand diagram.

$$e=\varepsilon_{qe}/(4/3\pi)^{1/2}\cdot\begin{matrix}-c_x^{1/2} & \cdot_{-j}c_y^{1/2} & \cdot_j c_z^{1/2}\\ {}_jc_x^{1/2} & \cdot -c^{1/2^y} & \cdot_{-j}c_z^{1/2}\\ {}_{-j}c_x^{1/2} & \cdot_j c_y^{1/2} & \cdot -c_z^{1/2}\end{matrix} \tag{8}$$

[0130] The three dimensional Argand diagram also accounts for chirality and indeed the up and down spin of the electron. There are two other "real" primary coordinates in the Argand diagram, these are themselves the partial mirror images of the above coordinates (i.e. x, −y, z and −x, y, −z).

In particular the y axis is of the opposite sign, thus in particles the y axis is in the downward direction, to form down spin particles and in anti-particles in the up direction, to form the antiparticle The up spin electron is given by eq. 8 and hence the down spin electron $(-e|\downarrow \leqq)$ is given by the equation

$$-el\downarrow\rangle = \varepsilon_{qe}/(4/3\pi)^{1/2} \cdot \begin{matrix} c_x^{1/2} \cdot_{-j} c_y^{1/2} \cdot_j c_z^{1/2} \\ {}_j c_x^{1/2} \cdot c_y^{1/2} \cdot_{-j} c_z^{1/2} \\ {}_{-j} c_x^{1/2} \cdot_j c_y^{1/2} \cdot c_z^{1/2} \end{matrix} \tag{9}$$

[0131] Thus the three dimensioanl Argand diagram accounts directly for the presence of antiparticles and the spin up and spin down particles seen in nature. It also accounts for the necessity of the electron to form a square root spherical object, as complex space depends on $\sqrt{-1}$, otherwise known as j.

[0132] Electron Pairing and Superconductivity

[0133] As the quintessences making up the electron are in a square root conformation, each of these quasi electrons would have a tendancy to pair to form an entire sphere.

[0134] The square root sphere structure of electrons with up and down spins can now superimpose to produce a complete sphere of varying extents. This produces electron pairing as seen at the atomic and molecular levels. It also accounts for the Pauli exclusion principle. This pairing thus accounts for the reactivity of the valence electrons and the electron probability densities, which in turn accounts for the existence of chemistry.(ref 8)

[0135] Furthermore, it is possible to account directly for superconductivity from first principles. For if both the complex and real vectors of the electron combine completely, the product of an up and down spin electron form a perfect superimposed sphere with radius c, with a charge of 2.59×10^{-38} Coulombs, denoted by the formula:

$$e^2 = \frac{\varepsilon o}{3(4/3\pi c^3)} = 2.59\times10^{-38} C \tag{6b}$$

[0136] As with standard superconducting theory, superconductivity can be explained by the formation of "Cooper" electron pairs, where the electrons are forced to pair by the presence of positve crystal charge in particular formation, at supercooled temperatures. In addition the electron pair now forms a stable entity whose angular momentum cancels.(ref 8)

[0137] It additionally becomes clear that the charge of two seperate electrons (2e) is 3.2×10^{-19} C, but the charge of the combined electrons (e^2) is 2.59×10^{-38} C. This electron pair thus appears to have 19 orders of magnitude less charge than the electron and in turn 19 orders of magnitude less resistance. It is this effective reduction in charge and in turn resistance which may account for superconductivity. When observed directly any electrical interaction with the Cooper electron pair will, however, result in the release of the full charge of both electrons, so that the full electrical charge put in will be equal to that coming out of the apparatus.

[0138] The Fine Structure Constant

[0139] Intruiginty from our knowledge of the electron we can further define the term α, the fine structure constant; from the structure of the electron. Thus as the standard term $\alpha=e^2/hc.47\pi\epsilon_0$; substituting the term $e^2=\epsilon_0/3(4/3\pi c^3)$ (eq. 6) and $h=m_q c^2$ (eq. 1) we find:

$$2\pi/\alpha=m_q[3(4/3\pi c^3)]^2$$

[0140] *or

$$2\pi/\alpha=m_q e^4/\epsilon_0{}^2$$

[0141] For brevity we may represent the quasi electron structure as $(4/3\pi c^3)=\Theta$; to signify its threefold symmetry, thus

$$2\pi/\alpha=m_q[3\Theta]^2 \tag{10}$$

[0142] Indicating that the fine structure constant of the electron (α) is indeed related to its dimensional structure. Again taking into account the effects of gravity the fine structure constant can be derived from first principles to nine decimal places (see Gravity and the Charge of the Electron).

[0143] Fundamental Forces and Particle Structure

[0144] In order to understand the fundamental forces and the nature of fundamental particles, an overview is required. Thus, there are three major forces; strong, electro-weak and gravity, each mediated by three force particles the gluon, photon and graviton respectively. These in turn, influence three types of particle, the quark, lepton, and by general relativity space-time itself. Each of these are composed of particles with multiples of charge of ⅓, which are themselves in three generations, and are present in three dimensions of real space. It is important that a comprehensive view of nature explains this threefold symmetry.

[0145] Using the Standard Model of particles, it is well accepted there exist quark particle charges of −⅓, −⅔ and +⅓ and +⅔ in quarks and anti-quarks. Given that each particle is made up of three quarks the presence of these fractional charges support the association of the fractional charges in this way to form three dimensional charged particles. In stable particles each of the three quarks would have a vector in one dimension, giving the three quarks together an existence in three dimensional visible space time. The particles that bind the quarks (gluons) are themselves required, in stable particles, to have three different colour charges, one colour in each dimension, for the particle to exist in three dimensional space-time. Furthermore, there are three generations of quarks (and indeed leptons).

[0146] The Standard Model (or a modification of this) and in particular the observation of quarks and indeed quasi electrons with fractional charge of ⅓ and ⅔ in both cases, indicates that particles are constituted from the equivalent of three of these quasi particles to form an electron and quarks to form baryons. In the normal three dimensions the energy would be carried by the particle, However, because each particle is constituted of three quasi particles and in each quasi particle or quark one visible dimension would be the direction vector, in the other two hidden dimensions of each vector the waves would carry energy. Thus each particle would be associated with vibration, which would account for wave particle duality and Heisenbergs uncertainty principle in three dimensional visible space-time.

[0147] These observations lead us directly to the previous postulate that the structure of the electron is composed of none other than three (root) spheres, and that this equation for the electron allows the determination of the charge of the electron from first principles, thus:

$$e=\frac{\varepsilon 0^{1/2}}{[3(4/3\pi c^{3})]} \tag{6}$$

[0148] In addition the mass of the proton (m_p) can be directly calculated from the ratio of the mass (m_e) of the electron, given by the equation:

$$m_e/m_p=5.45\times10^{-4}=3(\pi/c^{1/2}) \tag{11}$$

[0149] Strictly we should write, $m_e/(m_p+m_e)=3\pi c^{1/2}$; which is much more elegant.

[0150] Which now gives

$$m_e/m_p=1/(c^{1/2}/3\pi-1)=5.4462\times10^{-3}$$

[0151] This is in very close agreement with the experimentally derived ratio of the proton to electron masses which is also; 5.4462×10^{-3}

[0152] Thus the correlation factor between theory and experiment has a maximum error <0.00001 (ref 17)

[0153] If we combine equation 3: $e=[\epsilon_0/3(4/3\pi c^3)]^{1/2}$ and equation 13: $m_e/m_p=3(\pi/c^{1/2})$ the positive charge of the proton (e_p) is given by:

$$e_p=[\epsilon_0/3(4/3\pi c^3)]^{1/2}\times m_e.3(\pi/c^{1/2})/m_p=e \tag{12}$$

[0154] The stable nuclear proton conformation can thus be represented by the short form equation:

$$p=3^{+}(4/3\pi c^3)^{1/2}.3(\pi/c^{1/2}) \tag{13}$$

[0155] This forms a stable 3×3 conformation as with the stable electron structure.

[0156] Importantly the term $(\pi/c^{1/2})$ is the 90° solution to the Shrödinger wave equation for an electron confined in a space with radius c!(ref 8)

[0157] Thus the standard equation for an electron confined in a one dimensional box is given by:

$$E\psi(x)=-\frac{h^2}{2m}\frac{d^2\psi(x)}{dx^2}$$

[0158] If the one dimensional box has a length 2L the quantum amplitude (A) can only be non zero between x=0 and X=2L and the standard solution for the amplitude is none other than:

$$A=(1/L^{1/2})$$

[0159] Thus in one dimension the standard solution to the Shrödinger wave equation is:

$$\psi(x)=(\pi/L^{/12})\sin x/L$$

[0160] Thus not only is the electron charge derived from the equation for three spheres each with a radius of c (eq. 3); but the proton mass and charge can also be derived from the standard solution to the Shrödinger wave equation for a an electron confined in a space of radius c!(ref 8)

[0161] The term $(\pi/c^{1/2})$ itself would thus most logically represent the gluon which is present in the proton. These gluons would bind the quasi electrons together to form the fundamental particles

[0162] The masses of all the known particles, including the up and down quarks, the W boson, the muon, charm, strange, the tauon, truth and beauty can thus also be derived from first principles in this fashion, and have the quasi electron as their basic constituent particle (see Appendix 1).

[0163] Thus the structure of the muon (μ) can also be derived from the ratio of the mass of the electron (m_e) and the mass of the muon (m_μ):

$$m_e/m_\mu=4.7\times10^{-3}=/c^{1/3}$$

[0164] Thus

$$[\mu=\epsilon_0{}^{1/2}.m_e/m_\mu.3^{-}(4/3\pi c^3)^{1/2}.(\pi/c^{1/3})$$

[0165] Where the charge of the muon is in this equation equivalent to that of the electron e. In this case $(\pi c/^{1/3})$ can be considered to represent a specific high energy photon. Thus the structure of the muon, written in short form is:

$$\mu=3^{-}(4/3\pi c^3)^{1/2}.(\pi/c^{1/3}) \tag{14}$$

[0166] Moreover the structure of the tauon can be calculated from the ratio of the mass of the electron and that of the Tauon (1.79 Mev);

[0167] Thus

$$0.511\ \text{Mev}/1.79\ \text{Gev}=2.85\times10^{-4}$$

$$m_e/m_\tau=(\pi/c)^{1/3}.(\pi/c)^{1/9}=2.85\times10^{-4}$$

[0168] As the charge of the tauon is equivalent to the charge of the electron, hence the structure of the Tauon is given by the above equation

$$e_\tau=\epsilon_{qe}.m_e/m_\tau.3^{-}(4/3\pi c^3)^{1/2}.(\pi/c)^{1/3}.(\pi/c)^{1/9}=e$$

[0169] This equation accurately predicts the charge −1; and mass of the Tauon (~1.78 Gev)(ref 1). Thus the structure of the Tauon can in short form be given by the equation

$$\tau=-3(4/3\pi c^3)^{1/2}.(\pi/c)^{1/3}.(\pi/c)^{1/9} \tag{15}$$

[0170] Furthermore a more exact value for the mass of the muon and tauon can be deduced by taking into account the gravitational field of the Earth (see Gravity and the Charge of the Electron pg. 17) in a similar way to identifying the exact charge of the electron. In addition it may be necessary to take into account a possible mass value of the neutrino to arrive at a precisely accurate mass value of the muon and tauon. Nevertheless, the mathematical proof of these short form equations lies in the fact that they can very closely identify the charge and the masses of these particles, from first principles, as in equations (6, 13-27, see Appendix 1).

[0171] Overall the mathematical geometrical structure of all the particles can be derived from the quasi electron, which is in turn derived from quintessence. Thus, the short form particle structures can now be derived from first principles. This includes the quasi electron (qe) and electron (e), from which the quarks (u,d) and in turn the stable proton (p) and stable neutron (n) and alpha particle (α) respectively are derived. The general structure of the force carrying bosons the photon. (g) and the gluon (γ) and the intermediate vector boson (W) can be given. It will also intriguingly be possible to derive, according to their generation, the structure of the strange (s) charm (c), beauty (b, or bottom) and truth (t or top) quarks directly from the structure of the muon (μ) and Tauon (τ) respectively.

[0172] Using the term $\Theta=(\frac{4}{3}\pi c^3)$, where, –/+ represents the charge of the quasi electron, we find:

[0173] 1st Generation:

$$q_e={}^{-}\Theta^{1/2} \tag{6c}$$

$$e=3^{-1/2} \tag{6}$$

$$d={}^{-}\Theta^{1/2}.3(\pi/c^{1/2}) \tag{16}$$

$$u=2^{+}\Theta^{1/2}.3(\pi/c^{1/2}) \tag{17}$$

$$s={}^{-}\Theta^{1/2}.3(\pi/c)^{1/2} \tag{18}$$

[0174] 2nd Generation

$$\mu=3^{-}\Theta^{1/2}.(\pi/c^{1/3}) \tag{14}$$

$$c=2^{+}\Theta^{1/2}.(\pi/c^{1/3}).(\pi/c^{1/4}) \tag{19}$$

$$b={}^{-}\Theta^{1/2}.(\pi/c/^{1/9}).(\pi/c)^{1/4} \tag{20}$$

[0175] 3rd Generation

$$\tau=3^{-}\Theta^{1/2}.(\pi/c)^{1/3}.(\pi/c)^{1/9} \tag{15}$$

$$t=2^{+}\Theta^{1/2}.(\pi/c)^{1/3}.(\pi/c)^{1/9}.(\pi/c)^{1/4} \tag{21 }$$

[0176] Particle Gluons (g):

$$g_1=(\pi/c^{1/2}) \tag{22}$$

$$g_2=(\pi/c)^{1/2} \tag{22a}$$

$$g_3=(\pi/c^{1/4}) \tag{22b}$$

$$g_4=(\pi/c)^{1/4} \tag{22c}$$

[0177] Particle Photons (γ):

$$\gamma_1=(\pi/c^{1/3}) \tag{23}$$

$$\gamma_2=(\pi/c)^{1/3} \tag{23a}$$

$$\gamma_3=(\pi/c^{1/9}) \tag{23b}$$

$$\gamma_4=(\pi/c)^{1/9} \tag{23c}$$

[0178] Intermediate Vector Boson ($W^{+/-}$):

$$W^{+}=3^{+}\Theta^{1/2}.2(\pi/c^{1/6})^6 \tag{24}$$

$$W^{-}=3^{-}\Theta^{1/2}.2(\pi/c^{1/6})^6 \tag{25}$$

[0179] Stable* Proton:

$$p=3^{+}\Theta^{1/2}.3(\pi/c^{1/2}) \tag{13}$$

[0180] Stable* Neutron:

Stable* Neutron:

$$n=\begin{matrix} {}^{+}\Theta\downarrow^{1/2}\cdot 2(3\pi/c^{1/2})\cdot{}^{-}\Theta\uparrow^{1/2} \\ {}^{+}\Theta\downarrow^{1/2}\cdot 2(3\pi/c^{1/2})\cdot{}^{-}\Theta\uparrow^{1/2} \\ {}^{+}\Theta\downarrow^{1/2}\cdot 2(3\pi/c^{1/2})\cdot\Theta^{-}\uparrow^{1/2} \end{matrix} \tag{26}$$

[0181] Alpha particle (α):

Alpha particle (α):

$$\alpha=\begin{matrix} [3^{+}\Theta]\downarrow^{1/2}\cdot 2(3\pi/c^{1/2})\cdot[3^{-}\Theta]\uparrow^{1/2} \\ [3^{+}\Theta]\uparrow^{1/2}\cdot 2(3\pi/c^{1/2})\cdot[3^{+}\Theta]\downarrow^{1/2} \\ [3^{-}\Theta]\downarrow^{1/2}\cdot 2(3\pi/c^{1/2})\cdot[3^{+}\Theta]\uparrow^{1/2} \end{matrix} \tag{27}$$

[0182] The mathematical proof for these structures and their decay mechanisms is lengthy and is thus fully contained in Appendix 1. All the particle structures are accurately mathematically defined by the masses of these particles.$^{(\text{ref }17)}$

[0183] The structure of these particles all contain the quasi electron and thus the metric tensor structure necessary in the formulation of the gravitational equations is sustained. The respective forces created by the gluon and the photon are important as they tell us the behaviour of matter and also lead to the likely structure of the graviton

[0184] Particle Spin and Size

[0185] The significance of the electron, composed of three spheres each with a radius of 1/c, is not immediately clear, but can be understood if the frequency of rotation of the electron is also taken into account. Knowing the structure of the electron has led us to deduce its charge and thus may lead us estimate its size and spin. Thus these observations might be used to calculate the radius and rate of rotation of the electron.

[0186] Let us suppose, that nature is truly beautiful, and that the radius of the fundamental quasi electron is indeed 1/c, and in turn the radius was balanced by the velocity of rotation 2π/c. This can be directly confirmed mathematically by taking into account the known spin of the electron, h/4π. Thus the actual spin of the electron may be calculated form the known energy of the spin.

[0187] The radius of the electron is not up till now known, but the radius of a quark has been estimated, and this is the radius derived from deep inelastic collisions of the proton. These estimates reveal a radius of approx. $r_p=1.18\times10^{-15}$ m.$^{(\text{ref }13)}$ This value may be used to assist in confirming the spin of the proton in revolutions per sec. (revs) and in turn the spin and size of the electron. Firstly we may proceed to estimate the spin of the proton. Thus as h=E.t (Joules×sec) and E.t=F.d.t (Joules×sec), then the spin;

$$h/4\pi=F.d.t \tag{28}$$

[0188] As F=ma, where $a=(\text{revs. }2\pi)^2 r_{qu}$ and m=the mass of the proton, then

$$h/4\pi=m(\text{revs}.2\pi)^2 r_p.\text{d.t}$$

[0189] The actual distance (d) travelled in a circle of half integer spin in 1 second is: $\text{revs}.\pi r_p$, thus:

$$h/4=m(\text{revs}.2\pi)^3 r_p^{\,2}/2$$

[0190] Hence:

$$\text{revs}={}^3\surd[h/m(2\pi)^4 r_p^{\,2}]$$

[0191] Taking the effective mass the proton as 1.6726×10^{-27} kg, then the rate of spin of the proton in revolutions/sec is:

$$\text{revs}=5.65\times10^{6}\text{ cycles/sec}$$

---

* Stable nucleonic neutron and proton conformations differ slightly from the Standard Model, this is due to the sharing of quasi electron and quasi positron particles within the nucleus, which allows stabalisation of these particles by the formation of stable 3x3 structures. The Standard conformations which describe non-nucleonic neutrons and protons are additionally given in Appendix 1.

[0192] From the frequency of the specific rotation of the proton, given the half integer spin associated with the proton, we can thus mathematically confirm the relationship between the radius of a particle and its spin:

$$r_p \times \text{revs}/2 = 1/c \tag{29}$$

[0193] Furthermore, the fundamental radius of 1/c seen in geometric structure the quasi electron, is also reflected in the rotation rate and radius for the proton, thus as above $1/c \div \frac{1}{2}\text{revs} = 1.18\times10^{-15}$ m. Moreover, this means the actual half integer velocity of rotation is none other than 2π/c in metres/sec. So that the particle is in harmonic balance.

[0194] Using the fundamental formula h/4π=F.d.t, it is possible to obtain accurate estimates of the radius and spin rates of the electron, or indeed any particle, using the same principle of harmonic balance. Using the formula:

$$\text{revs} = {}^3\surd(h/m\ (2\pi)^4 r_e^{\,2}]$$

[0195] It appears there are two unknowns, the radius if the electron and its revolution rate, however, in accordance with the equation, $r_p = 2/c.\text{revs}$, which gives the revolution rate of the proton, the same principle may also be used for the electron, by substituting $r_e = 2/c.\text{revs}$, such that:

$$\text{revs} = hc^2/4m_e(2\pi)^4 \tag{30}$$

[0196] Taking the mass of the electron 9.109382×10^{-31} kg, the rate of revolution of the electron is:

$$\text{revs} = 1.048\times10^{10} \text{ cycles/sec}$$

[0197] Which gives a predicted radius of the electron as

$$r_e = 6.366\times10^{-19} \text{ m.}$$

[0198] So the half integer rotation velocity (revs.πr) is 2π/c !, for the electron in keeping with the harmonic balance of the electron.

[0199] The same principle may be used to obtain an accurate estimate of the spin and radius of the muon, or any other particle. Using the above formula

$$\text{revs} = hc^2/4m_e(2\pi)^4 \tag{31}$$

[0200] Then as the mass of the muon is 1.8823×10^{-28} kg then the revs of the muon $f_\mu = 5.070\times10^7$ cycles/sec and the radius (r_μ) is thus 1.316×10^{-16} m.

[0201] It is now possible to begin to explain how the muon and the other subatomic particles are formed. If a quasi electron is complexed with another structure the total geometric structure needs to maintain harmonic balance. So the frequency of rotation would need to match geometric sructure with which the quasi electron was complexed

[0202] Intriguingly we find asymptotic convergeance for the formulas for frequency and mass occurs, when the geometric structure complexed with the quasi electron has the structure represented by $(\pi/c^{1/3})$ [giving the frequency divided by two, because the single integer spin of the force carrying particles compares to a half integer spin for the muon]. So that

$$(f_\gamma 3\pi/2)^{1/3} = f_\mu$$

[0203] When the ratio of the masses of the electron (m_e) and muon (m_μ) are related, such that:

$$m_e(3\pi/c^{1/3}) = m_\mu$$

[0204] Indeed we find that (allowing for the neutrino) this ratio is very close to the actual ratio of the mass of the electron to the mass of the muon, determined experimentally.

[0205] Furthermore, we have seen that these geometric structures, representing harmonics of the speed of light, which either match the frequency or the amplitude of vibration of the quasi electron, mathematically define the masses of the particles and the fundamental forces of Nature.

Part III

Quantum Gravity

[0206] Quantum General Relativity

[0207] Given the overall energy "complex" energy tensor structure of the electron and the metric tensor, assumed in general relativity, the quantum nature of gravity itself can now be explored. The spherical complex tensor for the electron and the positron (eq. 4, 5, 6) give the mathematical quantum structure and energy tensor for all the other particles (see appendix 2). Together with the time dimension these nine space dimensions account for the 10 parameters present in the metric tensor necessary to formulate the equations for gravity using Riemann geometry and thus forms the basis of quantum gravity. Intriguingly the metric tensor at each point in space time is required to consist of a collection of ten numbers, Consequently, ten dimensional space-time hypotheses, such as this or superstring theory, do automatically yield general relativity.(ref 3)

[0208] Furthermore, the mathematical representation of the graviton and the gravitational constant may be directly estimated from the knowledge of the mass and radius of quintessence. Thence, the force of the vibrations of quintessence lead directly to quantum gravity.

[0209] The radius of quintessence should be approximately in keeping with the Planck length estimate (r), which is conventionally derived from the standard dimensional equation:

$$r_q^{\,2} \approx Gh/c^3 \tag{32}$$

[0210] Given the nine spacial parameters present in the metric tensor, used in general relativity we find that the actual formula for $r_q^{\,2}$ is mathematically in agreement with theory when:

$$9r_q^{\,2} = Gh/c^3 \tag{33}$$

[0211] This again supports the 9 dimensional view of space and the size of the vibrations of quintessence can thus be estimated.

$$r_q = 1.35\times10^{-35} m \tag{33a}$$

[0212] This value is in agreement with the Planck length. Indeed if the above equation is correct then we find that we can derive the standard equation for the general relativistic increase in radius, r', (eq. 34) directly from first principles and arrive at a more fundamental equation for quantum gravity. As

$$r' = G.M/3c^2 \tag{34}$$

[0213] By substituting eq 33) into equation 34, a fundamental relationship between r' and M is obtained.

$$r'/3r_q^{\,2} = GMc^3/Ghc^2 = Mc/h$$

[0214] And substituting the quintessential equation, $h=m^qc^2$ (eq. 1) then:

$$r'/3r_q^2=M/m_qc=n_q/c \tag{35}$$

[0215] Hence the ratio of the change in radius to that of the radius of quintessence squared, is proportional, by a factor of c, to the ratio of the mass M of an object to that of the mass of quintessence, effectively the number of quintessences. Thus the change in radius, r' due to gravitation, is related to none other than the ratio of the mass and radius of an object to the mass and the square of the radius of quintessence. Thus again the gravitational change in radius is directly related to the number of quintessences.

[0216] Naturally, this would be exactly what would be logically expected if quintessence, like the equation for the charge of the electron (eq. 6) forms from a root sphere. Thus the change in spacial radius of a normal sphere is dependant on the square of the quintessential radius.

[0217] This increase in apparent radius represents none other than the (gravitational) binding energy for quintessence.

[0218] The meaning of the above dimensional equation (33) might itself be further understood by substituting the mass of quintessence (where $m_q=h/c^2$) into the equation. Thus in nine dimensions the gravitational constant (G) may be more logically given as,

$$9(\pi r_q^2/m_q)=G\pi/c \tag{36}$$

[0219] Where πr_q^2 is the cross sectional area of quintessence and m_q is the effective mass of quintessence, and thus $(\pi r_q^2/m_q)$ represents the effective mass per unit area which quintessence exerts. This equation reduces to:

$$9r_q^2/m_q=G/c \tag{37}$$

[0220] From this we may derive the standard general relativistic relationship for the apparent change in radius (r') around a mass (M), from an understanding of the mass (m_q) and number (n_q) of qiuntessences. As $m_q=M/n_q$, then:

$$3r_q^2=G.M/3c.n_q \tag{38}$$

[0221] Then if

$$n_q=r'c/3r_q^2 \tag{39}$$

[0222] thus directly substituting for n_q in eq 38:

$$r'=G.M/3c^2 \tag{34}$$

[0223] The importance of this is that the gravitational change in radius now logically derives from equation 36, which describes the gravitational force as resulting directly from the mass of quintessence exerted/per unit area of quintessence.

$$9(\pi r_q^2/m_q)=G\pi/c \tag{36}$$

[0224] Thus equation 34 is the conventional equation for the general relativistic increase in radius (r') in a gravitational field, which is here derived from the underlying nature of quintessence. Thus the gravitational constant is derived from the mass and radius of vibration squared of quintessence from first principles.

[0225] Indeed it is apparent that a more fundamental equation for gravitation now exists, for equation (39) is mathematically accurate and numerically agrees with eq. 34:

$$r'/3r_q^2=n_q/c \tag{39}$$

[0226] These equations may be readily mathematically verified. If in accordance with standard general relativity, the apparent increase in radius r' is:

$$r'=GM/3c^2 \tag{34}$$

[0227] Then given that the mass of the Earth is 5.974×10^{24} kg;

$$r'=1.478\times10^{-3}\ \text{m}$$

[0228] Accordingly if $r'=3r_q^2n_q/c$; (eq.39). Given the number of quintessences (n_q) constituting the Earth is M_E/m_q, then

$$n_q=5.9745\times10^{24}/7.3725\times10^{-51}=8.104\times10^{74}$$

[0229] As $r_q^2=1.823\times10^{-70}$ (eq. 33a) then:

$$r'=1.478\times10^{-3}\ \text{m}$$

[0230] Thus equation 39 gives the same answer as the standard equation and may be understood on a logical basis. Indeed the meaning of c in the equation may be understood as it has been previously shown as being the basis for the radius of matter (eq. 6). Hence the general relativistic change in radius, r', is none other than the effective binding energy for quintessence.

[0231] Quantum Gravity and Wave Particle Duality

[0232] Quantum gravity can now be readily linked with quantum mechanics, indeed any observations which are self consistent must be able to do so easily.

[0233] The frequency of light has been previously derived

$$f=E/h=n_q$$

[0234] Thus the formula for the frequency of light (E=hf) has previously been explained theoretically by the simple observation that the frequency is determined quite directly from the number of quintessences (n_q) within the photon. The same principle has also been shown to apply to matter.

[0235] Let us now follow these equations for matter by calculating the wavelength of a photon from the Gravitational constant as an example; and also as a test of these observations and to demonstrate that the gravitational equations can also apply to the quantum world.

[0236] If

$$n_q=r'c/3r_q^2 \tag{39a}$$

[0237] where r' is the is the general relativistic increase in radius, and r_q is the radius of quintessence (eq. 33). Where $f=n_q$, substituting for n_q, then the frequency of the photon f_γ(where $\beta=1$) is given by:

$$f_\gamma=r'c/3r_q^2$$

[0238] Uing the standard equation, $r'=GM/3c^2$ (eq. 34); we may substitute for r', thus we have:

$$f_\gamma=GM/9r_q^2c$$

[0239]

Thus

$$f_\gamma=\frac{G}{9r_q^2c^3}\cdot m_\gamma c^2$$

and as $E=m_\gamma c^2$;

-continued

$$f_\gamma = \frac{GE}{9r_q^2c^3} \tag{40}$$

[0240] Indeed as $9r_q^2=Gh/c^3$, then

$$f_\gamma=E/h=n_q$$

[0241] It is possible to also demonstrate that the same relationship holds for the wave equation for matter. If we take the relativistic wave energy of matter, which has been previously derived,

$$f=\beta^2 n_q$$

[0242] This includes the term for the number of quintessences flowing through the electron, in the complex vectors of space-time, to give the relativistic electron momentum (ρ) and a term for the rest mass, thus substituting into (40)

$$f_\gamma = \frac{GE}{9r_q^2c^3} \tag{40}$$

[0243] As $f=\beta^2 n_q$ for matter then the equation expands to:

$$f_m = \frac{G}{9r_q^2c^3}\cdot\beta^2E$$

As $\lambda = v/f$, then

$$\lambda = \frac{9r_q^2c^3v}{G\beta^2E} \tag{41}$$

[0244] Then the equation again reduces to:

$$G = \frac{9r_q^2c^4}{\lambda\beta E} \tag{3}$$

[0245] Equations 3, 40 and 41 are important as they show that the quantum wavelength of any particle of rest mass m can be derived from the gravitational constant G. Thus linking quantum mechanics to quantum gravity.

[0246] It is therefore important to confirm the numerical accuracy of the above equation (40). We can do this by comparing the result to the standard computation of the de Broglie equation, in a range where de Broglie itself is likely to be most accurate; which according to these observations is in the low energy range (see section on Wave Particle Duality).

[0247] If we take an electron with an energy of 0.1 KeV the wavelength is conventionally given (where the kinetic energy of the electron E_K is given by the product of the charge of the electron (C) and the potential applied eV=0.1 Kev), by the standard equation:

$$\lambda=h/p=h/(E_K.2m_0)^{1/2}$$

[0248] thus

$$\lambda=6.63\times10^{-34}/[1.602\times10^{-19}\times1\times10^{2}\times18.22\times10^{-31}]^{1/2}$$

[0249] hence

$$\lambda\cong1.23\times10^{-10}\,m$$

[0250] Using

$$\lambda = \frac{9r_q^2c^4}{G\beta E} \tag{3a}$$

[0251] Where $E=\gamma m_0c^2$

[0252] At 0.1 Kev, electron velocity is 6×10^6 r/sec, thus $\beta=2\times10^{-2}$ and $\gamma=1/(1-v^2/c^2)^{1/2}=1.0002$ Thus:

$$\lambda = \frac{9\times1.82\times10^{-70}\times80.78\times10^{32}}{6.67\times10^{-11}\times\beta\times1.0002\times9.11\times10^{-31}\times8.998\times10^{16}}$$

$$\lambda = \underline{1.21\times10^{-10}m}$$

[0253] Divergence between the de Broglie equation and the above equation (2) occurs at intermediate and high energies where it is generally accepted that the standard de Broglie equation may be less accurate (ref 18). The values for eq. 2 and de Broglie are compared to recent experiments, which demonstrate a relativistic curvilinear plot for wavelengths of matter in keeping with eq. 40.(ref 1)

[0254] The de Broglie equation in the non-relativistic format yields a simple log/linear scale, which is not in keeping with relativity; whereas eq. 3 is dependent on relativity and mathematically accounts for both relativity in calculating the wavelength. Indeed recent experiment on quantum tunnelling through a wire mesh strongly suggests that the relationship between energy and wavelength is relativisticalty curvilinear.(ref 1) Furthermore equation 3a suggests a fundamental relationship between energy (E), relative velocity (v/c =β), gravity (G) and the quantum wavelength (λ)

$$\lambda = \frac{9r_q^2c^4}{G\beta E} \tag{3a}$$

[0255] Indeed as ${}^9r_q^2$ =Gh/c^3, then

$$\lambda=hc/\beta E \tag{2}$$

[0256] Equation 2 is the very same as the Universal wave equation derived form first principles for the wavelength of tight and matter, which allowed a relativistic solution to the equations for wave particle duality (see Wave Particle Duality). This now indicates that these quintessential equations are compatible with relativity, quantum mechanics and quantum gravity.

[0257] Graviton Structure

[0258] From these observations, if the value for the gravitational constant is substituted into the equation (35) we may now estimate the probable geomteric structure of the gravtion, which is the force particle mediating gravity by acting on quintessence. Thus the Gravitational constant has been previously derived from the vibration of quintessence by the equation:

$$G.(\pi/c)=9(\pi r_q^2/m_q) \tag{36}$$

[0259] This is in accurate agreement with the value for G (6.67×10^{-11} N m^2 kg^{-2}). This suggests that the most probable mathematical representation of the graviton (ϕ), the third force carrying particle is

$$\phi=(\pi/c). \tag{42}$$

[0260] Thus the gravitational constant (G) can be given by the mass and radius of quintessence and the structure of the graviton

$$G=9\pi r_q^{\,2}/\phi.m_q \tag{43}$$

[0261] This shows the gravitational force to be related to the fundamental radius of quintessence space time, and the graviton.

[0262] Quantised General Relativity

[0263] The classical general relativistic formula, as given by Einstein is:

$$R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R=-\kappa T_{\mu\nu}$$

[0264] Where R is effectively the curvature of space-time, $R_{\mu\nu}$ denotes the contracted Riemann tensor of curvature and $T_{\mu\nu}$ is the "energy tensor" of matter.(ref 3)

[0265] If we substitute the energy tensor matrix of the electron (eq. 9) x time, for the energy tensor of matter $T_{\mu\nu}$; and the metric tensor of the space-time lattice x time for the contracted Riemann tensor we can arrive at the same solutions for general relativity.

[0266] Furthermore, in his published paper on General Relativity, Einstein(ref 3) defined the constant κ as:

$$\kappa=8\pi G/c^2$$

[0267] Therefore Einstein's equation should be written as

$$R_{\mu\nu}-1/2g_{\mu\nu}R=-\frac{8\pi G}{c^2}\cdot T_{\mu\nu} \tag{43}$$

[0268] Einstein himself was apparently not happy about the right hand component of the equation. However, we find that this part of the equation can now be explained and quantised by substituting the gravitational constant, $G=9\pi r_q^{\,2}/\phi.m_q$, (eq. 42a),

[0269] Giving:

$$R_{\mu\nu}-1/2g_{\mu\nu}R=-\frac{8\pi^2\cdot 9r_q^2}{\varphi\cdot m_q\cdot c^2}\cdot T_{\mu\nu} \tag{44}$$

[0270] By substituting $m_q.c^2=h$, and further substituting $h=h/2\pi$, we arrive at a quantised solution to Einstein's equations. Where A_q is the surface area of quintessence ($A_q=4\pi r_q^{\,2}$); ϕ is the graviton [ϕ=(π/c)] and h is Plancks constant. thus:

$$R_{\mu\nu}-1/2g_{\mu\nu}R=-\frac{9A_q}{\varphi\cdot\hbar}\cdot T_{\mu\nu} \tag{45}$$

[0271] The gravitational equation can now be further understood on a logical basis. The term $A_q=(4\pi r_q^{\,2})$, where $(4\pi r_q^{\,2})$ represents standard term for the surface area of a sphere of quintessence for the 9 space dimensions of the space time lattice, h is the energy content of quintessence x time and ϕ is the graviton, thus the right hand term now represents a true "metric energy tensor" of matter.

[0272] This leads directly to the standard solution to the field equations, for the general relativistic increase in radius r' of an object, where A is the surface area of a sphere of a given mass M, such that

$$r'=\surd(A/4\pi)-r=GM/3c^2 \tag{34}$$

[0273] Furthermore, although equation 45, gives the same solutions as Einstein's equation, which is essentially correct, the difference is that the equation is now dependant upon Planck's constant (h), and moreover the radius of quintessence, which now defines a quantised solution to the equations.

[0274] Graviton Force Characteristics

[0275] Similar to the photon, the previously derived equation (42) for the graviton [(ϕ=(π/c)] appears to also mathematically represent a helical ringlet of quintessence, but with a spin of 2. For the photon, taking the direction of motion as the x vector and its axis of spin also as the x vector, would account for the electromagnetic force and its attraction and repulsion characteristics. In the case of the gluon component ($\pi/c^{1/2}$), if the direction vector is x, then the axis of spin would be in the y vector, the same as quasi electrons, accounting for the particle binding characteristics of the gluon force. (see; Gluon structure and force characteristics. Appendix 2). In the case of the graviton, if the direction of motion was in the x vector, the graviton spin axis would be in the z vector thus, as will be demonstrated, accounting for the gravitational force.

[0276] The spin axis of the graviton can also be derived using the known characteristics of the electron. If an electron is travelling in the x direction, then its spin axis is determined by the by the sign of the ${}_{j}y$ vector (up or down). This view is in agreement with conventional theory, which indicates that the electron spin is similar to a rotating planet orbiting the sun, (the electron even appears to have orbital precession). As the electron passes through the space-time lattice, this spin would generate the formation of gravitons. This would occur as a result of the ejection of the excess quintessence passing through the electron. As the electron spins, the ejection of these gravitons would occur at a tangent to the electron's direction of motion. The ejection of the gravitons would occur, similarly to the ejection of energy of a pulsar or quasar, through the equivalent of the north and south poles of the electron. Thus, propelling the graviton in the direction of the electrons y vector. The ejection of the graviton would re-orientate and impart a specific angular momentum to the gravitons which would thus end up spinning on its own z axis. If for instance the graviton is released from an up spin electron the graviton will be rotating clockwise and its leading edge will displace quintessence downwards. In turn this will provide an upwards force.

[0277] This picture accounts for Fleming's left hand rule, is logical and provides an explanation for the magnetic force around a wire. According, to the left hand rule if the direction of the current is in the x vector, the magnetic field is in the z vector, and the force is upwards, in the y vector, in accordance with the above model. Therefore, this particular spin axis and the structure of the graviton results in its force characteristics. As the graviton is very small compared to the electron and both have different rather rapid spin axis it is difficult for these to bind and interact. Nevertheless, because the graviton has a spin of 2, and as it spin axis is perpendicular to its direction of motion, in the z vector, it readily displaces space-time quintessence to produce gravity. Thus because the graviton is able to displace space-time, it is capable of escaping a black hole. How else could the effects of gravity be felt beyond a black hole?

[0278] Quantum Gravity and Electromagnetism

[0279] With the above electron model of graviton production (see pp. 32), the nature of magnetism can be understood from first principles, Furthermore, the presence of a space-time lattice links relativity, and the forces of gravity with the electromagnetic and other forces of Nature. Indeed, evidence for these links may first date back to the 1820's, when Andre Ampere first defined the Amp. The force of attraction between two parallel wires 1 metre apart each carrying 1 Amp in a vacuum was defined as none other than the permeability of free space (2×10^{-7} N per metre of conductor). Thus conventionally the magnetic field strength around a long straight wire is given as:

$$B=\mu_0 I/2\pi r$$

[0280] Where I is the current and μ_0 is the permeability of free space ($4\pi\times10^{-7}$ N A^{-2}).

[0281] The attraction between two wires both carrying negative charge is, however, counterintuitive as negative charges should repel. A conventional explanation overcomes this by invoking the presence of a magnetic field which is created by the current by the production of virtual photons. Thus we appear to have an explanation for the effects of magnetism which involves virtual photons, however, these photons are not observed. More accurately, according to conventional special relativity the magnetic field is none other than the electric field viewed relativistically(ref 19).

[0282] A more satisfactory explanation, therefore, lies in the interaction between the electrons and the space time lattice. The moving electrons in the two wires interact with the lattice to produce gravitons; which are in phase when both streams of electrons are travelling in the same direction. The gravitonic waves interact constructively to disperse the space time lattice between the wires and induce an attractive force between the two wires, which produces in effect the permeability of free space. Thus this force results from the vibration of quintessence itself.

[0283] Conversly in two wires with current going in opposite directions the graviton waves are in anti-phase and would interact destructively between the wires. The gravitonic waves travelling radially outward from the wires would, however, disperse the lattice outside the two wires and produce apparent repulsion between the wires, which is exactly what is observed. These effects of electricity suggest that gravitons act as waves and that phase is important.

[0284] This effect is also seen with the north and south poles of ferromagnets. Nevertheless, with matter other than iron, cobalt or nickel, the graviton emission cannot be phased as the atoms are unable to align and magnets do not appear to exist with other materials.

[0285] In ordinary magnetic system the release of gravitons from the north pole would be exactly balanced by those released from the south pole of the magnet and hence there would be no nett force on the magnet until an external magnet or electrical current were applied.

[0286] Overall the magnitude of the forces in electrical systems where electrical conduction occurs are well defined by the permeability and permittivity of free space μ_0, and ϵ_0. Where v is the constant velocity of the charge and ϵ is the electric field produced by the charge.

$$B=[\mu_o.\epsilon_0]v.\epsilon$$

[0287] These observations suggest that the forces of electricity which produce magnetism are indeed related to the permittivity and permeability of free space and that these quantities are exerted by an apparent vacuum. Thus the effects of magnetism could be explained by none other than the phased effects of gravitational waves on the space time lattice.

[0288] Electromagnetism is of further interest to quantum gravity, particularly if we combine the standard equations, $B=\mu_0 I/2\pi r$ and $B=[\mu_o.\epsilon_0]v.\epsilon$, substituting for B we have:

$$2\pi r=I/\epsilon_0.v.\epsilon \tag{46}$$

[0289] Thus $2\pi r$ is proportional to the inverse of ϵ_0. Thus as space time is dispersed by gravitons the permittivity field will increase in the same way capacitance increases with separation of plates. Because of the inverse relationship between ϵ_0 and $2\pi r$, as ϵ_0 increases the circumference of a circle and the apparent ratio of π is to r, will appear to diminish in accordance with general relativity. This not an actual diminution in the circumference of a circle but the effective reduction of the resistance to motion in a circular path in this field.

[0290] Incidentally, the above observations, also lead us directly to Shrödinger's formula for the average equilibrium distance (r) between an electron with charge (e) in orbit around a proton, which is conventionally given by:

$$r=n^2 4\pi\hbar^2\epsilon_0/me^2$$

[0291] Where ϵ_o is again the permittivity of free space, m is the mass of the electron and n is an orbital integer, h is Planck's constant and e is the charge of the electron. Furthermore if $e=[\epsilon_0/3(4/3\pi c^3)]^{1/2}$ (eq.3); then the equation at n=1, for the electron orbital radius elegantly simplifies to:

$$r=4h^2c^3/m$$

[0292] Hence the orbital radius of the electron is related to spin of the electron (h) and its mass (m).

[0293] Quantum Gravity and the Charge of the Electron

[0294] The equation for the charge of the electron (eq. 1) contains the term so (permittivity of free space) which according to these observations should vary in a gravitational field.

$$e=\left[\frac{\varepsilon o}{3(4/3\pi c^3)}\right]^{1/2} \tag{6a}$$

[0295] If we combine the standard equations, $B=[\mu_0 I/2\pi r$ and $B=(\mu_o.\epsilon_0]v.\epsilon$, substituting for B we have:

$$2\pi r=I/\epsilon_0.v.\epsilon \tag{46}$$

[0296] Thus $2\pi r$ is proportional to the inverse of ϵ_0. Thus as space time is dispersed by gravitons the permittivity field will increase in the same way capacitance increases with separation of plates. Because of the inverse relationship between ϵ_0 and $2\pi r$, as ϵ_0 increases the circumferance of a circle and the apparent ratio of π to r, will appear to diminish in accordance with general relativity.

[0297] Thus ϵ_0 rises when space-time is dispersed by the gravitons that produce the gravitational field, This occurs in a similar way to the process by which capacitance increases with separation of plates in a capacitor.

[0298] Nevertheless, as c is a constant and as $c=[\mu_0.\epsilon_0]^{-1/2}$, then if ϵ_o rises then μ_0 falls. This is entirely consistent as μ_0, which represents the force that quintessence exerts, would be reduced if the quintessence space time lattice is dispersed.

[0299] Furthermore, as $\mu_0=4\pi \times 10^{-7}$ N A^{-2}; then as μ_0 falls, then the appareent ration π to r, also falls in a gravitational field. This is largely the same as stating, as does general relativity, that the apparent radius r', rises in a gravitational field. So this view is consistent with general relativity.

[0300] Nevertheless, to derive an exact value for the charge of the electron we must account for gravity in the above equation. We will take the specific example of the Earth's gravitational field in order to obtain the exact value for the electron. If in accordance with standard general relativity, the apparent increase in radius r' is:

$$r'=GM/3c^2 \tag{34}$$

[0301] Then given that the mass of the Earth is 5.9745×10^{24} kg; then

$$r'=1.47864\times 10^{-3}\ m$$

[0302] thus

$$2\pi r'=9.29057\times 10^{-3}$$

[0303] Which is the incremental factor by which ϵ_0 must increase in Earth's gravitational field. So to correct ϵ_0 to account for gravity, ϵ_0 must be divided by the incremental factor, 2πr'. Similarly as effectively π decreases in a gravitational field, to correct π to account for gravity it must be multiplied by this incrememtal factor. So the equation for an electron in a zero gravtational field is:

$$e=\left(\left[\frac{\varepsilon o}{3(4/3\pi c^3)}\right]\right)^{1/2}\div(1+2\pi r')=1.6022\times 10^{-19}\ C \tag{6b}$$

[0304] This now gives the charge of the electron as measured in a zero gravitational field as 1.6022×10^{-19} C, which is the same as that measured on Earth. Notably these observations appear to suggest that the charge of the electron is the same irrespective of the gravitiational field.

[0305] Virtually unlimited degrees of accuracy for the charge of the electron and for the fine structure constant (α), may be achieved by taking into account 2nd and nth order gravitometric effects. Thus if we take into account the effect of gravity upon the radius of the Earth it is also important to take into account an effect a upon the instruments with which we measure quantities, this would be a second order gravitometric effect. Thus taking into account 2nd order effects (r"), we have a very small, but nevertheless relevant change, such that: r"=r'(1+2r'). Thus $2r''=9.3180486\times 10^{-3}$, and thus:

$$e=\left(\left[\frac{\varepsilon o}{3(4/3\pi c^3)}\right]\right)^{1/2}\div(1+2\pi r'')=1.6021765\times 10^{-19}\ C \tag{6c}$$

[0306] This agrees exactly to the nearest 7 decimal places with the maximum accuracy of the experimental value for the charge of the electron.(ref 15) Furthermore by taking into account the nth order gravitometric effect, it is theoretically possible to predict accuracy for the charge of the electron to 3n decimal places. This mathematically accuracy confirms the structure of the electron from first principles and indeed the theoretical effects of gravity on the permittivity of free space (ϵ_0).

[0307] This returns us directly to the fine structure constant for the electron which is conventionally given by: $\alpha=e^2/\hbar c.4\pi\epsilon_0$. If $e^2=\epsilon_0/3(\tfrac{4}{3}\pi c^3)$, accordingly the quintessential equation for α is structurally given by: $2\pi/\alpha=m_q[3\Theta]^2$ (where $\Theta=\tfrac{4}{3}\pi c^3$; see The Structure of the Electron and Matter), we must now take into account the effects of gravity, as above, thus:

$$\alpha 2\pi/m_q[3\Theta]^2\div(1+2\pi r'')^2=0.007297353$$

[0308] Where the gravitational term for the increase in radius r" allows the mathematical derivation of α, and the above equation is in agreement with the conventional experimental value for α=0.007297353 to the nearest 9 decimal places.(ref 15)

[0309] Hence the term $(1+2\pi r'')^2$ is in accordance with these observations for the effect of gravity on electromagnetic forces. To a maximum accuracy governed by current knowledge of the mass of the Earth and the Gravitational constant and thus the term for the gravitational increase in radius r'. These observations can also be used to accurately predict the magnetic moment of the electron (see Quantum Gravity and the Electron Magnetic Moment).

[0310] Thus the presence of the fine structure constant can now be further understood, by deriving the constant from first principles; specifically from the actual dimensional conformation for the charge of the electron: $e=[\epsilon^0/3(\tfrac{4}{3}\pi c^3)]^{1/2}$ (eq. 6).

[0311] Overall the fine structure constant α (allowing for the term r' which is the general relativistic increase in the radius of the Earth due to gravitation) is given by none other than the formula for the mass of quintessence and from the structure of the electron, which can now be derived from first principles to seven decimal places or more.

[0312] Quantum Gravity and the Electron Magnetic Moment

[0313] The theoretical origin and nature of magnetism remains obscure in current electromagnetic theory. An explanation suggests these magnetic effects are produced by photons, although no photons have ever been observed. To get round this difficulty it is postulated by physics that magnetism results from "virtual" photons. However, Maxwell's equation for electromagnetism states that the photon has no nett magnetic effect.

$$\delta B_x/\delta x+\delta B_y/\delta y+\delta B_z/\delta z=0$$

[0314] Thus magnetism could not, by the above standard equation, be derived from a photon real or virtual.

[0315] In addition observational data suggests that black holes have powerful magnetic fields and as in theory photons are unable to escape from black holes (except for small quantites in the form of Hawking radiation), it would be difficult to explain these magnetic fields on the basis of photon emission.

[0316] Einstein postulated that magnetism was merely due to special relativity (ref 17). The postulate for the nature of magnetism in these current observations, states that the magnetic force results from relativity due to none other than the phased emission of gravitons (why postulate two invisible forces, magnetism and gravity, when one, the graviton, will do). This view as previously discussed (Quantum Gravity and Electromagnetism) is entirely compatible with standard relativity.(ref 19) Thus with the graviton origin of magnetism, the equation for the magnetic moment of the electron should have an expression in terms of quintessence and in turn the gravitational force and in particular the graviton.

[0317] The standard term for the magnetic moment of the Bohr Magneton (SIB) is:

$$\mu B=eh/4\pi m_e$$

[0318] In standard quantum mechanics the Bohr Magneton, μB, however, needs to be corrected to agree with experiment. The "correction factor" is termed "ϵ"; where $\epsilon=(\alpha/2\pi)-0.328\alpha^2/\pi^2)=0.001159641$. Thus theory reveals μ_e, the magnetic moment of the electron where:

$$\mu_e=(eh/4\pi m_e)[1+(\alpha/2\pi)-0.328\alpha^2/\pi^2)]$$

[0319] The conventional derivation of the term ϵ above, is given from the fine structure constant, $(\alpha/2\pi)$ which is theoretically consistent. However, a rather arbitary mathematical correction term; $0.328\alpha^2/\pi^2$ needs to be used in this standard equation. This appears ad hoc and needless to say, more accurate measurements show , the electron magnetic moment to the Bohr magneton ratio, $1+\epsilon=1.001159652$, which suggests the correction factor is indeed incorrect. Nevertheless, this correction factor is essential for "renormalisation" and thus for quantum mechanics to work.

[0320] Quantum gravity readily explains the discrepancy between the theoretical Bohr Magneton (μB) and the actual measured magnetic moment of the electron (μ_e). In accordance with the above chapter (Quantum Gravity and the Charge of the Electron)

[0321] Thus the significant mathematical discrepancies can be removed by accounting for the effects of quantum gravity.

[0322] Thus taking the charge of the electron (e), using the equation for the Bohr magneton and the effects of quantum gravity such that gravitational change in radius is r". The magnetic moment of the electron is given by:

$$\mu_e=(eh/4\pi m_e)(1+[\alpha/2\pi\div(1+r'')])$$

[0323] This gives an electron magnetic moment to Bohr magneton ratio of 1.00115968. Thus the mathematical term for the magnetic moment of the electron is given, avoiding the arbitary and dubious term $0.328\alpha^2/\pi^2$ used in the standard equation, simply by accounting for quantum gravity.

[0324] It is now possible to unite the equations for gravity and magnetism by substituting the fundamental key equations of quantum gravity. Thus if: $h=3m_qc^2$ (eq 1b) and $m_e=m_q.n_q$ (eq. 2). Then we can express the magnetic moment of any particle with the charge of the electron, including the proton, in terms of the number of quintessences (n_q) in that particle.

$$\mu B=ec^2/{}^4\!/_3\pi n_q \tag{47}$$

[0325] Given that the postulated structure of the graviton is: ($\phi=\pi/c$ (eq. 42), then substituting we have

$$\mu B=3ec/4\phi n_q \tag{48}$$

[0326] Showing that the equations for the magnetic moment are compatible with the gravitational equations given earlier. Principally, the quintessential equations now allow the determination of the magnetic moment of any charged object from the equation for the graviton and directly from the number of quintessences it contains. In conventional physics the magnetic moment of the electron requires a correction factor, $(1+(\alpha/2\pi c)-0.328\alpha^2/\pi^2))$, to derive the correct experimental value. These observations herein, indicate that the correction factor is more logically (1+r"), where r" is the general relativistic increase in radius around a gravitational body. This suggests that magnetism is not only affected by gravity, but can, as shown as above, be derived using the quantum gravitational equations.

[0327] Quantum Gravity and Special Relativity

[0328] Ordinary matter passing through the lattice would produce gravitons which would interact with space-time as described by general relativity. The quantity of gravitons would be determined by the apparent mass and in turn these would apparently curve space time. The geometry of this "curvature" is elegantly described by general relativity using Riemann geometry, specifically using metric tensors. Intriguingly the metric tensor is not a single number, but at each point in space time it is required to consist of a collection of ten numbers, Consequently, ten dimensional space-time hypotheses, such as this or superstring theory, may automatically yield general relativity. (see quantum general relativity p. 36)

[0329] General relativity is indeed very elegant, nevertheless there was a logical step yet to answer. That is, how do gravitons shape space time? This can now be readily answerd by considering the interaction of a three dimensional space time lattice with gravitons themselves to produce the effects of gravity. The effects of gravity are as such to compel a body in motion towards the gravitational object and to a much smaller extent visa versa. This effect can only be produced If gravitons repel quintessence (the constituents of the 3D lattice). Indeed, it has been stated that in order to explain cosmic inflation and the "flatness" of the Universe that quintessence must shun (or be shunned by) matter (ref 5,6,7).

[0330] In descriptive terms a body close to a large mass will have a tendency to move toward it because the three dimensional lattice would be less dense as it approached the surface of the large mass. Overall there would be less resistance to motion in the direction of the large mass, and the motion in this direction would be facilitated by the vibration of quintessence.

[0331] In general relativity the principle governing motion is the geodesic of least distance, this can be re-expressed using similar equations using least action. Furthermore, the concept of motion due to the vibrations of quintessence is more logically and experimentally compelling.

[0332] These observations can now be used to link general and special relativity. Thus as we approach the speed of light, the mass of an object travelling through the space-time lattice would approach infinity, directly because the number of quintessences passing through a body would increase with increasing velocity, hence the equation:

$$m'=m_0/(1-v^2/c^2)^{1/2} \quad (50)$$

[0333] or

$$m'=m_0/(1-v^2[\mu_0.\epsilon_0])^{1/2} \quad (50a)$$

[0334] In turn this would generate increasing gravitons and accordingly this would explain the observed effects of special relativity. Time itself is due to passage through the space-time lattice, and where the space-time lattice is dispersed by gravitons, time and length are reduced with increasing velocity and hence increasing space-time lattice dispersion, similar to the way in which gravity alters space-time

[0335] As a result:

$$t'=t(1-v^2/c^2)^{1/2},\ l'=l(1-v^2/c^2)^{1/2}$$

[0336] Thus resulting in the effects of special relativity.

[0337] Quintessence and Black Holes

[0338] To address the relationship of the space-time lattice to gravity directly, it is important to discuss the concept of quintessence with regard to general relativitivistic equations. The standard general relativistic equation for the apparent increase in radius (r) due to the curvature of space time around a gravitational object, which has also been previously derived from first principles (eq. 36), is:

$$r'=G.M/3c^2 \quad (34)$$

[0339] This can also thus be written as:

$$3r'=G.M[\mu_0.\epsilon_0] \quad (51)$$

[0340] This standard equation, is in keeping with the above observations. Specifically, as the mass increases, ϵ_0 increases, in turn the radius will appear to increase (relative to π).

[0341] The above observations now allow us to examine the effects with regard to the interior of black holes themselves. The event horizon would represent a critical density for quintessence, in which light could not escape. The Schwarzschild radius would now be given by:

$$R_s=2GM[\mu_0.\epsilon_0]$$

[0342] The event horizon will occur at the point at which there is less resistance to circular motion than motion in a straight or partially curved line. Given that π is proportional to $1/\epsilon_0$. The event horizon should occur when the permittivity has increased by a factor of π.

[0343] Effectively because the permittivity of free space rises, π decreases. This is entirely in keeping with general relativity which predicts the effective change in the ratio of the radius to the circumference as given by the conventional equation, where r', is again the apparent change in radius.

$$r=G.M/3c^2$$

[0344] Hence π will effectively decrease as we approach the event horizon of a black hole, When π decreases to 1, the circular circumference is equal to the diameter and moreover, inside this limit it is shorter for light to travel in a circle. Thus light cannot escape the event horizon.

[0345] This can give us great insights into the workings of space-time, for flat Euclidean space the standard equation is:

$$e^{i\pi}=-1$$

[0346] In accordance with general relativity, the ratio of the radius to the circumference changes in a gravitational field, and effectively π=1, at the event horizon, thus the boundary condition for the shape of space-time at the event horizon now has the direct equation:

$$e^{f}$$

[0347] Within a black hole as the permittivity of space increases by a factor of 2π an object within it will complete two rotations rather than travel in a straight line. In effect exceeding the speed of light by 2π. Hence, the condition for space-time is represented by the equation:

$$e^{i/2}$$

[0348] Thus an increase in the permittivity of free space by a minimum factor of π, to produce a black hole is estimated to result from an increase in mass by a factor of approx. 10^6 (the ratio of the mass of the earth and that of a putative black hole).

[0349] Continuing with the subject of a black hole, according to the model inside the black hole, the gravitons produced by the matter present would be in equilibrium with the density of the space-time lattice. Increasing the rate of rotation of the matter in the black hole for instance would thus increase the production of gravitons and its effective mass and increase the radius of the event horizon. A density gradient of the space-time lattice would continue to exist within the black hole. Progressively closer to the centre of a black hole matter itself would be increasingly compressed and the spherical structure of the quasi electron would be predicted to collapse. This collapse would result in the formation of an exotic form of matter in the form of pure quintessence in a black hole.

[0350] This pure quintessence would produce the singularity at the centre of the black hole. The larger the black hole in terms of mass the more pure quintessence would exist at its core.

[0351] Quintessence and the Big Bang

[0352] Quintessence theory not only predicts the occurrence of the Big Bang, but allows a prediction for the value of the entire mass of the Universe, from first principles.

[0353] In accordance with quintessence theory the big bang resulted from the explosion of an immense black hole singularity, which was constituted from pure quintessence.

[0354] On the basis of quintessence, there will be a critical mass for Big Bang event; thus if entire space-time between quintessence is compacted so that no further quintessence can be accommodated, the addition of further quintessence would destabilise the immense black hole, resulting in the Big Bang.

[0355] It is possible to predict this critical mass, using the radius of quintessence as a benchmark. Given the nine spatial parameters present in the metric tensor, used in general relativity we find that the actual formula for the radius of quintessence; r^{q2}, is mathematically in agreement with general relativistic theory when:

$$9r_q^2=Gh/C^3 \tag{33}$$

[0356] This again supports the 9 dimensional view of space (so crucial in superstring theory). Moreover, the size of the vibrations of quintessence can thus be calculated as:

$$r_q=1.35\times10^{-35}m \tag{33a}$$

[0357] The volume of each quintessence is thus:

$$\tfrac{4}{3}\pi r_q^3=1.0306\times10^{-104}\,m^3 \tag{33b}$$

[0358] So to be accommodated within unit volume of space time, with no intervening apparent space time, (given that each of 9 overlapping quintessences are required) would require approx.

9×10^{104} quintessences

[0359] As the mass of quintessence is $m^q=h/c^2=7.373\times10^{-51}$ kg.sec (eq. 1). Then the mass of of the Universe, to two decimal places, is:

$$1.18\times10^{53}\text{ kg} \tag{33c}$$

[0360] This is in close agreement with a recent estimate of the mass of the Universe from COBE and other satellite data, which estimates the mass to be 100 trillion trillion trillion trillion tonnes (10^{53} kg). (ref 20)

[0361] Moreover, the early formation of the galaxies can be readily explained, it is likely that in such a big bang some very small black holes might have prevailed and that these formed the seeds of the galaxies we see today.

[0362] The event horizon, calculated from the Schwartzschild radius, of such an immense black hole is $\sim10^{26}$ m, which would have allowed Guth's inflationary component to the early expansion of the Universe.

[0363] In addition, inflation may result directly from the observation that once electrons have formed from the primordial soup of quintessence, they emit gravitons which in turn repel space time, which might also result in another cosmic inflationary cycle.

[0364] Most importantly quintessence theory explains the Big Bang from first principles and and is capable of accurately predicting the mass of the Universe.

[0365] The Nature of Energy

[0366] Foremostly, these observations allow a fundamental understanding of energy. The quantum physical, minimum component of energy is Planck's constant; h. To define the minimal component of mass, using the standard energy equivalence formula; $E=mc^2$, such a minimal mass (m_q) would be required to have the value equivalent to; $m_q=h/c^2$ (1). The total mass of a system (m) would then be; $m=m_q n_q$, where n_q is the number of these minimal units. Thence, the total energy of a system can be derived from the minimal energy; h, multiplied by the number of these energy units (n_q). Thus as, $E=mc^2$, then also $E=m_q n_q c^2$ and substituting $m_q=h/c^2$, the energy equivalence formula has the more logical formulation;

$$E=hn_q \tag{1a}$$

[0367] Thus the energy of a system is equivalent to the minimal energy unit; h, multiplied by the number of those minimal energy units (n_q).

[0368] This leads directly to a deeper understanding of wave particle duality and the wave nature of matter.

[0369] This is encapsulated by the quintessential energy formulae

[0370] As conventionally $\beta E/c=p$, then

$$\lambda=h/p=hc/\beta E \tag{2}$$

[0371] and

$$E=hn_q \tag{1a}$$

[0372] then

$$\lambda=c/\beta n_q \tag{2b}$$

[0373] Importantly, as indicated by equation (2b), energy having no quintessence; would have a wavelength of infinity. Specifically pure energy containing no quintessences, would have a lambda of infinity. According to quantum mechanics an infinite wavelength would result in the probability of that energy being anywhere. As energy itself has no electrical charge it would not be impeded by the permittivity and permeability of the three dimensional space-time lattice. Moreover, energy would not be detectable in three dimensional space-time, unless it interacted with matter, as in the EPR experiments. Indeed, energy is not observed when not bound to any form of mass or particle.

[0374] Thus equation 2b, takes us to our original assertion regarding the existence of pure energy.

-Energy is not bound by the space-time lattice-

[0375] Thus, as the EPR experiments suggest the existence of energy separate from matter and thus separate from the three dimensional space-time lattice, it is interesting to find that experiment suggests the existence of free energy in a continuum separate from space time and matter to produce the effects of quantum teleportation.(refs 10, 11, 12, 13)

[0376] This is not, however, teleportation across an additional dimension, this is a term to describe in partially familiar terms the dissociation of energy from the three dimensional space-time lattice. As time is inextricably linked to each dimension of space, the effects of energy would be inextricably linked to the events, such as the creation of virtual particles, we see interacting within space-time. It is unlikely that observers have any direct day to day experience to explain quantum events. Nevertheless, quintessence theory may have given us a window into the hitherto hidden workings of the Universe. Thereby, the mystery of the uniformity of the Universe, across distances which the speed of light could not apparently traverse, is readily explained by the fact that the free energy contained in the Universe is not bound by the space-time lattice.

[0377] In the case of light, due to the exceedingly small masses involved, there would be relatively easy exchange of matter with free energy within a photon. This would make the photon the ideal experimental tool to look for energy which is not bound by matter and in turn energy which is not bound in space-time. Indeed, very recently Furusawa et al. have reported to have observed the transference of energy as photons from A to B, without those photons traversing space-time(ref 10). This finding which has been supported using other experimental techniques(refs 11, 12, 13,), is very important as it suggests the existence of such free energy.

[0378] Overall, quintessence theory gives an a priori explanation for the concept of mass, the elementary particles, the forces of nature and quantum effects. It can equally be used logically to explain the inner physics of a black hole, the missing mass in the Galaxy, the expansion of the Universe, Guth's inflationary theory and predicts the Big Bang, from first principles.

Prior Art

[0379] Antigravity Systems

[0380] Recent research at NASA indicates that an antigravity device may be feasible. This research, however, has been undertaken at NASA without fundamental knowledge of the theoretical basis of quantum gravity. In these experiments the use of radio frequency (RF) pulses, such as in Nuclear Magnetic resonance, on rotating superconducting discs has been attempted, to produce antigravity effects.(ref 21)

[0381] However, radio frequency pulses are a weak form of electromagnetic radiation and are therefore unlikely to provide sufficient energy to produce measurable antigravity effects.

[0382] The warp drive technology described in this application is entirely different to previous prior art. The most important difference relates to the nature of the invention. In this application specifications are given for a direct drive system, as opposed to a gravity shield device, which other inventions claim. There is of course a vast difference as a shield device, merely reduces the force of gravity, but does not directly produce drive.

[0383] The references specifically relating to the shield effect are:

SCHNURER:	WO 98/23976	14 Nov. 1996
MODANESE + SCHNURER:	arXiv:gr-qc/9612022 v4,	19 Feb. 1998
PODKLETNOV:	arXiv:cond-mat/9701174 v3,	16 Sep. 1997
PODKLETNOV + NIEMINEN:	Physica C, Vol 203, 441–444,	9 Sep. 1992

[0384] Other devices produce an attractive effect, as in an electromagnet magnet, which is again entirely different to the direct production of drive.

[0385] HOOPER U.S. Pat. No. 3,610,971 Apr. 15, 1969

[0386] In these and other antigravity devices the use of equal and opposite currents in a superconductor to produce anti-gravity effects are disclosed, as in:

[0387] BETTELS DE19832001 Jul. 16, 1988

[0388] In this application for Warp Drive technology, there is a differential current applied. Thus the differential speed of the electrons will, in accordance with relativity, produce differential mass, which will in turn result in a differential graviton release. Thus this invention is very different to previous disclosed inventions, as it is the differential graviton release, which produces the effective warping of space-time, to produce a direct warp drive system, which operates by the warping of space-time.

SUMMARY OF THE INVENTION

[0389] Principles of Warp Drive Systems

[0390] The theoretical understanding of quantum gravity allows the design of warp drive systems from first principles. It is unlikely that gravitons can be controlled in a precise way using current technology. Nevertheless, an understanding of three dimensional space-time and matter, does allow the design of elementary warp drive systems. That is, systems whose propulsion rely on warping space-time as opposed to the ejection of material to provide thrust.

[0391] The background for these systems are already partially understood and quintessence theory allows their further development. For instance, the formation of black holes, represents a crucial step in understanding the mechanisms that underlay gravitational physics and in turn the warping of space-time. The existence of black holes, implies that the warping of space-time allows a powerful system for driving matter. In this instance the space-time lattice would be repelled by gravitons in such a way as to disperse space-time quintessence in a circular fashion, producing an event horizon. Similarly if we suppose that gravitons could be controlled and collimated in a single direction similar to a laser using photons, it would be possible to focus such a graviton beam ahead of an object. This in turn would dissipate the space-time lattice in front of that object, thereby allowing the theoretical potential for what is termed Warp drive.

[0392] Essentially, these are systems whose propulsion relies on warping space-time, as opposed to the ejection of material to provide thrust. This does not, of necessity, mean that superluminal velocities are produced, but that the drive is based on the warping of space-time in accordance with general relativity. However, it would in the future theoretically be possible for an object to achieve speeds greater than that of light, as superluminal velocities produced by the warping of space-time do not contravene general relativity.

[0393] Indeed according to the standard model, if the horizon distance is taken as the maximum distance which light could have travelled, then two points on opposite sides of the Universe are as far apart as ninety times the horizon distance, which indicates that components of the Universe themselves have effectively far exceeded the speed of light.(ref 20a)

[0394] With current technology warp drive systems (not of necessity superluminal) could be achieved by rotating intermediate sized superconducting masses at ultracentrifugational velocities. The release of gravitons could be controlled by differentially governing the electron flow through these masses with the use of powerful electric currents. In turn the differential direction of flow of gravitons would determine the direction of motion through the space time lattice.

[0395] In order to design a mechanism for elementary warp drive we may utilise either a normal conductive material or preferably a superconducting material, or a combination of both. It may be possible to use any shape, such as a disc, cylinder or preferably a sphere. It would also be possible to use dual counter rotating discs to minimize the net angular momentum of the system. In the case of a sphere, this is rotated along its horizontal axis in a clockwise fashion at ultracentrifugational speeds. A differential current is then applied from left to right in such a way as to pass through the entire sphere in this direction.

[0396] The electric charge in the upper half of the sphere would be maximised. The electrons will have a vector in the left to right direction as the sphere spins clockwise. If a maximised current is applied to the sphere in the same direction this will result in a increase in the velocity of the electrons relative to the centre of gravity of the rotating object, due to the flow of current. In turn, according special relativity and to the space-time lattice model, this wilt result in an increase in the relativistic mass if the electrons and in turn by general relativity an increase in the release of gravitons.

[0397] Conversley in the lower half of the sphere the electrons wilt have a vector of motion in the right to left direction due to the spin of the sphere. This will be relativistically slowed by the differential current applied in the same direction as the current above, and hence in the opposite direction to the direction of rotation. The charge can be separately applied and adjusted to ensure that the electrons are relativistically stationary relative to the centre of gravity. In turn this will minimize the relativistic mass and result in a decrease in the release of gravitons for the lower half of the sphere.

[0398] The overall result will be a greater release of gravitons in the upward direction and a lesser release of gravitons in the downward direction. The effect will be enhanced by the use of a multi-phasic current simultaneously applied. This will result in the release of multiphasic gravitons which will disperse space time above the sphere with increase in density below the sphere which will effectively reverse the effects of Earths gravity. Importantly this effect can also be produced with the use radio frequency pulses. This obviates the need for commutator devices. Nevertheless, the radio frequency pulses must be designed to produce a change in the spin of the particle to enhance the release of gravitons in the upward direction, and moreover with current technology theses RF pulses contain too little energy to effect a significant change in the gravitational field.

[0399] With the use of large currents the drift velocity of the electrons could be greatly increased. The use of superconducting devices would greatly enhance the efficiency of such systems as the electrical resistance is virtually eliminated. Thereby allowing large currents to be used with minimum total power output.

[0400] The technological crux of the device is to produce electrons of high speed and hence high relativistic mass in the top half of the sphere, whilst producing low speed and thus low mass electrons at the bottom of the sphere, in accordance with special relativity. The imbalance in the rotating sphere will be continuously present impelling the device upwards. In effect the differential current flow, will produce differential graviton production and in turn, by general relativity, the warping of space-time

TECHNICAL EXAMPLE 1

[0401] A small scale device, which produces warp thrust can be readily constructed using current technology.

[0402] This wilt require a sphere, cylinder or disc of superconducting material. A dual disc either co rotating or counter rotating could also be used, counter rotating discs have the advantage that the total angular momentum is effectively zero. Any of the superconductors may be used such as Ag_2F, C_8K or SnTe. However for their tensile strength and/or magnitude of critical current density (J_c), Nb_3Sn or Ag_2F is recommended. Normal conductive material may be used but due to electrical resistance the power required (see superconductivity theory pg. 9) would be in the order of 10^{19} times greater.

[0403] Taking a sphere of Nb_3Sn, made hollow to reduce the mass of the sphere, with an external radius 0.25 m and an internal radius of 0.20 m; the total volume of Nb_3Sn required would be 3.19×10^4 cm^3. As Nb_3Sn has a density of 7.86 g/cm^3 the total mass of superconducting material would hence be 250 kg.

[0404] This sphere is required to be mounted upon two ultracentrifugational spindles which themselves are on the horizontal axis. The untracentrifugational spindles would be powered to rotate the sphere to produce an enhanced g force, as in a centrifuge.

[0405] The superconductive sphere would require a supercooled jacket of liquid Helium, and would require to be in a vacuum. The power source, motors and refrigeration system should be placed outside the supercooled jacket. In addition these ancillary elements of the device may be either placed outside the system as a whole, or within it, if the device is designed as a self contained vehicle. The device may also be mounted upon vertical rails to experimentally demonstrate vertical lift. (See **FIGS. 1, 2**)

[0406] A high current needs to be applied, by means of a commutator device, to produce a current passing through the sphere. If the negative terminal were placed on the left of the sphere and the positive terminal on the right this would produce electrons flowing from left to right. The commutator device would be split to allow a smaller current to be applied to the lower half of the sphere.

[0407] The power expenditure for this is reduced by using a superconducting material. If the sphere were rotating from left to right i.e clockwise the drift velocity of the electrons would be advanced in the upper half of the sphere, and retarded in the lower half of the sphere by the differential current induced flow of the electrons.

[0408] According to relativity, as defined in the theoretical sections above, it is accepted that the mass of the electrons in the upper half, relative to the centre of gravity, of the sphere can thus be increased by using the standard relativistic formula:

$$m'=m_0/(1-v^2/c^2)^{1/2} \tag{50}$$

[0409] Thus by inducing a relative drift velocity difference, the mass of the electrons at the top of the sphere can be increased, and the mass of the electrons at the bottom of the sphere can be reduced relative to the centre of gravity.

[0410] If the object is set to rotate, centrifugal forces can be used to increase this mass difference. This can be used to increase the net momentum of the object in the upwards direction, to produce upwards thrust.

[0411] Given that there are approximately 10^{23} per cm^3 free electrons in a niobium conductor then in the total sphere there will be approx. $10^{23}\times3.19\times10^{4}$ $cm^3=3.19\times10^{27}$ free electrons in the sphere. If each electron weighs 9.11×10^{-31} kg; then the total mass of free electrons in the upper half of the sphere is 1.453 grams.

[0412] If we take the maximum zero voltage current density (J_c) in Nb_3Sn or another suitable superconducting device as approximately 10×10^{14} Amps/m^2, the drift velocity in the electrons can be calculated. Given that the average surface area of the hollow sphere is approximately 350 cm^2 or 3.5×10^{-2} m^2, then the max current applicable is 3.5×10^{13} Amps. Thus the drift velocity of the electrons would be 6.3×10^{6} m/sec. The velocity of the electrons in the upper part of the sphere would be advanced relative to the centre of gravity. The relative velocity to this centre would be equivalent to the velocity of the sphere plus the velocity of the electrons in the upper half.

[0413] If the drift velocity of the electrons in the lower half were separately, induced by the current so that these electrons were stationary relative to the centre of gravity, those in the lower half of the sphere would be relatively retarded. The current applied to the lower half of the sphere to make the lower electrons, relative to the centre of gravity, stationary (given a rotation rate of 3.14×10^{4} revs/sec) would be 2.46×10^{11} Amps.

[0414] This differential current would allow the electrons to have a differential velocity and in turn a different mass relative to the centre of gravity. The relative velocity would be approximately 6.3×10^{6} m/sec. Using the standard equation for relativity (50) would give a relative mass difference ratio of 2.21×10^{-4}.

[0415] The effective total mass difference between the upper and lower half of the sphere would thus be the mass of the electrons times the relative mass difference which is 1.45 grammes$\times2.21\times10^{-4}=3.2\times10^{-7}$ kg.

[0416] Overall, as the acceleration due to gravity is 9.81 m/sec^2; to produce upwards lift on a sphere of 250 kg in Earth's gravitational field would require a force of about 2,500N. In order to produce in excess of this force the acceleration produced by the ultracentrifuge device would need to be about=8.76×10^{9} m/sec^2. Thus according to the formula $a=(\omega2\pi r)^2/r$ (where the average radius, r of the device is 0.225 m) the rotation rate required would be 3.142×10^{4} revs/sec or approximately 1,880,000 revs/min to produce sufficient acceleration to allow the device to completely self levitate. Thus as F=ma, then the force produced by the device is

$$3.2\times10^{-7}\ \text{kg}\times8.76\times10^{9}\ \text{m/sec}^2=2{,}800\ \text{N}$$

[0417] A detectable anti-gravtiy effect could, however, be realised at much lower spin rates. Standard ultracentrifugation devices can rotate at 600,000 revs/min. Thus even at standard ultracentrifuge speeds the g forces produced would be sufficient to produce approximately 12.8% lift, using these anti-gravity techniques.

TECHNICAL EXAMPLE 2

[0418] For this example we will use a large scale device, which would therefore deliver sufficient thrust to power a vehicle beyond Earth's gravitational field. Under terrestrial conditions the entire devise would be required to be encased in a liquid helium jacket, with an internal vacuum to reduce friction due to air. However, the tatter two constraints need not be in place if the device is to be used in space as, very conveniently, space is both a vacuum, and when shielded from the rays of the sun the ambient temperature is less than that of liquid helium. Indeed the large mass of the superconductor becomes less relevant as weightlessness exists in space, so that the size of the object may be increased substantially.

[0419] Taking a sphere of 1 meter in radius of a superconductive material. The volume is $\frac{4}{3}\pi r^3=4.2\times10^{6}$ cm^3. If the high temperature superconductors are used such as $YBa_2Cu_3O_7$, or $Nd_{1.85}Ce_{0.15}CuO_4$, which are ceramic in nature, the specific density is approximately 3 g/cm^3, Hence the total mass of the sphere would be approximately 12,600 kg or 12.6 metric tonnes. If the standard superconductors are used, for instance Nb_3Sn with a density of 7.86 g/cm^3 the sphere becomes heavier at 33 metric tonnes

[0420] This sphere is required to be mounted upon two ultracentrifugational spindles which themselves are on the horizontal axis. The untracentrifugational spindles would be powered to rotate the sphere to produce an enhanced g force, as in a centrifuge. The sphere would need to be carefully constructed to be exactly balanced in all directions.

[0421] Nevertheless, a high current does still need to be applied, by means of a commutator device, to produce a current. If the negative terminal were placed on the left of the sphere and the positive terminal on the right this would produce electrons flowing from left to right. The power expenditure for this is reduced by using a superconducting material. If the sphere were rotating from left to right i.e clockwise the drift velocity of the electrons would be advanced in the upper half of the sphere, and retarded in the lower half of the sphere by the current induced flow of the electrons. The split commutator device allows a differential current to be applied to the upper and lower halves of the sphere respectively.

[0422] According to the model in the preceding sections this will result in an increase in the mass of the electrons in the upper sphere and hence an increase in graviton production in the upper half of the sphere. Indeed, this model is in agreement mathematically with general and special relativity.

[0423] Given that there are approximately 10^{23} per cm^3 free electrons in a copper conductor then in the total sphere there will be approx. $10^{23}\times4.2\times10^{6}=4.2\times10^{29}$ free electrons in the sphere. If each electron weighs 9.11×10^{-31} kg; then the total mass of free electrons in the upper part of the sphere is 0.191 kg.

[0424] If we take the maximum zero voltage current density (J_c) in Nb_3Sn, or another suitable superconducting device, as approximately 10×10^{14} Amps/m^2, the drift velocity in the electrons can be calculated. Given that the average surface area of the sphere is approximately 1.57×m^2, then the max current applicable is 1.57×10^{15} Amps. Thus the drift velocity of the electrons would be 6.3×10^{6} m/sec.

[0425] At the same time the current in the lower half of the sphere would be adjusted to retard the electrons so that they remain stationary relative to the centre of gravity. If the sphere were rotating at 2.6×10^4 revs/sec this would require a current of 2.04×10^{13} Amps. This differential current would induce a differential velocity in the electrons and in turn a mass difference.

[0426] By inducing a difference of mass as little as one part in a thousand in these electrons we can achieve significant lift. Thus by the above equation if we induce a drift velocity of 6.3×10^6 m/sec (as in technical example 1), given that the velocity of the electrons in the upper part of the sphere would be advanced relative to the centre of gravity and those in the lower half of the sphere relatively stationary; then the relative velocity would be approximately 6.3×10^6 m/sec. Thus between the upper and lower half of the electrons we increase the relative mass of the electrons in the upper half by a factor of 2.21×10^{-4}, as in technical example 1.

[0427] Thus the difference in mass in the upper half of the sphere due to the moving electrons would be 0.191 kg$\times2.21\times10^{-4}=4.22\times10^{-5}$ kg., relative to the lower half electrons. This difference is enhanced by producing a centrifugal force by rotating the sphere.

[0428] If we rotate the large superconducting sphere at 2.6×10^4 revs/sec or 1,560,000 revs/min we can achieve accelerations of 1.33×10^{10} m/sec^2 (given an average radius of 0.5 m) by the formula $a=(\omega2\pi r)^2/r$.

[0429] Thus as F=ma, where m is the relative mass difference of the electrons the thrust produced in the device would be equivalent to:

$$4.22\times10^{-5}\ \text{kg}\times1.33\times10^{10}\ \text{m/sec}^2=5.88\times10^5\ \text{N}$$

[0430] Thus in Earth's gravitational field this force would be capable of levitating 60 metric tonnes.

[0431] The estimated current required to produce the drift velocity of the electrons for this effect would be approx. 1.57×10^{15} Amps. The power usage is given by; $P=I^2R$, under non-superconducting conditions, given that Niobium has a resistivity of 15.2×10^{-8} Ω/m the sphere would have a resistance of approximately 10^{-7} Ω. The power used by a non-superconducting device would thus be 2.46×10^{23} Watts. However, due to superconductivity, under standard theory, the device has zero resistance and would use no power$^{(\text{ref }8)}$. Nevertheless, if we wish to use the model described above for superconductivity (see superconductivity, page 6) there would be a practical power consumption which would nevertheless still be low. Thus the power consumption would be in the order of 1.6×10^{-19} less than the standard power usage, at 3.94×10^4 Watts.

[0432] If Nb_3Sn is used to make the sphere the mass of the sphere would be 33 metric tonnes. If the mass of the ancillary ship were 27 metric tonnes, giving a total of 60 metric tonnes the acceleration produced would be 9.81 m/sec^2. This would be equivalent to the force of gravity on Earth. Thus an artificial gravitational field would be incidentally created which would be exactly equivalent to that on Earth. Thus the device could also be used for the production of artificial gravity.

APPENDIX 1

Particle Physics

In the overall picture, it is generally accepted that there are three major forces; strong, electro-weak and gravity, each mediated by three bosons the, gluon, photon and graviton. These in turn are known to influence three types of particle, the quark, lepton and by General Relativity, space-time itself . Each of these appear to be composed of three particles with multiples of charge of 1/3 and exist in three dimensions of space time, which are themselves in three generations.

The nature of particles thus, may be revealed by their structure which occurs generally in multiples of three. Three quarks in the case of baryons are neccessary to make up a particle such as the proton or neutron. The particles that bind the quarks (gluons) are themselves required, in stable particles, to have three different colour charges, one colour in each dimension, for the particle to exist in three dimensional space-time. Furthermore, there are three generations of leptons an indeed quarks.

Nevertheless, the Standard model itself appears unable to explain the existence of these three generations of particles or the fundamental properties of sub atomic particles. Furthermore, there appear now to be a total of 60 fundamental particles each with their own fundamental properties which are arbitarily defined to fit the data. Furthermore, the Standard model only partially explains the decay process of the each particle and it does not explain their masses accurately.

For instance the mass of the up (u), anti-up (_) down (d) and anti-down (_) quarks are currently estimated at 0.35 Gev each (ref 17) (although some controversy exists about about this basic value). The mass of the proton constituted of three quarks, uud, is 0.9383 Gev, which is only approximately equivalent to the total mass of the three quarks (1.05 Gev).

Furthermore, the very process by which the subatomic particles decay cannot be explained by the "fundamental" constituents, the quarks. Indeed the known hadron decay processes ultimately always end up producing either an electron or positron. Indeed the quarks have not been experimentally seen, suggesting that other particles may underlay their fundamental structure.

Electron Structure

Given the presence of a quasi electron with a charge of 1/3 e (ref 4,5), then three of the quasi electron vortices, as previously described (ref 5), could form an entire electron in three dimensional visible space time. Nevertheless, each would have energy and hence a wave function which would be present in the other vectors. This electron could thus follow the probability functions as described by the Shrödinger wave equation for ψ (otherwise termed as "essence" by Shrödinger).

It is of importance therefore to note that the charge of the electron (e) in Coulombs (C), (where ε_0 is the permittivity of free space), can be derived from first principles by the equation:

$$e = \left[\frac{\varepsilon_0}{3\,(4/3\pi c^3)}\right]^{1/2} = 1.61 \times 10^{-19}\ \mathrm{C} \qquad (6a)$$

Equation (1) has a number of special implications, which have been previously discussed (p 6)

The dimensions of the equation can be readily resolved by considering **each of the three vector dimensions.** The exact dimensions of the equation need to be considered in the light of the nature of space-time itself, and this is fully addressed on pages 82-84. These dimensional equations help explain the nature of matter. Indeed the equation for the electron may underly the structure of the subatomic particles and may be necessary for the full understanding of gravity.

Additionally, the square root sphere structure of the electron accounts for its 1/2 integer spin. This square root structure also forms the basis of the electron pair bonding. According to the above equation (6), the electron will tend to form a "complete" electron sphere, thereby explaining how the presence of pair bonding occurs.

Thus from the equation (3) for the square root structure of the single electron; it is clear that the product of two such spheres will tend to form a complete sphere, where:

$$e^2 = \varepsilon_0/3(4/3\pi c^3) \qquad (6b)$$

In addition, the equation for the quasi electron can be directly derived from eq. 1.

Thus in one dimension, a single quasi electron (q_e) with charge of 1/3e, can be mathematically represented, (where the permittivitty of free space for each quasi electron is $\varepsilon_{qe} = [\varepsilon_0/3^3]^{1/2}$, by the equation :

$$q_e = \varepsilon_{qe}/(4/3\pi c^3)^{1/2} = 1/3e \qquad (6c)$$

Thereby accounting for the experimental observation of the 1/3e charge of the quasi electron(ref 4,5). Overall knowing the structure of the quasi electron may lead to the knowledge of the structures of the the other subatomic particles such as the quarks.

First Generation Quark Structure

From these observations we may now examine the derivation of the mathematical nature of the proton and in turn the structure of the quarks. Using the above equations (eq. 6) enables a far more accurate derivation of the mass of quarks from first principles.

To do so requires calculating the ratio of the mass of the electron ($m_e = 0.511$ Mev) to the mass of the proton ($m_p = 938.3$ Mev)

Thus $m_e/m_p = 0.511/938.3 = 5.44 \times 10^{-4}$

and $m_e/m_p = 5.44 \times 10^{-4} = (3\pi/c^{1/2})$

Intruigingly, the term $(3\pi/c^{1/2})$ mathematically determines the ratio of the mass of the electron (m_e) to the proton (m_p). Indeed it is this ratio that also leads to the mathemaical derivation of the structure of the quarks

Given the mass of the proton is; 938.3 Mev, then according to the Standard Model, as there are three quarks in the proton (uud) of virtually equal mass, then the effective mass of each quark would be more accurately given as 312.8 Mev.

Thence, the structure of the quarks can now be defined by the ratio of the masses of the quasi electron to the quark. If the mass of the quasi electron is given as one third the mass of the electron

$$m_{qe} = 0.511 \text{ Mev} \div 3 = 0.17033 \text{ Mev}$$

Then the ratio of the mass of the down quark (d), to the quasi electron is:

$$0.17033 \text{ Mev}/ 312.8 \text{ Mev} = 5.44 \times 10^{-4}$$

We also find that this same ratio of the mass of the quasi electron to the mass of the down quark (m_d) is mathematically equivalent to $3(\pi/c^{1/2})$, thus:

$$m_{qe}/m_d = 5.44 \times 10^{-4} = 3(\pi/c^{1/2})$$

From here the equation for the charge and structure of the down quark can be accurately defined by combining the equation for the structure of the quasi electron $q_e = \varepsilon_{qe}/(4/3\pi c^3)^{1/2}$ and the ratio of the masses of the quasi electron to the down quark (19); hence

$$d = \varepsilon_{qe}.m_d/m_{qe}.(4/3\pi c^3)^{1/2}.3(\pi/c^{1/2})$$

Hence in short form

$$d = (4/3\pi c^3)^{1/2}.3(\pi/c^{1/2}) \qquad (16)$$

Giving the charge of the down quark as -1/3; and its estimated mass as 312.8 Mev.

Thus not only can the electron charge be derived from the equation for three spheres each with a radius of c (equation 3) ; but the mass, charge and internal structure of the down quark can also be derived from the term $3(\pi/c^{1/2})$.

Furthermore, the significance of the term $3(\pi/c^{1/2})$ in this equation, is revealed by the Shrödinger wave equation. Thus $(\pi/c^{1/2})$ is none other than the solution to the Shrödinger wave equation for the amplitude of an electron confined in a space with radius c (ref 6). If c is the space defined by the equation for the structure of the quasi electron. Then the standard Shrödinger equation for an object confined in a space c is given by the following standard calculation (ref 8):

$$E\psi(x) = -\frac{h^2}{2m} . \frac{d^2\psi(x)}{dx^2}$$

If one of the dimensions of space has a length 2L the amplitude can only be non zero between $x = 0$ and $X = 2L$ and the standard solution for the amplitude can be determined by using the constant (A) where:

$$A = (1/L^{1/2})$$

Thus in one dimension the solution to the Shrödinger wave equation for the amplitude of oscillation is conventionally given by (ref 8) :

$$\psi(x) = (\pi/L^{1/2}) \quad \sin x/L$$

Substiuting L for c gives the one dimensional equation (when $\sin x/L = 1$) as $\pi/c^{1/2}$. Thus at 90°to the dimensional confinement of the quasi electron, the standard solution to the Shrödinger wave equation yields an amplitude of $\pi/c^{1/2}$, which in three dimensions gives:

$$3(\pi/c^{1/2}) \qquad (22)$$

Thus the term $3(\pi/c^{1/2})$ is not only the ratio of the mass of the electron to the proton but is also none other than the standard solution to the Shrödinger wave equation for an object confined in a space c $^{(\text{ref } 8)}$.

From here the charge and structure of the up quark (u) can now be derived in a similar way to that of the up quark from the mass of two quasi positrons [m_{qp} = 2(0.170 Mev)] and the mass of the up quark (m_u) by the equation,

$$2m_{qP}/m_U = 2(5.44 \times 10^{-4}) = 2(3\pi/c^{1/2})$$

Thus:

$$u = \varepsilon_{qe} .2m_{qP}/m_u.2^{+}(4/3\pi c^3)^{1/2}.(3\pi/c^{1/2})$$

Hence in short form

$$u = 2^{+}(4/3\pi c^3)^{1/2}.(3\pi/c^{1/2}) \qquad (17)$$

Giving the charge of the up quark as +2/3, and its mass as 312.8 Mev

Overall the mathematical structure of all the particles can be shown to be derived from the quasi electron. Thus, the first generation particle structures can now be derived from first principles. This includes the quasi electron (qe) and electron (e), from which the quarks (u,d) and in turn the proton (p) respectively are derived.

Thus the first generation of particles of the Standard model are given in short form as:

$$qe = {}^{-}(4/3\pi c^3)^{1/2} \qquad (6a)$$

$$e = 3^{-}(4/3\pi c^3)^{1/2} \qquad (6)$$

$$d = {}^{-}(4/3\pi c^3)^{1/2}.3(\pi/c^{1/2}) \qquad (16)$$

$$u = 2^{+}(4/3\pi c^3)^{1/2}.3(\pi/c^{1/2}) \qquad (27)$$

The respective force binding particle, the gluon, for the first generation of particles is thus given in short form as:

$$g = 3(\pi/c^{1/2}) \qquad (22)$$

According to these equations, it is apparent that the down and up quarks appear be constituted from quasi electrons and three component gluons, and are themselves not fundamental particles.

Indeed, from these observations it is clear that the structure of the electron, may underly the structure of all the elementary particles and nucleons, as it forms a "perfect sphere" based on c. As will also be seen later, as this structure forms the basis of matter, it may itself underpin the theory of gravity. Furthermore, these obsevations will allow the mathematical estimation of the mass and size of the elementary particles, including the second and third generation particles

Second and Third Genration Lepton Structure

Thus the structure of the muon (μ) can also be derived from the ratio of the mass of the electron (m_e) and the mass of the muon (m_μ):

$$m_e/m_\mu = 4.7 \times 10^{-3} = \pi/c^{1/3}$$

Thus $$\mu = \varepsilon_0^{1/2}.m_e/m_\mu. 3^{-}(4/3\pi c^3)^{1/2}.(\pi/c^{1/3})$$

Where the charge of the muon is in this equation equivalent to that of the electron e. In this case ($\pi/c^{1/3}$) can be considered to represent a specific high energy photon. Thus the structure of the muon, written in short form is:

$$\mu = 3^{-}(4/3\pi c^3)^{1/2}.(\pi/c^{1/3}) \qquad (14)$$

Moreover the structure of the tauon can be calculated from the ratio of the mass of the electron and that of the Tauon (1.79 Mev);

Thus $$0.511 \text{ Mev}/1.79 \text{ Gev} = 2.85 \times 10^{-4}$$

$$m_e/m_\tau = (\pi/c)^{1/3}.(\pi/c)^{1/9} = 2.85 \times 10^{-4}$$

As the charge of the tauon is equivalent to the charge of the electron, hence the structure of the Tauon is given by the above equation

$$e_\tau = \varepsilon_{qe}.m_e/\ m_\tau.3^{-}(4/3\pi c^3)^{1/2}.(\pi/c)^{1/3}.(\pi/c)^{1/9} = e$$

This equation accurately predicts the charge -1; and mass of the Tauon (~1.78 Gev) (ref 14). Thus the structure of the Tauon can in short form be given by the equation

$$\tau = {}^{-}3(4/3\pi c^3)^{1/2}.(\pi/c)^{1/3}.(\pi/c)^{1/9} \qquad (15)$$

Overall the mathematical structure of all the particles can be derived from the quasi electron. Thus, the particle structures can now be derived from first principles. This includes the quasi electron (qe) and electron (e), from which the quarks (u,d) and in turn the proton (p) respectively are derived. The general structure of the force carrying partilcles the photon (γ) and the gluon (g) can be given. It will also intriguingly be possible to derive the structure of the charm (c), beauty (b, or bottom) and truth (t or top) quarks directly from the sructure of the muon (μ) and Tauon (τ).

The calculations underlying these equations can further accurately reflect their measured values, particularly if the effects of gravity are taken into account (see Quantum Gravity and the Charge of the Electron.p 30). In addition the neutrino has not been included in the equation as its mass is considered to be very small. The mathematical proof of these short form equations nevertheless lies in the fact that they can closely identify the charge and the masses of these particles, from first principles, as in equations (6, 13-27). (ref 17)

Second and third generation Quark Structures

From here it is possible to derive the basic structure of the remaining quarks including the strange (s) quark.

Thus the ratio of the masses of the quasi electron (m_{qe} = 0.17033 Mev) and the strange quark (m_s = 0.555 Gev) are given by the equation:

$$m_{qe}/m_s = 0.17033\ \text{Mev}/\ 0.555\ \text{Gev} = 3.07 \times 10^{-4}$$

Furthermore the structure of the gluon in the strange particle, can now be accurately derived by the term; $3(\pi/c)^{1/2}$,(n.b. change of brackets in the gluon term).

$$m_{qe}/m_s = 3.07 \times 10^{-4} = 3(\pi/c)^{1/2}$$

Thus the equation for the charge and structure of the strange quark (s) is:

$$s = \varepsilon_{qe}.m_{qe}/m_s\ ^{-}(4/3\pi c^3)^{1/2}.3(\pi/c)^{1/2} \qquad (18)$$

This can be written in short form as:

$$s = {}^{-}(4/3\pi c^3)^{1/2}.3(\pi/c)^{1/2}$$

Giving the strange quark a charge of -1/3 and a mass of 0.555 Gev, in agreement with current estimates of its mass - 0.55 Gev (ref 14). This again shows that strange quark is derived from a quasi electron and a gluon.

Thus given that the particle Ω^- is made of three strange quarks (sss) we may derive its mass as 3 x 0.555 Gev = 1.67 Gev, which is in accurate aggreement with its known mass of Ω^- = 1.67 Gev (ref 17)

From here it is possible to define the structure of the other quarks, using the same first principles. As the charmed quark (c), beauty (b, otherwise termed bottom) and the truth (t, otherwise termed the top quark), clearly belong to the second and third generation of particles, they should be mathematically based on the formula for the second and third generation lepton particles, specifically the muon and tauon.

Indeed the charm derives from the second generation of particles and therefore mathematically does qappear to contain the basic muon structure, as given by the equation:

$$\mu = 3(4/3\pi c^3)^{1/2}.(\pi/c^{1/3}) \qquad (14)$$

As before, the ratio of the masses of the quasi electron and that of the charm quark gives the equation from the structure of the charm (c).Thus the structure of charm can be determined from the ratio of the mass of two quasi electrons; m_{qe} = 2(0.17033) Mev, and the mass of the charmed quark (m_c = 1.518 Gev)

Thus if: $2m_{qe}/m_c = 2(0.17033\ \text{Mev})/1.518\ \text{Gev} = 2(1.122 \times 10^{-4})$

and $2m_{qe}/m_c = 2(1.122 \times 10^{-4}) = 2(\pi/c^{1/3}).(\pi/c^{1/4})$

Then the structure of the charm quark (c) is:

$$c = \varepsilon_{qe}.2m_{qe}/m_c\ .2^{+}(4/3\pi c^3).^{1/2}(\pi/c^{1/3}).(\pi/c^{1/4})$$

or in short form

$$c = 2^{+}(4/3\pi c^{3})^{1/2}.(\pi/c^{1/3}).(\pi/c^{1/4}) \qquad (19)$$

Thus the charmed quark has a charge of +2/3 and a mass of 1.518 Gev in accordance with the previous approximate estimate of its mass ~ 1.5 Gev. (ref 17) and is derived from a muon and a higher order gluon. This gluon itself $(\pi/c^{1/4})$, is the solution to the Shrödinger wave equation for the amplitude of an electron confined in a space of length $c^{1/2}$ (ref 8). Thus this gluon now exactly matches the amplitude of the quasi electron.

Furthermore, from this we can deduce the mass of the particle the J/ψ, whose structure is predicted to consist of a charm and anticharm particle in orbit around each other, with a minimum mass (including orbital energy) of ~ 3.1 Gev.

From here we can in a similar fashion determine beauty (b). To be consistant, as beauty is a third generation quark it should be defined by the component structure of the Tauon. In this case, this is the smaller component of the Tauon, which is mathemtaically represented by:

$$3(4/3\pi c^{3})^{1/2}.(\pi/c^{1/9})$$

Indeed the structure can again be determined by the ratio of the mass of the quasi electron electron (m_{qe}) to beauty (m_b). Such that the ratio can be given by the equation

$$m_{qe}/m_b = 3.673 \times 10^{-4} = (\pi/c^{1/9}).(\pi/c)^{1/2}$$

Thus the structure of beauty is given by the equation

$$b = \varepsilon_{qe}.m_{qe}/m_b.(4/3\pi c^{3})^{1/2}.(\pi/c^{1/9}).(\pi/c)^{1/2}$$

To confirm this structure, the mass of the beauty quark can be given by the following calculation

$$m_{qe}/m_b = 3.673 \times 10^{-4} = 0.17033 \text{ Mev} / 4.64 \text{ Gev}$$

and

$$m_{qe}/m_b = 3.673 \times 10^{-4} = (\pi/c^{1/9}).(\pi/c)^{1/2}$$

Thus the short form of beauty is:

$$b = {}^{-}(4/3\pi c^3)^{1/2}.(\pi/c^{1/9}).(\pi/c)^{1/2} \qquad (20)$$

Thus giving beauty a charge of -1/3, and a mass of 4.64Gev, which is in agreement with its current estimated mass of ~ 4.7 Gev (ref 17).

To complete the structure of the quarks, we can finally proceed to deduce the truth. As the top quark or truth (t) belongs to the third generation of quarks, it is also based on the structure of the Tauon. Thus mathematically the truth, primarily consists of the third generation, tauon.

$$\tau = 3(4/3\pi c^3)^{1/2}.(\pi/c)^{1/3}.(\pi/c)^{1/9} \qquad (15)$$

The geometrical structure can be confirmed from the mathematical ratio of the mass of the electron to mass of truth (m_t), hence the equation for the top quark is given by the ratio of the mass of the electron to mass of the truth .

Thus as $m_t = 176$ Gev

and as $m_e = 0.511$ Mev

$$m_e/m_t = 0.511 \text{ Mev}/176 \text{ Gev} = 2.89 \times 10^{-6}$$

and

$$m_e/m_t = 2.89 \times 10^{-6} = (\pi/c)^{1/3}.(\pi/c)^{1/9}.(\pi/c)^{1/4}$$

As each quasi positron is associated with each of the gluon complexes, then the equation for truth is thus predicted by the equation:

$$t = 2(4/3\pi c^3)^{1/2+}.(\pi/c)^{1/3}.(\pi/c)^{1/9}.(\pi/c)^{1/4} \qquad (21)$$

Which gives the mass of the truth as 176 Gev and a charge of +2/3e. Which is in agreement with the known mass of the truth; 176 ± 12 Gev; the most accurate estimate of the mass of the truth quark being; 175.6 Gev.(ref 22)

Thus the basic structure of the truth, is that of a pair of quasi anti-tauons complexed with a gluon moiety which has an amplitude that, according to the Shrödinger wave equation (3), exactly matches the amplitude of the quasi electron (ref 8).

Mathematically the probability of arriving at all such tightly conformed equations (3, 16-30), the particles being related only to π and c and its specific powers, **by chance would have odds of millions to one against.!** Furthermore these mathematical structures are derived from none other than solutions to the Shrödinger wave equation itself (ref 8).

Moerover, the structures of the second and third generation particles, the muon (μ), and charm (c); tauon (τ) and truth (t) quarks have a **notable symmetry**, with their masses and charges accurately given by the following short form equations:

$$\mu = 3^{-}(4/3\pi c^{3})^{1/2}.(\pi/c^{1/3}) \tag{14}$$

$$c = 2^{+}(4/3\pi c^{3})^{1/2}.(\pi/c^{1/3}).(\pi/c^{1/4}) \tag{19}$$

$$\tau = 3^{-}(4/3\pi c^{3})^{1/2}.(\pi/c)^{1/3}.(\pi/c)^{1/9} \tag{15}$$

$$t = 2^{+}(4/3\pi c^{3})^{1/2}.(\pi/c)^{1/3}.(\pi/c)^{1/9}.(\pi/c)^{1/4} \tag{21}$$

Hence, from these equations it can be deduced that the fundamental structure of the quasi electron probablistically underpins the mathematical structure of the fundamental particles, and that these structures are all related to one another either by complexing with a photon $[(\pi/c)^{1/3}]$ which matches the frequency of the quasi electron or a gluon $[(\pi/c^{1/4})]$ which according to the Shrödinger wave equation accurately matches the amplitude of the quasi electron. This explains the fundamental observations that final decay products of hadrons always lead to the production of an electron or positron and explains why quarks have not been individually seen.

Particle Decay and the Electroweak Force.

In order to corroborate the estimated quark structures these structures need to explain in detail the pattern of decay of the quarks themselves. We are thus required to explain particle decay from first principles including the structures which are formed in these decays, such as the mediator of the electro weak force; the Intermediate Vector Boson. To see if this is possible the decay of the truth (t, top quark) will be examined. According to experiment the truth quark splits into two particles, beauty (b, bottom quark) and the intermediate vector boson (W^{+}). It is apparent that the truth may yield beauty.(ref 22)

Thus:

$$2^{+}(4/3\pi c^{3})^{1/2}.(\pi/c)^{1/3}.(\pi/c)^{1/9}.(\pi/c)^{1/4}=t$$

$$\Rightarrow \quad (4/3\pi c^{3})^{1/2}.(\pi/c^{1/9}).(\pi/c)^{1/2} \quad = b$$

Thus the two gluon structures, $(\pi/c)^{1/4}$; in the truth combine to form the more energetic gluon, $(\pi/c)^{1/2}$; present in beauty, thus:

$$2(\pi/c)^{1/4}\Rightarrow(\pi/c)^{1/4}.(\pi/c)^{1/4}\_\;(\pi/c)^{1/2}$$

and the photon term, $(\pi/c)^{1/9}$; decays into the less energetic photon $(\pi/c^{1/9})$ present in beauty. Thus the structures in the truth account for the structures in beauty, and are produecd via the production of a quasi electron and quasi positron.

The remaining terms are thus 3 quasi positrons, two high energy photons and a remaining low energy photon

$$3^{+}(4/3\pi c^{3})^{1/2}$$

and

$$2(\pi/c)^{1/3}$$

and

$$(\pi/c)^{1/9}$$

There is also a very considerable amount of energy remaining, the energy difference between beauty (b) and truth (t) being ~ 170 Gev.

This energy difference allows the transformation of the two high energy photons and the low energy photon to two intermediate vector photons, given by the equation.

$$2(\pi/c^{1/6})^{6}$$

This leads to the formation of the intermediate vector boson (W^{+}), which is the mediator of the electroweak nuclear force.

$$W^{+}=3^{+}(4/3\pi c^{3})^{1/2}.2(\pi/c^{1/6})^{6} \qquad (24)$$

which is also equivalent to

$$W^+ = 3^+ (4/3\pi c^3)^{1/2}.2(\pi/c^{1/6})^3.(\pi/c^{1/6})^3$$

This probable structure for the (W^+), can again be confirmed by the ratio of the mass of the electron to the intermediate vector boson (W^+), with an estimated mass of 80 Gev $^{(\text{ref } 17)}$. Thus the ratio of the masses is given by

$$m_e/m_{W+} = 0.511 \text{ Mev}/ 80 \text{ Gev} = 6.4 \times 10^{-6}$$

and

$$m_e/m_{W+} = 6.4 \times 10^{-6} = 2(\pi/c^{1/6})^6$$

Thus

$$W^+ = \varepsilon_{qe}\text{-}m_e/m_{W+}.3^+(4/3\pi c^3)^{1/2}.2(\pi/c^{1/6})^6$$

thus in short form:

$$W^+ = 3^+ (4/3\pi c^3)^{1/2}.2(\pi/c^{1/6})^6$$

The true elegance of the structure of the intermediate vector boson can now be further explored. The W^+ can be considered to represent a unification of the electromagnetic force and the weak force, in keeping with the theoretical predictions of Salam and Weinberg for the nature of the electroweak force. Thus, the intermediate vector photons can either decay leptonically or via gluons; as is demonstrated by the following interconversions.

$W^+ = 3^+ (4/3\pi c^3)^{1/2}.2(\pi/c^{1/6})^6$ _ $3^+ (4/3\pi c^3)^{1/2}.2(\pi/c^{1/6})^3.(\pi/c^{1/6})^3$

If we examine the further decay of the three dimensional photon structure, $(\pi/c^{1/6})^3$; we have:

$(\pi/c^{1/6})^3 \Rightarrow [(\pi/c)^{1/6}]^3$ _ $(\pi/c)^{1/2}$

Where the term $(\pi/c)^{1/2}$ is none other than the term for a first generation gluon.

Thus the W^+ can either decay leptonically or into first generation quarks where it can yield an up (u) and an anti-down quark (_), thus:

$$W^+ = 3^+ (4/3\pi c^3)^{1/2}.2(\pi/c^{1/6})^3.(\pi/c^{1/6})^3 \Rightarrow 3^+(4/3\pi c^3)^{1/2}.2[(\pi/c)^{1/6}]^3.[(\pi/c)^{1/6}]^3$$

And

$3^+(4/3\pi c^3)^{1/2}.2[(\pi/c)^{1/6}]^3.[(\pi/c)^{1/6}]^3$ _ $3^+(4/3\pi c^3)^{1/2}.2(\pi/c)^{1/2}.(\pi/c)^{1/2}$

thus W^+ decays to:

$$\Rightarrow \quad 2^{+}(4/3\pi c^{3})^{1/2}.3(\pi/c^{1/2}) = u$$

and

$$\Rightarrow \quad {}^{+}(4/3\pi c^{3})^{1/2}.3(\pi/c^{1/2}) = \_d$$

Leptonically the W^{+} decay yields the standard products including the positron the neutrino (v), and energy in the form of photons thus

$$W^{+} = 3^{+}(4/3\pi c^{3})^{1/2}.2(\pi/c^{1/6})^{6} \_ 3^{+}(4/3\pi c^{3})^{1/2}.2(\pi/c^{1/6})^{2}.(\pi/c^{1/6})^{2}.(\pi/c^{1/6})^{2}$$

and

$$\Rightarrow \quad 3^{+}(4/3\pi c^{3})^{1/2}.2[(\pi/c)^{1/6}]^{2}.[(\pi/c)^{1/6}]^{2}.[(\pi/c)^{1/6}]^{2}$$

$$\_ \quad 3^{+}(4/3\pi c^{3})^{1/2}.2(\pi/c)^{1/3}.(\pi/c)^{1/3}.(\pi/c)^{1/3}$$

Thus W^{+} may also decay into the positron the neutrino (v), and energy in the form of photons thus:

$$\Rightarrow \quad 3^{+}(4/3\pi c^{3})^{1/2} = e^{+}$$

and

$$\Rightarrow \quad v$$

and

$$n(\pi/c)^{1/3}$$

Thus according to standard experimental observations the truth quark yields the beauty quark and the intermediate vector boson. In turn from these observations we may elegantly reveal the structure of truth and beauty and the intermediate vector boson and its decay pathways from first mathematical principles. (ref 18)

Gluon Structure and Force Characteristics

There are in the Standard model (or in this modification) three colour charges for the gluons and quarks, each of these represents each of three vectors x, y and z which can arise in different combinations. Thus the three gluon colours are required to be together (one in each dimension x, y and z) in baryons to form a three dimensional objects. There are also noted to be 8 gluons, these in turn gives rise to all the known baryons and mesons

The structure of these gluons have been previously mathematically identified, and this structure mathematically represents a helical ringlet. These are as previously given: $(\pi/c)^{1/2}$; $(\pi/c)^{1/4}$; $(\pi/c^{1/2})$; $(\pi/c^{1/4})$ [see eq. 25- 25c]

These four mathematical structures and their antigluons now readily account for the eight gluons known to exist, these in turn account for the structure of all the quarks as given by equations 3, 15-30, which in turn give rise to all the baryons and mesons as described by the Standard model.

Recent studies show that the proton conventionally consisting of the "fundamental" quarks (uud); can by bombardment with high energy electrons produce a Kaon which consists of an anti-up and a strange quark (s_). As the standard model predicts quarks are fundamental, the proton should not contain a strange quark. The conventional explanation for this resides in the proposal that the proton contains virtual strange particles. Indeed to explain these new observations each proton is now postulated by some physicists to contain 21 or more quarks, 3 of which are real and 18 or more which are required to be virtual! (ref 23).

In the revised Standard model presented here, the process by which a Kaon is produced is readily understood by the structural interconversion of the gluon in the up quark $(\pi/c^{1/2})$; to the gluon in the strange quark $(\pi/c)^{1/2}$,(n.b change of brackets)

Furthermore the most recent experiments on the internal structure of the proton has shown that gluons can seperate and recombine. This surprising recent observation actually arises from first principles when the mathemtatical structure of the gluon is understood. Thus mathematically the gluon structures may easily recombine and separate by the equation.(ref 24)

$$(\pi/c)^{1/2} = (\pi/c)^{1/4}.(\pi/c)^{1/4}$$

Hence this mathematical structure of the gluon allows the observed "recombination" and separation to occur in a way which is readily understood.

In addition, the structure of the gluons can be explained on the basis that these structures match the wave function of the quasi electron, thus the gluons match the amplitude of the electron as given by the Shrödinger wave equation.(ref 8)

Furthermore, the equations for the gluon present in the quark can explain their force characteristics. In the case of the gluon component $(\pi/c^{1/2})$; which mathematically represents a helical ringlet; if the direction vector is x, then the axis of spin would need be 90° to this, in the y vector in order to match the amplitude of vibration of the quasi elelctron/ positron. Thus this spin vector is known to be the same as that of electrons. This would account for the particle binding characteristics of the gluon force to the quasi electron. Given that the gluon would mathematically require the same spin axis as that of the quasi electron or positron, the gluon would be strongly associated with these particles and the force required to part them would increase with distance, which explains the characteristics seen experimentally with the gluon force.

Moreover, as will be shown, the photonic component of the subatomic particles given in the above equations matches exactly the frequency of rotation of the electron.

Nucleon Structure

The mass (m_p) and the internal mathematical structure of the proton can now be directly calculated form the quarks. Thus the structure of the proton according to the Standard model derives the combination of 2 up quarks and one down quark (p = uud). As each quark has an estimated mass of 312.8 Mev (see eq. 13) then the proton (p) has in accordance with the Standard model a mass of 3 x 312.8 = 938.3 Mev and a charge of +1, and may be given by the mathematical structure of the up and down quarks according to eq. 16,17; and is thus given by:

$$p = \begin{array}{c} 2^{+}(4/3\pi c^{3})^{1/2}.(3\pi/c^{1/2}) \\ {}^{-}(4/3\pi c^{3})^{1/2}.(3\pi/c^{1/2}) \\ 2^{+}(4/3\pi c^{3})^{1/2}.(3\pi/c^{1/2}) \end{array}$$

Which, gives the proton a net charge +1 and a mass of 938 Mev.

Fifteen component gluons are required to be present in the proton, three gluons associated with each quasi electron particle.

Therefore the structure of the proton can be mathematically derived by combining the structure of the quasi electrons and the term $(\pi/c^{1/2})$. Thus the structure of the neutron and indeed the overall structure of the nuclei can be determined.

Similarly the structure of the neutron is given by the Standard model (n = udd), which can be written in short form as:

$$n = \begin{array}{c} {}^{-}(4/3\pi c^{3})^{1/2}.3(\pi/c^{1/2}) \\ 2^{+}(4/3\pi c^{3})^{1/2}.3(\pi/c^{1/2}) \\ {}^{-}(4/3\pi c^{3})^{1/2}.3(\pi/c^{1/2}) \end{array}$$

The mystery of the stability of the neutron may now be addressed. A neutron outside the nucleus has a half life of approximately 6 seconds. Inside the nucleus however, the stability of the neutron, provides the atom with its stability. The three dimensional structures described may now provide the clues to this stability.

The compostion of the quarks now logically explains the apparent "fluid" characteristics of the subatomic particles, and some free exchange of the quasi electrons and gluons would be expected to occur within a nucleus.

This would occur to to give the structures stability and this in turn would allow the decresed binding energies seen in these structures. As has been shown by the metric tensor structure of matter (eq. 4-6), the preferred conformation is a 3 x 3 structure.

This can be acheived by donating a quasi electron and a quasi positron along with three associated gluons each; from the proton structure to that of the neutron. The proton would then contain three positrons and a total of nine gluons

The stable nuclear proton conformation would now be :

$$p = 3^{+}(4/3\pi c^{3})^{1/2}.3(\pi/c^{1/2}) \qquad (13)$$

This can form a stable 3 x 3 conformation.

Thus the probable structure of the stable nucleonic neutron (n), would now contain a total of three quasi positrons and three quasi electrons along with their respective eighteen gluons and is represented by an equation, which also forms a 3 x 3 conformation:

$$n = \begin{matrix} {}^{+}(4/3\pi c^{3})\downarrow^{1/2}.2(3\pi/c^{1/2}).{}^{-}(4/3\pi c^{3})\uparrow^{1/2} \\ {}^{+}(4/3\pi c^{3})\downarrow^{1/2}.2(3\pi/c^{1/2}).{}^{-}(4/3\pi c^{3})\uparrow^{1/2} \\ {}^{+}(4/3\pi c^{3})\downarrow^{1/2}.2(3\pi/c^{1/2}).{}^{-}(4/3\pi c^{3})\uparrow^{1/2} \end{matrix} \qquad (26)$$

Viewed in three dimensions the configuration will allow the positron components of the neutron to be placed interiorly and the electron components exteriorly, in keeping with experiments that suggest that the neutron core is positive an the exterior is negative. This polarised structure in turn will enable the neutron to bind the protons of the nucleus together.

Alpha Particle Structure

Importantly it is now possible to **derive the mathematical structure of the helium nucleus** (α particle) containing two protons and two neutrons, along the same lines. The proposed basic conformation is again 3 x 3.

$$\alpha = \begin{array}{l} [3(4/3\pi c^3)]\downarrow^{1/2+}.2(3\pi/c^{1/2}).\ [3(4/3\pi c^3)]\uparrow^{1/2-} \\ [3(4/3\pi c^3)]\uparrow^{1/2+}.2(3\pi/c^{1/2}).\ [3(4/3\pi c^3)]\downarrow^{1/2+} \\ [3(4/3\pi c^3)]\downarrow^{1/2-}.2(3\pi/c^{1/2}).\ [3(4/3\pi c^3)]\uparrow^{1/2+} \end{array} \quad (27)$$

This mathematical expression for the structure of the α particle allows the **calculation of its binding energy from first principles!** Each pair of particles is still associated with the equivalent of 18 gluons each, that is $3^{1/2}.2(3\pi/c^{1/2}).3^{1/2}$. However, the internal structure has altered, to effectivly reduce the number of quasi electrons. Thus as a result of triplet pairing, which has occurred in order to make up the helium nucleus, we are left with a reduced number quasi electrons.

Hence, the number of quasi electrons is reduced by a factor of 3 as a result each of the three pairings of the six quasi electron groups. This means the reduction in the equivalent number of quasi electron masses (including the electrical kinetic energy of the complex vectors) is effectively 3^3 x 6. Each of these weighs 1/3 the rest mass of the electron. The total energy therefore liberated is equivalent to 27 x 6 quasi electron masses (approximately 27.5 Mev). This agrees with the difference in mass between the constituent protons and neutrons and the rest mass of the helium nucleus, thus accounting for the binding energy of the helium nucleus.

This structure is also important for the understanding of the internal structure of atomic nucleii, The alpha particle appears to form a natural sphere as indicated by the metric tensor structure (see eq. 6-9). In turn the basic structure of atomic nucleii appear to form in multiples of alpha particles; the carbon atom for instance forms from three, separate, such helium nuclei spheres.

Furthermore knowledge, of the structure of the α particle may be of considerable importance in the understanding **of nuclear fusion** and the eventual harnessing of this technology.

Neutrino Structure

Whilst it has been possible to derive the structure of the fundamental particles form their known masses, the mass of the neutrino remains uncertain. However, recent evidence does support the contention that the neutrino does indeed have rest mass, which is a starting point.

If the neutrino does have rest mass this suggests it has a basic spherical structure like the electron, and will therefore be subject to relativistic changes in mass. Again experiment confirms this relativistic component.

It is also reasonably certain that whilst the neutrino mass is smaller than that of the electron, its size is also likely to be smaller, and on the basis of what is known of the radius of the electron (see Particle Spin and Size) the essential radius is also likely to be related to c.

Indeed as the electron radius is 1/c x 2/spin, where the intrinsic radius is 1/c. Then by deduction the neutrino radius is related to $1/c^2$. Thus the most likely equation for the neutrino is:

$$e\nu = \left[\frac{\varepsilon_0}{3\left[4/3\pi(c^2)^3\right]}\right]^{1/2} = \left[\frac{\varepsilon_0}{3\left[4/3\pi c^6\right)}\right]^{1/2} \qquad (6d)$$

This of course endows the neutrino with a very small charge of: 9.71×10^{-32} C. (which reduces to 9.42×10^{-63} C, when the neutrino forms a Cooper neutrino pair) Nevertheless, if the neutrino has mass, like all other elementary particles we should not be surprised if it also has a charge. Furthermore we may use the above formula to estimate the neutrino mass itself.

Without concerning ourselves too much at this point about the exact origin of the nature of the mass of the electron (which is addressed in chap. Quintessence and Electron Structure). It is reasonable to postulate that, if the radius is related to $1/c^2$ then the number of quintessences in the neutrino will be equivalent to the square root of the number of quintessences in the electron. Indeed this means that the number of quintessences in the electron neutrino is 1.925×10^{10} quintessences. Knowing the mass of quintessence; 2.4575×10^{-51} kg, gives us the mass of the electron neutrino (with an estimated error range of: x/÷ 3), as:

$$m_{ve} = 4.731 \times 10^{-41} \text{ kg} = 0.0000265 \text{ ev}$$

We may also deduce the masses of the mu neutrino and the tau neutrino from first principles. Interestingly we also know that the μυ and the ταυ neutrino can convert an electron to a muon or tauon respectively (if the neutrino has sufficient relativistic energy). So it is likely that the geometric elements present in the muon and tauon are also present in their respective neutrinos. Thus the short inverted form equations are given by:

$$\nu_\mu = 3^{-}(4/3\pi c^6)^{1/2}/(\pi/c^{1/3}) \qquad (6d)$$

$$\nu_\tau = 3^{-}(4/3\pi c^6)^{1/2}/(\pi/c)^{1/3}.(\pi/c)^{1/9} \qquad (6d)$$

These mathematical suffixes are merely the same geometric components, which combine with the quasi electron to form the muon and tau. In the same way the ratio of the masses of the muon and tau to the electron, were deduced from first principles (see Fundamental Forces and Particle Structure) these equations predict the respective masses of these neutrinos (with an error range of x/÷ 3), as:

$$\nu_\mu = 0.00546 \text{ ev} \quad \text{and} \quad \nu_\tau = 0.0927 \text{ ev}$$

Indeed recent evidence from the super Kami0kande experiment supports these observations with regard the tau neutrino, and places its mass in the range 0.1- 0.03 ev.

Quintessence and Internal Electron Structure

The reason for the value for the mass of the electron, has till now remained a mystery. The internal structure of the electron is therefore of considerable importance as it may elucidate the underlying reason for the mass of the electron. In turn we may derive the mass of all the particles from first principles, as these themselves derive from the structure of the electron (see previous chapter; Fundamental Forces and Particle Structure). Additionally, it may also be possible to estimate the radius of quintessence, from the structure of the electron. Quintessence in turn is fundamental to the formulation of quantum gravity.

We may begin with the formula for the charge of the electron

$$e = \varepsilon_{qe}/3(4/3\pi c^3)^{1/2} \qquad (6)$$

This demonstrates a square root spherical structure. From here we may deduce the most logical internal structure of the quasi electron as derived from quintessence. Indeed if the quasi electron has a square root spherical structure, this needs to be accounted for when calculating the number of quintessence it contains to account for the square root sphere conformation of the electron.

The structure can be defined using the equation for the surface area of a sphere: $A=4\pi r^2$. Thus the total volume of a single outer layer of quintessence in the quasi electron will be the surface area of the quasi electron x the average diameter of quintessence [if quintessence is oscillating the average diameter is $(2r_q + 0)/2 = r_q$].

Taking r_e as the radius of the electron and r_q as the radius of quintessence, given that there are three quasi electrons in the total electron, the volume of a single outer layer of quintessence (V) is:

$$V = (4\pi r_e^2)^{1/2} x \ (r_q/3)^{1/2}$$

Taking the volume of a single quintessence (V_q); based on a square root electron sphere, then:

$$V_q = (4/3\pi r_q^3)^{1/2}$$

The total number of quintessences in a single outer layer will thus be the total volume divided by the volume of a single quintessence itself: V/V_q. Thus the number of quintessences (n_q) in a single outer layer is:

$$n_q = (4\pi r_{qe}^2)^{1/2} x \ (r_q/3(4/3\pi r_q^3)^{1/2}$$

This very elegantly reduces to

$$n_q = r_e/r_q$$

To be further precise (after all quantum dynamics has to be elegant), the total number of quintessences in the electron will thus be determined by the number of quintessences in a single outer electron layer x the quantum amplitude. The quantum amplitude has been previously calculated using the standard Shrödinger wave equation (see Fundamental Forces and Particle Physics) and is given by $(c^{1/2}/\pi)$. Taking the root mean square of the quantum amplitude, the number of quintessences (n_q) in the electron is:

$$n_q = r_e/r_q \ x \ \sqrt{2}(c^{1/2}/\pi).$$

As the radius of the electron is estimated as, $r_e = 6.4 \times 10^{-19}$ m. (see Particle spin and Size), and the radius of quintessence is, $r_q = 1.35 \times 10^{-35}$ m, this allows us to estimate the number of quintessences in the electron itself using the above equation, thus:

$$\mathbf{n_q = 1.2355 \ x \ 10^{20} \ m}$$

Thus, given that m_q = 7.373 x 10^{-51} kg, the mass of the electron is given by

$$m_e = m_q \ x \ n_q, = 9.10938 \ x \ 10^{-31} \ kg.$$

Using the radius for quintessence (the estimated Planck length) then the number of quintessences present in the electron and thus its mass, is explained.

Therefore, these observations postulate that the internal structure of the electron consists of a square root sphere, composed of a single outer layer of quintessence x the root mean square of the its quantum amplitude.

The electron is itself composed of quintessence with a velocity of c and has an intrinsic radius of 1/c x 1/spin in metres; and in accordance with its half integer spin $h/4\pi$; its spin velocity is $2\pi/c$ in metres/sec, (see Particle Spin and Size) which in vortex mechanics gives it the ideal harmonic balance with quintessence space time.

Thus given the surface area of the quasi electron in order to fill the outer layer allowing for the quantum amplitude of oscillation would require 1.235×10^{20} quintessences. The total number of quintessence required to fill the whole electron would thus be $n_q = 1.235 \times 10^{20}$. Overall the number of quintessences required to make up the electron can be verified from the internal structure of the electron and the radius of quintessence. Thus if r_q is equivalent to the given Planck length, then accordingly the number of quintessences in the electron is calculated as 1.235×10^{20}, which gives the mass of the electron $m_e = m_q \times n_q$, hence again:

$$m_e = m_q \times n_q, = 9.10938 \times 10^{-31} \text{ kg}.$$

Intriguingly the derivation of the structure of the electron reflects the volume of an ordinary sphere

$$V = 4\pi r^2 \times r/3 = 4/3\pi r^3$$

This underpins a fundamental concept, mass determines the shape of space-time and in turn the shape determines the amount of mass within it.

It is important to have a working understanding of why the mass of the electron should be what it is from its basic structure. This understanding can now be derived from the radius of quintessence, which also corresponds with the Planck length.

Dimensional Equations for Quintessence

As $h = 3m_qc^2$ and the dimensions of h are $[ML^2T^{-2}.T]$ and those of c^2 are $[L^2.T^{-2}]$ the equational dimensions of quintessence are [M][T], and the dimensions relating to the number of quintessences n_q is $[T^{-1}]$. Clearly therefore,

$$[M] = [M][T] \times [T^{-1}]$$

and overall

$$M = m_q \times n_q.$$

Dimensional Equations for the Electron

The dimensions of the equation for the electron, in the light of quintessence, can now be fully examined, if

$$e^2 = \frac{\varepsilon o}{3\ (4/3\pi c^3)} \qquad (6b)$$

Then taking the dimensions of the equation, e^2 is the charge $[q^2]$, in Coulombs; ε_0 is the permittivity of free space, in Capacitance [C] per metre [L]; and [c] is the speed of light, in [L/T].

Thus the above equation using dimensions, is given by

$$[q^2] = \frac{[C]}{[L][c^3]}$$

In this case the velocity is given as the velocity of light c, then L = c.T ; then

$$[q^2] = \frac{[C]}{[c^4][T]}$$

As $q^2 = C^2V^2$, where V is volts, then:

$$[C^2][V^2] = \frac{[C]}{[c^4][T]}$$

and

$$[CV^2] = \frac{1}{[c^4][T]}$$

as $E = CV^2$, then

$$E = \frac{1}{[c^4][T]}$$

and as $E = mc^2$, then

$$mc^2 = \frac{1}{[c^4][T]}$$

and hence

$$m = \frac{1}{[c^6][T]}$$

The significance of this dimensional analysis, in the first instance, appears obscure. Nevertheless, it reveals the very nature of matter and energy. Thus the equation for the mass of a structure can be represented by **$1/c^6$ which represents the six complex vectors of matter.**

Interestingly, given it is known that $1/c^2 = [\mu_o.\varepsilon_0]$, then

$$m = [\mu_o.\varepsilon_0]^3/T$$

Where μ_o. and ε_0 are again the permeability and permittivity of free space, quantities that are inherently caused by the vibration of quintessence. Thus mass itself is the result of the vibration of quintessence in the six complex vectors (each represented by the fundamental properties of quintessence μ_o or ε_0 .Hence, the equation for the quasi electron mathematically and geometrically forms the "perfect" three dimensional sphere with its mass accounted for by its six complex vectors.

Energy associated with this matter is in turn is also caused by the the vibration of quintessence (including that in the complex vectors, i) such that:

$$E = [\mu_o.\varepsilon_0]^2/T$$

The dimensions for the equation for energy can also be written as:

$$E = \frac{1}{c^3.L}$$

Which indicates that the energy of matter is again related to the dimensions of the structure of the electron $\varepsilon_0 /(4/3\pi c^3)$ and the permittivity of free space in capacitance per meter (C/L)

Overall the equation for the quasi electron and its dimensions, gives us the origin of mass and energy. The equation for the electron then forms the key to the understanding of the nature of the relationship between matter and space-time.

Indeed we may now even derive Newton's second law of motion (F = ma, and thus the other fundamental laws of physics) from first principles to confirm the validity of the above.

If:

$$E = \frac{1}{[c^4.][T]}$$

Which is dimensionally equivalent to:

$$E = \frac{[T^3]}{[L^4]}$$

then as F = E/L:

$$F = \frac{[T^3]}{[L^5]}$$

and thus

$$F = \frac{[T^6]}{[L^6.T]} \cdot \frac{[L]}{[T^2]}$$

Substituting for m from eq. 1g

$$m = \frac{[T^6]}{[L^6][T]}$$

Then $F = ma$

Thus taking into account the whole of these observations enables the laws of motion and the equations for energy and mass and their equivalence to be derived from geometric first principles.

REFERENCES

[0433] 1. Mesoscopic behaviour of the neutral Fermi gas ^{3}He confined in quantum wires. S. L. Phillipson, A. M. Genault, S. N Fisher, G. R. Pickett, P. J. Y Thibault. Nature, 395: 578-580.

[0434] 2. Direct Observation of d-orbital Holes and Cu-Cu Bonding in Cu_2. J. M. Zuo et al. Nature 401: 49-52.

[0435] 3. The Foundation of the General Theory of Relativity, A Einstein. Translated from "Die Grundlage der Allgemeinen Relativitätstheorie". Annalen der Physik, 49, 1916. Published in, Einstein, The Principle of Relativity, Dover publications, 1952.

[0436] 4. Sidelights on Relativity. Ether and the Theory of Relativity Albert Einstein. Dover Publications 1983.

[0437] 5. Inflation is dead long live inflation. George Musser. Scientific American, July 1998, p 9.

[0438] 6. The flip side of the Universe. George Musser. Scientific American, September 1998, p 13.

[0439] 7. The Fifth Element, Marcus Chown. New Scientist, Apr. 3, 1999, pp 29-32

[0440] 8. The Quantum Universe. Tony Hey and Patrick Walters. Cambridge University Press 1987. (Solving the Shrödinger equation pp. 165-167).

[0441] 9. Complex Numbers, Mathematics for Engineers. W. Bolton; Longman Scientific & Technical; Chap. Introducing Complex Numbers, p 1-11.

[0442] 10. Unconditional Quantum Teleportation. A. Furusawa, I. Sorensen, S. L. Braunstein, C. A. Fuchs, H. J Kimble, E. S. Polzik. Science, Oct. 23, 1998. Vol. 282: 706-709.

[0443] 11. Complete quantum teleportation using nuclear magnetic resonance. M. A. Neilsen, E. Knill, R. Laflamme. Nature, Nov. 5, 1998. Vol. 396: 52-55.

[0444] 12. Experimental quantum teleportation. D. Bouwmeester et al. Nature, 390: 575-579 (1997).

REFERENCES (cont.)

[0445] 13. Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein-Podolski-Rosen channels. D. Boschi, S. Branca, F. De Martini, L. Hardy, S. Popescu. Phys. Rev. Lett. 80: 1121-1125 (1998).

[0446] 14. W. Tittel, J. Brendel, H. Zhuniden, N. Gisin, Phys. Rev. Lett. 81: 3563-3566. (1998)

[0447] 15. Into the Vortex. Charles Seife. New Scientist, Oct. 24, 1998. p 7.

[0448] 16. Direct observation of a fractional charge. R de Picciotto, M Reznikov, M. Heibium, V Umansky, G. Bunin, D. Mahalu. Nature, 389: 162-164.

[0449] 17. Tables of Physical and Chemical Constants. Kay & Laby. Longman Scientific & Technical.

[0450] 18. Physics in Focus. Michael Brimicombe. Section, Wave Mechanics; p 521. Nelson

[0451] 19. Special Relativity. A. P. French. Nelson. Chapter 8, Special Relativity and Electricty, pp 256.

[0452] 20. Universe In Balance. Jeff Peterson. New Scientist, 168: p 27

[0453] 20*a*. The Accelerating Universe. Mario Livio. John Wiley and Sons, Inc. (2000)

[0454] 21. Escape from Earth, Charles Seife, New Scientist, Feb. 6, 1999. p 6.

[0455] 22. The Discovery of the Top Quark. Tony Liss and Paul Tipton. Scientific American; September 1997, pp. 26-41.

[0456] 23. The Mystery of the Nucleon Spin. Klaus Rith and Andreas Schaifer. Scientific American: July 1999, pp 42-47.

[0457] 24. And you're glue. Frank Wilczek. Nature; 400, pp 21-23.

1. A method of generating a force on a body, wherein a differential electron flow through the body in rotation is directed so as to simultaneously pass through said body in its direction of rotation and contrary to its direction of rotation to release a directed flow of gravitons.

2. A method of accelerating a body by generating a force acting upon it by the method of claim 1.

3. A method according to claim 1 or **2**, wherein electrical currents are passed simultaneously through said body in its direction of rotation and contrary to its direction of rotation.

4. A method according to any one of claims 1 to 3, wherein the body is a least partly of superconducting material.

5. A device for generating a force sufficient to accelerate a body, the device comprising;

the body in the form of an electrically conducting mass, means for rotating said mass and means for passing electrical currents simultaneously through said mass in its direction of rotation and contrary to its direction of rotation.

6. A device according to claim 5, wherein the mass is spherical, cylindrical or consists of a dual co-rotating or counter rotating disc.

7. A device according to claim 5, wherein the mass is cylindrical.

8. A device according to claim 5, wherein the mass consists of a disc.

9. A device according to claim 5, wherein the mass consists of dual co-rotating or counter rotating discs.

* * * * *

E. H. ARMSTRONG.
WIRELESS RECEIVING SYSTEM.
APPLICATION FILED OCT. 29, 1913.

1,113,149.

Patented Oct. 6, 1914.
3 SHEETS—SHEET 1.

Fig. 1,

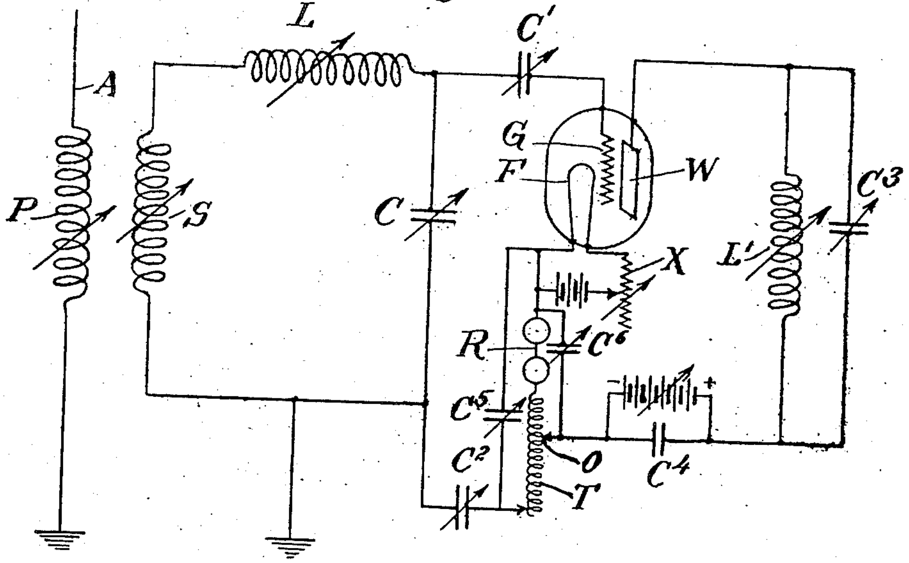

L
C′
A
G
F
W
C3
P
S
C
X
L′
R
C6
C5
C2
O
C4
T

Fig. 2,

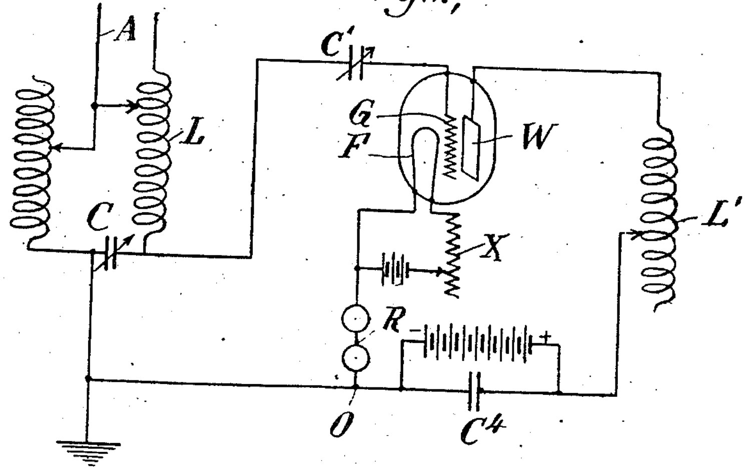

A
C′
G
F
W
L
C
X
L′
R
O
C4

WITNESSES
L Bates
C. B. Schroeder

Edwin H. Armstrong
INVENTOR
BY
Pennie, Davis & Goldsborough
ATTORNEYS

E. H. ARMSTRONG.
WIRELESS RECEIVING SYSTEM.
APPLICATION FILED OCT. 29, 1913.

1,113,149.

Patented Oct. 6, 1914.
3 SHEETS—SHEET 2.

Fig. 3,

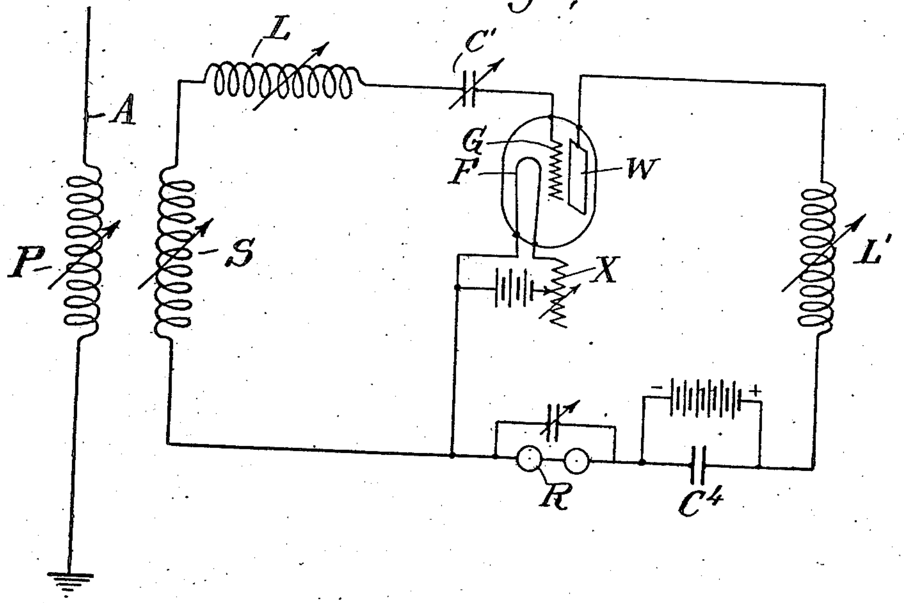
L
C′
A
G
F
W
P
S
X
L′
−
+
R
C4

Fig. 4,

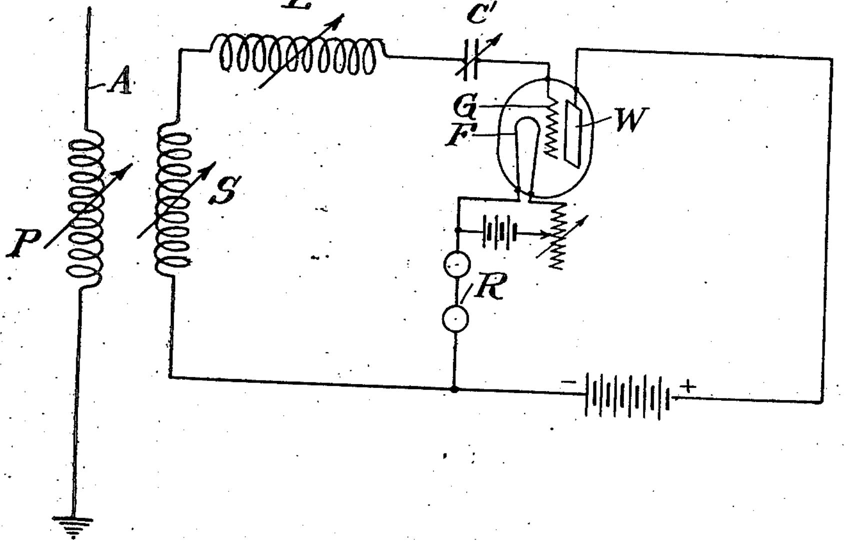

WITNESSES
L. Bates
C. B. Schroeder

INVENTOR
Edwin H. Armstrong
BY
Pennie, Davis & Goldsborough
ATTORNEYS

E. H. ARMSTRONG.

WIRELESS RECEIVING SYSTEM.

APPLICATION FILED OCT. 29, 1913.

1,113,149.

Patented Oct. 6, 1914.

3 SHEETS—SHEET 3.

Fig. 5,

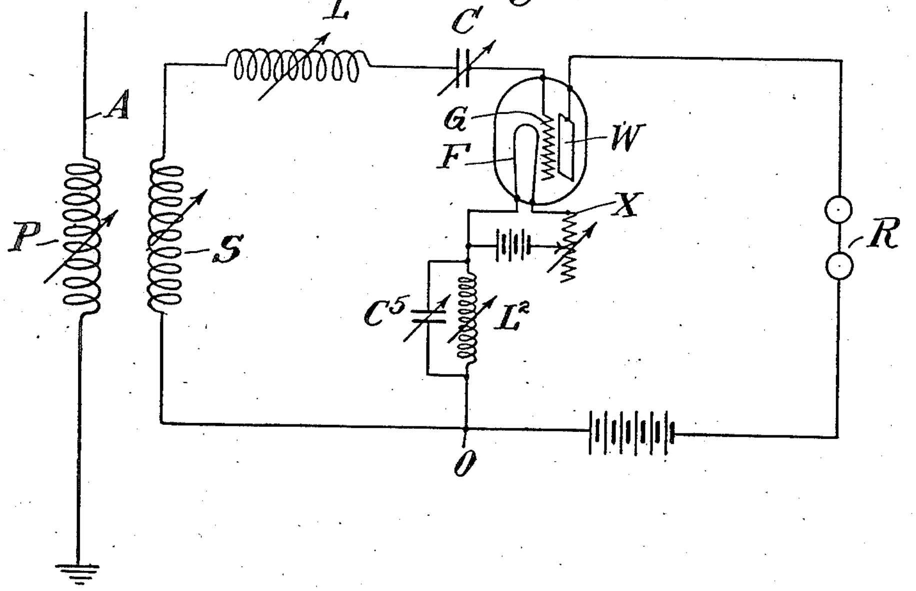
L
C
A
G
W
F
X
P
S
R
C5
L2
O

Fig. 6

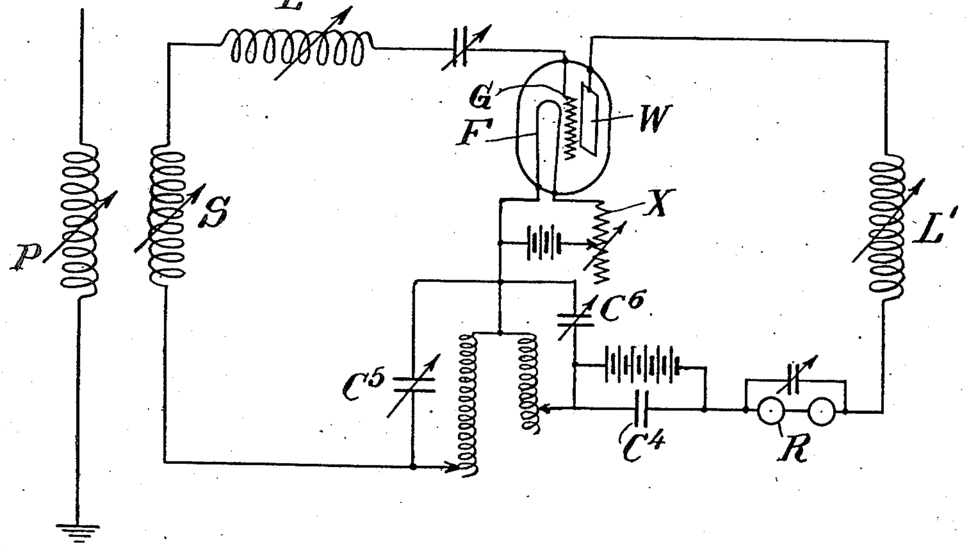
L
G
W
F
X
S
P
L'
C6
C5
C4
R

WITNESSES

L. Bates

C. B. Schroeder

INVENTOR

Edwin H. Armstrong

BY

Pennie, Davis & Goldsborough

ATTORNEYS

UNITED STATES PATENT OFFICE.

EDWIN H. ARMSTRONG, OF YONKERS, NEW YORK,

WIRELESS RECEIVING SYSTEM.

1,113,149. Specification of Letters Patent. Patented Oct. 6, 1914.

Application filed October 29, 1913. Serial No. 797,947.

To all whom it may concern:

Be it known that I, Edwin H. Armstrong, a citizen of the United States, residing at 1032 Warburton avenue, Yonkers, county of Westchester, State of New York, have invented certain new and useful Improvements in Wireless Receiving Systems; and I do hereby declare the following to be a full, clear, and exact description of the invention, such as will enable others skilled in the art to which it appertains to make and use the same.

The present invention relates to improvements in the arrangement and connections of electrical apparatus at the receiving station of a wireless system, and particularly a system of this kind in which a so-called "audion" is used as the Hertzian wave detector; the object being to amplify the effect of the received waves upon the current in the telephone or other receiving circuit, to increase the loudness and definition of the sounds in the telephone or other receiver, whereby more reliable communication may be established, or a greater distance of transmission becomes possible. To this end I have modified and improved upon the arrangement of the receiving circuits in a manner which will appear fully from the following description taken in connection with the accompanying drawings. As a preliminary, it is to be noted that my improved arrangement corresponds with the ordinary arrangement of circuits in connection with an audion detector to the extent that it comprises two interlinked circuits; a tuned receiving circuit in which the audion grid is included, and which will be hereinafter referred to as the "tuned grid circuit", and a circuit including a battery or other source of direct current and the "wing" of the audion, and which will be hereinafter referred to as the "wing circuit". As is usual, the two circuits are interlinked by connecting the hot filament of the audion to the point of junction of the tuned grid circuit and the wing circuit. I depart, however, from the customary arrangement of these circuits in a manner which may, for convenience of description, be classified by analysis under three heads; firstly, the provision of means, or the arrangement of the apparatus, to impart resonance to the wing circuit so that it is capable of sustaining oscillations corresponding to the oscillations in the tuned grid circuit; secondly, the provision of means supplementing the electrostatic coupling of the audion to facilitate the transfer of energy from the wing circuit to the grid circuit, thereby reinforcing the high frequency oscillations in the grid circuit, and thirdly, the introduction into the wing circuit of an inductance through which the direct current of the wing circuit flows, and which is so related to the grid circuit that the maintaining electromotive-force across the terminals of the inductance due to reduction of the direct current, is effective in the tuned grid circuit to increase the grid charge and consequently to further reduce the current in the wing circuit and in the telephones. By a further extension of this idea, the effect of the maintaining electromotive-force upon the grid current may be augmented by the use of a transformer in a manner which will be understood from the following description.

Figure 1 illustrates the arrangement and connection of apparatus with which I have thus far obtained the best results and which embodies in combination the several features of improvement which I have invented or discovered. Fig. 2 represents a like arrangement with the exception that there is no transformer for augmenting the effect of the maintaining electromotive-force, and several condensers, which may advantageously be employed but which are not essential, are eliminated. Fig. 3 illustrates an arrangement in which the advantages of my invention are only partially present, the inductance which produces the maintaining electromotive-force effective on the tuned grid circuit being eliminated. Fig. 4 illustrates an arrangement in which inductance, in this case the inductance of the telephones, is employed for producing the maintaining electromotive force effective on the tuned grid circuit, but the wing circuit is not resonant. Fig. 5 illustrates an arrangement in which an inductance replaces the telephones in that portion of the connections which is common to the two circuits, and the telephones are put in the wing circuit, and Fig. 6 illustrates an arrangement in which a double winding transformer is used, the primary being located in the wing circuit.

Referring particularly to Fig. 1, A represents the ordinary grounded aerial connected to the primary P of an oscillation transformer, the secondary S of which is connected as usual in the tuned grid circuit, this circuit also including the inductance L and preferably a shunted capacity C as is usual. Between the inductance L and the audion grid G is located a condenser C' adapted to receive and hold the charge which accumulates on the grid as a result of the received oscillations. The grid lies within the audion in the path of current in the wing circuit and the grid circuit is connected to the wing circuit at the junction point O, so that the two circuits are interlinked from the grid to that point. Between the junction point O and the filament of the audion I insert the telephone receivers R and the primary of the auto-transformer T. From junction point O the tuned grid circuit is completed through the secondary of the auto-transformer and condenser C^2 back to the secondary S; and the condenser C^5 is connected as shown in shunt to the telephones and the secondary of the auto-transformer. The wing circuit may be traced from the positive terminal of the battery through the inductance L' to the wing W and through the connections which are common to the two circuits, including the current path through the audion to the filament F, the telephone receivers R, and the primary of the auto-transformer and back to the negative terminal of the battery. The inductance L' is preferably shunted by a condenser C^3 and a condenser C^4 is placed across the battery terminals to afford a path of low impedance for the high frequency oscillation. The telephones R and the auto-transformer T are shunted by a condenser C^5, which affords a path for the high frequency oscillations in the grid circuit; and the telephones R and the primary of the auto-transformer are shunted by a capacity C^6, which affords a path for the high frequency oscillations in the wing circuit, and I find that the audion is made more stable and shows less tendency to become a high frequency generator and to set up oscillations in the interlinked circuits, if the tuned grid circuit is grounded as indicated. Each of the pieces of apparatus crossed with an arrow on the drawing is continuously variable, that is, may be varied by infinitesimal increments, and the condenser C^4 is preferably made adjustable by steps. I find that with such an arrangement of apparatus, and by properly adjusting the reactances, signals which are scarcely audible with the ordinary audion connection can be amplified to a point where they are too strong for, and "paralyze" the most stable audions that I have been able to obtain. The present understanding and modes of explanation of the phenomena which present themselves in such an arrangement is such that any theory of operation which may be advanced in regard to them is merely an attempt to explain the results attained in language which will be understandable to those skilled in the art; and it is with this idea in mind that the following description is written.

Upon reference to Fig. 2 of the drawings, which is an attempt to simplify the illustration of the essential elements of my invention, it will be observed that the auto-transformer T, and the condensers C^2, C^3, C^5, and C^6, are omitted. In this case the capacity of the telephone cords is sufficient to by-pass the high frequency oscillations.

The effect of tuning the wing circuit to the received high frequency oscillations by such an arrangement of apparatus as is illustrated in Fig. 3, may be explained as follows: In the ordinary audion connection, the wing W will be constantly maintained at the same potential with respect to the filament F and will constitute a surface of positive potential within the audion and having a constant tendency to absorb electrons. When the wing circuit is made resonant in accordance with my invention, the initial received oscillations in the grid circuit set up corresponding oscillations in the wing circuit so that a negative charge on the grid is accompanied by an increase of the potential of the wing, and a positive charge on the grid is accompanied by a decrease of the potential of the wing. The absorption of electrons by the grid and the building up of the charge in the condenser C' occurs only when the potential of the grid is positive with respect to the potential of the filament, and during these periods the potential of the wing with respect to the filament is reduced, thereby reducing its absorption of electrons and increasing the absorption of electrons by the grid. The charge thus built up in the condenser C' by absorption at the grid is entrapped and is an effective counter-electromotive-force in the wing circuit to reduce the current in the telephones. It is understood that on each successive wave train this effect is repeated, it being probable, according to my observations, that the full condenser charge is not built up by the first wave of a train but is gradually built up during the first portion of the wave train; and because of the high resistance of the discharge path of the condenser C′, this charge outlasts the duration of the wave train.

Referring now to Fig. 4, we may examine the effect of an inductance inserted in the wing circuit and related to the grid circuit as above defined. The reduction of current through the telephones, due to the received oscillations, produces a maintaining electromotive-force across the terminals of the telephones R which is in the same direction as the battery electromotive-force, and, because of the fact that the telephones are in that portion of the connections which is common to the two circuits, this maintaining electromotive-force is effective upon the tuned grid circuit to charge the condenser C′ in the same sense that the flow of electrons from the filament to the grid would charge the condenser. I have found, as a practical matter, that the arrangement of apparatus shown in Fig. 4 is effective to materially increase the loudness of the signals, but obviously better effects are obtained by combining the arrangements of Figs. 3 and 4 to make up the arrangement of Fig. 2. In this figure an auto-transformer L. is used instead of the two-coil transformer P. S., and in Figs. 2 and 4 the capacity of the telephone cords affords a path of low impedance for the passage of the high frequency oscillations about the inductance of the telephones. It will be observed, that this capacity of the telephone cords in Figs. 2 and 4, and likewise the capacity C^5 in Fig. 5 and the distributed capacity of the transformers in Figs. 1 and 6, is in each case common to the grid circuit and the wing circuit and constitutes an electrostatic coupling facilitating the transfer of energy from the wing circuit to the grid circuit and increasing the effect upon the grid of high frequency pulsations in the wing circuit. This effect occurs whether the wing circuit is tuned or not, but obviously the transfer of energy is increased by tuning the two circuits alike. I have discovered, however, that the beneficial effect may be still further increased by the interposition of a transformer in such a way as to increase the effect of the maintaining electromotive-force due to the reduction of current through the telephone receivers, and such a transformer is shown at T in Fig. 1. With the foregoing discussion in mind the effect of this transformer will at once be apparent. The inductance of the primary of the transformer is of course added to and increases the effect of the inductance of the telephones, but beyond this the maintaining electromotive-force across the primary is transformed into a higher electromotive-force in the tuned grid circuit. I have found that a ratio of transformation of 2 to 1 is sufficient in most cases when the inductance of the primary is equal to the inductance of the telephones, since a greater ratio causes the audion to become a high frequency generator setting up disturbing oscillations in the grid and wing circuits. The condenser C^5 by-passes the high frequency oscillations in the grid circuit about the telephones and the auto-transformer T, and the condenser C^6 by-passes the oscillations in the wing circuit about the telephones and the primary of the auto-transformer. The capacity of the telephone cords may, however, be relied upon to by-pass the oscillations about the telephones and condensers C^5 and C^6 utilized for by-passing the oscillations about the transformer only. The manner in which the different reactances will be adjusted so as to tune the two circuits to one another will be at once understood by those skilled in the art, and it will also be understood that the adjustable resistance X may be varied to get the proper temperature of the filament. Furthermore, it will be understood from what has been said that the ratio of transformation of the transformer should be adjusted to get the maximum signals without causing the audion to generate oscillations. The condenser C^5 has the additional effect, as I have found, to steady the audion and enable it to withstand greater variations of potential of the grid G and wing W without becoming a generator of disturbing oscillations in the grid and wing circuits. Particularly when working with long waves it is advisable to use this condenser in every case. I also find that the introduction of a condenser C^2 in the place indicated in Fig. 1, that is, between the auto-transformer and the secondary S, has the effect to prevent buzzing or humming in the telephones.

From what has been said, it will be understood that the inductance of the telephones may be utilized as the generator of the maintaining electromotive-force by inserting the telephones in the connections common to the two circuits, as shown in Figs. 1, 2 and 4; but a like result will be obtained by locating the telephones in the wing circuit as indicated in Fig. 5, and placing a suitable inductance as L^2 in that portion of the connections which is common to the two circuits. Indeed I find that the beneficial effect of an inductance so located can be obtained, as would be expected, by utilizing the primary of the transformer alone, as indicated in Fig. 6. In this case a two coil transformer is shown in place of an auto-transformer and the variations of current in the primary induce potential variations in the secondary which may be given any value by properly selecting the ratio of transformation.

Inasmuch as the purpose of the transformer is to magnify the effect of the maintaining electromotive-force in the inductance, it will be obvious that by increasing this inductance, the transformer may be rendered unnecessary, and some simplification of the whole arrangement may be attained by using telephone receivers of extremely high inductance and locating them in the position shown in Fig. 2, without any transformer.

The purpose of the condenser C^3 in Fig. 1 is to enable the wing circuit to be made resonant with the grid circuit when long waves are to be received. The capacity of the audion and the distributed capacity of the inductance L' are such that up to 2000 meters wave length resonance can be obtained in the wing circuit with reasonable values of inductance. For waves longer than 2000 meters resonance can be obtained in the wing circuit without the parallel condenser, but only by unduly increasing L'.

For the purpose of completing the disclosure of my invention, and to facilitate its practical application by those skilled in the art, I give below the values of the constants which I have found to work well for a wave length of 1800 meters, using an antenna having a capacity of .0012 microfarads and a standard McCandless audion. The inductance of the primary P was 760 microhenries and the combined inductance of the secondary S and the coil L was 3660 microhenries. The capacity of C was about .00025 microfarads, the capacity of C^2, .01 microfarads, and of C^4, 2 microfarads. The values of C' and L' are largely dependent upon one another and upon the amplitude of the received oscillations as well as the characteristics of the audion. These are the two elements which are chiefly relied upon by the operator to bring the system into condition to give the best results in receiving the signals from any particular station. The value of L' will be in the neighborhood of 10,000 microhenries, and of C' between .00003 and .0001 microfarads. If the inductance L' is not used, the value of C' can be much larger. I have usually employed about .003 microfarads as a matter of convenience; but there does not seem to be any particular advantage in increasing the capacity of C' beyond .0005 microfarads. The batteries, and the resistance X, are as used in the standard audion set. The way to tune this set is to cut out L', set C' at .00005 microfarads, and then adjust P, S, L and C as in the ordinary audion set until signals are strongest. The inductance L' is then gradually cut in and the strength of the signals will increase many times until a point is reached where the signals lose distinctness and there is a loud hiss in the telephones. This indicates that the audion is generating high frequency oscillations in the grid and wing circuits, and the inductance L' should be set at a point just below that at which this occurs. If there is no hiss as L' is increased, and the signals pass through a maximum of strength and begin to fall off, then L' should be set at the point of maximum strength of signals and C' should be increased to a point just below that at which the hiss appears. Under these circumstances, the increase of C' will be accompanied by an increase of the strength of the signals, the maximum strength being obtained in each case just below that point at which the audion begins to act as a generator of high frequency oscillations.

I find that the capacity required to bypass the oscillations about the inductance which is common to the two circuits, and about the transformer, if the transformer is used, is relatively small; and when the telephone receivers are used as the inductance at this point, the capacity of the telephone cords is sufficient. With greater wave lengths, however, when it becomes necessary to increase the tuning inductances L and L', the audion begins to be unstable, and I find that its stability can be restored by the use of a considerably larger capacity C^5. For example, with wave lengths of 4000 to 7500 meters I use a capacity C^5 of from .001 to .003 microfarads.

Having thus described my invention, what I claim is:

1. An audion wireless receiving system having a resonant wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, the resonant grid circuit having a capacity so related to the grid as to receive and retain the charge which accumulates thereon.

2. An audion wireless receiving system having a resonant wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, the resonant grid circuit having a capacity so related to the grid as to receive and retain the charge which accumulates thereon, and an inductance through which the current in the wing circuit flows, the grid circuit including connections for making effective upon that circuit the potential variations resulting from a change of current in the wing circuit.

3. An audion wireless receiving system having a resonant wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, the resonant grid circuit having a capacity so related to the grid as to receive and retain the charge which accumulates thereon, and an inductance in that portion of the connections which is common to the two circuits.

4. An audion wireless receiving system having a resonant wing circuit interlinked with a resonant grid circuit upon which the

received oscillations are impressed, the resonant grid circuit having a capacity so related to the grid as to receive and retain the charge which accumulates thereon, an inductance in that portion of the connections which is common to the two circuits, and a transformer having its primary in the wing circuit and its secondary in the grid circuit.

5. An audion wireless receiving system having a resonant wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, the resonant grid circuit having a capacity so related to the grid as to receive and retain the charge which accumulates thereon, and a telephone receiver in that portion of the connections which is common to the two circuits.

6. An audion wireless receiving system having a resonant wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, the resonant grid circuit having a capacity so related to the grid as to receive and retain the charge which accumulates thereon, a telephone receiver in that portion of the connections which is common to the two circuits, and a transformer having its primary in the wing circuit and its secondary in the grid circuit.

7. An audion wireless receiving system having a resonant wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, the resonant grid circuit having a capacity so related to the grid as to receive and retain the charge which accumulates thereon, an inductance in that portion of the connections which is common to the two circuits a transformer having its primary in the wing circuit and its secondary in the grid circuit, and condensers affording paths of low impedance about the inductance and transformer for the high frequency oscillations in both circuits.

8. An audion wireless receiving system having a wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, and an inductance in that portion of the connections which is common to the two circuits.

9. An audion wireless receiving system having a wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, and an inductance through which the current in the wing circuit flows, the grid circuit including connections for making effective upon that circuit the potential variations resulting from a change of current in the wing circuit.

10. An audion wireless receiving system having a wing circuit interlinked with a resonant grid circuit, an inductance in that portion of the connections which is common to the two circuits, and a transformer having its primary in the wing circuit and its secondary in the grid circuit.

11. An audion wireless receiving system having a wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, an inductance in that portion of the connections which is common to the two circuits, a transformer having its primary in the wing circuit and its secondary in the grid circuit, and condensers affording paths of low impedance about the inductance and transformer for the high frequency oscillations in both circuits.

12. An audion wireless receiving system having a wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, and a high inductance telephone receiver in that portion of the connections which is common to the two circuits.

13. An audion wireless receiving system having a wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, a capacity between that portion of the grid circuit upon which the oscillations are impressed and the grid, and a second capacity between the point of junction of the grid circuit with the wing circuit and that portion of the grid circuit upon which the oscillations are impressed.

14. An audion wireless receiving system having a wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, an inductance in that portion of the connections which is common to the two circuits, and a stabilizing capacity shunted about said inductance.

15. An audion wireless receiving system having a wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, and means supplementing the coupling of the audion to facilitate transfer of energy from the wing circuit to the grid circuit, whereby the effect upon the grid of high frequency pulsations in the wing circuit is increased.

16. An audion wireless receiving system having a resonant wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, and means supplementing the coupling of the audion to facilitate transfer of energy from the wing circuit to the grid circuit, whereby the effect upon the grid of high frequency pulsations in the wing circuit is increased.

17. An audion wireless receiving system having a wing circuit interlinked with a resonant grid circuit upon which received

oscillations are impressed and an electrostatic coupling between the circuits supplementing the coupling of the audion to facilitate transfer of energy from the wing circuit to the grid circuit, whereby the effect upon the grid of high frequency pulsations in the wing circuit is increased.

18. An audion wireless receiving system having a resonant wing circuit interlinked with a resonant grid circuit upon which received oscillations are impressed and an electrostatic coupling between the circuits supplementing the coupling of the audion to facilitate transfer of energy from the wing circuit to the grid circuit, whereby the effect upon the grid of high frequency pulsations in the wing circuit is increased.

In testimony whereof I affix my signature, in presence of two witnesses.

EDWIN H. ARMSTRONG.

Witnesses:
William H. Davis,
John C. Pennie.

Zapper Schematics
No claims are made by KeelyNet for any of these plans
They are provided for informational purposes only!
KeelyNet 01/10/02

Simple Zapper Schematic

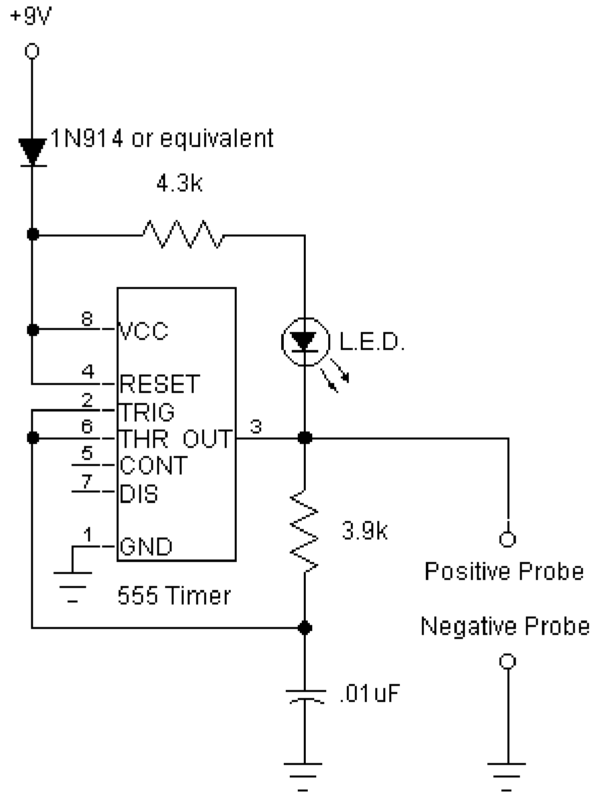
+9V
1N914 or equivalent
4.3k
8 VCC
4 RESET
2 TRIG
6 THR OUT 3
5 CONT
7 DIS
1 GND
555 Timer
L.E.D.
3.9k
Positive Probe
Negative Probe
.01uF

The Research of Dr. Hulda Regehr Clark, Ph. D., N.D.

I have recently (9/15/2000) come across the research of Dr. Hulda Regehr Clark, Ph. D., N.D. In her books, she states that many diseases, among which are cancer and HIV/AIDS, can be cured by the use of, among other things, certain herbs and a frequency generator called a "Zapper" that was invented by her and her son.

Speaking of mothers :), my mother, 62, had fatigue and arthritic symptoms in her hands before I found out about and told her about Dr. Hulda Clark. She is a seamstress, so the arthritic symptoms hindered her work. After doing the Parasite Cleanses and using the Zapper, she doesn't have to take naps like before and her hands are flexible again. She has done the complete Parasite Cleanse only twice but she Zaps practically every day.

Simple Zapper Information

Schematic updated on 12/21/2001

Simple Zapper Information text updated on 12/21/2001.

I forgot to mention that the Simple Zapper that I'm using has a 1 Megohm potentiometer instead of the 2.2k ohm resistor in order to vary the output frequency. If you use a potentiometer with the circuit put it in series with the 2.2k ohm resistor because the chip gets hot when there is too little resistance.

The duty-cycle is pretty consistent with a range of 51% to 57%. With the oscilloscope's frequency analyzer in storage mode, I hold the Zapper probes and put the oscilloscope probes somewhere on my body and then I vary the frequency of the Zapper using the 1 Megohm potentiometer.

When testing my body on a spectrum analyzer while using the Simple Zapper with an output of around 5 Volts (tested on a DC voltmeter), I noticed that the frequencies around 4,000 hz to around 18,000 hz had larger peaks telling me that my body at the time was more conductive (less reactive?) and/or less resistive to frequencies in this range at the given voltage.

With my body being more conductive (less reactive?) at these frequencies, I believe that there is a better dispersal of the Simple Zapper's output within my body. I did this recent test with one Zapper probe in each hand (as opposed to both in one hand) and the oscilloscope probes (RCA to 1/4" phone plug adaptor) in my mouth near my wisdom teeth with my lips around the RCA ground (like a lollipop).

The oscilloscope's frequency analyzer was in storage mode. With the Simple Zapper circuit on this page, change the 2.2k ohm resistor to 4.3k ohm if you want a ~15,000 hz frequency. I have recently built a couple of Zappers with optimized output frequencies of around 11,000 hz. I haven't tested anyone using the Simple Zapper other than myself, so there is the possibility that your body's response may differ.

For more information, check out my Simple Zapper webpage at:

www.members.aol.com/mas1911/index.html

or for really interesting free circuit schematics:

www.members.tripod.com/~mas1911/zappers20011221tripod.html

You can write the Author of Simple Zapper if you have comments or questions.

The following document, written by Luke Parrish, provides instructions on building an astable multivibrator in DC pulse mode, identical in output to the Zapper, described in The Cure For All Diseases by Dr. Hulda Clark.

Plans For Zapper

This page tells how to make a Hulda Clark-style "Parasite Zapper". It's a simpler and more efficient version, but works just as well. We make no theraputic claims for this device. It is for experimental purposes only. We provide a ready made version for those who prefer not to spend so much time building it.

Materials

- A CD4069 hex inverter (Radio Shack link)
- 1 1M resistor, 1 10k resistor, and 1 1k resistor (Radio Shack link) (Resistor color code)
- 1 1000 pF capacitor (Radio Shack link)

NOTE: The size of the resistors and capacitors can vary, because they determine the frequency of the zapper, which doesn't really matter much according to Hulda Clark.

The following parts may be substituted according to preference, though something similar will be needed:
- Solderless breadboard (Radio Shack link)
- 9-volt battery
- 4 alligator clip-leads sets Radio Shack link
- 2 copper pipes (hand-holds)

Instructions

Plug the components into the solderless breadboard according to the following schematic. Make sure none of the wires touch each other. To hook up the power, attach an alligator clip-lead to each side of the capacitor that stretches from hot to ground. Clip the one that goes to hot to the positive terminal on the battery and the one that goes to ground to the negative terminal. The two hand-holds are attached with the other two clip-leads. (One goes to ground, the other to the end of a resistor that goes to pin 12.)

Schematic

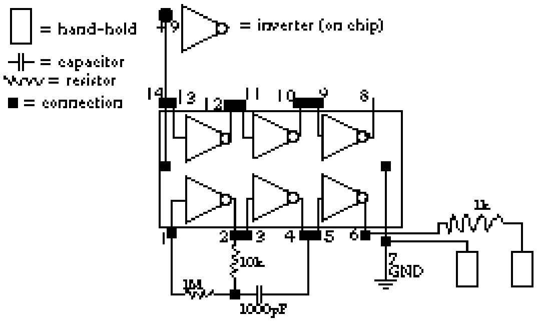

Schematic most recently updated 5-17-2001

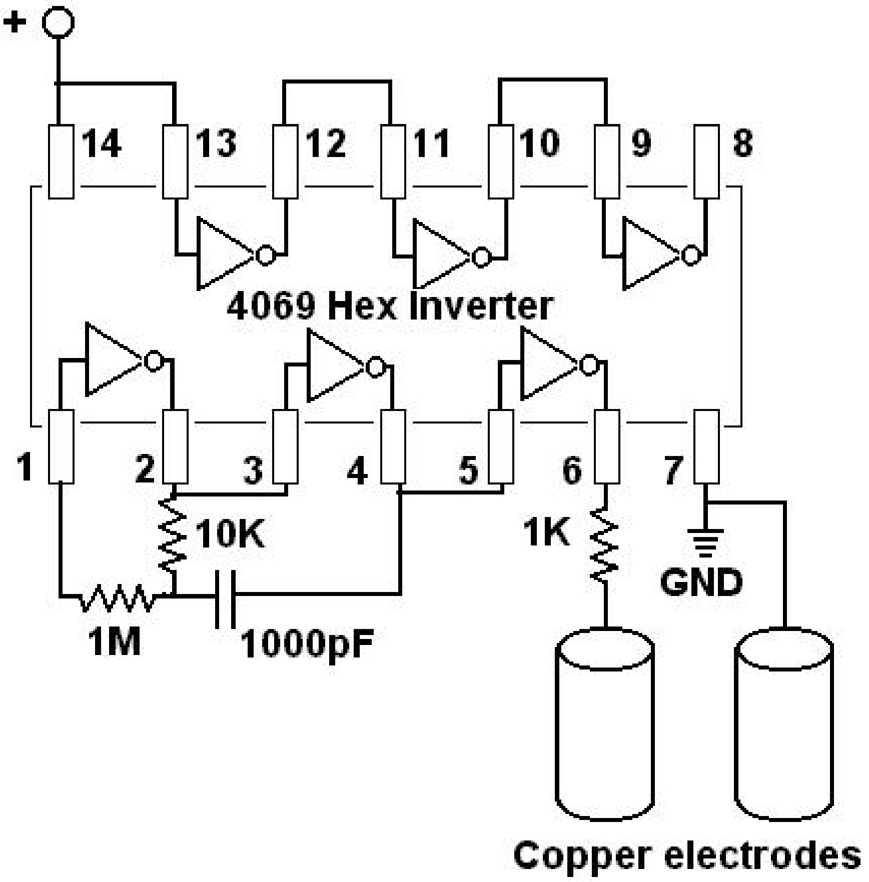

The above instructions are not the only way to do it. You can, for example, solder all the components directly to a 14-pin socket, like we do in our ready made version. This is a lot more trouble and more time-consuming for most people, which is why we recommend a breadboard. The resistors and capacitor don't all have to be the exact sizes mentioned here to make a working Zapper. However, any variation in capacitance or resistivity will change the frequency. (There are certain limits on how high of frequencies the chip can handle, so don't stray too far from the specified values.) If you have an oscilliscope, that's handy to check if it's working.

Ready made Zappers

Free plans from Zapperlab.com

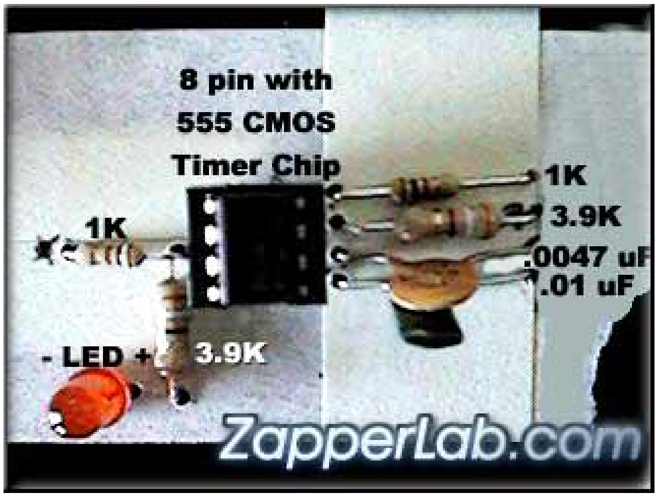

Here we have the easy step-by-step instructions for building your very own Hulda Clark Zapper reprinted with permission from her books (provided below). For advanced electricians there is a schematic for building a Hulda Clark Zapper as well. If you do not know anyone that could put one together for you, you can now order the zapper Hulda Clark uses as described in "The Cure For All Cancers" and "The Cure For All Diseases" at: www.SyncroZap.com. The Dr. Clark Zapper can be sold only as an experimental device, and no claim can be made by us legally as to a diagnosis, cure or treatment for any medical condition or disease.

Build A Dr. Clark Zapper

[Order A Dr. Clark Zapper]()

Instructions for building your very own zapper can be found reprinted below from the books by Dr. Hulda Clark. If you would rather order a pre-built zapper, just click [here]().

How To Build A Zapper from *The Cure For All Cancers*:

The Cure For All Cancers © Copyright 1993 by Hulda Regehr Clark, Ph.D., N.D.

Get from Radio Shack, (serial number included to make it easy).
Order these parts from their web site at:
www.RadioShack.com

___ Black plastic Project Box Radio Shack # 270-1809

___ 9 volt battery

___ 9 volt battery clips: Radio Shack # 270-325

___ On off switch: Radio Shack # 275-624A

___ 1K Ohm resistor: RS# 271-1321

___ 3.9K Ohm resistor: RS# 271-1123

___ low current red LED: RS# 276-044

___ .0047 uF capacitor: RS# 272-130

___ .01 uF capacitor: RS# 272-1065

___ 555 CMOS timer chip: RS# 276-1723

___ 8 pin wire wrapping socket for CMOS chip: RS# 276-1988

___ two packs of Microclip test jumpers: RS# 278-017

___ one pack of 14" alligator clip leads RS# 278-1156C

Get from any Hardware Store:

2 Bolts 1/8" diameter, 2" long with 4 nuts and washers.

2 pieces of ¾" copper pipe, cut to 4 inches long each.

Editor's Note:

Remove 8 Pins From A 16 Pin Wire Wrap Socket

The **8 pin wire-wrapping socket** has been discontinued at many Radio Shacks and readers have asked what to do. Some say it is still available on their website at www.RadioShack.com, if not:

There are a couple options. One can still get the **16 pin wire-**

wrapping socket. Pull out the bottom grouping of 8 pins, and you now have an 8 pin socket. Just make sure when you plug the 555 Timer into it, you line it up with the remaining 8 pins. Very simple.

There is also a **Low Profile 8 pin socket**. The pins are so small one must solder to them.

The wire-wrap socket has long pins for easy twisting onto other wires. The low profile is for those that solder electronics.

If you are soldering your zapper together get the low profile 8 pin socket. Since the socket is just a saddle for the 555 Timer, it is not a vital part of the circuit. It's just a way to put long or short pins on the 555 Timer for easy twisting. If you cannot find one at your local Radio Shack, simply get the 16 pin wire-wrap socket, and pull out 8 pins. -editor.

Assembling The Zapper

If you have tools such as a drill, needle nose pliers, and small drill bits, buy one of the plastic project boxes on the list, otherwise build your zapper in a shoe box, or a box half the size of a shoe box.

1. You will be using the lid of the shoe box or plastic lid of the project box to mount the components. Save the box to enclose the finished project.

2. Pierce two holes near the ends of the lid. Enlarge the holes with a pen or pencil until the bolts would fit through. Mount the bolts on the outside about half way through the holes so there is a washer and nut holding it in place on both sides. Tighten. Label one hole "grounding bolt" on the inside and outside.

3. Mount the 555 chip in the wire wrap socket. Find the "top end" of the chip by searching the outside surface carefully for a cookie-shaped bite or hole taken out of it. Align the chip with the socket and very gently squeeze the pins of the chip into the socket until they click in place.

4. Make 8 pinholes to fit the wire wrap socket. Enlarge them slightly with a sharp pencil. Mount it on the outside. Write in the numbers of the pins (connections) on both the outside and inside, starting with number one to the left of the "cookie bite" as seen from outside. After number 4, cross over to number 5 and continue. Number 8 will be across from number 1. The pins are numbered like this:

1	8
2	7

3 6

4 5

5. Pierce two holes ½ inch apart very near to pins 5,6,7, and 8. They should be less than 1/8 inch away. (Or, one end of each component can share a hole with the 555 chip.) Mount the .01 uF capacitor near pin 5 on the outside. On the inside connect pin 5 to one end of this capacitor by simply twisting them together. Loop the capacitor wire around the pin first; then twist with the long-nose pliers until you have made a tight connection. Bend the other wire from the capacitor flat against the inside of the shoe box lid. Label it .01 on the outside and inside. Mount the .0047 uF capacitor near pin 6. On the inside twist the capacitor wire around the pin. Flatten the wire from the other end and label it .0047. Mount the 3.9 K Ohm resistor near pin 7, connecting it on the inside to the pin. Flatten the wire on the other end and label it 3.9. Mount the 1 K Ohm resistor and connect it similarly to pin 8 and label it 1K.

6. Pierce two holes ½ inch apart next to pin 3 (again, you can share the hole for pin 3 if you wish), in the direction of the bolt. Mount the other 1 K Ohm resistor and label inside and outside. Twist the connections together and flatten the remaining wire. This resistor protects the circuit if you should accidentally short the terminals. Mount the 3.9 K Ohm resistor downward. One end can go in the same hole as the 1K resistor near pin 3. Twist that end around pin 3 which already has the 1K resistor attached to it. Flatten the far end. Label.

7. Next to the 3.9 K Ohm resistor pierce two holes ¼ inch apart for the LED. Notice that the LED has a positive and a negative connection. The longer wire is the positive (anode). Mount the LED on the outside and bend back the wires, labeling them + and - on the inside.

8. Near the top pierce a hole for the toggle switch. Enlarge it until the shaft fits through from the inside. Remove nut and washer from switch before mounting. You may need to trim away some paper with a serrated knife before replacing washer and nut on the outside. Tighten.

9. Next to the switch pierce two holes for the wires from the battery holder and poke them through. Attach the battery and tape it to the outside.

Now To Connect Everything

Twist free ends of the two capacitors .01 and .0047 together. Connect to Grounding Bolt using an alligator clip.

Bend pin 2 and pin 6 together inward, using an alligator clip, catch them and connect to free end of 3.9K Ohm (by pin 7).

Alligator clip Pin 7 to free end 1K Ohm (near pin 8)

Using two microclips Pin 8 and pin 4 to one end of switch (use hole to attach both microclips).

Free end of 1K Ohm (by pin 3) to the Bolt using an alligator clip.

Alligator clip the free end of 3.9K Ohm (by pin 3) to plus end (long) of LED.

Minus end (short) of LED to Grounding Bolt with an alligator clip.

Alligator clip pin 1 to the Grounding Bolt.

Alligator clip Black battery wire to Grounding Bolt.

Micro clip the Red battery wire to free end of switch.

Hook up the battery, and the light should turn on, click the switch if it does not, check connections if LED does not light up.

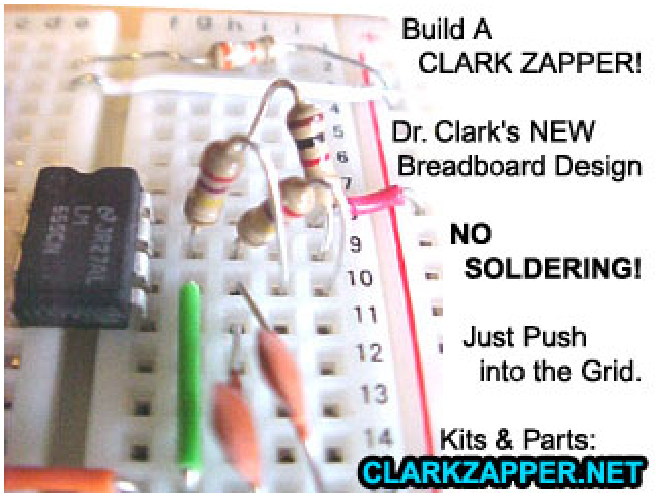

Attach a long lead wire to the grounding bolt & bolt, and then to the copper handles. Wrap handles in paper towel, clip with lead wire, wet handle when you are ready to zap.)

How To Zap

Wrap handles in wet paper towel, or wet cotton sleeves, and hold in each hand. If you have wrist cuffs, run them under water to wet them, attach to each wrist.

Attach handles or wrist cuffs to zapper.

Turn zapper on, zapping for 7 minutes.

At end of 7 minute zapping rest 20 minutes.

Do this a minimum of three 7 minute zapping sessions, with a 20 minute break in between each session. Some people zap five times in extreme cases.

Order A Dr. Hulda Clark Zapper - Click Here!

Learn about Dr Clark's latest zapping technique at: www.PlateZapping.com

All information is for educational purposes only. These statements have not been evaluated by the FDA. Not intended to diagnose, treat, prevent, mitigate or cure disease.

This Circuit designed by Geoff Clark and approved by Dr. Hulda Clark

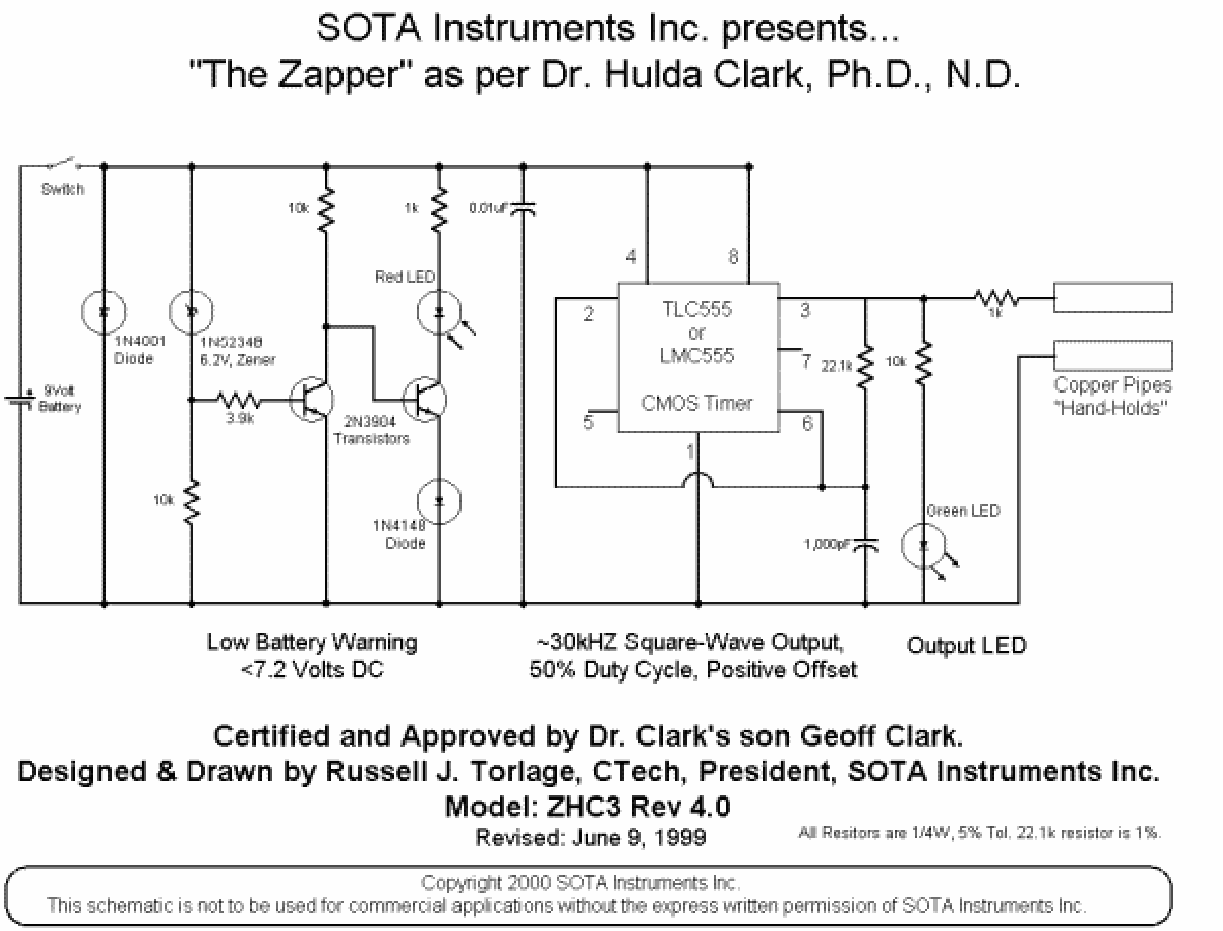

SOTA Instruments Inc. presents...
"The Zapper" as per Dr. Hulda Clark, Ph.D., N.D.

Switch
9Volt Battery
1N4001 Diode
1N5234B 6.2V, Zener
10k
3.9k
10k
1k
Red LED
2N3904 Transistors
1N4148 Diode
0.01uF
4
8
2
3
5
6
7
1
TLC555 or LMC555
CMOS Timer
22.1k
10k
1,000pF
Green LED
1k
Copper Pipes "Hand-Holds"
Low Battery Warning <7.2 Volts DC
~30kHZ Square-Wave Output, 50% Duty Cycle, Positive Offset
Output LED

Certified and Approved by Dr. Clark's son Geoff Clark.
Designed & Drawn by Russell J. Torlage, CTech, President, SOTA Instruments Inc.
Model: ZHC3 Rev 4.0
Revised: June 9, 1999

All Resitors are 1/4W, 5% Tol. 22.1k resistor is 1%.

Copyright 2000 SOTA Instruments Inc.
This schematic is not to be used for commercial applications without the express written permission of SOTA Instruments Inc.

If you found this file useful or interesting, please consider a donation or a purchase to help keep KeelyNet online and providing free information. Even a dollar will help. Others sell it, we prefer to share it, thanks!

PayPal DONATE

US 20100201133A1

(19) **United States**

(12) **Patent Application Publication**

MESLER

(10) **Pub. No.: US 2010/0201133 A1**

(43) **Pub. Date: Aug. 12, 2010**

(54) **ZERO POINT ENERGY ROTATOR TRANSDUCER AND ASSOCIATED METHODS**

(76) Inventor: **PAUL R. MESLER**, EMERY, UT (US)

Correspondence Address:
THORPE NORTH & WESTERN, LLP.
P.O. Box 1219
SANDY, UT 84091-1219 (US)

(21) Appl. No.: **12/541,007**

(22) Filed: **Aug. 13, 2009**

Related U.S. Application Data

(60) Provisional application No. 61/088,913, filed on Aug. 14, 2008.

Publication Classification

(51) **Int. Cl.**

H02K 7/18	(2006.01)
F03G 7/10	(2006.01)

(52) **U.S. Cl.** .... **290/1 C**; 290/1 R; 74/DIG.009

(57) **ABSTRACT**

An energy conversion system for extracting energy from the zero-point field of space is disclosed and described. This system includes a primary transducer operatively associated with a secondary transducer. The primary transducer can include a substrate having an arm rotatably coupled to the substrate, while the substrate can be confined to move along a linear oscillatory path. The secondary transducer converts kinetic energy of the substrate and arm to usable energy. The multi-mode motion of the substrate and arm combination is sufficient to impart a momentum influx to rotational motion of the arm.

The momentum influx can then be converted to usable energy such as mechanical, electrical and/or thermal energy. As described in more detail herein, the momentum influx appears to be a result of radiation pressure due to scattering in part of the electro-magnetic momentum-energy contained within the zero-point field of space opposite to centripetal acceleration of the arm.

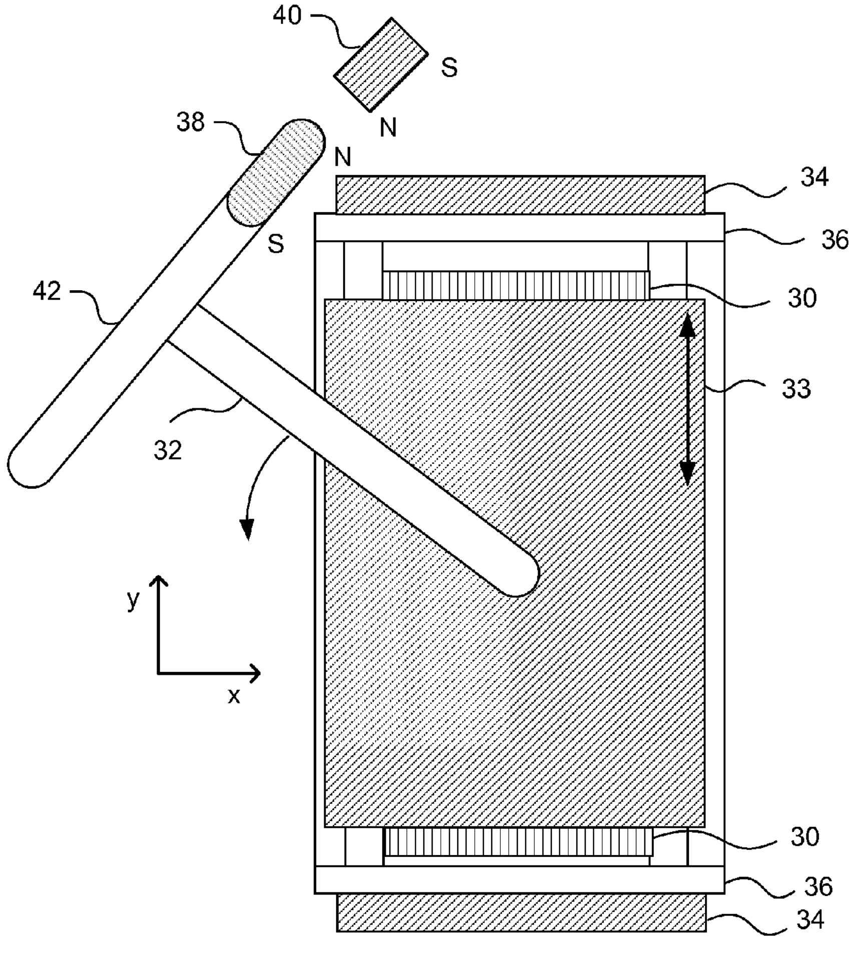

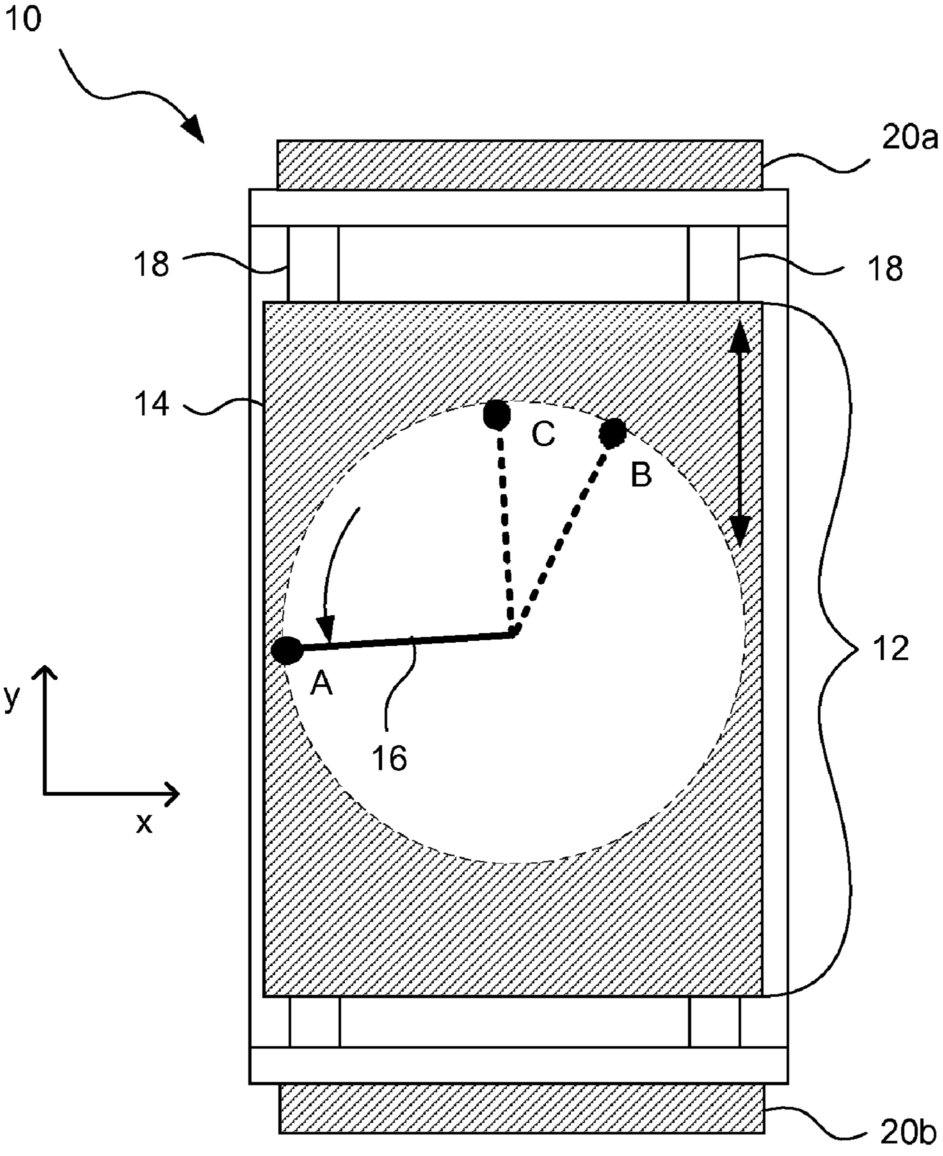
10
20a
18
18
14
C
B
A
12
16
y
x
20b

FIG. 1

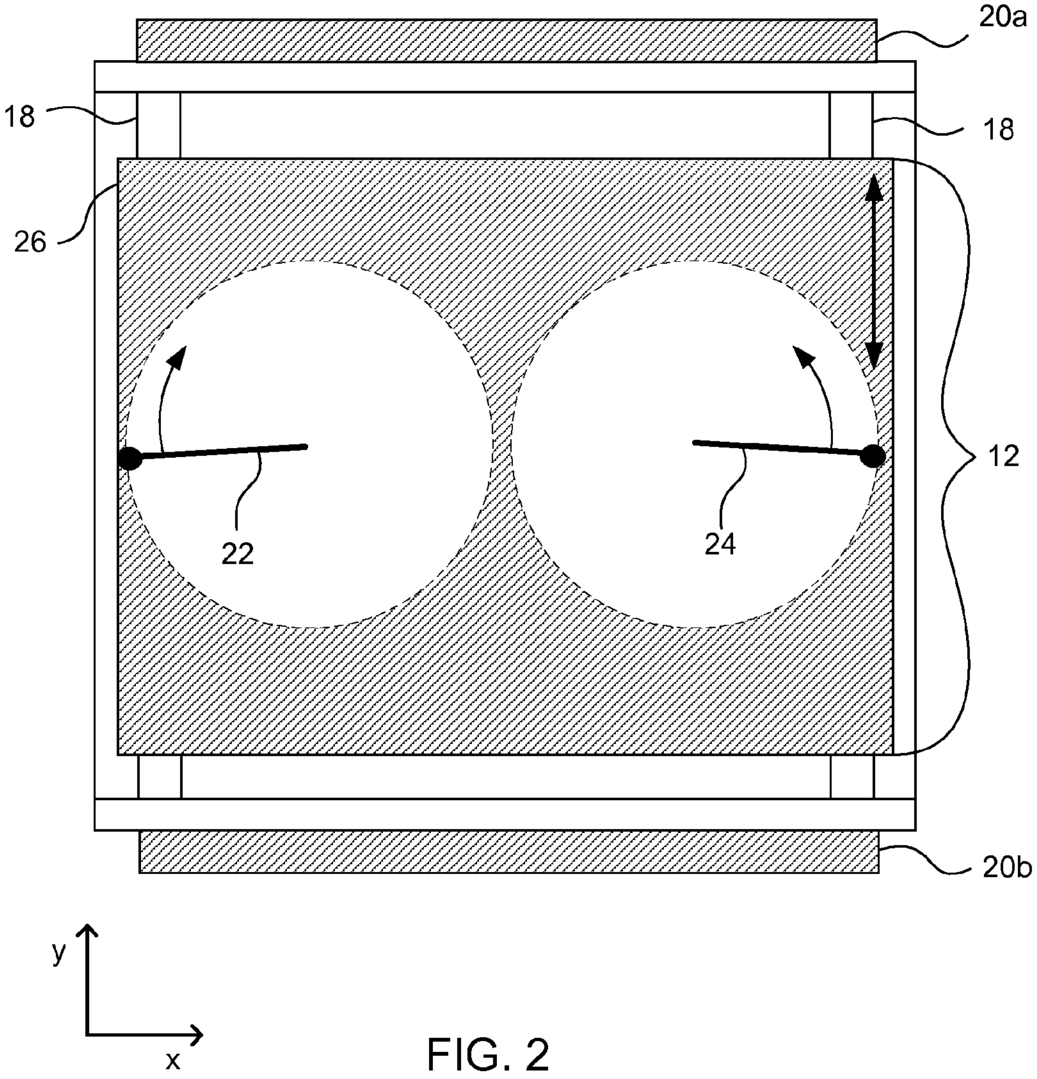
20a
18
18
26
12
22
24
20b
y
x

FIG. 2

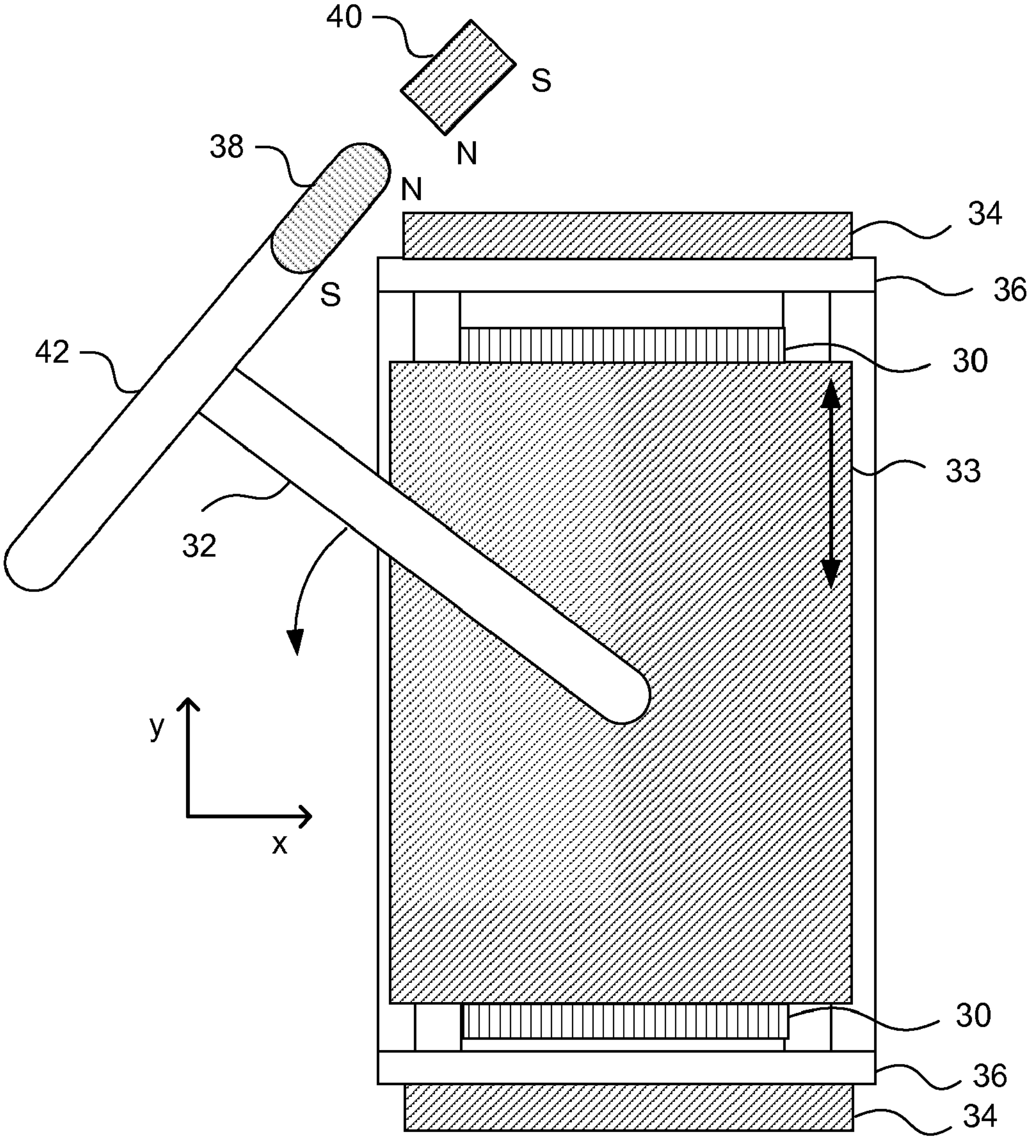
40
S
N
38
N
S
34
36
42
30
33
32
y
x
30
36
34

FIG. 3

ZERO POINT ENERGY ROTATOR TRANSDUCER AND ASSOCIATED METHODS

RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 61/088,913, filed Aug. 14, 2008 which is incorporated herein by reference.

FIELD OF THE INVENTION

[0002] This invention relates to generation of energy from the zero-point energy field of space or electromagnetic quantum vacuum. More specifically, the invention relates to a device useful for extraction of useful work via a sliding rotator. Therefore, the present invention relates generally to the fields of physics, quantum physics, and electromagnetism.

BACKGROUND OF THE INVENTION

[0003] The now ubiquitous drive for energy independence has provided a new set of incentives and appreciation for development of environmentally and economically responsible energy resources. Conventional discussions regarding such energy resources often revolve around fossil fuels, nuclear, and "alternative energy sources" such as solar, wind, bio fuels, hydrogen, fuel cells, and the like. Each of these energy resources currently comes with various benefits and drawbacks and often entail trade-offs between perceived pollution and environmental problems, costs, efficiencies, and unintended consequences (e.g. corn demand for use as fuel versus food).

[0004] Since at least the 1960's some theoretical physicists have proposed the concept of zero-point energy which suggests a background energy field throughout space. Historically, these physicists have been at least partially marginalized by mainstream academic thought. However, over the past decade or so, such theories have begun to gain significant traction in terms of recognition and available proofs to support these theories. Although this developing area of physics has experienced progress, practical applications of such theories remain severely limited. One example of such progress in practical applications is the work of Bernard Haisch and others in capitalizing on the Casimir effect. Such efforts thus far remain relegated to the quantum level. Although a wide variety of purported devices and claims have been made, no macroscale efforts have yet proven to yield useful energy gains.

SUMMARY OF THE INVENTION

[0005] In light of the relative infancy of the field and deficiencies noted above, the present invention provides an energy conversion system for extracting energy from the zero-point field of space. This system includes a primary transducer operatively associated with a secondary transducer. The primary transducer can include a substrate having an arm rotatably coupled to the substrate, while the substrate can be confined to move along a linear oscillatory path. The secondary transducer converts kinetic energy of the substrate and arm to usable energy. The multi-mode motion of the substrate and arm combination is sufficient to impart a momentum influx to a rotational motion of the arm. The momentum influx can then be converted to usable energy such as mechanical, electrical and/or thermal energy. As described in more detail herein, the momentum influx appears to be a result of radiation pressure resulting from scattering in part of the electro-magnetic momentum-energy contained within the zero-point field of space opposite to centripetal acceleration of the arm.

[0006] There has thus been outlined, rather broadly, the more important features of the invention so that the detailed description thereof that follows may be better understood, and so that the present contribution to the art may be better appreciated. Other features of the present invention will become clearer from the following detailed description of the invention, taken with the accompanying drawings and claims, or may be learned by the practice of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] The present invention will become more fully apparent from the following description and appended claims, taken in conjunction with the accompanying drawings. Understanding that these drawings merely depict exemplary embodiments of the present invention and they are, therefore, not to be considered limiting of its scope. It will be readily appreciated that the components of the present invention, as generally described and illustrated in the figures herein, could be arranged, sized, and designed in a wide variety of different configurations. In particular, relative dimensions may vary and are generally not presented in an optimized configuration but are rather provided for clarity and convenience in describing the invention generally. Nonetheless, the invention will be described and explained with additional specificity and detail through the use of the accompanying drawings in which:

[0008] FIG. **1** is a schematic of an energy conversion system in accordance with one embodiment of the present invention.

[0009] FIG. **2** is a schematic of an energy conversion system having two rotating arms, one rotating clockwise and the other counterclockwise, in accordance with another embodiment of the present invention.

[0010] FIG. **3** is a schematic of an energy conversion system illustrating an electro-magnetic mechanism for accelerating the arm in accordance with yet another embodiment of the present invention.

DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS

[0011] The following detailed description of exemplary embodiments of the invention makes reference to the accompanying drawings, which form a part hereof and in which are shown, by way of illustration, exemplary embodiments in which the invention may be practiced. While these exemplary embodiments are described in sufficient detail to enable those skilled in the art to practice the invention, it should be understood that other embodiments may be realized and that various changes to the invention may be made without departing from the spirit and scope of the present invention. Thus, the following more detailed description of the embodiments of the present invention is not intended to limit the scope of the invention, as claimed, but is presented for purposes of illustration only and not limitation to describe the features and characteristics of the present invention, to set forth the best mode of operation of the invention, and to sufficiently enable one skilled in the art to practice the invention. Accordingly, the scope of the present invention is to be defined solely by the appended claims.

[0012] The following detailed description and exemplary embodiments of the invention will be best understood by reference to the accompanying drawings, wherein the elements and features of the invention are designated by numerals throughout.

Definitions

[0013] In describing and claiming the present invention, the following terminology will be used.

[0014] The singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "an arm" includes reference to one or more of such members and reference to "converting" refers to one or more such steps.

[0015] As used herein with respect to an identified property or circumstance, "substantially" refers to a degree of deviation that is sufficiently small so as to not measurably detract from the identified property or circumstance. The exact degree of deviation allowable may in some cases depend on the specific context.

[0016] As used herein, "adjacent" refers to the proximity of two structures or elements. Particularly, elements that are identified as being "adjacent" may be either abutting or connected. Such elements may also be near or close to each other without necessarily contacting each other. The exact degree of proximity may in some cases depend on the specific context.

[0017] As used herein, a plurality of items, structural elements, compositional elements, and/or materials may be presented in a common list for convenience. However, these lists should be construed as though each member of the list is individually identified as a separate and unique member. Thus, no individual member of such list should be construed as a de facto equivalent of any other member of the same list solely based on their presentation in a common group without indications to the contrary.

[0018] Concentrations, amounts, and other numerical data may be presented herein in a range format. It is to be understood that such range format is used merely for convenience and brevity and should be interpreted flexibly to include not only the numerical values explicitly recited as the limits of the range, but also to include all the individual numerical values or sub-ranges encompassed within that range as if each numerical value and sub-range is explicitly recited. For example, a numerical range of about 1 to about 4.5 should be interpreted to include not only the explicitly recited limits of 1 to about 4.5, but also to include individual numerals such as 2, 3, 4, and sub-ranges such as 1 to 3, 2 to 4, etc. The same principle applies to ranges reciting only one numerical value, such as "less than about 4.5," which should be interpreted to include all of the above-recited values and ranges. Further, such an interpretation should apply regardless of the breadth of the range or the characteristic being described.

[0019] In the present disclosure, the term "preferably" or "preferred" is non-exclusive where it is intended to mean "preferably, but not limited to." Any steps recited in any method or process claims may be executed in any order and are not limited to the order presented in the claims. Means-plus-function or step-plus-function limitations will only be employed where for a specific claim limitation all of the following conditions are present in that limitation: a) "means for" or "step for" is expressly recited; and b) a corresponding function is expressly recited. The structure, material or acts that support the means-plus function are expressly recited in the description herein. Accordingly, the scope of the invention should be determined solely by the appended claims and their legal equivalents, rather than by the descriptions and examples given herein.

Embodiments of the Invention

[0020] Referring now to FIG. **1**, an energy conversion system **10** is shown generally for extracting energy from the zero-point field of space. The system can include a primary transducer **12** which is designed to extract energy from the zero-point field of space in the form of kinetic energy. Multi-mode motion can be induced in a primary transducer along an oscillatory path. The multi-mode motion can include a primary translational motion and a secondary motion such that abrupt changes in the primary translational motion are sufficient to impart a momentum influx to the secondary motion. In one embodiment, the primary transducer can include a substrate **14** having an arm **16** rotatably coupled to the substrate. The substrate is generally confined to move along an oscillatory path. A secondary transducer **20***a* can be operatively associated with the primary transducer to covert the momentum influx into usable energy. Thus, the systems of the present invention provide an energy conversion system for extracting energy from the zero-point field of space. Generally, the system includes a means for extracting momentum-energy from vacuum by the action of inertia and a means for converting the momentum-energy to usable energy operatively associated with the means for extracting momentum-energy.

[0021] More particularly, the substrate **14** can be any suitable member upon which a rotating arm **16** can be coupled or otherwise mounted. The substrate can be guided along the oscillatory path by any suitable mechanism such as, but not limited to, tracks, walls or the like. FIG. **1** shows a linear oscillatory path defined by a pair of tracks **18** along which the substrate can move. The substrate can move along the tracks via bearings, wheels, magnetically repulsive members, or any other suitable approach. Typically, it is most desirable to minimize friction between the substrate and tracks in order to maximize energy transfer to the secondary transducer. The substrate and arm can be formed of any suitable material which can be readily chosen based on designed parameters to meet various property requirements such as density, fracture strength, toughness, etc. Furthermore, the shape and contours of the substrate and arm can also be designed to minimize aerodynamic resistance due to the significant acceleration forces per unit of time experienced by the primary transducer.

[0022] In one aspect of the present invention, the substrate has a substrate mass (m_2) and the arm has an arm mass (m_1) such that a ratio of m_2 to m_1 is chosen to maximize a gain in kinetic energy (ΔKE) of the substrate. The rotating arm acts as the primary momentum influx extraction mechanism which experiences an inertial force acting opposite to the centripetal force acting on the rotating arm. A fraction of the inertial force is manifest at an axis of rotation which is rigidly attached to the substrate and which constrains the arm to rotate around this axis. The y-component of this force acting on the axis, denoted as the centrifugal reactive force, imparts an impulse on the substrate in the positive y-direction with respect to an x-y coordinate system when the arm rotates within the first quadrant of the said x-y coordinate system or imparts an impulse on the substrate in the negative y-direction when the arm rotates within the third quadrant of the said x-y coordinate system, such that the substrate experiences a net timerate influx of momentum from the zero-point energy field of space or an increase in the momentum of the substrate in either the positive or negative y-direction, depending on which quadrant the arm is rotating in.

[0023] The oscillatory path can be a linear track **18** which defined on both ends by a pair of resilient stationary members **20***a* and **20***b* oriented on either end of the linear track. During operation of the system, the primary transducer oscillates back and forth between the two stationary members. The approach at each end is accompanied by a transfer of kinetic energy of the primary transducer (e.g. substrate and arm) to the secondary transducer. The primary transducer can thus be coupled to the secondary transducer in a manner which allows for conversion of this kinetic energy to the secondary transducer. For example, this can be accomplished by providing an at least partially or substantially inelastic collision with the resilient stationary members. As discussed in more detail below, various mechanisms can be effectively used to convert kinetic energy of the primary transducer into usable energy, e.g. via mechanical, thermal, and/or electromagnetic features.

[0024] As can be readily appreciated, the arm and substrate need an initial input of energy from an external source to begin movement. The source of energy to initiate rotation of the arm can come from a conventional energy source, such as a battery. This external energy can be transferred to the arm such as by a coaxial motor, magnetic accelerator, external flywheel, or the like. For example, a drive motor can be operatively associated with the arm and configured to initiate rotational motion of the arm and sustain the rotational motion at a predetermined critical angular velocity. Another option is to provide a permanent magnet on the arm, e.g. at the free end, and a complimentary magnetic induction coil at a location external to the primary transducer. In this manner, an electric current applied to the induction coil creates a magnetic field sufficient to act on the permanent magnet and accelerate the arm. After a critical angular velocity is reached, the external source of energy (such as a battery, the power grid, etc.) for the motor or other suitable initiator mechanism can be switched to an alternative energy storage device such as collected energy.

[0025] Referring now to FIG. **2**, an alternative embodiment of the present invention is shown including an arm **22** and a secondary arm **24** each rotatably coupled to the substrate. In one embodiment, the secondary arm has an axis of rotation spaced apart from an axis of rotation of the arm **22**. Alternatively, the secondary arm could share a common axis of rotation. In either case, the secondary arm is coupled to the arm such that both the arm and the secondary arm rotate at a common angular velocity and opposite rotational directions. Typically, the secondary arm also has a substantially identical mass and dimensions to that of the arm. In other words, the arms swing forward and back in the same cycle, with common y-coordinates and mirror x-coordinates.

[0026] In one embodiment, a motor can drive the first arm, and the second arm can be driven by gears linked to the first arm, such that they are in sync with each other, rotating at the same angular speed but in opposite rotational directions. In such cases, the cosines of the rotation angle of each arm will always be equal in magnitude, but opposite in sign. Thus, the x-components of the centrifugal reactive force acting on each axis would be equal and opposite in direction, and therefore, they would cancel each other. As a result, the substrate having two rotating arms will not impart compressive forces on the corresponding linear track along the x-axis during operation of the system.

[0027] The kinetic energy of the primary transducer can be converted to usable energy twice for every full rotation of the arm(s), i.e. once between the 0° position and the 90° position and once between the 180° and the 270° position. The optimal position of the arm at impact of the substrate with the stationary member can vary but is generally in these ranges. Specifically, as the arm approaches 90° or 270° the gain approaches 0 (see Example, Equation 3). Similarly, the momentum of the substrate at 0° and 180° is also 0 so the gain at those points would also be 0. A suitable secondary transducer can thus be coupled with the primary transducer as a means for converting kinetic and momentum-energy of the substrate and arm into usable energy such as electrical, mechanical, or thermal energy. The secondary transducer can be an electromagnetic induction system, mechanical gear system, heat transfer system, piezoelectric system, or the like. For example, the means for converting the momentum-energy can be an induction coil system, heat dissipater, magnetic brake, friction brake, or linear to rotational motion translator. A heat dissipater can include a substantially inelastic collision between the substrate and a set of bumpers. Heat generated by collision can be converted to energy through heat transfer to produce steam which can be used to drive a turbine, or other similar thermo-electric conversion devices.

[0028] In one specific embodiment of the present invention, the secondary transducer includes at least one induction coil system including a magnet and a complimentary conductive coil. The magnet can be physically coupled to the substrate and oriented to approach the complimentary conductive coil sufficient to induce electrical current therein during movement of the substrate along the oscillatory path. A similar magnet system can be incorporated into the arm as well. Such a secondary transducer can also be utilized as the initiator system by switching the energy source, i.e. from an external battery to energy derived from the zero-point energy field. FIG. **3** illustrates a simplified version of such an electromagnetic-kinetic energy conversion system for the secondary transducer. In particular, a permanent magnet **30** can be coupled to each collision end of the substrate **33** having the arm **32** rotatably coupled thereto. An external induction coil **34** can be oriented at a position (or positions) to allow the kinetic energy of the substrate to be progressively converted to electrical energy at each approach of the substrate to the stationary members **36**. In another embodiment, a piezoelectric material such as a crystal or ceramic can be coupled to the stationary member. The force applied by the substrate to the piezoelectric material can generate an electric potential in response to the applied mechanical stress. Alternatively, the secondary transducer can include a mechanical coupling between the substrate and an energy translator. The energy translator can be a linear-to-rotational motion drive, heat surfacing, or rack and pinion, although other mechanisms may also be suitable.

[0029] After the substrate collides with either of the stationary members, the arm's rotational speed can be replenished by a suitable arm accelerator. For example, an electro-magnet system can be oriented with a permanent magnet **38** at the free end of the arm with a corresponding external induction coil **40**. In this case, the arm can have a t-shape to allow the permanent magnet to be oriented substantially in-line with the magnetic field produced by the external electro-magnet (although opposite signs). This allows for a maximum transfer of kinetic energy to the arm from the electro-magnet acceleration system. Alternatively, the permanent magnet and induction coil can be swapped. Typically, the t-portion **42** can have substantially equal mass on either end such that the center of mass lies along the arm axis. This electro-magnet can impart an impulsive torque to the arm in either the second or fourth quadrant (or at any point right after the collision of the substrate with a stationary member) as the permanent magnet attached to the arm passes closely to the electro-magnet. The advantage of this approach is that the arm's rotation will be maintained substantially by electro-magnetism without additional gears or mechanical losses due to the friction of such gears. This replenishment of the rotational speed of the arm can optionally be achieved by a motor in the second and fourth quadrant (or at any point right after the collision of the substrate with a stationary member) by using pulsed technology, wherein the power to the motor is only turned on in the second or fourth quadrant.

[0030] In operation, the movement of the substrate can follow a linear oscillatory path having two endpoints while the arm(s) follows a cyclic path such that translational motion of the substrate in a y-direction along the oscillatory path substantially coincides with translational motion of the arm in the y-direction. Furthermore, the arm has two theoretical maximum energy transfer positions, the positions occurring twice for every cycle of rotation of the arm. Actual maximum energy transfer positions can be empirically determined and a collision swing angle adjusted to obtain a substantially maximum in the usable energy, depending on the specific design, i.e. relative masses, arm length, rotational velocity, etc.

[0031] There are at least two factors that impact the overall maximum energy gain of the system. For example, the substrate's final kinetic energy before collision with a stationary member partly determines the maximum energy gain of the system. Maximum energy gain is also determined by the angular displacement of the arm at the time of the collision, the angular velocity of the arm, the linear velocity of the substrate prior to collision and the relative masses of the arm and substrate.

[0032] During operation of the system, a full cycle of rotation of the rotator can be divided into the four conventional quadrants counter-clockwise. Upon start-up, the arm is initiated with a battery driven motor (or equivalent, such as an electro-magnet system) to reach a terminal velocity while holding the slider substrate stationary. The terminal velocity is determined by the friction in the system and the power transfer capabilities of the device that delivers power to the rotator to keep it rotating. After the rotator has reached this constant velocity, integrated sensors in the system release the slider just prior to the rotator arriving at its 0 degree position. The slider will begin to accelerate when the y-component of the centrifugal reactive force acting on the physical vertical axis attached to the slider is greater than any static friction force between the slider and its linear track. The angle of the rotator with respect to the x-axis at which angle the slider begins to accelerate is denoted as phi. During this first quadrant rotation of the rotator, the power to the motor can be optionally turned off so that the rotator continues to rotate on its own rotational inertia.

[0033] The distance between the front and back secondary transducers or stationary members can be adjusted such that the slider "collides" with this secondary front transducer when the rotator has rotated an angle less than 90 degrees (e.g. 15° to 75° although other ranges may be useful such as about)45° with respect to the x-axis which gives a maximum transfer of energy.

[0034] In one embodiment, the secondary transducer can comprise a linear generator. The linear generator can include a coil or set of coils receiving the magnets attached to the slider, and all of the electronic and electrical circuitry associated with the linear generator must be able to handle a brief impulse of power if the slider is to come to a near stop quickly. The linear generator can be designed of sufficiently robust materials and capacity to handle this power surge. Alternatively, the secondary transducer can include a thermal conversion system or a piezoelectric device, as previously discussed. More particularly, the stationary members and substrate can be designed to provide a substantially inelastic collision such that much of the kinetic energy of the substrate is converted to heat or electricity. In a thermal conversion process, the substrate and stationary member can accordingly be thermally linked to a water reservoir such that water is heated to steam, which drives an electrical generator.

[0035] A thermal conversion process may be more effective for a large scale energy conversion system. A piezoelectric conversion process may be more useful for a small scale energy conversion system, such as an energy conversion system used in a micro-machine or other type of micro-electro-mechanical system (MEMS).

[0036] After the slider comes to a complete stop after the collision of the slider with a stationary member a pulsed power source can be used to replenish the speed of the rotator back to its initial speed. There will typically be a loss in speed of the rotator due to friction and a replenishment of the speed of the rotator must happen twice every cycle after every collision of the slider with a stationary member.

[0037] Based on the guidance provided herein, a speed of the rotational motion, relative masses of the arm and the substrate, geometry of the arm, and the geometry of the substrate can all be designed such that zero-point energy is utilized to maintain the movement of the substrate. Non-limiting examples of specific variables which can be adjusted include aerodynamic shape of the arm and substrate, length of the arm, materials of construction, density of substrate and arm, and the like.

[0038] Theoretical Background

[0039] Without being bound to any particular theory, it is thought that the principles of operation of the present invention are governed by a theoretical backdrop described in more detail below. In summary, this theoretical backdrop indicates that the momentum influx to the primary transducer is a result of radiation pressure due to scattering in part of the electromagnetic momentum-energy contained within the zero-point field of space opposite to the centripetal acceleration of the arm. The systems of the present invention can be sized and incorporated into a wide variety of applications which are currently occupied by conventional energy sources, e.g. batteries, power generators, fuel cells, and so forth. A more in depth description of one theory follows; however, it is understood that the following description is not exclusive, exhaustive or definitive of the underlying inventive concept.

[0040] We begin a discussion with a rotating point particle with a mass of m_1, and denote it as the rotator. One end of the rotator is attached to a vertical axis by a mass-less wire of length r, so that the rotator rotates freely counter-clockwise around this axis with two degrees of freedom in a plane parallel to the earth. The vertical axis is rigidly attached to a second object denoted as the slider with mass m_2. The slider is constrained to move along a linear track with one degree of freedom, coincident to the y-axis of an x-y Cartesian coordinate system. Initially, the axis of rotation is set at the origin of the coordinate system.

[0041] The rotator is aligned coincident with the x-axis of the x-y coordinate system, such that the rotator is located at 180 degrees with respect to the x-axis. The slider presses against a back bumper so that the slider cannot move in the negative y-direction. An impulse is applied on the rotator in the negative y-direction, causing the rotator to rotate counterclockwise. The rotator rotates until it passes the 0 degrees point, and the slider moves in the positive y-direction at some angle after this 0 degrees point. Immediately prior to the movement of the slider, the initial tangential velocity of the rotator is denoted as v_i. A short time interval later as the slider moves in the positive y-direction, the slider is allowed to make an inelastic collision with a front bumper at some angle θ of the rotator with respect to the x-axis. The tangential velocity of the rotator prior to the collision of the slider is denoted as v_α. The post-collision tangential velocity of the rotator is denoted as v_f. The linear velocity of the slider prior to the collision of the slider is denoted as v_s.

[0042] The initial tangential kinetic energy of the rotator is:

$$KE_{rotator\ initial}\frac{1}{2}m_1v_i^2 \quad (1).$$

[0043] With respect to the laboratory frame this represents the initial energy of the rotator-slider system. As the slider accelerates forward in the positive y-direction under the action of the y-component of the centrifugal reactive force on the axis, both the slider and the rotator acquire an equal instantaneous linear velocity in the positive y-direction, and therefore, a linear kinetic energy with respect to the laboratory frame. By the conservation of energy, the tangential kinetic energy of the rotator must decrease, such that:

$$\Delta KE_{rotator}=KE_{linear\ slider}+KE_{linear\ rotator} \quad (2).$$

Or:

$$\frac{1}{2}m_1v_i^2-\frac{1}{2}m_1v_\alpha^2=\frac{1}{2}m_2v_s^2+\frac{1}{2}m_1v_s^2 \quad (2a).$$

[0044] Rearranging, we have:

$$\frac{1}{2}m_1v_i^2-\frac{1}{2}m_1v_\alpha^2=\frac{1}{2}m_2v_s^2+\frac{1}{2}m_1v_s^2 \quad (3).$$

[0045] That is, by the conservation of energy the pre-collision total kinetic energy observed in the system must equal the initial kinetic energy of the rotator.

[0046] Upon collision of the slider with the front bumper, the kinetic energy of the slider equal to $\frac{1}{2} m_2 v_s^2$ is essentially converted to thermal energy. The linear kinetic energy of the rotator is accounted for as follows.

[0047] Taking θ as the angle of the rotator at the time of collision, and by application of the Pythagorean Theorem, we have:

$$\frac{1}{2}m_1v_s^2=\frac{1}{2}m_1\sin^2\theta v_s^2+\frac{1}{2}m_1\cos^2\theta v_s^2 \quad (4).$$

[0048] The first term on the right side of the above equation represents the kinetic energy of the slider transformed to thermal energy upon collision of the slider. It is reasonable to assume that the energy described by the second term transfers and adds to the tangential kinetic energy of the rotator.

[0049] The tangential kinetic energy of the rotator just prior to the collision is:

$$KE_{rotator\ \alpha}=\frac{1}{2}m_1v_\alpha^2 \quad (5).$$

[0050] Adding the second term on the right of equation 4 to equation 5 to obtain the final post-collision tangential kinetic energy of the rotator, the answer would be:

$$KE_{rotator\ post\ collision}=\frac{1}{2}m_1v_\alpha^2+\frac{1}{2}m_1\cos^2\theta v_s^2 \quad (6).$$

[0051] However, the above equation is incorrect because it was assumed the second term could be used in equation 4 to obtain the final result. This equation, derived from the Pythagorean Theorem, is correct from a conservation of energy perspective, but it cannot be used for post-collision analysis. The post-collision tangential kinetic energy of the rotator must first be obtained by vector addition of vector entities.

[0052] The correct procedure is as follows. Just prior to the collision of the slider, the rotator has a linear velocity in the positive y-direction equal to the linear velocity of the slider. Thus, the rotator has a linear momentum in the positive y-direction equal to:

$$p_{rotator}=m_1v_s \quad (7).$$

[0053] At collision, a part of this momentum is transferred to the front bumper or the earth, and the amount of momentum transferred is equal to:

$$p_{bumper}=\sin\theta m_1v_s \quad (8).$$

[0054] By the conservation of angular momentum the remaining momentum is transferred to the rotator, such that there is an increase in the post-collision angular momentum of the rotator, equal to:

$$m_1v_fr=m_1v_\alpha r+m_1\cos\theta v_sr \quad (9).$$

[0055] Dividing both sides of the above equation by m_1 and r, we obtain:

$$v_f=v_\alpha+\cos\theta v_s \quad (10).$$

[0056] Using the above, the correct expression for the final post-collision tangential kinetic energy of the rotator is derived, equal to:

$$KE_{rotator\ pst\ collision}=\frac{1}{2}m_1v_\alpha^2+m_1v_\alpha\cos\theta v_s+\frac{1}{2}m_1\cos^2\theta v_s^2 \quad (11).$$

[0057] Now, recalling equation 6, we have:

$$KE_{rotator\ post\ collision}=\frac{1}{2}m_1v_\alpha^2+\frac{1}{2}m_1\cos^2\theta v_s^2$$

[0058] Subtracting the above two equations, the difference is obtained as:

$$m_1v_\alpha\cos\theta v_s$$

[0059] Recalling equation 3:

$$\frac{1}{2}m_1v_i^2=\frac{1}{2}m_1v_\alpha^2+\frac{1}{2}m_2v_s^2+\frac{1}{2}m_1v_s^2$$

[0060] Substitute equation 4 into the last term of the above, giving:

$$\frac{1}{2}m_1v_i^2=\frac{1}{2}m_1v_\alpha^2+\frac{1}{2}m_2v_s^2+\frac{1}{2}m_1\sin^2\theta v_s^2+\frac{1}{2}m_1\cos^2\theta v_s^2$$

[0061] Regroup the above as follows:

$$\frac{1}{2}m_1v_i^2=\left(\frac{1}{2}m_2v_s^2+\frac{1}{2}m_1\sin^2\theta v_s^2\right)+\left(\frac{1}{2}m_1v_\alpha^2+\frac{1}{2}m_1\cos^2\theta v_s^2\right) \quad (12).$$

[0062] This is the correct equation for the total energy of the system prior to the collision of the slider. Now take equation 12 and substitute equation 11 for the last two terms in the last parentheses, giving the total energy of the system after the collision of the slider:

$$KE_{system\ final}=\left(\frac{1}{2}m_2v_s{}^2{}_{te}+\frac{1}{2}m_1\sin^2\theta v_s{}^2{}_{te}\right)+\left(\frac{1}{2}m_1v_\alpha^2+m_1v_\alpha\cos\theta v_s+\frac{1}{2}m_1\cos^2\theta v_s^2\right) \quad (13).$$

[0063] Where, the subscript represents the thermal energy transformed from kinetic energy after the collision of the slider.

[0064] The left side of equation 12 represents the initial tangential kinetic energy of the rotator or the initial total energy of the system prior to collision. Equation 13 represents the total energy of the system after the collision. Clearly, there has been a gain in the total energy of the system by the amount of:

$$\Delta E = m_1 v_\alpha \cos\theta v_s$$

[0065] Substituting v_β for $\cos\theta v_s$, the above can be expressed as:

$$\Delta E = m_1 v_\alpha v_\beta$$

[0066] By the first law of thermodynamics there must be an influx of energy-momentum to the system to account for this gain in energy. One possible explanation for the influx may be found in a paper that postulates inertia is due to a time rate change of incoming momentum flux to an accelerating object from the vacuum energy-momentum of space (A. Rueda & B. Haisch, *Contribution to inertial mass by reaction of the vacuum to accelerated motion*, Foundations of Physics, Vol. 28, No. 7, pp. 1057-1108 (1998) which is incorporated herein by reference). The paper states:

[0067] We demonstrate that from the point of view of a nearby inertial observer there exists a net energy and momentum flux (Poynting vector) of ZPF radiation transiting the accelerating object in a direction necessarily opposite to the acceleration vector. The scattering opacity of the object to the transiting flux creates the back-reaction force customarily called the inertial reaction.

[0068] If the above is true, then at the time of the collision of the slider an influx of momentum from the vacuum transits the rotator in the positive y-direction, opposite to the instantaneous linear de-acceleration of the rotator and slider in the negative y-direction. A portion of this momentum, equal to $m_1 \cos\theta v_s r$, transfers to the rotator and increases its angular momentum by an amount found in equation 9, which was derived as a consequence of the conservation of angular momentum.

[0069] It has been known since the last century that energy has inertia implied by Einstein's equation $E=m_0c^2$. Rearranging the equation, we see the mass equivalent of energy expressed as $E/c^2=m_0$. That is, energy in the form of radiation can interact with physical objects as if the energy had the properties of material objects—momentum, etc. Hence, if we assume space is filled with an irrepressible average minimum vacuum energy as predicted by the Heisenberg uncertainty principle in part composed of electro-magnetic radiation, the interaction of accelerating objects with the photons comprising this radiation could indeed create a reaction force. For example, a photon contains momentum equal to:

$$p = hv/c$$

[0070] It is conceivable that photons "colliding" with the rotator opposite to the instantaneous linear de-acceleration of the rotator could transfer part of their momentum to the rotator. And if this were true, along with a transfer of momentum, energy would necessarily transfer also. For example, in the relativistic expression relating the relationship between kinetic energy and momentum we have:

$$(K+m_0c^2)^2 = (pc)^2 + (m_0c^2)^2$$

[0071] Reduced to the Newtonian limit, the above can be expressed as:

$$K = (p)^2/(2m_0)$$

[0072] Hence, any increase in the angular momentum of the rotator due to this transfer of momentum from the vacuum must be accompanied by an increase in kinetic energy with respect to a laboratory frame.

[0073] The above explanation is more plausible than any Machian explanation. One may incorrectly argue that the increase in energy is due to linear frame-dragging, the so-called Lens-Thirring effect, predicted by the general theory of relativity. To date, the Gravity Probe B experiment has not conclusively proven this effect, but even if it did, it would not have any bearing on this discussion. The frame-dragging effect is only observed in the de-accelerating frames of the rotator and the slider as the slider collides with the front bumper. It is not observed in the frame of the laboratory because the mass of the universe is obviously not accelerating with respect to this frame. To insist this effect must apply to the laboratory frame would violate the principle of relativity by implying the de-accelerating frames of the rotator and slider are preferred frames and the effects observed in these frames must be imposed in the frame of the laboratory. An explanation that only applies to the laboratory frame is demonstrated below.

Example

[0074] An experiment was performed to test if energy gain occurred in a rotator-slider system. The experiment tested two hypotheses. The momentum influx hypothesis (MIH) asserted there would be a gain in the post-collision rotational energy of the rotator, and therefore, a gain in the energy of the system as a whole by the action of an influx of energy-momentum to the system via inertia. The null hypothesis or classical hypothesis (CH) asserted there would be no gain in energy of the post-collision rotational energy of the rotator, and therefore, there would be no gain in the system as a whole. The null hypothesis was designated as the classical hypothesis because it takes the classical Newtonian assumption that inertia is an intrinsic property of matter, and therefore, could never affect a gain in energy of any system. The momentum influx hypothesis takes the position that inertia is an extrinsic property of matter a link to the vacuum energy of space under the right conditions. In the experiment the rotator is a physical object. The equations derived in this paper were derived for a point particle. Thus, we must modify the equations previously demonstrated by including the moment of inertia of the rotator.

[0075] Expressed mathematically, the CH maintains the following:

$$\tfrac{1}{2}I\omega_i^2 = \tfrac{1}{2}m_2v_s^2 + \tfrac{1}{2}m_1\sin^2\theta v_s^2 + \left(\tfrac{1}{2}I\omega_\alpha^2 + \tfrac{1}{2}m_1\cos^2\theta v_s^2\right) \quad (1).$$

[0076] The above essentially states that the post-collision energy of the rotator, the terms in parentheses, when added to the other terms, equals the initial rotational kinetic energy of the system. Thus, the energy of the system is conserved.

[0077] The MIH is expressed mathematically as:

$$E_{final} = \frac{1}{2}m_2v_s + \frac{1}{2}m_1\sin^2\theta v_s^2 + \left(\frac{1}{2}I\omega_\alpha^2 + \frac{I\omega_\alpha\cos 0\theta v_s}{r} + \frac{I\cos^2\theta v_s^2}{2r^2}\right). \quad (2)$$

[0078] The terms in parentheses above represent the post-collision energy of the rotator derived from the conservation of angular momentum. When added to the remaining terms, the above is greater than equation 1 by the following, which expresses the gain in energy of the system:

$$\Delta E = \frac{I\omega_\alpha \cos\theta v_s}{r} + \frac{I\cos^2\theta v_s^2}{2r^2} - \frac{m_1\cos^2\theta v_s^2}{2}. \tag{3}$$

[0079] Where:

[0080] I=the moment of inertia of the rotator

[0081] ω_α=the angular velocity of the rotator prior to collision

[0082] θ=the angle of the rotator at the time of collision

[0083] v_s=the linear velocity of the slider prior to collision

[0084] r=the distance from the axis to the center of mass of the rotator

[0085] In the experiment the post-collision energy of the rotator was measured and compared to the MIH and CH prediction. The collected data is in Table 1 while the results are presented in Table 2.

[0086] In Table 1 and Table 2 the following terms are defined:

[0087] ω_α=the measured angular velocity of the rotator prior to collision

[0088] θ=the measured angle of the rotator at the time of collision

[0089] v_s=the measured linear velocity of the slider prior to collision

[0090] ω_i=the measured initial angular velocity of the rotator

[0091] ω_f=the measured final post-collision angular velocity of the rotator

[0092] ke_i=the measured initial kinetic energy of the rotator

[0093] ke_α=the measured kinetic energy of the rotator prior to collision

[0094] ke_f=the measured final kinetic energy of the rotator after collision

[0095] m_1=the mass of the rotator

[0096] m_2=the mass of the slider

[0097] I=the moment of inertia of the rotator

[0098] r=the radial distance from the axis to the center of mass of the rotator

[0099] m_1 was 3.89 kg, m_2 was 1.97 kg, l was 0.403 kg m^2, and r was 0.314 m.

TABLE 1

ω_α	θ	v_s	ω_i	ω_f	ke_i	ke_α
8.80	32.0	0.294	9.81	9.72	19.4	15.6
8.80	34.5	0.450	9.54	9.44	18.3	15.6
7.30	39.0	0.540	8.89	8.71	15.9	10.7
7.33	45.0	0.622	9.62	9.17	18.6	10.8

TABLE 2

ke_f(J)	MIH Prediction (J)	CH Prediction (J)	Net over CH (J)
19.0	20.2	15.73	3.3
18.0	21.5	15.87	2.1
15.3	16.2	11.08	4.2
16.9	16.3	11.20	5.7

[0100] As can be seen, the net gain over CH predicted values was from about 13% to about 50%.

[0101] The foregoing detailed description describes the invention with reference to specific exemplary embodiments. However, it will be appreciated that various modifications and changes can be made without departing from the scope of the present invention as set forth in the appended claims. The detailed description and accompanying drawings are to be regarded as merely illustrative, rather than as restrictive, and all such modifications or changes, if any, are intended to fall within the scope of the present invention as described and set forth herein.

What is claimed is:

1. An energy conversion system for extracting energy from the zero-point field of space, comprising:

a) a primary transducer including a substrate having an arm rotatably coupled to the substrate, said substrate being confined to move along an oscillatory path; and

b) a secondary transducer operatively associated with the primary transducer, the secondary transducer converting kinetic energy of the substrate and arm to usable energy.

2. The system of claim **1**, wherein the oscillatory path is a linear track defined by a pair of resilient stationary members oriented on both ends of the linear track.

3. The system of claim **1**, further including a drive motor operatively associated with the arm and configured to initiate rotational motion of the arm and sustain the rotational motion at a critical angular velocity.

4. The system of claim **2**, wherein the substrate has a substrate mass (m_2) and the arm has an arm mass (m_1) such that a ratio of m_2 to m_1 is chosen to maximize a gain in kinetic energy (ΔKE) of the substrate and the arm.

5. The system of claim **1**, further comprising a secondary arm rotatably coupled to the substrate and having a secondary axis of rotation spaced apart from an axis of rotation of the arm, said secondary arm being coupled to the arm such that both the arm and the secondary arm rotate at a common angular velocity and opposite rotational directions, and wherein the secondary arm has a substantially identical mass and dimensions to that of the arm.

6. The system of claim **5**, wherein the secondary arm rotates about a common axis of rotation with the arm.

7. The system of claim **1**, wherein the secondary transducer includes at least one induction coil system including a magnet and a complimentary conductive coil, such that the magnet is physically coupled to the substrate and oriented to approach the complimentary conductive coil sufficient to induce electrical current therein during movement of the substrate along the oscillatory path.

8. The system of claim **1**, wherein the secondary transducer includes piezoelectric material operable to convert a force applied by the substrate to the secondary transducer to an electric potential in response to applied mechanical stress.

9. The system of claim **1**, wherein the secondary transducer includes a mechanical coupling between the substrate and an energy translator.

10. The system of claim **9**, wherein the energy translator is a linear to rotational motion drive, heat surfacing, or rack and pinion.

11. The system of claim **1**, wherein the usable energy is mechanical, thermal and/or electrical energy.

12. An energy conversion system for extracting energy from the zero-point field of space, comprising a means for extracting momentum-energy from vacuum by the action of inertia and a means for converting the momentum-energy to usable energy operatively associated with the means for extracting momentum-energy.

13. The system of claim **12**, wherein the means for extracting momentum-energy is a substrate having an arm rotatably coupled to the substrate which substrate is confined to move along an oscillatory path.

14. The system of claim **12**, wherein the means for converting the momentum-energy is an induction coil system, heat dissipater, magnetic brake, friction brake, or linear to rotational motion translator.

15. A method of extracting energy from the zero-point field of space, comprising:

a) inducing rotational motion of an arm rotatably mounted to a substrate which is confined to move along an oscillatory path and wherein the rotational motion is sufficient to impart movement of the substrate along the oscillatory path; and

b) converting momentum-energy of the substrate to usable energy.

16. The method of claim **15**, wherein a speed of the rotational motion, relative masses of the arm and the substrate, geometry of the arm, and the geometry of the substrate are designed such that zero-point energy is utilized to maintain the movement of the substrate.

17. The method of claim **15**, wherein the inducing is accomplished by a drive motor.

18. The method of claim **15**, wherein the movement of the substrate follows a linear oscillatory path having two endpoints and the arm follows a cyclic path such that translational motion of the substrate in a y-direction along the oscillatory path substantially coincides with translational motion of the arm in the y-direction.

19. The method of claim **18**, wherein the arm has two theoretical maximum energy transfer positions that occur twice each full cycle of the arm, corresponding to two angles of the arm with respect to an x-axis, which angles occur simultaneously at the time of collision of the substrate with a stationary member.

20. The method of claim **15**, wherein the step of converting is accomplished by inductive coupling of conductive wire and at least one magnet.

21. The method of claim **15**, wherein the usable energy is mechanical, thermal and/or electrical energy.

22. The method of claim **15**, further comprising inducing rotational motion in a secondary arm in an opposite direction to that of the arm such that motion along the oscillatory path of each of the arm and the secondary arm substantially coincide with each other.

23. A method of extracting energy from the zero-point field of space, comprising:

a) inducing multi-mode motion in a primary transducer along an oscillatory path, wherein the multi-mode motion includes a primary translational motion and a secondary motion such that abrupt changes in the primary translational motion are sufficient to impart a momentum influx to the secondary motion; and

b) converting the momentum influx to usable energy.

24. The method of claim **23**, wherein the momentum influx is a result of radiation pressure due to scattering in part of the electro-magnetic momentum-energy contained within the zero-point field of space opposite to centripetal acceleration of the arm.

* * * * *

US 20110073715A1

(19) **United States**

(12) **Patent Application Publication**

Macaulife

(10) **Pub. No.: US 2011/0073715 A1**

(43) **Pub. Date: Mar. 31, 2011**

(54) **METHOD AND APPARATUS FOR GENERATING PROPELLANTLESS PROPULSION**

(76) Inventor: **Neville Macaulife**, Oak View, CA (US)

(21) Appl. No.: **12/586,661**

(22) Filed: **Sep. 25, 2009**

Publication Classification

(51) **Int. Cl.**
B64G 1/40 (2006.01)

(52) **U.S. Cl.** **244/171.5**

(57) **ABSTRACT**

Existing spacecraft accelerate by ejecting propellant, and VTOL craft hover by creating violent downdrafts of air. We describe a propellantless propulsion system that can achieve practical spaceflight and hovering flight by exchanging momentum with normally unobservable quantum particles that pervade all of space.

High-polarizability nanoparticles with electromagnetic resonances in the petahertz range are formed into nanoantennas, and exposed to high fields in the terahertz range. The fields increase the density of naturally occurring vacuum fluctuation (VF) petahertz oscillations in the nanoantennas. The nanoantennas are assembled into arrays, such that their naturally occurring zero-point oscillations are amplified by the terahertz fields, producing enhanced oscillations in VF densities during each petahertz oscillation period that interfere constructively in the intended direction of emission. This causes the VF oscillations to partially convert to travelling waves that emit from the system, creating an imbalance in VF pressure and urging the device to move in the direction of least pressure.

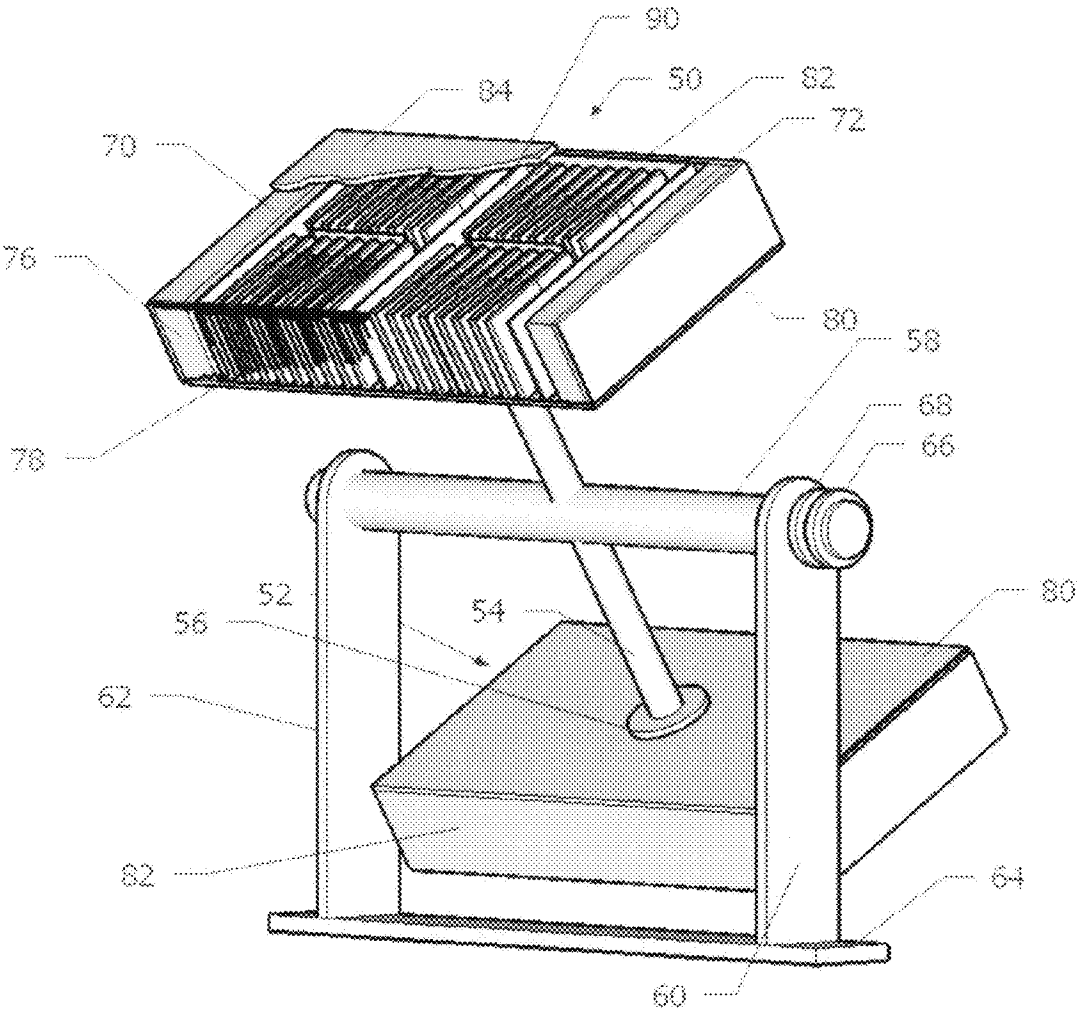

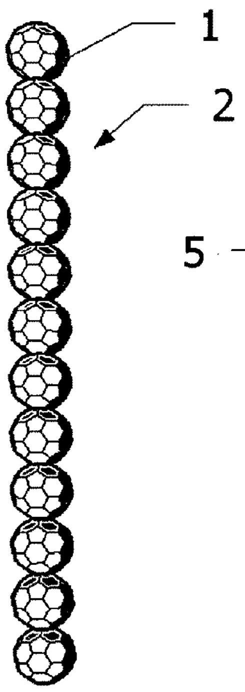

FIG. 1A

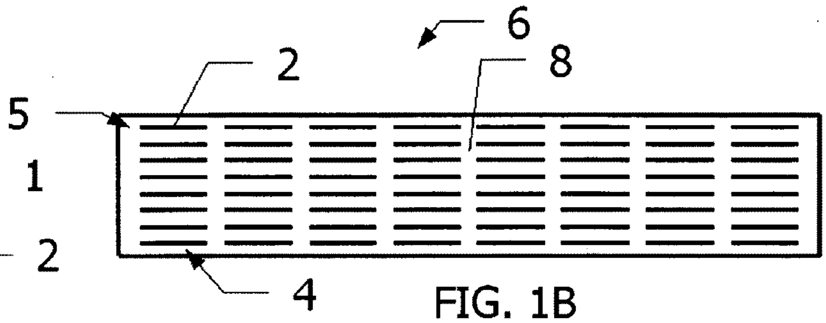

FIG. 1B

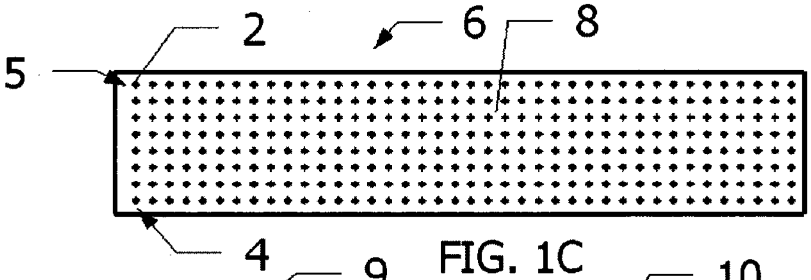

FIG. 1C

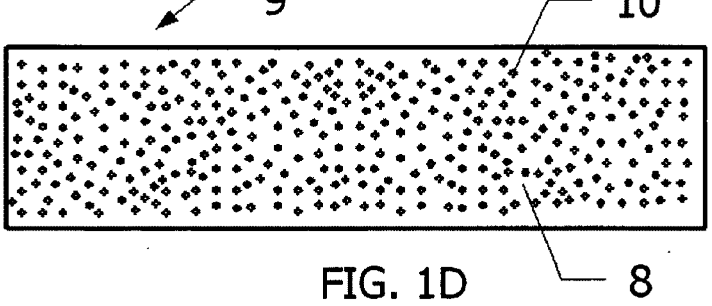

FIG. 1D

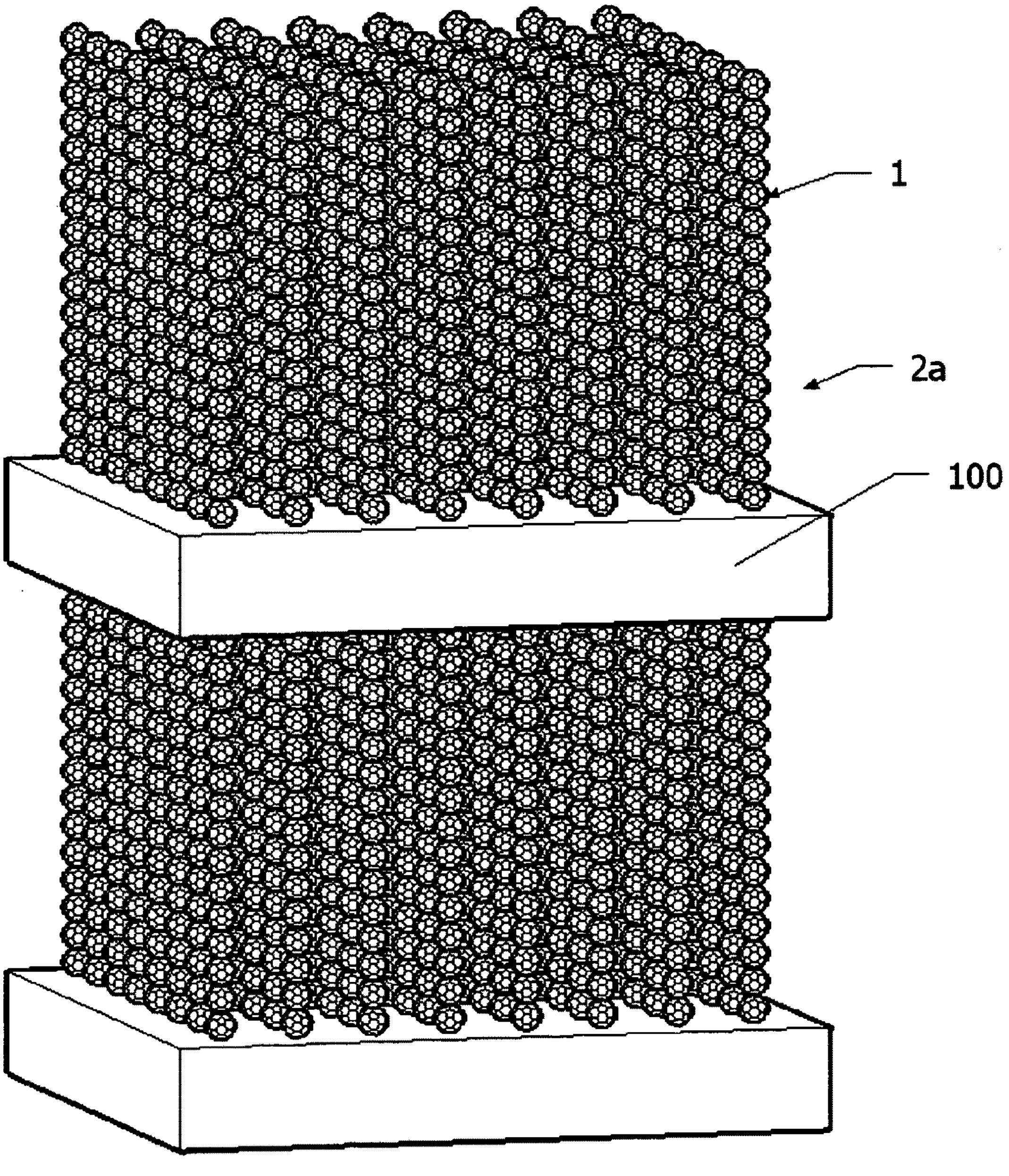

FIG. 1E

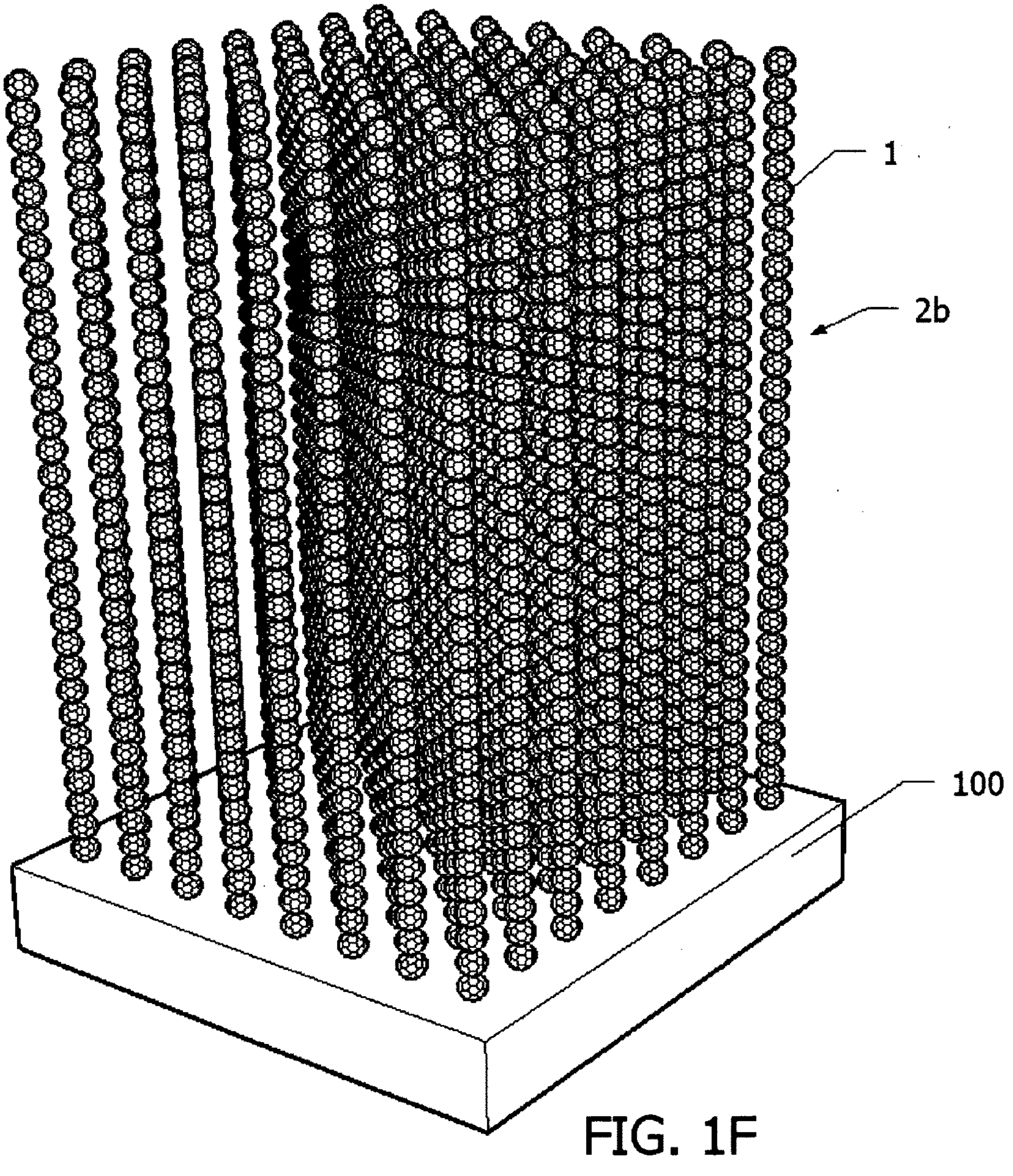

FIG. 1F

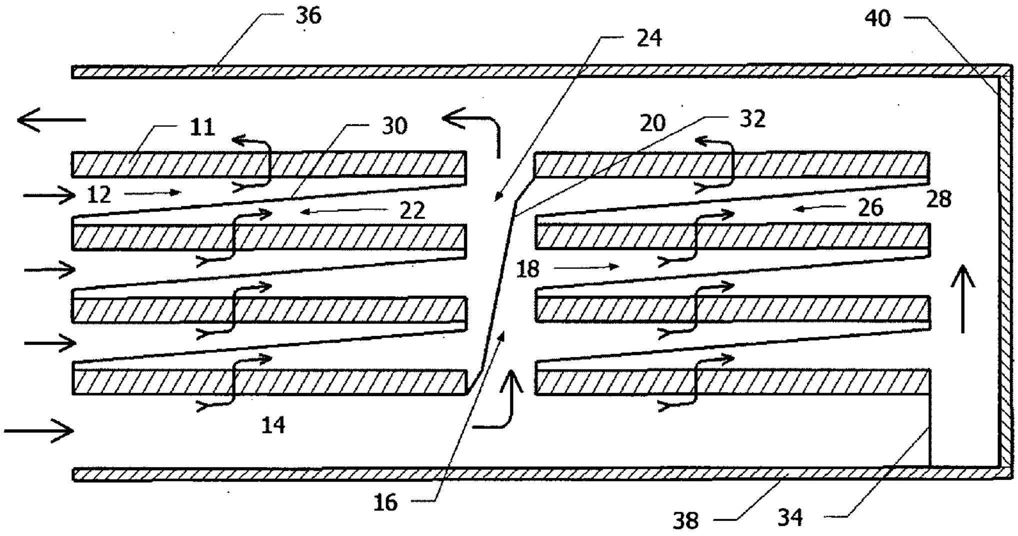
36
40
24
11
30
20
32
12
22
26
28
18
14
16
38
34

FIG. 2

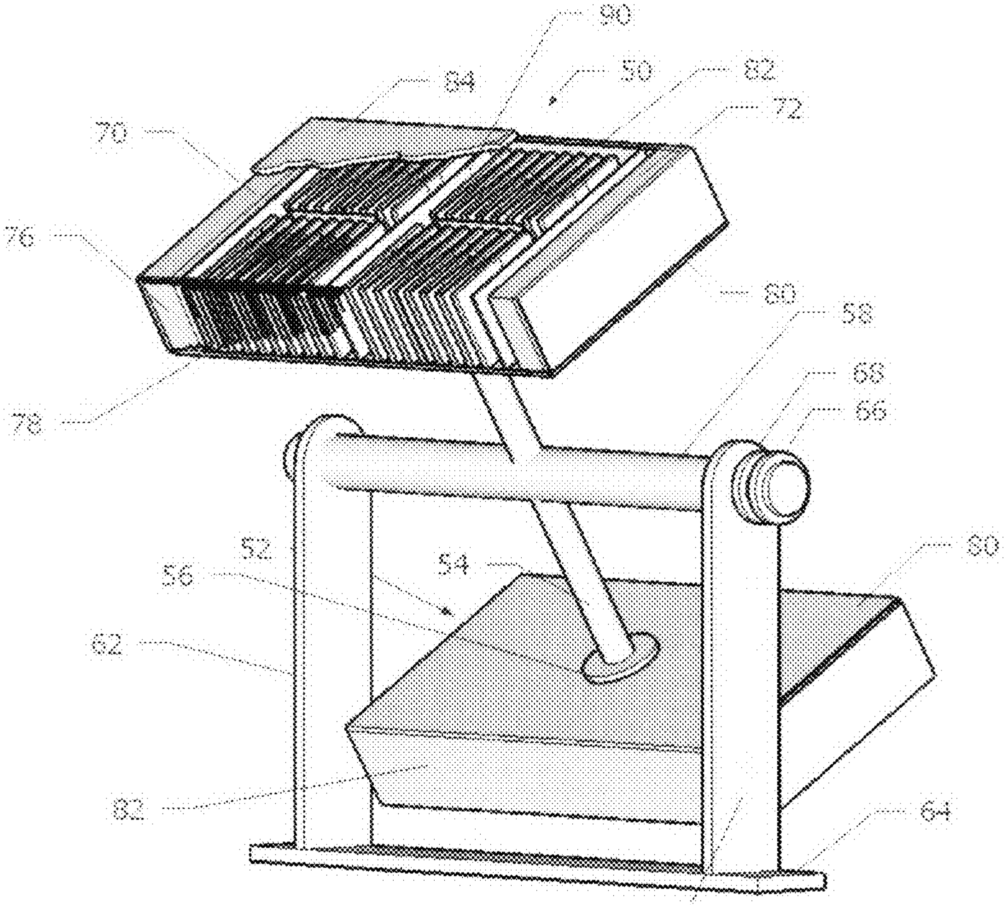
90
84
50
82
72
70
76
80
58
68
66
78
52
54
56
80
62
82
64
60

FIG. 3

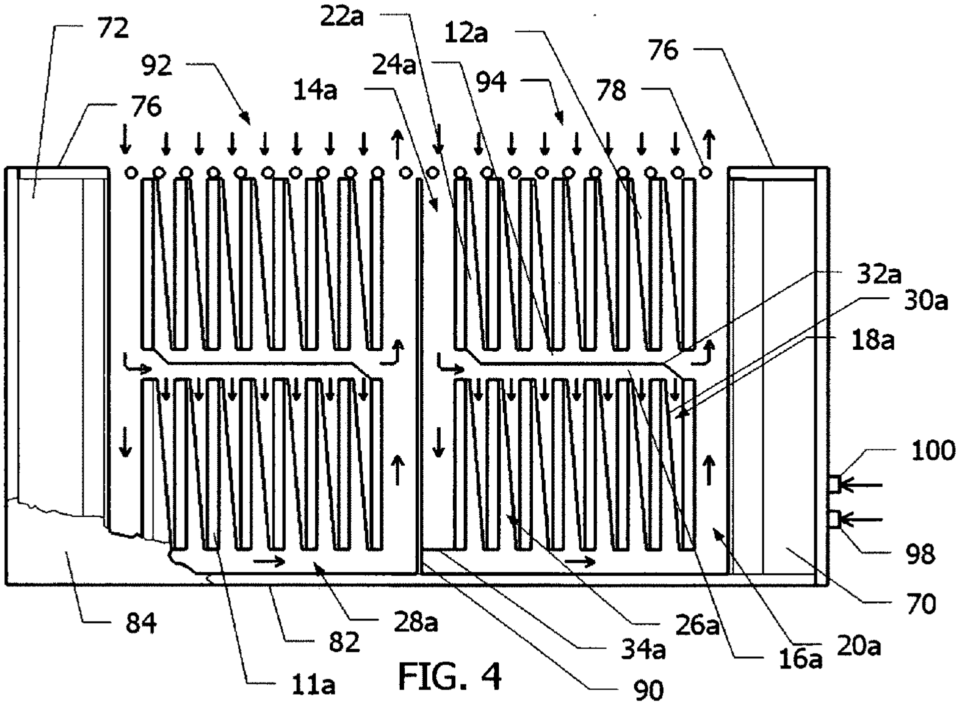

FIG. 4

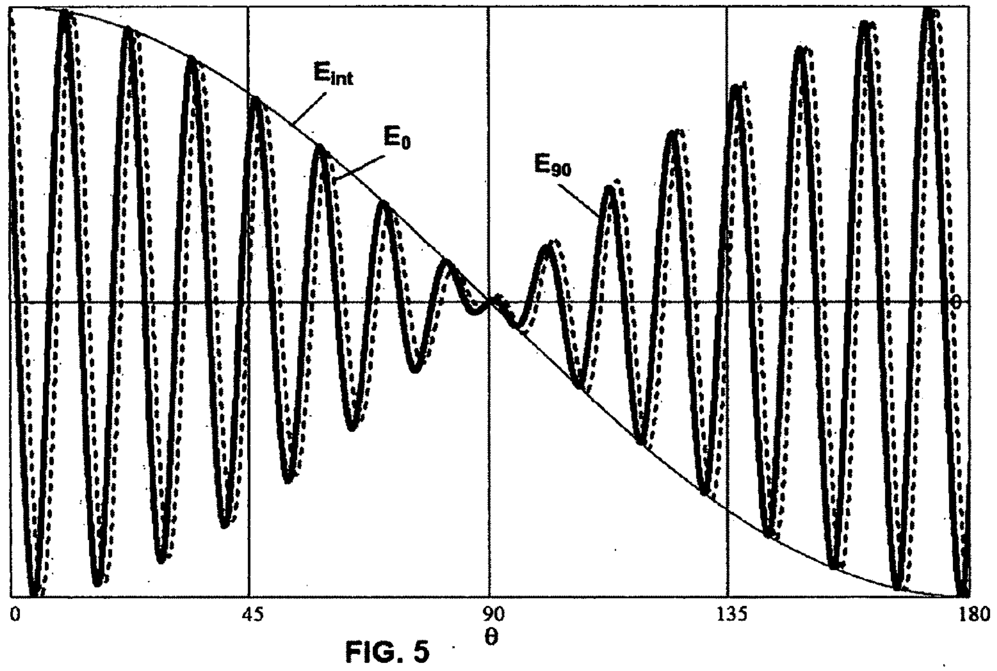

FIG. 5

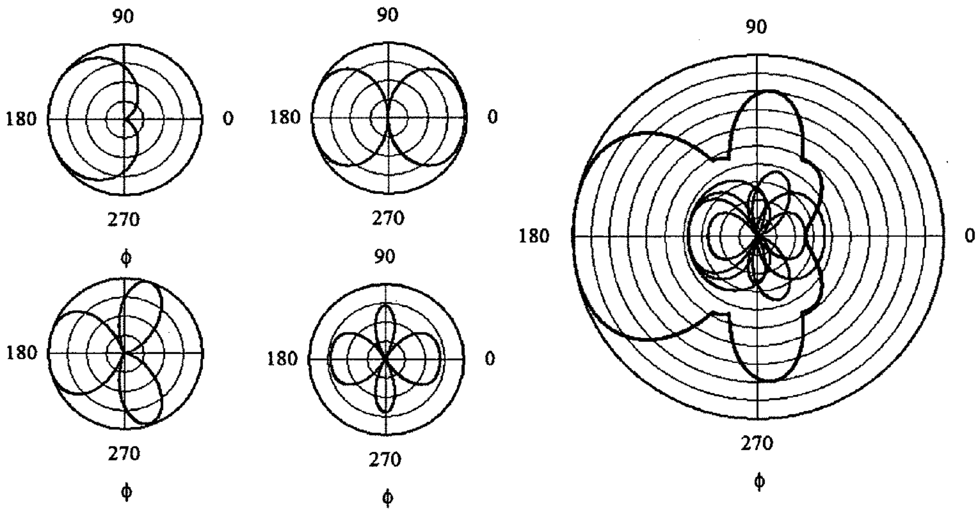

FIG. 6

METHOD AND APPARATUS FOR GENERATING PROPELLANTLESS PROPULSION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] Not Applicable

FEDERALLY SPONSORED RESEARCH

[0002] Not Applicable

SEQUENCE LISTING OR PROGRAM

[0003] Not Applicable

BACKGROUND OF THE INVENTION

[0004] Apparatus is described for achieving propellantless propulsion of space and air vehicles, by causing the naturally occurring radiation pressure (RP) of virtual quantum particles to become greater on one side of the claimed device than the opposite side, thus causing movement in the direction of least pressure. The apparatus may also provide hovering flight of airborne vehicles without requiring violent downdrafts of air, or exert force on various objects as may be useful for industrial processes or meet other needs.

BRIEF SUMMARY OF THE INVENTION

[0005] Hitherto, space vehicles have been propelled by the ejection of propellant; and aircraft, by thrusting against air. The application presents viable alternatives whereby such vehicles may achieve motion by absorbing from free space, amplifying, and re-emitting unidirectionally, the virtual radiation of so-called virtual particles such as vacuum fluctuations (VFs). The claimed means of doing so are as follows:

1. Electromagnetic oscillators generate above zero-point VF densities in high-polarizability nanoparticles such as buckyballs or other fullerenes. The nanoparticles being formed into nanoantennas, either singly, or in combination, by, for example, joining about twelve C3840 buckyballs to form full-wave linear petahertz nanoantennas.
2. A numerically large number of such nanoantennas are formed into 2D arrays, which are then stacked to form 3D arrays.
3. The nanoantennas are so spaced apart as to oscillate in a phase relationship such that they interfere constructively in a given direction here termed the forward direction, and destructively to the rear and sides.
4. This transforms the forward-directed VF oscillations from standing waves to travelling waves, which then move or shunt along the nanoantennas, combining their amplitudes as they do so, until they emit from the final tier of nanoantennas.
5. The VF flow creates a below zero-point VF pressure in the rearward nanoantennas, causing a replenishing flow of VF waves to be drawn into the device from the rearward free space surroundings, thus allowing the forward flow of VFs through the device to continue for as long as the device is appropriately energized.
6. The device thus produces thrust via the generation of radiation pressure somewhat analogously to a photon rocket. However, photons require hv joules to create them, but only supply hv/c newtons of radiation pressure, where h is Planck's constant, v the photon frequency, and c the speed of light in free space. This means that the thrust to power ratio is an abysmal 1/c newtons per joule at best. It is hard to imagine anything less efficient, for even moving a single skateboard would require the entire output of numerous power stations to be converted into photons! By contrast, VFs are freely provided by nature, and, in principle, require no unrecoverable energy to create them except in so far as their creation generates pressure. It follows that the propulsive efficiency of a VF radiating device can, in principle, approach a limit of one newton of thrust per joule.

[0006] We wish to point out that while we have described VFs as travelling through space much as do photons, we consider this to be only a mental image, since the behavior of VFs in free space is an unobservable; it being only their interactions with matter that are capable of direct detection and measurement.

[0007] What is known, and what is claimed here as an operating principle of the invention, is that VFs possess a momentum of ½hv/c, and their unidirectional emission from the forward antennas in the claimed device must create a reactive force acting on the device, urging it rearward.

[0008] It will be shown that the claimed device permits electrical equipment oscillating in, for example, the terahertz frequency range, to generate VF emissions in the petahertz range.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] FIG. **1**A shows twelve high-polarizability buckyballs combined to form a nanoantenna.

[0010] FIG. **1**B shows in side view, sixty-four FIG. **1**A nanoantennas formed into a 3D array or stack, comprising eight vertical columns, each with eight horizontal tiers.

[0011] FIG. **1**C shows the stack of FIG. **1**B in front view.

[0012] FIG. **1**D shows a stack in side-view, in which the nanoantennas are randomly spaced.

[0013] FIG. **1**E shows antennas formed from buckyballs in two 8×8 tiers.

[0014] FIG. **1**F shows an 8×8 matrix of nanoantennas constructed from buckyballs, mounted on a substrate to form long-wire antennas.

[0015] FIG. **2** shows eight stacks as FIGS. **1**B & **1**C combined together to form a "thrust generator," by which is meant an assemblage of stacks so assembled, constructed, and energized by electromagnetic devices, as to emit VF radiation unidirectionally.

[0016] FIG. **3** shows schematically how a motor may be constructed by mounting suitably modified FIG. **2** thrust generators around a rotatable shaft, such that their radiative outputs are directed tangentially to the shaft.

[0017] FIG. **4** Shows FIG. **2** thrust generators powered by an electromagnetic terahertz frequency generator.

[0018] FIG. **5** shows an electromagnetic generator output waveform, and the waveforms it generates in two nanoantennas that are oscillating in a particular phase relationship with each other.

[0019] FIG. **6** shows typical polar field diagrams for two antennas separated by a given number of wavelengths. The larger diagram sums their combined output.

REFERENCES

(a) Background References

[0020] 1. Milonni, Peter W, The Quantum Vacuum, an Introduction to Quantum Electrodynamics, Academic, 1994.

[0021] 2. Milonni, P W, Cook, R J; Goggin, M E; Radiation Pressure from the Vacuum: Physical Interpretation of the Casimir Force, Phys. Rev., A38, p1621, 1988.

[0022] 3. Bohm, David, Quantum Theory, 1951, p415; Dover reprint, 1989.

[0023] 4. Bonin, Keith D & Kresin, Vitaly V; Electric-Dipole Polarizabilities of Atoms, Molecules and Clusters, World Scientific, 1997.

[0024] 5. Sopova, V; Ford, L H; Casimir Force between a Small Dielectric Sphere and a Dielectric Wall, arXiv: quant-ph/0406021 v1 1 Jun. 2004.

[0025] 6. Gueorguiev, G. K; Pacheco, J M; and Tomanek, David; Quantum-Size Effects in the polarizability of Carbon tubules, Physical Review Letters, 92, p215501-1 to 4, 2004.

[0026] 7. Casimir, H B G; Polder, D; The *Influence of Retardation on the London-Van der Waals Forces*, Phys. Rev., **73**, p360-372, 1948.

[0027] 8. Power, E A; *Introductory Quantum Electrodynamics*, American Elsevier, 1965.

[0028] 9. Ford, L H; Svaiter, N V; *Amplification of Vacuum Fluctuations*, http://www.physics.umd.edu/rgroups/ep/yskim/boston/ford.pdf.

[0029] 10. Levin, Frank S; Micha, David A; *Long-Range Casimir Forces, Theory and Recent Experiments on Atomic Systems*; Plenum, 1993.

[0030] 11. Power, p136.

[0031] 12. Milonni, p49.

[0032] 13. Lide, David R, Ed., Handbook of Chemistry and Physics, 85th edition, CRC Press, 2004-2005; 10-167-182, 12-17-18.

[0033] 14. Milonni, p2.

[0034] 15. Amusia, M Ya et al, Dramatic distortion of the 4d giant resonance by the C60 fullerene shell. J. Phys. B: At. Mol. Opt. Phys. 38, 2005, p L169-L173.

[0035] 16. Farson, D F; Choi, H W; & Rokhlin, S I; Electric discharges between platinum nanoprobe tips and gold films at nanometre gap lengths;* Nanotechnology 16 (2005) p1-8.

[0036] 17. Patten, Van P G; Noll, J D; Myrick, M L; Spark-gap atomic emission spectroscopy 2. Improvements in resolution; J. Vac. Sci. Technol. B 15(2), March/April 1997.

[0037] 18. Kraus & Markefka, p830-835, and p848-850.

[0038] 19. Karalis, Aristeidis; Joannopoulos, J. D; Soljacic, Marin; Efficient wireless non-radiative mid-range energy transfer. 19 pages. arxiv.org/pdf/physics/0611063v2. Also, for material on evanescent transmission see: Brady, J. J, Brick, R. 0, Pearson, M. D; Penetration of Microwaves into Rare Medium in Total Reflection. Journal Optical Society of America, 50, p1080-1084, 1960. And: Raad, Bechara A; Jacobs, Ira; Tunneling Through a Total Internal Reflection Layer, IEEE Transactions of Education, 35, p112-114, 1992.

[0039] 20. Kraus, John D & Marhefka, Ronald J; Antennas for All Applications; Third Edition, Tata McGraw-Hill, 2003.

[0040] 21. Corso, Martina et al; Boron Nitride Nanomesh, Science, 303, p217, 9th Jan., 2004.

[0041] 22. Kaganovskii, Yu S et al; Langmuir-Blodgett films of the fullerene C60; JETP Lett., 60, No. 5, Sep. 10th, 370-374, 1994.

[0042] 23. Hecht, David, S, et al; Electronic properties of carbon nanotubes/fabric composites; Current Applied Physics, xxx (2005) xxx-xxx. Online at www.sciencedirect.com.

[0043] 24. Maroulis, George; Atoms, Molecules and Clusters in Electric Fields, Imperial College Press, 2006.

[0044] 25. Casimir, H. B. G; Proc. K. Ned. Akad. Wet.; 51, p793, 1948.

[0045] 26. Hushwater, V; Radiation approach to the repulsive Casimir force; arXiv:quant-ph/9909084v1, September, 1999.

[0046] 27. Bertsch, George F et al; Collective Plasmon Excitations in C60 Clusters; Physical Review Letters, 67, No. 19, 4 November, 1991; available online.

[0047] 28. Wikipedia online encyclopedia; Terahertz radiation.

[0048] 29. Ozyuzer, L et al; Emission of Coherent THz Radiation from Superconductors; Science, 318, 23 November, 2007; 1291-1293; available online.

[0049] 30. Bower, Chris et al; A Micromachined Vacuum Triode Using a Carbon Nanotube Cold Cathode; IEEE Transactions on Electron Devices; 49, No. 8, August, 2002; p1478-1483; available online.

[0050] 31. Zhu, Jinfeng et al; Simulation of Field Emission Micro-Triode Based on Coaxial Nanostructure; J Infrared Milli Terahertz Waves (2009) 30:94-102; available online.

[0051] 32. Wang, C. P. Michael and Gates, D. Byron; Directed assembly of nanowires; Materials Today, May, 2009; p34-43.

[0052] 33 Barlow, H. E. M; Bending waveguides—breakthrough in communications; Discovery, August, 1965.

[0053] 34 Scmitt, Ron; Understanding electromagnetic fields and antenna radiation takes (almost) no math; ww.ednmag.com. Milonni, p220.

[0054] 36 Kraus & Markefka; 0548.

(b) General Review References

[0055] Bonin, Keith D & Kresin, Vitaly V; Electric-Dipole Polarizabilities of Atoms, Molecules and Clusters, World Scientific, 1997.

[0056] Kraus, John D & Marhefka, Ronald J; Antennas for All Applications; Third Edition, Tata McGraw-Hill, 2003.

[0057] Levin, Frank S; Micha, David A; Long-Range Casimir Forces, Theory and Recent Experiments on Atomic Systems; Plenum, 1993.

[0058] Maroulis, George; Atoms, Molecules and Clusters in Electric Fields, Imperial College Press, 2006.

[0059] Milonni, Peter W, The Quantum Vacuum, an Introduction to Quantum Electrodynamics, Academic, 1994.

Special Terms Used

[0060] Tier: A 2D nanoantenna array.

Stack: A 3D nanoantenna array.

Thrust generator or TG: A device in which a number of stacks are assembled such that they combine their radiative outputs.

VFs: Vacuum fluctuations.

Forward direction: The direction in which the radiative output of the stacks combines constructively. Opposite: Rearward direction, associated with destructive interference, or redirection of output.

DETAILED DESCRIPTION OF THE INVENTION

Introduction

[0061] It is well established that quantum entities known as vacuum fluctuations (VFs), pervade all of free space at a mean energy density known as the zero-point. The VFs are, in fact, virtual photons, the word virtual denoting that their transitions do not conserve energy, and thus render their energy inaccessible [1]. This assumption of inaccessibility is being challenged by some ingenious proposals, however, what concerns us here is the known fact that VFs possess momentum which is constantly traded back-and-forth with real-world particles [1, 2, 5, 8-10, 12, 26]. This indicates that one-waying VFs in guiding structures may produce unidirectional propulsive effects.

[0062] It may seem unphysical to suggest that VF momentum can be accessible while VF energy is not. However, while the energy is held to be inaccessible because the vacuum has no lower state the VFs can fall into (i.e, there can be no change of state), linear momentum is equivalent to a harmonic oscillation of zero hertz and is not comparably restricted.

[0063] One might also note that the term virtual does not mean unreal! In fact, so-called virtual transitions have very real effects. To name but one example, the Van der Waals attractions between molecules—the very glue that holds our entire world together—arises from virtual transitions [3].

[0064] We require no weird physics to explain the operating principles of the claimed devices. We first consider the known fact that the relationship for the interaction of an induced dipole with an electric field is provided by $E_\alpha=(-\frac{1}{2}\alpha E^2)$, where

E_α is the potential energy in joules, E is the electrical field strength in volts per meter, and α the electrical polarizability of the dipole in Cm^2/V [1, 4, 5, 9-11]. The units are SI, and may be applicable to both to static and dynamic polarizabilities [5]. The minus sign (discarded from here on), indicates the resulting force is attractive.

[0065] What is particularly relevant for our purposes is that this term provides the potential energy that appears between a polarizable particle and a conductive wall due to the action of VFs alone [1, 7-11]. It can also produce the force between two polarizable particles that is likewise due to VFs, as will be developed below.

[0066] These matters suggested to the applicant that if giant-polarizability nanoparticles were formed into 3D nanoantenna arrays, and their E_α energy components cycled up and down by externally applied fields, this might produce rapid oscillations of their associated VF field densities, causing the nanoantennas to emit VFs unidirectionally, in rather the same way that conventional antenna arrays [20] emit radar and signal-carrying beams when driven in phase-quadrature.

[0067] Before proceeding with practical details, we address the following theoretical concerns.

The Reality of Vacuum Fluctuation Pressure and Momentum Exchange.

[0068] It is a known fact that VFs possess a momentum of ½hv/c newton per fluctuation that can exert pressure on real-world particles [1, 2, 10, 14, 15, 25, 26]. Further, VFs are widely believed to exist independently of matter [1, 2, 9, 11, 12, 26], and thus exert a pressure on any object reflecting a VF field. This pressure equals 2u for incident radiation reflected from a plane surface on one side, where u is the free-space or zero-point energy density. The pressure reduces to u if the radiation is evenly divided between both sides of the reflector, and equals u/3 for an isotropic distribution, in which case both sides are referred to by definition [2, 12, 14].

[0069] It follows that if any means is found to create a greater VF pressure one side of an object than the other, the object will be urged to move in the direction of least pressure. This is the simple principle underlying the proposed propulsive apparatus.

[0070] Some analysts have a reluctance to accept—despite the overwhelming number of papers that now routinely define vacuum quantum electrodynamics in terms of vacuum fluctuation theory—that fluctuating vacuum fields exist in the absence of physical sources. However, Peter Milonni, while discussing laser linewidths in his definitive book *The Quantum Vacuum [*1], refers to "the external vacuum field" transmitting through a laser mirror, undergoing amplification as it propagates to the second mirror, and then being reflected (p194). He then writes:

> **[0071]** "It is worth noting again that our calculation of the Schawlow-Townes linewidth assumes that the vacuum field is transmitted, reflected, and amplified in the same way as fields of "real" photons, i.e., fields defined in terms of excited states of the field. The transmission and reflection properties of the vacuum field are consistent with the fact (see {P.M's} Sections 2.4 and 6.2) that the spatial variations of the quantized field are the same as those of the classical field subject to the same boundary conditions. The fact that the vacuum field may be amplified follows from the fact that its contribution to the laser linewidth is more or less interchangeable with the contribution from spontaneous emission: If spontaneously emitted radiation inside the cavity is amplified by the gain medium, then so too must the vacuum field entering the cavity." (p197-198.)

[0072] It may be particularly noted, that Milonni describes the VFs as being amplified, reflected, and transmitted without undergoing any conversion into real photons. This directly contravenes the expectations of some pundits that VFs transmitting from such apparatus would so convert into photons.

[0073] Another commonly raised objection to the idea that VFs possess radiation pressure independently of matter, is that the Casimir effect can be negative.

[0074] For those unfamiliar with the subject matter, we explain that the Casimir effect or force is most typically demonstrated when two metal plates are separated from each other by a submicron distance. When this is done, a small attractive force is observed, the reason being, generally speaking, is that there are less VF modes between the plates urging them apart, than there are VF modes outside the plates urging them together. Thus the latter VF modes produce the greater radiation pressure, resulting in a net inward force.

[0075] However, there are a number of geometries and configurations in which the pressure in small gaps exceeds the external pressure. This has led to the dubious argument that the pressure is a special case of the van der Waals molecular forces. But lately Hushwater [27], has demonstrated that, in such cases, the negative pressure is due to there being a greater number of VF modes within the gap than without.

[0076] Lifshitz has also shown, in a seminal paper, that intermolecular forces may cease to be additive when the dielectric constants are high [35].

[0077] We therefore claim the VFs will exert pressure on the claimed devices. By way of example, we look ahead to the device shown in FIG. **2**. It is not necessary to understand how it works at this time, but merely accept that the arrows show the flow of VFs through the apparatus; and that the apparatus will amplify or pressurize the VFs when appropriately energized. We assume isotropic VF distributions, and consider only horizontal pressures since all others will balance out. We determine the forces acting on the device as follows:

[0078] (1). If VFs enter the device from the left as shown, pass through the stacks as **10**, and exit back to the left without amplification or loss, they will produce a pressure of u/3 newton per square meter urging the device to the right. This pressure will, of course, equal the zero-point pressure of free space.

[0079] (2). Some VFs that would otherwise enter the device are scattered back by the left-facing edges of cover plates **42** & **38**, the four leftmost stacks as **10**, and the stub-like leftmost vertical sections of the three leftmost partitions as **30**. These scatterings will exert an ambient pressure of u/3 N/m^2 urging the device to the right.

[0080] (3). VFs approaching from the right, and encountering the right-hand cover plate **40** will exert the same pressure of u/3 urging the device to the left.

[0081] (4). The above pressures clearly sum to zero because the left-hand and right-hand areas are identical. Thus, the net pressure=(1)+(2)−(3)=0.

[0082] (5). However, if the VFs entering the device as (1) are amplified by their passage through the stacks as **10**, their pressure will clearly rise above ambient or zero-point levels. For example, if pressure (1) is doubled by amplification, the resultant net pressure will be approximately 2×(1)+(2)−(3)=(1)≅u/3 N/m^2, urging the device to the right. The "nearly equals" is inserted because the emitted distribution will no longer be isotropic.

A Small Sample of Relevant Patents Issued by the US Patent Office

[0083] The US Patent Office has granted a number of patents to inventors attempting to harness VFs or zero-point energy to create new energy sources and the like. The applicant does not consider any to anticipate his own claims, but offers the following as a small sample.

1. In 1996 Dr. Frank Meade was granted U.S. Pat. No. 5,590,031 which claimed two spherical antennas resonating passively at different frequencies, that gathered VF radiation from free space, thereby generating a VF radiation beat-frequency to be supplied to electrical devices and perform work. Thus the USPO recognized the possibility that VF radiation exists freely in space and that it may be harnessed for useful purposes, even if there is an appearance of building a "free energy" device. A matter that does not concern us in the present application.

2. In 1987 Federico Capasso was granted U.S. Pat. No. 4,704,622 which claimed devices that could harness VFs to enhance the tunneling abilities of electrons in solid-state devices utilizing quantum wells. Although the present application makes no claims involving VFs in such electronic devices, which operate by electron-hole formations, such a VF aid to tunneling is similar in principle to the present inventor's claim that VFs can shift progressively along chained nanoantennas, especially those that lie within each others near-fields.

3. In 1977 Grigg was granted U.S. Pat. No. 4,004,210 for devices that utilized zpe pressure to drive fluctuations through electronic devices in a cold to hot direction. He insisted for many years that such a thing need not violate the second law, and he may have gained some support, since a multi-national company is reported to have a "nanorectifier" program which at least sniffs at such feasibilities. Again, the US Patent Office has recognized the possibility that VF radiation pressure may perform work, and we note that accepting that VF pressure may drive energy from cold to hot is a far greater leap than to accept that appropriate devices can simply transform VF oscillations from standing waves to travelling waves and emit from the claimed apparatus. However, it is to be understood that the present inventor makes no claim for any device that violates physical law.

4. In 2002 F. Pinto was granted U.S. Pat. No. 6,477,028 which claimed apparatus for storing "Casimir force energy" (i.e, VF radiation) by, in one case, linking two "Casimir force generating boundaries" while varying the physical parameter of the environment separating them. This is another example of the US Patent Office accepting the possibility that VF radiation can be treated as a working medium for producing useful effects.

Giant Polarizability

[0084] Returning to the earlier discussion regarding the construction of nanoantennas using materials possessing large electrical polarizabilities, we find that carbon fullerenes such as buckyballs [6], and materials such as $Na_{14}F_{13}$ possess so-called giant polarizabilities that can be some three orders of magnitude greater than the run of the mill values customarily found in handbooks [13]. We note particularly that Gueorguiev et al [6], have calculated that a 3840-atom buckyball has what they term an α_{LCAO} figure (based on calculations involving linear combinations of atomic orbitals), of 33,731 $Å^3$, or 3.75×10^{-36} Cm^2/V in the SI system. Therefore, by way of example, we consider the use of these buckyballs to construct the nanoantennas.

The Antenna Stacks and their Construction

[0085] FIG. **1**A shows twelve buckyballs as **1** mounted together to construct a linear nanoantenna **2**. FIG. **1**B shows, in front view, sixty-four such nanoantennas as **2** formed into eight columns as **4**, each with eight horizontal (2D) tiers as **5**, to create a 3D nanoantenna array or stack **6**. The antennas are embedded in a dielectric or aerogel substance **8**.

[0086] FIG. **1**C shows stack **6** in side view, revealing forty of the previous eight-tier columns as **4**. It is envisaged that such stacks might consist of about 26 tiers as **5** each consisting of ~10^{16} antennas per square meter, all emitting broadside in the preferred case (i.e, vertically in FIGS. **1**B & **1**C), and embedded in a dielectric or aerogel substance **8**.

[0087] FIG. **1**D shows a stack **9** in which can be seen about sixty spherical nanoantennas as **10** spaced randomly, and embedded in a dielectric or aerogel substance **8**. It will be described later.

[0088] FIG. **1**E shows 128 antennas as **2*a*** formed from buckyballs as **1**, in two 8×8 tiers which may be stacked—as many as required—between substrates as **100**, two of which are shown. An advantage of such an assembly is that there is no requirement to embed the antennas in a dielectric. Such nano-structures can be caused to self-assemble, as is explained below.

[0089] FIG. **1**F shows an 8×8 matrix of nanoantennas as **2***b*, formed from buckyballs as **1**, mounted on substrate **100**, in extended lengths that may act operatively as what are known to those versed in the arts, as long-wire antennas. Thinner versions of substrate **100** may be used to brace longer antennas at regular intervals.

[0090] Buckyballs may be readily formed into a well-ordered 2D array by attaching them to a nanomesh. One such suitable nanomesh consists of a single layer of boron nitride, another is graphene. A boron nitride mesh can be formed by self-assembly if an operator exposes a rhodium or ruthenium surface to borazene at high temperatures in an ultra-high vacuum [21]. Such nanomeshes have a remarkable ability to trap fullerenes, and also other particles such as molecules, nanoparticles, and metallic clusters, in their pore matrix, such that each pore seats one buckyball.

[0091] A nanomesh described in the above reference had a periodicity of 3 nanometers, and a 2 nanometer hole diameter. However, similar nanomeshes can now be tailored to meet a wide range of geometric, and also electrical specifications, such as whether conductive or insulating parameters are required.

[0092] A second method by which an operator may build a 2D buckyball array on a nanomesh is by spraying [23]. For example, purified arc buckyballs such as may be purchased from Carbon Solutions, California, USA, can be added to an aqueous solution of 1% sodium dodecyl sulfate (SDS), at a concentration of 0.02 to 0.05 mg buckyballs per ml of solution. The resulting suspension is mixed by sonication for one hour, and then sprayed onto the nanomesh in a fine mist by means of a Paasche airbrush. It is necessary to heat the nanomesh to 100 degrees centigrade to prevent the formation of large droplets. It is then soaked in water to remove residual SDS.

[0093] An operator may also produce 2D buckyball arrays by dissolving the buckyballs in benzene, and depositing them on the surface of doubly-distilled water [22]. The solvent is allowed to evaporate, leaving behind a self-assembled film of buckyballs. The film is then compressed by a barrier that is moved at microscopic speed, applying a pressure of about 0.02 N/m to the film as it does so. The film can then be transferred to a substrate using the Schaefer method, in which the film is moved horizontally by contact with a hydrophobic plate.

[0094] It will be apparent that such 2D buckyball arrays may be readily mounted one above another to form the stacks required for the present application. This requires that the spacing between each 2D array be such as may best develop the electrical phase relationships to be described later. This may be accomplished by using much the same airbrush spraying technique as revealed above to coat the buckyball arrays with a dilute solution cadmium telluride, using such amount as will produce a dielectric film of the desired thickness once the solvent has evaporated. A second 2D buckyball array may then be placed on the film, and the above operations repeated until the desired number of tiers has been formed.

[0095] Although FIG. **1**A shows the use of only a single column of buckyballs to construct an antenna, it will be understood that multiple parallel columns may be used to provide a thicker antenna, thereby, for example, increasing the bandwidth. Alternately, multiple buckyballs or other fullerenes might be used singly, or clustered to form spherical antennas. Thereby, for example, abolishing any need for the axial alignment that non-spherical antennas may require. Such particles may also operate as point-source antennas.

[0096] Buckyballs link together naturally to form chains as FIG. **1**A when formed into a suspension in suitable solvents. Droplets of such a suspension may then be formed onto a substrate, subjected to DNA hybridization, and finally immobilized with random orientations either on or between Au electrodes. The nano-chains may then be straightened naturally by the shear-forces generated by an evaporating droplet [32].

[0097] Alternately, the substrate may be etched with linear cavities of suitable length and width to trap buckyballs and form them into antennas as FIG. **1**A. The cavities being spaced apart lengthwise and sidewise as necessary to form tiers as **5** FIG. **1**B & **1**C. The substrate may then be rinsed clean of all other fullerene material. At which time a similar substrate may be mounted atop it, and such stacking of substrates continued until a 3D stack of suitable tier count has been assembled, as described above.

The Thrust Generator

[0098] FIG. **2** shows in schematic, seriously not to scale side-view, the mid-cross-section of a device we term a thrust generator (TG). The figure shows eight stacks as **11**. The arrows show paths taken by the VFs as they flow through the TG. It can be seen that they appear from the left, and flow into the stacks either through three input channels as **12** or lower channel **14**, vertical channel **16**, and three input channels as **18**.

[0099] After passing through the stacks and becoming pressurized (powering the device is described in the next section), the VFs enter horizontal channel **20** at the top of the figure, either directly, or via the three horizontal output channels as **22**, and the vertical channel **24**, or alternately via three output channels as **26**, and vertical channel **28**. They then exit the device to the left.

[0100] Six slanted partitions as **30**, slanted vertical partition **32**, and vertical partition **34** prevent the VF pressure from back-flowing to the input regions after emerging from the stacks. (Note that the stacks are one-way devices, and cannot themselves backflow).

[0101] The partitions also prevent the VFs emerging from a lower stack from passing into the one above—a matter that would reduce thrust generation efficiency, as is explained later.

[0102] The top, bottom, and rear covers **36**, **38**, & **40** further prevent VFs from leaking in or out of the system. It will be understood that there are also two side covers preventing any sideways leakage, both covers being in sealing contact with the full length of covers **36**, **38**, & **40**. One cover faces the viewer, but cannot be shown since the view is cross-sectional, the other is at the rear of the drawing, but is omitted for clarity.

[0103] The various partitions and cover plates comprise jointly the first and second radiation guiding structures mentioned in the Claims.

[0104] The TG can also generate thrust if the VFs passing through the stacks flow straight through the device from left to right instead of entering from the left and reversing back upon themselves as shown. This can be arranged by removing right-hand cover-plate **40**, and placing a partition similar to partition **34** at the left opening of topmost channel **20** to seal it off.

[0105] It will be apparent that the permanent direction of thrust can be reversed simply by rotating the device horizontally until the emission is to the right instead of to the left as shown in the drawing. To ensure balanced thrust during the transition, the thrust might be generated by an even number of identical TGs, half of which rotate clockwise, and the other half, anti-clockwise.

The Terahertz Generator

[0106] For the purposes of example and calculation, the TGs are here described as being powered by rapid oscillation of external fields in the form of THz radiation. Lower frequencies will not be efficient, and higher infrared frequencies may produce heating problems. However, the generation of THz radiation is a comparatively new field, and although there are a number of somewhat standardized systems to choose from that have undergone significant development, non of them seem very satisfactory. These systems include gyrotrons; backward wave oscillators; far-infrared, quantum, and free-electron lasers; and synchrotrons [28].

[0107] Unfortunately, those that can develop significant power, such as gyrotrons, are inefficient and bulky, while others are currently limited to milliwatt outputs, and/or can only operate at 50K or worse. All emit only narrow beams.

[0108] Fortunately, two new choices have now been demonstrated. Both involve the development of nano-emitters formed into 2D matrices with huge packing densities, in rather similar fashion to the nanoantenna tiers **5** of FIGS. **1**B & **1**C, and can potentially produce square-meter beams from flat, pancake-like substrates, if required. The first uses superconducting Josephson junctions, and is able to generate THz radiation [29]. However, the temperature is limited to 50K, and despite the high packing densities, the output power density still appears to be low.

[0109] The second choice, developed at UCSD, involves micromachining nano-vacuum cold triodes and similarly forming high packing-density matrices. Emission power densities exceeding 100 W/cm^2 may be possible at room temperature. The frequency was limited to 10 Ghz in 2002[30]. However, a later paper [31] indicates that frequencies of 10 THz should be possible. We therefore select the UCSD system as our currently preferred THz emitter with which to power the TGs.

Generating Rotary Motion

[0110] It will be apparent that if TGs are mounted on a rotatable shaft they will capable of generating rotary motion around the shaft. FIG. **3** shows a device configured for this purpose: The drawing shows TG assemblies **50** & **52**, each of which contains two pairs of TGs separated by a central partition as **90**. The TGs are geometrically tweaked but conceptually identical versions of that shown in FIG. **2**. They can be seen more clearly in FIG. **4** which is described below. They are butt-joined to the end-faces of rod **54** via flanges as **56**. Rod **54** is centrally secured to rotatable shaft **58**, which is mounted on support arms **60** and **62**, which are attached to base **64**.

[0111] The two slip-rings **66** and **68** deliver electrical energy from a conventional power supply to a THz generator **70** similar to the UCSD device, via conventional brushes and supply-wiring that is not shown. The generator emits THz radiation directly into the TGs, causing them to amplify and unidirectionally emit VF radiation as previously described. The radiation is emitted tangentially to shaft **58**, and thus urges the rotatably mounted TG assemblies **50** & **52** to rotate clockwise (top TG moving away from viewer) in the drawing.

[0112] The residual THz power that exits the TGs to the right in the drawing, passes into structure **72** which is so contrived as to either absorb the radiation and dissipate it as heat, or convert it back to electrical energy and return it to THz generator **70**.

[0113] The VFs enter the TGs from free space through the exposed leftmost face of TG **50** in the drawing, and thus also through the unseen exposed rightmost face of TG **52**. These faces are covered by mesh filters consisting of partially shown frames as **76** and wires as **78** which prevent the escape of THz radiation but freely transmits PHz radiation; it being the case that the mesh filters, in combination with partitions **80** (bottommost for TG assembly **50**), partitions **82** (rearmost for TG assembly **50**), and partitions **84** (topmost for TG assembly **50**), function as a THz waveguide. From which it follows that the filters are so contrived as to present essentially the same reflective characteristics to the THz emissions as do the three partitions.

[0114] Suitable mesh filters can be supplied by Microtech Instruments located in Oregon, USA. If the output of THz generator **70** is polarized either S-wise or L-wise, it can be blocked by evenly spaced microwires with a diameter to spacing ratio of about three or four, having appropriate axial alignment to the THz beam. An unpolarized beam will require a mesh.

[0115] FIG. **4** shows the FIG. **3** TGs in top schematic view. Central partition **90** separates two identical TGs **92** & **94**. The end cross-sections of mesh wires as **78** can be seen, as can the side sections of mesh frame **76**. Two sockets **98** & **100** supply the generator with electrical energy from a conventional power supply, as indicated by the two arrows.

[0116] As stated, the TGs are conceptually identical to that shown in FIG. **2** with minor changes in dimensions. etc. Thus, numbering as in FIG. **2**, there are 32 stacks as **11***a*, with diagonal partitions as **30***a*, and **32***a*, and partitions as **34***a*; the same arrangement of main channels as **14***a*, **16***a*, **20***a*, **24***a*, and **28***a*; and smaller channels as **12***a* and **22***a*.

Buckyball Resonance

[0117] The nanoantennas described earlier may be tuned to have a natural resonance at a frequency of about 2 PHz (10^{15} Hz). This should be perfectly feasible since, for example, C60 buckyballs are known to have a giant resonance (with a surprisingly large bandwidth) at about 2 Phz [15]. Bertsch et al also predict "a giant collective resonance at an unusually high energy of ~20 eV" (about 5 PHz), for C60 buckyball clusters [27]. It is also well known that fullerene resonances can be readily enhanced, or tweaked this way or that, by the insertion of additional elements in the form of one or two atoms [4, 24]. The very intense fields created by the THz generator can amplify these PHz oscillations in accordance with the $\frac{1}{2}\alpha E^2$ relationship described earlier, and thus increase the density of the VF PHz fields associated with the buckyballs, in sync with the field oscillations.

[0118] We assign the generator field intensity E_{gen} a peak value of 3E+5 V/m, this being the most likely level at which polarization saturation will occur in the nanoantennas, i.e, the point at which polarizability ceases to increase very much with increasing field strength. It might be thought that fields this high will cause electrical breakdown in the airgaps in the TGs. However, it has been found that micro and nano airgaps have a very high resistance to breakdown, this being because their short path lengths prevent ion-cascading. Also discharges are limited by work-function type effects [16, 17].

Obtaining Constructive Interference

[0119] Antenna arrays emit unidirectionally because they are (typically) spaced a quarter-wavelength apart along their emission axes (the exact distance is usually tweaked for best results), and so driven that there is a time-lag of 90° between each successive antenna in that direction (i.e, they are driven in phase-quadrature), causing the antenna emissions to interfere constructively, and thus combine in the forward direction or direction of declining phase angle; while interfering destructively in the reverse direction, and very largely so to the sides.

[0120] This can best be seen in FIG. **5** in which E_{int}, denoted by the thin solid line, is that part of the generator voltage E_{gen} which appears directly across the buckyballs which form the nanoantennas. Its amplitude is given by

$$E_{int} := \left(\frac{3\cdot\varepsilon_s}{2\cdot\varepsilon_{bb}+\varepsilon_s}\right)\cdot E_{gen},$$

where $\in_{bb}=4$ is the dielectric constant of the buckyballs, and $\in_S=4$ is that of the stacks, for both the generator input frequency and the full VF bandwidth. It follows that

$$\left(\frac{3\cdot\varepsilon_s}{2\cdot\varepsilon_{bb}+\varepsilon_s}\right) = 1,$$

in which case $E_{int}=E_{gen}=3E+5$ V/m.

[0121] Also in FIG. **5**, waveforms E_0 and E_{90} show the nominal enhancement above zero-point level of the petahertz-frequency voltages induced in the nanoantennas by E_{int}. E_0 is a waveform that is initially in phase with E_{int} and is denoted by the thick solid line, where $E_0=\cos(\theta)\cos(\theta\cdot r)$, and r is the ratio of the generator stack VF wavelength $\lambda_{ip.s}$ to the average VF stack wavelength $\lambda_{av.s}$, where $\lambda_{ip.s}/\lambda_{av.s}=30$.

[0122] E_{90}, denoted by the broken line, is a waveform that leads E_0 by 90 degrees due to the phase-quad spacing of the nanoantennas, and is thus given by $E_{90}=\cos(\theta)\cos(\theta\cdot r+\pi/4)$.

[0123] In accordance with the principle of pattern multiplication, E_0 and E_{90} flow through any given stack in the form of travelling waves. They combine their amplitudes, the phase difference between them disappearing in the forward direction of flow. It can thus be seen that the stacks behave somewhat as do amplitude modulators, with E_{int} simulating a carrier wave, while the induced VF waves behave analogously to signal waves.

[0124] For the purpose of example we set the tier-to-tier spacing between the antennas (center-to-center) at $0.8\lambda_{av.s}$. This rather large spacing is necessitated by the fact that the center-to-center distance between two C3840 buckyballs placed in contact is about $0.65\lambda_{av.s}$

[0125] When antennas are stacked this close together relative to their diameter, those skilled in the arts usually determine the exact distance of separation required to produce constructive interference in the forward direction by simple experimentation. It is likely that the best experimentally determined figure will not differ significantly from that assumed above [36].

[0126] Short gap lengths have the advantage that all antenna-antenna transitions take place within the reactive near-field distance in which—by definition—the transitioning particles remain in the virtual state [18, 19, 34]. The nominal reactive near-field distance of full-wave antennas is $\lambda/2\pi$, so the combined overlap distance of two such antennas$=\lambda/\pi=0.318\lambda$. We note that this is about twice the edge-to-edge gap length between the antennas, as developed above.

Calculations

[0127] We set the maximum VF frequency v2 that we propose to transmit unidirectionally through the stacks at 10 PHz. This lies slightly above the ultraviolet transmission fade of lithium fluoride, which we consider to be a suitable dielectric for the stacks; but since the thickness of the stacks is only about 170 nm, the fade should not be a problem.

[0128] We assume a modest VF bandwidth=(v2−v1)/½(v2+v1)=0.2, where v1 is the minimum frequency and so equals 8.2 PHz. The average frequency v_{av} thus equals 9.1 PHz, corresponding to a free-space wavelength $\lambda_{av.o}=33$ nm. We have assumed that the permittivity of the stacks $\in_S=4$, we further assume it is lossless, in which case the v_{av} wavelength in the stacks$=\lambda_{av.s}=(\lambda_{av.o}/\sqrt{\in_s})=16.5$ nm. This figure provides the principle measuring stick for the stacks. We now set the generator frequency v_g at 3 THz, which therefore has a stack wavelength$=\lambda_{ip.s}=(c/v_g\sqrt{\in_S})=500$ nm.

[0129] Recalling that, as mentioned above, the average THz to PHz wavelength ratio=30, it is apparent that the phase difference between any two neighboring nanoantennas should best be considered only an average figure, and that a certain variance in individual phase measurements is to be expected for any given antenna pair.

[0130] However, this is not a serious matter: FIG. **6** shows that unidirectional beam-forming of antenna pairs need not depend critically on the antenna phasing having any particular precise value. Here polar graphs labeled ¼, ½, ¾, & 1 show field amplitudes of two half-wave antennas separated by the same numbers of wavelengths, and driven respectively with phase difference of 90, 180, 270, and 360 degrees. These fields are combined additively in larger graph Σ, which shows the combined fields to be stronger to the left of the diagram than the right by a factor of about 2:1. It follows that reasonable variances in antenna phasing will not seriously affect the beaming efficiency of the TGs. Note that when large numbers of antennas are stacked together, as in the TGs, the rightward and side emissions shown in the above diagrams will be almost zero, since the diagrams are correct for only two antennas.

[0131] It should be particularly noted that the graphs—particularly graph Σ—indicate that if antennas are randomly spaced in a stack as FIG. **1**D, they will still generate significant unidirectional emissions, as long as their maximum spacing is limited to about one wavelength.

[0132] This indicates that stacks suitable for use in TGs might be formed, by, for example, mixing buckyballs into a solvent into which a dielectric has been dissolved, and spraying the solution onto a substrate with an airbrush as described earlier, but this time using a buckyball concentration sufficient to produce an average spacing within the sprayed dielectric of about a half wavelength, once the solvent has evaporated. The amount of spray used should be sufficient to build a stack of whatever final thickness is desired, since it will not be formed by layering as before. The FIG. **1**D stack, formed in this way, should emit radiation rather as shown in the FIG. **6** Σ combined-fields figure, excepting that, as stated, there will be no emission in the rearward direction, and very little to the sides.

[0133] We can now calculate the PHz emission rates: We have mentioned that when E_0 and E_{90} flow through a stack in the form of travelling waves, they combine their amplitudes. It follows that since there are 26 nanoantennas in each stack, the nominal total forward flow=26 E_0 (or for that matter, =26 E_{90}). which, in either case, equals the peak value of E_{int}.

[0134] We therefore define the average nominal energy flow along any single column of antennas in any given stack as

$$\varepsilon_{an.av} = \frac{1}{2}\alpha_{C3840} \cdot 26 \cdot \left(\frac{E_{int.pk}}{\sqrt{2}}\right)_{17}^{216} = 2.2E\text{-}24 \text{ J},$$

where $E_{int.pk}$ is, of course, the peak value of E_{int}. Losses will be dealt with later.

[0135] We assume that this virtual energy is emitted in the form of VFs. Therefore, to obtain the total nominal virtual energy output E_{st} per second of a one-square meter stack, and thus the total power of the emitted VFs, we multiply $\in_{an.av}$ by the number of antennas per tier=$n_{a.tr}$=1.47E+16, and the generator frequency $v._{av}$. Thus:

$E_{st}=\in_{an.t}\times n_{a.tr}\times v_{av}=2.9E+8$ Watt per stack per square meter.

[0136] This energy is virtual, and, of course, has no particular observable effects in and of itself. However, since it is the energy of the emitted VFs (½hv per fluctuation), we have only to divide by c to obtain the radiation pressure (hv/2c per fluctuation), which will have real effects as explained earlier. Thus:

[0137] $p_{rp}=E_{st}/c=0.98$ newton per square-meter stack. We now divide by the constant of gravitation to get kilograms:

[0138] $p_{rp}/g=0.1$ kg per square-meter stack, disregarding losses, which will be discussed in the next section.

[0139] If this figure seems high, we should bear in mind that it is produced by ~$\mathbf{10}^{17}$ nanoantennas possessing giant polarizabilities, driven at near breakdown-limited field strengths, and operating at unprecedented THz frequencies.

[0140] Finally, we estimate the thrust output for a really large TG: And since, as already stated, the thickness of a 26 tier stack is only about 170 nm, we can readily pack about 10^6 square-meter stacks one above another, to form a one-cubic meter TG. The thrust output of such a structure=p_{m3}=0.1×10^6=10^5 kg.

[0141] It is instructive to compare p_{kg} to the pressure p_{freeS} produced by free space VF densities acting on one side of a one square-meter mirror. It is determined by first calculating the energy density of free space E_{freeS} [22]:

$$E_{FreeS} = \left[\frac{h'}{8 \cdot \pi^2 \cdot c^3} \cdot (\omega_2^4 - \omega_1^4)\right] = 4.26E + 5 \text{ J/m}^3,$$

from which the pressure, as described earlier:

$$F_{FreeS} = \left(\frac{1}{3} \cdot \frac{E_{FreeS}}{g_m}\right) = 14490 \text{ kg/m}^2,$$

[0142] where ω2 & ω1 are v2 and v1 expressed in radians, h'=h/2π is the gravitational constant, and the 1/3 is an adjustment for isotropic distribution. This frequency range lies beyond the opacity region of many materials, but is, of course, fully balanced by the internal structural strengths of materials with higher frequency absorptions—the structural strengths being themselves due to internal VFs and other virtual photons.

[0143] We might note that the ratio of stack radiation pressure p_{ip} to p_{freeS} is only about (7×10^{-6}):**1**. This means that the stacks shown in FIGS. **2**, **3**, & **4** will need only proportionately narrow channels (as **12**, **22**, & **14**, **20**, **28**, **24**, **16**; FIG. **2** and their FIG. **4** counterparts), set between them to provide for VF flow.

Losses

[0144] We identify the principle losses and their associated decimal multiplier as follows:

1. Imperfect transparency of the dielectric in the stacks: ≅×0.75.
2. Unwanted VF scattering in the stacks and losses due to imperfect antenna positioning: ≅×(1/3).
3. Reduction in effective radiation pressure due to beam spread to either side of normal: ≅[1−cos(30)]=×0.846.
4. Lack of proportionality between increase of nanoantenna VF potential energy with $E._{int}^2$ due to polarization saturation: ≅×(2/3).
5. Finally we estimate inefficiencies in the transmission of VFs from antenna to antenna in the tiers, by making a reasonable assumption that the antennas transmit their outputs one to another in the forward direction with a decimal efficiency η=0.95. The output of the first antenna will then=η^1=0.95, the combined output of the first and second will=($\eta^1+\eta^2$)=1.853, and so on. This can be solved for any number of tiers by the summand below, which shows the gain g_n for any number of tiers n_{ts} per stack. An assumption that n_{ts}=26 looks good, in which case:

$$g_n = \left[\sum_{n_{t.s}=(n_{t.s})}^{0} \eta^{n_{t.s}}\right]_{=14}.$$

[0145] We therefore add a loss factor ≅×(14/26).

6. Finally we add ×0.7 for miscellaneous losses.

[0146] The summed losses therefore

[0147] ≅(0.75×(1/3)×0.846×(2/3)×(14/26)×0.7)=0.053.

[0148] Adjusting by this factor, we find that the net pressure per stack

[0149] =$p_{rp.net}$=0.1×0.053=0.0053 kg/m^2,

and similarly, the net pressure developed by a cubic meter TG=5300 kg.

Closing Remarks

[0150] It is to be understood that the calculations presented here are only intended as a rough guide to the performance of the apparatus. And further, that even though numerous characteristics and advantages of the present invention have been set forth in the above description, together with details of the structure and function of the invention, the disclosure is illustrative only, and changes may be made in detail, especially in matters of shape, size and arrangement of parts within the principles of the invention, to the full extent indicated by the broad general meaning of the terms in which the appended claims are expressed.

[0151] Further, the theories given above as to why VFs may usefully exchange momentum with propulsive devices, are provided in the spirit of presenting information as best the inventor is able, but it should be understood that the operation of the devices described here do not necessarily depend on the theories being correct as stated.

What is claimed is:

1. A method for producing thrust in a mass without ejection of tangible propellant, said method comprising the following steps;

- a) absorbing from free space the virtual radiation of so-called virtual particles such as vacuum fluctuations by means of antenna-like structures;
- b) causing said virtual radiation to flow along a plurality of said antenna-like structures whilst its field intensities are caused to be amplified;
- c) causing said virtual radiation with said amplified field intensities to be returned to free-space or emitted by said antenna-like structures;
- d) said emitting of said virtual radiation being further caused to occur in an essentially unidirectional or beam-like pattern;
- e) it being well known to those versed in the arts that said virtual radiation generates real radiation pressure, it is thus the case that said unidirectional emission of said virtual radiation causes said radiation pressure to urge said mass to move in accordance with Newton's third law, whereby the force exerted on said mass by said radiation pressure, and the reverse, are simultaneous; equal in magnitude; opposite in direction; and in the same line;
- f) steps (a) through (d) are accomplished by causing said antenna-like structures to be:
  - (1) appropriately structured such that they are tuned and suitably resonant in the targeted frequency range of said virtual radiation;
  - (2) appropriately spaced apart or contiguous, such that they are suitably linked one to another with regard to efficiency of radiative transfer, and so as to cause such particular electromagnetic phase differences between neighboring said antenna-like structures as may be required;
  - (3) appropriately energized by rapidly oscillating externally generated electromagnetic fields, such that said rapid oscillations of external fields cause said field intensities of said virtual radiation to vary cyclically.

2. The method recited in claim **1** wherein the following means are provided;

- a) first means for converting to travelling waves, the standing waves, otherwise known as oscillations, of said virtual particles, that are known to exist resonantly in numerous forms of matter;
- b) second means to provide that said numerous forms of matter take the form of nanoparticles, examples of which are fullerenes, nanotubes, spherical nano-clusters, and the like; along which said travelling waves may flow, growing in strength due to said amplification of field intensities as they do so, in approximate proportion to the number of said nanoparticles;
- c) third means whereby said travelling waves, on reaching an end-point of said nanoparticles, here termed the emitting end, are caused to undergo said emission into free space;
- d) it being the case that said flow of said travelling waves, and their said emission from said emitting end of said nanoparticles, causes said field intensities to become depleted to below zero-point level at or near another end-point, here termed the receiving end, at or near which said flow begins;
- e) it being further the case that said absorbing of said virtual radiation from free space occurs at or near said receiving end, thereby combining with, and contributing additively to, said flow of said travelling waves to said emitting end of said nanoparticles; thus restoring said depletion below zero-point level of said virtual radiation, and increasing it to above zero-point level;
- f) it being understood that while said depletion and said increasing of said virtual radiation is described in cause and effect terms for clarity, they are essentially simultaneous;
- g) it therefore being the case that said absorption; said flow of said travelling waves to said emitting end; and resultant said emission, may continue for as long as suitable apparatus for accomplishing the several means of this claim is appropriately energized.

3. Apparatus for accomplishing the goals of claim **2**, comprising:

- a) said nanoparticles such as may possess large electrical polarizabilities, particularly those formed from one of the general classes of materials possessing so-called giant polarizabilities; two such examples being various fullerenes and $Na_{14}F_{13}$ molecules;
- b) whereby said nanoparticles are so formed and positioned in said apparatus, as to function operatively as nanoantennas, such as may be best suited for said absorbing, said amplifying, and said emission of said virtual radiation;
- c) by said nanoantennas, is meant electromagnetic antennas, of comparable size to the wavelengths of said virtual radiation, said wavelengths being sub-micrometer in length; it being understood that said nanoantennas may also be linked contiguously one to another to form what is known to those versed in the arts as long-wire antennas;
- d) first means of assembling said nanoantennas into 2D arrays, by which is meant a single layer of said nanoantennas, multiple in number, that are formed into rows and columns in a single plane;
- e) second means of assembling said 2D arrays by a stacking process to form 3D arrays; the spacing between said nanoantennas in said 2D and said 3D arrays being such as may best permit said nanoantennas to accomplish said absorption, said amplification, and said unidirectional emission, of said virtual radiation, when energized by said rapid oscillations of external fields;
- f) to better illustrate, we mention that one well known example of said spacing of antennas in beaming arrays is that which permits said antennas to operate in phase-quadrature.

4. means of assembling said 3D arrays of claim **3** in a modified format, said modification being that said nanoantennas placed within it are spaced apart somewhat randomly instead of being arranged in a regular order as disclosed in **3**(*d*) and **3**(*e*); it being the case that the construction of said modified 3D array may thus be simplified;

a) it being further noted that the maximum spacing between said nanoantennas in said modified 3D arrays should be of the order of a wavelength of said virtual radiation.

5. Means whereby said 3D arrays of claims **3** may be placed one above another, or otherwise in close proximity to each other, while being separated by first radiation guiding structures, such that said emissions of one said 3D array is prevented from becoming the input of another; this being undesirable because such series feeding may lower the efficiency of said emission of said virtual radiation.

6. Means whereby said 3D arrays of claims **3** are housed within second radiation guiding structures that may be minor additions to said first radiation guiding structures, and which so direct said virtual radiation entering said first and second guiding structures, that said virtual radiation is emitted in the direction from which it first entered;

a) that is to say, that if, by way of example, said virtual radiation enters said first and said second guiding structures from the left, it also exits to the left, thus being scattered back about 180 degrees; thereby approximately doubling said radiation pressure acting on said first and said second guiding structures;

7. Apparatus for generating said rapid oscillations of external fields of claim **3**, for the purpose of energizing said 2D and said 3D arrays, whereby said rapid oscillations of external fields cause said amplification of said field intensities of said virtual radiation in said nanoantennas;

a) said apparatus for generating said rapid oscillations of external fields consisting of a high-packing density matrix of nano-emitters such as nano-vacuum cold triodes or the like, able to emit beamed electromagnetic radiation at terahertz frequencies or greater; said beams being capable of covering large areas of a square-meter or more when required for such applications as thrusting heavy vehicles;

b) means for developing electrical phase differences in said oscillations of said virtual radiation in said 2D and said 3D arrays, whereby said emissions of neighboring said nanoantennas interfere constructively in a preferred direction, and interfere destructively, or nearly so, in all other directions.

8. Means whereby said apparatus of claim **7** for generating said rapid oscillations of external fields may operate at a frequency lower than the frequency of said oscillations of said nanoantennas, while non-the-less causing said emissions of said nanoantennas at a higher frequency;

a) for example, said apparatus for generating said rapid oscillations of external fields may operate in the terahertz, i.e, 10^{14} hertz frequency range, while causing changes in the amplitude of said oscillations of said virtual radiation of said nanoantennas in the petahertz, i.e, 10^{15} hertz range, thereby producing said emissions at said petahertz range.

9. Means whereby said 3D arrays of claim **3** are rotatably mounted on pivoting means such as a shaft seated on bearings, such that said virtual radiation emitted by said 3D arrays is tangentially directed around said pivoting means, whereby it urges rotation of said 3D arrays around said pivoting means;

a) it being the case that said rotation may convey rotational energy to conventional work-performing devices, by means of gear trains or the like;

b) it being further the case that apparatus for causing said rapid oscillations of external fields may likewise be rotatably mounted on said pivoting means, so as to be in suitable close proximity to said 3D arrays;

c) wherein said apparatus for generating said rapid oscillations of external fields may be energized by conventional power supplies; said power being transmitted by such means as brushes and slip-rings mounted on said pivoting means; inductive transfer of power from a stationary coil to a rotating one mounted on said pivoting means; electrodes immersed in mercury pools, and the like.

* * * * *

US007379286B2

(12) **United States Patent**
Haisch et al.

(10) **Patent No.: US 7,379,286 B2**
(45) **Date of Patent: May 27, 2008**

(54) **QUANTUM VACUUM ENERGY EXTRACTION**

(75) Inventors: **Bernard Haisch**, Redwood City, CA (US); **Garret Moddel**, Boulder, CO (US)

(73) Assignee: **Jovion Corporation**, Menlo Park, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 430 days.

(21) Appl. No.: **11/236,142**

(22) Filed: **Sep. 26, 2005**

(65) **Prior Publication Data**

US 2007/0241470 A1 Oct. 18, 2007

(51) **Int. Cl.**
H02N 11/00 (2006.01)
H02M 1/00 (2007.01)

(52) **U.S. Cl.** **361/233**; 361/226; 363/178; 327/603; 318/16; 318/555; 318/558; 307/151

(58) **Field of Classification Search** .... None
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,018,180 A	5/1991	Shoulders	
5,590,031 A *	12/1996	Mead, Jr. et al.	363/8
6,477,028 B1 *	11/2002	Pinto	361/233
6,593,566 B1	7/2003	Pinto	
6,665,167 B2 *	12/2003	Pinto	361/233

OTHER PUBLICATIONS

Boyer, T. H. 1975, Random Electrodynamics: The Theory of Classical Electrodynamics with Classical Zero-Point Radiation Field, Phys. Rev. D, 11, 790.

Cole, D. C. and Puthoff, H. E. 1993, Extracting energy and heat from the vacuum, Phys. Rev. E, 48, 2, 1562.

Cole, D. C. and Zou, Yi 2003, Quantum Mechanical Ground State of Hydrogen Obtained from Classical Electrodynamics, Physics Letters A, vol. 317, No. 1-2, pp. 14-20 (Oct. 13, 2003), quant-ph/0307154.

(Continued)

Primary Examiner—Nikita Wells
(74) *Attorney, Agent, or Firm*—Pritzkau Patent Group, LLC

(57) **ABSTRACT**

A system is disclosed for converting energy from the electromagnetic quantum vacuum available at any point in the universe to usable energy in the form of heat, electricity, mechanical energy or other forms of power. By suppressing electromagnetic quantum vacuum energy at appropriate frequencies a change may be effected in the electron energy levels which will result in the emission or release of energy. Mode suppression of electromagnetic quantum vacuum radiation is known to take place in Casimir cavities. A Casimir cavity refers to any region in which electromagnetic modes are suppressed or restricted. When atoms enter into suitable micro Casimir cavities a decrease in the orbital energies of electrons in atoms will thus occur. Such energy will be captured in the claimed devices. Upon emergence form such micro Casimir cavities the atoms will be re-energized by the ambient electromagnetic quantum vacuum. In this way energy is extracted locally and replenished globally from and by the electromagnetic quantum vacuum. This process may be repeated an unlimited number of times. This process is also consistent with the conservation of energy in that all usable energy does come at the expense of the energy content of the electromagnetic quantum vacuum. Similar effects may be produced by acting upon molecular bonds. Devices are described in which gas is recycled through a multiplicity of Casimir cavities. The disclosed devices are scalable in size and energy output for applications ranging from replacements for small batteries to power plant sized generators of electricity.

20 Claims, 7 Drawing Sheets

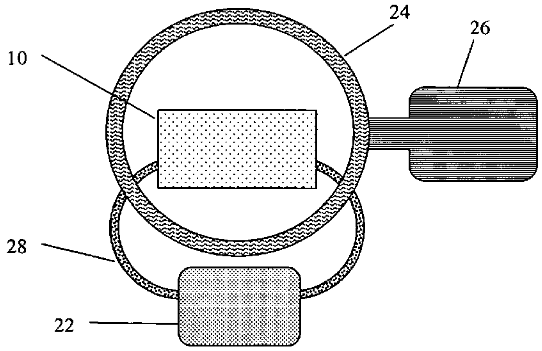

OTHER PUBLICATIONS

Cole, D. C. and Zou, Yi 2004, Analysis of Orbital Decay Time for the Classical Hydrogen Atom Interacting with Circularly Polarized Electromagnetic Radiation, Phys. Rev. E. 69 (1), 016601, pp. 1-12 (2004).

Forward, R. 1984, Extracting electrical energy from the vacuum by cohesion of charged foliated conductors, Phys. Rev. B, 30, 4, 1700.

Goho, A., "Diatom Menagerie," Science News, vol. 166, Jul. 17, 2004a, pp. 42-44, and references mentioned therein.

Goho, A., "Rattle and Hum; Molecular machinery makes yeast cells purr," Science News, vol. 166, Aug. 21, 2004b, p. 116, and references therein.

Milonni, P. W. The Quantum Vacuum: An Introduction to Quantum Electrodynamics, Sect. 3.3, Academic Press, 1994.

Park, W., and Summers, C. J., "Extraordinary refraction and dispersion in 2D photonic crystal slabs", Opt. Lett., vol. 27, (2002) p. 1397.

Plöl, A. and Kräuter, G., "Wafer direct bonding: tailoring adhesion between brittle materials", Mater. Sci. Eng. R, vol. 25 (1999) p. 1.

Puthoff, H. E., Ground State of Hydrogen as a Zero-Point-Fluctuation-Determined State, Phys. Rev. D 35, 3266, 1987.

Roy, Subrata et al., "Modeling gas flow through microchannels and nanopores," Journal of Applied Physics, vol. 93, No. 8, Apr. 15, 2003, pp. 4870-4879.

* cited by examiner

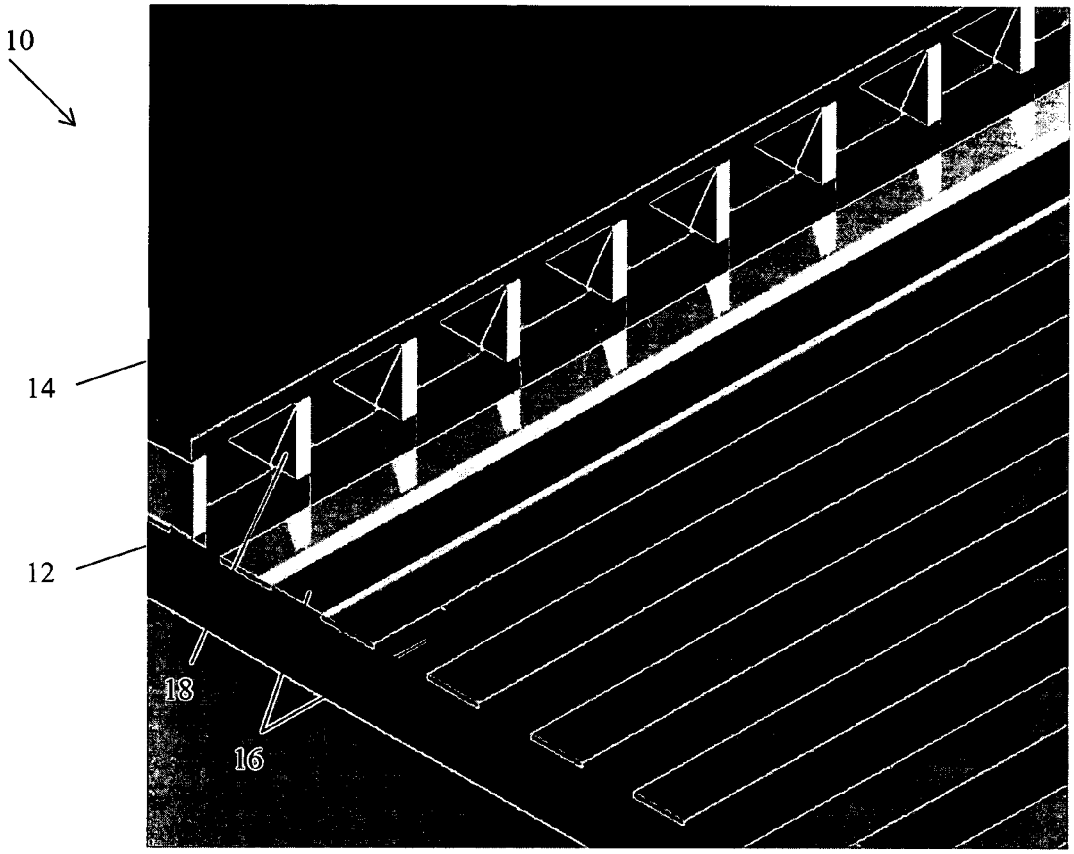
10
14
12
18
16

FIG. 1

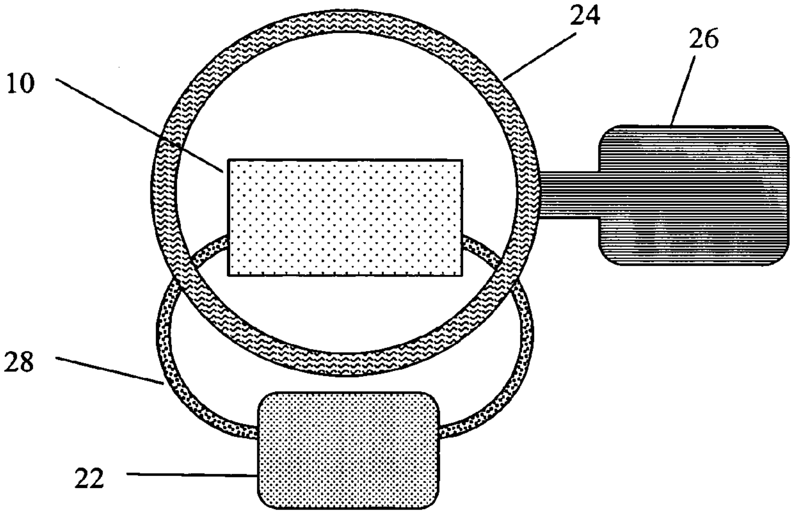
24
26
10
28
22

FIG. 2

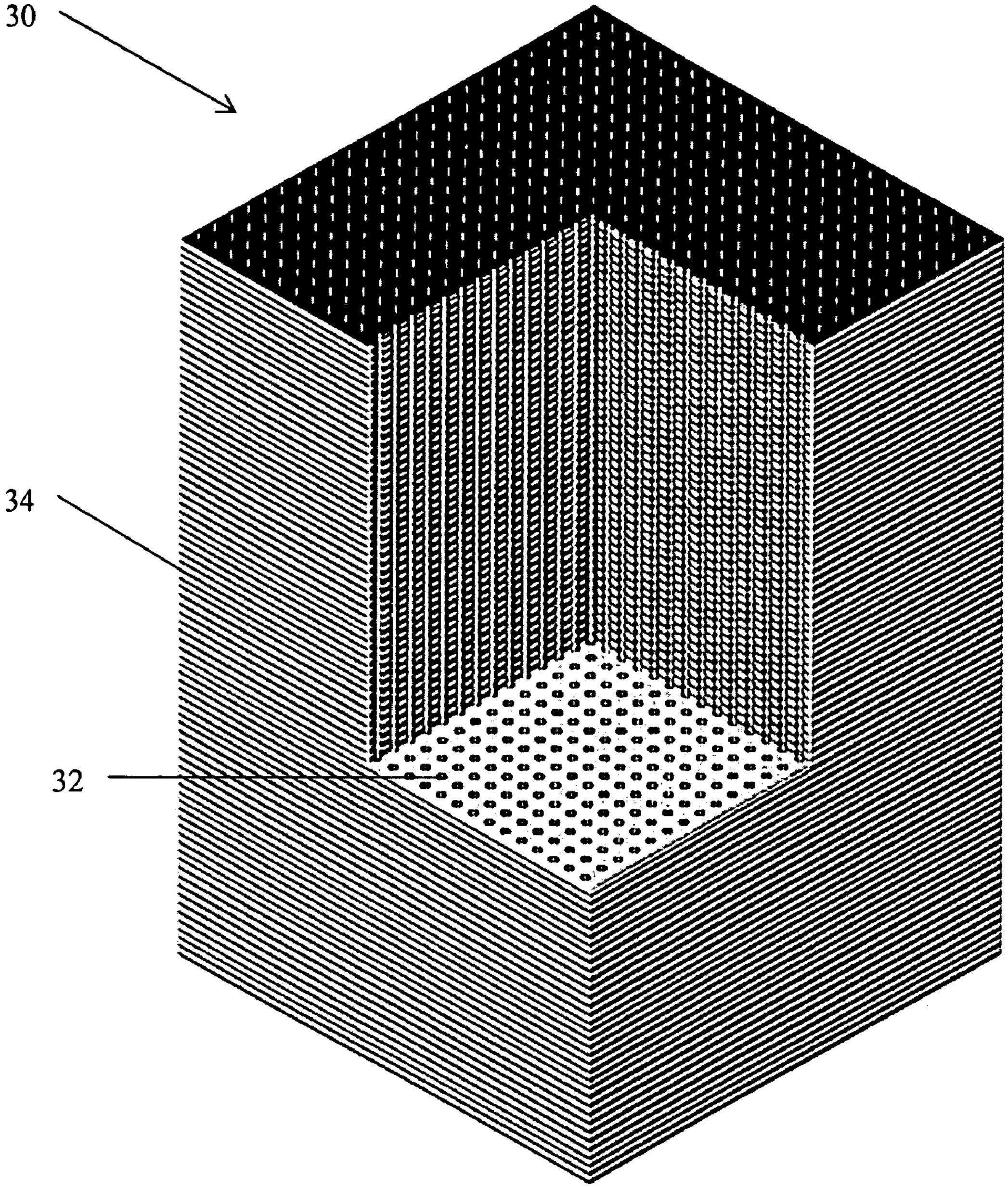

FIG. 3

FIG. 4A

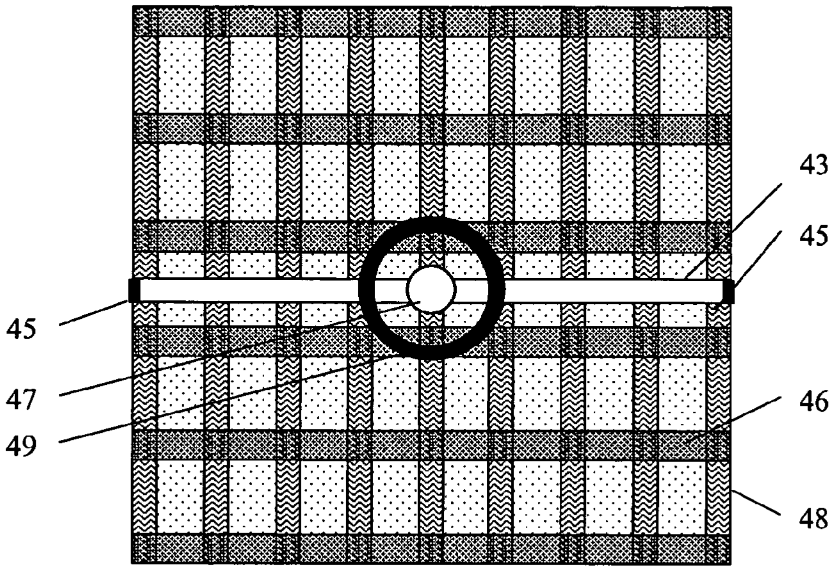
43
45
45
47
49
46
48

FIG. 4B

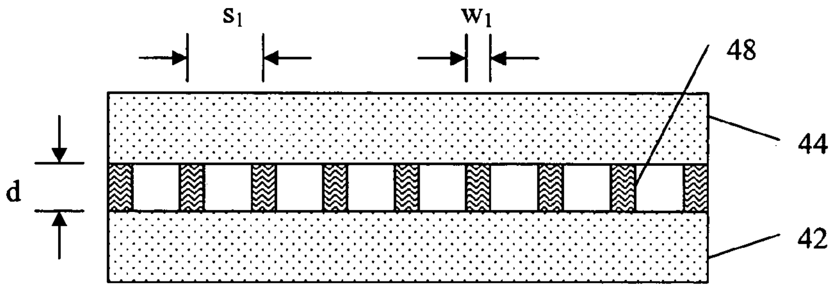
s_1
w_1
48
44
d
42

FIG. 4C

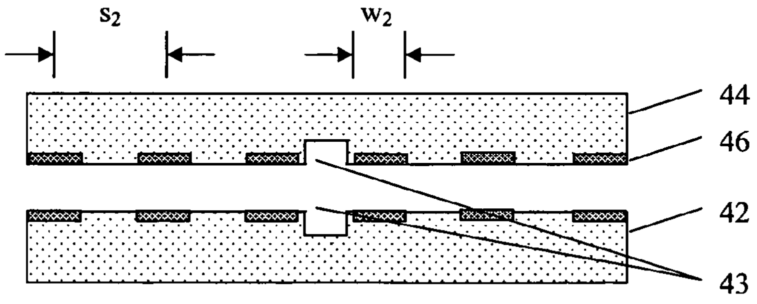
s_2
w_2
44
46
42
43

FIG. 4D

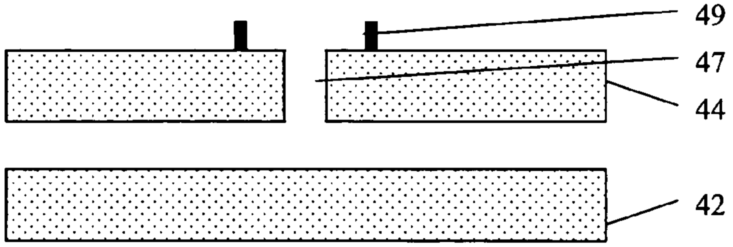

FIG. 5A

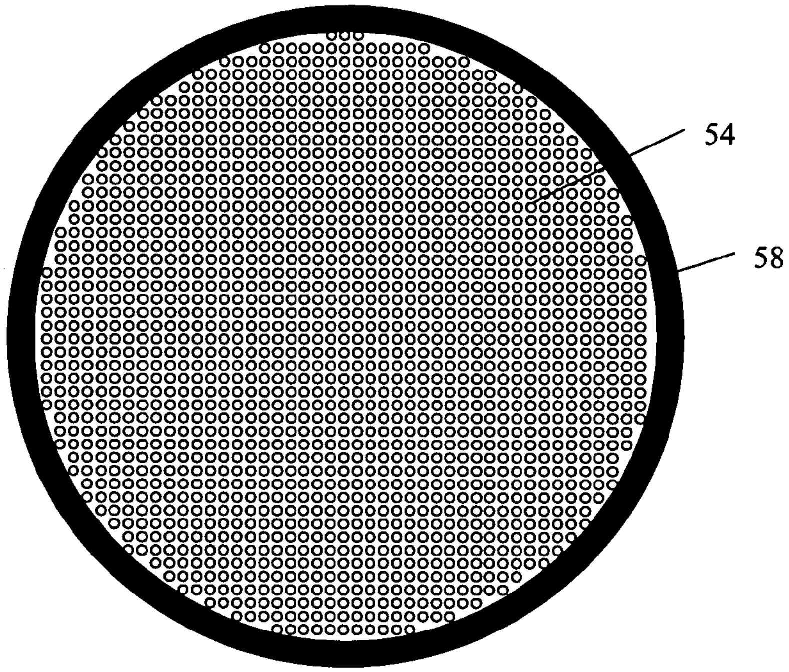

FIG. 5B

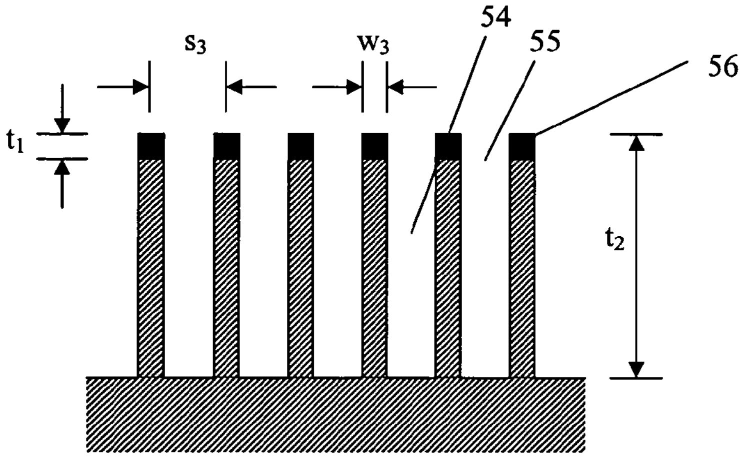
s_3
w_3
54
55
56
t_1
t_2

FIG. 5C

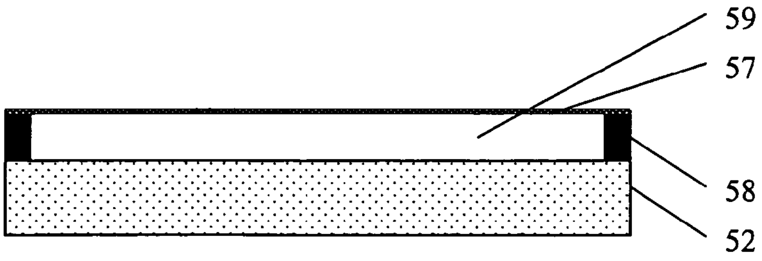

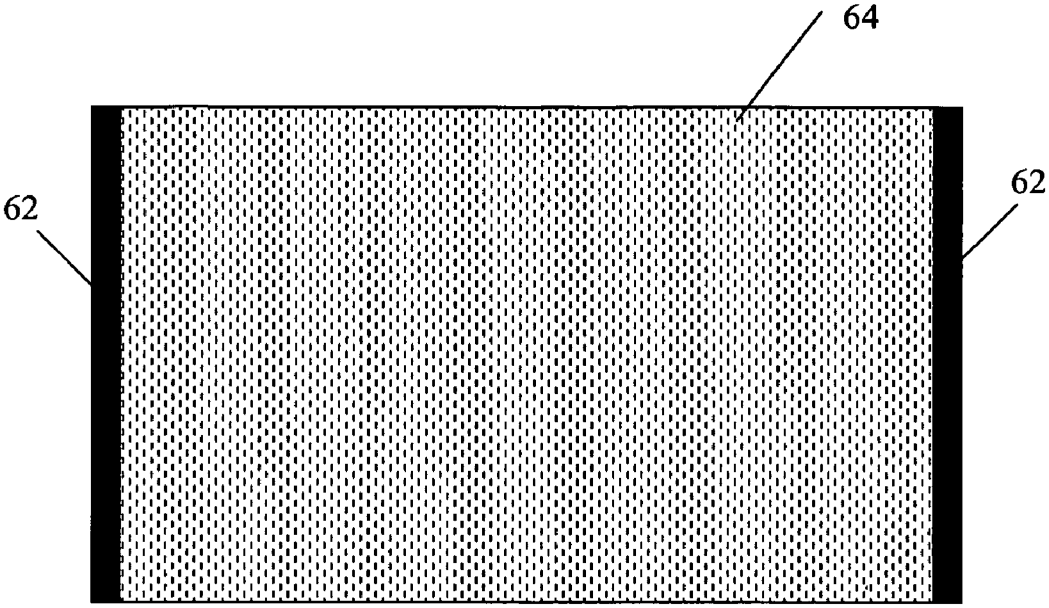

FIG. 6A

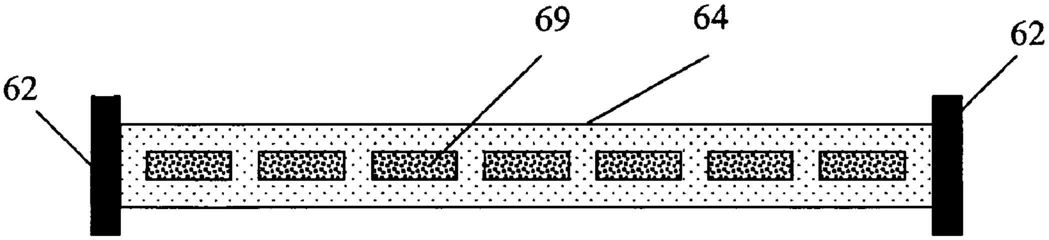

FIG. 6B

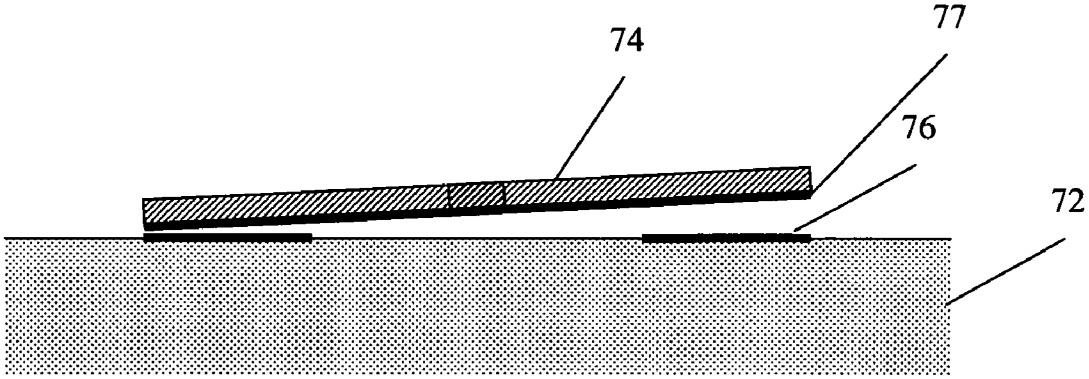

FIG. 7A

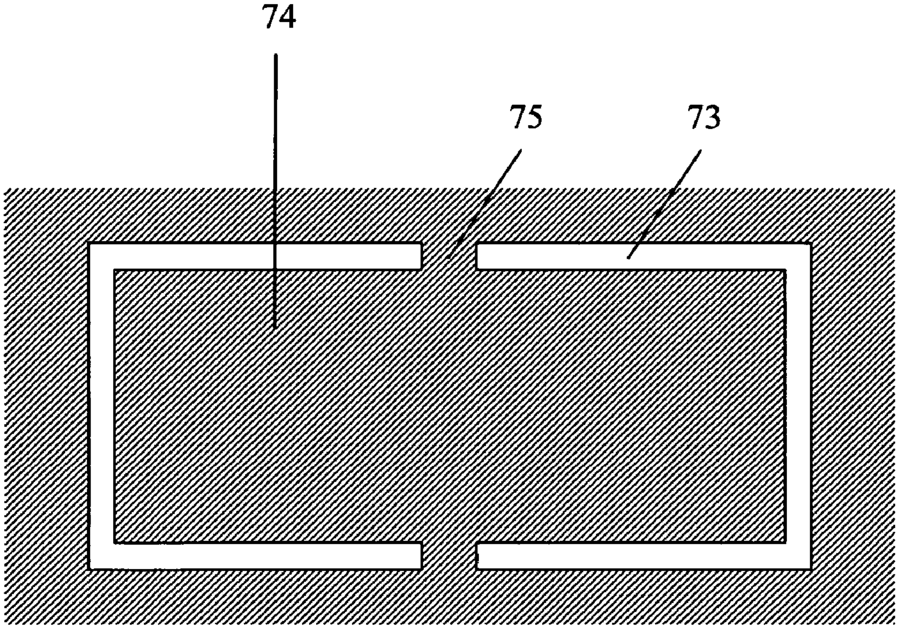

FIG. 7B

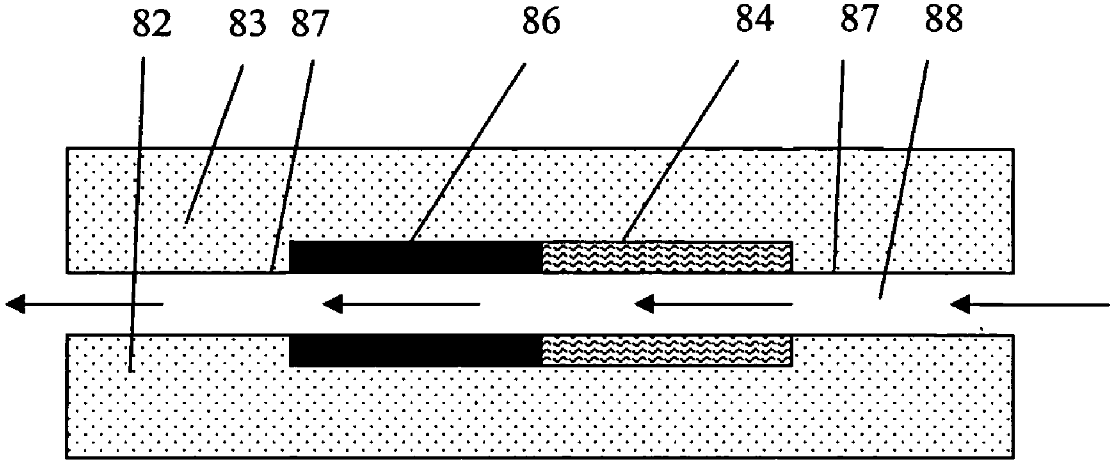
82
83
87
86
84
87
88

FIG. 8

QUANTUM VACUUM ENERGY EXTRACTION

BACKGROUND OF THE INVENTION

Max Planck proposed the concept of zero-point energy in 1912. The idea was then studied by Albert Einstein and Otto Stern in 1913. In 1916 Walther Nernst proposed that the Universe was filled with zero-point energy. The modern field of stochastic electrodynamics is based upon these ideas.

At that same time the structure and stability of the atom were puzzles. The Rutherford model of the atom was based on analogy to the motions of planets (electrons) around the Sun (the nucleus). However this was not feasible. The orbiting electron(s) would emit Larmor radiation, quickly losing energy and thus spiraling into the nucleus on time scales less than one-trillionth of a second, thereby rendering stable matter impossible. It is now known within the context of stochastic electrodynamics (SED) theory that a possible solution involves the absorption of zero-point energy. It was shown in 1975 by Boyer that the simplest possible atom and atomic state, the hydrogen atom in its ground state, would be in a state of equilibrium between Larmor radiation and absorption of zero-point energy at the correct radius for a classical Rutherford hydrogen atom.

Since this solution was not known in 1913, Niels Bohr followed a different path by simply postulating that only discrete energy levels were available to the electron in an atom. This line of reasoning let to the development of quantum theory in the 1920s. The concept of classical zero-point energy was forgotten for a decade. However the same concept found itself reborn in a quantum context in 1927 with the formulation of the Heisenberg uncertainty principle. According to the principle, the minimum energy of a harmonic oscillator has the value hf/2, where h is Planck's constant and f is the frequency. It is thus impossible to remove this last amount of random energy from an oscillating system.

Since the electromagnetic field also must be quantized in quantum theory, a parallel is drawn between the properties of a quantum oscillator and the waves of the electromagnetic field. It is concluded that the minimum energy of any possible mode of the electromagnetic field, consisting of frequency, propagation direction and polarization state, is hf/2. Multiplying this energy by all possible modes of the field gives rise to the electromagnetic quantum vacuum, which has identical properties—energy density and spectrum—to the classical zero-point energy studied by Planck, Einstein, Stern and Nernst a decade previously.

The line of inquiry involving classical physics plus the addition of a classical zero-point field was reopened in the 1960s by Trevor Marshall and Timothy Boyer and has been named stochastic electrodynamics (SED). SED asks the question: "Which quantum properties, processes or laws can be explained in terms of classical physics with the only addition being a zero-point electromagnetic field." Two of the early successes were a classical derivation of the blackbody spectrum (i.e. one not involving quantum physics) and the discovery that a classically orbiting electron in a hydrogen atom emitting Larmor radiation but absorbing zero-point radiation would have an equilibrium orbit at the classical Bohr radius. An initial approach to this problem by Timothy Boyer (1975) was perfected by H. E. Puthoff (1987). Their analyses treated the orbiting electron as a harmonic oscillator.

This result underwent a major new development with the recent work of Daniel Cole and Y. Zou which simulated the orbit of a classical electron in a true Coulomb field of a hydrogen nucleus and found that such a realistic electron would find itself in a range of distances from the nucleus, in agreement with quantum mechanics, owing to the random nature of the emission and absorption processes. The mean position is at the correct Bohr radius, but the actual distribution of positions very precisely duplicates the electron probability distribution of the corresponding Schrödinger equation in which the electron is regarded as being represented by a wave function. (In the SED representation the electron is "smeared out" not because it is a wave function, but because as a point-like particle it is subject to the continuous perturbations of the electromagnetic quantum vacuum fluctuations.)

A clear consequence of this theory is that a reduction of the electromagnetic quantum vacuum at the frequency corresponding to the orbit of the electron will result in a decay of the orbit since there will thereby be an imbalance in the Larmor radiation vs. absorption.

The electromagnetic quantum vacuum energy spectrum is proportional to the cube of the frequency. If the vacuum energy is suppressed at the frequency of the "normal" orbit of the electron, this will cause the electron to spiral inward to a higher frequency orbit. In this fashion it will then encounter a new equilibrium situation with the electromagnetic quantum vacuum energy spectrum owing to that spectrum's increase with the cube of the frequency.

If the SED interpretation is correct for the hydrogen atom as the analyses of Boyer, Puthoff, Cole and Zou indicate, it must apply as well to all other atoms and their multi-electron configurations. In that case, a transition of an electron from an excited state to a lower energy state involves a rapid decay from one stable orbit to another, not an instantaneous quantum jump. The details of the bases for stability of electron orbits has yet to be determined by SED theory, but the logical extrapolation from the single-electron hydrogen case is clear: electron orbits in all atoms must be determined by an emission vs. absorption balance and thus are subject to modification involving mode suppression of the electromagnetic zero-point field at appropriate frequencies.

It is claimed that modification of electron orbits is in essence the same process as natural transition between energy levels of electrons in atoms and therefore that the energy released in such a process can be captured in the same way as ordinary transition energy.

By moving an atom into and out of a microstructure that suppresses appropriate modes of the electromagnetic quantum vacuum, an extraction of energy from the electromagnetic quantum vacuum may be accomplished. This can be done with micro Casimir cavities.

The electromagnetic quantum vacuum as a real source of energy is indicated by the Lamb shift between s and p levels in hydrogen, van der Waals forces, the Aharanov-Bohm effect, and noise in electronic circuits.

However the most important effect of the electromagnetic quantum vacuum is the existence of the Casimir Force, a force between parallel conducting plates which may be interpreted as a radiation pressure effect of electromagnetic quantum vacuum energy. Electromagnetic waves in a cavity whose walls are conducting are constrained to certain wavelengths for reasons having to do with transverse component boundary conditions on the wall surfaces. As a result, in a Casimir cavity between parallel plates there will be, in effect, an exclusion of radiation modes whose wavelengths are longer than the separation of the plates. An overpressure of electromagnetic quantum vacuum radiation on the outside then pushes the plates together. An extensive literature exists on the Casimir force and the reality of the force has moved from laboratory experimentation to micro-electro-mechanical structures (MEMS) technology both as a problem (so-called "stiction") and as a possible control mechanism.

The exclusion of modes does not begin all at once at the wavelength equivalent to the plate separation, d. Mode suppression will be strongest for wavelengths of d or greater, but will begin to occur as well for wavelengths falling in between the "stairway" d/n, with the effect diminishing as n increases. We propose to use the partial suppression of modes for wavelengths shorter than d occurring in this fashion in order to be able to employ Casimir cavities of the maximum possible physical size.

Researchers have shown that thermodynamic laws are not violated when energy is "extracted" from zero-point energy, as energy is still conserved and the second law is not violated. Cole and Puthoff have carried out and published thermodynamic analyses showing that there is no violation. Indeed, a thought experiment by Forward (1984) showed a simple, but not practical, energy extraction experiment.

In the stochastic electrodynamics (SED) interpretation of the hydrogen atom, the ground state is interpreted as effectively equivalent to a classically orbiting electron whose velocity is c/137. The orbit is stable at the Bohr radius owing to a balance between classical electromagnetic emission and absorption from the electromagnetic zero-point field. This view, first obtained by Boyer (1975) and subsequently refined by Puthoff (1987) has been further strengthened by the detailed simulations of Cole and Zou (2003, 2004) which demonstrate that the stochastic motions of the electron in this interpretation reproduce the probability density distribution of the Schrödinger wave function. Note that one apparent difference between this interpretation and that of quantum mechanics is that in quantum mechanics the 1s state of the electron is regarded as having zero angular momentum, whereas in the SED interpretation the electron has an instantaneous angular momentum of $mcr/137=h/2\pi$. However SED simulations by Nickisch have shown that the time-averaged angular momentum is zero just as in the quantum case owing to the zero-point perturbations on the orbital plane. Thus averaged over enough "orbits" this "classical electron" will fill a spherical symmetric volume around the nucleus having the same radial probability density as the Schrödinger wave function and zero net angular momentum, completely consistent with quantum behavior.

The Bohr radius of the atom in the SED view is 0.529 A (Angstroms). This implies that the wavelength of zero-point radiation responsible for sustaining the orbit is $2*\pi*0.529*137=455$ A (0.0455 microns). It is claimed that suppression of zero-point radiation at this wavelength and shorter in a Casimir cavity will result in the decay of the electron to a lower energy state determined by a new balance between classical emission of an accelerated charge and zero-point radiation at $\lambda<455$ A, where λ depends on the Casimir plate separation, d. Note that the tail end of the quantum probability density of the electron (as well as the SED simulation of Cole and Zou) extends beyond five Bohr radii, so that some change in the energy balance could be accomplished even at considerably longer wavelengths of perhaps 0.1 microns-0.2 microns

Since the frequency of this orbit is 6.6×10^{15} s^{-1}, no matter how quickly the atom is injected into a Casimir cavity the process will be a slow one as experienced by the orbiting electron. We therefore assume that the decay to a new sub-Bohr ground state will involve gradual release of energy in the form of heat, rather than a sudden optical radiation signature.

Since the binding energy of the electron is 13.6 eV, we assume that the amount of energy released in this process would be on the order of 1 to 10 eV for injection of the hydrogen atom into a Casimir cavity of d=250 A or thereabouts (and perhaps even a larger cavity as noted above). Upon exiting the cavity the electron would absorb energy from the zero-point field and be re-excited to its normal state. The energy (heat) extracted in the process comes at the expense of the zero-point field, which in the SED interpretation flows at the speed of light throughout the Universe. We are in effect extracting energy locally and replenishing it globally. Imagine extracting thimbles-full of water from the ocean. Yes, the ocean is being depleted thereby, but no practical consequences ensue.

Since naturally occurring hydrogen at standard temperature and pressure (STP) is a two-atom molecule, a dissociation process would need to precede an injection of hydrogen atoms into a Casimir cavity. We avoid this complication and take advantage of multi-electron modification by working with monatomic (noble) gases which also have the advantage of being safe and inexpensive.

We work with naturally occurring monatomic gases for three reasons:

(1) No dissociation process is required.

(2) Heavier element atoms are approximately two to four times larger than hydrogen and thus can utilize and be affected by a larger Casimir cavity which is easier to fabricate.

(3) Heavier elements have numerous outer shell electrons, several of which may be simultaneously affected by the reduction of zero-point radiation in a Casimir cavity.

The following five noble gases are potentially suitable:
He (Z=2, r=1.2 A)
Ne (Z=10, r=1.3 A),
Ar (Z=18, r=1.6 A)
Kr (Z=36, r=1.8 A)
Xe (Z=54, r=2.05 A).

All of these elements contain ns electrons. He has two 1s electrons. Ne has two each of 1s and 2s electrons. Ar has two each of 1s, 2s, and 3s electrons. Kr has two of each of 1s, 2s, 3s, and 4s electrons. Xe has two of each of 1s, 2s, 3s, 4s and 5s electrons.

Assuming an outermost electron which is completely shielded by the other electrons (a crude assumption), its orbital velocity would scale as $r^{-1/2}$ (the familiar Keplerian period squared proportional to semi-major axis cubed relationship) and thus λ proportional to r/v) will scale as $r^{3/2}$. If that is the case, then the larger radii translate as $r^{3/2}$ into larger Casimir cavities having an effect on the energetics of the outer electron shells. We would therefore expect that a Casimir cavity having d=0.1 microns (or perhaps even as large as one micron would have an effect on reducing the energy levels of the outermost pair of s electrons . . . and possibly also p electrons and intermediate shell s electrons as well.

It is reasonable to expect that a 0.1 microns Casimir cavity would result in a release of 1 to 10 eV for each injection of a He, Ne, Ar, Kr or Xe atom into such a cavity.

According to a Jordan Maclay, who has done theoretical Casimir cavity calculations, a long cylindrical cavity results in an inward force on the cavity. In the "exclusion of modes" interpretation of the Casimir force, this implies that a cylindrical cavity of diameter 0.1 micron would yield the desired decay of outer shell electrons and subsequent release of energy.

It is now recognized that an electromagnetic quantum vacuum field is formally necessary for atomic stability in conventional quantum theory (Milonni 1994). In the field of physics known as stochastic electrodynamics, this concept has been shown by theory and simulations to underlie the ground state of the electron in the hydrogen atom. The classical Bohr orbit is determined by a balance of Larmor emission and absorption of energy from the zero-point fluctuations of the electromagnetic quantum vacuum in SED theory. It follows that upon suppression of appropriate zero-point fluctuations the balance will be upset causing the electron to decay to a lower energy level not ordinarily found in nature with a release of energy during this transition. A Casimir cavity of the proper dimensions can accomplish this suppression of zero-point fluctuations. A Casimir cavity refers to any region in which electromagnetic modes are suppressed or restricted. Upon entering such a properly designed Casimir cavity the electron energy level will shift and energy will be released. Upon exiting the Casimir cavity the electron will return to its customary state by absorbing energy from the ambient zero-point fluctuations. This permits an energy extraction cycle to be achieved at the expense of the zero-point fluctuations. Although it has not yet been proven theoretically, a similar balance of Larmor emission and absorption of energy from the zero-point fluctuations must underlie the electron states of all atoms, not just hydrogen, permitting any atom to be used as a catalyst for extraction of zero-point energy (the energy associated with the zero-point fluctuations). An analogous process is also believed to underlie molecular bonds, yielding a similar energy extraction cycle.

The following is a list of patents that deal with related phenomena:

U.S. Pat. No. 5,018,180, Energy conversion using high charge density, Kenneth R. Shoulders. This concerns the production of charge clusters in spark discharges. It is conjectured that the electrostatic repulsion of charges is overcome in charge clusters by a Casimir-like force. This invention does not deal with energy release from atoms in Casimir cavities and is therefore not relevant to the present invention.

U.S. Pat. No. 5,590,031, System for converting electromagnetic radiation energy to electrical energy, Franklin B. Mead and Jack Nachamkin. This invention does not deal with energy release from atoms in Casimir cavities and is therefore not relevant to the present invention.

U.S. Pat. No. 6,477,028, Method and apparatus for energy extraction, Fabrizio Pinto. Proposes to vary one or more of a variety of physical factors that affect the Casimir force, or by altering any of a variety of environmental factors that affect such physical factors and thereby render a Casimir force system as non-conservative. This invention does not deal with energy release from atoms in Casimir cavities and is therefore not relevant to the present invention.

U.S. Pat. No. 6,593,566, Method and apparatus for energy extraction, Fabrizio Pinto. A method and apparatus for accelerating and a decelerating particles based on particle surface interactions. This invention does not deal with energy release from atoms in Casimir cavities and is therefore not relevant to the present invention.

U.S. Pat. No. 6,665,167, Method for energy extraction-I, Fabrizio Pinto. Similar to U.S. Pat. No. 6,477,028. This invention does not deal with energy release from atoms in Casimir cavities and is therefore not relevant to the present invention.

SUMMARY OF THE INVENTION

A system is disclosed for converting part of the energy of the electromagnetic quantum vacuum available at any point in the universe to usable energy in the form of heat, electricity, mechanical energy or other forms of power. This is accomplished using an effect on the electron configurations of atoms predicted by the theory of stochastic electrodynamics (SED). Within the context of SED theory it is predicted that the electron energy levels in atoms are determined by a balance of Larmor radiation vs. absorption of radiative energy from the electromagnetic quantum vacuum. By suppressing electromagnetic quantum vacuum energy at appropriate frequencies a change may be effected in the electron energy levels which will result in the emission or release of energy. This change in energies is analogous to a standard emission of a photon as an electron makes a transition from an excited to a lower energy state, but on a longer time scale and with the change being a continuous one rather than a “jump” from one energy level to another. Mode suppression of electromagnetic quantum vacuum radiation is known to take place in Casimir cavities. A Casimir cavity refers to any region in which electromagnetic modes are suppressed or restricted. When atoms enter into suitable micro Casimir cavities a decrease in the orbital energies of electrons in atoms will thus occur, with the effect being most pronounced for outer shell electrons. Such energy will be captured in the claimed devices. Upon emergence form such micro Casimir cavities the atoms will be re-energized by the ambient electromagnetic quantum vacuum. In this way energy is extracted locally and replenished globally from and by the electromagnetic quantum vacuum. This process may be repeated an unlimited number of times. This process is also consistent with the conservation of energy in that all usable energy does come at the expense of the energy content of the electromagnetic quantum vacuum. Two example variations of a system are disclosed that permit multiple extractions of electromagnetic quantum vacuum energy during passage of a gas through a series of micro Casimir cavities and that operate in a self-sustaining, recycling fashion. Similar effects may be produced by acting upon molecular bonds. The disclosed devices are scalable in size and energy output for applications ranging from replacements for small batteries to power plant sized generators of electricity. Since the electromagnetic quantum vacuum is thought to permeate the entire Universe, devices drawing power from the electromagnetic quantum vacuum in the fashion claimed will be effectively inexhaustible sources of power.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention may be understood by reference to the following detailed description taken in conjunction with the drawings briefly described below.

FIG. **1** is a diagrammatic illustration of a set of channels each containing a multiplicity of Casimir cavities in accordance with the present invention.

FIG. **2** is a diagrammatic illustration of a system for converting quantum vacuum energy into locally usable power in accordance with the present invention.

FIG. **3** is a diagrammatic illustration of a block of tunnels each containing a multiplicity of Casimir cavities in accordance with the present invention.

FIGS. **4A-4D** are diagrammatic illustrations of Casimir channels in bonded wafers in accordance with the present invention.

FIGS. **5**A-**5**C are diagrammatic illustrations of a device for oscillating a fluid though Casimir channels in accordance with the present invention.

FIGS. **6**A and **6**B are diagrammatic illustrations of a device switching the reflecting characteristics of walls of Casimir cavities in accordance with the present invention.

FIGS. **7**A and **7**B are diagrammatic illustrations of a device Casimir cavity wall spacing in accordance with the present invention.

FIG. **8** is a diagrammatic illustration of a device incorporating asymmetric Casimir cavity entry and exits in accordance with the present invention.

DETAILED DESCRIPTION

The first embodiments of this concept utilize Casimir cavities consisting of volumes through which, or in and out of which, gases flow, and which on the size scales of atoms appear as regions bounded by parallel plates of conducting material in which the plate scales are much larger than the plate separations; or by cylinders of conducting material in which the lengths of the cylinders are much larger than the diameters. It is claimed that other forms of Casimir cavity are capable of producing a similar effect, and the term Casimir cavity will be used to designate any volume capable of mode suppression of the zero-point field. The necessary condition is that the mode suppression ability of the Casimir cavity be matched to the electron energy levels in such a way as to result in a significant difference of the electron energy levels inside vs. outside the cavity.

These embodiments demonstrate the following concepts:

A method, comprising: (a) use of a device including a series of Casimir cavities and causing a specific gas to flow through the cavities, said Casimir cavities being configured and said specific gas being selected such that as the gas flows through the cavities energy is released from the gas; and (b) means for collecting at least some of said released gas.

A method, comprising: (a) providing a device including at least one Casimir cavity and causing a specific gas to enter and then exit the cavity, said Casimir cavity being configured and said specific gas being selected such that when the gas is caused to enter the cavity, energy is released from the gas; and (b) means for collecting at least some of said released energy.

A means for effecting changes in the electron configurations. A system for converting part of the energy of the electromagnetic quantum vacuum available at any point in the Universe to usable energy in the form of heat, electricity, mechanical energy or other forms of power.

A means for effecting changes in the electron configurations in the process of which energy is released.

A means for allowing the electron configurations to be re-energized by exposure to the ambient electromagnetic quantum vacuum radiation.

The use of microstructures consisting of many pairs of alternating Casimir cavities and regions in which the electromagnetic quantum vacuum radiation freely propagates.

The use of conducting strips on facing pairs of plates so that atoms go through alternating regions in which they are exposed to the full electromagnetic quantum vacuum spectrum, and regions in which part of the spectrum is blocked. The result is that they dump (or radiate) the energy difference into the local medium.

The use of spacers to separate the layer pairs.

The use of multiple conducting strips to amplify the effect (hugely).

The stacking of such plates with strips on both sides so that the top of one pair becomes the bottom of the next, each with identical conducting strips which form Casimir cavities with their partner strips in each pair.

The use of sandwiched layers of alternating conducting and non-conducting plates having micron sized thicknesses in which micron or submicron diameter holes are introduced by etching or some other method.

The stacking, co-registration and alignment of such sandwiched layers to produce many parallel Casimir tunnels having alternating Casimir and non-conducting segments.

The use of multiple segments to amplify the effect (hugely).

The use of monatomic gases as the medium in such a system.

The use of molecular gases in such a system for the purpose of modifying molecular bonds with the attendant release of energy.

A closed recycling system in which these processes take place.

Fabricatable and workable configuration and dimensions but with the claims not limited to these specific embodiments.

A means whereby the flow of gas is initiated and maintained in a closed system.

A means whereby the energy released from the electron orbital changes is converted into usable energy in the form of heat, electricity, mechanical energy or other forms of power.

Casimir Channels

This embodiment shown in FIG. **1** involves two square parallel plates **12** and **14**, 10×10 cm in size for illustration. On each one lay down 5000 conducting strips **16** that are 10 microns in width and the full 10 cm in length, separated by 10 microns non-conducting strips. Perpendicular to the strips deposit a spacer material **18** at 0.1 to 1 cm intervals with a height of 0.1 microns. Put the plates face to face and align the strips so as to form 5000 Casimir strips.

If we assume a gas flow rate of 10 cm/s parallel to the spacers and perpendicular to the strips, this would result in 1.3×10^{20} transitions/s.

An energy release of 1 to 10 eV per transition corresponds to 21 to 210 watts of energy release for the entire Casimir cavity. A stacked set of 10 or more such layers could be fabricated yielding 210 to 2100 watts for a 10×10×10 cm block.

This may be directly converted into electricity using a thermophotovoltaic process, or indirectly by using a heat exchanger. As in the previous embodiment, one means of capturing the emitted radiation is to surround the apparatus with a water bath.

The dimensions above are solely examples. The device may be scaled to both smaller and significantly larger dimensions.

The essential components of an energy generating device of this sort shown in FIG. **2** are:

(1) An array of parallel Casimir channels with conducting strips **10**

(2) A pump **22** providing continuous recycling of gas through the tunnels

(3) A means **24** for capturing the emitted energy

(4) A thermal photovoltaic, heat exchanger or other device **26** capable of converting output heat into electricity or other usable forms of power.

A desirable property of the system is its ability to radiate the accumulated energy locally and absorb it globally. Thus surprisingly the means **24** for capturing the emitted energy can capture the emitted energy without hindering the capture of the quantum vacuum energy by the gas. This is due to the fact that the vacuum field permeates all space and cannot be blocked. (Note that the reason that Casimir cavities have reduced vacuum energy modes is not that they block it, but rather that because of destructive interference they do not allow some of the electromagnetic modes to exist in their interior.) A second reason that the means **24** does not block the capture of the quantum vacuum energy is that the absorbed energy is dominantly shorter wavelength electromagnetic modes that are not absorbed by the means **24**, whereas the radiated energy can be longer wavelengths for which the means **24** has a much larger absorption coefficient. Such is the case, for example, when the means **24** comprises a water bath.

The first two components will be enclosed in sealed structure. The third and fourth components may be interior or exterior to this structure.

A variation on the above device consists of stacking plates such that the top of one pair becomes the bottom in the next pair, etc.

Casimir Tunnels

One embodiment of the concept shown in FIG. **3** is multiple, parallel, 0.1 micron diameter Casimir tunnels. If we let the length of the cylinder be 100 times the width, this results in z=10 microns for the length of the Casimir tunnel. We propose a segmented tunnel consisting of alternating conducting and non-conducting materials, each 10 microns in length. In a length of 1 cm, there could be 500 such pairs in segments, resulting in 500 energy releases events (each yielding 1 to 10 eV) for each transit of an atom through the entire 1 cm-long segmented Casimir tunnel.

Consider a one cubic cm "Casimir Block" that is built up of 10 micron thick alternating layers as shown in FIG. **3**. Assume that tunnels **32** of 0.1 micron diameter could be drilled through the cube perpendicular to the layers **34** (this is not physically possible, of course; tunnel manufacture must be done differently). Ten percent of the cross section comprises entrance to some 1.3 billion tunnels. The amount of energy released would be proportional to the flow rate of the gas through these tunnels.

A flow rate of 10 cm s^{-1} through a total cross sectional area of 0.1 cm^2 yields 1 cm^3 of gas per second flowing through the tunnels, which at STP would be 2.7×10^{19} atoms. A very simple sealed, closed-loop pumping system could maintain such a continuous gas flow. Since each atom interacts 500 times during its passage, there would be 1.3×10^{22} transitions per second in the entire cube of one cubic centimeter. An energy release of 1 to 10 eV per transition corresponds to 2150 to 21500 watts of energy release for the entire Casimir cube of segmented tunnels.

Obviously it is not possible to drill 1.3 billion tunnels having diameters of 0.1 microns. However it is feasible to use microchip technology to etch holes into the individual layers first and then assemble the stack. Extremely fine coregistration and alignment of stacks will need to be accomplished.

This may be directly converted into electricity using a thermophotovoltaic process, or indirectly by using a heat exchanger.

One means to capture the emitted energy is to surround the apparatus with a water bath. Water absorbs infrared radiation very effectively. For the wavelength range of 2 microns to 200 microns, the absorption coefficient of water is greater than 10 cm^{-1}. Therefore a layer of water that is 1 mm thick and surrounds the apparatus will be sufficient to absorb nearly all the emitted infrared radiation. The water will be heated, and that heat converted into the desired form of energy.

The dimensions above are solely examples. The device may be scaled to both smaller and significantly larger dimensions.

The essential components of an energy generating device of this sort are:

(1) An array of parallel segmented Casimir tunnels **32**

(2) A pump **22** providing continuous recycling of gas through the tunnels

(3) A means **24** for capturing the emitted energy

(4) A thermal photovoltaic, heat exchanger or other device **26** capable of converting output heat into electricity or other usable forms of power.

The first two components will be enclosed in sealed structure. The third and fourth components may be interior or exterior to this structure.

Casimir Channels in Bonded Wafers

The basic concept of the present invention is to flow gas into and out from multiple Casimir cavities. When the gas is outside of a Casimir cavity, a wide range of quantum mechanical vacuum electromagnetic modes are available to interact with the gas's atomic electronic orbital states. When the gas passes into a Casimir cavity the range of available modes is restricted and the gas sheds some of its electromagnetic energy such that this energy is available locally. When the gas once again flows out from the Casimir cavity, the gas's atomic electronic orbital state energy is recharged from quantum mechanical vacuum fields. Thus energy is harvested globally and delivered locally.

The configuration for a basic device comprising bonded wafers is shown in FIGS. **4**A-**4**D. A top view is shown in FIG. **4**. The device is 1 sq. cm. As seen from the south edge **41** in FIG. **4**B, it consists of two substrates **42** and **44** separated by a series of spacers which extend across the device from the south to the north side. These spacers have a height d, a width w_1, and a center-to-center spacing s_1. The thin gaps delineated by the spacers **48** extend to openings at the south edge of the device, as seen in FIG. **4**B and the north edge. As seen from the east edge in FIG. **4**C, the upper **44** and lower **42** substrates are each coated with conducting stripes **46** that extend from the east edge to the west edge. These stripes are discontinuous, such that the discontinuity occurs at each region where the stripe is intersected by a spacer **48**. These stripes have a width w_2 and a center-to-center spacing s_2. In the central region of the device there is a region of both substrates that has been removed to form a conduit **43** from close to the east edge to close to the west edge. This conduit does not extend all the way to the edges, but is instead sealed **45** at each end, as shown in FIG. **4**A. Finally, as seen in FIG. **4**D, which shows an east view of the central cross section, and in FIG. **4**A, a hole **47** extends through the upper substrate. This hole connects to the conduit **43** shown in FIGS. **4**A and **4**C. As can also be seen in FIGS. **4**A and **4**D, a connector ring **49** that surrounds the hole is affixed to the upper substrate.

For the device to function, gas tubing **28**, shown in FIG. **2**, is attached to the connector ring **49** extending from the upper substrate, forming a sealed connection. Pressurized gas flows through the tubing and the hole **47** in the upper substrate into the conduit **43** between the substrates. From the conduit **43** the gas flows from the central region through the gap between the substrates to the north and south edges. The spacers guide the gas so that it flows alternately between regions coated with the conducting stripes **46** and regions that are not coated with these stripes, until it reaches the north and south edges, at which point it escapes from the gap between the substrates. The escaped gas is captured in a surrounding enclosure, not shown, and pumped back through the tubing **28** into the hole at the top center of the device, forming a close-loop system. In this way the gas is passed through multiple Casimir cavities. The gas atoms or molecules absorb energy from the surrounding electromagnetic field when they are in the non-conducting region and then release a portion of their energy as they enter the gap between the conductive coatings, i.e., in the Casimir cavity.

The apparatus is surrounded by a means **24** to capture the released energy, such as a water bath, shown in FIG. **2**. Water absorbs infrared radiation very effectively. For the wavelength range of 2 microns to 200 microns, the absorption coefficient of water is greater than 10 cm^{-1}. Therefore a layer of water that is 1 mm thick and surrounds the apparatus is sufficient to absorb a large proportion of the emitted infrared radiation, providing thermal energy to heat the water. That energy can be used directly as heating source, or converted into the desired form of energy, by means **26** well known to those skilled in the art.

The materials and dimensions in the preferred embodiment are as follows. The upper **44** and lower **42** substrates are sapphire, which is transparent to much of the ambient electromagnetic spectrum, is thermally conductive, and is rigid and robust. The thickness of each substrate is 250 microns. The conducting regions **46** are formed by standard photolithography known to those skilled in the art. The width of each conducting stripe, w_2, is 2 microns, and separated by a 2 micron nonconducting region, to form a center-to-center spacing S_L of 4 microns. The stripe has gaps where the spacers **48** are to be formed. The conductive coating **46** is platinum, having a thickness of 40 nm. The spacers **48** consist of silicon dioxide, deposited and patterned by standard means known by those skilled in the art. The total spacer height, d, is 200 nm, its width, w_1, is 5 microns, and the center-to-center spacing, s_1, is 0.5 mm. The spacers are formed by depositing 100 nm thick layers on each substrate, and then joining them. The central conduit regions **43** are cut into the substrates using a standard diamond saw. The cuts are 100 microns in width and 50 microns in depth, forming a conduit that is approximately a 100 micron square. The hole **47** drilled through the upper substrate has a diameter of 1 mm, and is surrounded by a ring having a diameter of 2.5 mm. The ring **49** is affixed to the upper substrate by epoxy. The substrates are pressure bonded together by direct bonding (Plöl, 1999), with the bond forming between the silicon dioxide spacers layers on each substrate.

The steps in the device fabrication that are not described explicitly are well known to those skilled in the art.

Following the calculations presented in the background section, the power produced by a single such device is estimated to be between 1 and 10 watts for an input pressure of 8 atmospheres.

Pumping gas through the Casimir pores requires power. We examine how much power is required, as a check that it is not more than is produced by the device. Consider a Casimir block that contains 200 nm diameter pores over a 1 cm^2 area, having a thickness of 1 cm and a porosity of 0.25. We find the pressure and power required to produce a flux of 1 cm^3 per second at standard temperature and pressure (STP):

According to FIG. **10**(*a*) in a paper by Roy et al. (1993) a pressure drop of 760 torr (equal to one atmosphere) results from a flow of approximately 5 mol/m^2-s through a thickness of 60 microns, which corresponds to a gas velocity of 10 cm/s. Reducing the velocity by a factor of ten, making the appropriate unit conversions and multiplying the result by the thickness ratio of 1 cm (10^4 microns) divided by 60 microns gives the result that a pressure of 1700 Pa, corresponding to 17 atmospheres, is required to produce the desired gas flow. Multiplying this by the gas flux of 1 cm^3 s^{-1} results in a required power of 1.7 milliwatts. These results are only approximate, as temperature and structural variations through the Casimir pores are expected to produce resistance which will then require a somewhat greater pressure. In any case the required power of approximately 1.7 milliwatts is much lower than our estimate of 2.2 to 22 kilowatts of power release, and so much more power is produced than is used to produce the gas flow.

It is to be understood that the dimensions and materials can be varied greatly and still be part of this invention. The following is a list of some such variations, but it is far from exhaustive:

i. The substrates may be other insulating or partially conducting materials, such as silicon, glass, ceramic, plastic, etc.
ii. The conducting stripes can be formed of other conductors, such as copper, aluminum, gold, sliver, silicides, transparent conductors such as indium tin oxide, etc.
iii. Instead of depositing the stripes so that they protrude from the surface and potentially interfere with the gas flow, they may be recessed, either by etching recesses into which the conductors are deposited, or by using planarization techniques to coat an insulating layer between the stripes, using techniques well known in the industry.
iv. The spacer materials can be formed from polymers used, for example, as photoresist and electron-beam resist, from metals, and other materials.
v. Instead of depositing spacers they may be formed by the etching of one or both of the substrates to form grooves.
vi. The spacer height may be from 1 nm to many microns.
vii. The substrates may be bonded by pressure bonding or the use of adhesives, such as cyanoacrylics.
viii. The dimensions of the overall structure may be varied from the distance between a single pair of spacers and conductor/nonconductor region to large plates that are many meters in width.
ix. The individual devices may be sandwiched together to form thick structures. For example, in place of the 250 micron thick substrates, micro-sheet having a thickness of 50 microns or far less may be used so that dense structures are formed.
x. The working fluid may be a wide variety of gases, in addition to the noble gases described earlier, so that all mentions of gas atoms may be extended to molecules of various types.
xi. The working fluid may be a liquid, so that all mentions of gases and gas atoms may be extended to liquids of various types. For operation within approximately of 100° C., one possible liquid is ethylene glycol. For high temperature operation, the liquid can be sodium.

xii. Micro-motors formed using micro-electro-mechanical systems (MEMS) technology can be used to pump the gas through the channels.

xiii. The Casimir cavities may be composed of carbon nanotubes.

xiv. The pattern may be formed using self-assembled layers.

xv. The device may incorporate a naturally formed structure. For example, diatom shells (Goho, 2004a) consist of silicon dioxide patterned with features, including holes, that are tens of nanometers in size. They can be coated as needed with conductors to form Casimir cavities.

xvi. The water bath may be replaced with any other material or device that absorbs substantially the released energy wavelengths. Such materials include glass, organic polymers, thermophotovoltaic devices, among many possibilities known to those skilled in the art.

xvii. Rather than surrounding the entire apparatus, the absorbing material may be placed in the apparatus, for example coating the channels through which the gas flows. Such placement can allow the absorber to reside within roughly an emission wavelength of the gas that is releasing the energy.

Gas Oscillating Through Casimir Channels

The device described in the previous embodiment exposes the gas atoms to a very large number of transitions between Casimir cavity regions (between conducting layers) and exposed regions (without the conducting layers) by pumping them across multiple transitions. Instead of pumping gas through the device, gas atoms can simply be oscillated back and forth between Casimir cavity and exposed regions.

A simple way to visualize this, but not necessarily the most efficient working device, is to consider the device of FIGS. **4**A-**4**D, but with the gaps sealed at the north and south edges. Instead of connecting to tubing via the connector ring, the ring is sealed with a thin metal diaphragm. Before sealing the device it is filled with the desired working gas. An ultrasonic transducer is then mated to the diaphragm. When the ultrasonic transducer is powered, it rapidly compresses and decompressed the gas, causing it to oscillate back and forth between Casimir and exposed regions.

A vertical oscillatory flow device is shown in FIGS. **5**A-**5**C. FIG. **5**A shows a top view, in which many small holes **54** are formed in the substrate surface. The device is surrounded by a connector ring **58**. A magnified cross section of the holes is shown in FIG. **5**B. The holes **54** have a diameter d, a center-to-center spacing s, a depth t_2, and the thickness of a conducting region **56** at the surface is t_1. A central cross section of the entire device is shown in FIG. **5**C. It shows the substrate (holes and conducting layer not shown), the connector ring at the periphery, and a thin diaphragm **57** attached to the top of the connector ring.

The gap and holes are filled with the chosen working gas **59**. An ultrasonic transducer or other source of high frequency vibrations is placed in contact with the diaphragm **57** and powered. This produces gas pressure oscillations that force gas atoms past the Casimir region **55** formed at the top of each hole, alternately in upward and downward directions. Instead of a single conducting layer at the top, multiple alternating conducting and non-conducting layers can be formed at the top of the holes, to multiply the effect. As in the embodiment of FIGS. **4**A-**4**D, the apparatus is surrounded by a means for absorbing the released energy, such as a water bath **24**.

The device is fabricated as follows. The conducting layer **56** is deposited using vacuum deposition, such as sputtering, or from a liquid by anodic or electroless deposition. The layers are patterned by methods known to those skilled in the art, such as electron-beam lithography or photolithography. Alternatively, the holes **54** can be formed using self-assembled monolayers to create the lithography mask, as known to those skilled in the art. The holes are etched to a high aspect ratio, e.g., ratio of depth-to-diameter of 20, such as by ion milling. The outer ring **58** is attached using epoxy, the region is filled with the desired working gas **59**, and the diaphragm **57** is attached with epoxy.

The materials and dimensions in the preferred embodiment are as follows. The substrate **52** is sapphire, and has diameter of 2.54 cm and a thickness of 250 microns. The conducting layer **56** is aluminum, of thickness t_1 of 1 micron. The hole **54** depth t_2 is 4 microns. The hole diameter d is 0.2 microns and center-to-center spacing s is 0.3 microns.

It is to be understood that the shape, dimensions, modulation techniques and materials can be varied greatly and still be part of this invention. The following is a list of some such variations, but it is far from exhaustive:

i. The Casimir cavities may be composed of carbon nanotubes.

ii. The working fluid may be a wide variety of gases, in addition to the noble gases described earlier, so that all mentions of gas atoms may be extended to molecules of various types.

iii. The working fluid may be a liquid, so that all mentions of gases and gas atoms may be extended to liquids of various types. For operation of up to approximately 100° C., one possible liquid is ethylene glycol. For high temperature operation, the liquid can be sodium.

iv. Instead of actively causing the gas atoms to oscillate into and out from the Casimir cavity regions, the oscillations can result from ambient thermal vibrations (e.g., Brownian motion).

v. The configuration of the device can be similar to that of the MEMS device of FIGS. **7**A and **7**B (described as part of a later embodiment), such that the working gas is pushed back and forth between the left-hand and right-hand regions.

vi. The pattern may be formed using self-assembled layers.

vii. The device may incorporate a naturally formed structure. For example, diatom shells consist of silicon dioxide patterned with features, including holes, that are tens of nanometers in size. They can be coated as needed with conductors to form Casimir cavities.

viii. The pumping can be driven by a naturally occurring mechanism. For example, some yeast cell have been found to naturally vibrate at 1.6 kHz (Goho, 2004b). This could be used to cause a gas to oscillate back and forth between Casimir cavity and exposed regions.

ix. The water bath may be replaced with any other material or device that absorbs substantially the released energy wavelengths. Such materials include glass, organic polymers, thermophotovoltaic devices, among many possibilities known to those skilled in the art.

x. Rather than surrounding the entire apparatus, the absorbing material may be placed in the apparatus, for example coating the channels through which the gas flows. Such placement can allow the absorber to reside within roughly an emission wavelength of the gas that is releasing the energy.

Casimir Cavities in Flexible Polymer

Rather than moving the working gas by flowing it (FIGS. **4**A-**4**D) or vibrating it into and out of a Casimir cavity (FIGS. **5**A-**5**C), the cavity wall characteristics can be switched, which results in a shift in the cavity's allowed modes. This produces the same result of tapping vacuum electromagnetic energy that the flowing gas device of the embodiment of FIGS. **4**A-**4**D produces. One way to accomplish this is to put the working gas into gaps formed in flexible photonic crystals. A photonic crystal blocks and passes bands of electromagnetic radiation, where the band wavelength ranges depend upon the material properties and spacing of small repeated structures. A flexible photonic crystal can be formed by embedding an array or rigid objects, such as silicon pillars, in a thin film of flexible polymer. The electromagnetic (or optical) properties of such two-dimensional slab photonic crystal structures is well known to those skilled in the art (Park, 2002).

FIGS. **6**A and **6**B show such a photonic crystal device. FIG. **6**A is a top view, showing metal supports **62** at both ends of a polymer film **64**. The rigid pillars that form the phonic crystal are buried in the polymer. As the film is stretched in the plane of the paper, the pillar spacing in the plane normal to the paper is decreased, which changes the electromagnetic passband. FIG. **6**B is an edge view showing the supports **62**, the polymer film **64**, and gaps in the film that are filled with the working gas **69**. (For clarity, the pillars are not shown.) The gap size is sufficiently narrow to produce a significant Casimir effect, e.g., 200 nm. The length or width need to be sufficiently small to maintain the narrow gap, e.g., 1 micron. The stretching takes place by attaching one support to a stationary object and attaching the other support to a vibrator, such as a piezoelectric crystal, which itself may be attached on its opposing side to another stationary support. As in the embodiment of FIGS. **4**A-**5**D, the apparatus is surrounded by a means for absorbing the released energy, such as a water bath **24**.

It is to be understood that the shape, dimensions, modulation techniques and materials can be varied greatly and still be part of this invention. The following is a list of some such variations, but it is far from exhaustive:

i. Instead of stretching the polymer, it can be modulated with an acoustic signal through the air, or through a liquid that surrounds it.

ii. Instead of stretching the polymer, it can be modulated with an ambient thermal vibrations. As the working gas and the structure heats up, the vibrations increase.

iii. The polymer embedded with rigid pillars may be formed into small spheres that are filled with the working gas. These spheres can fill or partially fill a volume in which the pressure is modulated, either by enclosing the volume and modulating the pressure in the entire volume, by passing an acoustic signal through the volume, or by thermal vibrations. This modulation causes the passband of the photonic crystal that surrounds the working gas to vary. Although the shape of the device is substantially different from that of FIGS. **6**A-**6**B, the function is the same.

iv. The working fluid may be a wide variety of gases, in addition to the noble gases described earlier, so that all mentions of gas atoms may be extended to molecules of various types.

v. The working fluid may be a liquid, so that all mentions of gases and gas atoms may be extended to liquids of various types. For operation of up to approximately 100° C., one possible liquid is ethylene glycol. For high temperature operation, the liquid can be sodium.

vi. The water bath may be replaced with any other material or device that absorbs substantially the released energy wavelengths. Such materials include glass, organic polymers, thermophotovoltaic devices, among many possibilities known to those skilled in the art.

vii. Rather than surrounding the entire apparatus, the absorbing material may be placed in the apparatus, for example in the polymer film through which the gas flows. Such placement can allow the absorber to reside within roughly an emission wavelength of the gas that is releasing the energy.

Modulating Casimir Cavity Wall Spacing

Rather than moving the working gas by flowing it (FIGS. **4**A-**4**D), vibrating it into and out of a Casimir cavity (FIGS. **5**A-**5**C), or switching the characteristics of walls of the cavity to change the passbands (FIGS. **6**A and **6**B), the spacing between the cavity walls can be modulated. This produces the same result of tapping zero point energy that the flowing gas device of the previous embodiments produce. One way to accomplish this is to put the working gas into gaps formed in micro-electro-mechanical systems (MEMS).

MEMS technology makes use of semiconductor lithography techniques to build miniature mechanical devices. The Casimir effect has already been found to be in evidence in MEMS devices. In 2001, Chan and co-workers at Bell Labs Lucent Technologies first demonstrated the effect of the Casimir force in a MEMS device. A gold coated sphere was brought close to a MEMS seesaw paddle, consisting of a polysilicon plate suspended above a substrate on thin torsion rods. The Bell Labs researchers demonstrated the effect of the Casimir force in rocking the plate.

In the current invention we make use of MEMS technology to modulate the spacing between Casimir cavity walls. (Note that we are not making use of the Casimir force to change this spacing, as was done in the Bell Labs demonstration.) The basic MEMS device used to accomplish this is shown in FIGS. **7**A and **7**B. A side view is shown in FIG. **7**A. Two conducting electrodes **76** are shown on the substrate. A pivoting polysilicon plate **74** is shown suspended above the substrate **72**. A conducting layer **77** is formed on the underside of this plate. A top view is shown in FIG. **7**B. The pivoting plate **74** forms the central rectangular region, which is surrounded by a gap **73**. The pivoting arm **75** connects this plate to the surrounding region at the top and bottom of the rectangle. As in the earlier embodiments, the apparatus is surrounded by a means **24** for absorbing the released energy, such as a water bath

The device functions as follows. The working gas fills the region between the pivoting plate **74** and the substrate **72**. A voltage is applied first between the pivoting plate and the left-hand electrode. This causes the distance between the left side of the plate and the substrate to diminish, thereby changing the dimensions of the Casimir cavity formed by these two surfaces. Then the voltage is instead applied between the pivoting plate and the right-hand electrode. This causes the plate to pivot, such that the distance between the right side of the plate and the substrate diminishes, thereby changing the dimensions of the Casimir cavity formed by these two surfaces. The voltage is switched alternately between these two electrodes, causing the plate to oscillate back and forth. The oscillating action is greatly enhanced by the torsion of the pivots, so that very little energy is required to maintain the oscillation.

The techniques to fabricate such a MEMS device is well known to those skilled in the art.

It is to be understood that the shape, dimensions, modulation techniques and materials can be varied greatly and still be part of this invention. The following is a list of some such variations, but it is far from exhaustive:

i. Instead of using a MEMS device, the Casimir cavity can be formed between a substrate and a suspended conducting sheet. A similar technology has been used to form electrostatic acoustic speakers, albeit with larger spacings.

ii. Gaps can be formed in a polymer, with both sides of the gap coated with a conductor and the gap filled with a working gas. The polymer can then be stretched, as in the embodiment of FIGS. **6**A and **6**B, such that the spacing of the Casimir cavity formed by the two conductors is modulated. A figure of this would appear much like that depicted in FIG. **6**B

viii. Instead of stretching the polymer, it can be modulated with an acoustic signal through the air, or through a liquid that surrounds it.

ix. Instead of stretching the polymer, it can be modulated with an ambient thermal vibrations. As the working gas and the structure heat up, the vibrations will increase.

x. The polymer coated on its interior surface with a conductor may be formed into small spheres that are filled with the working gas. These spheres can fill a volume in which the pressure is modulated, either by enclosing the volume and modulating the pressure in the entire volume, by passing an acoustic signal through the volume, or by thermal vibrations. This modulation causes the spacing of the Casimir cavity in which the working gas is contained to vary. Although the shape of the device is substantially different from that of FIGS. **7**A and **7**B, the function is the same.

xi. The working fluid may be a wide variety of gases, in addition to the noble gases described earlier, so that all mentions of gas atoms may be extended to molecules of various types.

xii. The working fluid may be a liquid, so that all mentions of gases and gas atoms may be extended to liquids of various types. For operation of up to approximately 100° C., one possible liquid is ethylene glycol. For high temperature operation, the liquid can be sodium.

xiii. The water bath may be replaced with any other material or device that absorbs substantially the released energy wavelengths. Such materials include glass, organic polymers, thermophotovoltaic devices, among many possibilities known to those skilled in the art.

xiv. Rather than surrounding the entire apparatus, the absorbing material may be placed in the apparatus, for example coating the substrate and cap of the region containing the gas. Such placement can allow the absorber to reside within roughly an emission wavelength of the gas that is releasing the energy.

We note that the MEMS device of FIGS. **7**A and **7**B can also be used to move the working gas back and forth between the left-hand and right-hand regions. This function is consistent with the embodiment of FIGS. **5**A-**5**C, in which the working gas is vibrated into and out of a Casimir cavity.

Assymetric Casimir Cavity Entry and Exits Including Absorbing Means

As a prelude to this embodiment, we review the processes involved in the present invention. A general concept of this entire invention is that a gas that is in equilibrium with the ambient electromagnetic modes, which include the vacuum field (also known as the zero point field), is caused to enter a Casimir cavity. For the purposes of this entire invention a Casimir cavity is defined as any region in which the electromagnetic modes are restricted. Upon approaching this region, the electromagnetic modes that the space supports are restricted and the energy of the electron orbitals of the gas atoms is reduced. As a consequence of this reduction the excess energy is emitted and absorbed by the apparatus, providing heat energy. By the time the atoms are in the Casimir cavity, nearly all the excess energy has been radiated (unless the gas flow is extremely fast). The gas atoms pass through the Casimir cavity, and upon emerging from this region to a region that supports a broader range of electromagnetic modes, the energy of the electron orbitals of the gas atoms is again allowed to rise to its previous value. The compensation for the energy deficit is provided from the ambient electromagnetic modes.

One of the tenets of the current invention is that excess energy released when the gas approaches the Casimir cavity is delivered locally and that the energy deficit that must be compensated for when it emerges from the cavity is supplied from global sources. In this way the ambient electromagnetic field is tapped to provide usable energy. There may be conditions in which it is possible that the excess energy release and the deficit energy supply are both local, in which case no net energy is provided. Similarly, there may be conditions in which it is possible that the excess energy release and the deficit energy supply are both global, in which case again no net energy is provided. To avoid these possibilities, we provide an asymmetry in the apparatus to ensure that the excess energy is released locally and that the energy deficit is supplied globally.

The concept of embodiment is shown in FIG. **8**. This figure depicts a channel **88**, similar to that shown in some of the earlier embodiments. Gas is constricted between two substrates **82** and **83** and flows through the channel in the direction of the arrows. As in the previous cases, gas flows from a region in which the substrate is not coated **87** with a conducting layer to a region in which it is **86**. The difference here is that an intermediate region **84** is provided in which the substrates are coated with an absorbing layer. This absorbing region absorbs the excess energy that is radiated from the atoms as they approach the Casimir cavity (conducting) region. The absorbing region is not substantially conducting, and therefore does not substantially restrict the electromagnetic modes that are supported in the region. Upon exiting the Casimir cavity (conducting) region, the atoms pass immediately into another region with no absorbing region **87**. Thus upon approaching the Casimir cavity the atoms are forced to deliver their excess energy locally because it is absorbed by the absorbing region **87**. Upon emerging from the Casimir cavity the gas atoms are forced to supply their energy deficit non-locally, i.e., globally, because there is no local source for this energy.

As an option, a further aspect of this invention is to situate the absorbing region within roughly one emission wavelength of the gas atoms at the time that they are emitting. No such layer is provided within such a distance when the gas atoms emerge from the Casmir cavity and are supplied with energy. The substrate is chosen such that it does not absorb the emission wavelengths.

The absorbing layers may comprise glass (amorphous silicon dioxide, usually with impurities), and the substrate may comprise sapphire. The glass has a much broader absorption band in the far infrared than does the sapphire. A wide range of other materials may be provided to form the absorbing layers and non-absorbing or less absorbing substrate. Such materials are known to those skilled in the art, and are available in tables and handbooks.

The sequence of regions depicted in FIG. **8** may be repeated to form the sort of multiply striped structure described in the embodiment of FIGS. **4**A-**4**D.

The dimensions of the channel and the apparatus are approximately the same as those of embodiment of FIGS. **4**A-**4**D. Similarly the attachments to provide for gas flow, the spacers, and other aspects of the apparatus are similar to those described in embodiment of FIGS. **4**A-**4**D. The conducting layer length is chosen so that the emerging atoms do not have substantial access to radiation emitted from the absorbing regions. Note that, unlike embodiment of FIG. **2**, it is not necessary to surround the apparatus with a means for absorbing the released energy **24**, such as a water bath.

The device fabrication is not described explicitly as it is well known to those skilled in the art.

It is to be understood that the dimensions and materials can be varied greatly and still be part of this invention. The following is a list of some such variations, but it is far from exhaustive:

i. The substrates may be other insulating or partially conducting materials, such as silicon, glass, ceramic, plastic, etc.

ii. The conducting stripes can be formed of other conductors, such as copper, aluminum, gold, sliver, silicides, transparent conductors such as indium tin oxide, etc.

iii. The stripes may be recessed in the substrate or they protrude from the surface.

iv. The individual devices may be sandwiched together to form thick structures. For example, in place of the 250 micron thick substrates, micro-sheet having a thickness of 50 microns or far less may be used so that dense structures are formed.

v. The working fluid may be a wide variety of gases, in addition to the noble gases described earlier, so that all mentions of gas atoms may be extended to molecules of various types.

vi. The working fluid may be a liquid, so that all mentions of gases and gas atoms may be extended to liquids of various types. For operation within approximately of 100° C., one possible liquid is ethylene glycol. For high temperature operation, the liquid can be sodium.

vii. Micro-motors formed using micro-electro-mechanical systems (MEMS) technology can be used to pump the gas through the channels.

viii. The Casimir cavities may be composed of carbon nanotubes.

ix. The pattern may be formed using self-assembled layers.

The device may incorporate a naturally formed structure. For example, diatom shells consisting of silicon dioxide patterned with features, including holes, that are tens of nanometers in size. To the extent necessary, these can be coated as needed with conductors to form Casimir cavities.

REFERENCES

Boyer, T. H. 1975, Random Electrodynamics: The Theory of Classical Electrodynamics with Classical Zero-Point Radiation Field, Phys. Rev. D, 11, 790.

Cole, D. C. and Puthoff, H. E. 1993, Extracting energy and heat from the vacuum, Phys. Rev. E, 48, 2, 1562.

Cole, D. C. and Zou, Yi 2003, Quantum Mechanical Ground State of Hydrogen Obtained from Classical Electrodynamics, Physics Letters A, Vol. 317, No. 1-2, pp. 14-20 (13 Oct. 2003), quant-ph/0307154.

Cole, D. C. and Zou, Yi 2004, Analysis of Orbital Decay Time for the Classical Hydrogen Atom Interacting with Circularly Polarized Electromagnetic Radiation, Phys. Rev. E. 69 (1), 016601, pp. 1-12 (2004).

Forward, R. 1984, Extracting electrical energy from the vacuum by cohesion of charged foliated conductors, Phys. Rev. B, 30, 4, 1700.

Goho, A., "Diatom Menagerie," Science News, Vol. 166, Jul. 17, 2004a, pp. 42-44, and references mentioned therein.

Goho, A., "Rattle and Hum; Molecular machinery makes yeast cells purr," Science News, Vol. 166, Aug. 21, 2004b, p. 116, and references therein.

Milonni, P. W. The Quantum Vacuum: An Introduction to Quantum Electrodynamics, Sect. 3.3, Academic Press, 1994.

Park, W., and Summers, C. J., "Extraordinary refraction and dispersion in 2D photonic crystal slabs", Opt. Lett., Vol 27, (2002) p. 1397.

Plöl, A. and Kräuter, G., "Wafer direct bonding: tailoring adhesion between brittle materials", Mater. Sci. Eng. R, Vol 25 (1999) p. 1.

Puthoff, H. E., Ground State of Hydrogen as a Zero-Point-Fluctuation-Determined State, Phys. Rev. D 35, 3266, 1987.

Roy, Subrata et al., "Modeling gas flow through microchannels and nanopores," Journal of Applied Physics, Vol. 93, No. 8, Apr. 15, 2003, pp. 4870-4879.

What is claimed is:

1. A system for extracting and collecting electromagnetic radiation from the ambient surroundings, comprising:

(a) a supply of fluid characterized by its ability to (i) take in electromagnetic radiation from the ambient surroundings and (ii) release at least some of said energy when the fluid is caused to pass into a Casimir cavity;

(b) a first arrangement configured to collect at least some of the electromagnetic radiation released by said fluid;

(c) a second arrangement including means defining a given path for containing said fluid along said path;

(d) a third arrangement including a Casimir cavity positioned within said given path and cooperating with said second arrangement such that said fluid is caused to pass into and out of the cavity as the fluid is contained along said given path, said Casimir cavity being positioned in sufficient communication with the ambient surroundings and with said first arrangement so as to (i) cause said fluid containing electromagnetic energy taken from the ambient surroundings to release at least some of said energy to said first arrangement when the fluid passes into said cavity and (ii) to again take in electromagnetic energy from the ambient surroundings when the fluid passes out of said cavity.

2. A system according to claim **1** wherein said second arrangement is configured such that said fluid is caused to flow along said path into and out of said Casimir cavity.

3. A system according to claim **1** wherein said second and third arrangements are configured such that said Casimir cavity is caused to move with respect to said fluid such that the fluid is in turn caused to pass into and out of said Casimir cavity.

4. A system according to claim **3** wherein said third arrangements configured so as to cause said Casimir cavity to move back and forth between first and second spaced apart positions.

5. A system according to claim **1** wherein said means defining said given path defines a closed passageway for containing said fluid and wherein said second arrangement is configured such that the same fluid is caused to cycle into and out of said Casimir cavity.

6. A system according to claim **5** wherein said passageway defines a looped path and wherein said second arrangement includes a mechanism configured to cause said fluid to flow around said path through said passageway into and out of said Casimir cavity.

7. A system according to claim **5** wherein said second arrangement includes a mechanism for causing said fluid to flow back and forth through said passageway into and out of said Casimir cavity.

8. A system according to claim **7** wherein said absorbing material is a liquid.

9. A system according to claim **8** wherein said liquid material is water.

10. A system according to claim **1** wherein said fluid is a gas.

11. A system according to claim **10** wherein said gas is a monatomic gas.

12. A system according to claim **10** wherein said gas is a molecular gas.

13. A system according to claim **1** wherein said first arrangement includes a container of material for absorbing electromagnetic energy, said absorbing material surrounding at least said Casimir cavity.

14. A system according to claim **1** wherein said Casimir cavity includes opposing walls and wherein said third arrangement is configured so as to cause the position of said Casimir cavity walls to move back and forth between first and second spaced positions.

15. A system for extracting and collecting electromagnetic energy from the ambient electromagnetic quantum vacuum, comprising:

(a) a first arrangement defining at least one Casimir cavity configured to cause gas containing electromagnetic energy obtained from the ambient electromagnetic quantum vacuum to release at least some of said energy when said gas is passed into said cavity;

(b) a second arrangement located in the ambient electromagnetic quantum vacuum and including a source of said gas and a mechanism cooperating with said first arrangement so as to cause said gas to pass from the ambient electromagnetic quantum vacuum into said Casimir cavity and then out of said cavity and back into the ambient electromagnetic quantum vacuum, whereby the gas when passing into said Casimir cavity releases at least some of its energy and then, upon passing back into the ambient electromagnetic quantum vacuum, again takes in electromagnetic energy from the ambient electromagnetic quantum vacuum, said means and said first arrangement cooperating with one another such that said fluid passes into and out of said Casimir cavity by relative movement between the cavity and gas; and

(c) a third arrangement for capturing at least some of the electromagnetic energy released by said fluid, said third arrangement including means located in a position with respect to said Casimir cavity such that at least some of the electromagnetic energy released by said gas is captured by said absorber.

16. A system, comprising:

(a) a first arrangement including a number of Casimir cavities, each of which is configured to cause fluid containing electromagnetic energy obtained from the ambient surroundings to release at least some of said energy when said fluid is passed into said cavity;

(b) a second arrangement located in the ambient surroundings and including a source of said fluid and means cooperating with said first arrangement for causing said fluid to pass from the ambient surroundings into each of said Casimir cavities and then out of the cavity and back into the ambient surroundings, whereby the fluid when passing into said Casimir cavities releases at least some of its energy and then, upon passing back into the ambient surroundings, again takes in electromagnetic energy from the ambient surroundings, said means and said first arrangement cooperating with one another such that said fluid passes into and out of said Casimir cavities by relative movement between the cavities and fluid; and

(c) a third arrangement for capturing at least some of the electromagnetic energy released by said fluid.

17. A system according to claim **16** wherein said means includes at least one fluid passageway extending from the ambient surroundings into and though said Casimir cavities and back into the ambient surroundings and wherein said Casimir cavities are defined by a series of conducting strips located within said passageway, said series of conducting strips including a first group of spaced apart strips located on one side of the passageway and a second groups of spaced apart strips on a opposite side of said passageway in alignment with respective strips of said first group, each of said aligned pair of strips being positioned relative to one another to produce a Casimir cavity.

18. A method, comprising:

(a) providing a first arrangement defining at least one Casimir cavity configured to cause fluid containing electromagnetic energy obtained from the ambient surroundings to release at least some of said energy when said fluid is passed into said cavity;

(b) providing a source of said fluid;

(c) causing said fluid to pass from the ambient surroundings into said Casimir cavity and then out of said cavity and back into the ambient surroundings such that the fluid when passing into said Casimir cavity releases at least some of its energy and then, upon passing back into the ambient surroundings, again takes in electromagnetic energy from the ambient surroundings, said fluid being cause to pass into and out of said Casimir cavity by relative movement between the cavity and fluid; and

(c) capturing at least some of the electromagnetic energy released by said fluid.

19. A method of extracting and collecting electromagnetic radiation from the ambient surroundings, comprising:

(a) providing a supply of fluid characterized by its ability to (i) take in electromagnetic radiation from the ambient surroundings and (ii) release at least some of said energy when the fluid is caused to pass into a Casimir cavity;

(b) providing a first arrangement configured to collect at least some of the electromagnetic radiation released by said fluid;

(c) providing a second arrangement including means defining a given path for containing said fluid along said path;

(d) providing a third arrangement including a Casimir cavity positioned within said given path;

(e) causing said fluid to pass into and out of the cavity as the fluid is contained along said given path; and

(f) positioning said Casimir cavity in sufficient communication with the ambient surroundings and with said first arrangement so as to (i) cause said fluid containing electromagnetic energy taken from the ambient surroundings to release at least some of said energy to said first arrangement when the fluid passes into said cavity and (ii) to again take in electromagnetic energy from the ambient surroundings when the fluid passes out of said cavity.

20. A system, comprising:

(a) a first arrangement defining at least one mechanism designed to cause the atoms and molecules making up a given fluid containing electromagnetic energy obtained from the ambient surroundings to change in configuration in a way which releases at least some of said energy when said fluid is passed into said mechanism;

(b) a second arrangement located in the ambient surroundings and including a source of said fluid and means cooperating with said first arrangement for causing said fluid to pass from the ambient surroundings into said mechanism and then out of said mechanism and back into the ambient surroundings, whereby the fluid when passing into said mechanism releases at least some of its energy and then, upon passing back into the ambient surroundings, again takes in electromagnetic energy from the ambient surroundings, said means and said first arrangement cooperating with one another such that said fluid passes into and out of said mechanism by relative movement between the mechanism and fluid; and

(c) a third arrangement for capturing at least some of the electromagnetic energy released by said fluid.

* * * * *

FM Crystal Radios?

copyright © 2006 by Larry J Solomon

I have heard, even from a physicist, that it is impossible to build FM crystal radios. On the other hand some experimenters claim that they have built them. This argument intrigued me to try and build an FM crystal radio, which I have done successfully. To my surprise, the result is an astounding performer, pulling in four local stations in Tucson. When connected as a receiver to a good sound system the sound fidelity is as good or better than more expensive AM radios. In fact, it sounds "high-fidelity".

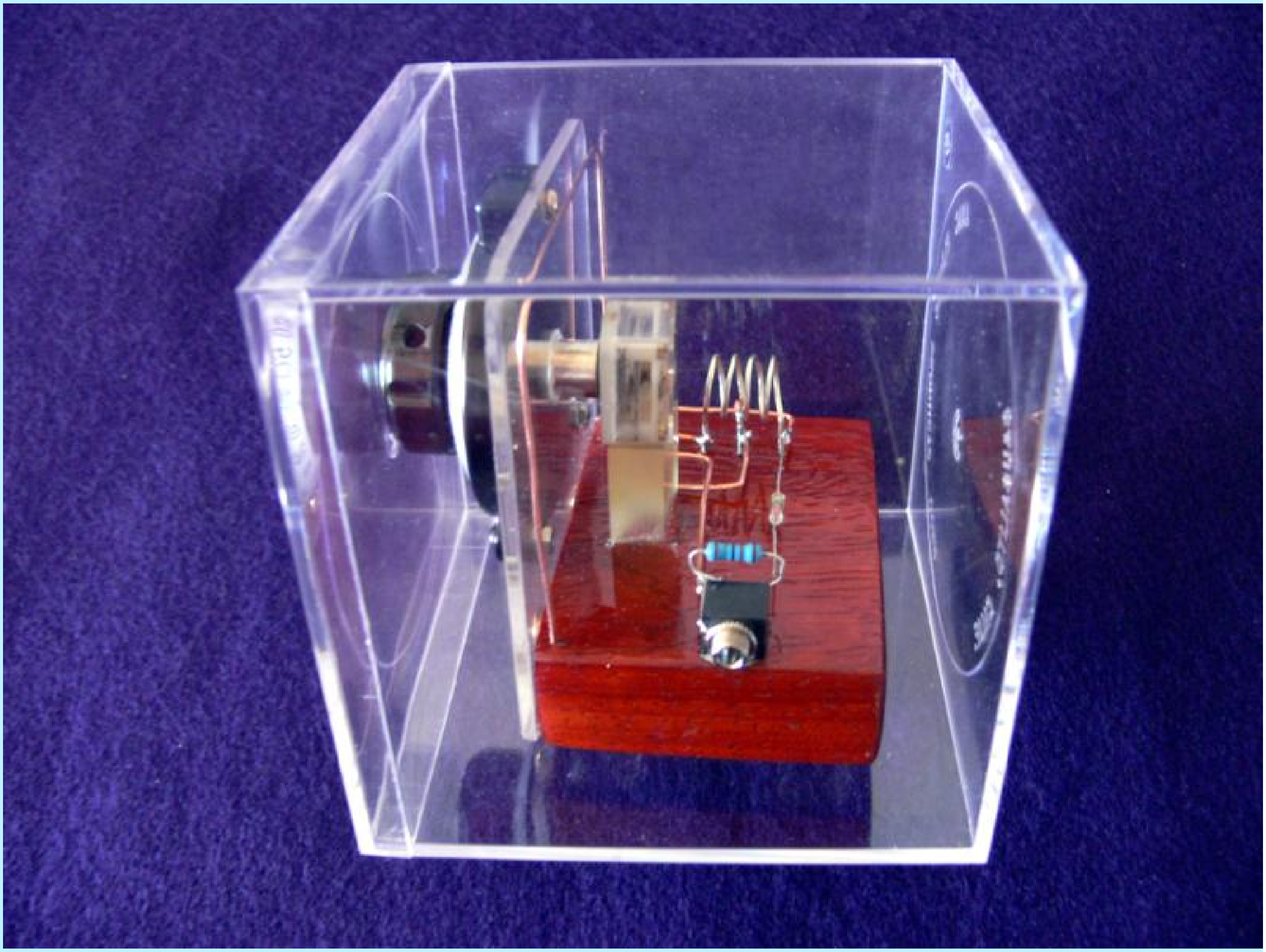

This picture shows the Solomon FM Crystal Set in an acrylic display case. I made the set specifically to fit inside this case (the case came first).

My definition of a crystal radio is one that is not powered, except by the radio transmission itself and employs a crystal detector. So, it should work without any batteries or AC power. An FM crystal receiver must be able to detect and receive FM signals well enough to be heard in earphones without any such extra power.

This FM receiver is an amazing performer. It has crystal clear reception (pun intended), good sensitivity, but only fair selectivity. This set was a discovery for me. I started out by designing and building the normal AM sets. Then one day while testing the "Mystery" set (see my other web links), to my surprise, in addition to the expected panoply of AM stations, I heard a very faint signal that I could not tune out. At first, it seemed too weak to identify. When I tuned out all the AM stations, I was astonished to hear the announcement "KiiM FM, 99.5"! This is a

country music FM station here in Tucson. It was all over the dial, untunable, but the much louder AM signals masked it when they were tuned in.

I set myself the task of trying to improve the FM reception. I tried some simple circuit modifications that did not seem to improve anything. Then I connected a dipole antenna instead of the AM antenna I normally use. Suddenly, the FM signal was much clearer, although still weak. By using the audio output and sound system amplifier, I was even more amazed that four different FM stations came in loud (or rather medium) and clear. I found that changing the telescoping antenna length and position I could tune the stations in and out. They were KRQ, KLPX, KiiM, and KHYT all local FM stations with transmitters nearby. Their reception was also affected by the length and position of the audio output cable.

After doing some research, I discovered that there was a physical theory that claimed that FM reception was possible and even probable using the same circuit as an AM receiver. The theory is called "slope detection". So, I set out to find circuit improvements. A web search yielded little, mostly theory. But there was enough information that I thought I could make some modifications to the AM circuits to make them more tunable to FM signals and less tunable to AM. Since FM operates at higher frequencies, all I had to do, I thought, was make the coil and caps smaller. After much "tinkering" I arrived at the current circuit.

The circuit looks identical to a classic AM crystal circuit but is even simpler to build. The components were reduced in dimension to resonate at higher frequencies. This was done by experimenting with smaller and smaller coils and capacitors. The antenna is also much reduced in size (from that of AM) to resonate at higher frequencies (the antenna is crucial). The air variable capacitor I used has two trimmers in it which should be adjusted for best reception.I have found that a commonly available vernier dial and knob will fit the capacitor nicely. See end of article for a picture of the variable. C3 is a ceramic capacitor of 18 pf, but may be anywhere from 10 to 50pf. A detected FM signal is converted to AM due to an effect called slope detection that modulates amplitude.

This FM Crystal Set works best near the transmitter (I have not tested it beyond about 10 miles). Secondly, the sound level is quiet, especially without an amplifier. A quiet room is needed for listening with earphones. One must be willing to move the set around to find a location for the best reception of signals. However, in addition to listening with high impedance earphones (crystal or otherwise), the set can be connected directly to an audio amplifier's low level magnetic input which can then play amplified through a sound system at any volume -- sounds GREAT. In fact, I recommend starting tests with the FM crystal set by connecting it to the low-level phono inputs of a receiver or preamplifier. (Nowadays, many receivers don't even have a phono input!) That way you can crank up the volume, which makes it more likely to find the FM stations. If no signals are detected, I also recommend connecting an external "rabbit ear" antenna or hanging a short wire (12 inches or so) in various positions next to the internal antenna. The variable length of rabbitt ears can help to tune in stations.

No additional wiring or antenna is necessary (the antenna is optimized in length for FM.)

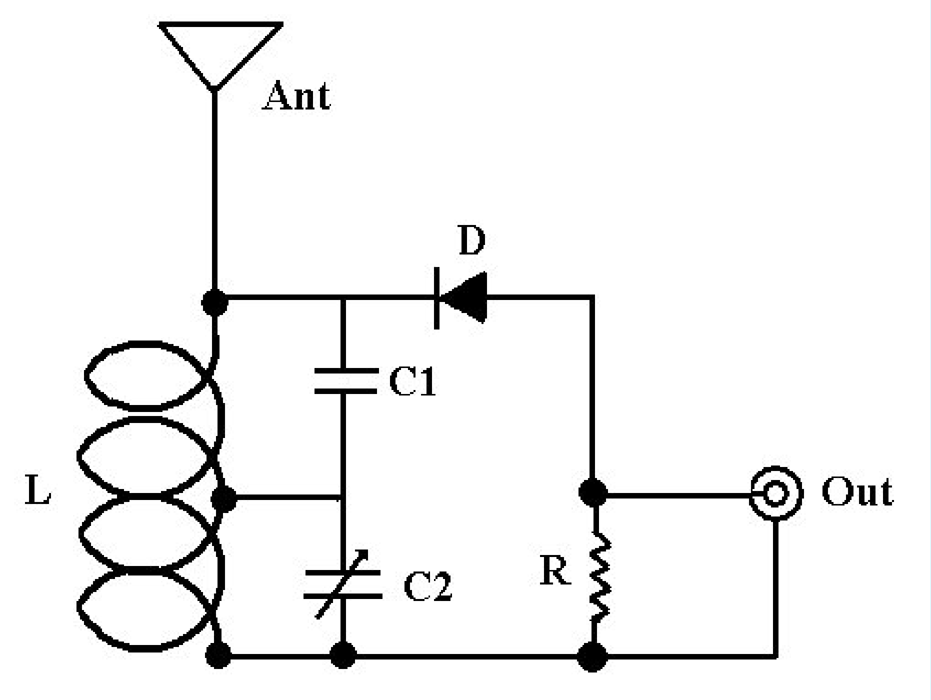

L - 4 turns #18 copper or silver wire, 12mm inside diameter, tapped at 2.5 turns
Ant - 7 inches of #18 bare copper wire
C1 - 18 pf ceramic capacitor
C2 - 50 pf air variable capacitor
D - 1N34 diode or rock crystal
R - 150K resistor

The diode is tapped directly to the antenna. The vernier dial fits directly on the tuning capacitor. The antenna parallels the perimeter of the acrylic face plate. "Military style" #18 AWG wiring is used without any insulation. It is important to keep the components physically close together. The component specifications are the same as in circuit #2. The coil is silver rather than copper, but copper does just as well. I think that the contrast of the silver and copper is beautiful. The coil was wrapped around a Sharpie Permanent Marker, then slipped off and expanded slightly. The wooden base is made from lacquered, polyurethane padouk.

I consider this set a work of art as well as science and think it is the most elegant crystal receiver I have created. I love the contrast of the silver coil, the copper antenna, the clear acrylic faceplate, the black vernier dial, the white and transparent variable capacitor, and the subtle colorings on the resistor, the diode, and the lucite base. Yet the circuit is so ridiculously simple that some will not believe it is possible without building it themselves. No shielding is necessary, and there is no problem with hand capacitance. However, the output cable position may affect reception sensitivity.

Photos of wired circuit

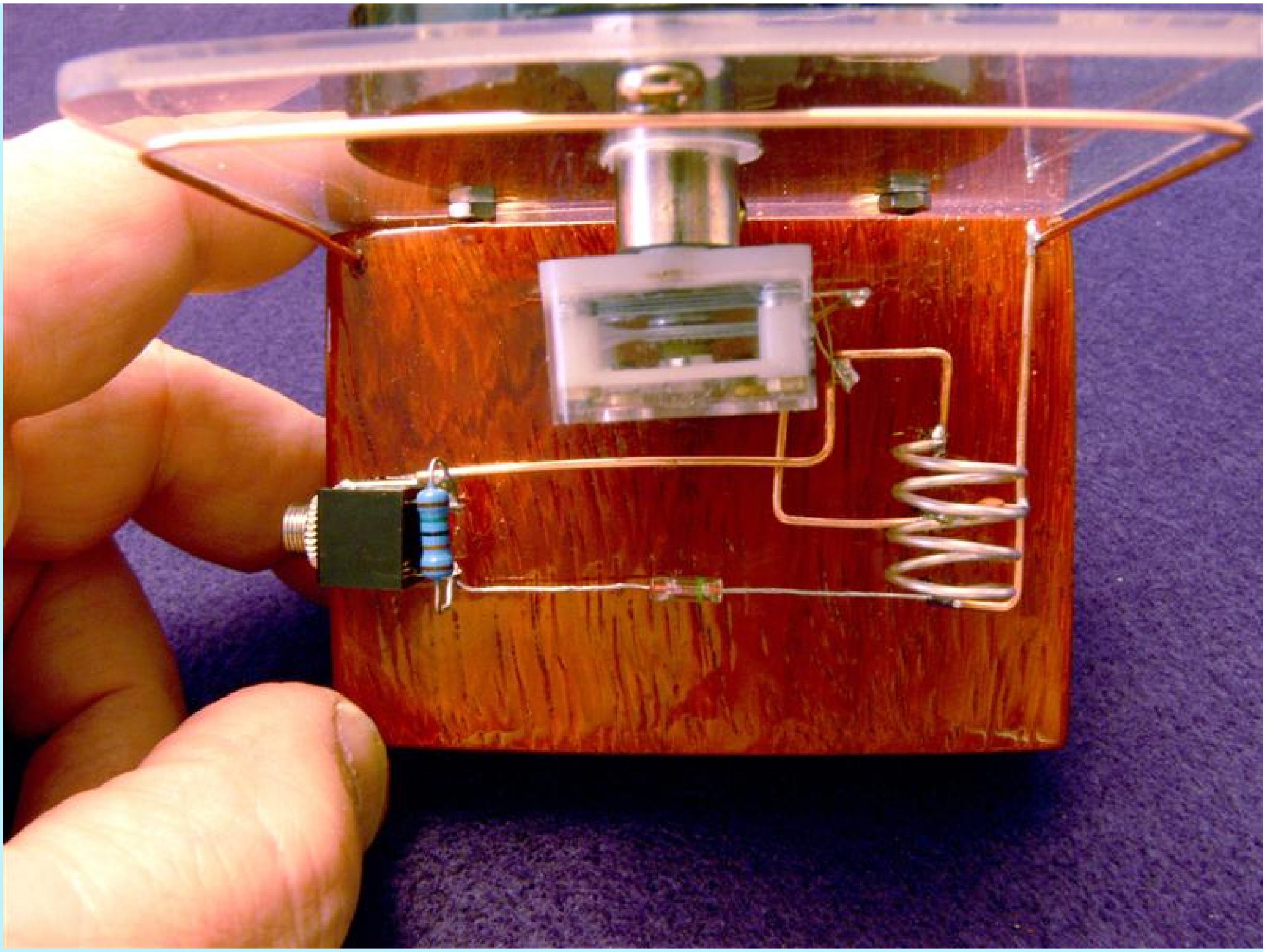

A hand is included in this photograph to show scale. Note the military style wiring, diode, and antenna. I wanted the wiring to create a modern design similar to a Mondrian painting. Not only is this set beautiful, it works! No power and no long antenna! It looks like a work of fiction.

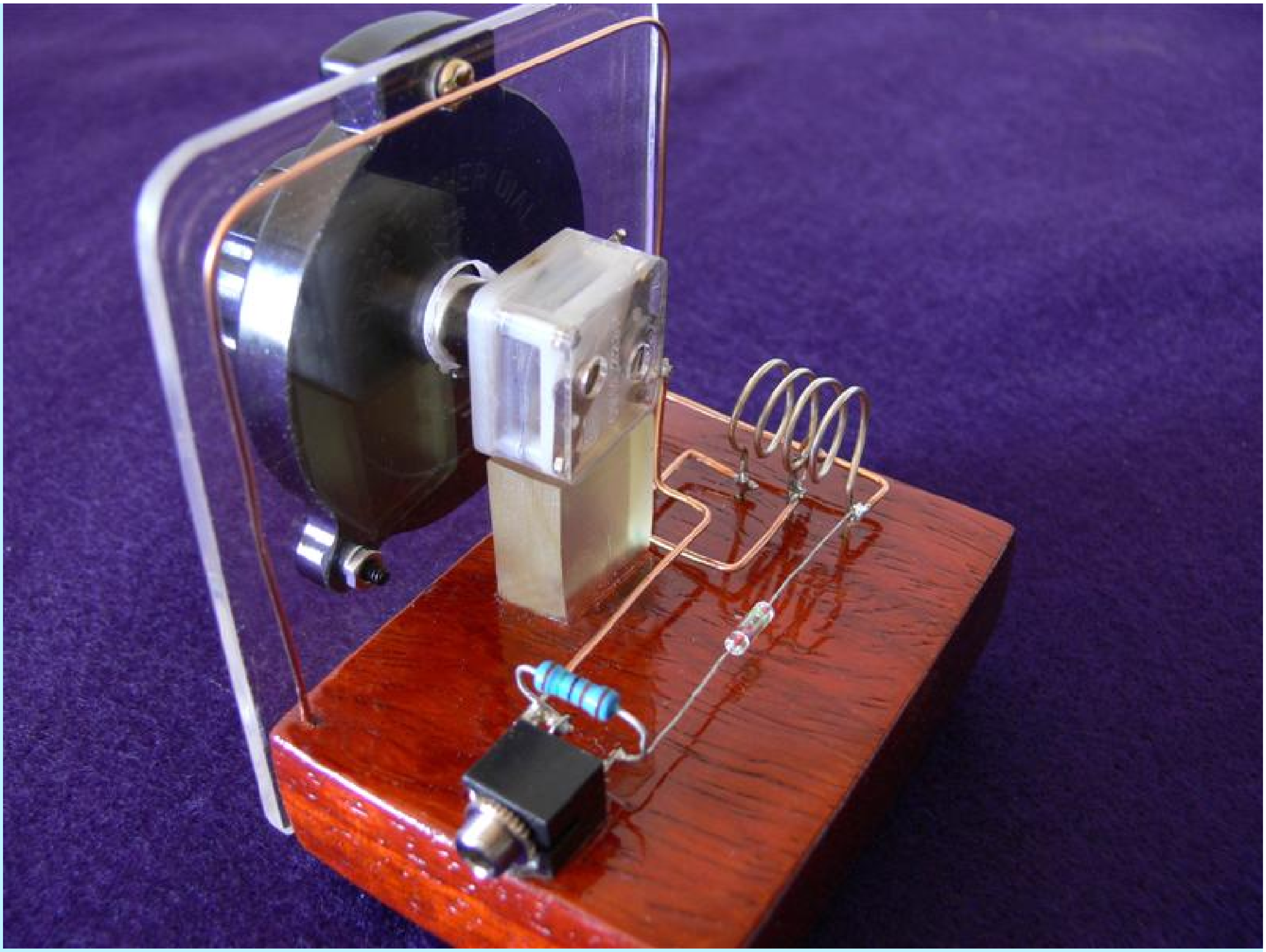

Is this thing imaginary -- science fiction? Well, imagination did play a part, but it is definitely not science fiction. This shot shows the elegance of the FM set best, I think. There is only one resistor and one fixed capacitor.

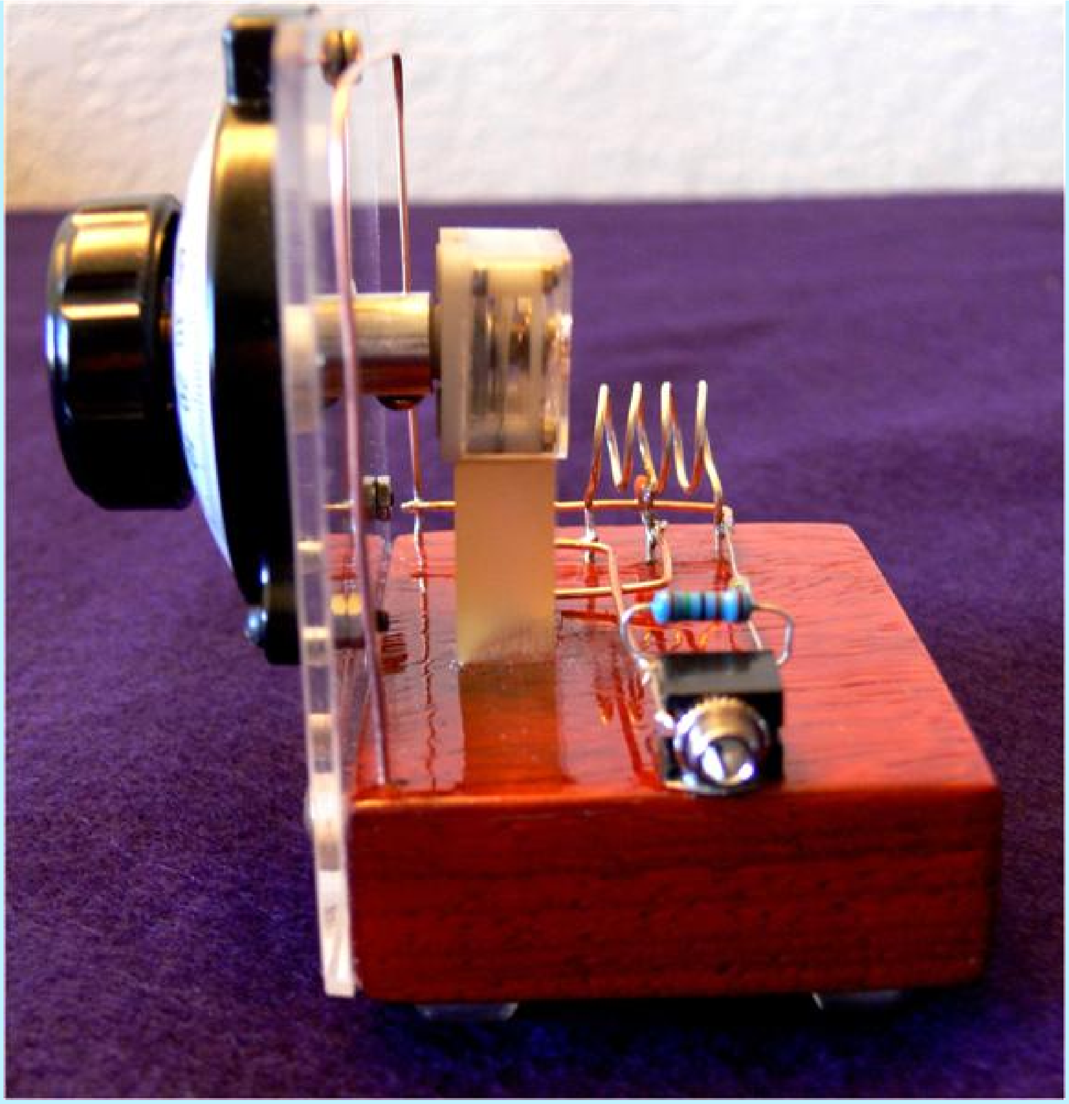

The inside of the tuning capacitor and the phono jack/output can be seen here. Can you spot the fixed ceramic capacitor? Note the polished edge of the face plate and the reflection in the wooden base.

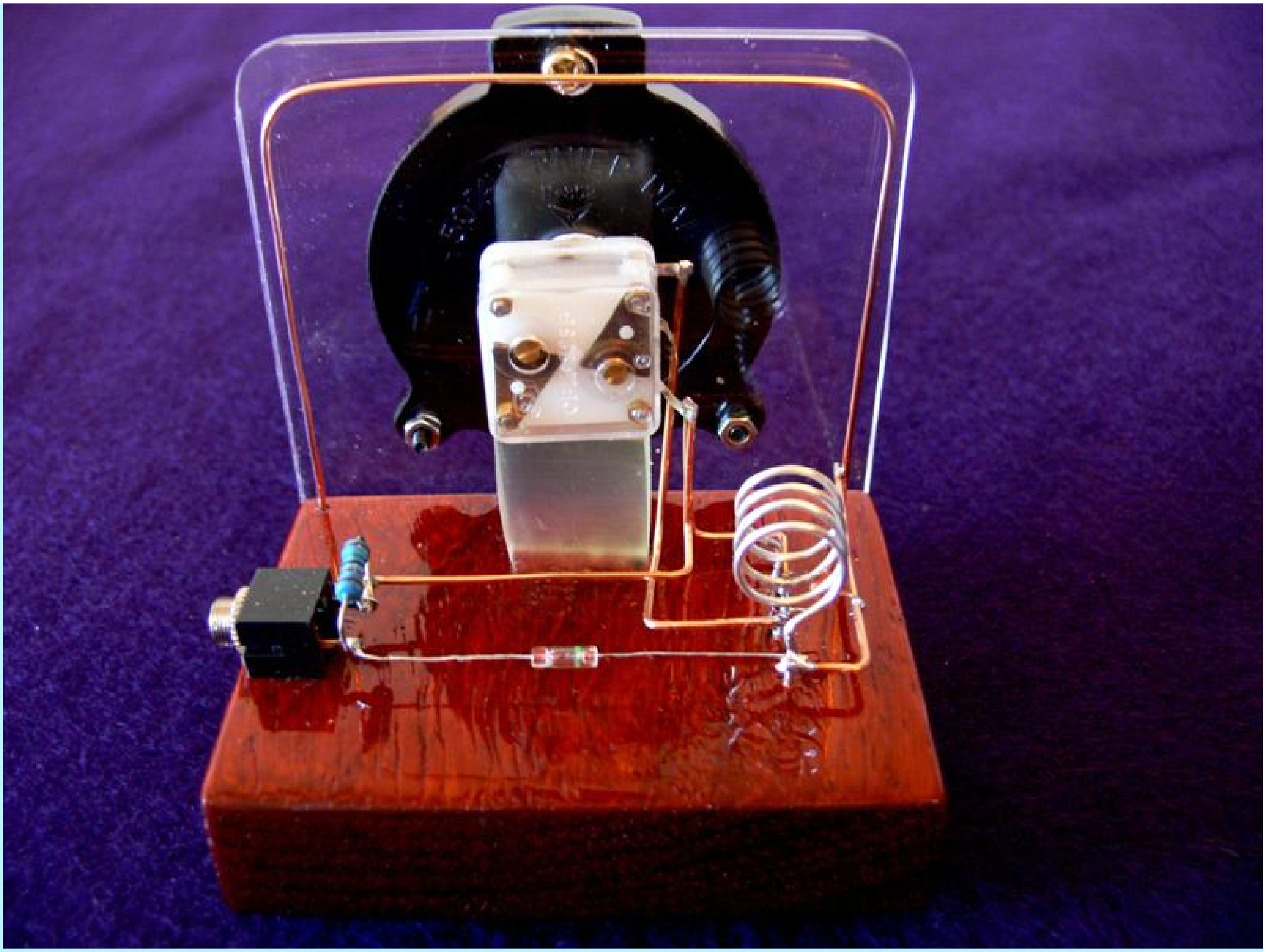

A quarter-inch piece of lucite was fitted under the tuning capacitor to anchor it. Note the two tiny trimmers on the back of the tuning capacitor. Brass screws were used to enhance appearance.

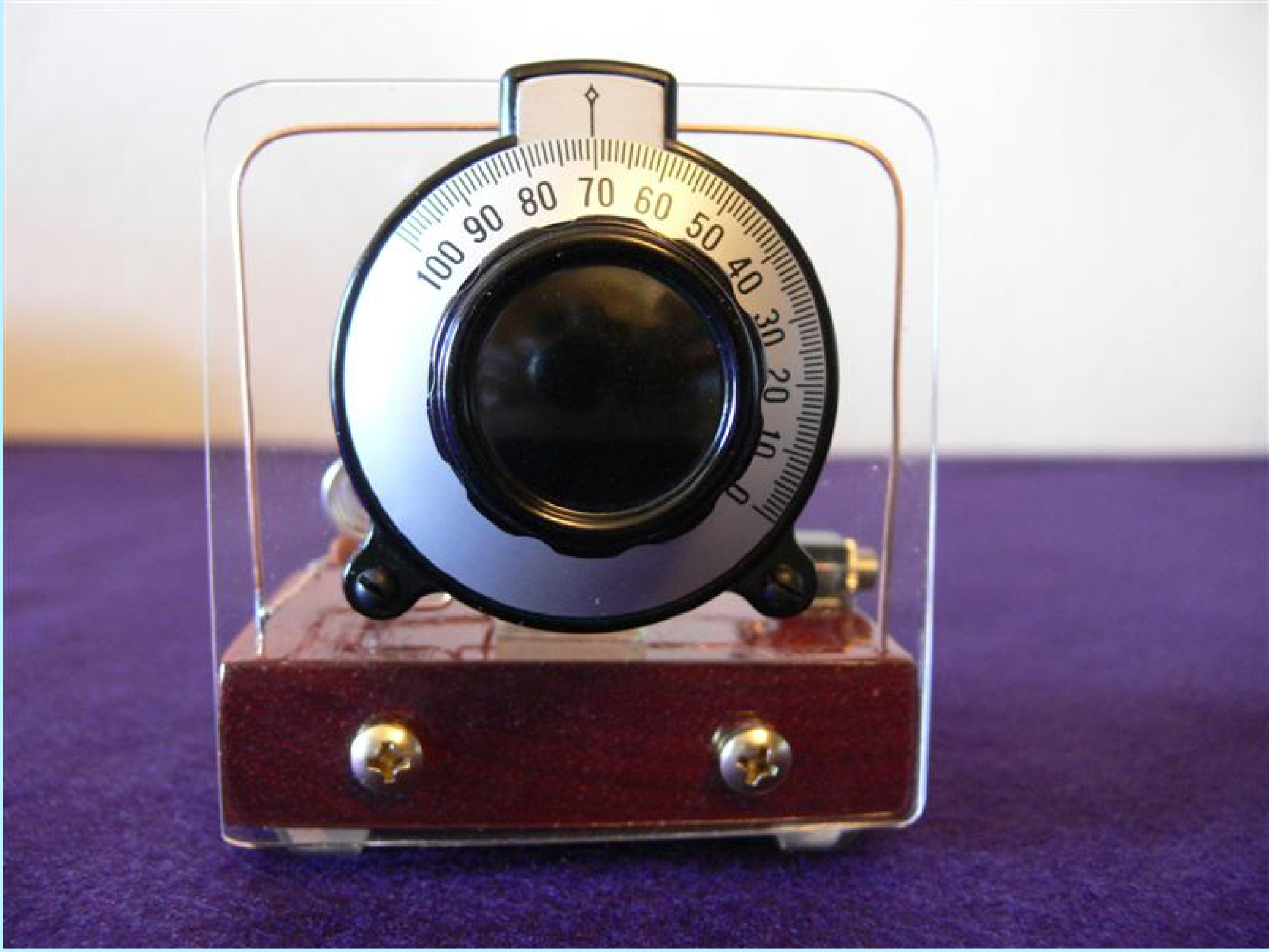

The vernier dial is large to accomodate ease of tuning, and the vernier makes it easy to separate stations. Two golden (brass) wood screws fix the face plate to the base. Holes for the face plate were made with special plastic drills, but ordinary drills may be used if drilled very SLOWLY. The knob is removable.

FM Crystal Circuit #2

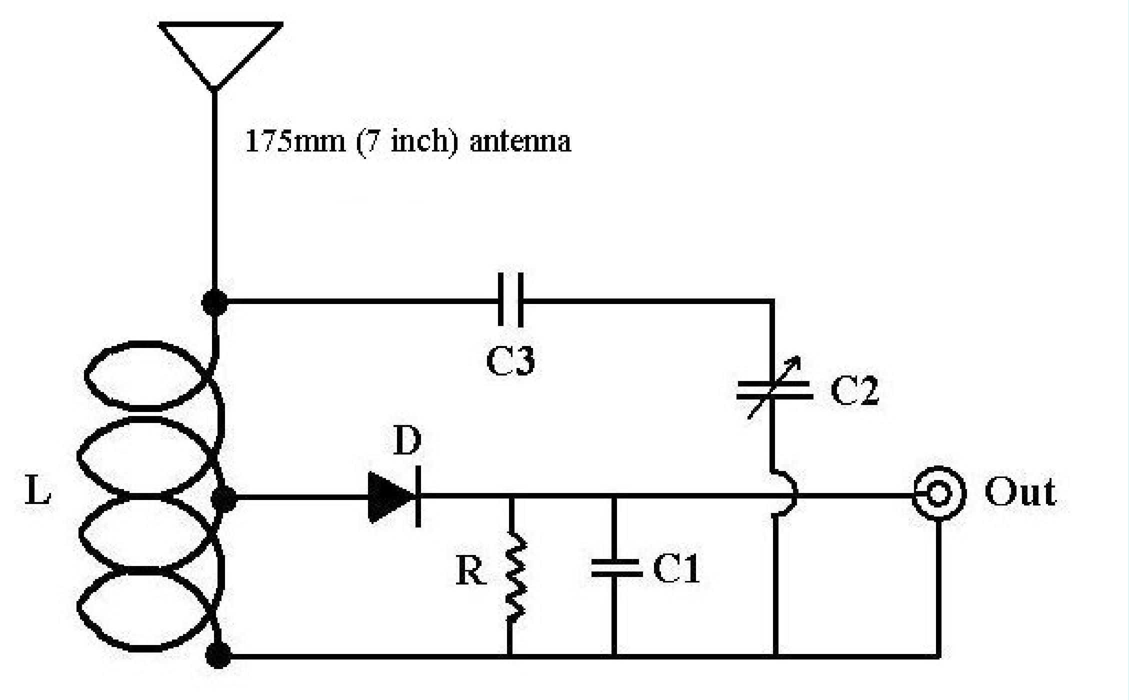

L - 5 turns AWG#18 bare copper or silver wire, 12mm inside diameter, tapped at 2.5 turns
D - 1N34 or rock crystal diode
C1 - 82 pf capacitor
C2 - 80 pf air variable capacitor
C3 - 18 pf capacitor
R - 150K resistor

The following photographs show the circuit wired with the handmade Saturn Dial. and knob. It is perhaps not as visually striking as set No. 1, but it works just as well. In fact, this set was the original version. Notice that all the wiring and coil are copper.

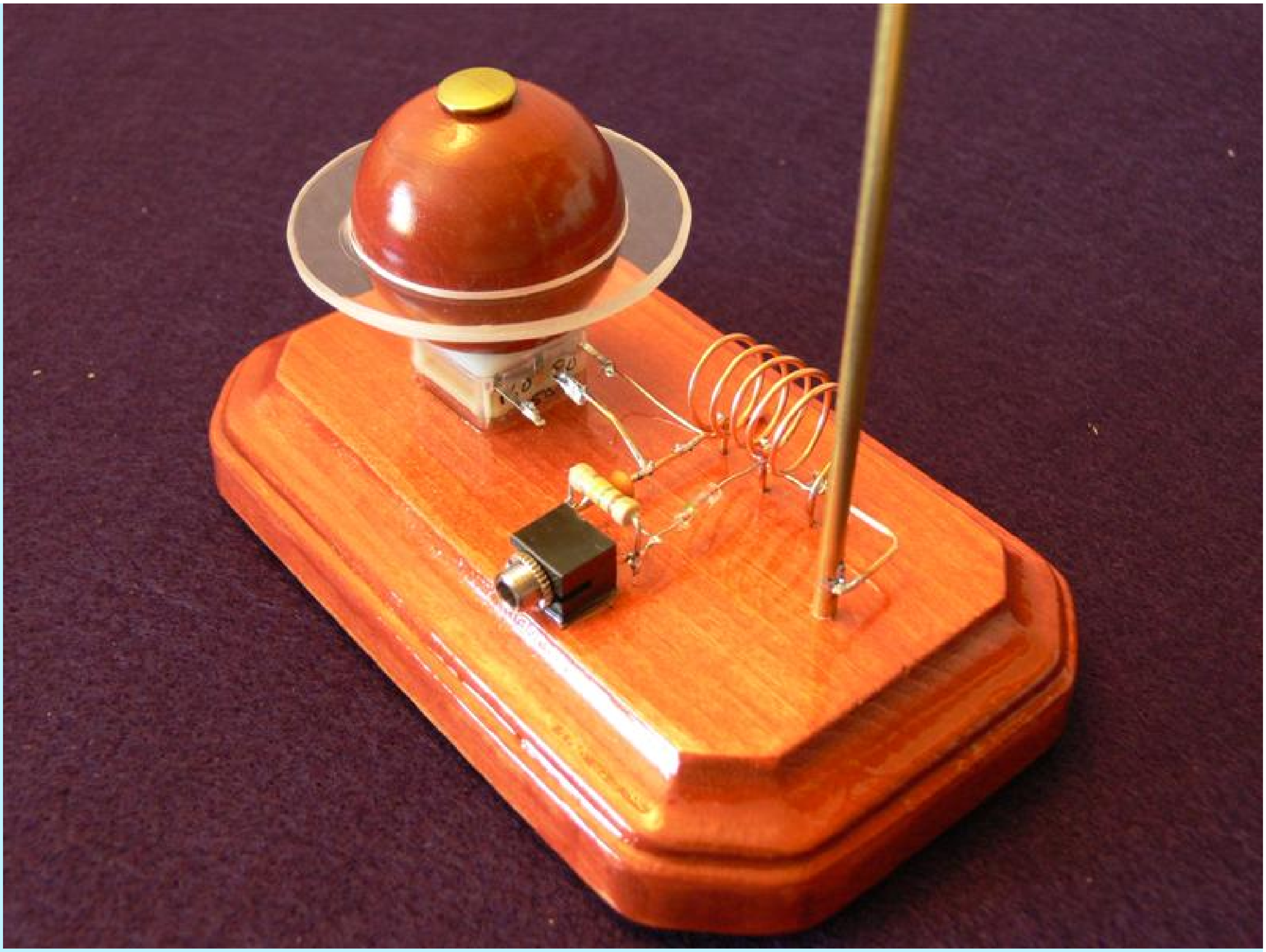

The Saturn dial and knob were fashioned from a "doll's head" from Michael's Arts and Crafts, a piece of lucite cut with two circle cutters, and a brass paper fastener. The knob is fixed to the tuning capacitor with a small machine screw that fits in the hole below the brass fastener. The most difficult part of this was fashioning "Saturn's rings". This must be done very carefully and slowly. The inside edge should be cut slightly undersized and then sanded with a drum sander to fit snugly. The outside edges can be sanded with fine sandpaper and polished with a plastic polisher.

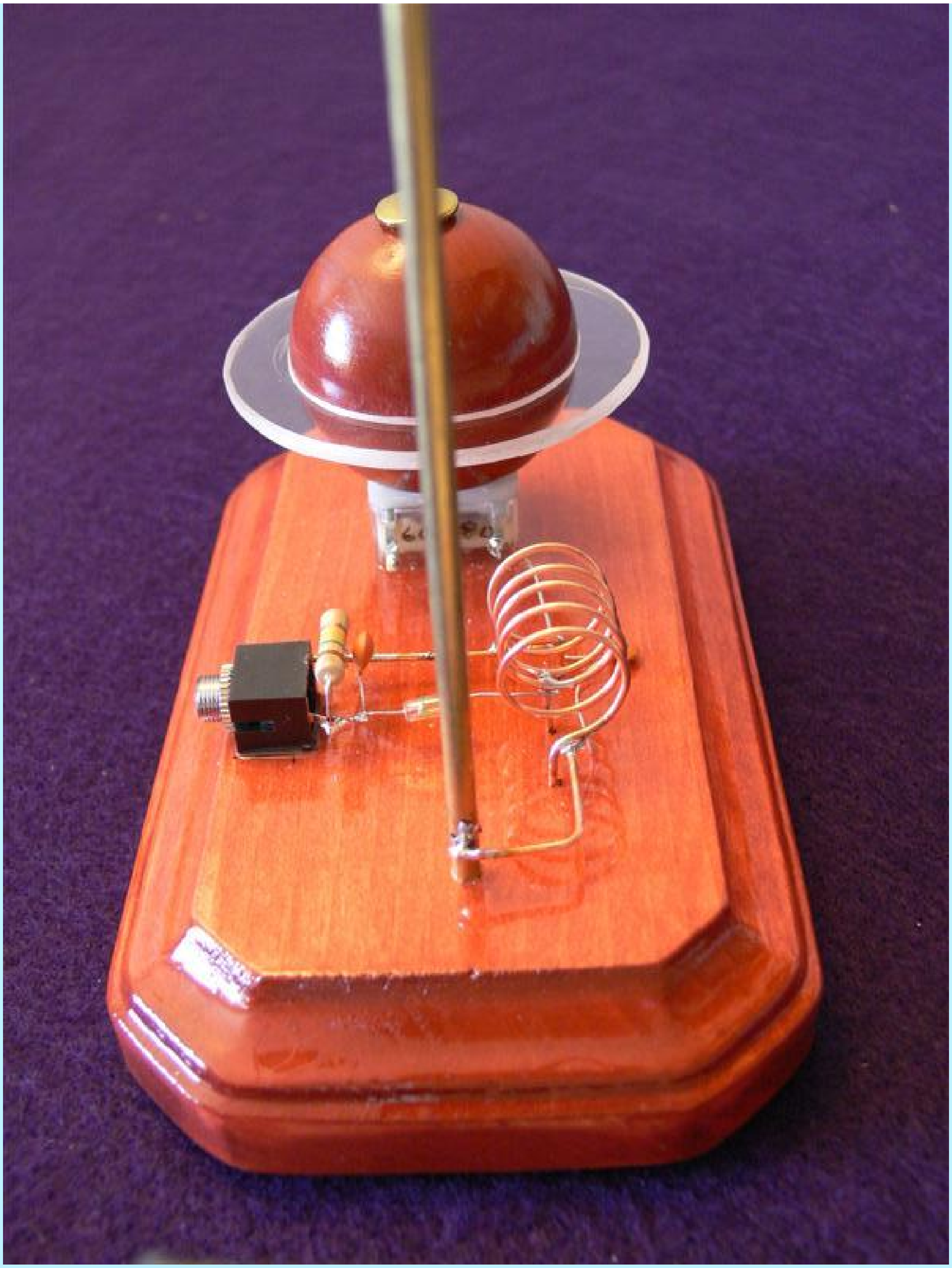

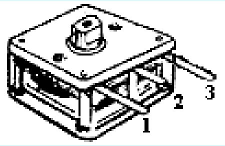

The air variable capacitor may be obtained from Electronix Express at http://www.elexp.com/. Part number 14VCRF10-280P. The 80 pf side is recommended for the second circuit, contacts 2-3. Contacts 1 and 3 were used for the first circuit (50pf).

- OSC: 5-59 pf
- ANT: 5-142 pf
- OSC and ANT Trimmer 10pf range

Home -- If you want to contact me go to the home page and go through the Music Clinic link.

28th June 2013

FM Crystal Radio

Build a battery free FM crystal radio using 1mm silver plated copper wire

[http://2.bp.blogspot.com/-LyTLmJ0D5ao/Uc2SwAO-02I/AAAAAAAABL4/jkHYmvIimcE/s1600/IMG_2535.JPG]

This FM radio runs without battery and you can listen to FM radio by just

placing this close to your window.
If reception is not good, you can attach an external FM antenna (or with antenna amplifier), instead of using the 1.2M radio antenna.

Parts you will need:
A Diode that works in high frequency (200Mhz or above, e.g. AA112, 1N34A, 1N48, Iss86 or Iss106.
A 0-12 pf AIR Variable capacitor - you can get a 0-22pf one, then remove all the copper plates and leave just 4.
50pf and 180pf fixed ceramic capacitor
100K 1/4W resistor
33uH Choke coil.
Silver platted copper wire 1mm diameter
1.2meter radio antenna.
3.5mm mini headphone jack
A high impedance 4K Ohm earphone or a crystal earphone.

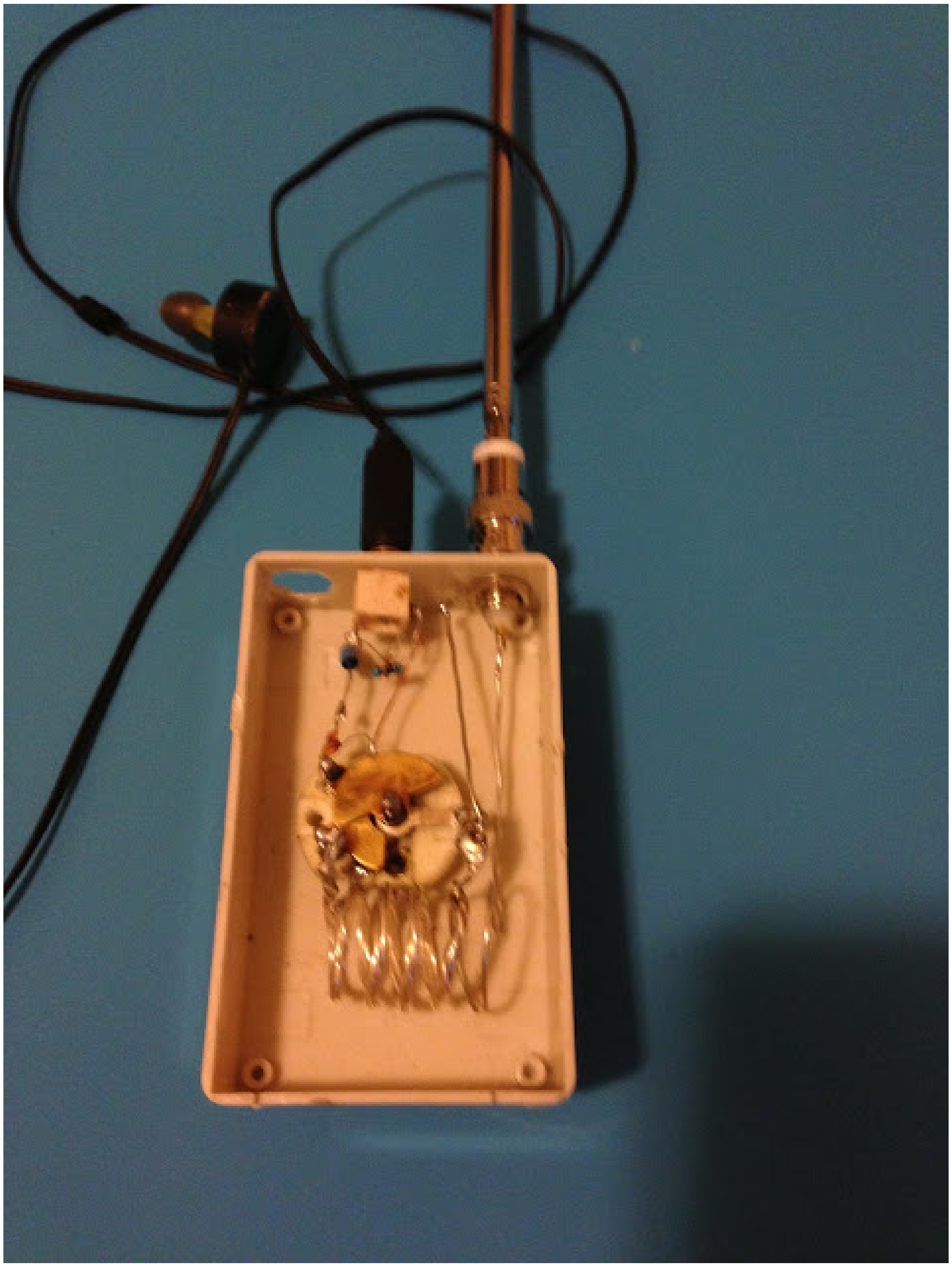

[http://3.bp.blogspot.com/-NhzpbldzMyU/Uc2SwLCh3LI/AAAAAAAABL8/J7ym0UWhqqk/s1600/IMG_2536.JPG]

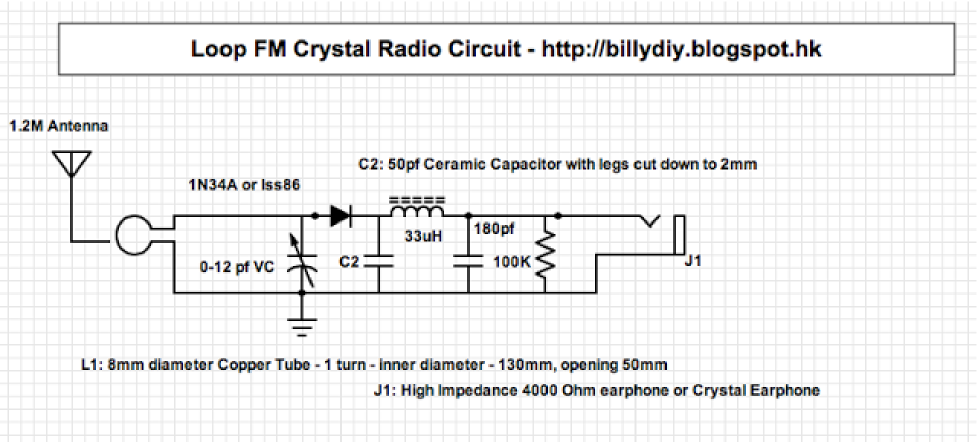
Loop FM Crystal Radio Circuit - http://billydiy.blogspot.hk
1.2M Antenna
C2: 50pf Ceramic Capacitor with legs cut down to 2mm
1N34A or Iss86
33uH
180pf
100K
0-12 pf VC
C2
J1
L1: 8mm diameter Copper Tube - 1 turn - inner diameter - 130mm, opening 50mm
J1: High Impedance 4000 Ohm earphone or Crystal Earphone

[http://2.bp.blogspot.com/-84ZQcpL-iDM/Uc2DeEbsVzI/AAAAAAAABLE/vF4um1tHpws/s763/Screen+Shot+2013-06-28+at+8.36.46+PM.png]

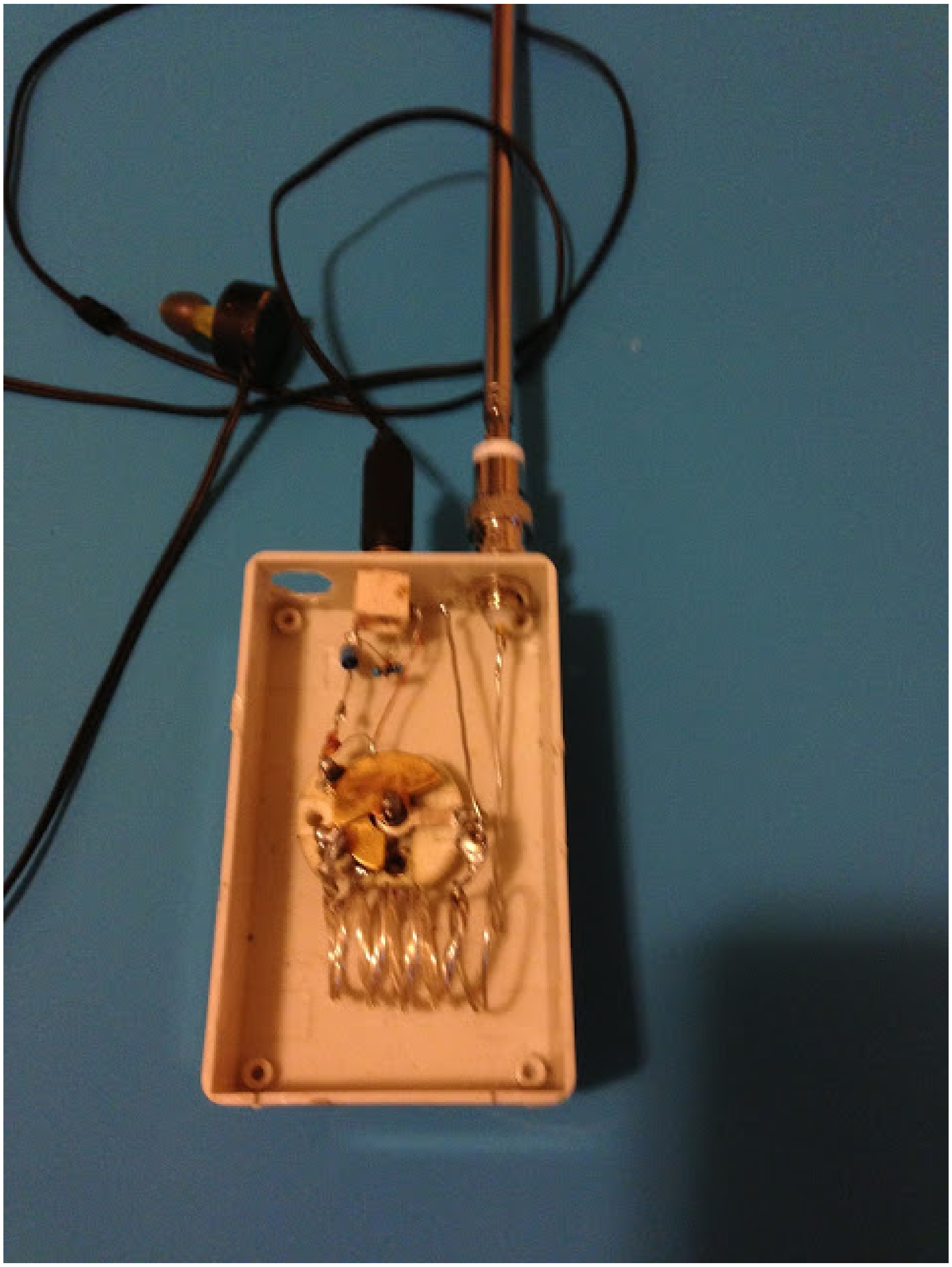

[http://3.bp.blogspot.com/-NhzpbldzMyU/Uc2SwLCh3LI/AAAAAAAABL0/eRdHXo7f22w/s1600/IMG_2536.JPG]

Posted 28th June 2013 by Billy

Download This To PDF
Free to Download and Convert. Get It Instantly, Download Now.
Go to fromdoctopdf.com

0 Add a comment

Enter your comment...

Comment as: Unknown (Goo ▼ **Sign out**

Publish **Preview** ☐ Notify me

- ☎ +61-3-90179663
- My Account
  - Register
  - Login
- Wish List (0)
- Shopping Cart
- Checkout

GaliGear
Electronics Store

Search

0 item(s) - $0.00

Your shopping cart is empty!

Categories

- Electronic Components
  - Breadboard/PCBs (5)
  - Capacitors (53)
  - Diodes (18)
  - Integrated Circuits (72)
  - Leds (46)
  - Oscillators (15)
  - Other (34)
  - Regulators (9)
  - Resistors (120)
  - SMT / SMD Parts (22)
  - Switches (5)
  - Transistors (29)

  See All Electronic Components
- Home and Security
- Measuring/Testing Tools
- Mixed Component Bags
- Surveillance & Monitoring

FM Crystal Radio Circuit

FM Crystal Radio Circuit

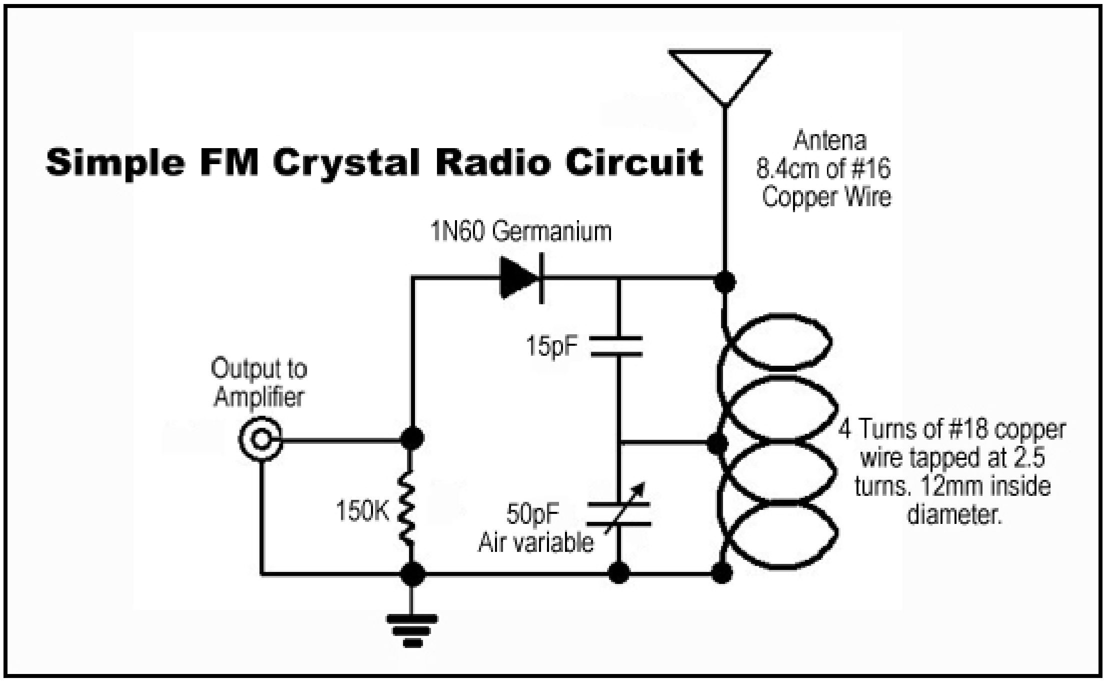
Simple FM Crystal Radio Circuit
Antena
8.4cm of #16
Copper Wire
1N60 Germanium
15pF
Output to
Amplifier
4 Turns of #18 copper
wire tapped at 2.5
turns. 12mm inside
diameter.
150K
50pF
Air variable

Parts List (some of these parts you can buy from our online store):

- 1N60 Germanium Diode
- 15pF Ceramic Capacitor
- 50pF Variable Capacitor
- 150K Ohm Resistor
- #16 & #18 Copper wires

This simple FM Crystal circuit works best near the transmitter (around 15 kilometers range). You must move the circuit around to find the best signal reception.

You can listen to FM radio stations by connecting high impedance earphones (crystal preferable) to its output or otherwise connect the circuit directly to an audio amplifier Aux input.

It is recommended to connect a "rabbit ear" antenna or hanging a short wire (15cm or so) in various positions next to the internal antenna if no signal is detected. The variable length of rabbit ears can help to tune in more stations.

Enjoy!

Information

- Our Return Policy
- About Us
- Delivery Information

- Privacy Policy
- Terms & Conditions

Customer Service

- Contact Us
- Returns
- Site Map

Extras

- Brands
- Gift Vouchers
- Affiliates
- Specials

My Account

- My Account
- Order History
- Wish List
- Newsletter

Gali Gear Electronics ©2017, Melbourne Australia

FM Crystal Set

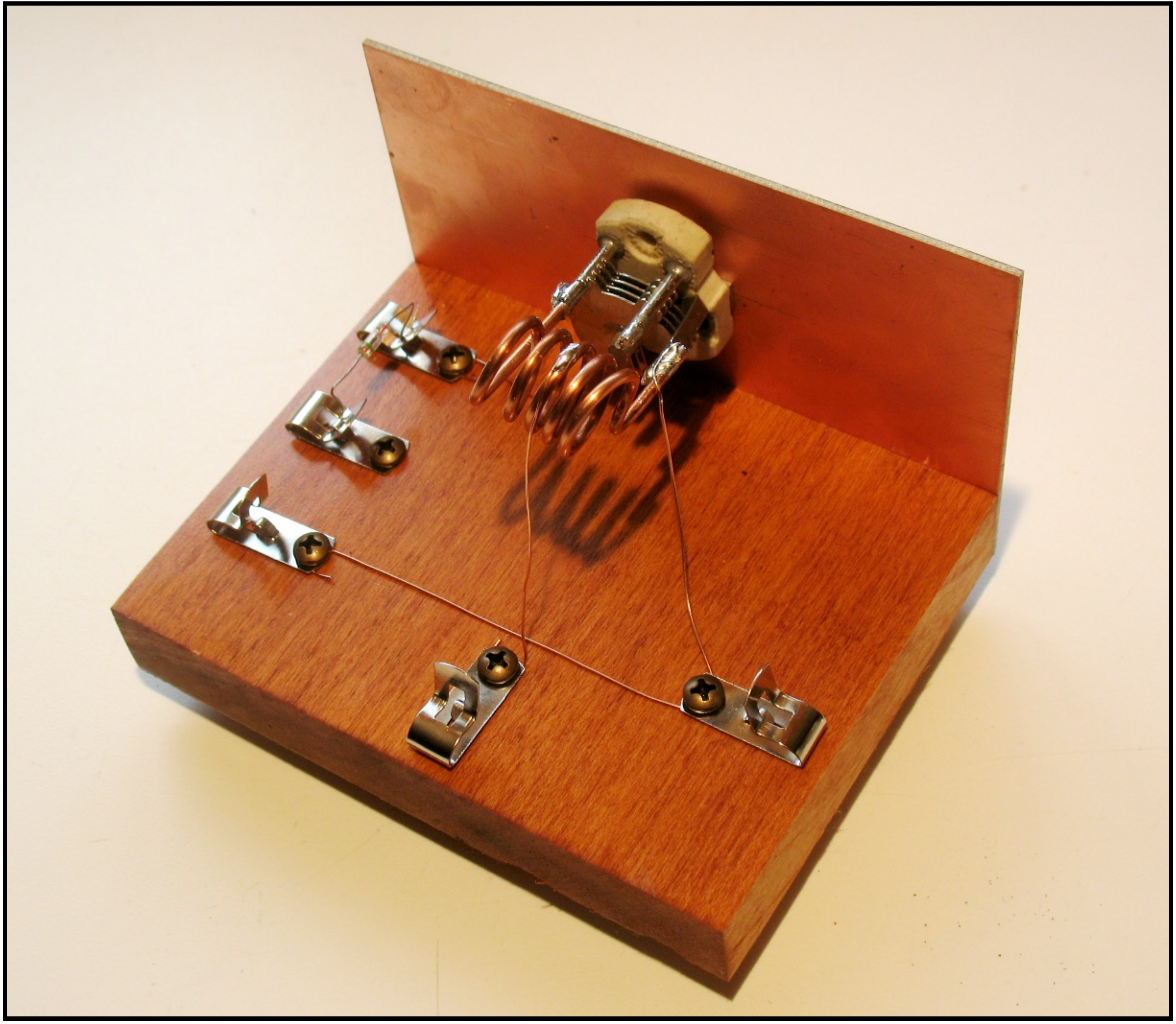

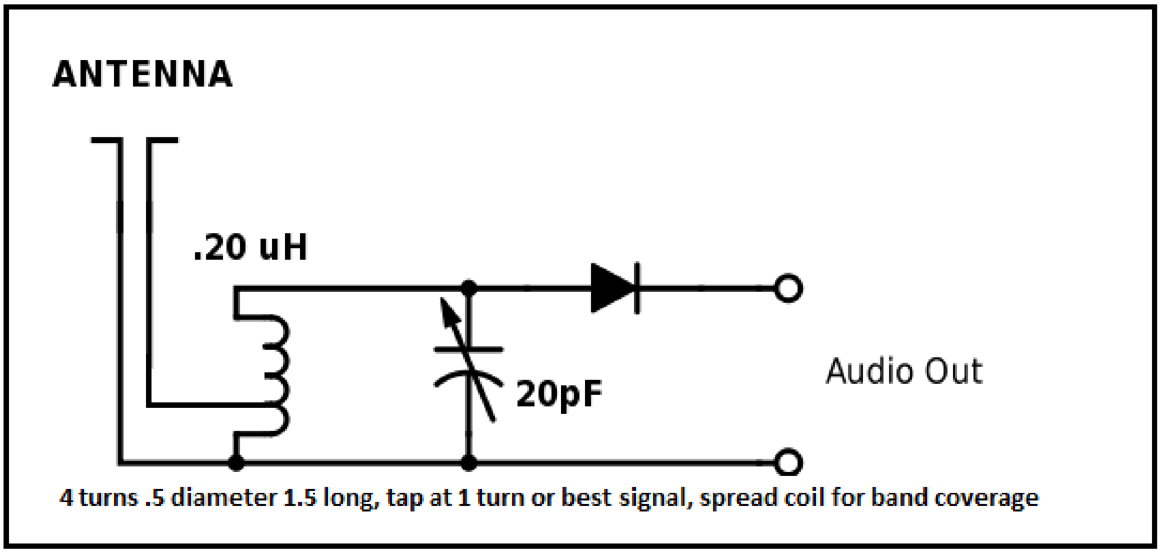
ANTENNA
.20 uH
20pF
Audio Out
4 turns .5 diameter 1.5 long, tap at 1 turn or best signal, spread coil for band coverage

By slope detection you can get FM. If you live close to the stations you should do well with Hi Z phones. I use an audio amp, I am far from the stations. I get three.

US006104349A

United States Patent [19]

Cohen

[11] **Patent Number: 6,104,349**

[45] **Date of Patent: *Aug. 15, 2000**

[54] **TUNING FRACTAL ANTENNAS AND FRACTAL RESONATORS**

[76] Inventor: **Nathan Cohen**, 21 Ledgewood Pl., Belmont, Mass. 02178

[*] Notice: This patent issued on a continued prosecution application filed under 37 CFR 1.53(d), and is subject to the twenty year patent term provisions of 35 U.S.C. 154(a)(2).

[21] Appl. No.: **08/967,372**

[22] Filed: **Nov. 7, 1997**

Related U.S. Application Data

[63] Continuation of application No. 08/609,514, Mar. 1, 1996, abandoned, which is a continuation-in-part of application No. 08/512,954, Aug. 9, 1995, abandoned.

[51] **Int. Cl.**^{7} **H01Q 1/24**; H01Q 1/36

[52] **U.S. Cl.** **343/702**; 343/741; 343/792.5; 343/787; 343/806; 343/846

[58] **Field of Search** 343/702, 795, 343/806, 792.5, 787, 788, 741, 846; 333/219; H01Q 1/24, 1/36, 1/38

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,079,602	2/1963	Duhamel et al.	343/792.5
3,249,946	5/1966	Flanagan	343/792.5
3,810,183	5/1974	Krutsinger et al.	343/708
3,811,128	5/1974	Munson	343/787
4,318,109	3/1982	Weathers	343/806
5,164,738	11/1992	Walter et al.	343/792.5

OTHER PUBLICATIONS

Pfeiffer, A., "The Pfeiffer Quad Antenna System", *QST*, pp. 28–30 (1994).

Primary Examiner—Michael C. Wimer
Attorney, Agent, or Firm—Flehr Hohbach Test Albritton & Herbert LLP

[57] **ABSTRACT**

A first fractal antenna of iteration N≧2 in free space exhibits characteristics including at least one resonant frequency and bandwidth. Spacing-apart the first fractal conductive element from a conductive element by a distance Δ, non-planarly or otherwise, preferably ≦0.05λ for non-planar separation for frequencies of interest decreases resonant frequency and/or introduces new resonant frequencies, widens the bandwidth, or both, for the resultant antenna system. The conductive element may itself be a fractal antenna, which if rotated relative to the first fractal antenna will alter or tune at least one characteristic of the antenna system. Forming a cut anywhere in the first fractal antenna causes new and different resonant nodes to appear. The antenna system may be tuned by cutting-off a portion of the first fractal antenna, typically increasing resonant frequency. A region of ground plane may be formed adjacent the antenna system, to form a sandwich-like system that is readily tuned. Resonator systems as well as antenna systems may be tuned using is disclosed methodology.

18 Claims, 19 Drawing Sheets

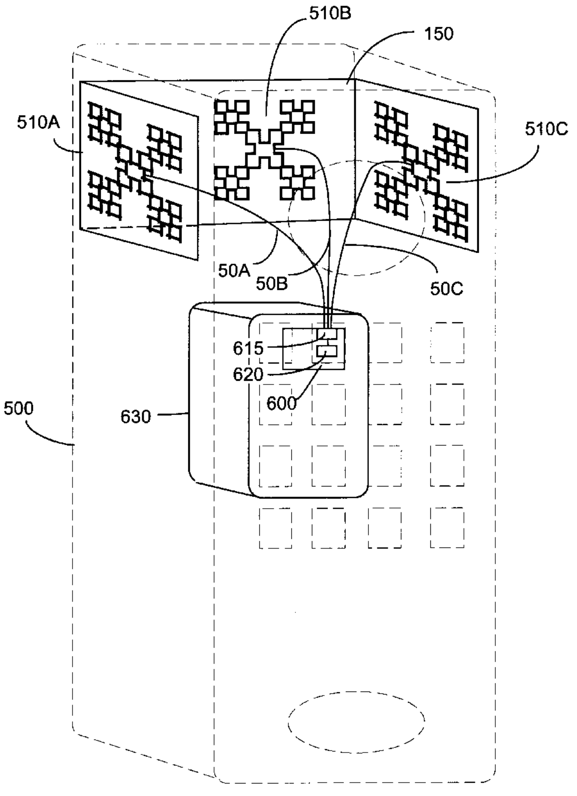

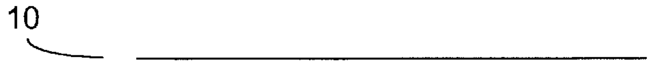

FIGURE 1A (PRIOR ART)

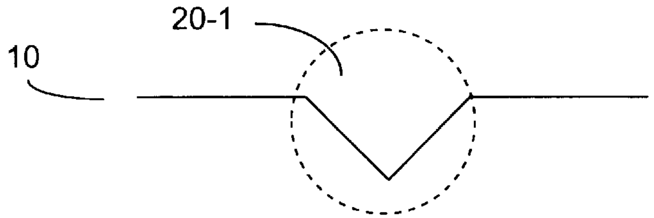

FIGURE 1B (PRIOR ART)

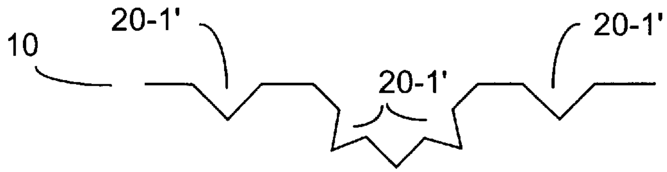
20-1'
10
20-1'
20-1'

FIGURE 1C (PRIOR ART)

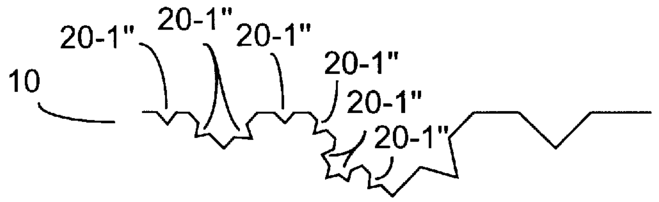
20-1"
20-1"
20-1"
20-1"
20-1"
20-1"
10

FIGURE 1D (PRIOR ART)

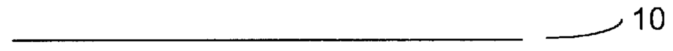

FIGURE 2A (PRIOR ART)

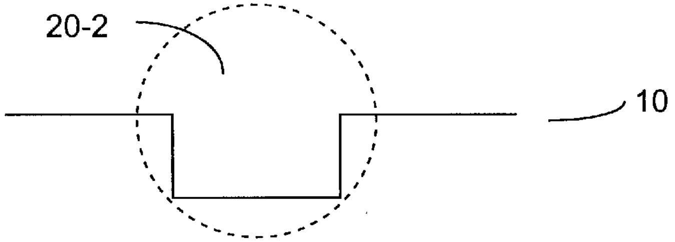

FIGURE 2B (PRIOR ART)

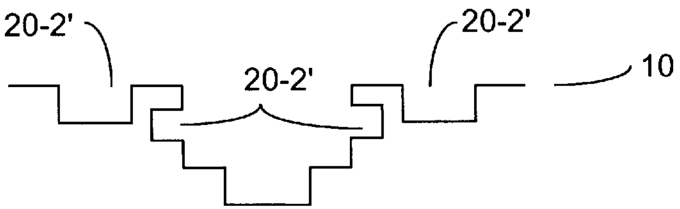
20-2'
20-2'
20-2'
10

FIGURE 2C (PRIOR ART)

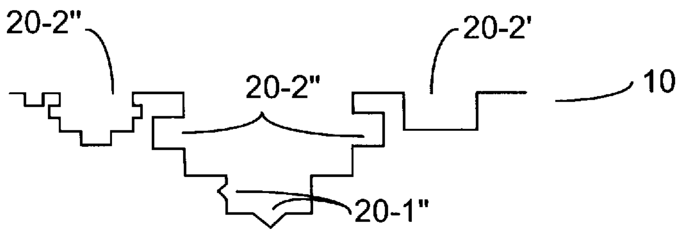
20-2"
20-2'
20-2"
10
20-1"

FIGURE 2D (PRIOR ART)

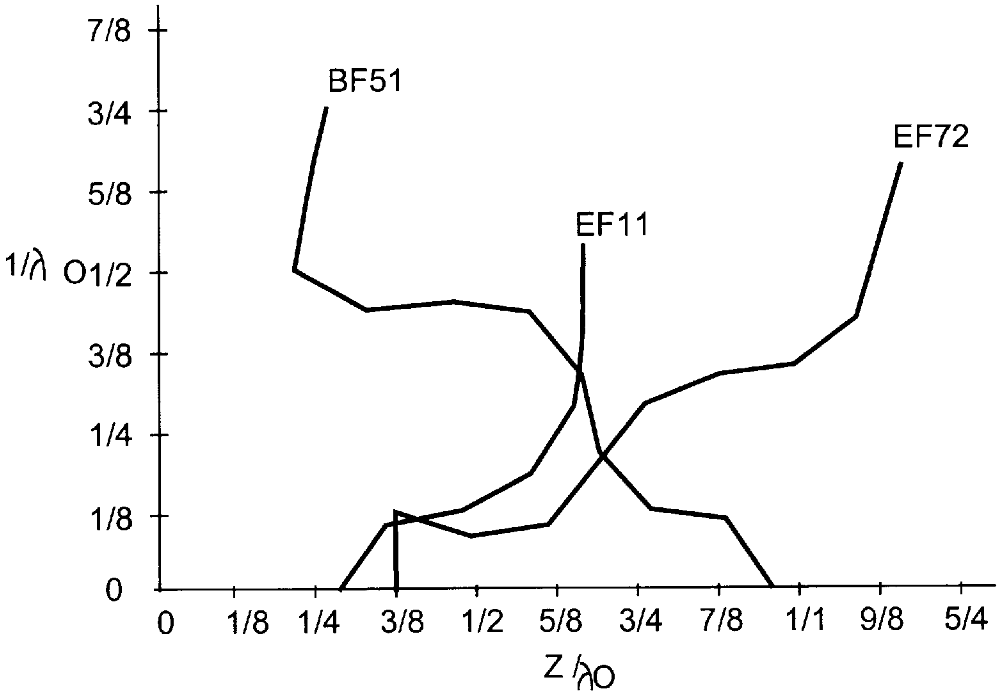

FIGURE 3 (PRIOR ART)

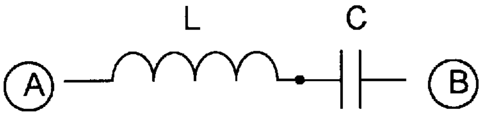

FIGURE 4A (PRIOR ART)

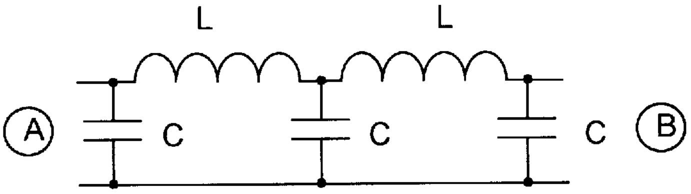

FIGURE 4B (PRIOR ART)

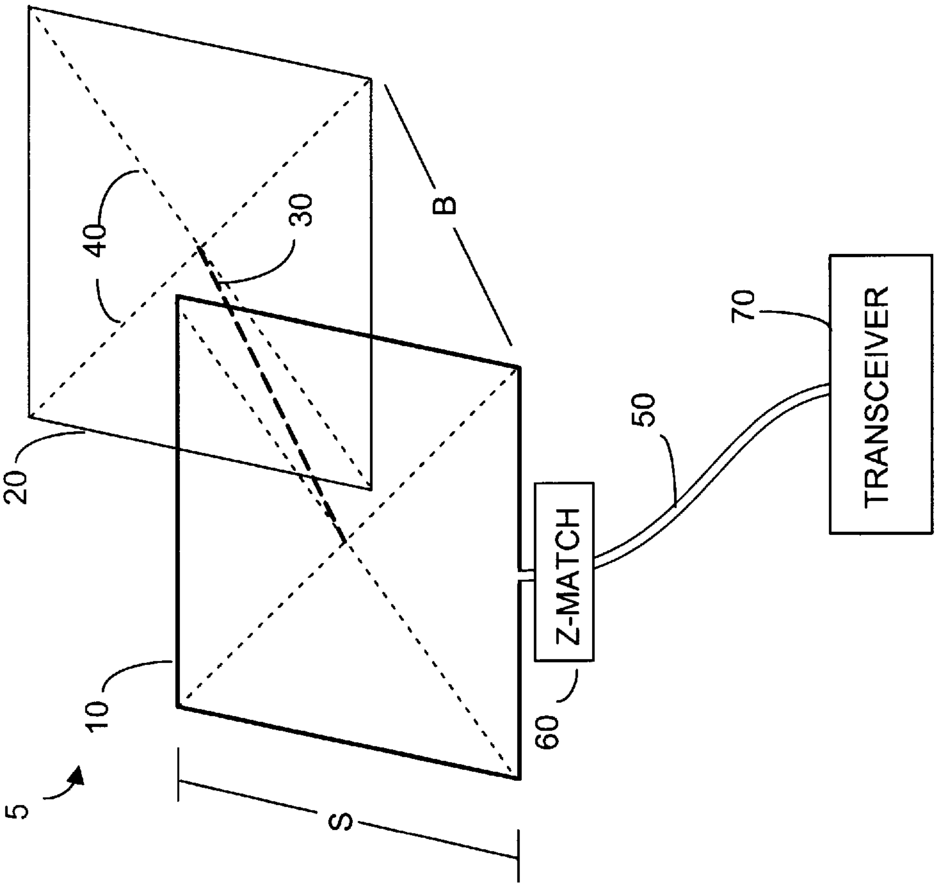
5
10
20
40
30
B
S
60
Z-MATCH
50
70
TRANSCEIVER

FIGURE 5A
(PRIOR ART)

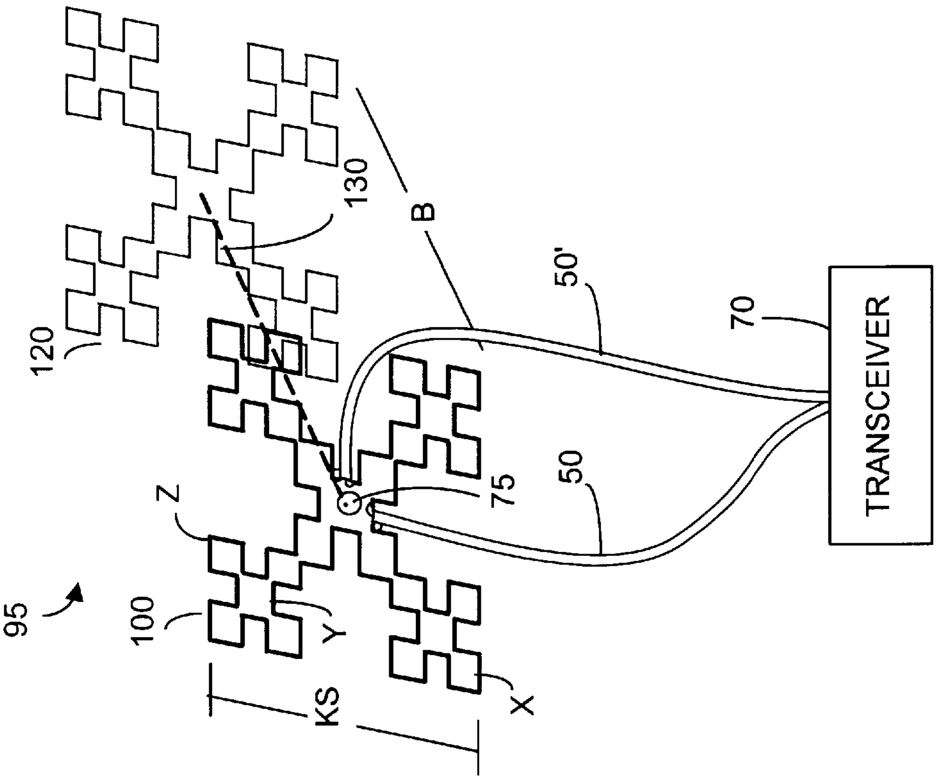
95
100
Z
120
130
B
Y
X
KS
75
50
50'
70
TRANSCEIVER

FIGURE 5B

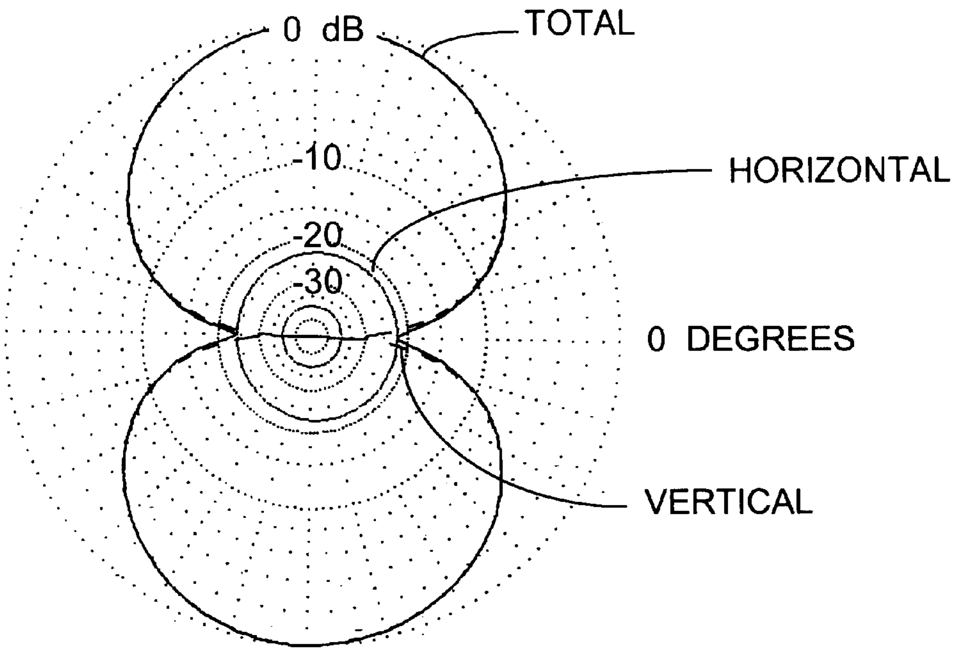

FIGURE 6

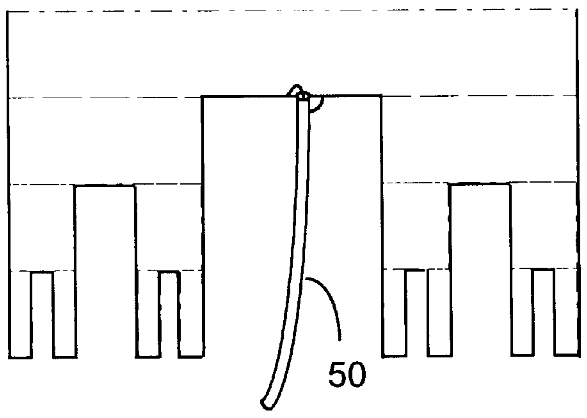

FIGURE 7A

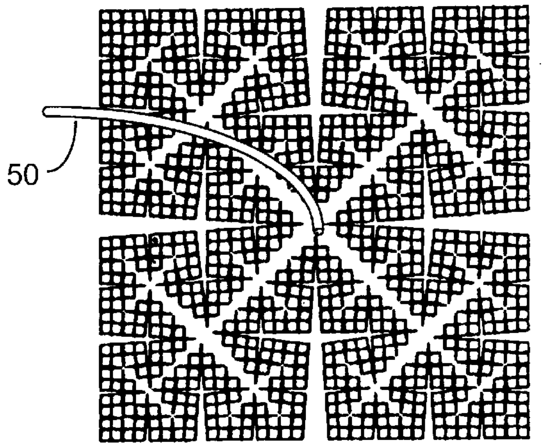

FIGURE 7B

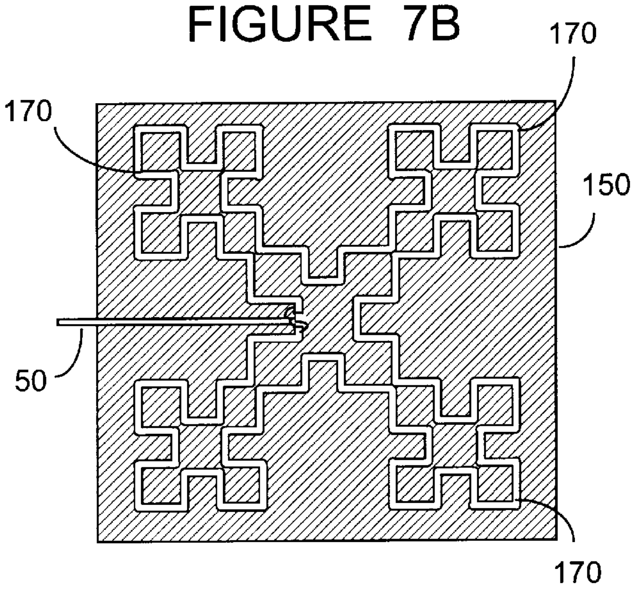
170
170
150
50
170

FIGURE 7C-1

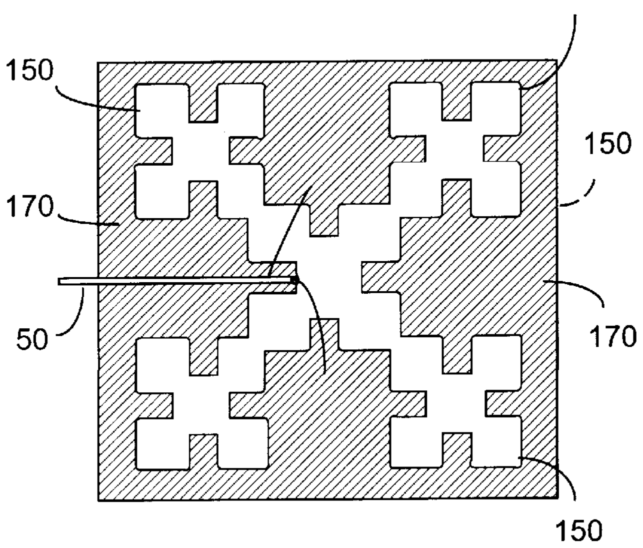
150
150
150
170
50
170
150

FIGURE 7C-2

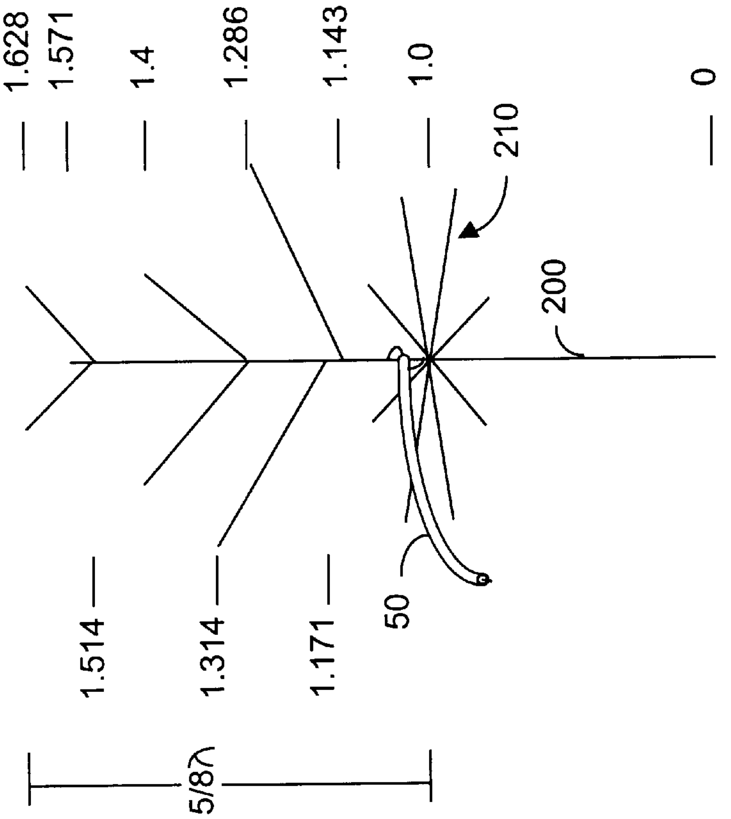
1.628
1.571
1.4
1.286
1.143
1.0
0
210
200
1.514
1.314
1.171
50
5/8λ

FIGURE 7D

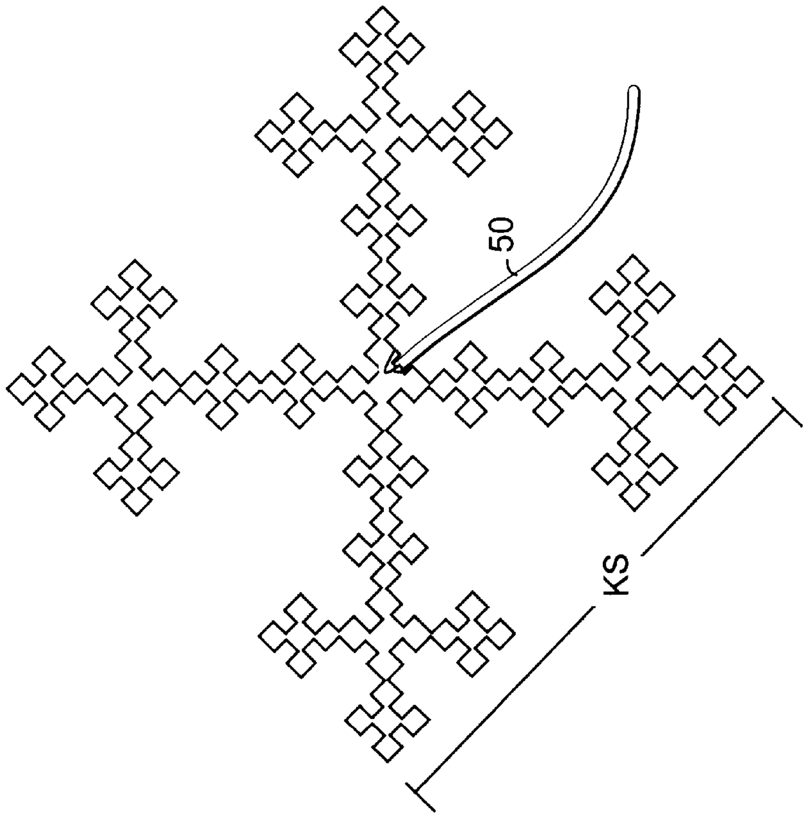

FIGURE 7E

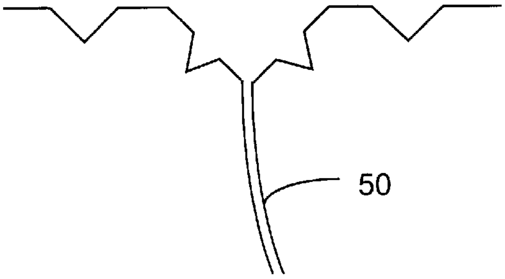

FIGURE 7F

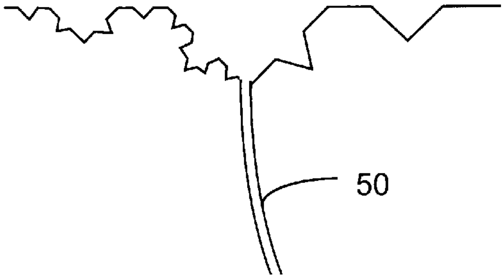

FIGURE 7G

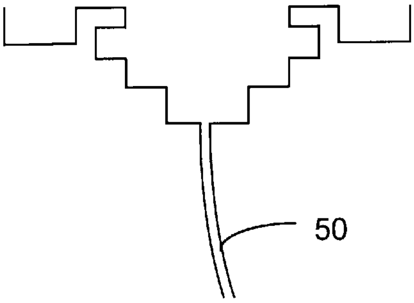

FIGURE 7H

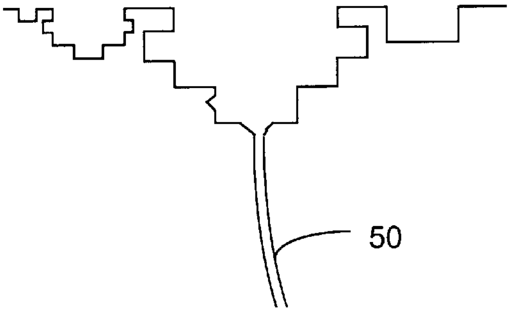

FIGURE 7I

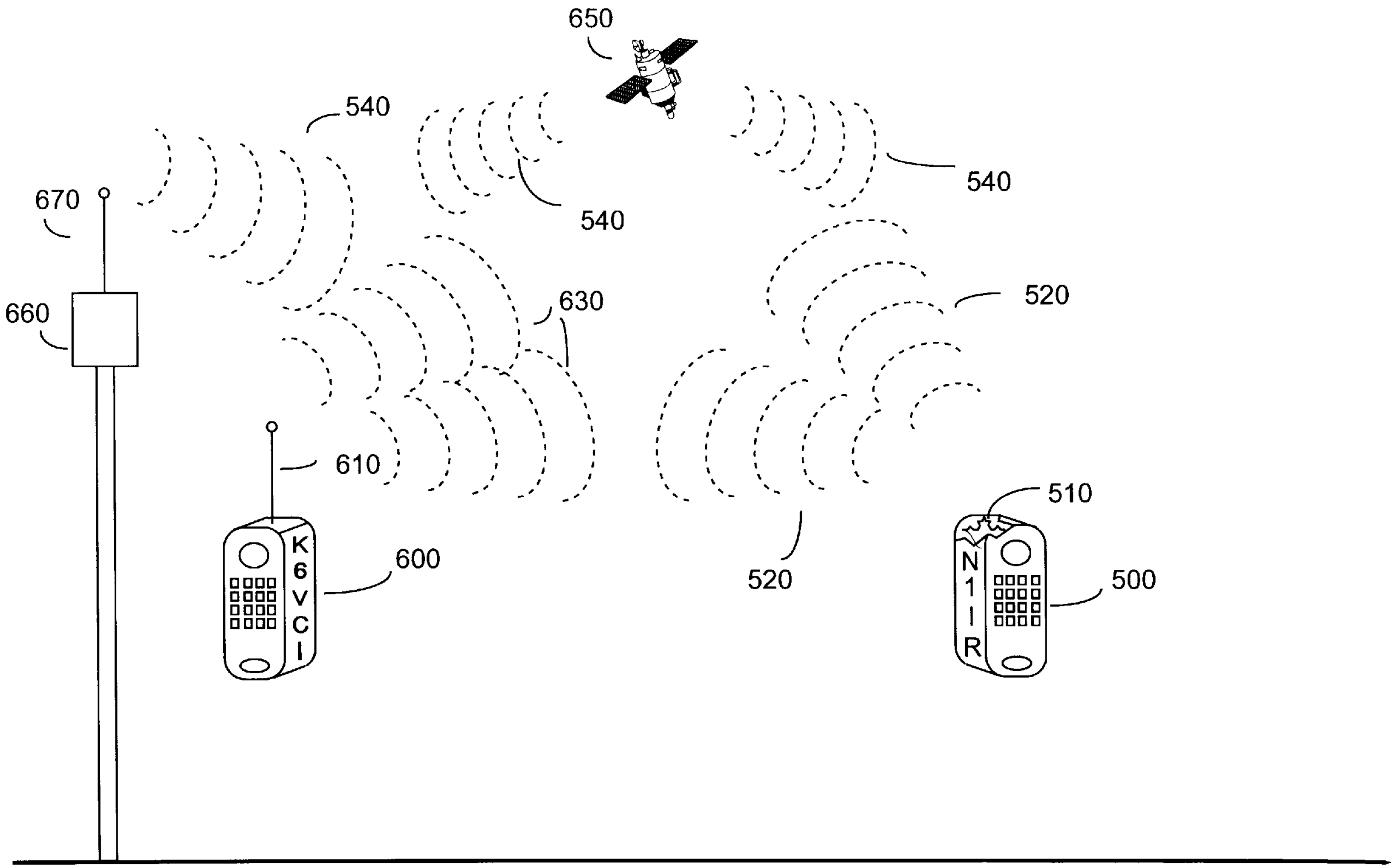
650
540
540
540
670
660
630
520
610
510
600
520
500
K
6
V
C
I
N
1
I
R

FIGURE 8A

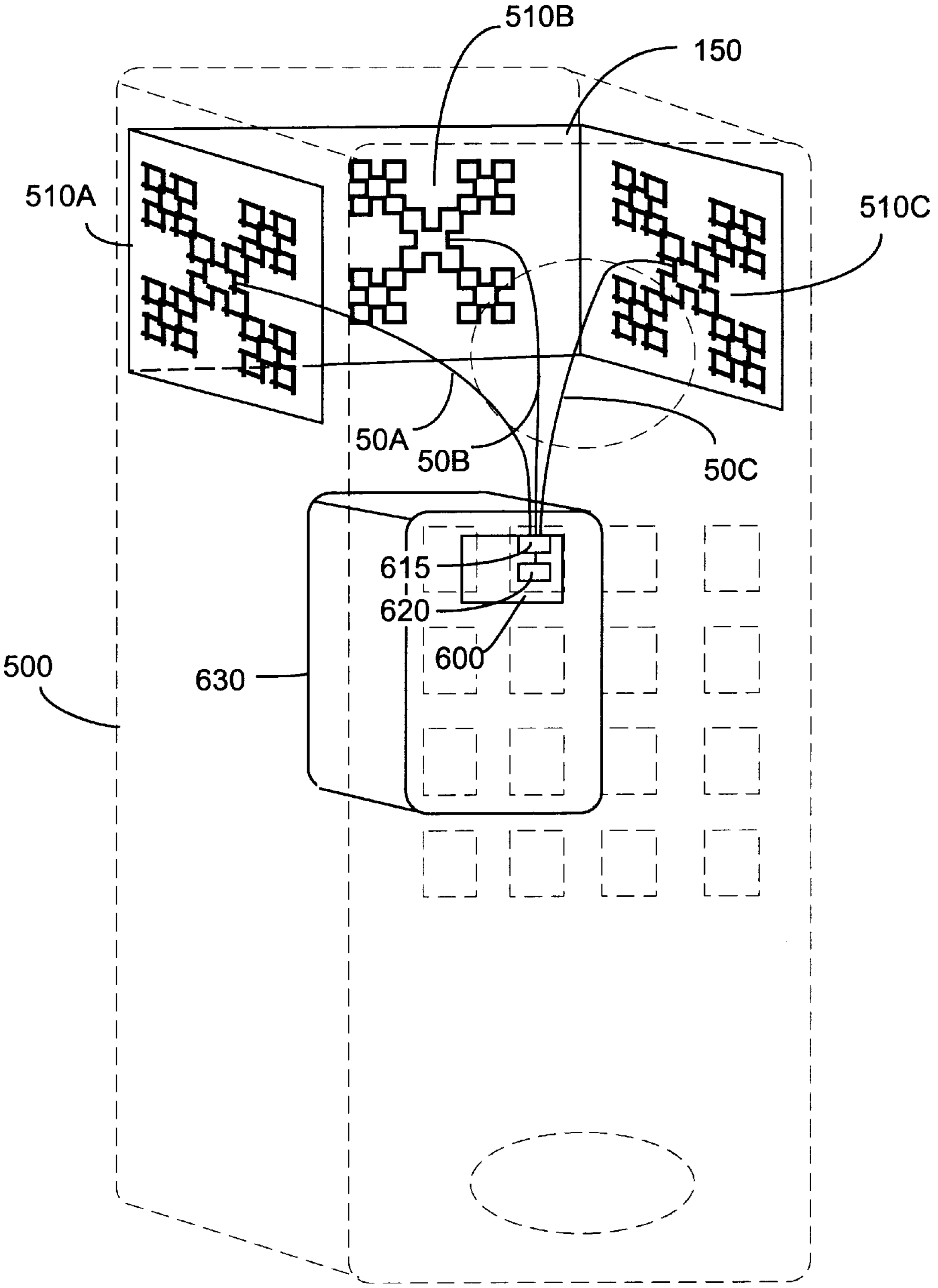
510B
150
510A
510C
50A
50B
50C
615
620
600
500
630

FIGURE 8B

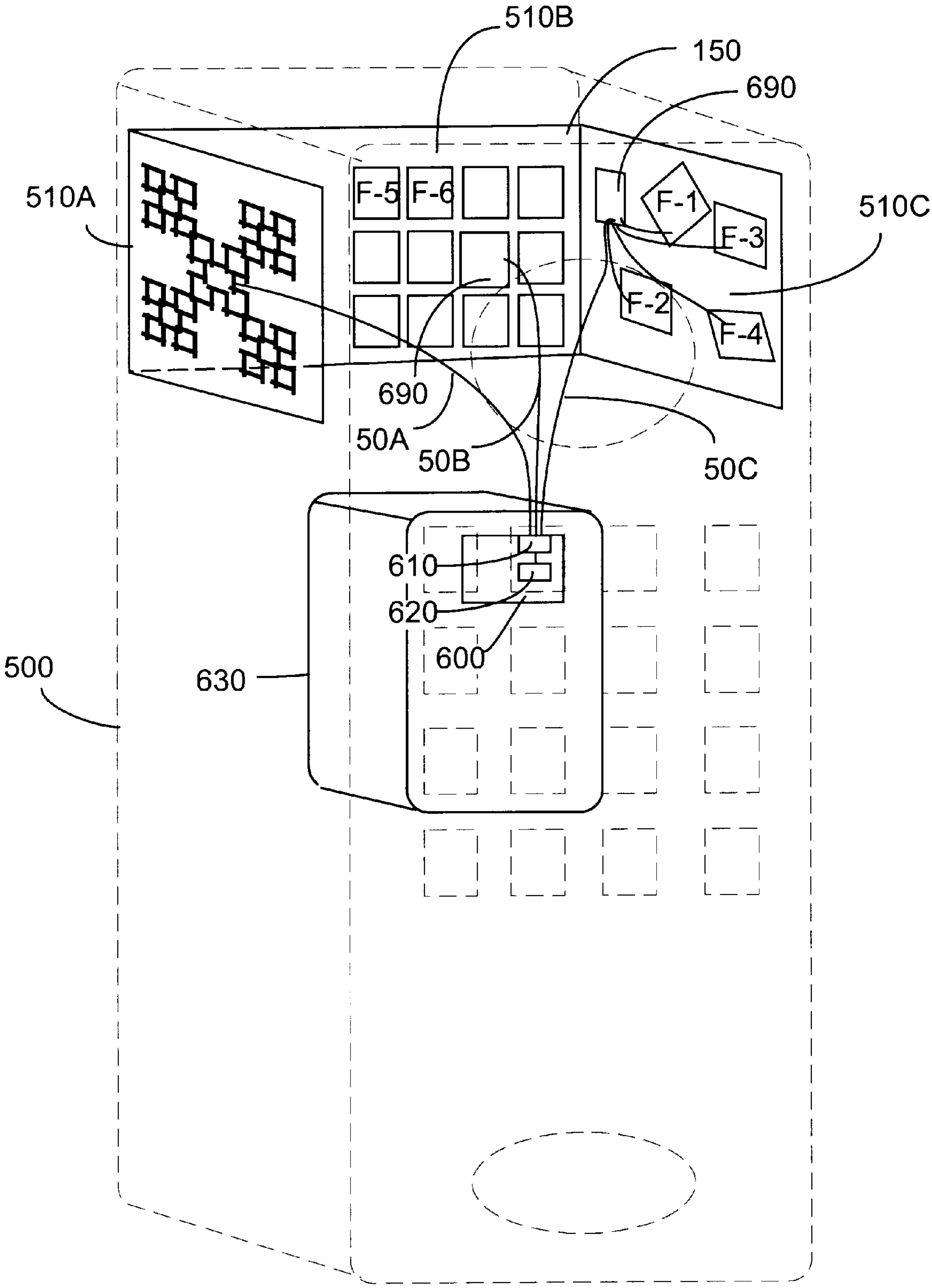
F-5
F-6
F-1
F-3
F-2
F-4

FIGURE 8C

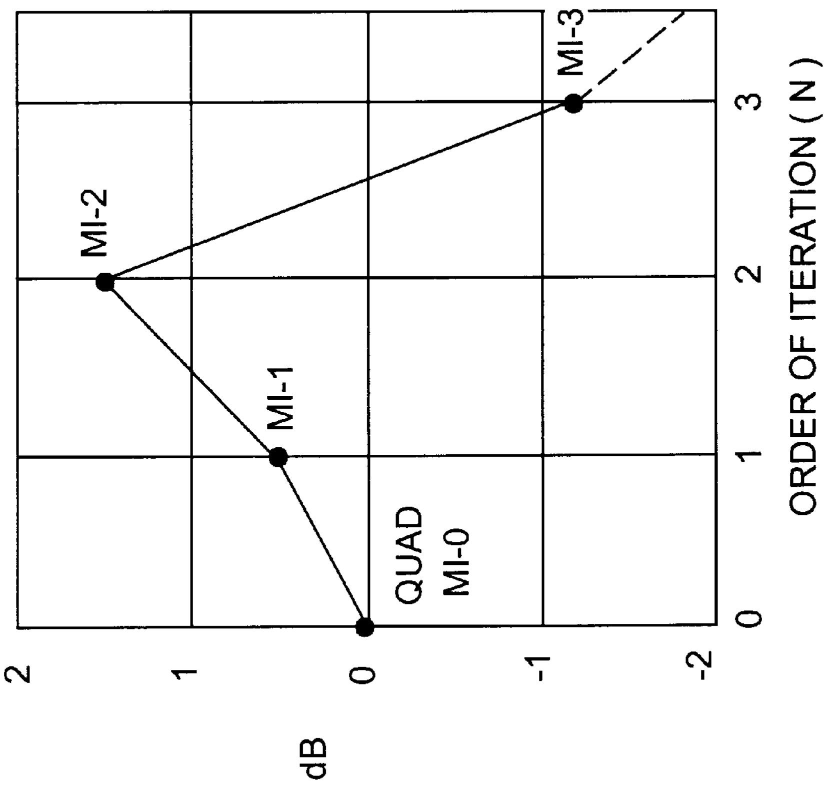

FIGURE 9A

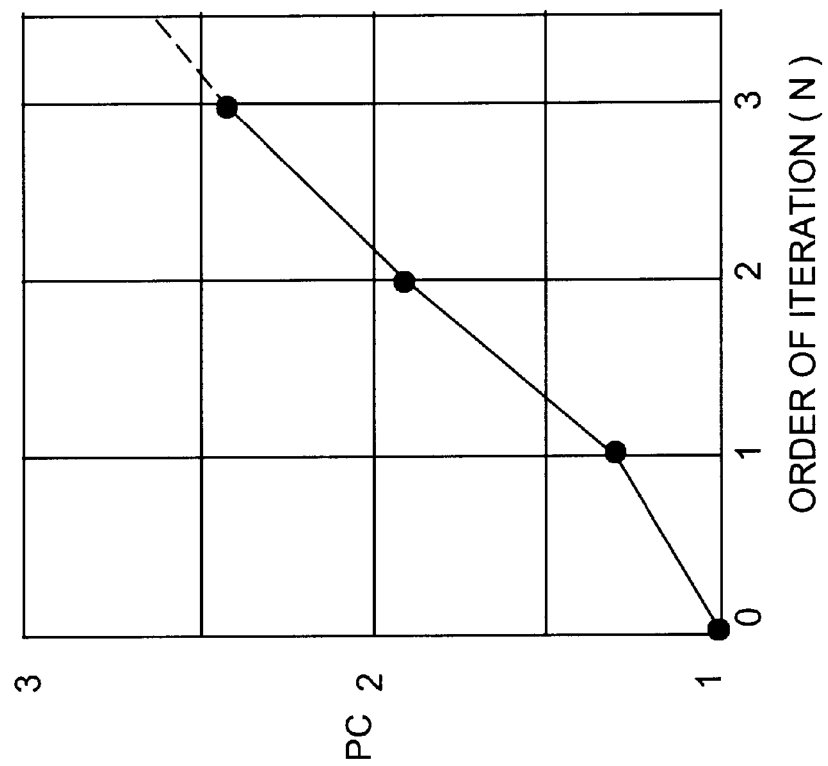

FIGURE 9B

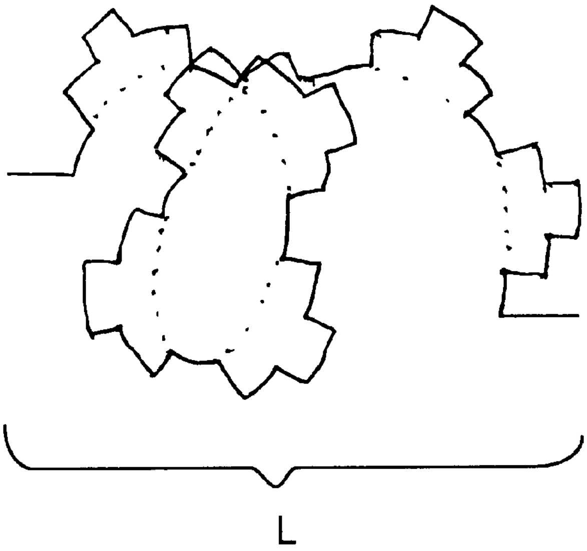

FIGURE 10A

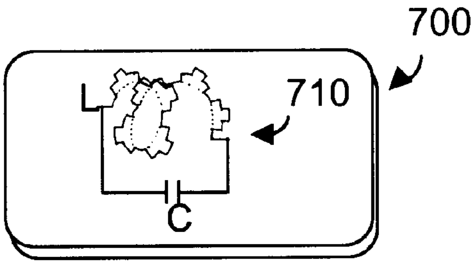

FIGURE 10B

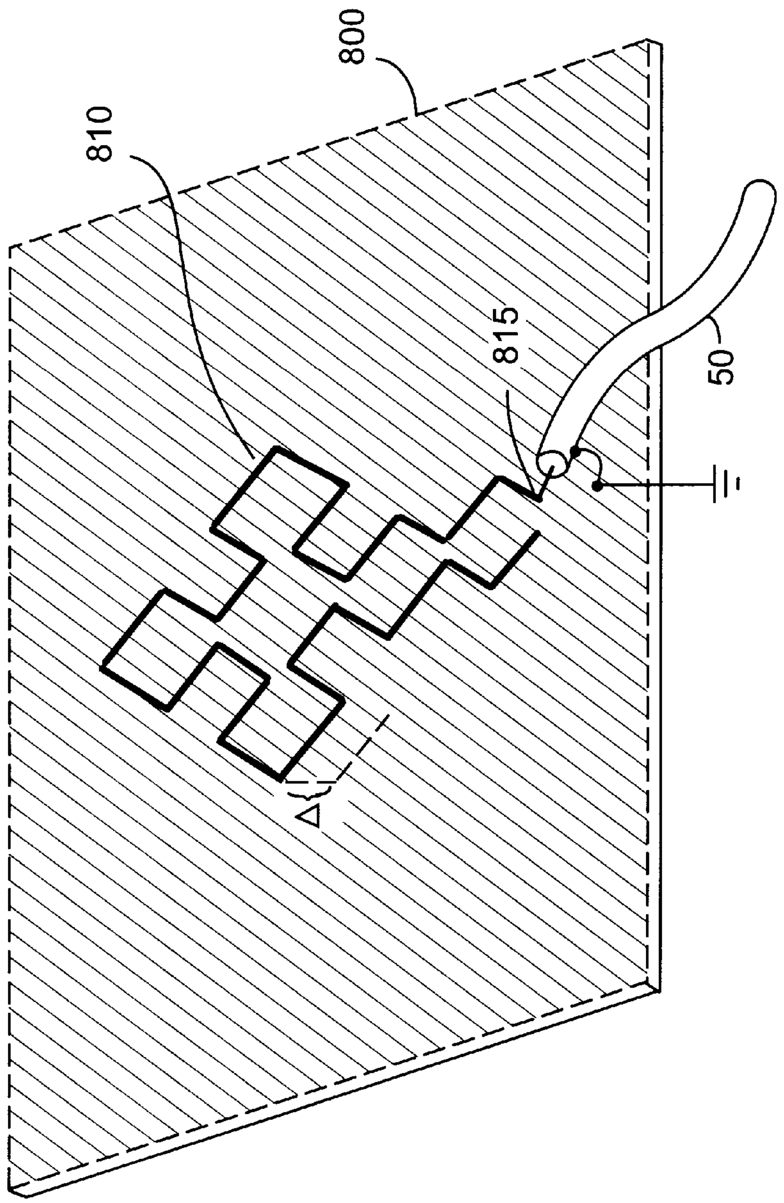
800
810
815
50
Δ

FIGURE 11A

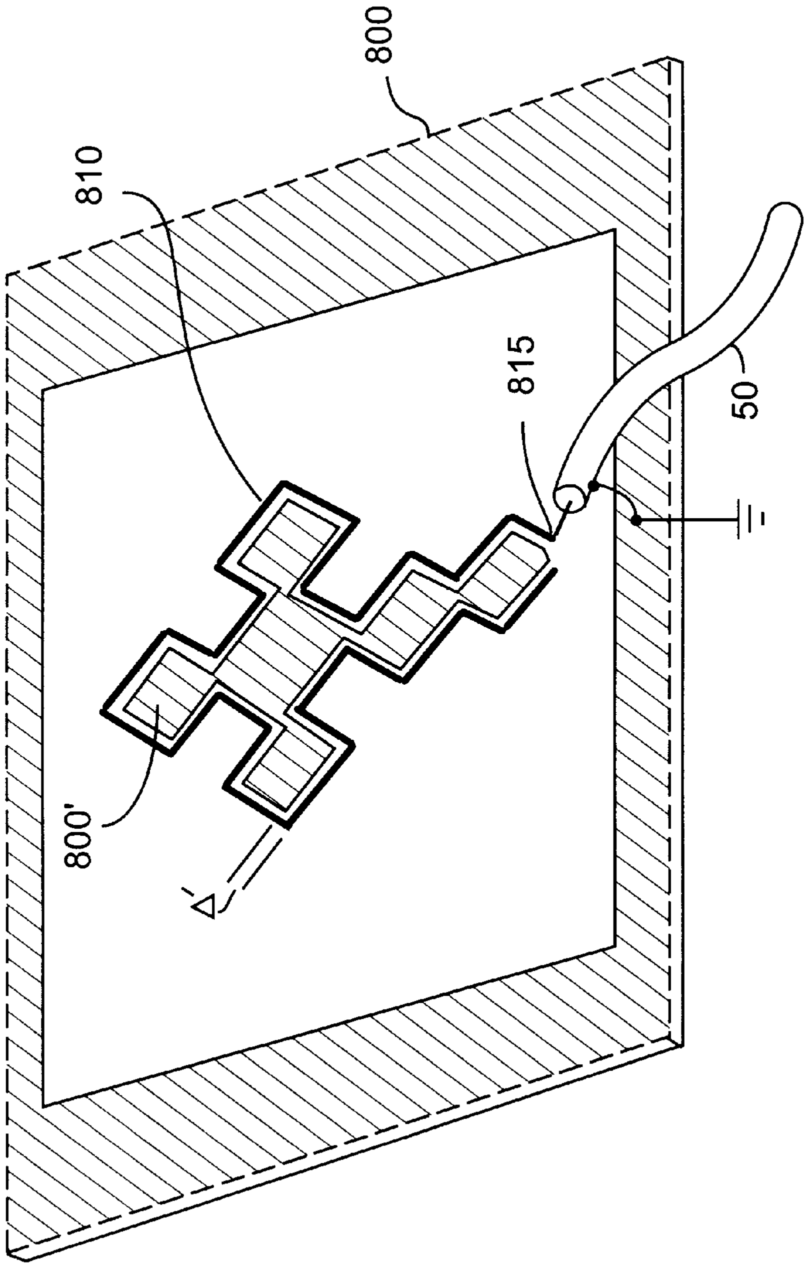
800
810
815
50
800'
Δ'

FIGURE 11B

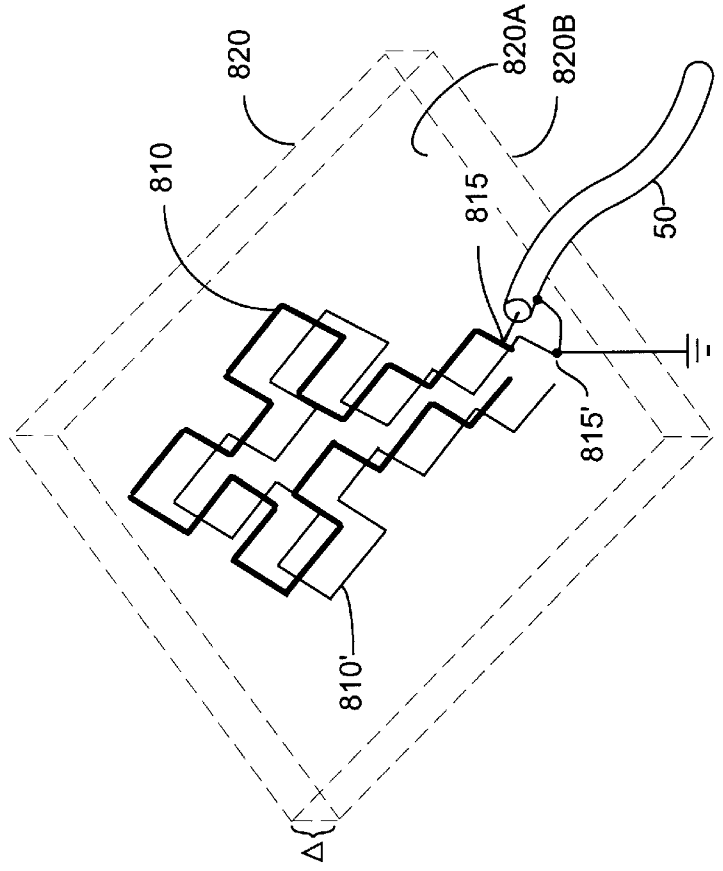

820
820A
820B
810
815
50
815'
810'
Δ

FIGURE 12A

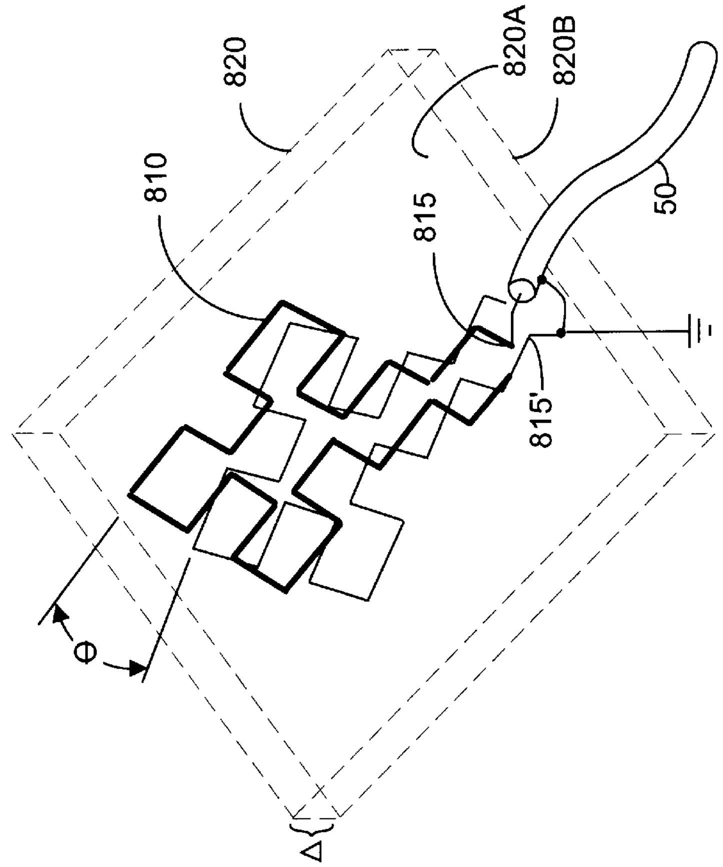

820
820A
820B
810
815
50
815'
Θ
Δ

FIGURE 12B

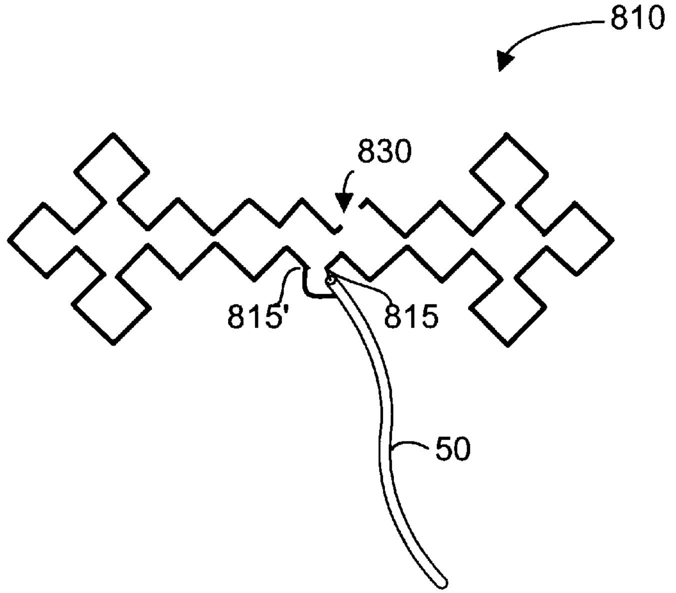
810
830
815'
815
50

FIGURE 13A

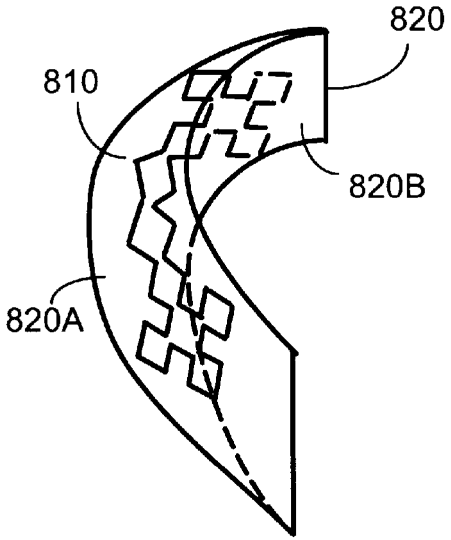
820
810
820B
820A

FIGURE 13B

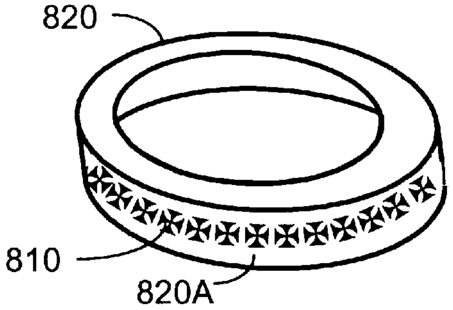
820
810
820A

FIGURE 13C

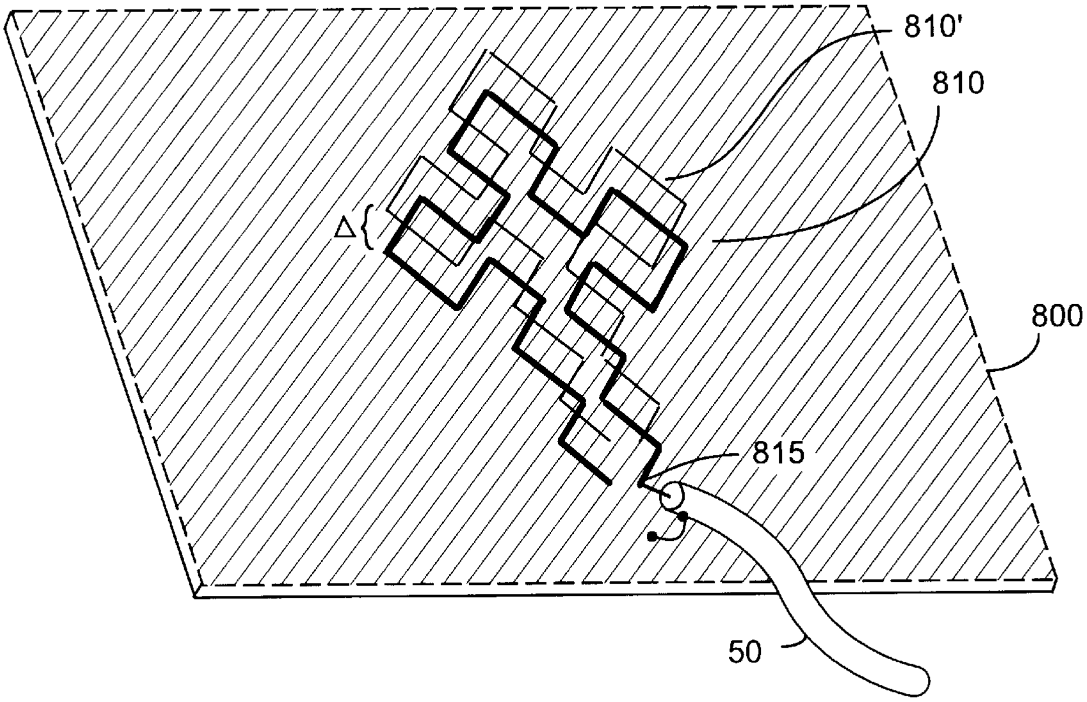
810'
810
Δ
800
815
50

FIGURE 14A

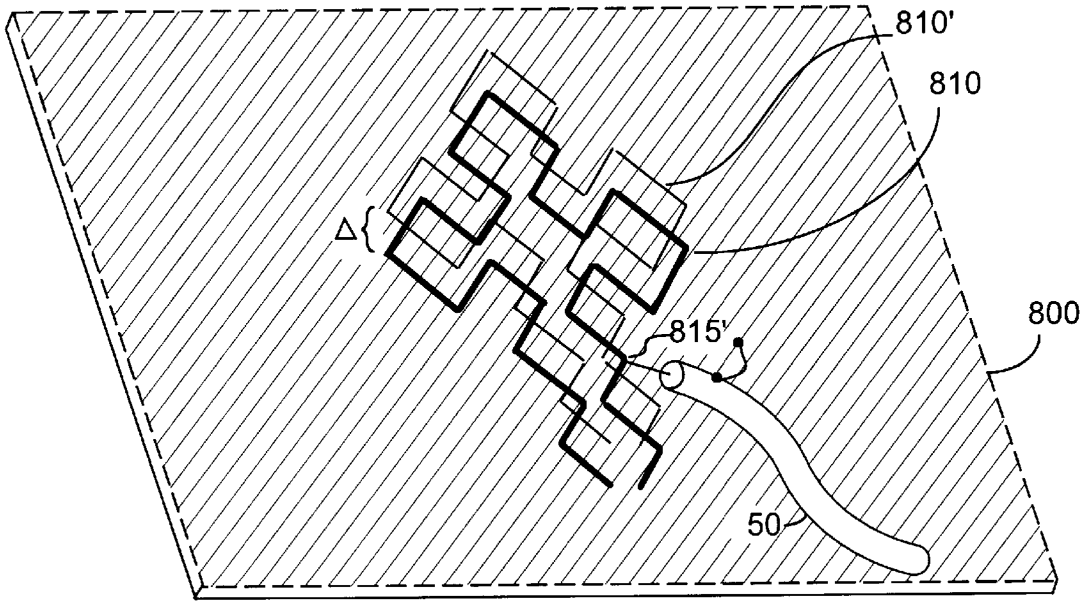
810'
810
Δ
800
815'
50

FIGURE 14B

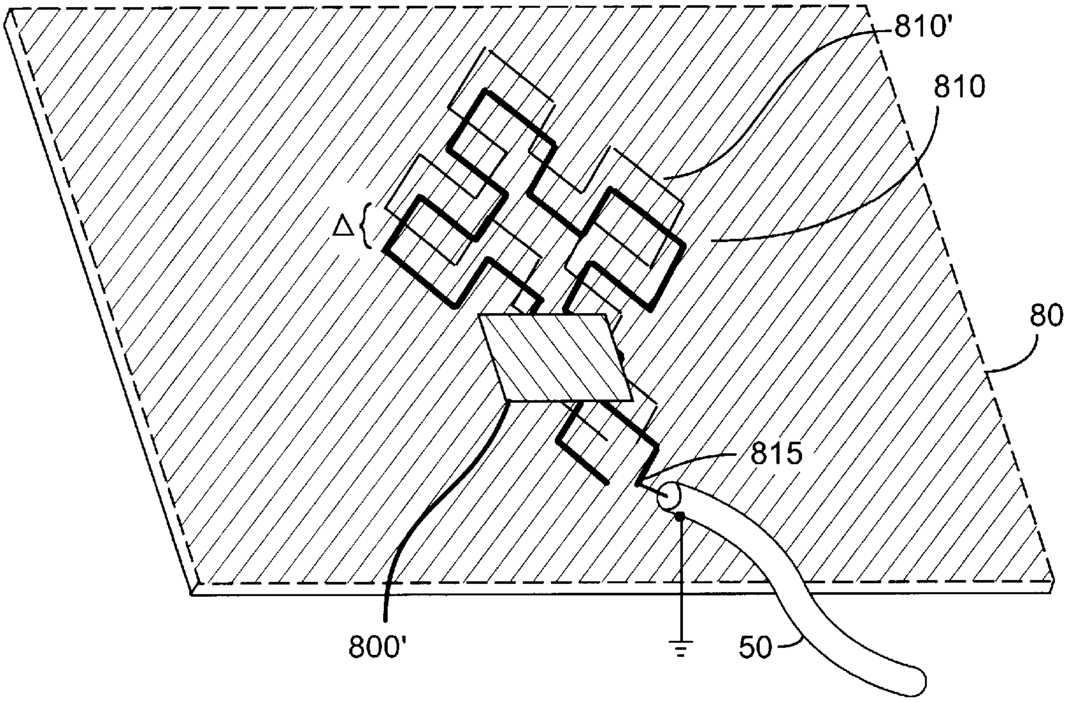
810'
810
Δ
800
815
800'
50

FIGURE 14C

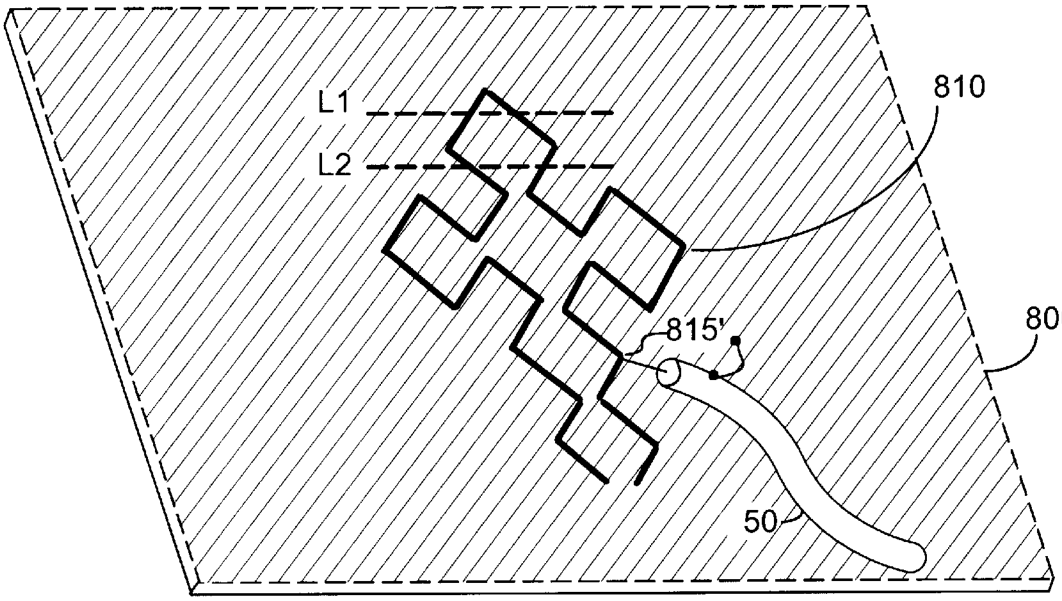
L1
L2
810
800
815'
50

FIGURE 14D

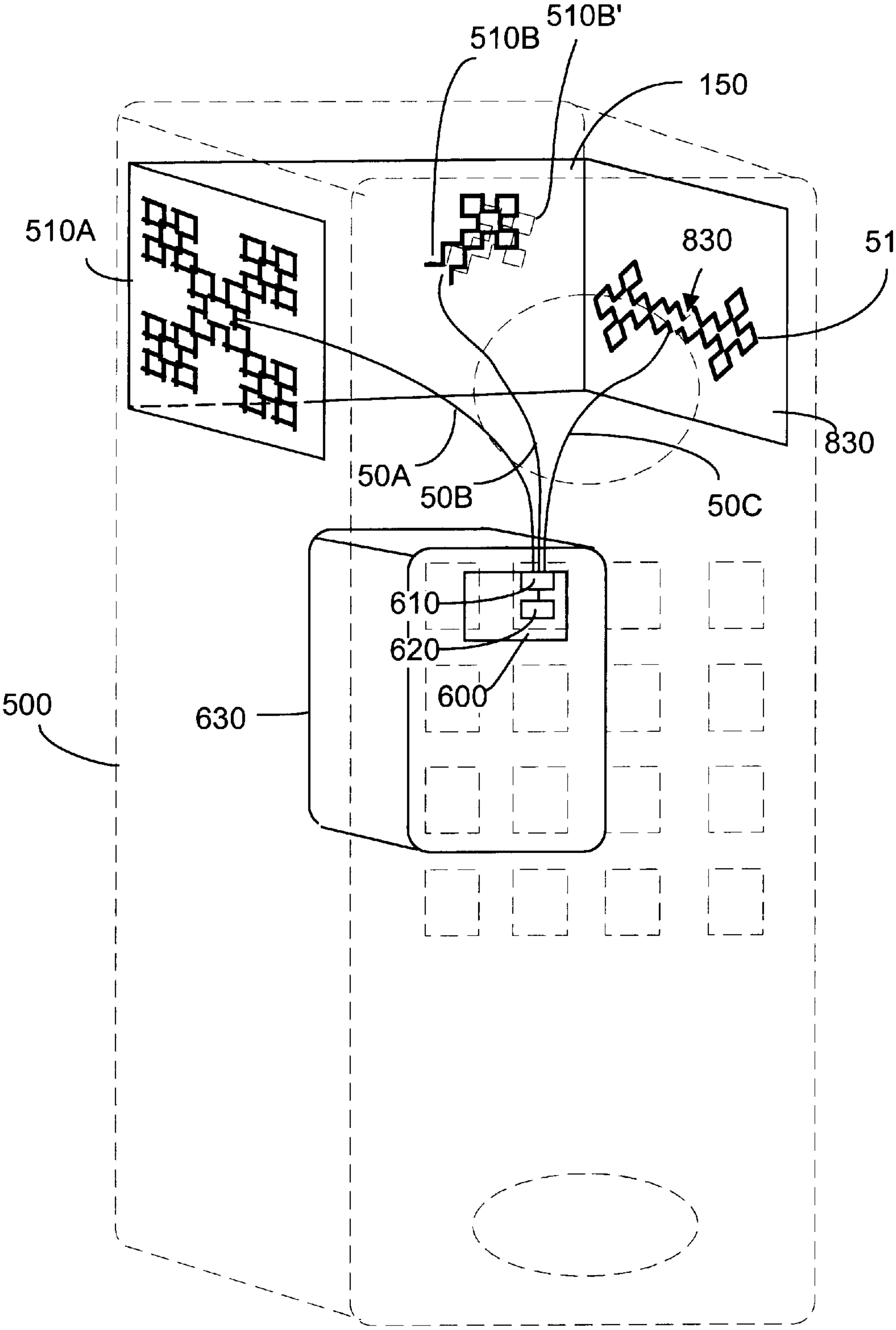
510B'
510B
150
510A
830
510C
830
50A
50B
50C
610
620
600
500
630

FIGURE 15

TUNING FRACTAL ANTENNAS AND FRACTAL RESONATORS

RELATION TO PREVIOUSLY FILED PATENT APPLICATION

This is a continuation, of application Ser. No. 08/609,514 filed Mar. 1, 1996 now abandoned, which is a continuation-in-part application of applicant's patent application Ser. No. 08/512,954 entitled FRACTAL ANTENNAS AND FRACTAL RESONATORS, filed on Aug. 9, 1995, abandoned.

FIELD OF THE INVENTION

The present invention relates to antennas and resonators, and more specifically to tuning non-Euclidian antennas and non-Euclidian resonators.

BACKGROUND OF THE INVENTION

Antenna are used to radiate and/or receive typically electromagnetic signals, preferably with antenna gain, directivity, and efficiency. Practical antenna design traditionally involves trade-offs between various parameters, including antenna gain, size, efficiency, and bandwidth.

Antenna design has historically been dominated by Euclidean geometry. In such designs, the closed antenna area is directly proportional to the antenna perimeter. For example, if one doubles the length of an Euclidean square (or "quad") antenna, the enclosed area of the antenna quadruples. Classical antenna design has dealt with planes, circles, triangles, squares, ellipses, rectangles, hemispheres, paraboloids, and the like, (as well as lines). Similarly, resonators, typically capacitors ("C") coupled in series and/or parallel with inductors ("L"), traditionally are implemented with Euclidian inductors.

With respect to antennas, prior art design philosophy has been to pick a Euclidean geometric construction, e.g., a quad, and to explore its radiation characteristics, especially with emphasis on frequency resonance and power patterns. The unfortunate result is that antenna design has far too long concentrated on the ease of antenna construction, rather than on the underlying electromagnetics.

Many prior art antennas are based upon closed-loop or island shapes. Experience has long demonstrated that small sized antennas, including loops, do not work well, one reason being that radiation resistance ("R") decreases sharply when the antenna size is shortened. A small sized loop, or even a short dipole, will exhibit a radiation pattern of ½λ and ¼λ, respectively, if the radiation resistance R is not swamped by substantially larger ohmic ("O") losses. Ohmic losses can be minimized using impedance matching networks, which can be expensive and difficult to use. But although even impedance matched small loop antennas can exhibit 50% to 85% efficiencies, their bandwidth is inherently narrow, with very high Q, e.g., Q>50. As used herein, Q is defined as (transmitted or received frequency)/(3 dB bandwidth).

As noted, it is well known experimentally that radiation resistance R drops rapidly with small area Euclidean antennas. However, the theoretical basis is not generally known, and any present understanding (or misunderstanding) appears to stem from research by J. Kraus, noted in *Antennas* (Ed. 1), McGraw Hill, New York (1950), in which a circular loop antenna with uniform current was examined. Kraus' loop exhibited a gain with a surprising limit of 1.8 dB over an isotropic radiator as loop area fells below that of a loop

having a 1λ-squared aperture. For small loops of area $A<\lambda^2/100$, radiation resistance R was given by:

$$R = K \cdot \left(\frac{A}{\lambda^2}\right)^2$$

where K is a constant, A is the enclosed area of the loop, and λ is wavelength. Unfortunately, radiation resistance R can all too readily be less than 1 Ω for a small loop antenna.

From his circular loop research Kraus generalized that calculations could be defined by antenna area rather than antenna perimeter, and that his analysis should be correct for small loops of any geometric shape. Kraus' early research and conclusions that small-sized antennas will exhibit a relatively large ohmic resistance O and a relatively small radiation resistance R, such that resultant low efficiency defeats the use of the small antenna have been widely accepted. In fact, some researchers have actually proposed reducing ohmic resistance O to 0 Ω by constructing small antennas from superconducting material, to promote efficiency.

As noted, prior art antenna and resonator design has traditionally concentrated on geometry that is Euclidean. However, one non-Euclidean geometry is fractal geometry. Fractal geometry may be grouped into random fractals, which are also termed chaotic or Brownian fractals and include a random noise components, such as depicted in FIG. **3**, or deterministic fractals such as shown in FIG. **1**C.

In deterministic fractal geometry, a self-similar structure results from the repetition of a design or motif (or "generator"), on a series of different size scales. One well known treatise in this field is *Fractals. Endlessly Repeated Geometrical Figures,* by Hans Lauwerier, Princeton University Press (1991), which treatise applicant refers to and incorporates herein by reference.

FIGS. **1**A–**2**D depict the development of some elementary forms of fractals. In FIG. **1**A, a base element **10** is shown as a straight line, although a curve could instead be used. In FIG. **1**B, a so-called Koch fractal motif or generator **20-1**, here a triangle, is inserted into base element **10**, to form a first order iteration ("N") design, e.g., N=1. In FIG. **1**C, a second order N=2 iteration design results from replicating the triangle motif **20-1** into each segment of FIG. **1**B, but where the **20-1**' version has been differently scaled, here reduced in size. As noted in the Lauwerier treatise, in its replication, the motif may be rotated, translated, scaled in dimension, or a combination of any of these characteristics. Thus, as used herein, second order of iteration or N=2 means the fundamental motif has been replicated, after rotation, translation, scaling (or a combination of each) into the first order iteration pattern. A higher order, e.g., N=3, iteration means a third fractal pattern has been generated by including yet another rotation, translation, and/or scaling of the first order motif.

In FIG. **1**D, a portion of FIG. **1**C has been subjected to a further iteration (N=3) in which scaled-down versions **20-1** of the triangle motif **20-1** have been inserted into each segment of the left half of FIG. **1**C. FIGS. **2**A–**2**C follow what has been described with respect to FIGS. **1**A–**1**C, except that a rectangular motif **20-2** has been adopted, which motif is denoted **20-2**' in FIG. **2**C, and **20-2**" in FIG. **2**D. FIG. **2**D shows a pattern in which a portion of the left-hand side is an N=3 iteration of the **20-2** rectangle motif, and in which the center portion of the figure now includes another motif, here a **20-1** type triangle motif, and in which the right-hand side of the figure remains an N=2 iteration.

Traditionally, non-Euclidean designs including random fractals have been understood to exhibit antiresonance characteristics with mechanical vibrations. It is known in the art to attempt to use non-Euclidean random designs at lower frequency regimes to absorb, or at least not reflect sound due to the antiresonance characteristics. For example, M. Schroeder in *Fractals, Chaos, Power Laws* (1992), W. H. Freeman, New York discloses the use of presumably random or chaotic fractals in designing sound blocking diffusers for recording studios and auditoriums.

Experimentation with non-Euclidean structures has also been undertaken with respect to electromagnetic waves, including radio antennas. In one experiment, Y. Kim and D. Jaggard in *The Fractal Random Array,* Proc. IEEE 74, 1278–1280 (1986) spread-out antenna elements in a sparse microwave array, to minimize sidelobe energy without having to use an excessive number of elements. But Kim and Jaggard did not apply a fractal condition to the antenna elements, and test results were not necessarily better than any other techniques, including a totally random spreading of antenna elements. More significantly, the resultant array was not smaller than a conventional Euclidean design.

Prior art spiral antennas, cone antennas, and V-shaped antennas may be considered as a continuous, deterministic first order fractal, whose motif continuously expands as distance increases from a central point. A log-periodic antenna may be considered a type of continuous fractal in that it is fabricated from a radially expanding structure. However, log periodic antennas do not utilize the antenna perimeter for radiation, but instead rely upon an arc-like opening angle in the antenna geometry. Such opening angle is an angle that defines the size-scale of the log-periodic structure, which structure is proportional to the distance from the antenna center multiplied by the opening angle. Further, known log-periodic antennas are not necessarily smaller than conventional driven element-parasitic element antenna designs of similar gain.

Unintentionally, first order fractals have been used to distort the shape of dipole and vertical antennas to increase gain, the shapes being defined as a Brownian-type of chaotic fractals. See F. Landstorfer and R. Sacher, *Optimisation of Wire Antennas,* J. Wiley, New York (1985). FIG. **3** depicts three bent-vertical antennas developed by Landstorfer and Sacher through trial and error, the plots showing the actual vertical antennas as a function of x-axis and y-axis coordinates that are a function of wavelength. The "EF" and "BF" nomenclature in FIG. **3** refer respectively to end-fire and back-fire radiation patterns of the resultant bent-vertical antennas.

First order fractals have also been used to reduce horn-type antenna geometry, in which a double-ridge horn configuration is used to decrease resonant frequency. See J. Kraus in *Antennas,* McGraw Hill, New York (1885). The use of rectangular, box-like, and triangular shapes as impedance-matching loading elements to shorten antenna element dimensions is also known in the art.

Whether intentional or not, such prior art attempts to use a quasi-fractal or fractal motif in an antenna employ at best a first order iteration fractal. By first iteration it is meant that one Euclidian structure is loaded with another Euclidean structure in a repetitive fashion, using the same size for repetition. FIG. **1**C, for example, is not first order because the **20-1**' triangles have been shrunk with respect to the size of the first motif **20-1**.

Prior art antenna design does not attempt to exploit multiple scale self-similarity of real fractals. This is hardly surprising in view of the accepted conventional wisdom that because such antennas would be anti-resonators, and/or if suitably shrunken would exhibit so small a radiation resistance R, that the substantially higher ohmic losses O would result in too low an antenna efficiency for any practical use. Further, it is probably not possible to mathematically predict such an antenna design, and high order iteration fractal antennas would be increasingly difficult to fabricate and erect, in practice.

FIGS. **4**A and **4**B depict respective prior art series and parallel type resonator configurations, comprising capacitors C and Euclidean inductors L. In the series configuration of FIG. **4**A, a notch-filter characteristic is presented in that the impedance from port A to port B is high except at frequencies approaching resonance, determined by $1/\sqrt{(LC)}$.

In the distributed parallel configuration of FIG. **4**B, a low-pass filter characteristic is created in that at frequencies below resonance, there is a relatively low impedance path from port A to port B, but at frequencies greater than resonant frequency, signals at port A are shunted to ground (e.g., common terminals of capacitors C), and a high impedance path is presented between port A and port B. Of course, a single parallel LC configuration may also be created by removing (e.g., short-circuiting) the rightmost inductor L and right two capacitors C, in which case port B would be located at the bottom end of the leftmost capacitor C.

In FIGS. **4**A and **4**B, inductors L are Euclidean in that increasing the effective area captured by the inductors increases with increasing geometry of the inductors, e.g., more or larger inductive windings or, if not cylindrical, traces comprising inductance. In such prior art configurations as FIGS. **4**A and **4**B, the presence of Euclidean inductors L ensures a predictable relationship between L, C and frequencies of resonance.

Applicant's above-noted FRACTAL ANTENNA AND FRACTAL RESONATORS patent application provides a design methodology that can produce smaller-scale antennas that exhibit at least as much gain, directivity, and efficiency as larger Euclidean counterparts. Such design approach should exploit the multiple scale self-similarity of real fractals, including $N>2$ iteration order fractals. Further, as respects resonators, said application discloses a non-Euclidean resonator whose presence in a resonating configuration can create frequencies of resonance beyond those normally presented in series and/or parallel LC configurations.

However, there is a need for a simple mechanism to tune and/or otherwise adjust such antennas and resonators.

The present invention provides such mechanisms.

SUMMARY OF THE INVENTION

The present invention tunes fractal antenna systems and resonator systems, preferably designed according to applicant's above-reference patent application, by placing an active (or driven) fractal antenna or resonator a distance Δ from a second conductor. Such disposition of the antenna and second conductor advantageously lowers resonant frequencies and widens bandwidth for the fractal antenna. In some embodiments, the fractal antenna and second conductor are non-coplanar and λ is the separation distance therebetween, preferably $\leqq 0.05\lambda$ for the frequency of interest $(1/\lambda)$. In other embodiments, the fractal antenna and second conductive element may be planar, in which case λ a separation distance, measured on the common plane.

The second conductor may in fact be a second fractal antenna of like or unlike configuration as the active antenna. Varying the distance Δ tunes the active antenna and thus the overall system. Further, if the second element, preferably a fractal antenna, is angularly rotated relative to the active antenna, resonant frequencies of the active antenna may be varied.

Providing a cut in the fractal antenna results in new and different resonant nodes, including resonant nodes having perimeter compression parameters, defined below, ranging from about three to ten. If desired, a portion of a fractal antenna may be cutaway and removed so as to tune the antenna by increasing resonance(s).

Tunable fractal antenna systems need not be planar, according to the present invention. Fabricating a fractal antenna around a form such as a torroid ring, or forming the fractal antenna on a flexible substrate that is curved about itself results in field self-proximity that produces resonant frequency shifts. A fractal antenna and a conductive element may each be formed as a curved surface or even as a torroid-shape, and placed in sufficiently close proximity to each other to provide a useful tuning and system characteristic altering mechanism.

In the various embodiments, more than two elements may be used, and tuning may be accomplished by varying one or more of the parameters associated with one or more elements.

Preferably fractal antennas and resonators so tuned are designed according to applicant's above-referenced patent application, which provides an antenna having at least one element whose shape, at least is part, is substantially a deterministic fractal of iteration order $N \geqq 2$. Using fractal geometry, the antenna element has a self-similar structure resulting from the repetition of a design or motif (or "generator") that is replicated using rotation, and/or translation, and/or scaling. The fractal element will have x-axis, y-axis coordinates for a next iteration N+1 defined by $x_{N+1}=f(x_N, yb_N)$ and $y_{N+1}=g(x_N, y_N$, where x_N, y_N define coordinates for a preceding iteration, and where f(x,y) and g(x,y) are functions defining the fractal motif and behavior.

In contrast to Euclidean geometric antenna design, applicant's deterministic fractal antenna elements have a perimeter that is not directly proportional to area. For a given perimeter dimension, the enclosed area of a multi-iteration fractal will always be as small or smaller than the area of a corresponding conventional Euclidean antenna.

A fractal antenna has a fractal ratio limit dimension D given by log(L)/log(r), where L and r are one-dimensional antenna element lengths before and after fractalization, respectively.

As used with the present invention, a fractal antenna perimeter compression parameter (PC) is defined as:

$$PC = \frac{\text{full-sized antenna element length}}{\text{fractal-reduced antenna element length}}$$

where:

$$PC=A\cdot\log\,[N(D+C)]$$

in which A and C are constant coefficients for a given fractal motif, N is an iteration number, and D is the fractal dimension, defined above.

Radiation resistance (R) of a fractal antenna decreases as a small power of the perimeter compression (PC), with a fractal loop or island always exhibiting a substantially higher radiation resistance than a small Euclidean loop antenna of equal size. In the present invention, deterministic fractals are used wherein A and C have large values, and thus provide the greatest and most rapid element-size shrinkage. A fractal antenna according to the present invention will exhibit an increased effective wavelength.

The number of resonant nodes of a fractal loop-shaped antenna according to the present invention increases as the iteration number N and is at least as large as the number of resonant nodes of an Euclidean island with the same area. Further, resonant frequencies of a fractal antenna include frequencies that are not harmonically related.

A fractal antenna according to the present invention is smaller than its Euclidean counterpart but provides at least as much gain and frequencies of resonance and provides essentially a 50 Ω termination impedance at its lowest resonant frequency. Further, the fractal antenna exhibits non-harmonically frequencies of resonance, a low Q and resultant good bandwidth, acceptable standing wave ratio ("SWR"), a radiation impedance that is frequency dependent, and high efficiencies. Fractal inductors of first or higher iteration order may also be provided in LC resonators, to provide additional resonant frequencies including non-harmonically related frequencies.

Other features and advantages of the invention will appear from the following description in which the preferred embodiments have been set forth in detail, in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1**A depicts a base element for an antenna or an inductor, according to the prior art;

FIG. **1**B depicts a triangular-shaped Koch fractal motif, according to the prior art;

FIG. **1**C depicts a second-iteration fractal using the motif of FIG. **1**B, according to the prior art;

FIG. **1**D depicts a third-iteration fractal using the motif of FIG. **1**B, according to the prior art;

FIG. **2**A depicts a base element for an antenna or an inductor, according to the prior art;

FIG. **2**B depicts a rectangular-shaped Minkowski fractal motif, according to the prior art;

FIG. **2**C depicts a second-iteration fractal using the motif of FIG. **2**B, according to the prior art;

FIG. **2**D depicts a fractal configuration including a third-order using the motif of FIG. **2**B, as well as the motif of FIG. **1**B, according to the prior art;

FIG. **3** depicts bent-vertical chaotic fractal antennas, according to the prior art;

FIG. **4**A depicts a series L-C resonator, according to the prior art;

FIG. **4**B depicts a distributed parallel L-C resonator, according to the prior art;

FIG. **5**A depicts an Euclidean quad antenna system, according to the prior art;

FIG. **5**B depicts a second-order Minkowski island fractal quad antenna, according to the present invention;

FIG. **6** depicts an ELNEC-generated free-space radiation pattern for an MI-2 fractal antenna, according to the present invention;

FIG. **7**A depicts a Cantor-comb fractal dipole antenna, according to the present invention;

FIG. **7**B depicts a torn square fractal quad antenna, according to the present invention;

FIG. **7**C-**1** depicts a second iteration Minkowski (MI-2) printed circuit fractal antenna, according to the present invention;

FIG. **7**C-**2** depicts a second iteration Minkowski (MI-2) slot fractal antenna, according to the present invention;

FIG. **7**D depicts a deterministic dendrite fractal vertical antenna, according to the present invention;

FIG. **7**E depicts a third iteration Minkowski island (MI-3) fractal quad antenna, according to the present invention;

FIG. **7**F depicts a second iteration Koch fractal dipole, according to the present invention;

FIG. **7**G depicts a third iteration dipole, according to the present invention;

FIG. **7**H depicts a second iteration Minkowski fractal dipole, according to the present invention;

FIG. **7**I depicts a third iteration multi-fractal dipole, according to the present invention;

FIG. **8**A depicts a generic system in which a passive or active electronic system communicates using a fractal antenna, according to the present invention;

FIG. **8**B depicts a communication system in which several fractal antennas are electronically selected for best performance, according to the present invention;

FIG. **8**C depicts a communication system in which electronically steerable arrays of fractal antennas are electronically selected for best performance, according to the present invention;

FIG. **9**A depicts fractal antenna gain as a function of iteration order N, according to the present invention;

FIG. **9**B depicts perimeter compression PC as a function of iteration order N for fractal antennas, according to the present invention;

FIG. **10**A depicts a fractal inductor for use in a fractal resonator, according to the present invention;

FIG. **10**B depicts a credit card sized security device utilizing a fractal resonator, according to the present invention;

FIG. **11**A depicts an embodiment in which a fractal antenna is spaced-apart a distance Δ from a conductor element to vary resonant properties and radiation characteristics of the antenna, according to the present invention;

FIG. **11**B depicts an embodiment in which a fractal antenna is coplanar with a ground plane and is spaced-apart a distance Δ' from a coplanar passive parasitic element to vary resonant properties and radiation characteristics of the antenna, according to the present invention;

FIG. **12**A depicts spacing-apart first and second fractal antennas a distance Δ to decrease resonance and create additional resonant frequencies for the active or driven antenna, according to the present invention;

FIG. **12**B depicts relative angular rotation between spaced-apart first and second fractal antennas Δ to vary resonant frequencies of the active or driven antenna, according to the present invention;

FIG. **13**A depicts cutting a fractal antenna or resonator to create different resonant nodes and to alter perimeter compression, according to the present invention;

FIG. **13**B depicts forming a non-planar fractal antenna or resonator on a flexible substrate that is curved to shift resonant frequency, apparently due to self-proximity electromagnetic fields, according to the present invention;

FIG. **13**C depicts forming a fractal antenna or resonator on a curved torroidal form to shift resonant frequency, apparently due to self-proximity electromagnetic fields, according to the present invention;

FIG. **14**A depicts forming a fractal antenna or resonator in which the conductive element is not attached to the system coaxial or other feedline, according to the present invention;

FIG. **14**B depicts a system similar to FIG. **14**A, but demonstrates that the driven fractal antenna may be coupled to the system coaxial or other feedline at any point along the antenna, according to the present invention;

FIG. **14**C depicts an embodiment in which a supplemental ground plane is disposed adjacent a portion of the driven fractal antenna and conductive element, forming a sandwich-like system, according to the present invention;

FIG. **14**D depicts an embodiment in which a fractal antenna system is tuned by cutting away a portion of the driven antenna, according to the present invention;

FIG. **15** depicts a communication system similar to that of FIG. **8**A, in which several fractal antennas are tunable and are electronically selected for best performance, according to the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

In overview, the present invention provides an antenna having at least one element whose shape, at least is part, is substantially a fractal of iteration order $N \geqq 2$. The resultant antenna is smaller than its Euclidean counterpart, provides a 50 Ω termination impedance, exhibits at least as much gain and more frequencies of resonance than its Euclidean counterpart, including non-harmonically related frequencies of resonance, exhibits a low Q and resultant good bandwidth, acceptable SWR, a radiation impedance that is frequency dependent, and high efficiencies.

In contrast to Euclidean geometric antenna design, fractal antenna elements according to the present invention have a perimeter that is not directly proportional to area. For a given perimeter dimension, the enclosed area of a multi-iteration fractal area will always be at least as small as any Euclidean area.

Using fractal geometry, the antenna element has a self-similar structure resulting from the repetition of a design or motif (or "generator"), which motif is replicated using rotation, translation, and/or scaling (or any combination thereof). The fractal portion of the element has x-axis, y-axis coordinates for a next iteration N+1 defined by $x_{N+1}=f(x_N, yb_N)$ and $y_{N+1}=g(x_N, y_N)$, where x_N, y_N are coordinates of a preceding iteration, and where f(x,y) and g(x,y) are functions defining the fractal motif and behavior.

For example, fractals of the Julia set may be represented by the form:

$$x_{N+1}=x_N^2-y_N^2+a$$

$$y_{N+1}=2x_N \cdot y_N=b$$

In complex notation, the above may be represented as:

$$z_{N+1}=z_N^2+c$$

Although it is apparent that fractals can comprise a wide variety of forms for functions f(x,y) and g(x,y), it is the iterative nature and the direct relation between structure or morphology on different size scales that uniquely distinguish f(x,y) and g(x,y) from non-fractal forms. Many references including the Lauwerier treatise set forth equations appropriate for f(x,y) and g(x,y).

Iteration (N) is defined as the application of a fractal motif over one size scale. Thus, the repetition of a single size scale of a motif is not a fractal as that term is used herein. Multi-fractals may of course be implemented, in which a motif is changed for different iterations, but eventually at least one motif is repeated in another iteration.

An overall appreciation of the present invention may be obtained by comparing FIGS. **5A** and **5B**. FIG. **5A** shows a conventional Euclidean quad antenna **5** having a driven element **10** whose four sides are each 0.25λlong, for a total perimeter of 1λ, where λ is the frequency of interest.

Euclidean element **10** has an impedance of perhaps 130 Ω, which impedance decreases if a parasitic quad element **20** is spaced apart on a boom **30** by a distance B of 0.1λ to 0.25λ. Parasitic element **20** is also sized S=0.25λ on a side, and its presence can improve directivity of the resultant two-element quad antenna. Element **10** is depicted in FIG. **5A** with heavier lines than element **20**, solely to avoid confusion in understanding the figure. Non-conductive spreaders **40** are used to help hold element **10** together and element **20** together.

Because of the relatively large drive impedance, driven element **10** is coupled to an impedance matching network or device **60**, whose output impedance is approximately 50 Ω. A typically 50 Ω coaxial cable **50** couples device **60** to a transceiver **70** or other active or passive electronic equipment **70**.

As used herein, the term transceiver shall mean a piece of electronic equipment that can transmit, receive, or transmit and receive an electromagnetic signal via an antenna, such as the quad antenna shown in FIG. **5A** or **5B**. As such, the term transceiver includes without limitation a transmitter, a receiver, a transmitter-receiver, a cellular telephone, a wireless telephone, a pager, a wireless computer local area network ("LAN") communicator, a passive resonant unit used by stores as part of an anti-theft system in which transceiver **70** contains a resonant circuit that is blown or not-blown by an electronic signal at time of purchase of the item to which transceiver **70** is affixed, resonant sensors and transponders, and the like.

Further, since antennas according to the present invention can receive incoming radiation and coupled the same as alternating current into a cable, it will be appreciated that fractal antennas may be used to intercept incoming light radiation and to provide a corresponding alternating current. For example, a photocell antenna defining a fractal, or indeed a plurality or array of fractals, would be expected to output more current in response to incoming light than would a photocell of the same overall array size. FIG. **5B** depicts a fractal quad antenna **95**, designed to resonant at the same frequency as the larger prior art antenna **5** shown in FIG. **5A**. Driven element **100** is seen to be a second order fractal, here a so-called Minkowski island fractal, although any of numerous other fractal configurations could instead be used, including without limitation, Koch, torn square, Mandelbrot, Caley tree, monkey's swing, Sierpinski gasket, and Cantor gasket geometry.

If one were to measure the amount of conductive wire or conductive trace comprising the perimeter of element **10** or element **20**, it would be perhaps 40% greater than the 1.0λ for the Euclidean quad of FIG. **5A**. However, for fractal antenna **95**, the physical straight length of one element side KS will be substantially smaller, and for the N=2 fractal antenna shown in FIG. **5B**, KS≈0.13λ (in air), compared with K≈0.25λ for prior art antenna **5**.

However, although the actual perimeter length of element **100** is greater than the 1λ perimeter of prior art element **10**, the area within antenna element **100** is substantially less than the S^2 area of prior art element **10**. As noted, this area independence from perimeter is a characteristic of a deterministic fractal. Boom length B for antenna **95** will be slightly different from length B for prior art antenna **5** shown in FIG. **5A**. In FIG. **5B**, a parasitic element **120**, which preferably is similar to driven element **100** but need not be, may be attached to boom **130**. For ease of illustration FIG. **5B** does not depict non-conductive spreaders, such as spreaders **40** shown in FIG. **5A**, which help hold element **100** together and element **120** together. Further, for ease of understanding the figure, element **10** is drawn with heavier lines than element **120**, to avoid confusion in the portion of the figure in which elements **100** and **120** appear overlapped.

An impedance matching device **60** is advantageously unnecessary for the fractal antenna of FIG. **5B**, as the driving impedance of element **100** is about 50 Ω, e.g., a perfect match for cable **50** if reflector element **120** is absent, and about 35Ω, still an acceptable impedance match for cable **50**, if element **120** is present. Antenna **95** may be fed by cable **50** essentially anywhere in element **100**, e.g., including locations X, Y, Z, among others, with no substantial change in the termination impedance. With cable **50** connected as shown, antenna **95** will exhibit horizontal polarization. If vertical polarization is desired, connection may be made as shown by cable **50'**. If desired, cables **50** and **50'** may both be present, and an electronic switching device **75** at the antenna end of these cables can short-out one of the cables. If cable **50** is shorted out at the antenna, vertical polarization results, and if instead cable **50'** is shorted out at the antenna, horizontal polarization results.

As shown by Table **3** herein, fractal quad **95** exhibits about 1.5 dB gain relative to Euclidean quad 10. Thus, transmitting power output by transceiver **70** may be cut by perhaps 40% and yet the system of FIG. **5B** will still perform no worse than the prior art system of FIG. **5A**. Further, as shown by Table 1, the fractal antenna of FIG. **5B** exhibits more resonance frequencies than the antenna of FIG. **5B**, and also exhibits some resonant frequencies that are not harmonically related to each other. As shown by Table 3, antenna **95** has efficiency exceeding about 92% and exhibits an excellent SWR of about 1.2:1. As shown by Table 5, applicant's fractal quad antenna exhibits a relatively low value of Q. This result is surprising in view of conventional prior art wisdom to the effect that small loop antennas will exhibit high Q.

In short, that fractal quad **95** works at all is surprising in view of the prior art (mis)understanding as to the nature of radiation resistance R and ohmic losses O. Indeed, the prior art would predict that because the fractal antenna of FIG. **5B** is smaller than the conventional antenna of FIG. **5A**, efficiency would suffer due to an anticipated decrease in radiation resistance R. Further, it would have been expected that Q would be unduly high for a fractal quad antenna.

FIG. **6** is an ELNEC-generated free-space radiation pattern for a second-iteration Minkowski fractal antenna, an antenna similar to what is shown in FIG. **5B** with the parasitic element **120** omitted. The frequency of interest was 42.3 MHz, and a 1.5:1 SWR was used. In FIG. **6**, the outer ring represents 2.091 dBi, and a maximum gain of 2.091 dBi. (ELNEC is a graphics/PC version of MININEC, which is a PC version of NEC.) In practice, however, the data shown in FIG. **6** were conservative in that a gain of 4.8 dB above an isotropic reference radiator was actually obtained. The error in the gain figures associated with FIG. **6** presumably is due to roundoff and other limitations inherent in the ELNEC program. Nonetheless, FIG. **6** is believed to accurately depict the relative gain radiation pattern of a single element Minkowski (MI-2) fractal quad according to the present invention.

FIG. **7A** depicts a third iteration Cantor-comb fractal dipole antenna, according to the present invention. Generation of a Cantor-comb involves trisecting a basic shape, e.g., a rectangle, and providing a rectangle of one-third of the basic shape on the ends of the basic shape. The new smaller rectangles are then trisected, and the process repeated. FIG. **7B** is modelled after the Lauwerier treatise, and depicts a single element torn-sheet fractal quad antenna.

FIG. **7C-1** depicts a printed circuit antenna, in which the antenna is fabricated using printed circuit or semiconductor fabrication techniques. For ease of understanding, the etched-away non-conductive portion of the printed circuit board **150** is shown cross-hatched, and the copper or other conductive traces **170** are shown without cross-hatching.

Applicant notes that while various corners of the Minkowski rectangle motif may appear to be touching in this and perhaps other figures herein, in fact no touching occurs. Further, it is understood that it suffices if an element according to the present invention is substantially a fractal. By this it is meant that a deviation of less than perhaps 10% from a perfectly drawn and implemented fractal will still provide adequate fractal-like performance, based upon actual measurements conducted by applicant.

The substrate **150** is covered by a conductive layer of material **170** that is etched away or otherwise removed in areas other than the fractal design, to expose the substrate **150**. The remaining conductive trace portion **170** defines a fractal antenna, a second iteration Minkowski slot antenna in FIG. **7C**. Substrate **150** may be a silicon wafer, a rigid or a flexible plastic-like material, perhaps Mylar™ material, or the non-conductive portion of a printed circuit board. Overlayer **170** may be deposited doped polysilicon for a semiconductor substrate **150**, or copper for a printed circuit board substrate.

FIG. **7C-2** depicts a slot antenna version of what was shown in FIG. **7C-2**, wherein the conductive portion **170** (shown cross-hatched in FIG. **7C-2**) surrounds and defines a fractal-shape of non-conductive substrate **150**. Electrical connection to the slot antenna is made with a coaxial or other cable **50**, whose inner and outer conductors make contact as shown.

In FIGS. **7C-1** and **7C-2**, the substrate or plastic-like material in such constructions can contribute a dielectric effect that may alter somewhat the performance of a fractal antenna by reducing resonant frequency, which increases perimeter compression PC.

Those skilled in the art will appreciate that by virtue of the relatively large amount of conducting material (as contrasted to a thin wire), antenna efficiency is promoted in a slot configuration. Of course a printed circuit board or substrate-type construction could be used to implement a non-slot fractal antenna, e.g, in which the fractal motif is fabricated as a conductive trace and the remainder of the conductive material is etched away or otherwise removed. Thus, in FIG. **7C-1**, for example, if the cross-hatched surface now represents non-conductive material, and the non-cross hatched material represents conductive material, a printed circuit board or substrate-implemented wire-type fractal antenna results.

Printed circuit board and/or substrate-implemented fractal antennas are especially useful at frequencies of 80 MHz or higher, whereat fractal dimensions indeed become small. A 2 M MI-3 fractal antenna (e.g., FIG. **7E**) will measure about 5.5" (14 cm) on a side KS, and an MI-2 fractal antenna (e.g., FIG. **5B**) will be about 7" (17.5 cm) per side KS. As will be seen from FIG. **9A**, an MI-3 antenna suffers a slight loss in gain relative to an MI-2 antenna, but offers substantial size reduction.

Applicant has fabricated an MI-2 Minkowski island fractal antenna for operation in the 850–900 MHz cellular telephone band. The antenna was fabricated on a printed circuit board and measured about 1.2" (3 cm) on a side KS. The antenna was sufficiently small to fit inside applicant's cellular telephone, and performed as well as if the normal attachable "rubber-ducky" whip antenna were still attached. The antenna was found on the side to obtain desired vertical polarization, but could be fed anywhere on the element with 50 Ω impedance still being inherently present. Applicant also fabricated on a printed circuit board an MI-3 Minkowski island fractal quad, whose side dimension KS was about 0.8" (2 cm), the antenna again being inserted inside the cellular telephone. The MI-3 antenna appeared to work as well as the normal whip antenna, which was not attached. Again, any slight gain loss in going from MI-2 to MI-3 (e.g., perhaps 1 dB loss relative to an MI-0 reference quad, or 3 dB los relative to an MI-2) is more than offset by the resultant shrinkage in size. At satellite telephone frequencies of 1650 MHz or so, the dimensions would be approximately halved again. FIGS. **8**A, **8**B and **8**C depict preferred embodiments for such antennas.

FIG. **7D** depicts a 2 M dendrite deterministic fractal antenna that includes a slight amount of randomness. The vertical arrays of numbers depict wavelengths relative to 0λ, at the lower end of the trunk-like element **200**. Eight radial-like elements **210** are disposed at 1.0λ, and various other elements are disposed vertically in a plane along the length of element **200**. The antenna was fabricated using 12 gauge copper wire and was found to exhibit a surprising 20 dBi gain, which is at least 10 dB better than any antenna twice the size of what is shown in FIG. **7D**. Although superficially the vertical of FIG. **7D** may appear analogous to a log-periodic antenna, a fractal vertical according to the present invention does not rely upon an opening angle, in stark contrast to prior art log periodic designs.

FIG. **7E** depicts a third iteration Minkowski island quad antenna (denoted herein as MI-3). The orthogonal line segments associated with the rectangular Minkowski motif make this configuration especially acceptable to numerical study using ELNEC and other numerical tools using moments for estimating power patterns, among other modelling schemes. In testing various fractal antennas, applicant formed the opinion that the right angles present in the Minkowski motif are especially suitable for electromagnetic frequencies.

With respect to the MI-3 fractal of FIG. **7E**, applicant discovered that the antenna becomes a vertical if the center led of coaxial cable **50** is connected anywhere to the fractal, but the outer coaxial braid-shield is left unconnected at the antenna end. (At the transceiver end, the outer shield is connected to ground.) Not only do fractal antenna islands perform as vertical antennas when the center conductor of cable **50** is attached to but one side of the island and the braid is left ungrounded at the antenna, but resonance frequencies for the antenna so coupled are substantially reduced. For example, a 2" (5 cm) sized MI-3 fractal antenna resonated at 70 MHz when so coupled, which is equivalent to a perimeter compression PC≈20.

FIG. **7F** depicts a second iteration Koch fractal dipole, and FIG. **7G** a third iteration dipole. FIG. **7H** depicts a second iteration Minkowski fractal dipole, and FIG. **7I** a third iteration multi-fractal dipole. Depending upon the frequencies of interest, these antennas may be fabricated by bending wire, or by etching or otherwise forming traces on a substrate. Each of these dipoles provides substantially 50 Ω termination impedance to which coaxial cable **50** may be directly coupled without any impedance matching device. It is understood in these figures that the center conductor of cable **50** is attached to one side of the fractal dipole, and the braid outer shield to the other side.

FIG. **8**A depicts a generalized system in which a transceiver **500** is coupled to a fractal antenna system **510** to send electromagnetic radiation **520** and/or receive electromagnetic radiation **540**. A second transceiver **600** shown equipped with a conventional whip-like vertical antenna **610** also sends electromagnetic energy **630** and/or receives electromagnetic energy **540**.

If transceivers **500, 600** are communication devices such as transmitter-receivers, wireless telephones, pagers, or the like, a communications repeating unit such as a satellite **650** and/or a ground base repeater unit **660** coupled to an antenna **670**, or indeed to a fractal antenna according to the present invention, may be present.

Alteratively, antenna **510** in transceiver **500** could be a passive LC resonator fabricated on an integrated circuit microchip, or other similarly small sized substrate, attached to a valuable item to be protected. Transceiver **600**, or indeed unit **660** would then be an electromagnetic transmitter outputting energy at the frequency of resonance, a unit typically located near the cash register checkout area of a store or at an exit.

Depending upon whether fractal antenna-resonator **510** is designed to "blow" (e.g., become open circuit) or to "short" (e.g., become a close circuit) in the transceiver **500** will or will not reflect back electromagnetic energy **540** or **630** to a receiver associated with transceiver **600**. In this fashion, the unauthorized relocation of antenna **510** and/or transceiver **500** can be signalled by transceiver **600**.

FIG. **8**B depicts a transceiver **500** equipped with a plurality of fractal antennas, here shown as **510**A, **510**B, **510**C coupled by respective cables **50**A, **50**B, **50**C to electronics **600** within unit **500**. In the embodiment shown, the antennas are fabricated on a conformal, flexible substrate **150**, e.g., Mylar™ material or the like, upon which the antennas per se may be implemented by printing fractal patterns using conductive ink, by copper deposition, among other methods including printed circuit board and semiconductor fabrication techniques. A flexible such substrate may be conformed to a rectangular, cylindrical or other shape as necessary.

In the embodiment of FIG. **8**B, unit **500** is a handheld transceiver, and antennas **510**A, **510**B, **510**C preferably are fed for vertical polarization, as shown. An electronic circuit **610** is coupled by cables **50**A, **50**B, **50**C to the antennas, and samples incoming signals to discern which fractal antenna, e.g., **510**A, **510**B, **510**C is presently most optimally aligned with the transmitting station, perhaps a unit **600** or **650** or **660** as shown in FIG. **8**A. This determination may be made by examining signal strength from each of the antennas. An electronic circuit **620** then selects the presently best oriented antenna, and couples such antenna to the input of the receiver and output of the transmitter portion, collectively **630**, of unit **500**. It is understood that the selection of the best antenna is dynamic and can change as, for example, a user of **500** perhaps walks about holding the unit, or the transmitting source moves, or due to other changing conditions. In a cellular or a wireless telephone application, the result is more reliable communication, with the advantage that the fractal antennas can be sufficiently small-sized as to fit totally within the casing of unit **500**. Further, if a flexible substrate is used, the antennas may be wrapped about portions of the internal casing, as shown.

An additional advantage of the embodiment of FIG. **8**B is that the user of unit **500** may be physically distanced from the antennas by a greater distance that if a conventional external whip antenna were used. Although medical evidence attempting to link cancer with exposure to electromagnetic radiation from handheld transceivers is still inconclusive, the embodiment of FIG. **8**B appears to minimize any such risk.

FIG. **8**C depicts yet another embodiment wherein some or all of the antenna systems **510**A, **510**B, **510**C may include electronically steerable arrays, including arrays of fractal antennas of differing sizes and polarization orientations. Antenna system **510**C, for example may include similarly designed fractal antennas, e.g., antenna F-**3** and F-**4**, which are differently oriented from each other. Other antennas within system **510**C may be different in design from either of F-**3**, F-**4**. Fractal antenna F-**1** may be a dipole for example. Leads from the various antennas in system **510**C may be coupled to an integrated circuit **690**, mounted on substrate **150**. Circuit **690** can determine relative optimum choice between the antennas comprising system **510**C, and output via cable **50**C to electronics **600** associated with the transmitter and/or receiver portion of unit **630**.

Another antenna system **510**B may include a steerable array of identical fractal antennas, including fractal antenna F-**5** and F-**6**. An integrated circuit **690** is coupled to each of the antennas in the array, and dynamically selects the best antenna for signal strength and coupled such antenna via cable **50**B to electronics **600**. A third antenna system **510**A may be different from or identical to either of system **510**B and **510**C.

Although FIG. **8**C depicts a unit **500** that may be handheld, unit **500** could in fact be a communications system for use on a desk or a field mountable unit, perhaps unit **660** as shown in FIG. **8**A.

For ease of antenna matching to a transceiver load, resonance of a fractal antenna was defined as a total impedance falling between about 20 Ω to 200 Ω, and the antenna was required to exhibit medium to high Q, e.g., frequency/Δfrequency. In practice, applicants' various fractal antennas were found to resonate in at least one position of the antenna feedpoint, e.g., the point at which coupling was made to the antenna. Further, multi-iteration fractals according to the present invention were found to resonate at multiple frequencies, including frequencies that were non-harmonically related.

Contrary to conventional wisdom, applicant found that island-shaped fractals (e.g., a closed loop-like configuration) do not exhibit significant drops in radiation resistance R for decreasing antenna size. As described herein, fractal antennas were constructed with dimensions of less than 12" across (30.48 cm) and yet resonated in a desired 60 MHz to 100 MHz frequency band.

Applicant further discovered that antenna perimeters do not correspond to lengths that would be anticipated from measured resonant frequencies, with actual lengths being longer than expected. This increase in element length appears to be a property of fractals as radiators, and not a result of geometric construction. A similar lengthening effect was reported by Pfeiffer when constructing a full-sized quad antenna using a first order fractal, see A. Pfeiffer, *The Pfeiffer Quad Antenna System,* QST, p. 28–32 (March 1994).

If L is the total initial one-dimensional length of a fractal pre-motif application, and r is the one-dimensional length post-motif application, the resultant fractal dimension D (actually a ratio limit) is:

$$D=\log(L)/\log(r)$$

With reference to FIG. **1**A, for example, the length of FIG. **1**A represents L, whereas the sum of the four line segments comprising the Koch fractal of FIG. **1**B represents r.

Unlike mathematical fractals, fractal antennas are not characterized solely by the ratio D. In practice D is not a good predictor of how much smaller a fractal design antenna may be because D does not incorporate the perimeter lengthening of an antenna radiating element.

Because D is not an especially useful predictive parameter in fractal antenna design, a new parameter "perimeter compression" ("PC") shall be used, where:

$$PC = \frac{\text{full-sized antenna element length}}{\text{fractal-reduced antenna element length}}$$

In the above equation, measurements are made at the fractal-resonating element's lowest resonant frequency. Thus, for a full-sized antenna according to the prior art PC=1, while PC=3 represents a fractal antenna according to the present invention, in which an element side has been reduced by a factor of three.

Perimeter compression may be empirically represented using the fractal dimension D as follows:

$$PC = A \cdot \log[N(D+C)]$$

where A and C are constant coefficients for a given fractal motif, N is an iteration number, and D is the fractal dimension, defined above.

It is seen that for each fractal, PC becomes asymptotic to a real number and yet does not approach infinity even as the iteration number N becomes very large. Stated differently, the PC of a fractal radiator asymptotically approaches a non-infinite limit in a finite number of fractal iterations. This result is not a representation of a purely geometric fractal.

That some fractals are better resonating elements than other fractals follows because optimized fractal antennas approach their asymptotic PCs in fewer iterations than non-optimized fractal antennas. Thus, better fractals for antennas will have large values for A and C, and will provide the greatest and most rapid element-size shrinkage. Fractal used may be deterministic or chaotic. Deterministic fractals have a motif that replicates at a 100% level on all size scales, whereas chaotic fractals include a random noise component.

Applicant found that radiation resistance of a fractal antenna decreases as a small power of the perimeter compression (PC), with a fractal island always exhibiting a substantially higher radiation resistance than a small Euclidean loop antenna of equal size.

Further, it appears that the number of resonant nodes of a fractal island increase as the iteration number (N) and is always greater than or equal to the number of resonant nodes of an Euclidean island with the same area.

Finally, it appears that a fractal resonator has an increased effective wavelength.

The above findings will now be applied to experiments conducted by applicant with fractal resonators shaped into closed-loops or islands. Prior art antenna analysis would predict no resonance points, but as shown below, such is not the case.

A Minkowski motif is depicted in FIGS. **2B–2D**, **5B**, **7C** and **7E**. The Minkowski motif selected was a three-sided box (e.g., **20-2** in FIG. **2B**) placed atop a line segment. The box sides may be any arbitrary length, e.g, perhaps a box height and width of 2 units with the two remaining base sides being of length three units (see FIG. **2B**). For such a configuration, the fractal dimension D is as follows:

$$D = \frac{\log(L)}{\log(r)} = \frac{\log(12)}{\log(8)} = \frac{1.08}{0.90} = 1.20$$

It will be appreciated that D=1.2 is not especially high when compared to other deterministic fractals.

Applying the motif to the line segment may be most simply expressed by a piecewise function f(x) as follows:

$$\begin{aligned} f(x) &= 0 && 0 \geq x \geq \frac{3x_{\max}}{8} \\ f(x) &= \frac{1}{4x_{\max}} && \frac{3x_{\max}}{8} \geq x \geq \frac{5x_{\max}}{8} \\ f(x) &= 0 && \frac{5x_{\max}}{8} \geq x \geq x_{\max} \end{aligned}$$

where x_{max} is the largest continuous value of x on the line segment.

A second iteration may be expressed as $f(x)_2$ relative to the first iteration $f(x)_1$ by:

$$f(x)_2 = f(x)_1 + f(x)$$

where x_{max} is defined in the above-noted piecewise function. Note that each separate horizontal line segment will have a different lower value of x and x_{max}. Relevant offsets from zero may be entered as needed, and vertical segments may be "boxed" by 90° rotation and application of the above methodology.

As shown by FIGS. **5B** and **7E**, a Minkowski fractal quickly begins to appear like a Moorish design pattern. However, each successive iteration consumes more perimeter, thus reducing the overall length of an orthogonal line segment. Four box or rectangle-like fractals of the same iteration number N may be combined to create a Minkowski fractal island, and a resultant "fractalized" cubical quad.

An ELNEC simulation was used as a guide to far-field power patterns, resonant frequencies, and SWRs of Minkowski Island fractal antennas up to iteration N=2. Analysis for N>2 was not undertaken due to inadequacies in the test equipment available to applicant.

The following tables summarize applicant's ELNEC simulated fractal antenna designs undertaken to derive lowest frequency resonances and power patterns, to and including iteration N=2. All designs were constructed on the x,y axis, and for each iteration the outer length was maintained at 42" (106.7 cm).

Table 1, below, summarizes ELNEC-derived far field radiation patterns for Minkowski island quad antennas for each iteration for the first four resonances. In Table 1, each iteration is designed as MI-N for Minkowski Island of iteration N. Note that the frequency of lowest resonance decreased with the fractal Minkowski Island antennas, as compared to a prior art quad antenna. Stated differently, for a given resonant frequency, a fractal Minkowski Island antenna will be smaller than a conventional quad antenna.

TABLE 1

Antenna	Res. Freq. (MHz)	Gain (dBi)	SWR	PC (for 1st)	Direction
Ref. Quad	76	3.3	2.5	1	Broadside
	144	2.8	5.3	—	Endfire

TABLE 1-continued

Antenna	Res. Freq. (MHz)	Gain (dBi)	SWR	PC (for 1st)	Direction
	220	3.1	5.2	—	Endfire
	294	5.4	4.5	—	Endfire
MI-1	55	2.6	1.1	1.38	Broadside
	101	3.7	1.4	—	Endfire
	142	3.5	5.5	—	Endfire
	198	2.7	3.3	—	Broadside
MI-2	43.2	2.1	1.5	1.79	Broadside
	85.5	4.3	1.8	—	Endfire
	102	2.7	4.0	—	Endfire
	116	1.4	5.4	—	Broadside

It is apparent from Table 1 that Minkowski island fractal antennas are multi-resonant structures having virtually the same gain as larger, full-sized conventional quad antennas. Gain figures in Table 1 are for "free-space" in the absence of any ground plane, but simulations over a perfect ground at 1λ yielded similar gain results. Understandably, there will be some inaccuracy in the ELNEC results due to round-off and undersampling of pulses, among other factors.

Table 2 presents the ratio of resonant ELNEC-derived frequencies for the first four resonance nodes referred to in Table 1.

TABLE 2

Antenna	SWR	SWR	SWR	SWR
Ref. Quad (MI-0)	1:1	1:1.89	1:2.89	3.86:1
MI-1	1:1	1:1.83	1:2.58	3.6:1
MI-2	1:1	2.02:1	2.41:1	2.74:1

Tables 1 and 2 confirm the shrinking of a fractal-designed antenna, and the increase in the number of resonance points. In the above simulations, the fractal MI-2 antenna exhibited four resonance nodes before the prior art reference quad exhibited its second resonance. Near fields in antennas are very important, as they are combined in multiple-element antennas to achieve high gain arrays. Unfortunately, programming limitations inherent in ELNEC preclude serious near field investigation. However, as described later herein, applicant has designed and constructed several different high gain fractal arrays that exploit the near field.

Applicant fabricated three Minkowski Island fractal antennas from aluminum #8 and/or thinner #12 galvanized groundwire. The antennas were designed so the lowest operating frequency fell close to a desired frequency in the 2 M (144 MHz) amateur radio band to facilitate relative gain measurements using 2 M FM repeater stations. The antennas were mounted for vertical polarization and placed so their center points were the highest practical point above the mounting platform. For gain comparisons, a vertical ground plane having three reference radials, and a reference quad were constructed, using the same sized wire as the fractal antenna being tested. Measurements were made in the receiving mode.

Multi-path reception was minimized by careful placement of the antennas. Low height effects were reduced and free space testing approximated by mounting the antenna test platform at the edge of a third-store window, affording a 3.5λ height above ground, and line of sight to the repeater, 45 miles (28 kM) distant. The antennas were stuck out of the window about 0.8λ from any metallic objects and testing was repeated on five occasions from different windows on the same floor, with test results being consistent within ½ dB for each trial.

Each antenna was attached to a short piece of 9913 50 Ω coaxial cable, fed at right angles to the antenna. A 2 M transceiver was coupled with 9913 coaxial cable to two precision attenuators to the antenna under test. The transceiver S-meter was coupled to a volt-ohm meter to provide signal strength measurements The attenuators were used to insert initial threshold to avoid problems associated with non-linear S-meter readings, and with S-meter saturation in the presence of full squelch quieting.

Each antenna was quickly switched in for volt-ohmmeter measurement, with attenuation added or subtracted to obtain the same meter reading as experienced with the reference quad. All readings were corrected for SWR attenuation. For the reference quad, the SWR was 2.4:1 for 120 Ω impedance, and for the fractal quad antennas SWR was less than 1.5:1 at resonance. The lack of a suitable noise bridge for 2 M precluded efficiency measurements for the various antennas. Understandably, anechoic chamber testing would provide even more useful measurements.

For each antenna, relative forward gain and optimized physical orientation were measured. No attempt was made to correct for launch-angle, or to measure power patterns other than to demonstrate the broadside nature of the gain. Difference of ½ dB produced noticeable S-meter deflections, and differences of several dB produced substantial meter deflection. Removal of the antenna from the receiver resulted in a 20^{+}dB drop in received signal strength. In this fashion, system distortions in readings were cancelled out to provide more meaningful results. Table 3 summarizes these results.

TABLE 3

Antenna	PC	PL	SWR	Cor. Gain (dB)	Sidelength (λ)
Quad	1	1	2.4:1	0	0.25
¼ wave	1	—	1.5:1	−1.5	0.25
MI-1	1.3	1.2	1.3:1	1.5	0.13
MI-2	1.9	1.4	1.3:1	1.5	0.13
MI-3	2.4	1.7	1:1	−1.2	0.10

It is apparent from Table 3 that for the vertical configurations under test, a fractal quad according to the present invention either exceeded the gain of the prior art test quad, or had a gain deviation of not more than 1 dB from the test quad. Clearly, prior art cubical (square) quad antennas are not optimized for gain. Fractally shrinking a cubical quad by a factor of two will increase the gain, and further shrinking will exhibit modest losses of 1–2 dB.

Versions of a MI-2 and MI-3 fractal quad antennas were constructed for the 6 M (50 MHz) radio amateur band. An RX 50 Ω noise bridge was attached between these antennas and a transceiver. The receiver was nulled at about 54 MHz and the noise bridge was calibrated with 5 Ω and 10 Ω resistors. Table 4 below summarizes the results, in which almost no reactance was seen.

TABLE 4

Antenna	SWR	Z (Ω)	O (Ω)	E (%)
Quad (MI-0)	2.4:1	120	5–10	92–96
MI-2	1.2:1	60	≦5	≧92
MI-3	1.1:1	55	≦5	≧91

In Table 4, efficiency (E) was defined as 100%*(R/Z), where Z was the measured impedance, and R was Z minus ohmic impedance and reactive impedances (0). As shown in Table 4, fractal MI-2 and MI-3 antennas with their low

≦1.2:1 SWR and low ohmic and reactive impedance provide extremely high efficiencies, $90^{+}\%$. These findings are indeed surprising in view of prior art teachings stemming from early Euclidean small loop geometries. In fact, Table 4 strongly suggests that prior art associations of low radiation impedances for small loops must be abandoned in general, to be invoked only when discussing small Euclidean loops. Applicant's MI-3 antenna was indeed micro-sized, being dimensioned at about 0.1λ per side, an area of about $\lambda^2/1{,}000$, and yet did not signal the onset of inefficiency long thought to accompany smaller sized antennas.

However the 6M efficiency data do not explain the fact that the MI-3 fractal antenna had a gain drop of almost 3 dB relative to the MI-2 fractal antenna. The low ohmic impedances of ≦5 Ω strongly suggest that the explanation is other than inefficiency, small antenna size notwithstanding. It is quite possible that near field diffraction effects occur at higher iterations that result in gain loss. However, the smaller antenna sizes achieved by higher iterations appear to warrant the small loss in gain.

Using fractal techniques, however, 2 M quad antennas dimensioned smaller than 3" (7.6 cm) on a side, as well as 20 M (14 MHz) quads smaller than 3' (1 m) on a side can be realized. Economically of greater interest, fractal antennas constructed for cellular telephone frequencies (850 MHz) could be sized smaller than 0.5" (1.2 cm). As shown by FIGS. **8**B and **8**C, several such antenna, each oriented differently could be fabricated within the curved or rectilinear case of a cellular or wireless telephone, with the antenna outputs coupled to a circuit for coupling to the most optimally directed of the antennas for the signal then being received. The resultant antenna system would be smaller than the "rubber-ducky" type antennas now used by cellular telephones, but would have improved characteristics as well.

Similarly, fractal-designed antennas could be used in handheld military walkie-talkie transceivers, global positioning systems, satellites, transponders, wireless communication and computer networks, remote and/or robotic control systems, among other applications.

Although the fractal Minkowski island antenna has been described herein, other fractal motifs are also useful, as well as non-island fractal configurations.

Table 5 demonstrates bandwidths ("BW") and multi-frequency resonances of the MI-2 and MI-3 antennas described, as well as Qs, for each node found for 6 M versions between 30 MHz and 175 MHz. Irrespective of resonant frequency SWR, the bandwidths shown are SWR 3:1 values. Q values shown were estimated by dividing resonant frequency by the 3:1 SWR BW. Frequency ratio is the relative scaling of resonance nodes.

TABLE 5

Antenna	Freq. (MHz)	Freq. Ratio	SWR	3:1 BW	Q
MI-3	53.0	1	1:1	6.4	8.3
	80.1	1.5:1	1.1:1	4.5	17.8
	121.0	2.3:1	2.4:1	6.8	17.7
MI-2	54.0	1	1:1	3.6	15.0
	95.8	1.8:1	1.1:1	7.3	13.1
	126.5	2.3:1	2.4:1	9.4	13.4

The Q values in Table 5 reflect that MI-2 and MI-3 fractal antennas are multiband. These antennas do not display the very high Qs seen in small tuned Euclidean loops, and there appears not to exist a mathematical application to electromagnetics for predicting these resonances or Qs. One approach might be to estimate scalar and vector potentials in Maxwell's equations by regarding each Minkowski Island iteration as a series of vertical and horizontal line segments with offset positions. Summation of these segments will lead to a Poynting vector calculation and power pattern that may be especially useful in better predicting fractal antenna characteristics and optimized shapes.

In practice, actual Minkowski Island fractal antennas seem to perform slightly better than their ELNEC predictions, most likely due to inconsistencies in ELNEC modelling or ratios of resonant frequencies, PCs, SWRs and gains.

Those skilled in the art will appreciate that fractal multiband antenna arrays may also be constructed. The resultant arrays will be smaller than their Euclidean counterparts, will present less wind area, and will be mechanically rotatable with a smaller antenna rotator.

Further, fractal antenna configurations using other than Minkowski islands or loops may be implemented. Table 6 shows the highest iteration number N for other fractal configurations that were found by applicant to resonant on at least one frequency.

TABLE 6

Fractal	Maximum Iteration
Koch	5
Torn Square	4
Minkowski	3
Mandelbrot	4
Caley Tree	4
Monkey's Swing	3
Sierpinski Gasket	3
Cantor Gasket	3

FIG. **9**A depicts gain relative to an Euclidean quad (e.g., an MI-0) configuration as a function of iteration value N. (It is understood that an Euclidean quad exhibits 1.5 dB gain relative to a standard reference dipole.) For first and second order iterations, the gain of a fractal quad increases relative to an Euclidean quad. However, beyond second order, gain drops off relative to an Euclidean quad. Applicant believes that near field electromagnetic energy diffraction-type cancellations may account for the gain loss for N>2. Possibly the far smaller areas found in fractal antennas according to the present invention bring this diffraction phenomenon into sharper focus.

n practice, applicant could not physically bend wire for a 4th or 5th iteration 2 M Minkowski fractal antenna, although at lower frequencies the larger antenna sizes would not present this problem. However, at higher frequencies, printed circuitry techniques, semiconductor fabrication techniques as well as machine-construction could readily produce N=4, N=5, and higher order iterations fractal antennas.

In practice, a Minkowski island fractal antenna should reach the theoretical gain limit of about 1.7 dB seen for sub-wavelength Euclidean loops, but N will be higher than 3. Conservatively, however, an N=4 Minkowski Island fractal quad antenna should provide a PC=3 value without exhibiting substantial inefficiency.

FIG. **9**B depicts perimeter compression (PC) as a function of iteration order N for a Minkowski island fractal configuration. A conventional Euclidean quad (MI-0) has PC=1 (e.g., no compression), and as iteration increases, PC increases. Note that as N increases and approaches 6, PC approaches a finite real number asymptotically, as predicted. Thus, fractal Minkowski Island antennas beyond iteration N=6 may exhibit diminishing returns for the increase in iteration.

It will be appreciated that the non-harmonic resonant frequency characteristic of a fractal antenna according to the present invention may be used in a system in which the frequency signature of the antenna must be recognized to pass a security test. For example, at suitably high frequencies, perhaps several hundred MHz, a fractal antenna could be implemented within an identification credit card. When the card is used, a transmitter associated with a credit card reader can electronically sample the frequency resonance of the antenna within the credit card. If and only if the credit card antenna responds with the appropriate frequency signature pattern expected may the credit card be used, e.g., for purchase or to permit the owner entrance into an otherwise secured area.

FIG. **10**A depicts a fractal inductor L according to the present invention. In contrast to a prior art inductor, the winding or traces with which L is fabricated define, at least in part, a fractal. The resultant inductor is physically smaller than its Euclidean counterpart. Inductor L may be used to form a resonator, including resonators such as shown in FIGS. **4**A and **4**B. As such, an integrated circuit or other suitably small package including fractal resonators could be used as part of a security system in which electromagnetic radiation, perhaps from transmitter **600** or **660** in FIG. **8**A will blow, or perhaps not blow, an LC resonator circuit containing the fractal antenna. Such applications are described elsewhere herein and may include a credit card sized unit **700**, as shown in FIG. **10**B, in which an LC fractal resonator **710** is implemented. (Card **700** is depicted in FIG. **10**B as though its upper surface were transparent.)

The foregoing description has largely replicated what has been set forth in applicant's above-noted FRACTAL ANTENNAS AND FRACTAL RESONATORS patent application. The following section will set forth methods and techniques for tuning such fractal antennas and resonators. In the following description, although the expression "antenna" may be used in referring to a preferably fractal element, in practice what is being described is an antenna or filter-resonator system. As such, an "antenna" can be made to behave as through it were a filter, e.g., passing certain frequencies and rejecting other frequencies (or the converse).

In one group of embodiments, applicant has discovered that disposing a fractal antenna a distance Δ that is in close proximity (e.g., less than about 0.05λ for the frequency of interest) from a conductor advantageously can change the resonant properties and radiation characteristics of the antenna (relative to such properties and characteristics when such close proximity does not exist, e.g., when the spaced-apart distance is relatively great. For example, in FIG. **11**A a conductive surface **800** is disposed a distance Δ behind or beneath a fractal antenna **810**, which in FIG. **11**A is a single arm of an MI-2 fractal antenna. Of course other fractal configurations such as disclosed herein could be used instead of the MI-1 configuration shown, and non-planar configurations may also be used. Fractal antenna **810** preferably is fed with coaxial cable feedline **50**, whose center conductor is attached to one end **815** of the fractal antenna, and whose outer shield is grounded to the conductive plane **800**. As described herein, great flexibility in connecting the antenna system shown to a preferably coaxial feedline exists. Termination impedance is approximately of similar magnitudes as described earlier herein.

In the configuration shown, the relative close proximity between conductive sheet **800** and fractal antenna **810** lowers the resonant frequencies and widens the bandwidth of antenna **810**. The conductive sheet **800** may be a plane of metal, the upper copper surface of a printed circuit board, a region of conductive material perhaps sprayed onto the housing of a device employing the antenna, for example the interior of a transceiver housing **500**, such as shown in FIGS. **8**A, **8**B, **8**C, and **15**.

The relationship between Δ, wherein $\Delta \leqq 0.05\lambda$, and resonant properties and radiation characteristics of a fractal antenna system is generally logarithmic. That is, resonant frequency decreases logarithmically with decreasing separation Δ.

FIG. **11**B shows an embodiment in which a preferably fractal antenna **810** lies in the same plane as a ground plane **800** but is separated therefrom by an insulating region, and in which a passive or parasitic element **800'** is disposed "within" and spaced-apart a distance Δ' from the antenna, and also being coplanar. For example, the embodiment of FIG. **11**B may be fabricated from a single piece of printed circuit board material in which copper (or other conductive material) remains to define the groundplane **800**, the antenna **810**, and the parasitic element **800'**, the remaining portions of the original material having been etched away to form the "moat-like" regions separating regions **800**, **810**, and **800'**. Changing the shape and/or size of element **800'** and/or the coplanar spaced-apart distance Δ' tunes the antenna system shown. For example, for a center frequency in the 900 MHz range, element **800'** measured about 63 mm×8 mm, and elements **810** and **800** each measured about 25 mm×12 mm. In general, element **800** should be at least as large as the preferably fractal antenna **810**. For this configuration, the system shown exhibited a bandwidth of about 200 MHz, and could be made to exhibit characteristics of a bandpass filter and/or band rejection filter. In this embodiment, a coaxial feedline **50** was used, in which the center lead was coupled to antenna **810**, and the ground shield lead was coupled to groundplane **800**. In FIG. **11**B, the inner perimeter of groundplane region **800** is shown as being rectangularly shaped. If desired, this inner perimeter could be moved closer to the outer perimeter of preferably fractal antenna **810**, and could in fact define a perimeter shape that follows the perimeter shape of antenna **810**. In such an embodiment, the perimeter of the inner conductive region **800'** and the inner perimeter of the ground plane region **800** would each follow the shape of antenna **810**. Based upon experiments to date, it is applicant's belief that moving the inner perimeter of ground plane region **800** sufficiently close to antenna **810** could also affect the characteristics of the overall antenna/resonator system.

Referring now to FIG. **12**A, if the conductive surface **800** is replaced with a second fractal antenna **810'**, which is spaced-apart a distance Δ that preferably does not exceed about 0.05λ, resonances for the radiating fractal antenna **810** are lowered and advantageously new resonant frequencies emerge. For ease of fabrication, it may be desired to construct antenna **810** on the upper or first surface **820**A of a substrate **820**, and to construct antenna **810'** on the lower or second surface **820**B of the same substrate. The substrate could be doubled-side printed circuit board type material, if desired, wherein antennas **810**, **810'** are fabricated using printed circuit type techniques. The substrate thickness Δ is selected to provide the desired performance for antenna **810** at the frequency of interest. Substrate **820** may, for example, be a non-conductive film, flexible or otherwise. To avoid cluttering FIGS. **12**A and **12**B, substrate **820** is drawn with phantom lines, as if the substrate were transparent.

Preferably, the center conductor of coaxial cable **50** is connected to one end **815** of antenna **810**, and the outer conductor of cable **50** is connected to a free end **815'** of antenna **810**', which is regarded as ground, although other feedline connections may be used. Although FIG. **12**A depicts antenna **810**' as being substantially identical to antenna **810**, the two antennas could in fact have different configurations.

Applicant has discovered that if the second antenna **810**' is rotated some angle θ relative to antenna **810**, the resonant frequencies of antenna **810** may be varied, analogously to tuning a variable capacitor. Thus, in FIG. **12**B, antenna **810** is tuned by rotating antenna **810**' relative to antenna **810** (or the converse, or by rotating each antenna). If desired, substrate **820** could comprise two substrates each having thickness Δ/2 and pivotally connected together, e.g., with a non-conductive rivet, so as to permit rotation of the substrates and thus relative rotation of the two antennas. Those skilled in the mechanical arts will appreciate that a variety of "tuning" mechanisms could be implemented to permit fine control over the angle θ in response, for example, to rotation of a tunable shaft.

Referring now to FIG. **13**A, applicant has discovered that creating at least one cut or opening **830** in a fractal antenna **810** (here comprising two legs of an MI-2 antenna) results in new and entirely different resonant nodes for the antenna. Further, these nodes can have perimeter compression (PC) ranging from perhaps three to about ten. The precise location of cut **830** on the fractal antenna or resonator does not appear to be critical.

FIGS. **13**B and **13**C depict a self-proximity characteristic of fractal antennas and resonators that may advantageously be used to create a desired frequency resonant shift. In FIG. **13**B, a fractal antenna **810** is fabricated on a first surface **820**A of a flexible substrate **820**, whose second surface **820**B does not contain an antenna or other conductor in this embodiment. Curving substrate **820**, which may be a flexible film, appears to cause electromagnetic fields associated with antenna **810** to be sufficiently in self-proximity so as to shift resonant frequencies. Such self-proximity antennas or resonators may be referred to a com-cyl devices. The extent of curvature may be controlled where a flexible substrate or substrate-less fractal antenna and/or conductive element is present, to control or tune frequency dependent characteristics of the resultant system. Com-cyl embodiments could include a concentrically or eccentrically disposed fractal antenna and conductive element. Such embodiments may include telescopic elements, whose extent of "overlap" may be telescopically adjusted by contracting or lengthening the overall configuration to tune the characteristics of the resultant system. Further, more than two elements could be provided.

In FIG. **13**C, a fractal antenna **810** is formed on the outer surface **820**A of a filled substrate **820**, which may be a ferrite core. The resultant com-cyl antenna appears to exhibit self-proximity such that desired shifts in resonant frequency are produced. The geometry of the core **820**, e.g., the extent of curvature (e.g., radius in this embodiment) relative to the size of antenna **810** may be used to determine frequency shifts.

In FIG. **14**A, an antenna or resonator system is shown in which the non-driven fractal antenna **810**' is not connected to the preferably coaxial feedline **50**. The ground shield portion of feedline **50** is coupled to the groundplane conductive element **800**, but is not otherwise connected to a system ground. Of course fractal antenna **810**' could be angularly rotated relative to driven antenna **810**, it could be a different configuration than antenna **810** including having a different iteration N, and indeed could incorporate other features disclosed herein (e.g., a cut).

FIG. **14**B demonstrates that the driven antenna **810** may be coupled to the feedline **50** at any point **815**', and not necessarily at an end point **815** as was shown in FIG. **14**A.

In the embodiment of FIG. **14**C, a second ground plane element **800**' is disposed adjacent at least a portion of the system comprising driven antenna **810**, passive antenna **810**', and the underlying conductive planar element **800**. The presence, location, geometry, and distance associated with second ground plane element **800**' from the underlying elements **810, 810', 800** permit tuning characteristics of the overall antenna or resonator system. In the multi-element sandwich-like configuration shown, the ground shield of conductor **50** is connected to a system ground but not to either ground plane **800** or **800**'. Of course more than three elements could be used to form a tunable system according to the present invention.

FIG. **14**D shows a single fractal antenna spaced apart from an underlying ground plane **800** a distance Δ, in which a region of antenna **800** is cutaway to increase resonance. In FIG. **14**D, for example, L1 denotes a cutline, denoting that portions of antenna **810** above (in the Figure drawn) L1 are cutaway and removed. So doing will increase the frequencies of resonance associated with the remaining antenna or resonator system. On the other hand, if portions of antenna **810** above cutline L2 are cutaway and removed, still higher resonances will result. Selectively cutting or etching away portions of antenna **810** permit tuning characteristics of the remaining system.

FIG. **15** depicts an embodiment somewhat similar to what has been described with respect to FIG. **8**B or FIG. **8**C. Once again unit **500** is a handheld transceiver, and includes fractal antennas **510**A, **510**B–**510**B', **510**C. Antennas **510**B–**510**B' are similar to what has been described with respect to FIGS. **12**A–**12**B. Antennas **510**B–**510**B' are fractal antennas, not necessarily MI-2 configuration as shown, and are spaced-apart a distance a and, in FIG. **13**, are rotationally displaced. Collectively, the spaced-apart distance and relative rotational displacement permits tuning the characteristics of the driven antenna, here antenna **510**B. In FIG. **14**, antenna **510**A is drawn with phantom lines to better distinguish it from spaced-apart antenna **510**B. Of course passive conductor **510**B' could instead be a solid conductor such as described with respect to FIG. **11**A. Such conductor may be implemented by spraying the inner surface of the housing for unit **500** adjacent antenna **510**B with conductive paint.

In FIG. **15**, antenna **510**C is similar to what has been described with respect to FIG. **13**A, in that a cut **830** is made in the antenna, for tuning purposes. Although antenna **510**A is shown similar to what was shown in FIG. **8**B, antenna **510**A could, if desired, be formed on a curved substrate similar to FIG. **13**B or **13**C. While FIGS. **13**A-**1**–**13**C and **14**C show at least two different techniques for tuning antennas according to the present invention, it will be understood that a common technique could instead be used. By that it is meant that any or all of antennas **510**A, **510**B–**510**B', **510**C could include a cut, or be spaced-apart a controllable distance Δ, or be rotatable relative to a spaced-apart conductor.

As described with respect to FIG. **8**B, an electronic circuit **615** may be coupled by cables **50**A, **50**B, **50**C to the antennas, and samples incoming signals to discern which fractal antenna, e.g., **510**A, **510**B–**510**B', **510**C is presently most optimally aligned with the transmitting station, perhaps a unit **600** or **650** or **660** as shown in FIG. **8**A. This determination may be made by examining signal strength from each of the antennas. An electronic circuit **620** then selects the presently best oriented antenna, and couples such antenna to the input of the receiver and output of the transmitter portion, collectively **630**, of unit **500**. It is understood that the selection of the best antenna is dynamic and can change as, for example, a user of **500** perhaps walks about holding the unit, or the transmitting source moves, or due to other changing conditions. In a cellular or a wireless telephone application, the result is more reliable communication, with the advantage that the fractal antennas can be sufficiently small-sized as to fit totally within the casing of unit **500**. Further, if a flexible substrate is used, the antennas may be wrapped about portions of the internal casing, as shown.

An additional advantage of the embodiment of FIG. **8**B is that the user of unit **500** may be physically distanced from the antennas by a greater distance that if a conventional external whip antenna were used. Although medical evidence attempting to link cancer with exposure to electromagnetic radiation from handheld transceivers is still inconclusive, the embodiment of FIG. **8**B appears to minimize any such risk.

Modifications and variations may be made to the disclosed embodiments without departing from the subject and spirit of the invention as defined by the following claims. While common fractal families include Koch, Minkowski, Julia, diffusion limited aggregates, fractal trees, Mandelbrot, the present invention may be practiced with other fractals as well.

What is claimed is:

1. An antenna system comprising:

a first fractal antenna including a first element having a portion that includes:

at least a first motif selected from a family consisting of (i) Koch, (ii) Minkowski, (iii) Cantor, (iv) torn square, (v) Mandelbrot. (vi) Caley tree, (vii) monkey's swing, (viii) Sierpinski gasket, and (ix) Julia;

and a first replication of said first motif and a second replication of said first motif such that a point chosen on a geometric figure represented by said first motif will result in a corresponding point on said first replication and on said second replication of said first motif; wherein there exists at least one non-straight line locus connecting each said point;

wherein a replication of said first motif is a change selected from a group consisting of (a) a rotation and change of scale of said first motif, (b) a linear displacement translation and a change of scale of said first motif, and (c) a rotation and a linear displacement translation and a change of scale of said first motif, and wherein said first fractal antenna is characterized in space by at least one resonant frequency and by a bandwidth; and

a conductive element, spaced-apart from said first fractal antenna by a distance Δ chosen to vary at least one characteristic of said antenna system, at a desired frequency c/λ, where c is velocity of light, selected from a group consisting of (i) said resonant frequency, and (ii) said bandwidth.

2. The antenna system of claim **1**, wherein said distance Δ has at least one characteristic selected from a group consisting of (a) said first fractal antenna and said conductive element are spaced-apart in different planes by a distance $\Delta \leqq 0.05\lambda$ where λ is wavelength at a resonant frequency of said antenna system in free space, (b) said first fractal antenna and said conductive element are spaced-apart in a common plane by said distance Δ, and (c) at least one of said first fractal antenna and said conductive element is non-planar and said distance Δ defines a closest distance separating said first fractal antenna from said conductive element.

3. The antenna system of claim **1**, wherein said conductive element is selected from a group consisting of (a) a planar conductor, (b) a second fractal antenna defined by an iteration identical to said first fractal antenna, (c) a second fractal antenna defined by an iteration N identical to said first fractal antenna but having a different configuration, (d) a second fractal antenna defined by at least a third-order iteration, and (e) a second fractal antenna defined by at least a third-order iteration and having a configuration similar to a configuration of said first fractal antenna.

4. The antenna system of claim **1**, wherein said conductive element is a second fractal antenna that includes a second element having a portion that includes at least a second motif and a first replication of said second motif and a second replication of said second motif such that a point chosen on a geometric figure represented by said second motif will result in a corresponding point on said first replication and on said second replication of said second motif; wherein there exists at least one locus connecting each said point that does not define a straight line; and

wherein a replication of said second motif is a change selected from a group consisting of (a) a rotation and change of scale of said second motif, (b) a linear displacement translation and a change of scale of said second motif, and (c) a rotation and a linear displacement translation and a change of scale of said second motif;

wherein one of said first fractal antenna and said second fractal antenna is rotatable through a relative angle θ therebetween to vary at least one characteristic of said antenna system.

5. The antenna system of claim **1**, wherein said first fractal antenna includes a region that defines a cut such that at least one said resonant frequency of said antenna system is varied by said cut.

6. The antenna system of claim **1**, wherein said first fractal antenna is formed non-planarly to vary at least one said resonant frequency of said antenna system.

7. The antenna system of claim **1**, wherein said antenna system includes a ferrite core about which said first fractal antenna is formed.

8. The antenna system of claim **1**, further including a feedline having a center conductor and a ground conductor; wherein:

a portion of said first fractal antenna is connected to said center conductor of said feedline, and a portion of said conductive element is connected to said ground conductor of said feedline.

9. The antenna system of claim **4**, further including a feedline having a center conductor and a ground conductor; wherein:

a portion of said first fractal antenna is connected to said center conductor of said feedline, and said ground conductor of said feedline is connected to at least one region of said system defined by (a) a portion of said second fractal antenna, (b) a system ground, and (c) said conductive element and a system ground.

10. The antenna system of claim **1**, further including:

a ground plane member;

said ground member disposed sufficiently close to said first fractal antenna so as to vary a characteristic of said antenna system.

11. The antenna system of claim **1**, wherein said first motif has x-axis, y-axis coordinates for a next iteration N+1 defined by $x_{N+1}=f(x_N, y_N)$ and $y_{N+1}=g(x_N, y_N)$, where x_N, y_N are coordinates for iteration N, and where f(x,y) and g(x,y) are functions defining said first motif.

12. The antenna system of claim **1**, wherein said first fractal antenna has a perimeter compression parameter (PC) defined by:

$$PC = \frac{\text{full-sized antenna element length}}{\text{fractal-reduced antenna element length}}$$

where:

$$PC = A \cdot \log[N(D+C)]$$

in which A and C are constant coefficients for a given said first motif, N is an iteration number, and D is a fractal dimension given by log(L)/log(r), where L and r are one-dimensional antenna element lengths before and after fractalization, respectively.

13. A tunable fractal antenna system coupleable to a transceiver unit, the antenna comprising:

a first fractal antenna including:

a first element having a portion that includes at least a first motif selected from a family consisting of (i) Koch, (ii) Minkowski, (iii) Cantor, (iv) torn square, (v) Mandelbrot, (vi) Caley tree, (vii) monkey's swing, (viii) Sierpinski gasket, and (ix) Julia;

and a first replication of said first motif and a second replication of said first motif such that a point chosen on a geometric figure represented by said first motif will result in a corresponding point on said first replication and on said second replication of said first motif; wherein there exists at least one non-straight line locus connecting each said point;

wherein a replication of said first motif is a change selected from a group consisting of (a) a rotation and change of scale of said first motif, (b) a linear displacement translation and a change of scale of said first motif, and (c) a rotation and a linear displacement translation and a change of scale of said first motif; and

a conductive element, spaced-apart from said first fractal antenna by a distance Δ chosen to vary at least one characteristic of said antenna system, at a desired frequency c/λ, where c is velocity of light, selected from the group consisting of (i) resonant frequency of said antenna system, and (ii) bandwidth of said antenna system.

14. The tunable fractal antenna system of claim **13**, wherein said conductive element is selected from a group consisting of (a) a planar conductor, (b) a second fractal antenna defined by an iteration identical to said first fractal antenna, (c) a second fractal antenna defined by an iteration identical to said first fractal antenna but having a different configuration, (d) a second fractal antenna defined by at least a third-order iteration, (e) a second fractal antenna defined by at least a third-order iteration and having a configuration similar to said first fractal antenna, (f) a second fractal antenna angularly rotated relative to said first fractal antenna, and (g) a second fractal antenna having a portion that defines a cut.

15. The tunable fractal antenna system of claim **13**, further including a feedline;

wherein said antenna system is tunable by varying at least one parameter selected from a group consisting of (a) said distance Δ, (b) relative rotation between said first fractal antenna and said conductive element, (c) location at which a center lead of said feedline is coupled to said first fractal antenna, (d) location of a cut in said first fractal antenna, and (e) size of a cut in said first fractal antenna.

16. The tunable fractal antenna system of claim **13**, wherein said distance Δ has at least one characteristic selected from a group consisting of (a) said first fractal antenna and said conductive element are spaced-apart in different planes by a distance $\Delta \leqq 0.05\lambda$ where λ is wavelength at a resonant frequency of said antenna system in free space, (b) said first fractal antenna and said conductive element are spaced-apart in a common plane by said distance Δ, and (c) at least one of said first fractal antenna and said conductive element is non-planar and said distance Δ defines a closest distance separating said first fractal antenna from said conductive element.

17. A method of tuning a fractal resonator system, comprising:

disposing a first fractal member a distance Δ from a conductive element;

said first fractal member including a first element having a portion that includes at least a first motif selected from a family consisting of (i) Koch, (ii) Minkowski, (iii) Cantor, (iv) torn square, (v) Mandelbrot, (vi) Caley tree, (vii) monkey's swing, (viii) Sierpinski gasket, and (ix) Julia; and a first replication of said first motif and a second replication of said first motif such that a point chosen on a geometric figure represented by said first motif will result in a corresponding point on said first replication and on said second replication of said first motif; wherein there exists at least one non-straight line locus connecting each said point;

wherein a replication of said first motif is a change selected from a group consisting of (a) a rotation and change of scale of said first motif, (b) a linear displacement translation and a change of scale of said first motif, and (c) a rotation and a linear displacement translation and a change of scale of said first motif; and

tuning said fractal resonator system by modifying at least one parameter selected from a group consisting of (a) a magnitude of said distance Δ, (b) an angular orientation between said first fractal member and said conductive element, (c) extent of curvature associated with at least one of said first fractal member and said conductive element, (d) size of at least one of said first fractal member and said conductive element, (e) shape of at least one of said first fractal member and said conductive element, (f) location of said conductive element relative to said first fractal member, (g) existence of a cut defined in at least one of said first fractal member and said conductive element, and (g) location of feedline coupling to at least one of said first fractal member and said conductive element.

18. The method of claim **17**, wherein said system is an antenna system.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,104,349

DATED : August 15, 2000

INVENTOR(S) : Cohen

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 24, line 51, after "FIGS.", delete"13A-1-13C" and insert therefor – 13A–13C –.

Signed and Sealed this

Twenty-ninth Day of May, 2001

Attest:

Nicholas P. Godici

NICHOLAS P. GODICI

Attesting Officer

Acting Director of the United States Patent and Trademark Office

US006127977A

United States Patent [19]

Cohen

[11] **Patent Number: 6,127,977**

[45] **Date of Patent: Oct. 3, 2000**

[54] **MICROSTRIP PATCH ANTENNA WITH FRACTAL STRUCTURE**

[76] Inventor: **Nathan Cohen**, 2 Ledgewood Pl., Belmont, Mass. 02178

[21] Appl. No.: **08/965,914**

[22] Filed: **Nov. 7, 1997**

Related U.S. Application Data

[60] Provisional application No. 60/030,633, Nov. 8, 1996.

[51] **Int. Cl.**^{7} **H01Q 1/38**

[52] **U.S. Cl.** .... **343/700 MS**; 343/792.5

[58] **Field of Search** .... 343/700 MS File, 343/792.5; H01Q 1/38

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,652,889	3/1987	Bizouard et al.	343/700 MS
5,006,858	4/1991	Shirosaka	343/700
5,111,211	5/1992	Dahlberg	343/700 MS
5,313,216	5/1994	Wang et al.	343/700 MS
5,453,752	9/1995	Wang et al.	343/700 MS

Primary Examiner—Michael C. Wimer
Attorney, Agent, or Firm—Flehr Hohbach Test Albritton & Herbert LLP

[57] **ABSTRACT**

A microstrip patch antenna having reduced size is implementing by providing a substrate having on one surface a conductive fractal pattern, and having on the other surface a conductive pattern that may (but need not) also be a fractal pattern. The fractal pattern is of order N≧1, and if fractal patterns are formed on each substrate surface, the fractal family and fractal iteration number may be different. So fractalizing at least one conductive surface permits reduction of substrate dimension may be reduced to one-eighth wavelength.

26 Claims, 1 Drawing Sheet

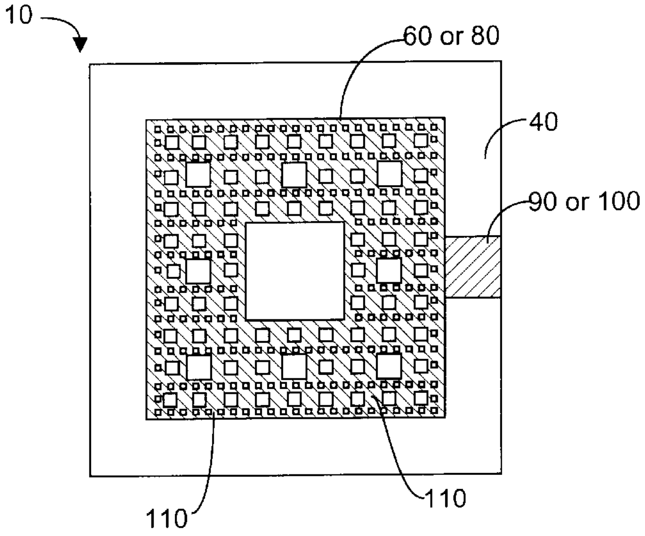

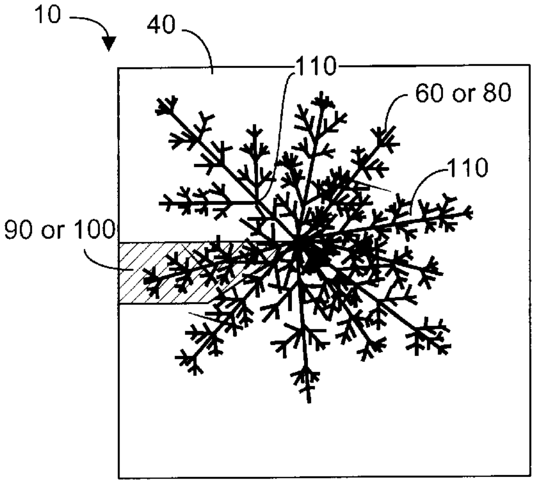

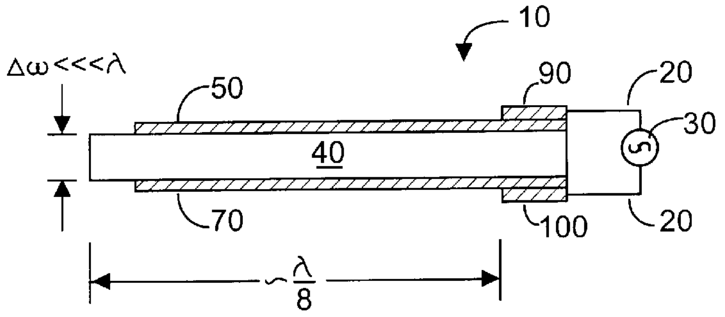
10
Δω<<<λ
50
90
20
30
40
70
100
20
~λ/8

FIG. 1

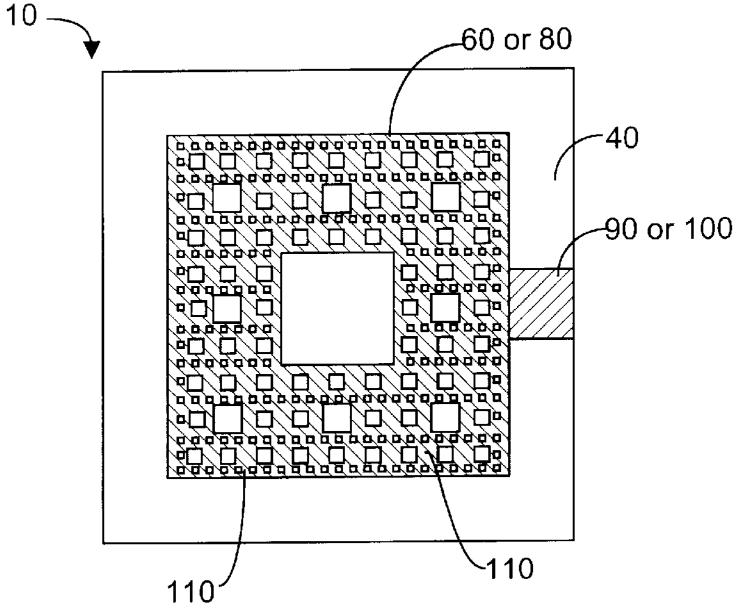
10
60 or 80
40
90 or 100
110
110

FIG. 2

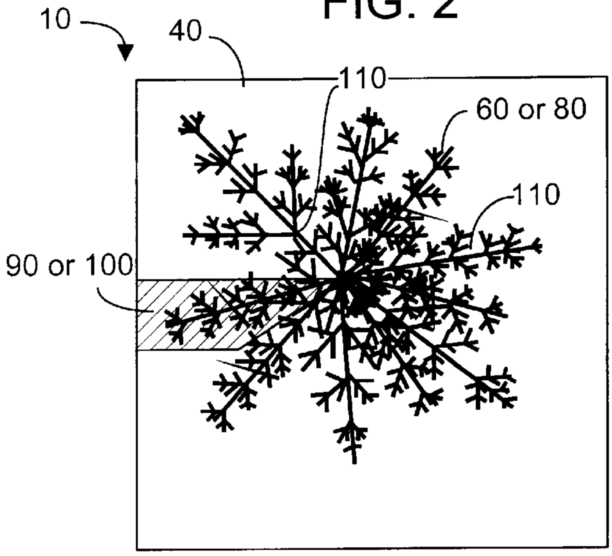
10
40
110
60 or 80
110
90 or 100

FIG. 3

MICROSTRIP PATCH ANTENNA WITH FRACTAL STRUCTURE

RELATION TO PREVIOUSLY FILED PATENT APPLICATIONS

This application claims priority from applicant's U.S. provisional patent application No. 60/030,633 filed Nov. 8, 1996 entitled "MICROSTRIP PATCH ANTENNAE INCORPORATING 1 AND/OR 2 SIDES OF FRACTAL STRUCTURE ELEMENTS".

Applicant also refers to and incorporates herein by reference applicant's U.S. application Ser. No. 08/649,825 filed May 17, 1996 entitled "FRACTAL ANTENNA GROUND COUNTERPOISE, GROUND PLANES, AND LOADING ELEMENTS", now abandoned, applicant's patent application Ser. No. 08,609,514 filed Mar. 1, 1996 entitled "TUNING FRACTAL ANTENNAS AND FRACTAL RESONATORS", now abandoned, and applicant's patent application Ser. No. 08/512,954 filed Aug. 9, 1995 entitled "FRACTAL ANTENNAS AND FRACTAL RESONATORS", now abandoned.

FIELD OF THE INVENTION

The present invention relates to microstrip patch antennas and more specifically to providing such antennas with fractal structure elements.

BACKGROUND OF THE INVENTION

Antenna are used to radiate and/or receive typically electromagnetic signals, preferably with antenna gain, directivity, and efficiency. Practical antenna design traditionally involves trade-offs between various parameters, including antenna gain, size, efficiency, and bandwidth.

Antenna design has historically been dominated by Euclidean geometry. In such designs, the closed antenna area is directly proportional to the antenna perimeter. For example, if one doubles the length of an Euclidean square (or "quad") antenna, the enclosed area of the antenna quadruples. Classical antenna design has dealt with planes, circles, triangles, squares, ellipses, rectangles, hemispheres, paraboloids, and the like, (as well as lines).

With respect to antennas, prior art design philosophy has been to pick a Euclidean geometric construction, e.g., a quad, and to explore its radiation characteristics, especially with emphasis on frequency resonance and power patterns. The unfortunate result is that antenna design has far too long concentrated on the ease of antenna construction, rather than on the underlying electromagnetics.

Many prior art antennas are based upon closed-loop or island shapes. Experience has long demonstrated that small sized antennas, including loops, do not work well, one reason being that radiation resistance ("R") decreases sharply when the antenna size is shortened. A small sized loop, or even a short dipole, will exhibit a radiation pattern of ½λ and ¼λ, respectively, if the radiation resistance R is not swamped by substantially larger ohmic ("O") losses. Ohmic losses can be minimized using impedance matching networks, which can be expensive and difficult to use. But although even impedance matched small loop antennas can exhibit 50% to 85% efficiencies, their bandwidth is inherently narrow, with very high Q, e.g., Q>50. As used herein, Q is defined as (transmitted or received frequency)/(3 dB bandwidth).

Applicant's above-referenced co-pending patent applications depict examples of fractal geometry, which geometry may be grouped into random fractals, which are also termed chaotic or Brownian fractals and include a random noise components, or deterministic fractals.

In deterministic fractal geometry, a self-similar structure results from the repetition of a design or motif (or "generator"), on a series of different size scales. One well known treatise in this field is *Fractals, Endlessly Repeated Geometrical Figures*, by Hans Lauwerier, Princeton University Press (1991), which treatise applicant refers to and incorporates herein by reference. Lauwerier notes that in its replication, the motif may be rotated, translated, scaled in dimension, or a combination of any of these characteristics. Thus, as used herein, second order of iteration or N=2 means the fundamental motif has been replicated, after rotation, translation, scaling (or a combination of each) into the first order iteration pattern. A higher order, e.g., N=3, iteration means a third fractal pattern has been generated by including yet another rotation, translation, and/or scaling of the first order motif.

Unintentionally, first order fractals have been used to distort the shape of dipole and vertical antennas to increase gain, the shapes being defined as a Brownian-type of chaotic fractals. See F. Landstorfer and R. Sacher, *Optimisation of Wire Antennas*, J. Wiley, New York (1985).

So-called microstrip patch antennas have traditionally been fabricated as two spaced-apart metal surfaces separated by a small width dielectric. The sides are dimensioned typically one-quarter wavelength or one-half wavelength at the frequency of interest. One surface is typically a simple euclidean structure such as a circle, a square, while the other side is a ground plane.

Attempting to reduce the physical size of such an antenna for a given frequency typically results in a poor feedpoint match (e.g., to coaxial or other feed cable), poor radiation bandwidth, among other difficulties.

Prior art antenna design does not attempt to exploit multiple scale self-similarity of real fractals. This is hardly surprising in view of the accepted conventional wisdom that because such antennas would be anti-resonators, and/or if suitably shrunken would exhibit so small a radiation resistance R, that the substantially higher ohmic losses O would result in too low an antenna efficiency for any practical use. Further, it is probably not possible to mathematically predict such an antenna design, and high order iteration fractal antennas would be increasingly difficult to fabricate and erect, in practice.

Thus, the use of fractals, especially higher order fractals, in fabricating microstrip patch antennas has not been investigated in the prior art.

Applicant's above-noted FRACTAL ANTENNA AND FRACTAL RESONATORS patent application provided a design methodology to produce smaller-scale antennas that exhibit at least as much gain, directivity, and efficiency as larger Euclidean counterparts. Such design approach should exploit the multiple scale self-similarity of real fractals, including N≧2 iteration order fractals. Further, said application disclosed a non-Euclidean resonator whose presence in a resonating configuration can create frequencies of resonance beyond those normally presented in series and/or parallel LC configurations. Applicant's above-noted TUNING FRACTAL ANTENNAS AND FRACTAL RESONATORS patent application provided devices and methods for tuning and/or adjusting such antennas and resonators. Said application further disclosed the use of non-Euclidean resonators whose presence in a resonating configuration could create frequencies of resonance beyond those normally presented in series and/or parallel LC configurations.

However, such antenna design approaches and tuning approaches should also be useable with microstrip patch antennas and elements for such antennas. Thus, there is a need for a method by which microstrip patch antennas could be made smaller without sacrificing antenna bandwidth, while preserving good feedpoint impedance matching, and while maintaining acceptable gain and frequency characteristics.

The present invention provides such microstrip patch antennas, and elements for such antennas.

SUMMARY OF THE INVENTION

The present invention provides a microstrip patch antenna comprising spaced-apart first and second conductive surfaces separated by a dielectric material. The dielectric material thickness preferably is substantially less than one wavelength for the frequency of interest.

At least one of the surfaces is fabricated to define a fractal pattern of first or higher iteration order. Overall dimensions of the surfaces may be reduced below the one-quarter to one-half wavelength commonly found in the prior art.

Radio frequency feedline coupling to the microstrip patch antenna may be made at a location on the antenna pattern structure, or through a conductive feedtab strip that may be fabricated along with the conductive pattern on one or both surfaces of the antenna. The resultant antenna may be sized smaller than a non-fractal counterpart (e.g., approximately one-eighth wavelength provides good performance at about 900 MHz.) while preserving good, preferably 50Ω, feedpoint impedance. Further bandwidth can actually be increased, and resonant frequency lowered.

Other features and advantages of the invention will appear from the following description in which the preferred embodiments have been set forth in detail, in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a sideview of a microstrip patch antenna with at least one fractal element, according to the present invention;

FIG. **2** is a top plan view of an exemplary fractal element (a Sierpinski square gasket, including an optional feedtab, according to the present invention;

FIG. **3** is a top plan view of an exemplary alternative fractal element (a diffusion limited aggregate), including an optional feed pad, according to the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

In overview, the present invention provides a microstrip patch antenna with at least one element whose shape, at least is part, is substantially a fractal of iteration order $N \geqq 1$. The resultant antenna is smaller than its Euclidean counterpart, provides close to 50Ω termination impedance, exhibits acceptable gain, increased bandwidth, and decreased resonant frequency than its Euclidean counterpart.

In contrast to Euclidean geometric antenna design, a fractal antenna ground counterpoise according to the present invention has a perimeter that is not directly proportional to area. For a given perimeter dimension, the enclosed area of a multi-iteration fractal area will always be at least as small as any Euclidean area.

Using fractal geometry, the ground element has a self-similar structure resulting from the repetition of a design or motif (or "generator"), which motif is replicated using rotation, translation, and/or scaling (or any combination thereof). The fractal portion of the element has x-axis, y-axis coordinates for a next iteration N+1 defined by $x_{N+1}=f(x_N, yb_N)$ and $y_{N+1}=g(x_N, y_N)$, where x_N, y_N are coordinates of a preceding iteration, and where f(x,y) and g(x,y) are functions defining the fractal motif and behavior.

For example, fractals of the Julia set may be represented by the form:

$$x_{N+1}=x_N^2-y_N^2+a$$

$$y_{N+1}=2x_N \cdot y_N=b$$

In complex notation, the above may be represented as:

$$Z_{N+1}=Z_N^2+C$$

Although it is apparent that fractals can comprise a wide variety of forms for functions f(x,y) and g(x,y), it is the iterative nature and the direct relation between structure or morphology on different size scales that uniquely distinguish f(x,y) and g(x,y) from non-fractal forms. Many references including the Lauwerier treatise set forth equations appropriate for f(x,y) and g(x,y).

Iteration (N) is defined as the application of a fractal motif over one size scale. Thus, the repetition of a single size scale of a motif is not a fractal as that term is used herein. Multi-fractals may of course be implemented, in which a motif is changed for different iterations, but eventually at least one motif is repeated in another iteration.

Referring now to FIG. **1**, a microstrip patch antenna **10** according to the present invention is shown coupled by coaxial or other cable (or equivalent) **20** to a source of radio frequency **30**. Antenna **10** comprises a substrate **40** whose top-to-bottom thickness is preferably substantially less than one wavelength at the frequency of interest, e.g., the radio frequency or band of radio frequencies coupled by cable **20** to antenna **10**. Preferably the effective dimension of substrate is one-eighth wavelength at such frequency.

On its first surface, substrate **40** is initially covered by a conductive layer of material **50** that is etched away or otherwise removed in areas other than the desired fractal pattern **(60)** design, to expose the substrate. The remaining conductive trace portion defines a fractal element, according to the present invention.

Similarly on its second surface, substrate **40** is initially covered by a conductive layer of material **70** that is selectively removed so as to leave a desired pattern **(80)** that may also be a fractal pattern, according to the present invention. Alternatively, conductive material defining the desired patterns **60**, **80** could be deposited upon substrate **40**, rather than beginning fabrication with a substrate clad or otherwise having conductive surfaces, portions of which are removed.

Preferably feedtabs **90** and **100** are coupled, respectively, to edge regions of the first and second surfaces of substrate **40** to facilitate electrical radio frequency coupling between cable **20** and patterns **60** and/or **80**. These feedtabs preferably are etched using the same conductive material originally found on the upper or lower surfaces of substrate **40**, or may otherwise be formed using techniques known to those skilled in the relevant art. If patterns **60** and **80** are deposited rather than etched, then feedtabs **90**, **100** may be deposited at the same fabrication step.

Substrate **40** is a non-conductive material, and by way of example may be a silicon wafer, a rigid or a flexible plastic-like material, perhaps Mylar™ material, or the non-conductive portion of a printed circuit board, paper, epoxy, among other materials. The original conductive material on the first and/or second surfaces may be deposited doped polysilicon for a semiconductor substrate **40**, or copper (or other conductor) for a printed circuit board substrate.

FIG. **2** is a plan view of one surface of antenna **10** (it matters not which), and depicts a first iteration fractal conductive pattern, although a fractal pattern with higher than first iteration could instead be used. The pattern shown in FIG. **2** is often referred to as a Siepinski (square) gasket pattern. A margin is shown in FIG. **2** between the outer perimeter of the pattern and the edge of the substrate; however no such margin is required. Although FIG. **2** shows inclusion of feedtab **90** or **100**, radio frequency feed may be made elsewhere on the surface, for example at any point **110**.

If the fractal pattern of FIG. **2** represents one surface of antenna **10**, the opposite surface need not define a fractal pattern, but may in fact do so. For example, one surface may define a fractal pattern and the opposite surface may be entirely conductive, or may define on the substrate a conductive circle, etc. If the pattern on the opposite surface is also a fractal, there is no requirement that it be the same iteration fractal as is defined on the first surface, or that it be the same fractal type. While common fractal families include Koch, Minkowski, Julia, diffusion limited aggregates, fractal trees, Mandelbrot, microstrip patch antennas with fractal element(s) according to the present invention may be implemented with other fractals as well.

FIG. **3** depicts a pattern **60** or **80** in which a different fractal pattern is defined, a so-called diffusion limited aggregate pattern. It is understood, however, that according to the present invention, a great variety of fractal patterns of first or higher iteration may be defined on the first and/or second surface of antenna **10**. In FIG. **3**, while a feedtab **90** or **100** is shown, it is again understood that radio frequency feed may be made essentially anywhere on the fractal pattern, e.g., at a point **110**.

In one embodiment, applicant fabricated an antenna **10** having sides dimensioned to about one-eighth wavelength for a frequency of about 900 MHz. Those skilled in the art will readily appreciate that a microstrip patch antenna dimensioned to one-eighth wavelength is substantially smaller than prior art non-fractal microstrip patch antennas, in which dimensions are one-quarter or one-half wavelength in size. At 900 MHz, bandwidth was about 5% to about 8% of nominal frequency. Gain and matching impedance were acceptable, and indeed substantially 50Ω impedance is realized without the need for impedance transforming devices.

Modifications and variations may be made to the disclosed embodiments without departing from the subject and spirit of the invention as defined by the following claims. It will be appreciated, for example, that the present invention may be implemented and adjusted and used in ways described in any of applicant's referenced co-pending applications.

What is claimed is:

1. A microstrip patch antenna including:

a substrate having spaced-apart first and second surfaces, said substrate having a thickness substantially smaller than a wavelength at a frequency to be coupled to said antenna;

a conductive pattern defining a fractal of iteration order N disposed on the first surface, wherein said fractal is defined as a superposition over at least N=1 interations of a motiff, an iteration being placement of said motif upon a base figure through at least one positioning selected from a group consisting of (i) rotation, (ii) stretching, and (iii) translation;

wherein said motif is selected from a group consisting of (i) Koch, (ii) Minkowski, (iii) Cantor, (iv) torn square, (v) Mandelbrot, (vi) Caley tree, (vii) monkey's swing, (viii) Sierpinski gasket, and (ix) Julia; and

a conductive pattern disposed on the second surface.

2. The antenna of claim **1**, wherein said pattern on said second surface defines a fractal.

3. The antenna of claim **1**, wherein said motif has x-axis, y-axis coordinates for a next iteration N+1 defined by $x_{N+1}=f(x_N, y_N)$ and $y_{N+1}=g(x_N, y_N)$, where x_N, y_N are coordinates for iteration N, and where f(x,y) and g(x,y) are functions defining said motif.

4. The antenna of claim **1**, wherein said antenna has a perimeter compression parameter (PC) defined by:

$$PC = \frac{\text{full-sized antenna element length}}{\text{fractal-reduced antenna element length}}$$

where:

$$PC = A \cdot \log [N(D+C)]$$

in which A and C are constant coefficients for a given said motif, N is an iteration number, and D is a fractal dimension given by log(L)/log(r), where L and r are one-dimensional antenna element lengths before and after fractalization, respectively.

5. The antenna of claim **1**, in which said antenna is fabricated in a manner selected from the group consisting of (i) forming upon an insulator substrate a conductive layer defining said fractal, (ii) forming upon a flexible insulator substrate a conductive layer defining said fractal; (iii) forming upon a semiconductor substrate a layer of conductive material to define said fractal, and (iv) removing from a substrate having a surface covered with conductive material a portion of said conductive material to form said fractal.

6. The antenna of claim **1**, wherein said substrate is sized to less than one-quarter wavelength at a frequency of radio frequency signals to be coupled to said antenna.

7. The antenna of claim **1**, wherein said substrate is sized to approximately one-eighth wavelength at a frequency of radio frequency signals to be coupled to said antenna.

8. The antenna of claim **7**, wherein said frequency is approximately 900 MHz.

9. A microstrip patch antenna including:

a substrate having spaced-apart first and second surfaces, said substrate having a thickness substantially smaller than a wavelength at a frequency to be coupled to said antenna;

a conductive pattern defining a fractal of iteration order N disposed on the first surface, where said fractal is defined as a superposition over at least N=1 interations of a motiff, an iteration being placement of said motif upon a base figure through at least one positioning selected from a group consisting of (i) rotation, (ii) stretching, and (iii) translation;

wherein said antenna has a perimeter compression parameter (PC) defined by:

$$PC = \frac{\text{full-sized antenna element length}}{\text{fractal-reduced antenna element length}}$$

where:

$$PC = A \cdot \log [N(D+C)]$$

in which A and C are constant coefficients for a given said motif, N is an iteration number, and D is a fractal dimension given by log(L)/log(r), where L and r are one-dimensional antenna element lengths before and after fractalization respectively; and

a conductive pattern disposed on the second surface.

10. The antenna of claim **9**, wherein said motif is selected from a family consisting of (i) Koch, (ii) Minkowski, (iii) Cantor, (iv) torn square, (v) Mandelbrot, (vi) Caley tree, (vii) monkey's swing, (viii) Sierpinski gasket, and (ix) Julia.

11. A method of fabricating a microstrip patch antenna, the method including the following steps:

(a) providing a substrate having spaced-apart first and second surfaces and having a substrate thickness substantially smaller than a wavelength at a frequency to be coupled to said antenna;

(b) disposing on the first surface of said substrate a conductive pattern defining a fractal of iteration order N formed; and

(c) disposing on the second surface of said substrate a conductive pattern;

wherein said motif is selected from a family consisting of (i) Koch, (ii) Minkowski, (iii) Cantor, (iv) torn square, (v) Mandelbrot, (vi) Caley tree, (vii) monkey's swing, (viii) Sierpinski gasket, and (ix) Julia.

12. The method of claim **11**, wherein at step (c) said conductive pattern is formed so as to define a fractal.

13. The method of claim **11**, wherein at step (b), said fractal on said first surface is defined as a superposition over at least N=1 iterations of a motif, an iteration being placement of said motif upon a base figure through at least one positioning selected from the group consisting of (i) rotation, (ii) stretching, and (iii) translation.

14. The method of claim **11**, wherein said motif has x-axis, y-axis coordinates for a next iteration N+1 defined by $x_{N+1}=f(x_N, y_N)$ and $y_{N+1}=g(x_N, y_N)$, where x_N, y_N are coordinates for iteration N, and where f(x,y) and g(x,y) are functions defining said motif.

15. The antenna of claim **9**, wherein said antenna is fabricated in a manner selected from the group consisting of (i) forming upon an insulator substrate a conductive layer defining said fractal, (ii) forming upon a flexible insulator substrate a conductive layer defining said fractal; (iii) forming upon a semiconductor substrate a layer of conductive material to define said fractal, and (iv) removing from a substrate having a surface covered with conductive material a portion of said conductive material to form said fractal.

16. The method of claim **11**, wherein said antenna has a perimeter compression parameter (PC) defined by:

$$PC = \frac{\text{full-sized antenna element length}}{\text{fractal-reduced antenna element length}}$$

where:

$$PC=A\cdot\log\,[N(D+C)]$$

in which A and C are constant coefficients for a given said motif, N is an iteration number, and D is a fractal dimension given by log(L)/log(r), where L and r are one-dimensional antenna element lengths before and after fractalization, respectively.

17. The method of claim **11**, in which said antenna is fabricated in a manner selected from the group consisting of (i) forming upon an insulator substrate a conductive layer defining said fractal, (ii) forming upon a flexible insulator substrate a conductive layer defining said fractal; (iii) forming upon a semiconductor substrate a layer of conductive material to define said fractal, and (iv) providing a substrate having a surface covered with conductive material, and removing a portion of said conductive material to form said fractal.

18. The method of claim **11**, wherein said substrate is sized to less than one-quarter wavelength at a frequency of radio frequency signals to be coupled to said antenna.

19. The method of claim **11**, wherein at step (a) said substrate is sized to approximately one-eighth wavelength at a frequency of radio frequency signals to be coupled to said antenna.

20. The method of claim **19**, wherein said frequency is approximately 900 MHz.

21. A method of fabricating a microstrip patch antenna, the method including the following steps:

(a) providing a substrate having spaced-apart first and second surfaces and having a substrate thickness substantially smaller than a wavelength at a frequency to be coupled to said antenna;

(b) disposing on the first surface of said substrate a conductive pattern defining a fractal of iteration order N formed; and

(c) disposing on the second surface of said substrate a conductive pattern;

wherein said antenna has a perimeter compression parameter (PC) defined by:

$$PC = \frac{\text{full-sized antenna element length}}{\text{fractal-reduced antenna element length}}$$

where:

$$PC=A\cdot\log\,[N(D+C)]$$

in which A and C are constant coefficients for a given said motif, N is an iteration number, and D is a fractal dimension given by log(L)/log(r), where L and r are one-dimensional antenna element lengths before and after fractalization, respectively.

22. The method of claim **21**, wherein at step (c) said conductive pattern is formed so as to define a fractal.

23. The method of claim **21**, wherein said antenna is fabricated in a manner selected from the group consisting of (i) forming upon an insulator substrate a conductive layer defining said fractal, (ii) forming upon a flexible insulator substrate a conductive layer defining said fractal; (iii) forming upon a semiconductor substrate a layer of conductive material to define said fractal, and (iv) providing a substrate having a surface covered with conductive material, and removing a portion of said conductive material to form said fractal.

24. The method of claim **21**, wherein said substrate is sized to less than one-quarter wavelength at a frequency of radio frequency signals to be coupled to said antenna.

25. The method of claim **21**, wherein at step (a) said substrate is sized to approximately one-eighth wavelength at a frequency of radio frequency signals to be coupled to said antenna.

26. The method of claim **25**, wherein said frequency is approximately 900 MHz.

* * * * *

US006140975A

United States Patent [19]

Cohen

[11] **Patent Number: 6,140,975**

[45] **Date of Patent: *Oct. 31, 2000**

[54] **FRACTAL ANTENNA GROUND COUNTERPOISE, GROUND PLANES, AND LOADING ELEMENTS**

[76] Inventor: **Nathan Cohen**, 2 Ledgewood Pl., Belmont, Mass. 02178

[*] Notice: This patent issued on a continued prosecution application filed under 37 CFR 1.53(d), and is subject to the twenty year patent term provisions of 35 U.S.C. 154(a)(2).

[21] Appl. No.: **08/967,375**

[22] Filed: **Nov. 7, 1997**

Related U.S. Application Data

[63] Continuation of application No. 08/649,825, May 17, 1996, abandoned, which is a continuation of application No. 08/609,514, Mar. 1, 1996, abandoned, which is a continuation-in-part of application No. 08/512,954, Aug. 9, 1995, abandoned.

[51] **Int. Cl.**^{7} **H01Q 1/48**

[52] **U.S. Cl.** **343/846**; 343/848; 343/899; 343/792.5

[58] **Field of Search** 343/752, 829, 343/830, 846, 848, 792.5, 899; H01Q 1/48

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,079,602	2/1963	Du Hamel et al.	343/792.5
4,318,109	3/1982	Weathers	343/806
4,358,769	11/1982	Tada et al.	343/741
5,164,738	11/1992	Walter et al.	343/792.5
5,608,413	3/1997	Macdonald	343/702

OTHER PUBLICATIONS

Pfeiffer, A., "The Pfeiffer Quad Antenna System", *QST*, pp. 28–30 (1994).

Primary Examiner—Michael C. Wimer
Attorney, Agent, or Firm—Flehr, Hohbach, Test, Albritton & Herbert LLP

[57] **ABSTRACT**

An antenna system includes a ground counterpoise or a top-hat located load assembly having at least one element whose physical shape is at least partially defined as a first or higher iteration deterministic fractal. The resultant ground counterpoise may rely upon an opening angle for performance, and produces a more compact antenna system relative to prior art non-Euclidean ground counterpoise elements. A vertical antenna system may be fabricated with fractal ground elements and a vertical element that may also be a fractal. Gain characteristics of antenna systems utilizing a fractal ground counterpoise are no worse than prior art, larger, systems, and exhibit improved vertical polarization characteristics, and a termination impedance of about 30Ω. A vertical antenna system preferably includes vertically spaced-apart fractal conductive and passive elements, and one or more fractal ground elements. The resultant antenna system may be tuned by rotating the vertical elements relative to each other, and/or by varying the spaced-apart distance therebetween. Fractalized ground counterpoise elements may be fabricated on a flexible printed circuit substrate, and/or placed within the support mount of a cellular telephone car antenna. A vertical antenna having a fractalized top-hat loading assembly advantageously reduces resonant frequency, size and area of the loading assembly, without substantial penalty in performance.

19 Claims, 24 Drawing Sheets

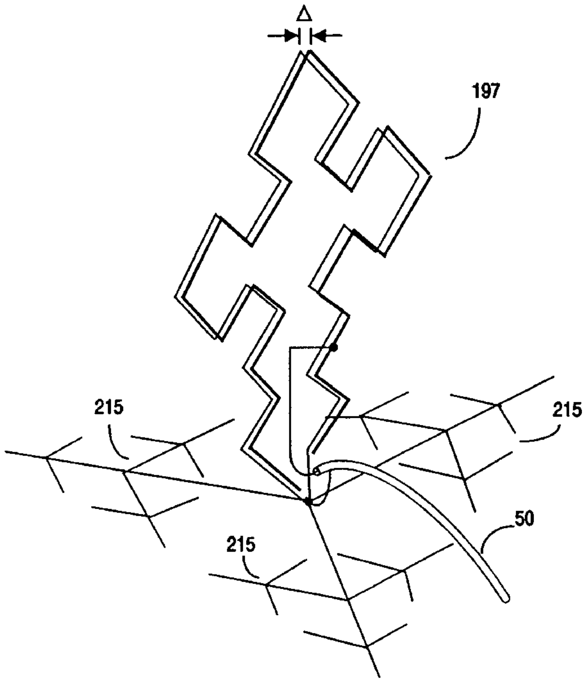

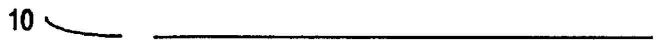

FIGURE 1A (PRIOR ART)

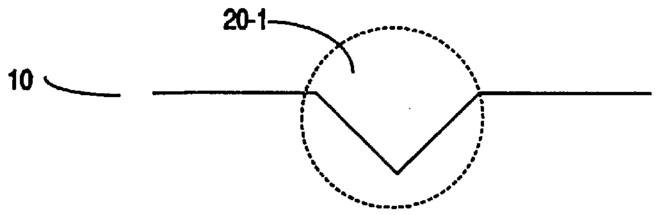

FIGURE 1B (PRIOR ART)

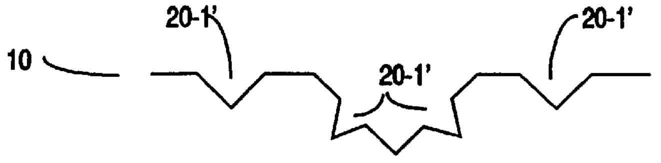
10
20-1'
20-1'
20-1'

FIGURE 1C (PRIOR ART)

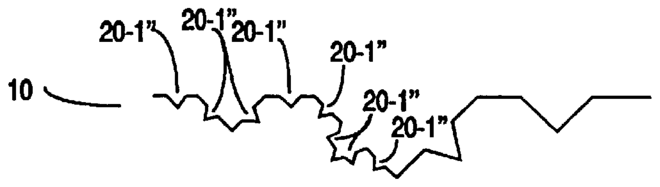
10
20-1"
20-1"
20-1"
20-1"
20-1"
20-1"

FIGURE 1D (PRIOR ART)

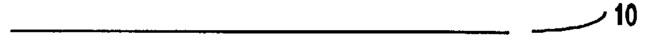

FIGURE 2A (PRIOR ART)

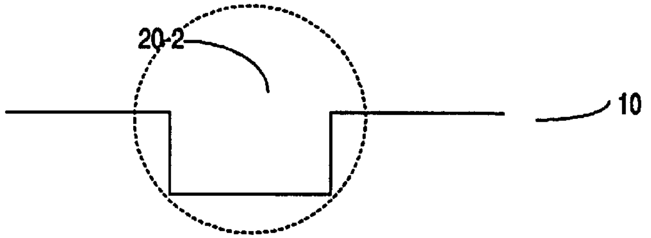

FIGURE 2B (PRIOR ART)

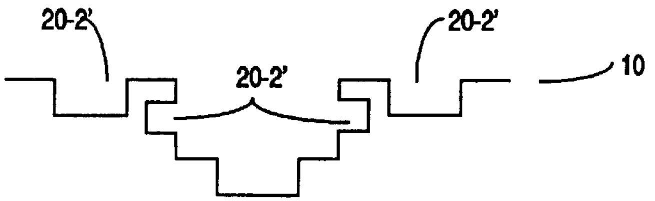
20-2'
20-2'
20-2'
10

FIGURE 2C (PRIOR ART)

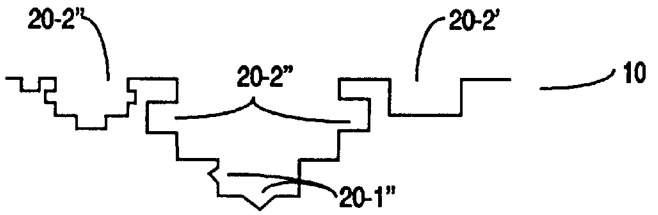
20-2"
20-2"
20-2'
10
20-1"

FIGURE 2D (PRIOR ART)

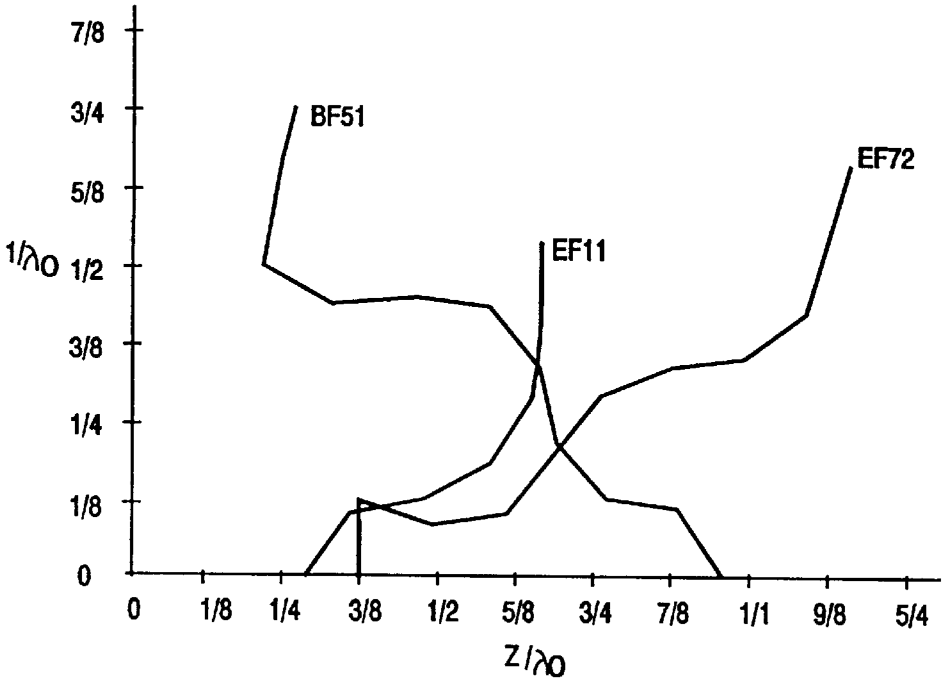

FIGURE 3 (PRIOR ART)

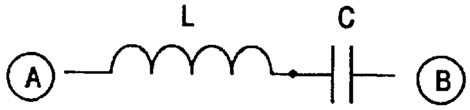

FIGURE 4A (PRIOR ART)

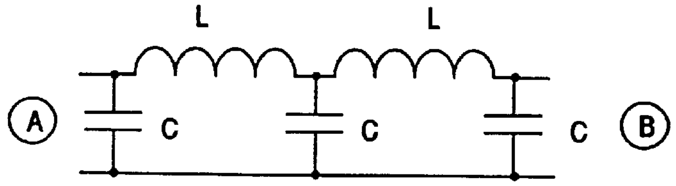

FIGURE 4B (PRIOR ART)

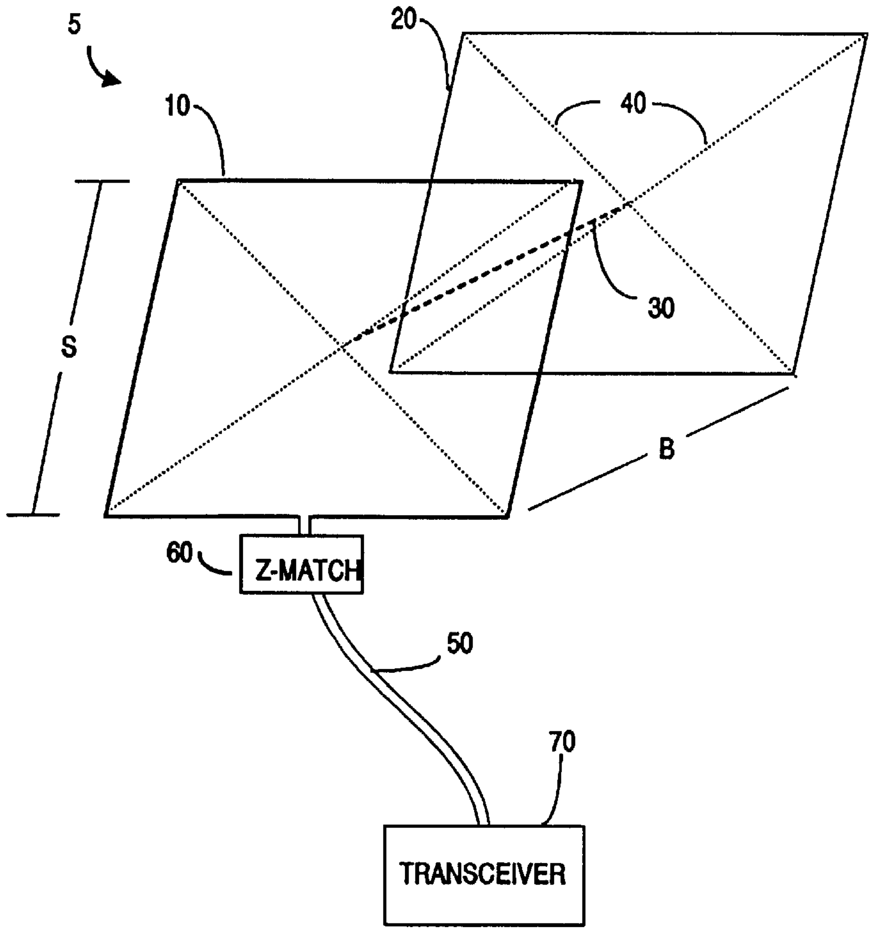
5
20
10
40
S
30
B
60
Z-MATCH
50
70
TRANSCEIVER

FIGURE 5A
(PRIOR ART)

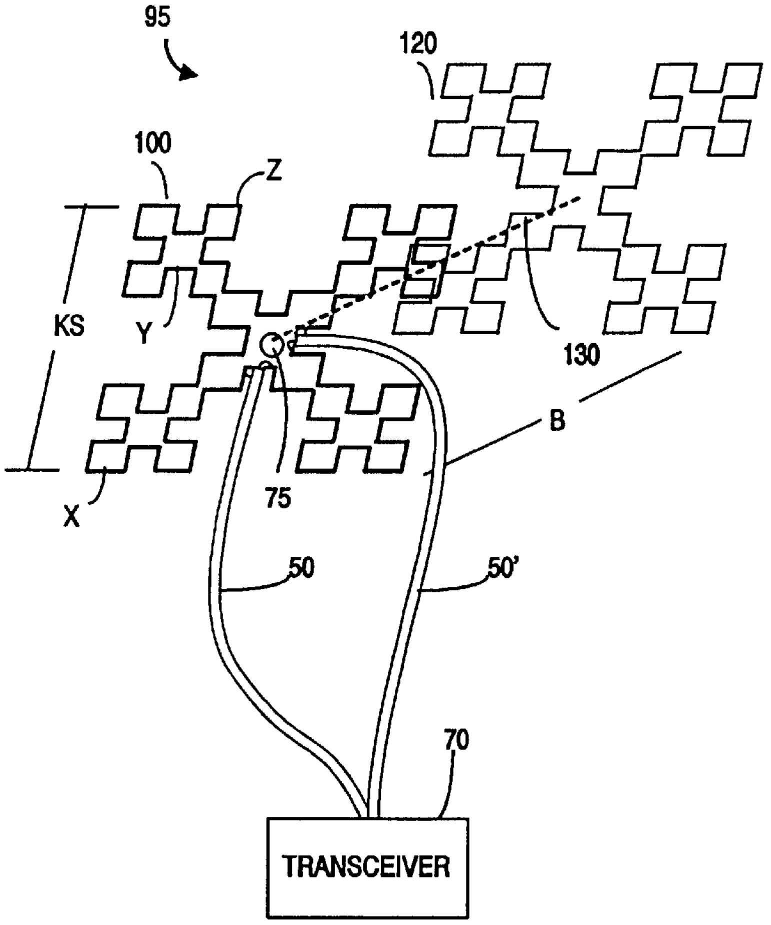
95
120
100
Z
KS
Y
130
B
75
X
50
50'
70
TRANSCEIVER

FIGURE 5B

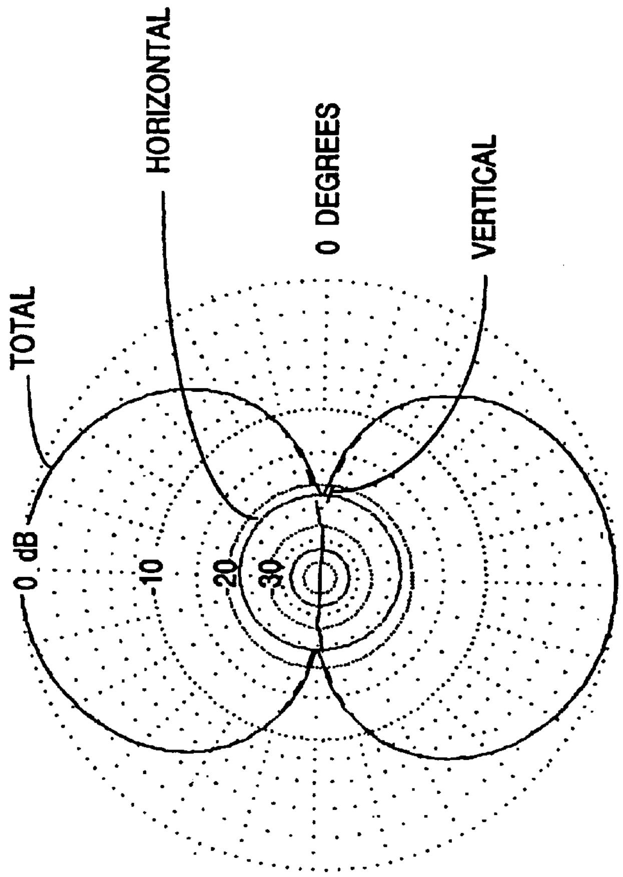

FIGURE 6

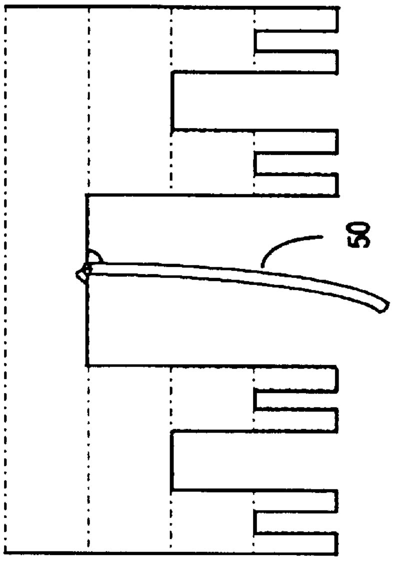

FIGURE 7A

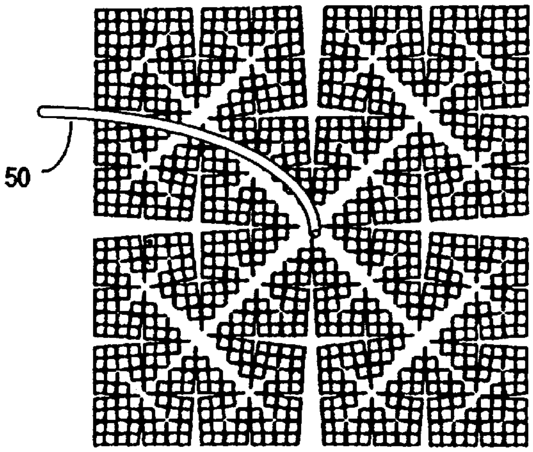

FIGURE 7B

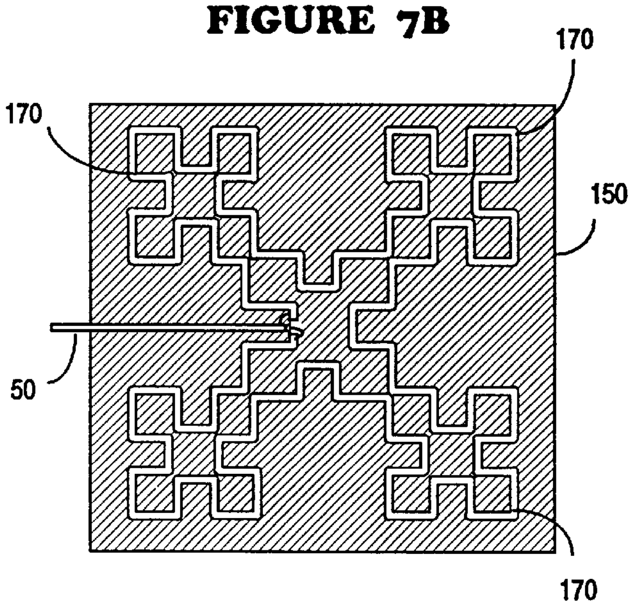
170
170
150
50
170

FIGURE 7C-1

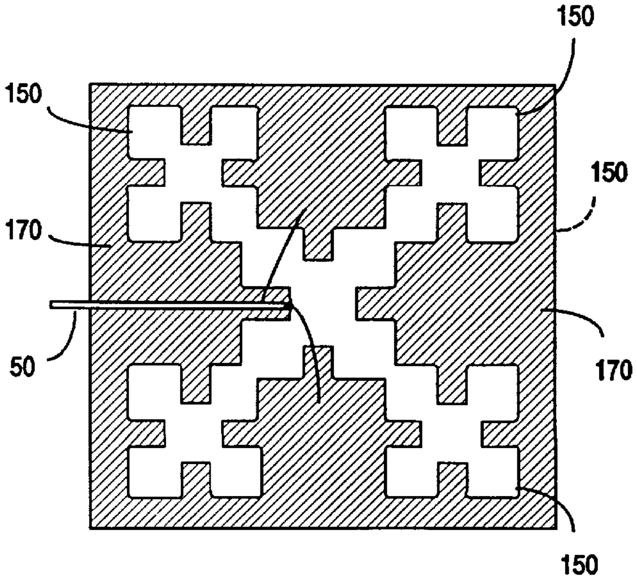
150
150
150
170
50
170
150

FIGURE 7C-2

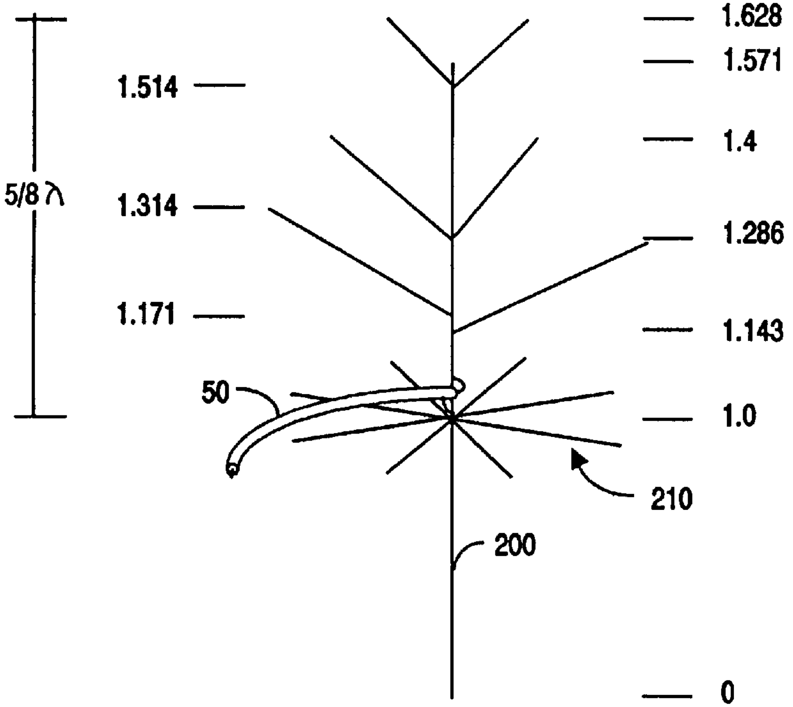
5/8 λ
1.514
1.314
1.171
50
1.628
1.571
1.4
1.286
1.143
1.0
210
200
0

FIGURE 7D

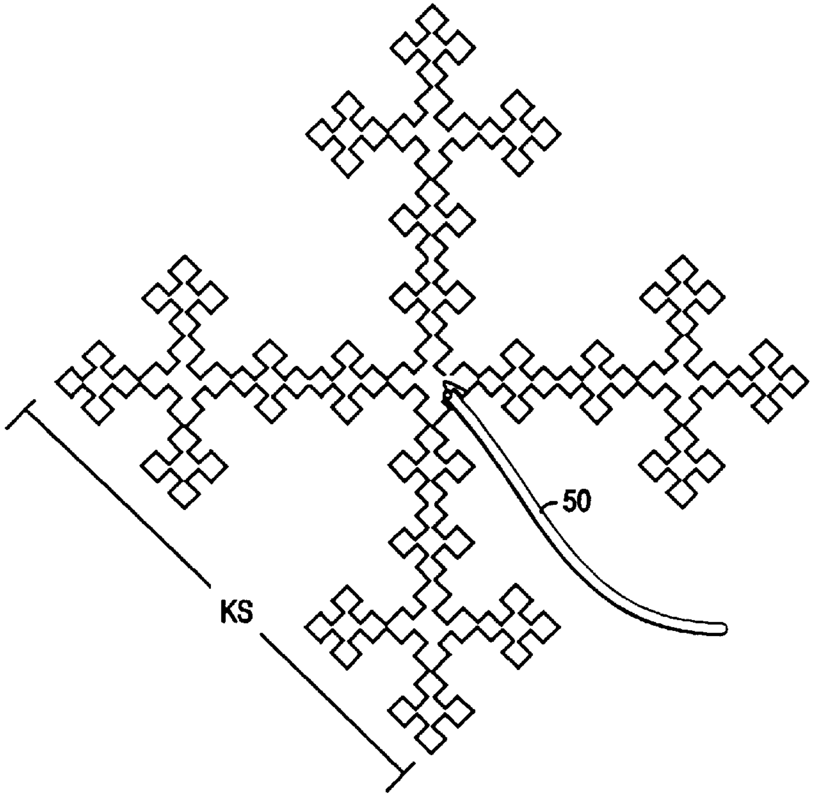

FIGURE 7E

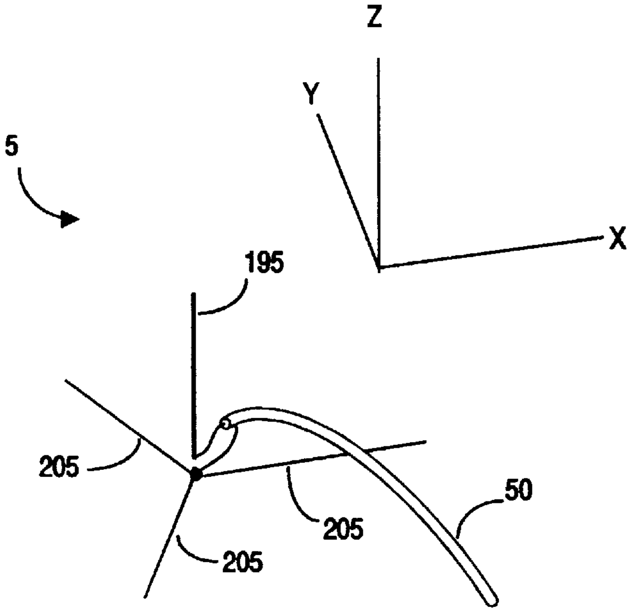
Z
Y
X
5
195
205
205
205
50

FIGURE 7D-1A

(PRIOR ART)

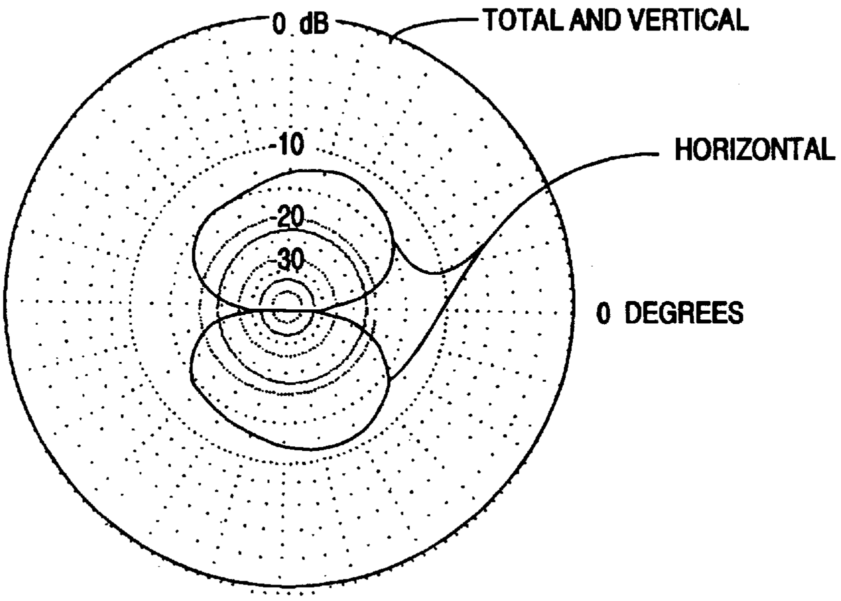

FIGURE 7D-1B

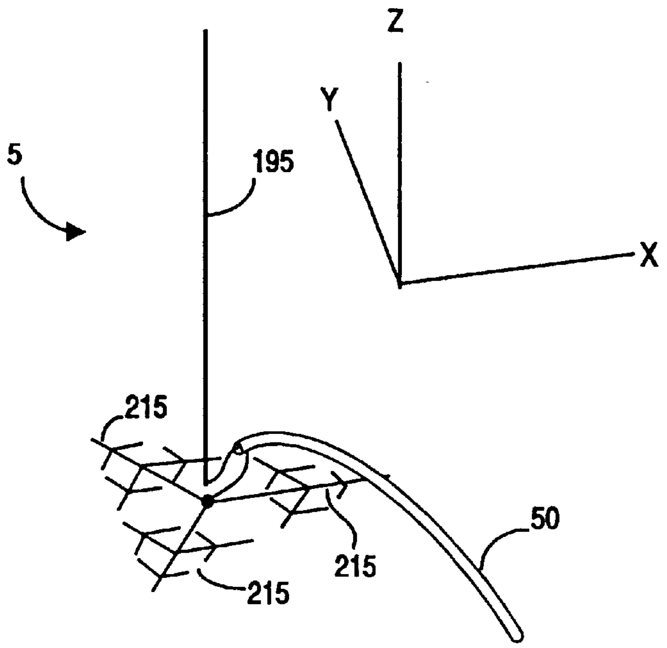
Z
Y
X
5
195
215
215
215
215
50

FIGURE 7D-2A

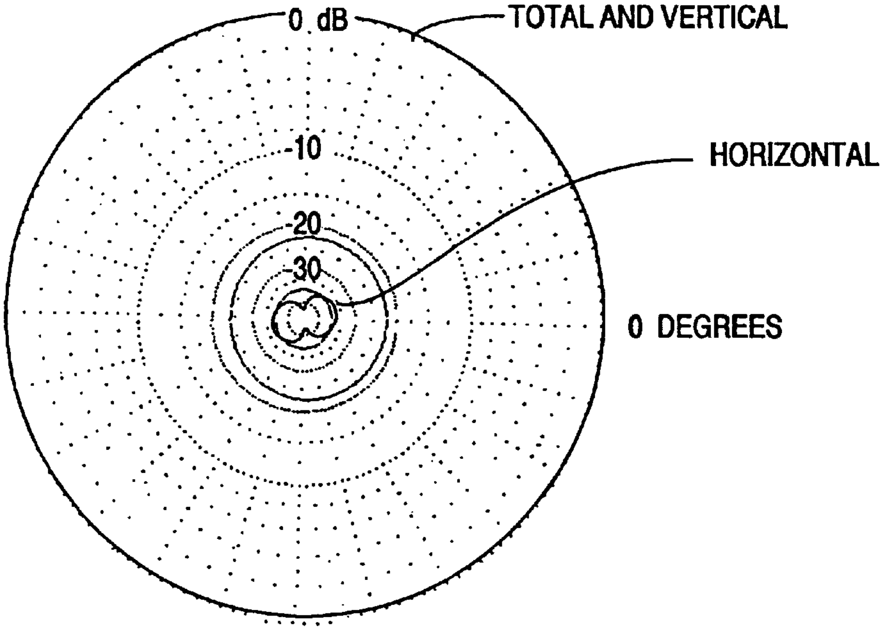

FIGURE 7D-2B

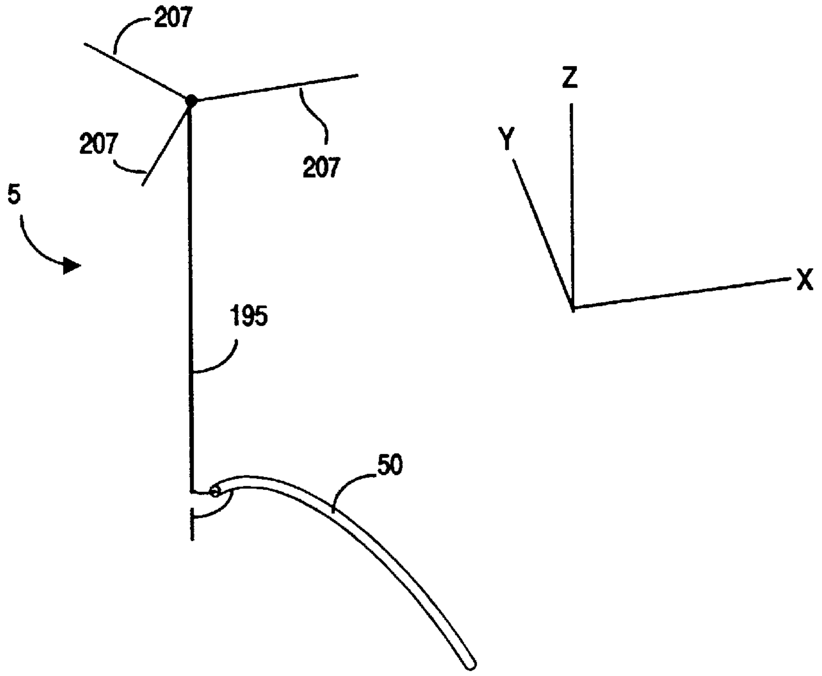
207
207
207
5
195
50
Z
Y
X

FIGURE 7D-3A
(PRIOR ART)

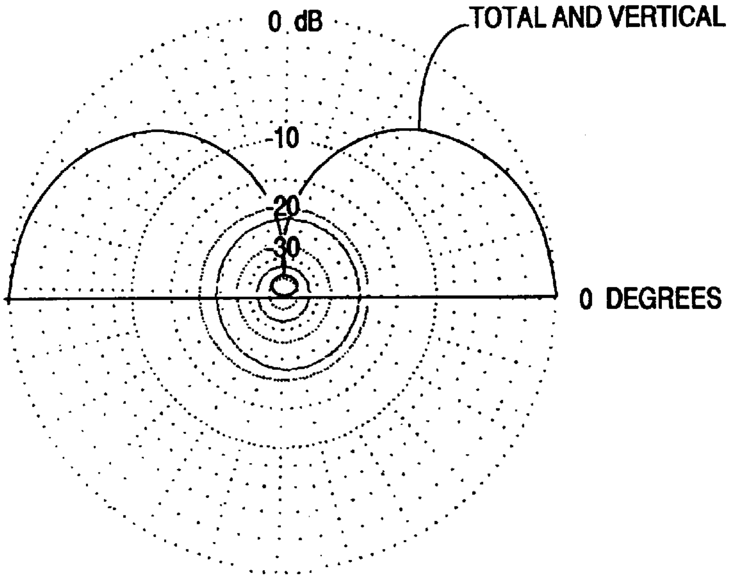

FIGURE 7D-3B
(PRIOR ART)

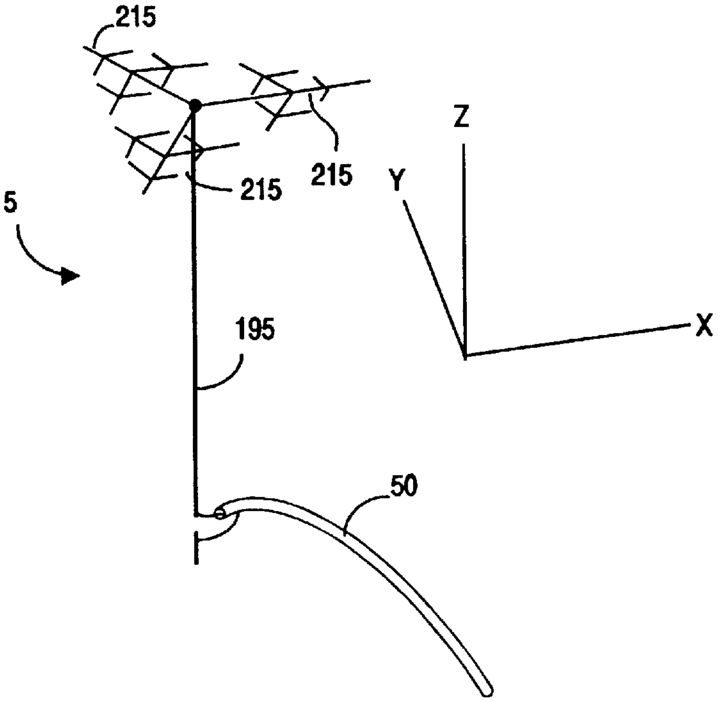
215
215
215
215
5
195
50
Z
Y
X

FIGURE 7D-4A

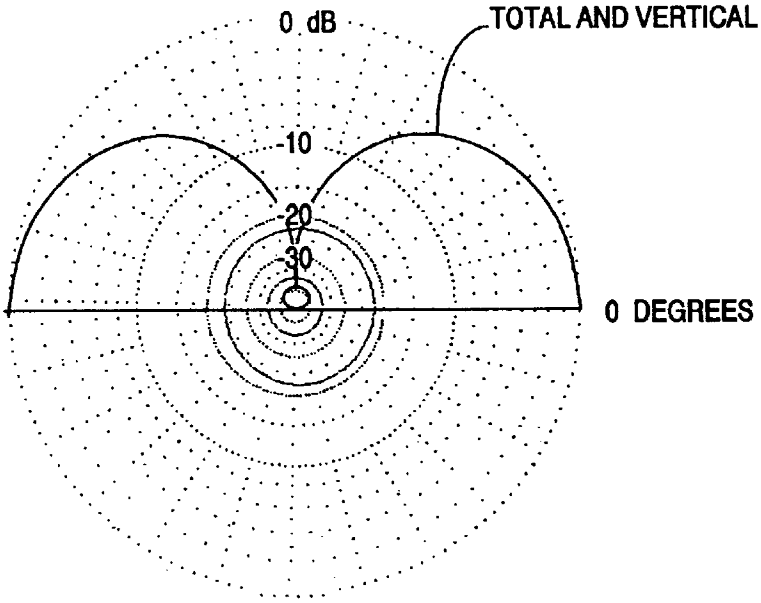

FIGURE 7D-4B

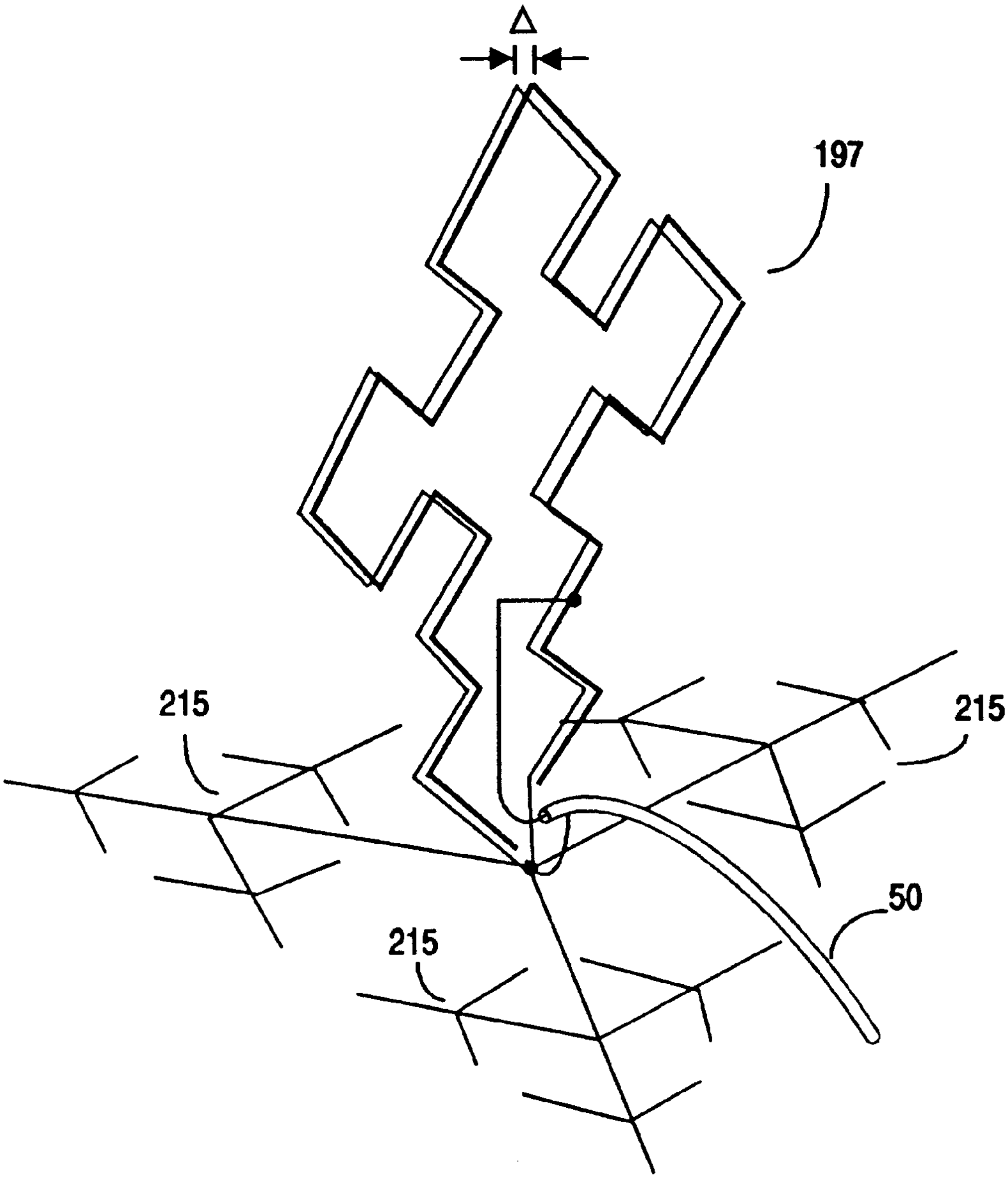
Δ
197
215
215
215
50

FIGURE 7D-5

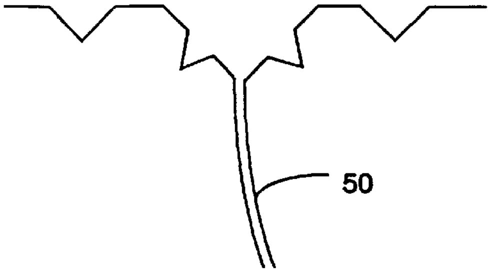

FIGURE 7F

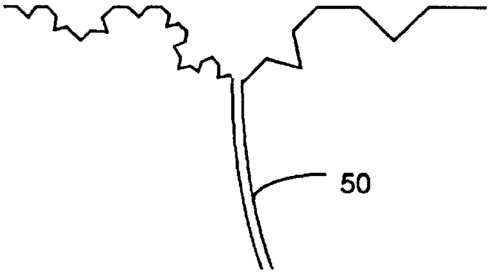

FIGURE 7G

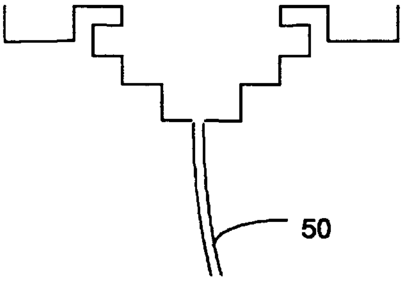

FIGURE 7H

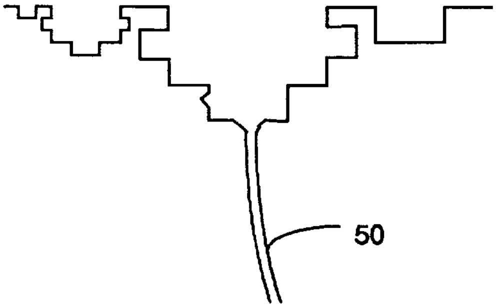

FIGURE 7I

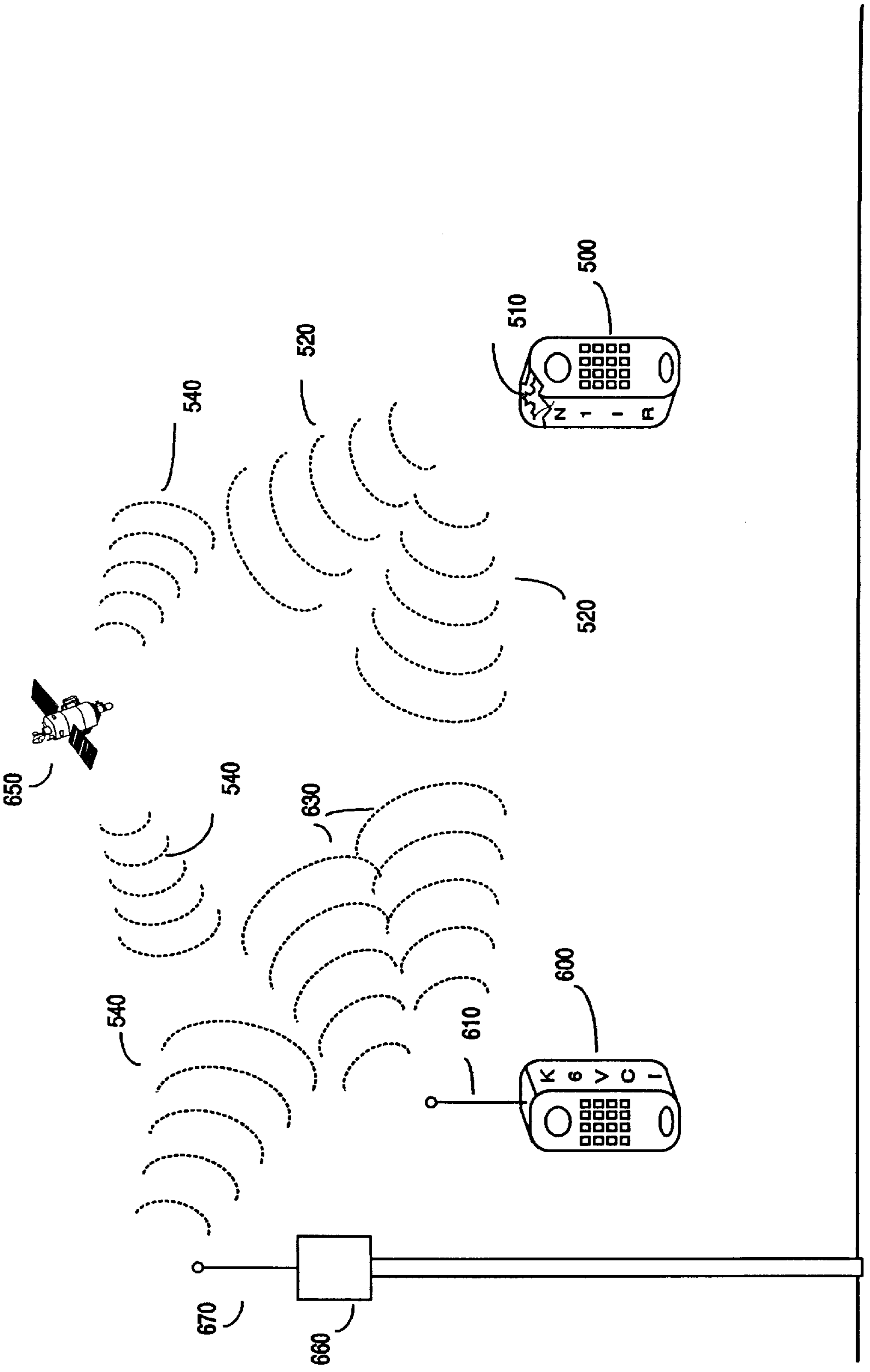
500
510
520
540
520
650
540
630
540
600
610
K 6 V C I
670
660

FIGURE 8A

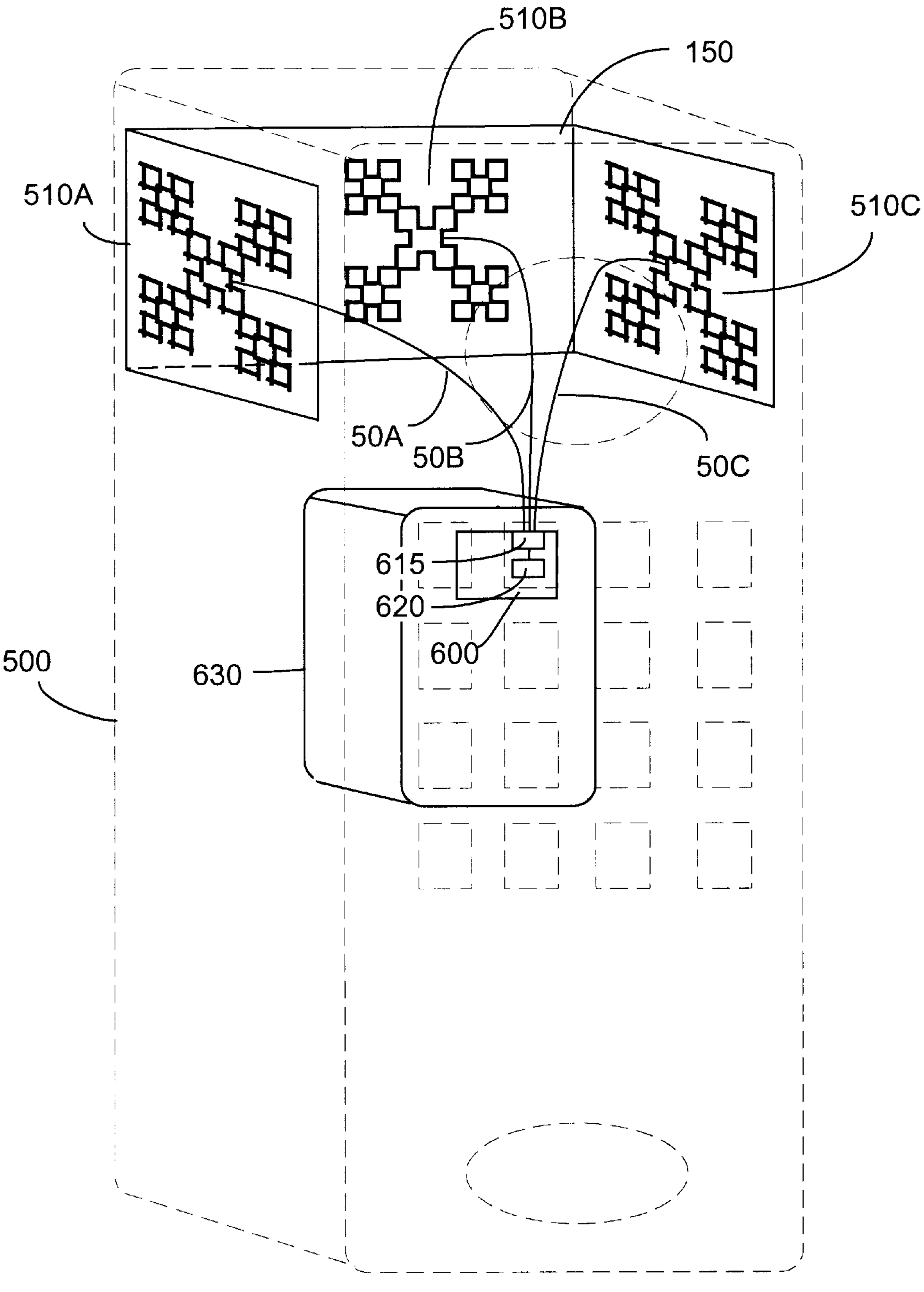
510B
150
510A
510C
50A
50B
50C
615
620
600
500
630

FIGURE 8B

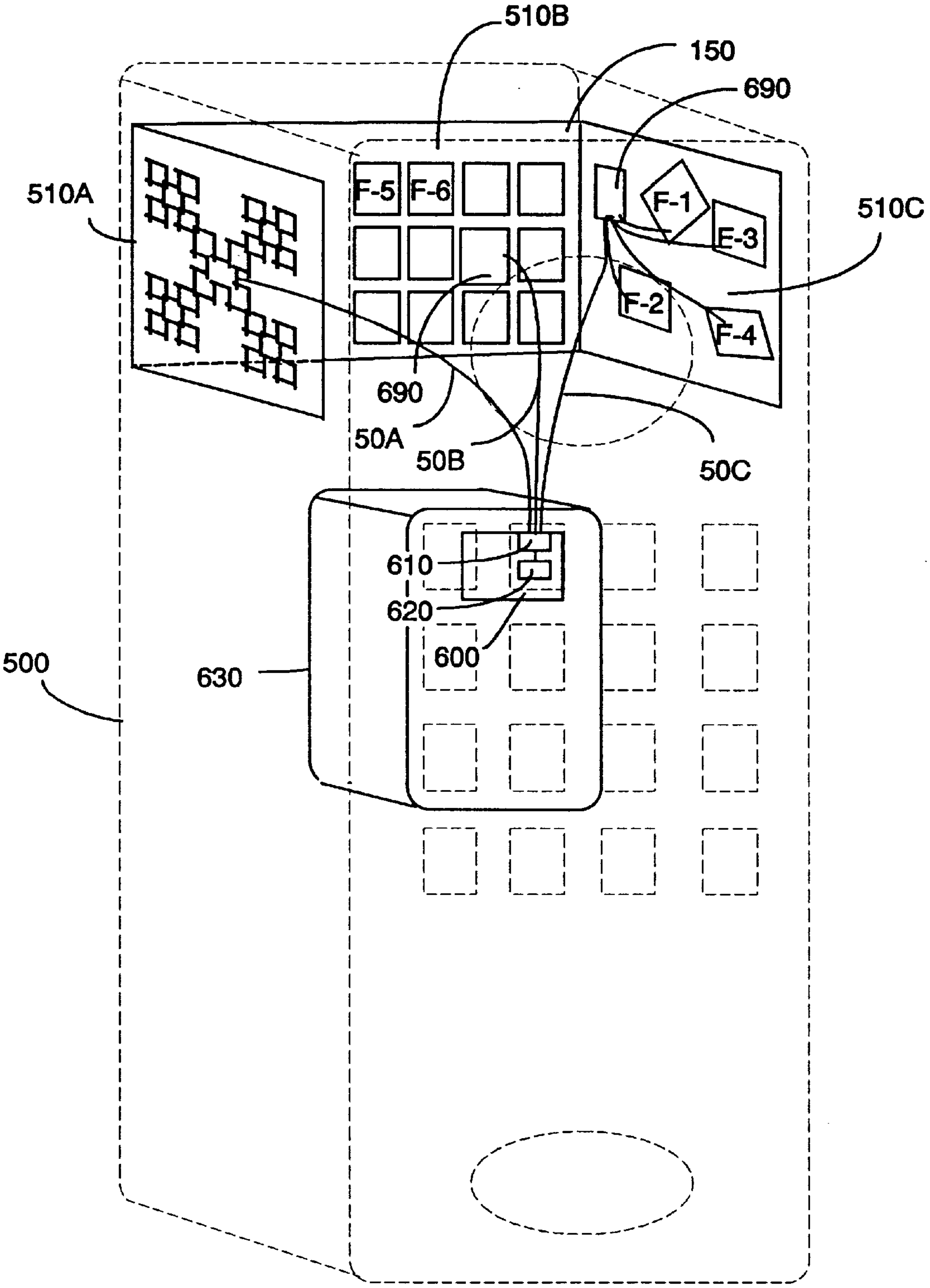
F-5
F-6
F-1
E-3
F-2
F-4

FIGURE 8C

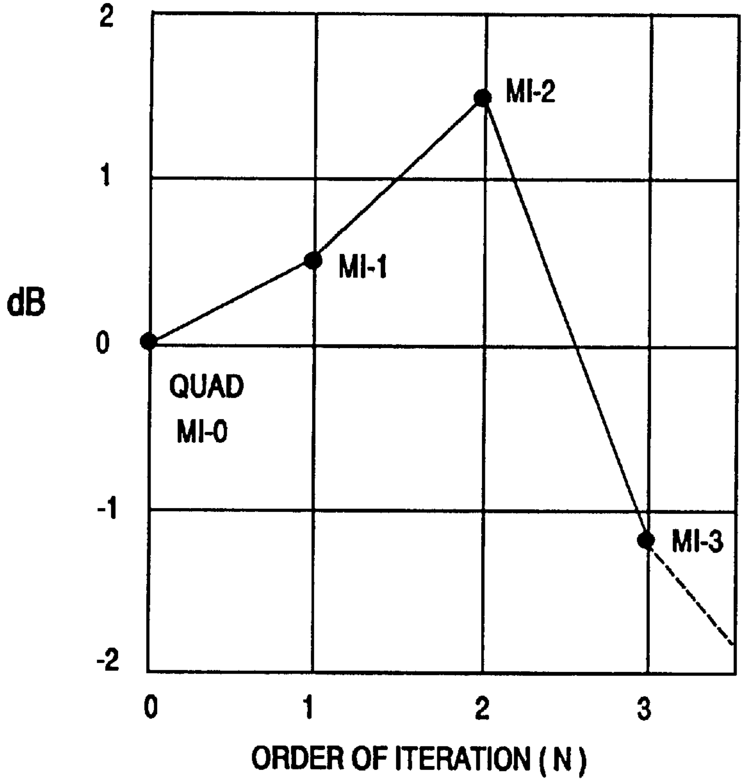

FIGURE 9A

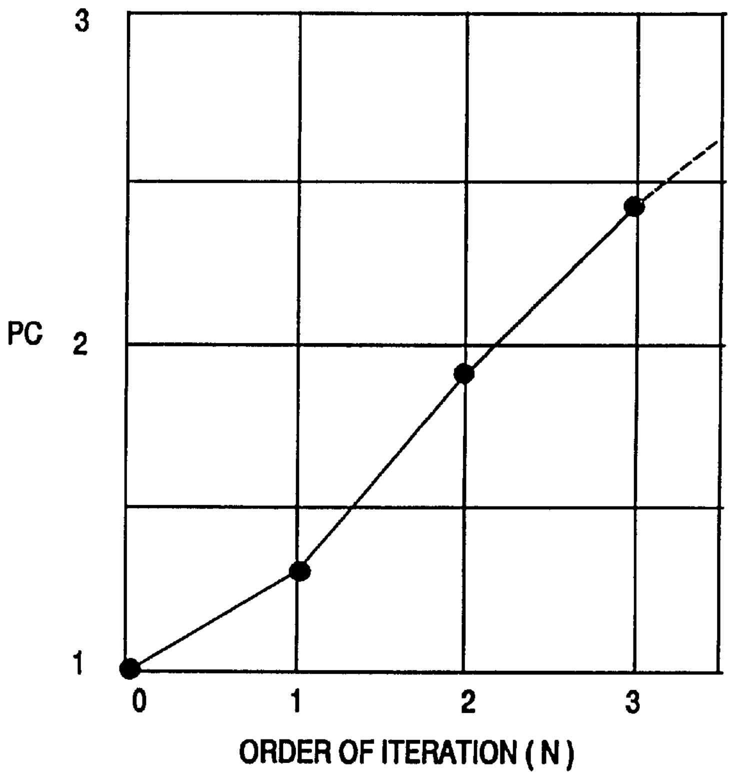

FIGURE 9B

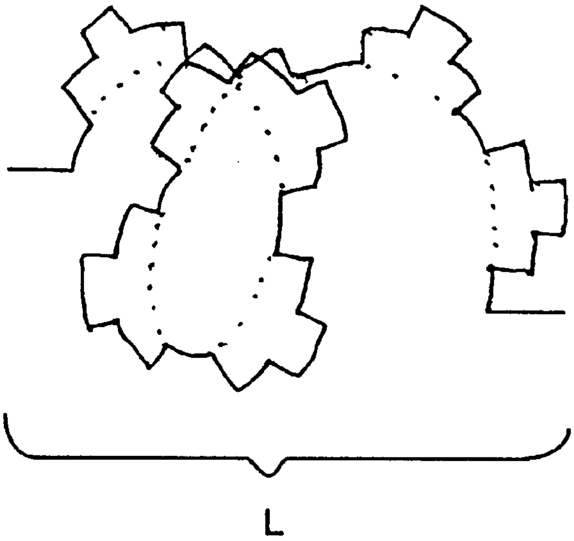

FIGURE 10A

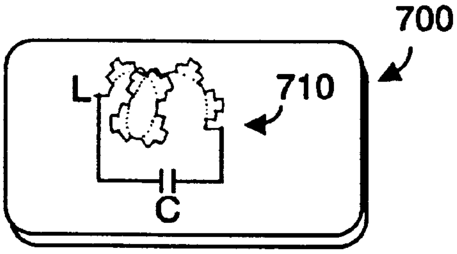

FIGURE 10B

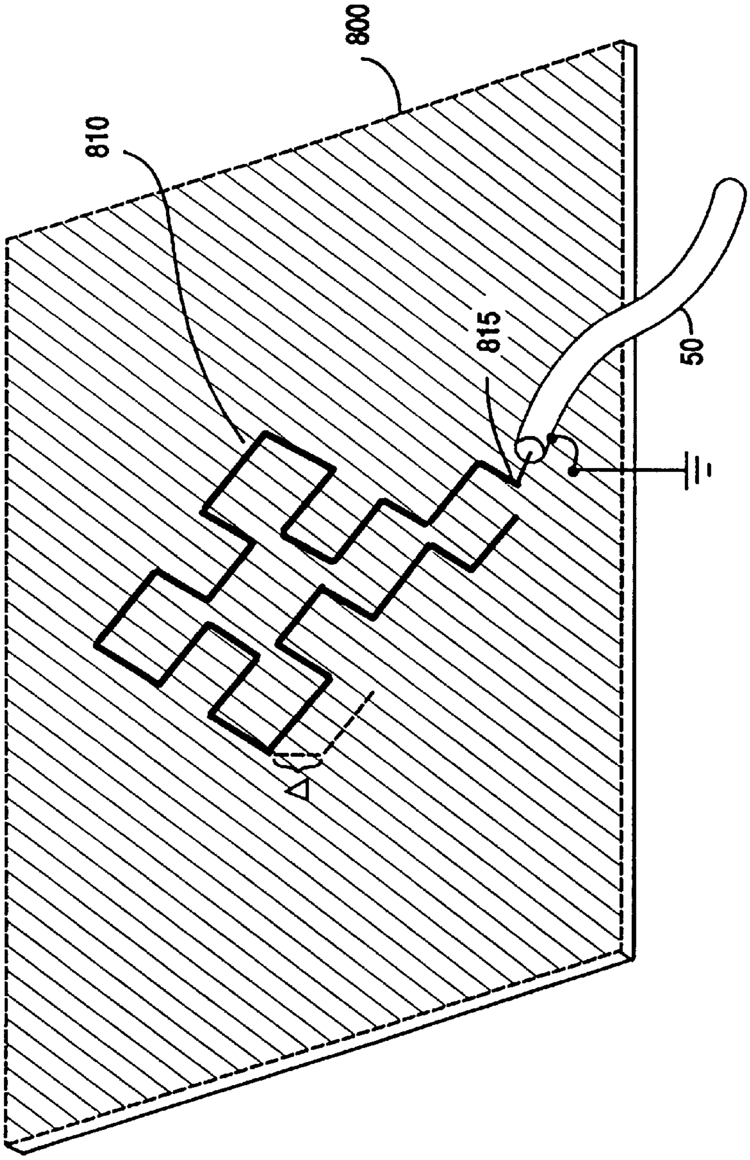
800
810
815
50
Δ

FIGURE 11A

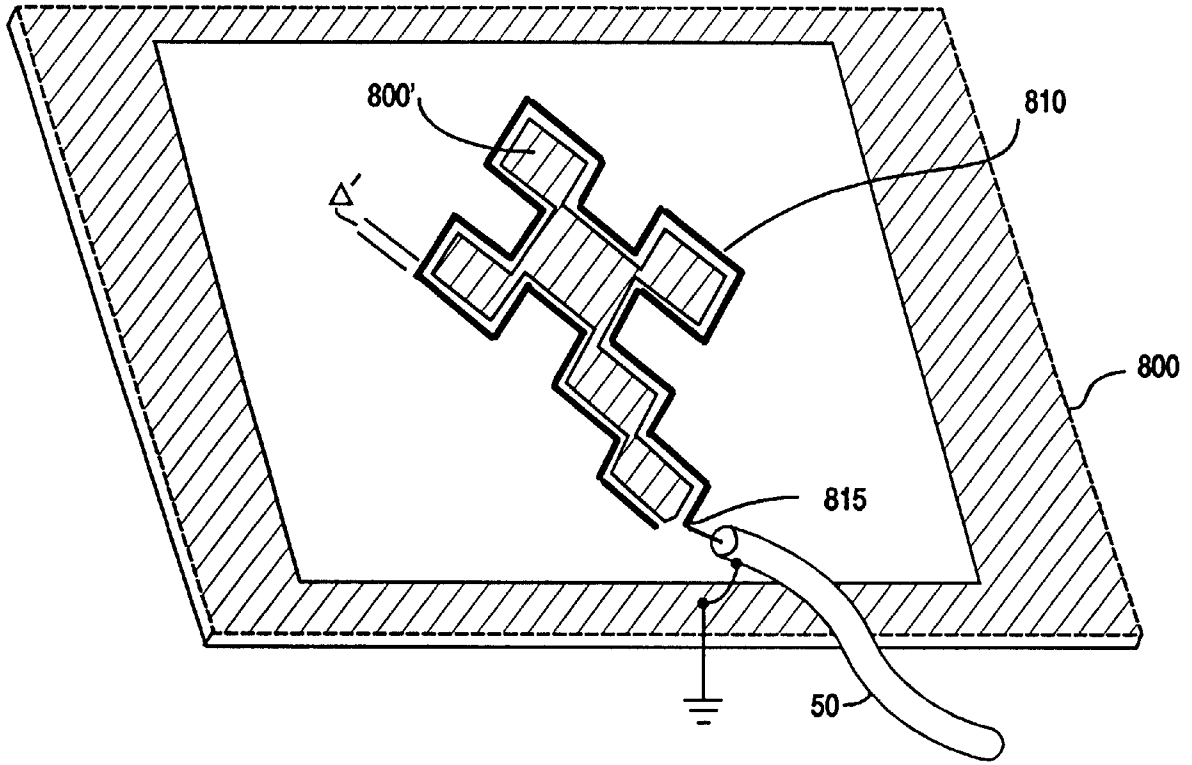
800'
810
800
815
50

FIGURE 11B

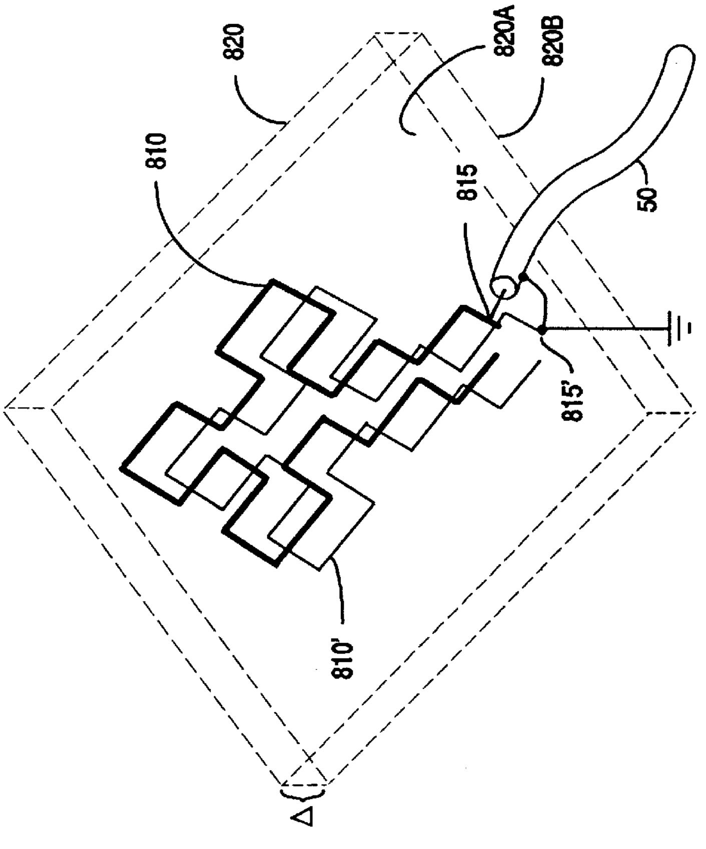
820
820A
820B
810
815
50
815'
810'
Δ

FIGURE 12A

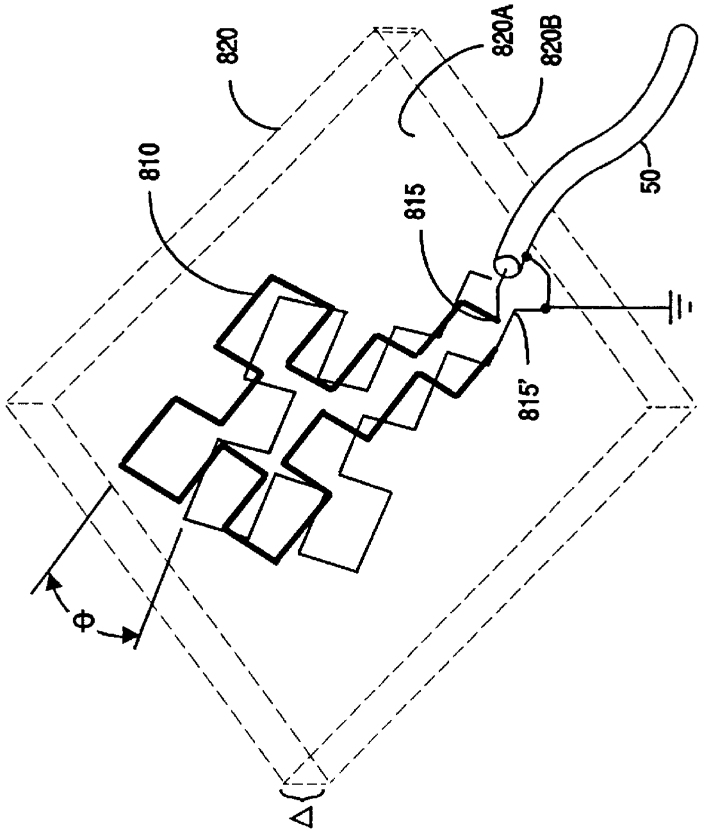
820
820A
820B
810
815
50
815'
Φ
Δ

FIGURE 12B

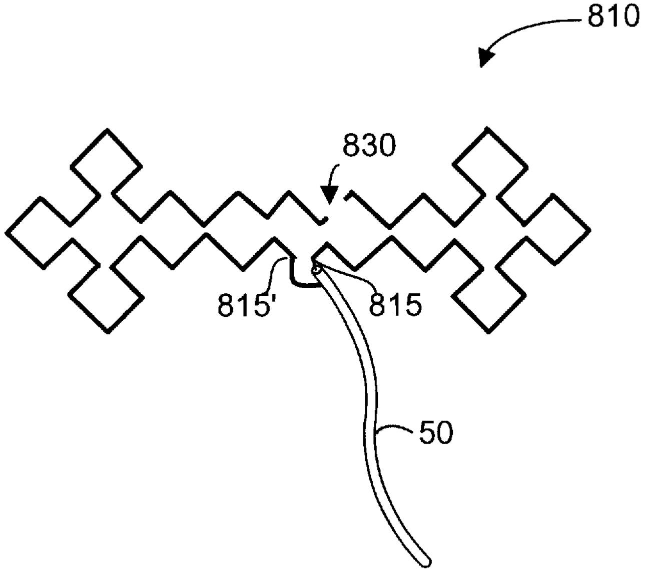
810
830
815'
815
50

FIGURE 13A

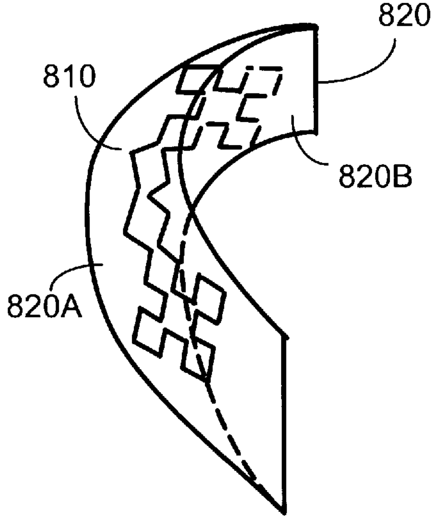
820
810
820B
820A

FIGURE 13B

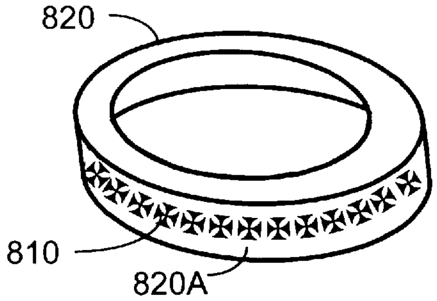
820
810
820A

FIGURE 13C

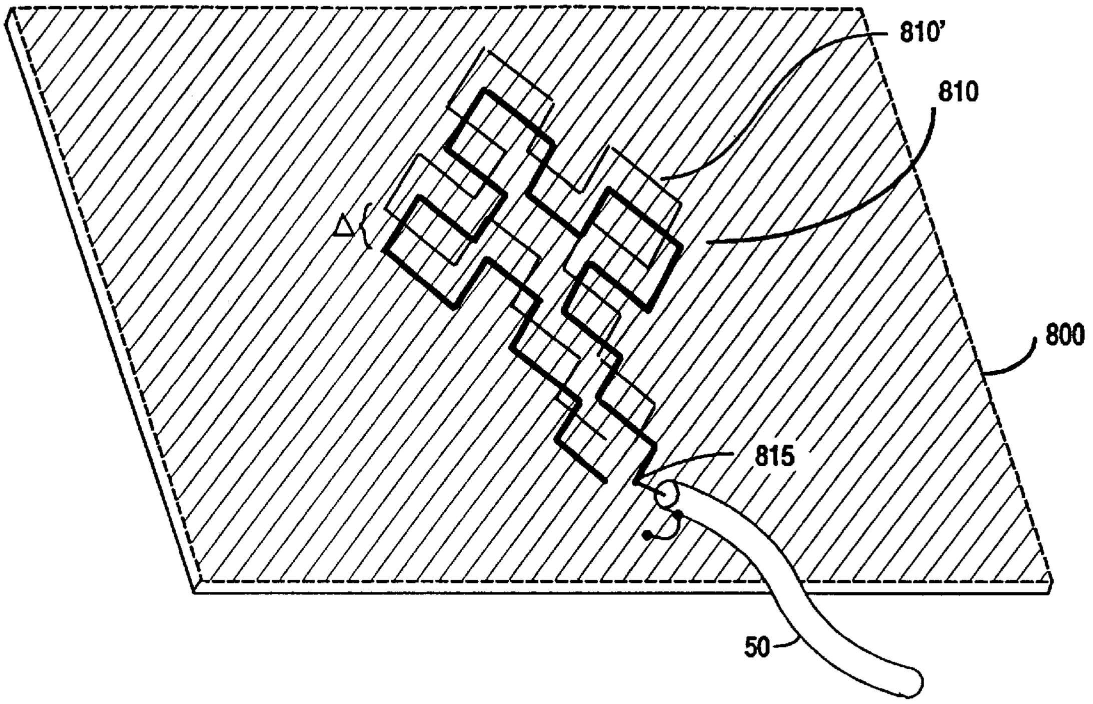
810'
810
Δ
800
815
50

FIGURE 14A

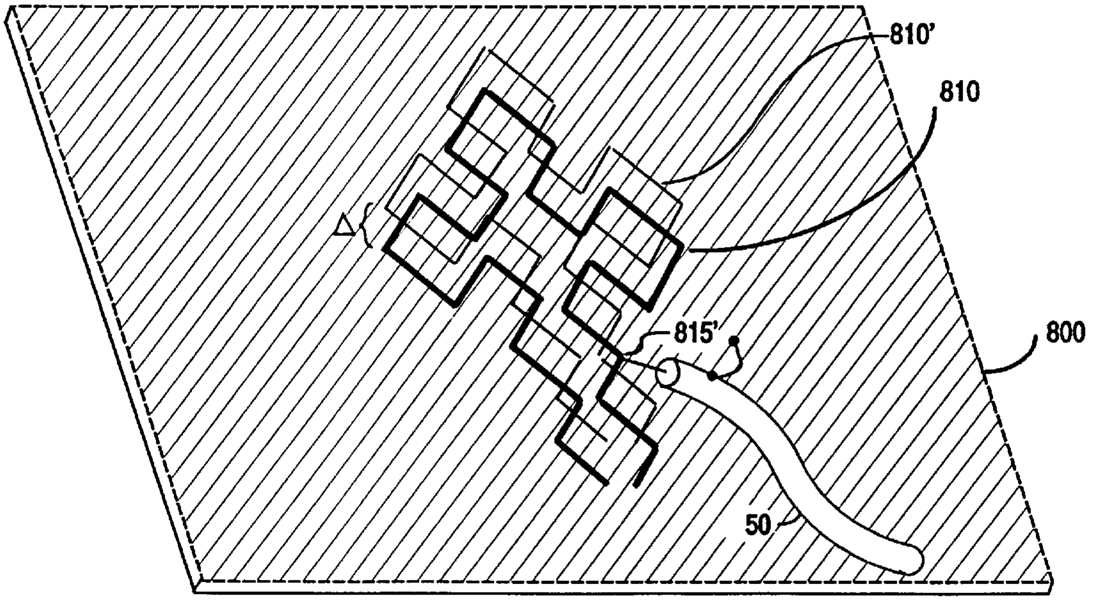
810'
810
Δ
815'
800
50

FIGURE 14B

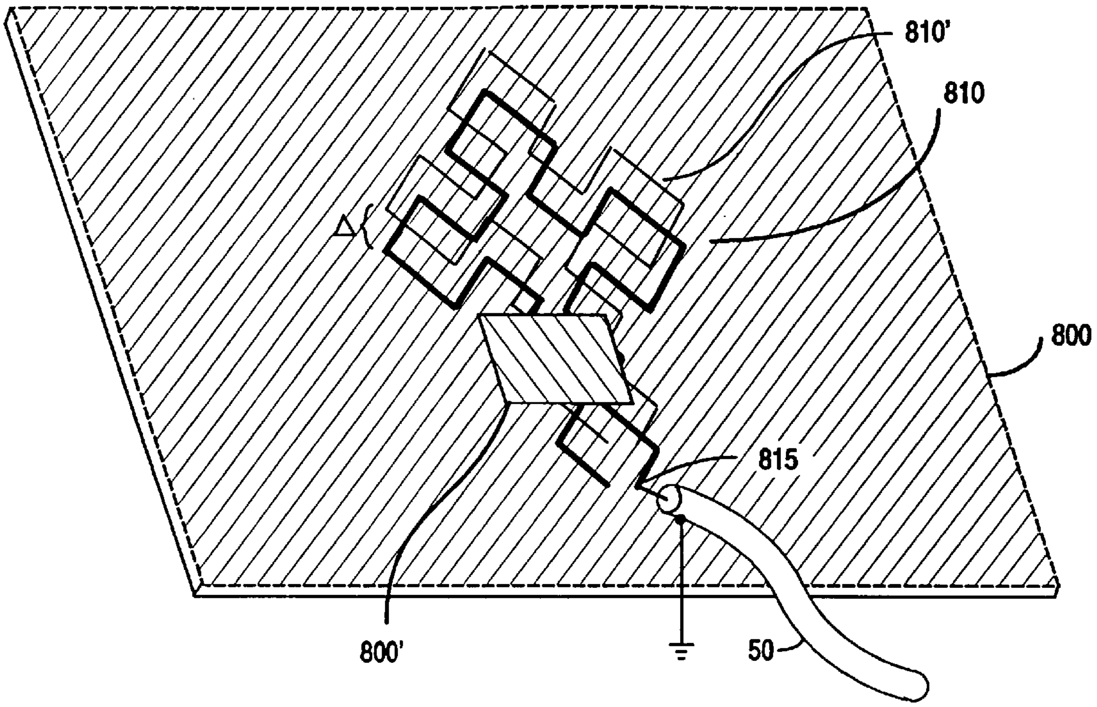
810'
810
Δ
800
815
800'
50

FIGURE 14C

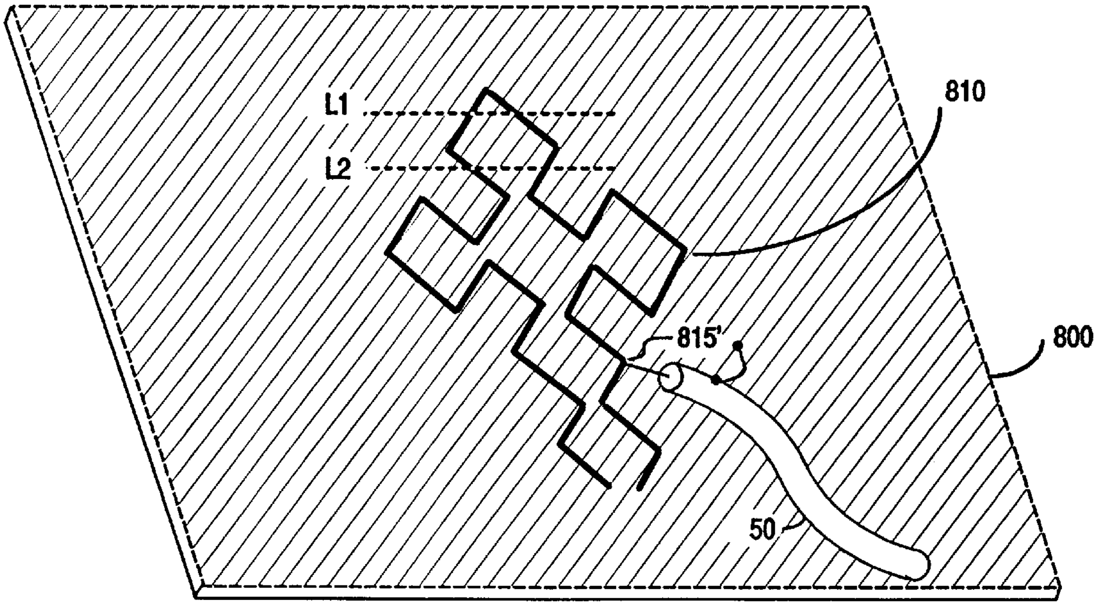
L1
L2
810
815'
800
50

FIGURE 14D

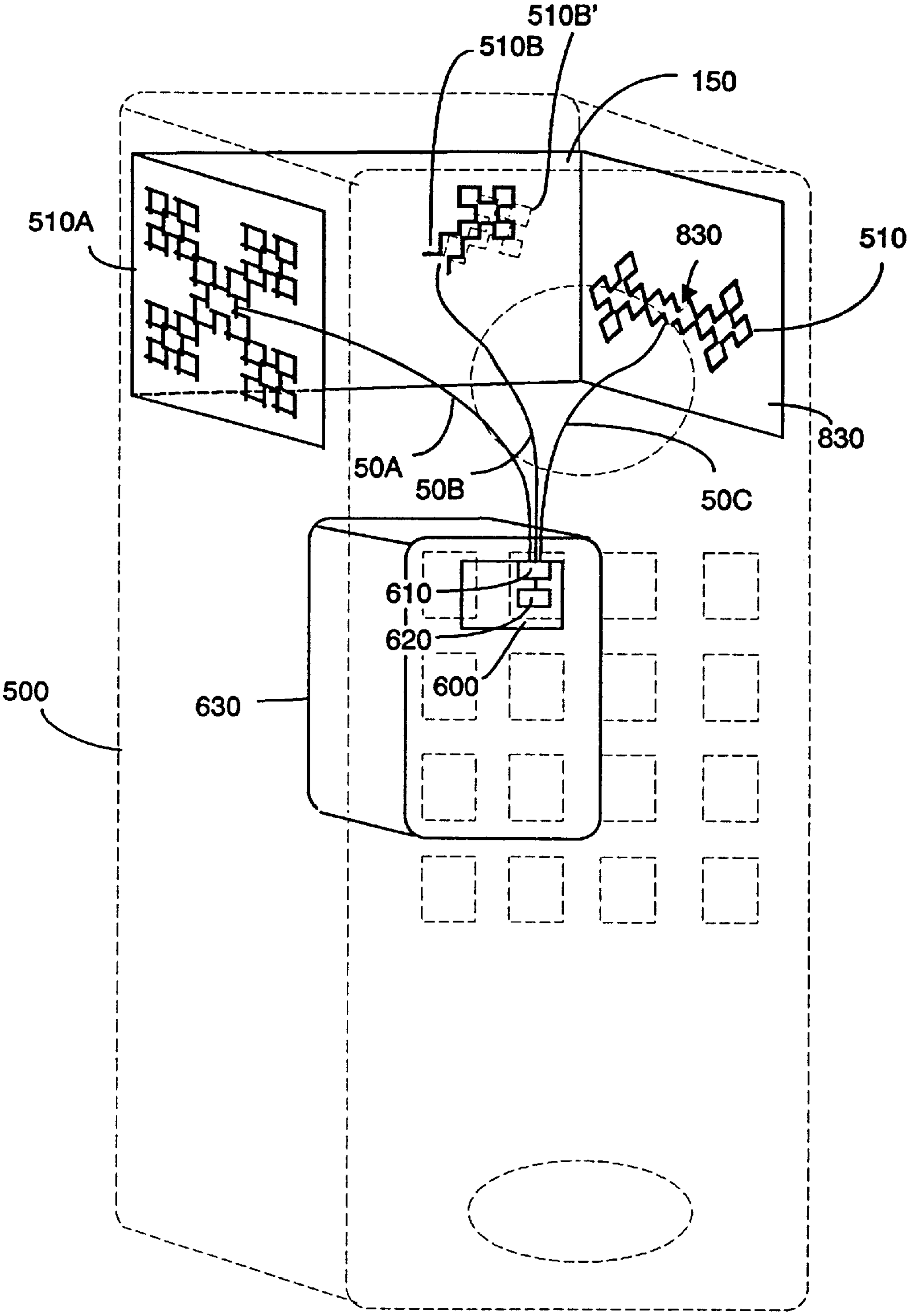
510B'
510B
150
510A
830
510C
830
50A
50B
50C
610
620
600
500
630

FIGURE 15

FRACTAL ANTENNA GROUND COUNTERPOISE, GROUND PLANES, AND LOADING ELEMENTS

RELATION TO PREVIOUSLY FILED PATENT APPLICATIONS

This is a continuation of application Ser. No. 08/649,825 filed May 17, 1996 is now abandoned.

This application is a continuing application from applicant's pending patent application Ser. No. 08,609,514 entitled TUNING FRACTAL ANTENNAS AND FRACTAL RESONATORS, filed on Mar. 1, 1996, abandoned, which itself is a continuation-in-part of applicant's pending patent application Ser. No. 08/512,954 entitled FRACTAL ANTENNAS AND FRACTAL RESONATORS, filed on Aug. 9, 1995, now abandoned.

FIELD OF THE INVENTION

The present invention relates to antennas and resonators, and specifically to designing and tuning non-Euclidian antenna ground radials, ground counterpoise or planes, top-loading elements, and antennas using such elements.

BACKGROUND OF THE INVENTION

Antenna are used to radiate and/or receive typically electromagnetic signals, preferably with antenna gain, directivity, and efficiency. Practical antenna design traditionally involves trade-offs between various parameters, including antenna gain, size, efficiency, and bandwidth.

Antenna design has historically been dominated by Euclidean geometry. In such designs, the closed antenna area is directly proportional to the antenna perimeter. For example, if one doubles the length of an Euclidean square (or "quad") antenna, the enclosed area of the antenna quadruples. Classical antenna design has dealt with planes, circles, triangles, squares, ellipses, ectangles, hemispheres, paraboloids, and the like, (as ell as lines). Similarly, resonators, typically apacitors ("C") coupled in series and/or parallel with inductors ("L"), traditionally are implemented with Euclidian inductors.

With respect to antennas, prior art design philosophy has been to pick a Euclidean geometric construction, e.g., a quad, and to explore its radiation characteristics, especially with emphasis on frequency resonance and power patterns. The unfortunate result is that antenna design has far too long concentrated on the ease of antenna construction, rather than on the underlying electromagnetics.

Many prior art antennas are based upon closed-loop or island shapes. Experience has long demonstrated that small sized antennas, including loops, do not work well, one reason being that radiation resistance ("R") decreases sharply when the antenna size is shortened. A small sized loop, or even a short dipole, will exhibit a radiation pattern of ½ λ and ¼ λ, respectively, if the radiation resistance R is not swamped by substantially larger ohmic ("O") losses. Ohmic losses can be minimized using impedance matching networks, which can be expensive and difficult to use. But although even impedance matched small loop antennas can exhibit 50% to 85% efficiencies, their bandwidth is inherently narrow, with very high Q, e.g., $Q>50$. As used herein, Q is defined as (transmitted or received frequency)/(3 dB bandwidth).

As noted, it is well known experimentally that radiation resistance R drops rapidly with small area Euclidean antennas. However, the theoretical basis is not generally known, and any present understanding (or misunderstanding) appears to stem from research by J. Kraus, noted in *Antennas* (Ed. 1), McGraw Hill, New York (1950), in which a circular loop antenna with uniform current was examined. Kraus' loop exhibited a gain with a surprising limit of 1.8 dB over an isotropic radiator as loop area fells below that of a loop having a 1 λ-squared aperture. For small loops of area $A<\lambda^2/100$, radiation resistance R was given by:

$$R = K\cdot\left(\frac{A}{\lambda^2}\right)^2$$

where K is a constant, A is the enclosed area of the loop, and λ is wavelength. Unfortunately, radiation resistance R can all too readily be less than 1 Ω for a small loop antenna.

From his circular loop research Kraus generalized that calculations could be defined by antenna area rather than antenna perimeter, and that his analysis should be correct for small loops of any geometric shape. Kraus' early research and conclusions that small-sized antennas will exhibit a relatively large ohmic resistance O and a relatively small radiation resistance R, such that resultant low efficiency defeats the use of the small antenna have been widely accepted. In fact, some researchers have actually proposed reducing ohmic resistance O to 0 Ω by constructing small antennas from superconducting material, to promote efficiency.

As noted, prior art antenna and resonator design has traditionally concentrated on geometry that is Euclidean. However, one non-Euclidian geometry is fractal geometry. Fractal geometry may be grouped into random fractals, which are also termed chaotic or Brownian fractals and include a random noise components, such as depicted in FIG. **3**, or deterministic fractals such as shown in FIG. **1**C.

In deterministic fractal geometry, a self-similar structure results from the repetition of a design or motif (or "generator"), on a series of different size scales. One well known treatise in this field is *Fractals, Endlessly Repeated Geometrical Figures*, by Hans Lauwerier, Princeton University Press (1991), which treatise applicant refers to and incorporates herein by reference.

FIGS. **1**A–**2**D depict the development of some elementary forms of fractals. In FIG. **1**A, a base element **10** is shown as a straight line, although a curve could instead be used. In FIG. **1**B, a so-called Koch fractal motif or generator **20-1**, here a triangle, is inserted into base element **10**, to form a first order iteration ("N") design, e.g., N=1. In FIG. **1**C, a second order N=2 iteration design results from replicating the triangle motif **20-1** into each segment of FIG. **1**B, but where the **20-1**' version has been differently scaled, here reduced in size. As noted in the Lauwerier treatise, in its replication, the motif may be rotated, translated, scaled in dimension, or a combination of any of these characteristics. Thus, as used herein, second order of iteration or N=2 means the fundamental motif has been replicated, after rotation, translation, scaling (or a combination of each) into the first order iteration pattern. A higher order, e.g., N=3, iteration means a third fractal pattern has been generated by including yet another rotation, translation, and/or scaling of the first order motif.

In FIG. **1**D, a portion of FIG. **1**C has been subjected to a further iteration (N=3) in which scaled-down versions **20-1**' of the triangle motif **20-1** have been inserted into each segment of the left half of FIG. **1**C. FIGS. **2**A–**2**C follow what has been described with respect to FIGS. **1**A–**1**C, except that a rectangular motif **20-2** has been adopted which motif is denoted **20-2**' in FIG. **2**C, and **20-2**" in FIG. **2**D.

FIG. 2D shows a pattern in which a portion of the left-hand side is an N=3 iteration of the **20-2** rectangle motif, and in which the center portion of the figure now includes another motif, here a **20-1** type triangle motif, and in which the right-hand side of the figure remains an N=2 iteration.

Traditionally, non-Euclidean designs including random fractals have been understood to exhibit antiresonance characteristics with mechanical vibrations. It is known in the art to attempt to use non-Euclidean random designs at lower frequency regimes to absorb, or at least not reflect sound due to the antiresonance characteristics. For example, M. Schroeder in *Fractals, Chaos, Power Laws* (1992), W. H. Freeman, New York discloses the use of presumably random or chaotic fractals in designing sound blocking diffusers for recording studios and auditoriums.

Experimentation with non-Euclidean structures has also been undertaken with respect to electromagnetic waves, including radio antennas. In one experiment, Y. Kim and D. Jaggard in *The Fractal Random Array*, Proc. IEEE 74, 1278–1280 (1986) spread-out antenna elements in a sparse microwave array, to minimize sidelobe energy without having to use an excessive number of elements. But Kim and Jaggard did not apply a fractal condition to the antenna elements, and test results were not necessarily better than any other techniques, including a totally random spreading of antenna elements. More significantly, the resultant array was not smaller than a conventional Euclidean design.

Prior art spiral antennas, cone antennas, and V-shaped antennas may be considered as a continuous, deterministic first order fractal, whose motif continuously expands as distance increases from a central point. A log-periodic antenna may be considered a type of continuous fractal in that it is fabricated from a radially expanding structure. However, log periodic antennas do not utilize he antenna perimeter for radiation, but instead rely upon an arc-like opening angle in the antenna geometry.

Such opening angle is an angle that defines the size-scale of the log-periodic structure, which structure is proportional to the distance from the antenna center multiplied by the opening angle. Further, known log-periodic antennas are not necessarily smaller than conventional driven element-parasitic element antenna designs of similar gain.

Unintentionally, first order fractals have been used to distort the shape of dipole and vertical antennas to increase gain, the shapes being defined as a Brownian-type of chaotic fractals. See F. Landstorfer and R. Sacher, *Optimisation of Wire Antennas*, J. Wiley, New York (1985). FIG. **3** depicts three bent-vertical antennas developed by Landstorfer and Sacher through trial and error, the plots showing the actual vertical antennas as a function of x-axis and y-axis coordinates that are a function of wavelength. The “EF” and “BF” nomenclature in FIG. **3** refer respectively to end-fire and back-fire radiation patterns of the resultant bent-vertical antennas.

First order fractals have also been used to reduce horn-type antenna geometry, in which a double-ridge horn configuration is used to decrease resonant frequency. See J. Kraus in *Antennas*, McGraw Hill, New York (1885). The use of rectangular, box-like, and triangular shapes as impedance-matching loading elements to shorten antenna element dimensions is also known in the art.

Whether intentional or not, such prior art attempts to use a quasi-fractal or fractal motif in an antenna employ at best a first order iteration fractal. By first iteration it is meant that one Euclidian structure is loaded with another Euclidean structure in a repetitive fashion, using the same size for repetition. FIG. 1C, for example, is not first order because the **20-1**' triangles have been shrunk with respect to the size of the first motif **20-1**.

Prior art antenna design does not attempt to exploit multiple scale self-similarity of real fractals. This is hardly surprising in view of the accepted conventional wisdom that because such antennas would be anti-resonators, and/or if suitably shrunken would exhibit so small a radiation resistance R, that the substantially higher ohmic losses O would result in too low an antenna efficiency for any practical use. Further, it is probably not possible to mathematically predict such an antenna design, and high order iteration fractal antennas would be increasingly difficult to fabricate and erect, in practice.

FIGS. **4**A and **4**B depict respective prior art series and parallel type resonator configurations, comprising capacitors C and Euclidean inductors L. In the series configuration of FIG. **4**A, a notch-filter characteristic is presented in that the impedance from port A to port B is high except at frequencies approaching resonance, determined by $1/\surd(LC)$.

In the distributed parallel configuration of FIG. **4**B, a low-pass filter characteristic is created in that at frequencies below resonance, there is a relatively low impedance path from port A to port B, but at frequencies greater than resonant frequency, signals at port A are shunted to ground (e.g., common terminals of capacitors C), and a high impedance path is presented between port A and port B. Of course, a single parallel LC configuration may also be created by removing (e.g., short-circuiting) the rightmost inductor L and right two capacitors C, in which case port B would be located at the bottom end of the leftmost capacitor C.

In FIGS. **4**A and **4**B, inductors L are Euclidean in that increasing the effective area captured by the inductors increases with increasing geometry of the inductors, e.g., more or larger inductive windings or, if not cylindrical, traces comprising inductance. In such prior art configurations as FIGS. **4**A and **4**B, the presence of Euclidean inductors L ensures a predictable relationship between L, C and frequencies of resonance.

Applicant’s above-noted FRACTAL ANTENNA AND FRACTAL RESONATORS patent application provided a design methodology to produce smaller-scale antennas that exhibit at least as much gain, directivity, and efficiency as larger Euclidean counterparts. Such design approach should exploit the multiple scale self-similarity of real fractals, including $N \geqq 2$ iteration order fractals. Further, said application disclosed a non-Euclidean resonator whose presence in a resonating configuration can create frequencies of resonance beyond those normally presented in series and/or parallel LC configurations. Applicant’s above-noted TUNING FRACTAL ANTENNAS AND FRACTAL RESONATORS patent application provided devices and methods for tuning and/or adjusting such antennas and resonators. Said application further disclosed the use of non-Euclidean resonators whose presence in a resonating configuration could create frequencies of resonance beyond those normally presented in series and/or parallel LC configurations.

However, such antenna design approaches and tuning approaches should also be useable with vertical antennas, permitting the downscaling of one or more radial ground plane elements, and/or ground planes, and/or ground counterpoises, and/or top-hat loading elements.

The present invention provides such antennas, radial ground plane elements, ground planes, ground counterpoises, and top-hat loading elements, as well as methods for their design.

SUMMARY OF THE INVENTION

The present invention provides an antenna with a ground plane or ground counterpoise system that has at least one element whose shape, at least is part, is substantially a deterministic fractal of iteration order $N \geqq 1$. (The term "ground counterpoise" will be understood to include a ground plane, and/or at least one ground element.) Using fractal geometry, the antenna ground counterpoise has a self-similar structure resulting from the repetition of a design or motif (or "generator") that is replicated using rotation, and/or translation, and/or scaling. The fractal element will have x-axis, y-axis coordinates for a next iteration N+1 defined by $x_{N+1}=f(x_N, yb_N)$ and $y_{N+1}=g(x_N, y_N)$, where x_N, y_N define coordinates for a preceding iteration, and where f(x,y) and g(x,y) are functions defining the fractal motif and behavior. In another aspect, a vertical antenna is top-loaded with a so-called top-hat assembly that includes at least one fractal element. A fractalized top-hat assembly advantageously reduces resonant frequency, as well as the physical size and area required for the top-hat assembly.

In contrast to Euclidean geometric antenna design, deterministic fractal elements according to the present invention have a perimeter that is not directly proportional to area. For a given perimeter dimension, the enclosed area of a multi-iteration fractal will always be as small or smaller than the area of a corresponding conventional Euclidean element.

A fractal antenna has a fractal ratio limit dimension D given by log(L)/log(r), where L and r are one-dimensional antenna element lengths before and after fractalization, respectively.

As used herein, a fractal antenna perimeter compression parameter (PC) is defined as:

$$PC = \frac{\text{full-sized antenna element length}}{\text{fractal-reduced antenna element length}}$$

where:

$$PC = A \cdot \log[N(D+C)]$$

in which A and C are constant coefficients for a given fractal motif, N is an iteration number, and D is the fractal dimension, defined above.

Radiation resistance (R) of a fractal antenna decreases as a small power of the perimeter compression (PC), with a fractal loop or island always exhibiting a substantially higher radiation resistance than a small Euclidean loop antenna of equal size. In the present invention, deterministic fractals are used wherein A and C have large values, and thus provide the greatest and most rapid element-size shrinkage. A fractal antenna according to the present invention will exhibit an increased effective wavelength.

The number of resonant nodes of a fractal loop-shaped antenna increases as the iteration number N and is at least as large as the number of resonant nodes of an Euclidean island with the same area. Further, resonant frequencies of a fractal antenna include frequencies that are not harmonically related.

An antenna including a fractal ground counterpoise according to the present invention is smaller than its Euclidean counterpart but provides at least as much gain and frequencies of resonance and provides a reasonable termination impedance at its lowest resonant frequency. Such an antenna system can exhibit non-harmonically frequencies of resonance, a low Q and resultant good bandwidth, acceptable standing wave ratio ("SWR"), and a radiation impedance that is frequency dependent, and high efficiencies.

With respect to vertical antennas, the present invention enables such antennas to be realized with a smaller vertical element, and/or with smaller ground counterpoise, e.g., ground plane radial elements, and/or ground plane. The ground counterpoise element(s) are fractalized with $N \geqq 1$. In a preferred embodiment, the vertical element is also a fractal system, preferably comprising first and second spaced-apart fractal elements.

A fractal antenna system having a fractal ground counterpoise and a fractal vertical preferably is tuned according to applicant's above-referenced TUNING FRACTAL ANTENNAS AND FRACTAL RESONATORS, by placing an active (or driven) fractal antenna or resonator a distance Δ from a second conductor. Such disposition of the antenna and second conductor advantageously lowers resonant frequencies and widens bandwidth for the fractal antenna. In some embodiments, the fractal antenna and second conductor are non-coplanar and λ is the separation distance therebetween, preferably $\leqq 0.05\ \lambda$ for the frequency of interest ($1/\lambda$). In other embodiments, the fractal antenna and second conductive element may be planar, in which case λ a separation distance, measured on the common plane. In another embodiment, an antenna is loaded with a fractal "top-hat" assembly, which can provide substantial reduction in antenna size.

The second conductor may in fact be a second fractal antenna of like or unlike configuration as the active antenna. Varying the distance Δ tunes the active antenna and thus the overall system. Further, if the second element, preferably a fractal antenna, is angularly rotated relative to the active antenna, resonant frequencies of the active antenna may be varied.

Providing a cut in the fractal antenna results in new and different resonant nodes, including resonant nodes having perimeter compression parameters, defined below, ranging from about three to ten. If desired, a portion of a fractal antenna may be cutaway and removed so as to tune the antenna by increasing resonance(s).

Tunable antenna systems with a fractal ground counterpoise need not be planar, according to the present invention. Fabricating the antenna system around a form such as a torroid ring, or forming the fractal antenna on a flexible substrate that is curved about itself results in field self-proximity that produces resonant frequency shifts. A fractal antenna and a conductive element may each be formed as a curved surface or even as a torroid-shape, and placed in sufficiently close proximity to each other to provide a useful tuning and system characteristic altering mechanism.

In the various embodiments, more than two elements may be used, and tuning may be accomplished by varying one or more of the parameters associated with one or more elements.

Other features and advantages of the invention will appear from the following description in which the preferred embodiments have been set forth in detail, in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1**A depicts a base element for an antenna or an inductor, according to the prior art;

FIG. **1**B depicts a triangular-shaped Koch fractal motif, according to the prior art;

FIG. **1**C depicts a second-iteration fractal using the motif of FIG. **1**B, according to the prior art;

FIG. **1**D depicts a third-iteration fractal using the motif of FIG. **1**B, according to the prior art;

FIG. **2**A depicts a base element for an antenna or an inductor, according to the prior art;

FIG. 2B depicts a rectangular-shaped Minkowski fractal motif, according to the prior art;

FIG. 2C depicts a second-iteration fractal using the motif of FIG. 2B, according to the prior art;

FIG. 2D depicts a fractal configuration including a third-order using the motif of FIG. 2B, as well as the motif of FIG. 1B, according to the prior art;

FIG. 3 depicts bent-vertical chaotic fractal antennas, according to the prior art;

FIG. 4A depicts a series L-C resonator, according to the prior art;

FIG. 4B depicts a distributed parallel L-C resonator, according to the prior art;

FIG. 5A depicts an Euclidean quad antenna system, according to the prior art;

FIG. 5B depicts a second-order Minkowski island fractal quad antenna, according to the present invention;

FIG. 6 depicts an ELNEC-generated free-space radiation pattern for an MI-2 fractal antenna, according to the present invention;

FIG. 7A depicts a Cantor-comb fractal dipole antenna, according to the present invention;

FIG. 7B depicts a torn square fractal quad antenna, according to the present invention;

FIG. 7C-1 depicts a second iteration Minkowski (MI-2) printed circuit fractal antenna, according to the present invention;

FIG. 7C-2 depicts a second iteration Minkowski (MI-2) slot fractal antenna, according to the present invention;

FIG. 7D depicts a deterministic dendrite fractal vertical antenna, according to the present invention; FIG. 7D-1A depicts a 0.25 λ vertical antenna with three 0.25 λ radial ground elements, according to the prior art;

FIG. 7D-1B depicts the gain pattern for the antenna of FIG. 7D-1A;

FIG. 7D-2A depicts a 0.25 λ vertical antenna with three fractal radial ground elements according to the present invention;

FIG. 7D-2B depicts the gain pattern for the antenna of FIG. 7D-2A;

FIG. 7D-3A depicts a “top-hat” loaded antenna, according to the prior art;

FIG. 7D-3B depicts the gain pattern for the antenna of FIG. 7D-3A;

FIG. 7D-4A depicts a ternary fractal “top-hat” loaded antenna, according the present invention;

FIG. 7D-4B depicts the gain pattern for the antenna of FIG. 7D-4A;

FIG. 7D-5 depicts an antenna having a fractal vertical element and fractal radial ground elements, according to the present invention;

FIG. 7E depicts a third iteration Minkowski island (MI-3) fractal quad antenna, according to the present invention;

FIG. 7F depicts a second iteration Koch fractal dipole, according to the present invention;

FIG. 7G depicts a third iteration dipole, according to the present invention;

FIG. 7H depicts a second iteration Minkowski fractal dipole, according to the present invention;

FIG. 7I depicts a third iteration multi-fractal dipole, according to the present invention;

FIG. 8A depicts a generic system in which a passive or active electronic system communicates using a fractal antenna, according to the present invention;

FIG. 8B depicts a communication system in which several fractal antennas including a vertical antenna with a fractal ground counterpoise are electronically selected for best performance, according to the present invention;

FIG. 8C depicts a communication system in which electronically steerable arrays of fractal antennas are electronically selected for best performance, according to the present invention;

FIG. 9A depicts fractal antenna gain as a function of iteration order N, according to the present invention;

FIG. 9B depicts perimeter compression PC as a function of iteration order N for fractal antennas, according to the present invention;

FIG. 10A depicts a fractal inductor for use in a fractal resonator, according to the present invention;

FIG. 10B depicts a credit card sized security device utilizing a fractal resonator, according to the present invention;

FIG. 11A depicts an embodiment in which a fractal antenna is spaced-apart a distance Δ from a conductor element to vary resonant properties and radiation characteristics of the antenna, according to the present invention;

FIG. 11B depicts an embodiment in which a fractal antenna is coplanar with a ground plane and is spaced-apart a distance Δ' from a coplanar passive parasitic element to vary resonant properties and radiation characteristics of the antenna, according to the present invention;

FIG. 12A depicts spacing-apart first and second fractal antennas a distance Δ to decrease resonance and create additional resonant frequencies for the active or driven antenna, according to the present invention;

FIG. 12B depicts relative angular rotation between spaced-apart first and second fractal antennas Δ to vary resonant frequencies of the active or driven antenna, according to the present invention;

FIG. 13A depicts cutting a fractal antenna or resonator to create different resonant nodes and to alter perimeter compression, according to the present invention;

FIG. 13B depicts forming a non-planar fractal antenna or resonator on a flexible substrate that is curved to shift resonant frequency, apparently due to self-proximity electromagnetic fields, according to the present invention;

FIG. 13C depicts forming a fractal antenna or resonator on a curved torroidal form to shift resonant frequency, apparently due to self-proximity electromagnetic fields, according to the present invention;

FIG. 14A depicts forming a fractal antenna or resonator in which the conductive element is not attached to the system coaxial or other feedline, according to the present invention;

FIG. 14B depicts a system similar to FIG. 14A, but demonstrates that the driven fractal antenna may be coupled to the system coaxial or other feedline at any point along the antenna, according to the present invention;

FIG. 14C depicts an embodiment in which a supplemental ground plane is disposed adjacent a portion of the driven fractal antenna and conductive element, forming a sandwich-like system, according to the present invention;

FIG. 14D depicts an embodiment in which a fractal antenna system is tuned by cutting away a portion of the driven antenna, according to the present invention;

FIG. 15 depicts a communication system similar to that of FIG. 8A, in which several fractal antennas are tunable and are electronically selected for best performance, according to the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

In overview, the present invention provides an antenna system with a fractal ground counterpoise, e.g., a counterpoise and/or ground plane and/or ground element having at least one element whose shape, at least is part, is substantially a fractal of iteration order $N \geqq 1$. The resultant antenna is smaller than its Euclidean counterpart, provides close to 50 Ω termination impedance, exhibits at least as much gain and more frequencies of resonance than its Euclidean counterpart, including non-harmonically related frequencies of resonance, exhibits a low Q and resultant good bandwidth, acceptable SWR, a radiation impedance that is frequency dependent, and high efficiencies.

In contrast to Euclidean geometric antenna design, a fractal antenna ground counterpoise according to the present invention has a perimeter that is not directly proportional to area. For a given perimeter dimension, the enclosed area of a multi-iteration fractal area will always be at least as small as any Euclidean area. Using fractal geometry, the ground element has a self-similar structure resulting from the repetition of a design or motif (or "generator"), which motif is replicated using rotation, translation, and/or scaling (or any combination thereof). The fractal portion of the element has x-axis, y-axis coordinates for a next iteration N+1 defined by $x_{N+1}=f(x_N, yb_N)$ and $Y_{N+1}=g(x_N, y_N)$, where x_N, y_N are coordinates of a preceding iteration, and where f(x,y) and g(x,y) are functions defining the fractal motif and behavior. For example, fractals of the Julia set may be represented by the form:

$$x_{N+1}=x_N^2-y_N^2+a$$

$$y_{N+1}=2x_N \cdot y_N=b$$

In complex notation, the above may be represented as:

$$z_{N+1}=z_N^2+c$$

Although it is apparent that fractals can comprise a wide variety of forms for functions f(x,y) and g(x,y), it is the iterative nature and the direct relation between structure or morphology on different size scales that uniquely distinguish f(x,y) and g(x,y) from non-fractal forms. Many references including the Lauwerier treatise set forth equations appropriate for f(x,y) and g(x,y).

Iteration (N) is defined as the application of a fractal motif over one size scale. Thus, the repetition of a single size scale of a motif is not a fractal as that term is used herein. Multi-fractals may of course be implemented, in which a motif is changed for different iterations, but eventually at least one motif is repeated in another iteration.

An overall appreciation of the present invention may be obtained by comparing FIGS. **5**A and **5**B. FIG. **5**A shows a conventional Euclidean quad antenna **5** having a driven element **10** whose four sides are each 0.25 λ long, for a total perimeter of 1 λ, where λ is the frequency of interest.

Euclidean element **10** has an impedance of perhaps 130 Ω, which impedance decreases if a parasitic quad element **20** is spaced apart on a boom **30** by a distance B of 0.1 λ to 0.25 λ. Parasitic element **20** is also-sized S=0.25 λ on a side, and its presence can improve directivity of the resultant two-element quad antenna. Element **10** is depicted in FIG. **5**A with heavier lines than element **20**, solely to avoid confusion in understanding the figure. Non-conductive spreaders **40** are used to help hold element **10** together and element **20** together.

Because of the relatively large drive impedance, driven element **10** is coupled to an impedance matching network or device **60**, whose output impedance is approximately 50 Ω. A typically 50 Ω coaxial cable **50** couples device **60** to a transceiver **70** or other active or passive electronic equipment **70**.

As used herein, the term transceiver shall mean a piece of electronic equipment that can transmit, receive, or transmit and receive an electromagnetic signal via an antenna, such as the quad antenna shown in FIGS. **5**A or **5**B. As such, the term transceiver includes without limitation a transmitter, a receiver, a transmitter-receiver, a cellular telephone, a wireless telephone, a pager, a wireless computer local area network ("LAN") communicator, a passive resonant unit used by stores as part of an anti-theft system in which transceiver **70** contains a resonant circuit that is blown or not-blown by an electronic signal at time of purchase of the item to which transceiver **70** is affixed, resonant sensors and transponders, and the like.

Further, since antennas according to the present invention can receive incoming radiation and coupled the same as alternating current into a cable, it will be appreciated that fractal antennas may be used to intercept incoming light radiation and to provide a corresponding alternating current. For example, a photocell antenna defining a fractal, or indeed a plurality or array of fractals, would be expected to output more current in response to incoming light than would a photocell of the same overall array size. FIG. **5**B depicts a fractal quad antenna **95**, designed to resonant at the same frequency as the larger prior art antenna **5** shown in FIG. **5**A. Driven element **100** is seen to be a second order fractal, here a so-called Minkowski island fractal, although any of numerous other fractal configurations could instead be used, including without limitation, Koch, torn square, Mandelbrot, Caley tree, monkey's swing, Sierpinski gasket, and Cantor gasket geometry.

If one were to measure to the amount of conductive wire or conductive trace comprising the perimeter of element **10** or element **20**, it would be perhaps 40% greater than the 1.0 λ for the Euclidean quad of FIG. **5**A. However, for fractal antenna **95**, the physical straight length of one element side KS will be substantially smaller, and for the N=2 fractal antenna shown in FIG. **5**B, KS≈0.13 λ (in air), compared with K≈0.25 λ for prior art antenna **5**.

However, although the actual perimeter length of element **100** is greater than the 1 λ perimeter of prior art element **10**, the area within antenna element **100** is substantially less than the S^2 area of prior art element **10**. As noted, this area independence from perimeter is a characteristic of a deterministic fractal. Boom length B for antenna **95** will be slightly different from length B for prior art antenna **5** shown in FIG. **5**A. In FIG. **5**B, a parasitic element **120**, which preferably is similar to driven element **100** but need not be, may be attached to boom **130**. For ease of illustration FIG. **5**B does not depict non-conductive spreaders, such as spreaders **40** shown in FIG. **5**A, which help hold element **100** together and element **120** together. Further, for ease of understanding the figure, element **10** is drawn with heavier lines than element **120**, to avoid confusion in the portion of the figure in which elements **100** and **120** appear overlapped.

An impedance matching device **60** is advantageously unnecessary for the fractal antenna of FIG. **5**B, as the driving impedance of element **100** is about 50 Ω, e.g., a perfect match for cable **50** if reflector element **120** is absent, and about 35 Ω, still an acceptable impedance match for cable **50**, if element **120** is present. Antenna **95** may be fed by cable **50** essentially anywhere in element **100**, e.g., including locations X, Y, Z, among others, with no substantial change in the termination impedance. With cable **50** connected as shown, antenna **95** will exhibit horizontal polarization. If vertical polarization is desired, connection may be made as shown by cable **50**'. If desired, cables **50** and **50**' may both be present, and an electronic switching device **75** at the antenna end of these cables can short-out one of the cables. If cable **50** is shorted out at the antenna, vertical polarization results, and if instead cable **50**' is shorted out at the antenna, horizontal polarization results.

As shown by Table 3 herein, fractal quad **95** exhibits about 1.5 dB gain relative to Euclidean quad **10**. Thus, transmitting power output by transceiver **70** may be cut by perhaps 40% and yet the system of FIG. **5**B will still perform no worse than the prior art system of FIG. **5**A. Further, as shown by Table 1, the fractal antenna of FIG. **5**B exhibits more resonance frequencies than the antenna of FIG. **5**B, and also exhibits some resonant frequencies that are not harmonically related to each other. As shown by Table 3, antenna **95** has efficiency exceeding about 92% and exhibits an excellent SWR of about 1.2:1. As shown by Table 5, applicant's fractal quad antenna exhibits a relatively low value of Q. This result is surprising in view of conventional prior art wisdom to the effect that small loop antennas will exhibit high Q.

In short, that fractal quad **95** works at all is surprising in view of the prior art (mis)understanding as to the nature of radiation resistance R and ohmic losses O. Indeed, the prior art would predict that because the fractal antenna of FIG. **5**B is smaller than the conventional antenna of FIG. **5**A, efficiency would suffer due to an anticipated decrease in radiation resistance R. Further, it would have been expected that Q would be unduly high for a fractal quad antenna.

FIG. **6** is an ELNEC-generated free-space radiation pattern for a second-iteration Minkowski fractal antenna, an antenna similar to what is shown in FIG. **5**B with the parasitic element **120** omitted. The frequency of interest was 42.3 MHz, and a 1.5:1 SWR was used. In FIG. **6**, the outer ring represents 2.091 dBi, and a maximum gain of 2.091 dBi. (ELNEC is a graphics/PC version of MININEC, which is a PC version of NEC.) In practice, however, the data shown in FIG. **6** were conservative in that a gain of 4.8 dB above an isotropic reference radiator was actually obtained. The error in the gain figures associated with FIG. **6** presumably is due to roundoff and other limitations inherent in the ELNEC program. Nonetheless, FIG. **6** is believed to accurately depict the relative gain radiation pattern of a single element Minkowski (MI-2) fractal quad according to the present invention.

FIG. **7**A depicts a third iteration Cantor-comb fractal dipole antenna, according to the present invention. Generation of a Cantor-comb involves trisecting a basic shape, e.g., a rectangle, and providing a rectangle of one-third of the basic shape on the ends of the basic shape. The new smaller rectangles are then trisected, and the process repeated. FIG. **7**B is modelled after the Lauwerier treatise, and depicts a single element torn-sheet fractal quad antenna.

As described later herein, the fractal element shown in FIG. **7**B may be used as a ground counterpoise for an antenna system, for example, for a vertical antenna. In such application, the center conductor of cable **50** would be coupled to the lower end of the vertical antenna element (not shown, but which itself may be a fractal), and the ground shield of cable **50** would be coupled to the fractal element shown in FIG. **7**B. The fractal groundpoise may be substantially smaller than a conventional 0.25 λ ground system, without detriment to gain, coupling impedance, and vertical polarization characteristics of the antenna system.

FIG. **7C-1** depicts a printed circuit antenna, in which the antenna is fabricated using printed circuit or semiconductor fabrication techniques. For ease of understanding, the etched-away non-conductive portion of the printed circuit board **150** is shown cross-hatched, and the copper or other conductive traces **170** are shown without cross-hatching.

Applicant notes that while various corners of the Minkowski rectangle motif may appear to be touching in this and perhaps other figures herein, in fact no touching occurs. Further, it is understood that it suffices if an element according to the present invention is substantially a fractal. By this it is meant that a deviation of less than perhaps 10% from a perfectly drawn and implemented fractal will still provide adequate fractal-like performance, based upon actual measurements conducted by applicant.

The substrate **150** is covered by a conductive layer of material **170** that is etched away or otherwise removed in areas other than the fractal design, to expose the substrate **150**. The remaining conductive trace portion **170** defines a fractal antenna, a second iteration Minkowski slot antenna in FIG. **7C-1**. Substrate **150** may be a silicon wafer, a rigid or a flexible plastic-like material, perhaps Mylar™ material, or the non-conductive portion of a printed circuit board. Overlayer **170** may be deposited doped polysilicon for a semiconductor substrate **150**, or copper for a printed circuit board substrate.

If desired, the fractal structure shown in FIG. **7C-1** could be utilized as a fractal ground counterpoise for an antenna system, for example a vertical antenna. The fractal ground counterpoise may be fabricated using smaller dimensions than a conventional prior art system employing typically 0.25 λ ground radials or elements. If the structure shown in FIG. **7C-1** is used as a ground counterpoise, the center lead of cable **50** would be coupled to the vertical element (not shown), and the ground shield would be coupled to the fractal structure shown.

FIG. **7C-2** depicts a slot antenna version of what was shown in FIG. **7C-2**, wherein the conductive portion **170** (shown cross-hatched in FIG. **7C-2**) surrounds and defines a fractal-shape of non-conductive substrate **150**. Electrical connection to the slot antenna is made with a coaxial or other cable **50**, whose inner and outer conductors make contact as shown.

In FIGS. **7C-1** and **7C-2**, the substrate or plastic-like material in such constructions can contribute a dielectric effect that may alter somewhat the performance of a fractal antenna by reducing resonant frequency, which increases perimeter compression PC.

Those skilled in the art will appreciate that by virtue of the relatively large amount of conducting material (as contrasted to a thin wire), antenna efficiency is promoted in a slot configuration. Of course a printed circuit board or substrate-type construction could be used to implement a non-slot fractal antenna, e.g, in which the fractal motif is fabricated as a conductive trace and the remainder of the conductive material is etched away or otherwise removed. Thus, in FIG. **7C-1**, for example, if the cross-hatched surface now represents non-conductive material, and the non-cross hatched material represents conductive material, a printed circuit board or substrate-implemented wire-type fractal antenna results.

Printed circuit board and/or substrate-implemented fractal antennas are especially useful at frequencies of 80 MHz or higher, whereat fractal dimensions indeed become small. A 2 M MI-3 fractal antenna (e.g., FIG. **7**E) will measure about 5.5" (14 cm) on a side KS, and an MI-2 fractal antenna (e.g., FIG. **5**B) will be about 7" (17.5 cm) per side KS. As will be seen from FIG. **9**A, an MI-3 antenna suffers a slight loss in gain relative to an MI-2 antenna, but offers substantial size reduction.

Applicant has fabricated an MI-2 Minkowski island fractal antenna for operation in the 850–900 MHz cellular telephone band. The antenna was fabricated on a printed circuit board and measured about 1.2" (3 cm) on a side KS. The antenna was sufficiently small to fit inside applicant's cellular telephone, and performed as well as if the normal attachable "rubber-ducky" whip antenna were still attached. The antenna was found on the side to obtain desired vertical polarization, but could be fed anywhere on the element with 50 Ω impedance still being inherently present. Applicant also fabricated on a printed circuit board an MI-3 Minkowski island fractal quad, whose side dimension KS was about 0.8" (2 cm), the antenna again being inserted inside the cellular telephone. The MI-3 antenna appeared to work as well as the normal whip antenna, which was not attached. Again, any slight gain loss in going from MI-2 to MI-3 (e.g., perhaps 1 dB loss relative to an MI-0 reference quad, or 3 dB los relative to an MI-2) is more than offset by the resultant shrinkage in size. At satellite telephone frequencies of 1650 MHz or so, the dimensions would be approximately halved again. FIGS. **8**A, **8**B and **8**C depict preferred embodiments for such antennas.

FIG. **7**D depicts a 2 M dendrite deterministic fractal antenna that includes a slight amount of randomness. The vertical arrays of numbers depict wavelengths relative to 0 λ, at the lower end of the trunk-like element **200**. Eight radial-like elements **210** are disposed at 1.0 λ, and various other elements are disposed vertically in a plane along the length of element **200**. The antenna was fabricated using 12 gauge copper wire and was found to exhibit a surprising 20 dBi gain, which is at least 10 dB better than any antenna twice the size of what is shown in FIG. **7**D. Although superficially the vertical of FIG. **7**D may appear analogous to a log-periodic antenna, a fractal vertical according to the present invention does not rely upon an opening angle, in stark contrast to prior art log periodic designs.

FIGS. **7**D-**1**A and **7**D-**1**B depict a conventional vertical antenna **5**, comprising a 0.25 λ long vertical element **195**, and three 0.25 λ long ground plane radials **205**. Antenna **5** is fed using coaxial cable **50** in conventional fashion, the antenna impedance being on the order of about 24 Ω. Antenna efficiency may be improved by adding additional radial elements **205**, however doing so frequently requires more space than is conveniently available. In other configurations, a ground plane or counterpoise may be used without radials, e.g., earth or the metal body of an automobile in the case of a vehicular-mounted antenna. The 0° elevation angle azimuth plot of FIG. **7**D-**1**B depicts the undesirably large horizontal polarization components (the "figure eight" pattern) exhibited by this prior art vertical system, with vertical and total gain being about 1.45 dBi.

FIG. **7**D-**2**A depicts an antenna system **5** according to the present invention as including a vertical element **195** and a fractalized ground counterpoise system comprising, in this example, three dendrite fractal ground radials **215**. The ground radials are coupled to the ground shield on cable **50**, whereas the center lead of cable **50** is coupled to the vertical element **195**. Of course, other fractal configurations may be used instead, and a different number of ground radials may also be used.

In the azimuth plot of FIG. **7**D-**2**B, the elevation angle is 0°, and each fractal ground radial element is only about 0.087 λ. The maximum gain, at the outermost ring in the figure, is 1.83 dBi and the input impedance is about 30 Ω. Note in FIG. **7**D-**2**B that relatively little energy is radiated horizontally, and nearly all of the energy is radiated vertically, a desirable characteristic for a vertical antenna. It will be appreciated that the 0.087 λ dimensions of fractal ground plane elements **215** are substantially physically smaller than the 0.25 λ elements **205** in the prior art system of FIG. **7**D-**1**A. However, the radiation pattern for the system of FIG. **7**D-**2**A is actually better than that of the larger prior art system.

FIG. **7**D-**3**A depicts a so-called "top-hat" loaded vertical antenna **5**, according to the prior art. Antenna **5** includes a vertical element **195** and, in the example shown, a top-hat assembly comprising three spokes **207** located at the antenna top. The antenna is fed in conventional fashion with coaxial cable **50**. FIG. **7**D-**3**B depicts the radiation pattern for the conventional top-hat loaded antenna of FIG. **7**D-**3**A.

FIG. **7**D-**4**A depicts a "top-hat" antenna **5** that includes a vertical element **195** whose top is loaded by a top-hat assembly including fractalized radial spokes **215**. Antenna **5** may be fed in conventional fashion by coaxial cable **50**. For the same vertical length of element **195** as was used in FIG. **7**D-**3**A, the use of fractal radial spokes **215** advantageously decreases resonant frequency by 20%. In addition, the size of the "top-hat" assembly may be reduced by about 20%, and the area required for the "top-hat" assembly may be reduced by about 35%. These reductions are advantageous in that the fractalized top-hat antenna of FIG. **7**D-**4**A can require less material to fabricate, thus reducing manufacturing cost, weight, and wind resistance, relative to a prior art top-hat configuration. According to the present invention, it suffices if at least one of the elements in the top-hat assembly has a physical shaped defined at least in part by a fractal. Of course, more or less than three spokes may be used, and other fractal configurations may also be used, including combinations of fractal and non-fractal elements, as well as different types of fractal elements.

FIG. **7**D-**4**B represents the radiation pattern for the fractalized top-hat antenna of FIG. **7**D-**4**A. A comparison of FIGS. **7**D-**4**B and **7**D-**3**B confirms that there is no real performance penalty associated with using the fractalized configuration. Thus, the above-noted savings in cost, weight, and wind resistance are essentially penalty free.

FIG. **7**D-**5** depicts an antenna system according to the present invention, in which fractal ground elements **215** and a fractal vertical element **197** are both used. Fractal antenna elements **215** are preferably about 0.087 λ, and element **197** is about λ/12. Fractal vertical element **197** preferably comprises a pair of spaced-apart elements such as generally described with respect to FIGS. **11**A, **12**A, **12**B, **13**B, **14**A, **14**B, and **14**C. It is to be understood, however, that the salient feature of element **197** in FIG. **7**D-**3** is not its specific shape, but rather that it defines a fractal, and preferably a pair of spaced-apart fractal elements. It is solely for ease of illustration that the fractal elements shown in FIGS. **7**D-**3**, **11**A, **12**A, **12**B, **13**B, **14**A, **14**B, **14**C, and **14**D are similarly drawn. Further, the fractal—fractal antenna system shown in FIG. **7**D-**3** is preferably tuned by varying the spaced-apart distance Δ, and/or by rotating the spaced-apart elements relative to one another, and/or by forming a "cut" in an element, as described hereinafter with respect to various of FIGS. **11**A, **12**A, **12**B, **13**B, **14**A, **14**B, **14**C and **14**D.

FIG. **7**E depicts a third iteration Minkowski island quad antenna (denoted herein as MI-3). The orthogonal line segments associated with the rectangular Minkowski motif make this configuration especially acceptable to numerical study using ELNEC and other numerical tools using moments for estimating power patterns, among other modelling schemes. In testing various fractal antennas, applicant formed the opinion that the right angles present in the Minkowski motif are especially suitable for electromagnetic frequencies.

With respect to the MI-3 fractal of FIG. 7E, applicant discovered that the antenna becomes a vertical if the center led of coaxial cable **50** is connected anywhere to the fractal, but the outer coaxial braid-shield is left unconnected at the antenna end. (At the transceiver end, the outer shield is connected to ground.) Not only do fractal antenna islands perform as vertical antennas when the center conductor of cable **50** is attached to but one side of the island and the braid is left ungrounded at the antenna, but resonance frequencies for the antenna so coupled are substantially reduced. For example, a 2" (5 cm) sized MI-3 fractal antenna resonated at 70 MHz when so coupled, which is equivalent to a perimeter compression PC≈20.

FIG. **7**F depicts a second iteration Koch fractal dipole, and FIG. **7**G a third iteration dipole. FIG. **7**H depicts a second iteration Minkowski fractal dipole, and FIG. **7**I a third iteration multi-fractal dipole. Depending upon the frequencies of interest, these antennas may be fabricated by bending wire, or by etching or otherwise forming traces on a substrate. Each of these dipoles provides substantially 50 Ω termination impedance to which coaxial cable **50** may be directly coupled without any impedance matching device. It is understood in these figures that the center conductor of cable **50** is attached to one side of the fractal dipole, and the braid outer shield to the other side.

A fractal ground counterpoise may be fabricated using fractal element as shown in any (or all) of FIGS. 7E–7I. Thus, in FIGS. 7D-2A and 7D-**3**, fractal ground radial elements **215** are understood to depict any fractal of iteration order N≧1. Further, such fractals may, but need not be, defined by an opening angle.

FIG. **8**A depicts a generalized system in which a transceiver **500** is coupled to a fractal antenna system **510** to send electromagnetic radiation **520** and/or receive electromagnetic radiation **540**. A second transceiver **600** shown equipped with a conventional whip-like vertical antenna **610** also sends electromagnetic energy **630** and/or receives electromagnetic energy **540**.

Fractal antenna system **510** may include a fractal ground counterpoise and/or fractal antenna element, as described earlier herein. As noted in the case of a vertical antenna element, the overall size of the resulting antenna system is substantially smaller than what may be achieved with a prior art ground counterpoise system. Further, the fractal ground counterpoise system may be fabricated on a flexible substrate that is rolled or otherwise formed to fit within a case such as contains transceiver **500**. The resultant antenna ground system exhibits improved efficiency and power distribution pattern relative to a prior art system that may somehow be fit into an equivalent amount of area.

If transceivers **500, 600** are communication devices such as transmitter-receivers, wireless telephones, pagers, or the like, a communications repeating unit such as a satellite **650** and/or a ground base repeater unit **660** coupled to an antenna **670**, or indeed to a fractal antenna according to the present invention, may be present.

Alteratively, antenna **510** in transceiver **500** could be a passive LC resonator fabricated on an integrated circuit microchip, or other similarly small sized substrate, attached to a valuable item to be protected. Transceiver **600,** or indeed unit **660** would then be an electromagnetic transmitter outputting energy at the frequency of resonance, a unit typically located near the cash register checkout area of a store or at an exit. Depending upon whether fractal antenna-resonator **510** is designed to "blow" (e.g., become open circuit) or to "short" (e.g., become a close circuit) in the transceiver **500** will or will not reflect back electromagnetic energy **540** or **630** to a receiver associated with transceiver **600**. In this fashion, the unauthorized relocation of antenna **510** and/or transceiver **500** can be signalled by transceiver **600**.

FIG. **8**B depicts a transceiver **500** equipped with a plurality of fractal antennas, here shown as **510A, 510B, 510C** and **510D** coupled by respective cables **50A, 50B, 50C, 50D** to electronics **600** within unit **500**. In the embodiment shown, one of more of these antenna elements is are fabricated on a conformal, flexible substrate **150**, e.g., Mylar™ material or the like, upon which the antennas per se may be implemented by printing fractal patterns using conductive ink, by copper deposition, among other methods including printed circuit board and semiconductor fabrication techniques. A flexible such substrate may be conformed to a rectangular, cylindrical or other shape as necessary.

In the embodiment of FIG. **8**B, unit **500** is a handheld transceiver, and antennas **510A, 510B, 510C, 510D** preferably are fed for vertical polarization, as shown. Element **510D** may, for example, be a fractal ground counterpoise system for a vertical antenna element, shown in phantom as element **193** (which element may itself be a fractal to further reduce dimensions).

An electronic circuit **615** is coupled by cables **50A, 50B, 50C** to the antennas, and samples incoming signals to discern which fractal antenna system, e.g., **510A, 510B, 510C, 510D** is presently most optimally aligned with the transmitting station, perhaps a unit **600** or **650** or **660** as shown in FIG. **8**A. This determination may be made by examining signal strength from each of the antennas. An electronic circuit **620** then selects the presently best oriented antenna, and couples such antenna to the input of the receiver and output of the transmitter portion, collectively **630,** of unit **500**. It is understood that the selection of the best antenna is dynamic and can change as, for example, a user of **500** perhaps walks about holding the unit, or the transmitting source moves, or due to other changing conditions. In a cellular or a wireless telephone application, the result is more reliable communication, with the advantage that the fractal antennas can be sufficiently small-sized as to fit totally within the casing of unit **500**. Further, if a flexible substrate is used, the antennas may be wrapped about portions of the internal casing, as shown.

An additional advantage of the embodiment of FIG. **8**B is that the user of unit **500** may be physically distanced from the antennas by a greater distance that if a conventional external whip antenna were used. Although medical evidence attempting to link cancer with exposure to electromagnetic radiation from handheld transceivers is still inconclusive, the embodiment of FIG. **8**B appears to minimize any such risk. Although FIG. **8**B depicts a vertical antenna **193** and a fractal ground counterpoise **510D**, it is understood that antenna **193** could represent a cellular antenna on a motor vehicle, the groundpoise for which is fractal unit **510D**. Further, as noted, vertical element **193** may itself be a fractal.

FIG. **8**C depicts yet another embodiment wherein some or all of the antenna systems **510A, 510B, 510C** may include electronically steerable arrays, including arrays of fractal antennas of differing sizes and polarization orientations. Antenna system **510C**, for example may include similarly designed fractal antennas, e.g., antenna F-**3** and F-**4**, which are differently oriented from each other. Other antennas within system **510C** may be different in design from either of F-**3**, F-**4**. Fractal antenna F-**1** may be a dipole for example. Leads from the various antennas in system **510C** may be coupled to an integrated circuit **690**, mounted on substrate

150. Circuit **690** can determine relative optimum choice between the antennas comprising system **510**C, and output via cable **50**C to electronics **600** associated with the transmitter and/or receiver portion **630** of unit **630**. Of course, the embodiment of FIG. **8**C could also include the vertical antenna element **193** and fractal ground counterpoise **510**D, depicted in FIG. **8**B.

Another antenna system **510**B may include a steerable array of identical fractal antennas, including fractal antenna F-**5** and F-**6**. An integrated circuit **690** is coupled to each of the antennas in the array, and dynamically selects the best antenna for signal strength and coupled such antenna via cable **50**B to electronics **600**. A third antenna system **510**A may be different from or identical to either of system **510**B and **510**C.

Although FIG. **8**C depicts a unit **500** that may be handheld, unit **500** could in fact be a communications system for use on a desk or a field mountable unit, perhaps unit **660** as shown in FIG. **8**A.

For ease of antenna matching to a transceiver load, resonance of a fractal antenna was defined as a total impedance falling between about 20 Ω to 200 Ω, and the antenna was required to exhibit medium to high Q, e.g., frequency/Δfrequency. In practice, applicants' various fractal antennas were found to resonate in at least one position of the antenna feedpoint, e.g., the point at which coupling was made to the antenna. Further, multi-iteration fractals according to the present invention were found to resonate at multiple frequencies, including frequencies that were non-harmonically related.

Contrary to conventional wisdom, applicant found that island-shaped fractals (e.g., a closed loop-like configuration) do not exhibit significant drops in radiation resistance R for decreasing antenna size. As described herein, fractal antennas were constructed. with dimensions of less than 12" across (30.48 cm) and yet resonated in a desired 60 MHz to 100 MHz frequency band.

Applicant further discovered that antenna perimeters do not correspond to lengths that would be anticipated from measured resonant frequencies, with actual lengths being longer than expected. This increase in element length appears to be a property of fractals as radiators, and not a result of geometric construction. A similar lengthening effect was reported by Pfeiffer when constructing a full-sized quad antenna using a first order fractal, see A. Pfeiffer, *The Pfeiffer Quad Antenna System*, QST, p. 28–32 (March 1994).

If L is the total initial one-dimensional length of a fractal pre-motif application, and r is the one-dimensional length post-motif application, the resultant fractal dimension D (actually a ratio limit) is:

$$D=\log(L)/\log(r)$$

With reference to FIG. **1**A, for example, the length of FIG. **1**A represents L, whereas the sum of the four line segments comprising the Koch fractal of FIG. **1**B represents r.

Unlike mathematical fractals, fractal antennas are not characterized solely by the ratio D. In practice D is not a good predictor of how much smaller a fractal design antenna may be because D does not incorporate the perimeter lengthening of an antenna radiating element.

Because D is not an especially useful predictive parameter in fractal antenna design, a new parameter "perimeter compression" ("PC") shall be used, where:

$$PC = \frac{\text{full-sized antenna element length}}{\text{fractal-reduced antenna element length}}$$

In above equation, measurements are made at the fractal-resonating element's lowest resonant frequency. Thus, for a full-sized antenna according to the prior art PC=1, while PC=3 represents a fractal antenna according to the present invention, in which an element side has been reduced by a factor of three.

Perimeter compression may be empirically represented using the fractal dimension D as follows:

$$PC=A\cdot\log[N(D+C)]$$

where A and C are constant coefficients for a given fractal motif, N is an iteration number, and D is the fractal dimension, defined above.

It is seen that for each fractal, PC becomes asymptotic to a real number and yet does not approach infinity even as the iteration number N becomes very large. Stated differently, the PC of a fractal radiator asymptotically approaches a non-infinite limit in a finite number of fractal iterations. This result is not a representation of a purely geometric fractal.

That some fractals are better resonating elements than other fractals follows because optimized fractal antennas approach their asymptotic PCs in fewer iterations than non-optimized fractal antennas. Thus, better fractals for antennas will have large values for A and C, and will provide the greatest and most rapid element-size shrinkage. Fractal used may be deterministic or chaotic. Deterministic fractals have a motif that replicates at a 100% level on all size scales, whereas chaotic fractals include a random noise component.

Applicant found that radiation resistance of a fractal antenna decreases as a small power of the perimeter compression (PC), with a fractal island always exhibiting a substantially higher radiation resistance than a small Euclidean loop antenna of equal size.

Further, it appears that the number of resonant nodes of a fractal island increase as the iteration number (N) and is always greater than or equal to the number of resonant nodes of an Euclidean island with the same area. Finally, it appears that a fractal resonator has an increased effective wavelength.

The above findings will now be applied to experiments conducted by applicant with fractal resonators shaped into closed-loops or islands. Prior art antenna analysis would predict no resonance points, but as shown below, such is not the case.

A Minkowski motif is depicted in FIGS. **2**B–**2**D, **5**B, **7**C and **7**E. The Minkowski motif selected was a three-sided box (e.g., **20-2** in FIG. **2**B) placed atop a line segment. The box sides may be any arbitrary length, e.g, perhaps a box height and width of 2 units with the two remaining base sides being of length three units (see FIG. **2**B). For such a configuration, the fractal dimension D is as follows:

$$D = \frac{\log(L)}{\log(r)} = \frac{\log(12)}{\log(8)} = \frac{1.08}{0.90} = 1.20$$

It will be appreciated that D=1.2 is not especially high when compared to other deterministic fractals.

Applying the motif to the line segment may be most simply expressed by a piecewise function f(x) as follows:

$$f(x)=0 \qquad 0 \geq x \geq \frac{3x_{\max}}{8}$$

$$f(x)=\frac{1}{4x_{\max}} \qquad \frac{3x_{\max}}{8} \geq x \geq \frac{5x_{\max}}{8}$$

$$f(x)=0 \qquad \frac{5x_{\max}}{8} \geq x \geq x_{\max}$$

where x_{max} is the largest continuous value of x on the line segment.

A second iteration may be expressed as $f(x)_2$ relative to the first iteration $f(x)_1$ by:

$$f(x)_2=f(x)_1+f(x)$$

where x_{max} is defined in the above-noted piecewise function. Note that each separate horizontal line segment will have a different lower value of x and x_{max}. Relevant offsets from zero may be entered as needed, and vertical segments may be "boxed" by 90° rotation and application of the above methodology.

As shown by FIGS. **5**B and **7**E, a Minkowski fractal quickly begins to appear like a Moorish design pattern. However, each successive iteration consumes more perimeter, thus reducing the overall length of an orthogonal line segment. Four box or rectangle-like fractals of the same iteration number N may be combined to create a Minkowski fractal island, and a resultant "fractalized" cubical quad.

An ELNEC simulation was used as a guide to far-field power patterns, resonant frequencies, and SWRs of Minkowski Island fractal antennas up to iteration N=2. Analysis for N>2 was not undertaken due to inadequacies in the test equipment available to applicant.

The following tables summarize applicant's ELNEC simulated fractal antenna designs undertaken to derive lowest frequency resonances and power patterns, to and including iteration N=2. All designs were constructed on the x,y axis, and for each iteration the outer length was maintained at 42" (106.7 cm).

Table 1, below, summarizes ELNEC-derived far field radiation patterns for Minkowski island quad antennas for each iteration for the first four resonances. In Table 1, each iteration is designed as MI-N for Minkowski Island of iteration N. Note that the frequency of lowest resonance decreased with the fractal Minkowski Island antennas, as compared to a prior art quad antenna. Stated differently, for a given resonant frequency, a fractal Minkowski Island antenna will be smaller than a conventional quad antenna.

TABLE 1

Antenna	Res. Freq. (MHz)	Gain (dBi)	SWR	PC (for 1st)	Direction
Ref. Quad	76	3.3	2.5	1	Broadside
	144	2.8	5.3	—	Endfire
	220	3.1	5.2	—	Endfire
	294	5.4	4.5	—	Endfire
MI-1	55	2.6	1.1	1.38	Broadside
	101	3.7	1.4	—	Endfire
	142	3.5	5.5	—	Endfire
	198	2.7	3.3	—	Broadside
MI-2	43.2	2.1	1.5	1.79	Broadside
	85.5	4.3	1.8	—	Endfire
	102	2.7	4.0	—	Endfire
	116	1.4	5.4	—	Broadside

It is apparent from Table 1 that Minkowski island fractal antennas are multi-resonant structures having virtually the same gain as larger, full-sized conventional quad antennas. Gain figures in Table 1 are for "free-space" in the absence of any ground plane, but simulations over a perfect ground at 1 λ yielded similar gain results. Understandably, there will be some inaccuracy in the ELNEC results due to round-off and undersampling of pulses, among other factors.

Table 2 presents the ratio of resonant ELNEC-derived-frequencies for the first four resonance nodes referred to in Table 1.

TABLE 2

Antenna	SWR	SWR	SWR	SWR
Ref. Quad (MI-0)	1:1	1:1.89	1:2.89	3.86:1
MI-1	1:1	1:1.83	1;2.58	3.6:1
MI-2	1:1	2.02:1	2.41:1	2.74:1

Tables 1 and 2 confirm the shrinking of a fractal-designed antenna, and the increase in the number of resonance points. In the above simulations, the fractal MI-2 antenna exhibited four resonance nodes before the prior art reference quad exhibited its second resonance. Near fields in antennas are very important, as they are combined in multiple-element antennas to achieve high gain arrays. Unfortunately, programming limitations inherent in ELNEC preclude serious near field investigation. However, as described later herein, applicant has designed and constructed several different high gain fractal arrays that exploit the near field.

Applicant fabricated three Minkowski Island fractal antennas from aluminum #8 and/or thinner #12 galvanized groundwire. The antennas were designed so the lowest operating frequency fell close to a desired frequency in the 2 M (144 MHz) amateur radio band to facilitate relative gain measurements using 2 M FM repeater stations. The antennas were mounted for vertical polarization and placed so their center points were the highest practical point above the mounting platform. For gain comparisons, a vertical ground plane having three reference radials, and a reference quad were constructed, using the same sized wire as the fractal antenna being tested. Measurements were made in the receiving mode.

Multi-path reception was minimized by careful placement of the antennas. Low height effects were reduced and free space testing approximated by mounting the antenna test platform at the edge of a third-store window, affording a 3.5 λ height above ground, and line of sight to the repeater, 45 miles (28 kM) distant. The antennas were stuck out of the window about 0.8 λ from any metallic objects and testing was repeated on five occasions from different windows on the same floor, with test results being consistent within ½ dB for each trial.

Each antenna was attached to a short piece of 9913 50 Ω coaxial cable, fed at right angles to the antenna. A 2 M transceiver was coupled with 9913 coaxial cable to two precision attenuators to the antenna under test. The transceiver S-meter was coupled to a volt-ohm meter to provide signal strength measurements The attenuators were used to insert initial threshold to avoid problems associated with non-linear S-meter readings, and with S-meter saturation in the presence of full squelch quieting.

Each antenna was quickly switched in for volt-ohmmeter measurement,with attenuation added or subtracted to obtain the same meter reading as experienced with the reference quad. All readings were corrected for SWR attenuation. For the reference quad, the SWR was 2.4:1 for 120 Ω impedance, and for the fractal quad antennas SWR was less than 1.5:1 at resonance. The lack of a suitable noise bridge for 2 M precluded efficiency measurements for the various antennas. Understandably, anechoic chamber testing would provide even more useful measurements.

For each antenna, relative forward gain and optimized physical orientation were measured. No attempt was made to correct for launch-angle, or to measure power patterns other than to demonstrate the broadside nature of the gain. Difference of ½ dB produced noticeable S-meter deflections, and differences of several dB produced substantial meter deflection. Removal of the antenna from the receiver resulted in a 20^+ dB drop in received signal strength. In this fashion, system distortions in readings were cancelled out to provide more meaningful results. Table 3 summarizes these results.

TABLE 3

Antenna	PC	PL	SWR	Cor. Gain (dB)	Sidelength (λ)
Quad	1	1	2.4:1	0	0.25
¼ wave	1	—	1.5:1	−1.5	0.25
MI-1	1.3	1.2	1.3:1	1.5	0.13
MI-2	1.9	1.4	1.3:1	1.5	0.13
MI-3	2.4	1.7	1:1	−1.2	0.10

It is apparent from Table 3 that for the vertical configurations under test, a fractal quad according to the present invention either exceeded the gain of the prior art test quad, or had a gain deviation of not more than 1 dB from the test quad. Clearly, prior art cubical (square) quad antennas are not optimized for gain. Fractally shrinking a cubical quad by a factor of two will increase the gain, and further shrinking will exhibit modest losses of 1–2 dB.

Versions of a MI-2 and MI-3 fractal quad antennas were constructed for the 6 M (50 MHz) radio amateur band. An RX 50 Ω noise bridge was attached between these antennas and a transceiver. The receiver was nulled at about 54 MHz and the noise bridge was calibrated with 5 Ω and 10 Ω resistors. Table 4 below summarizes the results, in which almost no reactance was seen.

TABLE 4

Antenna	SWR	Z (Ω)	O (Ω)	E (%)
Quad (MI-0)	2.4:1	120	5–10	92–96
MI-2	1.2:1	60	≦5	≧92
MI-3	1.1:1	55	≦5	≧91

In Table 4, efficiency (E) was defined as 100%*(R/Z), where Z was the measured impedance, and R was Z minus ohmic impedance and reactive impedances (0). As shown in Table 4, fractal MI-2 and MI-3 antennas with their low ≦1.2:1 SWR and low ohmic and reactive impedance provide extremely high efficiencies, 90^+%. These findings are indeed surprising in view of prior art teachings stemming from early Euclidean small loop geometries. In fact, Table 4 strongly suggests that prior art associations of low radiation impedances for small loops must be abandoned in general, to be invoked only when discussing small Euclidean loops. Applicant's MI-3 antenna was indeed micro-sized, being dimensioned at about 0.1 λ per side, an area of about $\lambda^2/1,000$, and yet did not signal the onset of inefficiency long thought to accompany smaller sized antennas.

However the 6M efficiency data do not explain the fact that the MI-3 fractal antenna had a gain drop of almost 3 dB relative to the MI-2 fractal antenna. The low ohmic impedances of ≦5 Ω strongly suggest that the explanation is other than inefficiency, small antenna size notwithstanding. It is quite possible that near field diffraction effects occur at higher iterations that result in gain loss. However, the smaller antenna sizes achieved by higher iterations appear to warrant the small loss in gain.

Using fractal techniques, however, 2 M quad antennas dimensioned smaller than 3" (7.6 cm) on a side, as well as 20 M (14 MHz) quads smaller than 3' (1 m) on a side can be realized. Economically of greater interest, fractal antennas constructed for cellular telephone frequencies (850 MHz) could be sized smaller than 0.5" (1.2 cm). As shown by FIGS. **8**B and **8**C, several such antenna, each oriented differently could be fabricated within the curved or rectilinear case of a cellular or wireless telephone, with the antenna outputs coupled to a circuit for coupling to the most optimally directed of the antennas for the signal then being received. The resultant antenna system would be smaller than the "rubber-ducky" type antennas now used by cellular telephones, but would have improved characteristics as well.

Similarly, fractal-designed antennas could be used in handheld military walkie-talkie transceivers, global positioning systems, satellites, transponders, wireless communication and computer networks, remote and/or robotic control systems, among other applications.

Although the fractal Minkowski island antenna has been described herein, other fractal motifs are also useful, as well as non-island fractal configurations.

Table 5 demonstrates bandwidths ("BW") and multi-frequency resonances of the MI-2 and MI-3 antennas described, as well as Qs, for each node found for 6 M versions between 30 MHz and 175 MHz. Irrespective of resonant frequency SWR, the bandwidths shown are SWR 3:1 values. Q values shown were estimated by dividing resonant frequency by the 3:1 SWR BW. Frequency ratio is the relative scaling of resonance nodes.

TABLE 5

Antenna	Freq. (MHz)	Freq. Ratio	SWR	3:1 BW	Q
MI-3	53.0	1	1:1	6.4	8.3
	80.1	1.5:1	1.1:1	4.5	17.8
	121.0	2.3:1	2.4:1	6.8	17.7
MI-2	54.0	1	1:1	3.6	15.0
	95.8	1.8:1	1.1:1	7.3	13.1
	126.5	2.3:1	2.4:1	9.4	13.4

The Q values in Table 5 reflect that MI-2 and MI-3 fractal antennas are multiband. These antennas do not display the very high Qs seen in small tuned Euclidean loops, and there appears not to exist a mathematical application to electromagnetics for predicting these resonances or Qs. One approach might be to estimate scalar and vector potentials in Maxwell's equations by regarding each Minkowski Island iteration as a series of vertical and horizontal line segments with offset positions. Summation of these segments will lead to a Pointing vector calculation and power pattern that may be especially useful in better predicting fractal antenna characteristics and optimized shapes.

In practice, actual Minkowski Island fractal antennas seem to perform slightly better than their ELNEC predictions, most likely due to inconsistencies in ELNEC modelling or ratios of resonant frequencies, PCs, SWRs and gains.

Those skilled in the art will appreciate that fractal multiband antenna arrays may also be constructed. The resultant arrays will be smaller than their Euclidean counterparts, will present less wind area, and will be mechanically rotatable with a smaller antenna rotator.

Further, fractal antenna configurations using other than Minkowski islands or loops may be implemented. Table 6 shows the highest iteration number N for other fractal configurations that were found by applicant to resonant on at least one frequency.

TABLE 6

Fractal	Maximum Iteration
Koch	5
Torn Square	4
Minkowski	3
Mandelbrot	4
Caley Tree	4
Monkey's Swing	3
Sierpinski Gasket	3
Cantor Gasket	3

FIG. **9**A depicts gain relative to an Euclidean quad (e.g., an MI-0) configuration as a function of iteration value N. (It is understood that an Euclidean quad exhibits 1.5 dB gain relative to a standard reference dipole.) For first and second order iterations, the gain of a fractal quad increases relative to an Euclidean quad. However, beyond second order, gain drops off relative to an Euclidean quad. Applicant believes that near field electromagnetic energy diffraction-type cancellations may account for the gain loss for N>2. Possibly the far smaller areas found in fractal antennas according to the present invention bring this diffraction phenomenon into sharper focus.

In practice, applicant could not physically bend wire for a 4th or 5th iteration 2 M Minkowski fractal antenna, although at lower frequencies the larger antenna sizes would not present this problem. However, at higher frequencies, printed circuitry techniques, semiconductor fabrication techniques as well as machine-construction could readily produce N=4, N=5, and higher order iterations fractal antennas.

In practice, a Minkowski island fractal antenna should reach the theoretical gain limit of about 1.7 dB seen for sub-wavelength Euclidean loops, but N will be higher than 3. Conservatively, however, an N=4 Minkowski Island fractal quad antenna should provide a PC=3 value without exhibiting substantial inefficiency.

FIG. **9**B depicts perimeter compression (PC) as a function of iteration order N for a Minkowski island fractal configuration. A conventional Euclidean quad (MI-0) has PC=1 (e.g., no compression), and as iteration increases, PC increases. Note that as N increases and approaches 6, PC approaches a finite real number asymptotically, as predicted. Thus, fractal Minkowski Island antennas beyond iteration N=6 may exhibit diminishing returns for the increase in iteration.

It will be appreciated that the non-harmonic resonant frequency characteristic of a fractal antenna according to the present invention may be used in a system in which the frequency signature of the antenna must be recognized to pass a security test. For example, at suitably high frequencies, perhaps several hundred MHz, a fractal antenna could be implemented within an identification credit card. When the card is used, a transmitter associated with a credit card reader can electronically sample the frequency resonance of the antenna within the credit card. If and only if the credit card antenna responds with the appropriate frequency signature pattern expected may the credit card be used, e.g., for purchase or to permit the owner entrance into an otherwise secured area.

FIG. **10**A depicts a fractal inductor L according to the present invention. In contrast to a prior art inductor, the winding or traces with which L is fabricated define, at least in part, a fractal. The resultant inductor is physically smaller than its Euclidean counterpart. Inductor L may be used to form a resonator, including resonators such as shown in FIGS. **4**A and **4**B. As such, an integrated circuit or other suitably small package including fractal resonators could be used as part of a security system in which electromagnetic radiation, perhaps from transmitter **600** or **660** in FIG. **8**A will blow, or perhaps not blow, an LC resonator circuit containing the fractal antenna. Such applications are described elsewhere herein and may include a credit card sized unit **700**, as shown in FIG. **10**B, in which an LC fractal resonator **710** is implemented. (Card **700** is depicted in FIG. **10**B as though its upper surface were transparent.).

The foregoing description has largely replicated what has been set forth in applicant's above-noted FRACTAL ANTENNAS AND FRACTAL RESONATORS patent application. The following section will set forth methods and techniques for tuning such fractal antennas and resonators. In the following description, although the expression "antenna" may be used in referring to a preferably fractal element, in practice what is being described is an antenna or filter-resonator system. As such, an "antenna" can be made to behave as through it were a filter, e.g., passing certain frequencies and rejecting other frequencies (or the converse).

In one group of embodiments, applicant has discovered that disposing a fractal antenna a distance Δ that is in close proximity (e.g., less than about 0.05 λ for the frequency of interest) from a conductor advantageously can change the resonant properties and radiation characteristics of the antenna (relative to such properties and characteristics when such close proximity does not exist, e.g., when the spaced-apart distance is relatively great. For example, in FIG. **11**A a conductive surface **800** is disposed a distance Δ behind or beneath a fractal antenna **810**, which in FIG. **11**A is a single arm of an MI-2 fractal antenna. Of course other fractal configurations such as disclosed herein could be used instead of the MI-1 configuration shown, and non-planar configurations may also be used. Fractal antenna **810** preferably is fed with coaxial cable feedline **50**, whose center conductor is attached to one end **815** of the fractal antenna, and whose outer shield is grounded to the conductive plane **800**. As described herein, great flexibility in connecting the antenna system shown to a preferably coaxial feedline exists. Termination impedance is approximately of similar magnitudes as described earlier herein.

In the configuration shown, the relative close proximity between conductive sheet **800** and fractal antenna **810** lowers the resonant frequencies and widens the bandwidth of antenna **810**. The conductive sheet **800** may be a plane of metal, the upper copper surface of a printed circuit board, a region of conductive material perhaps sprayed onto the housing of a device employing the antenna, for example the interior of a transceiver housing **500**, such as shown in FIGS. **8**A, **8**B, **8**C, and **15**.

The relationship between Δ, wherein $\Delta \leqq 0.05\ \lambda$, and resonant properties and radiation characteristics of a fractal antenna system is generally logarithmic. That is, resonant frequency decreases logarithmically with decreasing separation Δ.

FIG. **11**B shows an embodiment in which a preferably fractal antenna **810** lies in the same plane as a ground plane **800** but is separated therefrom by an insulating region, and in which a passive or parasitic element **800'** is disposed "within" and spaced-apart a distance Δ' from the antenna, and also being coplanar. For example, the embodiment of

FIG. **11**B may be fabricated from a single piece of printed circuit board material in which copper (or other conductive material) remains to define the groundplane **800**, the antenna **810**, and the parasitic element **800'**, the remaining portions of the original-material having been etched away to form the "moat-like" regions separating regions **800, 810**, and **800'**. changing the shape and/or size of element **800'** and/or the coplanar spaced-apart distance Δ' tunes the antenna system shown. For example, for a center frequency in the 900 MHz range, element **800'** measured about 63 mm×8 mm, and elements **810** and **800** each measured about 25 mm×12 mm. In general, element **800** should be at least as large as the preferably fractal antenna **810**. For this configuration, the system shown exhibited a bandwidth of about 200 MHz, and could be made to exhibit characteristics of a bandpass filter and/or band rejection filter. In this embodiment, a coaxial feedline **50** was used, in which the center lead was coupled to antenna **810**, and the ground shield lead was coupled to groundplane **800**. In FIG. **11**B, the inner perimeter of groundplane region **800** is shown as being rectangularly shaped. If desired, this inner perimeter could be moved closer to the outer perimeter of preferably fractal antenna **810**, and could in fact define a perimeter shape that follows the perimeter shape of antenna **810**. In such an embodiment, the perimeter of the inner conductive region **800'** and the inner perimeter of the ground plane region **800** would each follow the shape of antenna **810**. Based upon experiments to date, it is applicant's belief that moving the inner perimeter of ground plane region **800** sufficiently close to antenna **810** could also affect the characteristics of the overall antenna/resonator system.

Referring now to FIG. **12**A, if the conductive surface **800** is replaced with a second fractal antenna **810'**, which is spaced-apart a distance Δ that preferably does not exceed about 0.05 λ, resonances for the radiating fractal antenna **810** are lowered and advantageously new resonant frequencies emerge. For ease of fabrication, it may be desired to construct antenna **810** on the upper or first surface **820**A of a substrate **820**, and to construct antenna **810'** on the lower or second surface **820**B of the same substrate. The substrate could be doubled-side printed circuit board type material, if desired, wherein antennas **810, 810'** are fabricated using printed circuit type techniques. The substrate thickness Δ is selected to provide the desired performance for antenna **810** at the frequency of interest. Substrate **820** may, for example, be a non-conductive film, flexible or otherwise. To avoid cluttering FIGS. **12**A and **12**B, substrate **820** is drawn with phantom lines, as if the substrate were transparent.

As noted earlier, the fractal spaced-apart structure depicted in FIGS. **12**A and **12**B may instead be used to form a fractal element in a vertical antenna system, preferably including a fractal ground counterpoise, such as was described with respect to FIG. **8**D-**3**.

Preferably, the center conductor of coaxial cable **50** is connected to one end **815** of antenna **810**, and the outer conductor of cable **50** is connected to a free end **815'** of antenna **810'**, which is regarded as ground, although other feedline connections may be used. Although FIG. **12**A depicts antenna **810'** as being substantially identical to antenna **810**, the two antennas could in fact have different configurations.

Applicant has discovered that if the second antenna **810'** is rotated some angle θ relative to antenna **810**, the resonant frequencies of antenna **810** may be varied, analogously to tuning a variable capacitor. Thus, in FIG. **12**B, antenna **810** is tuned by rotating antenna **810'** relative to antenna **810** (or the converse, or by rotating each antenna). If desired, substrate **820** could comprise two substrates each having thickness Δ/2 and pivotally connected together, e.g., with a non-conductive rivet, so as to permit rotation of the substrates and thus relative rotation of the two antennas. Those skilled in the mechanical arts will appreciate that a variety of "tuning" mechanisms could be implemented to permit fine control over the angle θ in response, for example, to rotation of a tunable shaft.

Referring now to FIG. **13**A, applicant has discovered that creating at least one cut or opening **830** in a fractal antenna **810** (here comprising two legs of an MI-2 antenna) results in new and entirely different resonant nodes for the antenna. Further, these nodes can have perimeter compression (PC) ranging from perhaps three to about ten. The precise location of cut **830** on the fractal antenna or resonator does not appear to be critical.

FIGS. **13**B and **13**C depict a self-proximity characteristic of fractal antennas and resonators that may advantageously be used to create a desired frequency resonant shift. In FIG. **13**B, a fractal antenna **810** is fabricated on a first surface **820**A of a flexible substrate **820**, whose second surface **820**B does not contain an antenna or other conductor in this embodiment. Curving substrate **820**, which may be a flexible film, appears to cause electromagnetic fields associated with antenna **810** to be sufficiently in self-proximity so as to shift resonant frequencies. Such self-proximity antennas or resonators may be referred to a com-cyl devices. The extent of curvature may be controlled where a flexible substrate or substrate-less fractal antenna and/or conductive element is present, to control or tune frequency dependent characteristics of the resultant system. Com-cyl embodiments could include a concentrically or eccentrically disposed fractal antenna and conductive element. Such embodiments may include telescopic elements, whose extent of "overlap" may be telescopically adjusted by contracting or lengthening the overall configuration to tune the characteristics of the resultant system. Further, more than two elements could be provided.

In FIG. **13**C, a fractal antenna **810** is formed on the outer surface **820**A of a filled substrate **820**, which may be a ferrite core. The resultant com-cyl antenna appears to exhibit self-proximity such that desired shifts in resonant frequency are produced. The geometry of the core **820**, e.g., the extent of curvature (e.g., radius in this embodiment) relative to the size of antenna **810** may be used to determine frequency shifts.

In FIG. **14**A, an antenna or resonator system is shown in which the non-driven fractal antenna **810'** is not connected to the preferably coaxial feedline **50**. The ground shield portion of feedline **50** is coupled to the groundplane conductive element **800**, but is not otherwise connected to a system ground. of course fractal antenna **810'** could be angularly rotated relative to driven antenna **810**, it could be a different configuration than antenna **810** including having a different iteration N, and indeed could incorporate other features disclosed herein (e.g., a cut).

FIG. **14**B demonstrates that the driven antenna **810** may be coupled to the feedline **50** at any point **815'**, and not necessarily at an end point **815** as was shown in FIG. **14**A.

In the embodiment of FIG. **14**C, a second ground plane element **800'** is disposed adjacent at least a portion of the system comprising driven antenna **810**, passive antenna **810'**, and the underlying conductive planar element **800**. The presence, location, geometry, and distance associated with second ground plane element **800'** from the underlying elements **810, 810', 800** permit tuning characteristics of the overall antenna or resonator system. In the multi-element sandwich-like configuration shown, the ground shield of conductor **50** is connected to a system ground but not to either ground plane **800** or **800'**. Of course more than three elements could be used to form a tunable system according to the present invention.

FIG. **14**D shows a single fractal antenna spaced apart from an underlying ground plane **800** a distance Δ, in which a region of antenna **800** is cutaway to increase resonance. In FIG. **14**D, for example, L1 denotes a cutline, denoting that portions of antenna **810** above (in the Figure drawn) L1 are cutaway and removed. So doing will increase the frequencies of resonance associated with the remaining antenna or resonator system. On the other hand, if portions of antenna **810** above cutline L2 are cutaway and removed, still higher resonances will result. Selectively cutting or etching away portions of antenna **810** permit tuning characteristics of the remaining system.

As noted, fractal elements similar to what is generically depicted in FIGS. **14**A–**14**D may be used to form a fractal vertical element in a fractal vertical antenna system, such as was described with respect to FIG. **7**D-**3**.

FIG. **15** depicts an embodiment somewhat similar to what has been described with respect to FIG. **8**B or FIG. **8**C. Once again unit **500** is a handheld transceiver, and includes fractal antennas **510**A, **510**B–**510**B', **510**C. It is again understood that a vertical antenna such as elements **193** and fractal counterpoise **510**D (shown in FIG. **8**B) may be provided. Antennas **510**B–**510**B' are similar to what has been described with respect to FIGS. **12**A–**12**B. Antennas **510**B–**510**B' are fractal antennas, not necessarily MI-2 configuration as shown, and are spaced-apart a distance Δ and, in FIG. **13**, are rotationally displaced. Collectively, the spaced-apart distance and relative rotational displacement permits tuning the characteristics of the driven antenna, here antenna **510**B. In FIG. **14**, antenna **510**A is drawn with phantom lines to better distinguish it from spaced-apart antenna **510**B. Of course passive conductor **510**B' could instead be a solid conductor such as described with respect to FIG. **11**A. Such conductor may be implemented by spraying the inner surface of the housing for unit **500** adjacent antenna **510**B with conductive paint.

In FIG. **15**, antenna **510**C is similar to what has been described with respect to FIG. **13**A, in that a cut **830** is made in the antenna, for tuning purposes. Although antenna **510**A is shown similar to what was shown in FIG. **8**B, antenna **510**A could, if desired, be formed on a curved substrate similar to FIGS. **13**B or **13**C. While FIG. **15** shows at least two different techniques for tuning antennas according to the present invention, it will be understood that a common technique could instead be used. By that it is meant that any or all of antennas **510**A, **510**B–**510**B', **510**C could include a cut, or be spaced-apart a controllable distance Δ, or be rotatable relative to a spaced-apart conductor.

As described with respect to FIG. **8**B, an electronic circuit **610** may be coupled by cables **50**A, **50**B, **50**C to the antennas, and samples incoming signals to discern which fractal antenna, e.g., **510**A, **510**B–**510**B', **510**C (and, if present, antenna **510**D-**197**) is presently most optimally aligned with the transmitting station, perhaps a unit **600** or **650** or **660** as shown in FIG. **8**A. This determination may be made by examining signal strength from each of the antennas. An electronic circuit **620** then selects the presently best oriented antenna, and couples such antenna to the input of the receiver and output of the transmitter portion, collectively **630**, of unit **500**. It is understood that the selection of the best antenna is dynamic and can change as, for example, a user of **500** perhaps walks about holding the unit, or the transmitting source moves, or due to other changing conditions. In a cellular or a wireless telephone application, the result is more reliable communication, with the advantage that the fractal antennas can be sufficiently small-sized as to fit totally within the casing of unit **500**. Further, if a flexible substrate is used, the antennas may be wrapped about portions of the internal casing, as shown.

An additional advantage of the embodiment of FIG. **8**B is that the user of unit **500** may be physically distanced from the antennas by a greater distance that if a conventional external whip antenna were used. Although medical evidence attempting to link cancer with exposure to electromagnetic radiation from handheld transceivers is still inconclusive, the embodiment of FIG. **8**B appears to minimize any such risk.

Modifications and variations may be made to the disclosed embodiments without departing from the subject and spirit of the invention as defined by the following claims. While common fractal families include Koch, Minkowski, Julia, diffusion limited aggregates, fractal trees, Mandelbrot, ground counterpoise elements and/or top-hat loading elements according to the present invention may be implemented with other fractals as well.

What is claimed is:

1. An antenna system including:

a driven element, and

a fractal counterpoise element having a portion that includes at least a first motif and a first replication of said first motif and a second replication of said first motif such that a point chosen on a geometric figure represented by said first motif will result in a corresponding point on said first replication and on said second replication of said first motif; wherein there exists at least one non-straight line locus connecting each said point;

wherein a replication of said first motif is a change selected from a group consisting of (a) a rotation and change of scale of said first motif, (b) a linear displacement translation and a change of scale of said first motif, and (c) a rotation and a linear displacement translation and a change of scale of said first motif; and

wherein said fractal counterpoise element has a perimeter compression parameter (PC) defined by:

PC=(full-sized antenna element length)/(fractal-reduced antenna element length)

in which PC=A·log(N(D+C)). A and C are constant coefficients for said first motif, N is an iteration number, and D is a fractal dimension given by log(L)/log(r), where L and r are one-dimensional fractal counterpoise element lengths before and after fractalization, respectively.

2. The antenna system of claim **1**, wherein said first motif is selected from a group consisting of (i) Koch, (ii) Minkowski, (iii) Cantor, (iv) torn square, (v) Mandelbrot, (vi) Caley tree, (vii) monkey's swing, (viii) Sierpinski gasket, and (ix) Julia.

3. The antenna system of claim **1**, wherein said driven element includes an element having a portion that includes at least a second motif and a first replication of said second motif and a second replication of said second motif such that a point chosen on a geometric figure represented by said second motif will result in a corresponding point on said first replication and on said second replication of said second motif; wherein there exists at least one non-straight line locus connecting each said point; and wherein a replication of said second motif is a change selected from a group consisting of (a) a rotation and change of scale of said second motif, (b) a linear displacement translation and a change of scale of said second motif, and (c) a rotation and a linear displacement translation and a change of scale of said second motif.

4. The antenna system of claim **1**, wherein said driven element includes a spaced-apart second element having a portion that includes at least a third motif and a first replication of said third motif and a second replication of said third motif such that a point chosen on a geometric figure represented by said third motif will result in a corresponding point on said first replication and on said second replication of said third motif; wherein there exists at least one non-straight line locus connecting each said point; and

wherein a replication of said third motif is a change selected from a group consisting of (a) a rotation and change of scale of said third motif, (b) a linear displacement translation and a change of scale of said third motif, and (c) a rotation and a linear displacement translation and a change of scale of said third motif.

5. The antenna system of claim **3**, wherein at least one of said first motif and said second motif is selected from a group consisting of (i) Koch, (ii) Minkowski, (iii) Cantor, (iv) torn square, (v) Mandelbrot, (vi) Caley tree, (vii) monkey's swing, (viii) Sierpinski gasket, and (ix) Julia.

6. The antenna system of claim **1**, wherein said first motif has x-axis, y-axis coordinates for a next iteration N+1 defined by $x_{N+1}=f(x_N, y_N)$ and $y_{N+1}=g(x_N, y_N)$, where x_N, y_N are coordinates for iteration N, and where f(x,y) and g(x,y) are functions defining said first motif.

7. The antenna system of claim **4**, wherein at least one of said first motif and said third motif is selected from a group consisting of (i) Koch, (ii) Minkowski, (iii) Cantor, (iv) torn square, (v) Mandelbrot, (vi) Caley tree, (vii) monkey's swing, (viii) Sierpinski gasket, and (ix) Julia.

8. The antenna system of claim **1**, in which said fractal counterpoise element is fabricated in a manner selected from a group consisting of (i) shaping conductive wire to form said fractal counterpoise element, (ii) forming upon an insulator substrate a conductive layer defining traces shaped to form said fractal counterpoise element, (iii) forming upon a flexible insulator substrate conductive traces shaped to form said fractal counterpoise element, and (iv) forming upon a semiconductor substrate a layer of conductive material shaped to form said fractal counterpoise element.

9. The antenna system of claim **1**, wherein:

said antenna system is a vertical antenna system;

said fractal counterpoise element includes three dendrite elements that each have an overall length approximating 0.087 λ where λ is wavelength at a resonant frequency of interest; and wherein gain of said antenna system is within at least 1 dB of unity.

10. A fractal antenna coupleable to a transceiver unit, the fractal antenna comprising:

a driven element, and

a ground counterpoise system including at least one fractal counterpoise element having a portion that includes at least a first motif and a first replication of said first motif and a second replication of said first motif such that a point chosen on a geometric figure represented by said first motif will result in a corresponding point on said first replication and on said second replication of said first motif; wherein there exists at least one non-straight line locus connecting each said point;

wherein a replication of said first motif is a change selected from a group consisting of (a) a rotation and change of scale of said first motif, (b) a linear displacement translation and a change of scale of said first motif, and (c) a rotation and a linear displacement translation and a change of scale of said first motif; and

wherein said first motif is selected from a group consisting of (i) Koch, (ii) Minkowski, (iii) Cantor, (iv) torn square, (v) Mandelbrot, (vi) Caley tree, (vii) monkey's swing, (viii) Sierpinski gasket, and (ix) Julia.

11. The fractal antenna of claim **10**, wherein said fractal counterpoise element is fabricated in a manner selected from a group consisting of (i) shaping conductive wire to form said fractal counterpoise element, (ii) forming upon an insulator substrate a conductive layer defining traces shaped to form said fractal counterpoise element, (iii) forming upon a flexible insulator substrate conductive traces shaped to form said fractal counterpoise element, and (iv) forming upon a semiconductor substrate a layer of conductive material shaped to form said fractal counterpoise element.

12. The fractal antenna of claim **10**, wherein said first motif has x-axis, y-axis coordinates for a next iteration N+1 defined by $x_{N+1}=f(x_N, y_N)$ and $y_{N+1}=g(x_N, y_N)$, where x_N, y_N are coordinates for iteration N, and where f(x,y) and g(x,y) are functions defining said first motif.

13. The fractal antenna of claim **10**, wherein said ground counterpoise system has a perimeter compression parameter (PC) defined by:

$$PC = \frac{\text{full-sized antenna element length}}{\text{fractal-reduced antenna element length}}$$

where:

$$PC = A \cdot \log[N(D+C)]$$

in which A and C are constant coefficients for said first motif, N is an iteration number, and D is a fractal dimension given by log(L)/log(r), where L and r are one-dimensional fractal counterpoise system lengths before and after fractalization, respectively.

14. The fractal antenna of claim **10**, wherein:

said transceiver unit is hand holdable in size;

said antenna is mounted within a housing of said transceiver unit; and

said fractal counterpoise element is fabricated in a manner selected from a group consisting of (i) shaping conductive wire to form said fractal counterpoise element, (ii) forming upon an insulator substrate a conductive layer defining traces shaped to form said fractal counterpoise element, (iii) forming upon a flexible insulator substrate conductive traces shaped to form said fractal counterpoise element, and (iv) forming upon a semiconductor substrate a layer of conductive material shaped to form said fractal counterpoise element.

15. The fractal antenna of claim **10**, wherein said transceiver includes a plurality of such fractal antennas arrayed in at least one configuration selected from a group consisting of (i) an array of substantially identical said fractal antennas coupled to an electronic circuit that dynamically selects a chosen one of said fractal antennas to be electronically coupled to said transceiver unit, (ii) an array of substantially identical such fractal antennas coupled to an electronic circuit that dynamically selects a chosen one of said fractal antennas to be electronically coupled to said transceiver unit wherein at least two fractal antennas in said array differ in antenna orientation from other fractal antennas in said array, (iii) a plurality of said fractal antennas wherein at least two said fractal antennas have elements that differ from elements in other said fractal antennas.

16. The antenna of claim **10**, wherein said driven element has a portion that includes at least a second motif provided by a fractal generator, and a conductive element spaced-apart from said driven element by a distance Δ chosen to vary at least one characteristic of said antenna, at a desired frequency c/λ where λ and wavelength at a frequency of interest and c is velocity of light, selected from the group consisting of (i) resonant frequency, and (ii) bandwidth.

17. The antenna of claim **16**, wherein said antenna is tunable by varying at least one parameter selected from a group consisting of (a) said distance Δ, (b) relative rotation between at least a part of said ground counterpoise system and said portion of said driven element, (c) location at which a feedline center lead is coupled to said portion of said driven element, (d) location of a cut in said portion of said driven element, and (e) size of a cut-away region of said portion of said driven element.

18. A top-hat loaded antenna, comprising:

a vertical element having an upper end and a lower end; and

a top-hat assembly electrically coupled to said upper end of said vertical element;

wherein said top-hat assembly includes a top-hat element having a portion that includes a first motif and a first replication of said first motif and a second replication of said first motif such that a point chosen on a geometric figure represented by said first motif will result in a corresponding point on said first replication and on said second replication of said first motif; wherein there exists at least one non-straight line locus connecting each said point;

wherein a replication of said first motif is a change selected from a group consisting of (a) a rotation and change of scale of said first motif, (b) a linear displacement translation and a change of scale of said first motif, and (c) a rotation and a linear displacement translation and a change of scale of said first motif; and

wherein said top-hat element has a perimeter compression parameter (PC) defined by:

PC=(full-sized antenna element length)/(fractal-reduced antenna element length)

in which PC=A·log(N(D+C)), A and C are constant coefficients for said first motif, N is an iteration number, and D is a fractal dimension given by log(L)/log(r), where L and r are one-dimensional fractal top-hat element lengths before and after fractalization, respectively.

19. The antenna of claim **18**, wherein said first motif is selected from a group consisting of (i) Koch, (ii) Minkowski, (iii) Cantor, (iv) torn square, (v) Mandelbrot, (vi) Caley tree, (vii) monkey's swing, (viii) Sierpinski gasket, and (ix) Julia.

* * * * *

US007019695B2

(12) **United States Patent**
Cohen

(10) **Patent No.: US 7,019,695 B2**
(45) **Date of Patent: *Mar. 28, 2006**

---

(54) **FRACTAL ANTENNA GROUND COUNTERPOISE, GROUND PLANES, AND LOADING ELEMENTS AND MICROSTRIP PATCH ANTENNAS WITH FRACTAL STRUCTURE**

(76) Inventor: **Nathan Cohen**, 21 Ledgewood Pl., Belmont, MA (US) 02178

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **10/287,240**

(22) Filed: **Nov. 4, 2002**

(65) **Prior Publication Data**

US 2003/0151556 A1 Aug. 14, 2003

Related U.S. Application Data

(63) Continuation of application No. 09/677,645, filed on Oct. 3, 2000, now Pat. No. 6,476,766, which is a continuation of application No. 08/967,375, filed on Nov. 7, 1997, now Pat. No. 6,140,975.

(51) **Int. Cl.**
H01Q 1/38 (2006.01)

(52) **U.S. Cl.** .... **343/700 MS**; 343/846

(58) **Field of Classification Search** .... 343/792.5, 343/700 MS, 846, 752
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,079,602	A *	2/1963	DuHamel et al.	343/792.5
3,249,946	A	5/1966	Flanagan	
3,689,929	A	9/1972	Moody	
3,810,183	A	5/1974	Krutsinger et al.	
3,811,128	A	5/1974	Munson	
4,318,109	A	3/1982	Weathers	
4,358,769	A	11/1982	Tada et al.	
4,381,566	A	4/1983	Kane	
4,652,889	A	3/1987	Bizouard et al.	
4,656,482	A	4/1987	Peng	
5,006,858	A	4/1991	Shirosaka	
5,111,211	A	5/1992	Dahlberg	
5,164,738	A	11/1992	Walter et al.	
5,313,216	A *	5/1994	Wang et al.	343/700 MS

(Continued)

OTHER PUBLICATIONS

Pfeiffer, A., “The Pfeiffer Quad Antenna System”, QST, pp. 28-30, 1994.

(Continued)

Primary Examiner—Michael C. Wilmer
(74) *Attorney, Agent, or Firm*—McDermott Will & Emery LLP

(57) **ABSTRACT**

An antenna system includes a fractalized element that may be a ground counterpoise, a top-hat located load assembly, or a microstrip patch antenna having at least one element whose physical shape is at least partially defined as a first or higher iteration deterministic fractal. The resultant fractal element may rely upon an opening angle for performance, and is more compact than non-Euclidean ground counterpoise elements or the like. A vertical antenna system includes a vertical element that may also be a fractal, and a vertical antenna can include vertically spaced-apart fractal conductive and passive elements, and at least one fractal ground element. Various antenna configurations may be fabricated on opposite surfaces of a substrate, including a flexible substrate, and may be tuned by rotating elements relative to each other, and/or by varying the spaced-apart distance therebetween. Fractalized ground counterpoise elements and/or microstrip patch antenna systems may be fabricated on a flexible printed circuit substrate, and/or placed within the support mount of a cellular telephone car antenna.

14 Claims, 25 Drawing Sheets

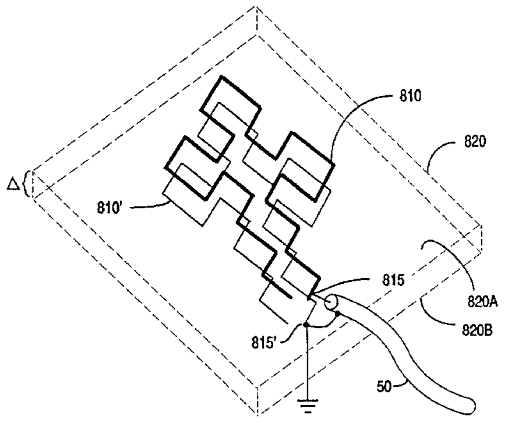

U.S. PATENT DOCUMENTS

5,355,318	A	10/1994	Dionnet et al.	
5,363,114	A	11/1994	Shoemaker	
5,453,752	A	9/1995	Wang	
5,608,413	A *	3/1997	Macdonald	343/700 MS
5,844,525	A	12/1998	Hayes et al.	
5,995,064	A	11/1999	Yanagisawa et al.	
6,104,349	A	8/2000	Cohen	
6,127,977	A	10/2000	Cohen	
6,476,766	B1 *	11/2002	Cohen	343/700 MS

OTHER PUBLICATIONS

Kim, et al., “The Fractal Random Array”, IEEE, vol. 74, No. 9, pp. 1278-1280, Sep. 1986.

* cited by examiner

FIGURE 1A (PRIOR ART)

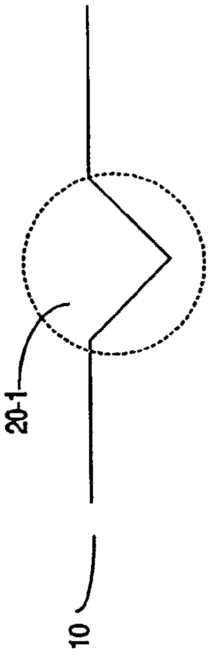

FIGURE 1B (PRIOR ART)

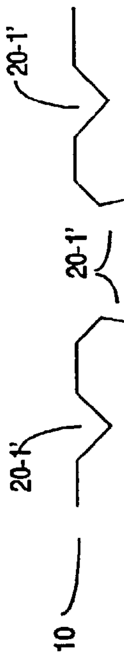
10
20-1'
20-1'
20-1'

FIGURE 1C (PRIOR ART)

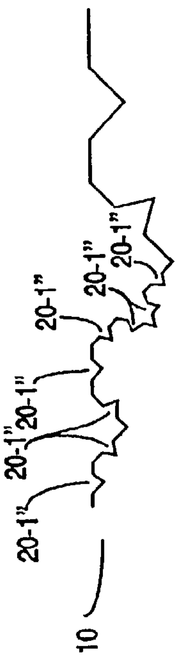
10
20-1"
20-1"
20-1"
20-1"
20-1"
20-1"
20-1"

FIGURE 1D (PRIOR ART)

FIGURE 2A (PRIOR ART)

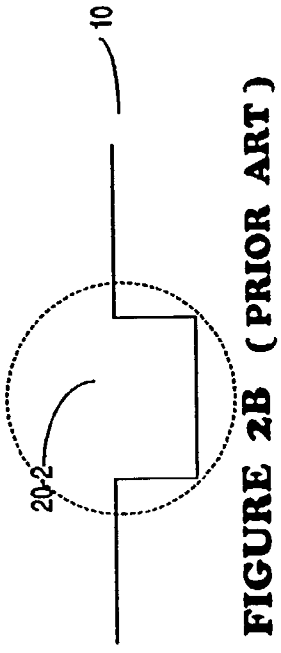

FIGURE 2B (PRIOR ART)

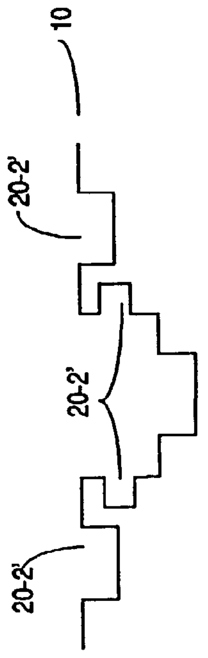
10
20-2'
20-2'
20-2'

FIGURE 2C (PRIOR ART)

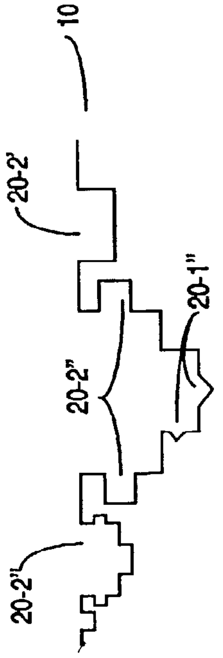
10
20-2'
20-2"
20-1"
20-2"

FIGURE 2D (PRIOR ART)

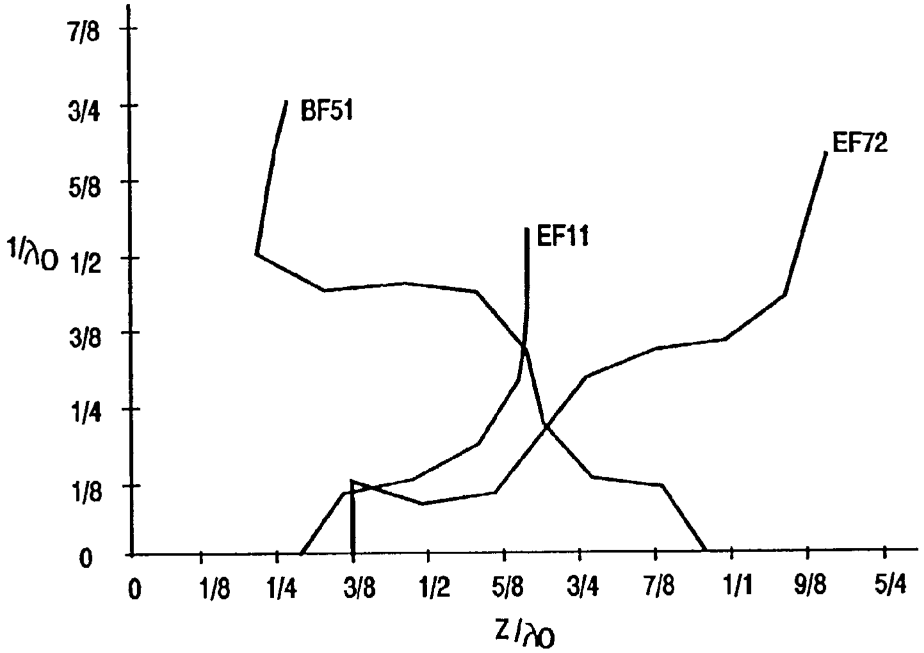

FIGURE 3 (PRIOR ART)

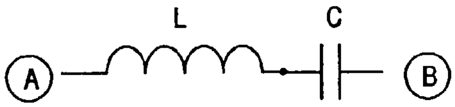

FIGURE 4A (PRIOR ART)

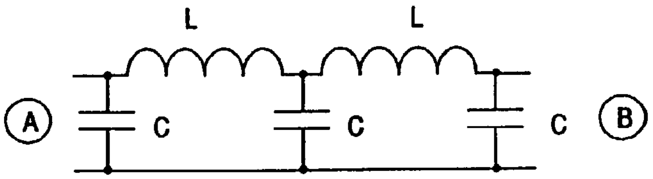

FIGURE 4B (PRIOR ART)

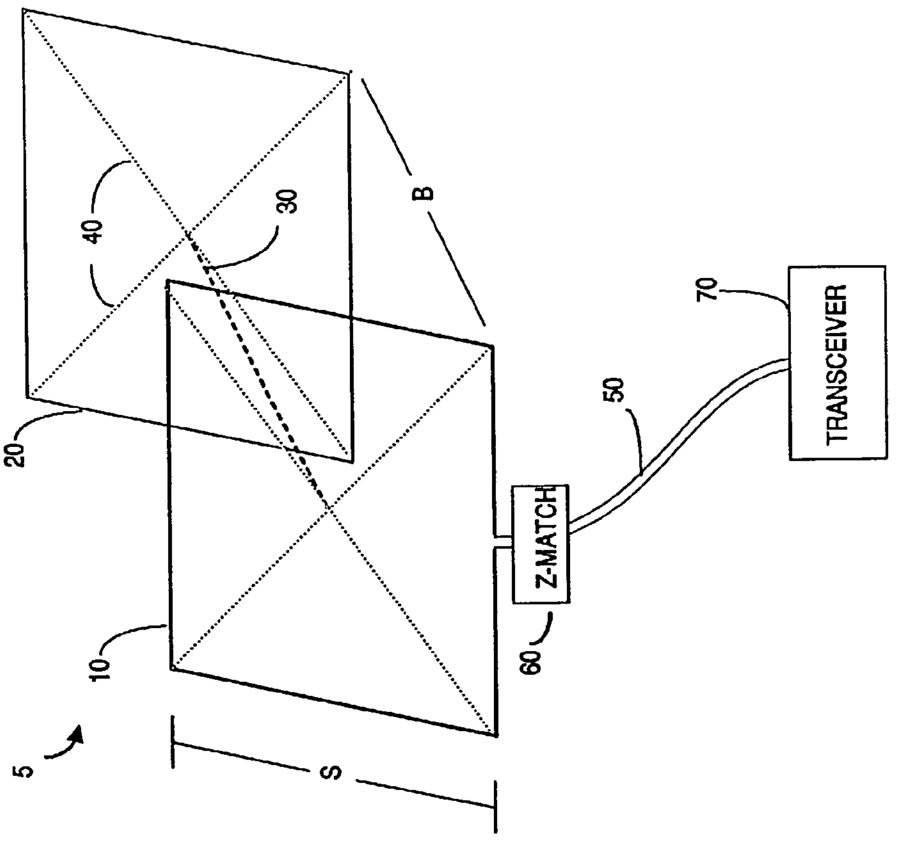
5
10
20
40
30
B
S
60
Z-MATCH
50
70
TRANSCEIVER

FIGURE 5A
(PRIOR ART)

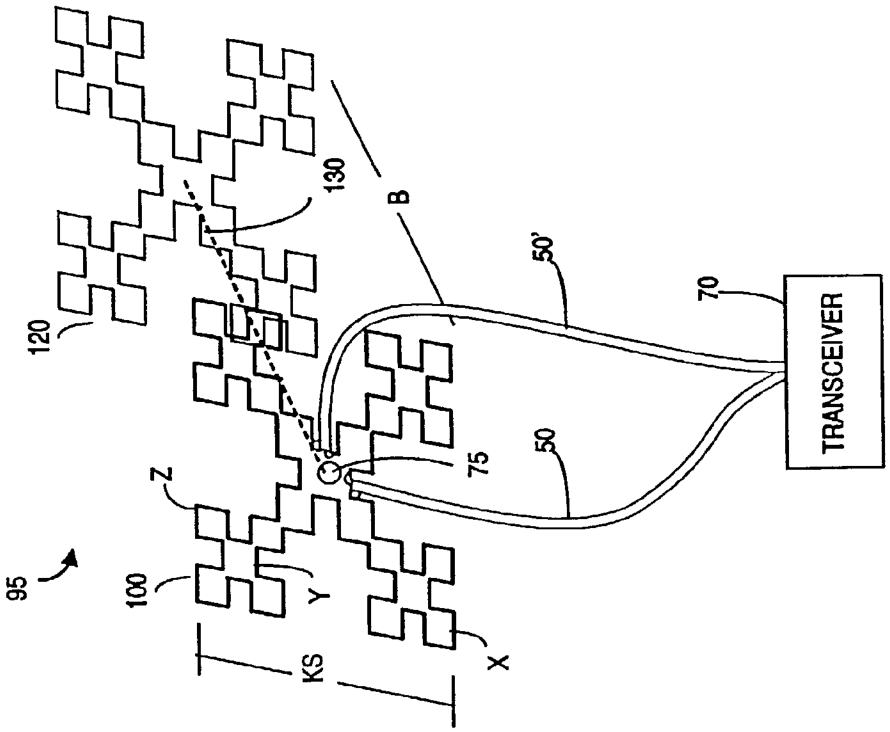
95
100
Z
Y
X
KS
75
50
50'
120
130
B
70
TRANSCEIVER

FIGURE 5B

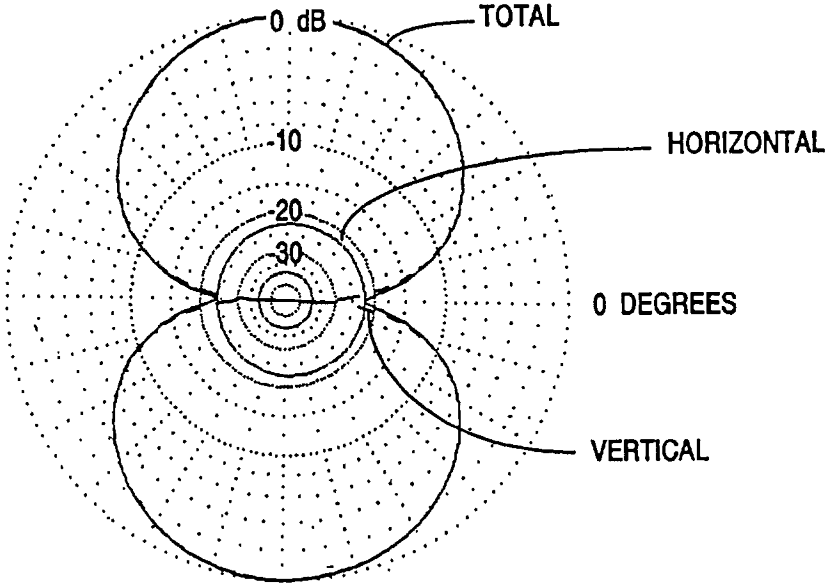

FIGURE 6

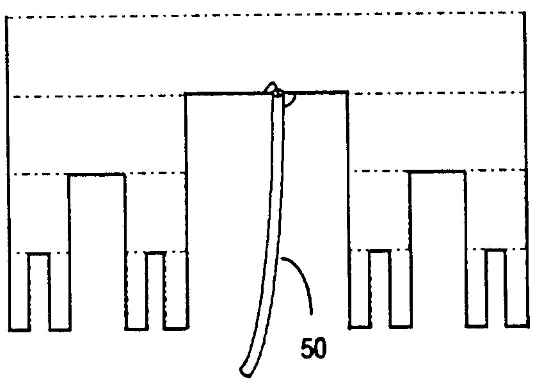

FIGURE 7A

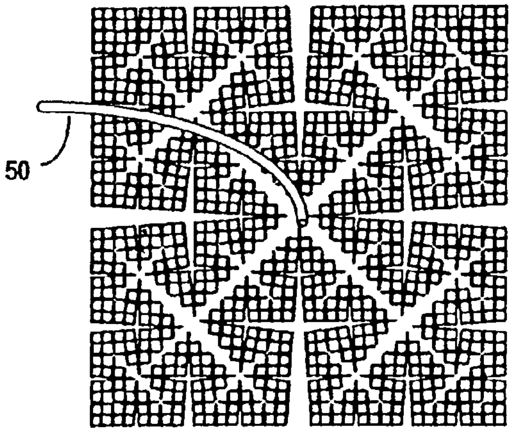

FIGURE 7B

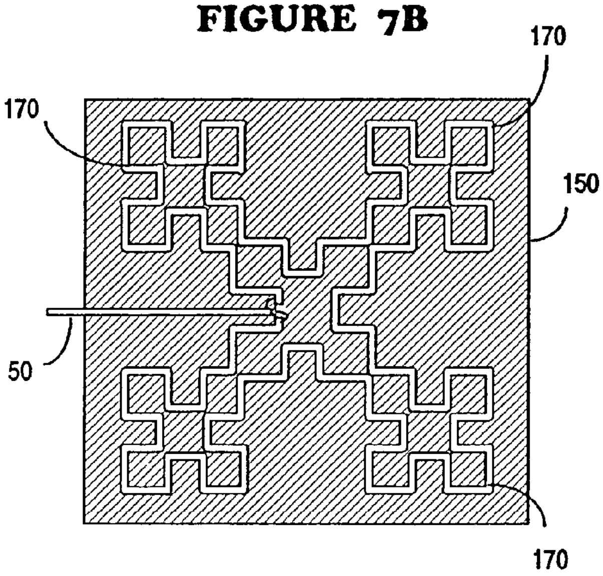
170
170
150
50
170

FIGURE 7C-1

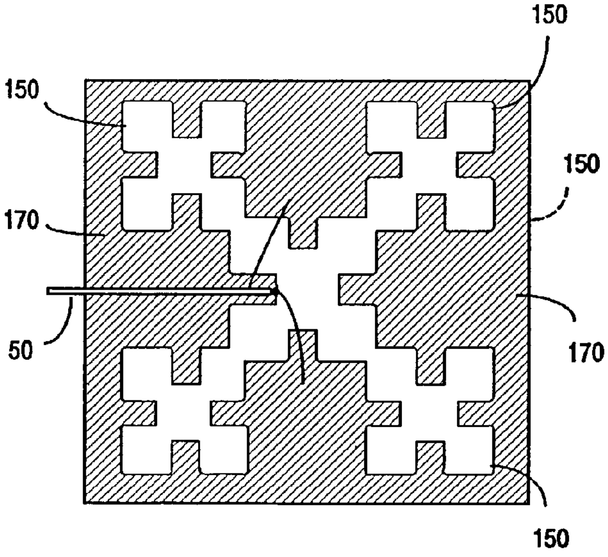
150
150
150
170
50
170
150

FIGURE 7C-2

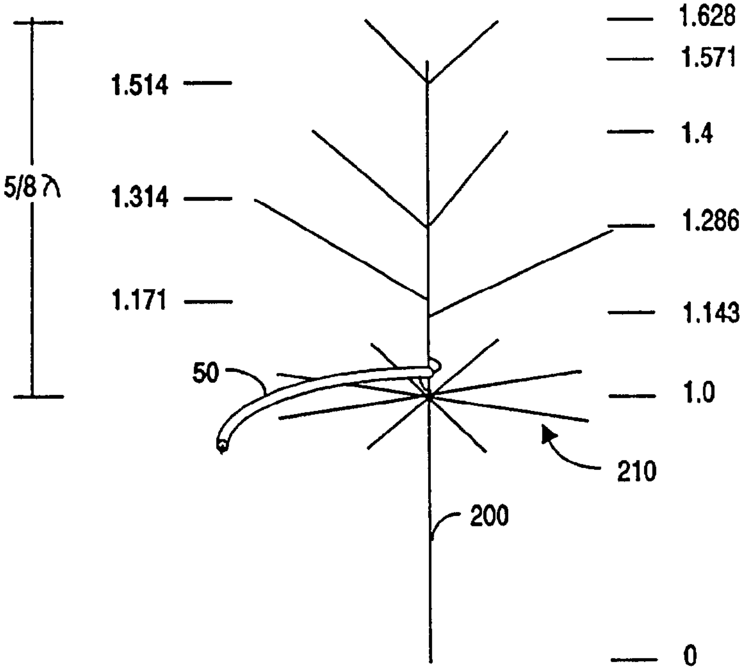
1.628
1.571
1.514
1.4
5/8 λ
1.314
1.286
1.171
1.143
50
1.0
210
200
0

FIGURE 7D

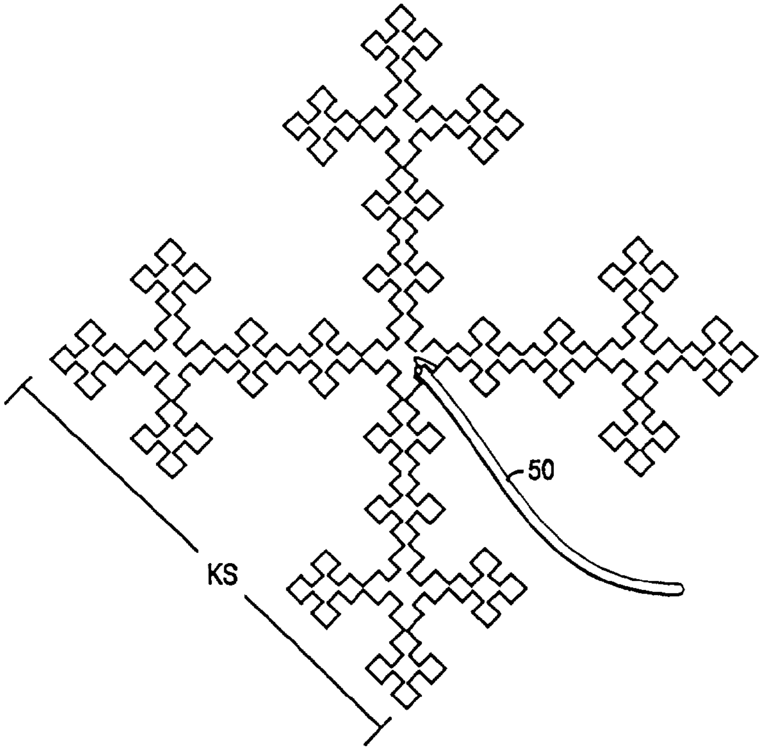

FIGURE 7E

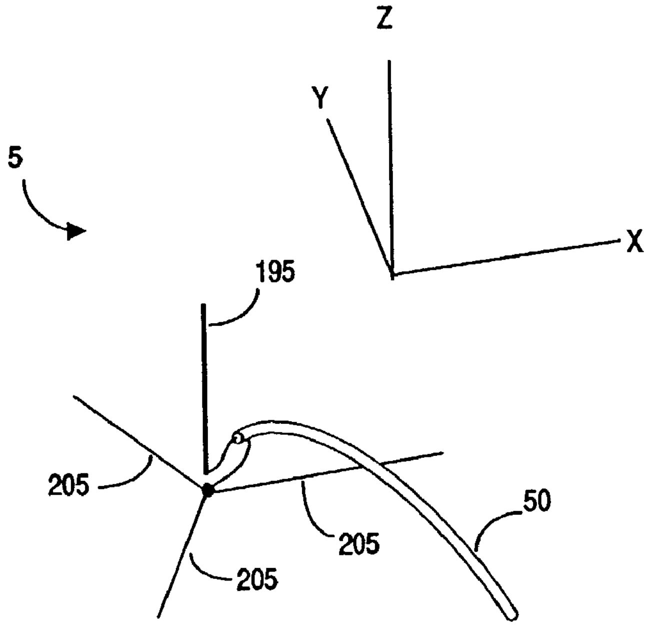
Z
Y
X
5
195
205
205
205
50

FIGURE 7D-1A

(PRIOR ART)

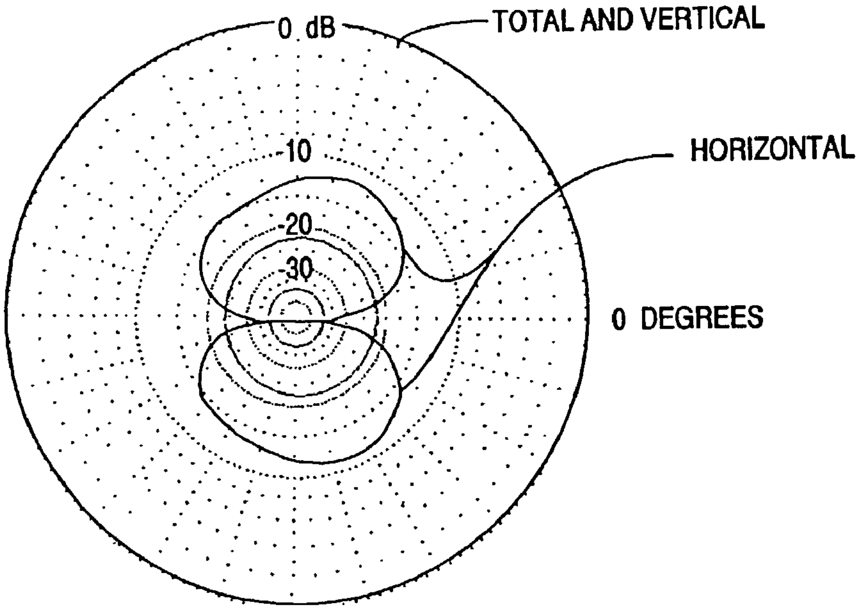

FIGURE 7D-1B

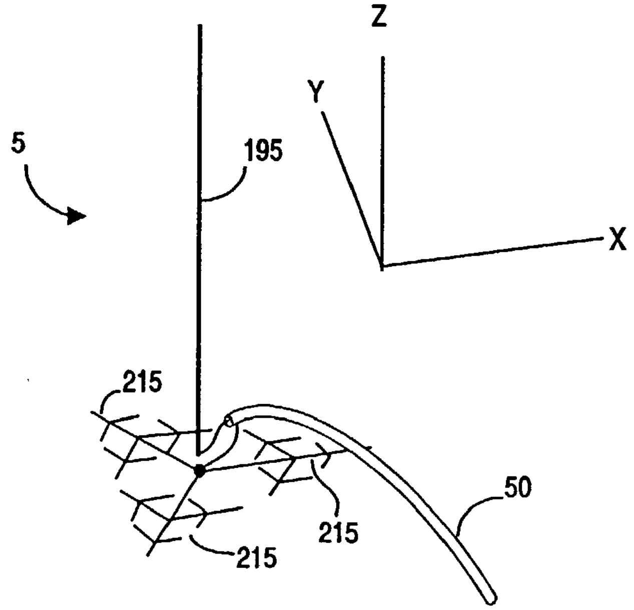
Z
Y
X
5
195
215
215
215
215
50

FIGURE 7D-2A

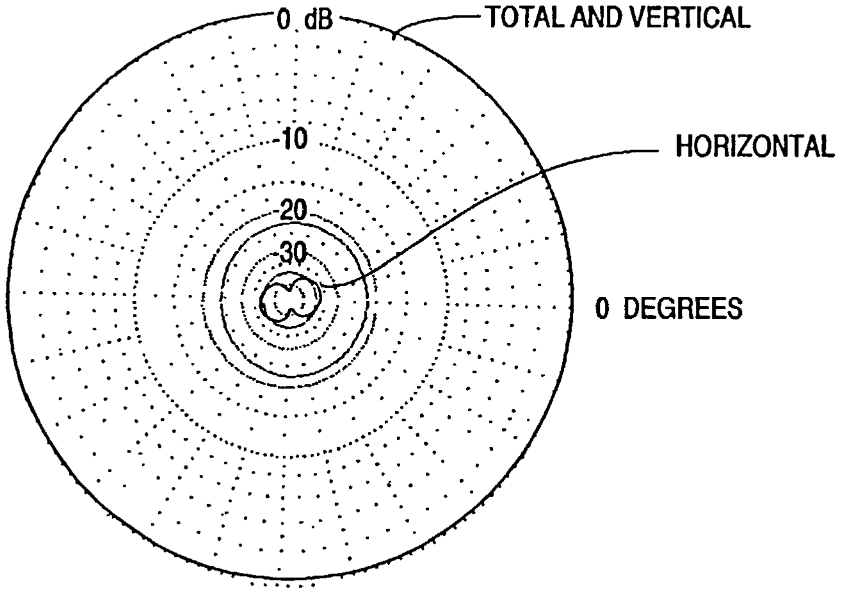

FIGURE 7D-2B

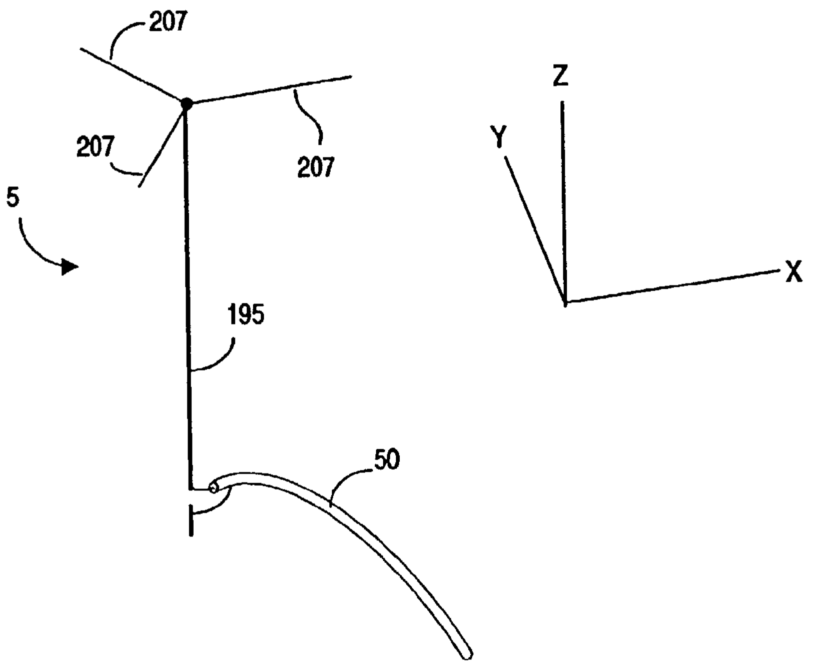
207
207
207
5
195
50
Z
Y
X

FIGURE 7D-3A

(PRIOR ART)

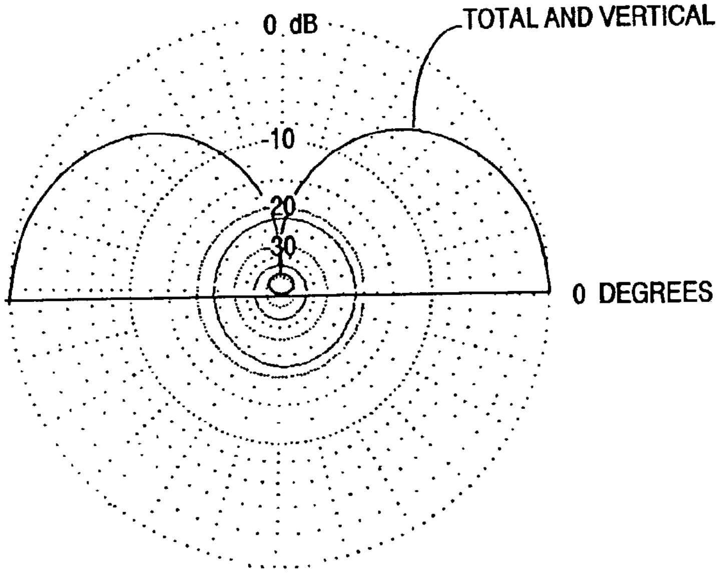

FIGURE 7D-3B

(PRIOR ART)

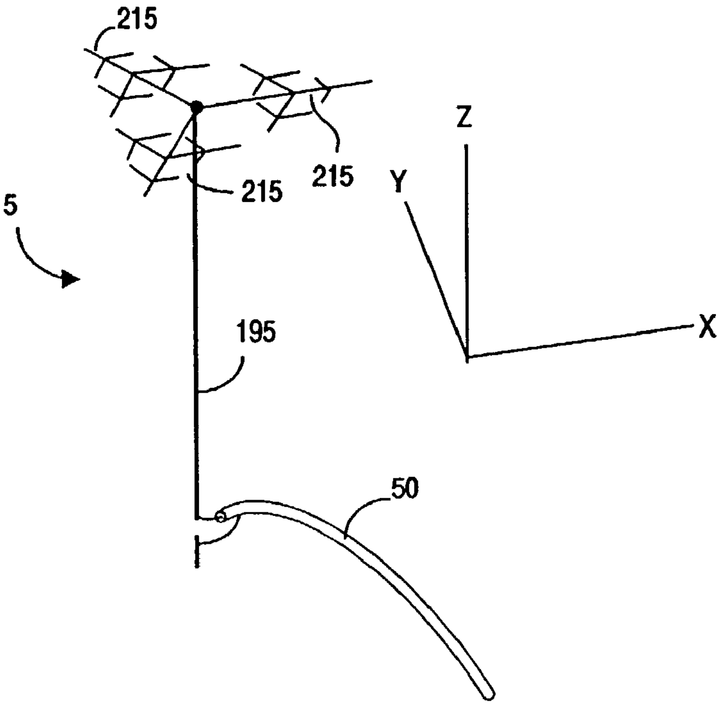
215
215
215
5
195
50
Z
Y
X

FIGURE 7D-4A

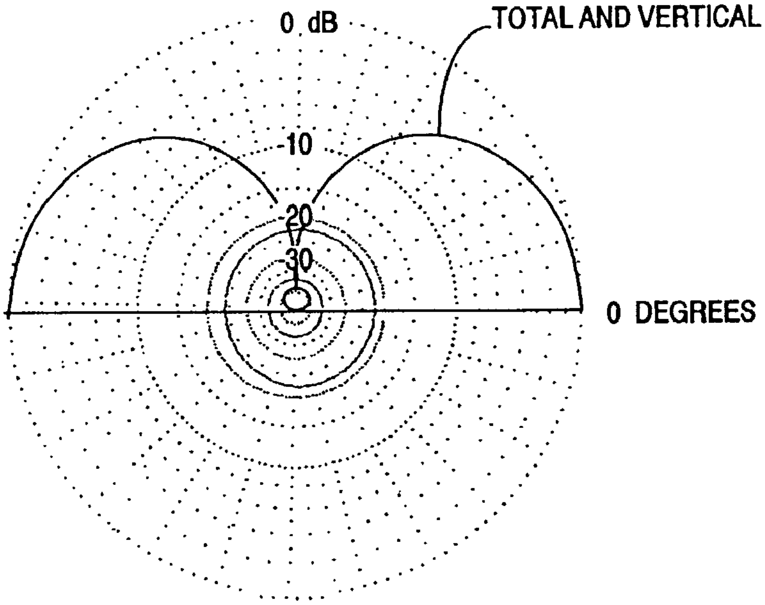

FIGURE 7D-4B

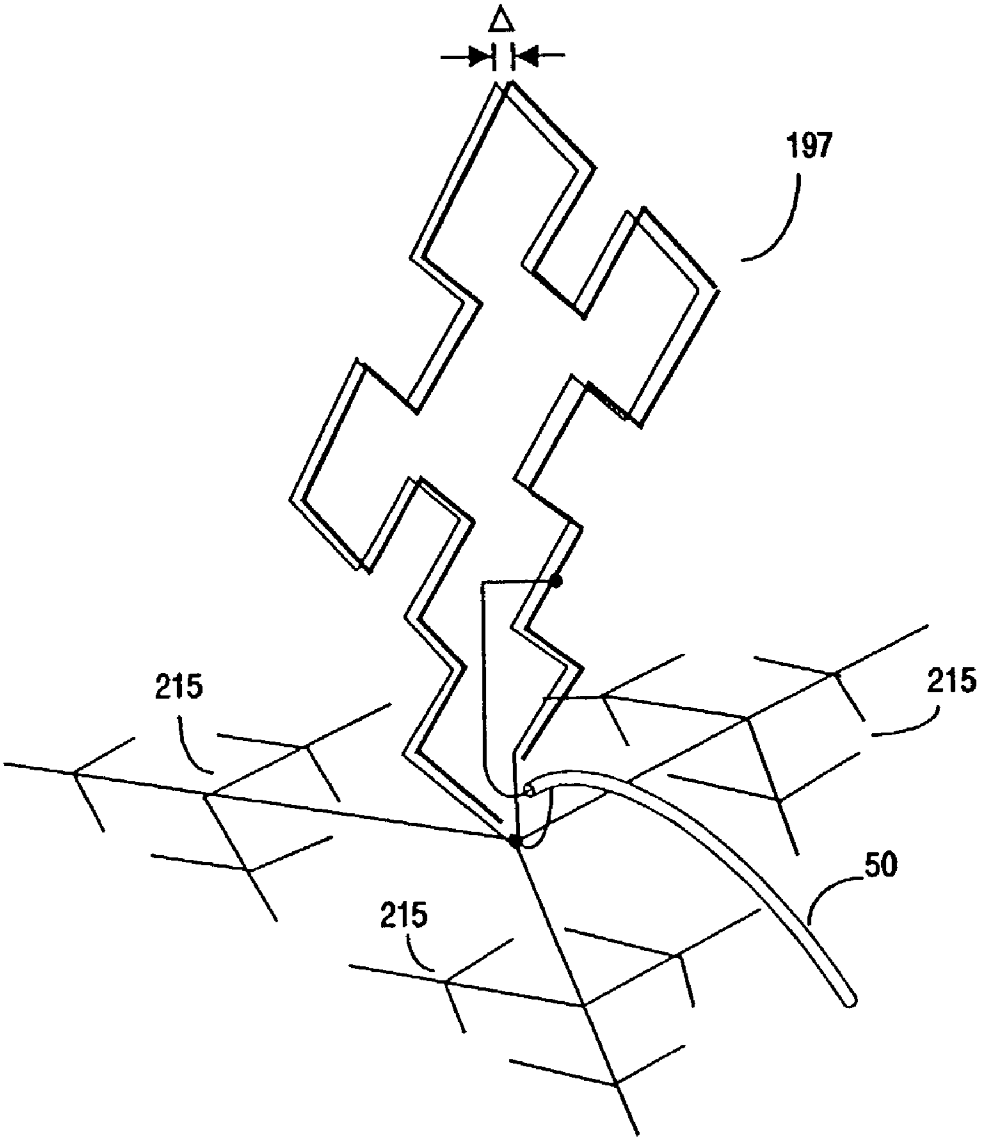
Δ
197
215
215
215
50

FIGURE 7D-5

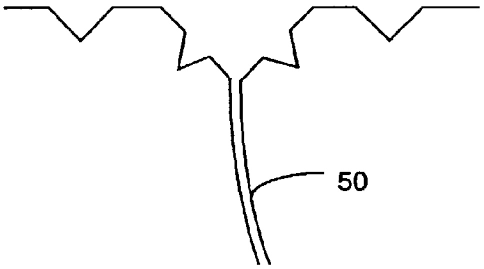

FIGURE 7F

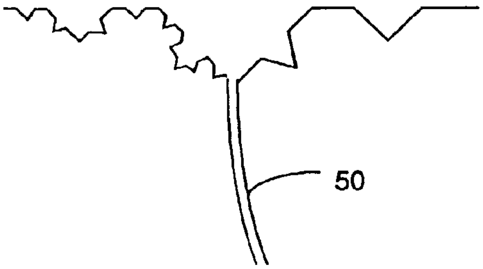

FIGURE 7G

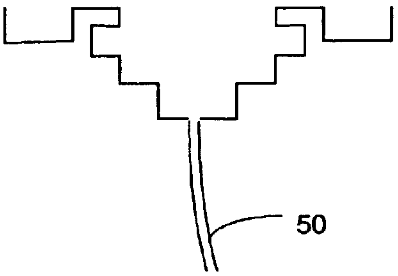

FIGURE 7H

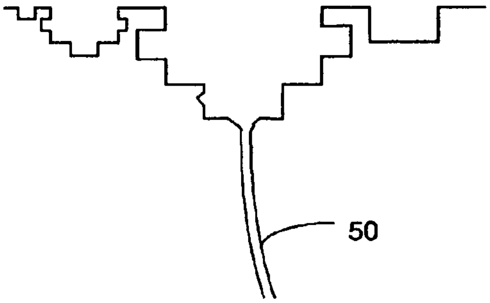

FIGURE 7I

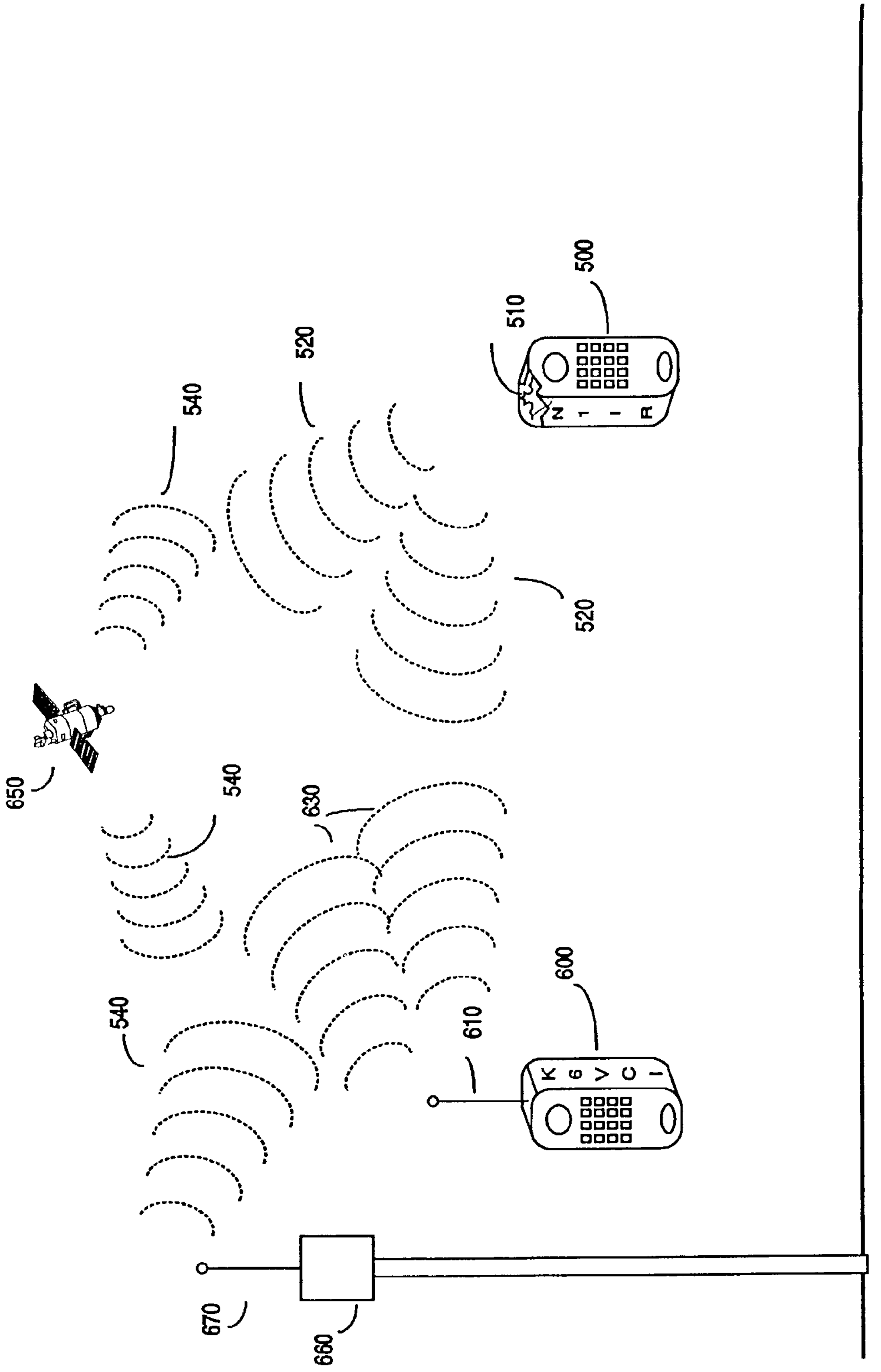
500
510
520
540
520
650
540
630
540
600
610
670
660

FIGURE 8A

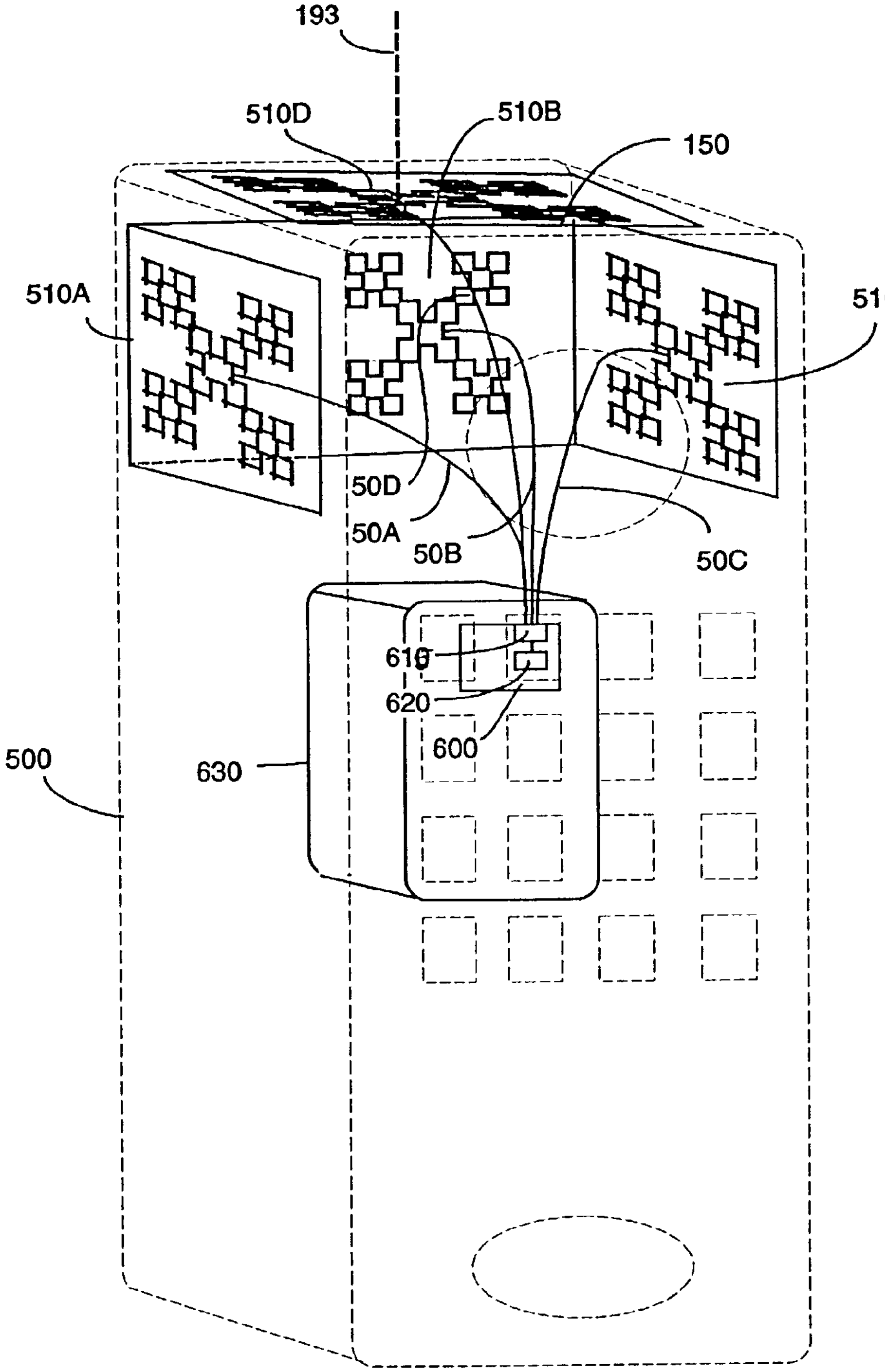
193
510D
510B
150
510A
510C
50D
50A
50B
50C
610
620
600
500
630

FIGURE 8B

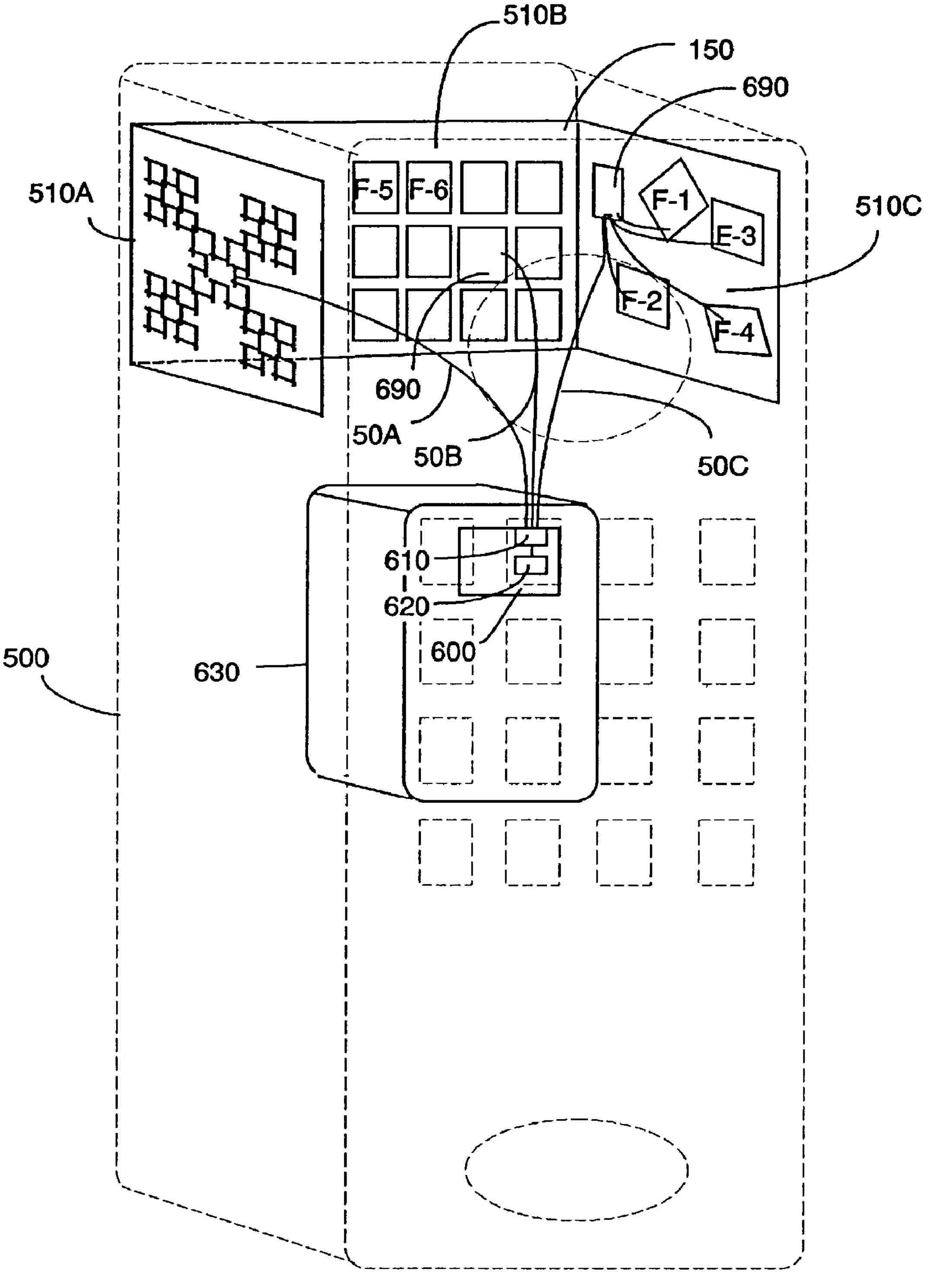
F-5
F-6
F-1
E-3
F-2
F-4

FIGURE 8C

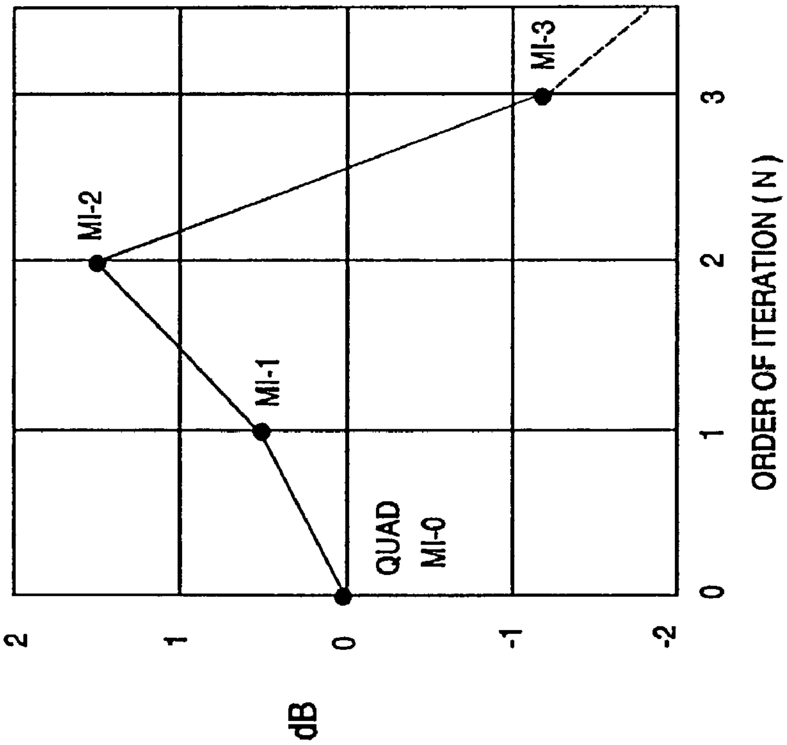

FIGURE 9A

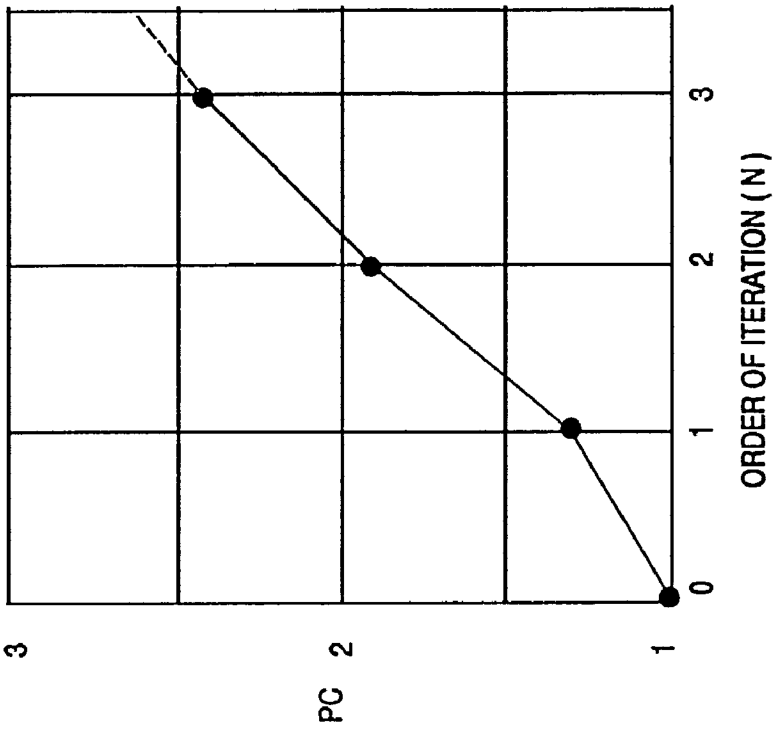

FIGURE 9B

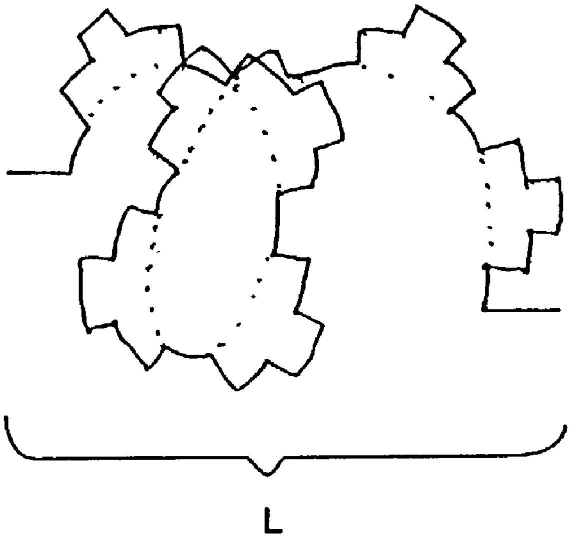

FIGURE 10A

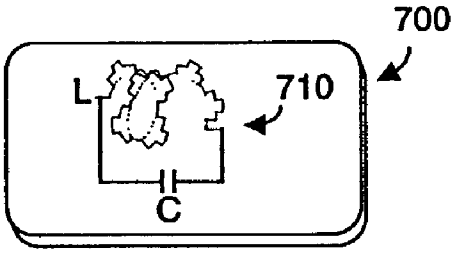

FIGURE 10B

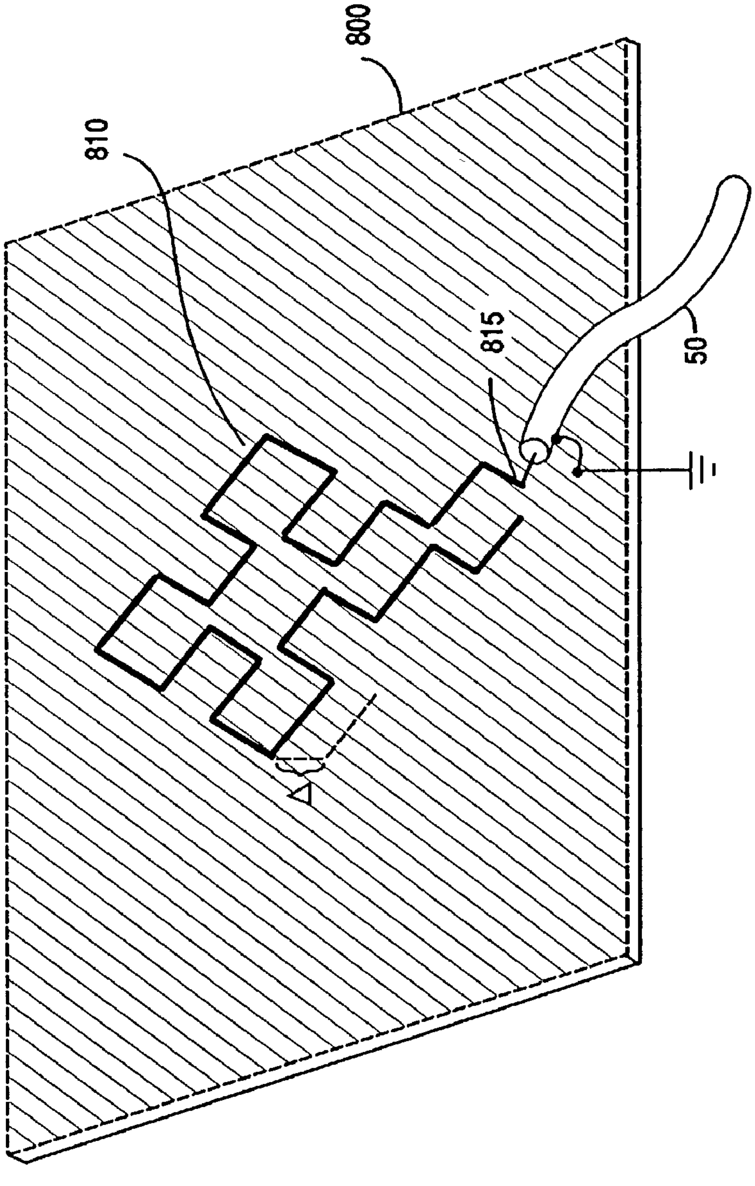
800
810
815
50
Δ

FIGURE 11A

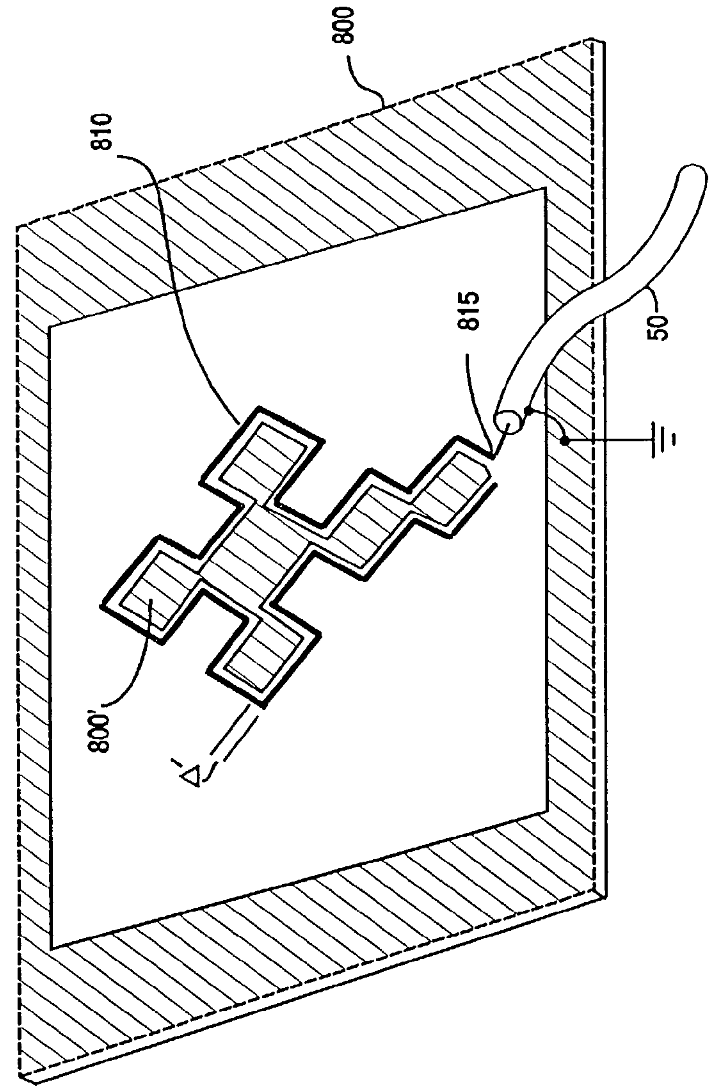
800
810
815
50
800'

FIGURE 11B

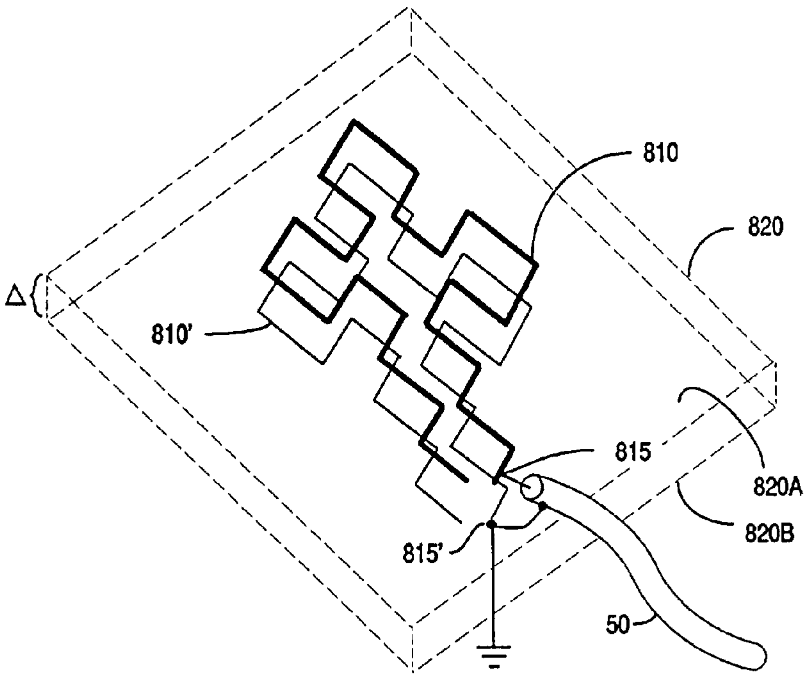
810
820
Δ
810'
815
820A
820B
815'
50

FIGURE 12A

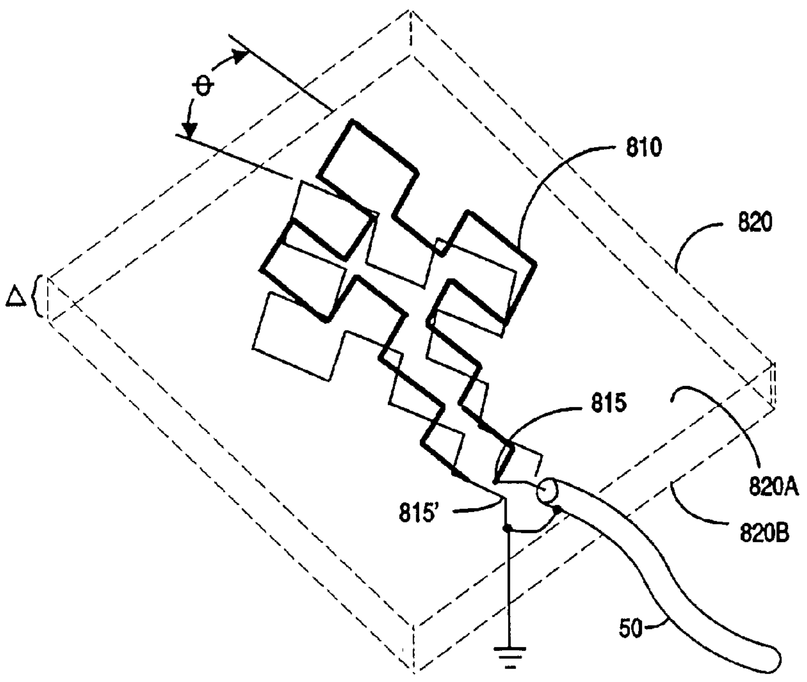
θ
810
820
Δ
815
820A
815'
820B
50

FIGURE 12B

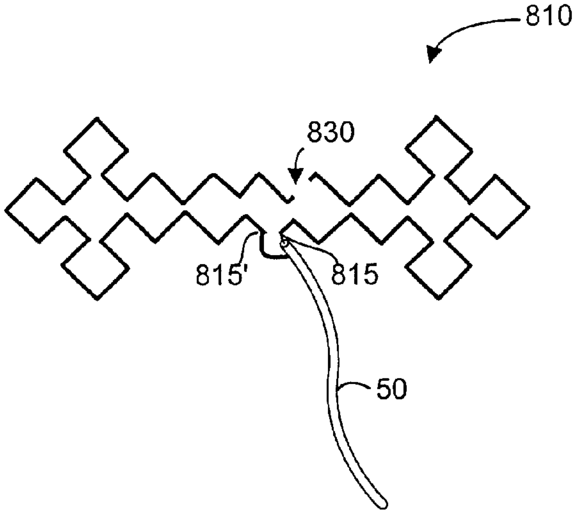
810
830
815'
815
50

FIGURE 13A

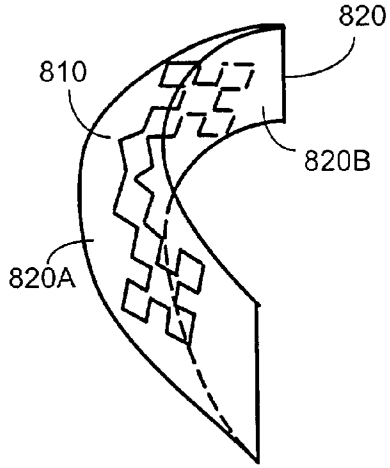
820
810
820B
820A

FIGURE 13B

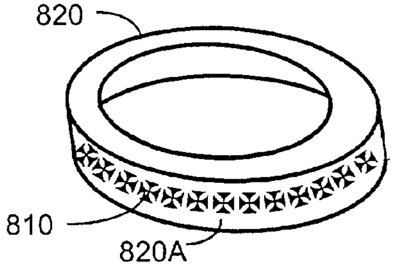
820
810
820A

FIGURE 13C

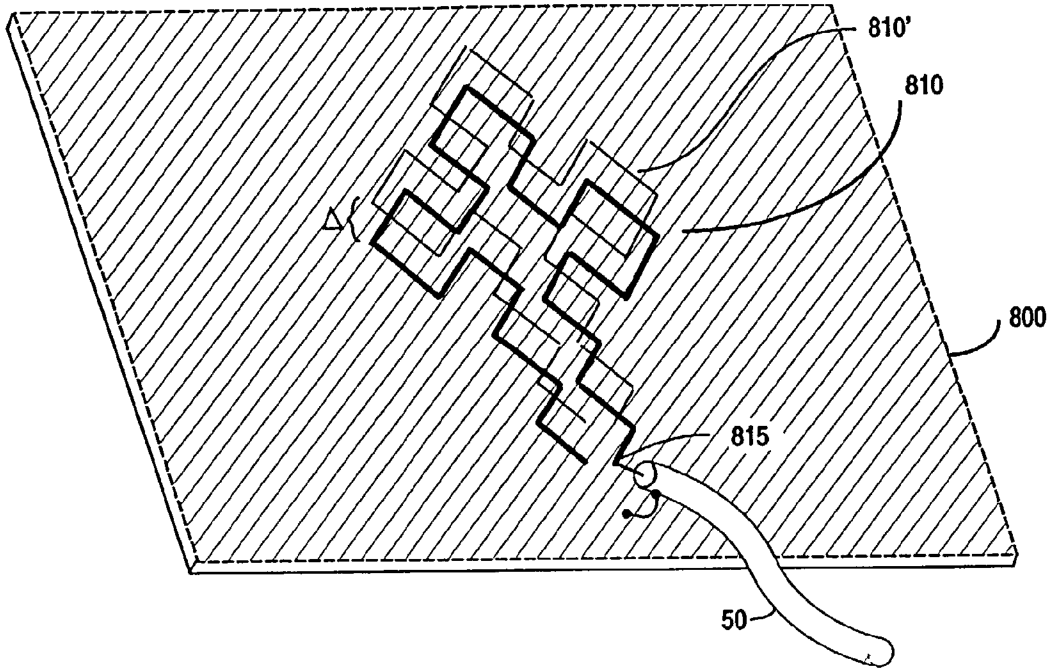
810'
810
Δ
800
815
50

FIGURE 14A

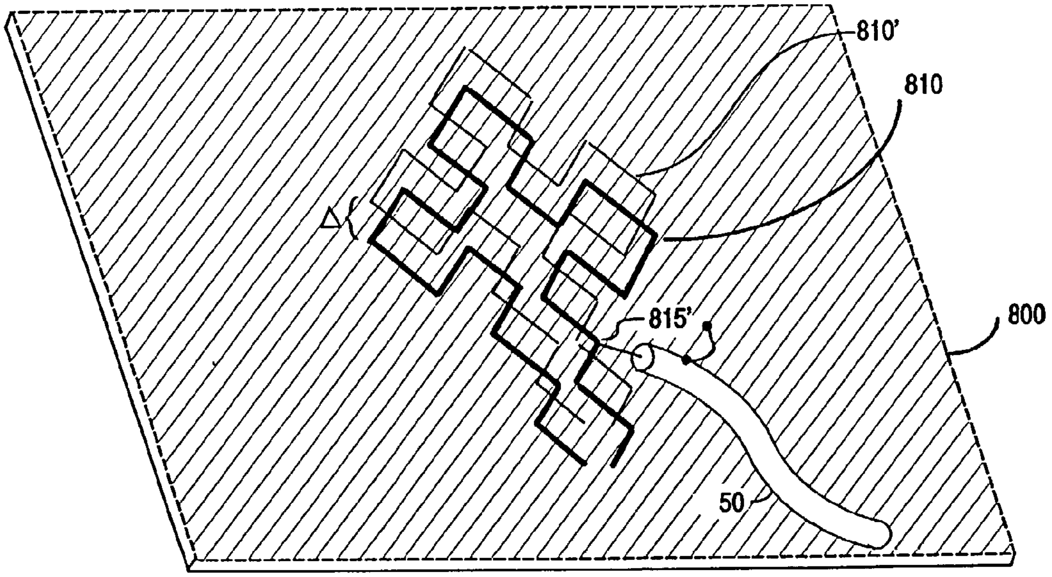
810'
810
Δ
815'
800
50

FIGURE 14B

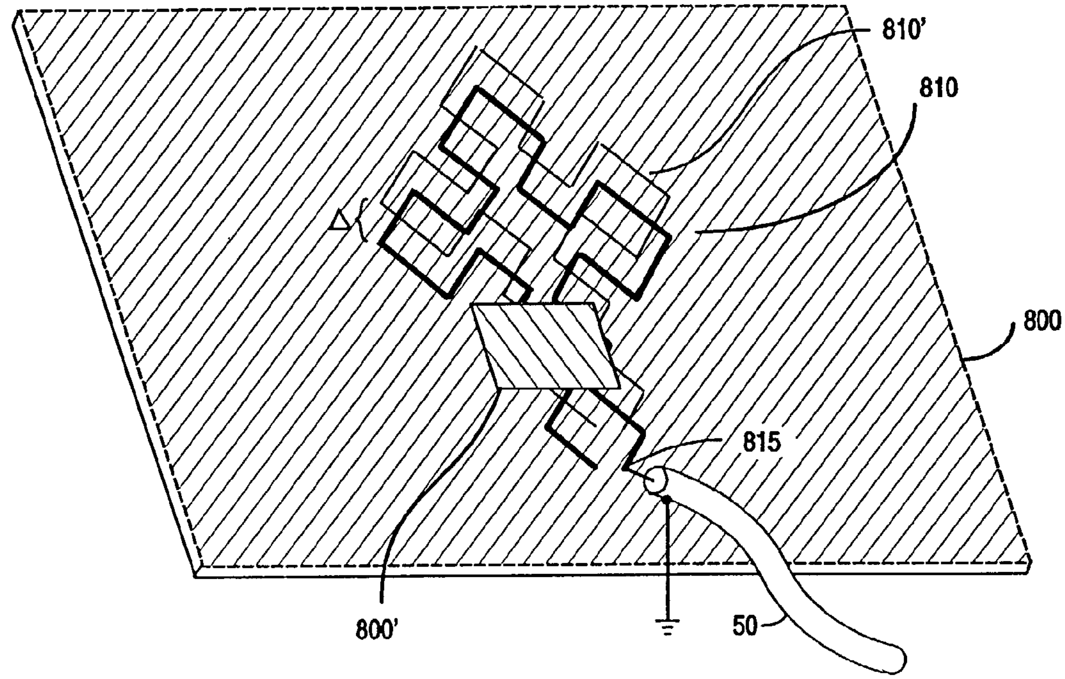
810'
810
Δ
800
815
800'
50

FIGURE 14C

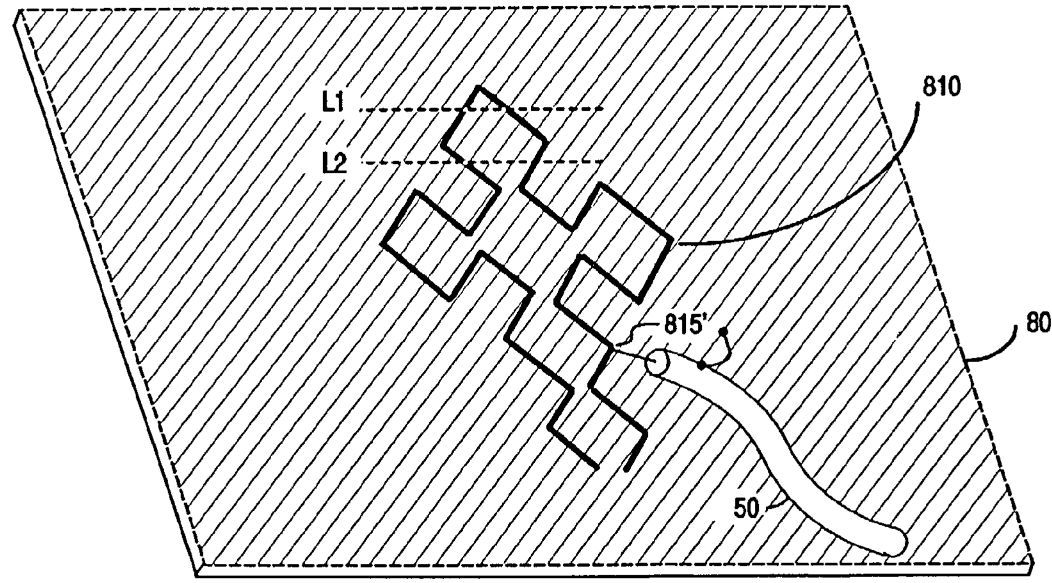
L1
L2
810
815'
800
50

FIGURE 14D

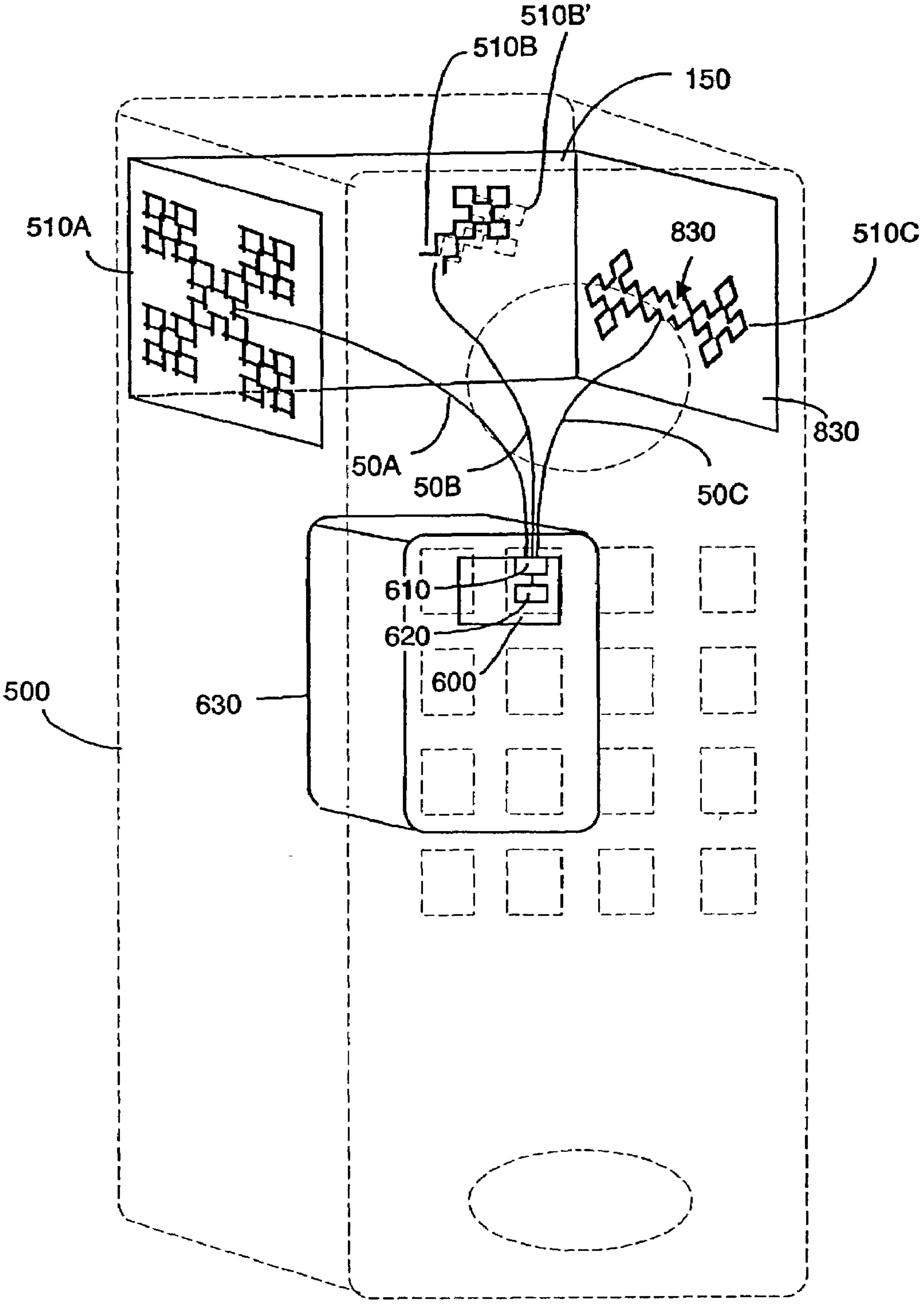
510B'
510B
150
510A
830
510C
830
50A
50B
50C
610
620
600
500
630

FIGURE 15

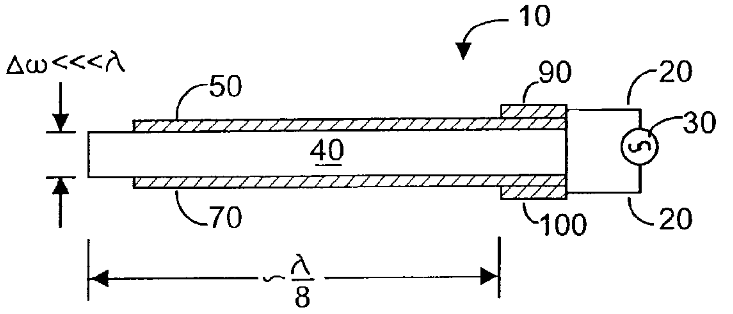
10
Δω<<<λ
50
90
20
30
40
70
100
20
~λ/8

FIG. 16

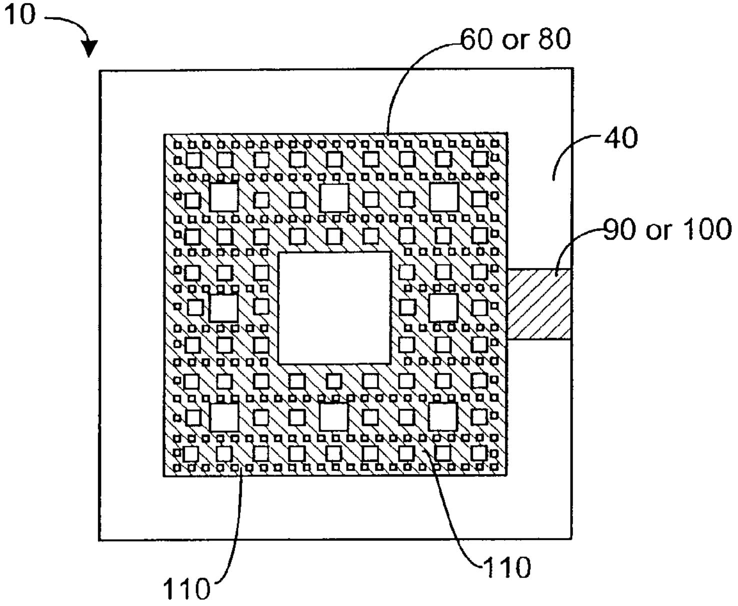
10
60 or 80
40
90 or 100
110
110

FIG. 17

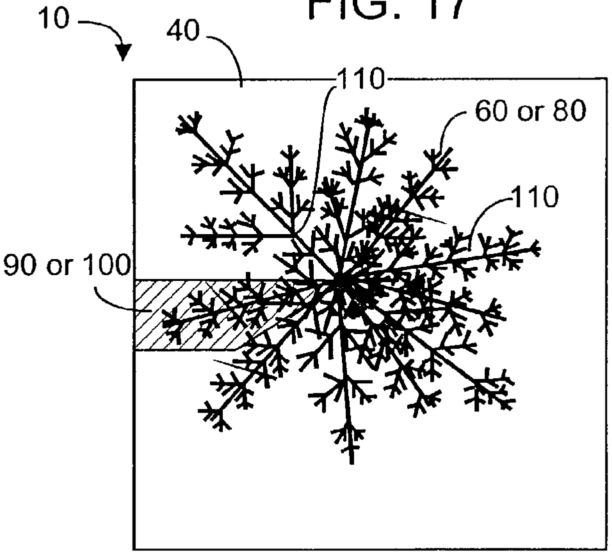
10
40
110
60 or 80
110
90 or 100

FIG. 18

FRACTAL ANTENNA GROUND COUNTERPOISE, GROUND PLANES, AND LOADING ELEMENTS AND MICROSTRIP PATCH ANTENNAS WITH FRACTAL STRUCTURE

RELATION TO PREVIOUSLY FILED PATENT APPLICATIONS

This application is a continuing application from applicant's co-pending patent application Ser. No. 09/677,645 entitled Fractal Antenna Ground Counterpoise, Ground Planes, And Loading Elements And Microstrip Patch Antennas With Fractal Structure, filed 3 Oct. 2000, which in turn is a continuing application of application Ser. No. 08/967, 375 entitled Fractal Antenna Ground Counterpoise, Ground Planes, And Loading Elements, filed 7 Nov. 1997, and from applicant's patent application Ser. No. 08/965,914 entitled Microstrip Patch Antennas With Fractal Structure, filed 7 Nov. 1997, issued as U.S. Pat. No. 6,127,977 (3 Oct. 2000). Applicant incorporate by reference herein his U.S. Pat. No. 6,104,349 (15 Aug. 2000) entitled Tuning Fractal Antennas and Fractal Resonators.

FIELD OF THE INVENTION

The present invention relates to antennas and resonators, and microstrip patch antennas, and specifically to designing and tuning non-Euclidian antenna ground radials, ground counterpoise or planes, top-loading elements, and antennas using such elements and to providing microstrip patch antennas with fractal structure elements.

BACKGROUND OF THE INVENTION

Antenna are used to radiate and/or receive typically electromagnetic signals, preferably with antenna gain, directivity, and efficiency. Practical antenna design traditionally involves trade-offs between various parameters, including antenna gain, size, efficiency, and bandwidth.

Antenna design has historically been dominated by Euclidean geometry. In such designs, the closed antenna area is directly proportional to the antenna perimeter. For example, if one doubles the length of an Euclidean square (or "quad") antenna, the enclosed area of the antenna quadruples. Classical antenna design has dealt with planes, circles, triangles, squares, ellipses, rectangles, hemispheres, paraboloids, and the like, (as well as lines). Similarly, resonators, typically capacitors ("C") coupled in series and/or parallel with inductors ("L"), traditionally are implemented with Euclidian inductors.

With respect to antennas, prior art design philosophy has been to pick a Euclidean geometric construction, e.g., a quad, and to explore its radiation characteristics, especially with emphasis on frequency resonance and power patterns. The unfortunate result is that antenna design has far too long concentrated on the ease of antenna construction, rather than on the underlying electromagnetics.

Many prior art antennas are based upon closed-loop or island shapes. Experience has long demonstrated that small sized antennas, including loops, do not work well, one reason being that radiation resistance ("R") decreases sharply when the antenna size is shortened. A small sized loop, or even a short dipole, will exhibit a radiation pattern of ½λ and ¼λ, respectively, if the radiation resistance R is not swamped by substantially larger ohmic ("O") losses. Ohmic losses can be minimized using impedance matching networks, which can be expensive and difficult to use. But although even impedance matched small loop antennas can exhibit 50% to 85% efficiencies, their bandwidth is inherently narrow, with very high Q, e.g., Q>50. As used herein, Q is defined as (transmitted or received frequency)/(3 dB bandwidth).

As noted, it is well known experimentally that radiation resistance R drops rapidly with small area Euclidean antennas. However, the theoretical basis is not generally known, and any present understanding (or misunderstanding) appears to stem from research by J. Kraus, noted in *Antennas* (Ed. 1), McGraw Hill, New York (1950), in which a circular loop antenna with uniform current was examined. Kraus' loop exhibited a gain with a surprising limit of 1.8 dB over an isotropic radiator as loop area fells below that of a loop having a 1λ-squared aperture. For small loops of area $A<\lambda^2/100$, radiation resistance R was given by:

$$R = K\cdot\left(\frac{A}{\lambda^2}\right)^2$$

where K is a constant, A is the enclosed area of the loop, and λ is wavelength. Unfortunately, radiation resistance R can all too readily be less than 1Ω for a small loop antenna.

From his circular loop research Kraus generalized that calculations could be defined by antenna area rather than antenna perimeter, and that his analysis should be correct for small loops of any geometric shape. Kraus' early research and conclusions that small-sized antennas will exhibit a relatively large ohmic resistance O and a relatively small radiation resistance R, such that resultant low efficiency defeats the use of the small antenna have been widely accepted. In fact, some researchers have actually proposed reducing ohmic resistance O to 0Ω by constructing small antennas from superconducting material, to promote efficiency.

As noted, prior art antenna and resonator design has traditionally concentrated on geometry that is Euclidean. However, one non-Euclidian geometry is fractal geometry. Fractal geometry may be grouped into random fractals, which are also termed chaotic or Brownian fractals and include a random noise components, such as depicted in FIG. **3**, or deterministic fractals such as shown in FIG. **1**C.

In deterministic fractal geometry, a self-similar structure results from the repetition of a design or motif (or "generator"), on a series of different size scales. One well known treatise in this field is *Fractals, Endlessly Repeated Geometrical Figures,* by Hans Lauwerier, Princeton University Press (1991), which treatise applicant refers to and incorporates herein by reference.

FIGS. **1**A–**2**D depict the development of some elementary forms of fractals. In FIG. **1**A, a base element **10** is shown as a straight line, although a curve could instead be used. In FIG. **1**B, a so-called Koch fractal motif or generator **20-1**, here a triangle, is inserted into base element **10**, to form a first order iteration ("N") design, e.g., N=1. In FIG. **1**C, a second order N=2 iteration design results from replicating the triangle motif **20-1** into each segment of FIG. **1**B, but where the **20-1'** version has been differently scaled, here reduced in size. As noted in the Lauwerier treatise, in its replication, the motif may be rotated, translated, scaled in dimension, or a combination of any of these characteristics. Thus, as used herein, second order of iteration or N=2 means the fundamental motif has been replicated, after rotation, translation, scaling (or a combination of each) into the first order iteration pattern. A higher order, e.g., N=3, iteration means a third fractal pattern has been generated by including yet another rotation; translation, and/or scaling of the first order motif.

In FIG. **1**D, a portion of FIG. **1**C has been subjected to a further iteration (N=3) in which scaled-down versions of the triangle motif **20-1** have been inserted into each segment of the left half of FIG. **1**C. FIGS. **2**A–2C follow what has been described with respect to FIGS. **1**A–1C, except that a rectangular motif **20-2** has been adopted. FIG. **2**D shows a pattern in which a portion of the left-hand side is an N=3 iteration of the **20-2** rectangle motif, and in which the center portion of the figure now includes another motif, here a **20-1** type triangle motif, and in which the right-hand side of the figure remains an N=2 iteration.

Traditionally, non-Euclidean designs including random fractals have been understood to exhibit antiresonance characteristics with mechanical vibrations. It is known in the art to attempt to use non-Euclidean random designs at lower frequency regimes to absorb, or at least not reflect sound due to the antiresonance characteristics. For example, M. Schroeder in *Fractals, Chaos. Power Laws* (1992), W. H. Freeman, New York discloses the use of presumably random or chaotic fractals in designing sound blocking diffusers for recording studios and auditoriums.

Experimentation with non-Euclidean structures has also been undertaken with respect to electromagnetic waves, including radio antennas. In one experiment, Y. Kim and D. Jaggard in *The Fractal Random Array*, Proc. IEEE 74, 1278–1280 (1986) spread-out antenna elements in a sparse microwave array, to minimize sidelobe energy without having to use an excessive number of elements. But Kim and Jaggard did not apply a fractal condition to the antenna elements, and test results were not necessarily better than any other techniques, including a totally random spreading of antenna elements. More significantly, the resultant array was not smaller than a conventional Euclidean design.

Prior art spiral antennas, cone antennas, and V-shaped antennas may be considered as a continuous, deterministic first order fractal, whose motif continuously expands as distance increases from a central point. A log-periodic antenna may be considered a type of continuous fractal in that it is fabricated from a radially expanding structure. However, log periodic antennas do not utilize the antenna perimeter for radiation, but instead rely upon an arc-like opening angle in the antenna geometry. Such opening angle is an angle that defines the size-scale of the log-periodic structure, which structure is proportional to the distance from the antenna center multiplied by the opening angle. Further, known log-periodic antennas are not necessarily smaller than conventional driven element-parasitic element antenna designs of similar gain.

Unintentionally, first order fractals have been used to distort the shape of dipole and vertical antennas to increase gain, the shapes being defined as a Brownian-type of chaotic fractals. See F. Landstorfer and R. Sacher, *Optimisation of Wire Antennas,* J. Wiley, New York (1985). FIG. **3** depicts three bent-vertical antennas developed by Landstorfer and Sacher through trial and error, the plots showing the actual vertical antennas as a function of x-axis and y-axis coordinates that are a function of wavelength. The “EF” and “BF” nomenclature in FIG. **3** refer respectively to end-fire and back-fire radiation patterns of the resultant bent-vertical antennas.

First order fractals have also been used to reduce horn-type antenna geometry, in which a double-ridge horn configuration is used to decrease resonant frequency. See J. Kraus in *Antennas,* McGraw Hill, New York (1885). The use of rectangular, box-like, and triangular shapes as impedance-matching loading elements to shorten antenna element dimensions is also known in the art.

Whether intentional or not, such prior art attempts to use a quasi-fractal or fractal motif in an antenna employ at best a first order iteration fractal. By first iteration it is meant that one Euclidian structure is loaded with another Euclidean structure in a repetitive fashion, using the same size for repetition. FIG. **1**C, for example, is not first order because the **20-1'** triangles have been shrunk with respect to the size of the first motif **20-1**.

So-called microstrip patch antennas have traditionally been fabricated as two spaced-apart metal surfaces separated by a small width dielectric. The sides are dimensioned typically one-quarter wavelength or one-half wavelength at the frequency of interest. One surface is typically a simple Euclidean structure such as a circle, a square, while the other side is a ground plane. Attempting to reduce the physical size of such an antenna for a given frequency typically results in a poor feedpoint match (e.g., to coaxial or other feed cable), poor radiation bandwidth, among other difficulties.

Prior art antenna design does not attempt to exploit multiple scale self-similarity of real fractals. This is hardly surprising in view of the accepted conventional wisdom that because such antennas would be anti-resonators, and/or if suitably shrunken would exhibit so small a radiation resistance R, that the substantially higher ohmic losses **0** would result in too low an antenna efficiency for any practical use. Further, it is probably not possible to mathematically predict such an antenna design, and high order iteration fractal antennas would be increasingly difficult to fabricate and erect, in practice. The use of fractals, especially higher order fractals, in fabricating microstrip patch antennas has not been investigated in the prior art.

FIGS. **4**A and **4**B depict respective prior art series and parallel type resonator configurations, comprising capacitors C and Euclidean inductors L. In the series configuration of FIG. **4**A, a notch-filter characteristic is presented in that the impedance from port A to port B is high except at frequencies approaching resonance, determined by $1/\sqrt{(LC)}$.

In the distributed parallel configuration of FIG. **4**B, a low-pass filter characteristic is created in that at frequencies below resonance, there is a relatively low impedance path from port A to port B, but at frequencies greater than resonant frequency, signals at port A are shunted to ground (e.g., common terminals of capacitors C), and a high impedance path is presented between port A and port B. Of course, a single parallel LC configuration may also be created by removing (e.g., short-circuiting) the rightmost inductor L and right two capacitors C, in which case port B would be located at the bottom end of the leftmost capacitor C.

In FIGS. **4**A and **4**B, inductors L are Euclidean in that increasing the effective area captured by the inductors increases with increasing geometry of the inductors, e.g., more or larger inductive windings or, if not cylindrical, traces comprising inductance. In such prior art configurations as FIGS. **4**A and **4**B, the presence of Euclidean inductors L ensures a predictable relationship between L, C and frequencies of resonance.

Applicant’s cited applications provide design methodologies to produce smaller-scale antennas that can exhibit at least as much gain, directivity, and efficiency as larger Euclidean counterparts. Such design approach should exploit the multiple scale self-similarity of real fractals, including $N \geqq 2$ iteration order fractals. Further, said application disclosed a non-Euclidean resonator whose presence in a resonating configuration can create frequencies of resonance beyond those normally presented in series and/or parallel LC configurations. Applicant's above-noted TUNING FRACTAL ANTENNAS AND FRACTAL RESONATORS patent disclosed devices and methods for tuning and/or adjusting such antennas and resonators. This patent further disclosed the use of non-Euclidean resonators whose presence in a resonating configuration could create frequencies of resonance beyond those normally presented in series and/or parallel LC configurations.

However, such antenna design approaches and tuning approaches should also be useable with vertical antennas, permitting the downscaling of one or more radial ground plane elements, and/or ground planes, and/or ground counterpoises, and/or top-hat loading elements. Further, such antenna design approaches and tuning approaches should also be useable with microstrip patch antennas and elements for such antennas. Thus, there is a need for a method by which microstrip patch antennas could be made smaller without sacrificing antenna bandwidth, while preserving good feedpoint impedance matching, and while maintaining acceptable gain and frequency characteristics.

The present invention provides such antennas, radial ground plane elements, ground planes, ground counterpoises, and top-hat loading elements, as well as methods for their design, and further provides such microstrip patch antennas, and elements for such antennas.

SUMMARY OF THE INVENTION

In one aspect, the present invention provides an antenna with a ground plane or ground counterpoise system that has at least one element whose shape, at least is part, is substantially a deterministic fractal of iteration order $N \geqq 1$. (The term "ground counterpoise" will be understood to include a ground plane, and/or at least one ground element.) Using fractal geometry, the antenna ground counterpoise has a self-similar structure resulting from the repetition of a design or motif (or "generator") that is replicated using rotation, and/or translation, and/or scaling. The fractal element will have x-axis, y-axis coordinates for a next iteration N+1 defined by $x_{N+1}=f(x_N, yb_N)$ and $y_{N+1}=g(x_N, y_N$, where x_N, y_N define coordinates for a preceding iteration, and where f(x,y) and g(x,y) are functions defining the fractal motif and behavior. In another aspect, a vertical antenna is top-loaded with a so-called top-hat assembly that includes at least one fractal element. A fractalized top-hat, assembly advantageously reduces resonant frequency, as well as the physical size and area required for the top-hat assembly.

In contrast to Euclidean geometric antenna design, deterministic fractal elements according to the present invention have a perimeter that is not directly proportional to area. For a given perimeter dimension, the enclosed area of a multi-iteration fractal will always be as small or smaller than the area of a corresponding conventional Euclidean element.

A fractal antenna has a fractal ratio limit dimension D given by log(L)/log(r), where L and r are one-dimensional antenna element lengths before and after fractalization, respectively.

As used herein, a fractal antenna perimeter compression parameter (PC) is defined as:

$$PC = \frac{\text{full-sized antenna element length}}{\text{fractal-reduced antenna element length}}$$

where:

$$PC = A \cdot \log[N(D+C)]$$

in which A and C are constant coefficients for a given fractal motif, N is an iteration number, and D is the fractal dimension, defined above.

Radiation resistance (R) of a fractal antenna decreases as a small power of the perimeter compression (PC), with a fractal loop or island always exhibiting a substantially higher radiation resistance than a small Euclidean loop antenna of equal size. In the present invention, deterministic fractals are used wherein A and C have large values, and thus provide the greatest and most rapid element-size shrinkage. A fractal antenna according to the present invention will exhibit an increased effective wavelength.

The number of resonant nodes of a fractal loop-shaped antenna increases as the iteration number N and is at least as large as the number of resonant nodes of an Euclidean island with the same area. Further, resonant frequencies of a fractal antenna include frequencies that are not harmonically related.

An antenna including a fractal ground counterpoise according to the present invention is smaller than its Euclidean counterpart but provides at least as much gain and frequencies of resonance and provides a reasonable termination impedance at its lowest resonant frequency. Such an antenna system can exhibit non-harmonically frequencies of resonance, a low Q and resultant good bandwidth, acceptable standing wave ratio ("SWR"), and a radiation impedance that is frequency dependent, and high efficiencies.

With respect to vertical antennas, the present invention enables such antennas to be realized with a smaller vertical element, and/or with smaller ground counterpoise, e.g., ground plane radial elements, and/or ground plane. The ground counterpoise element(s) are fractalized with $N \geqq 1$. In a preferred embodiment, the vertical element is also a fractal system, preferably comprising first and second spaced-apart fractal elements.

A fractal antenna system having a fractal ground counterpoise and a fractal vertical preferably is tuned according to applicant's above-referenced TUNING FRACTAL ANTENNAS AND FRACTAL RESONATORS patent, by placing an active (or driven) fractal antenna or resonator a distance Δ from a second conductor. Such disposition of the antenna and second conductor advantageously lowers resonant frequencies and widens bandwidth for the fractal antenna. In some embodiments, the fractal antenna and second conductor are non-coplanar and λ is the separation distance therebetween, preferably $\leqq 0.05\lambda$ for the frequency of interest ($1/\lambda$). In other embodiments, the fractal antenna and second conductive element may be planar, in which case λ a separation distance, measured on the common plane. In another embodiment, an antenna is loaded with a fractal "top-hat" assembly, which can provide substantial reduction in antenna size.

The second conductor may in fact be a second fractal antenna of like or unlike configuration as the active antenna.

Varying the distance Δ tunes the active antenna and thus the overall system. Further, if the second element, preferably a fractal antenna, is angularly rotated relative to the active antenna, resonant frequencies of the active antenna may be varied.

Providing a cut in the fractal antenna results in new and different resonant nodes, including resonant nodes having perimeter compression parameters, defined below, ranging from about three to ten. If desired, a portion of a fractal antenna may be cutaway and removed so as to tune the antenna by increasing resonance(s).

Tunable antenna systems with a fractal ground counterpoise need not be planar, according to the present invention. Fabricating the antenna system around a form such as a toroid ring, or forming the fractal antenna on a flexible substrate that is curved about itself results in field self-proximity that produces resonant frequency shifts. A fractal antenna and a conductive element may each be formed as a curved surface or even as a toroid-shape, and placed in sufficiently close proximity to each other to provide a useful tuning and system characteristic altering mechanism.

In the various embodiments, more than two elements may be used, and tuning may be accomplished by varying one or more of the parameters associated with one or more elements.

In a second aspect, the present invention provides a microstrip patch antenna comprising spaced-apart first and second conductive surfaces separated by a dielectric material. The dielectric material thickness preferably is substantially less than one wavelength for the frequency of interest.

At least one of the surfaces is fabricated to define a fractal pattern of first or higher iteration order. Overall dimensions of the surfaces may be reduced below the one-quarter to one-half wavelength commonly found in the prior art.

Radio frequency feedline coupling to the microstrip patch antenna may be made at a location on the antenna pattern structure, or through a conductive feedtab strip that may be fabricated along with the conductive pattern on one or both surfaces of the antenna. The resultant antenna may be sized smaller than a non-fractal counterpart (e.g., approximately one-eighth wavelength provides good performance at about 900 MHz.) while preserving good, preferably 50Ω, feed-point impedance. Further bandwidth can actually be increased, and resonant frequency lowered.

Components from the generally-described first and second aspects of the present invention may be combined.

Other features and advantages of the invention will appear from the following description in which the preferred embodiments have been set forth in detail, in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1**A depicts a base element for an antenna or an inductor, according to the prior art;

FIG. **1**B depicts a triangular-shaped Koch fractal motif, according to the prior art;

FIG. **1**C depicts a second-iteration fractal using the motif of FIG. **1**B, according to the prior art;

FIG. **1**D depicts a third-iteration fractal using the motif of FIG. **1**B, according to the prior art;

FIG. **2**A depicts a base element for an antenna or an inductor, according to the prior art;

FIG. **2**B depicts a rectangular-shaped Minkowski fractal motif, according to the prior art;

FIG. **2**C depicts a second-iteration fractal using the motif of FIG. **2**B, according to the prior art;

FIG. **2**D depicts a fractal configuration including a third-order using the motif of FIG. **2**B, as well as the motif of FIG. **1**B, according to the prior art;

FIG. **3** depicts bent-vertical chaotic fractal antennas, according to the prior art;

FIG. **4**A depicts a series L-C resonator, according to the prior art;

FIG. **4**B depicts a distributed parallel L-C resonator, according to the prior art;

FIG. **5**A depicts an Euclidean quad antenna system, according to the prior art;

FIG. **5**B depicts a second-order Minkowski island fractal quad antenna, according to the present invention;

FIG. **6** depicts an ELNEC-generated free-space radiation pattern for an MI-2 fractal antenna, according to the present invention;

FIG. **7**A depicts a Cantor-comb fractal dipole antenna, according to the present invention;

FIG. **7**B depicts a torn square fractal quad antenna, according to the present invention;

FIG. **7**C-**1** depicts a second iteration Minkowski (MI-2) printed circuit fractal antenna, according to the present invention;

FIG. **7**C-**2** depicts a second iteration Minkowski (MI-2) slot fractal antenna, according to the present invention;

FIG. **7**D depicts a deterministic dendrite fractal vertical antenna, according to the present invention;

FIG. **7**D-**1**A depicts a 0.25λ vertical antenna with three 0.25λ radial ground elements, according to the prior art;

FIG. **7**D-**1**B depicts the gain pattern for the antenna of FIG. **7**D-**1**A;

FIG. **7**D-**2**A depicts a 0.25λ vertical antenna with three fractal radial ground elements according to the present invention;

FIG. **7**D-**2**B depicts the gain pattern for the antenna of FIG. **7**D-**2**A;

FIG. **7**D-**3**A depicts a "top-hat" loaded antenna, according to the prior art;

FIG. **7**D-**3**B depicts the gain pattern for the antenna of FIG. **7**D-**3**A;

FIG. **7**D-**4**A depicts a ternary fractal "top-hat" loaded antenna, according the present invention;

FIG. **7**D-**4**B depicts the gain pattern for the antenna of FIG. **7**D-**4**A;

FIG. **7**D-**5** depicts an antenna having a fractal vertical element and fractal radial ground elements, according to the present invention;

FIG. **7**E depicts a third iteration Minkowski island (MI-3) fractal quad antenna, according to the present invention;

FIG. **7**F depicts a second iteration Koch fractal dipole, according to the present invention;

FIG. **7**G depicts a third iteration dipole, according to the present invention;

FIG. **7**H depicts a second iteration Minkowski fractal dipole, according to the present invention;

FIG. **7**I depicts a third iteration multi-fractal dipole, according to the present invention;

FIG. **8**A depicts a generic system in which a passive or active -electronic system communicates using a fractal antenna, according to the present invention;

FIG. **8**B depicts a communication system in which several fractal antennas including a vertical antenna with a fractal ground counterpoise are electronically selected for best performance, according to the present invention;

FIG. **8**C depicts a communication system in which electronically steerable arrays of fractal antennas are electronically selected for best performance, according to the present invention;

FIG. **9**A depicts fractal antenna gain as a function of iteration order N, according to the present invention;

FIG. **9**B depicts perimeter compression PC as a function of iteration order N for fractal antennas, according to the present invention;

FIG. **10**A depicts a fractal inductor for use in a fractal resonator, according to the present invention;

FIG. **10**B depicts a credit card sized security device utilizing a fractal resonator, according to the present invention;

FIG. **11**A depicts an embodiment in which a fractal antenna is spaced-apart a distance Δ from a conductor element to vary resonant properties and radiation characteristics of the antenna, according to the present invention;

FIG. **11**B depicts an embodiment in which a fractal antenna is coplanar with a ground plane and is spaced-apart a distance Δ' from a coplanar passive parasitic element to vary resonant properties and radiation characteristics of the antenna, according to the present invention;

FIG. **12**A depicts spacing-apart first and second fractal antennas a distance Δ to decrease resonance and create additional resonant frequencies for the active or driven antenna, according to the present invention;

FIG. **12**B depicts relative angular rotation between spaced-apart first and second fractal antennas Δ to vary resonant frequencies of the active or driven antenna, according to the present invention;

FIG. **13**A depicts cutting a fractal antenna or resonator to create different resonant nodes and to alter perimeter compression, according to the present invention;

FIG. **13**B depicts forming a non-planar fractal antenna or resonator on a flexible substrate that is curved to shift resonant frequency, apparently due to self-proximity electromagnetic fields, according to the present invention;

FIG. **13**C depicts forming a fractal antenna or resonator on a curved torroidal form to shift resonant frequency, apparently due to self-proximity electromagnetic fields, according to the present invention;

FIG. **14**A depicts forming a fractal antenna or resonator in which the conductive element is not attached to the system coaxial or other feedline, according to the present invention;

FIG. **14**B depicts a system similar to FIG. **14**A, but demonstrates that the driven fractal antenna may be coupled to the system coaxial or other feedline at any point along the antenna, according to the present invention;

FIG. **14**C depicts an embodiment in which a supplemental ground plane is disposed adjacent a portion of the driven fractal antenna and conductive element, forming a sandwich-like system, according to the present invention;

FIG. **14**D depicts an embodiment in which a fractal antenna system is tuned by cutting away a portion of the driven antenna, according to the present invention;

FIG. **15** depicts a communication system similar to that of FIG. **8**A, in which several fractal antennas are tunable and are electronically selected for best performance, according to the present invention;

FIG. **16** is a sideview of a microstrip patch antenna with at least one fractal element, according to the present invention;

FIG. **17** is a top plan view of an exemplary fractal element (a Sierpinski square gasket, including an optional feedtab, according to the present invention; and

FIG. **18** is a top plan view of an exemplary alternative fractal element (a diffusion limited aggregate), including an optional feed pad, according to the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

In overview, in one aspect, the present invention provides an antenna system with a fractal ground counterpoise, e.g., a counterpoise and/or ground plane and/or ground element having at least one element whose shape, at least is part, is substantially a fractal of iteration order $N \geqq 1$. The resultant antenna is smaller than its Euclidean counterpart, provides close to 50Ω termination impedance, exhibits at least as much gain and more frequencies of resonance than its Euclidean counterpart, including non-harmonically related frequencies of resonance, exhibits a low Q and resultant good bandwidth, acceptable SWR, a radiation impedance that is frequency dependent, and high efficiencies.

In another aspect, the present invention provides a microstrip patch antenna with at least one element whose shape, at least is part, is substantially a fractal of iteration order $N \geqq 1$. The resultant antenna is smaller than its Euclidean counterpart, provides close to 50Ω termination impedance, exhibits acceptable gain, increased bandwidth, and decreased resonant frequency than its Euclidean counterpart.

In contrast to Euclidean geometric antenna design, a fractal element including a fractal antenna ground counterpoise according to the present invention has a perimeter that is not directly proportional to area. For a given perimeter dimension, the enclosed area of a multi-iteration fractal area will always be at least as small as any Euclidean area.

Using fractal geometry, the ground element has a self-similar structure resulting from the repetition of a design or motif (or "generator"), which motif is replicated using rotation, translation, and/or scaling (or any combination thereof). The fractal portion of the element has x-axis, y-axis coordinates for a next iteration N+1 defined by $x_{N+1}=f(x_N, yb_N)$ and $y_{N+1}=g(x_N, y_N)$, where x_N, y_N are coordinates of a preceding iteration, and where f(x,y) and g(x,y) are functions defining the fractal motif and behavior.

For example, fractals of the Julia set may be represented by the form:

$$x_{N+1}=x_N^2-y_N^2+a$$

$$y_{N+1}=2x_N \cdot y_N=b$$

In complex notation, the above may be represented as:

$$z_{N+1}=z_N^2+c$$

Although it is apparent that fractals can comprise a wide variety of forms for functions f(x,y) and g(x,y), it is the iterative nature and the direct relation between structure or morphology on different size scales that uniquely distinguish f(x,y) and g(x,y) from non-fractal forms. Many references including the Lauwerier treatise set forth equations appropriate for f(x,y) and g(x,y).

Iteration (N) is defined as the application of a fractal motif over one size scale. Thus, the repetition of a single size scale of a motif is not a fractal as that term is used herein. Multi-fractals may of course be implemented, in which a motif is changed for different iterations, but eventually at least one motif is repeated in another iteration.

The first aspect of the present invention will now be described with reference to FIGS. **5A-15**, and then the second aspect will be described with reference to FIGS. **16–18**. As noted, elements of each aspect may be combined.

An overall appreciation of the present invention, and especially the first aspect thereof, may be obtained by comparing FIGS. **5**A and **5**B. FIG. **5**A shows a conventional Euclidean quad antenna **5** having a driven element **10** whose four sides are each 0.25λ long, for a total perimeter of 1λ, where λ is the frequency of interest.

Euclidean element **10** has an impedance of perhaps 130Ω, which impedance decreases if a parasitic quad element **20** is spaced apart on a boom **30** by a distance B of 0.1λ to 0.25λ. Parasitic element **20** is also sized S=0.25λ on a side, and its presence can improve directivity of the resultant two-element quad antenna. Element **10** is depicted in FIG. **5**A with heavier lines than element **20**, solely to avoid confusion in understanding the figure. Non-conductive spreaders **40** are used to help hold element **10** together and element **20** together.

Because of the relatively large drive impedance, driven element **10** is coupled to an impedance matching network or device **60**, whose output impedance is approximately 50Ω. A typically 50Ω coaxial cable **50** couples device **60** to a transceiver **70** or other active or passive electronic equipment **70**.

As used herein, the term transceiver shall mean a piece of electronic equipment that can transmit, receive, or transmit and receive an electromagnetic signal via an antenna, such as the quad antenna shown in FIG. **5**A or **5**B. As such, the term transceiver includes without limitation a transmitter, a receiver, a transmitter-receiver, a cellular telephone, a wireless telephone, a pager, a wireless computer local area network ("LAN") communicator, a passive resonant unit used by stores as part of an anti-theft system in which transceiver **70** contains a resonant circuit that is blown or not-blown by an electronic signal at time of purchase of the item to which transceiver **70** is affixed, resonant sensors and transponders, and the like.

Further, since antennas according to the present invention can receive incoming radiation and coupled the same as alternating current into a cable, it will be appreciated that fractal antennas may be used to intercept incoming light radiation and to provide a corresponding alternating current. For example, a photocell antenna defining a fractal, or indeed a plurality or array of fractals, would be expected to output more current in response to incoming light than would a photocell of the same overall array size.

FIG. **5**B depicts a fractal quad antenna **95**, designed to resonant at the same frequency as the larger prior art antenna **5** shown in FIG. **5**A. Driven element **100** is seen to be a second order fractal, here a so-called Minkowski island fractal, although any of numerous other fractal configurations could instead be used, including without limitation, Koch, torn square, Mandelbrot, Caley tree, monkey's swing, Sierpinski gasket, and Cantor gasket geometry.

If one were to measure to the amount of conductive wire or conductive trace comprising the perimeter of element **40**, it would be perhaps 40% greater than the 1.0λ for the Euclidean quad of FIG. **5**A. However, for fractal antenna **95**, the physical straight length of one element side KS will be substantially smaller, and for the N=2 fractal antenna shown in FIG. **5**B, KS≈0.13λ (in air), compared with K≈0.25λ for prior art antenna **5**.

However, although the actual perimeter length of element **100** is greater than the 1λ perimeter of prior art element **10**, the area within antenna element **100** is substantially less than the S^2 area of prior art element **10**. As noted, this area independence from perimeter is a characteristic of a deterministic fractal. Boom length B for antenna **95** will be slightly different from length B for prior art antenna **5** shown in FIG. **4**A. In FIG. **5**B, a parasitic element **120**, which preferably is similar to driven element **100** but need not be, may be attached to boom **130**. For ease of illustration FIG. **5**B does not depict non-conductive spreaders, such as spreaders **40** shown in FIG. **4**A, which help hold element **100** together and element **120** together. Further, for ease of understanding the figure, element **10** is drawn with heavier lines than element **120**, to avoid confusion in the portion of the figure in which elements **100** and **120** appear overlapped.

An impedance matching device **60** is advantageously unnecessary for the fractal antenna of FIG. **5**B, as the driving impedance of element **100** is about 50Ω, e.g., a perfect match for cable **50** if reflector element **120** is absent, and about 35Ω, still an acceptable impedance match for cable **50**, if element **120** is present. Antenna **95** may be fed by cable **50** essentially anywhere in element **100**, e.g., including locations X, Y, Z, among others, with no substantial change in the termination impedance. With cable **50** connected as shown, antenna **95** will exhibit horizontal polarization. If vertical polarization is desired, connection may be made as shown by cable **50'**. If desired, cables **50** and **50'** may both be present, and an electronic switching device **75** at the antenna end of these cables can short-out one of the cables. If cable **50** is shorted out at the antenna, vertical polarization results, and if instead cable **50'** is shorted out at the antenna, horizontal polarization results.

As shown by Table 3 herein, fractal quad **95** exhibits about 1.5 dB gain relative to Euclidean quad **10**. Thus, transmitting power output by transceiver **70** may be cut by perhaps 40% and yet the system of FIG. **5**B will still perform no worse than the prior art system of FIG. **5**A. Further, as shown by Table 1, the fractal antenna of FIG. **5**B exhibits more resonance frequencies than the antenna of FIG. **5**B, and also exhibits some resonant frequencies that are not harmonically related to each other. As shown by Table 3, antenna **95** has efficiency exceeding about 92% and exhibits an excellent SWR of about 1.2:1. As shown by Table 5, applicant's fractal quad antenna exhibits a relatively low value of Q. This result is surprising in view of conventional prior art wisdom to the effect that small loop antennas will exhibit high Q.

In short, that fractal quad **95** works at all is surprising in view of the prior art (mis)understanding as to the nature of radiation resistance R and ohmic losses O. Indeed, the prior art would predict that because the fractal antenna of FIG. **5**B is smaller than the conventional antenna of FIG. **5**A, efficiency would suffer due to an anticipated decrease in radiation resistance R. Further, it would have been expected that Q would be unduly high for a fractal quad antenna.

FIG. **6** is an ELNEC-generated free-space radiation pattern for a second-iteration Minkowski fractal antenna, an antenna similar to what is shown in FIG. **5**B with the parasitic element **120** omitted. The frequency of interest was 42.3 MHz, and a 1.5:1 SWR was used. In FIG. **6**, the outer ring represents 2.091 dBi, and a maximum gain of 2.091 dBi. (ELNEC is a graphics/PC version of MININEC, which is a PC version of NEC.) In practice, however, the data shown in FIG. **6** were conservative in that a gain of 4.8 dB above an isotropic reference radiator was actually obtained. The error in the gain figures associated with FIG. **6** presumably is due to roundoff and other limitations inherent in the ELNEC program. Nonetheless, FIG. **6** is believed to accurately depict the relative gain radiation pattern of a single element Minkowski (MI-2) fractal quad according to the present invention.

FIG. **7**A depicts a third iteration Cantor-comb fractal dipole antenna, according to the present invention. Generation of a Cantor-comb involves trisecting a basic shape, e.g., a rectangle, and providing a rectangle of one-third of the basic shape on the ends of the basic shape. The new smaller rectangles are then trisected, and the process repeated. FIG. **7**B is modeled after the Lauwerier treatise, and depicts a single element torn-sheet fractal quad antenna.

As described later herein, the fractal element shown in FIG. **7**B may be used as a ground counterpoise for an antenna system, for example, for a vertical antenna. In such application, the center conductor of cable **50** would be coupled to the lower end of the vertical antenna element (not shown, but which itself may be a fractal), and the ground shield of cable **50** would be coupled to the fractal element shown in FIG. **7**B. The fractal ground poise may be substantially smaller than a conventional 0.25λ ground system, without detriment to gain, coupling impedance, and vertical polarization characteristics of the antenna system.

FIG. **7**C-**1** depicts a printed circuit antenna, in which the antenna is fabricated using printed circuit or semiconductor fabrication techniques. For ease of understanding, the etched-away non-conductive portion of the printed circuit board **150** is shown cross-hatched, and the copper or other conductive traces **170** are shown without cross-hatching.

Applicant notes that while various corners of the Minkowski rectangle motif may appear to be touching in this and perhaps other figures herein, in fact no touching occurs. Further, it is understood that it suffices if an element according to the present invention is substantially a fractal. By this it is meant that a deviation of less than perhaps 10% from a perfectly drawn and implemented fractal will still provide adequate fractal-like performance, based upon actual measurements conducted by applicant.

The substrate **150** is covered by a conductive layer of material **170** that is etched away or otherwise removed in areas other than the fractal design, to expose the substrate **150**. The remaining conductive trace portion **170** defines a fractal antenna, a second iteration Minkowski slot antenna in FIG. **7**C-**1**. Substrate **150** may be a silicon wafer, a rigid or a flexible plastic-like material, perhaps Mylar™ material, or the non-conductive portion of a printed circuit board. Overlayer **170** may be deposited doped polysilicon for a semiconductor substrate **150**, or copper for a printed circuit board substrate.

If desired, the fractal structure shown in FIG. **7**C-**1** could be utilized as a fractal ground counterpoise for an antenna system, for example a vertical antenna. The fractal ground counterpoise may be fabricated using smaller dimensions than a conventional prior art system employing typically 0.25λ ground radials or elements. If the structure shown in FIG. **7**C-**1** is used as a ground counterpoise, the center lead of cable **50** would be coupled to the vertical element (not shown), and the ground shield would be coupled to the fractal structure shown.

FIG. **7**C-**2** depicts a slot antenna version of what was shown in FIG. **7**C-**2**, wherein the conductive portion **170** (shown cross-hatched in FIG. **7**C-**2**) surrounds and defines a fractal-shape of non-conductive substrate **150**. Electrical connection to the slot antenna is made with a coaxial or other cable **50**, whose inner and outer conductors make contact as shown.

In FIGS. **7**C-**1** and **7**C-**2**, the substrate or plastic-like material in such constructions can contribute a dielectric effect that may alter somewhat the performance of a fractal antenna by reducing resonant frequency, which increases perimeter compression PC.

Those skilled in the art will appreciate that by virtue of the relatively large amount of conducting material (as contrasted to a thin wire), antenna efficiency is promoted in a slot configuration. Of course a printed circuit board or substrate-type construction could be used to implement a non-slot fractal antenna, e.g, in which the fractal motif is fabricated as a conductive trace and the remainder of the conductive material is etched away or otherwise removed. Thus, in FIG. **7**C, if the cross-hatched surface now represents non-conductive material, and the non-cross hatched material represents conductive material, a printed circuit board or substrate-implemented wire-type fractal antenna results.

Printed circuit board and/or substrate-implemented fractal antennas are especially useful at frequencies of 80 MHz or higher, whereat fractal dimensions indeed become small A 2 M MI-3 fractal antenna (e.g., FIG. **7**E) will measure about 5.5" (14 cm) on a side KS, and an MI-2 fractal antenna (e.g., FIG. **5**B) will about 7" (17.5 cm) per side KS. As will be seen from FIG. **8**A, an MI-3 antenna suffers a slight loss in gain relative to an MI-2 antenna, but offers substantial size reduction.

Applicant has fabricated an MI-2 Minkowski island fractal antenna for operation in the 850–900 MHz cellular telephone band. The antenna was fabricated on a printed circuit board and measured about 1.2" (3 cm) on a side KS. The antenna was sufficiently small to fit inside applicant's cellular telephone, and performed as well as if the normal attachable “rubber-ducky” whip antenna were still attached. The antenna was found on the side to obtain desired vertical polarization, but could be fed anywhere on the element with 50Ω impedance still being inherently present. Applicant also fabricated on a printed circuit board an MI-3 Minkowski island fractal quad, whose side dimension KS was about 0.8" (2 cm), the antenna again being inserted inside the cellular telephone. The MI-3 antenna appeared to work as well as the normal whip antenna, which was not attached. Again, any slight gain loss in going from MI-2 to MI-3 (e.g., perhaps 1 dB loss relative to an MI-0 reference quad, or 3 dB los relative to an MI-2) is more than offset by the resultant shrinkage in size. At satellite telephone frequencies of 1650 MHz or so, the dimensions would be approximated halved again. FIGS. **8**A, **8**B and **8**C depict preferred embodiments for such antennas.

FIG. **7**D depicts a 2 M dendrite deterministic fractal antenna that includes a slight amount of randomness. The vertical arrays of numbers depict wavelengths relative to 0λ, at the lower end of the trunk-like element **200**. Eight radial-like elements **210** are disposed at 1.0λ, and various other elements are disposed vertically in a plane along the length of element **200**. The antenna was fabricated using 12 gauge copper wire and was found to exhibit a surprising 20 dBi gain, which is at least 10 dB better than any antenna twice the size of what is shown in FIG. **7**D. Although superficially the vertical of FIG. **7**D may appear analogous to a log-periodic antenna, a fractal vertical according to the present invention does not rely upon an opening angle, in stark contrast to prior art log periodic designs.

FIGS. **7**D-**1**A and **7**D-**1**B depict a conventional vertical antenna **5**, comprising a 0.25λ long vertical element **195**, and three 0.25λ long ground plane radials **205**. Antenna **5** is fed using coaxial cable **50** in conventional fashion, the antenna impedance being on the order of about 24Ω. Antenna efficiency may be improved by adding additional radial elements **205**, however doing so frequently requires more space than is conveniently available. In other configurations, a ground plane or counterpoise may be used without radials, e.g., earth or the metal body of an automobile in the case of a vehicular-mounted antenna. The 0° elevation angle azimuth plot of FIG. **7**D-**1**B depicts the undesirably large horizontal polarization components (the “figure eight” pattern) exhibited by this prior art vertical system, with vertical and total gain being about 1.45 dBi.

FIG. **7D-2**A depicts an antenna system **5** according to the present invention as including a vertical element **195** and a fractalized ground counterpoise system comprising, in this example, three dendrite fractal ground radials **215**. The ground radials are coupled to the ground shield on cable **50**, whereas the center lead of cable **50** is coupled to the vertical element **195**. Of course, other fractal configurations may be used instead, and a different number of ground radials may also be used.

In the azimuth plot of FIG. **7D-2**B, the elevation angle is 0°, and each fractal ground radial element is only about 0.087λ. The maximum gain, at the outermost ring in the figure, is 1.83 dBi and the input impedance is about 30Ω. Note in FIG. **7D-2**B that relatively little energy is radiated horizontally, and nearly all of the energy is radiated vertically, a desirable characteristic for a vertical antenna. It will be appreciated that the 0.087λ dimensions of fractal ground plane elements **215** are substantially physically smaller than the 0.25λ elements **205** in the prior art system of FIG. **7D-1**A. However, the radiation pattern for the system of FIG. **7D-2**A is actually better than that of the larger prior art system.

FIG. **7D-3**A depicts a so-called “top-hat” loaded vertical antenna **5**, according to the prior art. Antenna **5** includes a vertical element **195** and, in the example shown, a top-hat assembly comprising three spokes **207** located at the antenna top. The antenna is fed in conventional fashion with coaxial cable **50**. FIG. **7D-3**B depicts the radiation pattern for the conventional top-hat loaded antenna of FIG. **7D-3**A.

FIG. **7D-4**A depicts a “top-hat” antenna **5** that includes a vertical element **195** whose top is loaded by a top-hat assembly including fractalized radial spokes **215**. Antenna **5** may be fed in conventional fashion by coaxial cable **50**. For the same vertical length of element **195** as was used in FIG. **7D-3**A, the use of fractal radial spokes **215** advantageously decreases resonant frequency by 35%. In addition, the size of the “top-hat” assembly may be reduced by about 35%, and the area required for the “top-hat” assembly may be reduced by about 35%. These reductions are advantageous in that the fractalized top-hat antenna of FIG. **7D-4**A can require less material to fabricate, thus reducing manufacturing cost, weight, and wind resistance, relative to a prior art top-hat configuration. According to the present invention, it suffices if at least one of the elements in the top-hat assembly has a physical shaped defined at least in part by a fractal. Of course, more or less than three spokes may be used, and other fractal configurations may also be used, including combinations of fractal and non-fractal elements, as well as different types of fractal elements.

FIG. **7D-4**B represents the radiation pattern for the fractalized top-hat antenna of FIG. **7D-4**A. A comparison of FIGS. **7D-4**B and **7D-3**B confirms that there is no real performance penalty associated with using the fractalized configuration. Thus, the above-noted savings in cost, weight, and wind resistance are essentially penalty free.

FIG. **7D-5** depicts an antenna system according to the present invention, in which fractal ground elements **215** and a fractal vertical element **197** are both used. Fractal antenna elements **215** are preferably about 0.087λ, and element **197** is about λ/12. Fractal vertical element **197** preferably comprises a pair of spaced-apart elements such as generally described with respect to FIGS. **11**A, **12**A, **12**B, **13**B, **14**A, **14**B, and **14**C. It is to be understood, however, that the salient feature of element **197** in FIG. **7D-3** is not its specific shape, but rather that it defines a fractal, and preferably a pair of spaced-apart fractal elements. It is solely for ease of illustration that the fractal elements shown in FIGS. **7D-3**, **11**A, **12**A, **12**B, **13**B, **14**A, **14**B, **14**C, and **14**D are similarly drawn. Further, the fractal-fractal antenna system shown in FIG. **7D-3** is preferably tuned by varying the spaced-apart distance Δ, and/or by rotating the spaced-apart elements relative to one another, and/or by forming a “cut” in an element, as described hereinafter with respect to various of FIGS. **11**A, **12**A, **12**B, **13**B, **14**A, **14**B, **14**C and **14**D.

FIG. **7**E depicts a third iteration Minkowski island quad antenna (denoted herein as MI-3). The orthogonal line segments associated with the rectangular Minkowski motif make this configuration especially acceptable to numerical study using ELNEC and other numerical tools using moments for estimating power patterns, among other modeling schemes. In testing various fractal antennas, applicant formed the opinion that the right angles present in the Minkowski motif are especially suitable for electromagnetic frequencies.

With respect to the MI-3 fractal of FIG. **7**E, applicant discovered that the antenna becomes a vertical if the center led of coaxial cable **50** is connected anywhere to the fractal, but the outer coaxial braid-shield is left unconnected at the antenna end. (At the transceiver end, the outer shield is connected to ground.) Not only do fractal antenna islands perform as vertical antennas when the center conductor of cable **50** is attached to but one side of the island and the braid is left ungrounded at the antenna, but resonance frequencies for the antenna so coupled are substantially reduced. For example, a 2" (5 cm) sized MI-3 fractal antenna resonated at 70 MHz when so coupled, which is equivalent to a perimeter compression PC≈20.

FIG. **7**F depicts a second iteration Koch fractal dipole, and FIG. **7**G a third iteration dipole. FIG. **7**H depicts a second iteration Minkowski fractal dipole, and FIG. **71** a third iteration multi-fractal dipole. Depending upon the frequencies of interest, these antennas may be fabricated by bending wire, or by etching or otherwise forming traces on a substrate. Each of these dipoles provides substantially 50Ω termination impedance to which coaxial cable **50** may be directly coupled without any impedance matching device. It is understood in these figures that the center conductor of cable **50** is attached to one side of the fractal dipole, and the braid outer shield to the other side.

A fractal ground counterpoise may be fabricated using fractal element as shown in any (or all) of FIGS. **7**E-**7**I. Thus, in FIGS. **7D-2**A and **7D-3**, fractal ground radial elements **215** are understood to depict any fractal of iteration order N≧1. Further, such fractals may, but need not be, defined by an opening angle.

FIG. **8**A depicts a generalized system in which a transceiver **500** is coupled to a fractal antenna system **510** to send electromagnetic radiation **520** and/or receive electromagnetic radiation **540**. A second transceiver **600** shown equipped with a conventional whip-like vertical antenna **610** also sends electromagnetic energy **630** and/or receives electromagnetic energy **540**.

Fractal antenna system **510** may include a fractal ground counterpoise and/or fractal antenna element, as described earlier herein, and/or a microstrip patch antenna with fractal structure, as described later herein. As noted in the case of a vertical antenna element, the overall size of the resulting antenna system is substantially smaller than what may be achieved with a prior art ground counterpoise system. Further, the fractal ground counterpoise system may be fabricated on a flexible substrate that is rolled, curved, or otherwise formed to fit within a case such as contains transceiver **500**. The resultant antenna ground system exhibits improved efficiency and power distribution pattern relative to a prior art system that may somehow be fit into an equivalent amount of area.

If transceivers **500**, **600** are communication devices such as transmitter-receivers, wireless telephones, pagers, or the like, a communications repeating unit such as a satellite **650** and/or a ground base repeater unit **660** coupled to an antenna **670**, or indeed to a fractal antenna according to the present invention, may be present.

Alternatively, antenna **510** in transceiver **500** could be a passive LC resonator fabricated on an integrated circuit microchip, or other similarly small sized substrate, attached to a valuable item to be protected. Configurations such as shown in exemplary FIGS. **16**–**18** may also be fabricated for antenna **510**, according to the present invention. Transceiver **600**, or indeed unit **660** would then be an electromagnetic transmitter outputting energy at the frequency of resonance, a unit typically located near the cash register checkout area of a store or at an exit. Depending upon whether fractal antenna-resonator **510** is designed to "blow" (e.g., become open circuit) or to "short" (e.g., become a close circuit) in the transceiver **500** will or will not reflect back electromagnetic energy **540** or **6300** to a receiver associated with transceiver **600**. In this fashion, the unauthorized relocation of antenna **510** and/or transceiver **500** can be signaled by transceiver **600**.

FIG. **8**B depicts a transceiver **500** equipped with a plurality of fractal antennas, here shown as **510**A, **510**B, **510**C and **510**D coupled by respective cables **50**A, **50**B, **50**C, **50**D to electronics **600** within unit **500**. In the embodiment shown, one of more of these antenna elements is are fabricated on a conformal, flexible substrate **150**, e.g., Mylar™ material or the like, upon which the antennas per se may be implemented by printing fractal patterns using conductive ink, by copper deposition, among other methods including printed circuit board and semiconductor fabrication techniques. A flexible such substrate may be conformed to a rectangular, cylindrical or other shape as necessary.

In the embodiment of FIG. **8**B, unit **500** is a handheld transceiver, and antennas **510**A, **510**B, **510**C, **510**D preferably are fed for vertical polarization, as shown. Element **510**D may, for example, be a fractal ground counterpoise system for a vertical antenna element, shown in phantom as element **193** (which element may itself be a fractal to further reduce dimensions).

An electronic circuit **610** is coupled by cables **50**A, **50**B, **50**C to the antennas, and samples incoming signals to discern which fractal antenna system, e.g., **510**A, **510**B, **510**C, **510**D is presently most optimally aligned with the transmitting station, perhaps a unit **600** or **650** or **670** as shown in FIG. **8**A. This determination may be made by examining signal strength from each of the antennas. An electronic circuit **620** then selects the presently best oriented antenna, and couples such antenna to the input of the receiver and output of the transmitter portion, collectively **630**, of unit **500**. It is understood that the selection of the best antenna is dynamic and can change as, for example, a user of **500** perhaps walks about holding the unit, or the transmitting source moves, or due to other changing conditions. In a cellular or a wireless telephone application, the result is more reliable communication, with the advantage that the fractal antennas can be sufficiently small-sized as to fit totally within the casing of unit **500**. Further, if a flexible substrate is used, the antennas may be wrapped about portions of the internal casing, as shown.

An additional advantage of the embodiment of FIG. **8**B is that the user of unit **500** may be physically distanced from the antennas by a greater distance that if a conventional external whip antenna were used. Although medical evidence attempting to link cancer with exposure to electromagnetic radiation from handheld transceivers is still inconclusive, the embodiment of FIG. **8**B appears to minimize any such risk. Although FIG. **8**B depicts a vertical antenna **193** and a fractal ground counterpoise **510**D, it is understood that antenna **193** could represent a cellular antenna on a motor vehicle, the ground poise for which is fractal unit **510**D. Further, as noted, vertical element **193** may itself be a fractal.

FIG. **8**C depicts yet another embodiment wherein some or all of the antenna systems **510**A, **510**B, **510**C may include electronically steerable arrays, including arrays of fractal antennas of differing sizes and polarization orientations. Antenna system **510**C, for example may include similarly designed fractal antennas, e.g., antenna F-**3** and F-**4**, which are differently oriented from each other. Other antennas within system **510**C may be different in design from either of F-**3**, F-**4**. Fractal antenna F-**1** may be a dipole for example. Leads from the various antennas in system **510**C may be coupled to an integrated circuit **690**, mounted on substrate **150**. Circuit **690** can determine relative optimum choice between the antennas comprising system **510**C, and output via cable **50**C to electronics **660** associated with the transmitter and/or receiver portion **630** of unit **630**. Of course, the embodiment of FIG. **8**C could also include the vertical antenna element **193** and fractal ground counterpoise **510**D, depicted in FIG. **8**B.

Another antenna system **510**B may include a steerable array of identical fractal antennas, including fractal antenna F-**5** and F-**6**. An integrated circuit **690** is coupled to each of the antennas in the array, and dynamically selects the best antenna for signal strength and coupled such antenna via cable **50**B to electronics **600**. A third antenna system **510**A may be different from or identical to either of system **510**B and **510**C.

Although FIG. **8**C depicts a unit **500** that may be handheld, unit **500** could in fact be a communications system for use on a desk or a field mountable unit, perhaps unit **660** as shown in FIG. **8**A.

For ease of antenna matching to a transceiver load, resonance of a fractal antenna was defined as a total impedance falling between about 20Ω to 200Ω, and the antenna was required to exhibit medium to high Q, e.g., frequency/Δfrequency. In practice, applicants' various fractal antennas were found to resonate in at least one position of the antenna feedpoint, e.g., the point at which coupling was made to the antenna. Further, multi-iteration fractals according to the present invention were found to resonate at multiple frequencies, including frequencies that were non-harmonically related.

Contrary to conventional wisdom, applicant found that island-shaped fractals (e.g., a closed loop-like configuration) do not exhibit significant drops in radiation resistance R for decreasing antenna size. As described herein, fractal antennas were constructed with dimensions of less than 12" across (30.48 cm) and yet resonated in a desired 60 MHz to 100 MHz frequency band.

Applicant further discovered that antenna perimeters do not correspond to lengths that would be anticipated from measured resonant frequencies, with actual lengths being longer than expected. This increase in element length appears to be a property of fractals as radiators, and not a result of geometric construction. A similar lengthening effect was reported by Pfeiffer when constructing a full-sized quad antenna using a first order fractal, see A. Pfeiffer, *The Pfeiffer Quad Antenna System,* QST, p. 28–32 (March 1994).

If L is the total initial one-dimensional length of a fractal pre-motif application, and r is the one-dimensional length post-motif application, the resultant fractal dimension D (actually a ratio limit) is:

$$D=\log(L)/\log(r)$$

With reference to FIG. **1**A, for example, the length of FIG. **1**A represents L, whereas the sum of the four line segments comprising the Koch fractal of FIG. **1**B represents r.

Unlike mathematical fractals, fractal antennas are not characterized solely by the ratio D. In practice D is not a good predictor of how much smaller a fractal design antenna may be because D does not incorporate the perimeter lengthening of an antenna radiating element.

Because D is not an especially useful predictive parameter in fractal antenna design, a new parameter "perimeter compression" ("PC") shall be used, where:

$$PC=\frac{\text{full-sized antenna element length}}{\text{fractal-reduced antenna element length}}$$

In the above equation, measurements are made at the fractal-resonating element's lowest resonant frequency. Thus, for a full-sized antenna according to the prior art PC=1, while PC=3 represents a fractal antenna according to the present invention, in which an element side has been reduced by a factor of three.

Perimeter compression may be empirically represented using the fractal dimension D as follows:

$$PC=A\cdot\log[N(D+C)]$$

where A and C are constant coefficients for a given fractal motif, N is an iteration number, and D is the fractal dimension, defined above.

It is seen that for each fractal, PC becomes asymptotic to a real number and yet does not approach infinity even as the iteration number N becomes very large. Stated differently, the PC of a fractal radiator asymptotically approaches a non-infinite limit in a finite number of fractal iterations. This result is not a representation of a purely geometric fractal.

That some fractals are better resonating elements than other fractals follows because optimized fractal antennas approach their asymptotic PCs in fewer iterations than non-optimized fractal antennas. Thus, better fractals for antennas will have large values for A and C, and will provide the greatest and most rapid element-size shrinkage. Fractal used may be deterministic or chaotic. Deterministic fractals have a motif that replicates at a 100% level on all size scales, whereas chaotic fractals include a random noise component.

Applicant found that radiation resistance of a fractal antenna decreases as a small power of the perimeter compression (PC), with a fractal island always exhibiting a substantially higher radiation resistance than a small Euclidean loop antenna of equal size.

Further, it appears that the number of resonant nodes of a fractal island increase as the iteration number (N) and is always greater than or equal to the number of resonant nodes of an Euclidean island with the same area. Finally, it appears that a fractal resonator has an increased effective wavelength.

The above findings will now be applied to experiments conducted by applicant with fractal resonators shaped into closed-loops or islands. Prior art antenna analysis would predict no resonance points, but as shown below, such is not the case.

A Minkowski motif is depicted in FIGS. **2**B–**2**D, **5**B, **7**C and **7**E. The Minkowski motif selected was a three-sided box (e.g., **20-2** in FIG. **2**B) placed atop a line segment. The box sides may be any arbitrary length, e.g, perhaps a box height and width of 2 units with the two remaining base sides being of length three units (see FIG. **2**B). For such a configuration, the fractal dimension D is as follows:

$$D=\frac{\log(L)}{\log(r)}=\frac{\log(12)}{\log(8)}=\frac{1.08}{0.90}=1.20$$

It will be appreciated that D=1.2 is not especially high when compared to other deterministic fractals.

Applying the motif to the line segment may be most simply expressed by a piecewise function f(x) as follows:

$$\begin{array}{ll} f(x)=0 & 0\geq x\geq\frac{3x_{\max}}{8} \\ f(x)=\frac{1}{4x_{\max}} & \frac{3x_{\max}}{8}\geq x\geq\frac{5x_{\max}}{8} \\ f(x)=0 & \frac{5x_{\max}}{8}\geq x\geq x_{\max} \end{array}$$

where x_{max} is the largest continuous value of x on the line segment.

A second iteration may be expressed as $f(x)_2$ relative to the first iteration $f(x)_1$ by:

$$f(x)_2=f(x)_1+f(x)$$

where x_{max} is defined in the above-noted piecewise function. Note that each separate horizontal line segment will have a different lower value of x and x_{max}. Relevant offsets from zero may be entered as needed, and vertical segments may be "boxed" by 90° rotation and application of the above methodology.

As shown by FIGS. **5**B and **7**E, a Minkowski fractal quickly begins to appear like a Moorish design pattern. However, each successive iteration consumes more perimeter, thus reducing the overall length of an orthogonal line segment. Four box or rectangle-like fractals of the same iteration number N may be combined to create a Minkowski fractal island, and a resultant "fractalized" cubical quad.

An ELNEC simulation was used as a guide to far-field power patterns, resonant frequencies, and SWRs of Minkowski Island fractal antennas up to iteration N=2. Analysis for N>2 was not undertaken due to inadequacies in the test equipment available to applicant.

The following tables summarize applicant's ELNEC simulated fractal antenna designs undertaken to derive lowest frequency resonances and power patterns, to and including iteration N=2. All designs were constructed on the x,y axis, and for each iteration the outer length was maintained at 42" (106.7 cm).

Table 1, below, summarizes ELNEC-derived far field radiation patterns for Minkowski island quad antennas for each iteration for the first four resonances. In Table 1, each iteration is designed as MI-N for Minkowski Island of iteration N. Note that the frequency of lowest resonance decreased with the fractal Minkowski Island antennas, as compared to a prior art quad antenna. Stated differently, for a given resonant frequency, a fractal Minkowski Island antenna will be smaller than a conventional quad antenna.

TABLE 1

Antenna	Res. Freq. MHz	Gain (dBi)	SWR	PC (for 1st)	Direction
Ref. Quad	76	3.3	2.5	1	Broadside
	144	2.8	5.3	—	Endfire
	220	3.1	5.2	—	Endfire
	294	5.4	4.5	—	Endfire
MI-1	55	2.6	1.1	1.38	Broadside
	101	3.7	1.4	—	Endfire
	142	3.5	5.5	—	Endfire
	198	2.7	3.3	—	Broadside
MI-2	43.2	2.1	1.5	1.79	Broadfire
	85.5	4.3	1.8	—	Endfire
	102	2.7	4.0	—	Endfire
	116	1.4	5.4	—	Broadside

It is apparent from Table 1 that Minkowski island fractal antennas are multi-resonant structures having virtually the same gain as larger, full-sized conventional quad antennas. Gain figures in Table 1 are for "free-space" in the absence of any ground plane, but simulations over a perfect ground at 1λ yielded similar gain results. Understandably, there will be some inaccuracy in the ELNEC results due to round-off and undersampling of pulses, among other factors.

Table 2 presents the ratio of resonant ELNEC-derived frequencies for the first four resonance nodes referred to in Table 1.

TABLE 2

Antenna	SWR	SWR	SWR	SWR
Ref. Quad (MI-0)	1:1	1:1.89	1:2.89	3.86:1
MI-1	1:1	1:1.83	1;2.58	3.6:1
MI-2	1:1	2.02:1	2.41:1	2.74:1

Tables 1 and 2 confirm the shrinking of a fractal-designed antenna, and the increase in the number of resonance points. In the above simulations, the fractal MI-2 antenna exhibited four resonance nodes before the prior art reference quad exhibited its second resonance. Near fields in antennas are very important, as they are combined in multiple-element antennas to achieve high gain arrays. Unfortunately, programming limitations inherent in ELNEC preclude serious near field investigation. However, as described later herein, applicant has designed and constructed several different high gain fractal arrays that exploit the near field.

Applicant fabricated three Minkowski Island fractal antennas from aluminum #8 and/or thinner #12 galvanized groundwire. The antennas were designed so the lowest operating frequency fell close to a desired frequency in the 2 M (144 MHz) amateur radio band to facilitate relative gain measurements using 2 M FM repeater stations. The antennas were mounted for vertical polarization and placed so their center points were the highest practical point above the mounting platform. For gain comparisons, a vertical ground plane having three reference radials, and a reference quad were constructed, using the same sized wire as the fractal antenna being tested. Measurements were made in the receiving mode.

Multi-path reception was minimized by careful placement of the antennas. Low height effects were reduced and free space testing approximated by mounting the antenna test platform at the edge of a third-store window, affording a 3.5λ height above ground, and line of sight to the repeater, 45 miles (28 Km) distant. The antennas were stuck out of the window about 0.8λ from any metallic objects and testing was repeated on five occasions from different windows on the same floor, with test results being consistent within ½ dB for each trial.

Each antenna was attached to a short piece of 9913 50Ω coaxial cable, fed at right angles to the antenna. A 2 M transceiver was coupled with 9913 coaxial cable to two precision attenuators to the antenna under test. The transceiver S-meter was coupled to a volt-ohm meter to provide signal strength measurements The attenuators were used to insert initial threshold to avoid problems associated with non-linear S-meter readings, and with S-meter saturation in the presence of full squelch quieting.

Each antenna was quickly switched in for volt-ohmmeter measurement, with attenuation added or subtracted to obtain the same meter reading as experienced with the reference quad. All readings were corrected for SWR attenuation. For the reference quad, the SWR was 2.4:1 for 120Ω impedance, and for the fractal quad antennas SWR was less than 1.5:1 at resonance. The lack of a suitable noise bridge for 2 M precluded efficiency measurements for the various antennas. Understandably, anechoic chamber testing would provide even more useful measurements.

For each antenna, relative forward gain and optimized physical orientation were measured. No attempt was made to correct for launch-angle, or to measure power patterns other than to demonstrate the broadside nature of the gain. Difference of ½ dB produced noticeable S-meter deflections, and differences of several dB produced substantial meter deflection. Removal of the antenna from the receiver resulted in a 20^{+} dB drop in received signal strength. In this fashion, system distortions in readings were cancelled out to provide more meaningful results. Table 3 summarizes these results.

TABLE 3

Antenna	PC	PL	SWR	Cor. Gain (dB)	Sidelength (λ)
Quad	1	1	2.4:1	0	0.25
1/4 wave	1	—	1.5:1	−1.5	0.25
MI-1	1.3	1.2	1.3:1	1.5	0.13
MI-2	1.9	1.4	1.3:1	1.5	0.13
MI-3	2.4	1.7	1:1	−1.2	0.10

It is apparent from Table 3 that for the vertical configurations under test, a fractal quad according to the present invention either exceeded the gain of the prior art test quad, or had a gain deviation of not more than 1 dB from the test quad. Clearly, prior art cubical (square) quad antennas are not optimized for gain. Fractally shrinking a cubical quad by a factor of two will increase the gain, and further shrinking will exhibit modest losses of 1–2 dB.

Versions of a MI-2 and MI-3 fractal quad antennas were constructed for the 6 M (50 MHz) radio amateur band. An RX 50Ω noise bridge was attached between these antennas and a transceiver. The receiver was nulled at about 54 MHz and the noise bridge was calibrated with 5Ω and 10Ω resistors. Table 4 below summarizes the results, in which almost no reactance was seen.

TABLE 4

Antenna	SWR	Z(Ω)	O(Ω)	E(%)
Quad (MI-0)	2.4:1	120	5–10	92–96
MI-2	1.2:1	60	≦5	≧92
MI-3	1.1:1	55	≦5	≧91

In Table 4, efficiency (E) was defined as 100%*(R/Z), where Z was the measured impedance, and R was Z minus ohmic impedance and reactive impedances (0). As shown in Table 4, fractal MI-2 and MI-3 antennas with their low ≦1.2:1 SWR and low ohmic and reactive impedance provide extremely high efficiencies, $90^{+}\%$. These findings are indeed surprising in view of prior art teachings stemming from early Euclidean small loop geometries. In fact, Table 4 strongly suggests that prior art associations of low radiation impedances for small loops must be abandoned in general, to be invoked only when discussing small Euclidean loops. Applicant's MI-3 antenna was indeed micro-sized, being dimensioned at about 0.1λ per side, an area of about $\lambda^2/1{,}000$, and yet did not signal the onset of inefficiency long thought to accompany smaller sized antennas.

However the 6M efficiency data do not explain the fact that the MI-3 fractal antenna had a gain drop of almost 3 dB relative to the MI-2 fractal antenna. The low ohmic impedances of ≦5Ω strongly suggest that the explanation is other than inefficiency, small antenna size notwithstanding. It is quite possible that near field diffraction effects occur at higher iterations that result in gain loss. However, the smaller antenna sizes achieved by higher iterations appear to warrant the small loss in gain.

Using fractal techniques, however, 2 M quad antennas dimensioned smaller than 3" (7.6 cm) on a side, as well as 20 M (14 MHz) quads smaller than 3' (1 m) on a side can be realized. Economically of greater interest, fractal antennas constructed for cellular telephone frequencies (850 MHz) could be sized smaller than 0.5" (1.2 cm). As shown by FIGS. **8**B and **8**C, several such antenna, each oriented differently could be fabricated within the curved or rectilinear case of a cellular or wireless telephone, with the antenna outputs coupled to a circuit for coupling to the most optimally directed of the antennas for the signal then being received. The resultant antenna system would be smaller than the "rubber-ducky" type antennas now used by cellular telephones, but would have improved characteristics as well.

Similarly, fractal-designed antennas could be used in handheld military walkie-talkie transceivers, global positioning systems, satellites, transponders, wireless communication and computer networks, remote and/or robotic control systems, among other applications. Although the fractal Minkowski island antenna has been described herein, other fractal motifs are also useful, as well as non-island fractal configurations.

Table 5 demonstrates bandwidths ("BW") and multi-frequency resonances of the MI-2 and MI-3 antennas described, as well as Qs, for each node found for 6 M versions between 30 MHz and 175 MHz. Irrespective of resonant frequency SWR, the bandwidths shown are SWR 3:1 values. Q values shown were estimated by dividing resonant frequency by the 3:1 SWR BW. Frequency ratio is the relative scaling of resonance nodes.

TABLE 5

Antenna	Freq. (MHz)	Freq. Ratio	SWR	3:1 BW	Q
MI-3	53.0	1	1:1	6.4	8.3
	80.1	1.5:1	1.1:1	4.5	17.8
	121.0	2.3:1	2.4:1	6.8	17.7
MI-2	54.0	1	1:1	3.6	15.0
	95.8	1.8:1	1.1:1	7.3	13.1
	126.5	2.3:1	2.4:1	9.4	13.4

The Q values in Table 5 reflect that MI-2 and MI-3 fractal antennas are multiband. These antennas do not display the very high Qs seen in small tuned Euclidean loops, and there is a lack of a mathematical application to electromagnetics to predict these resonances or Qs. One approach might be to estimate scalar and vector potentials in Maxwell's equations by regarding each Minkowski Island iteration as a series of vertical and horizontal line segments with offset positions. Summation of these segments will lead to a Poynting vector calculation and power pattern that may be especially useful in better predicting fractal antenna characteristics and optimized shapes. In practice, actual Minkowski Island fractal antennas seem to perform slightly better than their ELNEC predictions, most likely due to inconsistencies in ELNEC modeling or ratios of resonant frequencies, PCs, SWRs and gains.

Those skilled in the art will appreciate that fractal multiband antenna arrays may also be constructed. Such arrays will be smaller and present less wind area than their Euclidean counterparts, and are mechanically rotatable with a smaller antenna rotator. Fractal antenna configurations using other than Minkowski islands or loops may be implemented. Table 6 shows the highest iteration number N for other fractal configurations that were found by applicant to resonant on at least one frequency.

TABLE 6

Fractal	Maximum Iteration
Koch	5
Torn Square	4
Minkowski	3
Mandelbrot	4
Caley Tree	4
Monkey's Swing	3
Sierpinski Gasket	3
Cantor Gasket	3

FIG. **9**A depicts gain relative to an Euclidean quad (e.g., an MI-0) configuration as a function of iteration value N. (It is understood that an Euclidean quad exhibits 1.5 dB gain relative to a standard reference dipole.) For first and second order iterations, the gain of a fractal quad increases relative to an Euclidean quad. However, beyond second order, gain drops off relative to an Euclidean quad. Applicant believes that near field electromagnetic energy diffraction-type cancellations may account for the gain loss for N>2. Possibly the far smaller areas found in fractal antennas according to the present invention-bring this diffraction phenomenon into sharper focus.

n practice, applicant could not physically bend wire for a 4th or 5th iteration 2 M Minkowski fractal antenna, although at lower frequencies the larger antenna sizes would not present this problem. However, at higher frequencies, printed circuitry techniques, semiconductor fabrication techniques as well as machine-construction could readily produce N=4, N=5, and higher order iterations fractal antennas.

In practice, a Minkowski island fractal antenna should reach the theoretical gain limit of about 1.7 dB seen for sub-wavelength Euclidean loops, but N will be higher than 3. Conservatively, however, an N=4 Minkowski Island fractal quad antenna should provide a PC=3 value without exhibiting substantial inefficiency.

FIG. **9**B depicts perimeter compression (PC) as a function of iteration order N for a Minkowski island fractal configuration. A conventional Euclidean quad (MI-0) has PC=1 (e.g., no compression), and as iteration increases, PC increases. Note that as N increases and approaches 6, PC approaches a finite real number asymptotically, as predicted.

Thus, fractal Minkowski Island antennas beyond iteration N=6 may exhibit diminishing returns for the increase in iteration.

It will be appreciated that the non-harmonic resonant frequency characteristic of a fractal antenna according to the present invention may be used in a system in which the frequency signature of the antenna must be recognized to pass a security test. For example, at suitably high frequencies, perhaps several hundred MHz, a fractal antenna could be implemented within an identification credit card. When the card is used, a transmitter associated with a credit card reader can electronically sample the frequency resonance of the antenna within the credit card. If and only if the credit card antenna responds with the appropriate frequency signature pattern expected may the credit card be used, e.g., for purchase or to permit the owner entrance into an otherwise secured area.

FIG. **10**A depicts a fractal inductor L according to the present invention. In contrast to a prior art inductor, the winding or traces with which L is fabricated define, at least in part, a fractal. The resultant inductor is physically smaller than its Euclidean counterpart. Inductor L may be used to form a resonator, including resonators such as shown in FIGS. **4**A and **4**B. As such, an integrated circuit or other suitably small package including fractal resonators could be used as part of a security system in which electromagnetic radiation, perhaps from transmitter **600** or **660** in FIG. **8**A will blow, or perhaps not blow, an LC resonator circuit containing the fractal antenna. Such applications are described elsewhere herein and may include a credit card sized unit **700**, as shown in FIG. **10**B, in which an LC fractal resonator **710** is implemented. (Card **700** is depicted in FIG. **10**B as though its upper surface were transparent.).

The foregoing description has largely replicated what has been set forth in applicant's above-noted FRACTAL ANTENNAS AND FRACTAL RESONATORS patent application. The following section will set forth methods and techniques for tuning such fractal antennas and resonators. In the following description, although the expression "antenna" may be used in referring to a preferably fractal element, in practice what is being described is an antenna or filter-resonator system. As such, an "antenna" can be made to behave as through it were a filter, e.g., passing certain frequencies and rejecting other frequencies (or the converse).

In one group of embodiments, applicant has discovered that disposing a fractal antenna a distance Δ that is in close proximity (e.g., less than about 0.05λ for the frequency of interest) from a conductor advantageously can change the resonant properties and radiation characteristics of the antenna (relative to such properties and characteristics when such close proximity does not exist, e.g., when the spaced-apart distance is relatively great. For example, in FIG. **11**A a conductive surface **800** is disposed a distance Δ behind or beneath a fractal antenna **810**, which in FIG. **11**A is a single arm of an MI-2 fractal antenna. Of course other fractal configurations such as disclosed herein could be used instead of the MI-1 configuration shown, and non-planar configurations may also be used. Fractal antenna **810** preferably is fed with coaxial cable feedline **50**, whose center conductor is attached to one end **815** of the fractal antenna, and whose outer shield is grounded to the conductive plane **800**. As described herein, great flexibility in connecting the antenna system shown to a preferably coaxial feedline exists. Termination impedance is approximately of similar magnitudes as described earlier herein.

In the configuration shown, the relative close proximity between conductive sheet **800** and fractal antenna **810** lowers the resonant frequencies and widens the bandwidth of antenna **810**. The conductive sheet **800** may be a plane of metal, the upper copper surface of a printed circuit board, a region of conductive material perhaps sprayed onto the housing of a device employing the antenna, for example the interior of a transceiver housing **500**, such as shown in FIGS. **8**A, **8**B, **8**C, and **15**.

The relationship between Δ, wherein $\Delta \leqq 0.05\lambda$, and resonant properties and radiation characteristics of a fractal antenna system is generally logarithmic. That is, resonant frequency decreases logarithmically with decreasing separation Δ.

FIG. **11**B shows an embodiment in which a preferably fractal antenna **810** lies in the same plane as a ground plane **800** but is separated therefrom by an insulating region, and in which a passive or parasitic element **800'** is disposed "within" and spaced-apart a distance Δ' from the antenna, and also being coplanar. For example, the embodiment of FIG. **11**B may be fabricated from a single piece of printed circuit board material in which copper (or other conductive material) remains to define the groundplane **800**, the antenna **810**, and the parasitic element **800'**, the remaining portions of the original material having been etched away to form the "moat-like" regions separating regions **800**, **810**, and **800'**. Changing the shape and/or size of element **800'** and/or the coplanar spaced-apart distance Δ' tunes the antenna system shown. For example, for a center frequency in the 900 MHz range, element **800'** measured about 63 mm×8 mm, and elements **810** and **800** each measured about 25 mm×12 mm.

In general, element **800** should be at least as large as the preferably fractal antenna **810**. For this configuration, the system shown exhibited a bandwidth of about 200 MHz, and could be made to exhibit characteristics of a bandpass filter and/or band rejection filter. In this embodiment, a coaxial feedline **50** was used, in which the center lead was coupled to antenna **810**, and the ground shield lead was coupled to groundplane **800**. In FIG. **11**B, the inner perimeter of groundplane region **800** is shown as being rectangularly shaped. If desired, this inner perimeter could be moved closer to the outer perimeter of preferably fractal antenna **810**, and could in fact define a perimeter shape that follows the perimeter shape of antenna **810**. In such an embodiment, the perimeter of the inner conductive region **800'** and the inner perimeter of the ground plane region **800** would each follow the shape of antenna **810**. Based upon experiments to date, it is applicant's belief that moving the inner perimeter of ground plane region **800** sufficiently close to antenna **810** could also affect the characteristics of the overall antenna/resonator system.

Referring now to FIG. **12**A, if the conductive surface **800** is replaced with a second fractal antenna **810'**, which is spaced-apart a distance Δ that preferably does not exceed about 0.05λ, resonances for the radiating fractal antenna **810** are lowered and advantageously new resonant frequencies emerge. For ease of fabrication, it may be desired to construct antenna **810** on the upper or first surface **820**A of a substrate **820**, and to construct antenna **810'** on the lower or second surface **820**B of the same substrate. The substrate could be doubled-side printed circuit board type material, if desired, wherein antennas **810**, **810'** are fabricated using printed circuit type techniques. The substrate thickness Δ is selected to provide the desired performance for antenna **810** at the frequency of interest. Substrate **820** may, for example, be a non-conductive film, flexible or otherwise. To avoid cluttering FIGS. **12**A and **12**B, substrate **820** is drawn with phantom lines, as if the substrate were transparent.

As noted earlier, the fractal spaced-apart structure depicted in FIGS. **12**A and **12**B may instead be used to form a fractal element in a vertical antenna system, preferably including a fractal ground counterpoise, such as was described with respect to FIG. **8**D-**3**.

Preferably, the center conductor of coaxial cable **50** is connected to one end **815** of antenna **810**, and the outer conductor of cable **50** is connected to a free end **815'** of antenna **810'**,which is regarded as ground, although other feedline connections may be used. Although FIG. **12**A depicts antenna **810'** as being substantially identical to antenna **810**, the two antennas could in fact have different configurations.

Applicant has discovered that if the second antenna **810'** is rotated some angle Θ relative to antenna **810**, the resonant frequencies of antenna **810** may be varied, analogously to tuning a variable capacitor. Thus, in FIG. **12**B, antenna **810** is tuned by rotating antenna **810'** relative to antenna **810** (or the converse, or by rotating each antenna). If desired, substrate **820** could comprise two substrates each having thickness Δ/2 and pivotally connected together, e.g., with a non-conductive rivet, so as to permit rotation of the substrates and thus relative rotation of the two antennas. Those skilled in the mechanical arts will appreciate that a variety of "tuning" mechanisms could be implemented to permit fine control over the angle Θ in response, for example, to rotation of a tunable shaft.

Referring now to FIG. **13**A, applicant has discovered that creating at least one cut or opening **830** in a fractal antenna **810** (here comprising two legs of an MI-2 antenna) results in new and entirely different resonant nodes for the antenna. Further, these nodes can have perimeter compression (PC) ranging from perhaps three to about ten. The precise location of cut **830** on the fractal antenna or resonator does not appear to be critical.

FIGS. **13**B and **13**C depict a self-proximity characteristic of fractal antennas and resonators that may advantageously be used to create a desired frequency resonant shift. In FIG. **13**B, a fractal antenna **810** is fabricated on a first surface **820**A of a flexible substrate **820**, whose second surface **820**B does not contain an antenna or other conductor in this embodiment. Curving substrate **820**, which may be a flexible film, appears to cause electromagnetic fields associated with antenna **810** to be sufficiently in self-proximity so as to shift resonant frequencies. Such self-proximity antennas or resonators may be referred to a com-cyl devices. The extent of curvature may be controlled where a flexible substrate or substrate-less fractal antenna and/or conductive element is present, to control or tune frequency dependent characteristics of the resultant system. Com-cyl embodiments could include a concentrically or eccentrically disposed fractal antenna and conductive element. Such embodiments may include telescopic elements, whose extent of "overlap" may be telescopically adjusted by contracting or lengthening the overall configuration to tune the characteristics of the resultant system. Further, more than two elements could be provided.

In FIG. **13**C, a fractal antenna **810** is formed on the outer surface **820**A of a filled substrate **820**, which may be a ferrite core. The resultant com-cyl antenna appears to exhibit self-proximity such that desired shifts in resonant frequency are produced. The geometry of the core **820**, e.g., the extent of curvature (e.g., radius in this embodiment) relative to the size of antenna **810** may be used to determine frequency shifts.

In FIG. **14**A, an antenna or resonator system is shown in which the non-driven fractal antenna **810'** is not connected to the preferably coaxial feedline **50**. The ground shield portion of feedline **50** is coupled to the groundplane conductive element **800**, but is not otherwise connected to a system ground. Of course fractal antenna **810'** could be angularly rotated relative to driven antenna **810**, it could be a different configuration than antenna **810** including having a different iteration N, and indeed could incorporate other features disclosed herein (e.g., a cut).

FIG. **14**B demonstrates that the driven antenna **810** may be coupled to the feedline **50** at any point **815'**, and not necessarily at an end point **8'5** as was shown in FIG. **14**A.

In the embodiment of FIG. **14**C, a second ground plane element **800'** is disposed adjacent at least a portion of the system comprising driven antenna **810**, passive antenna **810'**, and the underlying conductive planar element **800**. The presence, location, geometry, and distance associated with second ground plane element **800'** from the underlying elements **810**, **810'**, **800** permit tuning characteristics of the overall antenna or resonator system. In the multi-element sandwich-like configuration shown, the ground shield of conductor **50** is connected to a system ground but not to either ground plane **800** or **800'**. Of course more than three elements could be used to form a tunable system according to the present invention.

FIG. **14**D shows a single fractal antenna spaced apart from an underlying ground plane **800** a distance Δ, in which a region of antenna **800** is cutaway to increase resonance. In FIG. **14**D, for example, L**1** denotes a cutline, denoting that portions of antenna **810** above (in the Figure drawn) L**1** are cutaway and removed. So doing will increase the frequencies of resonance associated with the remaining antenna or resonator system. On the other hand, if portions of antenna **810** above cutline L**2** are cutaway and removed, still higher resonances will result. Selectively cutting or etching away portions of antenna **810** permit tuning characteristics of the remaining system.

As noted, fractal elements similar to what is generically depicted in FIGS. **14**A–**14**D may be used to form a fractal vertical element in a fractal vertical antenna system, such as was described with respect to FIG. **7**D-**3**.

FIG. **15** depicts an embodiment somewhat similar to what has been described with respect to FIG. **8**B or FIG. **8**C. Once again unit **500** is a handheld transceiver, and includes fractal antennas **510**A, **510**B–**510**B', **510**C. It is again understood that a vertical antenna such as elements **193** and fractal counterpoise **510**D (shown in FIG. **8**B) may be provided. Antennas **510**B–**510**B' are similar to what has been described with respect to FIGS. **12**A–**12**B. Antennas **510**B–**510**B' are fractal antennas, not necessarily MI-2 configuration as shown, and are spaced-apart a distance Δ and, in FIG. **13**, are rotationally displaced. Collectively, the spaced-apart distance and relative rotational displacement permits tuning the characteristics of the driven antenna, here antenna **510**B. In FIG. **14**, antenna **510**A is drawn with phantom lines to better distinguish it from spaced-apart antenna **510**B. Of course passive conductor **510**B' could instead be a solid conductor such as described with respect to FIG. **11**A. Such conductor may be implemented by spraying the inner surface of the housing for unit **500** adjacent antenna **510**B with conductive paint.

In FIG. **15**, antenna **510**C is similar to what has been described with respect to FIG. **13**A, in that a cut **830** is made in the antenna, for tuning purposes. Although antenna **510**A is shown similar to what was shown in FIG. **8**B, antenna **510**A could, if desired, be formed on a curved substrate similar to FIGS. **13**B or **13**C. While FIG. **15** shows at least two different techniques for tuning antennas according to the present invention, it will be understood that a common technique could instead be used. By that it is meant that any or all of antennas **510**A, **510**B–**510**B', **510**C could include a cut, or be spaced-apart a controllable distance Δ, or be rotatable relative to a spaced-apart conductor.

As described with respect to FIG. **8**B, an electronic circuit **610** may be coupled by cables **50**A, **50**B, **50**C to the antennas, and samples incoming signals to discern which fractal antenna, e.g., **510**A, **510**B–**510**B', **510**C (and, if present, antenna **510**D-**197**) is presently most optimally aligned with the transmitting station, perhaps a unit **600** or **650** or **670** as shown in FIG. **8**A. This determination may be made by examining signal strength from each of the antennas. An electronic circuit **620** then selects the presently best oriented antenna, and couples such antenna to the input of the receiver and output of the transmitter portion, collectively **630**, of unit **500**. It is understood that the selection of the best antenna is dynamic and can change as, for example, a user of **500** perhaps walks about holding the unit, or the transmitting source moves, or due to other changing conditions. In a cellular or a wireless telephone application, the result is more reliable communication, with the advantage that the fractal antennas can be sufficiently small-sized as to fit totally within the casing of unit **500**. Further, if a flexible substrate is used, the antennas may be wrapped about portions of the internal casing, as shown.

An additional advantage of the embodiment of FIG. **8**B is that the user of unit **500** may be physically distanced from the antennas by a greater distance that if a conventional external whip antenna were used. Although medical evidence attempting to link cancer with exposure to electromagnetic radiation from handheld transceivers is still inconclusive, the embodiment of FIG. **8**B appears to minimize any such risk.

Turning now to FIGS. **16**–**18**, in FIG. **16**, a microstrip patch antenna **10** according to the present invention is shown coupled by coaxial or other cable (or equivalent) **20** to a source of radio frequency **30**. Antenna **10** comprises a substrate **40** whose top-to-bottom thickness is preferably substantially less than one wavelength at the frequency of interest, e.g., the radio frequency or band of radio frequencies coupled by cable **20** to antenna **10**. Preferably the effective dimension of substrate is one-eighth wavelength at such frequency.

On its first surface, substrate **40** is initially covered by a conductive layer of material **50** that is etched away or otherwise removed in areas other than the desired fractal pattern (**60**) design, to expose the substrate. The remaining conductive trace portion defines a fractal element, according to the present invention.

Similarly on its second surface, substrate **40** is initially covered by a conductive layer of material **70** that is selectively removed so as to leave a desired pattern (**80**) that may also be a fractal pattern, according to the present invention. Alternatively, conductive material defining the desired patterns **60**, **80** could be deposited upon substrate **40**, rather than beginning fabrication with a substrate clad or otherwise having conductive surfaces, portions of which are removed.

Preferably feedtabs **90** and **100** are coupled, respectively, to edge regions of the first and second surfaces of substrate **40** to facilitate electrical radio frequency coupling between cable **20** and patterns **60** and/or **80**. These feedtabs preferably are etched using the same conductive material originally found on the upper or lower surfaces of substrate **40**, or may otherwise be formed using techniques known to those skilled in the relevant art. If patterns **60** and **80** are deposited rather than etched, then feedtabs **90**, **100** may be deposited at the same fabrication step.

Substrate **40** is a non-conductive material, and by way of example may be a silicon wafer, a rigid or a flexible plastic-like material, perhaps Mylar™ material, or the non-conductive portion of a printed circuit board, paper, epoxy, among other materials. The original conductive material on the first and/or second surfaces may be deposited doped polysilicon for a semiconductor substrate **40**, or copper (or other conductor) for a printed circuit board substrate.

FIG. **17** is a plan view of one surface of antenna **10** (it matters not which), and depicts a first iteration fractal conductive pattern, although a fractal pattern with higher than first iteration could instead be used. The pattern shown in FIG. **2** is often referred to as a Sierpinski (square) gasket pattern. A margin is shown in FIG. **17** between the outer perimeter of the pattern and the edge of the substrate; however no such margin is required. Although FIG. **2** shows inclusion of feedtab **90** or **100**, radio frequency feed may be made elsewhere on the surface, for example at any point **110**.

If the fractal pattern of FIG. **17** represents one surface of antenna **10**, the opposite surface need not define a fractal pattern, but may in fact do so. For example, one surface may define a fractal pattern and the opposite surface may be entirely conductive, or may define on the substrate a conductive circle, etc. If the pattern on the opposite surface is also a fractal, there is no requirement that it be the same iteration fractal as is defined on the first surface, or that it be the same fractal type. While common fractal families include Koch, Minkowski, Julia, diffusion limited aggregates, fractal trees, Mandelbrot, microstrip patch antennas with fractal element(s) according to the present invention may be implemented with other fractals as well.

FIG. **18** depicts a pattern **60** or **80** in which a different fractal pattern is defined, a so-called diffusion limited aggregate pattern. It is understood, however, that according to the present invention, a great variety of fractal patterns of first or higher iteration may be defined on the first and/or second surface of antenna **10**. In FIG. **18**, while a feedtab **90** or **100** is shown, it is again understood that radio frequency feed may be made essentially anywhere on the fractal pattern, e.g., at a point **110**.

In one embodiment, applicant fabricated an antenna **10** having sides dimensioned to about one-eighth wavelength for a frequency of about 900 MHz. Those skilled in the art will readily appreciate that a microstrip patch antenna dimensioned to one-eighth wavelength is substantially smaller than prior art non-fractal microstrip patch antennas, in which dimensions are one-quarter or one-half wavelength in size. At 900 MHz, bandwidth was about 5% to about 8% of nominal frequency. Gain and matching impedance were acceptable, and indeed substantially 50Ω impedance is realized without the need for impedance transforming devices.

Modifications and variations may be made to the disclosed embodiments without departing from the subject and spirit of the invention as defined by the following claims. While common fractal families include Koch, Minkowski, Julia, diffusion limited aggregates, fractal trees, Mandelbrot, ground counterpoise elements and/or top-hat loading elements according to the present invention may be implemented with other fractals as well.

What is claimed is:

1. An antenna system including:

a driven element, and

at least one element a portion of which is a fractal element selected from a group consisting of (a) a fractal counterpoise, and (b) a microstrip patch element; wherein the fractal element is defined as a superposition over at least N=1 iterations of a fractal generator motif, an iteration being placement of said fractal generator motif upon a base figure through at least one positioning selected from the group consisting of (i) rotation, (ii) stretching, and (iii) translation,

wherein said system includes a substrate having (1) spaced-apart first and second surfaces and (2) a substrate thickness substantially less than a wavelength at a frequency to be coupled to said antenna system; and said driven element is fabricated on the first surface of said substrate.

2. The antenna system of claim **1**, wherein said substrate is flexible.

3. The antenna system of claim **1**, wherein said fractal generator motif has x-axis, y-axis coordinates for a next iteration N+1 defined by $X_{N+1}=f(x_N, y_N)$ and $y_{N+1}=g(x_N, y_N)$, where x_N, y_N are coordinates for iteration N, and where f(x,y) and g(x,y) are functions defining said fractal generator motif and behavior.

4. The antenna system of claim **3**, wherein said fractal generator motif is selected from a family consisting of (i)-Koch, (ii) Minkowski, (iii-) Cantor, (iv) torn square, (v) Mandelbrot, (vi) Caley tree, (vii) monkey's swing, (viii) Sierpinski gasket, and (ix) Julia.

5. The antenna system of claim **1**, in which said fractal counterpoise is fabricated in a manner selected from the group consisting of (i) shaping conductive wire into said fractal, (ii) forming upon an insulator substrate a conductive layer defining traces shaped to form said fractal, (iii) forming upon a flexible insulator substrate conductive traces shaped to form said fractal; and (iv) forming upon a semiconductor substrate a layer of conductive material shaped to form said fractal.

6. The antenna system of claim **1**, wherein said system is an antenna selected from a group consisting of (i) a fractal quad, (ii) an $N\geqq3$ iteration fractal quad, (iii) a Minkowski fractal quad, (iv) a dipole, (vi) a vertical, and (vii) a microstrip patch antenna.

7. A fractal antenna coupleable to a transceiver unit, the antenna comprising:

a driven element, and

at least one fractal element selected from a group consisting of (a) a fractal counterpoise, and (b) a microstrip patch element, said fractal element having a physical shape defined substantially as a deterministic fractal of iteration $N\geqq2$ for at least a portion of said element;

wherein said antenna includes a substrate having (1) spaced-apart first and second surfaces and a substrate thickness substantially less than a wavelength at a frequency to be coupled to said antenna system: and said driven element is fabricated on the first surface of said substrate.

8. The antenna of claim **7**, wherein said fractal is defined as a superposition over at least N=2 iterations of a fractal generator motif, an iteration being placement of said fractal generator motif upon a base figure through at least one positioning selected from the group consisting of (i) rotation, (ii) stretching, and (iii) translation.

9. The antenna of claim **8**, wherein said driven element has at least one characteristic selected from a group consisting of (a) said driven element defined by a fractal, (b) said driven element is formed as a microstrip patch antenna, (c) said driven element is formed on a substrate, and (c) said driven element is formed on a flexible substrate.

10. The antenna of claim **7**, in which said transceiver unit is hand holdable in size, and wherein said antenna is mounted within a housing of said transceiver unit, and said antenna is fabricated in a manner selected from the group consisting of (i) shaping conductive wire into said fractal, (ii) forming upon an insulator substrate a conductive layer defining traces shaped to form said fractal, (iii) forming upon a flexible insulator substrate conductive traces shaped to form said fractal; and (iv) forming upon a semiconductor substrate a layer of conductive material shaped to form said fractal.

11. The antenna of claim **7**, wherein said transceiver includes a plurality of said antennas in at least one configuration selected from the group consisting of (i) an array of substantially identical said antennas coupled to an electronic circuit that dynamically selects a chosen one of said antennas to be coupled to said transceiver unit, (ii) an array of substantially identical said antennas coupled to an electronic circuit that dynamically selects a chosen one of said antennas to be coupled to said transceiver unit, at least two antennas in said array having orientation differing from other antennas in said plurality, (iii) a plurality of antennas in which at least two antennas have elements differing from elements in other of said antennas.

12. The antenna of claim **7**, wherein said driven element includes a fractal element and a conductive element, spaced-apart from said fractal element by a distance Δ sufficiently small at a frequency of interest λ to decrease said at least one resonant frequency, to widen said bandwidth, or to cause a combination thereof.

13. The antenna of claim **12**, wherein said antenna is tunable by varying at least one parameter selected from the group consisting of (a) said distance Δ, (b) relative rotation between said first and said second fractal antenna, (c) location at which a feedline center lead is coupled to said first fractal antenna, (d) location of a cut in said first fractal antenna, and (e) size of a region of said first fractal antenna cutaway and removed.

14. A top-hat loaded antenna, comprising:

a vertical element having an upper end and a lower end; and

a top-hat assembly electrically coupled to said upper end of said vertical element;

wherein said top-hat assembly includes at least one element whose physical shape is defined substantially as a deterministic fractal of iteration $N\geqq2$ for at least a portion of said element.

* * * * *

US007830319B2

(12) **United States Patent**
Cohen et al.

(10) **Patent No.: US 7,830,319 B2**
(45) **Date of Patent: Nov. 9, 2010**

(54) **WIDEBAND ANTENNA SYSTEM FOR GARMENTS**

(76) Inventors: **Nathan Cohen**, 2 Ledgewood Pl., Belmont, MA (US) 02478; **David Moschella**, 195 Follen Rd., Lexington, MA (US) 02421

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 143 days.

(21) Appl. No.: **12/118,957**

(22) Filed: **May 12, 2008**

(65) **Prior Publication Data**

US 2009/0153420 A1 Jun. 18, 2009

Related U.S. Application Data

(63) Continuation of application No. 11/210,978, filed on Aug. 24, 2005, now abandoned.

(60) Provisional application No. 60/603,882, filed on Aug. 24, 2004.

(51) **Int. Cl.**
H01Q 1/12 (2006.01)

(52) **U.S. Cl.** .... **343/718**; 343/700 MS

(58) **Field of Classification Search** .... 343/718, 343/702, 700 MS, 846
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,079,602	A	2/1963	Hamel et al.
3,249,946	A	5/1966	Flanagan
3,689,929	A	9/1972	Moody
3,810,183	A	5/1974	Krutsinger et al.
3,811,128	A	5/1974	Munson
4,318,109	A	3/1982	Weathers
4,358,769	A	11/1982	Tada et al.
4,381,566	A	4/1983	Kane
4,652,889	A	3/1987	Bizouard et al.
4,656,482	A	4/1987	Peng
5,006,858	A	4/1991	Shirosaka
5,111,211	A	5/1992	Dahlberg
5,164,738	A	11/1992	Walter et al.
5,313,216	A	5/1994	Wang et al.
5,355,318	A	10/1994	Dionnet et al.
5,363,114	A	11/1994	Shoemaker
5,453,752	A	9/1995	Wang
5,608,413	A	3/1997	MacDonald
5,844,525	A	12/1998	Hayes et al.
5,995,064	A	11/1999	Yanagisawa et al.
6,104,349	A	8/2000	Cohen
6,127,977	A	10/2000	Cohen
6,140,975	A	10/2000	Cohen
6,323,809	B1	11/2001	Maloney et al.
6,356,238	B1	3/2002	Gainer et al.

(Continued)

OTHER PUBLICATIONS

Kim, Y., Jaggard, D., "The Fractal Random Array", Proc. IEEE 74, 1278-1280 (1986).

(Continued)

Primary Examiner—HoangAnh T Le
(74) *Attorney, Agent, or Firm*—McDermott Will & Emery LLP

(57) **ABSTRACT**

A portable antenna system includes an antenna that is substantially defined by one or more portions that include electrically conductive self-similar extensions. The system also includes an article of clothing in which the antenna is attached to a surface of the article of clothing such that electrically conductive self-similar extensions extend across the surface of the article of clothing.

10 Claims, 4 Drawing Sheets

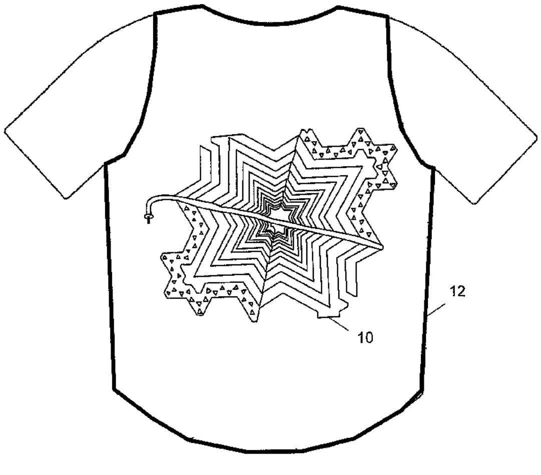

U.S. PATENT DOCUMENTS

6,476,766	B1	11/2002	Cohen
6,603,440	B2	8/2003	Howard
6,867,740	B2	3/2005	Goodyear
6,947,764	B2	9/2005	Carillo et al.
6,985,122	B2	1/2006	Cohen
2003/0214389	A1	11/2003	Arneson et al.
2003/0230614	A1	12/2003	Karenga

OTHER PUBLICATIONS

Pfeiffer, A., “The Pfeiffer Quad Antenna System,” QST., pp. 28-31, (1994).

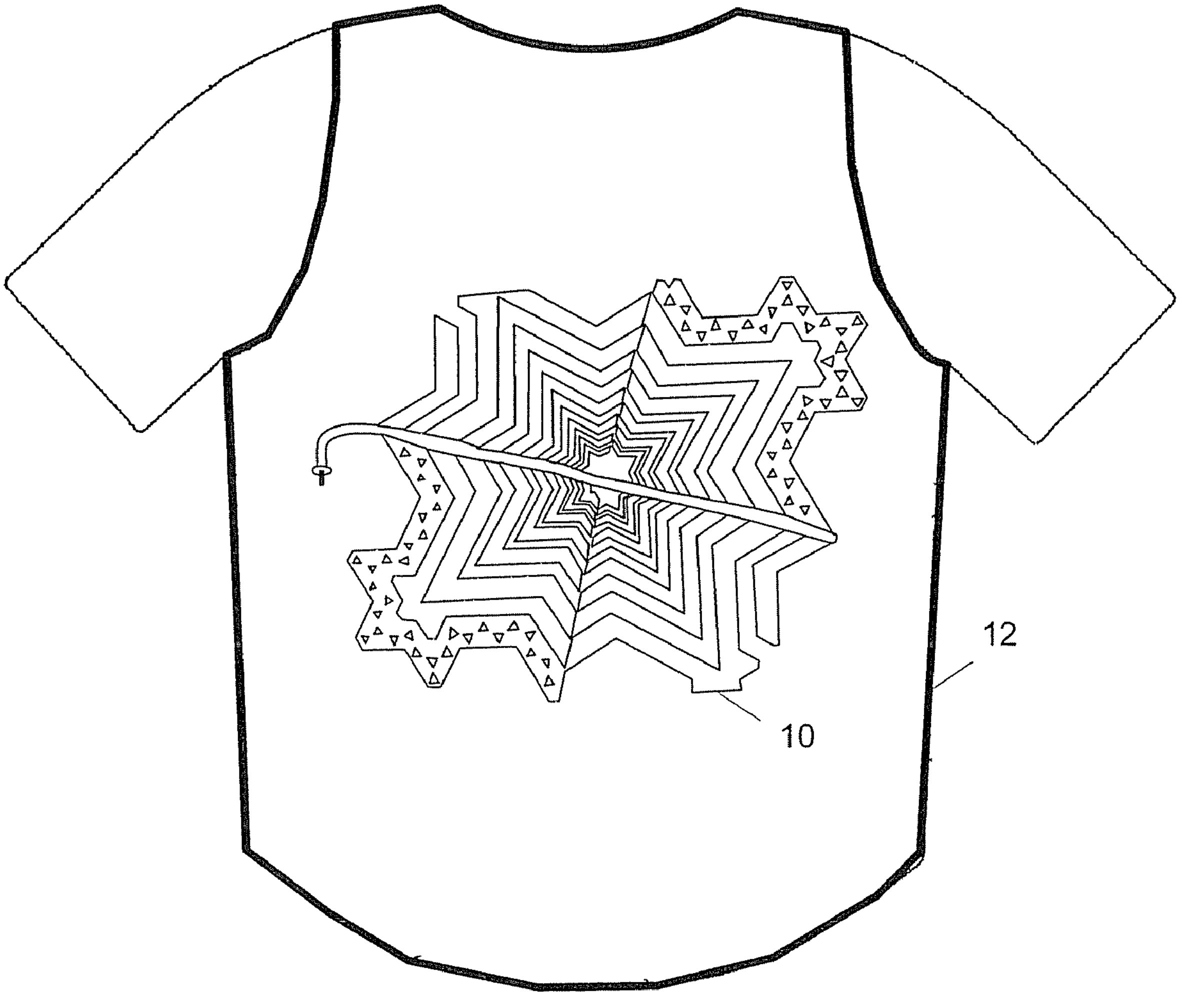

FIG. 1

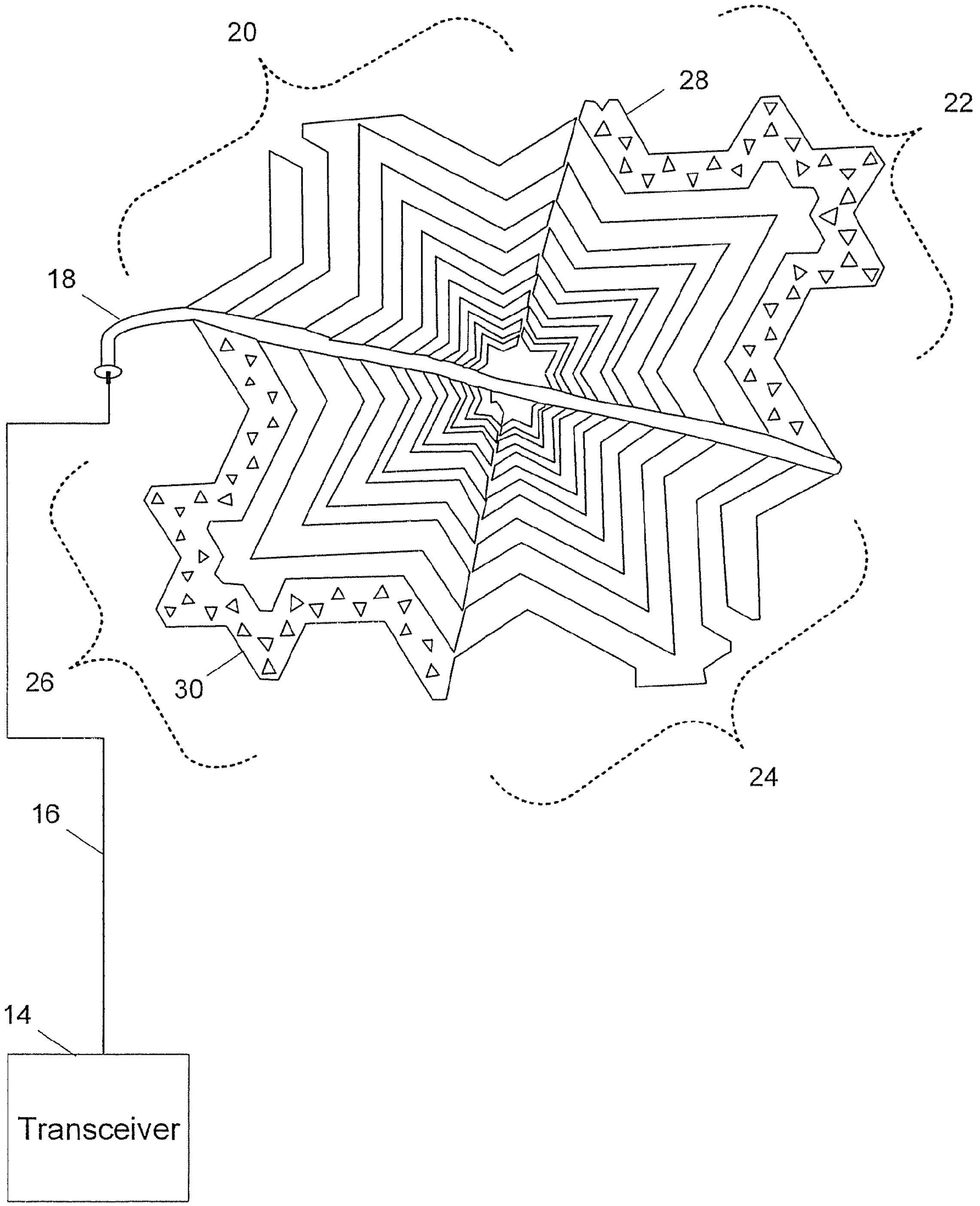
20
28
22
18
26
30
24
16
14
Transceiver

FIG. 2

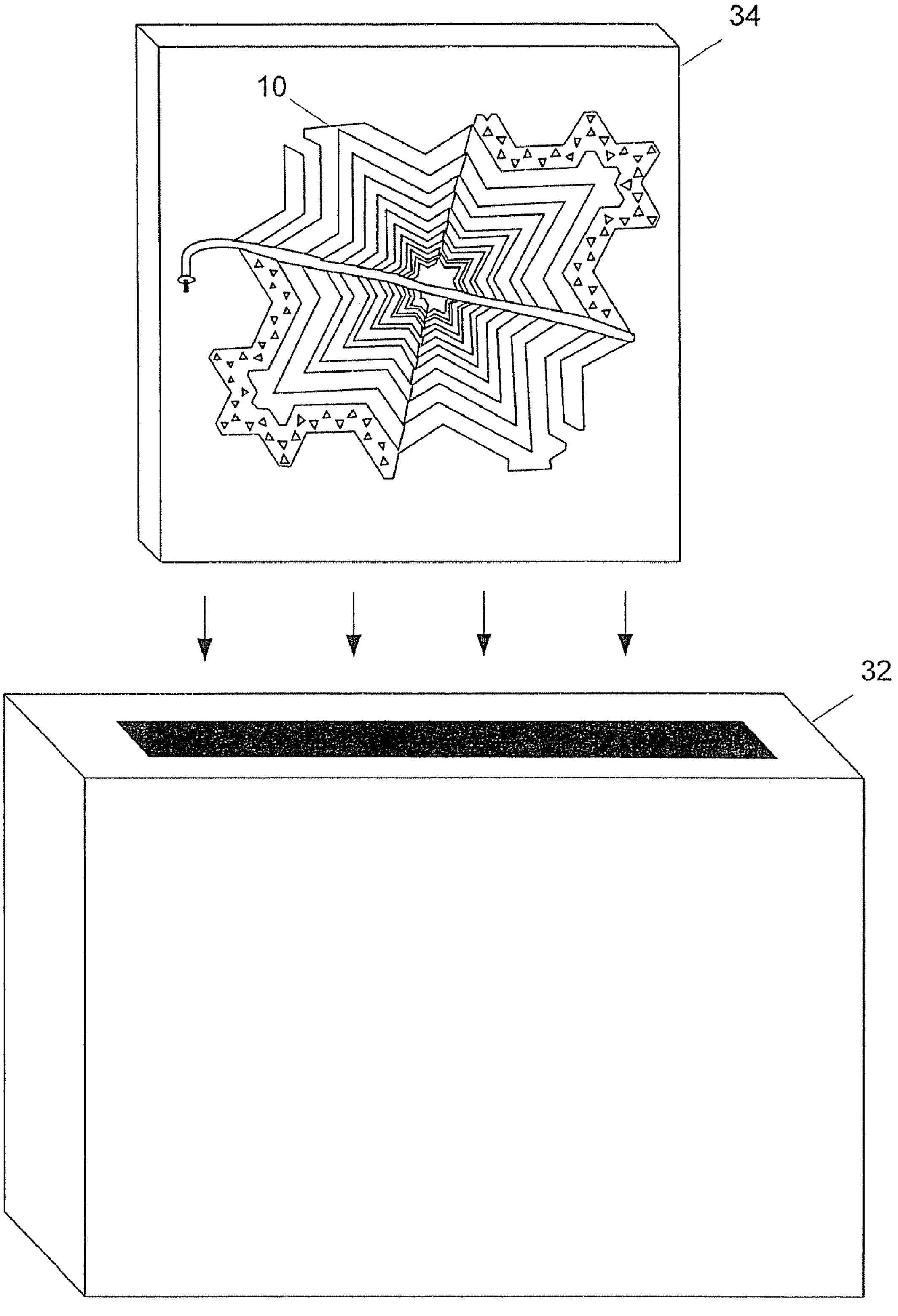

FIG. 3

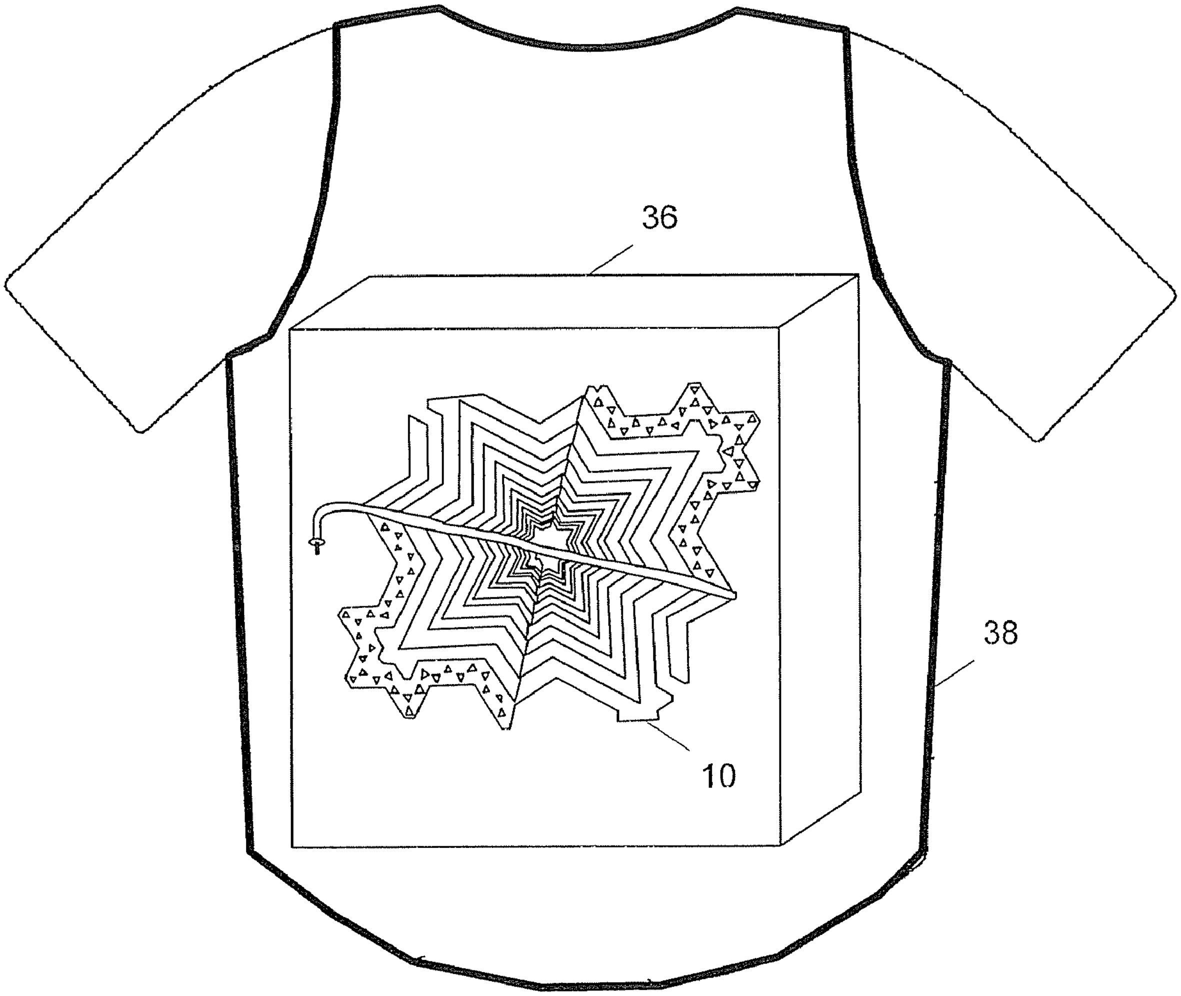

FIG. 4

WIDEBAND ANTENNA SYSTEM FOR GARMENTS

RELATED APPLICATIONS AND TECHNICAL FIELD

This application is a continuation of U.S. patent application Ser. No. 11/210,978 filed 24 Aug. 2005, which claims priority to U.S. Provisional Patent Application Ser. No. 60/603,882, filed Aug. 24, 2004, the entire contents of both of which application are incorporated herein by reference. This application is also a continuation-in-part of U.S. patent application Ser. No. 11/778,734 (FRTK-1CN6) filed 17 Jul. 2007, which is a continuation of U.S. patent application Ser. No. 10/243,444 (FRTK-1CN5) filed 13 Sep. 2002, which is a continuation of U.S. application Ser. No. 08/512,954 (FRTK-1) filed 9 Aug. 1995, now issued as U.S. Pat. No. 6,452,553; this application is also a continuation-in-part of U.S. patent application Ser. No. 11/390,323 (FRTK-3CN2CN) filed 27 Mar. 2006, which is a continuation of U.S. patent application Ser. No. 10/287,240 (FRTK-3CN2) filed 4 Nov. 2002, which in turn is a continuation of U.S. patent application Ser. No. 09/677,645 (FRTK-3CN) filed 3 Oct. 2000, which in turn is a continuation of both U.S. patent application Ser. No. 08/967,375 (FRTK-1CN4) filed 7 Nov. 1997 and U.S. patent application Ser. No. 08/965,914 (FRTK-3) filed 7 Nov. 1997, issued as U.S. Pat. No. 6,127,977 (3 Oct. 2000); this application is also a continuation-in-part of U.S. patent application Ser. No. 11/867,284 (FRTK-6CN2) filed 4 Oct. 2007, which is a continuation of U.S. patent application Ser. No. 11/327,982 (FRTK-6CN) filed 9 Jan. 2006, which is a continuation of U.S. patent application Ser. No. 10/971,815 (FRTK-6) filed Oct. 22, 2004 now issued as U.S. Pat. No. 6,985,122, which claimed priority to U.S. Provisional Patent Application Ser. No. 60/513,497, filed Oct. 22, 2003.

This application is also related to the following U.S. application, of common assignee, and the contents of which are incorporated herein in their entirety by reference: "Antenna System for Radio Frequency Identification," U.S. patent application Ser. No. 10/971,815 (FRTK-6) filed 22 Oct. 2004.

This disclosure relates to antenna systems and, more particularly, to wideband antennas that are incorporated into garments.

BACKGROUND

Antennas are used to typically radiate and/or receive electromagnetic signals, preferably with antenna gain, directivity, and efficiency. Practical antenna design traditionally involves trade-offs between various parameters, including antenna gain, size, efficiency, and bandwidth.

Antenna design has historically been dominated by Euclidean geometry. In such designs, the closed area of the antenna is directly proportional to the antenna perimeter. For example, if one doubles the length of an Euclidean square (or "quad") antenna, the enclosed area of the antenna quadruples. Classical antenna design has dealt with planes, circles, triangles, squares, ellipses, rectangles, hemispheres, paraboloids, and the like.

With respect to antennas, prior art design philosophy has been to pick a Euclidean geometric construction, e.g., a quad, and to explore its radiation characteristics, especially with emphasis on frequency resonance and power patterns. Unfortunately antenna design has concentrated on the ease of antenna construction, rather than on the underlying electromagnetics, which can cause a reduction in antenna performance.

Antenna systems that incorporate a Euclidean geometry include man-portable communication antennas such as monopole antennas. Typically these types of antennas include a wire or rod that may be extended to a deployed position that is located above the antenna carrier's head. As such, these extendable antennas may provide a visual signature that may disclose the location of the person carrying the antenna (such as a soldier in the field). Additionally, these antennas implement a monopole design that typically exhibit a narrow instantaneous bandwidth.

SUMMARY OF THE DISCLOSURE

In accordance with an aspect of the disclosure, a portable antenna system includes an antenna that is substantially defined by one or more portions that include electrically conductive self-similar extensions. The system also includes an article of clothing in which the antenna is attached to a surface of the article of clothing such that electrically conductive self-similar extensions extend across the surface of the article of clothing.

In one embodiment, the self-similar extensions may include two or more angular bends. The system may further include a co-planar feed connected to the antenna for transmitting and/or receiving electromagnetic signals through the antenna. Each self-similar extension may incorporate a fractal geometry. Furthermore, the antenna may transmit and/or receive electromagnetic energy across a spectral bandwidth that is defined by a ratio of at least 5:1. The system may also include a dielectric plate to which the antenna may be mounted. The dielectric plate may capable of deflecting projectiles. The antenna may be mounted to various locations on clothing. For example, the antenna may be mounted on an internal clothing layer or to an exterior surface of the article of clothing. Various articles of clothing may be used, for example, the article of clothing may be a vest.

In accordance with another aspect, a portable antenna system includes an antenna that is substantially defined by one or more portions that include electrically conductive self-similar extensions. The portable antenna system also includes a pouch, in which the antenna is contained. The pouch is also configured for mounting to a clothing surface.

In one embodiment, the system may further include a plate upon which the pouch is positioned such that the plate separates the antenna from the body of a person wearing clothing that includes the clothing surface. The self-similar extensions may include two or more angular bends. The system may also include a co-planar feed that is connected to the antenna for transmitting and/or receiving electromagnetic signals. Each self-similar extension may incorporate a fractal geometry. The pouch may include a layer of foam dielectric material or a layer of solid dielectric material. The pouch may include a fibrous dielectric material such as Tyvek™. The plate may include a projectile deflecting material.

In accordance with another aspect, a portable antenna system includes an antenna that is substantially defined one or more portions that include electrically conductive self-similar extensions. The system also includes a plate in which the antenna is mounted upon, and a garment in which the plate is attached to a clothing surface included in the garment.

In one embodiment, the plate may include a projectile deflecting material and/or a dielectric material. The garment may be a vest. The plate may be attached to a surface of the garment such that when worn, the antenna extends across the back of the person wearing the garment. Each self-similar extension may incorporate a fractal geometry. The antenna may transmit and/or receive electromagnetic energy across a spectral bandwidth that is defined by a ratio of at least 5:1.

Additional advantages and aspects of the present disclosure will become readily apparent to those skilled in the art from the following detailed description, wherein embodiments of the present invention are shown and described, simply by way of illustration of the best mode contemplated for practicing the present invention. As will be described, the present disclosure is capable of other and different embodiments, and its several details are susceptible of modification in various obvious respects, all without departing from the spirit of the present disclosure. Accordingly, the drawings and description are to be regarded as illustrative in nature, and not as limitative.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a diagrammatic view of a wideband antenna mounted to a garment.

FIG. **2** is a diagrammatic view of the wideband antenna shown in FIG. **1**.

FIG. **3** is a diagrammatic view of a pouch that holds the wideband antenna and may be mounted to the garment shown in FIG. **1**.

FIG. **4** is a diagrammatic view of wideband antenna embedded into a projectile deflecting plate that is mounted on a garment.

DETAILED DESCRIPTION OF THE EMBODIMENTS

Referring to FIG. **1**, an antenna **10** is mounted conformal to a surface of a garment. In particular, antenna **10** is mounted to the back of a vest **12**, however, in other arrangements the antenna **10** may be mounted to other types of garments such as shirts, coats, parkas, etc. By mounting antenna **10** to the back of vest **12**, a fully integrated antenna is provided for various applications such as combat wear for military personnel. In some arrangements antenna **10** may be incorporated into a military "flak" vest or other similar military clothing known in the art for protecting soldiers in hazardous situations. Typically a flak vest is produced from light-weight material and includes conducting regions formed from a metalized cloth. Such cloth may be formed of a copper coated polyester fabric that is commercially available from Flectron Metalized Materials of St. Louis, Mo. However, any materials known in the art of clothing design and tailoring may be used to produce vest **12**.

In this particular implementation, due to materials and production procedures, antenna **10** is opaque at visual wavelengths. However, in other implementations, antenna **10** may be substantially transparent at wavelengths in the visual portion of the electromagnetic spectrum. To mount antenna **10** conformal to vest **12**, the antenna predominately extends in two dimensions (i.e., length and width) and is relatively thin to provide flexibility in movement. Rather than mounting antenna **10** directly to the outer surface of vest **12**, the antenna may be embedded within one or more cloth layers of the vest. Some of these layers may be designed for particular capabilities, such as a bullet-proof layer or other types of projectile (e.g., flak) defection. For example, antenna **10** may be partially or fully embedded in one or more dielectric layer that are incorporated into the vest for bullet and/or flak deflection. A portion or all of this dielectric material may include one or more layers of foam or solid dielectric material. These layers of dielectric material may further be partially or fully embedded within another material. For example, antenna **10** may be embedded in a dielectric plate that is then wrapped around a fibrous dielectric material such as Tyvek®, which is produced by Dupont of Wilmington, Del.

Rather than incorporating antenna **10** into the clothing material of vest **12** (or other type of clothing article), the antenna may be incorporated into a pouch or other similar article capable of holding the antenna. By using a pouch, a person such as a soldier can position the antenna on various locations on his or her person. For example, a soldier may position the pouch on his chest or on his back to provide appropriate signal transmission and/or reception performance with other troops, a base, etc.

Along with being incorporated into an article of clothing or a pouch, antenna **10** is designed with a self-similar geometry that provides broad frequency coverage for signal transmission and/or reception. In general the self-similar shape is defined as a fractal geometry. Fractal geometry may be grouped into random fractals, which are also termed chaotic or Brownian fractals and include a random noise components, or deterministic fractals. Fractals typically have a statistical self-similarity at all resolutions and are generated by an infinitely recursive process. For example, a so-called Koch fractal may be produced with N iterations (e.g., N=1, N=2, etc.). One or more other types of fractal geometries may also be incorporated into the design of antenna **10**.

By incorporating the fractal geometry into electrically conductive and non-conductive portions of antenna **10**, the length and width of the conductive and non-conductive portions of the antenna is increased due to the nature of the fractal pattern. However, while the lengths and widths increase, the overall footprint area of antenna **10** is relatively small. By providing longer conductive paths, antenna **10** can perform over a broad frequency band. For example, the size reduction (relative to a wavelength) for the lowest frequency of operation approximately has a ratio of approximately 15:1 to 20:1.

Antenna **10** provides wideband frequency coverage for transmitting and/or receiving electromagnetic signals. For example, bandwidths ratios of 5:1 or larger may be supported by antenna **10**. For this lower ratio (i.e., 5:1) antenna **10** may perform at frequencies within a broad frequency band, for example, of approximately 3000 Mega Hertz (MHz) to 15,000 MHz. However, it should be appreciated that performance within other frequency bands may be achieved. Thus, antenna **10** is capable of transmitting and receiving electromagnetic signals over a broad frequency range.

Referring to FIG. **2**, antenna **10** is connected to a transceiver **14** over a conductor **16** (e.g., a cable, conducting trace, wire, etc.). By connecting to antenna **10**, transceiver **14** may send signals to the antenna for transmission or receive signals collected by the antenna. Typically to send and receive signals (and improve the gain of antenna **10**), transceiver **10** includes a low noise amplifier (LNA) and a power amplifier (PA). To connect conductor **16** to antenna **10**, a co-planar feed **18** is electrically connected to the antenna that also provides wideband performance. In some arrangements, a matching network is included in co-planar feed **18** to reduce signal dropouts (known as "suckouts") that are located within particular portions of the spectrum. Various techniques known to one skilled in the art of electronics and antenna system design may be implemented to connect connector **16** to antenna **10**. For example, an electrically conductive epoxy may be used to provide an adhesive connection with appropriate electrical conductivity. Additionally, in some embodiments other electromagnetic and electronic devices and components may be connected to co-planar feed **18**. For example, a power divider may be connected between conductor **16** and co-planar feed **18**.

In this exemplary fractal antenna design, antenna **10** includes an electrically conductive portion and a non-conductive portion. In particular, antenna **10** includes four sections **20**, **22**, **24**, **26** that include electrically conductive and non-conductive portions that implement a self-similar pattern (e.g., a fractal geometry). Both the conductive and non-conductive portions include extensions that include multiple angular bends to incorporate the self-similar pattern. In this example, each extension includes at least two angular bends. However, in other embodiments more angular bends may be incorporated to produce a similar fractal geometry or a different type of self-similar pattern.

In addition to incorporating a self-similar pattern into the conductive and non-conductive extensions, one or more self-similar patterns may be incorporated into the individual extensions. In this exemplary design, triangular holes are cut into two extensions **28** and **30** that are respectively included in section **22** and **26** of antenna **10**. Along with being distributed throughout each extension in a self-similar manner, each individual triangular hole may implement a fractal geometry.

Various types of conductive materials may be used to produce the electrically conductive portion (i.e., self-similar extensions) of antenna **10**. For example, various types of metallic material such as metallic tape, metallic paint, metallic ink or powder, metallic film, or other similar materials capable of conducting electricity may be selected. In this particular example, the electrically conductive portion of antenna **10** is produced from an electrically conductive coating that covers a non-conductive substrate. To produce the shape of the self-similar extensions, a laser or other type of cutting device may be used to ablate the conductive coating and from the non-conductive substrate.

By exposing portions of the non-conductive substrate, a boundary of the outer-most self-similar extensions is defined by a portion of the substrate. Additionally, exposed segments of the substrate define boundaries of the self-similar extensions. Various types of non-conductive materials may be used as a substrate to define the boundaries of the conductive portions of antenna **10**. For example, these materials may include insulators (e.g., air, etc.), dielectrics (e.g., glass, fiberglass, plastics, etc.), semiconductors, and other materials that impede the flow of electricity.

In some embodiments, the non-conductive portions of antenna **10** are produced from a high quality plastic or fiberglass that is structurally sturdy and may be processed (e.g., shaped) relatively quickly. Along with impeding current flow, the non-conductive material also typically provides structural support to the conductive portion of antenna **10**. To provide such support, the non-conductive materials may include materials typically used for support and/or re-enforce other materials. To protect antenna **10** (and provide structural support), a visually transparent (or semi-transparent) material may cover the conductive and non-conductive portions of the antenna. For example, both sides of antenna **10** may be covered by a transparent laminate that is applied with a thermal transfer. The electrically conductive portion and the non-conductive may also be cover by similar or dissimilar material. For example, one laminate may be used to cover the conductive portion of antenna **10** while another laminate is used to cover the non-conductive portion. These different laminates may be used to approximately match the optical appearance of both portions. Multiple layers of materials may also be used to cover the portions of antenna **10**. For example, one layer of laminate may be applied to the electrically-conductive portions of antenna **10** and two or more layers of laminate may be applied to the non-conductive portions to match the optical appearances of the entire antenna.

In this exemplary design, the four portions **20**-**26** are configured to provide a dipole response pattern for transmission and/or reception. Alternatively, other antenna designs may be implemented (e.g., a phased array design, etc.) independent or in combination with the dipole design provided in the figure. To expand the frequency coverage of antenna **10**, additional structure may be included in the antenna. For example, one or more conductors (e.g., conductive traces, wires, etc.) may be attached to some (or all) of the self-similar extensions. By including these conductive attachments, the frequency coverage of antenna may be significantly extended. For example, for this exemplary design, the frequency coverage may extend to relatively low frequencies.

Antenna **10** may be implemented into various types of antenna systems known to one skilled in the art of antenna design and antenna system design. In one scenario, antenna **10** may be used to transfer radio frequency (RF) signals among people such as military personnel in the field, various types of instillations (e.g., bases, etc.), and/or telecommunication equipment (e.g., wireless telephones, cellular telephones, satellites, etc.).

Along with wideband frequency coverage for broadband operations, by incorporating a fractal geometry into antenna **10** to increase conductive trace length and width, antenna losses are reduced. By reducing antenna loss, the output impedance of antenna **10** is held to a nearly constant value across the operating range of the antenna. For example, a 50-ohm output impedance may be provided by antenna **10** across the operational frequency band.

Referring to FIG. **3**, a pouch **32** is shown in which antenna **10** may be inserted. In this particular arrangement, antenna **10** is mounted to a plate **34** that provides structural support. By inserting antenna **10** in pouch **32**, the antenna may be positioned upon various locations of a person that is carrying the pouch. For example, pouch **32** may be attached to the front, back, or side of vest or other type of clothing worn over the torso. Along with wearing pouch **32** external to a piece of clothing or garment, the pouch may be worn under a garment or inserted into between clothing layers of a garment. As mentioned above, various types of material may be incorporated into the pouch. For example, pouch **32** may include one or more layers of foam or solid dielectric material. Fibrous material such as Tyvek™ may also be implemented to cover or wrap around antenna **10**. To attach pouch **32** to a garment of a piece of clothing, various techniques known in the art of clothing design and tailoring may be implemented. For example, Velcro™, straps, hooks, or other similar materials and/or mechanisms may implemented for attaching the pouch. In this exemplary design, one antenna (i.e., antenna **10**) is inserted into pouch **32**, however, in other implementation, a pouch may be produced that is capable of holding two or more antennas to increase directional coverage. Furthermore, by including electronic equipment such as a power divider in pouch **32**, signals may be split among the multiple antennas. Structural plate **34** may be produced from various materials, for example, the plate may be produced from one or more dielectric materials (e.g., ceramic). In addition to providing structural support, plate **34** may also increase the distance between antenna **10** and the body for the person (e.g., a soldier) that is carrying pouch **32**. For example, pouch **32** maybe positioned on the back of a person such that plate **34** provides a separation distance between antenna **10** and the person's back. This separation distance increases the electric distance between the person and antenna **10** and thereby reduces the interference effects caused by the person's body. By decreasing this interference, performance improves for antenna **10**. In this exemplary design, antenna **10** and plate **34** are inserted into pouch **32** that is positioned on a person's body (e.g., back, chest, etc.). However, in other designs plate **34** may be positioned without the need of pouch **32**.

Referring to FIG. **4**, antenna **10** is embedded in a structural plate **36** that is attached to the back of a vest **38**. Similar to plate **34** (shown in FIG. **4**), structural plate **36** also separates antenna **10** from the body of the person wearing vest **38**. By providing this separation, the performance of antenna **10** improves since the separation reduces the interference effects of the person's body. Also, by implementing various types of material into plate **36**, additionally capabilities may be provided. For example, projectile deflection materials known to one skilled in the art of armor design and personnel protection technology may be incorporated into plate **36**. Various types of bullet deflecting and/or flak deflecting materials may be incorporated into the exterior surface or inner layers of plate **36**.

A number of implementations have been described. Nevertheless, it will be understood that various modifications may be made. Accordingly, other implementations are within the scope of the following claims.

What is claimed is:

1. A portable antenna system comprising:

an antenna that includes (i) an electrically conductive portion including electrically conductive traces each including a fractal pattern based on an acute angle, and (ii) an electrically non-conductive portion that structurally supports the electrically conductive portion; and

an article of clothing, wherein the antenna is fixedly attached to a surface ofthe article of clothing such that electrically conductive traces extend across the surface of the article of clothing.

2. The portable antenna system of claim **1**, wherein the traces include two or more angular bends.

3. The portable antenna system of claim **1**, further comprising:

a co-planar feed connected to the antenna for transmitting and/or receiving electromagnetic signals through the antenna.

4. The portable antenna system of claim **1**, wherein the antenna is configured to transmit electromagnetic energy across a spectral bandwidth that is defined by a ratio of at least 5:1.

5. The portable antenna system of claim **1**, wherein the antenna is configured to receive electromagnetic energy across a spectral bandwidth that is defined by a ratio of at least 5:1.

6. The portable antenna system of claim **1**, further comprising:

a dielectric plate to which the antenna is mounted.

7. The portable antenna system of claim **6**, wherein the dielectric plate is capable of deflecting projectiles.

8. The portable antenna system of claim **1**, wherein the antenna is mounted on an internal clothing layer of the article of clothing.

9. The portable antenna system of claim **1**, wherein the antenna is mounted to an exterior surface of the article of clothing.

10. The portable antenna system of claim **1**, wherein the article of clothing is a vest.

* * * * *

United States Patent [19]
Gray

BEST AVAILABLE COPY

[11] **3,890,548**
[45] **June 17, 1975**

[54] PULSED CAPACITOR DISCHARGE ELECTRIC ENGINE

[75] Inventor: **Edwin V. Gray**, Northridge, Calif.

[73] Assignee: **Evgray Enterprises, Inc.**, Van Nuys, Calif.

[22] Filed: **Nov. 2, 1973**

[21] Appl. No.: **412,415**

[52] **U.S. Cl.** **318/139**; 318/254; 318/439; 310/46
[51] **Int. Cl.** **H02p 5/00**
[58] **Field of Search** 310/46, 5, 6; 318/194, 318/439, 254, 139; 320/1; 307/110

[56] **References Cited**

UNITED STATES PATENTS

2,085,708	6/1937	Spencer	318/194
2,800,619	7/1957	Brunt	318/194
3,579,074	5/1971	Roberts	320/1
3,619,638	11/1971	Phinney	307/110

OTHER PUBLICATIONS

Frungel, *High Speed Pulse Technology*, Academic Press Inc., 1965, pp. 140–148.

Primary Examiner—Robert K. Schaefer
Assistant Examiner—John J. Feldhaus
Attorney, Agent, or Firm—Gerald L. Price

[57] ABSTRACT

There is disclosed herein an electric machine or engine in which a rotor cage having an array of electromagnets is rotatable in an array of electromagnets, or fixed electromagnets are juxtaposed against movable ones. The coils of the electromagnets are connected in the discharge path of capacitors charged to relatively high voltage and discharged through the electromagnetic coils when selected rotor and stator elements are in alignment, or when the fixed electromagnets and movable electromagnets are juxtaposed. The discharge occurs across spark gaps disclosed in alignment with respect to the desired juxtaposition of the selected movable and stationary electromagnets. The capacitor discharges occur simultaneously through juxtaposed stationary movable electromagnets wound so that their respective cores are in magnetic repulsion polarity, thus resulting in the forced motion of movable electromagnetic elements away from the juxtaposed stationary electromagnetic elements at the discharge, thereby achieving motion. In an engine, the discharges occur successively across selected ones of the gaps to maintain continuous rotation. Capacitors are recharged between successive alignment positions of particular rotor and stator electromagnets of the engine.

18 Claims, 19 Drawing Figures

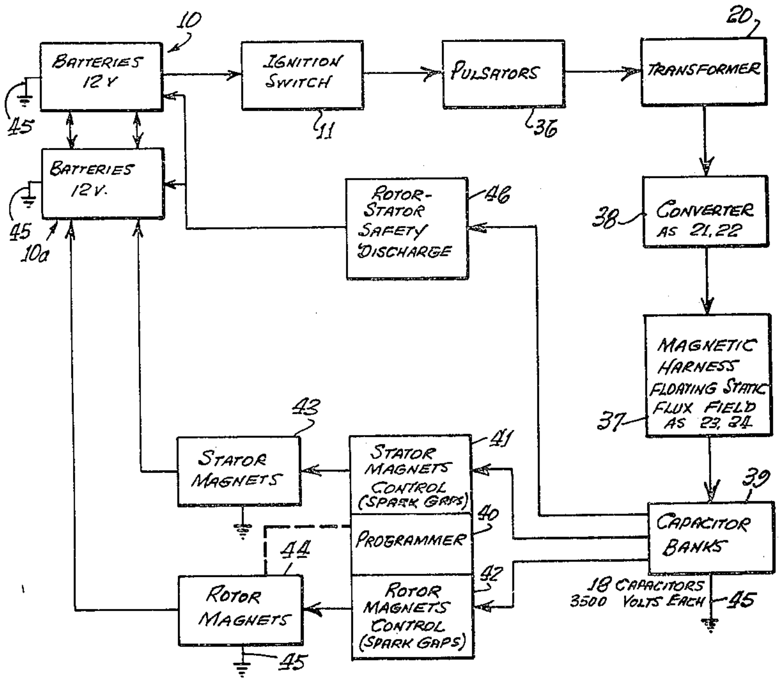
10
BATTERIES 12 V
IGNITION SWITCH
11
PULSATORS
36
20
TRANSFORMER
45
BATTERIES 12 V.
45
10a
ROTOR-STATOR SAFETY DISCHARGE
46
38
CONVERTER AS 21, 22
MAGNETIC HARNESS FLOATING STATIC FLUX FIELD AS 23, 24
37
43
STATOR MAGNETS
STATOR MAGNETS CONTROL (SPARK GAPS)
41
40
PROGRAMMER
39
CAPACITOR BANKS
44
42
ROTOR MAGNETS
ROTOR MAGNETS CONTROL (SPARK GAPS)
18 CAPACITORS 3500 VOLTS EACH
45
45

FIG. 1.

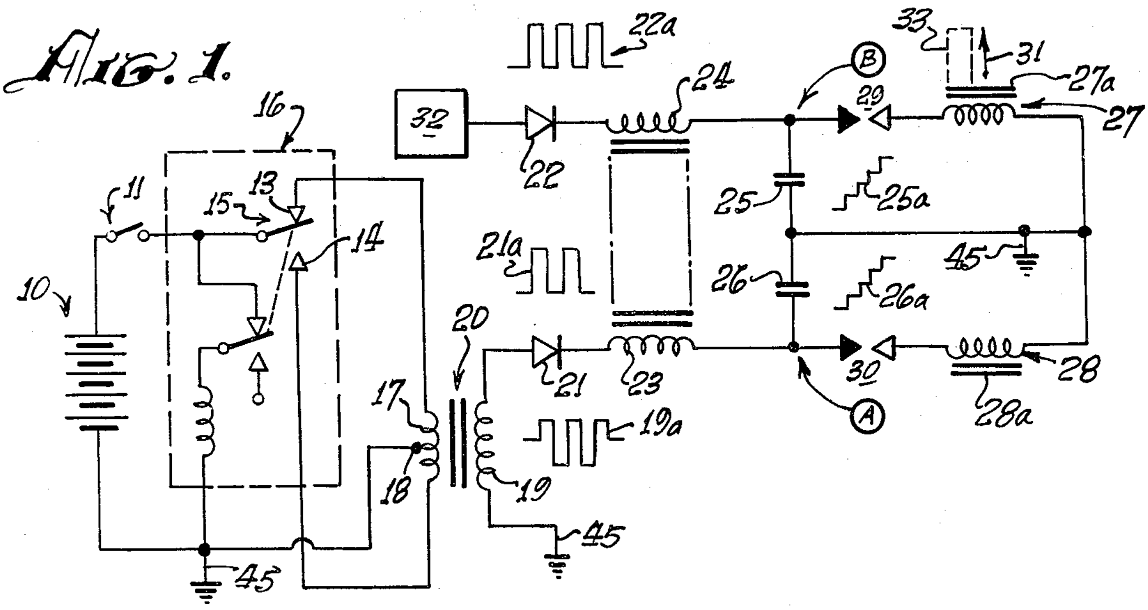
A
B

FIG. 2.

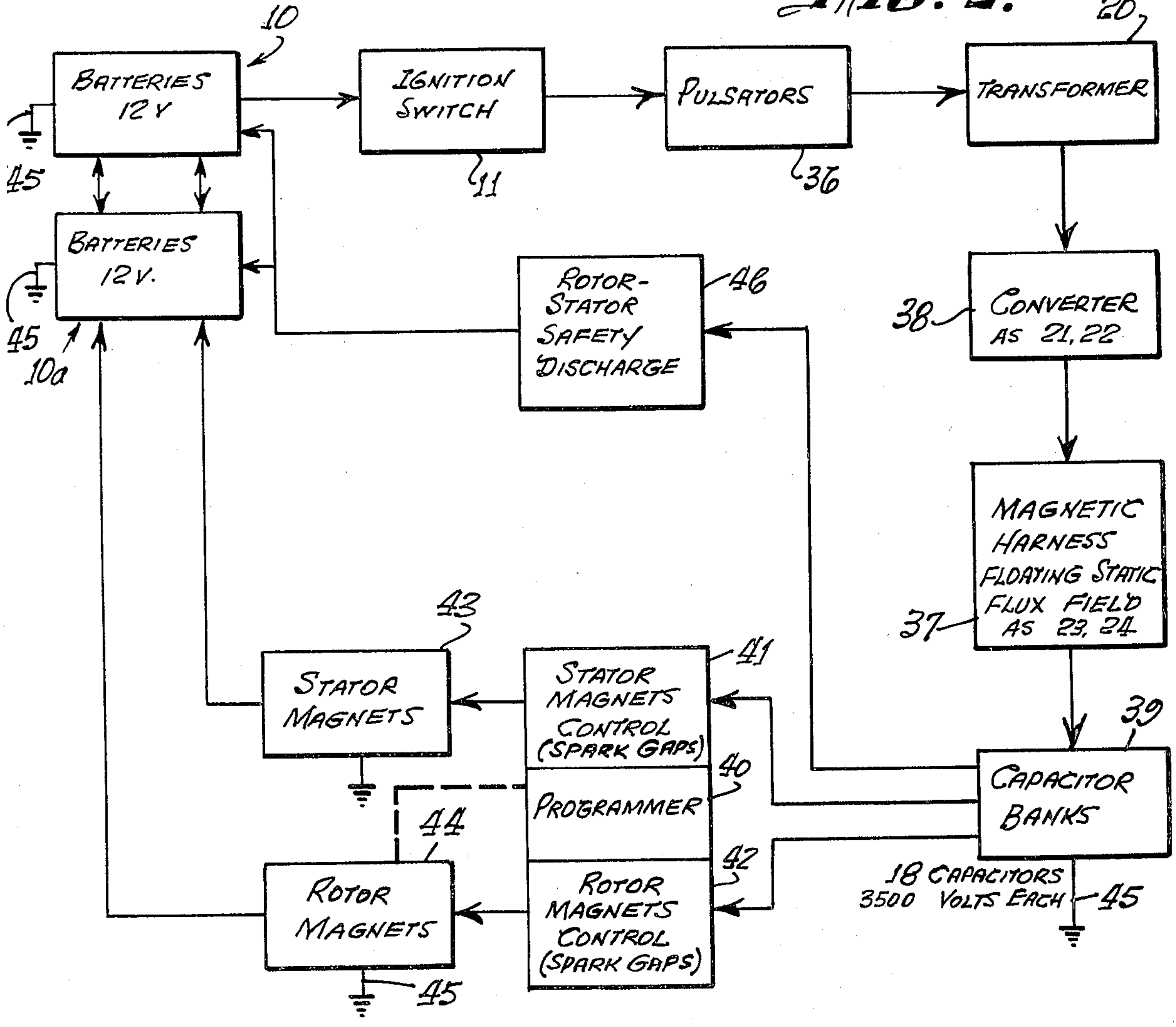
BATTERIES 12 V
10
BATTERIES 12 V.
10a
45
45
IGNITION SWITCH
11
PULSATORS
36
TRANSFORMER
20
CONVERTER AS 21, 22
38
MAGNETIC HARNESS FLOATING STATIC FLUX FIELD AS 23, 24
37
ROTOR-STATOR SAFETY DISCHARGE
46
CAPACITOR BANKS
39
18 CAPACITORS 3500 VOLTS EACH
45
STATOR MAGNETS
43
STATOR MAGNETS CONTROL (SPARK GAPS)
41
PROGRAMMER
40
ROTOR MAGNETS CONTROL (SPARK GAPS)
42
ROTOR MAGNETS
44
45

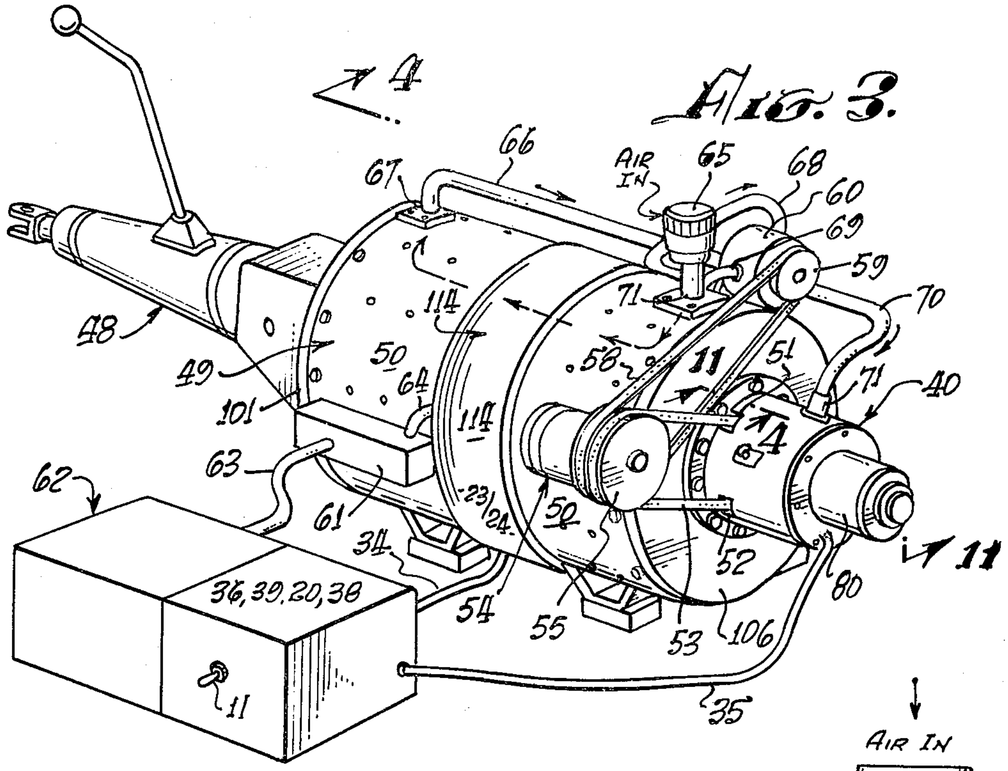
Fig. 3.
AIR IN
23/24
36,39,20,38

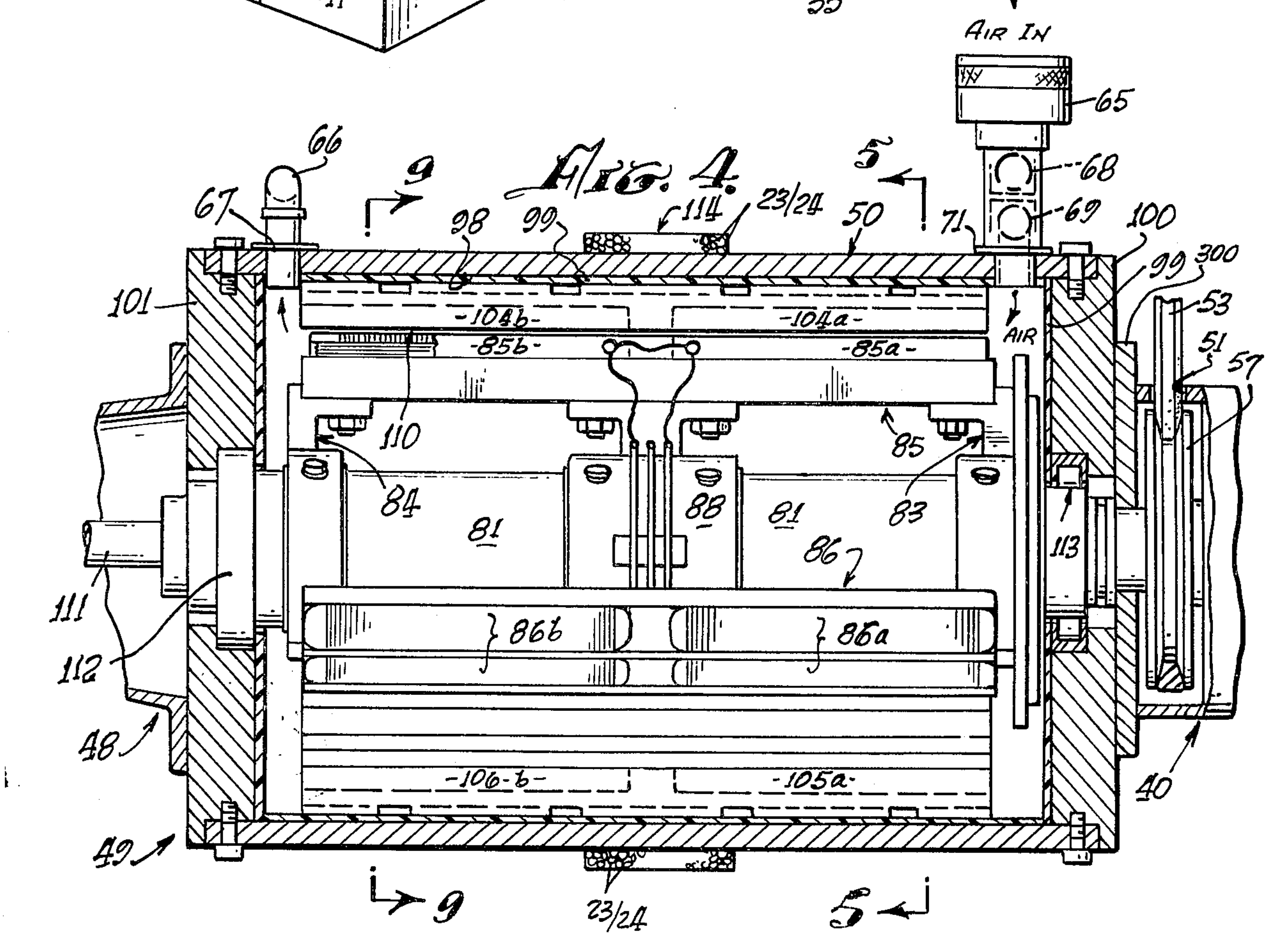
AIR IN
Fig. 4.
23/24
AIR
-104b-
-104a-
-85b-
-85a-
-106-b-
-105a-
23/24

Fig. 5.

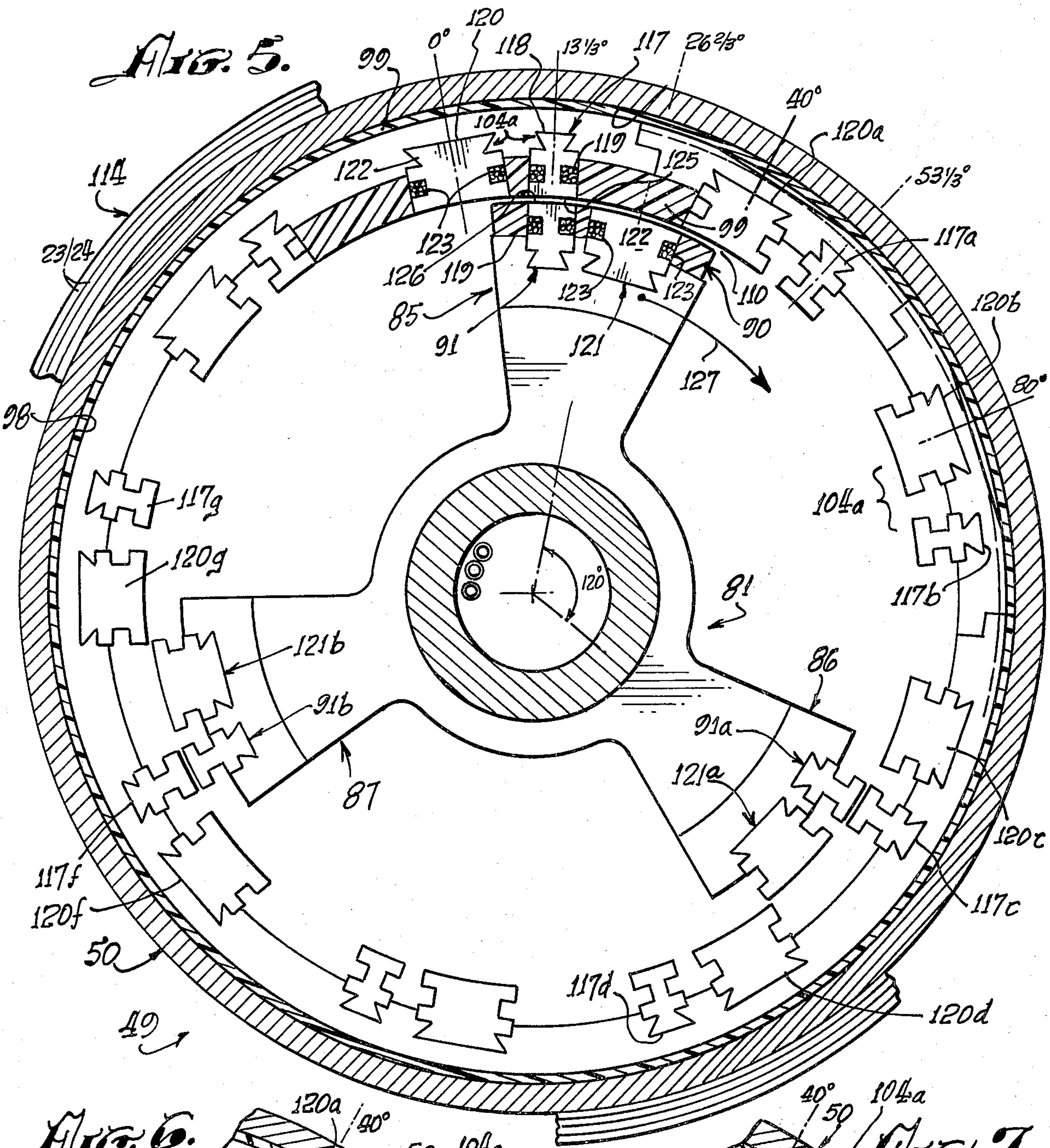
99
0°
120
118
13⅓°
117
26⅔°
104a
119
122
40°
120a
125
53⅓°
114
123
99
117a
23/24
126
119
122
123
110
123
85
91
121
90
120b
127
80°
98
104a
117g
120g
120°
117b
81
121b
86
91b
91a
121a
87
120c
117f
120f
117c
50
117d
120d
49

Fig. 6.

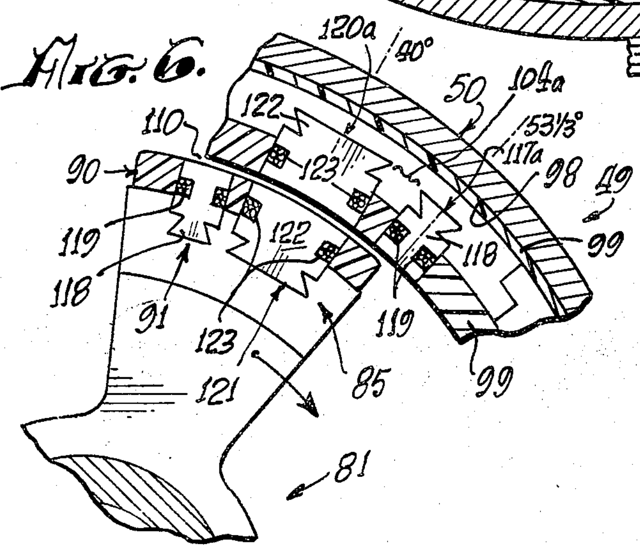
120a
40°
122
50
104a
110
53⅓°
90
123
117a
98
49
119
122
99
118
118
119
91
123
121
85
99
81

Fig. 7.

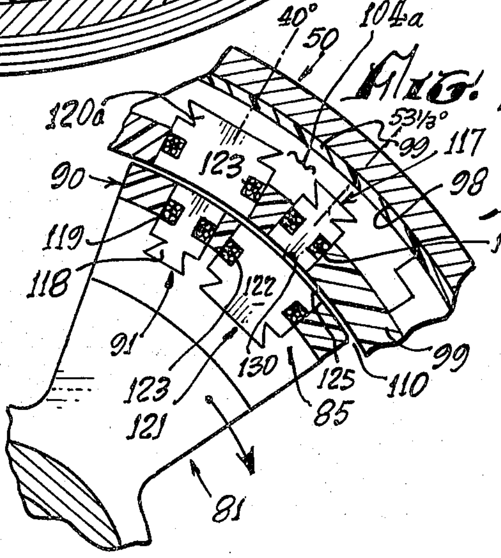
40°
104a
50
120a
53⅓°
99
117
90
123
98
49
119
118
118
122
91
123
130
125
110
99
121
85
81

FIG. 8.

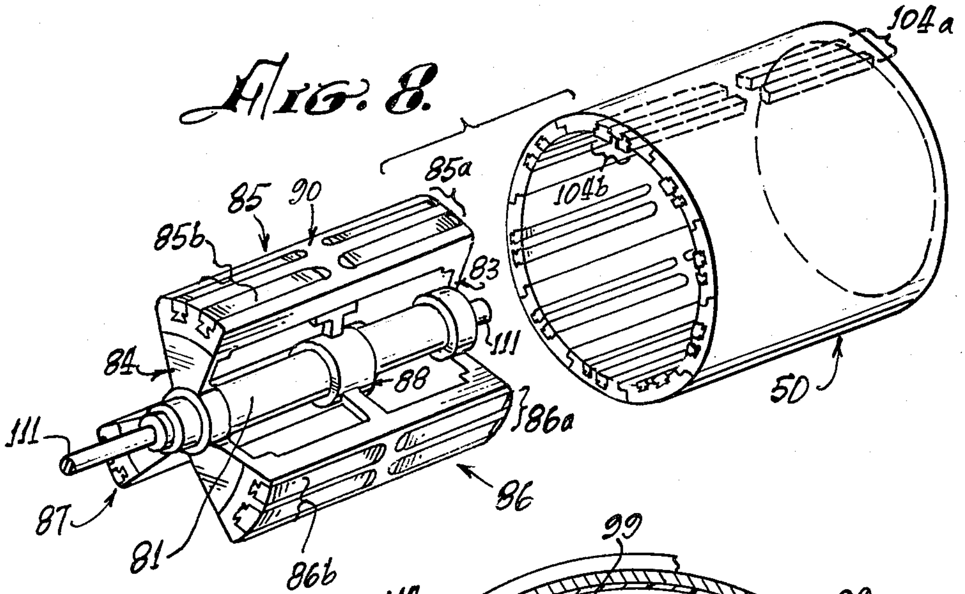
104a
104b
85a
85
90
85b
83
111
84
88
111
86a
87
81
86b
86
50

FIG. 9.

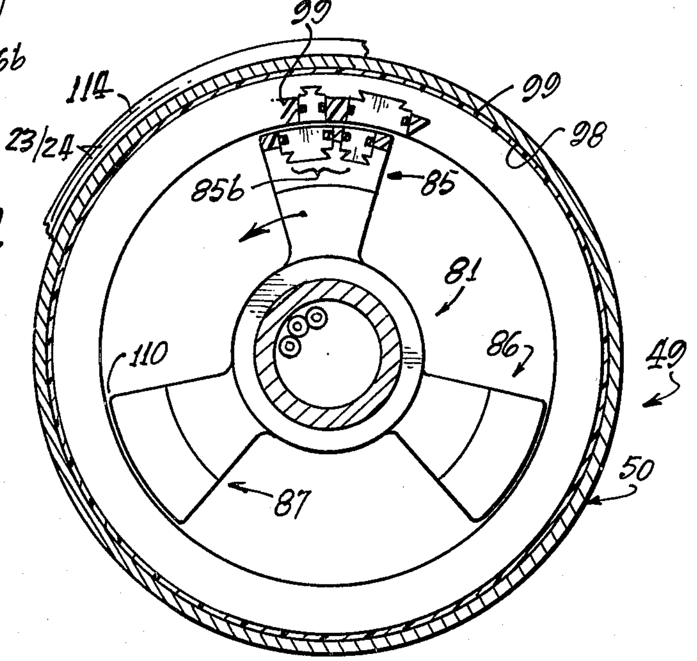
99
114
99
98
23/24
85b
85
81
110
86
49
87
50

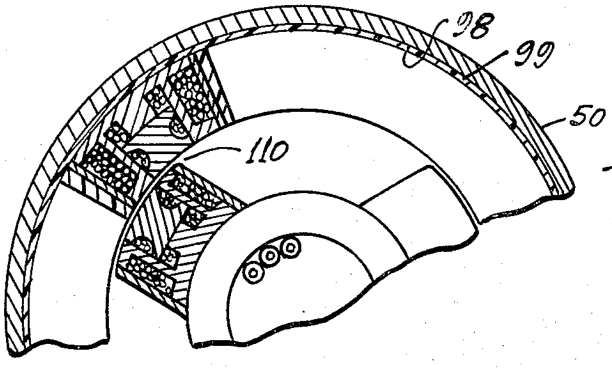

FIG. 10.

FIG. 11.

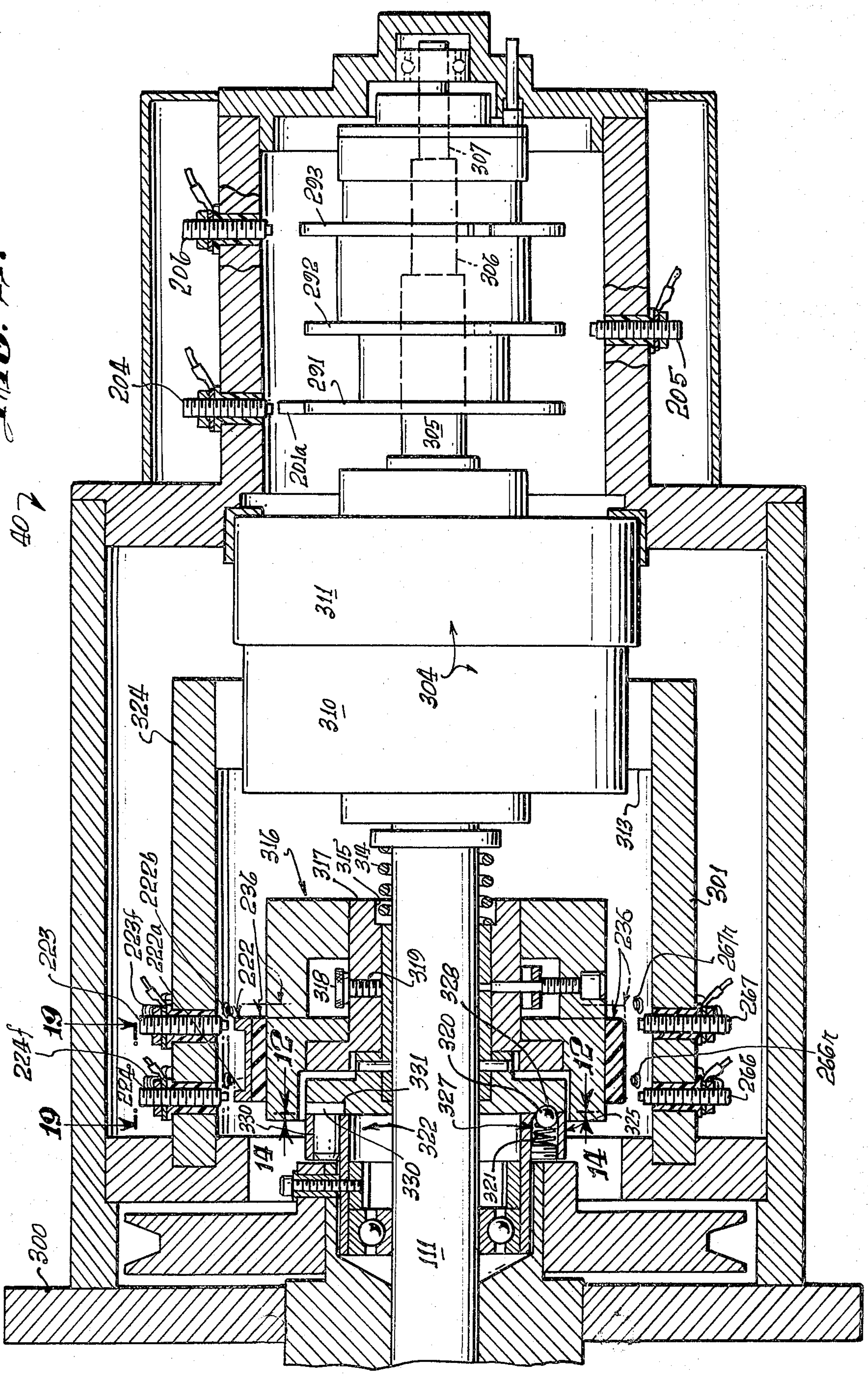

FIG. 12.

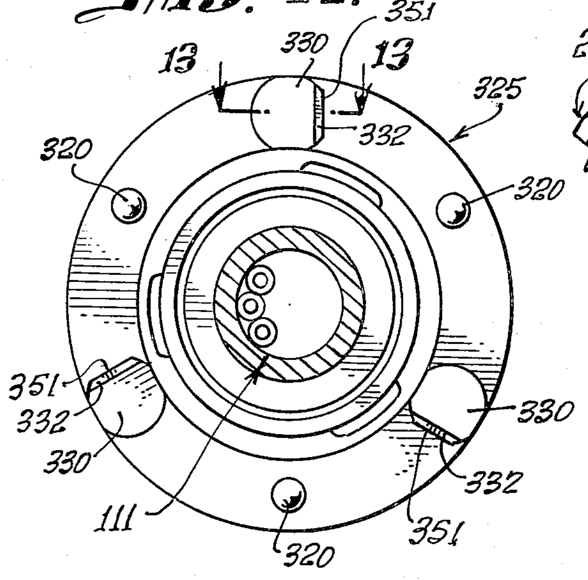
13
330
351
13
325
332
320
320
351
332
330
330
332
111
320
351

FIG. 14.

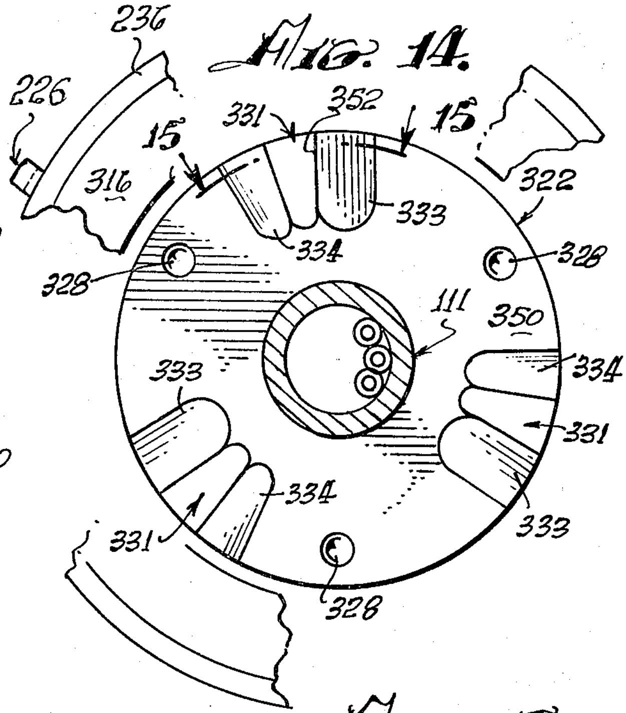
236
226
15
331
352
15
316
333
322
334
328
328
111
350
333
334
331
331
334
333
328

FIG. 13.

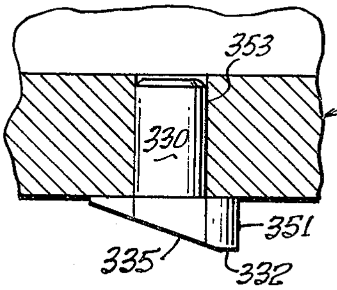
353
325
330
351
335
332

FIG. 15.

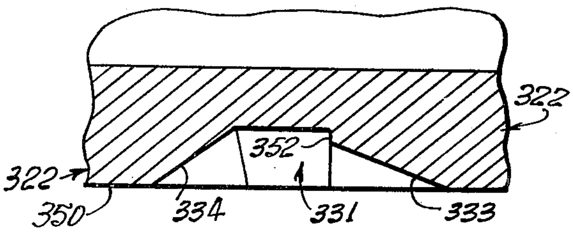
322
352
322
350
334
331
333

FIG. 16.

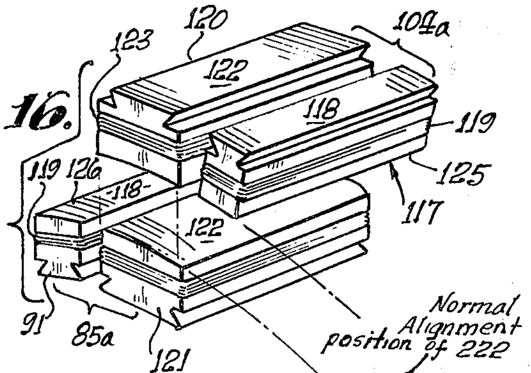
123
120
104a
122
118
119
119
126
118
125
117
122
91
85a
121

FIG. 17.

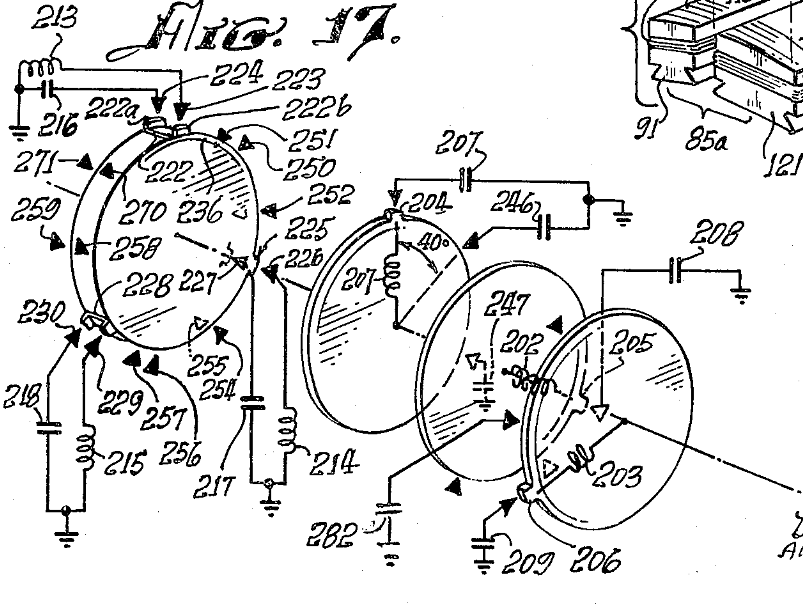
213
224
223
216
222a
222b
251
271
222
250
259
270
236
252
258
225
228
227
226
230
255
218
229
257
254
256
215
217
214
207
204
246
40°
201
208
247
202
205
203
282
209
206

FIG. 19.

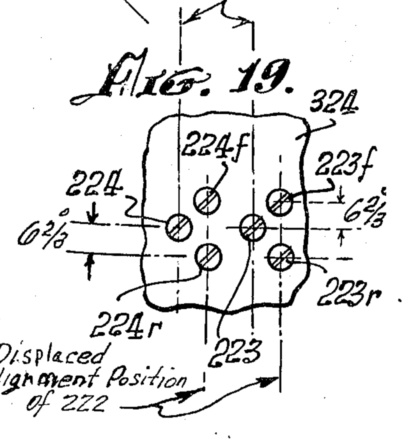
Normal Alignment position of 222
324
224f
223f
224
6⅔°
6⅔°
223r
224r
223
Displaced Alignment Position of 222

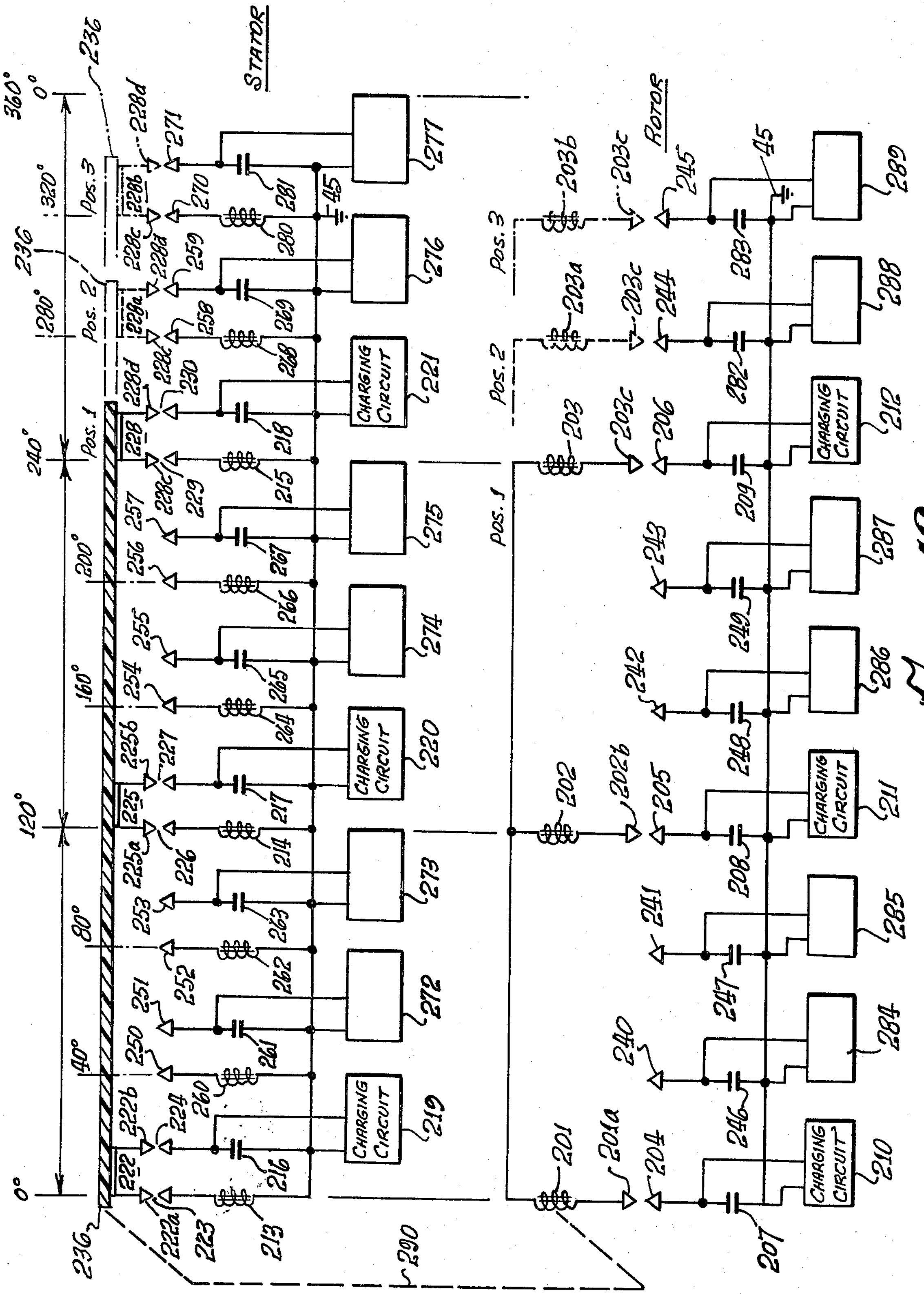
STATOR
ROTOR
360°
0°
320°
280°
240°
200°
160°
120°
80°
40°
0°
Pos. 1
Pos. 2
Pos. 3
CHARGING CIRCUIT

FIG. 18.

PULSED CAPACITOR DISCHARGE ELECTRIC ENGINE

BACKGROUND OF THE INVENTION

1. Field of the Invention

There is no known engine or motor operated on the principle of the present invention, that a capacitor charged to a relatively high voltage from a low-voltage d-c source is discharged across a spark gap to provide current through motor drive coils in the discharge path, these being solenoids which generate motion by magnetic repulsion of juxtaposed pairs of cores. The solenoids are preferably configured in motor and stator assemblies to effect motion of the rotor element with respect to the stator.

The present invention utilizes this principle to provide a rotary motion machine or engine which can develop considerable torque through the magnetic repulsion action of rotor and stator cores wound with coils through which capacitors are discharged synchronously with the positioning of the rotor coils opposite particular stator coils. Similarly, a linear action can be achieved with a stationary electromagnet juxtaposed against a movable electromagnet and the movable electromagnet can perform work with a tool or piston attached thereto.

A novel control mechanism is associated with the rotor is the engine to position discharge elements appropriately to create the desired discharge through the electromagnet coils when the juxtaposed rotor and stator electromagnets are in alignment. The electromagnets in the stator and rotor are so arranged that the control mechanism can advance or retard the discharge points relative to rotor-stator positions for control of rotational speed.

The discharge overshoot or back e.m.f. from the collapsing fields in the coils from the capacitor discharge is used to energize external batteries for conservation of power. The recovered energy thus stored may be used to operate equipment associated with the engine or motive force producing device.

The engine or rotary electric machine of the invention is believed to operate on the principle of conservation of energy, in that once rotation is achieved, current is needed only at the instant of a capacitor discharge in order to advance the rotor. The rotor moves to the next discharge point on the inertia of the repulsion action. The capacitor is recharged during the interval and stores the energy until discharge at the next rotor-stator coil coincidence. Thus, the new engine produces torque and stores the excess energy for subsequent use.

In a linear motion device according to the invention, only a single pulse discharge is needed to perform work.

The applications of the engine include use as an electric automotive engine which is economical and which can regenerate a part of the energy consumed to provide power for other loads in the automotive electric vehicle. As a linear actuator an economical use of power is possible because each stroke will result from a single discharge pulse of a capacitor through a coil.

2. Prior Art

Heretofore, electric engines or motors have operated on the principle that a conductor carrying a current in a magnetic field tends to move perpendicularly to that field; the electromagnetic torque developed by an armature or rotating portion of the motor is proportional to the magnetic flux in the stationary field and to the armature current.

In direct current motors the field is created by current through two or more field coils disposed in opposing magnetic relationship in the motor casing, while current through a rotatable armature positioned in the field is alternatingly reversed in polarity to provide continuous motion. The polarity reversing mechanism is a commutator. Some d-c motors have their field windings electrically in parallel with the rotor armature winding and are called "shunt-wound" motors. Other d-c motors have field and armature windings connected in series. In both series and shunt motors commutators are used for reversing the magnetic polarity of the armature to maintain rotation within the field.

A third type of d-c motor utilizes a permanent magnet field so that the operating current passes only through the armature winding. Such motors also use polarity reversing commutators to maintain direction of rotation. Reversal of direction of motion is effected by reversing the polarity of applied d-c potential.

Control of speed of d-c motors is accomplished basically by decrease or increase of magnetic field flux or the current through the armature. Either or both of these effects can be accomplished by raising or lowering the applied potential. In shunt motors, a series resistance may be varied to produce speed changes. In a permanent magnet motor or series motor, speed variation is best accomplished by voltage variation with a variable resistance in series with the motor d-c supply.

In alternating current motors, as is well known, a rotating magnetic field is created in the stator, and the rotor may be wound with as many poles as there are in the stator, with terminals connected with slip rings, or the rotor may consist of solid bars shorted by rings on each end to form a "squirrel cage" configuration. The speed of an a-c motor depends on the frequency of the applied a-c energy, if the motor is synchronous.

"Universal" motors are operable on either a-c or d-c energy.

In stepping motors, a rotor is moved from one pole to the next adjacent pole with each application of current, the rotor remaining at that position until a next application of current. This is accomplished by switching the current on and off or by pulsing the current. Examples of stepping motors are described in U.S. Pat. No. 3,467,902 to Shimizu, et al., U.S. Pat. No. 3,462,667 to Jackson, and U.S. Pat. No. 3,599,069 to Welch.

Operation of the a-c and d-c motors described above involves the consumption of substantial electric current. These motors can generate electric current when driven externally by a mechanical force. External energy to rotate the generator rotors can be provided by hydroelectric and steam sources or by other electric motors. In some of these systems, a d-c motor source drives an a-c generator for conversion of d-c energy to a-c energy or a d-c motor may drive a d-c generator which delivers a higher voltage than the source.

An extensive prior art search by the applicant uncovered no capacitor-discharge-operated motor resembling that of the present invention. All motors of the patents located in the search employed direct electrical connection between coils and electric power sources. Where selective switching is involved, semiconductor devices are employed, such as silicon-controlled rectifiers. Capacitors are used only for starting and phasing purposes, and not for basic motor operation from the discharge thereof, as in this invention.

SUMMARY OF THE INVENTION

This invention relates to electric motors or engines, and more particularly to a new electric machine including electromagnetic poles in a stator configuration and electromagnetic poles in a rotor configuration wherein in one form thereof the rotor is rotatable within the stator configuration and where both are energized by capacitor discharges through rotor and stator electromagnets at the instant of the alignment of a rotor electromagnet with a stator electromagnet. The rotor electromagnet is repelled from the stator electromagnet by the discharge of the capacitor through the coils of both the stator and rotor electromagnets at the same instant.

In an exemplary rotary engine according to this invention, rotor electromagnets may be disposed 120° apart on a central shaft and major stator electromagnets may be disposed 40° apart in the motor housing about the stator periphery. Other combinations of rotor elements and stator elements may be utilized to increase torque or rotation rate.

In another form, a second electromagnet is positioned to one side of each of the major stator electromagnets on a center line 13⅓° from the center line of the stator magnet, and these are excited in a predetermined pattern or sequence. Similarly to one side of each major rotor electromagnet is a second electromagnet spaced on a 13⅓° center line from the major rotor electromagnet. Electromagnets in both the rotor and stator assemblies are identical, the individual electromagnets of each being aligned axially and the coils of each being wired so that each rotor electromagnetic pole will have the same magnetic polarity as the electromagnet in the stator with which it is aligned and which it is confronting at the time of discharge of the capacitor.

Charging of the discharge capacitor or capacitors is accomplished by an electrical switching circuit wherein electrical energy from a battery or other source of d-c potential may be applied in alternating polarity to ignition coils or other voltage step-up arrangements from which a high voltage d-c potential is derived through rectification by diodes.

The capacitor charging circuit comprises a pair of high frequency switchers which feed respective automotive-type ignition coils employed as step-up transformers. The "secondary" of each of the ignition coils provides a high voltage square wave to a half-wave rectifier to generate a high voltage output pulse of d-c energy with each switching alternation of the high frequency switcher. Only one polarity is used so that a unidirectional pulse is applied to the capacitor bank being charged.

Successive unidirectional pulses are accumulated on the capacitor or capacitor bank until discharged. Discharge of the bank of capacitors occurs across a spark gap by arc-over. The gap spacing determines the voltage at which discharge or arc-over occurs. An array of gaps is created by fixed elements in the engine housing and moving elements positioned on the rotor shaft. At the instant when the moving gap elements are positioned opposite fixed elements during the rotor rotation, a discharge occurs through the coils of the aligned rotor and stator electromagnets to produce the repulsion action between the stator and rotor electromagnet cores.

A plurality of fixed gap elements are arrayed in the motor housing to correspond to the locations of the stator electromagnets in the housing. The rotor gap elements correspond to the positions of the rotor electromagnets on the rotor so that at the instant of correct alignment of the gaps the capacitors are discharged to produce the necessary current through the stator and rotor coils to cause the electromagnets to repel one another.

The charging circuits are arranged in pairs, and are such that the discharge occurs through both rotor and stator windings of the electromagnets, which are opposite one another when the spark gap elements are aligned and arc-over.

The speed of the rotor can be changed by means of a clutch mechanism associated with the rotor. The clutch shifts the positions of the rotor gap elements so that the discharge will energize the stator coils in a manner to advance or retard the time of discharge with respect to the normal rotor/stator alignment positions. The discharge through the rotor and stator then occurs when the rotor has passed the stator 6⅔° for speed advance.

By causing the discharge to occur when the rotor position is approaching the stator, the repulsion pulse occurs 6⅔° before the alignment position of the rotor and stator electromagnets, thus slowing the speed.

The clutch mechanism for aligning capacitor discharge gaps for discharge is described as a control head. It may be likened to a firing control mechanism in an automotive combustion engine in that it "fires" the electromagnets and provides a return of any discharge overshoot potential back to the battery or other energy source.

The action of the control head is extremely fast. From the foregoing description, it can be anticipated that an increase in the speed or a decrease in speed of rotation can occur within the period in which the rotor electromagnet moves between any pair of adjacently located electromagnets in the stator assembly, which are 40° apart in the exemplary engine according to the invention. Thus, speed changes can be effected in a maximum of one-ninth of a revolution.

The rotor speed-changing action of the control head and its structure are believed to be further novel features of the invention, in that they maintain normal 120° firing positions during uniform speed or rotation conditions, but shift to ±6⅔° longer or shorter intervals for speed change by the novel shift mechanism in the rotor clutch assembly.

Accordingly, the preferred embodiment of this invention is an electric rotary engine wherein motor torque is developed by discharge of high potential from a bank of capacitors through stator and rotor electromagnet coils when the electromagnets are in alignment. The capacitors are charged from batteries by a switching mechanism, and are discharged across spark gaps set to achieve the discharge of the capacitor charge voltage through the electromagnetic coils when the gaps and predetermined rotor and stator electromagnet pairs are in alignment.

Exemplary embodiments of the invention are herein illustrated and described. These exemplary illustrations and description should not be construed as limiting the invention to the embodiments shown, because those skilled in the arts appertaining to the invention may conceive of other embodiments in the light of the description within the ambit of the appended claims.

BREIF DESCRIPTION OF THE DRAWINGS

FIG. **1** is an explanatory schematic diagram of a capacitor charging and discharging circuit utilized in the present invention;

FIG. **2** is a block diagram of an exemplary engine system according to the invention;

FIG. **3** is a perspective view of a typical engine system according to the invention, coupled to an automotive transmission;

FIG. **4** is an axial sectional view taken at line **4—4** in FIG. **3**;

FIG. **5** is a sectional view taken at line **5—5** in FIG. **4**;

FIGS. **6** and **7** are fragmentary sectional views, corresponding to a portion of FIG. **5**, illustrating successive advanced positions of the engine rotor therein;

FIG. **8** is an exploded perspective view of the rotor and stator of the engine of FIGS. **3** and **4**;

FIG. **9** is a cross-sectional view taken at line **9—9** of FIG. **4**;

FIG. **10** is a partial sectional view, similar to the view of FIG. **9**, illustrating a different configuration of electromagnets in another engine embodiment of the invention;

FIG. **11** is a sectional view taken at line **11—11** in FIG. **3**, illustrating the control head or novel speed change controlling system of the engine;

FIG. **12** is a sectional view, taken at line **12—12** in FIG. **11**, showing a clutch plate utilized in the speed change control system of FIG. **11**;

FIG. **13** is a fragmentary view, taken at line **13—13** in FIG. **12**;

FIG. **14** is a sectional view, taken at line **14—14** in FIG. **11**, showing a clutch plate which cooperates with the clutch plate of FIG. **12**;

FIG. **15** is a fragmentary sectional view taken at line **15—15** of FIG. **13**;

FIG. **16** is a perspective view of electromagnets utilized in the present invention;

FIG. **17** is a schematic diagram showing cooperating mechanical and electrical features of the programmer portion of the invention;

FIG. **18** is an electrical schematic diagram of an engine according to the invention, showing the electrical relationships of the electromagnetic components embodying a new principle of the invention; and

FIG. **19** is a developed view, taken at line **19—19** of FIG. **11**, showing the locations of displaced spark gap elements of the speed changing mechanism of an engine according to the invention.

DESCRIPTION OF THE PREFERRED EMBODIMENT

As hereinbefore mentioned, the basic principle of operation of the engine of the invention is the discharge of a capacitor, across a spark gap through an inductor. When a pair of inductors is used, and the respective magnetic cores thereof are arranged opposite and another in magnetic polarity repulsion relation, the discharge through them causes the cores to repel each other with considerable force.

Referring to the electrical schematic diagram of FIG. **1**, a battery **10** energizes a pulse-producing vibrator mechanism **16**, which may be of the magnetic type incorporating an armature **15** moving between contacts **13** and **14**, or of the transistor type (not shown) with which a high frequency bipolar pulsed output is produced in primary **17** of transformer **20**. The pulse amplitude is stepped-up in secondary **19** of transformer **20**. Wave form **19***a* represents the bidirectional or bipolar pulsed output. A diode rectifier **21** produces a unidirectional pulse train, as indicated at **21***a*, to charge capacitor **26**. A delay coil **23** is connected in series with the unipolar pulsed output to capacitor **26**. Successive unidirectional pulses of wave **21***a* charge capacitor **26** to a high level, as indicated at **26***a*, until the voltage amplitude at point A reaches the breakdown potential of spark gap **30**. At the breakdown of spark gap **30**, capacitor **26** discharges across the arc created through the inductor coil **28**. A current pulse is produced which magnetizes core **28***a*. Simultaneously, another substantially identical charging system **32** produces a discharge through inductor **27** across spark gap **29** to magnetize core **27***a*. Cores **28***a*, **27***a* are wound with coils **28**, **27** so that their magnetic polarities are the same. As the cores **27***a*, **28***a* confront one another, they tend to fly apart when the discharge occurs through coils **27** and **28** because of repulsion of identical magnetic poles, as indicated by arrow **31**. If core **28***a* is fixed or stationary and core **27***a* is movable, then core **27***a* may have tools **33** attached to it to perform work when the capacitor discharges.

Referring to FIGS. **1** and **2**, a d-c electrical source or battery **10** energizes pulsators **36** (including at least two vibrators **16** as previously described) when switch **11** between the battery **10** and pulsator **36** is closed, to apply relatively high frequency pulses to the primaries of transformers **20**. The secondaries of transformers **20** are step-up windings which apply bipolar pulses, such as pulses **19***a* (FIG. **1**) to the diodes in converter **38**. The rectified unidirectional pulsating output of each of the diodes in converter **38** is passed through delay coils **23**, **24**, thus forming a harness **37** wound about the case of the engine, as hereinafter described, which is believed to provide a static floating flux field. The outputs from delay lines **37** drive respective capacitors in banks **39** to charge the capacitors therein to a relatively high charge potential. A programmer and rotor and stator magnet control array **40**, **41**, **42** is formed by spark gaps positioned, as hereinafter described, so that at predetermined positions of the rotor during rotation of the engine, as hereinafter described, selected capacitors of capacitor banks **39** will discharge across the spark gaps through the rotor and stator electromagnets **43**, **44**. The converters **38**, magnetic harness **37**, capacitor banks **39**, programmer **40**, and controls **41**, **42** from a series circuit path across the secondaries of transformers **20** to the ground, or point of reference potential, **45**. The capacitor banks **39** are discharged across the spark gaps of programmer **40** (the rotor and stator magnet controls **41**, **42**). The discharge occurs through the coils of stator and rotor electromagnets **43**, **44** to ground **45**. Stator and rotor electromagnets are similar to those shown at **27**, **27***a*, **28**, **28***a* in FIG. **1**.

The discharge through the coils of stator and rotor electromagnets **43**, **44** is accompanied by a discharge overshoot or return pulse, the output of which is applied in an appropriate polarity to a secondary battery **10***a* to store this excess energy. The overshoot pulse returns to battery **10***a* because after discharge the only path open is that to battery **10***a*, since the gaps in **40**, 41 and **42** have broken down, because the capacitors in banks **39** are discharged and have not yet recovered the high voltage charge from the high frequency pulsers **36** and converter rectifier units **38.**

In the event of a misfire in the programmer control circuits **40, 41, 42,** the capacitors are discharged through a rotor safety discharge circuit **46** and returned to batteries **10–10***a*, adding to their capacity. The circuit **46** is connected between the capacitor banks **39** and batteries **10, 10***a*.

Referring to FIG. **3**, a motor or engine **49** according to the present invention is shown connected with an automotive transmission **48.** The transmission **48** represents one of many forms of loads to which the engine may be applied. A motor housing **50** encases the operating mechanism hereinafter described. The programmer **40** is axially mounted at one end of this housing. Through apertures **51, 52** a belt **53** couples to a pulley **57** (not shown in this view) and to an alternator **54** attached to housing **50.** A pulley **55** on the alternator has two grooves, one for belt **53** to the drive pulley **58** on the shaft (not shown) of the engine **49,** and the other for a belt **58** coupled to a pulley **59** on a pump **60** attached to housing **50.** A terminal box **61** on the housing interconnects means between the battery assembly **62** and motor **49** via cables **63** and **64.**

An intake **65** for air is coupled to pump **60** via piping **68, 69** and from pump **60** via tubing or piping **66, 70** to the interior of housing **50** via coupling flanges **67** and **71.** The air flow tends to cool the engine, and the air may preferably be maintained at a constant temperature and humidity so that a constant spark gap discharge condition is maintained. A clutch mechanism **80** is provided on programmer **40.**

Referring to FIGS. **4, 5** and **9,** rotor **81** has spider assemblies **83, 84** with three electromagnet coil assembly sets mounted thereon, two of which are shown in FIG. **4,** on **85** at **85***a* and **85***b*, and on **86** at **86***a* and **86***b*. One of the third electromagnet coil assemblies, designated **87***a*, is shown in FIG. **5**, viewed from the shaft end. As more clearly shown in the perspective view of FIG. **8,** a third spider assembly **88** provides added rigidity and a central support for the rotor mechanism on shaft **81.**

The electromagnet sets **85***a* and **85***b*, 86*a* and **86***b*, **87***a* and **87***b*, disposed on rotor **81** and spiders **83, 84,** and **88** each comprise pairs of front units **85***a*, 86*a*, 87*a* and pairs of rear units **85***b*, 86*b*, 87*b*. Each pair consists of a major electromagnet and a minor electromagnet, as hereinafter described, which are embedded in an insulating material **90,** which insulates the electromagnet coil assemblies from one another and secures the electromagnets rigidly in place on the spider/rotor cage **81,** 83, 84, 88.

The interior wall **98** of housing **50** is coated with an electrically insulating material **99** in which are embedded electromagnet coils, as hereinafter described, and the interiors of end plates **100, 101** of the housing **50.** On the insulating surface **98** of housing **50** is mounted a series of stator electromagnet pairs **104***a*, identical with electromagnet pairs **85***a*, 86*a*, 87*a*, etc. Electromagnet pairs such as **104***a* or **105***a* are disposed every 40° about the interior of housing **50** to form a stator which cooperates with the rotor **81–88.** An air gap **110** of very close tolerance is defined between the rotor and stator electromagnets, and air from pump **65** flows through this gap.

As shown in FIG. **8,** the electromagnet assemblies, such as **85** through **87,** of the rotor and magnet assemblies, such as **104***a* in the stator are so embedded in their respective insulative plastic carriers (rotor and stator) that they are smoothly rounded in a concave contour for the stator, and in a convex contour on the rotor to permit smooth and continuous rotation of rotor **81** in stator housing **50.** The air gap **110** is uniform at all positions of any rotor element within the stator assembly, as is clearly shown in FIG. **16.**

The rotor **81** and spiders **83, 84, 88** are rigidly mounted on a shaft **111** journaled in bearing assemblies **112, 113** which are of conventional type, for easy rotation of the rotor on shaft **111** within housing **50.**

Around the central outer surface of housing **50** are wound a number of turns of wire **23, 24** to provide a static flux coil **114** as hereinbefore described, which is a delay line, as previously described.

FIGS. **5, 6, 7** and **9** are cross-sectional views of the rotor assembly **81–88,** arranged to show the positioning and alignment of the rotor and stator electromagnet coil assemblies at successive stages of the rotation of rotor **81–88** through a portion of a cycle of operation thereof. For example, in FIG. **5** the rotor assembly **81–88** is shown so positioned that a minor rotor electromagnet assembly **91** is aligned with a minor stator electromagnet assembly **117.**

As shown in further detail in FIG. **16,** minor electromagnet assembly **117** consists of an iron core **118,** grooved so that there may be wound thereabout a coil of wire **119.** Core **118** is the same in stator electromagnet **117** as it is in rotor electromagnet **91.**

As a position 13⅓° to the right of rotor electromagnet **91,** as viewed in FIGS. **5** and **16,** there is a second or major rotor electromagnet **121** which has a winding **123** about its core **122.** The electromagnets **91, 121** are the pair **85***a* of FIGS. **4** and **8.**

At a position 13⅓° to the left of stator electromagnet **117,** as viewed in FIG. **5,** there is a second or major stator electromagnet **120** whose core **122** is of the same configuration as core **122** of rotor electromagnet **121.** A winding **123** about core **122** of electromagnet **120** is of the same character as winding **123** on electromagnet **121.**

Electromagnet assembly pair **85***a* on the rotor is identical in configuration with that of the electromagnet stator assembly pair **104***a* except for the position reversal of the elements **117–120** and **91–121** of the respective pairs.

There are nine pairs of electromagnets **120–117** (**104***a*) disposed at 40° intervals about the interior of housing **50.** The center line of core **122** of electromagnet **120** is positioned 13⅓° to the left of the center line of the core **118** of electromagnet **117.** Three pairs of electromagnets **85***a*, 86*a*, 87*a* are provided on rotor assembly **81–88** as shown in FIG. **5.**

Other combinations are possible, but the number of electromagnets in the rotor should always be an integral fraction of the number of electromagnets in the stator. As shown in FIG. **8,** for the rotor assembly **85***a*, 85*b*, there are three of each of the front and back pairs of electromagnetic assemblies. Similarly, as shown in FIGS. 4 and **8,** there are nine front and back pairs of electromagnets in the stator such as **104***a* and **104***b*.

In order to best understand the operation of the rotor **81–88** rotating within the stator housing **50** of an engine according to this invention, the positions of rotor electromagnets **91** and stator electromagnets **117** are initially exactly in line at the 13⅓° peripheral starting position marked on the vertical center line of FIG. **5.** The winding direction of the coils of these magnets is such that a d-c current through the coils **119** will produce a particular identical magnetic polarity on each of the juxtaposed surfaces **125** of magnet **117,** and **126** of magnet **91,** (FIG. **5**). FIGS. **16** and **6** illustrate the next step in the motion wherein the two major electromagnets, **120** in the stator and **121** in the rotor, are in alignment.

When the d-c discharges from the appropriate capacitors in banks **39** occur simultaneously across spark gaps through the coils **119** of electromagnets **117** and **91,** at the instant of their alignment, their cores **118** will repel one another to cause rotor assembly **81–88** to rotate clockwise in the direction indicated by arrow **127.** The system does not move in the reverse direction because it has been started in the clockwise direction by the alternator motor **54** shown in FIG. **3,** or by some other starter means. If started counterclockwise, the rotor will continue to move counterclockwise.

As hereinbefore noted, the discharge of any capacitor occurs over a very short interval across its associated spark gap, and the resulting magnetic repulsion action imparts motion to the rotor. The discharge event occurs when electromagnets **117** and **91** are in alignment. As shown in FIG. **5,** rotor electromagnet **91*a*** is aligned with stator electromagnet **117*c*,** and rotor electromagnet **91*b*** is aligned with stator electromagnet **117*e*** at the same time that similar electromagnets **117** and **91** are aligned. A discharge occurs through all six of these electromagnets simultaneously (that is, **117,** 91; **117*c*,** 91*a*, 117*e* and **91*b***). A capacitor and a spark gap are required for each coil of each electromagnet. Where, as in the assembly shown in FIG. **8,** front and back pairs are used, both the axial in-line front and back coils are energized simultaneously by the discharge from a single capacitor or from a bank of paralleled capacitors such as **25, 26** (FIG. **1**). Although FIGS. **4** and **8** indicate the use of front and back electromagnets, it should be evident that only a single electromagnet in any stator position and a corresponding single electromagnet in the rotor position, may be utilized to accomplish the repulsion action of the rotor with respect to the stator. As stated, each electromagnet requires a discharge from a single capacitor or capacitor bank across a spark gap for it to be energized, and the magnetic polarity of the juxtaposed magnetic core faces must be the same, in order to effect the repulsive action to produce the rotary motion.

Referring to FIGS. **5** and **6,** the repulsion action causes the rotor to move 13⅓° clockwise, while electromagnets **91, 91*a*** and **91*b*** move away from electromagnets **117, 117*c*** and **117*e*** to bring electromagnets **121, 121*a*** and **121*b*** into respective alignment with electromagnets **120*a*,** 120*d* and **120*f*.** At this time, a capacitor discharge across a spark-gap into their coils **123** occurs, thus moving the rotor. Another 13⅓° ahead, as shown in FIG. **7,** major electromagnets **121, 121*a*** and **121*b*** come into alignment with minor electromagnets **117*a*,** 117*d* and **117*f*,** at which time a discharge occurs to repeat the repulsion action, this action continuing as long as d-c power is applied to the system to charge the capacitors in the capacitor banks.

FIG. **18** further illustrates the sequencing of the capacitor discharges across appropriate spark gap terminal pairs. Nine single stator coils and three single rotor coils are shown with their respective interconnections with the spark gaps and capacitors with which they are associated for discharge. When the appropriate spark gap terminals are aligned, at the points in the positioning of the rotor assembly for most effective repulsion action of juxtaposed electromagnet cores, the discharge of the appropriate charged capacitors across the associated spark gap occurs through the respective coils. The capacitors are discharged in sets of three through sets of three coils at each discharge position, as the rotor moves through the rotor positions. In FIG. **18,** the rotor electromagnets are positioned linearly, rather than on a circular base, to show the electrical action of an electric engine according to the invention. These motor electromagnets **201, 202,** and **203** are aligned with stator electromagnets **213, 214** and **215** at 0°, 120° and 450° respectively. The stator electromagnets are correspondingly shown in a linear schematic as if rolled out of the stator assembly and laid side by side. For clarity of description, the capacitors associated with the rotor operation **207, 208, 209** and **246, 247,** 248, 249, 282 and **283** are arranged in vertical alignment with the respective positions of the rotor coils **201, 202, 203** as they move from left to right, this corresponding to clockwise rotation of the rotor. The stator coils **213, 214, 215, 260, 261, 262, 263, 264, 265,** 266, etc. and capacitor combinations are arranged side by side, again to facilitate description.

An insulative disc **236** (shown in FIG. **17** as a disc, but opened out linearly in FIG. **18**) has mounted thereon three gap terminal blocks **222, 225** and **228.** Each block is rectangularly U-shaped, and each interconnects two terminals with the base of the U. Block **222** has terminals **222*a*** and **222*b*,** block **225** has terminals **225*a*** and **225*b*,** and block **228** has terminals **228*c*** and **228*d*.** When insulative disc **230** is part of the rotor, as indicated by mechanical linkage **290,** it can be seen that terminal U **222** creates a pair of gaps with gap terminals **223** and **224,** respectively. Thus, when the voltage on capacitor **216** from charging circuit **219** is of a value which will arc over the air spaces between **222*a*** and **223,** and between **222*b*** and **224,** the capacitor **216** will discharge into the coil of electromagnet **213** to ground. Similarly gap terminal U **225** forms a dual spark gap with gap terminals **226** and **227** to result in arc-over when the voltage on capacitor **217,** charged by charging circuit **220,** discharges into the coil of electromagnet **214.** Also, U-gap terminal **228** with terminals **228*c*** and **228*d*,** creates a spark gap with terminals **229** and **230** to discharge capacitor **218,** charged by charging circuit **221,** into coil **215.** At the same time, rotor coils, **201, 202** and **203** across gaps **201*a*–204,** 202*b*–**205** and **203*c*–206** each receives a discharge from respective capacitors **207, 208,** and **209.**

When the electromagnet coils **213, 214, 215** and **201, 202, 203** are energized, the repulsion action causes the rotor assembly to move to position **2** where a new simultaneous group of discharges occurs into rotor coils **201, 202,** and **203** from capacitors **246, 248,** and **282** across gaps **201*a*–240, 202*b*–242** and **203*c*–244.** Simultaneously, because gap-U-elements **222, 225** and **228** have also moved to position **2** with the rotor assembly, capacitor **261** is discharged through electromagnet coil **260,** capacitor **265** is discharged through electromagnet coil **264,** and capacitor **269** is discharged through electromagnet coil **268** in alignment with position 2 of the rotor electromagnet coils, thus to cause the rotor electromagnets to move to position 3 where the discharge pattern is repeated now with capacitors **247, 249** and **283** discharging through the rotor electromagnet coils **201, 202, 203,** and the capacitors **263, 267** and **281** discharging respectively through stator electromagnet coils **262, 266** and **280.**

After each discharge the charging circuits **219–221** and **272–277** for the stator capacitors, and **210–212** and **284–289** for the rotor capacitors, are operated continuously from a battery source, as described earlier with reference to FIG. **1,** to constantly recharge the capacitors to which each is connected. Those versed in the art will appreciate that, as each capacitor discharges across an associated spark gap, the resulting drop in potential across the gap renders the gap an open circuit until such time as the capacitor can recharge to the arc-over level of the gap. This recharge to a discharge potential occurs before a rotor element arrives at the next position in question.

The mechanical schematic diagram of FIG. **17** further clarifies the operation of the spark-gap discharge programming system. A forward disc **236** of an electrically insulative material, has thereon the set of U-shaped gap terminal connectors previously described. These are positioned at 0°, 120° and 240° respectively. In FIG. **17,** schematic representations of the position of the coil and capacitor arrangements at the start of a cycle are shown to correspond to the description hereinabove with reference to FIG. **18.** Accordingly, the coil and capacitor combinations **213**/216, 214/217, and **215**/218 are shown connected with their gap terminals, respectively, **223**/224, 226/227 and **229**/230. On the rotor coil and capacitor connection three separate discs **291, 292** and **293** are shown, each with a single gap terminal. The discs **291–293** are rotated so as to position their respective gap terminals **201***a*, 202*b* and **203***c* at 120° increments, with the 0° position corresponding to the 0° position of U-gap terminal **222** on disc **230.**

Representative gap terminals are shown about the peripheries of discs **230, 291–293** to clearly indicate how, as the discs turn in unison, the gap alignments correspond so that three rotor coil/capacitor positions always line up with three stator coil/capacitors at 120° intervals about the rotary path, producing an alignment every 40°, there being nine stator coils. Thus there are three simultaneous discharges into stator coils and three into rotor coils at each 40° position. Nine positions displaced 40° apart provide a total of 27 discharge points for capacitors into the rotor coils and 27 discharge points for capacitors into the stator coils in one revolution.

It will be understood that, as illustrated in FIGS. **17** and **18,** nine individual electromagnet coils are shown in the stator and three in the rotor, in order to show in its simplest form how the three rotor electromagnets are stepped forward from alignment with three of the stator electromagnets, when the appropriate spark gaps are in alignment, to effect the discharge of capacitors through juxtaposed pairs of rotor/stator electromagnets. The repulsion moves the rotor electromagnet from the stator electromagnet to the next alignment position advanced at an arc of 40° from the preceding alignment position. In the interval until another rotor electromagnet, 120° removed, is aligned with the stator electromagnet that had just been excited, the associated capacitor is recharged to a potential which will cause the spark gap to break down to produce another discharge of this capacitor on the next revolution. Thus the rotor moves from one position to the next, with capacitor discharges occurring each 40° of rotation, a total of nine per revolution. It should be obvious that, with other rotor/stator combinations, the number of electromagnet coincidences and spark-gap discharges will vary. For example, with the coil pairs shown in FIGS. **4** through **8,** a total of 27 discharges will occur. Although there are 18 stator electromagnets and three rotor electromagnets, the discharge pattern is determined by the specific spark gap arrangement.

The rotor/stator configuration of FIGS. **5** and **8,** involving the major and minor pairs of electromagnets such as **85***a* and **104***a* (the terms "minor" and "major" referring to the difference in size of the elements), include nine pairs of electromagnets in the stator, such as **104***a*, with three electromagnet pairs of the rotor, such as **85***a*. Because of the 13⅓° separation between the major and minor electromagnets in the rotor pair **85***a*, with the same separation of minor and major electromagnets of the stator pair **104***a*, the sequence of rotation and discharge described above, with respect to the illustrative example of FIG. **5,** involves the following:

1. A minor element **117** of stator pair **104***a* is aligned with the minor element **91** of rotor pair **85***a*. On the discharge, this moves the rotor ahead 13⅓°.
2. The major rotor element **122** of the pair **85***a* now is aligned with the major stator element **120***b* of the next stator electromagnet pair, in the stator array as shown in FIG. **6.** On the discharge, the rotor moves ahead 13⅓°.
3. This brings the minor rotor electromagnet **91** into alignment with the major stator electromagnet **120***b* of pair **104***d*, and the major electromagnet **122** (just discharged) of pair **85***a* into alignment with minor electromagnet **117***b* of pair **104***d*, and the rotor spark gap ements into alignment with a different position of gap elements connected with capacitors not discharged in the previous position of the rotor. It should be remembered at this point that it is the positioning of a rotatable spark gap array, similar to that illustrated in FIGS. **17** and **18,** which controls the time of discharge of capacitors connected to these gap terminals. Therefore, any electromagnet can be energized twice successively from separate capacitors as the rotor brings appropriate gap terminals into alignment with the coil terminals of a particular electromagnet.

Thus, although major electromagnet **120***b* of pair **104***d* has just been energized as described above, it can now be energized again along with minor rotor electromagnet **91** in step **3,** because the rotor moved to a new set of terminals of the spark gap arrays connected with capacitors not yet discharged. These capacitors now discharge through rotor electromagnet **91** and stator electromagnet **120***b*, causing rotor to move ahead another 13⅓°, thus again aligning two minor electromagnets again, these being **117***b* of stator pair **104***d* and **91** of rotor pair **85***a*. The rotor has now moved 40° since step **1** above. The sequence is repeated, and it is to be noted that at each 13⅓° step, the discharges rotate the rotor another 13⅓°. There are 27 steps per revolution with nine stator coil pairs. The discharge sequence is not uniform, as is shown in Table I. In the stator, three major electromagnets 120° apart are energized twice in sequence followed by a hiatus of one step while three minor electromagnets of the stator, 120° apart, are energized during the hiatus. In the rotor the major electromagnets are energized during a hiatus step following two minor electromagnet energization steps. A total of 27 energizations are thus accomplished in the nine pairs of coils of the stator.

In Table I, the leftmost column depicts the location of each rotor arm **85, 86, 87** at an arbitrarily selected step No. **1** position. For example, in step **1** rotor arm **85** has a minor stator and minor rotor electromagnet in alignment for capacitors to discharge through them simultaneously at the 13⅓° position.

TABLE I

CAPACITOR DISCHARGE SEQUENCE FOR ONE REVOLUTION OF ROTOR IN SYSTEM OF FIGS. 5 AND 8

Step No. Rotor Arm				Rotor Electromagnet		Stator Electromagnet	
87	85	86	Rotor Angle	Minor	Major	Minor	Major
10	1	19	13 1/3°	x		x	
11	2	20	26 2/3°		x		x
12	3	21	40°	x			x
13	4	22	53 1/3°	x		x	
14	5	23	66 2/3°		x		x
15	6	24	80°	x			x
16	7	25	93 1/3°	x		x	
17	8	26	106 2/3°		x		x
18	9	27	120°	x			x
19	10	1	133 1/3°	x		x	
20	11	2	146 2/3°		x		x
21	12	3	160°	x			x
22	13	4	173 1/3°	x		x	
23	14	5	186 2/3°		x		x
24	15	6	200°	x			x
25	16	7	213 1/3°	x		x	
26	17	8	226 2/3°		x		x
27	18	9	240°	x			x
1	19	10	253 1/3°	x		x	
2	20	11	266 2/3°		x		x
3	21	12	280°	x			x
4	22	13	293 1/3°	x		x	
5	23	14	306 2/3°		x		x
6	24	15	320°	x			x
7	25	16	333 1/3°	x		x	
8	26	17	346 2/3°		x		x
9	27	18	360°	x			x

Similarly, in step **1** rotor arm **86** is at the 133⅓° position with a minor rotor and minor stator electromagnet in alignment for discharge. Simultaneously, rotor arm **87** is at the 253⅓° position with a minor rotor and minor stator in alignment for capacitor discharge therethrough. The other steps of the sequence are apparent from Table I, for each position of the three rotor arms at any step and the juxtapositions of respective stator and rotor electromagnet elements at that position.

In the simplified motor arrangement shown in schematic form in FIG. **18**, with single electromagnet configuration the alignment is uniform and the discharge sequences follow sequentially.

As hereinbefore mentioned, a change in speed is effected by displacing the stator spark gap terminals on the rotor (shown at **236** in FIGS. **17** and **18**) either counter-clockwise or clockwise 6⅔° so that the discharge position of the stator electromagnets is displaced 6⅔° either clockwise or counter-clockwise of a rotor electromagnet at the time of discharge. Referring to FIGS. **11** to **15**, the simultaneous discharge of selected capacitors into the electromagnets so displaced results in a deceleration at the time the rotor electromagnet is just approaching the associated stator, or an acceleration if the rotor electromagnet is just leaving its associated stator electromagnet. In each event, there is a repulsive reaction between the stator and rotor electromagnets, so that if the rotor is approaching the stator, there is a slow-down and if the rotor is receding from the stator there is a speed-up.

Referring to FIG. **11**, clutch mechanism **304** about shaft **111** is operated electromagnetically in conventional manner to displace the spark-gap mechanism **236** which is operated normally in appropriate matching alignment with the rotor spark-gap discs **291, 292,** 293. Clutch **304** has a fixed drive element **311** containing an electromagnetic drive coil (not shown) and a motor element **310** which, when the electromagnetic drive coil is energized, can be operated by a direct current. The operation of motor element **310** brings into operation spark gap elements **224*r***, 223*r* or **223*f***, 224*f* of the system shown in FIGS. **4, 5** and **8**, as illustrated in FIG. **19.**

The fixed stator coil spark gap terminal pairs **223,** 224 and **266, 267** are arrayed about a cylindrical frame **322** which is fabricated in insulative material. In the illustrative example of FIGS. **17** and **18**, there are nine such spark gap terminal pairs about the periphery of cylinder frame **324.** In the engine of FIGS. **4** to **8**, a total of 27 such spark gap pairs are involved. In addition, although not shown in the drawing, there are also pairs of terminals, such as **223*r*** or *f*, 224*r* or *f* and **266*r*** or *f*, **267*r*** or *f*, displaced 6⅔° on either side of the pairs **223, 224** or **266, 267** and all other pairs in the spark gap array, the letters *r* and *f* denoting "retard" and "faster." The latter displaced pairs of used in the control of speed of the engine rotor. The displaced pairs not shown are involved with the operation of cluth **304**, the speed changing control element.

Clutch **304** is associated with shaft **111** in that the movable element **310** draws clutch disc element **316** on shaft **111** away from clutch disc element **322** when energized by a voltage of appropriate polarity applied to its motor electromagnet **311.** Such clutch drives are well known in the art.

The clutch mechanism **304** of FIGS. **11** and **19**, when not energized is in the configuration shown in FIG. **11.** The energized configuration of clutch **304** is not specifically illustrated. Upon energization, spark-gap element **222** on disc **236** is displaced rightward, as viewed in FIG. **11**, by broken lines **236X**, into alignment with the positions of fixed spark-gap terminals **223*f***, 224*f* and **267*r***, **266*r***. When the disc is in position **236X**, the flattened edge **332** of pin **330** in disc **325** rides on surface **350** of disc **322.** Normally, the flattened edges **351** of pins **330** are engaged against the flat edge **352** in recess **331** of disc **322.** The displacement of disc **322** on shaft **111** is effected by the action of clutch **304** against spring **314** (FIG. **11**). An electric switch (not shown) of clutch mechanism **304** energizes it from a d-c power source, and has two positions, one for deceleration and one for acceleration. In either position, clutch **304** is engaged to pull clutch disc **322** from clutch disc **325**, momentarily. For the decelerate or the accelerate position, the displaced alignment of spark gap elements **222** is with the **224*f***, 223*f* and the **224*r***, 223*r* spark-gap terminal elements. However, only the **224*f***, 223*f* spark-gap elements are switched into operation with appropriate capacitors for the accelerate position, while in the decelerate position only the **223*r*** and **224*r*** spark-gap elements are switched into the circuit with their associated capacitors.

Of course, when insulative disc **236** is displaced by clutch **304**, its gap terminals **222, 225** and **228** (FIGS.

14 and 18) are all displaced into the alignment position of **236X** so as to engage the *r* and *f* lines of fixed spark gap elements. Although the accelerate and decelerate positions of disc **236** are the same, it is the switching into operation of the **223, 224** or **266, 267** exemplary *r* or *f* pairs of terminals which determine whether speed up or slow down action of the rotor will occur.

The momentary displacement of clutch disc **322** from clutch disc **325** results in rotation of disc **325** about disc **322** through an angle of 120°. The detent ball and spring mechanism **320, 321** in disc **325** positions itself between one detent dimple **328** and a succeeding one **328** at a position 120° away on disc **325.** As stated, flat **332** of pin **330** rides on surface **350** of disc **322,** and pin **330** leaves the pin-holding groove **331**/352 along ramp **333** in disc **322** during the momentary lifting of disc **322** by clutch **304.** Pin **330** falls back into the next groove **331** at a point 120° further on about disc **322.** Pin **330** falls into place in groove **331** on ramp **334.** Pins **330** are rotatable in their sockets **353,** so that for either clockwise or counterclockwise rotation, the flat **351** will engage the flat **352** by the particular ramp it encounters.

The deceleration or acceleration due to the action of clutch **304** thus occurs within a 120° interval of rotation of disc **325.** Disc **322** during this interval may only move a fraction of this arc.

There has been described hereinabove an electromotive engine system wherein at least one electromagnet is in a fixed position and a second electromagnet of similar configuration is juxtaposed with it in a magnetic polarity relationship such that, when the cores of the electromagnets are energized, the juxtaposed core faces repel one another. One core being fixed and the second core being free to move, any attachments to the second electromagnet core will move with it. Hence, if a plurality of fixed cores are positioned about a circular confining housing, and, within the housing, cores on a shaft are free to move, the shaft is rotationally urged each time the juxtaposed fixed and rotatable cores are in alignment and energized. Both the fixed and the movable cores are connected to spark gap terminal elements, and the associated other terminal elements of the spark gaps are connected to capacitors which are changed to a high voltage from pulsed unipolar signal generators. These capacitors are discharged through the electromagnets across the spark gaps. By switching selected groups of capacitors into selected pairs of spark gap elements displaced from the normal uniformly positioned elements for discharge through the electromagnets, the rotor of the circular array systems is accelerated and decelerated.

By confining a fixed electromagnet array in a linear configuration with a linearly movable electromagnet to which a working tool is attached, exciting the juxtaposed pairs of electromagnets by capacitor discharge results in the generation of linear force for such tools as punch presses or for discharging projectiles with considerable energy.

The inventor claims:

1. An electric engine comprings:

a housing;

an array of electromagnets uniformly spaced in said housing to form a stator;

a rotor cage on a shaft journaled in and rotatable within said housing and within said stator, said rotor cage having thereon a spaced array of electromagnets similar to said stator electromagnets and in number comprising an integral fraction of the number of electromagnets in said stator array;

each of the electromagnets of said stator and of said rotor having a magnetizable core of particular configuration and each being wound with a coil such that a pulse of unidirectional electric current through said coil magnetizes the respective core thereof to a particular magnetic polarity, and the faces of rotor cores juxtaposing selected stator cores are magnetized to the same polarity, the juxtaposed cores thereby tending to repel one another, one lead of each of the stator and rotor coils being connected to a common terminal, the other lead of each of said coils being connected to a gap terminal, the gap terminals of said rotor coils being on the rotor and equal in number to the number of coils thereon and matching the positions of said rotor electromagnets thereon, the gap terminals of said stator being equal in number to the number of coils on the stator and disposed uniformly about said stator to match the positions of said stator electromagnets within said housing;

a first array of capacitors each having a terminal in common with the common coil terminal of said stator electromagnets, and each capacitor having its other terminal connected to a gap terminal arrayed adjacent the gap terminal of an electromagnet associated therewith;

a second array of capacitors, each having a terminal in common with said common terminal of said rotor electromagnet coils but equal in number to the number of capacitors in said stator array, the other terminals of said capacitors in said second array being connected to gap terminals arrayed about said housing so as to be in axial alignment with said stator gap terminal positions and being alignable with said rotor gap terminals as said rotor is rotated in said housing and respective gap terminals of said rotor coils pass each second array capacitor gap terminals at a predetermined gap distance;

gap coupling terminals on said rotor equal in number to the number of rotor electromagnet coils and positioned to match the rotor electromagnet positions on said rotor, the gap coupling terminals being rotatable with said rotor so as to pass said adjacent stator coil and associated stator capacitor gap terminal at a predetermined distance therefrom;

a plurality of capacitor charging circuits connected respectively across each of said capacitors in both said first and said second arrays of capacitors for charging each of said capacitors to a predetermined high d-c potential;

a first source of unidirectional electric potential connected to each of said capacitor charging circuits for energizing said charging circuits; and

a second unidirectional electric potential source connected to said electromagnets of said rotor and said stator of such polarity as to receive a charge from the inverse inductive discharge of the electromagnet coils as their fields collapse following the discharge of each capacitor through a rotor or stator electromagnet coil,

whereby whenever a rotor electromagnet is aligned opposite a stator electromagnet, the rotor coil gap terminal of that electromagnet is opposite an associated second capacitor array gap terminal, and a gap coupling terminal of said rotor is aligned opposite the stator electromagnet coil gap terminal and associated first capacitor gap terminal, the capacitors discharge the charge thereon across the gaps through their associated electromagnet coils to magnetize their respective juxtaposed electromagnet cores to cause them to repel one another, thus aligning a succeeding pair of rotor and stator electromagnets for capacitor discharge across their respective gaps, to cause them to repel one another, alignments rotor rotation within the housing continuously bringing successive rotor-stator electromagnets into alignment for discharge of the capacitors therethrough to produce rotary motion of the rotor on said rotor shaft continuously so long as energy from said first source of unidirectional electric potential is applied to said charging circuits to recharge said capacitors after each discharge.

2. In an electric engine having a rotor comprising electromagnetic coil means rotatable within a stator comprising similar electromagnetic coil means, said electromagnetic coil means being polarized for magnetic repulsion;

capacitor means electrically coupled across successive spark gaps to selected ones of said coils of said stator and all of the coils of said rotor;

charging means connected to said capacitor means for charging said capacitor means to an electrical charge potential sufficient to cause arcing across said spark gaps to result in the discharge of said capacitor means through the electromagnetic coils so that the electromagnetic coil means repel one another; and

a unidirectional electric power source connected to said charging means to energize said charging means to continue charging said capacitor means following each discharge whereby the rotor of said engine is maintained in rotation by the successive discharges of said capacitor means across successive spark gaps into said electromagnetic coil means.

3. An electric engine according to claim **2,** wherein:

the charging means includes electronic square core oscillators connected to said unidirectional electric power source and includes step-up means and a rectifier to produce a substantial voltage step up from the voltage of said power source.

4. An electric engine according to claim **2,** wherein:

the charging means includes a vibrator connected to said power source, and step up transformer and rectifier means to provide a high voltage for charging said capacitor means.

5. A motive force producing means comprising:

at least a first electromagnet means including at least one coil wound about a core,

at least a second electromagnet means including at least one coil wound about a core similar to said first core,

the respective cores being positioned adjacent to one another so that the magnetic polarities of the adjacent core surfaces are the same when a unidirectional electric current is passed through the coils,

at least one capacitor means having one terminal thereof connected to one terminal of both of said electromagnet coils,

the other terminal of said capacitor means being connected to one terminal of a spark gap means, the other terminals of the coils of both said first and said second electromagnet means being connected to the other terminal of said spark gap means,

at least one unidirectional pulse charging means connected to said capacitor means to charge said capacitor means to a relatively high potential sufficient to arc across said spark gap means at predetermined spacing of said gap terminals, and

a source of unidirectional potential connected to said charging circuit to energize said charging means,

whereby upon application of current from said potential source to said charging means the successive pulses generated thereby charge said capacitor means to a voltage level sufficient to arc across said spark gap means to produce a discharge path for said capacitor means through said coils to cause said electromagnet means to repel one another with a substantial force.

6. A motive force producing means according to claim **5,** wherein:

said first electromagnet means is secured in a relatively stable housing, and said second electromagnet means is connected with and freely movable relative to said stable housing, and has utilization means connected thereto for performing work therewith when said capacitor means discharges through said coils of said electromagnet means.

7. A motive force producing means according to claim **6,** wherein said utilization means is a motor rotor coupled with said second electromagnet means and said first electromagnet means is a stator.

8. A motive force producing means according to claim **6,** wherein said utilization means is a piston attached to said second electromagnet means and is movable therewith to produce hammer-like blows when said capacitor means discharges through said electromagnet means.

9. In an electromotive force generating system as disclosed, means for accelerating or decelerating the motion of a force generating system, said means comprising:

at least two juxtaposed electromagnetic core elements, one fixed and one movable, including coils wound thereabout to provide a repulsion tendency when said cores are energized,

spark gap termimals connected with said coils,

capacitor means connected with said spark gap terminals to discharge across said spark gap terminals through said coils when a charge of sufficient voltage level appears across said capacitor means, thus to energize said juxtaposed electromagnets to induce said juxtaposed electromagnet cores to repel one another,

charging means connected to said capacitors for charging them to said sufficient voltage level, and

selective positioning means coupled with said spark gap terminals and with at least said movable electromagnet core to cause selective displacement of said movable core with respect to said fixed core to accelerate or decelerate the motion of the movable core relative to the fixed core.

10. An electromotive force generating system according to claim **9,** wherein:

said juxtaposed electromagnetic cores include a plurality of fixed cores and a smaller number of movable cores, said smaller number being an integral fraction of the number of fixed cores, and

said selective positioning means is an electromagnetic clutch coupled with said smaller number of movable cores for movement therewith, and includes selective displacement means coupled with said spark gap terminals connected with selected capacitors in said capacitor means and selected combinations of coils in said plurality of fixed electromagnets.

11. The method of generating motive power comprising the steps of:

a. positioning similar electromagnets in juxtaposed relationship with their respective cores arranged for repulsion when said electromagnets are energized,

b. charging capacitors to a relatively high potential, and

c. discharging said capacitors simultaneously through said electromagnets across spark gaps set to break down at said relatively high potential,

thereby to cause said similar electromagnets to repel one another with considerable force.

12. The method of generating motive power defined in claim **11,** wherein in said positioning step at least one of said electromagnets is maintained in a fixed position and another electromagnet is free to move relative to said fixed electromagnet.

13. The method of generating motive power according to claim **11,** wherein:

the charging step includes the charging of capacitors to a relatively high potential from a pulsed unipolar source of electric energy.

14. In an electromagnetic capacitor discharge engine including movable electromagnets and fixed electromagnets, said movable electromagnets being movable into polar alignment with said fixed electromagnets, capacitor means, means for charging said capacitor means, and means for discharging said charged capacitor means through said fixed and movable electromagnets to polarize aligned fixed and polarized electromagnets for magnetic repulsion, an acceleration and deceleration control means comprising:

first selective means for momentarily delaying the discharge of the capacitors until the movable electromagnets in said engine have begun to recede from the fixed electromagnets, in order to accelerate the motion of said movable electromagnets by the added impetus of the repulsion, and

second selective means for momentarily accelerating the discharge of the capacitors to occur at a point in the motion of the movable electromagnets where said movable electromagnets are approaching said fixed electromagnets to decelerate the motion of said movable electromagnet by the tendency to repel the approaching electromagnet by the fixed electromagnet.

15. An electric engine, comprising:

fixed electromagnets;

movable electromagnets movable into alignment with said fixed electromagnets;

capacitor means;

means for charging said capacitor means; and

means for discharging said charged capacitor means through said fixed and movable electromagnets to polarize said aligned fixed and movable electromagnets for magnetic repulsion.

16. An electric engine as recited in claim **15,** wherein: said means for discharging said charged capacitor means comprises voltage breakdown switch means.

17. An electric engine as recited in claim **16,** wherein:

said voltage breakdown switch means includes at least one terminal movable with at least one of said movable electromagnets for breaking down when said at least one of said movable electromagnets is in alignment with a said fixed electromagnet.

18. An electric engine as recited in claim **17,** wherein:

said voltage breakdown switch means comprises a spark gap means.

* * * * *

United States Patent [19]
Whiteley

[11] **3,723,891**
[45] **Mar. 27, 1973**

[54] FREQUENCY TO VOLTAGE CONVERTER

[75] Inventor: **Thomas B. Whiteley**, Riverside, Calif.

[73] Assignee: **The United States of America as represented by the Secretary of the Navy**

[22] Filed: **May 12, 1971**

[21] Appl. No.: **143,678**

[52] **U.S. Cl.** **329/103,** 307/233, 307/246, 328/67, 328/151, 329/126
[51] **Int. Cl.** **H03d 3/04**
[58] **Field of Search** 329/103, 1, 2, 126; 307/246, 307/233; 328/67, 151

[56] **References Cited**

UNITED STATES PATENTS

3,582,799	6/1971	Reid	307/233 X
2,834,883	5/1958	Lukoff	328/151
2,978,615	4/1961	Chater	307/233 X
3,363,113	1/1968	Bedingfield	328/151 X
3,371,291	2/1968	Forrest et al.	307/233 X
3,593,171	7/1971	Gassmann	307/233 X

Primary Examiner—Alfred L. Brody
Attorney—R. S. Sciascia, G. J. Rubens, J. W. McLaren and T. M. Phillips

[57] **ABSTRACT**

A frequency to voltage converter including a switching circuit responsive to an input signal at each zero crossing in the positive direction to charge a capacitor to a peak negative voltage and discharging through a resistor towards a positive voltage until the next zero crossing in the positive direction. The peak voltage charge is coupled to a holding circuit to provide a D.C. voltage proportional to the frequency of the input signal.

8 Claims, 3 Drawing Figures

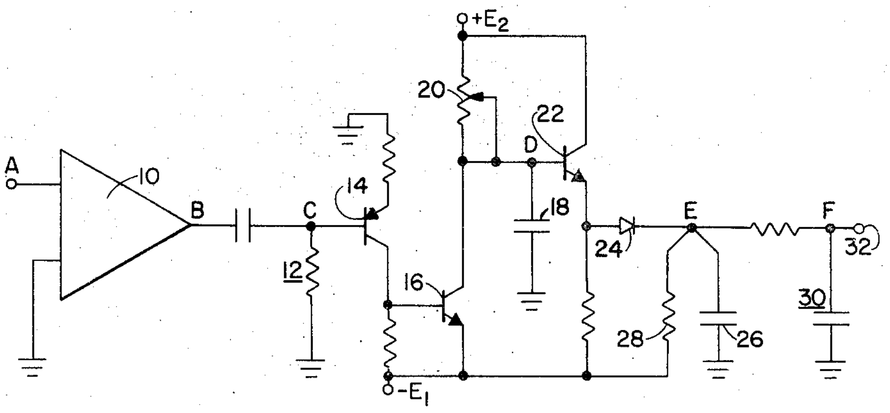

Patented March 27, 1973 **3,723,891**

2 Sheets-Sheet 1

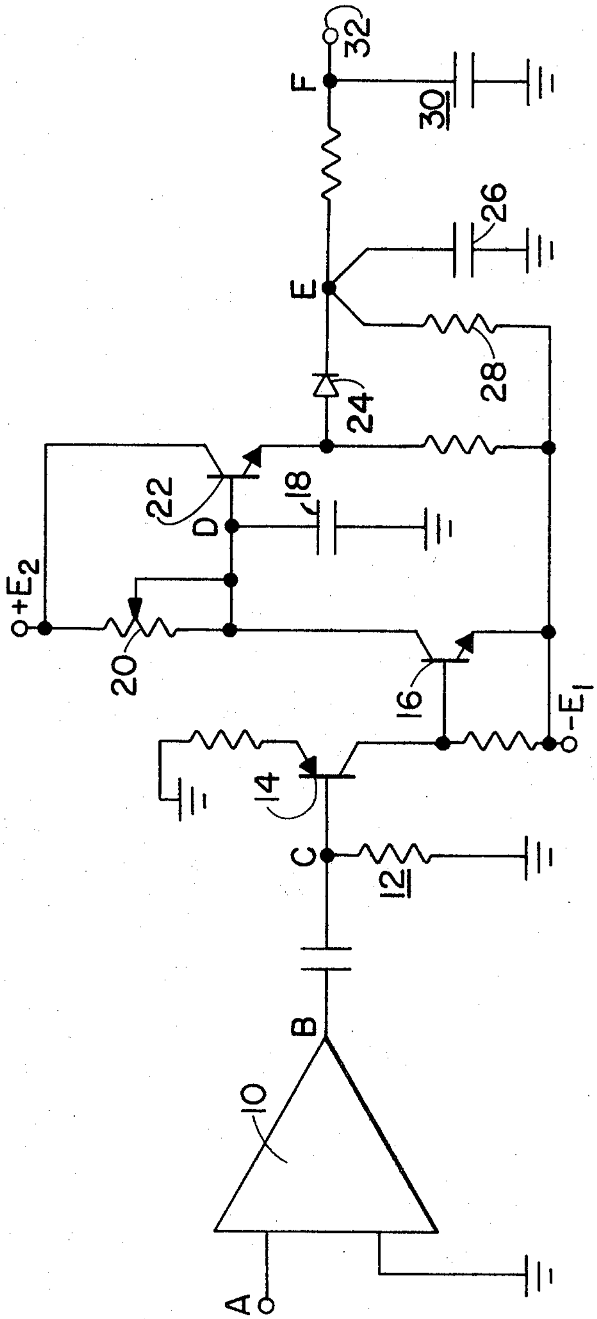
32
F
30
26
E
28
24
22
18
D
+E2
20
16
-E1
14
C
12
B
10
A

FIG. 1.

INVENTOR.
THOMAS B. WHITELEY
BY G. M. Phillips
George J. Rubens

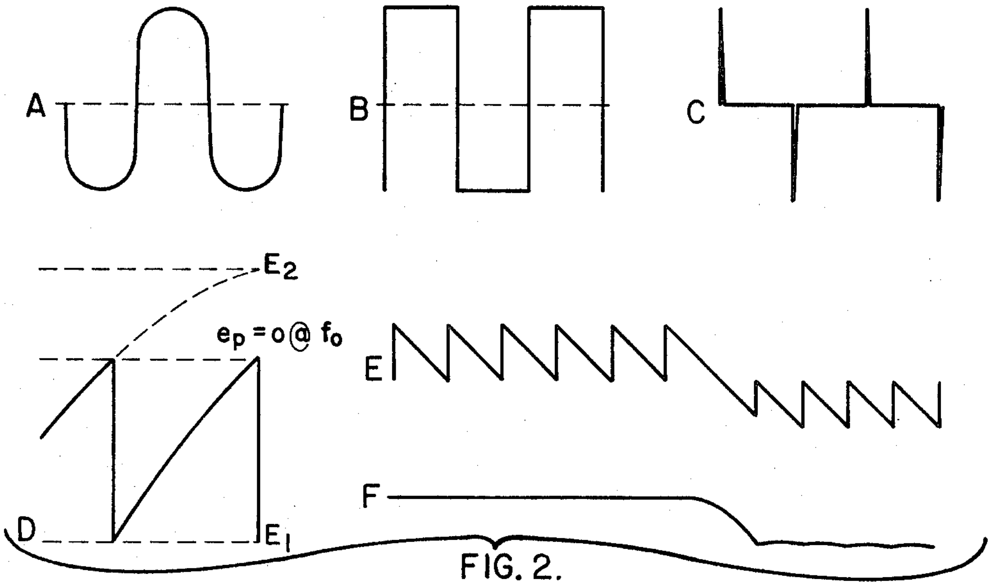
A
B
C
E2
$e_p = 0 @ f_0$
E
D
E1
F

FIG. 2.

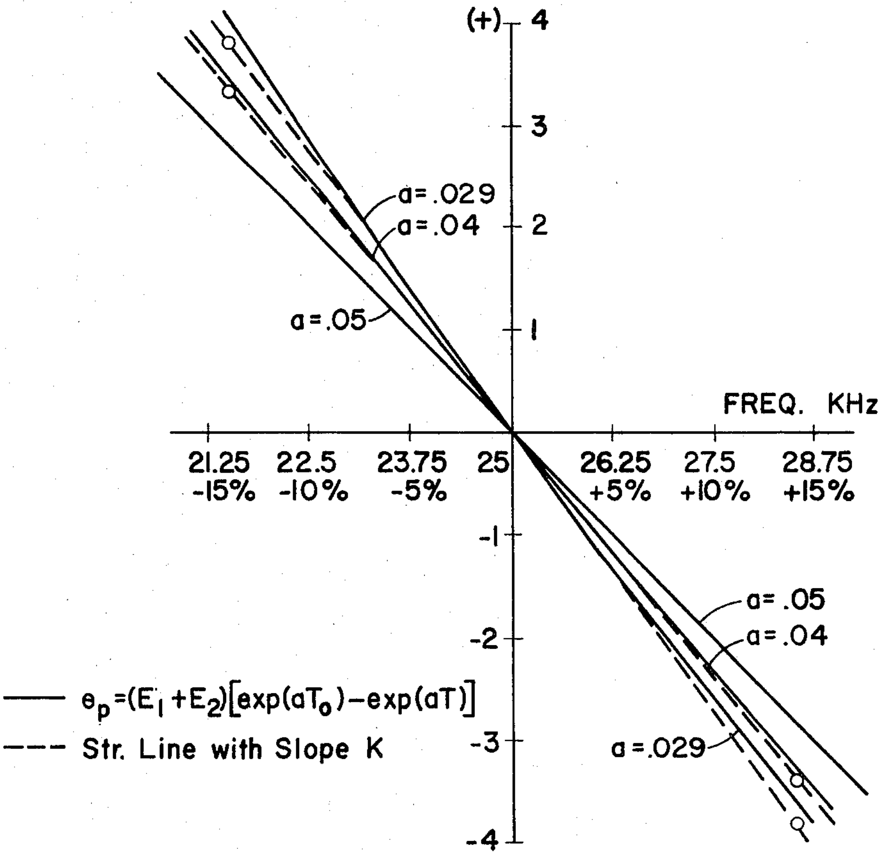

FIG. 3.

INVENTOR.
THOMAS B. WHITELEY
BY G. M. Phillips
George J. Rubens

FREQUENCY TO VOLTAGE CONVERTER

STATEMENT OF GOVERNMENT INTEREST

The invention described herein may be manufactured and used by or for the Government of the United States of America for governmental purposes without the payment of any royalties thereon or therefor.

BACKGROUND OF THE INVENTION

The invention relates to frequency to voltage conversion systems and more particularly to the concept of providing an exponential charging curve beginning at each crossover in the positive direction of a sinusoidal input signal.

Various methods for achieving an indication of the frequency of an alternating current signal have previously been devised. Such prior art circuits are not suitable where precise performance over a large dynamic range and time lag is important.

SUMMARY OF THE INVENTION

In the preferred form of the invention, an input signal (approximately sinusoidal) is shaped and differentiated by a very short time constant R.C. network. A switching circuit responsive to the negative pulse corresponding to the positive-going zero-crossing of the input signal, charges a capacitor to a peak negative voltage and is then discharged through a resistor toward a positive voltage. The peaks of the discharge waveform are sampled and then held constant until the next peak.

Other objects and many of the attendant advantages of this invention will be readily appreciated as the same becomes better understood by reference to the following detailed description when considered in connection with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a schematic circuit diagram of one embodiment of the invention;

FIG. 2 is a representation of voltages in various parts of the circuit of FIG. 1; and

FIG. 3 is a graph showing the effect of variance from the computed values for best linearity.

DESCRIPTION OF THE PREFERRED EMBODIMENT

It is well known that a linear ramp function generated at a zero crossover of an input sine wave, and allowed to rise until the next crossover, has a peak amplitude that is a measure of the length of time between crossovers. Instead of a linear ramp, the present invention utilizes an exponential curve, generated at successive crossovers of the input signal.

To obtain a voltage proportional to the frequency, f, which is the inverse of the period, T, the expression

$$e_1 = K\,(1/T_O - 1/T)$$

represents a voltage proportional to the variation of $f = 1/T$ about the center frequency $fo = 1/T_O$. K is the slope constant relating voltage and frequency, e.g. V/MHz.

A capacitor discharging through a resistance, R, from an initial voltage, $-E_1$, toward a supply voltage, $+E_2$, has an exponential discharge voltage represented by

$$e = E - (E_1 + E_2)\; exp\,(^{-at})$$

where $a = 1/Rc$.

Then at $t = T$,

$$ep = E_2 - (E_1 + E_2)\; exp\,(^{-aT})$$

where e_p is the capacitor voltage reached after discharging for a time period T.

Considered as functions of a variable T, e is a hyperbolic curve, and e_p is an exponential curve. By optimum choice of parameters, the two curves can be made to very nearly coincide over a considerable range near $T = T_O$.

The peak voltage e_p then is an instantaneous measure of the inverse period $1/T$, i.e., the frequency of the input signal.

It is obvious that $e_1 = O$ at $T = T_O$. In order that $e_p = 0$ at $T = T_O$,

$$E_2 - (E_1 + E_2)\; exp\,(^{-aT_O}) = 0$$

$$E_2 = (E_1 + E_2)\; exp\,(^{-aT_O})$$

For best coincidence of the two curves the first and second derivatives with respect to T are equated at $T = T_0$

$$e_1' = K/T^2$$

$$e_1'' = -2K/T^3 = (-2/T)\, e_1'$$

$$e_p' = a\,(E_1 + E_2)\; exp\,(-aT)$$

$$e_p'' = {}^{-ae}p'$$

From $e_1' = e_p =$,

$$K = aT_O^2(E_1 + e_2)\; exp\,(-aT_O) \text{ and}$$

from the second derivative,

$$a = 2/T_O$$

In an example for a discriminator with 25 KHz center frequency

$$To = 40\mu s$$

$$a = 2/40 = 0.050$$

$$RC = 20\mu s$$

$$exp\,(-aT_O) = 0.13534$$

$$\text{Assume } (E_1 + E_2) = 75V$$

$$E_2 = 0.1353 \times 75 = 10.1V$$

$$E_1 = 64.9V$$

The table below and the plotted voltage vs. frequency graph (FIG. 3) illustrate the effect of variance from the computed values for best linearity.

TABLE I

a	exp(−aT$_O$)	E$_1$	E$_2$	K	Frequency range linearity within 2%
0.05	0.13534	65V	10V	812	±20%
0.04	0.20190	60V	15V	969	±10%
0.029	0.31187	51.5V	23.5V	1090	± 6%

The circuit of this invention is illustrated in FIG. 1 with the waveform at various points shown in FIG. 2.

Input signal, A, is applied to high gain limiting amplifier **10** which produces a square wave output, B. Square wave B, is differentiated by short time constant RC network **12** to produce waveform C. Limiter **10** and differentiator **12** should produce pulses (waveform C) of sufficient amplitude to trigger transistor **14**, having a width not greater than one percent of T_0 and constant in amplitude and width over the expected amplitude range of input waveform, A. Transistor **14** is a PNP type which conducts only on negative pulses from differentiator **12**. The negative pulses of waveform C correspond to the positive going zero crossings of the input signal. Transistor **16** is a NPN type having high breakdown voltage and current capability, conducts heavily when transistor **14** is conducting. Capacitor **18** is charged rapidly to voltage $-E_1$, and then released and allowed to discharge through resistor **20** towards $+E_2$ as shown by waveform D. The discharge of capacitor **18** is arrested by the next charging pulse, and the sequence repeats. Resistor **20** is provided with a trim adjustment so that the operation of the circuit can be adjusted precisely to zero at center frequency.

The peaks of waveform D are coupled by emitter follower **22** and silicon diode **24** to holding capacitor **26**. The sawtooth voltage dropoff between peaks (waveform E) is caused by discharging of capacitor **26** through resistor **28**. The dropoff rate should be selected to allow the output to follow the maximum expected rate of change of frequency. RC filter **30** coupled between holding capacitor **26** and output terminal **32** removes most of the carrier frequency sawtooth from the output (waveform F).

Obviously many modifications and variations of the present invention are possible in the light of the above teachings. It is therefore to be understood that within the scope of the appended claims the invention may be practiced otherwise than as specifically described.

What is claimed is:

1. A frequency to voltage converter for producing a direct current voltage having a magnitude proportional to the frequency of an alternating input signal comprising:
   a. an input terminal for receiving an input signal the frequency of which is to be measured;
   b. a capacitor;
   c. first means coupled to said capacitor for charging said capacitor to a known voltage;
   d. discharge circuit means coupled to said capacitor and to said input terminal for discharging said capacitor to a voltage that is an instantaneous measure of the frequency of the input signal; and
   e. sampling and holding circuit means coupled to said capacitor for maintaining an output voltage proportional to the frequency of said input signal.

2. The frequency to voltage converter of claim **1** wherein said first means includes a circuit responsive to said input signal passing through zero in the positive direction to stop the discharge of said capacitor and start the charging of said capacitor.

3. The frequency to voltage converter of claim **1** wherein said capacitor charging includes a negative voltage source so that said capacitor is charged to a known negative voltage.

4. The frequency to voltage converter of claim **1** wherein said discharge circuit means includes a positive voltage source so that the capacitor is discharged towards said positive voltage source.

5. The frequency to voltage converter of claim **4** wherein said sampling and holding circuit samples and holds the positive peak voltage to which said capacitor is discharged.

6. The frequency to voltage converter of claim **1** said first means includes:
   a. a negative voltage source;
   b. switch circuit means responsive to said input signal passing through zero in the positive direction for charging said capacitor to the magnitude of said negative voltage source.

7. The converter of claim **6** wherein said switch circuit means includes:
   a. a differentiator for generating negative pulses when said input signal passes through zero in a positive direction;
   b. a first transistor coupled to said differentiator and being responsive to said negative pulse for generating a trigger pulse;
   c. a second transistor having a high breakdown voltage and a high current capability in circuit with said first transistor, said negative voltage source and said capacitor for connecting said capacitor to said negative voltage source in response to said trigger pulse.

8. The converter of claim **4** wherein said discharge circuit means includes:
   a. a positive voltage source;
   b. an adjustable resistor connecting said capacitor to said positive voltage source so that said capacitor will discharge toward said positive voltage source.

* * * * *

June 4, 1968 P. T. FARNSWORTH **3,386,883**

METHOD AND APPARATUS FOR PRODUCING NUCLEAR-FUSION REACTIONS

Filed May 13, 1966 16 Sheets-Sheet 1

FIG. 1

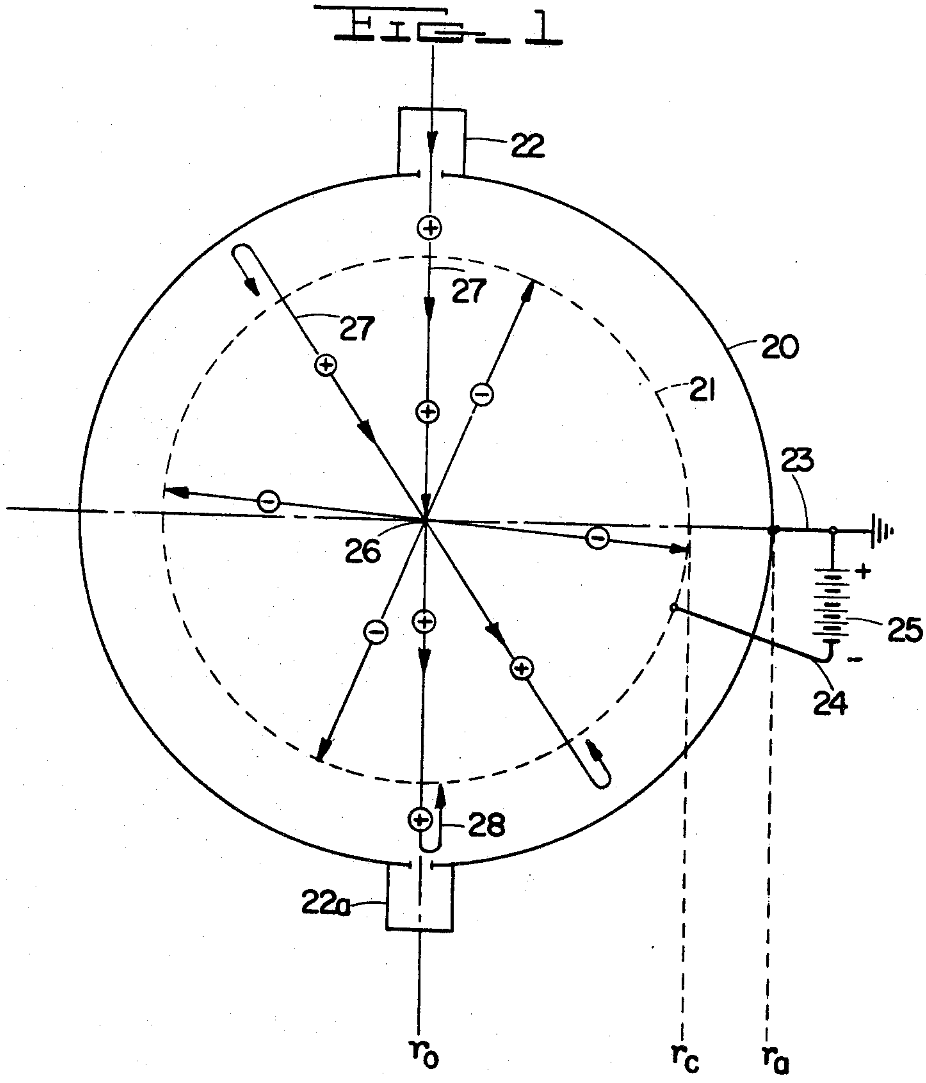
22
+
27
27
20
+
-
21
+
23
-
26
-
+
-
25
24
+
-
+
28
22a
r_o
r_c
r_a

INVENTOR
PHILO T. FARNSWORTH
BY Wood, Gust & Irish

ATTORNEYS

FIG. 2

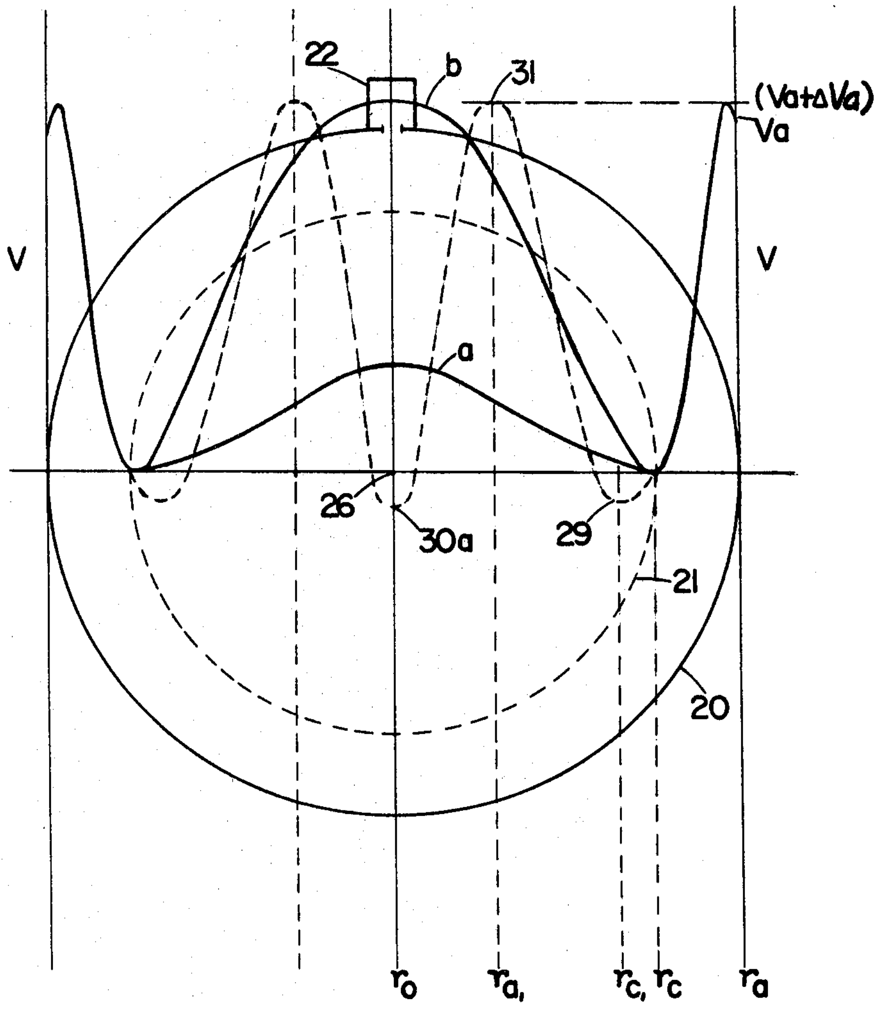
22
b
31
(Va+ΔVa)
Va
V
V
a
26
30a
29
21
20
r₀
rₐ₁
r_c₁
r_c
rₐ

INVENTOR
PHILO T. FARNSWORTH

BY Hood, Gust & Irish

ATTORNEYS

FIG. 3

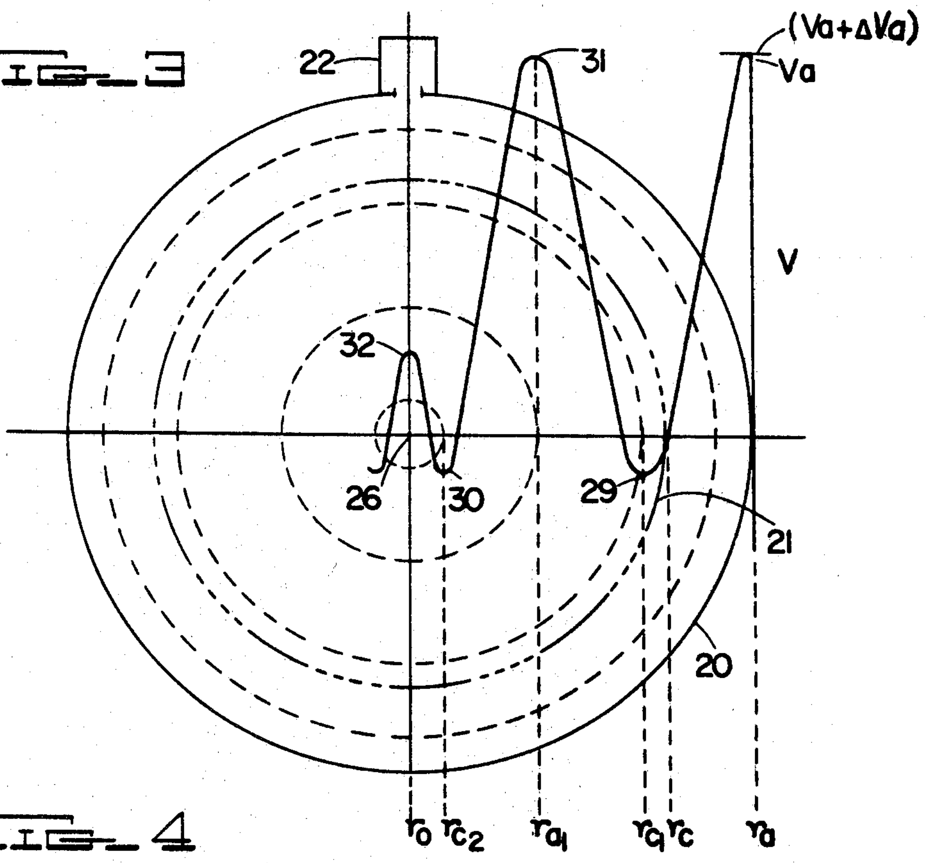
(Va+ΔVa)
Va
22
31
V
32
26
30
29
21
20
r_0
r_{c_2}
r_{a_1}
r_{c_1}
r_c
r_a

FIG. 4

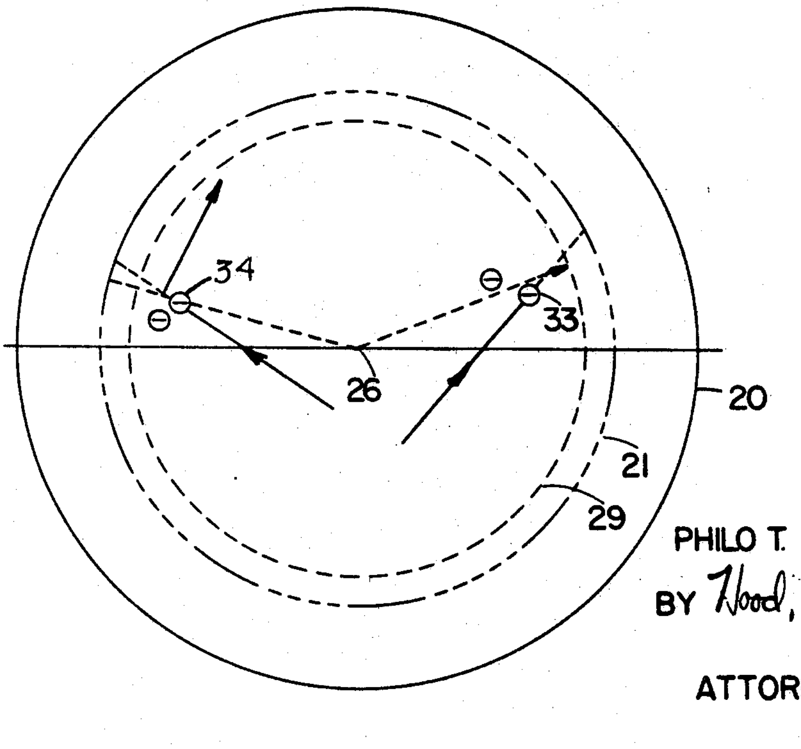
34
⊖
⊖
33
26
20
21
29

INVENTOR
PHILO T. FARNSWORTH

BY Hood, Gust & Irish

ATTORNEYS

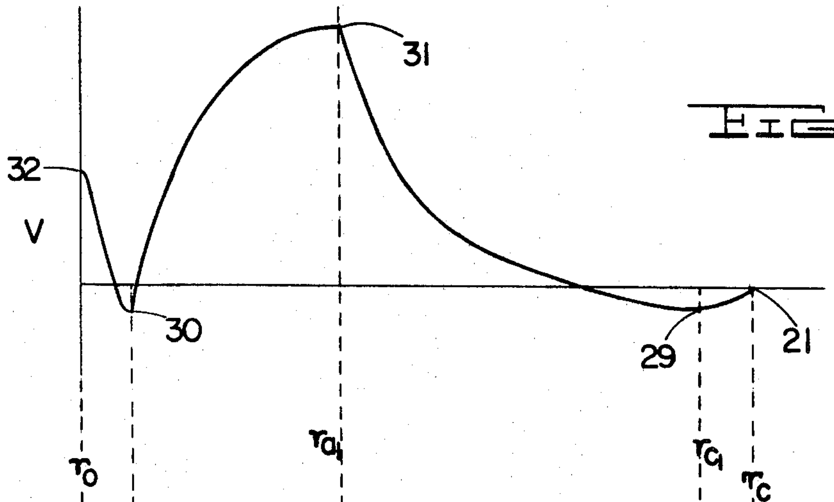

FIG. 5a

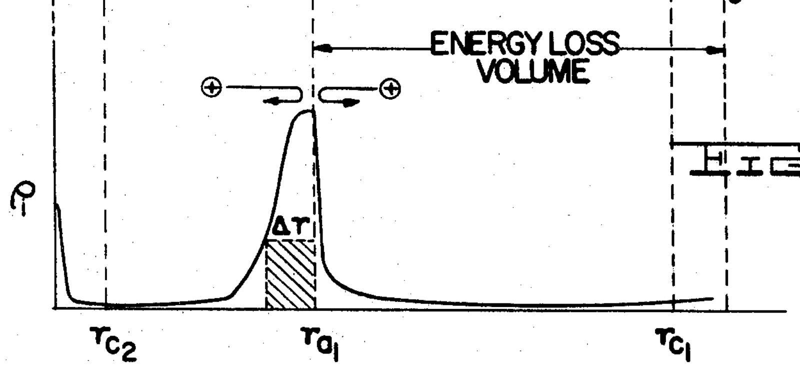

FIG. 5b

FIG. 10

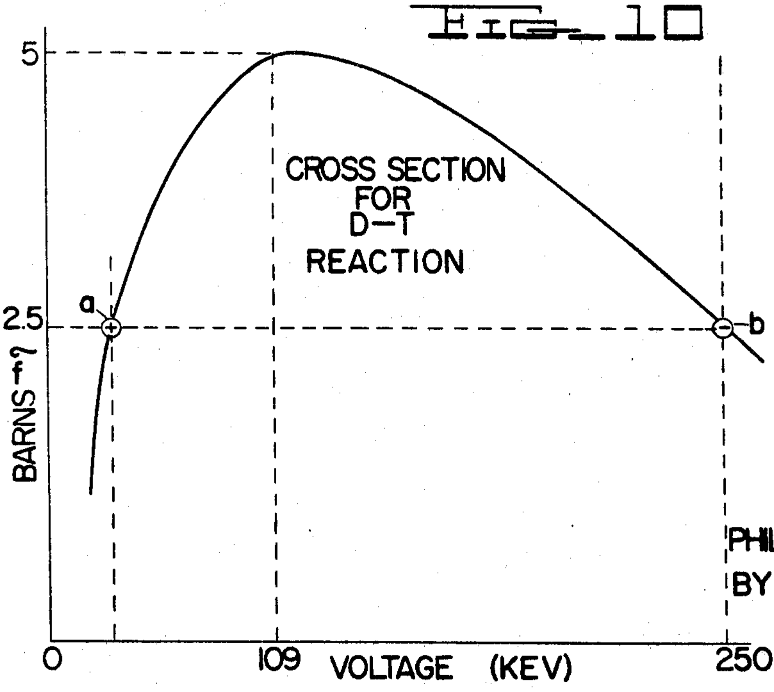

INVENTOR
PHILO T. FARNSWORTH
BY Hood, Gust & Irish
ATTORNEYS

June 4, 1968 P. T. FARNSWORTH **3,386,883**

METHOD AND APPARATUS FOR PRODUCING NUCLEAR-FUSION REACTIONS

Filed May 13, 1966 16 Sheets-Sheet 5

FIG. 6

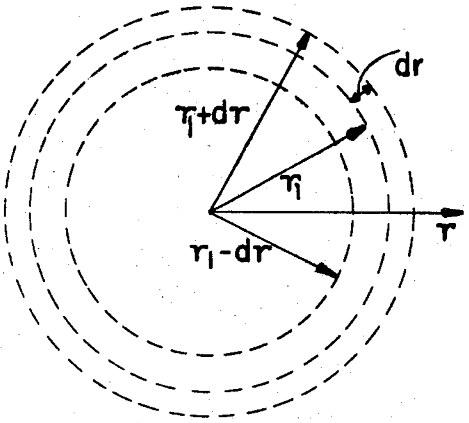

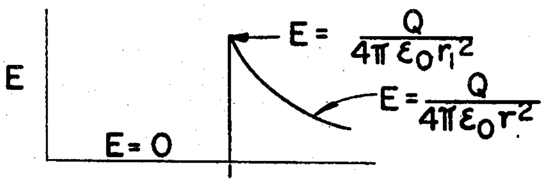

6(a)

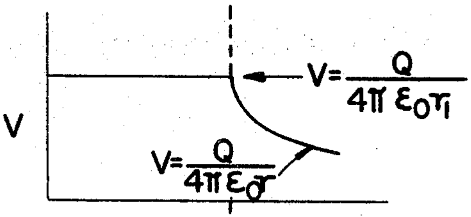

6(b)

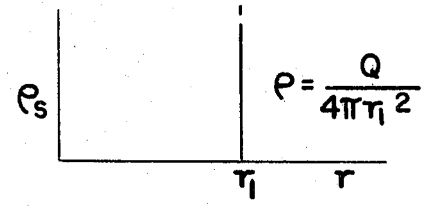

6(c)

INVENTOR
PHILO T. FARNSWORTH
BY Hood, Gust & Irish
ATTORNEYS

FIG. 7

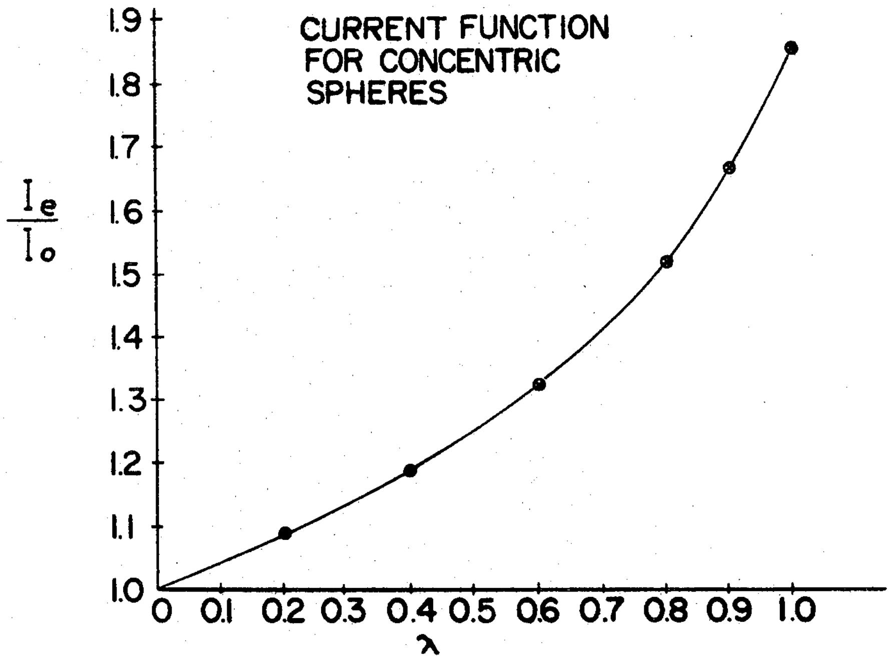

INVENTOR
PHILO T. FARNSWORTH
BY Hood Gust & Irish

ATTORNEYS

FIG. 8

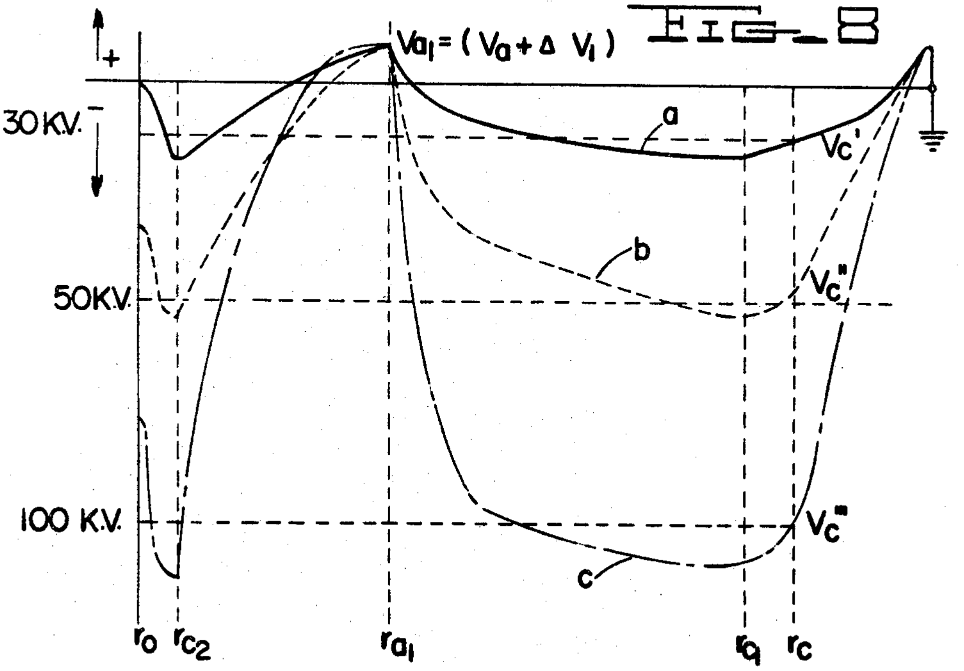

FIG. 9

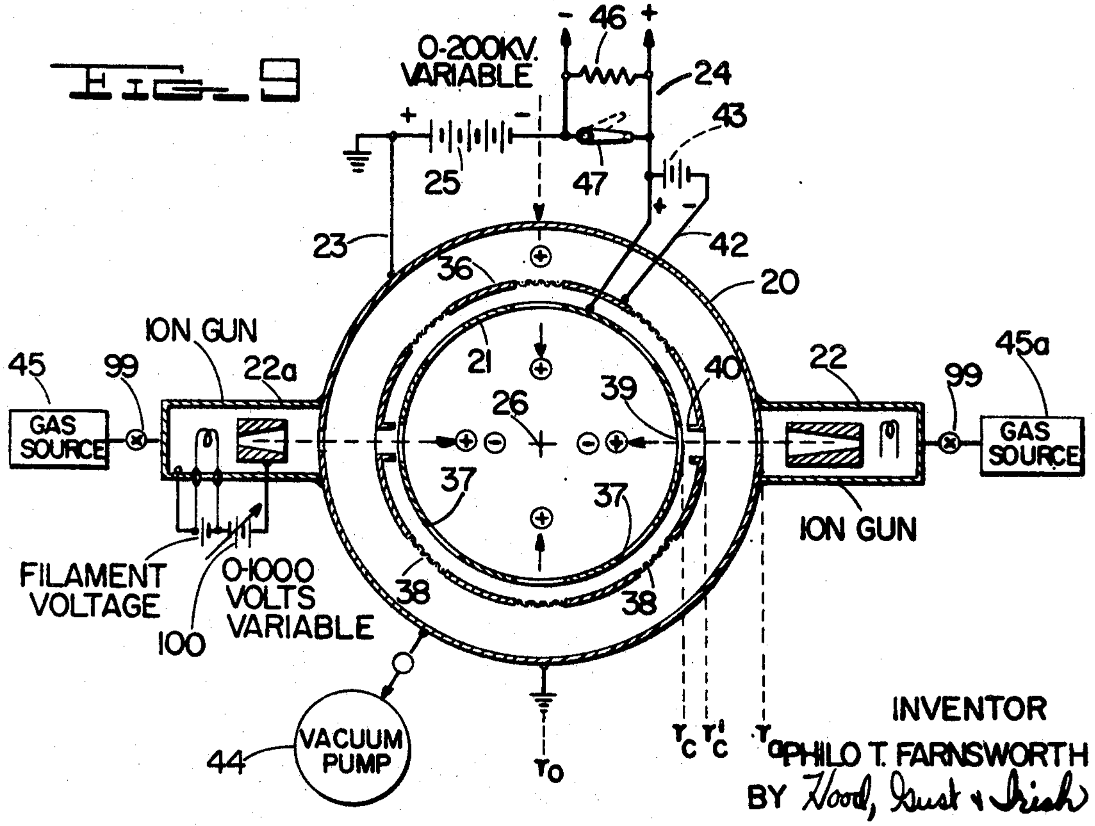
0-200KV. VARIABLE
− 46 +
24
43
+
−
25
47
+ −
23
42
20
36
ION GUN
45
99
22a
21
26
39
40
22
45a
99
GAS SOURCE
GAS SOURCE
37
37
ION GUN
FILAMENT VOLTAGE
0-1000 VOLTS VARIABLE
100
38
38
44
VACUUM PUMP
r0
rc
r'c
ra

INVENTOR
PHILO T. FARNSWORTH
BY Wood, Guest & Irish
ATTORNEYS

June 4, 1968 P. T. FARNSWORTH 3,386,883

METHOD AND APPARATUS FOR PRODUCING NUCLEAR-FUSION REACTIONS

Filed May 13, 1966 16 Sheets-Sheet 8

FIG. 11

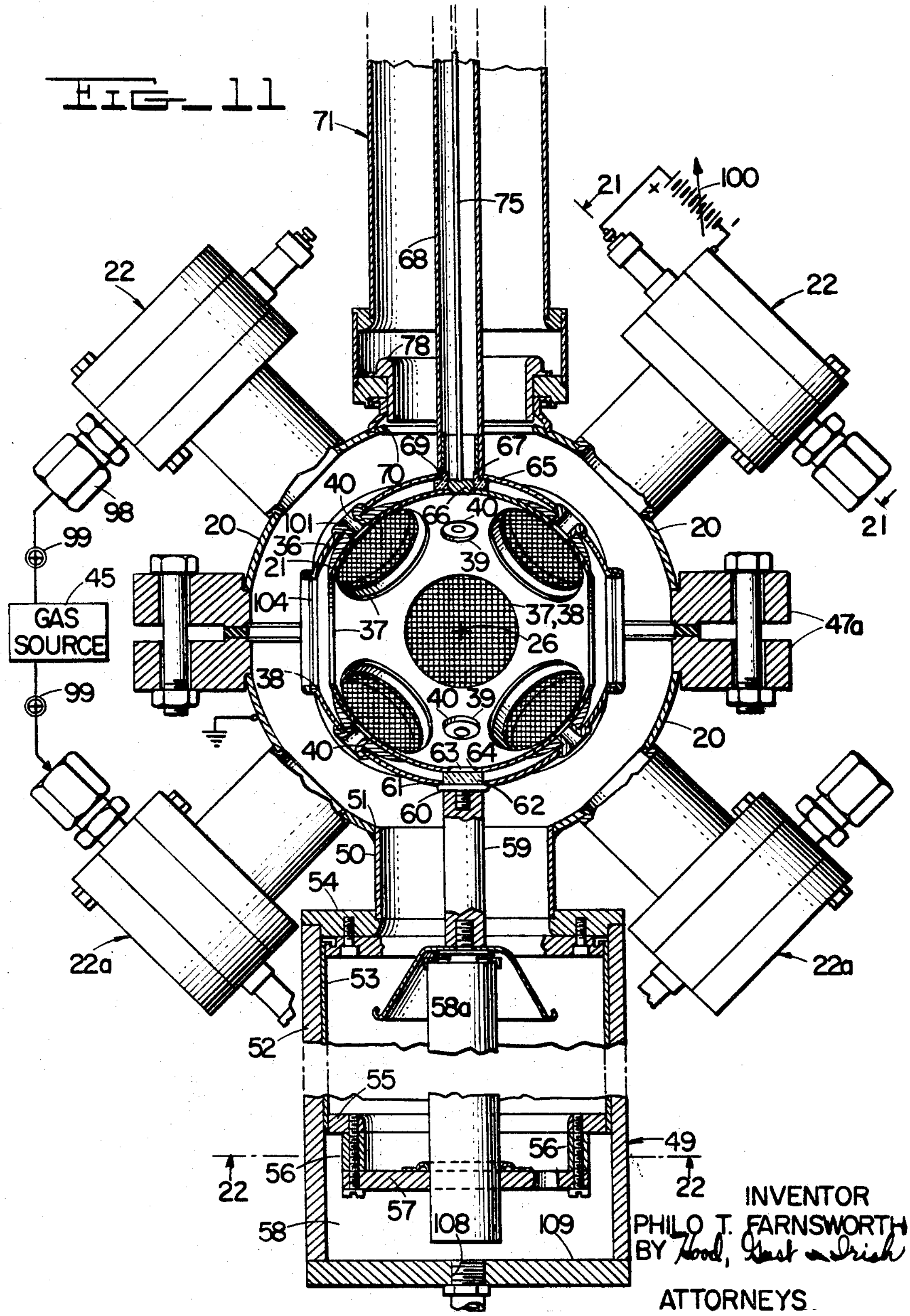

71
75
68
78
21
100
22
22
98
99
45
GAS SOURCE
99
69
67
65
70
40
66
40
20
101
36
21
39
104
37,38
26
37
38
40
39
40
63
64
61
62
60
51
50
59
54
47a
20
20
21
22a
53
58a
52
22a
55
56
56
49
57
108
109
58
22
22

INVENTOR
PHILO T. FARNSWORTH
BY Hood, Gust & Irish
ATTORNEYS

June 4, 1968 P. T. FARNSWORTH **3,386,883**

METHOD AND APPARATUS FOR PRODUCING NUCLEAR-FUSION REACTIONS

Filed May 13, 1966 16 Sheets-Sheet 9

FIG. 12

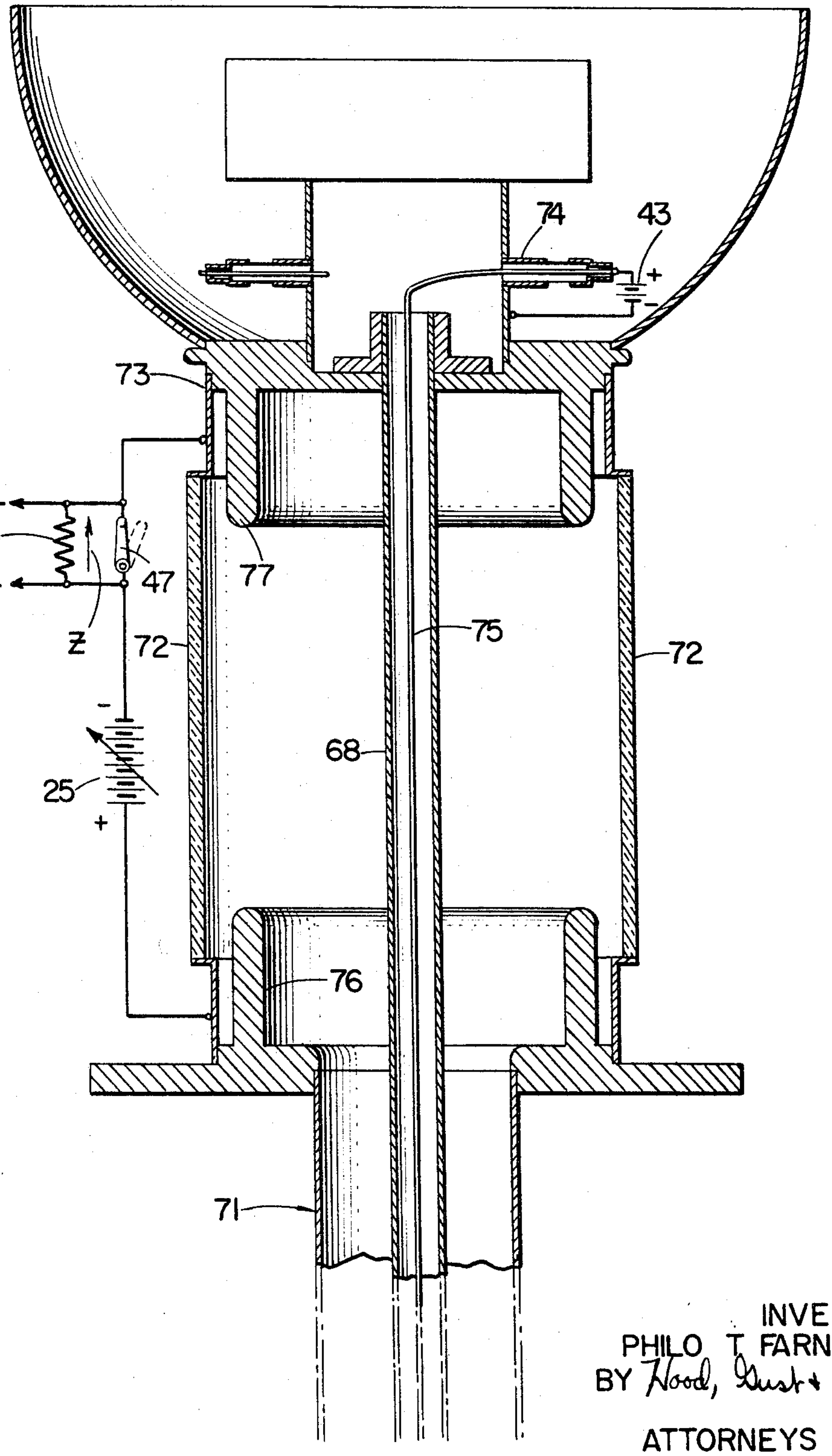
74
43
+
-
73
-
46
+
47
77
Z
72
75
72
68
-
25
+
76
71

INVENTOR
PHILO T. FARNSWORTH
BY Wood, Gust & Irish
ATTORNEYS

June 4, 1968 P. T. FARNSWORTH 3,386,883

METHOD AND APPARATUS FOR PRODUCING NUCLEAR-FUSION REACTIONS

Filed May 13, 1966 16 Sheets-Sheet 10

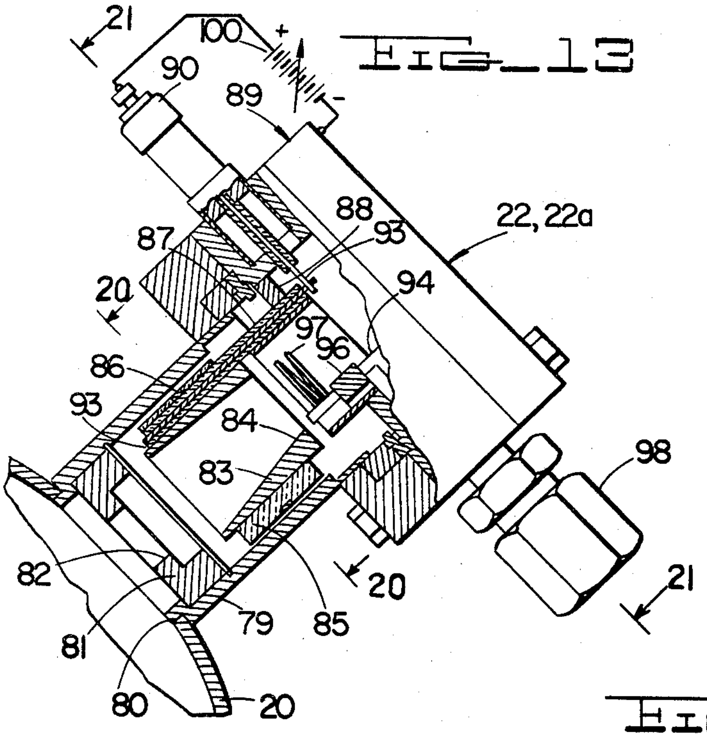
FIG. 13

21
100
+
-
90
89
88
93
22,22a
87
94
20
97
96
86
93
84
83
98
82
20
21
79
85
81
80
20

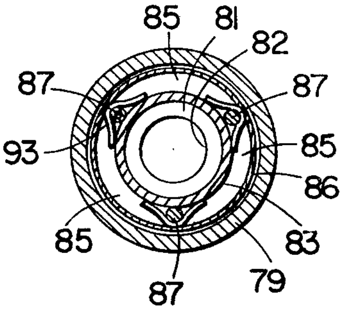
FIG. 20

85
81
82
87
87
93
85
86
85
83
87
79

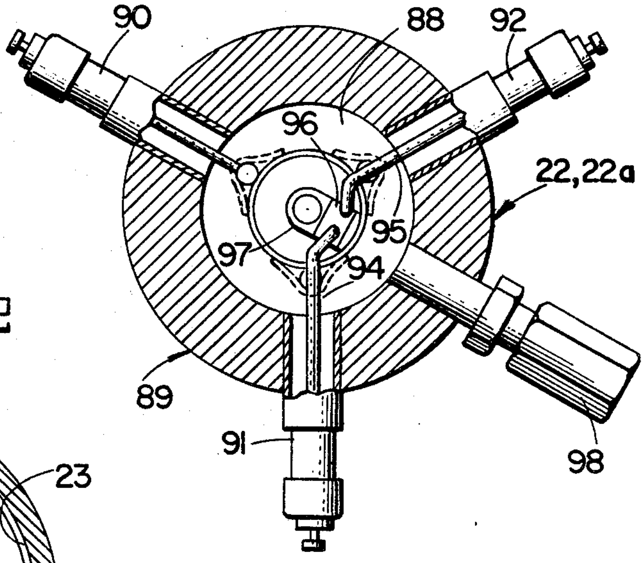
FIG. 21

90
88
92
96
22,22a
95
97
94
89
91
98

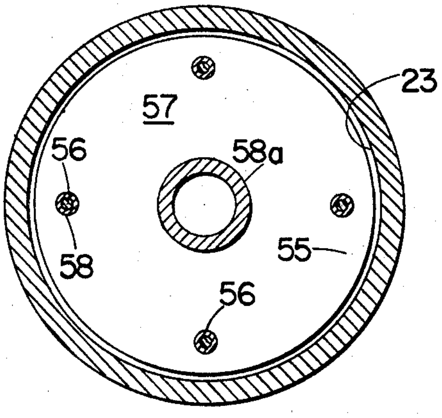
FIG. 22

23
57
56
58a
58
55
56

INVENTOR
PHILO T. FARNSWORTH
BY Wood, Gust & Irish
ATTORNEYS

FIG. 14

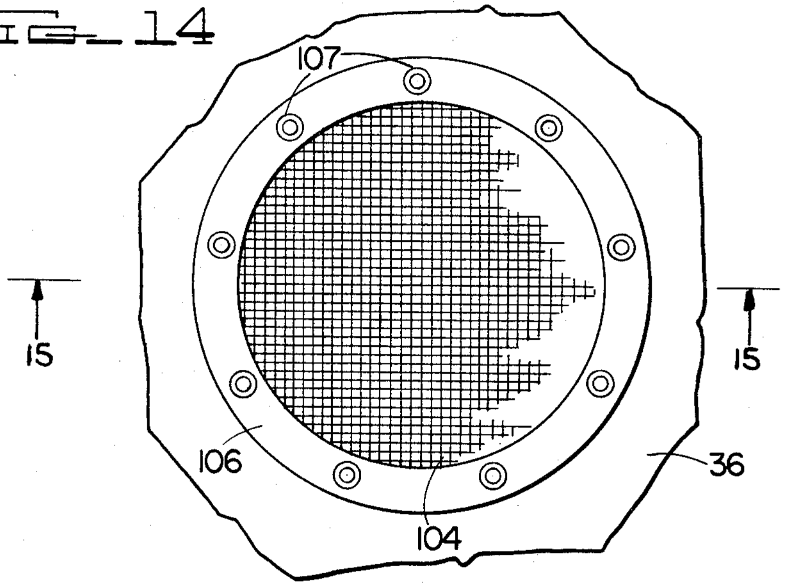
107
15
15
106
36
104

FIG. 15

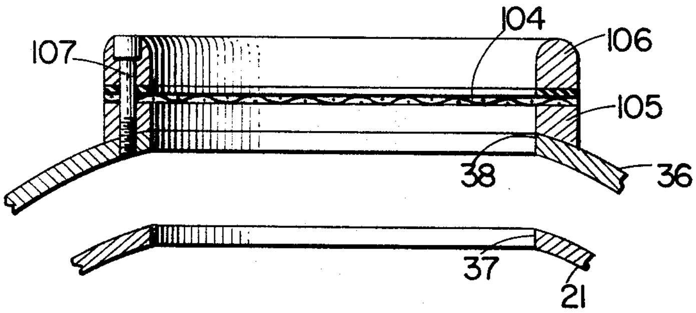
107
104
106
105
38
36
37
21

FIG. 16

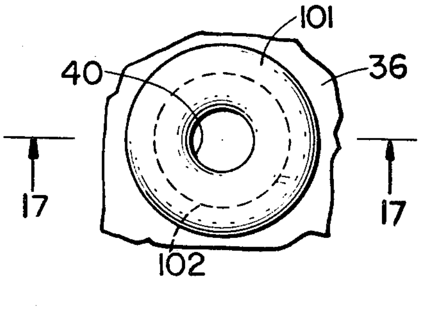
101
40
36
17
17
102

FIG. 17

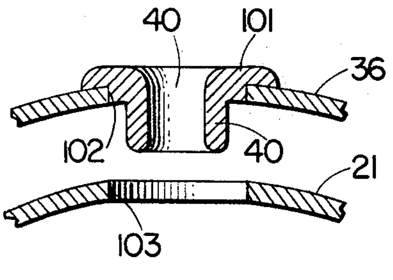
40
101
36
102
40
21
103

INVENTOR
PHILO T. FARNSWORTH
BY Hood Gust & Irish
ATTORNEYS

June 4, 1968 P. T. FARNSWORTH 3,386,883

METHOD AND APPARATUS FOR PRODUCING NUCLEAR-FUSION REACTIONS

Filed May 13, 1966 16 Sheets-Sheet 10

FIG. 18a

0-180° PLANE

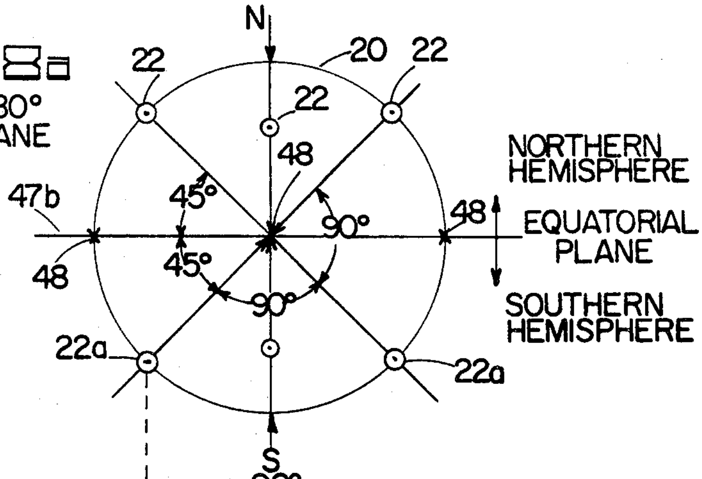
N
20
22
22
22
48
45°
47b
45°
90°
48
48
90°
NORTHERN HEMISPHERE
EQUATORIAL PLANE
SOUTHERN HEMISPHERE
22a
22a
S

FIG. 18b

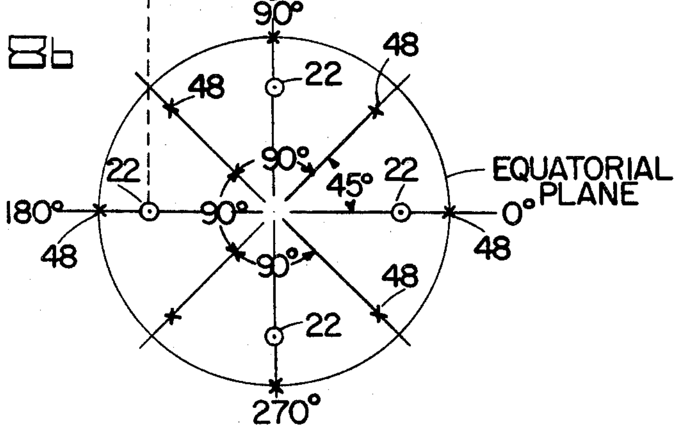
90°
48
48
22
90°
45°
22
22
EQUATORIAL PLANE
180°
90°
0°
48
48
90°
48
22
270°

FIG. 19

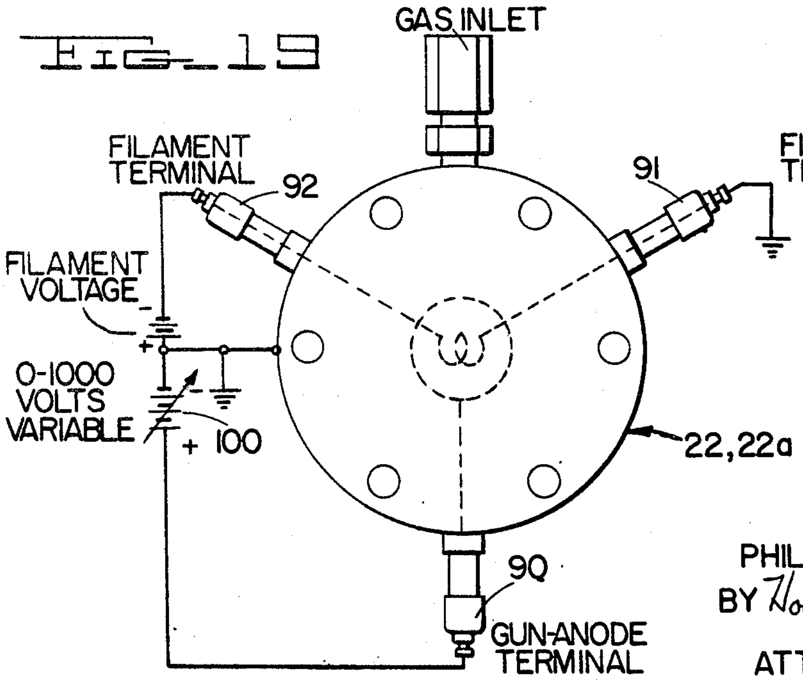
GAS INLET
FILAMENT TERMINAL
92
91
FILAMENT TERMINAL
FILAMENT VOLTAGE
0-1000 VOLTS VARIABLE
100
22,22a
90
GUN-ANODE TERMINAL

INVENTOR
PHILO T. FARNSWORTH
BY Hood, Gust & Irish
ATTORNEYS

June 4, 1968 P. T. FARNSWORTH 3,386,883

METHOD AND APPARATUS FOR PRODUCING NUCLEAR-FUSION REACTIONS

Filed May 13, 1966 16 Sheets-Sheet 13

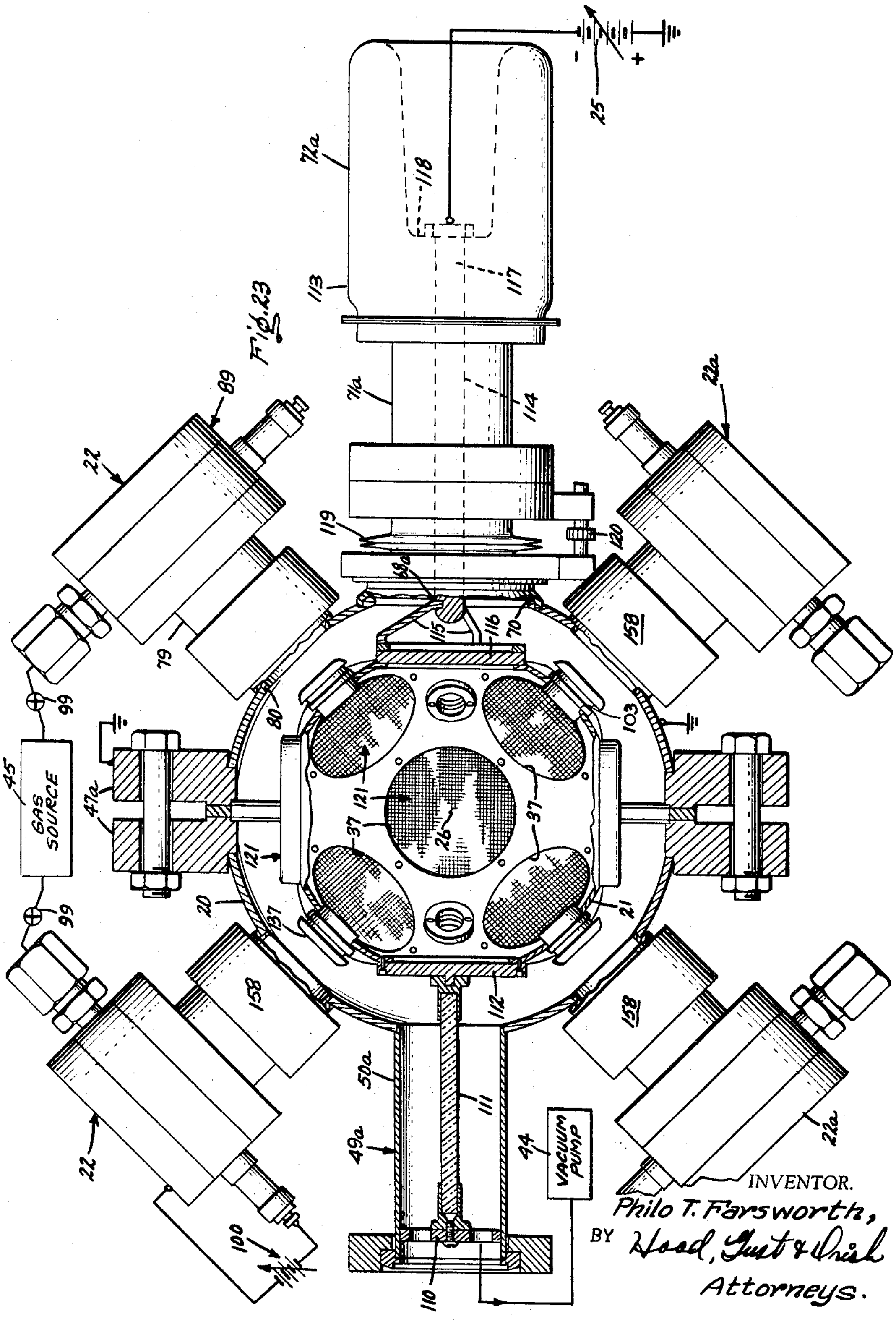

25
-
+
72a
118
117
113
71a
114
89
22
22a
119
120
69a
115
116
70
79
158
80
103
99
45
GAS SOURCE
47a
121
37
26
20
21
112
158
50a
111
49a
44
VACUUM PUMP
22
22a
100
110

Fig. 23

INVENTOR.
Philo T. Farnsworth,
BY Hood, Gust & Irish
Attorneys.

June 4, 1968 P. T. FARNSWORTH **3,386,883**

METHOD AND APPARATUS FOR PRODUCING NUCLEAR-FUSION REACTIONS

Filed May 13, 1966 16 Sheets-Sheet 14

Fig. 25

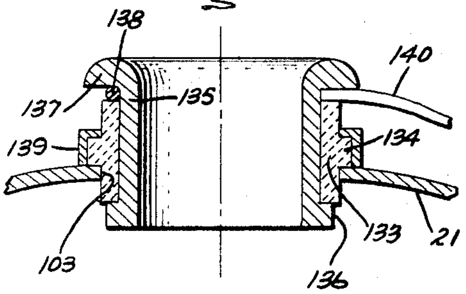
138
140
137
135
139
134
133
21
103
136

Fig. 29

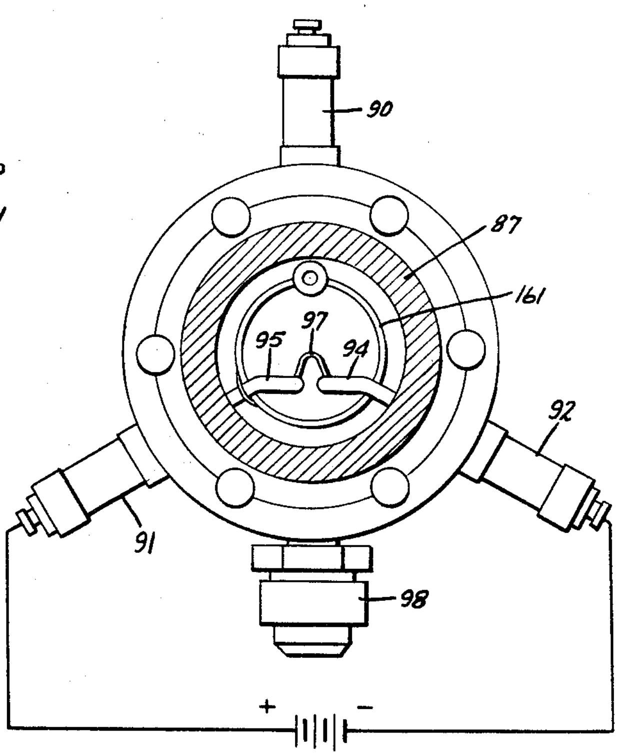
90
87
161
97
95
94
92
91
98
+
−

Fig. 24

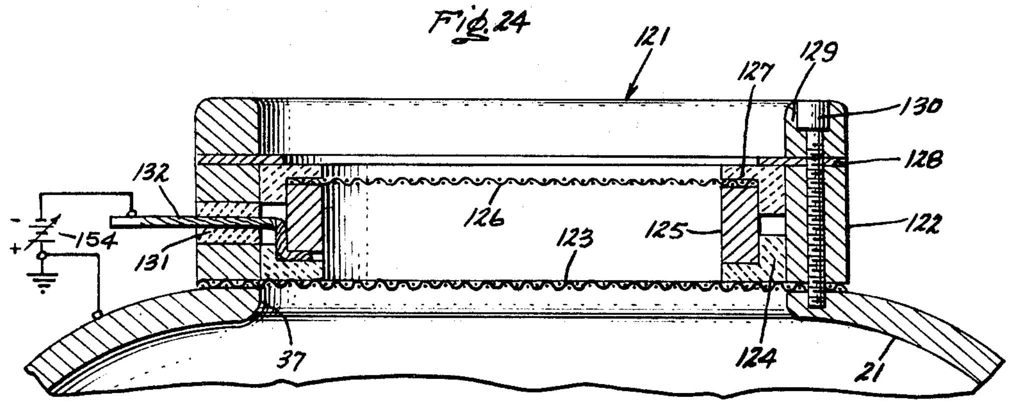
121
129
127
130
128
132
126
125
122
-
+
154
131
123
124
37
21

INVENTOR.
Philo T. Farnsworth,
BY Hood, Gust & Irish
Attorneys.

June 4, 1968 P. T. FARNSWORTH **3,386,883**

METHOD AND APPARATUS FOR PRODUCING NUCLEAR-FUSION REACTIONS

Filed May 13, 1966 16 Sheets-Sheet 15

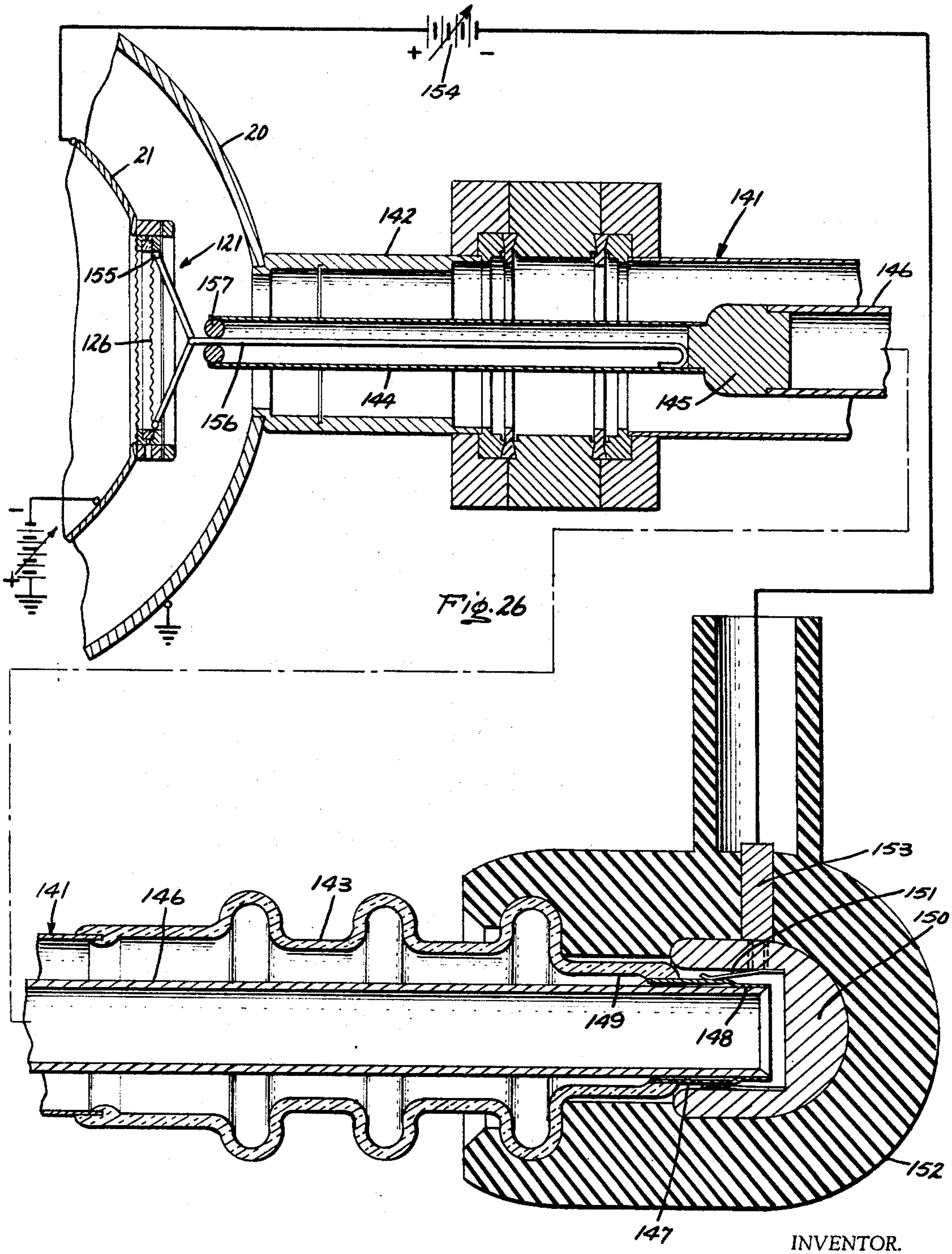
+
-
154
20
21
121
155
157
126
156
142
141
146
144
145
-
+
141
146
143
153
151
150
149
148
152
147

Fig. 26

INVENTOR.
Philo T. Farnsworth,
BY Hood, Gust & Irish
Attorneys.

METHOD AND APPARATUS FOR PRODUCING NUCLEAR-FUSION REACTIONS

Filed May 13, 1966 16 Sheets-Sheet 16

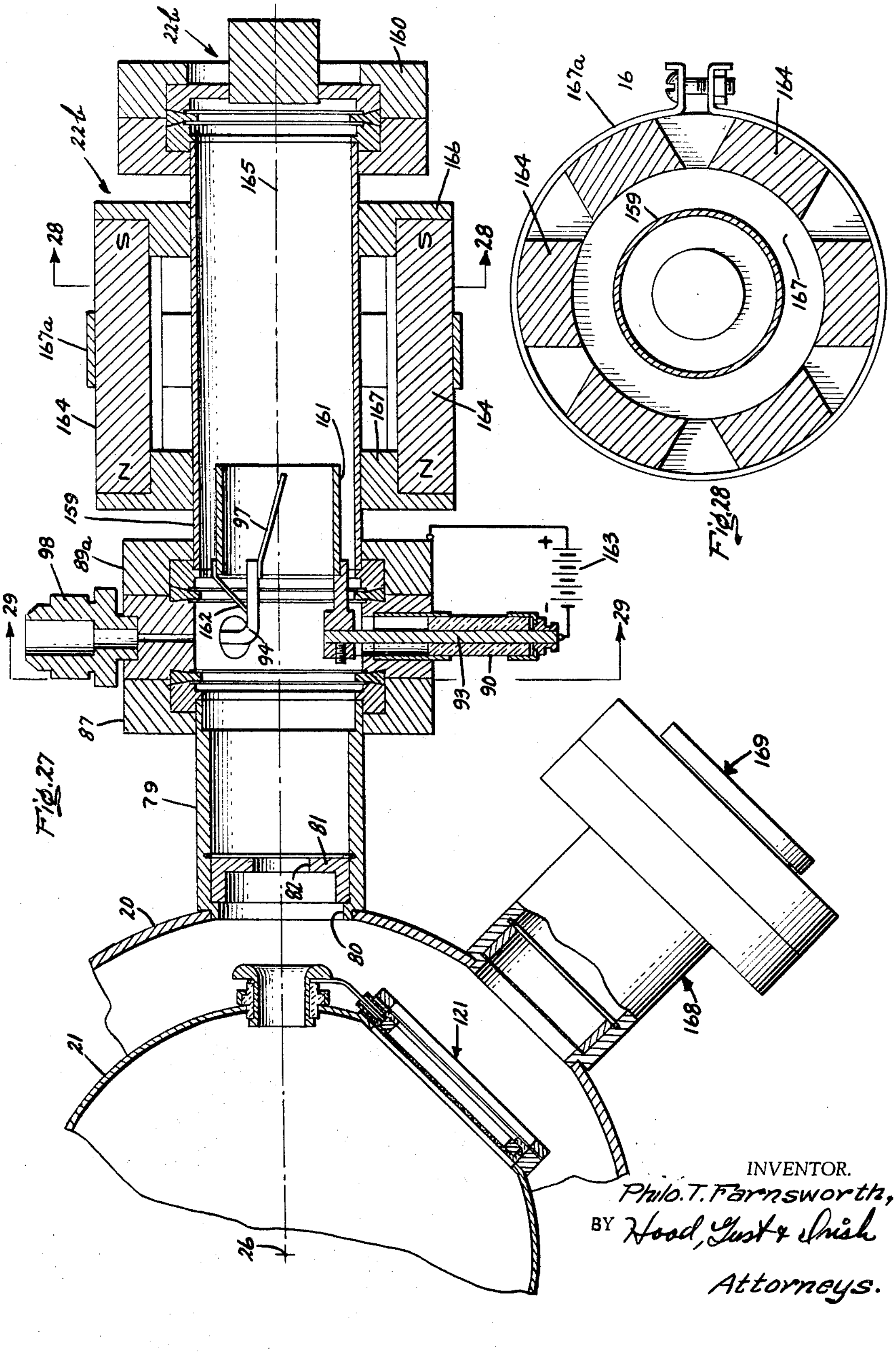
22b
160
22b
166
165
28
28
S
S
167a
164
164
N
N
161
167
159
98
89a
97
162
94
+
-
163
29
29
93
90
87
79
81
82
20
80
121
169
168
21
26

Fig. 27

167a
16
164
164
159
167

Fig. 28

INVENTOR.
Philo T. Farnsworth,
BY Hood, Gust & Irish
Attorneys.

United States Patent Office
3,386,883
Patented June 4, 1968

3,386,883
METHOD AND APPARATUS FOR PRODUCING NUCLEAR-FUSION REACTIONS

Philo T. Farnsworth, Fort Wayne, Ind., assignor to International Telephone and Telegraph Corporation, a corporation of Delaware

Filed May 13, 1966, Ser. No. 549,849
19 Claims. (Cl. 176—1)

ABSTRACT OF THE DISCLOSURE

A method and apparatus for producing controlled nuclear-fusion reaction by use of self-generated electric fields and inertial ionized-gas containment. The apparatus comprises a spherical anode which concentrically surrounds a cathode. A plurality of ion guns are mounted on the exterior of the anode in spherically spaced and diametrically aligned relationship such that the beam axes intersect at the center of the cathode. Appropriate apertures are provided in the cathode for passage of the ions. Other apertures permit passage of positively charged particles outwardly from the cathode interior, but are biased negatively to prevent the flow of electrons into the interelectrode space.

A voltage is applied between the anode and cathode. Ions from the guns are propelled and focused into the center of the cathode establishing in the cathode interior a series of concentric spherical sheaths of alternating maxima and minima potentials called "virtual electrodes." The ions in the innermost "virtual electrode" have fusion energies, and are contained at a density sufficient to produce a self-sustained fusion reaction.

I. *Introduction*

The present invention relates to a method and apparatus for producing nuclear-fusion reactions, and more particularly to a method and apparatus for producing controlled nuclear-fusion reactions by use of self-generated electric fields and inertial ionized gas containment.

In the well-known fusion reactions, nuclei of two light elements are combined to form a nucleus of a single heavier element, together with a release of the excess binding energy in the form of sub-atomic particles (neutrons and protons). Before the positively charged nuclei can be brought close enough together for fusion to take place, sufficient energy must be supplied to overcome the forces of electrostatic repulsion between them. There are many possible reactions involving the combination of two light nuclei which are accompanied by the release of energy, but hydrogen isotopes (deuterium and tritium) and helium, under the proper circumstances, are considered to be the most likely to produce controllable fusion reactions. Examples of these reactions are:

(1) ${}_1D^2+{}_1D^2\rightarrow{}_2He^3$ (.82 mev.)$+{}_0n^1$ (2.45 mev.)
(2) ${}_1D^2+{}_1D^2\rightarrow{}_1T^3$ (1.01 mev.)$+{}_1p^1$ (3.02 mev.)
(3) ${}_1D^2+{}_1T^3\rightarrow{}_2He^4$ (3.5 mev.)$+{}_0n^1$ (14.1 mev.)
(4) ${}_1D^2+{}_2He^3\rightarrow{}_2He^4$ (3.6 mev.)$+{}_1p^1$ (14.7 mev.)

It has been determined that to produce a "self-sustained fusion reaction" (release of more energy from the reaction than is required to produce it) the density of the fusionable particles must be maintained at a high order. It is generally accepted that if such a density could be so contained, the other problems involved in producing a self-sustained fusion reaction, principally that of raising the particle energy-levels high enough to overcome their repelling forces, could be solved. Most suggestions and proposals for plasma containment employ very high magnetic fields; these include the pinched discharge, the

Stellerator, the magnetic mirror, the Astron, and the like. This invention departs widely from those approaches by utilizing self-generated electric fields for containing the ionized gases. Through the use of such electric fields, many, if not most, of the complex problems inherent in the magnetic-field devices have been overcome.

Methods and apparatus capable of producing such continuous reactions are disclosed and claimed in my application Ser. No. 165,639, filed Jan. 11, 1962, entitled, "Electric Discharge Device for Producing Interactions Between Nuclei," now U.S. Patent No. 3,258,402, issued on June 28, 1966. Generally speaking, one form of the invention of my aforesaid application is of spherical geometry, in which an electron-emissive cathode concentrically surrounds a shell-like anode having an inner concentric cavity or space. The anode is permeable to the flow of atomic particles, while the cathode is not. In operation, an electrical discharge, composed of high-order magnitude electron and ion currents in the space enveloped by the anode, produces a radial potential distribution which is, generally speaking, a minimum adjacent to the center of the anode cavity and a maximum adjacent to the anode wall. Ions created at points intermediate the center and the anode wall fall toward and oscillate through the center at velocities which are dependent upon the operating potentials, the potentials developed within the anodic space and the potentials at which the ions are born. With developed potentials of sufficiently high magnitude, the ions are propelled at nuclear reacting energies, so that ion collisions occur at the center and thus produce nuclear-fusion reactions.

The present invention structurally differs fundamentally therefrom in the respect that the cathode and anode elements are reversed, and the anode therefore surrounds the cathode. The cathode is substantially impervious to the flow of electrons therethrough and is reasonably permeable to the free flow of positively charged particles, while the anode shell is not. An electrical gaseous discharge developed within the device causes the concentration of electrons and ions toward and into the central structureless zone of the volume enclosed by the cathode, thereby producing in said zone concentric, abrupt, shell-like, potential barriers of alternating polarities or virtual cathodes and anodes. High-energy ions within the space of the structure are propelled toward and through these virtual electrodes to the geometric center of the structure, there producing, in concert with the low-energy ions created in that region, an ion concentration of extremely high density. When the space charge in the center fully develops, the potential difference between the innermost virtual cathode and the first adjacent virtual anode reaches a magnitude sufficient to propel ions through the center at fusion-reacting energies. Atomic and sub-atomic products produced by these reactions are available for the production of useful power. The present invention involves novel and improved means and methods for confining the ionized gas particles, regardless of the charge; compressing them into small volumes with high densities; and then maintaining them in a stable condition for a prolonged and continuous period of time.

It is, therefore, an object of the present invention to provide a method and means for compressing and confining an ionized gas discharge by the use of self-generated electric fields.

It is another object of the present invention to provide an electric-field method and means for producing a stable electric discharge.

It is still another object of this invention to provide a method and apparatus for converting the energy released from the nuclear reactions into useful heat energy.

It is another object of this invention to provide a method and apparatus for converting positively-charged particles from the nuclear-fusion reaction directly into electrical power.

It is still another object of this invention to provide a source of electrical power and heat in which the kinetic energy of the products of the nuclear-fusion reactions are converted directly into electrical power with or without accompanying neutrons.

II. *Description of theoretical device*

In the accomplishment of the above and related objects, an electric discharge device has an anode which surrounds a cathode that electrically encloses a structureless space or volume. The cathode is rendered practically impervious to electrons but reasonably pervious to ions. The anode and cathode are uniquely designed and assembled to form an electron-optical system wherein the cathodic space is maintained completely filled with the electrical discharge. Traversing electrons, under the influence of the electron optice, follow radial-like paths through the cathodic space and, as the cathode is quite impervious to electrons, they are thereby kept from being intercepted in appreciable numbers by the surrounding anode. Electron energy spread is thus maintained at minimum values, and there is developed a high-magnitude electron circulatory current which serves to produce the required potential gradient in the cathodic space.

A variety of fusion reactions are possible, with certain of the reactions being useful in directly generating electrical power. The charged particles of these particular fusion reactions are emitted at sufficient kinetic energies to overcome the decelerating field of the anode. This results in the particles, charged positively, performing work against the anode field, and thereby augmenting the energy stored in that field. The remainder of the energy is converted into heat at the anode, but the deposit of the positive charges thereon results in their conversion directly into electrical energy or power. Energy converted into heat also may be utilized in the generation of power.

The above-mentioned and other features and objects of this invention, and the manner of attaining them, will become more apparent and the invention itself will be best understood by reference to the following description of embodiments of this invention taken in conjunction with the accompanying drawings, wherein:

FIG. 1 is a diagrammatic illustration of an embodiment of the present invention used in explaining the theory of operation;

FIGS. 2 and 3 are potential distribution curves used in explaining the operating principles of this invention;

FIG. 4 is a simplified diagram used in explaining the theory of charged particle scattering;

FIGS. *5a, 5b,* 6, *6a, 6b, 6c,* 7 and 8 are graphs used in explaining the principles of this invention;

FIG. 9 is a further diagrammatic illustration of an embodiment of this invention (more elaborate than FIG. 1) used in explaining the theory of operation;

FIG. 10 is a graph of the deuterium-tritium reaction used in explaining the operation of this invention;

FIG. 11 is a partial diametral cross-section of a working embodiment of this invention;

FIG. 12 is an axial section of one part of the terminal and supporting structure for the embodiment of FIG. 11;

FIG. 13 is a fragmentary axial section of one ion gun of FIG. 11;

FIG. 14 is an enlarged fragmentary plan view of a cathode biasing screen assembly of FIG. 11;

FIG. 15 is a cross-section taken along section line **15—15** of FIG. 14;

FIG. 16 is an enlarged fragmentary plan view of an ion aperture assembly of FIG. 11;

FIG. 17 is a cross-section taken along section line **17—17** of FIG. 16;

FIGS. 18*a* and 18*b* are diagrams used in explaining the relative positions of the ion guns and the screen-covered apertures in the cathode;

FIG. 19 is an end view of an ion gun;

FIGS. 20 and 21 are cross-sections taken along section lines 20—20 and 21—21, respectively, of FIG. 13;

FIG. 22 is a cross-section taken along section line **22—22** of FIG. 11;

FIG. 23 is a partial cross-section of another working embodiment of this invention;

FIG. 24 is an enlarged fragmentary cross-section of the biasing screen assembly mounted on the cathode shown in FIG. 23;

FIG. 25 is an enlarged fragmentary cross-section of the ion-aperture assembly mounted on the cathode shown in FIG. 23;

FIG. 26 is an enlarged fragmentary cross-section of the terminal assembly for applying the biasing voltage to the biasing screen assembly of FIGS. 23 and 24;

FIG. 27 is an enlarged, fragmentary sectional view of another embodiment of an ion gun assembly;

FIG. 28 is a cross-section taken substantially along section line **28—28** of FIG. 27; and

FIG. 29 is a cross-section taken substantially along section line **29—29** of FIG. 27.

III. *Simplified explanation of operation*

It has been accepted that fusion reactions, which will yield more energy than that supplied to produce the effect, will occur only in high-density plasmas in which the overall kinetic energies of the ions composing the gas are high enough to overcome their mutual repelling forces. Most approaches proposed or under study to produce these conditions employ powerful external magnets to contain and compress the plasma into a small volume to increase the density and temperature (energy) of the particles therein. Such magnetic fields thus far have been successful in confining the plasma particles in the required high-density configuration for only extremely short (microseconds) fusion reactions. The time required to produce the favorable conditions (interval between pulses) is so great that energy output is extremely small when compared with the energy input.

This invention differs fundamentally from those in that charge-particles are compressed into a suitably dense configuration through the action of an "electrodynamic lens" established by confined bi-polar charges in a volume of free space. Such bi-polar charges are developed into a spherical configuration in such a way as to establish a plurality of concentric spherical potential sheaths radially spaced, these sheaths having large potential differences therebetween. The maxima and minima potential sheaths alternate radially and are characterized, respectively, as virtual anodes and cathodes. A virtual cathode is innermost and substantially coincident with the geometric center.

From this arrangement to virtual anodes and cathodes, a radial potential distribution obtains in a spherical space, with one potential minimum being near the center. Thus positively charged particles in this space fall toward and through the center; hence, in effect they are focused onto the center. The focusing forces may be considered as resulting from bi-polar (electron-ion) charge optics.

Bi-polar charges moving radially within the space produce a radial potential distribution as described above. It is, therefore, convenient to characterize these bi-polar charge optics as "Poisson Optics" inasmuch as the solution of Poisson's differential equation (given later on) for the radial potential distribution resulting from bi-polar charges in a spherical configuration reveals the phenomenon of the aforedescribed virtual anodes and cathodes. The electric field which obtains within this free space occupied by these virtual electrodes I call a "poissor."

The trajectories of the charged particles are determined by the composite field or poissor which is produced by the aggregate of the fields of the oppositely charged particles. When a positively charged particle (ion) of zero velocity falls from a point just inside a virtual anode (which defines a spherical positive-potential boundary), it is accelerated toward the adjacent virtual cathode and, in passing beyond it, decelerates toward a condition of zero velocity at an equipotential boundary of that from which it started, and then repeats this velocity rate of change from zero to maximum, and then again to zero in the opposite direction. Electrons likewise will oscillate through virtual anodes between negative equipotential barriers. Since one polarity of charged particle approaches its maximum velocity while the other polarity approaches its minimum, the two species maintain their separate identities and spaced charges. Hence, it is seen that the particles are trapped inertially between such equipotential boundaries; accordingly, I call my method for containing the nuclear-fusion reacting particles "inertial containment." The ions which are created on the inward sides of the boundaries of the spherical virtual anodes and which oscillate through or near the center of the device produce the required density thereat to satisfy the conditions for a self-sustaining nuclear fusion reaction. Thus, the charged particles are literally compressed into the required density in the central region by the process of inertial containment, and the present device utilizes this process.

Referring to the drawings, and more particularly to FIG. 1, an evacuated spherical electron-tube structure is shown which comprises a spherical, anode shell **20** at radius r_a, enclosing a sperical, cathode shell **21** of radius r_c, and an ion gun **22** which is mounted on the outside of the anode **20** as shown. The anode and cathode electrodes are concentrically arranged as shown. The cathode **21** and its field are substantially impervious to electron flow to the anode **20**, but is pervious to the flow of positively charged particles such as ions. In this theoretical exemplification, the cathode **21** may be considered as an open-mesh electrode formed of metallic screen or the like, which is constructed of a material, the surface of which emits electrons upon bombardment by ions or electrons and preferably is photoelectric in the ultra-violet spectral region. Suitable connections are made to the electrodes, a lead **23** being connected to the anode **20** for applying a positive potential thereto and another lead **24** being connected from the cathode **21** for connection to a negative potential terminal. A power source, such as a battery **25**, delivering a suitably high voltage is connected as shown to the leads **23** and **24**. In the preferred embodiment of this invention, the anode **20** is operated at ground potential.

In one embodiment of this invention, ions of a suitable nuclear-fusion reactive gas are introduced into the space enclosed by the cathode **21**. An electrical discharge composed of high-magnitude electron and ion currents forms in the cathodic volume and develops a difference of potential which is, generally speaking, a minimum near the geometric center **26** and a maximum adjacent to the anode **20**, with one or more potential maxima (virtual anodes) and minima (virtual cathodes) concentrically enclosed within the cathode **21**. Ions created at points adjacent to potential maxima (virtual anodes) fall toward and oscillate through adjacent potential minima (virtual cathodes) with energies equivalent to the potentials at the points where they start their journeys. These ions, falling inwardly toward the center **26** from regions of high potentials, will be propelled at velocities (energies) which are sufficient to overcome the repelling forces of other high-energy ions and also the slower (target) ions born in the region near the center **26** and collide, thus producing nuclear-fusion reactions.

In order to obtain ions with fusion-reacting energies, the sizes of the virtual cathodes and anodes must be maintained such that the boundaries thereof are well defined, and the potential differences therebetween are pronounced. This is achieved by methods which are explained in detail in the written description later following.

IV. *Description of poissor (virtual-electrode system) formation*

With suitable potentials applied to the electrodes, and with the ion gun **22** energized, ions of a fusion reactive gas are directed into the space enclosed by cathode **21**, where they will be accelerated radially inwardly toward the geometric center **26**. Since there is a potential gradient between the anode **20** and cathode **21**, the positively charged ions, as indicated by the numeral **27**, will be accelerated by the cathode **21**. Inasmuch as the latter is substantially ion permeable, each ion **27** will continue its transit and, if the optics are considered to be perfect, it will pass through the center **26** and travel onwardly in a straight line until the repelling field of the anode **20** at about the point **28** is reached. At this point, the direction of travel of the ion will be reversed, and it will repeat the diametric excursion to the opposite side of the device adjacent to the anode **20**, passing through the center **26** enroute.

Because the potential within the cathode is uniform (assuming no other ions or electrons are in existence), ions **27** will experience no velocity-changing force while traveling therethrough. Thus, an ion may be considered as starting its travel at or near a given point on the anode, accelerating toward the cathode, traveling with constant velocity through the cathodic space, and then decelerating from the cathode to the vicinity of the anode where its velocity becomes zero and it reverses its direction of travel. This ion will continue its oscillatory motion until it is lost by one of several competing actions, which will be explained later on. The significance of this single-ion concept is two-fold: firstly, it recognizes that the normal space potential inside cathode **21** is uniform at the value of the cathode, and an ion traveling thereacross does so with uniform velocity; and secondly, that the ion will oscillate within the space enclosed by anode **20** until ultimately collected by the cathode **21**.

Now let it be assumed that two ions leave the anode simultaneously from diametrically oposite points, such as from ion guns **22** and **22***a*. Each of these ions will be propelled radially inwardly toward the exact center **26** of the cathodic space so that they would collide at that point in the absence of any other forces. Inasmuch as each ion is positively charged, it will exert a repelling force upon the other, such that their respective velocities will be progressively decreased until they nearly reach the exact center **26**, where they will have given up all of their energies and stop. Under the repelling influence of their respective fields, they will reverse their direction of travel and be accelerated outwardly. In a practical embodiment, however, the ions will experience a diverting effect and will pass each other at minimum velocity rather than stopping. Upon passing through the cathode **21**, the ions will be decelerated by the anode **20** field until they stop adjacent to the anode **20**, whereupon the cycle is repeated. Even though the unipotential space inside the cathode **21** exerts no force on a single ion passing therethrough, two ions approaching each other along a diametral path experience coulomb repulsion and velocity changes which serve to create a positive electric field in the cathodic space, the maximum effect occurring in the central region where the velocity is the least. This may be properly described as a space-charge effect.

Now assume that a copious quantity of ions are introduced into the space immediately adjacent to the anode **20** from a number of symmetrically disposed ion guns. These ions will converge toward the center **26** at progressively decreasing velocities until they reach a minimum velocity and thereafter diverge outwardly along essentially the same diameters, accelerating until they pass out through the cathode **21**. The ions contribute a positive charge to the cathodic space which increases progressively as the coulomb forces become effective and reach a maximum within center **26**; as the ions leave the center **26**, they absorb energy from that field on their passage to the cathode **21**. Thus, surrounding the cathodic center **26** (and inside the cathode **21**), a virtual anode will be produced which can be made to have a potential which can exceed that potential (V_a) of the anode **20** by a factor ΔV_a equal to the average voltage increase which is required by the ions from the ion guns to penetrate the anode **20** field. The ions, therefore, oscillate back and forth through the permeable cathode **21** until lost by one of several competing actions which will be explained later, since they do not re-enter the anode **20**.

The establishment of the ion space-charge inside the cathode **21** may be better understood by reference to the graph of FIG. 2 wherein the abscissa represents the radius of the device and the ordinate represents the potential distribution therein. The magnitude of the ion space-charge is dependent upon the amount of space-charge current flowing in the cathodic space. For a minute quantity of current, the positive charge contribution is small, and the potential at the center would appear as shown by curve *a*. A larger current will produce a more positive potential distribution such as curve *b*. Greater or lesser currents will change the potential at the center **26** correspondingly.

Since the cathode **21** is not completely permeable to the flow of ions therethrough, ions will impact it and dislodge secondary and Auger neutralization electrons thereby. (During operation of the device, electrons also will be supplied from the photoelectric cathode **21** by ultra-violet radiation from Bremsstrahlung and recombination.) Electrons, to be able to leave the cathode **21**, must be excited to a potential ($V_c+\Delta V_c$), V_c being the potential of the cathode **21**. The electrons so emitted by cathode **21** will leave with a Fermi-Dirac distribution of velocities and the potential gradient near the cathode **21** thereby becomes more negative by the factor ΔV_c than that of cathode **21**, and the potential curve moves down until it has a minimum at a radius r_{c_1}, close to that of the cathode **21**, at which the potential corresponds to the average velocity of emission. For this condition, most of the electrons are slowed down until they come to rest, thus creating a virtual cathode **29** at a radius r_{c1}, and from which the electrons may flow in either direction, i.e., either return to the cathode **21** or flow into the central cathodic volume.

Some electrons, under the attractive influence of the positive potential gradient (virtual anode) created in the central volume by the ions and the repelling influence of the virtual cathode **29**, will be accelerated toward the center **26** until their mutually repelling coulomb forces predominate. They then will be decelerated, giving up kinetic energy to the field, and ultimately may be scattered through large angles, returning to the virtual cathode **29**, where they will repeat the cycle. Since the electron density is the greatest in the small spherical volume enveloping the center **26**, the negative space-charge contribution by the electrons will be the greatest in that region and a potential minimum or a crater will develop in the center of the positive gradient *b* (FIG. 2) which will increase in depth until a virtual cathode **30***a*, with a potential equal to that of virtual cathode **29**, is created at the center **26**, radius r_0. This action is accompanied by an outward displacement of the peak of the positive potential gradient to a radius r_{a_1}, where a spherical potential sheath or virtual anode **31** is formed.

The description thus far has disregarded the presence of neutral gas molecules within the device. When neutral gas is present, it initially will be distributed equally throughout the volume enclosed by anode **20**. The electrons and ions oscillating within the cathode **21** space will strike the neutral gas molecules, creating additional electrons and ions which, under the influence of the electric fields, will be set into oscillatory motion, the lengths of their respective paths being determined by the potentials at the points within the cathodic space where they were created. For example, suppose an ion is born in the outward (cathode) side of the virtual anode **31** (FIG. 2); it will be propelled through the virtual cathode **29** and impact the cathode **21**, thereby producing additional electrons. The electron which was released during the creation of the ion will pass through the virtual anode **31**, and be brought to rest in the vicinity of the virtual cathode **30***a* where, neglecting any deviation, it will reverse its direction and return to the point where created. Now suppose the ion and electron were created on the inside of virtual anode **31** (FIG. 2). The ion will move inwardly, being accelerated toward and through the virtual cathode **30***a*, and will come to rest at the positive barrier on the opposite side of the center **26**, and then oscillate between those barriers. The electron will be propelled through the virtual anode **31** and will go into oscillation between the virtual cathode **29** and the virtual cathode **30***a*. By far the greatest number of ions (and electrons) will be produced near the virtual cathode **30***a*, because the probability of ionization by electron impact is greater the slower the electron and because the electron density is the greatest in the virtual cathode.

As the process continues, the ion and electron space currents tend to build up a small positive potential maximum within the virtual cathode **30***a* (FIG. 2) by virtue of the ions that are created in that region and which oscillate through it. Thus, a vestigial virtual anode **32** (FIG. 3) is formed in the center **26**, and the virtual cathode **30***a* moves to a new location **30** at radius r_{c_2}. When neutral, gas molecules are present in the virtual cathode, ionization will occur. The electrons so created will be nearly indistinguishable from and react in the same manner as those in the virtual cathode, i.e., they will be accelerated outwardly toward a virtual anode. The associated ions will be born with the very low kinetic energy of the neutral particles and therefore will remain in the vicinity of the virtual cathode, neutralizing some of the space charge. These combined actions of electrons and ions will result in smaller negative space charge which decreases the radius of the virtual cathode. The new radial potential distribution is shown in FIG. 3. It will be noted that an increase in the electron space-charge current within the virtual anode **32** could result in the formation of a new virtual cathode therein. This process of adding virtual electrodes, theoretically, could continue until a point of confusion is reached where they merge and the field collapses. In actual practice, however, the number of virtual electrodes in a poissor may, and will, be controlled. In the practical exemplifications described herein, only two virtual cathodes (**29** and **30**), one virtual anode **31** and the vestigial anode **32** will be assumed to exist.

It must be borne in mind that in an actual operating device neither the ions nor the electrons are traveling precisely radially; moreover, there is a velocity spread within both the ion and electron groups due to interparticle scattering. As a result, the points where ions and electrons reverse their direction of motion do not lie on mathematically thin surfaces; accordingly, the virtual anodes and cathodes have finite radial thicknesses which are called "potential gradient sheaths" or simply "sheaths."

At this point, it will be appreciated that the space charge buildup inside the cathode **21** has resulted in the development of concentric potential gradient sheaths which serve in confining negatively and positively charged particles to movement in paths, almost all of which intersect at the center **26**, while others oscillate along path lengths which are determined by the energy levels at which the charged particles were created. When a high kinetic energy ion (projectile) passes near a low kinetic energy ion (target) in the central volume, scattering occurs. An energy transfer takes place therebetween which sets the target ion into oscillation and tends to trap the projectile. The positively charged particles for fusion reactions thus are compressed into such a densely packed volume concentric with the geometric center **26** that the probability of those particles being fused is high.

The manifestation of a poissor (virtual-electrode system) by radiation of light within spherical geometries similar to the device described above has been observed through viewing ports and studied photographically, on many occasions.

V. *Other atomic processes in the discharge*

In the immediately preceding discussion, the ions and electrons have been considered as moving past each other, interacting only through their respective space charges. There are a number of other atomic processes which can occur, however, and it will be shown in the following that the design of the apparatus can be such as to minimize some of these effects and use others to good advantage.

As has already been explained, the high-energy ions or electrons, in the absence of outside influences, would oscillate radially through the space of the anode **20** or the cathode **21** (FIG. 3). The ion transit time between the limits of its travel, i.e., the time which an ion takes for one traversal of its path, is proportional to the ion path length and inversely proportional to its average velocity. This ion oscillatory action will, in the presence of electrons and neutral gas, continue only until one of four possible actions occurs:

(a) the ion path is altered by the "scattering process";
(b) the ion "captures" an electron from a neutral atom, called charge exchange, or recombines with an electron and becomes a neutral atom;
(c) the ion is fused by a nuclear reaction; or,
(d) the ion is captured by the cathode or the anode wall.

A scattering event is a collision of or near collision of two or more particles in which the energies and momentums of the individual particles are altered, but total energy and momentum are conserved. In an ionized gas, ion-ion scattering and electron-electron deflections from coulomb forces are of primary importance because energy is most readily transferred between particles of similar masses. Electron-ion deflections involve much less energy transfer because of the large mass difference between the two species.

"Scattering" is the total statistical effect of the probability of electrostatic deflection of like particles with boh radius and distance as variables; i.e., the deflecting of particles of like charge by the coulomb repelling forces. Ion scattering takes place mostly in the virtual anodes where the kinetic energies are low and the ion densities are many fold greater than anywhere else along the ions' paths. Scattering in the virtual anodes cannot result in a large energy transfer. When an ion is scattered in a virtual anode and is accelerated toward a virtual cathode, it gains high kinetic energy and its path is essentially radial. (As the ions and electrons are oscillating within a spherical enclosure, their current densities increase radially inwardly as $1/r^2$ where r is the distance from the geometric center **26** to the point where the density is measured.) An ion will be deflected from its path as it approaches another ion if it is unable to approach closely enough (within the radius of fusion, r_f) for the sub-atomic attractive nuclear forces to predominate. This particular distance between two ion (r_f) where scattering is replaced by fusion, can be calculated if the fusion cross-section for the particular ion is known. For example, if we assume that the fusion cross-section (σ_f) for tritium ions with deuterium ions is:

$$5\times10^{-24}\text{ cm.}^2$$

then

$$\pi(r_f)^2=\sigma_f=5\times10^{-24}$$

$$r_f=\sqrt{\frac{5\times10^{-24}}{\pi}}\simeq1.26\times10^{-12}\text{ cm.} \tag{5}$$

If scattering is too great, it will not be possible to maintain the high ion density at the center **26** necessary to assure the required fusion probability. In other words, the effect of scattering is to increase the size of the minute volume of the central vestigial virtual anode **32**. Thus it becomes advantageous to limit the number of ion trips through the center, for a given particle energy, in order to prevent an increase in that volume by scattering.

The deflection angle will depend upon how closely the two ions approach each other, this angle being directly proportional to the density of the target ions bound within the innermost virtual cathode **30** (FIG. 3) and inversely proportional to the square of the energy of the scattered ion. Ions which approach each other at a distance less than r_f will fuse; those that are unable to close within that range will be deflected. The scattering angle will be greater the closer the ions come to each other. The critical angle where fusion occurs, called the "angle of fusion," can be calculated using Rutherford's formula. Neglecting multiple scattering for the tritium-deuterium ions, this angle turns out to be 58°; thus, scattering will take place only when the angles are less than 58°. It can be shown mathematically that the probability of the scattering angle exceeding 8° is only about 10%.

As an ion has a much shorter wavelength than an electron, its atomic dimensions are much smaller. Two important consequences follow:

(a) Diffraction effects will not occur in elastic scattering. Thus, over a major part of angular range, elastic scattering will be the same as that given by the classical theory. (At very small angles, the quantum theory will give better results.)

(b) Elastic scattering will be confined to very small angles. The scattered intensity will decrease with an increase of scattering angle, and the scattering is concentrated to small angles as the incident energy is increased: the total scattering will decrease with increasing energy.

For elastic collisions, the cross-section will be small if:

$$\frac{r}{v}\times de\gg\frac{h}{2\pi} \tag{6}$$

where

de=ionization energy
h=Planck's constant
v=velocity of the ions
r=radius of ion deflection

The greater the value of de, the greater will be the energy at which maximum inelastic cross-section will occur. Because of small values of de, charge-exchange is the most important inelastic collision process.

An ion can be neutralized by an encounter with an electron. This can happen in two different ways: an ion can collide with a neutral atom and rob it of its electron ("charge-exchange"), or the ion may collide with the free electron and recombine to form a neutral atom. The probability of recombination is extremely unlikely compared to charge exchange (see Brown, "Basic Data of Plasma Physics," Wiley, 1959, page 104) and will occur only if the two colliding particles are traveling in the same direction at about the same velocity. Since the charge-exchange cross-section is greater than 10^{-16} cm.^{2} between 1 kev. and 50 kev., charge exchange will occur over a wide range. The gentler the slope of the potential gradient, the longer will be the ion path length within which charge-exchange can occur. The only region in which recombination can occur is close to the virtual cathodes **29** and **30** where the electron velocity is lowest and ion velocity is highest.

When charge-exchange occurs, there is very little energy transfer between the two particles other than the transfer of an electron from one nucleus to the other. The old ion, which becomes a neutral atom, retains almost all of its velocity, and, since it is no longer influenced by the electric field, it simply travels onwardly until it is lost at the tube wall. The newly formed ion has the energy corresponding to the potential at the point in space where it was created. It begins an oscillatory motion of its own through the center 26, or travels to the cathode 21 with no indication of having been previously deflected. The newly formed ion may be considered to be continuing to increase the average number of trips started by the old ion, but with the disadvantage of having suffered an energy loss. The ion charge-exchange cycle of the inner side of the virtual anode 31 may be repeated several times before the energy loss becomes so great that fusion will not occur.

When recombination occurs, a neutral atom results, which travels outwardly until it is either reionized by another electron or ion or migrates to the tube walls. Recombination results in a loss of energy through radiation as well as a loss of both ions and electrons from the space charge; it is fortunate that the probability of this event occurring is so low and thus represents a negligible energy loss. This condition is achieved by maintaining the paths of the electrons as nearly radial as possible. One means for obtaining this radial motion is to hold the velocity spread of the electrons to a minimum.

Control of charge-exchange by limiting ion trips

Ion scattering will be limited if the number of ion trips through the center of the device can be limited. It will now be shown that this number of ion trips hence charge-exchange can be controlled.

The extent of charge-exchange may be calculated from the relationship which expresses that probability per trip P_c as follows:

$$P_c = \rho_n \times \sigma_c \times L \tag{7}$$

in which:

L=the length of one ion path in centimeters;

ρ_n=the neutral gas density in particles per cubic centimeter; and

σ_c=the charge-exchange cross-section in cm.^{2}.

The highest possible value for P_c is unity (1), since this means that the event of charge-exchange occurs. The obvious minimum value of P_c is zero (0). From the formula, if the calculated value for P_c is 10^{-6}, this means that an ion makes on an average 10^6 oscillatory trips before the energy loss of the ion by charge-exchange becomes so great that it will not be able to reach a fusion radius. Values for charge-exchange cross-section may be obtained from the curve of Jackson & Schriff in "Physical Review," vol. 89, No. 2, page 359, 1953.

In the case at hand, it is desirable to limit the number of ion trips to less than one million during the occurrence of from four (4) to six (6) charge-exchange cycles. If an average of four (4) captures per 10^6 is assumed, then:

$$P_c = \frac{4}{10^6} = 4 \times 10^{-6}$$

Now assuming as an example that the path length is 0.135 centimeter and the value of σ_c is 10^{-16} cm.^{2}, then the value of the neutral gas density ρ_n may be derived from Equation 7 as follows:

$$\rho_n = \frac{P_c}{\sigma_c \times L} = \frac{4 \times 10^{-6}}{10^{-16} \times 0.135} = \frac{1}{250{,}000 \times 10^{-16} \times 0.135}$$
$$\simeq 3 \times 10^{11} \text{ particles/cm.}^3$$

Thus, it is readily observed that the control of charge-exchange, which in turn governs to a major extent the shape of the potential distribution, is achieved by adjustment of gas density as charge-exchange cross-section is inversely proportional to gas density. Thus it is desirable to work at the lowest practicable gas density.

Maintaining radial thinness of virtual cathode by controlling electron scattering

During space charge build-up, electrons initially have energies for oscillating within the confines of the cathode 21 along radial paths which pass through the center 26. These electrons suffer the same type of deflective forces as approaching ions, such that meeting electrons scatter. Electrons approaching each other at or near the center 26 deflect along different paths which are also radial. However, if the electrons should meet at points other than the center 26, they will deflect along new paths which may become either more radial or less radial. As the electron transits continue, the non-radial electrons can and do meet with other electrons, producing further non-radial and radial electrons such that eventually, after a relatively large number of electron transits, electron movements inside the cathode 21 tend to become, in some measure, non-radial. FIG. 4 shows a simplified example of this deflection for two different locations of an electron (33 and 34) with respect to an electron traveling along a non-radial path. In FIG. 4, the electron 33 is deflected along the new (and radial) path, while electron 34 is deflected along a more non-radial path. If the deflection angle is large, some of the electrons could, unless otherwise prevented, acquire sufficient energy to penetrate the virtual cathode 29 and the cathode 21 and be absorbed by the anode 20 or to enter the cathode 21 and thus be removed from the field; others will accumulate orbital momentum.

The required condition for producing sharply resolved virtual anodes and cathodes is initially achieved by controlling the number of electron trips within the cathodic space. This condition may be represented by the equation:

$$I_e = 2\overline{N}_e \times i_{e2c} \tag{8}$$

where

I_e=the circulatory electron current through the virtual anode 31 inside the cathode 21 (FIG. 3),

i_{e2c}=the electron current emitted from the inner surface of the cathode 21 (FIG. 3),

$\overline{N}_e$=the average number of electron trips ($2\overline{N}_e$ is the average number of round trips).

If a particle's path is through the center, the distance from its starting point to the point where it commences its return journey is the equivalent of two trips or one "round trip." Where the path of the particle is away from the center, the distance from its point of origin to its reversal point is one trip. It is desired that the factor $\overline{N}_e$ fall somewhere between the figures of 10^3 and 10^4. If $\overline{N}_e$ becomes too large, the virtual cathodes and anodes will not form properly. Thus, by preventing the number of electron trips from becoming excessive, the non-radial component of the electron space-current can be held to a minimum.

The number of electron trips may be made relatively small by maintaining a high electron current density within the cathodic space. Since the maximum value of the circulatory electron current I_e is limited by space-charge saturation, which is a function of the cathode voltage gradient, I_e (Equation 8) may be made to approach its saturation limit by increasing the cathode current i_{e2c}, thereby confining a very large number of electrons within the cathode 21 which make but few round trips. Accordingly, few orbital electrons will be formed. In a practical device, the electron current can be increased by the use of electron emitters within the cathode 21.

The paths of the orbiting electrons are elliptical, and each such electron is therefore subjected to a precessional force which causes the electron's path to sweep a spherical surface inside the cathode. These electrons contribute to the electron circulatory current and also to the fields of the virtual cathodes. Because of their elliptical paths, with non-radial velocity components, orbital electrons cannot approach the virtual cathodes as closely as radial electrons. Thus, their maximum space charge contributions occur farther from the virtual cathodes than those from radials, and the virtual cathodes will tend to be displaced inwardly. Excessively large numbers of orbital electrons (greater than 10% of the electron circulatory current) therefore are undesirable as they prevent the maintaining of sharply defined virtual electrodes. In actual practice, the orbital electrons appear not to exceed that limit.

Orbital electrons, in addition to their contribution to the electron circulatory current, further reduce the probability of recombination. In the fusion-reaction volume enclosed by the virtual cathode **30**, this becomes extremely important as it greatly reduces the loss of fusion-reactive ions from that region. In the region outside of the virtual cathode **30**, the orbital electrons, as previously shown, not only contribute to the field which maintains the virtual cathode **29**, but also displaces that cathode inwardly toward the virtual anode **31** and thereby reduce greatly, any cusps or singularities in the electric field pattern of the virtual cathode **29** produced by the anode **20** field penetrating through the apertures in the cathode **21**.

VI. Particle flow

Further consideration of the process involved in the present invention requires investigation of the space-charge optics and the particle flow. The laws by which charged particles move from an emitter to a collector without regard to the direction of voltage between electrodes have been described by Langmuir and Blodgett ("Physical Review," vol. 24, No. 1, pp. 49–59, June 1924). That particular derivation was made to calculate the currents limited by space-charge between concentric spheres (either electron or ion but not in the presence of each other). For this invention, however, bi-polar charges exist within the spherical electrodes, and the Poisson's equation becomes:

$$\frac{d^2V}{dr^2}+\frac{2}{r}\left(\frac{dV}{dr}\right)=4\pi(\rho_e-\rho_i) \tag{10}$$

where V is the potential at any radius r, and ρ_e and ρ_i are the electron and ion volumetric space-charge densities.

Equation 10, together with the energy equations

$$\tfrac{1}{2}M_e v_e^2=eV, \text{ and } \tfrac{1}{2}M_i v_i^2=e(V_b-V)$$

where M is the mass of the charged particle, e the charge of the electron, v the velocity of the charged particle at any potential V, and b is the point of ion entry (which in this invention is considered to be the anode since most of the charged particles are assumed to be formed at or very near to their respective electrodes); and the density, charge and velocity relationships

$$J_e=\rho_e v_e, \text{ and } J_i=\rho_i v_i$$

(where J is the charge of the space-charge) are sufficient to determine all the relations involved in the electron and ion space-charge flows; i.e., all motions are assumed to be radial in this basic determination.

By following the general approach employed by Langmuir in a later paper, "The Interaction of Electron and Positive Ion Space Charges in Cathode Sheaths," ("Physical Review," vol. 33, page 954, June 1929), Equation 10 reduces to

$$3\mu\frac{d^2\mu}{d\gamma^2}+\left(\frac{d\mu}{d\gamma}\right)^2+3\mu\left(\frac{d\mu}{d\gamma}\right)=1-\gamma\left[\frac{\mu^{4/3}}{\mu_b{}^{4/3}-\mu^{4/3}}\right]^{1/2} \tag{11}$$

where

$$\lambda=\frac{I_i}{I_e}\sqrt{\frac{M_i}{M_e}} \tag{12}$$

I_i=the circulatory ion current;
I_e=the circulatory electron current in the presence of ions;
$\gamma=ln(r/r_0)$, (r_0 being the radius of the emitting surface); and
$\mu=f(\gamma)$, (corresponding to the α of Langmuir and Blodgett and becomes α when λ equals zero). In this ideal treatment μ_b is taken at the anode and is thus a constant.

A quantity similar to λ was used by Langmuir in his June 1929 referenced paper which he called α. He showed that the potential gradient at the surface of the anode was proportional to $(1-\alpha)^{1/2}$ and so became imaginary if $\alpha>1$. In other words, if α were to become greater than unity, ions would be withdrawn faster than injected. While those calculations referred to a plane-parallel discharge, the same limitation has been found to apply to the spherical concentric shells if the area used in evaluating the current density J from the total current I is taken to be the area of a surface four-fifths of the distance from the emitter to the collector. (See I. Langmuir and K. T. Compton, "Review of Modern Physics," Vol. III, No. 2, April 1931) where one is confronted with radial flow, however, λ may be larger than unity and still satisfy the boundary conditions.

Once I_e is selected, then λ and I_i are uniquely determined; the ratio of the radii is likewise uniquely determined.

It will be noted, from Equation 12, that when one is concerned with a fixed ratio of the radii of the virtual electrodes, λ has but one value and therefore becomes a constant. When λ is zero (no ion current present), Langmuir and Blodgett's α may be substituted for μ in Equation 11 and reduces to:

$$3\alpha\frac{d^2\alpha}{d\gamma^2}+\left(\frac{d\alpha}{d\gamma}\right)^2+3\alpha\left(\frac{d\alpha}{d\gamma}\right)=1 \tag{13}$$

which, as should be, is the Langmuir and Blodgett equation.

Spangenberg (book "Vacuum Tubes," 1st ed. 1948, § 8.4, p. 181) submits that the equation for the electron space-charge flow of current for concentric spherical electrodes in the absence of ions is:

$$I_0=\frac{29.34\times10^{-6}\times V^{3/2}}{\alpha^2} \tag{14}$$

where V is the potential difference between the spherical electrodes and α is a function of $\lambda=ln\ (r/r_0)$ given by the Maclaurin series:

$$\alpha=\gamma-0.3\gamma^2+0.075\gamma^3-0.0013\gamma^4+\ .\ .\ . \tag{14a}$$

The values for α are valid whether the cathode is external or internal and can be found in tabular form in Table II of the referenced Langmuir and Blodgett paper. Two functions designated as α^2 and $(-\alpha)^2$ are listed. The first applies to the case where the voltage V is evaluated at a point P outside the electron emitting electrode; the second where P is inside that source.

Germane to this invention, Langmuir's logic may be extended into the determination of electron and ion circulatory currents in the present of each other. As was the case for the Langmuir and Blodgett Equation 13, Equation 11 probably cannot be integrated directly, but a solution was obtained in the same manner as used by Langmuir and Blodgett. The electron circulatory current in the presence of ions took the form:

$$I_e=\frac{29.34\times10^{-6}\times V^{3/2}}{\mu^2} \tag{15}$$

The values for μ being obtained by expanding the Taylor's series:

(15a)

$$\mu=f(\gamma)=\gamma_0+\frac{\gamma+\gamma_0}{1}f'(\gamma_0)+\frac{(\gamma-\gamma_0)^2}{2}f''(\gamma_0)+\frac{(\gamma+\gamma_0)^3}{3}f'''(\gamma_0)-\ .\ .\ .$$

Equation 15*a* is expanded near zero to avoid infinite terms in the higher derivatives.

It is apparent from Equation 15 that the electron circulatory current I_e is independent of the actual sizes of the spheres since the radii enter only as a ratio.

Knowing the value of λ and the circulatory electron current from Equation 12, circulatory ion current I_i can be determined. Thus:

$$I_i = I_e \times \lambda \times \sqrt{\frac{M_e}{M_i}} = \frac{29.34 \times 10^{-6} \times V^{3/2}}{(-\mu)^2 \sqrt{\frac{M_e}{M_i}}} \lambda \tag{16}$$

It was at this point in the solution of the Poisson equation for bi-polar charges in a spherical geometry that I was led to conclude, and which was later verified by experimentation, that a phenomenon of alternating virtual cathodes and anodes could be made to exist in the volume enclosed by the inner hollow electrode of a fusor. Furthermore, to satisfy the Poisson's equation, λ would have to be greater than unity when a virtual cathode enclosed a virtual anode, such as shown by r_{c_1} and r_{a_1} (FIG 3). A "singularity" or mathematical discontinuity appears to exist at the virtual anode. Likewise, such a "singularity" appears to exist at the virtual cathode r_{c_2}, enclosed by the virtual anode r_{a_1}. In this type of system "a space-charge saturated current cannot be produced on the side of an emitting surface which radiates outwardly." The radial potential distribution for the fusor of this invention is shown in FIG. *5a*. That such "singularities" do exist in the fusor of this invention can be likened to the case of a charge uniformly distributed over an imaginary spherical shell.

If one assumes that a positive charge Q is uniformly distributed over an imaginary spherical shell of radius r_1, Gauss' law for electrostatic charges may be applied to calculate the field intensity (E) and the potential (V) at any radius r inside and outside of the radius r_{a_1} (FIG. 5). Additionally, the surface charge density ρ_s may be calculated as a function of the radial distance r. (See Sec. 1–20, Chapter 1, Part I, of "Electromagnetics" by John D. Kraus, McGraw–Hill Electrical and Electronics Engineering Series, 1st Ed., 1953.) It is found that the field intensity E is zero (in the absence of electric charges) for any radius inside the shell and the absolute potential at a radius outside the shell is given by

$$V = \frac{Q}{4\pi r \xi_0} \tag{17}$$

where Q is the positive charge in coulombs; and ξ_0 is the permittivity of air in farads per meter. Since the field intensity is zero inside the shell, the potential within the shell will be constant. The surface density ρ_s is zero everywhere except at the radius of the shell where it has the value

$$\rho_s = \frac{Q}{4\pi r_1^2} \tag{18}$$

The variation of E, V and ρ_s is shown by FIG. 6, curves (*a*), (*b*) and (*c*).

The similarity of the radial potential gradients of FIGS. *5a* and *6b* is marked. Inasmuch as the gradient (dV/dr) at r_1 (FIG. 6) and r_{a_1} (FIG. *5a*) is zero, and a condition for space-charge saturation at r_{a_1} is defined by an infinitely thin spherical shell at that radius, it is reasonable to assume that Equations 17 and 18 can be applied to the treatment of a radial potential distribution such as shown by FIG. *5a*, even though that field is not produced by bound charges. It is helpful to identify the referenced emitting surfaces by subscripts (cathode *c* and numeral or anode *a* and numeral) which produced the circulating current (electron or ion) being calculated.

Since the circulatory currents through all virtual electrodes have constant values for all radii between the limits of travel thereof, it becomes fundamental that:

$$\frac{I_e}{I_o} = k^2 \tag{19}$$

where k^2 is a constant independent of the voltages at any radii. Using Equations 14 and 15, Equation 19 may be written in the form:

$$k^2 = \frac{\dfrac{29.34 \times 10^{-6} \times V_a^{3/2}}{-\mu_c^2}}{\dfrac{29.34 \times 10^{-6} \times V_a^{3/2}}{-\alpha_c^2}} = \frac{-\alpha_c^2}{-\mu_c^2} \tag{20}$$

when the real cathode *c* is the emitting surface.

Langmuir's Equation 12, et seq., in his paper of June, 1929, teaches a method for computing values of I_e/I_o for plane-parallel electrodes for any values of his α (which corresponds to my λ for the spherical-electrode geometry).

When the ratio of the radii r_{a_1}/r_{c_1} (FIG. 3) approaches unity, the spherical geometry approaches the plane-parallel geometry so closely that the potential gradient at r_{c_1} becomes zero and I_e/I_o (and k^2) takes the same value as Langmuir's for α equal to unity, which is 1.8605. This means that the electron circulatory current in the presence of unlimited ions will be 1.8605 times that for the same applied potential in the absence of ions.

When λ becomes zero, i.e., no ion current present, I_e/I_o again is the same as Langmuir's α and I_e/I_o (and k^2) which is unity.

Thus, one observes that with the limits for Langmuir's α and my λ being identical, it holds that α and λ will be the same for intermediate values. Accordingly, I have computed and plotted in FIG. 7 the values of I_e/I_o for various values of λ so as to permit obtaining any intermediate values readily.

Experimentally, one may apply a desired potential between the real anode **20** at radius r_a and the real cathode **21** at radius r_c (FIG. 3) and measure the current I_o therebetween. Ions from the ion sources are then injected and measurements of the resulting currents I_e recorded. The ratios of I_e/I_o may then be calculated and the corresponding values for λ obtained from FIG. 7.

Thus, having values for I_e/I_o (or k^2), I_e and λ, one may compute values for the corresponding I_i $(-\alpha)^2$, and $(-\mu)^2$, keeping in mind that the virtual electrodes cannot emit outwardly and only the negative values for α and μ are valid. With these data, one has the means for obtaining the radial potential distribution.

The radius of the virtual anode r_{a_1} (FIG. *5a*) is generally known quite accurately from experimental data of a variety of kinds and largely by heuristic reasoning based on those data. The actual value of r_{a_1} (and the resulting ratio r_{c_1}/r_{a_1}) depends upon the total energy losses, which will be taken up later, required to maintain the desired poissor and the magnitude of the high-energy ion current which flows through (or into) the real cathode **21** (radius r_c). The approximate values for r_{a_1} have been determined visually for experimental fusors.

In the design of a working embodiment of a fusor, it is essential that the volume enclosed by the innermost virtual cathode, such as **30** at radius r_{c_2} (FIG. *5a*) be adequate to confine a sufficient total number of fusion particles (at a volumetric density of unity fusion probability) to assure the generation of a useful amount of power. It has been found that if the ratio of the radii of the adjacent pair of virtual electrodes is much greater than 3, the inner virtual cathode may not form; if the ratio of radii is too small, the circulatory currents required to maintain the poissor become excessively large.

In an actual working embodiment of a fusor, the radii of the virtual anodes and cathodes can be computed, as well as the values of the currents and the radial potential distribution outside of the virtual cathode **30** at radius r_{c_2}. The radial potential within the aforementioned virtual cathode has not been computed because as the radii of the balance of the virtual electrode system become much samller, errors in ion and electron optics begin to reveal themselves to the extent that it is not known whether or not the system continues to repeat itself. Since fusion does occur in the small volume in the fusor's center, the space-charge conditions prevailing therein have therefore been determined by logic and heuristic reasoning.

In FIG. 3 is illustrated diagrammatically one embodiment of this invention consisting of an anode **20**, a cathode **21** and an ion gun **22**. Enclosed within the cathode **21** is shown a poissor with a virtual cathode **29** at a radius of r_{c_1}, a virtual anode **31** at a radius of r_{a_1}, a virtual cathode **30** at a radius of r_{c2}, and a vestigial virtual anode **32** at a radius of r_o. To understand the operation of the fusor, it is important to distinguish between the currents which flow through the apertures of the real cathode **21**, referred to herein as "mesh currents" (i_m), and the current which flows into the real cathode **21**, termed "cathode current" (i_c).

The ion "mesh current" (i_{im}) is produced by the ions:

(a) injected into the cathode **21** space from the ion guns;
(b) released by the anode **20** under bombardment by electrons and neutral gas molecules;
(c) created in the volume between the anode **20** and the cathode **21**; and,
(d) created by the ionization process within the volume r_{a_1} to r_{c_1}.

Those ions which originate on the anode **20** will be collected by the cathode **21**, or returned to the anode **20** after but one round trip; those few ions created on the inward side of the virtual anode **31** ($r \leqq r_{a_1}$) which have sufficient energy to penetrate therethrough will be collected by the anode **20** or the cathode **21** after only a few round trips; the other ions will oscillate from the virtual anode **31** through the cathode **21** apertures until collected by the cathode **21**. The potential of the virtual anode **31** will build up to that of the anode **20**.

The electron "mesh current" (i_{em}) which initially may be quite large and is extremely useful in the formation of the virtual electrode system, will become so small in operation as to be negligible. It is made up of those electrons emitted by the inner surface of the real cathode **21** and which flow through the cathode **21** apertures to the anode **20**.

The total mesh current through the cathode **21** apertures, therefore, is the algebraic sum of the ion mesh current and the electron mesh current. Thus,

$$i_m = i_{im} + i_{em} \tag{21}$$

The net cathode **21** current i_c is the algebraic sum of the ion current into, and the electron currents into and out of it. Ions flowing into and electrons flowing out of the cathode **21** as usual are positive currents to the cathode **21**. Thus,

$$i_c = i_{ic} + i_{e2c} - i_{e1c} + i_{ec}' \tag{22}$$

where i_{ic} is the ion current into the cathode; i_{e1c} is the electron current into the inside of the cathode; i_{e2c} is the electron current out of the inner surface of the cathode; and i_{ec}' is the electron current from the outside surface of the cathode to the anode.

The space current i_s into or out of the volume enclosed by the cathode **21** is the sum of the mesh current and the cathode **21** current, so

$$i_s = i_m + i_c \tag{23}$$

The general condition for an equilibrium potential field occurs when i_s becomes zero, as no charges are then being added or removed from the space. Accordingly,

$$i_m = i_c = 0 \tag{24}$$

and when i_c is zero, the power consumed by the device is a minimum.

It is the ion mesh current i_{im} into the central cathode **21** volume which maintains the density of the virtual anode **31**, and which causes any free charges therein to be oscillated radially. As the inward flow of ions continues, the positive potential at the virtual anode **31** increases until all the inward flow of ions is cancelled by the outward flow to the cathode **21**. The net ion mesh current then becomes zero. Neglecting the minute electron mesh current i_{em}, when all the electrons that are emitted from the inside of the cathode **21**, as well as all of the ions and electrons generated within the cathode **21** volume, return to the cathode **21**, it is apparent from Equation 22 that only the electron current, i_{ec}', remains to flow to the anode **20**. Thus, for i_c to become zero, i_{ec}' must likewise be brought to zero. Such an equilibrium potential field can be obtained when there are sufficient ions reaching the anode **20** to compensate for i_{ec}'.

Those ions from the virtual anode **31** which have energies equivalent to or greater than V_a will reach the anode **20** and remove an equal or greater number of electrons therefrom. With a sufficient number of such ions reaching the anode **20**, the component of cathode **20** current, i_{ec}' can be reduced to zero.

In the disclosure thus far, a space-charge saturation of the cathode volume has been assumed. The ions which are created near the virtual anode **31** ($r \leqq r_{a_1}$), where the ionization process is a maximum, have energies approaching eV_{a_1}. They traverse the inner virtual cathode **30** and the vestigial virtual anode **32** at fusion-reacting energies; this is an essential operating condition. The bulk of the ions which make up the ion mesh current i_{im} are created in the volume outside the virtual anode **31** ($r \geqq r_{a_1}$) and, like those ions from the ion guns **22** and the anode **20**, are ultimately collected by the cathode **21**; thus, if space-charge saturation could exist in this region, it would produce such an extremely large energy loss that nuclear-fusion reactions could not obtain. Fortunately, as previously shown, ion space-charge saturation cannot exist in this region, and operating conditions are maintained such that the ion circulatory current is just adequate to assure a virtual anode **31** of the desired size. Thus, the ideal radial potential distribution for the device of this invention (sometimes referred to herein as a "fusor") is as shown in FIG. 5*a* where the poissor is maintained by ion space-charge saturation in the volume $r \leqq r_{a_1}$ and a non-space charge (dissymmetrical) field exists in the cathode **21** volume where r is greater than r_{a_1}. λ, therefore, has the value which gives ion space-charge saturation only in the region adjacent to the radiating surface of the virtual anode **31** where the potential gradient is zero; electron space-charge saturation obtains in the region just inside of r_{c_2}.

Within a virtual anode **31** (FIG. 5*a*), each positive charge must be matched by an electron. Most of the flux lines from the ions at a radius $r \leqq r_{a_1}$ terminate in the virtual cathode **30**; where $r \geqq r_{a_1}$, most of the flux lines terminate in the virtual cathode **29**. All the charged particles in the cathode **21** volume are free to move and are in oscillation. The virtual anode **31** at r_{a_1}, the locus of maximum ion density, has the potential practically equivalent to V_a, and very few ions are formed inside that virtual anode with energies greater than eV_{a_1}. Thus, to maintain the potential V_{a_1}, it is only necessary that there be an equal number of flux lines from the inner of r_{a_1} to r_{c_2} as there are outward to r_{c_1} (FIG. 5*a*). Ions with lesser energies than eV_{a_1} will reach their reversal loci at radii greater than r_{a_1} when $r \geqq r_{a_1}$ or less than r_{a_1} when $r \leqq r_{a_1}$ and those ion densities will be less than at r_{a_1}. If a few ions are generated inside of r_{a_1} with energies greater than eV_{a_1}, they will pass through the virtual anode **31** and fall to the cathode **21**, and in some instances to the anode **20**; they do not contribute to the ion circulatory current. The virtual anode **31** thus acts as an effective hi-pass ion filter for ions with energies greater than eV_{a_1}. Inside the virtual anode **31** where $r \leqq r_{a_1}$, the ions travel slowly for a relatively large portion of their radial paths as they approach their respective reversal loci; the opposite is true outside the radius r_{a_1} where the ions are decelerated rapidly as they approach r_{a_1}, giving up their energies within a very short distance from r_{a_1}.

The radial potential V_r is a function of the radius r and the time t; i.e., $V_r=f(r,t)$. The buildup of the poissor inside the radius r_{a_1}, together with the creation of the asymmetrical field outward from r_{a_1}, is shown diagrammatically by FIG. 8. The anode **20** and the cathode **21** initially are at ground potential as shown. If a negative potential of V'_c (curve *a*, FIG. 8) of approximately −30 kv. is applied to cathode **21**, and ions from the ion guns are injected into the cathode **21** volume concurrently, they begin to charge that volume positively. This charge will increase until the potential at some internal radius r_{a_1} is sufficient to return the ions to the cathode **21** at the same rate at which they enter the cathodic **21** volume. This equilibrium may be expressed by the relation

$$i_{\text{im}}=i_{\text{ic}} \tag{25}$$

As soon as the potential inside the cathode **21** becomes positive, electrons emitted from the inner surface of cathode **21** will be accelerated toward and oscillate through the central volume and quickly establish virtual cathodes at r_{c_1} and r_{c_2}. With the virtual cathodes formed, the position of the virtual anode between them is determined. The equilibrium radius and the potential of the virtual anode at r_{a_1} requires a somewhat longer buildup time than the formation of the virtual cathodes at r_{c_1} and r_{c_2} because of the longer ion transit time.

When the negative potential on the cathode **21** is abruptly increased from V_c' to $V'_c{}'$ to $V'_c{}''$, similar sequences of events occur. Each of the subsequent equilibrium radial potential gradients is shown by curves *b* and *c* of FIG. 8.

To determine the position of the virtual electrodes, it is only necessary to establish the radial density distribution of electrons and ions. The radial potential distribution already has been so calculated. The virtual anode locus will lie close to the surface where the ion density is the greatest. The virtual cathodes will lie at the termini of the electron paths where the electron densities are the highest.

A practical embodiment of this invention by which the aforesaid conditions for a self-sustaining nuclear-fusion reaction can be produced is shown diagrammatically by FIG. 9. A spherical anode **20** of radius r_a encloses dual concentric cathodes **21** and **36** of radii r_c and r_c', respectively, and ion guns **22** and **22***a* mounted on the outside of anode **20**. The concentric cathodes **21** and **36**, which are constructed of an electron-emitting material such as stainless steel, molybdenum or the like, and which have an openness of about 70% by virtue of spherically spaced, registered apertures **37** and **38**, are supported within the anode **20** by suitable insulators (not shown). The concentricity of the cathodes **21** and **36** is maintained by suitable insulators therebetween (not shown). Attached to the outer cathode **36** on the axis passing through the ion gun **22** is a conical sleeve **40** through which the intense pencil-like ion beam is focused. Suitable connections are made between the unidirectional power supply **25** and the electrodes through leads **23**, **24**, and **42**. The anode **20** is maintained at or near ground potential as shown, while negative potentials are applied to the cathodes **21** and **36** from a supply **43**, with the outer being more negative than the inner.

In operation, most of the ions from the ion guns **22** and **22***a* are utilized to establish the initial virtual anode in the central volume of cathode **21** as previously described. Electrons from cathode **21** are prevented from reaching the anode **20** by the negatively biased outer cathode **36**. The early removal of ions from the non-space-charge saturated field is greatly enhanced by the second cathode **36**, and only those positive charges with energies greater than V_a have any likelihood of avoiding capture on their first outward pass from the virtual anode **31** (FIG. 5*a*). Ions captured by the outer cathode **36** (FIG. 9), will create electrons therefrom which either will enter the cathodic space or be returned by the outermost virtual cathode.

VII. *Losses*

In order to form and sustain a radial-potential distribution such as is shown in FIG. 5*a*, it is first necessary to compensate for all losses. For a typical small fusor, these losses, which represent about 10% of the total power created, in their order of importance and magnitude are:

(1) Heat loss—7%
(2) Charge interchange—2%
(3) Power required to create electrons—⅔ of 1%
(4) Bremsstrahlung—⅓ of 1%

Heat loss.—In the center of an inner virtual cathode in a poissor, the ions from the inner side of the adjacent virtual anode are traversing the center from all directions. As they pass near other ions, they will experience some deflection which will result in small deviations from radial paths and thus they will appear to be in random motion. This is the equivalent of a very high apparent temperature, about 1.6 ×10⁹ °K. for 100 kev. or about 11,600 °K./c.v. (Maxwell-Boltzman constant). This apparent temperature reduces to about 2×10^6 degrees K. in a distance of a few millimeters from the center. Any neutral atom which passes through this central region will be heated to an extremely high temperature. It will experience a cryogenic pumping action that will expel it to one of the real electrodes where, upon impact, it will give up its kinetic energy (heat), penetrate the surface to a depth determined by its kinetic energy, then diffuse out in a random direction following the cosine law. There is very little likelihood of its again reaching the center as it will remain in the vicinity of the electrode until ionization takes place.

It becomes very serious when too many high-energy neutral atoms are intercepted by the cathode as the cooling thereof is most difficult. On the other hand, cooling of the anode is relatively easy. The cooling of the cathode may be accomplished by conduction through its electrical connections to the power supply. The size of the cathode does affect the cooling problem, as increasing the size reduces the heating per unit area by the ratio of $1/r^2$ since the circulating currents are independent of the radii of the spheres.

Rutherford's scattering equation shows how the neutron count can be made to increase with an increase in applied voltage, for: the current rises by the three-halves power; and the particle density may be increased as it is proportional to four times the energy (or voltage) squared. Accordingly, the heat loss will decrease with an increase in voltage as there will be less scattering. For example, refer to FIG. 10, which shows a typical neutron cross-section curve for an equal mixture of deuterium and tritium. The operating points *a* and *b* are for the same fusion cross-section (σ_f), but from the point of view of the heat losses, *b* provides for more efficient performance, for:

$$\frac{N_{\frac{\theta}{2}}}{N_o}=\frac{\pi\rho_c{}^4}{4w^2}\cot\theta \tag{26}$$

where

$$N_{\frac{\theta}{2}}$$

is the number of ions deflected by half angles; N_o is the total number of ions; and W is the energy ($mv.^2=2W$); ($(mv.^2)^2=4W^2$) which shows that the lower the voltage the greater the losses.

Charge interchange.—Ions travel relatively slowly within the radial thickness adjacent to the virtual anode due to the repelling forces of their like charges. Therefore, there is a fairly high probability that a charge interchange with a neutral atom may occur. This loss is discussed in my Patent No. 3,258,402, previously mentioned. This loss becomes sizeable if there are too many gas molecules present. The gas density (pressure) must be kept as low as is possible to maintain a high fusion probability. This affects the fusor's design characteristics. In practice, the fusor is operated at a neutral gas density of 3.535×10^{13} molecules or less, such that

ρ_n is the normal gas density of 2.687×10^{19} molecules/cc. at normal temperature (20° C.) and pressure (760 mm. of Hg); and

ρ_i the operating gas density, in ions per cc., is

$$\frac{2.687\times10^{19}}{760\times10^{-3}}\times10^{-6}=3.535\times10^{13}\text{ ions/cc.}$$

at a working pressure of one micron (10^{-3} mm. Hg).

As the power level of a fusor is increased, more ion current will be required from the ion guns, which in turn will increase the radius of the virtual anode **31** (FIG. *5a*). This will mean a larger volume which, for ion space-charge saturation, will result in an increase of the total number of the ions contained therein. The greater power input (ions) will increase the power loss, thus again increasing the radius of the virtual anode until equilibrium is reached where the power input equals the power dissipated.

Bremsstrahlung.—When high-energy electrons are deflected on their passage through the virtual anode **31**, they radiate at right angles to their directions of motion. As the electrons have momentum which must be conserved, there will occur an equal and opposite reaction when an electron changes its direction of travel. Since these reactions take place after the electron changes its direction, they are non-radial. These non-radial electrons broaden the virtual cathode **30** with a movement toward the center **26**, and thus increase the electron temperatures. The so-called "radiation dampening" effect of the excess ions (positive nuclei) in the virtual anode tends to reduce the Bremsstrahlung losses to less than 1% of the power available in the virtual anode. With the value of the ion density in the virtual anode, the electron scattering can be computed by the well-known Rutherford formula for scattering.

VIII. *Fusion reactions*

The fusor, as shown diagrammatically in FIG. 9, is initially evacuated to a pressure of from 10^{-9} to 10^{-10} millimeters of mercury by means of a vacuum pump **44**, the anode **20** being the hermetically sealed envelope. Potentials are applied to the various electrodes as already explained and a fusion-reactive fuel, in the form of gas, is introduced into the ion guns **22**, **22*a*** from sources **45** and **45*a*** at a pressure of from 10^{-6} to 10^{-4} millimeters of mercury. This fuel gas is ionized by the guns **22**, **22*a*** and formed into ion beams directed diametrically toward center **26**. A poissor forms, as previously explained.

High energy ions oscillate through the region of the center **26** along diametral paths, terminating adjacent to the virtual anode **31** (FIG. *5a*). Other lower energy ions follow shorter diametral paths within the virtual anode **31** through the same region, thereby contributing to the density of ionic particles in the center. With the virtual anode **31** potential high enough (100 kv.), collisions of the high energy ions with others in the center **26** region will produce nuclear-fusion reactions.

There are a number of nuclear-fusion reactions which are possible. Fusion reactions of an exothermic nature will be specifically considered as one type of reaction which this invention seeks to attain. Power production by such reactions is theoretically proportional to both the amount of energy released per reaction and the number of reactions taking place per unit time. The number of reactions per unit time is obtained from the product of the nuclear cross-section, which expresses the probability for a specific nuclear reaction to occur, the number of ions in the center **26** region, and the number of projectile particles passing through this region from the anode **20** and the virtual anode **31** (FIG. *5a*) per unit time. The nuclear cross-section, or the probability that a nuclear reaction will occur, is a function of the velocity or energy of the interacting particles, which in this device is a function of the potential difference between the anode **20**, the virtual anode **31** and the center **26** (FIG. 3). From the above, it is clear that good power-producing reactions require a large nuclear cross-section and also a large amount of input energy.

One reaction which theoretically is outstanding in meeting the above requirement is the tritium reaction with deuterium. The nuclear equation for this process is:

(3) ${}_1D^2+{}_1T^3\rightarrow{}_2He^4$ (3.5 mev.)$+{}_0n^1$ (14.1 mev.)

This states that a triton plus a deuteron plus the sum of their kinetic energies results in a nuclear reaction whose products are helium 4, a neutron, and the sum of the excess binding energy and the kinetic energy possessed by the original trition and deuteron. The reaction energy released is 17.6 million electron volts (mev.) for the sample above. This value is large compared with the values for other possible reactions, which in most cases yield 3 or 4 mev. The nuclear cross-section for the reaction shown in Equation 3 peaks at a value of about 5×10^{-24} cm.^{2} for projectile energies of 100 kiloelectron volts (kev.). This cross-section value is about 10^2 times larger than that for most competing reactions when compared at the same projectile energies.

Additional possible reactions are the following:

(1) ${}_1D^2+{}_1D^2\rightarrow{}_2He^3$ (.82 mev.)$+{}_0n^1$ (2.45 mev.)

(2) ${}_1D^2+{}_1D^2\rightarrow{}_1T^3$ (1.01 mev.)$+{}_1p^1$ (3.02 mev.)

(4) ${}_1D^2+{}_2He^3\rightarrow{}_2He^4$ (3.6 mev.)$+{}_1p^1$ (14.7 mev.)

Reactions 1 and 2 have a lower value of released energy and a lower 100 kev. cross-section value than Reaction 3. Reaction 4 has about the same energy value but a lower 100 kev. cross-section value.

IX. *Fusion products in the discharge*

As already explained, when an external voltage is applied between the anode **20** and the cathode **21**, **36** (FIG. 9), a system of virtual anodes and cathodes (poissor) is formed within the cathodic space. The radial potential distribution of the poissor was shown in FIG. *5a*. It is within the minute volume enclosed by the virtual cathode **30**, where the ion density is the greatest and the energies (velocities) of the projectile ions a maximum, that the probability of fusion is the highest.

It was shown by Equations 1, 2, 3 and 4 that, using only six deuterons, it is possible to trigger four different nuclear-fusion reactions, and these reactions result in the creation of five new kinds of particles having a total energy of 43.2 million electron volts (mev.). Four of these five particles (${}_1p^1$, ${}_1T^3$, ${}_2He^3$ and ${}_2He^4$) are positively charged, and carry 26.65 mev.; the fifth is a neutral particle, the neutron, and the two produced by the two reactions carry 16.55 mev. There are two important aspects of this energy distribution between the charged and uncharged particles. Firstly, the neutrons will escape from the reacting system and dissipate their energy elsewhere as heat, because being uncharged, they easily pass through material walls. Only the charged particles will be retained within the reacting region and be available to compensate for energy losses (radiation, recombination, etc.) and to sustain the nuclear-fusion reactions. Secondly, it is only the energy of the charged particles, which represents 64% of the energy released by the four reactions, that is available to be converted directly into electrical power. It should be noted that only those ions which are created by the fusion reaction will have sufficient energy to produce electrical energy.

All of the particles formed by the nuclear-fusion reaction are ejected radially outward from the reaction volume with the high energies shown by Equations 1 through 4. These energies (velocities) are more than adequate to propel them through the virtual anode **31** (FIG. 5*a*). Under the influence of the real cathode **21, 36** (FIG. 9), they will be accelerated on their outward journey until passing through the cathode apertures **37, 38**, where they will be decelerated by the anodic field and either brought to rest short of the anode **20** or impact the anode. If particles do not reach the anode **20**, they will then reverse their direction of travel and go into oscillation through the center. Those which do strike the anode **20** will dissipate their remaining energies as heat and at the same time deposit their charges on the anode. A few particles may strike the cathode and be captured, with an accompanying release of secondary electrons.

With deuterium only in the fusor, for two D—D reactions (Equations 2 and 3) will take place at essentially equal rates. The positively charged products of these reactions which go into oscillation increase the ion circulatory current, thus increasing the ion density in the reaction space and hence the probability of fusion. The oscillating reactive particles ($_1T^3$ and $_2He^3$) may fuse with deuterons to produce the reactions shown by Equations 1 and 4, or may experience scattering, capture or recombination with electrons and become neutral atoms. The non-fusionable particles ($_1p^1$ and $_2He^4$) contribute to the fusion process only by the creation of fusion-reactive ions through the capture of electrons from fuel-gas atoms and by their contribution to the space-charge; ultimately, they will become neutral gas atoms. The neutral gas atoms will drift toward the electrode walls, where eventually they will be scavenged from the reaction system by the vacuum pump 44 (FIG. 9).

It has been stated that some of the charged particles will reach the real anode **20** when ejected by the nuclear-fusion reactions. When this occurs, the particles remove enough electrons from the anode **20** to neutralize their positive charges and form gas atoms. This is the equivalent of producing a current which flows from the anode **20** to the battery **25** (FIG. 9) (power supply). In other words, the charged particles have been coverted directly into electrical energy. This is one of the unique features of this invention.

Insertion of a suitable resistor **46** (FIG. 9) in series with the power lead **24** becomes a means for coupling the converted electrical power to an external load. A switch **47** is used to connect selectively this resistor **46** into the series arrangement.

It will be seen from Equations 1 through 4 that as many positive charges enter the fusion region as are removed therefrom. In other words, nuclear charges are conserved. What the fusion reactions do accomplish is the formation of new particles with the release of the excess binding energy of the original ions and, as the original energy must be conserved, the excess is applied to the new particles. Accordingly, the new particles leave the reaction volume at extremely high velocities. The replacement ions which enter the reaction volume do so at the comparatively low velocities of the order of 100 to 200 kev. Thus, there is an extremely short period when the potential in the center is less than prior to the fusion reaction which, in effect, tends momentarily to increase the negative potential of the virtual cathode **30** as well as decrease the virtual anode **31**, and there is an associated increase in the electron current. This effect produces an increase in the radii of the virtual electrodes until the original potential of the vestigial anode **32** (FIG. 5*a*) can be restored. Thus, there is a periodic expanding and contracting of the poissor volume by the very high frequency oscillation of the inertially contained particles.

When the fusion reaction commences, there will be more positive charges leaving the center than flowing in and this condition will continue until steady state is reached. This would tend to develop what might be termed a "potential well" (more negative potential) at the virtual cathode **30**. Such a condition could not exist, as current would then flow from the real cathode **21, 36** (FIG. 10) to the virtual cathode **30**. Accordingly, the virtual cathode **30** must be returned to the real cathode potential within an extremely short period of time. This is accomplished by an equivalent increase in the potential of the virtual anode **31**. The loss of positively charged particles from the center produces a positive bias on the virtual cathode **30** which is equivalent to an increase in the potential of the virtual anode **31**. As these increases are cumulative until the steady state condition is produced, it is seen that the virtual anode **31** (FIG. 5*a*) reaches the potential of the real anode **20**.

The positively charged particles formed by the fusion reaction perform work against the field of the virtual anode **31** as they pass through it. In so doing, their velocities are reduced somewhat, thus increasing the field and contributing to the potential of the virtual anode. They are then accelerated by the cathode, regaining energy.

It is the potential of the virtual cathode **30** (FIG. 3) which contributes to the energy of the positively charged particles which oscillate through the center **26**. As the energies of the reacting particles increase, within the resonance limits of the nuclear-fusion cross-section, the probability of fusion increases. Electrons which are created within the virtual anode **31** are ejected radially outwardly with sufficient energy to overcome the repelling field of the real cathode.

If one were to lose control of the input density of the fuel-gas atoms, it might appear that the process of building up the potential of the virtual anode **31** could continue unabated and might well result in a catastrophic accident. Such is not the case, however, as there are "built-in" safety processes which act to prevent such a build-up. The principal control is the cross-section resonance phenomenon which occurs in nuclear reactions. The cross-section curves for D—D (total), D—T and D—HE^3 reactions, such as contained in "Physical Review," vol. 88, p. 468, 1952, by Connor, Bonner and Smith, and also "Physical Review," vol. 88, p. 473, 1952, by the same authors, show that the D—T curve exhibits a maximum at an energy of 110 kilo-electron volts (kev.), where it falls off sharply with increasing deuteron energy. In the same energy range, the D—He^3 cross-sections are appreciably less than the D—D (total) and the D—T values, but the curve is rising rapidly with increasing deuteron energy and it crosses the D—D (total) curve at 120 kev.; it approaches the D—T curve at 500 kev., its peak, and then falls off rapidly. For example, in the range of deuteron energies from 120 to 200 kev., the D—T cross-sections decrease 1 barn (10^{-24} cm.^{2}) while the D—D (total) and the D—He^3 cross-sections increase only about 0.5 barn, an overall decrease of approximately 1 barn. After the D—He^3 cross-sections pass their maximum at 500 kev., the decrease in the overall cross-sections is much more pronounced. Thus, it is apparent that the build-up in the potential of the virtual anode will be checked at deuteron energies greater than about 150 to 200 kev.

Another built-in safety process which assists in preventing a "run-away" by the so-called "potential well" is the widening of the negative potential sheath (virtual cathode **30**) (FIG. 5*a*), as the radius is changed by the oscillating particles, which produces a decrease in the fusion cross-sections.

The conversion of the positively charged particles into electrical energy at the anode has been touched upon briefly. This phase will now be covered in some detail. In order to prevent the over-heating of the anode, it is highly desirable that most of the particle energy be given up to the decelerating virtual and real anode fields. The particles would reach the anode **20** with very much less than their original energies and thus dissipate less heat to it with a significant increase in the overall efficiency. The ideal situation would permit the charges to reach the anode **20** (FIG. 9) with little or no remaining energy, deposit their charges and drift away as neutral gas atoms. Obviously, with such a wide range of energies, 0.82 to

14.7 mev., this condition cannot be achieved. Therefore, one may envision two different operating situations; namely, (a) be satisfied with the maximum electrical conversion that can be obtained with the minimum heating of the anode **20**, or, (b) operate the fusor as a heat machine to convert heat and charged particles into the maximum amount of electric power.

In the first situation, it becomes necessary to restrict, as nearly as possible, the fusion reactions to those of Equations 2 and 3, where the energies range from 0.8 to 3.02 mev., and to operate the anode at a potential which will produce a deceleration of the 0.8 mev. particles such that they arrive at the anode **20** with just sufficient energy to deposit their charges. The 3.02 mev. particles under this condition will impact the anode **20** and convert its remaining energy into heat. Inasmuch as approximately 50% of the D—D reactions will follow the proton branch, it is not possible to eliminate that higher energy particle. The reactions of Equations 1 and 4 may be reduced appreciably by operating the fusor at a lower than normal gas density, thereby removing the neutral helium-three and tritium atoms by the use of a large volume vacuum pump **44** before they have an opportunity to fuse.

In direct contrast, if the device is to be operated as a heat generator, the D—T and D—He^3 reactions must be encouraged. Since the contribution of the D—T and D—He^3 reactions could be quite significant in the direct conversion to electricity, it might prove advantageous in some applications to produce tritium by bombarding lithium-six with the neutrons:

$$(27)\qquad {}_3Li^6+{}_0n^1\rightarrow{}_2He^4+{}_1T^3+4.6\text{ mev.}$$

and then utilize the tritium as a fuel gas.

If enough positively charged particles can be made to reach the anode **20** to reverse the current flow from the battery **25**, it then becomes possible to remove the applied anode (cathode) voltage **25** while continuing the nuclear-fusion reactions. If I and V_a are the measured current and anode voltage before fusion, then after fusion commences the input power in watts will be

$$(28)\qquad W=V_a\times\left[I-\langle F\rangle e(A-B)\right]$$

where $\langle F\rangle$ is the fusion expectancy; e is the electric charge; A is the interception factor of the positively charged particles by the anode; and B is the interception factor of the positively charged particles by the cathode. Since the cathode **21, 36** is almost pervious to positively charged particles, A will be very much greater than B and when the value of

$$\left[-\langle F\rangle e(A-B)\right]$$

becomes greater than $I(V_a)$, the fusor will become truly self-sustaining. This condition will be reached without the aid of an external load. To remove electrical energy in excess of that required to drive the fusor, it only becomes necessary to add an external load **46** such as shown in FIG. 9.

Working embodiment

A working embodiment of this invention as shown in FIG. 9 is illustrated in substantial detail in FIGS. 11 through 22. In these figures, like numerals indicate like parts. The anode **20** is of spherical construction, being fabricated of two hemispherical shells of stainless steel hermetically sealed together at the equators by means of suitable annular sealing flanges **47***a*. Concentrically positioned inside the anode **20** are the two cathode shells **36** and **21**, respectively, which are also assembled of hemispherical shells of stainless steel. The means for supporting these electrodes concentrically will be described in detail hereinafter.

Eight ion guns **22, 22***a* are mounted on the exterior of the anode **20** in spherically spaced and diametrically aligned relationship, these guns all having beam axes which intersect at the center **26** of the fusor. The precise location of these ion guns **22, 22***a* is explained in connection with the spherical diagrams of FIGS. 18*a* and 18*b* which show the angular relationship therebetween. Each of the ion guns is indicated by a circle with a center dot indicating the spot at which the gun axis intersects a spherical surface defined by the anode **20**. FIG. 18*a* illustrates a spherical surface corresponding to the anode and cathode shells with the plane of the drawing being the same as the cross-sectional plane of FIG. 11. This latter plane includes the north and south poles of the sphere. The line **47***b* indicates the equatorial plane at right angles to the plane of the drawing. In FIG. 11, this equatorial plane would pass midway between the annular flanges **47***a*. The guns **22** and **22***a* are located in the planes spaced 45° from the equatorial plane **47***b* as shown, these planes passing through the tube center **26** and being rotated about the equatorial diameter which is perpendicular to the plane of the drawing. As shown in FIG. 18*b*, which is a right angle projection of the spherical surface of FIG. 18*a*, there are four such guns **22** spaced 90° apart on the spherical surface and on the same meridian. In terms of angular displacement, these guns are indicated as lying on the 0°, 90°, 180° and 270° lines. Within each plane containing guns, each gun center is at the intersection of a radius 45° above the equator and the spherical surface, certain of the radii lying in the plane of FIG. 18*a*.

Four ion guns **22** are so located in the upper or northern hemisphere of the sphere while four other guns **22***a* are located allochirally in the southern hemisphere.

The apertures **37** and **38** in the cathode shells **21** and **36**, respectively, are in twelve registered pairs, each pair **37** and **38** being diagrammatically indicated by an X mark which is assigned numeral **48** in FIGS. 18*a* and 18*b*. The locations of the ion guns and the apertures **48** are indicated as being on the same spherical surface for showing the angular relation therebetween. There are four screen aperture pairs **48** on the equator as shown in FIG. 18*b* at the 0°, 90°, 180° and 270° lines. Planes intersecting the north and south poles which are spaced from each other by 90° contain the remaining aperture pairs **48**, these planes being spaced 45° from the 0°–180° plane which includes the north and south poles also. In these individual 90° planes, on a radius at an angle of 35°16′ with respect to the equatorial plane **47***b*, lie the remaining aperture pairs **48**. The spherical surface containing the aperture pairs **48** may be regarded as that of one of the cathode shells **21** and **36**.

With the equators of the cathode and anode shells lying in the same equatorial plane, the aperture pairs **48** are thus spherically offset in a regular angular pattern from the ion guns **22** and **22***a*.

The anode **20** shell, as well as the cathode shells **21** and **36**, are supported in position by a tubular mounting assembly indicated generally by the numeral **49**. This mounting assembly includes a stainless steel pipe section **50** perimetrically welded at one end to an opening **51** in the anode shell **20**. Coaxially and hermetically secured to the opposite end of the tube **50** is a larger diameter supporting tube or pipe section **52**. Inside this larger tube **52** is a smaller diameter sleeve **53** having a sliding fit therewith, the upper end portion of sleeve **53** being secured to the radial flange **54** which joins the two pipe sections **50** and **52**.

At the bottom end of the sleeve **53** is coaxially secured a mounting ring **55** from which depends four metal posts **56** spaced 90° apart around the axis of the pipe section **52**. A metal disc **57** is coaxially secured to the bottom ends of the posts **56** by means of suitable screws **58**, this disc **57** having secured to the central portion thereof a ceramic post **58***a*. This post **58***a* is coaxially positioned inside the pipe section **52** as well as the smaller pipe section 50 by means of the mounting sleeve, rings and posts just described.

Secured to the upper end of the post 58*a* and coaxially extending thereabove is a metal post 59 having threadedly connected thereto a disc-like mounting stud 60 which is welded into an opening 61 in the outer cathode shell 36. Abutting against the flat surface of this stud 60 is a ceramic or the like spacer 62 having a boss 63 intimately fitting into a locating aperture 64 in the inner cathode shell 21. The spacer 62 is sandwiched between the two shells 21 and 36, thereby determining the spacing therebetween.

The post 59 is of such length with respect to the pipe section 50, for example, as to position the two cathode shells 21 and 36 concentrically inside the anode shell 20. This post 59 may be regarded as coinciding with the south poles of the cathode and anode shells.

At the north poles of these shells is located further supporting structure as well as a terminal assembly for connecting the operating voltages to the electrodes. This assembly includes an annular spacer 65 of suitable insulating material, such as boron nitride, positioned between the two shells 21 and 36 so as to determine the spacing therebetween. This spacing is the same as that between the south poles. Coaxially secured inside this spacer 65 is a conductive terminal button 66 which is intimately conductively fitted into a companion aperture in the shell 21. This aperture coincides with the north pole of the shell.

Telescoped over an upwardly projecting portion 67 of the spacer 65 is an elongated tubular element or conductor 68 which intimately fits into and makes conductive contact with an aperture 69 in the outer shell 36.

This conductor 68 passes diametrically outwardly through the relatively large insulating aperture 70 in the anode 20 in spaced relation therewith, this spacing being large enough to provide insulation at the high voltage at which the fusor is operated. A conductive tubular assembly 71 is rigidly and hermetically sealed to the exterior of the anode 20 in registry with the aperture 70 and extends coaxially outwardly with respect to the conductor 68. At the upper end of this assembly 71 is secured a relatively long, cylindrically shaped insulator 72 (FIG. 12) which has secured to the upper end thereof a conductive terminal assembly 73. This terminal assembly 73 is provided with a suitable connection terminal 74 by means of which cathode bias from the battery 43 may be connected to a wire 75 which extends coaxially through the conductor 68 and connects to the terminal button 66. This tubular assembly 71, 72 and 73 which coaxially surrounds the conductor 68 is hermetically sealed such that the space between the conductor 68 and the assembly 71, 72 and 73 may be evacuated and held at the same pressure as the interior of the anode 20.

Inasmuch as high operating potentials prevail between the conductor 68 and the outer tubular assembly 71, 72 and 73, suitable corona shields 76, 77 and 78, of annular configuration, are used for the purpose of preventing corona discharge. Other configurations for the terminal and supporting assembly may be used without departing from the spirit and scope of this invention, the requirement for such an assembly being that it provide the necessary physical support for the fusor anode and cathode elements and also be capable of supplying the anode and cathode elements with the operating voltages without producing corona discharges or field emission.

All of the ion guns 22 and 22*a* are identically constructed such that a description of one will suffice for all. For this purpose, reference is made to FIGS. 11, 13, 19, 20 and 21. Each ion gun includes a conductive housing or sleeve 79 of stainless steel or the like which is perimetrically sealed to an opening 80 in the anode shell 20. The sleeve 79 is positioned such that its axis coincides with a diameter of the anode 20. Coaxially secured internally to the sleeve 79 and against the anode 20 is an aperture plate 81 having a beam-defining aperture 82. Coaxially positioned inside the sleeve 79 in radially spaced and insulated relation therewith is an anode sleeve 83 having a frusto-conically shaped bore 84, the smaller diameter end of this bore being positioned as shown. Three arcuately shaped and circumferentially spaced insulators 85 surround the anode 83 in spaced relation therewith. For securing these insulators 85 to the anode 83, a thin metal ribbon 86 fits into a shallow circumferentially groove and binds the three insulators 85 against the anode 83.

Axial positional support for the anode 83 is provided by means of three ceramic, axially extending posts 87 which are fastened to the outer surface of the anode 83 at points spaced circumferentially equally part. One such post 87 is shown in FIG. 13. Considering this latter figure along with FIGS. 20 and 21, these three posts 87 are secured to a metallic mounting ring 88 which in turn is secured at its outer periphery to a cap assembly indicated by the numeral 89. This cap assembly 89 is composed of suitably machined metal parts formed to a substantially cylindrical shape of somewhat larger diameter than the sleeve 79, this cap assembly 89 being hermetically sealed to the sleeve 79.

As shown more clearly in FIGS. 19 and 21, three power supply terminals 90, 91 and 92 are mounted in the cap assembly 89 and are spaced circumferentially apart 120°. The terminal 90 has a wire conductor 93 internally thereof which passes coaxially through one of the supporting posts 87 and connects to the anode 83. The terminals 91 and 92 have suitably strong, bar-like conductors 94 and 95 which project into the central portion of the gun and pass through a block 96 of insulation for maintaining the spacing therebetween. At the innermost ends of these conductors 94 and 95 is connected a thermionic filament 97, this filament being in the form of a circular loop of wire located coaxially with respect to the axis of the ion gun. As shown in FIG. 11, this filament 97 is situated to the rear of the anode 83.

A pipe fitting 98 is also secured to the cap assembly 89 midway between the two filament terminals 91 and 92 and communicates with the inside of the ion gun as shown. By means of this pipe connection 98, fusion-reactive gas may be admitted to the interior of the ion guns from a suitable source 45 (FIG. 9) by the adjustment of series-connected hand valves 99.

The anode 83 is maintained insulated from the surrounding sleeve 79 such that a power supply 100 may be connected between this anode 83 and the sleeve 79 as well as the anode 20. In FIGS. 11 and 13, such a power supply 100 is indicated as being connected with the positive electrode to the terminal 90 and the negative electrode to ground and the housing 79 of the ion gun. This power supply 100 in an operating embodiment of this invention is designed to supply a voltage which is variable between the limits of 0 and 1,000. A suitable voltage connected between the two terminals 91 and 92 (FIG. 21) serves to energize the filament 97.

Each ion gun 22 in operation generates and emits a beam of ions which is directed toward the fusor center 26. In order to approach this center 26, the beam must penetrate the cathode structure 36, 21, and for this purpose a suitable aperture assembly is provided therein. This assembly is indicated in enlarged detail in FIGS. 16 and 17 and is shown in proper coaxial relationship with respect to the ion guns in FIG. 11. Each aperture assembly includes an annular insert 101 having smoothly radiused corners, which tightly conductively fits into a companion aperture 102 in the shell 36. The shell 21 is provided with an aperture 103 of the same diameter as aperture 102 in precise radial registry therewith.

The beam emitted from the respective ion gun 22 passes through the annulus 101 and the aperture 103. Each aperture 38 in the cathode shell 36 is covered with a biasing screen indicated by the numeral 104 (FIGS. 14 and

15), this biasing screen 104 being secured in position by means of a first mounting ring 107 secured to the outside of the shell 36 in registry with the opening 38 and another ring 106 which clamps the screen 104 to the ring 105. The aperture 37 in the shell 21 is of a size equal to that of the aperture 38. The two rings 105 and 106 are firmly, conductively secured to the shell 36 by means of a series of circumferentially spaced screws 107.

Except where otherwise stated, the structure of FIG. 11 is fabricated of metal and primarily of stainless steel. The cathode shell 21 should be of a material which is photoelectric, and two suitable materials having the necessary properties are stainless steel and molybdenum. Other material obviously may be used without departing from the spirit and scope of this invention.

Inasmuch as the dimensions of the structure of FIG. 11 are relatively small in comparison with the operating voltages which are used, it is of course necessary to take the usual precautions in connection with preventing field emission and corona discharge. Suitable corona shields in various places are used as explained in some detail hereinabove, and all sharp edges, corners, points and the like, from which field emission could occur, are avoided.

The ion guns 22 and 22*a* are disposed in pairs which are diametrically opposite each other. Additionally, the axes of these opposite pairs are set to be coincident such that the ion beams emitted thereby will occupy essentially the same pencil-like space. This space passes through the fusor center 26.

The operation of the device of FIG. 11 is essentially as has already been described hereinbefore in connection with FIG. 9. To start with, all operating voltages are turned "Off" and a suitable vacuum pump 44 is connected to the interior of the anode shell 20. Referring to FIG. 11, such a vacuum pump is connected to an outlet port 108 in the cap 109 which seals the end of pipe section 52. Thus, gas inside the anode 20 as well as the cathode shells 21 and 36 may be evacuated therefrom, through the pipe sections 50 and 52, through webs in the disc 57 and out of the port 108.

The pump 44 is allowed to operate for a sufficient period of time and the fusor is otherwise outgassed so as to remove contaminates which could interfere with operation.

After outgassing and evacuating the fusor to a pressure of from 10^{-9} to 10^{-10} millimeters of mercury, a suitable fusion-reactive gas, such as deuterium, from gas sources 45 and 45*a* (FIGS. 9 and 11), is admitted to the ion guns 22 and 22*a* by adjustment of the valves 99. For this purpose, these valves 99 are opened just sufficiently to admit small quantities of gas which will raise the pressure to from 10^{-6} to 10^{-4} millimeters of mercury. However, just prior to admitting this gas to the ion guns, operating voltages are applied to the fusor as previously explained. In this connection, in order to have complete control of the fusor operation, it is desirable that all voltages be applied progressively starting from zero voltage. Initially, the voltage is applied to the terminals 91 and 92 for heating the filaments 97 of all of the ion guns 22 and 22*a*. Next, the voltages of ion gun power supplies 100 are increased progressively starting with zero voltage until a sufficiently high difference of potential is applied between the anodes 83 and the respective filaments 97 so as to ionize the fusion reactive gas that has been admitted and to form ion beams. Next, the power supply 25 is steadily increased in voltage until fusion reactions are noted. The detection of neutron counts by usual methods can serve as such an indication. The power supplies 100 are adjusted until a satisfactory degree of ionization as well as suitable intense and compact ion beams are obtained which are directed toward the fusor center 26. Instrumental in the formation and acceleration of these ion beams is, of course, the potential applied to the anode 20 and the cathode 21, 36.

The vacuum pump 44 connected to the port 108 may be kept operating continuously so as to maintain the neutral gas pressure within the fusor to suitably low levels, and, additionally, the gas valves 99 may be opened slightly or may be opened and closed intermittently for the purpose of supplying necessary quantities of fuel gas to the fusor in order to obtain the reactions desired. By adjusting the high voltage supply 25, the reaction rate can be changed. Similarly, adjustment of the ion gun anode voltage can affect the reaction rate. Increasing and reducing, selectively, the quantity of neutral gas admitted to the fusor also affects the reaction rate. Therefore, desired reaction rates may be controlled by adjustment of the various circuit and system parameters. A gas pressure of about 10^{-5} mm. Hg continuously maintained has been found suitable for producing fusion reactions.

As shown in FIG. 12, the switch 47 which is connected in series with the high voltage supply 25 is closed, and the neutron-producing reactions just described are obtained. If it is desired to obtain an electrical output from the fusor, it is only necessary to open the switch 47 to place the resistor 46 in series with the high voltage supply. As the reaction rate increases within the fusor and reaches a suitably high value, a reverse current is developed in the circuit of the anode 20 in the direction of the arrow Z adjacent to the resistor 46. This results in the development of a voltaged drop across the resistor 46. This reverse current Z will develop when the voltage generated on the anode 20 exceeds that of the supply 25. Once this self-sustaining operation is achieved by the generation of this anode voltage, the high voltage supply 25 may be taken out of the circuit by first disconnecting the battery and then providing a conductive or essentially short circuit in place of the supply 25. However, such a circuit change is not necessary if the supply 25 is adapted to conduct current in the reverse direction.

In order to form the poissor in the center of the cathode 21, 36, it is, as previously explained, necessary to have both ions and electrons. These ions are supplied primarily by the ion guns 22 and 22*a*. The electrons are in part provided by photoemission from the internal surface of the cathode shell 21. The ions from the ion guns 22 and 22*a* establish the initial virtual anode as previously explained in the central volume of the cathode shell 21. The poissor thereupon forms, as previously explained. Electron flow from the inside of the cathode shell 21 to the anode 20 is inhibited by reason of the fact that the cathode shell 36 is biased negatively with respect to the shell 21. This bias is supplied by the power supply 43 and is adjusted to a value at which the external anode 20 current is reduced to a desired minimum. The screens 104 over the openings 38 in the outer cathode shell 36 provide the desired equal potential surface which is more negative than the potential on the cathode shell 21. This inhibits the escape of electrons from the inside of the shell 21, thereby reducing losses due to the interception of such electrons by the anode 20. Also, the screens 104 prevent the positive field from the anode 20 from entering the apertures 37 of the cathode shell 21 and upsetting the spherical configuration of the equal potential surface therein, as previously explained.

The various aperture pairs 37 and 38 are arranged in diametric opposition for radial ion traversal. A cathode shell 21 and 36 construction of about 70% openness provides adequate ion permeability.

Alternative working embodiment

In FIGS. 22 through 26 is illustrated an alternative working embodiment wherein like numerals indicate like parts. Basically, the difference between this embodiment and the one just described and shown in FIG. 11 resides in the use of a single cathode shell assembly 21 instead of the double-shell arrangement 21 and 36 of FIG. 11. The construction of the fusor of FIG. 23 to the extent that it differs from the arrangement of FIG. 11 will be described in detail in the following.

The cathode shell **21** of FIG. 23 is mounted in concentric relation with respect to the anode **20** by means of the pedestal assembly **49***a* which corresponds to the assembly 49 of the fusor of FIG. 11. This assembly **49***a* includes a stainless steel or the like supporting tube **50***a* secured at one end to the anode **20** as shown. An apertured disc **110** is perimetrically fastened inside the tube **50***a* in the position shown and has secured thereto a ceramic post **111** which projects coaxially through the tube **50***a* to a position just inside the anode **20**. The opposite end of the ceramic post **111** is secured to a mounting plate **112** suitably rigidly secured to the cathode shell **21**. All the parts just described are sufficiently strong and rigid to position securely the cathode **21** with respect to the anode **20**.

The left-hand end of the tube **50***a* of FIG. 23 is adapted to have a vacuum pump connected thereto the same as has already been described in connection with the preceding embodiments.

The high voltage terminal assembly **72***a* corresponds to the assembly 72 of FIG. 11. This assembly **72***a* comprises a conductive tubular assembly **71***a* capped at its right-hand end as shown in FIG. 23 with a bottle-like insulator **113** which is sealed in vacuum-tight relation to the tubular assembly **71***a*. Extending coaxially through the assembly **71***a* and the insulator **113** is a rod-like electrode **114** which corresponds to the tubular element **68** of FIGS. 11 and 12. The inner end **68***a* of this electrode **114** has rigid conductors **115** connected to a mounting plate **116** secured to the cathode **21** as shown. The opposite end **117** of the electrode **114** penetrates in vacuum-sealed relation the re-entrant end wall portion **118** of the insulator **113** to have connected thereto the high voltage leads shown diagrammatically in FIG. 23. The parts **71***a*, **113**, **114**, **115** and **116** are sufficiently strong and rigid as to assist in centering and securing the cathode **21** within the anode **20**. A suitable bellows structure **119** is incorporated within the assembly **71***a* along with a threaded micrometer adjustment **120** for tilting the assembly **71***a*, **113** slightly so as to position adjustably the cathode **21** within the anode **20**. The total support for the cathode is provided by the two assemblies **49***a* and **71***a*.

The cathode **21** itself may be made identical to the cathode **21** of FIG. 11, apertures **37** and **103** being provided therein. The biasing function of the outer shell **36** of FIG. 11, however, is performed by a biasing screen assembly shown in enlarged detail in FIG. 24 and identified generally by numeral **121**. This assembly includes an annular spacer **122** of stainless steel or the like which has secured to the underside thereof a fine mesh, conductive screen **123** which full covers the aperture **37**. Concentrically positioned on the peripheral portion of the screen **123** is an annular insulator **124** L-shaped in cross-section as shown. Seated on the insulator **124** is another spacer **125** of some suitable conductor such as stainless steel which has secured to the upper side thereof a second screen **126** like the other screen **123**. The screen **126** is arranged parallel to but is spaced from the screen **123** for performing the purposes previously explained in connection with the screens of FIGS. 14 and 15 which are attached to the outer cathode shell **36**.

Fitted over the spacer **125** is another annular insulator **127** of L-shaped cross-section which engages the main spacer **122** for holding the second spacer **125** in concentric position.

Superposed on and in engagement with the spacer **122** and the insulator **127** is a flat annular retaining ring **128** which clamps the parts **124**, **125** and **127** into assembly. An annular cap **129** fits over the retaining ring **128** as shown, and a plurality of circumferentially spaced screws **130** passing through the parts **129**, **128**, **122** clamp the latter together and secure the same to the cathode **21**.

A sleeve **131** of insulating material is radially secured within a companion aperture in the main spacer **122** as shown and receives therethrough a conductive connecting wire or lead **132**. This lead at its inner end is conductively connected to the spacer **125** by being clamped between the spacer **125** and the insulator **124** as shown. Each of the apertures **37** in the cathode **21** is provided with a biasing screen assembly **121** as just described and the leads **132** of all of these assemblies are connected together so that the biasing potentials applied to all of the screens **126** will be the same.

As in the fusor of FIG. 11, each of the cathode apertures **103** of the fusor of FIG. 23 is coaxially aligned with the beam axis of the respective ion guns **22** and **22***a*. However, in each of the apertures **103** of the fusor of FIG. 23 is provided a biasing-electrode assembly shown in enlarged detail in FIG. 25. This assembly includes a tubular sleeve **133** of ceramic having an enlarged annular projection **134** thereon. The sleeve **133** intimately fits into the cathode opening **103** with the projection **134** seating against the outer cathode surface. A tubular metal insert or electrode **135** is snugly received by the sleeve **133** as shown and is provided with an end portion **136** which engages the bottom end of the sleeve **133** as shown. The upper end of the insert **135** is flared outwardly into a smoothly radiused flange **137** which overlaps the upper end of the sleeve **133**. A suitable wire-like conductor **138** is wrapped around the insert **135** between the flange **137** and the upper end of the sleeve **133**. This wire **138** is of such diameter as to have a wedging fit between the flange **137** and the sleeve **133** thereby locking the electrode **135** to the sleeve **133**. The assembly described thus far is finally secured to the cathode shell **21** by means of a clamping ring **139** fitted over the projection **134** and welded to the shell **21**.

Either as a part of the wire **138** or otherwise secured thereto is a connecting lead **140** for applying a biasing potential to the electrode **135**. As will now be apparent, this electrode **135** is insulated from the cathode **21**.

All of the ion-aperture assemblies as shown in FIG. 25 are interconnected by leads **140**. All of these leads **140** are interconnected with the leads **132** of the biasing screen assemblies **121** so that the electrodes **135** will have the same biasing potential applied thereto as the various screen **126**.

Biasing potential is connected to the biasing electrodes **135** (FIGS. 24 and 25) by means of the assembly shown in FIG. 26. This assembly includes an evacuated tubular housing, indicated generally by the numeral **141**, composed of tubular housing **142** of metal hermetically sealed at the inner end to anode **20** and at the outer end to a glass insulator **143** also of tubular configuration. Coaxially within the tubular housing **141** is a center conductor assembly composed of a metal tube **144** which is hermetically sealed at its outer end with a metallic plug **145**. A tubular conductor **146** is also sealed at one end to the plug **145** and receives on the other end a sleeve-like terminal **147** hermetically sealed at **148** to the conductor **146** and at the end **149** to the outer end of the insulator **143**. The space between the outer wall **142**, **143** and the conductor assembly **144**, **145**, **146** is vacuum-tight and in communication with the fusor interior. A metal cap **150** conductively carrying a sleeve-like contactor **151** telescopes over and conductively engages the terminal **147**. The cap **150** is molded into a protective rubber cover **152** which carries a conductive pin **153** connected at one end to the cap **150** and projects outwardly to a point at which a connection may be made to an external biasing circuit. This biasing circuit is indicated as being in the form of a battery **154** having its positive terminal connected to the cathode shell **21** and the negative terminal to the pin **153**.

A loop of spring wire indicated by the numeral **155** conductively engages the peripheral portions of one of the screens **126** and is formed by convoluting one end portion of a wire **156** which extends through the conductive tube **144**. This wire **156** is of such length that the upper end thereof engages the conductive plug **145** and makes conductive contact therewith. The wire **156** is maintained in position within the tube **144** by means of a small metal ring **157** fitted into the end of the tube **144**. It will now be seen that by providing one of these assemblies **141** on the fusor of FIG. 23, biasing potential may be applied to all of the screens **126** and ion-aperture electrodes **135**.

Referring to the ion guns **22** and **22***a* of the fusor of FIG. 23, it should be noted that annular magnets **158** are provided thereon which encircle the sleeve elements **79**. These magnets **158** have one pole adjacent to the anode **20** and the other pole outwardly therefrom. They are positioned with respect to the anode **83** (FIG. 13) such as to lengthen the travel of the electrons emitted by the cathode **97** axially of the gun thereby increasing the probability of ionization by each electron. This provides an increase in ion generation over the electrostatic arrangement shown in FIG. 13. In all other respects, the ion guns **22** and **22***a* of FIG. 23 are like that shown and described in connection with FIG. 13.

The operation of a fusor of FIG. 23 is the same as that of the fusor of FIG. 11 in almost all respects. The same techniques and operating voltages are applied, and all of the voltages are made variable so as to provide convenient control of fusor operation. As in the case of the fusor of FIG. 11, the bias on the screens **126** and the electrodes **135** and supplied by the battery **154**, for example, is adjusted to a value at which the external anode circuit current is reduced to a desired minimum. The screens **126** over the openings **37** in the cathode shell **21** provide the desired equipotential surface which is more negative than the potential on the cathode shell **21** as well as the inner screens **123**. This inhibits the escape of electrons from the inside of the shell **21** thereby reducing losses due to the interception of such electrons by the anode **20**. Also, the screens **126** prevent the positive field from the anode **20** from entering the apertures **37** and disturbing the spherical configuration of the poissor therein, as explained previously.

The electrodes **135** form the same biasing function, these being more negative than the cathode shell **21**, thereby setting up equipotential surfaces which inhibit the outward flow of electrons from the interior of the cathode shell **21**.

Alternative ion gun

In FIGS. 27 through 29 is illustrated an alternative ion gun **22***b* which may be substituted for the guns **22** and **22***a* of the preceding embodiments. Comparing this gun **22***b* with the one of FIG. 13, like numerals will indicate like parts. Instead of providing a cap such as **89** in FIG. 13 for sealing off the end of the gun, a ring flange **89***a*, like the flange **87**, is used for mounting one end of a tubular anode extension, indicated by the numeral **159**. The outer end of this extension **159** is hermetically sealed by means of a cap structure **160**. Thus, it will be noted that the body or envelope of the gun **22***b* which includes the parts **79** and **159** is considerably longer than the bodies of the guns **22** and **22***a*.

Carried by the wire conductor **93** of the terminal **90** is a tubular electrode **161** which corresponds to some extent to a control grid in an ordinary vacuum tube. This sleeve **161** is secured to the conductor **93** and additionally has a conductive connection **162** to one of the filament leads **94**. The filament **97** as well as the electrode **161** are biased negatively by means of a battery **163**, the negative terminal being connected to the conductor **93** and the positive terminal to the body of the gun **22***b*. This body **79**, **159** constitutes a conductive extension of the anode **20** such that it is operated at the same potential as the anode **20**. Surrounding the anode extension **159** is a series of bar magnets **164** arranged with the longitudinal axis thereof parallel to the axis **165** of the gun **22***b*. Supporting these bar magnets **164** are two axially spaced mounting rings **166** and **167**, these rings being made of iron or some equally suitable, highly magnetic material. A band **167***a* clamps these magnets onto rings **166** and **167** and retains them in assembly. There are six such magnets used in the arrangement of FIG. 27 as shown more clearly in FIG. 28; however, either a greater or fewer number may be used without departing from the scope of this invention.

In operating the guns **22***b*, the same conditions of fusor operation as explained previously are observed. More specifically, with respect to the gun **22***b*, a suitable filament voltage is applied to the filament **97** for the generation of ionizing electrons. Since the filament **97** is substantially contained within the length of the electrode **161**, the electrons in order to travel to the anode **159** must move axially of the gun. By reason of the focusing influence of the magnets **164**, the electrons will be directed rightwardly as viewed in FIG. 27 in a spiral path until they are eventually collected by the right-hand end portion of the anode **159**. Fusion-reactive gas admitted to the gun via the fitting **98** will be ionized by these electrons, such ions thereupon coming under the influence of the more negative electrode **161** as well as the fusor cathode **21**. The ions eventually are accelerated through the aperture **82** and form into beams which travel paths intersecting the tube center **26**.

Higher ion currents may be obtained from the ion guns **22***b* than those of FIG. 11 inasmuch as the ionizing electrons emitted from the filament **97** (FIG. 27) are accelerated through longer paths before interception by the anode; hence, they have a greater probability of producing ions before being intercepted.

In the following are given dimensions and values for system and circuit parameters for the embodiment of this invention as shown in FIG. 23, these being given by way of illustration only and not by way of limitation.

Dimensions

Internal diameter of anode shell **20**	inches	6
Thickness of anode shell **20**	do	0.093
Anode **20** material		[1] 304
Internal diameter of cathode shell **21**	inches	4.38
Thickness of shell **21**	do	0.060
Cathode **21** material		[1] 304
Size of openings **37**	inches diameter	1.75
Size of openings **103**	do	0.625
Inner diameter of rings **124**, **125**, **127**	inches	1.500
Size of screens **123** and **126**	mesh	[2] 15
Spacing between screens **123** and **126**	inch	0.140
Overall height of assembly **121** radially beyond cathode **21**	inch	0.320
Inner diameter of electrode **135**	do	0.437
Length of electrode **135**	do	0.500
Distance electrode **135** projects radially outwardly from cathode **21**	inch	0.312
Size of opening **80**	do	1.125
Outer diameter of sleeve **79**	do	1.500
Inner diameter of aperture **82**	do	0.500
Internal axial length of gun **22**, **22***a*	do	2.875
Length of anode **84**	do	1.000
Minimum inner diameter of anode **84**	do	0.625
Maximum inner diameter of anode **84**	do	0.825
Spacing of filament **97** from the adjacent end of anode **84**	inch	0.125
Space between anode **84** and sleeve **79**	do	0.187
Thickness of aperture ring **81**	do	0.375
Distance between ring **81** and anode **84**	do	0.500

[1] Stainless steel.
[2] Woven tungsten.

For a neutron count of 10^9, the following operating parameters for the fusor of FIG. 23 may be used as an approximation. Adjustments of voltages and dimensions may be required for obtaining this precise count.

Fusion-reactive gas	Deuterium and tritium, equal portions.
Neutral gas pressure in fusor	2 microns.
Anode **20** circuit current	30 milliamperes.
Anode **20** applied voltage	100 kilvolts.
Voltage of supply **100** as applied to the anode **84**	250 volts.
Voltage applied to the filament **97**	6 volts.
Filament **97** current	7 amperes.
Bias voltage on screens **126**	500 volts.

A viewing port assembly **168** is attached to the anode **20** as shown in FIG. 27 for the purpose of permitting observation, either directly or indirectly by means of closed circuit television equipment, of the reaction which occurs in the center, the outer end of the assembly **168** being vacuum sealed by reason of a window **169**.

While there have been described above the principles of this invention in connection with specific apparatus, it is to be clearly understood that this description is made only by way of example and not as a limitation to the scope of the invention.

What is claimed is:

1. In a device for producing nuclear reactions, anode means, cathode means inside said anode means, said cathode means defining a volume centrally located with respect to both said anode and cathode means, said volume being free of tangible structure, said cathode means being open to the flow of gaseous particles therethrough, a fusion-reactive gas within said volume, means supplying a potential between said cathode and anode means for establishing an electric field therebetween, means including said cathode and anode means and said potential-supplying means for accelerating ions of said gas to fusion-reacting energies along converging radial paths in said volume, and means for inhibiting flow of electrons from said volume to said anode means, said inhibiting means including means for biasing the open portions of cathode means more negatively than the cathode means itself.

2. The device of claim **1** wherein said volume is substantially spherical in shape and said paths have a common intersection in the central portion of said volume, and further including means maintaining said gas at sub-atmospheric pressure.

3. The device of claim **1**, said cathode means being a generally spherically shaped metallic shell having a plurality of openings therein.

4. The device of claim **3** wherein said anode means is a generally spherically shaped metallic shell concentrically surrounding said cathode shell, said inhibiting means including a plurality of biasing electrodes, one electrode for each opening, said biasing electrodes having portions peripherally coextensive with said openings, respectively, and means for applying biasing potential to said biasing electrodes for establishing said equipotential surfaces.

5. The device of claim **4**, including means connected to said anode shell for generating and injecting charge-particles into the space between said cathode and anode means, the magnitude of said electric field being adequate to direct at least a portion of said charge-particles through at least a portion of said openings and into said volume.

6. The device of claim **4**, including a plurality of ion guns carried by said anode shell in radial alignment with respective ones of said openings, means including said ion guns for directing ions through said openings and into said volume, whereby ions from said guns may be projected toward the opposite side of said anode shell.

7. The device of claim **6** wherein the first openings diametrically aligned with said ion guns, respectively, are free of structure and the remaining second openings are covered with biasing screens, respectively, means insulating said screens from the cathode shell next adjacent to said volume, the biasing electrodes for said first openings including annular elements coaxially fitted into the respective cathode shell openings, said annular elements being insulated from said cathode shell.

8. The device of claim **7** wherein each of said second openings is covered with two screens, the first of said screens being conductively mounted on said cathode shell, each second opening further having an annular insulating device coaxially surrounding the same and operatively mounted on the outer side of said cathode shell, each insulating device having a second of said screens mounted thereon spaced from the first screen, and connections to said second screens for biasing the same negatively with respect to said cathode shell.

9. The device of claim **6**, wherein the means for maintaining said gas at sub-atmospheric pressure includes a supply of neutral fusion-reactive gas connected to said ion guns and also means for controlling the pressure of said gas as admitted to said ion guns in the range of from 10^{-3} to 10^{-8} millimeters of mercury, said ion guns and said anode shell being hermetically sealed and in communication with each other, said controlling means including a vacuum pump for exhausting gas from within the space of said anode shell.

10. The device of claim **2**, said anode means incudling a generally spherical anode shell, said cathode means including two generally spherical concentric cathode shells spaced apart a predetermined distance, said two cathode shells having a plurality of openings therein which are in radial registry, said cathode shells being concentric with respect to said anode shell, means for biasing the outer one of said cathode shells more negative than the inner one.

11. The device of claim **10**, including means connected to said anode shell for generating and injecting charge-particles into the space between said cathode and anode shells, the magnitude of said electric field being adequate to direct at least a portion of said charge-particles through at least a portion of said openings and into said volume.

12. The device of claim **11**, including a plurality of ion guns carried by said anode shell in radial alignment with respective ones of said openings, means including said ion guns for directing ions through said openings and into said volume, whereby ions from said guns may be projected toward the opposite side of said anode shell.

13. The device of claim **12**, wherein the first openings diametrically are free of structure and the remaining second openings of the outer cathode shell are covered with biasing screens, respectively, said screens being conductively secured to said outer cathode shell, said means for maintaining said gas at sub-atmospheric pressure includes a supply of neutral fusion-reactive gas connected to said ion guns and also means for controlling the pressure of said gas admitted to said guns in the range of from 10^{-3} to 10^{-8} millimeters of mercury, said ion guns and said anode shell being hermetically sealed and in communication with each other, said controlling means including a vacuum pump for exhausting gas from within the space of said anode shell.

14. The device of claim **13** including means connected to said anode means for coupling electrical power therefrom.

15. In a device for producing nuclear reactions comprising anode means, cathode means inside said anode means, said cathode means defining a volume free of tangible structure, a fusion-reactive gas within said volume, said cathode means having openings for the traversal therethrough of charge-particles, means including said anode and cathode means and ions of said gas forming a space charge in the central portion of said volume, said space charge including electrons and ions of said gas so distributed as to produce a substantially spherical electric field which confines target ions in a point-like region in said central portion and oscillates projectile ions through said region at nuclear reacting energies, whereby fusion reactions result from collisions between said projectile ions and between said projectile ions and target ions, and means for biasing the openings in said cathode means negatively with respect to said cathode means for inhibiting the flow of electrons from the volume inside said cathode means to said anode means.

16. The device of claim **15** wherein said space charge is arranged in a configuration which provides a spherical virtual anode concentrically surrounding a spherical virtual cathode, said cathode means including a generally spherical shell concentrically surrounding said virtual anode thereby providing an electric field therebetween which exerts a force outwardly on a positively charged particle in a direction radially toward said cathode means, the potential of said virtual anode being of a magnitude which propels ions inwardly and oscillates them as projectiles through said virtual cathode at nuclear-reacting energies, whereby collisions of ions in the region of said virtual cathode produces nuclear reactions.

17. The device of claim **15** wherein said gas is of a composition that reactant products of fusion include positively charged particles which are propelled outwardly against the field of said anode means with sufficient energy to transfer the charges thereof to said anode means thereby generating electrical power, and circuit means connected to said anode means for coupling the generated electrical power therefrom to a utilization device.

18. The device of claim **17** wherein said anode means includes a generally spherical conductive anode shell, said cathode means includes a generally spherical conductive cathode shell concentrically positioned inside said anode shell, said cathode shell having said openings therein, said positively charged particles passing through said openings in transit from said volume toward said anode shell, and said circuit means including a load impedance connected in series between said anode and cathode means whereby said generated electrical power may be coupled from said impedance to a utilization device.

19. The device of claim **17** wherein said space charge-forming means includes a power supply connected to said anode and cathode means for applying an operating voltage thereto; the value of said operating voltage, the composition of said gas and the number of gaseous particles within said volume being so related that the fusion reactions are self-sustaining and contribute more power to said circuit means than is consumed in initiating such reactions.

References Cited

UNITED STATES PATENTS

3,258,402 6/1966 Farnsworth ----------- 176—1

REUBEN EPSTEIN, *Primary Examiner.*

rexresearch.com

G. Patrick FLANAGAN

Neurophone

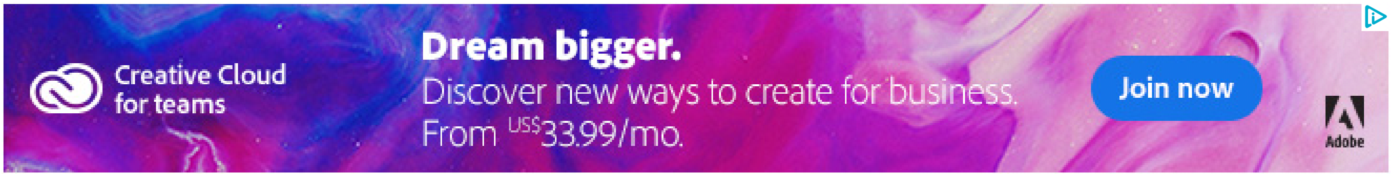
Creative Cloud for teams
Dream bigger.
Discover new ways to create for business.
From US$33.99/mo.
Join now
Adobe

Become a data scientist, without quitting your job.
THINKFUL

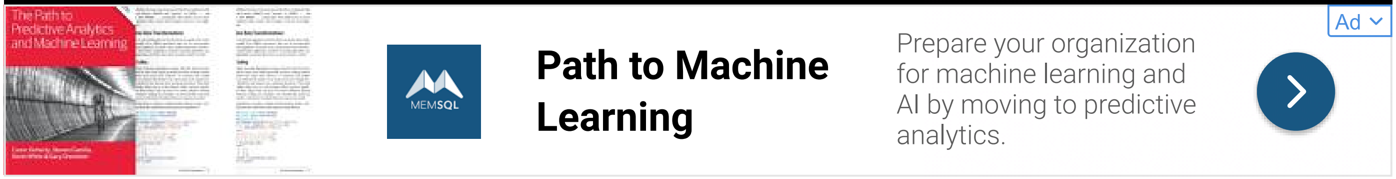
Ad
MEMSQL
Path to Machine Learning
Prepare your organization for machine learning and AI by moving to predictive analytics.

G. Harry Stine*Analog Magazine* (Feb 1980): "Bio Cybernetics Revisited"
G. H. Stine: *Analog Magazine* (July 1979): "Bio Cybernetics II"
Sheldon C. Deal: Meridian Balancing by Use of the Neurophone
G. Patrick Flanagan: "Operation of the Neurophone on the Nervous System"
G. P. Flanagan: The Neurophone Mark XI Manual (1979)
G. P. Flanagan: "History of the Neurowave Signal Development"
William Moeser: *Life Magazine* (14 September 1962) ~ "Whiz Kid, Hands Down"
The Neurophone Model 50 ~ The Thinkman
Anonymous: "The Pacinian Corpuscle"
Anon.: "Neurophone Breakthrough"

T.E. Bearden: *Excalibur Briefing* (Excerpt)
G. P. Flanagan: *Pyramid Power* (Excerpt, Chapter 3)
Eddy Taylor: "Patrick Flanagan's Neurophone --- Hope for the Deaf and Superlearning for All"
G. P. Flanagan: US Patent # 3,393,279 ~ "Nervous System Excitation Device"
G.P. Flanagan: US Patent # 3,647,970 ~ "Method & Apparatus for Simplifying Speech Waveforms"
Tools for Wellness Ad
Neurophone.com Ad
Rick Andersen: "Another Look at the Neurophone"

***Analog Magazine* (Feb 1980) ~**

Bio Cybernetics Revisited
by
G. Harry Stine

When Jerry Pournelli and I started to write these columns as a vis-a-vis experiment, we hoped that sooner or later one of us would write something that the other could tee off on and thus get a good controversy going that would present "alternate views." It hasent worked out that way because Jerry and I have discovered that we think very much alike on most matters. We have not been able to stir up controversy between one another. But we did stir up controversy among the readers.

When I wrote the columns on biocybernetics in the May and July 1979 issues, my objective was more or less than science reporting with some synthesis included. I wanted to point out that a true interface between human brains and electronic computers was not only possible but that the basic experiments have been done and the foundation technologies already existed, albeit in the case of the Flanagan neurophone the data had been generally unknown for 16 years. I reported on the basis of what I knew at that time. I requested that no one write for additional data because I didnt have any additional data at that time. I do now.

These two columns have resulted in an inordinate amount of mail from readers which makes the editor happy because he knows people are reading this. Some of the letters were anticipated, some were not. Some were from people who were deaf and wanted the Flanagan neurophone at any price. Others wanted the neurophone to use in attempts to contact loved family members in a coma. There were the usual letters from Ph.Ds who should know better than to make some of the statements they did. And I got letters with real up to date information...

You may obtain copies of his two neurophone patents by writing to the U.S. Patent Office and ordering Patent No. 3,393,279 granted July 16, 1968 and Patent No. 3,647,970 granted March 7, 1972.

The rights to Patent No. 3,393,279 are owned by Intelectron, Inc. of New York City. The FDA will not permit it to be marketed because it uses a low frequency r-f carrier, and they are very sensitive these days about the biological effects of r-f raditation.

But Pat Flanagan is an inventor. With his original patent rights in the hands of another company and with an FDA ban on its production, Flanagan developed the second neurophone device covered by Patent No. 3,647,970 which does not use an r-f carrier and does not involve high voltages. At the behest of the Department of Defense, the Department of Commerce placed Flanagan's new device under secrecy order No. 756,124 dated August 28, 1968. Flanagan was able to get this secrecy order rescinded in 1972 to permit the patent to be issued.

I sat in Flanagan's study and listened to the new neurophone on July 26, 1979. It works even better than the original one. It operates from eight Type AA alkaline penlight cells. The audio signal is double differentiated and converted to a 60 volt square wave which is then run through a zero crossing detector. The sensor electrodes are one inch diameter plates made from lead zirconium titanate insulated on one side with epoxy that also seals the attachment of the lead from the neurophone.

With one electrode placed on the soft skin of my calf and the other on my chest, I heard the audio information from the tape recorder input in my head. I listened to both speech and music, and the fidelity was outstanding. I had a cassette tape recorder running two feet from me while this was going on; there is no evidence of the audio information recorded on the cassette by the recorder; I was the only entity hearing the neurophone, and it was therefore not producing and sound waves that the microphone of the recorder could pick up.

As I knew 17 years ago, the Flanagan neurophone works. I dont give a doodly damn what all you bloody experts out there have presumptuously stated in your outraged fan letters to me. I am forced to repeat the immortal words of Dr. J. C. Warren voiced after performing the first surgery with a patient under ether anesthetic administrated by Dr. William Thomas Green Morton in Boston on October 16, 1846:

"Gentlemen, this is no humbug!"

For the benefit of all you who wrote me, I gave your letters to Flanagan. There is no need for me to be in the loop. There is no need for Flanagan to defend himself nor does he need help in doing so; he has a device that works. In the eyes of some sober, respected academicians, Flanagan is perhaps flamboyant and different, but that must not be allowed to detract from the fact that he is a successful inventor and, in common with many others of his unconventional and freewheeling nature in the history of science and technology, working on the fringes of scientific research in areas that others can't or won't touch.

The most fascinating aspect of this whole affair has been the confirmation and terrible realization that the human race apparently has not progressed beyond the state of affairs a century ago when "legitimate" scientist dug in to defend their sacred turf....and made damned fools of themselves in the process. They also retarded the development of technology that would have relieved human suffering and generally improved the human condition. These scientific brahmins still exist; they are no different from the predecessors of a hundred years ago; and they exhibit an appalling lack of knowledge and understanding of the history of their own field of expertise!

Arthur C. Clarke's law is still very much in force.

Jerry Pournelle was 100% correct in his June 1979 "The Alternate View."

I would like to extend my profound thanks and grateful appreciation to those scientific brahmnis who wrote to me and to the editor about the neurophone and the human computer interface. You have greatly contributed to the contents of my "Utter Bilge" file. This is my collection of presumptuous statements from people who should know better than to make such statements in the first place. The file title derived from in infamous statement of Sir Richard Van Der Riet Wolley: "Space travel is utter bilge!"

Scientific research and technical progress is not and has never has been the exclusive turf of any select group of people. There has always been a role for the amateur investigator, and still exists today. Before the "official" establishment of any new area of scientific endeavor, everyone who works in that area must, by definition, be an amateur. Even in the established fields of scientific endeavor, important discoveries have been made by amateur investigators. Good scientists and I have the pleasure of knowing many would never reject out of hand any amateur findings and will react with enthusiasm and interest to such findings. It is only poor scientists who are insecure that will react negatively. Would there were more good scientists!

Bioelectronics today is about the same situation as electricity about a century ago. The field is ripe for experimentation. Certainly there is danger; there always is. But that is no reason for suppressing information, as one PhD. correspondent demanded be done, for fear of someone electrocuting himself by building a neurophone in his basement shop. People still get hurt in steam boiler explosions. As a matter of fact, people still get burned handling the ancient technology of a campfire. These accidents merely strengthen the statement of Herbert George Wells, "History is a race between education and catastrophe," but down on the personal level.

To paraphrase Harold Laski, science by experts means, after a time, science in the interest of experts.....

Analog Magazine **(July 1979) ~**

Bio Cybernetics II

by
G. Harry Stine

Last time around, two issues ago, I reported on the process made by DARPA (Defense Advanced Research Projects Agency) concerning computer recognition of human EEG patterns. I stated that this was a major breakthrough in the development of the true intelligence amplifier wherein the human nervous system was linked directly with the electronic circuitry of a large general purpose computer. Soon, the crystalline circuitry of the computer will be able to "understand" the colloidal circuitry of the human nervous system by direct linkage through electromagnetic fields.

And I left you hanging by your fingernails on a figurative cliff overhanging an ammonia sea swarming with methane monsters, so to speak.

I stated that seventeen years ago the breakthrough had been made that would permit the colloidal circuitry of the nervous system to "understand" the crystalline circuitry of the electronic computer. In other words, if the computer signal was in the proper form, the human nervous system could pick up and decipher the electronic computer signal through an electromagnetic field.

In early 1962, I was working as assistant director of research for the Huyck Corporation in Milford, Connecticut. Dr. William O. Davis was the director of research, and he had been charged by the company to look for new products to insure that Huyck would be in business twenty five years hence. There were no caveats placed on us at that time. We could look at or for anything....and we did. Then as now I have all sorts of pipelines into the back rooms of various research and development organizations here and abroad; a science fact or science fiction writer really cant function properly without these contacts. My most valuable Little Black Spy was the former editor of this magazine. John W. Campbell, who in turn had an incredible network of Little Black Spies.

Campbell told me of an 18 year old boy in Texas who had invented a spectacular new hearing aid. I made a telephone call, talked to the boy, and was on a plane to Houston within days.

G. Patrick Flanagan of Ballaire Texas was a boy genius. In addition to being an outstanding gymnast and a pilot, this high school graduate had stumbled upon a technique of introducing audio information directly into the human nervous system without loudspeakers or earphones, and without direct electrical connection between his gadget and the nervous system.

On July 24, 1962 in Bellaire, Texas in Flanagan's shop in the attic of his parents home, I personally witnessed proof-of-principle demonstration of a direct linkage between a crystalline electronic circuit and the colloidal system of the human nervous system.

The device called the "neurophone" by Flanagan, was a very simple gadget. It used those ancient devices known as vacuum tubes. It used a 6L6G tube running as a 35 Kilohertz oscillator; the output of the oscillator was amplitude-modulated by the output of ant single channel hi-fi system through suitable impedance matching transformers, etc. This amplitude modulated 35 kilohertz signal then went through a cheapie stepup transformer so that the output was very high voltage (about 4 kilovolts, as I remember) but at very high impedance. Flanagan fed this signal through a length of ordinary TV antenna twin lead to a pair of rubber pads about six inches in diameter that he had filched from a "relaxicisor" muscle relaxing device. Each pad consisted of a 1/16 inch thick rubber sheet, a piece of copper window screen to which was soldered one of the wires of the twin lead, and the rubber covering of the relaxicisor pad. The rubber insulation kept one from getting zapped, although the shock wasnt painful....just annoying.

Put a signal into the neurophone input, peak the frequency, peak the modulation to 100%, and then hold the two pads to your body....anywhere.

And you heard music playing in your head!

Although the pads worked better and you got a louder signal with the pads on your bare skin, it would work through a layer of clothing.

And it worked regardless of where you put the two pads! You could put one on your shoulder and the other on the sole of your foot...and you would still hear whatever program material was being played into the neurophone form the hi-fi system.

I can freely describe this today because (a) I dont think Flanagan's patent application ever got accepted and I dont believe the patent ever issued. (b) Huyck Corporation is off doing its classic thing of making paper machine parts and isnt interested in the slightest, and (c) I think the time is just about right to get going on this little gadget again.

Flanagan had succeeded where others had failed. Dr Henry Puharich had succeeded in developing a similar device, but the pads were one inch copper plates that had to be rubbed along the skin to achieve the "fricative effect" to permit Puharich's gadget to work.

In a series of long and complex experiments conducted under my supervision at Huyck and by Dr Wayne Batteau at Tuffs, it was conclusively shown that the neurophone effect not the result of (a) the pads acting as electrostatic loudspeakers, (b) the pads exciting bone conduction of the sound, or (c) the pads activating the eighth cranial nerve. Although I tried the neurophone an a nerve deaf medical doctor at Columbia University in 1965, I kept the experiment running for only about ten minutes. Dr Wayne Batteau tried it at Tufts and succeed because the brain of his nerve deaf subject had "forgotten how to hear" and needed something over an hour of application of the neurophone signal in order to begin to hear in his brain again! We know now that the Flanagan neurophone operated by direct linkage of the electronic circuitry with the nervous system through electromagnetic fields. We know now that the nervous system will pick up any signal and send it to the brain where the brain recognizes the signal according to the sensory data it represents, then switches the signal to the proper sector of the brain responsible for processing such signals. But we dont know how Flanagan accidentally happened to crack the neural code for audio data.

I can assure you that the Flanagan neurophone is no hoax. Many responsible people experienced it. Just before his death, Dr William O. Davis gave me one of the neurophones we had built at Huyck; I still have it, but a two hour search through my uncatalogued archives of scientific memorabilia and trivia failed to locate it this morning. Im going to dig it out and try it again for fun. And because it does represent a breakthrough that will permit electronic computers to communicate directly to our nervous systems...by putting on a beanie, and not plugging a cable into a socket implanted on the top of our heads!

The DARPA work with computers deciphering human EEG signals, plus the Flanagan neurophone capable of permitting a computer to communicate directly with a human, both make inevitable the final computer; the use of the computer as a extension to the human brain as an intelligence amplifier for the human being.

Now, what happened with the Flanagan neurophone, and why havent you heard more about it? I am not sure that it is because of the reaction of Big Science as discussed by Jerry Pournelle last month. There were several factors involved.

First of all, it didnt look like Huyck could get a patent position, so they wanted out. They also discovered that the Huyck Corporate Research Laboratories were producing more inventions, patents, and potential new products than the company could possibly absorb. Huyck also got into production trouble with several products that have come along before Corporate Research entered the picture. So Huyck dropped the entire corporate research program, writing off about 2.5 million spent over a five year period. They also convinced themselves that they could not afford to develop the marketing know-how for any of the new products we were coming up with. They succumbed, as have so many other small companies, to the "Harvard Business School Syndrome": Dont develop new technology and products yourself, but wait until somebody else does and then buy the whole schmear, product, production, and marketing know-how in a big package. My employment agreement was written in a way that I wasnt sure what I could talk about....then.

Dr. Wayne Batteau suffered a fatal heart attack in Hawaii while diving with dolphins. He had also done considerable work in that area.

I dont know where Flanagan is. He worked for NASA in Houston for a time in computer programming. The last I knew, he was somewhere in California as one of the top experts pyramidology. I suspect I will hear from him as a result of this, and Ill welcome it.

Big Science did not kill the neurophone. It was complacent, and it really didnt know anything about the device. Academia was not really consulted. The neurophone has been dormant for fourteen years now because of the failure of nerve and imagination in the sales and marketing department.

But, now, there appears to be a need for the neurophone to take the final step toward direct communication between ourselves and our crystalline symbiotes/servants, the electronic computers.

Yes, I am still interested in this area. I am interested in anything and everything. But do not-repeat, repeat do not write to me or telephone me and ask for free information. I am a writer and a consultant, and my business is information...and the only thing of real value that you or I have in this world is time. I didnt write this soliciting business; I wrote it to report to you that the intelligence amplifier is just around the corner and awaiting you bright people out there to put it together right so that humans stay in control of intelligence amplifiers.

Now, what can we do in the future to integrate the creative process of "invention" more fully into the process of research, development, production, and marketing? Hang in there......

KeelyNet BBS (15 October 1993): Dr-Deal.ASC ~
This File is shared with Keelynet courtesy of Terry Bastian

Meridian Balancing by Use of the Neurophone

Sheldon C. Deal, D.C., N.D.

Abstract ~

A series of experements were done by muscle testing paitents for meridian imbalences and then strengthing the muscle indicators, indicating balanced meridians, by running a time-domain amplifies signal through the patient's skin.

The skin is embryonically the source of all special senses. Our skin contains more sensors for heat, touch, pain, etc., than any other part of the human anatomy. The human ear evolves embryonically out of the convolutions of the skin of an embryo in the mothers uterus. Basically, the skin is the oldest evolutonary nervous system sensor. Another way to look at this, is the special sense of tough has progressively extended over the entire body by way of skin and evolutionary wise our other special senses should follow suit and extend over our entire body in time. The skin then, since it is the precursor of the ears, should also be capable of hearing, or to put it another way, it should beable to transmit the sense of hearing to the brain by means other than the 8th Crainial Nerve.

The skin is both piezo-electric and opto-electric, that is, when the skin is stimulated by an electric or photon field, it vibrates. If it is mechanically stimulated, it will generate its own electric and photon field. In Russia, blind people have been trained to "see" with their fingertips and in Czechoslavakia, deaf people have been trained to "hear" with their fingertips.

For the purpose of our experement, we used a Neurophone, a device developed by Dr Pat Flanagan, author of the book, *Pyramid Power*. The device uses hyperspacial nested modulation technology. Briefly the device takes a complex signal such as the sound of an orchestra and electriclly processes it. First, the signal is passed into a section which clips everything into a series of square waves, remarkably analogous to the sort of clipped waves Lisitsyn confirms are the carriers of the human brain waves. Next, the square waves are differentiated, since these are finite spikes with real non-zero rise times and decay times rather than theoretical constructs, a series of noisy spikes results the second differentiator section. From here, the noisy spikes are introduced to special contact electrodes which are placed on the skin. The electrodes are fabricated of zirconium titanate imbedded in acrylic plastic tiles to provide maximum impedance match to the skin itself.

In experements done at Tufts University near Boston by Dr Dwight Batteau in Man-Dolphin communications and assisted by Dr Flanagan, it was established that the nervous system uses time ratios as major sources of intelligent information. It was found that speech intelligibility was contained in time dominant ratios of the speech waveform. Speech quality was found to be contained in dominant frequency ratios. So the nervous system is designed to recognise two distinct paramaters: the time domain and the frequency domain.

Baised partially upon work done at Tufts University, it is known that the human voice does not depend upon frequencies. People who have had their larynix removed can use an "artificial larynix", a buzz generator held against the side of the throat. Word information is formed by the action of the jaw, tongue, teeth, glottis, and the nasal cavaties. The human nose and mouth form a variable time dely generator. Thus, the basic audio information our brains evolved to decipher, the human voice, is dependent not upon frequency but upon the time rate of change nature of sound caused by time delays imposed by the mouth and nasal passages.

The Neurophone makes use of these research results by suppressing the frequency domain and amplifying the time domain of the incomming signal. This is one reason why the Neurophone sounds so scratchy when one first begins to listen to it. The electronic circuitry of the Neurophone presents audio information to the skin in the way the skin originally evolved to recieve and decode the information eons ago. Thus the Neurphone directly inputs information into the brain and nervous system, and even the mind, bypassing all the normal sensory systems that lie between the mind/brain loop and the outside environment.

Since all of the acupuncture meridians are present on the surface of the skin, we found some very interesting affects when the skin was activated by the Neurophone. All the subjects of the experement were muscle tested fo r one major muscle for each of the twelve meridians bilaterally. The meridians were recorded as under active for muscle weak in the clear and over active for muscles that became weak only by stimulation of the alarm point for its respective meridian. The subjects were then connected to the Neurophone by placing the electrodes on the skin of the forehead just lateral ans superior to the eyes. K-Y Jelly was used to enhance the contact between the skin and the electrodes. A tape recording of pink noise was then played through the Neurophone for a period of two minutes and then the muscles were retested after removal of the electrodes.

Conclusion ~

Table 1 represents the totals of 10 different paitents who went thorugh the experiment. Out of the 240 meridiand tested, 85.8% of the unbalanced ones found, were corrected. The Neurophone appears to have a beneficial effect on the body. On subesquent visits, the meridians were remaining in balance two weeks later. Some of the many comments from the paitents included: cessations of seizures, inproved sleep, total relaxation, improved respiration, better concentration, and cessation of pain.

Additional Research ~

To be certain the Neurophone was not operating by bone conduction, a simple experement was done, known as the Batteau test in honor of the hearing research of the late Dr Dwight Batteau who developed the test during the Neurophone evaluation at Tufts University.

Two seperate channels of auditory information were used. One channel goes through a set of ordinary headphones, the other goes through the Neurophone. One specific frequency is played through the headphone channel. Another frequency slightly different is played through the Neurophone circuitry to the transducer disks. If the Neurophone was producing hearing by bone conduction, the two slightly different frequencies would "mix" in the bone structures of the inner ear, producing a discernible "beat frequency" which is the difference between the two frequencies. With the Neurophone, this "beat" is heard only at very high volume levels in both channels, levels at which the Neurophone is probably producing bone conduction by the strong vibration of the skin under each transducer disk. However, the beat frequency should theoreitcally be heard at all volume levels; it is not heard at normal neurophonic listining levels.

Looking Ahead ~

Although the present experement was a relatively simple one, the implications of future possibilities are tremendous. The following ideas may well serve as topics of future research papers and I invite all the members of the I.C.A.K. to pursue them, either on their own or in conjunction with me. Some of the possibilities for which ther is already some evidence are:

1. Increase in telepathic awareness.

2. A brain/mind link between two or more people. T.E. Bearden has developed a mathematical formula that indicates that the combined mind power of a group of people will multiply exponentially if these people are linked in a unitary consciousness.

3. Subliminal learning may be accomplished by the subject material being played through the Neurophone at a low volume level. No conscious effort is made to learn the material. This could even be accomlished while the student is asleep.

4. Conscious learning may be accomplished by simultaneously listening to tape recorded data by means of Neurophone and headphones. In this way, the learning centers of the brain are being accessed by at least two seperate channels.

5. Positive subliminal programs to alter undesirable habits: i.e., weight control, smoking control, and creation of positive mental attitude.

6. Control the aging process by means of positive cellular programming via the Neurophone.

7. Totally deaf people being able to the recorded audio.

Huyck Research Center
209 Greenwich Ave., Stamford, CT
28 Jun 1963

Operation of the Neurophone on the Nervous System

by
G. Patrick Flanagan

Definitions: CRT = Critical Reaction Time

In a purely rate of onset pulse information system, the pulse rate of the system and the pulse amplitude of the system would remain the same. The things which will be varied, are the slope of the leading edge of the pulse, and inversely the width.

In this system, the amplitude,(e) of the pulse remains the same; so does the repeition rate of the signal. The only things to be changed are the slope and thus the width of the top of the pulse. From now on, the slope in degrees will be represented by (à), and the width at the top will be represented by (T). We can then formulate an empirical formula for the relationship of à to T. (T measured in usec). T = K/à as à = 45 degrees, T = 0 or T = max time, à = 0 degrees. K = conversion factor.

Therefore T also equals I (instantaneous audio information).

The width of the entering pulse with no information must be equal to the CRT of the nervous system at any given instance. Then the slope of the pulse and thus the width at the top are varied, so is the response of the system. A pulse of the same amplitude but of the wrong width for the CRT of the neuron will only partially load the neuron. The amount or amplitude of the loaded signal in the neuron will be proportional to the width of the pulse as it is varied.

Now, if a sine wave is loaded into the system, the frequency of the wave must be varied so that the width of the wave will coincide with the varying CRT of the loaded system, otherwise the sine wave will have no effect on the neuron, i.e., very little.

It has been shown by my experimentation with the effects of adrenergia and cholinergia on the nervous system, that as the body goes into the state of adrenergia, the frequency of the carrier must be increased to meet the smaller CRT of the nerves introduced by the addition of adrenaline and acetylcholine into the system. And, as the body goes into a state of colinergia, the opposite is true, the frequency must be lowered, due to the increased CRT. Otherwise no information can be transmitted into the system.

In adrenergia, the excited state, the K ion mobility is increased and the CRT is smaller....thus requiring a higher frequency. In the case of cholinergia, the opposite is true.

Lets now analyze a 30 KHZ sine wave as applied to the nervous system and see what the nervous system sees. The sine wave is clipped at e and the width of the top of the clipped wave is < 15 usec. Now, if the nervous system at this instant has a CRT of approx 15 usec perfect loading will occur.

Now, if the CRT of the nervous system at this moment should change to 20 usec, the carrier would have to be shifted to 25 KHz to obtain the same response. This increase in CRT would indicate that the person had gone into a state of colinergia. The new CRT of the neuron would now cause the neuron to be improperly loaded at 30 KHZ, but properly loaded at 25 KHz.

Now we will look and see what the effects of modulated and unmodulated carrier have on the system at the neurons CRT. According to the Encylopedia Britannica, the output frequency generated by a neuron is proportional to the amplitude of the stimulus, until the stimulus reaches a certain level and the frequency then remains the same is.. the clipping level of the system. The completely loaded neuron will then be receiving maximum stimulus and will then be generating a frequency of its own in proportion to the loaded voltage. At this loading in the unmodulated carrier, a series of meaningless pulses are being generated and sent to the brain where they are probably rejected and passed off as random noise.

Now let us look at a tone modulated carrier (AM Signal) and see what happens with some information input. The pulse width at the top (clipping) now varies with change in carrier amplitude at the audio tone modulation frequency. The loading voltage of the neuron will now vary with the audio rate....the varying width of the top of the pulse now follows a pattern of widths in time with the audio tone.

The overall process: (Audio) I = varies T and which vary e loading in neuron which generates freq to the brain. We now have a system that detects and interprets rate of change into the neuron's own coding system. Or, as applied to a purely rate of onset modulation system with a pulse input instead of a sine wave input. We now have a coded conversion system for introducing information into the nervous system.

KeelyNet BBS (15 October 1993) Neuroph11.ASC ~
This File is shared with Keelynet courtesy of Terry Bastian

The Neurophone Mark XI Manual
(1979)

G. Patrick Flanagan, Ph.D.

This year the neurophone concept has now reached maturity. It was developed 21 years ago this year (1979). The entire history of the neurophone reads like a James Bond mystery novel. We are currently writing a full length book on the neurophone story, and expect to publish this next year. For the present, we will give you a synopsis of the neurophone story, and bring the reader up to date on the current development of the neurophone Mk XI, the eleventh model in a long chain of discovery. following the basic history of the neurophone, we will give instructions for the set up and use of the neurophone.

The first neurophone was developed when I was 14 years old. A brief description of the original device is given in the book PYRAMID POWER. The original neurophone patent #3,393,279, describes a high voltage, high frequency, amplitude modulated radio oscillator. This device made the listener into a radio receiver. It transmitted sound information to the user by means of small insulated electrodes. The 3000 volt signal produced an electric field of extremely low power density. As the carrier frequency was in the vicinity of 50 Khz, and the capacitance of the electrodes was therefore extremely low. The above electrical conditions of low power density indicate that the original device was safe for use. However, if the electrode surfaces were scratched, the user would experience a mild, but annoying mild electric shock.

When I originally received the idea for the first neurophone, I was a child of 14, and did not have funds to hire a patent attorney. One of my fathers friends was a patent attorney for Shell Development Corporation, and agreed to help me write my own application. Over the years, I became very familiar with patent law, as the patent examiner in Washington said the neurophone could not possibly work, as it defied ALL present laws of neurophysiology. The patent fight over the neurophone went on for many years, finally resulting in complete rejection by the examiner, closing the file forever with no recourse.

In a brilliant move, my patent attorney (I could afford one by now) suggested that we go to Washington and demonstrate the neurophone to the examiner in a final attempt to gain approval. When we arrived at the patent office, the examiner had a real surprise waiting for us, he had arranged for a deaf employee of the patent office to be present for the demonstration! The deaf gentleman was nerve deaf in one ear, and almost totally deaf in the other ear. When the deaf employee heard high frequency sounds for the first time in 15 years, he wept for joy. The examiner then reopened the neurophone file, and issued the patent giving the original filing date.

By the time the patent was actually issued, a company in New York owned by Joseph Lawrence, and Andreja Puharich was infringing on my patent. It seems that they had applied for a similar patent several months after I had applied for mine. As I was the first to file, and had pursued the patent, I won the device. The small New York company bought the patent, and by means of legal maneuvers, succeeded in causing me to lose a small fortune in the process. This was due to their issuance of lettered stock which tumbled to a fraction of its original value by the time the SEC allowed me to sell it.

We have actually gone far ahead of our story, so we will backtrack a little. during the years in which I battled the patent office for the original patent many events took place. When I was 17 years old, LIFE magazine called and wanted to do an article on the device. When I was 14, I had received national publicity on the front pages of 300 newspapers. After the LIFE article came out, I was offered up to $8,000,000 for the patent rights. The press coverage was unbelievable. The Ive Got a Secret show(Gary Moore) invited me to New York, and millions of people saw the neurophone demonstrated on Bess Meyerson. It was during this show, when I first met Puharich and Lawrence.

Puharich invited me to lunch at the Twenty One club, along with G. Harry Stine, a scientist under the employ of a firm in Connecticut known as Huyck Corporation. Stine's company was interested in the rights to the neurophone.

When all the offers were in, my parents and my Houston attorney agreed that the neurophone should be given to Huyck. After examining the device for two years, Huyck said the device could not be patented, and dropped it into my hands once again. By this time, I was under extreme pressure, as I was once again given the task of pursuing the patent on my own limited funds. At the same time, Puharich and Lawrence kept urging me to give up, saying that they had me beat in the patent office.

In 1964, I was offered a position at Tufts University near Boston. Dr. Dwight Wayne Batteau, a professor at Tufts had an intense interest in the neurophone as it might play a role in man dolphin communications. Dr. Batteau had a small company known as Listening Inc., along with a brilliant electronics man: Steve Moshier. Listening had extensive contracts with the ONR (Office of Naval Research), and was subcontracting Tufts in some of its research and development work.

Our first task was to establish the limits and parameters of the neurophone as it would apply to dolphins. Our secondary task was to determine the physiological mechanism by which the device worked.

Dolphins have extremely large brains, and communicate on many levels. We considered that the neurophone would open a channel to the dolphin which was not previously accessible.

In our initial testing, we found that the neurophone produced a minute mechanical vibration in the skin under the electrodes. When one person was listening to the device, other people standing near the electrodes could hear the sounds coming from the skin. If a stethoscope was placed in contact with the users skin, the vibration could be heard loudly, and clearly. Our initial effort was to determine if this vibration was producing bone conduction, or was the skin vibration merely an artifact?

Doctor Batteau came up with an ingenious experiment to determine if the skin vibration was actually leading to bone conduction: If the neurophone were a separate acoustic channel into the brain, we should be able to play two discordant sounds into the brain of an ordinary person, one by neurophone, and one by earphones through the ears. If bone conduction were involved, the sounds would mix or beat in the bone structures of the inner ear, producing a discord. In the process of bone conduction, sounds are transmitted through bony skeletal structures to the inner ear, where the small bones normally vibrated by the ear drum are activated by skeletal vibration. This vibration is then encoded by the inner ear, and transmitted to the brain by means of the 8th cranial nerve. We are sure that the neurophone did make use of another channel, due to the fact that people with 8th cranial nerve damage had heard with the device.

In our experiment, we played one frequency of sound into the neurophone, and another frequency into the headphones through the ears. If bone conduction were involved, the sounds would mix as predicted in the inner ear, and a discord would be perceived by the test subjects (Tufts psychology students). If the channel to the brain were a separate one, the signals would be heard simultaneously with no discord. At low volume levels, the students were able to perceive the separate sounds with no mixing. At higher volume levels, a mixing could be heard.

We then ran the same experiment by playing the sounds through stereo headphones, with each signal going into separate ears. The results were exactly the same as the neurophone experiment. At low volume levels, the signals were indeed perceived apart from each other. At higher volume levels, the discord could be perceived.

The higher volume discordant mixing can be explained by neural intermodulation possibly taking place in the corpus collosum, the small nerve bundle which connects the left and right hemispheres of the brain together. We tested a number of Tufts students, and they all had identical results. Wayne Batteau was so excited, that he suggested we write a letter to Science, for publication. Our letter is duplicated in the appendix.

As the neurophone was now proven to be a separate brain communication channel, Doctor Batteau made an offer for rights to the device. He offered to give me a 1/3 interest in Listening, a 5% royalty, and position of Vice President in charge of research. The idea was to develop the neurophones potential in many areas, including its potential as an interface (electronic corpus collosum) between man and dolphin, and play EEG signals from a human brain into the dolphin; and to reverse the procedure.

Our work was progressing at a satisfactory rate, when tragedy struck Listening. Wayne Batteau died of a heart attack in shallow water while swimming near our dolphin facility on Oahu, Hawaii. When he died, the Navy canceled our contracts and Listening began to fold. Shortly after that, we demonstrated the neurophone to the patent examiner, and the rest is history.

While investigating man dolphin communications, we succeeded in developing a language translator, a device that translated human speech into dolphin language, and vic versa. This development required a thorough understanding of the nature of speech, and information theory. We made many efforts to model the nervous system, and succeeded in demonstrating that the nervous system uses time ratios as major sources of intelligent information. We then began to investigate timing ratios in speech patterns of humans and dolphins. During that period, we found that speech intelligibility was contained in time dominant ratios in the speech waveform. We found that speech quality was contained in dominant frequency ratios. So, the nervous system is designed to recognize two distinct parameters: the time domain, and the frequency domain.

As a result of the knowledge gained in this area, I designed a circuit which suppressed the frequency domain, while amplifying the time domain. This device, was so radical in approach, that I applied for a patent on it as a specialized speech processor. Six months after the patent was applied for, the National Security Agency placed the patent application under a top secrecy order #756,124. The order said that my invention was being suppressed in the interest of National Security. It further stated that I could be tried for treason and shot by firing squad if I revealed its working nature to anyone other than an authorized government representative. In other words they stole it! Needless to say I was very disappointed in the system. It took four year, and three law firms to sue for release of my invention. We won the battle, and the secrecy order was rescinded. Patent #3,647,970 was issued on the 7th of March 1972.

In 1974, two years after I wrote Pyramid Power, we spent the night in the Great Pyramid. A full description of that event will have to wait for the publication of our full length book on the neurophone. (see also "Gods of Aquarius" by Brad Stieger a starseed book) I can say that I had an experience of enlightenment, including a full blown Kundalini Release. After that experience, I became a trance medium. During one of my trances, the message came through that the speech processing patent which was under secrecy so long was in actuality the perfect neurophone circuit. Well, this was a surprise, as I had never tried the circuit in that way. I must admit, I really did not believe it could work as a neurophone. The message to try the speech processor as a neurophone kept coming and coming until I finally decided to try it out. When I tried it, I was in the suprise of my life! It worked! The result was the development of the present neurophone Mk XI, the finest neurophone to date, and it does not require the use of a radio frequency carrier wave.

The clue as to how the neurophone actually works, is contained in the skin vibration artifact which we discovered at Tufts University. The original neurophone used a high voltage amplitude modulated carrier wave to create a molecular vibration in the skin itself. The skin became the diaphragm of a biological electrostatic vibrator. The skin is piezo electric and opto-electric. That is, when the skin is stimulated by an electric field, or by a photon field, it will contract and vibrate with modulation of the field. If it is mechanically stimulated, it will generate its own electric field. In Russia, blind people have been trained to see with their fingertips; and in Czechoslavakia, deaf people have been trained to hear with their fingertips.

The skin is the largest most complex organ of the living system. As we develop in the womb, all organs of sense evolve from the skin. The skin involutes and convolutes to form eyes, ears, etc. Our research indicates that the skin itself has the latent potential of performing all functions of perception.

The neurophone stimulates and develops this latent ability. The skin is the organ which receives the signal from the neurophone, and converts the incoming signal into a modulated molecular vibration which is then interpreted as sound. We could theoretically stimulate the sense of sight in a similar way. As all acupuncture meridians are present on the surface of the skin, we have found that the neurophone stimulation balances all the acupuncture meridians by activating the skin!

The neurophone Mk XI converts incoming non-linear acoustic information into a time domain amplified signal. This signal is then transmitted to a pair of high dielectric constant ceramic electrodes which are placed in contact with the skin of the head. The electric field (approx 20 Volts RMS) interacts with the skin ceramic electrode combination to create a molecular vibration in the skin. This is then interpreted by the brain. The result is a new modality for coupling information to the brain, using the skin itself as the receptor. Bone conduction vibrators will not work as a neurophone, as the vibratory signal is to gross, the skin itself must vibrate in a synchronous mode in accordance with the time encoded information.

Doctor Christopher Hills, in his book Nuclear Evolution states that the skin is a second brain, and is the basic organ of spiritual and psychic evolution. He states that the skin can be trained to develop powers of perception such as telepathy, etc...

The neural information processing system of the human body is apparently extremely sensitive to time domain information. Doctor Batteau postulated that the nervous system incorporates delay line correlation technology to detect time varied information ratios. The neurophone Mk XI processing circuit processes the incoming complex non linear signal waveform, and amplifies the non linearities thus increasing the timing recognition pattern of the signal. In the process, the frequency domain is suppressed. The time rate of change of the incoming signal is thus amplified. This signal is so time dominant, that it can be hard clipped or run through a zero crossing detector without losing any intelligibility.

This time processed signal is then fed to the pair of high dielectric ceramic electrodes. The 20 volt average RMS signal does not require a radio carrier to work. As stated earlier, the original neurophone design had to actually work by brute force, due to the fact that the modulation signal was not processed to increase the time domain signal properties. As previously stated, the original neurophone produced typical voltages on the order of 3000 at a carrier frequency of 50000 hertz. As the skin is piezoelectric, and has a dielectric constant in the range of 12000, the neurophone electrodes are made of a ceramic material designed to provide a maximum impedance match to the skin itself. The entire skin electrode system is a piezoelectric resonator. Note: The ceramic electrodes are fragile and could break if dropped on a hard surface. Be very careful, they are expensive to replace.

To obtain the best results from your neurophone experience, you should spend at least 1/2 hour per day listening to a broad spectrum frequency source in a quiet, relaxing environment. It is best to listen with an increased blood flow to the brain. The preferred position is an inclined plane of 11 degrees with the head down. Testing has shown that most people go into deep alpha within 30 seconds when placed in this position. This state is the most receptive state to listen to the neurophone. The neurophone listener can build his own incline plane from a board six feet long, and at least 18 inches wide. The raised end of the board should be supported at a level of 14 inches above the floor.

The electrodes should be placed on the temples, directly behind and slightly above the eyes. Do not place on the hair. Although the electrodes will work perfectly well without electrode jelly, we suggest the use of EEG type electrode jelly, or KY jelly, as this improves impedance matching to the skin. If you use KY jelly as an electrode cream, smear an even coating over the black ceramic of each electrode, and place the electrodes in contact with the skin. The electrodes may be held in place by means of an EEG type rubber strap, or an ordinary 2 inch ACE bandage may be used.

After the electrode strap is in position on the head, insert the KY coated ceramic electrodes under the strap so the electrodes are held securely in contact with the temples. Later, you may desire to move the electrodes around to experience different sensations. Many neurophone listeners prefer to place one electrode in the center of the forehead, on the 3rd eye area, and to place the other one on the back of the neck, or on the hand or wrist.

The sound source for neurophone listening can be a cassette player, a radio, or a stereo HI Fi system. The neurophone should be driven from a headphone or a speaker output jack. Your neurophone is provided with an audio connector cable with a mini plug on the cassette end, and a 1/4 inch phone plug on the other end. This will fit most cassette players. If you want to drive the neurophone from another source you may have to obtain a different wire. Your local Radio Shack store will probably have the right one.

In using the neurophone, I generally adjust the sound level of the cassette machine to a comfortable listening level as heard through the built in loudspeaker of the machine. I then plug the mini plug into the earphone jack of the player, and plug the standard phone plug into the input jack on the neurophone. Plug the electrode phone plug into the neurophone output jack. Rotate the neurophone volume control slowly clockwise. The switch on the control will turn the unit on. The small LED panel lamp should glow at this time. Slowly turn the control up until you begin to hear the tape from your cassette player through the neurophone electrodes. Depending on the program material to which you are listening, the sound which you first hear through the neurophone will not sound like Hi Fi. This is due to two things: The sound you hear is time domain dominant. 2. As this is a new listening channel, the brain actually has missing processing capability. If we run a frequency sweep of the neurophone while listening, we will find that all of us have certain spectra which are entirely missing from our perceptual ability. That is, in the beginning we may hear a complex sound wave of one millisecond duration (1KHz), but miss entirely a sound of another domain. As we listen through the neurophone, the missing ranges are programmed into the brain. After listening for as little as 30 minutes, the sound begins to take on new qualities. The sound appears to move around in the head, and take on new dimension as we program our psychic brain centers to receive the new signal input. The more the neurophone is used, the clearer it gets. I recommend electronic music tapes in the beginning, such as the astral sound tape.

As the neurophone channel is but one channel into the brain, we have found that supplementing the neurophone channel with the frequency domain input by means of quality headphones is desirable. The reintroduction of the frequency domain through the headphones adds a new dimension of sound to the brain mind input system. The combination of neurophone plus headphone signals provides the best listening combination. For this purpose, the neurophone contains a headphone output jack which will accept monaural or stereo headphones. The second volume control from the left is used to control sound to the headphones. As the headphones are driven by the cassette recorder, the neurophone must be plugged into a speaker or headphone output jack. The neurophone plus headphone adjustment should be made as follows: 1. First adjust the sound level to the electrodes with the headphone control in the off position. 2. With the headphones in place, slowly adjust the level of the headphones to supplement the electrode signal. Sit back, or lay back close your eyes and let the signals carry you away.

In the near future: Source of Innergy will begin to produce cassette tapes designed to be used only with the neurophone. The tapes will cover many different categories from: Psychic Center Stimulation; to Subliminal Habit Modification Programs. We will notify neurophone owners as these tapes become available for purchase.

In the beginning, it is not necessary to use special tapes, as the object is to develop the latent channel through which the neurophone works. This may be done by listening to white noise (waterfalls) or your favorite music tapes. Neurophone stimulated perceptual enhancement occurs with increasing frequency as you use your neurophone. This experience is similar to the meditation experience of transcending. These periods of extremely clarity become more and more pronounced as you put hours on your neurophone. All changes in awareness are not gradual. All progress is in the form of discrete steps. What may appear to be a gradual altering of consciousness is actually a series of graduations. We may plod along thinking we are making no forward progress, and at the point in time where we feel we want to give up we experience quantum leap in awareness. One of the most common awareness changes with the neurophone effect is an increase in telepathic awareness. Although this cannot be turned on at will, instances of its occurrence will increase in frequency as time goes on. Please keep a diary of neurophone hours of listening, and make note of any change in awareness, dreaming, or unusual perception changes.

We would like all neurophone owners to send us a monthly research report or diary of experiences. This is important data which will enable us to share with you all experiences, and tune the neurophone experience. Keep note of your actual listening time, and listening material. If you experience any change in consciousness or awareness record this down. Others would like to share your experience.

Retired Lt. Col. Thomas Bearden is a nuclear physicist from George Washington University. He is one of the foremost experts in warfare strategy, and is an expert in the field of Soviet Psychotronics Research. Tom Bearden and I have made two mind links by means of the neurophone. Details of the brain mind link will be given in a later edition of the Innergy News, or the neurophone research journal. Col Bearden has developed a mathematical formula that indicates that the combined mind power of a group of people will multiply exponentially if these people are linked in a unitary consciousness. A small group of people could actually alter the entire course of mankind in an instant of time, if they were perfectly linked in a unified field of consciousness. Col Bearden believes that the neurophone is the key to linking consciousness. He believes it may hold the answer to solving all the problems of mankind.

This linking could be done in a number of ways. The Soviets have established that an EEG machine of only 16 channels could pick up the entire consciousness of an individual. All that is then necessary, is to feed the data into the mind of another by means of a multi channel neurophone. The neurophone would then become an electronic corpus collosum between the minds of two or more people.

At a given time in the near future, we will notify all neurophone owners to all try to listen to the neurophone at an exact time period everyday. This will be a simple attempt to unify the consciousness of the group. This type of experiment has been somewhat successful in the past, when used with group meditation. It may be that this experiment will speed up the increased awareness brought about by the use of the neurophone.

History of the Neurowave Signal Development

1. A step up audio transformer was attached to a hi-fi amplifier. The output voltage of the audio transformer was about 1,500 volts peak to peak. The perceived sound quality was very poor,highly distorted and very weak. The signal was perceived as being loudest and clearest when the amplifier was over driven and clipped square waves were being generated. The O'scope signal had ringing spikes or oscillations of a dampened wave at frequencies of 40 to 50kHz. (Flanagan, *Life Magazine* 14 Sep 1962)

2. An amplitude-modulated signal was then fed into a high frequency transformer that was flat in frequency response for an adjustable frequency in the 20-100kHz range. The output was a 2,000 volt peak to peak amplitude modulated carrier wave. (Flanagan Patent No. 3,393,279 (1968)

3. The audio signal is pulse width modulated on to a 50KHZ square wave carrier. Output is stepped up to a 50 volt square wave. This signal is applied to the body by means of piezo ceramic disks.(lead zirconium titanate) (Flanagan Patent No. 3,647,970 (1972)

4. The audio signal is pulse width modulated on to a 45KHZ square wave carrier then double differentiated(processed thru 2 differentiator circuits). Output is stepped up to a 60 volts RMS. This signal is applied to the body by means of piezo ceramic disks (Radio Shack Type)(Brass side to skin, red leads connected to circuit).(Info obtained from various researchers)

5. The audio signal is differentiated and passed into a section that clips everything into a series of square waves and then converted to 40 volt pulses which is then run through a zero crossing detector (comparator).The sensor electrodes are one inch diameter plates made from lead zirconium titanate (piezoelectric discs). (Extrema, US Patent No. 4,545,065)

6. Audio signal is modulated on to a 100KHZ carrier and power amplified then sent to an antenna. Those near the antenna percieves the sound if like thru earphones while those further away hear nothing, no contact with the antenna is made.(approx 1970's)Laser Sound System, Inc., 438 W. Cypress, Glendale, Ca 91204.....also Intelectron Corp., 432 W 57th St., New York, NY, 10036)

7. Another unit is based upon a miniaturized hi-power amplifier amplifying the audio information directly, no square waves or carrirers are used. The sensors are piezo-electric transducers on ceramic bases. Good quality sound (not quite high fidelity headphone sound) is experenced coming from the middle of your brain, more or less.(www.throne.com)(GROC BOX(R))

How to use nuro devices:

1. Connect the brass electrode plates (or bare ceramic) ideally to a point above the shoulders. The sides of the neck or about 1-2 cm out on the face from the center line of the ears are good locations.

2. Adjust the sound input device (stereo tape player or CD player) volume to its lowest setting and then increase the volume until it reaches the maximum volume possible without creating any sensation at the point of electrode contact. You should hear an external sound emanating from the location of the electrodes. This is normal.

3. Using ear plugs, close off the exterior ears during the first week of use. This gives you time to turn your brains attention to the interior sound.

4. For the initial week of use it is advised to use music with a wide band of frequency. Classical string instruments with the full range of orchestal sounds work well. Use the device one or more hours a day.

5. After the first week of use switch to whatever learning tapes you wish to use. It is recommended that you use an auto reversing tape player and use it one or more hours per day.

***Life Magazine* (14 September 1962) ~**

Whiz Kid, Hands Down

by William Moeser

The very young man above is standing on his head because he says it helps him think. It evidently does. Pat Flanagan, a 17 year old inventor from Bellaire, Texas, is already nipping at the heels of the venerable 30 and 40 year old scientist and inventors who built the remarkable structures seen on pages 54 to 65. Pat has just perfected a remarkable machine of his own which one day may help deaf people hear and blind people see. It may also earn him a million dollars. Pat treats his imminent collision with success with equanimity, for he reckons and who is to gain say him nowadays that the generation which will take over from the take over generation will find nothing is impossible.

Pat Flanagan is a unique and self spurred teenage boy who has forged his mind and body into the model of a mature and inquisitive scientist. At the same time he reflects the more standard teenage model; he is the twist champion of Bellaire a suburb of Houston a moderate party goer and girl chaiser, the holder of a private pilots licence, and a spectacular gymnast. espite his ability to function in two worlds, Pat leaves no doubt which one he favors. There are far too many kids my age who are willing to just get along. Pat is confident in his ability to do alot more than just get along.

His single minded belief in his abilities began with a compelling dream he had when he was 8 years old. In the dream I was told I had to learn all about physics and electronics, he says. And it told me I should help people. Already an athletic boy able to do 300 pushups a day, he thereupon set out to improve his mind. By the time he was 13 he was repairing television sets during summer vacations, trying to earn money to build an electronic laboratory in his attic.

Pat's restless imagination drove him to tireless sessions in his laboratory. To abet them he solicited a rare favor from his parents and his older brother Mike the privilage to experiment there undisturbed. One weekend last October, Pat started the experiment which led to the development of his particular fantastic machine. Starting with a radio transmitter he had designed himself, he tried modulating its waves to see if he could induce hearing in his nervous systemwithout going through the normal channels of hearing. He hooked his radio to a small transmitter which looked like an earmuff. After 34 hours of work, he stopped up his ears, put the earmuff to his head and found he still could hear.

I ran downstairs to tell somebody anybody. I woke my mom. She just rolled over and said to me, thats nice, Pat, but Ill listen to it in the morning. She did listen in the morning and alot of very important people have been listening to Pat ever since.

Pat calls his device the neurophone and the process it operates by neuroception. Essentially what it does, he thinks, is transmit electrical messages identical to those sounds generate through the bodys nervous system direct to the brain. Hence he can place the neurophones earmuff on someone's spine or solar plexus, plug up his subjects ears, and the person will still hear. Obviously if the neurophone in fact does what it seems to do, Pat has come a long way twoard short circuiting the bodys ordinary sensory processes and giving man, unprecedented access to his brain. Other inventors many with a lot more experience and facilities than Pat have been seeking such a device for years, and Pat explains his success verses their failure as a product of his own vigorous one man approach to science. I believe research in the problem of electronic hearing has been limited because inventors havent been able to use human subjets as guinea pigs. An animal cant tell you just what 'he heard or how clearly he heard it. But I was my own gunea pig and I wasnt restricted by the possibile bad effects, and I got the secret.

There is some question as to just what Pat has got even he has no firm knowledge of why his neurophone works but no question whatever that somehow he has onto something valuable. Several comanies have expressed interest in buying the rights to the neurophone and one Corpus Cristi firm has tentatively offered him a million dollars if the machine can be adapted to send visual images into the brains of blind people. Dr. William O. Davis of Stamford, Conn Hyuck Corporation, a research and development company which is also fascinated by the neurophone says, The ability to detect radio signals in the brain is a remarkable phenomenon. If we never learn more about Pat's invention, even if we never learn why it works, it certainly is a utilitarian breakthru which could help a number of people. Davis, who used to run the Air Forces basic research program, adds, its important to realize that young Flanagan had the necessary intuition to invent his neurophone. You make discoveries intuitively, in the same manner you would paint a picture or write a symphony.

Pat wants to go on to college, but he is worried about fettering his talent: I seek the knowledge college will provide, but I never want to be just satisfied with what someone else has written and done. He hopes, as his skills increase, to probe other recesses of mens mind. I believe someday the entire concept of medical pratice will be changed by electronics, he says. People will be treated electronically rather than with medicine. If God can make the earth and sky and the force that people and trees live, then inventing anything less than this should be relatively simple.

Statements like this one tend to prove a bit abrasive to Pat's classmates. Pats a wise guy, plenty cocky, and sure of himself, one says, but the bad part of it is hes just that much better at anything he sets his mind to do.

Pat claims this reaction does not bother him I want to be accepted, sure, but some people were cutout to go full tilt. Pats hands and mind are always going full tilt of late. The books strewn across his cluttred attic laboratory range from Zen to Karate to electronic jurnals to the Hidden Persuaders. Lights glow from a wave testing machine and he is working on a new way of tuning TV sets.

People think Ive accomplished so much in life, he says. They say what else can you do , and all that stuff. But I know where Im going and I know what I have to do. When I die I want to leave behind something which will greatly affect and help everyone.

KeelyNet (15 October 1993) Neurop50.ASC ~
This File is shared with Keelynet courtesy of Terry Bastian

The Neurophone Model 50 ~ The Thinkman
Principles of Operation

The Thinkman is the latest development in the series of neurophone devices first invented by Patrick Flanagan in 1959 when he was a prolific young inventor of 14. Since that time, considerable progress has been made in the development of improved neurophone devices, and the thinkman is the fiftieth neurophone system developed by Dr. Flanagan. A full and complete theory of the neurological, physiological and psychological operation of the neurophone is still being pieced together.

How does it work? How is it possible to hear without using ones ears? The current hypothesis is based upon the fact that the skin is embrionically the source of all our human sense organs. In fact, the skin itself contains more sensors, for heat, touch, pain, etc., than any other part of the human anatomy. The human ear evolves embryonically out of the convolutions of the skin of an embryo in the mothers womb. Basically, the skin is the oldest evolutionary nervous system sensor. Since it is the precursor of the ears, the skin should also be capable of hearing and, as the neurophone proves, the skin does indeed have this capability.

Neurologically, the human skin is both piezoelectric and optoelectric. This means that it produces minute electrical currents when vibrated or rubbed. Soviet and Czechoslovakian neurological research has also shown that the skin produces an electric current when stimulated by light.

As long ago as 1785, Charles Augustin DeCoulomb, a french physicist and an early experimenter with electricity, proved that an electrostatic field produces a measurable physical force. The neurophone processes audio information to produce a very weak 20 volt RMS electric field at each of its two transducer disks. This alternating electric field is changed as a function of the time rate of change of the audio signal coming into the neurophone. This minute electrical field actually causes microscopic vibrations of the skin under the transducer disks. Maximum coupling of the electric field to the skin is ensured by fabricating the transducer disks from Zirconium Titanate which possess the same dielectric constant as human skin. If you were to put an ordinary medical stethoscope on the skin next to one of the transducer disks while the neurophone is being used, you would be able to detect the vibrations of the skin created by the tiny electric field of the transducer disk.

Work done by Dr's Patrick Flanagan and Dwight Wayne Batteau at Tufts University during the years of 1964-1968 is the basis of the current theory regarding how the neurophone works. They discovered that the frequency content of the human voice had little to do with the brain's ability to recognize intelligence in human speech. For example, people who have had their larynx removed can use an artificial larynx, a buzz generator or low frequency vibrator held against the side of the throat. Words are formed totally by the action of the jaw, the tongue, the teeth, the glottis, and the nasal cavities. These cavities form a highly variable time delay encoding chamber. They found that the basic audio information which our brains evolved to decipher, the human voice, is dependent not upon frequency but upon the time rate of change nature of a sound caused by time delays imposed by the mouth and nasal passages. The neurophone makes use of these time delay codes by processing the incoming audio signal to remove the frequency component and leave only the time domain, the time rate of change information. This is one reason why the neurophone sounds so scratchy when one first begins to listen to it. Thus, the electronic circuitry presents audio information to the skin in the manner that the skin was originally designed to receive and decode such information eons ago.

But is it certain that the neurophone is not operating by bone conduction as are some other devices available today for listening experiments and enjoyment? A definitive experiment proving that bone conduction is not a cause of neurophonic hearing can be duplicated by anyone with the required simple equipment. This procedure is called the Batteau test, honoring the late hearing researcher, Dr Dwight Wayne Batteau, who developed the test during neurophone evaluation at Tufts University. Two separate channels of audio information are required. One channel goes through a set of ordinary headphones, the other goes through the neurophone. One specific frequency is played through the headphone channel. Another slightly different is played through the neurophone circuitry to the transducer disks. If the neurophone were producing hearing by bone conduction, the two slightly different frequencies would mix in the bone structure of the inner ear producing a discernible beat frequency. With the nerophone, this beat frequency is heard only at very high volume levels in both channels, levels at which the neurophone probably producing bone conduction by the strong vibration of the skin under each transducer disk. However the beat frequency should theoretically be heard at all volume levels and yet it is not heard at normal neurophonic listening levels. The neurophonic experience is therefore probably a new way to hear, using a new channel into the brain: the skin.

The neurophone is an electronic audio information processor designed and sold for experimental and entertainment purposes. The electronic circuitry of the neurophone accepts an input from any audio or Hi Fi system. It converts the audio signal into a digital like low voltage electrical signal that activates two small transducer disks. These two disks or electrodes may be placed anywhere on the bare skin of the listener. When the transducers are in contact with the skin, the audio signal will then be perceived in the persons head. There are no hazardous voltages or currents between the two sensor disks. There are no radio frequency carriers involved in the action of the neurophone. The neurophone thinkman operates from a 9 volt transistor radio type battery which is enclosed in its case.

The audio signal source for the neurophone may be a portable cassette tape player, a radio, or a stereo Hi Fi system. Most modern cassette tape machines and Hi Fi stereo systems have a headphone or external speaker output jack. The neurophone output goes through two 1 inch diameter transducer disks fabricated of Zirconium Titinate which are imbedded in acrylic plastic tiles to protect the brittle disks against breakage. The transducer disks are, nevertheless, still fragile and should be treated carefully because hard physical shocks will break them. If a transducer disk is broken by rough handling, a new set of transducer disks may be obtained from us. Insert the plug at the end of the transducer lead into the jack labeled electrode on the end of the thinkman.

Connect your neurophone to an audio source such as a cassette machine by means of an audio connecting cable. The small plug on your connecting cable will usually fit into the headphone output or external speaker jack on your machine, the other plug on your connecting cable will plug into the audio jack on the neurophone. Before plugging the cable into your cassette jack, adjust the output volume control on your machine to a 50% or higher level to ensure that the signal processor on your neurophone is receiving an adequate signal level. Turn on the neurophone by rotating the volume control in a clockwise direction. The red LED indicator light will glow, indicating the neurophone is on. Place the two transducer disks on the bare skin on either side of the forehead; They may be held in place by the elastic headband supplied with the unit. Turn the volume control clockwise to about mid way. Turn on your audio source so that the signal will now play through the neurophone. Place the electrode disks on the skin of your temples, plug up your ears, and the audio signal should be heard, appearing to exist in the middle of your head. If necessary, adjust the neurophone volume control until the signal is heard.

As you become more experienced in neurophone listening, you will find that you will be able to place the two transducer disks on the bare skin in spots other than the forehead. One neurophone listener reports excellent neurophone listening with one transducer disk on the soft flesh of one leg and the other transducer on the soft skin of the stomach. However, the neurophone must be operated at a higher volume setting to obtain the same level of neurophone hearing under such remote transducer locations. The ability of the transducer disks to function at a location of the skin remote from the head is partially explained by the current operating hypothesis.

If your neurophone begins to sound weak and the red LED indicator on the front panel begins to glow very dimly, replace the battery in the unit. Unfasten the four phillips head screws holding the bottom on the case and remove the bottom plate. Inside, you will find a battery clip and a 9 volt transistor battery. Replace the battery with a fresh one. For best results, use an alkaline battery which will provide more than 10 hours of neurophone use under normal conditions.

The neurophone processes audio information in such a way that the frequency domain is eliminated but the time domain is preserved. Studies with earlier versions of the neurophone by Tufts University for the US Navy as long ago as 1966 indicated that the brain initially may have difficulty in recognizing the neurophone signals due to gaps in perception. The user may not be able to perceive certain frequencies and time domains in portions of the audio spectrum. As you use your neurophone, these holes in your neurophone hearing process will disappear, as your brain learns to recognize these energies, and you will begin to hear neurophonic sound with full frequency and wide dynamic range.

The neurophone is an experimental listening device that may by continuous use stimulate and enhance dormant perceptive abilities in the user. For example, since the device is apparently stimulating pathways to the brain that are not used normally, it may increase intelligence, telepathic ability, and neural efficiency. Dr. Flanagan has used the neurophone longer than anyone in existence, and has developed extremely high neural efficiency scores. He believes this increase in brain efficiency is directly related to the use of the neurophone device. Many long term neurophone users report increased awareness, telepathic ability, out of body experiences, better memory, and increased auditory frequency range. A commercial pilot from Saudi Arabia reported a recovery of hearing damage due to aural trauma resulting from the loud aircraft engine noise as a result of his job. Please note however, that the neurophone is an experimental device, and has many uses that have been untapped. The neurophone owner will have to experiment and discover his own uses for the device. Dr Flanagan used his own neurophone to learn the Arabic language; actress Susan Strasberg found that she could learn her stage lines easier by playing them through the device. A recent book by G. Harry Stine, titled THE SILICON GODS by Dell, discusses the possibility of using the neurophone as a part of a mind computer link in which the power of the human mind could be amplified by millions of times. Stine says that the entire process if feasible right now, with the help of the neurophone and state of the art brain scan technology. Tom Bearden, in his book EXCALIBER BRIEFING says that the neurophone may eventually be used to successfully enable one or more people to do a perfect mind link in which the power of multiple minds may solve all world problems.

We hope that neurophone owners will join us in the task of discovering and using the potential of the device. The field of neurophone research is wide open, in the future we will see the development of neurophone software and hardware as well as neurophone user's league, our own neurophone magazine, and a host of other neurophone support devices. We invite all neurophone owners to share their experiences with us, so that we may share them with others.

TECH DATA: Model 50 ~ US Patent # 3,647,970

Input impedance: 5000 ohms
Max output voltage at disks: 20 volt RMS>
Battery: 9 volt
Battery life: 10 hours Alkaline cell
Output: Up to 12 pairs of disks
Control: Transducer volume
Operating temp: -40 to +120 F
Dimensions: 4 3/4"x 2 1/2"x 1 1/2"

KeelyNet (15 October 1993) Pacinian.ASC ~
This File is shared with Keelynet courtesy of Terry Bastian

The Pacinian Corpuscle

The skin contains many energy sensing mechanism, one such mechanism is known as the Pacinian Corpuscle. The PC is a special nerve ending that transforms mechanical vibrations or pressures into nerve impulses. Until this year, (1980) very little was known about thie transducer, and it was thought that the device could only work at low frequencies of stimulation.

Recent research by Fernando Grandori and Antonio Pedotti of Milan, Italy has shed a whole new light on this mechanism. (*IEEE Transaction on Biomedical Engineering,*, Vol BME-27, #10, Oct 1980) It now seems that the PC can react to very high frequencies, and responds best to a square wave stimulus.

The PC consists of a sensitive nerve ending surrounded by a cylindrical-like core structre consisting of closely packed membranes called lamellae. The core is surrounded by a second set of lamellae, in which the distance between each one increases from the innermost lamella toward the periphery of the corpuscle. The space between all the lameae is filled with a liquid whos mechanical properties can be considered similar to that of water. When a displacement of the outermost lamella is caused by exerting an appropriate pressure, this stimulus is transmitted to the core, and in turn, causes a compression of the innermost lamella. This system acts as a differentiating mechanical amplifier.

What is significant here, is that the PC responds best to a Squarewave. The above paper reveals that rate of change is more important than the amplitude of the pressure applied to the corpuscule. This indicates that this highly refined receptor is designed to detect time significance first, and pressure significance secondly.

If we take the Neurophone electrode disks and feed a 50 volt sine wave audio signal into them while placing them on the temples, the signal will be percieved weakly. If we feed the same signal as a square wave, the percieved is 10 times as loud as the sine wave. This tends to substantiate Dr Flanagan's theory reguarding the Neurophone perceptual mechanism as being timerate of change encoded.

PC are located over the entire skins surface, with greater concentrations on the fingertips and sexual organs. In 1981 Dr Flanagan intends to develop research projects which will be able to evaluate the PC theory.

KeelyNet BBS (15 October 1993) Theory.ASC ~
This File is shared with Keelynet courtesy of Terry Bastian

Neurophone Breakthrough

A Possible Mechanism for Neurophonic Action

The Neurophone is a new electronic invention that may enable us to hear by a completely new information channel to the brain. Ordinary hearing is the result of the stimulation of bones in the inner ear by means of vibration. Sound waves may reach these bones through ear canal via the ear drum, or by bone conduction in which sound waves are conducted to the inner ear vibrations in the crainial bones.

When the sound waves reach the inner ear, a vibration is set up in the cochlea which then converts the waves into nerve inpulses that travel up the 8th Crainial Nerve to the sound recognition centers of the brain.

In 1958, Dr Flanagan, then a child of 14 developed a radio transmitter that made the brain into a radio reciever. This device transmits acoustic information to the brain by means of radio waves into the skin, bypassing the 8th Crainial Nerve. When he applied for a patent on the device, the patent examiner rejected the whole thing saying that such a device would go against all known laws of science. ver the following years, Dr Flanagan fought against insurmontable odds to prove that the device did indeed work. In the meantime, LIFE magazine ran a major article on Flanagan and the Neurophone, naming him as one of the top ten scientists in the US at the age of 17! In a final desperate move Flanagan flew to the patent office with a model of his invention and successfully demonstrated the device on a deaf employee in the patent examiner's office. The deaf man heard music for the first time in 15 years and broke down into tears. The examiner declared that the Nurophone was indeed a basic patentable device and approved the patent for release. Patent # 3,393,279 dated 16 July 1968..........

In the years that Dr Flanagan fought to recieve deserved recognition by the patent office, he grew into manhood and was working on Man-Dolphin Communications for the US Navy when the patent was finally issued. While involved in Man-Dolphin research, he became interested in nerve signal information encoding, and began to develop electronic circuits that duplicated the process of pattern recognityion observed in the human nervous system. This work led to research in Cryptography. During that period he developed a top secret sound scrambler that was virtually impossible to decode. Part of the scrambler was based on his research into nerve encoding.

Dr Flanagan believed that the pattern of nerve encoding used in the human speech recognition system could be used to make a better Neurophone. He succeeded in perfecting an electronic curcuit that he believes duplicates the precise encoding of the Cochlea and 8th Crainial Nerve. When he applied for a patent on the new circuit, the patent aplication was immediately placed under top secrecy by the National Security Agency. The only explnation given at the time was that the circuit had potential uses in the defense of the country. Dr Flanagan was happy that the government considered that his device could be used in his country's defense. The only problem was that the government wanted the devise free, and he spent 14 years on it.

He hired attorneys and challenged the secrecy order for over five years. At the end of that period, the patent was released from secrecy and was approved for issue by the patent office. Patent # 3,647,970 dated 7 Mar 1972.

Dr Flanagan then perfected the circuit for another five years. This circuit recognizes time-relationships in the signal waveform, and generates a square wave that is time encoded. Dr Flanagan believes that the nervous system uses a complex deley line time recognition computational system that recognises time information. (50KHZ square wave pulse width audio modulation with double differentiator output)

In July of 1978, he successfully applied the Time Recognition Processor to his Neurophone. When an audio signal is processed through his circuit, it is converted into a form which he believes is an electronic analog of the nerve signal released from the human cochlea, but with one major difference; in the cochlea hundreds of nerves carry the time-encoded signal to the brain. In the case of the Neurophone, the full signal processing is complete and may be carried to the brain by alternate pathways. Through the skin itself.

In the original Neurophone, a 3000 volt amplitude modulated radio wave carried the signal to a pair of insulated electrodes that were placed on the head of the subject. In the present Neurophone, the voltage has been reduced to a 50 volt (maximum) square wave. This signal is applied to the body by means of ceramic disks.(zirconium titanate) The ceramic disks allow the energy field to affect the skin without a current flow. The small electric field causes the skin to vibrate internally in rhythim with the stimulation. The intea-dermal vibration can be heard by others if they place there own ears near the point of electrode skin contact. The vibration is not powerful enough however, to vibrate the bone below the skin surface.

For the past year, Dr Flanagan has been developing a theory which would explain how the Neurophone actually works in the body. Some mechanism must exist that transmits the information from skin to the brain.

The Neurophone has been out now for nearly five months, and we are starting to get reports back from layman and professional users. The inital results are exciting. Several people have reported that their ability to remember data is increasing. People who could not remember telephone numbers are becoming walking phone books!

The first professional research report has been written by Dr Sheldon Deal, D.C., N.D.; of Tucson, Az. Dr Deal is director of the Swan Clinic, and is current President of the International College of Applied Kinesiology.

Dr Deal's preliminary Neurophone research paper was recently presented at the I.C.A.K. convention in Detroit. With Dr Deal's permission, we are publishing his entire paper in INNERGY NEWS.

KeelyNet (15 October 1993) Bearden.ASC ~
This File is shared with Keelynet courtesy of Terry Bastian

Neurophone Extract from T.E. Bearden's Book *Excalibur Briefing*

Another device that uses the new hyperspatial, virtual state, nested modulation technology (and has done so for 17 or 18 years) is Dr. Pat Flanagan's neurophone. With brilliant insight and intuition far beyond that of science at the time, Pat invented and patented the instrument by the time he was seventeen years old. The neurophone is a device that, contrary to all present theory and knowledge, will directly pump the brain and reproduce sound and information directly in the brain and mind system, without going through the auditory system at all. A simplified diagram of Pat's improved neurophone is shown in the figure.

Briefly, the device takes a complex signal, such as the sound of an orchestra playing a musical interlude, and electrically processes it as shown in the figure. First the signal is passed into a section that clips everything into a series of square waves, remarkably analogous to the sort of clipped waves Lisitsyn confirms are the information carriers of the human brain. Next the square waves are differentiated, ielding a series of sharp spikes (note that these spikes retain the pulse time content of the clipped signal). These spikes are again differentiated, and since these are finite spikes with real nonzero rize times and decay times rather than theoretical constructs, a series of noisy spikes results from the second differentiator section. From here, the noisy spikes are introduced to special contact electrodes, one of which is normally placed on the forehead, while the other may be placed almost anywhere, including on the foot. Nowhere are any sound waves introduced to the head.

The square wave clipper section reduces the complex signals, their overtones, and their complex modulations to square waves, retaining the temporal content of the wave mix but not the waves themselves. The first and second differentiators highten or filter through the temporal content of the higher order differentiations, that is, they serve as a band pass filter unit to accent the time keying of the neutrinic and mind field portions or aspects of the signal. When these time spikes are then introduced across the body as pulsed voltages, they are modulated directly on the dendrite firings of the brain and nervous system, providing direct and pulsed modulation of the neutrnic and mind field component channels of the mind brain consciousness life loop itself. Thus the neurophone directly inputs information into the brain and nervous system, bypassing all the normal sensory systems that lie between the mind brain loop and outside environment.

KeelyNet (15 October 1993) History.ASC ~
This File is shared with Keelynet courtesy of Terry Bastian

Excerpt from Flanagan's book *Pyramid Power*

(8th Edition, 1980)

Chapter 3

Authors Background, The Neurophone

In 1958, when I was fourteen years old, a close friend of mine, Lou Macko, a television repairman in Houston, Texas told me of a most unusual phenomenon. He told me of a device that would transmit sound to the brain without using the ears.

The description of the device had been given to him by a complete stranger when he had visited Chicago. The device consisted of two copper mesh scrub pads with wires attached to them. The pads were placed in a pair of plastic bags to act as insulators. The wires were then connected to the high voltage side of an audio output transformer which was connected back to back with the output of an ordinary record player.

The result is shown in the figure.

The voltage from the Hi Fi was stepped up a bit and applied to the pads. If the pads were placed on the skin of the head in the vicinity of the temples, one would "hear" the sound from the Hi Fi as if it were coming from within the head itself.

The item at that time had no practical advantage as it was extremely distorted. A good deal of the sound seemed to be missing as if it were cut off.

I researched at the library and discovered the phenomenon was known as far back as 1800 and was discovered by Volta. It was called electrophonic hearing. It was believed that the phenomenon was merely the action of the muscles being electrically stimulated and affecting the bones in the ear by means of minute muscle vibration.

I examined the signal from the device by oscilloscope and discovered the sound came through in a blast when the transformer was overloaded and produced a sharp spike or ringing on the wave form. I soon discovered that the real information was coming through only when this effect occurred. This explained why only parts of the music and voice came through.

I reasoned that the true carrier of information was a radio signal due to the oscillation produced by shock excitation of the combined circuit of the transformer and the parallel tuned circuit formed by the electrodes and the human body. I started experimenting with the unit using a high frequency oscillator of my own design and discovered a resonance in the circuit around 40,000 cycles per second. I soon found that the resonant frequency would change abruptly with emotions and general body changes. The capacitance of the electrodes, therefore the dielectric constant of the skin, changed abruptly from the slightest outside stimulus. The dielectric constant of the skin would change by several orders of magnitude in a fraction of a second! After preliminary measurement of parameters, I designed the original Neurophone which is illustrated in my patent on the device. (#3,393,279)

The device was essentially a high voltage frequency modulated radio transmitter of low power. Its frequency was adjustable to correct for changes of resonance. The original unit was hand adjustable. Later units were automatically tuned for maximum resonance.

The addition of the resonant radio carrier wave made the difference. The sound from the device was fantastic, like sound from another world. The normal frequency response of the ear was extended beyond normal boundaries and there was no distortion.

I began experimenting with a number of people who were considered to be totally nerve deaf, that is, they were unable to hear anything, even with what is known as bone conduction. The results were spectacular. People who had not heard in years were now hearing sounds they only dreamed of ever recapturing. A patent was applied for, and my research continued.

The press learned of my discovery and articles appeared on the front pages of 300 newspapers, and it was picked up by news syndicates and international magazines around the world. LIFE did an article in Sept 14, 1962 page 69, 72. The results were as spectacular as the discovery. We received hundreds of thousands of letters from all over the world. Letters came from as far away as Tasmania addressed only to Pat Flanagan, Inventor, Texas, U.S.A. The U.S. Postal service did a spectacular job in getting the mail to the right party.

There were many Neurophone discoveries not made public. On occasions the device stimulated perfect telepathic contact between one or more persons, often with startling results. These telepathic incidents were kept under wraps for fear of ridicule.

KeleyNet BBS (15 October 1993) Meditate.ASC ~
This File is shared with Keelynet courtesy of Terry Bastian

As a result of his research, Dr Flanagan has designed the ideal meditation environment. This environment is a special room shielded from electromagnetic pollution. In this room are placed: an ERG (earth resonance generator), Ion Generator (caution-some ion gen produce pollution as well), an 8HZ pulsating high voltage field (sync to the ERG), and a Neurophone.....

Patrick Flanagan's Neurophone --- Hope for the Deaf and Superlearning for All

by
Eddy Taylor

In 1958, Dr. Patrick Flanagan invented the Neurophone in Bellaire, Texas when he was 14 years old. This electronic device transmits sound through the skin, by-passing normal hearing. A family friend who was a patent attorney for Shell Oil, helped Patrick submit a patent application. The patent examiners thought that this was just sound transfer through bone conduction and refused to issue a patent for 12 years. In a rare meeting in 1970, the patent office agreed to examine the Neurophone for themselves and meet

Patrick and his attorney. They both encountered a surprise. The examiner had a deaf employee attend the meeting to test the device. The man was totally nerve deaf in one ear and almost totally deaf in the other. Patrick showed him how to use the Neurophone and played a record of the famous Maria Callas singing an opera. As he was able to hear the undistorted beauty of her voice, the tears of joy streamed down his face.

For the first time in history the Patent Office reopened a file after it was officially closed and Patrick Flanagan received United States Patent Number 3,393,279. Prior to this, scientists thought that sound could only be perceived by the inner ear to the brain.

A Houston Post reporter had a relative who was nerve deaf from spinal meningitis. A test was arranged and when the device worked, an explosion of news coverage resulted in the story being carried on the international wire services.

How the genie was lost ~

In an agonizing infringement, a Defense Intelligence Agency sealed the patent and a national security order forbade Patrick from working on, or even talking about this invention. After many years and nearly 300 inventions later, Patrick received the Neurophone for public use. In a triumphant court case, the secrecy and the seizure of the device, by the Government Agency, was overruled and the genie was out of the bottle!

Electronic hearing and telepathy ~

To quote Patrick's Neurophone article in Nexus Magazine-Feb/March 1994, "All hearing aids stimulate tiny bones in the ear. In order for bone conduction to work, the cochlea or inner ear that connects to the 8th cranial nerve must function. People who are nerve deaf cannot hear through bone conduction because the nerves in the inner ear are not functional. A number of nerve deaf people who have had the entire inner ear removed by surgery have been able to hear with the Neurophone.

The Neurophone is really an electronic telepathy machine. Several tests prove that it bypasses the 8th cranial nerve or hearing nerve and transmits sound directly to the brain. This means that the Neurophone stimulates perception through a 7th or Alternate Sense!"

The Texas home of the Neurophone ~

The Neurophone became publicly available on a broad scale for the first time in July, 1996. An international Health Products company is now distributing the Neurophone. This corporation is based in Dallas, Texas and was established in 1976. A digitally computerized model is now being developed by Patrick and his wife Gael Crystal in their laboratory in Sedona, Arizona.

Turning on our neurophysiological potential for health, knowledge and awareness ~

One of the last statements by Nikola Tesla, inventor of alternating current, the radio and holder of well over 1,000 patents, was that he regretted not having done more in the science of electromedicine. In 1962, Patrick predicted that electromedicine would revolutionize conventional health care. One of the relatively unknown, silent revolutions has already taken place in the form of blood and lymph cleansing devices. These simple, yet powerful, electronic devices have been popularized by Bob Beck, physicist and bio electronic researcher and Hulda Clark, N.D.

In 1990, a remarkable discovery was made by Steven Kaali, MD and William Lyman, associate Professor of Pathology at Einstein College of Medicine in N.Y.C. It was shown that a minute current (50-100 micro amperes) can alter outer protein layers of the HIV virus and prevent it's attachment to receptor sites. (Science News March 30, 1991, pg. 207). The viruses loose the ability to make an enzyme crucial to their reproduction. This process may also reverse Epstein Barr (Chronic Fatigue Syndrome), Hepatitis and Herpes B. Responsible users of this technology who are HIV positive, may expect a Negative P24 surface antigen or PCR test (no more HIV detectable in blood) after 30 days. A simplified version of this unit now makes self help feasible. The potential to clean and potentize the blood banks of the world with this instrument is truly staggering. Bob Beck's Sept. 96 Explore Magazine Article notes a study on the life span of blood cells sealed under cover slips on microscope slides. While the average life of "normal" blood is about 4 days; blood cells treated with a mild microcurrent live for well over a month!

This very mild charging of the blood does not harm blood cells. Furthermore, the applied physics of Eddy Current (Lenz's law) demonstrates the neutralization of a myriad of parasites, viruses, microbes and fungi. It is important to realize that these organisms are critical co-factors, if not actual carriers, regarding the multitude of new deadly contagious diseases appearing throughout the world.

These blood cleansing devices are safe, with intelligent use, have been around for decades, and seem to have a variety of other benefits. You can do your own research on the scientific evidence by scanning (http:/www.electriciti.com/explore/Articles/Beck/HIVArticle.html). Be aware that there are now customized, blood and lymph cleansing units far better than Mr. Beck, Dr. Clark, and most others, describe and or advertise.

The big turn off ~

By contrast, the HAARP Project, a giant array of microwave towers in Anchorage Alaska, may not exactly raise awareness and improve human health. It's billion plus watts of power can impact an entire nation or even a complete hemisphere of the world. You may find the rest of the story through the writings of Nick Begich and other investigators. This brings new meaning to that old expression used in T.V. and Radio, "Don't touch that dial!" Nick is a life long educator, native of Alaska, and has been featured on thousands of radio programs in the last 4 years. He also is perhaps the most informed writer on the Patrick Flanagan story and the benefits of the Neurophone.

Education Automation Vacation ~

Scientific experiments indicate that effects of the Neurophone range from super learning, long term memory speed learning, relaxation, pain control and enhanced psychic abilities. The Neurophone's ability to transfer large amounts of information into long term memory may, alone, make the current model of education obsolete. We could advance beyond the problem of information overload and actually emphasize values, goals, strategy and a profound interdisciplinary approach to world problems. Positive solutions in an ever more complex world may themselves depend on the solution offered by the responsible use of the Neurophone.

Ma Bell ~

Studies on a special "head start" program, with pregnant women using the Neurophone have promising news. These women describe children with marked intelligence. This neuro-networking seems actually to nurture neuropathways for the fetus.

Extra sensory perceptions ~

All organs of perceptions evolve from the skin of the child within the womb. Theoretically, the skin could perform all sense perception. In Russia blind people have been taught 'to see' with their fingertips. Deaf people in Czechoslovakia have been instructed 'to hear' with their fingertips.

Details on the Neurophone, super learning and electromedicine ~

The intricacies of how the Neurophone is constructed, along with it's bio-physical- interactivity, is provided in the book, *Towards A New Alchemy*, by Nick Begich. For a broader background see *Super-Memory: The Revolution*, by Sheila Ostrander and Lynn Schroeder. The Explore More Magazine's March/April 1996 article, 'The Skin, Our Fifth Sense' is a real eye opener. For a historical view of related research read *Psychic Discoveries Behind The Iron Curtain* by Sheila Ostrander and Lynn Schroeder. A technical understanding of the bio-electronics of organisms in health and disease, will be found in the book, *The Body Electric- Electromagnetism and the Foundation of Life*, by Robert O. Becker, M.D. and Gary Selden, (Quill Publisher).

Dolphin speak ~

In a curious parallel to Roxanne Kremer's work on interspecies communication with the Pink Amazon River Dolphins, Dr. Flanagan has made impressive technical contributions. In February, 1968, he applied for a patent on a device for translating human speech into dolphin language and vice/versa. This was a result of studies with dolphins in the lagoon of a small island off the coast of Oahu, Hawaii. A vocabulary of 30 words was discovered before a startling intervention. Six months after applying for the patent, it was placed under secrecy order #756, 124 by a U.S.Government surveillance agency. Five years later, another hard-fought legal battle rescinded the suppression and patent #3,647,970 was granted on March 7th 1972. This speech processing patent is actually used as the circuit in the present version of the Neurophone.

The sound of the new millennium ~

"The Music of the Spheres and the Hemispheres", heard with the assist of the Neurophone, opens new doors for musicians and composers, as well as educators and students of life in general. In what may be seen as another motif on Sonic Bloom-like-phenomena, some theories hold that it has the potential for developing super plants and healing animals as well as people.

Mind meld ~

Experiments with a Neurophone mind link between two or more people seemingly makes possible learning directly from the mind of another. And I thought this was only in the SCI FI domain of Star Trek's Mr. Spock and certainly beyond the ken of Dr. Spock. Nick Begich describes the phenomenon as an electronic 'corpus collosum' between the minds of 2 people. The corpus collosum is the semi-permeable bridge between the 2 hemispheres of an individual's brain.

The Neurophone and other gems ~

The Neurophone also causes both sides of the brain to pulse in harmony creating an environment which may be ideal for learning. Use of the Neurophone tends to balance all the acupuncture meridians. In the near future the Flanagans will produce cassette tapes and CD's designed to be used only with the Neurophone. The tapes will cover categories as varied as Psychic Center Stimulation to Subliminal Learning Programs.

Other Flanagan products include Crystal Energy, (ultra colloidals --- each only 12 atoms wide --- of silica, magnesium, zinc, gold, silver, and titanium which have duplicated the structure and improved upon the health benefits of Hunza Water). The people of Hunza claim that their longevity, often living up to 130, stems primarily from their consumption of Hunza water, also known as 'glacial milk'. This technology has also been used to create what is truly the most unique spirulina, and other supplements, presently known.

A word to the wise: Winston Churchill said, "Most people occasionally stumble over the truth, but most pick themselves up and continue on as if nothing had ever happened". If you sense you've encountered some truths in this article don't use them as stumbling blocks. They may be stepping stones to far higher truths, if you dare to care and share. Data in this article can be legally offered only as "theoretical"; and no medical claims can be made or implied. See your health professional...

US Patent # 3,393,279

US Cl. 179/107 (16 July 1968)

Nervous System Excitation Device

Gillis Patrick Flanagan

This invention relates to electromagnetic excitation of the nervous system of a mammal and pertains more particularly to a method and apparatus for exciting the nervous system of a person with electromagnetic waves that are capable of causing that person to become conscious of information conveyed by the electromagnetic waves.

It is an object of the present invention to provide a means of initiating controllable responses of the neuro senses without applying pressure waves or stress waves to the ears or bones. Another object of this invention is to provide a means of causing a person to receive an aural perception of the sound corresponding to the audio modulation of radio frequency electromagnetic waves that are coupled with the nervous system of the person. These and other objects of this invention will be understood from the following drawings and description of the invention, wherein:

Figure 1 is a schematic illustration of one form of the present nervous system excitation device.

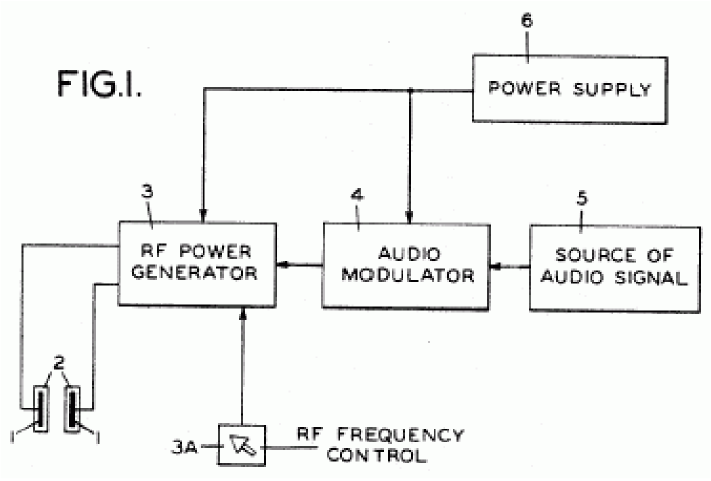
FIG.1.
6
POWER SUPPLY
3
RF POWER GENERATOR
4
AUDIO MODULATOR
5
SOURCE OF AUDIO SIGNAL
2
1
1
3A
RF FREQUENCY CONTROL

Figure 2 is a circuit diagram of one form of the present nervous system excitation device.

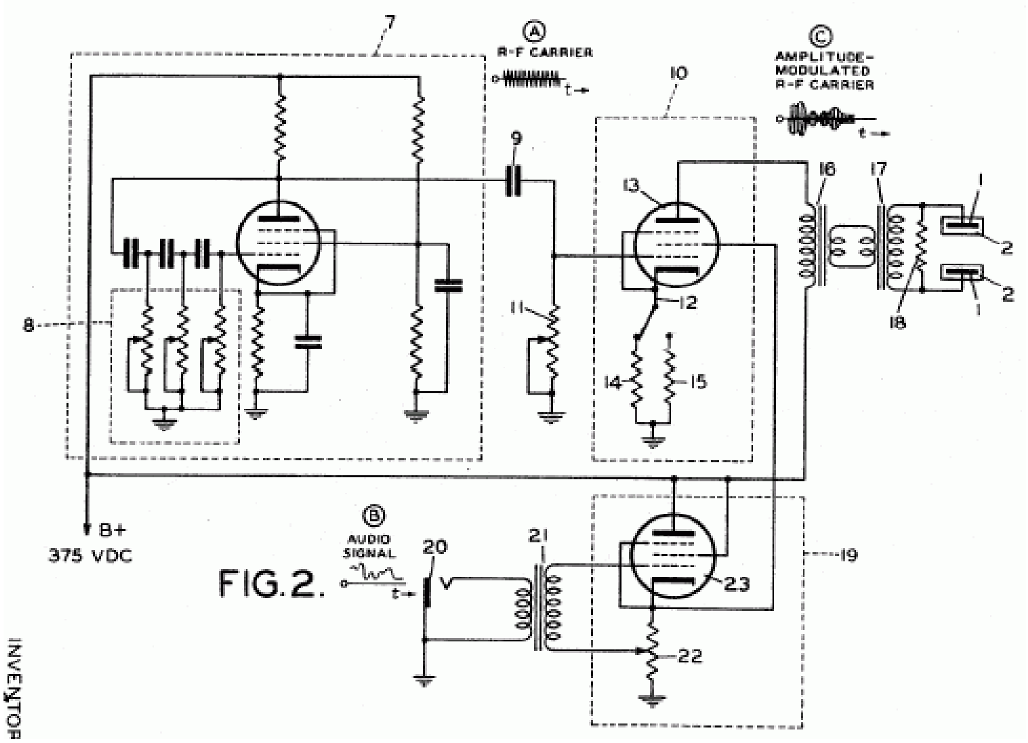
FIG.2.
7
Ⓐ
R-F CARRIER
t→
10
Ⓒ
AMPLITUDE-MODULATED R-F CARRIER
t→
9
13
16
17
1
2
2
1
18
12
11
8
14
15
B+
375 VDC
Ⓑ
AUDIO SIGNAL
t→
20
21
23
19
22
INVENTOR

Figure 3 is a diagrammatic view illustrating one form of field generator adapted to be used with the device of Figure 1.

FIG.3.

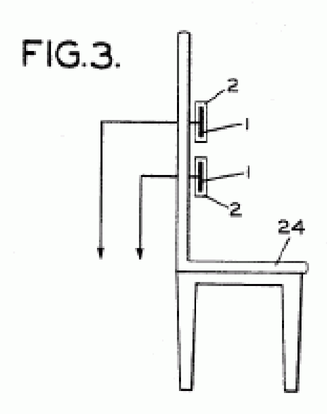

Figure 4 is a diagrammatic view illustrating another form of field generator adapted to be used with the device of Figure 1.

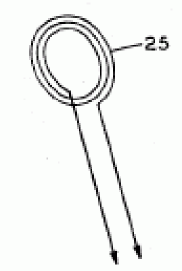

FIG.4.

The present invention involves the discovery that certain electromagnetic waves induce responses in the nervous systems of mammals. In human beings a response is produced when some or all of a person's nervous system is placed within a field of electromagnetic carrier waves of such a frequency, the nervous system is responsive to the modulation of the carrier waves. Each individual nervous system is at least somewhat selective in respect to the frequencies to which it is most responsive. A frequency to which the nervous system of a person is demonstratively responsive can be determined by varying the frequency of carrier waves that are modulated by an information signal, such as speech or music, and measuring the frequency of such waves that produce the sensation of hearing the sounds corresponding to the modulating signal.

In the method of the present invention, a response is initiated in the nervous system of a mammal by disposing at least a portion of that nervous system within a field of electromagnetic waves of a frequency above the aural range. In a preferred embodiment of this invention, at least a portion of the nervous system of a person is exposed to audio modulated electromagnetic waves having a radio frequency such that the person experiences the sensation of hearing, substantially free of distortion, the information which is conveyed by the audio modulation.

The present invention may be used as a hearing aid, as an aid to teaching speech to a person who was born deaf, as a means of communicating with persons in locations in which the noise level is high, as a device by which a person can listen to an audio signal that cannot be heard by others, etc.

As shown in Figure 1 of the drawing, in a preferred form of the invention, a field of electromagnetic waves is generated by a field generating means, such as a pair of electrodes 1. The electrodes 1 are preferably electrically insulated, for example by surrounding them with a suitable electrical insulating material 2, and are arranged to generate a field coupled with at least a portion of the nervous system of a person, for example by being placed near or along opposite sides of a person's head. The electrodes 1 can be placed in direct contact with the skin and the electrodes can be placed on or near various portions of the body, such portions preferably being near the spinal cord.

The electrodes 1 are electrically connected to a source of modulated electromagnetic waves inclusive of a radio frequency power amplifier and variable frequency oscillator, indicated in box 4, a source of audio signal, indicated in box 5, and a power supply for the signal source, modulator and amplifier, indicated by box 6. The variable frequency oscillator 3 is preferably provided with a manual radio frequency control means, indicated by box 3a. Numerous forms of the components, indicated in boxes 3 to 6, that provide suitable power and a source of modulated electromagnetic waves are presently known and the known devices can be suitably used as long as they are arranged to produce a relatively high voltage output that has a radio frequency above the audio range and is capable of being modulated by an audio signal or other signal adapted to be conveyed by the modulation of electromagnetic waves of such a frequency.

The modulation can suitably be effected by means of either an amplitude or frequency modulation of such electromagnetic waves. These waves preferably have a frequency in the range of about 20 kilocycles per second to about 200 kilocycles per second. The output of the source of modulated electromagnetic waves is preferably at least about 1 watt where the field generator comprises a pair of insulated electrodes placed on the head of a person. The extent to which a person is aurally perceptive to the output supplied at a given wattage is materially increased when at least one of the electrodes is placed in electrical contact with the body of the person.

In a preferred mode of operating the apparatus shown in Figure 1, the electrodes 1 are placed on the sides of the head of a person. The source 5 of audio signal is actuated to produce an audio signal corresponding to sounds recognizable by that person, and source 3 of modulated electromagnetic waves is actuated to couple the waves with the nervous system of that person. When control 3A is adjusted so that the frequency of the modulated waves is a frequency to which his nervous system is particularly responsive, the person to whom the field of such waves is applied has the sensation of hearing the sounds corresponding to the audio signal substantially free of distortion.

In the circuit shown in Figure 2, a phase shift type of carrier oscillation, generally designated by dotted rectangle 7, with a frequency control, generally designated by rectangle 8, is arranged to produce electromagnetic waves, shown at A, a frequency ranging from about 20 to 200 kilocycles per second. The oscillator output is coupled through capacitor 9 to a radio frequency power amplifier, generally designated by a dotted rectangle 10. Potentiometer 11, which is connected between capacitor 9 and ground, provides a means of adjusting the input to the amplifier. Switch 12, which is connected to the cathode of tube 13 of the amplifier, provides a means of switching between resistors 14 and 15 to vary the operating power characteristics of the tube.

The output of amplifier 10 is connected to transformer 16 which is coupled back-to-back with transformer 17. This arrangement of transformers provides an inductive load such that the amplifier yields a high voltage output is isolated from other components of the circuit. Resistor 18 connected across the output side of transformer 17 serves to reduce any dangerous voltage spikes which might be produced. The output side of transformer 17 is connected to a suitable field generator, which may comprise the electrodes 1 surrounded by insulating material 2.

The output of amplifier 10 is amplitude modulated by means of the modulator generally designated by dotted rectangle 19. A fluctuating electrical signal B, preferably of audio frequency, is applied to the modulator by means of input jack 20 and transformer 21. The output of the modulator varies the screen voltage of tube 13 of the amplifier so that the modulation envelope of the current oscillations C produced across the load of tube 13 correspond to the fluctuating signal B applied to the modulator.

Potentiometer 22 is connected to the cathode of tube 23 as the cathode resistor of tube 23. Potentiometer 22 is preferably adjusted so that the plate current of tube 13 of the amplifier so that the modulation envelope of the current oscillation C produced across the load of tube 13 correspond to the fluctuating signal B applied to the modulator.

Potentiometer 22 is connected to the cathode of tube 23 as the cathode resistor of tube 23. Potentiometer 22 is preferable adjusted so that the plate current of tube 13 is about half its normal maximum value. The fluctuating signal applied to modulator 19 is then adjusted to cause the plate current of tube 13 to vary between the maximum and minimum values so that a large current variation occurs in the load 16 of tube 13.

The apparatus shown in Figure 2 has been used to communicate speech and music to numerous persons including registered physicians. In these uses the electrodes 1, in the form of circular disc covered by a plastic insulation 2, were placed against the sides of the heads of the persons. When the electromagnetic waves were adjusted to a frequency to which persons having normal hearing were selectively responsive, none of these persons perceived any sensations of hearing or experienced any discomfort when no audio modulation was applied to the waves. When the waves were audio modulated with a speech or music signal, none of these persons experienced any discomfort, but they each had the sensation of listening to the transmitted information and hearing it at least as clearly as they would hear such information from an audible transmitter. When the same apparatus was similarly employed on a person whose hearing had been damaged to an extent requiring a hearing aid to hear normal conversation, that person heard the audio signal (with this hearing aid disconnected) and heard music with a better fidelity than that obtainable with this hearing aid.

Figure 3 shows an arrangement for mounting the field generating means in a position such that a portion of a person’s nervous system may be moved into and out of coupling with the field at the will of the person. In this arrangement, electrodes 1 surrounded by insulation 2 are mounted in vertical alignment along the back of a seating device, such as chair 24. When a person is seated and leaning back in the chair, portions of his nervous system are brought into coupling relationship with the field produced by electrodes 1.

Figure 4 shows an alternative arrangement of the field generating means. In this arrangement, inductive coil 25 is connected to the output of a suitable source of modulated electromagnetic waves and serves as a field generating means which is adapted to be placed around the head of a person.

It is to be understood that the above embodiments and examples have been presented for descriptive purposes and that, within the scope of the appended claims, the invention may be practiced otherwise than specifically illustrated and described.

US Patent # 3,647,970

US Cl. 179/1/5 (5 March 1972)

Method and Apparatus for Simplifying Speech Waveforms

Gillis P. Flanagan

Abstract ~

A speech waveform is converted to a constant amplitude square wave in which the transitions between the amplitude extremes are spaced so as to carry the speech information. The system includes a pair of tuned amplifier circuits which act as high-pass filters having a 6 decibel per octave slope from 0 to 15,ooo cycles followed by two stages, each comprised of an amplifier and clipper circuit. For converting the filtered waveform to a square wave. A radio transmitter and receiver having a plurality of separate channels within a conventional single side band transmitter bandwidth and a system for transmitting secure speech information are also disclosed.

Background of Invention ~

This invention relates generally to electronic processing of speech, and more particularly relates to a method and system for simplifying the speech waveform to facilitate transmission of the speech throughout various media without materially degrading intelligibility.

In the process of producing human speech, the voice box creates a series of sound pulses which reverberate within and are shaped by the upper throat and mouth cavity. The frequency of the pulses produced by the voice box primarily determines the frequency or pitch of the sound, while the shape of the mouth cavity reverberates and shapes the sound pulses to produce the speech information. The resulting speech waveform is very complex and highly redundant. If such a waveform is passed through a band-pass filter having a bandwidth significantly less than 3000 cycles per second, the speech becomes unintelligible. Thus, even the simplest voice communication channels require a substantial bandwidth. Heretofore it has been commonly believed that the speech information was contained in the amplitude as well as the frequency modulation of the speech waveform, When voice sounds are induced in a body of water or the earth, the many reverberations caused by the various velocity discontinuities make speech unintelligible over relatively short transmission lengths. Also, the complex speech waveform has made encoding or scrambling for secure transmissions, either by electromagnetic, electrical, or pressure waves, so impractical as to be very seldom used.

Summary of Invention Claimed ~

This invention is concerned with a method and system for simplifying a complex speech waveform so that it can be used for a multitude of applications. The simplified speech waveform may be passed through a narrow band-pass filter, thus permitting a greater number of communication channels within a given frequency band. The simplified speech waveform can be transmitted directly through the earth or water as a pressure wave and understand, either directly from the medium, or after simple amplification. The simplified waveform can be easily encoded by scrambling to provide secure voice communications. The simplified may be sued to operate machinery, produces more efficient public address systems and transmitters with greater range peak power for a given average power, and thus longer ranges.

In accordance with the present invention, the speech waveform is converted to a signal having substantially constant upper and lower levels with abrupt transitions from one level to the other, the abrupt transitions being in time correspondence to amplitude changes in the speech waveform that exceed a predetermined rate of change. This is accomplished by a system including a high-pass filter and means for converting the filtered waveform to a constant amplitude, substantially square wave.

More specifically, optimum results have been achieved by using a filter having a 12 decibel per octave slope from 0 to 15,000 cycles per second. In one specific embodiment, this filter is formed by a pair of tuned amplifier circuits each having a 6 decibel per octave slope within the frequency range of interest. In this embodiment, the speech waveform is preferably combined with a high frequency noise masking signal of lower amplitude prior to processing.

In accordance with another specific aspect of the invention, means for converting the filtered signal to a square wave comprises at least one amplifier followed by a clipper circuit.

The invention also contemplates a voice communication system having a plurality of separate channels within a bandwidth normally allotted for a single frequency, for example four channels within a bandwidth of 1500 cycles per second. In this system the processes speech is selectively passed through one of a plurality of narrow band-pass filters to a transmitter. The receiver has similar narrow band-pass filters so as to be selectively sensitive to transmissions in that pass band.

In accordance with another specific aspect of the invention, each transition of the square wave is converted to a pulse of predetermined amplitude and width, which is then converted into a plurality of pulses with predetermined time spacing. These pulses are then transmitted to a receiver where the plurality of spaced pulses are recombined as one pulse. The square wave is then reproduced from the combined pulses.

Brief Description of the Drawings ~

The novel features believed characteristic of this invnetion are ste forth in the appended claims. The invention itself, however, as well as other objects and advantages thereof, may bes be understood by reference to the following detailed description of illustrative embodiments, when read in conjunction with the accompanying drawings, wherein:

Figure 1 is a schematic block diagram of a system for processing a simplified speech waveform in accordance with the present invention;

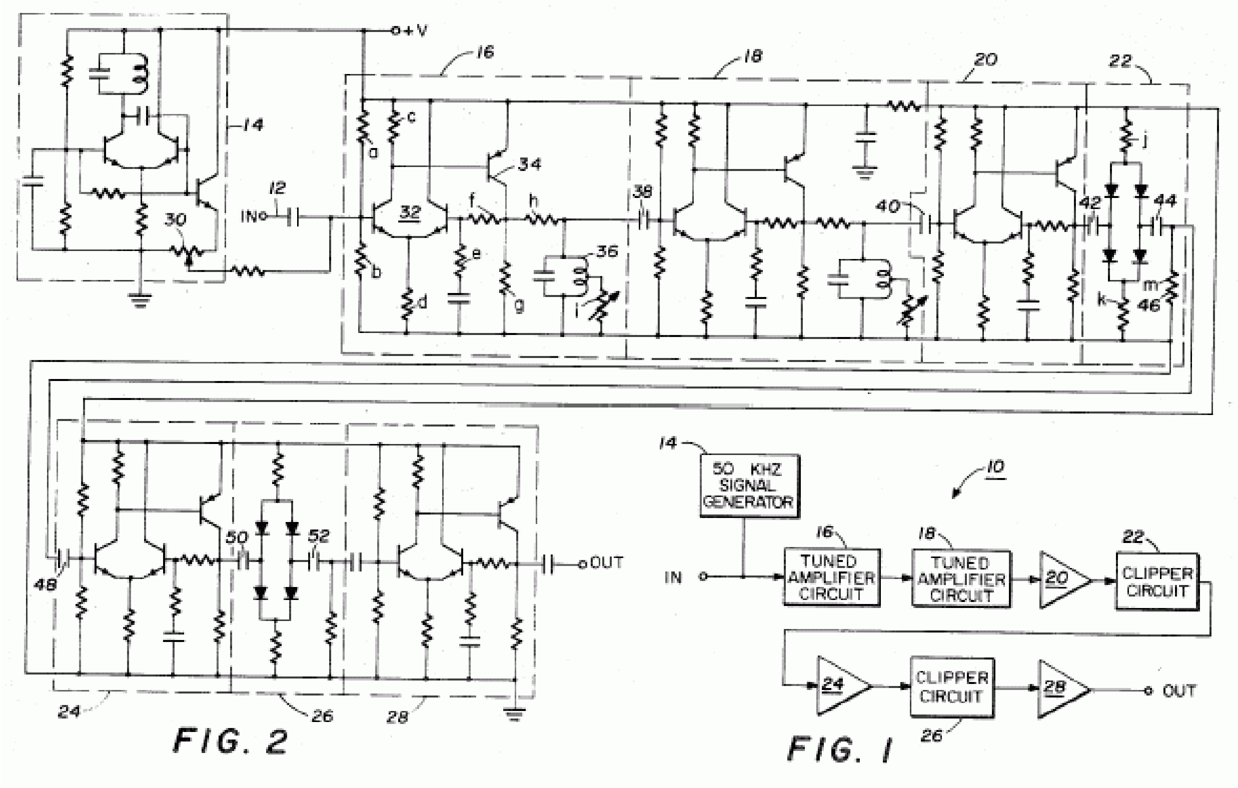
+V
16
18
20
22
14
12
IN
30
a
b
c
d
e
f
g
h
i
j
k
m
32
34
36
38
40
42
44
46
48
50
52
OUT
24
26
28
14
50 KHZ SIGNAL GENERATOR
10
16
TUNED AMPLIFIER CIRCUIT
18
TUNED AMPLIFIER CIRCUIT
20
22
CLIPPER CIRCUIT
IN
24
CLIPPER CIRCUIT
28
OUT
26

FIG. 2

FIG. 1

Figure 2 is a detailed schematic circuit diagram of the system in Figure 1;

Figure 3 is a schematic block diagram of a multichannel transmitter in accordance with the present invention.

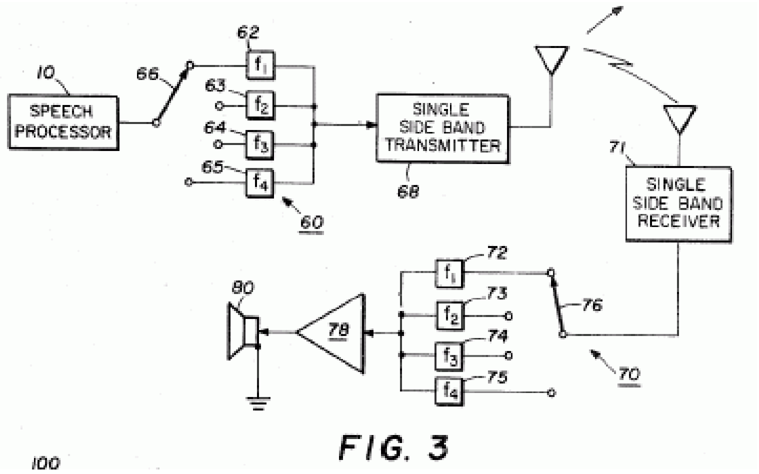
10
SPEECH PROCESSOR
66
62
f_1
63
f_2
64
f_3
65
f_4
60
SINGLE SIDE BAND TRANSMITTER
68
71
SINGLE SIDE BAND RECEIVER
72
f_1
73
f_2
74
f_3
75
f_4
76
70
78
80
100

FIG. 3

Figure 4 is a schematic block diagram of a system for transmitting and receiving scrambled speech waveforms in accordance with the present invention.

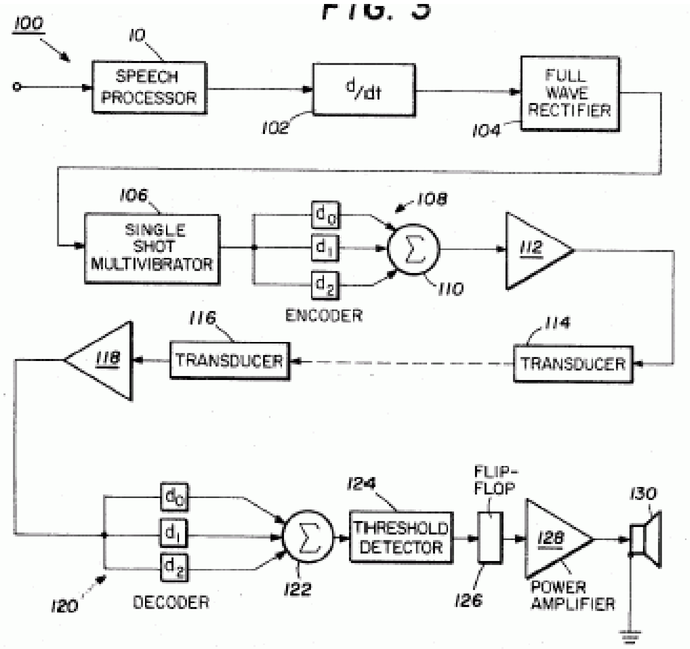

FIG. 4

Description of Preferred Embodiments ~

Referring now to the drawings, and in particular to Figure 1, a speech processor in accordance with the present invention is indicated generally by the reference numeral 10. The speech waveform is applied to the input 12 as a voltage signal derived from a microphone (not illustrated) or other suitable transducer. The speech waveform is summed with a much higher frequency, for example 50 KHz, masking signal produced by the signal generator 14. This signal is passed through a pair of tuned amplifier circuits 16 and 18. Each of the circuits 16 and 18 is a high-pass filter having a 6 decibel per octave slope from 0 to 15,000 cycles per second, thus providing a combined slope of 12 decibels per octave.

The filtered waveform is then passed through a circuit means for converting the filtered waveform to a square wave which is converting the filtered waveform to a square wave which is comprised of a first high gain amplifier 20, a first clipper circuit 22, a second high gain amplifier 24, and a second clipper circuit 26. The square wave is then passed through power amplifier 28 to the output in a form to drive a loudspeaker, transducer, radio transmitter or the like. When ultimately passed through a speaker, or other suitable transducer, the square wave is fairly intelligible. The square wave so produced has constant upper and lower levels, with very abrupt transitions between the upper and lower levels as a result of the two stages of amplification and clipping. The transitions occur in time correspondence to amplitude changes in the original speech waveform applied to the input 12 that exceed a predetermined rate of change so as to be passed through the high-pass filters 16 and 18.

A detailed circuit diagram of the system 1o us shown in Figure 2 wherein corresponding components are designated by corresponding reference numerals. Each of the components is of conventional design. The signal generator 14 has a variable load resistor 30 in the output stage which permits the amplitude of the masking signal to be adjusted to eliminate oscillations caused by noise. The amplitude of the masking signal should not be any greater than is required to prevent oscillation to minimize interference with the processing of the speech waveform. The tuned amplifier circuits 16 and 18 are of identical construction. Each is comprised of an amplifier having a differential input state 32 and a single output stage 34 which drives a tuned filter circuit 36. The tuned amplifier circuits 16 and 18 are coupled by capacitor 38, which of course also comprises an element of the filter. The amplifier 20 is identical to the amplifier portions of the tuned amplifiers 16 and 18, and is coupled to the output of tuned amplifiers 16 and 18, and is coupled to the output of tuned amplifier 18 by capacitor 40. The clipper circuit 22 is merely a diode bridge to the output of tuned amplifier 18 by capacitor 40. The clipper circuit 22 is merely a diode bridge coupled to the output of amplifier 20 by capacitor 42, followed by a filter comprised of capacitor 44 and resistor 46. The output of the clipper circuit 22 is coupled to the input of amplifier 24 by capacitor 50. Amplifier 28 is identical to amplifiers 20 and 24 and is coupled to the output of the clipper circuit 26 by capacitor 52.

In a typical embodiment of the circuit of Figure 2, the PNP transistors may be MPS3640 transistors, and the diodes may be IN914 diodes. The resistors have the following values in kilohms as referenced in circuits 16 and 22: a = 33, b = 33, c = 10, d = 33, c = 0.33, f = 33, g = 10, h = 10, I = 10, j = 100, k = 100, and m = 1.0. The capacitors are 10 microfarads, except for the capacitors in the LC tuned circuits which are 0.001 microfarads, All coils are 10 millihenrys.

The high-pas filters 16 and 18 may be of any suitable conventional circuit design, and may be a resistor-capacitor filter, a shortened delay line filter, or an inductor-capacitor filter, for example. The means for converting the filtered waveform to a square wave may also be any suitable conventional circuit such as a Schmidt trigger, or a very high gain amplifier which quickly saturates.

A multichannel speech transmission system in accordance with the present invention is indicated generally by the reference numeral 60 in the schematic block diagram of Figure 3. In the system 60, the speech processor 10 is selectively connectable to any one of four filters 62-65 by a selector switch 66. The outputs of the filters 62-65 are connected to the input of a conventional single side-band transmitter 68.

The filters 62-65 are narrow band-pass filters of any suitable conventional design having mutually exclusive pass bands of about 300 cycles centered at frequencies f1, f2, f3 and f4, and are grouped within a total bandwidth of about 1500 cycles, for example. Since 3000 cycles is a typical bandwidth for single side band transmitters operated for simple speech transmission, eight filters can be used if desired. The square wave produced by the speech processor10 may be selectively passed through any one of the narrow band-pass filters 62-65 without materially reducing its intelligibility.

The filtered square wave is transmitted by the conventionaltrnamitter68 to a conventional single side band receiver 70.The output of the receiver 70 is selectively connectable through filters 72-75 to a power amplifier 78 by a selector switch 76.The filters 72-75 have corresponding passbands centered at frequencies f1, f2, f3, and f4. The amplifier 78 may drive a speaker 80. Therefore, if the selector switch 76 of a particular receiving set 70 is set to the filter corresponding in frequency to the filter selected by switch 66 in the transmitter, the filtered square wave will be reproduced by the speaker 80 and will be sufficiently intelligible for nearly all voice communication purposes. However, if the selector switch 76 of a particular receiver is set to another frequency filter, no sound is produced by the speaker 80. Thus, the transmission system of Figure 3 provides four separate voice channels within the frequency band of 1500 cycles, or eight channels in the 3000 cycle bandwidth conventionally allotted for single side band operation. Of course,itisto be understood that the particular radio frequency is merely illustrative of the broader concept of the invention and that the same principles can be applied to transmissions through any media by electrical or electromagnetic waves.

A secure system for transmitting scrambled voice communications is indicated generally by the reference numeral 100 in Figure 4. Again the speech processor 10 is used to generate the square wave as heretofore described., The square wave is then passed through a differentiator 102 which produces a sharp spike pulse in time correspondence to each transition of the square wave. The sharp spike pulses have a polarity determined by the polarity of the transition and are therefore passed through a fullwave rectifier 104 which converts all of the spike pulses to the same polarity. The spike pulses are then used to trigger a single shot multivibrator 106 which produces a pulse of predetermined amplitude and time width in response to each spike pulse. The uniform pulses from the single shot multivibrator 106 are then passed through an encoder 108 which produces a plurality of pulses of corresponding width in a predetermined timed sequence in response to each input pulse. This may easily be accomplished by a plurality of parallel delay lines d0, d1, and d2 for transferring the pulses to point 110 at predetermined time intervals. The pulses are then amplified by an amplifier 112 which drives a transducer 114. The transducer 114 may induce the pulses in water, in the earth, or in any other propagating medium. Or, if desired, the transducer 114 can be replaced by a radio or other electromagnetic wave transmitter.

The transmitted pulses are received by an appropriate receiving transducer 116, which reproduces electrical pulses of corresponding width and amplitude. The received pulses are amplified by amplifier 118 and applied to a decoder 120. The decoder 120 is comprised of an identical number of delay lines of identical time relationship so that the three pulses are recombined as a single pulse summation point 122.

Each time that the three pulses occur at the same point in time, the sum of the pulses exceeds the threshold of a detector 124 which triggers a flip-flop 126. The output of the flip-flop is then a reproduction of the square wave originally produced by the speech processor 10. This square wave is then amplified by amplifier 128 to drive a speaker 130 and produce the voice communication. Reproduction of the voice communication can be accomplished only if the receiving decoder matches the transmitting encoder. The encoders and decoders can be easily interchanged in order to maintain secure transmissions.

Although preferred embodiments of the invention have been described in detail, it is to be understood that various changes, substitutions and alterations can be made therein without departing from the spirit and scope of the invention as defined by the appended claims.

[Claims not included here].

---

http://www.toolsforwellness.com/neurophone.html

Flanagan
Neurophone
DSP

click to enlarge

This fascinating technology was developed in 1958 by Patrick **Flanagan**. It was thought that sound was only heard by the inner ear; **Flanagan**, however, was able to demonstrate that the brain also hears sounds that vibrate on the skin.

To use this technology, you place the sensors on the forehead under a headband. Then plug the unit into your sound source, such as a CD player. The sensors cause your skin, the largest organ of the body, to vibrate to the music you’ve selected, much like a speaker vibrates.

By bypassing the ear, you are using a completely different part of the brain to process the sound creating new neural pathways. This technology is great for concentration while studying, learning languages, assisting in “new sound perception” great for listening to music in a whole new way, helping with meditation, relaxation, and healing.

When you first put the sensors on your skin, you feel a vibration and can actually hear the sound of the music you’ve selected. You can put the sensors on any part of your body, not just the forehead.

Just by wearing the sensors, you will notice a dramatic mood shift making you more positive, and also a change in your brain power, producing unwavering focus! The ultrasound it produces sends high frequency sound waves and their harmonics beyond the hearing threshold to the brain. It is these high frequency sounds that actually energize the brain, giving you more energy and focus. For music lovers, the clarity of the sound will astound you.

There are two products using this technology available, the **Neurophone** and the Echofone. Here are the differences:

Neurophone

Portability makes it useful for: Listening to lectures in the classroom ~ Using at work for powerful focusing ~ Focusing at sporting events

Echofone

Stereo Transducers makes it useful for: Serious musicians learning music ~ Music afficianados to experience incredible listening pleasure Sound therapy utilizing binaural beat frequencies

The **Neurophone** operates on an internal rechargeable battery making it completely portable, not requiring a plug. The internal pink sound generator makes it easy to use on-the-go because you don't need to plug it into a CD player.

The Echofone works in a similar manner, however is not portable. It is ideal however for music appreciation having a convenient control panel allowing you to easily mix the vibration intensity of the sensors with the volume intensity of the headphones.

---

http://www.neurophone.com

The New Neurophone GPF-1011 DSP

Dr. Patrick Flanagan's revolutionary ultrasonic neural stimulation instrument for brain entrainment, aiding in learning, relaxation and meditation

The NEW Neurophone® Golden Ratio Series is here!

The Neurophone® marks a culmination of innovation and engineering spanning 45 years. The GPF-1011 DSP offers a refined aesthetic design and advanced microprocessor control to create the finest experience in ultrasonic neuropathic stimulation. The Neurophone® is a precision scientific instrument with an extensive digital signal processor that encodes sound and modulates it into ultrasonic signals. The GPF-1011DSP offers the first, all digital Neurophone®. This profound digital engineering offers a quality of ultrasonic transmission that is more accurate and rich in harmonics than an analog representation, offering a more enriching and stimulating experience. Exciting new University studies are underway evaluating the Neurophone's® profound role in accelerated learning and memory retention. This instrument represents my life's work and I am greatly pleased to offer the new Neurophone® Model GPF-1011 DSP.

Contact us with your questions and comments!

---

http://home.dmv.com/~tbastian/files/cochlea1.txt

Another Look at the Neurophone

by

Rick Andersen

(6/21/97) Freely distributed for informational purposes

Dr. Pat Flanagan's Neurophone, in its several versions, has been the subject of intense interest to many experimenters. Later embodiments of his device appear to be based on the observation that our neural systems are set up to receive and decode external stimuli as sharp, transient spikes which represent a conversion of smoothly-varying analog sound pressure waves into time-coded "digital" nerve impulses. The associated propagation delays most likely do with sound what our TV screens do with serially-sequenced video waveforms: Convert them into a spatial distribution across the TV screen as a picture (sound waves encoded spatially across the brain cortex as a spatio-temporal pattern). The Neurophone is a first attempt at performing this complex conversion from frequency- to time-domain.

The idea of reducing full-fidelity speech/music into a differentiated, time-ratio coded version has many potential applications such as:

* Compression of audio or other analog signals
* Artifical cochlea for "androids"
* Alternative sensory stimulus ("hearing" through the skin) a la Neurophone
* Induction of information to the brain without contact (e.g., sharply-pulsed magnetic fields around the head such as Persinger's work

After reading the many files available on the Neurophone's history and many embodiments, as well as building a few experimental circuits myself, I've come to the conclusion that the simple "clipper/double-differentiator"-type circuit, as described by Tom Bearden in his book "Excalibur Briefing", is lacking as an effective way to convert sound to time-coded pulses, for the following reasons:

All sound waveforms (other than pure sine waves) are complex superpositions of several frequencies -- fundamentals and their harmonics, if they're periodic (pitched) sounds. The "tone color" is heavily dependent upon the relative amplitudes of the harmonics vs. the fundamental frequency.

Vowels, or diphthongs, in spoken language, are essentially very specific bandpass-filtered spectra. The vocal cords produce a harmonic-rich, raspy buzz, analogous to a sawtooth or pulse wave as seen on an oscilloscope. The mouth and nasal cavities act as resonators to selectively filter the buzz from the voice box into recognizable vowel sounds. "Fricatives" such as the "ssss" sound are unpitched bursts of air, "filtered" through the teeth, lips, etc. Fricatives contain many high frequencies clustered around 6-9 KHz, whereas a male voice is usually pitched (fundamental) around 200 Hz.

To run such a waveform into a clipper or squaring amplifier (comparator, Schmitt trigger) would be to take a very crude zero-crossing detection; the frequency with the largest amplitude is the fundamental... the higher harmonics "ride" upon it... so the fundamental is the frequency that gets to "call the shots" with the zero-crossing detector. So out comes a square wave, switching up or down through zero only at the grossest points of change along the waveform. We've already "hacked off" most of the "details"-- the information contained in the upper harmonics.

Now if we differentiate, and then "double-differentiate" this clipped wave, we will get narrow pulses which "track" the edges of the original complex sound wave ONLY WHERE IT CROSSED THROUGH ZERO on the scope. The problem is, we've clipped all the good stuff -- the harmonics -- off, and lost much of the information in the process. The speech now sounds like a person talking through a "kazoo" -- amusing, but not very useful.

How do we make square waves, and then narrow pulses (differentiated), out of a complex wave in such a way as to preserve DETAILED variations (upper harmonics)?

INTEGRATE, COMPARE, DIFFERENTIATE, and SUM

The method I am outlining here is a result of my own brainstorming on the problem; if I have stepped on anyone's feet by re-inventing your ideas, my apologies, but rest assured that I'm not looking for patent rights, just circuit functionality. Maybe I've discovered Flanagan's REAL method???...

First, we need to appreciate the fact that there are millions of neurons doing their thing in our auditory system, not just a few. Accordingly, I'm going to outline a simple circuit that will produce time-coded spikes, but I'm recommending that several of the circuit modules be built and chained together SEQUENTIALLY (analog sections), with their differentiated SPIKE outputs all SUMMED together and perhaps clipped as a group so that there is one amplitude for them all, but varying repetition rates -- sort of like Pulse-Position Modulation but more complex. This is the output that will connect to the piezo skin transducers after suitable amplification.

The cochlea has many, many neurons working on a serial sound "stream" at the same time. Some are positioned "up front" where the sound enters from the outside, some are down the trough a ways, and some are located at the far end. As such, there is a built-in time delay, or acoustic propagation delay, from one end of the membrane to the other. Carver Mead at Cal-Tech has attempted to model this by means of a series of transconductance amplifiers wired as simple delays with carefully-calibrated frequency "damping" per section.

My version models this with a simple R-C lowpass filter (integrator) and an op amp (buffer) per module. The first stage or module will have the highest cutoff frequency, approximately 5 KHz, and will set the resolution limit of the system. The output of the buffer, which reproduces the slightly low-pass filtered (and DELAYED) version of the original audio signal, goes on to the input of the next module.

The next module's R-C filter has a slightly lower cutoff frequency, say, 3 KHz. And so on, down to maybe 200 Hz. (Probably best to taper the cutoff frequencies LOGARITHMICALLY for economy and accuracy.)

Back to the first module, which had the highest cutoff frequency of the chain of modules we're building up... say, about 5KHz:

Across the series resistor forming part of our RC filter, we tap off two wires and run them into a comparator (op amp with no feedback resistor... = squaring amplifier). The input side of the resistor (before being low-pass filtered) goes to the (+) input of the comparator; the other lead of the resistor (which connects to the filter cap and then to the buffer input) goes to the (-) input of the comparator.

The comparator, therefore, constantly compares the two versions of the audio waveform at its terminals: the unfiltered/undelayed input vs. the filtered/delayed output side of the RC filter. When the input side is rising in amplitude, and exceeds the delayed output side, the comparator switches its output logic HIGH; when the input signal amplitude dips down below the delayed signal amplitude, the comparator switches LOW. THIS PRODUCES A VERY PRECISE "PEAK-CROSSING" DETECTOR WHICH TENDS TO SWITCH AT THE +/- PEAKS OF A SINE WAVE, rather than at the zero-crossings at the middle of the waveform.

Now we run this squared-off comparator output into a DIFFERENTIATOR --- a series cap (about .001uf) and a resistor to ground (10k) -- which hi-pass filters the square wave into sharp spikes having approximately a 10 uS time constant, nice and short. We have now generated a series of bipolar spikes whose position, relative to one another over time, is a function of the DIFFERENCE of amplitudes and phases of two slightly different copies of the original incoming sound wave.

A similar mechanism has been discovered within the retinal cells in the eye. Carver Mead has reproduced this one, too. I.e., a retina cell "computes" the [logarithm of the] DIFFERENCE between what it "sees" and what its immediate neighbors see -- and it tries to inhibit their outputs, producing automatic EDGE DETECTION which can be thought of as spatially high-pass filtering a visual scene; the result is that CHANGING brightnesses are emphasized. Play with a kitten and you'll notice a direct proportion between how fast you pull the yarn past its visual field and how badly the kitten feels it "has to" pounce on it. And frogs tend to ignore everything except quickly-moving black spots (flies) across its visual field.

This is likely the mechanism of neural coding by which we hear, also (or something like it). Living things tend to respond to CHANGES in sensory stimulus; long, monotonous stimuli lead to neural shutdown (boredom --> sleep) as any school teacher can tell you.

My version of Flanagan's Neurophone preserves several frequency-bands worth of harmonic information... but as time-domain spikes just as Flanagan says we need. And we don't hack the most interesting stuff off the top with this approach, like we do with simpler versions.

INTEGRATE, COMPARE, DIFFERENTIATE, and SUM -- A SUMMARY

Visualise the human cochlea as a long delay-line with damping (low-pass filtering) progressing as you go down the line. Picture a comparator every few millimeters, with its two inputs slightly separated, "sampling" across a small section of the delay line as the waves propagate down the line. An entire array of such comparators, with their outputs differentiated into10 uS spikes, perhaps "OR'd" together, would present a highly detailed PPM spike-train which, I predict, would produce an increased efficiency in frequency-to-time or analog-to-digital sound conversion for the next generation of homemade Neurophone-like devices.

HERE IS A QBASIC PROGRAM WHICH WILL RUN A SIMULATION OF THE ABOVE ON YOUR COMPUTER SCREEN.... drag your mouse across the following code, COPY it and SAVE AS a new file called COCHLEA1.BAS. Then, RUN it from your resident QBASIC environment.

The program generates a rippled, sawtooth-like "audio" waveform (white) made up of a fundamental and 10 harmonics, and then samples it several times, each time having a slightly lower cutoff frequency (longer time delay, more low-pass filtering). This is displayed in yellow.

The comparator output for each delay is plotted in green.

The superimposed differentiated spikes from all the comparators are plotted at the bottom of the screen in white.

Hope these ideas stimulate further research and development.

```
' ------------------------------------------------------------------
' cochlea1.bas  6/21/97 by Rick Andersen

DECLARE FUNCTION integrate (sample, cap, tc)
DECLARE FUNCTION differentiate (sample, cap, tc)
COMMON SHARED accumulator
CONST pi = 3.14159
SCREEN 9: CLS

FOR timeConstant = 10 TO 600 STEP 100
accumulator = 0

FOR a = 0 TO 22 STEP .01
wave = 0
FOR h = 1 TO 10
wave = wave + SIN(a * h) / h
NEXT h

lopass = integrate(wave, cap1, timeConstant)

IF wave > lopass THEN
trigger = 1
ELSE
trigger = 0
END IF

hipass = differentiate(trigger, cap2, 20)

PSET (a * 30, 50 - wave * 20), 15
PSET (a * 30, 50 - lopass * 20), 14
PSET (a * 30, 100 + timeConstant / 4 - trigger * 15), 2
PSET (a * 30, 270 - hipass * 20), 15

NEXT a
NEXT timeConstant

END

FUNCTION differentiate (sample, accumulator, tc) STATIC
fsample = tc
```

```
leakage = 1 - EXP(-2 * pi * 1 / fsample)
capAvg = leakage * accumulator
accumulator = accumulator - capAvg + sample
differentiate = sample - capAvg
END FUNCTION

FUNCTION integrate (sample, accumulator, tc) STATIC
fsample = tc
leakage = 1 - EXP(-2 * pi * 1 / fsample)
capAvg = leakage * accumulator
accumulator = accumulator - capAvg + sample
integrate = capAvg
END FUNCTION

' -------------------------------------------------------------------
```

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

Rex Research, POB 19250, Jean, NV 89019 USA

United States Patent [19]
Nelson et al.

[11] **3,728,937**
[45] **Apr. 24, 1973**

[54] **GAS FIRED CASELESS AMMUNITION RIFLE**

[76] Inventors: **Bruce C. Nelson**, 6304 Martin, Detroit, Mich. 48210; **Alphonse G. Scibor**, 6540 Longacre, Detroit, Mich. 48228

[22] Filed: **Apr. 30, 1971**

[21] Appl. No.: **138,925**

[52] **U.S. Cl.** **89/7**, 89/26
[51] **Int. Cl.** **F41f 1/04**, F41f 11/00
[58] **Field of Search** ....42/1 R, 84; 89/7, 89/26

[56] **References Cited**

UNITED STATES PATENTS

1,291,674	1/1919	Brannon	89/7
1,237,157	8/1917	Bellairs	89/7
2,094,854	10/1937	Smith	89/7
1,596,057	8/1926	Mallory	89/7
3,427,924	2/1969	Johnsen	42/84
1,290,596	1/1919	Lewis	89/7
3,343,455	9/1967	Lohr	89/7

Primary Examiner—Benjamin A. Borchelt
Assistant Examiner—C. T. Jordan
Attorney—Edward M. Apple

[57] **ABSTRACT**

A caseless ammunition, bolt action, rifle charged with gas and fired electrically. The firing chamber is sealed with a truncated cone insert, having thin walls which are contracted around the bolt by gas pressure, to seal against back pressure leakage.

1 Claim, 4 Drawing Figures

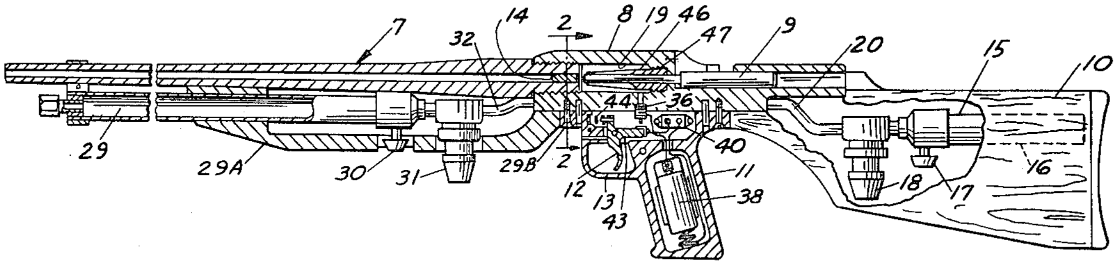

PATENTED APR 24 1973

3,728,937

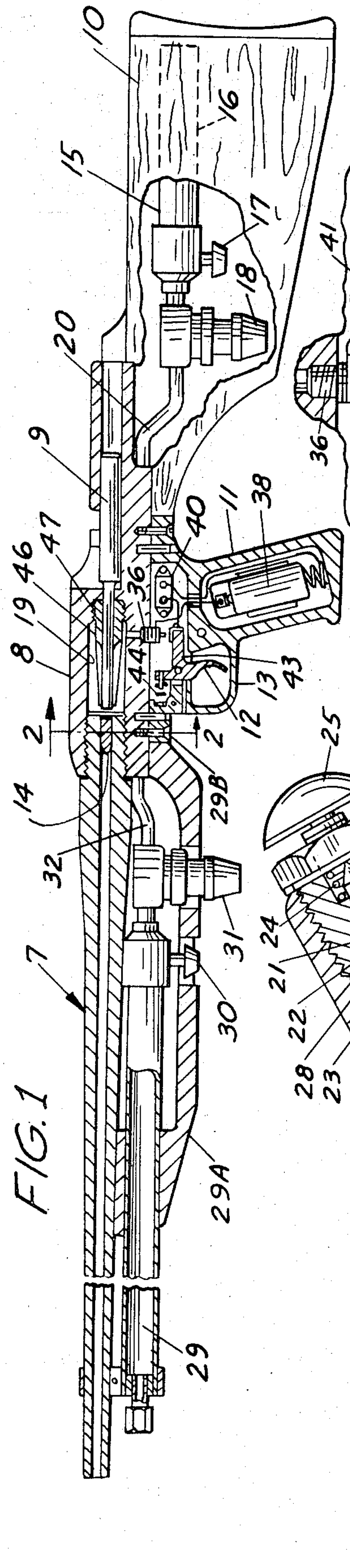
10
16
15
17
18
20
9
11
38
40
36
46
47
19
8
44
43
13
12
2
2
29B
14
32
31
30
7
29A
29

FIG. 1

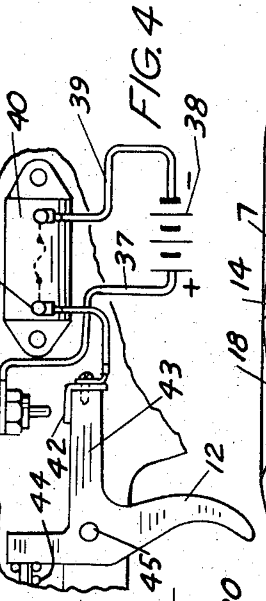
40
39
38
41
37
+
43
42
36
44
45
12

FIG. 4

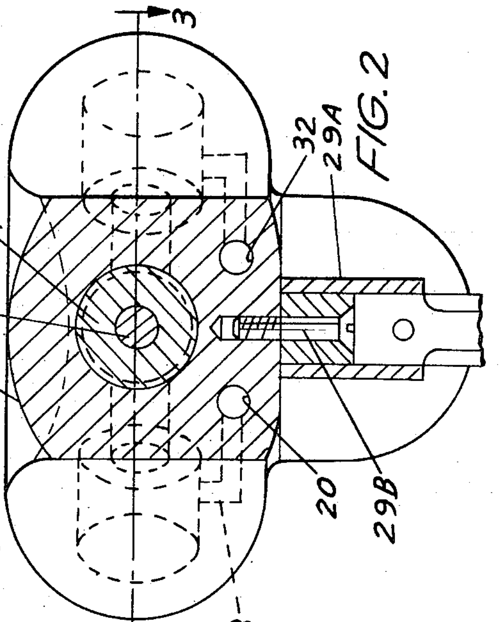
3
32
29A
7
14
18
20
29B
28
3

FIG. 2

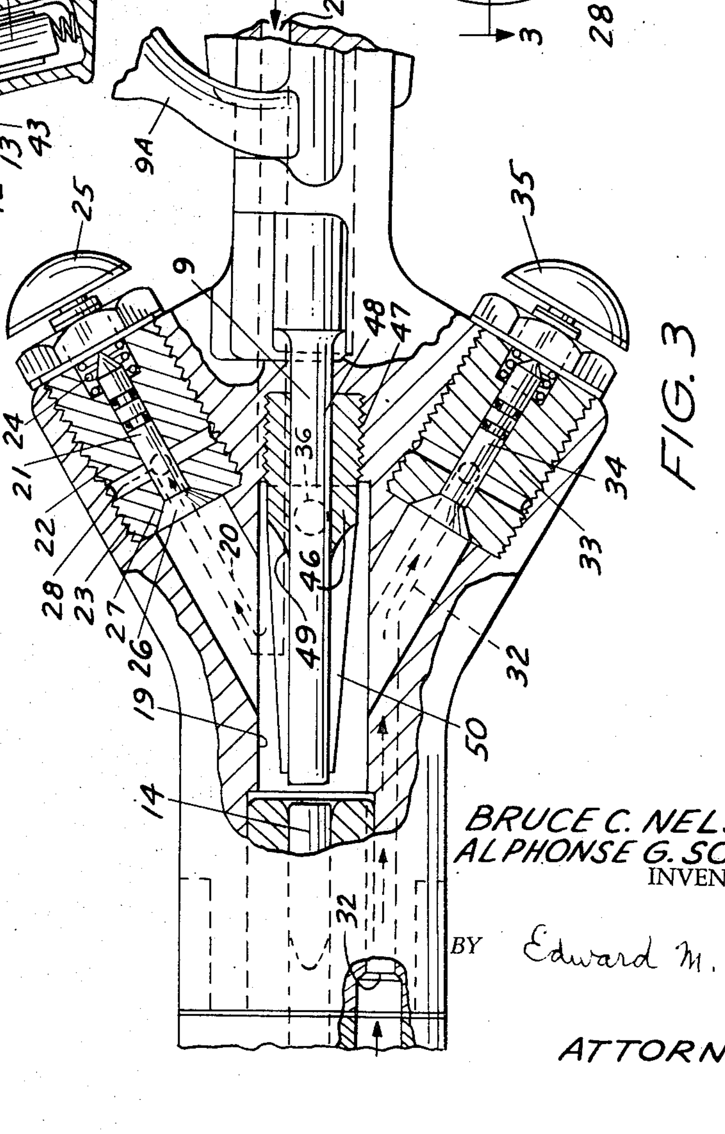
20
9A
25
35
9
48
47
24
21
22
28
23
27
26
36
20
46
49
19
50
32
34
33
14
32

FIG. 3

BRUCE C. NELSON
ALPHONSE G. SCIBOR
INVENTORS

BY Edward M. Apple

ATTORNEY

GAS FIRED CASELESS AMMUNITION RIFLE

This invention relates to a rifle designed to shoot caseless ammunition by charging the rifle with gas, and firing the charge electrically.

An object of the invention is to provide a rifle which will not jam after firing, because there is no empty cartridge shell or case to be ejected.

Another object of the invention is to gas charge, and electrically fire, a rifle in much the same manner as an internal combustion engine is operated. In this instance a projectile is substituted for the engine piston.

Another object of the invention is to provide a rifle which may be fired with hydrocarbons and oxygen, both in a gaseous state. This has various advantages over rifle using metal cases in that it is more economical to shoot and there is no need of buying expensive metal cased ammunition.

Another object of the invention is to provide a rifle which may quickly be adjusted to provide different amounts of foot pounds of muzzle energy, so that the same rifle can fire at long range, medium range, and for target practice with less wear and tear on the device and less fatigue to the operator.

Another object of the invention is to provide a device of the character indicated, which is provided with integrated receptacles for holding the gases which are to be mixed in the firing chamber.

Another object of the invention is to provide a device of the character indicated, which may be supported in the arms of the operator, or may be used as a light machine gun, in which event the gas tanks could be detached from the gun and carried separately.

Another object of the invention is to provide a device which may be mounted on a stationary member, or may be mounted on a mobile unit, or an airplane, with the tanks incorporated into the structure of the machine.

Another object of the invention is to provide a device of the character indicated, which fires clean and emits only CO_2 and H_2O.

Another object of the invention is to provide a rifle which may be fired electrically, or with caps, or with the compression of its own gases.

Another object of the invention is to provide a device of the character indicated, which is provided with a truncated cone insert in the firing chamber, which insert has comparatively thin walls which are contracted about the bolt by the exploding gases to prevent leakage of the gas around the bolt.

Another object of the invention is to provide a device of the character indicated, which may be manufactured and assembled inexpensively, in as much as it has few machined parts.

Another object of the invention is to provide a device of the character indicated, which will warm itself in cold weather and will facilitate its use under adverse weather conditions.

The foregoing and other objects and advantages of the invention will become more apparent as the description proceeds, reference being made from time to time to the accompanying drawing, forming part of the within disclosure, in which drawing:

FIG. 1 is a longitudinal, sectional view taken through a rifle embodying the invention.

FIG. 2 is an enlarged section taken substantially on the line 2—2 of FIG. 1.

FIG. 3 is a section taken substantially on the line 3—3 of FIG. 2.

FIG. 4 is an enlarged detail illustrating a typical wiring circuit when the device is fired electrically.

Referring now more particularly to the drawing, it will be understood that in the embodiment herein disclosed, the reference character **7** indicates the barrel of the rifle which has a breech block or receiver **8**, a bolt **9** having a handle **9A**, a stock **10**, hand grip **11**, trigger **12** and trigger guard **13**, all of which are conventional parts of a rifle and constitute no part of the invention except as they are combined with the elements hereinafter described.

The rifle embodying the invention is designed to shoot a bullet, or projectile **14**, which may be of any desired caliber and which is fed into the barrel **7**, by the bolt **9**, as in conventional, single action practice. The rifle may also be equipped to load the bullets automatically, although no claim is made to such mechanism in this application.

Instead of using a brass cartridge to carry the explosive charge, we intend to fire this rifle by means of a mixed charge of hydrocarbons gas (preferably butane) and compressed oxygen.

In the embodiment shown in FIG. **1** a compressed butane cylinder **15**, is housed in a suitable bore **16**, formed in the stock **10**. The cylinder **15** is provided with a shut off valve assembly **17** and a pressure regulator **18**. The cylinder **15** is connected to the firing chamber **19** through a passage way **20**, which passageway **20** is controlled by means of a valve assembly **21** and **22**, which is threadedly mounted on a lateral extension of the breech bolck **8** as at **23**. The valve member **21** is spring loaded as at **24**, and may be opened by striking the button **25** with the palm of a hand. The valve stem **21** has a head **26** which engages a seat **27** and when the head **26** is lifted from the seat **27**, the gas is permitted to enter the chamber **19** through the ports **28**.

A compressed oxygen cylinder **29** is supported by any suitable means **29A** and **29B**, under the barrel **7**. The oxygen cylinder **29** is provided with a shut off valve **30** and regulator mechanism **31** and communicates with the firing chamber **19** through a passageway **32** which is controlled by a valve assembly **33** and **34** having a palm button **35** similar to the mechanism **21–28** previously described in connection with the butane gas assembly.

Both the butane and oxygen being under compression, it is simply necessary to strike the buttons **25** and **35** to admit an explosive mixture into the firing chamber **19**. The explosive mixture in the firing chamber may be fired electrically, by means of a glow plug **36**, which is connected as at **37** to one side of a battery **38** (FIG. **4**), the other side of the battery **38** being connected as at **39** to a safety fuse **40**, which in turn is connected as at **41** to a contact element **42**, which is mounted on a lever **43**, which is part of the trigger mechanism **12**. The contact **42** is normally held away from the glow plug **36** by a spring **44**, which holds the trigger mechanism **12** in vertical position on the pivot **45** as shown in FIG. **4**.

The firing chamber **19** is provided with a frusto conical insert member **46**, which is threadedly received, as at **47**, in a suitable bore formed in the breech block **8**. The truncated cone member **46** has a central bore **48** in which is received a reduced portion of the bolt **9**. The member **46** is also provided with comparatively thin walls around the bolt **9**, as at **49**, so that upon the firing of the charge from the chamber **19** the back pressure of the explosive gases contract the walls of the member **46** tightly around the bolt **9**, to prevent escape of gases rearwardly from the chamber **19**. The member **46** is also provided with radial webs **50**, which trap the exploding gases and prevent them from swirling around the bolt **9** to more effectively exert pressure on the thin walls **49** to assist in sealing the firing chamber around the bolt area.

From the foregoing, it will be understood that by regulating the pressures in the passageways **20** and **32**, it will be possible to effect a greater or lesser explosive charge in the chamber **19**.

It is believed that the operation of the device is obvious from the foregoing description.

Having described our invention, what we claim and desire to secure by Letters Patent is:

1. In a rifle having a barrel, breech block, stock, hand grip and trigger assembly, the combination of a firing chamber in said breech block, means to seal said firing chamber against back pressure, means to feed caseless bullets to said barrel, a first compressed gas chamber, means to conduct gas from said first gas chamber to said firing chamber, a second compressed gas chamber in communication with said firing chamber, means to control the flow of gas from each of said gas chambers, and means to fire the mixture of gases in said firing chamber, to eject the bullets from said barrel, in which the means to seal said firing chamber include a truncated cone insert received therein, said cone having thin walls arranged to be compressed about the bullet feed means to seal the area around said feed means against gas leakage, upon firing the gas mixture in said firing chamber.

* * * * *

rexresearch.com

George SQUIER

Tree Antennas

Save up to $50 + free delivery on your first order* *See terms & conditions. SHOP NOW > Ad ×

Save up to $50 + free delivery on your first order* *See terms & conditions. SHOP NOW > Ad ×

Save up to $50 + free delivery on your first order* *See terms & conditions. SHOP NOW > Ad ×

General George O. Squier

***Scientific American* (July 14, 1919, p. 624) ~**

"With Trees For Ears"

A Wireless Station Within the Reach of Everybody

With a pair of receives to his ears, an amazed visitor to a certain radio station heard a high-toned hum which changed to a low growl, then skied to the upper reaches of the musical scale in a

faint, very faint buzz, as if some microscopic mosquito had had his song made audible. The operator rapidly rapidly turning the knobs on his couplers and condensers, raised his hand: suddenly, through the changing radio signals which were clamoring for attention together in the receivers came his voice; "There --- the loud, easily heard one is New Brunswick; the fainter, lower one is Nauen, in Germany".

If all this had taken place in the great Arlington station one would not have wondered, save perhaps at the inability to tune out all radio but Nauen. But it was a little portable house erected in thick woods near the edge of the District of Columbia and the signals were received through an oak tree for an antenna.

It is not a joke nor a scientific curiosity, this strange discovery of Gen. George O. Squire, Chief Signal Officer, that trees --- all trees, of all kinds and all heights, growing anywhere --- are nature's own wireless towers and antenna combined. The matter first came to his attention in 1904, through the use of trees as grounds for Army buzzer and telegraph and telephone sets, which, in perfectly dry ground and in a dry season, functioned poorly or not at all with ordinary grounds. Right then he began experiments with a view to seeing what possibilities, if any, the tree had as an aerial. But in 1904 radiotelegraphy was far more undeveloped than at present, and vacuum amplifying tubes were not thought of.

During the war the Signal Corps established a chain of special receiving stations in different localities to copy and record enemy and allied radio messages. Some of those stations were instructed to test the efficiency of growing trees as receiving antennae.

With the remarkably sensitive amplifiers now available, it was not only possible to receive signals from all the principle European stations through a tree, but it has developed beyond a theory and to a fact that a tree is as good as any man-made aerial, regardless of the size or extent of the latter, and better in the respect that it brings to the operator's ears far less static interference.

This is a rather broad statement, yet there beyond the Capital of the nation stands a little portable house, the oak tree, a small receiving set and a couple of enlisted men and an officer on duty; and the curious may, with permission, hear for themselves that the signals so received are neither faint not interrupted, but strong, full-toned dots and sashes even when they come from far-off Nauen. Page after page is copied daily from the propaganda material which Nauen sends out by the ream. Lyons, Poldhu, ships at sea, even the NC-4 on her way, are heard plainly. As for New Brunswick or nearby Arlington --- they deafen the listener if he is unwise enough to try to "take" them otherwise than with the phones lying on the tables.

It will puzzle the amateur as it has puzzled the experts, how a tree, which is certainly well grounded, can also be an insulated aerial. The method of getting the disturbances in potential from treetop to instrument is so simple as to be almost laughable. One climbs a tree to two-thirds of its height, drives a nail a couple of inches into the tree, hangs a wire therefrom, and attaches the wire to the receiving apparatus as if it were a regular lead-in from a lofty copper or aluminum aerial. Apparently some of the etheric disturbances passing from treetop to ground through the tree are diverted through the wire --- and the thermionic tube most efficiently does the rest.

It is interesting to learn that the tree behaves very much like any other aerial; it receives better in dry clear weather than in muggy, damp weather. It plucks messages from the ether more clearly at night than in the day. It is affected very little by rain. It is affected not at all by the presence of other trees; so far as has yet been ascertained it makes little difference whether one drives his nail in a tree in the forest or a lone tree on the plain. Certainly it makes no difference that amounts to anything whether the tree be just an ordinary tree or a giant; it was a 60-foot oak over which the very awe-struck correspondent heard Nauen telling a waiting world what good people the Germans really are. And to prove that it made no particular difference what kind of tree was used the officer in charge switched to a pine tree, which received equally well.

A dead tree will not do, and a tree not in leaf is not so sensitive as one in full foliage. It makes much difference where the nail is driven. General Squier calls the proper place the optimum point, and experimentally it has been determined that two-thirds of the distance from ground to top is the best place -- in a 60-foot tree, 40 feet from the ground.

One nail is sufficient, and it may be any kind of nail; but copper is preferred as not rusting. In practice, if a tree station is to be at all permanent, several nails would be driven and connected to the same wire, each additional nail up to 6 or 8 making the diverted current a little stronger. But 40 nails apparently produce no clearer signals than half a dozen.

The tree may serve as a receiving station for several sets, either connected in series with the same material or from separate terminals.

Some skeptics have expressed the belief that it was not the tree, but the wire leading to the nail in the tree which was the real aerial. The absurdity of thinking a 40-foot wire could receive the widely differing wave lengths which come through the tree station is obvious, but to set any doubt at rest, the wire to the tree has been hung to the nail by means of an insulator, when the signals immediately cease, only to come in as strong as ever just as soon as the connection is again established.

Just what will the tree do as a transmitting station for radio telegraphic messages has not been determined in the Signal Corps Experimental Laboratory. As those in charge express it, "The fact has been demonstrated, but the matter is still in laboratory stage only. What remains to be done now is to develop the best methods of using the demonstrated fact".

But it has already been shown that the tree can be used in wireless telephony and for short distances it has been shown that two-way telephonic communications is easily established through trees with remarkably low values of transmitting antenna current.

If a tree may be used to send wireless telephonic waves it seems not unreasonable to suppose that it will do so as easily with the telegraphic waves. At present the Signal Corps is at work on apparatus to test the possibilities of the tree as a transmitting station.

Just what this development of the art of radio telegraphy may mean has not yet been worked out. It is the history of most discoveries that their potentialities are hardly dreamed of when they are first made --- for instance the telephone, the electromagnet, the vacuum tube amplifier. But it seems fairly obvious that in war, at least, the tree receiving station opens up great possibilities.

True enough there are few trees which remain intact under shell fire, and doubtless with this possibility in mind the armies of the future (if there be such) will in action consider all trees as dangerous enemy aerial stations. But there will always be trees behind the lines and not all actions will be fought on bare ground. What would it have meant to the "lost battalion" to have had a tree wireless set along by which it could have heard that every effort was being made to find and relieve it, or by which it might have sent back messages supplementing that carried by the pigeon?

The greatest development, however, of the tree as the foundation for a receiving and possibly a sending station will come in peace uses. General Squier has written:

"In view of what has been accomplished in space telegraphy, it is difficult to predict to what extent this means of communication may ultimately be developed. If, as indicated in these experiments, the earth's surface is already generously provided with efficient antennae, which we have but to utilize for such communication, even over short distances, it is a fascinating thought to dwell upon in connection with the future development of the transmission of intelligence.

"Since a transmitting station is a central point for electromagnetc waves sent out in all directions over the surface of the earth, a large class of information, such as meteorological reports, crop reports, and general news items of interest to all, may in time be sent from central points, to be received in many places within the radius of influence of the signal station, and this, too, by the simplest form of apparatus."

The amateur wireless world will unquestionably take an intense interest in the tree radio work. At present, while the government has lifted the ban upon amateur aerials, it has not removed the structures against sending.

The aerial is always the greatest problem for the amateur. Lack of both money and material prevents him from erecting anything very large or of very great capacity. If any lad with a receiving set and some thermionic tubes can hook to a tree and take in any wavelength he can tune to, will not tree radio vastly increase the devotees of this particular variety of indoor sport? The matter is one of some importance, inasmuch as many valuable recruits to the radio world have come from amateur ranks, and many a radio engineer had got his first taste for the fascinating art through a homemade tuning coil and detector, under the attic roof. The greater the amateur wireless world, the quicker the development of the art as a whole.

Explorers, discoverers, engineers in far places, the forest service, the woodsman, all have use for the new development. Moreover the tree as an antenna offers unusual possibilities for the investigation of atmospheric phenomena and for what may be called the physics of botany (or the botany of physics) and perhaps is the road by which the unsolved puzzle of growth may be studied.

Meanwhile, it is a thought not without great power to move the sensitive imagination that every tree, growing everywhere, is a wireless tower and antenna and that, as General Squier says, "It is significant that a tree, possessing utility and natural strength, architectural beauty of design and endurance far superior to artificial structures prepared by man, should be able yet further to minister to his needs".

***Electrical Experimenter* (July 1919), p. 204**

Talking Through The Trees

by

Major-General George O. Squier

(Chief Signal Officer, US Army)

How Transatlantic Radio messages Are Copied Via Tree Antenna

As long ago as 1904, the author conducted some experiments with a view to utilizing growing trees as antennae for radio-telegraphy and discovered the efficacy, in a general way, of using a direct metallic contact to certain trees (principally Eucalyptus) to increase the audibility of radio signals. My attention was first called to this phenomenon during the course of summer maneuvers of the Army at Camp Atascadero, CA, where, due to the prevalence of the dry season of the soil, it was found that the regular Army buzzer telephone and telegraph sets were inoperative with any ordinary ground or earth but became operative when connected to a metallic nail driven in the trunk or roots of a live tree. This incident led the author to pursue the subject experimentally in the autumn of 1904 continuing the experiments to the range of frequencies than employed in radio-telegraphy.

Tree Antennae

In connection with the organization and development of Transatlantic radio reception, which was carried out during the period of the war to provide against the possibility of the interruption of the submarine cable system, the Signal Corps established a chain of special receiving stations in different parts of the United States to copy and record enemy and Allied radio messages from European stations for the information of our Army General Staff.

In the prosecution of this work, directions were given to the Signal Corps Laboratory at Camp Alfred Vail, Little Silver, NJ, and also to the experimental staff in Washington to test the efficiency of growing trees as receiving antenna, in connection with this service, using the vastly superior technique and facilities now represented in the radio Art as compared with the crude apparatus with which the discovery was made in 1904. With a collection of apparatus representing the most advanced state of the radio Art, the problem, as a war measure, was attacked anew and has now reached a point where a very brief outline of some of the physical results obtained should be presented in the interests of the development of the Art in general. Since the phenomena involved embrace a variety of physical problems rather than strictly engineering ones, these data are presented in the hope that our scientists may see in the experiments some points of departure for further research.

It was immediately discovered that with the sensitive amplifiers now in use it was possible to receive signals from the principal European stations by simply laying a small wire netting on the ground beneath the tree and connecting an insulated wire to a nail driven in the tree well within the outline of the tree top. See Figure 1. Messages having been received from England, France, Germany and Italy.

This encouraging first result justified a more careful examination of the phenomena and the most suitable arrangement of circuits for the purpose.

And Now For the Floraphone

The messages carried over this tree telephone and telegram system have been named by the writer. They are to be ‘floragrams’. The tree telephone is to be a ‘floraphone’; the tree telegraph a ‘floragraph’.

The discovery is now announced after experiments covering 15 years, beginning in California and continuing intermittently until the outbreak of the war, when they went forward with vigor as an emergency means of communication. The system was utilized during the war in listening-in on the German radio communication.

The final development took place in a small portable laboratory, purchased from a mail order house, capable of being carried to any place in the woods. A group of soldiers, taking turns at copying messages, assisted the writer in the development of the apparatus.

Without entering into the details of these preliminary experiments here it may be said that one fo the best receiving arrangements is found to be an elevated tree earth-terminal in the upper part of the tree top, and an earth consisting practically of several short pieces of insulated wire, sealed at the outer end, radiating out from a common center, and buried a few inches beneath the surface of the ground in the neighborhood of the tree. See Figure 2.

It was soon found that a tree-antenna could be used efficiently as a multiple radio receiving set over widely different wavelengths, see Figure 3A, receiving either from separate terminals at the same (shown dotted in Figure 3A) or different heights of the tree, or in series from the same terminal. See Figure 3B.

This same type of circuit was employed in an inverse manner for radio-telephonic transmitting purposes, although the experiments thus far have been limited to short distances. It was found that 2-way radio-telephonic communication was easily established with remarkably low values of transmitting antenna current. See diagram of this test illustrated at Figure 4.

The flexibility of this arrangement is very striking. The linking up of wire and wireless methods was found to be both convenient and efficient. Radio-telephonic messages from airplanes were readily received by the tree-antenna arrangement and transferred thence to the wire system of the city of Washington and finally received at any point desired.

Furthermore, radio-telephonic transmission through the tree-antenna was received by another tree-antenna, and automatically returned to the sender on a wire system, thus making a complete circuit. Illustrated at Figure 6.

Long distance reception on any wave length from all the larger European stations and from our ships at sea was easily accomplished and traffic copied on a 24 hour schedule by the regular enlisted operators of the Signal Corps. A small portable house serving as a field laboratory was erected in the midst of the forest area on Grant Road, Washington DC, and here was assembled a collection of amplifiers from the Army, navy, the British and French, and American manufacturers. With these facilities it was a matter of a few days to test out, at least in a superficial manner, a large number of proposed arrangements using trees as antenna.

Figure 7 shows how a number of trees can be connected up to give the effect of one large antenna.

In France officers of the Signal Corps, by the simple device of driving a spike into a tree-trunk to which connect the audion set which wireless operators use to make up wireless messages, it was found possible to listen in on communications between German airplanes and the German lines. Messages were thus intercepted in spots into which it would have been impossible to transport a field wireless apparatus. See Figure 8.

The Physics of Tree Wireless

We may regard the metallic electrode rigidly driven into the living organism of a tree, as described above, as a potential earth-terminal for the study of the potential distribution on the surface of the earth itself. It has been shown in these experiments that this metallic terminal intimately connected to the earth itself and a part thereof is subject to changes of potential representing the innumerable frequencies required by modern radio-telephony and telegraphy, as well as any other electrical disturbances which may occur on the surface of the earth or the atmosphere above the earth.

It has also been shown, as expected, that we can select from this composite one or more of the different frequencies by tuned electrical loop circuits suitably connected to this electrode and study each in turn, at will, just as color screens can select a particular component of white light. We may, indeed, by means of a highly insulated conductor bring this terminal directly to the laboratory and connect it immediately to the modern thermionic tube and amplify almost at will the particular effects we are studying…

We can consider that trees have been pieces of electrical apparatus from their beginning and with their manifold chains of living cells are absorbers, conductors and radiators of the long electromagnetic waves as used in the radio Art.

For our present purpose we may consider, therefore, a growing tree as a higly organized piece of living earth, to be used in the same manner as we now use the earth as a universal conductor for telephony and telegraphy and other electrical purposes…

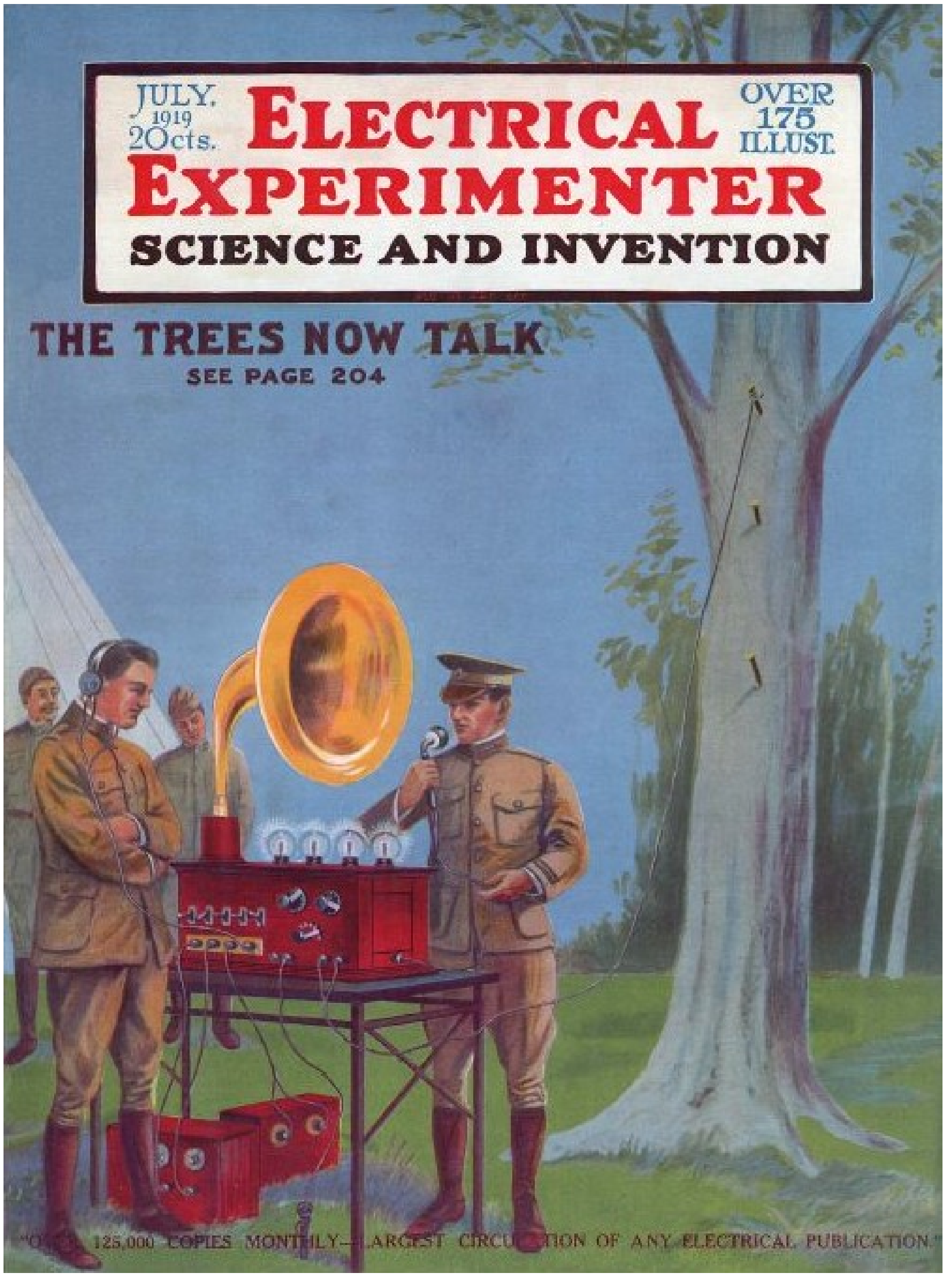
JULY, 1919 20cts.
ELECTRICAL EXPERIMENTER
OVER 175 ILLUST.
SCIENCE AND INVENTION
THE TREES NOW TALK
SEE PAGE 204
"OVER 125,000 COPIES MONTHLY—LARGEST CIRCULATION OF ANY ELECTRICAL PUBLICATION."

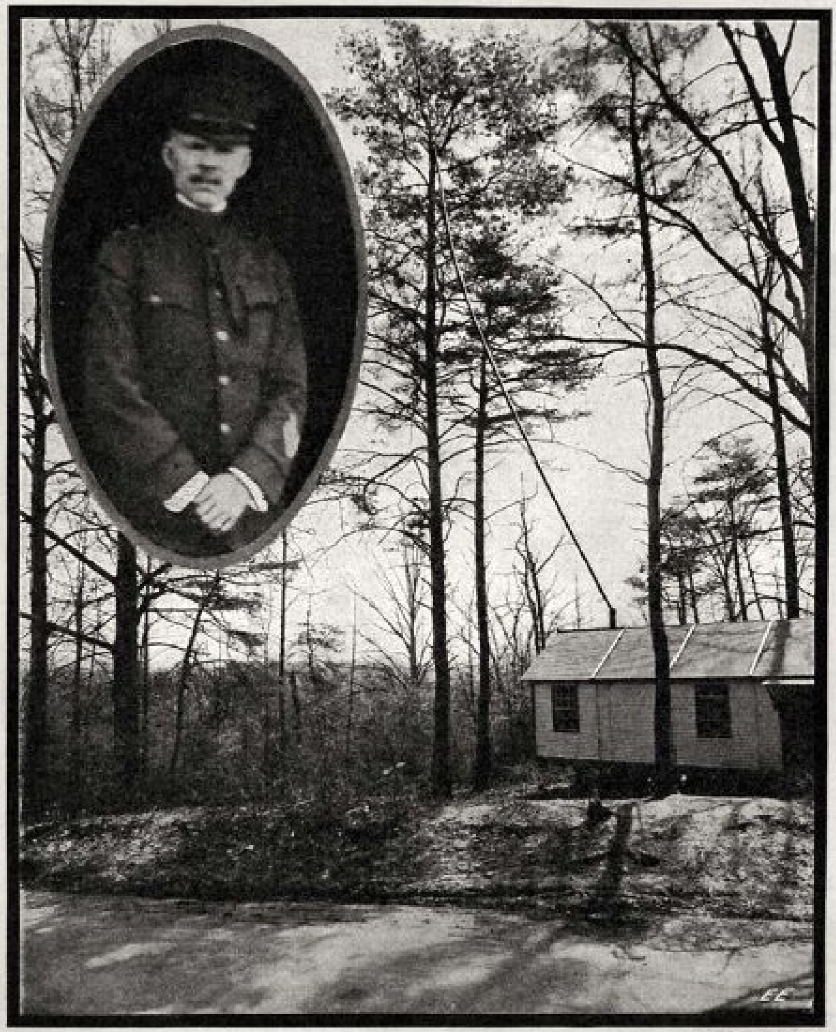

Signal Corps Photo

The U. S. Signal Corps Laboratory Near Washington, D. C., Where Remarkable Results Were Obtained Using Living Trees as Radio Antennae. Signals from Europe and Other Countries Were Easily Copied. Trees Have Been Successfully Used Also for Radiophone Transmission. Insert View: Major General George O. Squier, Inventor of "Tree Wireless."

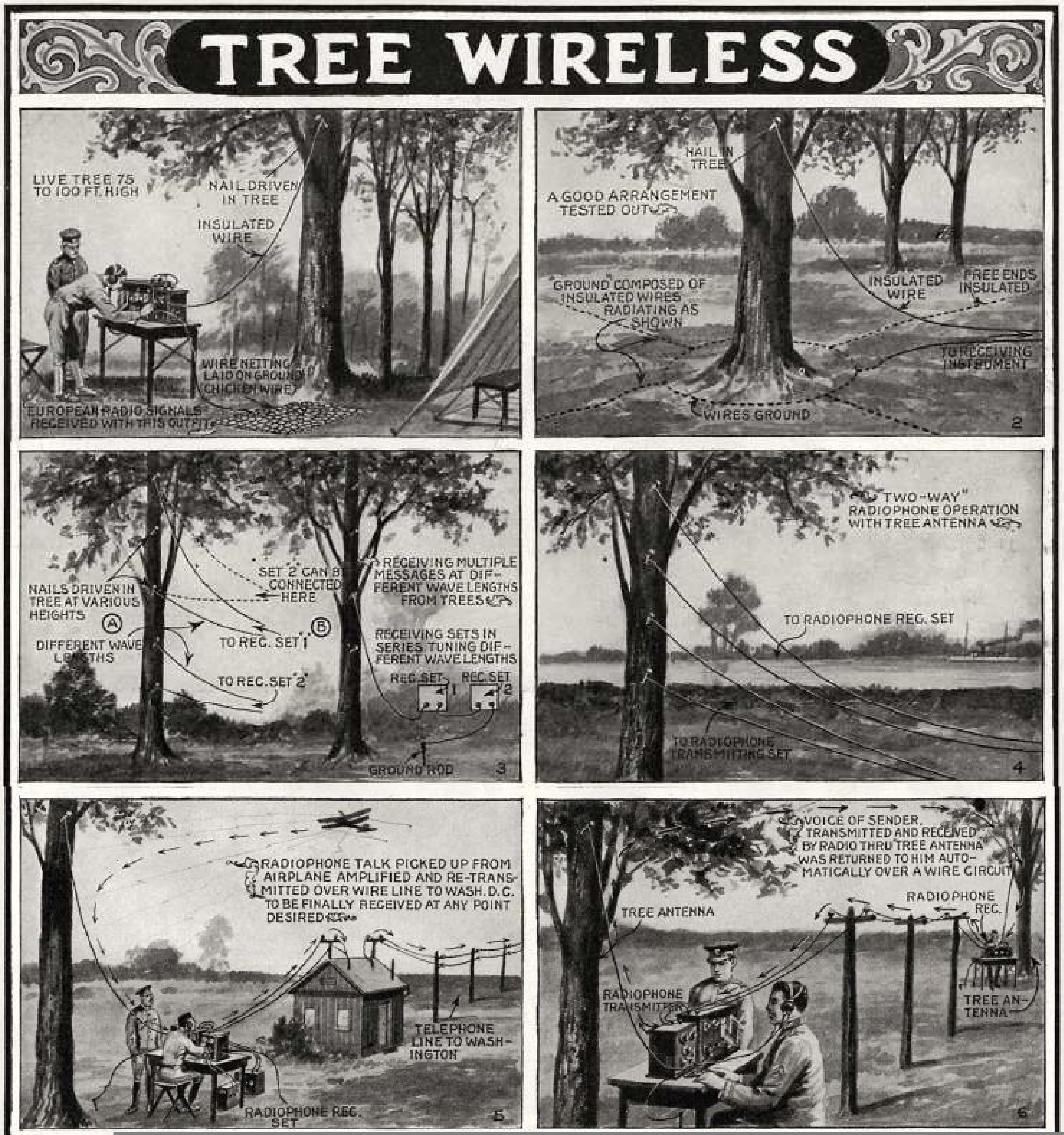
TREE WIRELESS
LIVE TREE 75 TO 100 FT. HIGH
NAIL DRIVEN IN TREE
INSULATED WIRE
WIRE NETTING LAID ON GROUND (CHICKEN WIRE)
"EUROPEAN RADIO SIGNALS" RECEIVED WITH THIS OUTFIT
NAIL IN TREE
A GOOD ARRANGEMENT TESTED OUT
"GROUND" COMPOSED OF INSULATED WIRES RADIATING AS SHOWN
INSULATED WIRE
FREE ENDS INSULATED
TO RECEIVING INSTRUMENT
WIRES GROUND
2
NAILS DRIVEN IN TREE AT VARIOUS HEIGHTS
Ⓐ
DIFFERENT WAVE LENGTHS
SET "2" CAN BE CONNECTED HERE
Ⓑ
TO REC. SET "1"
TO REC. SET "2"
RECEIVING MULTIPLE MESSAGES AT DIFFERENT WAVE LENGTHS FROM TREES
RECEIVING SETS IN SERIES, TUNING DIFFERENT WAVE LENGTHS
REC. SET 1
REC. SET 2
GROUND ROD
3
"TWO-WAY" RADIOPHONE OPERATION WITH TREE ANTENNA
TO RADIOPHONE REC. SET
TO RADIOPHONE TRANSMITTING SET
4
RADIOPHONE TALK PICKED UP FROM AIRPLANE AMPLIFIED AND RE-TRANSMITTED OVER WIRE LINE TO WASH. D.C. TO BE FINALLY RECEIVED AT ANY POINT DESIRED
TELEPHONE LINE TO WASHINGTON
RADIOPHONE REC. SET
5
VOICE OF SENDER, TRANSMITTED AND RECEIVED BY RADIO THRU TREE ANTENNA WAS RETURNED TO HIM AUTOMATICALLY OVER A WIRE CIRCUIT
RADIOPHONE REC.
TREE ANTENNA
RADIOPHONE TRANSMITTER
TREE ANTENNA
6

British Patent Specification # 149,917

Improvements in & Relating to Radio Communication Systems

George O. Squier

This invention relates to radio communication systems. More particularly, the invention relates to radio transmission and reception through the use of living vegetable organisms such as trees, plants, and the like.

As disclosed in the Specification of my prior Letters Patent No. 25,610 (1904), I have discovered heretofore that tall trees and like growing vegetation possessed electrical conductivity of a certain nature adapted for the reception of signaling electromagnetic waves and capable of forming a part of an antenna or serial with the use of a direct earth "ground" or equivalent point of connection with the tree, in the potential node region.

I have recently discovered that living vegetable organisms generally are adapted for transmission and reception of radio or high frequency oscillations, whether damped or undamped, with the use of a suitable counterpoise. I have further discovered that such living organisms are adapted for respectively transmitting or receiving a plurality of separate trains of radio or high frequency oscillations simultaneously, in the communication of either or both telephonic or telegraphic messages.

In such use of an antenna comprising living vegetable organism and a counterpoise, I have discovered that optimum results are obtained upon arranging the point of connection of the transmitting or the receiving set within a certain range of the tree, or other living vegetable organism; for trees indigenous in the temperate zones, I have discovered that such optimum point of connection is in a region approximately two-thirds of the height of the tree above the exposed surface of the earth.

In accordance with my invention for receiving radio trains of telephonic or telegraphic high frequency oscillations, a plurality of receiving sets attuned respectively to any desired frequency, may be connected by the same leading-in wire to the same tree, whereby such tree serves as the common receiving antenna for the respective individual receiving sets; and, similarly, for transmitting radio trains of telephonic or telegraphic high frequency oscillations, a plurality of transmitting sets respectively resonant to any desired frequency may be connected by a common lead-in wire to the same tree serving as the common antenna.

In the more preferred forms of my invention, I have devised suitable means for selective directional reception of radio oscillations of any desired definite frequency, by the use of a coil having the turns of its windings disposed in substantial parallelism and mounted to be rotated in a horizontal plane, whereby the received oscillations are restricted to the direction coincident with or parallel to a plane lying normal to any plane passing through any turn of the windings of the coil.

The counterpoise comprises preferably a suitable extent of conductor in the form of one or more lengths of wire disposed below the surface of the earth and suitable insulated therefrom; it is desirable to arrange a number of individual counterpoises, each extending rectilinearly in different directions to one another and to employ in directional selection the particular counterpoise extending substantially in the selected direction.

Further features and objects of the invention will be more fully understood from the following description and the accompanying drawings, in which:

Figure 1 is a diagrammatic vertical elevation of a preferred form of the invention;

Fig 1

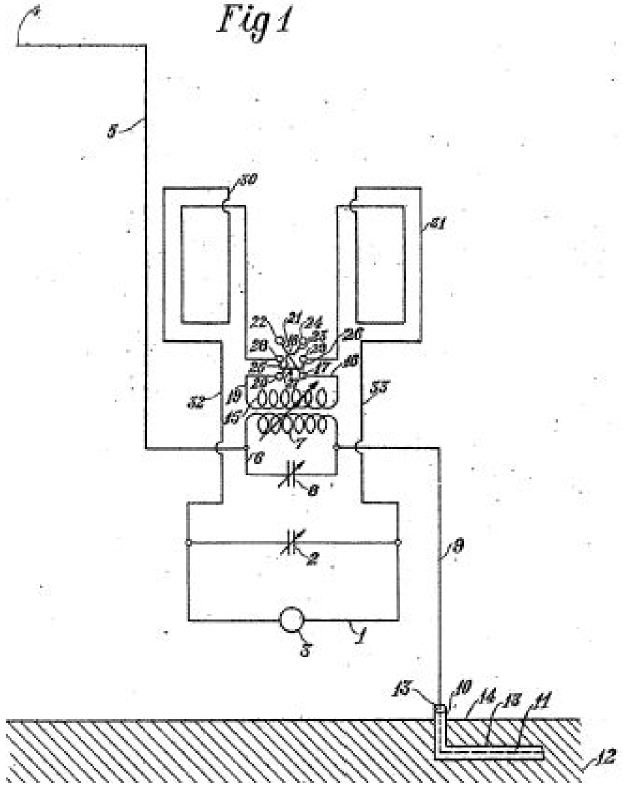
4
5
30
31
22
21
24
20
16
25
23
26
27
17
28
18
19
15
32
33
7
6
8
2
9
3
1
13
10
14
13
11
12

Figure 2 is a diagrammatic view of one preferred form of an arrangement of receiving sets, embodying my invention; and

Fig 2

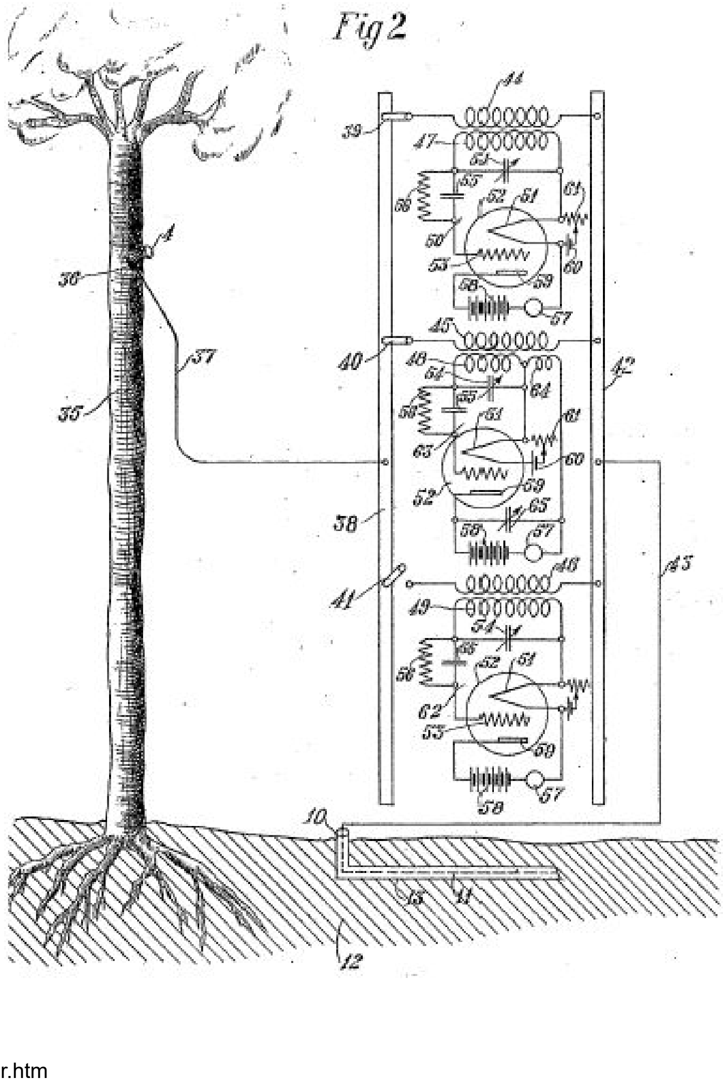
44
39
47
53
56
55
52
51
61
50
60
58
59
57
45
40
42
48
54
64
36
4
37
35
63
65
38
46
43
41
49
62
10
13
11
12

Figure 3 is a diagrammatic view of a preferred form of an arrangement of transmission sets embodying my invention.

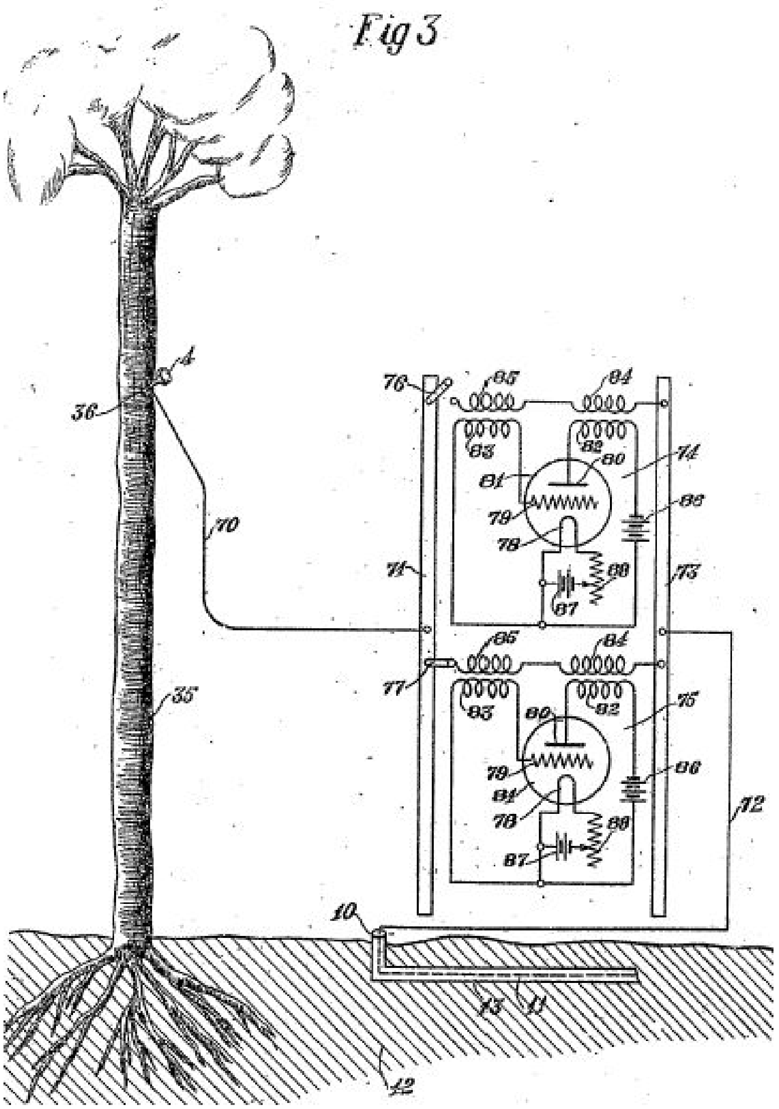
Fig 3

Referring to Figure 1, the circuit designated 1 represents any approved form of resonant receiving circuit comprising the variable condenser 2, and detector, rectifier or amplifier 3, arranged with a telephone or any suitable recording device in the usual manner. Such attuned receiving circuit 1 is suitably coupled through transformer windings or otherwise with proper regulatable reactance with the tree or other living vegetable organism and in Figure 1 I have indicated the location 4 as representing the location of a metallic nail, or equivalent, affixed to and extending within the body of such tree. I have discovered that optimum results for any particular tree are dependent upon the physical contour of the space occupied by the branches and leaves as well as the subterranean portions of the tree such as the roots and that generally the optimum results are attained by affixing the nail 4 within the region of the tree substantially two-thirds of the height of the tree above the surface of the earth.

To such nail 4 is connected one end of the lead-in wire 5, preferably enclosed in suitable insulation, and connected at its other end with the resonant circuit 6 comprising the adjustable inductance coil 7 and the adjustable condenser 8. The lead 9 connects the circuit 6 with the counterpoise 10, preferably comprising one or more metallic wires, 11, extending substantially rectilinearly and electrically insulated from the earth 12 by means of suitable insulation 13, as by enclosing such wire or wires 11 within a suitable insulating coating. Such coating 13 extends preferably an appreciable distance above the surface 14 of the earth 12. In the use of my invention with living antennae, I have arranged the counterpoise 10-13 extending in the rectilinear direction substantially parallel to the plane passing through the transmitting station, the receiving station and the center of the earth. It is advantageous to arrange a plurality of individual counterpoises 10 radiating in different rectilinear directions, with a suitable switching device for connecting the lead 9 with the particular counterpoise parallel to or most closely parallel to the selected direction.

Such coil 7 is preferably a primary coil suitably spacially coupled with an adjustable secondary coil 15 connected in the circuit comprising the variable condenser 2 and 3. In the specific form of the invention shown in Figure 1, such secondary coil 15 is provided with the lead wire 16 connected to the terminal 17 of the reversing switch 18 (shown diagrammatically) the other end of the secondary coil 15 being connected by the cross wire 20 of the reversing switch 18. Such terminal 17 is connected by the cross wire 21 with the terminal 24. The pivoted knife blades 25, 26, having the handle 27, are mounted in the usual manner on the central terminals 28, 29. The coil 30 represents diagrammatically a coil comprising a desired number of windings or successive turns of exposed or insulation covered wire, each turn lying substantially in a common vertical plane and the coil indicated at 31 represents a similar plurality of turns of windings mounted in substantially a common vertical plane. One terminal of the coil 30 is connected with one of the central terminals of the reversing switch 18, say terminal 28, and one terminal of the coil 31 is connected with the other central terminal 29, of the reversing switch 18. The other terminal of the coil 30 is connected by the wire 32 with the leads of the variable condenser 2 in shunt with the detector or amplifier 3, while the other terminal of the coil 31 is connected by the wire 33 to the opposite lead of the variable condenser 2 and the detector or amplifier 3.

Such set of coils 30, 31 are suitably mounted to be rotated horizontally, whereby all turns of the coils 30, 31, at any given position lie substantially in a common vertical plane. Such coils 30, 31 may be arranged on a common wooden or like insulating frame of cylindrical, rectangular or other desired contour on cross section.

By virtue of such an arrangement, the primary and secondary circuits are respectively directly exposed to and simultaneously electromagnetically acted upon by the incoming oscillations, and by means of the reversing switch 18 the oscillations generated in the coil 30 may be neutralized or accumulated relatively to the oscillation generated in the coil 31 to produce either the "barrage" or the "amplifying" effect.

By means of such "barrage" effect, the oscillations of any given wavelength directed in any absolute direction are cut out from the detector or amplifying circuit and enable oscillations of similar or the same wave length received in any definite direction to be conducted to the detector or amplifying circuit. By means of such accumulating effect of the coils 30, 31, and the switch 18, the waves of any definite wavelength received in any definite absolute direction are amplified and enable long distance reception at any wave length.

In practice, I have carried out the invention for reception by the use of either a single receiving set or a plurality of receiving sets. Such receiving set or sets may be arranged either to receive damped high frequency oscillations or undamped high frequency oscillations.

In Figure 2 I have illustrated diagrammatically one form of the application of the invention for the simultaneous reception of a plurality of different trains of high frequency oscillations.

The nail 4 is indicated as located in electrical connection with the tree 35 at an optimum point 36, the lead-in wire 37 being connected at its upper end with the nail 4 and at its lower end with the incoming connector bar 38. The connector 38 is provided with the individual switches 39-40-41, etc., corresponding to the number of individual receiving sets. The outgoing connector bar 42 is connected by the lead 43 with the counterpoise 10, preferably constructed and arranged as set forth hereinabove.

In suitable relation with the incoming connector bar 38 and outgoing connector bar 42 and the switches 39, 40, 41 are arranged the respective primary coils 44, 45, 46, etc., respectively coupled in any approved arrangement with the secondary coils 47, 48, 49.

The receiving set 50 is arranged for the reception of undamped waves and is shown of the vacuum tube valve type. The terminals of the secondary 47 are suitably connected in the input circuit including the filament cathode 51 of the vacuum valve 52 and the grid 53 and comprises the variable condenser 54 in shunt with the secondary 47, the condenser 55 and the grid leak resistance 56. The output circuit comprises the battery 58, the filament cathode 51, and the telephone receiver or other audible, or any visual indicator 57. The heating circuit of the filament cathode includes the battery 60 and the variable resistance 61.

The receiving set 62 is shown of a similar undamped oscillation receiving type and like elements are designated by like reference numbers.

The receiving set 63 is indicated of the damped oscillation receiving type and comprises the additional inductance turns 64 having one terminal connected to one terminal of the filament cathode 51 and its other terminal connected to the variable condenser 65 in shunt with the telephone receiver 57 and the storage battery 58. The remaining elements of the receiving set 63 correspond to like elements of the receiving set 50 and are designated by like reference numbers.

In Figure 2 I have shown the switch 39 in closed position with the primary inductance 44 of the receiving set 50, the switch 40 in closed position with the primary inductance 45 of the receiving set 63 and the switch 41 in open position with the primary inductance 46 of the receiving set 62, and accordingly enabling the reception and detection of undamped oscillations of the frequency to which the receiving set 50 is tuned and simultaneously the reception and detection of damped oscillations of the frequency to which the receiving set 53 is tuned.

Figure 3 shows one form of practical application of my invention for transmitting simultaneously radio oscillations, either telephonic or telegraphic, either damped or undamped, or both. The nail 4, lodged at an optimum point 36 in the tree 35, is connected by the lead 70 to the connector bar 71, and the counterpoise 10 is connected by the lead 72 to the connector bar 73.

The transmitting sets 74, 75, etc., of the desired number, are suitably arranged for single or multiple transmission, for which purpose the switches 76, 77, etc., are provided. The transmitting sets 74, 75 are shown of any approved type, such as the oscillating vacuum tube type comprising the filament cathode 78, the grid 79 and anode 80, arranged in the vacuum tube 81, suitably connected with the primary inductance 84, 85. The primary inductances 84, 85 are suitably connected to the connector bar 73 and the connector bar 71 through the respective switches 76, 77. The battery or other source of electric current is indicated at 86. The adjustable heating circuit of the cathode filament 79 is shown comprising the battery 87 and the variable resistance 88.

In Figure 3, the switch 77 is in closed position thus placing the transmitting set 75 in operative connection with the tree serving as the antenna. Upon closing the switch 76 the transmitting set 74 is similarly placed in operative connection with the tree antenna 35. Upon closure of both switches 76, 77 the tree serves as the antenna for the transmission of simultaneous trains of oscillations emitted by the respective transmitting sets 74, 75, modified by a key or telephone transmitter, or other suitable modulator, for the transmission of telegraphic or telephonic messages as desired.

In the use of trees or other living vegetable organisms serving as the antenna or a art thereof, I have discovered from tests that such tree possess impedance consisting of two components, resistance and reactance, the latter being usually condensive, for oscillation within the range of present day wavelengths. The tests also show that the apparent capacity of a tree serving as an antenna is substantially proportional to the height of contact of the lead wire connecting the transmission set with the tree.

The tests also show that the apparent resistance of a tree serving as an antenna is appreciably greater than the resistance of the ordinary metallic antenna, and it is accordingly desirable to modify the resistance or alter the design of the receivers, if it desired to increase the effectiveness of the energy of the electromagnetic radiation impinging on the tree; however, the ordinary receivers of present design may be used and are responsive to any wave length of radiation employed in present day practice.

Whereas I have described my invention by reference to specific forms thereof, it will be understood that many changes and modifications may be made without departing from the spirit of the invention as defined by the appended claims.

An alternative electric power generating system that draws energy from a seemingly unlikely yet abundant, eminently renewable and virtually free power source has been submitted for patenting by MagCap Engineering, LLC, Canton, Mass., in collaboration with Gordon W. Wadle, an inventor from Thomson, Ill. Wadle has invented a way to capture the energy generated by a living non- animal organism --- such as a tree.

Source: KeelyNet / Jerry Decker email (12/21/05)

"Unlimited Electric Energy from the Environment?"

Chris Lagadinos, president of MagCap, developed circuitry that converts this natural energy source into useable DC power capable of sustaining a continuous current to charge and maintain a battery at full charge.

"As unbelievable as it sounds, we've been able to demonstrate the feasibility of generating electricity in this manner," said Wadle.

"While the development is in its infancy, it has the potential to provide an unlimited supply of constant, clean energy without relying on fossil fuels, a power generating plant complex or an elaborate transmission network."

Wadle likened the invention to the discovery of electricity over 200 years ago when charged particles were harnessed to create an electric current. "Now we've learned that there is an immense, inexhaustible source of energy literally all around us that can be harnessed and converted into usable electric power," he said. Ultimately, it should prove to be more practical than solar energy or wind power, and certainly more affordable than fuel cells, he added.

Wadle said he got the original idea of harnessing a tree for electrical energy from studying lightening, more than 50 percent of which originates from the ground. This prompted him to develop the theories resulting in a method to access this power source.

Lagadinos then designed circuitry that filtered and amplified these energy emanations, creating a useable power source.

Basically, the existing system includes a metal rod embedded in the tree, a grounding rod driven into the ground, and the connecting circuitry, which filters and boosts the power output sufficient to charge a battery.

In its current experimental configuration, the demonstration system produces 2.1 volts, enough to continuously maintain a full charge in a nickel cadmium battery attached to an LED light. "Think of the environment as a battery, in this case," said Lagadinos, "with the tree as the positive pole and the grounding rod as the negative."

Lagadinos said the system could be enhanced enough to generate 12 volts and one amp of power, "a desirable power level that could be used to power just about anything," he said. It is enough power to charge batteries for any type of vehicle, including hybrids and electric cars, or to use with an AC converter to produce household power, he added. The LED industry is a prime example of a potential user of this power source.

While the basic concept of this invention -- using a tree to generate electric power -- seems too incredible to be true, Lagadinos said it can be demonstrated quite simply. "Simply drive an aluminum roofing nail through the bark and into the wood of a tree -- any tree -- approximately one half inch; drive a copper water pipe six or seven inches into the ground, then get a standard off-the-shelf digital volt meter and attach one probe to the pipe, the other to the nail and you'll get a reading of anywhere from 0.8 to 1.2 volts of DC power," he said.

"You can't do anything with it in that form because it is 'dirty' -- i.e. highly unstable and too weak to power anything," he added. In order to properly harness this potential energy source, MagCap devised two test circuits: one with three capacitors that were connected in parallel by means of a switch and charged to 0.7 volts each.

When fully charged they are switched to a series mode, multiplying the voltage to 2.1 volts and flashing an LED to show that sufficient power could be generated to produce a useable result. The second circuit included a filtering device to stabilize and "clean" the current so it could be used to charge and maintain a NiCad battery.

The battery then could be connected to the LED to keep the LED lit continuously. Wadle pointed out that there seems to be no limit to the amount of power that can be drawn from an individual tree, no matter how many "taps" are inserted -- each produces the same amount of energy, an average of 0.7 - 0.8 volts. Size of the tree also seems not to matter.

Interestingly, while conventional wisdom would seem to indicate that the tree draws much of its energy from photosynthesis via its leaves, the voltage output actually increases to 1.2-1.3 volts in the winter after the leaves have fallen.

Patents by George O. Squire:

Apparatus for using telephone-subscribers' loops for transmission and reception of broadcasting
SQUIER GEORGE O
EC: H04H1/08 IPC: H04H1/08; H04H1/04
US1742422
1930-01-07

Apparatus for effecting electrical signaling
SQUIER GEORGE O; LOUIS COHEN
EC: H04L27/06 IPC: H04L27/06; H04L27/06
US1608252
1926-11-23

Aerial attachment device for radiofrequency signal systems
SQUIER GEORGE
EC: H01Q1/46 IPC: H01Q1/46; H01Q1/44
Publication info: US1584197
1926-05-11

Electrical signaling
SQUIER GEORGE O; MAUBORGNE JOSEPH O
EC: H01F21/02 IPC: H01F21/02; H01F21/02
US1641608
1927-09-06

Apparatus for electrical signaling
SQUIER GEORGE
EC: H04B1/034 IPC: H04B1/034; H04B1/02
US1531629
1925-03-31

Improvements in the Art of Telegraphy
SQUIER GEORGE OWEN; AUSTIN LOUIS WINSLOW
EC: H04J1/00 IPC: H04J1/00; H04J1/00
GB108763
1917-08-23

Improvements in the Art of Electrical Signalling
SQUIER GEORGE; AUSTIN LOUIS WINSLOW
EC: H04J1/00 IPC: H04J1/00; H04J1/00
GB108230
1917-08-02

Improvements in Duplex and Multiplex Telegraphy
SQUIER GEORGE; COHEN LOUIS
EC: H04J1/00 IPC: H04J1/00; H04J1/00

GB103771
1917-02-08

Combined sound and light distributing apparatus
SQUIER GEORGE
EC: F21V33/00; H05K11/00 IPC: F21V33/00; H05K11/00; F21V33/00 (+1)
US1999579
1935-04-30

Radio signaling apparatus
SQUIER GEORGE
EC: H04B3/56 IPC: H04B3/56; H04B3/54
US1791541
1931-02-10

Improvements in or relating to Receiving Systems for Radio Telegraphy or Telephony
SQUIER GEORGE
EC: H01Q1/46; H04H1/04 IPC: H01Q1/46; H04H1/04; H01Q1/44 (+1)
GB191417487
1914-11-19

Improvements in or relating to Radio-telegraphic and Radio-telephonic Receiving Systems
SQUIER GEORGE
EC: H04M11/06 IPC: H04M11/06; H04M11/06
GB191403191
1915-02-08

Improvements in or connected with Multiplex Telephony and Telegraphy
SQUIER GEORGE
EC: H04J1/00 IPC: H04J1/00; H04J1/00
GB191030003
1911-12-27

Improvements in or relating to Receiving Systems for Radio Telegraphy or Telephony
SQUIER GEORGE
EC: H01Q1/46; H04H1/04 IPC: H01Q1/46; H04H1/04; H01Q1/44 (+1)
GB191417487
1914-11-19

Improvements in or relating to Radio-telegraphic and Radio-telephonic Receiving Systems
SQUIER GEORGE
EC: H04M11/06 IPC: H04M11/06; H04M11/06
GB191403191
1915-02-08

Improvements in or connected with Multiplex Telephony and Telegraphy
SQUIER GEORGE
EC: H04J1/00 IPC: H04J1/00; H04J1/00
GB191030003
1911-12-27

Improvements in or connected with Telegraphic Apparatus
CREHORE ALBERT CUSHING (US); SQUIER GEORGE
GB189919867
1900-10-03

Improvements in Telegraphic Apparatus
CREHORE ALBERT CUSHING (US); SQUIER GEORGE

GB189918980
1900-09-20

Improvements in Telegraphic Apparatus
CREHORE ALBERT CUSHING (US); SQUIER GEORGE
GB189917722
1900-09-01

Improvements in Submarine Cable Telegraphs
CREHORE ALBERT C (US); SQUIER GEORGE
GB189823556
1900-02-08

Method and apparatus for using telephone subscribers' loops for transmission and reception of broadcasting
SQUIER GEORGE
EC: H04B3/56; H04H1/04 IPC: H04B3/56; H04H1/04; H04B3/54 (+1)
FR644877
1927-10-16

TELEPHONE AND BROADCASTING SYSTEM
SQUIER GEORGE
CA285479
1928-12-11

HIGH FREQUENCY CURRENT SIGNALLING
SQUIER GEORGE
CA244330
1924-11-11

ELECTRIC SIGNALLING SYSTEM
SQUIER GEORGE
CA233299
1923-08-07

TREE TELEPHONY AND TELEGRAPHY
SQUIER GEORGE
CA232487
1923-07-03

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

The Global Atmospheric Electric Circuit

Giles Harrison

Department of Meteorology,
University of Reading, UK

Atmospheric Electricity - reading list

- J.A. Chalmers *Atmospheric Electricity*, 2nd edition, Pergamon Press (1967)
- H. Israël (1970) *Atmospheric Electricity vol1 (Fundamentals, Conductivity, Ions)* (Problems of Cosmic Physics vol 29), Israel Program for Scientific Translations, Jerusalem.
- H. Israël (1973) *Atmospheric Electricity vol2 (Fields, charges, currents)* (Problems of Cosmic Physics vol 29), Israel Program for Scientific Translations, Jerusalem.
- D.R. McGorman and W.D. Rust, *The Electrical Nature of Storms,* Oxford University Press, Oxford, 1998.
- Krider, E.P., et al *The Earth's Electrical Environment*, Studies in Geophysics, National Academy Press, Washington, 1986 http://www.nap.edu/openbook.php?record_id=898&page=R1
- Vladimir A. Rakov and Martin A. Uman *Lightning: Physics and Effects* Cambridge University Press (2007)

Other useful resources include sections 3.31 and 10.5 of R.G. Harrison's *Meteorological Measurements and Instrumentation*, *The Feynman Lectures in Physics* (Volume III) and the PPLATO teaching resources at http://www.cse.salford.ac.uk/profiles/gsmcdonald/pp/PPLATOResources/PPLATO.htm

Outline

1. Vertical current flow and the global circuit
2. Findings supporting the global circuit
3. Measurements of global circuit parameters
4. Variability in the global circuit
5. Applications and planetary global circuits

Atmospheric electricity

Atmospheric electricity is concerned with:

•electrical properties of the lower and upper atmosphere
•currents flowing in the **fair weather atmosphere**
•**ionisation** of atmospheric air by natural radioactivity and cosmic rays
•**charging of particles** and droplets, snow and hail
•**charges** separated by thunderstorms and disturbed weather
•**currents** flowing from thunderstorms
•**Electric fields** in thunderstorms, the upper atmosphere and the fair weather atmosphere
•initiation of **lightning**

0. Historical background and some key individuals

Benjamin Franklin (1706-1790) – *Statesman, a Founding Father of the US, and scientist*

Giambatista Beccaria (1716-1781) – *Professor of Experimental Physics, University of Turin*

Lord Kelvin (1824-1907) - *Experimental and mathematical physicist, University of Glasgow*

Victor Hess (1883-1964) – Austrian physicist, discoverer of cosmic rays (Nobel Prize 1936)

CTR Wilson (1869-1959) – Scottish physicist, inventor of the cloud chamber (Nobel Prize 1927)

Sir Edward Appelton (1892-1965) – English physicist and discoverer of the ionosphere (Nobel Prize, 1946)

1750s - Benjamin Franklin

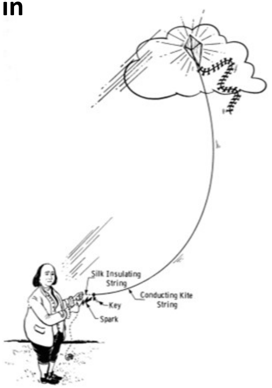

M.A. Uman, All about lightning, Dover, 1986

but.. debatable whether *he* actually conducted this experiment…

Scientific motivation: the origin of the "fair weather" field

Key finding: measurements by Lemonnier, Canton and others in 1750s detected positive electrification in fair (or "serene") weather conditions

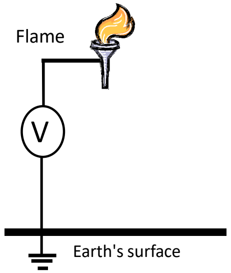

→ *Why, in fair weather (no thunderstorms present), using a flame probe sensor, does the voltmeter record a positive voltage?*

Lord Kelvin's standardised measurements

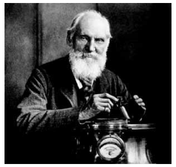

"The height of the match was 3 feet above the ground, and the observer at the electrometer lay on the ground to render the electrical influence of his own body on the match insensible. The result showed a difference of potentials between the earth (negative) and the air (positive) at the match equal to that of 115 elements of Daniel's battery."

29th meeting, British Association for the Advancement of Science, Aberdeen, September 1859

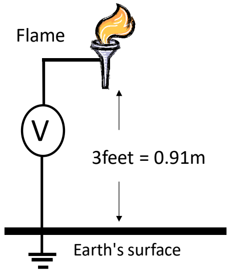

1 Daniel Cell (Zn/Cu) generates 1.08V

→ 115 Daniel cells = 124.2V

(vertical) **Potential Gradient**

Aberdeen (8am, 14th September, 1859)

= 124.2V / 0.91m = 137 Vm^{-1}

Kelvin’s interest in the variability

> *“There can be no doubt but the electric indications, when sufficiently studied, will be found important additions to our means for prognosticating the weather …[I] hope to soon see the atmospheric electrometer generally adopted as a useful and convenient weather-glass” (Lord Kelvin, 18th May 1860)*

Thomson, W. Reprint of papers on Electrostatics and Magnetism Macmillan, London, 1872

The variability which he observed probably challenged Kelvin’s world view, which may be the reason he sought to devise methods to investigate it – he was also central in developing telegraphy technology, which might have been affected by atmospheric disturbances

See: K.L. Aplin and R.G. Harrison, Lord Kelvin's atmospheric electricity measurements *History of Geo- and Space Sciences* **4**, 83-95, 2013 doi:10.5194/hgss-4-83-2013

Kelvin’s “water dropper” potential equaliser

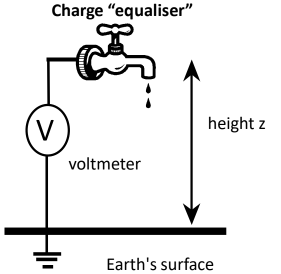
Charge “equaliser”
V
voltmeter
height z
Earth's surface

- provided continuous measurements used with photographic recording
- works by the exchange of charge between the air and the nozzle through the water droplets
- showed the PG was variable

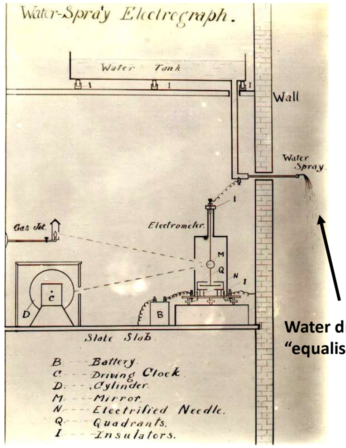
Water-Spray Electrograph.
Water Tank
Wall
Water Spray
Gas Jet.
Electrometer.
M
Q
N
I
C
D
B
Slate Slab.
B - - - Battery.
C - - - Driving Clock.
D - - - Cylinder.
M - - - Mirror.
N - - - Electrified Needle.
Q - - - Quadrants.
I - - - Insulators.

Water dropper “equaliser”

K.L. Aplin and R.G. Harrison, Lord Kelvin's atmospheric electricity measurements, *History of Geo- and Space Sciences* **4**, 83-95, 2013

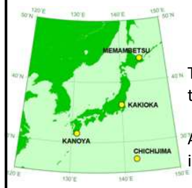
MEMAMBETSU
KAKIOKA
KANOYA
CHICHIJIMA
120°E
130°E
140°E
150°E
50°N
40°N
30°N

Kelvin water dropper at Kakioka Observatory

The Kelvin device was widely used, including for balloon studies and on the top of the Eiffel Tower

A water dropper sensor is still in use at Kakioka geophysical observatory in Japan

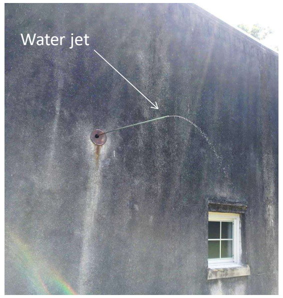

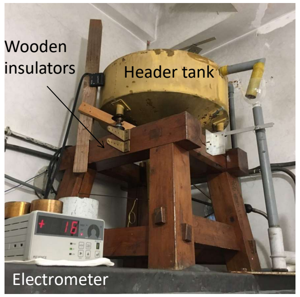
Wooden insulators
Header tank
Electrometer

(pictures from July 2018)

Paradigm shift around 1900

- Kelvin and others’ understanding of atmospheric electricity was based on an electrostatic view, perhaps informed by the Victorian view of an “inherited” earth system
- But…work on ionisation and current flow by C.T.R.Wilson (and, independently, by Elster and Geitel in Germany) led to the electrostatic viewpoint becoming untenable
- Evidence indicating continuing atmospheric current flow brought with it the concept of a varying, dynamic, terrestrial atmospheric electricity system

K.L. Aplin and R.G. Harrison, Lord Kelvin's atmospheric electricity measurements *History of Geo- and Space Sciences* **4**, 83-95, 2013 doi:10.5194/hgss-4-83-2013

1. Vertical current flow in the fair weather atmosphere

Measuring vertical current flow

Charles Thomas Rees "CTR" Wilson (1869-1959)

- **Nobel prize winner (1927) for invention of the cloud chamber**
- **also demonstrated ionisation in air (*c*1900), which implied atmospheric charge was continuously replenished through current flow**

"I remember the satisfaction I had when my work led to the fulfilment of my dream of isolating a portion of the earth's surface and measuring the charge upon it and the current flowing into it from the atmosphere."

Wilson C.T.R., 1960. Reminiscences of my early years. *Notes & Recollects Royal Society* 14, 2, 163-173

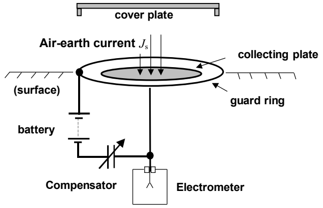

measured: **Potential Gradient** (PG) and **vertical current** density (J_c)

C.T.R. Wilson, 1906. On the measurement of the earth-air current and on the origin of atmospheric electricity. *Proc. Camb. Philos. Soc.* **13**, 6, 363-382; R.G. Harrison and W..J. Ingram, Air-earth current measurements at Kew, London, 1909-1979 *Atmos Res* **76**, (1-4), 49-64

Modern version: A.J. Bennett and R.G. Harrison, A simple atmospheric electrical instrument for educational use *Advances in Geosciences* **13**, 11-15 (2007)

Considering global current balance – Wilson's notebooks

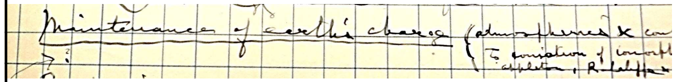

Maintenance of earth's charge { atmospherics × con
{ to ionisation of [illegible]
[illegible]

(A20)

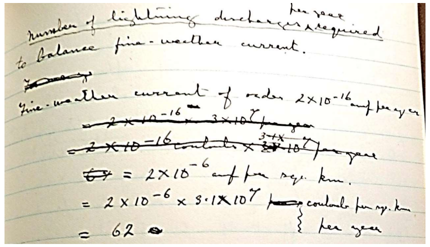

number of lightning discharges per year required
to balance fine-weather current.
~~[illegible]~~
fine-weather current of order 2×10^{-16} amp per sq cm
~~$2\times10^{-16} \times 3\times10^{7}$ per year~~
~~2×10^{-16} coulombs $\times$ $3{\cdot}1\times10^{7}$ per year~~
~~[illegible]~~ $= 2\times10^{-6}$ amp per sq. km.
$= 2\times10^{-6} \times 3{\cdot}1\times10^{7}$ coulombs per sq. km } per year
$= 62$

"Three lightning discharges upwards per square km per year of 20 coulombs would be sufficient.

It is not impossible that even in Cambridge there may be that number of lightning strokes per square km per year."

(B8)

Electrical conductivity of air, σ

Air is not a perfect insulator – it is weakly conducting due to presence of ions

Electrical conductivity describes ability of air to conduct electric current

Coulomb discovered this property in 1785– he noticed that charge on an object decayed slowly with time

Subsequent work by Elster and Geitel 1900, and Gerdien 1905 developed instrumentation for balloons to further characterise atmospheric ions and measure conductivity

Conductivity inferred from rate of decay of charge on electrode

=> Measurement principles still used today

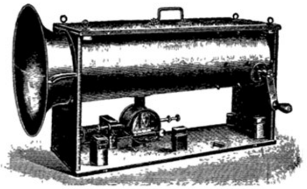

Balloon-borne conductivity apparatus from Gerdien (1905)

Ionisation measurements

Early 1900s, observations showed that charge decay occurred inside electroscopes (charge measuring devices working by mechanical deflection) that were well insulated from outside (Wilson, Elster and Geitel)

What was the source of ionisation?

Ionisation due to natural radioactivity was one source – originated at ground level, therefore ionisation should decrease with height

Balloon measurements during (1911-1913) by Victor Hess showed that ionisation decreased with height for first few hundred meters, then increased

Measurements during a solar eclipse showed that the sun was not the origin of the radiation

=> Discovery of galactic cosmic rays – the lower atmosphere's principal ionising agents

Vertical profile of air conductivity

Electrical conductivity of air results from the small ions it contains, generated by natural ion production from radioactivity and cosmic rays

$$\boxed{\sigma = 2\, n\, \mu\, e}$$

σ =total conductivity

n = ion number concentration

μ = ion mobility

e = electronic charge

Typically $\sigma \sim 10^{-14}\ \Omega^{-1}\ m^{-1}$ in surface atmospheric air, or using SI ~ 10 fS m^{-1}

Only **small** ions (d ~1nm) contribute to σ

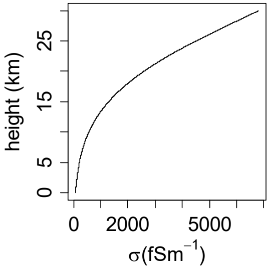

Ion production rate varies with height (decreasing terrestrial ionisation but increasing cosmic ionisation)

Ion mobility also varies with temperature and pressure

→ Conductivity profile

Columnar resistance, R_c

•Columnar resistance is the found from the reciprocal of conductivity σ(z)

•Expressed as an integrated resistance, of a unit area column of air:

$$\boxed{R_c = \int_0^{z_I} \frac{dz}{\sigma(z)}}$$

•z = altitude
•z_I = height of upper conducting layer

•R_c~ 50 to 300 PΩ m²

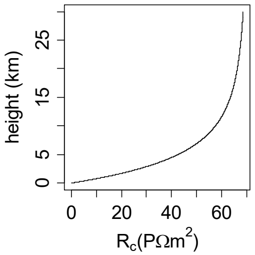

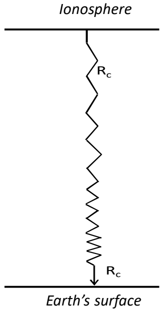

•most of R_c contribution is from lowest few km of atmosphere

Ionospheric potential - V_I

•The potential increases with height, to a maximum value essentially constant from the stratosphere to the ionosphere – known as the **ionospheric potential** V_I

•V_I is found from integrating the vertical profile of PG, measured by radiosonde balloons or aircraft, carrying a field mill or radioactive collectors. It is essentially a **global equipotential**

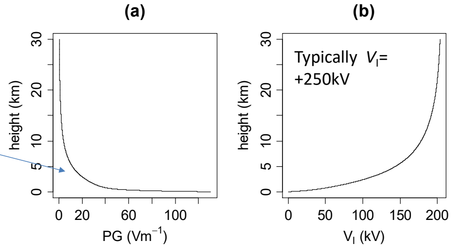

most variation in potential is within the lowest 10km

Typically V_I= +250kV

If the PG at height z is written as $F(z)$

$$V_I = \int_0^{z_I} F(z)\, dz$$

Local to global - Wilson's "global circuit" conceptual model

Wilson's "global atmospheric electric circuit", describes electrical characteristics of the atmosphere:

Source: ***Thunderstorms and rain clouds*** *transfer positive charge to the ionosphere in disturbed weather regions*

Sink: *In fair weather regions the air-earth current density J_c returns a small "conduction current" to Earth .*

- Atmosphere bounded by upper and lower "conducting" layers
- Global thunderstorms and rain clouds charge upper layer to ~ + 300kV (V_i)
- Vertical current density J_c (~ 2pAm^{-2}) of cluster ions flows in fair weather regions
- Unit area "Columnar resistance" R_c ~ 150PΩm^2

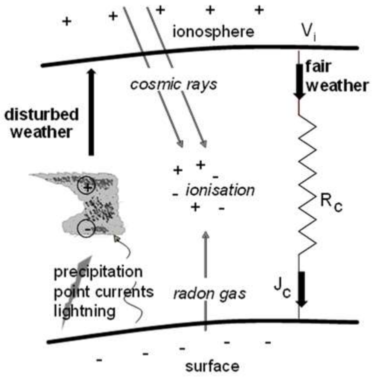

2. Findings supporting the global circuit concept

Ocean measurements: the *Galilee* and *Carnegie*

The *Galilee* and *Carnegie* were the Carnegie Institution of Washington's geophysical survey ships. The *Galilee* was ultimately unsuitable as some magnetic parts could not be replaced; the *Carnegie* was a replacement constructed specially, built from wood, copper and bronze, with an observing deck

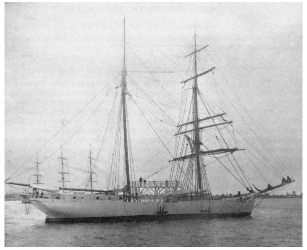

Galilee – brigantine built in 1891, as a fast packet between San Francisco and Tahiti. Chartered by CIW in 1905.

Carnegie under full sail in 1909 at launch. From 1909 to 1929 she covered 500,000km, and made the fastest circumnavigation of Antarctica.

Difficulties with early instrumentation

- Atmospheric electricity development work was delayed, until the final cruise of the *Galilee.*
- These first measurement attempts proved very difficult. In particular, the atmospheric electric field (Potential Gradient, PG) measurements…

> *"…seemed quite impracticable…the rolling of the ship, the flapping of the sails, and the varying position of the yards and boom under various sailing conditions all contributed to make the problem of reducing observations of potential-gradient to a uniform basis too complicated…"* [CIW175v3, p364]

- Of several atmospheric electricity instruments considered for the *Galilee* (e.g. the Ebert ion counter) only air conductivity measurements, using a Gerdien aspirated device, showed practical promise.
- Air ion measurements were made on cruise I of the *Carnegie* (1909-1910), alongside observations of "specific conductivity" and "radioactive content".

Ion content of the atmosphere, 1911 (CIW)

R.G. Harrison, The Carnegie Curve *Surv Geophys* **34**, 2, 209-232, DOI: 10.1007/s10712-012-9210-2 (2013)

Potential Gradient sensors

Principle: radioactive "collector" methods to acquire local potential

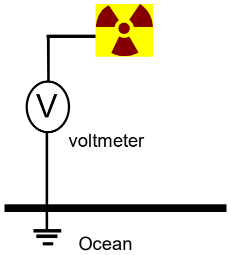

Carnegie cruise 2 (1910-1913)
Kidson and Johnston used two radium collectors suspended on a bamboo pole, extending aft from the stern rail; Hewlett tried ionium (^{232}Th) collectors

Carnegie cruise 3 (June 8th 1914 to Oct 21st 1914)
PG used ionium collectors suspended on a bamboo pole, extending aft from the stern taffrail; standardization achieved by simultaneous ship and shore observation on two occasions, at Reykjavik and Gardiners Bay: average PG for whole cruise 93 Vm^{-1}

Carnegie cruise 4 (1915-1916)
special "atmospheric-electric" house built aboard to reduce set up time, and keep instruments in stable conditions, but the ionium coated collector plate on bamboo pole took too long (~2min) to come within 1V of the final steady potential

Mechanical Potential Gradient sensor

A new faster electric field measurement method was developed, which also compensated for tilt of ship

-a "parasol" collector was constructed from gauze on a pivot arm, to allow it to be adjusted for ship angle.

-Rotated from an earthed (pointing downwards) position, to horizontal to horizon, and **change in electroscope** reading recorded. (Hourly data)

-sustained good quality insulation was not needed; used sulphur insulators. The handle's hard rubber insulation was more critical, and had to be kept clean. Daily leakage tests

-battery offset for electroscope could also be applied

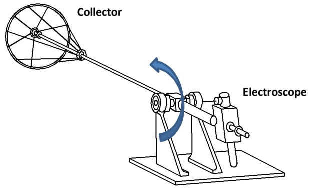

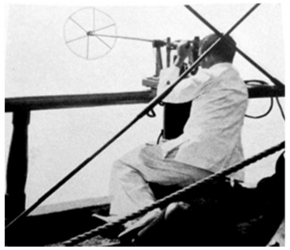

R.G. Harrison, The Carnegie Curve *Surv Geophys* **34**, 2, 209-232, DOI: 10.1007/s10712-012-9210-2 (2013)

Diurnal cycle results from Cruises 4, 5 and 6

By 1921, using harmonic analysis on the previous decade's data, the principal contributor to the PG's diurnal variation was a 24 hour component, and it was reported by its discoverer, the Carnegie Institution's S.J. Mauchly, that

> *"…the 24 hour Fourier wave was at the great majority of land stations in practical phase agreement on **universal time** with the prime daily wave over the oceans without regard to location.* [CIW175v5,p387]

This indicated that the origin of the potential gradient was not local

Mauchly SJ, Note on the diurnal variation of the atmospheric electric potential gradient *Phys Rev* n.s. **18**, 161-162 (1921); Mauchly SJ, On the diurnal variation of the potential gradient of atmospheric electricity *Terr Mag* **28**, 61-81 (1923)

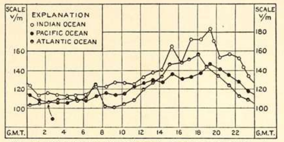

Fig. 24.—Mean Value of the Potential Gradient for the Different Oceans from Observations on the *Carnegie*, 1915–1921.

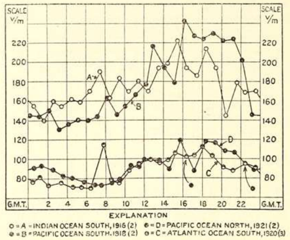

Fig. 25.—Comparison of Diurnal Variation of Potential Gradient for Same Time of Year in Different Oceans and in Different Latitudes.

Motivation for Cruise 7

Further frequent and widely distributed varying atmospheric electricity measurements were still sought

"…to settle the question whether such variations progress on the basis of universal time" (Ault, 1927).

A key improvement on the earlier cruises proposed was to employ continuous recording techniques, using a projection electrometer with photographic chart recording.

Cruise 7: oceanography and meteorology

Ault taking off "Nansen bottle", used to sample deep-sea water, 1928.

Pilot balloon weighing, prior to ascent, 1928

Ault releasing pilot balloon, to measure upper air currents, 1928.

Sighting on pilot balloon with sextant and theodolite, 1929

Ault in diving helmet, going down to disentangle bottom sample wire from hull, 1929

Carnegie Institution of Washington

Continuous PG recording on Cruise 7

- Initially a sensor with continuous recorder was planned for the masthead, but became impractical from flexing of the hemp rigging.
- Site adjacent to the parasol collector on the stern rail used instead.

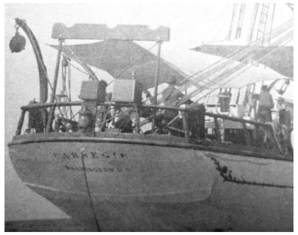

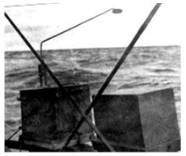

Bent collector rod used initially, projecting out over the water astern. (7th July 1928 to Nov 5th 1929)

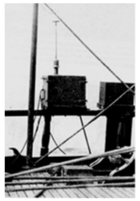

Vertical collector installed Nov 5th 1929.

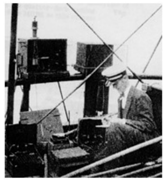

Collectors in amber, 75cm above weatherproof wooden box housing fibre electrometer, microscope, light source, and motorised photographic recorder

Cruise 7 - tragedy

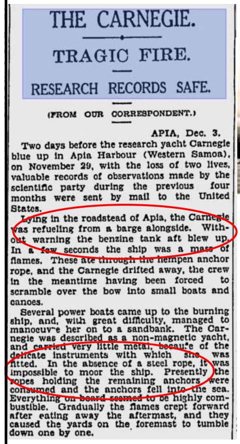

THE CARNEGIE.

TRAGIC FIRE.

RESEARCH RECORDS SAFE.

(FROM OUR CORRESPONDENT.)

APIA, Dec. 3.

Two days before the research yacht Carnegie blue up in Apia Harbour (Western Samoa), on November 29, with the loss of two lives, valuable records of observations made by the scientific party during the previous four months were sent by mail to the United States.

Lying in the roadstead of Apia, the Carnegie was refueling from a barge alongside. Without warning the benzine tank aft blew up. In a few seconds the ship was a mass of flames. These ate through the hempen anchor rope, and the Carnegie drifted away, the crew in the meantime having been forced to scramble over the bow into small boats and canoes.

Several power boats came up to the burning ship, and, with great difficulty, managed to manoeuvre her on to a sandbank. The Carnegie was described as a non-magnetic yacht, and carried very little metal, because of the delicate instruments with which she was fitted. In the absence of a steel rope, it was impossible to moor the ship. Presently the ropes holding the remaining anchors were consumed and the anchors fell into the sea. Everything on board seemed to be highly combustible. Gradually the flames crept forward after eating away the aftermast, and they caused the yards on the foremast to tumble down one by one.

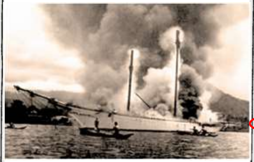

November 29th, 1929, Apia Harbour (*Carnegie Institution*)

"The story of individual endeavour and enterprise, of invention and accomplishment, cannot be told." Captain J.P. Ault (August 26th, 1929) *Carnegie Institution of Washington Publication* 568, page iii

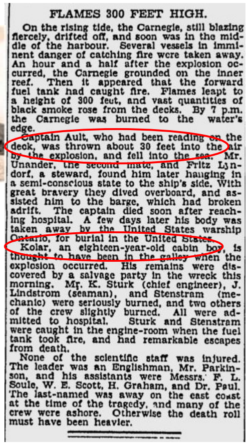

FLAMES 300 FEET HIGH.

On the rising tide, the Carnegie, still blazing fiercely, drifted off, and soon was in the middle of the harbour. Several vessels in imminent danger of catching fire were taken away. An hour and a half after the explosion occurred, the Carnegie grounded on the inner reef. Then it appeared that the forward fuel tank had caught fire. Flames leapt to a height of 300 feet, and vast quantities of black smoke rose from the decks. By 7 p.m. the Carnegie was burned to the water's edge.

Captain Ault, who had been reading on the deck, was thrown about 30 feet into the air by the explosion, and fell into the sea. Mr. Unander, the second mate, and Fritz Lyndorf, a steward, found him later hanging in a semi-conscious state to the ship's side. With great bravery they dived overboard, and assisted him to the barge, which had broken adrift. The captain died soon after reaching hospital. A few days later his body was taken away by the United States warship Ontario, for burial in the United States.

Kolar, an eighteen-year-old cabin boy, is thought to have been in the galley when the explosion occurred. His remains were discovered by a salvage party in the wreck this morning. Mr. K. Sturk (chief engineer), J. Lindstrom (seaman), and Stenstram (mechanic) were seriously burned, and two others of the crew slightly burned. All were admitted to hospital. Sturk and Stenstram were caught in the engine-room when the fuel tank took fire, and had remarkable escapes from death.

None of the scientific staff was injured. The leader was an Englishman, Mr. Parkinson, and his assistants were Messrs. F. L. Soule, W. E. Scott, H. Graham, and Dr. Paul. The last-named was away on the east coast at the time of the tragedy, and many of the crew were ashore. Otherwise the death roll must have been heavier.

The Sydney Morning Herald - Dec 13th, 1929

Cruise 7 observations

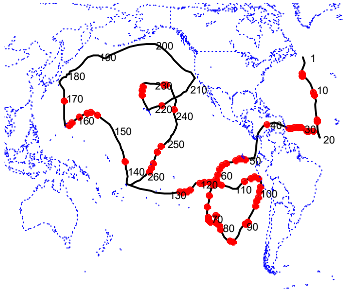

Number of days from outset marked: red points are those days which were subsequently identified as electrically undisturbed

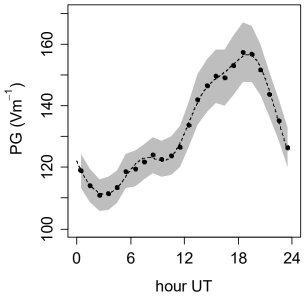

- •From Aug 7th 1928 to Nov 18th 1929 the Carnegie was on open sea for 317 days, yielding 181 days of continuous PG data
- •5 reduction factor experiments (Kitts Point, (Maryland), Engey Island, (Reykavik), Bridgetown, (Barbados) Easter Island and Apia, Samoa)

R.G. Harrison, The Carnegie Curve *Surv Geophys* **34**, 2, 209-232, DOI: 10.1007/s10712-012-9210-2 (2013)

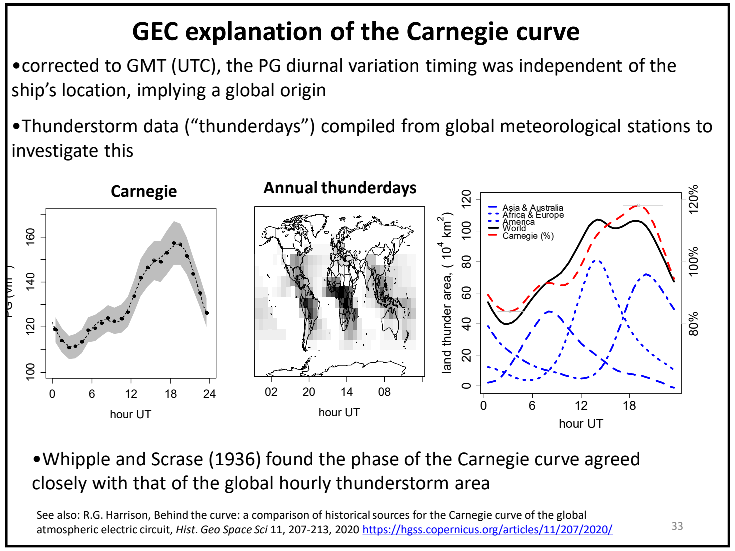

GEC explanation of the Carnegie curve

- corrected to GMT (UTC), the PG diurnal variation timing was independent of the ship's location, implying a global origin
- Thunderstorm data ("thunderdays") compiled from global meteorological stations to investigate this

- Whipple and Scrase (1936) found the phase of the Carnegie curve agreed closely with that of the global hourly thunderstorm area

See also: R.G. Harrison, Behind the curve: a comparison of historical sources for the Carnegie curve of the global atmospheric electric circuit, *Hist. Geo Space Sci* 11, 207-213, 2020 https://hgss.copernicus.org/articles/11/207/2020/

33

33

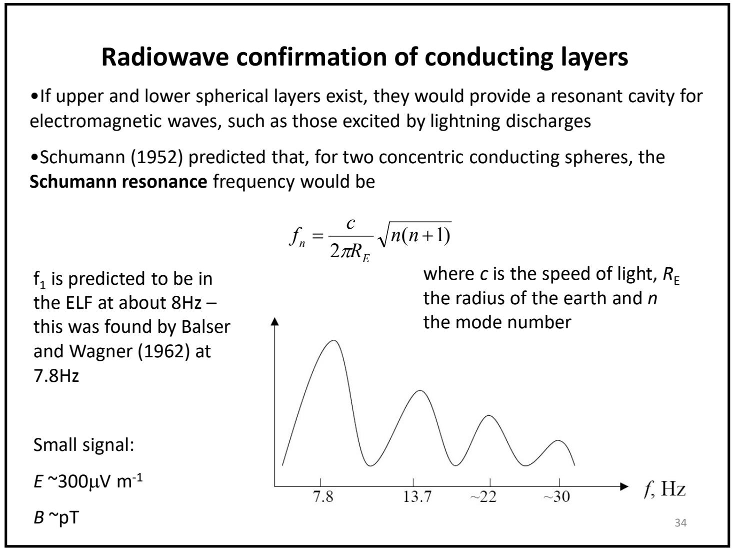

Radiowave confirmation of conducting layers

- If upper and lower spherical layers exist, they would provide a resonant cavity for electromagnetic waves, such as those excited by lightning discharges
- Schumann (1952) predicted that, for two concentric conducting spheres, the **Schumann resonance** frequency would be

$$f_n = \frac{c}{2\pi R_E}\sqrt{n(n+1)}$$

where *c* is the speed of light, R_E the radius of the earth and *n* the mode number

f_1 is predicted to be in the ELF at about 8Hz – this was found by Balser and Wagner (1962) at 7.8Hz

Small signal:

E ~300µV m^{-1}

B ~pT

34

34

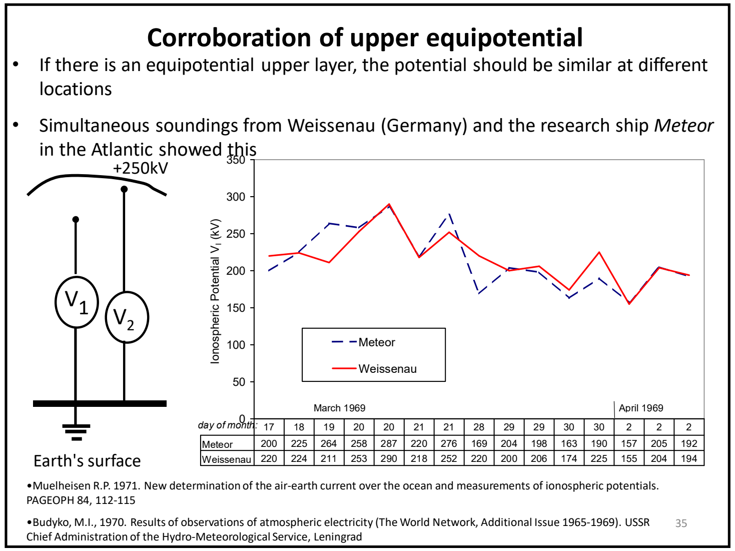

Corroboration of upper equipotential

- If there is an equipotential upper layer, the potential should be similar at different locations
- Simultaneous soundings from Weissenau (Germany) and the research ship *Meteor* in the Atlantic showed this

day of month:	17	18	19	20	20	21	21	28	29	29	30	30	2	2	2
Meteor	200	225	264	258	287	220	276	169	204	198	163	190	157	205	192
Weissenau	220	224	211	253	290	218	252	220	200	206	174	225	155	204	194

- Muelheisen R.P. 1971. New determination of the air-earth current over the ocean and measurements of ionospheric potentials. PAGEOPH 84, 112-115
- Budyko, M.I., 1970. Results of observations of atmospheric electricity (The World Network, Additional Issue 1965-1969). USSR Chief Administration of the Hydro-Meteorological Service, Leningrad

35

35

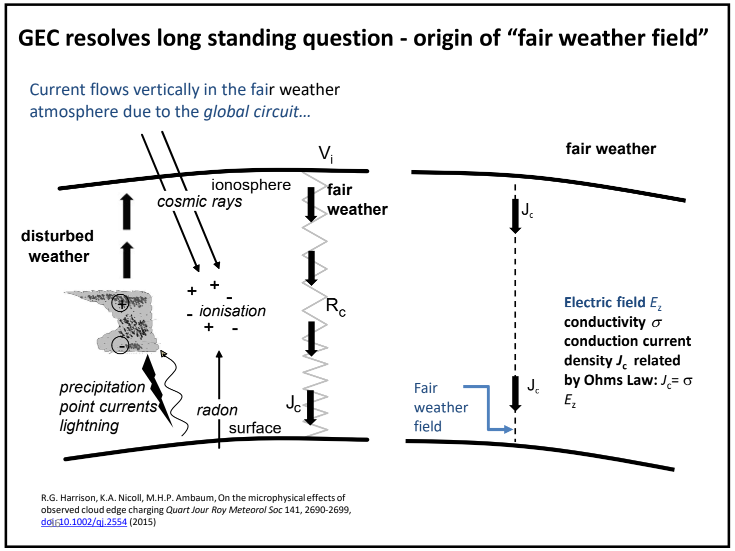

GEC resolves long standing question - origin of "fair weather field"

Current flows vertically in the fair weather atmosphere due to the *global circuit…*

Electric field E_z conductivity σ conduction current density J_c related by Ohms Law: $J_c = \sigma E_z$

R.G. Harrison, K.A. Nicoll, M.H.P. Ambaum, On the microphysical effects of observed cloud edge charging *Quart Jour Roy Meteorol Soc* 141, 2690-2699, doi:10.1002/qj.2554 (2015)

36

36

3. **Measurements of global circuit parameters**

1. Potential Gradient (field mill, long wire antenna)
2. Air-earth current (Wilson apparatus, Lerwick plate, and GDACCS)
3. Air conductivity
4. Fair weather classification

PG - a very variable quantity

- With **thunderclouds** present, the PG obviously becomes highly variable.
- **But convective showers** and rainfall also lead to considerable variability
- And…even with low level layer cloud - no ice and minimal convection - the PG can be quite variable – is it due to space weather, or earth weather? --> reliable series of measurements are needed.

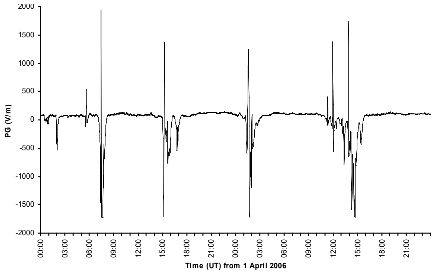

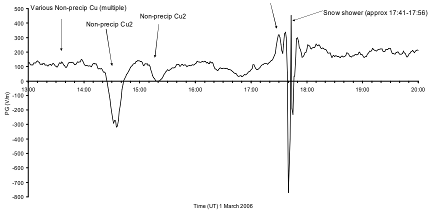

The field mill

•Robust and durable instrument for measuring atmospheric **electric field**

•Horizontal plate electrode in which a charge is induced by the atmospheric electric field

•This is compared with the charge induced under zero electric field conditions (when electrode covered).

=> Electrode is alternately covered and exposed by a mechanical shutter, which is driven by a motor – **synchronous detection** used to detect changes.

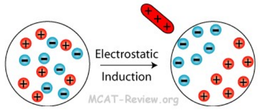

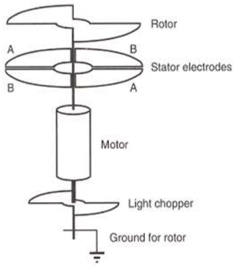

Operates continuously at the Observatory – see the real time graphs

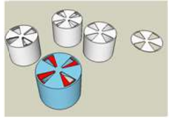

Commercial field mills

- Electric field mills now widely commercially available
- Not so for other atmospheric electricity instruments

Boltek EFM100

CAUTION
MOVING PARTS
BOLTEK

Campbell CS110

Chilworth JCI131

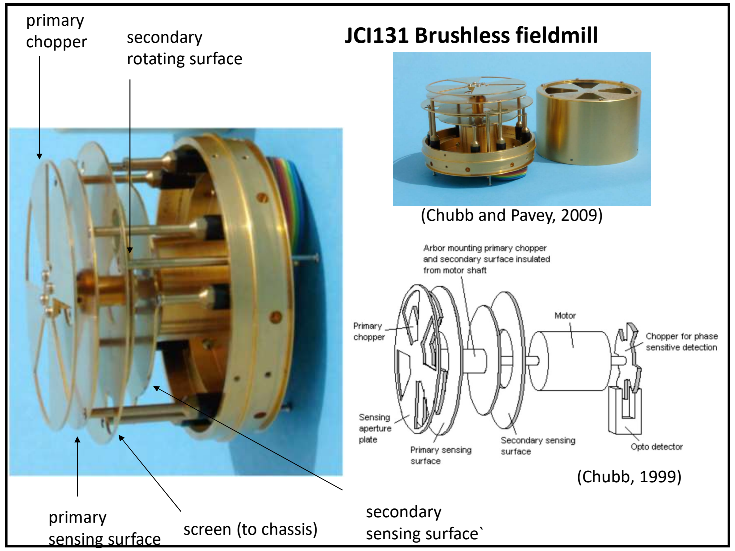

JCI131 Brushless fieldmill

primary
chopper
secondary
rotating surface
primary
sensing surface
screen (to chassis)
secondary
sensing surface`
Arbor mounting primary chopper and secondary surface insulated from motor shaft
Primary chopper
Motor
Chopper for phase sensitive detection
Sensing aperture plate
Primary sensing surface
Secondary sensing surface
Opto detector

(Chubb and Pavey, 2009)

(Chubb, 1999)

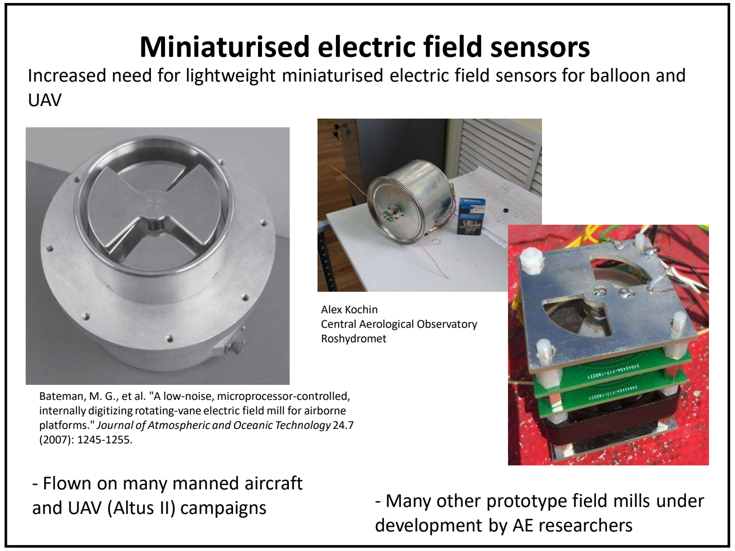

Miniaturised electric field sensors

Increased need for lightweight miniaturised electric field sensors for balloon and UAV

Bateman, M. G., et al. "A low-noise, microprocessor-controlled, internally digitizing rotating-vane electric field mill for airborne platforms." *Journal of Atmospheric and Oceanic Technology* 24.7 (2007): 1245-1255.

Alex Kochin
Central Aerological Observatory
Roshydromet

- Flown on many manned aircraft and UAV (Altus II) campaigns

- Many other prototype field mills under development by AE researchers

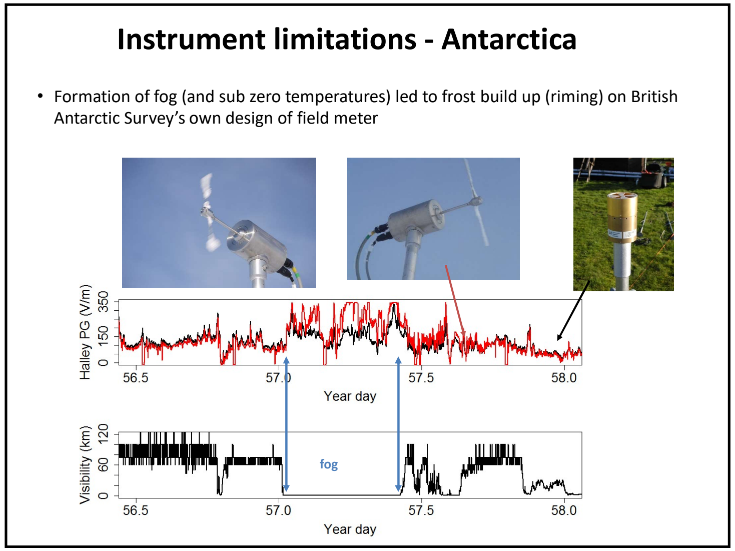

Instrument limitations - Antarctica

- Formation of fog (and sub zero temperatures) led to frost build up (riming) on British Antarctic Survey's own design of field meter

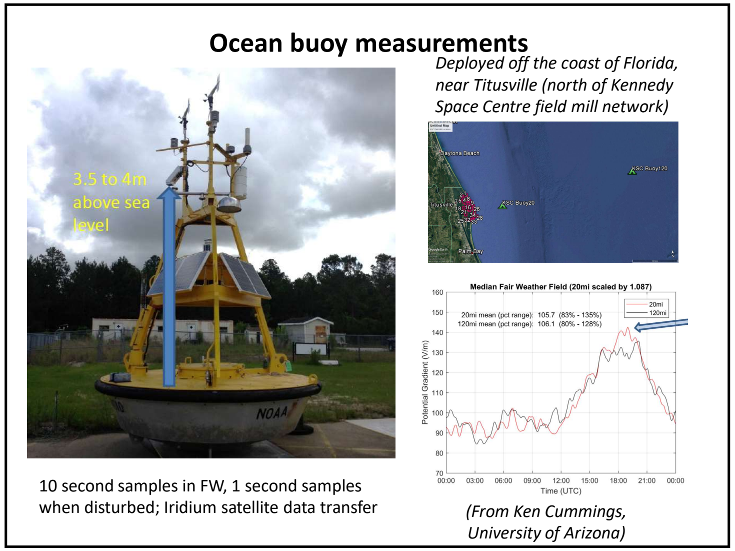

Ocean buoy measurements

Deployed off the coast of Florida, near Titusville (north of Kennedy Space Centre field mill network)

10 second samples in FW, 1 second samples when disturbed; Iridium satellite data transfer

(From Ken Cummings, University of Arizona)

Standardisation – e.g. corrections for ship distortion

- On the **Carnegie voyages**, a “reduction factor” was needed to correct for the distortion of the ship and its rigging
- comparison made with flat shore sites close to level with the sea, free from trees, using a 15 to 20m long horizontal “passive wire” antenna
- Difficulty was in getting ship within half-mile…so reduction factor was measured often and averaged
- Reduction factor practically constant for PG 120 to 480V m^{-1}, 2.85 for mainsail up and boom to port or starboard, and 3.77 for mainsail down, with boom over port crutch

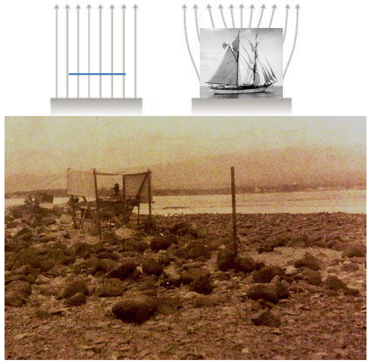

Potential Gradient “reduction-factor station”, Watson’s Island, Apia, Samoa (April 1929). Spiders caused frequent trouble by spinning webs in the cap of the shore electrometer and in the caps of the supporting insulators of the stretched wire system. “The spiders were numerous. Several hours of record were affected by the presence of spider webs.”

Passive wire antenna

- Ultra well-insulated horizontal wire positioned at known height naturally acquires the potential of the atmosphere

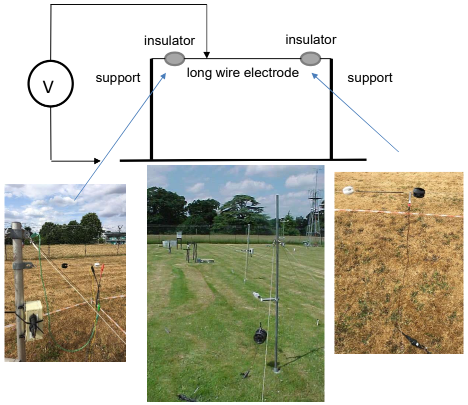

- Rate of acquisition of potential on the wire is not artificially accelerated by an equaliser
=> instrument is called the **passive wire**

- Wire potential measured using a high impedance voltmeter

- No calibration or geometrical corrections required as long as there are no distorting objects nearby =>value of PG at 1m measured using passive wire is considered absolute

Vertical array of passive antennas

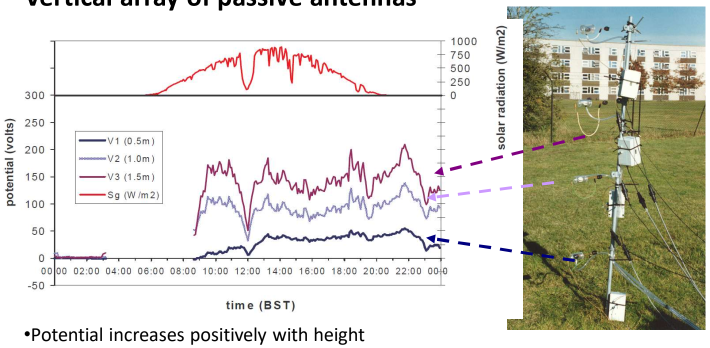

- Potential increases positively with height
- Continuous turbulent mixing of ions is required for effective operation
- Note the effect of a cloud passing around noon BST

J.F. Barlow and R.G. Harrison (1999) Turbulent transfer of space charge in the atmospheric surface layer, *Proceedings Int Conf Atmos Elect, Huntsville Alabama*

Cruise of the NRP* Sagres

*Navio da República Portuguesa

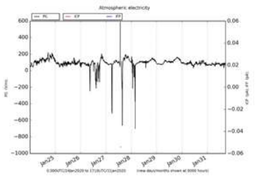

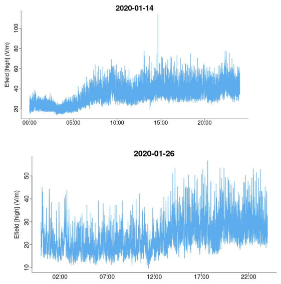

https://www.vesselfinder.com/vessels/PT-NAVY-SAGRES-IMO-0-MMSI-263141000

Conduction current – the Lerwick plate

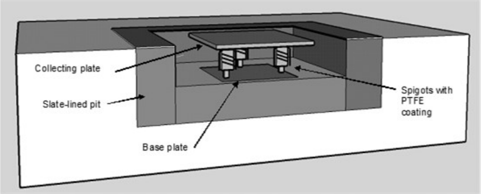
Collecting plate
Slate-lined pit
Base plate
Spigots with PTFE coating

Schematic of 0.5m² air-earth current collecting plate, operating flush with the ground in a slate-lined pit

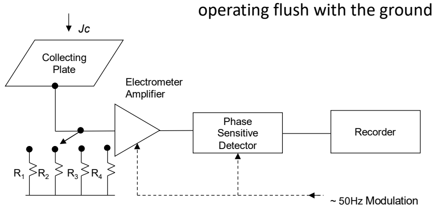
Jc
Collecting Plate
Electrometer Amplifier
Phase Sensitive Detector
Recorder
R1
R2
R3
R4
~ 50Hz Modulation

Block diagram of electronic system. (*Input resistor values:* R1 10GΩ, R2 25 GΩ, R3 50 GΩ, R4 100 GΩ.)

R.G. Harrison and K.A. Nicoll**,** Air-earth current density measurements at Lerwick; implications for seasonality in the global electric circuit *Atmospheric Research* **89***,* 1-2, 181-193, doi:10.1016/j.atmosres.2008.01.008

Vertical current density instrumentation

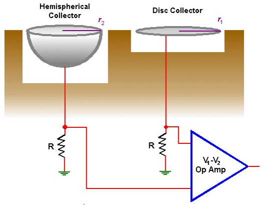
Hemispherical Collector
Disc Collector
r2
r1
R
R
V1-V2 Op Amp

Mukku, V. N. R., (1984). A design for eliminating displacement currents in the air-earth current measurement, *Journal of Physics* **17**(E) 629.

displacement current J_D – proportional to total area - and conduction current J_c – proportional to cross section

Reading “GDACCS” instrument uses two electrodes with different geometry

- heated, to repel water and insects

Bennett, A.J. and Harrison, R.G., (2008). Surface measurement system for the atmospheric electrical vertical conduction current density, with displacement current density correction, *J Atmos & Solar-Terrestrial Physics* **70** 1373–1381.

Reading

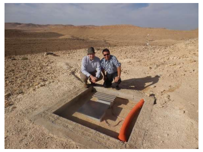

Negev

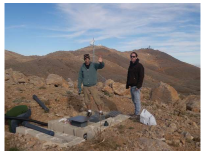

Mt Hermon

J_c and PG measured during fog

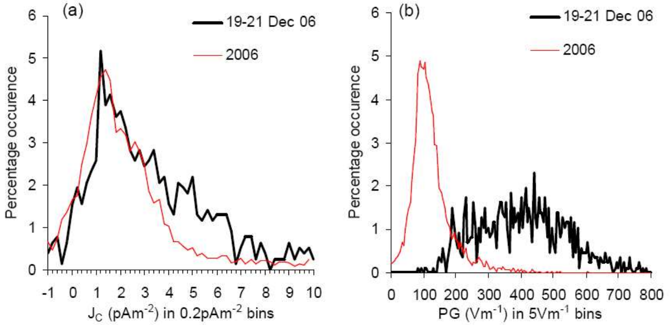

A.J.Bennett Measurements of Atmospheric electricity during different meteorological conditions, PhD Thesis, Reading, 2007

- Vertical current density affected little by fog layer
- PG greatly affected by low conductivity of foggy air

Changes in electrical parameters during fog

Quantity	Fair weather (FW)	Foggy	Ratio foggy/FW
Conduction current density	1.4 pAm^{-2}	1.2 pAm^{-2}	0.9
Potential Gradient	100 Vm^{-1}	400 Vm^{-1}	4.0
Air conductivity (calculated)	14 fSm^{-1}	3 fSm^{-1}	0.2

Table 5-1. Comparison of Ohm’s Law parameters in foggy and fair weather.

- PG in fog is ~4 to 5 times larger than FW
- PG measurements in fog can be used to estimate in fog conductivity

→ Conductivity in fog ~ 5 times smaller than clear air

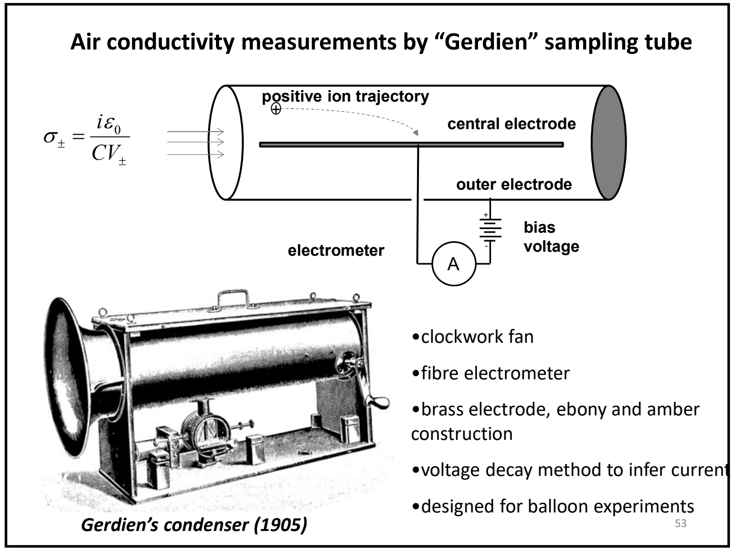

Air conductivity measurements by “Gerdien” sampling tube

$$\sigma_{\pm} = \frac{i\varepsilon_0}{CV_{\pm}}$$

positive ion trajectory
central electrode
outer electrode
bias voltage
electrometer
A

- clockwork fan
- fibre electrometer
- brass electrode, ebony and amber construction
- voltage decay method to infer current
- designed for balloon experiments

Gerdien’s condenser (1905)

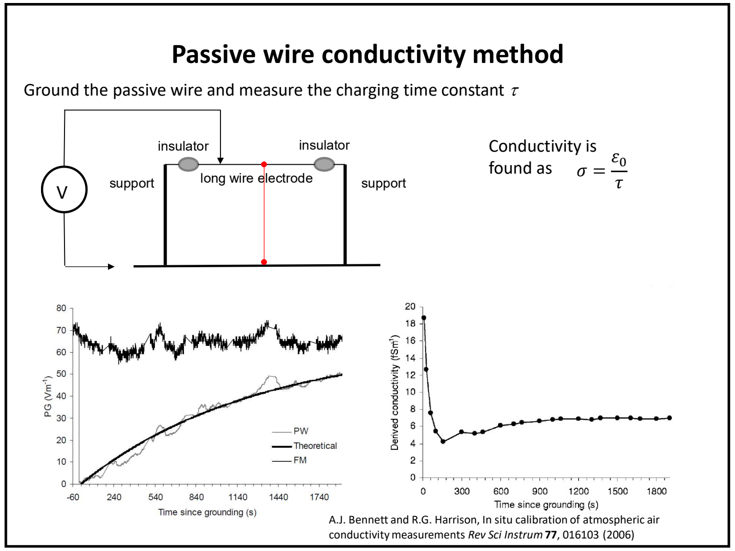

Passive wire conductivity method

Ground the passive wire and measure the charging time constant τ

Conductivity is found as $\sigma = \frac{\varepsilon_0}{\tau}$

V
support
insulator
long wire electrode
insulator
support
80
70
60
50
40
30
20
10
0
PG (Vm⁻¹)
-60
240
540
840
1140
1440
1740
Time since grounding (s)
PW
Theoretical
FM
20
18
16
14
12
10
8
6
4
2
0
Derived conductivity (fSm⁻¹)
0
300
600
900
1200
1500
1800
Time since grounding (s)

A.J. Bennett and R.G. Harrison, In situ calibration of atmospheric air conductivity measurements *Rev Sci Instrum* **77**, 016103 (2006)

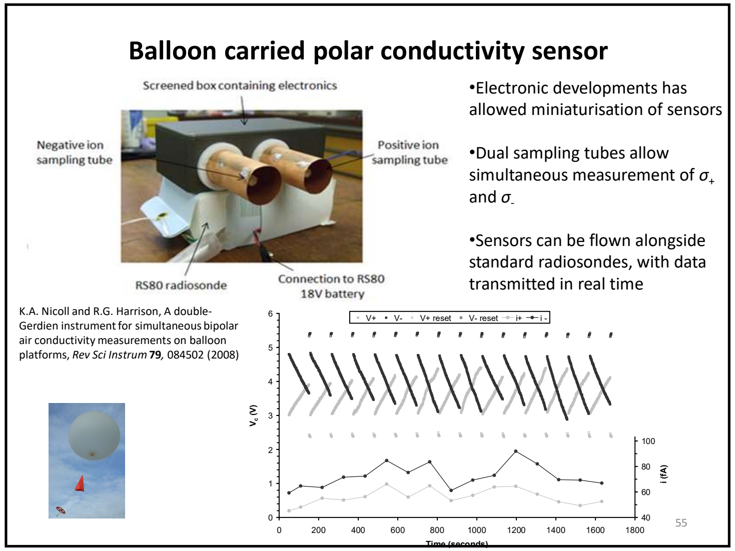

Balloon carried polar conductivity sensor

Screened box containing electronics
Negative ion sampling tube
Positive ion sampling tube
RS80 radiosonde
Connection to RS80 18V battery

- Electronic developments has allowed miniaturisation of sensors
- Dual sampling tubes allow simultaneous measurement of σ_+ and σ_-
- Sensors can be flown alongside standard radiosondes, with data transmitted in real time

K.A. Nicoll and R.G. Harrison, A double-Gerdien instrument for simultaneous bipolar air conductivity measurements on balloon platforms, *Rev Sci Instrum* **79**, 084502 (2008)

V+
V-
V+ reset
V- reset
i+
i -
6
5
4
3
2
1
0
Vc (V)
100
80
60
40
i (fA)
0
200
400
600
800
1000
1200
1400
1600
1800
Time (seconds)

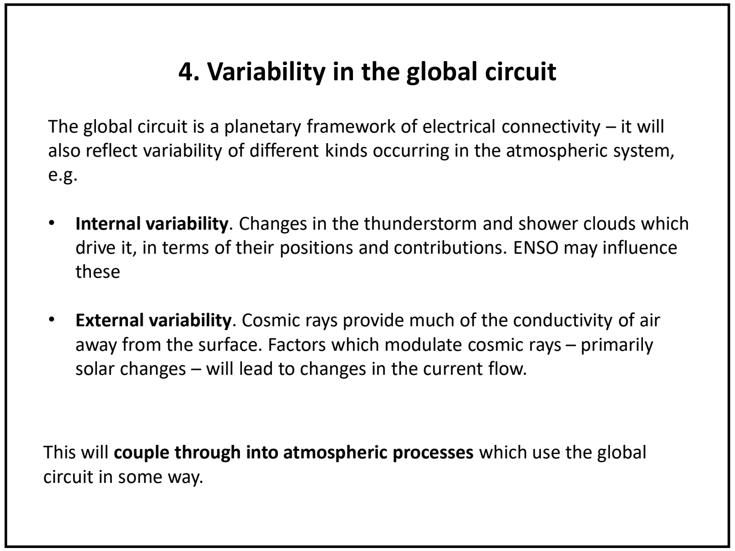

4. Variability in the global circuit

The global circuit is a planetary framework of electrical connectivity – it will also reflect variability of different kinds occurring in the atmospheric system, e.g.

- **Internal variability**. Changes in the thunderstorm and shower clouds which drive it, in terms of their positions and contributions. ENSO may influence these
- **External variability**. Cosmic rays provide much of the conductivity of air away from the surface. Factors which modulate cosmic rays – primarily solar changes – will lead to changes in the current flow.

This will **couple through into atmospheric processes** which use the global circuit in some way.

Global circuit variability

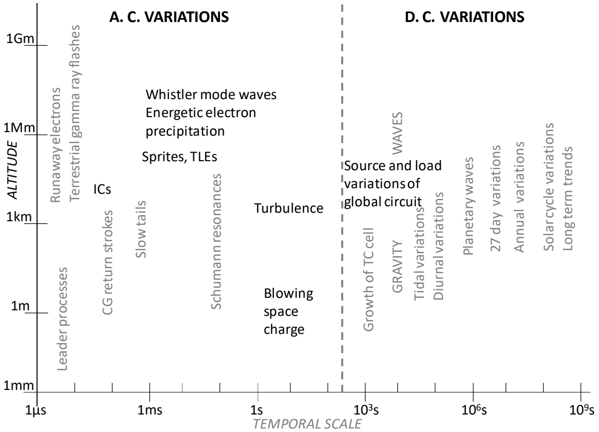
A. C. VARIATIONS
D. C. VARIATIONS
1Gm
1Mm
1km
1m
1mm
ALTITUDE
Runaway electrons
Terrestrial gamma ray flashes
Leader processes
ICs
CG return strokes
Slow tails
Whistler mode waves
Energetic electron
precipitation
Sprites, TLEs
Schumann resonances
Turbulence
Blowing
space
charge
Source and load
variations of
global circuit
WAVES
Growth of TC cell
GRAVITY
Tidal variations
Diurnal variations
Planetary waves
27 day variations
Annual variations
Solar cycle variations
Long term trends
1µs
1ms
1s
10³s
10⁶s
10⁹s
TEMPORAL SCALE

M.J. Rycroft, K.A. Nicoll, K.L. Aplin and R.G. Harrison, Recent advances in global electric circuit coupling between the space environment and the troposphere *J.Atmos Sol-Terr Phys* 90–91, 198–211 (2012),

ENSO effects on lightning

- •Intensity, flash rate and location of lightning varies with ENSO timescale
- •10% change in global lightning flash rate from La Niña to El Niño

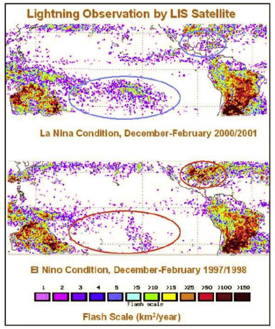
Lightning Observation by LIS Satellite
La Nina Condition, December-February 2000/2001
El Nino Condition, December-February 1997/1998
Flash Scale (km²/year)

More lightning in Pacific during(cool) La Niña phase than (warm) El Niño

Complex relationship between regional variations in lightning and El Niño, La Niña . i.e. very location dependent

Satori G., et al, Variability of global lightning activity on the ENSO time scale, Atmospheric Research, 91, 500-507, 2009

ENSO effects on surface PG

GEC integrates effects of all thunderstorms active around globe

Signals from large scale variations may be seen in distant PG measurements i.e. possible remote sensing method for electrically active storms

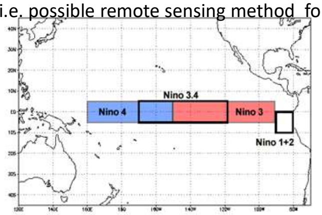
Nino 4
Nino 3.4
Nino 3
Nino 1+2

Lerwick PG measurements compared to Niño 3.4 index (SSTs from area in Pacific ocean, bounded by 5°N-5°S and 120°-170°W)

Image from www.ncdc.noaa.gov

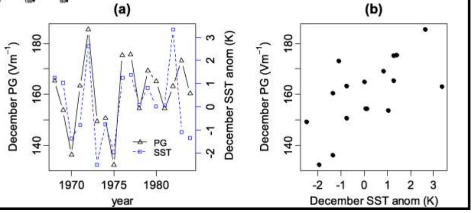

Remarkably strong correlation – PG large during El Niño and small during La Niña

Increase in GEC of 3.5% for 1K of SST anomaly

ENSO: Internal variability and Carnegie curve

El Niño Southern Oscillation shows a relationship with the Carnegie curve of the global circuit in atmospheric electricity – mean values increase with Pacific sea surface temperature

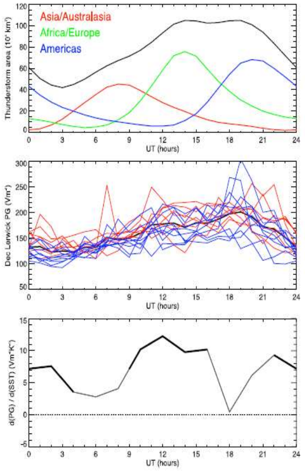

Red lines show Carnegie curve in atmospheric electricity at Lerwick (for December) during **El Niño** conditions, and *blue lines* during *La Niña*

→ sensitivity of ~ 5 Vm^{-1} K^{-1}

Greatest PG sensitivity to SST change is between 10-16UT => African lightning

R.G. Harrison, M. Joshi, K. Pascoe, Inferring convective responses to El Niño with atmospheric electricity measurements at Shetland *Environ Res Lett* 6 (2011) 044028

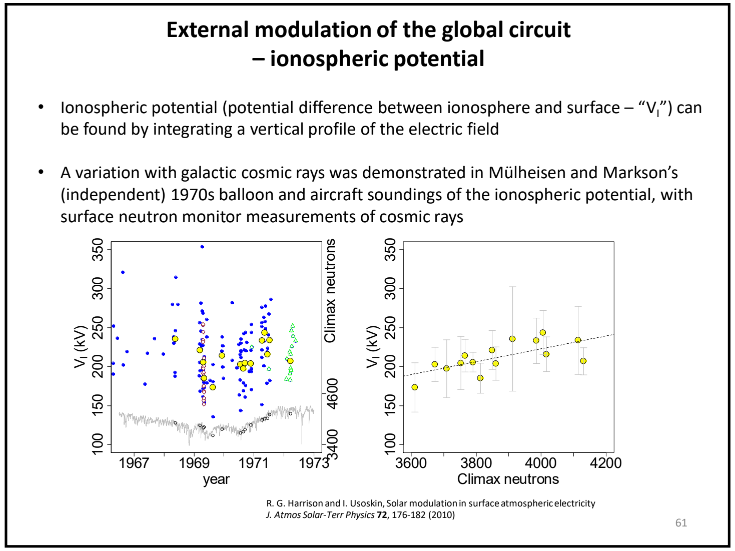

External modulation of the global circuit – ionospheric potential

- Ionospheric potential (potential difference between ionosphere and surface – "V_I") can be found by integrating a vertical profile of the electric field
- A variation with galactic cosmic rays was demonstrated in Mülheisen and Markson's (independent) 1970s balloon and aircraft soundings of the ionospheric potential, with surface neutron monitor measurements of cosmic rays

V_I (kV)
100
150
200
250
300
350
1967
1969
1971
1973
year
Climax neutrons
3400
4600
V_I (kV)
3600
3800
4000
4200
Climax neutrons

R. G. Harrison and I. Usoskin, Solar modulation in surface atmospheric electricity *J. Atmos Solar-Terr Physics* **72**, 176-182 (2010)

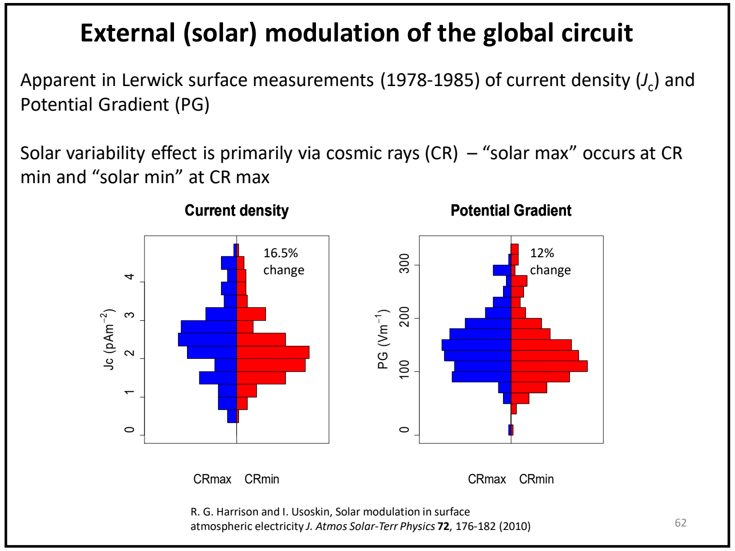

External (solar) modulation of the global circuit

Apparent in Lerwick surface measurements (1978-1985) of current density (J_c) and Potential Gradient (PG)

Solar variability effect is primarily via cosmic rays (CR) – "solar max" occurs at CR min and "solar min" at CR max

R. G. Harrison and I. Usoskin, Solar modulation in surface atmospheric electricity *J. Atmos Solar-Terr Physics* **72**, 176-182 (2010)

5. Applications and planetary global circuits

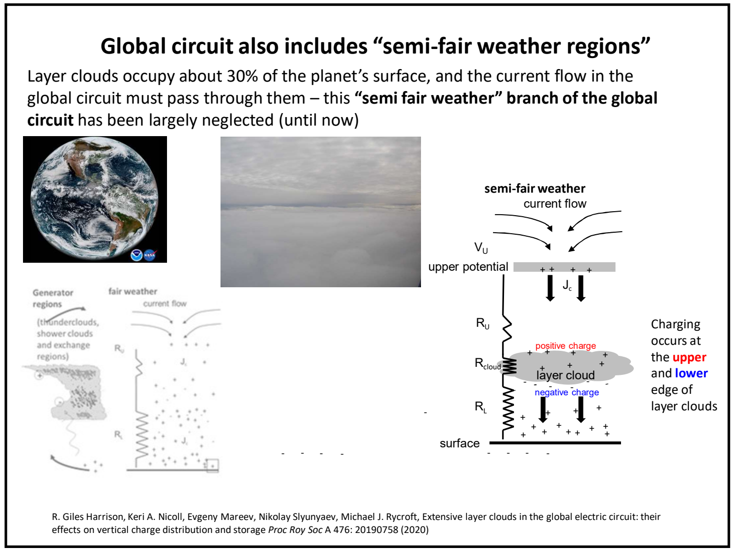

Global circuit also includes "semi-fair weather regions"

Layer clouds occupy about 30% of the planet's surface, and the current flow in the global circuit must pass through them – this **"semi fair weather" branch of the global circuit** has been largely neglected (until now)

Generator regions
(thundercloulds, shower clouds and exchange regions)
fair weather
current flow
R_U
J_c
R_L
semi-fair weather
current flow
V_U
upper potential
J_c
R_U
positive charge
R_{cloud}
layer cloud
negative charge
R_L
surface

Charging occurs at the **upper** and **lower** edge of layer clouds

R. Giles Harrison, Keri A. Nicoll, Evgeny Mareev, Nikolay Slyunyaev, Michael J. Rycroft, Extensive layer clouds in the global electric circuit: their effects on vertical charge distribution and storage *Proc Roy Soc* A 476: 20190758 (2020)

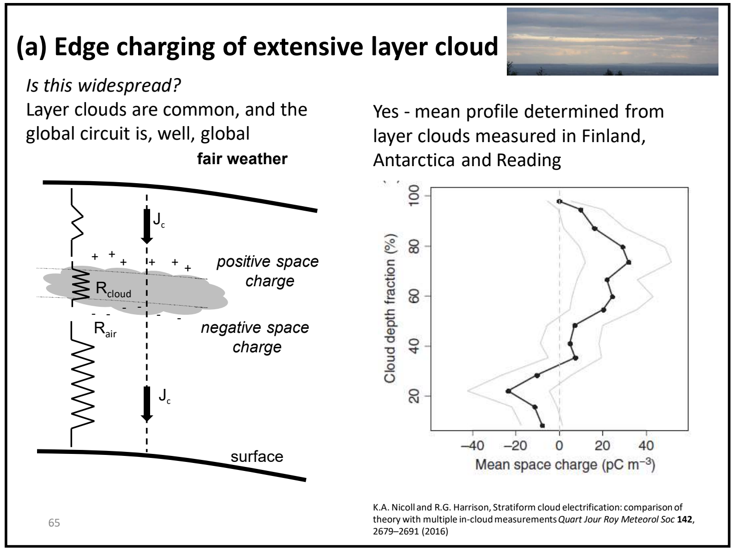

(a) Edge charging of extensive layer cloud

Is this widespread?
Layer clouds are common, and the global circuit is, well, global

Yes - mean profile determined from layer clouds measured in Finland, Antarctica and Reading

fair weather
J_c
positive space charge
R_{cloud}
R_{air}
negative space charge
J_c
surface
Cloud depth fraction (%)
100
80
60
40
20
−40
−20
0
20
40
Mean space charge (pC m^{-3})

K.A. Nicoll and R.G. Harrison, Stratiform cloud electrification: comparison of theory with multiple in-cloud measurements *Quart Jour Roy Meteorol Soc* **142**, 2679–2691 (2016)

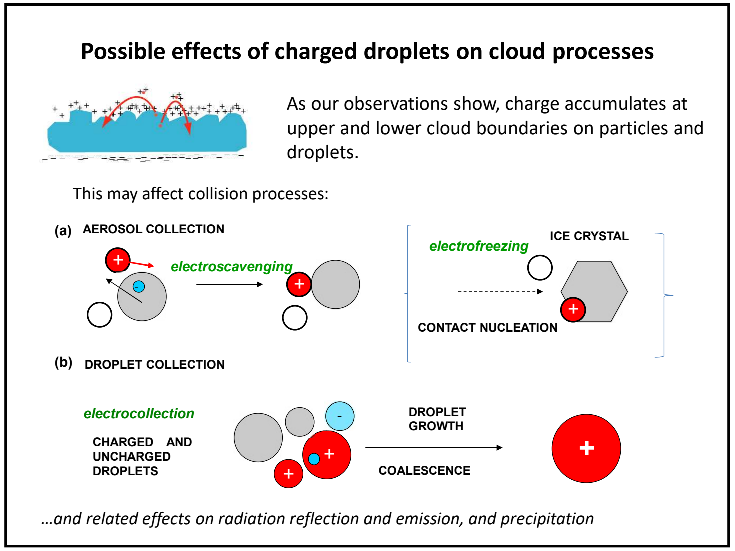

Possible effects of charged droplets on cloud processes

As our observations show, charge accumulates at upper and lower cloud boundaries on particles and droplets.

This may affect collision processes:

(a) AEROSOL COLLECTION
electroscavenging
electrofreezing
ICE CRYSTAL
CONTACT NUCLEATION
(b) DROPLET COLLECTION
electrocollection
CHARGED AND UNCHARGED DROPLETS
DROPLET GROWTH
COALESCENCE

…and related effects on radiation reflection and emission, and precipitation

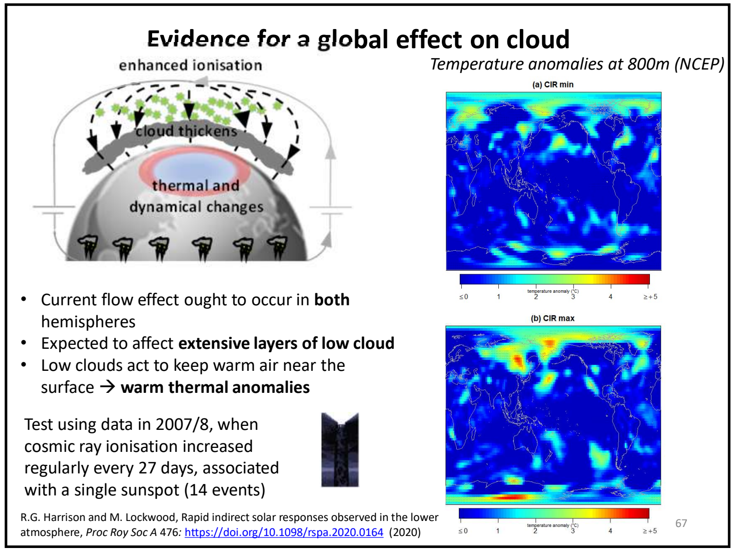

Evidence for a global effect on cloud

enhanced ionisation
cloud thickens
thermal and dynamical changes

Temperature anomalies at 800m (NCEP)

(a) CIR min
≤0
1
2
3
4
≥+5
temperature anomaly (°C)
(b) CIR max
≤0
1
2
3
4
≥+5
temperature anomaly (°C)

- Current flow effect ought to occur in **both** hemispheres
- Expected to affect **extensive layers of low cloud**
- Low clouds act to keep warm air near the surface → **warm thermal anomalies**

Test using data in 2007/8, when cosmic ray ionisation increased regularly every 27 days, associated with a single sunspot (14 events)

R.G. Harrison and M. Lockwood, Rapid indirect solar responses observed in the lower atmosphere, *Proc Roy Soc A* 476: https://doi.org/10.1098/rspa.2020.0164 (2020)

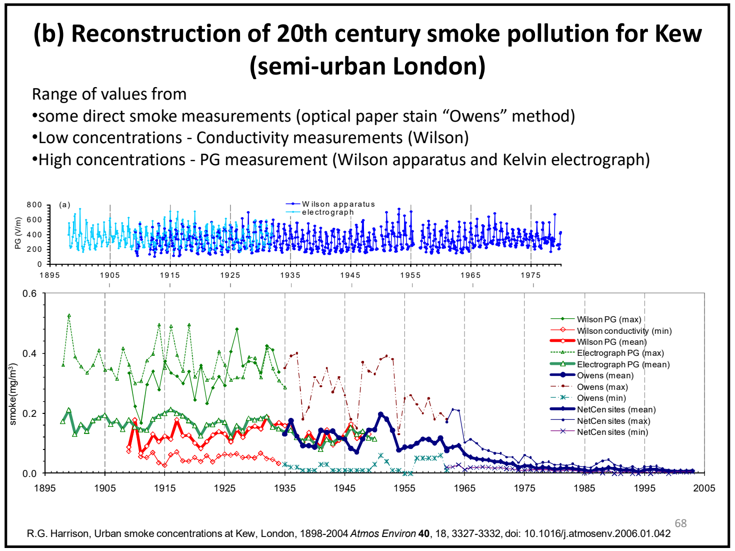

(b) Reconstruction of 20th century smoke pollution for Kew (semi-urban London)

Range of values from
•some direct smoke measurements (optical paper stain "Owens" method)
•Low concentrations - Conductivity measurements (Wilson)
•High concentrations - PG measurement (Wilson apparatus and Kelvin electrograph)

(a)
Wilson apparatus
electrograph
PG (V/m)
800
600
400
200
0
1895
1905
1915
1925
1935
1945
1955
1965
1975
smoke (mg/m³)
0.6
0.4
0.2
0.0
1895
1905
1915
1925
1935
1945
1955
1965
1975
1985
1995
2005
Wilson PG (max)
Wilson conductivity (min)
Wilson PG (mean)
Electrograph PG (max)
Electrograph PG (mean)
Owens (mean)
Owens (max)
Owens (min)
NetCen sites (mean)
NetCen sites (max)
NetCen sites (min)

R.G. Harrison, Urban smoke concentrations at Kew, London, 1898-2004 *Atmos Environ* **40**, 18, 3327-3332, doi: 10.1016/j.atmosenv.2006.01.042

(c) Detecting releases of artificial radioactivity

…whichs causes increase in surface air conductivity

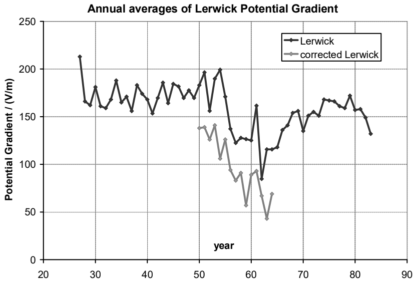

Radioactivity increases number of ions
=> Conductivity increases

$F = J_c/\sigma$

Annual averages of Fair Weather PG obtained at Lerwick, showing the period of atmospheric nuclear weapons testing 1956-1962.

(corrected values were derived by excluding "polluted" wind directions known to produce high PG measurements.)

Harrison, R.G., Twentieth century atmospheric electrical measurements at the observatories of Kew, Eskdalemuir and Lerwick *Weather* **58**, 11-19, (2003)

Observations of conductivity changes from radioactivity

Oxfordshire, direct measurements of ion production rate from ^{90}Sr

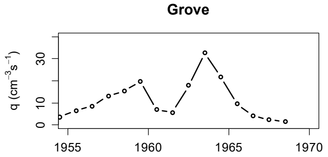

London, Wilson apparatus measurements of air conductivity

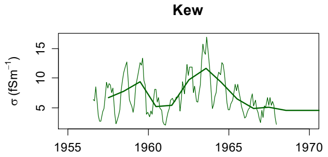

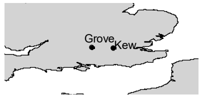

Gibson J.A.B, Richards J.E. and Docherty J. Nuclear radiation in the environment-beta and gamma ray does rates and air ionisation from 1951 to 1968. *JATP* **31**, 1183-1196 (1969)

R.G. Harrison and W.J. Ingram, 2005, Air-earth current measurements at Kew, London, 1909-1979 *Atmos Res* **76**, (1-4), 49-64

Fukushima reactor accident

Release on 14th March 2011

Kelvin water dropper sensor

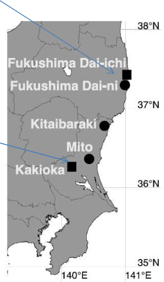

Figure 2. The locations of Fukushima Dai-ichi nuclear power plant (FNPP), Fukushima Dai-ni nuclear power plant (about 12 km south of FNPP), Kitaibaraki (about 70 km southwest of FNPP), Mito (about 130 km southwest of FNPP), and Kakioka (about 150 km southwest of FNPP).

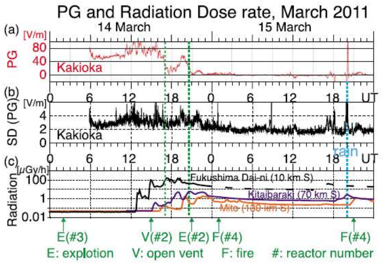

Figure 3. Two-day's data before and after the largest southbound release of the radioactive material from the Fukushima Dai-ichi nuclear power plant (FNPP) in March 2011: (a) 1-min averaged PG at Kakioka, (b) standard deviation of PG calculated from 1 Hz sampling data, and (c) 10 min resolution radiation dose rate data at three different stations between Kakioka and the FNPP. The green notes at the bottom denote incidents at the FNPP.

From Takeda et al, 2013, *Geophys Res Lett* 38, L15811, doi:10.1029/2011GL048511, 2011

Planetary global circuits -remote detection requirements

sensing possibilities:

requirement	Charge generation		Lower conductive surface or region	Upper conductive region	Vertical current flow
	Electrical discharges	Precipitation			
Schumann resonances	✓		✓	✓	
Radar		✓	✓		
Broadband radio emission	✓				
Optical	✓				

Detection of vertical current flow within an atmosphere is not amenable to remote sensing… historically, terrestrial method has been to install surface electrodes

K.L. Aplin, R.G. Harrison and M.J. Rycroft, *Space Science Reviews* **137**, 11-27 (2008)

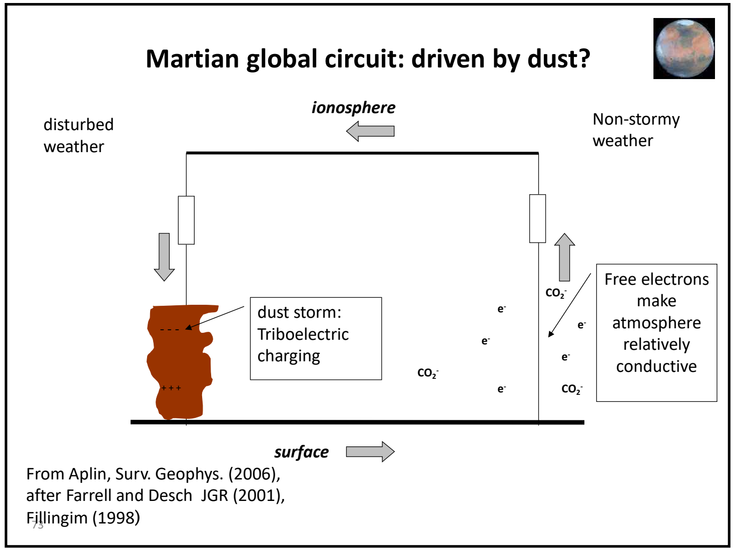

Martian global circuit: driven by dust?

disturbed
weather

Non-stormy
weather

dust storm:
Triboelectric
charging

Free electrons
make
atmosphere
relatively
conductive

From Aplin, Surv. Geophys. (2006),
after Farrell and Desch JGR (2001),
Fillingim (1998)

73

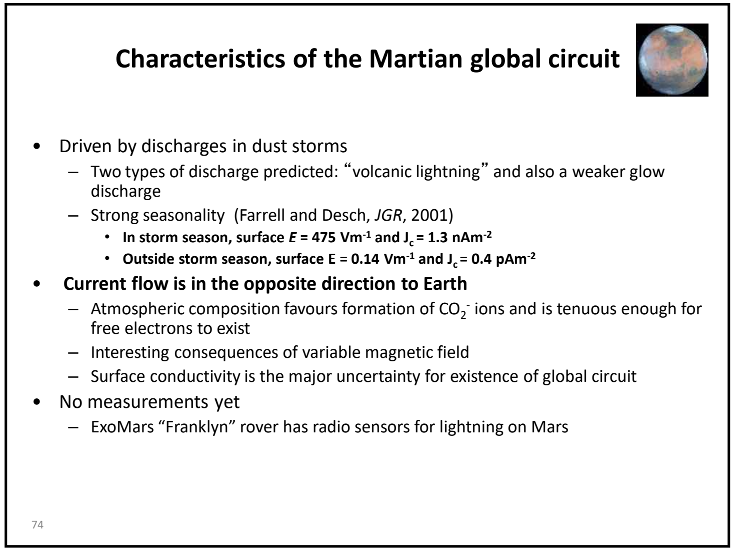

Characteristics of the Martian global circuit

- Driven by discharges in dust storms
  - Two types of discharge predicted: “volcanic lightning” and also a weaker glow discharge
  - Strong seasonality (Farrell and Desch, *JGR*, 2001)
    - **In storm season, surface *E* = 475 Vm^{-1} and J_c = 1.3 nAm^{-2}**
    - **Outside storm season, surface E = 0.14 Vm^{-1} and J_c = 0.4 pAm^{-2}**
- **Current flow is in the opposite direction to Earth**
  - Atmospheric composition favours formation of CO_2^- ions and is tenuous enough for free electrons to exist
  - Interesting consequences of variable magnetic field
  - Surface conductivity is the major uncertainty for existence of global circuit
- No measurements yet
  - ExoMars “Franklyn” rover has radio sensors for lightning on Mars

74

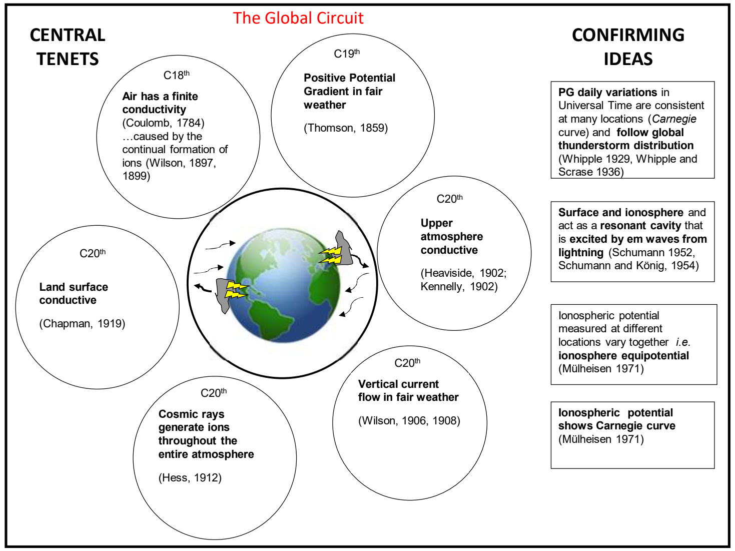

The Global Circuit

CENTRAL TENETS

C18th
Air has a finite conductivity (Coulomb, 1784) …caused by the continual formation of ions (Wilson, 1897, 1899)

C19th
Positive Potential Gradient in fair weather
(Thomson, 1859)

C20th
Upper atmosphere conductive
(Heaviside, 1902; Kennelly, 1902)

C20th
Land surface conductive
(Chapman, 1919)

C20th
Vertical current flow in fair weather
(Wilson, 1906, 1908)

C20th
Cosmic rays generate ions throughout the entire atmosphere
(Hess, 1912)

CONFIRMING IDEAS

PG daily variations in Universal Time are consistent at many locations (*Carnegie* curve) and **follow global thunderstorm distribution** (Whipple 1929, Whipple and Scrase 1936)

Surface and ionosphere and act as a **resonant cavity** that is **excited by em waves from lightning** (Schumann 1952, Schumann and König, 1954)

Ionospheric potential measured at different locations vary together *i.e.* **ionosphere equipotential** (Mülheisen 1971)

Ionospheric potential shows Carnegie curve (Mülheisen 1971)

75

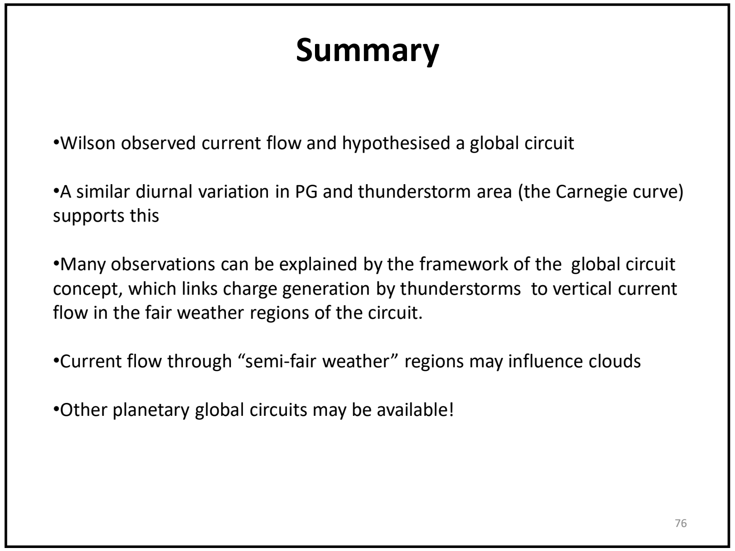

Summary

•Wilson observed current flow and hypothesised a global circuit

•A similar diurnal variation in PG and thunderstorm area (the Carnegie curve) supports this

•Many observations can be explained by the framework of the global circuit concept, which links charge generation by thunderstorms to vertical current flow in the fair weather regions of the circuit.

•Current flow through “semi-fair weather” regions may influence clouds

•Other planetary global circuits may be available!

76

PERGAMON

Journal of Atmospheric and Solar-Terrestrial Physics 62 (2000) 1563–1576

Journal of
ATMOSPHERIC AND
SOLAR-TERRESTRIAL
PHYSICS

www.elsevier.nl/locate/jastp

The global atmospheric electric circuit, solar activity and climate change

M.J. Rycroft[a,*], S. Israelsson[b], C. Price[c]

[a] International Space University, Boul. Gonthier d'Andernach, 67400 Illkirch, France
[b] Department of Earth Sciences, Uppsala University, Villavägen 16, SE-752 36 Uppsala, Sweden
[c] Department of Geophysics and Planetary Sciences, Tel Aviv University, Ramat Aviv 69978, Israel

Received 18 April 2000; accepted 14 June 2000

Abstract

The study of the global atmospheric electric circuit has advanced dramatically in the past 50 years. Large advances have been made in the areas of lightning and thunderstorm research, as related to the global circuit. We now have satellites looking down on the Earth continuously, supplying information on the temporal and spatial variability of lightning and thunderstorms. Thunderstorms are electric current generators, which drive electric currents up through the conducting atmosphere. They maintain the ionosphere at a potential of $\sim+250$ kV with respect to the Earth's surface. The global electric circuit is completed by currents $\sim$2 pA/m^2 flowing through the fair weather atmosphere, remote from thunderstorms, and by transient currents due to negative cloud-to-ground lightning discharges. The time constant of the circuit, $\sim > 2$ min, demonstrates that thunderstorms must occur continually to maintain the fair weather electric field. New discoveries have been made in the field of sprites, elves and blue jets, which may have a direct impact on the global circuit. Our knowledge of the global electric circuit modulated by solar effects has improved. Changes to the global circuit are associated with changes of conductivity linked with the time-varying presence of energetic charged particles, and the solar wind may influence the global electric circuit by inferred effects on cloud microphysics, temperature, and dynamics in the troposphere. We now have a better understanding of how the conductivity of the atmosphere is influenced by aerosols, and how this impacts our measurements of the fair-weather global circuit. The global atmospheric electric circuit is also beginning to be recognised by some climate researchers as a useful tool with which to study and monitor the Earth's changing climate. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Atmospheric electricity; Global circuit; Lightning; Climate change

1. Introduction to the global atmospheric electric circuit

One of the first, and certainly the most famous, demonstrations of a reproducible variation with Universal Time is the hourly average curve of electric potential gradient in the lowest atmosphere obtained over the world's oceans on the Carnegie cruises. These curves have long served as de facto standards (see Bering et al., 1998) with which to evaluate the viability of attempts to measure worldwide variations of electrical properties of the atmosphere near the Earth's surface.

The thunderstorm generator hypothesis proposed by Wilson (1920) was based on his observations that, beneath the thundercloud, negative charge is transferred to the Earth and above the thundercloud positive charge is transferred to the conducting upper atmosphere. A subsequent discovery was the close correlation between the diurnal (Universal Time) variation of the thunderstorm generator current (represented by the frequency of thunderstorm occurrence) and the load current (represented by the fair-weather ground electric field, or air-Earth current density), integrated over the surface of the Earth.

---

[*] Corresponding author. Tel.: 44-1223-353839; fax: 44-1223-303839.
E-mail address: michael.j.rycroft@ukgateway.net (M.J. Rycroft).

1364-6826/00/$ - see front matter © 2000 Elsevier Science Ltd. All rights reserved.
PII: S1364-6826(00)00112-7

By making measurements in regions where local diurnal variations of atmospheric conductivity are minor, by excluding data collected during periods of local meteorological influence, and by averaging to remove the day-to-day variability of global thunderstorm activity, a diurnal curve with a peak near 18 UT, a minimum near 03 UT and a range of some 30% of the mean was obtained. This is known as the global, 'fair-weather', diurnal variation, or the 'Carnegie curve'. Bering (1995) gives a comprehensive interpretation of recent advances in this area.

Humans are becoming increasingly concerned about changes — both natural and anthropogenic — in the Earth's environment. Particular attention is being paid to the atmosphere, not only the troposphere, the region of the Earth's weather and climate, but also the stratosphere — above it, which is where the ozone layer is. The springtime Antarctic ozone depletion, the so-called ozone hole, is the most popular example (see Farman et al., 1985; Rycroft, 1990; Solomon, 1999). Changes higher up, in the mesosphere, thermosphere and ionosphere may also be important.

How does atmospheric electricity affect man and his technological systems? Is our electrical environment changing significantly as a result of air pollution, the release of radioactive materials, the construction of high-voltage power lines, and other activities, or by energetic charged particle effects in the atmosphere? It is clear that modern technological advances may be seriously affected by various electrical processes in the atmosphere or in space, and also that man is beginning to affect the electrical environment in which he resides.

2. Physics of the global atmospheric circuit

Atmospheric electricity plays various roles in the highly coupled system representing the Earth's atmosphere and the near-Earth space environment. For further background information, see Herman and Goldberg (1978).

The conductivity of the fair-weather atmosphere near the surface is on the order of 10^{-14} mho/m, and it increases nearly exponentially with altitude up to 60 km, with a scale height of about 7 km. The main charge carriers below about 60 km are small positive and negative ions that are produced primarily by galactic cosmic rays. Above 60 km, free electrons become more important as charge carriers and their high mobility abruptly increases the conductivity throughout the mesosphere. Above 80 km, the conductivity becomes anisotropic because of the influence of the geomagnetic field, and there are diurnal variations due to solar photoionization processes. The arena for the subject is included in the system shown in Fig. 1, which shows the Earth at the centre, surrounded by the atmosphere, ionosphere, the van Allen belts and magnetosphere deformed by the flowing solar wind.

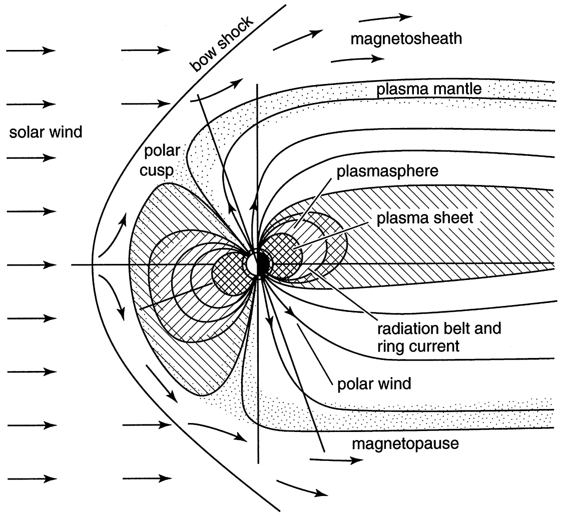

Fig. 1. Diagram showing the Earth at the centre, surrounded by the atmosphere, ionosphere, the van Allen radiation belts and magnetosphere deformed by the flowing solar wind (from Davies, 1990).

The area for the subject of atmospheric electricity is the electrically insulating concentric shell between the highly conducting Earth and the highly conducting ionosphere (see Fig. 2); this is surrounded by "magnetospheric action". Fig. 3, from Markson (1978), shows that the global electric circuit is believed to be driven by the upward current from thunderstorms, 1000 of them occurring around the globe at any one time. They occur mostly at low latitudes, over tropical land masses, in the local afternoon and evening. The more energetic thunderstorms reach higher into the atmosphere.

It has been known for over two centuries that the solid and liquid Earth and its atmosphere are almost permanently electrified. The surface has a net negative charge, and there is an equal and opposite positive charge distributed throughout the atmosphere above the surface. The fair-weather electric field is typically from 100 to 300 V/m at the surface; there are diurnal, seasonal, and other time variations in this field that are caused by many factors. The atmosphere has a finite conductivity that increases with altitude; this conductivity is maintained primarily by galactic cosmic ray ionization. Near the Earth's surface, the conductivity is large enough to dissipate any field in just 5–40 min (depending on the amount of pollution); therefore, the local electric field must be maintained by some almost continuous current source.

According to the classical picture of atmospheric electricity, the totality of thunderstorms acting together at any time charges the ionosphere to a potential of several hundred thousand Volts with respect to the Earth's surface. This potential difference drives a vertical electric conduction current downward from the ionosphere to the ground in all fair-weather regions of the globe. The fair weather electric conduction current varies according to the ionospheric potential difference and the columnar resitance between the ionosphere and the ground. Horizontal currents flow freely along the highly conducting Earth's surface and in the

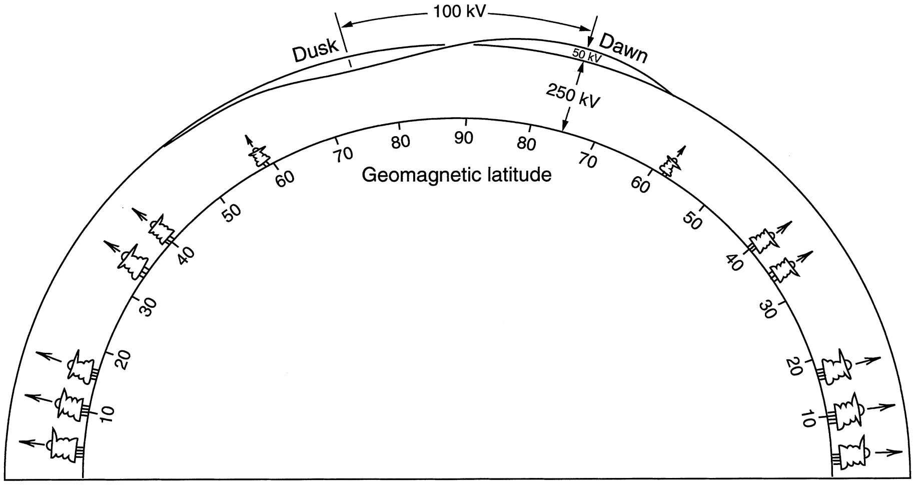

Fig. 2. Diagram showing the upward current to the ionosphere generated by thunderstorms; the ionosphere is approximately an equipotential surface, at +250 kV with respect to the Earth's surface (from Markson, 1983).

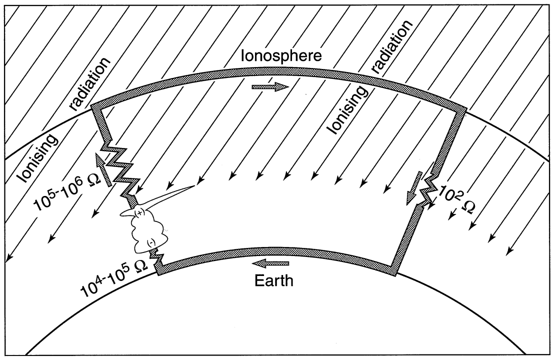

Fig. 3. Diagram of the global electric circuit (bold); the current flowing in the circuit is determined mainly by the current in the generator representing the sum of thunderstorms around the world, in the charging resistor above, and in the resistor representing more than the boundary layer below. Also important are increases in the ionising radiation which, from time to time, reduce the conductivity of the middle atmosphere (from Markson, 1978).

ionosphere. A current flows upward from a thunderstorm cloud top toward the ionosphere and also from the ground into the thunderstorm generator, closing the circuit. The global fair-weather load resistance is of the order of 100 Ω, as indicated in Fig. 3.

A few measurements have been made that give the magnitude of the current flowing upward over the whole area of a thundercloud (Kasemir, 1979). The currents range from 0.1 up to 6 A, with an average between about 0.5 and 1 A per thunderstorm cell (Blakeslee et al., 1989).

The topic is not as simple as it seems initially — there are many variations. It is a problem of global scale, but with spatial variations of important parameters, on the regional scale (e.g. in the geomagnetic latitude, or geographic longitude), on the local scale, and also with altitude. At higher geomagnetic latitudes, the conductivity of the air, σ, is larger due to the larger flux of ionising radiation there due to cosmic rays ($\sim > 1$ GeV ions) or to relativistic charged particle precipitation from the magnetosphere (> 1 GeV ions or > 1 MeV electrons). The cosmic ray fluxes are somewhat reduced at the time of a Forbush decrease and, for certain geomagnetic storms, the fluxes of magnetospheric origin are enhanced.

Fig. 3 illustrates one thunderstorm representing all 10^3 active thunderstorms occurring over less than 1% of the Earth's surface, and each generating $\sim$1 A. Thus, an upward current of $\sim$1 kA is driven through the "charging resistor", of 10^5–10^6 Ω. This current flows through the almost perfectly conducting ionosphere. The return current of 1 kA flows through the fair weather atmosphere, over 99% of Earth's surface remote from thunderstorms; the "discharging resistor", or load resistor, has the value $\sim 2 \times 10^2$ Ω globally.

It is generally convenient to regard the highly (but not perfectly) conducting ionosphere as an equipotential surface locally, at a potential of about +250 kV with respect to the Earth. However, the ionosphere is not an equipotential surface globally, because of

- the dawn-to-dusk potential difference (up to $\sim$100 kV) across the polar cap arising from the interaction of the solar wind with the magnetosphere (see Fig. 2), which causes the two-cell convection pattern of the polar F-region,
- potential differences (up to $\sim$100 kV) associated with auroral/magnetospheric processes,
- unipolar induction, due to the rotation of the Earth's magnetic field, so that the equator is at a potential of −91 kV with respect to the pole, in the centred dipole approximation, in an inertial frame of reference, and

- dynamo action in the ionosphere, such as the Sq process (up to ∼20 kV).

The theoretical foundations of the subject of atmospheric electricity are firm, being based on the laws of physics — the Navier–Stokes and continuity equations, and the laws of thermodynamics — which are required to understand and predict atmospheric motions. The atmospheric density decreases exponentially with increasing height, with a scale height, $H \sim 7$ km. The atmospheric conductivity σ increases with height due to the energy spectrum of the cosmic rays and the charged particles precipitating from the magnetosphere. By Ohm's law, which is valid for linear processes only,

$$J = \sigma E, \tag{1}$$

where J is the vertical electric current density and E is the vertical electric field. This leads us to expect that, in order to maintain current continuity, E is largest near the Earth's surface, where its value is, from observations, ∼130 V/m. It decreases upwards, exponentially. By Gauss' law, there is a space charge profile associated with this electric field.

Cho and Rycroft (1998) have presented a simple model profile for the atmospheric conductivity, ranging from near 10^{-13} mho/m near the surface to 10^{-7} mho/m at 80 km altitude in the lower ionosphere. Hale (1994) has presented more complex profiles, which show variations in both space and time.

Only a few mathematical models of global atmospheric electricity have appeared over the years (Kasemir, 1963, 1977; Hill, 1971; Hays and Roble, 1979; Volland, 1982; Ogawa, 1985). Since it is difficult to obtain global measurements to deduce the instantaneous properties of the global circuit, these models provide a convenient means of examining, through numerical experiments, the various interacting processes operating in the global circuit. The overall success of the models may be judged by how well they represent the observed atmospheric electrical properties at any place and time within the circuit.

Thunderstorms are extremely complex, and some simplifying assumptions must be made to represent their properties in a global model. The usual assumption is to consider thunderstorms as dipolar current sources, with a positive source in the could top and a negative source in the cloud bottom.

Ogawa (1985) has considered the simple equivalent circuit for the atmosphere shown in Fig. 4. With R_1 being the charging resistor mentioned earlier, R_2 the resistance of the thunderstorm generator where there exist potential differences ∼100 MV between the positively charged top of the thundercloud and the negatively charged bottom, and R_3 the resistance of the boundary layer (first few km) of the atmosphere, all of which are much greater than r, the resistance of the fair-weather atmosphere, it follows that

$$I = R_2 I_0/(R_1 + R_2 + R_3). \tag{2}$$

Eq. (2) relates the current through the fair-weather atmosphere, I, to the current in the thunderstorm generator, I_0.

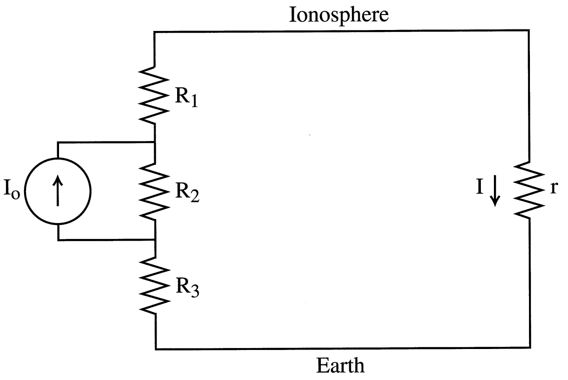

Fig. 4. A simplified equivalent circuit of the global electric circuit, showing the net thunderstorm generator current, I_0, and the fair weather current I (from Ogawa, 1985).

It is clear that R_1 and R_3 are particularly important in this relation; R_3 can be significantly reduced by

- point discharge currents due to the very large electric field below the thundercloud, in the region of updraughts, and
- increasing the height of the ground surface above mean sea level.

In this regard, measurements of I at a high-altitude (mountain) observatory remote from active thunderstorms can give some information on I_0 if R_1 and R_2 can be considered to be constant. Both may, however, be reduced at times of enhanced fluxes of energetic charged particles associated with enhanced geomagnetic activity. And the contribution to R_1 above an active, sprite-producing thundercloud may be greatly reduced due to the ionisation (see Cho and Rycroft, 1998) produced in the rarefied mesosphere by the large transient electric fields due to large, positive cloud-to-ground lightning discharges.

Fig. 5 is a new diagram, the upper part of which attempts to illustrate the electric currents flowing through different parts of the atmosphere, and including the magnetosphere above. The lower part of the diagram shows the equivalent circuit, with three different fair weather regions. One of these is for a high altitude part of the Earth where the profiles of J and E through the fair weather atmosphere will differ from those elsewhere.

From standard electrostatic theory, the capacitance C of the concentric shell of atmosphere between the Earth and the ionosphere, over one scale height of atmosphere rather than over the full height of the ionosphere, is given by

$$C = 4\pi\varepsilon_0 R_{\mathrm{E}}^2/H \sim 0.7 \text{ F}, \tag{3}$$

where R_{E} is the radius of the Earth and ε_0 is the permittivity of free space.

Hence the time constant for the atmospheric global electric circuit, t, is given by

$$\tau = Cr \sim 2 \text{ min}. \tag{4}$$

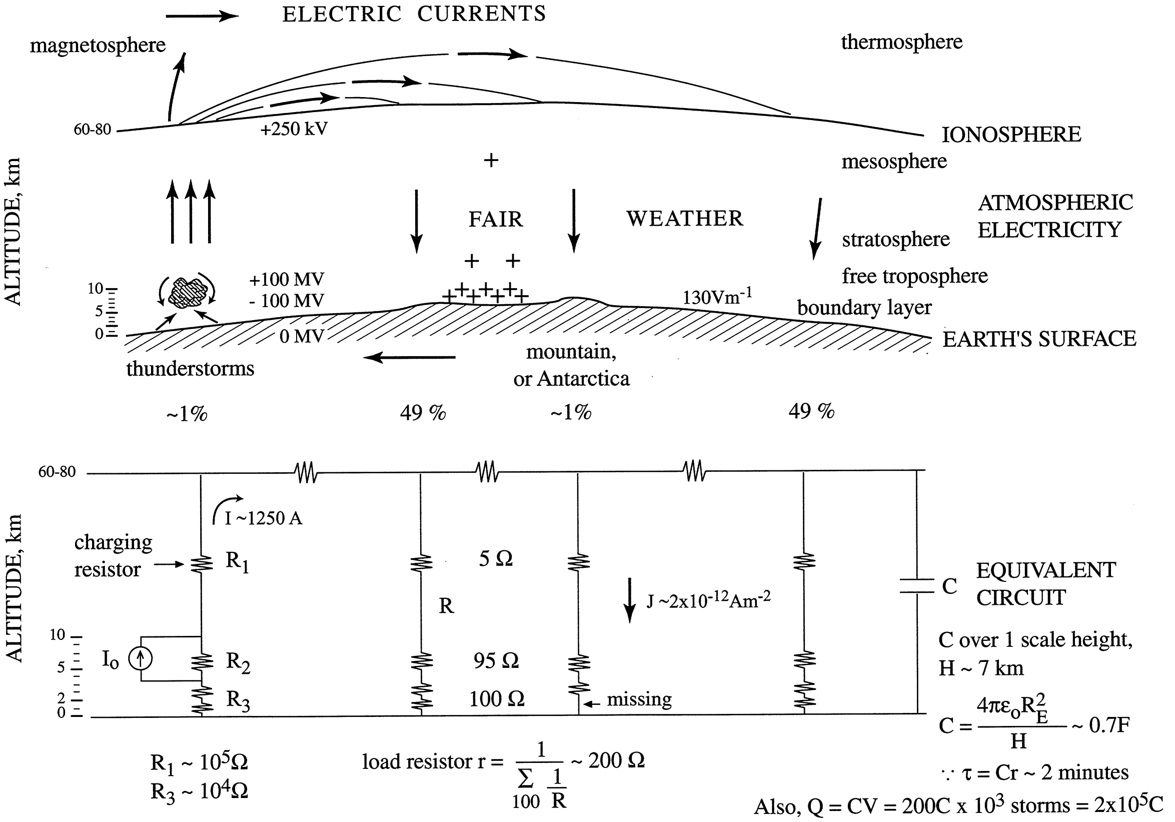
ELECTRIC CURRENTS
magnetosphere
thermosphere
60-80
+250 kV
IONOSPHERE
mesosphere
+
ALTITUDE, km
FAIR
WEATHER
ATMOSPHERIC ELECTRICITY
stratosphere
+ +
10
5
0
+100 MV
- 100 MV
+++++++
130Vm^{-1}
free troposphere
boundary layer
0 MV
EARTH'S SURFACE
thunderstorms
mountain, or Antarctica
~1%
49 %
~1%
49 %
60-80
I ~1250 A
charging resistor
R_1
5 Ω
R
$J \sim 2\times10^{-12}\,\mathrm{Am}^{-2}$
EQUIVALENT CIRCUIT
C
10
5
2
0
I_0
R_2
95 Ω
R_3
100 Ω
missing
C over 1 scale height, H ~ 7 km
$C = \frac{4\pi\varepsilon_0 R_E^2}{H} \sim 0.7\mathrm{F}$
$R_1 \sim 10^5\Omega$
$R_3 \sim 10^4\Omega$
load resistor $r = \frac{1}{\Sigma\frac{1}{100}\frac{1}{R}} \sim 200\,\Omega$
$\because \tau = Cr \sim 2$ minutes
Also, $Q = CV = 200\mathrm{C} \times 10^3$ storms $= 2\times10^5\mathrm{C}$

Fig. 5. (a) Top. Diagram, approximately to scale, showing electric currents (bold arrows) flowing up from the net thunderstorm generator over $\sim<1\%$ of the Earth's surface, through the ionosphere, the fair-weather atmosphere and the Earth, and closing as point discharge currents below the thundercloud. In the centre, the distribution of positive space charge is shown. (b) Bottom. The equivalent circuit for (a), showing components of the fair-weather resistor over different height ranges for 2–3 km above the surface; the boundary layer resistor is "missing" above a mountain or over Antarctica. Typical numerical values are also shown.

With a charge of 200 C associated with each thunderstorm, the total Q on the plates of the spherical condenser is 2×10^5 C. Thus, the energy associated with the global electric circuit is enormous; it is

$$W = CV^2/2 \sim 2\times10^{10}\ \mathrm{J}, \tag{5}$$

inserting $V = 250$ kV.

The electric current density through the fair-weather atmosphere, J, has the value $\sim 2\times10^{-12}$ A/m^2. Taking the conductivity of air at ground level, produced by extremely energetic cosmic rays and by radon from the ground, and due to aerosols, to be $\sim 2\times10^{-14}$ mho/m, the fair weather electric field at ground level is $\sim 10^2$ V/m, close to the near observed value of 130 V/m. At 20 km altitude, the fair-weather vertical electric field is ~ 1 V/m, and can be well measured from balloons. At 50 km altitude, it is only $\sim 10^{-2}$ V/m.

Following a Forbush decrease, the atmospheric conductivity everywhere could be reduced by $\sim 10\%$. If J is unchanged, the ionospheric potential, and hence also the fair weather electric field, may be increased by $\sim 10\%$. There is some observational evidence to support this concept (see Ogawa, 1985).

After a sprite above one of the thousand active thunderstorms, the ionospheric potential would reduce to 99.9% of its initial value for a millisecond or so. Thus, it is evident that sprites are unlikely to cause an observable effect on the fair-weather electric field.

To consider electrodynamic or electromagnetic effects rather than electrostatic phenomena, Maxwell's equations are involved. They relate electromagnetic fields to charge and current densities in a time-varying situation. Electromagnetic waves of appropriate frequencies can be generated, and these propagate away. The medium through which the energy propagates has a certain refractive index, which is equal to the square root of the relative electric permittivity.

The medium is treated as a conductor at frequencies (ω) such that $\omega \ll \sigma/\varepsilon_0$. Alternatively, the medium behaves as a leaky dielectric when $\omega \gg \sigma/\varepsilon_0$; that is the situation in the ionosphere for VLF waves and for higher frequency waves. However, for waves at Schumann resonance frequencies the lower ionosphere behaves like a conductor. Radiation produced by the current pulse due to a cloud-to-ground lightning discharge (i.e. due to the rapid destruction of the dipole moment formed by the negative charge in the bottom of a thundercloud and its positively charged image in the Earth) ranges from Schumann resonance frequencies up to ~100 MHz.

However, if there are no radiation effects to be considered, the current flowing is the sum of the conduction current and the displacement current. This is termed the Maxwell current. In the atmosphere, there can be other significant contributions to the current density, due to convection, turbulence, precipitation (or settling) or point discharges. Mentioned earlier, and sometimes called St. Elmo's fire, this last contributor to the current through the global atmospheric circuit may actually carry half the current (see Bering et al., 1998). By its very nature, it is variable, and not a great deal is known about it. It is a topic for further research.

A fundamental property of the global atmospheric electrical circuit is the electrical relaxation time at various altitudes, which is defined as the time the electric current takes to adjust to $1/e$ of its final value after an electric field is suddenly applied, assuming that the conductivity remains constant. At high altitudes, near 70 km, the relaxation time is about 10^{-4} s, increasing with decreasing altitude to about 4 s near 18 km and to about 5–40 min near the Earth's surface. The electrical relaxation time of the land surface of the Earth is about 10^{-5} s. The maximum value of about 40 min in the atmosphere near the Earth's surface is the characteristic time that the global circuit would take to discharge if all thunderstorm activity suddenly ceased. Measurements have never shown a complete absence of a fair-weather electric field for any length of time, thereby suggesting a continuous operation of thunderstorms and other generators in maintaining the currents flowing in the global circuit. For time variations longer than about 40 min a quasi-static approximation can be applied when modelling the electrical properties of the global circuit.

3. Solar and cosmic influences on atmospheric electricity

3.1. Solar cycle effects

Variable plasma phenomena from the Sun, such as gigantic coronal mass ejections, can create geomagnetic storms, periods of enhanced geomagnetic activity. Such phenomena exert a profound — often disruptive, and occasionally devastating — influence on humankind's technological systems, such as satellites, telecommunications systems, power lines and pipelines; they may even affect human health.

It is well known that the Sun is not a constant star but one whose radiative output (especially in the ultraviolet and X-ray spectral regions) varies over a solar cycle, of ~11 years period. (Taking account of the Sun's magnetic polarity, it is actually a ~22 year cycle.) The temperature, and hence the density, of the upper atmosphere (thermosphere) vary markedly over a solar cycle. For example, at ~400 km altitude where the International Space Station will operate, the density — and hence the drag force — is almost a hundred times greater at solar maximum than at solar minimum.

The solar magnetic field and its extension through the interplanetary medium (carried by the plasma of the solar wind) to the Earth's orbit and beyond, into the heliosphere, is markedly more variable at solar maximum than at solar minimum. Thus, the flux of galactic cosmic rays (scattered by the irregularities of the interplanetary magnetic field) entering the top of the atmosphere is some tens of percent larger at solar minimum that at solar maximum, whereas the flux of energetic (>0.1 GeV) charged particles (protons) from the Sun at solar minimum is typically many times less that its value at solar maximum. Both types of energetic charged particles and also energetic charged particles precipitating from the inner and outer magnetospheric radiation belts (see Fig. 1, from Davies, 1990) during geomagnetically active periods interact with the Earth's middle and lower atmosphere by depositing their energy in the atmosphere, by creating ionisation directly or via bremmstrahlung radiation, by altering its chemistry (particularly ozone, nitric oxide and sulphate aerosols, see Jackman et al., 1995), or by affecting the nucleation of water droplets to form clouds (see Tinsley, 1996, 2000). There is some recent evidence that energetic solar protons appear to produce small holes in the Arctic ozone layer. They may trigger the formation of cirrus clouds which decrease the flux of solar radiation to the Earth's surface and hence affect the evolution of weather systems (see Pudovkin and Babushkina, 1992), and also affect the Earth's radiation budget by contributing to the greenhouse effect (also see Svensmark and Friis-Christensen, 1997),

Ground-based radars and other instruments in Scandinavia and around the world observe the effects of the ever changing, million Ampere current systems in the auroral ionosphere at heights near 110 km. These cause changing geomagnetic fields which induce large and varying currents to flow in long conductors, such as electrical power lines or oil pipe lines, sometimes with catastrophic effects. These include exploding transformers and power cuts (outages) of several hours duration, or corrosion.

Thus atmospheric electricity plays various roles in the highly coupled system representing the Earth's atmosphere and the near-Earth space environment. There are temporal variations, on time scales varying from microseconds (in lightning discharges), to milliseconds (e.g. sprites), minutes to an hour (e.g. thunderstorm regeneration), hours to a day (e.g. diurnal variations), months (solar rotation and seasonal effects), and to a decade (solar cycle effects), as well as spatial variations.

3.2. Latitudinal effects

Solar influences on the geoelectric field differ between mid-latitudes and the polar regions (Tinsley, 1996).

Increased solar activity reduces galactic cosmic ray fluxes in mid-latitudes, thus reducing the atmospheric conductivity in this region, while at the same time solar protons may be "funnelled" by the Earth's magnetic field to polar regions resulting in an increased atmospheric conductivity there. A dawn-to-dusk potential difference is also applied across the polar regions as a result of the interaction of the solar wind and the Earth's magnetic field (see, for example, Tinsley and Heelis, 1993). The cross-polar potential results from the same interactions that generate auroral activity and typical lies in the range 30–60 kV (Fig. 2). It may, at times of increased solar activity, exceed 100 kV. The cross-polar potential may not be the dominant solar influence on the global geoelectric circuit, but it is of significant magnitude in the polar regions. The diurnal pattern generated as a station rotates beneath the cross-polar cap potential difference also aids its detection.

The global electric circuit is closed by currents flowing in the ionosphere. At high latitudes, within the polar cap, these currents are strongly influenced by parameters of the solar wind, particularly the direction of the interplanetary magnetic field. In the auroral zone, the ionosphere is markedly disturbed when the magnetosphere is buffeted by a cloud of solar wind of enhanced density which is associated with a huge coronal mass ejection event from the Sun. As a result, the potential difference across the polar cap ionosphere can vary markedly.

The measurement techniques applied in magnetospheric and ionospheric research enable the detection of strong, intrinsic effects of the solar wind/ionosphere on the electrical potential distribution and conductivity of the atmosphere, as presented by Michnowski (1998). These manifestations of the solar wind interaction with the magnetosphere and ionosphere are especially evident at high latitudes.

4. Aerosols and the global circuit

4.1. Planetary boundary layer (PBL) aerosols

The planetary boundary layer (PBL) is the lowest few kilometers of the atmosphere, where interactions with the surface, man, and the biosphere are the most pronounced. Near land surfaces the ionization is, beside galactic cosmic rays, produced by the decay of natural and artificial radioactivity. Electrical processes in the PBL are complex, highly variable, and span a tremendous range of space and time scales. The electrical variables respond to many of the lower atmospheric processes but usually have little influence on the phenomena to which they respond. Within the PBL, local turbulent fluctuations of space-charge density impose a time-varying electric field that is comparable in magnitude with, or even greater than, the electric field maintained by global thunderstorm activity. Since the PBL is the region of the atmosphere with the greatest resistance, it is this layer (as well as the thunderstorm generators) that control the currents in the global circuit. Boundary layer processes can have a substantial impact on the fields and currents appearing throughout the entire atmospheric column from the Earth to the ionosphere. Modelling of the atmospheric electricity in the PBL is difficult due to the complicated conditions in the PBL (see Hoppel et al., 1986).

Aerosol particles can change the Earth's radiation budget both directly through scattering and through absorption and indirectly through modification of cloud microphysical properties. It is clear that there is a great uncertainty in assessing the direct and indirect effect of aerosols on climate change. We must understand and be able to quantify the most important sources and characteristics of aerosols in order to assess their effects on climate change. The relatively short residence time of aerosols in the troposphere, of order of a few days, results in significant spatial and temporal variations in aerosol particle concentration, size and composition. The high variability of the atmospheric aerosol leads to one of the largest uncertainties of anthropogenic climate forcing. This variability is compounded by a lack of much information on the properties of the atmospheric aerosols prior to the late 1970's when commercial aerosol instrumentation became more widely available. Thus, it is difficult to assess the influence of anthropogenic aerosols on radiative forcing up to that time. However, a number of researchers have shown that atmospheric electricity parameters can act as a good indicator or surrogate.

Many factors affect the electrical conditions in the PBL. An important area of research is the study of atmospheric ions. According to their physical nature, atmospheric ions with mobilities above 0.5 $\mathrm{cm}^2\ \mathrm{V}^{-1}\ \mathrm{s}^{-1}$ are charged molecular clusters or cluster ions. Tammet et al. (1992), Horrak et al. (1998) and others have shown the importance of these charged particles in atmospheric electricity, particularly in relation to atmospheric conductivity and to space charge formation. Retalis and Retalis (1998) treated the effects of air pollution on the large ion concentration, which indicated the importance of polluted areas (like cities, etc.) on the electrical environment. The impact of Chernobyl in April 1986 on all atmospheric electrical parameters in some parts of Sweden has been significant directly after the accident and several months later (Israelsson and Knudsen, 1986). Martell (1985) and Israelsson et al. (1987) have shown the importance of the ionisation from natural and artificial radioactivity in the charging mechanism of clouds.

Even natural processes over the sea are of importance. Blanchard (1963) indicated that currents as large as hundreds of Amperes flow from the ocean surface into the atmosphere as a result of electrified droplets that are ejected from the bursting of small bubbles.

Modelling of convection currents has shown that they only become important in unstable mixed layers, where the turbulent transport time across the entire PBL can be comparable with the electrical relaxation time. Willett (1978) showed that, on average, the convection currents act as a local generator capable of reducing the total downward current density by as much as 44%.

In order to study the horizontal structure of vertical current densities in the PBL, experiments to examine the covariation of atmospheric vertical electric currents at different stations using long-wire antennas are discussed by Israelsson et al. (1994).

4.2. Volcanoes and aerosols above the PBL

The main source of ionisation in the mid-troposphere and stratosphere (above the PBL) is cosmic radiation. The ionisation rate depends on magnetic latitude and on solar activity. The composition of the ions that establish the bulk electrical properties is relatively unknown. The ion concentration is also affected by aerosols whose distributions are quite variable in both space and time.

The global electric circuit varies over the solar cycle associated with cosmic ray flux changes. Whilst the contributions of thunderstorms may not only be to maintain an upward DC current, an improved understanding is required of the return current (~ 2 pA/m^2) flowing through the "fair-weather" atmosphere as part of the global electric circuit, particularly for geographically restricted variations of atmospheric conductivity. Another type of perturbation to the atmosphere results from volcanoes. Their sulphate aerosols and dust particles could nucleate high-altitude clouds, and the sulphates change the atmospheric conductivity and hence the properties of the global electrical circuit.

The character of upper tropospheric aerosols can be temporarily disturbed not only by volcanic eruptions but also by forest fires, biomass burning, and large dust storms. In addition annual variations of their concentrations have also been observed (Hofmann et al., 1975). Another important temporal variation of tropospheric aerosols is associated with the so-called Arctic haze events. New evidence suggests that these events are characterized by high particle loading throughout a large portion of the troposphere. The impact of this extensive aerosol loading on atmospheric electrical parameters is yet to be determined.

Hogan and Mohnen (1979) reported the results of a global survey of aerosols in the troposphere and lower stratosphere. They found that the concentrations were more or less symmetrically distributed about the Earth. More measurements of this type could provide the basis for extrapolating local or isolated observations to characteristic worldwide values. Man-made global ionisation effects reported by Boeck (1983), who studied the effect of krypton 85 released into the atmosphere on the global conductivity, have to be included in the global electric circuit.

The morphology of stratospheric aerosols is dominated by a persistent structure frequently referred to as the 20-km sulphate layer, or Junge layer. It is now known that the character of this layer is highly affected by large volcanic eruptions (Gringel et al., 1986). These volcanic eruptions can result in a rather dramatic increase of the aerosol content up to the stratosphere. The extent to which these aerosols directly affect the atmospheric conductivity, small-ion concentration, and mobility should be investigated to a greater extent. On the other hand, the sulphuric acid content of the lower atmosphere is also drastically enhanced following volcanic eruptions and might influence the ion composition considerably. Air-Earth current density measurements seem to be consistent with the classical picture of the global circuit, with some exceptions. The question of whether there are other global generators in the lower atmosphere, in addition to thunderstorms, could probably be unravelled by ground-based and balloon-borne current measurements at different locations.

5. Thunderstorms and the global circuit

5.1. Spatial and temporal distribution of thunderstorms

In recent decades we have greatly advanced our knowledge regarding the spatial and temporal distribution of thunderstorms and lightning. In the 1960s the first satellite observations of lightning were obtained (Vorpahl et al., 1970). However, in the last five years tremendous advances have occurred with the launch of three separate lightning detectors in space (OTD, LIS and FORTE) (Christian and Latham, 1998; Jacobson et al., 1999). We now know that the vast majority of lightning activity is located over the continental regions of the globe, with little lightning ever being found over the oceans. The reason for this is still debated; however, it is related to the differences in the microphysics of oceanic and continental deep convection (Price and Rind, 1992). Observations have shown significant differences in the updraft intensities in storms, with the oceanic storms rarely having updraft intensities greater than 10 m/s. The intensity of the updrafts is very important for the electrification processes, due to the vertical transport of supercooled droplets and graupel particles (Williams et al., 1992).

As a result of the global distribution of lightning, there are three main centres of lightning activity over the three tropical continental landmasses (South America, Africa, and South East Asia/Maritime Continent). In addition to the tropical lightning, extratropical lightning activity plays a major role in the summer season in the northern hemisphere, resulting in the global lightning activity having a maximum in June–August. Since the daily lightning activity in the tropics maximizes late in the local afternoon (16–18 local time), the combination of spatial and temporal activity produces the well-defined diurnal cycle in measurements of the global electric circuit (the 'Carnegie curve').

Whipple (1929) showed the agreement between the diurnal cycle of the global circuit and the diurnal cycle of global thunderstorm activity. Markson (1986) showed that individual measurements of the ionospheric potential agree well with the 'Carnegie curve', confirming that lightning plays a major role in the global electric circuit. Clayton and Polk

(1977) also showed good agreement between the AC component of the global circuit (Schumann resonances, see below) and global lightning activity. All these studies showed that the global electric circuit has a maximum at approximately 1800 UT, and a minimum at 0300 UT (Price, 1993).

As mentioned above, global lightning activity has a seasonal cycle with a maximum in the northern hemisphere summer. This annual cycle is also observed in the various measures of the global electric circuit. Hence, monitoring the global electric circuit, either via the DC components (ionospheric potential, fair-weather current) or via the AC component (Schumann resonances) provides a good proxy for global lightning activity.

5.2. *Sprites*

In the last decade, a completely new discovery has been made in the field of atmospheric electricity. Optical flashes have been observed high above thunderstorms, reaching altitudes of up to 100 km. These luminous transient events lasting from 1 to 100 ms have been named elves, sprites and blue jets. In the last few years the research in this area has increased dramatically (see Sentman and Wescott, 1996; Rodger, 1999, and a Special Issue of JASTP (May/June 1998)). It is now believed that all three phenomena are independent of each other, although sometimes they may occur simultaneously. Most of the advances in this field have occurred as the result of dedicated field experiments in the United States. During the summer months, mesoscale convective systems develop over the United States, which provide a natural laboratory for studying sprites. Observations have been taken on the ground (Lyons, 1996), from aircraft (Sentman and Wescott, 1993), and from high-altitude balloons (Bering et al., 1999). It appears that elves occur with all lightning discharges. The electromagnetic pulse (emp) produced by the lightning propagates up into the stratosphere and mesosphere, producing large electric fields that result in a ring of light at the base of the ionosphere (Cho and Rycroft, 1998), similar to the airglow phenomenon. This ring expands outwards as the emp propagates away from the lightning source in the storm below. The elves last only a few milliseconds, and they can only be seen with very sensitive low-light-level cameras. The propagation of elves outward has been observed using high-speed photometers (Inan et al., 1997).

Sprites are the more spectacular events, and are caused by only a small fraction of the lightning discharges within the storm below. We now know that sprites occur simultaneously with positive cloud-to-ground lightning (Boccippio et al., 1995), although recently there has been some evidence of negative cloud-to-ground flashes producing sprites (Barrington-Leigh et al., 1999). Sprites are red in colour as a result of the excitation of the N_2^+ line. The lifetime of sprites is much longer than that of elves, up to 50 milliseconds, and they occupy volumes greater than 1000 km^3. The removal of charge from the cloud by the positive discharge is thought to result in the stressing of the upper atmosphere. The screening layer around the anvil of the cloud immediately after the ground discharge acts as the charged plate of a horizontal capacitor, producing large fields above the storm in the mesosphere. The enhanced field accelerates electrons in the low-density mesosphere, colliding with N_2 molecules, resulting in the emission of red light. The relaxation time of the screening layer is some tens of milliseconds, the time scale of the sprites.

Do sprites influence the global electric circuit? This is still an unanswered question. Although sprites occur in the upward branch of the global circuit above thunderstorms, hence influencing the conductivity of the upper atmosphere (the resistor R_1 in Fig. 5), they occur much less frequently than regular lightning. Observations indicate approximately 1 sprite for every 200 lightning discharges. Therefore, due to their low occurrence rate they may not pay a major role in the global electric circuit. This is a topic of future research.

5.3. *Schumann resonances*

A paper in this issue (Barr et al., 2000, this issue) describes in detail the history and development of Schumann resonances (SR). The SR are produced by electromagnetic radiation from lightning being trapped in the Earth-ionosphere waveguide. At extremely low frequencies (ELF), very little attenuation occurs, allowing these radio waves to propagate a few times around the globe before dissipating. Constructive interference produces standing resonant waves at $8, 14, 20, 26, \dots$ Hz, where the 8 Hz mode represents a wave with wavelength equal to the circumference of the Earth (40,000 km); see also Rycroft (1965) and Fuellekrug and Fraser-Smith (1996). Hence, similar to the DC fair-weather global circuit measurements, the SR can also be monitored from a single position on the Earth's surface.

Research on SR is twofold. The background SR signal intensities are well correlated with global lightning activity (Heckman et al., 1998); hence, the SR are used for studying the global variability of lightning activity (Satori, 1996; Nickolaenko et al., 1998). On the other hand, the transient SR signals have been shown to be well correlated with sprite activity (Boccippio et al., 1995; Cummer et al., 1998). Therefore, we are also able to use the SR measurements to study sprites on a global scale (see Fig. 6).

Unlike the DC circuit that is influenced by changes in the conductivity of the atmosphere, the SR are dependent only on global lightning activity. Hence, the AC component of the global circuit is better correlated with global lightning activity than is the DC component.

6. Climate change and the global electric circuit

In recent years a new and important application of the global electric circuit has become evident. The global

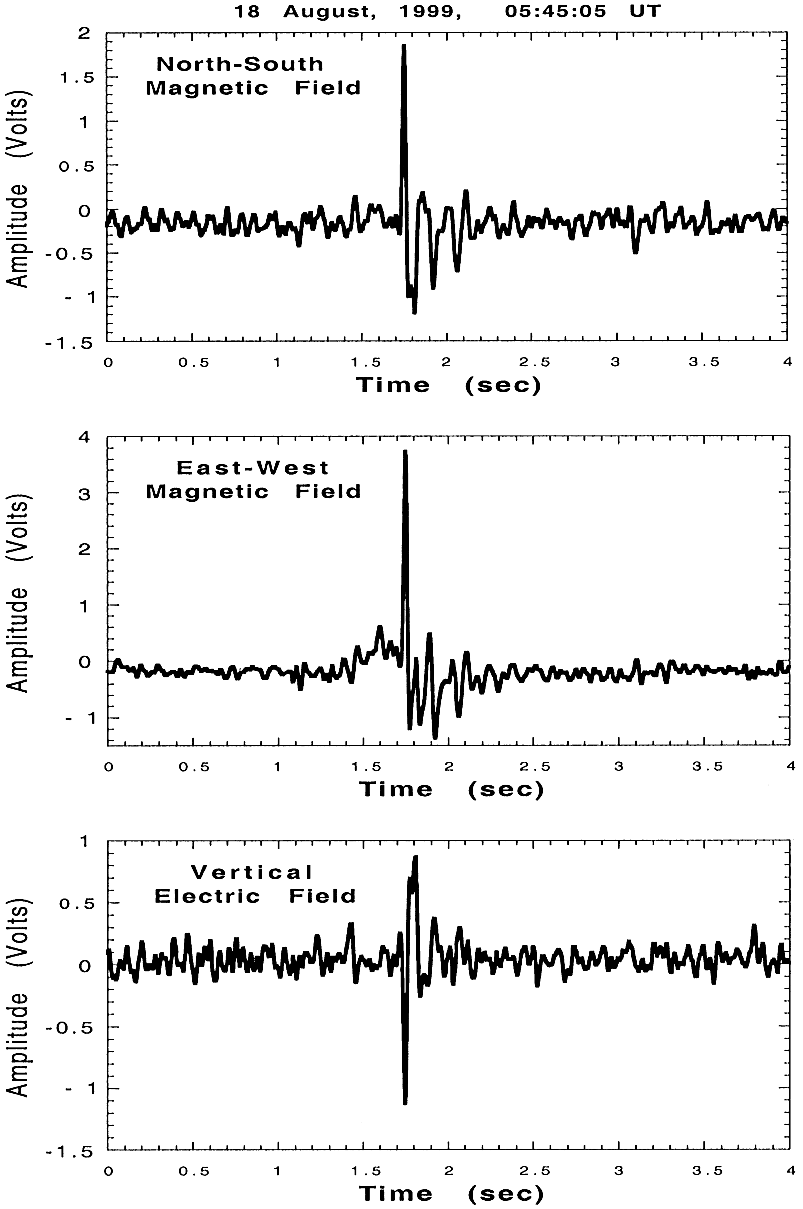

Fig. 6. The ELF radio signals observed in Israel associated with a sprite generated in the United States.

circuit is becoming an important tool in studying the Earth's climate, and climate change. Price and Rind (1990) first proposed the possibility that changes in the Earth's climate could result in more lightning activity around the planet. Since then global lightning activity and the global electric circuit have been shown to be closely related to various important climate parameters.

Williams (1992) showed the close relationship between tropical surface temperatures and the monthly variability of the SR measured in Rhode Island, USA. A similar study was conducted by Fuellekrug and Fraser-Smith (1997), showing a connection between ELF observations in Antarctica and Greenland and global surface temperatures. Price (1993) showed a good agreement between the diurnal surface temperature changes and the diurnal variability of the global electric circuit ('Carnegie curve'). This study also showed a strong link between the frequency/intensity of global deep convection and global surface temperatures. Markson and Price (1999) used direct measurements of the ionospheric potential to further demonstrate the positive correlations between global and tropical surface temperatures and the ionospheric potential. Recently, Reeve and Toumi (1999) used new satellite lightning observations to show the agreement between globally observed lightning activity and global temperatures. Finally, a recent study by Price (2000) has extended the above connections to additional climatic parameters such as upper tropospheric water vapour; the variability of upper tropospheric water vapour is closely linked to the variability of global lightning activity, measured via the SR.

Therefore, by monitoring the global electric circuit, we may be able to study, in a cheap and consistent way, the variability of global lightning activity, which is closely related to surface temperatures, tropical deep convection, rainfall, upper tropospheric water vapour, and other important parameters that affect the global climate system. The advantage of these measurements is that they can be made continuously, for many years, unlike satellite sensors that generally have a relatively short lifetime of only a few years, or even less.

7. Potential applications and recommended atmospheric electricity research

An increased interest in understanding the Earth's electrical environment has resulted from recent advances in different disciplines, along with the recognition that many of man's modern technological systems can be adversely affected by this environment. The cornerstone of our understanding of the Earth's electrical environment and the global electrical circuit is an integration of measurements, theory, and modelling. There should be a concerted effort of coordinated measurement compaigns, supported by critical laboratory experiments, theory, and numerical modelling of processes, to improve our understanding of the Earth's electrical environment.

It is also an opportunity to further investigate relationships/correlations between atmospheric aerosols and electrical parameters, with the advent of more widely available aerosol measurement capabilities. A proper understanding of the impact of aerosols on climate rests both on quality-assessed measurements, and on a predictive modelling capability which enhances our ability to understand complex and nonlinear interactions between aerosols and atmospheric electric parameters. An interesting and novel approach in linking atmospheric electricity parameters and lightning activity to the global electric circuit and to global warming provides the basis for other crucial modelling initiatives.

The global circuit of atmospheric electricity is influenced by meteorological processes. These influences are incorporated in the current that flows in the global circuit. As the electric conductivity of the troposphere and the stratosphere is so much lower than that of the ground and that of the ionosphere, the potential difference between these two highly conducting layers indicates the strength of the global circuit current and its variations with time. In measuring the electric field or the air-Earth current density at any one station, we may obtain information about those meteorological processes that, world-wide, influence the global circuit current. At the same time, the data obtained by measuring the electric field and/or the air-Earth current density are also influenced by local events at and above the measuring stations. These events may be generators for local electric circuits or may cause variations of resistances in the circuit. Therefore, we must basically consider that in the data thus obtained we are witnessing influences both from processes that occur over the entire globe at the same time, as well as from purely local events. If the scientists in atmospheric electricity had the chance to separate these contributions, we might have a tool of significance for global change problems.

The ionospheric potential, electric field and air-Earth current density measurements in the PBL with periods shorter than a few hours are usually attributed entirely to local sources, primarily turbulence and pollution. Recent advances that have been made in independent research areas examine the interrelations between them, and project how new knowledge could be applied for the benefit of mankind. They also indicate needs for new research and for the types of coordinated efforts that will provide significant new advances in basic understanding and in applications over the next few decades. They emphasize a need to consider the interactions between various atmospheric, ionospheric, and telluric current systems to achieve an overall understanding of global electrical phenomena.

Evaluating the present situation from this vantage point, we can perceive three main approaches from atmospheric electricity to the global change monitoring effort:

1. Measure local parameters at many places and attempt meaningful integrations for the results over areas of global importance, optimally for the whole globe.
2. Measure local parameters and follow local processes at locations where they may be representative for larger areas; the expectation is that this will be the case for the oceans far from their shores.
3. Find global contributions to local measurements, and derive global information from a rather small number of local stations.

The global detection of lightning is necessary to determine the global flashing rate and how this rate relates to other parameters in the global circuit. NASA has developed new optimal sensors that could be used to detect and locate lightning in the daytime or at night, and with continuous coverage by using satellites in geosynchronous orbits. These sensors are capable of measuring the spatial and temporal distribution of lightning over extended periods with good spatial resolution; they offer significant new opportunities for research and for many applications.

There are ample reasons for interest in global-scale atmospheric electrical phenomena. For example, valuable information about the distribution and temporal variability of horizontal potential differences in the ionosphere could be provided by monitoring the ionospheric potential simultaneously in different locations. Furthermore, the widely accepted relationship between global thunderstorm activity and ionospheric potential has yet to be verified on anything but the crudest statistical basis. From the present perspective, finally, a detailed knowledge of the forcing from the global circuit would be useful in evaluating the electrical response of the PBL.

Seen from our present state of knowledge and our present priorities, the following atmospheric electricity elements should be investigated. Are they changing over time? If so, how and why?

- Thunderstorm nature, frequency, and geographic distribution; assess secular changes in the number of thunderstorms and/or of lightning.
- Ionosphere-Earth potential difference; assess its changes by direct or indirect measurements of air-Earth current density at the ground or in the air.
- Atmospheric electrical conductivity and its changes at a selection of typical stations, with various local atmospheric and geographic conditions and of types of anthropogenic influences.

In a recently published book, the Board of Atmospheric Sciences and Climate of the US National Research Council (1998) has presented a few Recommended Atmospheric Electricity Research Topics. These are to

- Investigate the possibility that the global electrical circuit and global and regional lightning frequency might be an indicator of climate change.
- Determine mechanisms responsible for charge generation and separation in clouds to understand cloud formation mechanisms and elucidate the fundamental physics of lightning.
- Determine the nature and sources of middle-atmospheric discharges to (1) increase knowledge of these recently discovered phenomena and their possible association with severe weather, and (2) explore their effects on radio propagation and atmospheric chemistry.
- Quantify the production of oxides of nitrogen (NO_x) by lightning to better understand upper-troposphere ozone production or loss.

The European Science Foundation has recently established a Network on "Space weather and the Earth's weather electrodynamic and charged particle effects on the stratosphere and troposphere". The overall aim of that interdisciplinary study, now named Space Processes and Electrical Changes Influencing Atmospheric Layers (SPECIAL), is to

- review and investigate such subjects,
- generate and test some hypotheses on the effects of energetic charged particles on the atmosphere, and
- consider their effects on weather and climate (also see Feynman and Ruzmaikin, 1999).

Thus, it is evident that the "old" subject of atmospheric electricity is very much "alive and kicking" today. Indeed, in the next decade or so, it may experience a renaissance.

8. Conclusions and discussions

The study of atmospheric electricity and the global electric circuit has advanced dramatically in the past 50 years. Our knowledge of the global electric circuit modulated by solar effects has improved, although there are generally more questions than answers in this field of atmospheric electricity. Regarding aerosols and atmospheric ions, advances have been made in studying air pollution using atmospheric electricity, and we have a better understanding of how the conductivity of the atmosphere is influenced by aerosols, and how this impacts our measurements of the fair-weather global circuit.

Perhaps the largest advances have been made in the areas of lightning and thunderstorm research, as related to the global circuit. We now have satellites looking down on the Earth continuously, supplying information on the temporal and spatial variability of lightning and thunderstorms. New discoveries have been made in the field of sprites, elves and blue jets, which may have a direct impact on the global circuit and how we model the global circuit.

Finally, the global atmospheric electric circuit is now starting to be recognised by some climate researchers as a useful tool to study and monitor the Earth's changing climate. The global electric circuit is the only climate-related parameter that can be measured at a single location on the Earth's surface, and yet supply global information. This unique quality should encourage additional research in this field.

Although there have been dramatic advances in the field of sprites, lightning, and even the Schumann resonances over the last decade, there is a need for an increased understanding of the solar-terrestrial connection. How do changes in solar particles and cosmic radiation, and also the solar wind, influence the Earth's climate? How strong is this connection? Is it mainly a high-latitude connection, or can tropical thunderstorms be affected by changes in the solar wind flux? What about variations over the solar cycle? These and many more questions need to be addressed in future research.

Our need to assess environmental impacts on human kind's technological systems requires a better understanding of electrical processes in the Earth's atmosphere than we now possess. Further research is needed to understand better the natural electrical environment and its variability and to predict its future evolution.

References

Barr, R., Jones, D.L., Rodger, C.J., 2000. ELF and VLF Radio Waves. Journal of Atmospheric and Solar Terrestrial Physics 62, 1689–1718.

Barrington-Leigh, C.P., Inan, U.S., Stanley, M., Cummer, S.A., 1999. Sprites triggered by negative lightning discharges. Geophysical Research Letters 26, 3605–3608.

Bering III, E.A., 1995. The global circuit: global thermometer, weather by-product or climate modulator?. Reviews of Geophysics 33 (Supplement to Vol. 33, Part 2), 845–862.

Bering, E.A., Few, A.A., Benkrook, J.R., 1998. The global electric circuit. Physics Today 51, 24–29.

Bering, E.A., Benbrook, J.R., Garrett, J.A., Paredes, A., Wescott, E.M., Sentman, D.D., Stenbaek-Nielsen, H.C., Lyons, W.A., 1999. The 1999 Sprites Balloon Campaign, EOS Transactions of American Geophysical Union Fall Meeting, p. F216.

Blakeslee, R.J., Christian, H.J., Vonnegut, B., 1989. Electrical measurements over thunderstorms. Journal of Geophysical Research 94, 13,135–13,140.

Blanchard, D.C., 1963. Electrification of the atmosphere by particles from bubbles in the sea. Progress in Oceanography 1, 71–202.

Boccippio, D.J., Williams, E.R., Heckman, S.J., Lyons, W.A., Baker, I.T., Boldi, R., 1995. Sprites, ELF transients and positive ground strokes. Science 269, 1088–1091.

Boeck, W.L., 1983. Krypton-85 in the atmosphere. In: Ruhnke, L., Latham, J. (Eds.), Proceedings in Atmospheric Electricity. Deepak, Hampton, pp. 73–75.

Cho, M., Rycroft, M.J., 1998. Computer simulation of the electric field structure and optical emission from cloud-top to the ionosphere. Journal of Atmospheric and Solar Terrestrial Physics 60, 871–888.

Christian, H.J., Latham, J., 1998. Satellite measurements of global lightning. Quarterly Journal of the Royal Meteorological Society 124, 1771–1774.

Clayton, M., Polk, C., 1977. Diurnal variation and absolute intensity of world-wide lightning activity, September 1970 to May 1971. In: Dolezalek, H., Reiter, R. (Eds.), Electrical Processes in Atmospheres. Steinkopff, Darmstadt, pp. 440–449.

Cummer, S.A., Inan, U.S., Bell, T.F., Barrington-Leigh, C.P., 1998. ELF radiation produced by electrical currents in sprites. Geophysical Research Letters 25, 1281–1285.

Davies, K., 1990. Ionospheric Radio. Peter Peregrinus, London.

Farman, J.C., Gardiner, B.G., Shanklin, J.D., 1985. Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction. Nature 315, 207–210.

Feynman, J., Ruzmaikin, A., 1999. Modulation of cosmic ray precipitation related to climate. Geophysical Research Letters 26, 2057–2060.

Fuellekrug, M., Fraser-Smith, A.C., 1996. Further evidence for a global correlation of the Earth-ionosphere cavity resonances. Geophysical Research Letters 23, 2773–2776.

Fuellekrug, M., Fraser-Smith, A.C., 1997. Global lightning and climate variability inferred from ELF magnetic field variations. Geophysical Research Letters 24, 2411–2414.

Gringel, W., Rosen, J.M., Hofmann, D.J., 1986. Electrical structure from 0 to 30 km. In: The Earth's Electrical Environment, Studies in Geophysics. National Academy Press, USA, pp. 166–182.

Hays, P.B., Roble, R.G., 1979. A quasi-static model of global atmospheric electricity 1. The lower atmosphere. Journal of Geophysical Research 84, 3291–3305.

Hale, L.C., 1994. The coupling of ELF/ULF energy from lightning and MeV particles to the middle atmosphere, ionosphere and global circuit. Journal of Geophysical Research 99, 21,089–21,096.

Heckman, S.J., Williams, E.R., Boldi, B., 1998. Total global lightning inferred from Schumann resonance measurements. Journal of Geophysical Research 103, 31,775–31,779.

Herman, J.R., Goldberg, R.A., 1978. Sun, weather and climate, NASA Special Publication, SP-426, pp. 360.

Hill, R.D., 1971. Spherical capacitor hypothesis of the Earth's electric field. Pure and Applied Geophysics 84, 67–75.

Hofmann, D.J., Rosen, J.M., Pepin, T.J., Pinnick, R.G., 1975. Stratospheric aerosol measurements 1: time variations of northern mid-latitudes. Journal of Atmospheric Science 23, 1446–1456.

Hogan, A.W., Mohnen, V.A., 1979. On the global distributions of aerosols. Science 205, 1373–1375.

Hoppel, W., Anderson, R.V., Willett, J.C., 1986. Atmospheric electricity in the planetary boundary layer. In: The Earth's Electrical Environment, Studies in Geophysics. National Academy Press, USA, 149–165.

Horrak, U., Salm, J., Tammet, H., 1998. Burst of intermediate ions in atmospheric air. Journal of Geophysical Research 103, 13,909–13,915.

Inan, U.S., Barrington-Leigh, C., Hansen, S., Glukhov, V.S., Bell, T.F., Rairden, R., 1997. Rapid lateral expansion of optical luminosity in lightning-induced ionospheric flashes referred to as 'elves'. Geophysical Research Letters 24, 583–586.

Israelsson, S., Knudsen, E., 1986. Effects of radioactive fallout from a nuclear power plant accident on electrical parameters. Journal of Geophysical Research 91, 11,909–11,910.

Israelsson, S., Schutte, T., Pisler, E., Lundquist, S., 1987. Increased occurrence of lightning flashes in Sweden during 1986. Journal of Geophysical Research 92, 10,996–10,998.

Israelsson, S., Knudsen, E., Tammet, H., 1994. An experiment to examine the covariation of atmospheric electrical vertical currents at two separate stations. J. Atmospheric Electricity 14, 63–73.

Jackman, C.H., Cerniglia, M.C., Nielsen, J.E., Allen, D.J., Zawodny, J.M., McPeters, R.D., Douglass, A.R., Rosefield, J.E., Rood, B., 1995. Two-dimensional and three-dimensional model simulations, measurements and interpretation of the influence of the October 1989 solar proton events on the middle atmosphere. Journal of Geophysical Research 100, 11,641–11,660.

Jacobson, A.R., Knox, S.O., Franz, R., Enemark, D.C., 1999. FORTE observations of lightning radio-frequency signatures: capabilities and basic results. Radio Science 34 (2), 337–354.

Kasemir, H.W., 1963. On the theory of the atmospheric electric current flow, IV, Technical Report No. 2394, U.S. Army Electronics Research and Development Laboratories, Fort Monmouth, NJ, October.

Kasemir, H.W., 1977. Theoretical problems of the global atmospheric electric circuit. In: Dolezalek, H., Reiter, R. (Eds.), Electrical Processes in Atmospheres. Steinkopff, Darmstadt, pp. 423–438.

Kasemir, H.W., 1979. The atmospheric electric global circuit. In: Proceedings of Workshop on the Need for Lightning Observations from Space, NASA CP-2095, pp. 136–147.

Lyons, W., 1996. Sprite observations above the U.S. high plains in relation to their parent thunderstorm systems. Journal of Geophysical Research 101, 29,641–29,652.

Markson, R., 1978. Solar modulation of atmospheric electrification and possible implications for the Sun-weather relationship. Nature 273, 103–109.

Markson, R., 1983. Solar modulation of fair-weather and thunderstorm electrification and a proposed program to test an atmospheric electrical Sun-weather mechanism. In: McCormac, B.M. (Ed.), Weather and Climate Responses to Solar Variations. Colorado Associated University Press, Colorado, pp. 323–343.

Markson, R., 1986. Tropical convection, ionospheric potentials and global circuit variations. Nature 320, 588–594.

Markson, R., Price, C., 1999. Ionospheric potential as a proxy index for global temperatures. Atmospheric Research 51, 309–314.

Martell, E.A., 1985. Enhanced ion production in convective storms by transpired radon isotopes and their decay products. Journal of Geophysical Research 90, 5909–5916.

Michnowski, S., 1998. Solar wind influences on atmospheric electrical variables in polar regions. Journal of Geophysical Research 103, 13,939–13,948.

Nickolaenko, A.P., Satori, G., Zieger, B., Rabinowicz, L.M., Kudintseva, I.G., 1998. Parameters of global thunderstorm activity deduced from long-term Schumann resonance records. Journal of Atmospheric and Solar Terrestrial Physics 60, 387–399.

Ogawa, T., 1985. Fair-weather electricity. Journal of Geophysical Research 90 (4), 5951–5960.

Price, C., Rind, D., 1990. The effect of global warming on lightning frequencies, in: Proceedings of the 16th Conference on Severe Storms and Atmospheric Electricity, AMS, Kananaskis Park, Alberta, Canada, pp. 748–751.

Price, C., Rind, D., 1992. A simple lightning parameterization for calculating global lightning distributions. Journal of Geophysical Research 97, 9919–9933.

Price, C., 1993. Global surface temperatures and the atmospheric electrical circuit. Geophysical Research Letters 20, 1363–1366.

Price, C., 2000. Evidence for a link between global lightning activity and upper tropospheric water vapour. Nature 406, 290–293.

Pudovkin, M.I., Babushkina, S.V., 1992. Atmospheric transparency variations associated with geomagnetic disturbances. Journal of Atmospheric and Solar Terrestrial Physics 59, 1135–1138.

Recommended Atmospheric Electricity Research, 1998. The Atmospheric Sciences Entering the Twenty-First Century. National Academy Press, Washington, DC, p. 67.

Reeve, N., Toumi, R., 1999. Lightning activity as an indicator of climate change. Quarterly Journal of the Royal Meteorological Society 125, 893–903.

Retalis, D., Retalis, A., 1998. Effects of air pollution and wind on the large-ion concentration in the air above Athens. Journal of Geophysical Research 103, 13,927–13,932.

Rodger, C.J., 1999. Red sprites, upward lightning, and VLF perturbations. Reviews in Geophysics 37, 317–336.

Rycroft, M.J., 1965. Resonances of the Earth-ionosphere cavity observed at Cambridge, England, Radio Science. Journal of Research NBS 69D, 1071–1081.

Rycroft, M.J., 1990. The Antarctic atmosphere: a hot topic in a cold cauldron. Geographical Journal 156, 1–11.

Satori, G., 1996. Monitoring Schumann resonances — II. Daily and seasonal frequency variations. Journal of Atmospheric and Solar Terrestrial Physics 58, 1483–1488.

Sentman, D.D., Wescott, E.M., 1993. Observations of upper atmospheric optical flashes recorded from an aircraft. Geophysical Research Letters 20, 2857–2860.

Sentman, D.D., Wescott, E.M., 1996. Red sprites and blue jets: High-altitude optical emissions linked to lightning. EOS Transactions of American Geophysical Unit 77, 1–4.

Solomon, S., 1999. Stratospheric ozone depletion: a review of concepts and history. Reviews of Geophysics 37, 275–316.

Special Issue, 1998. Effects of thunderstorm activity on the upper atmosphere and ionosphere. Journal of Atmospheric and Solar-Terrestrial Physics 60, 667–973.

Svensmark, H., Friis-Christensen, E., 1997. Variation of cosmic ray flux and global cloud coverage — a missing link in solar climate relations. Journal of Atmospheric and Solar-Terrestrial Physics 59, 1225–1232.

Tammet, H., Iher, H., Salm, J., 1992. Spectrum of atmospheric ions in the mobility range 0.32–3.2 $cm^2/(V.s)$. Acta Comm. University Tartu 947, 39–49.

Tinsley, B.A., Heelis, R.A., 1993. Correlations of atmospheric dynamics with solar activity. Evidence for a connection via the solar wind, atmospheric electricity, and cloud microphysics. Journal of Geophysical Research 98D (6), 10,375–10,384.

Tinsley, B.A., 1996. Correlations of atmospheric dynamics with solar-wind-induced changes of air-Earth current density into cloud tops. Journal of Geophysical Research 101D (23), 29,701–29,714.

Tinsley, B.A., 2000. Influence of solar wind on the global electric circuit and inferred effects on cloud microphysics, temperature, and dynamics in the troposphere. Space Science Reviews, in press.

Volland, H., 1982. Quasi-electrostatic fields within the atmosphere. In: Volland, H. (Ed.),, CRC Handbook of Atmospherics, Vol. 1. CRC Press, Boca Raton, FL, pp. 65–109.

Vorpahl, J.A., Sparrow, J.G., Ney, E.P., 1970. Satellite observations of lightning. Science 169, 860–863.

Whipple, F.J.W., 1929. On the association of the diurnal variations of electric potential in fine weather with the distribution of thunderstorms over the globe. Quarterly Journal of the Royal Metereological Society 55, 1–17.

Willett, J.C., 1978. An analysis of the electrode effect in the limit of strong turbulent mixing. Journal of Geophysical Research 83, 402–408.

Williams, E.R., 1992. The Schumann resonance: a global thermometer. Science 256, 1184–1187.

Williams, E.R., Rutledge, S.A., Geotis, S.G., Renno, N., Rasmussen, E., Rickenbach, T., 1992. A radar and electrical study of tropical “hot towers”. Journal of Atmospheric Science 49, 1386–1395.

Wilson, C.T.R., 1920. Investigation on lightning discharges and on the electric field of thunderstorms. Philosophical Transactions of Royal Society of London A221, 73–115.

US005705763A

United States Patent [19]

Leon

[11] **Patent Number: 5,705,763**

[45] **Date of Patent: Jan. 6, 1998**

[54] **FIRE SELECTOR SYSTEM FOR SELECTING BETWEEN AUTOMATIC AND SEMI-AUTOMATIC OPERATION OF A GUN**

[76] Inventor: **Jorge A. Leon**, Av. Macuto, Res. Macaracuay, Apto. #4B, Macaracuay Caracas, Venezuela

[21] Appl. No.: **685,184**

[22] Filed: **Jul. 18, 1996**

[51] **Int. Cl.**^{6} **F41A 19/33**

[52] **U.S. Cl.** **89/140**; 89/128

[58] **Field of Search** 89/140, 141, 142, 89/128

[56] **References Cited**

U.S. PATENT DOCUMENTS

2,216,412	10/1940	Holek	89/142
2,432,486	12/1947	Patchett	89/142
2,572,872	10/1951	Ketterer	89/140
2,718,818	9/1955	Cuppini et al.	89/142
2,909,100	10/1959	Kennerley-Tayler	89/140
3,012,479	12/1961	Ruffell	89/142
3,021,763	2/1962	Beretta	89/140
3,358,560	12/1967	Ruffell	89/142
4,539,889	9/1985	Clock	89/147
5,355,768	10/1994	Felk	89/147

FOREIGN PATENT DOCUMENTS

1148550	12/1957	France	89/140

OTHER PUBLICATIONS

Wooldridge, J. "The Enforcer," Popular Mechanics, Sep., 1991 pp. 39–42.

Primary Examiner—Stephen M. Johnson
Attorney, Agent, or Firm—Heslin & Rothenberg, P.C.

[57] **ABSTRACT**

A fire selector system for selecting between automatic and semi-automatic operation of a gun is provided. The system is mountable on a reciprocating slide of the gun and into operative relationship with the trigger bar of the gun. A member of the selector system is movable between forward and rearward longitudinal positions. The forward longitudinal position allows the member to assume an activating longitudinal position for automatic operation of the gun. The rearward longitudinal position prevents the member from assuming its activating longitudinal position, thereby effecting semi-automatic operation of the gun. The member repetitively contacts a contact area of the trigger bar as it reciprocates with a reciprocating slide of the gun during automatic operation. Further, the device is easily installed and removed from the reciprocating slide depending on user requirements.

7 Claims, 6 Drawing Sheets

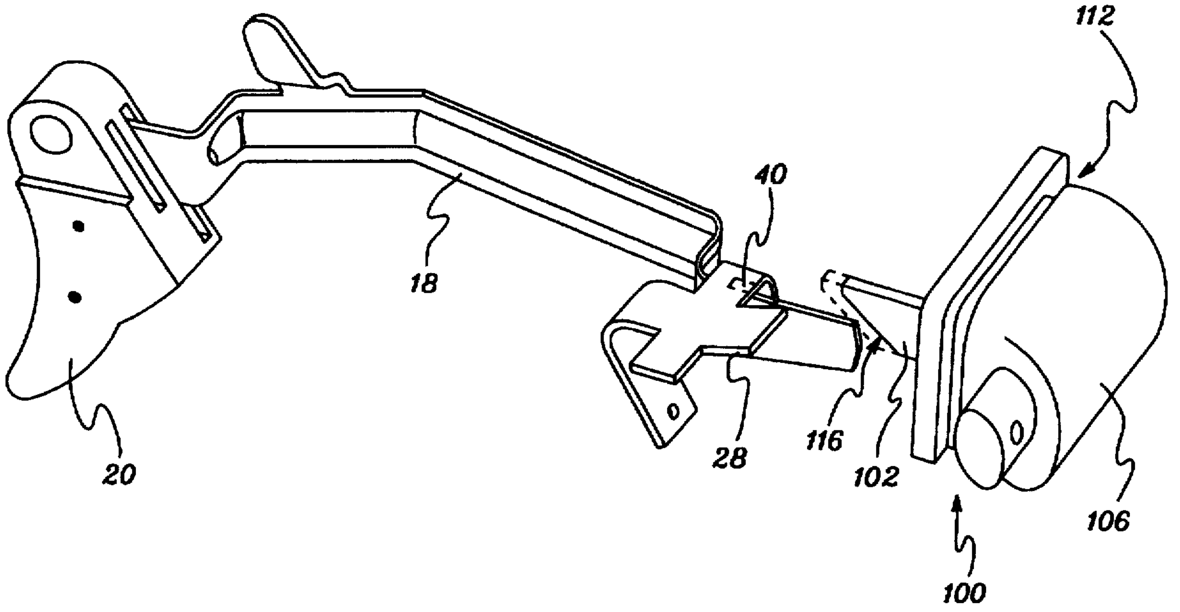

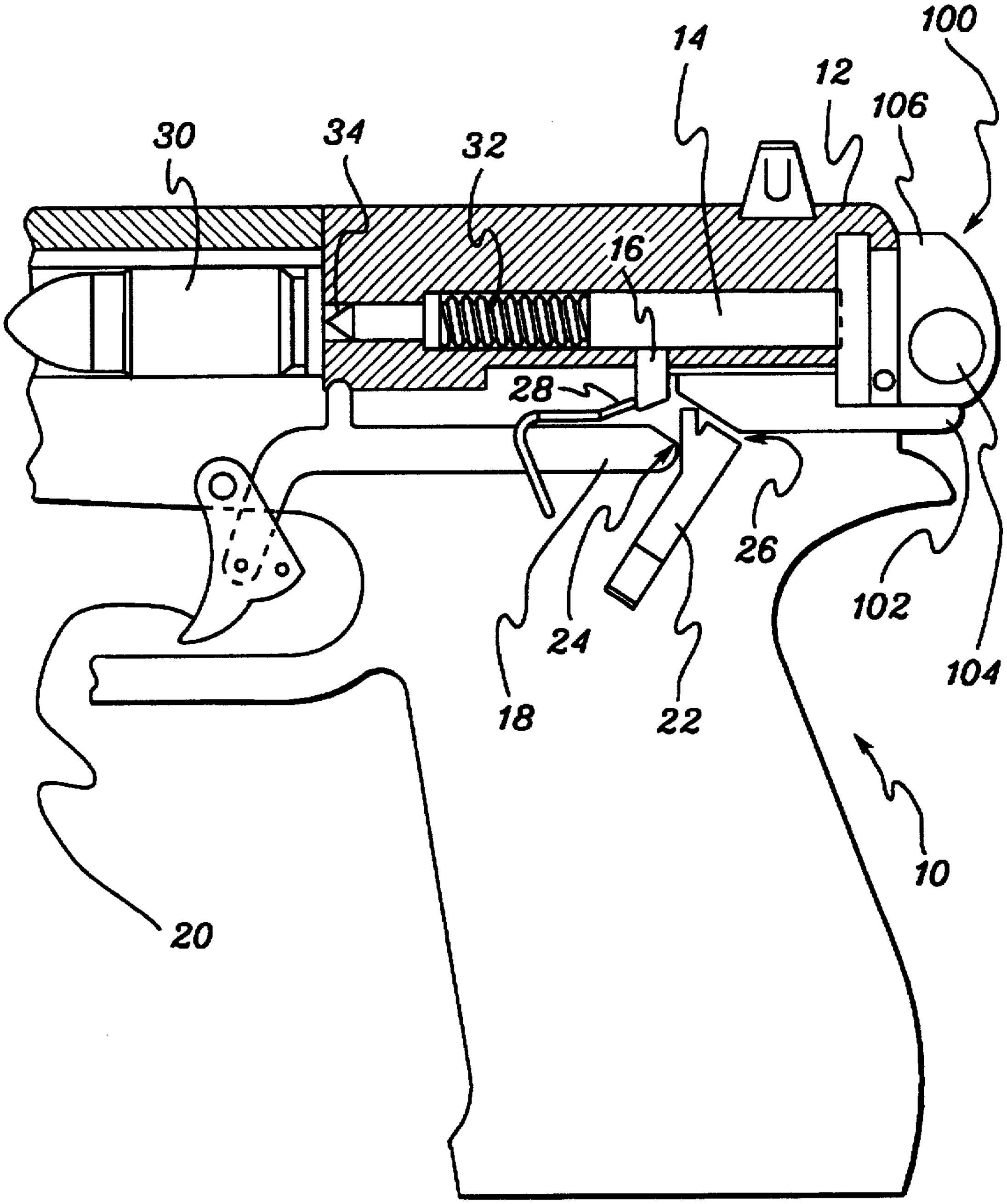
100
14
12
106
30
34
32
16
28
26
24
22
18
102
104
20
10

FIG. 1a

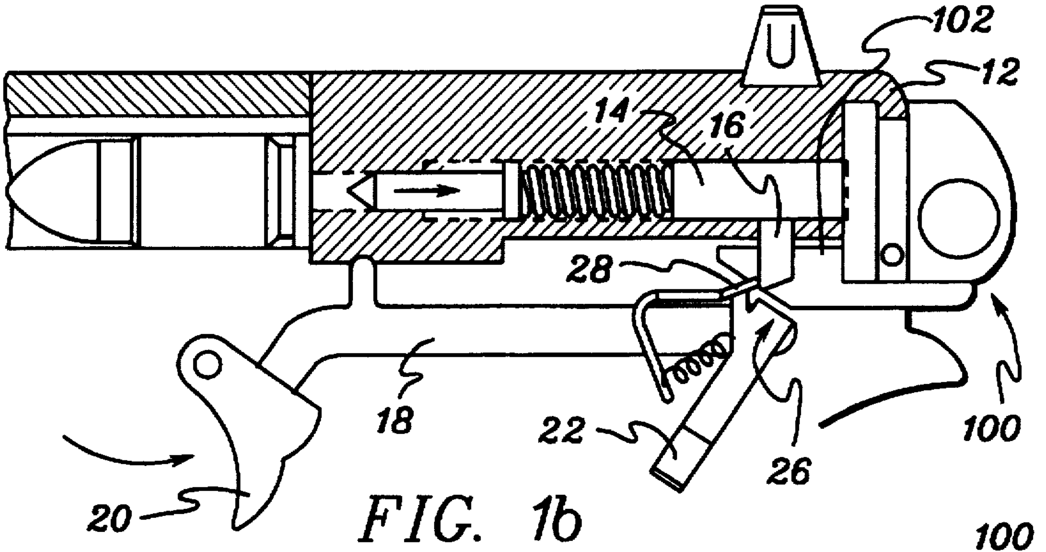
102
12
14
16
28
18
22
26
100
20

FIG. 1b

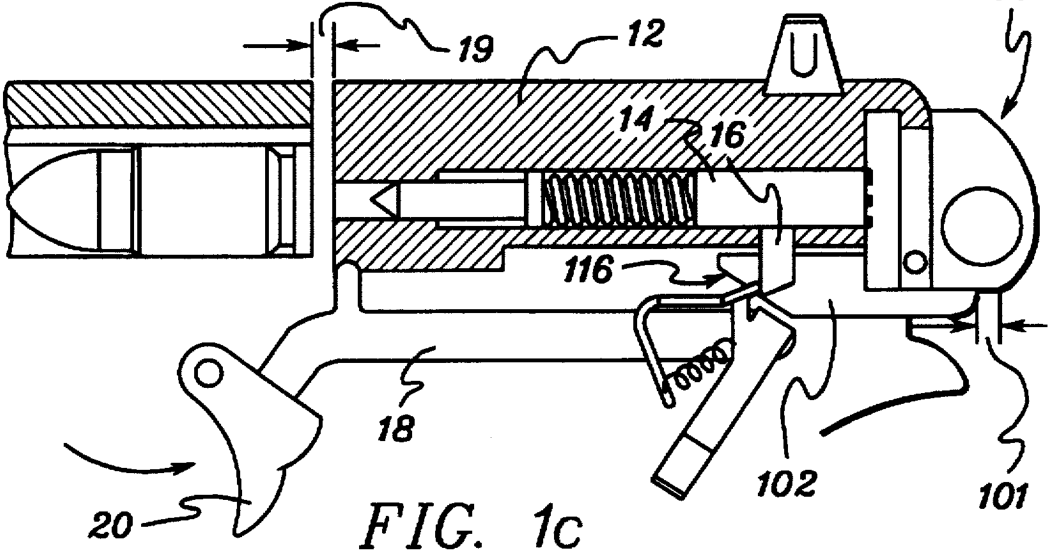
100
19
12
14
16
116
18
102
101
20

FIG. 1c

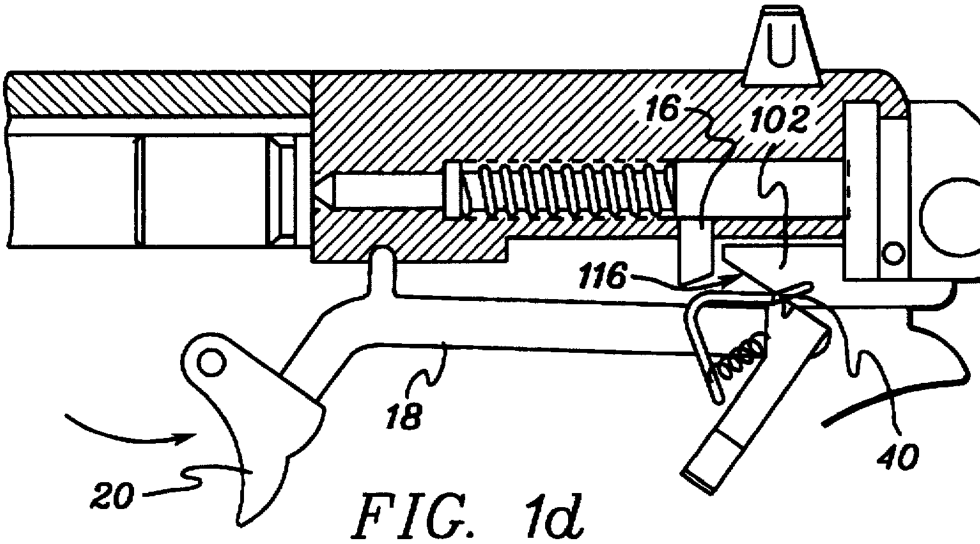
16
102
116
18
40
20

FIG. 1d

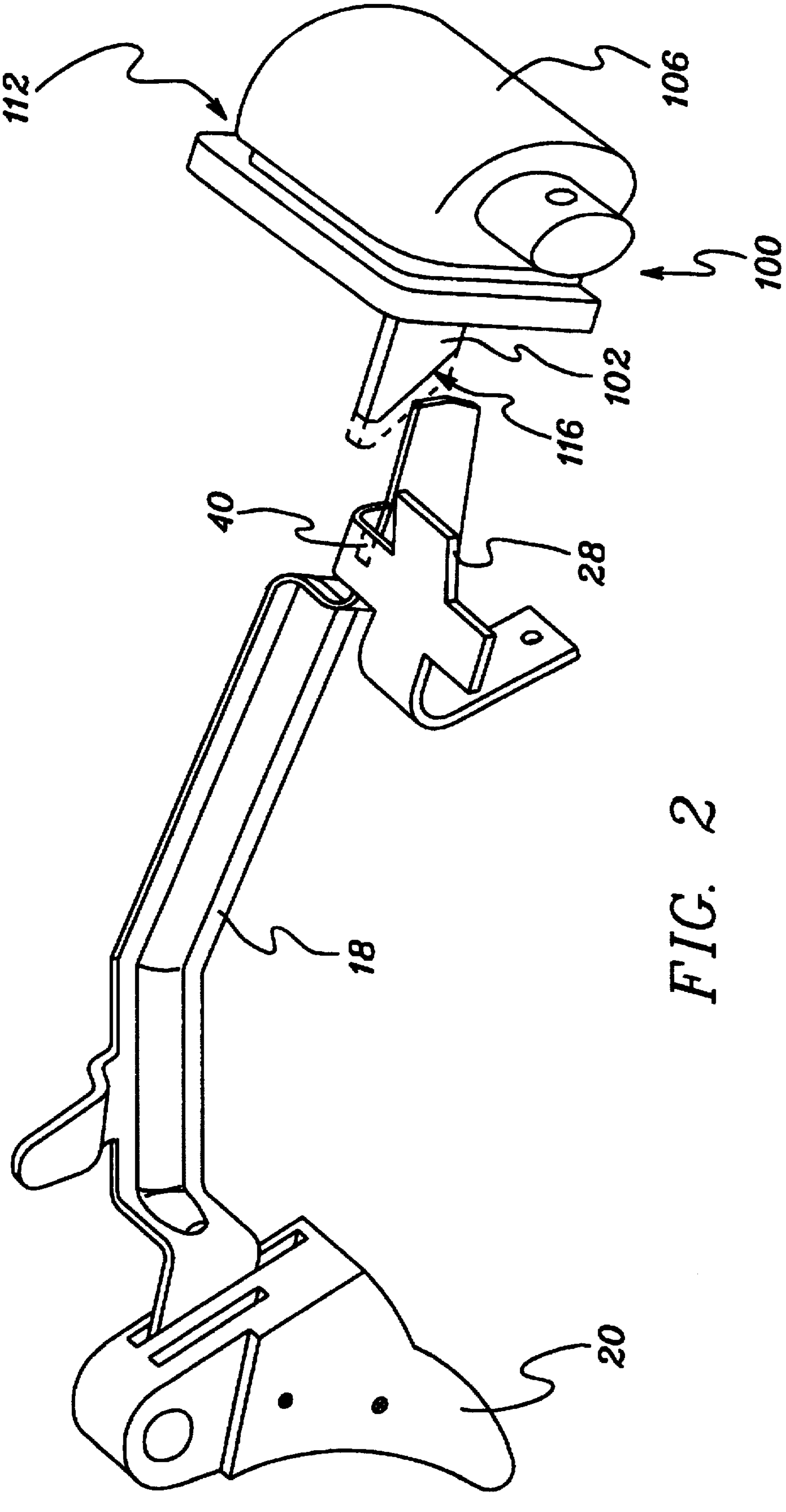
112
106
100
102
116
40
28
18
20

FIG. 2

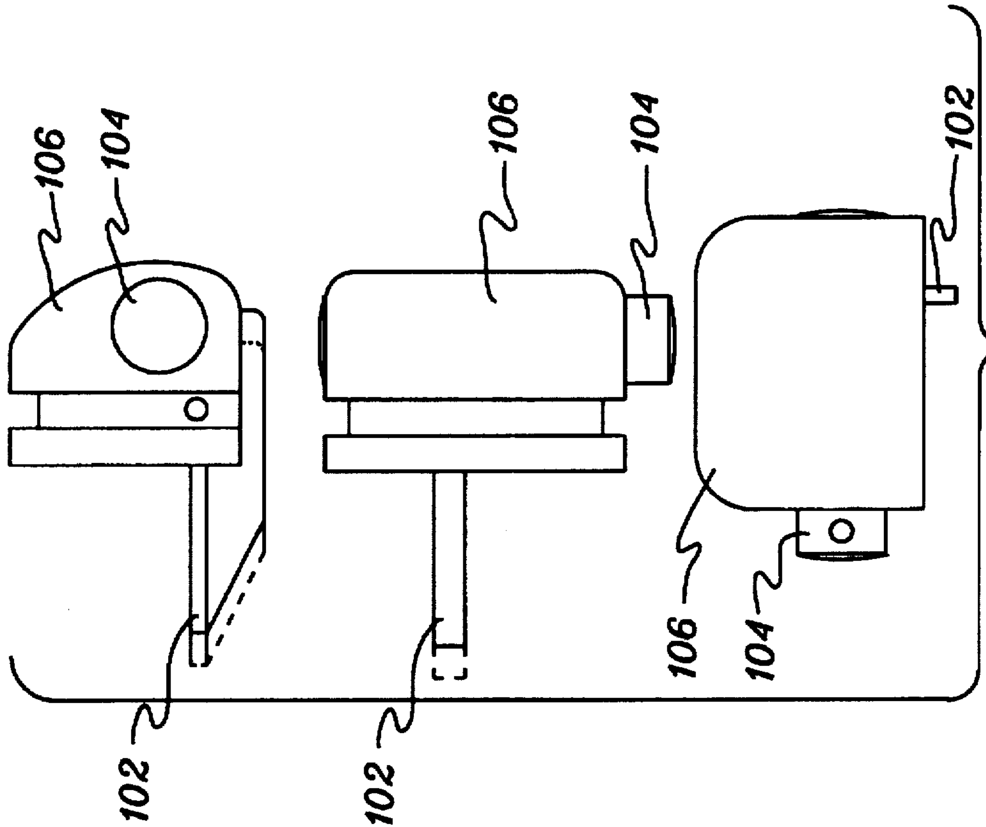

FIG. 3b

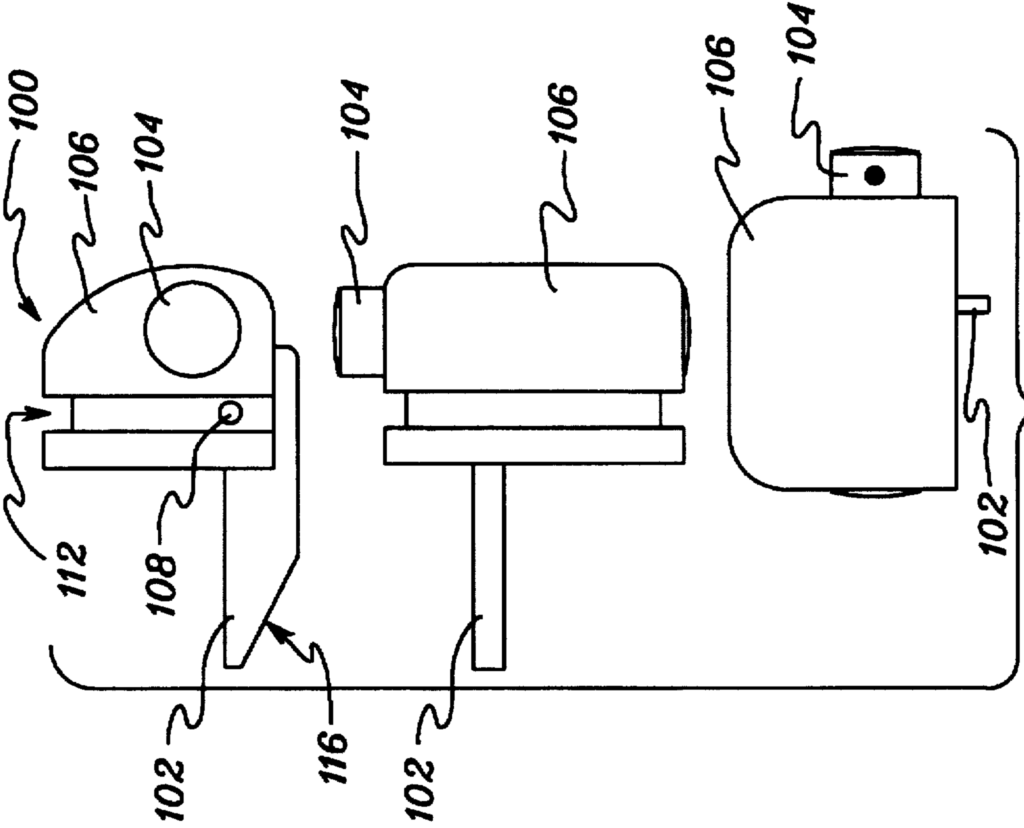

FIG. 3a

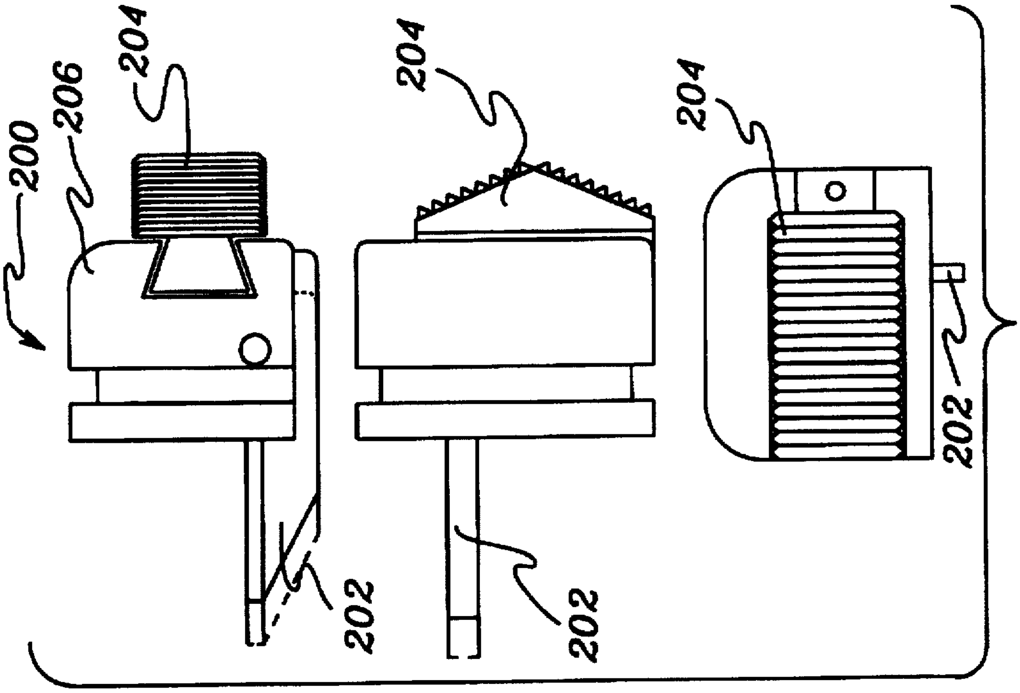

FIG. 4b

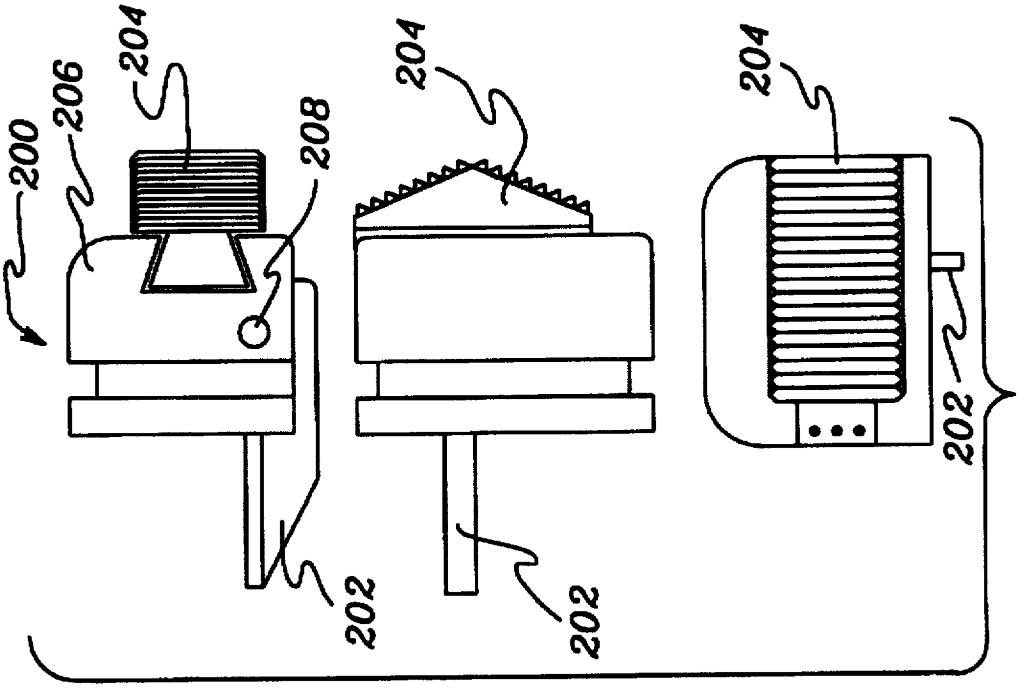

FIG. 4a

FIG. 5a

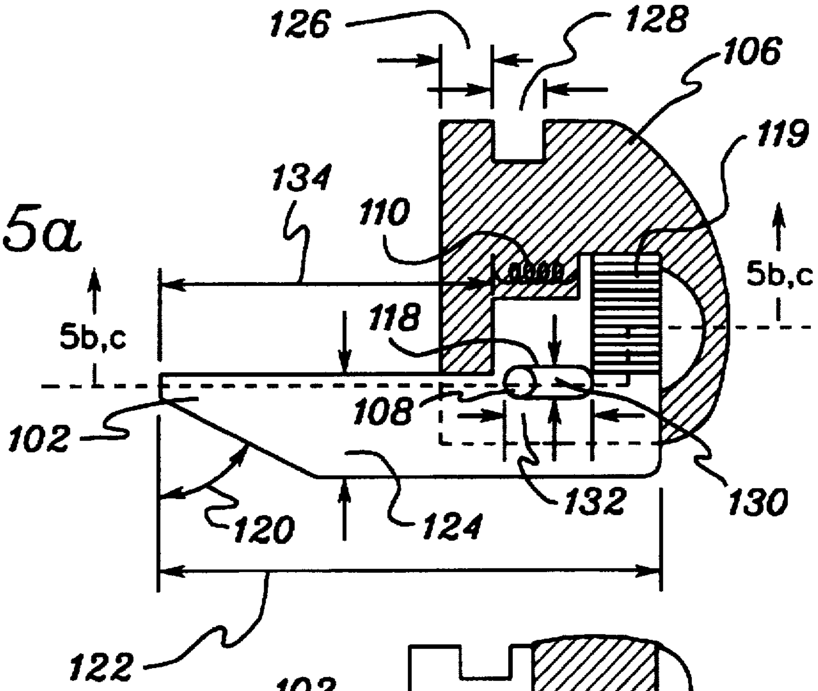
126
128
106
119
134
110
5b,c
118
5b,c
102
108
132
130
120
124
122

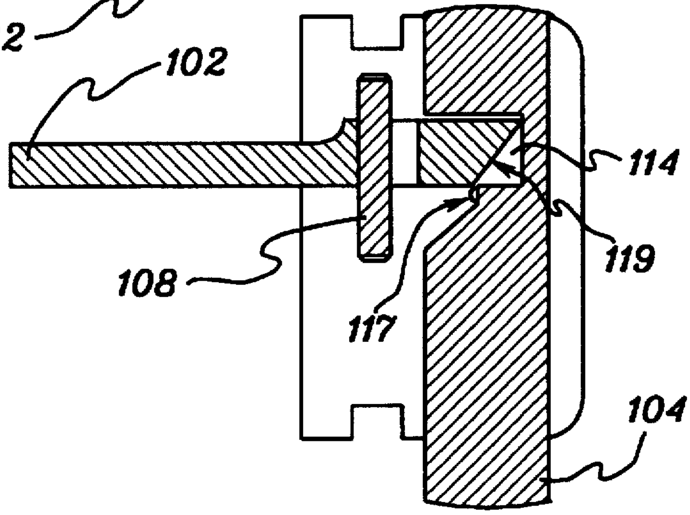
102
114
108
119
117
104

FIG. 5b

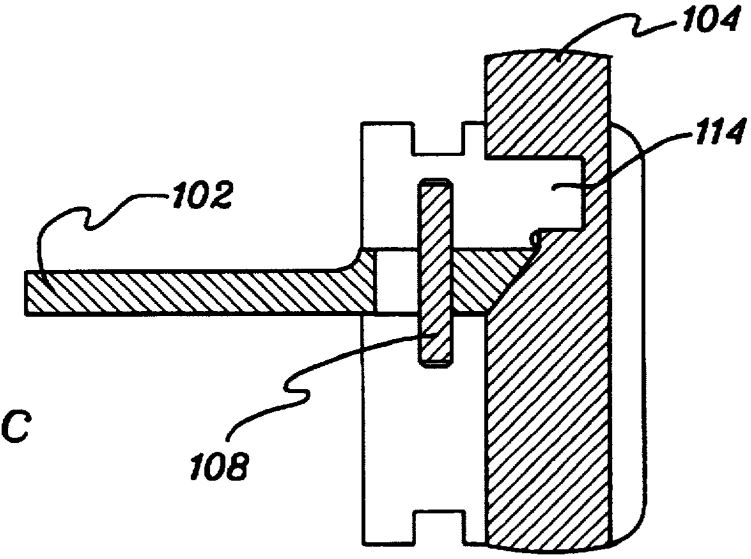
104
114
102
108

FIG. 5c

FIRE SELECTOR SYSTEM FOR SELECTING BETWEEN AUTOMATIC AND SEMI-AUTOMATIC OPERATION OF A GUN

TECHNICAL FIELD

This invention relates to semi-automatic and automatic firearms. More particularly, this invention is a fire selector system for selectively converting a firearm between semi-automatic and automatic operation.

BACKGROUND OF THE INVENTION

Many firearms are sold in configurations which provide semi-automatic operation or automatic operation. However, it is often desirable to convert an existing semi-automatic firearm to an automatic firearm, without significant modification thereof.

Having made an initial investment in a semi-automatic firearm, many users are inclined not to make an additional, potentially equal investment in a related firearm which provides automatic operation.

Various mechanisms have been proposed in the prior art for effecting selection between semi-automatic and automatic operation. For example, U.S. Pat. No. 3,012,479 to L. E. Ruffell discloses an abutment opposite a recess in a sear, and movable between three positions for safe, semi-automatic, and automatic operation. However, this device does not appear to convert an existing semi-automatic gun to automatic operation. U.S. Pat. No. 2,572,872 to S. J. Ketterer discloses an attachment for converting a semi-automatic rifle to full automatic, but which is dependent on pivoting motion of a lever, triggered by forward movement of a bolt. An arm of the lever thereby rotates a sear into a hammer releasing position to effect automatic operation. This pivoting action of the retrofitted attachment requires significant, independent movement of the attachment itself during automatic operation, which is undesirable as it may lead to failures, mis-fires, etc. Further, cuts to the stock of the rifle are necessary to install the attachment.

Thus, what is required is a simple device for converting a semi-automatic firearm into an automatic firearm. It is preferable that such a device have few moving parts, be easily installed and removed from the existing semi-automatic firearm, operate in a simple manner, and be inexpensive.

SUMMARY OF THE INVENTION

The shortcomings of the prior art are overcome by the present invention, which in one aspect is a device for converting a semi-automatic gun into an automatic gun. The gun has a trigger bar, a reciprocating slide, and a longitudinal firing mechanism. The trigger bar is normally mounted below the firing mechanism and is movable into a first position in which the trigger bar engages the firing mechanism and prevents forward motion of the firing mechanism following trigger activation. Semi-automatic operation is thus effected by this engagement. The converting device includes a member mountable on the reciprocating slide of the gun and into operative relationship with the trigger bar of the gun. The member is movable into an activating longitudinal position in which contact between the member and a contact area of the trigger bar occurs to urge the trigger bar away from its first position following trigger activation. Automatic operation of the gun is thus effected by this contact.

The member may be reciprocatable by a user between forward and rearward longitudinal positions. The forward longitudinal position allows the member to assume its activating longitudinal position for automatic operation of the gun. The rearward longitudinal position prevents the member from assuming its activating longitudinal position and thereby allows the trigger bar to assume and retain its first position following trigger activation thus effecting semi-automatic operation.

The device may include a switch for longitudinally moving the member between its forward and rearward longitudinal positions.

The device reciprocates with the reciprocating slide upon its mounting thereon, and, when in its forward longitudinal position, can assume its activating longitudinal position and repetitively contact the contact area of the trigger bar during its reciprocation with the reciprocating slide. The member does not assume its activating longitudinal position and is thus isolated from the contact area of the trigger bar during its reciprocation when the member is in its rearward longitudinal position.

The member may include a tapered tip for contacting the contact area of the trigger bar, the tapered tip gradually urging the trigger bar away from its first position as the member is moved into its activating longitudinal position.

As set forth in greater detail hereinbelow, the fire selector system of the present invention can be easily installed and removed from an existing semi-automatic firearm, thus providing the option of a convertible gun or a non-convertible gun. Further, the fire selector system has very few moving parts, operates in a simple manner and is relatively inexpensive to fabricate.

BRIEF DESCRIPTION OF DRAWINGS

The subject matter which is regarded as the invention is particularly pointed out and distinctly claimed in the concluding portion of the specification. The invention, however, both as to organization and method of practice, together with further objects and advantages thereof, may best be understood by reference to the following detailed description of the preferred embodiment(s) and the accompanying drawings in which:

FIG. 1*a* depicts a partial cross-section of an exemplary gun having the fire selector system of the present invention mounted therein, the gun being cocked and ready to fire a single round, the fire selector system being set for semi-automatic operation;

FIG. 1*b* depicts a position of the components of the gun of FIG. 1*a* at a point during a trigger activation;

FIG. 1*c* depicts the gun having the fire selector system therein, the fire selector system being set for automatic operation, the components shown at a point following trigger activation and during forward movement of the reciprocating slide;

FIG. 1*d* depicts the moment of firing of the gun of FIG. 1*c*;

FIG. 2 is a perspective view of the trigger bar and fire selector systems of FIGS. 1*a*–*d*, and further illustrates the positional relationship therebetween;

FIGS. 3*a*–*b* are various views of the fire selector system of the present invention in its extended (automatic) and retracted (semi-automatic) settings, respectively;

FIGS. 4*a*–*b* depict an alternate embodiment of the fire selector system in its extended and retracted positions, respectively; and

FIGS. 5*a*–*c* are cross-sectional views depicting additional detail of the fire selector system of FIGS. 1–3.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT(S)

With reference to FIGS. 1*a–d*, depicted therein is an exemplary pistol **10** having a fire selector system **100** installed thereon, according to the principles of the present invention. FIGS. 1*a–b* depict the fire selector system in its setting for semi-automatic operation. FIGS. 1*c–d* depict the fire selector system in its setting for automatic operation.

As shown in FIG. 1*a*, the components of the gun **10** relevant to the instant invention are as follows. Trigger **20**, shown in its unactivated position, is hingedly connected to a trigger bar **18**. A wing **28** of trigger bar **18** is engaging a leg **16** of firing pin **14** (or mechanism). Following backward motion of the trigger, or trigger activation, trigger bar **18** moves rearward and an inclined end **24** thereof contacts an inclined control surface **26** (perpendicular to the page) of a connector **22**. As end **24** reaches inclined control surface **26**, the trigger bar is pushed downward such that nose **16** is released and, urged by spring **32**, firing pin tip **34** will move forward, contact bullet **30**, and discharge the same.

Pursuant to the principles of the present invention, a fire selector system **100** is provided on the rear of the slide **12**. The system **100** has a housing **106** adapted for mounting on the slide, a switch **104** and a disconnector member **102**. Disconnector **102** is movable into either a forward or rearward longitudinal position using switch **104**. Disconnector **102** in FIGS. 1*a–b* is shown in its rearward longitudinal position, for semi-automatic operation. In this position, no interaction will take place between the fire selector system and the trigger bar of the gun. Therefore, as discussed above, normal semi-automatic operation will occur based on the interaction between trigger bar end **24** and control surface **26** upon activation of the trigger.

Such trigger activation is shown in FIG. 1*b*, wherein the components of the gun are shown at a moment during trigger activation. Wing **28** has pulled leg **16** backward to a point where the end **24** of trigger bar **18** (now obscured by connector **22**) is beginning to contact inclined control surface **26**. Upon further trigger pull, the control surface **26** will guide trigger bar **18** downward, thereby releasing leg **16** and bolt **14** and firing the gun. As discussed above, the disconnector **102** of fire selector system **100** is generally out of engagement with any of these firing components of the gun during semi-automatic operation. Thus, the gun functions in its normal, semi-automatic operation with disconnector **102** in its rearward longitudinal position. Discharge of the bullet and the subsequent backward motion of the slide **12**, and release of the trigger, will effect a re-engagement of wing **28** with leg **16**, in preparation for another trigger activation.

FIGS. 1*c–d* depict automatic operation of the gun, wherein the disconnector **102** of the fire selector system **100** is shown in its forward longitudinal position (i.e., moved a distance **101**). FIG. 1*c* depicts forward motion of the slide following recoil from a previous trigger activation. Inclined surface **116** of the forward tip of the disconnector **102** has been moved a sufficient distance **101** forward such that contact and downward motion of the trigger bar occurs. This contact occurs at a point **19** about 1 or 2 millimeters before the slide **12** reaches its normal forward position. The firing pin **14** is thus released at this point in the reciprocation of the slide and another shot is fired without a trigger activation. Control surface **116** of the disconnector **102** contacts the trigger bar **18** to effect this firing. FIG. 1*d* depicts the moment of firing at which time the control surface **116** of the disconnector **102** has made contact approximately at a point or contact area **40** of the trigger bar. The longitudinal position at which the disconnector begins to engage the trigger bar is referred to herein as the activating longitudinal position.

As long as the trigger is pulled, and ammunition is provided in the clip, the disconnector will repetitively disengage the firing pin by repetitively engaging the trigger bar and repetitive firing will result as the selector system reciprocates with the slide.

FIG. 2 depicts, in perspective form, the relevant components effecting automatic operation including trigger **20**, trigger bar **18** and selector system **100**. The inclined control surface **116** of disconnector **102** of the selector system **100** is designed to effect contact at about contact area **40** of trigger bar **18**, and gradually urge the trigger bar **18** downward as the disconnector moves forward.

FIG. 3*a* includes side, top and rear views of the fire selector system of the present invention, with the disconnector **102** in its extended (automatic) setting. As discussed above, the components of this system **100** include a disconnector **102**, housing **106** and a cylindrical switch **104** running laterally through the housing. A pin **108** can be provided to hold the disconnector **102** in housing **106**. Further, the housing **106** may have a flange **112** formed thereon to effect simple installation and removal of this selector system from the gun. FIG. 3*b* depicts the same views of the selector system with the disconnector in its retracted position, using the opposite switch setting. FIGS. 3*a* and 3*b* clearly show that a first switch position effects extension of the disconnector **102** thus effecting automatic operation, and a second switch position effects retraction of the disconnector **102** thus effecting semi-automatic operation.

FIG. 4*a* includes side, top and rear views of an alternate embodiment **200** of a selector system pursuant to the principles of the present invention, with the disconnector in its extended (automatic) setting. Disconnector **202**, housing **206** and pin **208** are similarly provided, however, a thumb switch **204** is provided in this embodiment rather than the cylindrical switch running laterally through the housing set forth in FIG. 3. FIG. 4*b* depicts the same views of the alternate embodiment **200** of the fire selector system with the disconnector in its retracted (semi-automatic) setting.

FIG. 5*a* is a sectional, side view of the fire selector system of FIGS. 3*a–b*, illustrating additional detail of the system. A spring **110** is shown therein which operates to urge the disconnector **102** into a rearward position. Pin **108** is shown positioned through a via **118** of the member **102**. The pin is shown in front of the via indicating that the disconnector **102** is in its retracted longitudinal position. An inclined surface **119** of the disconnector **102** facilitates movement of the disconnector **102** between its retracted and extended longitudinal positions, as illustrated in the sectional, top views of FIGS. 5*b–c*, which also show a cross-section of the laterally positioned switch **104**. Switch **104** includes a recess **114** into which disconnector **102** is urged by spring **110** in its retracted position. Further, a small ridge **117** is provided which, as shown in FIG. 5*c*, holds disconnector **102** in its extended longitudinal position. The inclined surface **119** of the disconnector facilitates movement into the extended position, and the inclined surfaces of ridge **117** similarly facilitate movement from the extended position into the retracted position.

Various dimensions of the fire selector system of FIGS. 5*a–c* are set forth in the table below.

120	54°
122	21.95 mm
124	4.65 mm
126	2 mm
128	2.45 mm
130	2.06 mm
132	4.81 mm
134	14.7 mm

An exemplary embodiment of the fire selector system is fabricated using glass filled reinforced polymer for the switch and housing, hardened steel for the disconnector and spring, and carbon steel for the pin.

The above-described selector system therefore has few moving parts, and further is composed of relatively inexpensive parts.

Using the above-described dimensions for the housing and disconnector, a fire selector system is provided to select between semi-automatic and automatic operation of a Glock pistol of the type described in U.S. Pat. No. 4,825,744, hereby incorporated herein by reference in its entirety. The fire selector system of the present invention can be adapted to replace a slide cover plate of that pistol at the rear of the slide, and it can thus be easily installed and removed from the pistol. In its installed position, the fire selector system provides the user with a choice between semi-automatic and automatic operation; and upon its removal, the gun operates in its normal semi-automatic mode.

While the invention has been described in detail herein in accordance with certain preferred embodiments thereof, many modifications and changes therein may be affected by those skilled in the art. Accordingly, it is intended by the following claims to cover all such modifications and changes as fall within the true spirit and scope of the invention.

What is claimed is:

1. A device for converting a semi-automatic gun into an automatic gun, the gun having a trigger bar, a reciprocating slide, and a longitudinal firing mechanism, the trigger bar mounted below the firing mechanism and movable into a first position in which the trigger bar engages the firing mechanism and prevents forward motion of the firing mechanism following trigger activation, thereby effecting semi-automatic operation, the device comprising:

a member mountable on the reciprocating slide of the gun and in operative relationship with the trigger bar of the gun, the member movable into an activating longitudinal position in which contact between the member and a contact area of the trigger bar occurs to urge the trigger bar away from its first position following trigger activation thereby effecting automatic operation of the gun, wherein the member is reciprocatable between forward and rearward longitudinal positions, the forward longitudinal position allowing the member to assume its activating longitudinal position for said automatic operation of the gun, the rearward longitudinal position preventing the member from assuming its activating longitudinal position, thereby allowing the trigger bar to assume and retain its first position following trigger activation to effect said semi-automatic operation of the gun;

and wherein the device further comprises a user switch coupled for longitudinally moving the member, the user switch having a first position corresponding to the forward longitudinal position of the member and a second position corresponding to the rearward longitudinal position of the member.

2. The device of claim **1**, further comprising a housing in which the member and the user switch are mounted, the housing adapted to be mountable on a rearward portion of the reciprocating slide of the gun with said member extending forward into the reciprocating slide and into the operative relationship with the trigger bar of the gun upon mounting of the housing on the rearward portion of the reciprocating slide.

3. The device of claim **2**, wherein the device reciprocates with the reciprocating slide upon its mounting thereon, the member assuming its activating longitudinal position and repetitively contacting the contact area of the trigger bar during its reciprocation with the reciprocating slide when the member is moved into its forward longitudinal position, the member not assuming its activating longitudinal position and being isolated from the contact area of the trigger bar during its reciprocation with the reciprocating slide when the member is moved into its rearward longitudinal position.

4. The device of claim **1**, wherein the user switch is elongate and runs laterally through the housing, and wherein the device comprises a spring for urging a rearward end of the member against a side of the user switch, the side of the user switch having a recess, such that:

in a first lateral switch position comprising said first position, the rearward end of the member is positioned outside of the recess and the member thereby assumes its forward longitudinal position, and

in a second lateral switch position comprising said second position, the rearward end of the member is urged into the recess and the member thereby assumes its rearward longitudinal position.

5. A device, mountable on a rearward portion of a semi-automatic gun, for selectively converting the gun between semi-automatic and automatic operation, the device comprising:

a housing adapted to be mountable on the rearward portion of the gun; and

a member extending from said housing and into the gun upon mounting of the device on the gun, the member movable between an extended longitudinal position and a retracted longitudinal position, wherein

in the extended longitudinal position a forward end of the member engages a trigger bar of the gun and urges the trigger bar out of engagement with a firing mechanism of the gun following trigger activation thereby effecting said automatic operation, and

in the retracted longitudinal position, the member is disengaged from the trigger bar of the gun following trigger activation, thereby effecting said semi-automatic operation;

and wherein the device further comprises a switch for moving the member between its extended and retracted longitudinal positions.

6. The device of claim **5**, wherein the forward end of the member is tapered, the tapered forward end gradually urging the trigger bar out of engagement with the firing mechanism when the member is in its extended longitudinal position.

7. A gun having a device therein for facilitating selection between semi-automatic and automatic operation of the gun, the gun having a trigger bar and a firing mechanism, the trigger bar being disposed below the firing mechanism and movable into an upper position in which the trigger bar engages the firing mechanism and prevents forward motion of the firing mechanism following trigger activation to effect said semi-automatic operation, the device comprising:

a member for placement in operative relationship with the trigger bar of the gun and moveable by a user between a forward and rearward longitudinal position, the rearward longitudinal position of the member allowing the trigger bar to assume its upper position following trigger activation thereby effecting said semi-automatic operation by allowing engagement between the trigger bar and the firing mechanism, the forward longitudinal position of the member effecting contact between the device and the trigger bar and urging the trigger bar

away from its upper position following trigger activation thereby effecting said automatic operation, wherein the device includes a user switch for longitudinally moving the member, the user switch having a first position corresponding to the forward longitudinal position of the member and a second position corresponding to the rearward longitudinal position of the member.

* * * * *

INSTALLATION PROCEDURE FOR THE GLOCK 9MM PISTOLS

All NFA Rules and Regulations Apply!

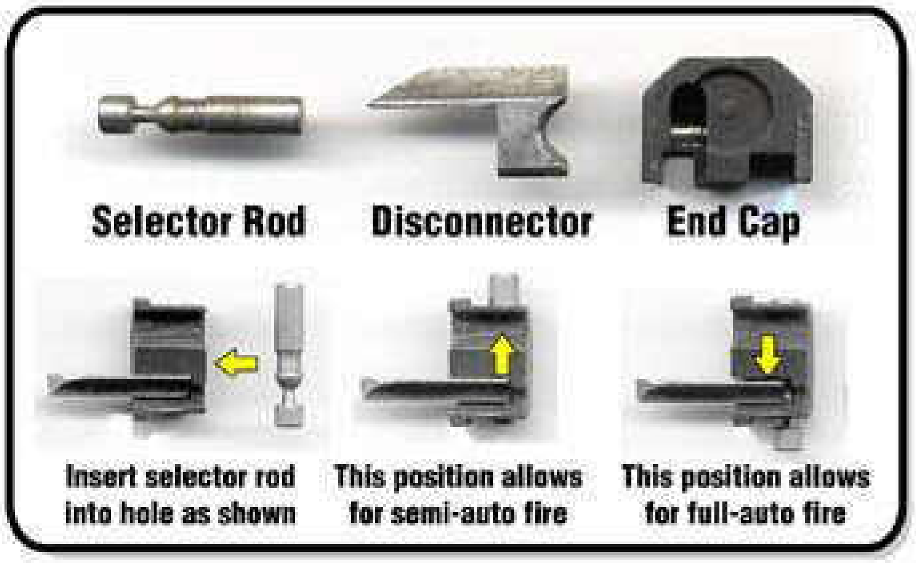
Selector Rod
Disconnector
End Cap
Insert selector rod into hole as shown
This position allows for semi-auto fire
This position allows for full-auto fire

Step (1) Assembly Of The Unit:

Your Glock device consists of three parts; Selector Rod, Disconnector, and an End Cap. These three items will replace your existing plastic "End Cap" located on the back most part of your Glock 9mm pistol. These pictures show the assembled unit along with the two positions for the "Selector Rod".

Please note that when the "Selector Rod" is set for full-auto fire, you are still able to control the amount of fire from the pistol through trigger pul.. A short, quick trigger pull will release a small burst while a lengthy pull will release as many as your Glock 9mm magazine can hold.

Proper care should be given to all parts of the device assembly as well as observing all BATF laws and regulations relevant to this device.

Step (2) Preparing For Installation:

Verify that you are working with a UNLOADED weapon by inspecting the chamber for loose rounds and any magazines that may be used in the process of installing the Glock device. Pull back the slide of your Glock with an EMPTY magazine installed. This places the weapon in the "Out of Ammo" mode. This will allow you to comfortably install the Glock device without disassembly of the weapon. Have a sturdy key or screwdriver handy for installation. Be careful not to damage any surrounding parts and above all, do not force the parts!.

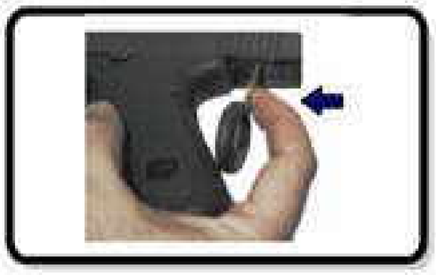

Step (3) Preparing For Installation:

Use you sturdy car key or a flathead screw driver between the firing pin notch and the firing pin jacket as shown in the illustration to the left. Push the key or screwdriver towards the front of the gun. This will allow you to remove the existing end cap

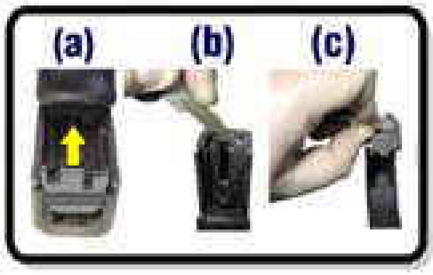

Step (4) Preparing For Installation:

While you are holding back the firing pin you can now easily remove the existing end cap as shown in illustration (a). After you have successfully removed the existing end cap, again use the key or screwdriver to push the firing pin jacket from the rear to allow your fully assembled Glock device to slip over and into the empty end cap slot as shown in illustration (b). Use the key or screwdriver again to push back the extractor plunger to allow the fully assembled device to completely slide into the end cap slot as shown in illustration (c).

Step (5) Preparing For Test Firing:

Pull and release the slide several times to allow the device to create a small channel for the disconnector on the frame. Your Glock pistol is now ready to test fire. Please see the illustrations above for setting the "Selector Rod" to allow for full-auto or semi-auto fire.

Troubleshooting:

If the pistol does not close or closes improperly simply disassemble the device from the weapon and replace with the standard "End Cap". If the slide now closes, the problem is with the disconnector. There are two possible reasons for this:

(1) The disconnector's gap has not formed. (Verify that there is a gap formed by the disconnector. If not, repeat step (5) after reassembly.

(2) The curved part of the disconnector does not allow the trigger to complete a cycle. There are two methods to correct this:

   (a) With a Dremel tool and a sanding stone you could enhance the disconnector's shape. (caution not to over-sand).

   (b) With a Dremel tool and a sanding cylinder you can smooth out the trigger bar's little notch (you should disassemble the pistol before doing this).

Installing the Device without selector bar or disconnector (and further troubleshooting):

If you wish to install the device without the selector rod and the disconnector you may still have difficulty with a non-closing slide. Try the following:

(1) Using sand paper, sand the left side (square protuberance on the top) allowing the device to enter with the slide over the frame

(2) Use a small file to file the device a small amount in the recesses opposite to the disconnector.

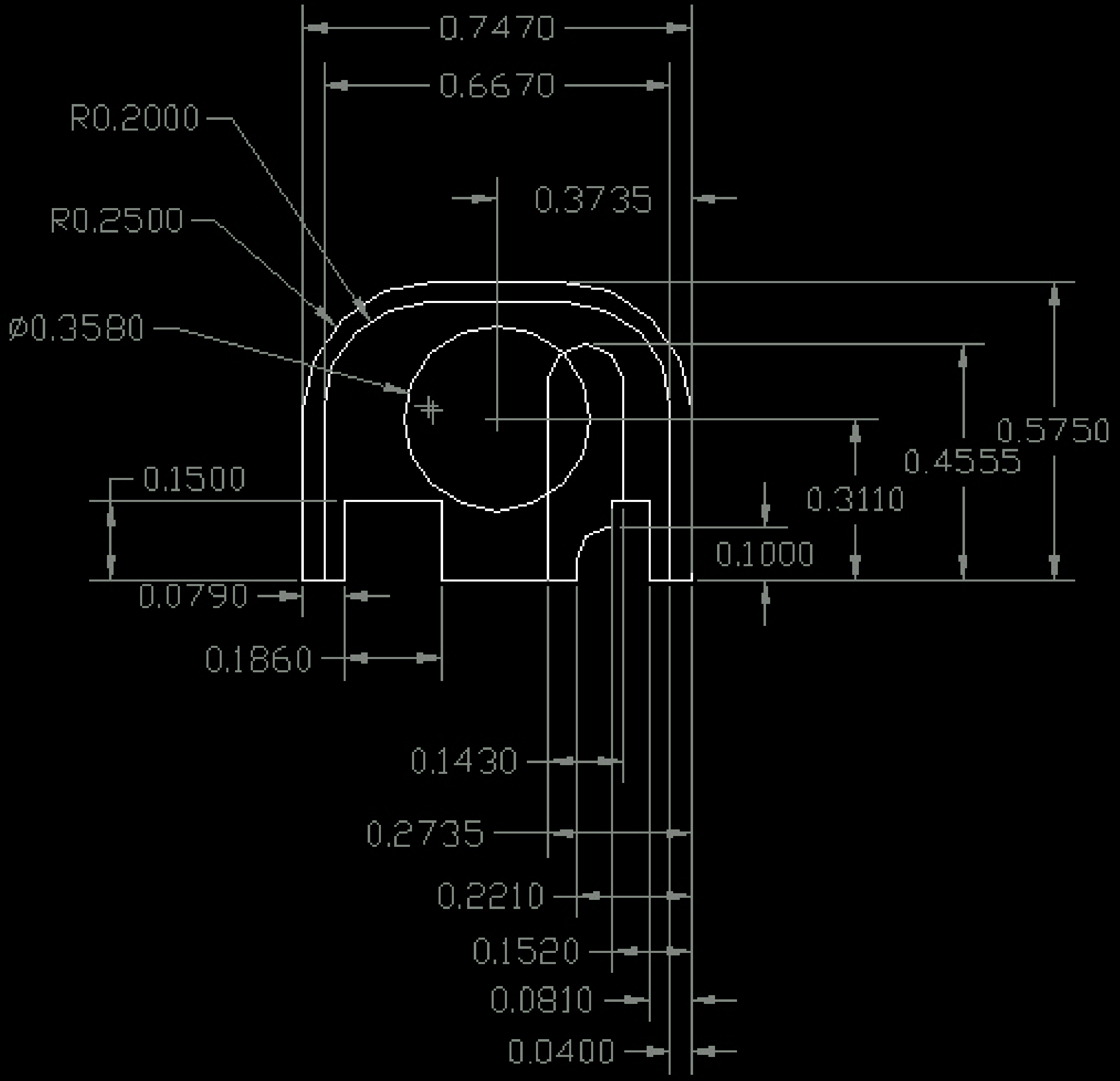
0.7470
0.6670
R0.2000
R0.2500
0.3735
ø0.3580
0.5750
0.4555
0.3110
0.1500
0.1000
0.0790
0.1860
0.1430
0.2735
0.2210
0.1520
0.0810
0.0400

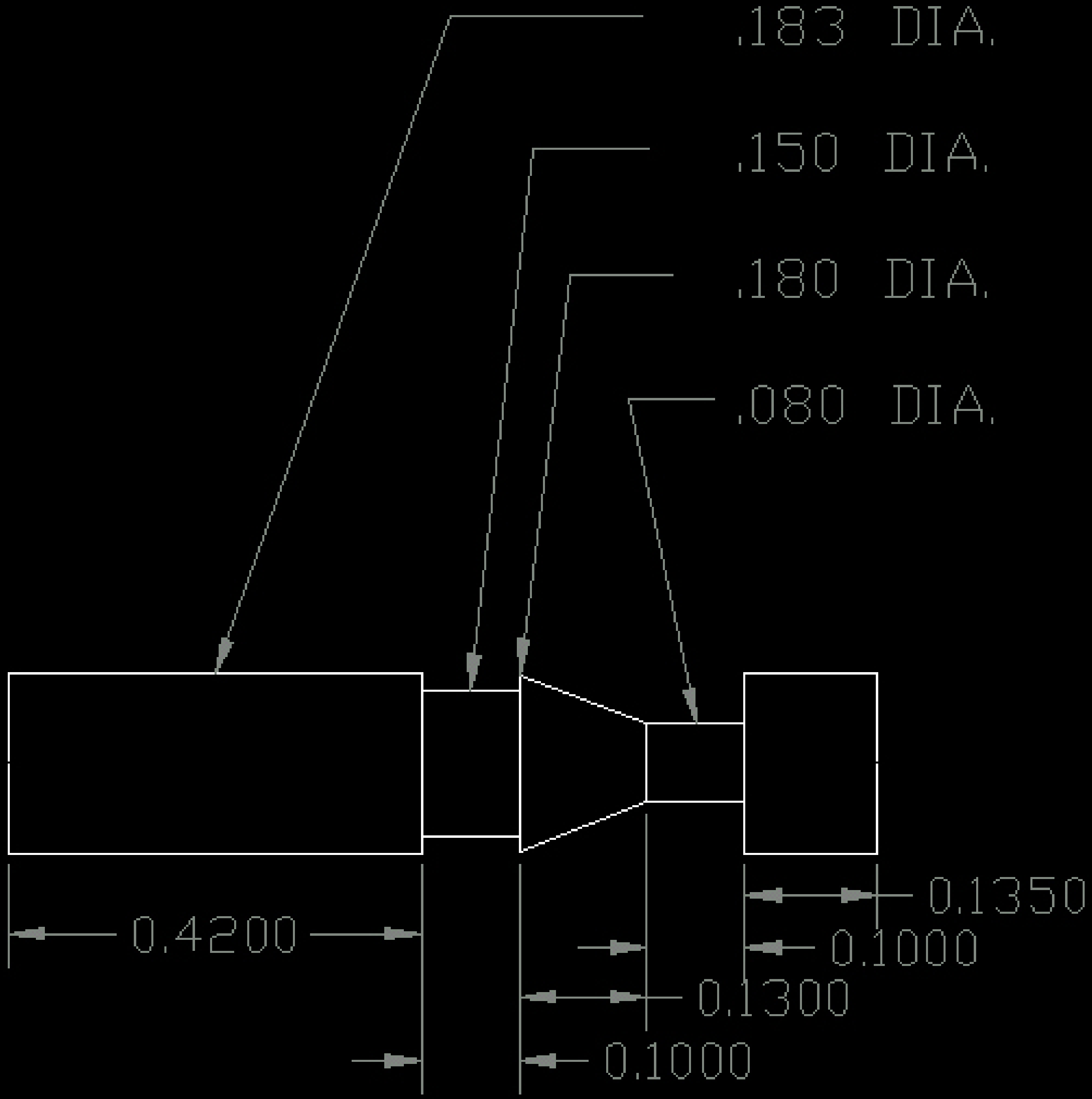
.183 DIA.
.150 DIA.
.180 DIA.
.080 DIA.
0.4200
0.1350
0.1000
0.1300
0.1000

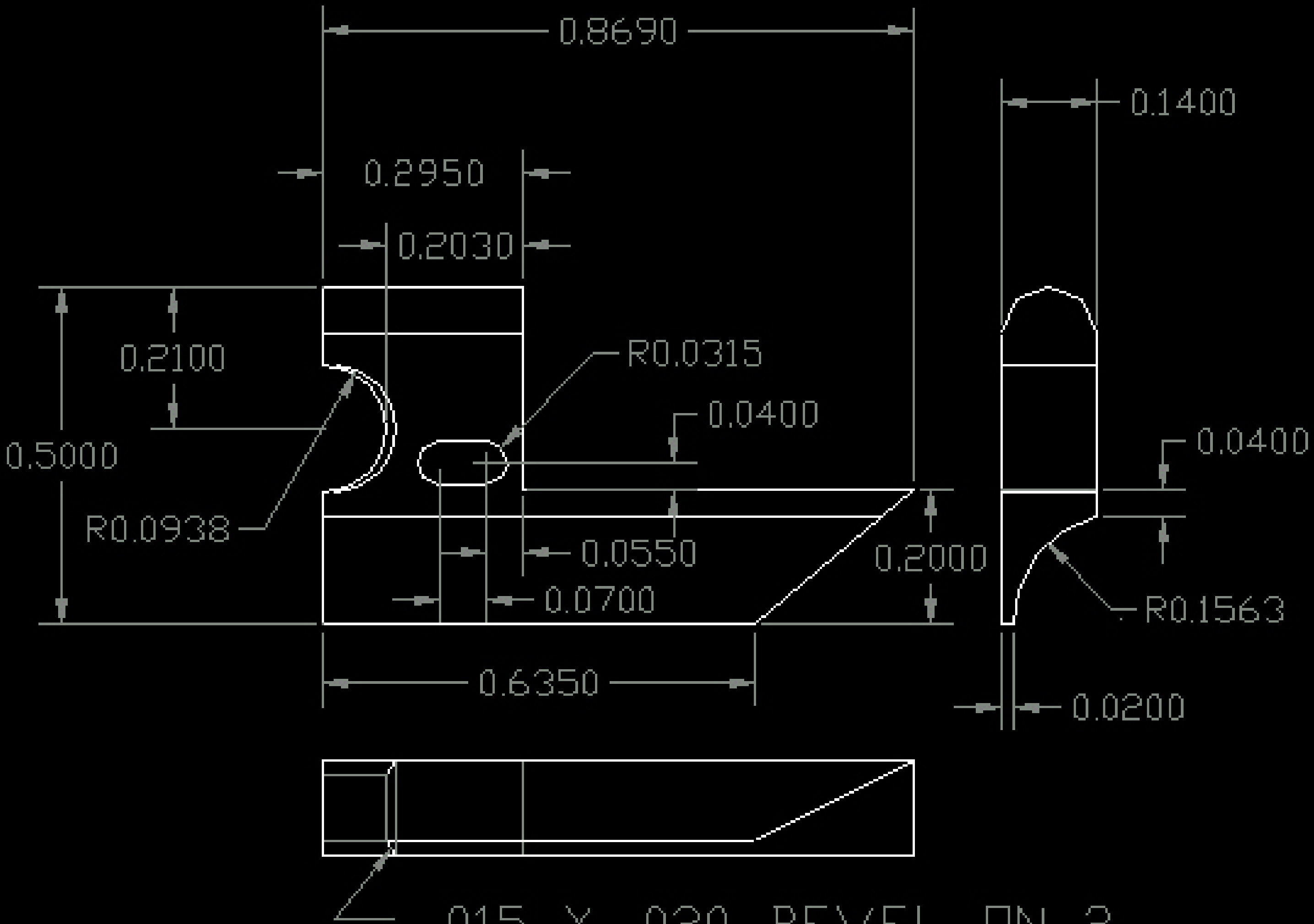
0.8690
0.1400
0.2950
0.2030
0.2100
R0.0315
0.0400
0.0400
0.5000
R0.0938
0.0550
0.2000
0.0700
R0.1563
0.6350
0.0200

.015 X .020 BEVEL ON 2
PLACES OF THE NOTCH.

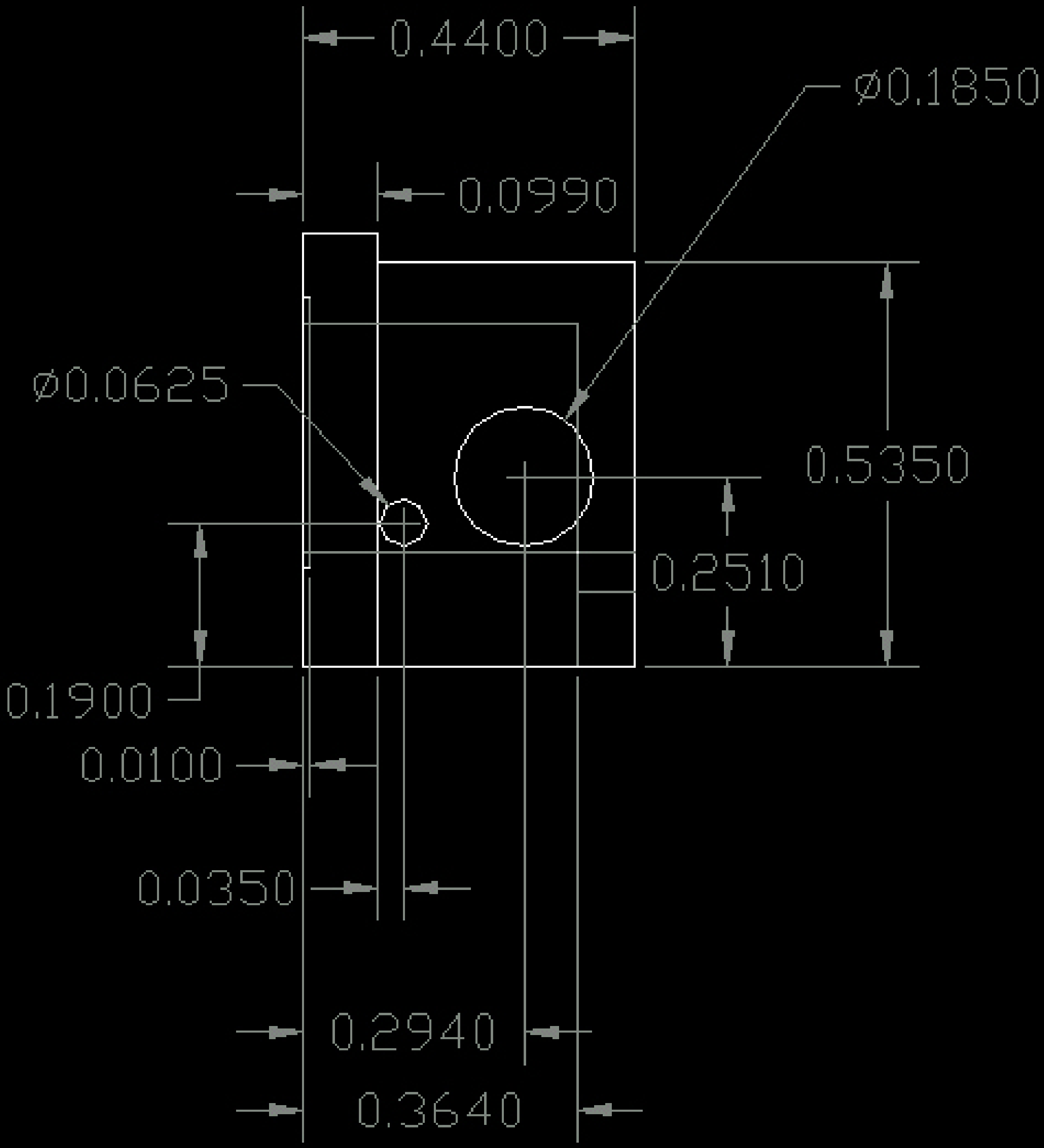
0.4400
Ø0.1850
0.0990
Ø0.0625
0.5350
0.2510
0.1900
0.0100
0.0350
0.2940
0.3640

US 20100207399A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2010/0207399 A1**

Grandics (43) **Pub. Date: Aug. 19, 2010**

(54) **PYRAMID ELECTRIC GENERATOR**

(76) Inventor: **Peter Grandics**, Carlsbad, CA (US)

Correspondence Address:
Law Offices of Michael B. Farber
1902 Wright Place, Suite 200
Carlsbad, CA 92008 (US)

(21) Appl. No.: **12/689,351**

(22) Filed: **Jan. 19, 2010**

Related U.S. Application Data

(63) Continuation-in-part of application No. 12/140,887, filed on Jun. 17, 2008, which is a continuation-in-part of application No. PCT/US07/69804, filed on May 25, 2007.

(60) Provisional application No. 60/818,360, filed on Jul. 3, 2006.

Publication Classification

(51) **Int. Cl.**
H02K 57/00 (2006.01)

(52) **U.S. Cl.** .... **290/1 R**

(57) **ABSTRACT**

A pyramid electric generator for harvesting the vibrational energies of Earth's atomic oscillators according to the present invention comprises: (1) an antenna/waveguide that is geometrically optimized; (2) a coil wound with an insulated conductor on a nonconductive coil form, the coil being connected at its top lead to the conducting surface of the antenna/waveguide such that the coil is connected near the point at which the electric field contacts the antenna/waveguide, the antenna/waveguide serving as a quasi-capacitive series element to provide a specific resonant frequency; (3) the bottom lead of the coil is connected to a driver operating in the LF or ELF bands. In another mode of operation, a secondary coil is positioned coaxially within the first coil acting as a resonant step-up transformer winding. The generator resonantly couples into specific frequencies of Earth's atomic oscillators and extracts electric energy thereof.

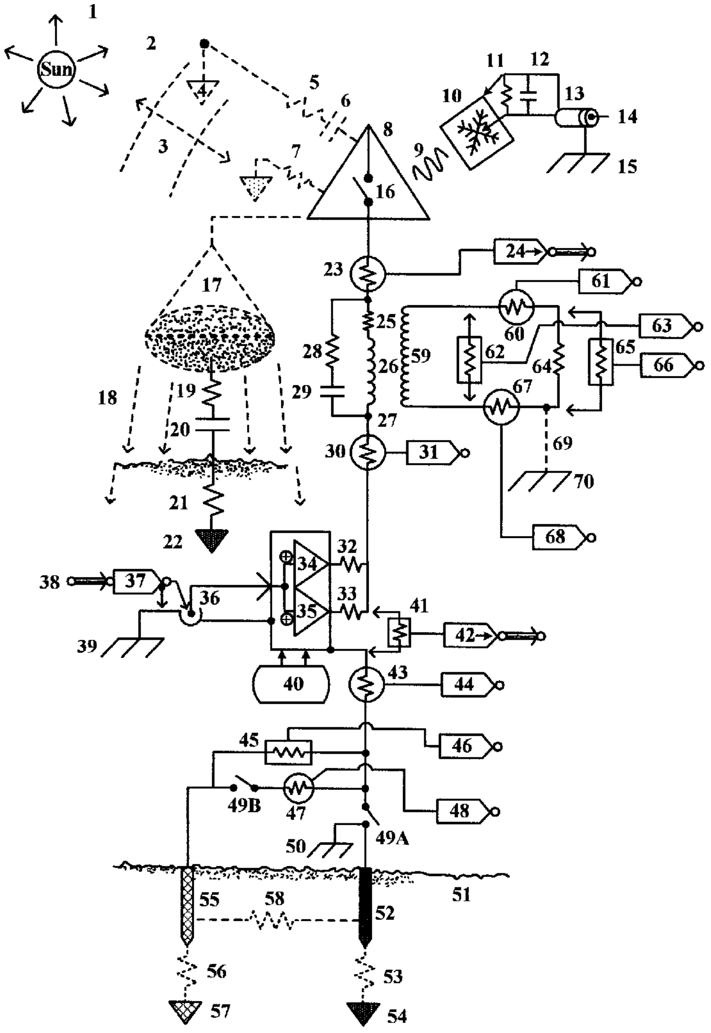

Figure 1.

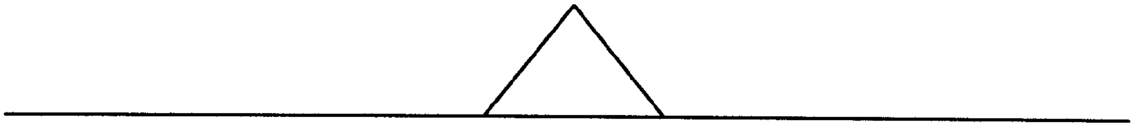

Figure 2.

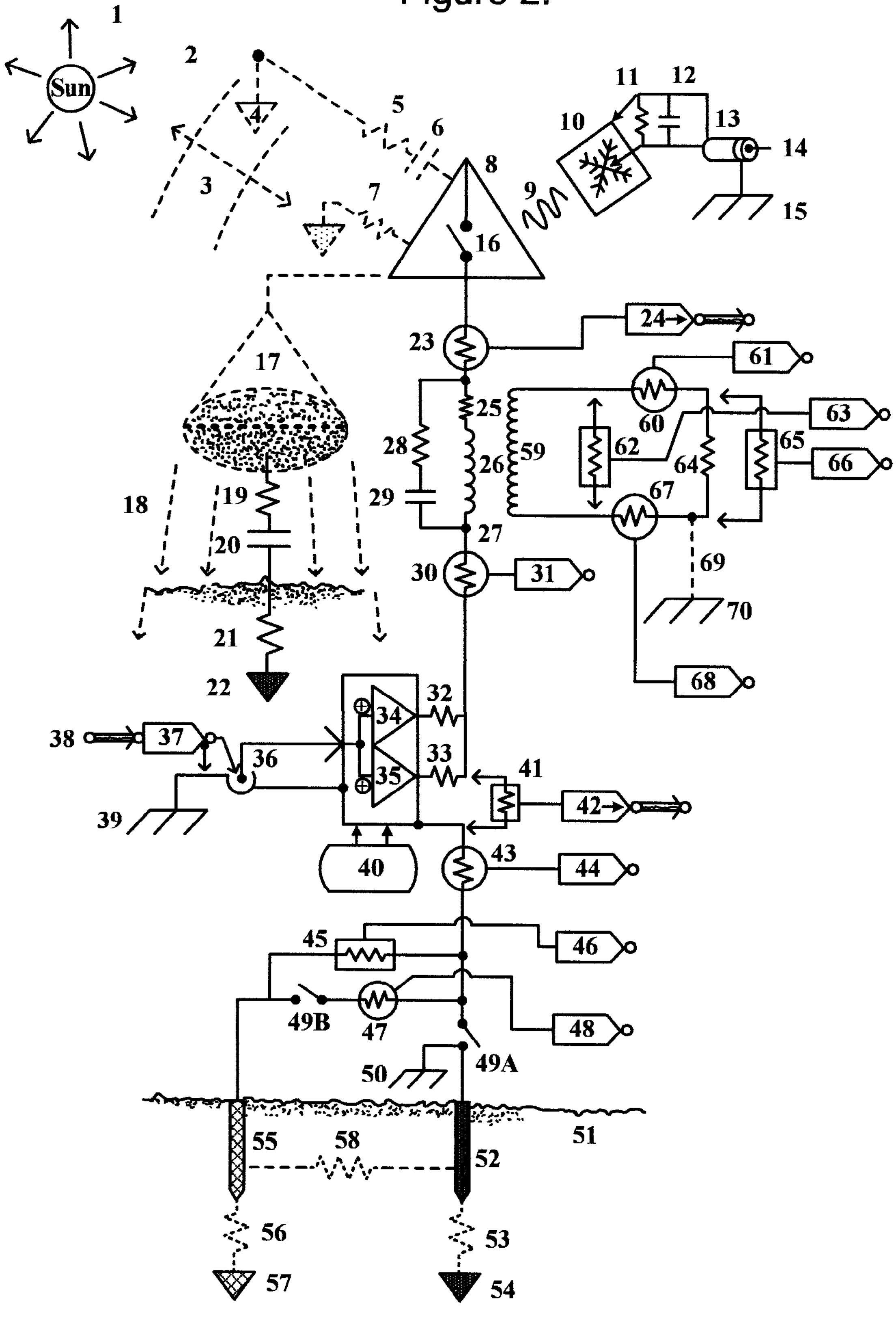
Sun

Figure 3.

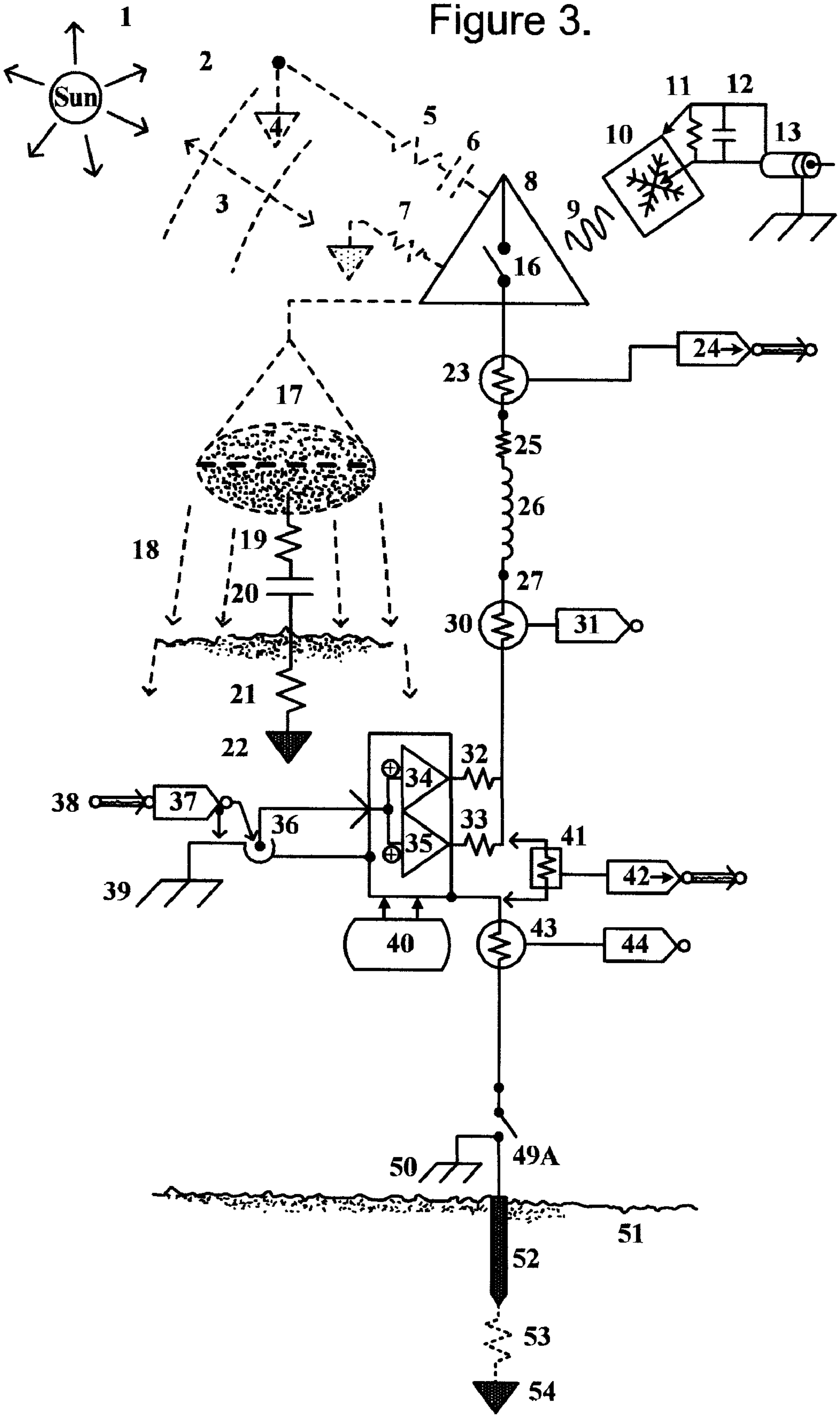
Sun

PYRAMID ELECTRIC GENERATOR

CROSS-REFERENCES

[0001] This application is a continuation-in-part of U.S. patent application Ser. No. 12/140,887 by Grandics, entitled "The Pyramid Electric Generator," and filed on Jun. 17, 2008, which was a continuation-in-part of PCT Application Serial No. PCT/U.S.07/69804 by Grandics, entitled "An Electric Power Converter for Extraction of Atmospheric Electrical Energy," designating the United States and filed on May 25, 2007, which in turn claimed priority from U.S. Provisional Application Ser. No. 60/818,360 by Grandics, also entitled "An Electric Power Converter for Extraction of Atmospheric Electrical Energy," and filed on Jul. 3, 2006. The specifications of these three prior applications are incorporated herein in their entirety by this reference.

BACKGROUND OF THE INVENTION

[0002] This invention relates to the generation of electrical power by drawing energy from Earth's electric field. Earth represents a resonant system in which both atmospheric and telluric electrical processes can be harvested. In U.S. Pat. No. 6,974,110, we have provided an apparatus and method for converting the electrostatic potential energy of Earth's atmosphere into usable power. The present invention aims to provide an electric generator that, in addition to tapping atmospheric electric processes, extracts electrical power from the vibrational energy of atomic oscillators by a novel resonant coupling method.

SUMMARY OF THE INVENTION

[0003] An electrical generator according to the present invention provides a new method of tapping Earth's electric energy field.

[0004] One aspect of the present invention is an apparatus for capturing both atmospheric and telluric electric energies, comprising:

[0005] (1) an antenna/waveguide that is geometrically optimized having a conducting surface; and

[0006] (2) a coil wound with an insulated conductor over a non-conductive coil form, the coil being connected to an AC or AC-DC driver at its bottom lead and, by its top lead to the conducting surface of the antenna/waveguide near the point where the electric field contacts the antenna/waveguide, with the antenna/waveguide serving as a quasi-capacitive series element to provide a specific resonant frequency;

wherein the generator absorbs impulses from Earth's electric oscillations; and

wherein the attracted energy manifests as high voltage sinusoidal waveforms in the coil, representing harvested electrical energy of atomic oscillators and measurable on the leads of the coil.

[0007] Preferably, the antenna/waveguide is of pyramidal shape. When the antenna/waveguide is of pyramidal shape, the apex/primary coil is preferably connected near or at the apex of the antenna/waveguide.

[0008] Another aspect of the invention is an electric generator for harvesting the energies of atomic oscillators comprising:

[0009] (1) an antenna/waveguide that is geometrically optimized having a conducting surface; and

[0010] (2) a primary (apex) coil wound with an insulated conductor over a non-conductive coil form, the coil being connected to an AC or AC-DC driver at its bottom lead and, by its top lead to the conducting surface of the antenna/waveguide near the point where the electric field contacts the antenna/waveguide, and the antenna/waveguide serving as a quasi-capacitive series element to provide a specific resonant frequency; and

[0011] (3) a secondary coil of smaller diameter than the primary (apex) coil having a greater number of turns than the primary coil, the secondary coil being positioned coaxially within the primary coil and acting as a resonant step-up transformer winding inductively coupled with the primary coil;

wherein the generator absorbs impulses from Earth's electric oscillations; and

wherein the attracted energy manifests as high voltage sinusoidal waveforms in the secondary coil, representing harvested electrical energy of atomic oscillators and measurable on the leads of the secondary coil.

[0012] Yet another aspect of the present invention is an electric generator for harvesting the electric emissions of atomic oscillations comprising an oscillator. In this aspect, the electric generator typically comprises:

[0013] (1) a primary (apex) coil wound with a conductor over a secondary coil, the primary coil being connected electrically at its bottom lead to a driver operating in the LF or ELF bands while connected by its top lead near the point at which the electric field contacts the conductive surface of the antenna/waveguide; and

[0014] (2) a secondary coil of smaller diameter than the primary (apex) coil having a greater number of turns than the primary coil, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding inductively coupled with the primary;

wherein said resonant step-up transformer's output voltage exceeds 500 V_{RMS};

wherein the electric generator attracts impulses from Earth's electric oscillations;

and wherein the attracted energy manifests as high voltage sinusoidal waveforms representing harvested electric energy of atomic oscillations in the secondary coil.

[0015] Another aspect of the present invention is a method of tapping Earth's electric energy field, comprising the steps of:

[0016] (1) positioning an electric generator according to the present invention as described above such that it is exposed to Earth's electric oscillations; and

[0017] (2) generating a sinusoidal voltage signal representing harvested electric energy of atomic oscillators by the operation of the generator.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] The following invention will be better understood with reference to the specification, appended claims, and accompanying drawings, where:

[0019] FIG. **1** is a graph showing the formation of electric vortices across and around the pyramid.

[0020] FIG. **2** is a graph showing the circuit diagram and measurement points of the pyramid generator.

[0021] FIG. **3** is a graph showing the minimum circuit diagram and relevant measurement points of the pyramid generator.

DETAILED DESCRIPTION OF THE INVENTION

[0022] One aspect of the invention is an apparatus for capturing atmospheric and telluric electric energies, comprising:

[0023] (1) an antenna/waveguide that is geometrically optimized having a conducting surface; and

[0024] (2) a coil wound with an insulated conductor over a non-conductive coil form, the coil being connected to an AC or AC-DC driver at its bottom lead and, by its top lead to the conducting surface of the antenna/waveguide near the point where the electric field contacts the antenna/waveguide, with the antenna/waveguide serving as a quasi-capacitive series element to provide a specific resonant frequency;

wherein the generator absorbs impulses from Earth's electric oscillations; and

wherein the attracted energy manifests as high voltage sinusoidal waveform in the coil, representing harvested electrical energy of atomic oscillators and measurable on the leads of the coil.

[0025] Typically, the electric generator stimulates and attracts the electric emissions of atomic oscillators by the use of high voltage. Typically, the antenna/waveguide is designed based on functions of phi and/or pi.

[0026] In one alternative, the antenna/waveguide is placed on an insulated base.

[0027] Typically, the antenna/waveguide is at a high voltage during the operation of the generator. Typically, the antenna/waveguide voltage is measured by using an E-field antenna.

[0028] The antenna/waveguide can be connected to an oscillator. The oscillator can be a resonant step-up transformer.

[0029] Preferably, the antenna/waveguide is of pyramidal shape. Preferably, the antenna/waveguide exhibits functions of phi and/or pi. However, any geometric form employing these functions can be used for the antenna/waveguide. When it is of pyramidal shape, the primary (apex) coil is preferably connected near or at the apex of the antenna/waveguide. The shape and size of the antenna/waveguide can be optimized depending on the intended use of the generator.

[0030] Typically, the generator further comprises an insulating base on which the antenna/waveguide is placed. The size and shape of the insulating base can also be optimized, depending on the intended use of the generator.

[0031] Typically, the output of the apex coil-antenna/waveguide series resonant circuit or that of the secondary coil is harvested. One way to do this is by connecting to a step-down transformer and rectifier-capacitor-load resistance. The optimal circuit can be selected by one of ordinary skill in the art to optimize power extraction. The load draws power from the generator.

[0032] The driver can supply AC or a combination of AC-DC power to activate the resonant circuit.

[0033] Typically, the antenna/waveguide is positioned in the general North-South direction. In this arrangement, when the antenna/waveguide is of pyramidal shape, the height of the pyramid can vary from about 0.1 m to about 1000 m. The optimal pyramid size is a function of the power requirements. However, a pyramid can be built as high as 1 km in height. The base surface area of the pyramid may range from about 0.01 m^2 to about 1,000,000 m^2. Naturally, other heights and base surface areas or shapes can be used for particular applications. The numbers of sides can also vary. When the pyramid is small (e.g., the 6.26 feet base length described in this application), there is an insufficient atmospheric potential gradient on its apex to trigger the attraction of energy. In such event, a driver is necessary to bring the pyramid to the electric potential necessary for its harvesting function.

[0034] Another embodiment of the present invention is an electric generator for harvesting the energies of atomic oscillators comprising:

[0035] (1) an antenna/waveguide that is geometrically optimized having a conducting surface; and

[0036] (2) a primary (apex) coil wound with an insulated conductor over a non-conductive coil form, the coil being connected to an AC or AC-DC driver at its bottom lead and, by its top lead to the conducting surface of the antenna/waveguide near the point where the electric field contacts the antenna/waveguide, and the antenna/waveguide serving as a quasi-capacitive series element to provide a specific resonant frequency; and

[0037] (3) a secondary coil of smaller diameter than the primary (apex) coil having a greater number of turns than the primary coil, the secondary coil being positioned coaxially within the primary coil and acting as a resonant step-up transformer winding inductively coupled with the apex coil;

wherein the generator absorbs impulses from Earth's electric oscillations; and

wherein the attracted energy manifests as high voltage sinusoidal waveforms in the secondary coil, representing harvested electrical energy of atomic oscillators and measurable on the leads of the secondary coil.

[0038] In this embodiment, more than one resonant step-up transformer winding can be used. As described above for the first embodiment, in this embodiment, the electric generator stimulates and attracts the electric emissions of atomic oscillators by the use of high voltage. Typically, in this embodiment, the antenna/waveguide is designed based on functions of phi and/or pi. In one alternative, in this embodiment, the antenna/waveguide is placed on an insulated base. Typically, in this embodiment, the generator further comprises an insulating base on which the antenna/waveguide is placed. Typically, in this embodiment, antenna/waveguide is at a high voltage during the operation of the generator. Typically, in this embodiment, the antenna/waveguide voltage is measured by using an E-field antenna. In this embodiment, the antenna/waveguide can be connected to an oscillator. The oscillator can be a resonant step-up transformer. Preferably, in this embodiment, the antenna/waveguide is of pyramidal shape. Preferably, in this embodiment, the antenna/waveguide exhibits functions of phi and/or pi. Typically, in this embodiment, the antenna/waveguide is positioned in the general North-South direction. In this embodiment, in this arrangement, when the antenna/waveguide is of pyramidal shape, the height of the pyramid can vary from about 0.1 m to about 1000 m. The optimal pyramid size is a function of the power requirements. However, a pyramid can be built as high as 1 km in height. In this embodiment, the base surface area of the pyramid may range from about 0.01 m^2 to about 1,000,000 m^2.

[0039] In this embodiment, typically, the primary (apex) coil of the resonant transformer is connected near the point where the electric field contacts the antenna/waveguide.

[0040] In this embodiment, typically, the bottom lead of the secondary coil is connected to ground. In this embodiment, typically, the primary (apex) coil is inductively coupled to the secondary coil.

[0041] In this embodiment, typically, the output of the secondary coil is connected to a load for drawing power from the generator.

[0042] Another embodiment of the present invention is an electric generator for harvesting the electric emissions of atomic oscillations comprising an oscillator.

[0043] In this embodiment, typically, the oscillator operates in the LF or ELF bands.

[0044] Typically, in this embodiment, the electric generator comprises:

[0045] (1) a primary (apex) coil wound with a conductor over a secondary coil, the primary (apex) coil being connected electrically at its bottom lead to a driver operating in the LF or ELF bands while connected by its top lead near the point at which the electric field contacts the conductive surface of the antenna/waveguide; and

[0046] (2) a secondary coil of smaller diameter than the primary coil having a greater number of turns than the primary coil, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding inductively coupled with the primary;

wherein said resonant step-up transformer's output voltage exceeds 500 V_{RMS};

wherein the electric generator attracts impulses from Earth's electric oscillations;

and wherein the attracted energy manifests as high voltage sinusoidal waveforms representing harvested electric energy of atomic oscillations in the secondary coil.

[0047] In this embodiment, typically, the output of the secondary coil is connected to a load for drawing power from the generator.

[0048] Another aspect of the invention is a method of tapping Earth's electric energy based on the following concepts:

[0049] (1) utilizing a pyramidal antenna/waveguide designed based on the functions of phi and/or pi;

[0050] (2) using the antenna/waveguide to absorb and focus specific frequencies of Earth's electric field;

[0051] (3) positioning a coil(s) connected to the pyramidal antenna/waveguide near or at its apex according to the present invention; and

[0052] (4) generating a high-voltage sinusoidal signal, representing harvested atmospheric and telluric electrical energy, by the operation of the system.

[0053] Typically, the harvested electric energy can be fed into a power grid for distribution, but smaller units capable of distributed power generation or for use as stand-alone generators are also feasible by the concept of this invention.

[0054] In this aspect, the method typically comprises a method of tapping Earth's electric oscillatory energies comprising the steps of:

[0055] (1) positioning an electric generator according to the present invention as described above such that it is exposed to Earth's electric oscillations; and

[0056] (2) generating a sinusoidal voltage signal representing harvested electric energy of atomic oscillators by the operation of the generator.

[0057] The invention is described by the following Example. This Example is included for illustrative purposes only and is not intended to limit the invention.

Example

[0058] We have introduced a new theory of space, energy and matter that predicted that electromagnetic interaction propagates through a carrier medium called the space lattice that is made up of pulsating, phi-based spiral vortices (STAR), the excitation of which generates propagating sinusoidal lines of force across the space lattice recorded as electromagnetic waves by our instruments [1,2]. We have also predicted that natural pressure gradients exist within the space lattice that can be used for power generation by inserting a capacitor into such gradient. Furthermore, we proposed that a pyramid-shaped capacitor is an optimally shaped device for tapping the energy gradient of the space lattice that manifests in the form of an electrical potential gradient in Earth's atmosphere. This was subsequently demonstrated [3-5]. We also found that the pyramid-shaped capacitor design should be based on functions of phi and/or pi [5] because electrical energy propagates along a phi-based spiral [6].

[0059] We suggested that elemental particles are composed of STAR oscillators that make up the space lattice by electromagnetic compression of its elemental cubes into pulsating conical/vortexual subunits that form the basis for all particles of matter [1,2]. We proposed that atomic and subatomic oscillators could vibrate in a very wide frequency range from sub-acoustic to extremely high EM frequencies [1,2]. This does not mean that material bodies can take or maintain a physical form at any energy level, but this allows the recycling of atoms into the space lattice as part of their natural evolution [2,6].

[0060] Atomic oscillators are constantly in motion, and such motion is electric because all motion and energy is fundamentally electric [2,6]. It is known that atomic bodies constantly exchange electric energy. Such energy flow can be harvested when the spiral physical nature of electric waves is recognized. Russell observed that energy moves during its generative cycle in spirals towards the higher potential [7]; therefore, we reasoned that an "attractor" of high electric potential must be provided to pull in the random electromagnetic emissions of atomic oscillators. The "attractor" must function as a phi-based antenna/waveguide to focus the phi-based electric emissions of atomic bodies into the apex of a vortex, the same method Nature uses for power multiplication [6,7]. Therefore, a phi-based pyramidal antenna/waveguide must be at a high voltage in order to perform its attractor function.

[0061] To achieve this goal, a geometrically optimized pyramidal antenna/waveguide must be connected to an oscillator that operates at a high voltage and is tuned to a suitably selected frequency.

[0062] The appropriate resonant frequency is dictated by practical considerations. Atomic oscillators can vibrate in a very wide frequency range [1, 2, 7, 8] but short wavelength radiations are readily re-absorbed by atoms and are rapidly attenuated; therefore, long wavelength electromagnetic emissions should be targeted. The operation of the pyramid generator is most practical in the ELF to the LF range. As electric radiation propagates in a vortex [2,7,8], the emitted electric vortex over the pyramidal antenna (FIG. **1**) increases its "size," and so it is capable of funneling atmospheric electromagnetic emissions of the same frequency back into the antenna/receiver. Thus, the pyramid electric generator could capture the energy emitted by both telluric and atmospheric atomic oscillators. We have tested these assumptions and demonstrated the basic principles of an electric generator.

Methods and Results

[0063] In our previous paper, we reported on a pyramid-shaped capacitor/antenna connected in series to a coil that form a resonant circuit for harvesting an expanded range of atmospheric and telluric electric energies [6]. As the signals in the pyramid energy harvester system are at extremely high voltages, we have developed a measurement and control system that allows secure remote transmission of test signals with great accuracy while removing all the hazards associated with direct connection of ultra-high voltages (over 50 kV AC) to the measurement and data-logging system. The voltage and current signals measured across a resistor are turned into a proportional electronic signal, which is then fed to a light emitting diode (opto transmitter) and sent through an optical cable to the instrumentation part of the system where it is converted back by an opto receiver module into an electronic signal for acquisition and processing by the National Instrument PXI 1042Q controller programmed with Labview 6. An ENI-1140LA amplifier was used to activate the oscillator at the resonant frequency.

[0064] For the experiments, we used a novel tetrahedral pyramid-shaped antenna/waveguide that expressed the golden mean in its ratios. The base length of the pyramidal antenna was 6.26 feet. The pyramid was built of a wooden frame and covered with 26 AWG triangular copper sheets on its sides that were electrically connected. It was placed on an insulating base and positioned in the general North-South direction.

[0065] The apex coil was wound on a HDPE cylindrical coil form using a 14 AWG insulated magnet wire with a monofilament spacer polyester twine (0.05" diameter) between the turns, and connected to the conducting surface of the pyramid near its apex. Thus, the apex coil is connected to the pyramidal antenna/waveguide acting as a quasi-capacitive series element to provide a specific resonant frequency by coupling to the ambient electromagnetic environment. The bottom lead of the apex coil was connected to a driver and grounded via the driver (FIG. **2**). The purpose of the monofilament spacer is to reduce interwinding capacitance in the apex coil by increasing distance between wire turns of the coil.

[0066] The parameters of the apex coil are shown in Table 1.

TABLE 1

Physical and electrical parameters of the apex coil

Number of turns of wire	290	
Coil height	0.87	m
Coil diameter	0.787	m
Wire diameter	1.6277	mm
Inductance	42.86	mH
Resistance	6.5	ohms

[0067] In the experiment, the controlled variable was the voltage, which was increased in steps by the controller until a maximum voltage in the system (permitted by circuit components and test equipment) was achieved for the given resonant frequency. The objective was to "drive the atmosphere and ground" by radiating a high voltage (~200 kV p-p), LF (~70 kHz) signal fed by a powerful signal generator at a resonant frequency determined by the pyramid as a quasi-capacitive series element connected to the apex coil and radiating to the local atmosphere and the ground.

[0068] The experimental data are shown in Table 2. An apparent power gain of nearly 170-fold was achieved in the resonant circuit relative to input power from the driver.

TABLE 2

Power measurements in the resonant pyramid-coil system.

Frequency (kHz)	Drive voltage (rms)	Drive current (rms)	Drive power (VA)	Drive power (W)
70.9	321.1 V	3.2 A	1125.54	735.5
Frequency (kHz)	Apex voltage (rms)	Apex current (rms)	Apex power (kVA)	Apex power (kW)
70.9	67.922 kV	2.903 A	196.621	40.393

[0069] A load may be connected to the system to draw power from it. The load may also be connected via a secondary coil inductively coupled to the apex coil. The load may be a resistor, a rectifier or storage capacitor powering a DC load. The high voltage, high frequency output can also be converted to the frequency and voltage of standard mains power.

[0070] A graphic representation of the system is shown in FIG. **2**. The pyramidal antenna/waveguide, placed on an insulating base, is coupled to the ambient electromagnetic environment and serves as an antenna/waveguide for the harvesting of atmospheric and telluric electromagnetic oscillations at resonance frequency (FIG. **1**). The atmospheric electric vortex possibly acts as an ionic antenna pulling in EM radiations from a large atmospheric domain.

[0071] The pyramid's apex is connected to a coil of high turn number (FIG. **2**). This apex coil wound on a nonconductive coil form forms a resonant circuit with the pyramid that acts as a quasi-capacitive series element. An RF power amplifier activates the apex coil. The measurement points are also indicated, displaying the positions of voltage and current sense resistors and opto-electric transmitters. A detailed description of the system follows in FIG. **2**.

[0072] The source of radiative and electromagnetic power to earth is the Sun **1**, which is harvested by effect of pyramid operations. Solar radiation is separated into light, heat, electromagnetic waves **2** by interaction with atmosphere, magnetosphere, ionosphere. "Electrosphere" **3** comprises of ionized, highly conductive air. It can be understood as a conductive spherical band about 60 mi above the surface of the earth. Virtual "Ground" **4** in electrosphere—a low impedance spherical area that in the circuit may be considered as a "ground" (for non-harvesting calculations) or a variable circuit source (for harvesting calculations). Effective resistive impedance **5** is from pyramid surface to "sky" or "electrosphere" (radiative impedance). Effective capacitive impedance **6** is from pyramid surface to "sky" or "electrosphere". Effective resistive (conductive) impedance **7** is from pyramid external surface to nearby atmosphere due to "corona" type ionic conduction through ionized air, caused by high peak voltages on the pyramid's surface. Conductive pyramidal transducer/energy harvester **8** is coupled to the ambient electromagnetic environment as above. Radiated "near field" electromagnetic field **9** from pyramid surface (toward "sky") is measured by E-field antenna **10**. The antenna does indirect measurement of pyramid surface voltage based on E-field pick-up of alternating current field within ~20 ft of pyramid surface, at a particular fixed distance and orientation. Impedance-matching resistor **11** couples E-field antenna to coaxial cable and impedance-matching capacitor **12** for coupling E-field antenna to coaxial cable. Coaxial cable **13** connects to measurement system interface with a "BNC" connector **14** to provide electrical interface to measurement system. System "ground" (common connection point) **15** is for measurement/control system components. Switch **16** is for disconnection of pyramid/antenna from pyramid driving electronics. Radiative "cone" **17** emits from inside and bottom surface of pyramid towards "ground" underneath the pyramid harvester system. Radiated E-M field **18** is directed from pyramid to ground surface directly underneath. Resistive impedance **19** of radiated E-M field ("virtual" component) is derived from measurement and calculation of E-M field model to ground area. Capacitive impedance **20** of radiated E-M field to "ground" area is a "virtual component" derived from measurement and calculation of E-M field model. Virtual resistance of radiated energy **21** from bottom of pyramid, is conducted through earth to "firm ground" potential (another "virtual component" derived from measurement and calculation). Pyramid ground **52** is an equipotential "zero-impedance" point in the earth which serves as the ground reference for both conducted (from pyramid ground stake **52**) and radiated **18**, **19**, **20**, **21** energy underneath the pyramid harvester device. In a non-harvesting model this ground point is always at zero potential. However, in an energy-harvesting model the point **22** serves as an AC voltage source, sending conducted energy back into the pyramid system through all the components that "drain" to it in the non-harvesting model. Sense resistor **23** for apex current monitor in pyramid energizing system is a low value resistor (typically 0.05 to 0.2 ohms) that allows a small voltage to develop, proportional to the current flowing to the pyramid apex from the energizing system. Apex current monitor opto transmitter **24** measures the voltage across resistor **23** and turns it into a proportional electronic signal, which is then fed to a light emitting diode and transmitted through an optical cable to the instrumentation part of the system. All the signals in the pyramid harvester system are at extremely high voltages, and the optical cable allows secure remote transmission of these signals to the measurement and data-logging system (not shown on this diagram). Resistance **25** of the apex excitation coil **26** is measured at DC. This is a "virtual" component, since the resistance is in the coil itself, and is not a separate component. Apex coil **26** translates the stimulation power signal from amplifiers **34**, **35** into a very high voltage at the resonant frequency established by the coil and intrinsic capacitance of the pyramid. Mutual inductance coupling exists between the apex coil and the harvest or secondary coil **27**. This is another "virtual component" consisting of the electromagnetic signal coupling between the two coils based on many different factors. Optional resistor **28** can be used for resonance or phase tuning of pyramid network. The main resonance is determined by apex coil **26** inductance and pyramid lumped capacitive element **8**. However, this will make the voltage at the junction of the pyramid (capacitor) and the apex coil **26** approximately 90 degrees out of phase with the current. One approach to reduce the phase angle and also to possibly tune the resonant frequency and the Q of the circuit is to insert an additional reactive element **29** in parallel into the apex coil **26**. The apex coil winding resistance is expressed here as resistor **25** in series with inductor **26**. The inserted element is an additional optional capacitor **29** in parallel with the apex coil. This capacitor develops an independent L-C resonant circuit, besides the L-C resonant circuit of the apex coil and pyramid. An anticipated function if used in the circuit would be to lower the resonant frequency. Drive current monitor current sense resistor **30** is for drive current opto transmitter **31**. Summing/protection resistors are placed in series with **32** drive **1** and **33** drive **2**. Power amplifiers **34**, **35** drive coil/pyramid resonant system. Power (drive) amplifier signal input **36** is fed through opto receiver **37**. The signal in the optical cable **38** is from the National Instruments control system for control of power amplifiers. Pyramid system ground (star-ground) connects to top of pyramid ground stake, after ground stake disconnect switches **39** and **50**. Power supply **40** is for pyramid driver amplifier. Resistor **41** is used for voltage measurement across power driver higher value terminals. Opto transmitter **42** is for voltage measurement across power amplifier and includes drawing of optical cable to measurement system. Current-sensing resistor **43** is for current measurement from pyramid amplifiers to ground post. Opto transmitter **44** is for transmitting ground current to control and measurement system. Voltage sense resistor **45** is between pyramid ground stake **52** and remote ground stake **55** (100 kohm-1 Mohm). Opto transmitter **46** is for voltage sense from pyramid ground **52** to remote ground **55**. Current sense resistor **47** from pyramid ground to remote ground is for opto transmitter **48**. When used, this closes switch **49**B to short out the voltage measurement **45** and also short out the intrinsic ground impedance **58** between pyramid ground and remote ground stakes. Ground post disconnect switch **49** is below the system star ground stake and is very dangerous when open. Main connection point for pyramid system is the star ground terminal **50**. Ground surface **51** is on which pyramid system is placed. Ground stake **52** for pyramid system is a 1" diameter copper rod, 8-12 ft deep. R absolute **53** is between ground stake and earth integrated from ground stake theoretically through entire earth. "Absolute ground" is located under pyramid and ground stake **54**. Remote ground stake **55** is about 200 feet away from pyramid system and is used as remote reference ground although a real remote ground would likely be much further away. R absolute **56** is from remote ground stake to absolute ground under remote stakes **57**. Ground impedance **58** between **52** and **55** stakes appears mostly resistive and is about 28-30 ohms based on AC measurements at 1 KHz. The remote ground setup is not required for the operation of the system but provides test data. Harvest/secondary coil **59** is an inductor wound with 3 to 6 times as many turns as the apex coil, and connected to it by mutual inductance **27**. Current sense resistor **60** is placed on top lead of harvest coil. Opto transmitter **61** is for top harvest coil current. Voltage sense resistor **62** that measures voltage across harvest coil is for opto transmitter **63**. Load resistor **64** is for secondary/harvest coil and the load resistor is optimized to draw the maximum power from the harvest/secondary coil such that the energy harvest efficiency of the entire system is optimized. This is the defining element of the system. Voltage sense resistor **65** for the opto transmitter for load voltage **66** measures the voltage across the load. Current sense resistor **67** is placed at bottom side of harvest coil. Opto transmitter **68** is for bottom lead current sense resistor **67**. A jumper is placed to system ground **69** that allows grounded operation of the harvest coil with system ground **70**. More than one step-up transformer coil can be used as energy harvester in the pyramid generator system.

[0073] The minimum system is shown in FIG. **3**. When a minimum system is used, a load can be directly connected to the pyramid's surface.

DISCUSSION

[0074] This invention demonstrates a novel approach to harvest Earth's electric energy. The masses of Earth and its atmosphere represent a coupled resonant system that is continually electrified by solar radiation. We have found that a pyramidal antenna designed based on the functions of phi and/or pi is optimal for the capture of atmospheric electrostatic discharge (ESD) impulses [3-5]. Atmospheric ESD is a product of solar radiation and is a wide bandwidth phenomenon. The pyramid as a wideband, non-resonant antenna is uniquely adapted to harvest the energy of atmospheric ESD.

[0075] This invention also points out an additional energizing mechanism at work during the operation of the pyramid electric generator. The observation is derived from our theory on energy, matter and space [1,2] as well as that of Russell [7]. Both Russell and we observed that spiral motion is a fundamental action of matter and that the vortex is the mechanism of power multiplication in Nature. A vortex concentrates power into its apex where the highest velocity of motion, the highest pressure and the highest electric potential resides [1, 2, 6, 7]. As electricity propagates along a pulsating phi-based spiral [6,7], a phi-based antenna/waveguide is suitable to focus its energy into the apex of a vortex inside the antenna. A phi-based pyramid is optimal for this purpose.

[0076] As atomic bodies can oscillate in a very wide frequency range [1, 2, 7, 8], a pyramidal antenna/waveguide coupled to an oscillator that is tuned to a suitable frequency can focus the "electric noise" from Earth's atomic oscillators into the apex of the pyramid at a high voltage. Russell observed that energy moves towards the higher potential during its generative cycle [7]; therefore, it is essential that the pyramidal antenna be at a high potential. Our results corroborated this assumption.

[0077] For practical reasons on our small pyramid, we have chosen a frequency above the audio range but below the radio frequency spectrum; our antenna was made to emit at the target wavelength. The wavelength of the 50-120 kHz range is 6 km to 25 km, thus allowing energy to be harvested over a large atmospheric domain. As electrical energy propagates in a vortex [6,7], the emitted electric vortex (ionic antenna) over the pyramidal antenna/waveguide increases its "size" and enables it to funnel the same frequency atmospheric electromagnetic radiations back into the antenna. Thus, the pyramid electric generator captures at the selected frequency the energy emitted by Earth's atomic oscillators.

[0078] The presence of an electric (ionic) vortex was demonstrated by radar testing of the space over a 44 m tall fiberglass pyramid located near Moscow [12]. The Russians found that the large ionized column of air over the area of the vertical axis of the pyramid had a width of about 500 m and reached an altitude of 2 km. It is remarkable that this effect was induced by a nonconductive pyramid surface and demonstrating a significant degree of atmospheric ionization even under fair weather conditions. Thus, a suitably sized large or an electrically activated small pyramid should open a low impedance path to higher elevations of relatively conductive atmospheric domains.

[0079] In conclusion, the total power that can be extracted from Earth's atomic oscillators must be extremely large, likely far exceeding current global electric generation capacity. In our experiment, we measured over 196 kVA power at the pyramid's apex at about 200 kV p-p voltage using a 6.26 feet base length pyramid. This power is 175 times greater than the power necessary to drive the pyramid to this operating voltage. By constructing the pyramid generator from appropriate high voltage components and using higher drive voltages, a significant increase in power output is envisioned. The power extraction will also be enhanced by further optimization of the designs of both the pyramid and the coil system.

[0080] The power output will also increase by employing larger pyramid structures and coils. As voltage is the primary factor in attracting power into the system, the necessary voltage can be provided by the vertical atmospheric potential on a tall pyramid. Since the atmospheric vertical potential gradient could go as high as 1200 V/m near Earth's surface under fair weather conditions [13], a pyramid height of 20-150 m seems sufficient to provide the apex voltage for a self-sustaining power generator. As the pyramid scales up volumetrically, a power generator pyramid of the size range of the GPG could likely have an output in the range of hundreds of gigawatts. Groups of several pyramid electric generators could be placed within specific geographical areas, thus combining their energy harvesting capacity.

REFERENCES

[0081] The following references are specifically applicable to the Example and are incorporated herein in their entirety by this reference; these references are referenced in the Example by the reference numbers assigned to them.

[0082] [1] Grandics, P. 2002, "The genesis of electromagnetic and gravitational forces." *J. New Energy,* 6, (3) 33-45.

[0083] [2] Grandics, P. 2007, "The genesis of fundamental forces acting at a distance and some practical derivations." *Infinite Energy,* 12, (71) 13-24.

[0084] [3] Grandics, P. 2000, "A method to capture atmospheric electrostatic energy", in *Proceedings of IEJ-ESA Joint Symposium on Electrostatics*, Kyoto University, Kyoto, Japan, pp. 355-361.

[0085] [4] Grandics, P. 2006, "A DC to RF converter for the capture of atmospheric electrostatic energy", in *Proceedings of the 5th Conference of the Societe Francaise D'Electrostatique, SFE* 2006, Grenoble, France pp. 279-284.

[0086] [5] Grandics, P. 2007, "Pyramidal Electric Transducer: A DC to RF Converter for the Capture of Atmospheric Electrostatic Energy." *Infinite Energy,* 13, (73) 20-27.

[0087] [6] Grandics, P. 2009, "The pyramid electric generator." *Infinite Energy,* 14, (84) 55-59.

[0088] [7] Russell, W. 1947, in *The Secret of Light,* 3rd ed. University of Science and Philosophy, Waynesboro, Va., pp. 218-288.

[0089] [8] Russell, W. 1957, in *Atomic Suicide?* 2nd ed. University of Science and Philosophy, Waynesboro, Va., pp. 129-153.

[0090] [9] Nishida, K, Kobayashi, N and Fukao, Y. 2000, "Resonant oscillations between the solid earth and the atmosphere", *Science* 287, (5461), 2244-2246.

[0091] [10] Tanimoto, T, Um, J, Nishida, K, and Kobayashi, N. 1998, "Earth's continuous oscillations observed on seismically quiet days", *Geophys Res Lett.* 25, 1553-1556.

[0092] [11] Tanimoto, T. 2001, "Continuous free oscillations: Atmosphere-solid earth coupling". *Ann Rev Earth Planet Sci.* 29, 563-584.

[0093] [12] http://www.pyramidoflife.com/eng/tests_experiments.html.

[0094] [13] Marshall, T C, Rust, W D, Stolzenburg, M, Roeder, WP, and Kriebel, P R. 1999, "A study of enhanced

fair-weather electric fields occurring soon after sunrise". *J Geophys Res.* 104, 24,455-24,469.

ADVANTAGES OF THE INVENTION

[0095] The present invention provides a new method of tapping Earth's electric energy pool and providing usable power that can be fed into the power grid or supplied by stand-alone power generators. It does so without the need for mechanical energy or the consumption of fossil fuel or the long-term risks associated with power generated by nuclear fission, including the risk of diversion of fissionable material to terrorist aims or the risk posed by the required long-term storage of spent nuclear fuel. Devices according to the present invention can operate virtually continuously with little to no maintenance.

[0096] Articles of manufacture and methods according to the present invention possess industrial applicability for the generation, storage, and transmission of electrical energy.

[0097] With respect to ranges of values, the invention encompasses each intervening value between the upper and lower limits of the range to at least a tenth of the lower limit's unit, unless the context clearly indicates otherwise. Moreover, the invention encompasses any other stated intervening values and ranges including either or both of the upper and lower limits of the range, unless specifically excluded from the stated range.

[0098] Unless defined otherwise, the meanings of all technical and scientific terms used herein are those commonly understood by one of ordinary skill in the art to which this invention belongs. One of ordinary skill in the art will also appreciate that any methods and materials similar or equivalent to those described herein can also be used to practice or test this invention.

[0099] The publications and patents discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

[0100] All the publications cited are incorporated herein by reference in their entireties, including all published patents, patent applications, literature references, as well as those publications that have been incorporated in those published documents. However, to the extent that any publication incorporated herein by reference refers to information to be published, applicants do not admit that any such information published after the filing date of this application to be prior art.

[0101] As used in this specification and in the appended claims, the singular forms include the plural forms. For example the terms "a," "an," and "the" include plural references unless the content clearly dictates otherwise. Additionally, the term "at least" preceding a series of elements is to be understood as referring to every element in the series. The inventions illustratively described herein can suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms "comprising," "including," "containing," etc. shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the future shown and described or any portion thereof, and it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the inventions herein disclosed can be resorted by those skilled in the art, and that such modifications and variations are considered to be within the scope of the inventions disclosed herein. The inventions have been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the scope of the generic disclosure also form part of these inventions. This includes the generic description of each invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised materials specifically resided therein. In addition, where features or aspects of an invention are described in terms of the Markush group, those schooled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group. It is also to be understood that the above description is intended to be illustrative and not restrictive. Many embodiments will be apparent to those of in the art upon reviewing the above description. The scope of the invention should therefore, be determined not with reference to the above description, but should instead be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled. Those skilled in the art will recognize, or will be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described. Such equivalents are intended to be encompassed by the following claims.

I claim:

1. An electric generator for harvesting the energies of atomic oscillators comprising:

(a) an antenna/waveguide that is geometrically optimized having a conducting surface; and

(b) a coil wound with an insulated conductor over a non-conductive coil form, the coil being connected to an AC or AC-DC driver at its bottom lead and, to the conducting surface of the antenna/waveguide by its top lead near the point where the electric field contacts the antenna/waveguide, with the antenna/waveguide serving as a quasi-capacitive series element to provide a specific resonant frequency;

wherein the generator absorbs impulses from Earth's electric oscillations; and

wherein the attracted energy manifests as high voltage sinusoidal waveforms in the coil, representing harvested electrical energy of atomic oscillators and measurable on the leads of the coil.

2. The electric generator of claim **1**, wherein the electric generator stimulates and attracts the electric emissions of atomic oscillators by the use of high voltage.

3. The electric generator of claim **1**, wherein the antenna/waveguide is designed based on functions of phi and/or pi.

4. The electric generator of claim **1**, wherein the antenna/waveguide is of pyramidal shape.

5. The electric generator of claim **1**, wherein the antenna/waveguide is placed on an insulating base.

6. The electric generator of claim **1**, wherein the antenna/waveguide is at a high voltage during the operation of the generator.

7. The electric generator of claim **1**, wherein the antenna/waveguide voltage is measured by using an E-field antenna.

8. The electric generator of claim **1**, wherein the antenna/waveguide is connected to an oscillator.

9. The electric generator of claim **8**, wherein the oscillator comprises a resonant step-up transformer.

10. The electric generator of claim **1**, wherein the height of the pyramid of the electric generator is from about 0.10 m to about 1000 m.

11. The electric generator of claim **1**, wherein the base surface area of the pyramid is from about 0.01 m^2 to about 1,000,000 m^2.

12. The electric generator of claim **1**, wherein the generator is positioned in the general North-South direction.

13. An electric generator for harvesting the energies of atomic oscillators comprising:

(a) an antenna/waveguide that is geometrically optimized having a conducting surface; and

(b) a primary (apex) coil wound with an insulated conductor over a non-conductive coil form, the coil being connected to an AC or AC-DC driver at its bottom lead and, to the conducting surface of the antenna/waveguide by its top lead near the point where the electric field contacts the antenna/waveguide, and the antenna/waveguide serving as a quasi-capacitive series element to provide a specific resonant frequency; and

(c) a secondary coil of smaller diameter than the primary (apex) coil having a greater number of turns than the primary coil, the secondary coil being positioned coaxially within the primary coil and acting as a resonant step-up transformer winding inductively coupled with the apex coil;

wherein the generator absorbs impulses from Earth's electric oscillations; and

wherein the attracted energy manifests as high voltage sinusoidal waveforms in the secondary coil, representing harvested electrical energy of atomic oscillators and measurable on the leads of the secondary coil.

14. The electric generator of claim **13**, in which more than one resonant step-up transformer winding is used.

15. The electric generator of claim **13**, wherein the electric generator stimulates and attracts the electric emissions of atomic oscillators by the use of high voltage.

16. The electric generator of claim **15**, wherein the high voltage is greater than 500 V_{RMS}.

17. The electric generator of claim **13**, wherein the antenna/waveguide is designed based on functions of phi and/or pi.

18. The electric generator of claim **13**, wherein the antenna/waveguide is of pyramidal shape.

19. The electric generator of claim **13**, wherein the antenna/waveguide is placed on an insulating base.

20. The electric generator of claim **13**, wherein the antenna/waveguide is at a high voltage during the operation of the generator.

21. The electric generator of claim **13**, wherein the antenna/waveguide voltage is measured by using an E-field antenna.

22. The electric generator of claim **13**, wherein the antenna/waveguide is connected to an oscillator.

23. The electric generator of claim **22**, wherein the oscillator comprises a resonant step-up transformer.

24. The electric generator of claim **13**, wherein the primary (apex) coil of the resonant transformer is connected near the point where the electric field contacts the antenna/waveguide.

25. The electric generator of claim **13**, wherein the bottom lead of the secondary coil is connected to ground.

26. The electric generator of claim **13**, wherein the primary (apex) coil is inductively coupled to the secondary coil.

27. The electric generator of claim **13**, wherein the output of the secondary coil is connected to a load.

28. The electric generator of claim **13**, wherein the height of the pyramid of the electric generator is from about 0.10 m to about 1000 m.

29. The electric generator of claim **13**, wherein the base surface area of the pyramid is from about 0.01 m^2 to about 1,000,000 m^2.

30. The electric generator of claim **13**, wherein the generator is positioned in the general North-South direction.

31. An electric generator for harvesting the electric emissions of atomic oscillations comprising an oscillator.

32. The electric generator of claim **31**, wherein the oscillator operates in the LF or ELF bands.

33. The electric generator of claim **31** wherein the electric generator comprises:

(a) a primary (apex) coil wound with a conductor over a secondary coil, the primary (apex) coil being connected electrically at its bottom lead to a driver operating in the LF or ELF bands while connected by its top lead near the point at which the electric field contacts the conductive surface of the antenna/waveguide; and

(b) a secondary coil of smaller diameter than the primary coil having a greater number of turns than the primary coil, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding inductively coupled with the primary;

wherein said resonant step-up transformer's output voltage exceeds 500 V_{RMS};

wherein the electric generator attracts impulses from Earth's electric oscillations;

and wherein the attracted energy manifests as high voltage sinusoidal waveforms representing harvested electric energy of atomic oscillations in the secondary coil.

34. The electric generator of claim **33**, wherein the output of the secondary coil is connected to a load for drawing power from the generator.

35. A method of tapping Earth's electric oscillatory energies comprising the steps of:

(a) positioning the electric generator of claim **1** such that it is exposed to Earth's electric oscillations; and

(b) generating a sinusoidal voltage signal representing harvested electric energy of atomic oscillators by the operation of the generator.

36. A method of tapping Earth's electric oscillatory energies comprising the steps of:

(a) positioning the electric generator of claim **2** such that it is exposed to Earth's electric oscillations; and

(b) generating a sinusoidal voltage signal representing harvested electric energy of atomic oscillators by the operation of the generator.

37. A method of tapping Earth's electric oscillatory energies comprising the steps of:

(a) positioning the electric generator of claim **32** such that it is exposed to Earth's electric oscillations; and

(b) generating a sinusoidal voltage signal representing harvested electric energy of atomic oscillators by the operation of the generator.

* * * * *

US008004250B2

(12) **United States Patent**
Grandics

(10) **Patent No.: US 8,004,250 B2**
(45) **Date of Patent: Aug. 23, 2011**

(54) **PYRAMID ELECTRIC GENERATOR**

(76) Inventor: **Peter Grandics**, Carlsbad, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 476 days.

(21) Appl. No.: **12/140,887**

(22) Filed: **Jun. 17, 2008**

(65) **Prior Publication Data**

US 2008/0315587 A1 Dec. 25, 2008

Related U.S. Application Data

(63) Continuation-in-part of application No. PCT/US2007/069804, filed on May 25, 2007.

(60) Provisional application No. 60/818,360, filed on Jul. 3, 2006.

(51) **Int. Cl.**
H02N 1/00 (2006.01)

(52) **U.S. Cl.** .... **322/27**; 290/1 R; 322/32; 322/2 A

(58) **Field of Classification Search** .... 372/32, 372/123; 290/1 R, 2, 1 A; 244/158 R, 166, 244/172; 322/27
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

1,540,998	A	1/1921	Plauson	
2,813,242	A	11/1957	Crump	
3,013,201	A	12/1961	Goldie	
4,127,804	A	11/1978	Breaux	
4,151,409	A	4/1979	O'Hare	
4,595,852	A	6/1986	Gundlach	
5,052,638	A	10/1991	Minovitch	
5,148,461	A *	9/1992	Shoulders	378/119
5,153,901	A *	10/1992	Shoulders	378/119
5,305,974	A	4/1994	Willis	
5,502,354	A *	3/1996	Correa et al.	315/111.01
5,813,217	A	9/1998	Beall	
5,966,986	A	10/1999	Laul	
6,089,511	A	7/2000	Rasmusson	
6,193,194	B1	2/2001	Minovitch	

(Continued)

FOREIGN PATENT DOCUMENTS

WO W001/49804 A2 7/2001

OTHER PUBLICATIONS

R.V. Anderson, in Electrical Processes in Atmospheres (H. Dolezalek & R. Reiter, eds., Steinkopff, Darmstadt, 1977), pp. 87-99.

(Continued)

Primary Examiner — T C Patel
Assistant Examiner — Iraj A Mohandesi
(74) *Attorney, Agent, or Firm* — Michael B. Farber

(57) **ABSTRACT**

A pyramid electric generator for harvesting the vibrational energies of Earth's atomic oscillators according to the present invention comprises: (1) an antenna/waveguide that is geometrically optimized; (2) a secondary coil wound with an insulated conductor on a nonconductive coil form, the coil being attached electrically to the conducting surface of the antenna/waveguide such that the secondary coil is attached near the point at which the electric field contacts the antenna/waveguide; (3) the antenna/waveguide connected with the secondary coil serving as a quasi-capacitive series element to provide a specific resonant frequency; and (4) a primary coil of a few turns wound around the secondary coil, the secondary coil being positioned coaxially within the primary coil and acting as a resonant step-up transformer winding, inductively coupled with the primary coil. The generator resonantly couples into specific frequencies of Earth's atomic oscillators and extracts electric energy therefrom.

19 Claims, 5 Drawing Sheets

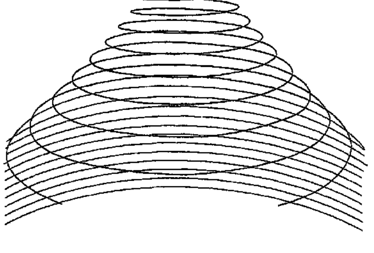

U.S. PATENT DOCUMENTS

6,974,110 B2 12/2005 Grandics

OTHER PUBLICATIONS

R.G. Roble & I. Tzur, in the Earth's Electrical Environment, Studies in Geophysics (National Academy Press, Washington, D.C. 1986), pp. 206-231.

R. Feynman, Lectures on Physics (Addison-Wesley, Inc., Palo Alto, California, 1964) v.2, ch.9, pp. 1-11.

W.R. Gringel et al., in the Earth's Electrical Environment, Studies in Geophysics (National Academy Press, Washington, D.C. 1986), pp. 166-182.

R.H. Holzworth et al., "Direct Measurement of Lower Atmospheric Vertical Potential Differences," Geophys. Res. Lett. 8: 783-786 (1981) .

R.H. Holzworth, "Hy-wire Measurement of Atmospheric Potential," J. Geophys. Res. 89: 1395-1401 (1984).

R.E. Orville, in the Earth's Electrical Environment, Studies in Geophysics (National Academy Press, Washington, D.C. 1986), pp. 23-29.

R.B. Standler & W.P. Winn, "Effects of Coronae on Electric Fields Beneath Thunderstorms," Quart. J.R. Met. Soc. 105: 285-302 (1979).

R.A. Oliva & C.W. Dale, "Circuit Analysis Methods", in DC Circuits, vol. II (Radio Shack, Fort Worth, Texas, 1979), pp. 13-8 to 13-12.

W.M. Hayt, Jr. & J.E. Kemmerly. "Engineering Circuit Analysis" (3rd ed., McGraw-Hill, New York, 1978), pp. 146-152 .

H. Christian et al., "Airborne and Ground Based Studies of Thunderstorms in the Vicinity of Langmuir Laboratory," Q. J. R. Meteorol. Soc. 106: 159-175 (1980).

T.C. Marshall & W.P. Winn, "Measurements of Charged Precipitation in a New Mexico Thunderstorm: Lower Positive Charge Centers," J. Geophys. Res. 87:. 7141-7157 (1982).

P. Grandics, "A Method to Capture Atmospheric Electrostatic Energy", in Proceedings of IEJ-ESA Joint Symposium on Electrostatics, (Kyoto University, Kyoto, Japan, 2000), pp. 355-361.

P. Grandics, "The Genesis of Fundamental Forces Acting at a Distance," Infinite Energy 12: 13-24 (2007).

P. Grandics, "The Genesis of Electromagnetic and Gravitational Forces," J. New Energy 6: 33-45 (2007).

R.A. Kerr, "Will the Arctic Ocean Lose All Its Ice," Science 286: 1828 (1999).

S.A. Laxon et al., "High Interannual Variability of Sea Ice Thickness in the Arctic Region," Nature 425: 947-950 (2003).

http://www.pyramidoflife.com/eng/tests_experiments.html.

T.C. Marshall & M. Stolzenburg, "Voltages Inside and Just Above Thunderstorms," J. Geophys. Res. 106: 4757-4768 (2001).

* cited by examiner

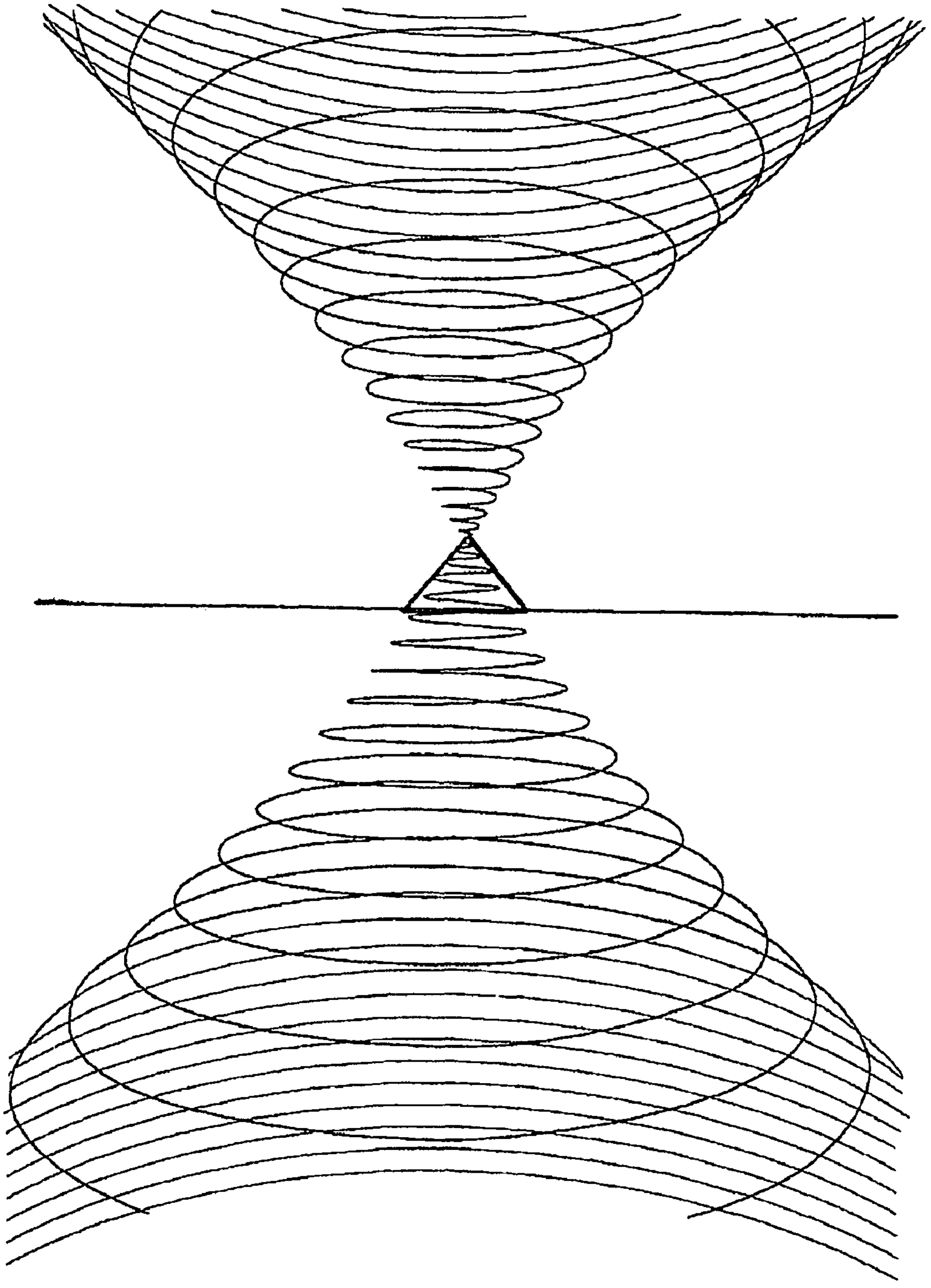

FIG. 1

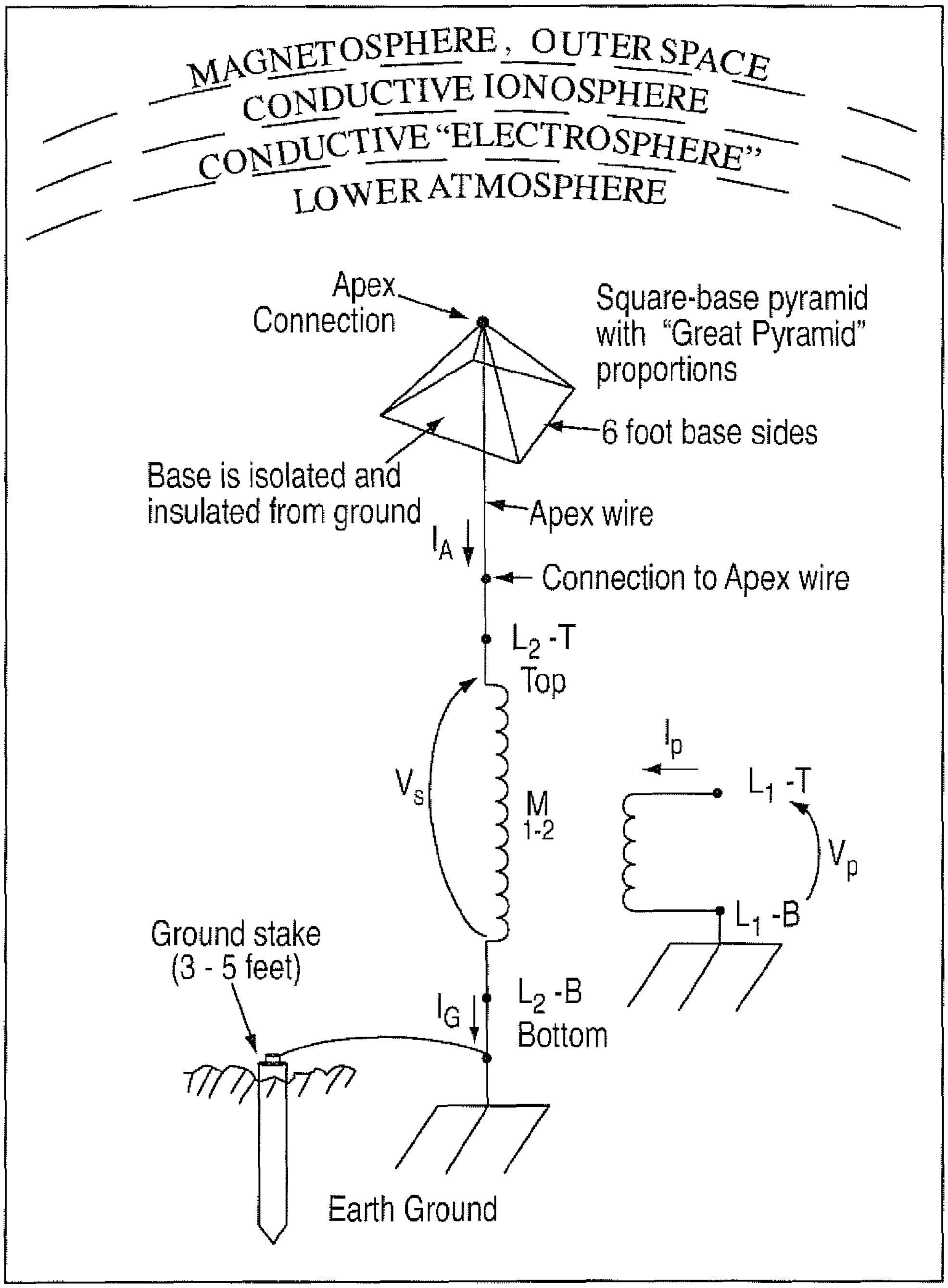
MAGNETOSPHERE, OUTER SPACE
CONDUCTIVE IONOSPHERE
CONDUCTIVE "ELECTROSPHERE"
LOWER ATMOSPHERE
Apex Connection
Square-base pyramid with "Great Pyramid" proportions
6 foot base sides
Base is isolated and insulated from ground
Apex wire
I_A
Connection to Apex wire
L_2 -T Top
V_s
M 1-2
I_p
L_1 -T
V_p
L_1 -B
Ground stake (3 - 5 feet)
I_G
L_2 -B Bottom
Earth Ground

V_p	Voltage across primary coil
I_p	Current into primary coil
I_A	Current in the apex wire
I_g	Current into ground wire
V_s	Voltage across secondary coil

FIG. 2

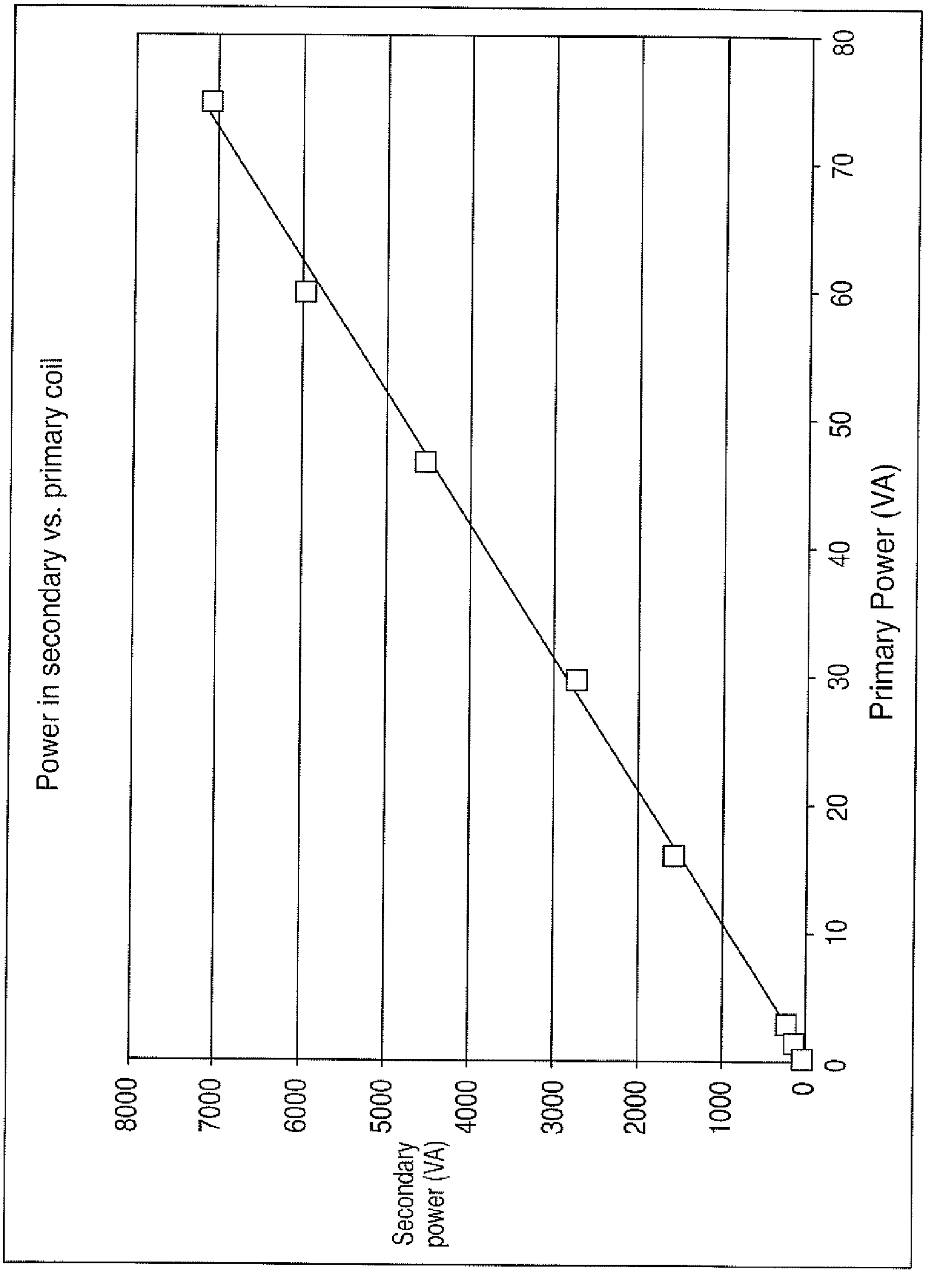
Power in secondary vs. primary coil
Secondary power (VA)
8000
7000
6000
5000
4000
3000
2000
1000
0
0
10
20
30
40
50
60
70
80
Primary Power (VA)

FIG. 3

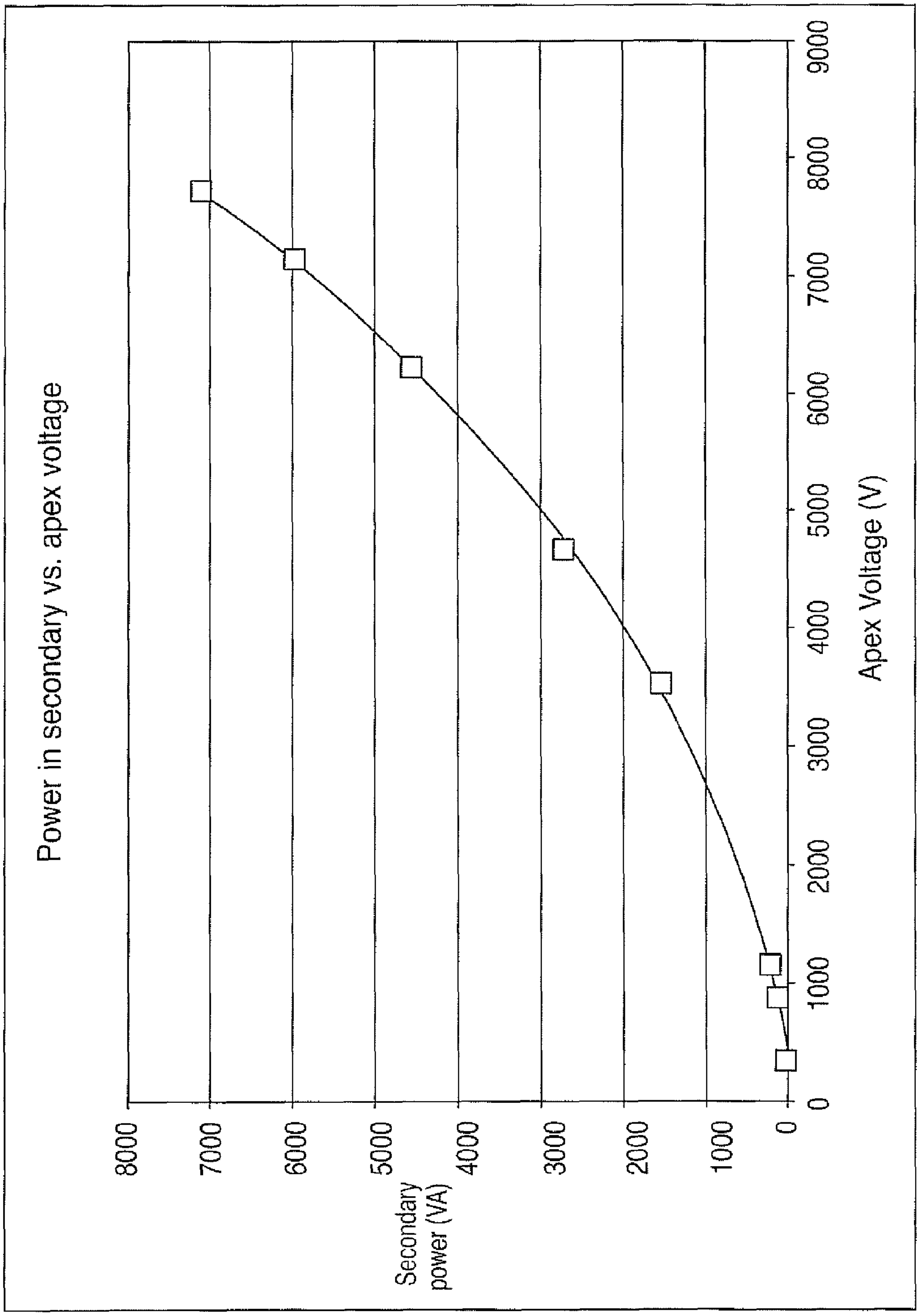
Power in secondary vs. apex voltage
Secondary power (VA)
8000
7000
6000
5000
4000
3000
2000
1000
0
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
Apex Voltage (V)

FIG. 4

FIG. 5

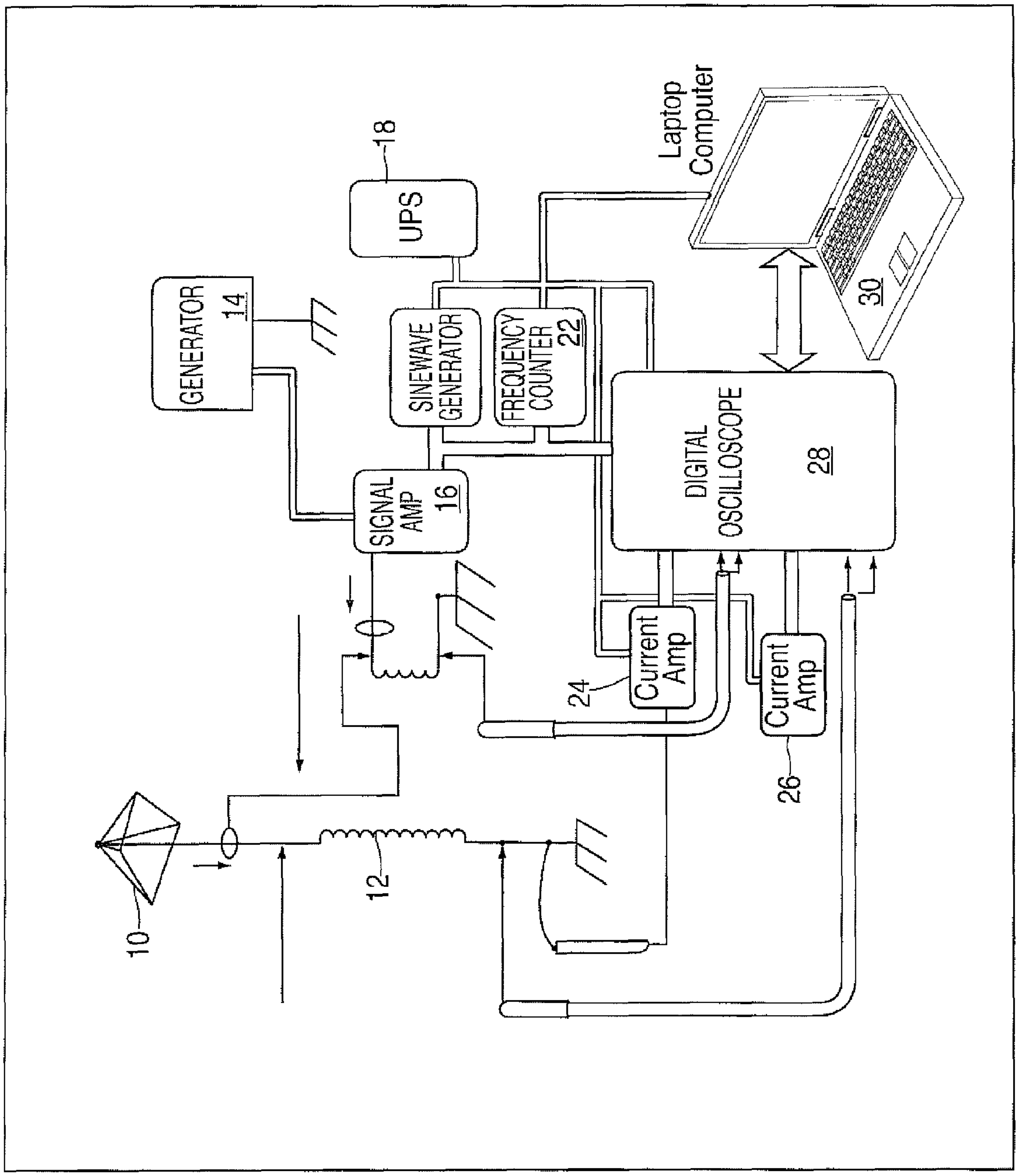
10
12
GENERATOR
14
SIGNAL AMP
16
UPS
18
SINEWAVE GENERATOR
FREQUENCY COUNTER
22
DIGITAL OSCILLOSCOPE
28
Current Amp
24
Current Amp
26
Laptop Computer
30

PYRAMID ELECTRIC GENERATOR

CROSS-REFERENCES

This application is a continuation-in-part of PCT Application Ser. No. PCT/US07/69804 by Grandics, entitled "An Electric Power Converter for Extraction of Atmospheric Electrical Energy," designating the United States and filed on May 25, 2007, which in turn claimed priority from U.S. Provisional Application Ser. No. 60/818,360 by Grandics, also entitled "An Electric Power Converter for Extraction of Atmospheric Electrical Energy," and filed on Jul. 3, 2006. The specifications of these two applications are incorporated herein in their entirety by this reference.

BACKGROUND OF THE INVENTION

This invention relates to the generation of electrical power by drawing energy from Earth's electric field. Earth represents a resonant system in which both atmospheric and telluric electrical processes can be harvested. In U.S. Pat. No. 6,974,110, we have provided an apparatus and method for converting the electrostatic potential energy of Earth's atmosphere. The present invention aims to provide an electric generator that, in addition to tapping atmospheric electric processes, extracts electrical power from the vibrational energy of atomic oscillators by a novel resonant coupling method.

SUMMARY OF THE INVENTION

An electrical generator according to the present invention provides a new method of tapping Earth's electric energy field.

One aspect of the present invention is an apparatus for capturing both atmospheric and telluric electric energies, comprising:

(1) an antenna/waveguide that is geometrically optimized, having a conducting surface;

(2) a primary coil wound with a conductor over a secondary coil, the coil being attached electrically to a driver operating in the LF radio band; and

(3) a secondary coil smaller in diameter than the primary coil having a greater length and a greater number of turns than the primary coil and having a first lead and a second lead, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding, inductively coupled with the primary coil and an external capacitor connected in parallel with the secondary coil to provide a specific resonant frequency; or

(4) a secondary coil smaller in diameter than the primary coil, having a greater length and a greater number of turns than the primary coil and having a first lead and a second lead, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding, inductively coupled with the primary coil and connected with the antenna/waveguide serving as a quasi-capacitive series element to provide a specific resonant frequency;

wherein the conducting surface of the antenna/waveguide is configured such that the secondary coil is connected near the point at which the electric field contacts the antenna/waveguide;

wherein the generator absorbs electromagnetic impulses from Earth's electric field; and wherein a high voltage sinusoidal waveform is generated in the secondary coil representing harvested atmospheric and telluric electrical energy and measurable on the leads of the secondary coil.

Preferably, the antenna/waveguide is of pyramidal shape. When the antenna/waveguide is of pyramidal shape, the secondary coil is preferably connected near or at the apex of the antenna/waveguide.

Another aspect of the present invention is an electric generator for harvesting the electric emissions of atomic oscillations comprising an oscillator. Typically, the oscillator operates in the LF or ELF bands.

Typically, in this aspect of the present invention, the generator comprises:

(1) a primary coil wound with a conductor over a secondary coil, the coil being connected electrically to a driver operating in the LF or ELF bands; and

(2) a secondary coil of smaller diameter than the primary coil having a greater length and a greater number of turns than the primary coil, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding inductively coupled with the primary;

wherein said resonant step-up transformer's output voltage exceeds 500 V_{RMS};

wherein the electric generator attracts impulses from Earth's electric oscillations; and

wherein the attracted energy manifests as high voltage sinusoidal waveforms representing harvested electric energy of atomic oscillations in the secondary coil and measurable on leads of the secondary coil.

Another aspect of the present invention is a method of tapping Earth's electric energy field, comprising these steps:

(1) positioning an electric generator according to the present invention as described above, so that it is exposed to a source of atmospheric and telluric electric energy; and

(2) producing a high-voltage sinusoidal signal representing electrical energy harvested by the operation of the generator.

BRIEF DESCRIPTION OF THE DRAWINGS

The following invention will be better understood with reference to the specification, appended claims, and accompanying drawings, where:

FIG. **1** is a graph showing the formation of electric vortices across and around the pyramid.

FIG. **2** is a graph showing the circuit diagram of the pyramid generator.

FIG. **3** is a graph showing the power in the secondary coil vs. the power in the primary coil.

FIG. **4** is a graph showing the power in the secondary coil vs. the apex voltage on the pyramid.

FIG. **5** is a graph showing the circuit diagram of the pyramid generator with the measurement points.

DETAILED DESCRIPTION OF THE INVENTION

One aspect of the present invention is an electric generator for attracting the electric emissions of atomic oscillations by the use of high voltage, to harvest electric energy therefrom.

One embodiment of this aspect of the invention is an apparatus for capturing atmospheric and telluric electric energies by harvesting the energies of atomic oscillators, comprising:

(1) an antenna/waveguide that is geometrically optimized, having a conducting surface;

(2) a primary coil wound with an insulated conductor, the coil being connected to an AC or AC-DC driver; and

(3) a secondary coil smaller in diameter than the primary coil, having a greater length and a greater number of turns than the primary coil and having a first lead and a second lead, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding,

inductively coupled with the first coil; and an external capacitor connected in parallel with the secondary coil to provide a specific resonant frequency; or

(4) a secondary coil smaller in diameter than the primary coil, having a greater length and a greater number of turns than the primary coil and having a first lead and a second lead, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding, inductively coupled with the first coil and connected with the antenna/waveguide serving as a quasi-capacitive series element to provide a specific resonant frequency, with the secondary coil attached near the point at which the electric field contacts the antenna/waveguide;

wherein the generator absorbs electromagnetic impulses from Earth's electric field; and

wherein a high-voltage, sinusoidal waveform is generated in the secondary coil representing harvested atmospheric and telluric electrical energy and measurable on leads of the secondary coil.

Preferably, the antenna/waveguide is of pyramidal shape. However, any geometric shape can be used for the antenna/waveguide. When it is of pyramidal shape, the secondary coil is preferably connected near or at the apex of the antenna/waveguide. The shape and size of the antenna/waveguide can be optimized depending on the intended use of the generator. Preferably, the antenna/waveguide exhibits ratios of pi and the Fibonacci number (phi).

Typically, the generator further comprises an insulated base on which the antenna/waveguide is placed. The size and shape of the insulated base can also be optimized, depending on the intended use of the generator. The second lead of the secondary coil is typically connected to earth ground.

Typically, the output of the secondary coil is harvested. This can be done by connecting to a rectifier-capacitor-load resistance, or by a third coil inductively coupled with the secondary coil. In one alternative, the output of the secondary coil is coupled to a load for drawing power from the generator. The optimal circuit can be selected by one of ordinary skill in the art to optimize power extraction. The load draws power from the generator. The load can be a resistor, a rectifier or a storage capacitor powering a DC load. Typically, the antenna/waveguide is at a high voltage during the operation of the generator. Typically, the antenna/waveguide voltage is greater than 500 V_{RMS}. The antenna/waveguide can be connected to an oscillator. The oscillator can comprise a resonant step-up transformer.

Typically, the antenna/waveguide is in resonance with the secondary coil. Typically, the primary coil is inductively coupled to the secondary coil.

Typically, the antenna/waveguide is positioned in a generally north-south direction to tap atmospheric and telluric electric energy. In this arrangement, when the antenna/waveguide is of pyramidal shape, the height of the pyramid can be from about 0.5 m to about 1000 m, more typically, from about 0.50 m to about 150 m. The optimal pyramid size is a function of the power requirements. However, a pyramid can be built even 1 km in height. Typically, the base surface area of the pyramid ranges from about 0.25 m^2 to about 1,000,000 m^2; more typically, from about 0.25 m^2 to about 25,000 m^2. Naturally, other heights and base surface areas can be used for particular applications. The numbers of sides can also vary. When the pyramid is small (e.g., 6-feet base length described in this application), there is an insufficient atmospheric potential on its apex to trigger the attraction of power. In such event, an AC or AC-DC driver is necessary to bring the pyramid to the electric potential necessary. A suitable voltage is typically greater than 500 V_{RMS}.

In one alternative, the primary coil is wound with a conductor over the secondary coil, the primary coil being connected electrically to a driver operating in the LF or ELF bands.

Another aspect of the present invention is an electric generator for harvesting the electric emissions of atomic oscillations comprising an oscillator. Typically, the oscillator operates in the LF or ELF bands.

Typically, in this aspect of the present invention, the generator comprises:

(1) a primary coil wound with a conductor over a secondary coil, the coil being connected electrically to a driver operating in the LF or ELF bands; and

(2) a secondary coil of smaller diameter than the primary coil having a greater length and a greater number of turns than the primary coil, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding inductively coupled with the primary;

wherein said resonant step-up transformer's output voltage exceeds 500 V_{RMS};

wherein the electric generator attracts impulses from Earth's electric oscillations; and

wherein the attracted energy manifests as high voltage sinusoidal waveforms representing harvested electric energy of atomic oscillations in the secondary coil and measurable on leads of the secondary coil.

In this alternative, the output of the secondary coil can be coupled to a load for drawing power from the generator.

Another aspect of the invention is a method of tapping Earth's electric energy based on the following concepts:

(1) utilizing a pyramidal antenna/waveguide designed based on the ratios of pi and phi;

(2) using the antenna/waveguide to absorb and focus specific frequencies of Earth's electric field;

(3) positioning a resonant coil system connected to said antenna/waveguide near or at its apex according to the present invention; and

(4) generating a high-voltage sinusoidal signal, representing harvested atmospheric and telluric electrical energy, by the operation of the coil system.

Typically, in this method, the method comprises the steps of:

(1) positioning an electric generator according to the present invention as described above such that it is exposed to Earth's electric oscillations; and

(2) generating a high voltage sinusoidal signal representing harvested electric energy of atomic oscillators by the operation of the generator.

Typically, the harvested electric energy can be fed into a power grid for distribution, but smaller units capable of distributed power generation or for use as stand-alone generators are also feasible by the concept of this invention.

The invention is described by the following Example. This Example is included for illustrative purposes only and is not intended to limit the invention.

EXAMPLE

We have introduced a new theory of space, energy and matter that predicted that electromagnetic interaction propagates through a carrier medium called the space lattice that is made up of pulsating, phi-based spiral vortices (STAR), the excitation of which generates propagating sinusoidal lines of force across the space lattice recorded as electromagnetic waves by our instruments [1,2]. We have also predicted that natural pressure gradients exist within the space lattice that can be used for power generation by inserting a capacitor into such gradient. Furthermore, we proposed that a pyramid-shaped capacitor is an optimally shaped device for tapping the energy gradient of the space lattice that manifests in the form of an electrical potential gradient in Earth's atmosphere. This was demonstrated subsequently [3-5]. We also found that the pyramid-shaped capacitor design should be based on ratios of pi and phi [5] because electrical energy propagates along a phi-based spiral [6].

We suggested that elemental particles are composed of STAR oscillators that make up the space lattice by electromagnetic compression of its elemental cubes into pulsating conical/vortexual subunits that form the basis for all particles of matter [1,2]. We proposed that atomic and subatomic oscillators could vibrate in a very wide frequency range from sub-acoustic to extremely high EM frequencies [1,2]. This does not mean that material bodies can take or maintain a physical form at any energy level, but this allows the recycling of atoms into the space lattice as part of their natural evolution [2,6].

Atomic oscillators are constantly in motion, and such motion is electric because all motion and energy is fundamentally electric [2,6]. It is known that atomic bodies constantly exchange electric energy. Such energy flow can be harvested when the spiral physical nature of electric waves is recognized. Russell observed that energy always moves during its generative cycle in spirals towards the higher potential [7]; therefore, we reasoned that an "attractor" of high electric potential must be provided to pull in the random electromagnetic emissions of atomic oscillators. The "attractor" must function as a phi-based antenna/waveguide to focus the phi-based electric emissions of atomic bodies into the apex of a vortex, the same method Nature uses for power multiplication [6,7]. Therefore, a phi-based pyramidal antenna/waveguide must be at a high voltage in order to perform its attractor function.

As all material manifestations are tonal [8], acoustic vibrations are always generated during the oscillations of atomic bodies, even though they are inaudible most of the time. Sometimes, however, they are detected; one such example is Earth's hum [9-11]. Earth's solid mass and its atmosphere represent a coupled resonant system, and is therefore our expanded target power source. Acoustic waves are also electrical [8] and can be harvested. To achieve this goal, the geometrically optimized pyramidal antenna/waveguide must be connected to an oscillator that operates at a high voltage and is tuned to a suitably selected frequency.

The appropriate resonant frequency is dictated by practical considerations. Atomic oscillators can vibrate in a very wide frequency range [1,2,6-8] but short wavelength radiations are readily re-absorbed by atoms and are rapidly attenuated; therefore, long wavelength electromagnetic emissions must be targeted. The operation of the pyramid generator is feasible in the ELF to the LF range. As electric radiation propagates in a vortex [2,6-8], the emitted electric vortex over the pyramidal antenna (FIG. **1**) increases its "size," and so it is capable of funneling atmospheric electromagnetic emissions of the same frequency back into the antenna/receiver. Thus, the pyramid electric generator could capture the energy emitted by both telluric and atmospheric atomic oscillators. We have tested these assumptions and demonstrated the basic principles of an electric generator.

Methods and Results

In our previous papers, we reported that a pyramid-shaped capacitor/antenna converts atmospheric electrostatic discharge impulses (ESD) into a periodic high-frequency signal that can be detected in an insulated coil placed in proximity of the capacitor [3,4]. Here, we report on further developments including a new design for the resonant coil system and an expanded range of energy sources targeted. A Tektronix TPS 2024 digital oscilloscope was used for signal acquisition and analysis, and Tektronix A503 current amplifiers were used with the Tektronix A6302 and A6303 probes for current measurement. High voltage was measured using a Tektronix P6015 probe. An ENI-1140LA amplifier was used to activate the primary coil of the resonant coil system at the selected frequency.

For the experiments, we used a geometrically optimized pyramid-shaped antenna/waveguide [3-5]. The pyramid expressed ratios of pi and phi and was approximately a 1:125 scaled down replica of the Great Pyramid of Giza (GPF), 6 feet base length. The pyramid was built of a wooden frame and covered with triangular copper sheets of 0.66 mm thickness on its sides that were electrically connected. It was placed on an insulating base and positioned in the general North-South direction.

The secondary coil (coil **2**) was wound with a 20 AWG insulated magnet wire with a monofilament nylon spacer (0.06 mm diameter) between the turns on a fiberglass cylindrical coil form, and connected to the conducting surface of the pyramid near its apex. Thus, the secondary coil was connected with the pyramidal antenna/waveguide as a quasi-capacitive series element to provide a specific resonant frequency with coupling to the surrounding electromagnetic environment. The other lead of coil **2** was grounded (FIG. **2**). The purpose of the monofilament spacer was to reduce inter-winding capacitance in coil **2** by increasing distance between wire turns of the coil.

A primary coil of larger diameter (coil **1**) was wound on or near the lower cylindrical volume of coil **2** with the lead connected to earth ground, and served as a driver coil inductively coupled with the secondary (coil **2**). The secondary was isolated from the primary by layers of silicon rubber and foam. The two coils make a resonant step-up transformer, the parameters of which are shown in Table 1.

TABLE 1

Physical and Electrical Parameters of the Resonant Transformer	
Secondary turns of wire	840 turns
Height	0.85 m
Wire diameter	0.00081 m
Primary turns of wire	5 + 9/10 turns
Height	0.21 m
Wire diameter	¼" O.D. copper pipe with .030" wall thickness
Primary	
Inductance	98 μH
Resistance	0.118 ohms
Secondary	
Inductance	24.16 mH
Resistance	24.42 ohms
Capacitance	1040 nF
Winding ratio	1:143
Mutual inductance	6.13 mH

In the experiment; the controlled variable was the voltage, which was increased in steps until the maximum voltage in the secondary (permitted by circuit components and test equipment) was achieved for the given resonant frequency. The objective was to "drive the atmosphere" by radiating a high voltage (~20 kV p-p), LF (40-120 kHz) signal developed by the high ratio step-up transformer and fed by a powerful signal generator at a resonant frequency either with an L-C resonant secondary coil, or a resonance determined by the pyramid as a quasi-capacitive series element connected to the secondary coil of the transformer and radiating to the local atmosphere. The data are shown in Table 2.

The experimental data are shown in Table 2. The current readings were also confirmed by using "current sensing" resistors (not shown). A apparent power gain of nearly 100-fold was achieved in the secondary coil relative to input power in the primary coil, and a linear increase of power in the secondary coil with the increase of input power into the primary coil (FIG. **3**). The power in the secondary coil exponentially increases with the pyramid apex voltage (FIG. **4**).

TABLE 2

Power Measurements in the Pyramid-Resonant Coil System

Primary								
Frequency (kHz)	83.00							
$V_{P\ RMS}$	3.97	6.38	9.86	20.3	26.7	33.8	39	40.8
$I_{P\ RMS}$ (mA)	87.6	222	300	792	1117	1389	1541	1838
$I_{P\ angle\ with\ respect\ to}V_P$	18°	24°	14°	22°	23°	21°	20°	28°
Total power (VA)	0.35	1.42	2.95	16.02	29.70	46.77	59.95	74.75
True power (W)	0.33	1.30	2.87	14.85	27.35	43.66	56.34	66.01
Secondary								
$V_{S\ RMS}$	341	872	1150	3540	4670	6220	7140	7710
$I_{A\ RMS}$ (mA)	62.2	148	193	438	579	728	834	919
$I_{A\ angle\ with\ respect\ to}V_S$	42°	43°	44°	52°	52°	54°	53°	52°
Total power (VA)	21.2	129.5	222.8	1552	2707.8	4530.2	5957.5	7087.3
True power (W)	15.7	94.7	160.3	956	1668.1	2664.5	3587.6	4366
Power sec/prim	61.1	91.4	75.4	96.9	91.2	96.9	99.4	94.8

A load may be connected to coil **2** to draw power from the system. The load may also be connected via a tertiary coil inductively coupled to the secondary coil. The load may be a resistor, a rectifier or storage capacitor powering a DC load.

A graphic representation of the test system is shown in FIG. **5**. The pyramidal antenna/waveguide, placed on an insulating base, is coupled to the surrounding electromagnetic environment and serves as an antenna/waveguide for the concentration of atmospheric and telluric electromagnetic oscillations at resonance frequency. As the flow of electric energy is always balanced [6], two vortices may cross the pyramidal antenna: one from the direction of the ground concentrating into its apex, and another one from its apex propagating into the atmosphere (FIG. **1**). The atmospheric electric vortex possibly acts as an ionic antenna pulling in EM radiations from a large atmospheric domain. In FIG. **5**, the antenna/waveguide **10** is attached to a coil **12**. To drive the system, a RF generator **14** is connected to a signal amplifier **16**. The signal amplifier **16** is also connected to an uninterruptible power supply (UPS) **18** through a sinewave generator **20** and a frequency counter **22**. Outputs from the sinewave generator **20** and frequency counter **22** are also connected to a first current amplifier **24**. The output from the antenna/waveguide **10** and coil **12** are also connected to a second current amplifier **26**. The sinewave generator **20**, frequency counter **22**, first current amplifier **24**, and second current amplifier **26** are connected to a digital oscilloscope **28**. Outputs from the antenna/waveguide **10** and coil **12** are also connected to the digital oscilloscope **28**. A laptop computer **30** controls the sinewave generator **20** and frequency counter **22**.

The pyramid's apex is attached to a coil of high turn number (FIGS. **2** & **4**). This secondary coil wound on a nonconductive coil form serves as a step-up transformer, and forms a resonant circuit with the pyramid that acts as a quasi-capacitive series element. The secondary coil is activated by a signal from an RF generator via a primary coil of a few turn numbers inductively coupled to coil **2**. The measurement points are also indicated, displaying the positions of voltage and current probes. The data are captured by a digital oscilloscope and recorded on a laptop computer.

Discussion

This invention demonstrates a novel approach to harvest Earth's electric energy. The masses of Earth and its atmosphere represent a coupled resonant system that is continually electrified by solar radiation. We have found that a pyramidal antenna designed based on the ratios of phi and pi is optimal for the capture of atmospheric electrostatic discharge (ESD) impulses [3-5]. Atmospheric ESD is a product of solar radiation and is a wide bandwidth phenomenon. The pyramid as a wideband, non-resonant antenna is uniquely adapted to harvest the energy of atmospheric ESD.

This invention also points out an additional energizing mechanism at work during the operation of the pyramid electric generator. The observation is derived from our theory on energy, matter and space [1,2] as well as that of Russell [6,7]. Both Russell and we observed that spiral motion is a fundamental action of matter and that the vortex is the mechanism of power multiplication in Nature. A vortex concentrates power into its apex where the highest velocity of motion, the highest pressure and the highest electric potential resides [1,2,6,7]. As electricity propagates along a pulsating phi-based spiral [6,7], a phi-based antenna/waveguide is suitable to focus its energy into the apex of a vortex inside the antenna. A phi-based pyramid is optimal for this purpose.

As atomic bodies can oscillate in a very wide frequency range [1,2,6-8], a pyramidal antenna/waveguide coupled to an oscillator that is tuned to a suitable frequency can focus the "electric noise" from Earth's atomic oscillators into the apex of the pyramid at a high voltage. Russell observed that energy moves towards the higher potential during its generative cycle [7]; therefore, it is essential that the pyramidal antenna be at a high potential. Our data corroborated this assumption.

To select a suitable frequency range, one can begin with the observation that all material manifestations are tonal [8], and that acoustic frequencies are always generated during the oscillations of atomic bodies even though they remain inaudible most of the time. Acoustic waves are also electric [8] and therefore Earth's enormous mass and its atmosphere represent a coupled resonant system that offer a vast source of power.

For practical reasons on our small pyramid, we have chosen a frequency above the audio range but below the radio frequency spectrum; our antenna was made to radiate at the target wavelength. The wavelength of the 50-120 kHz range is 6 km to 25 km, thus allowing energy to be harvested over a large atmospheric domain. As electrical energy propagates in a vortex [6,7], the emitted electric vortex (ionic antenna) over the pyramidal antenna/waveguide increases its "size" and enables it to funnel the same frequency atmospheric electromagnetic radiations back into the antenna. Thus, the pyramid electric generator captures at the selected frequency the energy emitted by Earth's atomic oscillators.

The presence of an electric (ionic) vortex was demonstrated by radar testing of the space over a 44 m tall fiberglass pyramid located near Moscow [12]. The Russians found that the large ionized column of air over the area of the vertical axis of the pyramid had a width of about 500 m and reached an altitude of 2 km. It is remarkable that this effect was induced by a nonconductive pyramid surface demonstrating a significant degree of atmospheric ionization even under fair weather conditions. Thus, a suitably sized large or an electrically activated small pyramid should open a low impedance path to higher elevations of relatively conductive atmospheric domains.

In conclusion, the total power that can be extracted from Earth's atomic oscillators must be extremely large, likely far exceeding current global electric generation capacity. In our experiment, we obtained over 7000 VA power at about 20 kV p-p pyramid voltage using a 6 feet base length pyramid. This power is nearly 100 times greater than the power necessary to drive the pyramid to the required operating voltage. By constructing the pyramid generator from dedicated high voltage components and using higher drive voltages, a significant increase in power output is envisioned. The power extraction will also be enhanced by further optimization of the designs of both the pyramid and the coil system.

The power output will also increase by employing larger pyramid structures and coils. As voltage is the primary factor in attracting power into the system, the necessary voltage can be provided by the vertical atmospheric potential on a tall pyramid. Since the atmospheric vertical potential gradient could go as high as 1200 V/m near Earth's surface under fair weather conditions [13], a pyramid height of 80-150 m seems sufficient to provide the apex voltage for a self-sustaining power generator. As the pyramid scales up volumetrically, a power generator pyramid of the size range of the GPG could likely have an output in the hundreds to thousands of megawatts range. Groups of several pyramid electric generators could be placed within specific geographical areas, thus combining their energy harvesting capacity.

References

The following references are specifically applicable to the Example and are incorporated herein by reference; these references are referenced in the Example by the reference numbers assigned to them.

[1] Grandics, P. 2002, "The genesis of electromagnetic and gravitational forces." *J. New Energy,* 6, (3) 33-45.

[2] Grandics, P. 2007, "The genesis of fundamental forces acting at a distance and some practical derivations." *Infinite Energy,* 12, (71) 13-24.

[3] Grandics, P. 2000, "A method to capture atmospheric electrostatic energy", in *Proceedings of IEJ-ESA Joint Symposium on Electrostatics*, Kyoto University, Kyoto, Japan, pp. 355-361.

[4] Grandics, P. 2006, "A DC to RF converter for the capture of atmospheric electrostatic energy", in *Proceedings of the 5_{th} Conference of the Societe Francaise D'Electrostatique, SFE* 2006, Grenoble, France pp. 279-284.

[5] Grandics, P. 2007, "Pyramidal Electric Transducer: A DC to RF Converter for the Capture of Atmospheric Electrostatic Energy." *Infinite Energy,* 13, (73) 20-27.

[6] Russell, W. 1947, in *The Secret of Light,* 3_{rd} ed. University of Science and Philosophy, Waynesboro, Va., pp. 218-288.

[7] Russell, W. 1957, in *Atomic Suicide?* 2_{nd} ed. University of Science and Philosophy, Waynesboro, Va., pp. 129-153.

[8] Russell, W. 1926, in *The Universal One*, University of Science and Philosophy, Waynesboro, Va., pp. 247-249.

[9] Nishida, K., Kobayashi, N. and Fukao, Y. 2000, "Resonant oscillations between the solid earth and the atmosphere", *Science* 287, (5461), 2244-2246.

[10] Tanimoto, T., Um, J., Nishida, K., and Kobayashi, N. 1998, "Earth's continuous oscillations observed on seismically quiet days", *Geophys. Res. Lett.* 25, 1553-1556.

[11] Tanimoto, T. 2001, "Continuous free oscillations: Atmosphere-solid earth coupling". *Ann. Rev. Earth Planet Sci.* 29, 563-584.

[12] http://www.pyramidoflife.com/eng/tests_experiments.html.

[13] Marshall, T. C., Rust, W. D., Stolzenburg, M., Roeder, W. P., and Kriebel, P. R. 1999, "A study of enhanced fair-weather electric fields occurring soon after sunrise". *J. Geophys. Res.* 104, 24,455-24, 469.

ADVANTAGES OF THE INVENTION

The present invention provides a new method of tapping Earth's electric energy and providing usable power that can be fed into a power grid or supplied by stand-alone power generators. It does so without the need for mechanical energy or the consumption of fossil fuel or the long-term risks associated with power generated by nuclear fission, including the risk of diversion of fissionable material to terrorist aims or the risk posed by the required long-term storage of spent nuclear fuel. Devices according to the present invention can operate virtually continuously with little to no maintenance.

With respect to ranges of values, the invention encompasses each intervening value between the upper and lower limits of the range to at least a tenth of the lower limit's unit, unless the context clearly indicates otherwise. Moreover, the invention encompasses any other stated intervening values and ranges including either or both of the upper and lower limits of the range, unless specifically excluded from the stated range.

Unless defined otherwise, the meanings of all technical and scientific terms used herein are those commonly understood by one of ordinary skill in the art to which this invention belongs. One of ordinary skill in the art will also appreciate that any methods and materials similar or equivalent to those described herein can also be used to practice or test this invention.

The publications and patents discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

All the publications cited are incorporated herein by reference in their entireties, including all published patents, patent applications, and literature references, as well as those publications that have been incorporated in those published documents. However, to the extent that any publication incorporated herein by reference refers to information to be published, applicants do not admit that any such information published after the filing date of this application is to be considered prior art.

As used in this specification and in the appended claims, the singular forms include the plural forms. For example the terms "a," "an," and "the" include plural references unless the content clearly dictates otherwise. Additionally, the term "at least" preceding a series of elements is to be understood as referring to every element in the series. The inventions illustratively described herein can suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms "comprising," "including," "containing," etc. shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the future shown and described or any portion thereof, and it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the inventions herein disclosed can be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of the inventions disclosed herein. The inventions have been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the scope of the generic disclosure also form part of these inventions. This includes the generic description of each invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised materials specifically resided therein. In addition, where features or aspects of an invention are described in terms of the Markush group, those schooled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group. It is also to be understood that the above description is intended to be illustrative and not restrictive. Many embodiments will be apparent to those of in the art upon reviewing the above description. The scope of the invention should therefore, be determined not with reference to the above description, but should instead be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled. Those skilled in the art will recognize, or will be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described. Such equivalents are intended to be encompassed by the following claims.

I claim:

1. An electric generator for harvesting the energies of atomic oscillators comprising:

(a) an antenna/waveguide that is geometrically optimized having a conducting surface;

(b) a primary coil wound with an insulated conductor over a secondary coil, the coil being connected to an AC or AC-DC driver; and

(c) a secondary coil of smaller diameter than the primary coil having a greater length and a greater number of turns than the primary coil and having a first lead and a second lead, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding inductively coupled with the primary and an external capacitor connected in parallel with the secondary coil to provide a specific resonant frequency; or

(d) a secondary coil of smaller diameter than the primary coil having a greater length and a greater number of turns than the primary coil and having a first lead and a second lead, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding inductively coupled with the primary and connected with the antenna/waveguide serving as a quasi-capacitive series element to provide a specific resonant frequency;

wherein the conducting surface of the antenna/waveguide is configured such that the secondary coil is attached near the point at which the electric field contacts the antenna/waveguide;

wherein the generator absorbs impulses from Earth's electric oscillations; and wherein the attracted energy manifests as high voltage sinusoidal waveforms in the secondary coil, representing harvested atmospheric and telluric electrical energy and measurable on the leads of the secondary coil.

2. The electric generator of claim 1, wherein a suitable voltage is greater than 500 V_{RMS}.

3. The electric generator of claim 1, wherein the antenna/waveguide is designed based on ratios of pi and phi.

4. The electric generator of claim 1, wherein the antenna/waveguide is of pyramidal shape.

5. The electric generator of claim 1, wherein the antenna/waveguide is placed on an insulated base.

6. The electric generator of claim 1, wherein the antenna/waveguide is at a high voltage during the operation of the generator.

7. The electric generator of claim 6, wherein the antenna/waveguide voltage is greater than 500 V_{RMS}.

8. The electric generator of claim 1, wherein the antenna/waveguide is connected to an oscillator.

9. The electric generator of claim 8, wherein the oscillator comprises a resonant step-up transformer.

10. The electric generator of claim 4, wherein the secondary coil of the transformer is connected near the apex of the antenna/waveguide.

11. The electric generator of claim 1 wherein the second lead of the secondary coil is connected to earth ground.

12. The electric generator of claim 1 wherein the primary coil is wound with a conductor over the secondary coil, the primary coil being connected electrically to a driver operating in the LF or ELF bands.

13. The electric generator of claim 1, wherein the primary coil is inductively coupled to the secondary coil.

14. An electric generator for harvesting the electric emissions of atomic oscillations of claim 1 comprising an oscillator.

15. The electric generator of claim 14, wherein said oscillator operates in the LF or ELF bands.

16. The electric generator of claim 14 wherein the generator comprises:

(a) a primary coil wound with a conductor over a secondary coil, the coil being connected electrically to a driver operating in the LF or ELF bands; and

(b) a secondary coil of smaller diameter than the primary coil having a greater length and a greater number of turns than the primary coil, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding inductively coupled with the primary; wherein said resonant step-up transformer's output voltage exceeds 500V_{RMS};

wherein the electric generator attracts impulses from Earth's electric oscillations; and wherein the attracted energy manifests as high voltage sinusoidal waveforms representing harvested electric energy of atomic oscillations in the secondary coil and measurable on leads of the secondary coil.

17. The electric generator of claim **4**, wherein the height of the pyramidal electric generator is from about 0.50 m to about 1000 m.

18. The electric generator of claim **4**, wherein the base surface area of the pyramid is from about 0.25 m^2 to about 1,000,000 m^2.

19. The electric generator of claim **1**, wherein the generator is positioned in a generally north-south direction.

* * * * *

US006974110B2

(12) United States Patent
Grandics

(10) **Patent No.: US 6,974,110 B2**
(45) **Date of Patent: Dec. 13, 2005**

(54) METHOD AND APPARATUS FOR CONVERTING ELECTROSTATIC POTENTIAL ENERGY

(76) Inventor: **Peter Grandics**, 5922 Farnsworth Ct., Carlsbad, CA (US) 90720

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 74 days.

(21) Appl. No.: **10/331,007**

(22) Filed: **Dec. 27, 2002**

(65) **Prior Publication Data**

US 2004/0011925 A1 Jan. 22, 2004

Related U.S. Application Data

(63) Continuation-in-part of application No. 09/732,110, filed on Dec. 7, 2000, now abandoned.
(60) Provisional application No. 60/169,378, filed on Dec. 7, 1999.

(51) **Int. Cl.**^{7} **F03H 5/00**
(52) **U.S. Cl.** **244/172**; 244/62
(58) **Field of Search** .... 244/172, 158 R, 244/169, 62, 53 R, 51

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,013,201	A	12/1961	Goldie
4,127,804	A	11/1978	Breaux
4,151,409	A	4/1979	O'Hare
4,595,852	A	6/1986	Gundlach
5,052,638	A	10/1991	Minovitch
5,305,974	A	4/1994	Willis
5,813,217	A	9/1998	Beall
5,947,421	A	9/1999	Beattie et al.
5,966,986	A	10/1999	Laul
6,089,511	A	7/2000	Rasmusson
6,193,194	B1	2/2001	Minovitch

OTHER PUBLICATIONS

DC Circuits vol. II, Circuit Analysis Methods, 1979.*
Engineering Circuit Analysis, Third Edition, 1978.*
R.V. Anderson, *in Electrical Processes in Atmospheres*, (H. Holezalek & R. Reiter, eds, Steinkopff, Darmstadt, 1977), pp. 87–99.
R.G. Roble & I. Tzur, *in The Earth's Electrical Environment, Studies in Geophysics* (National Academy Press, Washington, D.C. 1986), pp. 206–231.
R.P. Feynman, *Lectures on Physics* (Addison–Wesley, Inc., Palo Alto, California, 1964), v.2, ch. 9, pp. 1–11.
W. Gringel, J.M. Rosen, & D.I. Hofmann, *in The Earth's Electrical Environment, Studies in Geophysics* (National Academy Press, Washington, D.C. 1986), pp. 166–182.
R.H. Holzworth et al., "Direct Measurement of Lower Atmospheric Vertical Potential Differences," *Geophys. Res. Lett.* 8: 783–786 (1981).
R.H. Holzworth, "Hy–wire Measurement of Atmospheric Potential," *J. Geophys. Res.* 89: 1395–1401 (1984).
R.E. Orville, *in The Earth's Electrical Environment, Studies in Geophysics* (National Academy Press, Washington, D.C. 1986), pp. 23–29.
R.B. Standler & W.P. Winn, "Effects of Coronae on Electrid Fields Beneath Thunderstorms," *Quart. J.R. Met. Soc.* 105: 285–302 (1979).

* cited by examiner

Primary Examiner—Peter M. Poon
Assistant Examiner—Timothy D. Collins
(74) *Attorney, Agent, or Firm*—Catalyst Law Group, APC; Michael B. Faber, Esq.

(57) ABSTRACT

A new method is described to produce useful electrical energy from DC electrostatic fields using a pyramid-shaped capacitor. The system uses no moving parts and no mechanical energy is introduced. Also, when a pyramid-shaped electrode is charged with DC high voltage, a propulsive force is generated. This will allow the manufacture of vehicles capable of levitation and flight.

13 Claims, 3 Drawing Sheets

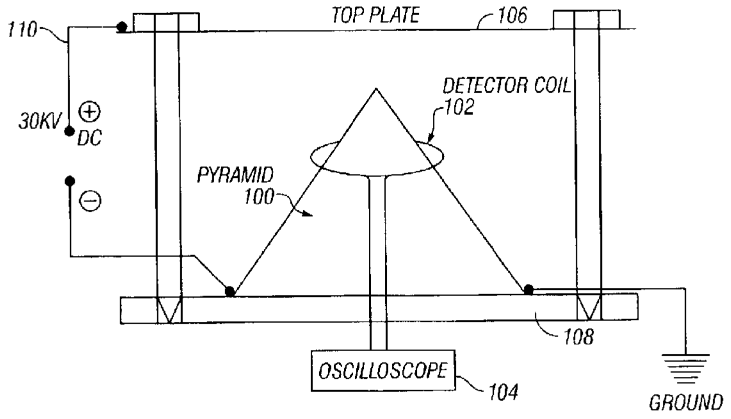

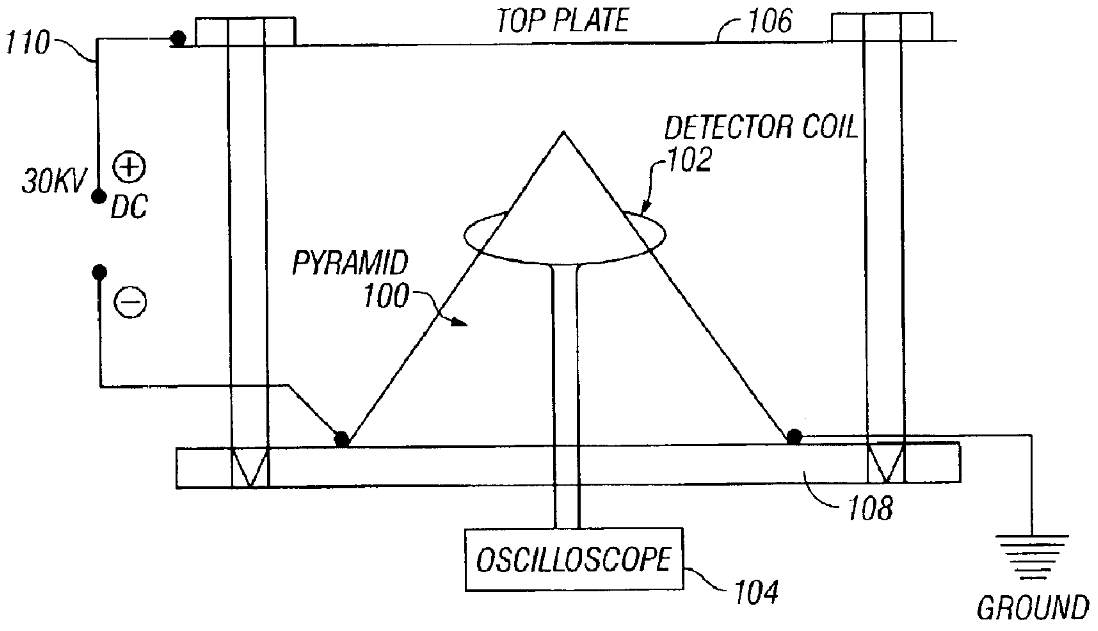
110
TOP PLATE
106
DETECTOR COIL
102
30KV
⊕
DC
⊖
PYRAMID
100
108
OSCILLOSCOPE
104
GROUND

FIG. 1

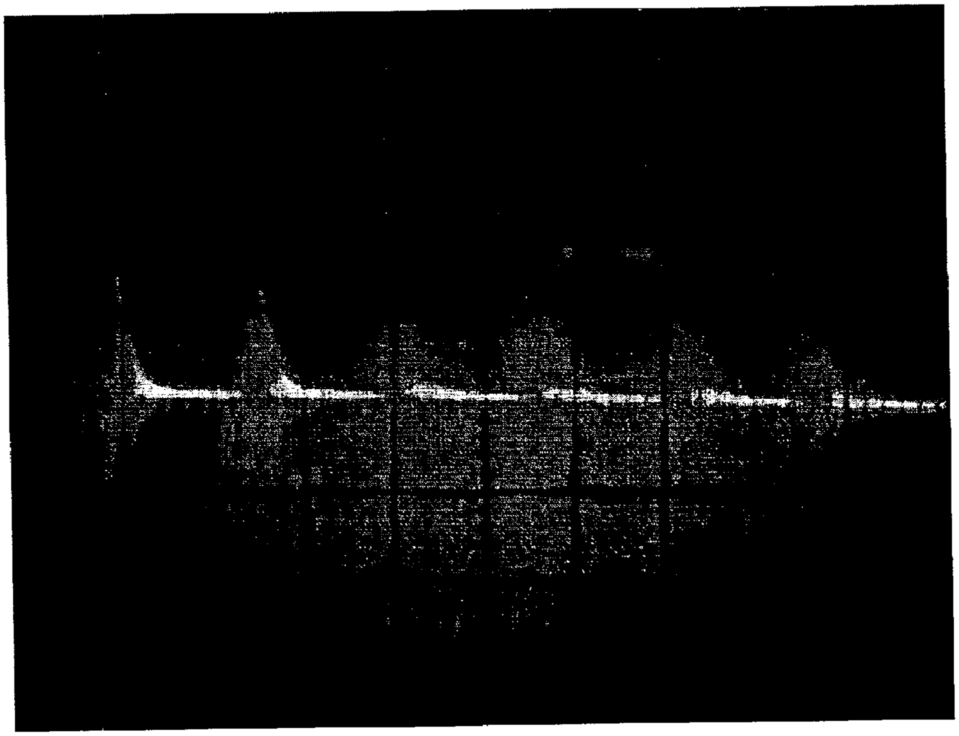

FIG. 2

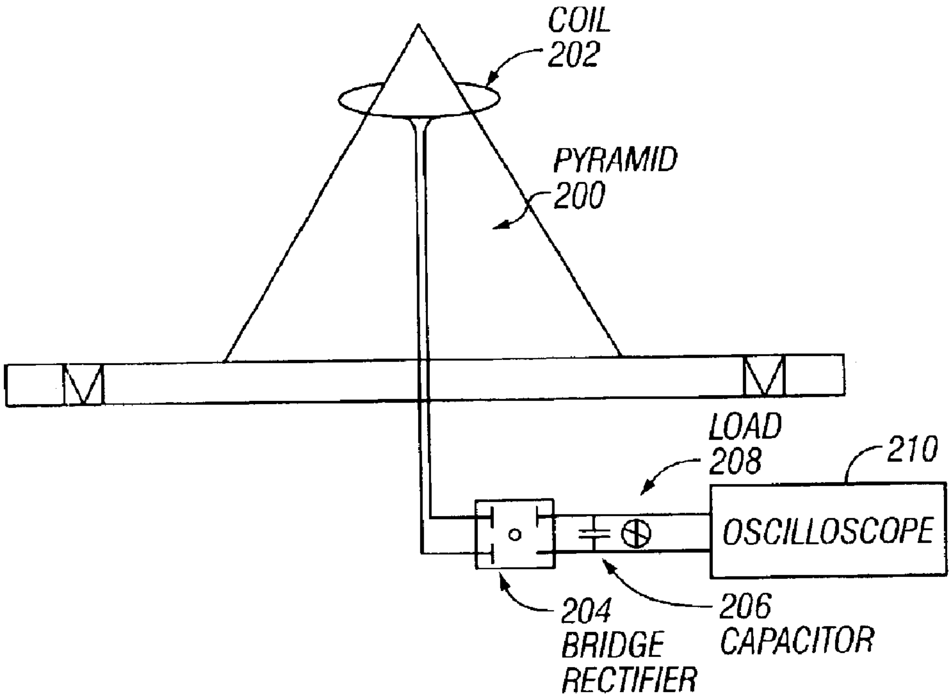
COIL
202
PYRAMID
200
LOAD
208
210
OSCILLOSCOPE
204
BRIDGE
RECTIFIER
206
CAPACITOR

FIG. 3

METHOD AND APPARATUS FOR CONVERTING ELECTROSTATIC POTENTIAL ENERGY

CROSS-REFERENCES

This application is a continuation-in-part of application Ser. No. 09/732,110, filed Dec. 7, 2000, now abandoned, entitled "Method and Apparatus for Converting Electrostatic Potential Energy," by Peter Grandics, which claims the benefit of Provisional application Ser. No. 60/169,378, filed Dec. 7, 1999, which is incorporated herein in its entirety by this reference.

BACKGROUND OF THE INVENTION

This invention relates to the generation of electrical power by drawing energy from an electrostatic potential field. The conversion of DC electrostatic energy into useable electrical energy by electrostatic generators is already described in the prior art as disclosed in U.S. Pat. Nos. 3,013,201, 4,127,804, 4,151,409 and 4,595,852. Generally, such prior art electrostatic generators utilize mechanical energy to separate charges and thus contain complex mechanics that is difficult to scale up for a high output system. Therefore, the present invention aims to provide an electrostatic generator in which electrical power is derived exclusively from the energy of DC electrostatic fields without the input of mechanical power.

SUMMARY OF THE INVENTION

Pursuant to this invention a simple technique is described to convert the energy of a DC electrostatic field into an alternating current by wrapping a coil around the pyramid. The resulting AC current can be rectified and used for practical purposes. A pyramid-shaped capacitor can also be used in an inverse mode of operation for the generation of propulsive force.

Accordingly, one embodiment of the present invention is a method for converting DC electrostatic energy into usable electrical energy, the method comprising the steps of:

(1) providing a capacitor of pyramidal shape;

(2) placing an insulated coil on the surface of the capacitor, the coil having leads;

(3) attaching a rectifier to the leads of the coil, the rectifier having leads; and

(4) attaching a capacitor or a battery to the leads of the rectifier so that DC electrostatic energy is converted into usable electrical energy.

Another embodiment of the present invention comprises a method for converting DC electrostatic energy into propulsive force comprising the steps of:

(1) providing a first capacitor of pyramidal shape;

(2) placing an insulated conductive tip on the point of the capacitor;

(3) providing a static generator for charging the capacitor, and

(4) attaching 3 ball-shaped smaller capacitors to the bottom of the pyramidal-shaped capacitor; and

(5) providing means to extend or retract the ball-shaped capacitors from the pyramidal-shaped capacitor, so DC electrostatic energy is converted into propulsive force by the action of the first capacitor and the 3 ball-shaped smaller capacitors provide directional control.

BRIEF DESCRIPTION OF THE DRAWINGS

The following invention will become better understood with reference to the specification, appended claims, and accompanying drawings, where:

FIG. **1** is a diagram demonstrating the pyramid experimental setup;

FIG. **2** is the pyramid coil signal as detected by oscilloscope; and

FIG. **3** is a diagram demonstrating power transmission on the pyramid setup.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Pursuant to this invention, a novel method is described to convert DC electrostatic energy into an AC current that can be rectified and used for practical purposes. The shape of the capacitor and the body of such device is designed to convert the DC electrostatic energy into the AC current for maximal effect.

A pyramidal or conical shape is preferred for one of the capacitor electrodes (FIG. **1**). In FIG. **1**, a detector coil **102** is provided that is in turn connected to an oscilloscope **104**. The coil surrounds the metallic pyramid **100**. In the experimental setup shown, the field is established between a top plate **106** and the pyramid **100** by using a ground **108** connected to a source of DC electrical energy **110**. When a high voltage DC field (30 kV) is established on such capacitor, a regularly repeating, clocklike signal is detected in the coil placed on the pyramid's surface (FIG. **2**). This is an unexpected observation as corona discharges are irregular by nature.

The alternating current from the coil can be rectified and used for practical purposes. If a suitable DC electrostatic field could be found in nature, this principle would be useful in tapping the energy of such field. To test for this possibility, I have measured the rectified signal from the coil without an external power source. The rectified coil output was collected in a capacitor and voltage measured in 1 h intervals. The voltage measured is significantly higher if the capacitor electrode is pyramid-shaped as opposed to a box-shaped electrode of the same height and volume. When the pyramid is placed into a Faraday cage, the signal is abolished (see details in the Example). The data has demonstrated in principle that with this experimental setup electrical energy can be extracted from the Earth's electrostatic field. The Earth's surface and the ionosphere substitutes for the two charged electrodes, which exhibit negative and positive polarities, respectively.

Atmospheric electricity manifests as a buildup of electrostatic energy, a phenomenon that permanently electrifies our environment (Anderson, R. V. in Electrical Processes in Atmospheres, H. Holezalek and R. Reiter, eds., (Steinkopff, Darmstadt, 1977) pp. 87–99). The voltage gradient between the Earth's surface and the ionosphere is thought to be maintained by the electrical activity of the troposphere as well as the solar wind-coupled magnetospheric dynamo (Roble, R. G. Tzur, I. in The Earth's Electrical Environment, Studies in Geophysics (National Academy Press, Washington D.C., 1986) pp. 206–231). The Earth's electrostatic field is never depleted, despite the constant discharges of energy. There are two segments of atmospheric electrical processes that potentially could be tapped, fair weather and thunderstorm related atmospheric electrical phenomena.

The Earth's voltage field creates a potential difference of about 400,000 V (Feynman, R. P. Lectures on Physics (Addison Wesley, Inc., Palo Alto, Calif., 1964) v.2, Chapter 9), with a voltage gradient of about 200–300 V/m around the surface of the Earth (W. Gringell, J. M. Rosen, D. I. Hofmann, in The Earth's Electrical Environment, Studies in Geophysics (National Academy Press, Washington D.C.,

1986) pp. 166–182). Given this voltage drop of 200–300 V/m, the desired 30 kV potential can be obtained at a pyramid height of about 100–150 m.

Due to the evolving nature of our understanding of atmospheric electrical processes, we do not have a clear picture of the magnitude of electrical energy generated (and dissipated) in the global atmospheric electrical circuit by localized convective phenomena. Direct measurement of the atmospheric vertical potential difference using tethered balloons have shown that under fair weather conditions and at low altitudes (150–550 m), short circuit currents on the order of 10 μA were obtained with a single wire collector yielding a power of about 1 W (Holzworth, R. H. et al., Direct measurement of lower atmospheric vertical potential differences, Geophys. Res. Lett. 8, 783–786 (1981), and Holzworth, R. H. Hy-wire measurement of atmospheric potential, J. Geophys. Res. 89, 1395–1401 (1984)).

These data remarkably demonstrated that it is possible to couple to a large portion of the atmosphere and also gave an indirect proof of the presence of a fair weather convective generation process. Since at a low altitude of 1500 m, the entire atmospheric vertical electric potential (400 kV) could be bridged, it may even be possible to couple right into ionospheric electrical processes (Holzworth, R. H. et al., Direct measurement of lower atmospheric vertical potential differences, Geophys. Res. Lett. 8,783–786 (1981)). This would significantly expand the available energy pool for this device.

The other possibility is to tap localized peaks of atmospheric electrical activity arising from thunderstorm activity. The approximate order of magnitude of electrostatic energy generated in the troposphere could be estimated by the amount of energy dissipated by lightning discharges. The estimates for the total number of lightning strikes over the entire surface of the Earth vary from 100 to 300 per second (Orville, R. E. in The Earth's Electrical Environment, Studies in Geophysics (National Academy Press. Washington D.C., 1986) pp. 23–29). The estimate for the power discharged in an average lightning event is 10^{12} watts over a time period of about 0.2 sec (Christian, H. J. and McCook, M. A. in A lightning primer at http://thunder.msfc.nasa.gov/). This yields a total of $2–6\times10^{13}$ W-sec of electrical energy spent by lightning discharges globally in each second, or approximately 4×10^{17} W-h per year. Over 90% of all lightning occur over landmasses. The distribution of tropospheric electrical activity is highly concentrated, the major production zones being Central Africa, the south central United States and the Amazon Basin (Miller, T. L. Global lighting activity at http://www.ghcc.msfc.nasa.gov/rotating/otd_oval_full.html). This suggests that the density of atmospheric electrical activity in these zones may be high enough to tap.

The current annual global electricity production by the world's power generating industry is $1,22\times10^{16}$ W-h (Seth, A. Nuclear power and its role in global electricity and energy, Institute for Energy and Environmental Research, at http://www.ieer.org/ensec/no-1/glbnrg.html (1997)) which is only a small fraction, about 1⁄30th, of the electricity generated and discharged naturally by lightning activity. As lightning represents only peak charge density events, the localized total charge production of thunderstorms is necessarily underestimated by a factor, the magnitude of which is still unknown.

These indicate that atmospheric electricity, if harnessed, could meet all the energy needs of mankind. This invention describes a new electrical phenomenon that allows the conversion of electrostatic energy into an alternating current.

Research on the average electrical power generated and dissipated by the Earth's global electrical circuit suggests that about 2000 thunderstorms are in progress at any given time, and that the average storm current delivered to the ionosphere is of the order of one ampere (Roble, R. G. Tzur, I. in The Earth's Electrical Environment, Studies in Geophysics (National Academy Press, Washington D.C., 1986) pp. 206–231). The potential difference between the Earth and the ionosphere is of the order of 500 kV; therefore, the total source power is of the order $2000\times1\times5\times10^{5}$, or 10_9 watts over the entire planet. Thunderstorms produce an average of one to several nanoamperes per square meter at the surface (Standler, R. B. and Winn, W. P. Quart. J. Met. Soc. 105 285 (1979)). These estimates would suggest that there might not be sufficient energy to tap.

However, the limitation of these low, average Earth surface current densities is not applicable to the pyramid collector for several reasons. In thunderstorms, terawatts of electrical power are generated over relatively long periods of time. To capture this electric power and prevent its dissipation, an effective "charge sink" is necessary. The sub-optimal geometry of the Earth's surface terrain and its relatively low conductivity produce an ineffective sink which leads to small ground surface current densities. A metal pyramid, however, due to its optimal geometry and construction acts as an effective charge sink.

FIG. **3** shows power transmission on the pyramid setup. In FIG. **3**, a pyramid **200** as described above is in contact with a coil **202** and a bridge rectifier **204**. The latter is charging capacitor **206** which feeds load **208**. Oscilloscope **210** measures the voltage drop on the load.

Accordingly, one embodiment of the invention is a method for converting DC electrostatic energy into usable electrical energy, the method comprising the steps of:

(1) providing a capacitor of pyramidal shape;

(2) placing an insulated coil on the surface of the capacitor, the coil having leads;

(3) attaching a rectifier to the leads of the coil, the rectifier having leads; and

(4) attaching a capacitor or a battery to the leads of the rectifier so that DC electrostatic energy is converted into usable electrical energy.

As detailed below, typically, a clock-like signal is detected from the leads of the coil. The signal can be rectified, pooled in a capacitor, or pooled in a battery.

A production pyramid, with a projected 40,000 m^2 base surface area, a height of 150 m, and a metallic composition would provide a far more effective charge sink than the surrounding ground surface. As current always flows in the direction of least resistance, this would lead to a concentrated flow of current from the thunderclouds into the pyramid. As a result, a column of air above the surface of the pyramid would become highly ionized and much more conductive than the air in the surrounding area. This would ensure a rapid transfer of charges from the thunderclouds into the pyramid. Therefore, the current flowing into the pyramid would likely be very strong. As a result, lightning activity could be reduced or abolished. In addition, since the threshold effect causes exponential power amplification, such a pyramid could produce very large quantities of electrical energy.

The generation of thunderstorm-related electrical energy in the atmosphere shows a high degree of concentration into three major geographical zones (Miller, T. L. Global lightning activity at http://www.ghcc.msfc.nasa.gov/rotating/otd_oval_full.html). By placing energy-producing pyramids into these zones of high-energy electrical activity, vast quantities of electricity could be obtained. Atmospheric electricity would be a renewable, clean energy source, and could give us the opportunity to begin reversing the process of global warming caused by anthropogenic greenhouse gas emissions. This process, which now produces well-measurable results, threatens to upset our ecosystem with unpredictable consequences (Kerr, R. A. Science 286 1828 (1999)).

Electrostatic energy could also be used to produce a propulsive force. A capacitor made into the shape of a pyramid or cone is suitable. The conductive elements of the capacitor are layered into the pyramidal or conical segment of the hull of such craft. Optimally, there should be several layers of negatively charged conductors on the outside while a single layer of positively charged conductor is in the inside of the craft. All conductive elements should be carefully insulated in order to maintain the charges inside the capacitor.

In the center of the craft, a conductive, insulated pole is mounted such that the pole extends out into the exterior at the top of the craft. The tip of that pole has a metal ball mounted on which is also positively charged. When all capacitor plates are charged, a pressure differential is induced and above the craft the pressure will be higher than at the bottom of the craft. This pressure differential is unstable and progresses towards an equilibrium. It will propel the craft forward in order to eliminate the pressure gradient. This requires that a sufficiently high voltage (>100,000 V) be applied. The capacitors incorporated into the hull will give control over the lift.

To obtain directional control over all three axes of movement, three capacitors need to be mounted at the bottom of the craft placed 120° apart from each other. These capacitors could take the shape of a ball and would be mounted on retractable shafts. The top half of the capacitors would be positively charged while the bottom half is negatively charged. When in flight, the capacitors are retracted and charged based on the needs of directional control so that the ball-shaped capacitors provide lateral movement control for the vehicle. To achieve forward movement, the charges on one capacitor facing the direction of the desired path should be reduced relative to the other two capacitors. This would make the craft tilt forward and start moving in that direction. An alternative method of lateral directional control could be obtained by dividing the capacitor bank inside the hull into segments which segments would be charged individually.

Upon landing, the three ball capacitors would be protracted on their shafts and serve as landing gear. On landing, they would be discharged because the earth is the greater ground. On launching, the main capacitors of the hull would be activated leading to takeoff. The three-capacitor landing gear would then be retracted and charged up so that they can function for lateral directional control after takeoff. On-board static generators would supply voltage required for the operation of the craft.

Accordingly, another embodiment of the invention comprises a method for converting DC electrostatic energy into propulsive force comprising the steps of:

(1) providing a first capacitor of pyramidal shape;

(2) placing an insulated conductive tip on the point of the capacitor;

(3) providing a static generator for charging the capacitor, and

(4) attaching 3 ball-shaped smaller capacitors to the bottom of the pyramidal-shaped capacitor; and

(5) providing means to extend or retract the ball-shaped capacitors from the pyramidal-shaped capacitor, so DC electrostatic energy is converted into propulsive force by the action of the first capacitor and the 3 ball-shaped smaller capacitors provide directional control.

Inertial forces pose great obstacles to improving the performance of current propulsion systems. The most significant aspect of this type of propulsion would be the loss of inertia during acceleration. The passengers of such craft would feel no forces acting upon their bodies even if the craft achieved enormous accelerations.

This overall design is suitable for the development of a wide range of transportation vehicles, from automobile-type to aircraft and potentially to spacecraft. With such technology, we could eliminate a major factor in environmental pollution which place a great stress on our ecosystem. In the United States, transportation vehicles produce 66% of the greenhouse gases.

The subject invention offers a solution by providing a renewable, non-polluting electricity generation method as well as an electricity-based transportation method capable of levitation and flight.

The following Example illustrates the features and advantages of the subject invention. Accordingly, it is to be understood that the description in this disclosure is to facilitate comprehension of the invention and should not be construed to limit the scope thereof as persons skilled in the art can, in light of this disclosure, generate additional embodiments without exceeding the scope or departing from the spirit of the claimed invention.

EXAMPLE 1

Demonstration of the pyramid generator: For the experiments, I have selected a one-foot base length foam pyramid from a pyramid vendor (The Pyramid Project, Ft. Wayne, Ind.). The outside of the pyramid was covered with aluminum foil. The pyramid was placed on a 2'×2'insulating polyethylene platform equipped with an adjustable height 2'×2'size aluminum top plate, 1/16" thick. The height of the aluminum plate was adjusted as needed and a gap of 1 1/4" between the plate and the tip of the pyramid was used in the experiments. In some experiments, an aluminum pyramid was used with a wall thickness of 1/16".

A high voltage (HV) CRT power source producing 30 kV DC was taken from a color monitor. I have assumed that an actual energy-producing pyramid should be relatively high in order to obtain a large voltage drop from its tip to the ground. Therefore, assuming a height of 100–150 m for a life-size pyramid and a voltage drop of 200–300 V/m near the surface of the Earth, the 30 kV is in the range of the voltage-drop expected for the height of a life size pyramid.

The positive pole was attached to the top aluminum plate. This simulated the positive charge of the atmosphere. One corner of the pyramid was attached to the negative pole of the high voltage power source, while the opposite corner of the pyramid was grounded. This setup served as a model for the electrostatic field distribution around a potential life-sized pyramid. As controls, either a 1'×1' sheet of aluminum foil or an aluminum foil-covered box, having the main dimensions of the test pyramid (1'×1'×7 5/8"), was used as a negative pole. The detector coils were made by winding up a 24 gauge enamel-coated magnet wire (20 turns, approximately 8 cm in diameter). A Tektronix high-frequency oscilloscope, Model no. 2236 was used for signal acquisition and analysis.

The first sets of experiments were control measurements with a box of the same height and base length as the test pyramid. The detector coil was placed on the top of the box. Measurements were taken with or without the high voltage applied. One corner of the box was attached to the HV power source (negative pole) and the opposite corner to the ground.

The same arrangement was used for the flat square (1'×1') foil. The peak-to-peak signal amplitude for the box was 8 mV and the signal frequency was 2 MHz. For the flat foil sheet, the signal amplitude was 12 mV with a frequency of 1.43 MHz. The signal form was of a decaying sine wave.

When high voltage was applied to these shapes, signal amplitude of 14 mV was obtained for the flat sheet and of 16 mV for the box. The signal frequency was 1.54 MHz for the flat sheet and 2 MHz for the box. The waveforms were of decaying sine waves in all these experiments.

When the pyramid was tested without HV, the peak-to-peak signal amplitude was measured at 60 mV with a frequency of 2 MHz. When the high voltage was applied, the signal amplitude increased up to 180–200 mV, while the frequency remained at 2 MHz. The pyramid produced signal intensity significantly higher than the controls. The signal is regularly repeating, clock-like in nature (FIG. **2**). When a metal (aluminum) pyramid of the same size (wall thickness 1⁄16" inch) was tested in the same high voltage field using the same detection coil, a voltage of 1–1.5 V was detected at the frequency of 2 MHz.

To collect energy from the coil, a bridge rectifier (1000 V peak voltage at 6 A) was attached to the leads of the coil. The rectified current was fed into a capacitor (1500 μF, 250 V DC max.), and a direct current of 45 V was obtained. This has demonstrated a simple method to convert electrostatic energy into a continuous direct current. An 8-turn coil having an output of 200–300 V AC (peak-to-peak) was also used for energy conversion. The rectified current from the 8-turn coil powered a 0.186 W light bulb (FIG. **3**). The bridge rectifier is optimally made of fast-recovery diodes.

Even in the absence of an externally applied voltage, current is always present in the circuit. Charge builds up in the capacitor and 1 V was obtained overnight using the 20-turn coil setup. Over 48 h, a voltage of 5 V was measured. Faraday shielding practically abolished the phenomenon.

The preferred shape of the pyramid as a charge collector was again demonstrated in further experiments using the 8-turn coil-bridge rectifier-capacitor (1,500 μF) assembly placed on the 1 ft base length aluminum pyramid. A same volume and height aluminum box was used as control. Charging times of 1–2 h were used under fair weather conditions. For the pyramid, 550 mV was measured on the capacitor while on the box 100 mV was obtained. This demonstrated the superiority of the pyramidal shape in capturing atmospheric electrostatic energy. It also demonstrated that we could tap into the electrostatic field of the atmosphere and draw electric energy. For the collection of energy, a battery could substitute for the capacitor.

I claim:

1. A method for converting DC electrostatic energy into usable electrical energy, the method comprising the steps of:

(a) providing a first capacitor of pyramidal shape, the first capacitor capable of being charged by and storing DC electrostatic energy;

(b) placing an insulated coil on the surface of the first capacitor, the coil having leads;

(c) attaching a rectifier to the leads of the coil, the rectifier having leads; and

(d) attaching a second capacitor or battery to the leads of the rectifier so that DC electrostatic energy stored by the capacitor of pyramidal shape is converted into usable electric energy.

2. The method of claim **1** wherein a clock-like signal is detected from the leads of the coil.

3. The method of claim **1** wherein the signal is rectified.

4. The method of claim **1** wherein the signal is pooled in the second capacitor.

5. The method of claim **1** wherein the rectified signal is pooled in the battery.

6. A method for converting DC electrostatic energy into propulsive force comprising the steps of:

(a) providing a first capacitor of pyramidal shape, the first capacitor capable of being charged by and storing DC electrostatic energy;

(b) placing an insulated conductive tip on the point of the first capacitor;

(c) providing a static generator for charging the first capacitor;

(d) attaching 3 ball-shaped smaller second, third, and fourth capacitors to the bottom of the pyramidal-shaped first capacitor; and

(e) providing means to extend or retract the ball-shaped second, third, and fourth capacitors from the pyramidal-shaped first capacitor, so DC electrostatic energy stored by the first capacitor of pyramidal shape is converted into propulsive force by the action of the first capacitor and the 3 ball-shaped smaller second, third, and fourth capacitors provide directional control.

7. The method of claim **6** wherein the energy stored in the pyramidal first capacitor is used for the propulsion of a vehicle.

8. The method of claim **6** wherein the pyramidal first capacitor is negatively charged.

9. The method of claim **6** wherein the insulated conductive tip placed on the point of the pyramidal first capacitor is positively charged.

10. The method of claim **6** wherein the ball-shaped second, third, and fourth capacitors provide lateral movement control for the vehicle.

11. The method of claim **6** wherein the ball-shaped second, third, and fourth capacitors serve as landing pods.

12. The method of claim **6** wherein the ball-shaped second, third, and fourth capacitors are retracted after take-off.

13. The method of claim **10** wherein the ball-shaped second, third, and fourth capacitors are charged for lateral directional control after takeoff.

* * * * *

US008004250B2

(12) United States Patent
Grandics

(10) **Patent No.:** **US 8,004,250 B2**
(45) **Date of Patent:** **Aug. 23, 2011**

(54) **PYRAMID ELECTRIC GENERATOR**

(76) Inventor: **Peter Grandics**, Carlsbad, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 476 days.

(21) Appl. No.: **12/140,887**

(22) Filed: **Jun. 17, 2008**

(65) **Prior Publication Data**

US 2008/0315587 A1 Dec. 25, 2008

Related U.S. Application Data

(63) Continuation-in-part of application No. PCT/US2007/069804, filed on May 25, 2007.

(60) Provisional application No. 60/818,360, filed on Jul. 3, 2006.

(51) **Int. Cl.**
H02N 1/00 (2006.01)

(52) **U.S. Cl.** .... **322/27**; 290/1 R; 322/32; 322/2 A

(58) **Field of Classification Search** .... 372/32, 372/123; 290/1 R, 2, 1 A; 244/158 R, 166, 244/172; 322/27
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

1,540,998 A		1/1921	Plauson	
2,813,242 A		11/1957	Crump	
3,013,201 A		12/1961	Goldie	
4,127,804 A		11/1978	Breaux	
4,151,409 A		4/1979	O'Hare	
4,595,852 A		6/1986	Gundlach	
5,052,638 A		10/1991	Minovitch	
5,148,461 A	*	9/1992	Shoulders	378/119
5,153,901 A	*	10/1992	Shoulders	378/119
5,305,974 A		4/1994	Willis	
5,502,354 A	*	3/1996	Correa et al.	315/111.01
5,813,217 A		9/1998	Beall	
5,966,986 A		10/1999	Laul	
6,089,511 A		7/2000	Rasmusson	
6,193,194 B1		2/2001	Minovitch	

(Continued)

FOREIGN PATENT DOCUMENTS

WO	W001/49804 A2	7/2001

OTHER PUBLICATIONS

R.V. Anderson, in Electrical Processes in Atmospheres (H. Dolezalek & R. Reiter, eds., Steinkopff, Darmstadt, 1977), pp. 87-99.

(Continued)

Primary Examiner — T C Patel
Assistant Examiner — Iraj A Mohandesi
(74) *Attorney, Agent, or Firm* — Michael B. Farber

(57) **ABSTRACT**

A pyramid electric generator for harvesting the vibrational energies of Earth's atomic oscillators according to the present invention comprises: (1) an antenna/waveguide that is geometrically optimized; (2) a secondary coil wound with an insulated conductor on a nonconductive coil form, the coil being attached electrically to the conducting surface of the antenna/waveguide such that the secondary coil is attached near the point at which the electric field contacts the antenna/waveguide; (3) the antenna/waveguide connected with the secondary coil serving as a quasi-capacitive series element to provide a specific resonant frequency; and (4) a primary coil of a few turns wound around the secondary coil, the secondary coil being positioned coaxially within the primary coil and acting as a resonant step-up transformer winding, inductively coupled with the primary coil. The generator resonantly couples into specific frequencies of Earth's atomic oscillators and extracts electric energy therefrom.

19 Claims, 5 Drawing Sheets

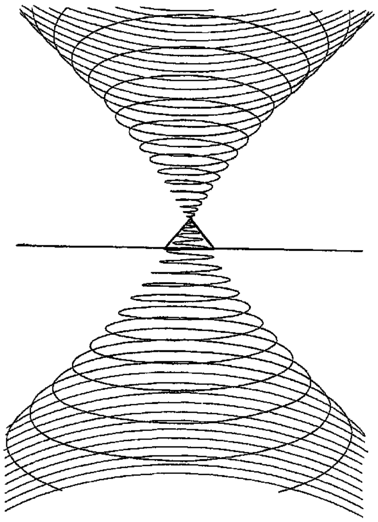

U.S. PATENT DOCUMENTS

6,974,110 B2 12/2005 Grandics

OTHER PUBLICATIONS

R.G. Roble & I. Tzur, in the Earth's Electrical Environment, Studies in Geophysics (National Academy Press, Washington, D.C. 1986), pp. 206-231.

R. Feynman, Lectures on Physics (Addison-Wesley, Inc., Palo Alto, California, 1964) v.2, ch.9, pp. 1-11.

W.R. Gringel et al., in the Earth's Electrical Environment, Studies in Geophysics (National Academy Press, Washington, D.C. 1986), pp. 166-182.

R.H. Holzworth et al., "Direct Measurement of Lower Atmospheric Vertical Potential Differences," Geophys. Res. Lett. 8: 783-786 (1981) .

R.H. Holzworth, "Hy-wire Measurement of Atmospheric Potential," J. Geophys. Res. 89: 1395-1401 (1984).

R.E. Orville, in the Earth's Electrical Environment, Studies in Geophysics (National Academy Press, Washington, D.C. 1986), pp. 23-29.

R.B. Standler & W.P. Winn, "Effects of Coronae on Electric Fields Beneath Thunderstorms," Quart. J.R. Met. Soc. 105: 285-302 (1979).

R.A. Oliva & C.W. Dale, "Circuit Analysis Methods", in DC Circuits, vol. II (Radio Shack, Fort Worth, Texas, 1979), pp. 13-8 to 13-12.

W.M. Hayt, Jr. & J.E. Kemmerly. "Engineering Circuit Analysis" (3rd ed., McGraw-Hill, New York, 1978), pp. 146-152 .

H. Christian et al., "Airborne and Ground Based Studies of Thunderstorms in the Vicinity of Langmuir Laboratory," Q. J. R. Meteorol. Soc. 106: 159-175 (1980).

T.C. Marshall & W.P. Winn, "Measurements of Charged Precipitation in a New Mexico Thunderstorm: Lower Positive Charge Centers," J. Geophys. Res. 87:. 7141-7157 (1982).

P. Grandics, "A Method to Capture Atmospheric Electrostatic Energy", in Proceedings of IEJ-ESA Joint Symposium on Electrostatics, (Kyoto University, Kyoto, Japan, 2000), pp. 355-361.

P. Grandics, "The Genesis of Fundamental Forces Acting at a Distance," Infinite Energy 12: 13-24 (2007).

P. Grandics, "The Genesis of Electromagnetic and Gravitational Forces," J. New Energy 6: 33-45 (2007).

R.A. Kerr, "Will the Arctic Ocean Lose All Its Ice," Science 286: 1828 (1999).

S.A. Laxon et al., "High Interannual Variability of Sea Ice Thickness in the Arctic Region," Nature 425: 947-950 (2003).

http://www.pyramidoflife.com/eng/tests_experiments.html.

T.C. Marshall & M. Stolzenburg, "Voltages Inside and Just Above Thunderstorms," J. Geophys. Res. 106: 4757-4768 (2001).

* cited by examiner

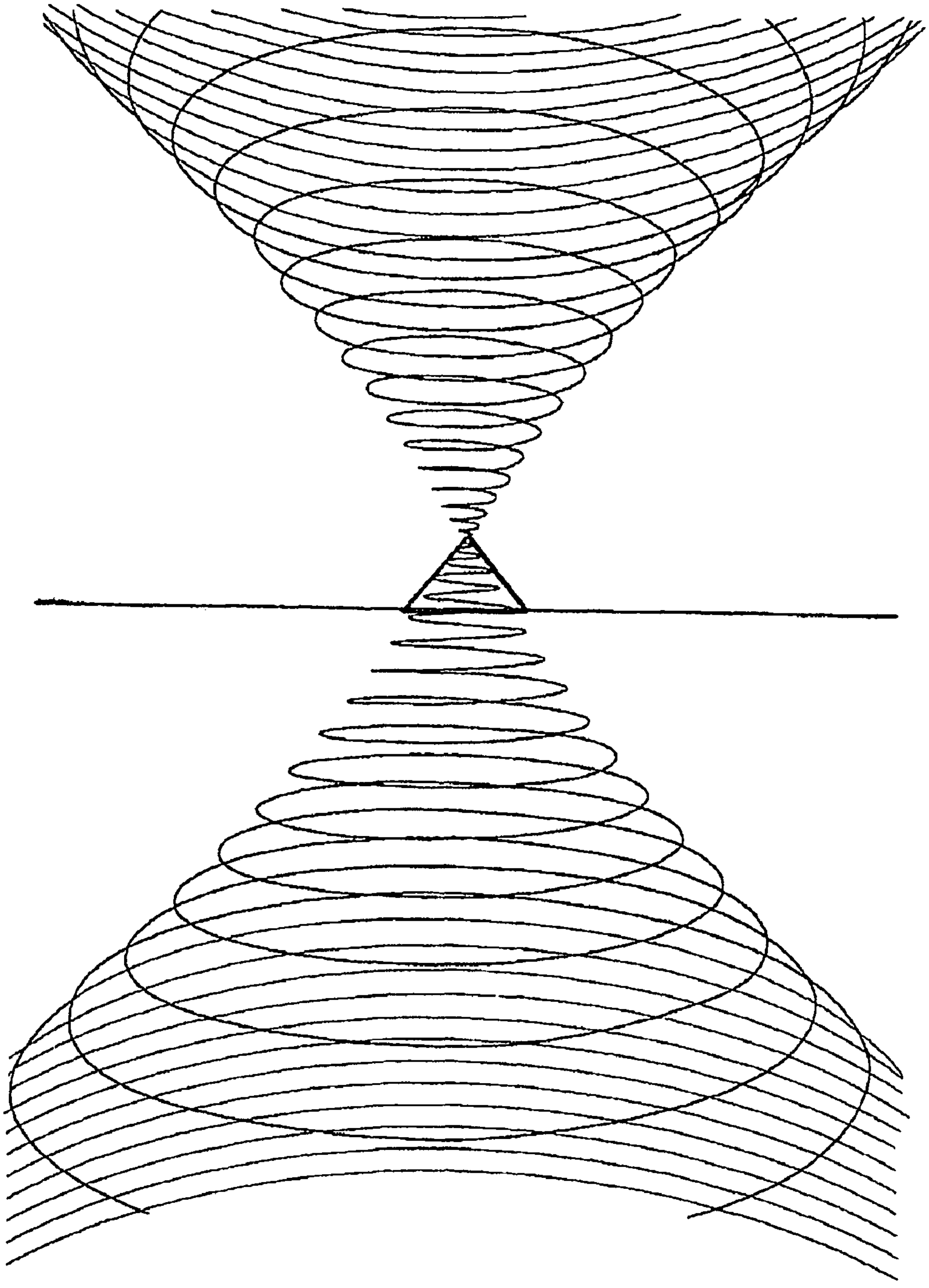

FIG. 1

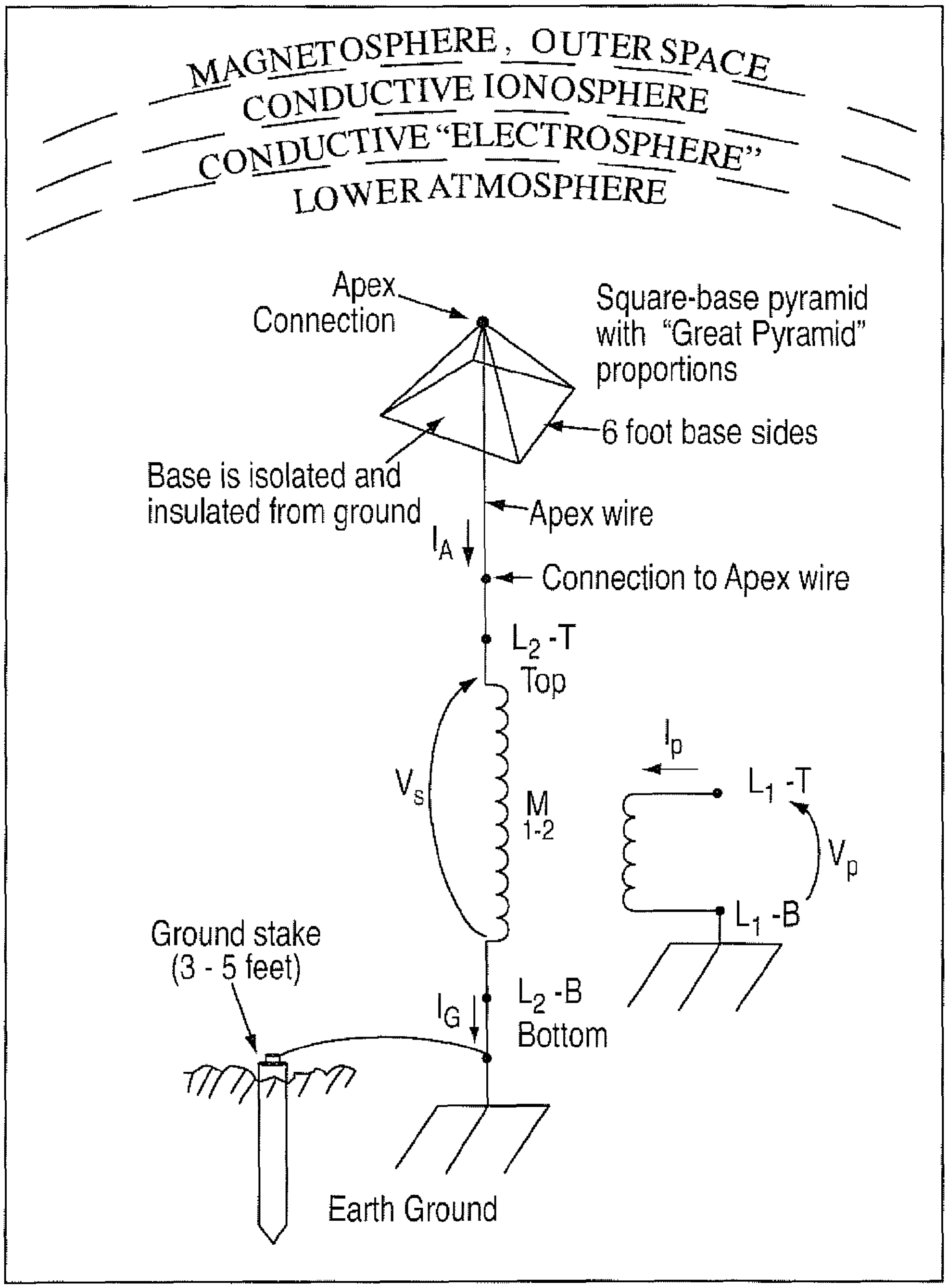
MAGNETOSPHERE, OUTER SPACE
CONDUCTIVE IONOSPHERE
CONDUCTIVE "ELECTROSPHERE"
LOWER ATMOSPHERE
Apex Connection
Square-base pyramid with "Great Pyramid" proportions
6 foot base sides
Base is isolated and insulated from ground
Apex wire
I_A
Connection to Apex wire
L_2 -T Top
V_s
M 1-2
I_p
L_1 -T
V_p
L_1 -B
Ground stake (3 - 5 feet)
I_G
L_2 -B Bottom
Earth Ground

V_p	Voltage across primary coil
I_p	Current into primary coil
I_A	Current in the apex wire
I_g	Current into ground wire
V_s	Voltage across secondary coil

FIG. 2

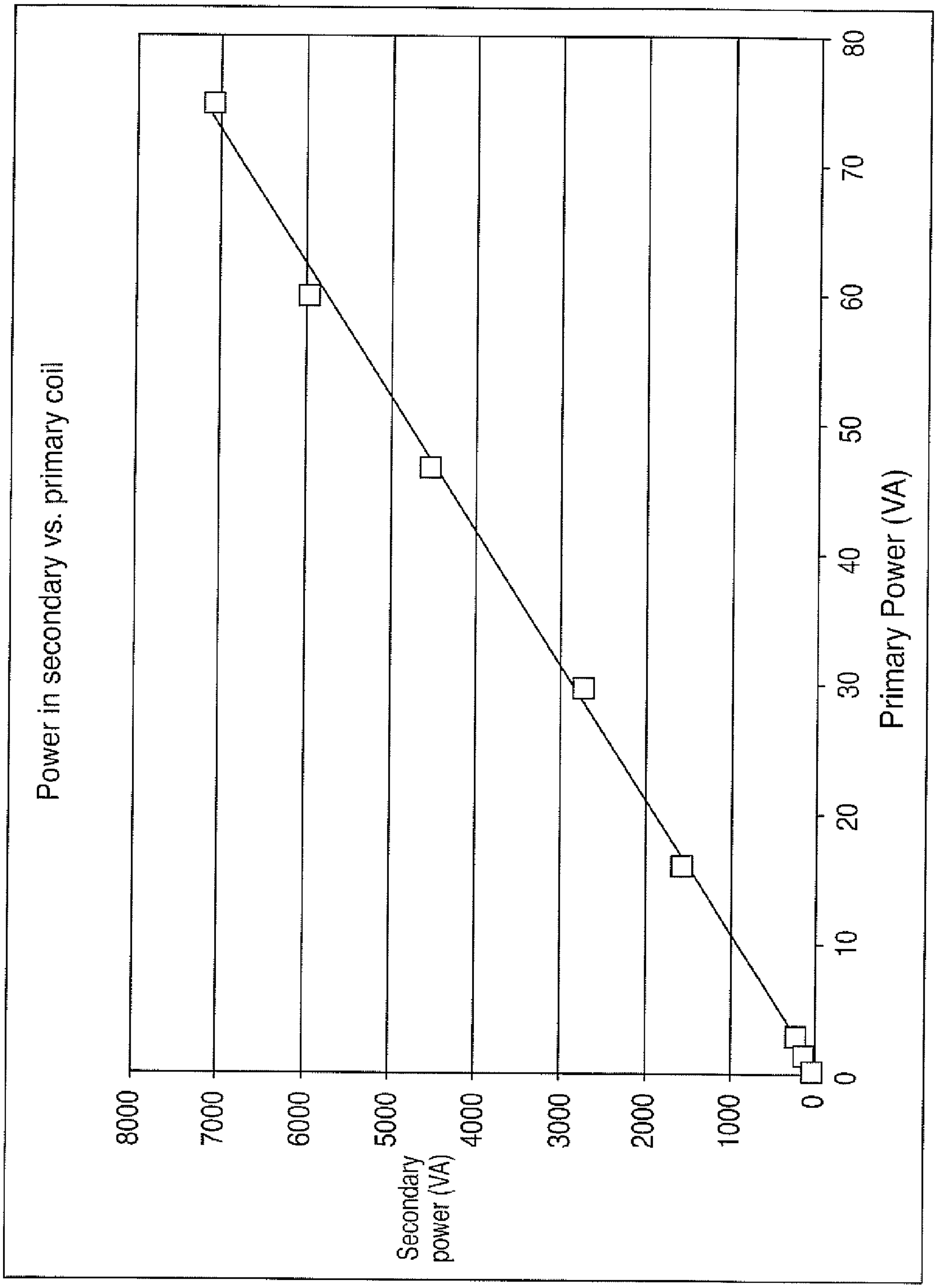
Power in secondary vs. primary coil
Secondary power (VA)
8000
7000
6000
5000
4000
3000
2000
1000
0
0
10
20
30
40
50
60
70
80
Primary Power (VA)

FIG. 3

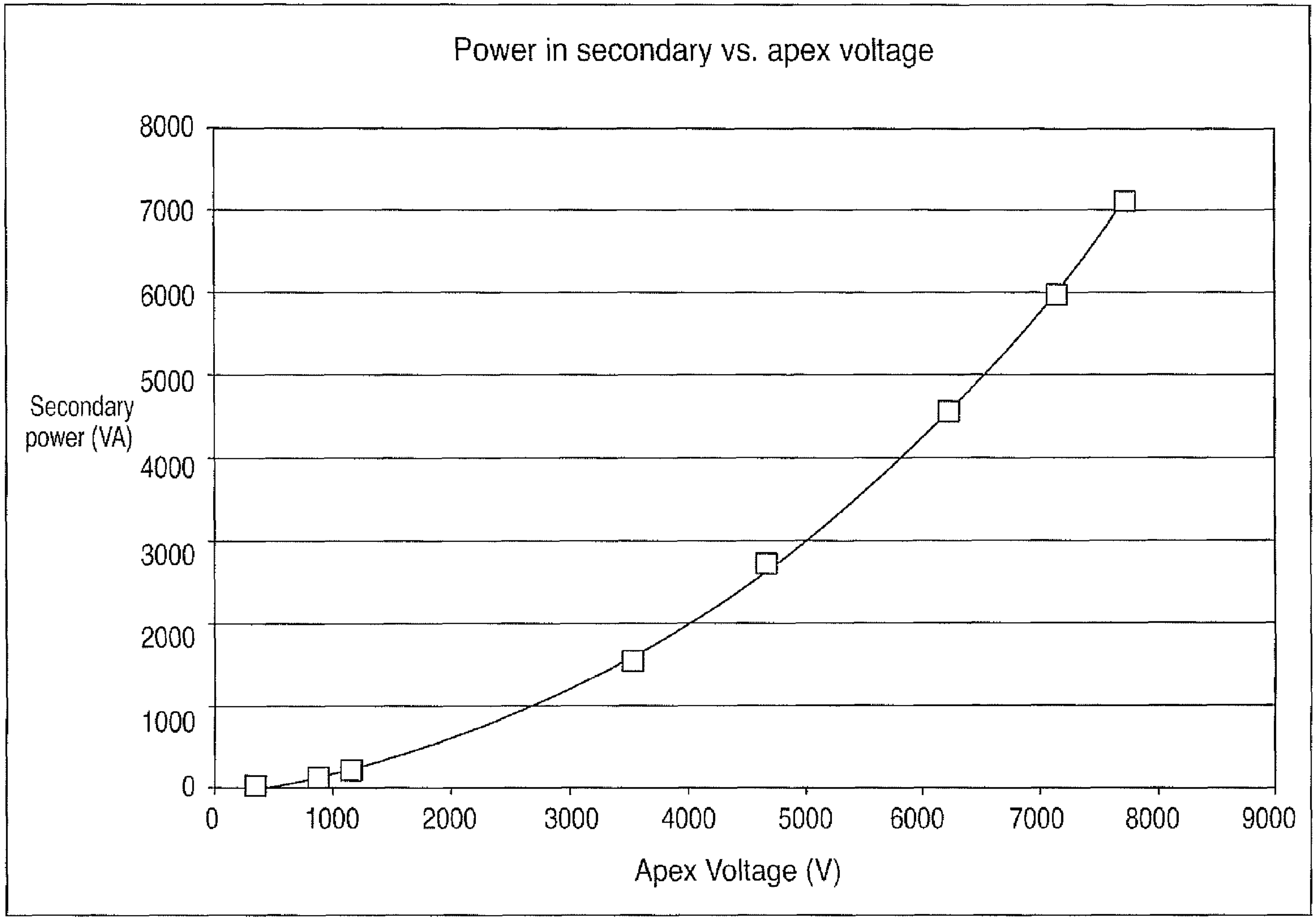
Power in secondary vs. apex voltage
Secondary power (VA)
8000
7000
6000
5000
4000
3000
2000
1000
0
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
Apex Voltage (V)

FIG. 4

FIG. 5

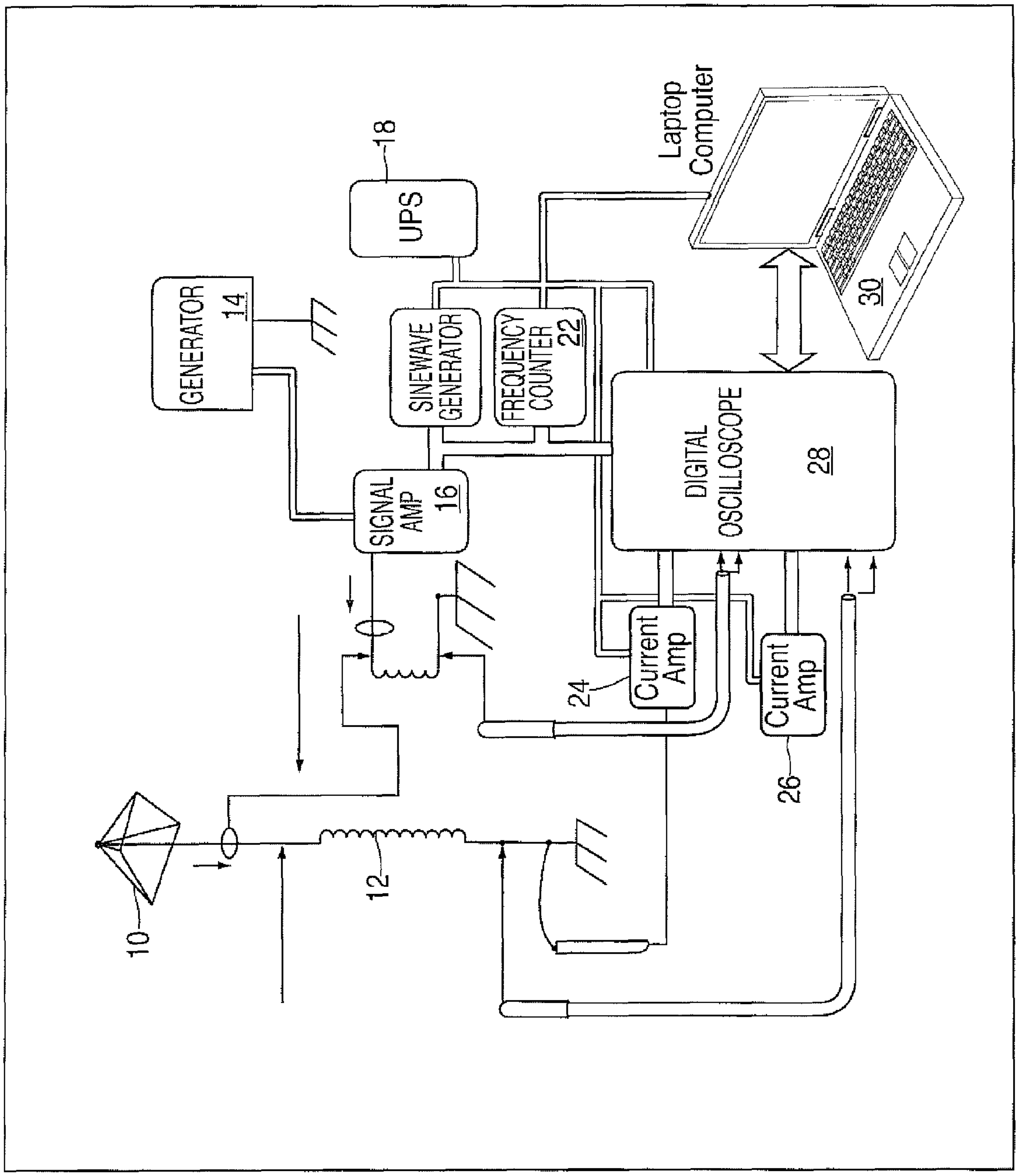
10
12
GENERATOR
14
SIGNAL AMP
16
UPS
18
SINEWAVE GENERATOR
FREQUENCY COUNTER
22
DIGITAL OSCILLOSCOPE
28
Current Amp
24
Current Amp
26
Laptop Computer
30

PYRAMID ELECTRIC GENERATOR

CROSS-REFERENCES

This application is a continuation-in-part of PCT Application Ser. No. PCT/US07/69804 by Grandics, entitled "An Electric Power Converter for Extraction of Atmospheric Electrical Energy," designating the United States and filed on May 25, 2007, which in turn claimed priority from U.S. Provisional Application Ser. No. 60/818,360 by Grandics, also entitled "An Electric Power Converter for Extraction of Atmospheric Electrical Energy," and filed on Jul. 3, 2006. The specifications of these two applications are incorporated herein in their entirety by this reference.

BACKGROUND OF THE INVENTION

This invention relates to the generation of electrical power by drawing energy from Earth's electric field. Earth represents a resonant system in which both atmospheric and telluric electrical processes can be harvested. In U.S. Pat. No. 6,974,110, we have provided an apparatus and method for converting the electrostatic potential energy of Earth's atmosphere. The present invention aims to provide an electric generator that, in addition to tapping atmospheric electric processes, extracts electrical power from the vibrational energy of atomic oscillators by a novel resonant coupling method.

SUMMARY OF THE INVENTION

An electrical generator according to the present invention provides a new method of tapping Earth's electric energy field.

One aspect of the present invention is an apparatus for capturing both atmospheric and telluric electric energies, comprising:

(1) an antenna/waveguide that is geometrically optimized, having a conducting surface;

(2) a primary coil wound with a conductor over a secondary coil, the coil being attached electrically to a driver operating in the LF radio band; and

(3) a secondary coil smaller in diameter than the primary coil having a greater length and a greater number of turns than the primary coil and having a first lead and a second lead, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding, inductively coupled with the primary coil and an external capacitor connected in parallel with the secondary coil to provide a specific resonant frequency; or

(4) a secondary coil smaller in diameter than the primary coil, having a greater length and a greater number of turns than the primary coil and having a first lead and a second lead, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding, inductively coupled with the primary coil and connected with the antenna/waveguide serving as a quasi-capacitive series element to provide a specific resonant frequency;

wherein the conducting surface of the antenna/waveguide is configured such that the secondary coil is connected near the point at which the electric field contacts the antenna/waveguide;

wherein the generator absorbs electromagnetic impulses from Earth's electric field; and wherein a high voltage sinusoidal waveform is generated in the secondary coil representing harvested atmospheric and telluric electrical energy and measurable on the leads of the secondary coil.

Preferably, the antenna/waveguide is of pyramidal shape. When the antenna/waveguide is of pyramidal shape, the secondary coil is preferably connected near or at the apex of the antenna/waveguide.

Another aspect of the present invention is an electric generator for harvesting the electric emissions of atomic oscillations comprising an oscillator. Typically, the oscillator operates in the LF or ELF bands.

Typically, in this aspect of the present invention, the generator comprises:

(1) a primary coil wound with a conductor over a secondary coil, the coil being connected electrically to a driver operating in the LF or ELF bands; and

(2) a secondary coil of smaller diameter than the primary coil having a greater length and a greater number of turns than the primary coil, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding inductively coupled with the primary;

wherein said resonant step-up transformer's output voltage exceeds 500 V_{RMS};

wherein the electric generator attracts impulses from Earth's electric oscillations; and

wherein the attracted energy manifests as high voltage sinusoidal waveforms representing harvested electric energy of atomic oscillations in the secondary coil and measurable on leads of the secondary coil.

Another aspect of the present invention is a method of tapping Earth's electric energy field, comprising these steps:

(1) positioning an electric generator according to the present invention as described above, so that it is exposed to a source of atmospheric and telluric electric energy; and

(2) producing a high-voltage sinusoidal signal representing electrical energy harvested by the operation of the generator.

BRIEF DESCRIPTION OF THE DRAWINGS

The following invention will be better understood with reference to the specification, appended claims, and accompanying drawings, where:

FIG. **1** is a graph showing the formation of electric vortices across and around the pyramid.

FIG. **2** is a graph showing the circuit diagram of the pyramid generator.

FIG. **3** is a graph showing the power in the secondary coil vs. the power in the primary coil.

FIG. **4** is a graph showing the power in the secondary coil vs. the apex voltage on the pyramid.

FIG. **5** is a graph showing the circuit diagram of the pyramid generator with the measurement points.

DETAILED DESCRIPTION OF THE INVENTION

One aspect of the present invention is an electric generator for attracting the electric emissions of atomic oscillations by the use of high voltage, to harvest electric energy therefrom.

One embodiment of this aspect of the invention is an apparatus for capturing atmospheric and telluric electric energies by harvesting the energies of atomic oscillators, comprising:

(1) an antenna/waveguide that is geometrically optimized, having a conducting surface;

(2) a primary coil wound with an insulated conductor, the coil being connected to an AC or AC-DC driver; and

(3) a secondary coil smaller in diameter than the primary coil, having a greater length and a greater number of turns than the primary coil and having a first lead and a second lead, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding,

inductively coupled with the first coil; and an external capacitor connected in parallel with the secondary coil to provide a specific resonant frequency; or

(4) a secondary coil smaller in diameter than the primary coil, having a greater length and a greater number of turns than the primary coil and having a first lead and a second lead, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding, inductively coupled with the first coil and connected with the antenna/waveguide serving as a quasi-capacitive series element to provide a specific resonant frequency, with the secondary coil attached near the point at which the electric field contacts the antenna/waveguide;

wherein the generator absorbs electromagnetic impulses from Earth's electric field; and

wherein a high-voltage, sinusoidal waveform is generated in the secondary coil representing harvested atmospheric and telluric electrical energy and measurable on leads of the secondary coil.

Preferably, the antenna/waveguide is of pyramidal shape. However, any geometric shape can be used for the antenna/waveguide. When it is of pyramidal shape, the secondary coil is preferably connected near or at the apex of the antenna/waveguide. The shape and size of the antenna/waveguide can be optimized depending on the intended use of the generator. Preferably, the antenna/waveguide exhibits ratios of pi and the Fibonacci number (phi).

Typically, the generator further comprises an insulated base on which the antenna/waveguide is placed. The size and shape of the insulated base can also be optimized, depending on the intended use of the generator. The second lead of the secondary coil is typically connected to earth ground.

Typically, the output of the secondary coil is harvested. This can be done by connecting to a rectifier-capacitor-load resistance, or by a third coil inductively coupled with the secondary coil. In one alternative, the output of the secondary coil is coupled to a load for drawing power from the generator. The optimal circuit can be selected by one of ordinary skill in the art to optimize power extraction. The load draws power from the generator. The load can be a resistor, a rectifier or a storage capacitor powering a DC load. Typically, the antenna/waveguide is at a high voltage during the operation of the generator. Typically, the antenna/waveguide voltage is greater than 500 V_{RMS}. The antenna/waveguide can be connected to an oscillator. The oscillator can comprise a resonant step-up transformer.

Typically, the antenna/waveguide is in resonance with the secondary coil. Typically, the primary coil is inductively coupled to the secondary coil.

Typically, the antenna/waveguide is positioned in a generally north-south direction to tap atmospheric and telluric electric energy. In this arrangement, when the antenna/waveguide is of pyramidal shape, the height of the pyramid can be from about 0.5 m to about 1000 m, more typically, from about 0.50 m to about 150 m. The optimal pyramid size is a function of the power requirements. However, a pyramid can be built even 1 km in height. Typically, the base surface area of the pyramid ranges from about 0.25 m^2 to about 1,000,000 m^2; more typically, from about 0.25 m^2 to about 25,000 m^2. Naturally, other heights and base surface areas can be used for particular applications. The numbers of sides can also vary. When the pyramid is small (e.g., 6-feet base length described in this application), there is an insufficient atmospheric potential on its apex to trigger the attraction of power. In such event, an AC or AC-DC driver is necessary to bring the pyramid to the electric potential necessary. A suitable voltage is typically greater than 500 V_{RMS}.

In one alternative, the primary coil is wound with a conductor over the secondary coil, the primary coil being connected electrically to a driver operating in the LF or ELF bands.

Another aspect of the present invention is an electric generator for harvesting the electric emissions of atomic oscillations comprising an oscillator. Typically, the oscillator operates in the LF or ELF bands.

Typically, in this aspect of the present invention, the generator comprises:

(1) a primary coil wound with a conductor over a secondary coil, the coil being connected electrically to a driver operating in the LF or ELF bands; and

(2) a secondary coil of smaller diameter than the primary coil having a greater length and a greater number of turns than the primary coil, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding inductively coupled with the primary;

wherein said resonant step-up transformer's output voltage exceeds 500 V_{RMS};

wherein the electric generator attracts impulses from Earth's electric oscillations; and

wherein the attracted energy manifests as high voltage sinusoidal waveforms representing harvested electric energy of atomic oscillations in the secondary coil and measurable on leads of the secondary coil.

In this alternative, the output of the secondary coil can be coupled to a load for drawing power from the generator.

Another aspect of the invention is a method of tapping Earth's electric energy based on the following concepts:

(1) utilizing a pyramidal antenna/waveguide designed based on the ratios of pi and phi;

(2) using the antenna/waveguide to absorb and focus specific frequencies of Earth's electric field;

(3) positioning a resonant coil system connected to said antenna/waveguide near or at its apex according to the present invention; and

(4) generating a high-voltage sinusoidal signal, representing harvested atmospheric and telluric electrical energy, by the operation of the coil system.

Typically, in this method, the method comprises the steps of:

(1) positioning an electric generator according to the present invention as described above such that it is exposed to Earth's electric oscillations; and

(2) generating a high voltage sinusoidal signal representing harvested electric energy of atomic oscillators by the operation of the generator.

Typically, the harvested electric energy can be fed into a power grid for distribution, but smaller units capable of distributed power generation or for use as stand-alone generators are also feasible by the concept of this invention.

The invention is described by the following Example. This Example is included for illustrative purposes only and is not intended to limit the invention.

EXAMPLE

We have introduced a new theory of space, energy and matter that predicted that electromagnetic interaction propagates through a carrier medium called the space lattice that is made up of pulsating, phi-based spiral vortices (STAR), the excitation of which generates propagating sinusoidal lines of force across the space lattice recorded as electromagnetic waves by our instruments [1,2]. We have also predicted that natural pressure gradients exist within the space lattice that can be used for power generation by inserting a capacitor into such gradient. Furthermore, we proposed that a pyramid-shaped capacitor is an optimally shaped device for tapping the energy gradient of the space lattice that manifests in the form of an electrical potential gradient in Earth's atmosphere. This was demonstrated subsequently [3-5]. We also found that the pyramid-shaped capacitor design should be based on ratios of pi and phi [5] because electrical energy propagates along a phi-based spiral [6].

We suggested that elemental particles are composed of STAR oscillators that make up the space lattice by electromagnetic compression of its elemental cubes into pulsating conical/vortexual subunits that form the basis for all particles of matter [1,2]. We proposed that atomic and subatomic oscillators could vibrate in a very wide frequency range from sub-acoustic to extremely high EM frequencies [1,2]. This does not mean that material bodies can take or maintain a physical form at any energy level, but this allows the recycling of atoms into the space lattice as part of their natural evolution [2,6].

Atomic oscillators are constantly in motion, and such motion is electric because all motion and energy is fundamentally electric [2,6]. It is known that atomic bodies constantly exchange electric energy. Such energy flow can be harvested when the spiral physical nature of electric waves is recognized. Russell observed that energy always moves during its generative cycle in spirals towards the higher potential [7]; therefore, we reasoned that an "attractor" of high electric potential must be provided to pull in the random electromagnetic emissions of atomic oscillators. The "attractor" must function as a phi-based antenna/waveguide to focus the phi-based electric emissions of atomic bodies into the apex of a vortex, the same method Nature uses for power multiplication [6,7]. Therefore, a phi-based pyramidal antenna/waveguide must be at a high voltage in order to perform its attractor function.

As all material manifestations are tonal [8], acoustic vibrations are always generated during the oscillations of atomic bodies, even though they are inaudible most of the time. Sometimes, however, they are detected; one such example is Earth's hum [9-11]. Earth's solid mass and its atmosphere represent a coupled resonant system, and is therefore our expanded target power source. Acoustic waves are also electrical [8] and can be harvested. To achieve this goal, the geometrically optimized pyramidal antenna/waveguide must be connected to an oscillator that operates at a high voltage and is tuned to a suitably selected frequency.

The appropriate resonant frequency is dictated by practical considerations. Atomic oscillators can vibrate in a very wide frequency range [1,2,6-8] but short wavelength radiations are readily re-absorbed by atoms and are rapidly attenuated; therefore, long wavelength electromagnetic emissions must be targeted. The operation of the pyramid generator is feasible in the ELF to the LF range. As electric radiation propagates in a vortex [2,6-8], the emitted electric vortex over the pyramidal antenna (FIG. **1**) increases its "size," and so it is capable of funneling atmospheric electromagnetic emissions of the same frequency back into the antenna/receiver. Thus, the pyramid electric generator could capture the energy emitted by both telluric and atmospheric atomic oscillators. We have tested these assumptions and demonstrated the basic principles of an electric generator.

Methods and Results

In our previous papers, we reported that a pyramid-shaped capacitor/antenna converts atmospheric electrostatic discharge impulses (ESD) into a periodic high-frequency signal that can be detected in an insulated coil placed in proximity of the capacitor [3,4]. Here, we report on further developments including a new design for the resonant coil system and an expanded range of energy sources targeted. A Tektronix TPS 2024 digital oscilloscope was used for signal acquisition and analysis, and Tektronix A503 current amplifiers were used with the Tektronix A6302 and A6303 probes for current measurement. High voltage was measured using a Tektronix P6015 probe. An ENI-1140LA amplifier was used to activate the primary coil of the resonant coil system at the selected frequency.

For the experiments, we used a geometrically optimized pyramid-shaped antenna/waveguide [3-5]. The pyramid expressed ratios of pi and phi and was approximately a 1:125 scaled down replica of the Great Pyramid of Giza (GPF), 6 feet base length. The pyramid was built of a wooden frame and covered with triangular copper sheets of 0.66 mm thickness on its sides that were electrically connected. It was placed on an insulating base and positioned in the general North-South direction.

The secondary coil (coil **2**) was wound with a 20 AWG insulated magnet wire with a monofilament nylon spacer (0.06 mm diameter) between the turns on a fiberglass cylindrical coil form, and connected to the conducting surface of the pyramid near its apex. Thus, the secondary coil was connected with the pyramidal antenna/waveguide as a quasi-capacitive series element to provide a specific resonant frequency with coupling to the surrounding electromagnetic environment. The other lead of coil **2** was grounded (FIG. **2**). The purpose of the monofilament spacer was to reduce inter-winding capacitance in coil **2** by increasing distance between wire turns of the coil.

A primary coil of larger diameter (coil **1**) was wound on or near the lower cylindrical volume of coil **2** with the lead connected to earth ground, and served as a driver coil inductively coupled with the secondary (coil **2**). The secondary was isolated from the primary by layers of silicon rubber and foam. The two coils make a resonant step-up transformer, the parameters of which are shown in Table 1.

TABLE 1

Physical and Electrical Parameters of the Resonant Transformer	
Secondary turns of wire	840 turns
Height	0.85 m
Wire diameter	0.00081 m
Primary turns of wire	5 + 9/10 turns
Height	0.21 m
Wire diameter	¼" O.D. copper pipe with .030" wall thickness
Primary	
Inductance	98 μH
Resistance	0.118 ohms
Secondary	
Inductance	24.16 mH
Resistance	24.42 ohms
Capacitance	1040 nF
Winding ratio	1:143
Mutual inductance	6.13 mH

In the experiment; the controlled variable was the voltage, which was increased in steps until the maximum voltage in the secondary (permitted by circuit components and test equipment) was achieved for the given resonant frequency. The objective was to "drive the atmosphere" by radiating a high voltage (~20 kV p-p), LF (40-120 kHz) signal developed by the high ratio step-up transformer and fed by a powerful signal generator at a resonant frequency either with an L-C resonant secondary coil, or a resonance determined by the pyramid as a quasi-capacitive series element connected to the secondary coil of the transformer and radiating to the local atmosphere. The data are shown in Table 2.

The experimental data are shown in Table 2. The current readings were also confirmed by using "current sensing" resistors (not shown). A apparent power gain of nearly 100-fold was achieved in the secondary coil relative to input power in the primary coil, and a linear increase of power in the secondary coil with the increase of input power into the primary coil (FIG. **3**). The power in the secondary coil exponentially increases with the pyramid apex voltage (FIG. **4**).

TABLE 2

Power Measurements in the Pyramid-Resonant Coil System

Primary								
Frequency (kHz)	83.00							
$V_{P\ RMS}$	3.97	6.38	9.86	20.3	26.7	33.8	39	40.8
$I_{P\ RMS}$ (mA)	87.6	222	300	792	1117	1389	1541	1838
$I_{P\ angle\ with\ respect\ to}\ V_P$	18°	24°	14°	22°	23°	21°	20°	28°
Total power (VA)	0.35	1.42	2.95	16.02	29.70	46.77	59.95	74.75
True power (W)	0.33	1.30	2.87	14.85	27.35	43.66	56.34	66.01
Secondary								
$V_{S\ RMS}$	341	872	1150	3540	4670	6220	7140	7710
$I_{A\ RMS}$ (mA)	62.2	148	193	438	579	728	834	919
$I_{A\ angle\ with\ respect\ to}\ V_S$	42°	43°	44°	52°	52°	54°	53°	52°
Total power (VA)	21.2	129.5	222.8	1552	2707.8	4530.2	5957.5	7087.3
True power (W)	15.7	94.7	160.3	956	1668.1	2664.5	3587.6	4366
Power sec/prim	61.1	91.4	75.4	96.9	91.2	96.9	99.4	94.8

A load may be connected to coil **2** to draw power from the system. The load may also be connected via a tertiary coil inductively coupled to the secondary coil. The load may be a resistor, a rectifier or storage capacitor powering a DC load.

A graphic representation of the test system is shown in FIG. **5**. The pyramidal antenna/waveguide, placed on an insulating base, is coupled to the surrounding electromagnetic environment and serves as an antenna/waveguide for the concentration of atmospheric and telluric electromagnetic oscillations at resonance frequency. As the flow of electric energy is always balanced [6], two vortices may cross the pyramidal antenna: one from the direction of the ground concentrating into its apex, and another one from its apex propagating into the atmosphere (FIG. **1**). The atmospheric electric vortex possibly acts as an ionic antenna pulling in EM radiations from a large atmospheric domain. In FIG. **5**, the antenna/waveguide **10** is attached to a coil **12**. To drive the system, a RF generator **14** is connected to a signal amplifier **16**. The signal amplifier **16** is also connected to an uninterruptible power supply (UPS) **18** through a sinewave generator **20** and a frequency counter **22**. Outputs from the sinewave generator **20** and frequency counter **22** are also connected to a first current amplifier **24**. The output from the antenna/waveguide **10** and coil **12** are also connected to a second current amplifier **26**. The sinewave generator **20**, frequency counter **22**, first current amplifier **24**, and second current amplifier **26** are connected to a digital oscilloscope **28**. Outputs from the antenna/waveguide **10** and coil **12** are also connected to the digital oscilloscope **28**. A laptop computer **30** controls the sinewave generator **20** and frequency counter **22**.

The pyramid's apex is attached to a coil of high turn number (FIGS. **2** & **4**). This secondary coil wound on a nonconductive coil form serves as a step-up transformer, and forms a resonant circuit with the pyramid that acts as a quasi-capacitive series element. The secondary coil is activated by a signal from an RF generator via a primary coil of a few turn numbers inductively coupled to coil **2**. The measurement points are also indicated, displaying the positions of voltage and current probes. The data are captured by a digital oscilloscope and recorded on a laptop computer.

Discussion

This invention demonstrates a novel approach to harvest Earth's electric energy. The masses of Earth and its atmosphere represent a coupled resonant system that is continually electrified by solar radiation. We have found that a pyramidal antenna designed based on the ratios of phi and pi is optimal for the capture of atmospheric electrostatic discharge (ESD) impulses [3-5]. Atmospheric ESD is a product of solar radiation and is a wide bandwidth phenomenon. The pyramid as a wideband, non-resonant antenna is uniquely adapted to harvest the energy of atmospheric ESD.

This invention also points out an additional energizing mechanism at work during the operation of the pyramid electric generator. The observation is derived from our theory on energy, matter and space [1,2] as well as that of Russell [6,7]. Both Russell and we observed that spiral motion is a fundamental action of matter and that the vortex is the mechanism of power multiplication in Nature. A vortex concentrates power into its apex where the highest velocity of motion, the highest pressure and the highest electric potential resides [1,2,6,7]. As electricity propagates along a pulsating phi-based spiral [6,7], a phi-based antenna/waveguide is suitable to focus its energy into the apex of a vortex inside the antenna. A phi-based pyramid is optimal for this purpose.

As atomic bodies can oscillate in a very wide frequency range [1,2,6-8], a pyramidal antenna/waveguide coupled to an oscillator that is tuned to a suitable frequency can focus the "electric noise" from Earth's atomic oscillators into the apex of the pyramid at a high voltage. Russell observed that energy moves towards the higher potential during its generative cycle [7]; therefore, it is essential that the pyramidal antenna be at a high potential. Our data corroborated this assumption.

To select a suitable frequency range, one can begin with the observation that all material manifestations are tonal [8], and that acoustic frequencies are always generated during the oscillations of atomic bodies even though they remain inaudible most of the time. Acoustic waves are also electric [8] and therefore Earth's enormous mass and its atmosphere represent a coupled resonant system that offer a vast source of power.

For practical reasons on our small pyramid, we have chosen a frequency above the audio range but below the radio frequency spectrum; our antenna was made to radiate at the target wavelength. The wavelength of the 50-120 kHz range is 6 km to 25 km, thus allowing energy to be harvested over a large atmospheric domain. As electrical energy propagates in a vortex [6,7], the emitted electric vortex (ionic antenna) over the pyramidal antenna/waveguide increases its "size" and enables it to funnel the same frequency atmospheric electromagnetic radiations back into the antenna. Thus, the pyramid electric generator captures at the selected frequency the energy emitted by Earth's atomic oscillators.

The presence of an electric (ionic) vortex was demonstrated by radar testing of the space over a 44 m tall fiberglass pyramid located near Moscow [12]. The Russians found that the large ionized column of air over the area of the vertical axis of the pyramid had a width of about 500 m and reached an altitude of 2 km. It is remarkable that this effect was induced by a nonconductive pyramid surface demonstrating a significant degree of atmospheric ionization even under fair weather conditions. Thus, a suitably sized large or an electrically activated small pyramid should open a low impedance path to higher elevations of relatively conductive atmospheric domains.

In conclusion, the total power that can be extracted from Earth's atomic oscillators must be extremely large, likely far exceeding current global electric generation capacity. In our experiment, we obtained over 7000 VA power at about 20 kV p-p pyramid voltage using a 6 feet base length pyramid. This power is nearly 100 times greater than the power necessary to drive the pyramid to the required operating voltage. By constructing the pyramid generator from dedicated high voltage components and using higher drive voltages, a significant increase in power output is envisioned. The power extraction will also be enhanced by further optimization of the designs of both the pyramid and the coil system.

The power output will also increase by employing larger pyramid structures and coils. As voltage is the primary factor in attracting power into the system, the necessary voltage can be provided by the vertical atmospheric potential on a tall pyramid. Since the atmospheric vertical potential gradient could go as high as 1200 V/m near Earth's surface under fair weather conditions [13], a pyramid height of 80-150 m seems sufficient to provide the apex voltage for a self-sustaining power generator. As the pyramid scales up volumetrically, a power generator pyramid of the size range of the GPG could likely have an output in the hundreds to thousands of megawatts range. Groups of several pyramid electric generators could be placed within specific geographical areas, thus combining their energy harvesting capacity.

References

The following references are specifically applicable to the Example and are incorporated herein by reference; these references are referenced in the Example by the reference numbers assigned to them.

[1] Grandics, P. 2002, "The genesis of electromagnetic and gravitational forces." *J. New Energy,* 6, (3) 33-45.

[2] Grandics, P. 2007, "The genesis of fundamental forces acting at a distance and some practical derivations." *Infinite Energy,* 12, (71) 13-24.

[3] Grandics, P. 2000, "A method to capture atmospheric electrostatic energy", in *Proceedings of IEJ-ESA Joint Symposium on Electrostatics*, Kyoto University, Kyoto, Japan, pp. 355-361.

[4] Grandics, P. 2006, "A DC to RF converter for the capture of atmospheric electrostatic energy", in *Proceedings of the* 5_{th} *Conference of the Societe Francaise D'Electrostatique, SFE* 2006, Grenoble, France pp. 279-284.

[5] Grandics, P. 2007, "Pyramidal Electric Transducer: A DC to RF Converter for the Capture of Atmospheric Electrostatic Energy." *Infinite Energy,* 13, (73) 20-27.

[6] Russell, W. 1947, in *The Secret of Light,* 3_{rd} ed. University of Science and Philosophy, Waynesboro, Va., pp. 218-288.

[7] Russell, W. 1957, in *Atomic Suicide?* 2_{nd} ed. University of Science and Philosophy, Waynesboro, Va., pp. 129-153.

[8] Russell, W. 1926, in *The Universal One*, University of Science and Philosophy, Waynesboro, Va., pp. 247-249.

[9] Nishida, K., Kobayashi, N. and Fukao, Y. 2000, "Resonant oscillations between the solid earth and the atmosphere", *Science* 287, (5461), 2244-2246.

[10] Tanimoto, T., Um, J., Nishida, K., and Kobayashi, N. 1998, "Earth's continuous oscillations observed on seismically quiet days", *Geophys. Res. Lett.* 25, 1553-1556.

[11] Tanimoto, T. 2001, "Continuous free oscillations: Atmosphere-solid earth coupling". *Ann. Rev. Earth Planet Sci.* 29, 563-584.

[12] http://www.pyramidoflife.com/eng/tests_experiments.html.

[13] Marshall, T. C., Rust, W. D., Stolzenburg, M., Roeder, W. P., and Kriebel, P. R. 1999, "A study of enhanced fair-weather electric fields occurring soon after sunrise". *J. Geophys. Res.* 104, 24,455-24, 469.

ADVANTAGES OF THE INVENTION

The present invention provides a new method of tapping Earth's electric energy and providing usable power that can be fed into a power grid or supplied by stand-alone power generators. It does so without the need for mechanical energy or the consumption of fossil fuel or the long-term risks associated with power generated by nuclear fission, including the risk of diversion of fissionable material to terrorist aims or the risk posed by the required long-term storage of spent nuclear fuel. Devices according to the present invention can operate virtually continuously with little to no maintenance.

With respect to ranges of values, the invention encompasses each intervening value between the upper and lower limits of the range to at least a tenth of the lower limit's unit, unless the context clearly indicates otherwise. Moreover, the invention encompasses any other stated intervening values and ranges including either or both of the upper and lower limits of the range, unless specifically excluded from the stated range.

Unless defined otherwise, the meanings of all technical and scientific terms used herein are those commonly understood by one of ordinary skill in the art to which this invention belongs. One of ordinary skill in the art will also appreciate that any methods and materials similar or equivalent to those described herein can also be used to practice or test this invention.

The publications and patents discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

All the publications cited are incorporated herein by reference in their entireties, including all published patents, patent applications, and literature references, as well as those publications that have been incorporated in those published documents. However, to the extent that any publication incorporated herein by reference refers to information to be published, applicants do not admit that any such information published after the filing date of this application is to be considered prior art.

As used in this specification and in the appended claims, the singular forms include the plural forms. For example the terms "a," "an," and "the" include plural references unless the content clearly dictates otherwise. Additionally, the term "at least" preceding a series of elements is to be understood as referring to every element in the series. The inventions illustratively described herein can suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms "comprising," "including," "containing," etc. shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the future shown and described or any portion thereof, and it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the inventions herein disclosed can be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of the inventions disclosed herein. The inventions have been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the scope of the generic disclosure also form part of these inventions. This includes the generic description of each invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised materials specifically resided therein. In addition, where features or aspects of an invention are described in terms of the Markush group, those schooled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group. It is also to be understood that the above description is intended to be illustrative and not restrictive. Many embodiments will be apparent to those of in the art upon reviewing the above description. The scope of the invention should therefore, be determined not with reference to the above description, but should instead be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled. Those skilled in the art will recognize, or will be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described. Such equivalents are intended to be encompassed by the following claims.

I claim:

1. An electric generator for harvesting the energies of atomic oscillators comprising:

(a) an antenna/waveguide that is geometrically optimized having a conducting surface;

(b) a primary coil wound with an insulated conductor over a secondary coil, the coil being connected to an AC or AC-DC driver; and

(c) a secondary coil of smaller diameter than the primary coil having a greater length and a greater number of turns than the primary coil and having a first lead and a second lead, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding inductively coupled with the primary and an external capacitor connected in parallel with the secondary coil to provide a specific resonant frequency; or

(d) a secondary coil of smaller diameter than the primary coil having a greater length and a greater number of turns than the primary coil and having a first lead and a second lead, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding inductively coupled with the primary and connected with the antenna/waveguide serving as a quasi-capacitive series element to provide a specific resonant frequency;

wherein the conducting surface of the antenna/waveguide is configured such that the secondary coil is attached near the point at which the electric field contacts the antenna/waveguide;

wherein the generator absorbs impulses from Earth's electric oscillations; and wherein the attracted energy manifests as high voltage sinusoidal waveforms in the secondary coil, representing harvested atmospheric and telluric electrical energy and measurable on the leads of the secondary coil.

2. The electric generator of claim **1**, wherein a suitable voltage is greater than 500 V_{RMS}.

3. The electric generator of claim **1**, wherein the antenna/waveguide is designed based on ratios of pi and phi.

4. The electric generator of claim **1**, wherein the antenna/waveguide is of pyramidal shape.

5. The electric generator of claim **1**, wherein the antenna/waveguide is placed on an insulated base.

6. The electric generator of claim **1**, wherein the antenna/waveguide is at a high voltage during the operation of the generator.

7. The electric generator of claim **6**, wherein the antenna/waveguide voltage is greater than 500 V_{RMS}.

8. The electric generator of claim **1**, wherein the antenna/waveguide is connected to an oscillator.

9. The electric generator of claim **8**, wherein the oscillator comprises a resonant step-up transformer.

10. The electric generator of claim **4**, wherein the secondary coil of the transformer is connected near the apex of the antenna/waveguide.

11. The electric generator of claim **1** wherein the second lead of the secondary coil is connected to earth ground.

12. The electric generator of claim **1** wherein the primary coil is wound with a conductor over the secondary coil, the primary coil being connected electrically to a driver operating in the LF or ELF bands.

13. The electric generator of claim **1**, wherein the primary coil is inductively coupled to the secondary coil.

14. An electric generator for harvesting the electric emissions of atomic oscillations of claim **1** comprising an oscillator.

15. The electric generator of claim **14**, wherein said oscillator operates in the LF or ELF bands.

16. The electric generator of claim **14** wherein the generator comprises:

(a) a primary coil wound with a conductor over a secondary coil, the coil being connected electrically to a driver operating in the LF or ELF bands; and

(b) a secondary coil of smaller diameter than the primary coil having a greater length and a greater number of turns than the primary coil, the secondary coil being positioned coaxially within the first coil and acting as a resonant step-up transformer winding inductively coupled with the primary; wherein said resonant step-up transformer's output voltage exceeds 500V_{RMS};

wherein the electric generator attracts impulses from Earth's electric oscillations; and wherein the attracted energy manifests as high voltage sinusoidal waveforms representing harvested electric energy of atomic oscillations in the secondary coil and measurable on leads of the secondary coil.

17. The electric generator of claim **4**, wherein the height of the pyramidal electric generator is from about 0.50 m to about 1000 m.

18. The electric generator of claim **4**, wherein the base surface area of the pyramid is from about 0.25 m^2 to about 1,000,000 m^2.

19. The electric generator of claim **1**, wherein the generator is positioned in a generally north-south direction.

* * * * *

US 20160029467A1

(19) **United States**

(12) **Patent Application Publication**

LANE

(10) **Pub. No.: US 2016/0029467 A1**

(43) **Pub. Date: Jan. 28, 2016**

(54) **CHARGED PARTICLE INDUCTION FROM IONOSPHERE TO GROUND**

(71) Applicant: **Glenn Lane Family Limited Liability Limited Partnership**, Summerfield, FL (US)

(72) Inventor: **GLENN E. LANE**, SUMMERFIELD, FL (US)

(21) Appl. No.: **14/829,079**

(22) Filed: **Aug. 18, 2015**

Related U.S. Application Data

(63) Continuation of application No. 14/197,698, filed on Mar. 5, 2014, now Pat. No. 9,112,435, which is a continuation of application No. 12/917,447, filed on Nov. 1, 2010, now Pat. No. 8,693,160.

(60) Provisional application No. 61/280,131, filed on Oct. 31, 2009.

Publication Classification

(51) **Int. Cl.**

H05F 7/00	(2006.01)
H02N 99/00	(2006.01)

(52) **U.S. Cl.**

CPC . ***H05F 7/00*** (2013.01); ***H02N 99/00*** (2013.01)

(57) **ABSTRACT**

A charged particle induction apparatus and method comprising a high power light emitting means, such as a laser array, in operable communication with a high energy output means to accomplish initiation of at least two concentric plasma channels in atmosphere extending from the Earth's surface to the charge-rich upper atmosphere, including the ionosphere, for the transmission of charged particles therethrough to ground using the surrounding atmosphere as an insulator. The transmitted energy is drawn down (due to an artificially created potential) through the conductive plasma channels to collection means.

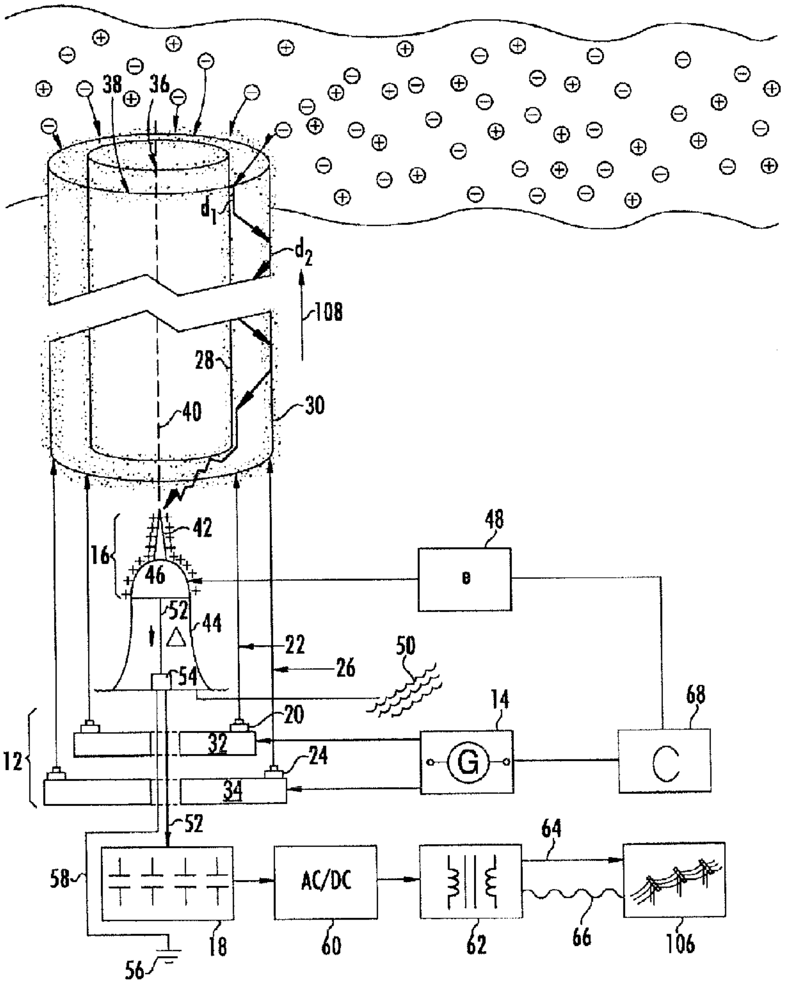

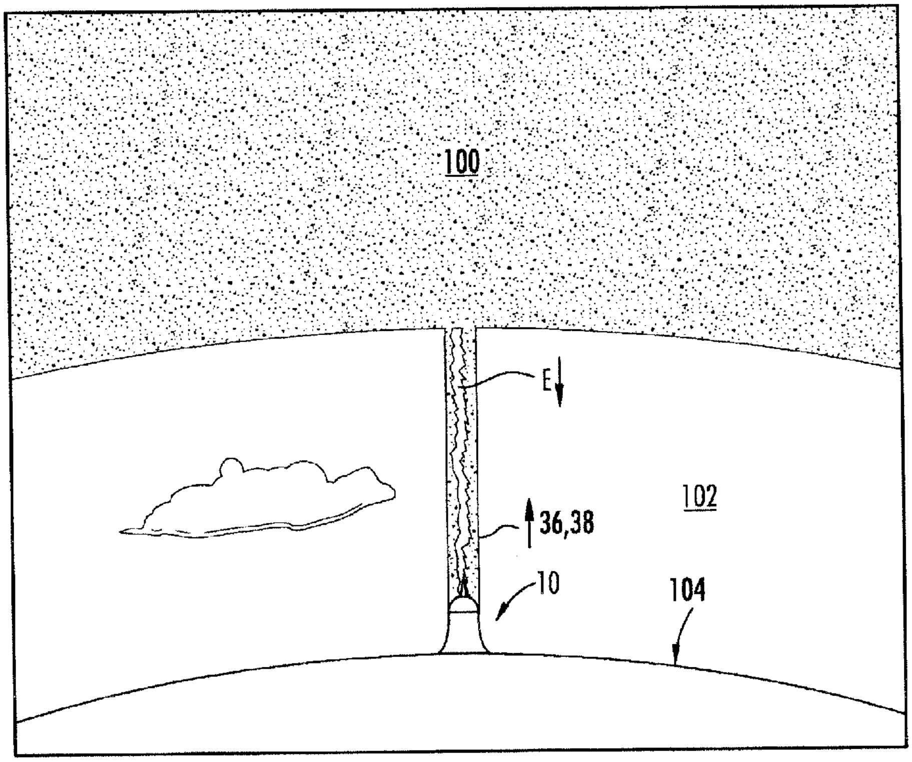
100
E
102
36,38
10
104

FIG. 1

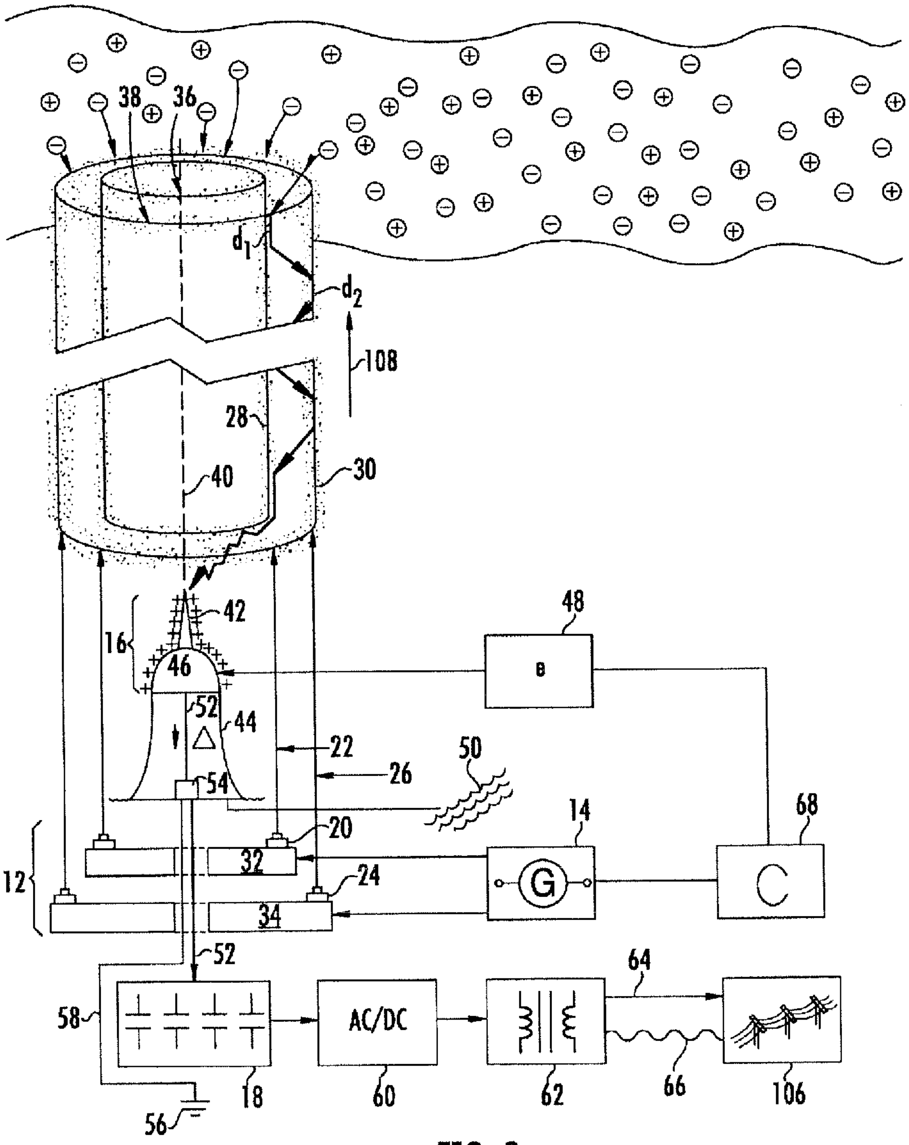
d1
d2
B
G
C
AC/DC

FIG. 2

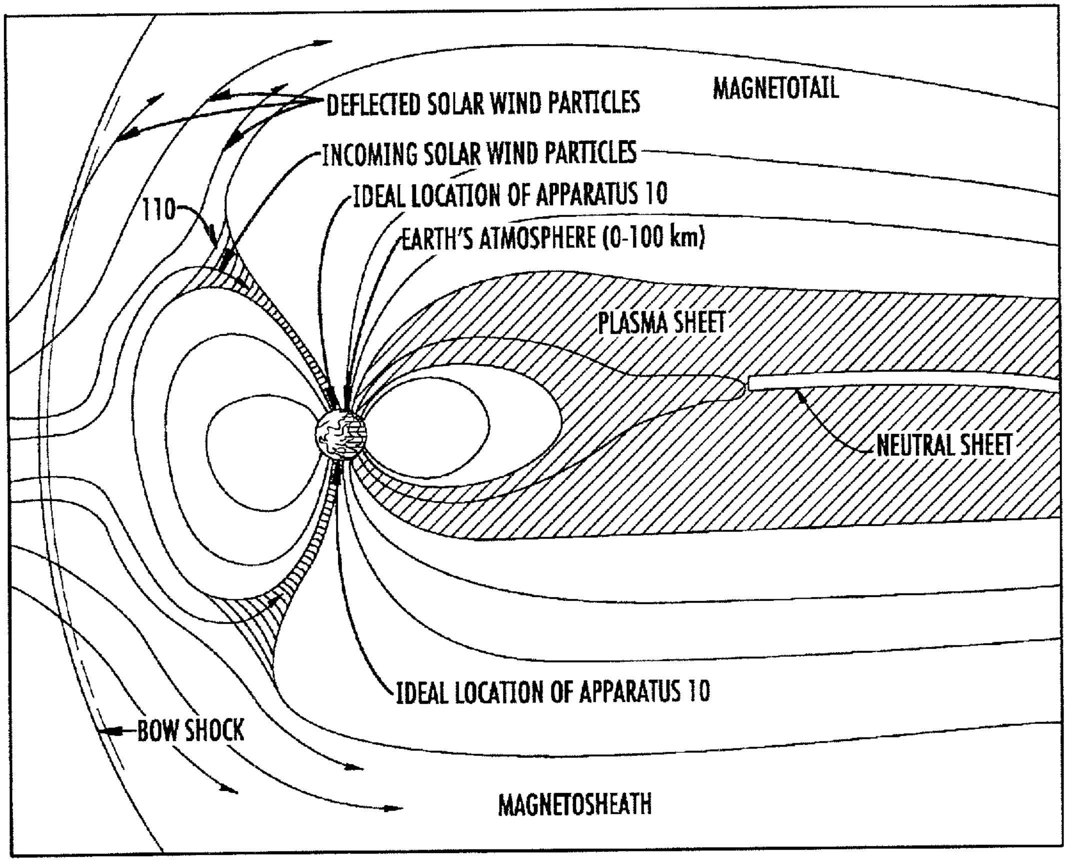
DEFLECTED SOLAR WIND PARTICLES
MAGNETOTAIL
INCOMING SOLAR WIND PARTICLES
110
IDEAL LOCATION OF APPARATUS 10
EARTH'S ATMOSPHERE (0-100 km)
PLASMA SHEET
NEUTRAL SHEET
IDEAL LOCATION OF APPARATUS 10
BOW SHOCK
MAGNETOSHEATH

FIG. 3

CHARGED PARTICLE INDUCTION FROM IONOSPHERE TO GROUND

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present application is a continuation of U.S. patent application Ser. No. 14/197,698, filed Mar. 5, 2014; which is a continuation of U.S. patent application Ser. No. 12/917,447, filed Nov. 1, 2010 (now U.S. Pat. No. 8,693,160) which claims the benefit of U.S. Provisional Application Ser. No. 61/280,131, filed Oct. 31, 2009; all of which are hereby incorporated by reference herein in their entirety, including any figures, tables, or drawings.

FIELD OF THE INVENTION

[0002] The present invention relates to the transmission of ionospheric energy through at least two "ionosphere-to-ground" coaxial laser-induced plasma channels formed through the Earth's atmosphere. As used herein, the term "ionospheric energy" means charged particles, namely ions and electrons, in the ionosphere. As used herein, the term "channel" means a conduit having walls comprised of ionized or partially ionized gas, the channel being substantially ring-shaped in cross-section, the walls serving as the conductive path through which charged particles are induced to ground.

BACKGROUND OF THE INVENTION

[0003] The need to develop alternative clean renewable energy supplies is becoming critical. Problems of climate change and global warming due to carbon emissions are becoming an accepted fact. The dramatic rise in the cost of petro-chemical energy is due to decreased supplies and increase in world demand. There is increasing political instability of regions of the world that have large oil supplies. There are geo-political and economic ramifications of being dependent on imported energy or on being economically dependent on the export of energy from dwindling energy reserves.

[0004] Heretofore, space solar power has been considered as an alternative clean energy source. Space solar power is clean, inexhaustible, available 24 hours a day, and has the potential to generate as much energy as terrestrial power plants. However, because of the filtering effect of the Earth's atmospheric gases (air acts as an insulator), and for other reasons such as reflection and absorption, only a fraction of solar energy in any given area reaches Earth. The average solar power per unit area outside Earth's atmosphere during any given time period is about 136% that available on Earth's surface during direct sunlight (1336 W/m2). Various apparatus and methods have been considered, and some developed, to compensate for this shortcoming.

[0005] A very recent proposed solution is described in pending U.S. Patent Application No. 2009/0171477 of Neyfeh et al. which describes systems and methods that employ high-intensity lasers to set up a thin plasma sheet, also called a waveguide or "hot shell", in the atmosphere as a function of beam intensity and geometry. A laser beam can be spread and directed with physical optics to generate a thin inverted cone-shaped hot shell waveguide in the atmosphere. According to Neyfeh, the hot shell of the waveguide has a different index of refraction from that of the surrounding air layers and as such serves to internally reflect portions of the entering solar ray entering an aperture in the hot shell, toward the tip of the cone and a solar energy storage component positioned there, thus providing a virtual solar energy concentration system. The Neyfeh et al. systems and methods, therefore, endeavor to make the best of the filtered solar energy present in Earth's atmosphere.

[0006] Another proposed solution is to collect the Sun's energy in space, where it is more concentrated, via satellite-based or lunar-based collection stations. The energy must then be transmitted from the collection point in space to the Earth's surface. Since wires extending from Earth's surface to an orbiting satellite would be impractical, many space-based solar power designs have proposed the use of microwave or laser beams for wireless power transmission. The collecting station would convert solar energy into electrical energy, which would then be used to power a microwave emitter or laser directed at a collector on the Earth's surface. There are numerous technical, political, legal and economic challenges to building space solar power stations. Limitations in photovoltaic technology and the difficulty of building large structures in space are some of the issues. In the past, launch costs have been prohibitive. In addition to the relatively high costs involved with this method, other problems include cumulative radiation damage or micrometeoroid impacts.

[0007] The underlying physics of wireless power transmission resembles that of wireless communications, but with an important difference. Unlike information transfer, where the percentage of received power must be only sufficiently high enough to recover the signal, wireless power transmission places a critical emphasis on the maximum amount of possible energy transfer and conversion efficiency. Ideally, a wireless power transmission system would have the ability to transmit any amount of power to any point in space, but practical limitations such as conversion efficiencies at the source and the receiver, and disturbances in the transmission medium will always limit the performance of an implemented system.

[0008] For instance, as recently as 2006, the highest amount of energy obtained from a receiver when testing high intensity lasers to illuminate vertical multi junction solar cells developed by NASA-GRC was 23.7778 watts. One vertical multi junction solar cell was able to achieve a power density of 13.6 watts per cm^2, at a conversion efficiency of 24%. Although these experiments confirm that the VMJ technology can withstand and utilize the high intensity laser energy without damage or significant reduction in previously known conversion efficiencies, wireless power transmission via high intensity laser beaming technology is not yet ready for high order energy transfer. In short, the main limiting factors to the laser power beaming system are the conversion efficiencies of the laser (electrical to photonic) and the photovoltaic cells (photonic to electrical). Although there are a variety of photovoltaic cells on the market approaching conversion efficiencies of 40% (such as triple junction cells), these technologies cannot operate at intensities 1000 times or greater than that of the sun.

[0009] Based on the above, it is clear that there is a strong desire to harness the energy existing in both our atmosphere and beyond, but it is apparent that the various methods and apparatus designed to capture and utilize this energy source are attendant with various, problems, shortcomings and limitations.

SUMMARY OF THE INVENTION

[0010] In order to meet the need in the art, the subject invention presents an entirely different approach to "tapping" this vast energy source. At least one preferred embodiment of the invention is discussed below.

[0011] The Earth's surface—ocean and solid—and the ionosphere are highly conductive. The atmosphere conducts electricity because of the presence of positive and negative ions plus free electrons. Conductivity is poor near sea level but increases rapidly with height up to the ionosphere; also it is greater at polar latitudes than equatorial. Between ground and the Ionosphere, the atmosphere acts an insulator.

[0012] The ionosphere begins at a height of about 50 km (30 miles) above the Earth's surface, but it is most distinct and important above 80 km (50 miles). In the photochemical ionosphere (which is mostly contained within the thermosphere but also partly extends into the neighboring mesosphere), cosmic radiation of high-energy sub-atomic particles and the absorption of much of the solar ultraviolet radiation separates negative electrons from oxygen and nitrogen molecules. The ions and free electrons move rapidly under the influence of electrical forces—the ionospheric wind—and the ionosphere is highly conductive in this region. During fair weather there is an electric potential difference of 250,000 to 500,000 volts between the ionosphere and the Earth's surface, the surface being negative relative to the ionosphere. This, incidentally, gives rise to the fair weather current, which is a steady flow of electrons from the surface at about one microwatt per square meter. The subject invention is directed generally toward the exploitation of this electric potential difference, and more specifically to a system, method and apparatus for inducing a flow of charged particles from the charged upper atmosphere, including but not limited to the ionosphere, to ground through concentric laser induced plasma channels ("con-channels").

[0013] It is widely recognized that a laser beam of suitable wavelength can penetrate through a gas medium over great distances and will establish a partially ionized trail therethrough. Examples of laser systems used to produce a thin ionizing beam through the atmosphere are taught in U.S. Pat. No. 3,719,829 issued to Vaill in 1973 entitled, Laser Beam Techniques and U.S. Pat. No. 5,726,855 issued to Maourou et al. in 1998 and entitled, Apparatus and Method for Enabling the Creation of Multiple Extended Conduction Paths in the Atmosphere. Mourou et al. teaches an apparatus and method for enabling the creation of multiple laser generated extended conduction paths in the atmosphere to safely control the discharge of lightning strikes to a grounding tower. It is also well known in the art that a plasma channel has a low electrical resistance and, once formed, will permit continuous current flow if the energy source that heats the plasma can be maintained. Unlike a normal electrical conductor, the resistance (and voltage drop) across an unconfined plasma channel decreases with increasing current flow, a property called negative resistance. Because plasmas are conductive and respond to electric and magnetic fields and can be efficient sources of radiation, they are usable in numerous applications where such control is needed or when special sources of energy or radiation are required.

[0014] Unfortunately, it has been observed that a highly-ionized path will tend to lose its straight-line configuration after creation and will begin to wander transversely in serpentine configurations, which wandering, once started, will grow progressively worse. Moreover, charges carried in plasma channels tend to self constrict due to magnetic forces stemming from the current flowing through the plasma. This phenomenon is known as the "plasma pinch". Upon constricting of the charge as a result of the pinch, the current will then follow the path of least resistance to ground which is by its nature unpredictable and dangerous.

[0015] Applicant's methods and apparatus overcome these shortcomings through the creation of at least two concentric conductive paths to and from which the wandering current may "jump", in alternating fashion, thereby maintaining a continuous flow of energy in the desired direction from the ionosphere to the terrestrial target collection apparatus. According to the present invention there is thus provided a ground-to-ionosphere laser-induced conduction path of concentric plasma channels through Earth's atmosphere along which charged particles within the ionosphere may be transmitted to Earth. The term "atmosphere" as used herein means the mass of air surrounding the Earth. The term "plasma" as used herein means a combination of neutrons, ions, electrons and fields that have conductive and collective effects.

[0016] The subject apparatus more specifically comprises high power light emitting means, such as a laser array, in operable communication with a high energy output means to accomplish initiation of at least two concentric plasma channels from the Earth's surface to the ionosphere for the transmission of ionospheric energy therethrough. The path need only remain open for relatively short intervals to transmit high orders of energy. The transmitted energy is drawn down through the conductive plasma channels to at least one intervening electrode mounted to a support structure. Charges flow down through the channels to the electrode as they approach ground and the conductive electrode carries the energy away from the laser assembly to avoid damage to or destruction thereof. The at least one electrode will then pass these charges onto a capacitor bank or other storage means to store them. The at least one electrode is mounted to a support structure housing heat dissipation means well known in the nuclear fission art. The light emitting means is preferably insulated from and/or located remotely from the at least one electrode in order to protect the former from damage which could result if the high order energy being transmitted through the plasma channels was not intercepted by the electrode(s). Ideal locations for the subject apparatus are also described.

[0017] There has thus been outlined, rather broadly, the more important components and features of the invention in order that the detailed description thereof that follows may be better understood, and in order that the present contribution to the art may be better appreciated. There are, of course, additional features of the invention that will be described hereinafter and which will form the subject matter of the claims appended hereto. In this respect, before explaining at least one embodiment of the invention in detail, it is to be understood that the invention is not limited in its application to the details of construction and to the arrangements of the components set forth in the following description or illustrated in the drawings. The invention is capable of other embodiments and of being practiced and carried out in various ways. Also, it is to be understood that the phraseology and terminology employed herein are for the purpose of description and should not be regarded as limiting. As such, those skilled in the art will appreciate that the conception, upon which this disclosure is based, may readily be utilized as a basis for the designing of other structures, methods and systems for carrying out the several purposes of the present invention. It is important, therefore, that the claims be regarded as including such equivalent constructions insofar as they do not depart from the spirit and scope of the present invention.

[0018] Further, the purpose of the foregoing abstract is to enable the U.S. Patent and Trademark Office and the public generally, and especially the scientists, engineers and practitioners in the art who are not familiar with patent or legal terms or phraseology, to determine quickly from a cursory inspection the nature and essence of the technical disclosure of the application. The abstract is neither intended to define the invention of the application, which is measured by the claims, nor is it intended to be limiting as to the scope of the invention in any way.

[0019] For a better understanding of the invention, its advantages and the specific objects attained by its uses, reference should be had to the accompanying drawings and descriptive matter in which there is illustrated a preferred embodiment of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0020] The invention will be better understood and objects other than those set forth above will become apparent when consideration is given to the following detailed description thereof. Such description makes reference to the annexed drawings wherein:

[0021] The invention will be better understood and objects other than those set forth above will become apparent when consideration is given to the following detailed description thereof. Such description makes reference to the annexed drawings wherein:

[0022] FIG. **1** is a highly simplified schematical view of energy being transmitted from the ionosphere to ground via concentric laser-induced plasma channels produced by the subject apparatus in atmosphere (not-to-scale);

[0023] FIG. **2** is simple schematic illustration of the subject apparatus for inducing the flow of charged particles from the ionosphere to ground through concentric laser induced plasma channels (not-to-scale); and

[0024] FIG. **3** is a simplified schematical view of the earth with a magnetic field (force) line along which the present invention may be carried out (not-to-scale).

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

[0025] Reference is now made to FIG. **1** in which there is illustrated a highly simplified schematic view of energy E being transmitted from the ionosphere **100** through atmosphere **102** via concentric plasma channels **36**,**38** generated by the subject apparatus which is designated generally by reference numeral **10**. Apparatus **10** is situated on the earth's surface **104** at preferred locations herein described.

[0026] Referring to FIG. **2**, apparatus **10** is comprised of three primary components: 1) a wireless conductive channel producing means **12** or "atmospheric auger" connected to an energy source **14**, 2) charged particle receiving means **16** and 3) energy storage means **18**. Each primary component is described in detail below, together with secondary components of the subject apparatus.

[0027] Wireless conductive channel producing means **12** (hereinafter "plasma channel producing means") is comprised of first ionizing beam emitting means **20** used to create a first ionizing beam **22**, and second ionizing beam emitting means **24** used to create a second ionizing beam **26**. The term "ionizing beam emitting means" as used herein includes not only presently known lasers, but also other light sources which will excite ionization in a medium. Lasers utilize the natural oscillations of atoms or molecules between energy levels for generating beams **22**,**26** of highly amplified and coherent electromagnetic radiation of one or more discrete frequencies, which in turn generate first and second conductive paths **28**,**30**, respectively. The laser means used to establish conductive paths **28**,**30** should be selected with regard to energy, pulsewidth and wavelength so that it provides control of ionization concentration in each plasma channel. In a gas medium it is important that the ionizing beams **22**,**26** are controlled for optimum transmission of a charge.

[0028] Technologies that facilitate the creation of the subject con-channels include recent advancements in ultra-high intensity pulsed lasers and their transmission optics which together can generate a plasma field and have precise control over the field's geometry. Information regarding the manipulation and transmission of high powered lasers is generally classified due to the nature of their application. It is well known however, that successful high power laser systems are currently operation, and it follows that the condition and control optics are in place to handle such a system. According to an embodiment described herein, this technology is adapted to generate the con-channels. Ultra-high power pulsed or continuous lasers are now available in the Megawatt class, such as the Airborne Laser (ABL) built by Northrop Grumman. The lasers employed by the subject apparatus and methods should also by capable of varying the chirp of the input pulse to achieve control of the filamentation over horizontal kilometric distances. By way of example, one embodiment may include a Ti:sapphire CPA (Chirp Pulse Amplification) laser system operating at 10 Hz and delivering a maximum output power of 0.2 TW per pulse. The pulse duration approximates 130-135 fs (FWHM). The beam may be truncated by a circular aperture having a desired radius in order to have a more stable beam shape with a perfect or near-perfect circular symmetry.

[0029] Those skilled in the art will readily recognize that there are two main mechanisms responsible for the photoionization of air: avalanche ionization and multiphoton ionization. In avalanche ionization, or sometimes known as cascade ionization, the bound electrons in the valence shell of air molecules usually have an energy bandgap greater than the incident laser photon energy. However, a small number of free electrons exist. These free electrons may be generated through high altitude ionization of air molecules, radicals, break-up of aerosol molecules, thermal ionization or multiphoton ionization. These free electrons constitute the required seed electrons for avalanche ionization. Although these free electrons may exist in small number initially and have low kinetic energy, through a process called inverse bremsstrahlung, they can be accelerated to gain enough kinetic energy to impact ionize a molecule in collision, resulting in more free but slower electrons. The whole process repeats itself with more electrons gaining higher kinetic energy, exceeding the ionization potential of the bound electrons, to impact even more molecules, leading to an avalanche with the resultant formation of a plasma. In this inverse bremsstrahlung process, the free electrons absorb photons through the laser field when they collide with atoms or molecules. A free electron does not absorb laser energy directly without any collision. This is because the free electron simply quivers in the oscillating laser field, and when averaged over an optical cycle, it does not gain any net energy. The other dominant photoionization mechanism is the multi photon ionization process. Through this process, bound electrons may absorb several laser photons simultaneously to overcome their ionization potential. Multiphoton ionization depends largely on the laser operating frequency. Either method may be employed with the subject application.

[0030] In order to transform first and second conductive paths **28**,**30** into the desired cylindrical-shaped first and second plasma channels **36**,**38** it is necessary to cause their movement around a common central axis **40**. First and second beam moving means **32**,**34** are provided for this purpose and are comprised of either or a combination of mechanical or optical means. The manner in which at least two concentric plasma channels are generated using first and second ionizing beams **22**,**26** is not critical provided the properties (i.e., length, ionization intensity, uniformity, etc.) of each channel is controllable. Applicant's co-pending U.S. patent application entitled, Wireless Transmission Of Energy Through Concentric Laser-Induced Plasma Channels In Atmosphere, filed contemporaneously with the instant application and incorporated herein in its entirety, teaches at least one method of producing at least two concentric plasma channels. As previously stated, the term "channel" means a conduit, substantially ring-shaped in cross-section, having walls comprised of partially ionized gas ("plasma"), each wall being created by the movement of first and second conductive paths **28**,**30**, respectively. First plasma channel **36** and second plasma channel **38** share a common central longitudinal axis; they are concentric channels (or co-axial from a three-dimensional linear perspective). In alternate embodiments, the outermost channel and any channels between it and the innermost channel may be shaped like a funnel or frustum, narrower at point of origin and progressing outward as it extends into the atmosphere. U.S. Pat. No. 6,377,436 issued to Margolin on Apr. 23, 2002 and entitled, Microwave Transmission Using a Laser-Generated Plasma Beam Waveguide, which said reference is incorporated herein in its entirety to the extent it is not inconsistent with the explicit teachings of this specification, teaches various directed energy beam systems in the form of channels or "shells". Although the Margolin shells are used as a guide through which microwaves are transmitted rather than high order energy, the means by which the shells are created are instructive and find utility in the instant case. It should be appreciated that multiple concentric plasma channels may be created with at least two being the minimum required and at least three channels being preferred.

[0031] With continued reference to FIG. **2**, power to each ionizing beam emitting means is supplied by energy source **14**. The voltage required is related to and increases with the length of the path to be ionized. The energy in Joules required to excite the gas molecules in the path into ionization can be calculated in advance knowing the density of the molecules and the approximate diameter and length of the conductive path over which the energy from the ionosphere is to be transmitted. By way of example, energy source **14** may be comprised of a Marx generator which is capable of generating a high-voltage pulse. Voltages produced may exceed 2 million volts.

[0032] Charged particle collection means **16** comprises a charged particle attraction structure. Such structure includes at least one elongated electrode **42** generally somewhat similar to the lightning rods which are normally used to direct the electrical energy to ground when protecting buildings and the like. Electrode **42** is mounted to a support structure **44** generally, and to a conductive hemispherical dome **46** thereof specifically. A charge indicated by symbol + is applied to the at least one electrode **42** and/or dome **46** via charging means **48** which can impart a charge via conduction or induction (friction being impractical). Note that the + symbol is used as a substitute for the more conventional symbol "e" which of course can be either positive or negative. The hemispherical shape allows for the widest variance of charge wandering during energy transfer to the Earth's surface. This orb-like structure may include openings through its ionosphere-facing surface to allow the transmission of at least one ionizing beam therethrough with the actual light emitting means located remotely enough from the orb surface to mitigate against damage from expected energy loads.

[0033] The induction of ionospheric energy utilizing the subject apparatus **10** may be likened to the induction of a lighting strike, only of greater magnitude. A lightning strike discharges about 1-10 billion joules of energy and produces a current of some 30,000-50,000 amps, which heats the surrounding air to over 20,000 degrees Celsius. A single lightning bolt unleashes as much energy as blowing up a ton of TNT. Because of the enormous amount of heat attendant with the collection of ionospheric energy, even over a relatively short period of time, heat dissipation means are associated with the subject apparatus **10**. In one embodiment, heat dissipation means may be housed within support structure **44** as indicated by the "down arrow delta" symbol. Alternatively, heat dissipation means may be connected to support structure **44** but remote therefrom. Guidance for the design of a safe and efficient cooling system may be found in the nuclear science field of art for example, or in other power generation facilities. The main types (mechanisms) of heat transfer to be employed are 1) Wet cooling towers or simply cooling towers which operate on the principle of evaporation wherein the working fluid and the evaporated fluid (usually H_2O) are one and the same; 2) Dry coolers which operate by heat transfer through a surface that separates the working fluid from ambient air, such as in a heat exchanger, utilizing convective heat transfer (they do not use evaporation); and 3) Fluid coolers which are hybrids that pass the working fluid through a tube bundle, upon which clean water is sprayed and a fan-induced draft applied; the resulting heat transfer performance is much closer to that of a wet cooling tower, with the advantage provided by a dry cooler of protecting the working fluid from environmental exposure. In all applications, water may be tapped from a nearby water source **50** or other fluids used. Steam turbines (not shown) may be utilized to extract thermal energy from pressurized steam and convert it into rotary motion for the generation of electricity which may be utilized locally to power system components and/or fed to a local power grid **106**. Note that the release of steam in the ionospheric energy induction process will interfere with the photo-ionization of surrounding air and is therefore undesirable.

[0034] Once ionospheric particles are attracted to charged particle collection means **16** in the above described manner, the resulting electrical current is directed through the heat dissipation means via a first wire conductor **52** downwardly through support structure **44** to a surge arrestor assembly designated generally by reference numeral **54**. Such surge arrestor assembly is connected directly to ground **56** and has a switching assembly (not shown for purposes of clarity). The switch assembly is a structural adaptation of the surge arrestor assembly **54** serves automatically or manually to direct current flow either to energy storage means **18** via second wire conductor **58** when the conductive path is "open" or to ground **56** when the conduction path is closed. Energy storage means **18** may be comprised of, for example, a capacitor bank as illustrated or some other suitable means such as a collection coil of superconductive material as has been proposed in the art.

[0035] Various systems, methods and apparatus may be employed to convert the energy stored in energy storage means **18** to use. By way of example only, when energy stored within energy storage means **18** is required, a suitable switching apparatus (not shown for purposes of clarity) may be activated and the stored energy passed through a converter apparatus **60** structurally adapted to convert current coming from energy storage means **18** from D.C. to A.C. The A.C. current may then be passed through a transformer **62** for the purpose of regulating voltage before the current is directed into the power grid **106** for practical distribution and use. Energy may be transmitted from transformer **62** to power grid **106** via conventional power take-off means **64** and/or wirelessly **66** in a manner described in applicant's aforementioned co-pending application.

[0036] In operation, ionizing beam emitting means may be applied in continuous or, more likely, pulsed fashion, as determined by controller **68** used to activate energy source **14**. Alternatively, at least one ionizing beam emitting means may be operated in continuous fashion while at least one other ionizing beam emitting means may be operated in pulsed fashion. The intensity of both may additionally be modulated. In the event that the emitting means provides outputs which are continuous the control circuit will comprise merely on-off means, but where pulsed lasers and/or a pulsed energy source are used, the control circuit will comprise a trigger, suitable for firing each ionizing beam emitting, and also for controlling the rate at which each repeat their outputs. Pulsed systems have the advantage of providing quiescent intervals between their output pulses during which the ionized plasma in the portion of the beam which is located between the electrodes can die out to permit the system to start all over again and re-establish a straight-line highly conductive path of narrow and well defined proportions. The periodic extinguishing of each channel in a pulsed system is beneficial since it allows unstable ionization to die out so that it can be reestablished in a new channel which is both straight, and highly defined and concentrated. The subject apparatus provides a means of assuring that at least one plasma channel **36**,**38**, etc. is always open at any point in time thereby assuring that the energy to be transmitted always has a pre-defined conductive path to travel upon to the target collection means **16**. For instance, while plasma channel **32** is extinguished, controller **68** assures that plasma channel **36** is active. As previously stated, the periodic discharges from the at least two ionizing beam emitting means can be synchronized to produce, in the aggregate, a constant directional plasma channel between the ionosphere and ground. The overall transition of the Earth's atmosphere within the concentric laser induced plasma channels from insulator to conductor is in correlation to the Bose-Hubbard (boson Hubbard) model. In contrast to relatively low order energy transfer at colder temperatures demonstrated by fermion conductors, the subject invention provides high order energy transfer at ambient temperatures using Hubbard's principles as applied to bosons.

[0037] FIG. **2** illustrates the manner in which energy is transmitted from its point of origin in the ionosphere **104** to collection means **16**. First and second plasma channels **36**,**38** comprise a plasma which is highly conductive and which is propagating along directional arrow **108**. Prior to transmission of ionospheric charged particles, designated by circumscribed minus signs, through plasma channels **36**,**38**, the channels remain fairly well defined for a usable interval of time. However, when the energy to be transmitted is introduced into and travels along a given plasma channel, the charge will eventually begin to self constrict due to magnetic forces stemming from the current flowing through the plasma causing the charge to close upon itself and leaving the energy being transmitted with no clear path to follow, potentially striking unintended targets.

[0038] The employment of at least two, and preferably at least three, concentric plasma channels in accordance with the subject invention mitigates such occurrences by providing adjacent conductive paths which ionospheric energy can "jump to" when losing its point-to-point direction. After the charge transitions to its new conductive path, the last-travelled path is regenerated by its corresponding ionization beam emitting means. The subject apparatus and methods, therefore, provide that at least one conductive path is always open and in close proximity to a neighboring current-carrying plasma channel.

[0039] In the example illustrated in FIG. **2**, a potential exists between free charged particles in the ionosphere, on the one hand, and charged electrode **42** and dome **46** on the other. First and second plasma channels **36**,**38** provide a bridge between them for the transmission of the former to the latter. Charged particles from the ionosphere are drawn into the walls of whichever plasma channel is open at the time (in the instant illustration, plasma channel **36**, and begin to travel downwardly there along a first distance or time interval d_1 until the current is pinched and can no longer travel along its current channel. At that time the errant current will seek out a new path of least resistance on the way to collection means **16**. The closest path of least resistance will be the then open neighboring channel **38** onto which the current will "jump" and then travel a second distance or time interval d_2. As energy traverses its new channel the previously traversed plasma channel **36** will be reformed serving as the next adjacent "jump site" for current to find when it again experiences pinch. The charge will jump from channel to channel, traveling a distance or time interval along each, in alternating or "stair-step" down fashion until reaching energy collection means **16**. Once energy collection means **16** reaches capacity, power to the light emitting means may be relinquished and the magnetic spin of the transferred charged particles will close the field. It is possible, of course, that in certain circumstances a particular charge will remain open while traversing a single channel to the target. What is important is that there is always a predefined alternate directional conductive path in close proximity for the energy to travel upon.

[0040] As previously mentioned, the ionosphere is most distinct above 80 km (50 miles) where the ions and free electrons move rapidly under the influence of electrical forces-the ionospheric wind-and where the ionosphere is highly conductive. During fair weather there is an electric potential difference of 250,000 to 500,000 volts between the ionosphere and the Earth's surface. Accordingly, construction of the subject apparatus in locations which facilitate tapping this region of the atmosphere are desirable, particularly at high elevations where the air to be ionized contains a lower percentage of contaminants which adversely affect propagation of conductive plasma channels.

[0041] Additionally, reference now being made to FIG. **3**, in the upper regions of the ionosphere, beginning several hundred kilometers above Earth's surface and extending tens of thousands of kilometers into space, is the magnetosphere, a region where the behavior of charged particles is strongly affected by the magnetic fields of Earth and the Sun. The solar wind is a stream of charged particles ejected from the upper atmosphere of the sun. It consists mostly of electrons and protons with energies of about 1 keV. As the solar wind approaches Earth's well-developed magnetic field, the particles are deflected by the Lorentz force. This region, known as the magnetosphere, causes the particles to travel around the planet rather than bombarding the atmosphere or surface. However some of the charged particles leak through this magnetic barrier and become trapped in the Van Allen radiation belt. A smaller number of particles from the solar wind manage to travel, as though on an electromagnetic energy transmission line, to the Earth's upper atmosphere and ionosphere in the auroral zones. More specifically, the charged particles travel through funnel-like openings (cusps **110**) at the North and South Poles, releasing tremendous energy when they hit the upper atmosphere. The Northern and Southern Lights (auroras) are the visible evidence of this energy transfer from the Sun to the Earth. These regions, approximately 13 latitude 14 n longitude, and where magnetic field lines intersect the Earth, are believed to be ideal for employment of the subject apparatus **10**. It may be advantageous to locate the subject system and apparatus at a dwindling or soon to be decommissioned mine, because of pre-existing environmental permits and useful tunnels.

[0042] Although the present invention has been described with reference to the particular embodiments herein set forth, it is understood that the present disclosure has been made only by way of example and that numerous changes in details of construction may be resorted to without departing from the spirit and scope of the invention. Thus, the scope of the invention should not be limited by the foregoing specifications, but rather only by the scope of the claims appended hereto.

1. An apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground, comprising:

a. means for generating a first conductive channel through the atmosphere and in communication with the charge-rich outer atmosphere, said first conductive channel being substantially ring-shaped in cross-section and comprised of a wall of partially ionized atmosphere; said first conductive channel having a diameter;

b. means for generating a second conductive channel through the atmosphere and in communication with the charge-rich outer atmosphere, said second conductive channel being substantially ring-shaped in cross-section and comprised of a wall of partially ionized atmosphere; said second conductive channel having a diameter greater than said diameter of said first conductive channel; said first conductive channel and said second conductive channel being concentric; and

c. charged particle collection means associated with said means for generating said first and said second conductive channels.

2. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **1**, wherein said charged particle collection means is comprised of a at least one electrode attached to a support structure.

3. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **2**, wherein said support structure comprises a conductive hemispherical dome.

4. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **2**, wherein one or more of said at least one electrode and said support structure are charged.

5. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **3**, wherein one or more of said at least one electrode and said hemispherical dome are charged.

6. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **1**, wherein said second conductive channel is funnel shaped being more narrow at its point of origin.

7. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **2**, wherein said second conductive channel is funnel shaped being more narrow at its point of origin.

8. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **3**, wherein said second conductive channel is funnel shaped being more narrow at its point of origin.

9. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **4**, wherein said second conductive channel is funnel shaped being more narrow at its point of origin.

10. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **5**, wherein said second conductive channel is funnel shaped being more narrow at its point of origin.

11. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **1**, further including means for momentarily or periodically extinguishing said first conductive channel, said second conductive channel, or asynchronously both.

12. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **2**, further including means for momentarily or periodically extinguishing said first conductive channel, said second conductive channel, or asynchronously both.

13. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **3**, further including means for momentarily or periodically extinguishing said first conductive channel, said second conductive channel, or asynchronously both.

14. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **4**, further including means for momentarily or periodically extinguishing said first conductive channel, said second conductive channel, or asynchronously both.

15. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **5**, further including means for momentarily or periodically extinguishing said first conductive channel, said second conductive channel, or asynchronously both.

16. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **6**, further including means for momentarily or periodically extinguishing said first conductive channel, said second conductive channel, or asynchronously both.

17. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **7**, further including means for momentarily or periodically extinguishing said first conductive channel, said second conductive channel, or asynchronously both.

18. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **8**, further including means for momentarily or periodically extinguishing said first conductive channel, said second conductive channel, or asynchronously both.

19. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **9**, further including means for momentarily or periodically extinguishing said first conductive channel, said second conductive channel, or asynchronously both.

20. The apparatus for inducing the flow of charged particles from the charge-rich outer atmosphere to ground of claim **10**, further including means for momentarily or periodically extinguishing said first conductive channel, said second conductive channel, or asynchronously both.

* * * * *

United States Patent [19]
Eastlund

[11] **Patent Number: 4,686,605**
[45] **Date of Patent: Aug. 11, 1987**

[54] METHOD AND APPARATUS FOR ALTERING A REGION IN THE EARTH'S ATMOSPHERE, IONOSPHERE, AND/OR MAGNETOSPHERE

[75] Inventor: **Bernard J. Eastlund**, Spring, Tex.
[73] Assignee: **APTI, Inc.**, Los Angeles, Calif.
[21] Appl. No.: **690,333**
[22] Filed: **Jan. 10, 1985**
[51] **Int. Cl.**^{4} **H05B 6/64**; H05C 3/00; H05H 1/46
[52] **U.S. Cl.** .... **361/231**; 89/1.11; 380/59; 244/158 R
[58] **Field of Search** .... 361/230, 231; 244/158 R; 376/100; 89/1.11; 380/59

[56] **References Cited**

PUBLICATIONS

Liberty Magazine, (2/35) p. 7 N. Tesla.
New York Times (9/22/40) Section 2, p. 7 W. L. Laurence.
New York Times (12/8/15) p. 8 Col. 3.

Primary Examiner—Salvatore Cangialosi
Attorney, Agent, or Firm—Roderick W. MacDonald

[57] ABSTRACT

A method and apparatus for altering at least one selected region which normally exists above the earth's surface. The region is excited by electron cyclotron resonance heating to thereby increase its charged particle density. In one embodiment, circularly polarized electromagnetic radiation is transmitted upward in a direction substantially parallel to and along a field line which extends through the region of plasma to be altered. The radiation is transmitted at a frequency which excites electron cyclotron resonance to heat and accelerate the charged particles. This increase in energy can cause ionization of neutral particles which are then absorbed as part of the region thereby increasing the charged particle density of the region.

15 Claims, 5 Drawing Figures

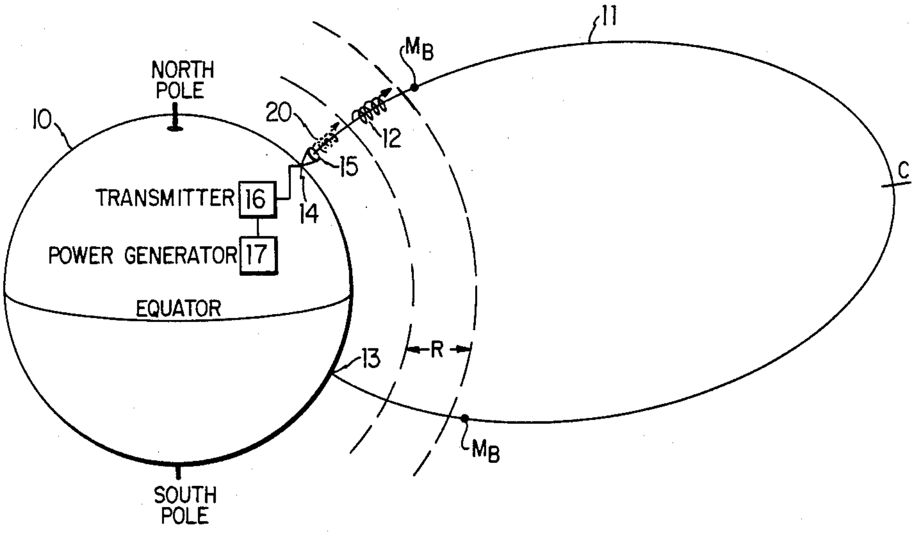

FIG. 1

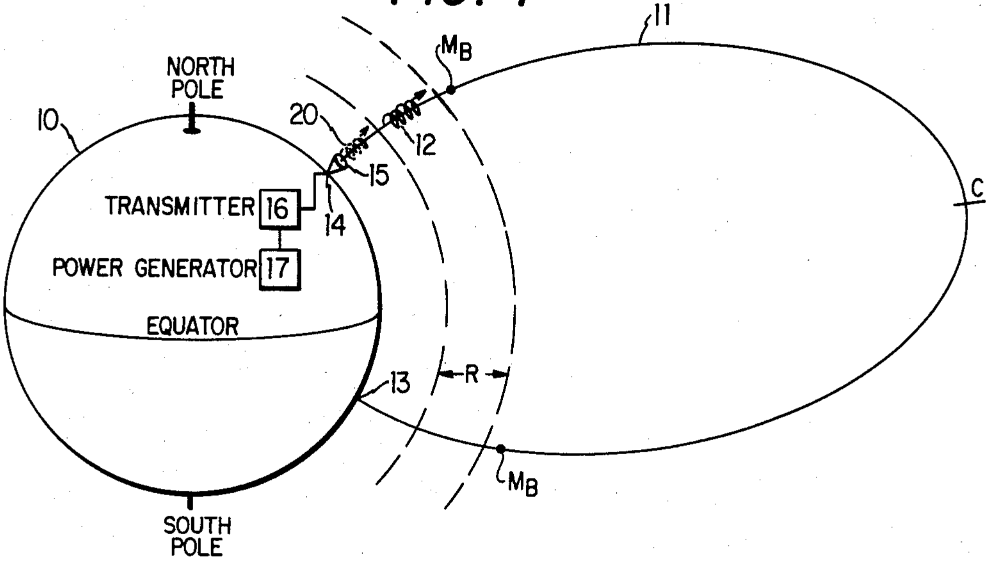
11
MB
NORTH POLE
10
20
12
15
C
TRANSMITTER
16
14
POWER GENERATOR
17
EQUATOR
13
R
MB
SOUTH POLE

FIG. 2

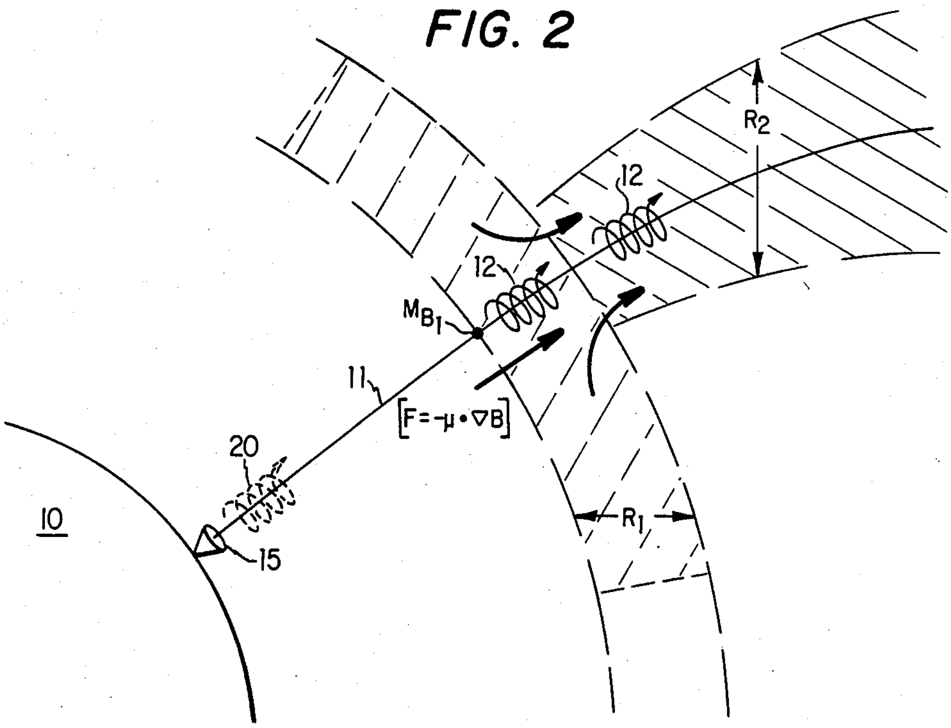
R2
MB1
[F=-μ•∇B]
R1

FIG. 3

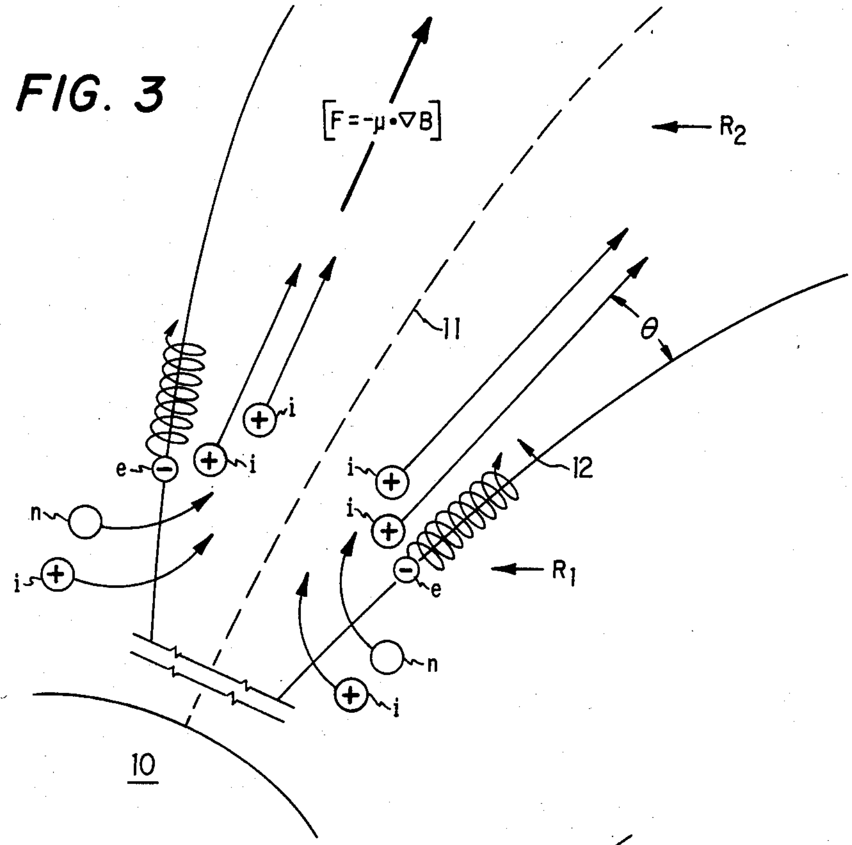
[F = -μ•∇B]
R$_2$
11
θ
12
R$_1$
e
i
i
i
i
i
i
n
n
e
10

FIG. 4

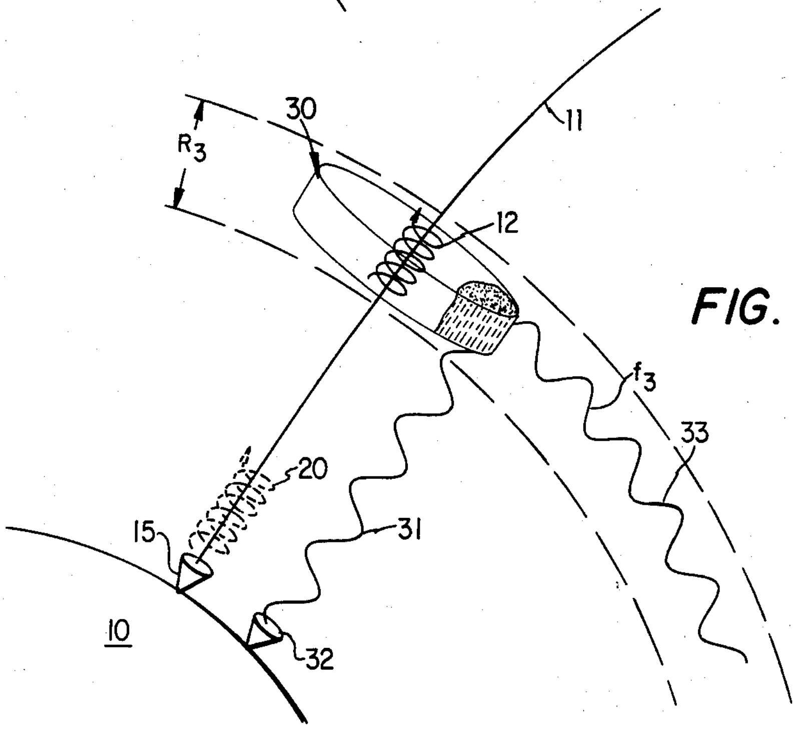
30
R$_3$
11
12
f$_3$
33
20
15
31
32
10

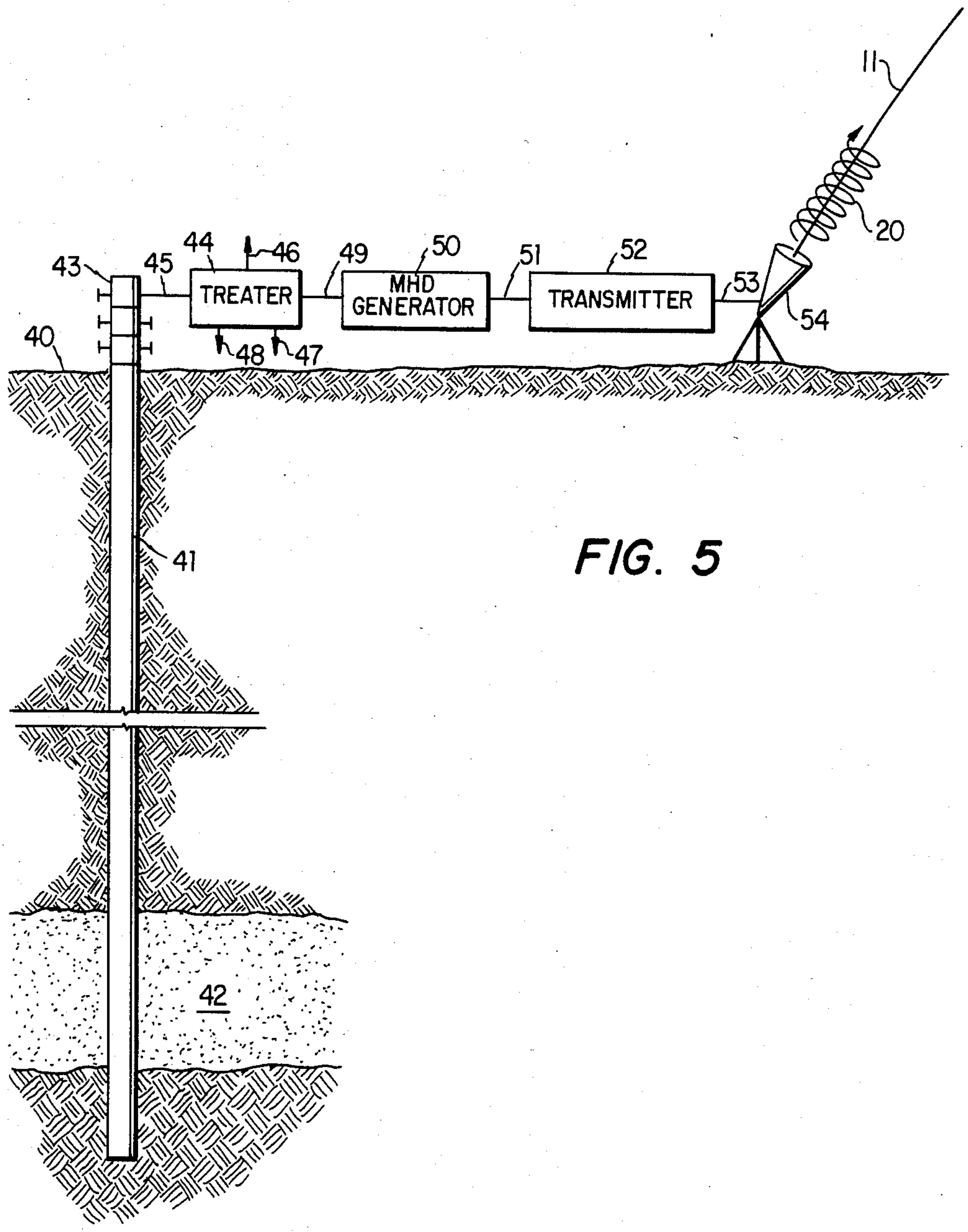
TREATER
MHD GENERATOR
TRANSMITTER

FIG. 5

METHOD AND APPARATUS FOR ALTERING A REGION IN THE EARTH'S ATMOSPHERE, IONOSPHERE, AND/OR MAGNETOSPHERE

DESCRIPTION

1. Technical Field

This invention relates to a method and apparatus for altering at least one selected region normally existing above the earth's surface and more particularly relates to a method and apparatus for altering said at least one region by initially transmitting electromagnetic radiation from the earth's surface essentially parallel to and along naturally-occurring, divergent magnetic field lines which extend from the earth's surface through the region or regions to be altered.

2. Background Art

In the late 1950's, it was discovered that naturally-occuring belts exist at high altitudes above the earth's surface, and it is now established that these belts result from charged electrons and ions becoming trapped along the magnetic lines of force (field lines) of the earth's essentially dipole magnetic field. The trapped electrons and ions are confined along the field lines between two magnetic mirrors which exist at spaced apart points along those field lines. The trapped electrons and ions move in helical paths around their particular field lines and "bounce" back and forth between the magnetic mirrors. These trapped electrons and ions can oscillate along the field lines for long periods of time.

In the past several years, substantial effort has been made to understand and explain the phenomena involved in belts of trapped electrons and ions, and to explore possible ways to control and use these phenomena for beneficial purposes. For example, in the late 1950's and early 1960's both the United States and U.S.S.R. detonated a series of nuclear devices of various yields to generate large numbers of charged particles at various altitudes, e.g., 200 kilometers (km) or greater. This was done in order to establish and study artifical belts of trapped electrons and ions. These experiments established that at least some of the extraneous electrons and ions from the detonated devices did become trapped along field lines in the earth's magnetosphere to form artificial belts which were stable for prolonged periods of time. For a discussion of these experiments see "The Radiation Belt and Magnetosphere", W. N. Hess, Blaisdell Publishing Co., 1968, pps. 155 et sec.

Other proposals which have been advanced for altering existing belts of trapped electrons and ions and/or establishing similar artificial belts include injecting charged particles from a satellite carrying a payload of radioactive beta-decay material or alpha emitters; and injecting charged particles from a satellite-borne electron accelerator. Still another approach is described in U.S. Pat. No. 4,042,196 wherein a low energy ionized gas, e.g., hydrogen, is released from a synchronous orbiting satellite near the apex of a radiation belt which is naturally-occurring in the earth's magnetosphere to produce a substantial increase in energetic particle precipitation and, under certain conditions, produce a limit in the number of particles that can be stably trapped. This precipitation effect arises from an enhancement of the whistler-mode and ion-cyclotron mode interactions that result from the ionized gas or "cold plasma" injection.

It has also been proposed to release large clouds of barium in the magnetosphere so that photoionization will increase the cold plasma density, thereby producing electron precipitation through enhanced whistler-mode interactions.

However, in all of the above-mentioned approaches, the mechanisms involved in triggering the change in the trapped particle phenomena must be actually positioned within the affected zone, e.g., the magnetosphere, before they can be actuated to effect the desired change.

The earth's ionosphere is not considered to be a "trapped" belt since there are few trapped particles therein. The term "trapped" herein refers to situations where the force of gravity on the trapped particles is balanced by magnetic forces rather than hydrostatic or collisional forces. The charged electrons and ions in the ionosphere also follow helical paths around magnetic field lines within the ionosphere but are not trapped between mirrors, as in the case of the trapped belts in the magnetosphere, since the gravitational force on the particles is balanced by collisional or hydrostatic forces.

In recent years, a number of experiments have actually been carried out to modify the ionosphere in some controlled manner to investigate the possibility of a beneficial result. For detailed discussions of these operations see the following papers: (1) Ionospheric Modification Theory; G. Meltz and F. W. Perkins; (2) The Platteville High Power Facility; Carrol et al.; (3) Arecibo Heating Experiments; W. E. Gordon and H. C. Carlson, Jr.; and (4) Ionospheric Heating by Powerful Radio Waves; Meltz et al., all published in Radio Science, Vol. 9, No. 11, November, 1974, at pages 885–888; 889–894; 1041–1047; and 1049–1063, respectively, all of which are incorporated herein by reference. In such experiments, certain regions of the ionosphere are heated to change the electron density and temperature within these regions. This is accomplished by transmitting from earth-based antennae high frequency electromagnetic radiation at a substantial angle to, not parallel to, the ionosphere's magnetic field to heat the ionospheric particles primarily by ohmic heating. The electron temperature of the ionosphere has been raised by hundreds of degrees in these experiments, and electrons with several electron volts of energy have been produced in numbers sufficient to enhance airglow. Electron concentrations have been reduced by a few percent, due to expansion of the plasma as a result of increased temperature.

In the Elmo Bumpy Torus (EBT), a controlled fusion device at the Oak Ridge National Laboratory, all heating is provided by microwaves at the electron cyclotron resonance interaction. A ring of hot electrons is formed at the earth's surface in the magnetic mirror by a combination of electron cyclotron resonance and stochastic heating. In the EBT, the ring electrons are produced with an average "temperature" of 250 kilo electron volts or kev (2.5×10^9K) and a plasma beta between 0.1 and 0.4; see, "A Theoretical Study of Electron—Cyclotron Absorption in Elmo Bumpy Torus", Batchelor and Goldfinger, Nuclear Fusion, Vol. 20, No. 4 (1980) pps. 403–418.

Electron cyclotron resonance heating has been used in experiments on the earth's surface to produce and accelerate plasmas in a diverging magnetic field. Kosmahl et al. showed that power was transferred from the electromagnetic waves and that a fully ionized plasma

was accelerated with a divergence angle of roughly 13 degrees. Optimum neutral gas density was 1.7×10^{14} per cubic centimeter; see, "Plasma Acceleration with Microwaves Near Cyclotron Resonance", Kosmahl et al., Journal of Applied Physics, Vol. 38, No. 12, Nov., 1967, pps. 4576–4582.

DISCLOSURE OF THE INVENTION

The present invention provides a method and apparatus for altering at least one selected region which normally exists above the earth's surface. The region is excited by electron cyclotron resonance heating of electrons which are already present and/or artifically created in the region to thereby increase the charged particle energy and ultimately the density of the region.

In one embodiment this is done by transmitting circularly polarized electromagnetic radiation from the earth's surface at or near the location where a naturally-occurring dipole magnetic field (force) line intersects the earth's surface. Right hand circular polarization is used in the northern hemisphere and left hand circular polarization is used in the southern hemisphere. The radiation is deliberately transmitted at the outset in a direction substantially parallel to and along a field line which extends upwardly through the region to be altered. The radiation is transmitted at a frequency which is based on the gyrofrequency of the charged particles and which, when applied to the at least one region, excites electron cyclotron resonance within the region or regions to heat and accelerate the charged particles in their respective helical paths around and along the field line. Sufficient energy is employed to cause ionization of neutral particles (molecules of oxygen, nitrogen and the like, particulates, etc.) which then become a part of the region thereby increasing the charged particle density of the region. This effect can further be enhanced by providing artificial particles, e.g., electrons, ions, etc., directly into the region to be affected from a rocket, satellite, or the like to supplement the particles in the naturally-occurring plasma. These artificial particles are also ionized by the transmitted electromagnetic radiation thereby increasing charged particle density of the resulting plasma in the region.

In another embodiment of the invention, electron cyclotron resonance heating is carried out in the selected region or regions at sufficient power levels to allow a plasma present in the region to generate a mirror force which forces the charged electrons of the altered plasma upward along the force line to an altitude which is higher than the original altitude. In this case the relevant mirror points are at the base of the altered region or regions. The charged electrons drag ions with them as well as other particles that may be present. Sufficient power, e.g., 10^{15} joules, can be applied so that the altered plasma can be trapped on the field line between mirror points and will oscillate in space for prolonged periods of time. By this embodiment, a plume of altered plasma can be established at selected locations for communication modification or other purposes.

In another embodiment, this invention is used to alter at least one selected region of plasma in the ionosphere to establish a defined layer of plasma having an increased charged particle density. Once this layer is established, and while maintaining the transmission of the main beam of circularly polarized electromagnetic radiation, the main beam is modulated and/or at least one second different, modulated electromagnetic radiation beam is transmitted from at least one separate source at a different frequency which will be absorbed in the plasma layer. The amplitude of the frequency of the main beam and/or the second beam or beams is modulated in resonance with at least one known oscillation mode in the selected region or regions to excite the known oscillation mode to propagate a known frequency wave or waves throughout the ionosphere.

BRIEF DESCRIPTION OF THE DRAWINGS

The actual construction, operation, and apparent advantages of this invention will be better understood by referring to the drawings in which like numerals identify like parts and in which:

FIG. **1** is a simplified schematical view of the earth (not to scale) with a magnetic field (force) line along which the present invention is carried out;

FIG. **2** is one embodiment within the present invention in which a selected region of plasma is raised to a higher altitude;

FIG. **3** is a simplified, idealized representation of a physical phenomenon involved in the present invention; and

FIG. **4** is a schematic view of another embodiment within the present invention.

FIG. **5** is a schematic view of an apparatus embodiment within this invention.

BEST MODES FOR CARRYING OUT THE INVENTION

The earth's magnetic field is somewhat analogous to a dipole bar magnet. As such, the earth's magnetic field contains numerous divergent field or force lines, each line intersecting the earth's surface at points on opposite sides of the Equator. The field lines which intersect the earth's surface near the poles have apexes which lie at the furthest points in the earth's magnetosphere while those closest to the Equator have apexes which reach only the lower portion of the magnetosphere.

At various altitudes above the earth's surface, e.g., in both the ionosphere and the magnetosphere, plasma is naturally present along these field lines. This plasma consists of equal numbers of positively and negatively charged particles (i.e., electrons and ions) which are guided by the field line. It is well established that a charged particle in a magnetic field gyrates about field lines, the center of gyration at any instance being called the "guiding center" of the particle. As the gyrating particle moves along a field line in a uniform field, it will follow a helical path about its guiding center, hence linear motion, and will remain on the field line. Electrons and ions both follow helical paths around a field line but rotate in opposite directions. The frequencies at which the electrons and ions rotate about the field line are called gyromagnetic frequencies or cyclotron frequencies because they are identical with the expression for the angular frequencies of gyration of particles in a cyclotron. The cyclotron frequency of ions in a given magnetic field is less than that of electrons, in inverse proportion to their masses.

If the particles which form the plasma along the earth's field lines continued to move with a constant pitch angle, often designated "alpha", they would soon impact on the earth's surface. Pitch angle alpha is defined as the angle between the direction of the earth's magnetic field and the velocity (V) of the particle. However, in converging force fields, the pitch angle does change in such a way as to allow the particle to

turn around and avoid impact. Consider a particle moving along a field line down toward the earth. It moves into a region of increasing magnetic field strength and therefore sine alpha increases. But sine alpha can only increase to 1.0, at which point, the particle turns around and starts moving up along the field line, and alpha decreases. The point at which the particle turns around is called the mirror point, and there alpha equals ninety degrees. This process is repeated at the other end of the field line where the same magnetic field strength value B, namely Bm, exists. The particle again turns around and this is called the "conjugate point" of the original mirror point. The particle is therefore trapped and bounces between the two magnetic mirrors. The particle can continue oscillating in space in this manner for long periods of time. The actual place where a particle will mirror can be calculated from the following:

$$\sin^2 \text{alpha}_o = B_o/B_m \quad (1)$$

wherein:

alpha_o = equatorial pitch angle of particle

B_o = equatorial field strength on a particular field line

B_m = field strength at the mirror point

Recent discoveries have established that there are substantial regions of naturally trapped particles in space which are commonly called "trapped radiation belts". These belts occur at altitudes greater than about 500 km and accordingly lie in the magnetosphere and mostly above the ionosphere.

The ionosphere, while it may overlap some of the trapped-particle belts, is a region in which hydrostatic forces govern its particle distribution in the gravitational field. Particle motion within the ionosphere is governed by both hydrodynamic and electrodynamic forces. While there are few trapped particles in the ionosphere, nevertheless, plasma is present along field lines in the ionosphere. The charged particles which form this plasma move between collisions with other particles along similar helical paths around the field lines and although a particular particle may diffuse downward into the earth's lower atmosphere or lose energy and diverge from its original field line due to collisions with other particles, these charged particles are normally replaced by other available charged particles or by particles that are ionized by collision with said particle. The electron density (N_e) of the plasma will vary with the actual conditions and locations involved. Also, neutral particles, ions, and electrons are present in proximity to the field lines.

The production of enhanced ionization will also alter the distribution of atomic and molecular constituents of the atmosphere, most notably through increased atomic nitrogen concentration. The upper atmosphere is normally rich in atomic oxygen (the dominant atmospheric constituent above 200 km altitude), but atomic nitrogen is normally relatively rare. This can be expected to manifest itself in increased airglow, among other effects.

As known in plasma physics, the characteristics of a plasma can be altered by adding energy to the charged particles or by ionizing or exciting additional particles to increase the density of the plasma. One way to do this is by heating the plasma which can be accomplished in different ways, e.g., ohmic, magnetic compression, shock waves, magnetic pumping, electron cyclotron resonance, and the like.

Since electron cyclotron resonance heating is involved in the present invention, a brief discussion of same is in order. Increasing the energy of electrons in a plasma by invoking electron cyclotron resonance heating, is based on a principle similar to that utilized to accelerate charged particles in a cyclotron. If a plasma is confined by a static axial magnetic field of strength B, the charged particles will gyrate about the lines of force with a frequency given, in hertz, as $f_g = 1.54 \times 10^3 B/A$, where: B = magnetic field strength in gauss, and A = mass number of the ion.

Suppose a time-varying field of this frequency is superimposed on the static field B confining the plasma, by passage of a radiofrequency current through a coil which is concentric with that producing the axial field, then in each half-cycle of their rotation about the field lines, the charged particles acquire energy from the oscillating electric field associated with the radio frequency. For example, if B is 10,000 gauss, the frequency of the field which is in resonance with protons in a plasma is 15.4 megahertz.

As applied to electrons, electron cyclotron resonance heating requires an oscillating field having a definite frequency determined by the strength of the confining field. The radio-frequency radiation produces time-varying fields (electric and magnetic), and the electric field accelerates the charged particle. The energized electrons share their energy with ions and neutrals by undergoing collisions with these particles, thereby effectively raising the temperature of the electrons, ions, and neutrals. The apportionment of energy among these species is determined by collision frequencies. For a more detailed understanding of the physics involved, see "Controlled Thermonuclear Reactions", Glasstone and Lovberg, D. Van Nostrand Company, Inc., Princeton, N.J., 1960 and "The Radiation Belt and Magnetosphere", Hess, Blaisdell Publishing Company, 1968, both of which are incorporated herein by reference.

Referring now to the drawings, the present invention provides a method and apparatus for altering at least one region of plasma which lies along a field line, particularly when it passes through the ionosphere and/or magnetosphere. FIG. **1** is a simplified illustration of the earth **10** and one of its dipole magnetic force or field lines **11**. As will be understood, line **11** may be any one of the numerous naturally existing field lines and the actual geographical locations **13** and **14** of line **11** will be chosen based on a particular operation to be carried out. The actual locations at which field lines intersect the earth's surface is documented and is readily ascertainable by those skilled in the art.

Line **11** passes through region R which lies at an altitude above the earth's surface. A wide range of altitudes are useful given the power that can be employed by the practice of this invention. The electron cyclotron resonance heating effect can be made to act on electrons anywhere above the surface of the earth. These electrons may be already present in the atmosphere, ionosphere, and/or magnetosphere of the earth, or can be artificially generated by a variety of means such as x-ray beams, charged particle beams, lasers, the plasma sheath surrounding an object such as a missile or meteor, and the like. Further, artificial particles, e.g., electrons, ions, etc., can be injected directly into region R from an earth-launched rocket or orbiting satellite carrying, for example, a payload of radioactive beta-decay material; alpha emitters; an electron accelerator; and/or ionized gases such as hydrogen; see U.S. Pat. No. 4,042,196. The altitude can be greater than about 50 km if desired,

e.g., can be from about 50 km to about 800 km, and, accordingly may lie in either the ionosphere or the magnetosphere or both. As explained above, plasma will be present along line **11** within region R and is represented by the helical line **12**. Plasma **12** is comprised of charged particles (i.e., electrons and ions) which rotate about opposing helical paths along line **11**.

Antenna **15** is positioned as close as is practical to the location **14** where line **11** intersects the earth's surface. Antenna **15** may be of any known construction for high directionality, for example, a phased array, beam spread angle (θ) type. See "The MST Radar at Poker Flat, Alaska", Radio Science, Vol. 15, No. 2, Mar.-Apr. 1980, pps. 213–223, which is incorporated herein by reference. Antenna **15** is coupled to transmitter **16** which generates a beam of high frequency electromagnetic radiation at a wide range of discrete frequencies, e.g., from about 20 to about 1800 kilohertz (kHz).

Transmitter **16** is powered by power generator means **17** which is preferably comprised of one or more large, commercial electrical generators. Some embodiments of the present invention require large amounts of power, e.g., up to 10^9 to 10^{11} watts, in continuous wave or pulsed power. Generation of the needed power is within the state of the art. Although the electrical generators necessary for the practice of the invention can be powered in any known manner, for example, by nuclear reactors, hydroelectric facilities, hydrocarbon fuels, and the like, this invention, because of its very large power requirement in certain applications, is particularly adapted for use with certain types of fuel sources which naturally occur at strategic geographical locations around the earth. For example, large reserves of hydrocarbons (oil and natural gas) exist in Alaska and Canada. In northern Alaska, particularly the North Slope region, large reserves are currently readily available. Alaska and northern Canada also are ideally located geographically as to magnetic latitudes. Alaska provides easy access to magnetic field lines that are especially suited to the practice of this invention, since many field lines which extend to desirable altitudes for this invention intersect the earth in Alaska. Thus, in Alaska, there is a unique combination of large, accessible fuel sources at desirable field line intersections. Further, a particularly desirable fuel source for the generation of very large amounts of electricity is present in Alaska in abundance, this source being natural gas. The presence of very large amounts of clean-burning natural gas in Alaskan latitudes, particularly on the North Slope, and the availability of magnetohydrodynamic (MHD), gas turbine, fuel cell, electrogasdynamic (EGD) electric generators which operate very efficiently with natural gas provide an ideal power source for the unprecedented power requirements of certain of the applications of this invention. For a more detailed discussion of the various means for generating electricity from hydrocarbon fuels, see "Electrical Aspects of Combustion", Lawton and Weinberg, Clarendon Press, 1969. For example, it is possible to generate the electricity directly at the high frequency needed to drive the antenna system. To do this, typically the velocity of flow of the combustion gases (v), past magnetic field perturbation of dimension d (in the case of MHD), follow the rule:

$$v=df$$

where f is the frequency at which electricity is generated. Thus, if $v=1.78\times10^6$ cm/sec and $d=1$ cm then electricity would be generated at a frequency of 1.78 mHz.

Put another way, in Alaska, the right type of fuel (natural gas) is naturally present in large amounts and at just the right magnetic latitudes for the most efficient practice of this invention, a truly unique combination of circumstances. Desirable magnetic latitudes for the practice of this invention interest the earth's surface both northerly and southerly of the equator, particularly desirable latitudes being those, both northerly and southerly, which correspond in magnitude with the magnetic latitudes that encompass Alaska.

Referring now to FIG. **2** a first ambodiment is illustrated where a selected region R_1 of plasma **12** is altered by electron cyclotron resonance heating to accelerate the electrons of plasma **12**, which are following helical paths along field line **11**.

To accomplish this result, electromagnetic radiation is transmitted at the outset, essentially parallel to line **11** via antenna **15** as right hand circularly polarized radiation wave **20**. Wave **20** has a frequency which will excite electron cyclotron resonance with plasma **12** at its initial or original altitude. This frequency will vary depending on the electron cyclotron resonance of region R_1 which, in turn, can be determined from available data based on the altitudes of region R_1, the particular field line **11** being used, the strength of the earth's magnetic field, etc. Frequencies of from about 20 to about 7200 kHz, preferably from about 20 to about 1800 kHz can be employed. Also, for any given application, there will be a threshhold (minimum power level) which is needed to produce the desired result. The minimum power level is a function of the level of plasma production and movement required, taking into consideration any loss processes that may be dominant in a particular plasma or propagation path.

As electron cyclotron resonance is established in plasma **12**, energy is transferred from the electromagnetic radiation **20** into plasma **12** to heat and accelerate the electrons therein and, subsequently, ions and neutral particles. As this process continues, neutral particles which are present within R_1 are ionized and absorbed into plasma **12** and this increases the electron and ion densities of plasma **12**. As the electron energy is raised to values of about 1 kilo electron volt (kev), the generated mirror force (explained below) will direct the excited plasma **12** upward along line **11** to form a plume R_2 at an altitude higher than that of R_1.

Plasma acceleration results from the force on an electron produced by a nonuniform static magnetic field ($\overline{B}$). The force, called the mirror force, is given by

$$F=-\mu\nabla B \quad (2)$$

where μ is the electron magnetic moment and $\nabla\overline{B}$ is the gradient of the magnetic field, μ being further defined as:

$$W_{\perp}/B=mV_{\perp}^2/2B$$

where $W_{\perp}$ is the kinetic energy in the direction perpendicular to that of the magnetic field lines and B is the magnetic field strength at the line of force on which the guiding center of the particle is located. The force as represented by equation (2) is the force which is responsible for a particle obeying equation (1).

Since the magnetic field is divergent in region R_1, it can be shown that the plasma will move upwardly from the heating region as shown in FIG. **1** and further it can be shown that

$$\tfrac{1}{2}M_eV_{e\perp}^2(x) \simeq \tfrac{1}{2}M_eV_{e\perp}^2(Y) + \tfrac{1}{2}M_iV_{i\parallel}^2(Y) \quad (3)$$

where the left hand side is the initial electron transverse kinetic energy; the first term on the right is the transverse electron kinetic energy at some point (Y) in the expanded field region, while the final term is the ion kinetic energy parallel to B at point (Y). This last term is what constitutes the desired ion flow. It is produced by an electrostatic field set up by electrons which are accelerated according to Equation (2) in the divergent field region and pulls ions along with them. Equation (3) ignores electron kinetic energy parallel to B because $V_{e\parallel} \approx V_{i\parallel}$, so the bulk of parallel kinetic energy resides in the ions because of their greater masses. For example, if an electromagnetic energy flux of from about 1 to about 10 watts per square centimeter is applied to region R, whose altitude is 115 km, a plasma having a density (N_e) of 10^{12} per cubic centimeter will be generated and moved upward to region R_2 which has an altitude of about 1000 km. The movement of electrons in the plasma is due to the mirror force while the ions are moved by ambipolar diffusion (which results from the electrostatic field). This effectively "lifts" a layer of plasma **12** from the ionosphere and/or magnetosphere to a higher elevation R_2. The total energy required to create a plasma with a base area of 3 square kilometers and a height of 1000 km is about 3×10^{13} joules.

FIG. **3** is an idealized representation of movement of plasma **12** upon excitation by electron cyclotron resonance within the earth's divergent force field. Electrons (e) are accelerated to velocities required to generate the necessary mirror force to cause their upward movement. At the same time neutral particles (n) which are present along line **11** in region $R_1$are ionized and become part of plasma **12**. As electrons (e) move upward along line **11**, they drag ions (i) and neutrals (n) with them but at an angle θ of about 13 degrees to field line **11**. Also, any particulates that may be present in region R_1, will be swept upwardly with the plasma. As the charged particles of plasma **12** move upward, other particles such as neutrals within or below R_1, move in to replace the upwardly moving particles. These neutrals, under some conditions, can drag with them charged particles.

For example, as a plasma moves upward, other particles at the same altitude as the plasma move horizontally into the region to replace the rising plasma and to form new plasma. The kinetic energy developed by said other particles as they move horizontally is, for example, on the same order of magnitude as the total zonal kinetic energy of stratospheric winds known to exist.

Referring again to FIG. **2**, plasma **12** in region R_1 is moved upward along field line **11**. The plasma **12** will then form a plume (cross-hatched area in FIG. **2**) which will be relatively stable for prolonged periods of time. The exact period of time will vary widely and be determined by gravitational forces and a combination of radiative and diffusive loss terms. In the previous detailed example, the calculations were based on forming a plume by producing 0^+energies of 2 ev/particle. About 10 ev per particle would be required to expand plasma **12** to apex point C (FIG. **1**). There at least some of the particles of plasma **12** will be trapped and will oscillate between mirror points along field line **11**. This oscillation will then allow additional heating of the trapped plasma **12** by stochastic heating which is associated with trapped and oscillating particles. See "A New Mechanism for Accelerating Electrons in the Outer Ionosphere" by R. A. Helliwell and T. F. Bell, Journal of Geophysical Research, Vol. 65, No. 6, June, 1960. This is preferably carried out at an altitude of at least 500 km.

The plasma of the typical example might be employed to modify or disrupt microwave transmissions of satellites. If less than total black-out of transmission is desired (e.g., scrambling by phase shifting digital signals), the density of the plasma (N_e) need only be at least about 10^6 per cubic centimeter for a plasma orginating at an altitude of from about 250 to about 400 km and accordingly less energy (i.e., electromagnetic radiation), e.g., 10^8 joules need be provided. Likewise, if the density $N_e$is on the order of 10^8, a properly positioned plume will provide a reflecting surface for VHF waves and can be used to enhance, interfere with, or otherwise modify communication transmissions. It can be seen from the foregoing that by appropriate application of various aspects of this invention at strategic locations and with adequate power sources, a means and method is provided to cause interference with or even total disruption of communications over a very large portion of the earth. This invention could be employed to disrupt not only land based communications, both civilian and military, but also airborne communications and sea communications (both surface and subsurface). This would have significant military implications, particularly as a barrier to or confusing factor for hostile missiles or airplanes. The belt or belts of enhanced ionization produced by the method and apparatus of this invention, particularly if set up over Northern Alaska and Canada, could be employed as an early warning device, as well as a communications disruption medium. Further, the simple ability to produce such a situation in a practical time period can by itself be a deterring force to hostile action. The ideal combination of suitable field lines intersecting the earth's surface at the point where substantial fuel sources are available for generation of very large quantitities of electromagnetic power, such as the North Slope of Alaska, provides the wherewithal to accomplish the foregoing in a practical time period, e.g., strategic requirements could necessitate achieving the desired altered regions in time periods of two minutes or less and this is achievable with this invention, especially when the combination of natural gas and magnetohydrodynamic, gas turbine, fuel cell and/or EGD electric generators are employed at the point where the useful field lines intersect the earth's surface. One feature of this invention which satisfies a basic requirement of a weapon system, i.e., continuous checking of operability, is that small amounts of power can be generated for operability checking purposes. Further, in the exploitation of this invention, since the main electromagnetic beam which generates the enhanced ionized belt of this invention can be modulated itself and/or one or more additional electromagnetic radiation waves can be impinged on the ionized region formed by this invention as will be described in greater detail herein after with respect to FIG. **4**, a substantial amount of randomly modulated signals of very large power magnitude can be generated in a highly nonlinear mode. This can cause confusion of or interference with or even complete disruption of guidance systems employed by

even the most sophisticated of airplanes and missiles. The ability to employ and transmit over very wide areas of the earth a plurality of electromagnetic waves of varying frequencies and to change same at will in a random manner, provides a unique ability to interfere with all modes of communications, land, sea, and/or air, at the same time. Because of the unique juxtaposition of usable fuel source at the point where desirable field lines intersect the earth's surface, such wide ranging and complete communication interference can be achieved in a resonably short period of time. Because of the mirroring phenomenon discussed hereinabove, it can also be prolonged for substantial time periods so that it would not be a mere transient effect that could simply be waited out by an opposing force. Thus, this invention provides the ability to put unprecedented amounts of power in the earth's atmosphere at strategic locations and to maintain the power injection level, particularly if random pulsing is employed, in a manner far more precise and better controlled than heretofore accomplished by the prior art, particularly by the detonation of nuclear devices of various yeilds at various altitudes. Where the prior art approaches yielded merely transitory effects, the unique combination of fuel and desirable field lines at the point where the fuel occurs allows the establishment of, compared to prior art approaches, precisely controlled and long-lasting effects which cannot, practically speaking, simply be waited out. Further, by knowing the frequencies of the various electromagnetic beams employed in the practice of this invention, it is possible not only to interfere with third party communications but to take advantage of one or more such beams to carry out a communications network even though the rest of the world's communications are disrupted. Put another way, what is used to disrupt another's communications can be employed by one knowledgeable of this invention as a communications network at the same time. In addition, once one's own communication network is established, the far-reaching extent of the effects of this invention could be employed to pick up communication signals of other for intelligence purposes. Thus, it can be seen that the disrupting effects achievable by this invention can be employed to benefit by the party who is practicing this invention since knowledge of the various electromagnetic waves being employed and how they will vary in frequency and magnitude can be used to an advantage for positive communication and eavesdropping purposes at the same time. However, this invention is not limited to locations where the fuel source naturally exists or where desirable field lines naturally intersect the earth's surface. For example, fuel, particularly hydrocarbon fuel, can be transported by pipeline and the like to the location where the invention is to be practiced.

FIG. 4 illustrates another embodiment wherein a selected region of plasma R_3which lies within the earth's ionosphere is altered to increase the density thereof whereby a relatively stable layer **30** of relatively dense plasma is maintained within region R_3. Electromagnetic radiation is transmitted at the outset essentially parallel to field line **11** via antenna **15** as a right hand circularly polarized wave and at a frequency (e.g., 1.78 megahertz when the magnetic field at the desired altitude is 0.66 gauss) capable of exciting electron cyclotron resonance in plasma **12** at the particular altitude of plasma **12**. This causes heating of the particles (electrons, ions, neutrals, and particulates) and ionization of the uncharged particles adjacent line **11**, all of which are absorbed into plasma **12** to increase the density thereof. The power transmitted, e.g., 2×10^6 watts for up to 2 minutes heating time, is less than that required to generate the mirror force F required to move plasma **12** upward as in the previous embodiment.

While continuing to transmit electromagnetic radiation **20** from antenna **15**, a second electromagnetic radiation beam **31**, which is at a defined frequency different from the radiation from antenna **15**, is transmitted from one or more second sources via antenna **32** into layer **30** and is absorbed into a portion of layer **30** (cross-hatched area in FIG. **4**). The electromagnetic radiation wave from antenna **32** is amplitude modulated to match a known mode of oscillation f_3 in layer **30**. This creates a resonance in layer **30** which excites a new plasma wave **33** which also has a frequency of f_3 and which then propogates through the ionosphere. Wave **33** can be used to improve or disrupt communications or both depending on what is desired in a particular application. Of course, more than one new wave **33** can be generated and the various new waves can be modulated at will and in a highly nonlinear fashion.

FIG. **5** shows apparatus useful in this invention, particularly when those applications of this invention are employed which require extremely large amounts of power. In FIG. **5** there is shown the earth's surface **40** with a well **41** extending downwardly thereinto until it penetrates hydrocarbon producing reservoir **42**. Hydrocarbon reservoir **42** produces natural gas alone or in combination with crude oil. Hydrocarbons are produced from reservoir **42** through well **41** and wellhead **43** to a treating system **44** by way of pipe **45**. In treater **44**, desirable liquids such as crude oil and gas condensates are separated and recovered by way of pipe **46** while undesirable gases and liquids such as water, H_2S, and the like are separated by way of pipe **47**. Desirable gases such as carbon dioxide are separated by way of pipe **48**, and the remaining natural gas stream is removed from treater **44** by way of pipe **49** for storage in conventional tankage means (not shown) for future use and/or use in an electrical generator such as a magnetohydrodynamic, gas turbine, fuel cell or EGD generator **50**. Any desired number and combination of different types of electric generators can be employed in the practice of this invention. The natural gas is burned in generator **50** to produce substantial quantities of electricity which is then stored and/or passed by way of wire **51** to a transmitter **52** which generates the electromagnetic radiation to be used in the method of this invention. The electromagnetic radiation is then passed by way of wire **53** to antenna **54** which is located at or near the end of field line **11**. Antenna **54** sends circularly polarized radiation wave **20** upwards along field line **11** to carry out the various methods of this invention as described hereinabove.

Of course, the fuel source need not be used in its naturally-occurring state but could first be converted to another second energy source form such as hydrogen, hydrazine and the like, and electricity then generated from said second energy source form.

It can be seen from the foregoing that when desirable field line **11** intersects earth's surface **40** at or near a large naturally-occurring hydrocarbon source **42**, exceedingly large amounts of power can be very efficiently produced and transmitted in the direction of field lines. This is particularly so when the fuel source is natural gas and magnetohydrodynamic generators are employed. Further, this can all be accomplished in a

relatively small physical area when there is the unique coincidence of fuel source **42** and desirable field line **11**. Of course, only one set of equipment is shown in FIG. **5** for sake of simplicity. For a large hydrocarbon reservoir **42**, a plurality of wells **41** can be employed to feed one or more storage means and/or treaters and as large a number of generators **55** as needed to power one or more transmitters **52** and one or more antennas **54**. Since all of the apparatus **44** through **54** can be employed and used essentially at the sight where naturally-occurring fuel source **42** is located, all the necessary electromagnetic radiation **20** is generated essentially at the same location as fuel source **42**. This provides for a maximum amount of usable electromagnetic radiation **20** since there are no significant storage or transportation losses to be incurred. In other words, the apparatus is brought to the sight of the fuel source where desirable field line **11** intersects the earth's surface **40** on or near the geographical location of fuel source **42**, fuel source **42** being at a desirable magnetic latitude for the practice of this invention, for example, Alaska.

The generation of electricity by motion of a conducting fluid through a magnetic field, i.e., magnetohydrodynamics (MHD), provides a method of electric power generation without moving mechanical parts and when the conducting fluid is a plasma formed by combustion of a fuel such as natural gas, an idealized combination of apparatus is realized since the very clean-burning natural gas forms the conducting plasma in an efficient manner and the thus formed plasma, when passed through a magnetic field, generates electricity in a very efficient manner. Thus, the use of fuel source **42** to generate a plasma by combustion thereof for the generation of electricity essentially at the site of occurrence of the fuel source is unique and ideal when high power levels are required and desirable field lines **11** intersect the earth's surface **40** at or near the site of fuel source **42**. A particular advantage for MHD generators is that they can be made to generate large amounts of power with a small volume, light weight device. For example, a 1000 megawatt MHD generator can be construed using superconducting magnets to weigh roughly 42,000 pounds and can be readily air lifted.

This invention has a phenomenal variety of possible ramifications and potential future developments. As alluded to earlier, missile or aircraft destruction, deflection, or confusion could result, particularly when relativistic particles are employed. Also, large regions of the atmosphere could be lifted to an unexpectedly high altitude so that missiles encounter unexpected and unplanned drag forces with resultant destruction or deflection of same. Weather modification is possible by, for example, altering upper atmosphere wind patterns or altering solar absorption patterns by constructing one or more plumes of atmospheric particles which will act as a lens or focusing device. Also as alluded to earlier, molecular modifications of the atmosphere can take place so that positive environmental effects can be achieved. Besides actually changing the molecular composition of an atmospheric region, a particular molecule or molecules can be chosen for increased presence. For example, ozone, nitrogen, etc. concentrations in the atmosphere could be artificially increased. Similarly, environmental enhancement could be achieved by causing the breakup of various chemical entities such as carbon dioxide, carbon monoxide, nitrous oxides, and the like. Transportation of entities can also be realized when advantage is taken of the drag effects caused by regions of the atmosphere moving up along diverging field lines. Small micron sized particles can be then transported, and, under certain circumstances and with the availability of sufficient energy, larger particles or objects could be similarly affected. Particles with desired characteristics such as tackiness, reflectivity, absorptivity, etc., can be transported for specific purposes or effects. For example, a plume of tacky particles could be established to increase the drag on a missile or satellite passing therethrough. Even plumes of plasma having substantially less charged particle density than described above will produce drag effects on missiles which will affect a lightweight (dummy) missile in a manner substantially different than a heavy (live) missile and this affect can be used to distinguish between the two types of missiles. A moving plume could also serve as a means for supplying a space station or for focusing vast amount of sunlight on selected portions of the earth. Surveys of global scope could also be realized because the earth's natural magnetic field could be significantly altered in a controlled manner by plasma beta effects resulting in, for example, improved magnetotelluric surveys. Electromagnetic pulse defenses are also possible. The earth's magnetic field could be decreased or disrupted at appropriate altitudes to modify or eliminate the magnetic field in high Compton electron generation (e.g., from high altitude nuclear bursts) regions. High intensity, well controlled electrical fields can be provided in selected locations for various purposes. For example, the plasma sheath surrounding a missile or satellite could be used as a trigger for activating such a high intensity field to destroy the missile or satellite. Further, irregularities can be created in the ionosphere which will interfere with the normal operation of various types of radar, e.g., synthetic aperture radar. The present invention can also be used to create artificial belts of trapped particles which in turn can be studied to determine the stability of such parties. Still further, plumes in accordance with the present invention can be formed to simulate and/or perform the same functions as performed by the detonation of a "heave" type nuclear device without actually having to detonate such a device. Thus it can be seen that the ramifications are numerous, far-reaching, and exceedingly varied in usefulness.

I claim:

1. A method for altering at least one region normally existing above the earth's surface with electromagnetic radiation using naturally-occurring and diverging magnetic field lines of the earth comprising transmitting first electromagnetic radiation at a frequency between 20 and 7200 kHz from the earth's surface, said transmitting being conducted essentially at the outset of transmission substantially parallel to and along at least one of said field lines, adjusting the frequency of said first radiation to a value which will excite electron cyclotron resonance at an initial elevation at least 50 km above the earth's surface, whereby in the region in which said electron cyclotron resonance takes place heating, further ionization, and movement of both charged and neutral particles is effected, said cyclotron resonance excitation of said region is continued until the electron concentration of said region reaches a value of at least 10^6 per cubic centimeter and has an ion energy of at least 2 ev.

2. The method of claim **1** including the step of providing artificial particles in said at least one region which are excited by said electron cyclotron resonance.

3. The method of claim **2** wherein said artificial particles are provided by injecting same into said at least one region from an orbiting satellite.

4. The method of claim **1** wherein said threshold excitation of electron cyclotron resonance is about 1 watt per cubic centimeter and is sufficient to cause movement of a plasma region along said diverging magnetic field lines to an altitude higher than the altitude at which said excitation was initiated.

5. The method of claim **4** wherein said rising plasma region pulls with it a substantial portion of neutral particles of the atmosphere which exist in or near said plasma region.

6. The method of claim **1** wherein there is provided at least one separate source of second electromagnetic radiation, said second radiation having at least one frequency different from said first radiation, impinging said at least one second radiation on said region while said region is undergoing electron cyclotron resonance excitation caused by said first radiation.

7. The method of claim **6** wherein said second radiation has a frequency which is absorbed by said region.

8. The method of claim **6** wherein said region is plasma in the ionosphere and said second radiation excites plasma waves within said ionosphere.

9. The method of claim **8** wherein said electron concentration reaches a value of at least 10^{12} per cubic centimeter.

10. The method of claim **8** wherein said excitation of electron cyclotron resonance is initially carried out within the ionosphere and is continued for a time sufficient to allow said region to rise above said ionosphere.

11. The method of claim **1** wherein said excitation of electron cyclotron resonance is carried out above about 500 kilometers and for a time of from 0.1 to 1200 seconds such that multiple heating of said plasma region is achieved by means of stochastic heating in the magnetosphere.

12. The method of claim **1** wherein said first electromagnetic radiation is right hand circularly polarized in the northern hemisphere and left hand circularly polarized in the southern hemisphere.

13. The method of claim **1** wherein said electromagnetic radiation is generated at the site of a naturally-occurring hydrocarbon fuel source, said fuel source being located in at least one of northerly or southerly magnetic latitudes.

14. The method of claim **13** wherein said fuel source is natural gas and electricity for generating said electromagnetic radiation is obtained by burning said natural gas in at least one of magnetohydrodynamic, gas turbine, fuel cell, and EGD electric generators located at the site where said natural gas naturally occurs in the earth.

15. The method of claim **14** wherein said site of natural gas is within the magnetic latitudes that encompass Alaska.

* * * * *

United States Patent [19]
Eastlund

[11] **Patent Number:** **4,686,605**
[45] **Date of Patent:** **Aug. 11, 1987**

[54] **METHOD AND APPARATUS FOR ALTERING A REGION IN THE EARTH'S ATMOSPHERE, IONOSPHERE, AND/OR MAGNETOSPHERE**

[75] Inventor: **Bernard J. Eastlund**, Spring, Tex.

[73] Assignee: **APTI, Inc.**, Los Angeles, Calif.

[21] Appl. No.: **690,333**

[22] Filed: **Jan. 10, 1985**

[51] **Int. Cl.**^{4} **H05B 6/64**; H05C 3/00; H05H 1/46

[52] **U.S. Cl.** .... **361/231**; 89/1.11; 380/59; 244/158 R

[58] **Field of Search** .... 361/230, 231; 244/158 R; 376/100; 89/1.11; 380/59

[56] **References Cited**

PUBLICATIONS

Liberty Magazine, (2/35) p. 7 N. Tesla.
New York Times (9/22/40) Section 2, p. 7 W. L. Laurence.
New York Times (12/8/15) p. 8 Col. 3.

Primary Examiner—Salvatore Cangialosi
Attorney, Agent, or Firm—Roderick W. MacDonald

[57] **ABSTRACT**

A method and apparatus for altering at least one selected region which normally exists above the earth's surface. The region is excited by electron cyclotron resonance heating to thereby increase its charged particle density. In one embodiment, circularly polarized electromagnetic radiation is transmitted upward in a direction substantially parallel to and along a field line which extends through the region of plasma to be altered. The radiation is transmitted at a frequency which excites electron cyclotron resonance to heat and accelerate the charged particles. This increase in energy can cause ionization of neutral particles which are then absorbed as part of the region thereby increasing the charged particle density of the region.

15 Claims, 5 Drawing Figures

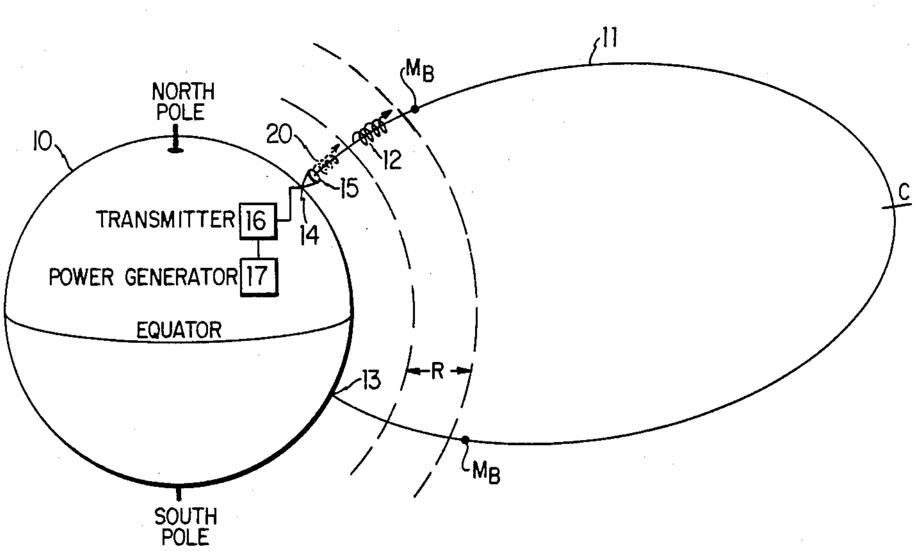

FIG. 1

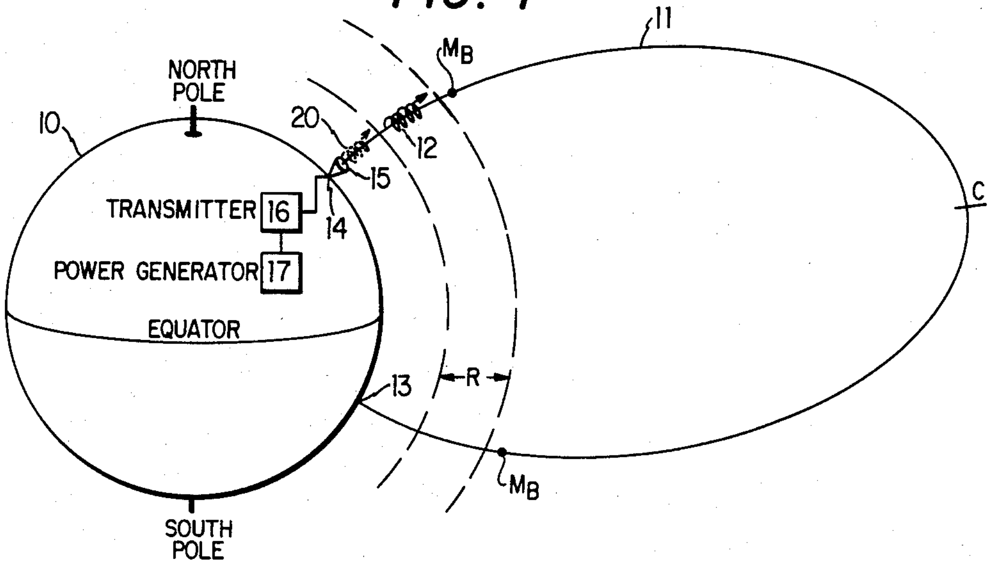
11
MB
NORTH POLE
10
20
12
15
C
TRANSMITTER 16
14
POWER GENERATOR 17
EQUATOR
13
R
MB
SOUTH POLE

FIG. 2

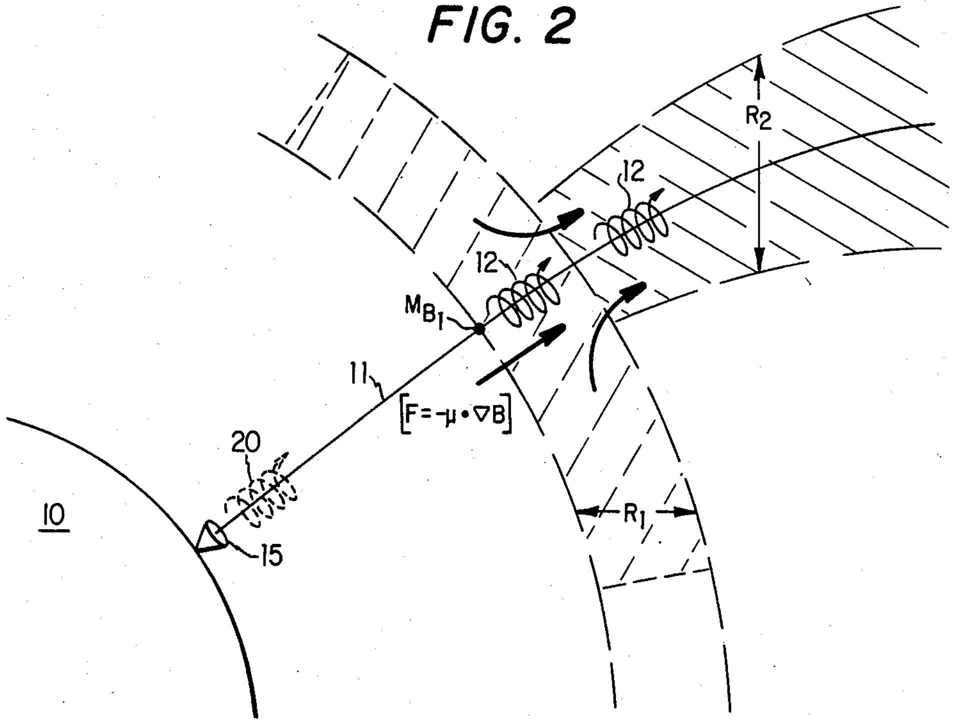
R2
MB1
[F=-μ•∇B]
R1

FIG. 3

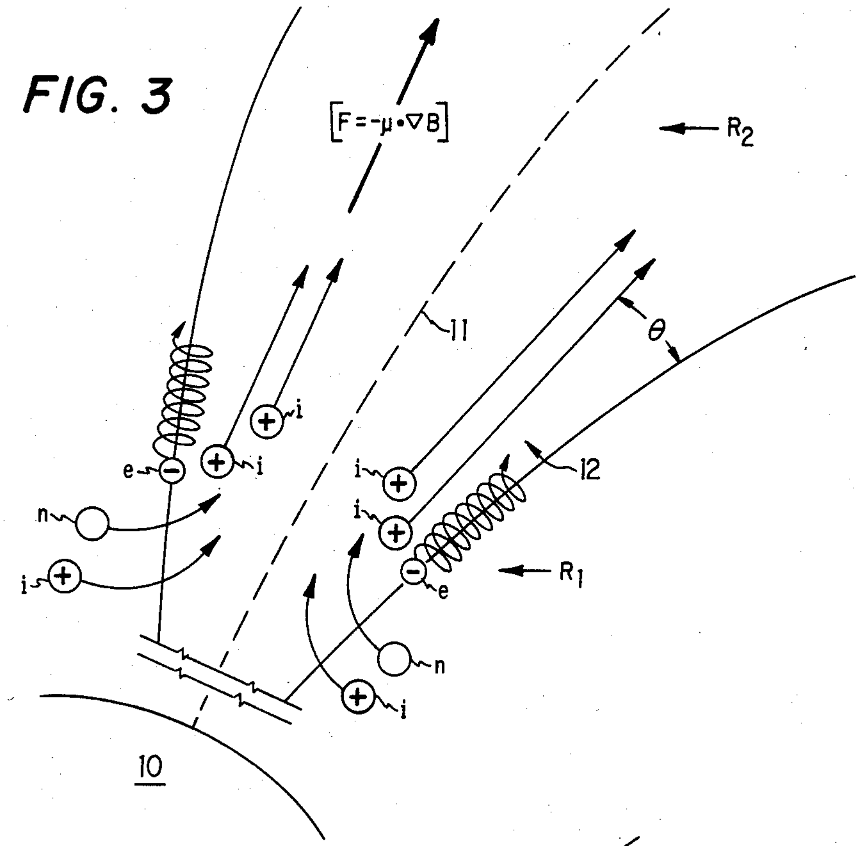

[F = -μ·∇B]
R2
11
θ
i
e
i
n
i
i
i
12
R1
e
n
i
10

FIG. 4

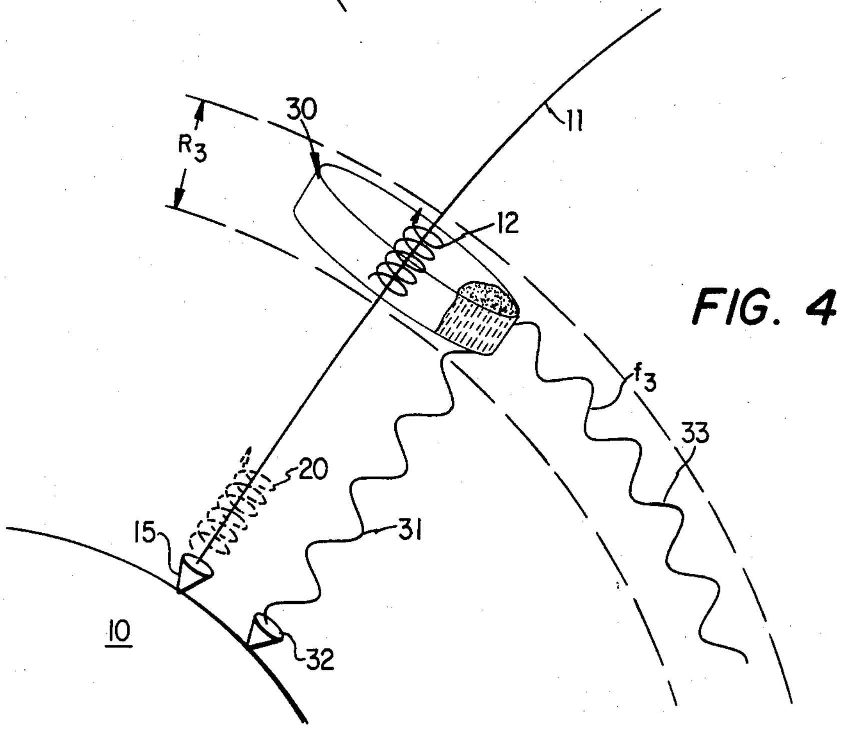

30
R3
11
12
f3
33
20
15
31
10
32

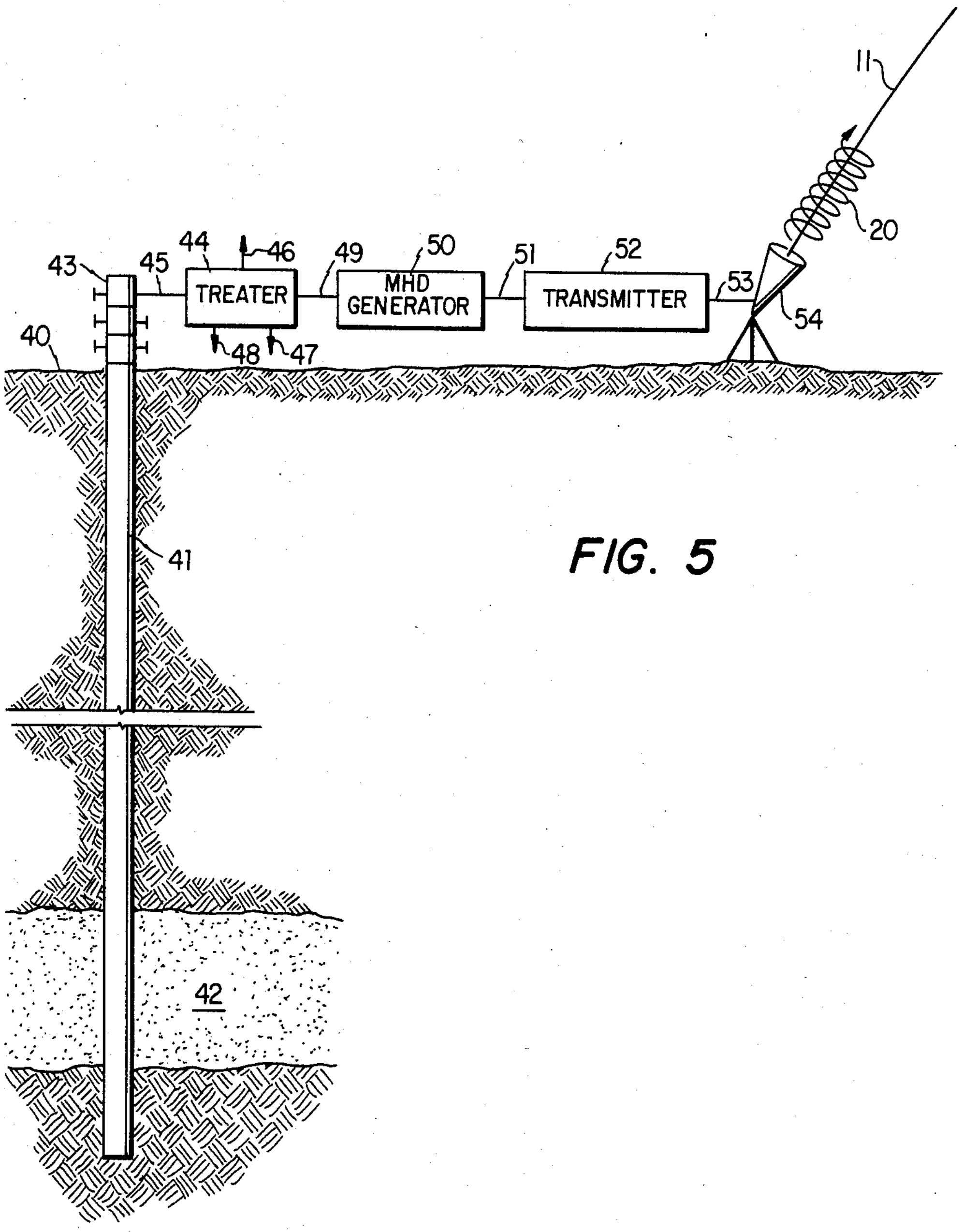
TREATER
MHD GENERATOR
TRANSMITTER

FIG. 5

METHOD AND APPARATUS FOR ALTERING A REGION IN THE EARTH'S ATMOSPHERE, IONOSPHERE, AND/OR MAGNETOSPHERE

DESCRIPTION

1. Technical Field

This invention relates to a method and apparatus for altering at least one selected region normally existing above the earth's surface and more particularly relates to a method and apparatus for altering said at least one region by initially transmitting electromagnetic radiation from the earth's surface essentially parallel to and along naturally-occurring, divergent magnetic field lines which extend from the earth's surface through the region or regions to be altered.

2. Background Art

In the late 1950's, it was discovered that naturally-occuring belts exist at high altitudes above the earth's surface, and it is now established that these belts result from charged electrons and ions becoming trapped along the magnetic lines of force (field lines) of the earth's essentially dipole magnetic field. The trapped electrons and ions are confined along the field lines between two magnetic mirrors which exist at spaced apart points along those field lines. The trapped electrons and ions move in helical paths around their particular field lines and "bounce" back and forth between the magnetic mirrors. These trapped electrons and ions can oscillate along the field lines for long periods of time.

In the past several years, substantial effort has been made to understand and explain the phenomena involved in belts of trapped electrons and ions, and to explore possible ways to control and use these phenomena for beneficial purposes. For example, in the late 1950's and early 1960's both the United States and U.S.S.R. detonated a series of nuclear devices of various yields to generate large numbers of charged particles at various altitudes, e.g., 200 kilometers (km) or greater. This was done in order to establish and study artifical belts of trapped electrons and ions. These experiments established that at least some of the extraneous electrons and ions from the detonated devices did become trapped along field lines in the earth's magnetosphere to form artificial belts which were stable for prolonged periods of time. For a discussion of these experiments see "The Radiation Belt and Magnetosphere", W. N. Hess, Blaisdell Publishing Co., 1968, pps. 155 et sec.

Other proposals which have been advanced for altering existing belts of trapped electrons and ions and/or establishing similar artificial belts include injecting charged particles from a satellite carrying a payload of radioactive beta-decay material or alpha emitters; and injecting charged particles from a satellite-borne electron accelerator. Still another approach is described in U.S. Pat. No. 4,042,196 wherein a low energy ionized gas, e.g., hydrogen, is released from a synchronous orbiting satellite near the apex of a radiation belt which is naturally-occurring in the earth's magnetosphere to produce a substantial increase in energetic particle precipitation and, under certain conditions, produce a limit in the number of particles that can be stably trapped. This precipitation effect arises from an enhancement of the whistler-mode and ion-cyclotron mode interactions that result from the ionized gas or "cold plasma" injection.

It has also been proposed to release large clouds of barium in the magnetosphere so that photoionization will increase the cold plasma density, thereby producing electron precipitation through enhanced whistler-mode interactions.

However, in all of the above-mentioned approaches, the mechanisms involved in triggering the change in the trapped particle phenomena must be actually positioned within the affected zone, e.g., the magnetosphere, before they can be actuated to effect the desired change.

The earth's ionosphere is not considered to be a "trapped" belt since there are few trapped particles therein. The term "trapped" herein refers to situations where the force of gravity on the trapped particles is balanced by magnetic forces rather than hydrostatic or collisional forces. The charged electrons and ions in the ionosphere also follow helical paths around magnetic field lines within the ionosphere but are not trapped between mirrors, as in the case of the trapped belts in the magnetosphere, since the gravitational force on the particles is balanced by collisional or hydrostatic forces.

In recent years, a number of experiments have actually been carried out to modify the ionosphere in some controlled manner to investigate the possibility of a beneficial result. For detailed discussions of these operations see the following papers: (1) Ionospheric Modification Theory; G. Meltz and F. W. Perkins; (2) The Platteville High Power Facility; Carrol et al.; (3) Arecibo Heating Experiments; W. E. Gordon and H. C. Carlson, Jr.; and (4) Ionospheric Heating by Powerful Radio Waves; Meltz et al., all published in Radio Science, Vol. 9, No. 11, November, 1974, at pages 885–888; 889–894; 1041–1047; and 1049–1063, respectively, all of which are incorporated herein by reference. In such experiments, certain regions of the ionosphere are heated to change the electron density and temperature within these regions. This is accomplished by transmitting from earth-based antennae high frequency electromagnetic radiation at a substantial angle to, not parallel to, the ionosphere's magnetic field to heat the ionospheric particles primarily by ohmic heating. The electron temperature of the ionosphere has been raised by hundreds of degrees in these experiments, and electrons with several electron volts of energy have been produced in numbers sufficient to enhance airglow. Electron concentrations have been reduced by a few percent, due to expansion of the plasma as a result of increased temperature.

In the Elmo Bumpy Torus (EBT), a controlled fusion device at the Oak Ridge National Laboratory, all heating is provided by microwaves at the electron cyclotron resonance interaction. A ring of hot electrons is formed at the earth's surface in the magnetic mirror by a combination of electron cyclotron resonance and stochastic heating. In the EBT, the ring electrons are produced with an average "temperature" of 250 kilo electron volts or kev (2.5×10^9K) and a plasma beta between 0.1 and 0.4; see, "A Theoretical Study of Electron—Cyclotron Absorption in Elmo Bumpy Torus", Batchelor and Goldfinger, Nuclear Fusion, Vol. 20, No. 4 (1980) pps. 403–418.

Electron cyclotron resonance heating has been used in experiments on the earth's surface to produce and accelerate plasmas in a diverging magnetic field. Kosmahl et al. showed that power was transferred from the electromagnetic waves and that a fully ionized plasma

was accelerated with a divergence angle of roughly 13 degrees. Optimum neutral gas density was 1.7×10^{14} per cubic centimeter; see, "Plasma Acceleration with Microwaves Near Cyclotron Resonance", Kosmahl et al., Journal of Applied Physics, Vol. 38, No. 12, Nov., 1967, pps. 4576–4582.

DISCLOSURE OF THE INVENTION

The present invention provides a method and apparatus for altering at least one selected region which normally exists above the earth's surface. The region is excited by electron cyclotron resonance heating of electrons which are already present and/or artifically created in the region to thereby increase the charged particle energy and ultimately the density of the region.

In one embodiment this is done by transmitting circularly polarized electromagnetic radiation from the earth's surface at or near the location where a naturally-occurring dipole magnetic field (force) line intersects the earth's surface. Right hand circular polarization is used in the northern hemisphere and left hand circular polarization is used in the southern hemisphere. The radiation is deliberately transmitted at the outset in a direction substantially parallel to and along a field line which extends upwardly through the region to be altered. The radiation is transmitted at a frequency which is based on the gyrofrequency of the charged particles and which, when applied to the at least one region, excites electron cyclotron resonance within the region or regions to heat and accelerate the charged particles in their respective helical paths around and along the field line. Sufficient energy is employed to cause ionization of neutral particles (molecules of oxygen, nitrogen and the like, particulates, etc.) which then become a part of the region thereby increasing the charged particle density of the region. This effect can further be enhanced by providing artificial particles, e.g., electrons, ions, etc., directly into the region to be affected from a rocket, satellite, or the like to supplement the particles in the naturally-occurring plasma. These artificial particles are also ionized by the transmitted electromagnetic radiation thereby increasing charged particle density of the resulting plasma in the region.

In another embodiment of the invention, electron cyclotron resonance heating is carried out in the selected region or regions at sufficient power levels to allow a plasma present in the region to generate a mirror force which forces the charged electrons of the altered plasma upward along the force line to an altitude which is higher than the original altitude. In this case the relevant mirror points are at the base of the altered region or regions. The charged electrons drag ions with them as well as other particles that may be present. Sufficient power, e.g., 10^{15} joules, can be applied so that the altered plasma can be trapped on the field line between mirror points and will oscillate in space for prolonged periods of time. By this embodiment, a plume of altered plasma can be established at selected locations for communication modification or other purposes.

In another embodiment, this invention is used to alter at least one selected region of plasma in the ionosphere to establish a defined layer of plasma having an increased charged particle density. Once this layer is established, and while maintaining the transmission of the main beam of circularly polarized electromagnetic radiation, the main beam is modulated and/or at least one second different, modulated electromagnetic radiation beam is transmitted from at least one separate source at a different frequency which will be absorbed in the plasma layer. The amplitude of the frequency of the main beam and/or the second beam or beams is modulated in resonance with at least one known oscillation mode in the selected region or regions to excite the known oscillation mode to propagate a known frequency wave or waves throughout the ionosphere.

BRIEF DESCRIPTION OF THE DRAWINGS

The actual construction, operation, and apparent advantages of this invention will be better understood by referring to the drawings in which like numerals identify like parts and in which:

FIG. **1** is a simplified schematical view of the earth (not to scale) with a magnetic field (force) line along which the present invention is carried out;

FIG. **2** is one embodiment within the present invention in which a selected region of plasma is raised to a higher altitude;

FIG. **3** is a simplified, idealized representation of a physical phenomenon involved in the present invention; and

FIG. **4** is a schematic view of another embodiment within the present invention.

FIG. **5** is a schematic view of an apparatus embodiment within this invention.

BEST MODES FOR CARRYING OUT THE INVENTION

The earth's magnetic field is somewhat analogous to a dipole bar magnet. As such, the earth's magnetic field contains numerous divergent field or force lines, each line intersecting the earth's surface at points on opposite sides of the Equator. The field lines which intersect the earth's surface near the poles have apexes which lie at the furthest points in the earth's magnetosphere while those closest to the Equator have apexes which reach only the lower portion of the magnetosphere.

At various altitudes above the earth's surface, e.g., in both the ionosphere and the magnetosphere, plasma is naturally present along these field lines. This plasma consists of equal numbers of positively and negatively charged particles (i.e., electrons and ions) which are guided by the field line. It is well established that a charged particle in a magnetic field gyrates about field lines, the center of gyration at any instance being called the "guiding center" of the particle. As the gyrating particle moves along a field line in a uniform field, it will follow a helical path about its guiding center, hence linear motion, and will remain on the field line. Electrons and ions both follow helical paths around a field line but rotate in opposite directions. The frequencies at which the electrons and ions rotate about the field line are called gyromagnetic frequencies or cyclotron frequencies because they are identical with the expression for the angular frequencies of gyration of particles in a cyclotron. The cyclotron frequency of ions in a given magnetic field is less than that of electrons, in inverse proportion to their masses.

If the particles which form the plasma along the earth's field lines continued to move with a constant pitch angle, often designated "alpha", they would soon impact on the earth's surface. Pitch angle alpha is defined as the angle between the direction of the earth's magnetic field and the velocity (V) of the particle. However, in converging force fields, the pitch angle does change in such a way as to allow the particle to

turn around and avoid impact. Consider a particle moving along a field line down toward the earth. It moves into a region of increasing magnetic field strength and therefore sine alpha increases. But sine alpha can only increase to 1.0, at which point, the particle turns around and starts moving up along the field line, and alpha decreases. The point at which the particle turns around is called the mirror point, and there alpha equals ninety degrees. This process is repeated at the other end of the field line where the same magnetic field strength value B, namely Bm, exists. The particle again turns around and this is called the "conjugate point" of the original mirror point. The particle is therefore trapped and bounces between the two magnetic mirrors. The particle can continue oscillating in space in this manner for long periods of time. The actual place where a particle will mirror can be calculated from the following:

$$\sin^2 \text{alpha}_o = B_o/B_m \tag{1}$$

wherein:

alpha$_o$=equatorial pitch angle of particle

B_o=equatorial field strength on a particular field line

B_m=field strength at the mirror point

Recent discoveries have established that there are substantial regions of naturally trapped particles in space which are commonly called "trapped radiation belts". These belts occur at altitudes greater than about 500 km and accordingly lie in the magnetosphere and mostly above the ionosphere.

The ionosphere, while it may overlap some of the trapped-particle belts, is a region in which hydrostatic forces govern its particle distribution in the gravitational field. Particle motion within the ionosphere is governed by both hydrodynamic and electrodynamic forces. While there are few trapped particles in the ionosphere, nevertheless, plasma is present along field lines in the ionosphere. The charged particles which form this plasma move between collisions with other particles along similar helical paths around the field lines and although a particular particle may diffuse downward into the earth's lower atmosphere or lose energy and diverge from its original field line due to collisions with other particles, these charged particles are normally replaced by other available charged particles or by particles that are ionized by collision with said particle. The electron density (N_e) of the plasma will vary with the actual conditions and locations involved. Also, neutral particles, ions, and electrons are present in proximity to the field lines.

The production of enhanced ionization will also alter the distribution of atomic and molecular constituents of the atmosphere, most notably through increased atomic nitrogen concentration. The upper atmosphere is normally rich in atomic oxygen (the dominant atmospheric constituent above 200 km altitude), but atomic nitrogen is normally relatively rare. This can be expected to manifest itself in increased airglow, among other effects.

As known in plasma physics, the characteristics of a plasma can be altered by adding energy to the charged particles or by ionizing or exciting additional particles to increase the density of the plasma. One way to do this is by heating the plasma which can be accomplished in different ways, e.g., ohmic, magnetic compression, shock waves, magnetic pumping, electron cyclotron resonance, and the like.

Since electron cyclotron resonance heating is involved in the present invention, a brief discussion of same is in order. Increasing the energy of electrons in a plasma by invoking electron cyclotron resonance heating, is based on a principle similar to that utilized to accelerate charged particles in a cyclotron. If a plasma is confined by a static axial magnetic field of strength B, the charged particles will gyrate about the lines of force with a frequency given, in hertz, as $f_g = 1.54 \times 10^3 B/A$, where: B=magnetic field strength in gauss, and A=mass number of the ion.

Suppose a time-varying field of this frequency is superimposed on the static field B confining the plasma, by passage of a radiofrequency current through a coil which is concentric with that producing the axial field, then in each half-cycle of their rotation about the field lines, the charged particles acquire energy from the oscillating electric field associated with the radio frequency. For example, if B is 10,000 gauss, the frequency of the field which is in resonance with protons in a plasma is 15.4 megahertz.

As applied to electrons, electron cyclotron resonance heating requires an oscillating field having a definite frequency determined by the strength of the confining field. The radio-frequency radiation produces time-varying fields (electric and magnetic), and the electric field accelerates the charged particle. The energized electrons share their energy with ions and neutrals by undergoing collisions with these particles, thereby effectively raising the temperature of the electrons, ions, and neutrals. The apportionment of energy among these species is determined by collision frequencies. For a more detailed understanding of the physics involved, see "Controlled Thermonuclear Reactions", Glasstone and Lovberg, D. Van Nostrand Company, Inc., Princeton, N.J., 1960 and "The Radiation Belt and Magnetosphere", Hess, Blaisdell Publishing Company, 1968, both of which are incorporated herein by reference.

Referring now to the drawings, the present invention provides a method and apparatus for altering at least one region of plasma which lies along a field line, particularly when it passes through the ionosphere and/or magnetosphere. FIG. **1** is a simplified illustration of the earth **10** and one of its dipole magnetic force or field lines **11**. As will be understood, line **11** may be any one of the numerous naturally existing field lines and the actual geographical locations **13** and **14** of line **11** will be chosen based on a particular operation to be carried out. The actual locations at which field lines intersect the earth's surface is documented and is readily ascertainable by those skilled in the art.

Line **11** passes through region R which lies at an altitude above the earth's surface. A wide range of altitudes are useful given the power that can be employed by the practice of this invention. The electron cyclotron resonance heating effect can be made to act on electrons anywhere above the surface of the earth. These electrons may be already present in the atmosphere, ionosphere, and/or magnetosphere of the earth, or can be artificially generated by a variety of means such as x-ray beams, charged particle beams, lasers, the plasma sheath surrounding an object such as a missile or meteor, and the like. Further, artificial particles, e.g., electrons, ions, etc., can be injected directly into region R from an earth-launched rocket or orbiting satellite carrying, for example, a payload of radioactive beta-decay material; alpha emitters; an electron accelerator; and/or ionized gases such as hydrogen; see U.S. Pat. No. 4,042,196. The altitude can be greater than about 50 km if desired,

e.g., can be from about 50 km to about 800 km, and, accordingly may lie in either the ionosphere or the magnetosphere or both. As explained above, plasma will be present along line **11** within region R and is represented by the helical line **12**. Plasma **12** is comprised of charged particles (i.e., electrons and ions) which rotate about opposing helical paths along line **11**.

Antenna **15** is positioned as close as is practical to the location **14** where line **11** intersects the earth's surface. Antenna **15** may be of any known construction for high directionality, for example, a phased array, beam spread angle (θ) type. See "The MST Radar at Poker Flat, Alaska", Radio Science, Vol. 15, No. 2, Mar.-Apr. 1980, pps. 213–223, which is incorporated herein by reference. Antenna **15** is coupled to transmitter **16** which generates a beam of high frequency electromagnetic radiation at a wide range of discrete frequencies, e.g., from about 20 to about 1800 kilohertz (kHz).

Transmitter **16** is powered by power generator means **17** which is preferably comprised of one or more large, commercial electrical generators. Some embodiments of the present invention require large amounts of power, e.g., up to 10^9 to 10^{11} watts, in continuous wave or pulsed power. Generation of the needed power is within the state of the art. Although the electrical generators necessary for the practice of the invention can be powered in any known manner, for example, by nuclear reactors, hydroelectric facilities, hydrocarbon fuels, and the like, this invention, because of its very large power requirement in certain applications, is particularly adapted for use with certain types of fuel sources which naturally occur at strategic geographical locations around the earth. For example, large reserves of hydrocarbons (oil and natural gas) exist in Alaska and Canada. In northern Alaska, particularly the North Slope region, large reserves are currently readily available. Alaska and northern Canada also are ideally located geographically as to magnetic latitudes. Alaska provides easy access to magnetic field lines that are especially suited to the practice of this invention, since many field lines which extend to desirable altitudes for this invention intersect the earth in Alaska. Thus, in Alaska, there is a unique combination of large, accessible fuel sources at desirable field line intersections. Further, a particularly desirable fuel source for the generation of very large amounts of electricity is present in Alaska in abundance, this source being natural gas. The presence of very large amounts of clean-burning natural gas in Alaskan latitudes, particularly on the North Slope, and the availability of magnetohydrodynamic (MHD), gas turbine, fuel cell, electrogasdynamic (EGD) electric generators which operate very efficiently with natural gas provide an ideal power source for the unprecedented power requirements of certain of the applications of this invention. For a more detailed discussion of the various means for generating electricity from hydrocarbon fuels, see "Electrical Aspects of Combustion", Lawton and Weinberg, Clarendon Press, 1969. For example, it is possible to generate the electricity directly at the high frequency needed to drive the antenna system. To do this, typically the velocity of flow of the combustion gases (v), past magnetic field perturbation of dimension d (in the case of MHD), follow the rule:

$$v=df$$

where f is the frequency at which electricity is generated. Thus, if $v=1.78\times10^6$ cm/sec and d=1 cm then electricity would be generated at a frequency of 1.78 mHz.

Put another way, in Alaska, the right type of fuel (natural gas) is naturally present in large amounts and at just the right magnetic latitudes for the most efficient practice of this invention, a truly unique combination of circumstances. Desirable magnetic latitudes for the practice of this invention interest the earth's surface both northerly and southerly of the equator, particularly desirable latitudes being those, both northerly and southerly, which correspond in magnitude with the magnetic latitudes that encompass Alaska.

Referring now to FIG. 2 a first ambodiment is illustrated where a selected region R_1 of plasma **12** is altered by electron cyclotron resonance heating to accelerate the electrons of plasma **12**, which are following helical paths along field line **11**.

To accomplish this result, electromagnetic radiation is transmitted at the outset, essentially parallel to line **11** via antenna **15** as right hand circularly polarized radiation wave **20**. Wave **20** has a frequency which will excite electron cyclotron resonance with plasma **12** at its initial or original altitude. This frequency will vary depending on the electron cyclotron resonance of region R_1 which, in turn, can be determined from available data based on the altitudes of region R_1, the particular field line **11** being used, the strength of the earth's magnetic field, etc. Frequencies of from about 20 to about 7200 kHz, preferably from about 20 to about 1800 kHz can be employed. Also, for any given application, there will be a threshhold (minimum power level) which is needed to produce the desired result. The minimum power level is a function of the level of plasma production and movement required, taking into consideration any loss processes that may be dominant in a particular plasma or propagation path.

As electron cyclotron resonance is established in plasma **12**, energy is transferred from the electromagnetic radiation **20** into plasma **12** to heat and accelerate the electrons therein and, subsequently, ions and neutral particles. As this process continues, neutral particles which are present within R_1 are ionized and absorbed into plasma **12** and this increases the electron and ion densities of plasma **12**. As the electron energy is raised to values of about 1 kilo electron volt (kev), the generated mirror force (explained below) will direct the excited plasma **12** upward along line **11** to form a plume R_2 at an altitude higher than that of R_1.

Plasma acceleration results from the force on an electron produced by a nonuniform static magnetic field ($\overline{B}$). The force, called the mirror force, is given by

$$F=-\mu\nabla B \qquad (2)$$

where μ is the electron magnetic moment and $\nabla\overline{B}$ is the gradient of the magnetic field, μ being further defined as:

$$W_\perp/B=mV_\perp^2/2B$$

where $W_\perp$ is the kinetic energy in the direction perpendicular to that of the magnetic field lines and B is the magnetic field strength at the line of force on which the guiding center of the particle is located. The force as represented by equation (2) is the force which is responsible for a particle obeying equation (1).

Since the magnetic field is divergent in region R_1, it can be shown that the plasma will move upwardly from the heating region as shown in FIG. **1** and further it can be shown that

$$\tfrac{1}{2}M_eV_{e\perp}^2(x) \approx \tfrac{1}{2}M_eV_{e\perp}^2(Y) + \tfrac{1}{2}M_iV_{i\parallel}^2(Y) \tag{3}$$

where the left hand side is the initial electron transverse kinetic energy; the first term on the right is the transverse electron kinetic energy at some point (Y) in the expanded field region, while the final term is the ion kinetic energy parallel to B at point (Y). This last term is what constitutes the desired ion flow. It is produced by an electrostatic field set up by electrons which are accelerated according to Equation (2) in the divergent field region and pulls ions along with them. Equation (3) ignores electron kinetic energy parallel to B because $V_{e\parallel} \approx V_{i\parallel}$, so the bulk of parallel kinetic energy resides in the ions because of their greater masses. For example, if an electromagnetic energy flux of from about 1 to about 10 watts per square centimeter is applied to region R, whose altitude is 115 km, a plasma having a density (N_e) of 10^{12} per cubic centimeter will be generated and moved upward to region R_2 which has an altitude of about 1000 km. The movement of electrons in the plasma is due to the mirror force while the ions are moved by ambipolar diffusion (which results from the electrostatic field). This effectively "lifts" a layer of plasma **12** from the ionosphere and/or magnetosphere to a higher elevation R_2. The total energy required to create a plasma with a base area of 3 square kilometers and a height of 1000 km is about 3×10^{13} joules.

FIG. **3** is an idealized representation of movement of plasma **12** upon excitation by electron cyclotron resonance within the earth's divergent force field. Electrons (e) are accelerated to velocities required to generate the necessary mirror force to cause their upward movement. At the same time neutral particles (n) which are present along line **11** in region $R_1$are ionized and become part of plasma **12**. As electrons (e) move upward along line **11**, they drag ions (i) and neutrals (n) with them but at an angle θ of about 13 degrees to field line **11**. Also, any particulates that may be present in region R_1, will be swept upwardly with the plasma. As the charged particles of plasma **12** move upward, other particles such as neutrals within or below R_1, move in to replace the upwardly moving particles. These neutrals, under some conditions, can drag with them charged particles.

For example, as a plasma moves upward, other particles at the same altitude as the plasma move horizontally into the region to replace the rising plasma and to form new plasma. The kinetic energy developed by said other particles as they move horizontally is, for example, on the same order of magnitude as the total zonal kinetic energy of stratospheric winds known to exist.

Referring again to FIG. **2**, plasma **12** in region R_1 is moved upward along field line **11**. The plasma **12** will then form a plume (cross-hatched area in FIG. **2**) which will be relatively stable for prolonged periods of time. The exact period of time will vary widely and be determined by gravitational forces and a combination of radiative and diffusive loss terms. In the previous detailed example, the calculations were based on forming a plume by producing 0+energies of 2 ev/particle. About 10 ev per particle would be required to expand plasma **12** to apex point C (FIG. **1**). There at least some of the particles of plasma **12** will be trapped and will oscillate between mirror points along field line **11**. This oscillation will then allow additional heating of the trapped plasma **12** by stochastic heating which is associated with trapped and oscillating particles. See "A New Mechanism for Accelerating Electrons in the Outer Ionosphere" by R. A. Helliwell and T. F. Bell, Journal of Geophysical Research, Vol. 65, No. 6, June, 1960. This is preferably carried out at an altitude of at least 500 km.

The plasma of the typical example might be employed to modify or disrupt microwave transmissions of satellites. If less than total black-out of transmission is desired (e.g., scrambling by phase shifting digital signals), the density of the plasma (N_e) need only be at least about 10^6 per cubic centimeter for a plasma orginating at an altitude of from about 250 to about 400 km and accordingly less energy (i.e., electromagnetic radiation), e.g., 10^8 joules need be provided. Likewise, if the density $N_e$is on the order of 10^8, a properly positioned plume will provide a reflecting surface for VHF waves and can be used to enhance, interfere with, or otherwise modify communication transmissions. It can be seen from the foregoing that by appropriate application of various aspects of this invention at strategic locations and with adequate power sources, a means and method is provided to cause interference with or even total disruption of communications over a very large portion of the earth. This invention could be employed to disrupt not only land based communications, both civilian and military, but also airborne communications and sea communications (both surface and subsurface). This would have significant military implications, particularly as a barrier to or confusing factor for hostile missiles or airplanes. The belt or belts of enhanced ionization produced by the method and apparatus of this invention, particularly if set up over Northern Alaska and Canada, could be employed as an early warning device, as well as a communications disruption medium. Further, the simple ability to produce such a situation in a practical time period can by itself be a deterring force to hostile action. The ideal combination of suitable field lines intersecting the earth's surface at the point where substantial fuel sources are available for generation of very large quantitities of electromagnetic power, such as the North Slope of Alaska, provides the wherewithal to accomplish the foregoing in a practical time period, e.g., strategic requirements could necessitate achieving the desired altered regions in time periods of two minutes or less and this is achievable with this invention, especially when the combination of natural gas and magnetohydrodynamic, gas turbine, fuel cell and/or EGD electric generators are employed at the point where the useful field lines intersect the earth's surface. One feature of this invention which satisfies a basic requirement of a weapon system, i.e., continuous checking of operability, is that small amounts of power can be generated for operability checking purposes. Further, in the exploitation of this invention, since the main electromagnetic beam which generates the enhanced ionized belt of this invention can be modulated itself and/or one or more additional electromagnetic radiation waves can be impinged on the ionized region formed by this invention as will be described in greater detail herein after with respect to FIG. **4**, a substantial amount of randomly modulated signals of very large power magnitude can be generated in a highly nonlinear mode. This can cause confusion of or interference with or even complete disruption of guidance systems employed by

even the most sophisticated of airplanes and missiles. The ability to employ and transmit over very wide areas of the earth a plurality of electromagnetic waves of varying frequencies and to change same at will in a random manner, provides a unique ability to interfere with all modes of communications, land, sea, and/or air, at the same time. Because of the unique juxtaposition of usable fuel source at the point where desirable field lines intersect the earth's surface, such wide ranging and complete communication interference can be achieved in a resonably short period of time. Because of the mirroring phenomenon discussed hereinabove, it can also be prolonged for substantial time periods so that it would not be a mere transient effect that could simply be waited out by an opposing force. Thus, this invention provides the ability to put unprecedented amounts of power in the earth's atmosphere at strategic locations and to maintain the power injection level, particularly if random pulsing is employed, in a manner far more precise and better controlled than heretofore accomplished by the prior art, particularly by the detonation of nuclear devices of various yeilds at various altitudes. Where the prior art approaches yielded merely transitory effects, the unique combination of fuel and desirable field lines at the point where the fuel occurs allows the establishment of, compared to prior art approaches, precisely controlled and long-lasting effects which cannot, practically speaking, simply be waited out. Further, by knowing the frequencies of the various electromagnetic beams employed in the practice of this invention, it is possible not only to interfere with third party communications but to take advantage of one or more such beams to carry out a communications network even though the rest of the world's communications are disrupted. Put another way, what is used to disrupt another's communications can be employed by one knowledgeable of this invention as a communications network at the same time. In addition, once one's own communication network is established, the far-reaching extent of the effects of this invention could be employed to pick up communication signals of other for intelligence purposes. Thus, it can be seen that the disrupting effects achievable by this invention can be employed to benefit by the party who is practicing this invention since knowledge of the various electromagnetic waves being employed and how they will vary in frequency and magnitude can be used to an advantage for positive communication and eavesdropping purposes at the same time. However, this invention is not limited to locations where the fuel source naturally exists or where desirable field lines naturally intersect the earth's surface. For example, fuel, particularly hydrocarbon fuel, can be transported by pipeline and the like to the location where the invention is to be practiced.

FIG. 4 illustrates another embodiment wherein a selected region of plasma R_3which lies within the earth's ionosphere is altered to increase the density thereof whereby a relatively stable layer **30** of relatively dense plasma is maintained within region R_3. Electromagnetic radiation is transmitted at the outset essentially parallel to field line **11** via antenna **15** as a right hand circularly polarized wave and at a frequency (e.g., 1.78 megahertz when the magnetic field at the desired altitude is 0.66 gauss) capable of exciting electron cyclotron resonance in plasma **12** at the particular altitude of plasma **12**. This causes heating of the particles (electrons, ions, neutrals, and particulates) and ionization of the uncharged particles adjacent line **11**, all of which are absorbed into plasma **12** to increase the density thereof. The power transmitted, e.g., 2×10^6 watts for up to 2 minutes heating time, is less than that required to generate the mirror force F required to move plasma **12** upward as in the previous embodiment.

While continuing to transmit electromagnetic radiation **20** from antenna **15**, a second electromagnetic radiation beam **31**, which is at a defined frequency different from the radiation from antenna **15**, is transmitted from one or more second sources via antenna **32** into layer **30** and is absorbed into a portion of layer **30** (cross-hatched area in FIG. 4). The electromagnetic radiation wave from antenna **32** is amplitude modulated to match a known mode of oscillation f_3 in layer **30**. This creates a resonance in layer **30** which excites a new plasma wave **33** which also has a frequency of f_3 and which then propogates through the ionosphere. Wave **33** can be used to improve or disrupt communications or both depending on what is desired in a particular application. Of course, more than one new wave **33** can be generated and the various new waves can be modulated at will and in a highly nonlinear fashion.

FIG. **5** shows apparatus useful in this invention, particularly when those applications of this invention are employed which require extremely large amounts of power. In FIG. **5** there is shown the earth's surface **40** with a well **41** extending downwardly thereinto until it penetrates hydrocarbon producing reservoir **42**. Hydrocarbon reservoir **42** produces natural gas alone or in combination with crude oil. Hydrocarbons are produced from reservoir **42** through well **41** and wellhead **43** to a treating system **44** by way of pipe **45**. In treater **44**, desirable liquids such as crude oil and gas condensates are separated and recovered by way of pipe **46** while undesirable gases and liquids such as water, H_2S, and the like are separated by way of pipe **47**. Desirable gases such as carbon dioxide are separated by way of pipe **48**, and the remaining natural gas stream is removed from treater **44** by way of pipe **49** for storage in conventional tankage means (not shown) for future use and/or use in an electrical generator such as a magnetohydrodynamic, gas turbine, fuel cell or EGD generator **50**. Any desired number and combination of different types of electric generators can be employed in the practice of this invention. The natural gas is burned in generator **50** to produce substantial quantities of electricity which is then stored and/or passed by way of wire **51** to a transmitter **52** which generates the electromagnetic radiation to be used in the method of this invention. The electromagnetic radiation is then passed by way of wire **53** to antenna **54** which is located at or near the end of field line **11**. Antenna **54** sends circularly polarized radiation wave **20** upwards along field line **11** to carry out the various methods of this invention as described hereinabove.

Of course, the fuel source need not be used in its naturally-occurring state but could first be converted to another second energy source form such as hydrogen, hydrazine and the like, and electricity then generated from said second energy source form.

It can be seen from the foregoing that when desirable field line **11** intersects earth's surface **40** at or near a large naturally-occurring hydrocarbon source **42**, exceedingly large amounts of power can be very efficiently produced and transmitted in the direction of field lines. This is particularly so when the fuel source is natural gas and magnetohydrodynamic generators are employed. Further, this can all be accomplished in a

relatively small physical area when there is the unique coincidence of fuel source **42** and desirable field line **11**. Of course, only one set of equipment is shown in FIG. **5** for sake of simplicity. For a large hydrocarbon reservoir **42**, a plurality of wells **41** can be employed to feed one or more storage means and/or treaters and as large a number of generators **55** as needed to power one or more transmitters **52** and one or more antennas **54**. Since all of the apparatus **44** through **54** can be employed and used essentially at the sight where naturally-occurring fuel source **42** is located, all the necessary electromagnetic radiation **20** is generated essentially at the same location as fuel source **42**. This provides for a maximum amount of usable electromagnetic radiation **20** since there are no significant storage or transportation losses to be incurred. In other words, the apparatus is brought to the sight of the fuel source where desirable field line **11** intersects the earth's surface **40** on or near the geographical location of fuel source **42**, fuel source **42** being at a desirable magnetic latitude for the practice of this invention, for example, Alaska.

The generation of electricity by motion of a conducting fluid through a magnetic field, i.e., magnetohydrodynamics (MHD), provides a method of electric power generation without moving mechanical parts and when the conducting fluid is a plasma formed by combustion of a fuel such as natural gas, an idealized combination of apparatus is realized since the very clean-burning natural gas forms the conducting plasma in an efficient manner and the thus formed plasma, when passed through a magnetic field, generates electricity in a very efficient manner. Thus, the use of fuel source **42** to generate a plasma by combustion thereof for the generation of electricity essentially at the site of occurrence of the fuel source is unique and ideal when high power levels are required and desirable field lines **11** intersect the earth's surface **40** at or near the site of fuel source **42**. A particular advantage for MHD generators is that they can be made to generate large amounts of power with a small volume, light weight device. For example, a 1000 megawatt MHD generator can be construed using superconducting magnets to weigh roughly 42,000 pounds and can be readily air lifted.

This invention has a phenomenal variety of possible ramifications and potential future developments. As alluded to earlier, missile or aircraft destruction, deflection, or confusion could result, particularly when relativistic particles are employed. Also, large regions of the atmosphere could be lifted to an unexpectedly high altitude so that missiles encounter unexpected and unplanned drag forces with resultant destruction or deflection of same. Weather modification is possible by, for example, altering upper atmosphere wind patterns or altering solar absorption patterns by constructing one or more plumes of atmospheric particles which will act as a lens or focusing device. Also as alluded to earlier, molecular modifications of the atmosphere can take place so that positive environmental effects can be achieved. Besides actually changing the molecular composition of an atmospheric region, a particular molecule or molecules can be chosen for increased presence. For example, ozone, nitrogen, etc. concentrations in the atmosphere could be artificially increased. Similarly, environmental enhancement could be achieved by causing the breakup of various chemical entities such as carbon dioxide, carbon monoxide, nitrous oxides, and the like. Transportation of entities can also be realized when advantage is taken of the drag effects caused by regions of the atmosphere moving up along diverging field lines. Small micron sized particles can be then transported, and, under certain circumstances and with the availability of sufficient energy, larger particles or objects could be similarly affected. Particles with desired characteristics such as tackiness, reflectivity, absorptivity, etc., can be transported for specific purposes or effects. For example, a plume of tacky particles could be established to increase the drag on a missile or satellite passing therethrough. Even plumes of plasma having substantially less charged particle density than described above will produce drag effects on missiles which will affect a lightweight (dummy) missile in a manner substantially different than a heavy (live) missile and this affect can be used to distinguish between the two types of missiles. A moving plume could also serve as a means for supplying a space station or for focusing vast amount of sunlight on selected portions of the earth. Surveys of global scope could also be realized because the earth's natural magnetic field could be significantly altered in a controlled manner by plasma beta effects resulting in, for example, improved magnetotelluric surveys. Electromagnetic pulse defenses are also possible. The earth's magnetic field could be decreased or disrupted at appropriate altitudes to modify or eliminate the magnetic field in high Compton electron generation (e.g., from high altitude nuclear bursts) regions. High intensity, well controlled electrical fields can be provided in selected locations for various purposes. For example, the plasma sheath surrounding a missile or satellite could be used as a trigger for activating such a high intensity field to destroy the missile or satellite. Further, irregularities can be created in the ionosphere which will interfere with the normal operation of various types of radar, e.g., synthetic aperture radar. The present invention can also be used to create artificial belts of trapped particles which in turn can be studied to determine the stability of such parties. Still further, plumes in accordance with the present invention can be formed to simulate and/or perform the same functions as performed by the detonation of a "heave" type nuclear device without actually having to detonate such a device. Thus it can be seen that the ramifications are numerous, far-reaching, and exceedingly varied in usefulness.

I claim:

1. A method for altering at least one region normally existing above the earth's surface with electromagnetic radiation using naturally-occurring and diverging magnetic field lines of the earth comprising transmitting first electromagnetic radiation at a frequency between 20 and 7200 kHz from the earth's surface, said transmitting being conducted essentially at the outset of transmission substantially parallel to and along at least one of said field lines, adjusting the frequency of said first radiation to a value which will excite electron cyclotron resonance at an initial elevation at least 50 km above the earth's surface, whereby in the region in which said electron cyclotron resonance takes place heating, further ionization, and movement of both charged and neutral particles is effected, said cyclotron resonance excitation of said region is continued until the electron concentration of said region reaches a value of at least 10^6 per cubic centimeter and has an ion energy of at least 2 ev.

2. The method of claim **1** including the step of providing artificial particles in said at least one region which are excited by said electron cyclotron resonance.

3. The method of claim **2** wherein said artificial particles are provided by injecting same into said at least one region from an orbiting satellite.

4. The method of claim **1** wherein said threshold excitation of electron cyclotron resonance is about 1 watt per cubic centimeter and is sufficient to cause movement of a plasma region along said diverging magnetic field lines to an altitude higher than the altitude at which said excitation was initiated.

5. The method of claim **4** wherein said rising plasma region pulls with it a substantial portion of neutral particles of the atmosphere which exist in or near said plasma region.

6. The method of claim **1** wherein there is provided at least one separate source of second electromagnetic radiation, said second radiation having at least one frequency different from said first radiation, impinging said at least one second radiation on said region while said region is undergoing electron cyclotron resonance excitation caused by said first radiation.

7. The method of claim **6** wherein said second radiation has a frequency which is absorbed by said region.

8. The method of claim **6** wherein said region is plasma in the ionosphere and said second radiation excites plasma waves within said ionosphere.

9. The method of claim **8** wherein said electron concentration reaches a value of at least 10^{12} per cubic centimeter.

10. The method of claim **8** wherein said excitation of electron cyclotron resonance is initially carried out within the ionosphere and is continued for a time sufficient to allow said region to rise above said ionosphere.

11. The method of claim **1** wherein said excitation of electron cyclotron resonance is carried out above about 500 kilometers and for a time of from 0.1 to 1200 seconds such that multiple heating of said plasma region is achieved by means of stochastic heating in the magnetosphere.

12. The method of claim **1** wherein said first electromagnetic radiation is right hand circularly polarized in the northern hemisphere and left hand circularly polarized in the southern hemisphere.

13. The method of claim **1** wherein said electromagnetic radiation is generated at the site of a naturally-occurring hydrocarbon fuel source, said fuel source being located in at least one of northerly or southerly magnetic latitudes.

14. The method of claim **13** wherein said fuel source is natural gas and electricity for generating said electromagnetic radiation is obtained by burning said natural gas in at least one of magnetohydrodynamic, gas turbine, fuel cell, and EGD electric generators located at the site where said natural gas naturally occurs in the earth.

15. The method of claim **14** wherein said site of natural gas is within the magnetic latitudes that encompass Alaska.

* * * * *

United States Patent [19]

Koert

[11] **Patent Number: 5,041,834**

[45] **Date of Patent: Aug. 20, 1991**

[54] **ARTIFICIAL IONOSPHERIC MIRROR COMPOSED OF A PLASMA LAYER WHICH CAN BE TILTED**

[75] Inventor: **Peter Koert**, Washington, D.C.

[73] Assignee: **APTI, Inc.**, Washington, D.C.

[21] Appl. No.: **524,435**

[22] Filed: **May 17, 1990**

[51] **Int. Cl.**^{5} **H04B 7/00**; H01Q 3/22

[52] **U.S. Cl.** **342/367**; 342/372

[58] **Field of Search** 342/367, 353, 371, 372; 455/64

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,445,844	5/1969	Grossi et al.	342/367
4,253,190	2/1981	Csonka	455/12
4,686,605	8/1987	Eastlund	361/231
4,712,155	12/1987	Eastlund et al.	361/231
4,817,495	4/1989	Drobot	89/1.11

Primary Examiner—Gregory C. Issing

Attorney, Agent, or Firm—Foley & Lardner

[57] **ABSTRACT**

This invention relates to generation of a Artificial Ionospheric Mirror (AIM), or a plasma layer in the atmosphere. The AIM is used like the ionosphere to reflect RF energy over great distances. A tiltable AIM is created by a heater antenna controlled in phase and frequency. The heater antenna phase shift scans a beam to paint a plasma layer. Frequency is changed to refocus at continually higher altitudes to tilt the plasma layer.

16 Claims, 15 Drawing Sheets

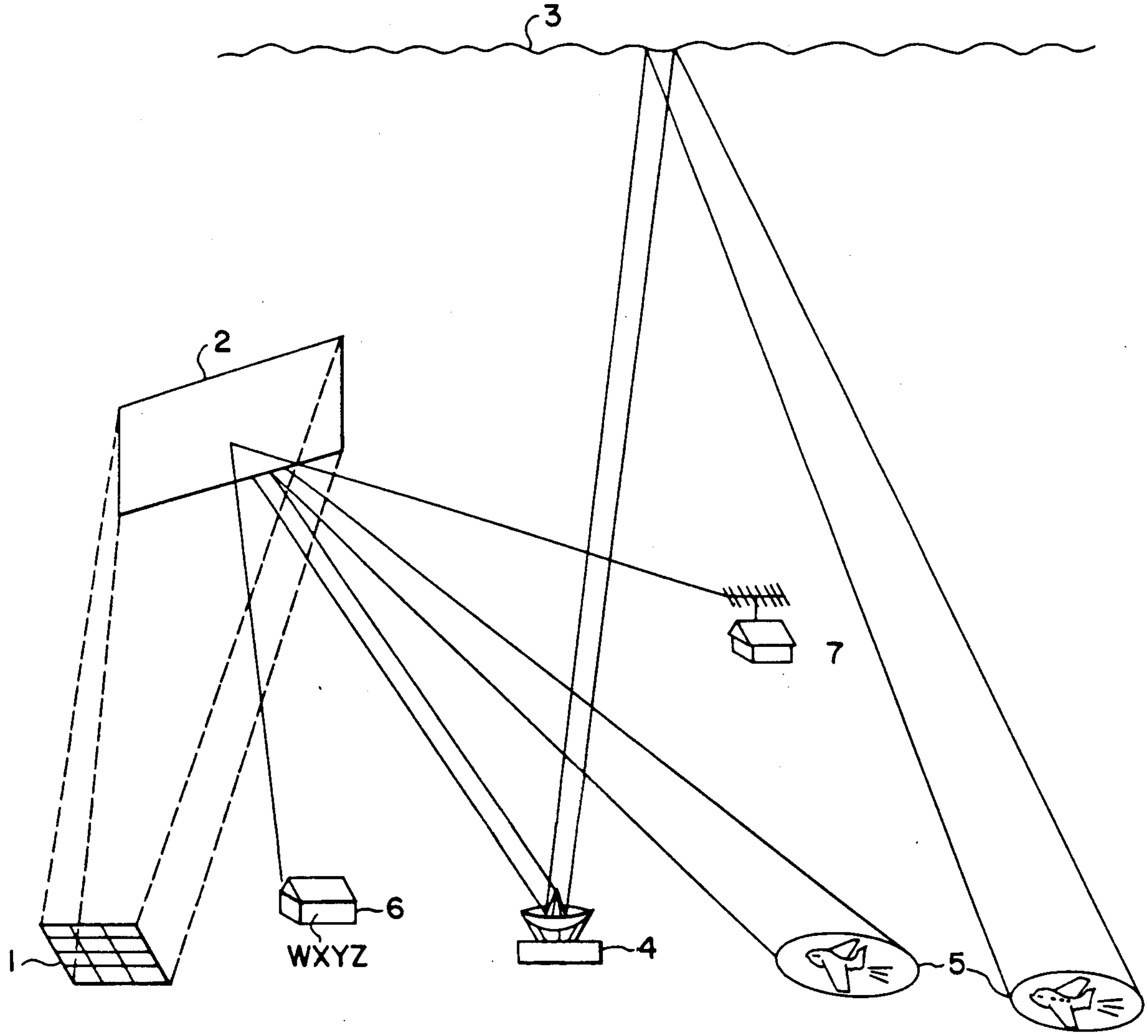

FIG.1

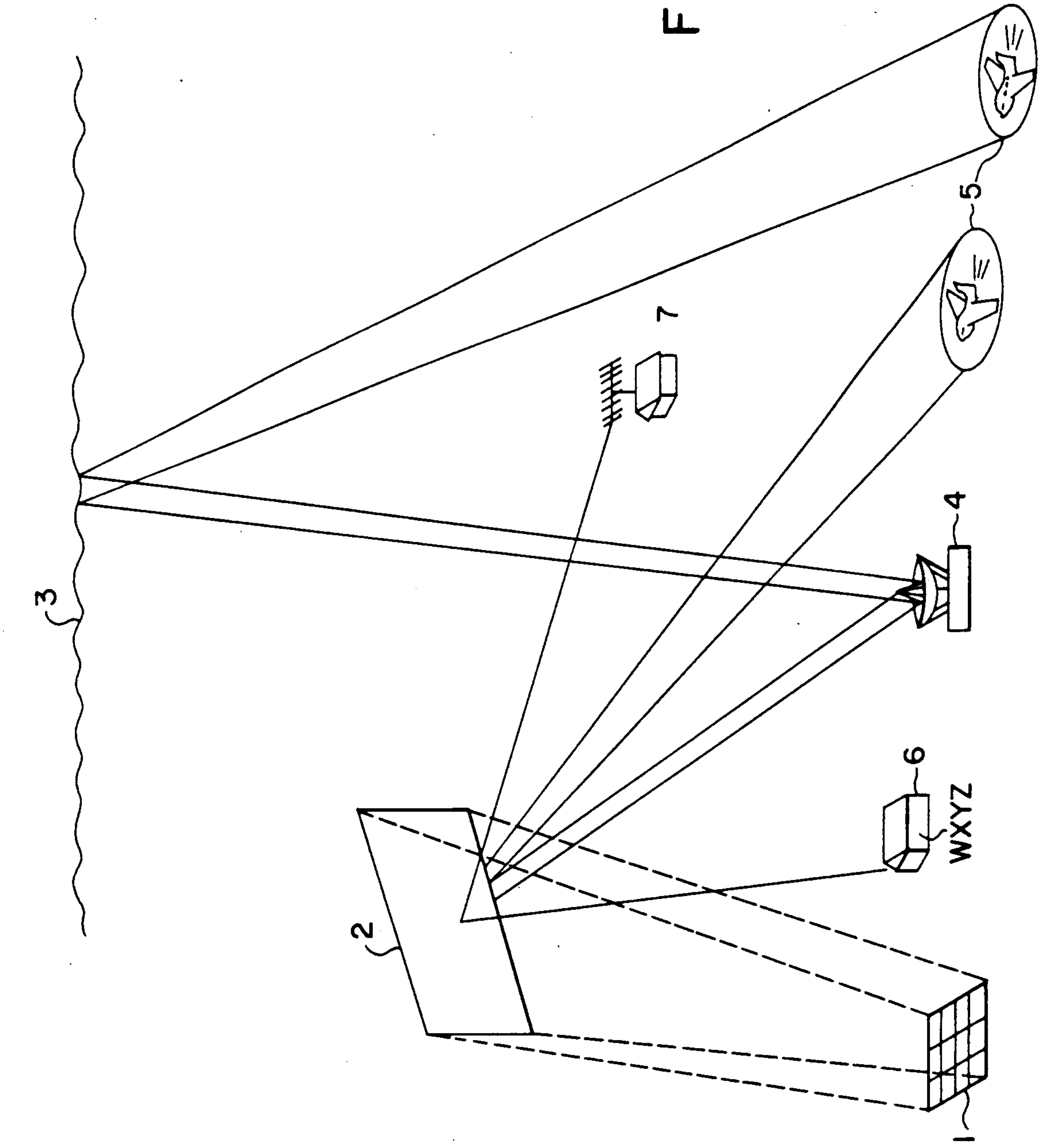
1
2
3
4
5
5
6
7
WXYZ

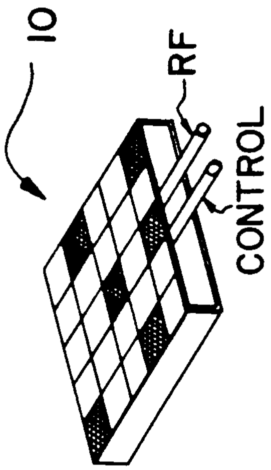
10
RF
CONTROL

FIG. 2b

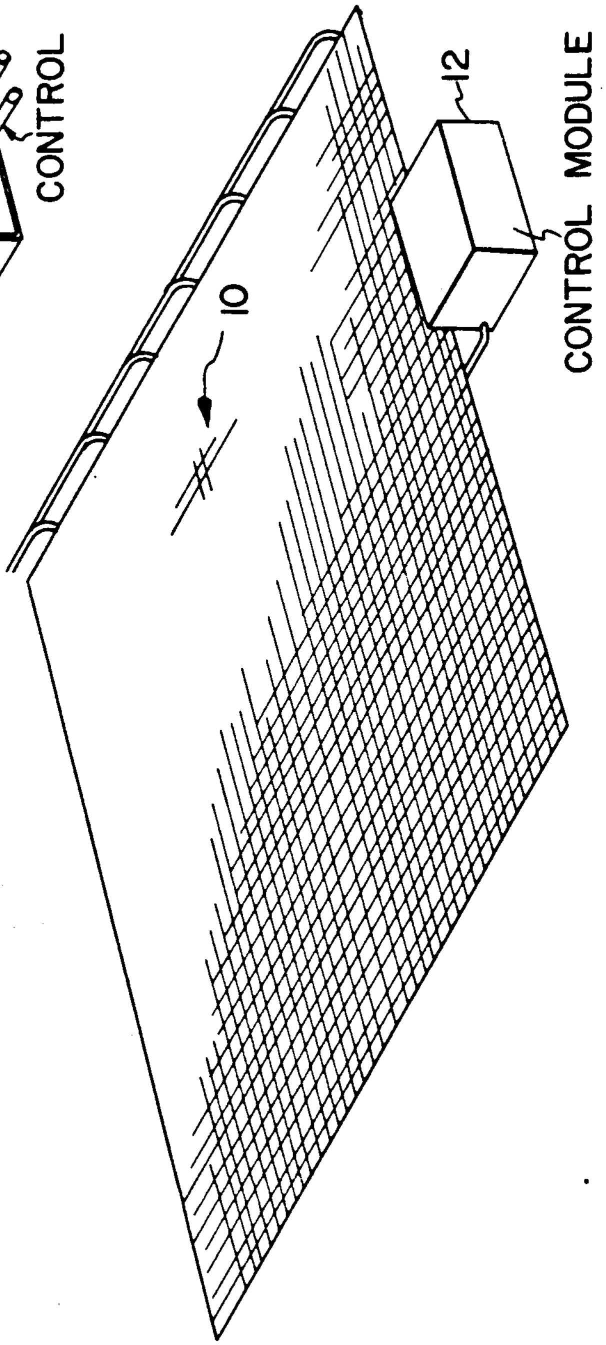
12
CONTROL MODULE
10

FIG. 2a

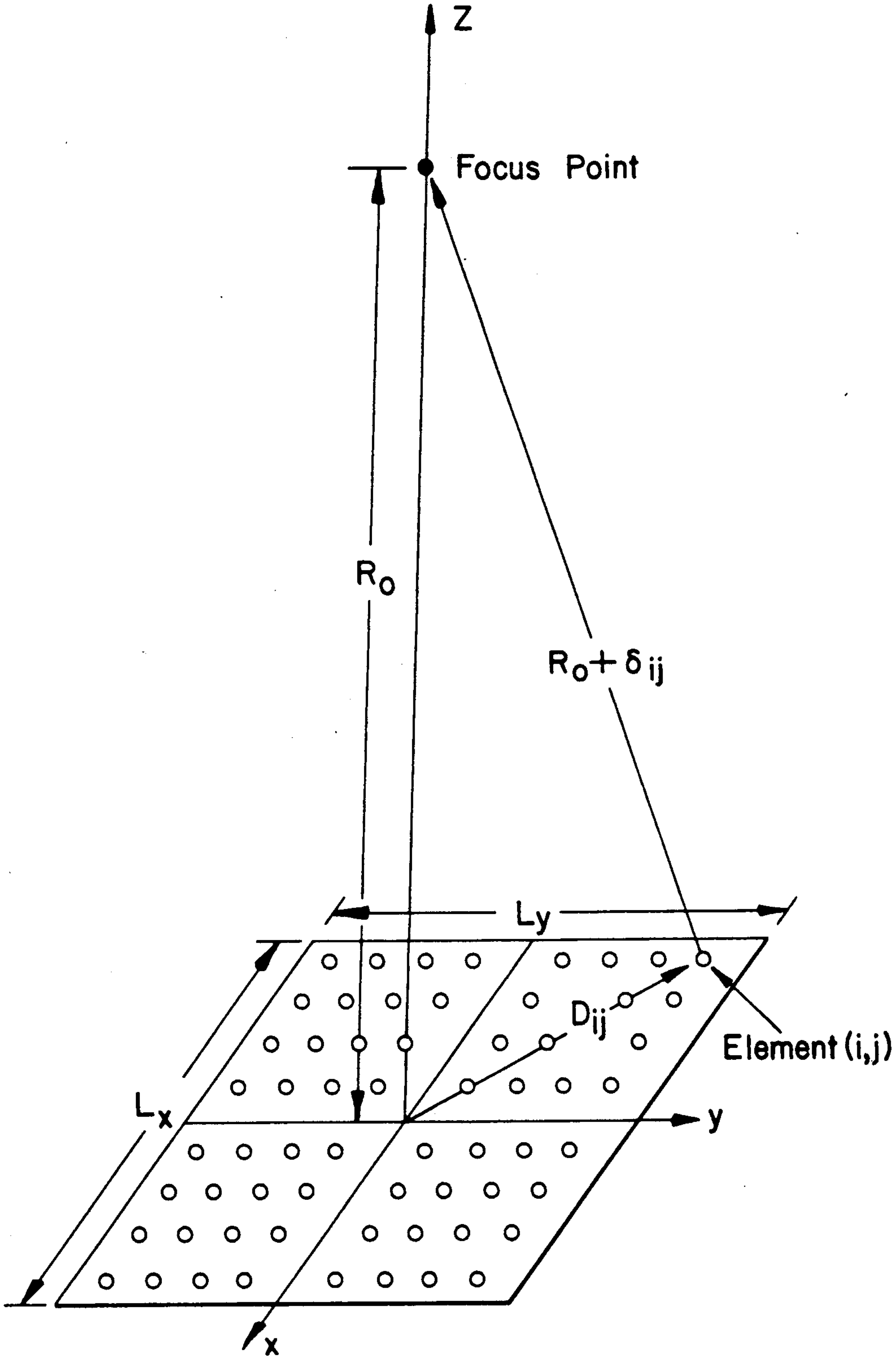
Z
Focus Point
R_o
$R_o+\delta_{ij}$
L_y
L_x
D_{ij}
Element (i,j)
y
x

FIG. 3

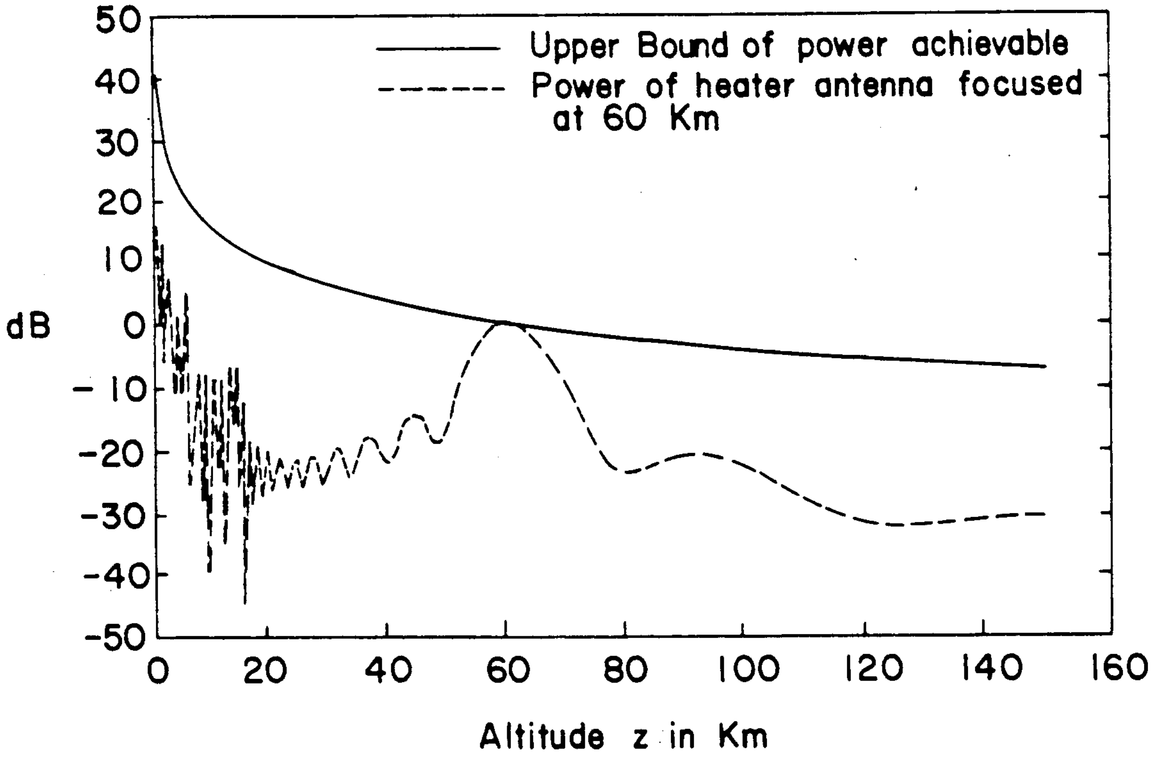
Upper Bound of power achievable
Power of heater antenna focused at 60 Km
dB
50
40
30
20
10
0
- 10
-20
-30
-40
-50
0
20
40
60
80
100
120
140
160
Altitude z in Km

FIG. 4

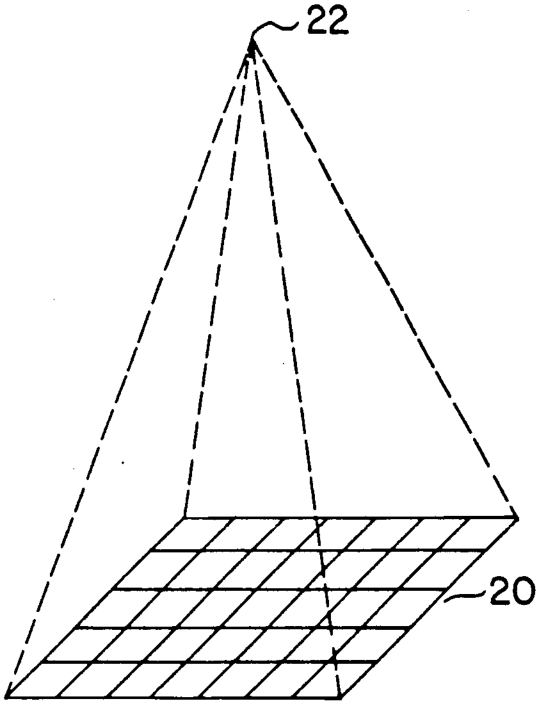

FIG.5

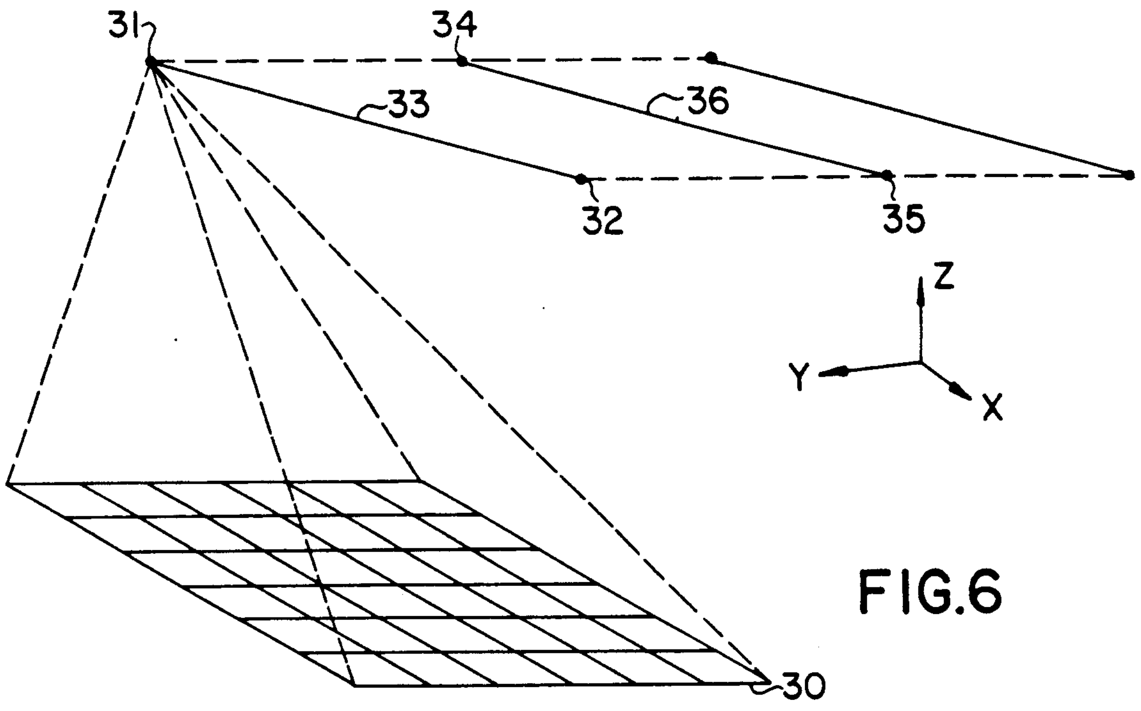
31
34
33
36
32
35
Z
Y
X
30

FIG.6

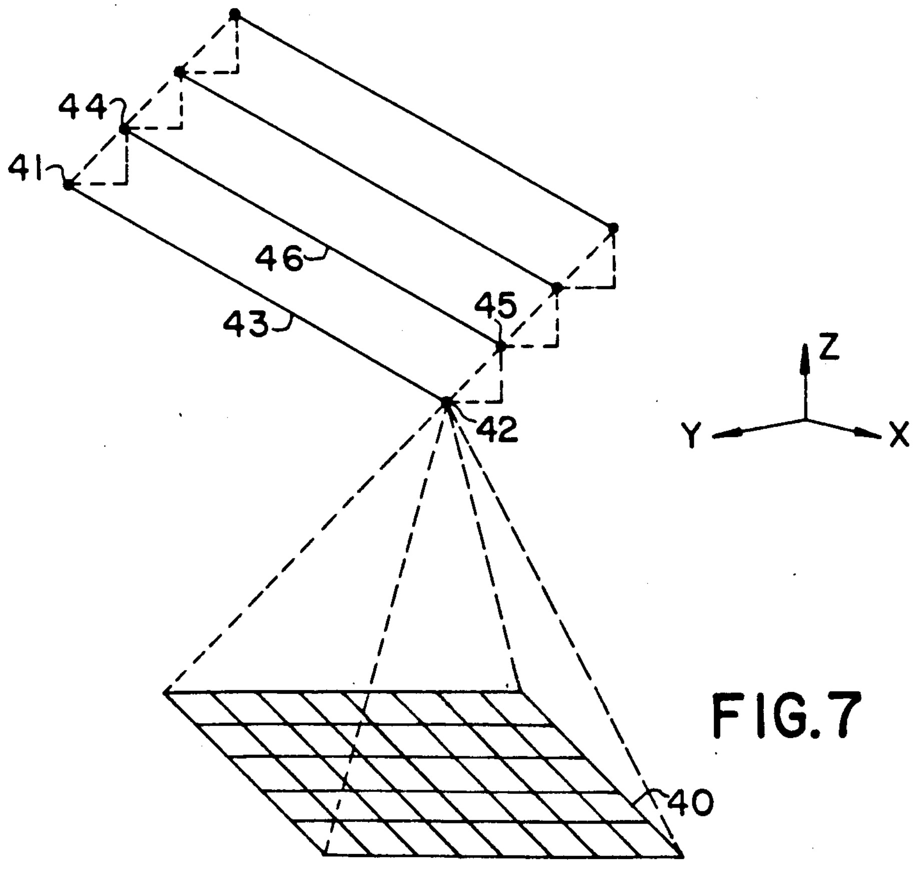
44
41
46
43
45
42
Z
Y
X
40

FIG.7

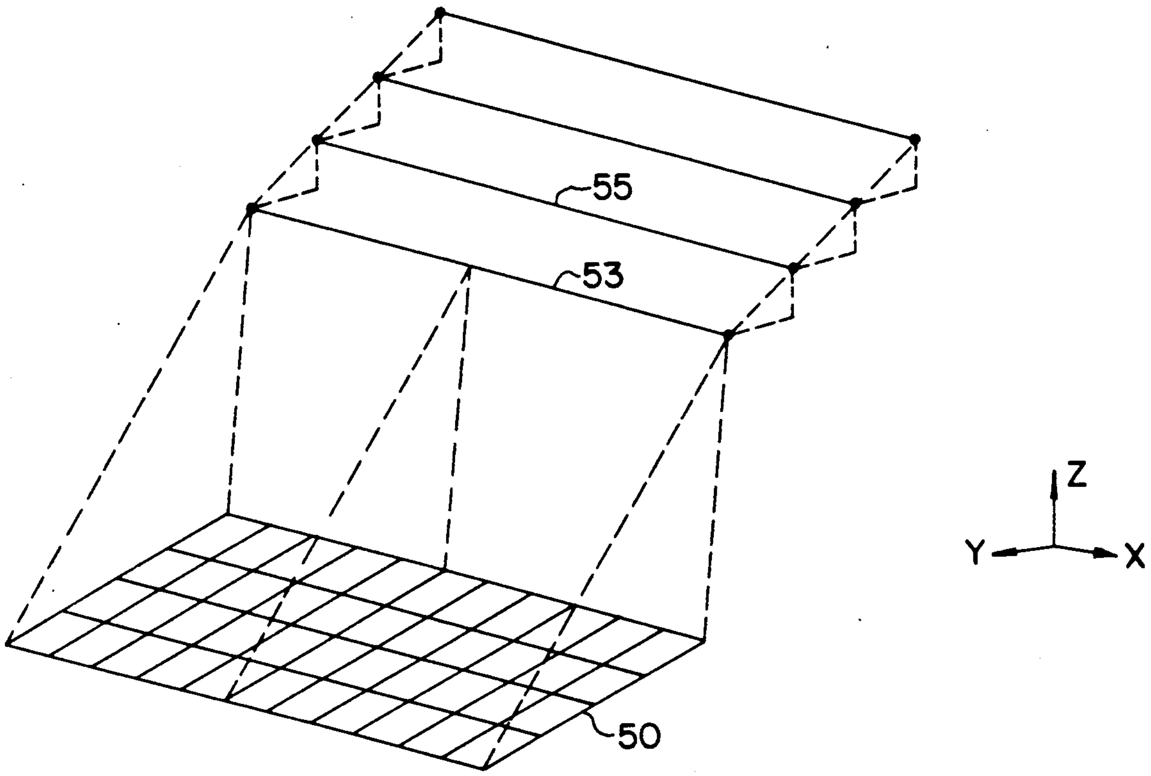

FIG.8

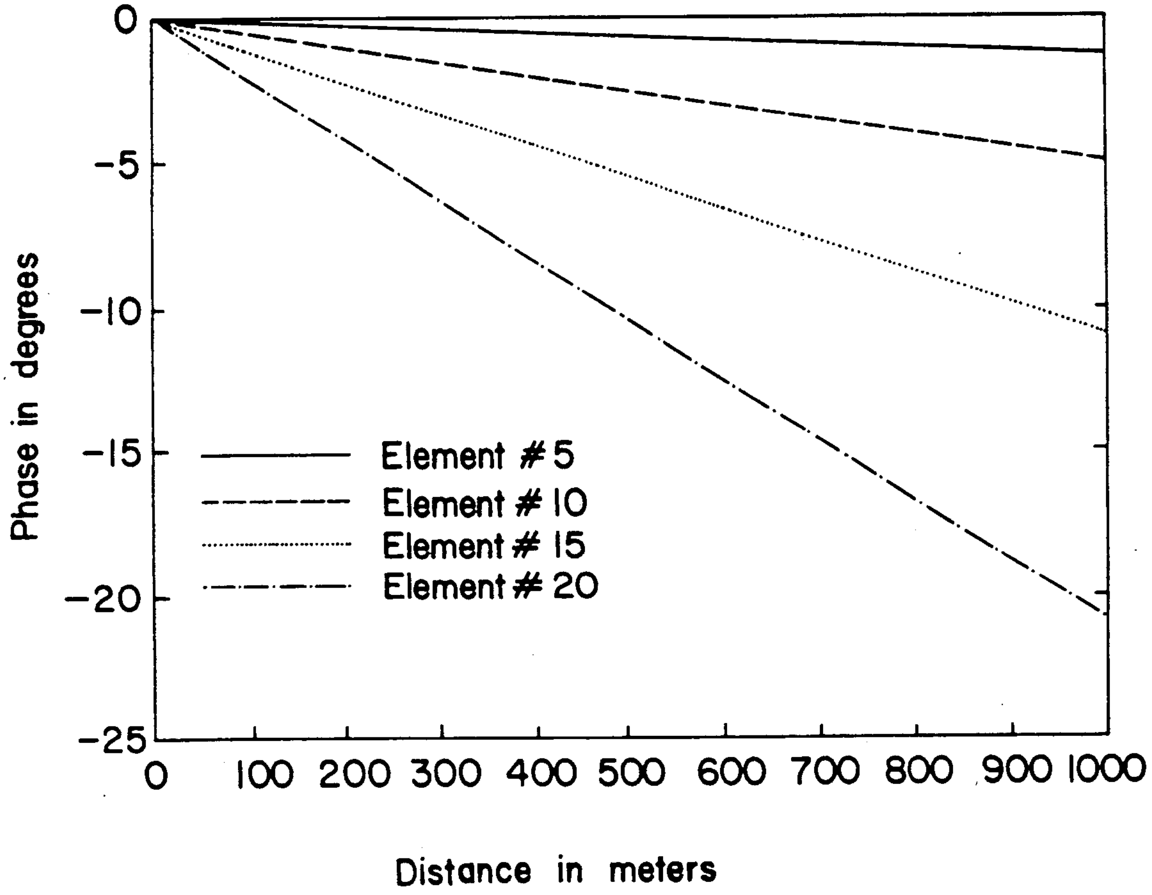

Phase Corrections to move focal point from 60Km to 61Km

FIG.9

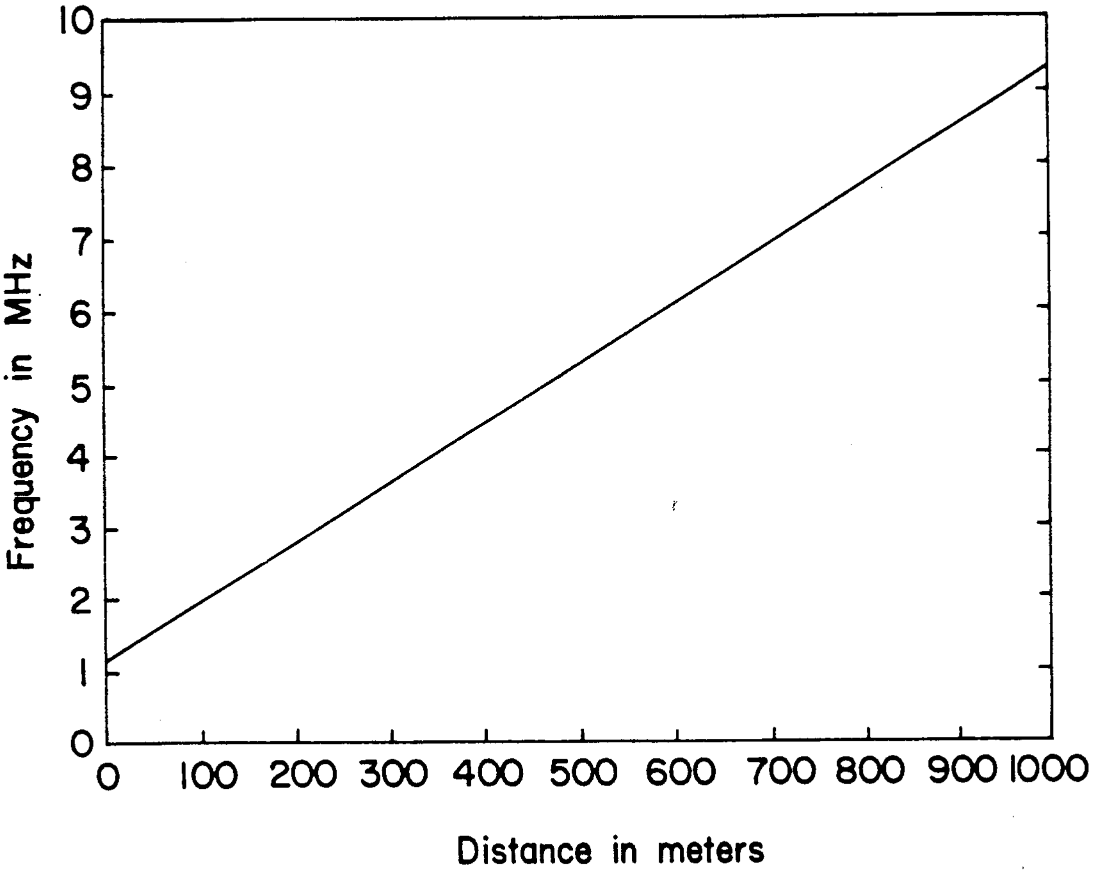

Frequency correction to move focal point from 60Km to 61Km

FIG.10

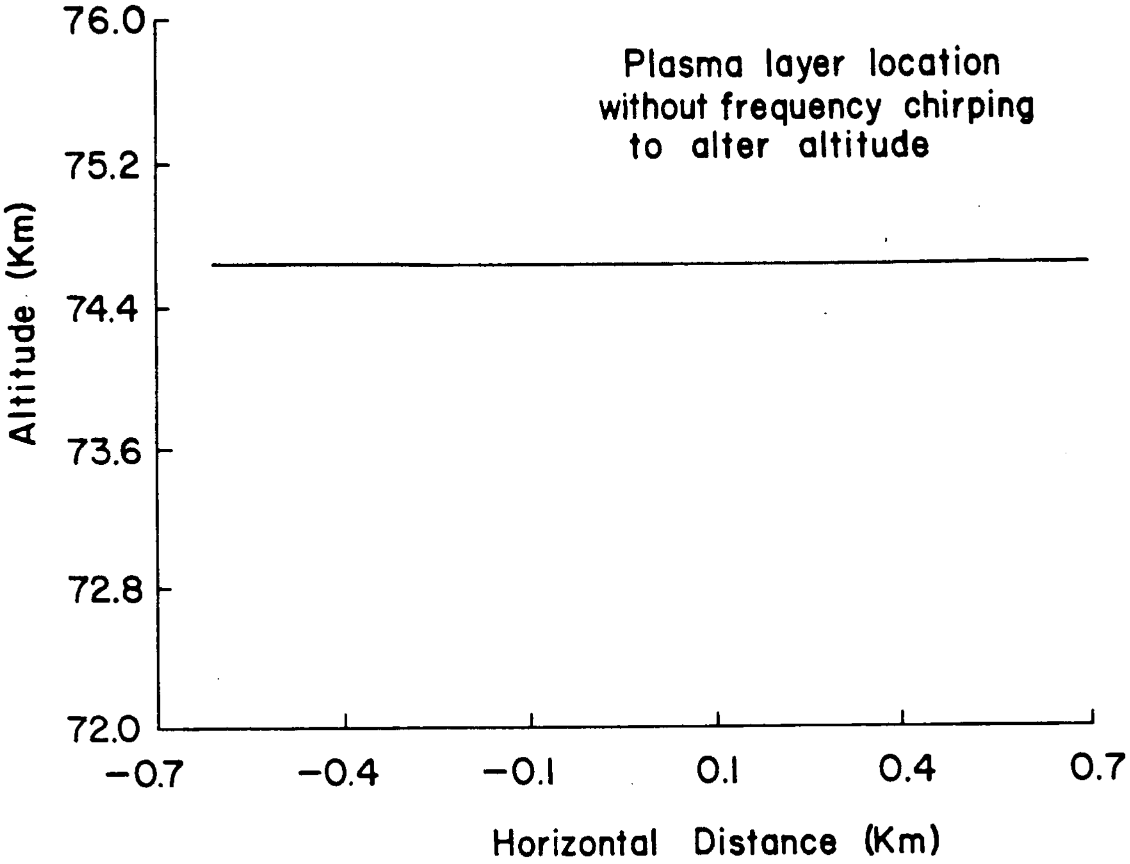

Heater focusing at 75Km with 550 MHz frequency and 154 dBW

FIG. 11

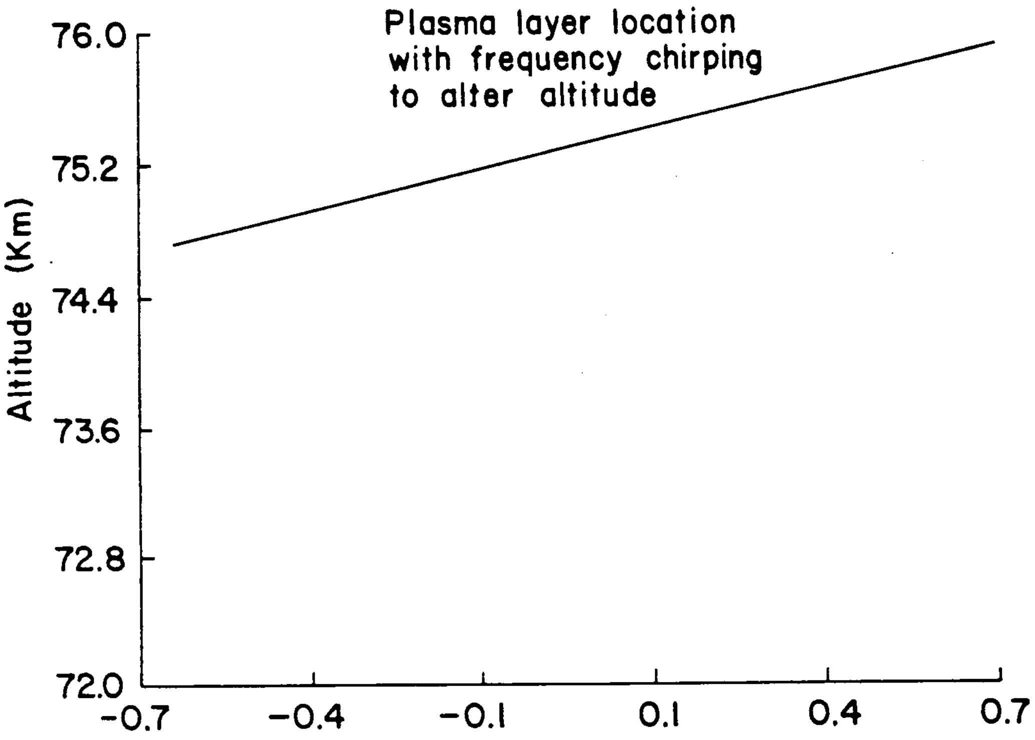

Horizontal Distance (Km)

Heater Focusing at 75Km with 550 to 559.375 MHz frequency and 154 dBW

FIG. 12

FIG. 13

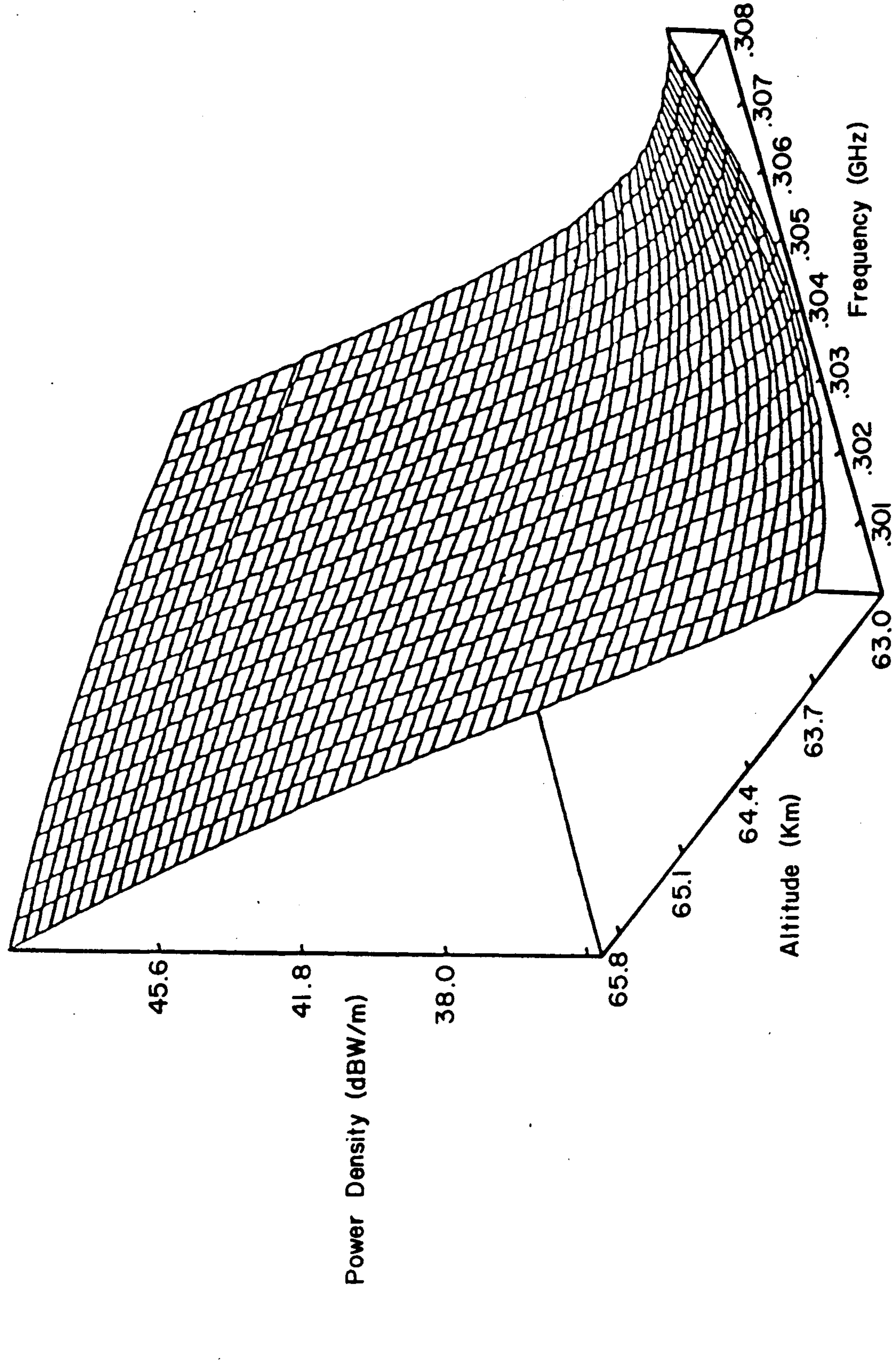

Effects of Frequency Chirping: Heater focused at 70 km with 300-308 MHz frequency chirping

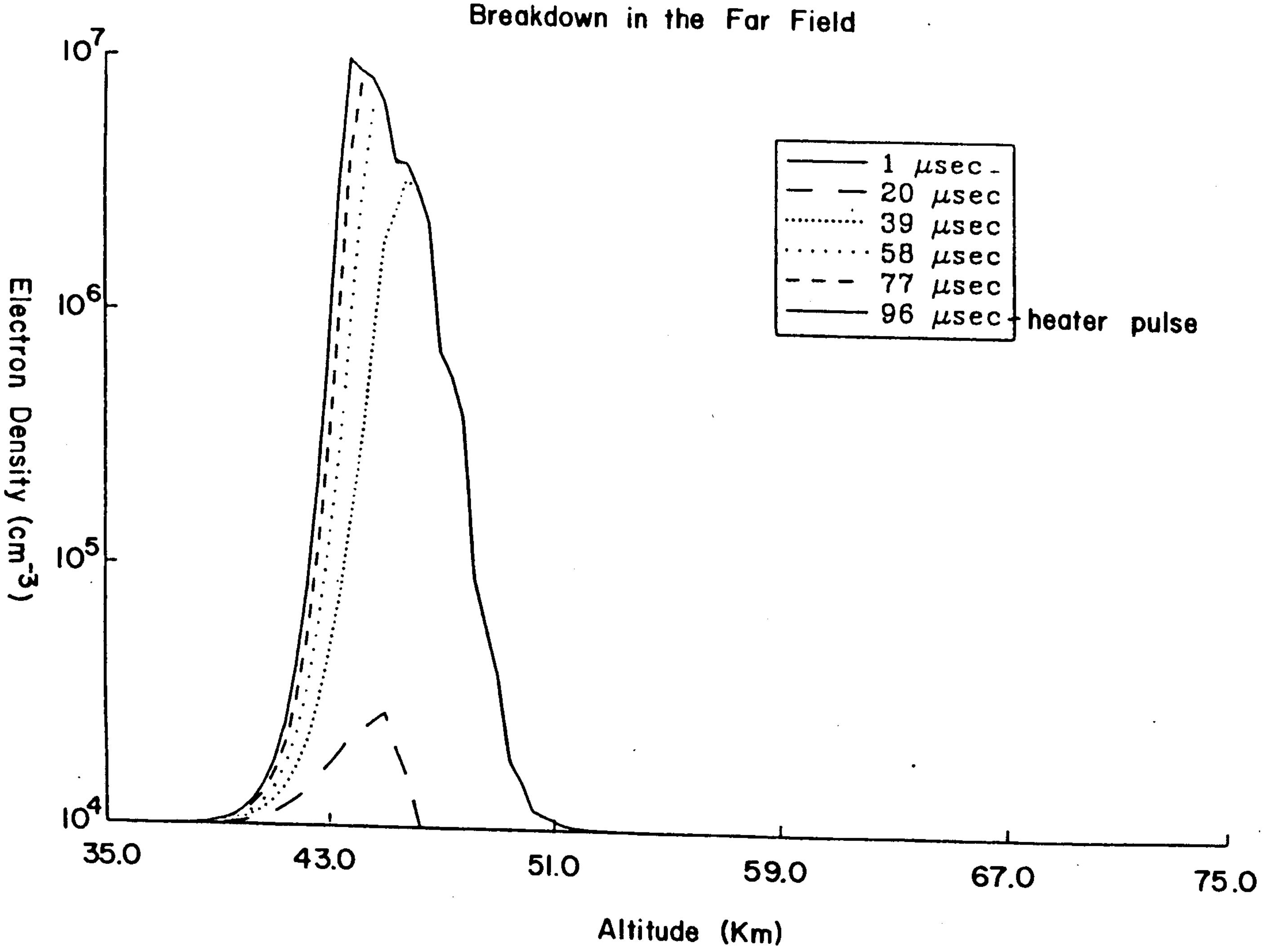

FIG.14

Vertical Electron Density Profile for a Non-Focused Heater Array, ERP=154dBW, frequency=300 MHz

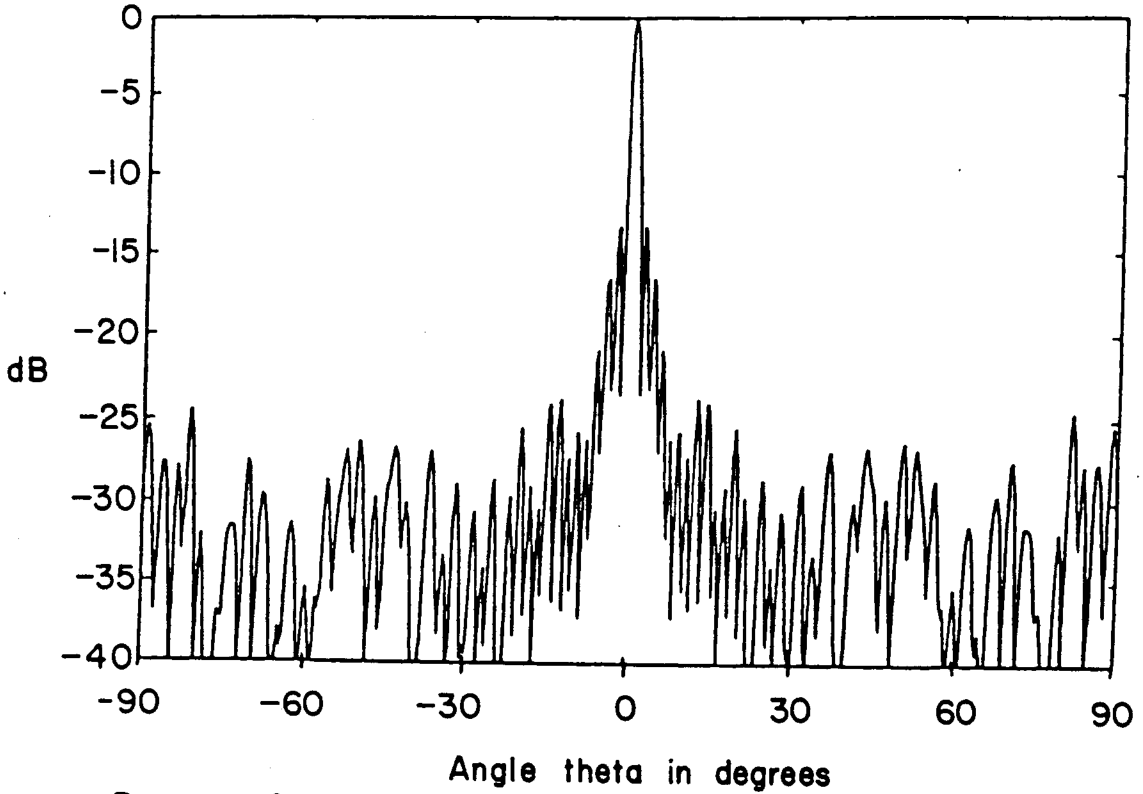

Pattern of a thinned array antenna with randomized spacing containing no grating lobes

FIG. 15

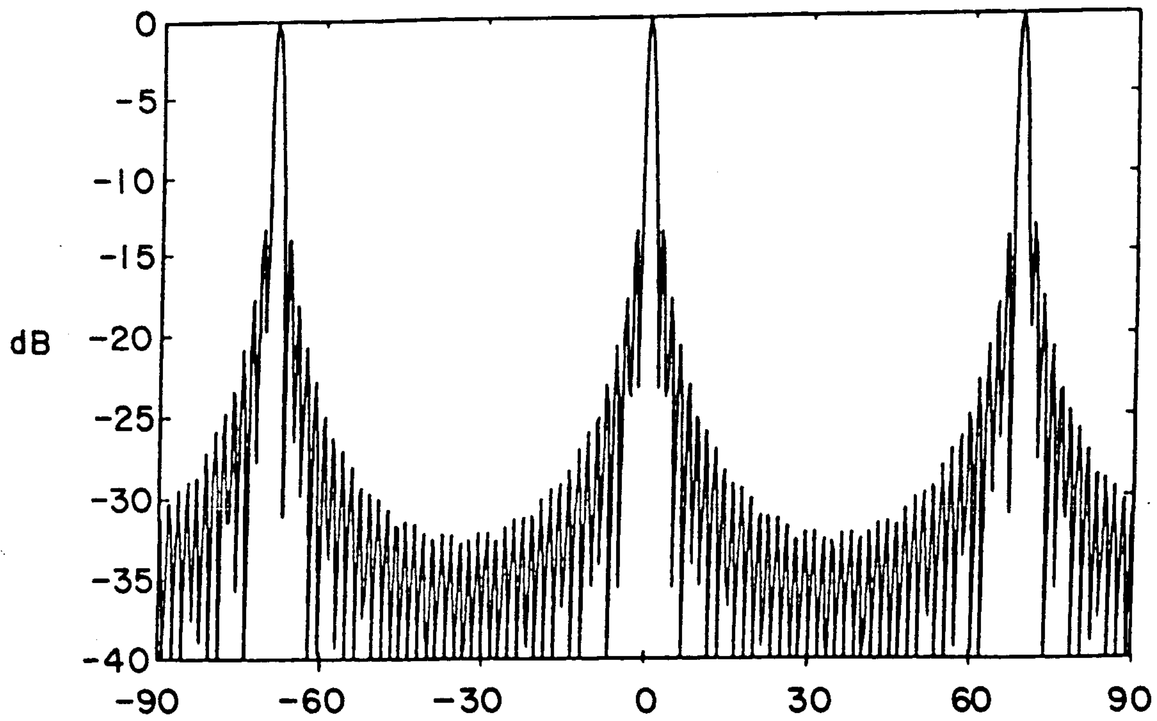

Angle theta in degrees

Pattern of a thinned array antenna with uniform spacing containing grating lobes

FIG. 16

ARTIFICIAL IONOSPHERIC MIRROR COMPOSED OF A PLASMA LAYER WHICH CAN BE TILTED

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates to generation of a Artificial Ionospheric Mirror (AIM), or a plasma layer in the atmosphere. The AIM is used like the ionosphere to reflect RF energy over great distances.

2. Description of the Related Art

In the past, the technique of using the ionosphere as a mirror to reflect radio waves, or RF energy, has given Ham Radio operators the ability to send transmissions over long distances. This technique has also provided radar systems the ability to look "over the horizon." Variations and fluctuations in the ionosphere, however, can render the effectiveness of such communications uncertain. Thus, the desirability of creating controllable plasma layers in the atmosphere for communications purposes has been recognized. See, for example, U.S. Pat. No. 4,686,605 issued to Eastlund and U.S. Pat. No. 4,712,155 issued to Eastlund et al.

Previous experiments directed toward creating plasma layers for communications have suffered from the inability to control the inclination of the plasma layer so that signals could be transmitted and received from various ranges. In other words, while one could create a plasma layer in the atmosphere at a lower altitude than the ionosphere, point to point communications would be limited in range based on the reflection angles of the transmitted and reflected signals.

SUMMARY AND OBJECTS OF THE INVENTION

In view of the limitations of the related art it is an object of this invention to generate a plasma layer that could be angled or tilted with respect to the horizon in order to affect signal transmission range.

The present invention provides a system and method for generating a plasma layer at controlled altitudes and inclinations that acts as an artificial ionospheric mirror (AIM) to reflect RF signals. The AIM increases the range and predictability with which RF energy may be reflected off the AIM for communications purposes. More specifically, a tiltable AIM is created by a heater antenna controlled in phase and frequency. The heater antenna phase shift scans a beam to paint a plasma layer. The heater antenna continually refocuses at a higher altitudes by frequency shifting to tilt the plasma layer.

BRIEF DESCRIPTION OF THE DRAWINGS

Further details of the present invention are explained with the help of the attached drawings in which:

FIG. **1** shows creation of an AIM by a heater antenna and use of the AIM for tracking aircraft and reflecting radio waves.

FIG. **2** shows a typical heater array.

FIG. **3** shows the spacial relationship for a heater array used in defining heater array focusing.

FIG. **4** is a graph showing that power is at its upper bound at the antenna focal point.

FIG. **5** shows generation of plasma by a heater array.

FIG. **6** illustrates generation of a plasma layer by scanning a heater antenna.

FIG. **7** illustrates generation of a tilted plasma layer by scanning and refocusing a heater antenna.

FIG. **8** shows generation of a plasma layer using a heater antenna to scan with a line rather than a point.

FIG. **9** shows the phase corrections to move the antenna focal point from **60** Km to **61** Km.

FIG. **10** shows the frequency corrections to move the antenna focal point from **60** Km to **61** Km.

FIG. **11** is a plot of altitude v. distance location of plasma without frequency chirping.

FIG. **12** is a plot of altitude v. distance location of plasma with frequency chirping.

FIG. **13** shows the power density change after refocusing using frequency chirping .

FIG. **14** graphs the free electron density v. altitude for an unfocused array.

FIG. **15** shows an antenna power pattern without grating lobes.

FIG. **16** shows an antenna power pattern with grating lobes.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

FIG. **1** illustrates the creation and use of An Artificial Ionospheric Mirror (AIM) for tracking aircraft and reflecting radio waves. A heater antenna **1** radiates power causing avalanche ionization or breakdown releasing free electrons in the atmosphere to generate the AIM **2**. The heater antenna **1** is an array which can be used to focus energy at varying altitudes and elevations to tilt the AIM **2** using phase and frequency control. The AIM **2** simulates the ionosphere **3** which is also used to detect "over the horizon targets" **5**. In addition, the AIM **2** can reflect radio signals transmitted from a transmitter **6** to a receiver **7** over long distances.

A typical heater antenna is shown in FIG. **2**. It consists of an array of multiple active radiating elements **10** having their individual phase and frequency controlled from a control module **12**. The radiating element **10** is used here to represent all possible antennas, including, but not limited to, dipoles, slots, small or large horns, log-periodic antennas, large parabolic reflectors, etc.

FIG. **3** shows the spatial relationship for a focused heater array. To have the electric fields from all of the array elements focus, or arrive in phase at a distance R_o in the near field of the array, it is necessary to correct the phase of each element to compensate for the phase delay difference from the center element due to the additional phase path W_{ij}. If R_o is much larger than the maximum D_{ij} in FIG. **3**, then the phase delay can be approximated in wavelength to be:

$$W_{ij}=(D_{ij})^2/(4*R*g) \quad (1)$$

where g is the wavelength of the heater frequency. Equation 1 is referred to as the quadratic phase error. If this error is less than g/8 when the element (i,j) is on the outer edge of the array, then the distance R_o is said to be in the far field of the array.

In order to focus the array at R_o, it is necessary to have several wavelengths of phase error from the outer elements of the array. That is, the term "focus" is used in this context to mean that the electric field from the array is concentrated in a desired spatial region.

FIG. **4** shows the degree of focusing that can be accomplished. This is a vertical pattern of an array whose elements have been phase shifted to focus at 60 Km. The array has 400 elements with a total width and length of 2000 g. The peak of the pattern is determined by the $1/R_o^2$ dependence.

The AIM ionization layer is created by using this focused power to ionize an area in the atmosphere, as shown in FIG. 5. The microwave breakdown of air occurs where free electrons gain enough energy from an electric field to generate additional free electrons until no more can be generated, thereby resulting in avalanche ionization, or breakdown. This causes the generation of a plasma layer **21**. For example, a pulse of power from the heater begins to propagate in the z direction shown in FIG. **5**. As the field propagates, more free electrons are generated. A breakdown point descends vertically from the focal point of the propagating field giving thickness to the ionized layer, or plasma layer, until all ionization stabilizes. This "clamping" creates a thin vertical plasma layer.

Simulation results show that when an array **20** is focused at a point **22**, electric field power peaks at the focal point. Simulation results shows that given a focused microwave source avalanche ionization, or breakdown will occur at a power level 3–10 dB below the focal point power level.

To create an AIM, the heater array is focused at a desired altitude to maximize power at a point and thereby generate plasma. The heater antenna then "scans" the phase of each array element to move the focal point.

FIG. **6** illustrates creation of a noninclined AIM layer. The heater array **30** is first focused at point **31**. The heater array scans horizontally by phase shifting to a point **32** creating an avalanche ionization line **33**. Next, the heater array scans from a point **34** to a point **35** creating another avalanche ionization line **36**. The heater array continues this process to create an ionization plane or AIM layer.

In order to form an inclined AIM cloud, each new ionization line must occur at a slightly higher altitude. By altering the phase or frequency of the array elements, the focal point can be moved up in altitude, as described below.

FIG. **7** illustrates creation of an inclined AIM. The heater array **40** is first focused at point **41**. The heater array scans along the x direction to point **42** to generate avalanche ionization along line **43**. Next, as in creation of a non-inclined AIM, the heater array scans along the x and y directions directly below point **44**. The heater array **40** alters either phase or frequency to refocus to a higher altitude in the z direction to the point **44**. The heater array then scans along the x axis to point **45** to create the avalanche ionization line **46**. The heater array continues this process to create a tilted ionization plane or tilted AIM layer.

FIG. **8** shows that the preferred method of generating a plasma layer uses a heater antenna to scan with a line rather than a point. Scanning using a line is preferred since an AIM can be created in the atmosphere in less time. To create lines of ionization rather than points, a rectangular array **50** is used. In the array **50**, radiating elements are focused only along the plane of the long dimension of the rectangular array, creating a line of ionization **53**. The array is then scanned along the x-y axis and in altitude along the z axis to create another ionization line **55**. More ionization lines are similarly generated to form a tilted AIM layer.

In order to create a tilted AIM it is necessary to refocus the heater array at successively higher altitudes. Moving the focal point by changing the phase of each element of the heater in a very precise manner is not practical. Moving the focal point away from the initial location requires changing the phase on each element. The phase change required is near the rms tolerance level, typically 1 degree. FIG. **9** shows the required phase corrections to move the focal point from 60 Km to 61 km. Elements **5**, **10**, **15**, and **20** have distances 5d, 10d, 15d, and 20d, respectively from the center of the antenna, where d = 25 meters It is clear from FIG. **9** that it is impractical to alter numerous antenna element phases to move the focal point to create tilted patches for AIM applications. 2000 elements may be required here to generate enough power to ionize the atmosphere.

The second method of refocusing is accomplished by first setting the phases of all elements for the initial focal point and then moving the focal point by changing the frequency rather than the phase. This frequency chirping method is less precise, but easier for hardware implementation because precise phases for 2000 elements need not be changed. FIG. **9** shows the required phase corrections to move the focal point from 60 Km to 61 Km. FIG. **10** shows that the focal point can be moved 100 meters by increasing the frequency approximately 1 Mhz. The resulting focal point power levels are not completely optimized, but simulation shows that there is less than a 0.1 db difference between the frequency shifted peaks and those obtained by phasing.

Tilting the AIM using frequency chirping is practical to achieve in a real system. FIG. **11** shows the plasma layer location with no frequency chirping. FIG. **12** shows the plasma location of the same heater creating a tilted AIM by increasing frequency from 550 MHz to 559.375 MHz while scanning horizontally. The result is a smooth patch with a 45 degree inclination.

While it is true that frequency chirping does not achieve the same power as phase focusing at the higher altitude, the difference for small frequency chirps is negligible. FIG. **13** shows actual power density data generated by a 300 MHz heater focused at 70 km with the frequency chirped to 308 MHz.

In the far field region, power meets its upper bound without focusing. For a far field or unfocused array, there is no way to raise the ionization altitude or create a tilted AIM. Ionization takes place at a point where there is enough power to initiate breakdown and where there is low enough neutral density (i.e. pressure). This usually occurs between 40 and 50 km altitude as shown in FIG. **14**. Consequently, a near field focused antenna is required to create a controlled AIM.

The focused pattern is a picture of constructive and destructive interference of the fields from the elements of the array. Other interference positions, or grating lobes, outside the focal point occur when some of the array elements add up in phase. The power of grating lobes can be kept below that of the main lobe, or focal point, by having a large number of elements in the array and spreading them out over the array aperture. This is called thinning the array. For square arrays having 400 elements or more grating lobes can be kept down by 20 db or more from the focal point.

The degree of focusing depends on the ratio of focal range to aperture size. The half power width from peak "V" can be approximated as:

$$V = 2*g*(R_o/L)^2 \qquad (2)$$

where L is the length of the array which is assumed square for equation 2. The power gradient at the half power point "grad(P)" can be approximated as:

$$grad(P)=10/V(db/meter) \quad (3)$$

For an AIM it is desirable that the power gradient be high because this directly determines the gradient of the electron density of the generated ionized cloud. The electron density must be high to avoid RF losses caused by absorption. Hence V be small, preferably less than 2 Km. A heater frequency of 300 MHz and a focal distance of 70 Km would project an aperture size greater than 2 Km. Note in equation 2 that array size scales with the square root of frequency.

Since a near field antenna is required, the near field of the heater antenna may be required to extend to reach distant points. This is accomplished by increasing the array size. It may not be economically feasible to fill this entire aperture with elements, hence a thinned array is utilized.

If a thinned array had its elements uniformly distributed, there would be many grating lobes in the radiation pattern of the array. These grating lobes can be eliminated by randomly spacing elements. However, random spacing puts power from the grating lobes into the average side lobe level. If no new elements are introduced when the aperture is increased, then the peak power of the main lobe remains constant and the main lobe receives less of the total power as its beamwidth decreases. In order to preserve the efficiency of the heater array, grating lobes must be utilized in creating the AIM cloud or the array can not be heavily thinned. FIG. **15** shows an array with uniform spacing having grating lobes. FIG. **16** shows an array with randomized spacing which eliminates the grating lobes.

Although the invention has been described above with particular reference to certain preferred embodiments thereof, it will be understood that modifications and variations are possible within the spirit and scope of the appended claims.

What is claimed is:

1. A method for generating an AIM, comprising the steps of:

(a) creating avalanche ionization in the atmosphere using a heater antenna;

(b) refocusing said heater antenna to alter the altitude of said avalanche ionization by frequency shifting said heater antenna; and

(c) scanning said heater antenna to paint an avalanche ionization layer.

2. A method for generating an AIM as claimed in claim **1** wherein said heater antenna is focused in the near field.

3. An apparatus for generating an AIM comprising:

(a) a phased array heater antenna which is focused at an altitude to cause an avalanche ionization area to be created in the atmosphere;

(b) means for controlling frequency of individual radiators of said phased array heater antenna to

refocus said altitude of said avalanche ionization area; and

(c) means for controlling phase of the individual radiators to scan said phased array heater antenna.

4. An apparatus for generating an AIM as claimed in claim **3** wherein said phased array heater antenna is focused to cause said avalanche ionization area to be substantially a line.

5. An apparatus for generating an AIM as claimed in claim **4** wherein said means for controlling phase moves said line substantially at a constant altitude and said means for controlling frequency moves said line to different altitudes.

6. An apparatus for generating an AIM as claimed in claim **4** wherein said phased array heater antenna is a rectangular array and said line is formed parallel to a long dimension of said rectangular array.

7. An apparatus for generating an AIM as claimed in claim **3** wherein said phased array heater antenna is focused to cause said avalanche ionization area to be substantially a point.

8. An apparatus for generating an AIM as claimed in claim **7** wherein said means for controlling the phase moves said point substantially at the same altitude and said means for controlling frequency moves said point to different altitudes.

9. An apparatus for generating an AIM as claimed in claim **3** wherein said phased array heater antenna is focused in the near field.

10. A method of generating an AIM comprising the steps of:

(a) focusing a phased array heater antenna at an altitude to cause an avalanche ionization area to be created in the atmosphere;

(b) controlling the frequency of individual radiators of said phased array heater antenna to refocus said altitude of said avalanche ionization area;

(c) controlling phase of the individual radiators to scan said phased array heater antenna.

11. A method of generating an AIM as claimed in claim **10** wherein said step of focusing causes said avalanche ionization are to be substantially a line.

12. A method of generating an AIM as claimed in claim **11** wherein said step of controlling phase moves said line substantially at a constant altitude and said step of controlling frequency moves said line to different altitudes.

13. A method of generating an AIM as claimed in claim **11** wherein said phased array heater antenna is a rectangular array and said line is formed parallel to a long dimension of said rectangular array.

14. A method of generating an AIM as claimed in claim **10** wherein said step of focusing causes said avalanche ionization area to be substantially a point.

15. A method of generating an AIM as claimed in claim **14** wherein said step of controlling phase moves said point substantially at the same altitude and said step of controlling frequency moves said point to different altitudes.

16. A method of generating an AIM as claimed in claim **10** wherein said step of focusing is performed in the near field.

* * * * *

Covert Geopolitics

DISCLOSURE, **HAARP / MANMADE DISASTERS**

US Air Force Admits They Can Control Weather

JUNE 6, 2015 | ECLINIK LEARNING | 66 COMMENTS

The US Air Force and DARPA would like us to believe that they have stopped using HAARP in Alaska for research and experiment. Even then, we all know that there are other HAARP systems out there in the form of radar communication and surveillance systems that are rigged on top of mobile platforms that are deployable anywhere in the world.

All they need to do is twist a button to change the frequency to microwave range and increase transmission power enough to reach and heat up the atmosphere above the target.

Unnatural weather patterns suggest HAARP is still here.

Air Force Bombshell: Admits They Can Control Weather – HAARP

(https://eclinik.files.wordpress.com/2015/03/haarp-28.jpg)

While HAARP and weather control has been called a conspiracy theory by the mainstream media and government officials, during a Senate hearing on Wednesday, David Walker, deputy assistant secretary of the Air Force for science, technology and engineering, dropped a bombshell (http://www.adn.com/2014/05/14/3470442/air-force-prepares-to-dismantle.html?sp=/99/100/&ihp=1) in answer to a question asked by Lisa Murkowski in relation to the dismantling of the $300 million High Frequency Active Auroral Research Program in Gakona this summer.

Walker said this is "not an area that we have any need for in the future" and it would not be a good use of Air Force research funds to keep HAARP going. "We're moving on to other ways of managing the ionosphere, which the HAARP was really designed to do," he said. "**To inject energy into the ionosphere to be able to actually control it. But that work has been completed.**"

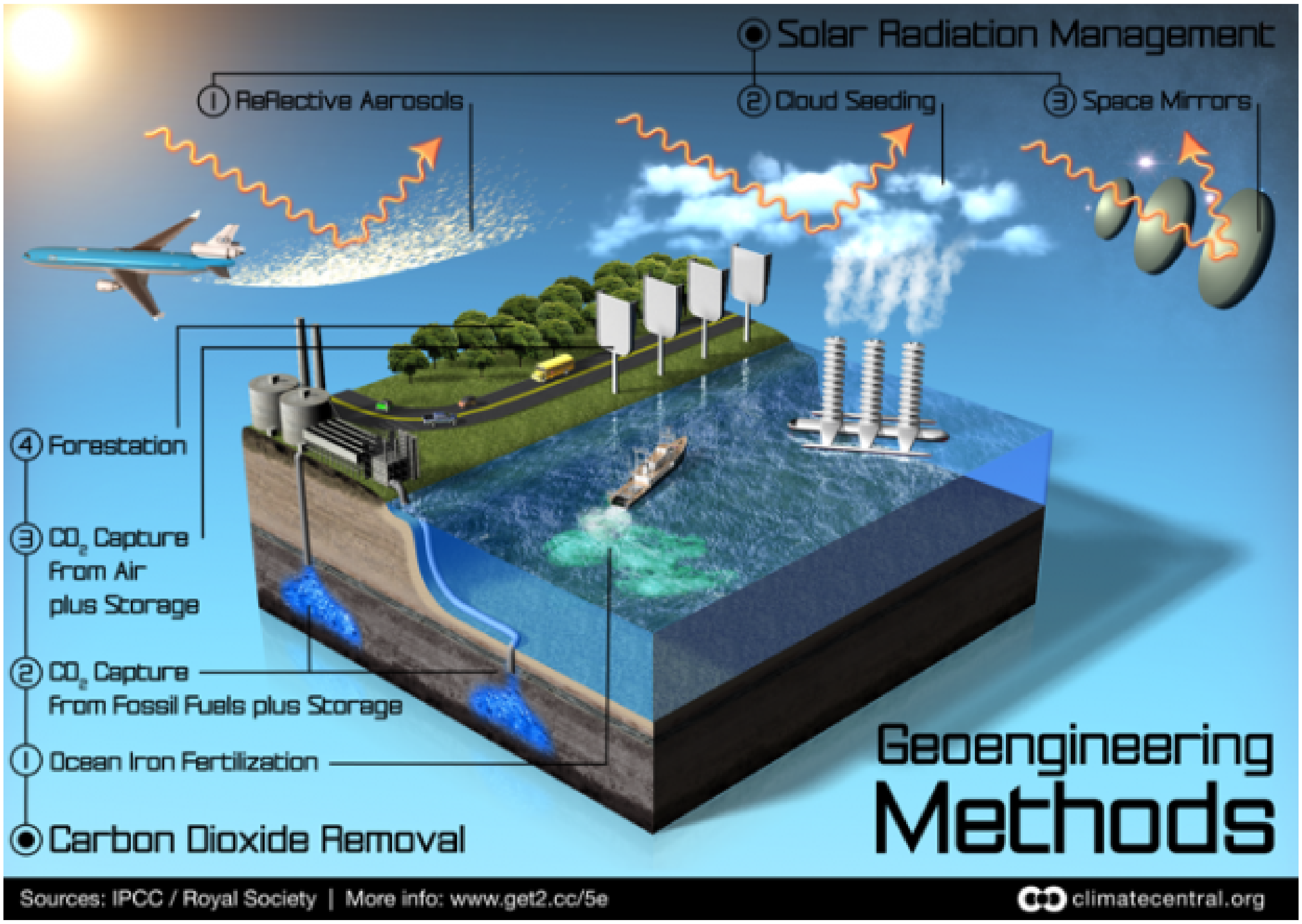
Solar Radiation Management
① Reflective Aerosols
② Cloud Seeding
③ Space Mirrors
④ Forestation
③ CO_2 Capture From Air plus Storage
② CO_2 Capture From Fossil Fuels plus Storage
① Ocean Iron Fertilization
Carbon Dioxide Removal
Geoengineering Methods
Sources: IPCC / Royal Society | More info: www.get2.cc/5e
climatecentral.org

Many believe HAARP was created and has been used for weather control, with enough juice to trigger hurricanes, tornadoes and earthquakes and comments such as this bring about the question of whether conspiracy theorists are more on target than anyone has admitted to date.

This is not the first time a public official has acknowledged that HAARP and weather control is not only possible, but has been and continues to be, used as a "super weapon," as evidenced by a statement in 1997 by former U.S. Defense Secretary William Cohen (http://www.cheniere.org/correspondence/022501.htm), where he said *"Others [terrorists] are engaging even in an eco-type of terrorism whereby they can alter the climate, set off earthquakes, volcanoes remotely through the use of electromagnetic waves… So there are plenty of ingenious minds out there that are at work finding ways in which they can wreak terror upon other nations…It's real, and that's the reason why we have to intensify our [counterterrorism] efforts."*

Is it still just a conspiracy theory if public officials admit it is true?

Watch the question and answer session below:

source (http://woundedamericanwarrior.com/air-force-bombshell-admits-they-can-control-weather-haarp/#) »

Are they really dismantling the Alaska based HAARP?

How about the floating HAARP transmitters masquerading as mobile radars used in the Philippines?

The US Air Force and DARPA would like us to believe that they have stopped using HAARP in Alaska for research and experiment. Even then, we all know that there are other HAARP systems out there in the form of radar communication and surveillance systems that are rigged on top of mobile platforms that are deployable anywhere in the world.

All they need to do is twist a button to change the frequency to microwave range and increase transmission power enough to reach and heat up the atmosphere above the target."

The technology is covered under **US Patent 4,686,605** on the "*Method and Apparatus for Altering a Region in the Earth's Atmosphere, Ionosphere, and/or Magnetosphere.*"

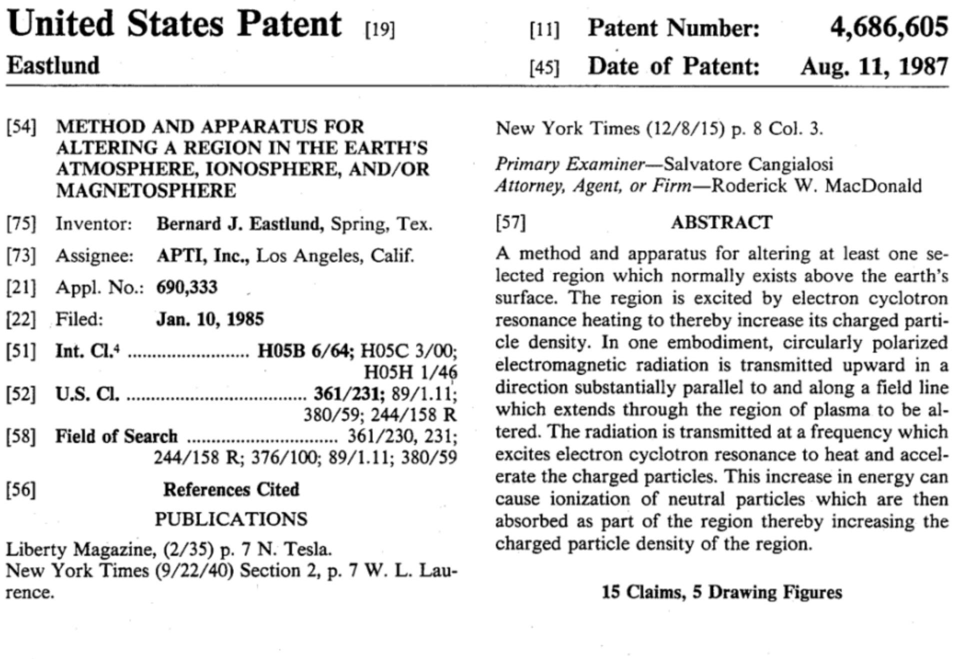

United States Patent [19]
Eastlund

[11] **Patent Number:** **4,686,605**
[45] **Date of Patent:** **Aug. 11, 1987**

[54] **METHOD AND APPARATUS FOR ALTERING A REGION IN THE EARTH'S ATMOSPHERE, IONOSPHERE, AND/OR MAGNETOSPHERE**

[75] Inventor: **Bernard J. Eastlund**, Spring, Tex.

[73] Assignee: **APTI, Inc.**, Los Angeles, Calif.

[21] Appl. No.: **690,333**

[22] Filed: **Jan. 10, 1985**

[51] **Int. Cl.⁴** **H05B 6/64;** H05C 3/00; H05H 1/46

[52] **U.S. Cl.** .... **361/231;** 89/1.11; 380/59; 244/158 R

[58] **Field of Search** .... 361/230, 231; 244/158 R; 376/100; 89/1.11; 380/59

[56] **References Cited**

PUBLICATIONS

Liberty Magazine, (2/35) p. 7 N. Tesla.
New York Times (9/22/40) Section 2, p. 7 W. L. Laurence.
New York Times (12/8/15) p. 8 Col. 3.

Primary Examiner—Salvatore Cangialosi
Attorney, Agent, or Firm—Roderick W. MacDonald

[57] **ABSTRACT**

A method and apparatus for altering at least one selected region which normally exists above the earth's surface. The region is excited by electron cyclotron resonance heating to thereby increase its charged particle density. In one embodiment, circularly polarized electromagnetic radiation is transmitted upward in a direction substantially parallel to and along a field line which extends through the region of plasma to be altered. The radiation is transmitted at a frequency which excites electron cyclotron resonance to heat and accelerate the charged particles. This increase in energy can cause ionization of neutral particles which are then absorbed as part of the region thereby increasing the charged particle density of the region.

15 Claims, 5 Drawing Figures

(https://eclinik.files.wordpress.com/2014/02/us-patent-4686605.png)

- https://patentimages.storage.googleapis.com/pdfs/US4686605.pdf (https://patentimages.storage.googleapis.com/pdfs/US4686605.pdf)
- U.S. Attacked Philippines with HAARP (https://geopolitics.co/2014/02/04/u-s-attacked-philippines-using-haarp/)

LEYTE SURVIVE

(https://eclinik.files.wordpress.com/2014/02/us-haarped-phil1.jpg)

(https://geopolitics.co/2015/06/06/us-air-force-admits-they-can-control-weather/130401144228-sbx-1-us-navy-0401-horizontal-gallery/)

(https://geopolitics.co/2015/06/06/us-air-force-admits-they-can-control-weather/attachment/781140641989246129/)

(https://geopolitics.co/2015/06/06/us-air-force-admits-they-can-control-weather/sbx_haarp-magnetometer1/)

(https://geopolitics.co/2015/06/06/us-air-force-admits-they-can-control-weather/espiral-en-canada-haarp/)

(https://geopolitics.co/2015/06/06/us-air-force-admits-they-can-control-weather/haarp7/)

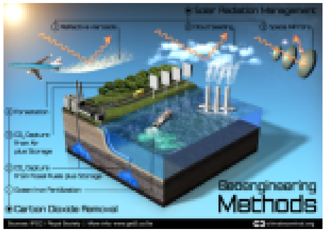

(https://geopolitics.co/2015/06/06/us-air-force-admits-they-can-control-weather/60aec-geoengineering-methods-infographic/)
Just three units of these mobile HAARP platforms, masquerading as radars, can make the hurricane steerable by a common method known as *triangulation*.

One of the significant sources of funds for the fascist Nazionist Jesuit (https://geopolitics.co/2015/03/11/hidden-history-of-the-incredibly-evil-khazarian-mafia/)Khazarian (https://geopolitics.co/2015/03/11/hidden-history-of-the-incredibly-evil-khazarian-mafia/) Mafia (https://geopolitics.co/2015/03/11/hidden-history-of-the-incredibly-evil-khazarian-mafia/) is the healthcare industry which registered a whopping $3.09 trillion in 2014 (http://www.plunkettresearch.com/health-care-medical-market-research/industry-trends), and is projected to soar to $3.57 trillion in 2017, in the US alone.

We can help take down the Dark Cabal by avoiding Khazarian pharmaceutical drugs, defeat any viral attack and scaremongering easily by knowing how to build our own comprehensive antiviral system. Find more about it here (http://myeclinik.com/).

ALASKA **GEOENGINEERING** **HAARP** **US AIR FORCE** **WEATHER WEAPON**

66 thoughts on “US Air Force Admits They Can Control Weather”

1. **Senatssekretär FREISTAAT DANZIG** says:
   JUNE 6, 2015 AT 5:58 PM

   1

   3

   I
   RATE THIS

Reblogged this on Aussiedlerbetreuung und Behinderten – Fragen and commented:

Glück, Auf, meine Heimat!

1. **Jeremy Tripi** says:
   JUNE 11, 2015 AT 11:38 AM

   16

   57

   I
   RATE THIS

   And if you believe this report you are the worlds biggest ass clown out ive spent 13 yrs in the army and nothing can or has ever been known to control the weather and ive carried a high top secret clearance

   1. **Joe Mama** says:
      JUNE 13, 2015 AT 5:01 AM

      82

      4

      I
      RATE THIS

      “High top security clearance”? Rebok or Nike?

   2. **knows your a joke** says:
      JUNE 26, 2015 AT 11:06 PM

      31

      3

      I
      RATE THIS

      High top clearance huh. You were in the guard for four years and faked a back injury as soon as there was any wind of a possible deployment. You have never been any where and you live at home with your dad. Your whole life us a lie and if you keep telling these lies the stolen valor act will catch up with you. Just admit to everyone you are full of it and move on. What would your family think of you if they knew all the lies you have told.

3. **Jerry** says:
   NOVEMBER 20, 2015 AT 10:29 AM

   33

   3

   I
   RATE THIS

   Your doing your job well as a disinformation agent … Well done soldier , now stand down and let the truth out !

4. **Carol** says:
   DECEMBER 18, 2015 AT 5:23 AM

   7

   3

   I
   RATE THIS

   13 years is not a long time…..

5. **doggie duty** says:
   DECEMBER 20, 2015 AT 3:44 AM

   30

   1

   I
   RATE THIS

   You carry a "HIGH" top secret security clearance? That in itself lends to you know nothing about security clearance. And secondly, even if you DO have some level of a security clearance, that does not give you knowledge of every program in the United States. Gimme a break dude. That is laughable.

6. **Me** says:
   FEBRUARY 4, 2016 AT 4:45 AM

   12

   1

I
RATE THIS

BULLSHIT…if you believe that Sir , your either Ignorant or not privleged enough to be able to recieve classified info…from the people you work for . The HAARP, is for Real ive bern keeping an eye on this reversed radio waves antenna for over 20 yrs. its just so sad that the govorment has used it on its own country, numerous times..
That last time was 2015 the largest Hurricane to ever hit this planet … in MEXICO…

7. **Chaz Maztolo** says:
   MARCH 18, 2016 AT 5:22 AM

   13

   2

   I
   RATE THIS

   Top secret clearance? Aren't you the janitor over at the local middle school…

8. **Candi Workman** says:
   JUNE 1, 2016 AT 9:31 AM

   9

   0

   I
   RATE THIS

   Are you kidding me? Of course you are going to say that, you were military. With top secret clearance. Do we think you would say anything else? Jesus!

9. **Colin R Brophy** says:
   JUNE 9, 2016 AT 8:37 PM

   9

   0

   I
   RATE THIS

   Need to know! Well, you have a clearance so you should know that…. your service like my service doesn't make you know anything….

10. **Bill Connor** says:
    AUGUST 18, 2016 AT 11:31 PM

    2

    1

    I
    RATE THIS

    Project Popeye ? Look it up

11. **Bill Connor** says:
    AUGUST 18, 2016 AT 11:32 PM

    5

    2

    I
    RATE THIS

    Look up Project Popeye ,. we've known how to control the weather a long time now

12. **Prime** says:
    AUGUST 22, 2016 AT 4:40 AM

    7

    4

    I
    RATE THIS

    Did it say army? Nope! Air force has all the intelligent people thats why you have to score high to join but the other branches accept the uninitelligent people like yourself!!! GTFOHWTBS

13. **Zack** says:
    AUGUST 25, 2016 AT 8:17 AM

    2

    1

    I
    RATE THIS

Just because you have a TSSC, doesn't mean you know everything. just like there are things the President doesn't know. If you truly believe this is bogus the HOAH and Charlie Mike. No need to insult people because you believe it fake.

14. **Rodewaryer** says:
    AUGUST 26, 2016 AT 10:25 AM

    4

    1

    I
    RATE THIS

    Correct, HAARP was for influencing the ionosphere but for communication, not weather modification.

15. **Paul Horn (@angelriver)** says:
    OCTOBER 8, 2016 AT 1:11 AM

    3

    0

    I
    RATE THIS

    I've never seen you at any of our meetings!

16. **Star Walker** says:
    OCTOBER 8, 2016 AT 3:40 AM

    2

    1

    I
    RATE THIS

    Whatever, you brainwashed troll.

17. **Micmiller** says:
    OCTOBER 8, 2016 AT 7:08 AM

    1

    0

I
RATE THIS

If you carry a high security clearance you would tell people. For the simple fact you make yourself a target. What ever makes you feel cool

18. **Virginia** says:
OCTOBER 8, 2016 AT 2:34 PM

1

0

I
RATE THIS

There are 37 levels of 'Above Top Secret' so you do not apparently "have a need to know."

19. **Anonofyourbizness** says:
OCTOBER 8, 2016 AT 7:35 PM

1

0

I
RATE THIS

Oh ya? What's your MOS mr. Top secret?

20. **Conspiracists** says:
OCTOBER 9, 2016 AT 10:58 PM

0

0

I
RATE THIS

That's just what you want us to think lol

21. **Steve Paige** says:
OCTOBER 10, 2016 AT 11:02 AM

2

0

I
RATE THIS

Most people, regardless of clearance, absolutely do not have the "need to know" anything outside of their current job. And that "need to know" leaves when you do.

22. **joefox** says:
    OCTOBER 11, 2016 AT 4:08 AM

    0

    0

    I
    RATE THIS

    WOW REALLY WHAT WAS IT??

2. **Deb Ward (@DebWard628)** says:
   JANUARY 19, 2016 AT 1:26 PM

   1

   0

   I
   RATE THIS

   they are making more and no, they didn't close it down..also click on the links right above comment section https://www.youtube.com/watch?v=IG82OfX5WVs

3. **Peggy Severson** says:
   JUNE 5, 2016 AT 9:52 AM

   4

   0

   I
   RATE THIS

   They may not be using HAARP, but are certainly chemtrailing our Alaskan skies. They are trying to control the weather, which isn't what we pay them for.

1. **Nicole Leatherman** says:
   OCTOBER 10, 2016 AT 4:39 AM

   3

   1

   I
   RATE THIS

   I agree with you Peggy. They have over 2 dozen FEMA DEATH Camps on American Soil, with 40 Million Coffins ready. We have ISIS Terror Compounds as well,.. I believe it's 35 now. They are all armed and ready for Obama to give them the word to attack. There are over 2200+ Mosques in America – the one's that have been raided, Law Enforcement found tunnels under the Mosques leading to massive amounts of Guns, Rockets, Bombs,.. I'm more than positive every Mosque, No go zone and individual homes that these "refugees" stole,… Are all Armed. Obama and Clinton made sure they were. The ultimate goal is to DEPOPULATE 3 Billion lives in the Western World. They will allow 500 Million to stay alive – which in turn they will be either sex slaves for the women and children, and labor for the able bodied men. Speak up and get this out – with Clinton wanting to bring in 65,000 more and forecasting 100,000 more each year – even if we don't have what's already here causing Destruction on Epic Proportions – Look at Sweden, France, Germany, London, Australia,..etc – you just can't look at the developments happening and by now, haven't figured out that these "Peaceful Refugees" are Terrorist Soldiers with one intention, to kill all Western Civilization and and turn the world into a Islam/Muslim place where their Barbaric Sharia Law will rule. Sharia Law will put women especially at the worst situation of all,.. they will be downgraded to below a rat, beatings are within the law, Rapes and "Honor Killings" will be enforced. If you get "accused" of doing anything, you can be hung, stoned to death, burned to death, beheaded,.. what ever they feel like doing – they will do it without even 1 second of hesitation. Girls 12 – up will more than likely be forced to marry one of these Soldiers, who will then beat, rape, belittle and maybe even killed. Boys – Muslims are Pedophiles, they use boys for pleasure there is no age range as they have killed children as young as a few weeks. They also make arranged marriages for the young girls, toddlers actually – usually 4 and up, where in a lot of cases, they end up with a man 3 – 5 times her age. Unfortunately they are brutally raped and sodomized to the point of death on that very marriage night. If they manage to live – their reproductive functions are very likely destroyed. If they cannot use you to have babies,.. they will probably get sold into sex slavery and will eventually die. From what I understand about the Agenda – they are going to take all the little children from families and be raised by the Government. Those child care facilities are more than likely going to be run by men and women who are very disturbed and are Pedophiles. You can figure out what's in store for those poor babies. Men will be for the most part slaughtered. The one's allowed to live – slaves. If you look up FEMA Executive orders – you will see that you are being run by the Government – they can displace you, take your home, car, anything you cherish and probably throw you in a FEMA DEATH

Camp and or killed,.. they can do this for NO particular reason,.. it's you with the clothes on your back. You will no longer be in charge of what you eat, drink or amounts of it. They will enforce MANDATORY "vaccines" which will be poison in some form. This is no joke – this is happening. Think about it,.. does bringing in 65,000 more of these people that want us all dead, make any sense? To a SANE person – of course not. Obama and Clinton are not on our side. Look up Agenda 21 Depopulation. All of the politicians, leaders, etc. that are FOR OPEN BORDERS are all corrupt and under the impression that they will be immune to being a part of the Depopulation amount of 3 Billion that must be exterminated. They aren't, once they have done what Obama and Clinton wants, they have no use. They are dead to. (those people deserve it). Well my friend – I will Pray to God for GOOD to win over EVIL, please do the same. We are at the point of TRUST NO ONE. Never let anyone take your weapons, if you see someone that looks shady – SHOOT. You Must Profile people now. Don't wait til' they pound down your door – SHOOT. God Bless the World! In Jesus name AMEN. Pray friend.

4. **Samuli Alatalo** says:
JUNE 10, 2016 AT 8:02 AM

1

1

I
RATE THIS

We say no for haarp project!

https://www.facebook.com/We-say-no-for-Haarp-Project-366463017307/

5. **diginess** says:
JUNE 13, 2016 AT 5:21 AM

3

1

I
RATE THIS

To the best of my knowledge, the ionisphere only carries charged particles which affect how radio waves are reflected. I'm not aware of any way of controlling the weather with this. There may be some edge cases where clouds can be created or moved a bit with energy from radars / microwave, but that's not really that much, and ultimately, you aren't getting as much back as you put in.

   1. **stvjns** says:
   JULY 1, 2016 AT 2:24 PM

2

0

I
RATE THIS

The layers surrounding our planet are heated by the charged particles and that creates weather effects. The patents for this technology demonstrate many many militarily “useful” effects, including opening unique communication channels while simultaneously blocking other communication channels, generating large area brainwave entrainments, a radar-like ability to show activity deep in the ocean and even underground, and more effects. Look into it, it’s interesting.

2. **Donnar** says:
OCTOBER 7, 2016 AT 9:16 AM

1

1

I
RATE THIS

Yes that’s correct. Weather is created in the lowest layer called the troposphere where the clouds exist and weather conditions are formed. I learned this in high school in a meteorology class.

6. **caleb slaaen** says:
AUGUST 13, 2016 AT 11:42 PM

1

0

I
RATE THIS

people have eyes but cant see and people have ears but cant hear do a little research and deploy yourself as a a disciple
if life is such an amusement then pick your side
faith is he who dont lie and the devil meaning evil cant stand the beginning nor the end (genisis and revalation)
so if life is hidden from us its because of our actions where not corrected so we end up w/ deceiving people
we must reconstruct our lives yet if people would buy into their future like counting your

blessings before they hatch we could possibly live with god (god is the future) think about it the word convenience is rude and their is so much evil behind that word (i dont want to get into that) but patience opens a view of opportunities
understand the american dream has been controlled for years
if its as simple as feeding a cow grass why not let domestic animals roam they where put here for what to own (wrong)
weve been owned and mentally controlled (a pattern has occured)
we the people have explored and found electricity awesome right
understanding it has been difficult for some people yet we choose to hang it over our heads like candy we want it but dont care who made it so convenient right
FYI if u didnt know static electricity pulls water (plastic materials is the moon possibly and wool is mostly the earth trees,life) no one really knows what the moons really made of but i do know the moon pulls the water like static would so thats my best answer molecules on the surface
so the phenomenon lives on as our lives could be at jeopardy
so much $ has been invested in this ask donald trump his uncle and father met nikola tesla also trump family was involved in creating a static generator (van-degraaff generator)possible controlling the storm would be a better generator my brother and i have ideas on that conversation w/some more understanding and info of errors its like starting at the bottom w all previous mistakes to happen again government info must be replayed because getting info from them is like history replaying itself hidden from the main agenda us the people if i could create a storm wouldnt u think some idiotic person try it (hidden from us isnt good and us knowing isnt good either)idk im not a

7. **Mike** says:
   OCTOBER 7, 2016 AT 10:00 AM

   2

   1

   I
   RATE THIS

   OK 1st off the ball on the red ship is the x ban radar. I watched it pearl harbor in 2006 ish. This article is so full of half truths it's not funny. So a radio Antena can transmitted power into the ionisphere to create a high pressure system as it bubbles the ionisphere. News flash hurricanes are low pressure system lol

8. **Jesus Christ** says:
   OCTOBER 9, 2016 AT 11:00 PM

   0

   0

I
RATE THIS

Whore of The Bible here, sir. The Hurricane? Did you do this? Did you flood your own land? What about the other countries? THUS SAYETH THE LORD, “I DO THIS. YOU DO NOT.”

9. **Al** says:
OCTOBER 22, 2016 AT 8:36 AM

0

0

I
RATE THIS

The Air Force gave HAARP to the University of Alaska. They recently had an open house for the public: http://gi.alaska.edu/facilities/haarp

And, yes, we have been manipulating the weather for awhile. One civilian example is LA. They have been cloud seeding since the 1950s, per The LA Times: http://www.latimes.com/local/lanow/la-me-ln-l-a-officials-seeded-clouds-during-el-nino-storm-in-hopes-of-more-rain-20160308-story,amp.html

2. **Johnson** says:
JUNE 8, 2015 AT 7:45 AM

1

0

I
RATE THIS

Reblogged this on Johnsono ne'Blog'as.

3. **alfatah69** says:
JUNE 8, 2015 AT 3:54 PM

2

0

I
RATE THIS

Reblogged this on LIBYA AGAINST SUPER POWER MEDIA.

1. **Janie** says:
   OCTOBER 3, 2015 AT 11:38 AM

   4

   0

   I
   RATE THIS

   I'm so glad that Libyans are waking up to this.
   I was no fan of Muammar Gaddafi, but he was murdered because he didn't cow tow to the corrupt elites in the West, America and Israel, and now he's dead we have the first ever *World Bank* squatting in Libya. Screams volumes…

   1. **Anthracinus** says:
      OCTOBER 9, 2016 AT 5:36 AM

      0

      0

      I
      RATE THIS

      *kowtow

4. **WakeUpCitizens** says:
   JUNE 9, 2015 AT 6:26 AM

   6

   0

   I
   RATE THIS

   Just wait till the world wakes up and discovers that HAARP, and the GWEN Towers around the U.S. can alter your mind as well!!

   1. **Old Woman** says:
      FEBRUARY 4, 2016 AT 5:02 AM

      5

1

I
RATE THIS

Thats the truth…they arent going to admit to the fact that the big hurricane that hit Mexico this last yr. 2015 and was said to be the biggest hurricane ever to hit this planet , i live in Wadhington State and the day before it hit Mexico the chemtrails in the sky here was so massive , it was unbelievable. So at that point i knew the haarp had been used and im convinced they caused the hurricane. Ive been studing the HAARP. For over 20 yrs. Its a reversed radio waves Antenna that DOES mess with the weather as well as can single out a person and mess with their mind. Seriously there is also one in Florida too. … Hurricane Catrina… so I say to Our Goverment, who do you think you are messing with. MOTHER NATURE??? Admit what you assholes have done and let the chips fall where they may.

Wake up AMERICA.. Its time to take action.

1. **Brandon** says:
   MARCH 14, 2016 AT 12:56 PM

   1

   0

   I
   RATE THIS

   If a plane left tails like they do, we would hear them down here. Idk if these are related. (Most likly) But they are dusting us with something. Becareful who you say that too. You can get in trouble. They'll lock you up and give you "experimental" drugs.

5. **Michael James** says:
   JUNE 12, 2015 AT 1:59 PM

   11

   2

   I
   RATE THIS

   Walker said "We're moving on to other ways of managing the ionosphere, which the HAARP was really designed to do"

   The Air Force admits that HAARP affects the ionosphere, what has this to do with weather control.

HAARP is a means to manage the ionosphere, not the troposphere.
The ionosphere is the layer of the Earth's atmosphere that is ionized by solar and cosmic radiation. It lies 75-1000 km (46-621 miles) above the Earth.

The troposphere is the lowest layer of Earth's atmosphere and site of all weather on Earth. The troposphere is bonded on the top by a layer of air called the tropopause, which separates the troposphere from the stratosphere, and on bottom by the surface of the Earth.

All weather occurs only in the troposphere.

HAARP affects only the ionosphere.

1. **Joyce Herder** says:
   JULY 3, 2015 AT 4:21 AM

   3

   0

   I
   RATE THIS

   Thank you for having some brains

2. 0 **eClinik Learning** says:
   DECEMBER 17, 2015 AT 1:20 PM

   3

   0

   I
   RATE THIS

   The distance traveled by an electrical signal depends on its electrical pressure [V], i.e. the higher the voltage [V] the farther the distance reached by that signal.

   If your 3-volt remote control could change your TV channel it doesn't mean you can change your neighbor's TV channel across the road.

   If HAARP could reach the ionosphere, why not the troposphere which, as you said, is nearer?

   All the HAARP operator needs to do is lower the transmission voltage from 200 million volts to 100 or 50 million volts by twisting a single knob.

A Tesla "death ray" may need a different frequency with the same amount of voltage as HAARP, but your car's spark plug only needs 40kv to 100kv to force that quarter inch gap of atmosphere to become conductive.

By comparison, FM radio transmitters broadcast at 50kw power while HAARP is about 4 billion watts according to its patent.

6. **New Comer** says:
   JUNE 13, 2015 AT 12:59 AM

   11

   0

   I
   RATE THIS

   If the USAF does control the weather, why do we have people starving because they cannot grow food due to drought? Why do we in the U.S. Have states like California in a man made drought?
   I'm sure someone somewhere is making tons of Money. Its amazing how greed is showing it's self. I believe in making money but not at the expense of taking life human life.

   1. **Dr. Agnew** says:
      JUNE 16, 2015 AT 6:58 AM

      1

      0

      I
      RATE THIS

      Oh, I think he has figured it out.

   2. **Donita Stepke-Thompson** says:
      DECEMBER 19, 2015 AT 4:30 AM

      14

      3

      I
      RATE THIS

Yes, it has caused the drought here. Check out satellite images of the Pacific by the west coast. We are bombarded, then it drifts in — along with the streams of chemicals they spray on us in our mountains and valleys. I guess they've decided to create rain for us, because it's been pouring for a week or more, which we need…but it also makes us wonder wtf is raining down on us?!! People laugh at me and others for sharing this stuff. Whatever. Keep sleeping, sheep. Keep drinking your fluoride water. Keep watching that Fox "news" like my asshole neighbor and friends and family do. The attacks no longer phase me. Screw them all. This whole thing goes deeper than just chemtrails and weather manipulation. All of us now have metallic nanoparticles inside of us. Do your ears ring constantly? Loudly? Mine do, and have for a long time. It's like a high-pitched radio frequency that is incessant! Jump down that rabbit hole if you dare, but I warn you — it will blow your freaking mind what they have in store for us.

1. **amdarela** says:
   OCTOBER 14, 2016 AT 4:02 AM

   0

   0

   I
   RATE THIS

   I need to find out more about the high pitched sound I'm hearing. What causes it? It started a year ago. Anyplace you can recommend to research? Links? Thanks

7. **Janie** says:
   OCTOBER 3, 2015 AT 11:53 AM

   5

   0

   I
   RATE THIS

   Why people don't believe that such contraptions have been created to alter weather patterns astonishes me. Mankind is still a basic animal, but many are quite intelligent enough to make these systems, so for those who are either unwilling or too stupid to comprehend that this is real, then I suggest a refresher course of secondary school…And you can bet that certain governments are using these systems to profit from other countries suffering, the US, UK Israel and they're ilk being the biggest offenders.

8. Pingback: We Need Economic System Change Summit, not Climate Change | Covert Geopolitics
9. Pingback: U.S. Attacked Philippines with HAARP | Covert Geopolitics
10. **Vaughan** says:

APRIL 28, 2016 AT 5:16 PM

1

0

I
RATE THIS

Raw & Real Data: 100 years ago the Wright Bros first flight was called a hoax by some scientist because they said it was not possible.🙂 Next google “Jet Fuel additives” than read the MSDS sheets on them. There known as Polymer Fuels and additives that work well with HAARP to make artificial clouds. I also know an ex RAAF that says he never saw anything out of normal while serving for 8 years but agrees the Clouds are weird over the skies in Melbourne, AU these days.

11. **Mike B** says:
AUGUST 8, 2016 AT 3:20 AM

2

0

I
RATE THIS

Is this where you pass out tinfoil hats? No? Why? I would have thought you had an overabundance of the shiny little things.

12. **Mike B** says:
AUGUST 8, 2016 AT 3:23 AM

2

0

I
RATE THIS

“SENSIBLE COMMENTS ARE HIGHLY APPRECIATED.”

….though mostly absent.

13. **William L. Finucan** says:
AUGUST 21, 2016 AT 2:21 AM

2

0

I
RATE THIS

…..AFFECT IS NOT CONTROL…. Boll weevils affect cotton plants…they don't control them.

14. **Me** says:
    AUGUST 21, 2016 AT 8:24 PM

    3

    0

    I
    RATE THIS

    If chemtrails were sprinkling stuff from the air, wouldn't the govt folks be affected as well? That's like pooping in your own cage and then living in it….it doesn't make sense

15. **Tomeeka** says:
    OCTOBER 7, 2016 AT 6:30 AM

    0

    0

    I
    RATE THIS

    Uhh "high top secret security clearance" means jack shit if you dont have what the military calls the need to know. You are only allowed the information that is top secret to the level that allows you to get your job done. Top secret clearance does not mean you all the sudden know all of the top secret info. How do I know? Well not only do they tell us this I can also say so because I have a top secret clearance. Theres no one waiting to brief you on everything going down on every level as soon as you get a clearance. Its a need to know basis.

16. **Michael** says:
    OCTOBER 7, 2016 AT 7:09 PM

    1

    0

    I
    RATE THIS

I'm not ameteorologist, but I am familiar with enough physics that I find this merits questions. First: define "controllable." Even if there were some means of influencing a hurricane's path, they wouldn't just be able to reverse it, because of inertia. Second: define "control." When people say "control the weather," it conjures up images of TPTB deciding just how the weather should act that day, and implementing it. This ability to micromanage is not possible: for one, the Navier-Stokes equations, which model fluid flow, by necessity also model turbulence (it would make a lousy model if they didn't describe all known phenomena). This means the equations (which, by the way, haven't been solved) have chaotic solutions, and so we can't always even predict the result of a given action. Thisis why I find the language misleading; but "influence the weather" doesn't sound nearly so ominous, does it?

17. **jockwilliamdoubleday** says:
    OCTOBER 7, 2016 AT 7:20 PM

    0

    0

    I
    RATE THIS

    There are many H.A.A.R.P. arrays on U.S. (Rothschild) military ships around the world. So closing one array is nothing. And we must wonder WHY they are closing it. I'm sure it's for even more nefarious reasons than keeping it running.

18. **Greg** says:
    OCTOBER 8, 2016 AT 12:56 AM

    0

    0

    I
    RATE THIS

    Ya'll are a ruthless bunch. Lol

19. **Marc Lloyd** says:
    OCTOBER 8, 2016 AT 8:57 PM

    0

    0

    I
    RATE THIS

Weather modification has been used for warfare purposes since at least the 1960's – Operation Popeye: cloud-seeding was used over the Ho Chi Minh trail in Vietnam to slow down the enemy and that's not even the tip of the iceberg. It's all there in the declassified documents for anyone who has a brain to see. It's not a "conspiracy theory" – it's been openly acknowledged for decades.

20. **Mkie** says:
    OCTOBER 9, 2016 AT 12:51 AM

    0

    0

    I
    RATE THIS

    This and many other forms of skullduggery has been happening since 95..may 95..
    WAKE UP

21. **oscar m. recina** says:
    OCTOBER 15, 2016 AT 1:59 PM

    0

    0

    I
    RATE THIS

    if so why hasn't it controlled hurricanes?

22. **Frank J** says:
    OCTOBER 20, 2016 AT 12:42 PM

    0

    0

    I
    RATE THIS

    Hey oscar,

    These radar installations and mobile ship radar platforms are what is creating and directing the direction of the hurricanes. Not saying every event is man made, but lots have questionable aspects and can be linked to the x band ship being deployed to the Philippines as there are 2 HAARP installations in Japan. The patent uses three beams to steer the storm cell.

Michael, ultimately influence and control are virtually the same- it's more so the context that you view the term "control". The weather works on its own, just like your mind isn't telling you to breathe. The point is that if they would like a "natural" event to Happen, they can make it happen – when and where. Sure they can't turn a hurricane around, but why would you even try? How would that look natural lol. Also the point is to create a weakened country so they need to lean on ol Uncle Sam and their friends at the world bank and IMF. Similar events to this recently would be the sudden earth quake that struck China's super computer…. wow, how fortunate for the USA if the worlds fastest super computer(China's) were critically damaged or even just taken offline for awhile

23. **anna jordan** says:
DECEMBER 5, 2016 AT 2:35 AM

0

0

I
RATE THIS

https://en.wikipedia.org/wiki/Operation_Popeye

Nov. 8, 1966 H. W. KASEMIR ETAL **3,284,005**

WEATHER CONTROL BY ARTIFICIAL MEANS

Filed Sept. 23, 1964

FIG. 1

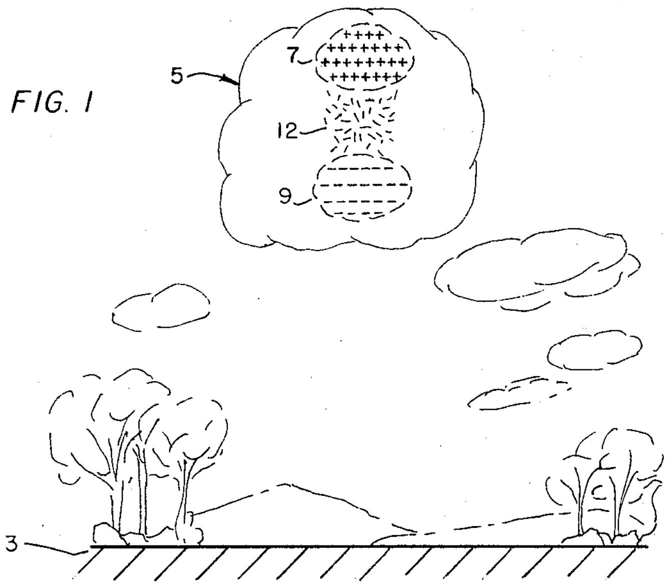

FIG. 2

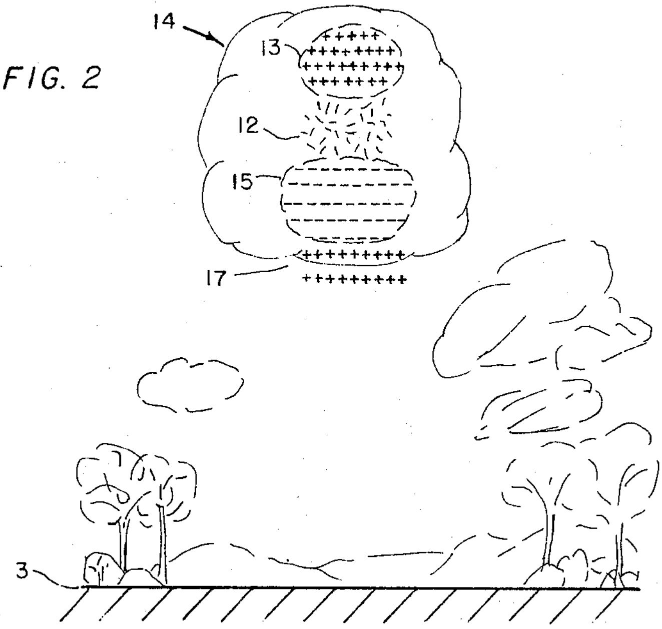

INVENTORS,
HEINZ W. KASEMIR &
HELMUT K. WEICKMANN.

BY Harry M. Saragovitz,
Edward J. Kelly, Herbert Berl
& Julian C. Keppler ATTORNEYS

3,284,005
WEATHER CONTROL BY ARTIFICIAL MEANS

Heinz W. Kasemir, Neptune, and Helmut K. Weickmann, Interlaken, N.J., assignors to the United States of America as represented by the Secretary of the Army

Filed Sept. 23, 1964, Ser. No. 398,816
3 Claims. (Cl. 239—2)

The invention described herein may be manufactured and used by or for the Government for governmental purposes, without the payment of any royalty thereon.

The present invention relates to weather control and more particularly to a method and means for changing the natural electric charge distribution or pattern of thunderstorms and thereby controlling the discharge of lightning. Depending on the way in which the charge distribution is modified, lightning strokes may be inhibited in cases where the natural conditions would normally result in lightning and in other cases the natural charge distribution which would not normally result in lightning may be modified to induce lightning discharge. According to the invention, the natural space charge distribution of the thunderstorm is modified by introducing in regions of strong electric potential gradient a plurality of needle-shaped conductors. The interaction of the electric field and the pointed conductors will produce point discharge or corona. The corona discharge produces a large number of positive and negative ions which are capable of supporting heavy currents needed to hold the electric field below that value at which lightning discharge is initiated. In other cases, the change in the natural space charge distribution caused by the introduction of these conductors can result in an increased electric field gradient in adjacent regions of the atmosphere which will tend to induce lightning discharge in these adjacent regions.

It is therefore an object of the invention to modify atmospheric electric conditions by artificial means.

It is a further object of the invention to control lightning discharge by the seeding of a selected area of the atmosphere with a plurality of needle-shaped conductors.

These and other objects and advantages of the invention will become apparent from the following detailed description and drawing, in which

FIGS. 1 and 2 are pictorial diagrams of two different types of atmospheric conditions to which the invention may be applied.

In FIG. 1, the thundercloud 5 floats in the atmosphere above the earth 3. The cloud contains areas of positive and negative electrical space charge 7 and 9, respectively. A certain electric field gradient will exist in the region between these space charges depending on the spacing thereof and the charge magnitude. Under natural conditions, prior to the lightning discharge, the electric field will cause a limited amount of ionization or glow discharge in the region between the space charges 7 and 9 by corona discharge from precipitation particles in this region. This natural ionization will support or carry electrical current between the charged areas 7 and 9 and therefore will tend to discharge or bleed off the space charge. However, because of the generally spherical shape and the small size of the precipitation particles, they are not efficient ion producers. If the natural rate of charge buildup of areas 7 and 9 is greater than the charge dissipation due to the natural ionization, the potential gradient will increase until intracloud lightning occurs which of course substantially neutralizes the space charge. In accordance with the invention, the region between the space charges 7 and 9 is seeded with needle-like conductors 12. These conductors are made at least 1 centimeter in length and are therefore substantially larger than the natural precipitation particles. The size of the conductors together with their needle-like shape make them more efficient producers of atmospheric ionization due to corona discharge than the natural precipitation particles. The increased ionization in effect raises the atmospheric conductivity to such an extent that the space charge can be bled off or neutralized as fast as it naturally builds up. The result is a slow, controlled dissipation or neutralization of the cloud space charge rather than the rapid and sometimes destructive lightning stroke.

FIG. 2 illustrates a situation in which ground lightning strokes as well as intracloud strokes are possible. In thundercloud 14 is an area 13 of positive charge, a larger area 15 of negative charge and a somewhat smaller area 17 of positive charge near the lower boundary of the cloud. Under natural conditions a small intracloud stroke may start between areas 15 and 17 and may grow until it reaches ground, becoming a ground stroke, thereby carrying negative charge from region 15 to earth and consequently tending to dissipate the area of negative charge. If, however, an intracloud lightning stroke takes place first between positive area 13 and negative area 15, the negative charge may thereby be dissipated to such an extent that no ground stroke will be possible. Thus the intracloud strokes between areas 13 and 15 will use up the negative charge of 15 and thereby lessen the probability of ground strokes. By seeding the region between 13 and 15 with a plurality of needle-shaped conductors 12, the intracloud lightning will be inhibited as explained in connection with FIG. 1 and this will leave a larger negative charge at region 15 which is then available for producing ground strokes. If it is desired to inhibit instead of induce ground strokes, the region between space charge areas 15 and 17 would be seeded with the needle-like conductors. Also, by seeding both of the above-mentioned regions, both ground and intracloud strokes would be inhibited.

The needle-like conductors 12 may comprise metallized nylon threads of from 1 centimeter to 1 inch in length. The metallized nylon is lighter than solid metal needles of the same size and therefore will have a smaller terminal velocity and will remain in the air longer than would all-metal conductors. The conductors may be dispensed at the proper place in the thundercloud by an airplane or a rocket after measurements have been made to determine the most advantageous place for seeding. In order to achieve the desired results, it is necessary that at least several million of these conductors be seeded in each thundercloud.

The present invention may be put to several practical uses. For example, forest fire prevention, reduction of radio interference due to lightning, or as a tactical military weapon for either harassing the enemy or protecting friendly forces from adverse weather conditions.

While the invention has been described in connection with an illustrative embodiment, the principles disclosed herein are of general application; hence the invention should be limited only by the scope of the appended claims.

What is claimed is:

1. The method of inhibiting lightning discharge comprising the seeding of a region between two areas of opposite type electric space charge with a plurality of metallized nylon threads, said nylon threads being at least one centimeter in length.

2. The method of controlling lightning discharge comprising increasing the natural rate of ionization in a region of high electric field by the seeding of said region with a plurality of metallized nylon threads, said nylon threads being at least one centimeter in length.

3. The method of inducing ground lightning strokes from a thundercloud which includes an area of negative electric charge with areas of positive electric charge above and below said negative electric charge, comprising, inhibiting intracloud discharge in the region between said area of negative electric charge and the upper one of said areas of positive electric charge by seeding said region with a plurality of metallized nylon threads, said nylon threads being at least one centimeter in length.

References Cited by the Examiner

UNITED STATES PATENTS

911,260	2/1909	Pennock	239—14
1,284,982	11/1918	Balsillie	239—2
3,019,989	2/1962	Vonnegut	239—2

OTHER REFERENCES

The Washington Post, Sunday August 20, 1916 "Latest Progress in Turning Rain On and Off at Will," 1 page.

Final Report of The Advisory Committee on Weather Control, vol. II, 1957, pp. 105–125 "Project Skyfire."

EVERETT W. KIRBY, *Primary Examiner.*

TRAPPED ELECTROMAGNETIC RADIATION COMMUNICATIONS SYSTEM

Original Filed Sept. 11, 1963

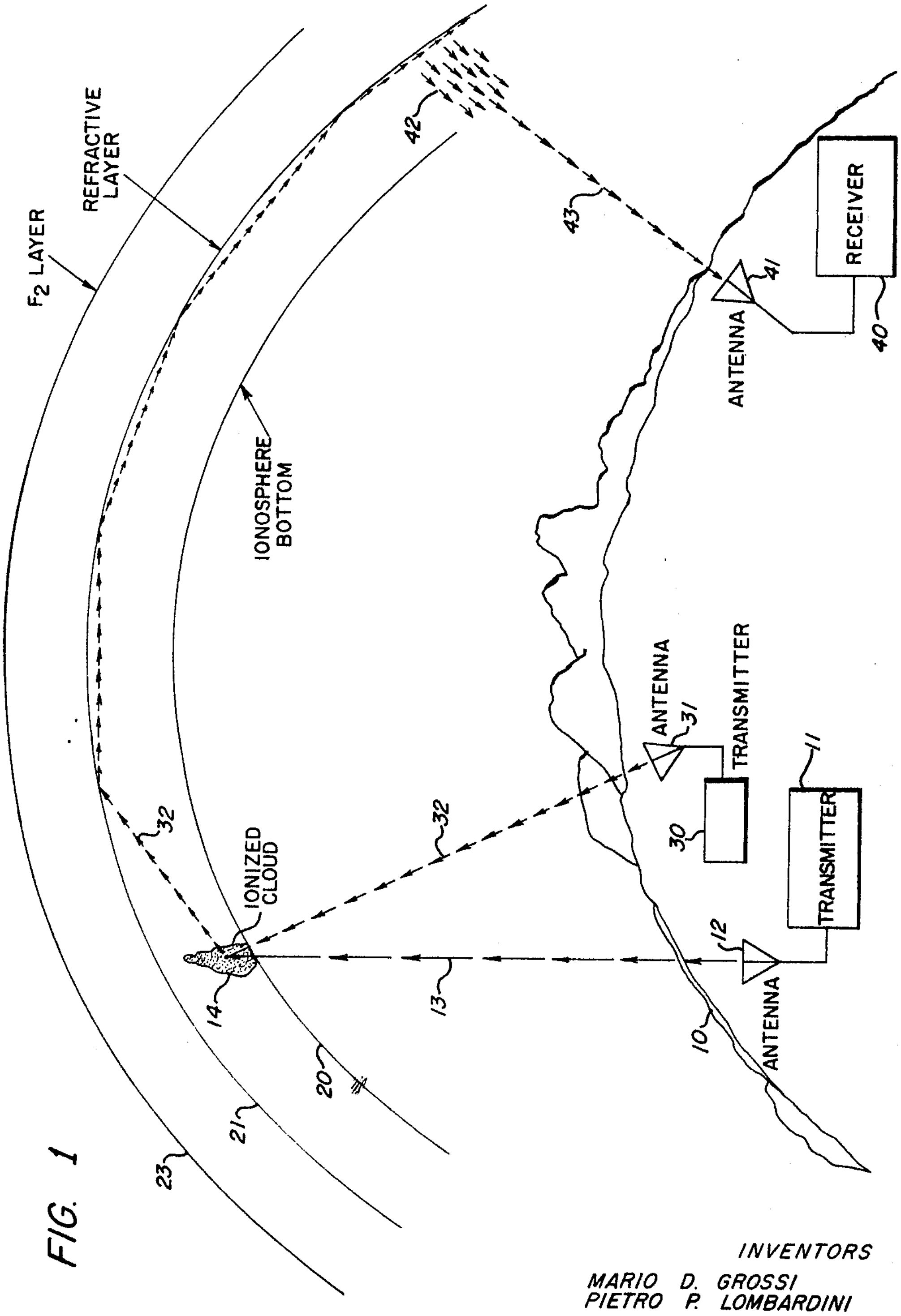
F2 LAYER
REFRACTIVE LAYER
IONOSPHERE BOTTOM
IONIZED CLOUD
RECEIVER
ANTENNA
TRANSMITTER
ANTENNA
TRANSMITTER
ANTENNA
42
43
41
40
31
32
32
30
11
12
14
13
10
20
21
23

FIG. 1

INVENTORS
MARIO D. GROSSI
PIETRO P. LOMBARDINI

BY Donald Brown
AGENT

May 20, 1969 M. D. GROSSI ET AL **3,445,844**

TRAPPED ELECTROMAGNETIC RADIATION COMMUNICATIONS SYSTEM

Original Filed Sept. 11, 1963 Sheet 2 of 4

FIG. 2

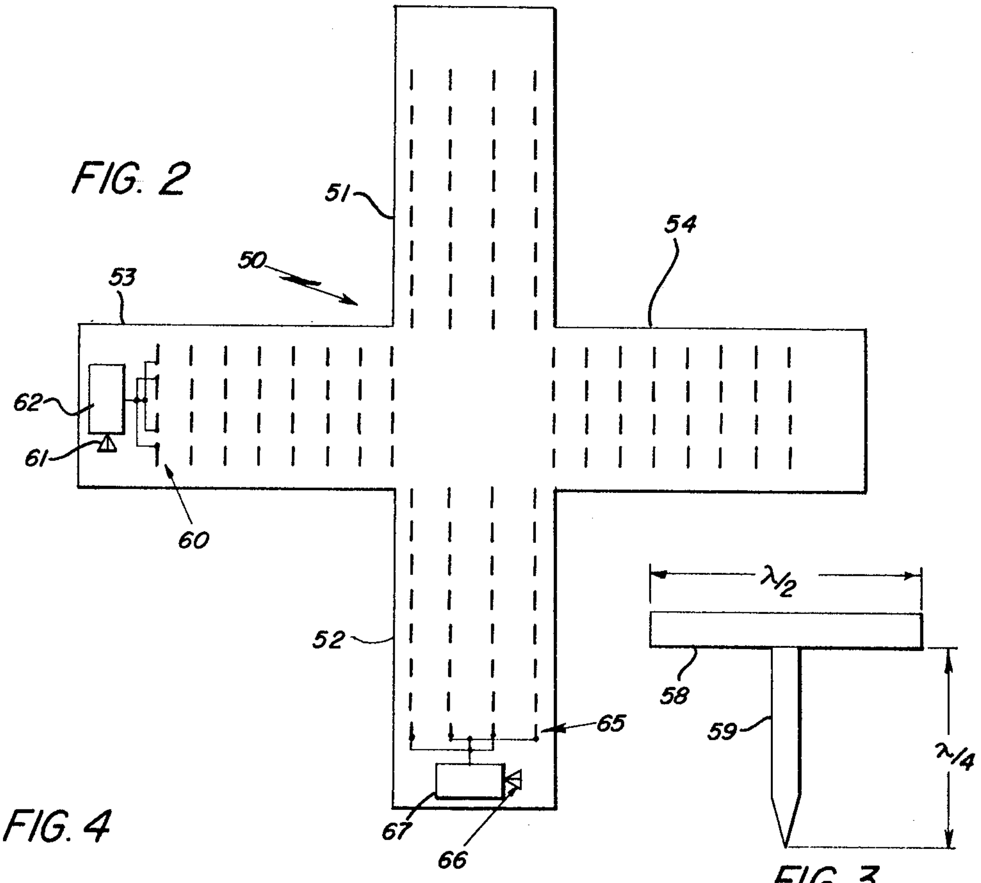

51
50
54
53
62
61
60
52
65
67
66
λ/2
58
59
λ/4

FIG. 3

FIG. 4

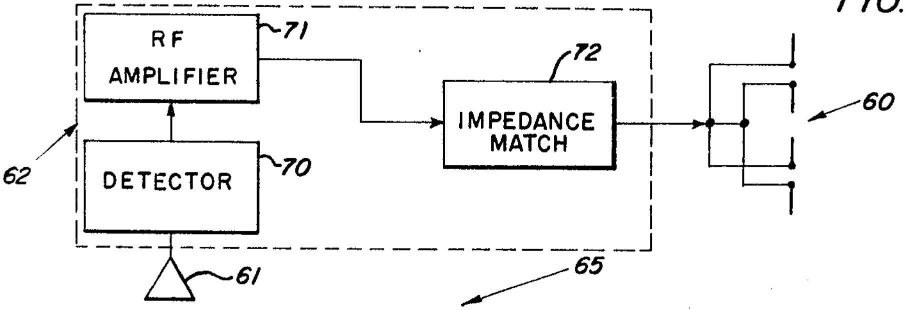

RF AMPLIFIER
71
72
IMPEDANCE MATCH
60
62
DETECTOR
70
61
ANTENNA HORN

FIG. 5

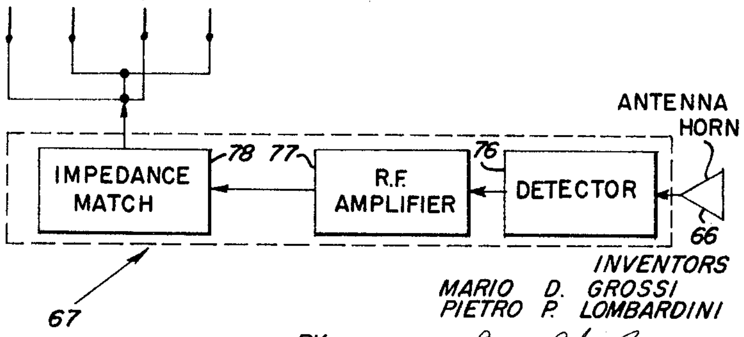

65
ANTENNA HORN
IMPEDANCE MATCH
78
77
R.F. AMPLIFIER
76
DETECTOR
66
67

INVENTORS
MARIO D. GROSSI
PIETRO P. LOMBARDINI

BY Donald Brown
AGENT

May 20, 1969 M. D. GROSSI ETAL 3,445,844

TRAPPED ELECTROMAGNETIC RADIATION COMMUNICATIONS SYSTEM

Original Filed Sept. 11, 1963 Sheet 3 of 4

FIG. 6

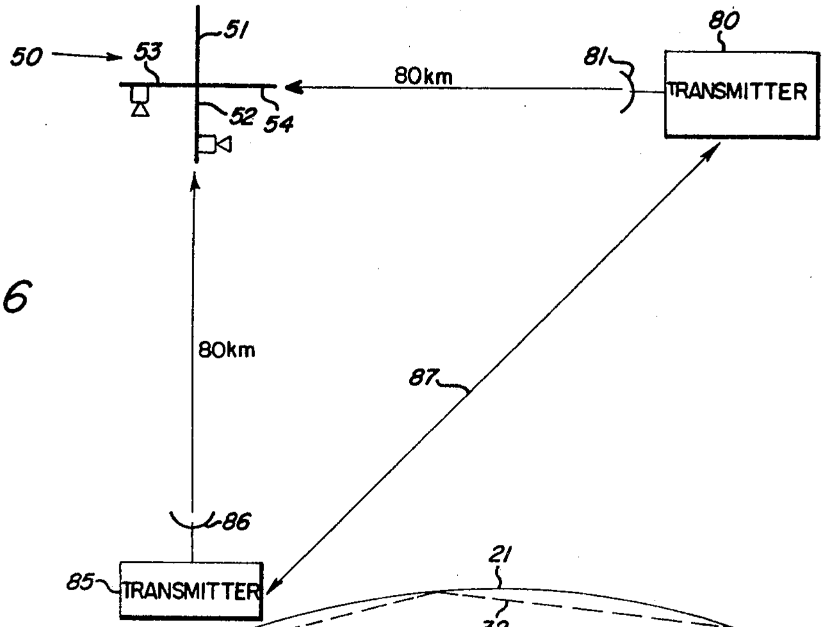
50
53
51
52
54
80km
81
80
TRANSMITTER
80km
87
86
85
TRANSMITTER

FIG. 8

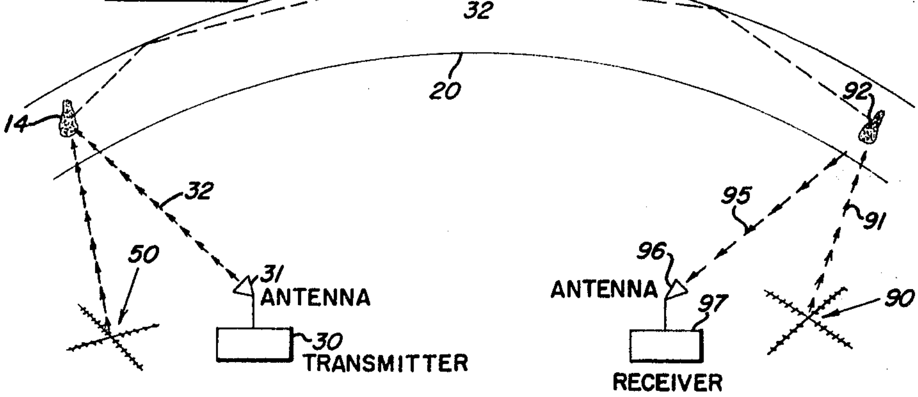
21
32
20
14
92
32
95
91
50
96
31
ANTENNA
ANTENNA
97
90
30
TRANSMITTER
RECEIVER

FIG. 9

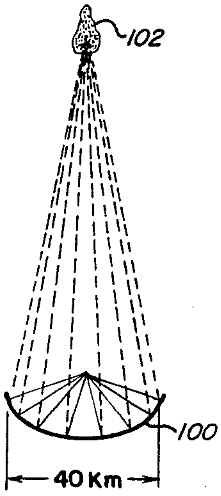
102
100
40 Km

INVENTORS
MARIO D. GROSSI
PIETRO P. LOMBARDINI

BY Donald Brown
AGENT

FIG. 7

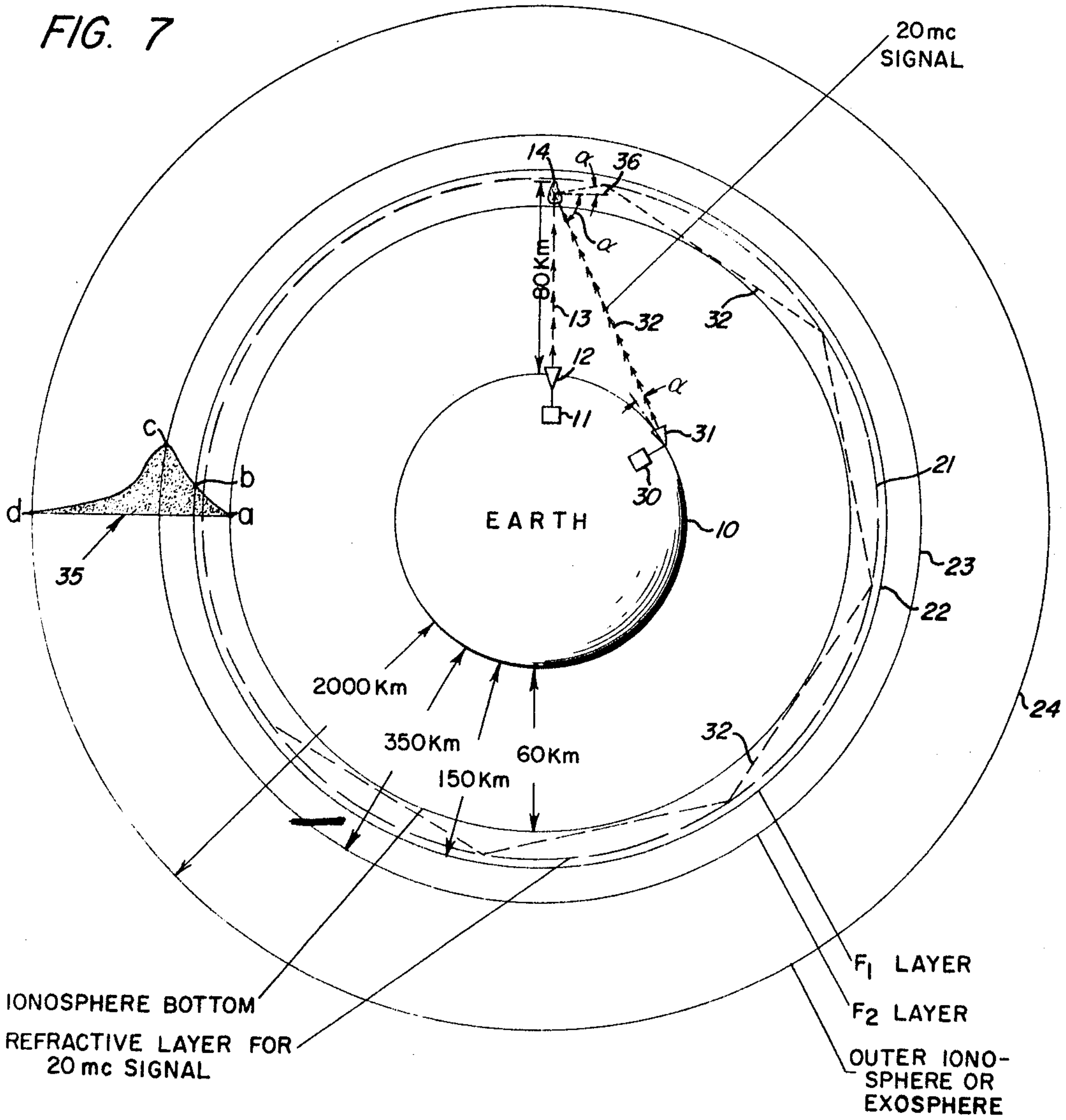
20mc SIGNAL
14
α
36
α
80Km
13
32
32
12
α
11
31
30
EARTH
10
21
23
22
24
c
b
d
a
35
2000Km
350Km
60Km
150Km
32
F1 LAYER
F2 LAYER
OUTER IONO-SPHERE OR EXOSPHERE
IONOSPHERE BOTTOM
REFRACTIVE LAYER FOR 20 mc SIGNAL

INVENTORS
MARIO D. GROSSI
PIETRO P. LOMBARDINI

BY Donald Brown
AGENT

United States Patent Office

3,445,844

Patented May 20, 1969

3,445,844

TRAPPED ELECTROMAGNETIC RADIATION COMMUNICATIONS SYSTEM

Mario D. Grossi, Cambridge, Mass., and Pietro P. Lombardini, Philadelphia, Pa., assignors to Raytheon Company, Lexington, Mass., a corporation of Delaware

Continuation of application Ser. No. 308,172, Sept. 11, 1963. This application Jan. 11, 1968, Ser. No. 697,244

Int. Cl. H04b *7/00*

U.S. Cl. 343—100 **57 Claims**

This invention relates to communication systems and more particularly to the use of trapped electromagnetic radiation to provide an over the horizon long distance communications network or system.

This application is a continuation of application Ser. No. 308,172 filed Sept. 11, 1963, now abandoned.

Ever since the advent of the radio, various attempts have been made to provide a long range communications system. During the World War I era attempts were made to obtain a reliable world wide communications system utilizing very low frequencies, but due to atmospheric noises and bandwidth limitations, more desirable techniques were required. In more recent years, radio propagation over long distances has been attempted by systems employing "one hop" or "bounce" reflective techniques for scattering V.H.F., U.H.F., and S.H.F. signals. These particular systems have required the use of large amounts of power in order to obtain reception at distant points. Additionally, scattering techniques have suffered from the occurrence of atmospheric instabilities which have caused periodic fading of the transmitted signals. Although scattering techniques have been applied successfully over relatively short distances, such as between the continent of Europe and islands in the Mediterranean, and additionally in Alaska and Northern Canada, long distance communications have not been possible.

Other global communication systems not relying on ground based links have also been attempted. For example, the Telstar orbiting satellite and the orbiting satellite communications system have been proposed as a solution to this problem. Although Telstar has proved that a global communication system is feasible, utilizing an orbiting satellite or a group of orbiting satellites, these systems are quite costly, particularly due to the requirement of at least forty orbiting satellites in order to obtain coverage of the entire world. Additionally, problems of reliability of the repeater equipment utnlized in the satellites will necessitate that additional repeater satellites be periodically placed in orbit as a substitute for an inoperable satellite. Additionally, these satellite systems rely on basically a "one hop" line-of-sight technique between two points, and thus forty satellites are required as a minimum for complete world wide coverage.

Further, "one hop" reflective long distance communication systems have been proposed utilizing a ground created reflective zone or layer of an ionized atmosphere. This reflective layer could be obtained using seeding techniques or possibly by the injection of neutrons to form a reflective cloud in the atmosphere. In spite of the fact that communications have been successfully attempted utilizing seeding for "one hop" point-to-point over the horizon reflective communications links, seeded ionized clouds have been difficult to control, particularly due to the sun's adverse effect. Although all the aforementioned systems have had some success, either their cost or complexity has produced a requirement for a more inexpensive and reliable system to obtain world wide communications.

Accordingly, it is the principal object of this invention to provide a new, useful, and improved highly reliable world wide communications system.

Another object of this invention is to provide an over

the horizon communications system which utilizes a waveguide duct in the ionosphere to trap electromagnetic radiation.

Still another object of this invention is to create an electromagnetically generated ionized cloud to provide a means for injecting electromagnetic radiation into an ionospheric waveguide duct.

A further object of this invention is to provide a means for extracting electromagnetic radiation trapped within an ionospheric waveguide duct.

In accordance with the communications system of this invention, electromagentic radiation is artificially injected into and trapped within an atmospheric waveguide duct. The trapped radiation then propogates within said atmospheric waveguide duct. The trapped radiation then propogates within said atmospheric waveguide duct around the earth. Devices positioned around the earth's periphery are utilized to receive a portion of trapped radiation either by detecting radiation scattered by natural or artificially generated scaters positioned within the waveguide duct.

Other features and objectives of the communications system of this invention will become apparent from the following description taken in conjunction with the following drawings, wherein:

FIG. 1 illustrates schematically a communications system following the principals of this invention including means for generating an artificial scatterer to inject electromagnetic radiation into an ionospheric waveguide duct, and means for receiving a portion of said radiation reflected by natural scatterers.

FIG. 2 illustrates a top plan view of a configuration of an antenna for generating an artificial scatterer.

FIG. 3 is a side elevational view of one of dipole elements of the antenna of FIG. 2.

FIG. 4 illustrates schematically the transmitter and the dipole antenna arrangement for the legs numbered **53** and **54** of the antenna of FIG. 2.

FIG. 5 illustrates schematically the transmitter and the dipole antenna arrangement for the legs numbered **51** and **52** of the antenna of FIG. 2.

FIG. 6 illustrates schematically a means for properly phasing the output radiation from the antenna of FIG. 2 in order to generate an artificial scatterer positioned within the ionosphere.

FIG. 7 is a pictorial representation of the earth surrounded by ionospheric layers for the purpose of explaining the presence of ionospheric waveguide ducts.

FIG. 8 is a schematic diagram of the communications system of this invention including means for generating an artificial scatterer to extract electromagnetic radiation from an ionosphereic waveguide duct; and

FIG. 9 is a side schematic view of an ellipsoid antenna which could be used to generate an artificial scatterer within the ionosphere.

FIG. 1 illustrates a long distance communications system following the principles of this invention and which comprises a transmitter **11** disposed at a point on the earth **10**. The transmitter **11** will be described in conjunction with an explanation of FIGS. 4 and 5. Transmitter **11** provides electromagnetic energy to antenna **12**. Antenna **12** then focuses this energy or radiation, shown as arrows **13**, to form an ionized cloud or scatterer **14**, within the ionospheric region of the atmosphere. A suitable antenna for generating the ionized cloud **14** will be described at a later time in conjunction with the explanation of FIGS. 2, 3, 4 and 5 of the accompanying drawings. The ionized cloud **14** is shown positioned slightly above the ionospheric bottom layer **20**, but substantially below the F_2 layer **23** of the ionospheric region.

The ionized cloud **14** is believed to be generated in the following manner: electrons that already exist in the ionospheric region of the upper atmosphere are accelerated by electromagnetic radiation which is focused from the ground. These electrons are accelerated by the radiation to a degree such that their kinetic energy reaches the level needed for the occurrence of ionizing collisions. Cloud **14** is a result of these ionizing collisions. Scattering from this cloud takes place due to the discontinuity between this zone of enhanced ionization and the surrounding medium. The shape and the properties of the scatterer can be adjusted by controlling the properties of the antenna used, the amount of power provided by the antenna, and the phase of the energy emitted by various portions of the antenna. For example, the preferred embodiment of this invention utilizes a transmitter **11** and antenna **12** operating at a frequency of 50 megacycles, and which provides 26 megawatts of peak power and 260 kilowatts of average power. Antenna **12** focuses the energy to produce a cloud **14** approximately 80 kilometers above the earth's surface. The ionized cloud formed has a mean width of approximately 15 meters and a length of approximately 60 meters. Due to the focusing, the ionized cloud has a shape which comes to a point at its greatest distance from the earth's surface and tapers in the manner of a cone at its widest point. The ionized cloud of this invention also has a variation of the electric field in the cloud versus the height from the antenna which is positive. The antenna used is in a near-field arrangement. In other words, with increasing distance in the resulting cloud **14** there is an increasing field. This does not happen with an unfocused antenna of the far-field type. An unfocused antenna has a radiation pattern which spreads at greater distances from the antenna and its electric field decreases with increasing range. In our case $dE/dh>0$ where E is the E field vector and h is the distance from the antenna above the earth's surface. It is to be noted that the life of our scatterer or cloud persists essentially as long as the ground transmitter is on. This constitutes a basic advantage with respect to any other suggested artificial scatterer, such as that which could be produced by seeding. Accordingly, this system utilizing an ionized cloud which is generated from the earth and which is controllable from the earth is highly superior to any technique which is completely out of the control of the operators.

Referring once again to FIG. 1, a transmitter **30** is shown for providing a communications signal which is to carry information from a point somewhere on the earth's periphery which is in proximity to transmitter **11**. The transmitter **30** could be any of the common A.M. communication transmitters, such as those which operate in the M.F., V.H.F. and H.F. frequency bands. The transmitter **30** provides this communication signal to antenna **31**. The antenna **31** is of a directional type in order that the communication signal can be directed towards the ionized cloud or artificial scatterer **14**. Directional antennas, such as rhombic or Yagi, could be utilized for this purpose. Assume now that the transmitter **30** provides a 20 megacycle communications signal having a 500 kilowatt peak output power to the antenna **31**. This signal will then be directed as shown by the arrowed lines **32**, and be refracted off the ionized cloud or scatterer **14**. This signal **32** will be directed through the ionospheric waveguide duct, due to a refractive electron layer **21** which curves spherically around and is concentric with the earth's periphery. This refractive layer **21** causes the communication signal **32** to be trapped within an ionospheric waveguide duct which has the refractive layer **21** as its upper boundary, substantially the ionospheric bottom as its lower boundary, and which is concentric with the earth's periphery. The explanation of the presence of this refractive layer and the duct will be given at a later time in conjunction with a description of FIG. 7.

Assuming now a receiving station, such as receiver **40** and antenna **41**, located at a site which is at a long distance from transmitter **30**, is required to receive information transmitted by transmitter **30**. Communication reception at the receiving station is made possible by natural scatterers within the ionospheric layer, such as produced by natural discontinuities in the electron density within the ionosphere. These natural scatterers cause power leakages in the flow of the communications signal **32** through the ionospheric waveguide duct. These power leakages are shown as arrows **42**. By utilizing the antenna **41** along with a receiver **40**, it is thus possible to detect the power leakages due to scattering caused by natural scatterers within the ionosphere. This is shown diagrammatically as arrow **43** being received by antenna **41**. The antenna **41** could be of the Yagi type. The recetiver **40** may be of the double conversion superheterodyne type utilizing two local oscillators and two I_F stages. Thus, it is seen that once a communications signal is injected into the ionosphere and is then trapped in the waveguide duct of the ionopshere, a portion of this trapped signal can then be received after traveling over long distances. It is to be noted that this type of communications link does not depend on local minima or maxima in the ionopshere electron density profile. Furthermore, it is to be noted that this system does not rely upon the reflection of a signal off an ionized cloud and back down to a second location on the earth in order to obtain "one hop" point-to-point communications. Rather, the communications system presented here utilizes ionospherically trapped electromagnetic radiation in order to obtain point-to-point over the horizon communications.

Referring now to FIGS. 2, 3, 4, 5 and 6 for the description of a transmitter and antenna system shown as transmitter **11** and antenna **12** in FIG. 1. FIG. 2 illustrates the configuration of a Mills Cross antenna for generating the ionized cloud or artificial scatterer **14**. This antenna is of that type which is common to the radio astronomy field. A description of this type of antenna can be found in the book "Radio Astronomy," page 75, authored by J. L. Pawsey and R. M. Bracewell, published by Oxford at the Clarendon Press, under the date of 1955 in Oxford, England. An arrangement of this antenna is shown in a top plan view in FIG. 2. This Mills Cross antenna **50** is shown having two legs **51** and **52** and two legs **53** and **54**. The width of each of the legs is 12 meters and the length of each leg is 19 kilometers. The total antenna of this invention compriess 104,000 dipoles. A side elevational view of one of the dipoles is shown in FIG. 3 and has a radiating portion **58** which when mounted is parallel to the earth's surface and a vertical portion **59** which is used to mount each of these dipole antennas in the earth's surface. The horizontal radiating portion **58** is designed with a radiating surface equal to $\lambda/_2$ and the mounting portion **59** has a dimension of $\lambda/_4$. Each of the legs **51**, **52**, **53** and **54** have a total of 26,000 dipoles mounted four across to obtain the configuration shown in FIG. 2. Each of the legs has 6500 rows of four dipole antennas each.

The Mills Cross antenna **50** has the rows of dipoles comprising legs **53** and **54** connected in the same manner. For the purposes of explanation, a four dipole row in leg **53** will be described. This four dipole row is shown as **60** and has a transmitter **62** providing power to be radiated by each of the dipoles. A receiving antenna horn **61** is shown coupled to the transmitter **60** for the purposes of phasing the output radiation from each of the dipole antennas shown in **60**. The description of the phasing of the radiation of this antenna will be given in conjunction with an explanation of FIG. 6.

In order to show the connection of the dipole antennas of the legs **51** and **52**, a four row dipole antenna configuration **65** is singled out for purposes of description. Dipole antenna row **65** is shown coupled to a transmitter **67** which has an antenna horn **66** which is also used for the purposes of phasing the signal to be radiated by the dipole row **65**. Each of the transmitters of the Mills Cross antenna configuration provides a final stage power output of approximately one kilowatt peak power or 10 watts of average power. Since there are a total of 26,000 transmitters in this antenna configuration **50**, each providing one kilowatt, a total output power of approximately 26 megawatts peak power is obtained.

Referring now to FIG. 4 for a description of the transmiter **62**, a detector **70** is shown coupled to the antenna horn **61** for the purpose of receiving a phasing signal to be retransmitted. The signal to be retransmitted is applied to an R.F. amplifier **71** in order to amplify the detected signal to be retransmitted by the dipole configuration shown as **60**. The R.F. amplifier final stage is of the normal tetrode type of amplifier. R.F. amplifier **71** provides an amplified signal which is coupled through an impedance matching device **72** to feed the dipole antenna configuration **60**, as schematically shown in FIG. 4. The impedance matching device is utilized in order to couple the R.F. antenna to the dipole configuration **60**.

Referring now to FIG. 5, there is disclosed the transmitter **67**, having a detector **76**, R.F. amplifier **77**, and an impedance matching device **78** coupled between an antenna horn **66** and the dipole radiating configuration **65**. The apparatus of FIG. 5 is of the same type and operates similarly as does the apparatus of FIG. 4.

Referring now to FIG. 6, there is shown a scheme for phasing radiations from each of the dipole antennas of the Mills Cross antenna configuration **50**. Although it is possible to pre-set the phase of each of the transmitted signals which emanate from the dipole antennas by individually adjusting the phase of the signal provided by each of the transmitters of the antenna configuration **50**, a more effective technique for obtaining the proper phase relationship is shown in FIG. 6. This simplified technique is used to obtain the phase relationship to produce a cloud or artificial scatterer approximately 80 kilometers above the earth's surface. This technique utilizes two transmitters, **80** and **85**, which are slaved together by way of a ground link **87**. The transmtiters, **80** and **85**, simultaneously transmit a signal which emanates in the same phase relationship from each of these transmitters. By placing transmitter **80** perpendicular to and at a distance of approximately 80 kilometers from the legs **51** and **52** and transmitting this signal by way of a parabolic antenna configuration, and by simultaneously positioning transmitter **85** perpendicular to and at a distance of 80 kilometers from legs **53** and **54**, and transmitting a signal from a parabolic antenna **86**, it is possible to phase the signals radiated by the dipoles of the antenna **50** in order to produce the ionized cloud **14**. Thus, for example, by amplitude modulating a microwave frequency signal with a 50 megacycle per second signal, and by detecting this modulation with a detector, such as detector **70** or **76** of FIGS. 4 and 5, and then amplifying this 50 megacycle detected signal, the proper phase relationship for producing the ionized cloud is obtained. The proper phase is obtained since the phase of arrival of the 50 megacycle modulated microwave signal differs from point-to-point along the legs of the antenna configuration **50** in the same phase relationship that is required to recompose the field at 80 kilometers of height.

FIG. 7 is a pictorial representation of the earth surrounded by the ionospheric layers for the purpose of explaining the ionospheric waveguide ducts. For purposes of obtaining a reference from the earth's surface there is presented a pictorial diagram (not to be considered as drawn to scale) showing an ionospheric bottom layer **20** approximately 60 kilometers above the earth's surface, an F_1 layer substantially 150 kilometers above the earth's surface, an F_2 layer approximately 350 kilometers above the earth's surface, and outer ionosphere or exosphere about 2000 kilometers above the earth's surface. The ionosphere has the shape of a spherical shell and is concentric with the earth's surface.

Referring now to the portion of FIG. 7 numbered **35**, this stippled area represents the density of the electrons in the various regions of the ionosphere. For example, point *a* represents an electron density of approximately 10^3 electrons per cubic centimeter, which occurs at the ionospheric bottom. Point *b* represents the electron density at the F_1 layer, which is approximately equal to 10^5 electrons per cubic centimeter. Point *c* depicts the maximum electron density, which is presumed to exist at the F_2 layer of the ionosphere. At this point, an electron density of approximately 10^6 electrons per cubic centimeter is believed to exist. Point *d* on the tail end of the exponentially decaying region between points *c* and *d* represents an electron density of approximately 10^3 electrons per cubic centimeter which is considered to exist at the exposure.

Assume now that a transmitter, such as **11**, and an antenna **12**, which was previously disclosed with relation to FIGS. 1–6, generates an artificial scatterer **14** at approximately 80 kilometers above the earth's surface. Further, assume that a transmitter **30** and an antenna **31** directs a 20 megacycle communications signal in such a manner that the signal leaves the earth at an angle α of approximately 10° so that it is injected at an angle of α degrees into an ionospheric waveguide duct which lies between the bottom layer **20** and the F_2 layer **23**. The upper limit of the ionspheric waveguide duct, shown as the refractive layer **21** within the ionosphere, is a function of the frequency of the transmitted communications signal and the ionospheric conditions. The index of refraction of the ionosphere refractive layer is, among other things, a function of the local electron density of the ionospheric region and of the frequency of the transmitted communications signal. When the index of refraction approaches 0, refractive bending of a ray toward the earth's surface will take place. However, a trapped ray will not land on the earth's surface. For a 20 megacycle communications signal, a refractive layer **21** theorized at 130 kilometers, comprises the upper limit of the wave guidance duct. The lower limit of the ionospheric waveguide duct is a function of the angle α. In the situation required for trapping of the signal **32**, the lower limit is substantially bounded by the ionospheric bottom for an α which permits trapping. Therefore, for a 20 megacycle signal, the ionospheric waveguide duct will lie in the region between substantially 60 kilometers at its lower limit to 130 kilometers at its upper limit. For other frequencies, the waveguide duct will vary in accordance with the frequency of the communications signal.

Referring once again to the communications signal **32**, it can be seen that this signal **32** will be refracted from the artificial scatterer **14**, and be injected into the ionospheric duct, described previously. The signal after entering the duct will be trapped, due to the signal **32** being refracted periodically at points along the upper shell-like refractive layer **21**.

FIG. 8 is a schematic diagram of the communication system of this invention including a second artificial scatterer, which is placed in the ionospheric duct in order to extract a comunications signal trapped within the duct. A transmitter **30** and an antenna **31** is shown providing a communications signal **32** which is refracted by a scatterer **14** generated by a Mills Cross antenna **50**. This signal **32** is then periodically refracted by a refractive layer **21** and travels parallel to the earth's surface. A second Mills Cross type antenna emitting radiation, shown as arrowed lines **91**, places a second artificial scatterer **92** within the ionospheric duct which lies substantially between the ionospheric bottom **20** and the refractive layer **21**. This cloud **92** intercepts the communications signal **32** and reflects a major portion thereof. This is shown as arrowed lines **95** which are received by an antenna **96** and a receiver **97**. Although this particular embodiment is more costly than that which is shown in FIG. 1, due to the use of two artificial scatterers **14** and **92**, this embodiment will provide a communications system which will be operable over long distances at higher data rates. It is to be noted that this two scatter system is not of the "one-hop" reflective type. This system does not require that the artificial scatterers **14** and **92** be within line-of-sight of each other in order that there be over the horizon communications.

Referring now to FIG. 9, an alternate embodiment of an antenna is shown for producing an ionized cloud or scatterer, such as cloud **102**. The antenna **100** is of the ellipsoidic variety and is designed with a diameter of 40 kilometers to radiate and focus energy to produce the cloud **102**.

Although the preferred embodiment shows the use of separate antennas and transmitters to produce the cloud and to provide the communications signal, it is to be understood that a single station comprising a dual purpose antenna and transmitter could be used. Furthermore, it is to be understood that although a 20 megacycle communication signal is used in the preferred embodiment, other frequencies such as M.F., H.F., and V.H.F. frequencies of between approximately a few megacycles to 70 megacycles could be used. For example, recent tests have shown that communications between Massachusetts and Italy are possible utilizing this technique. Instead of generating an artificial scatterer, ionized meteor trails were utilized as scatterers to inject a 23 megacycle signal into an ionospheric waveguide duct. Receivers positioned in Italy detected a portion of the signal trapped within the duct. Additionally, this technique could be applied to satellite-borne receivers orbiting within the ionospheric waveguide duct or between ground stations and satellites. In addition, it is possible to use the trapped radiation technique of this invention in other regions of the atmosphere, such as in the troposphere. A tropospheric communications system would permit the propagation of frequencies in the V.H.F., U.H.F. and S.H.F. bands. Accordingly, it is desired that this invention not be limited except as defined by the appended claims.

We claim:

1. A communications system comprising:

means for generating and controlling an ionized cloud in the ionosphere, said means including first and second transmitters for simultaneously transmitting signals in the same phase relationship, ground link means for slaving said transmitters together, and antenna means for focusing on and accelerating ionospheric electrons with the electromagnetic radiation signals from said transmitters to produce ionizing collisions;

transmitting means for transmitting a communications signal directed at said ionized cloud, said cloud injecting said communications signal into and propagating said communications signal within an ionospheric signal trapping waveguide duct encircling the earth; and

means for receiving at least a portion of said communications signal trapped in the duct.

2. A system comprising:

means for transmitting a signal;

means for electromagnetically generating a scatterer within a given zone of the ionosphere, said scatterer injecting at least a portion of said signal into and causing said portion to propagate within an ionospheric signal trapping waveguide duct encircling the earth; and

means for receiving at least a portion of said signal trapped within said waveguide duct.

3. A system comprising:

means for transmitting a signal into a given zone of the atmosphere;

means for generating a scatterer in a given zone of the atmosphere, said scatterer causing said signal to be injected into and propagated within an atmospheric signal trapping duct to encircle the earth, and

means for detecting at least a portion of said signal trapped within said duct.

4. A system comprising:

means for providing a signal having a frequency below approximately 7 megacycles;

means for injecting at least a portion of said signal into a trapping zone of the ionosphere encircling the earth and defined by the ionospheric bottom layer and a refractive layer which is a function of the frequency of said signal and for causing said portion to propagate within said trapping zone around the earth; and

means for detecting at least a portion of said injected signal.

5. A communications system comprising:

means for providing a signal;

means for electromagnetically generating a scatterer to inject said signal into and to propagate said signal within a given signal trapping zone of the atmosphere encircling the earth; and

means for detecting at least a portion of said injected signal.

6. A communications system in accordance with claim **5** wherein said means for electromagnetically generating a scatterer includes an antenna for focusing on and accelerating ionospheric electrons with electromagnetic radiation signals to produce ionizing collisions.

7. A communications system in accordance with calim **6** wherein said antenna comprises a Mills Cross antenna having first and second perpendicular legs and includes means for phasing said antenna to produce said scatterer comprising first and second transmitters for simultaneously transmitting signals in the same phase relationship, each of said transmitters being perpendicular to and spaced apart from a different one of said legs, and ground link means for slaving said transmitters together.

8. A communications system comprising:

transmitting means for transmitting a communications signal;

transmitting means for electrromagnetically generating a scatterer in a given signal trapping zone in the ionosphere;

an antenna included in said scatterer generating means said scatterer being provided to inject said communications signal into and to propagate said communications signal within a given signal trapping zone of the ionosphere encircling the earth;

means for focusing on and acceleratnig ionospheric electrons with electromagnetic radiation signals to produce ionizing collisions;

means for phasing said antenna to produce said scatterer, said phasing means including first and second transmitters for simultaneously transmitting signals in the same phase relationship;

ground link means for slaving said transmitters together; and

means for receiving at least a portion of said communications signal which has been injected into and propagated within the given signal trapping zone.

9. A system comprising:

means for generating and controlling an ionized cloud including transmitter means for producing signals to accelerate ionospheric electrons and for stopping producing said signals to cause the removal of said cloud;

means for transmitting a signal directed at said cloud, said cloud injecting said signal into and propagating said signal within an ionospheric trapping waveguide duct encircling the earth; and

means for receiving at least a portion of said signal injected into said duct.

10. A system in accordance with claim **9** wherein said means for generating a cloud includes a near-field antenna producing a positive gradient electric field within said cloud at increasing distances from said antenna.

11. A communications system comprising:

means for providing a signal;

means for electromagnetically generating a scatterer within a region of the ionosphere below the ionospheric F_2 layer, said scatterer providing means for the injection and the porpagating said signal in a trapping zone of the ionosphere defined by the ionosphere bottom layer and a refractive layer which is a function of the frequency of said signal; and

means for receiving at least a portion of said trapped signal.

12. A communications system comprising:

means for transmitting a communications signal;

means for electromagnetically generating and controlling a scatterer within a given region of the ionosphere below the ionosphere F_2 layer, said scatterer being provided to inject said communication signal into and to propagate said communications signal within a trapping zone of the ionosphere defined by the ionospheric bottom layer and a refractive layer which is a function of the frequency of said communications signal;

said scatterer generating means including first and second transmiters for simultaneously transmitting signals in the same phase relationship, ground link means for slaving said transmitters together, and a near-field antenna for producing a positive gradient electric filed within the scatterer at increasing distances from the antenna; and

means for receiving at least a portion of the communications signal in said trapping zone.

13. A system in accordance with claim **11** including a second electromagnetically generated scatterer positioned within the zone of the ionosphere defined by the ionospheric bottom layer and said refractive layer, said scatterer providing means for extracting at least a portion of said signal trapped within said ionosphere.

14. A communications system comprising:

means for providing a signal having a frequency between a few megacycles to about 70 megacycles;

means for electromagnetically generating a scatterer within a given region of the ionosphere including transmitter means for producing signals to accelerate ionospheric electrons, said scatterer providing for the trapping of and the propagating of said signal within a given region of the ionosphere defined by the ionospheric bottom layer and a refractive layer which is a function of the frequency of said signal;

means for electromagnetically generating a second scatterer within a given region of the ionosphere to extract at least a portion of said signal including transmitter means for producing signals to accelerate ionospheric electrons; and

means for detecting at least a portion of said extracted signal.

15. A communications system comprising:

means for generating and controlling an ionized cloud in the ionosphere, said means including one transmitting means for simultaneously transmitting signals in a suitable phase relationship to achieve focusing of the radio frequency generated by the transmitting means, means for slaving the signals from said transmitting means and antenna means for focusing on and accelerating ionospheric electrons with the electromagnetic radiation signals from said transmitting means to produce ionizing collisions;

another transmitting means for transmitting a communications signal directed at said ionized cloud, said cloud injecting said communications signal into and propagating said communications signal within an ionospheric signal trapping waveguide duct encircling the earth; and

means for receiving at least a portion of said communications signal trapped in the duct.

16. A system comprising:

means for transmitting a signal;

means for electromagnetically generating a scatterer for injecting at least a portion of said signal into a trapping zone of the ionosphere encircling the earth, said zone defined by a lower boundary which is a function of the angle of injection of the signal into the zone and a upper refractive layer boundary which is a function of the frequency of said signal and for causing said portion to propagate within said trapping zone around the earth without any lower boundary reflections; and

means for receiving at least a portion of said signal trapped within said zone.

17. A system comprising:

means for transmitting a signal;

means for electromagnetically generating a scatterer for injecting at least a portion of said signal into a trapping zone of the ionosphere, said zone limited by an upper refractive layer which is a function of the frequency of said signal and for causing said portion to propagate within said trapping zone around the earth with reflections only off of the upper refractive layer; and

means for receiving at least a portion of said signal trapped within said zone.

18. A system comprising:

means for transmitting a signal;

a scatterer injecting at least a portion of said signal into a trapping zone of the ionosphere encircling the earth, said zone defined by a lower boundary which is a function of the angle of injection of the signal into the zone and an upper refractive layer boundary which is a function of the frequenncy of said signal and for causing said portion to propagate within said trapping zone around the earth without any lower boundary reflections; and

means for detecting at least a portion of said injected signal.

19. A system comprising:

means for transmitting a signal;

a scatter injecting at least a portion of said signal into a trapping zone of the ionosphere, said zone limited by an upper refractive layer which is a function of the frequency of said signal and causing said portion to propagate within said trapping zone around the earth with reflections only off of the upper refractive layer; and

means for detecting at least a portion of said injected signal.

20. A communications system comprising:

means for providing a signal;

means for electromagnetically generating a permanent and controllable scatterer to inject said signal into and to propagate said signal within a given signal trapping zone of the ionosphere encircling the earth, and

means for detecting at least a portion of said injected signal.

21. A method of communicating comprising the steps of:

electromagnetically generating a scatterer in a given zone of the atmosphere;

transmitting a signal into said scatterer in such manner that said signal is injected into and propagated within an atmospheric signal trapping duct encircling the earth; and

detecting at least a portion of said signal trapped within said duct.

22. A method of communicating comprising the steps of:

electromagnetically generating a scatterer within a given zone of the ionosphere;

transmitting a signal into said scatterer in such manner that at least a portion of said signal is injected into and propagated within an ionospheric signal trapping waveguide duct encircling the earth; and

receiving at last a portion of said signal trapped within said waveguide duct.

23. A method of communicating comprising the steps of:
- transmitting a communications signal;
- injecting at least a portion of said signal into a signal trapping zone of the atmosphere encircling the earth, said zone defined by a lower boundary which is a functon of the angle of injection of the signal into the zone and an upper shell-like refractive layer boundary which is a function of said signal, said signal being periodically refracted at points along the upper shell-like refractive layer without any lower boundary reflections as it is propagated approximately parallel to the earth's surface; and
- receiving at least a portion of the signal propagating within said trapping zone.

24. A method of communicting using trapped electromagnetic radiation comprising the steps of:
- transmitting a signal having a frequency below approximately 70 megacycles;
- injecting at least a portion of said signal into a trapping zone of the ionosphere encircling the earth and defined by the ionosphere bottom layer and a refractive layer which is a function of the frequency of said signal;
- propagating said signal within said trapping zone around the earth; and
- detecting at least a portion of said propagated signal.

25. A method of communicating using ionizing atmospheric collisions comprising the steps of:
- electromagnetically generating a scatterer in a given zone of the atmosphere;
- providing a communications signal which said scatterer injects into and propagates within the given signal trapping zone encircling the earth; and
- detecting at least a portion of said propagated signal.

26. A method in accordance with Claim 25 wherein the step of electromagnetically generating the scatterer includes:
- focusing electromagnetically radiation signals on atmospheric electrons within said given zone; and
- accelerating the electrons with the electromagnetic radiation signals to produce ionizing collisions.

27. A method in accordance with Claim 26 wherein the steps of focusing and accelerating include:
- simultaneously transmitting identically phased signals;
- slaving said phased signals together; and
- phasing an antenna to achieve the focusing and accelerating of the electrons to produce the ionizing collisions.

28. A method of communicating comprising tne steps of:
- generating an ionized cloud with electromagnetic radiation signals;
- controlling said cloud by said signals;
- transmitting a communication signal directed at said cloud such that said communication signal is injected into and propagated within an ionospheric signal trapping waveguide duct encircling the earth; and
- receiving at least a portion of said communication signal which has been injected into and propagated within said duct.

29. A method of communicating comprising the steps of:
- electromagnetically generating a scatterer within a region of the ionosphere below the ionospheric F_2 layer;
- transmitting a communication signal into said scatterer to inject into and propagate said signal within a signal trapping zone of the ionosphere defined by the ionospheric bottom layer and a refractive layer which is a function of the signal frequency; and
- receiving at least a portion of said trapped signal.

30. The method of claim 29 wherein said receiving step includes:
- electromagnetically generating a second scatterer positioned within the zone of the ionosphere defined by the ionospheric bottom layer and said refractive layer in order to provide means for extracting at least a portion of said trapped signal.

31. A method of communicating comprising the steps of:
- electromagnetically generating a scatterer within a given region of the ionosphere;
- transmitting a communications signal having a frequency between a few megacycles to about 70 megacycles into said scatter causing the signal to be trapped and propagated within the given region of the ionosphere defined by the ionospheric bottom layer and refractive layer which is a function of the frequency of the signal;
- electromagnetically generating a second scatterer within the given region of the ionosphere in order to permit extraction of at least a partion of said signal; and
- detecting at least a portion of said extracted signal.

32. A method of communicating comprising the steps of:
- electromagnetically generating an ionized cloud in the ionosphere;
- controlling said cloud by simultaneously transmitting signals in a suitable phase relationship to achieve focusing of the radio frequency generated, slaving the transmitted signals together and focusing on and accelerating ionospheric electrons with the electromagnetic radiation signals to produce ionizing collisions;
- transmitting a communications signal directed at said ionized cloud to inject into and propagate said communications signal within an ionospheric signal trapping waveguide duct encircling the earth; and
- receiving at least a portion of said communications signal trapped in the duct.

33. A method of communicating comprising the steps of:
- electromagnetically generating a scatterer within a given region of the ionosphere;
- transmitting a communications signal into said scatterer to inject at least a portion of said signal into and propagate it within a trapping zone of the ionosphere defined by a lower boundary which is a function of the angle of injection of the signal into the zone and an upper refractive layer boundary which is a function of the frequency of said signal, the propagation occurring without any lower boundary reflections; and
- receiving at least a portion of said signal trapped within said zone.

34. A method of communicating comprising the steps of:
- electromagnetically generating a scatterer within a given region of the ionosphere;
- transmitting a communications signals into said scatterer to inject at least a portion of said signal into and propagate it within a trapping zone of the ionosphere limited by an upper refractive layer which is a function of the frequency of said signal, the propagation occurring with reflections only off of the upper refractive layer; and
- receiving at least a portion of said signal trapped within said zone.

35. A method of communicating comprising the steps of:
- electromagnetically generating a permanent and controllable scatterer in a given zone of the atmosphere;
- directing a communications signal into said scatterer to inject said signal into and to propagate said signal within the given zone of the atmosphere; and
- detecting at least a portion of said injected signal.

36. A communications system as set forth in claim **15** wherein said antenna is employed in a near-field arrangement.

37. A communications system as set forth in claim **36** where:

said antenna comprising a Mills Cross antenna having first and second sets of perpendicular legs; and

said one transmitting means comprises first and second transmitters for simultaneously transmitting signals in the same phase relationship, each of said transmitters being perpendicular to and spaced apart from a different one of said legs.

38. A communications system as set forth in claim **37** wherein:

each of said antenna legs has a width of approximately twelve meters and a length of approximately nineteen kilometers;

said antenna has a horizontal radiating surface equal to one half wavelength and a mounting portion of one quarter wavelength; and

said first and second transmitters each being spared approximately 80 kilometers from the corresponding legs of said antenna and each having a parabolic transmitting antenna.

39. A communications system as set forth in claim **38** wherein:

said ionized cloud is located approximately 80 kilometers above the earth's surface, has a mean width of approximately 15 meters, a length of approximately 60 meters and tapers in a cone-shaped fashion from its widest portion to a point at its greatest distance from the earth's surface.

40. A communicating system as set forth in claim **15** wherein:

said another transmitting means transmits a communications signal having a frequency between a few megacycles to about 70 megacycles; and

said transmitting means including a transmitter and an antenna, said antenna being of the directional type.

41. A communications system as set forth in claim **40** wherein:

said antenna is a rhombic antenna.

42. A communications system as set forth in claim **40** wherein:

said antenna is a yagi antenna.

43. A communications system as set forth in claim **40** wherein:

said directional antenna is positioned so that the communications signal leaves the earth at an angle α of approximately 10° and the angle of injection into said duct is approximatetly α.

44. A communications system as set forth in claim **15** wherein:

said receiving means includes a receiver, said receiver being a double conversion superheterodyne receiver having two oscillators and two IF stages and a receiving antenna.

45. A communications system as set forth in claim **44** wherein:

said receiving antenna being a yagi antenna.

46. A communications system comprising:

means for generating and controlling an ionized cloud in the ionosphere, said means including one transmitting means for simultaneously transmitting signals in a suitable phase relationship to achieve focusing of the radio frequency generated by the transmitting means, means for slaving the signals from said transmitting means and antenna means for focusing on and accelerating ionospheric electrons with the electromagnetic radiation signals from said transmitting means to produce ionizing collisions;

another transmitting means for transmitting a communications signal directed at said ionized cloud, said cloud injecting said communications signal into and propagating said communications signal within an ionospheric signal trapping waveguide duct encircling the earth;

means for generating and controlling a second ionized cloud within said duct to extract at least a portion of said signal including means for producing signals to accelerate ionospheric electrons; and

means for detecting at least a portion of the signal intercepted by said second cloud.

47. A communications system comprising:

means for generating and controlling an ionized cloud within a given zone of the ionosphere, said means including,

one transmitting means comprising first and second transmitters for simultaneously transmitting electromagnetic radiation signals in the same phase relationship to achieve focusing of the radio frequency generated by the transmitting means,

means for slaving the signals from said transmitters,

antenna means for focusing on and accelerating ionospheric electrons with the signals from said transmitting means to produce ionizing collisions, said means comprising an antenna employed in a near-field arrangement and being a Mills Cross antenna having first and second sets of perpendicular legs, each of said transmitters being perpendicular and spaced apart from a different one of said sets of legs;

another transmitting means including a transmitter and a directional antenna for transmitting a communications signal having a frequency between a few megacycles to about 70 megacycles and directed at said ionized cloud, said cloud injecting at least a portion of said communications signal into and propagating within an ionospheric signal trapping zone encircling tne earth, said zone defined by a lower boundary which is a function of the angle of injection of the signal into the zone and an upper shell-like refractive layer boundary which is a function of the frequency of said communications signal, said signal being periodically refracted at points along the upper shell-like refractive layer without any lower boundary reflections as it is propagated approximately parallel to the earths surface;

means for generating and controlling a second ionized cloud within said zone to extract at least a portion of said communications signal including means for producing signals to accelerate ionospheric electrons; and

means for receiving at least a portion of said communications signal trapped witnin said zone.

48. A communications system as set forth in claim **47** wherein:

each of said antenna legs has a width of approximately 12 meters and a length of approximately 19 kilometers;

said antenna has a horizontal radiating surface equal to one half wavelength and a mounting portion of one quarter wavelength; and

said first and second transmitters each being spaced approximately 80 kilometers from the corresponding legs of said antenna and each having a parabolic transmitting antenna.

49. A communications system as set forth in claim **48** wherein:

said ionized cloud is located approximately 80 kilometers above the earths surface, has a mean width of approximately 15 meters, a length of approximately 60 meters and tapers in a cone-shaped fashion from its widest portion to a point at its greatest distance from the earth's surface.

50. A communications system as set forth in claim **47** wherein:

said directional antenna is positioned so that the communications signal leaves the earth at an angle α of approximately 10° and the angle of injection into said zone is approximately α.

51. A communications system as set forth in claim **47** wherein:

said receivng means includes a receiver, said receiver being a double conversion superheterodyne receiver having two oscillators and two IF stages and a receiving antenna.

52. A communications system as set forth in claim **51** wherein:

said receiving antenna being a yagi antenna.

53. A communications system as set forth in claim **47** wherein:

said second cloud generating and controlling means includes,

one transmitting means comprising first and second transmitters for simultaneously transmitting electromagnetic radiation signals in the same phase relationship to achieve focusing of the radio frequency generated by the transmitting means,

means for slaving the signals from said transmitters;

antenna means for focusing on and accelerating ionospheric electrons with the signals from said transmitting means to produce ionizing collisions, said means comprising an antenna employed in a near-field arrangement and being a Mills Cross antenna having first and second sets of perpendicular legs, each of said transmitters being perpendicular and spaced apart from a different one of said sets of legs.

54. A communications system as set fortn in claim **49** wherein:

said Mills Cross antennas each operate at a frequency of approximately 50 megacycles and provide approximately 26 negawatts of peak power and 260 kilowatts of average power.

55. A communications system as set forth in claim **47** wherein:

both said cloud generating and controlling means as well as said communications signal transmitting and receiving means are located near the earths surface.

56. A communications system as set forth in claim **47** wherein:

the lower boundary of said zone is substantially bounded by the ionospheric bottom layer which is approximately 60 kilometers above the earths surface and the upper boundary is substantially bounded by the ionospheric F_2 layer, which is approximately 350 kilometers above the earths surface.

57. A communications system as set forth in claim **36** wherein:

said antenna is an ellipsoidic antenna having a diameter of approximately 40 kilometers.

References Cited

UNITED STATES PATENTS

2,871,344	1/1959	Busibries	343—100.4
3,030,500	4/1962	Katzin	343—100
3,189,901	5/1965	Cutolo	343—100

OTHER REFERENCES

Terman, Frederick E.: Electronic and Radio Engineering, fourth ed., 1955, pp. 834–5 relied on.

RODNEY D. BENNETT, Jr., *Primary Examiner.*

J. P. MORRIS, *Assistant Examiner.*

United States Patent [19]
Bass

[11] **Patent Number:** **4,999,637**
[45] **Date of Patent:** **Mar. 12, 1991**

[54] CREATION OF ARTIFICIAL IONIZATION CLOUDS ABOVE THE EARTH

[75] Inventor: **Ronald M. Bass**, Houston, Tex.
[73] Assignee: **APTI, Inc.**, Washington, D.C.

[21] Appl. No.: **49,881**
[22] Filed: **May 14, 1987**
[51] **Int. Cl.^{5}** **H04B 7/00**
[52] **U.S. Cl.** **342/367**; 342/5
[58] **Field of Search** 361/231; 342/352, 5, 342/367; 376/123, 124

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,133,250	5/1964	Molmud	342/352
3,174,150	1/1965	Sferrazza et al.	342/352
3,189,901	6/1965	Cutolo	342/350
3,300,721	1/1967	Seaton	342/352
3,445,844	5/1969	Grossi et al.	342/367
3,518,670	6/1970	Miller	342/5
3,882,393	5/1975	Epstein	455/59
4,035,726	7/1977	Brice et al.	342/352
4,686,605	8/1987	Eastlund	361/231
4,712,155	12/1987	Eastlund et al.	361/231

OTHER PUBLICATIONS

Radio Science, vol. 15, No. 2, pp. 213–223, (4/80), "MST Radar at Poker Flat, Alaska", Balsley et al.

Primary Examiner—Salvatore Cangialosi
Attorney, Agent, or Firm—Drude Faulconer

[57] **ABSTRACT**

A method for forming a cloud of artificial ionization above the earth by initially heating the resident plasma at a desired altitude with electromagnetic radiation having a frequency approximately the same as that of the ambient plasma. As the plasma frequency increases due to heating, the radiation frequency is also increased until the final maintenance frequency is attained.

11 Claims, 1 Drawing Sheet

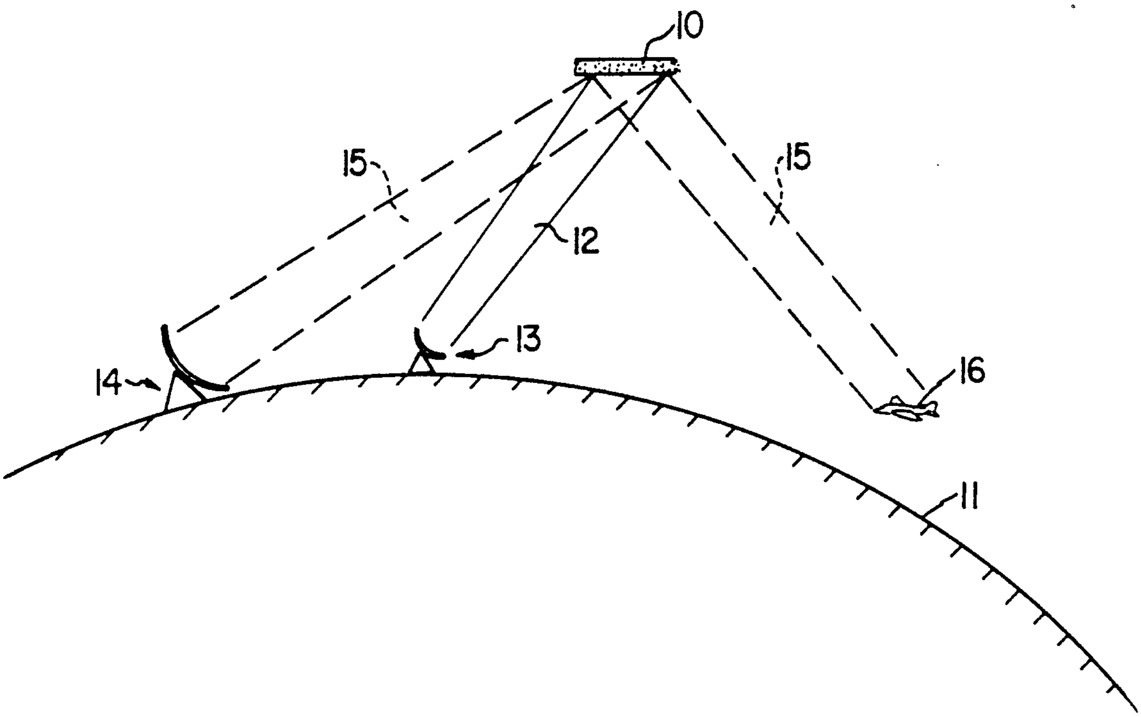

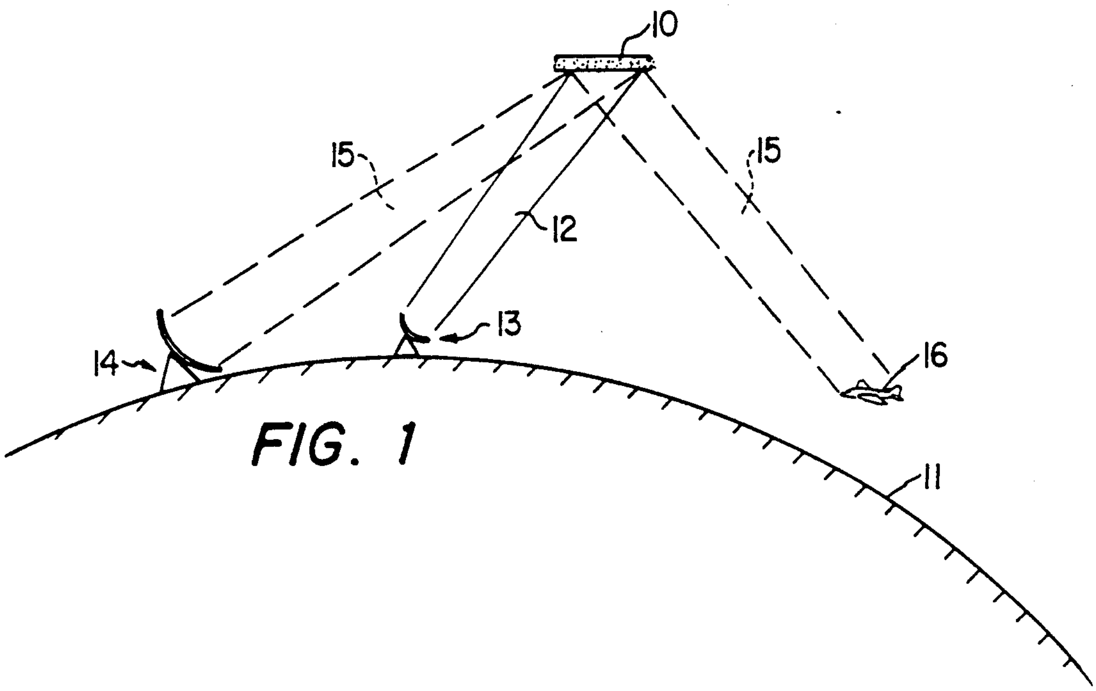
10
15
15
12
13
14
16
11

FIG. 1

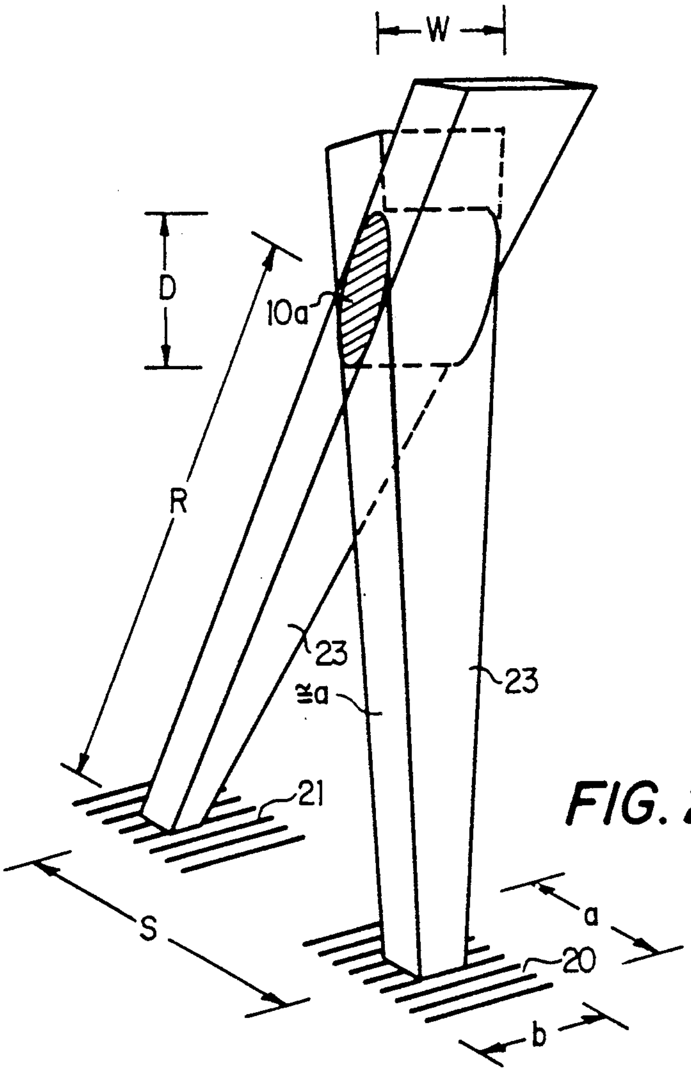
W
D
10a
R
23
23
≅a
21
S
a
20
b

FIG. 2

CREATION OF ARTIFICIAL IONIZATION CLOUDS ABOVE THE EARTH

DESCRIPTION

1. Technical Field

The present invention relates to a method for establishing a patch or cloud of artificial ionization above the earth and more particularly, relates to a method of forming a cloud of artificial ionization by heating naturally occurring plasma with electromagnetic energy which is transmitted from the Earth's surface at a variable, increasing frequency.

2. Background Act

Certain communication and radar systems operate by "bouncing" transmitted and/or reflected signals off of naturally-occurring layers of ionization in the ionosphere. One known system using this technique is "over-the-horizon" (OTH) radar. By bouncing or reflecting the signals off an ionized layer, the signals can actually travel "over-the-horizon", thereby substantially increasing the range of the system.

However, while present OTH systems are capable of detecting objects at long range (e.g. strategic threats), they are not well suited for detecting "close-in" objects (e.g. missiles at 1000 kilometers or less). One problem lies in the fact that as the beam angle of the radar is increased from the horizontal, the frequency of the beam must be lowered in order to achieve refraction at a more nearly normal incidence. As this frequency is lowered, the antenna system gain is reduced and the radar cross section decreases for small close-in object. These effects act to set a minimum range for the OTH system.

Another major problem with present OTH systems is related to the low radar cross section of small targets at typical OTH operating frequencies. These objects having small cross-sections produce a weak return signal even when the object is within the range of the OTH radar since the OTH system is normally designed for objects having much larger cross-sections, e.g. large aircraft.

A still further problem encountered by present OTH radar is directly related to the unstable conditions in the ionosphere which widely vary depending on seasonal, diurnal, and/or sunspot cycles. Accordingly, the operating frequency of present OTH radar systems has to be constantly adjusted to allow for the varying ionospheric conditions which may vary so much at times that the OTH system is rendered inoperable.

Several techniques have been proposed to overcome some of the shortfalls of present OTH radar systems. One known technique is disclosed in U.S. Pat. No. 3,445,844 wherein a cloud of artificial ionization is formed above the earth to serve as a layer for redirecting communication signals. The cloud is formed by "breakdown", i.e. creation of a high level flux of free electrons (i.e. plasma) at a desired altitude by focusing electromagnetic energy thereon to heat a localized region or area. The electromagnetic energy heats and accelerate the electrons in the resident plasma to a degree such that their kinetic energy reaches the level required for the occurrence of ionizing collisions. Scattering from a cloud so formed takes place due to the discontinuity between this zone of enhanced ionization and the surrounding medium.

A cloud formed in accordance with the method disclosed in U.S. Pat. No. 3,445,844 will provide a good reflection layer for OTH radar and like systems. However, when a cloud is formed by breakdown as in the mentioned patent, the plasma frequency of the cloud quickly adjusts to the hating frequency and breakdown is initiated over the entire area of the cloud. By initiating and forming the cloud with radiation having the same high frequency as that required for continued maintenance of the cloud once formed, a substantial amount of power is required, much of which is reflected or passes through the cloud while it is being formed and, accordingly, is wasted.

BRIEF DESCRIPTION OF THE DRAWINGS

The actual construction, operation, and apparent advantages of this invention will be better understood by referring to the drawings in which like numerals identify like parts and in which:

FIG. 1 is a simplified schematical view of a system for forming a cloud of artificial ionization above the earth for bouncing signals over-the-horizon in accordance with the present invention; and

FIG. 2 is a schematic illustration of a cross-beam radiation transmission system for forming a cloud of artificial ionization in accordance with the present invention.

DISCLOSURE OF THE INVENTION

The present invention provides a method for forming a patch or cloud of artificial ionization at an altitude above the earth wherein no substantial amounts of power are wasted in forming and maintaining the cloud.

More specifically, the present invention provides a method wherein variable frequency heating is used to form a cloud of artificial ionization. This is accomplished by initially heating the resident plasma at the selected altitude by transmitting electromagnetic radiation from the earth at an initial frequency which is approximately the same frequency as the ambient plasma frequency. This radiation, being of the same frequency as the plasma, will be efficiently absorbed with relatively little being reflected from or passed through the ambient plasma. The radiation heats the plasma and accelerates the free electrons in the plasma thereby increasing the plasma frequency.

The frequency of the plasma is monitored by radar or the like and, as it increases, the frequency of the radiation being transmitted also increases, preferably in a manner where the radiation frequency continues to substantially match the increasing plasma frequency. The radiation frequency is continuously increased until the final maintenance frequency is attained, at which time, the transmission of electromagnetic radiation is continued at the final frequency to maintain the integrity of the cloud. The final plasma frequency (i.e. maintenance frequency) is selected so that it is always greater than the frequency of any signals (e.g. communications, radar) that are expected to be "bounced" off the cloud once the cloud is used for its intended purpose.

To further conserve power and to reduce the power required for carrying out this invention, the initial radiation is weakly focused so that only the center area of the plasma within the cloud are will undergo initial heating. The focus of the radiation is contracted as the radiation frequency is increased until the entire cloud are is heated by the radiation. The radiation can be transmitted by a single antenna system or by two spaced antenna systems positioned so that their beams intersect at said altitude to thereby form the cloud.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Referring more particularly to the drawings, FIG. **1** is a illustration of how the present invention is utilized with an over-the-horizon (OTH) radar system. A cloud **10** of artificial ionization is formed at an altitude above the surface of the earth **11** by transmitting electromagnetic energy **12** from an antenna system **13**. An OTH radar system **14** transmits and receives signals **15** that are reflected off cloud **10** to detect a target **16** that is "over-the-horizon" as will be understood in the art.

In accordance with the present invention, cloud **10** is formed by variable frequency heating. The degree of ionization in the ionosphere depends on the electron temperature of the average energy of the free electrons; i.e. plasma, at a particularly altitude. The electron energy can be increased by absorption of incident electromagnetic radiation. This, in turn, increases the degree of ionization (i.e. number of free electrons and ions per unit volume). The higher the electron density of an ionized layer, the higher the frequency of radio or radar waves that can be reflected from that layer in communications or radar applications.

To raise the electron temperature efficiently, it is necessary to irradiate the ionosphere at or near the frequency of the plasma that is naturally present at the altitude of interest. "Plasma frequency" is defined roughly as the highest frequency that will be reflected from a particular altitude in the ionosphere and this frequency will increase as the electron density increases. However, if the frequency of the incident radiation used to heat the plasma is much higher or lower than the plasma frequency a large portion of the radiation will not be absorbed but will be reflected or passed through the heating zone and will be wasted.

In the present method, a target altitude at which cloud **10** is to be formed is selected and the frequency of the natural plasma is determined by tracking radar or the like. Heating is begun by transmitting electromagnetic radiation **12** from antenna system **13** at substantially the same frequency as that of the plasma resident at the target altitude. Radiation **12** is absorbed efficiently by the plasma at the target altitude which increase the electron density thereof which, in turn, increases the frequency of the plasma frequency. As the plasma frequency increases, the frequency of radiation **12** is increased to approximately match the increasing plasma frequency. By constantly tracking the increasing plasma frequency and adjusting the frequency of the heating radiation accordingly, almost all, if not all, of the power being used to heat the ionospheric electrons is absorbed efficiently into the plasma over the entire heating cycle. In this way, the electron density (degree of ionization) can be raised to the desired level without any substantial waste of power which otherwise would be considerable.

In the present method, the initial heating is carried out with the radiation **13** being broadly focused so that the central area within cloud **10** is being heated. As the plasma frequency within the smaller area is increased, the focused area of radiation **13** is contracted until the entire final area of cloud **10** is being heated by the radiation **13**. This minimizes the initial heating power requirements and results in a further substantial reduction in the overall power requirements for forming cloud **10**.

The antenna system **13** required to transmit radiation **12** in the present invention may be of any known construction having high direct inability capabilities; for example, a phased array, beam spread angle (O) type see U.S. Pat. No. 3,445,844 and the "The MST RADAR at Poker Flat, Ak.," Radio Science, Vol. 15, No. 2, March-April, 1980; pps. 213–223, both of which are incorporated herein by reference. A phased array antenna generating a steerable focused beam can be assembled at a single site or two phased array antenna systems **20, 21** (FIG. **2**) may be spaced from each of other to generate two coherent beams **23, 24** of radiation that cross each other at the selected altitude to form cloud **10***a*.

The key beam geometric parameters of a two antenna system are shown in FIG. **2**. The beam of radiation from each antenna is assumed to diverge in the azimuthal plane (b) and to be collimated in the elevational plane (a). This assumption is appropriate for a backscatter cloud that must be inclined about 45 degrees from the horizontal. For forward scattering, with cloud **10***a* either directly overhead or downrange from an antenna, the beam will diverge in both planes and "b" would be used for both antenna. The following relationships can be used to calculate the various parameters of the system of FIG. **2**:

$$a \cong \sqrt{2\lambda_p R} \qquad (1)$$

wherein:

a=length of side of antenna array in elevational plane.

λp=Wavelength of Radiation Frequency

R=Actual distance from array to cloud **10***a*.

$$b \cong \frac{\lambda_p R}{W} \qquad (2)$$

wherein:

b=length of side of antenna array in azimuthal plane

W=Width of cloud **10***a*

$$S = \frac{Ra}{D} \qquad (3)$$

S=distance between arrays.

D=Depth of cloud **10***a*.

While the size and characteristics of a particular cloud **10** will vary depending on its application and actual conditions under which it is formed, the following example will serve to better illustrate the present invention. A cloud **10** (FIG. **1**) is to formed at an altitude of 90 kilometers (km) and is to have a final area of 1 square km or larger with a thickness of from 3 to 10 meters. The resident plasma density at 90 km is typically on the order of 10^6/cubic cm. A square, phased array antenna system **13** of 100 meters on a side, beams electromagnetic radiation 12 (i.e. power) at an initial or starting frequency to initiate heating of the resident plasma. Antenna **13** is focused so that only the plasma at the center of cloud **10** will initially be heated, thereby requiring substantially less power than if the entire 1 square km area of the cloud was originally heated.

Since the plasma frequency of the ambient plasma is approximately 9 megahertz (MH_z), about 70 MW of power will be required. As the plasma heats, the frequency thereof increased and is tracked by groundbased radar. As the frequency of the plasma increases, the frequency of radiation **12** is increased accordingly until a frequency of 15 MHz is reached. The power requirement at this point will have dropped to approximately 27 MW.

At this point, the focus of the antenna is contracted to cover a greater, if not all, of the area of cloud **10**, and radiation **12** (i.e. heating) is now applied over the entire area. Due to the increased heating area, the power requirement is temporarily increased to approximately 33 MW but will quickly decrease as the radiation frequency continues to increase until the final maintenance frequency of 300 MH_z is reached. The power requirement for the final maintenance frequency is approximately 7 MW and final electron or plasma density will be about 10^9/cubic cm.

The integrity of cloud **10** will be maintained as long as radiation **12** is transmitted thereto at the final maintenance frequency. The final maintenance frequency for any particular cloud is interrelated to the frequency of the radar or other communication signal which is to be bounced off that cloud. That is, it is generally preferably to have the plasma frequency of the cloud substantially higher than the radar frequency. This will ensure a high degree of reflection for the radar. Also, irregularities in the plasma density may form in the cloud. These will commonly be spaced about one heater wavelength apart. This spacing should be relatively small compared to a radar wavelength to avoid excessive scattering of the radar signal in undesirable directions.

What is claimed is:

1. A method of forming a cloud of artificial ionization at an altitude above the earth, said method comprising:

initiating heating of the resident plasma at said altitude by transmitting electromagnetic radiation from the earth to said altitude at an initial frequency which is approximately the same as the original frequency of said resident plasma; and

increasing said frequency of said electromagnetic radiation as said frequency of said resident plasma increases, until a final maintenance frequency is attained, said maintenance frequency being t or above the plasma frequency necessary to provide a plasma having an electron density capable of reflecting communication or like signals which come into contact with said plasma.

2. The method of claim **1** including:

defocusing said electromagnetic radiation so only the center area of said cloud is initially heated; and

contracting the focus of said electromagnetic radiation as the frequency of said radiation is adjusted until the entire area of said cloud is heated.

3. The method of claim **1** wherein said electromagnetic radiation is transmitted by a single antenna system.

4. The method of claim **1** wherein said electromagnetic radiation is transmitted by two antenna systems, each spaced from the other, and inclined whereby the beams of said electromagnetic radiation transmitted from said systems will intersect each other at said altitude.

5. A variable frequency heating method for forming a cloud of artificial ionization at an altitude above the earth, said method comprising:

transmitting electromagnetic radiation form the earth to said altitude at an initial frequency which is approximately the same as the original frequency of the plasma naturally present at said altitude;

focusing said electromagnetic radiation to heat said plasma to thereby accelerate the free electrons therein thereby increasing the frequency of said plasma;

monitoring the frequency of said plasma as it increases;

increasing the frequency of said electromagnetic radiation as said frequency of said plasma increases;

continuing to increase said frequency of said electromagnetic radiation until a final desired maintenance frequency is attained; said final desired frequency being at or above the plasma frequency necessary to provide a plasma having an electron density capable of reflecting communication signals or the like which come into contact with said plasma;

and continuing to transmit said electromagnetic radiation at said final frequency to maintain the integrity of said cloud.

6. The method of claim **5** wherein said final frequency is greater than the frequency of any communication and/or radar signals expected to be reflected by said cloud.

7. The method of claim **6** wherein said electromagnetic radiation is initially focused whereby only the center area of said plasma is initially heated.

8. The method of claim **7** including:

contracting the focus of said electromagnetic radiation as the said frequency of said radiation is increased whereby the entire area of said cloud is heated.

9. The method of claim **8** wherein said frequency of said electromagnetic radiation is increased to approximately match said increasing frequency of said plasma.

10. The method of claim **9** wherein said electromagnetic radiation is transmitted by a single antenna system.

11. The method of claim **9** wherein said electromagnetic radiation is transmitted by two antenna systems, each spaced from the other, and inclined whereby the beams of said electromagnetic radiation transmitted from said systems will intersect each other at said altitude.

* * * * *

United States Patent [19]
Koert

[11] **Patent Number:** **5,041,834**
[45] **Date of Patent:** **Aug. 20, 1991**

[54] **ARTIFICIAL IONOSPHERIC MIRROR COMPOSED OF A PLASMA LAYER WHICH CAN BE TILTED**

[75] Inventor: **Peter Koert**, Washington, D.C.

[73] Assignee: **APTI, Inc.**, Washington, D.C.

[21] Appl. No.: **524,435**

[22] Filed: **May 17, 1990**

[51] **Int. Cl.**^{5} **H04B 7/00**; H01Q 3/22
[52] **U.S. Cl.** **342/367**; 342/372
[58] **Field of Search** 342/367, 353, 371, 372; 455/64

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,445,844 5/1969 Grossi et al. 342/367
4,253,190 2/1981 Csonka 455/12
4,686,605 8/1987 Eastlund 361/231
4,712,155 12/1987 Eastlund et al. 361/231
4,817,495 4/1989 Drobot 89/1.11

Primary Examiner—Gregory C. Issing
Attorney, Agent, or Firm—Foley & Lardner

[57] **ABSTRACT**

This invention relates to generation of a Artificial Ionospheric Mirror (AIM), or a plasma layer in the atmosphere. The AIM is used like the ionosphere to reflect RF energy over great distances. A tiltable AIM is created by a heater antenna controlled in phase and frequency. The heater antenna phase shift scans a beam to paint a plasma layer. Frequency is changed to refocus at continually higher altitudes to tilt the plasma layer.

16 Claims, 15 Drawing Sheets

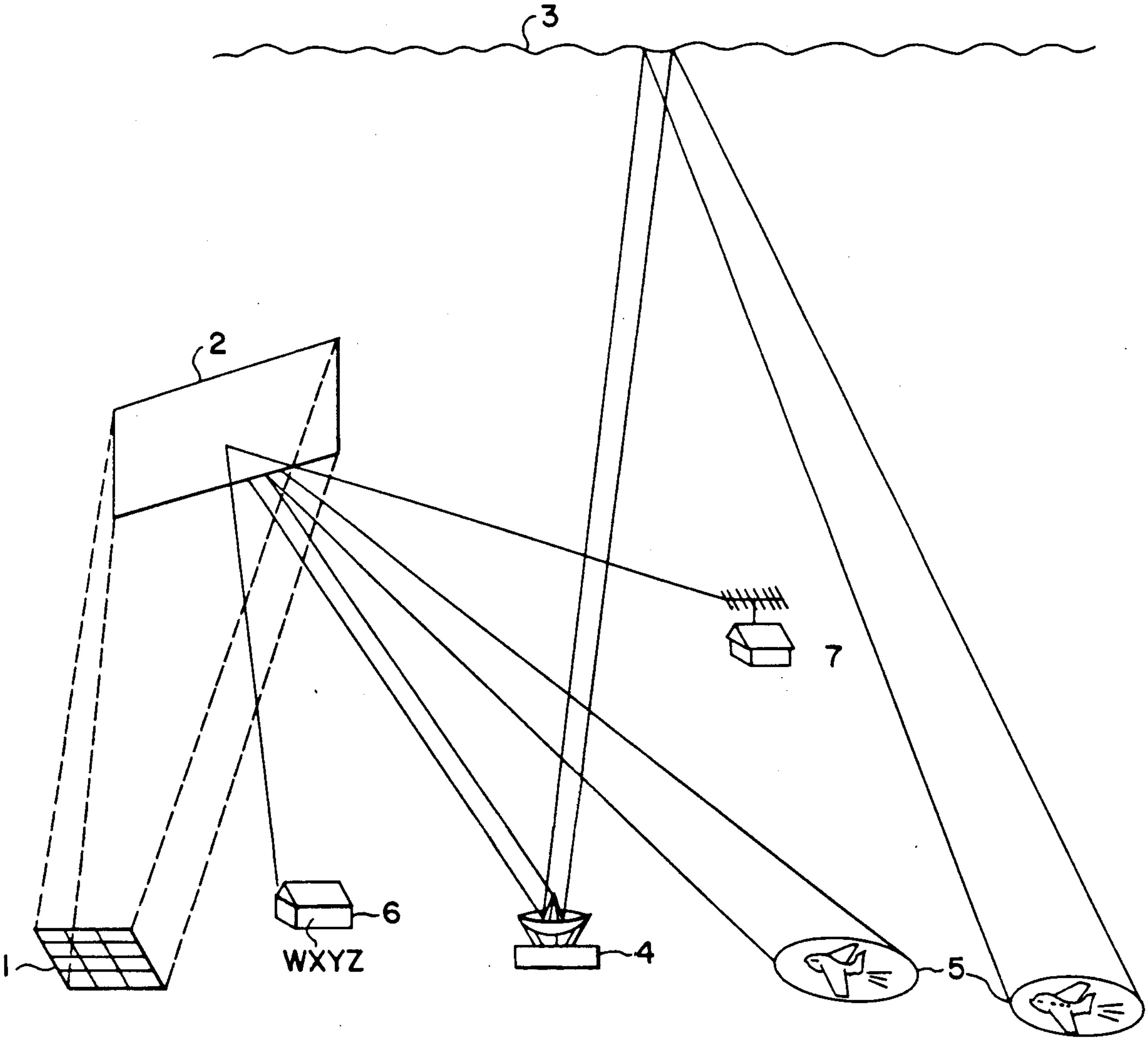

FIG. 1

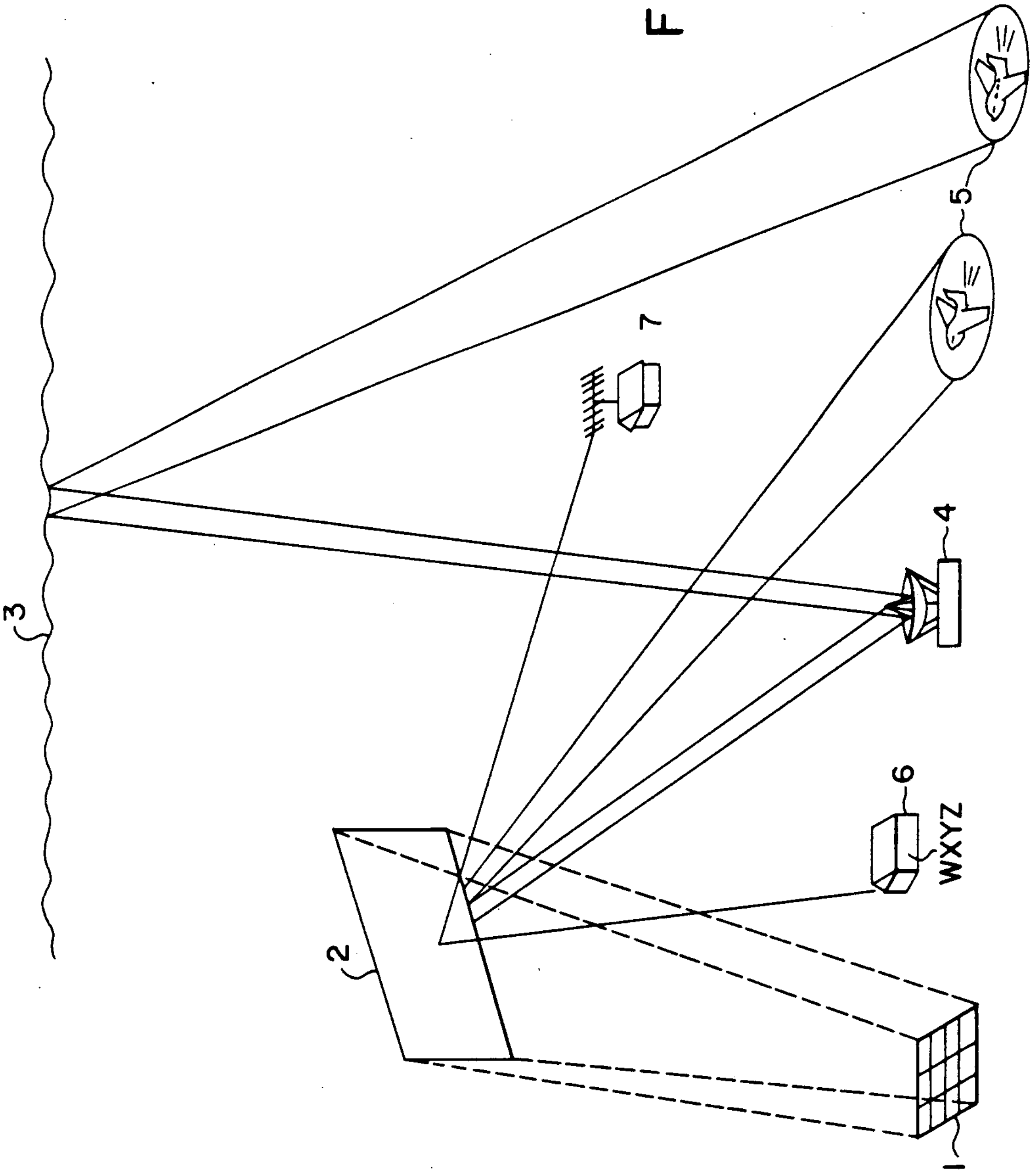
1
2
3
4
5
5
6
7
WXYZ

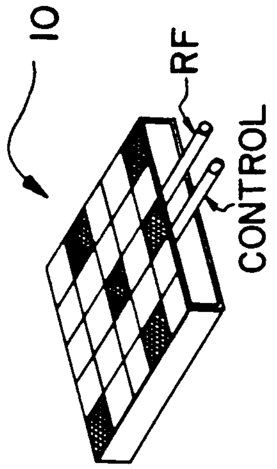
10
RF
CONTROL

FIG. 2b

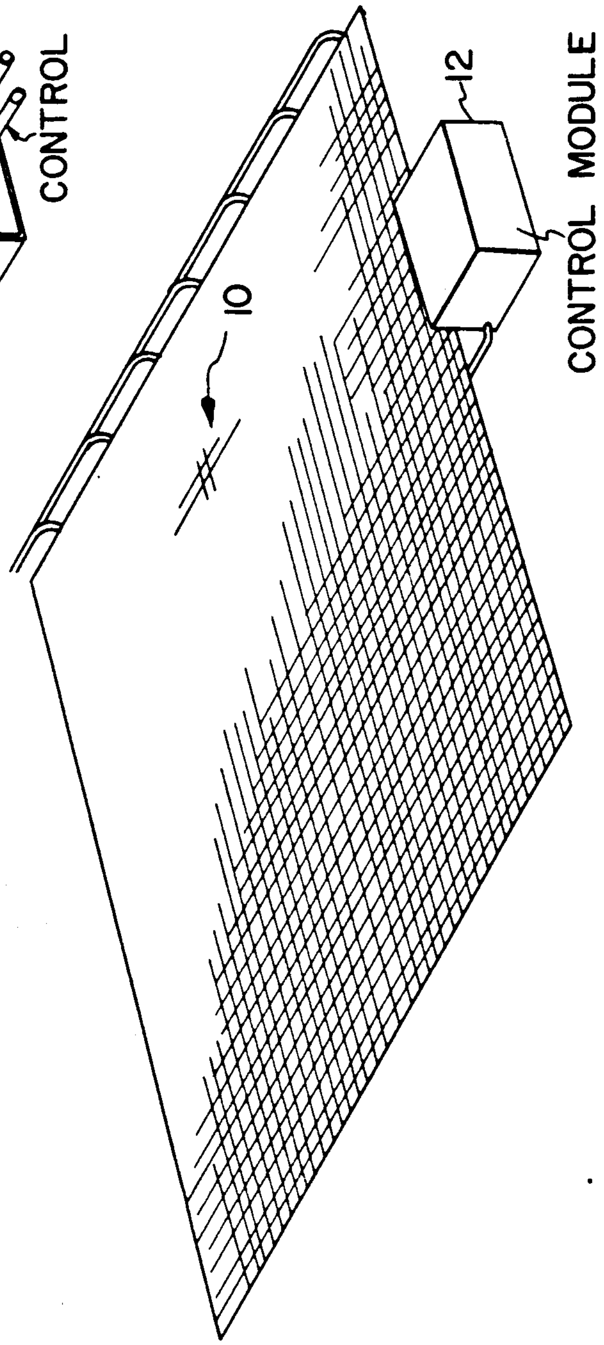
12
CONTROL MODULE
10

FIG. 2a

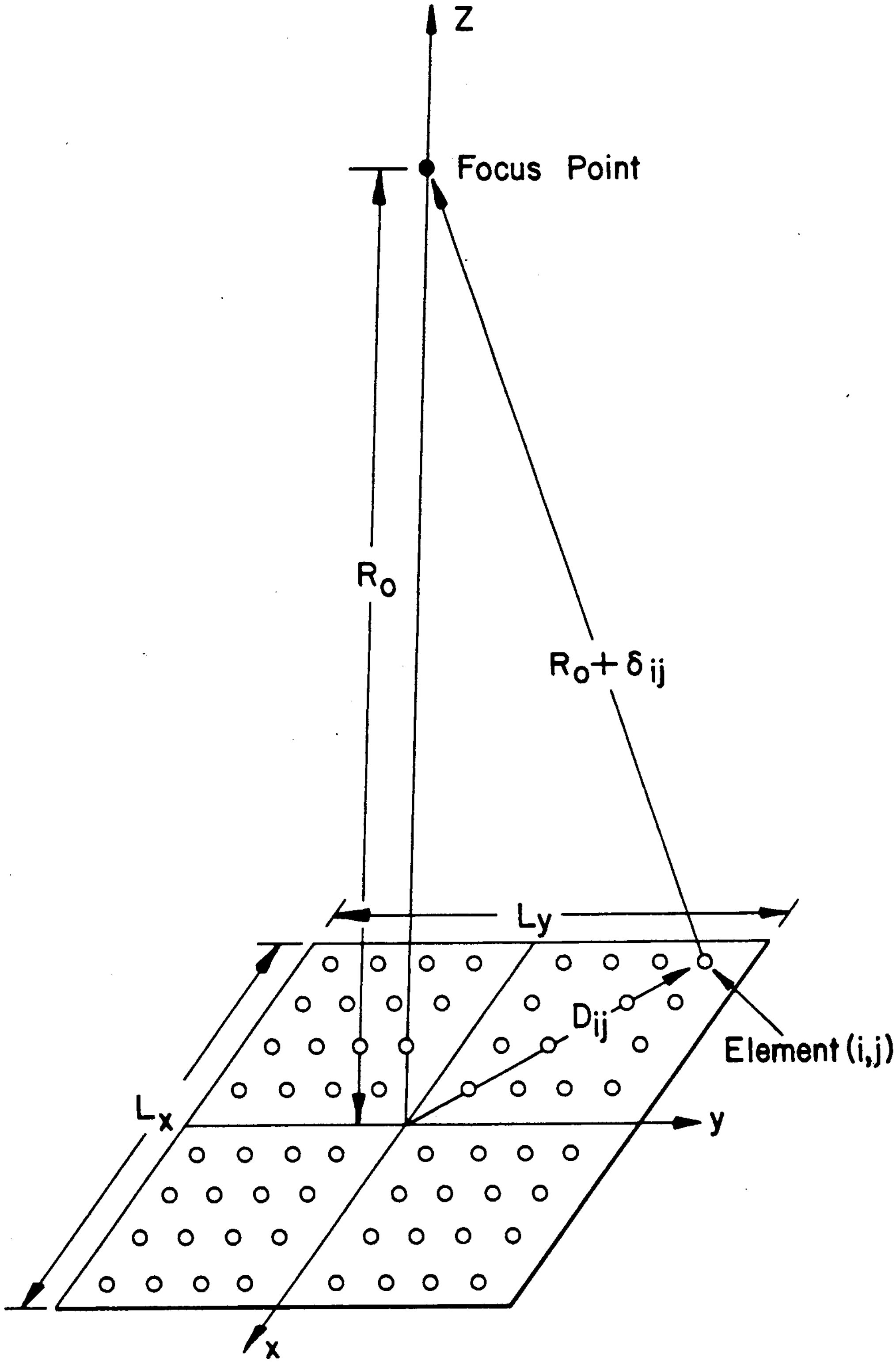
Z
Focus Point
R_o
$R_o + \delta_{ij}$
L_y
L_x
D_{ij}
Element (i,j)
y
x

FIG. 3

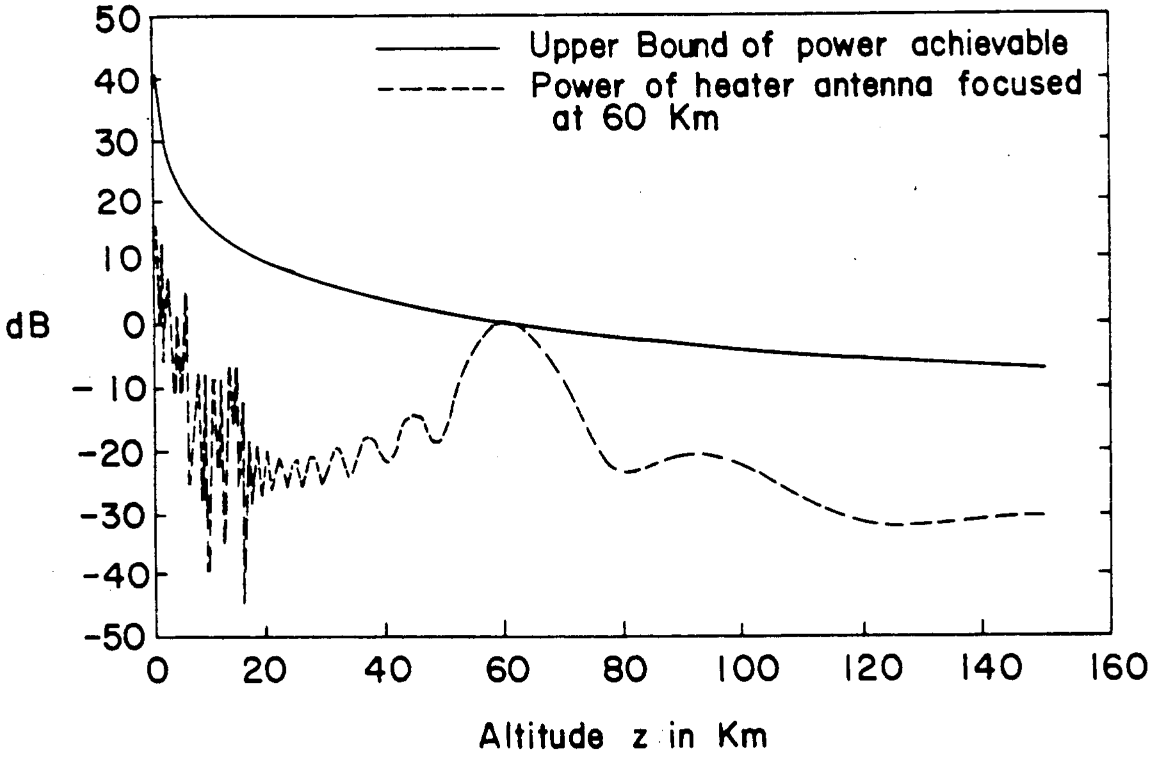
50
40
30
20
10
0
- 10
-20
-30
-40
-50
dB
0
20
40
60
80
100
120
140
160
Altitude z in Km
Upper Bound of power achievable
Power of heater antenna focused at 60 Km

FIG. 4

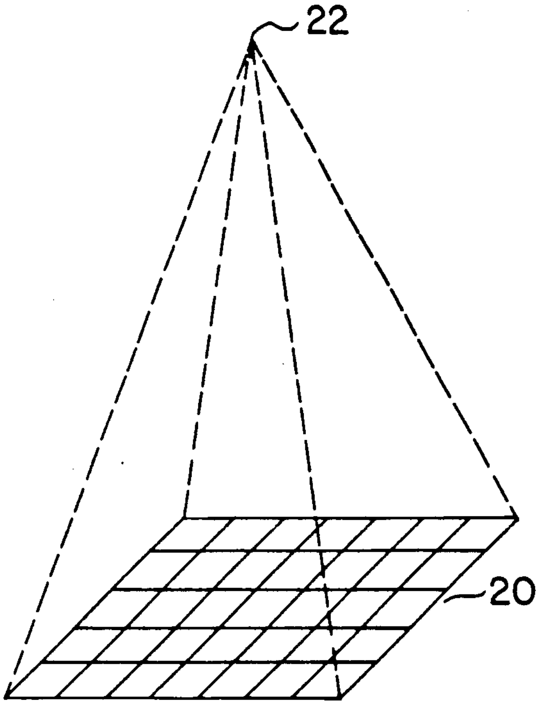

FIG.5

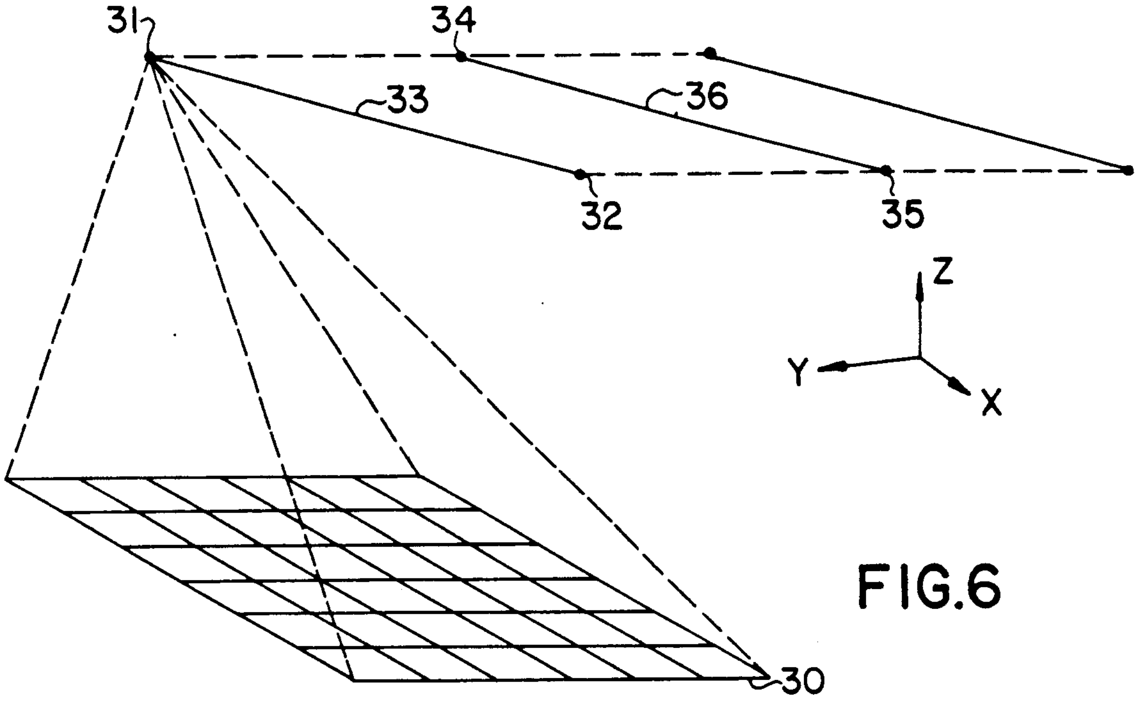
31
34
33
36
32
35
Z
Y
X
30

FIG.6

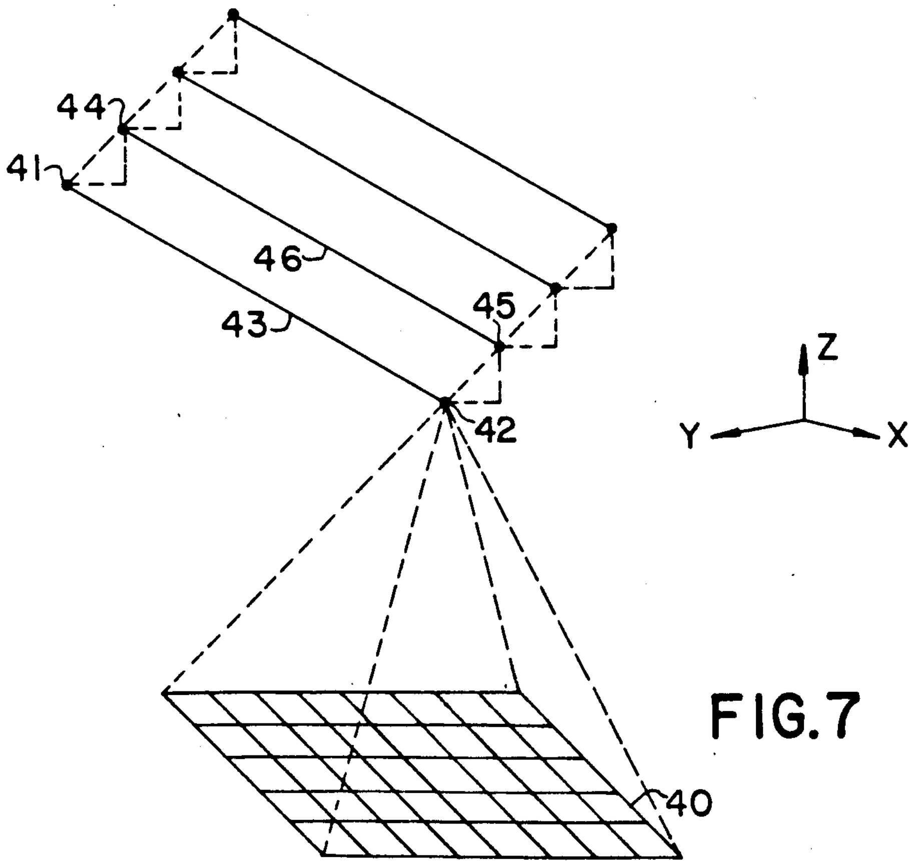
44
41
46
45
43
42
Z
Y
X
40

FIG.7

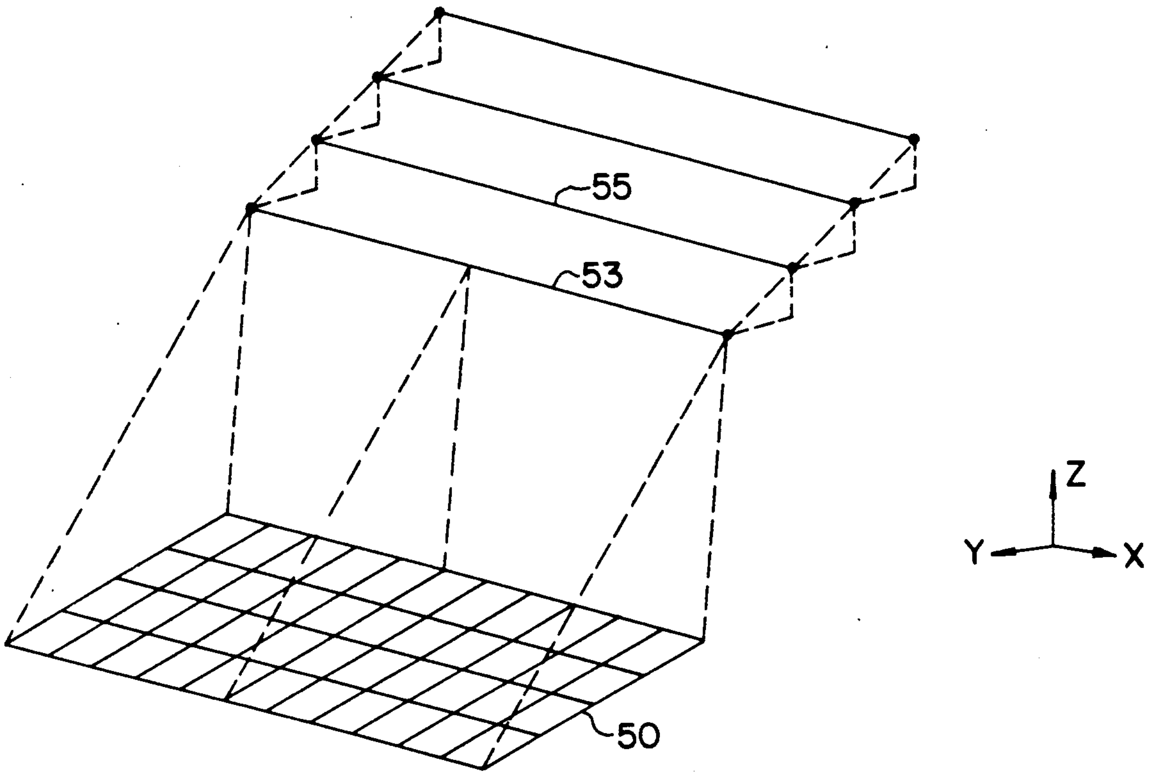

FIG.8

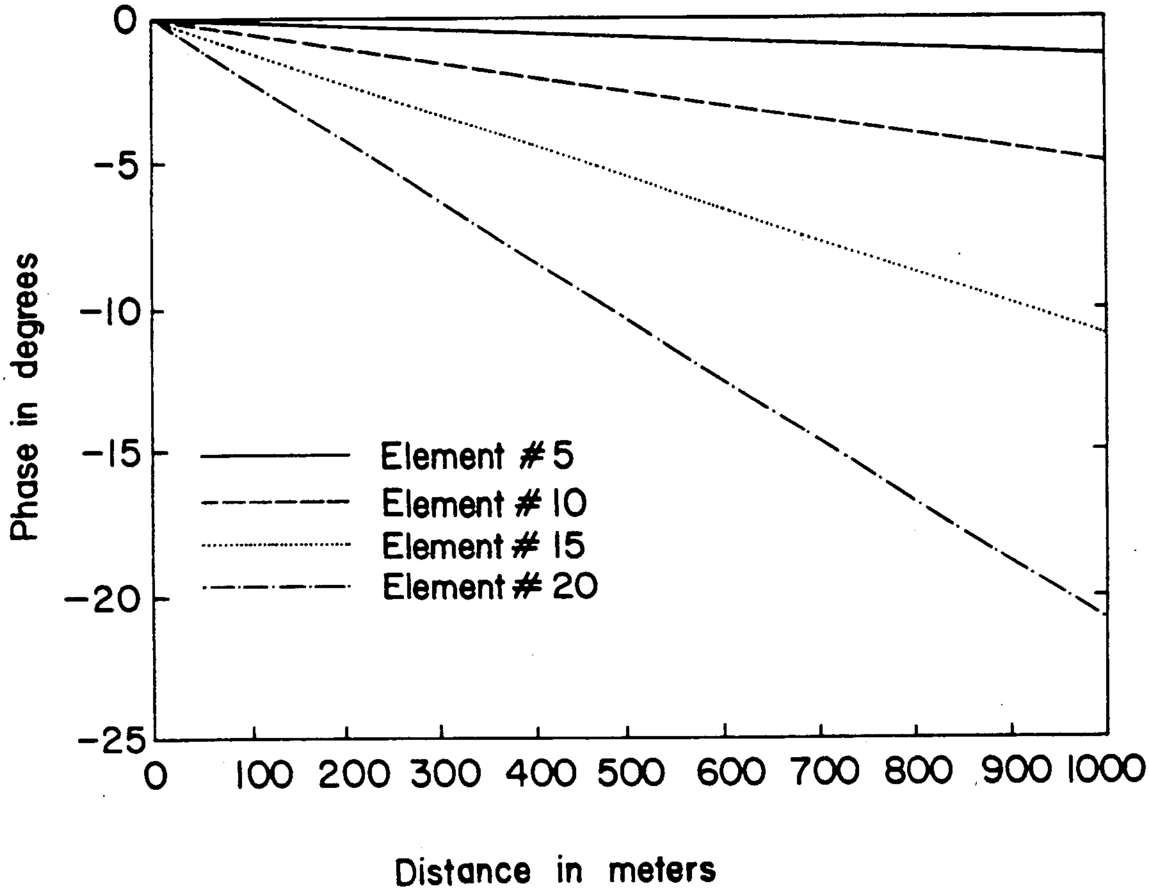

Phase Corrections to move focal point from 60Km to 61Km

FIG.9

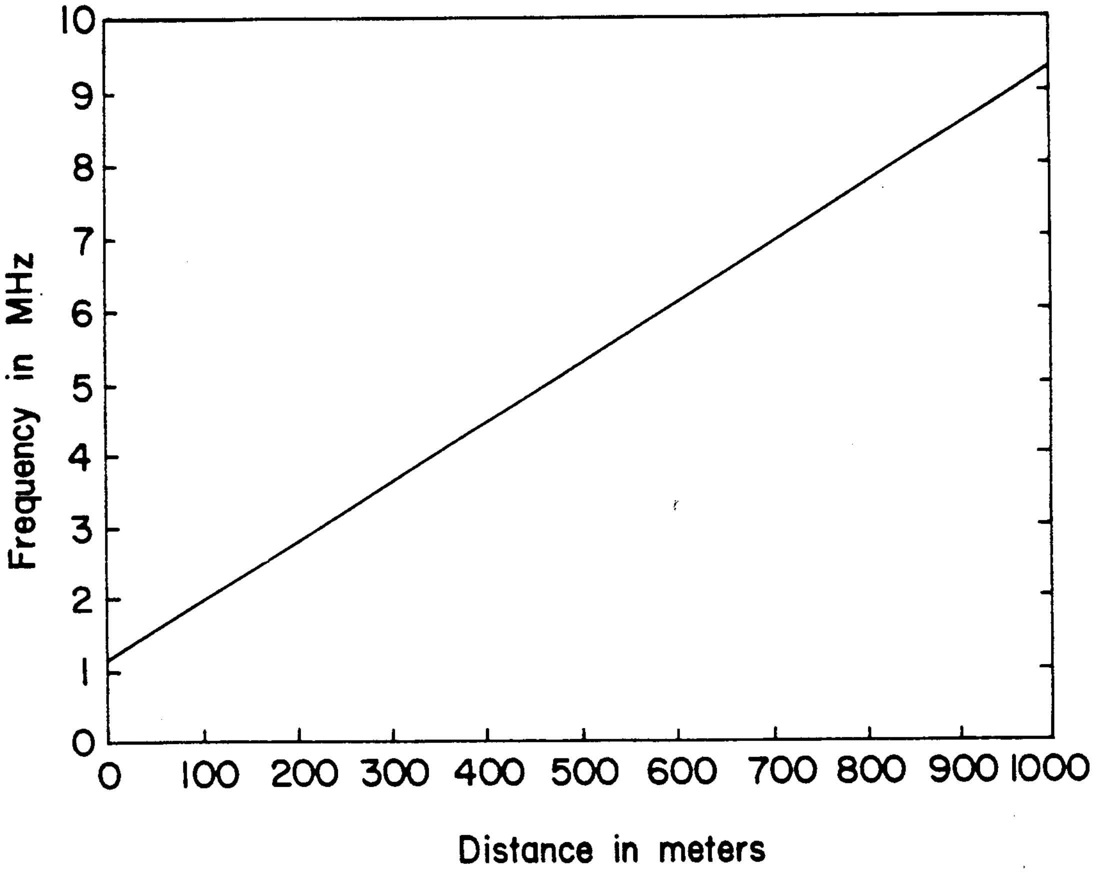

Frequency correction to move focal point from 60Km to 61Km

FIG. 10

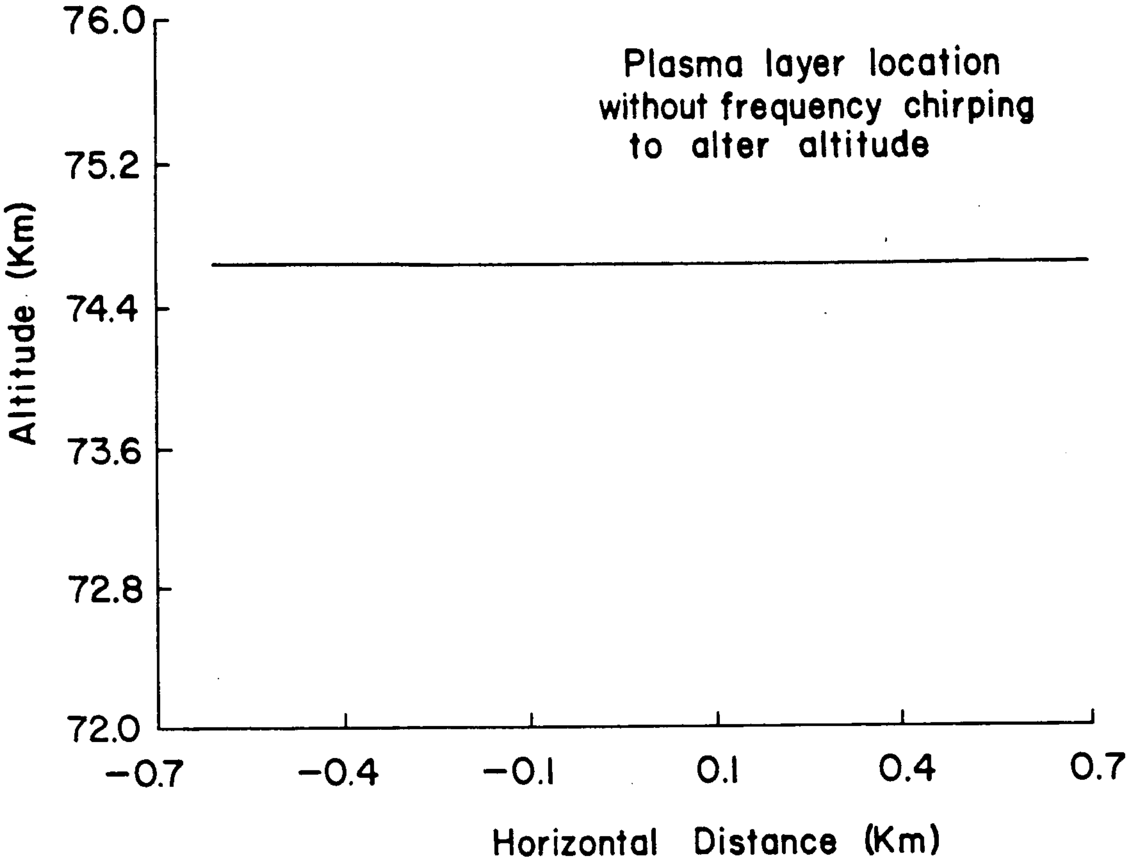

Heater focusing at 75Km with 550 MHz frequency and 154 dBW

FIG. 11

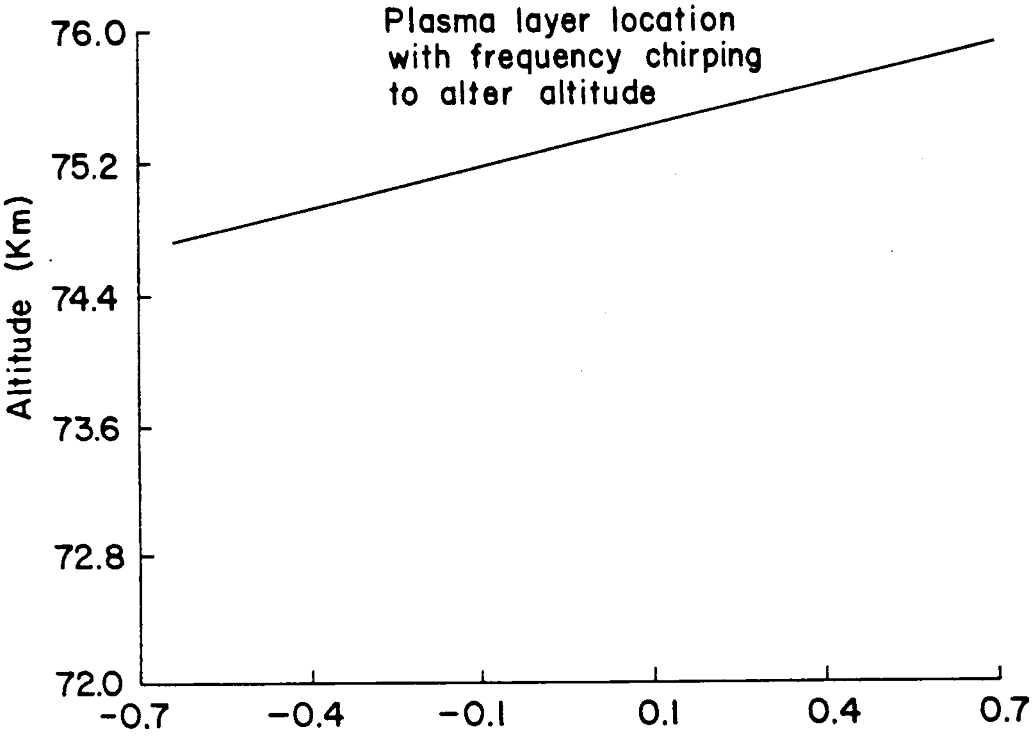

Heater Focusing at 75Km with 550 to 559.375 MHz frequency and 154 dBW

FIG. 12

FIG. 13

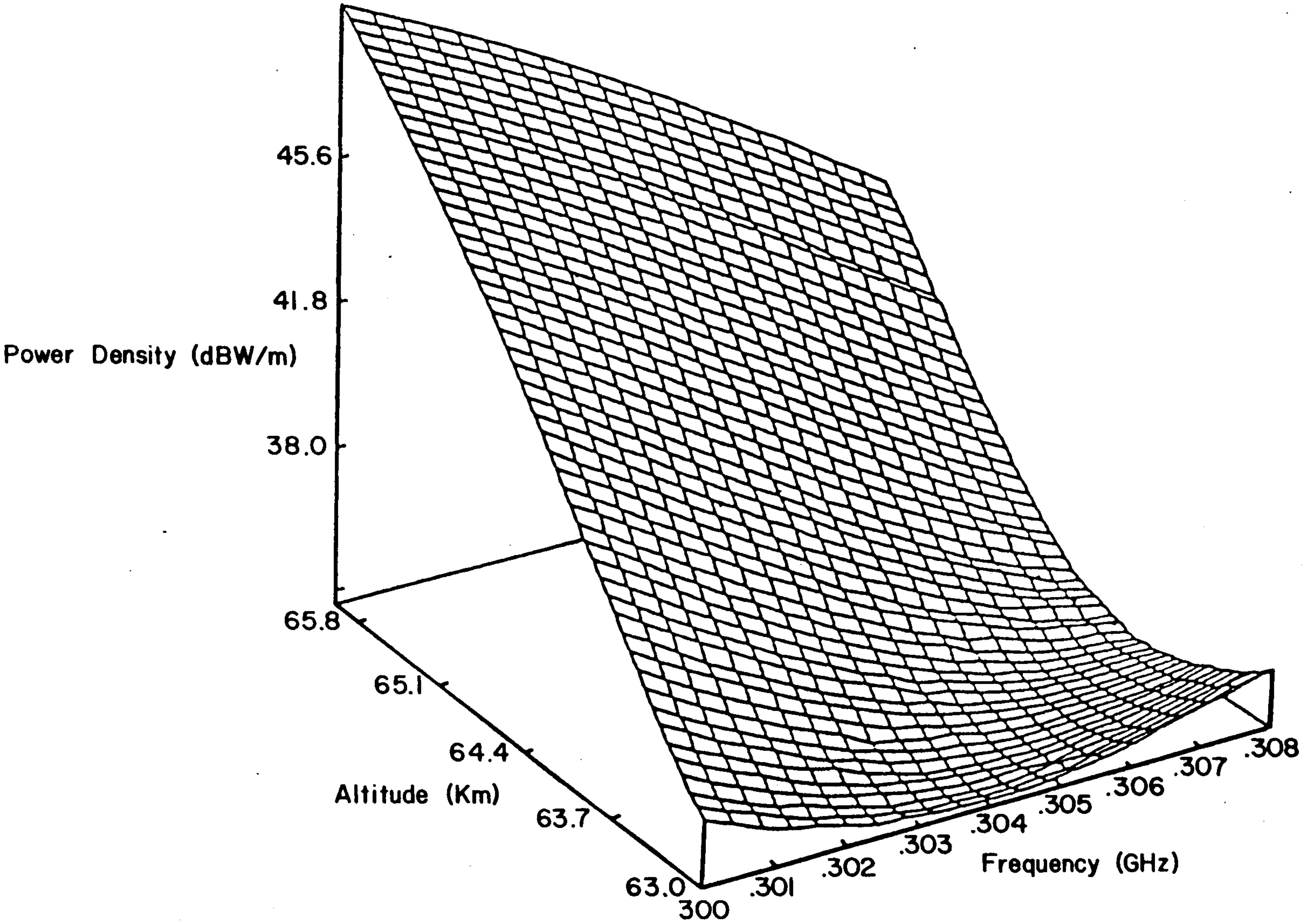

Effects of Frequency Chirping: Heater focused at 70 km with 300-308 MHz frequency chirping

FIG. 14

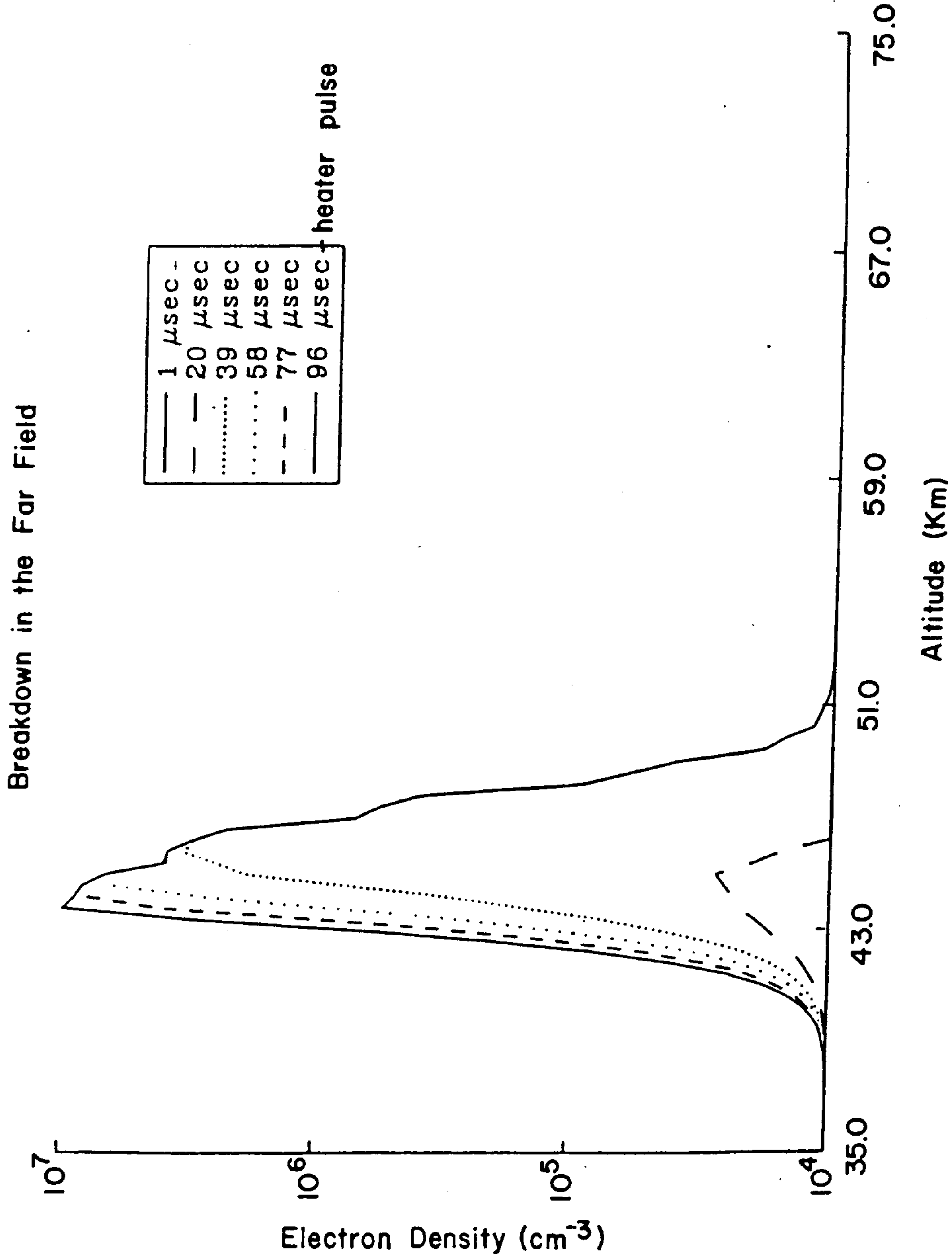

Vertical Electron Density Profile for a Non-Focused Heater Array, ERP=154dBW, frequency=300 MHz

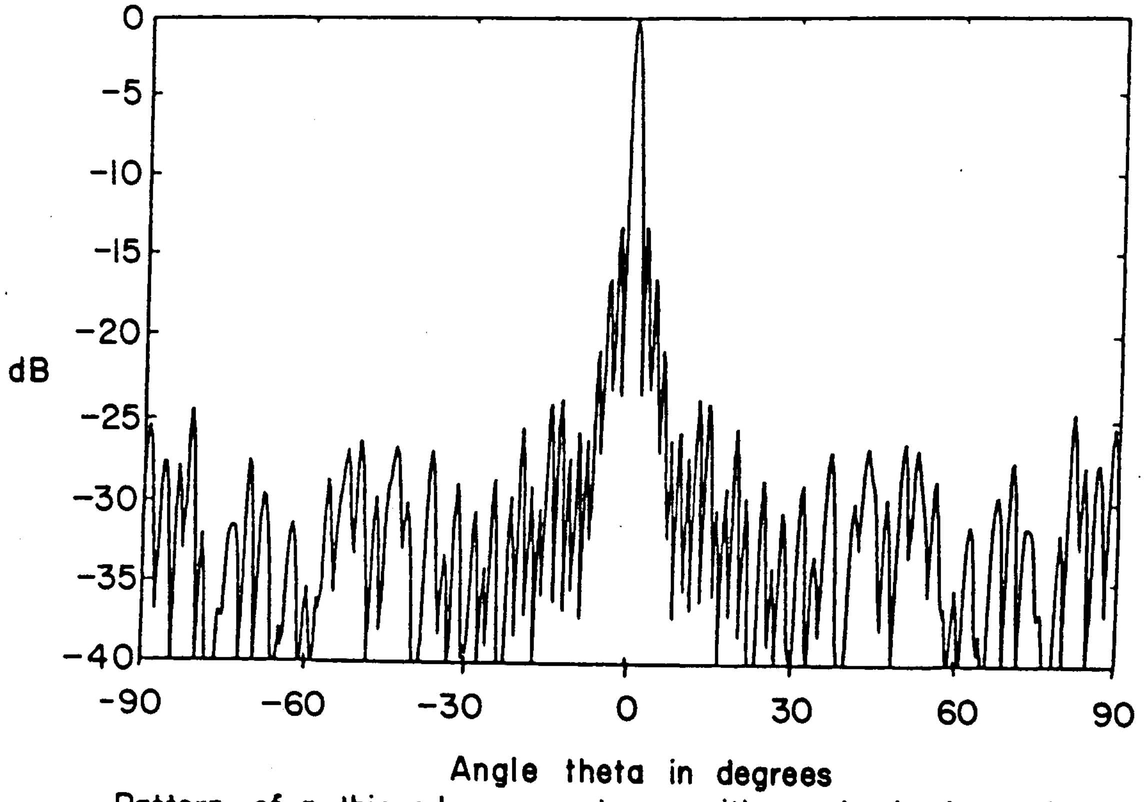

Pattern of a thinned array antenna with randomized spacing containing no grating lobes

FIG.15

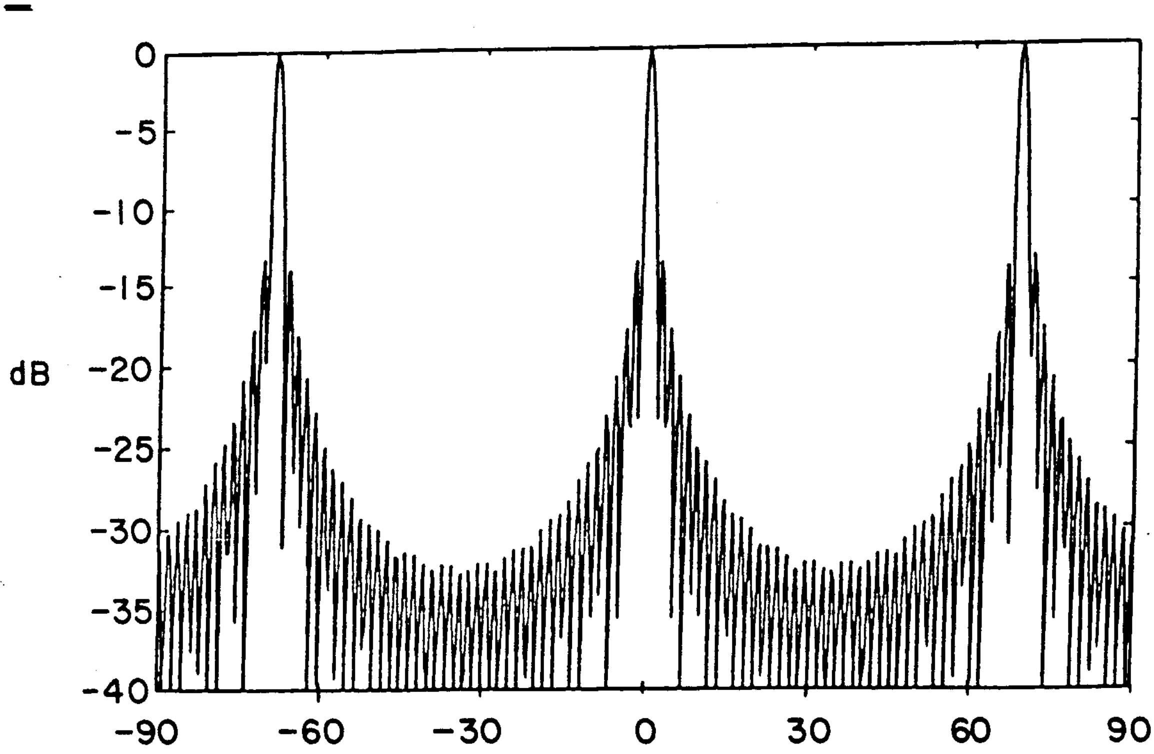

Angle theta in degrees
Pattern of a thinned array antenna with uniform spacing containing grating lobes

FIG.16

ARTIFICIAL IONOSPHERIC MIRROR COMPOSED OF A PLASMA LAYER WHICH CAN BE TILTED

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates to generation of a Artificial Ionospheric Mirror (AIM), or a plasma layer in the atmosphere. The AIM is used like the ionosphere to reflect RF energy over great distances.

2. Description of the Related Art

In the past, the technique of using the ionosphere as a mirror to reflect radio waves, or RF energy, has given Ham Radio operators the ability to send transmissions over long distances. This technique has also provided radar systems the ability to look "over the horizon." Variations and fluctuations in the ionosphere, however, can render the effectiveness of such communications uncertain. Thus, the desirability of creating controllable plasma layers in the atmosphere for communications purposes has been recognized. See, for example, U.S. Pat. No. 4,686,605 issued to Eastlund and U.S. Pat. No. 4,712,155 issued to Eastlund et al.

Previous experiments directed toward creating plasma layers for communications have suffered from the inability to control the inclination of the plasma layer so that signals could be transmitted and received from various ranges. In other words, while one could create a plasma layer in the atmosphere at a lower altitude than the ionosphere, point to point communications would be limited in range based on the reflection angles of the transmitted and reflected signals.

SUMMARY AND OBJECTS OF THE INVENTION

In view of the limitations of the related art it is an object of this invention to generate a plasma layer that could be angled or tilted with respect to the horizon in order to affect signal transmission range.

The present invention provides a system and method for generating a plasma layer at controlled altitudes and inclinations that acts as an artificial ionospheric mirror (AIM) to reflect RF signals. The AIM increases the range and predictability with which RF energy may be reflected off the AIM for communications purposes. More specifically, a tiltable AIM is created by a heater antenna controlled in phase and frequency. The heater antenna phase shift scans a beam to paint a plasma layer. The heater antenna continually refocuses at a higher altitudes by frequency shifting to tilt the plasma layer.

BRIEF DESCRIPTION OF THE DRAWINGS

Further details of the present invention are explained with the help of the attached drawings in which:

FIG. **1** shows creation of an AIM by a heater antenna and use of the AIM for tracking aircraft and reflecting radio waves.

FIG. 2 shows a typical heater array.

FIG. 3 shows the spacial relationship for a heater array used in defining heater array focusing.

FIG. **4** is a graph showing that power is at its upper bound at the antenna focal point.

FIG. 5 shows generation of plasma by a heater array.

FIG. 6 illustrates generation of a plasma layer by scanning a heater antenna.

FIG. 7 illustrates generation of a tilted plasma layer by scanning and refocusing a heater antenna.

FIG. **8** shows generation of a plasma layer using a heater antenna to scan with a line rather than a point.

FIG. **9** shows the phase corrections to move the antenna focal point from **60** Km to **61** Km.

FIG. **10** shows the frequency corrections to move the antenna focal point from **60** Km to **61** Km.

FIG. **11** is a plot of altitude v. distance location of plasma without frequency chirping.

FIG. **12** is a plot of altitude v. distance location of plasma with frequency chirping.

FIG. **13** shows the power density change after refocusing using frequency chirping .

FIG. **14** graphs the free electron density v. altitude for an unfocused array.

FIG. **15** shows an antenna power pattern without grating lobes.

FIG. **16** shows an antenna power pattern with grating lobes.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

FIG. **1** illustrates the creation and use of An Artificial Ionospheric Mirror (AIM) for tracking aircraft and reflecting radio waves. A heater antenna **1** radiates power causing avalanche ionization or breakdown releasing free electrons in the atmosphere to generate the AIM **2**. The heater antenna **1** is an array which can be used to focus energy at varying altitudes and elevations to tilt the AIM **2** using phase and frequency control. The AIM **2** simulates the ionosphere **3** which is also used to detect "over the horizon targets" **5**. In addition, the AIM **2** can reflect radio signals transmitted from a transmitter **6** to a receiver **7** over long distances.

A typical heater antenna is shown in FIG. **2**. It consists of an array of multiple active radiating elements **10** having their individual phase and frequency controlled from a control module **12**. The radiating element **10** is used here to represent all possible antennas, including, but not limited to, dipoles, slots, small or large horns, log-periodic antennas, large parabolic reflectors, etc.

FIG. **3** shows the spatial relationship for a focused heater array. To have the electric fields from all of the array elements focus, or arrive in phase at a distance R_o in the near field of the array, it is necessary to correct the phase of each element to compensate for the phase delay difference from the center element due to the additional phase path W_{ij}. If R_o is much larger than the maximum D_{ij} in FIG. **3**, then the phase delay can be approximated in wavelength to be:

$$W_{ij}=(D_{ij})^2/(4*R*g) \quad (1)$$

where g is the wavelength of the heater frequency. Equation 1 is referred to as the quadratic phase error. If this error is less than g/8 when the element (i,j) is on the outer edge of the array, then the distance R_o is said to be in the far field of the array.

In order to focus the array at R_o, it is necessary to have several wavelengths of phase error from the outer elements of the array. That is, the term "focus" is used in this context to mean that the electric field from the array is concentrated in a desired spatial region.

FIG. **4** shows the degree of focusing that can be accomplished. This is a vertical pattern of an array whose elements have been phase shifted to focus at 60 Km. The array has 400 elements with a total width and length of 2000 g. The peak of the pattern is determined by the $1/R_o^2$ dependence.

The AIM ionization layer is created by using this focused power to ionize an area in the atmosphere, as shown in FIG. 5. The microwave breakdown of air occurs where free electrons gain enough energy from an electric field to generate additional free electrons until no more can be generated, thereby resulting in avalanche ionization, or breakdown. This causes the generation of a plasma layer **21**. For example, a pulse of power from the heater begins to propagate in the z direction shown in FIG. **5**. As the field propagates, more free electrons are generated. A breakdown point descends vertically from the focal point of the propagating field giving thickness to the ionized layer, or plasma layer, until all ionization stabilizes. This "clamping" creates a thin vertical plasma layer.

Simulation results show that when an array **20** is focused at a point **22**, electric field power peaks at the focal point. Simulation results shows that given a focused microwave source avalanche ionization, or breakdown will occur at a power level 3–10 dB below the focal point power level.

To create an AIM, the heater array is focused at a desired altitude to maximize power at a point and thereby generate plasma. The heater antenna then "scans" the phase of each array element to move the focal point.

FIG. **6** illustrates creation of a noninclined AIM layer. The heater array **30** is first focused at point **31**. The heater array scans horizontally by phase shifting to a point **32** creating an avalanche ionization line **33**. Next, the heater array scans from a point **34** to a point **35** creating another avalanche ionization line **36**. The heater array continues this process to create an ionization plane or AIM layer.

In order to form an inclined AIM cloud, each new ionization line must occur at a slightly higher altitude. By altering the phase or frequency of the array elements, the focal point can be moved up in altitude, as described below.

FIG. **7** illustrates creation of an inclined AIM. The heater array **40** is first focused at point **41**. The heater array scans along the x direction to point **42** to generate avalanche ionization along line **43**. Next, as in creation of a non-inclined AIM, the heater array scans along the x and y directions directly below point **44**. The heater array **40** alters either phase or frequency to refocus to a higher altitude in the z direction to the point **44**. The heater array then scans along the x axis to point **45** to create the avalanche ionization line **46**. The heater array continues this process to create a tilted ionization plane or tilted AIM layer.

FIG. **8** shows that the preferred method of generating a plasma layer uses a heater antenna to scan with a line rather than a point. Scanning using a line is preferred since an AIM can be created in the atmosphere in less time. To create lines of ionization rather than points, a rectangular array **50** is used. In the array **50**, radiating elements are focused only along the plane of the long dimension of the rectangular array, creating a line of ionization **53**. The array is then scanned along the x-y axis and in altitude along the z axis to create another ionization line **55**. More ionization lines are similarly generated to form a tilted AIM layer.

In order to create a tilted AIM it is necessary to refocus the heater array at successively higher altitudes. Moving the focal point by changing the phase of each element of the heater in a very precise manner is not practical. Moving the focal point away from the initial location requires changing the phase on each element. The phase change required is near the rms tolerance level, typically 1 degree. FIG. **9** shows the required phase corrections to move the focal point from 60 Km to 61 km. Elements **5**, **10**, **15**, and **20** have distances 5d, 10d, 15d, and 20d, respectively from the center of the antenna, where d=25 meters It is clear from FIG. **9** that it is impractical to alter numerous antenna element phases to move the focal point to create tilted patches for AIM applications. 2000 elements may be required here to generate enough power to ionize the atmosphere.

The second method of refocusing is accomplished by first setting the phases of all elements for the initial focal point and then moving the focal point by changing the frequency rather than the phase. This frequency chirping method is less precise, but easier for hardware implementation because precise phases for 2000 elements need not be changed. FIG. **9** shows the required phase corrections to move the focal point from 60 Km to 61 Km. FIG. **10** shows that the focal point can be moved 100 meters by increasing the frequency approximately 1 Mhz. The resulting focal point power levels are not completely optimized, but simulation shows that there is less than a 0.1 db difference between the frequency shifted peaks and those obtained by phasing.

Tilting the AIM using frequency chirping is practical to achieve in a real system. FIG. **11** shows the plasma layer location with no frequency chirping. FIG. **12** shows the plasma location of the same heater creating a tilted AIM by increasing frequency from 550 MHz to 559.375 MHz while scanning horizontally. The result is a smooth patch with a 45 degree inclination.

While it is true that frequency chirping does not achieve the same power as phase focusing at the higher altitude, the difference for small frequency chirps is negligible. FIG. **13** shows actual power density data generated by a 300 MHz heater focused at 70 km with the frequency chirped to 308 MHz.

In the far field region, power meets its upper bound without focusing. For a far field or unfocused array, there is no way to raise the ionization altitude or create a tilted AIM. Ionization takes place at a point where there is enough power to initiate breakdown and where there is low enough neutral density (i.e. pressure). This usually occurs between 40 and 50 km altitude as shown in FIG. **14**. Consequently, a near field focused antenna is required to create a controlled AIM.

The focused pattern is a picture of constructive and destructive interference of the fields from the elements of the array. Other interference positions, or grating lobes, outside the focal point occur when some of the array elements add up in phase. The power of grating lobes can be kept below that of the main lobe, or focal point, by having a large number of elements in the array and spreading them out over the array aperture. This is called thinning the array. For square arrays having 400 elements or more grating lobes can be kept down by 20 db or more from the focal point.

The degree of focusing depends on the ratio of focal range to aperture size. The half power width from peak "V" can be approximated as:

$$V=2*g*(R_o/L)^2 \quad (2)$$

where L is the length of the array which is assumed square for equation 2. The power gradient at the half power point "grad(P)" can be approximated as:

$$grad(P)=10/V(db/meter) \qquad (3)$$

For an AIM it is desirable that the power gradient be high because this directly determines the gradient of the electron density of the generated ionized cloud. The electron density must be high to avoid RF losses caused by absorption. Hence V be small, preferably less than 2 Km. A heater frequency of 300 MHz and a focal distance of 70 Km would project an aperture size greater than 2 Km. Note in equation 2 that array size scales with the square root of frequency.

Since a near field antenna is required, the near field of the heater antenna may be required to extend to reach distant points. This is accomplished by increasing the array size. It may not be economically feasible to fill this entire aperture with elements, hence a thinned array is utilized.

If a thinned array had its elements uniformly distributed, there would be many grating lobes in the radiation pattern of the array. These grating lobes can be eliminated by randomly spacing elements. However, random spacing puts power from the grating lobes into the average side lobe level. If no new elements are introduced when the aperture is increased, then the peak power of the main lobe remains constant and the main lobe receives less of the total power as its beamwidth decreases. In order to preserve the efficiency of the heater array, grating lobes must be utilized in creating the AIM cloud or the array can not be heavily thinned. FIG. **15** shows an array with uniform spacing having grating lobes. FIG. **16** shows an array with randomized spacing which eliminates the grating lobes.

Although the invention has been described above with particular reference to certain preferred embodiments thereof, it will be understood that modifications and variations are possible within the spirit and scope of the appended claims.

What is claimed is:

1. A method for generating an AIM, comprising the steps of:

(a) creating avalanche ionization in the atmosphere using a heater antenna;

(b) refocusing said heater antenna to alter the altitude of said avalanche ionization by frequency shifting said heater antenna; and

(c) scanning said heater antenna to paint an avalanche ionization layer.

2. A method for generating an AIM as claimed in claim **1** wherein said heater antenna is focused in the near field.

3. An apparatus for generating an AIM comprising:

(a) a phased array heater antenna which is focused at an altitude to cause an avalanche ionization area to be created in the atmosphere;

(b) means for controlling frequency of individual radiators of said phased array heater antenna to refocus said altitude of said avalanche ionization area; and

(c) means for controlling phase of the individual radiators to scan said phased array heater antenna.

4. An apparatus for generating an AIM as claimed in claim **3** wherein said phased array heater antenna is focused to cause said avalanche ionization area to be substantially a line.

5. An apparatus for generating an AIM as claimed in claim **4** wherein said means for controlling phase moves said line substantially at a constant altitude and said means for controlling frequency moves said line to different altitudes.

6. An apparatus for generating an AIM as claimed in claim **4** wherein said phased array heater antenna is a rectangular array and said line is formed parallel to a long dimension of said rectangular array.

7. An apparatus for generating an AIM as claimed in claim **3** wherein said phased array heater antenna is focused to cause said avalanche ionization area to be substantially a point.

8. An apparatus for generating an AIM as claimed in claim **7** wherein said means for controlling the phase moves said point substantially at the same altitude and said means for controlling frequency moves said point to different altitudes.

9. An apparatus for generating an AIM as claimed in claim **3** wherein said phased array heater antenna is focused in the near field.

10. A method of generating an AIM comprising the steps of:

(a) focusing a phased array heater antenna at an altitude to cause an avalanche ionization area to be created in the atmosphere;

(b) controlling the frequency of individual radiators of said phased array heater antenna to refocus said altitude of said avalanche ionization area;

(c) controlling phase of the individual radiators to scan said phased array heater antenna.

11. A method of generating an AIM as claimed in claim **10** wherein said step of focusing causes said avalanche ionization are to be substantially a line.

12. A method of generating an AIM as claimed in claim **11** wherein said step of controlling phase moves said line substantially at a constant altitude and said step of controlling frequency moves said line to different altitudes.

13. A method of generating an AIM as claimed in claim **11** wherein said phased array heater antenna is a rectangular array and said line is formed parallel to a long dimension of said rectangular array.

14. A method of generating an AIM as claimed in claim **10** wherein said step of focusing causes said avalanche ionization area to be substantially a point.

15. A method of generating an AIM as claimed in claim **14** wherein said step of controlling phase moves said point substantially at the same altitude and said step of controlling frequency moves said point to different altitudes.

16. A method of generating an AIM as claimed in claim **10** wherein said step of focusing is performed in the near field.

* * * * *

US 20080191580A1

(19) **United States**

(12) **Patent Application Publication**

Deyo

(10) **Pub. No.: US 2008/0191580 A1**

(43) **Pub. Date: Aug. 14, 2008**

(54) HARMONIC ENERGY EXCHANGE DEVICE

(76) Inventor: **Harold Stanley Deyo**, Pueblo West, CO (US)

Correspondence Address:
HAROLD S DEYO, JR.
241 S. DOS PALOS DRIVE
PUEBLO WEST, CO 81007

(21) Appl. No.: **11/675,060**

(22) Filed: **Feb. 14, 2007**

Publication Classification

(51) **Int. Cl.**
H02N 11/00 (2006.01)

(52) **U.S. Cl.** **310/300**

(57) ABSTRACT

This invention converts inertial impulses into electric currents. Specifically, it converts impulses created by the impacts of high-energy particles from the Sun and other cosmic sources into the Earth's Magnetosphere and the varying D, E, F1 and F2 layers of its Ionosphere to controlled electric currents. This invention presents a new method of utilizing energy from the Sun and other sources of high energy articles as a virtually, inexhaustible, alternative-energy source for the world.

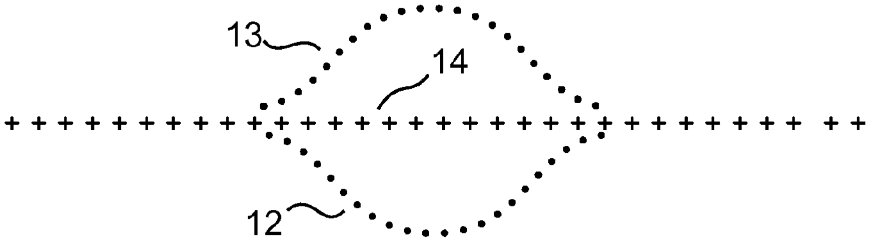

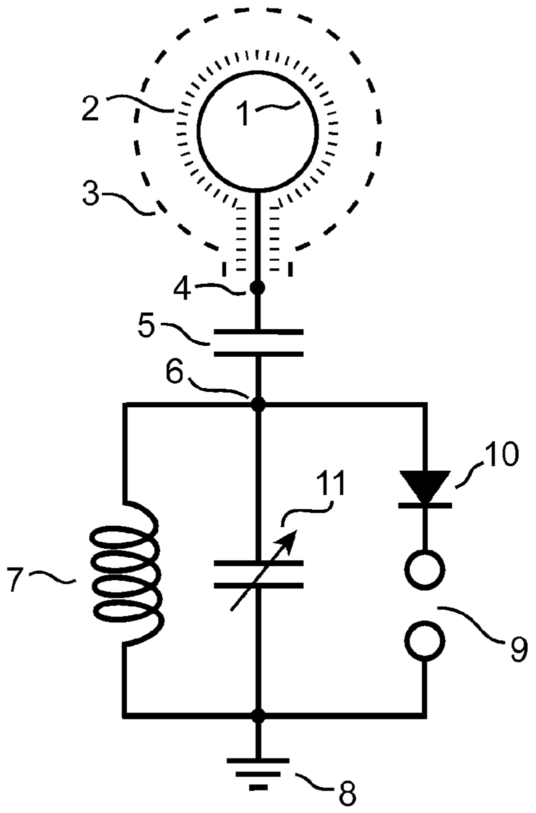

FIGURE 1

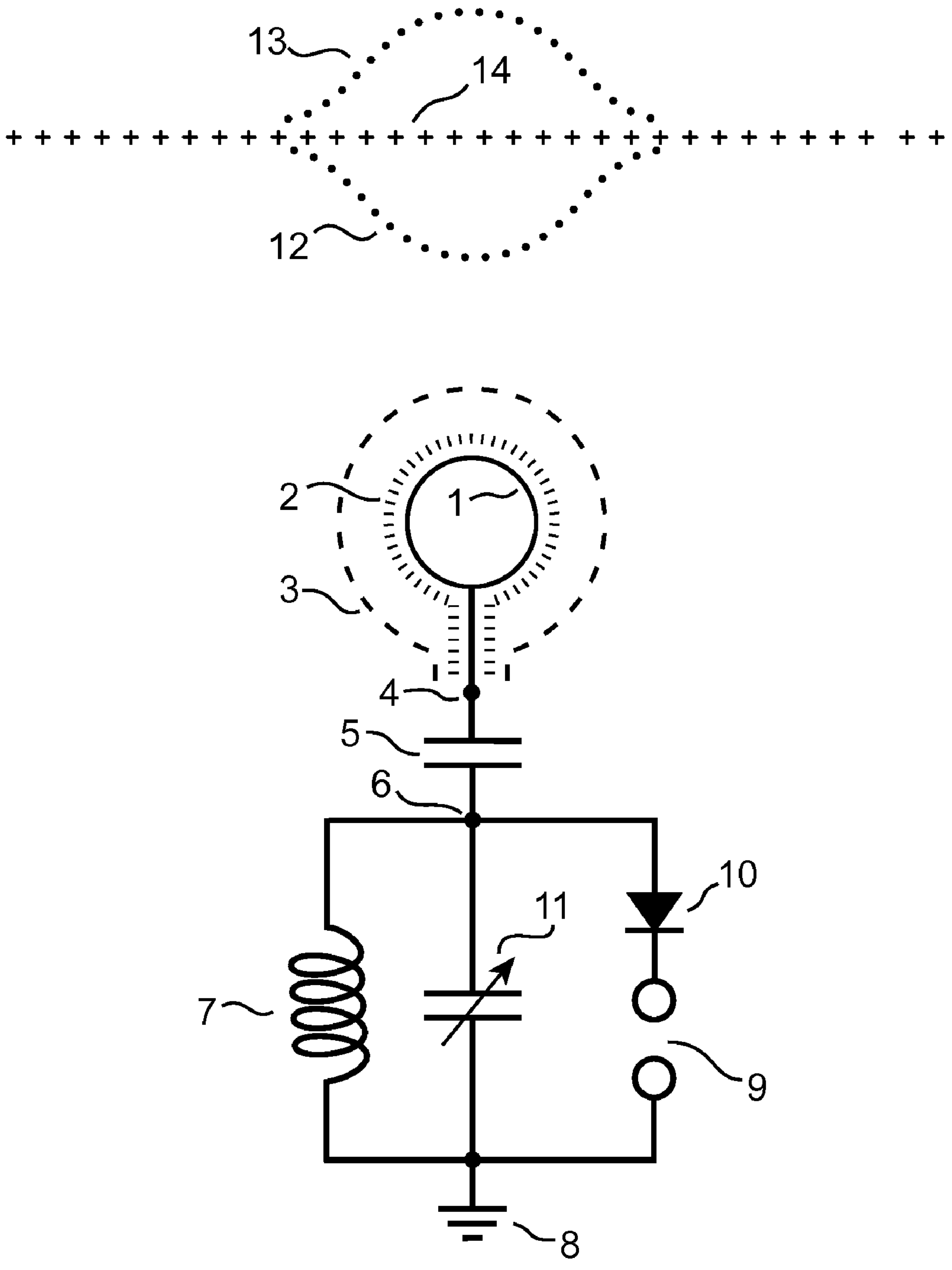
13
14
12
2
1
3
4
5
6
10
11
7
9
8

FIGURE 2

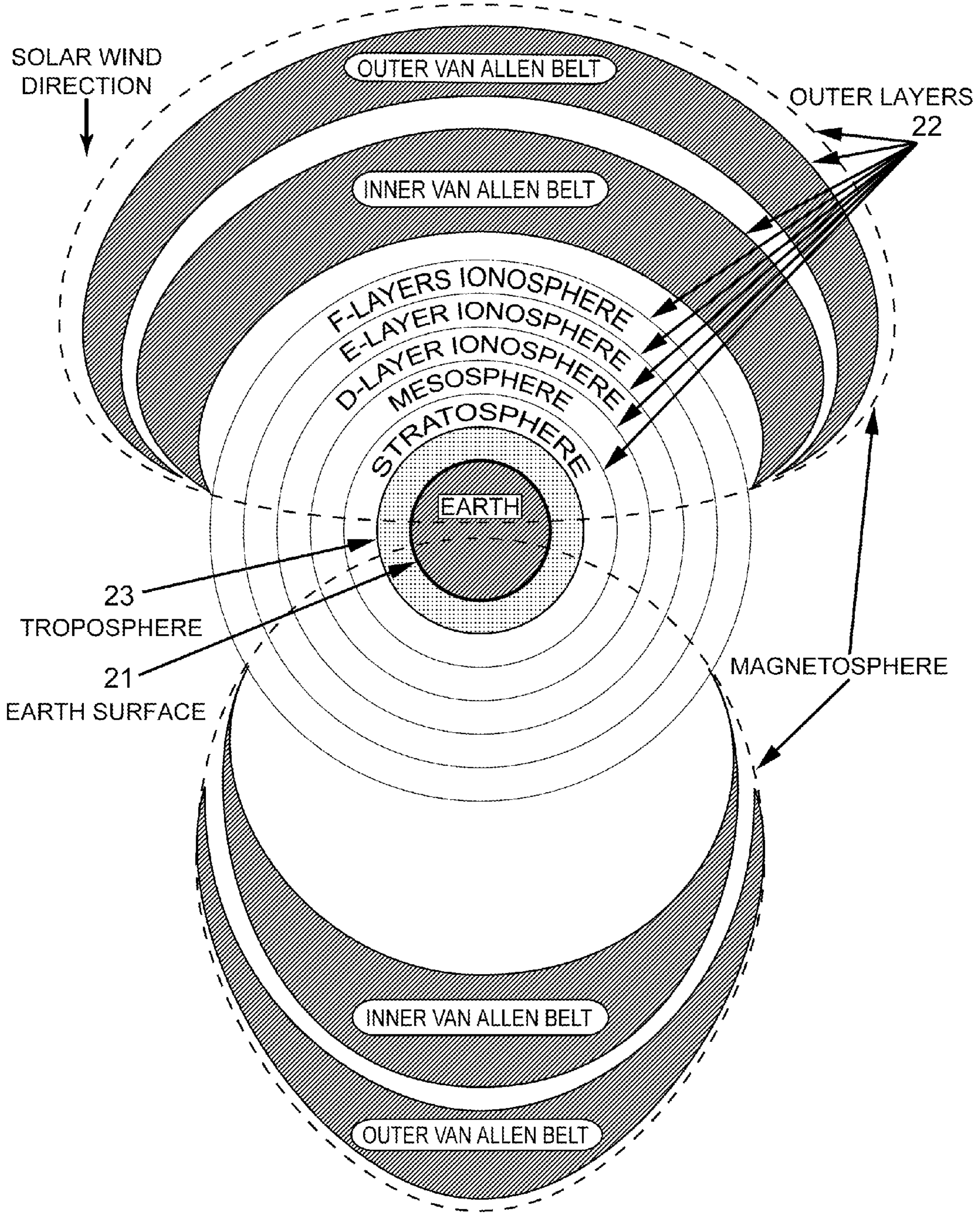
SOLAR WIND
DIRECTION
OUTER VAN ALLEN BELT
OUTER LAYERS
22
INNER VAN ALLEN BELT
F-LAYERS IONOSPHERE
E-LAYER IONOSPHERE
D-LAYER IONOSPHERE
MESOSPHERE
STRATOSPHERE
EARTH
23
TROPOSPHERE
21
EARTH SURFACE
MAGNETOSPHERE
INNER VAN ALLEN BELT
OUTER VAN ALLEN BELT

FIGURE 3

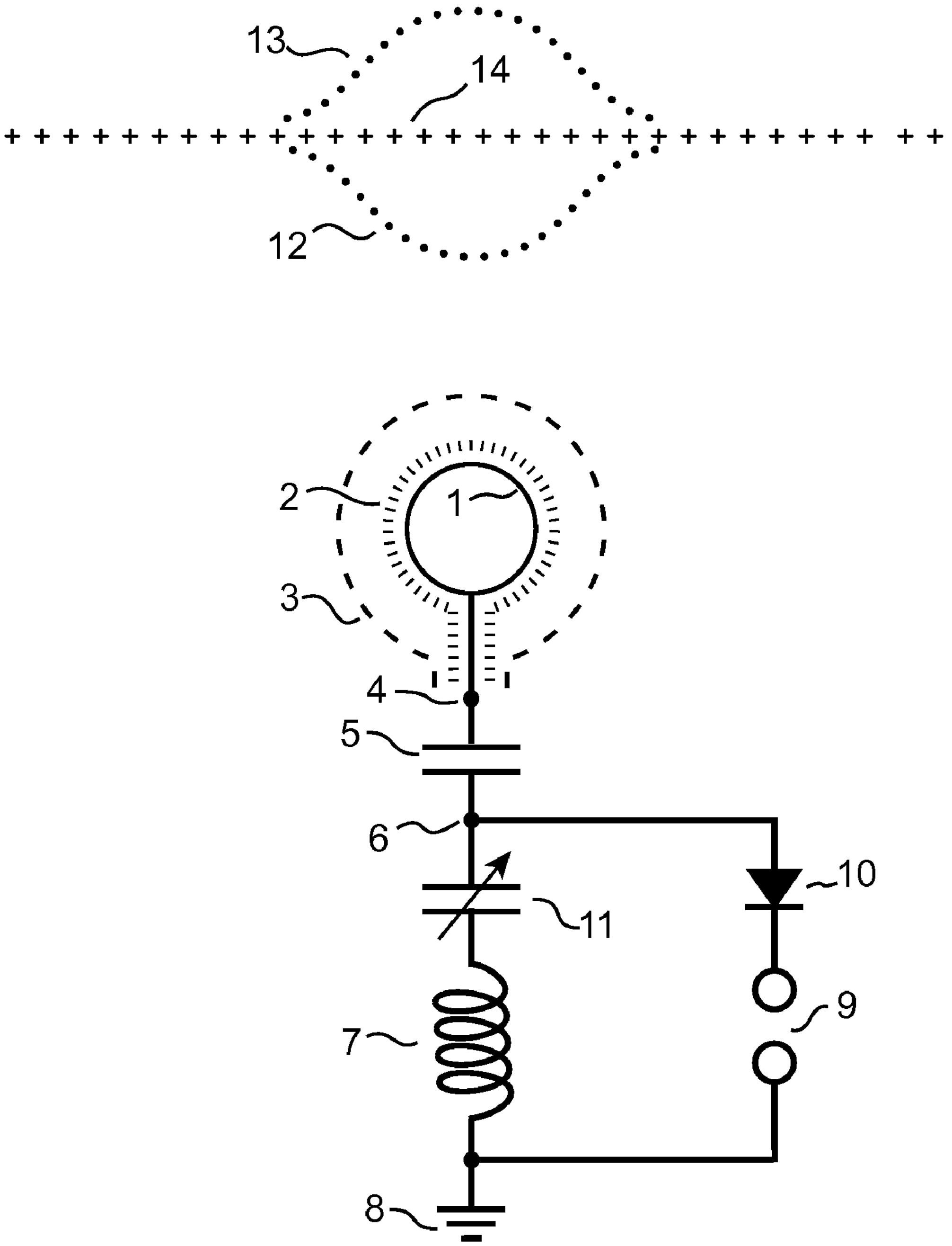
13
14
12
2
1
3
4
5
6
10
11
9
7
8

HARMONIC ENERGY EXCHANGE DEVICE

BACKGROUND OF THE INVENTION

[0001] 1. Field of Invention

[0002] This invention relates to the conversion of impact energies created by the collision of high-speed cosmic particles and electromagnetic radiations with "Earth's Outer Layers" to produce inertial waves in the dielectric Troposphere which are subsequently converted into electricity by this invention.

[0003] The term "Earth's Outer Layers" refers to: Earth's Magnetosphere, Van Allen Belts, Ionosphere, Mesosphere, and Stratosphere as illustrated in FIG. **2**

[0004] 2. Description of the Prior Art

[0005] One day while reading an old scientific magazine I saw a small article on the research of Dr. John Trump of MIT (the basis for an electrostatic generator patented later by Onezime P. Breaux in U.S. Pat. No. 4,127,804). The article talked about a series of experiments which Dr. Trump had performed with a 2-plate, capacitor in a vacuum dielectric. Briefly, he discovered that by varying the distance between the two plates one could produce current flows in one direction or the other between either of the plates and ground.

[0006] The Solar Wind strikes "Earth's Outer Layers" constantly as do many other cosmic particles. At any given instant of time, the vector product of these impacts produces either a net pressure increase or decrease in the Troposphere. This creates random waves of pressure in the "Earth's Outer Layers" as one charged "plate" moves in relation to the oppositely-charged Earth's surface as the other "plate". This variation of pressure in the "Earth's Outer Layers" is equivalent to moving an outer "plate" back and forth toward the Troposphere and the Earth's surface as the inner "plate"—thus giving rise to variations in voltage on both "plates". In this case the Troposphere (see FIG. **2**) acts as the dielectric medium instead of the vacuum in the Trump experiments. Furthermore, the Troposphere, itself, is also compressed and decompressed by these random waves of pressure on the "Earth's Outer Layers". Thus, I reasoned if one could create a charged envelope or field of sufficient voltage within the Troposphere, one could convert these random pressure waves in the dielectric Troposphere into current flow on the Earth's surface.

[0007] As "Earth's Outer Layers" surround the planet, any impact waves propagate throughout all of them when they are present—even to the nighttime side of the planet. Thus, I also reasoned one could extract power from these impacts anywhere on Earth's surface or in its atmosphere whether it be day or night. The pressure waves on the dark side of Earth would be less energetic than those occurring on the daytime side because the nighttime layers of the Ionosphere are compressed so much that the D-Layer of the Ionosphere disappears at night and the F1 and F2-Layers of the Ionosphere combine into one F-Layer. I calculated the available energy from these impact waves would be significantly less by 30-45% depending upon one's location on the night side of the planet in respect to the terminator.

[0008] Many years ago around 1900 Dr Nikola Tesla built and tested a device to extract energy from the Sun using an apparatus which forms part of my own invention (see U.S. Pat. Nos. 685,957 an 685,958 dated 5 Nov. 1901).

[0009] Subsequently, Dr. Thomas Henry Moray, another inventor and student of these Tesla patents made a device which apparently accessed the same source of energy but with a method different than my own. As Dr. Moray was never granted a patent for his device I cannot be certain that his work is considered prior art but I list it here as part of my due diligence. Apparently, he had trouble explaining the source of energy his device was converting to the patent reviewer and was, thus, not granted a patent for his impressive work. Dr. Moray's public disclosure of certain aspects of his invention are public knowledge through his published lectures and his book, "The Sea of Energy in Which the Earth Floats'—published in 1931.

[0010] Since the beginning of the 20th Century mankind has been looking for new sources of electric power to feed the ever-increasing energy demands of the human population. In the last half of the 19th Century coal, whale oil, human and animal labor, moving water, wind and wood were main sources of energy. However, in the first quarter of the 20th Century mankind began to use more electricity produced by hydroelectric generators and other generators producing electricity by combustion of fossil fuels. With increasing use of gasoline engines and the increasing use of electrical appliances the demand for electricity accelerated the addition of heat and greenhouse gases into the Earth's ecosystem.

[0011] In the last half of the 20th Century many forward-thinking individuals began looking for alternate energy sources to reduce the emissions of greenhouse gases. All of the alternate energy systems investigated and/or developed in the last 56 years have had significant drawbacks to their usage as the main energy supply for the needs of humanity.

[0012] Some of those alternate energy systems were nowhere near as efficient as the existing systems. Some of the alternate energy systems still introduced extra heat into the environment like the existing fossil fuel and nuclear energy systems. And, yet, others produced even more greenhouse gases than the existing energy systems. Some of the new systems used food resources to produce combustible fuels which, consequently, reduced the food resources of the planet. Among the leaders in the new energy sources was and is nuclear energy which produces extra heat, toxic wastes and materials which can be used in nuclear weapons. Hydroelectric energy sources do not add heat to the ecosystem but they are not an inexhaustible resource as they depend upon rainfall which is diminishing as global warming increases. Wind energy resources do not add heat to the ecosystem but they are not constant and require expensive conversion and storage equipment.

[0013] The ideal energy source for mankind will not add heat to the planet's ecosystem, will not produce by-product green house gases; will not deplete our food and water resources; will not produce toxic wastes; however, it will be portable, self-replenishing and constantly available in real time anywhere on Earth and in useful quantities.

[0014] Our major source of natural energy on Earth is the Sun. Energy from the Sun enters the Earth's ecosphere by particle bombardment, radiated electromagnetic energy, gravitational variations and magnetic processes. Mankind has developed various methods of capturing energy from the Sun already. Silicon-based solar cells convert light in a narrow bandwidth into electron flows. These cells are—at most—only 50% efficient and only work when the Sun is shining on them and at an optimum angle. Wind turbines only work when there is wind to power them and are—for the most part—not portable. Devices to convert wave action in the oceans only work when there are waves created by the wind and tidal effects caused by gravitational variations; and they are not portable enough to be used in common transportation devices.

[0015] I concluded if mankind could find an inexpensive method of obtaining electricity from the Sun in some form that worked both day and night and which was already flowing into the Earth's ecosystem, we might be able to develop a clean and self-replenishing energy source which would add no extra heat to our ecosystem, would not alter our planetary albedo and might well be made in a portable form.

SUMMARY OF THE INVENTION

[0016] The Harmonic Energy Exchange Device (or "HEED") offers an interim solution to Earth's dwindling fossil fuel resources. It utilizes a previously untapped energy resource which exists throughout the known Universe. By the very nature of the invention, it only interrupts the flow of energy from the Sun and all the other cosmic sources of high-energy particles on their normal journey to the Earth by way of The Outer Layers.

[0017] The invention does not add energy to the normal thermodynamic equilibrium of the planet and its associated outer layers. The invention does not produce wastes—toxic or otherwise. It does not produce harmful gases; and in some embodiments it could even assist in the rebuilding of our thinning ozone shield. Use of this invention will eventually remove mankind's dependence on fossil fuels and create new occupations. It will be cheap enough that homeowners will be able to install their own HEED and sell back energy to the grid until such time as every home has their own HEED and civilization has completely weaned itself from the use of fossil fuels and nuclear energy in its current form.

BRIEF DESCRIPTION OF DRAWING

[0018] FIG. **1**: Is the preferred embodiment of the circuit diagram of my invention incorporating references to external stimuli which are used to activate the circuit. It represents a parallel tuned circuit to store the energy.

[0019] FIG. **2**: Is an illustration of the Earth and it immediate environment as a global capacitor where The Outer Layers form one plate and the Earth's surface form the other plate and both are separated by the Troposphere as a dielectric medium.

[0020] FIG. **3**: Is an alternate embodiment of the circuit diagram of my invention incorporating references to external stimuli which are used to activate the circuit. It represents the use of a series tuned circuit to store the energy.

DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT OF THE INVENTION

[0021] Be it known that I, Harold Stanley Deyo, Jr, a citizen of both The United States of America and The Commonwealth of Australia, residing in the community of Pueblo West in Pueblo County, Colorado have invented an Harmonic Energy Exchange Device which converts dynamic pressures in the ambient media around Earth to controlled electric currents.

[0022] This invention as represented in FIG. **1** is unique in that it is designed to extract electricity from random pressure waves propagated in the Troposphere **14** by impacts of the solar wind and other cosmic particles with the Earth's "The Outer Layers" **22** of FIG. **2**.

[0023] In FIG. **1**, the Troposphere **14** in its rest state is represented by the line of "+" marks **14**. The convergent or compression state **12** of the random waves in the Troposphere **14** is represented by the lower line of dots while the divergent or decompression state **13** of the random waves in the Troposphere **14** is represented by the upper line of dots. The preferred embodiment of my invention FIG. **1** is a method of coupling a parallel, resonant, electrical circuit to these random pressure waves to extract electricity from them.

[0024] As Dr John Trump's research and Onezime P. Breaux's U.S. Pat. No. 4,127,804 show, when one plate of a charged capacitor is moved closer to or farther from the other plate of that capacitor a voltage change appears on both plates of that capacitor. Furthermore, the electric field gradient between those plates changes as the plates are moved in this manner. Thus they have described a method of converting mechanical energy into electrical current by simply varying the plate spacing over time which can be expressed as ds/dt where "ds" is the change in spacing and "dt" is the change in time.

[0025] As illustrated in FIG. **2**, The Outer Layers can be viewed as one plate of a capacitor comprised of the Earth's surface itself as one plate **21** and Earth's outer layers as the other plate **22** where the Earth's Troposphere **23** serves as the dielectric medium separating the two plates. There is a charge between these plates which varies extremely rapidly but not with a constant period. At any given instant, the vector product of all the impacts by the cosmic particles with The Outer Layers **22** will create a pressure wave in The Outer Layers **22** which will manifest in the Troposphere **23**. Whatever that vector product is, it will alter the field gradient in the dielectric Troposphere **23**. That alteration will either increase or decrease the effective spacing between "plates" **22** and **23**.

[0026] In order to capture and convert the resulting voltage variations produced at or near the Earth's surface **21** of FIG. **2**, this invention creates its own, localized stress field **3** of FIG. **1** established in the Troposphere **23** of FIG. **2** with a charged, conductive surface **1** of FIG. **1** which is encased in high-voltage insulation **2** of FIG. **1** and extending into the Troposphere **23** of FIG. **2**.

[0027] In FIG. **1**, a series of high-voltage, starting pulses is applied across points **4** and **6** on opposite sides of a capacitor **5** to create the localized stress field **3**. As these high-voltage starting pulses are applied, the parallel resonant circuit formed by inductor coil **7** and variable capacitor **11** both referenced to ground **8** is stimulated into resonance within the bandwidth determined by the values of these circuit components. Tuning of this circuit is effected through variable capacitor **11**. The high voltage charge on conductive surface **1** of FIG. **1** is maintained by the parallel tuned circuit formed by inductor coil **7** and variable capacitor **11**.

[0028] Then as the random, pressure waves propagate throughout the Troposphere **23** of FIG. **2**, the localized stress field **3** of FIG. **1** is oscillated by the compression wave front **12** and the decompression wave front **13**—both of FIG. **1** which creates voltage changes on capacitor **5**.

[0029] In FIG. **1**, the resulting voltages changes on capacitor **5** will add energy to the parallel resonant circuit formed by inductor coil **7** and variable capacitor **11** which acts as a tank circuit to store the energy which has been passed to it. As energy builds in the parallel circuit the voltage of the circuit rises until a spark discharge occurs across the gap **9** also referenced to ground **8**. The current will only discharge in one direction as determined by diode **10**. The circuit will work without diode **10** but a diode is used here as one method to keep from draining all the energy from the tank circuit when a discharge occurs.

[0030] This resonant circuit can be tuned to various bandwidths to maximize the efficiency of the conversion process depending upon location of the device, time of day, temperature variations, relative humidity and other variables in the ambient Troposphere **23** of FIG. **2** around the conductive surface **1** of FIG. **1**. A resistive load to extract power from the resonant circuit can be used instead of the spark gap **9** of FIG. **1**.

[0031] This preferred embodiment uses a parallel, tuned circuit to access a wide range of frequencies usually to be found in the range of 4.5 to 7 MHz. This range encompasses the major, naturally-occurring, resonant frequencies found in the ionosphere.

[0032] A second embodiment of this invention as shown in FIG. **3** replaces the parallel tuned circuit formed by inductor coil **7** and variable capacitor **11** in FIG. **1** placed in parallel to each other and referenced to ground **8** in FIG. **1**. This second embodiment forms a series tuned circuit formed by inductor coil **7** and variable capacitor **11** in FIG. **3** placed in series to each other and referenced to ground **8** of FIG. **3**. It differs from the preferred embodiment only in the placement of variable capacitor **11** of FIG. **3** so that it is in series with coil **7** of FIG. **3**.

[0033] This embodiment limits the frequency range and, hence, the energy the system will store when compared to the preferred embodiment. It produces higher voltages across the spark gap **9** of FIG. **3** than those produced in the preferred embodiment across the spark gap **9** of FIG. **1**.

1. An apparatus for coupling to the pressure waves created in the Earth's Magnetosphere and the varying D, E, F1 and F2 layers of its Ionosphere by the impacts of various high-energy particles from both our Sun and other cosmic sources.

2. An electrical circuit for harmonically converting the pressure waves described in claim (**1**) into periodic electric current discharges.

* * * * *

rexresearch.com

Harry STOCKMAN

RF / Microwave Motors

Adobe MAKE ANYTHING MORE CREATIVE WITH CREATIVE CLOUD
Students save 60% ›

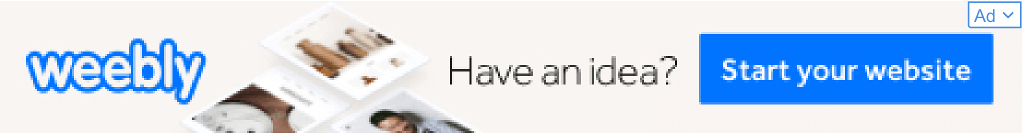
Ad
weebly Have an idea? Start your website

Ad
THE HITCHHIKER'S GUIDE TO PCB DESIGN
Find the Answers to Life, the Universe, and Everything PCB Get the Guide
EMA Design Automation

These motors can be driven by radio power. Here are several articles detailing the design and operation of push-pull, parametric, Poynting Vector and other RF/ microwave motors of extremely simple construction.

(Click on images to enlarge)

http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/5/31066/01445336.pdf?isnumber=31066&prod=JNL&arnumber=1445336&arSt=+1380&ared=+1381&arAuthor=Garnier%2C+R.C.%3B+Ishii%2C+K.

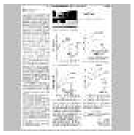

R.W. Steinhoff & T.K. Ishii : *IEEE Transactions on Microwave Theory & Techniques* 52 (11), p. 438 (September 1966); "Pickup Antennas for Waveguide Motors".

---

http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1448878

...

R.J. Dewey : *Proc. IEEE* 57 (2) : 248-9 (February 1969) : "Microwave Motor"

Harry E. Stockman : *Proc. IRE* (September 1963), pp. 1253-4 ; "Parametric Motor Energized by Radio-Frequency Field"

Harry E. Stockman : *Proc. IEEE* (September 1961) [fragment]

Wayne Anderson : *Proc. IEEE* (August 1969), pp. 1440-1 : "Microwave-to-Light Motors"

...

R. C. Garnier : *Proc. IEEE* (November 1964), pp. 1963-4 : "Microwave Motor"

John Schuder : *Proc. IEEE* (February 1963), pp. 399-400 : "Parametric Motor Energized by Radio-Frequency Field"

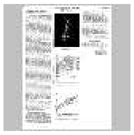

R.C. Garnier & T.K. Ishii : *Proc. IEEE* (February 1965), p. 178 ; " Microwave Motor Utilizing a Double Antenna and a Double Coil"

US3233157
Electric Field Motor

1966-02-01
STOCKMAN HARRY E
Classification: - international: H02N1/00; H02N1/00 - European: H02N1/00B1

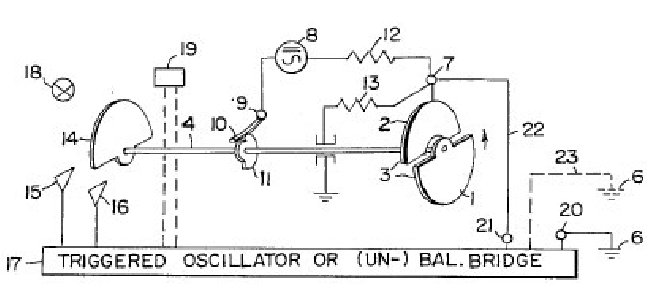
18
19
8
12
7
9
13
10
4
2
14
22
11
3
1
23
6
15
16
21
20
6
17
TRIGGERED OSCILLATOR OR (UN-) BAL. BRIDGE

FIG.1

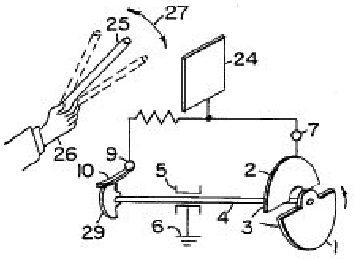
25
27
24
7
26
9
10
5
2
29
6
4
3
1

FIG. 2

...

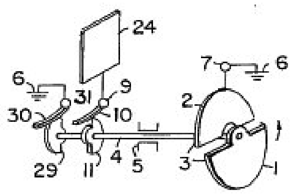
24
7
6
6
31
9
30
10
2
29
11
4
5
3
1

FIG. 3

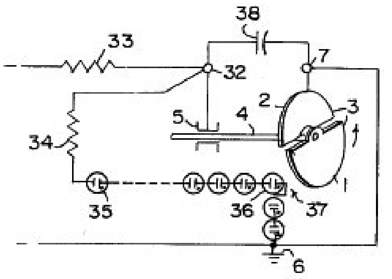
38
33
32
7
2
3
5
4
34
1
35
36
37
6

FIG.4

...

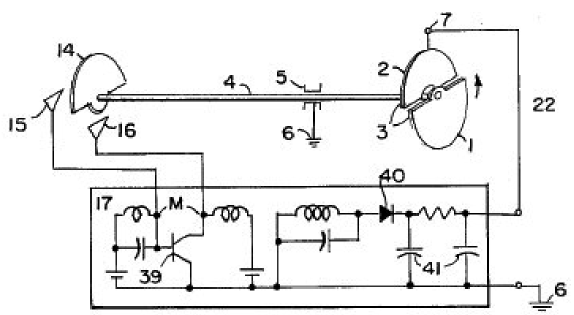
7
14
4
5
2
15
16
6
3
1
22
40
17
M
39
41
6

FIG.5

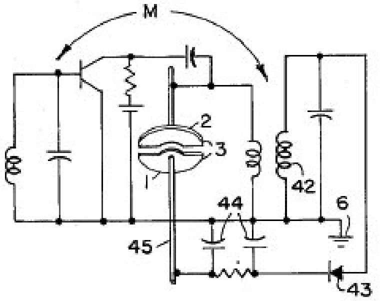
M
2
3
1
42
44
6
45
43

FIG.6

...

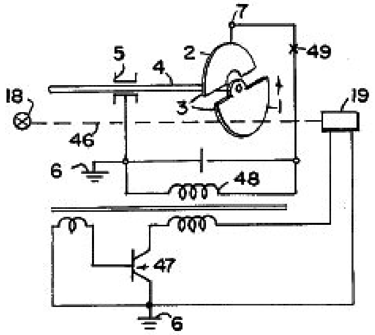
7
5
2
4
49
18
3
1
19
46
6
48
47
6

FIG.7

US3249944
Electromagnetic Transduction Utilizing Radiant Energy

1966-05-03
STOCKMAN HARRY E
Classification: - international: G01S13/75; G01S13/00 - European: G01S13/75

FIG.1

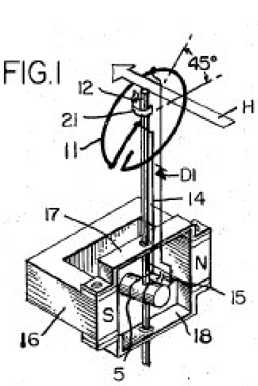
45°
12
H
21
11
D1
14
17
N
15
S
16
18
5

...

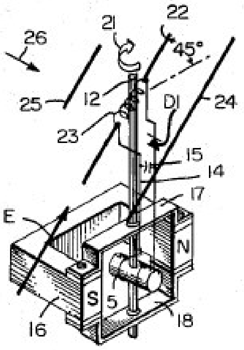
21
22
26
45°
25
12
D1
24
23
15
14
17
E
N
S
5
16
18

FIG.2

...

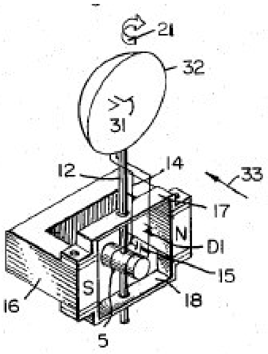
21
32
31
12
14
33
17
N
D1
15
16
18
5

FIG.3

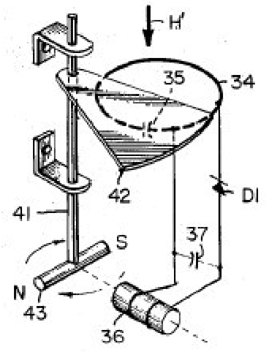
H'
35
34
42
D1
41
37
S
N
43
36

FIG.4

...

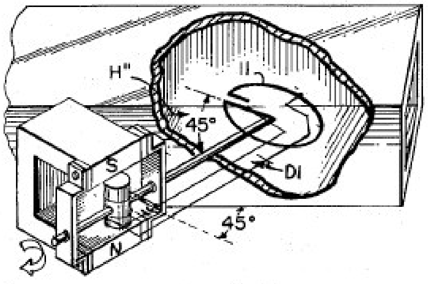
H"
11
45°
S
D1
N
45°

FIG.5

US3241018
Tunnel Diode Electromechanical Movement

STOCKMAN HARRY E
1966-03-15

FIG.1

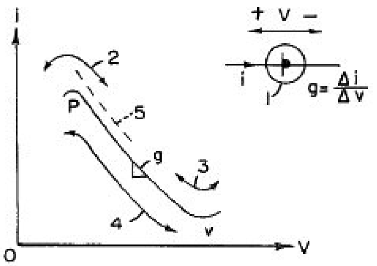

...

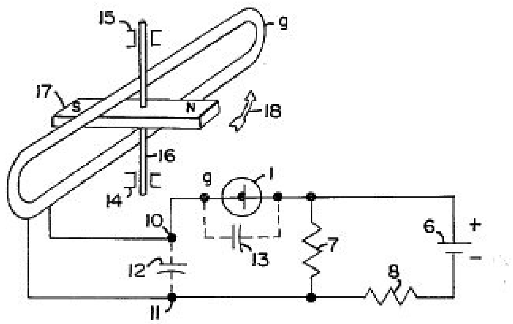
15
g
17
S
N
18
16
14
10
g
1
13
7
6
+
−
12
8
11

FIG. 2

FIG.3

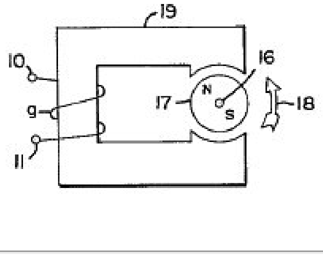
19
10
16
g
17
N
S
18
11

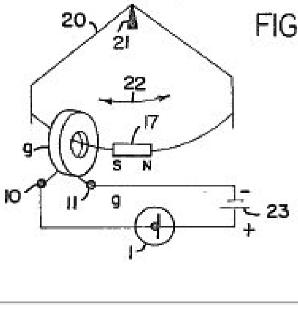
20
21
22
17
g
S N
10
11
g
23
−
+
1

FIG.4

...

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

Truly Global Daily News
Interpreted by Experts

Posted on January 12, 2010 by Tessa Henderson

Harvesting power from WiFi signals

RCA
Airnergy
WiFi Hotspot
Power Harvesting
Battery Pack

At the recent 2010 International Consumer Electronics Show (CES), RCA introduced the Airnergy Charger which they claim is able to convert the energy in WiFi signals into electricity to power small devices. RCA demonstrated at CES that the Airnergy Charger was able to charge a BlackBerry from 30% to full charge in around 90 minutes using only ambient WiFi signals at the show for power.

Perovskite Photovoltaics 2016-2026

The pocket sized Airnergy Charger has an internal battery for electricity storage. Power can be extracted from WiFi signals as the charger is carried around, or it can be left near a WiFi router at night to collect and store power for use the next day. Although this device would be a boon to the energy harvesting industry, some detractors suggest that the device would not generate a useful amount of power unless placed in the immediate vicinity of a high power broadcast.

RCA
Airnergy
WiFi Hotspot
Power Harvester

RCA aims to produce a battery with WiFi harvesting technology but in the meantime they say the USB Airnergy Charger will be available later this year for US$40.

Parasitic use of WiFi signals has already been employed by so-called WiFi Real Time Locating Systems RTLS ("Second Generation Active RFID") as detailed in the IDTechEx report Real Time Locating Systems 2009-2019. Here principles such as Received Signal Strength Indication RSSI from several WiFi nodes are used in conjuction with radio mapping to permit the tag to be located in 3D. However, the tag always contains a battery because the WiFi emissions are inadequate for the necessary data processing and transmission of derived information.

Thermoelectric Energy Harvesting 2016-2026

Image source TechCat

For more read : Energy Harvesting and Storage for Electronic Devices 2009-2019

and attend Energy Harvesting & Storage Europe and Wireless Sensor Networks & RTLS Summit 2010

Tags: Energy Harvesting

Print Comment Forward Tweet Like Pin reddit Plus Share

Stay updated with developments: subscribe to our newsletter

PREVIOUS NEXT

Authored By: Tessa Henderson

Related Posts

Embedded wireless power based on RF energy harvesting

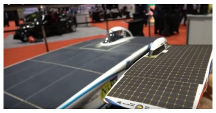

IDTechEx Show! Features Winning Solar Racing Car USA 2016

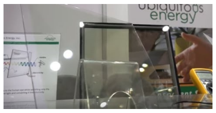

Transparent Solar Technolgy

Vibration Harvesting Technology by Star Micronics

Related IDTechEx Events

Printed Electronics EUROPE
Berlin, Germany
10 - 11 May 2017

ENERGY HARVESTING EUROPE OFF-GRID RENEWABLE ENERGY
Berlin, Germany
10 - 11 May 2017

Wearable EUROPE
Berlin, Germany
10 - 11 May 2017

Graphene & 2D Materials
Berlin, Germany
10 - 11 May 2017

More IDTechEx Events

Related Research Reports from IDTechEx

Sensors for Robotics: Technologies, Markets and Forecasts 2017-2027

Printed, Organic & Flexible Electronics Forecasts, Players & Opportunities 2017-2027

Electric Boats and Ships 2017-2027

Electric Buses 2017-2027

More IDTechEx Reports

Leave a Comment

Your email address will not be published.

Name *

Email *

Comment *

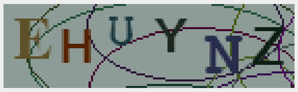

Please enter the numbers and/or letters in the image.

Required fields are marked *

Post Comment

Stay updated with developments: subscribe to our newsletter

Search...

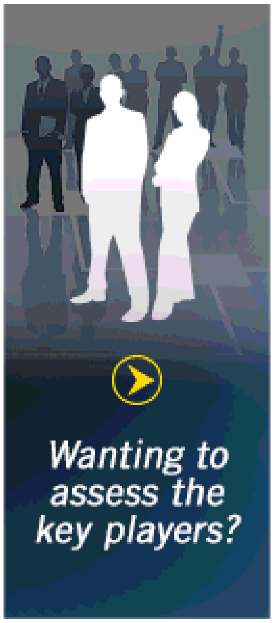
Wanting to assess the key players?

"A significant part of the $0.7tr EV market will be fully autonomous"

Popular Recent Comments

Researchers improve production of thin, efficient solar cells

Researchers are making strides in material development that may open up new possibilities for thin, wearable electronics and other vital power applications.

Self-healing via ion-dipole interaction
Transparent, self-healing artificial muscle
Healed

Self-healing, transparent, highly stretchable material

Scientists have developed a transparent, self-healing, highly stretchable conductive material that can be electrically activated to power artificial muscles and could be used to improve batteries,…

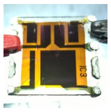

Solved: mystery holding back development of next-gen solar cells

Scientists have identified an unexpected cause of poor performance in a new class of flexible and cheap solar cells, bringing them closer to market.

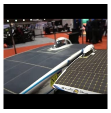

IDTechEx Show! Features Winning Solar Racing Car USA 2016

Video Interview with the Nuon Solar Team at the IDTechEx Show! in Santa Clara, USA (November 16-17 2016)

Solar powered plane to reach the edge of space

he SolarStratos Mission began in 2014 and is the brainchild of Raphaël Domjan, eco-adventurer. He will be flying to an altitude of over 80,000 feet on board the first two-seater solar aircraft, and…

IDTECHEX RESEARCH

Electric Boats and Ships 2017-2027

Electric Motors for Electric Vehicles 2017-2027

Solid-State and Polymer Batteries 2017-2027: Technology, Markets, Forecasts

Micro EVs, e-Bikes, e-Scooters, e-Motorbikes, Mobility for Disabled 2017-2027

IDTECHEX TV

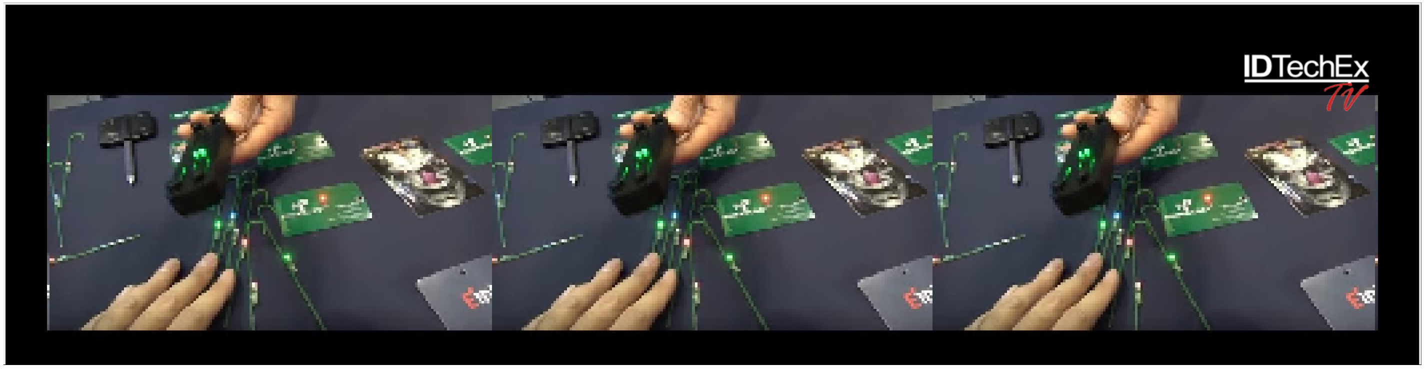

Embedded wireless power based on RF energy harvesting

Video Interview with Powercast at the IDTechEx Show! in Santa Clara, USA (November 16-17 2016)

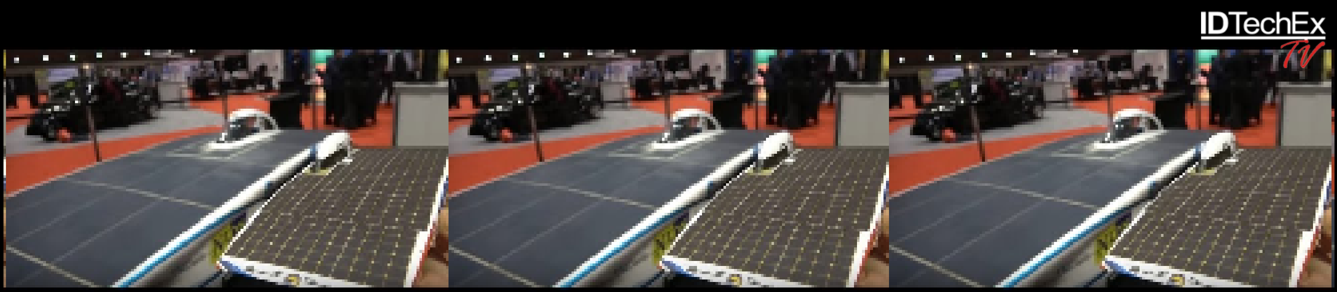

IDTechEx Show! Features Winning Solar Racing Car USA 2016

Video Interview with the Nuon Solar Team at the IDTechEx Show! in Santa Clara, USA (November 16-17 2016)

ENERGY HARVESTING JOURNAL TAGS

Small Devices | Small Devices | Energy Harvesting

- Site Map
- Submit Press Release
- Contact Us
- About Us
- Authors
- Privacy Policy
- Terms & Conditions

ISSN: 2399-4436 © Copyright 2017 - IDTechEx

Sign-up or Login

Joe Flynn's Parallel Path Magnetic Technology
Tim Harwood © 2003

There is a widespread opinion common to the mainstream academic community and also to various alternative scientific forums, that some kind of exotic new physics will be required to design and implement over-unity technologies. That is to say electrical motors, electrical generators, or other apparatus, which produce an excess of magnetic force or energy above the value actually inputted.

However, this has recently been experimentally demonstrated not to be the case, as I have validated myself in simple $20 experiments undertaken at home with parts bought from the local hardware store. It is the purpose of this article to educate readers that with care, thought, and a little work, it can be demonstrated that existing textbook physical law, freely allows for the extraction of excess electrical energy from magnetic systems.

The credit for this ground breaking research goes to Joe Flynn, who has been engaged in magnetic flux research for over 25 years now. His work is long standing, comprehensive, and in later years, well funded. It is reported $7m has been spent to date, with over $1m alone developing a revolutionary high performance magnetic motor. His equipment is validated, and apparently already in mass production for select customers. Since many lines of research have been formulated and explored by Joe Flynn, the following article presents only a brief summary of some of his best art apparatus, but is nonetheless sufficient to convey the basic ideas, and provide a framework within which one can undertake experiments.

Principals of Operation

The first illustration Figure 1 is taken from Joe Flynn's US patent 6,246,561, and explains a simple magnetic force multiplication experiment, which forms the basis for the Flynn magnetic art. If the windings on either side of the central magnet, which are normally connected in series, are properly pulsed, the field of the permanent magnet in the center, will be diverted to the opposite side of the core flux path provided. Or in alternative language, the side of the core that is pulsed, is demagnetized, relative to the field of the permanent magnet used in the apparatus. This is elementary textbook physics anyone can understand.

So what is surprising about this apparently simple apparatus, is that the armature on the side of the flux core, will contain 1.75 times more units of magnetic force, than could be manifested by the electrical input to the apparatus alone. Since the ability to arbitrarily move force from one point to another is the basis for motion or work, however simplistic, we therefore have a basis for a system that can be developed for practical technological purposes. Expressed in alternative language, we also have the capability to engineer a time varying magnetic field, without the need for moving parts, which will allow development of systems that output electrical energy. Both capabilities are highly desirable, and offer substantial opportunity for technical development.

Figure 1

U.S. Patent Jun. 12, 2001 Sheet 11 of 32 US 6,246,561 B1

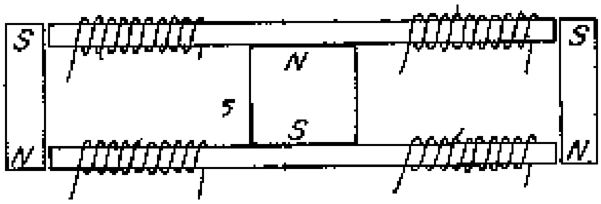

Fig. 16 B

1.75 times more force is delivered to the legs of the core
Than is provided by the electrical input to the control coils

Expanding upon this basic experiment, there is a second simple and logical improvement in layout illustrated in Figure 2, which should be obvious, but has been shown not to be the case. In this instance, the pulse is centrally located, and a dual flux field layout employed, which both demagnetizes the core relative to one magnet, and magnetizes it relative to the other. Since the two actions are complementary, the input required to manifest the flux switching effect is halved, therefore doubling 'efficiency.'

Figure 2

U.S. Patent Jun. 12, 2001 Sheet 30 of 32 US 6,246,561 B1

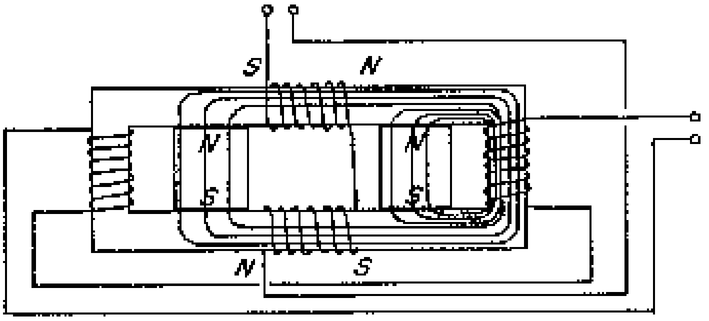

Fig. 45 Z

3.47 times more force is delivered to the legs of the core
Than is provided by the electrical input to the control coils

It should be noted that while the efficiency is doubled, the absolute output may not be significantly improved. This is because the major weakness of this effect and technology is flux saturation of the core, with values depending upon the specific properties of the B-H curve of the core material employed, limiting the absolute output of both layouts the same.

The previous statements are not required to be taken on trust, and simple experiments have been proposed by Joe Flynn, such that anyone can validate this effect for themselves. Figure 3 is a simple experiment taken from the Flynn website, that can be used to validate the principals put forward in this article.

Figure 3

Simple Magnetic Force Multiplication Experiments

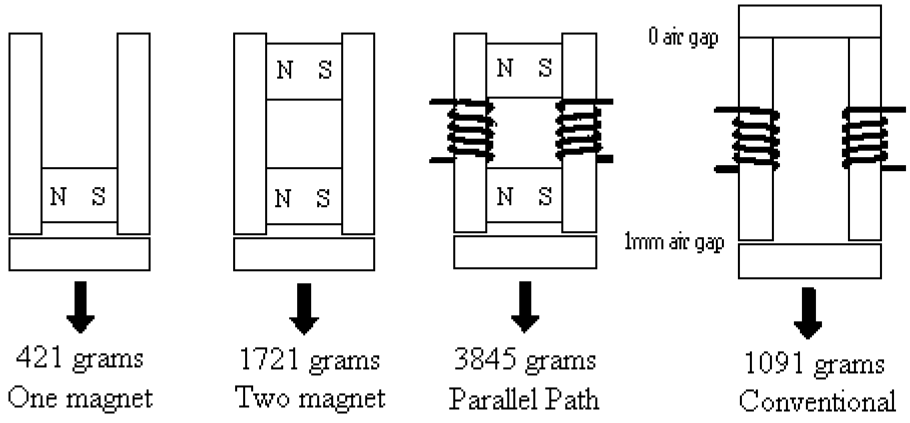
N S
N S
N S
N S
N S
0 air gap
1mm air gap
421 grams
One magnet
1721 grams
Two magnet
3845 grams
Parallel Path
1091 grams
Conventional

An even simpler non electrical flux experiment was proposed by GM in the Parallel Path Egroup. My apparatus is illustrated in Figure 4. It is no more than magnets and steel staple strips, bought from a local hardware store for a total of under $10. The Parallel Path effect can be replicated with identical apparatus, at only a slight increase in cost and complexity, with the addition of a simple 12v polarity reversible power supply, such as those commonly sold to power computer speakers, among other household applications.

Figure 4

In PP Egroup GM proposed a simple experiment to illustrate how small changes in layout can alter magnetic force in cores Experiment done with $20 of parts from local hardware store

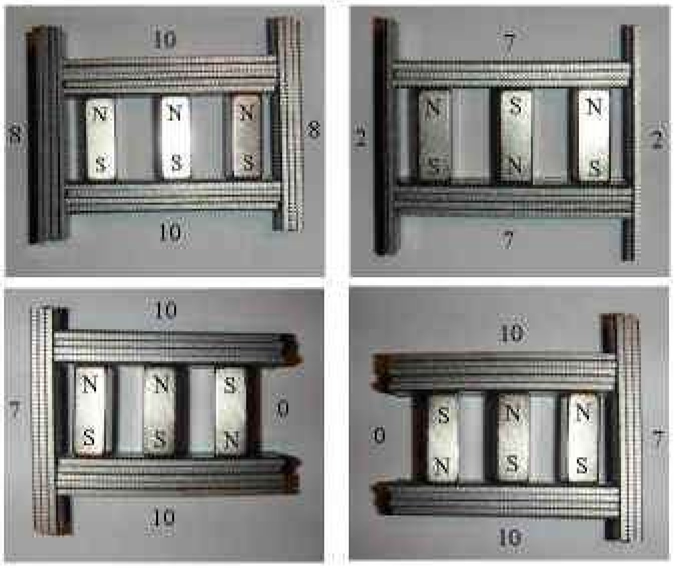
10
8
8
N N N
S S S
10
7
2
2
N S N
S N S
7
10
7
0
N N S
S S N
10
10
0
7
S N N
N S S
10

Conservation of Energy / Field Potential

One of the aspects of the Flynn technology people find most difficult to understand, is how you can have a device that delivers 3.47 times more units of magnetic force than is electrically inputted, yet not violate accepted principals of text book physics, as stated in the introduction. The key point here is magnetic fields do not gain energy - they are conservative. You can only ever obtain less energy from a magnetic field minus looses, than is in fact present. I feel this apparent puzzle can not be better explained, than by reference to Joe Flynn's own analysis of the experiment presented in Figure 3.

'*Since the Parallel Path System produced 3.47 times more force than the conventional system, with the same electrical input, it appears to violate conservation, this is only true when observed from a traditional view point. The system contains three flux producing sources (2 magnets and an electromagnet) which together are capable of producing a far greater force than is actually produced. All of the flux sources together can produce a force of 13.11 units, therefore in the physical sense a loss of 1 - (9.01 / 13.11) = 31% is realized.*'

So the system is 347 % efficient, in terms of delivered magnetic force compared to net electrical input, yet still conforms to the accepted physical principals of field conversation, by being only 69 % efficient, in terms of the fields present in the system. However surprising this result may appear, the analysis presented is in outline correct, with the difference between fields present in the system, and net electrical input, being the important concept presented.

Losses in the System

In order to optimise flux cores, an appreciation of the physics that underlies the transfer of flux within a core is required. The normal magnetization curve, or B-H curve, is a mathematical relationship between applied field intensity H, and resultant flux density manifested in the core B. It varies according to core material, and the curve will shift, if there is a starting magnetism within the core, such as that provided by the field of a permanent magnet. If the start magnetism is excessive, the core is saturated, and will not properly respond to the applied force H. A simple B-H curve is illustrated in Figure 5.

Figure 5

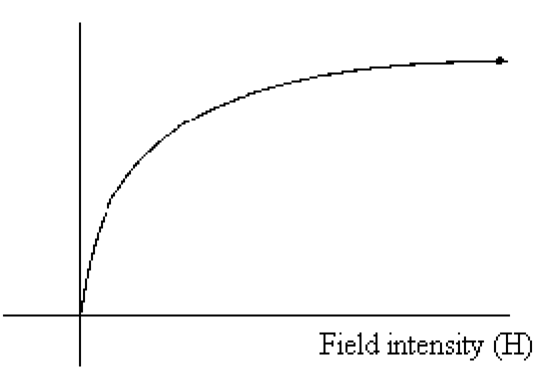

Hysteresis is a delay between applied magnetic force H, and resultant flux density B, that again varies according to material type. It also manifests as a delay between the termination of force H, and the manifestation of flux density B. So, the system will not turn on instantly, and will not turn off instantly, in simple terms. This is because the magnetic memory of the core, means a flux vector remains within it, even when the application of magnetic force H has been terminated. If we apply a reversed force H to the core, the basic B-H curve is now expanded as in Figure 6, with the memory effect also illustrated.

Figure 6

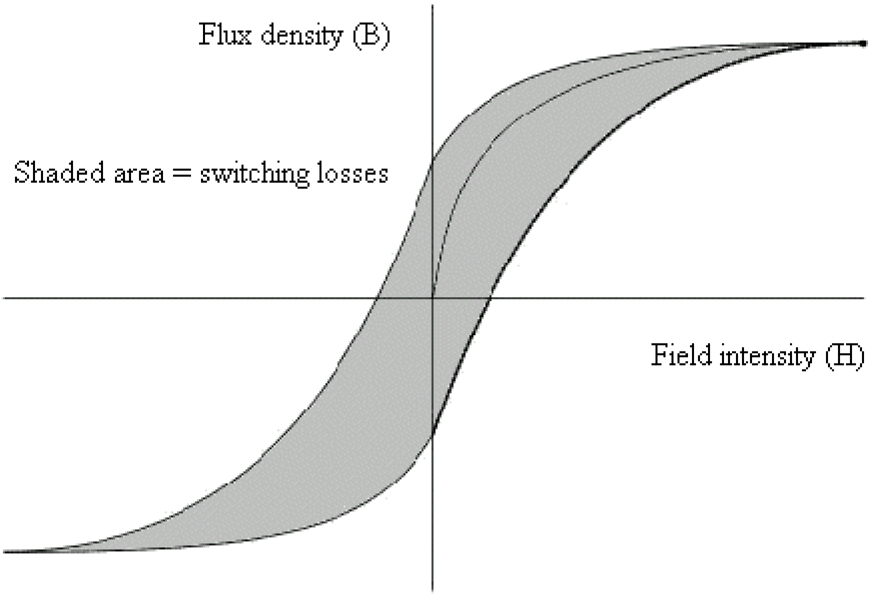

In order to return to the initial switched state, the remnance magnetism must now be overcome, hence input once in operation, will be greater than that required for the very first pulse. The area within the hysteresis curve gives a rough estimate for the amount of wasted energy, and along with other conventional sources of losses resultant in flux transfer within a core, is what reduces the efficiency of flux cores from maximum values of 2, or 4, down to values such as 1.75 or 3.47, typically.

Motor Apparatus

Although numerous practical applications abound for this effect, electric motor design remains the most outstanding opportunity. To this extent, again a few simple images, should be sufficient to explain how the basic flux switching apparatus, can be turned into a highly efficient electrical motor.

The first motor shown in Figure 7 is one I have proposed to validate the flux switching effect at a most basic level. It illustrates the point made in the Flynn patent, that the armature of the core can be removed, and replaced with a motor flux path. This first motor is not claimed to be highly efficient, but it helps one to understand how the transition from simple flux core to motor takes place.

Figure 7

Simple Parallel Path Technology Demonstrator Motor

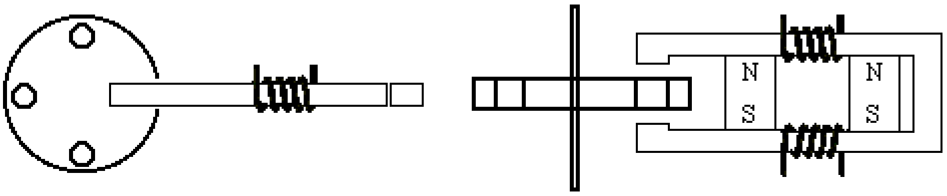

Designed to demonstrate core principals, not provide over-unity
4 magnetically permeable cores placed on rotor section
Flux is switched through core section on approach to armature section
When armature section is in register with rotor cores
Flux is switched back, and rotor core free wheels away from register

The next motor shown in Figure 8 is again taken directly from the Flynn patent, and illustrates the next intermediate step to motor design. The fields of the permanent magnets are alternatively switched from one side of the surrounding flux cores to the other, alternately interacting with N and S poles on the rotor, imparting motion to the central rotor shaft.

Figure 8

U.S. Patent

Jun. 12, 2001

Sheet 34 of 52

US 6,246,561 B1

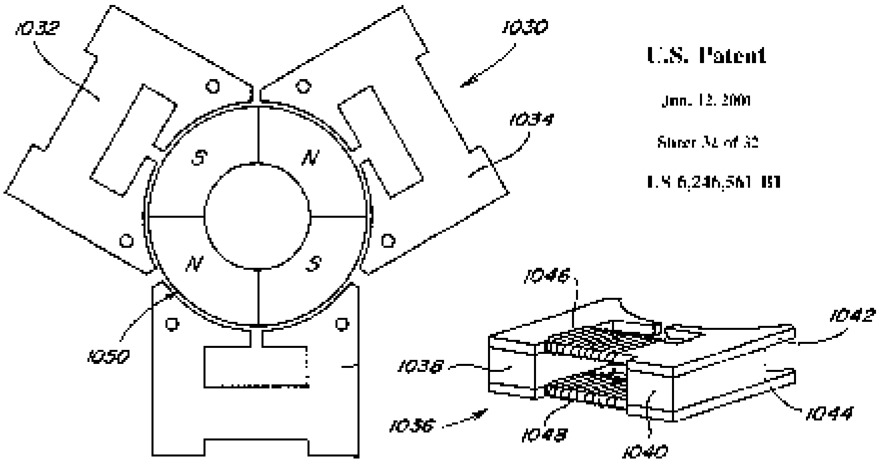

Fig. 49

Fig. 50

With proper financial support, and the facilities to have metglas cores custom moulded, Joe Flynn was able to develop his final best art, shown in Figure 9. No detailed performance numbers have been released for this motor, whose precise characteristics remain proprietary to Joe Flynn at this time. But the optimisation is so expert, it is stated to posses certain exotic properties, such as cool ambient operation, even during prolonged periods of continuous load. This 'cold running' is said to be of great interest to the American military, as it offers excellent stealth performance characteristics.

Figure 9
Optimal Flux Core Motor Construction

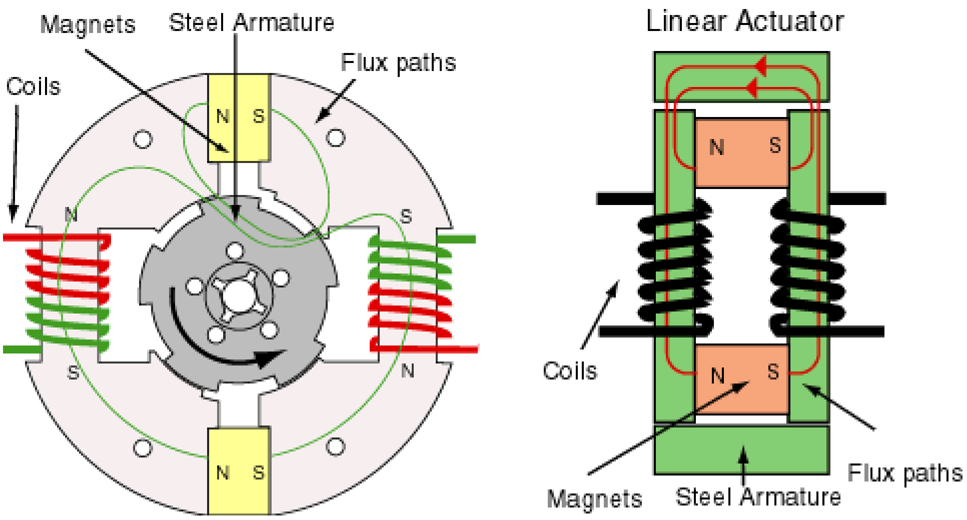

Electrical Apparatus

Many readers will no doubt have noticed the similarity of the first illustration presented in this document, to the so called 'Tom Bearden MEG.' This is fair comment, and Joe Flynn has always highlighted this issue. However, it has been commonly stated Joe Flynn simply developed mechanical apparatus, and the MEG with its electrical functionality, is distinct art, more advanced than the mechanical Flynn apparatus. However, this is shown not to be the case by a careful examination of the Flynn patent, in which the following is stated in the 'Power Conversion' section:

'*The construction shown in FIG. 45A utilizes four control coils and a single permanent magnet and the construction shown in FIG. 45X uses two control coils and two permanent magnets. The flux that would normally be supplied by a primary winding is supplied by the static flux of the permanent magnet or magnets and the control coils convert this static flux into a time varying flux in a novel way. Both arrangements use two secondary coils, the secondary coils are placed in the region of the continuous flux path that would be occupied by an armature or rotor in the linear or rotary arrangements. The regions of the flux paths that perform work are the same in all cases.*'

'*By alternating the polarity of the control coils during one cycle, one working region experiences an increasing flux and the opposite region experiences a decreasing flux and during the next cycle the opposite occurs. This results in the induction of a voltage in the secondary coils that is decided by the magnitude of the change in flux in the working region and the time in which this change occurs. The novelty of this discovery is that the primary flux inducing the voltage in the secondary coils is supplied by the permanent magnet or magnets and is far greater than the flux supplied by the control coils.*'

Figure 10

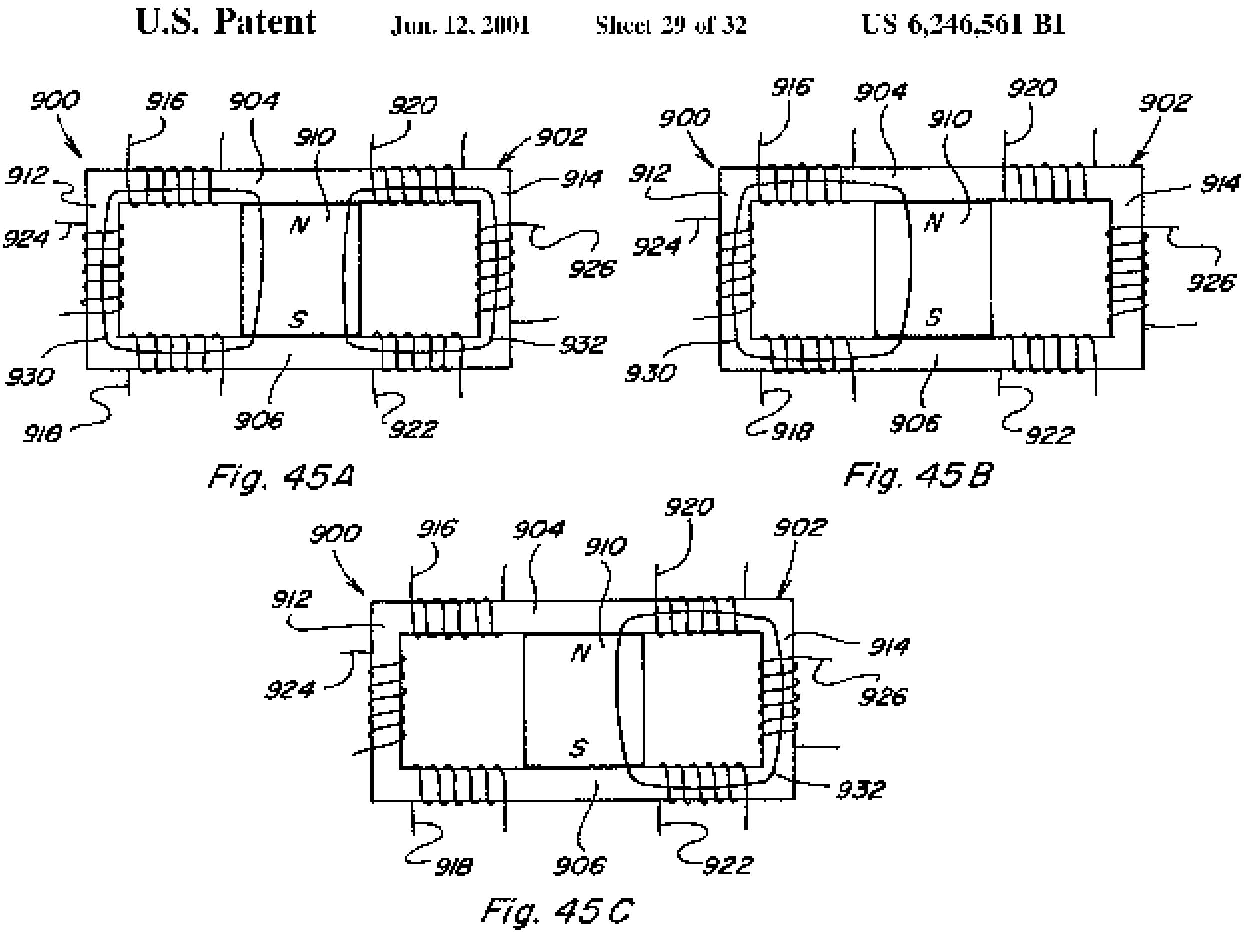

Figure 10 taken directly from the Flynn patent makes the point even clearer. Joe Flynn has also stated that his intellectual property rights will be robustly defended, by legal action if necessary, and he regards himself and his company as being in possession of exclusive rights to the so called 'MEG' unit.

As regards replication of electrical output orientated flux core devices, certain important details need to be stated. For example grade 8 ceramic magnets should be used, so as to avoid flux saturation of the core. A basic error, many early experimenters wasted time on. The requirement for strong magnets to obtain over-unity results, is as much of a myth, as the idea new physics is required.

But perhaps the greatest trade secret of the electrical devices, one which several lengthy non disclosure agreements are required to be signed before it can be disclosed, is that the input and output circuits must be closed in series. The disclosure of this technique, amounts to putting the basic MEG methodology fully into the public domain.

The reason for this circuitry requirement is obvious enough, with only a little analysis. If the output circuit is closed when the input circuit is activated, then the input energy simply leaks into the output circuit, as in an ordinary transformer. So no flux switching effect is manifested, and the field of the permanent magnet is static in time. Thus you have an ordinary transformer, with reduced efficiency, because of the core flux saturation effect provided by the permanent magnet.

This is one of the most important points to make about the Flynn apparatus. If you approach it as if it is a normal piece of scientific equipment, then proper optimisation is not greatly problematic. For example more turns on the output coils, simply means more voltage and less current, exactly as standard textbook equations predict. Increased input voltage enables faster switching speeds, a consequent greater rate of change of magnetic flux, resulting in higher absolute output, but only up to the flux saturation limit of the core material. Generally, problems only occur, if you imagine the effect is based upon exotic scalar type or vacuum energy physics, when in fact it is ordinary flux manipulation within a core.

Present Status of the Flynn Project

Initially Joe Flynn was remarkably open about his work and research. However, since performing a working demonstration of various advanced hardware samples for the American Department of Defence, little has been heard.

I want to clearly emphasize I do not speak for Flynn Research, nor am I in any way connected with Flynn Research, and by consequence, have no inside information whatsoever as to the present status of the project. But we all sincerely hope, that the project has not been swallowed whole by the American deep black military industrial research complex.

However, even if this is the case, it does not mean the technology is lost. Extensive and generous details have been provided by Joe Flynn of his research, both in his patent deposition, website, and other comments, such as to enable persons of scientific training and skill, to replicate the effects stated.

While replication of the electrical effect remains extremely demanding, the mechanical apparatus is very easy both to understand and replicate. There is no reason why professional scientists and home experimenters alike, can not build Flynn type flux core motors, and explore over-unity flux manipulation for themselves. The future has already arrived, and it is simpler and cheaper than anyone dared imagine to be possible.

References

http://www.flynnresearch.net/, Flynn website
http://www.rexresearch.com/meg/meg.htm, MEG US patent 6,362,718

Jools' Razor Blade Crystal Radio

Foxhole Radio

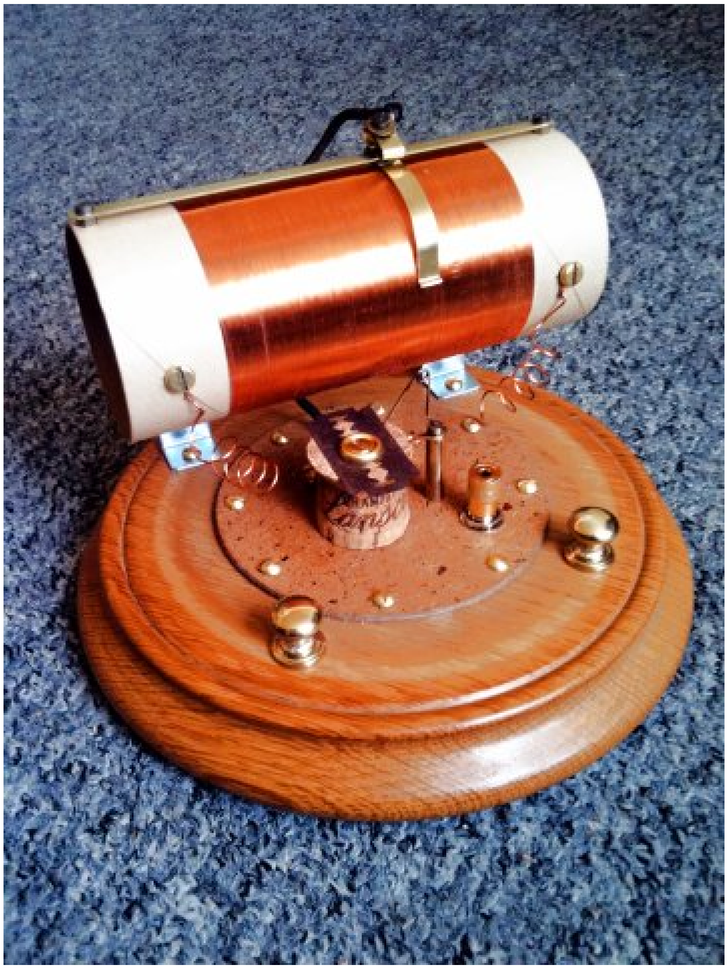

A- (javascript:changeFontSize(-1)) A+ (javascript:changeFontSize(1))

I am strictly a 'newby' in that my speciality is foxhole radios. If you might be interested in those, I have a whole load of photos of various contraptions I have built that work with razor blades and such like. I wanted to get right back to the beginning of the science and have been successful in building sets that work with a collection of household items.

I believe that the magic begins with the core science, and when you hear music and voices coming through on a set made from the contents of a kitchen drawer, it makes you realise how amazing the whole genre is.

In time, as I get the mechanical ability to make more 'beautiful' radios I will take more photos.

I am not a ham radio member, but I have been a 'commercial radio' broadcaster / presenter. So, I'm a bit of an oddball,

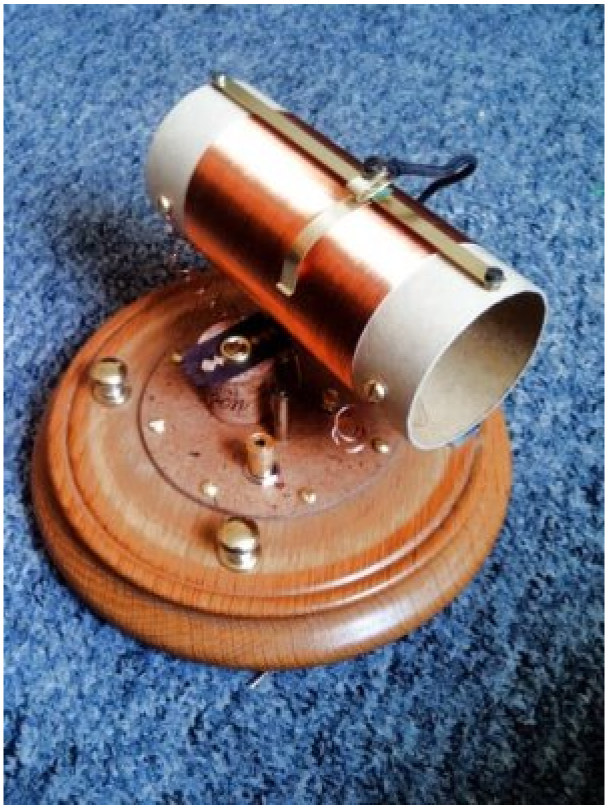

You can see from the schematic that it's based on a 1920's design and is a very basic design. However, with the variable coil wired into the aerial, there is a fair degree of tuning available however, the local BBC station is pretty dominant throughout as I only live about two miles away from their main transmitter. I took a 'short cut' route by linking the aerial wire directly into the roof TV aerial socket (not a particularly safe idea, don't try this at home) and got a surprisingly good, interference-free sound.

The audio quality is slightly sibilant but on the whole, very listenable. I uploaded a couple of sound files to demonstrate just how clear the sound is - Sound File 1 (http://audioboo.fm/boos/484097-listening-on-a-razor-radio) Sound File 2 (http://audioboo.fm/boos/484118-interacting-with-local-radio-using-the-power-of-a-basic-home-made-radio) I emailed the show and the DJ gave me a name check, just to prove it was working.

In terms of construction, I used things I could find in my shed (a cheeseboard, nuts, bolts, brackets, door handles) and the whole thing took about two days to complete. The only tools required were: a drill, a soldering iron, a screwdriver set, a pair of pointed pliers and a pair of scissors.

The luckiest find was a pack of 'Gillette Blue' razor blades which I picked up at a car boot sale for fifty pence, but I found it very hard to locate the 'sweet spot' (which was usually on or around the lettering) so I eventually used a rusty blade from my dad's old shaver. This worked best when one contact (the mounting post) was linked directly to the flat rusty surface and the 'cat's whisker' (a brass safety pin) touched the blade edge. This arrangement worked perfectly at any point along the blade so long as a tight contact was made. Other references describe using the point of a hard pencil 'lead' mounted on the safety pin as the contact point but I found this to be not only fiddly but quite ineffective (and the lead kept breaking).

I mounted it with a standard headphone socket so that my daughter could play it through her computer (and so I could make recordings of the various tests and programmes I heard). That was a decision of convenience.

I don't have a website dedicated to radios, but I am a writer and in the middle of a novel about a radio star from the 40s and 50s. The best link (if at all) to use would be to the hub of that creative pursuit - http://www.kooshrecords.co.uk/reg/ (http://www.kooshrecords.co.uk/reg/) Also - happy to show this email address as a contact.

As I get chance to take other photos of these 'foxhole' radios, I'll send them over for your perusal. There's some 'wacky' stuff! I hope you like the 'enthusiasm' and boy scout innocence inherent in these designs (which was my intention: to make a device capable of hearing radio without any deep scientific knowledge, skills, equipment or resources. The kind of radios that boys in the 20s would have made.)

Best regards
Jools
jools@bramwellslater.org (mailto:jools@bramwellslater.org)

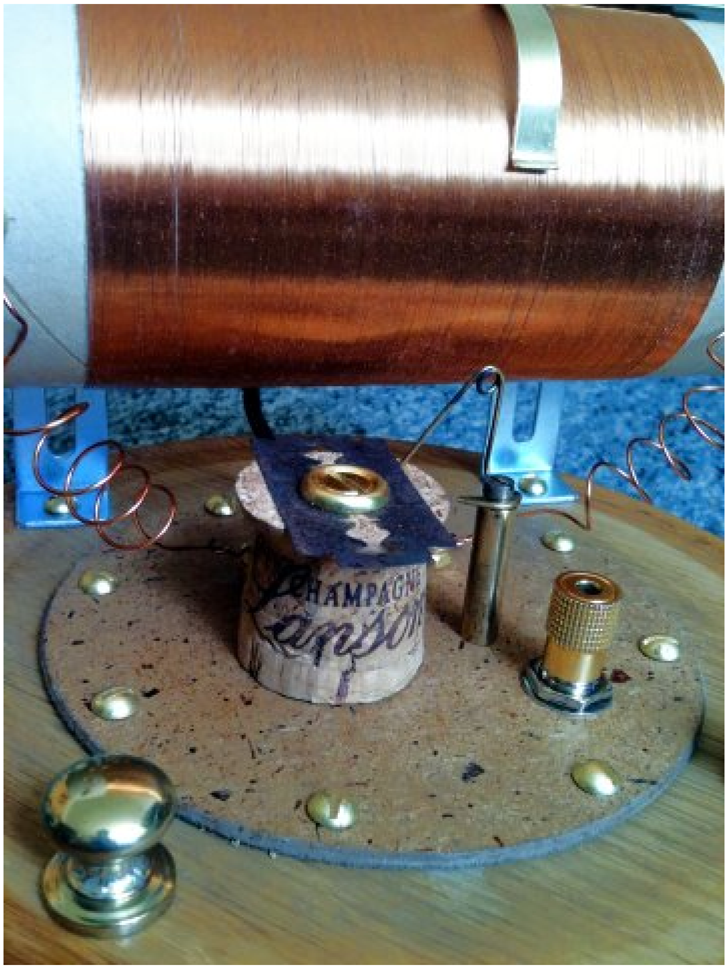

Close Up View of the Detector

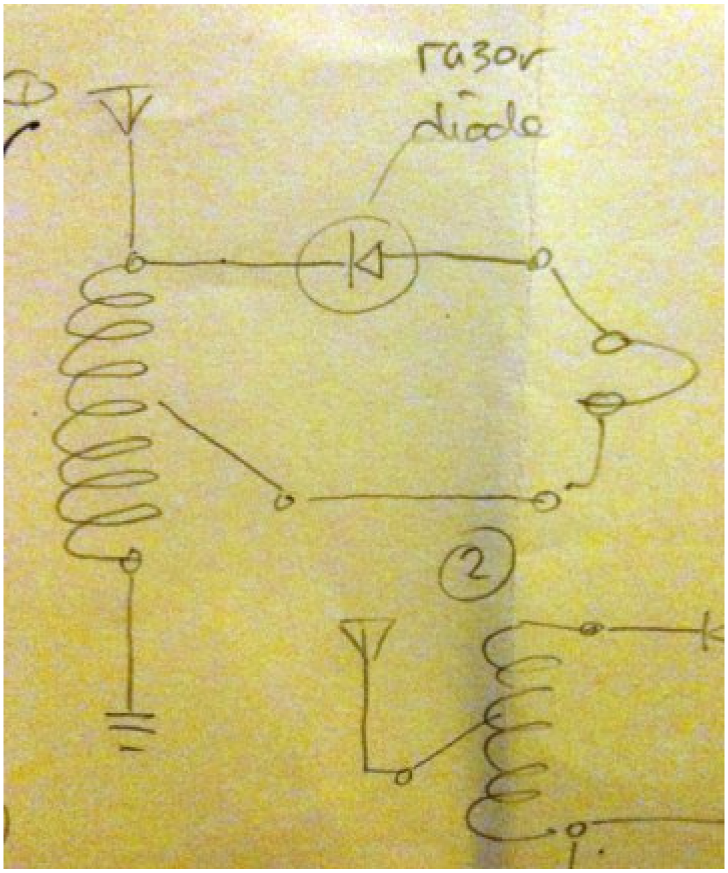

Schematic

---

rexresearch.com

Herman ANDERSON

Water Fuel System

Lifetime Warranty
Hydrogen kits, starts at $199 Same day shipping, great results!
MileageShop.com

Field Proven Fuel Economy
Discover how to save over 1000$ per truck per year !
www.total-fe.com

Auto Spark Plugs
More Horsepower & Fuel Efficiency Improved Performance. Find Out More
www.DensoAftermarket.com

Ads by Google

---

http://waterpoweredcar.com/herman.html

GOT
H2?
Photo by Dave Sieg © 2004

Herman P. ANDERSON
(1918-2004)

Herman P. Anderson Technologies, LLC
PO Box 2014
Brentwood, TN 37024-2014

Herman and his water-powered Chevy Cavalier.

by James Allen

He could drive it, but he was not allowed to sell any, or manufacture his discovery in the State of Tennessee. Why not?

Herman's special patented sparkplugs, similar to Stan Meyer's plugs.

Herman has consulted with Nasa and the US Airforce on some of our country's most important Top Secret Projects including the first US satellite in space, the SR-71 Blackbird, the Stealth Fighter/Bomber and Starwars. Herman worked closely with Dr. Wernher von Braun testing hydrogen powered rocket engines, and he also worked with engineers at the now famous Skunk Works, the Jet Propulsion Laboratory, JPL, and Cal Tech. While in the Air Force during WWII, Herman served as a fighter pilot, a flight instructor and bombardier trainer. He was certified in and flew 23 different airplanes during this period.

Herman believed in a Hydrogen Future and FREE fuel!

Herman's 1971 Ford LTD converted to run on hydrogen or gasoline, on display in the Water Fuel Museum in Lexington, KY. In a video interview with the inventor, he explains that ambient air is mixed with hydrogen (not oxygen) and a micron-sized fog of water mist (to mimic gasoline's burn), introduced in the same way propane gas is on a propane conversion. In fact, the LP regulator and LP aircleaner assembly are the same as on a propane vehicle. The butterfly in the carb, as in an LP system, now serves as the intake air control. The rest of the function of the gasoline carb is not used. He claimed greater power than gasoline, and 38 miles per gallon of water. As far as quantity of hydrogen produced, he produced enough gas from his electrolytic chamber to need a cut-off switch connected to a pressure gauge to stop gas production when it was no longer needed while driving. In other words, instead of producing more or less gas on demand, as some systems do, it appears his system always produced the same amount, but was turned on and off as needed. He emphasized that deuterium (heavy water) was essential to his approach, doubling the density of the hydrogen, making it twice as powerful. He also used high voltage, 70,000 volts to be exact, from two custom-made coils, to effect rapid separation of the component gases, a process which he calls 'radiolysis.' The 70,000 volts constitute what he calls a 'soft' x-ray, not radioactive but in need of shielding, in-between a microwave oven and a 'hard' x-ray. He was authorized by the state of Tennessee to drive his car as they recognized that he knew what he was doing with such technology. I view his prototype LTD as good proof of a watercar for those who doubt that the technology is even possible, but not necessarily the best approach for you and me to pursue. He accomplished his invention along the lines of his expertise, the field of ionics; others have approached it differently with less controversial technology and have been equally successful.

My thoughts about Herman

Herman's 1978 Fort LTD came to be known as The Green Machine. Using a modified propane carburetor, Herman converted the LTD to run on hydrogen in 1980. Due to the low mass energy density of hydrogen, Herman was only able to get a little more than half power from the engine with this traditional induction system.

Because Herman only used a single cell, I believe you need to do the 3 cells hooked up in series like Archie Blue, Dingel's and Peter's design. Use - 3 - 4"dia. SS tubes , 8" long (cells) welded together. Put your series of SS plates inside each cell. gas lines coming out the top through a check value to the bottom of the next cell. Use rubber caps ends with hose clamps so if it ever blows it just blows the caps off and the water is drained and your mild explosion is over with.

You need more molecules of hydrogen clustered together as they pass thru 3 cells then into the IMPCO propane fuel injection carburetor. In gasoline SAE words a higher octane of hydrogen (ortho hydrogen) before it enters the fuel injection chamber. first H2 and O, then H4 and O2 then H8 and O4?? Who is the mathematician here? I'm not him. Peter can tell you.

There is para-hydrogen and ortho-hydrogen molecules. Ortho-hydrogen molecules will cluster and stick together, as they pass through 3 cells or more they cluster together more and more, their proton travels in the same direction, when finally sent to the combustion chamber they have 2-1/2 times the explosive power than gasoline!

Para-hydrogen are hydrogen molecules that will run together but have repelling protons to each other, therefore the clustering does not happen, they are less combustive than ortho-hydrogen. If you produce ortho hydrogen, then this also creates deuterium in which is what you want. deuterium build up on the cylinder walls, values helps oxidation from happening.

Now Stan used one tall wide cell with 3 tubes inside one large tube. 12" diva. and about 24" tall. Go look at his files.

To increase the power of the hydrogen, he irradiated it making deuterium. The extra proton was more than enough to allow the engine to achieve full power. It is good to get Deuterium built up in your piston walls, from the hydrogen so the water exhaust will not rust out the motor. They say AMSCO synthetic oil helps deal with the water in the oil better, helps it vaporize the water molecules.

Unfortunately, state officials decided the radioactive source under the hood would not be safe for widespread use. They did however allow Herman to continue driving the car and he put nearly 30, 000 miles on the car running on hydrogen. Realizing the technology could be improved, Herman went back to the drawing board.

It is kinda of ironic in that is exactly what a Ballard fuel cell is also doing, when in use in the new Hybrid Cars. The Ballard Fuel Cell is radioactively produce hydrogen to burn mixed with the air and the gasoline. The new hybrid cars are just as safe, except they use 1/2 gas and 1/2 electric. The Hybrids also contain radioactive batteries. The only difference was Herman's Wonderful car ran for FREE and with the new hybrid cars, you still have to PURCHASE the GASOLINE.

Yes the Mean Green Dancing Herman Machine was in the heart of every reader passing by with a BRAIN. Addicted to oil you drive to work, paying the largest corp. the largest profit in American history ever! Last quarter of 2004's profit from Exxon/Mobil exceeded the record for the largest quarter profit of 8 billion dollars. We are all guilty, unless you walk to work ride a bike, or enjoy an electric train.

The Green Machine also employed onboard electrolysis, meaning that to fill the gas tank, Herman would pour in distilled water.

and a little se cert Electrolyte KOH (1-5% solution to water, is all)

Yes there is a car that "runs on water." The car is now parked in the research facilities at MTSU.

Just like the TESLA statue at Niagara Falls, NY, it is way off to the side. And Edison's bust is real big in the middle with lights shinning on him. and he didn't even invent AC electricity, he was angry with Tesla's invention of AC, because it made his DC look non efficient. Today in all our households, we use AC electricity. It could have also been FREE wireless electricity. J.P. Morgan quit funding Tesla's FREE electricity Tower, because he could not find where to put the meter.

Next 10,000 people with water powered cars can join the STAN MEYERS 10.000 DUNE BUGGY PLANET MARATHON CLUB.

> All truth passes through three stages: First, it is ridiculed; Second, it is violently opposed; and Third, it is accepted as self-evident. -- Arthur Schopenhauer (1788-1860)

Fuel System for Internal Combustion System and Adapter for Use in Same

USP # 5,852,993

1998-12-29

Applicant: HERMAN P ANDERSON TECHNOLOGIES (US)
Classification: - international: F02B43/00; F02M21/02; F02B43/00; F02M21/02; (IPC1-7): F02M57/06; - European: F02B43/00; F02M21/02A
Also published as: US6119651 (A1)

Abstract -- The internal combustion engine fuel system described includes a structure for mixing the alternative fuel, preferably hydrogen, with oxygen in ambient air to stratify the fuel. The system includes an adapter, and the adapter includes a housing mounted between spark plug and cylinder of the internal combustion engine. A plug is placed within the housing. The plug has ridges or grooves on its outer surface that act as mixing structures. Thus, when hydrogen is introduced into the adapter housing it is mixed with ambient oxygen within the chamber as it flows over the plug. The mixing structures in the housing create a vortexing action as the hydrogen flows over the plug and towards the cylinder of the engine. An electrode protrudes from the plug towards the cylinder. The electrode is preferably platinum and generates the necessary spark to create combustion of the hydrogen/air mixture adjacent to the cylinder to thereby power the cylinder in the engine. A platinum electrode is preferably used because it enhances a catalytic conversion of combustion by-products to more environmentally compatible products.

Inventors: Anderson; Herman P. (Brentwood, TN)
Assignee: Herman P. Anderson Technologies, LLC (Brentwood, TN)

Current U.S. Class: 123/297 ; 123/527; 123/DIG.12
Current International Class: F02B 43/00 (20060101); F02M 21/02 (20060101); F02M 057/06 ()
U.S. Patent Documents

3926169	December 1975	Leshner et al.
3980061	September 1976	Mcalister
4319552	March 1982	Sauer et al.

4343272	August 1982	Buck
4383198	May 1983	Hosking
4448160	May 1984	Vosper
4546740	October 1985	Clements et al.
5531199	July 1996	Bryant et al.

Description

BACKGROUND OF THE INVENTION

The present invention relates generally to a fuel system for internal combustion engines that enables the invention to utilize alternative fuels; and, more particularly to an improved fuel system for an internal combustion engine and an adapter for use in same.

The problems of air pollution caused by automobiles, lawn mowers and other vehicles that have internal combustion engines powered by gasoline is well documented. This problem is particularly described in an article in the San Diego Union Tribune, Wednesday, Sep. 25, 1996 and in Popular Science, August, 1996. Indeed, under the Clean Air Act, in the next few years cars that do not run on standard fossil fuels like gasoline will be required in several states. Currently, there are no vehicles available to meet this need.

Thus, it will be appreciated by those skilled in the art that there has been a push to seek other fuels for use in internal combustion engines. These fuels include methane, propane and gasohol.

Hydrogen has also been characterized as a promising fuel source. See Department of Energy publication entitled "Hydrogen Fuel", printed in 1978, U.S. Government Printing Office, U.S. Department of Energy, Office of Public Affairs, Washington, D.C. 20585. Particularly, the graph presented in FIG. 6, which is copied from this publication, shows the potential for hydrogen in BTUs/Lb. However, the mass energy density for hydrogen is low; and thus, hydrogen needs to be compressed before its energy potential can be realized. No prior art fuel system has been able to capture the energy of hydrogen.

Several fuel systems for use with hydrogen have been proposed. Examples are described in U.S. Pat. Nos. 4,167,919; 4,253,428; 4,016,836; 4,178,882; 5,222,993; 5,085,176; and 5,085,176. The problem with these prior art fuel systems is they cannot can take advantage of the alternative fuels to generate sufficient power from the fuels to drive an internal combustion engine for use in a lawn mower, automobile or other vehicle. As noted above, this is a particular problem when hydrogen is used as a fuel. Additionally, the prior art systems do not address the problem of back fire, i.e. undesired explosion of the fuel in the cylinder during the exhaust stroke of the internal combustion engine.

What is needed, then, is a fuel system for an internal combustion engine that provides a way to generate sufficient power from the alternative fuels and that reduces backfire problems. Such a system is lacking in the prior art.

SUMMARY OF THE INVENTION

It is an object of this invention to provide an improved fuel system for use with alternative fuels.

It is a further object of this invention to provide an improved fuel system for use particularly with hydrogen, and particularly to enhance the mass energy density of hydrogen.

It is still a further object of this invention to provide an adapter for use in a fuel system for an internal combustion engine that enhances the power obtained from the alternative fuels.

It is still a further object of this invention to provide an improved fuel system for an internal combustion engine that does not present the environmental hazards of a standard gasoline engine by providing reduced or zero emission of environmentally harmful gases.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that can use an alternative fuel and can use gasoline to facilitate the transition from gasoline to alternative fuels.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that reduces backfire problems.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that reduces backfire problems by cooling the electrode in the spark plug with the hydrogen gas fuel.

It is still a further object of this invention to provide an adapter for use in a fuel system for an internal combustion engine that enhances the power obtained from the alternative fuels.

It is yet a further object of this invention to provide a direct injection fuel system wherein fuel is injected directly into the cylinder.

It is yet a further object of this invention to provide a way to convert the standard induction fuel system of a lawnmower to a direct injection fuel system wherein fuel is injected directly into the cylinder.

It is yet a further object of this invention to provide a way to convert the indirect injection fuel system of an automobile to a direct injection fuel system wherein fuel is injected directly into the cylinder.

The internal combustion engine fuel system of this invention includes a structure for mixing the alternative fuel, preferably hydrogen, with oxygen in ambient air to stratify the fuel. The term "alternative fuels" is known in the art and is meant to refer to fuels other than conventional gasoline.

The invention comprises an adapter, the adapter including a housing mounted between spark plug and cylinder of the internal combustion engine. A plug is placed within the housing. The plug has ridges or grooves on its outer surface that act as mixing structures. Thus, when hydrogen is introduced into the adapter housing it is mixed with ambient oxygen within the chamber as it flows over the plug. The mixing structures on the plug create a vortexing action as the hydrogen flows over the plug and towards the cylinder of the engine.

An electrode protrudes from the plug towards the cylinder. The electrode is preferably platinum and generates the necessary spark to create combustion of the hydrogen/air mixture adjacent to the cylinder to thereby power the cylinder in the engine. A platinum electrode is preferably used because it enhances a catalytic conversion of combustion by-products to more environmentally compatible products.

Thus, the combination of the compression action generated by the conventional piston and push rod within the cylinder of the engine in conjunction with the hydrogen/air mixing structure that vortexes the hydrogen/air create compressed hydrogen that is mixed with a combustion facilitator, i.e. oxygen in the ambient air to enable the system of this invention to capture the abundant energy found in hydrogen fuel.

The enhanced mixing capabilities of this invention is also adaptable for use with other alternative fuels such as natural gas, alcohol, propane and the like.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a plan view of the fuel system of this invention with a lawn mower shown in phantom.

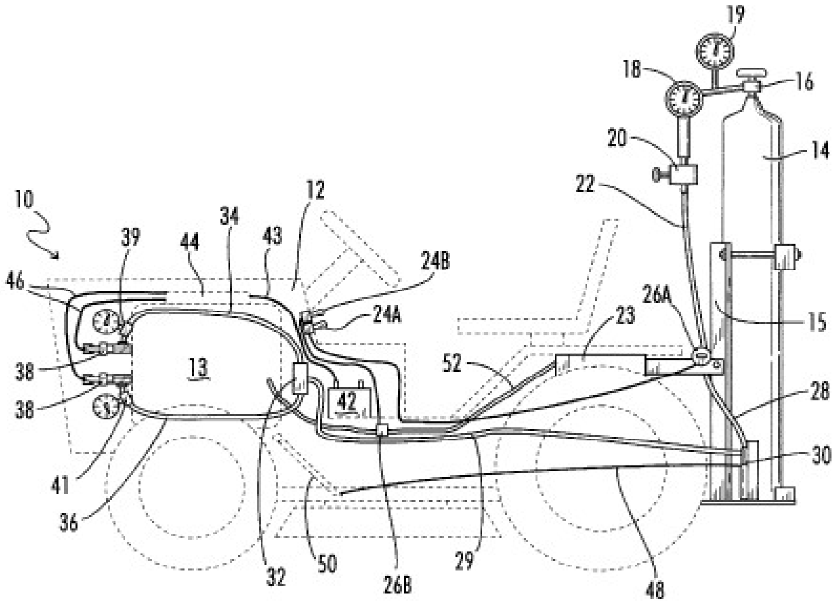
19
18
16
20
14
22
10
12
34
39
44
43
24B
46
24A
26A
23
15
52
13
38
38
42
28
30
41
36
32
50
26B
29
48

FIG. 1

FIG. 2 is a perspective view of the adapter of the fuel system of this invention.

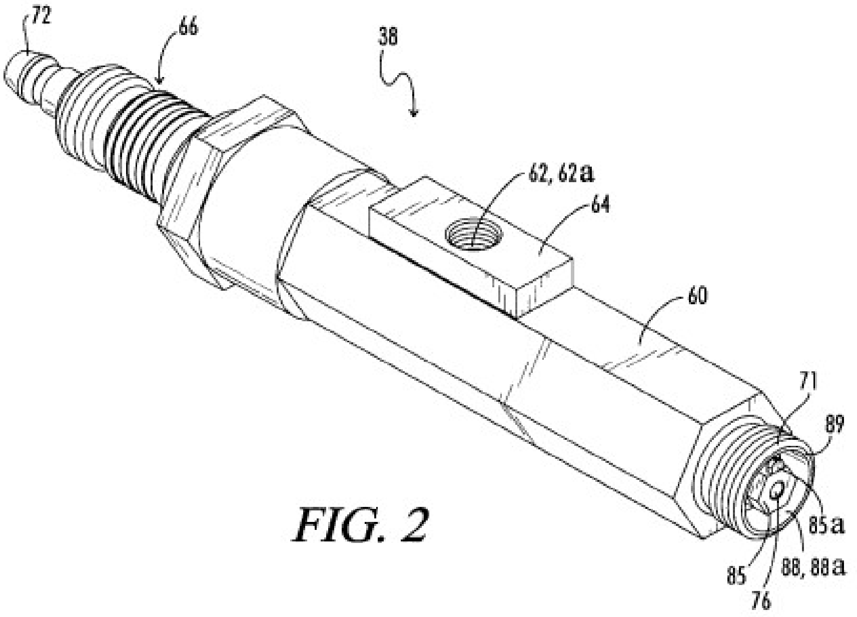
72
66
38
62, 62a
64
60
71
89
85a
85
88, 88a
76

FIG. 2

FIG. 3 is an exploded perspective view of the adapter of the fuel system of this invention.

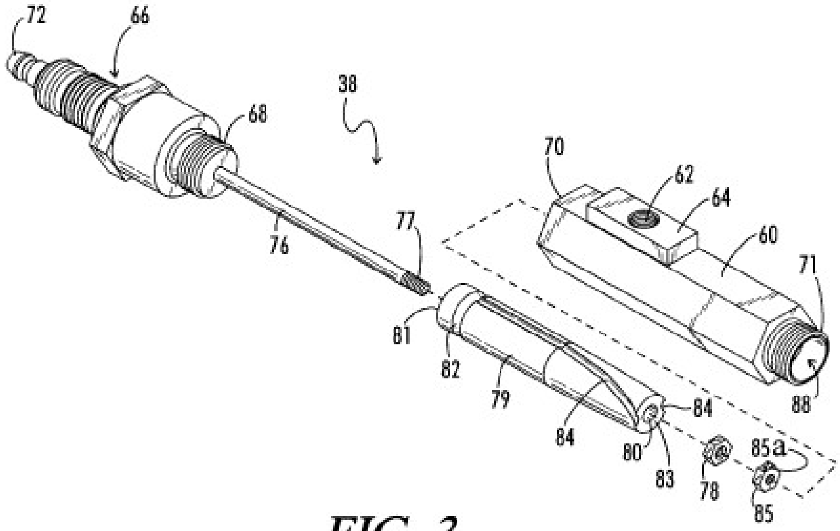
72
66
38
68
70
62
64
60
71
88
77
76
81
82
79
84
84
80
83
78
85a
85

FIG. 3

FIG. 4 is a cross-sectional view of the adapter of the fuel system of this invention.

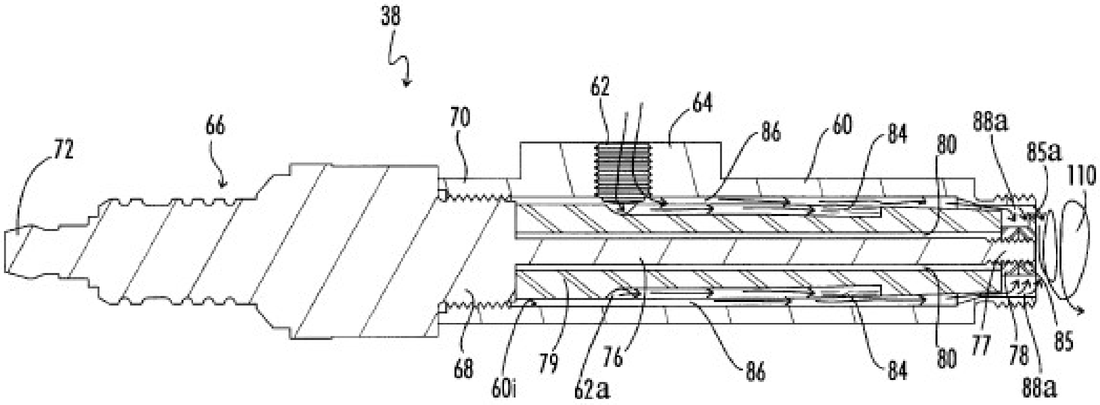
38
72
66
70
62
64
86
60
84
80
88a
85a
110
68
60i
79
62a
76
86
84
80
77
78
85
88a

FIG. 4

FIGS. 5A-5D are a schematic of the 4 cycle combustion in an internal combustion engine including the adapter of this invention.

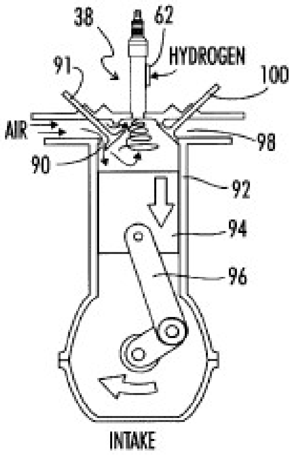
38
62
HYDROGEN
91
100
AIR
98
90
92
94
96
INTAKE

FIG. 5A

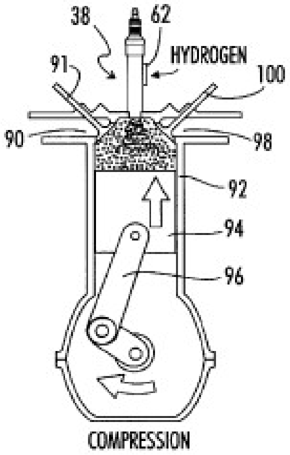
38
62
HYDROGEN
91
100
90
98
92
94
96
COMPRESSION

FIG. 5B

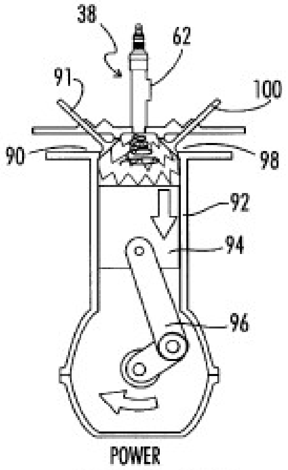
38
62
91
100
90
98
92
94
96
POWER

FIG. 5C

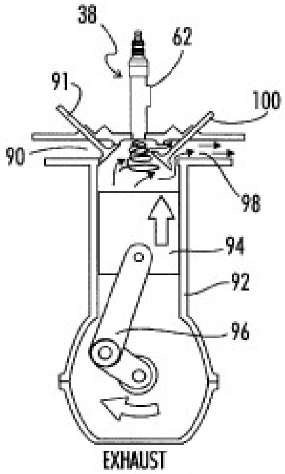
38
62
91
100
90
98
94
92
96
EXHAUST

FIG. 5D

FIG. 6 is a graph which shows the energy potential for hydrogen in BTUs/Lb.

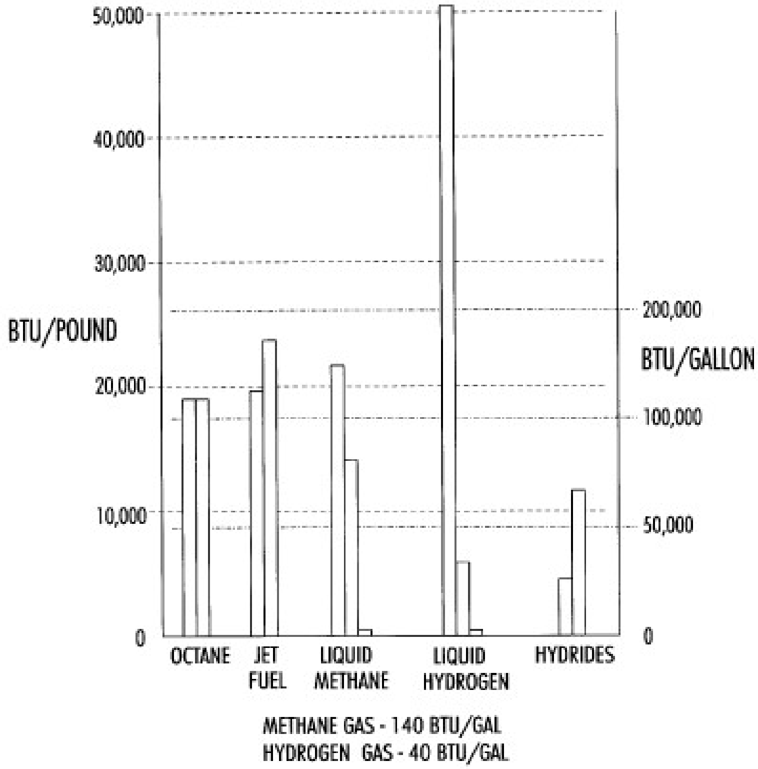

FIG. 6

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring now to the drawings, wherein like reference numerals refer to like parts throughout, the fuel system is referred to generally at 10. The fuel system of this invention is contemplated for use with an internal combustion power system, such as an internal combustion engine or such as a home heating system. Fuel system 10 is portrayed in this case as connected to a standard lawn mower. The lawn mower is shown in phantom at 12 in FIG. 1. The lawn mower is a conventional lawn mower as can be purchased from a company such as Murray of Ohio. The internal combustion engine in the lawn mower is a standard 4-cycle engine with two cylinders and develops 18 horsepower. A suitable example can be purchased from Briggs & Stratton, Milwaukee, Wis., 53201.

Referring particularly to FIG. 1, system 10 includes fuel tank 14 that is mounted to lawn mower 12 via mounting assembly 15. In the preferred embodiment, fuel tank 14 includes hydrogen gas. Accordingly, for the purposes of the foregoing detailed description, hydrogen gas will be described.

A line 22 begins at fuel tank 14 and proceeds to back-flow valve 16. Back-flow valve 16 is a standard valve and is available from AirCo Gases Company, 575 Mountain Avenue, Murray Hill, N.J. 07974, part no. WCS CV-4M. PSI tank gauge 19 is inserted in line 22 between back-flow valve 16 and flow pressure gauge 18. PSI tank gauge 19 is available from AirCo Gases Company also, part no. WCS CV-5M, and serves as an indicator of fuel remaining in the tank 14. Flow pressure gauge 18 is used to monitor the flow of the hydrogen in pressure units. The flow is started using valve 16. The optimal flow pressure is 25 pounds of flow pressure from the tank 14 to the cylinder of the engine.

The flow of hydrogen is regulated via valve 20 that is part of gauge 18. Thus, when valve 20 is opened, hydrogen proceeds through line section 22 to hydrogen flow solenoid 26A. The user can access the hydrogen fuel by flipping hydrogen flow switch 24A into the "on" mode so that hydrogen gas in line 22 can proceed through hydrogen flow solenoid 26A and into line section 28.

Throttle 30 regulates the flow of hydrogen from line section 28 into line section 29 and ultimately, to engine 13. Throttle 30 is connected to throttle pedal 50 via includes throttle cable 48. Throttle pedal 50 also operates the butterfly valves in the carburetor in engine 13 in a conventional manner.

As an option, the throttle can be controlled electronically by a digital fuel controller product. Such a product is commercially available from Autotronic Controls Corp., 1490 Henry Brennan Drive, El Paso, Tex. 79936.

Line portion 29 proceeds through splitter 32 and is split into upper line 34 and lower line 36. Lines 34 and 36 proceed to adapters 38. Additional back-flow prevention valves 39 and 41 are mounted between lines 34 and 36 and adapters 38. Braided Teflon.RTM. is the preferred material for the fuel lines of the system.

Battery 42 is connected via battery cable 43 to magneto 44. Magneto 44 then controls the current through spark plug lines 46 to adapters 38 in a conventional manner. Thus, the regulation of the spark that ignites the fuel in the cylinder of engine 13 is conventional. Additionally, the intake of air into the engine 13 is accomplished through the standard carburetor system of engine 13. In a car, the sparking of the spark plugs can be controlled by the standard distributor/battery ignition system.

A second solenoid 26B controls the flow of gasoline from tank 23 through gasoline line 52. Solenoid 26B is controlled by switch 24B. Throttle 30 is connected to gasoline line 52 and can also be used to control the flow of gasoline to engine 13. Thus, this invention can be used with both an alternative fuel and gasoline.

Referring now to FIGS. 2, 3 and 4, the adapter 38 of this invention is shown in more detail. Adapter 38 includes housing 60. A threaded port 62 is formed into housing 60 through port block 64. Hydrogen lines are threadably connected to the adapter 38 via port 62. Port 62 is also referred to as an entrance port.

Adapter 38 also includes spark plug 66 which is effectively a standard spark plug except for the modification to the electrode as described below. Spark plug 66 includes a threaded end 68 which is threadably received in the threaded upper end 70 of housing 60. Spark plug wires 46 (as seen in FIG. 1) are connected to connector end 72 of spark plug base 66.

Electrode 76 protrudes from spark plug 66 and is of a length sufficient so that the threaded tip 77 (best seen in FIGS. 3 and 4) of the electrode 76 is proximate to the cylinder of the internal combustion engine when the adapter 38 is mounted. In one embodiment the electrode 76 is co-terminus with an exit port 88 of the adapter 38. This is most clearly shown in FIG. 4. Electrode 76 is preferably platinum to enhance catalysis of combustion by-products, as will be more fully discussed below.

Adapter 38 also includes plug 79. Plug 79 includes an axial internal void space 80. Thus, plug 79 fits over electrode 76 via internal void space 80. Plug 79 preferably comprises a ceramic material that acts as a good insulator. A suitable source for ceramic plug 79 is AM/PRO Machinery, Inc. 134 Church Street, Plain City, Ohio 43064-1321.

An important aspect of this invention is found on the external surface of plug 79. An annular groove 82 is formed on the surface of plug 79 near its distal end 81 so as to be proximal to the port 62 where hydrogen enters housing 60. Two stratifying grooves 84 proceed along the surface of plug 79 until they reach its distal end 83. It is these mixing structures, i.e. annual groove 82 and stratifying grooves 84, that facilitate the improved power generation from alternative fuels as provided by this invention. It will be apparent to those skilled in the analogous arts that the mixing structure may be formed on an inner surface 60i of the adapter housing 60. Alternatively, a combination of vortex producing structures may be attached to, and preferably integral with, both the plug 79 and the adapter housing inner surface 60i. Also, any conventional vortex producing means would be sufficient, i.e. grooves, per se, are not required. An entrance area 62a proximate the entrance port 62 is typically defined between the housing inner surface 60i and the external surface of plug 79. The external surface of plug 79 typically includes the grooves 84. So, the entrance area 62a is the volume of space within the housing, proximate the entrance port 62, which the gas will be able to occupy. This is most clearly shown in FIG. 4.

On threaded tip 77 of electrode 76, there are mounted a spacer nut 78 and a spark gap nut 85. As best seen in FIG. 2, the spark gap nut 85 includes a protrusion 85a that provides the spark gap between electrode 76 and periphery 89 of the exit port 88 of housing 60. The exit port 88 includes an exit area 88a. This is shown most clearly in FIGS. 2 and 4. The exit area 88a is no larger than the entrance area 62a. If the exit area 88a is smaller than the entrance area 62a, the pressure on the gas will increase. This increases the rate at which the gas is expelled from the plug 79, typically out through grooves 84. The electrode 76 is substantially co-terminus with the plug distal end 83 (seen most clearly in FIG. 2). Thus, the electrode 76 is substantially co-terminus with the exit port 88 or exit area 88a. Note that tip of the electrode 77 is preferably proximate with the exit area 88a because this is the location of greatest mixing. This is most clearly seen in FIG. 4. Preferably, the spark gap nut 85 is brass and the preferred spark gap distance is 25/1000 of an inch.

As best seen in FIG. 4, where arrows represent gas flow, hydrogen gas flows into chamber 60 via port 62. The hydrogen gas cools electrode 76, which facilitates combustion. Stratifying grooves 84, preferably there are two, act to mix the hydrogen that is flowing into housing 60 with ambient air therein. Though, more or less vortex creating grooves (a.k.a. stratifying grooves) may be desired. Typically, this will depend on the relative dimensions of the grooves and diameter of the plug as well as the pressure and velocity of the mixture in the housing. The mixing occurs within void space 86 of housing 60 and creates a vortexing or tornado action. Plug 79 tapers as it proceeds from its distal end 81 to its proximal end 83. Thus, proximal end 83 is smaller in diameter than distal end 81. This provides additional space within void space 86 for the mixing of the hydrogen and air. The stratified hydrogen/air are twisting in a vortex fashion when leaving housing 60 via exit port 88. The vortexing action produced by grooves 84 on the plug 79 creates a "tornado" 110 of hydrogen/air with a hollow center that generates a vacuum. Similar effect, or vortex flow path, would result were the grooves to be placed on the housing. The vacuum propels the hydrogen/air mixture into the cylinder of the engine for combustion. Such mixing cannot be accomplished in a conventional induction system.

Referring now to FIG. 5, a schematic of the system of this invention is portrayed, with arrows representing gas flow. Hydrogen flows in through port 62 through adapter 38 and into cylinder 92. Air enters cylinder 92 through air intake 90 and air intake valve 91. The stratified hydrogen/air mixture is compressed in the compression stroke of the engine via piston 94 and push rod 96, which are standard structures in an internal combustion engine. In the power stroke shown in FIG. 5-C, a spark is generated in the compressed stratified hydrogen/air mixture ignites and drives piston 94 and push rod 96 to power the engine.

The mass energy density of hydrogen is enhanced through the increased pressure on the hydrogen exerted by piston 94 on the hydrogen/air mixture within the cylinder 92 in the uptake stroke in an internal combustion engine. The pressurized and well-mixed hydrogen/air mixture is rapidly ignited through the spark because hydrogen has a fast flame front. This reduces problems with timing. Thus, the energy store found in hydrogen is captured through this system via the compression of the hydrogen gas within cylinder 92 and via the thorough mixing of hydrogen and air provided by adapter 38.

In addition, in the exhaust phase in a four cycle engine a second spark is generated from the plug to ensure complete combustion of by-products. This is enhanced by using an electrode made of platinum. Exhaust leaves cylinder 92 through exhaust port 98. Exhaust port 98 is closed during other phases by exhaust outlet valve 100.

In the exhaust stroke, because air valve 91 is closed, the substantial component of the gases in cylinder 92 is hydrogen. This is also accomplished due to the fact that in the system of this invention, the hydrogen is continuously flowing. When this spark interacts with what is predominately hydrogen, a plasma is formed. This plasma blocks back-flow of air from exhaust port 98. Prevention of back-flow of air prevents the hydrogen fuel system from backfiring, which is another common problem of hydrogen fuel systems. In effect, then, the plasma formed by the hydrogen acts as a virtual valve in that the plasma blocks the back-flow of air.

When the plasma of the hydrogen is generated as described above, UV radiation is generated which also facilitates consumption of combustion by-products. This increases the environmental friendliness of the system.

In addition to hydrogen the following fuels can be used in this system: natural gas; alcohol; gasohol; propane and buthane. A mixture of hydrogen and natural gas can also be used. As noted above, gasoline can be used in this system when solenoid 26 is used to stop the flow of hydrogen.

Further, it is contemplated that an adapter that is described herein for use with an internal combustion engine can also be used with a home heating system such as a home heating system that may currently use natural gas. The vortexing action created by the plug would enable appropriate power to be captured from the hydrogen fuel so as to effectively heat a home.

Thus, the apparatus and system of this invention provides an effective means to use alternative fuels. The use of alternative fuels can enhance the environment by reducing the use of conventional gasoline and the environmentally detrimental by-products caused by the consumption thereof.

Thus, although there have been described particular embodiments of the present invention of a new and useful "Fuel System For Internal Combustion System And Adapter For Use In Same", it is not intended that such references be construed as limitations upon the scope of this invention except as set forth in the following claims. Further, although there have been described certain dimensions used in the preferred embodiment, it is not intended that such dimensions be construed as limitations upon the scope of this invention except as set forth in the following claims.

United States Patent 6,119,651

September 19, 2000

Hydrogen Powered Vehicle, Internal Combustion Engine, and Spark Plug for use in same

Abstract --- The internal combustion engine fuel system described includes a structure for mixing the alternative fuel, preferably hydrogen, with oxygen in ambient air to stratify the fuel. The system includes an adapter, and the adapter includes a housing mounted between spark plug and cylinder of the internal combustion engine. A plug is placed within the housing. The plug has ridges or grooves on its outer surface that act as mixing structures. Thus, when hydrogen is introduced into the adapter housing it is mixed with ambient oxygen within the chamber as it flows over the plug. The mixing structures in the housing creates a vortexing action as the hydrogen flows over the plug

and towards the cylinder of the engine. An electrode protrudes from the plug towards the cylinder. The electrode is preferably platinum and generates the necessary spark to create combustion of the hydrogen/air mixture adjacent to the cylinder to thereby power the cylinder in the engine. A platinum electrode is preferably used because it enhances a catalytic conversion of combustion by-products to more environmentally compatible products. The present invention also teaches a spark plug producing a hotter spark for a hydrogen fuel system. Also taught is a hydrogen powered vehicle with reduced emissions by producing a spark during the power stroke and the exhaust stroke. Methods for reducing exhaust pollution are also taught. Methods of reducing exhaust pollution by generating a plasma are taught as well.

Inventors: Anderson; Herman P. (Brentwood, TN)
Assignee: Herman P. Anderson Technologies, LLP

Current U.S. Class: 123/297 ; 123/635; 123/DIG.12
Current International Class: F02M 21/02 (20060101); F02B 43/00 (20060101); F02M 057/06 ()
Field of Search: 123/3,DIG.12,527,297,636,637,635,169EL 313/120

References Cited [Referenced By]

U.S. Patent Documents

2208030	July 1940	Holmes
3165099	January 1965	Vanderpoel
3926169	December 1975	Leshner et al.
3980061	September 1976	McAlister
4164912	August 1979	Beyler
4186712	February 1980	Fitzner et al.
4319552	March 1982	Sauer et al.
4343272	August 1982	Buck
4383198	May 1983	Hosking
4448160	May 1984	Vosper
4820957	April 1989	Zivkovich
5852993	December 1998	Anderson

Foreign Patent Documents

686943	Aug., 1930	FR

Description

BACKGROUND OF THE INVENTION

The present invention relates generally to a fuel system for internal combustion engines that enables the invention to utilize alternative fuels; and, more particularly to an improved fuel system for an internal combustion engine and an adapter and spark plug for use in same.

The problems of air pollution caused by automobiles, lawn mowers and other vehicles that have internal combustion engines powered by gasoline is well documented. This problem is particularly described in an article in the San Diego Union Tribune, Wednesday, Sep. 25, 1996 and in Popular Science, August 1996. Indeed, under the Clean Air Act, in the next few years cars that do not run on standard fossil fuels like gasoline will be required in several states. Currently, there are no vehicles available to meet this need.

Thus, it will be appreciated by those skilled in the art that there has been a push to seek other fuels for use in internal combustion engines. These fuels include methane, propane and gasohol.

Hydrogen has also been characterized as a promising fuel source. See Department of Energy publication entitled "Hydrogen Fuel", printed in 1978, U.S. Government Printing Office, U.S. Department of Energy, Office of Public Affairs, Washington, DC 20585. Particularly, the graph presented in FIG. 6, which is copied from this publication, shows the potential for hydrogen in BTUs/Lb. However, the mass energy density for hydrogen is low; and thus, hydrogen needs to be compressed before its energy potential can be realized. Hydrogen is a "dry fuel" that is more difficult to ignite than "wet fuels." It becomes more difficult to ignite when it is cooled. However, cooling is preferred in most systems to avoid premature ignition of the fuel. Other systems deliver the hydrogen under pressure to prevent ignition in the supply line, which is a form of pre-ignition. No prior art fuel system has been able to capture the energy of hydrogen.

Several fuel systems for use with hydrogen have been proposed. Examples are described in U.S. Pat. Nos. 4,167,919; 4,253,428; 4,016,836; 4,178,882; 5,222,993; 5,085,176; and 5,085,176. The problem with these prior art fuel systems is they cannot take advantage of the alternative fuels to generate sufficient power from the fuels to drive an internal combustion engine for use in a lawn mower, automobile or other vehicle. As noted above, this is a particular problem when hydrogen is used as a fuel. Additionally, the prior art systems do not address the problem of back fire, i.e. undesired explosion of the fuel in the cylinder during the exhaust stroke of the internal combustion engine.

Premature ignition or backfire is often a problem with hydrogen fueled systems. U.S. Pat. No. 4,383,198 (Hosking) teaches a fuel injection spark plug with a pre-cooling chamber to avoid premature ignition indentination of a gaseous fuel supplied to the spark plug. U.S. Pat. No. 4,383,198, Hosking, entitled "Fuel Injection Spark Plug", issued May 10, 1983, is hereby incorporated herein by reference. However, cooling hydrogen makes it more difficult to ignite. This is because hydrogen is a dry fuel whereas gas is generally considered a wet fuel. Thus to reduce fuel pollution one prefers to utilize a fuel such as hydrogen and ignite it with a hotter spark.

One patent discussing very high energy (VHE) ignition systems is U.S. Pat. No. 5,207,208 issued to Ward, entitled "Integrated Convertor High Power CD Ignition", issued May 4, 1993. U.S. Pat. No. 5,207,208 is hereby incorporated herein by reference. Pat. '208 teaches very high energy ignition systems, for instance on the order of 36K volts. It teaches voltage doubling utilizing pulsating sparks of peak current on the order of two amps. While 36K volts may at times be adequate, it is preferable to achieve a higher voltage in the ignition chamber when using hydrogen.

It is known in the art to use a direct ignition system, also referred to as a distributorless ignition system, to increase the spark plug voltage. Such direct ignition systems use a "waste spark" method of spark distribution. Each cylinder is paired with its opposing cylinder in the firing order (1-4, 2-3, on a four cylinder, 1-4, 2-5, 3-6 on a V6) so that one cylinder on compression fires simultaneously with its opposing cylinder on exhaust stroke. Since the cylinder on exhaust stroke requires very little of the available voltage to fire its plug, most of the voltage is used to fire the cylinder on the compression stroke.

U.S. Pat. No. 4,462,380 discusses a distributorless ignition system. U.S. Pat. No. 4,462,380 entitled "Enhanced Spark Energy Distributorless Ignition System" issued Jul. 31, 1984 by Asik is hereby incorporated herein by reference. Pat. '380 teaches a distributorless ignition system of an internal combustion engine which has a supplementary spark energy module to increase spark energy. Each high voltage terminal is connected to a single spark plug and each ignition coil primary is alternately energized and quickly de-energized, producing opposite polarity ignition voltages at each coil terminal. As a result, pairs of spark plugs are alternately fired, with each firing pair occurring in a compression or exhaust stroke and thereby providing the proper ignition to the engine. Asik asserts that firing a spark plug does not affect engine performance or emissions. As taught by the instant invention this is incorrect for hydrogen fuels when the exhaust is sparked at a sufficient voltage.

U.S. Pat. No. 5,146,882 (Brinkman) teaches use of specific wet fuel mixtures in conjunction with direct ignition systems (D.I.S.). It teaches that technical breakdown voltages for typical spark plugs are on the order of 15-20 kV and DIS typically yield between 30 kV-40 kV. U.S. Pat. No. 5,146,882, entitled "Method And Apparatus For Cold Starting A Spark Ignited Internal Combustion Engine Fueled With An Alcohol-Based Fuel Mixture", by Brinkman, et al., issued Sep. 15, 1992, is hereby incorporated herein by reference. The previously referenced prior art discusses alternate fuels and fuel systems but does not teach a method of achieving a sufficient voltage with a hydrogen system.

What is needed, then, is a fuel system for an internal combustion engine that provides a way to generate sufficient power from the alternative fuels and that reduces backfire problems. Such a system is lacking in the prior art. Also lacking in prior art is a system utilizing alternate fuels with reduced pollution capabilities. Also lacking is a way of generating sufficient voltage spark to ignite the alternate fuel.

The following U.S. Patents discuss spark plugs designed for gas internal combustion engines to reduce radio frequency noise and interference. U.S. Pat. No. 4,224,554, entitled "Spark Plug Having A Low Noise Lever" issued to Nishio et al., Sep. 23, 1980 is hereby incorporated herein by reference. U.S. Pat. No. 4,029,990 entitled "Spark Plug Construction" issued to Nagy et al., Jun. 14, 1977 is hereby incorporated herein by reference.

SUMMARY OF THE INVENTION

It is an object of this invention to provide an improved fuel system for use with alternative fuels.

It is a further object of this invention to provide an improved fuel system for use particularly with hydrogen, and particularly to enhance the mass energy density of hydrogen.

It is an object to provide a hotter sparking plug to ignite the fuel.

It is still a further object of this invention to provide an adapter for use in a fuel system for an internal combustion engine that enhances the power obtained from the alternative fuels.

It is still a further object of this invention to provide an improved fuel system for an internal combustion engine that does not present the environmental hazards of a standard gasoline engine by providing reduced or zero emission of environmentally harmful gases.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that can use an alternative fuel and can use gasoline to facilitate the transition from gasoline to alternative fuels.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that reduces backfire problems.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that reduces backfire problems by cooling the electrode in the spark plug with the hydrogen gas fuel.

It is an object to reduce backfire by providing hydrogen at a sufficient pressure to reduce backfire and igniting it with a hotter spark (i.e. a higher voltage).

It is still a further object of this invention to provide an adapter for use in a fuel system for an internal combustion engine that enhances the power obtained from the alternative fuels.

It is a further object to generate a plasma field in a cylinder to burn residual exhaust and reduce emissions.

It is yet a further object of this invention to provide a direct injection fuel system wherein fuel is injected directly into the cylinder.

It is an object to provide a "2-fire" direct injection system, wherein the fuel is ignited during the power stroke and the exhaust stroke.

It is yet a further object of this invention to provide a way to convert the standard induction fuel system of a lawnmower to a direct injection fuel system wherein fuel is injected directly into the cylinder.

It is yet a further object of this invention to provide a way to convert the indirect injection fuel system of an automobile to a direct injection fuel system wherein fuel is injected directly into the cylinder.

The internal combustion engine fuel system of this invention includes a structure for mixing the alternative fuel, preferably hydrogen, with oxygen in ambient air to stratify the fuel. The term "alternative fuels" is known in the art and is meant to refer to fuels other than conventional gasoline.

The invention in one embodiment comprises an adapter, the adapter including a housing mounted between spark plug and cylinder of the internal combustion engine. A plug is placed within the housing. The plug has ridges or grooves on its outer surface that act as mixing structures. Thus, when hydrogen is introduced into the adapter housing it is mixed with ambient oxygen within the chamber as it flows over the plug. The mixing structures on the plug create a vortexing action as the hydrogen flows over the plug and towards the cylinder of the engine.

An electrode protrudes from the plug towards the cylinder. The electrode is preferably platinum and generates the necessary spark to create combustion of the hydrogen/air mixture adjacent to the cylinder to thereby power the cylinder in the engine. A platinum electrode is preferably used because it enhances a catalytic conversion of combustion by-products to more environmentally compatible products.

Thus, the combination of the compression action generated by the conventional piston and push rod within the cylinder of the engine in conjunction with the hydrogen/air mixing structure that vortexes the hydrogen/air create compressed hydrogen that is mixed with a combustion facilitator, i.e. oxygen in the ambient air to enable the system of this invention to capture the abundant energy found in hydrogen fuel.

The enhanced mixing capabilities of this invention is also adaptable for use with other alternative fuels such as natural gas, alcohol, propane and the like.

One embodiment of the present invention is for a hydrogen powered vehicle including an internal combustion engine operably connected to power the vehicle and adapted to connect to a hydrogen supply line. The internal combustion engine comprises a cylinder operably connected to the supply line; a piston operably disposed in the cylinder; and a spark plug having an electrode assembly in operable sparking communication with the cylinder, wherein the spark plug is capable of producing a spark of at least 45 kV.

Preferably the spark plug comprises a housing having a sparking end and a hollow chamber containing the electrode assembly. The electrode assembly comprises an electrode having a connector end and a sparking end, wherein the sparking end is positioned adjacent to the housing sparking end. The electrode assembly preferably comprises a coil electrically connected to the electrode connector.

Preferably the housing comprises a fuel inlet port in fluid communication with the supply line, and an insulator surrounding the electrode assembly, wherein the insulator is positioned in the hollow chamber. Preferably the insulator comprises an outer surface defining a first fuel channel having a first respective fuel exit port, wherein the first fuel channel is in fluid communication with the fuel inlet port and the fuel exit port exits into the cylinder.

The present invention also includes a method of operating an internal combustion engine comprising supplying hydrogen fuel to the engine and creating a spark in the cylinder during the power stroke and the exhaust stroke, wherein the spark created during the power stroke is at least 45 kV. In one embodiment the spark is at least 70 kV. In certain embodiments the method comprises generating plasma during the exhaust stroke. The exhaust fuel converts to an exhaust emission and coalesces with the plasma, thereby reducing fuel pollution.

Accordingly, it is an object in the invention to provide a spark plug producing a sufficiently hot spark to ignite hydrogen fuel, including cooled hydrogen fuel.

Another object of the present invention is to provide a device for reducing fuel exhaust emissions.

Another object of the present invention is to provide a cleaner burning engine.

Another object is to provide a spark plug for reducing emissions.

Another object of the present invention is to provide a device functional with dry fuels. Use of the term hydrogen, as discussed herein, is intended to encompass such dry fuels. However the term as used excludes such wet fuels as gas and alcohol.

Another object of the present invention is to provide an integrated fuel system adapted to provide these and objectives.

Other objects and advantages of the present invention will be apparent to those of skill in the art by reference to the teachings disclosed herein, including the attached drawings, detailed description of exemplary embodiments, and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a plan view of the fuel system of this invention with a lawn mower shown in phantom.

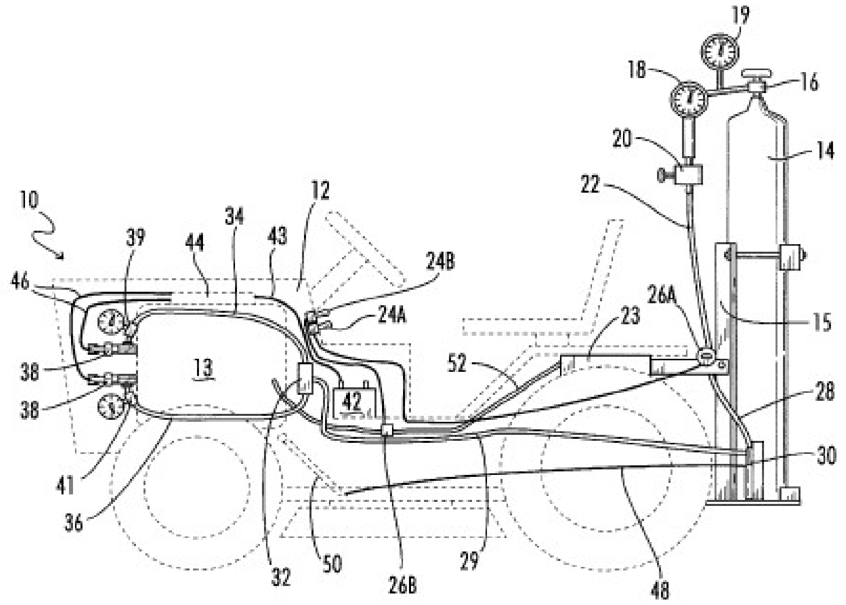
19
18
16
20
14
22
10
12
34
39
44
43
24B
46
26A
24A
23
15
13
52
38
38
42
28
30
41
36
32
50
26B
29
48

FIG. 1

FIG. 2 is a perspective view of the adapter of the fuel system of this invention.

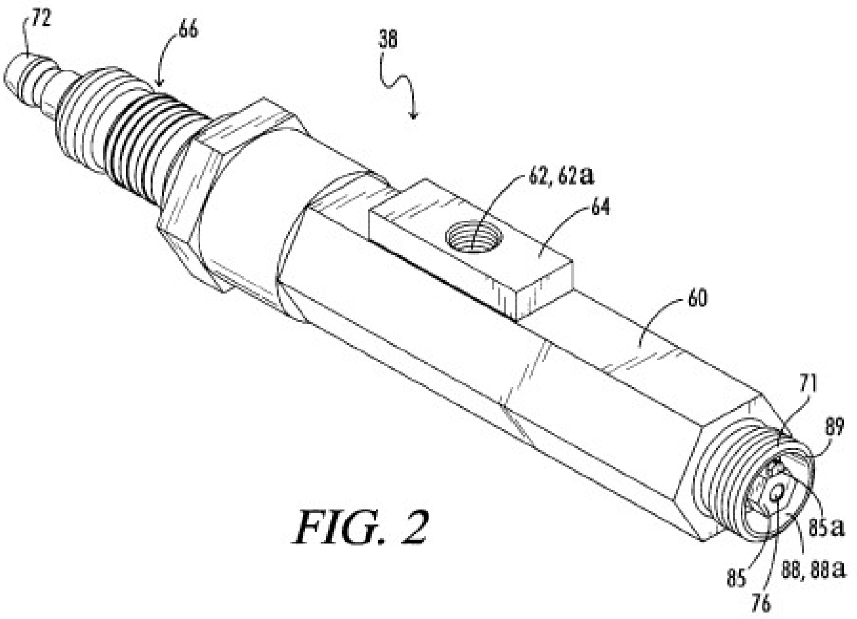
72
66
38
62, 62a
64
60
71
89
85a
85
88, 88a
76

FIG. 2

FIG. 3 is an exploded perspective view of the adapter of the fuel system of this invention.

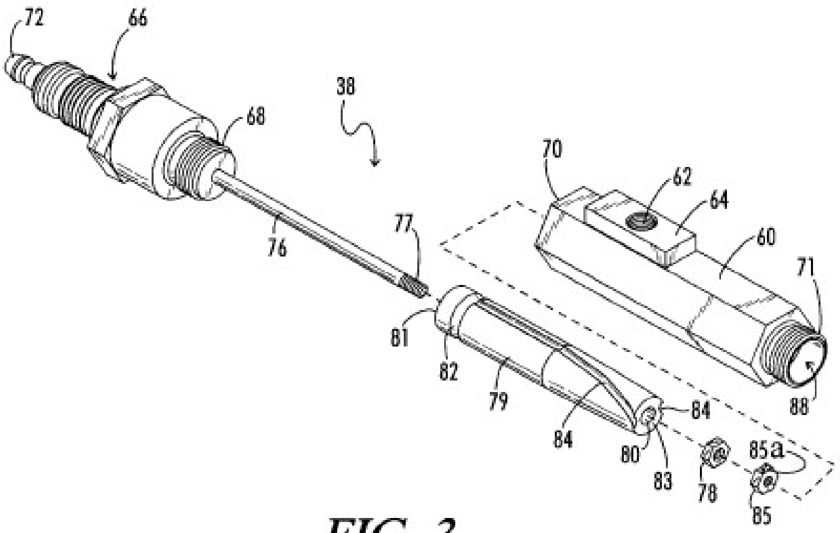
72
66
38
68
70
62
64
60
71
77
76
81
82
79
84
84
80
83
78
85a
85
88

FIG. 3

FIG. 4 is a cross-sectional view of the adapter of the fuel system of this invention.

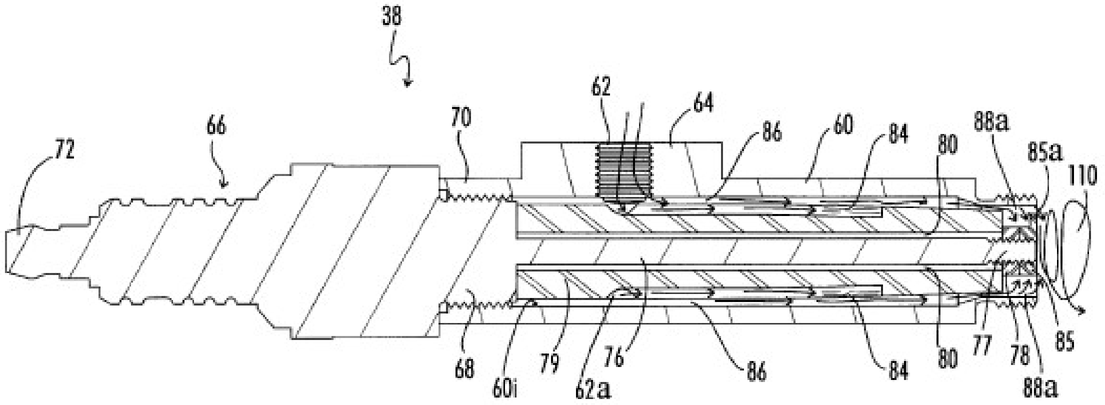
38
62
64
70
66
72
86
60
84
80
88a
85a
110
68
60i
79
62a
76
86
84
80
77
78
88a
85

FIG. 4

FIGS. 5A-5D are a schematic of the 4 cycle combustion in an internal combustion engine including the adapter of this invention.

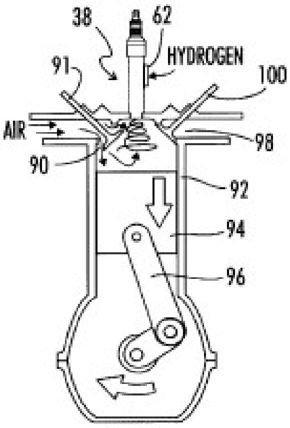
38
62
HYDROGEN
91
100
AIR
98
90
92
94
96

INTAKE

FIG. 5A

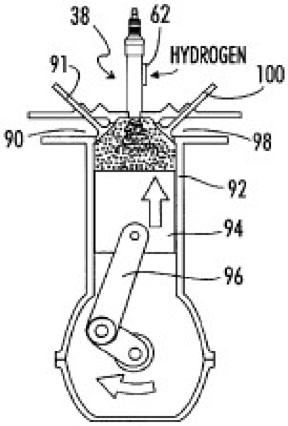
38
62
HYDROGEN
91
100
90
98
92
94
96

COMPRESSION

FIG. 5B

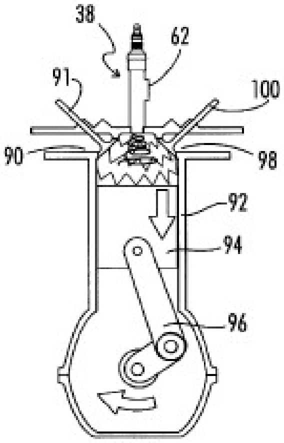
38
62
91
100
90
98
92
94
96

POWER

FIG. 5C

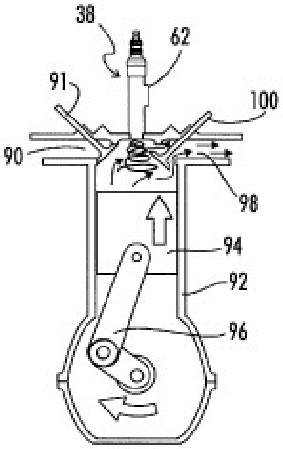
38
62
91
100
90
98
94
92
96

EXHAUST

FIG. 5D

FIG. 6 is a graph which shows the energy potential for hydrogen in BTUs/Lb.

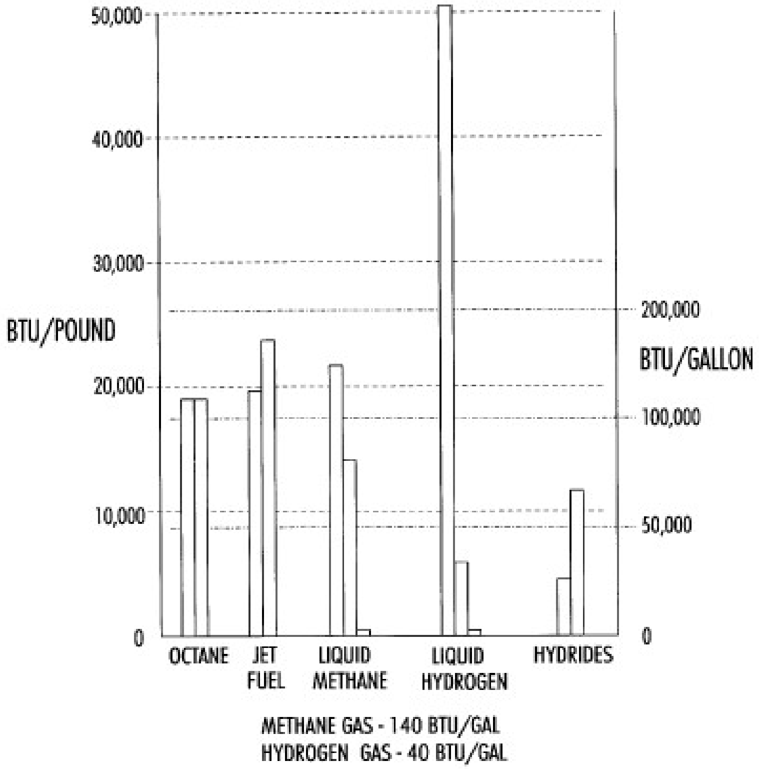

FIG. 6

FIG. 7 shows a vehicle of the present invention utilizing hydrogen fuel.

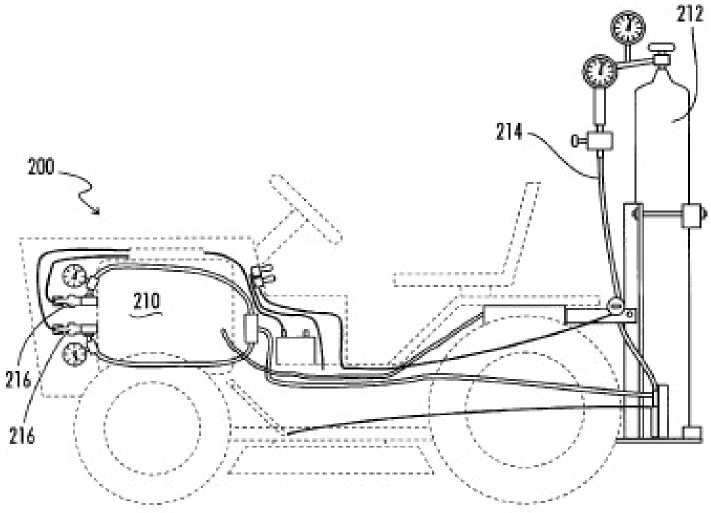

FIG. 7

FIG. 8 depicts an intake stroke of a cylinder connected to a hydrogen source.

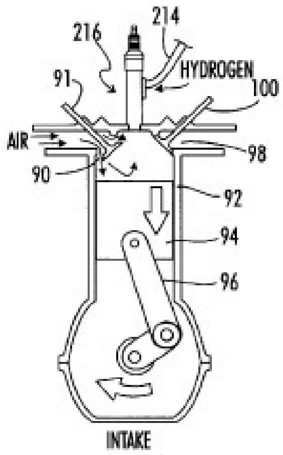
214
216
HYDROGEN
91
100
AIR
98
90
92
94
96
INTAKE

FIG. 8

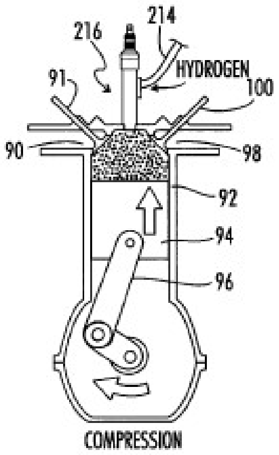
214
216
HYDROGEN
91
100
90
98
92
94
96
COMPRESSION

FIG. 9

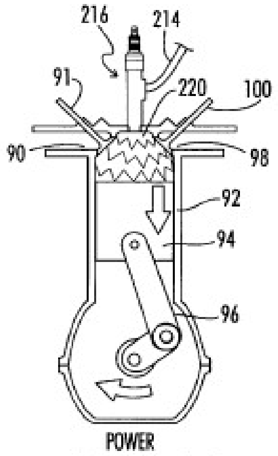
216
214
91
220
100
90
98
92
94
96
POWER

FIG. 10

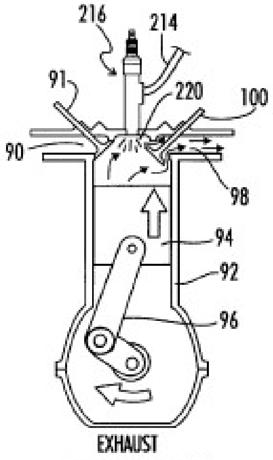
216
214
91
220
100
90
98
94
92
96
EXHAUST

FIG. 11

FIG. 9 depicts a compression stroke of a cylinder connected to a hydrogen source.
FIG. 10 depicts a power stroke of a cylinder connected to a hydrogen source. The spark plug shown in FIG. 10 is depicted sparking.
FIG. 11 depicts a cylinder in an exhaust stroke connected to a hydrogen source, wherein the spark plug is sparking during the exhaust stroke.

FIG. 12 shows an elevated side view of a spark plug of the present invention.

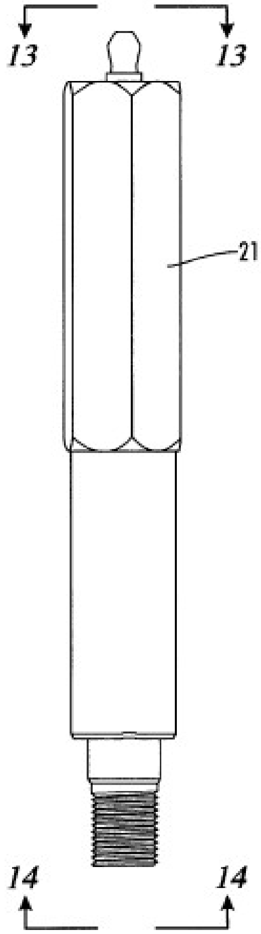

FIG. 12

FIG. 13 shows a plan view in the direction of view line 13--13 of the spark plug shown in FIG. 12.

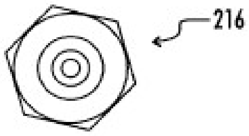

FIG. 13

FIG. 14 depicts a plan view in the direction of view line 14--14 of the spark plug shown in FIG. 12.

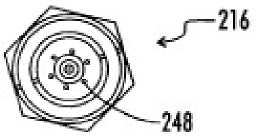

FIG. 14

FIG. 15 depicts an elevated side section view of the spark plug shown in FIG. 12.

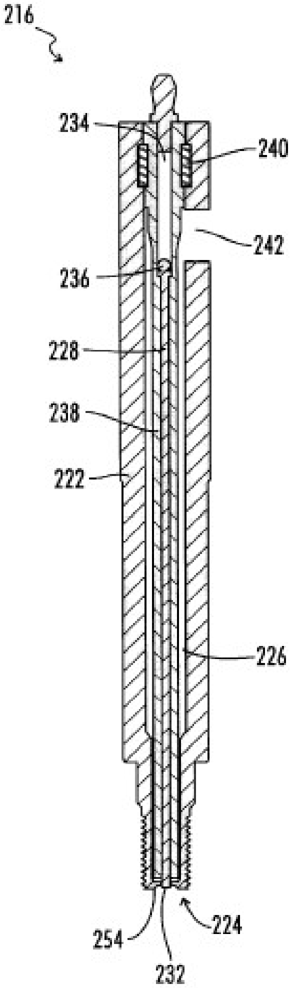
216
234
240
242
236
228
238
222
226
254
224
232

FIG. 15

FIG. 16 shows a electrode assembly which is disposed in the hollow chamber depicted in FIG. 15.

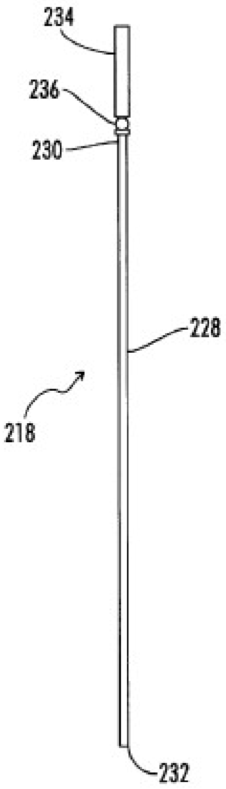

FIG. 16

FIG. 17 shows a section view of the spark plug housing clearly showing the hollow chamber and a torridal ground electrode.

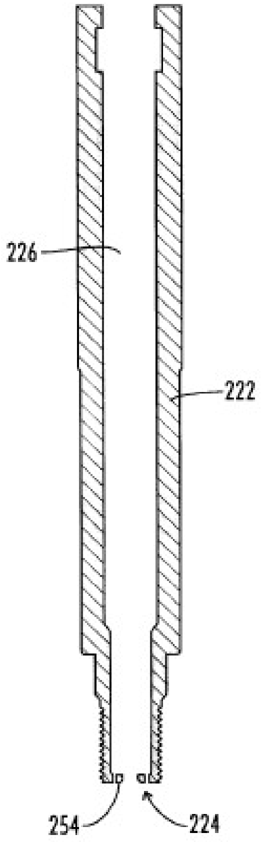

FIG. 17

FIG. 18 depicts an elevated side view of the spark plug housing and a fuel inlet port.

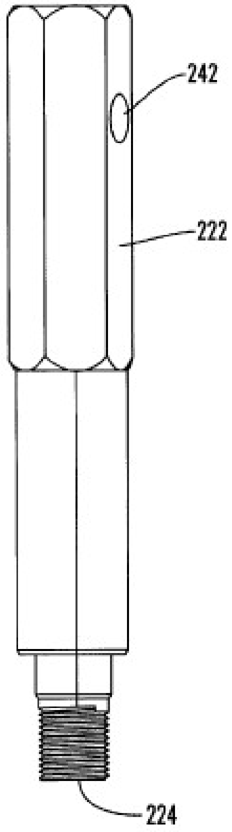

FIG. 18

FIG. 19 depicts an insulator comprising fuel channels and fuel exit ports.

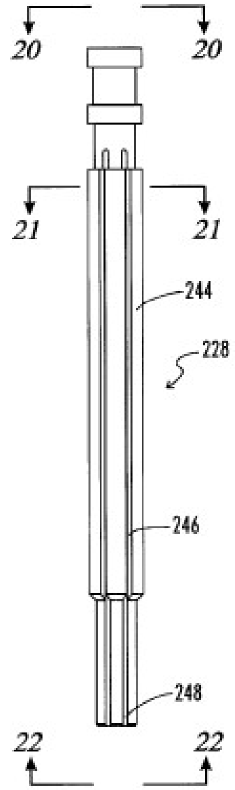
20
20
21
21
244
228
246
248
22
22

FIG. 19

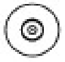

FIG. 20

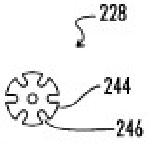

FIG. 21

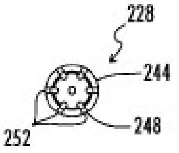
228
244
252
248

FIG. 22

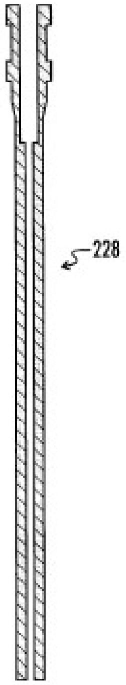

FIG. 23

FIG. 20 is a top plan view of the insulator shown in FIG. 19 viewed in the direction of view line 20--20.

FIG. 21 is a cross section view of the insulator shown in FIG. 19 along section line 21--21.

FIG. 22 is a cross section view of the insulator shown in FIG. 19 viewed in the direction of view line 22--22.

FIG. 23 is an elevated section view of the insulator shown in FIG. 19.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring now to the drawings, wherein like reference numerals refer to like parts throughout, the fuel system is referred to generally at 10. The fuel system of this invention is contemplated for use with an internal combustion power system, such as an internal combustion engine or such as a home heating system. Fuel system 10 is portrayed in this case as connected to a standard lawn mower. The lawn mower is shown in phantom at 12 in FIG. 1. The lawn mower is a conventional lawn mower as can be purchased from a company such as Murray of Ohio. The internal combustion engine in the lawn mower is a standard 4-cycle engine with two cylinders and develops 18 horsepower. A suitable example can be purchased from Briggs & Stratton, Milwaukee, Wis. 53201.

Referring particularly to FIG. 1, system 10 includes fuel tank 14 that is mounted to lawn mower 12 via mounting assembly 15. In the preferred embodiment, fuel tank 14 includes hydrogen gas. Accordingly, for the purposes of the foregoing detailed description, hydrogen gas will be described.

A line 22 begins at fuel tank 14 and proceeds to back-flow valve 16. Back-flow valve 16 is a standard valve and is available from AirCo Gases Company, 575 Mountain Avenue, Murray Hill, N.J. 07974, part no. WCS CV-4M. PSI tank gauge 19 is inserted in line 22 between back-flow valve 16 and flow pressure gauge 18. PSI tank gauge 19 is available from AirCo Gases Company also, part no. WCS CV5M, and serves as an indicator of fuel remaining in the tank 14. Flow pressure gauge 18 is used to monitor the flow of the hydrogen in pressure units. The flow is started using valve 16. The optimal flow pressure is 25 pounds of flow pressure from the tank 14 to the cylinder of the engine.

The flow of hydrogen is regulated via valve 20 that is part of gauge 18. Thus, when valve 20 is opened, hydrogen proceeds through line section 22 to hydrogen flow solenoid 26A. The user can access the hydrogen fuel by flipping hydrogen flow switch 24A into the "on" mode so that hydrogen gas in line 22 can proceed through hydrogen flow solenoid 26A and into line section 28.

Throttle 30 regulates the flow of hydrogen from line section 28 into line section 29 and ultimately, to engine 13. Throttle 30 is connected to throttle pedal 50 via includes throttle cable 48. Throttle pedal 50 also operates the butterfly valves in the carburetor in engine 13 in a conventional manner.

As an option, the throttle can be controlled electronically by a digital

fuel controller product. Such a product is commercially available from Autotronic Controls Corp., 1490 Henry Brennan Drive, El Paso, Tex. 79936.

Line portion 29 proceeds through splitter 32 and is split into upper line 34 and lower line 36. Lines 34 and 36 proceed to adapters 38. Additional back-flow prevention valves 39 and 41 are mounted between lines 34 and 36 and adapters 38. Braided Teflon.RTM. is the preferred material for the fuel lines of the system.

Battery 42 is connected via battery cable 43 to magneto 44. Magneto 44 then controls the current through spark plug lines 46 to adapters 38 in a conventional manner. Thus, the regulation of the spark that ignites the fuel in the cylinder of engine 13 is conventional. Additionally, the intake of air into the engine 13 is accomplished through the standard carburetor system of engine 13. In a car, the sparking of the spark plugs can be controlled by the standard distributor/battery ignition system.

A second solenoid 26B controls the flow of gasoline from tank 23 through gasoline line 52. Solenoid 26B is controlled by switch 24B. Throttle 30 is connected to gasoline line 52 and can also be used to control the flow of gasoline to engine 13. Thus, this invention can be used with both an alternative fuel and gasoline.

Referring now to FIGS. 2, 3 and 4, the adapter 38 of this invention is shown in more detail. Adapter 38 includes housing 60. A threaded port 62 is formed into housing 60 through port block 64. Hydrogen lines are threadably connected to the adapter 38 via port 62. Port 62 is also referred to as an entrance port. The entrance port has an entrance area 62a.

Adapter 38 also includes spark plug 66 which is effectively a standard spark plug except for the modification to the electrode as described below. Spark plug 66 includes a threaded end 68 which is threadably received in the threaded upper end 70 of housing 60. Spark plug wires 46 (as seen in FIG. 1) are connected to connector end 72 of spark plug base 66.

Electrode 76 protrudes from spark plug 66 and is of a length sufficient so that the threaded tip 77 (best seen in FIGS. 3 and 4) of the electrode 76 is proximate to the cylinder of the internal combustion engine when the adapter 38 is mounted. In one embodiment the electrode 76 is coterminus with an exit port 88 of the adapter 38. This most clearly is shown in FIG. 4. Electrode 76 is preferably platinum to enhance catalysis of combustion by-products, as will be more fully discussed below.

Adapter 38 also includes plug 79. Plug 79 includes an axial internal void space 80. Thus, plug 79 fits over electrode 76 via internal void space 80. Plug 79 preferably comprises a ceramic material that acts as a good insulator. A suitable source for ceramic plug 79 is AM/PRO Machinery, Inc. 134 Church Street, Plain City, Ohio 43064-1321.

An important aspect of this invention is found on the external surface of plug 79. An annular groove 82 is formed on the surface of plug 79 near its distal end 81 so as to be proximal to the port 62 where hydrogen enters housing 60. Two stratifying grooves 84 proceed along the surface of plug 79 until they reach its distal end 83. It is these mixing structures, i.e. annual groove 82 and stratifying grooves 84, that facilitate the improved power generation from alternative fuels as provided by this invention. It will be apparent to those skilled in the analogous arts that mixing structure may be formed in the inner surface of the adapter housing 60.

Alternatively a combination of vortex producing structures is integral with both the plug 79 and adapter housing 60. Also any conventional vortex producing means should be sufficient groves, per se, are not required.

On threaded tip 77 of electrode 76, there are mounted a spacer nut 78 and a spark gap nut 85. As best seen in FIG. 2, the spark gap nut 85 includes a protrusion 85a that provides the spark gap between electrode 76 and periphery 89 of the exit port 88 of housing 60. The exit port 88 includes an exit area 88a. Preferably, the exit area 88a is no larger than the entrance area 62a, though this is not critical. Preferably, the spark gap nut 85 is brass and the preferred spark gap distance is 25/1000 of an inch. Of course, the electrode may be coterminus with the plug distal end 83.

As best seen in FIG. 4, where arrows represent gas flow, hydrogen gas flows into chamber 60 via port 62. The hydrogen gas cools electrode 76, which facilitates combustion. Stratifying grooves 84, preferably there are two, act to mix the hydrogen that is flowing into housing 60 with ambient air therein. Though more or less vortex creating grooves (stratifying grooves) may be desired. Typically this will depend on the relative dimensions of the grooves and diameter of the plug as well as the pressure or velocity of the mixture in the housing. The mixing occurs within void space 86 of housing 60 and creates a vortexing or tornado action. Plug 79 tapers as it proceeds from its distal end 81 to its proximal end 83. Thus, proximal end 83 is smaller in diameter than distal end 81. This provides additional space within void space 86 for the mixing of the hydrogen and air. The stratified hydrogen/air are twisting in a vortex fashion when leaving housing 60 via exit port 88. The vortexing action produced by grooves 84 on the plug 79 creates a "tornado" 110 of hydrogen/air with a hollow center that generates a vacuum. Similar effect would result were t he grooves to be placed on the housing. The vacuum propels the hydrogen/air mixture into the cylinder of the engine for combustion. Such mixing cannot be accomplished in a conventional induction system.

Referring now to FIG. 5, a schematic of the system of this invention is portrayed, with arrows representing gas flow. Hydrogen flows in through port 62 through adapter 38 and into cylinder 92. Air enters cylinder 92 through air intake 90 and air intake valve 91. The stratified hydrogen/air mixture is compressed in the compression stroke of the engine via piston 94 and push rod 96, which are standard structures in an internal combustion engine. In the power stroke shown in FIG. 5-C, a spark is generated in the compressed stratified hydrogen/air mixture ignites and drives piston 94 and push rod 96 to power the engine.

The mass energy density of hydrogen is enhanced through the increased pressure on the hydrogen exerted by piston 94 on the hydrogen/air mixture within the cylinder 92 in the uptake stroke in an internal combustion engine. The pressurized and well-mixed hydrogen/air mixture is rapidly ignited through the spark because hydrogen has a fast flame front. This reduces problems with timing. Thus, the energy store found in hydrogen is captured through this system via the compression of the hydrogen gas within cylinder 92 and via the thorough mixing of hydrogen and air provided by adapter 38.

In addition, in the exhaust phase in a four cycle engine a second spark is generated from the plug to ensure complete combustion of by-products. This is enhanced by using an electrode made of platinum. Exhaust leaves cylinder 92 through exhaust port 98. Exhaust port 98 is closed during other phases by exhaust outlet valve 100.

In the exhaust stroke, because air valve 91 is closed, the substantial component of the gases in cylinder 92 is hydrogen. This is also accomplished due to the fact that in the system of this invention, the hydrogen is continuously flowing. When this spark interacts with what is predominately hydrogen, a plasma is formed. This plasma blocks back-flow of air from exhaust port 98. Prevention of back-flow of air prevents the hydrogen fuel system from backfiring, which is another common problem of hydrogen fuel systems. In effect, then, the plasma formed by the hydrogen acts as a virtual valve in that the plasma blocks the back-flow of air.

When the plasma of the hydrogen is generated as described above, UV radiation is generated which also facilitates consumption of combustion by-products. This increases the environmental friendliness of the system.

In addition to hydrogen the following fuels can be used in this system: natural gas; alcohol; gasohol; propane and buthane. A mixture of hydrogen and natural gas can also be used. As noted above, gasoline can be used in this system when solenoid 26 is used to stop the flow of hydrogen.

Further, it is contemplated that an adapter that is described herein for use with an internal combustion engine can also be used with a home heating system such as a home heating system that may currently use natural gas. The vortexing action created by the plug would enable appropriate power to be captured from the hydrogen fuel so as to effectively heat a home.

Another embodiment of the invention is a hydrogen powered vehicle 200 shown in FIG. 7. The hydrogen powered vehicle 200 includes an internal combustion engine 210 operably connected to power the vehicle 200 and adapted to connect to a hydrogen supply 212 through a supply line 214.

Referring to FIG. 8, the internal combustion engine 210 comprises a cylinder 92 operably connected to the supply line 214. A piston 94 is operably disposed in the cylinder 92. A spark plug 216 having an electrode assembly 218 (not shown in FIG. 8) is in operable sparking communication with the cylinder 92, wherein the spark plug 216 is capable of producing a spark 220 (shown in FIG. 10) of at least 45 kilovolts. FIGS. 8-11 are substantially similar to FIGS. 5A-5D depicting schematic representations of a 4-cycle piston-cylinder engine. Other piston and cylinder embodiments will be apparent to those with skill in the art. FIG. 8 represents an intake stroke. FIG. 9 depicts a compression stroke. FIG. 10 depicts a power stroke, in which the spark plug 216 sparks. FIG. 11 depicts an exhaust stroke, where the spark plug 216 is sparking. Refer to FIGS. 5A-5D discussed earlier for further description of similar components of FIGS. 8-11.

FIG. 12 shows one embodiment of the spark plug 216 in an elevated side view. FIG. 13 shows a plan view of the spark plug 216 along line 13--13 shown in FIG. 12. FIG. 14 shows a plan view of the spark plug 216 in the direction of line 14--14 shown in FIG. 12.

FIG. 15 depicts a section view of the spark plug 216. FIG. 15, section view, depicts a representative spark plug 216. The spark plug shown in FIG. 15 comprises a housing 222 having a sparking end 224 and a hollow chamber 226 containing the electrode assembly 218.

FIG. 16 depicts the electrode assembly 218 removed from the hollow chamber 226. The electrode assembly 218 depicted in FIG. 16 comprises an electrode 228. The electrode 228 has a connector end 230 and a sparking end 232. As shown in FIG. 15 the sparking end 232 of the electrode 228 is positioned adjacent to the housing sparking end 224. The electrode assembly further comprises a coil 234 electrically connected to the electrode connector end 230.

Referring to FIGS. 10 and 11, wherein the piston 94 moves through a power stroke in FIG. 10 and an exhaust stroke in FIG. 11, the spark plug 216 is shown to fire during the power stroke and the exhaust stroke.

The electrode assembly 218 shown in FIG. 16 comprises a steel ball 236 positioned between the coil 234 and the electrode connector end 230. Other operable ball-coil-electrode arrangements will be apparent to those of skill in the art.

The spark plug 216 shown in FIG. 15 also comprises an insulator 238 surrounding the electrode assembly 218 and positioned in the hollow chamber 226. A ceramic pore 240 is shown stabilizing the insulator 238 in the hollow chamber 226. Other methods of locating the insulator in the hollow chamber will be apparent to those of skill in the art.

An elevated section view of the housing 222 is shown in FIG. 17. An elevated side view of the housing 222 is shown in FIG. 18. In the exemplary embodiment of the housing 222 shown in FIG. 18, the housing 222 comprises a fuel inlet port 242 in fluid communication with the supply line 214 (not shown). FIG. 19 depicts an exemplary embodiment of the insulator 238. The insulator 238, shown in FIG. 19, comprises an outer surface 244 defining a first fuel channel 246 having a first respective fuel exit port 248. The first fuel channel 246 is in fluid communication with the fuel inlet port 242. The fuel exit port 248 exits into the cylinder 292.

FIG. 20 shows a plan view of the insulator 228 shown in FIG. 19 along line 20--20. FIG. 21

shows a section view of the insulator 228 shown in FIG. 19 viewed along section line 21--21. FIG. 22 depicts a bottom plan view of the insulator 228 shown in FIG. 19 along line 22--22. FIG. 23 shows a section view of the insulator 228 cut through section line 23--23 shown in FIG. 19.

In the embodiment of the insulator 228 shown in FIGS. 19-22, the insulator outer surface 244 defines a plurality of fuel channels 250, including the first fuel channel 246. The fuel channels 250 include a plurality of respective fuel exit ports 252, including the first fuel exit port 248. This is shown clearly in FIG. 22. Preferably, the plurality of fuel channel exit ports 252 are spaced equally, or equi-distant, about the electrode 228. This is shown well in FIGS. 21 and 22.

Referring to FIGS. 15 and 17, one embodiment of the housing 222 sparking and 224 comprises a torridal ground electrode 254 positioned about the electrode sparking end 232 (shown in FIG. 15). Preferably the housing sparking end 224 and the electrode sparking end 232 are co-terminous. In one embodiment the first respective fuel exit port 248 is co-terminous with the housing sparking end 224 and the electrode sparking end 232.

More generally the present invention includes an internal combustion engine 210 adapted to operate on hydrogen fuel supplied through a supply line 214. Typically the engine comprises a first cylinder 92 in fluid communication with the supply line 214; a first piston 94 disposed in the first cylinder 92 to move through a power stroke and an exhaust stroke (See FIGS. 10 and 11) in the first cylinder 92. The engine 210 also comprises a spark plug 216 having a housing 222 including a sparking end 224 and an electrode assembly 218 located in the housing 222, wherein the sparking end 224 is in operable communication with the cylinder 92 and capable of producing a spark 220 of at least 45 kV. Preferably, the spark plug 216 fires during the power stroke and the exhaust stroke. In one embodiment the housing sparking end 224 comprises a torridal ground electrode 254. The electrode assembly 218 comprises an electrode 228 having a sparking end 230 capable of sparking toward the torridal ground electrode 254. Preferably the electrode assembly 218 comprises a coil 234 in electrical communication with the electrode 228.

In one embodiment the electrode assembly 218 comprises an electrode 228 having a sparking end 230 and a coil 234 electrically connected to the electrode 228 opposite the electrode sparking end 230. This is the embodiment shown in FIG. 15. Typically the spark plug 216 comprises an insulator 228 surrounding the electrode assembly 218. Preferably the insulator 228 includes a fuel channel 246 having a fuel exit port 248 near the electrode sparking end 232.

It will be apparent to those with skill in the art that the present invention also comprises a method of operating an internal combustion engine 210 comprising the steps of supplying hydrogen fuel to the engine 210; moving a piston 94 in a cylinder 92 through a power stroke and an exhaust stroke; creating a spark 220 in the cylinder 92 during the power stroke and the exhaust stroke, wherein the spark created during the power stroke is at least 45 kV; and supplying the fuel to a location sufficiently near the spark to ignite the fuel. This is shown well in FIGS. 10, 15 and 19.

The present invention also includes a method wherein the step of creating a spark 220 comprises passing an electric current (not shown) through an electrode assembly 218 including a coil 234 electrically connected to an electrode 218 having a sparking end 232; insulating the electrode assembly 218; and arcing the electric current from the electrode sparking end 232 to a ground electrode 254. The ground electrode 254 as shown in FIGS. 15 and 17 a toroidal ground electrode 254.

In certain embodiments the step of creating a spark 220 includes creating a spark 220 having at least 70 kV. Creating a spark of sufficient voltage will generally reduce the exhaust pollution. Typically this happens by generating a plasma or plasma field during the exhaust stroke. Thus, one method of the present invention further comprises the steps of generating plasma during the exhaust stroke; and allowing exhaust fuel converted to an exhaust state to coalesce with the plasma. This reduces exhaust pollution.

Thus, the apparatus and system of this invention provides an effective means to use alternative fuels. The use of alternative fuels can enhance

the environment by reducing the use of conventional gasoline and the environmentally detrimental by-products caused by the consumption thereof.

Thus, although there have been described particular embodiments of the present invention of a new and useful "Hydrogen Powered Vehicle, Internal Combustion Engine, And Spark Plug For Use In Same", it is not intended that such references be construed as limitations upon the scope of this invention except as set forth in the following claims. Further, although there have been described certain dimensions used in the preferred embodiment, it is not intended that such dimensions be construed as limitations upon the scope of this invention except as set forth in the following claims.

Increase Your Mileage Now
30-80% Better Mileage Less Emission Starting at $375 Fast Shipping.
www.moneysavingdevice.com

UPS Hybrid Trucks
UPS, has the Largest Fleet of Low Emission Shipping Vehicles.
www.sustainability.ups.com

Top 10 Best Hybrids
New Top Rated and Performing Hybrid Vehicles Here!
www.AutoAndHybrid.com

Ads by Google

Your Support Maintains this Service -- and Your Survival ...

The Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...

Everything @ rexresearch.com, plus the Bonus Files CD !

ORDER PAGE

<< $13, Postpaid Anywhere >>

Rex Research, POB 19250, Jean, NV 89019 USA

rexresearch.com

Herman ANDERSON

Water Fuel System

Fuel Cell Plastic Parts Specializing in custom Fuel Cell molded or machined parts www.littlestarplastics.com
CFP® Certification Online Your Path to CFP® Starts Today. 3 Free Lessons. $1410 Off. Act Now! UGA.FinancialPlannerProg
6-Core 4-Way SuperServers Energy-efficient Supermicro Servers Featuring Intel® Xeon® Processors www.Supermicro.com
Ads by Google

---

http://waterpoweredcar.com/herman.html

GOT H2?
Photo by Dave Sieg © 2004

Herman P. ANDERSON
(1918-2004)

Herman P. Anderson Technologies, LLC
PO Box 2014
Brentwood, TN 37024-2014

Herman and his water-powered Chevy Cavalier.

by James Allen

He could drive it, but he was not allowed to sell any, or manufacture his discovery in the State of Tennessee. Why not?

Herman's special patented sparkplugs, similar to Stan Meyer's plugs.

Herman has consulted with Nasa and the US Airforce on some of our country's most important Top Secret Projects including the first US satellite in space, the SR-71 Blackbird, the Stealth Fighter/Bomber and Starwars. Herman worked closely with Dr. Wernher von Braun testing hydrogen powered rocket engines, and he also worked with engineers at the now famous Skunk Works, the Jet Propulsion Laboratory, JPL, and Cal Tech. While in the Air Force during WWII, Herman served as a fighter pilot, a flight instructor and bombardier trainer. He was certified in and flew 23 different airplanes during this period.

Herman believed in a Hydrogen Future and FREE fuel!

Herman's 1971 Ford LTD converted to run on hydrogen or gasoline, on display in the Water Fuel Museum in Lexington, KY. In a video interview with the inventor, he explains that ambient air is mixed with hydrogen (not oxygen) and a micron-sized fog of water mist (to mimic gasoline's burn), introduced in the same way propane gas is on a propane conversion. In fact, the LP regulator and LP aircleaner assembly are the same as on a propane vehicle. The butterfly in the carb, as in an LP system, now serves as the intake air control. The rest of the function of the gasoline carb is not used. He claimed greater power than gasoline, and 38 miles per gallon of water. As far as quantity of hydrogen produced, he produced enough gas from his electrolytic chamber to need a cut-off switch connected to a pressure gauge to stop gas production when it was no longer needed while driving. In other words, instead of producing more or less gas on demand, as some systems do, it appears his system always produced the same amount, but was turned on and off as needed. He emphasized that deuterium (heavy water) was essential to his approach, doubling the density of the hydrogen, making it twice as powerful. He also used high voltage, 70,000 volts to be exact, from two custom-made coils, to effect rapid separation of the component gases, a process which he calls 'radiolysis.' The 70,000 volts constitute what he calls a 'soft' x-ray, not radioactive but in need of shielding, in-between a microwave oven and a 'hard' x-ray. He was authorized by the state of Tennessee to drive his car as they recognized that he knew what he was doing with such technology. I view his prototype LTD as good proof of a watercar for those who doubt that the technology is even possible, but not necessarily the best approach for you and me to pursue. He accomplished his invention along the lines of his expertise, the field of ionics; others have approached it differently with less controversial technology and have been equally successful.

My thoughts about Herman

Herman's 1978 Fort LTD came to be known as The Green Machine. Using a modified propane carburetor, Herman converted the LTD to run on hydrogen in 1980. Due to the low mass energy density of hydrogen, Herman was only able to get a little more than half power from the engine with this traditional induction system.

Because Herman only used a single cell, I believe you need to do the 3 cells hooked up in series like Archie Blue, Dingel's and Peter's design. Use - 3 - 4"dia. SS tubes , 8" long (cells) welded together. Put your series of SS plates inside each cell. gas lines coming out the top through a check value to the bottom of the next cell. Use rubber caps ends with hose clamps so if it ever blows it just blows the caps off and the water is drained and your mild explosion is over with.

You need more molecules of hydrogen clustered together as they pass thru 3 cells then into the IMPCO propane fuel injection carburetor. In gasoline SAE words a higher octane of hydrogen (ortho hydrogen) before it enters the fuel injection chamber. first H2 and O, then H4 and O2 then H8 and O4?? Who is the mathematician here? I'm not him. Peter can tell you.

There is para-hydrogen and ortho-hydrogen molecules. Ortho-hydrogen molecules will cluster and stick together, as they pass through 3 cells or more they cluster together more and more, their proton travels in the same direction, when finally sent to the combustion chamber they have 2-1/2 times the explosive power than gasoline!

Para-hydrogen are hydrogen molecules that will run together but have repelling protons to each other, therefore the clustering does not happen, they are less combustive than ortho-hydrogen. If you produce ortho hydrogen, then this also creates deuterium in which is what you want. deuterium build up on the cylinder walls, values helps oxidation from happening.

Now Stan used one tall wide cell with 3 tubes inside one large tube. 12" diva. and about 24" tall. Go look at his files.

To increase the power of the hydrogen, he irradiated it making deuterium. The extra proton was more than enough to allow the engine to achieve full power. It is good to get Deuterium built up in your piston walls, from the hydrogen so the water exhaust will not rust out the motor. They say AMSCO synthetic oil helps deal with the water in the oil better, helps it vaporize the water molecules.

Unfortunately, state officials decided the radioactive source under the hood would not be safe for widespread use. They did however allow Herman to continue driving the car and he put nearly 30, 000 miles on the car running on hydrogen. Realizing the technology could be improved, Herman went back to the drawing board.

It is kinda of ironic in that is exactly what a Ballard fuel cell is also doing, when in use in the new Hybrid Cars. The Ballard Fuel Cell is radioactively produce hydrogen to burn mixed with the air and the gasoline. The new hybrid cars are just as safe, except they use 1/2 gas and 1/2 electric. The Hybrids also contain radioactive batteries. The only difference was Herman's Wonderful car ran for FREE and with the new hybrid cars, you still have to PURCHASE the GASOLINE.

Yes the Mean Green Dancing Herman Machine was in the heart of every reader passing by with a BRAIN. Addicted to oil you drive to work, paying the largest corp. the largest profit in American history ever! Last quarter of 2004's profit from Exxon/Mobil exceeded the record for the largest quarter profit of 8 billion dollars. We are all guilty, unless you walk to work ride a bike, or enjoy an electric train.

The Green Machine also employed onboard electrolysis, meaning that to fill the gas tank, Herman would pour in distilled water.

and a little se cert Electrolyte KOH (1-5% solution to water, is all)

Yes there is a car that "runs on water." The car is now parked in the research facilities at MTSU.

Just like the TESLA statue at Niagara Falls, NY, it is way off to the side. And Edison's bust is real big in the middle with lights shinning on him. and he didn't even invent AC electricity, he was angry with Tesla's invention of AC, because it made his DC look non efficient. Today in all our households, we use AC electricity. It could have also been FREE wireless electricity. J.P. Morgan quit funding Tesla's FREE electricity Tower, because he could not find where to put the meter.

Next 10,000 people with water powered cars can join the STAN MEYERS 10.000 DUNE BUGGY PLANET MARATHON CLUB.

All truth passes through three stages: First, it is ridiculed; Second, it is violently opposed; and
Third, it is accepted as self-evident. -- Arthur Schopenhauer (1788-1860)

Fuel System for Internal Combustion System and Adapter for Use in Same

USP # 5,852,993

1998-12-29

Applicant: HERMAN P ANDERSON TECHNOLOGIES (US)
Classification: - international: F02B43/00; F02M21/02; F02B43/00; F02M21/02; (IPC1-7): F02M57/06; - European: F02B43/00; F02M21/02A
Also published as: US6119651 (A1)

Abstract -- The internal combustion engine fuel system described includes a structure for mixing the alternative fuel, preferably hydrogen, with oxygen in ambient air to stratify the fuel. The system includes an adapter, and the adapter includes a housing mounted between spark plug and cylinder of the internal combustion engine. A plug is placed within the housing. The plug has ridges or grooves on its outer surface that act as mixing structures. Thus, when hydrogen is introduced into the adapter housing it is mixed with ambient oxygen within the chamber as it flows over the plug. The mixing structures in the housing create a vortexing action as the hydrogen flows over the plug and towards the cylinder of the engine. An electrode protrudes from the plug towards the cylinder. The electrode is preferably platinum and generates the necessary spark to create combustion of the hydrogen/air mixture adjacent to the cylinder to thereby power the cylinder in the engine. A platinum electrode is preferably used because it enhances a catalytic conversion of combustion by-products to more environmentally compatible products.

Inventors: Anderson; Herman P. (Brentwood, TN)
Assignee: Herman P. Anderson Technologies, LLC (Brentwood, TN)

Current U.S. Class: 123/297 ; 123/527; 123/DIG.12
Current International Class: F02B 43/00 (20060101); F02M 21/02 (20060101); F02M 057/06 ()
U.S. Patent Documents

3926169	December 1975	Leshner et al.
3980061	September 1976	Mcalister
4319552	March 1982	Sauer et al.
4343272	August 1982	Buck
4383198	May 1983	Hosking
4448160	May 1984	Vosper
4546740	October 1985	Clements et al.
5531199	July 1996	Bryant et al.

Description

BACKGROUND OF THE INVENTION

The present invention relates generally to a fuel system for internal combustion engines that enables the invention to utilize alternative fuels; and, more particularly to an improved fuel system for an internal combustion engine and an adapter for use in same.

The problems of air pollution caused by automobiles, lawn mowers and other vehicles that have internal combustion engines powered by gasoline is well documented. This problem is particularly described in an article in the San Diego Union Tribune, Wednesday, Sep. 25, 1996 and in Popular Science, August, 1996. Indeed, under the Clean Air Act, in the next few years cars that do not run on standard fossil fuels like gasoline will be required in several states. Currently, there are no vehicles available to meet this need.

Thus, it will be appreciated by those skilled in the art that there has been a push to seek other fuels for use in internal combustion engines. These fuels include methane, propane and gasohol.

Hydrogen has also been characterized as a promising fuel source. See Department of Energy publication entitled "Hydrogen Fuel", printed in 1978, U.S. Government Printing Office, U.S. Department of Energy, Office of Public Affairs, Washington, D.C. 20585. Particularly, the graph presented in FIG. 6, which is copied from this publication, shows the potential for hydrogen in BTUs/Lb. However, the mass energy density for hydrogen is low; and thus, hydrogen needs to be compressed before its energy potential can be realized. No prior art fuel system has been able to capture the energy of hydrogen.

Several fuel systems for use with hydrogen have been proposed. Examples are described in U.S. Pat. Nos. 4,167,919; 4,253,428; 4,016,836; 4,178,882; 5,222,993; 5,085,176; and 5,085,176. The problem with these prior art fuel systems is they cannot can take advantage of the alternative fuels to generate sufficient power from the fuels to drive an internal combustion engine for use in a lawn mower, automobile or other vehicle. As noted above, this is a particular problem when hydrogen is used as a fuel. Additionally, the prior art systems do not address the problem of back fire, i.e. undesired explosion of the fuel in the cylinder during the exhaust stroke of the internal combustion engine.

What is needed, then, is a fuel system for an internal combustion engine that provides a way to generate sufficient power from the alternative fuels and that reduces backfire problems. Such a system is lacking in the prior art.

SUMMARY OF THE INVENTION

It is an object of this invention to provide an improved fuel system for use with alternative fuels.

It is a further object of this invention to provide an improved fuel system for use particularly with hydrogen, and particularly to enhance the mass energy density of hydrogen.

It is still a further object of this invention to provide an adapter for use in a fuel system for an internal combustion engine that enhances the power obtained from the alternative fuels.

It is still a further object of this invention to provide an improved fuel system for an internal combustion engine that does not present the environmental hazards of a standard gasoline engine by providing reduced or zero emission of environmentally harmful gases.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that can use an alternative fuel and can use gasoline to facilitate the transition from gasoline to alternative fuels.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that reduces backfire problems.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that reduces backfire problems by cooling the electrode in the spark plug with the hydrogen gas fuel.

It is still a further object of this invention to provide an adapter for use in a fuel system for an internal combustion engine that enhances the power obtained from the alternative fuels.

It is yet a further object of this invention to provide a direct injection fuel system wherein fuel is injected directly into the cylinder.

It is yet a further object of this invention to provide a way to convert the standard induction fuel system of a lawnmower to a direct injection fuel system wherein fuel is injected directly into the cylinder.

It is yet a further object of this invention to provide a way to convert the indirect injection fuel system of an automobile to a direct injection fuel system wherein fuel is injected directly into the cylinder.

The internal combustion engine fuel system of this invention includes a structure for mixing the alternative fuel, preferably hydrogen, with oxygen in ambient air to stratify the fuel. The term "alternative fuels" is known in the art and is meant to refer to fuels other than conventional gasoline.

The invention comprises an adapter, the adapter including a housing mounted between spark plug and cylinder of the internal combustion engine. A plug is placed within the housing. The plug has ridges or grooves on its outer surface that act as mixing structures. Thus, when hydrogen is introduced into the adapter housing it is mixed with ambient oxygen within the chamber as it flows over the plug. The mixing structures on the plug create a vortexing action as the hydrogen flows over the plug and towards the cylinder of the engine.

An electrode protrudes from the plug towards the cylinder. The electrode is preferably platinum and generates the necessary spark to create combustion of the hydrogen/air mixture adjacent to the cylinder to thereby power the cylinder in the engine. A platinum electrode is preferably used because it enhances a catalytic conversion of combustion by-products to more environmentally compatible products.

Thus, the combination of the compression action generated by the conventional piston and push rod within the cylinder of the engine in conjunction with the hydrogen/air mixing structure that vortexes the hydrogen/air create compressed hydrogen that is mixed with a combustion facilitator, i.e. oxygen in the ambient air to enable the system of this invention to capture the abundant energy found in hydrogen fuel.

The enhanced mixing capabilities of this invention is also adaptable for use with other alternative fuels such as natural gas, alcohol, propane and the like.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a plan view of the fuel system of this invention with a lawn mower shown in phantom.

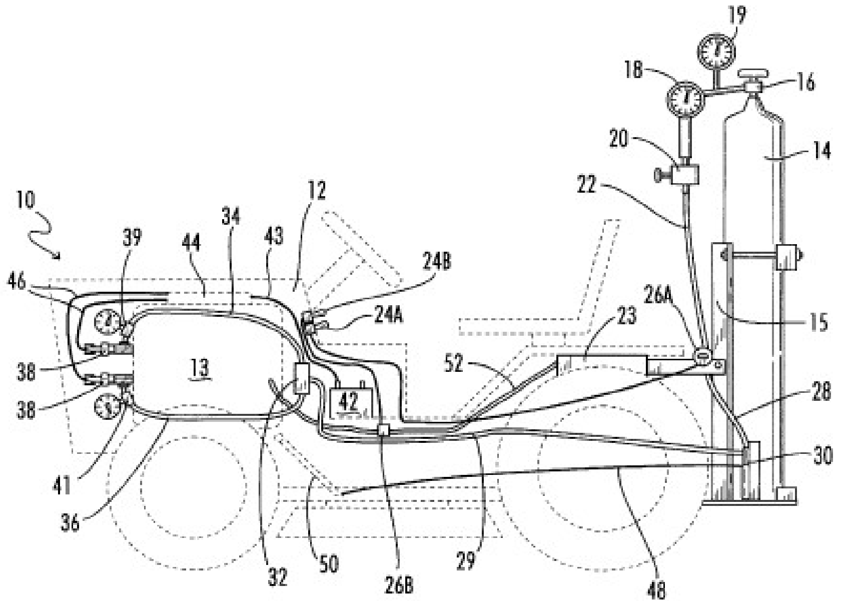
10
46
39
44
34
43
12
24B
24A
38
38
13
42
41
36
32
50
26B
29
48
52
23
26A
15
28
30
19
18
16
20
14
22

FIG. 1

FIG. 2 is a perspective view of the adapter of the fuel system of this invention.

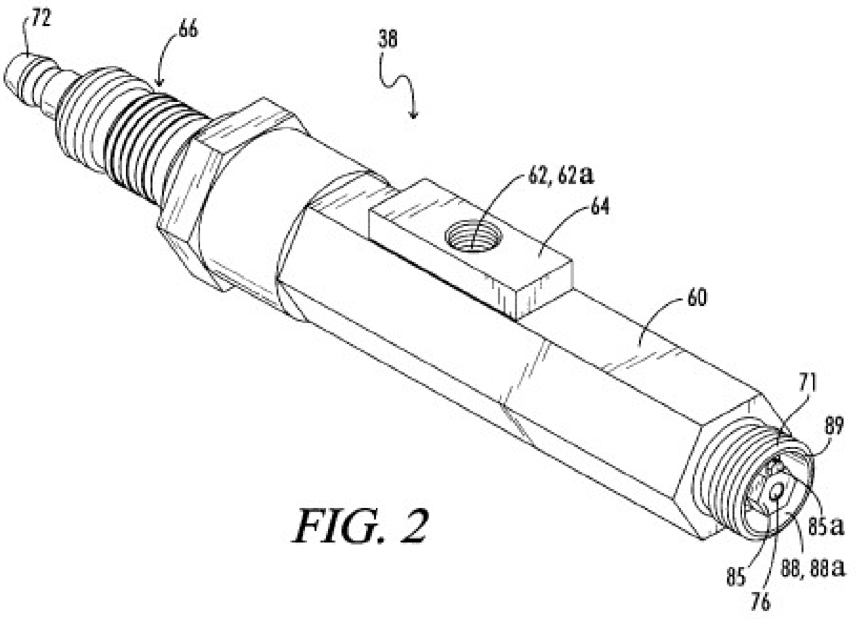
72
66
38
62, 62a
64
60
71
89
85a
88, 88a
85
76

FIG. 2

FIG. 3 is an exploded perspective view of the adapter of the fuel system of this invention.

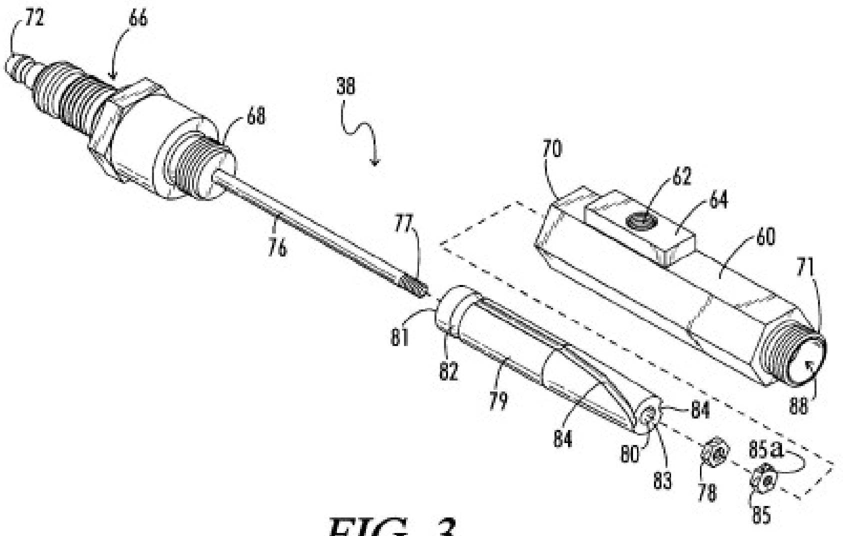
72
66
38
68
70
62
64
60
71
77
76
81
82
79
84
84
80
83
78
85a
85
88

FIG. 3

FIG. 4 is a cross-sectional view of the adapter of the fuel system of this invention.

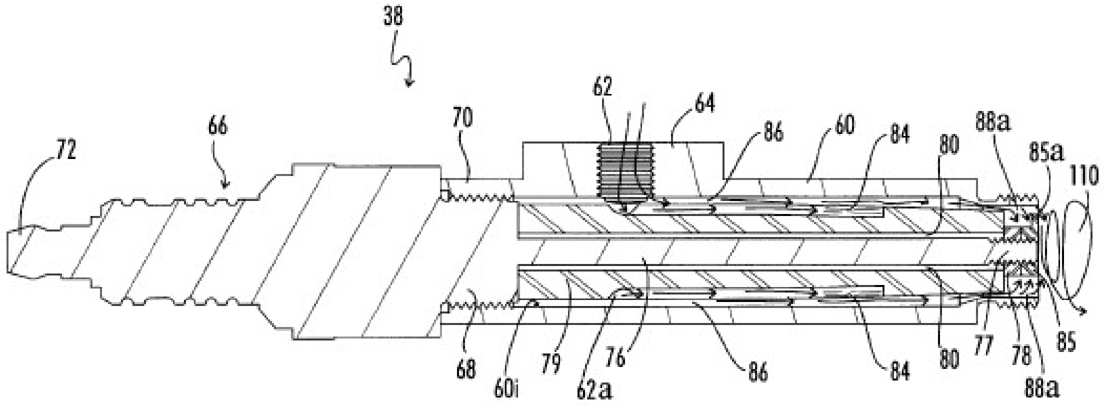
38
62
64
66
72
70
86
60
84
88a
80
85a
110
68
60i
79
62a
76
86
84
80
77
78
88a
85

FIG. 4

FIGS. 5A-5D are a schematic of the 4 cycle combustion in an internal combustion engine including the adapter of this invention.

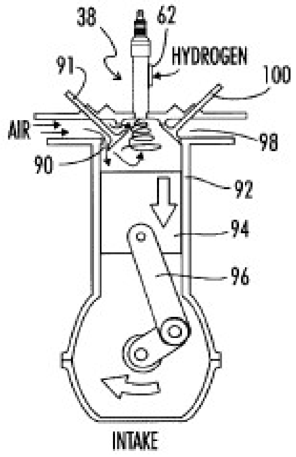
38
62
HYDROGEN
91
100
AIR
98
90
92
94
96
INTAKE

FIG. 5A

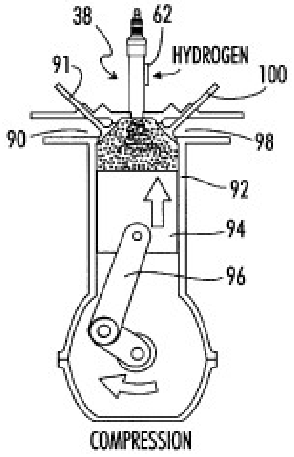
38
62
HYDROGEN
91
100
90
98
92
94
96
COMPRESSION

FIG. 5B

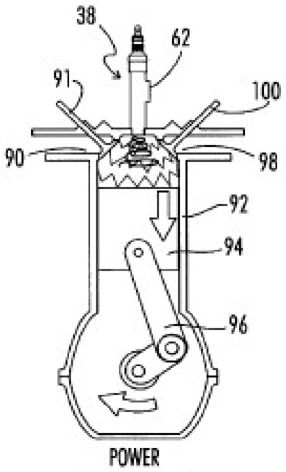
38
62
91
100
90
98
92
94
96
POWER

FIG. 5C

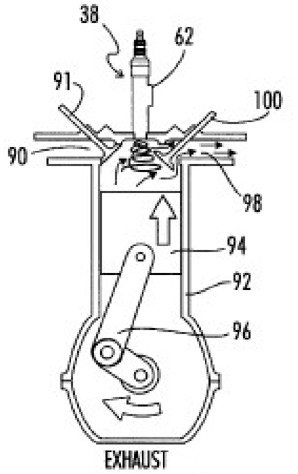
38
62
91
100
90
98
94
92
96
EXHAUST

FIG. 5D

FIG. 6 is a graph which shows the energy potential for hydrogen in BTUs/Lb.

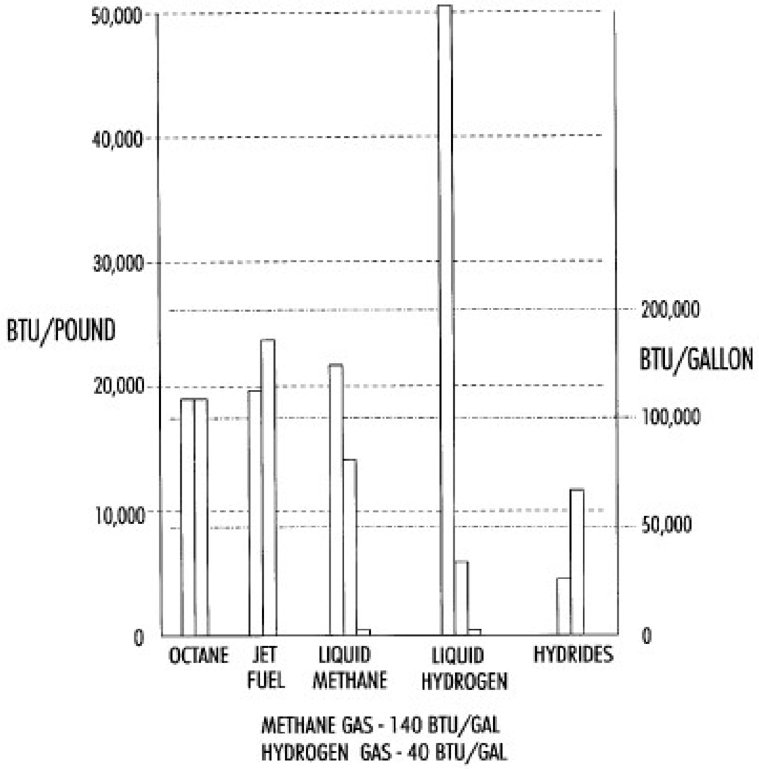

FIG. 6

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring now to the drawings, wherein like reference numerals refer to like parts throughout, the fuel system is referred to generally at 10. The fuel system of this invention is contemplated for use with an internal combustion power system, such as an internal combustion engine or such as a home heating system. Fuel system 10 is portrayed in this case as connected to a standard lawn mower. The lawn mower is shown in phantom at 12 in FIG. 1. The lawn mower is a conventional lawn mower as can be purchased from a company such as Murray of Ohio. The internal combustion engine in the lawn mower is a standard 4-cycle engine with two cylinders and develops 18 horsepower. A suitable example can be purchased from Briggs & Stratton, Milwaukee, Wis., 53201.

Referring particularly to FIG. 1, system 10 includes fuel tank 14 that is mounted to lawn mower 12 via mounting assembly 15. In the preferred embodiment, fuel tank 14 includes hydrogen gas. Accordingly, for the purposes of the foregoing detailed description, hydrogen gas will be described.

A line 22 begins at fuel tank 14 and proceeds to back-flow valve 16. Back-flow valve 16 is a standard valve and is available from AirCo Gases Company, 575 Mountain Avenue, Murray Hill, N.J. 07974, part no. WCS CV-4M. PSI tank gauge 19 is inserted in line 22 between back-flow valve 16 and flow pressure gauge 18. PSI tank gauge 19 is available from AirCo Gases Company also, part no. WCS CV-5M, and serves as an indicator of fuel remaining in the tank 14. Flow pressure gauge 18 is used to monitor the flow of the hydrogen in pressure units. The flow is started using valve 16. The optimal flow pressure is 25 pounds of flow pressure from the tank 14 to the cylinder of the engine.

The flow of hydrogen is regulated via valve 20 that is part of gauge 18. Thus, when valve 20 is opened, hydrogen proceeds through line section 22 to hydrogen flow solenoid 26A. The user can access the hydrogen fuel by flipping hydrogen flow switch 24A into the "on" mode so that hydrogen gas in line 22 can proceed through hydrogen flow solenoid 26A and into line section 28.

Throttle 30 regulates the flow of hydrogen from line section 28 into line section 29 and ultimately, to engine 13.

Throttle 30 is connected to throttle pedal 50 via includes throttle cable 48. Throttle pedal 50 also operates the butterfly valves in the carburetor in engine 13 in a conventional manner.

As an option, the throttle can be controlled electronically by a digital fuel controller product. Such a product is commercially available from Autotronic Controls Corp., 1490 Henry Brennan Drive, El Paso, Tex. 79936.

Line portion 29 proceeds through splitter 32 and is split into upper line 34 and lower line 36. Lines 34 and 36 proceed to adapters 38. Additional back-flow prevention valves 39 and 41 are mounted between lines 34 and 36 and adapters 38. Braided Teflon.RTM. is the preferred material for the fuel lines of the system.

Battery 42 is connected via battery cable 43 to magneto 44. Magneto 44 then controls the current through spark plug lines 46 to adapters 38 in a conventional manner. Thus, the regulation of the spark that ignites the fuel in the cylinder of engine 13 is conventional. Additionally, the intake of air into the engine 13 is accomplished through the standard carburetor system of engine 13. In a car, the sparking of the spark plugs can be controlled by the standard distributor/battery ignition system.

A second solenoid 26B controls the flow of gasoline from tank 23 through gasoline line 52. Solenoid 26B is controlled by switch 24B. Throttle 30 is connected to gasoline line 52 and can also be used to control the flow of gasoline to engine 13. Thus, this invention can be used with both an alternative fuel and gasoline.

Referring now to FIGS. 2, 3 and 4, the adapter 38 of this invention is shown in more detail. Adapter 38 includes housing 60. A threaded port 62 is formed into housing 60 through port block 64. Hydrogen lines are threadably connected to the adapter 38 via port 62. Port 62 is also referred to as an entrance port.

Adapter 38 also includes spark plug 66 which is effectively a standard spark plug except for the modification to the electrode as described below. Spark plug 66 includes a threaded end 68 which is threadably received in the threaded upper end 70 of housing 60. Spark plug wires 46 (as seen in FIG. 1) are connected to connector end 72 of spark plug base 66.

Electrode 76 protrudes from spark plug 66 and is of a length sufficient so that the threaded tip 77 (best seen in FIGS. 3 and 4) of the electrode 76 is proximate to the cylinder of the internal combustion engine when the adapter 38 is mounted. In one embodiment the electrode 76 is co-terminus with an exit port 88 of the adapter 38. This is most clearly shown in FIG. 4. Electrode 76 is preferably platinum to enhance catalysis of combustion by-products, as will be more fully discussed below.

Adapter 38 also includes plug 79. Plug 79 includes an axial internal void space 80. Thus, plug 79 fits over electrode 76 via internal void space 80. Plug 79 preferably comprises a ceramic material that acts as a good insulator. A suitable source for ceramic plug 79 is AM/PRO Machinery, Inc. 134 Church Street, Plain City, Ohio 43064-1321.

An important aspect of this invention is found on the external surface of plug 79. An annular groove 82 is formed on the surface of plug 79 near its distal end 81 so as to be proximal to the port 62 where hydrogen enters housing 60. Two stratifying grooves 84 proceed along the surface of plug 79 until they reach its distal end 83. It is these mixing structures, i.e. annual groove 82 and stratifying grooves 84, that facilitate the improved power generation from alternative fuels as provided by this invention. It will be apparent to those skilled in the analogous arts that the mixing structure may be formed on an inner surface 60i of the adapter housing 60. Alternatively, a combination of vortex producing structures may be attached to, and preferably integral with, both the plug 79 and the adapter housing inner surface 60i. Also, any conventional vortex producing means would be sufficient, i.e. grooves, per se, are not required. An entrance area 62a proximate the entrance port 62 is typically defined between the housing inner surface 60i and the external surface of plug 79. The external surface of plug 79 typically includes the grooves 84. So, the entrance area 62a is the volume of space within the housing, proximate the entrance port 62, which the gas will be able to occupy. This is most clearly shown in FIG. 4.

On threaded tip 77 of electrode 76, there are mounted a spacer nut 78 and a spark gap nut 85. As best seen in FIG. 2, the spark gap nut 85 includes a protrusion 85a that provides the spark gap between electrode 76 and periphery 89 of the exit port 88 of housing 60. The exit port 88 includes an exit area 88a. This is shown most clearly in FIGS. 2 and 4. The exit area 88a is no larger than the entrance area 62a. If the exit area 88a is smaller than the entrance area 62a, the pressure on the gas will increase. This increases the rate at which the gas is expelled from the plug 79, typically out through grooves 84. The electrode 76 is substantially co-terminus with the plug distal end 83 (seen most clearly in FIG. 2). Thus, the electrode 76 is substantially co-terminus with the exit port 88 or exit area 88a. Note that tip of the electrode 77 is preferably proximate with the exit area 88a because this is the location of greatest mixing. This is most clearly seen in FIG. 4. Preferably, the spark gap nut 85 is brass and the preferred spark gap distance is 25/1000 of an inch.

As best seen in FIG. 4, where arrows represent gas flow, hydrogen gas flows into chamber 60 via port 62. The hydrogen gas cools electrode 76, which facilitates combustion. Stratifying grooves 84, preferably there are two, act to mix the hydrogen that is flowing into housing 60 with ambient air therein. Though, more or less vortex creating grooves (a.k.a. stratifying grooves) may be desired. Typically, this will depend on the relative dimensions of the grooves and diameter of the plug as well as the pressure and velocity of the mixture in the housing. The mixing occurs within void space 86 of housing 60 and creates a vortexing or tornado action. Plug 79 tapers as it proceeds from its distal end 81 to its proximal end 83. Thus, proximal end 83 is smaller in diameter than distal end 81. This provides additional space within void space 86 for the mixing of the hydrogen and air. The stratified hydrogen/air are twisting in a vortex fashion when leaving housing 60 via exit port 88. The vortexing action produced by grooves 84 on the plug 79 creates a "tornado" 110 of hydrogen/air with a hollow center that generates a vacuum. Similar effect, or vortex flow path, would result were the grooves to be placed on the housing. The vacuum propels the hydrogen/air mixture into the cylinder of the engine for combustion. Such mixing cannot be accomplished in a conventional induction system.

Referring now to FIG. 5, a schematic of the system of this invention is portrayed, with arrows representing gas flow. Hydrogen flows in through port 62 through adapter 38 and into cylinder 92. Air enters cylinder 92 through air intake 90 and air intake valve 91. The stratified hydrogen/air mixture is compressed in the compression stroke of the engine via piston 94 and push rod 96, which are standard structures in an internal combustion engine. In the power stroke shown in FIG. 5-C, a spark is generated in the compressed stratified hydrogen/air mixture ignites and drives piston 94 and push rod 96 to power the engine.

The mass energy density of hydrogen is enhanced through the increased pressure on the hydrogen exerted by piston 94 on the hydrogen/air mixture within the cylinder 92 in the uptake stroke in an internal combustion engine. The pressurized and well-mixed hydrogen/air mixture is rapidly ignited through the spark because hydrogen has a fast flame front. This reduces problems with timing. Thus, the energy store found in hydrogen is captured through this system via the compression of the hydrogen gas within cylinder 92 and via the thorough mixing of hydrogen and air provided by adapter 38.

In addition, in the exhaust phase in a four cycle engine a second spark is generated from the plug to ensure complete combustion of by-products. This is enhanced by using an electrode made of platinum. Exhaust leaves cylinder 92 through exhaust port 98. Exhaust port 98 is closed during other phases by exhaust outlet valve 100.

In the exhaust stroke, because air valve 91 is closed, the substantial component of the gases in cylinder 92 is hydrogen. This is also accomplished due to the fact that in the system of this invention, the hydrogen is continuously flowing. When this spark interacts with what is predominately hydrogen, a plasma is formed. This plasma blocks back-flow of air from exhaust port 98. Prevention of back-flow of air prevents the hydrogen fuel system from backfiring, which is another common problem of hydrogen fuel systems. In effect, then, the plasma formed by the hydrogen acts as a virtual valve in that the plasma blocks the back-flow of air.

When the plasma of the hydrogen is generated as described above, UV radiation is generated which also facilitates consumption of combustion by-products. This increases the environmental friendliness of the system.

In addition to hydrogen the following fuels can be used in this system: natural gas; alcohol; gasohol; propane and buthane. A mixture of hydrogen and natural gas can also be used. As noted above, gasoline can be used in this system when solenoid 26 is used to stop the flow of hydrogen.

Further, it is contemplated that an adapter that is described herein for use with an internal combustion engine can also be used with a home heating system such as a home heating system that may currently use natural gas. The vortexing action created by the plug would enable appropriate power to be captured from the hydrogen fuel so as to effectively heat a home.

Thus, the apparatus and system of this invention provides an effective means to use alternative fuels. The use of alternative fuels can enhance the environment by reducing the use of conventional gasoline and the environmentally detrimental by-products caused by the consumption thereof.

Thus, although there have been described particular embodiments of the present invention of a new and useful "Fuel System For Internal Combustion System And Adapter For Use In Same", it is not intended that such references be construed as limitations upon the scope of this invention except as set forth in the following claims. Further, although there have been described certain dimensions used in the preferred embodiment, it is not intended that such dimensions be construed as limitations upon the scope of this invention except as set forth in the following claims.

United States Patent 6,119,651

September 19, 2000

Hydrogen Powered Vehicle, Internal Combustion Engine, and Spark Plug for use in same

Abstract --- The internal combustion engine fuel system described includes a structure for mixing the alternative fuel, preferably hydrogen, with oxygen in ambient air to stratify the fuel. The system includes an adapter, and the adapter includes a housing mounted between spark plug and cylinder of the internal combustion engine. A plug is placed within the housing. The plug has ridges or grooves on its outer surface that act as mixing structures. Thus, when hydrogen is introduced into the adapter housing it is mixed with ambient oxygen within the chamber as it flows over the plug. The mixing structures in the housing creates a vortexing action as the hydrogen flows over the plug and towards the cylinder of the engine. An electrode protrudes from the plug towards the cylinder. The electrode is preferably platinum and generates the necessary spark to create combustion of the hydrogen/air mixture adjacent to the cylinder to thereby power the cylinder in the engine. A platinum electrode is preferably used because it enhances a catalytic conversion of combustion by-products to more environmentally compatible products. The present invention also teaches a spark plug producing a hotter spark for a hydrogen fuel system. Also taught is a hydrogen powered vehicle with reduced emissions by producing a spark during the power stroke and the exhaust stroke. Methods for reducing exhaust pollution are also taught. Methods of reducing exhaust pollution by generating a plasma are taught as well.

Inventors: Anderson; Herman P. (Brentwood, TN)
Assignee: Herman P. Anderson Technologies, LLP

Current U.S. Class: 123/297 ; 123/635; 123/DIG.12
Current International Class: F02M 21/02 (20060101); F02B 43/00 (20060101); F02M 057/06 ()
Field of Search: 123/3,DIG.12,527,297,636,637,635,169EL 313/120

References Cited [Referenced By]

U.S. Patent Documents

2208030	July 1940	Holmes
3165099	January 1965	Vanderpoel
3926169	December 1975	Leshner et al.
3980061	September 1976	McAlister
4164912	August 1979	Beyler
4186712	February 1980	Fitzner et al.
4319552	March 1982	Sauer et al.
4343272	August 1982	Buck
4383198	May 1983	Hosking
4448160	May 1984	Vosper
4820957	April 1989	Zivkovich
5852993	December 1998	Anderson

Foreign Patent Documents

686943	Aug., 1930	FR

Description

BACKGROUND OF THE INVENTION

The present invention relates generally to a fuel system for internal combustion engines that enables the invention to utilize alternative fuels; and, more particularly to an improved fuel system for an internal combustion engine and an adapter and spark plug for use in same.

The problems of air pollution caused by automobiles, lawn mowers and other vehicles that have internal combustion engines powered by gasoline is well documented. This problem is particularly described in an article in the San Diego Union Tribune, Wednesday, Sep. 25, 1996 and in Popular Science, August 1996. Indeed, under the Clean Air Act, in the next few years cars that do not run on standard fossil fuels like gasoline will be required in several states. Currently, there are no vehicles available to meet this need.

Thus, it will be appreciated by those skilled in the art that there has been a push to seek other fuels for use in internal combustion engines. These fuels include methane, propane and gasohol.

Hydrogen has also been characterized as a promising fuel source. See Department of Energy publication entitled "Hydrogen Fuel", printed in 1978, U.S. Government Printing Office, U.S. Department of Energy, Office of Public Affairs, Washington, DC 20585. Particularly, the graph presented in FIG. 6, which is copied from this publication, shows the potential for hydrogen in BTUs/Lb. However, the mass energy density for hydrogen is low; and thus, hydrogen needs to be compressed before its energy potential can be realized. Hydrogen is a "dry fuel" that is more difficult to ignite than "wet fuels." It becomes more difficult to ignite when it is cooled. However, cooling is preferred in most systems to avoid premature ignition of the fuel. Other systems deliver the hydrogen under pressure to prevent ignition in the supply line, which is a form of pre-ignition. No prior art fuel system has been able to capture the energy of hydrogen.

Several fuel systems for use with hydrogen have been proposed. Examples are described in U.S. Pat. Nos. 4,167,919; 4,253,428; 4,016,836; 4,178,882; 5,222,993; 5,085,176; and 5,085,176. The problem with these prior art fuel systems is they cannot take advantage of the alternative fuels to generate sufficient power from the fuels to drive an internal combustion engine for use in a lawn mower, automobile or other vehicle. As noted above, this is a particular problem when hydrogen is used as a fuel. Additionally, the prior art systems do not address the problem of back fire, i.e. undesired explosion of the fuel in the cylinder during the exhaust stroke of the internal combustion engine.

Premature ignition or backfire is often a problem with hydrogen fueled systems. U.S. Pat. No. 4,383,198

(Hosking) teaches a fuel injection spark plug with a pre-cooling chamber to avoid premature ignition indentination of a gaseous fuel supplied to the spark plug. U.S. Pat. No. 4,383,198, Hosking, entitled "Fuel Injection Spark Plug", issued May 10, 1983, is hereby incorporated herein by reference. However, cooling hydrogen makes it more difficult to ignite. This is because hydrogen is a dry fuel whereas gas is generally considered a wet fuel. Thus to reduce fuel pollution one prefers to utilize a fuel such as hydrogen and ignite it with a hotter spark.

One patent discussing very high energy (VHE) ignition systems is U.S. Pat. No. 5,207,208 issued to Ward, entitled "Integrated Convertor High Power CD Ignition", issued May 4, 1993. U.S. Pat. No. 5,207,208 is hereby incorporated herein by reference. Pat. '208 teaches very high energy ignition systems, for instance on the order of 36K volts. It teaches voltage doubling utilizing pulsating sparks of peak current on the order of two amps. While 36K volts may at times be adequate, it is preferable to achieve a higher voltage in the ignition chamber when using hydrogen.

It is known in the art to use a direct ignition system, also referred to as a distributorless ignition system, to increase the spark plug voltage. Such direct ignition systems use a "waste spark" method of spark distribution. Each cylinder is paired with its opposing cylinder in the firing order (1-4, 2-3, on a four cylinder, 1-4, 2-5, 3-6 on a V6) so that one cylinder on compression fires simultaneously with its opposing cylinder on exhaust stroke. Since the cylinder on exhaust stroke requires very little of the available voltage to fire its plug, most of the voltage is used to fire the cylinder on the compression stroke.

U.S. Pat. No. 4,462,380 discusses a distributorless ignition system. U.S. Pat. No. 4,462,380 entitled "Enhanced Spark Energy Distributorless Ignition System" issued Jul. 31, 1984 by Asik is hereby incorporated herein by reference. Pat. '380 teaches a distributorless ignition system of an internal combustion engine which has a supplementary spark energy module to increase spark energy. Each high voltage terminal is connected to a single spark plug and each ignition coil primary is alternately energized and quickly de-energized, producing opposite polarity ignition voltages at each coil terminal. As a result, pairs of spark plugs are alternately fired, with each firing pair occurring in a compression or exhaust stroke and thereby providing the proper ignition to the engine. Asik asserts that firing a spark plug does not affect engine performance or emissions. As taught by the instant invention this is incorrect for hydrogen fuels when the exhaust is sparked at a sufficient voltage.

U.S. Pat. No. 5,146,882 (Brinkman) teaches use of specific wet fuel mixtures in conjunction with direct ignition systems (D.I.S.). It teaches that technical breakdown voltages for typical spark plugs are on the order of 15-20 kV and DIS typically yield between 30 kV-40 kV. U.S. Pat. No. 5,146,882, entitled "Method And Apparatus For Cold Starting A Spark Ignited Internal Combustion Engine Fueled With An Alcohol-Based Fuel Mixture", by Brinkman, et al., issued Sep. 15, 1992, is hereby incorporated herein by reference. The previously referenced prior art discusses alternate fuels and fuel systems but does not teach a method of achieving a sufficient voltage with a hydrogen system.

What is needed, then, is a fuel system for an internal combustion engine that provides a way to generate sufficient power from the alternative fuels and that reduces backfire problems. Such a system is lacking in the prior art. Also lacking in prior art is a system utilizing alternate fuels with reduced pollution capabilities. Also lacking is a way of generating sufficient voltage spark to ignite the alternate fuel.

The following U.S. Patents discuss spark plugs designed for gas internal combustion engines to reduce radio frequency noise and interference. U.S. Pat. No. 4,224,554, entitled "Spark Plug Having A Low Noise Lever" issued to Nishio et al., Sep. 23, 1980 is hereby incorporated herein by reference. U.S. Pat. No. 4,029,990 entitled "Spark Plug Construction" issued to Nagy et al., Jun. 14, 1977 is hereby incorporated herein by reference.

SUMMARY OF THE INVENTION

It is an object of this invention to provide an improved fuel system for use with alternative fuels.

It is a further object of this invention to provide an improved fuel system for use particularly with hydrogen, and particularly to enhance the mass energy density of hydrogen.

It is an object to provide a hotter sparking plug to ignite the fuel.

It is still a further object of this invention to provide an adapter for use in a fuel system for an internal combustion engine that enhances the power obtained from the alternative fuels.

It is still a further object of this invention to provide an improved fuel system for an internal combustion engine that does not present the environmental hazards of a standard gasoline engine by providing reduced or zero emission of environmentally harmful gases.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that can use an alternative fuel and can use gasoline to facilitate the transition from gasoline to alternative fuels.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that reduces backfire problems.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that reduces backfire problems by cooling the electrode in the spark plug with the hydrogen gas fuel.

It is an object to reduce backfire by providing hydrogen at a sufficient pressure to reduce backfire and igniting it with a hotter spark (i.e. a higher voltage).

It is still a further object of this invention to provide an adapter for use in a fuel system for an internal combustion engine that enhances the power obtained from the alternative fuels.

It is a further object to generate a plasma field in a cylinder to burn residual exhaust and reduce emissions.

It is yet a further object of this invention to provide a direct injection fuel system wherein fuel is injected directly into the cylinder.

It is an object to provide a "2-fire" direct injection system, wherein the fuel is ignited during the power stroke and the exhaust stroke.

It is yet a further object of this invention to provide a way to convert the standard induction fuel system of a lawnmower to a direct injection fuel system wherein fuel is injected directly into the cylinder.

It is yet a further object of this invention to provide a way to convert the indirect injection fuel system of an automobile to a direct injection fuel system wherein fuel is injected directly into the cylinder.

The internal combustion engine fuel system of this invention includes a structure for mixing the alternative fuel, preferably hydrogen, with oxygen in ambient air to stratify the fuel. The term "alternative fuels" is known in the art and is meant to refer to fuels other than conventional gasoline.

The invention in one embodiment comprises an adapter, the adapter including a housing mounted between spark plug and cylinder of the internal combustion engine. A plug is placed within the housing. The plug has ridges or grooves on its outer surface that act as mixing structures. Thus, when hydrogen is introduced into the adapter housing it is mixed with ambient oxygen within the chamber as it flows over the plug. The mixing structures on the plug create a vortexing action as the hydrogen flows over the plug and towards the cylinder of the engine.

An electrode protrudes from the plug towards the cylinder. The electrode is preferably platinum and generates the necessary spark to create combustion of the hydrogen/air mixture adjacent to the cylinder to thereby power the cylinder in the engine. A platinum electrode is preferably used because it enhances a catalytic conversion of combustion by-products to more environmentally compatible products.

Thus, the combination of the compression action generated by the conventional piston and push rod within the cylinder of the engine in conjunction with the hydrogen/air mixing structure that vortexes the hydrogen/air create compressed hydrogen that is mixed with a combustion facilitator, i.e. oxygen in the ambient air to enable the system of this invention to capture the abundant energy found in hydrogen fuel.

The enhanced mixing capabilities of this invention is also adaptable for use with other alternative fuels such as natural gas, alcohol, propane and the like.

One embodiment of the present invention is for a hydrogen powered vehicle including an internal combustion engine operably connected to power the vehicle and adapted to connect to a hydrogen supply line. The internal combustion engine comprises a cylinder operably connected to the supply line; a piston operably disposed in the cylinder; and a spark plug having an electrode assembly in operable sparking communication with the cylinder, wherein the spark plug is capable of producing a spark of at least 45 kV.

Preferably the spark plug comprises a housing having a sparking end and a hollow chamber containing the electrode assembly. The electrode assembly comprises an electrode having a connector end and a sparking end, wherein the sparking end is positioned adjacent to the housing sparking end. The electrode assembly preferably comprises a coil electrically connected to the electrode connector.

Preferably the housing comprises a fuel inlet port in fluid communication with the supply line, and an insulator surrounding the electrode assembly, wherein the insulator is positioned in the hollow chamber. Preferably the insulator comprises an outer surface defining a first fuel channel having a first respective fuel exit port, wherein the first fuel channel is in fluid communication with the fuel inlet port and the fuel exit port exits into the cylinder.

The present invention also includes a method of operating an internal combustion engine comprising supplying hydrogen fuel to the engine and creating a spark in the cylinder during the power stroke and the exhaust stroke, wherein the spark created during the power stroke is at least 45 kV. In one embodiment the spark is at least 70 kV. In certain embodiments the method comprises generating plasma during the exhaust stroke. The exhaust fuel converts to an exhaust emission and coalesces with the plasma, thereby reducing fuel pollution.

Accordingly, it is an object in the invention to provide a spark plug producing a sufficiently hot spark to ignite hydrogen fuel, including cooled hydrogen fuel.

Another object of the present invention is to provide a device for reducing fuel exhaust emissions.

Another object of the present invention is to provide a cleaner burning engine.

Another object is to provide a spark plug for reducing emissions.

Another object of the present invention is to provide a device functional with dry fuels. Use of the term hydrogen, as discussed herein, is intended to encompass such dry fuels. However the term as used excludes such wet fuels as gas and alcohol.

Another object of the present invention is to provide an integrated fuel system adapted to provide these and objectives.

Other objects and advantages of the present invention will be apparent to those of skill in the art by reference to the teachings disclosed herein, including the attached drawings, detailed description of exemplary embodiments, and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a plan view of the fuel system of this invention with a lawn mower shown in phantom.

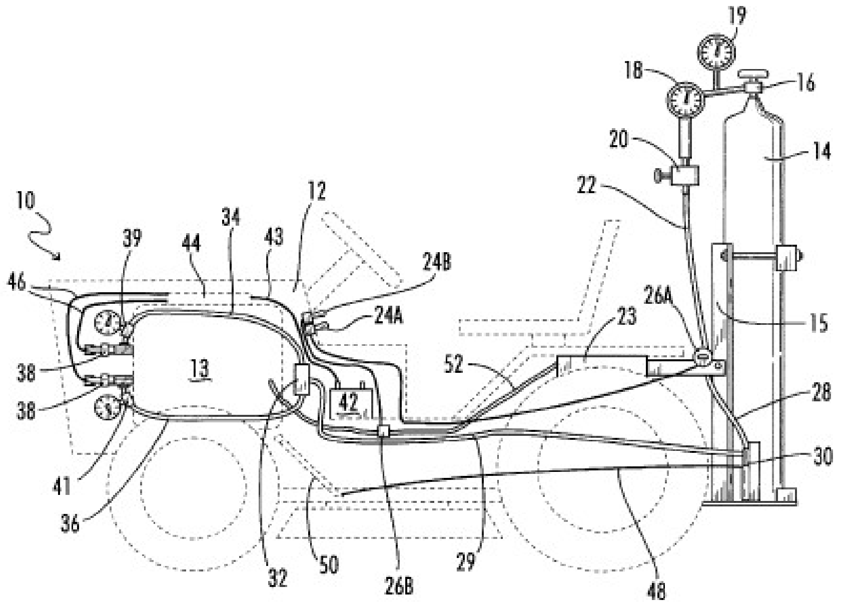
19
18
16
20
14
22
12
10
34
43
39
44
24B
46
26A
24A
23
15
13
52
38
42
28
38
38
30
41
36
32
50
26B
29
48

FIG. 1

FIG. 2 is a perspective view of the adapter of the fuel system of this invention.

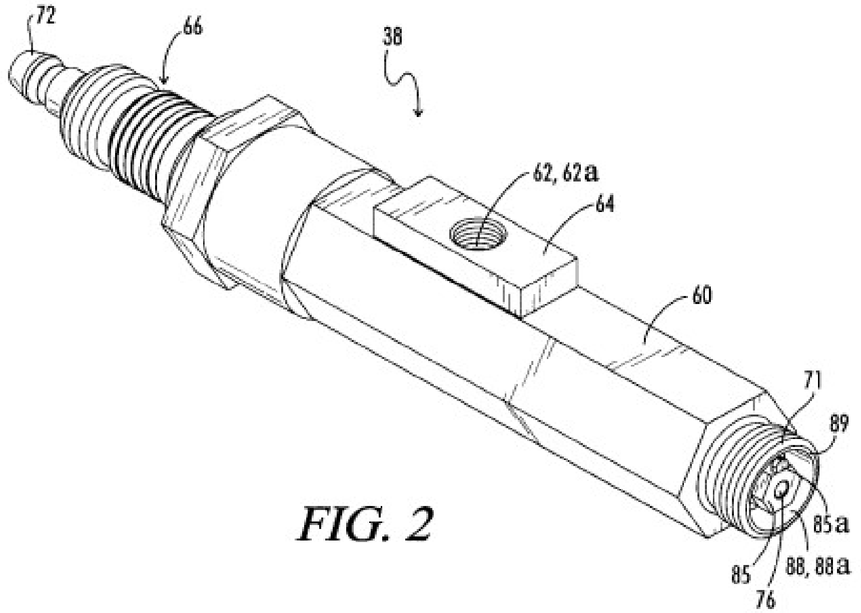
72
66
38
62, 62a
64
60
71
89
85a
85
88, 88a
76

FIG. 2

FIG. 3 is an exploded perspective view of the adapter of the fuel system of this invention.

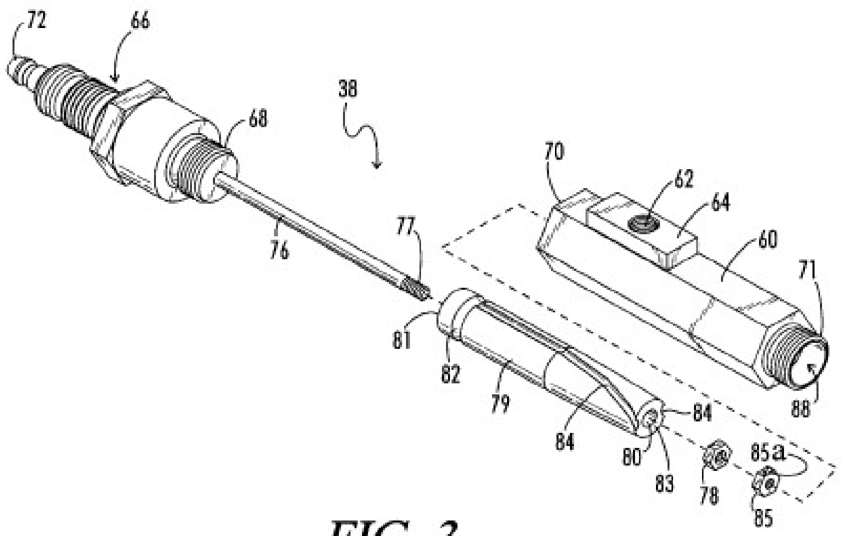
72
66
38
68
70
62
64
60
77
76
71
81
82
79
84
88
84
80
83
78
85a
85

FIG. 3

FIG. 4 is a cross-sectional view of the adapter of the fuel system of this invention.

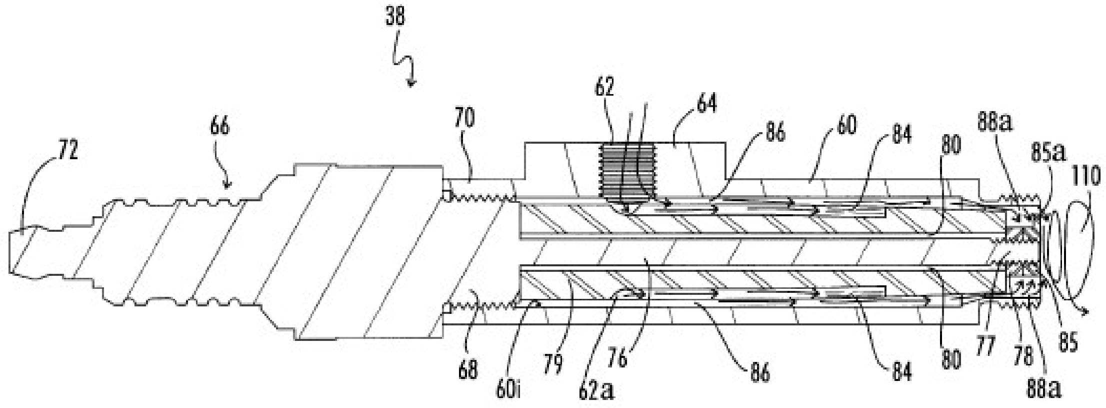
38
62
64
86
60
84
80
88a
85a
110
72
66
70
68
60i
79
62a
76
86
84
80
77
78
88a
85

FIG. 4

FIGS. 5A-5D are a schematic of the 4 cycle combustion in an internal combustion engine including the adapter of this invention.

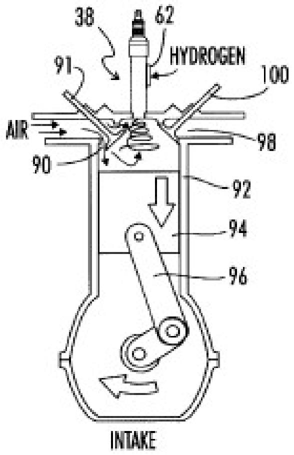
38
62
HYDROGEN
91
100
AIR
98
90
92
94
96
INTAKE

FIG. 5A

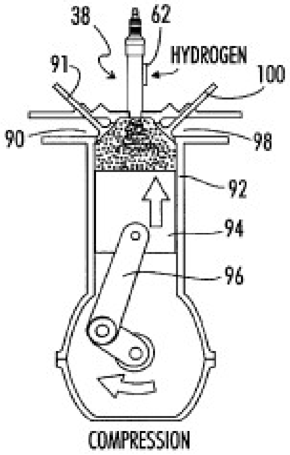
38
62
HYDROGEN
91
100
90
98
92
94
96
COMPRESSION

FIG. 5B

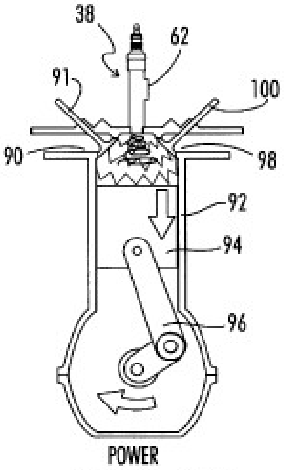
38
62
91
100
90
98
92
94
96
POWER

FIG. 5C

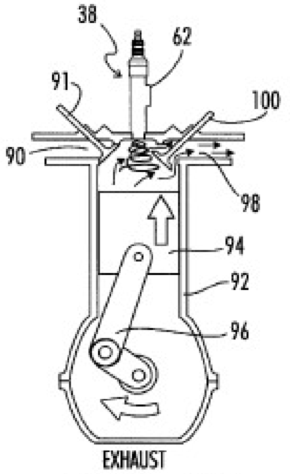
38
62
91
100
90
98
94
92
96
EXHAUST

FIG. 5D

FIG. 6 is a graph which shows the energy potential for hydrogen in BTUs/Lb.

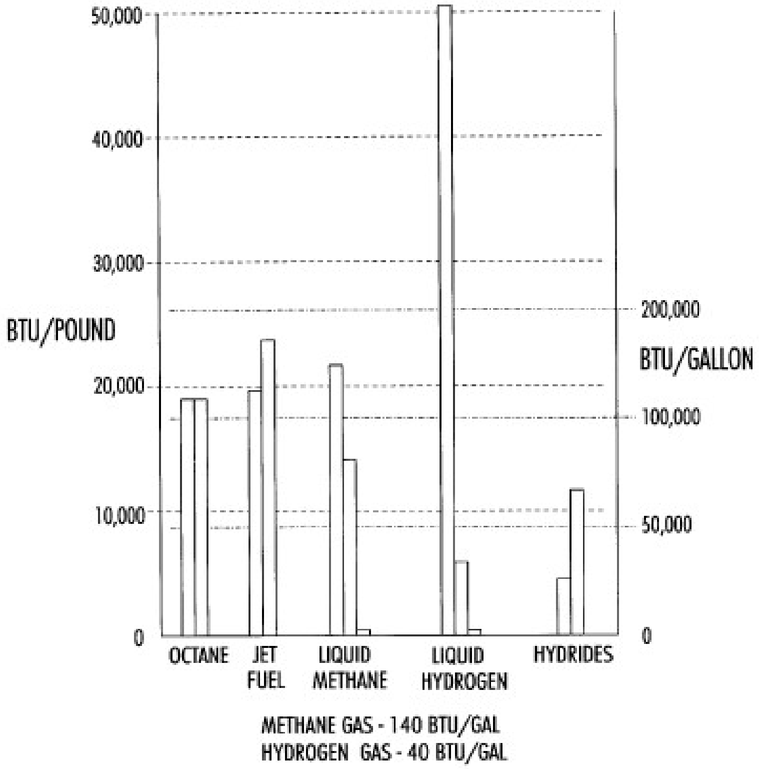

FIG. 6

FIG. 7 shows a vehicle of the present invention utilizing hydrogen fuel.

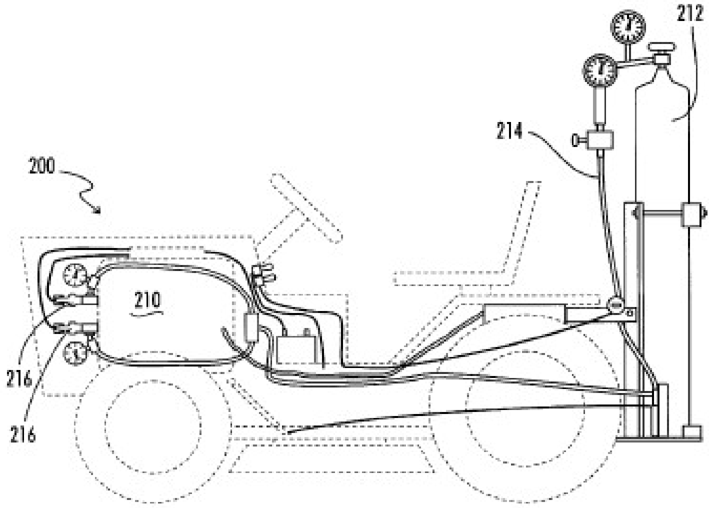

FIG. 7

FIG. 8 depicts an intake stroke of a cylinder connected to a hydrogen source.

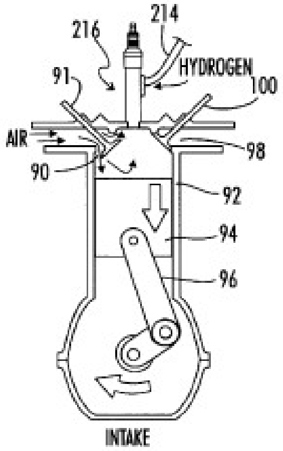
216
214
HYDROGEN
91
100
AIR
98
90
92
94
96
INTAKE

FIG. 8

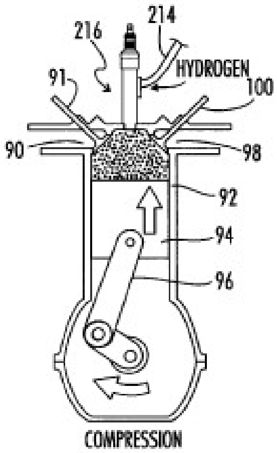
216
214
HYDROGEN
91
100
90
98
92
94
96
COMPRESSION

FIG. 9

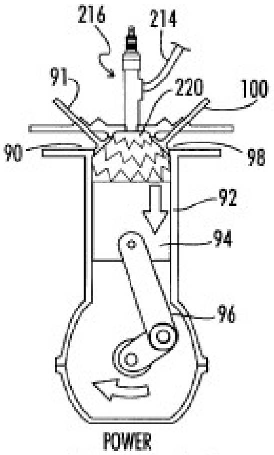
216
214
91
220
100
90
98
92
94
96
POWER

FIG. 10

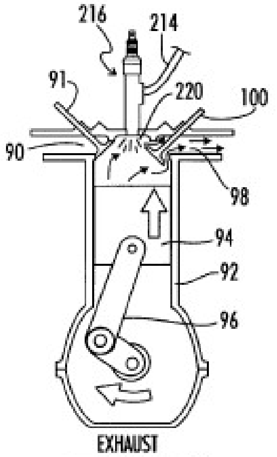
216
214
91
220
100
90
98
94
92
96
EXHAUST

FIG. 11

FIG. 9 depicts a compression stroke of a cylinder connected to a hydrogen source.
FIG. 10 depicts a power stroke of a cylinder connected to a hydrogen source. The spark plug shown in FIG. 10 is depicted sparking.
FIG. 11 depicts a cylinder in an exhaust stroke connected to a hydrogen source, wherein the spark plug is sparking during the exhaust stroke.

FIG. 12 shows an elevated side view of a spark plug of the present invention.

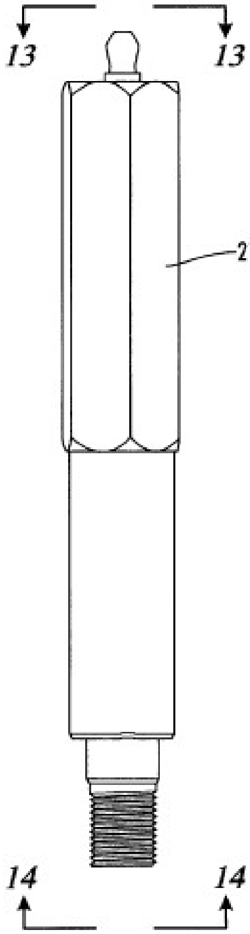

FIG. 12

FIG. 13 shows a plan view in the direction of view line 13--13 of the spark plug shown in FIG. 12.

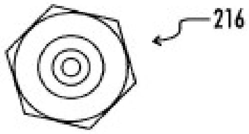

FIG. 13

FIG. 14 depicts a plan view in the direction of view line 14--14 of the spark plug shown in FIG. 12.

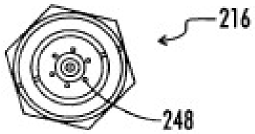

FIG. 14

FIG. 15 depicts an elevated side section view of the spark plug shown in FIG. 12.

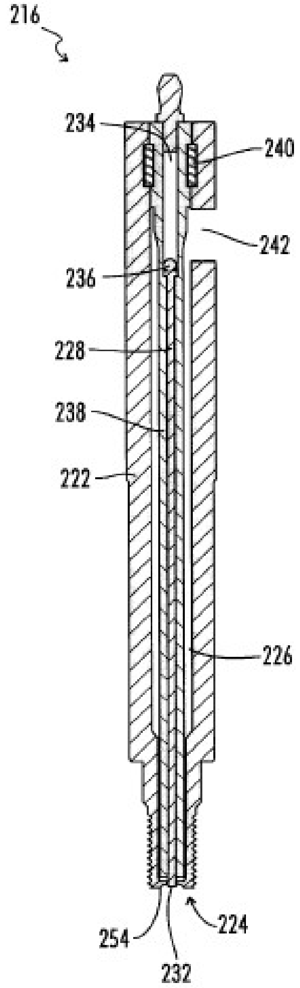
216
234
240
242
236
228
238
222
226
254
224
232

FIG. 15

FIG. 16 shows a electrode assembly which is disposed in the hollow chamber depicted in FIG. 15.

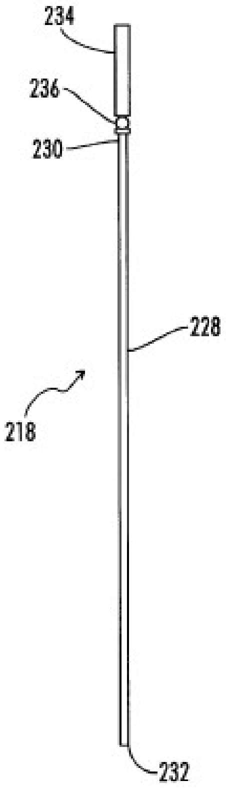

FIG. 16

FIG. 17 shows a section view of the spark plug housing clearly showing the hollow chamber and a torridal ground electrode.

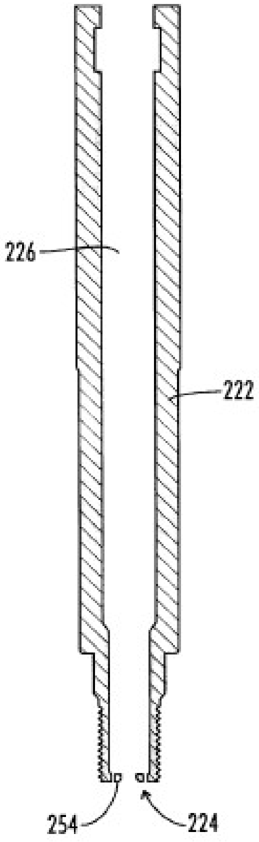

FIG. 17

FIG. 18 depicts an elevated side view of the spark plug housing and a fuel inlet port.

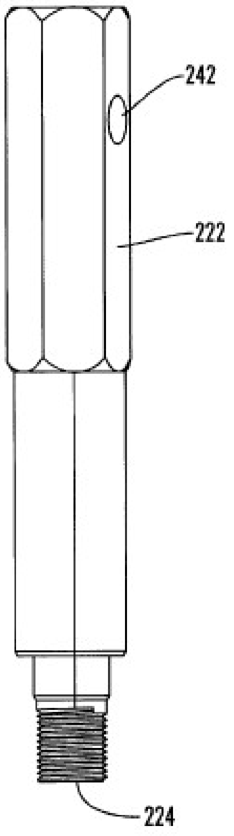

FIG. 18

FIG. 19 depicts an insulator comprising fuel channels and fuel exit ports.

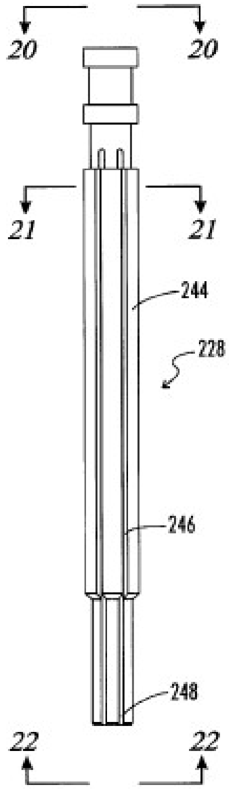
20
20
21
21
244
228
246
248
22
22

FIG. 19

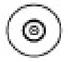

FIG. 20

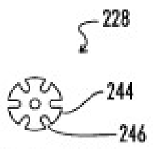

FIG. 21

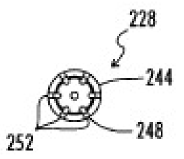
228
244
252
248

FIG. 22

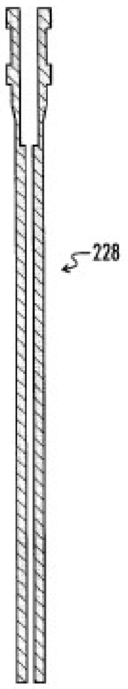

FIG. 23

FIG. 20 is a top plan view of the insulator shown in FIG. 19 viewed in the direction of view line 20--20.
FIG. 21 is a cross section view of the insulator shown in FIG. 19 along section line 21--21.
FIG. 22 is a cross section view of the insulator shown in FIG. 19 viewed in the direction of view line 22--22.
FIG. 23 is an elevated section view of the insulator shown in FIG. 19.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring now to the drawings, wherein like reference numerals refer to like parts throughout, the fuel system is referred to generally at 10. The fuel system of this invention is contemplated for use with an internal combustion power system, such as an internal combustion engine or such as a home heating system. Fuel system 10 is portrayed in this case as connected to a standard lawn mower. The lawn mower is shown in phantom at 12 in FIG. 1. The lawn mower is a conventional lawn mower as can be purchased from a company such as Murray of Ohio. The internal combustion engine in the lawn mower is a standard 4-cycle engine with two cylinders and develops 18 horsepower. A suitable example can be purchased from Briggs & Stratton, Milwaukee, Wis. 53201.

Referring particularly to FIG. 1, system 10 includes fuel tank 14 that is mounted to lawn mower 12 via mounting assembly 15. In the preferred embodiment, fuel tank 14 includes hydrogen gas. Accordingly, for the purposes of the foregoing detailed description, hydrogen gas will be described.

A line 22 begins at fuel tank 14 and proceeds to back-flow valve 16. Back-flow valve 16 is a standard valve and is available from AirCo Gases Company, 575 Mountain Avenue, Murray Hill, N.J. 07974, part no. WCS CV-4M. PSI tank gauge 19 is inserted in line 22 between back-flow valve 16 and flow pressure gauge 18. PSI tank gauge 19 is available from AirCo Gases Company also, part no. WCS CV5M, and serves as an indicator of fuel remaining in the tank 14. Flow pressure gauge 18 is used to monitor the flow of the hydrogen in pressure units. The flow is started using valve 16. The optimal flow pressure is 25 pounds of flow pressure from the tank 14 to the cylinder of the engine.

The flow of hydrogen is regulated via valve 20 that is part of gauge 18. Thus, when valve 20 is opened, hydrogen proceeds through line section 22 to hydrogen flow solenoid 26A. The user can access the hydrogen fuel by flipping hydrogen flow switch 24A into the "on" mode so that hydrogen gas in line 22 can proceed through hydrogen flow solenoid 26A and into line section 28.

Throttle 30 regulates the flow of hydrogen from line section 28 into line section 29 and ultimately, to engine 13. Throttle 30 is connected to throttle pedal 50 via includes throttle cable 48. Throttle pedal 50 also operates the butterfly valves in the carburetor in engine 13 in a conventional manner.

As an option, the throttle can be controlled electronically by a digital

fuel controller product. Such a product is commercially available from Autotronic Controls Corp., 1490 Henry Brennan Drive, El Paso, Tex. 79936.

Line portion 29 proceeds through splitter 32 and is split into upper line 34 and lower line 36. Lines 34 and 36 proceed to adapters 38. Additional back-flow prevention valves 39 and 41 are mounted between lines 34 and 36 and adapters 38. Braided Teflon.RTM. is the preferred material for the fuel lines of the system.

Battery 42 is connected via battery cable 43 to magneto 44. Magneto 44 then controls the current through spark plug lines 46 to adapters 38 in a conventional manner. Thus, the regulation of the spark that ignites the fuel in the cylinder of engine 13 is conventional. Additionally, the intake of air into the engine 13 is accomplished through the standard carburetor system of engine 13. In a car, the sparking of the spark plugs can be controlled by the standard distributor/battery ignition system.

A second solenoid 26B controls the flow of gasoline from tank 23 through gasoline line 52. Solenoid 26B is controlled by switch 24B. Throttle 30 is connected to gasoline line 52 and can also be used to control the flow of gasoline to engine 13. Thus, this invention can be used with both an alternative fuel and gasoline.

Referring now to FIGS. 2, 3 and 4, the adapter 38 of this invention is shown in more detail. Adapter 38 includes housing 60. A threaded port 62 is formed into housing 60 through port block 64. Hydrogen lines are threadably connected to the adapter 38 via port 62. Port 62 is also referred to as an entrance port. The entrance port has an entrance area 62a.

Adapter 38 also includes spark plug 66 which is effectively a standard spark plug except for the modification to the electrode as described below. Spark plug 66 includes a threaded end 68 which is threadably received in the threaded upper end 70 of housing 60. Spark plug wires 46 (as seen in FIG. 1) are connected to connector end 72 of spark plug base 66.

Electrode 76 protrudes from spark plug 66 and is of a length sufficient so that the threaded tip 77 (best seen in FIGS. 3 and 4) of the electrode 76 is proximate to the cylinder of the internal combustion engine when the adapter 38 is mounted. In one embodiment the electrode 76 is coterminus with an exit port 88 of the adapter 38. This most clearly is shown in FIG. 4. Electrode 76 is preferably platinum to enhance catalysis of combustion by-products, as will be more fully discussed below.

Adapter 38 also includes plug 79. Plug 79 includes an axial internal void space 80. Thus, plug 79 fits over electrode 76 via internal void space 80. Plug 79 preferably comprises a ceramic material that acts as a good insulator. A suitable source for ceramic plug 79 is AM/PRO Machinery, Inc. 134 Church Street, Plain City, Ohio 43064-1321.

An important aspect of this invention is found on the external surface of plug 79. An annular groove 82 is formed on the surface of plug 79 near its distal end 81 so as to be proximal to the port 62 where hydrogen enters housing 60. Two stratifying grooves 84 proceed along the surface of plug 79 until they reach its distal end 83. It is these mixing structures, i.e. annual groove 82 and stratifying grooves 84, that facilitate the improved power generation from alternative fuels as provided by this invention. It will be apparent to those skilled in the analogous arts that mixing structure may be formed in the inner surface of the adapter housing 60.

Alternatively a combination of vortex producing structures is integral with both the plug 79 and adapter housing 60. Also any conventional vortex producing means should be sufficient groves, per se, are not required.

On threaded tip 77 of electrode 76, there are mounted a spacer nut 78 and a spark gap nut 85. As best seen in FIG. 2, the spark gap nut 85 includes a protrusion 85a that provides the spark gap between electrode 76 and periphery 89 of the exit port 88 of housing 60. The exit port 88 includes an exit area 88a. Preferably, the exit area 88a is no larger than the entrance area 62a, though this is not critical. Preferably, the spark gap nut 85 is brass and the preferred spark gap distance is 25/1000 of an inch. Of course, the electrode may be coterminus with the plug distal end 83.

As best seen in FIG. 4, where arrows represent gas flow, hydrogen gas flows into chamber 60 via port 62. The hydrogen gas cools electrode 76, which facilitates combustion. Stratifying grooves 84, preferably there are two, act to mix the hydrogen that is flowing into housing 60 with ambient air therein. Though more or less vortex creating grooves (stratifying grooves) may be desired. Typically this will depend on the relative dimensions of the grooves and diameter of the plug as well as the pressure or velocity of the mixture in the housing. The mixing occurs within void space 86 of housing 60 and creates a vortexing or tornado action. Plug 79 tapers as it proceeds from its distal end 81 to its proximal end 83. Thus, proximal end 83 is smaller in diameter than distal end 81. This provides additional space within void space 86 for the mixing of the hydrogen and air. The stratified hydrogen/air are twisting in a vortex fashion when leaving housing 60 via exit port 88. The vortexing action produced by grooves 84 on the plug 79 creates a "tornado" 110 of hydrogen/air with a hollow center that generates a vacuum. Similar effect would result were t he grooves to be placed on the housing. The vacuum propels the hydrogen/air mixture into the cylinder of the engine for combustion. Such mixing cannot be accomplished in a conventional induction system.

Referring now to FIG. 5, a schematic of the system of this invention is portrayed, with arrows representing gas flow. Hydrogen flows in through port 62 through adapter 38 and into cylinder 92. Air enters cylinder 92 through air intake 90 and air intake valve 91. The stratified hydrogen/air mixture is compressed in the compression stroke of the engine via piston 94 and push rod 96, which are standard structures in an internal combustion engine. In the power stroke shown in FIG. 5-C, a spark is generated in the compressed stratified hydrogen/air mixture ignites and drives piston 94 and push rod 96 to power the engine.

The mass energy density of hydrogen is enhanced through the increased pressure on the hydrogen exerted by piston 94 on the hydrogen/air mixture within the cylinder 92 in the uptake stroke in an internal combustion engine. The pressurized and well-mixed hydrogen/air mixture is rapidly ignited through the spark because hydrogen has a fast flame front. This reduces problems with timing. Thus, the energy store found in hydrogen is captured through this system via the compression of the hydrogen gas within cylinder 92 and via the thorough mixing of hydrogen and air provided by adapter 38.

In addition, in the exhaust phase in a four cycle engine a second spark is generated from the plug to ensure complete combustion of by-products. This is enhanced by using an electrode made of platinum. Exhaust leaves cylinder 92 through exhaust port 98. Exhaust port 98 is closed during other phases by exhaust outlet valve 100.

In the exhaust stroke, because air valve 91 is closed, the substantial component of the gases in cylinder 92 is hydrogen. This is also accomplished due to the fact that in the system of this invention, the hydrogen is continuously flowing. When this spark interacts with what is predominately hydrogen, a plasma is formed. This plasma blocks back-flow of air from exhaust port 98. Prevention of back-flow of air prevents the hydrogen fuel system from backfiring, which is another common problem of hydrogen fuel systems. In effect, then, the plasma formed by the hydrogen acts as a virtual valve in that the plasma blocks the back-flow of air.

When the plasma of the hydrogen is generated as described above, UV radiation is generated which also facilitates consumption of combustion by-products. This increases the environmental friendliness of the system.

In addition to hydrogen the following fuels can be used in this system: natural gas; alcohol; gasohol; propane and buthane. A mixture of hydrogen and natural gas can also be used. As noted above, gasoline can be used in this system when solenoid 26 is used to stop the flow of hydrogen.

Further, it is contemplated that an adapter that is described herein for use with an internal combustion engine can also be used with a home heating system such as a home heating system that may currently use natural gas. The vortexing action created by the plug would enable appropriate power to be captured from the hydrogen fuel so as to effectively heat a home.

Another embodiment of the invention is a hydrogen powered vehicle 200 shown in FIG. 7. The hydrogen powered vehicle 200 includes an internal combustion engine 210 operably connected to power the vehicle 200 and adapted to connect to a hydrogen supply 212 through a supply line 214.

Referring to FIG. 8, the internal combustion engine 210 comprises a cylinder 92 operably connected to the supply line 214. A piston 94 is operably disposed in the cylinder 92. A spark plug 216 having an electrode assembly 218 (not shown in FIG. 8) is in operable sparking communication with the cylinder 92, wherein the spark plug 216 is capable of producing a spark 220 (shown in FIG. 10) of at least 45 kilovolts. FIGS. 8-11 are substantially similar to FIGS. 5A-5D depicting schematic representations of a 4-cycle piston-cylinder engine. Other piston and cylinder embodiments will be apparent to those with skill in the art. FIG. 8 represents an intake stroke. FIG. 9 depicts a compression stroke. FIG. 10 depicts a power stroke, in which the spark plug 216 sparks. FIG. 11 depicts an exhaust stroke, where the spark plug 216 is sparking. Refer to FIGS. 5A-5D discussed earlier for further description of similar components of FIGS. 8-11.

FIG. 12 shows one embodiment of the spark plug 216 in an elevated side view. FIG. 13 shows a plan view of the spark plug 216 along line 13--13 shown in FIG. 12. FIG. 14 shows a plan view of the spark plug 216 in the direction of line 14--14 shown in FIG. 12.

FIG. 15 depicts a section view of the spark plug 216. FIG. 15, section view, depicts a representative spark plug 216. The spark plug shown in FIG. 15 comprises a housing 222 having a sparking end 224 and a hollow chamber 226 containing the electrode assembly 218.

FIG. 16 depicts the electrode assembly 218 removed from the hollow chamber 226. The electrode assembly 218 depicted in FIG. 16 comprises an electrode 228. The electrode 228 has a connector end 230 and a sparking end 232. As shown in FIG. 15 the sparking end 232 of the electrode 228 is positioned adjacent to the housing sparking end 224. The electrode assembly further comprises a coil 234 electrically connected to the electrode connector end 230.

Referring to FIGS. 10 and 11, wherein the piston 94 moves through a power stroke in FIG. 10 and an exhaust stroke in FIG. 11, the spark plug 216 is shown to fire during the power stroke and the exhaust stroke.

The electrode assembly 218 shown in FIG. 16 comprises a steel ball 236 positioned between the coil 234 and the electrode connector end 230. Other operable ball-coil-electrode arrangements will be apparent to those of skill in the art.

The spark plug 216 shown in FIG. 15 also comprises an insulator 238 surrounding the electrode assembly 218 and positioned in the hollow chamber 226. A ceramic pore 240 is shown stabilizing the insulator 238 in the hollow chamber 226. Other methods of locating the insulator in the hollow chamber will be apparent to those of skill in the art.

An elevated section view of the housing 222 is shown in FIG. 17. An elevated side view of the housing 222 is shown in FIG. 18. In the exemplary embodiment of the housing 222 shown in FIG. 18, the housing 222 comprises a fuel inlet port 242 in fluid communication with the supply line 214 (not shown). FIG. 19 depicts an exemplary embodiment of the insulator 238. The insulator 238, shown in FIG. 19, comprises an outer surface 244 defining a first fuel channel 246 having a first respective fuel exit port 248. The first fuel channel 246 is in fluid communication with the fuel inlet port 242. The fuel exit port 248 exits into the cylinder 292.

FIG. 20 shows a plan view of the insulator 228 shown in FIG. 19 along line 20--20. FIG. 21 shows a section view of the insulator 228 shown in FIG. 19 viewed along section line 21--21. FIG. 22 depicts a bottom plan view of the insulator 228 shown in FIG. 19 along line 22--22. FIG. 23 shows a section view of the insulator 228 cut through section line 23--23 shown in FIG. 19.

In the embodiment of the insulator 228 shown in FIGS. 19-22, the insulator outer surface 244 defines a plurality of fuel channels 250, including the first fuel channel 246. The fuel channels 250 include a plurality of respective fuel exit ports 252, including the first fuel exit port 248. This is shown clearly in FIG. 22. Preferably, the plurality of fuel channel exit ports 252 are spaced equally, or equi-distant, about the electrode 228. This is shown well in FIGS. 21 and 22.

Referring to FIGS. 15 and 17, one embodiment of the housing 222 sparking and 224 comprises a torridal ground electrode 254 positioned about the electrode sparking end 232 (shown in FIG. 15). Preferably the housing sparking end 224 and the electrode sparking end 232 are co-terminous. In one embodiment the first respective fuel exit port 248 is co-terminous with the housing sparking end 224 and the electrode sparking end 232.

More generally the present invention includes an internal combustion engine 210 adapted to operate on hydrogen fuel supplied through a supply line 214. Typically the engine comprises a first cylinder 92 in fluid communication with the supply line 214; a first piston 94 disposed in the first cylinder 92 to move through a power stroke and an exhaust stroke (See FIGS. 10 and 11) in the first cylinder 92. The engine 210 also comprises a spark plug 216 having a housing 222 including a sparking end 224 and an electrode assembly 218 located in the housing 222, wherein the sparking end 224 is in operable communication with the cylinder 92 and capable of producing a spark 220 of at least 45 kV. Preferably, the spark plug 216 fires during the power stroke and the exhaust stroke. In one embodiment the housing sparking end 224 comprises a torridal ground electrode 254. The electrode assembly 218 comprises an electrode 228 having a sparking end 230 capable of sparking toward the torridal ground electrode 254. Preferably the electrode assembly 218 comprises a coil 234 in electrical communication with the electrode 228.

In one embodiment the electrode assembly 218 comprises an electrode 228 having a sparking end 230 and a coil 234 electrically connected to the electrode 228 opposite the electrode sparking end 230. This is the embodiment shown in FIG. 15. Typically the spark plug 216 comprises an insulator 228 surrounding the electrode assembly 218. Preferably the insulator 228 includes a fuel channel 246 having a fuel exit port 248 near the electrode sparking end 232.

It will be apparent to those with skill in the art that the present invention also comprises a method of operating an internal combustion engine 210 comprising the steps of supplying hydrogen fuel to the engine 210; moving a piston 94 in a cylinder 92 through a power stroke and an exhaust stroke; creating a spark 220 in the cylinder 92 during the power stroke and the exhaust stroke, wherein the spark created during the power stroke is at least 45 kV; and supplying the fuel to a location sufficiently near the spark to ignite the fuel. This is shown well in FIGS. 10, 15 and 19.

The present invention also includes a method wherein the step of creating a spark 220 comprises passing an electric current (not shown) through an electrode assembly 218 including a coil 234 electrically connected to an electrode 218 having a sparking end 232; insulating the electrode assembly 218; and arcing the electric current from the electrode sparking end 232 to a ground electrode 254. The ground electrode 254 as shown in FIGS. 15 and 17 a toroidal ground electrode 254.

In certain embodiments the step of creating a spark 220 includes creating a spark 220 having at least 70 kV. Creating a spark of sufficient voltage will generally reduce the exhaust pollution. Typically this happens by generating a plasma or plasma field during the exhaust stroke. Thus, one method of the present invention further comprises the steps of generating plasma during the exhaust stroke; and allowing exhaust fuel converted to an exhaust state to coalesce with the plasma. This reduces exhaust pollution.

Thus, the apparatus and system of this invention provides an effective means to use alternative fuels. The use of alternative fuels can enhance

the environment by reducing the use of conventional gasoline and the environmentally detrimental by-products caused by the consumption thereof.

Thus, although there have been described particular embodiments of the present invention of a new and useful "Hydrogen Powered Vehicle, Internal Combustion Engine, And Spark Plug For Use In Same", it is not intended that such references be construed as limitations upon the scope of this invention except as set forth in the following claims. Further, although there have been described certain dimensions used in the preferred embodiment, it is not intended that such dimensions be construed as limitations upon the scope of this invention except as set forth in the following claims.

Fuel Cell Plastic Parts Specializing in custom Fuel Cell molded or machined parts www.littlestarplastics.com
6-Core 4-Way SuperServers Energy-efficient Supermicro Servers Featuring Intel® Xeon® Processors www.Supermi
CFP® Certification Online Your Path to CFP® Starts Today. 3 Free Lessons. $1410 Off. Act Now! UGA.FinancialPla
Ads by Google

Your Support Maintains this Service -- and Your Survival ...

The Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Everything @ rexresearch.com, plus the Bonus Files CD !

ORDER PAGE

<< $13, Postpaid Anywhere >>

Rex Research, POB 19250, Jean, NV 89019 USA

(No Model.)

R. J. WALLACE.

BUSINESS COLLEGE CABINET.

No. 585,299. Patented June 29, 1897.

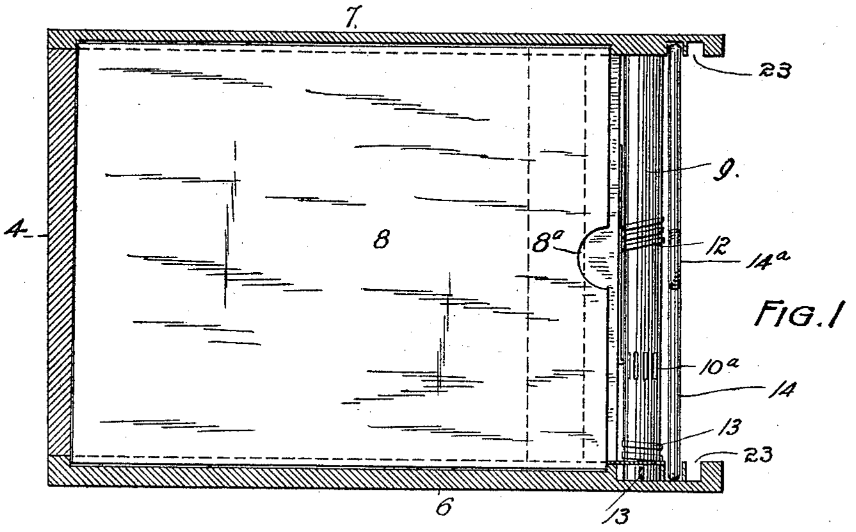
7
23
9
4
8
8a
12
14a
10a
14
13
23
6
13

FIG. 1

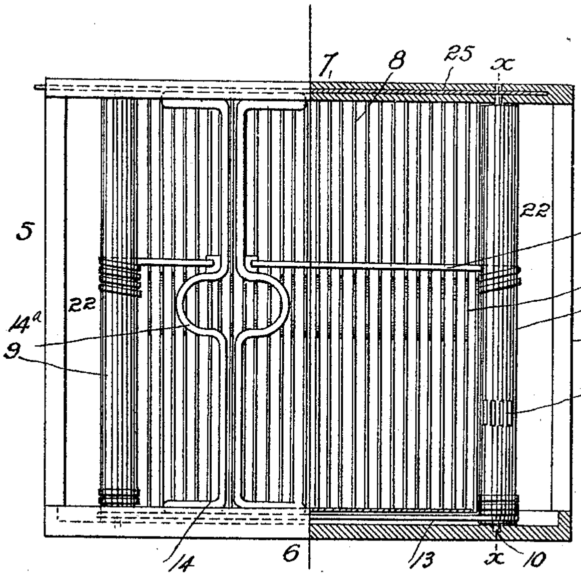
7
8
25
x
5
22
12
8a
14a
22
9
9
5
9
10a
14
6
13
x
10

FIG. 2

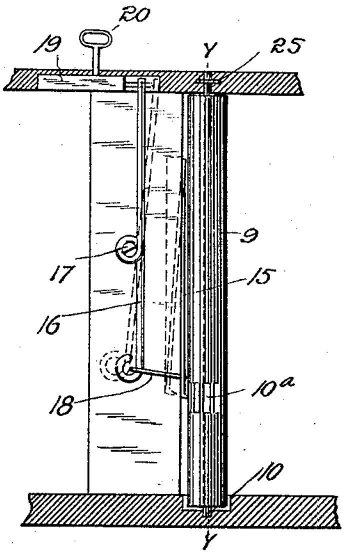
20
19
Y
25
9
17
15
16
18
10a
10
Y

FIG. 3

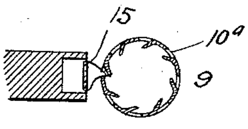

FIG. 4

Witnesses
G. J. Gallaudet
Edith Himsworth.

Inventor
R. J. Wallace.
By his Attorney

THE NORRIS PETERS CO., PHOTO-LITHO., WASHINGTON, D. C.

UNITED STATES PATENT OFFICE.

ROBERT J. WALLACE, OF DENVER, COLORADO.

BUSINESS-COLLEGE CABINET.

SPECIFICATION forming part of Letters Patent No. 585,299, dated June 29, 1897.

Application filed April 7, 1897. Serial No. 631,189. (No model.)

To all whom it may concern:

Be it known that I, ROBERT J. WALLACE, a citizen of the United States of America, residing at Denver, in the county of Arapahoe and State of Colorado, have invented certain new and useful Improvements in Business-College Cabinets; and I do declare the following to be a full, clear, and exact description of the invention, such as will enable others skilled in the art to which it appertains to make and use the same, reference being had to the accompanying drawings, and to the figures of reference marked thereon, which form a part of this specification.

My invention relates to improvements in cabinets for use in colleges and other institutions where bookkeeping and business methods are taught; and it consists of the features hereinafter described and claimed, all of which will be fully understood by reference to the accompanying drawings, in which is illustrated an embodiment thereof.

In the drawings, Figure 1 is a vertical longitudinal section taken through my improved cabinet. Fig. 2 is a front view of a cabinet, partly in section, taken on the line *y y*, Fig. 3. Fig. 3 is a fragmentary section illustrating the locking mechanism. This view is taken on the line *x x*, Fig. 2. Fig. 4 illustrates a detail of construction.

Similar reference-characters indicating corresponding parts in the views, let the numeral 5 designate the sides, 6 the bottom, and 7 the top, of the cabinet. These walls are preferably composed of wood, the top and bottom being interiorly grooved to receive the partitions 8, composed of cardboard or other suitable material. These partitions occupy positions parallel with the side walls 5, and they separate the inclosed space into a series of compartments for the reception of envelops containing instructions in business transactions, the method of entering them in books, the manner of preparing these papers, together with samples and forms. These envelops may also contain money, blank forms of business-papers for the student to fill out, business-papers already prepared in the business the student is doing or which is being done by the person for whom the student is keeping books. The envelops also contain model papers and books to aid the student in his work.

In the front end of the cabinet and located on opposite sides of the partitions 8 are two vertical rollers 9, whose extremities are respectively journaled or pivoted in the top and bottom of the cabinet. As shown in the drawings, the rollers consist of hollow cylinders, preferably composed of tin, having closed ends which are apertured to receive loose spindles 10, which are passed therethrough and protrude sufficiently to form journals which engage sockets in the top and bottom walls of the case or cabinet. The lower sockets do not extend entirely through the bottom 6. Hence the spindles cannot drop through or escape from the cabinet, case, or box. The upper sockets pass entirely through the top of the box, which is recessed or slotted to receive a thin strip 25 of metal or other material which intersects the sockets and passes just above the spindle extremities. This strip prevents the spindles from dropping out when the cabinet is turned over. It is, however, removable, and when taken out the spindles may be removed and inserted at will. When the axles or spindles are taken out, the rollers are released, allowing any adjustment or regulation of the parts which may be necessary or desirable.

The two rollers are connected by two tapes 12 and 13, composed of metal or any other suitable material. The extremities of each tape are made fast to the rollers. The tape 13 is attached to the roller extremities below the partitions 8. The tape 12 is passed under and the tape 13 over each roller, so that as the rollers are turned one tape is winding upon while the other is unwinding from each roller. The tape 12 is located about midway between the top and bottom of the cabinet and is connected with a vertical slotted slide 14, whose extremities are T-shaped and engage suitable guide-grooves formed in the top and bottom of the case. It is divided where it is attached to the slide, since it must not cross the slot therein. This slot is narrow and adapted to register successively with the compartments formed by the partitions 8 as the slide is adjusted or moved. Hence as the slide is moved from the left toward the right (referring to Fig. 2) each envelop may be removed in succession from its compartment by drawing it therefrom through the slot in the slide. The width of the slot, however, will only permit the removal of one envelop at a time. Hence when any envelop has been returned to its compartment and the slide moved to the next compartment the previously-used envelop cannot be again taken from its compartment until the slide is moved backward or in the opposite direction. The cabinet is provided with locking mechanism to prevent this return movement, as hereinafter described.

The slide 14 is provided intermediate its extremities with an enlargement 14^a, to permit the insertion of the thumb and finger for the purpose of removing the envelops as the slide is adjusted. The partitions are provided with recesses 8^c to enable the student to easily grasp the envelops whose forward extremities project into the said recesses. The tape 12 prevents the removal of any envelop until the slide is so adjusted that its slot registers with the compartment containing the envelop.

One of the rollers is provided with a series of short vertical slots 10^a, adapted to be engaged by a spring dog or pawl 15, mounted on a suitable support. The dog normally engages the slots 10^a of the roller and is so constructed that it will permit the roller to turn freely in one direction, but lock it against movement in the opposite direction. Provision is made, however, for disengaging this dog from the roller and holding it in the disengaged position when it is desired to reverse the movement of the rollers. This may be accomplished in any suitable manner. As shown in the drawings, a lever 16 is fulcrumed on the casing at 17. The lower extremity of this lever is connected with the spring-dog 15 by means of a link 18, while the upper extremity of the lever lies in the path of a bolt projecting from the casing 19 of any suitable locking mechanism. This bolt is controlled by a suitable key 20 in the ordinary manner. The key is carried by the teacher or instructor. When the student is at work, the parts are in the position shown in full lines in Fig. 3. Hence the slide can be moved only in one direction.

The dotted lines in Fig. 3 show the position of the parts when the roller is unlocked. This is accomplished by turning the key 20 in the casing 19 and throwing the bolt outward to engagement with the upper extremity of the lever 16. This engagement of the bolt actuates the lever sufficiently to disengage the dog 15 from the roller.

The slide 14 may be formed of a single piece of wire bent into the desired shape, or it may be of any other desired construction. The slots 10^a engaged by the locking-pawl 15 should all be of the same width and the same distance apart. The arrangement is preferably such that every time the dog or pawl 15 passes from one notch to another the slide moves from one compartment to another of the case or cabinet. If desired, the partitions 8 may be omitted and the envelops may be of sufficient width to engage the grooves in the top and bottom of the case.

In describing the use of the device I will assume that it is locked—that is to say, that the spring dog or pawl 15 and its controlling-lever are in the position shown in full lines in Fig. 3, or so that the slide 14 may be moved in one direction only. I will further assume that the student is using the cabinet and that the direction of movement is from the left toward the right, referring to Fig. 2. As the slide is moved so that its slot registers with any compartment of the case he removes the envelop in that compartment, uses its contents, makes the necessary entries in his book, replaces the envelop and contents in its compartment, and moves the slide to the next compartment toward the right, and so on until every envelop has been removed and its contents used. It will be observed that by virtue of the engagement of the pawl 15 with the slots 10^a of the roller (see Fig. 4) the slide cannot be moved toward the left until the roller is released from engagement with the dog. This is done by the teacher or instructor at the proper time by inserting and turning the key 20 in the lock-casing 19 and projecting the bolt, thus throwing the parts to the position shown by dotted lines in Fig. 3. The slide may then be moved backward toward the left at will.

Some of the advantages of my improved cabinet will now be described. The matter contained in the envelops can be reached, used, and replaced by the student, but he cannot anticipate what he has before him, neither can he turn to instructions he has passed over to make up neglected work, thus making the process of his work very similar to that practiced in an office and yet permitting full instructions for each transaction with model entries, books, and business papers for similar but not like transactions. It makes it convenient to have matter of different kinds for each student who is working in any part of his course of study, so that there cannot be any copying in the schoolroom; it permits any desired variation in the work by changing the instructions and matter of one or more envelops in a cabinet; it permits (by having several cabinets) of having only a small portion of the student's work before him at any one time, so that he is not discouraged by seeing what is ahead of him; it allows the student to start his business course as bookkeeper for a business man who transacts the business and instructs the student how to make the proper entries and business papers; it makes it practicable to have business papers and cash pass through the student's hands the same as in a first-class business house; it permits the directing of the student in his actual business practice without his being able to anticipate or to gather up neglected work gone over; it permits the using of the same instructions many times over, thus lessening expenses; it allows the student to keep his papers in compact form and free from dust; it permits complete and full instructions with reference to each step of the work, thus relieving the teacher to a certain extent and making the student more self-reliant, and, finally, it permits each school using the cabinet to adopt such a system of instructions in bookkeeping and of business forms as it thinks best.

Between the sides 5 of the case and the outermost partition 8 are two compartments 22 not controlled by the tape 12. These are for the student's private use. In the top and bottom of the case are formed two grooves 23 in front of the slotted slide. They are for the insertion of a sliding front end piece, (not shown,) which may be employed when the cabinet is not in use. This feature permits the closing of the cabinet perfectly tight, whereby all dust is excluded. The opposite end of the cabinet is tightly closed by a suitable wall 4.

Having thus described my invention, what I claim is—

1. A cabinet of the class described adapted to hold envelops or documents, a slotted device movably connected with the open end of the cabinet, and means connected with the slotted device for preventing the removal or insertion of papers except by way of the slot in said movable slide.

2. A case or cabinet of the class described adapted to contain business papers and being open at one end or side, a slotted device attached to the open end of the case and having an enlargement intermediate its extremities to permit the insertion of the thumb and finger in removing papers, and means connected with the slide for preventing the insertion or removal of papers except by way of the slot in the slide.

3. A cabinet of the class described adapted to hold envelops or other papers or documents, a slotted device movably connected with the open end of the cabinet, and means connected with the slotted device for preventing the removal or insertion of papers, except by way of the slot in the slide, and suitable means for locking the slotted slide against movement in one direction.

4. A cabinet or case adapted to contain business papers and open at one end or side, a slotted slide connected with the open end of the case, rollers pivoted in the case, and a ribbon, tape, or flexible band 12 connected with the slide and having its extremities attached to the rollers, said tape being adapted to prevent the insertion or removal of papers except by way of the slot in the slide.

5. A cabinet or case adapted to contain business papers and open at one side or end, a slotted slide connected with the open end of the case, rollers pivoted in the case, and a ribbon, tape, or flexible band 12 connected with the slide and having its extremities attached to the rollers, said tape being adapted to prevent the insertion or removal of papers except by way of the slot in the slide, and suitable means engaging one of the rollers for locking the slide against movement in one direction, and another tape connected with the rollers, but adapted to wind in a direction opposite from the tape 12, whereby as one tape is winding up on each roller, the other tape is unwinding from the same roller.

6. A cabinet or case adapted to contain business papers and open at one end or side, a slotted slide connected with the open end of the case, rollers pivoted in the case, and a ribbon, tape, or flexible band 12 connected with the slide and having its extremities attached to the rollers, said tape being adapted to prevent the insertion or removal of papers, except by way of the slot in the slide, suitable means engaging one of the rollers for locking the slide against movement in one direction, said means comprising a spring-pawl normally engaging a ratchet of the roller.

7. A cabinet or case adapted to contain business papers and open at one end or side, a slotted device connected with the open end of the case, rollers pivoted in the case, and a tape having its extremities attached to the rollers, said tape being adapted to prevent the insertion or removal of papers except by way of the slot in the slide, suitable means engaging one of the rollers for locking the slide against movement in one direction, said means comprising a spring-pawl normally engaging a ratchet in the roller, and suitable means for disengaging the pawl from the roller, comprising a lever connected with the pawl, and a key-operated bolt for actuating the lever.

8. A cabinet or case divided by suitable partitions into a series of compartments, said cabinet being open at one end, a slotted slide connected with the open end of the case, and means connected with the slide for preventing access to the compartments of the case except by way of the slot in the slide.

In testimony whereof I affix my signature in presence of two witnesses.

ROBERT J. WALLACE.

Witnesses:
G. J. Rollandet,
Edith Himsworth.

US006119651A

United States Patent [19]

Anderson

[11] **Patent Number: 6,119,651**

[45] **Date of Patent: Sep. 19, 2000**

[54] **HYDROGEN POWERED VEHICLE, INTERNAL COMBUSTION ENGINE, AND SPARK PLUG FOR USE IN SAME**

[75] Inventor: **Herman P. Anderson**, Brentwood, Tenn.

[73] Assignee: **Herman P. Anderson Technologies, LLP**, Brentwood, Tenn.

[21] Appl. No.: **09/220,286**

[22] Filed: **Dec. 23, 1998**

Related U.S. Application Data

[63] Continuation-in-part of application No. 08/905,862, Aug. 4, 1997, Pat. No. 5,852,993.

[51] **Int. Cl.**^{7} **F02M 57/06**

[52] **U.S. Cl.** **123/297**; 123/635; 123/DIG. 12

[58] **Field of Search** .... 123/3, DIG. 12, 123/527, 297, 636, 637, 635, 169 EL; 313/120

[56] **References Cited**

U.S. PATENT DOCUMENTS

2,208,030	7/1940	Holmes	123/169 EL
3,165,099	1/1965	Vanderpoel	123/637
3,926,169	12/1975	Leshner et al.	123/297
3,980,061	9/1976	McAlister	123/297
4,164,912	8/1979	Beyler	123/26
4,186,712	2/1980	Fitzner et al.	123/633
4,319,552	3/1982	Sauer et al.	123/297
4,343,272	8/1982	Buck	123/297
4,383,198	5/1983	Hosking	313/120
4,448,160	5/1984	Vosper	123/297
4,820,957	4/1989	Zivkovich	315/209 T
5,852,993	12/1998	Anderson	123/297

FOREIGN PATENT DOCUMENTS

686943	8/1930	France	123/297

Primary Examiner—Noah P. Kamen
Attorney, Agent, or Firm—Lucian Wayne Beavers Waddey & Patterson

[57] **ABSTRACT**

The internal combustion engine fuel system described includes a structure for mixing the alternative fuel, preferably hydrogen, with oxygen in ambient air to stratify the fuel. The system includes an adapter, and the adapter includes a housing mounted between spark plug and cylinder of the internal combustion engine. A plug is placed within the housing. The plug has ridges or grooves on its outer surface that act as mixing structures. Thus, when hydrogen is introduced into the adapter housing it is mixed with ambient oxygen within the chamber as it flows over the plug. The mixing structures in the housing creates a vortexing action as the hydrogen flows over the plug and towards the cylinder of the engine. An electrode protrudes from the plug towards the cylinder. The electrode is preferably platinum and generates the necessary spark to create combustion of the hydrogen/air mixture adjacent to the cylinder to thereby power the cylinder in the engine. A platinum electrode is preferably used because it enhances a catalytic conversion of combustion by-products to more environmentally compatible products. The present invention also teaches a spark plug producing a hotter spark for a hydrogen fuel system. Also taught is a hydrogen powered vehicle with reduced emissions by producing a spark during the power stroke and the exhaust stroke. Methods for reducing exhaust pollution are also taught. Methods of reducing exhaust pollution by generating a plasma are taught as well.

13 Claims, 12 Drawing Sheets

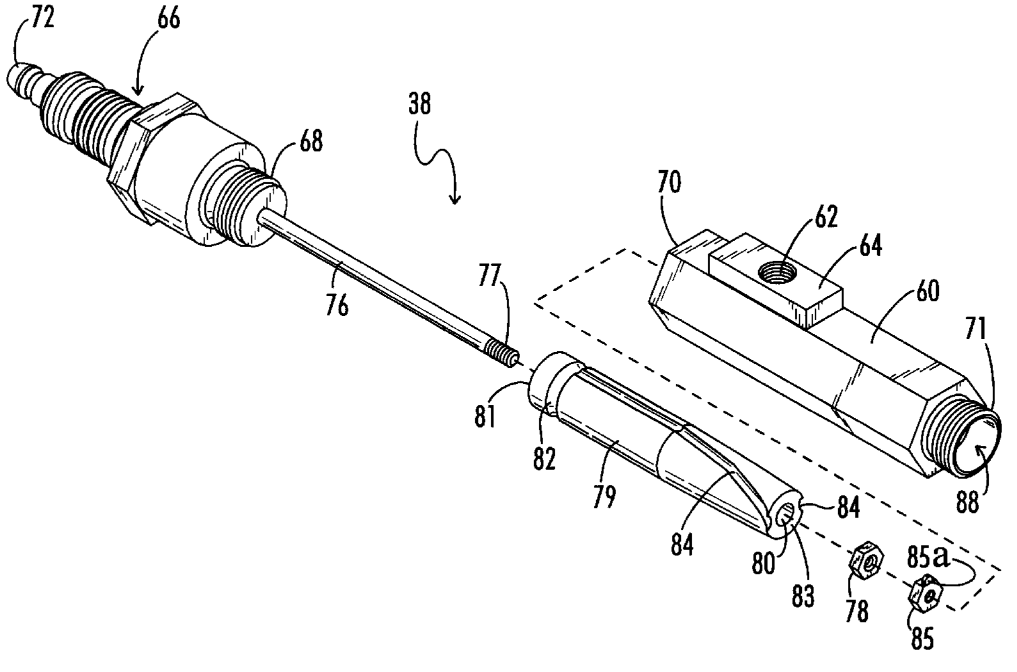

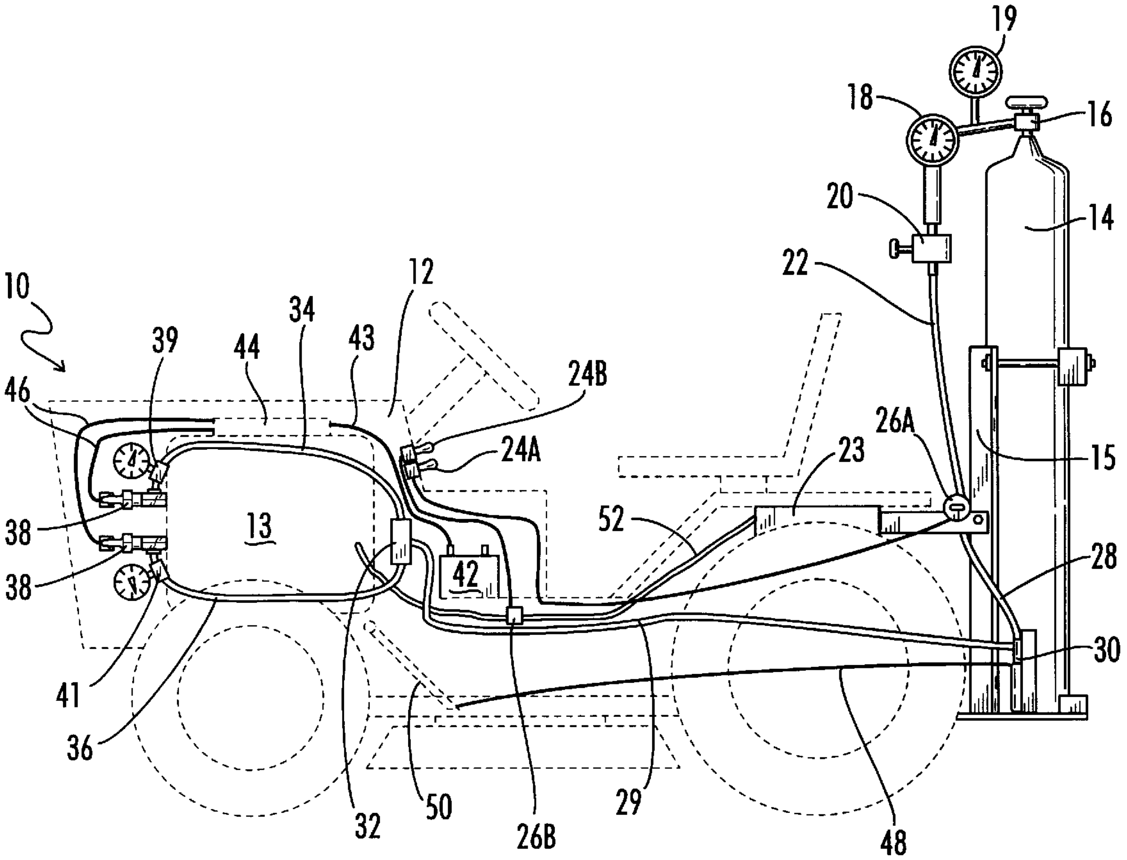

FIG. 1

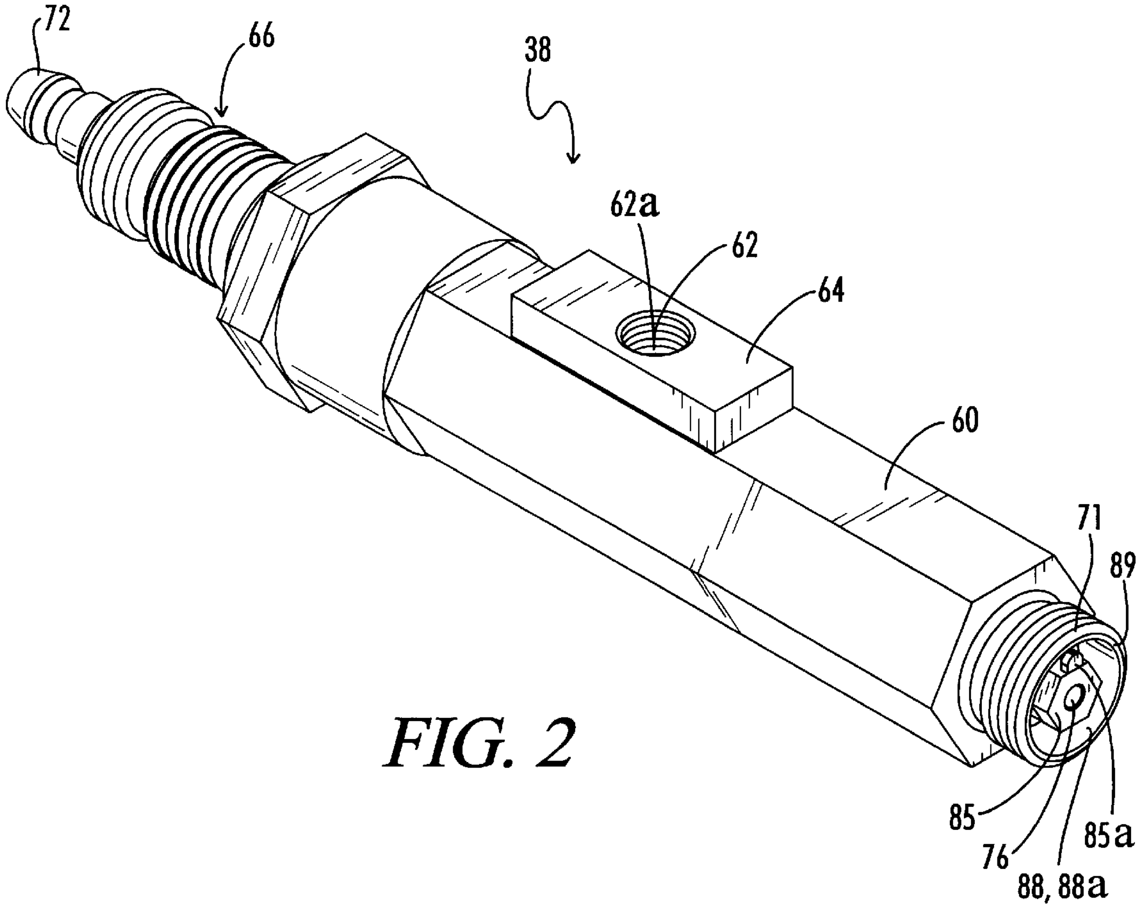
72
66
38
62a
62
64
60
71
89
85
76
85a
88, 88a

FIG. 2

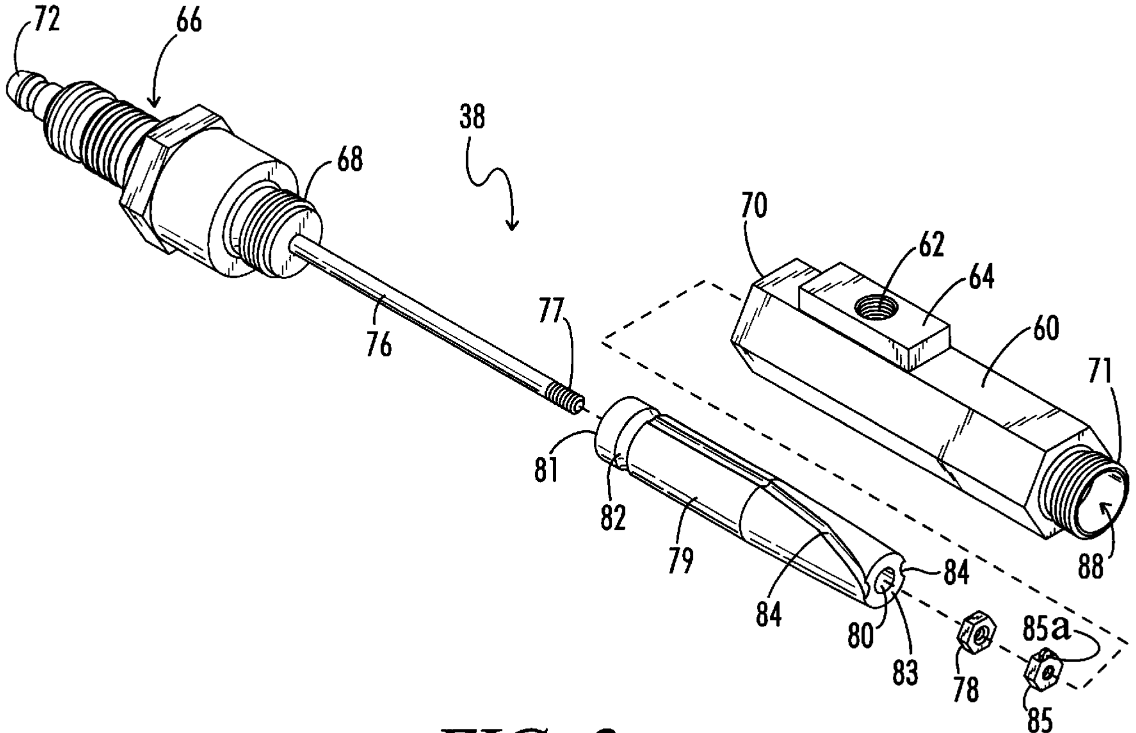
72
66
38
68
70
62
64
60
77
76
71
81
82
79
84
88
84
80
83
78
85a
85

FIG. 3

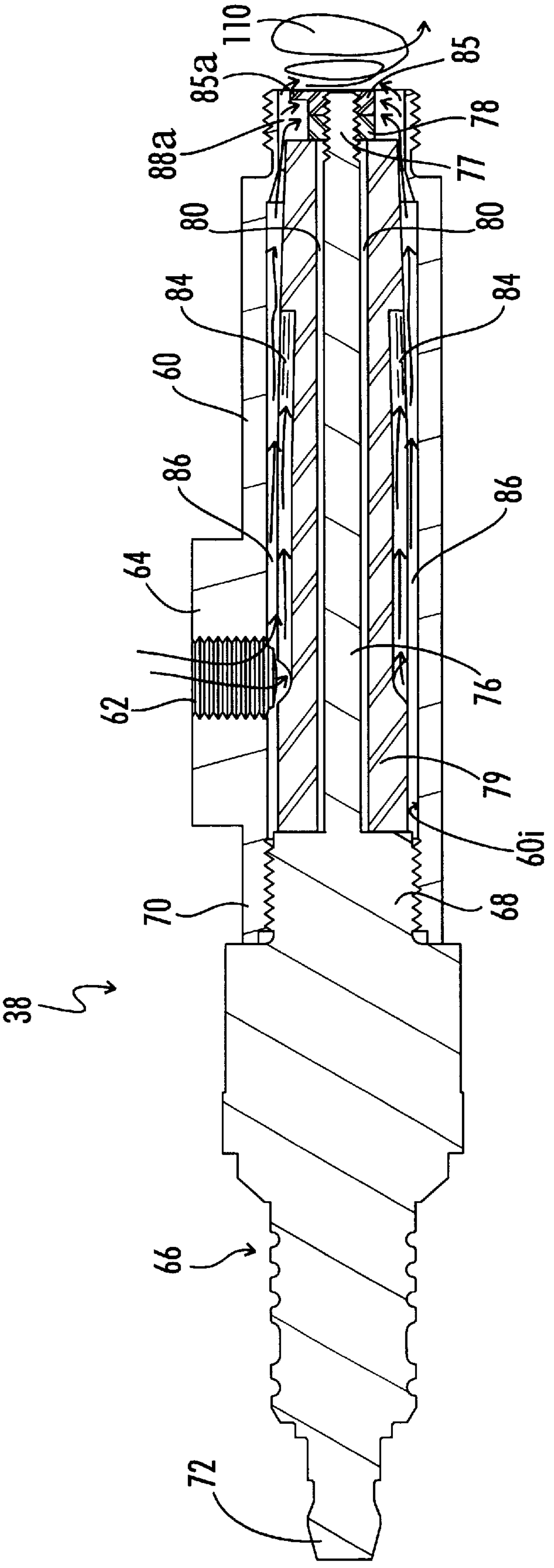

FIG. 4

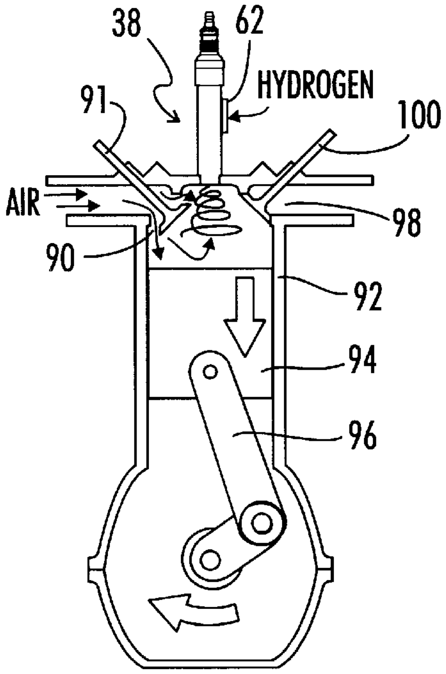
38
62
HYDROGEN
91
100
AIR
98
90
92
94
96

INTAKE

FIG. 5A

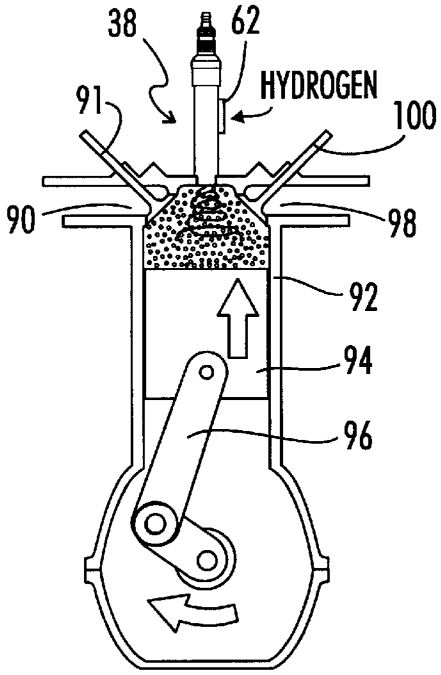
38
62
HYDROGEN
91
100
90
98
92
94
96

COMPRESSION

FIG. 5B

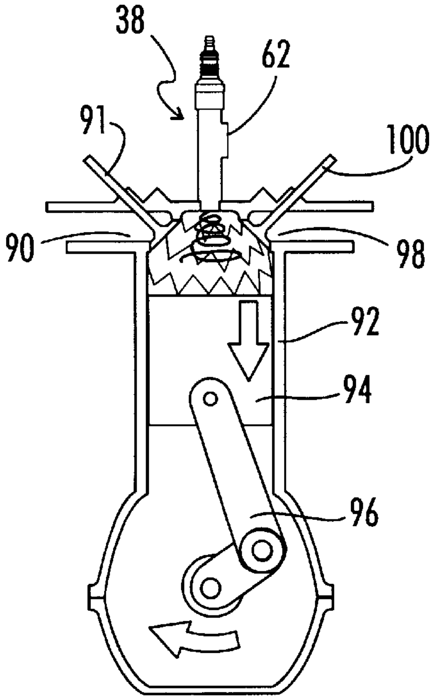
38
62
91
100
90
98
92
94
96

POWER

FIG. 5C

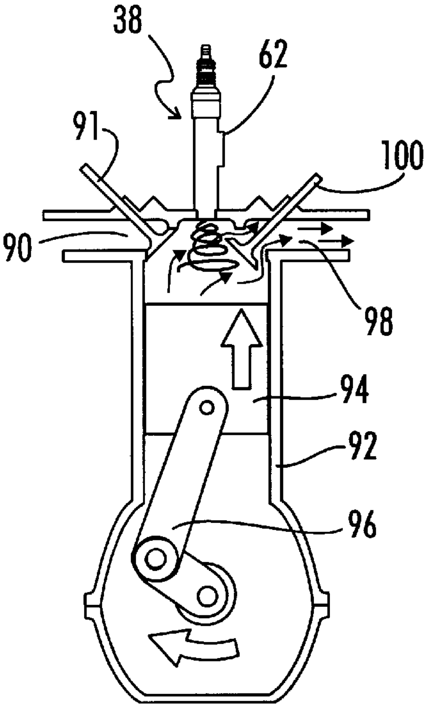
38
62
91
100
90
98
94
92
96

EXHAUST

FIG. 5D

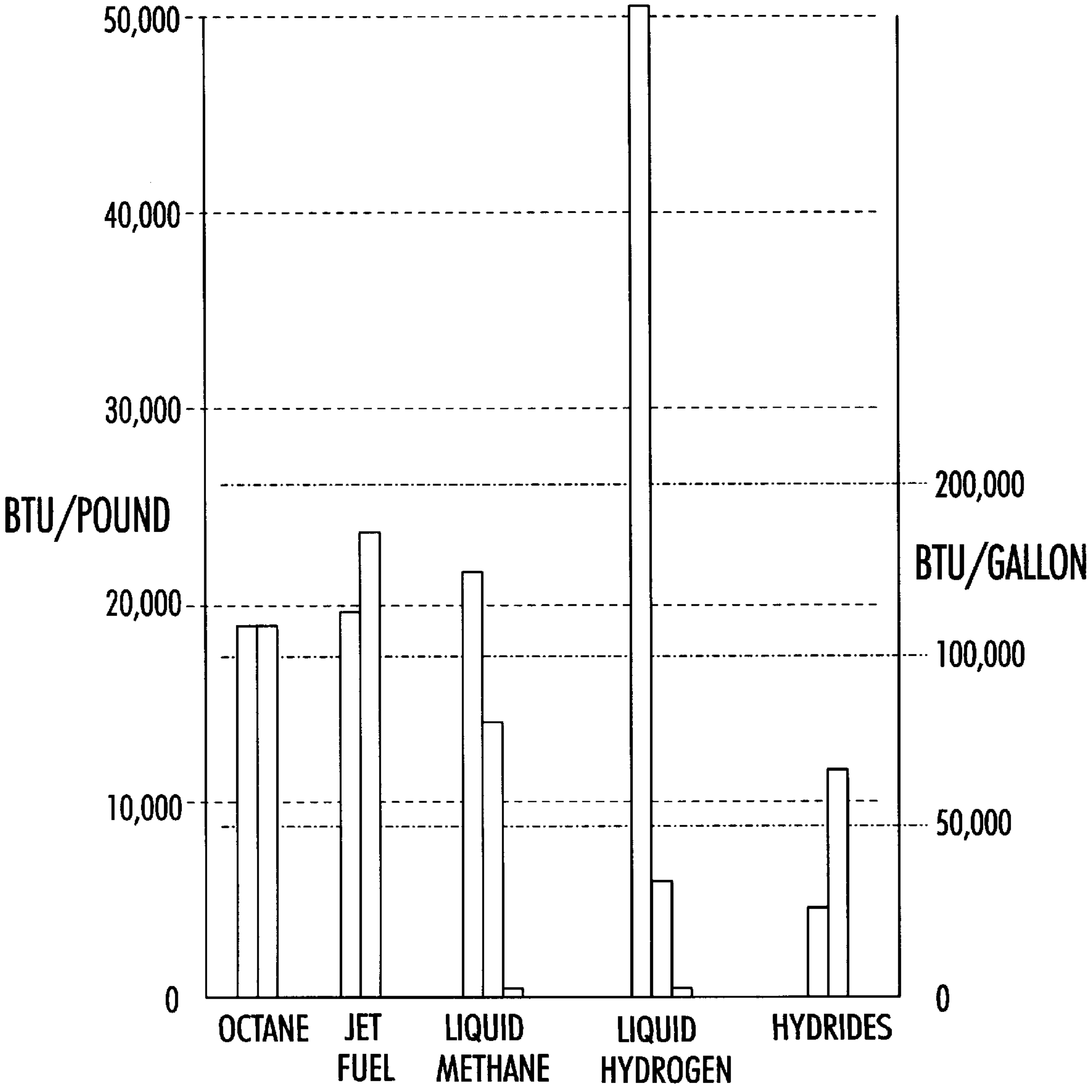

METHANE GAS - 140 BTU/GAL
HYDROGEN GAS - 40 BTU/GAL

FIG. 6

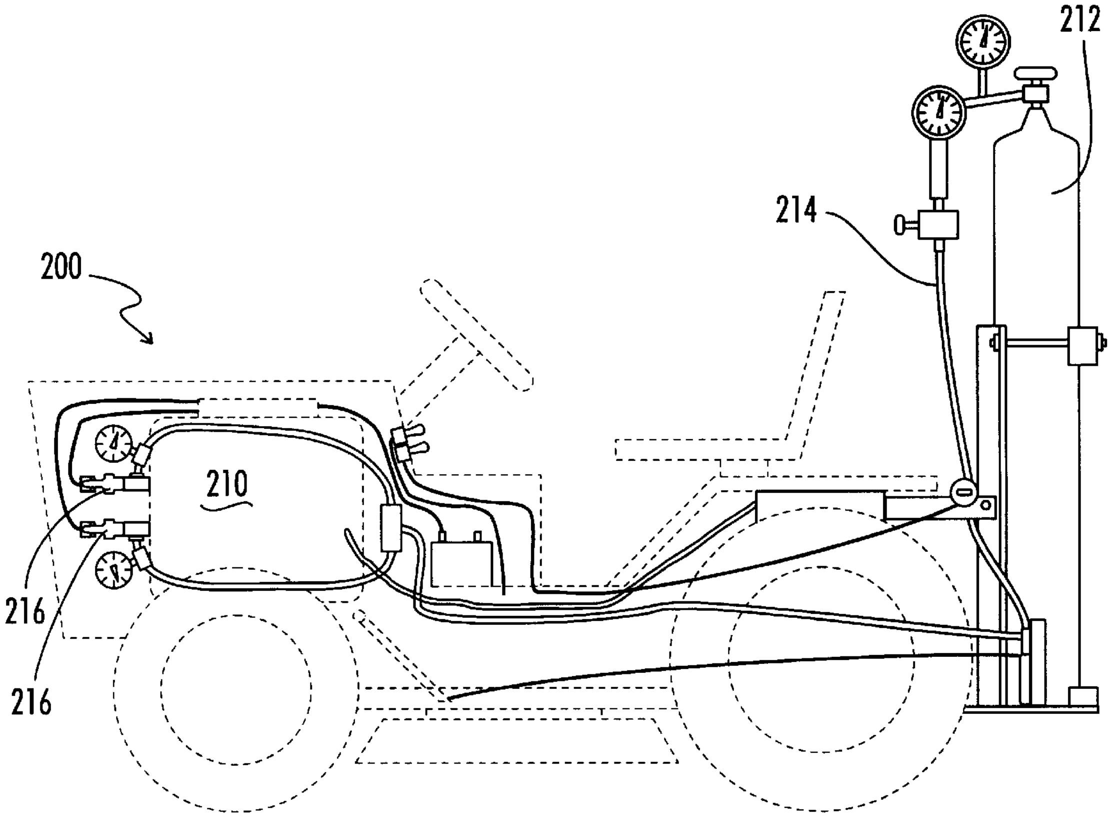
212
214
200
210
216
216

FIG. 7

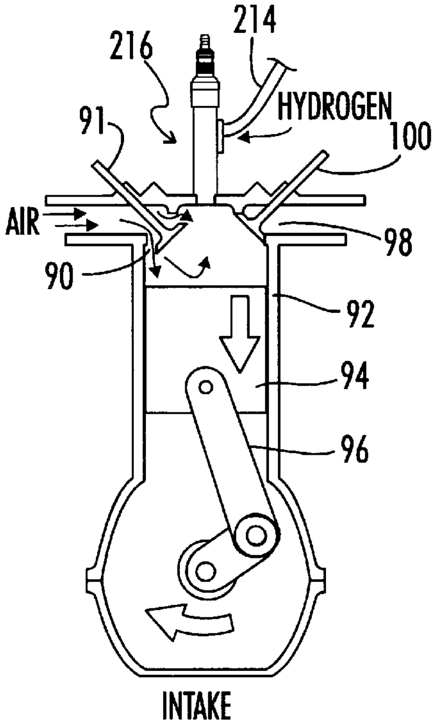
216
214
HYDROGEN
91
100
AIR
98
90
92
94
96
INTAKE

FIG. 8

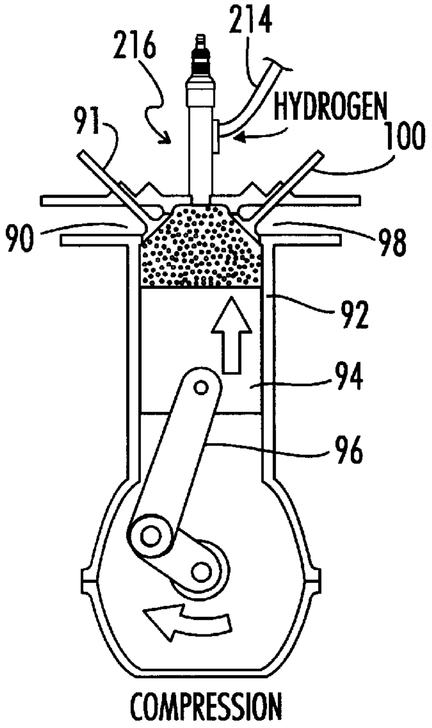
216
214
HYDROGEN
91
100
90
98
92
94
96
COMPRESSION

FIG. 9

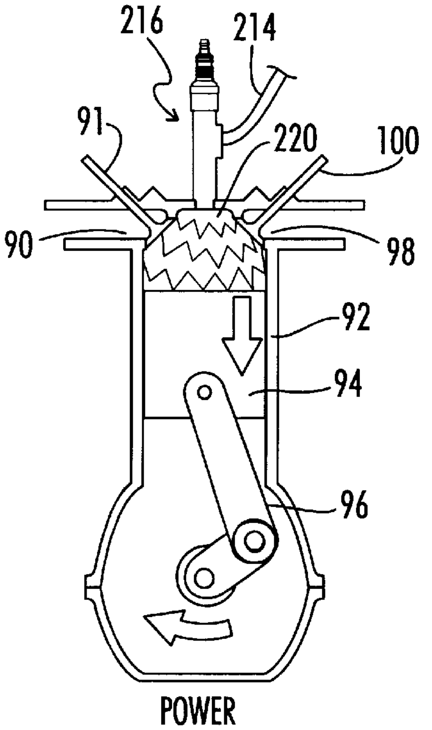
216
214
91
220
100
90
98
92
94
96
POWER

FIG. 10

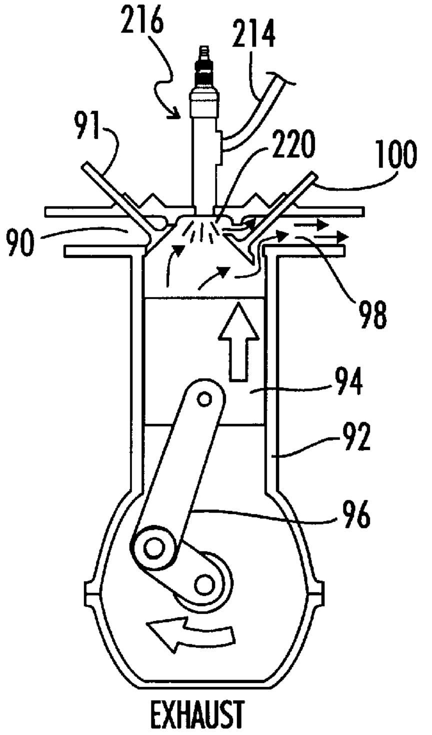
216
214
91
220
100
90
98
94
92
96
EXHAUST

FIG. 11

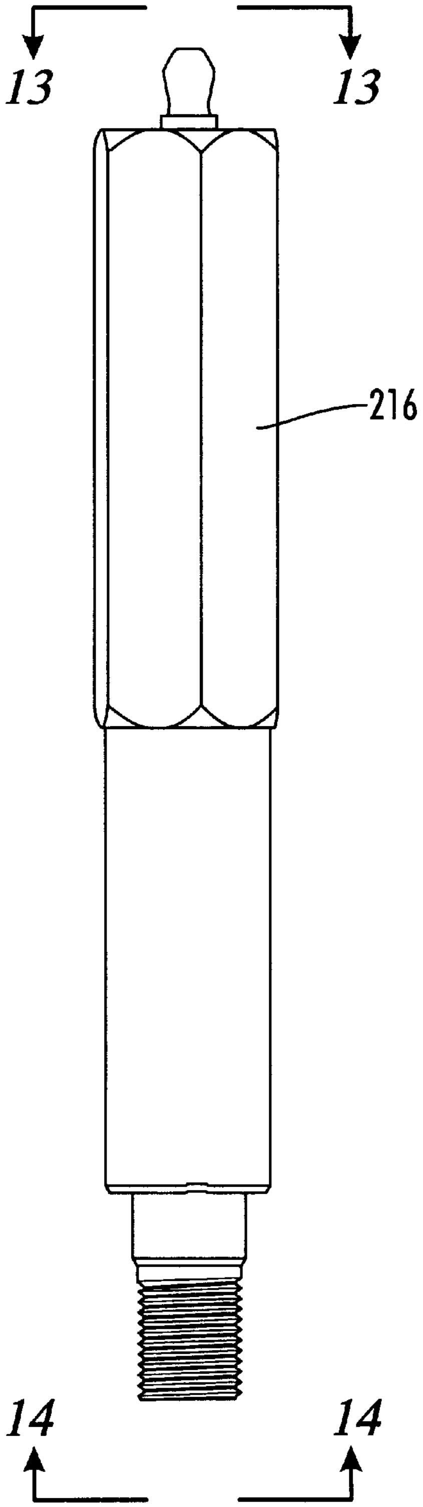
13
13
216
14
14

FIG. 12

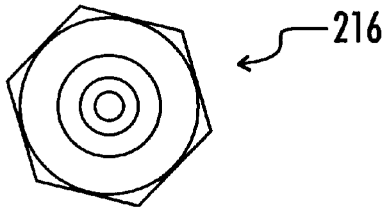

FIG. 13

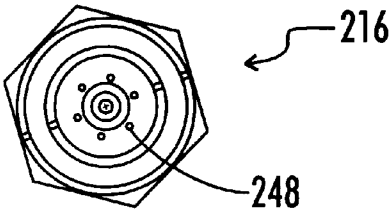

FIG. 14

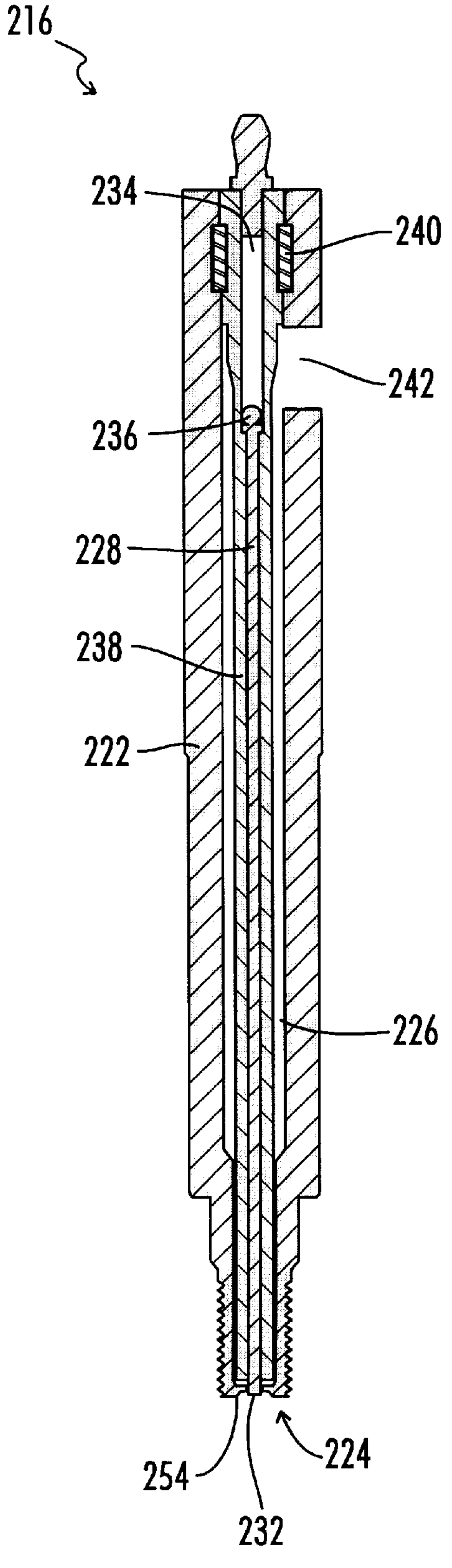

FIG. 15

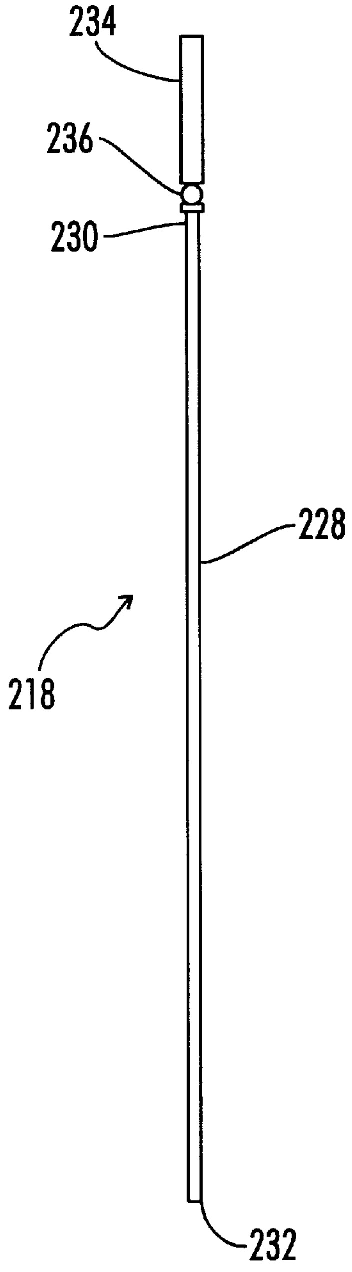

FIG. 16

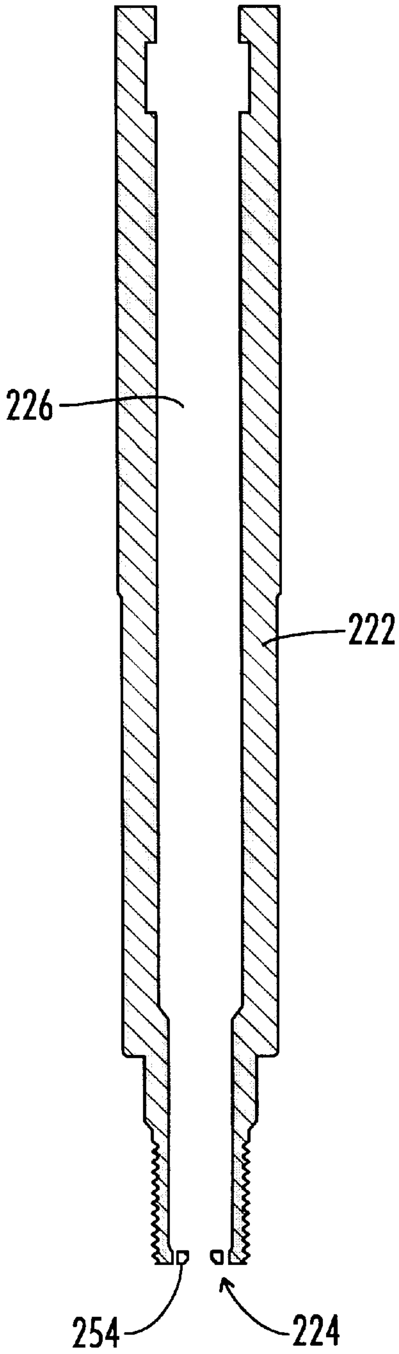
226
222
254
224

FIG. 17

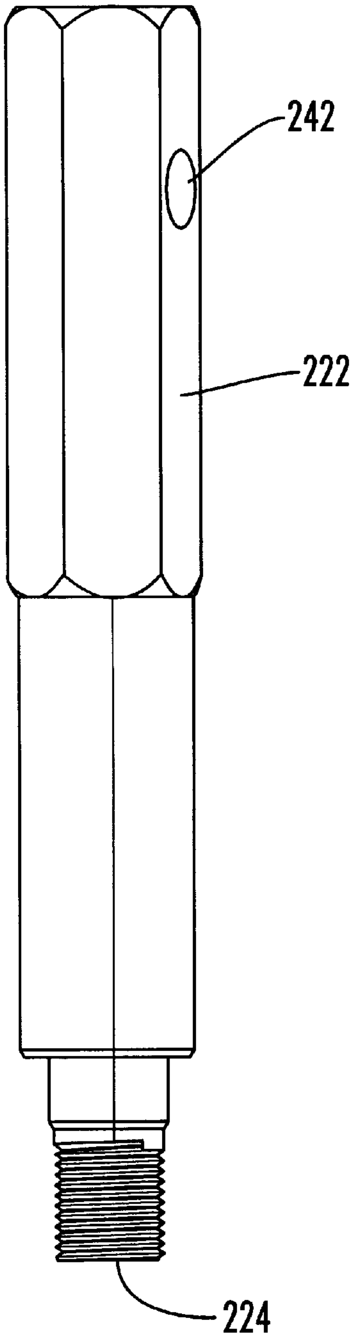

FIG. 18

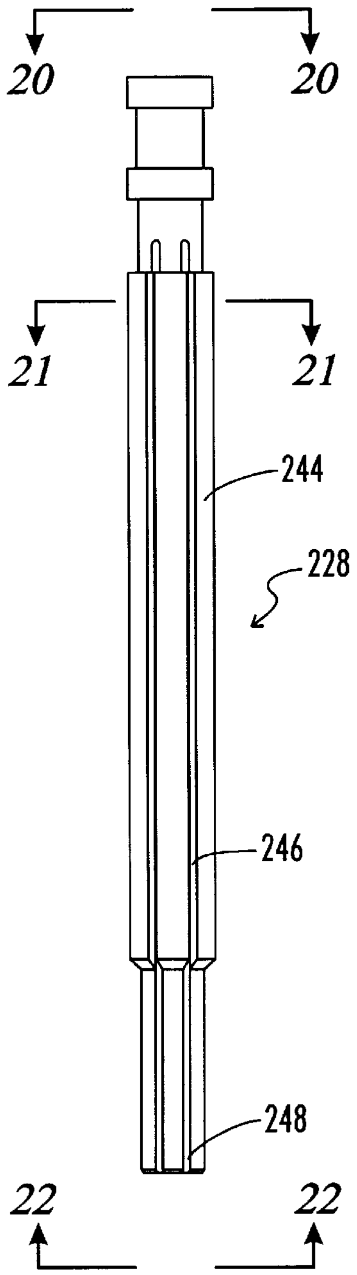
20
20
21
21
244
228
246
248
22
22

FIG. 19

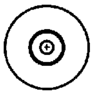

FIG. 20

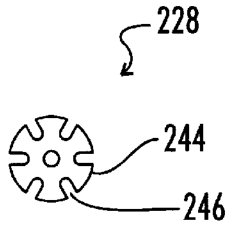

FIG. 21

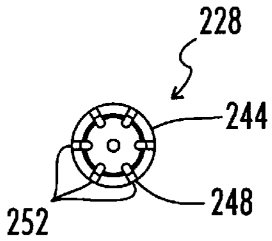
228
244
252
248

FIG. 22

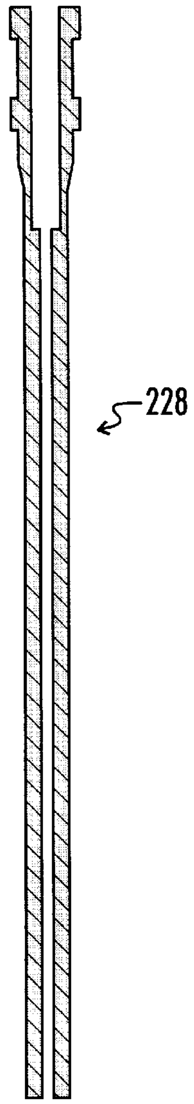

FIG. 23

HYDROGEN POWERED VEHICLE, INTERNAL COMBUSTION ENGINE, AND SPARK PLUG FOR USE IN SAME

This application is a continuation-in-part of U.S. patent application Ser. No. 08/905,862 filed Aug. 4, 1997 by Herman P. Anderson, entitled "Fuel System For Internal Combustion System And Adapter For Use In Same" , now U.S. Pat. No. 5,852,993. U.S. patent application Ser. No. 08/905,862 is hereby incorporated herein by reference.

BACKGROUND OF THE INVENTION

The present invention relates generally to a fuel system for internal combustion engines that enables the invention to utilize alternative fuels; and, more particularly to an improved fuel system for an internal combustion engine and an adapter and spark plug for use in same.

The problems of air pollution caused by automobiles, lawn mowers and other vehicles that have internal combustion engines powered by gasoline is well documented. This problem is particularly described in an article in the San Diego Union Tribune, Wednesday, Sep. 25, 1996 and in Popular Science, August 1996. Indeed, under the Clean Air Act, in the next few years cars that do not run on standard fossil fuels like gasoline will be required in several states. Currently, there are no vehicles available to meet this need.

Thus, it will be appreciated by those skilled in the art that there has been a push to seek other fuels for use in internal combustion engines. These fuels include methane, propane and gasohol.

Hydrogen has also been characterized as a promising fuel source. See Department of Energy publication entitled "Hydrogen Fuel", printed in 1978, U.S. Government Printing Office, U.S. Department of Energy, Office of Public Affairs, Washington, DC 20585. Particularly, the graph presented in FIG. **6**, which is copied from this publication, shows the potential for hydrogen in BTUs/Lb. However, the mass energy density for hydrogen is low; and thus, hydrogen needs to be compressed before its energy potential can be realized. Hydrogen is a "dry fuel" that is more difficult to ignite than "wet fuels." It becomes more difficult to ignite when it is cooled. However, cooling is preferred in most systems to avoid premature ignition of the fuel. Other systems deliver the hydrogen under pressure to prevent ignition in the supply line, which is a form of pre-ignition. No prior art fuel system has been able to capture the energy of hydrogen.

Several fuel systems for use with hydrogen have been proposed. Examples are described in U.S. Pat. Nos. 4,167,919; 4,253,428; 4,016,836; 4,178,882; 5,222,993; 5,085,176; and 5,085,176. The problem with these prior art fuel systems is they cannot take advantage of the alternative fuels to generate sufficient power from the fuels to drive an internal combustion engine for use in a lawn mower, automobile or other vehicle. As noted above, this is a particular problem when hydrogen is used as a fuel. Additionally, the prior art systems do not address the problem of back fire, i.e. undesired explosion of the fuel in the cylinder during the exhaust stroke of the internal combustion engine.

Premature ignition or backfire is often a problem with hydrogen fueled systems. U.S. Pat. No. 4,383,198 (Hosking) teaches a fuel injection spark plug with a pre-cooling chamber to avoid premature ignition indentination of a gaseous fuel supplied to the spark plug. U.S. Pat. No. 4,383,198, Hosking, entitled "Fuel Injection Spark Plug", issued May 10, 1983, is hereby incorporated herein by reference. However, cooling hydrogen makes it more difficult to ignite. This is because hydrogen is a dry fuel whereas gas is generally considered a wet fuel. Thus to reduce fuel pollution one prefers to utilize a fuel such as hydrogen and ignite it with a hotter spark.

One patent discussing very high energy (VHE) ignition systems is U.S. Pat. No. 5,207,208 issued to Ward, entitled "Integrated Convertor High Power CD Ignition", issued May 4, 1993. U.S. Pat. No. 5,207,208 is hereby incorporated herein by reference. Pat. '208 teaches very high energy ignition systems, for instance on the order of 36K volts. It teaches voltage doubling utilizing pulsating sparks of peak current on the order of two amps. While 36K volts may at times be adequate, it is preferable to achieve a higher voltage in the ignition chamber when using hydrogen.

It is known in the art to use a direct ignition system, also referred to as a distributorless ignition system, to increase the spark plug voltage. Such direct ignition systems use a "waste spark" method of spark distribution. Each cylinder is paired with its opposing cylinder in the firing order (1-4, 2-3, on a four cylinder, 1-4, 2-5, 3-6 on a V6) so that one cylinder on compression fires simultaneously with its opposing cylinder on exhaust stroke. Since the cylinder on exhaust stroke requires very little of the available voltage to fire its plug, most of the voltage is used to fire the cylinder on the compression stroke.

U.S. Pat. No. 4,462,380 discusses a distributorless ignition system. U.S. Pat. No. 4,462,380 entitled "Enhanced Spark Energy Distributorless Ignition System" issued Jul. 31, 1984 by Asik is hereby incorporated herein by reference. Pat. '380 teaches a distributorless ignition system of an internal combustion engine which has a supplementary spark energy module to increase spark energy. Each high voltage terminal is connected to a single spark plug and each ignition coil primary is alternately energized and quickly de-energized, producing opposite polarity ignition voltages at each coil terminal. As a result, pairs of spark plugs are alternately fired, with each firing pair occurring in a compression or exhaust stroke and thereby providing the proper ignition to the engine. Asik asserts that firing a spark plug does not affect engine performance or emissions. As taught by the instant invention this is incorrect for hydrogen fuels when the exhaust is sparked at a sufficient voltage.

U.S. Pat. No. 5,146,882 (Brinkman) teaches use of specific wet fuel mixtures in conjunction with direct ignition systems (D.I.S.). It teaches that technical breakdown voltages for typical spark plugs are on the order of 15–20 kV and DIS typically yield between 30 kV–40 kV. U.S. Pat. No. 5,146,882, entitled "Method And Apparatus For Cold Starting A Spark Ignited Internal Combustion Engine Fueled With An Alcohol-Based Fuel Mixture", by Brinkman, et al., issued Sep. 15, 1992, is hereby incorporated herein by reference. The previously referenced prior art discusses alternate fuels and fuel systems but does not teach a method of achieving a sufficient voltage with a hydrogen system.

What is needed, then, is a fuel system for an internal combustion engine that provides a way to generate sufficient power from the alternative fuels and that reduces backfire problems. Such a system is lacking in the prior art. Also lacking in prior art is a system utilizing alternate fuels with reduced pollution capabilities. Also lacking is a way of generating sufficient voltage spark to ignite the alternate fuel.

The following U.S. Patents discuss spark plugs designed for gas internal combustion engines to reduce radio frequency noise and interference. U.S. Pat. No. 4,224,554, entitled "Spark Plug Having A Low Noise Lever" issued to Nishio et al., Sep. 23, 1980 is hereby incorporated herein by reference. U.S. Pat. No. 4,029,990 entitled "Spark Plug Construction" issued to Nagy et al., Jun. 14, 1977 is hereby incorporated herein by reference.

SUMMARY OF THE INVENTION

It is an object of this invention to provide an improved fuel system for use with alternative fuels.

It is a further object of this invention to provide an improved fuel system for use particularly with hydrogen, and particularly to enhance the mass energy density of hydrogen.

It is an object to provide a hotter sparking plug to ignite the fuel.

It is still a further object of this invention to provide an adapter for use in a fuel system for an internal combustion engine that enhances the power obtained from the alternative fuels.

It is still a further object of this invention to provide an improved fuel system for an internal combustion engine that does not present the environmental hazards of a standard gasoline engine by providing reduced or zero emission of environmentally harmful gases.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that can use an alternative fuel and can use gasoline to facilitate the transition from gasoline to alternative fuels.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that reduces backfire problems.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that reduces backfire problems by cooling the electrode in the spark plug with the hydrogen gas fuel.

It is an object to reduce backfire by providing hydrogen at a sufficient pressure to reduce backfire and igniting it with a hotter spark (i.e. a higher voltage).

It is still a further object of this invention to provide an adapter for use in a fuel system for an internal combustion engine that enhances the power obtained from the alternative fuels.

It is a further object to generate a plasma field in a cylinder to burn residual exhaust and reduce emissions.

It is yet a further object of this invention to provide a direct injection fuel system wherein fuel is injected directly into the cylinder.

It is an object to provide a "2-fire" direct injection system, wherein the fuel is ignited during the power stroke and the exhaust stroke.

It is yet a further object of this invention to provide a way to convert the standard induction fuel system of a lawnmower to a direct injection fuel system wherein fuel is injected directly into the cylinder.

It is yet a further object of this invention to provide a way to convert the indirect injection fuel system of an automobile to a direct injection fuel system wherein fuel is injected directly into the cylinder.

The internal combustion engine fuel system of this invention includes a structure for mixing the alternative fuel, preferably hydrogen, with oxygen in ambient air to stratify the fuel. The term "alternative fuels" is known in the art and is meant to refer to fuels other than conventional gasoline.

The invention in one embodiment comprises an adapter, the adapter including a housing mounted between spark plug and cylinder of the internal combustion engine. A plug is placed within the housing. The plug has ridges or grooves on its outer surface that act as mixing structures. Thus, when hydrogen is introduced into the adapter housing it is mixed with ambient oxygen within the chamber as it flows over the plug. The mixing structures on the plug create a vortexing action as the hydrogen flows over the plug and towards the cylinder of the engine.

An electrode protrudes from the plug towards the cylinder. The electrode is preferably platinum and generates the necessary spark to create combustion of the hydrogen/air mixture adjacent to the cylinder to thereby power the cylinder in the engine. A platinum electrode is preferably used because it enhances a catalytic conversion of combustion by-products to more environmentally compatible products.

Thus, the combination of the compression action generated by the conventional piston and push rod within the cylinder of the engine in conjunction with the hydrogen/air mixing structure that vortexes the hydrogen/air create compressed hydrogen that is mixed with a combustion facilitator, i.e. oxygen in the ambient air to enable the system of this invention to capture the abundant energy found in hydrogen fuel.

The enhanced mixing capabilities of this invention is also adaptable for use with other alternative fuels such as natural gas, alcohol, propane and the like.

One embodiment of the present invention is for a hydrogen powered vehicle including an internal combustion engine operably connected to power the vehicle and adapted to connect to a hydrogen supply line. The internal combustion engine comprises a cylinder operably connected to the supply line; a piston operably disposed in the cylinder; and a spark plug having an electrode assembly in operable sparking communication with the cylinder, wherein the spark plug is capable of producing a spark of at least 45 kV.

Preferably the spark plug comprises a housing having a sparking end and a hollow chamber containing the electrode assembly. The electrode assembly comprises an electrode having a connector end and a sparking end, wherein the sparking end is positioned adjacent to the housing sparking end. The electrode assembly preferably comprises a coil electrically connected to the electrode connector.

Preferably the housing comprises a fuel inlet port in fluid communication with the supply line, and an insulator surrounding the electrode assembly, wherein the insulator is positioned in the hollow chamber. Preferably the insulator comprises an outer surface defining a first fuel channel having a first respective fuel exit port, wherein the first fuel channel is in fluid communication with the fuel inlet port and the fuel exit port exits into the cylinder.

The present invention also includes a method of operating an internal combustion engine comprising supplying hydrogen fuel to the engine and creating a spark in the cylinder during the power stroke and the exhaust stroke, wherein the spark created during the power stroke is at least 45 kV. In one embodiment the spark is at least 70 kV. In certain embodiments the method comprises generating plasma during the exhaust stroke. The exhaust fuel converts to an exhaust emission and coalesces with the plasma, thereby reducing fuel pollution.

Accordingly, it is an object in the invention to provide a spark plug producing a sufficiently hot spark to ignite hydrogen fuel, including cooled hydrogen fuel.

Another object of the present invention is to provide a device for reducing fuel exhaust emissions.

Another object of the present invention is to provide a cleaner burning engine.

Another object is to provide a spark plug for reducing emissions.

Another object of the present invention is to provide a device functional with dry fuels. Use of the term hydrogen, as discussed herein, is intended to encompass such dry fuels. However the term as used excludes such wet fuels as gas and alcohol.

Another object of the present invention is to provide an integrated fuel system adapted to provide these and objectives.

Other objects and advantages of the present invention will be apparent to those of skill in the art by reference to the teachings disclosed herein, including the attached drawings, detailed description of exemplary embodiments, and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a plan view of the fuel system of this invention with a lawn mower shown in phantom.

FIG. **2** is a perspective view of the adapter of the fuel system of this invention.

FIG. **3** is an exploded perspective view of the adapter of the fuel system of this invention.

FIG. **4** is a cross-sectional view of the adapter of the fuel system of this invention.

FIGS. **5A–5D** are a schematic of the 4 cycle combustion in an internal combustion engine including the adapter of this invention.

FIG. **6** is a graph which shows the energy potential for hydrogen in BTUs/Lb.

FIG. **7** shows a vehicle of the present invention utilizing hydrogen fuel.

FIG. **8** depicts an intake stroke of a cylinder connected to a hydrogen source.

FIG. **9** depicts a compression stroke of a cylinder connected to a hydrogen source.

FIG. **10** depicts a power stroke of a cylinder connected to a hydrogen source. The spark plug shown in FIG. **10** is depicted sparking.

FIG. **11** depicts a cylinder in an exhaust stroke connected to a hydrogen source, wherein the spark plug is sparking during the exhaust stroke.

FIG. **12** shows an elevated side view of a spark plug of the present invention.

FIG. **13** shows a plan view in the direction of view line **13—13** of the spark plug shown in FIG. **12**.

FIG. **14** depicts a plan view in the direction of view line **14—14** of the spark plug shown in FIG. **12**.

FIG. **15** depicts an elevated side section view of the spark plug shown in FIG. **12**.

FIG. **16** shows a electrode assembly which is disposed in the hollow chamber depicted in FIG. **15**.

FIG. **17** shows a section view of the spark plug housing clearly showing the hollow chamber and a torridal ground electrode.

FIG. **18** depicts an elevated side view of the spark plug housing and a fuel inlet port.

FIG. **19** depicts an insulator comprising fuel channels and fuel exit ports.

FIG. **20** is a top plan view of the insulator shown in FIG. **19** viewed in the direction of view line **20—20**.

FIG. **21** is a cross section view of the insulator shown in FIG. **19** along section line **21—21**.

FIG. **22** is a cross section view of the insulator shown in FIG. **19** viewed in the direction of view line **22—22**.

FIG. **23** is an elevated section view of the insulator shown in FIG. **19**.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring now to the drawings, wherein like reference numerals refer to like parts throughout, the fuel system is referred to generally at **10**. The fuel system of this invention is contemplated for use with an internal combustion power system, such as an internal combustion engine or such as a home heating system. Fuel system **10** is portrayed in this case as connected to a standard lawn mower. The lawn mower is shown in phantom at **12** in FIG. **1**. The lawn mower is a conventional lawn mower as can be purchased from a company such as Murray of Ohio. The internal combustion engine in the lawn mower is a standard 4-cycle engine with two cylinders and develops 18 horsepower. A suitable example can be purchased from Briggs & Stratton, Milwaukee, Wis. 53201.

Referring particularly to FIG. **1**, system **10** includes fuel tank **14** that is mounted to lawn mower **12** via mounting assembly **15**. In the preferred embodiment, fuel tank **14** includes hydrogen gas. Accordingly, for the purposes of the foregoing detailed description, hydrogen gas will be described.

A line **22** begins at fuel tank **14** and proceeds to back-flow valve **16**. Back-flow valve **16** is a standard valve and is available from AirCo Gases Company, 575 Mountain Avenue, Murray Hill, N.J. 07974, part no. WCS CV-4M. PSI tank gauge **19** is inserted in line **22** between back-flow valve **16** and flow pressure gauge **18**. PSI tank gauge **19** is available from AirCo Gases Company also, part no. WCS CV5M, and serves as an indicator of fuel remaining in the tank **14**. Flow pressure gauge **18** is used to monitor the flow of the hydrogen in pressure units. The flow is started using valve **16**. The optimal flow pressure is 25 pounds of flow pressure from the tank **14** to the cylinder of the engine.

The flow of hydrogen is regulated via valve **20** that is part of gauge **18**. Thus, when valve **20** is opened, hydrogen proceeds through line section **22** to hydrogen flow solenoid **26A**. The user can access the hydrogen fuel by flipping hydrogen flow switch **24A** into the "on" mode so that hydrogen gas in line **22** can proceed through hydrogen flow solenoid **26A** and into line section **28**.

Throttle **30** regulates the flow of hydrogen from line section **28** into line section **29** and ultimately, to engine **13**. Throttle **30** is connected to throttle pedal **50** via includes throttle cable **48**. Throttle pedal **50** also operates the butterfly valves in the carburetor in engine **13** in a conventional manner.

As an option, the throttle can be controlled electronically by a digital fuel controller product. Such a product is commercially available from Autotronic Controls Corp., 1490 Henry Brennan Drive, El Paso, Tex. 79936.

Line portion **29** proceeds through splitter **32** and is split into upper line **34** and lower line **36**. Lines **34** and **36** proceed to adapters **38**. Additional back-flow prevention valves **39** and **41** are mounted between lines **34** and **36** and adapters **38**. Braided Teflon® is the preferred material for the fuel lines of the system.

Battery **42** is connected via battery cable **43** to magneto **44**. Magneto **44** then controls the current through spark plug lines **46** to adapters **38** in a conventional manner. Thus, the regulation of the spark that ignites the fuel in the cylinder of engine **13** is conventional. Additionally, the intake of air into the engine **13** is accomplished through the standard carburetor system of engine **13**. In a car, the sparking of the spark plugs can be controlled by the standard distributor/battery ignition system.

A second solenoid **26**B controls the flow of gasoline from tank **23** through gasoline line **52**. Solenoid **26**B is controlled by switch **24**B. Throttle **30** is connected to gasoline line **52** and can also be used to control the flow of gasoline to engine **13**. Thus, this invention can be used with both an alternative fuel and gasoline.

Referring now to FIGS. **2**, **3** and **4**, the adapter **38** of this invention is shown in more detail. Adapter **38** includes housing **60**. A threaded port **62** is formed into housing **60** through port block **64**. Hydrogen lines are threadably connected to the adapter **38** via port **62**. Port **62** is also referred to as an entrance port. The entrance port has an entrance area **62***a*.

Adapter **38** also includes spark plug **66** which is effectively a standard spark plug except for the modification to the electrode as described below. Spark plug **66** includes a threaded end **68** which is threadably received in the threaded upper end **70** of housing **60**. Spark plug wires **46** (as seen in FIG. **1**) are connected to connector end **72** of spark plug base **66**.

Electrode **76** protrudes from spark plug **66** and is of a length sufficient so that the threaded tip **77** (best seen in FIGS. **3** and **4**) of the electrode **76** is proximate to the cylinder of the internal combustion engine when the adapter **38** is mounted. In one embodiment the electrode **76** is coterminus with an exit port **88** of the adapter **38**. This most clearly is shown in FIG. **4**. Electrode **76** is preferably platinum to enhance catalysis of combustion by-products, as will be more fully discussed below.

Adapter **38** also includes plug **79**. Plug **79** includes an axial internal void space **80**. Thus, plug **79** fits over electrode **76** via internal void space **80**. Plug **79** preferably comprises a ceramic material that acts as a good insulator. A suitable source for ceramic plug **79** is AM/PRO Machinery, Inc. 134 Church Street, Plain City, Ohio 43064-1321.

An important aspect of this invention is found on the external surface of plug **79**. An annular groove **82** is formed on the surface of plug **79** near its distal end **81** so as to be proximal to the port **62** where hydrogen enters housing **60**. Two stratifying grooves **84** proceed along the surface of plug **79** until they reach its distal end **83**. It is these mixing structures, i.e. annual groove **82** and stratifying grooves **84**, that facilitate the improved power generation from alternative fuels as provided by this invention. It will be apparent to those skilled in the analogous arts that mixing structure may be formed in the inner surface of the adapter housing **60**.

Alternatively a combination of vortex producing structures is integral with both the plug **79** and adapter housing **60**. Also any conventional vortex producing means should be sufficient groves, per se, are not required.

On threaded tip **77** of electrode **76**, there are mounted a spacer nut **78** and a spark gap nut **85**. As best seen in FIG. **2**, the spark gap nut **85** includes a protrusion **85***a* that provides the spark gap between electrode **76** and periphery **89** of the exit port **88** of housing **60**. The exit port **88** includes an exit area **88***a*. Preferably, the exit area **88***a* is no larger than the entrance area **62***a*, though this is not critical. Preferably, the spark gap nut **85** is brass and the preferred spark gap distance is $^{25}/_{1000}$ of an inch. Of course, the electrode may be coterminus with the plug distal end **83**.

As best seen in FIG. **4**, where arrows represent gas flow, hydrogen gas flows into chamber **60** via port **62**. The hydrogen gas cools electrode **76**, which facilitates combustion. Stratifying grooves **84**, preferably there are two, act to mix the hydrogen that is flowing into housing **60** with ambient air therein. Though more or less vortex creating grooves (stratifying grooves) may be desired. Typically this will depend on the relative dimensions of the grooves and diameter of the plug as well as the pressure or velocity of the mixture in the housing. The mixing occurs within void space **86** of housing **60** and creates a vortexing or tornado action. Plug **79** tapers as it proceeds from its distal end **81** to its proximal end **83**. Thus, proximal end **83** is smaller in diameter than distal end **81**. This provides additional space within void space **86** for the mixing of the hydrogen and air. The stratified hydrogen/air are twisting in a vortex fashion when leaving housing **60** via exit port **88**. The vortexing action produced by grooves **84** on the plug **79** creates a "tornado" **110** of hydrogen/air with a hollow center that generates a vacuum. Similar effect would result were t he grooves to be placed on the housing. The vacuum propels the hydrogen/air mixture into the cylinder of the engine for combustion. Such mixing cannot be accomplished in a conventional induction system.

Referring now to FIG. **5**, a schematic of the system of this invention is portrayed, with arrows representing gas flow. Hydrogen flows in through port **62** through adapter **38** and into cylinder **92**. Air enters cylinder **92** through air intake **90** and air intake valve **91**. The stratified hydrogen/air mixture is compressed in the compression stroke of the engine via piston **94** and push rod **96**, which are standard structures in an internal combustion engine. In the power stroke shown in FIG. **5**-C, a spark is generated in the compressed stratified hydrogen/air mixture ignites and drives piston **94** and push rod **96** to power the engine.

The mass energy density of hydrogen is enhanced through the increased pressure on the hydrogen exerted by piston **94** on the hydrogen/air mixture within the cylinder **92** in the uptake stroke in an internal combustion engine. The pressurized and well-mixed hydrogen/air mixture is rapidly ignited through the spark because hydrogen has a fast flame front. This reduces problems with timing. Thus, the energy store found in hydrogen is captured through this system via the compression of the hydrogen gas within cylinder **92** and via the thorough mixing of hydrogen and air provided by adapter **38**.

In addition, in the exhaust phase in a four cycle engine a second spark is generated from the plug to ensure complete combustion of by-products. This is enhanced by using an electrode made of platinum. Exhaust leaves cylinder **92** through exhaust port **98**. Exhaust port **98** is closed during other phases by exhaust outlet valve **100**.

In the exhaust stroke, because air valve **91** is closed, the substantial component of the gases in cylinder **92** is hydrogen. This is also accomplished due to the fact that in the system of this invention, the hydrogen is continuously flowing. When this spark interacts with what is predominately hydrogen, a plasma is formed. This plasma blocks back-flow of air from exhaust port **98**. Prevention of back-flow of air prevents the hydrogen fuel system from backfiring, which is another common problem of hydrogen fuel systems. In effect, then, the plasma formed by the hydrogen acts as a virtual valve in that the plasma blocks the back-flow of air.

When the plasma of the hydrogen is generated as described above, UV radiation is generated which also facilitates consumption of combustion by-products. This increases the environmental friendliness of the system.

In addition to hydrogen the following fuels can be used in this system: natural gas; alcohol; gasohol; propane and buthane. A mixture of hydrogen and natural gas can also be used. As noted above, gasoline can be used in this system when solenoid **26** is used to stop the flow of hydrogen.

Further, it is contemplated that an adapter that is described herein for use with an internal combustion engine can also be used with a home heating system such as a home heating system that may currently use natural gas. The vortexing action created by the plug would enable appropriate power to be captured from the hydrogen fuel so as to effectively heat a home.

Another embodiment of the invention is a hydrogen powered vehicle **200** shown in FIG. **7**. The hydrogen powered vehicle **200** includes an internal combustion engine **210** operably connected to power the vehicle **200** and adapted to connect to a hydrogen supply **212** through a supply line **214**.

Referring to FIG. **8**, the internal combustion engine **210** comprises a cylinder **92** operably connected to the supply line **214**. A piston **94** is operably disposed in the cylinder **92**. A spark plug **216** having an electrode assembly **218** (not shown in FIG. **8**) is in operable sparking communication with the cylinder **92**, wherein the spark plug **216** is capable of producing a spark **220** (shown in FIG. **10**) of at least 45 kilovolts. FIGS. **8**–**11** are substantially similar to FIGS. **5A**–**5D** depicting schematic representations of a 4-cycle piston-cylinder engine. Other piston and cylinder embodiments will be apparent to those with skill in the art. FIG. **8** represents an intake stroke. FIG. **9** depicts a compression stroke. FIG. **10** depicts a power stroke, in which the spark plug **216** sparks. FIG. **11** depicts an exhaust stroke, where the spark plug **216** is sparking. Refer to FIGS. **5A**–**5D** discussed earlier for further description of similar components of FIGS. **8**–**11**.

FIG. **12** shows one embodiment of the spark plug **216** in an elevated side view. FIG. **13** shows a plan view of the spark plug **216** along line **13**—**13** shown in FIG. **12**. FIG. **14** shows a plan view of the spark plug **216** in the direction of line **14**—**14** shown in FIG. **12**.

FIG. **15** depicts a section view of the spark plug **216**. FIG. **15**, section view, depicts a representative spark plug **216**. The spark plug shown in FIG. **15** comprises a housing **222** having a sparking end **224** and a hollow chamber **226** containing the electrode assembly **218**.

FIG. **16** depicts the electrode assembly **218** removed from the hollow chamber **226**. The electrode assembly **218** depicted in FIG. **16** comprises an electrode **228**. The electrode **228** has a connector end **230** and a sparking end **232**. As shown in FIG. **15** the sparking end **232** of the electrode **228** is positioned adjacent to the housing sparking end **224**. The electrode assembly further comprises a coil **234** electrically connected to the electrode connector end **230**.

Referring to FIGS. **10** and **11**, wherein the piston **94** moves through a power stroke in FIG. **10** and an exhaust stroke in FIG. **11**, the spark plug **216** is shown to fire during the power stroke and the exhaust stroke.

The electrode assembly **218** shown in FIG. **16** comprises a steel ball **236** positioned between the coil **234** and the electrode connector end **230**. Other operable ball-coil-electrode arrangements will be apparent to those of skill in the art.

The spark plug **216** shown in FIG. **15** also comprises an insulator **238** surrounding the electrode assembly **218** and positioned in the hollow chamber **226**. A ceramic pore **240** is shown stabilizing the insulator **238** in the hollow chamber **226**. Other methods of locating the insulator in the hollow chamber will be apparent to those of skill in the art.

An elevated section view of the housing **222** is shown in FIG. **17**. An elevated side view of the housing **222** is shown in FIG. **18**. In the exemplary embodiment of the housing **222** shown in FIG. **18**, the housing **222** comprises a fuel inlet port **242** in fluid communication with the supply line **214** (not shown). FIG. **19** depicts an exemplary embodiment of the insulator **238**. The insulator **238**, shown in FIG. **19**, comprises an outer surface **244** defining a first fuel channel **246** having a first respective fuel exit port **248**. The first fuel channel **246** is in fluid communication with the fuel inlet port **242**. The fuel exit port **248** exits into the cylinder **292**.

FIG. **20** shows a plan view of the insulator **228** shown in FIG. **19** along line **20**—**20**. FIG. **21** shows a section view of the insulator **228** shown in FIG. **19** viewed along section line **21**—**21**. FIG. **22** depicts a bottom plan view of the insulator **228** shown in FIG. **19** along line **22**—**22**. FIG. **23** shows a section view of the insulator **228** cut through section line **23**—**23** shown in FIG. **19**.

In the embodiment of the insulator **228** shown in FIGS. **19**–**22**, the insulator outer surface **244** defines a plurality of fuel channels **250**, including the first fuel channel **246**. The fuel channels **250** include a plurality of respective fuel exit ports **252**, including the first fuel exit port **248**. This is shown clearly in FIG. **22**. Preferably, the plurality of fuel channel exit ports **252** are spaced equally, or equi-distant, about the electrode **228**. This is shown well in FIGS. **21** and **22**.

Referring to FIGS. **15** and **17**, one embodiment of the housing **222** sparking and **224** comprises a torridal ground electrode **254** positioned about the electrode sparking end **232** (shown in FIG. **15**). Preferably the housing sparking end **224** and the electrode sparking end **232** are co-terminous. In one embodiment the first respective fuel exit port **248** is co-terminous with the housing sparking end **224** and the electrode sparking end **232**.

More generally the present invention includes an internal combustion engine **210** adapted to operate on hydrogen fuel supplied through a supply line **214**. Typically the engine comprises a first cylinder **92** in fluid communication with the supply line **214**; a first piston **94** disposed in the first cylinder **92** to move through a power stroke and an exhaust stroke (See FIGS. **10** and **11**) in the first cylinder **92**. The engine **210** also comprises a spark plug **216** having a housing **222** including a sparking end **224** and an electrode assembly **218** located in the housing **222**, wherein the sparking end **224** is in operable communication with the cylinder **92** and capable of producing a spark **220** of at least 45 kV. Preferably, the spark plug **216** fires during the power stroke and the exhaust stroke. In one embodiment the housing sparking end **224** comprises a torridal ground electrode **254**. The electrode assembly **218** comprises an electrode **228** having a sparking end **230** capable of sparking toward the torridal ground electrode **254**. Preferably the electrode assembly **218** comprises a coil **234** in electrical communication with the electrode **228**.

In one embodiment the electrode assembly **218** comprises an electrode **228** having a sparking end **230** and a coil **234** electrically connected to the electrode **228** opposite the electrode sparking end **230**. This is the embodiment shown in FIG. **15**. Typically the spark plug **216** comprises an insulator **228** surrounding the electrode assembly **218**. Preferably the insulator **228** includes a fuel channel **246** having a fuel exit port **248** near the electrode sparking end **232**.

It will be apparent to those with skill in the art that the present invention also comprises a method of operating an internal combustion engine **210** comprising the steps of supplying hydrogen fuel to the engine **210**; moving a piston **94** in a cylinder **92** through a power stroke and an exhaust stroke; creating a spark **220** in the cylinder **92** during the power stroke and the exhaust stroke, wherein the spark created during the power stroke is at least 45 kV; and supplying the fuel to a location sufficiently near the spark to ignite the fuel. This is shown well in FIGS. **10**, **15** and **19**.

The present invention also includes a method wherein the step of creating a spark **220** comprises passing an electric current (not shown) through an electrode assembly **218** including a coil **234** electrically connected to an electrode **218** having a sparking end **232**; insulating the electrode assembly **218**; and arcing the electric current from the electrode sparking end **232** to a ground electrode **254**. The ground electrode **254** as shown in FIGS. **15** and **17** a toroidal ground electrode **254**.

In certain embodiments the step of creating a spark **220** includes creating a spark **220** having at least 70 kV. Creating a spark of sufficient voltage will generally reduce the exhaust pollution. Typically this happens by generating a plasma or plasma field during the exhaust stroke. Thus, one method of the present invention further comprises the steps of generating plasma during the exhaust stroke; and allowing exhaust fuel converted to an exhaust state to coalesce with the plasma. This reduces exhaust pollution.

Thus, the apparatus and system of this invention provides an effective means to use alternative fuels. The use of alternative fuels can enhance the environment by reducing the use of conventional gasoline and the environmentally detrimental by-products caused by the consumption thereof.

Thus, although there have been described particular embodiments of the present invention of a new and useful "Hydrogen Powered Vehicle, Internal Combustion Engine, And Spark Plug For Use In Same", it is not intended that such references be construed as limitations upon the scope of this invention except as set forth in the following claims. Further, although there have been described certain dimensions used in the preferred embodiment, it is not intended that such dimensions be construed as limitations upon the scope of this invention except as set forth in the following claims.

What is claimed is:

1. A hydrogen powered vehicle including an internal combustion engine operably connected to power the vehicle and adapted to connect to a hydrogen supply through a supply line, the internal combustion engine comprising:

a cylinder operably connected to the supply line;

a piston operably disposed in the cylinder;

a spark plug having an electrode assembly in operable sparking communication with the cylinder, wherein the spark plug is capable of producing a spark of at least 45 kV, the spark plug including:

a housing having a sparking end and hollow chamber, the housing including a fuel inlet port in fluid communication with the supply line;

an electrode assembly, received in the hollow chamber and including an electrode having a connector end and a sparking end, the sparking end positioned near the housing sparking end;

an insulator surrounding the electrode assembly and positioned in the hollow chamber,

the insulator including an outer surface defining a first fuel channel having a first respective fuel exit port, wherein the first fuel channel is in fluid communication with the fuel inlet port and the fuel exit port exits in the cylinder.

2. The apparatus of claim **1**, wherein the piston moves through a power stroke and an exhaust stroke in the cylinder, and the spark plug fires during the power stroke and the exhaust stroke.

3. The apparatus of claim **2**, wherein:

the spark plug comprises a housing having a sparking end and hollow chamber containing the electrode assembly; and

the electrode assembly comprises an electrode having a connector end and a sparking end, the sparking end positioned adjacent to the housing sparking end and the electrode assembly further comprises a coil electrically connected to the electrode connector end.

4. The apparatus of claim **1**, wherein the electrode assembly comprises a steel ball positioned between a coil and the electrode connector end.

5. The apparatus of claim **3**, wherein the insulator outer surface defines a plurality of fuel channels, including the first fuel channel, wherein the fuel channels include a plurality of respective fuel exit ports, including the first respective fuel exit port.

6. The apparatus of claim **5**, wherein the plurality of fuel channel exit ports are spaced equally about the electrode.

7. The apparatus of claim **6**, wherein the housing sparking end comprises a toroidal ground electrode positioned about the electrode sparking end, and wherein the housing sparking end and the electrode sparking end are co-terminous.

8. The apparatus of claim **7**, wherein the housing sparking end comprises a toroidal ground electrode positioned about the electrode sparking end, and wherein the housing sparking end, the electrode sparking end, and the first respective fuel exit port are co-terminous.

9. An internal combustion engine adapted to operate on hydrogen fuel supplied through a supply line, the engine comprising:

a first cylinder in fluid communication with the supply line;

a first piston operably disposed in the first cylinder to move through a power stroke and an exhaust stroke in the first cylinder;

a spark plug having housing including a sparking end, and an electrode assembly located in the housing, wherein the sparking end is in operable communication with the cylinder and capable of producing a spark of at least 45 kV; and wherein

the spark plug fires during the power stroke and the exhaust stroke;

the spark plus comprises an insulator located in the housing around the electrode assembly;

the housing comprises a fuel inlet port in fluid communication with the supply line; and

the insulator includes a first fuel channel in fluid communication with the fuel inlet port, the first fuel channel having a respective first fuel exit port exiting into the cylinder.

10. The device of claim **9**, wherein the housing sparking end comprises a toroidal ground electrode, and wherein the electrode assembly comprises an electrode having a sparking end capable of sparking toward the toroidal ground electrode.

11. The device of claim **10**, wherein the electrode assembly comprises a coil in electrical communication with electrode.

12. The device of claim **11**, wherein the electrode sparking end, the toroidal ground electrode, and the first respective fuel exit port are co-terminus.

13. The device of claim **12**, wherein:

the electrode assembly comprises an electrode having a sparking end, and a coil electrically connected to the electrode opposite the electrode sparking end.

* * * * *

rexresearch.com

Hermann PLAUSON

Atmospheric Electrical Power (II)

Supplements PLAUSON, H. : Conversion of Atmsopheric Electricity

Related : Atmospheric / RF Electricity Collectors ** GUILLOT : Atmospheric Electric Generator ** JEFIMENKO : Electrostastic Motors ** ORLANDO : Atmospheric Electrical Generator ** TATE : Ambient Power Module ** Electrostatic Motors & Generators ** PERRIGO : Generator ** TENTZERIS : Ambient Energy Antenna ** WADLE : Tree Electricity **

---

https://en.wikipedia.org/wiki/Hermann_Plauson

Hermann Plauson

Hermann Plauson was an Estonian professor, engineer and inventor. Plauson investigated, among other things, the production of power based on atmospheric electricity.
Biography

Plauson was the director of the Fischer-Tropsch "Otto Traun Research Laboratories" in Hamburg, Germany during the Weimar Republic of the 1920s. He built on Nikola Tesla's idea for connecting machinery to the "wheelwork of nature". Plauson's patent no. 1540998 describe methods to convert alternating radiant static electricity into rectified continuous current pulses. His systems mainly involved needle & radium- coated electricity conducting balloons or aerostats which were connected to complex converting systems consisting of coils, capacitors, spark gaps etc. He developed an electrostatic generator called Plauson's converter. In 1920, Plauson published a research journal on the subject, titled "Production and Utilization of the Atmospheric Electricity" (Gr., Gewinnung und Verwertung der Atmosphärischen Elektrizität). A second, expanded edition of his journal was published later the same year. A copy of that edition is in the British Library at shelf mark 8754.b.36. A copy is also available in PDF-format, complete with Optical Character Recognition, through the books2ebooks service, here.

It is believed that Plauson was related to Gertrud Plauson (the exact relationship is unknown; she may have been his wife)...

Select Patents

https://www.google.com/patents/US1485706
US1485706
Method of carrying out electrochemical reactions and apparatus for the use therein

https://www.google.com/patents/US1540998
USP 1540998
Conversion of atmospheric electric energy

http://www.fischer-tropsch.org/primary_documents/patents/GB/gb309002.pdf
GB309002
Use of Beta-rays and X-rays in Synthesizing Liquid Hydro-carbons From Gaseous Compounds

Liquid hydrocarbons are synthesized by subjecting gaseous carbon compounds; such as are contained in water gas, coke oven gas, natural gas and gases produced by cracking heavy oils, to the action of rapidly moving electrons (#-rays) or the joint action of #-rays and X-rays. The reaction may be assisted by rotating electric and/or electromagmetic field and by catalyst in suspension in the reacting gases. The velocity of the rays should exceed one half the velocity of light and is suitably 200,000 to 260,000 kilometres per sec. The gases after exposure to the rays remain in a reactive condition the effect of which is obtained by circulating the gases under pressure. In addition to hydrocarbons, ketones, alcohols &c. are obtained particularly if the gases contain water vapour. Higher boiling point hydrocarbons may be obtained by subjecting those of lower boiling point to the radiations. According to an example water gas after subjection to radiations is compressed to 800-1000 atmospheres and maintained at this pressure for a time depending on conditions of radiation, temperature and the type of hydrocarbon to be produced. The hydrocarbon produced is mainly of saturated nature. The addition of ethylene, acetylene or other unsaturated hydrocarbons to the water gas shortens the process.

http://v3.espacenet.com/textdoc?DB=EPODOC&IDX=GB157263&F=0
Process and apparatus for converting static atmospheric electrical energy into dynamic electrical energy of any suitable high periodicity
GB157263

http://v3.espacenet.com/textdoc?DB=EPODOC&IDX=GB157262&F=0
Improvements in electric motors
GB157262

External articles

http://www.nuenergy.org/alt/GernsbackOnPlausonFebruary1922.htm

"Power from the Air". Science and Invention (Formerly Electrical Experimenter), Feb. 1922, no. 10. Vol IX, Whole No. 106

http://www.nuenergy.org/alt/PlausonMarch1922.htm

"Power from the Air". Science and invention (March 1922)

---

http://www.meridian-int-res.com/Energy/Atmospheric.htm

Historical Research

The leader in this field before the Second World War appears to have been Dr Hermann Plauson. Dr Plauson was an Estonian citizen who lived in Hamburg and Switzerland. He carried out experiments in Finland with aerostats manufactured from magnesium-aluminium alloy, covered with electrolytically deposited needles. The needles were further doped with a radium compound to increase local ionisation of the air. (This was the era in which the hands of watches were hand painted with radium to make them luminous in the dark). Zinc amalgam patches were also painted onto the aerostats. Plauson obtained a power output of between 0.72kW and 3.4kW from one and two aerostats 300m above ground level. Dr Plauson filed patents in the USA, Great Britain and Germany in the 1920s. His book "Gewinnung und Verwertung der Atmosphärischen Elektrizität" is the most detailed known account of the technology.

Other atmospheric electricity researchers contemporary to Dr Plauson included Walter Pennock and MW Dewey in the USA, Andor Palencsar in Hungary and Dr Heinrich Rudolph in Germany. Hippolyte Charles Vion in Paris predated them all, putting forward proposals in the 1850s and 1860s.

Heinrich Rudolph made an interesting contribution to the design of the aerostat collectors. In 1898 he designed an elliptical aerostat made up of faceted surfaces to minimise the effect of wind. The design bears a strong resemblance to Northrop's 2003 UCARS unamnned helicopter UAV project. The design uses the Coanda Effect to help keep the aerostat on station and minimise wind effects.

MIR's Research Programme

Since 1997 we have been carrying out theoretical research into conversion of atmospheric electricity into useable electrical power.

From a low level (5m high) simple zinc antenna we are able to obtain sufficient charge to light a number of white power LEDs. Further experimental investigations with metallic aerostat collectors and cavity resonant slow wave antennae concepts are ongoing.

A number of programmes are underway in developing countries to provide remote communities with LED lighting instead of conventional incandescent lighting. LEDs are a very efficient way to provide lighting for minimal electrical power requirements. Atmospheric electricity would be ideally suited to powering domestic LED lighting or low energy neon lighting.

There is every reason to believe from our experimental tests and earlier research by Plauson, Jefimenko and others that Atmospheric Electricity will be able to provide practically useful levels of electrical power particularly suited for off-grid applications in the developing world.

Advantages of Atmospheric Electricity

Simple and robust technology
Low Cost technology - much cheaper than photovoltaics or wind turbines
Available day and night in all weather conditions - in fact, more power is produced at night than during the day
Available at any point on the Earth's surface...

---

https://books.google.com/books/about/Gewinnung_und_Verwertung_der_atmosph%C3%A4ri.html?

id=t35yQwAACAAJ

Gewinnung und Verwertung der atmosphärischen Elektrizität, Beitrag zur Kenntnis ihrer Sammlung, Umwandlung und Verwendung...

by Hermann Plauson,...

Beysen und Maasch, 1920 - 75 pages

[PDF]

---

http://www.nuenergy.org/gewinnung-und-verwertung-der-atmospharischen-elektrizitat-deutsch/

Gewinnung und Verwertung der atmosphärischen Elektrizität

Beitrag zur Kenntnis ihrer Sammlung, Umwandlung und Verwendung Mit 82 Figuren auf 22 Tafeln

HERMANN PLAUSON, Harmburg 1920

Vorwort.

Motto: Wem es gelang in ein Geheimnis der Natur zu dringen Und dies in Menschenpflicht der Menschheit darzubringen, Der betet die Natur und ihrer Wunder Fülle an.Verfasser.

Es ist schwierig und wenig dankbar, ein Buch über ‘ein Gebiet zu schreiben, das bis heute fast noch gänzlich unberührt ist.

Obwohl die Lehre über atmosphärische Elektrizität an und für sich schon in gewissen Kreisen bekannt und einigermaßendurchforscht ist, auch sogar einige theoretische Abhandlungen über dieses Gebiet geschrieben sind, so ist umgekehrt auf dem Gebiete der technischen Gewinnung, Umwandlung und Verwertung von atmosphärischer Elektrizität bis jetzt nochnichts bekannt. Es kann sogar festgestellt werden, daß einige Wissenschaftler eine Gewinnung und Verwertung deratmosphärischen Elektrizität in der Praxis überhaupt bestreiten.

Wenn nun trotzdem der Verfasser dieses Wagnis unternommen hat, so trieb ihn vor allen Dingen die außerordentlicheWichtigkeit der Lösung dieser Frage für die Menscheit hierzu und dann die auf Grund von Versuchen erhaltenepersönliche Überzeugung, daß bei Verwendung aller wissenschaftlichen Errungenschaften und Forschungsergebnisse derletzten Jahre auf dem Gebiete der atmosphärischen und statischen Elektrizität, bei Benutzung der Erfahrungen auf demGebiete der drahtlosen Telegraphie und Telephonie, wie der Erzeugung von ungedämpften Schwingungen überhaupt,ferner unter Zugrundelegung der Lehre über radioaktive’Erscheinungen und deren Folgen viel größere Energiemengengesammelt werden können, als bisher nach den theoretischen Berechnungen zu erwarten war, die auf Grund desnormalen, vertikalen Leitungsstromes gelten die Erde, sowie auf Grund der Gesamtladung der Erde angestellt wurden.

s wurde vom Verfasser festgestellt, daß dies erreiCht wird sowohl durch Erhebung voll metallischen Ballons mitradioaktiven Substanzen und elektrolytisch hergestellten Nadeln usw., als auch durch geeignete weitere Verwendung undVmwandlungsart dieses Stromes (Umwandlung statischer Elektrizität in mehr oder weniger hochschwingendedynamische) sowie durch Verwendung geeigneter Kondensatorenbatterien und dereff Einschaltungsweise und zuletztdurch geeignete Schaltungsschemata. Versuche ergabeff aber, daß noch zahlreiche neue Vorfragen zu lösen waren, und datl der Ausbau des Problems, Verwertung der atmosphärischen Elektrizität, auf ganz neuen Grundlagen erfolgen mußte. Der Verfasser hofft durch diese Schrift den Beweis zu erbringen, daß er auch dieses erreichtoder wenigstens die Wege dazu angegeben hat.

Das zweite Ziel des Verfassers ist, durch baldiges Erscheinen dieses Buches breite Kreise der Wissenschaft und Elektrotechnik fürdiese Frage zu interessieren, und durch die im Buche angegebenen Anregungen möglichst viele Mitarbeiter zu gewinnen.

Es sei ferner noch gesagt, daß das Buch nicht eine abgeschlossene Forschungsarbeit zu sein beansprucht, sondern vielmehr als ersterVersuch, als Wegweiser in der vorliegenden Frage zu betrachten ist. Es bleibt sowohl in der Schreibart als auch in der Anordnung desMaterials noch viel zu wünschen übrig, da das Buch zum Teil aus Patentschriften des Verfassers entstanden ist, zum Teil aus einemVortrag über atmosphärische Elektrizität, den der Verfasser vor einer Versammlung von Fachleuten gehalten hat, um den Bau einerMusteranlage zu propagieren.

Veranlaßt durch das rege Interesse größerer Kreise für diese Frage, läßt der Verfasser dieses Buch sich seinen Weg selbst bahnen undbittet, etwa infolge der Eile sich fühlbar machenden Mängeln gegenüber nachsichtig zu sein.

Bei der Zusammenstellung dieses Buches wurde folgende Literaturbenutzt:

Mache & Schweidler: “Die atmosphärische Elektrizität”Dr. I. Zenneck: “Elektromagnetische.Schwingungen und drahtlose Telegraphie”
Dr. M. Dieckmann: “Experimentelle Untersuchungen aus dem Grenzgebietzwischen drahtloser Telegraphie und Luftelektrizität”I. Teil Teil (Zeitschrift fÜr Luftfahrt und Wissenschaft,2. Heft).

Für Leser, die dieses Gebiet nicht völlig beherrschen oder Einzelhejten darüber wissen wollen, sei zum besseren Verständnis dieserAbhandlung empfohlen, vorgenanntes Werk: Mache & Schweidler .Die “atmosphärischeElektrizität”, VerlagVieweg & Sohn,Braunschweig, durchzulesen.

Es ist dem Verfasser zum Schluß eine angenehme Pflicht, Herrn H. Otto Traun für die ihm erwiesenen Unterstützungen seineDankbarkeit auch an dieser Stelle auszudrücken.

Ferner möchte der Verfasser nichtversäumen, Herrn Max Thorn zu danken für die erste Anregung, die gesammelten Erfahrungen undKenntnisse der Allgemeinheit nicht länger vorzuenthalten, und für die Bemühungen zur Finanzierung einer hoffentlich irn Jahre 1920 zu erbauenden Versuchsanlage, wodurch dieWichtigkeit dieses Problems irn Großen praktisch vor Augen geführt werden soll.

Schließlich spricht der Verfasser Herrn Dr.-Ing. Gerhard Schmitt und seiner Assistentin, Fräulein Gertrud Hildenbrandt,für freundliche Übernahme der Korrektur, sowie Herrn Ing. Hans Koop für Ausfertigung der Zeichnungen an dieser Stelleseinen wärmsten Dank aus.

Der Verlagsbuchhandlung sei für die anerkennenswerte Ausstattung des Buches und besonders für die guteAusführungder Figurentafeln hiermit der besondere Dank des Verfassers ausgedrückt.

Der Verfasser.
Harnburg, irn Oktober 1919

Inhalt.

Allgemeiner Teil.
Historische Entwicklung
Neuere Literatur; Patente und deren Beurteilung

Ergebnisse der Vorversuche des Verfassers
Vergleich der Sammlung der atmosphärischen Elektrizität mit der der Sonnen energie
Grundprinzip der Umwandlung statischer .atmosphärischer Elektrizität in elektromanetische Schwingungen
Prinzip der Umwandlung elektromagnetischer Schwingungen in mechanische Energie
Resonanzmotoren und ihre Wirkungsweise
Schaltungsschemata für die gewonnene atmosphärische Elektrizität
Erzeugung von Schwingungen verschiedener Dämpfung
Unipolare Einschaltung der Resonanzmotoren
Antennensysteme zur Gewinnung der atmosphärischen Elektrizität Kondensatormotoren
Unipolare Einsch.ltung der Kondensatorenbatterien
Sammlerballons
Transformatoren zur Umwandlung von statischem Gleichstrom In gewöhnliche Wechselströme
Weitere Art der Erzeugung von Wechselströmen hoher Wechselz.hl
Leitungsschemata für hochschwingende Ströme
Wirtschaftlicher Teil.
Entwicklungsmöglichkeiten der Erflndung
Ausnutzbare Energiemengen
Theorelische Unterlagen
Elektrostatische Felder
Ionisation
Veftikaler Leitungsstrom
Gesamtldung der Erde
Entstehungsursachen des gewonnenen Gesamtstromes
usgleichstrom
ollektorstrom
Hallwachsstrom
Wirtschaftliche Ausblicke
Kostenanschlag für eine Anlage
entabilltät
Verhütung von B1itzschäden
Schlußwort

Partial machine translation (uncorrected)

http://peswiki.com/plauson-book-01

General part

The production and utilization of atmospheric electricity for the well-being and prospering of mankind is a goal of scholars and researchers; the implementation remained however until today a pious desire, and it did not succeed to somewhat really reach [WertvoJles] and for industrial purposes usable.

The cause for this is not to be looked for in the fact that the electricity freely available in atmospheric air is not to be exploited at all, but primarily in that the science before solution of this problem had to solve a set from preliminary questions to, by their clarifying only the possibility of an utilization of atmospheric electricity became at all conceivable.

Only in the last years their utilization could be made due to new experiences and investigations on that areas of the atmospheric electricity.

If we try many ages [zurückwand~rn] and to investigate the past according to data concerning atmospheric electricity then we must notice with being astonished the fact that nothing is new in the world that for a long time, for a very long time before the knowledge of the electricity at all admits the features of the atmospheric electricity was and that they, what appears still less [glaublich] were even used.

The knowledge of the effect and utilization of atmospheric electricity hand back already many thousand years, as we can infer from the holy writing. Thus we read all geweihten persons, with exception of the upper priest in the Bible Moses that the tabernacle in all-holiest of the temple Jehova possessed the power to kill when affecting by a thunderbolt or a fire. Furthermore we read that 40 priests had dared enter in the absence of Moses and Aaron, and that from the tabernacle lightning drove out, which killed all 40 priests. If we experience now far from the books of the kings the design of the new temple by king Salomon, then from the electrotechnical point of view can only come only to a conclusion that Moses and Aaron knew already at that time as one could collect an unknown godlike power from the tabernacle, as one had to [aurbewahren] the same, and as it worked, because of-today's point of view of the Wissenschart regarded the federal drawer can have been nothing else as a Leyden bottle or a condenser of very large capacity, which by means of arranged points was loaded with atmospheric electricity . The fact that the charge kept itself good and worked well is only the constant shop with atmospheric electricity to attribute as well as the dry climate in Palestine.

It comes out from the Biblical description that the ssundeslade out well isolationable noble wood had fitted made and also gold and such a thing from the inside and outside. Thereby all conditions were erffillt rur a good electrical condenser or a Leyden jar.

E.G. the gentleman writes the Moses forwards (2. Book Moses, Chap.25, verse 10-15)

Furthermore still is given the description of the building in the Chap.37-38. An electrical engineer could not make it better nowadays, wanted it follows a Leyden jar to build for the purpose, humans a [göttliche Kraft] the Ungeweihte to kill immediately can to lead before eyes or to pretend.

The charge of the federal drawer with atmospheric electricity in the all-holiest was obtained by many columns and golden points. In addition, the whole temple was, as from the description hervorgehl, provided with hundreds of gilded points, by presenting gold on zedern and acacia wood bars.

03

The fact that here really electrical discharges took place can do out following in 2. Book of the Chronicles, Chap.7, verse 3, described place to be taken DAuch saw falling down all children Israel the fire and the glory of the gentleman over the house, and fell to their knees, with the face to the earth and prayed on and thanked the sir, it good-naturedly are eternal and its mercy lich last. “People saw here the impact of lightning into the lightning conductor established on the temple. DI1ss the collection and Aufbewl1hrung of the atmospheric

Electricity much gefäh " to be rlich could, if someone did not understand with the federal drawer to deal, read 'we further in the 3. Book Moses, chapter 10, where the sons Aarons, Nl1dl1b and Ahibu wanted to bring, d~m Mr. Freudenfeuer, what the Mr. them not ordered hatteWir to read that a fire proceeded from the gentleman and it verzehrte, so that they died before the gentleman.

From everything it is evident that Moses were and its contemporaries the first connoisseurs and Ausnutzer of the atmospheric electricity. Were not natural them, how us, which admits to electrical laws separate only the effects of these mystischen Kraft. These features are probable the culturally more highly standing Egyptians even already admit been, and Moses may have bring along its knowledge of the priests from Egypt.

From this time to the attempts Frl1nklins with BJitz the production of atmospheric electricity made only small progress for arresters and Rimans attempts. With the study of the meteorology the large interest in the atmospheric electricity was again excited, and in the last 30 years several times the thought stepped to light to use it. Which was already done in this regard by other researchers, I try to describe in the connection hieran briefly, by lighting up this question I1m of best patents spent through in former times.

The first patent, that according to indication German realm Reich office, is that is given D.R.P. No. 98180 of the t9. January 1897. Inventor: Dr. Heinrich Rudolph in Stc Goarshausen A. RH. The Patentl1nspruch reads: Your by a kite chain balloon carried, from two steel tubes and many vertically stretched fine wires manufactured, large and nevertheless extraordinarily light net for the collection of atmospheric electricity together with for installing and for the acceptance of this net, occupied with needles I'rüher announced patents and their evaluation.

04

necessary mechanism, consisting of a number of posts with ever two axle camps and some small electric motors. “

Description.

Net for collecting atmospheric electricity.

The subject of available invention forms, with approximately 3.6 million needles an occupied, ungefäh~ 9 hectars large and net which can be held by a kite chain balloon in 1000~2000 rn heavy against 700 kg for height for accumulating atmospheric electricity together with the mechanisms erforderiichen for isolated derivative For the clarification of the state of affairs is emphasized that those.

For the clarification of the state of affairs is emphasized that those Franklin attempts the possibility of the derivative of atmospheric electricity for the earth long did that it concerns however with the voriiegenden invention particularly around the described mechanism, such, which suggests a commercially usable amount of energy of the derived atmospheric electricity. Which concerns the aids to the commercial utilization of the net described in being inferior, then is referred to the patent No. 98288 standing with available article in closest Zusammenhange. In the enclosed design a such net is together with the kite chain balloon in Fig. 1 in front view represented, Fig. a side view shows 2, while the Fig. 3-5 details represent.

A such net N (Fig. 1) can be manufactured by means of two ungerahr 300 rn long casting steel tubes R R, by 1 mm of wall thickness and 2 cm diameters, which will receive the same by Ineinanderschieben of individual pipes from production length and fastening by small steel pins with screwed on safeguard panels. These pipes are to form the two horizontal sides of the square net, between which about 1200 vertical bronze wires of 300 are to be stretched rn length and 0.04 qmm cross section in distances of 0,25 rn. Since rn much purifies the thin wires in distances of 0,1 and length sharpens needles of 1 cm carry must, then they are wound meanwhile before slowly on Rol) EN and the needles with the help of a springy eye open at their lower end are touched down. The eye must selbs (understandably the wire size so adapted to be that it wedges deQ wire and that the needles sit sufficient firmly. Also rolling up must happen sufficient loosely, so that the put on needles do not break. The main difficulty such a square network of 300 rn side length offers an installing and a removing the same, the whatever latter in Ausnahmerallen becomes necessary - can. Is necessary for it a number of man-high posts (p), in

05

Rn lineally and all two axle camps I are one behind the other set up to distances from 7 to 8 possess (Fig. 3). Built up on the same the two 300 rn are enough for Slahlrohre R R, and provided with the necessary hooks, by lelztere safeguard panels small by boreholes of the pipes put, and become screwed on!? n. Whereupon wires are to be stretched from Pfoslen to posts, on which the necessary number of roles with the needle-occupied bronze wires is. The ends of the same are raised to the hooks a Slahlrohres geknlipfl and the latter by the Fesselballon, which is to carry the net, so that the bronze wires of the roles complete themselves. So that the dlinne steel tube no bend experience, it is carried by 40 Gussslahldrähten, which hang on the hook for the chain cable and branch out toward their lower end three-way, so that they can be befesligt in distances of approximately 2.5 rn at the mentioned pipe. Dieseiben are in their length like that abzugleicheil that they form approximately an equal-leg triangle of 250 rn for height, whose basis is the pipe carried by them also they need to have only 1 qmm cross section against the center, during toward the ends of the pipe a Quersc--hnitt of 2 qmm necessarily isl.

If the needle-occupied bronze wires are completed, then they become to the zweile steel tube geknüpfl. After 31 bronze wires of 0,13 qmm Querschnilt without NadelQ, further on the opposite side of this pipe in distances of 10 rn, are beftlstigt for the derivative of the Elektrizilät, the net so because is lifted, until about 1000 rn the chain cable by the hoist of an anchor car it is completed, on which to the cable it fastens the bronze wires jointly but becomes only so loosely strained that the net down-hangs almost vertically, with its surface senkrechI the Windrichlung. If the net is to be removed again, then approximately in reverse way, only with the difference that thereby small electric motors at each second post step into activity, those will proceed by wheels with a removable Cutout, by which they can be touched down easily on the lower pipe, the latter in revolution to shift and the needle-occupied wires to the same to complete.

It can occur however the case that for the acceptance of the net in this way sufficient Zeil does not remain, if approaching storm with a larger wind force than 25 rn the rapid salvage of the Fesselballons behind a wind protection with the help of, the electrically operated Anke running on rails " rwagens makes advisable. Then \ ias net on a sufficient even must Place easily on the ground to be established,

06

similarly as one things materials in folds puts, then it could be lifted later again intact into air.

In order to be able to notice also a large wind force in 1000 to 2000 rn height with relative peace at the soil, a Federdynamometer put onable in any place without detachment of the chain cable must steadily the tension of the latter anteigen. If such a case of strong upper wind arises, then the bare short binding is naturally already sufficient to the safety device of the balloon and net

Which concerns the potential of the atmospheric electricity, then a potential gradient of 150 V for I is rn as always available to regard after numerous observations. For 1500 rn height would result in that 225000 V. If one wants to charge a battery with the received river from 20000 one behind the other switched accumulators, then it requires a charging voltage of 45-50000 V, for which straight still the isolation can be accomplished. From it a voltage drop of 175000 V in the environment of the net, which after the observations over the conductivity of air with grösserern potential gradient a sufficient river secures, results particularly due to the 3.6 million of fine points of the net and the constant continuation of the unloading air layers from the wind. So that the electricity collected by the net can be derived now in addition, isolated, about 100 rn a long silk rope T by sixfold security, which is protected from wetness by an india rubber casing, must be switched on, into the chain cable F, and, be led the electricity by a special cable G of 5 qmm for casting steel cross section and 2 qmm copper soul downward below the place, where the bronze wires are fastened to the chain cable.

While now the cable G after a well embodied, hollow iron mast M, is the chain cable leads to the electrically operated hoist Weines of anchor car fastened, which, according to kind the mobile crane with swivelling upper rack, provides jib and counterweight, on a circular rail gel ice with approximately 20 rn radius around the mentioned hollow mast to move can, so that it is able to follow each change of the wind direction. The jib of the anchor car serves to keep the lower horizontal part so high the Seilkurve described by the Fesselballon over that ground that it does not schJeift.

It is necessary that cables and chain cables are from each other distant in air always far for enough, in order to prevent an upper jumping of the high voltage stream. That is possible by a tauter tension of the easier Lejtungska~~ls. DeshaJb goes this, after it by a slip ring G carried by e1nem Olisolator is led, by an isolating india rubber pipe, by about 10 rn high mast down to a hand winch w present under oil, whose drum provided with steel axle is formed by porcelain or ebonite, so that one can with each change of the length of the chain cable for the case of zero wind or storm also the cable accordingly stop and again stretch

From the hollow mast M out the line goes underground through one. Channel up oil insulators under the gel ice for the anchor car away and then as high voltage air line after the battery station, which can be established also for several balloon stations together.

There a double battery of 20000 small accumulators each is to be set up, from which always a battery delivers river, while the other one is loaded. For this purpose one must be able to switch everyone one behind the other into approximately 300 groups at will neben¬ and. The accumulators must rest by groups on racks, which are carried by oil insulators.

From apparatuses are on a battery station except usual flat TAP-ready necessarily:

Erdschtuss with simultaneous interruption of the line to Battery.

Adjustable spark gap as Nebenstromkreis with ground fault to the announcement of interruptions.

fuses for the case of Kurzschi1uss of the battery.

Ammeter.

Pendulum electroscopes.

Electrostatic one dares to the innocuous making of line interruptions, by the same with exceed more eiper certain tension ground fault one manufactures and by a electromagnet in the ground conductor the line to the battery one interrupts. {In Fig. 4 is the electrostatic dare omitted.)

automatic commutator (Fig. 5) for the charging current, be standing from electromagnets (omitted in the design), which brings two levers with isolated connection to the

high voltage air line and to the earth with Poland the battery depending upon current direction in or to other sequence in connection.

The last device is particularly important for the utilization of the extraordinary, but sign after rapidly changing electrical tensions of the thunderstorm and heap clouds, that large Konduk gates and condensers of the atmospheric electricity.

By pie 3.6 million points of the net thereby the plant appears secured against each destructive spark discharge All apparatuses except 2, 4 " and 5 must have contacts from platinum and for preventing of sparking and fusion under oil are. All work on the high voltage transmission line before by means of laliger ebonite seaweeds ground fault is made, whereby for preventing of short-circuit the line must be interrupted at the same time to the battery thus the tension sinks on zero. On the balloon stations are necessary: the apparatuses I, 4, 5 and 6, the latters however without dare and only with electromagnet in that Ground conductor to the Unterbl echung (Fig.4). The Schienerigeleise must on. the ground conductor of the balloon station put its, füi' the case that once the cable puts over the anchor car or the chain cable affected.

If one takes now managing patent application under the magnifying glass of the criticism, then it must be stated that the Erfindungsgedanke for the following reasons cannot be lebensflihig:

The balloons must possess very large volume, since they are to carry d1e of large heavy nets and the strong rope. The latters must be strong, so that they are not torn up by the storm.

The expansion of such a plant over large surfaces - z. B. IQO qkm - is therefore: A) commercially impossible, because a great many balloons of this kind were needed, which would be enormously expensive, and B) practically not executably, because one cannot secure such large balloons in large heights sufficient before storm also the same, there made of things, ground very fast leaky. From it large dangers for humans, the result if such a balloon tear off and from leak from on earth fall, became a Kurzschussnt stand, the tremendous disaster, which death of many humans could entail.

For above reasons already this patent is not usable nevertheless is it interesting because of the attempt, which so collected electricity to the charge of a double battery of over 20000 small accumulators, from which the energy delivers, while the other one is loaded. Such a system would be the most ideal solution of the Transforming of atmospheric electricity into low tension usable for the technology. Because of the isolation difficulties it is however unfortunately impossible to load 20000 elements with approximately 50000 V 1 more user that the initial costs of the 40000 elements would be too expensive, completely apart from the fact that the maintenance of these batteries do not only require many people, but also with large danger for the supervision personnel it would be connected.

He same wants to obtain by the following patent claims:

procedures for the Nutzbarinachung of Atmospheric electricity, thereby characterized that the same by means of a be¬wegten, by a g~eignete HeizqueJ] e of catching body heated up it is caught and trans¬formiert by means of a rheostatischen machine, how to the charge of accumulators the Electricity thereby on constant Spal! nung to be held that the rheostatische Machine or a number of their plates is connected with a charge measurer (electrometer), that, as soon as dte rheostatische machine was loaded on a certain potential, knows by the excursion of its mobile part a contact closes, whereby, under switching of a relay, an electromagnet is perhaps brought to responding, the change-over of the rheostatischen machine effectuation

Device for the execution of the procedure, consisting of a double-walled LuftbaJ] on (1, ge¬kennzeichneten in requirement 1., 2), which by a MetaJ] net provided with needles is covered and when the Auf¬ and descending by tiltable shovels (9) in revolution one shifts, from a Hejz¬rohr (12), which is through tubes with upper and lower parts of the area (II), formed by the Dop~elwänden of the balloon, connected, aus' mjt the wire net in leading Ver¬bindung standing ball bearing (14), which with the taps des¬seiben leading connected cable (13), that with the KoJ] ek¬tor (17) provided winch (16) and the rheostatischen machine (18).

3. During the device characterized in requirement

2. the application of a Electromagnet, which is excited by a charge measurer (Elektro¬meter), connected with the rheostatischen machine, to the change-over of the rheostatischen machine, as soon as the potential achieved the same in-certain measure ".

Patent specification:Procedure and device for catching atmospheric electricity

Available invention trilft a procedure and a device for catching and deriving atmospheric Electricity, which make it possible, larger quantities of the same in usable: to receive it form.

The Erfindungsgedanke is based on your theory of the atmospheric electricity, after the the same by the condensation of

10

One forms for water vapours. According to this theory now the water droplets floating in the air layers are to be regarded as carriers of the electricity, and a rational system of the derivative of the atmospheric electricity must the same now of the water droplets ab¬leiten. This is way-reached after available invention on the following

With pointed needles of provided catching bodies of mög lchst large surface becomes into higher air-laminated? moved, however by a suitable heating device continual warms up. By the heat water droplets become! dey the catching bodies directly surrounded, the capacity the same evaporates decreases aJI¬mählich, and thus the tension of the charge rises.

It is now easy evidently that the whole charge of the droplets, as soon as the same evaporated, on which catching body Über¬gehen becomes, from which it be derived can. U m however the evaporated and their charge robbed droplets by new and loaded to replace, becomes the catching body the surrounding enclosed design schematically represented, in Pig. 6 and 7 by the opinion and the cut of the catching balloon. Pig. the deriving device, Pig.9 the regulation device, pi shows 8! (. 10 and II a shovel in their two positions, Pig. 12 and 13 the circuit way of the rheostatischen machine in their two air in relative motion shifts. A remark example of the invention article is up positions.

The pre-adjust consists of a double-walled balloon I, 2, the light wire nets, favourably from aluminum wire, occupied with one with needles, covered is. The balloon carries also the net 3, on which the ring 4 made from firm, but light material (wood, pipe etc.) is fastened. The latter carries the basket 6 by means of the cords 5. In the level of the ring 4 still the ring 7, that is by those easily swivelling gelagenen shovels (9) spread and by dle cords 8 in its position held w! about SI haufeln of covered framework with light material, whose turn in both directions is limited by the notices 10, consists. The shovels form with the vertical one appropriately an angle from 60 to 70 degrees. All wire nets, cords, rings and shovels can be covered? with small Metallna'deln, which are connected among themselves leading. Of the area II fühn of the deepest point the pipe 26, formed by dje Doppel¬wände of the balloon, too, which in the 'basket 6 untergebrach1en coiled tube 12, which is again with your in the upper Tei, 1 of the intermediate area II connected to flowing pipe 13. The coiled tube: becomes by means of a suitable heating source it heats up whereby between the Dappelwänden of the balloon continual warm gases vein air flow rotates.

Under the basket with the wire nets leiiend ver¬bundene ball bearings 14 are attached, its tap with sorgi: iltig isolated, easy and sufficient is connected for firm cables 15 leading.

On the surface of the earth the hoist 16 (Fig.6) is, by means of which the balloon can be down-pulled and brought to the rise, as soon as the internal balloon with illuminating gas or hydrogen is gefüljt.

With the increase of the balloon the carried out over liquid work can be stored in any accumulator and be made usable with the Niederziehen of the balloon. The end of the cable core is at one at the axle of the hoist isoJiert attached collector firmly, from which the electricity is derived by means of a sliding contact. Collecting the electricity takes place by means of the fact that the balloon is down-moved by means of the hoist continual up and.

With this movement the balloon receives a turn, which will not change its sense by the shovels 9, since during the transition of the rise to sinking the balloon or turned around the shovels are tilted by air resistance, and therefore the balloon the turning movement in the same sense to give. In order to avoid here a turn of the cable, the ball bearing 14 is intended. The fact that the turn of the balloon takes place with on and descending in the same direction illuminates from the following.

The shovels 9 are easily swivelling, whereby the axis of rotation does not fall into the center line the same. Hierdl1rch places itself the shovel with the downward course of the balloon in in Fig. 10 drawn position, and der' balloon is from right to left turned. During the upward course the shovel tilts, b it the lower notice 10 limited (Fig. I). IE one sees, takes place the drive from above after air flow arranged downward again from left to right, the balloon can thus its direction of rotation maintained. By this up and Niederbewegen and tricks of the balloon the purpose that the same with as much as possible water particles floating in air into contact comes, is reached.

Of collector rings 17 derived electricity nJn a too high and much too much a changing tension possesses, in order to be able to be veaertet directly. Since one maintains accumulators too ladan and only the easily adjustable river with an irregular Elektrizitätsquelle first the same further used, then one must to the shop of the accumulators the electricity on more constant, derived from the collector 17, stretches! to receive lng. search and those

12

Tension also into belleutend a lower transform. Since one has to do it however in this Fa1le with a direct current, then usual transformers can not be used for this purpose. In addition the Electricity in this case possesses a much to large tension, so that when using common transformers the largest part of the collected electricity would come again into loss. 'The only applicable method is for this purpose the transformation by means of the planning ashes rheostaischen machine, by this high-tension electricity almost completely without loss transforj11iert will can. Thereby the question of the transformation would be solved only the preservation of the constant tension remains for more. This is reached in the following way:

Rheostatische machine 18 or only one part of the plates stands for the latter change-over of the rheostatischen machine with any charge measurer (electrometer) in connection, whose mobile part closes one, contact, what is caused, how the same unloaded. The potential sinks, and the ElektromeIer takes its initial position, whereby the Slrom umschailung the causative Elektromagnelen becomes interrupted and the P'atten of the rheoslatischen machine tension geschaltel. Then now to the erneulen charge ferlig and is again unloaded with the beslimmten tension. This play repeats itself continual, so long the apparatus in the enterprise stehl.

To the easier understanding the pattern of the rheoslatischen machine is indicated. Fig. the load position shows 12. The panel 42 is connected with the Fesselballon, while the last panel is led to the earth. Fig. the unloading position shows 13. In Fig. the rheostalische machine is drawn 8 in the load position, and that simplicity half is not the connection of the panel of 42 with that contacts 17 gezeichnt just like also the contacts for the unloading position (Fig. 13) is omitted.

With the elektrostalischen Schaltapparat' (Fig. 9) combined effect of the mechanism is the following:

The contact roller 27 (Fig.) always one brings to 9 by the feather/spring 37 into the load position. In addition axle a Electromagnet 35 is gekeilt, and at the rack the firm magnet 36 attached on d1e. If now the electromagnet 35 is energized, it turns the contact roller under Überwindung of the spring action into the Enlladestellung. Damil the excitation 'umschallende? Electromagnets selbs, ttälig with a certain tension take place, is a elektrostalischer switch (Fig.9) indicate-indicating eight.

That mil the earth connected voucher of the letzlen or the voucher of any other condenser is not with the balls 29 and 32 connected. If now into the Ladestellung¬ the potential rises, then the balls push off each other more and more, until the mobile ball 32 provided with the fork 33 dips into two Queck¬silbernäpfchen and thus the river exciting from the battery 3 supplied and the swivelling electromagnets 35 closes or this current conclusion by means of relays cause lets.

One can regulate the amount of the tension, with which this current conclusion and thus the unloading of the rheostatischen machine take place, mittel~ the adjustable weights 38 and 39. The apparatus is provided to the protection with the glass fall 40. Within the fall also one is with Chlorkalzium gefUIlte bowl 41.

As long as the tension at the balls 29 and 32 remains constant, the outflow, concomitantly the excitation of the magnet continues the accumulators receive the utilizable stream. If the condensers delivered their charge, repulsive Kraft of the balls become smaller, with which the Quecksilberkontakt is interrupted, those. Excitation electromagnets stops and the feather/spring 37 D! e contact roller back into the load position brings, on which the play of neuern can begin. At larger electric charges which can be derived two rheostatische machines can work abaechselnd, so that, while the one will unload, which can be loaded others. It is now easily evident that, if the capacity of the rheostatischen machine is not changed switching will always take place via the electrometer with the same potential of the rheostatischen machine " and the number of plates, thus the transformation relationship the same remains there, then also the current draws dieseIbe tension led from the rheostatischen machine will have.

As long as the tension at the balls 29 and 32 remains constant, the outflow, concomitantly the excitation of the magnet continues the accumulators receive the utilizable stream. If the condensers delivered their charge, repulsive Kraft of the balls become smaller, with which the Quecksilberkontakt is interrupted, those. Excitation electromagnets stops and the feather/spring 37 D! e contact roller back into the load position brings, on which the play of neuern can begin. At larger Elek¬trizitätsmengen which can be derived two rheostatische machines can work ab~echselnd, so that, while the one will unload, which can be loaded others. It is now easily evident that, if the capacity of the rheostatischen machine is not changed switching will always take place via the electrometer with the same potential of the rheo¬statischen machine " and the number of plates, thus the transformation relationship the same remains there, then also the current draws dieseIbe tension led from the rheostatischen machine will have.

By the irregularities of the Elektrizitätsquelle the period, in which the siromstösse follows one another, is changed this does not.hat however, as soon as now the tension remains constant, on shop of the accumulators harmful influence. The river led from the rheostatischen machine can be ver¬wendet at times, where it is sufficient constant, directly with evasion of the accumulators. 24 and 25 is the wring leads, to-consumption-places either directly or to storage battery leading.“

14

Execution of the Erhilzung in practice, and it is at all questionable whether the desired result is to be obtained thereby in addition are the large balloons from things with heavy metal nets and - sharpen to and for itself a thing offering no prospects.

In the further the American patent is 1014719 from 16. January 1912 inventors: Walter L. Pennock, Philadelphia, Pennsylvania mentions. The patent claim reads in the translation about follows dermasseri: The patent protection becomes geforden for a collecting tank for electrical charges, which consists of a number of carriers and a metal net, which are held in such a way by these carriers that it is appropriate for particular the carrier mentioned above, which furthermore anchor devices possesses in one level and in approximately even distance, from everyone, which with gemein¬samen points after everyone the carrier mentioned goes out, which furthermore a device, the carriers in their normal position possesses to hold, and in addition a device attached between everyone of the mentioned carriers and the net, which permits each carrier to adapt fast and independently each changes of the air flow.

Also after this patent one lets balloons from silk and such materials, which carry metallic collecting TZE, ascend into air. As is shown by the patent claim, the patent protection sought for the kind of the execution of the balloon herab¬hängenden leader. Because of this small difference of the previous patents I refrain from a more exact explanation.

Would last be that D.R.P. 248580 of the 22.Juni 1912 Inventor: Heinrich Johannsen, Luebeck to state. its patent claim reads as follows: . Procedure for catching Luftelektrizilät by means of senk¬rechter or diagonal leaders, thereby ge marks that into the line a high-tension power source is switched on, for the purpose, the point resistance to lower. “ Patent specification. . Procedure for catching atmospheric electricity by means of more perpendicular or diagonal leaders.

15

Further nothing is well-known over the atmospheric electricity, apart from the fact, .dass in newspapers publications of shining successes of an inventor appear here and there, the one new invention to the Nutzbarmar: hung of atmospheric electricity to have ago made wants and for these purposes money searches.

Tried one, with these people in connection to set, then always turned out that either large sums had to be paid before, in order at all only something details too experienced, or the invention always existed in gleir: hen thoughts: A balloon from silk was suggested, for which a net with many needles should carry. In which way the Anlj! ge before lightning or overvoltage to be g~schützt or like the transformation of the Electricity in a technically useful form as well as their adjustment be made should, of it knew itself the gentlemen neither one seriously thought conception out to make, still for it proofs furnish.

From above-said it follows clearly that the attempts to win atmospheric electricity thereby that air balloons rise as collecting tanks are left, those on a non-metallic envelope of the ballon --herein all inventors agree - a metallic Ergebnie of the preliminary tests

16

Sammel1: 1etz carried, to the goal to lead could not. The balloon had to large loads to carry and received therefore to large extent. Large attack regions, so that surely tearing the whole collecting TZE was to be foreseen had storm, Regenund other atmospheric influences, without talking about the fact to that such large envelopes of the balloon lost dauarnd large quantities gas by diffusion. Thus its use and Rentabllitat were impossible, even if a collection of atmospheric electricity möglfch seems to in this way actually be.

The second error of the past attempts was that all inventors the atmospheric Electricity eusserst directly as statisihen direct current of high tension and Ieiner Stromsterke to use wanted. Nearly insurmountable difficulties of the isolation and adjustment face this application. Procedures, which permit, this kind Electricity into another, less gerährliche form, which was not technically usable commodity, too yerwandeln, yet well-known, which had everything only is found.

All this had to be found only.

Hauptgrllnde of the blherlgen failures

The principal reasons of the past failures can in the following general points to be summarized:

All inventors thought for attaching heavy metallic nets on individual air balloons from nonconductive, tear upable and the gas letting through materials as collecting tanks.

No means had been used, in order to increase the ionization of the collection coefficients of the collector nets by artificial means (radioactive substances, photoelectrically working substances, devices for the education of ultraviolet jets etc.).

No inventor has an attempt with balloon collecting tanks made from sheet metal.

such metallic balloon collecting tanks however rolgende important possess and aur other way not advantages which can be reached: aluminium the metallic coverings is impenetrably for helium and Wasserstoff' them places gJeichzeitig Erosse metallic, weather remainders of Collector surfaces.

Radioactive means etc. can be easily attached inside or outside, whereby the ionization and thus also the collected quantity of the atmospheric electricity is substantially increased.

Such balloon collecting tanks made of light alloy do not need to be large, since they must carry Dur their own small weight and that of the wring lead.

The whole system offers therefore storm and wind w~nig to attack regions and becomes resistant and stable.

One can lift and lower balloon easily by a hoist, So that all repairs, Nachrüllen etc., safely wlhrend the work are auaführbar.

17

4. No inventor recognized importance and meaning of an antenna collecting tank net from several individual collecting tanks, which are connected among themselves by electrical conductor, spread over the earth in air.

5. No inventor recognized the importance of the use such high capacities for these purposes until today as means for the increase of the capacity used with production of atmospheric electricity condenser batteries, whose correct engagement conditions aus¬gearbeitet, as well as at all. (Consequence of the fact is that to today only extremely small electric charges and these in a practically useless type of current were available.)

6. All inventors were always anxious to convert the collected static electricity of air directly into mechanical energy which encountered evenly insurmountable difficulties.

7. So far no devices were used or intended against overvoltages, lightning danger etc. by the inventors. Gleichfalls was possible an adjustment of such an energy in no way. It is impossible to use the atmospheric electricity as long as the obstacles given after 1 to 7 are not eliminated.

The collection of the atmospheric Electricity can with the comparison of the... also on hot summer days, in a boiler, its ifeen iitee: Surface the sunbeams is evenly suspended, the water of the suns to energy-to cooking bring, although the sun mix also large warmth delivers.

If one would concentrate those sunbeams however by large lenses, which must be substantially larger than the boiler surface, on essel, then one could bring the water not only to cooking, but even to evaporation. Without collection of the jets the Wärmedichte is not sufficient on given square a surface for heating.

The same applies in similar sense also to the atmospheric electricity. If one sets up only a Antenna, then only those parts of the atmospheric electricity are tightened, which are in the proximity of the antenna. Since the density and movement speed of the electrically charged particles of the Atmosphere are only small however, then also the induced tension will be in the antenna wire kJein and mit¬hin also the amperage.

The ring or netlike arranged, strongly ionizing collecting systems in accordance with available invention permit however to collect larger electric charges there it electrically charged particles over a surface of 1, 10, to tighten 100 or more qis fortunes. Must rate however here a number of other conditions considered, which are to be described on the basis designs later more near.

The collection of the atmospheric electricity by means of large antenna system has the comparison example mentioned above (collection of the solarly power) opposite the following two advantages, which understood tons use so far nobody recognized and: 1. the application of large collecting surfaces (the individually lenses and/or mirror surfaces, which find during utilization OF solarly heat to application, one can emergency as make large, as the collecting system) 2. the large capacity of the collecting apparatuses for the atmospheric electricity (by the lenses or hollow mirrors one can affect emergency as large of surfaces, as with the antenna system).

If one counts the other difficulties, which the utilization of the solar power in the way, like changes of the seasons, change of the position of the sun to the Zenith, still in addition, then it results that conditions for the collection of atmospheric electricity lie substantially more favorably. The latter is also not to those Application of compact collecting surfaces or at certain locations or such restrictions binds. It supports itself rather by arbitrary ring or netlike arranged wire communication, which are carried for set up strongly ionizing collecting antennas in arbitrary headroom of, ground of in even distances been suitable and these among themselves to connect.

The openings such shape of set up rings or nets know all the the vertical antennas are more largely its, the higher, since the whole line delivers neither warmth nor light shade. The static electricity is continued to store owing to its special characteristics by condenser effect in the antenna wire, and it is possible to distribute it evenly by g~eignete engagements of special condenser batteries for Jn the net and to receive kolossale charge capacities in this way, for which there is no analogy with the collection of the solar power.

For aforementioned reasons it can be counted on the fact that available procedure the production and Nutzb!}rmachung the atmospheric electricity really on a cheap and a comfortable Way will make possible. A lack, which adhered to all past inventions, was that that always one tried to transfer the electricity immediately into mechanical. Even if it e.g. succeeded to design an electrostatic generator from large strength to then would be nevertheless its application because of that

19

Difficulties of the saftey, as well as the isolation of the axles and other parts not possibly. Such uncommonly large tensions would be possible that short-circuit and spark copy could not be avoided. Likewise it is impossible, such large [unknown] to switch off (regate safely and or adjust for even load to feed engines with high-tension rivers foreseen from manifold other obstacles, which forbid at all. After available invention the atmospheric electricity becomes [unknown] of the new [unknown] into mecamsce nergle umgewan et, which differentiated EN aupt palnzips from the past inventions on this areas forms, [unknown], those separate the static electricity by antenna lines in more mospärlscher. Electricity in the form of direct current of very high tension and small amperage [unknown] runs, converted into electro-dynamic oscillation [unknown] energy of high Wechselzahl. Thereby many advantages it is enough and eliminated all disadvantages.

The electro-dynamic oscillation energy formed in closed, oszillatorischen electric circuits produces built machine in another electric circuit, e.g. in one been suitable, by resonance:. effect electromagnetic waves of desired size and mechanical effect. Furthermore the resonance effect of such rivers permits, , in simple and comfortable way a starting, the adjustment and the disconnection of such machines to cause, this takes place simply by means of tendency and/or detuning of the resonance in the transformer circle and the river flowing in the coil of the machine. Such rivers, except for engine enterprise, are further f Urverschiedene other areas derTech.nik directly zuyerwenden, e.g. : Lighting, production of warmth and Elektroche I1ie.

In addition one knows with such rivers a set of Apparatuses without direct firm current inlet, completely apart from the possibility, they feed f Ur wireless telegraphy and telephone to use. In practice large difficulties were to be overcome, because it was not well-known, how one should design such large machines, and how the electromagnetic oscillations necessary for the supply could be adjusted by so high Wechselzahl.

In the following a specification is to be given, how these questions were solved. In Fig. a simple pattern is represented 14 for the transformation of static electricity into dynamic energy of high frequency. For clarity of the design is calibrates an air antenna, but a lnfluenzmaschine accepted. 13 and 14 ssind collection combs .

20

the static electricity the lnfluepzmaschine. 7 and 8 is spark discharge contacts, 6 and 5 switched on condensers, 9 switched on inductive primary coil, 11 and 12 line ends of the secondary winding 10. If the disk of the static electrostatic generator is set by a mechanical Kraft in turn, then comb collects the positive, the A! 1dere the negative electricity.

The allocations of the condensers 5 and 6 are so long loaded, until the screen end becomes so large Spannungserhöhung that the spark gap will pierce 7 to 8. Since over spark gap 7 and 8, over condensers 6 and 5 and inductive resisted themselves 9 a closed electric circuit forms, so entsteheQ, as well-known, into this circles electromagnetic oscillations.

High Wechselzahl

The produced high-swinging rivers in primary circles induce rivers with the same frequency, jedoce of already completely electromagnetic nature in the secondary circle. The electromagnetic oscillations will maintain by new charges of the static electricity. If the number of turns of the primary and Sekundärkreises stands to each other in correct relation, which one can compute with correct application of the resonance coefficients (capacity, inductance and resistance), then one can convert the rivers of the primary circle with high tension in such arbitrarily low tension and higher amperage.

If oscillation unloadings become weaker in primary circles or to diminish completely, begin more wiedrer the charge of the condensers with static electricity, until the spark gap will again pierce. All of this repeats itself so long, when of the static machine electricity is produced by supply of mechanical energy. It is not to be maintained that this application of the static machine and the transformation of the produced electricity do not admit according to above pattern in former times were such remarks were described already often applied and. It is stressed only priority on the fact that in this invention first these.

Experimental assembly for the production of atmospheric electricity for practical purposes brought in application and that the machines (engines), necessary for such rivers, were designed, as well as their circuit and adjustment patterns. By this invention alone the possibility is given to make really atmospheric electricity energy source usable for the technology as practically comfortable 'without by the plant mortal danger for those people exists. It is further only by transforming the static atmospheric electricity been into electromagnetic oscillations possible, the difficulties of the isolation, the building construction, the adjustment, to become starting and switching gentleman off.

From Fig. is evident to 15, how the arrangement for production and transforming of the atmospheric electricity, shown in pattern 14, can be used. Particularly emphasized sei' the main difference between these and other earlier inventions. It consists of it that two parallel spark gaps are an¬gewendet, from which the one as work distance (7) and the other one for the safety device against overvoltages serves.

The latter consists of several individual Funkenstrecken with larger total distance than the work distance, arranged in series connection, and is bridged with very small capacities (aluminium, bl, Cl), which has the purpose to make the even sparking possible in the safeguard distance Continues to mean in Fig. 15 A an air antenna for the collection of atmospheric electricity (specification follows still), 13 connection of the safeguard spark gap with the earth, 5 and 6 condensers, 9 primary coil.

If now by antenna A the positive, atmospheric electricity is anxious with the negative charge of the earth to adjust, then the Luftzwischenraum between the spark gaps prevents itself this reconciliation. How evidently from the design, the resistance of the spark gap is 7 lower than with that! , ndern spark gap, the spark gaps switched from three into series exists and consequently one three times larger air intermediate distance to overcome has. , Thus, as long as the resistance of the spark gap does not become 7 Überlastet, unloadings take place only concerning this.

However if the tension increases by any influences so much that it could become dangerous for the allocation of the condensers 5 and 6 or for the isolation of the windings 9 and 10, then a non-inductive unloading takes place to the earth when correct regulating the resistance of the second spark gap by means of these, without the machine is endangered. ühne this second parallel arranged spark gap also larger resistance than the work spark gap is it possible to collect and make large quantities of electricity usable safely. In Fig. the simplest is selected 1,5 represented pattern and only, in order to explain the basic principle. For practice ~ind more complicated patterns necessarily, from those the most important for usevdifferent intended purposes are prepared and in the following to be described.

Explanation of the principle of the transformation electromagnet punch oscillations into meehanlsche energy

The impact of the closed oscillation circles, consisting of spark gap 1 and two condensers 5 and 6, the primary coil 9, like also the secondary winding 10, is exactly the same, as described 14 for pattern. Are further in Fig. 15 schematically two motor types represented, which serve high Wechselzahl for the transformation of electromagnetic oscillations received from static electricity into mechanical energy.

Tesla Engines

Also on that areas of the high-swinging rivers were to be solved difficult questions. So far only the engine system was Tesla admits (schematically in Fig. to 15 by 16 and 17 represented), on the features of the Hysteresis is based. The engine is compound from iron pins and metallic disks. If on the same electromagnetic oscillations of high Wechselzahl influence, a rotary movement is produced. The description of this pattern serves only for the explanation of the basic principle it does not have however in the execution of large machines any practical interest because of the impossibility of the adjustment and because of the low effectiveness

Resonant Engines

According to available invention all are overcome these difficulties by construction of a machine, which is applicable for elektro¬magnetische energy high change number of more or less absorbed nature. Mg. the first form of a such machine shows 15. The difference between the principle of the Bau'es of these engines opposite the used so far consists of the fact that the engine contains keide magnets and not alone on the principle is based to the magnetic induction (like all current engines and also the Tesla engines), but on combination of the pure static induction with the electromagnetic in the static field. Since the engine is intended for high-swinging rivers, then must be as iron-free as possible er' and manufactured made of well leading metal. It was now shown that for the supply of such engines without magnet metal rivers of more or less absorbed N cannot only be used atur, to switch off but that also due to their special characteristics opposite electromagnetic reesonance features the possibility is given, the engines simply and safely in and and to adjust. These motor types can be called about this resonance engines. In Pig. 15 attached engine consists in his main parts of individual oscillation circles briefly closed in itself. The electromagnetic oscillation poles both the stator (un¬beweglicher part of the engine) and the rotor (mobile part of the engine) are radially embedded or fastened in a suitable firm insulator. Sämlliche of oscillation circles become from the Hauplstromleitung of a eleklromagnetischen source of oscillation suitable frequency and more or less absorbed nature gespelsl. Everyone such oscillatory circuit exists for itself of a closed oscillation circle, given by self induction, Kapazitäl and Ohm's resistance. From it results:

#) a Mögiichkeit to operate also with electromagnetic “oscillations of high Wechselzahl engines of arbitrarily high horse power since a sufficient number of such oscillation circles can be planned parallel in the engine

#) a possibility of designing for the enterprise with eleklromagnetischen oscillations of high Wechselzahl also asynchronously working engines (for similar purposes, for those the current synchronous engines for usual alternating currents or. Three-phase alternating current to be used)

#) a possibility of being able to count not only on a desired number of horse power in advance to obtain but also a widerslandsiähige and a simply executable construction form

#) an ideal Ein and elimination and an adjustment of such Moloren by change of the resonance i.e. by change of the capacity or the

self induction coefficient of the energizing current the more or less good resonance of the individual oscillation circles in relation to the main stream is destroyed or manufactured.

Determined by attempts that the same machine works accordingly differently, ever more or less the mechanical components with their natural oscillations of the natural oscillations of the energizing current to differ. After a set of attempts the engine received in Fig. 15 drawn, simplified execution form. By the wring leads (14 and 15) the engine from the source of the electromagnetic Schwjngungen is fed by high Wechselzahl.

The wring leads are with the secondary transformer 10 over two condensers - A and 6a connected. In certain cases the condenser batteries ä and 6a can be replaced by a parallel switched battery 18 (dots drawn). The stator can be designed in different way. From the stator line (thick Linienzug 11 and I) e.g. twelve branchings go into twelve short circuit oscillation circles, which are radially into the stator surface embedded.

Everyone of these oscillation circles possesses a certain given capacity and self induction, which are so selected that they adapt to the wavelength of the energizing current. These twelve oscillation circles form the twelve electromagnetic fields. Between these twelve oscillation circles, which are connected with the main line, are twelve other oscillation circles, which are been correct on same wavelength, i.e. the same induction coefficients as well as those

same capacity possess, but are connected with the feeder not directly. The oscillation circles connected directly with the feeder are named b and other with A. generally speaking receive the stator twenty-four individual oscillation circles. The first twelve oscillation circles are however not directly metallically connected, 'separate by a cylindrical staff b a surface induction are caused. In order to describe before the saying more near, are in Fig. 16 four such individual oscillation circles and their connections with the main line 14 dargeslellt. Hierflir becomes in this invention a neuarliger condenser benutzl, which is characterized by the fact that it possesses three vouchers. As from Fig. 16 to see is, becomes the short circuit oscillation circle consisting of the two condenser vouchers I and 2, which are short circuit with a metal leader 4 of ineiner turn, by the drilte allocation 3 erregl, which is metallically with the Hauplleilung connected direkl. If the capacity and the Selbstinduktionskoeffizienl are so computed in this drilten line that the own oscillations are in resonance with the oscillations of the main line, then the maximum at electromagnetic energy is supplied to the short circuit oscillation circle...

Wavelength entsprichl.

In practice führl one the KoQdensalorbelege in the form of metallic tubes out (Fig. 17). As the third condenser surface, which is to be considered as exciters, slehl in the Mitle (in Fig. 15 by 3 and b, b, b bezeichnel) a Slab or a tube. Between the individual tubes and also between the exciter electrodes is a guler insulator. Thus one receives a very fesles and comfortable system from oscillation circles for high-swinging rivers both up the stator, and on the rotor discs. These must be made of well isolating material, best from particularly for this worked on art wood and or paper without structure, since usual Holzund paper pulps does not exhibit usually necessary firmness and Isolierrahigkeit.

After the execution described above one receives with good electrical insulating capacity a mechanical firmness, which to all requirements of the art of building of engines correspond, to t. All individual connecting wires 4 (Fig. 16) can also by a strong metallic ring 4 (Fig. 15}, which is fastened on the stator and serves as document for eJnzelne metallic staffs, replace will. Of course must be isolated for special cases of the leaders 4 by isolation disks from the ring. ,

Having such three pole oscillation circles with the same capacity a very high induction to exhibit, then instead of smooth metal cylinders spirals are rolled up in cylinder form, whereby the further execution of the individual induction condenser vouchers has to happen completely exactly according to pattern 16. In the last case one obtains the mechanical firmness thereby that one manufactures the insulating layer from ebonite, fiber or similar material with high firmness and rolls the turns up on such tubes, these however in metallic rings into the stator surface embeds firmly.

In Fig. 17 is the connections of the particulars oscillation circles shown if no third electrode is necessary, and the oscillation circle only for the purpose of the phase shift is used. In the same way two-pole oscillation circles are embedded also in the rotor disc, in the same amount of more ocler also only half as much as in the stator surface. They must possess however the the same capacity and self induction coefficient as the stator oscillation circles.

The individual oscillation circles in the rotor can either. from each other to be isolated, or, as in the description of stator. by a metal ring (4, Fig. 15) to be short circuit with one another. On Fig. 18 and 19 is perspectively represented, like individual S {: hwingungskreise onto the rotor disc to be installed and by rings short circuited. The individual oscillation circles exist (as from Fig. 17 evidently) from two into one another pushed, but from each other isolated metal cylinders (on Fig. 18 with 1 and 2 designation), those, by V erbindungsleitung3 short circuit, an oscillation circle closed in itself form.

The cylinders become with their two ends on the rotor basis (not drawn overview for the sake of) in two rings from IsoliermateriaJ (fiber etc.) with the taps fastens A, A, b, b.

In practice are this-struggle however not necessarily, since the tubes more.der whole length are simply embedded after into the cylindric base of the rotor from IsoIiermaterial, which is fastened a metal cylinder on the rotor axis, on dyrselben axle (isolates or also short circuit) are two metallic rings 6 and 7.

By the special feeder line 4 one half of the oscillation circles is in the same way metallically continued to connect with ring 6, and by the lines 4a the other half with ring 7. In addition, one can connect all ends with a ring or all with both, furthermore even all feeder lines can be switched off, if one regards the axle as a common connecting leader, want one the self induction of the individual oscillation circles to increase, then the rings must be isolated from the axle 6 and 7, and the feeder line 3 is not short circuit as on Fig. the connecting leader 3 can be put to 18, but her between the condenser vouchers I and 2 and the connecting leaders 4 and 4a with the ring 6 and 7 on the rotor axis into some WindungeR (Fig.1 9), here lies.

By Auewahl of suitable capacity and self induction the individual oscillation circles of the rotor must brought to the resonance. become with the energizing current on the stator, this type of rotor resembles very the famous asynchronischen type of the rotors used in three-phase engines, the so-called squirrel wheel. Also the connection of the individual Schwjngungskreise can take place parallel not only with one another, but also in series and Gruppenscha4ung, exactly the same as with usual three-phase engines, only that instead of direct leaders always individual oscillation circles from condenser surfaces and induction resistance are taken.

Effect. Upward indicated Vorerklärung more over those execution that welse OF the RH. l EGG. R “. D…. l “h. sonanz. eIDze nen ernente TM resonance engine incoming goods now mog 1C of acres, engines it. the fashion OF action OF thesis of engines ton describe (Fig. 15). Highly swinging electromagnetic rivers acres derived by LINE 14 and 15 tons the engines to and and excite the individually oscillation circles with three electrodes (NUMBER 3), present RK the stator, those, as were above already mentioned, into resonance tons been correct of acres ton the full load current oscillations (? /i or 11. Wavelength). Electromagnetic oscillations into the third electrode excite for their part all oscillation circles b b b, in then short circuit, oszillatorische, undamped oscillations develop. The latter happens in vollkomme.ner way then only if between the Erregerstromkreis and work condenser circle a full resonance exists, which is reached by suitable selection by capacity, self induction and resistance. Electromagnetic oscillations in such short circuit, undamped, oszillatorischen circles with three electrodes (b b b b… Fig. 15) excites for their part oscillations of same kind and length in the simple oscillation circles present at the stator (aaaa…). Since these oscillations are secondary, then they are shifted to the primary (in the oscillation circles b b b with three electrodes) wavelength.

To assist in the understanding become in the further description nomenclature D. S manufacturing costs D. EGG k D (b b b F '). ä the oscillation travel with rel e tro EN… industrial union. 15 primely RH gungskreise. Oscillation circles (oszillatorische circles) and the oscillation circles without third electrode (A A A…) secondary oscillation circles (or secondary oszillatorische circles) mentioned. In the first circle formed the oscillations are continued to call primary and in the second circle (A A A) as secondary oscillations furthermore the oscillations in the third electrode load oscillations and the Schwillgungen in the rotor rotor vibrations are called.

If now in the two groups of oscillations of A and b oscillation care produced, then one receives generally speaking to Zusammenhange on the stator an electromagnetic rotary field, similar to as one with usual two-phase alternating currents by current shift around 1. Period an electromagnetic rotary field receives. The individual closed primary oscillation circles form here the similar elektroinagnetischen change poles, and the secondary oscillations cause the necessary electromagnetic shifts of these poles. The electromagnetic rotary field is formed here not by electromagnets, but oscillation fields.

Each oscillation circle is to be regarded as an electromagnet fed by alternating current. This electromagnetic rotary field formed on the stator induces electromagnetic oscillations of same kind in on the rotor discs fastened and on same wavelength the tuned oscillation circles. As on the stator an electromagnetic rotary field is formed by cooperation of the individual electromagnetic oscillation circles, then also a same rotary field in the rotor is produced and thus obtains a torque after nearly exactly the same basic laws, as with usual wechseistrom engines.

The resonance engine begins to rotate and can work 1eisten. je' the resonance between the charge stream and the primary and secondary oscillation circles on the stator and on the rotor is better, the more ideally is the effectiveness and the initial torque. The engine can be adjusted by tuning and/or detuning of this resonance both and switched off and easily, by obtaining another Eigenschwingungsperiode and thus another wavelength and frequency by change of the capacity or self induction.

In Fig. the Zuführungsle exists 15! tung II from a simple circle line. In practice however it can be put into some turns to receive to umzwischen two oscillation circles a determining tension. In addition, one can get along with a turn, if one attaches an electromagnetic counterweight for the third electrode in the form of capacities (represented by S or M). Instead of producing by secondary induction in the stator a rotary field, one knows this through-direct induction also in the Schwingungskreise. For the purpose one branches a second ring circuit from the main line before the third electrode off (Fig.) and an adjustable Selbstinduktionsspule (16) in series connection attaches 21 to this branching a capacity or.

The river runs then from the Heuptleitung 15 partly without induction resistance into the ring circuit II and excites without delay a number of Schwin gungskreise (9, 10, 11, 12, 13, 14a, 15a and 16a), about the self induction resistance 16 by the ring circuit I. by the self induction this part around 1/e period is partly retarded (during correct attitude of the resistance) or around 1/e period ahead accelerated (if the phase shift by condensers is caused). Thus however another set of oscillation circles in the stator received (1,2, 3, 4, 5, 6, 7 and 8), of them charge around 1/e period is late or accelerated takes place.

Thus in the stator a rotary field is produced, that the rotor in rotation to bring and. Work to carry out can. The individual oscillation circles are as in Fig. 16 and 17 implemented, only becomes the individual groups of I to 8 (Fig. 21) by ring b and 9 to 16a by a second ring A short circuit, during after Fig. 15 all oscillation circles by a ring to be short circuit. It is possible, on this principle. to build not only asynchronous engines, but also-synchronously working. Then m, uss however the stator line from two main distribution board rings 14 and 15 to couple (Fig.20) exist (without them metallically with one another).

All oscillation circles must receive as with pattern 16 three electrodes.

The rotor can as in Fig. 15 remains oscillation-circles represented.

In Fig. 20 is only stator

In all patterns described so far ...

Alteneck windings to implement, both on the stator and on the rotor.

Then that part of the oscillation circle e.g. on the rotor, which forms side on the rotor disc the capacity, comes (on Fig. designates 17 with I and 2), diametrically more gegen¬über the feeder line which is on the other side of the rotor disc (Fig. 17 with 3 designates), or with multipolar rotors with desired bending. If one half of the rotor disc is suction-wound, then one winds the other half likewise, but in reverse sense, so that between the feeder lines the capacities forming pipes come here, whose feeder lines come the first half between the capacities rolled up first, similarly “, oie with the Haefner Alteneck system the direct current coil are implemented, only with the difference that the individual sections consist here of individual oscillation circles, there however only briefly closed wire coils.

In the same way like the rotor one knows also the stator wickelll. N ur the feeder lines at the end must become the side gebogeIl. Such types permit a special utilization of the Verbindungsleiters of the individual oscillation circles. It continues to be possible after this coil system to switch two or four etc.

Kondensatoren in series by on rotor (or stator) a surface opposite a condenser tube (in Fig.) instead of a connecting leader dieser' 8elege second, for which comes, designates 17 with I and 2 to be appropriate for first same condenser tube. Both condensers becomeal at the ends by lines switched into series with one another.

One has the possibility, in which kind the following new forms of Rotorwicklungen euszuführen:

#) in wreath/ring form, so that all ladder are connected with the axle (parallel system)

#) all condensers are switched to two series and such series again in 2, 3 or 4 and more parallel groups with the axle are connected.

#) all oscillation circles become with one another ge in series, switch etc.

Your Support Maintains This Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

Order Page

Source : Fonte: rexresearch.com

Hermann PLAUSON

Conversion of Atmospheric Electricity

Biography (Wikipedia)

Meridian International Research: Atmospheric Electricity Research

***Science & Invention* (Feb. 1922): "Power from the Air" (I)**

***Science & Invention* (March 1922): "Power from the Air" (II)**

Plauson's Patents (List)

H. Plauson: USP # 1,540,998 -- Conversion of Atmospheric Electricity

H. Plauson: British Patent # 157,262 -- Improvements in Electric Motors

H. Plauson: British Patent # 157,263 -- Process & Apparatus for Converting Static Atmospheric Electrical Energy into Dynamic Electrical Energy...

***Science & Invention* (June 1928) - "Harnessing Nature's Electricity"**

H. Plauson : British Patent # 299735 -- Process for Producing Rapidly Moving Electrons [PDF]

H. Plauson : *Gewinnung und Verwertung der Atmosphatischen Elecktrizitat* (1922) [PDF]

---

http://en.wikipedia.org/wiki/Hermann_Plauson

Biography

Hermann Plauson was an Estonian engineer and inventor. Plauson investigated the production of energy and power via atmospheric electricity.

Plauson was the director of the Fischer-Tropsch "Otto Traun Research Laboratories" in Hamburg, Germany during the Weimar Republic of the 1920s. He built on Nikola Tesla's idea for connecting machinery to the "wheelwork of nature". Plauson's US Patent # 1,540,998 describes methods to convert alternating radiant static electricity into rectified continuous current pulses. He developed the Plauson's converter, an electrostatic generator. In 1920, Plauson published a book titled "*Production and Utilization of the Atmospheric Electricity*" (*Gr., Gewinnung und Verwertung der Atmospharischen Elektrizitat*). A copy of this book is in the British Library.

It is believed that he was related to Gertrud Plauson (the exact relationship is unknown; she may be his wife).

"Power from the Air". *Science and Invention* , Feb. 1922, no. 10. Vol IX, Whole No. 106. New York. (nuenergy.org)
"Power from the Air". *Science and Invention* , March 1922.

***Science and Invention,* Vol. IX (106) #10 (February 1922)**

Power from the Air (I)

by

Hugo Gernsback

During the war there was developed in Germany a new art --- or science --- that bids fair to revolutionize our present means of obtaining power.

This art, which is as new now as wireless was 25 years ago, will attain proportions during the next 25 years that may appear fantastic today. The inventor of the new science, an engineer of note, Herr Hermann Plauson, has devoted years of labor to his researches and he has now actually in use small power plants, that generate electricity direct from the air, day and night, without interruption at practically no cost, once the plant is constructed.

We had occasion, in one of our former issues, to describe the system, roughly, from cabled dispatches, but complete information is available now. The amount of electrical power that resides in our atmosphere is astounding. Herr Plauson found in his experiments that a single balloon sent aloft to a height of 300 yards gave a constant current at 400 volts of 1.8 amperes, or in 24 hours over 17-1/4 kilowatts! By using two balloons in connection with a special condenser battery, the power obtained was 81-1/2 kilowatts in 24 hours. The actual current delivered was 6.8 amperes at 500 volts.

The best balloons used by the inventor are made of thin aluminum leaf. No fabric was used. A simple internal system of ribs, stays and wires, gives the balloon rigidity as well as a certain amount of elasticity. The balloon, when made airtight, is filled with hydrogen or better, with helium. It will then stay aloft for weeks at a time. The outer surface is dotted with extremely sharp pins, made sharp electrolytically. Ordinary pins did not prove good current collectors, as they lacked extreme sharpness. The pins themselves were made from amalgamated zinc, containing a radium preparation, in order to ionize the air. It was also found that by dotting the outer surface of the balloon with zinc-amalgam more current could be collected. Even better results were obtained with polonium amalgam. Plauson states that the function of these amalgams is purely photoelectric.

One hundred of such captive balloons, separated one hundred yards from each other, will give a steady yield of 200 horsepower. This is the minimum, because in the winter this figure increases up to 400 horsepower, due to the higher electrification of the atmosphere.

We need not go into the technic of how the current is finally made useable for industrial purposes, suffice it to say that the problem has been entirely solved by Herr Plauson. By using batteries of condensers, high tension transformers, etc., the current can be transformed to any form desires. Such as for lighting lamps, running motors, charging storage batteries, etc.

Plauson also invented a sort of electrostatic rotary transformer which gives alternating current without the use of condensers and transformers. Indeed, its output is very great, as it actually ‘sucks’ the current down rapidly from the collector balloons. There is no doubt that this invention will soon come into universal use all over the world. We will see the land dotted with captive balloons, particularly in the country and wherever water power does not abound. Indeed, the time is not distant when nearly all of our power will be derived from the atmosphere. So far it seems to be the cheapest form of power known, it being much cheaper even than water power --- the cheapest form of power known today. Not only that, but as the inventor points out, no devastating thunder storms occur near such aerial power plants, because the balloons act not only as lightning arresters, but they quickly discharge the biggest thunder cloud, safely and noiselessly through their grounded spark gaps.

Science & Invention **(March 1922), page 1006, 1007**

Power from the Air (II)

by

Hugo Gernsback

[For many years electrical engineers have endeavored to devise some means whereby it would become possible to utilize the free electrical energy ever present in the atmosphere, but they were not successful, as every now and then an extra heavy surge of static current would rush down the elevated conductor and endanger the lives of the experimenters, or else destroy the apparatus connected with it. A German engineer has, however, devised the somewhat elaborate scheme here shown in brief, and he has succeeded, at least so his report states, in safely extracting several kilowatts of electrical power from the atmosphere with metallic surfaced balloons, elevated to a height of only 1000 feet.]

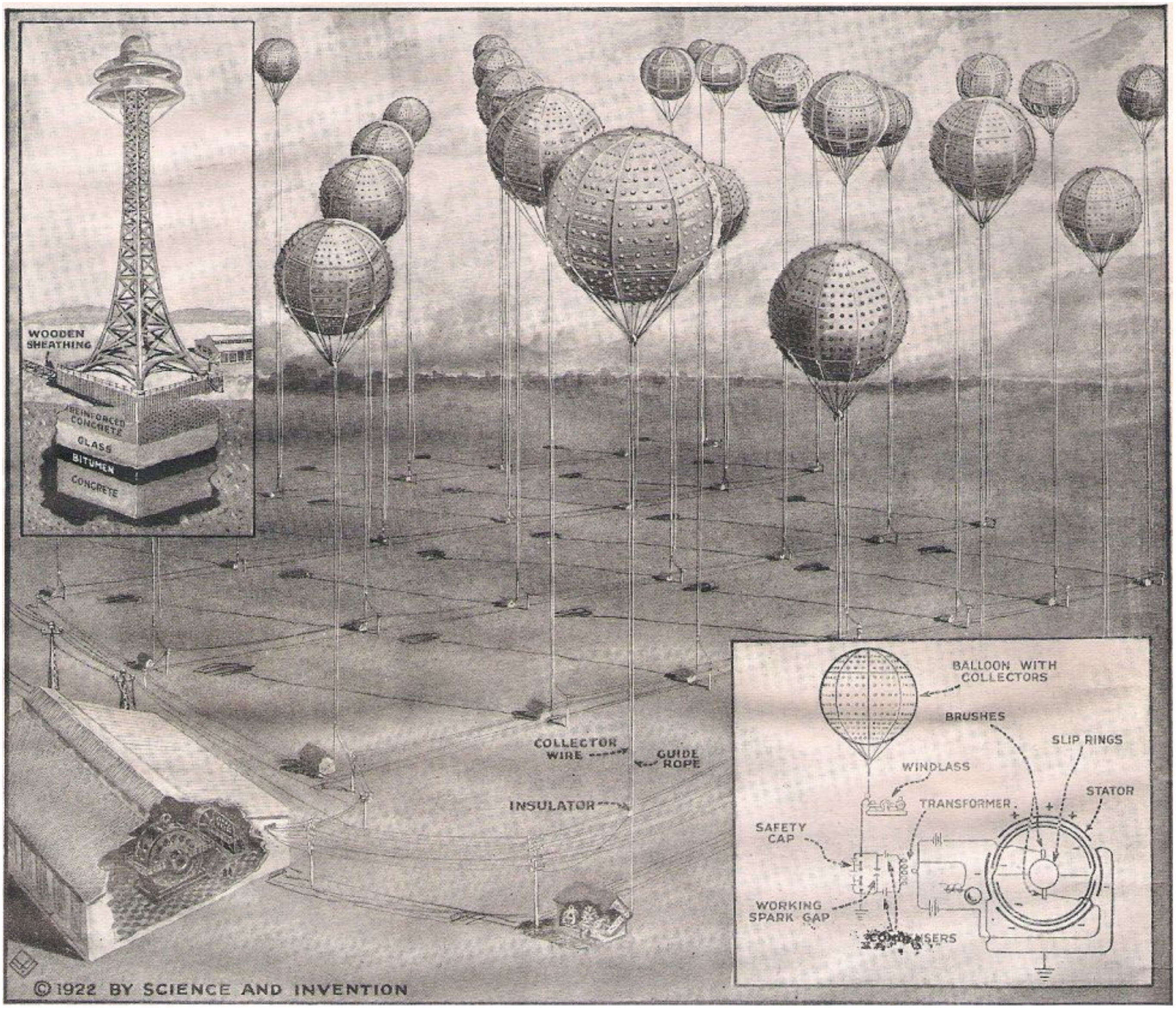

We have previously treated of the extraction of electrical energy from the atmosphere. The difference of electric potential in different parts of the atmosphere, and the difference between the upper air and earth make it a tempting proposition to obtain power from atmospheric electricity. The power would take the form of high potential difference with a discharge almost of a static nature. It has long appeared rather doubtful to conservative engineers, if such a source of power should really be available. Yet when we see the lightning flash, it certainly suggests very high power, even though the total of its energy may be small, on account of the small duration of the discharge It is not to the thunder storm that we look for getting power from the atmosphere, as the subject is now being seriously investigated. A German scientist, Hermann Plauson, has published a very elaborate work on this subject, and has investigated the use of kites, balloons and towers, for the utilization of the high potentials existing in the air at different altitudes, and has studied out the construction of motors to be operated by the peculiar type of discharge which will be obtained, if the projects are successfully carried out.

We will first speak of the methods used for collecting electricity from the upper air. The author cites several German patents. One of them shows the use of a kite balloon. The balloon is shown floating in the air, kite fashion, and from it hangs a great net or aerial for the collection of electricity. The conductor from the aerial leads to the ground station; quite an elaborate description is given of the net-work which the patentee proposes to have covered with needle points. A windlass takes in or pays out cable for the balloon, and the patentee claims that by sending the apparatus to a height of about one mile he will have 225,000 volts to draw upon. He then speaks o a battery of 20,000 cells in series, which will use up to 40,000 to 50,000 volts in the charging. This certainly provides for a reasonable large fall of potential.

But our author discards this idea and first suggests something more permanent. He proposes the erection of towers to be in the neighborhood of 1,000 feet high, or about the height of the Eiffel Tower. At the summit he has his collecting aerial. The appliance consists of a number of copper tubes; within each one he proposes to burn gas lamps, whose products of combustion will reach the aerial, a collecting net-work covering the tops of the tubes. One of his apprehensions is that if rain should wet his connections trouble might ensue, so he proposes a protection at the top in the shape of a great bell-like shield, resembling in his terms “a Siamese pagoda”. He also compares the form of the protection to that of a great petticoat insulator. Another of his difficulties is that he must have his tower insulated from the earth. He, therefore describes a complicated foundation for his structure. He proposes first to pour in at the bottom of the excavation a foundation of simple concrete. On this he places a layer of asphalt, and then a layer of cast glass, three to ten feet thick, and then comes a reinforced concrete foundation, to which the metallic foot of the tower is to be anchored. This foundation must rise at least seven feet above the ground level, and is to be boarded in on all sides to protect it from moisture. The author’s idea s to erect a number of these towers connected by a horizontal cable, to which the aerials for collection of potentials are secured.

The author strongly advocates balloons as collectors of the electric power of the air. These he depicts covered with spots. These spots indicate areas to be variously coated and prepared to collect potential from the atmosphere.

In the first place he describes the balloon as made of thin metallic leaf supported by internal ribs. Steel wires silver-plated, copper-plated, or aluminum-coated, run from the balloon to the pendant or junction ring. To this ring the tether cable is attached and runs to an insulated windlass on the surface of the earth. The balloon is to rise to an altitude varying from 300 feet to three miles.

The coating of the spots is to be of the thinnest amalgam, of mercury and gold, or zinc, or even polonium, perhaps only 1/2500 inch thick. All over the upper face of the balloon are numberless metal points. To prepare the needle-like wires, they are collected into bundles and are treated electrolytically in a bath, so as to be dissolved in part. This gives a sharp point and roughened surface, all adapted for collecting the electric energy. The points may be of copper, steel, or some hard metallic alloy. After this corrosion. As it may be termed, the wires are plated with gold or other of the so-called noble metals. It is advised that polonium or radium salts be added to the plating bath.

Dr Plauson devotes many pages of his book to describing his motor. This is a rotary motor including a stator and rotor and its peculiarity is that it contains no coils, develops no electromagnetic field properly speaking, but works by static excitation. One typical arrangement is shown in our illustration. The stator plates and rotor plates are concentric with each other, representing segments of cylinders. The alternation of negative and positive charged plates produces the rotation. In the connections there is included a safety spark gap to take care of dangerous potentials. Inductances and capacities are also used and indicated. It was found that the plates heated, owing to the Foucalt currents, and to overcome this, several methods of subdividing the stator and rotor plates, are described by the author.

The whole subject is quite captivating, and it really seems as if the utilization of the electricity of the air may be almost in sight. It would seem possible to carry out experiments in this direction by means of the Eiffel Tower, but of course, the trouble here is that the tower is grounded, and perfect insulation of the collecting surface is absolutely essential.

And now our author gives us some practical details. He says that on the Finland plains he carried out experiments with a balloon made of aluminum leaf with collecting needles of amalgamated zinc with a radium preparation as an ionizer. The surface of the balloon was sprinkled over with zinc amalgam. It was sent up to a height of 300 meters, early 1,000 feet, and was held by a copper-plated steel wire. A constant current of 1.8 amperes at an average of 400 volts potential difference was obtained. This gave nearly three-quarters of a kilowatt, or close to one horsepower. The collector of the balloon insulated from the earth showed a tension of 42,000 volts. By sending up a second balloon with an antenna to the same height at a distance of 100 meters from the first balloon, a current of over 3 amperes was obtained. Then by putting into the circuit a large condenser, whose capacity was equal to the surface capacity of both balloons, and of the antenna connections, the current rose to 6.8 amperes with about 500 volts mean tension. By the use of these two balloons, he eventually ran up the power to 3.4 kilowatts.

Science & Invention (June 1928)

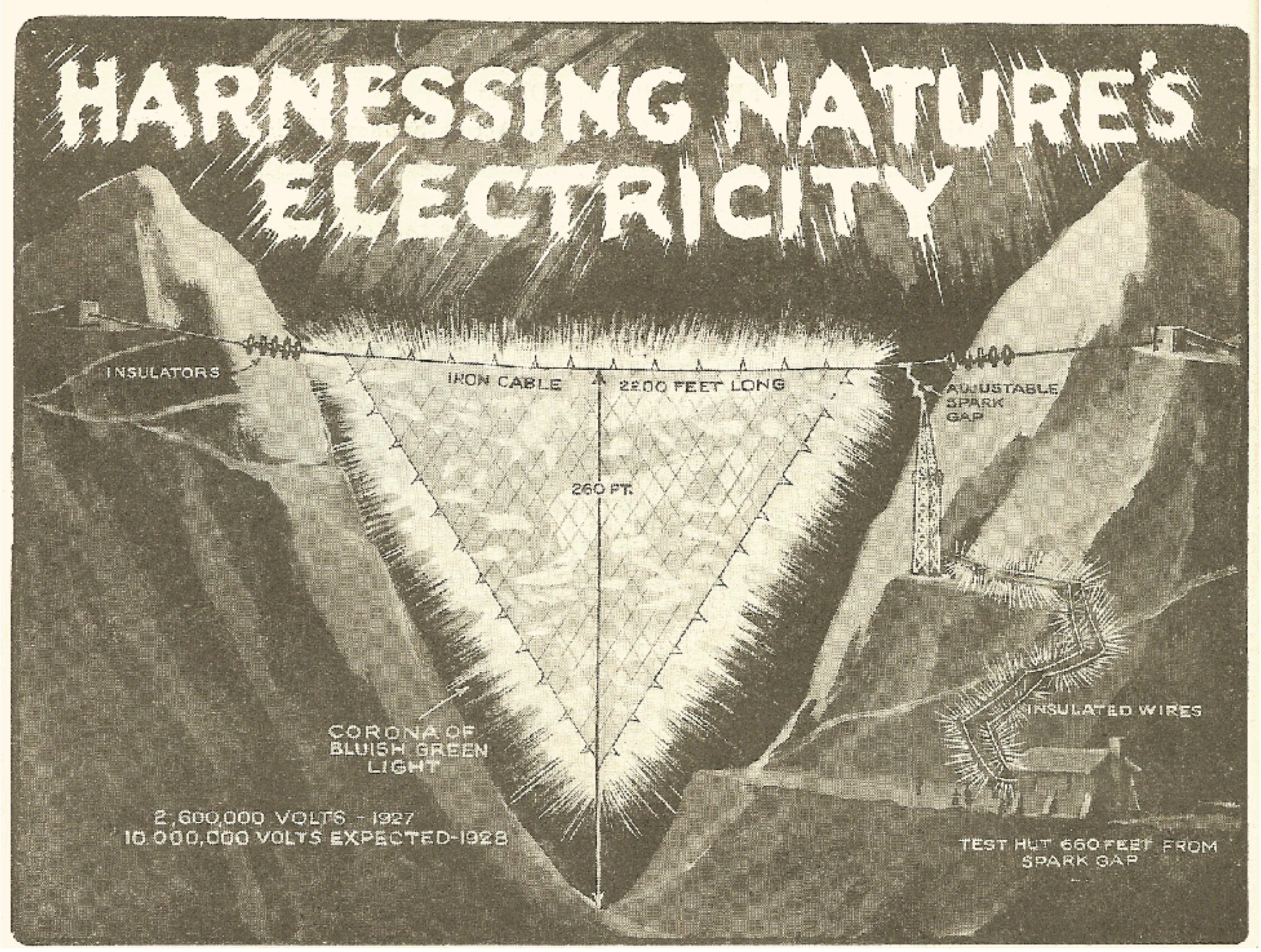
HARNESSING NATURE'S ELECTRICITY
INSULATORS
IRON CABLE
2200 FEET LONG
ADJUSTABLE SPARK GAP
260 FT.
CORONA OF BLUISH GREEN LIGHT
INSULATED WIRES
2,600,000 VOLTS - 1927
10,000,000 VOLTS EXPECTED-1928
TEST HUT 660 FEET FROM SPARK GAP

Remarkable European Experiments with Atmospheric Electrical Discharges with Potentials as High as 3,000,000 Volts

By HENRY TOWNSEND

elevation of 350 ft., and these students of natural electrical phenomena have found a very desirable location in the Alps, where they can suspend between one mountain and an adjacent one, a strong iron cable having a length of about 2,000 ft. This cable is about 250 feet above the intervening valley, and from it these daring engineers have suspended a coarsely woven wire net, which serves as an electrical capacity to gather the electricity from the atmosphere. As shown in the pictures, the wire net is supplied with numerous sharp points to aid in collecting the current from the air.

As the accompanying photographs of the actual apparatus and wire cable used last year clearly show, an adjustable spark gap of considerable length is provided. By adjusting this spark gap to various lengths, it is possible to judge the voltage of the discharge which leaps the gap at any moment. Mr. F. W. Peek, Jr., the well-known American worker in the realm of high voltage measurements, together with other engineers, have provided tabulated data and curves for various lengths of both needle and sphere type spark gaps. As one of the accompanying diagrams shows, it is a simple matter to calculate the voltage when a certain length of gap is used. The engineer first checks the length of the gap on the chart; he then follows a line horizontally from the gap length, to where it intersects with the angular line on the chart; and from the point of intersection he looks in a visual line downward to a place where the voltage is given. For needle spark gap measurements, the characteristic curve on the chart is practically a straight line, while for sphere gaps the characteristic curve on the voltage versus gap length, is a curved line. Those interested in high voltage measurements by means of the spark gap method can find the voltage-gap tables and charts in the *Standardization Rules* of the American Institute of Electrical Engineers. According to Mr. Peek's researches, the voltage per foot of atmospheric electrical discharges is about 100,000, while in laboratory measurements with A.C., transformer high potential discharges, the average voltage per foot of spark was found to be about 150,000 volts. The voltage of a lightning flash may

(*Continued on page* 156)

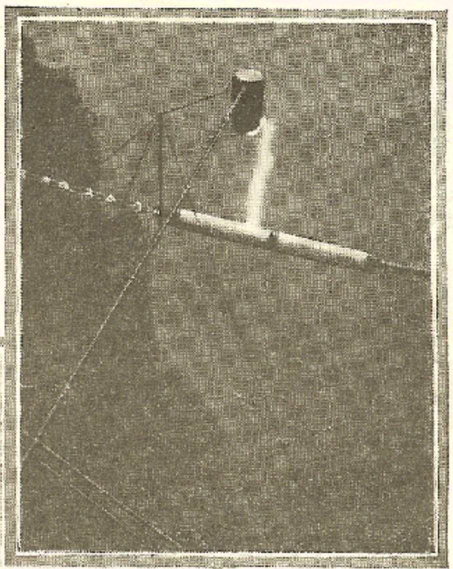

Actual photo above shows 13 ft., heavy spark obtained from the collecting net in the Alps by the German scientists. The voltage is about 2,000,000. The spark occurred once per second for 30 minutes.

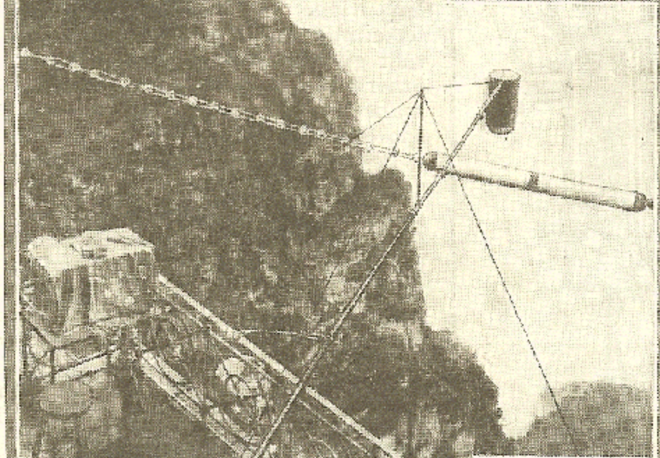

Photo, left, shows the adjustable spark gap used in the Alps. Notice the heavy electrode on the end of the adjustable arm to which the spark jumps.

Below we see 3,000,000-volt artificial lightning stroke produced in G. E. Laboratory at Pittsfield, Mass. Note man.

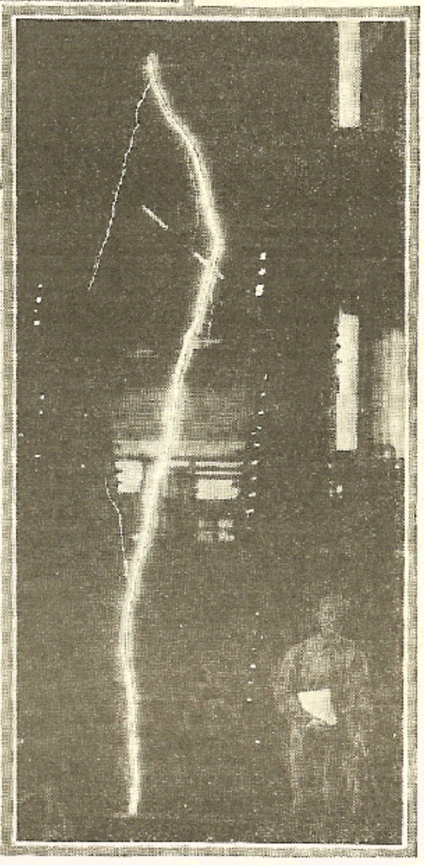

Actual photograph of the experimental "kite" used by the German experimenters in the Alps Mountains, for the purpose of accumulating high potential electrical discharges from the atmosphere. Note the size of the insulators.

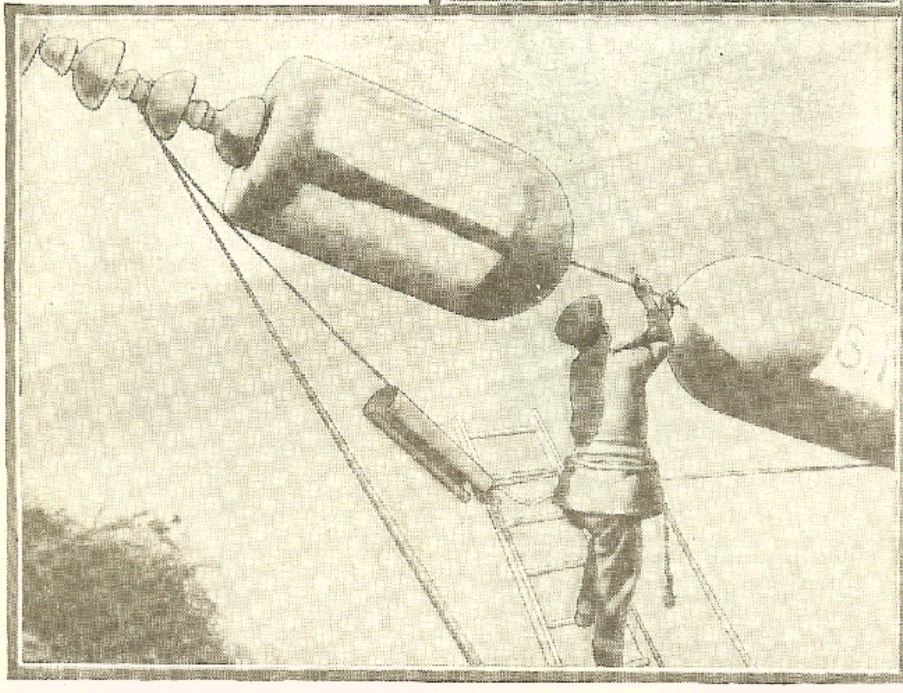

if-
you pledge
yourself to secrecy we will teach you the secrets of
Real Professional
Magic

Dr. Harlan Tarbell, internationally known magician and teacher of some of the world's greatest magicians.

Learn to be a Magician

For the first time in the history of Magic the age-old, sacredly guarded secrets of the Mystic Art are being revealed. Now at last you can learn to be a Real Professional Magician. Yes, and you can learn this wonderful art *easily and quickly* AT HOME. You are taught the mysterious *Principles* by which the Professional Magician works. You are taught everything from wonderful impromptu effects to massive stage illusions. But before you can study, you must sign the Magician's Solemn Pledge of Secrecy. This means that you are getting the *guarded secrets* of the Magic Profession. Think of that!

Earn $250 to $1000 a Month
There's a big demand for Magic entertainment. Magicians get big money. Dr. Tarbell gets as high as $250 for a half hour's work. Unlimited opportunities for you!

Low Prices Easy Terms

Write Today Send today for free Magic book, low prices and easy terms. Don't delay. Mail the coupon NOW!

Tarbell System, Inc., Studio A-142
1920 Sunnyside Avenue, Chicago, Ill.
Gentlemen: Without any obligation send me your free literature and information all about your wonderful "Tarbell Course in Magic." Also tell me about your Low Price Offer and Easy Payment Plan.

Name..

Address..

....Age........

HARNESSING NATURE'S ELECTRICITY

By HENRY TOWNSEND

(*Continued from page* 111)

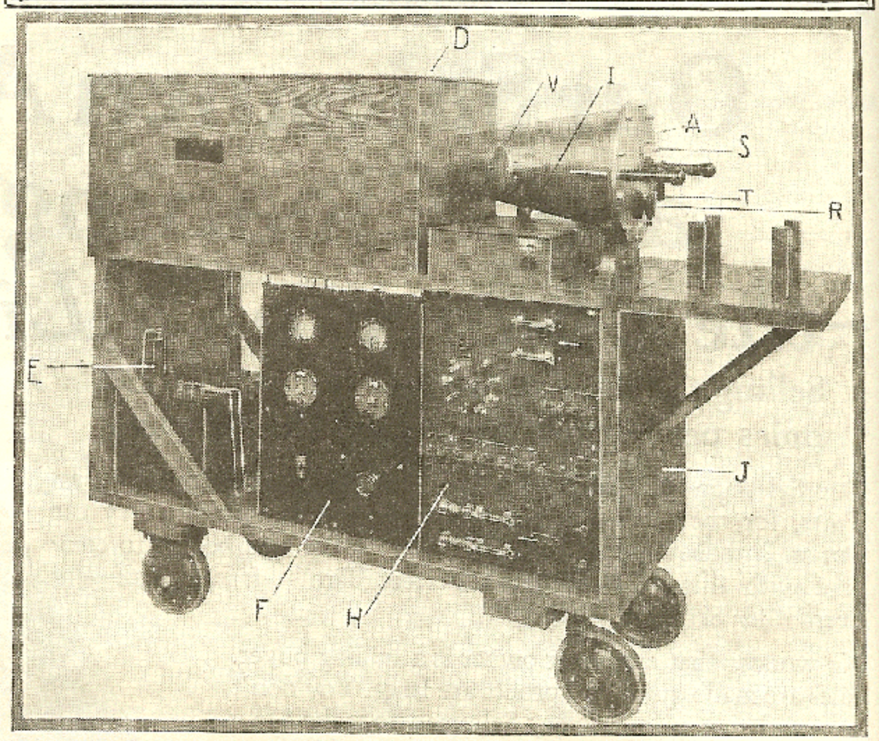

Remarkable cathode ray oscillograph used by Mr. F. W. Peck, Jr., in causing lightning to write its autograph with a pencil of electrons on a photographic film.

easily be 100,000,000 volts, as Mr. Peek has pointed out in one of his scientific papers.

WHY THE EXPERIMENTS ARE BEING MADE

ONE of the main reasons why these dangerous experiments are being carried on by the three young German scientists, whose names we have already learned, is because science believes that with a sufficiently high voltage, it will be possible for man to disintegrate the atom, and in this way make available a tremendous source of power as yet untapped. These experts have calculated that they will be able to obtain electrical energy in sufficient quantity from one of these powerful atmospheric discharges, to equal the Alpha rays obtained from 220 pounds of radium. As we have mentioned before, these experiments are of course fraught with great danger, and for that reason the experimenters seek refuge in a special lightning-proof hut, which is located about 600 ft., from the spark gap. When electrical storms are in the vicinity, it is especially important that the scientists keep within their protected fortress, for otherwise they would very probably be killed by a stray electrical discharge.

One of the peculiar things about this whole line of experiment is that the average layman does not realize perhaps, that there is a high electrical stress in the atmosphere on clear days, as well as when thunder storms are overhead. This fact has been known for a hundred years and more, and many years ago measurements of the various electrical potentials at increasing altitudes, were observed and measured by scientific investigators. There are a number of different ways in which these high electrical potentials found in the atmosphere can be measured; one of these methods involves the use of a calibrated spark gap. In this case the gap is set to a predetermined length, and when a discharge jumps this gap, the engineers know of course from previous experience and measurements, just what voltage is present. Another method of measuring extra high potentials, such as here encountered, requires the use of a static voltmeter, which involves the use of a stationary and of a movable or rotary set of metal plates, forming a condenser, to which an indicating needle is attached. For voltages above 2,000, static voltmeters have been used in a great many American central stations, and they have many desirable and useful characteristics. Of course as the voltage to be measured increases, the space between the quadrant shaped stationary and movable plates is increased and vice versa. The electrostatic field from voltages below 2,000 is not sufficient to warrant the use of a static voltmeter. Another method of measuring high potentials involves the use of the so-called vacuum tube voltmeter.

The general characteristics of the atmospheric electrical discharges, including lightning, have been measured and recorded by one of the newest scientific instruments, known as the cathode ray oscillograph. By means of this quite remarkable, high voltage scientific apparatus. Mr. Peck, one of the well-known General Electric Company staff of research engineers, has made some very interesting and remarkable discoveries concerning the nature of natural electrical discharges, particularly of lightning discharges. Many people will probably wonder why Mr. Peek and some of his colleagues in the engineering profession, including the three daring German students, Messrs. Brasch, Lange and Urban, play with such dangerous electrical discharges, and why they are at all interested in them. We explained previously why the German savants are intent on finding out all they can about these tremendous voltages obtained from the atmosphere, while Mr. Peck, we may say, also has a very practical reason for carrying on experiments with these death-dealing bolts of Thor. Mr. Peek has been for many years intent on finding out what causes the huge insulators on long distance, high potential transmission lines to break down when electrical storms break loose over these regions.

Plauson's Electrical Patents

USP # 1,540,998
Conversion of Atmospheric Electric Energy
6-09-1925

GB157262
Improvements in Electric Motors
1922-07-10

GB157263
Process and Apparatus for Converting Static Atmospheric Electrical Energy into Dynamic Electrical Energy of any Suitable High Periodicity
7-10-1922

British Patent # 299,735
Apparatus for Producing Rapidly Moving Electrons
7-15-1930

FI21227
Elektrisk uppvärmningsanordning
4-25-1946

Varmelegeme med elektriske varmemodstande
DK67691C
9-27-1948

FR877362
Dispositif de chauffage électrique
12-04-1942

DE734794
Elektrisches Heizsystem
4-24-1943

CH222509
Elektrischer Heizkörper zur Erwärmung von Flüssigkeiten
7-31-1942

DE738107
Elektrolyt fuer unmittelbare elektrische Warmwasser-Radiatorenheizung mit Elektroden
8-03-1943

DE433476
Verfahren zur Herstellung von Elektroden und Schleifkontakten fuer Dynamomaschinen
8-31-1926

CH94021
Elektrode und Verfahren zu deren Herstellung
4-01-1922

CA226423
Electrode for Electrolytic Apparatuses
11-21-1922

---

http://www.meridian-int-res.com/Energy/Atmospheric.htm

Atmospheric Electricity Research
[Excerpts]

In the nineteenth and early twentieth centuries, a large number of researchers investigated ways to extract electrical power from the Earth's ambient electric field.

The leader in this field was Dr Hermann Plauson who in the 1920s succeeded in generating significant quantities of electrical power comparable with modern solar photovoltaic systems of a similar scale...

The leader in this field before the Second World War appears to have been Dr Hermann Plauson. Dr Plauson was an Estonian citizen who lived in Hamburg and Switzerland. He carried out experiments in Finland with aerostats manufactured from magnesium-aluminium alloy, covered with electrolytically deposited needles. The needles were further doped with a radium compound to increase local ionisation of the air. (This was the era in which the hands of watches were hand painted with radium to make them luminous in the dark). Zinc amalgam patches were also painted onto the aerostats. Plauson obtained a power output of between 0.72kW and 3.4kW from one and two aerostats 300m above ground level. Dr Plauson filed patents in the USA, Great Britain and Germany in the 1920s. His book "Gewinnung und Verwertung der Atmosphärischen Elektrizität" is the most detailed known account of the technology.

Other atmospheric electricity researchers contemporary to Dr Plauson included Walter Pennock and MW Dewey in the USA, Andor Palencsar in Hungary and Dr Heinrich Rudolph in Germany. Hippolyte Charles Vion in Paris predated them all, putting forward proposals in the 1850s and 1860s.

Heinrich Rudolph made an interesting contribution to the design of the aerostat collectors. In 1898 he designed an elliptical aerostat made up of faceted surfaces to minimise the effect of wind. The design bears a strong resemblance to Northrop's 2003 UCARS unamnned helicopter UAV project. The design uses the Coanda Effect to help keep the aerostat on station and minimise wind effects.

In recent times, the only person who seems to have been active in this field is Dr Oleg Jefimenko. Dr Jefimenko carried out experiments on driving electrostatic motors from the Earth's electric field in the 1970s and has recently called for research into the neglected field of electrostatic motors to be renewed.

MIR's Research Programme

Since 1997 we have been carrying out theoretical research into conversion of atmospheric electricity into useable electrical power.

From a low level (5m high) simple zinc antenna we are able to obtain sufficient charge to light a number of white power LEDs. Further experimental investigations with metallic aerostat collectors and cavity resonant slow wave antennae concepts are ongoing...

Advantages of Atmospheric Electricity

Simple and robust technology
Low Cost technology - much cheaper than photovoltaics or wind turbines
Available day and night in all weather conditions - in fact, more power is produced at night than during the day
Available at any point on the Earth's surface

1. *Gewinnung und Verwertung der Atmosphärischen Elektrizität*, Dr Hermann Plauson, Hamburg, (1920)
2. Conversion of Atmospheric Electric Energy, USP 1,540,998, Dr Hermann Plauson, (1925)
3. Assembly for the Induction of Lightning into a Superconducting Magnetic Energy Storage System, USP 5,367,245 Goven Mims, (1994)
4. Electrostatic Motors are Powered by Electric Field of the Earth; CL Stong, *Scientific American*, (October 1974)
5. Operation of Electric Motors from the Atmospheric Electric Field; Dr Oleg Jefimenko, *American Journal of Physics*, vol. 39, July 1971.
6. *Electrostatic Motors: Their Principles, Types and Theory of Operation*; Dr Oleg Jefimenko, Electret Scientific, (1972).
7. Parametric Electric Machine, USP 4,622,510, Ferdinand Cap, (1986).

US Patent # 1,540,998

Conversion of Atmospheric Electric Energy

(9 June 1925)

Hermann PLAUSON

Be it known that I, Hermann Plauson, Estonian subject, residing in Hamburg, Germany, have invented certain new and useful improvements in the Conversion of Atmospheric Electric Energy, of which the following is a specification.

Methods of obtaining atmospheric electricity by means of metallic nettings set with spikes which are held by means of ordinary or anchored kite balloons made of fabric and filled with hydrogen, are in theory already known. Atmospheric electricity obtained in this way has been suggested to be used in the form of direct current for the charging of accumulators. This knowledge however is at present only theoretical as the conversion in practice has hitherto been a failure. No means are known of protecting the apparatus from destruction by lightning. The balloons used for collecting the charge must also me be made of very large size in order to be able to support the weight of the metallic netting and the heavy cable connections.

Instead of using heavy metallic netting as collectors attached to single air ballons of non-conducting materials which are liable to be torn and are permeable to the gas, it is proposed to use metallic balloon collectors which have the following important advantages ---

(a) The metallic cases are impenetrable to helium and hydrogen; they also represent large metallic weather-proof collecting surfaces.

(b) Radio active means the like may be easily applied internally or externally; whereby the ionization is considerable increased and therewith also the quantity of atmospheric electricity capable of being collected.

(c) Such balloon collectors of light metal do not require to be of large size as they have to carry only their own moderate weight, and that of the conducting cable or wire.

(d) The entire system therefore offers little surface for the action of storm and wind and is resistant and stable.

(e) Each balloon can be easily raised and lowered by means of a winch so that all repairs, recharging and the like can be carried out without danger during the operation.

It is further proposed to use a collecting aerial network of several separate collectors spread out in the air above the earth, which collectors are interconnected by electrical conductors.

According to this invention charges of atmospheric electricity are not directly converted into mechanical energy, and this forms the main difference from previous inventions, but the static electricity which runs to earth through aerial conductors in the form of direct current of very high voltage and low current strength is converted into electro-dynamic energy in the form of high frequency vibrations. Many advantages are thereby obtained and all disadvantages avoided.

The very high voltage of static electricity of a low current strength can be converted by this invention to voltages more suitable for technical purposes and of greater strength. By the use of closed oscillatory circuits it is possible to obtain electromagnetic waves of various amplitude and thereby to increase the degree of resonance of such current. Such resonance allows various values of inductance to be chosen whereby again the governing of the starting and stopping of machines driven thereby by simply tuning the resonance between coils of the machine and the transformer circuit forming the resonance can easily be obtained. Further, such currents have the property of being directly available for various uses, even without employing them for driving motors, of which there may be particularly mentioned, lighting, production of heat and use in electro-chemistry.

Further, with such currents a series of apparatus may be fed without direct current supply through conductors and also the electro-magnetic high frequency currents may be converted by means of special motors adapted for electro-magnetic oscillations into mechanical energy, or finally converted by special machines into alternating current of low frequency or even into direct current of high potential.

The invention is more particularly described with reference to the accompanying diagrams in which: ---

Figure 1 is and explanatory figure.

Figure 2 is a diagrammatic view of the simplest form.

Figure 3 shows a method of converting atmospheric electrical energy for use with motors.

Figure 4 is a diagram showing the use of protective means.

Figure 5 is a diagram of an arrangement for converting large current strengths.

Figure 6 is a diagram of an arrangement including controlling means.

Figure 7 shows means whereby the spark gap length can be adjusted.

Figure 8 shows a unipolar connection for the motor.

Figure 9 shows a weak coupled system suitable for use with small power motors.

Figures 10, 11, and 12 show modified arrangements.

Figure 13 shows a form of inductive coupling for the motor circuit.

Figure 14 is a modified form of Figure 13 with inductive coupling.

Figure 15 is an arrangement with non-inductive motor.

Figure 16 is an arrangement with coupling by condenser.

Figure 17, 18, and 19 are diagrams of further modifications.

Figure 20 shows a simple form in which the serial network is combined with special collectors.

Figure 21 shows diagrammatically an arrangement suitable for collecting large quantities of energy.

Figure 22 is a modified arrangement having two rings of collectors.

Figure 23 shows the connection for three rings of collectors.

Figure 24 shows a collecting balloon and diagram of its connection of condenser batteries.

Figure 25 and 26 show modified collector balloon arrangements.

Figure 27 shows a second method of connecting conductor for the balloon aerials.

Figure 28 shows an auto-transformer method of connection.

Figure 29 shows the simplest form of construction with incandescent cathode.

Figure 30 shows a form with cigar shaped balloon.

Figure 31 is a modified arrangement.

Figure 32 shows a form with cathode and electrode enclosed in a vacuum chamber.

Figure 33 is a modified form of Figure 32.

Figure 34 shows an arc light collector.

Figure 35 shows such an arrangement for alternating current.

Figure 36 shows an incandescent collector with Nernst lamp.

Figure 37 shows a form with a gas flame.

Fig. 1.

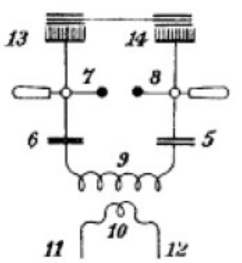

Fig. 2.

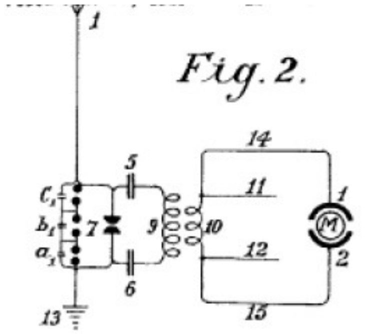

Figure 1 illustrates a simple diagram for converting static electricity into dynamic power of a high number of oscillations. For the sake of clearness in the drawings an influence machine is assumed to be employed and not an aerial antenna. 13 and 14 are combs for collecting the static electricity of the influence machine. 7 and 8 are spark discharging electrodes. 5 and 6 are condensers, 9 an inductive primary coil, 10 secondary coil, 11 and 12 ends of conductors of the secondary coil 10. When the disc of the static influence machine is rotated by mechanical means, the combs collect the electric charges one the positive and the other the negative, and charge the condensers 5 and 6 until such a high potential is formed across the spark gap 7-8, that the spark gap is jumped. As the spark gap 7-8 forms a closed circuit with condensers 5 and 6, and inductive resistance 9, as is well known, waves of high frequency electromagnetic oscillations will pass in this circuit.

The high frequency of the oscillations produced in the primary circuit induces waves of the same periodicity in the secondary circuit. Thus in the primary circuit electromagnetic oscillations are formed by the passage of the spark over the spark gap and these waves are maintained by fresh charges of static electricity.

By suitably selecting the ratio between the number of the coils in the primary and secondary circuits with regard to a correct application of the co-efficients of resonance (especially, inductance and resistance) the high voltage of the primary circuit may be suitably converted into low voltage and high current strength.

When the oscillatory discharges in the primary circuit becomes weaker or entirely cease, the condensers are charged again by the static electricity until the accumulated charge again breaks down the spark gap. All this is repeated as long as electricity is produced by the static machine employing mechanical energy.

An elementary form of the invention is shown in Figure 2 in which two spark gaps in parallel are used one of which may be termed the working gap 7 in Figure 2, whilst the second serves as a safety device for excess voltage and consists of a larger number of spark gaps than the working section, which gaps are arranged in series and are bridged by very small capacities as is illustrated in a, b, c, Figure 2 which allow of uniform sparking in the safety section.

In Figure 2 A is the aerial antenna for collecting charges of atmospheric electricity. 13 is the earth connection of the second part of the spark gap, 5 and 6 are condensers, 9 a primary coil. Now when through the aerial A the positive atmospheric electricity seeks to combine with the negative charge to earth, this is prevented by (the air gap between) the spark gaps. The resistance of the spark gap 7 is, as shown in the drawings, lower than that of the other safety section which consists of three spark gaps connected in series, and consequently a three times greater air resistance is offered by the latter.

So long, therefore, as the resistance of the spark gap 7 is not overloaded, so that the other spark gaps have an equal resistance with it the discharges take place only over spark gap 7. Should however the voltage be increased by and influences so that it might be dangerous for charging the condensers 5 and 6 or for the coil insulation 9 and 10 in consequence of break down, by a correct regulation of this spark gap the second spark gap can discharge free from inductive effects direct to earth without endangering the machine.

Without this second spark gap, arranged in parallel having a higher resistance than the working spark gap it is impossible to collect and render available large quantities of electrical energy.

The actions of this closed oscillation circuit consisting of spark gap 7, two condensers 5 and 6, primary coil 9, and also secondary coil 10 is exactly the same as the one described in Figure 1 with the arrangement of the static induction machine with the only difference that here the second spark gap is provided. The electromagnetic high frequency alternating current obtained can be tapped off from the conductors 11 and 12 for lighting and heating purposes. Special kinds of motors adapted for working with these peculiar electrical charges may be connected at 14 and 15 which can work with static electricity charges or with high frequency oscillations.

In addition to the use of spark gaps in parallel a second measure of security is also necessary for taking off the current. This precaution consists according to this invention, in the introduction of and method of connecting certain protective electromagnets or choking coils in the aerial circuit as shown by S in Figure 3.

A single electromagnet only having a core of the thinnest possible separate laminations is connected with the aerial.

In the case of high voltages in the aerial network or at places where there are frequent thunder storms, several such magnets may however be connected in series.

In the case of large units or plants several electromagnets can be employed in parallel or in series parallel.

The windings of these electromagnets may be simply connected in series with the aerials. In this case the winding preferably consists of several thin parallel wires, which make up together, the necessary section.

The winding may be made of primary and secondary windings in the form of a transformer. The primary windings will be then connected in series with the aerial network, and the secondary winding more or less short-circuited over a regulating resistance or an induction coil. In the latter case it is possible to regulate to a certain extent the effect of the choking coils. In the further description of the connecting and constructional diagrams the aerial electromagnet choke coil is indicated by a simple ring S.

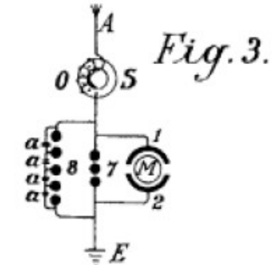
Fig. 3.

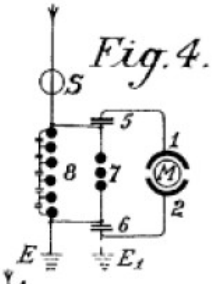
Fig. 4.

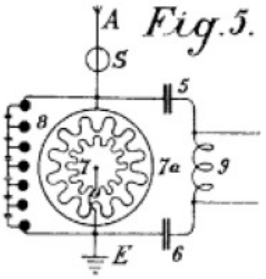
Fig. 5.

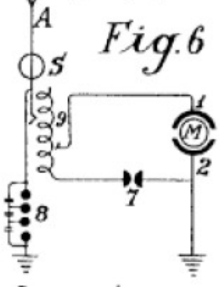
Fig. 6

Figure 3 shows the simplest way of converting atmospheric electricity into electromagnetic wave energy by the use of special motors adapted for high oscillatory currents or static charges of electrical energy. Recent improvements in motors for working with static charges and motors working by resonance, that is to say, having groups of tuned electromagnetic cooperating circuits render this possible but such do not form part of the present invention.

A motor adapted to operate with static charges will for the sake of simplicity be diagrammatically indicated by the two semicircles 1 and 2 and the rotor of the motor by a ring M (Figure 3). A is a vertical aerial or aerial network. S the safety choke or electromagnet with coil O as may be seen is connected with the aerial A. Adjacent the electromagnet S the aerial conductor is divided into three circuits, the circuit 8 giving the safety spark gap, the circuit 7 with the working spark gap, and then a circuit including the stator terminal 1, the rotor and stator terminal 2 at which a connection is made to the earth wire. The two spark gaps are also connected metallically with the earth wire. The method of working these diagrams is as follows:

The positive atmospheric electric charge collected tends to combine with the negative electricity (or earth electricity) connected with the earth wire. It travels along the aerial A through the electromagnet S without begin checked as it flows in the same direction as the direct current. Further, its progress is arrested by two sparks gaps placed in the way and the stator condenser surfaces. The stator condenser surfaces are charged until the charge is greater than the resistance of the spark gap 7, whereupon a spark springs over the spark gap 7 and an oscillatory charge is obtained as by means of the motor M, stator surfaces 1 and 2, and spark gap 7, a closed oscillation circuit is obtained for producing the electromagnetic oscillations. The motor here forms the capacity and the necessary inductance and resistance, which, as is well known, are necessary for converting static electricity into electromagnetic wave energy.

The discharge formed are converted into mechanical energy in special motors and cannot reach the aerial network by reason of the electromagnet or choke. If, however, when a spark springs over the spark gap 7 a greater quantity of atmospheric electricity tends to flow to earth, a counter voltage is induced in the electromagnet, which is greater the more rapidly and strongly the flow of current direct to the earth is. By the formation of this opposing voltage a sufficiently high resistance is offered to the flow of atmospheric electricity direct to earth to prevent a short circuit with the earth.

The circuit containing spark gap 8 having a different wave length which is not in resonance with the natural frequency of the motor, does not endanger the motor and serves as security against excess voltage, which, as practical experiments have shown, may still arise in certain cases, but can be conducted direct to earth through this spark gap.

In the diagram illustrated in Figure 4 the spark gap 7 is shunted across condensers 5 and 6 from the motor M. This construction affords mainly a better insulation of the motor against excess voltage and a uniform excitation through the spark gap 7.

In Figure 5 a diagram is illustrated for transforming large current strengths which may be employed direct without motors, for example, for lighting or heating purposes. The main difference is that here the spark gap consists of as star shaped disk 7 which can rotate on its own axis and is rotated by a motor opposite similarly fitted electrodes 7a. When separate points of stars face one another, discharges take place, thus forming an oscillation circuit over condensers 5 and 6, and inductance 9 for oscillatory discharges. It is evident that a motor may also be directly connected to the ends of the spiral 9.

The construction of the diagram shown in Figure 6 permits of the oscillation circuit of the motor being connected with an induction coil/ Here a regulating inductive resistance is introduced for counter-acting excess voltages in the motor. By cutting the separate coils 9 (coupled inductively to the aerial) in or out the inductive action on the motor may be more or less increased or variable aerial action may be exerted on the oscillation circuit.

Filed Jan. 13, 1921 12 Sheets-Sheet 2

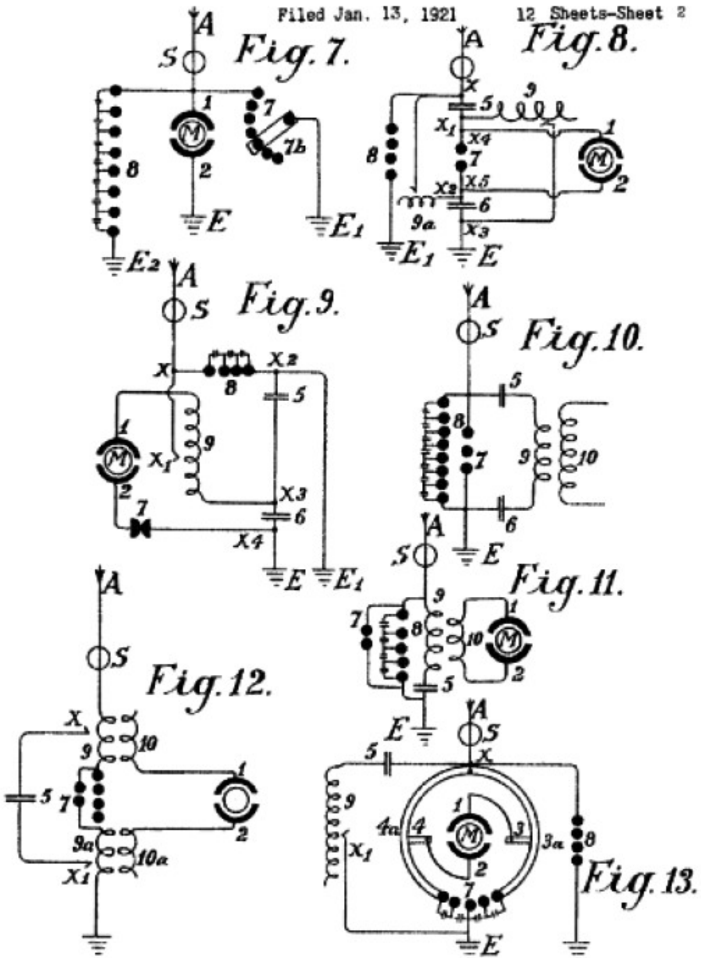

Fig. 7.

Fig. 8.

Fig. 9.

Fig. 10.

Fig. 11.

Fig. 12.

Fig. 13.

In Figure 7 the oscillation circuit is closed through the earth (E and E1). The spark gap 7 may be prolonged or shortened by more or fewer spark gaps being successively connected by means of a contact arm 7b.

Diagram 8 shows a unipolar connection of the motor with the aerial network. Here two oscillation circuits are closed through the same motor. The first oscillation circuit passes from aerial A through electromagnet S, point *x*, inductance 9a to the earth condenser 6 and further, over spark gap 7 to the aerial condenser 5 and back to *x*. The second oscillation circuit starts from the aerial condenser 5 at the point x^1 over the inductance 9 to the earth condenser 6 at the point x^3 and through the condenser 6 over the spark gap 7 back to x^1. The motor itself is inserted between the two points of the spark gap 7. From this arrangement slightly damped oscillation wave currents are produced.

In the diagram illustrated in Figure 9 a loosely coupled system of connections is illustrated which is assumed to be for small motors for measuring purposes. A indicates the aerial conductor, S the electromagnet in the aerial conductor, 9 the inductance, 7 the spark gap, 5 and 6 condensers, E the earth, M the motor, and 1 and 2 stator connections of the motor. The motor is directly metallically connected with the oscillation circuit.

In Figure 10 a purely inductive coupling is employed for the motor circuit. The motor is connected with the secondary wire 10 as may be seen in Figure 11 in a somewhat modified diagram connection. The same applies to the diagram of Figure 12.

The diagrams hitherto described preferably allow of motors of small and medium strength to be operated. For large aggregates, however, they are too inconvenient as the construction of two or more oscillation circuits for large amounts of energy is difficult; the governing is still more difficult and the danger in switching on or off is greater.

A means of overcoming such difficulties is shown in Figure 13. The oscillation circuit here runs starting from the point x over condenser 5, variable inductance 9, spark gap 7, and the two segments (3a and 4a) forming arms of a Wheatstone bridge, back to x, If the motor is connected by brushes 3 and 4 transversely to the two arms of the bridge as shown in the drawings, electromagnetic oscillations of equal sign are induced in the stator surfaces 1 and 2 and the motor does not revolve. If however the brushes 3 and 4 are moved in common with the conducting wires 1 and 2 which connect the brushes with the stator poles a certain alteration or displacement of the polarity is obtained and the motor commences to revolve.

The maximum action will result if one brush 3 comes on the central sparking contact 7 and the other brush 4 on the part *x*. They are however, usually in practice not brought on the central contact 7 but only held in the path of the bridge segments 4a and 3a in order not to connect the spark gaps with the motor oscillation circuit.

Filed Jan. 13, 1921 12 Sheets-Sheet

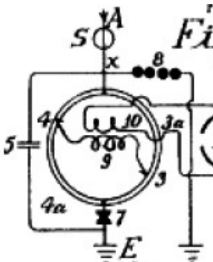
Fig. 14.

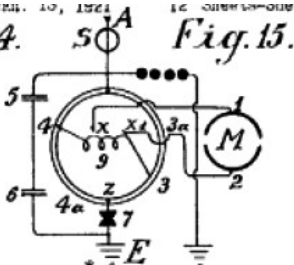
Fig. 15.

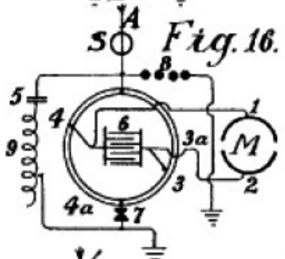
Fig. 16.

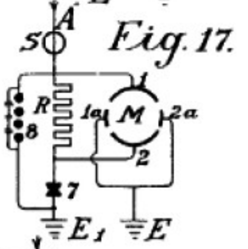
Fig. 17.

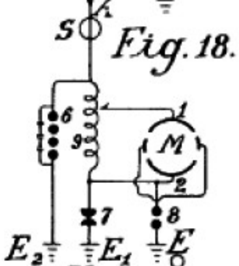
Fig. 18.

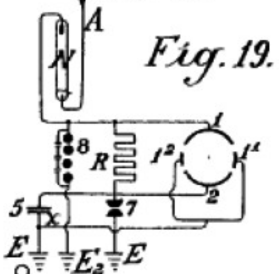
Fig. 19.

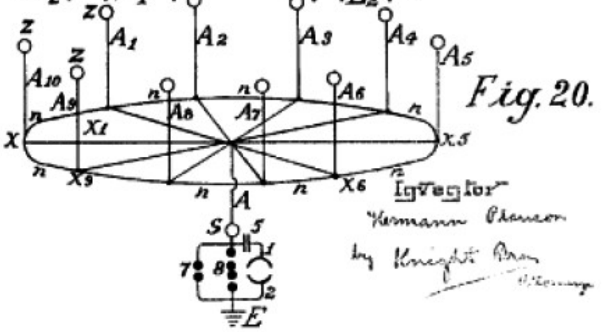
Fig. 20.

Inventor
Hermann Plauson
by Knight Bros
Attorneys

As however, the entire oscillation energy can thereby not act on the motor it is better to carry out the same system according to the diagram 14. The diagram 14 differs from the foregoing only by the motor not being directly metallically connected with the segments of the commutator, but only a primary coil 9 which induces in a secondary coil 10, current which feeds the motor M and takes the place of the rotor. By this arrangement a good transforming action is obtained, a loose coupling and also an oscillation circuit without a spark gap.

In Figure 15 the motor is not purely inductive as in 14, but directly metallically branched off from the primary coil (at x and x^1) after the principle of the auto-transformer.

In Figure 16 instead of an inductance a condenser 6 is in similar manner, and for the same object inserted between the segments 3a and 4a. This has the advantage that the segments 3a and 4a need not be made of solid metal but may consist of spiral coils whereby a more exact regulation is possible and further motors of high inductance may be employed.

The arrangements of Figures 17, 18 and 19 may be employed for use with resonance and particularly with induction condenser motors; between the large stator induction condenser surfaces, small reversing pole condenser surfaces, mall reversing pole condensers are connected, which, as may be seen from Figures 17, 18 and 19 are led together to earth. Such reversing poles have the advantage that with large quantities of electrical energy the spark formation between the separate oscillation circuits ceases.

Figure 19 shows a further method which prevents electromagnetic oscillations of high number of alternations formed in the oscillation circuit striking back to the aerial conductor. It is based on the well known principle that a mercury lamp, one electrode of which is formed of mercury, the other of solid metal such as steel allows an electric charge to pass in only one direction from the mercury to the steel and not vice versa. The mercury electrode of the vacuum tube N is therefore connected with the aerial conductor and the steel electrode with the oscillation circuit. From this it results that charges can pass only from the aerial through the vacuum tube to the oscillation circuit, but not vice versa. Oscillations which are formed on being transformed in the oscillation circuit cannot pass to the aerial conductor.

In practice these vacuum tubes must be connected behind an electromagnet as the latter alone affords no protection against the danger of lightning.

As regards the use of spark gaps, all arrangements as used for wireless telegraphy may be used. Of course the spark gaps in large machines must have a sufficiently large surface. In very large stations they are cooled in liquid carbonic acid or better still in liquid nitrogen or hydrogen; in most cases the cooling may also take place by means of liquefied low homologues of the metal series or by means of hydrocarbons the freezing point of which lies at between –90° C and –40° C. The spark gap casing must also be insulated and be of sufficient strength to be able to resist any pressure which may arise. Any undesirable excess super-pressure which may be formed must be automatically let off. I have employed wit very good results mercury electrodes which were frozen in liquid carbonic acid, the cooling being maintained during the operation from the outside through the walls.

Figure 20 is one of the simplest forms of construction of an aerial network in combination with collectors, transformers and the like illustrated diagrammatically. E is here the earth wire, 8 the safety spark gap, 7 the working spark gap, 1 and 2 the stator surfaces of the motor, 5 a condenser battery, S the protective magnet which is connected with the coil in aerial conductor, A^1 to A^{10} aerial antennae with collecting balloons, N horizontal collecting or connecting wire from which, to the center a number of connections run.

The actual collectors consist of metal sheaths preferably made of an aluminum magnesium alloy, and are filled with hydrogen or helium and are attached t copper plated steel wires. The size of the balloon is selected so that the actual weight of the balloon and the weight of the conducting wire is supported thereby. On top of the balloon aluminum spikes, made and gilded in a special manner hereinafter described, are arranged in order to produce a conductor action. Small quantities of radium preparations, more particularly polonium-ionium or meso-thorium preparations considerably increase the ionization, and therewith the action of these collectors.

In addition to metal balloons, fabric balloons which are superficially metal coated according to Schoop's metal spraying process, may also be employed. A metallic surface may also me produced by lacquering with metallic bronzes, preferably according to Schoop's spraying process or lacquering with metallic bronze powders in two electrical series of widely different metals, because thereby the collecting effect is considerably increased.

Instead of the ordinary round balloons, elongated cigar shaped ones may be employed. In order also to utilize the frictional energy of the wind, patches or strips of non-conducting substances which produce electricity by friction, may be attached to the metalized balloon surfaces. The wind will impart a portion of its energy in the form of frictional electricity, to the balloon casing, and thereby the collecting effect is substantially increased.

In practice however, very high towers (up to 300 meters is fully admissible) may be employed as antennae. In these towers copper tubes rise freely further above the top of the tower. A gas lamp secured against the wind is then lit at the point of the copper tube and a netting is secured to the copper tube over the flame of this lamp to form a collector. The gas is conveyed through the interior of the tube up to the summit. The copper tube must be absolutely protected from moisture at the place at which it enters the tower and also rain must be prevented running down the walls of the tower which might lead to a bad catastrophe. This is done by bell shaped enlargements which expand downwards, being arranged in the tower in the form of high voltage insulators of Siamese pagodas.

Special attention must be devoted to the foundations of such towers. They must be well insulated from the ground, which may be obtained by first embedding a layer of concrete in a box form to a sufficient depth in the ground and inserting in this an asphalt lining and then glass bricks cast about 1 or 2 meters in thickness. Over this in turn there is a ferro-concrete layer in which alone the metal foot of the tube is secured. This concrete block must be at least 2 meters from the ground and be fully protected at the sides by a wooden covering, from moisture. In the lower part of the tower a wood or glass house for the large condenser batteries or for the motors may be constructed. In order to lead the earth connection to the ground water, a well insulated pit constructed of vitreous brick, must be provided. Several such towers are erected at equal distances apart and connected with a horizontal conductor. The horizontal connecting wires may either run directly from tower to tower or be carried on bell shaped insulators similar to those in use for high voltage conductors. The width of the network may be of any suitable size and the connection of the motors can take place at any suitable places.

Fig. 21. Fig. 22.

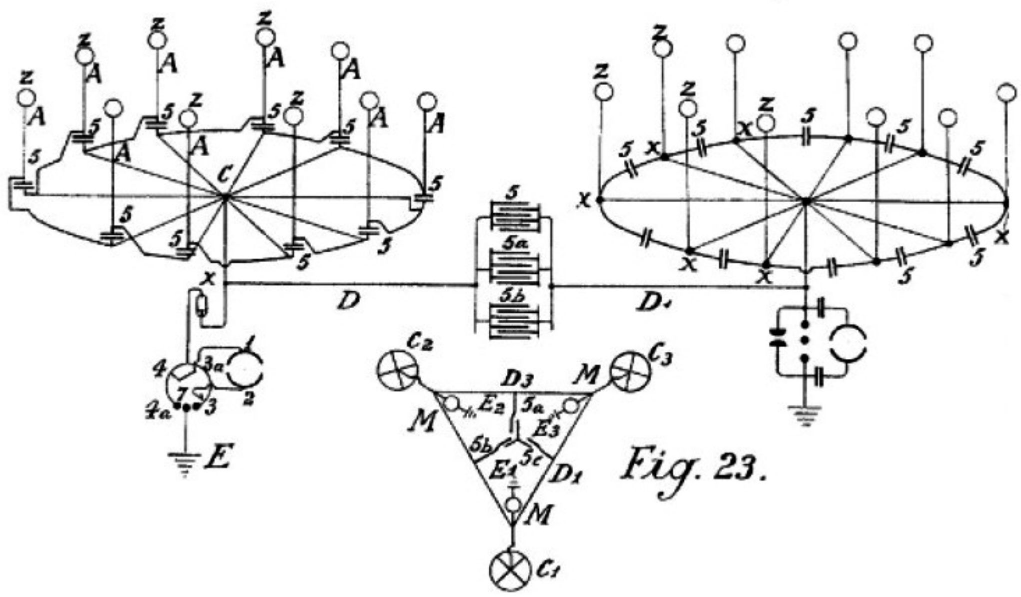

Fig. 23.

In order to collect large quantities of electricity with few aerials it is well to provide the aerial conductor with batteries of condensers as shown in Figures 21 and 22. In Figure 21 the batteries of condensers 5 are connected on the one hand with the aerial electricity collectors Z by the aerial conductor A, and on the other hand interconnected in series with an annular conductor from which horizontal conductors run to the connecting points C to which the earth wire is connected.

Figure 22 shows a similar arrangement, Should two such series of antennae rings be shown by a voltmeter to have a large difference of potential (for example, one in the mountains and one in the plain) or even of different polarity these differences may be compensated for by connecting sufficiently large condenser batteries (5, 5a, 5b) by means of Maji star conductors D and D^1. In Figure 23 a connection of three such rings of collectors to form a triangle with a central condenser battery is illustrated.

The condenser batteries of such large installations must be embedded in liquid gases or in liquids freezing at very low temperatures. In such cases a portion of the atmospheric energy must be employed for liquefying these gases. It is also preferable to employ pressure. By this means the condenser surfaces may be diminished, and still allow for large quantities of energy to be stored, secure against breakdown. For smaller installation the immersing of the condensers in well-insulated oil or the like, suffices. Solid substances on the other hand cannot be employed as insulators.

The arrangement in the diagrams hitherto described was always such that the condenser batteries were connected with both poles directly to the aerial condensers. An improved diagram of the connections for obtaining atmospheric electricity for the condenser batteries has however, been found to be very advantageous. This arrangement consists in that they are connected only by one pole (unipolar) to the collecting network. Such a method of arrangement is very important, as by means of it a constant current and an increase of the normal working pressure or voltage is obtained. If for example a collecting balloon aerial which is allowed to rise to a height of 300 meters, shows 40,000 volts above earth voltage, in practice it has been found that the working voltage (with a withdrawal of the power according to the method hereinbefore described by means of oscillating spark gaps and the like) is only about 400 volts. If however, the capacity of the condenser surfaces be increased, which capacity in the above mentioned case was equal to that of the collecting surface of the balloon aerials, to double the amount, by connecting the condenser batteries with only one pole, the voltage rises under an equal withdrawal of current up to and beyond 500 volts. This can only be ascribed to the favorable action of the connecting method.

In addition to this substantial improvement it has also been found preferable to insert double inductances with electromagnets and to place the capacities preferably between two such electromagnets. It has also been found that the useful action of such condensers can be further increased if an induction coil be connected as inductive resistance to the unconnected pole of the condenser, or still better if the condenser itself be made as an induction condenser. Such a condenser may be compared with a spring which when compressed carries in itself accumulated force, which it again gives off when released. In charging, a charge with reversed sign is formed at the other free condenser pole, and if through the spark gap a short circuit results, the accumulated energy is again given back since now new quantities of energy are induced at the condenser pole connected with the conductor network, which in fact charges with opposite signs to that at the free condenser pole. The new induced charges have of course the same sign as the collector network. The whole voltage energy in the aerial is thereby however increased. In the same space of time larger quantities of energy are accumulated than is the case without such inserted condenser batteries.

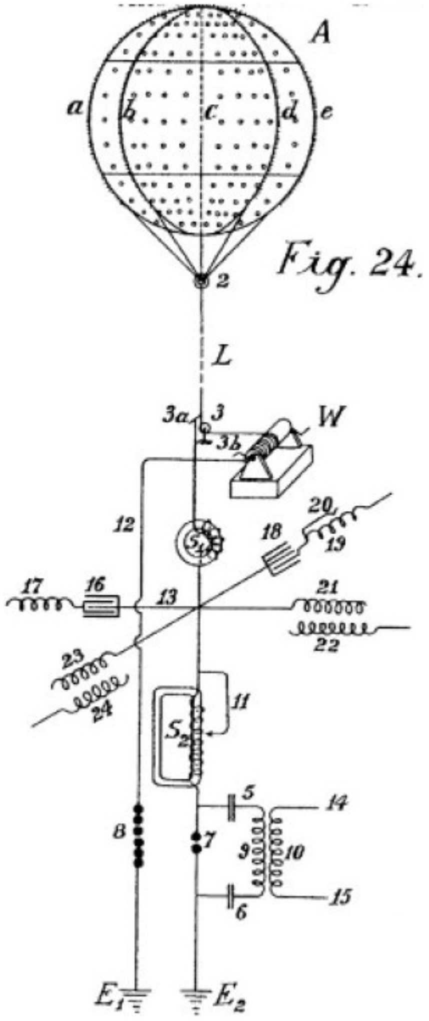

Fig. 24.

Fig. 25.

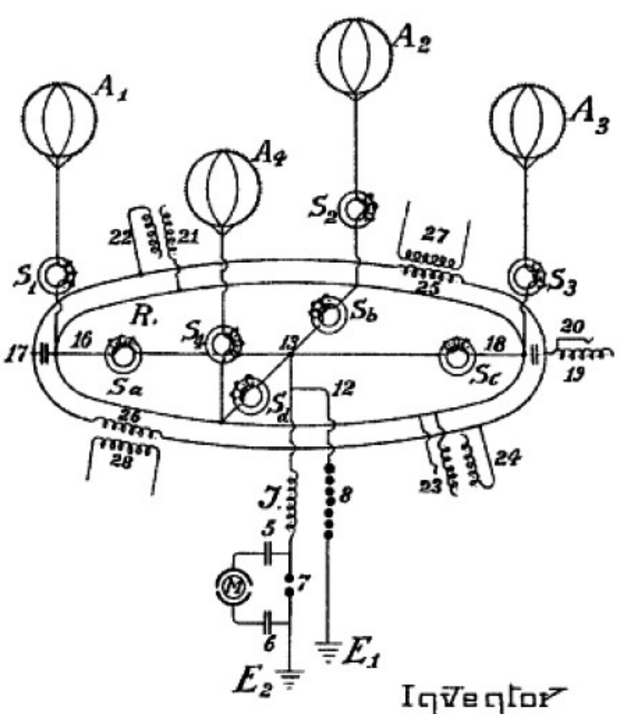

In Figures 24 and 25 two different diagrams of connections are more exactly illustrated, Figure 24 shows a collecting balloon and the diagram of the connections to earth. Figure 25 shows four collecting balloons and the parallel connection of the condenser batteries belonging thereto.

A is the collecting balloon made of an aluminum magnesium alloy (electron metal, magnalium) of a specific gravity of 1.8 and a thickness of plate 0.1 to 0.2 mm. Insider there are eight strong vertical ribs of T-shaped section about 10 to 20 mm in height and about 3 mm in thickness with the projecting part directed inwards (indicated by a, b, c, d and so forth); they are riveted together to form a firm skeleton and are stiffened in a horizontal direction by two cross ribs. The ribs are further connected with one another internally and transversely by means of thin steel wires, whereby the balloon obtains great power of resistance and elasticity. Rolled plates of 0.1 to 0.2 mm in thickness made of magnalium alloy are then either soldered or riveted on this skeleton so that a fully metallic casing with smooth external surface is obtained Well silvered or coppered aluminum plated steel wires run from each rib to the fastening ring 2. Further, the coppered steel hawser L preferably twisted out of separate thin wires (shown in dotted lines in Figure 24) and which must be long enough to allow the balloon to rise in the desired height, leads to a metal roller or pulley 3 and from thence to a winch W, well insulated from the earth. By means of this winch, the balloon, which is filled with hydrogen, or helium, can be allowed to rise to a suitable height (300 to 5000 meters) and brought to the ground for recharging or repairs.

The actual current is taken directly through a friction contact from the metal roller 3 or from the wire, or even from the winch or simultaneously from all three by means of brushes (3, 3a and 3b). Beyond the brushes the conductor is divided, the paths being: --- firstly over 12 to the safety spark gap 8, from thence to the earth conductor E^1, and secondly over electromagnet S^1, point 13, to a second loose electromagnet having an adjustable coil S^2, then to the spark gap 7 and to the second earth conductor E^2. The actual working circuit is formed through the spark gap 7, condensers 5 and 6, and through the primary coil 9; here the static electricity formed by oscillatory discharges is accumulated and converted into high frequency electromagnetic oscillations. Between the electromagnets S^1 and S^2 at the crossing point 13, four condenser batteries are introduced which are only indicated diagrammatically in the drawings each by one condenser. Two of these batteries (16 and 18) are made as plate condensers and prolonged by regulating induction coils or spirals 17 and 19 while the two others (21 and 23) are induction condensers. As may be seen from the drawings each of the four condenser batteries 16, 18, 21 and 23 is connected by only one pole to the aerial or to the collector conductor. The second poles 17, 19, 22 and 24 are open. In the case of plate condensers having no inductive resistance an induction coil is inserted. The object of such a spiral or coil is the displacement of phase by the induction current by 1/4 periods, whilst that of the charging current of the condenser poles which lie free in the air, works back to the collector aerial. The consequence of this is that in discharges in the collector aerial the back inductive action of the free poles allows a higher voltage to be maintained in the aerial collecting conductor than would otherwise be the case. It has also been found that such a back action has an extremely favorable effect on the wear of the contacts. Of course the inductive effect may be regulated at will within the limits of the size of the induction coil, the length of the coil in action being adjustable by means of wire connection without induction (see Figure 24, No. 20).

S^1 and S^2 may also be provided with such regulating devices in the case of S^2 (illustrated by 11). If excess voltage be formed it is conducted to earth through the wire 12 and spark gap 8 or through any other suitable apparatus, since this formation would be dangerous for the other apparatus.

The small circles on the collector balloon indicate places at which zinc amalgam or gold amalgam or other photoelectric acting metals in the form of small patches in extremely thin layers (0.01 to 0.05 mm in thickness) are applied to the entire balloon as well as in greater thickness to the conducting network. The capacity of the collector is thereby considerably strengthened at the surface. The greatest possible effect in collecting may be obtained by polonium amalgams and the like. On the surface of the collector balloon metal points or spikes are also fixed along the ribs, which spikes serve particularly for collecting the collector charge. Since it is well known that the resistance of the spikes is less the sharper the spike is, for this purpose it is therefore extremely important to employ as sharp spikes as possible. Experiments made as regards these have shown that the formation of the body of the spike or point also plays a large part, for example, spikes made of bars or rollers with smooth surface, have a many times greater point resistance as collector accumulator spikes than those with rough surfaces. Various kinds of spike bodies have been experimented with for the collector balloons hereinbefore mentions. The best results were given by spikes which were made in the following way. Fine points made of steel, copper, nickel, or copper and nickel alloys, were fastened together in bundles and then placed as anode with the points in a suitable electrolyte (preferably in hydrochloric acid or muriate of iron solutions) and so treated with weak current at 2 to 3 volts pressure. After 2 to 3 hours according to the thickness of the spikes or pins the points become extremely sharp and the bodies of the spikes have a rough surface. The bundle can then be removed and the acid washed off with water. The spikes are then placed as cathode in a batch consisting of solution of gold, platinum, iridium, palladium or wolfram salts or their compounds and coated a the cathode galvanically with a thin layer of precious metal, which must however be sufficiently firm to protect them from atmospheric oxidation.

Such spikes act at a 20-fold lower voltage almost as well as the best and finest points made by mechanical means. Still better results are obtained if polonium or radium salts are added to the galvanic bat when forming he protective layer or coating. Such pins have a low resistance at their points and even at one volt and still lower pressures have an excellent collector action.

In Figure 24 the three unconnected poles are not connected with one another in parallel. That is quite possible in practice without altering the principle of the free pole. It is also preferable to interconnect in parallel to a common collector network, a series of collecting aerials.

Figure 25 shows a diagram for such an installation. A^1, A^2, A^3, A^4 are four metal collector balloons with gold or platinum coated spikes which are electrolytically made in the presence of polonium emanations or radium salts, which spikes or needles are connected over four electromagnets S^1, S^2, S^3, S^4, through an annular conductor R. From this annular conductor four wires run over four further electromagnets Sa, Sb, Sc, Sd, to the connecting point 13. There the conductor is divided, one branch passing over 12 and the safety spark gap 8 to the earth at E^1, the other over inductive resistance J and working spark gap 7 to the earth at E^2. The working circuit, consisting of the condenser 5 and 6 and a resonance motor M, such as hereinbefore described, is connected in proximity round the sparking gap section 7.

Instead of directly connecting the condenser motor of course the primary circuit for high frequency oscillatory current may also be inserted.

The condenser batteries are connected by one pole to the annular conductor R and can be either inductionless (16 and 18) or made as induction condensers as shown by 21 and 23. The free poles of the inductionless condensers are indicated by 17 and 19, those of the induction condensers by 22 and 24. As may be seen from the drawings all these poles 17, 22, 19, 24 may be interconnected in parallel through a second annular conductor without any fear that thereby the principle of the free pole connection will be injured. In addition to the advantages already set forth the parallel connection also allows of an equalization of the working pressure in the entire collector network. Suitable constructed and calculated induction coils 25 and 26 may also be inserted in the annular conductor of the free poles, by means of which a circuit may be formed in the secondary coils 27 and 28 which allows current produced in this annular conductor by fluctuations of the charges of the like appearances to be measured or otherwise utilized.

According to what has been hereinbefore stated separate collector balloons may be connected at equidistant stations distributed over the entire country, either connected directly with one another metallically or by means of intermediate suitably connected condenser batteries through high voltage conductors insulated from earth. The static electricity is converted through a spark gap into dynamic energy of a high number of oscillations and may in such form be coupled as a source of energy b y means of a suitable method f connecting, various precautions being observed and with special regulations. The wires leading from the collector balloons have hitherto been connected through an annular conductor without this endless connection, which can be regarded as an endless induction coil, being able to exert any action on the whole conductor system.

It has now been found that if the network conductor connecting the aerial collector balloons with one another is not made as a simple annular conductor, but preferably short circuited in the form of coils over a condenser battery or spark gap or through thermionic tubes or valves or audions, then the total collecting network exhibits quite new properties. The collection of atmospheric electricity is thereby not only increased but an alternating field may be easily produced in the collector network. Further, the atmospheric electrical forces showing themselves in the higher regions may also be directly obtained by induction. In Figures 26 and 28 a form of construction is shown on the basis of which the further foundations of the method will be more particularly explained.

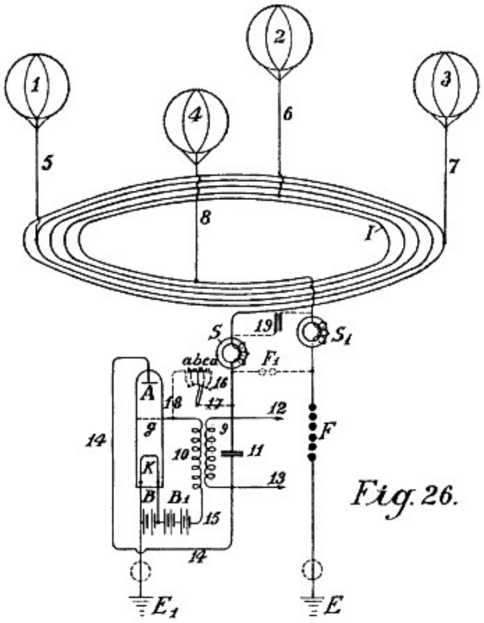

Fig. 26.

In Figure 26, 1, 2, 3, 4 are metal collector balloons, 5, 6, 7, 8 their metallic aerial conductors and I the actual collector network. This consists of five coils and is mounted on high voltage insulators in the air, in high voltage masts (or with a suitable construction of cable embedded in the earth). One coil has a diameter of 1 or 100 km or more. S and S^1 are two protective electromagnets, F the second safety section against excess voltage, E its earth conductor and E1 the earth conductor of the working section. When an absorption of static atmospheric electricity is effected through the four balloon collectors, the current in order to reach the earth connection E1 must flow spirally through the collector network over the electromagnet S, primary induction coil 9, conductor 14, anode A of the audion tube, incandescent cathode K, as the way over the electromagnet and safety spark gap F offers considerably greater resistance. Owing to the fact that the accumulated current flows in one direction, an electromagnetic alternating field is produced in the interior of the collector network coil, whereby the whole free electrons are directed more or less into the interior of the coil. An increased ionization of the atmosphere is thereby produced. In consequence of this the points mounted on the collector balloon show a considerably reduced resistance and therefore increased static charges between the points on the balloon and the surrounding atmosphere are produced. The result of this is a considerably increased collector effect.

A second effect which could not be obtained otherwise is obtained by the electromagnetic alternating field which running parallel to the earth surface, acts more or less with a diminishing or increasing effect on the earth magnetic field, whereby in the case of fluctuations in the current a return induction current of reversed sign is always produced in the collector coil by earth magnetism. Now if, however, a constantly pulsating continuous alternating field is produced as stated in the above collector network I, an alternating current of the same periodicity is produced also in the collecting network coil. As the same alternating field is further transmitted to the aerial balloon, the resistance to its points is thereby considerably reduced, whilst the collector action is considerably increased. A further advantage is that positive electrons which collect on the metal surfaces during the conversion into dynamic current produce a so-called drop of potential of the collector area. As an alternating field is present, the negative ions surrounding the collectors surfaces, when discharge of the collector surfaces takes place produce by the law of induction, an induction of reversed sign on the collector surface and so forth (that is to say again a positive charge). In addition to the advantages hereinbefore set forth, the construction of connecting conductors in coil form when of sufficiently large diameter, allows of a utilization of energy arising in higher regions also in the simplest way. As is well known electric discharges frequently take place at very great elevations which may be observed as St Elmo’s fire or northern lights. These energy quantities have no been available to be utilized up to now. By this invention all these kinds of energy, as they are of an electromagnetic nature and the direction of axis of the collector coils stands at right angles to the earth’s surface, can be more or less absorbed in the same way as a receiver in wireless telegraphy absorbs waves coming from a far distance. With a large diameter of the spiral it is possible to connect large surfaces and thereby to take up also large quantities of energy.

It is well known that large wireless stations in the summer months, and also in the tropics are very frequently unable to receive the signals in consequence of interruptions which are caused by atmospheric electricity, and this takes place with vertical coils of only 40 to 100 meters diameter. If on the contrary horizontal coils of 1 to 100 km diameter be employed very strong currents may be obtained through discharges which are constantly taking place in the atmosphere. Particularly in the tropics or still better in the polar regions where the northern lights are constantly present, large quantities of energy may probably be obtained in this way. A coil with several windings should act the best. In similar manner any alteration of the earth magnetism should act inductively on such a coil.

It is not at all unlikely that earthquakes and spots on the sun will also produce an induction in such collector coils of sufficient size. In similar manner this collector conductor will react on earth current more particularly when they are near the surface of the earth or even embedded in the earth. By combining the previous kind of current collectors so far as they are adapted for the improved system with the improved possibilities of obtaining current the quantities of free natural electricity which are to be obtained in the form of electricity are considerably increased.

In order to produce in the improved collector coil uniform current oscillations of an undamped nature so-called audion high vacuum or thermionic tubes of suitable connection are employed instead of the previously known spark gaps (Figure 26, Nos. 9-18). The main aerial current flows through electromagnet S (which in the case of a high number of alternations is not connected here but in the earth conductor E^1) and may be conveyed over the primary coils in the induction winding through wire 14 to the anode A of the high vacuum grid tube. Parallel with the induction resistance 9 a regulating capacity of suitable size, such as condenser 11 is inserted. In the lower part of the vacuum grid tube is arranged the incandescent filament or the cathode K which is fed through a battery B. From the battery B two branches run, one to the earth conductor E^1 and the other through battery B^1 and secondary coil 10 to the grid above *g* in the vacuum tube. By the method of connections shown in dotted lines, a desired voltage at the grid electrode *g* may also be produced through the wire 17 which is branched off from the main current conductor through switches 16 and some small condensers (*a, b, c, d*) connected in series, and conductor 18, without the battery B^1 being required.

The action of the entire system is somewhat as follows: --

On the connecting conductor of the aerial collector network being short circuited to earth, the condenser pole 11 is charged and slightly damped oscillations are formed in the short circuited existing oscillation circuit formed of the condenser 11 and self inductance 9. In consequence of the coupling through coil 10, fluctuations of voltage take place in the grid circuit 15 with the same frequency, which, fluctuations in turn influence the strength of the electrode current passing through the high vacuum amplifying tube and thus produce current fluctuations of the same frequency in the anode circuit. A permanent supply of energy to the oscillation circuits 9 and 10 consequently takes place, until a condition of balance is set up, in which the consumed oscillation energy is equal to that absorbed. Thereby constant undamped oscillations are now produced in the oscillation circuits 9-11.

For regular working of such oscillation producers high vacuum strengthening tubes are necessary and it is also necessary that the grid and anode voltages shall have a phase difference of 180° so that if the grid is negatively charged, then the anode is positively charged and vice versa. This necessary difference of phase may be obtained by most varied connections, for example, by placing the oscillation circuit in the grid circuit or by separating the oscillation circuit and inductive coupling from the anodes and the grid circuit and so forth.

A second important factor in this way of converting static atmospheric electricity into undamped oscillations is that care must be taken hat the grid and anode voltages have a certain relation to one another; the latter may be obtained by altering the coupling and a suitable selection of the self-induction in the grid circuit, or as shown by dotted lines 16, 17, 18 by means of a larger or smaller number of condensers of suitable size connected in series; in this case the battery B^1 may be omitted. With a suitable selection of the grid potential a glow discharge takes place between the grid *g* and the anode A, and accordingly at the grid there is a cathode drop and a dark space is formed. The size of this cathode drop is influenced by the ions which are emitted in the lower space in consequence of shock ionization of the incandescent cathodes K and pass through the grid in the upper space. On the other hand the number of ions passing through the grid is dependent on the voltage between the grid and the cathode. Thus is the grid voltage undergoes periodic fluctuations (as in the present case) the amount of the cathode drop at the grid fluctuates and consequently the internal resistance of the tube correspondingly fluctuates, so that when a back coupling of the feed circuit with the grid circuit takes place, the necessary means are afforded for producing undamped oscillations and of taking current, according to requirements from the collecting conductor.

The frequency of the undamped oscillations produced is with a suitably loose coupling equal to the self-frequency of the oscillation circuits 9 and 10. By a suitable selection of the self induction of the coil 9 and capacity 11 it is possible to extend from frequencies which produce electromagnetic oscillation of only a few meters wavelength down to the lowest practical alternating current frequency. For large installations a suitable number of frequency-producing tubes of the well-known high vacuum transmission tubes of 0.5 to 2 kw in size may be connected in parallel so that in this respect no difficulty exists.

The use of such tubes for producing undamped oscillations, and also the construction and method of inserting such transmission tubes in an accumulator or dynamo circuit is known and also that such oscillation producing tubes only work well at voltages of 1000 up to 4000 volts, so that on the contrary their use at lower voltages is considerably more difficult. By the use of high voltage static electricity this method of producing undamped oscillations as compared with that through spark gaps must be regarded as an ideal solution particularly for small installations of outputs of from 1 to 100 kw.

By the application of safety spark gaps, with interpolation of electromagnets, not only is short-circuiting avoided but also the taking up of current is regulated. Oscillation producers inserted in the above way form a constantly acting electromagnetic alternating field in the collector coil, whereby as already stated, a considerable accumulating effect takes place. The withdrawal wire or working wire is connected at 12 and 13, but current may be taken by means of a secondary coil which is firmly or movable mounted in any suitable way inside the large collector coil, i.e., in its electromagnetic alternating field, so long as the direction of its axis runs parallel with that of the main current collecting coil.

In producing undamped oscillations of a high frequency (50,000 per second or more) in the oscillation circuits 9 and 11, electromagnets S and S^1 must be inserted if the high frequency oscillations are not to penetrate the collector oil, between the oscillation producers and the collector coil. In all other cases they are connected shortly before the earthing (as in Figures 27 and 28).

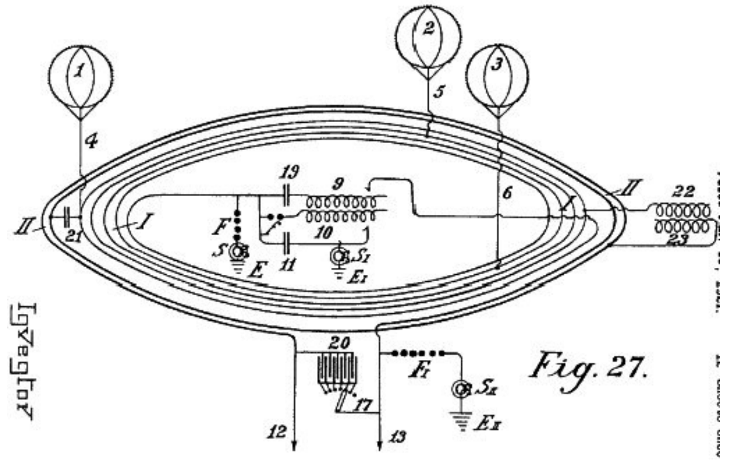

Fig. 27.

In Figure 27 a second method of construction of the connecting conductor of the balloon aerials is illustrated in the form of a coil. The main difference consists in that in addition to the connecting conductor I another annular conductor II is inserted parallel to the former on the high voltage masts in the air (or embedded as a cable in the earth) but both in te form of a coil. The connecting wire of the balloon aerials is indicated as a primary conductor and also as a current producing network; the other is the consumption network and is not in unipolar connection with the current producing network.

In Figure 27 the current producing network I is shown with three balloon collectors 1, 2, 3 and aerial conductors 4, 5, 6; it is short-circuited through condenser 19 and inductance 9. The oscillation forming circuit consists in this diagram of spark gap f, inductance 10, and condenser 11; the earth wire E, is connected to earth over electromagnet S^1. F s the safety spark gap which is also connected to earth through a second electromagnet S at E. On connecting up the condenser circuit 11 this is charged over the spark gap f whereby an oscillatory discharge is formed. This discharging current acts through inductance 10 on the inductively coupled secondary 9, whereby in the producing network a modification of the potential of the condenser 19 is produced. The consequence of this is that oscillations arise in the coil shaped producer network. These oscillations induce a current in the secondary circuit II, which has a smaller number of windings and a less resistance, the voltage of which, according to the proportion of the number of windings and of the ohmic resistance, is considerably lower whilst the current strength is greater.

In order to convert the current thus obtained into current of an undamped character, and to tune its wavelengths, a sufficiently large regulatable capacity 20 in inserted between the ends 12 and 13 of the secondary conductor II. Here also current may be taken without an earthy conductor, but it is advisable to insert a safety spark gap E1 and to connect this with the earth over an electromagnet S2.

The producer network may be connected with the working network II over an inductionless condenser 21 or over an induction condenser 22, 23. In this case the secondary conductor is unipolarly connected with the energy conductor.

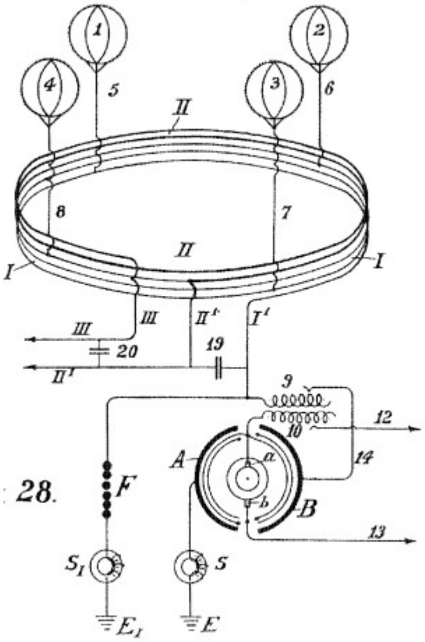

28.

In Figure 28 the connecting conductor between the separate accumulator balloons is carried out according to the autotransformer principle. The collecting coil connects four aerial balloons 1, 2, 3, 4, the windings of which are not made side by side but one above the other. In Figure 28 the collector coil I is shown with a thin line, the metallically connected prolongation coils II with a thick line. Between the ends I^1 and II^1 of the energy network I a regulating capacity 19 is inserted. The wire I^1 is connected with the output wire and with the spark gap F.

As transformer of the atmospheric electricity an arrangement is employed which consists in using rotary pairs of condensers in which the one stator surface B is connected with the main current, whilst the other A is connected with the earth pole. Between these pairs of short-circuited condensers are caused to rotate from which the converted current can be taken by means of two collector rings and brushes, in the form of an alternating current, the frequency of which is dependent on the number of balloons and the revolutions of the rotor. As the alternating current formed in the rotor can act, in this improved method of connection described in this invention, through coils 1 on the inductance 9, an increase or diminution of the feed current in I can be obtained according to the direction of the current by back induction. Current oscillations of uniform rhythm thereby result in the coil shaped windings of the produce network.

As the ends of this conductor are short-circuited through the regulatable condenser 19 these rhythms produce short-circuited undamped oscillations in the energy conductor, the periodicity and wave lengths of which oscillations can be adjusted according to desire by altering the capacity 19 to a given wavelength and therewith also to a given frequency. These currents may also be employed in this form directly as working current through the conductors II^1 and III. By inserting the condenser 20 a connection between these conductors may also be made, whereby harmonic oscillations of desired wavelengths are formed. By this means quite new effects as regards current distribution are obtained. The withdrawal of current can even take place without direct wire connection if, at a suitable point in the interior of the producing network (quite immaterially whether this has a diameter of 1 or 100 km) a coil tuned to these wavelengths and of the desired capacity is firmly or movably mounted in the aerial conductor in such a way that is axial direction is in parallel with that of the collector coil. In this case a current is induced in the producing network, the size of which is dependent on the total capacity and resistance and also on the periodicity employed. A possibility is thereby afforded in future, of taking energy from the producer network by wireless means. As thereby in addition to atmospheric electricity also magnetic earth currents and energy from the higher atmosphere (at leas partially may be simultaneously obtained, this last system for collecting the atmospheric energy is of particular importance for the future.

Of course everywhere instead of spark gaps grid vacuum tubes may be employed as producers for undamped oscillations. The separate coils of the producer network with large diameters may be connected with one another through separate conductors all in parallel or all in series or in groups in series. By regulating the number of oscillations and also the extent of the voltage more or less large collector coils of this kind may be employed. The coils may also be divided spirally over the entire section. The coils may be carried out in angular form or also in triangular, quadrangular, hexagonal or octagonal form.

Of course wires may be carried from a suitable place to the center or also laterally which serve the current waves as guides. This is necessary when the currents have to be conducted over mountains and valley and so forth. In all these cases the current must be converted into a current of suitable periodicity.

As already hereinbefore mentioned separate collecting balloons may be directly metallically interconnected at equidistant stations distributed over the entire country or may be connected by interpolation of suitable condenser batteries by means of high voltage conductors. The static electricity is converted through a spark gap into dynamic energy of a high number of oscillations, and could then in such forms, wit a suitable arrangement of the connections, observing various measures of protection, be employed as source of energy after separate or special regulation.

According to this invention in order to increase the collecting effect of the balloon in the aerial collector conductor or in the earth wire, radiating collectors are employed. These consist either of incandescent metal or oxide electrodes in the form of vacuum grid tubes, or electric arcs (mercury and the like electrodes), Nernst lamps, or finally flames of various kinds may be simply connected with the respective conductor.

It is well known that energy can be drawn of from a cathode consisting of an incandescent body opposite an anode charged with positive electricity (vacuum grid tube). Hitherto however, a cathode was always first directly placed opposite an anode, and secondly the system always consisted of a closed circuit.

Now if we dispense with the ordinary ideas in forming light or flame arcs in which a cathode must always stand directly opposite an anode, charged to a high potential or another body freely floating in the air, or regard the incandescent cathode only as a source of unipolar discharge (which represent group and point discharges in electrostatic machines similar to unipolar discharges), it may be ascertained that incandescent cathodes and less perfectly all incandescent radiators, flames and the like admit of relatively large current densities and allow large quantities of electric energy to radiate into the open space in the form of electron streams as transmitters.

The object of this invention is as described below, if such incandescent oxide electrodes or other incandescent radiators or flames are not freely suspended in space but connected metallically with the earth so that they can be charged with negative terrestrial electricity, these radiators possess the property of absorbing the free positive electrical charges contained in the air space surrounding them (that is to say of collecting them and conducting them to earth). They can therefore serve as collectors and have, in comparison to the action of the spikes, or points, a very large radius of action R; the effective capacity of these collectors is much greater than the geometrical capacity (R_{o-}) calculated in an electrostatic sense.

Now as our earth is surrounded as is well known with an electrostatic field and the difference of potential

$$\frac{\delta V}{\delta h}$$

of the earth field according to the latest investigations, is in summer about 60 to 100 volts and in winter 300 to 500 volts per meter of difference in height (δh), a simple calculation gives the result that when such a radiation collector or flame collector is arranged for example on the ground, and a second one is mounted vertically over it at a distance of 2000 meters and both are connected by a conducting cable, there is a difference in potential in summer of about 2,000,000 volts and in winter even of 6,000,000 volts and more.

According to Stefan Boltzmann's law of radiation, the quantity of energy which an incandescent surface (temperature T) of 1 sq cm radiates in a unit of time into the open air (temperature T_o) is expressed by the following formula:

$S = \delta (T^4 - T^4_o)$ watt./sq cm.

And the universal radiation constant δ is according to the latest researches of Ferry (*Annales de Chimie et de Physique*, 17: 267 [1909]) equal to 6.30 x 10^{-12} watt/sq cm.

Now if an in incandescent surface of 1 sq cm shows, as compared with the surrounding space a periodic fall of potential δV it radiates (independent of the current direction, that is to say of the sign) in accordance with the above formula, for example at a temperature of 3725° C an energy of 1.6 kw/sq cm/second. As for the radiation the same value can be calculated for the collection of energy, but reversed. Now, as carbon electrodes at the temperature of the electric arc support on the current basis a current density up to from 60 to 65 amperes per sq cm no difficulties will result in this direction in employing radiating collectors as accumulators.

If the earth be regarded as a cosmically insulated condenser in the sense of geometrical electrostatics *x* there results from the geometric (compare Edwald Rasch: *Das Elektrische Bogenlicht* [The Electric Arc Light], page 169) capacity of the earth according to Chwolson:

For negative charging 1.3 x 10^6 Coulomb

For negative potential V = 10 x 10^8 volts.

From this there results however, EJT = 24.7 x 10^{24} watt/sec. Now if it is desired to make a theoretic short circuit through an earthed flame collector this would represent an electric total work of about 79,500 x 10^{10} kilowatt years. As the earth must be regarded as a rotating mechanism which thermodynamically, electromagnetically and also kinematically coupled with the sun and star system by cosmic radiations and gravitation a diminution of the electric energy of the earth field is not to be feared. The energies which the incandescent collectors would withdraw from the earth field can only case by the withdrawal of motor work a lowering of the earth temperature (temperature $T_E = 300$) and reduce this to that of the world space (T = O) by using the entire energy. This is however not the case as the earth does not represent a cosmically entirely insulated system. On the contrary there is conveyed to the same according to the recent value corrected by Perry for the solar constants through the radiation from the sun an energy of 18,500 x 10^{10} kw. Accordingly, any lowering of the earth temperature (T_E) without a simultaneous lowering of the sun's temperature (T_s) would contradict Stefan Boltzmann's law of radiation.

$S = \delta (s^4 - T^4)$.

From this it must be concluded that if the earth temperature (T_E) sinks the total radiation S absorbed by the earth increases, and further also that the secular speed of cooling of the earth is directly dependent on that of the sun and the other radiators cosmically coupled with the sun and is connected most closely with these.

The incandescent radiation collectors may, according to this invention, be employed for collecting atmospheric electricity if they (1) are charged with the negative earth electricity (that is to say when they are directly connected by means of a metallic conductor with the earth) and (2) if large capacities (metal surfaces) charged with electricity are mounted opposite them as positive poles in the air. This is regarded as the main feature of the present invention as without these inventive ideas it would not be possible to collect with an incandescent collector, sufficiently large quantities of the electrical charges contained in the atmosphere as the technology requires; the radius of action of the flame collectors would also be too small, especially if it be considered that the very small surface density (energy density) (δ about = 2 x 7 . 10^9 St. E. per sq cm) does not allow of large quantities of charge being absorbed from the atmosphere.

x) Calculated according to Poisson's calculation:

$\Delta V = -4\pi\delta$; as here the alteration of the potential or potential gradients only takes place in the direction o the normal, this calculation assumes the simple form

$$\delta = \frac{1}{4\pi} \times \frac{\delta^2 V}{\delta n^2}$$

It has indeed already been proposed to employ flame collectors for collecting atmospheric electricity and it is known that their collecting effect is substantially greater opposite the points. It is however, not known that the quantities of current which could hitherto be obtained are too small for technical purposes. According to my experiments the reason for this is to be found in the too small capacities of the collector conductor poles. If such flame or radiating collectors have no or only small positive surfaces, their radius of action for large technical purposes is too small. If the incandescent collectors be constantly kept in movement in the air they may collect more according to the speed of the movement, but this again is not capable of being carried out in practice.

By this invention the collector effect is considerably increased by a body charged with a positive potential and of the best possible capacity being also held floating (without direct earth connection) opposite such an incandescent collector which is held floating in the air at a desired height. If for example, a collecting balloon of sheet metal or of metalized balloon fabric be caused to mount to 300 up to 3000 meters in the air and as positive pole it is brought opposite such a radiating collector connected by a conductor to the earth, quite different results are obtained.

The metallic balloon shell (with a large surface) is charged to a high potential by the atmospheric electricity. This potential is greater the higher the collecting balloon is above the incandescent collector. The positive electricity acts concentratedly on the anode floating in the air as it is attracted through the radiation shock ionization, proceeding from the incandescent cathode. The consequence of this is that the radius of action of the incandescent cathode collector is considerably increased and thereby also the collection effect of the collecting balloon surface. Further the large capacity of the anode floating in the air plays therefore an important part because it allows of the taking o large charges, and thereby a more uniform current is obtained even when there is a large consumption: this cannot be the case with small surfaces.

In the present case the metallic collecting balloon is a positive anode floating in the air and the end of the earth conductor of this balloon serves as positive pole surface opposite the surface of the radiating incandescent cathode, which in turn is charged with negative earth electricity being conductively connected to earth.

The process may be carried out by two such contacts (negative incandescent cathode and anode end of a capacity floating in the air) a condenser and an inductive resistance being switched on in parallel, whereby simultaneously undamped oscillations may be formed.

In very large installations it is advisable to connect two such radiating collectors in series. Thus an arc light incandescent cathode may be placed below on the open ground and an incandescent cathode which is heated by special electromagnetic currents be located high in the air. Of course for this the special vacuum Liebig tubes wit or without grids may also be employed. An ordinary arc lamp with oxide electrodes may be introduced on the ground and the positive pole is not directly connected with the collecting balloon, but through the upper incandescent cathode or over a condenser. The method of connecting the incandescent cathode floating in the air may be seen in Figures 29-33.

B is the air balloon, K a Cardan ring (connection with the hawser), C the balloon, L a good connecting cable, P a positive pole, N negative incandescent cathode, and E earth conductor.

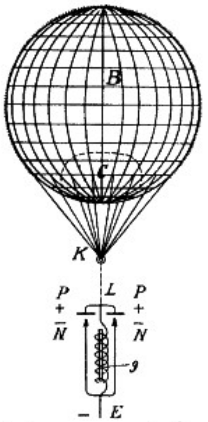

Fig. 29.

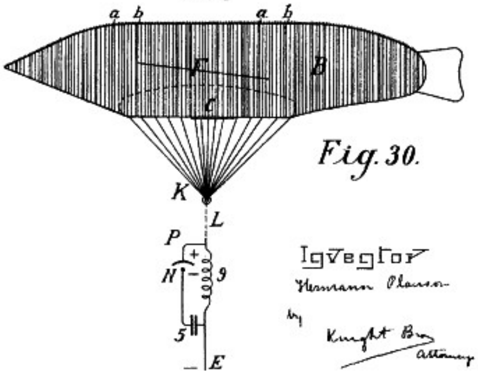

Fig. 30.

Inventor
Hermann Plauson
by
Knight Bros
attorney

Figure 29 represents the simplest form of construction. If electric oscillations are produced below on the ground by means of a carbon arc lamp or in other suitable way a considerably greater electric resistance is opposed to that in the direct way by inserting an electrical inductive resistance 9. Consequently between P and N, a voltage is formed, and as, over N and P only an inductionless ohmic resistance is present, a spark will spring over so long as the separate induction co-efficients and the like are correctly calculated. The consequence of this is that the oxide electrode (carbon or the like) is rendered incandescent and then shows as incandescent cathode an increased collecting effect. The positive poles must be substantially larger than the negative in order that they may not also become incandescent. As they are further connected with the large balloon area which has a large capacity and is charged at high voltage, an incandescent body which is held floating in the air and a positive pole which can collect large capacities is thereby obtained in the simplest way. The incandescent cathode is first caused to become incandescent by means of separate energy produced on the earth, and then maintained by the energy collected from the atmosphere.

Figure 30 only shows the difference that instead of a round balloon a cigar-shaped one (of metal or metalized fabric) may be employed and also a condenser 5 is inserted between the incandescent cathode and the earth conductor so that a short circuited oscillation circuit over P.N. 5 and 9 is obtained. This has the advantage that quite small quantities of electricity cause the cathode to become incandescent and much larger cathode bodies may be rendered incandescent.

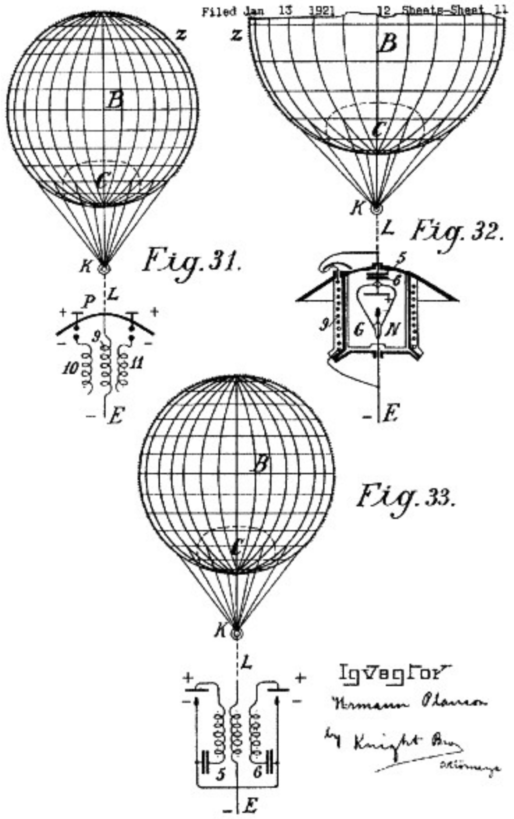

In this form of construction both the incandescent cathode and also the positive electrode may be enclosed in a vacuum chamber as may be seen in Figure 32. A cable L is carried well insulated through the cover of a vessel and ends in a condenser disc 5. The cover is arched in order to keep off the rain. The vessel is entirely or partially made of magnetic metal and well-insulated inside and outside. Opposite the disc 5 another disc 6 and on this again a positive pile of the vacuum tube g with the incandescent cathode (oxide electrode) N is arranged. The negative electrode is on the one hand connected with the earth conductor E, and on the other hand with the inductive resistance 9 which is also connected with the cable L with the positive pole and wound round the vessel in coils. The action is exactly the same as that in Figure 29 only instead of an open incandescent cathode one enclosed in vacuo is employed. As in such collectors only small bodies can be brought to incandescence in large installations a plurality of such vacuum tubes must be inserted in proximity to one another. According to the previous constructions Figures 31 and 33 are quite self evident without further explanations.

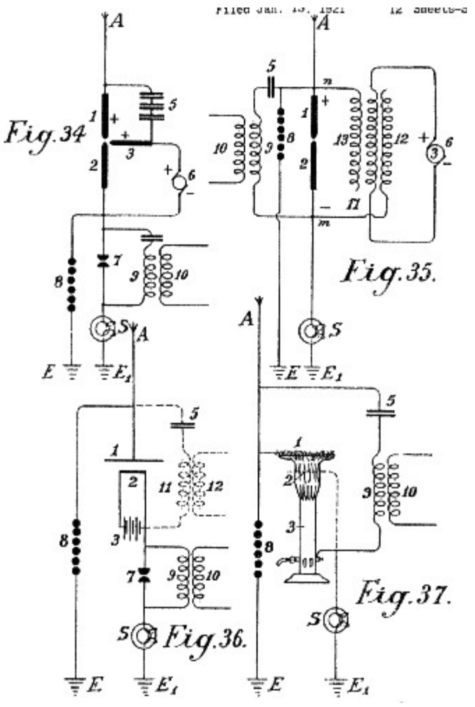

Figures 34-37 represent further diagrams of connections over radiating and flame collectors, and in fact, how they are to be arranged on the ground.

Figure 34 shows an arc light collector with oxide electrodes for direct current and its connection; Figure 35 a similar one for alternating current, Figure 36 an incandescent collector with a Nernst lamp and Figure 37 a similar one with a gas flame.

The positive pole 1 of the radiating collectors is always directly connected to the aerial collecting conductor A. In Figure 34 this is further connected over the condenser battery 5 with a second positive electrode 3. The direct current dynamo *b* produces current which flows over between the electrodes 3 and 2 as an arc light. On the formation of an arc the negative incandescent electrode 2 absorbs electricity from the positive poles standing opposite it and highly charged with atmospheric electricity and conveys the same to the working circuit. The spark gap 7, inductive resistance 9 and induction coil 10 are like the ones previously described. The protective electromagnet S guards the installation against earth circuiting, the safety spark gap 8 from excess voltage or overcharging.

In Figure 35 the connection is so far altered that the alternating current dynamo feeds the exciting coil 11 of the induction condenser. 12 is its negative and 13 its positive pole; if the coil 3 on the magnet core of the dynamo is correctly calculated and the periodicity of the alternating current is sufficiently high an arc light can be formed between the two poles 1 and 2. As the cathode 2 is connected with the negatively charged earth, and therefore always acts as a negative pole, a form of rectification of the alternating current produced by the dynamo 3 is obtained, the second half of the period is always suppressed. The working circuit may be carried out in the same way as in Figure 34; the working gap 7 may however be dispensed with, and instead thereof between the points m and n a condenser 5 and an induction resistance 9 may be inserted from which the current is taken inductively.

Figure 36 represents a form of construction similar to Figure 34 only that here instead of an arc lamp a Nernst incandescent body is employed. The Nernst lamp is fed through the battery 3. The working section is connected with the negative pole, the safety spark gap with the + poles. The working spark gap 7 may also be dispensed with and the current for it taken at 12 over the oscillation circuit 5, 11 (shown in dotted lines).

Flame collectors (Figure 37) may also be employed according to this invention. The wire network 1 is connected with the aerial collector conductor A and the burner with the earth. At the upper end of the latter, long points are provided which project into the flame. The positive electrode is connected with the negative over a condenser 5 and the induction coil 9 with the earth.

The novelty in this invention is firstly, the use of incandescent cathodes opposite positive poles which are connected with large metallic capacities as automatic collecting surfaces, (2) the connection of the incandescent cathodes with the earth whereby, in addition to the electricity conveyed to them from the battery or machine which causes the incandescing, also the negative charge of the earth potential is conveyed, and (3) the connection of the positive and negative poles of the radiating collectors over a condenser circuit alone or with the introduction of a suitable inductive resistance, whereby simultaneously an oscillatory oscillation circuit may be obtained. The collecting effect is by these methods quite considerably increased.

I declare that what I claim is: --- [Claims not included here]

British Patent # 157,262

(10 July 1922)

Improvements in Electric Motors

Hermann Plauson // Otto Traun's Forchungs-Laboratorium GmbH

This invention relates to that type of motor in which rotation is produced by means of the attraction and repulsion of surfaces carrying charges of electricity.

According to this invention a stator and rotor are formed of condensor surfaces and charges of electricity thereon imposed in the form of alternating currents of high frequency.

The invention is more particularly described with reference to the accompanying drawing in which: ---

Figure 1 shows a simple form of motor and feed.

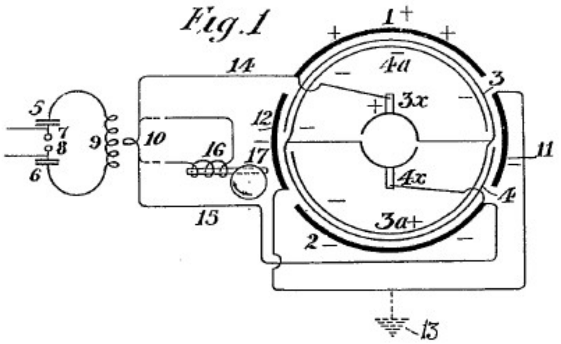
Fig.1
1 +
+
+
4a
3
14
3x
12
5
7
8
9
10
16
17
11
6
4x
4
15
3a+
2 –
13

Figure 2 is a modification of Figure 1.

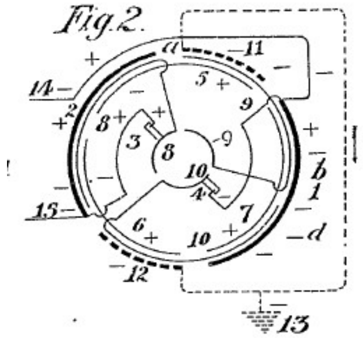
Fig.2.
a
–11
14–
2
5 +
9
8+
3
8
9
10
4
b
1
13–
6
7
10
–d
–12
13

Figure 3 shows one form of a spiral condenser surface.

Figure 4 shows a wire wound condenser surface.

Figure 5 is a diagram of one type of rotor,

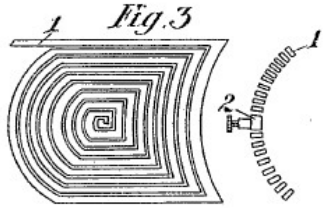
Fig.3

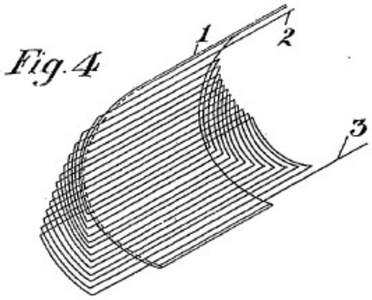
Fig.4

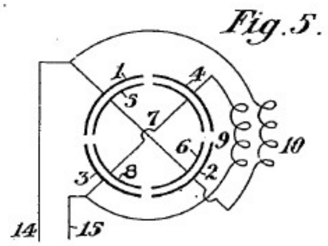
Fig.5.
1
4
5
7
9
10
6
3
8
2
14
15

The inner plates of the condenser 5 and 6 are charged from a spark gap 7, 8 connected to a source of energy of sufficiently high pressure (alternating or direct current), until the potential has risen so far that a spark springs over.

The spark gap 7, 8 forms with the condenser 5 and self-inductance 9 and condenser 6 a closed oscillatory circuit and alternating currents of high frequency will be produced in this circuit. The high frequency current produced in the primary circuit 9 excite by induction in the secondary circuit 10 currents of the same periodicity.

The improved type of motor is fed by the discharges produced by the induction in the secondary circuit.

Hitherto only Tesla's motor system (shown diagrammatically in Figure 1, 16 and 17) was known for this purpose. The above-mentioned diagram is only shown for illustrating the fundamental principle. It has however no practical interest for carrying out large machines by reason of the impossibility of the regulation and the low efficiency.

Now according to this process, all these defects are overcome by the construction of a machine which is applicable for high frequency currents and of a more or less undamped nature. The difference between the principle of construction of these motors as compared with those hitherto customary consists in that the motor is not based on the principle of magnetic induction only (as have been all motors hitherto and also Tesla's motors).

It has been fund that the machine constructed according to Figure 1 cannot only be fed directly with static electricity but if it is connected to a source of high frequency current it will operate.

The applicants call this new type of motors 'condenser motors' to differentiate them from hitherto existing types.

The simplest form of construction of such condenser motors is shown in Figure 1, and this motor may be fed with high frequency alternating currents.

At a given moment positive electricity is charged by means of the lead 14 to the stator surface 1 and to the brush 3x (Figure 1). The brush 3x is connected with the rotor condenser surface 3, so that both the stator surface 1 and also the rotor surface 3 is charged with positive electricity. The stator surface 1 and the rotor surface 3 being both charged with positive electricity and the second rotor surfaces 4 and 4a by brushes 4x with negative electricity, such motors can then be started by providing intermediate stator surfaces 11, 12, the earth connection 13 of one of which is broken by a switch (not shown) according to the direction of rotation desires, or alternatively the motor may be started by a separate source of alternating current in a manner similar to the starting of synchronous motors of known construction. After a half revolution of the rotor the brush 3x comes in contact with the second collector surface 4 so that now this surface is connected by the brush 3x with the stator surface 1 and the brush 4x with the collector surface 3. Consequently with a reverse direction of current through the second half of the oscillation period all the hereinbefore mentioned effects take place in the reverse direction which, however, produces no alteration in the direction of rotation because the dead points between two directions of oscillation are overcome by inertia.

Although this motor is easy to start it can only be employed for small experimental and measuring purposes because the stator and rotor surfaces are made of solid metal and are heated by Foucalt (eddy) currents. In spite of its simplicity and its unsuitability for use in practice it must however be regarded as a basic type for technical calculations.

The condenser motor shown in Figure 2 differs from Figure 1 by the rotor surfaces consisting of six condenser surfaces connected one behind the other in series and they are connected with three collector surfaces, so that at any one moment only two adjacent collector surfaces come under the two brushes (3 and 4). In its other actions it corresponds to Figure 1. The leads 14 and 15 may be connected either to the ends of the secondary coil 10 or directly with the source of energy. The outer thicker line indicates the stator surfaces 1 and 2 (that is to say the unmoving part of the motor), 11 and 12 shown by thick dotted lines means earthed additional poles of the stator, 8, 9, and 10 are the outer parts of the rotor condenser surfaces which in turn are connected with the collector surfaces 8, 9, and 10. 5, 6, and 7 are the inner parts of the condenser surfaces of the rotor and 3 and 4 are brushes.

Hitherto stator and rotor surfaces of compact metal have been spoken of. These however become highly heated with eddy currents and hardly yield 10-15% of useful effect. In examining into such small useful effects it was found that certain forms of metal sections in the stator and condenser surfaces highly increase these. It was then further found that if slots or notches be cut in the metal surfaces of the stator and rotor in the form of a spiral, not only was a higher useful effect possible, but also an easier starting and even a regulation could be obtained.

Experiments have shown that by such a form of construction it is possible to build a very useful motor for high frequency alternating currents more particularly those of an undamped nature.

If for example the system of construction of a stator shown in Figure 1, but four polar, be taken and the system of rotor construction shown diagrammatically in Figure 5, but with the form of construction of the condensers of the stator as well as of the rotor according to Figure 3, a condenser motor is obtained which works well in either direction for high frequency alternating current. It was also observed that the motors in such forms of construction were found to be more sensitive to resonance effects. Such a motor then works the best if stator and rotor surfaces have equal capacity and self-inductance so that the windings both in the stator and also in the rotor are in resonance.

A motor constructed according to the foregoing kind is already fully technically applicable. But even these motors have a series of faults, more particularly in their building construction. For example, the attachment of the spiral condenser surfaces both of the stator and of the rotor sown in Figure 3 are in practice difficult to carry out. Therefore in practice the condenser and stator surfaces are simply wound of wire or bands in the form shown in Figure 4. Such stator and rotor surfaces may, without further difficulty, be regarded as electromagnetic poles, although they are not made of iron as is the case in electromagnets. Such machines may be spoken of directly as motors for high frequency alternating currents in which the separate pole surfaces consists of wound induction condenser surfaces of which one is sound on the stator and the other on the rotor.

If the coil as shown in Figure 4 be made of well insulated wires the coil can be embedded in insulating material either for the stator or motor surfaces as has already been done in the case of ordinary single and multiphase motors. At the same time the possibility is afforded by increasing the number of turns to produce a greater or smaller alteration of the self induction co-efficients.

In Figure 5 is shown a modified construction of a rotor for a four pole motor consisting of four condenser surfaces 1, 2, 3, 4, of which 1 and 2 are connected though an iductance 9 coupled with the coil 10. Four inner surfaces 5, 6, 7, 8 are provided of which 5 and 6 are directly connected also 7 and 8 similarly connected.

The pairs of like poles are connected by wires 14 and 15 to the source of energy. By a suitable selection of the values of the reactance and capacity in these circuits resonance circuits may be formed.

Having now particularly described and ascertain the nature of our said invention and in what manner the same is to be performed, we declare that what we claim is: --- [Claims not included here]

British Patent # 157,263

Process & Apparatus for Converting Static Atmospheric Electrical Energy into Dynamic Electrical Energy of any Suitable High Periodicity

Hermann Plauson // H.O. Traun’s Forschungs-Laboratorium GmbH

Static aerial electricity in the form of direct current can be converted by using spark gaps and with the assistance of oscillatory circuits into dynamic electrical wave energy of a high number of alternations of a more or less undamped nature and in such form --- either direct or by means of a special kind of resonance or 'condenser motors' --- ready to be utilized for technical purposes as mechanical energy.

For small installations this system may be very well employed; about 100 horsepower may be stated as practical limit. In constructions of larger aggregate difficulties as regards the spark gaps however increase considerably. Further it is desirable to convert the accumulated currents of from 100 to 1000 periods which may then be used for the ordinary types of alternating current machines instead of into electromagnetic waves of a high number of alternations.

In experimenting with condenser motors the construction of which forms the object of British Patent # 157,262 it was observed that the rotor, if one pole of the stator surface be connected with the aerials collecting aerial electricity and the other pole with the earth, not only could act as a motor, but if vice versa the rotor connection with the stator be interrupted and the rotor caused to rotate by means of another motor, that when the brushes supply an alternating current the periodicity of which is dependent on the number of poles and the revolutions of the rotor. Such an apparatus may therefore be regarded as a transformer of static into dynamic electrical energy.

The invention is more particularly described with reference to the accompanying diagrams in which: ---

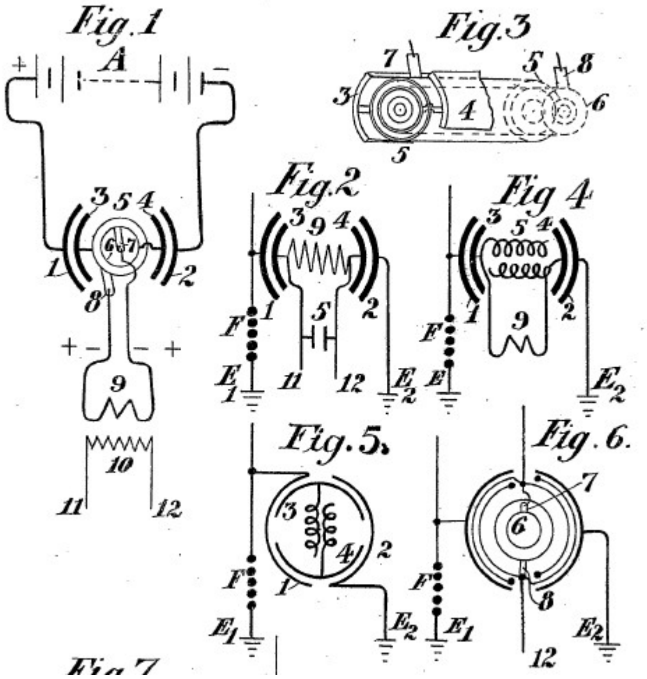

In Figure 1A is a strong accumulator battery, 1 and 2 are the outer poles of the transformer, consisting of simple metallic plates or are as shown in Figures 8-11, made of wire coils without an electromagnet being present. Between these poles an armature is revolubly mounted on a shaft, which armature also consists of two similar cylindrically curved plates 3 and 4. These are metallically connected with two collector rings 5 and 6 on which two brushes 7 and 8 freely run which again are short-circuited with one another over a primary coil 9. 10 is the secondary coil with the free ends 11 and 12. If through the accumulator battery the stator plate 1 is charged with positive electricity, it induces a charge of reverse sign on the rotor surface 3 which is connected by the brushes 7 and 8 over the primary coil 9 with the second rotor surface 4. This latter is therefore charged with positive electricity, which in turn induces negative electricity on the stator surface 2/ Up to this moment everything takes place in the same way as if two condensers were connected one behind the other in the current circuit A. If however, by means of mechanical power, this rotor be caused to rotate, the surface conditions are altered. After a quarter revolution the rotor plates are between the stator plates and therefore no condenser surface faces another. By this means however, the capacity of the entire system is reduced to a minimum and a change of current will also result in the main 9. Now if the rotor be turned further through 90 degrees by mechanical energy the rotor plate 3 comes opposite the stator plate 2 and the rotor plate 4 opposite the stator plate 1, so that then the rotor pates are in a field of reverse sign. A fresh charge of current in the reverse direction now runs through the primary coil 9. After a further half revolution the same action is repeated so that after a full revolution the initial condition is again produced. The result of such a revolution is an alternating current the periodicity of which is equal to the number of revolution. In practice of course not two poles but as many poles as possible would be employed because thereby the number of alternations would e considerably increased. The primary alternating current thus obtained induces in the secondary circuit an alternating current the potential of which is dependent on the winding of the coil. Figure 7 shows a multipolar machine.

If the stator surface 1, instead of being connected with the battery be connected with a collecting aerial network and the other stator surface 2 be directly earthed, but the rotor wich is otherwise constructed as hereinbefore, be rotated by a separate motor a much stronger alternating current results which is to be ascribed to the circumstance that a much higher potential can be charged on the pole surfaces of the stator by reason of the higher pressure of the static electricity than where accumulators are employed. By this means the transformer has of course much larger quantities of energy supplied to it.

Figure 2 shows a mode of connections. The stator surface 1 is connected with the aerial antennae which is connected through the safety spark gap F to earth at E1. The stator surface 2 is directly earthed at E2. The inner revoluble rotor surfaces 3 and 4 are interconnected by means of an induction coil which is constructed directly in the motor. The current is taken as in Figure 1 up to collector rings by means of brushes, which are not shown for the sake of clearness, and further conveyed through the conductors 11 and 12. Between these a condenser 5 may be inserted. There is thereby formed a short oscillatory oscillation circuit free from spark gaps, which circuit consists of the induction coil 9 and condenser 5 and is fed by the periodic charging current impulse. By this means the possibility is afforded of obtaining a kind of current which is characterized by longer periods and is undamped and oscillatory. Of course a simple alternating current may be obtained by cutting out the condenser.

Instead of the induction coil the condenser may also b constructed in the rotor. His can be carried out in such a way that its ends serve directly as collector rings for taking current through the brushes. In Figure 3 such a motor is sketched in perspective, 3 and 4 are the rotor surfaces, 5 and 6 are the condenser surfaces constructed to form part of the rotor consisting of two co-axial cylinders fitting one in the other in such a way that free room is left for the brushes 7 on one end of the condenser cylinder 6.

The condenser may be made in the form of a cylindrically wound spiral forming the capacity and reactance as shown in Figure 4. A further type of transformer is shown in Figure 5. The difference consists in the stator and rotor surfaces not only each assuming a quarter of the circuit but almost the half. By this means the space and the effective condenser surface is better utilized. Charge is produced only when the rotor surfaces face the full scope of the stator surfaces.

In addition a condition is obtained in which the stator surfaces are inductively connected by the rotor surfaces. The consequence of this is that an alternating current simultaneously results which is produced without sparking otherwise the connection is as before.

Figure 6 shows the alternation of the rotor surfaces; the rotor here consists of two cylindrical condenser plates arranged concentrically, each divided into two halves and connected so that half the inner cylinder is connected to half the outer. Such a machine shows the more complete transformer action.

Fig.7.

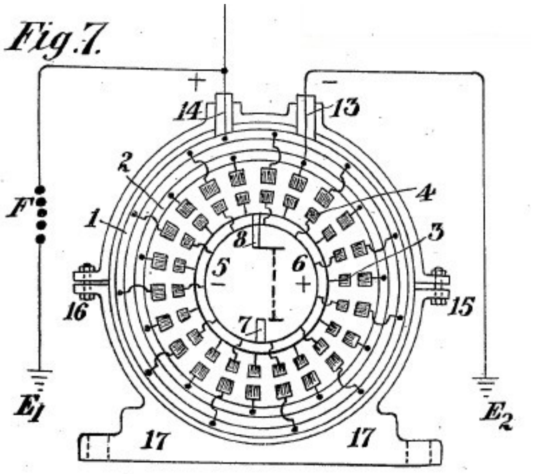

Figure 7 shows a four polar transformer. It consists of a metal casing, the lower half of which is fastened with the foundation plates 17 to the support or foundation. The upper half, the cover, is connected by bolts 15 and 16 firmly with the under part. This upper sleeve or casing is insulated from the under part. Two rings 1 and 2 are cylindrically constructed in the casing. The ring 1 is metallically connected with the collector aerial and the ring 2 with the earth. On both rings an equal number of stator surfaces are mounted side by side but well insulated from one another and thus form an electrostatic field similar to the electromagnetic in many alternating current machines. The rotor consists in similar manner of two rings 5 and 6 on which an equal number of rotor surfaces are fixed so that each stator surface faces a rotor surface. By the brushes 7 and 8 the alternating current formed is removed from the collector. The charge is conveyed by the conductor 14 to and by 13 away. If this rotor be then rotated by means of a motor the positive and negative fields precisely as in the case of magnetizing will alter and thereby an alternating current is formed in the rotor, the periods of which are dependent on the number of the poles and the revolutions per second.

At the commencement it was thought that this apparatus could only be regarded as alternating current converters, but it was soon found that much more energy was necessary to rotate the rotor than might be necessary to overcome the friction. It was then found that the considerable expenditure of energy for rotating the rotor was caused by a conductor being moved through strongly electrostatic fields since the electrostatic lines of force must be cut at right angles and that further in the conductors a stronger current arose than was otherwise to be expected. This apparatus must therefore not only be regarded as a transformer, but also as an energy producer, with the difference that the excitation here is obtained instead of by means of electromagnets, by static fields of high pressure. The entire system may, to some extent, be compared with a dynamo in which the excitation takes place by means of a fixed constant magnet. It was further ascertained that this way of using the atmospheric electricity produced a sort of suction on the collector network, and that thus suitable greater quantities of current could be obtained.

The effects which in this apparatus became evident are extremely interesting and open a prospect of being able to obtain here a great deal more. Merely that these transformer made it possible to transform suitable quantities of atmospheric electricity into alternating current of high or lower frequency (without the use of spark gaps) shows already the extreme utility of these apparatus. Should in future, the construction of larger aggregates be necessary the transformer installation may be constructed in such a way that motors which are fed by a current obtained from an installation with spark gaps produce a certain quantity of energy which may then be employed for producing current according to the last described system.

The results of the examinations made for this may be construed as follows.

(1) If solid electrodes (condenser surfaces, rotor and stator surfaces) are employed they become hot. This effect may be considerably reduced by cutting the electrodes in ribbed form Figure 8, but not entirely removed. This form allows the surface of the condenser plate to be enlarged or increased; the electrodes may be fastened in a simple manner on the under frame by perforation 1, 2, 3, 4, 5.

Fig.8

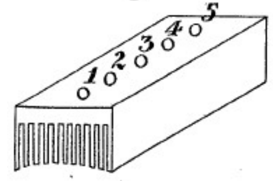

Fig. 9.

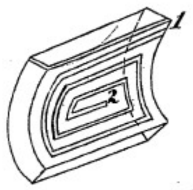

(2) If nicks or notches in spiral form as shown in Figure 9 seen from the side end in Figure 10 in section are employed, not only is the transformer effect greater but the poles yields also more current, but require greater quantities of energy for their movement than a simple commutator action would require.

Fig. 10

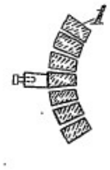

Fig. 11

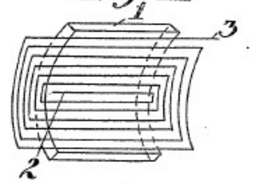

(3) The greatest effect is obtained if the rotor and stator surfaces are wound in flat spiral form of suitably thick wire, and in such a way that the inductive effect combining with the capacity is calculated in suitable proportion and this result is adapted to a suitable periodicity. In practice this is preferably done by the wire bent in spiral form being inserted in a separate vulcanite or hard rubber mass (see Figure 11) so that a smooth pole surface is formed similar to that in phase motors.

Fig.12

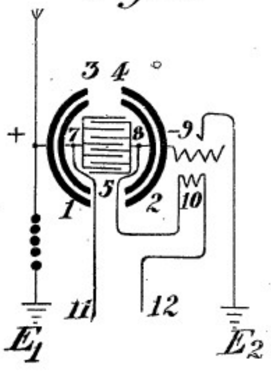

Fig. 13.

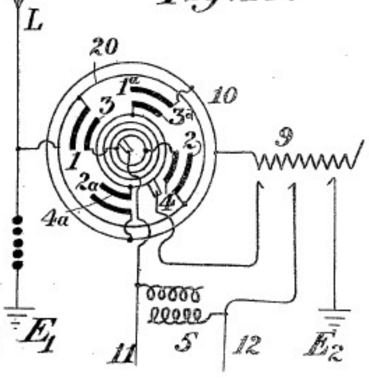

Regular undamped oscillations of a high frequency may however be produced if the converter be carried out in the manner shown in Figure 13. The aerial wire L is metallically connected with the ring 20. To this two pole surfaces 1 and 2 are connected. The inductive earth pole is also connected with a second ring 10 from which again two, poles 1a and 2a are branched off. Of course in similar manner any suitable number of poles may be branched of. In similar manner there are in the rotor two poles fastened to one another (3 and 4 and 3a and 4a) connected with separate collector rings. From these two rings the current is collected by means of two brushes. The induced alternating current is however directly metallically connected with an inductive earth stator conductor over an induction coil 9. Further a combined inductance and capacity 5 is inserted between the two wires 11 and 12 in parallel with the converter. By this means a sparkless oscillatory circuit is obtained which can act on the exciting current in the stator. This produces however, a periodic alteration of the charging quantities according to the oscillation curves of the rotor currents in consequence of which the stator charge also commences with resonance oscillations and if the stator and rotor surfaces are calculated to one another in such a way that they are adapted to form oscillations of waves of similar length the entire converter is caused to oscillate and furnishes undamped oscillations of a high number of alternations, but of periodically changed amplitude, the form of which is dependent on the amplitude of the main alternating current and is caused by the number of the poles and revolutions per second. Thus an alternating current of, for example, 100 periods is formed, the separate periods of which are formed by undamped oscillations of a higher number of alternations. In Figures 14-16 four other diagrams of converters are illustrated, the object of which is not to produce usual alternating current, but oscillations of high frequency.

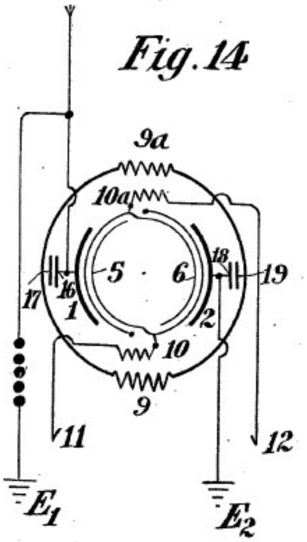
Fig. 14

The main difference of these systems from those previously described is that from the connection of the collecting aerials is made between the stator pole 1 (Figure 14) and one pole 16 of the condenser 17 and the earth connection between a second stator pole 2 and the pole 18 of the condenser 19. The other poles of these condensers 17 and 19 are short circuited through a ring over two inductive primary coils 9 and 9a with one another. The secondary coils form the rotor conductors 10 and 10a. The rotor itself is constructed in the manner shown in Figure 6 of two short-circuited plate condensers which may be wound as shown in Figure 11. In similar manner of course the stator surfaces may also be formed. The collector rings of the rotor with the two brushes for collecting current are here not shown in order to simplify the drawings. By the connection of the two condensers in the exiting circuit of the converter and also by the action of the alternating current produced in the rotor on the stator circuit, with a correct calculation of the capacity and the self-induction co-efficients a maximum action may be obtained. The kind of current produced will be similar to that described for Figure 12.

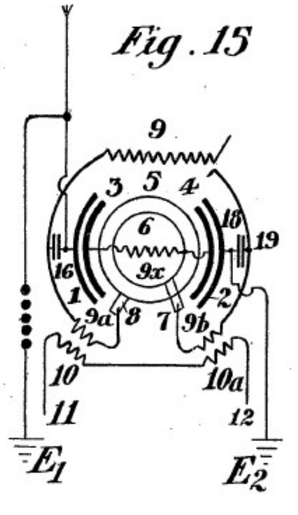
Fig. 15

The novelty of the converter illustrated in Figure 15 consists mainly in that the current resulting in the rotor is not directly employed, but only serves a exciter of the primary coils 9a and 9b. The working current is produced in the secondary coils 10 and 10a and further conducted through the conductors 11 and 12. The stator current may be brought by the regulatable inductive resistance 9 to the same resonance as the rotor current.

Fig. 16

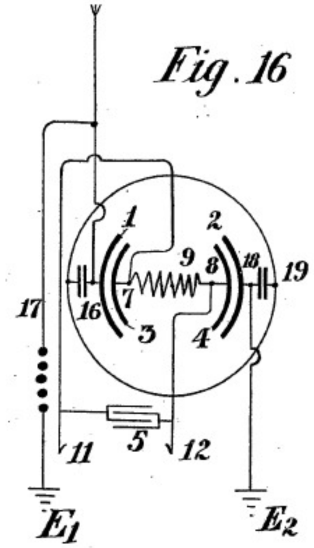

In Figure 16 a very similar system is shown to Figure 14. The condenser 5 is however connected in parallel with the converter; and by the inductive resistance constructed in the rotor a short circuited oscillatory circuit is formed which gives extraordinarily good results and is simple in construction.

The inductive resistance 9 may also instead of being constructed in the rotor be constructed as primary coil employed outside the rotor and short circuit the oscillatory circuit over the stator surfaces (see Figure 17).

Fig. 17.

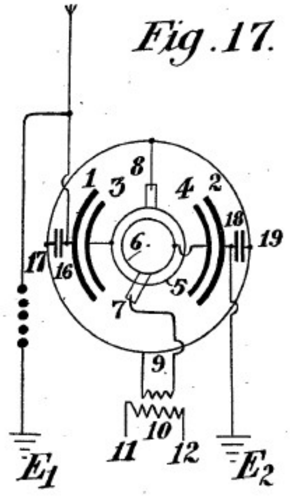

The last six types serve only for producing oscillations of a high number of alternations. If it be desired to obtain ordinary alternating current there complicated constructional arrangements are not required as the types illustrated in Figures 1 to 11 suffice. It is self-evident that these arrangements may be altered in various ways by means of different condenser surfaces in practice.

Having now particularly described and ascertained the nature of our said invention and in what manner the same is to be performed, we declare that what we claim is: --- [Claims not included here]

rexresearch.com

Hermann PLAUSON

Conversion of Atmospheric Electricity

Biography (Wikipedia)

Meridian International Research: Atmospheric Electricity Research

***Science & Invention* (Feb. 1922): "Power from the Air" (I)**

***Science & Invention* (March 1922): "Power from the Air" (II)**

Plauson's Patents (List)

H. Plauson: USP # 1,540,998 -- Conversion of Atmospheric Electricity

H. Plauson: British Patent # 157,262 -- Improvements in Electric Motors

H. Plauson: British Patent # 157,263 -- Process & Apparatus for Converting Static Atmospheric Electrical Energy into Dynamic Electrical Energy...

***Science & Invention* (June 1928) - "Harnessing Nature's Electricity"**

H. Plauson : British Patent # 299735 -- Process for Producing Rapidly Moving Electrons [PDF]

H. Plauson : *Gewinnung und Verwertung der Atmosphatischen Elecktrizitat* (1922) [PDF]

http://en.wikipedia.org/wiki/Hermann_Plauson

Biography

Hermann Plauson was an Estonian engineer and inventor. Plauson investigated the production of energy and power via atmospheric electricity.

Plauson was the director of the Fischer-Tropsch "Otto Traun Research Laboratories" in Hamburg, Germany during the Weimar Republic of the 1920s. He built on Nikola Tesla's idea for connecting machinery to the "wheelwork of nature". Plauson's US Patent # 1,540,998 describes methods to convert alternating radiant static electricity into rectified continuous current pulses. He developed the Plauson's converter, an electrostatic generator. In 1920, Plauson published a book titled "*Production and Utilization of the Atmospheric Electricity*" (*Gr., Gewinnung und Verwertung der Atmospharischen Elektrizitat*). A copy of this book is in the British Library.

It is believed that he was related to Gertrud Plauson (the exact relationship is unknown; she may be his wife).

"Power from the Air". *Science and Invention* , Feb. 1922, no. 10. Vol IX, Whole No. 106. New York. (nuenergy.org)
"Power from the Air". *Science and Invention* , March 1922.

***Science and Invention,* Vol. IX (106) #10 (February 1922)**

Power from the Air (I)

by

Hugo Gernsback

During the war there was developed in Germany a new art --- or science --- that bids fair to revolutionize our present means of obtaining power.

This art, which is as new now as wireless was 25 years ago, will attain proportions during the next 25 years that may appear fantastic today. The inventor of the new science, an engineer of note, Herr Hermann Plauson, has devoted years of labor to his researches and he has now actually in use small power plants, that generate electricity direct from the air, day and night, without interruption at practically no cost, once the plant is constructed.
We had occasion, in one of our former issues, to describe the system, roughly, from cabled dispatches, but complete information is

available now. The amount of electrical power that resides in our atmosphere is astounding. Herr Plauson found in his experiments that a single balloon sent aloft to a height of 300 yards gave a constant current at 400 volts of 1.8 amperes, or in 24 hours over 17-1/4 kilowatts! By using two balloons in connection with a special condenser battery, the power obtained was 81-1/2 kilowatts in 24 hours. The actual current delivered was 6.8 amperes at 500 volts.

The best balloons used by the inventor are made of thin aluminum leaf. No fabric was used. A simple internal system of ribs, stays and wires, gives the balloon rigidity as well as a certain amount of elasticity. The balloon, when made airtight, is filled with hydrogen or better, with helium. It will then stay aloft for weeks at a time. The outer surface is dotted with extremely sharp pins, made sharp electrolytically. Ordinary pins did not prove good current collectors, as they lacked extreme sharpness. The pins themselves were made from amalgamated zinc, containing a radium preparation, in order to ionize the air. It was also found that by dotting the outer surface of the balloon with zinc-amalgam more current could be collected. Even better results were obtained with polonium amalgam. Plauson states that the function of these amalgams is purely photoelectric.

One hundred of such captive balloons, separated one hundred yards from each other, will give a steady yield of 200 horsepower. This is the minimum, because in the winter this figure increases up to 400 horsepower, due to the higher electrification of the atmosphere.

We need not go into the technic of how the current is finally made useable for industrial purposes, suffice it to say that the problem has been entirely solved by Herr Plauson. By using batteries of condensers, high tension transformers, etc., the current can be transformed to any form desires. Such as for lighting lamps, running motors, charging storage batteries, etc.

Plauson also invented a sort of electrostatic rotary transformer which gives alternating current without the use of condensers and transformers. Indeed, its output is very great, as it actually ‘sucks’ the current down rapidly from the collector balloons. There is no doubt that this invention will soon come into universal use all over the world. We will see the land dotted with captive balloons, particularly in the country and wherever water power does not abound. Indeed, the time is not distant when nearly all of our power will be derived from the atmosphere. So far it seems to be the cheapest form of power known, it being much cheaper even than water power --- the cheapest form of power known today. Not only that, but as the inventor points out, no devastating thunder storms occur near such aerial power plants, because the balloons act not only as lightning arresters, but they quickly discharge the biggest thunder cloud, safely and noiselessly through their grounded spark gaps.

***Science & Invention* (March 1922), page 1006, 1007**

Power from the Air (II)

by

Hugo Gernsback

[For many years electrical engineers have endeavored to devise some means whereby it would become possible to utilize the free electrical energy ever present in the atmosphere, but they were not successful, as every now and then an extra heavy surge of static current would rush down the elevated conductor and endanger the lives of the experimenters, or else destroy the apparatus connected with it. A German engineer has, however, devised the somewhat elaborate scheme here shown in brief, and he has succeeded, at least so his report states, in safely extracting several kilowatts of electrical power from the atmosphere with metallic surfaced balloons, elevated to a height of only 1000 feet.]

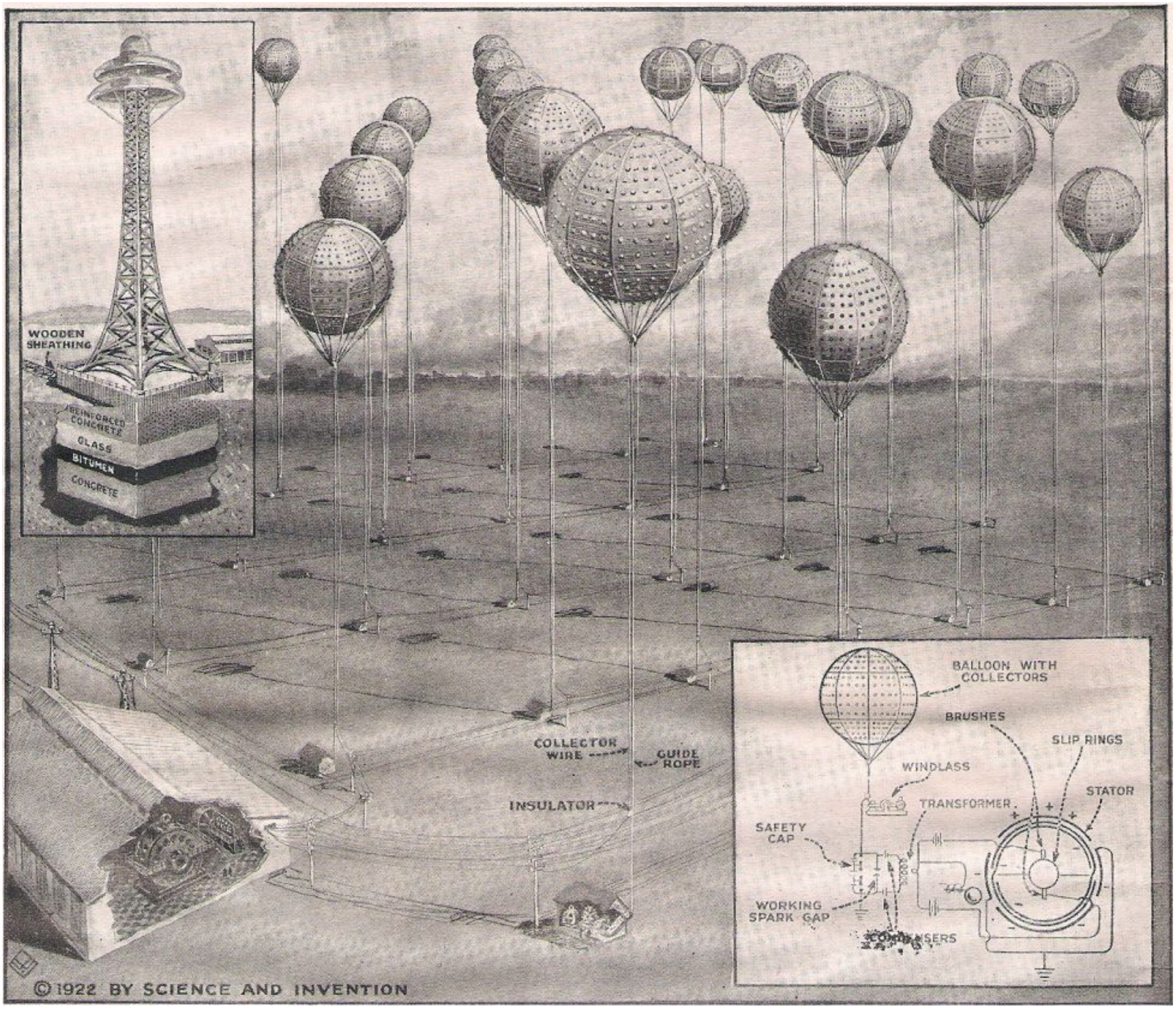

We have previously treated of the extraction of electrical energy from the atmosphere. The difference of electric potential in different parts of the atmosphere, and the difference between the upper air and earth make it a tempting proposition to obtain power from atmospheric electricity. The power would take the form of high potential difference with a discharge almost of a static nature. It has long appeared rather doubtful to conservative engineers, if such a source of power should really be available. Yet when we see the lightning flash, it certainly suggests very high power, even though the total of its energy may be small, on account of the small duration of the discharge It is not to the thunder storm that we look for getting power from the atmosphere, as the subject is now being seriously investigated. A German scientist, Hermann Plauson, has published a very elaborate work on this subject, and has investigated the use of kites, balloons and towers, for the utilization of the high potentials existing in the air at different altitudes, and has studied out the construction of motors to be operated by the peculiar type of discharge which will be obtained, if the projects are successfully carried out.

We will first speak of the methods used for collecting electricity from the upper air. The author cites several German patents. One of them shows the use of a kite balloon. The balloon is shown floating in the air, kite fashion, and from it hangs a great net or aerial for the collection of electricity. The conductor from the aerial leads to the ground station; quite an elaborate description is given of the net-work which the patentee proposes to have covered with needle points. A windlass takes in or pays out cable for the balloon, and the patentee claims that by sending the apparatus to a height of about one mile he will have 225,000 volts to draw upon. He then speaks o a battery of 20,000 cells in series, which will use up to 40,000 to 50,000 volts in the charging. This certainly provides for a reasonable large fall of potential.

But our author discards this idea and first suggests something more permanent. He proposes the erection of towers to be in the neighborhood of 1,000 feet high, or about the height of the Eiffel Tower. At the summit he has his collecting aerial. The appliance consists of a number of copper tubes; within each one he proposes to burn gas lamps, whose products of combustion will reach the aerial, a collecting net-work covering the tops of the tubes. One of his apprehensions is that if rain should wet his connections trouble might ensue, so he proposes a protection at the top in the shape of a great bell-like shield, resembling in his terms “a Siamese pagoda”. He also compares the form of the protection to that of a great petticoat insulator. Another of his difficulties is that he must have his tower insulated from the earth. He, therefore describes a complicated foundation for his structure. He proposes first to pour in at the bottom of the excavation a foundation of simple concrete. On this he places a layer of asphalt, and then a layer of cast glass, three to ten feet thick, and then comes a reinforced concrete foundation, to which the metallic foot of the tower is to be anchored. This foundation must rise at least seven feet above the ground level, and is to be boarded in on all sides to protect it from moisture. The author’s idea s to erect a number of these towers connected by a horizontal cable, to which the aerials for collection of potentials are secured.

The author strongly advocates balloons as collectors of the electric power of the air. These he depicts covered with spots. These spots indicate areas to be variously coated and prepared to collect potential from the atmosphere.

In the first place he describes the balloon as made of thin metallic leaf supported by internal ribs. Steel wires silver-plated, copper-plated, or aluminum-coated, run from the balloon to the pendant or junction ring. To this ring the tether cable is attached and runs to an insulated windlass on the surface of the earth. The balloon is to rise to an altitude varying from 300 feet to three miles.

The coating of the spots is to be of the thinnest amalgam, of mercury and gold, or zinc, or even polonium, perhaps only 1/2500 inch thick. All over the upper face of the balloon are numberless metal points. To prepare the needle-like wires, they are collected into bundles and are treated electrolytically in a bath, so as to be dissolved in part. This gives a sharp point and roughened surface, all adapted for collecting the electric energy. The points may be of copper, steel, or some hard metallic alloy. After this corrosion. As it may be termed, the wires are plated with gold or other of the so-called noble metals. It is advised that polonium or radium salts be added to the plating bath.

Dr Plauson devotes many pages of his book to describing his motor. This is a rotary motor including a stator and rotor and its peculiarity is that it contains no coils, develops no electromagnetic field properly speaking, but works by static excitation. One typical arrangement is shown in our illustration. The stator plates and rotor plates are concentric with each other, representing segments of cylinders. The alternation of negative and positive charged plates produces the rotation. In the connections there is included a safety spark gap to take care of dangerous potentials. Inductances and capacities are also used and indicated. It was found that the plates heated, owing to the Foucalt currents, and to overcome this, several methods of subdividing the stator and rotor plates, are described by the author.

The whole subject is quite captivating, and it really seems as if the utilization of the electricity of the air may be almost in sight. It would seem possible to carry out experiments in this direction by means of the Eiffel Tower, but of course, the trouble here is that the tower is grounded, and perfect insulation of the collecting surface is absolutely essential.

And now our author gives us some practical details. He says that on the Finland plains he carried out experiments with a balloon made of aluminum leaf with collecting needles of amalgamated zinc with a radium preparation as an ionizer. The surface of the balloon was sprinkled over with zinc amalgam. It was sent up to a height of 300 meters, early 1,000 feet, and was held by a copper-plated steel wire. A constant current of 1.8 amperes at an average of 400 volts potential difference was obtained. This gave nearly three-quarters of a kilowatt, or close to one horsepower. The collector of the balloon insulated from the earth showed a tension of 42,000 volts. By sending up a second balloon with an antenna to the same height at a distance of 100 meters from the first balloon, a current of over 3 amperes was obtained. Then by putting into the circuit a large condenser, whose capacity was equal to the surface capacity of both balloons, and of the antenna connections, the current rose to 6.8 amperes with about 500 volts mean tension. By the use of these two balloons, he eventually ran up the power to 3.4 kilowatts.

Science & Invention (June 1928)

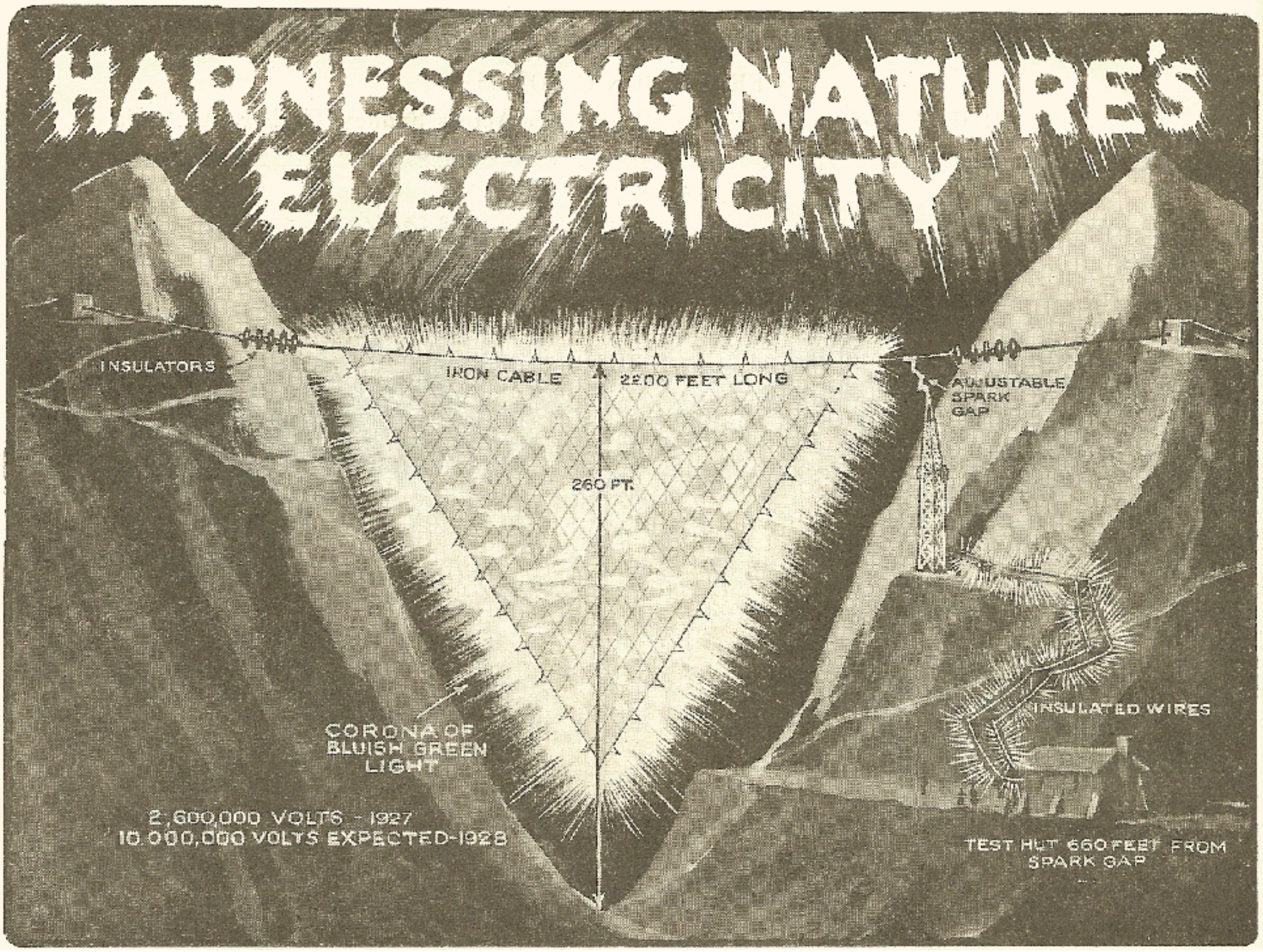
HARNESSING NATURE'S ELECTRICITY
INSULATORS
IRON CABLE
2200 FEET LONG
ADJUSTABLE SPARK GAP
260 FT.
CORONA OF BLUISH GREEN LIGHT
INSULATED WIRES
2,600,000 VOLTS - 1927
10.000,000 VOLTS EXPECTED-1928
TEST HUT 660 FEET FROM SPARK GAP

Remarkable European Experiments with Atmospheric Electrical Discharges with Potentials as High as 3,000,000 Volts

By HENRY TOWNSEND

elevation of 350 ft., and these students of natural electrical phenomena have found a very desirable location in the Alps, where they can suspend between one mountain and an adjacent one, a strong iron cable having a length of about 2,000 ft. This cable is about 250 feet above the intervening valley, and from it these daring engineers have suspended a coarsely woven wire net, which serves as an electrical capacity to gather the electricity from the atmosphere. As shown in the pictures, the wire net is supplied with numerous sharp points to aid in collecting the current from the air.

As the accompanying photographs of the actual apparatus and wire cable used last year clearly show, an adjustable spark gap of considerable length is provided. By adjusting this spark gap to various lengths, it is possible to judge the voltage of the discharge which leaps the gap at any moment. Mr. F. W. Peek, Jr., the well-known American worker in the realm of high voltage measurements, together with other engineers, have provided tabulated data and curves for various lengths of both needle and sphere type spark gaps. As one of the accompanying diagrams shows, it is a simple matter to calculate the voltage when a certain length of gap is used. The engineer first checks the length of the gap on the chart; he then follows a line horizontally from the gap length, to where it intersects with the angular line on the chart; and from the point of intersection he looks in a visual line downward to a place where the voltage is given. For needle spark gap measurements, the characteristic curve on the chart is practically a straight line, while for sphere gaps the characteristic curve on the voltage versus gap length, is a curved line. Those interested in high voltage measurements by means of the spark gap method can find the voltage-gap tables and charts in the *Standardization Rules* of the American Institute of Electrical Engineers. According to Mr. Peek's researches, the voltage per foot of atmospheric electrical discharges is about 100,000, while in laboratory measurements with A.C., transformer high potential discharges, the average voltage per foot of spark was found to be about 150,000 volts. The voltage of a lightning flash may

(*Continued on page* 156)

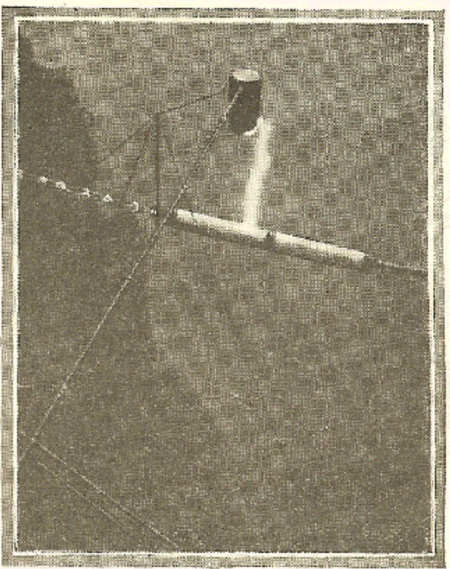

Actual photo above shows 13 ft., heavy spark obtained from the collecting net in the Alps by the German scientists. The voltage is about 2,000,000. The spark occurred once per second for 30 minutes.

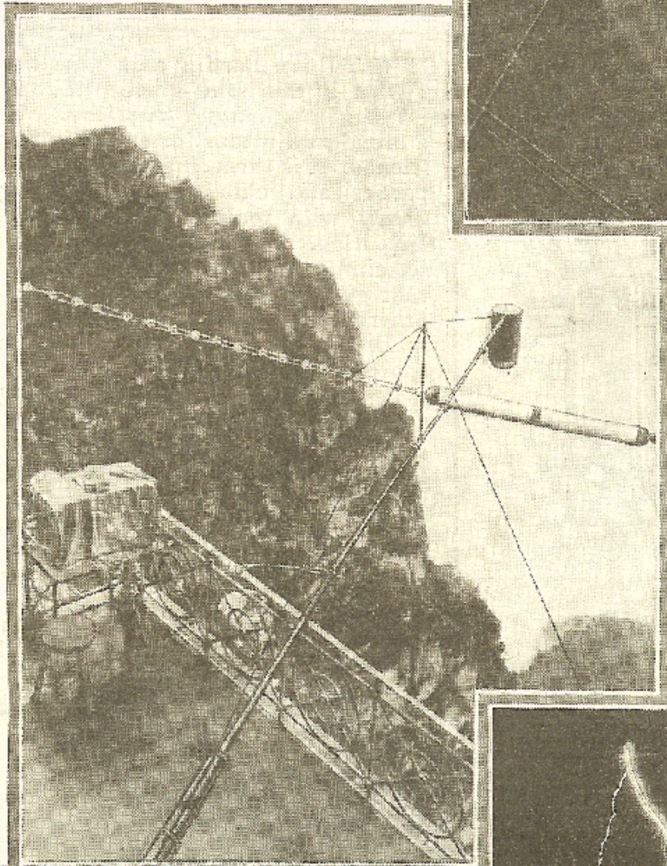

Photo, left, shows the adjustable spark gap used in the Alps. Notice the heavy electrode on the end of the adjustable arm to which the spark jumps.

Actual photograph of the experimental "kite" used by the German experimenters in the Alps Mountains, for the purpose of accumulating high potential electrical discharges from the atmosphere. Note the size of the insulators.

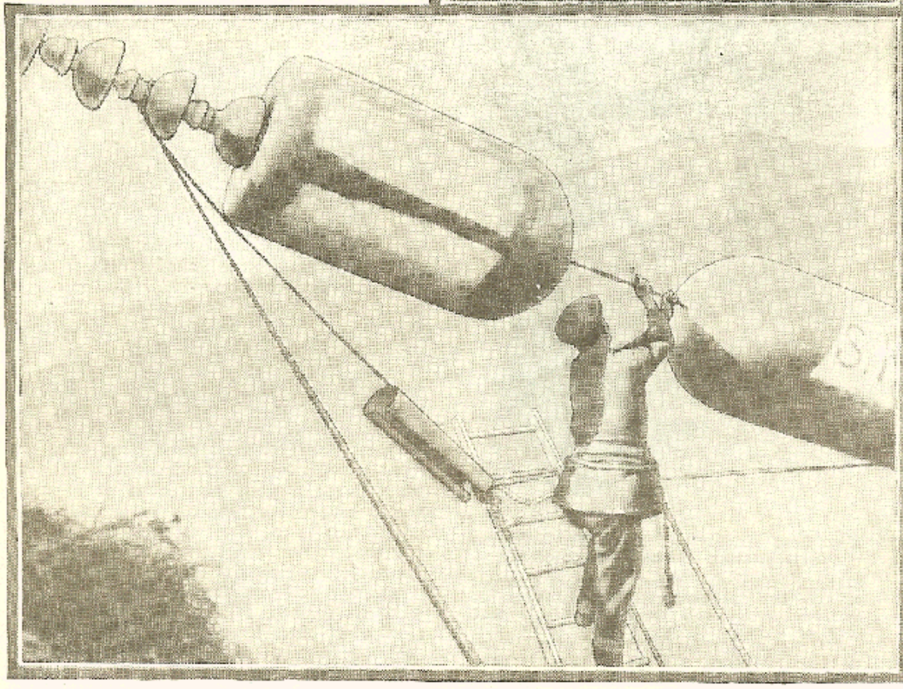

Below we see 3,000,000-volt artificial lightning stroke produced in G. E. Laboratory at Pittsfield, Mass. Note man.

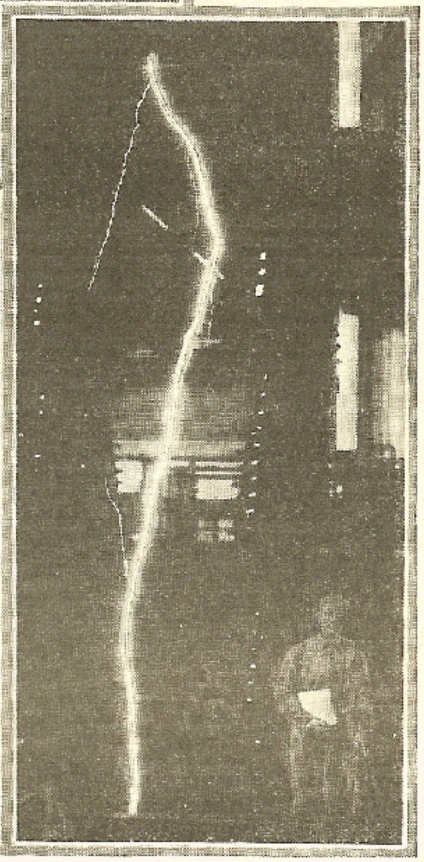

if-

you pledge

yourself to secrecy we will teach you the secrets of

Real Professional

Magic

Dr. Harlan Tarbell, internationally known magician and teacher of some of the world's greatest magicians.

Learn to be a Magician

For the first time in the history of Magic the age-old, sacredly guarded secrets of the Mystic Art are being revealed. Now at last you can learn to be a Real Professional Magician. Yes, and you can learn this wonderful art *easily and quickly* AT HOME. You are taught the mysterious *Principles* by which the Professional Magician works. You are taught everything from wonderful impromptu effects to massive stage illusions. But before you can study, you must sign the Magician's Solemn Pledge of Secrecy. This means that you are getting the *guarded secrets* of the Magic Profession. Think of that!

Earn $250 to $1000 a Month

There's a big demand for Magic entertainment. Magicians get big money. Dr. Tarbell gets as high as $250 for a half hour's work. Unlimited opportunities for you!

Low Prices Easy Terms

Write Today Send today for free Magic book, low prices and easy terms. Don't delay. Mail the coupon NOW!

Tarbell System, Inc., Studio A-142
1920 Sunnyside Avenue, Chicago, Ill.

Gentlemen: Without any obligation send me your free literature and information all about your wonderful "Tarbell Course in Magic." Also tell me about your Low Price Offer and Easy Payment Plan.

Name..

Address..

....*Age*........

HARNESSING NATURE'S ELECTRICITY

By HENRY TOWNSEND

(*Continued from page* 111)

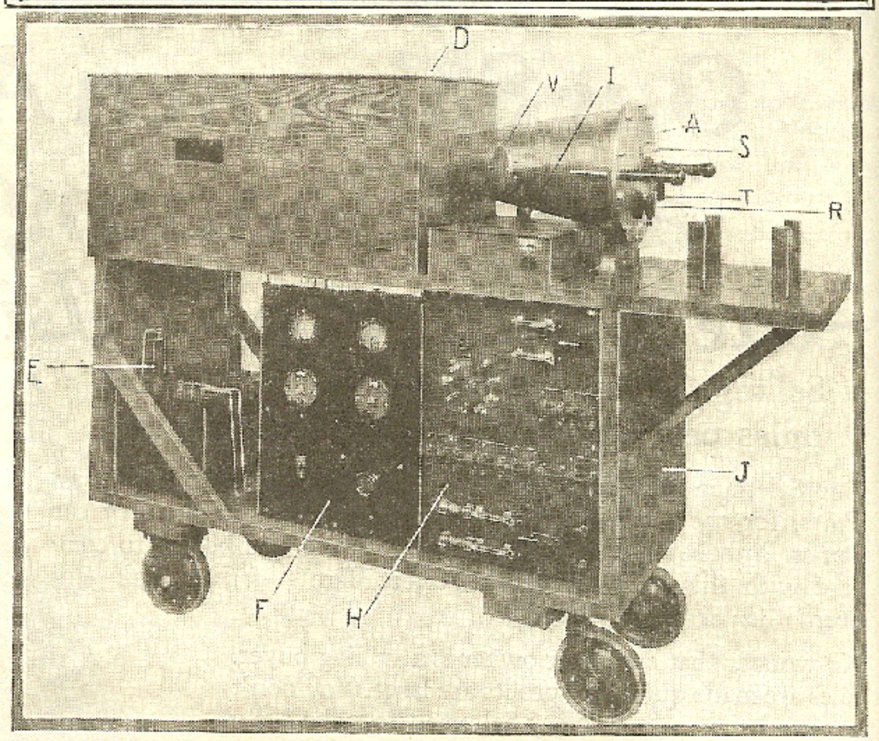

Remarkable cathode ray oscillograph used by Mr. F. W. Peck, Jr., in causing lightning to write its autograph with a pencil of electrons on a photographic film.

easily be 100,000,000 volts, as Mr. Peek has pointed out in one of his scientific papers.

WHY THE EXPERIMENTS ARE BEING MADE

ONE of the main reasons why these dangerous experiments are being carried on by the three young German scientists, whose names we have already learned, is because science believes that with a sufficiently high voltage, it will be possible for man to disintegrate the atom, and in this way make available a tremendous source of power as yet untapped. These experts have calculated that they will be able to obtain electrical energy in sufficient quantity from one of these powerful atmospheric discharges, to equal the Alpha rays obtained from 220 pounds of radium. As we have mentioned before, these experiments are of course fraught with great danger, and for that reason the experimenters seek refuge in a special lightning-proof hut, which is located about 600 ft., from the spark gap. When electrical storms are in the vicinity, it is especially important that the scientists keep within their protected fortress, for otherwise they would very probably be killed by a stray electrical discharge.

One of the peculiar things about this whole line of experiment is that the average layman does not realize perhaps, that there is a high electrical stress in the atmosphere on clear days, as well as when thunder storms are overhead. This fact has been known for a hundred years and more, and many years ago measurements of the various electrical potentials at increasing altitudes, were observed and measured by scientific investigators. There are a number of different ways in which these high electrical potentials found in the atmosphere can be measured; one of these methods involves the use of a calibrated spark gap. In this case the gap is set to a predetermined length, and when a discharge jumps this gap, the engineers know of course from previous experience and measurements, just what voltage is present. Another method of measuring extra high potentials, such as here encountered, requires the use of a static voltmeter, which involves the use of a stationary and of a movable or rotary set of metal plates, forming a condenser, to which an indicating needle is attached. For voltages above 2,000, static voltmeters have been used in a great many American central stations, and they have many desirable and useful characteristics. Of course as the voltage to be measured increases, the space between the quadrant shaped stationary and movable plates is increased and vice versa. The electrostatic field from voltages below 2,000 is not sufficient to warrant the use of a static voltmeter. Another method of measuring high potentials involves the use of the so-called vacuum tube voltmeter.

The general characteristics of the atmospheric electrical discharges, including lightning, have been measured and recorded by one of the newest scientific instruments, known as the cathode ray oscillograph. By means of this quite remarkable, high voltage scientific apparatus. Mr. Peck, one of the well-known General Electric Company staff of research engineers, has made some very interesting and remarkable discoveries concerning the nature of natural electrical discharges, particularly of lightning discharges. Many people will probably wonder why Mr. Peek and some of his colleagues in the engineering profession, including the three daring German students, Messrs. Brasch, Lange and Urban, play with such dangerous electrical discharges, and why they are at all interested in them. We explained previously why the German savants are intent on finding out all they can about these tremendous voltages obtained from the atmosphere, while Mr. Peck, we may say, also has a very practical reason for carrying on experiments with these death-dealing bolts of Thor. Mr. Peek has been for many years intent on finding out what causes the huge insulators on long distance, high potential transmission lines to break down when electrical storms break loose over these regions.

Plauson's Electrical Patents

USP # 1,540,998
Conversion of Atmospheric Electric Energy
6-09-1925

GB157262
Improvements in Electric Motors
1922-07-10

GB157263
Process and Apparatus for Converting Static Atmospheric Electrical Energy into Dynamic Electrical Energy of any Suitable High Periodicity
7-10-1922

British Patent # 299,735
Apparatus for Producing Rapidly Moving Electrons
7-15-1930

FI21227
Elektrisk uppvärmningsanordning
4-25-1946

Varmelegeme med elektriske varmemodstande
DK67691C
9-27-1948

FR877362
Dispositif de chauffage électrique
12-04-1942

DE734794
Elektrisches Heizsystem
4-24-1943

CH222509
Elektrischer Heizkörper zur Erwärmung von Flüssigkeiten
7-31-1942

DE738107
Elektrolyt fuer unmittelbare elektrische Warmwasser-Radiatorenheizung mit Elektroden
8-03-1943

DE433476
Verfahren zur Herstellung von Elektroden und Schleifkontakten fuer Dynamomaschinen
8-31-1926

CH94021
Elektrode und Verfahren zu deren Herstellung
4-01-1922

CA226423
Electrode for Electrolytic Apparatuses
11-21-1922

---

http://www.meridian-int-res.com/Energy/Atmospheric.htm

Atmospheric Electricity Research
[Excerpts]

In the nineteenth and early twentieth centuries, a large number of researchers investigated ways to extract electrical power from the Earth's ambient electric field.

The leader in this field was Dr Hermann Plauson who in the 1920s succeeded in generating significant quantities of electrical power comparable with modern solar photovoltaic systems of a similar scale...

The leader in this field before the Second World War appears to have been Dr Hermann Plauson. Dr Plauson was an Estonian citizen who lived in Hamburg and Switzerland. He carried out experiments in Finland with aerostats manufactured from magnesium-aluminium alloy, covered with electrolytically deposited needles. The needles were further doped with a radium compound to increase local ionisation of the air. (This was the era in which the hands of watches were hand painted with radium to make them luminous in the dark). Zinc amalgam patches were also painted onto the aerostats. Plauson obtained a power output of between 0.72kW and 3.4kW from one and two aerostats 300m above ground level. Dr Plauson filed patents in the USA, Great Britain and Germany in the 1920s. His book "Gewinnung und Verwertung der Atmosphärischen Elektrizität" is the most detailed known account of the technology.

Other atmospheric electricity researchers contemporary to Dr Plauson included Walter Pennock and MW Dewey in the USA, Andor Palencsar in Hungary and Dr Heinrich Rudolph in Germany. Hippolyte Charles Vion in Paris predated them all, putting forward proposals in the 1850s and 1860s.

Heinrich Rudolph made an interesting contribution to the design of the aerostat collectors. In 1898 he designed an elliptical aerostat made up of faceted surfaces to minimise the effect of wind. The design bears a strong resemblance to Northrop's 2003 UCARS unamnned helicopter UAV project. The design uses the Coanda Effect to help keep the aerostat on station and minimise wind effects.

In recent times, the only person who seems to have been active in this field is Dr Oleg Jefimenko. Dr Jefimenko carried out experiments on driving electrostatic motors from the Earth's electric field in the 1970s and has recently called for research into the neglected field of electrostatic motors to be renewed.

MIR's Research Programme

Since 1997 we have been carrying out theoretical research into conversion of atmospheric electricity into useable electrical power.

From a low level (5m high) simple zinc antenna we are able to obtain sufficient charge to light a number of white power LEDs. Further experimental investigations with metallic aerostat collectors and cavity resonant slow wave antennae concepts are ongoing...

Advantages of Atmospheric Electricity

Simple and robust technology
Low Cost technology - much cheaper than photovoltaics or wind turbines
Available day and night in all weather conditions - in fact, more power is produced at night than during the day
Available at any point on the Earth's surface

1. *Gewinnung und Verwertung der Atmosphärischen Elektrizität*, Dr Hermann Plauson, Hamburg, (1920)
2. Conversion of Atmospheric Electric Energy, USP 1,540,998, Dr Hermann Plauson, (1925)
3. Assembly for the Induction of Lightning into a Superconducting Magnetic Energy Storage System, USP 5,367,245 Goven Mims, (1994)
4. Electrostatic Motors are Powered by Electric Field of the Earth; CL Stong, *Scientific American*, (October 1974)
5. Operation of Electric Motors from the Atmospheric Electric Field; Dr Oleg Jefimenko, *American Journal of Physics*, vol. 39, July 1971.
6. *Electrostatic Motors: Their Principles, Types and Theory of Operation*; Dr Oleg Jefimenko, Electret Scientific, (1972).
7. Parametric Electric Machine, USP 4,622,510, Ferdinand Cap, (1986).

US Patent # 1,540,998

Conversion of Atmospheric Electric Energy

(9 June 1925)

Hermann PLAUSON

Be it known that I, Hermann Plauson, Estonian subject, residing in Hamburg, Germany, have invented certain new and useful improvements in the Conversion of Atmospheric Electric Energy, of which the following is a specification.

Methods of obtaining atmospheric electricity by means of metallic nettings set with spikes which are held by means of ordinary or anchored kite balloons made of fabric and filled with hydrogen, are in theory already known. Atmospheric electricity obtained in this way has been suggested to be used in the form of direct current for the charging of accumulators. This knowledge however is at present only theoretical as the conversion in practice has hitherto been a failure. No means are known of protecting the apparatus from destruction by lightning. The balloons used for collecting the charge must also me be made of very large size in order to be able to support the weight of the metallic netting and the heavy cable connections.

Instead of using heavy metallic netting as collectors attached to single air ballons of non-conducting materials which are liable to be torn and are permeable to the gas, it is proposed to use metallic balloon collectors which have the following important advantages ---

(a) The metallic cases are impenetrable to helium and hydrogen; they also represent large metallic weather-proof collecting surfaces.

(b) Radio active means the like may be easily applied internally or externally; whereby the ionization is considerable increased and therewith also the quantity of atmospheric electricity capable of being collected.

(c) Such balloon collectors of light metal do not require to be of large size as they have to carry only their own moderate weight, and that of the conducting cable or wire.

(d) The entire system therefore offers little surface for the action of storm and wind and is resistant and stable.

(e) Each balloon can be easily raised and lowered by means of a winch so that all repairs, recharging and the like can be carried out without danger during the operation.

It is further proposed to use a collecting aerial network of several separate collectors spread out in the air above the earth, which collectors are interconnected by electrical conductors.

According to this invention charges of atmospheric electricity are not directly converted into mechanical energy, and this forms the main difference from previous inventions, but the static electricity which runs to earth through aerial conductors in the form of direct current of very high voltage and low current strength is converted into electro-dynamic energy in the form of high frequency vibrations. Many advantages are thereby obtained and all disadvantages avoided.

The very high voltage of static electricity of a low current strength can be converted by this invention to voltages more suitable for technical purposes and of greater strength. By the use of closed oscillatory circuits it is possible to obtain electromagnetic waves of various amplitude and thereby to increase the degree of resonance of such current. Such resonance allows various values of inductance to be chosen whereby again the governing of the starting and stopping of machines driven thereby by simply tuning the resonance between coils of the machine and the transformer circuit forming the resonance can easily be obtained. Further, such currents have the property of being directly available for various uses, even without employing them for driving motors, of which there may be particularly mentioned, lighting, production of heat and use in electro-chemistry.

Further, with such currents a series of apparatus may be fed without direct current supply through conductors and also the electro-magnetic high frequency currents may be converted by means of special motors adapted for electro-magnetic oscillations into mechanical energy, or finally converted by special machines into alternating current of low frequency or even into direct current of high potential.

The invention is more particularly described with reference to the accompanying diagrams in which: ---

Figure 1 is and explanatory figure.

Figure 2 is a diagrammatic view of the simplest form.

Figure 3 shows a method of converting atmospheric electrical energy for use with motors.

Figure 4 is a diagram showing the use of protective means.

Figure 5 is a diagram of an arrangement for converting large current strengths.

Figure 6 is a diagram of an arrangement including controlling means.

Figure 7 shows means whereby the spark gap length can be adjusted.

Figure 8 shows a unipolar connection for the motor.

Figure 9 shows a weak coupled system suitable for use with small power motors.

Figures 10, 11, and 12 show modified arrangements.

Figure 13 shows a form of inductive coupling for the motor circuit.

Figure 14 is a modified form of Figure 13 with inductive coupling.

Figure 15 is an arrangement with non-inductive motor.

Figure 16 is an arrangement with coupling by condenser.

Figure 17, 18, and 19 are diagrams of further modifications.

Figure 20 shows a simple form in which the serial network is combined with special collectors.

Figure 21 shows diagrammatically an arrangement suitable for collecting large quantities of energy.

Figure 22 is a modified arrangement having two rings of collectors.

Figure 23 shows the connection for three rings of collectors.

Figure 24 shows a collecting balloon and diagram of its connection of condenser batteries.

Figure 25 and 26 show modified collector balloon arrangements.

Figure 27 shows a second method of connecting conductor for the balloon aerials.

Figure 28 shows an auto-transformer method of connection.

Figure 29 shows the simplest form of construction with incandescent cathode.

Figure 30 shows a form with cigar shaped balloon.

Figure 31 is a modified arrangement.

Figure 32 shows a form with cathode and electrode enclosed in a vacuum chamber.

Figure 33 is a modified form of Figure 32.

Figure 34 shows an arc light collector.

Figure 35 shows such an arrangement for alternating current.

Figure 36 shows an incandescent collector with Nernst lamp.

Figure 37 shows a form with a gas flame.

Fig. 1.

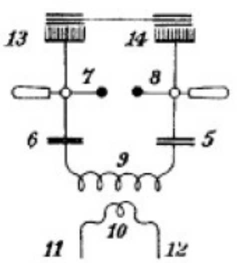

Fig. 2.

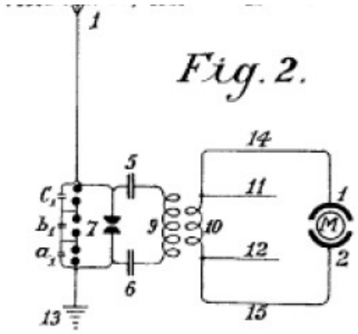

Figure 1 illustrates a simple diagram for converting static electricity into dynamic power of a high number of oscillations. For the sake of clearness in the drawings an influence machine is assumed to be employed and not an aerial antenna. 13 and 14 are combs for collecting the static electricity of the influence machine. 7 and 8 are spark discharging electrodes. 5 and 6 are condensers, 9 an inductive primary coil, 10 secondary coil, 11 and 12 ends of conductors of the secondary coil 10. When the disc of the static influence machine is rotated by mechanical means, the combs collect the electric charges one the positive and the other the negative, and charge the condensers 5 and 6 until such a high potential is formed across the spark gap 7-8, that the spark gap is jumped. As the spark gap 7-8 forms a closed circuit with condensers 5 and 6, and inductive resistance 9, as is well known, waves of high frequency electromagnetic oscillations will pass in this circuit.

The high frequency of the oscillations produced in the primary circuit induces waves of the same periodicity in the secondary circuit. Thus in the primary circuit electromagnetic oscillations are formed by the passage of the spark over the spark gap and these waves are maintained by fresh charges of static electricity.

By suitably selecting the ratio between the number of the coils in the primary and secondary circuits with regard to a correct application of the co-efficients of resonance (especially, inductance and resistance) the high voltage of the primary circuit may be suitably converted into low voltage and high current strength.

When the oscillatory discharges in the primary circuit becomes weaker or entirely cease, the condensers are charged again by the static electricity until the accumulated charge again breaks down the spark gap. All this is repeated as long as electricity is produced by the static machine employing mechanical energy.

An elementary form of the invention is shown in Figure 2 in which two spark gaps in parallel are used one of which may be termed the working gap 7 in Figure 2, whilst the second serves as a safety device for excess voltage and consists of a larger number of spark gaps than the working section, which gaps are arranged in series and are bridged by very small capacities as is illustrated in a, b, c, Figure 2 which allow of uniform sparking in the safety section.

In Figure 2 A is the aerial antenna for collecting charges of atmospheric electricity. 13 is the earth connection of the second part of the spark gap, 5 and 6 are condensers, 9 a primary coil. Now when through the aerial A the positive atmospheric electricity seeks to combine with the negative charge to earth, this is prevented by (the air gap between) the spark gaps. The resistance of the spark gap 7 is, as shown in the drawings, lower than that of the other safety section which consists of three spark gaps connected in series, and consequently a three times greater air resistance is offered by the latter.

So long, therefore, as the resistance of the spark gap 7 is not overloaded, so that the other spark gaps have an equal resistance with it the discharges take place only over spark gap 7. Should however the voltage be increased by and influences so that it might be dangerous for charging the condensers 5 and 6 or for the coil insulation 9 and 10 in consequence of break down, by a correct regulation of this spark gap the second spark gap can discharge free from inductive effects direct to earth without endangering the machine.

Without this second spark gap, arranged in parallel having a higher resistance than the working spark gap it is impossible to collect and render available large quantities of electrical energy.

The actions of this closed oscillation circuit consisting of spark gap 7, two condensers 5 and 6, primary coil 9, and also secondary coil 10 is exactly the same as the one described in Figure 1 with the arrangement of the static induction machine with the only difference that here the second spark gap is provided. The electromagnetic high frequency alternating current obtained can be tapped off from the conductors 11 and 12 for lighting and heating purposes. Special kinds of motors adapted for working with these peculiar electrical charges may be connected at 14 and 15 which can work with static electricity charges or with high frequency oscillations.

In addition to the use of spark gaps in parallel a second measure of security is also necessary for taking off the current. This precaution consists according to this invention, in the introduction of and method of connecting certain protective electromagnets or choking coils in the aerial circuit as shown by S in Figure 3.

A single electromagnet only having a core of the thinnest possible separate laminations is connected with the aerial.

In the case of high voltages in the aerial network or at places where there are frequent thunder storms, several such magnets may however be connected in series.

In the case of large units or plants several electromagnets can be employed in parallel or in series parallel.

The windings of these electromagnets may be simply connected in series with the aerials. In this case the winding preferably consists of several thin parallel wires, which make up together, the necessary section.

The winding may be made of primary and secondary windings in the form of a transformer. The primary windings will be then connected in series with the aerial network, and the secondary winding more or less short-circuited over a regulating resistance or an induction coil. In the latter case it is possible to regulate to a certain extent the effect of the choking coils. In the further description of the connecting and constructional diagrams the aerial electromagnet choke coil is indicated by a simple ring S.

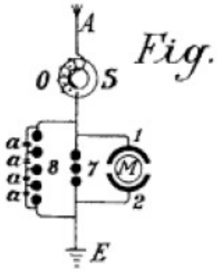
Fig. 3.

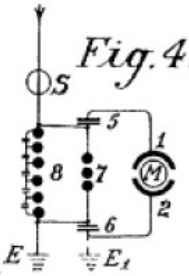
Fig. 4.

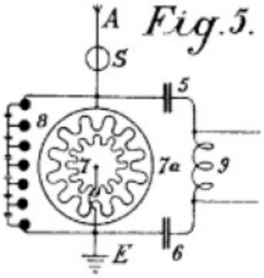
Fig. 5.

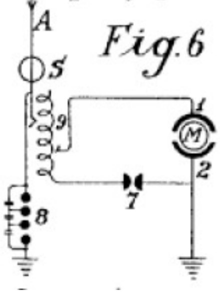
Fig. 6

Figure 3 shows the simplest way of converting atmospheric electricity into electromagnetic wave energy by the use of special motors adapted for high oscillatory currents or static charges of electrical energy. Recent improvements in motors for working with static charges and motors working by resonance, that is to say, having groups of tuned electromagnetic cooperating circuits render this possible but such do not form part of the present invention.

A motor adapted to operate with static charges will for the sake of simplicity be diagrammatically indicated by the two semicircles 1 and 2 and the rotor of the motor by a ring M (Figure 3). A is a vertical aerial or aerial network. S the safety choke or electromagnet with coil O as may be seen is connected with the aerial A. Adjacent the electromagnet S the aerial conductor is divided into three circuits, the circuit 8 giving the safety spark gap, the circuit 7 with the working spark gap, and then a circuit including the stator terminal 1, the rotor and stator terminal 2 at which a connection is made to the earth wire. The two spark gaps are also connected metallically with the earth wire. The method of working these diagrams is as follows:

The positive atmospheric electric charge collected tends to combine with the negative electricity (or earth electricity) connected with the earth wire. It travels along the aerial A through the electromagnet S without begin checked as it flows in the same direction as the direct current. Further, its progress is arrested by two sparks gaps placed in the way and the stator condenser surfaces. The stator condenser surfaces are charged until the charge is greater than the resistance of the spark gap 7, whereupon a spark springs over the spark gap 7 and an oscillatory charge is obtained as by means of the motor M, stator surfaces 1 and 2, and spark gap 7, a closed oscillation circuit is obtained for producing the electromagnetic oscillations. The motor here forms the capacity and the necessary inductance and resistance, which, as is well known, are necessary for converting static electricity into electromagnetic wave energy.

The discharge formed are converted into mechanical energy in special motors and cannot reach the aerial network by reason of the electromagnet or choke. If, however, when a spark springs over the spark gap 7 a greater quantity of atmospheric electricity tends to flow to earth, a counter voltage is induced in the electromagnet, which is greater the more rapidly and strongly the flow of current direct to the earth is. By the formation of this opposing voltage a sufficiently high resistance is offered to the flow of atmospheric electricity direct to earth to prevent a short circuit with the earth.

The circuit containing spark gap 8 having a different wave length which is not in resonance with the natural frequency of the motor, does not endanger the motor and serves as security against excess voltage, which, as practical experiments have shown, may still arise in certain cases, but can be conducted direct to earth through this spark gap.

In the diagram illustrated in Figure 4 the spark gap 7 is shunted across condensers 5 and 6 from the motor M. This construction affords mainly a better insulation of the motor against excess voltage and a uniform excitation through the spark gap 7.

In Figure 5 a diagram is illustrated for transforming large current strengths which may be employed direct without motors, for example, for lighting or heating purposes. The main difference is that here the spark gap consists of as star shaped disk 7 which can rotate on its own axis and is rotated by a motor opposite similarly fitted electrodes 7a. When separate points of stars face one another, discharges take place, thus forming an oscillation circuit over condensers 5 and 6, and inductance 9 for oscillatory discharges. It is evident that a motor may also be directly connected to the ends of the spiral 9.

The construction of the diagram shown in Figure 6 permits of the oscillation circuit of the motor being connected with an induction coil/ Here a regulating inductive resistance is introduced for counter-acting excess voltages in the motor. By cutting the separate coils 9 (coupled inductively to the aerial) in or out the inductive action on the motor may be more or less increased or variable aerial action may be exerted on the oscillation circuit.

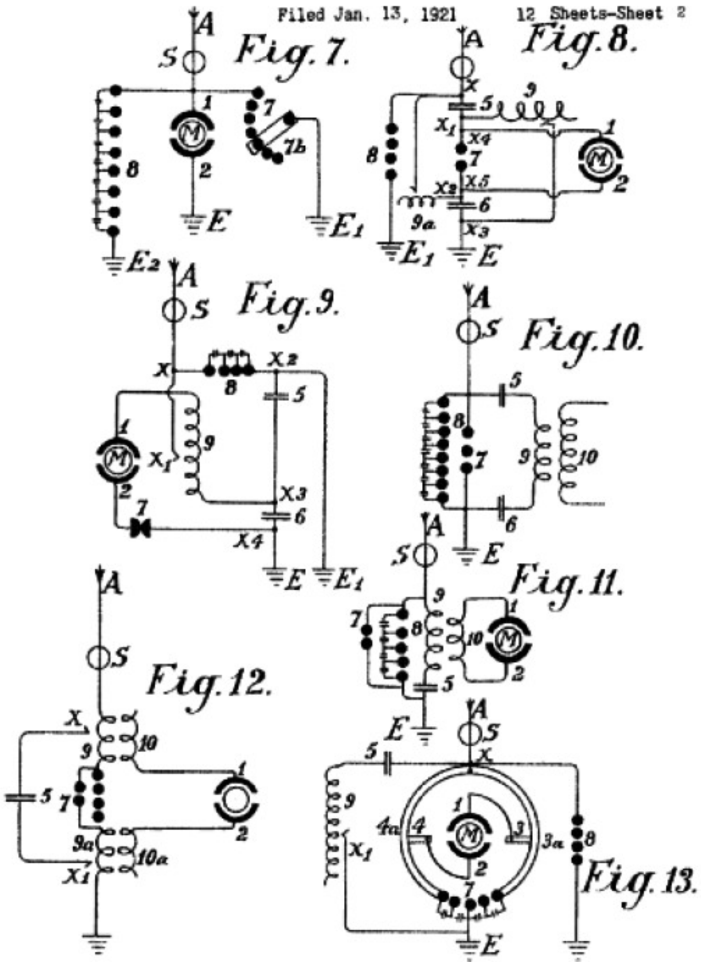

In Figure 7 the oscillation circuit is closed through the earth (E and E1). The spark gap 7 may be prolonged or shortened by more or fewer spark gaps being successively connected by means of a contact arm 7b.

Diagram 8 shows a unipolar connection of the motor with the aerial network. Here two oscillation circuits are closed through the same motor. The first oscillation circuit passes from aerial A through electromagnet S, point *x*, inductance 9a to the earth condenser 6 and further, over spark gap 7 to the aerial condenser 5 and back to *x*. The second oscillation circuit starts from the aerial condenser 5 at the point x^1 over the inductance 9 to the earth condenser 6 at the point x^3 and through the condenser 6 over the spark gap 7 back to x^1. The motor itself is inserted between the two points of the spark gap 7. From this arrangement slightly damped oscillation wave currents are produced.

In the diagram illustrated in Figure 9 a loosely coupled system of connections is illustrated which is assumed to be for small motors for measuring purposes. A indicates the aerial conductor, S the electromagnet in the aerial conductor, 9 the inductance, 7 the spark gap, 5 and 6 condensers, E the earth, M the motor, and 1 and 2 stator connections of the motor. The motor is directly metallically connected with the oscillation circuit.

In Figure 10 a purely inductive coupling is employed for the motor circuit. The motor is connected with the secondary wire 10 as may be seen in Figure 11 in a somewhat modified diagram connection. The same applies to the diagram of Figure 12.

The diagrams hitherto described preferably allow of motors of small and medium strength to be operated. For large aggregates, however, they are too inconvenient as the construction of two or more oscillation circuits for large amounts of energy is difficult; the governing is still more difficult and the danger in switching on or off is greater.

A means of overcoming such difficulties is shown in Figure 13. The oscillation circuit here runs starting from the point x over condenser 5, variable inductance 9, spark gap 7, and the two segments (3a and 4a) forming arms of a Wheatstone bridge, back to x, If the motor is connected by brushes 3 and 4 transversely to the two arms of the bridge as shown in the drawings, electromagnetic oscillations of equal sign are induced in the stator surfaces 1 and 2 and the motor does not revolve. If however the brushes 3 and 4 are moved in common with the conducting wires 1 and 2 which connect the brushes with the stator poles a certain alteration or displacement of the polarity is obtained and the motor commences to revolve.

The maximum action will result if one brush 3 comes on the central sparking contact 7 and the other brush 4 on the part *x*. They are however, usually in practice not brought on the central contact 7 but only held in the path of the bridge segments 4a and 3a in order not to connect the spark gaps with the motor oscillation circuit.

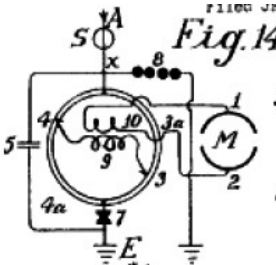

Fig. 14.

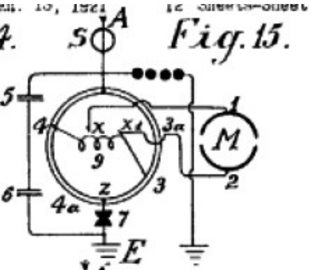

Fig. 15.

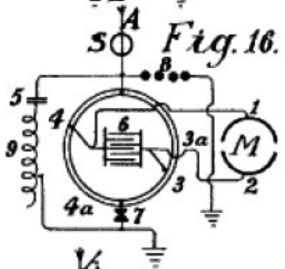

Fig. 16.

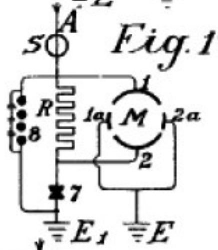

Fig. 17.

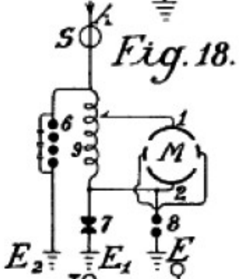

Fig. 18.

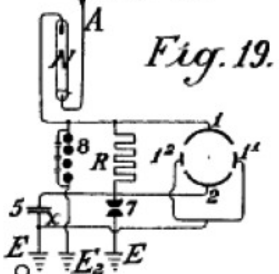

Fig. 19.

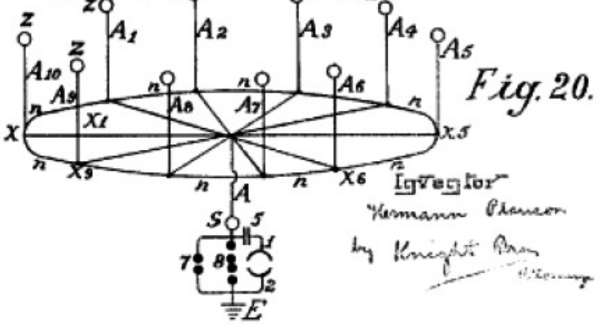

Fig. 20.

Inventor
Hermann Plauson
by Knight Bros
Attorneys

As however, the entire oscillation energy can thereby not act on the motor it is better to carry out the same system according to the diagram 14. The diagram 14 differs from the foregoing only by the motor not being directly metallically connected with the segments of the commutator, but only a primary coil 9 which induces in a secondary coil 10, current which feeds the motor M and takes the place of the rotor. By this arrangement a good transforming action is obtained, a loose coupling and also an oscillation circuit without a spark gap.

In Figure 15 the motor is not purely inductive as in 14, but directly metallically branched off from the primary coil (at x and x^1) after the principle of the auto-transformer.

In Figure 16 instead of an inductance a condenser 6 is in similar manner, and for the same object inserted between the segments 3a and 4a. This has the advantage that the segments 3a and 4a need not be made of solid metal but may consist of spiral coils whereby a more exact regulation is possible and further motors of high inductance may be employed.

The arrangements of Figures 17, 18 and 19 may be employed for use with resonance and particularly with induction condenser motors; between the large stator induction condenser surfaces, small reversing pole condenser surfaces, mall reversing pole condensers are connected, which, as may be seen from Figures 17, 18 and 19 are led together to earth. Such reversing poles have the advantage that with large quantities of electrical energy the spark formation between the separate oscillation circuits ceases.

Figure 19 shows a further method which prevents electromagnetic oscillations of high number of alternations formed in the oscillation circuit striking back to the aerial conductor. It is based on the well known principle that a mercury lamp, one electrode of which is formed of mercury, the other of solid metal such as steel allows an electric charge to pass in only one direction from the mercury to the steel and not vice versa. The mercury electrode of the vacuum tube N is therefore connected with the aerial conductor and the steel electrode with the oscillation circuit. From this it results that charges can pass only from the aerial through the vacuum tube to the oscillation circuit, but not vice versa. Oscillations which are formed on being transformed in the oscillation circuit cannot pass to the aerial conductor.

In practice these vacuum tubes must be connected behind an electromagnet as the latter alone affords no protection against the danger of lightning.

As regards the use of spark gaps, all arrangements as used for wireless telegraphy may be used. Of course the spark gaps in large machines must have a sufficiently large surface. In very large stations they are cooled in liquid carbonic acid or better still in liquid nitrogen or hydrogen; in most cases the cooling may also take place by means of liquefied low homologues of the metal series or by means of hydrocarbons the freezing point of which lies at between –90° C and –40° C. The spark gap casing must also be insulated and be of sufficient strength to be able to resist any pressure which may arise. Any undesirable excess super-pressure which may be formed must be automatically let off. I have employed wit very good results mercury electrodes which were frozen in liquid carbonic acid, the cooling being maintained during the operation from the outside through the walls.

Figure 20 is one of the simplest forms of construction of an aerial network in combination with collectors, transformers and the like illustrated diagrammatically. E is here the earth wire, 8 the safety spark gap, 7 the working spark gap, 1 and 2 the stator surfaces of the motor, 5 a condenser battery, S the protective magnet which is connected with the coil in aerial conductor, A^1 to A^{10} aerial antennae with collecting balloons, N horizontal collecting or connecting wire from which, to the center a number of connections run.

The actual collectors consist of metal sheaths preferably made of an aluminum magnesium alloy, and are filled with hydrogen or helium and are attached t copper plated steel wires. The size of the balloon is selected so that the actual weight of the balloon and the weight of the conducting wire is supported thereby. On top of the balloon aluminum spikes, made and gilded in a special manner hereinafter described, are arranged in order to produce a conductor action. Small quantities of radium preparations, more particularly polonium-ionium or meso-thorium preparations considerably increase the ionization, and therewith the action of these collectors.

In addition to metal balloons, fabric balloons which are superficially metal coated according to Schoop's metal spraying process, may also be employed. A metallic surface may also me produced by lacquering with metallic bronzes, preferably according to Schoop's spraying process or lacquering with metallic bronze powders in two electrical series of widely different metals, because thereby the collecting effect is considerably increased.

Instead of the ordinary round balloons, elongated cigar shaped ones may be employed. In order also to utilize the frictional energy of the wind, patches or strips of non-conducting substances which produce electricity by friction, may be attached to the metalized balloon surfaces. The wind will impart a portion of its energy in the form of frictional electricity, to the balloon casing, and thereby the collecting effect is substantially increased.

In practice however, very high towers (up to 300 meters is fully admissible) may be employed as antennae. In these towers copper tubes rise freely further above the top of the tower. A gas lamp secured against the wind is then lit at the point of the copper tube and a netting is secured to the copper tube over the flame of this lamp to form a collector. The gas is conveyed through the interior of the tube up to the summit. The copper tube must be absolutely protected from moisture at the place at which it enters the tower and also rain must be prevented running down the walls of the tower which might lead to a bad catastrophe. This is done by bell shaped enlargements which expand downwards, being arranged in the tower in the form of high voltage insulators of Siamese pagodas.

Special attention must be devoted to the foundations of such towers. They must be well insulated from the ground, which may be obtained by first embedding a layer of concrete in a box form to a sufficient depth in the ground and inserting in this an asphalt lining and then glass bricks cast about 1 or 2 meters in thickness. Over this in turn there is a ferro-concrete layer in which alone the metal foot of the tube is secured. This concrete block must be at least 2 meters from the ground and be fully protected at the sides by a wooden covering, from moisture. In the lower part of the tower a wood or glass house for the large condenser batteries or for the motors may be constructed. In order to lead the earth connection to the ground water, a well insulated pit constructed of vitreous brick, must be provided. Several such towers are erected at equal distances apart and connected with a horizontal conductor. The horizontal connecting wires may either run directly from tower to tower or be carried on bell shaped insulators similar to those in use for high voltage conductors. The width of the network may be of any suitable size and the connection of the motors can take place at any suitable places.

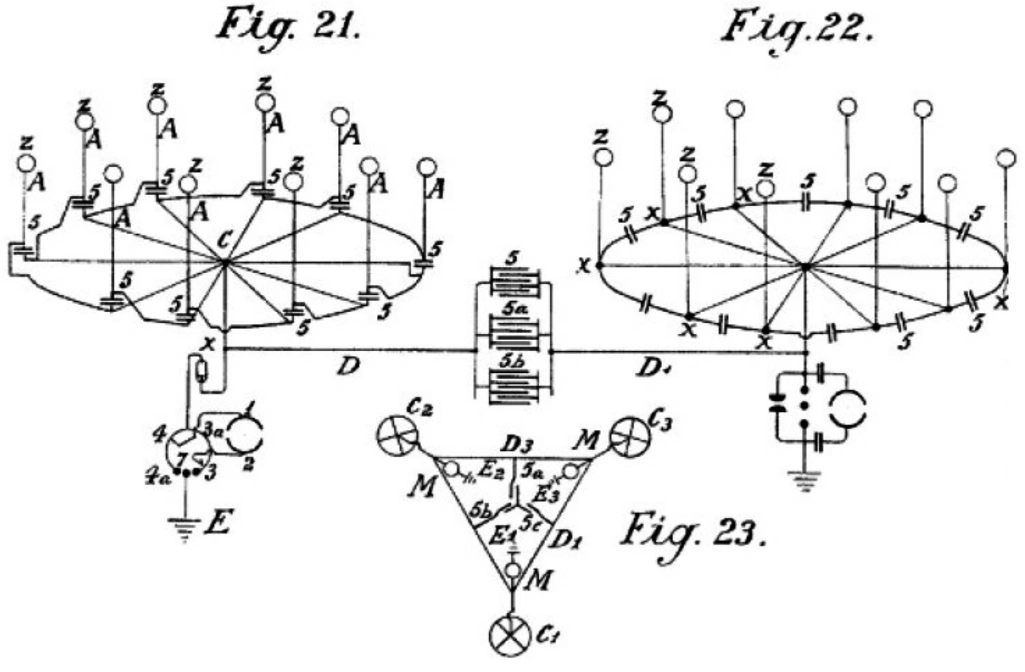

In order to collect large quantities of electricity with few aerials it is well to provide the aerial conductor with batteries of condensers as shown in Figures 21 and 22. In Figure 21 the batteries of condensers 5 are connected on the one hand with the aerial electricity collectors Z by the aerial conductor A, and on the other hand interconnected in series with an annular conductor from which horizontal conductors run to the connecting points C to which the earth wire is connected.

Figure 22 shows a similar arrangement, Should two such series of antennae rings be shown by a voltmeter to have a large difference of potential (for example, one in the mountains and one in the plain) or even of different polarity these differences may be compensated for by connecting sufficiently large condenser batteries (5, 5a, 5b) by means of Maji star conductors D and D^1. In Figure 23 a connection of three such rings of collectors to form a triangle with a central condenser battery is illustrated.

The condenser batteries of such large installations must be embedded in liquid gases or in liquids freezing at very low temperatures. In such cases a portion of the atmospheric energy must be employed for liquefying these gases. It is also preferable to employ pressure. By this means the condenser surfaces may be diminished, and still allow for large quantities of energy to be stored, secure against breakdown. For smaller installation the immersing of the condensers in well-insulated oil or the like, suffices. Solid substances on the other hand cannot be employed as insulators.

The arrangement in the diagrams hitherto described was always such that the condenser batteries were connected with both poles directly to the aerial condensers. An improved diagram of the connections for obtaining atmospheric electricity for the condenser batteries has however, been found to be very advantageous. This arrangement consists in that they are connected only by one pole (unipolar) to the collecting network. Such a method of arrangement is very important, as by means of it a constant current and an increase of the normal working pressure or voltage is obtained. If for example a collecting balloon aerial which is allowed to rise to a height of 300 meters, shows 40,000 volts above earth voltage, in practice it has been found that the working voltage (with a withdrawal of the power according to the method hereinbefore described by means of oscillating spark gaps and the like) is only about 400 volts. If however, the capacity of the condenser surfaces be increased, which capacity in the above mentioned case was equal to that of the collecting surface of the balloon aerials, to double the amount, by connecting the condenser batteries with only one pole, the voltage rises under an equal withdrawal of current up to and beyond 500 volts. This can only be ascribed to the favorable action of the connecting method.

In addition to this substantial improvement it has also been found preferable to insert double inductances with electromagnets and to place the capacities preferably between two such electromagnets. It has also been found that the useful action of such condensers can be further increased if an induction coil be connected as inductive resistance to the unconnected pole of the condenser, or still better if the condenser itself be made as an induction condenser. Such a condenser may be compared with a spring which when compressed carries in itself accumulated force, which it again gives off when released. In charging, a charge with reversed sign is formed at the other free condenser pole, and if through the spark gap a short circuit results, the accumulated energy is again given back since now new quantities of energy are induced at the condenser pole connected with the conductor network, which in fact charges with opposite signs to that at the free condenser pole. The new induced charges have of course the same sign as the collector network. The whole voltage energy in the aerial is thereby however increased. In the same space of time larger quantities of energy are accumulated than is the case without such inserted condenser batteries.

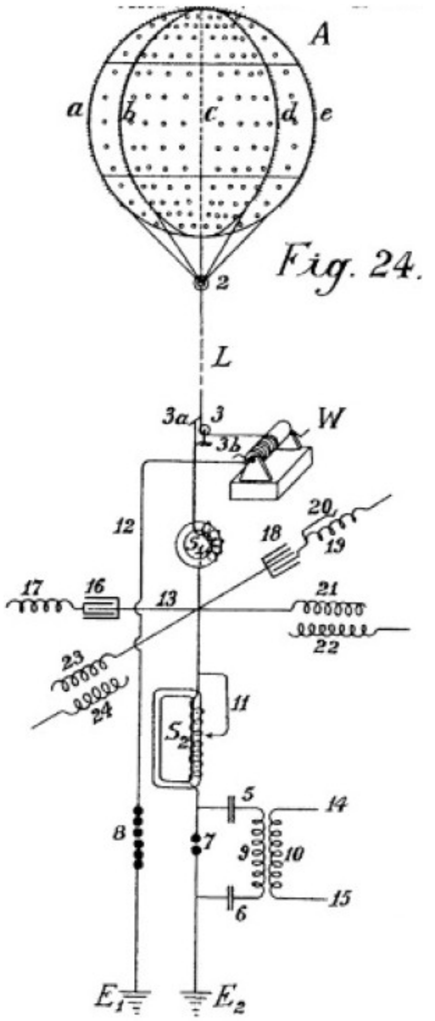
Fig. 24.

Fig. 25.

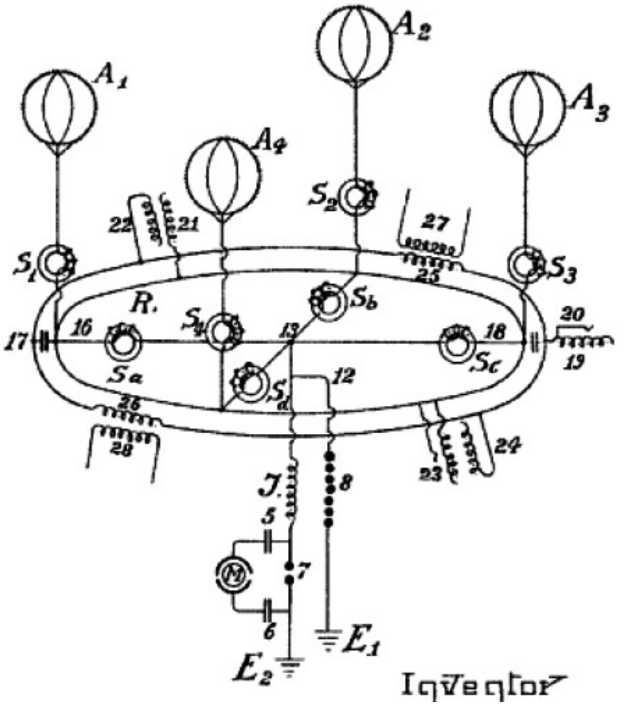

In Figures 24 and 25 two different diagrams of connections are more exactly illustrated, Figure 24 shows a collecting balloon and the diagram of the connections to earth. Figure 25 shows four collecting balloons and the parallel connection of the condenser batteries belonging thereto.

A is the collecting balloon made of an aluminum magnesium alloy (electron metal, magnalium) of a specific gravity of 1.8 and a thickness of plate 0.1 to 0.2 mm. Insider there are eight strong vertical ribs of T-shaped section about 10 to 20 mm in height and about 3 mm in thickness with the projecting part directed inwards (indicated by a, b, c, d and so forth); they are riveted together to form a firm skeleton and are stiffened in a horizontal direction by two cross ribs. The ribs are further connected with one another internally and transversely by means of thin steel wires, whereby the balloon obtains great power of resistance and elasticity. Rolled plates of 0.1 to 0.2 mm in thickness made of magnalium alloy are then either soldered or riveted on this skeleton so that a fully metallic casing with smooth external surface is obtained Well silvered or coppered aluminum plated steel wires run from each rib to the fastening ring 2. Further, the coppered steel hawser L preferably twisted out of separate thin wires (shown in dotted lines in Figure 24) and which must be long enough to allow the balloon to rise in the desired height, leads to a metal roller or pulley 3 and from thence to a winch W, well insulated from the earth. By means of this winch, the balloon, which is filled with hydrogen, or helium, can be allowed to rise to a suitable height (300 to 5000 meters) and brought to the ground for recharging or repairs.

The actual current is taken directly through a friction contact from the metal roller 3 or from the wire, or even from the winch or simultaneously from all three by means of brushes (3, 3a and 3b). Beyond the brushes the conductor is divided, the paths being: --- firstly over 12 to the safety spark gap 8, from thence to the earth conductor E^1, and secondly over electromagnet S^1, point 13, to a second loose electromagnet having an adjustable coil S^2, then to the spark gap 7 and to the second earth conductor E^2. The actual working circuit is formed through the spark gap 7, condensers 5 and 6, and through the primary coil 9; here the static electricity formed by oscillatory discharges is accumulated and converted into high frequency electromagnetic oscillations. Between the electromagnets S^1 and S^2 at the crossing point 13, four condenser batteries are introduced which are only indicated diagrammatically in the drawings each by one condenser. Two of these batteries (16 and 18) are made as plate condensers and prolonged by regulating induction coils or spirals 17 and 19 while the two others (21 and 23) are induction condensers. As may be seen from the drawings each of the four condenser batteries 16, 18, 21 and 23 is connected by only one pole to the aerial or to the collector conductor. The second poles 17, 19, 22 and 24 are open. In the case of plate condensers having no inductive resistance an induction coil is inserted. The object of such a spiral or coil is the displacement of phase by the induction current by 1/4 periods, whilst that of the charging current of the condenser poles which lie free in the air, works back to the collector aerial. The consequence of this is that in discharges in the collector aerial the back inductive action of the free poles allows a higher voltage to be maintained in the aerial collecting conductor than would otherwise be the case. It has also been found that such a back action has an extremely favorable effect on the wear of the contacts. Of course the inductive effect may be regulated at will within the limits of the size of the induction coil, the length of the coil in action being adjustable by means of wire connection without induction (see Figure 24, No. 20).

S^1 and S^2 may also be provided with such regulating devices in the case of S^2 (illustrated by 11). If excess voltage be formed it is conducted to earth through the wire 12 and spark gap 8 or through any other suitable apparatus, since this formation would be dangerous for the other apparatus.

The small circles on the collector balloon indicate places at which zinc amalgam or gold amalgam or other photoelectric acting metals in the form of small patches in extremely thin layers (0.01 to 0.05 mm in thickness) are applied to the entire balloon as well as in greater thickness to the conducting network. The capacity of the collector is thereby considerably strengthened at the surface. The greatest possible effect in collecting may be obtained by polonium amalgams and the like. On the surface of the collector balloon metal points or spikes are also fixed along the ribs, which spikes serve particularly for collecting the collector charge. Since it is well known that the resistance of the spikes is less the sharper the spike is, for this purpose it is therefore extremely important to employ as sharp spikes as possible. Experiments made as regards these have shown that the formation of the body of the spike or point also plays a large part, for example, spikes made of bars or rollers with smooth surface, have a many times greater point resistance as collector accumulator spikes than those with rough surfaces. Various kinds of spike bodies have been experimented with for the collector balloons hereinbefore mentions. The best results were given by spikes which were made in the following way. Fine points made of steel, copper, nickel, or copper and nickel alloys, were fastened together in bundles and then placed as anode with the points in a suitable electrolyte (preferably in hydrochloric acid or muriate of iron solutions) and so treated with weak current at 2 to 3 volts pressure. After 2 to 3 hours according to the thickness of the spikes or pins the points become extremely sharp and the bodies of the spikes have a rough surface. The bundle can then be removed and the acid washed off with water. The spikes are then placed as cathode in a batch consisting of solution of gold, platinum, iridium, palladium or wolfram salts or their compounds and coated a the cathode galvanically with a thin layer of precious metal, which must however be sufficiently firm to protect them from atmospheric oxidation.

Such spikes act at a 20-fold lower voltage almost as well as the best and finest points made by mechanical means. Still better results are obtained if polonium or radium salts are added to the galvanic bat when forming he protective layer or coating. Such pins have a low resistance at their points and even at one volt and still lower pressures have an excellent collector action.

In Figure 24 the three unconnected poles are not connected with one another in parallel. That is quite possible in practice without altering the principle of the free pole. It is also preferable to interconnect in parallel to a common collector network, a series of collecting aerials.

Figure 25 shows a diagram for such an installation. A^1, A^2, A^3, A^4 are four metal collector balloons with gold or platinum coated spikes which are electrolytically made in the presence of polonium emanations or radium salts, which spikes or needles are connected over four electromagnets S^1, S^2, S^3, S^4, through an annular conductor R. From this annular conductor four wires run over four further electromagnets Sa, Sb, Sc, Sd, to the connecting point 13. There the conductor is divided, one branch passing over 12 and the safety spark gap 8 to the earth at E^1, the other over inductive resistance J and working spark gap 7 to the earth at E^2. The working circuit, consisting of the condenser 5 and 6 and a resonance motor M, such as hereinbefore described, is connected in proximity round the sparking gap section 7.

Instead of directly connecting the condenser motor of course the primary circuit for high frequency oscillatory current may also be inserted.

The condenser batteries are connected by one pole to the annular conductor R and can be either inductionless (16 and 18) or made as induction condensers as shown by 21 and 23. The free poles of the inductionless condensers are indicated by 17 and 19, those of the induction condensers by 22 and 24. As may be seen from the drawings all these poles 17, 22, 19, 24 may be interconnected in parallel through a second annular conductor without any fear that thereby the principle of the free pole connection will be injured. In addition to the advantages already set forth the parallel connection also allows of an equalization of the working pressure in the entire collector network. Suitable constructed and calculated induction coils 25 and 26 may also be inserted in the annular conductor of the free poles, by means of which a circuit may be formed in the secondary coils 27 and 28 which allows current produced in this annular conductor by fluctuations of the charges of the like appearances to be measured or otherwise utilized.

According to what has been hereinbefore stated separate collector balloons may be connected at equidistant stations distributed over the entire country, either connected directly with one another metallically or by means of intermediate suitably connected condenser batteries through high voltage conductors insulated from earth. The static electricity is converted through a spark gap into dynamic energy of a high number of oscillations and may in such form be coupled as a source of energy b y means of a suitable method f connecting, various precautions being observed and with special regulations. The wires leading from the collector balloons have hitherto been connected through an annular conductor without this endless connection, which can be regarded as an endless induction coil, being able to exert any action on the whole conductor system.

It has now been found that if the network conductor connecting the aerial collector balloons with one another is not made as a simple annular conductor, but preferably short circuited in the form of coils over a condenser battery or spark gap or through thermionic tubes or valves or audions, then the total collecting network exhibits quite new properties. The collection of atmospheric electricity is thereby not only increased but an alternating field may be easily produced in the collector network. Further, the atmospheric electrical forces showing themselves in the higher regions may also be directly obtained by induction. In Figures 26 and 28 a form of construction is shown on the basis of which the further foundations of the method will be more particularly explained.

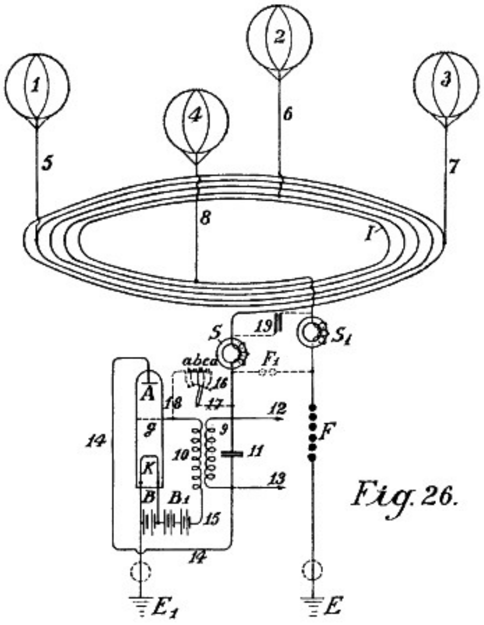

Fig. 26.

In Figure 26, 1, 2, 3, 4 are metal collector balloons, 5, 6, 7, 8 their metallic aerial conductors and I the actual collector network. This consists of five coils and is mounted on high voltage insulators in the air, in high voltage masts (or with a suitable construction of cable embedded in the earth). One coil has a diameter of 1 or 100 km or more. S and S^1 are two protective electromagnets, F the second safety section against excess voltage, E its earth conductor and E1 the earth conductor of the working section. When an absorption of static atmospheric electricity is effected through the four balloon collectors, the current in order to reach the earth connection E1 must flow spirally through the collector network over the electromagnet S, primary induction coil 9, conductor 14, anode A of the audion tube, incandescent cathode K, as the way over the electromagnet and safety spark gap F offers considerably greater resistance. Owing to the fact that the accumulated current flows in one direction, an electromagnetic alternating field is produced in the interior of the collector network coil, whereby the whole free electrons are directed more or less into the interior of the coil. An increased ionization of the atmosphere is thereby produced. In consequence of this the points mounted on the collector balloon show a considerably reduced resistance and therefore increased static charges between the points on the balloon and the surrounding atmosphere are produced. The result of this is a considerably increased collector effect.

A second effect which could not be obtained otherwise is obtained by the electromagnetic alternating field which running parallel to the earth surface, acts more or less with a diminishing or increasing effect on the earth magnetic field, whereby in the case of fluctuations in the current a return induction current of reversed sign is always produced in the collector coil by earth magnetism. Now if, however, a constantly pulsating continuous alternating field is produced as stated in the above collector network I, an alternating current of the same periodicity is produced also in the collecting network coil. As the same alternating field is further transmitted to the aerial balloon, the resistance to its points is thereby considerably reduced, whilst the collector action is considerably increased. A further advantage is that positive electrons which collect on the metal surfaces during the conversion into dynamic current produce a so-called drop of potential of the collector area. As an alternating field is present, the negative ions surrounding the collectors surfaces, when discharge of the collector surfaces takes place produce by the law of induction, an induction of reversed sign on the collector surface and so forth (that is to say again a positive charge). In addition to the advantages hereinbefore set forth, the construction of connecting conductors in coil form when of sufficiently large diameter, allows of a utilization of energy arising in higher regions also in the simplest way. As is well known electric discharges frequently take place at very great elevations which may be observed as St Elmo’s fire or northern lights. These energy quantities have no been available to be utilized up to now. By this invention all these kinds of energy, as they are of an electromagnetic nature and the direction of axis of the collector coils stands at right angles to the earth’s surface, can be more or less absorbed in the same way as a receiver in wireless telegraphy absorbs waves coming from a far distance. With a large diameter of the spiral it is possible to connect large surfaces and thereby to take up also large quantities of energy.

It is well known that large wireless stations in the summer months, and also in the tropics are very frequently unable to receive the signals in consequence of interruptions which are caused by atmospheric electricity, and this takes place with vertical coils of only 40 to 100 meters diameter. If on the contrary horizontal coils of 1 to 100 km diameter be employed very strong currents may be obtained through discharges which are constantly taking place in the atmosphere. Particularly in the tropics or still better in the polar regions where the northern lights are constantly present, large quantities of energy may probably be obtained in this way. A coil with several windings should act the best. In similar manner any alteration of the earth magnetism should act inductively on such a coil.

It is not at all unlikely that earthquakes and spots on the sun will also produce an induction in such collector coils of sufficient size. In similar manner this collector conductor will react on earth current more particularly when they are near the surface of the earth or even embedded in the earth. By combining the previous kind of current collectors so far as they are adapted for the improved system with the improved possibilities of obtaining current the quantities of free natural electricity which are to be obtained in the form of electricity are considerably increased.

In order to produce in the improved collector coil uniform current oscillations of an undamped nature so-called audion high vacuum or thermionic tubes of suitable connection are employed instead of the previously known spark gaps (Figure 26, Nos. 9-18). The main aerial current flows through electromagnet S (which in the case of a high number of alternations is not connected here but in the earth conductor E^1) and may be conveyed over the primary coils in the induction winding through wire 14 to the anode A of the high vacuum grid tube. Parallel with the induction resistance 9 a regulating capacity of suitable size, such as condenser 11 is inserted. In the lower part of the vacuum grid tube is arranged the incandescent filament or the cathode K which is fed through a battery B. From the battery B two branches run, one to the earth conductor E^1 and the other through battery B^1 and secondary coil 10 to the grid above *g* in the vacuum tube. By the method of connections shown in dotted lines, a desired voltage at the grid electrode *g* may also be produced through the wire 17 which is branched off from the main current conductor through switches 16 and some small condensers (*a, b, c, d*) connected in series, and conductor 18, without the battery B^1 being required.

The action of the entire system is somewhat as follows: --

On the connecting conductor of the aerial collector network being short circuited to earth, the condenser pole 11 is charged and slightly damped oscillations are formed in the short circuited existing oscillation circuit formed of the condenser 11 and self inductance 9. In consequence of the coupling through coil 10, fluctuations of voltage take place in the grid circuit 15 with the same frequency, which, fluctuations in turn influence the strength of the electrode current passing through the high vacuum amplifying tube and thus produce current fluctuations of the same frequency in the anode circuit. A permanent supply of energy to the oscillation circuits 9 and 10 consequently takes place, until a condition of balance is set up, in which the consumed oscillation energy is equal to that absorbed. Thereby constant undamped oscillations are now produced in the oscillation circuits 9-11.

For regular working of such oscillation producers high vacuum strengthening tubes are necessary and it is also necessary that the grid and anode voltages shall have a phase difference of 180° so that if the grid is negatively charged, then the anode is positively charged and vice versa. This necessary difference of phase may be obtained by most varied connections, for example, by placing the oscillation circuit in the grid circuit or by separating the oscillation circuit and inductive coupling from the anodes and the grid circuit and so forth.

A second important factor in this way of converting static atmospheric electricity into undamped oscillations is that care must be taken hat the grid and anode voltages have a certain relation to one another; the latter may be obtained by altering the coupling and a suitable selection of the self-induction in the grid circuit, or as shown by dotted lines 16, 17, 18 by means of a larger or smaller number of condensers of suitable size connected in series; in this case the battery B^1 may be omitted. With a suitable selection of the grid potential a glow discharge takes place between the grid *g* and the anode A, and accordingly at the grid there is a cathode drop and a dark space is formed. The size of this cathode drop is influenced by the ions which are emitted in the lower space in consequence of shock ionization of the incandescent cathodes K and pass through the grid in the upper space. On the other hand the number of ions passing through the grid is dependent on the voltage between the grid and the cathode. Thus is the grid voltage undergoes periodic fluctuations (as in the present case) the amount of the cathode drop at the grid fluctuates and consequently the internal resistance of the tube correspondingly fluctuates, so that when a back coupling of the feed circuit with the grid circuit takes place, the necessary means are afforded for producing undamped oscillations and of taking current, according to requirements from the collecting conductor.

The frequency of the undamped oscillations produced is with a suitably loose coupling equal to the self-frequency of the oscillation circuits 9 and 10. By a suitable selection of the self induction of the coil 9 and capacity 11 it is possible to extend from frequencies which produce electromagnetic oscillation of only a few meters wavelength down to the lowest practical alternating current frequency. For large installations a suitable number of frequency-producing tubes of the well-known high vacuum transmission tubes of 0.5 to 2 kw in size may be connected in parallel so that in this respect no difficulty exists.

The use of such tubes for producing undamped oscillations, and also the construction and method of inserting such transmission tubes in an accumulator or dynamo circuit is known and also that such oscillation producing tubes only work well at voltages of 1000 up to 4000 volts, so that on the contrary their use at lower voltages is considerably more difficult. By the use of high voltage static electricity this method of producing undamped oscillations as compared with that through spark gaps must be regarded as an ideal solution particularly for small installations of outputs of from 1 to 100 kw.

By the application of safety spark gaps, with interpolation of electromagnets, not only is short-circuiting avoided but also the taking up of current is regulated. Oscillation producers inserted in the above way form a constantly acting electromagnetic alternating field in the collector coil, whereby as already stated, a considerable accumulating effect takes place. The withdrawal wire or working wire is connected at 12 and 13, but current may be taken by means of a secondary coil which is firmly or movable mounted in any suitable way inside the large collector coil, i.e., in its electromagnetic alternating field, so long as the direction of its axis runs parallel with that of the main current collecting coil.

In producing undamped oscillations of a high frequency (50,000 per second or more) in the oscillation circuits 9 and 11, electromagnets S and S^1 must be inserted if the high frequency oscillations are not to penetrate the collector oil, between the oscillation producers and the collector coil. In all other cases they are connected shortly before the earthing (as in Figures 27 and 28).

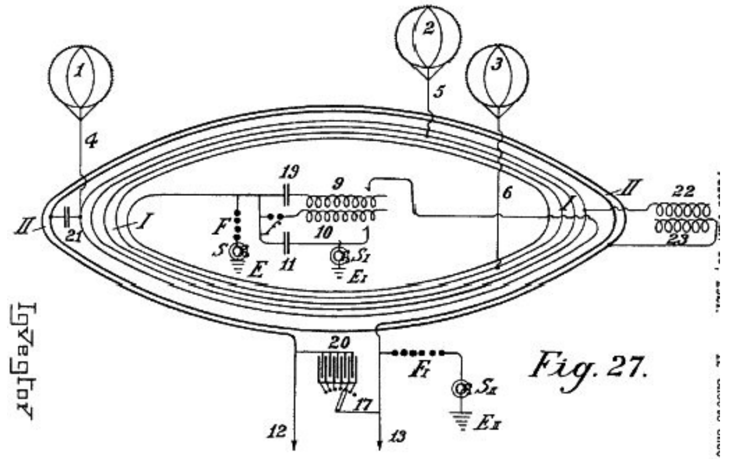

Fig. 27.

In Figure 27 a second method of construction of the connecting conductor of the balloon aerials is illustrated in the form of a coil. The main difference consists in that in addition to the connecting conductor I another annular conductor II is inserted parallel to the former on the high voltage masts in the air (or embedded as a cable in the earth) but both in te form of a coil. The connecting wire of the balloon aerials is indicated as a primary conductor and also as a current producing network; the other is the consumption network and is not in unipolar connection with the current producing network.

In Figure 27 the current producing network I is shown with three balloon collectors 1, 2, 3 and aerial conductors 4, 5, 6; it is short-circuited through condenser 19 and inductance 9. The oscillation forming circuit consists in this diagram of spark gap f, inductance 10, and condenser 11; the earth wire E, is connected to earth over electromagnet S^1. F s the safety spark gap which is also connected to earth through a second electromagnet S at E. On connecting up the condenser circuit 11 this is charged over the spark gap f whereby an oscillatory discharge is formed. This discharging current acts through inductance 10 on the inductively coupled secondary 9, whereby in the producing network a modification of the potential of the condenser 19 is produced. The consequence of this is that oscillations arise in the coil shaped producer network. These oscillations induce a current in the secondary circuit II, which has a smaller number of windings and a less resistance, the voltage of which, according to the proportion of the number of windings and of the ohmic resistance, is considerably lower whilst the current strength is greater.

In order to convert the current thus obtained into current of an undamped character, and to tune its wavelengths, a sufficiently large regulatable capacity 20 in inserted between the ends 12 and 13 of the secondary conductor II. Here also current may be taken without an earthy conductor, but it is advisable to insert a safety spark gap E1 and to connect this with the earth over an electromagnet S2.

The producer network may be connected with the working network II over an inductionless condenser 21 or over an induction condenser 22, 23. In this case the secondary conductor is unipolarly connected with the energy conductor.

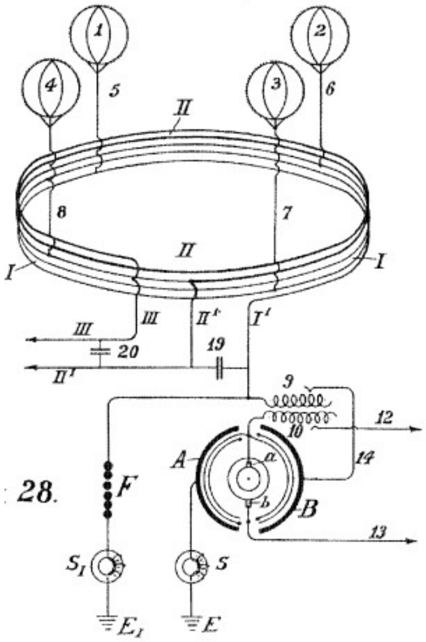

28.

In Figure 28 the connecting conductor between the separate accumulator balloons is carried out according to the autotransformer principle. The collecting coil connects four aerial balloons 1, 2, 3, 4, the windings of which are not made side by side but one above the other. In Figure 28 the collector coil I is shown with a thin line, the metallically connected prolongation coils II with a thick line. Between the ends I^1 and II^1 of the energy network I a regulating capacity 19 is inserted. The wire I^1 is connected with the output wire and with the spark gap F.

As transformer of the atmospheric electricity an arrangement is employed which consists in using rotary pairs of condensers in which the one stator surface B is connected with the main current, whilst the other A is connected with the earth pole. Between these pairs of short-circuited condensers are caused to rotate from which the converted current can be taken by means of two collector rings and brushes, in the form of an alternating current, the frequency of which is dependent on the number of balloons and the revolutions of the rotor. As the alternating current formed in the rotor can act, in this improved method of connection described in this invention, through coils 1 on the inductance 9, an increase or diminution of the feed current in I can be obtained according to the direction of the current by back induction. Current oscillations of uniform rhythm thereby result in the coil shaped windings of the produce network.

As the ends of this conductor are short-circuited through the regulatable condenser 19 these rhythms produce short-circuited undamped oscillations in the energy conductor, the periodicity and wave lengths of which oscillations can be adjusted according to desire by altering the capacity 19 to a given wavelength and therewith also to a given frequency. These currents may also be employed in this form directly as working current through the conductors II^1 and III. By inserting the condenser 20 a connection between these conductors may also be made, whereby harmonic oscillations of desired wavelengths are formed. By this means quite new effects as regards current distribution are obtained. The withdrawal of current can even take place without direct wire connection if, at a suitable point in the interior of the producing network (quite immaterially whether this has a diameter of 1 or 100 km) a coil tuned to these wavelengths and of the desired capacity is firmly or movably mounted in the aerial conductor in such a way that is axial direction is in parallel with that of the collector coil. In this case a current is induced in the producing network, the size of which is dependent on the total capacity and resistance and also on the periodicity employed. A possibility is thereby afforded in future, of taking energy from the producer network by wireless means. As thereby in addition to atmospheric electricity also magnetic earth currents and energy from the higher atmosphere (at leas partially may be simultaneously obtained, this last system for collecting the atmospheric energy is of particular importance for the future.

Of course everywhere instead of spark gaps grid vacuum tubes may be employed as producers for undamped oscillations. The separate coils of the producer network with large diameters may be connected with one another through separate conductors all in parallel or all in series or in groups in series. By regulating the number of oscillations and also the extent of the voltage more or less large collector coils of this kind may be employed. The coils may also be divided spirally over the entire section. The coils may be carried out in angular form or also in triangular, quadrangular, hexagonal or octagonal form.

Of course wires may be carried from a suitable place to the center or also laterally which serve the current waves as guides. This is necessary when the currents have to be conducted over mountains and valley and so forth. In all these cases the current must be converted into a current of suitable periodicity.

As already hereinbefore mentioned separate collecting balloons may be directly metallically interconnected at equidistant stations distributed over the entire country or may be connected by interpolation of suitable condenser batteries by means of high voltage conductors. The static electricity is converted through a spark gap into dynamic energy of a high number of oscillations, and could then in such forms, wit a suitable arrangement of the connections, observing various measures of protection, be employed as source of energy after separate or special regulation.

According to this invention in order to increase the collecting effect of the balloon in the aerial collector conductor or in the earth wire, radiating collectors are employed. These consist either of incandescent metal or oxide electrodes in the form of vacuum grid tubes, or electric arcs (mercury and the like electrodes), Nernst lamps, or finally flames of various kinds may be simply connected with the respective conductor.

It is well known that energy can be drawn of from a cathode consisting of an incandescent body opposite an anode charged with positive electricity (vacuum grid tube). Hitherto however, a cathode was always first directly placed opposite an anode, and secondly the system always consisted of a closed circuit.

Now if we dispense with the ordinary ideas in forming light or flame arcs in which a cathode must always stand directly opposite an anode, charged to a high potential or another body freely floating in the air, or regard the incandescent cathode only as a source of unipolar discharge (which represent group and point discharges in electrostatic machines similar to unipolar discharges), it may be ascertained that incandescent cathodes and less perfectly all incandescent radiators, flames and the like admit of relatively large current densities and allow large quantities of electric energy to radiate into the open space in the form of electron streams as transmitters.

The object of this invention is as described below, if such incandescent oxide electrodes or other incandescent radiators or flames are not freely suspended in space but connected metallically with the earth so that they can be charged with negative terrestrial electricity, these radiators possess the property of absorbing the free positive electrical charges contained in the air space surrounding them (that is to say of collecting them and conducting them to earth). They can therefore serve as collectors and have, in comparison to the action of the spikes, or points, a very large radius of action R; the effective capacity of these collectors is much greater than the geometrical capacity (R_{o-}) calculated in an electrostatic sense.

Now as our earth is surrounded as is well known with an electrostatic field and the difference of potential

$$\frac{\delta V}{\delta h}$$

of the earth field according to the latest investigations, is in summer about 60 to 100 volts and in winter 300 to 500 volts per meter of difference in height (δh), a simple calculation gives the result that when such a radiation collector or flame collector is arranged for example on the ground, and a second one is mounted vertically over it at a distance of 2000 meters and both are connected by a conducting cable, there is a difference in potential in summer of about 2,000,000 volts and in winter even of 6,000,000 volts and more.

According to Stefan Boltzmann's law of radiation, the quantity of energy which an incandescent surface (temperature T) of 1 sq cm radiates in a unit of time into the open air (temperature T_o) is expressed by the following formula:

$S = \delta \, (T^4 - T^4_o)$ watt./sq cm.

And the universal radiation constant δ is according to the latest researches of Ferry (*Annales de Chimie et de Physique*, 17: 267 [1909]) equal to 6.30 x 10^{-12} watt/sq cm.

Now if an in incandescent surface of 1 sq cm shows, as compared with the surrounding space a periodic fall of potential δV it radiates (independent of the current direction, that is to say of the sign) in accordance with the above formula, for example at a temperature of 3725° C an energy of 1.6 kw/sq cm/second. As for the radiation the same value can be calculated for the collection of energy, but reversed. Now, as carbon electrodes at the temperature of the electric arc support on the current basis a current density up to from 60 to 65 amperes per sq cm no difficulties will result in this direction in employing radiating collectors as accumulators.

If the earth be regarded as a cosmically insulated condenser in the sense of geometrical electrostatics *x* there results from the geometric (compare Edwald Rasch: *Das Elektrische Bogenlicht* [The Electric Arc Light], page 169) capacity of the earth according to Chwolson:

For negative charging 1.3 x 10^6 Coulomb

For negative potential V = 10 x 10^8 volts.

From this there results however, EJT = 24.7 x 10^{24} watt/sec. Now if it is desired to make a theoretic short circuit through an earthed flame collector this would represent an electric total work of about 79,500 x 10^{10} kilowatt years. As the earth must be regarded as a rotating mechanism which thermodynamically, electromagnetically and also kinematically coupled with the sun and star system by cosmic radiations and gravitation a diminution of the electric energy of the earth field is not to be feared. The energies which the incandescent collectors would withdraw from the earth field can only case by the withdrawal of motor work a lowering of the earth temperature (temperature $T_E = 300$) and reduce this to that of the world space (T = O) by using the entire energy. This is however not the case as the earth does not represent a cosmically entirely insulated system. On the contrary there is conveyed to the same according to the recent value corrected by Perry for the solar constants through the radiation from the sun an energy of 18,500 x 10^{10} kw. Accordingly, any lowering of the earth temperature (T_E) without a simultaneous lowering of the sun's temperature (T_s) would contradict Stefan Boltzmann's law of radiation.

$S = \delta \, (s^4 - T^4)$.

From this it must be concluded that if the earth temperature (T_E) sinks the total radiation S absorbed by the earth increases, and further also that the secular speed of cooling of the earth is directly dependent on that of the sun and the other radiators cosmically coupled with the sun and is connected most closely with these.

The incandescent radiation collectors may, according to this invention, be employed for collecting atmospheric electricity if they (1) are charged with the negative earth electricity (that is to say when they are directly connected by means of a metallic conductor with the earth) and (2) if large capacities (metal surfaces) charged with electricity are mounted opposite them as positive poles in the air. This is regarded as the main feature of the present invention as without these inventive ideas it would not be possible to collect with an incandescent collector, sufficiently large quantities of the electrical charges contained in the atmosphere as the technology requires; the radius of action of the flame collectors would also be too small, especially if it be considered that the very small surface density (energy density) (δ about = 2 x 7 . 10^9 St. E. per sq cm) does not allow of large quantities of charge being absorbed from the atmosphere.

x) Calculated according to Poisson's calculation:

$\Delta V = -4\pi\delta$; as here the alteration of the potential or potential gradients only takes place in the direction o the normal, this calculation assumes the simple form

$$\delta = \frac{1}{4\pi} \times \frac{\delta^2 V}{\delta n^2}$$

It has indeed already been proposed to employ flame collectors for collecting atmospheric electricity and it is known that their collecting effect is substantially greater opposite the points. It is however, not known that the quantities of current which could hitherto be obtained are too small for technical purposes. According to my experiments the reason for this is to be found in the too small capacities of the collector conductor poles. If such flame or radiating collectors have no or only small positive surfaces, their radius of action for large technical purposes is too small. If the incandescent collectors be constantly kept in movement in the air they may collect more according to the speed of the movement, but this again is not capable of being carried out in practice.

By this invention the collector effect is considerably increased by a body charged with a positive potential and of the best possible capacity being also held floating (without direct earth connection) opposite such an incandescent collector which is held floating in the air at a desired height. If for example, a collecting balloon of sheet metal or of metalized balloon fabric be caused to mount to 300 up to 3000 meters in the air and as positive pole it is brought opposite such a radiating collector connected by a conductor to the earth, quite different results are obtained.

The metallic balloon shell (with a large surface) is charged to a high potential by the atmospheric electricity. This potential is greater the higher the collecting balloon is above the incandescent collector. The positive electricity acts concentratedly on the anode floating in the air as it is attracted through the radiation shock ionization, proceeding from the incandescent cathode. The consequence of this is that the radius of action of the incandescent cathode collector is considerably increased and thereby also the collection effect of the collecting balloon surface. Further the large capacity of the anode floating in the air plays therefore an important part because it allows of the taking o large charges, and thereby a more uniform current is obtained even when there is a large consumption: this cannot be the case with small surfaces.

In the present case the metallic collecting balloon is a positive anode floating in the air and the end of the earth conductor of this balloon serves as positive pole surface opposite the surface of the radiating incandescent cathode, which in turn is charged with negative earth electricity being conductively connected to earth.

The process may be carried out by two such contacts (negative incandescent cathode and anode end of a capacity floating in the air) a condenser and an inductive resistance being switched on in parallel, whereby simultaneously undamped oscillations may be formed.

In very large installations it is advisable to connect two such radiating collectors in series. Thus an arc light incandescent cathode may be placed below on the open ground and an incandescent cathode which is heated by special electromagnetic currents be located high in the air. Of course for this the special vacuum Liebig tubes wit or without grids may also be employed. An ordinary arc lamp with oxide electrodes may be introduced on the ground and the positive pole is not directly connected with the collecting balloon, but through the upper incandescent cathode or over a condenser. The method of connecting the incandescent cathode floating in the air may be seen in Figures 29-33.

B is the air balloon, K a Cardan ring (connection with the hawser), C the balloon, L a good connecting cable, P a positive pole, N negative incandescent cathode, and E earth conductor.

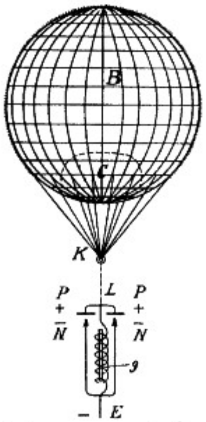

Fig. 29.

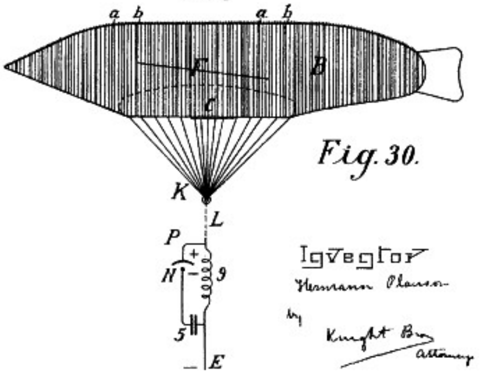

Fig. 30.

Figure 29 represents the simplest form of construction. If electric oscillations are produced below on the ground by means of a carbon arc lamp or in other suitable way a considerably greater electric resistance is opposed to that in the direct way by inserting an electrical inductive resistance 9. Consequently between P and N, a voltage is formed, and as, over N and P only an inductionless ohmic resistance is present, a spark will spring over so long as the separate induction co-efficients and the like are correctly calculated. The consequence of this is that the oxide electrode (carbon or the like) is rendered incandescent and then shows as incandescent cathode an increased collecting effect. The positive poles must be substantially larger than the negative in order that they may not also become incandescent. As they are further connected with the large balloon area which has a large capacity and is charged at high voltage, an incandescent body which is held floating in the air and a positive pole which can collect large capacities is thereby obtained in the simplest way. The incandescent cathode is first caused to become incandescent by means of separate energy produced on the earth, and then maintained by the energy collected from the atmosphere.

Figure 30 only shows the difference that instead of a round balloon a cigar-shaped one (of metal or metalized fabric) may be employed and also a condenser 5 is inserted between the incandescent cathode and the earth conductor so that a short circuited oscillation circuit over P.N. 5 and 9 is obtained. This has the advantage that quite small quantities of electricity cause the cathode to become incandescent and much larger cathode bodies may be rendered incandescent.

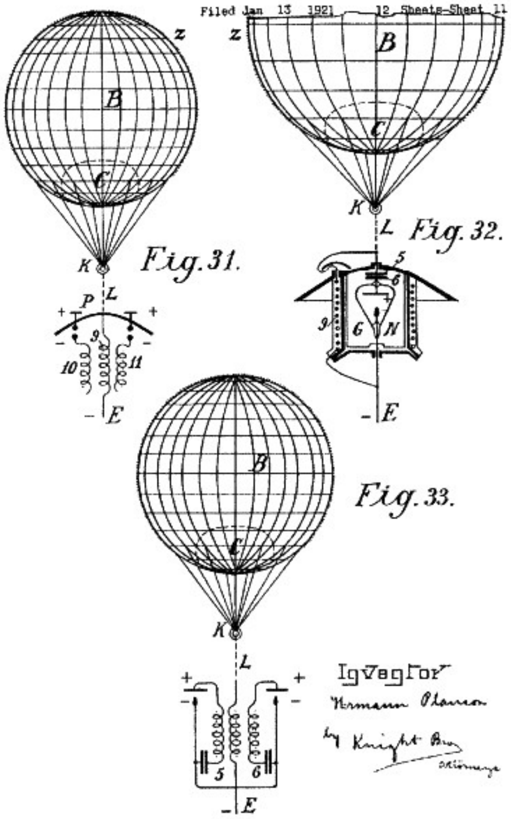

Fig. 31.

Fig. 32.

Fig. 33.

In this form of construction both the incandescent cathode and also the positive electrode may be enclosed in a vacuum chamber as may be seen in Figure 32. A cable L is carried well insulated through the cover of a vessel and ends in a condenser disc 5. The cover is arched in order to keep off the rain. The vessel is entirely or partially made of magnetic metal and well-insulated inside and outside. Opposite the disc 5 another disc 6 and on this again a positive pile of the vacuum tube g with the incandescent cathode (oxide electrode) N is arranged. The negative electrode is on the one hand connected with the earth conductor E, and on the other hand with the inductive resistance 9 which is also connected with the cable L with the positive pole and wound round the vessel in coils. The action is exactly the same as that in Figure 29 only instead of an open incandescent cathode one enclosed in vacuo is employed. As in such collectors only small bodies can be brought to incandescence in large installations a plurality of such vacuum tubes must be inserted in proximity to one another. According to the previous constructions Figures 31 and 33 are quite self evident without further explanations.

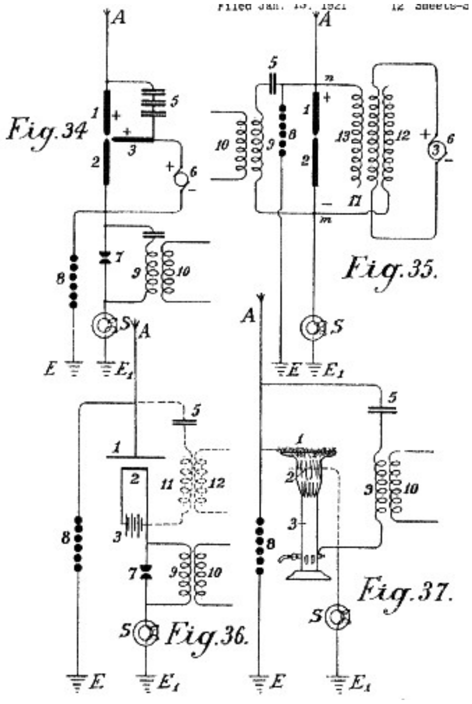

Figures 34-37 represent further diagrams of connections over radiating and flame collectors, and in fact, how they are to be arranged on the ground.

Figure 34 shows an arc light collector with oxide electrodes for direct current and its connection; Figure 35 a similar one for alternating current, Figure 36 an incandescent collector with a Nernst lamp and Figure 37 a similar one with a gas flame.

The positive pole 1 of the radiating collectors is always directly connected to the aerial collecting conductor A. In Figure 34 this is further connected over the condenser battery 5 with a second positive electrode 3. The direct current dynamo *b* produces current which flows over between the electrodes 3 and 2 as an arc light. On the formation of an arc the negative incandescent electrode 2 absorbs electricity from the positive poles standing opposite it and highly charged with atmospheric electricity and conveys the same to the working circuit. The spark gap 7, inductive resistance 9 and induction coil 10 are like the ones previously described. The protective electromagnet S guards the installation against earth circuiting, the safety spark gap 8 from excess voltage or overcharging.

In Figure 35 the connection is so far altered that the alternating current dynamo feeds the exciting coil 11 of the induction condenser. 12 is its negative and 13 its positive pole; if the coil 3 on the magnet core of the dynamo is correctly calculated and the periodicity of the alternating current is sufficiently high an arc light can be formed between the two poles 1 and 2. As the cathode 2 is connected with the negatively charged earth, and therefore always acts as a negative pole, a form of rectification of the alternating current produced by the dynamo 3 is obtained, the second half of the period is always suppressed. The working circuit may be carried out in the same way as in Figure 34; the working gap 7 may however be dispensed with, and instead thereof between the points m and n a condenser 5 and an induction resistance 9 may be inserted from which the current is taken inductively.

Figure 36 represents a form of construction similar to Figure 34 only that here instead of an arc lamp a Nernst incandescent body is employed. The Nernst lamp is fed through the battery 3. The working section is connected with the negative pole, the safety spark gap with the + poles. The working spark gap 7 may also be dispensed with and the current for it taken at 12 over the oscillation circuit 5, 11 (shown in dotted lines).

Flame collectors (Figure 37) may also be employed according to this invention. The wire network 1 is connected with the aerial collector conductor A and the burner with the earth. At the upper end of the latter, long points are provided which project into the flame. The positive electrode is connected with the negative over a condenser 5 and the induction coil 9 with the earth.

The novelty in this invention is firstly, the use of incandescent cathodes opposite positive poles which are connected with large metallic capacities as automatic collecting surfaces, (2) the connection of the incandescent cathodes with the earth whereby, in addition to the electricity conveyed to them from the battery or machine which causes the incandescing, also the negative charge of the earth potential is conveyed, and (3) the connection of the positive and negative poles of the radiating collectors over a condenser circuit alone or with the introduction of a suitable inductive resistance, whereby simultaneously an oscillatory oscillation circuit may be obtained. The collecting effect is by these methods quite considerably increased.

I declare that what I claim is: --- [Claims not included here]

British Patent # 157,262

(10 July 1922)

Improvements in Electric Motors

Hermann Plauson // Otto Traun’s Forchungs-Laboratorium GmbH

This invention relates to that type of motor in which rotation is produced by means of the attraction and repulsion of surfaces carrying charges of electricity.

According to this invention a stator and rotor are formed of condensor surfaces and charges of electricity thereon imposed in the form of alternating currents of high frequency.

The invention is more particularly described with reference to the accompanying drawing in which: ---

Figure 1 shows a simple form of motor and feed.

Fig.1

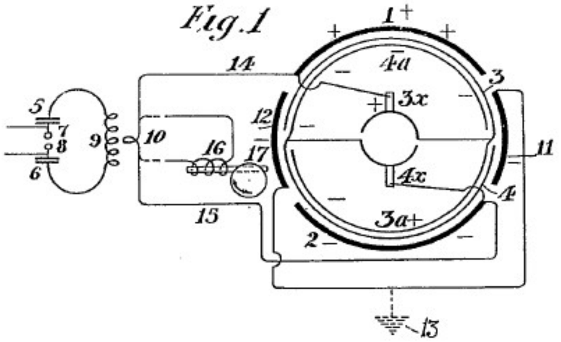
1+
14
4a
3
5
7
8
6
9
10
12
16
17
3x
11
4x
4
15
3a+
2
13

Figure 2 is a modification of Figure 1.

Fig.2.

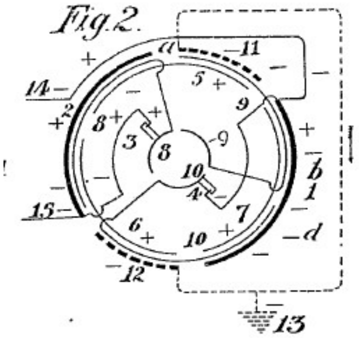
a
11
14
5
9
2
8
3
8
9
b
1
10
4
7
13
6
10
d
12
13

Figure 3 shows one form of a spiral condenser surface.

Figure 4 shows a wire wound condenser surface.

Figure 5 is a diagram of one type of rotor,

Fig.3

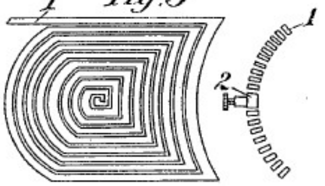

Fig.4

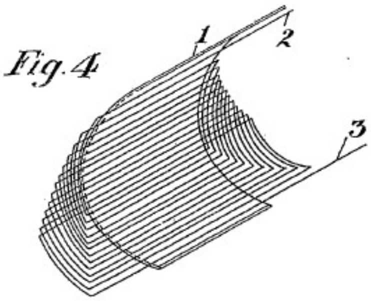

Fig.5.

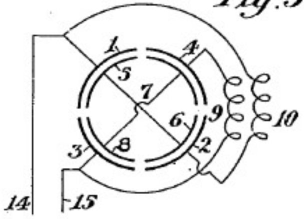
1
4
5
7
6
9
10
3
8
2
14
15

The inner plates of the condenser 5 and 6 are charged from a spark gap 7, 8 connected to a source of energy of sufficiently high pressure (alternating or direct current), until the potential has risen so far that a spark springs over.

The spark gap 7, 8 forms with the condenser 5 and self-inductance 9 and condenser 6 a closed oscillatory circuit and alternating currents of high frequency will be produced in this circuit. The high frequency current produced in the primary circuit 9 excite by induction in the secondary circuit 10 currents of the same periodicity.

The improved type of motor is fed by the discharges produced by the induction in the secondary circuit.

Hitherto only Tesla's motor system (shown diagrammatically in Figure 1, 16 and 17) was known for this purpose. The above-mentioned diagram is only shown for illustrating the fundamental principle. It has however no practical interest for carrying out large machines by reason of the impossibility of the regulation and the low efficiency.

Now according to this process, all these defects are overcome by the construction of a machine which is applicable for high frequency currents and of a more or less undamped nature. The difference between the principle of construction of these motors as compared with those hitherto customary consists in that the motor is not based on the principle of magnetic induction only (as have been all motors hitherto and also Tesla's motors).

It has been fund that the machine constructed according to Figure 1 cannot only be fed directly with static electricity but if it is connected to a source of high frequency current it will operate.

The applicants call this new type of motors 'condenser motors' to differentiate them from hitherto existing types.

The simplest form of construction of such condenser motors is shown in Figure 1, and this motor may be fed with high frequency alternating currents.

At a given moment positive electricity is charged by means of the lead 14 to the stator surface 1 and to the brush 3x (Figure 1). The brush 3x is connected with the rotor condenser surface 3, so that both the stator surface 1 and also the rotor surface 3 is charged with positive electricity. The stator surface 1 and the rotor surface 3 being both charged with positive electricity and the second rotor surfaces 4 and 4a by brushes 4x with negative electricity, such motors can then be started by providing intermediate stator surfaces 11, 12, the earth connection 13 of one of which is broken by a switch (not shown) according to the direction of rotation desires, or alternatively the motor may be started by a separate source of alternating current in a manner similar to the starting of synchronous motors of known construction. After a half revolution of the rotor the brush 3x comes in contact with the second collector surface 4 so that now this surface is connected by the brush 3x with the stator surface 1 and the brush 4x with the collector surface 3. Consequently with a reverse direction of current through the second half of the oscillation period all the hereinbefore mentioned effects take place in the reverse direction which, however, produces no alteration in the direction of rotation because the dead points between two directions of oscillation are overcome by inertia.

Although this motor is easy to start it can only be employed for small experimental and measuring purposes because the stator and rotor surfaces are made of solid metal and are heated by Foucalt (eddy) currents. In spite of its simplicity and its unsuitability for use in practice it must however be regarded as a basic type for technical calculations.

The condenser motor shown in Figure 2 differs from Figure 1 by the rotor surfaces consisting of six condenser surfaces connected one behind the other in series and they are connected with three collector surfaces, so that at any one moment only two adjacent collector surfaces come under the two brushes (3 and 4). In its other actions it corresponds to Figure 1. The leads 14 and 15 may be connected either to the ends of the secondary coil 10 or directly with the source of energy. The outer thicker line indicates the stator surfaces 1 and 2 (that is to say the unmoving part of the motor), 11 and 12 shown by thick dotted lines means earthed additional poles of the stator, 8, 9, and 10 are the outer parts of the rotor condenser surfaces which in turn are connected with the collector surfaces 8, 9, and 10. 5, 6, and 7 are the inner parts of the condenser surfaces of the rotor and 3 and 4 are brushes.

Hitherto stator and rotor surfaces of compact metal have been spoken of. These however become highly heated with eddy currents and hardly yield 10-15% of useful effect. In examining into such small useful effects it was found that certain forms of metal sections in the stator and condenser surfaces highly increase these. It was then further found that if slots or notches be cut in the metal surfaces of the stator and rotor in the form of a spiral, not only was a higher useful effect possible, but also an easier starting and even a regulation could be obtained.

Experiments have shown that by such a form of construction it is possible to build a very useful motor for high frequency alternating currents more particularly those of an undamped nature.

If for example the system of construction of a stator shown in Figure 1, but four polar, be taken and the system of rotor construction shown diagrammatically in Figure 5, but with the form of construction of the condensers of the stator as well as of the rotor according to Figure 3, a condenser motor is obtained which works well in either direction for high frequency alternating current. It was also observed that the motors in such forms of construction were found to be more sensitive to resonance effects. Such a motor then works the best if stator and rotor surfaces have equal capacity and self-inductance so that the windings both in the stator and also in the rotor are in resonance.

A motor constructed according to the foregoing kind is already fully technically applicable. But even these motors have a series of faults, more particularly in their building construction. For example, the attachment of the spiral condenser surfaces both of the stator and of the rotor sown in Figure 3 are in practice difficult to carry out. Therefore in practice the condenser and stator surfaces are simply wound of wire or bands in the form shown in Figure 4. Such stator and rotor surfaces may, without further difficulty, be regarded as electromagnetic poles, although they are not made of iron as is the case in electromagnets. Such machines may be spoken of directly as motors for high frequency alternating currents in which the separate pole surfaces consists of wound induction condenser surfaces of which one is sound on the stator and the other on the rotor.

If the coil as shown in Figure 4 be made of well insulated wires the coil can be embedded in insulating material either for the stator or motor surfaces as has already been done in the case of ordinary single and multiphase motors. At the same time the possibility is afforded by increasing the number of turns to produce a greater or smaller alteration of the self induction co-efficients.

In Figure 5 is shown a modified construction of a rotor for a four pole motor consisting of four condenser surfaces 1, 2, 3, 4, of which 1 and 2 are connected though an iductance 9 coupled with the coil 10. Four inner surfaces 5, 6, 7, 8 are provided of which 5 and 6 are directly connected also 7 and 8 similarly connected.

The pairs of like poles are connected by wires 14 and 15 to the source of energy. By a suitable selection of the values of the reactance and capacity in these circuits resonance circuits may be formed.

Having now particularly described and ascertain the nature of our said invention and in what manner the same is to be performed, we declare that what we claim is: --- [Claims not included here]

British Patent # 157,263

Process & Apparatus for Converting Static Atmospheric Electrical Energy into Dynamic Electrical Energy of any Suitable High Periodicity

Hermann Plauson // H.O. Traun's Forschungs-Laboratorium GmbH

Static aerial electricity in the form of direct current can be converted by using spark gaps and with the assistance of oscillatory circuits into dynamic electrical wave energy of a high number of alternations of a more or less undamped nature and in such form --- either direct or by means of a special kind of resonance or 'condenser motors' --- ready to be utilized for technical purposes as mechanical energy.

For small installations this system may be very well employed; about 100 horsepower may be stated as practical limit. In constructions of larger aggregate difficulties as regards the spark gaps however increase considerably. Further it is desirable to convert the accumulated currents of from 100 to 1000 periods which may then be used for the ordinary types of alternating current machines instead of into electromagnetic waves of a high number of alternations.

In experimenting with condenser motors the construction of which forms the object of British Patent # 157,262 it was observed that the rotor, if one pole of the stator surface be connected with the aerials collecting aerial electricity and the other pole with the earth, not only could act as a motor, but if vice versa the rotor connection with the stator be interrupted and the rotor caused to rotate by means of another motor, that when the brushes supply an alternating current the periodicity of which is dependent on the number of poles and the revolutions of the rotor. Such an apparatus may therefore be regarded as a transformer of static into dynamic electrical energy.

The invention is more particularly described with reference to the accompanying diagrams in which: ---

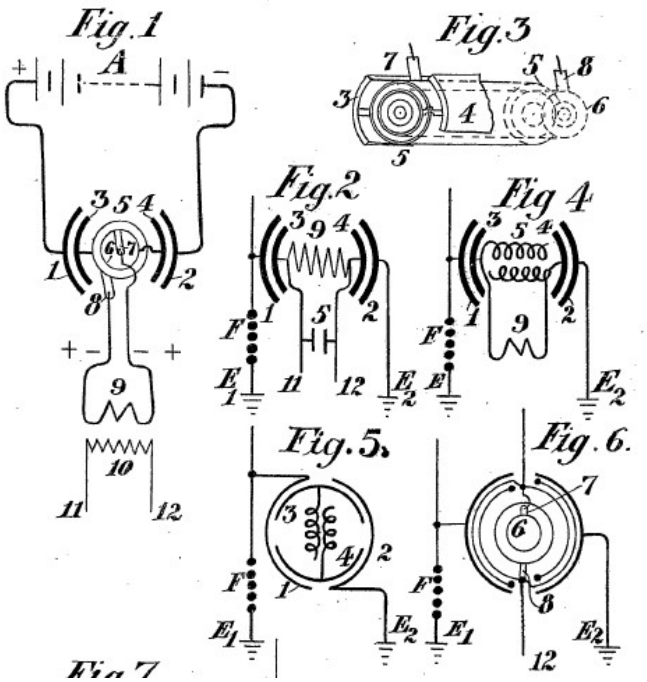

In Figure 1A is a strong accumulator battery, 1 and 2 are the outer poles of the transformer, consisting of simple metallic plates or are as shown in Figures 8-11, made of wire coils without an electromagnet being present. Between these poles an armature is revolubly mounted on a shaft, which armature also consists of two similar cylindrically curved plates 3 and 4. These are metallically connected with two collector rings 5 and 6 on which two brushes 7 and 8 freely run which again are short-circuited with one another over a primary coil 9. 10 is the secondary coil with the free ends 11 and 12. If through the accumulator battery the stator plate 1 is charged with positive electricity, it induces a charge of reverse sign on the rotor surface 3 which is connected by the brushes 7 and 8 over the primary coil 9 with the second rotor surface 4. This latter is therefore charged with positive electricity, which in turn induces negative electricity on the stator surface 2/ Up to this moment everything takes place in the same way as if two condensers were connected one behind the other in the current circuit A. If however, by means of mechanical power, this rotor be caused to rotate, the surface conditions are altered. After a quarter revolution the rotor plates are between the stator plates and therefore no condenser surface faces another. By this means however, the capacity of the entire system is reduced to a minimum and a change of current will also result in the main 9. Now if the rotor be turned further through 90 degrees by mechanical energy the rotor plate 3 comes opposite the stator plate 2 and the rotor plate 4 opposite the stator plate 1, so that then the rotor pates are in a field of reverse sign. A fresh charge of current in the reverse direction now runs through the primary coil 9. After a further half revolution the same action is repeated so that after a full revolution the initial condition is again produced. The result of such a revolution is an alternating current the periodicity of which is equal to the number of revolution. In practice of course not two poles but as many poles as possible would be employed because thereby the number of alternations would e considerably increased. The primary alternating current thus obtained induces in the secondary circuit an alternating current the potential of which is dependent on the winding of the coil. Figure 7 shows a multipolar machine.

If the stator surface 1, instead of being connected with the battery be connected with a collecting aerial network and the other stator surface 2 be directly earthed, but the rotor wich is otherwise constructed as hereinbefore, be rotated by a separate motor a much stronger alternating current results which is to be ascribed to the circumstance that a much higher potential can be charged on the pole surfaces of the stator by reason of the higher pressure of the static electricity than where accumulators are employed. By this means the transformer has of course much larger quantities of energy supplied to it.

Figure 2 shows a mode of connections. The stator surface 1 is connected with the aerial antennae which is connected through the safety spark gap F to earth at E1. The stator surface 2 is directly earthed at E2. The inner revoluble rotor surfaces 3 and 4 are interconnected by means of an induction coil which is constructed directly in the motor. The current is taken as in Figure 1 up to collector rings by means of brushes, which are not shown for the sake of clearness, and further conveyed through the conductors 11 and 12. Between these a condenser 5 may be inserted. There is thereby formed a short oscillatory oscillation circuit free from spark gaps, which circuit consists of the induction coil 9 and condenser 5 and is fed by the periodic charging current impulse. By this means the possibility is afforded of obtaining a kind of current which is characterized by longer periods and is undamped and oscillatory. Of course a simple alternating current may be obtained by cutting out the condenser.

Instead of the induction coil the condenser may also b constructed in the rotor. His can be carried out in such a way that its ends serve directly as collector rings for taking current through the brushes. In Figure 3 such a motor is sketched in perspective, 3 and 4 are the rotor surfaces, 5 and 6 are the condenser surfaces constructed to form part of the rotor consisting of two co-axial cylinders fitting one in the other in such a way that free room is left for the brushes 7 on one end of the condenser cylinder 6.

The condenser may be made in the form of a cylindrically wound spiral forming the capacity and reactance as shown in Figure 4. A further type of transformer is shown in Figure 5. The difference consists in the stator and rotor surfaces not only each assuming a quarter of the circuit but almost the half. By this means the space and the effective condenser surface is better utilized. Charge is produced only when the rotor surfaces face the full scope of the stator surfaces.

In addition a condition is obtained in which the stator surfaces are inductively connected by the rotor surfaces. The consequence of this is that an alternating current simultaneously results which is produced without sparking otherwise the connection is as before.

Figure 6 shows the alternation of the rotor surfaces; the rotor here consists of two cylindrical condenser plates arranged concentrically, each divided into two halves and connected so that half the inner cylinder is connected to half the outer. Such a machine shows the more complete transformer action.

Fig.7.

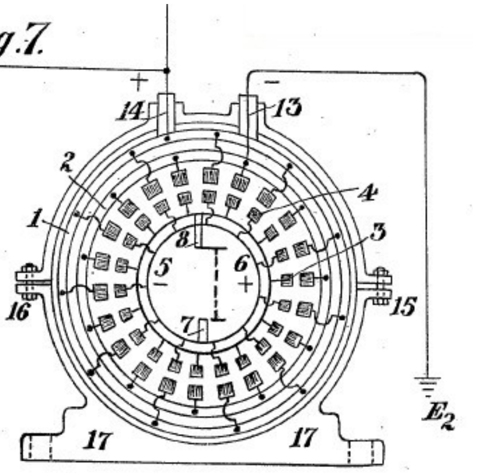

Figure 7 shows a four polar transformer. It consists of a metal casing, the lower half of which is fastened with the foundation plates 17 to the support or foundation. The upper half, the cover, is connected by bolts 15 and 16 firmly with the under part. This upper sleeve or casing is insulated from the under part. Two rings 1 and 2 are cylindrically constructed in the casing. The ring 1 is metallically connected with the collector aerial and the ring 2 with the earth. On both rings an equal number of stator surfaces are mounted side by side but well insulated from one another and thus form an electrostatic field similar to the electromagnetic in many alternating current machines. The rotor consists in similar manner of two rings 5 and 6 on which an equal number of rotor surfaces are fixed so that each stator surface faces a rotor surface. By the brushes 7 and 8 the alternating current formed is removed from the collector. The charge is conveyed by the conductor 14 to and by 13 away. If this rotor be then rotated by means of a motor the positive and negative fields precisely as in the case of magnetizing will alter and thereby an alternating current is formed in the rotor, the periods of which are dependent on the number of the poles and the revolutions per second.

At the commencement it was thought that this apparatus could only be regarded as alternating current converters, but it was soon found that much more energy was necessary to rotate the rotor than might be necessary to overcome the friction. It was then found that the considerable expenditure of energy for rotating the rotor was caused by a conductor being moved through strongly electrostatic fields since the electrostatic lines of force must be cut at right angles and that further in the conductors a stronger current arose than was otherwise to be expected. This apparatus must therefore not only be regarded as a transformer, but also as an energy producer, with the difference that the excitation here is obtained instead of by means of electromagnets, by static fields of high pressure. The entire system may, to some extent, be compared with a dynamo in which the excitation takes place by means of a fixed constant magnet. It was further ascertained that this way of using the atmospheric electricity produced a sort of suction on the collector network, and that thus suitable greater quantities of current could be obtained.

The effects which in this apparatus became evident are extremely interesting and open a prospect of being able to obtain here a great deal more. Merely that these transformer made it possible to transform suitable quantities of atmospheric electricity into alternating current of high or lower frequency (without the use of spark gaps) shows already the extreme utility of these apparatus. Should in future, the construction of larger aggregates be necessary the transformer installation may be constructed in such a way that motors which are fed by a current obtained from an installation with spark gaps produce a certain quantity of energy which may then be employed for producing current according to the last described system.

The results of the examinations made for this may be construed as follows.

(1) If solid electrodes (condenser surfaces, rotor and stator surfaces) are employed they become hot. This effect may be considerably reduced by cutting the electrodes in ribbed form Figure 8, but not entirely removed. This form allows the surface of the condenser plate to be enlarged or increased; the electrodes may be fastened in a simple manner on the under frame by perforation 1, 2, 3, 4, 5.

Fig.8

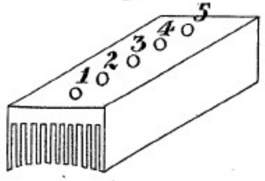

Fig. 9.

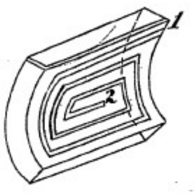

(2) If nicks or notches in spiral form as shown in Figure 9 seen from the side end in Figure 10 in section are employed, not only is the transformer effect greater but the poles yields also more current, but require greater quantities of energy for their movement than a simple commutator action would require.

Fig. 10

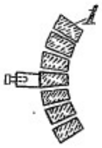

Fig. 11

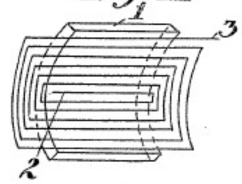

(3) The greatest effect is obtained if the rotor and stator surfaces are wound in flat spiral form of suitably thick wire, and in such a way that the inductive effect combining with the capacity is calculated in suitable proportion and this result is adapted to a suitable periodicity. In practice this is preferably done by the wire bent in spiral form being inserted in a separate vulcanite or hard rubber mass (see Figure 11) so that a smooth pole surface is formed similar to that in phase motors.

Fig.12

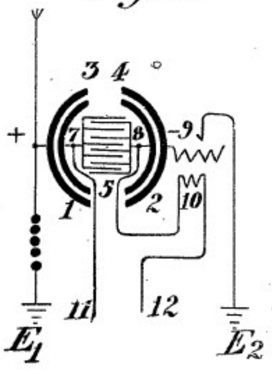

Fig. 13.

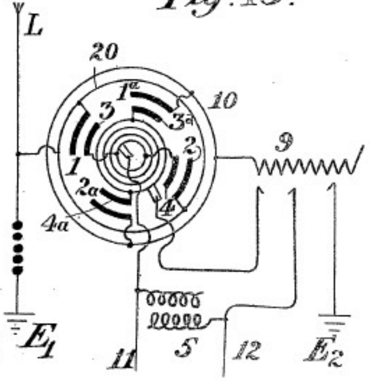

Regular undamped oscillations of a high frequency may however be produced if the converter be carried out in the manner shown in Figure 13. The aerial wire L is metallically connected with the ring 20. To this two pole surfaces 1 and 2 are connected. The inductive earth pole is also connected with a second ring 10 from which again two, poles 1a and 2a are branched off. Of course in similar manner any suitable number of poles may be branched of. In similar manner there are in the rotor two poles fastened to one another (3 and 4 and 3a and 4a) connected with separate collector rings. From these two rings the current is collected by means of two brushes. The induced alternating current is however directly metallically connected with an inductive earth stator conductor over an induction coil 9. Further a combined inductance and capacity 5 is inserted between the two wires 11 and 12 in parallel with the converter. By this means a sparkless oscillatory circuit is obtained which can act on the exciting current in the stator. This produces however, a periodic alteration of the charging quantities according to the oscillation curves of the rotor currents in consequence of which the stator charge also commences with resonance oscillations and if the stator and rotor surfaces are calculated to one another in such a way that they are adapted to form oscillations of waves of similar length the entire converter is caused to oscillate and furnishes undamped oscillations of a high number of alternations, but of periodically changed amplitude, the form of which is dependent on the amplitude of the main alternating current and is caused by the number of the poles and revolutions per second. Thus an alternating current of, for example, 100 periods is formed, the separate periods of which are formed by undamped oscillations of a higher number of alternations. In Figures 14-16 four other diagrams of converters are illustrated, the object of which is not to produce usual alternating current, but oscillations of high frequency.

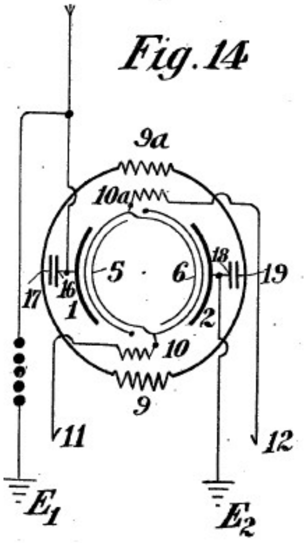
Fig. 14

The main difference of these systems from those previously described is that from the connection of the collecting aerials is made between the stator pole 1 (Figure 14) and one pole 16 of the condenser 17 and the earth connection between a second stator pole 2 and the pole 18 of the condenser 19. The other poles of these condensers 17 and 19 are short circuited through a ring over two inductive primary coils 9 and 9a with one another. The secondary coils form the rotor conductors 10 and 10a. The rotor itself is constructed in the manner shown in Figure 6 of two short-circuited plate condensers which may be wound as shown in Figure 11. In similar manner of course the stator surfaces may also be formed. The collector rings of the rotor with the two brushes for collecting current are here not shown in order to simplify the drawings. By the connection of the two condensers in the exiting circuit of the converter and also by the action of the alternating current produced in the rotor on the stator circuit, with a correct calculation of the capacity and the self-induction co-efficients a maximum action may be obtained. The kind of current produced will be similar to that described for Figure 12.

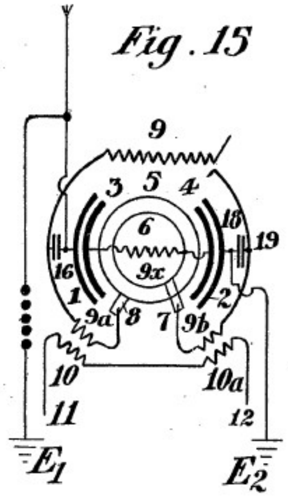
Fig . 15

The novelty of the converter illustrated in Figure 15 consists mainly in that the current resulting in the rotor is not directly employed, but only serves a exciter of the primary coils 9a and 9b. The working current is produced in the secondary coils 10 and 10a and further conducted through the conductors 11 and 12. The stator current may be brought by the regulatable inductive resistance 9 to the same resonance as the rotor current.

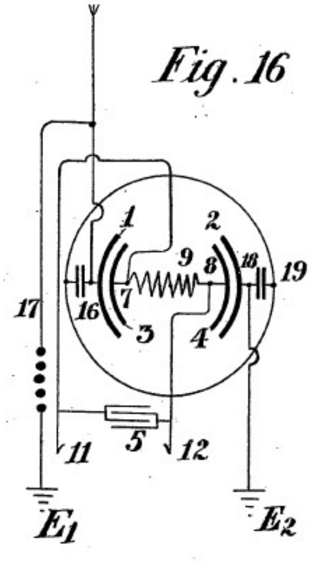

In Figure 16 a very similar system is shown to Figure 14. The condenser 5 is however connected in parallel with the converter; and by the inductive resistance constructed in the rotor a short circuited oscillatory circuit is formed which gives extraordinarily good results and is simple in construction.

The inductive resistance 9 may also instead of being constructed in the rotor be constructed as primary coil employed outside the rotor and short circuit the oscillatory circuit over the stator surfaces (see Figure 17).

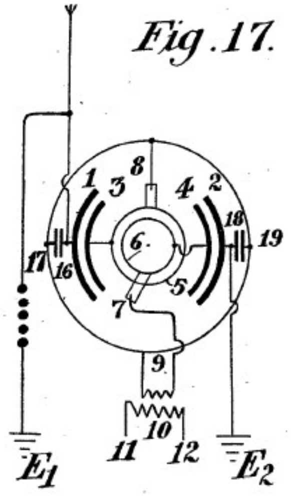

The last six types serve only for producing oscillations of a high number of alternations. If it be desired to obtain ordinary alternating current there complicated constructional arrangements are not required as the types illustrated in Figures 1 to 11 suffice. It is self-evident that these arrangements may be altered in various ways by means of different condenser surfaces in practice.
Having now particularly described and ascertained the nature of our said invention and in what manner the same is to be performed, we declare that what we claim is: --- [Claims not included here]

H. Plauson : *Gewinnung und Verwertung der Atmosphatischen Elecktrizitat* (1922)
[PDF]

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Thumb drive or Download !

ORDER PAGE

---

advertisement

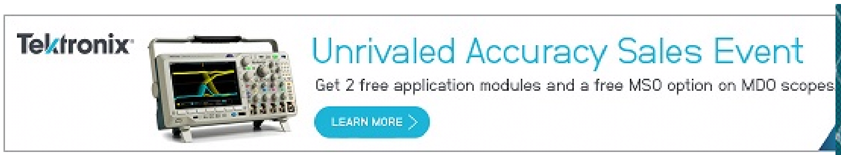
Tektronix
Unrivaled Accuracy Sales Event
Get 2 free application modules and a free MSO option on MDO scopes
LEARN MORE

AdChoices
Yearly subscription only $69 ~~$199~~
International weekly journal of science
SUBSCRIBE NOW
nature

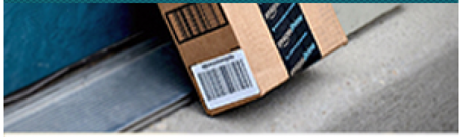
This little-known trick can save you money on Amazon

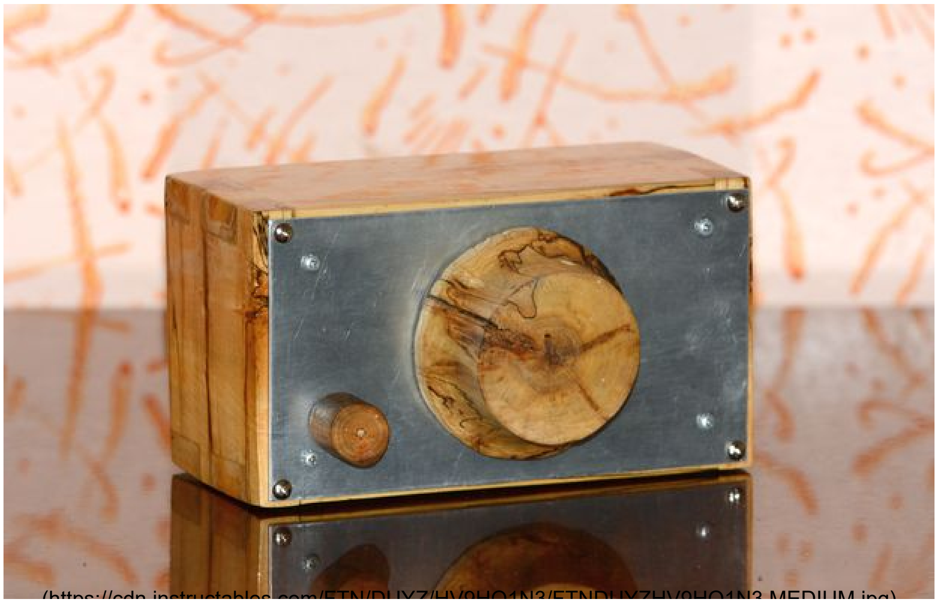
(https://cdn.instructables.com/FTN/DUYZ/HV9HO1N3/FTNDUYZHV9HO1N3.MEDIUM.jpg)

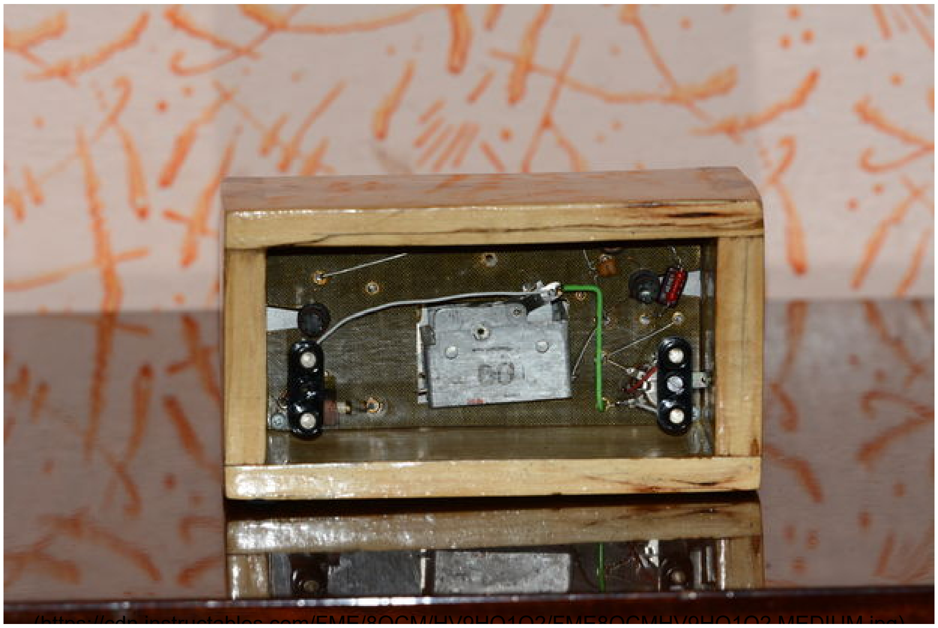
(https://cdn.instructables.com/FME/8OCM/HV9HO1O3/FME8OCMHV9HO1O3.MEDIUM.jpg)

Hello everyone!

About This Instructable

28,204 views
271 favorites

License: (cc) BY-NC-SA

DiggingFox (/member/DiggingFox/)
Follow 16
(/member/DiggingFox/)
Bio: I love simplicity, function and precision job!

I am big fan of old radios, and I am fascinating how precision were the radios making. 4 months ago, I was just looking for some interesting schematcs, and I found one! It's about 50 years old, but works perfectly, without any power supply. With the reciever you can listen to your favourite radio on middle waves. You just need some electronic components, long antenna, good ground and HI-Z earphones (or amplifier). To make it you don't need any money! All components I used were recycled! Please, respect the nature!!

Let's make it! :))

AdChoices
Yearly subscription only $69 ~~$199~~
nature
SEA CHANGES
International weekly journal of science
SUBSCRIBE NOW
nature

Step 1: Components

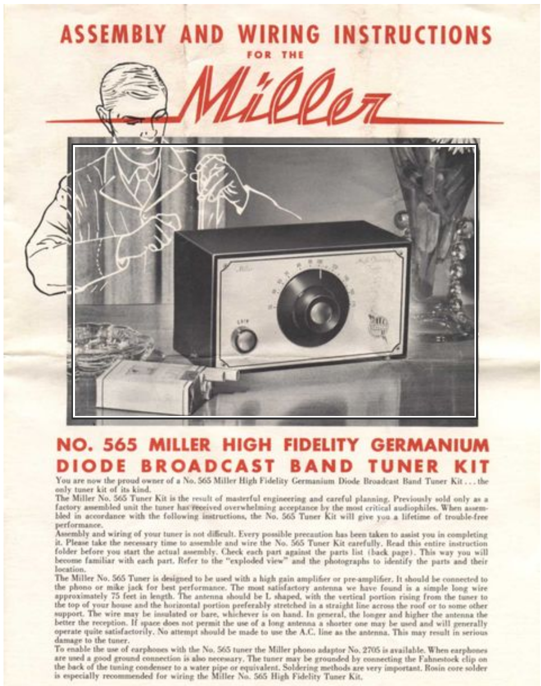
ASSEMBLY AND WIRING INSTRUCTIONS FOR THE Miller

NO. 565 MILLER HIGH FIDELITY GERMANIUM DIODE BROADCAST BAND TUNER KIT

You are now the proud owner of a No. 565 Miller High Fidelity Germanium Diode Broadcast Band Tuner Kit . . . the only tuner kit of its kind.

The Miller No. 565 Tuner Kit is the result of masterful engineering and careful planning. Previously sold only as a factory assembled unit the tuner has received overwhelming acceptance by the most critical audiophiles. When assembled in accordance with the following instructions, the No. 565 Tuner Kit will give you a lifetime of trouble-free performance.

Assembly and wiring of your tuner is not difficult. Every possible precaution has been taken to assist you in completing it. Please take the necessary time to assemble and wire the No. 565 Tuner Kit carefully. Read this entire instruction folder before you start the actual assembly. Check each part against the parts list (back page). This way you will become familiar with each part. Refer to the "exploded view" and the photographs to identify the parts and their location.

The Miller No. 565 Tuner is designed to be used with a high gain amplifier or pre-amplifier. It should be connected to the phono or mike jack for best performance. The most satisfactory antenna we have found is a simple long wire approximately 75 feet in length. The antenna should be L shaped, with the vertical portion rising from the tuner to the top of your house and the horizontal portion preferably stretched in a straight line across the roof or to some other support. The wire may be insulated or bare, whichever is on hand. In general, the longer and higher the antenna the better the reception. If space does not permit the use of a long antenna a shorter one may be used and will generally operate quite satisfactorily. No attempt should be made to use the A.C. line as the antenna. This may result in serious damage to the tuner.

To enable the use of earphones with the No. 565 tuner the Miller phono adaptor No. 2705 is available. When earphones are used a good ground connection is also necessary. The tuner may be grounded by connecting the Fahnestock clip on the back of the tuning condenser to a water pipe or equivalent. Soldering methods are very important. Rosin core solder is especially recommended for wiring the Miller No. 565 High Fidelity Tuner Kit.

(https://cdn.instructables.com/FB8/L1CY/HU842FKH/FB8

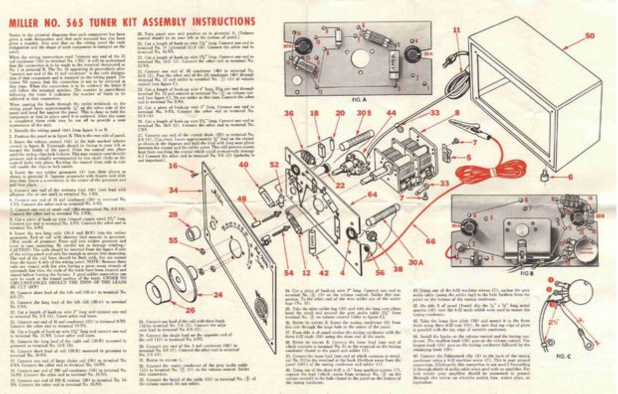
MILLER NO. 565 TUNER KIT ASSEMBLY INSTRUCTIONS

(https://cdn.instructables.com/FVB/UKAZ/HU842FKI/FVBUKAZHI

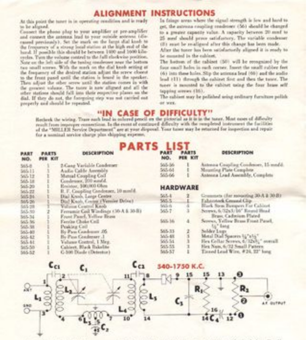
ALIGNMENT INSTRUCTIONS

"IN CASE OF DIFFICULTY"

PARTS LIST

J. W. MILLER COMPANY
5917 So. Main St., Los Angeles 3, California

(https://cdn.instructables.com/FR

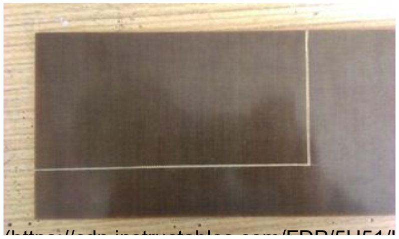
(https://cdn.instructables.com/FDB/EH51/I

(https://cdn.instructables.com/F03/Y68M/

(https://cdn.ins

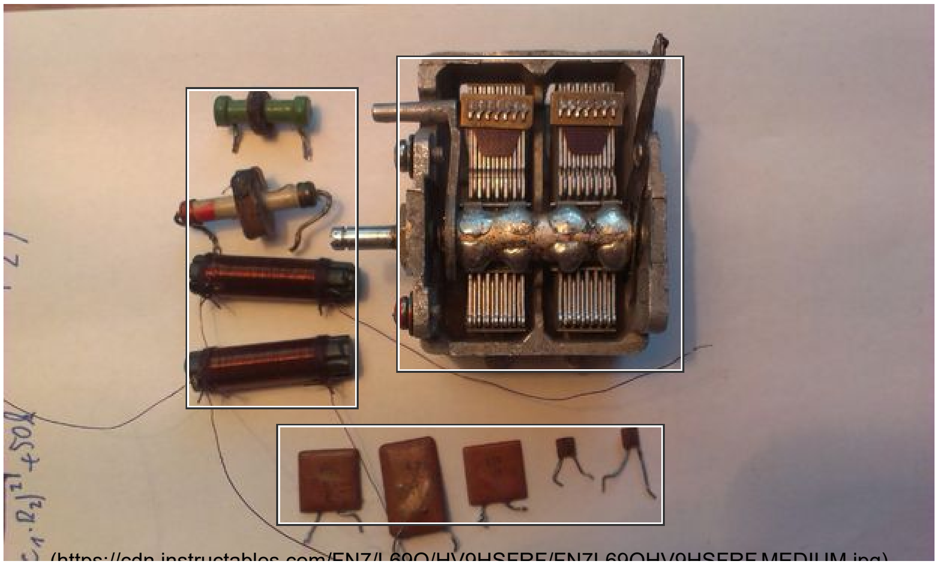

(https://cdn.instructables.com/FN7/L69O/HV9HSFRE/FN7L69OHV9HSFRE.MEDIUM.jpg)

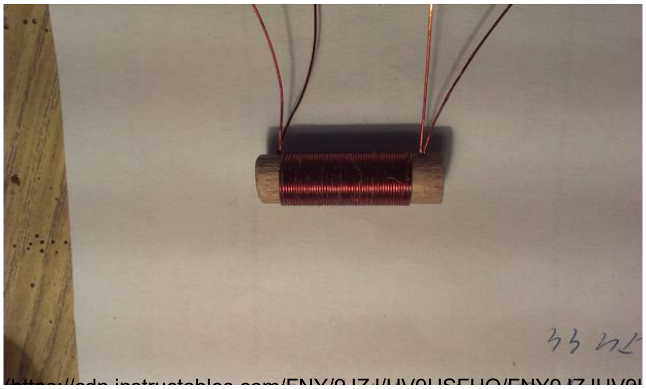

(https://cdn.instructables.com/FNY/0JZJ/HV9HSFHQ/FNY0JZJHV9H

(https://cdn.instructables.com/

AdChoices
Yearly subscription only $69
$199
nature
SEA CHANGES
International weekly journal of science
SUBSCRIBE NOW
nature

What you need:

- Epoxy board, or any non-conducting board (for motherboard)
- Metal board (for frontboard, I chose dural)
- Old ugly log (for case)

Electronic parts:

Resistors:

- R1 - 100 kOhms
- R2 - 1 MOhms potentiometer

Capacitators:

- C1 - Tuning capacitator 2 * 500 pF (from an old AM radio)
- C2 - 100 nF
- C3 - 200 pF
- C4 - 50 nF
- Cc1 - 15 pF
- Cc2 - 10 pF

Coils:

- L1 - 250 uH
- L2&L4 - 330 uH (100 turns with 0,21 mm wire on 7 mm in diameter and 29 mm lenght ferrite stick)
- L3 - 20 - 30 double turns on 8 mm in diameter 40 mm lenght wooden dowel, then solder left end of A turning and right end of B turning together
- L5 - 2,5 mH

One germanium diode

Note: Be precision while you are winding the coils. After winding is good to measure the inductive of coils. The inductive has to be the same. It is important for future quality receive.

Other:

- Some wires
- Some banana jacks (for antenna, groung, earphones..)
- Long (best 20 meters) antenna
- Good ground (metal heating is fine)
- High Z earphones (best about 1-2 kOhm or more)
- A lot of chocolate (for patience :))

AdChoices
Yearly subscription only $69 ~~$199~~
nature
SEA CHANGES
International weekly journal of science
SUBSCRIBE NOW
nature

Step 2: Motherboard

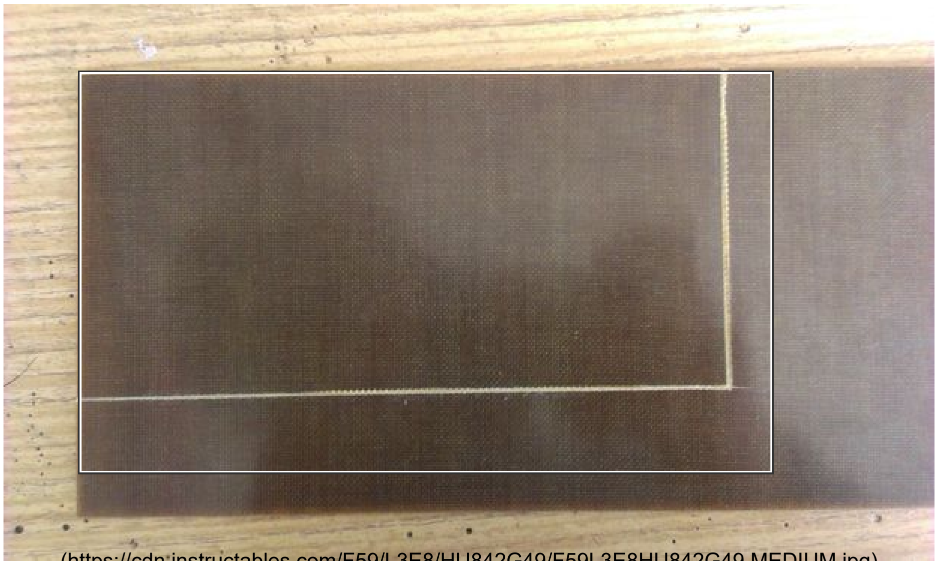
(https://cdn.instructables.com/F59/L3F8/HU842G49/F59L3F8HU842G49.MEDIUM.jpg)

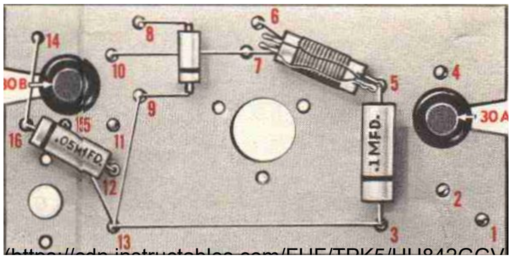

(https://cdn.instructables.com/FHE/TPK5/HU842CGV/

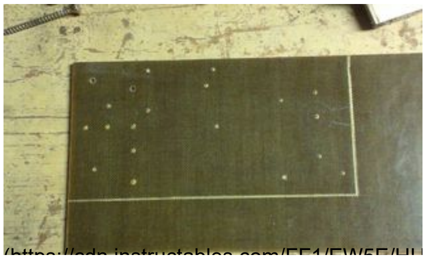
(https://cdn.instructables.com/FE1/EW5F/HU

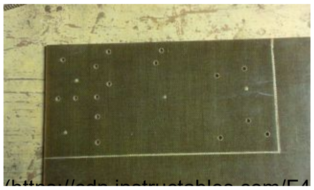
(https://cdn.instructables.com/F4

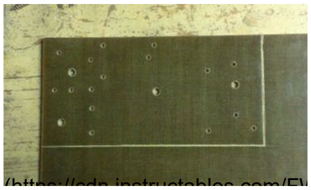
(https://cdn.instructables.com/F\

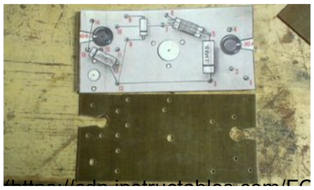
(https://cdn.instructables.com/FC

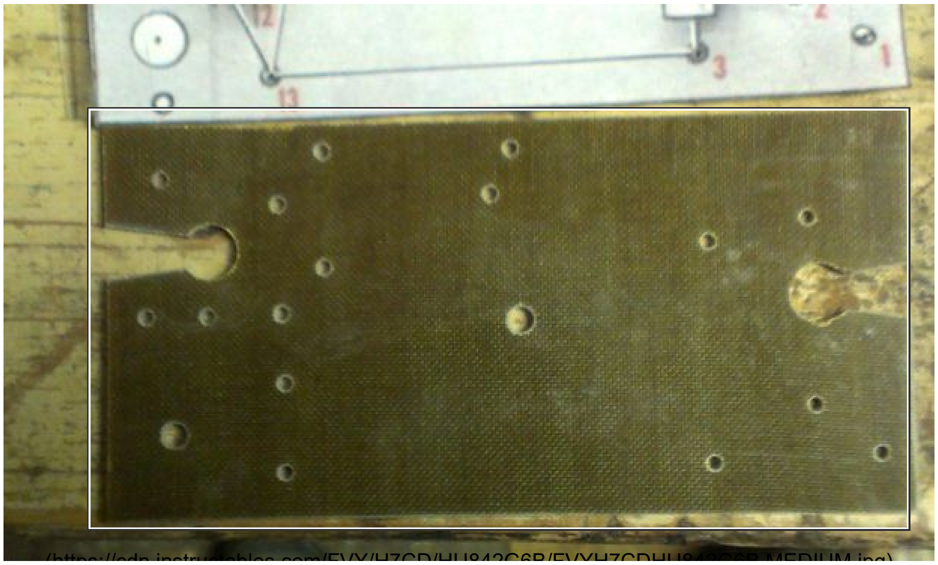
(https://cdn.instructables.com/FVY/H7CD/HU842G6B/FVYH7CDHU842G6B.MEDIUM.jpg)

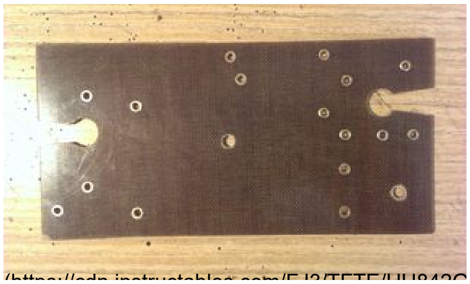
(https://cdn.instructables.com/FJ3/TFTF/HU842C

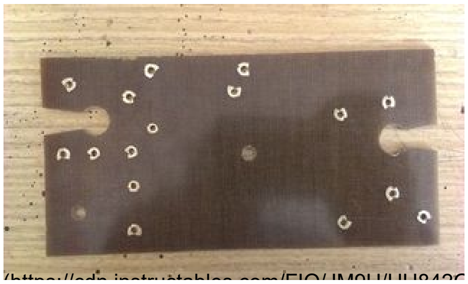
(https://cdn.instructables.com/FIO/JM9H/HU842C

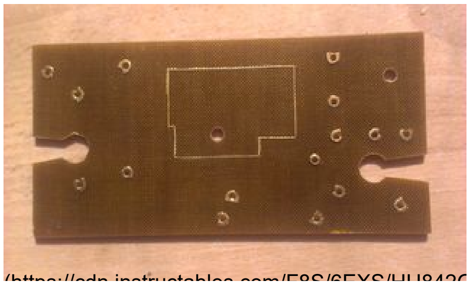
(https://cdn.instructables.com/F8S/6FXS/HU842C

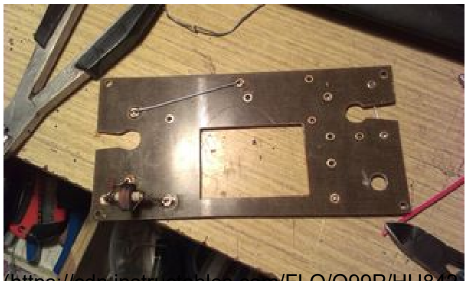
(https://cdn.instructables.com/FLQ/O99B/HU842

To make a motherboard for the reciever you need comprextit board (or any thin non-conducting board, wooden, platic...). Dimensions of the motherboard are 144 mm * 72 mm.

1. Line the cuprextit board.
2. Print the layout from original manual on paper.
3. Stick the printed layout on the coprextit board.
4. Mark the soldering holes with center punch.
5. Drill all holes you need,
6. Make holder for the ferrite core coils (L2, L4).
7. Rivet the soldering holes with hollow brass (or any solderable metal) rivets.
8. Make BIG hole for the tuning cap!
9. Be happy of the result, you made the motherboard!! :))

AdChoices
Yearly subscription only $69 ~~$199~~
International weekly journal of science
SUBSCRIBE NOW
nature

Step 3: Frontboard

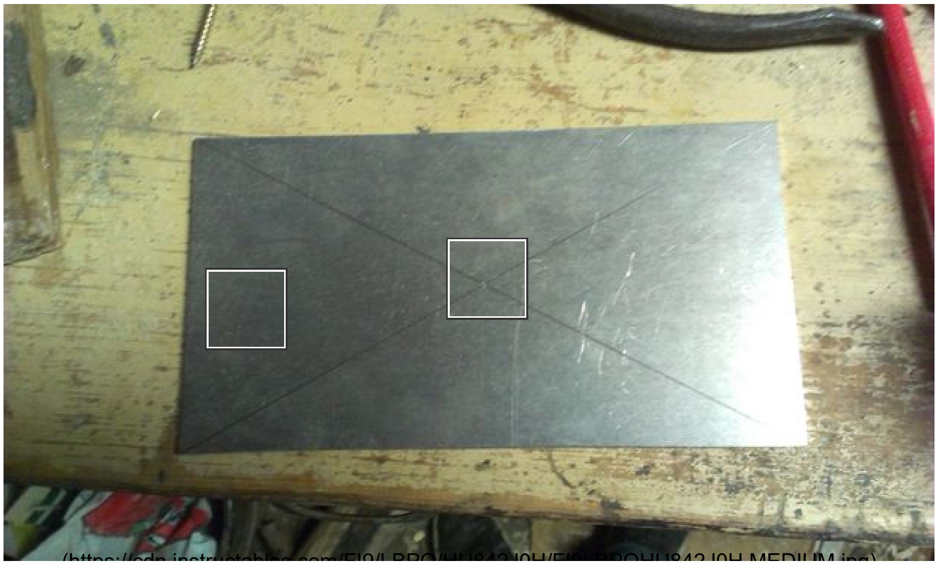
(https://cdn.instructables.com/FI9/LBRO/HU842J0H/FI9LBROHU842J0H.MEDIUM.jpg)

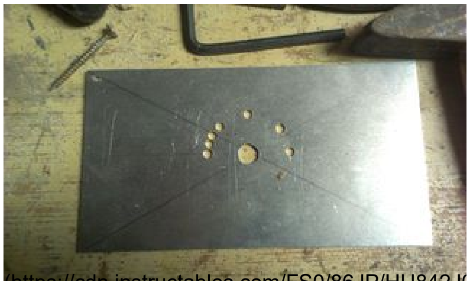
(https://cdn.instructables.com/FS0/86JB/HU842J

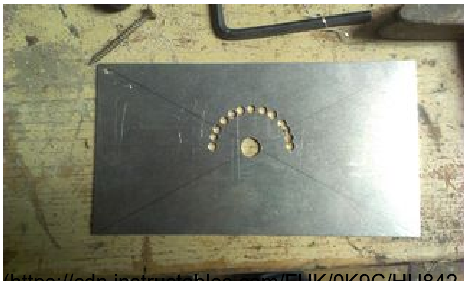
(https://cdn.instructables.com/FHK/0K9C/HU842

To make the frontboard I choosed dural board. It is not necessary, I just had what I had.

1. Frontboard dimensions are 170 mm * 90 mm.
2. Cut it!
3. Mark it in the center and drill hole.
4. Making some interesting things for good working cap (ellipse, holes..).
5. Make hole for potentiometer (by your motherboard).
6. More holes for future case.
7. By the Motherboard make into Frontboard 4 holes for some screws
8. Be proud od yourself! :))

Step 4: The Case

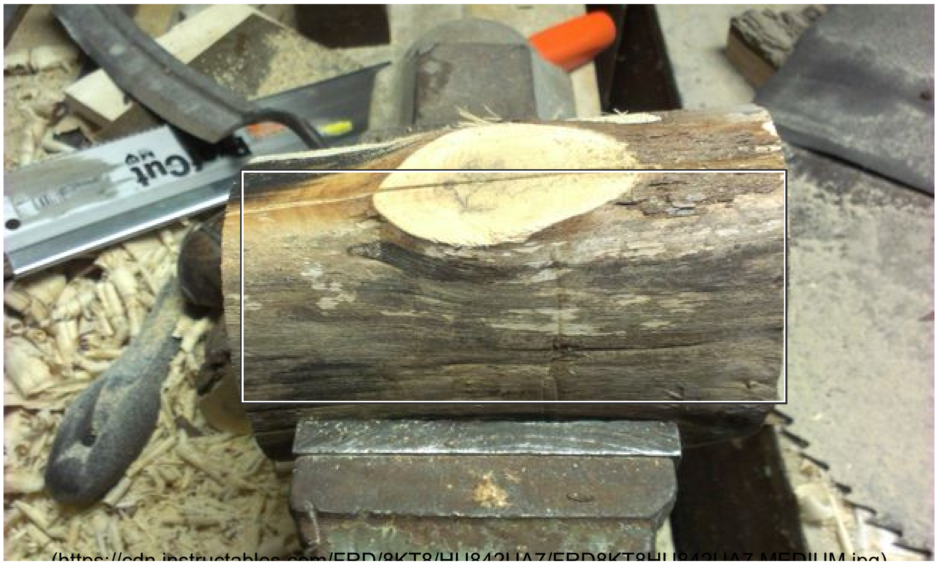

(https://cdn.instructables.com/FBD/8KT8/HU842UA7/FBD8KT8HU842UA7.MEDIUM.jpg)

AdChoices
Yearly subscription
only $69
$199
nature
SEA CHANGES
International weekly journal of science
SUBSCRIBE NOW
nature

AdChoices

Yearly subscription only $69 ~~$199~~

nature

SEA CHANGES

International weekly journal of science

SUBSCRIBE NOW

nature

(https://cdn.instructables.com/FN8/SADK/HU842HFH/FN8SADKHU842HFH.MEDIUM.jpg)

(https://cdn.instructables.com/F0K/VSHV/HU84

(https://cdn.instructables.com/F82/5OZ3/HU8

(https://cdn.instructables

(https://cdn.instructables.com/F

(https://cdn.instructables.com/FI

(https://cdn.instructables.com/FOY/DO3F/HI

AdChoices
Yearly subscription only $69 ~~$199~~
International weekly journal of science
SUBSCRIBE NOW
nature
nature
SEA CHANGES

(https://cdn.instructables.c

(https://cdn.instructables.com/FZB/5NMA/HU843URB/FZB5NMAHU843U

(https://cdn.instructables.com/FRF/O40MA

(https://cdn.ins

(https://cdn.instructables.com/F6D/O7N

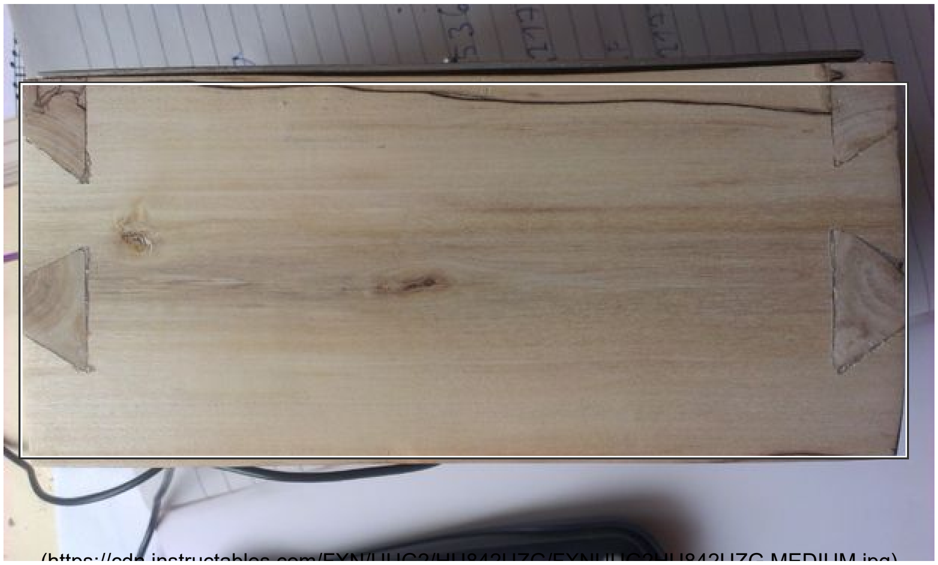
(https://cdn.instructables.com/FXN/UUG3/HU843UZG/FXNUUG3HU843UZG.MEDIUM.jpg)

(https://cdn.instructables.com/F35/KWYZ/HU842L

(https://cdn.instructables.com/FCL/JPRJ/HU842U

(https://cdn.instructables.com/FF

(https://cdn.instructables.com/F7

(https://cdn.instructables.com/FF

AdChoices
Yearly subscription only $69 ~~$199~~
International weekly journal of science
SUBSCRIBE NOW
nature
nature
SEA CHANGES

(https://cdn.instructables.com/FJG/Y725/HUS

(https://cdn.instructables.com/FBW/ID2Y/HU842VQ3/FBWID2YHU842V

(https://cdn.instructables.c

AdChoices

Yearly subscription only $69 ~~$199~~

International weekly journal of science

SUBSCRIBE NOW

nature

nature
SEA CHANGES
SCIENCE IN GERMANY
WASTED LIVES AND MONEY
LONG-DISTANCE INFORMATION

The original case is made from bakelite. I can't make it, so I decided to make it from the best material on the world, from the wood!

1. Find log (I found old, ugly log, to prove, that the ugly, old log with mushrooms init doesn't worth nothing, I didn't wanna to kill tree :((..please, respect the nature:)) BTW. I thing the tree is **Populus alba**.
2. Saw this log on right lenght. Side boards are 104 * 78 * 14 (lenght * width * thickness) in mm. Bottom and upper boards are 174 * 78 * 14 mm (first piece of log is shorter (104 mm), second one is longer (174mm)). (P.S. Be sure, this is aesthetics matters)
3. Saw it on Boards. Line the log with right dimensions from the middle of log. Once you have it, next step is comming!
4. Cut it on the one side, then in the middle, and on the end on the other side. Now you have 2 wooden boards. Repeat it with the second log.
5. Smooth it to soft and clean surface. (don't do it with sandpaper, it's not nice. I did it with drawknife (or jack-plane), and result was perfect)
6. Cut it on the right dimensions (lenght, width).
7. Sketch and saw dovetail joints (25 and 12,5 mm). Be precise, I choosed 2 dovetail joints on one side.
8. Assembly the BOX!
9. Time to puttying. The putty consist saw dust and glue for wood or paper (everybody knows that white liquid). The putty is for the holes in the case. After puttying, let it dray about 4-7 days, for sure, than sand it with sandpaper, but carefully. :)
10. So, we've got a case. Let's lacquer. I choosed "Boad lacquer", it has nice a bit orange color, for this old, ugly, white wood is good. Inside took only one layer. Outside took 4-6 layer of lacquer.
11. Let it dry, and be happy! :))

AdChoices
Yearly subscription only $69 $199
nature
SEA CHANGES
International weekly journal of science
SUBSCRIBE NOW
nature

Step 5: Electronic Part

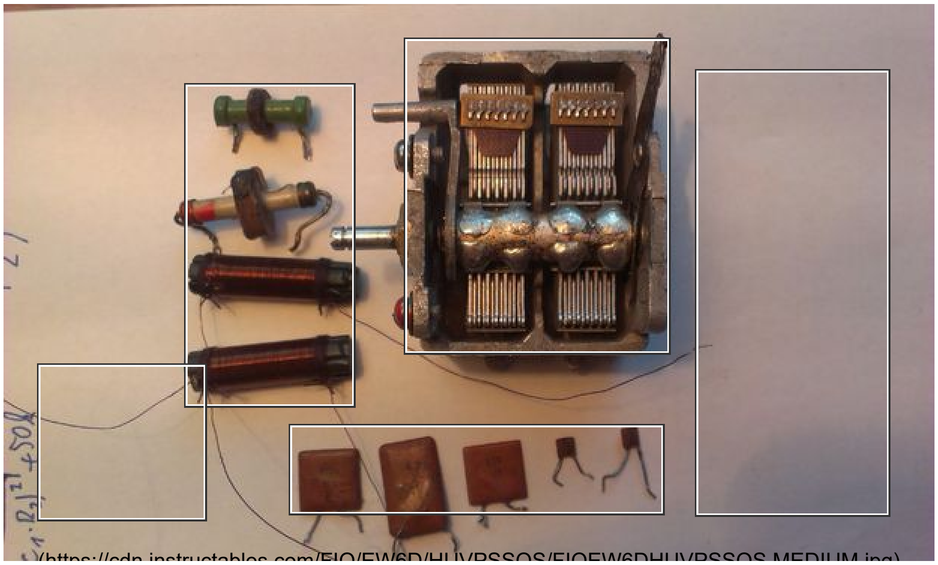
(https://cdn.instructables.com/FIQ/EW6D/HUVPSSOS/FIQEW6DHUVPSSOS.MEDIUM.jpg)

(https://cdn.instructables.com/F13/TC

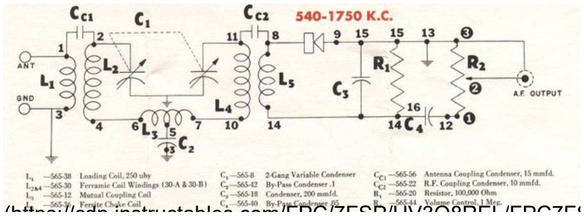
C_{C1}
C_1
C_{C2}
540-1750 K.C.
ANT
GND
L_1
L_2
L_3
L_4
L_5
C_2
C_3
C_4
R_1
R_2
A.F. OUTPUT
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
L_1 —565-38 Loading Coil, 250 uhy
$L_{2\&4}$ —565-30 Ferramic Coil Windings (30-A & 30-B)
L_3 —565-12 Mutual Coupling Coil
L_5 —565-36 Ferrite Choke Coil
C_1 —565-8 2-Gang Variable Condenser
C_2 —565-42 By-Pass Condenser .1
C_3 —565-18 Condenser, 200 mmfd.
C_4 —565-40 By-Pass Condenser .05
C_{C1} —565-56 Antenna Coupling Condenser, 15 mmfd.
C_{C2} —565-22 R.F. Coupling Condenser, 10 mmfd.
R_1 —565-20 Resistor, 100,000 Ohm
R_2 —565-44 Volume Control, 1 Meg.
(https://cdn.instructables.com/FRC/7ESD/IJV3O8BEL/FRC7E

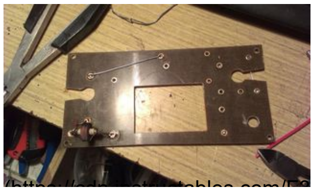
(https://cdn.instructables.com/F3

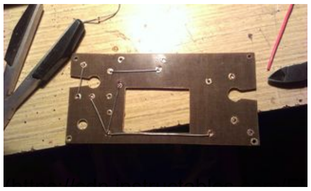

(https://cdn.instructables.com/FP

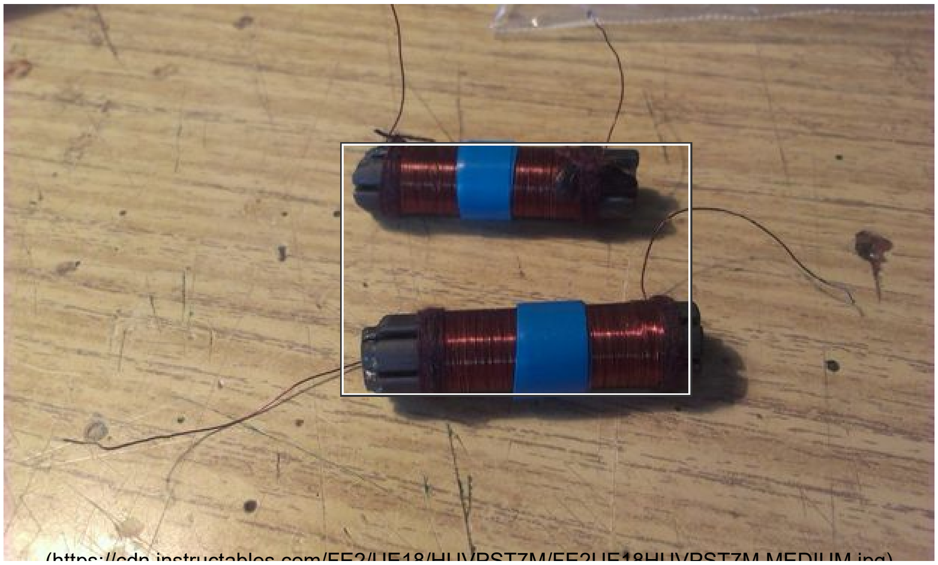

(https://cdn.instructables.com/FF2/UE18/HUVRSTZM/FF2UE18HUVRSTZM.MEDIUM.jpg)

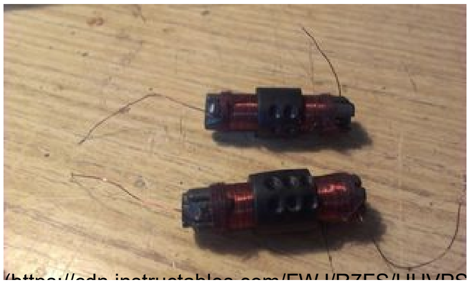

(https://cdn.instructables.com/FW1/RZES/HUVRS

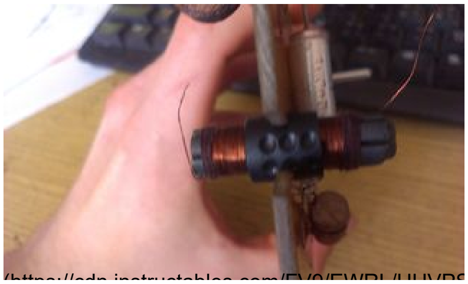

(https://cdn.instructables.com/FV9/FWBL/HUVRS

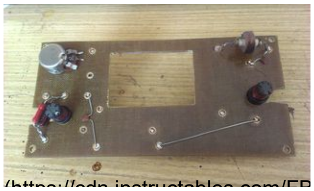

(https://cdn.instructables.com/FB

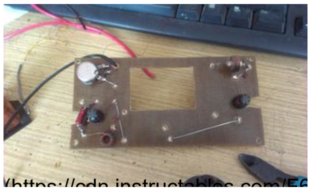

(https://cdn.instructables.com/F6

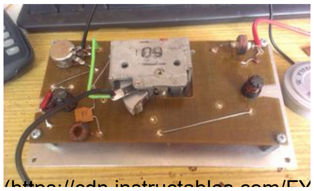

(https://cdn.instructables.com/FX

AdChoices
Yearly subscription only $69 $199
nature
SEA CHANGES
International weekly journal of science
SUBSCRIBE NOW
nature

Once you've got motherboard, frontboard and case for radio done, it's time on electronic part!

1. First, collect your electronic parts from step 1
2. Build your alfa version of crystal tuner (this step is not important, you can skip it!)
3. From original instruction lists you can find where every component has be placed. By numbers on schematics (this numbers are also on motherboard layout on original instructions), you will find every single soldering point on motherboard, so, just go on! :)
4. Now, you can solder and solder! First is good to solder wire interconnections. Then you can solder other components.
5. Another step is about coils holders. I didn't know how to solve this problem. In original, they use special rubber holders. I didn't have them. I used rubber holder from old pen, it works fine. Just look at pictures, it will say everything. :)
6. Finish funcional/electronic part, add some ground wires (on tuning cap for example in my case)
7. Done!..?

Step 6: Assembling

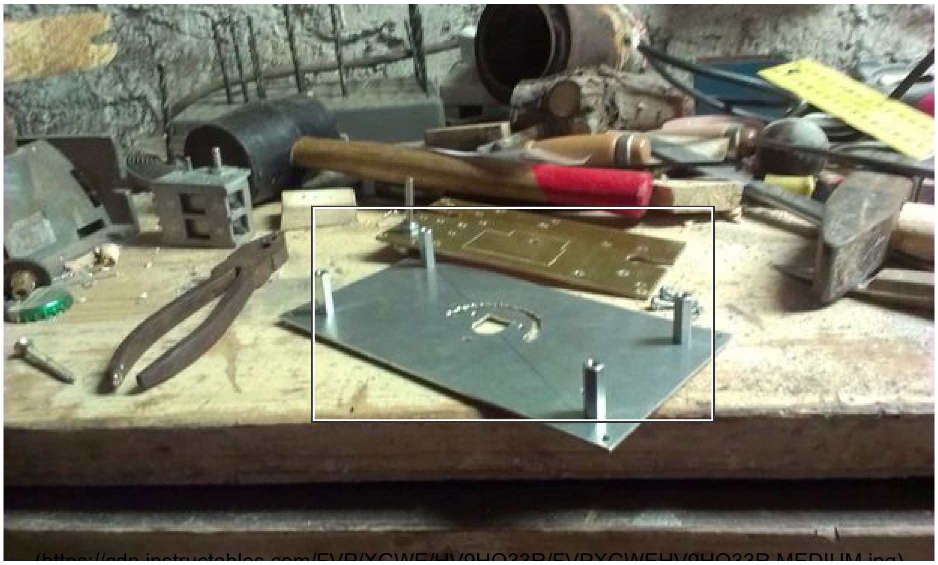
(https://cdn.instructables.com/FVB/XCWF/HV9HO33R/FVBXCWFHV9HO33R.MEDIUM.jpg)

(https://cdn.instructables.com/FB8/N04F/HV9HO

(https://cdn.instructables.com/FO2/JV9B/HV9HO

(https://cdn.instructables.com/FB

(https://cdn.instructables.com/FV

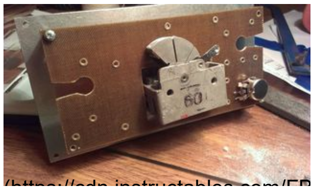
(https://cdn.instructables.com/FB

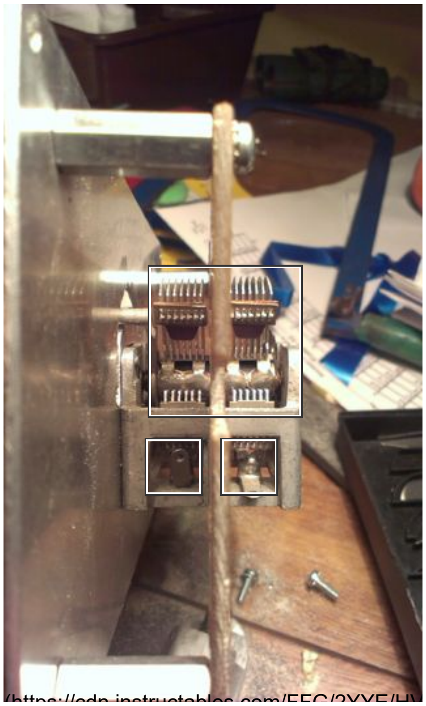
(https://cdn.instructables.com/FFC/2YXF/HV

AdChoices
Yearly subscription only $69 ~~$199~~
nature
SEA CHANGES
International weekly journal of science
SUBSCRIBE NOW
nature

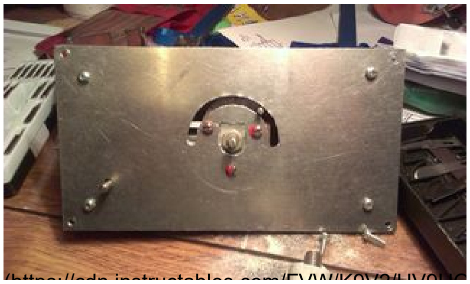

(https://cdn.instructables.com/FVW/K0Y2/HV9HQ

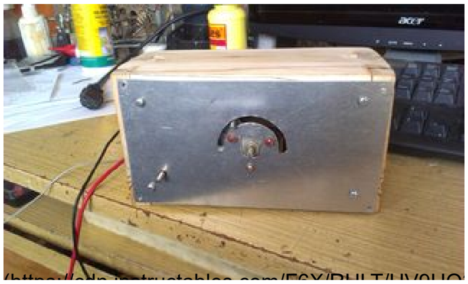

(https://cdn.instructables.com/F6X/RHLT/HV9HQ

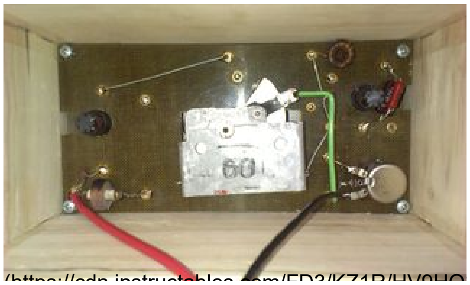

(https://cdn.instructables.com/FD3/KZ1B/HV9HQ

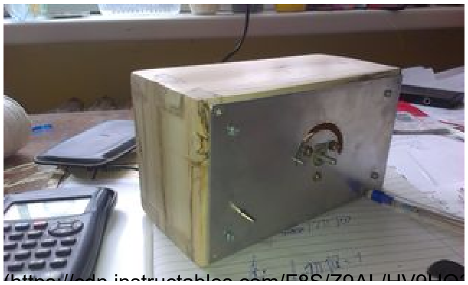

(https://cdn.instructables.com/F8S/Z9AL/HV9HQ

We have already finish, just assemly all parts into the case!

1. I chose solution with screwed mothorboard with frontboard, then frontboard is screwed with case. It's becaouse it is the siplest way for repearing/assembling/disassembling and also, I had not a good tuning cap for this case.
2. Pictures will say the same. :3

Step 7: Final Touches!

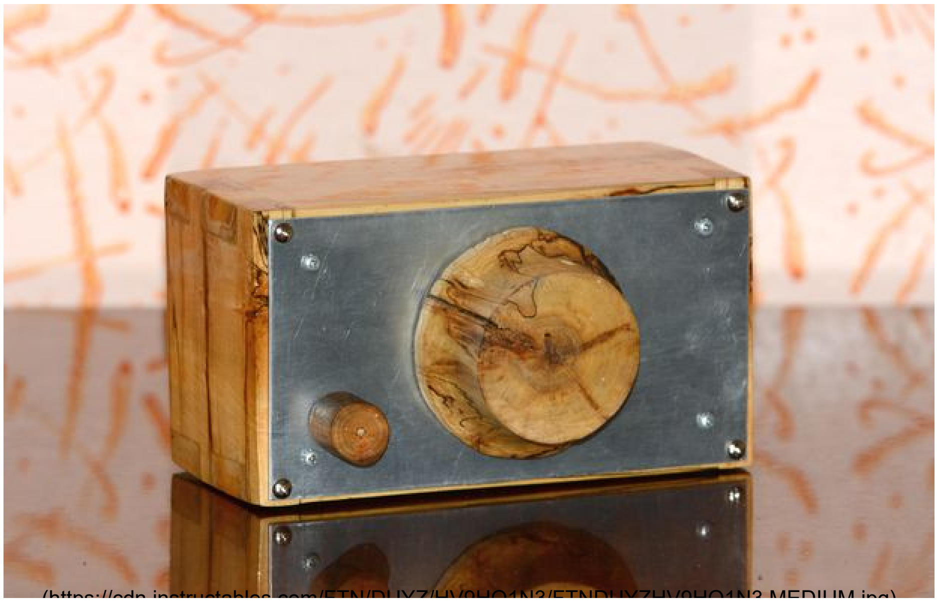

(https://cdn.instructables.com/FTN/DUYZ/HV9HQ1N3/FTNDUYZHV9HQ1N3.MEDIUM.jpg)

AdChoices
Yearly subscription
only $69
$199
nature
SEA CHANGES
International weekly journal of science

SUBSCRIBE NOW

nature

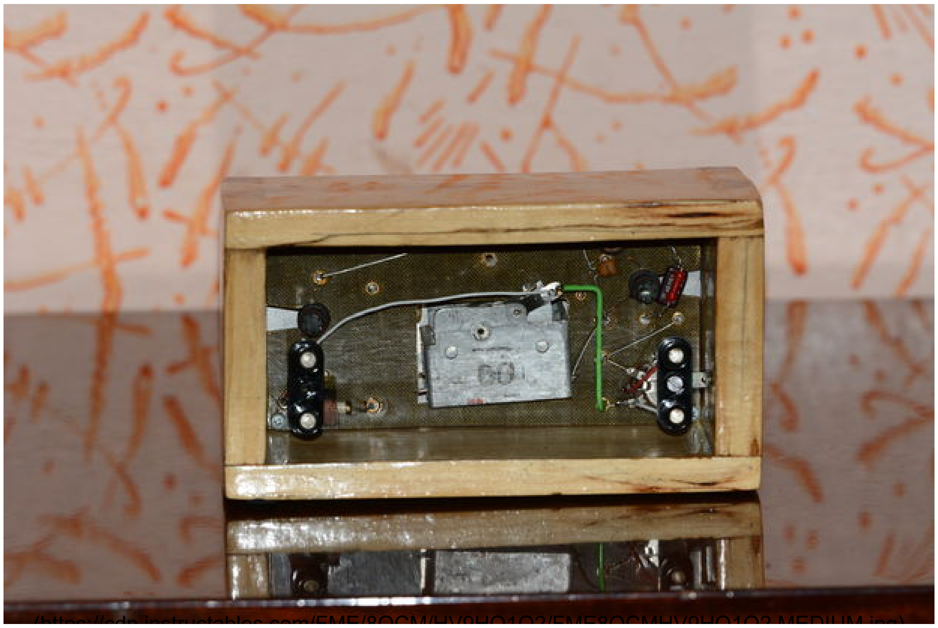

(https://cdn.instructables.com/FME/8OCM/HV9HQ1O2/FME8OCMHV9HQ1O2.MEDIUM.jpg)

AdChoices

Yearly subscription only $69 ~~$199~~

International weekly journal of science

SUBSCRIBE NOW

nature

This is the last step. You can add some buttons on potentiometer, or tuning cap, add some bannana jacks for antenna, ground and headphones and so on. That's only on you!

This is the end! Thanks for your time!

If you like my project, you can support me on facebook https://www.facebook.com/pages/D-Fox/236623253020... (https://www.facebook.com/pages/D-Fox/236623253020320?ref=hl) , or on instagram http://instagram.com/ondrej_fidra (http://instagram.com/ondrej_fidra) .

Respect the nature!

Peace!

advertisement

Make dead battery work

Never pay for new batteries again by using this 1 easy tip.

ezbatteryreconditioning.com

Make dead battery work

Never pay for new batteries again by using this 1 easy tip.

ezbatteryreconditioning.com

Comments

be nice

We have a be nice comment policy.
Please be positive and constructive.

I Made it!

Add Images

Post Comment

Colaris (/member/Colaris/) 2015-10-18 Reply

Instead of having the antenna drapped around, couldn't THAT be coiled around a tube also?

spark master (/member/spark+master/) 2014-05-22 Reply

Oh I fergit, if you have acess to formica or thin slate or anything non conductive you can build right on it. Stuff like this was built on wood, that was shellacked a few times or shellacked then waxed, or might be build on a chunk of (yikes) soapstone or an old tile.

But usually it was wood. As long as it was non ferrous, non conductive and not subject to water retention, it was good to go. On a few I just screwed the components to a block of scrap wood, then put it on an antenna or Arial .

A note on antenni, as ham's will tell you height makes maight and size matters.

The higher you get the antenna or the bigger it is the better it will work. Also if you get your antenna feed up high instead of one wire x feet long use 3 -4 all parallel that will snag more micro volts from the air giving you more energy into the detector circuits and headphones.

And

the better you ground this thing, the better it will work. For a really good ground bury a 1957 GM car (any one will do), at least ten feet below the ground attaching #8 stranded wire to it, back fill with lightly salted sand with a sprinkling of iron filings soak well with salted water and then put finished ground cover, water daily.

If that is unavailable,,,cold water pipes are sufficient.

they don't call me sparkie fer nut'n

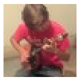

DiggingFox (/member/DiggingFox/) ▸ spark master (/member/spark+master/) 2014-05-23 Reply

Antenna and ground are very important. If you've got better antenna and ground, you've got better sound and selectivity! I've got about 20 meters long antenna in 30 meters, and it's fine (sometimes is very dangerous! when the storm comes!)

Thanks for your comments! 73!

spark master (/member/spark+master/) 2014-05-22 Reply

here is a page that has the same jpegs, but can be enlarged enough to read them

http:/ remove this /fmamradios.co remove this m/pics/Miller565-2-3.jpg

I think this version is nicer then the original

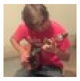

DiggingFox (/member/DiggingFox/) ▸ spark master (/member/spark+master/) 2014-05-23 Reply

If you look at the full size of the picture of instruction on this site, you can read them too! :) Thanks you!

spark master (/member/spark+master/) 2014-05-22 Reply

AdChoices
Yearly subscription only $69 ~~$199~~
nature
SEA CHANGES
International weekly journal of science
SUBSCRIBE NOW
nature

I have built a number of Xtal Radios and they are fun, but where does one purchase old huge variable caps?

You can make them (albeit they may be a tad crappy, or you may need to experiment to get the right capacitance), or if you find old dumped radios you are good to go. It has been years , but I thought the tuner was about 360 MFD's or so. Great project!

DiggingFox (/member/DiggingFox/) ▸ spark master (/member/spark+master/) 2014-05-23 Reply

Thanks! The best way to get the big caps is take it from old damaged radios (don't kill them!! old radios are jewelery!! look at the design!!)...but, this cap has about 300 mmFD on each side, 500 mmFD on each side is better!

manohar (/member/manohar/) 2014-05-20 Reply

could you please provide the circuit diagram for electronics.

DiggingFox (/member/DiggingFox/) ▸ manohar (/member/manohar/) Reply 2014-05-20

Hi! Circuit diagram and schematic is in step 1, in original instructions.

Tommy_tux (/member/Tommy_tux/) 2014-05-20 Reply

Very nice design... Maybe you can build in a web radio? Thanks for this idea...

DiggingFox (/member/DiggingFox/) ▸ Tommy_tux (/member/Tommy_tux/) Reply 2014-05-20

Tkanks you! Yea, I could, I am working on web radio with STM32, but it's too hard for me, I have to learn more about it :)

ASCAS (/member/ASCAS/) 2014-05-19 Reply

Man that sure is awesome!

DiggingFox (/member/DiggingFox/) ▸ ASCAS (/member/ASCAS/) Reply 2014-05-19

Thanks!

Akin Yildiz (/member/Akin+Yildiz/) 2014-05-18 Reply

very nice !

DiggingFox (/member/DiggingFox/) ▸ Akin Yildiz (/member/Akin+Yildiz/) Reply 2014-05-19

Thank you!

mikeasaurus (/member/mikeasaurus/) 2014-05-18 Reply

AdChoices
Yearly subscription only $69 ~~$199~~
International weekly journal of science
SUBSCRIBE NOW
nature
nature
SEA CHANGES

This looks great. It *wood* be fun to see this in a log cabin

---

DiggingFox (/member/DiggingFox/) ▸ mikeasaurus (/member/mikeasaurus/)

2014-05-18 Reply

Thanks! Haha, yea it *wood* be :)

---

plantprof (/member/plantprof/) ▸ DiggingFox (/member/DiggingFox/)

Reply 2014-05-18

Nice project and nicely presented. Love your sense of humor and I wish I could communicate as effectively in Czech as you can in English. Enjoyed reading and following your Instructable! I found myself smiling with admiration for what you made, and that you took time to share it here. Did you use that back saw that is visible in the photo to cut the pieces of wood? [That must have been fun!] And where did you you come across the old plans for the radio? [the internet?] How much chocolate is needed! :)

DiggingFox (/member/DiggingFox/) ▸ plantprof (/member/plantprof/)

Reply 2014-05-19

Thank you! yes, I used the saw visible on the fotos (with the orange handle), the plans are from the internet, of course :) I am a bit chocoholic, I ate about 10 bars od chocolate :D hope, you like it :)

---

andrea biffi (/member/andrea+biffi/)

2014-05-18 Reply

cool design!

DiggingFox (/member/DiggingFox/) ▸ andrea biffi (/member/andrea+biffi/)

Reply 2014-05-18

Thanks man!

More Comments

AdChoices
Yearly subscription
only $69
$199
nature
SEA CHANGES
International weekly journal of science
SUBSCRIBE NOW
nature

Hi-Fi Crystal AM Radio Tuner / No. 565 Miller

by DiggingFox (/member/DiggingFox/) in audio (/technology/audio/)

Download | (/id/Hi-Fi-Crystal-AM-radio-tuner-No-565-Miller/) | 7 Steps

Collection | I Made it! | Favorite | Share

advertisement

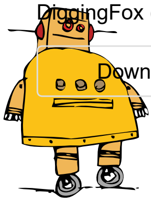

Newsletter

et your inbox help you discover our best projects, classes, and contests. Instructables will help you learn how to make anything!

enter email | I'm in!

About Us

Who We Are (/about/)
Advertise (/advertise/)
Contact (/about/contact.jsp)
Jobs (/community/Positions-available-at-Instructables/)
Help (/id/how-to-write-a-great-instructable/)

Find Us

Facebook (http://www.facebook.com/instructables)
Youtube (http://www.youtube.com/user/instructablestv)
Twitter (http://www.twitter.com/instructables)
Pinterest (http://www.pinterest.com/instructables)
Google+ (https://plus.google.com/+instructables)

Resources

For Teachers (/teachers/)
Residency Program (/pier9residency)
Gift Premium Account (/account/give?sourcea=footer)
Forums (/community/?categoryGroup=all&category=all)
Answers (/tag/type-question/?sort=RECENT)
Sitemap (/sitemap/)

Terms of Service (http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=21959721) |

Privacy Statement (http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=21292079) |
Legal Notices & Trademarks (http://usa.autodesk.com/legal-notices-trademarks/) | Mobile Site (https://www.instructables.com)
(http://usa.autodesk.com/adsk/servlet/pc/index?id=20781545&siteID=123112)
© 2017 Autodesk, Inc.

AdChoices
Yearly subscription only $69 ~~$199~~
nature
Evolutionary history of reef fish reveals how marine life diversifies near islands and seamounts
SEA CHANGES
SCIENCE IN GERMANY
WASTED LIVES AND MONEY
LONG-DISTANCE INFORMATION
International weekly journal of science
SUBSCRIBE NOW
nature

High Performance Crystal Radios

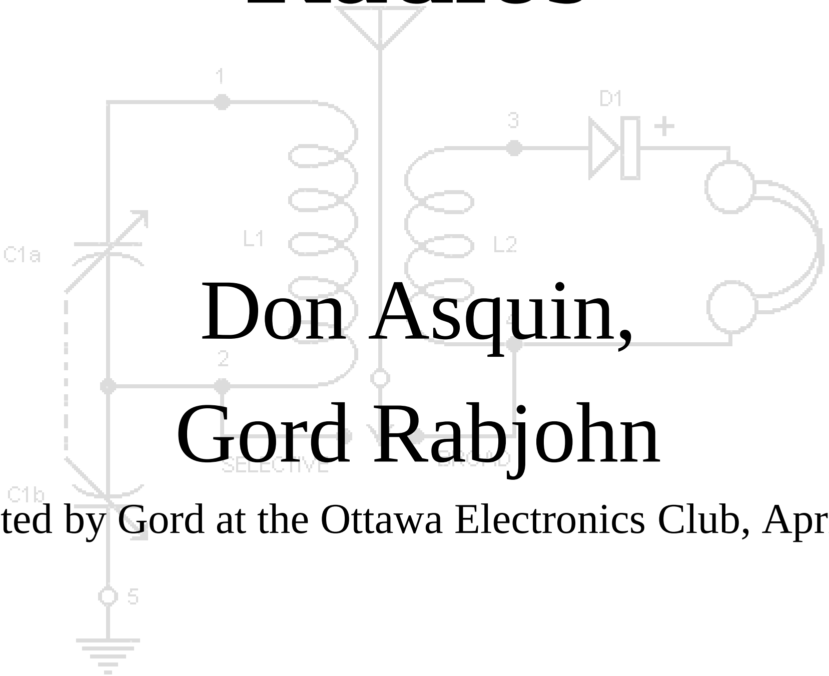
1
D1
3
+
C1a
L1
L2
2
SELECTIVE
C1b
5

Don Asquin,

Gord Rabjohn

(Presented by Gord at the Ottawa Electronics Club, April 2012)

Childhood History

www.crystalradio.net

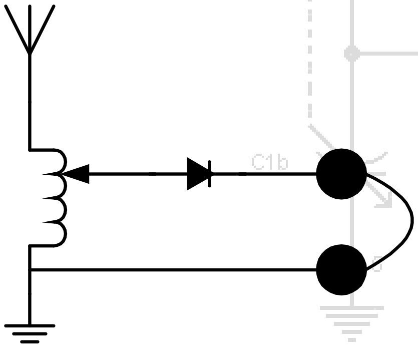

- Like many kids, I had frustrating childhood experiences with crystal radios!
- Simple crystal sets couldn’t separate multiple AM radio stations.
- Only local stations were received, and everything was received all at once.

Rediscovering History

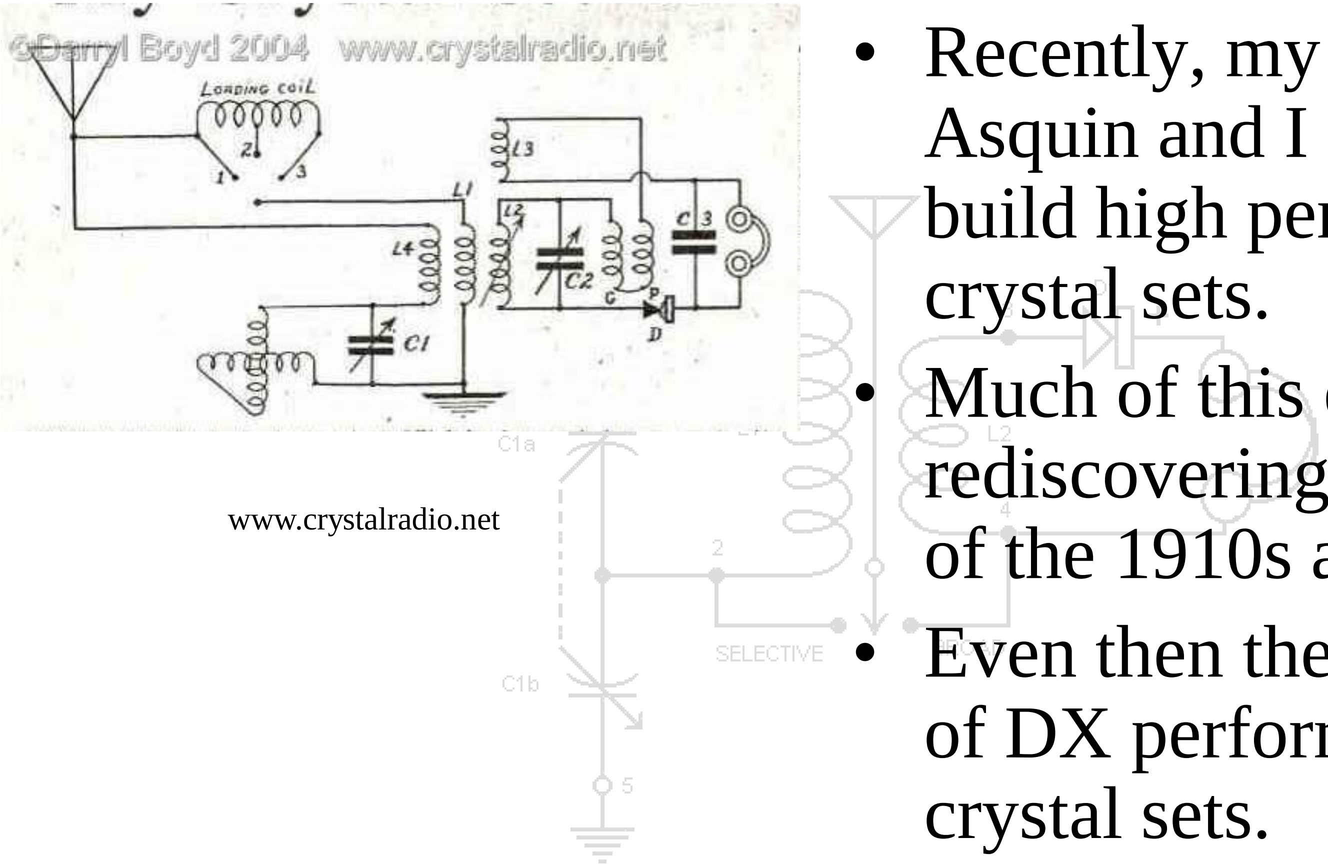
©Darryl Boyd 2004 www.crystalradio.net
LOADING COIL
1
2
3
L1
L2
L3
L4
C1
C2
C3
G
P
D
C1a
C1b
2
5
L2
4
SELECTIVE

www.crystalradio.net

- Recently, my friend Don Asquin and I decided to build high performance crystal sets.
- Much of this exploration is rediscovering the knowledge of the 1910s and early 1920s.
- Even then there were claims of DX performance with crystal sets.

How does a basic crystal set work?

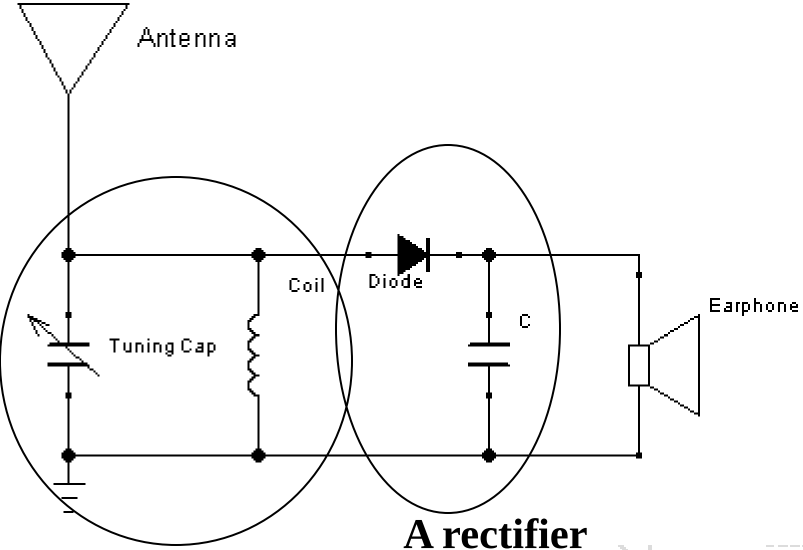
Antenna
Coil
Diode
Tuning Cap
C
Earphone

A rectifier

A filter

At low frequency, the inductor shorts the signal out. At high frequency the capacitor shorts the signal out. In-between, there is a magic point called "resonance"

- Antenna couples to the (usually) electric field in a radio wave.
- L-C selects the frequency.
- Diode rectifies it.
- Earphone makes it audible

AM modulated signal

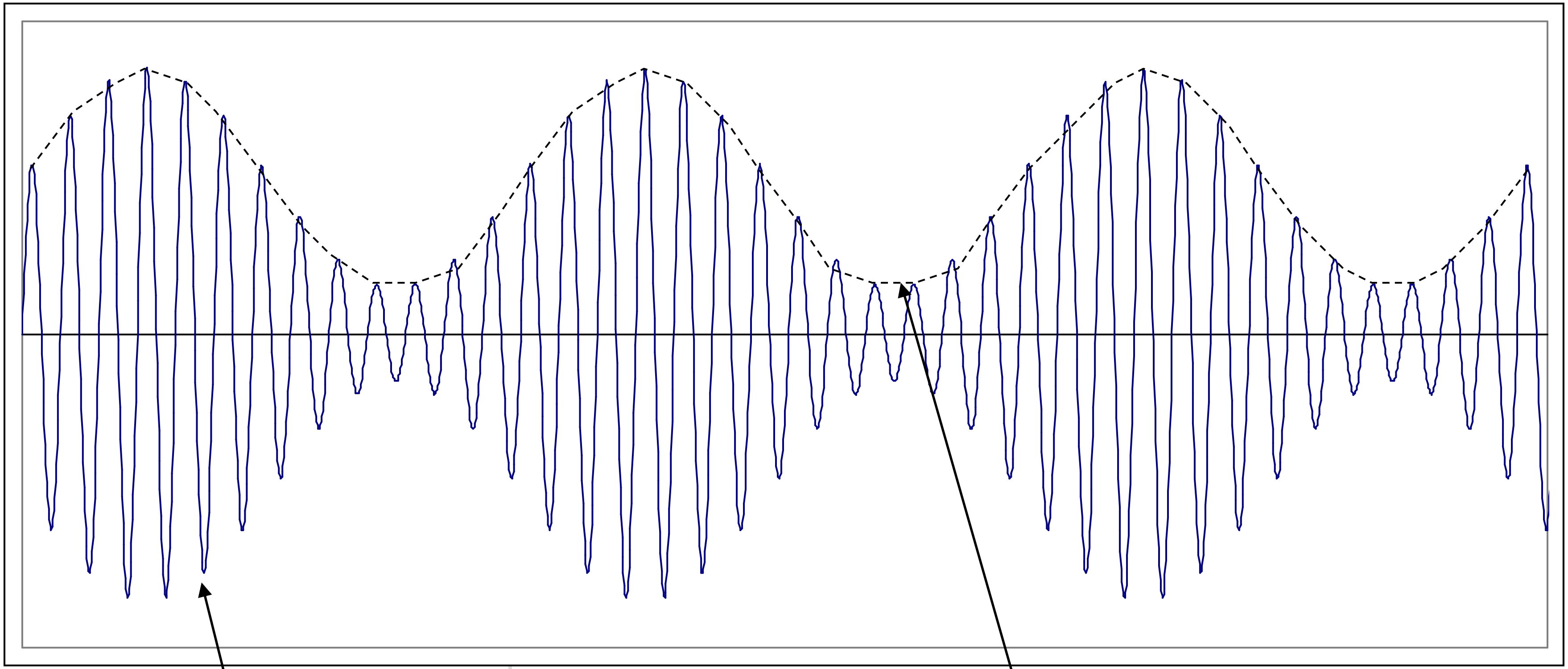

RF Carrier frequency (say 1310kHz): the filter selects only this frequency.

Rectified audio output, smoothed by the capacito

The Four Key Elements

ANTENNA
Federal Jr.
www.crystalradio.net

www.crystalradio.net

- Antenna
  - Antenna
  - Ground
- Tuner
  - Loose coupler
  - High Q coils
  - RF impedance matching
- Detector
- Audio Transducer
  - Audio Impedance matching

Antenna

- A long exterior antenna is crucial.
- Why? It’s not just about area… “watts per acre”. It’s also about giving the antenna a better “radiation impedance”.
  - The energy from an antenna, can be modeled as a voltage source in series with a source resistance (or a current source in parallel with a source conductance), and the antenna capacitance (or inductance).
  - Radiation resistance is the resistance of this apparent source of energy in an antenna.
  - The tuned circuit matches this impedance to the load (detector) for optimum energy transfer.
  - A small antenna will tend to look like a “difficult to match” impedance….

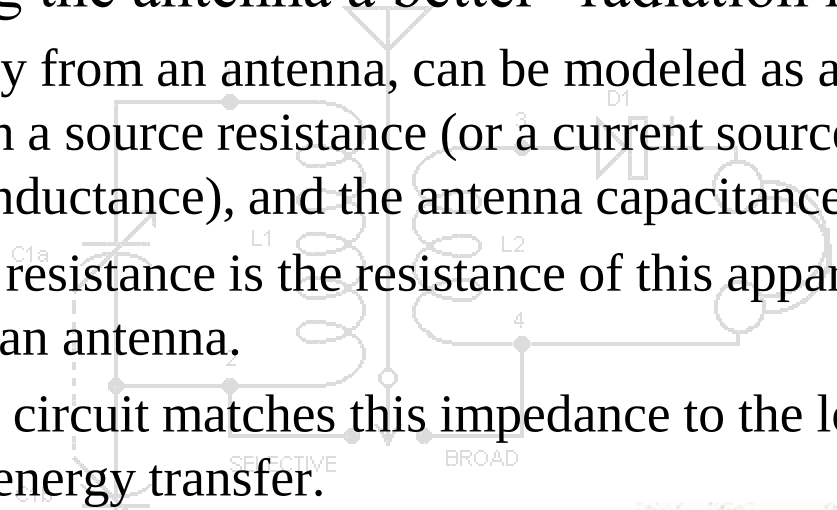
D1
3
L1
L2
C1a
4
2
SELECTIVE
BROAD
C1b
5

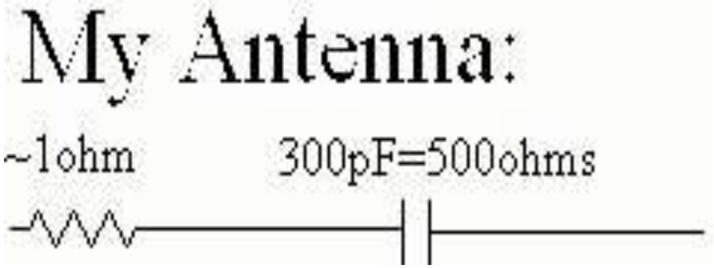
My Antenna:
~1ohm
300pF=500ohms

⏚ Ground ⏚

- The ground provides the completion of the current path for the radio signals.
- You can always attach a wire to a cold water pipe, but the best ground remains a copper plated rod pounded into the ground.
- Home Depot sells these rods for grounding electrical systems.
- If installed near a water tap, the ground can be kept moist and conductive.

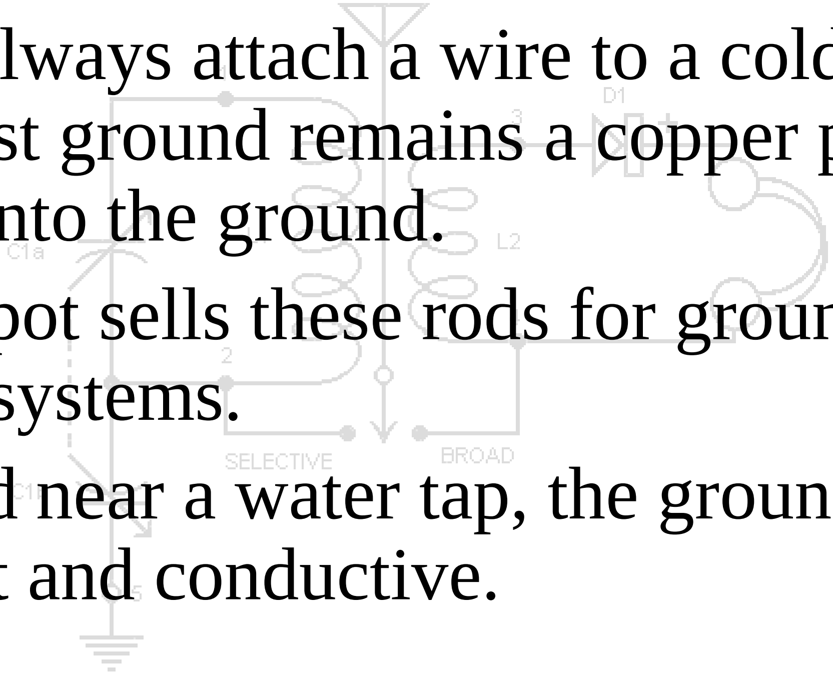

Tuner
-Antenna Tuner

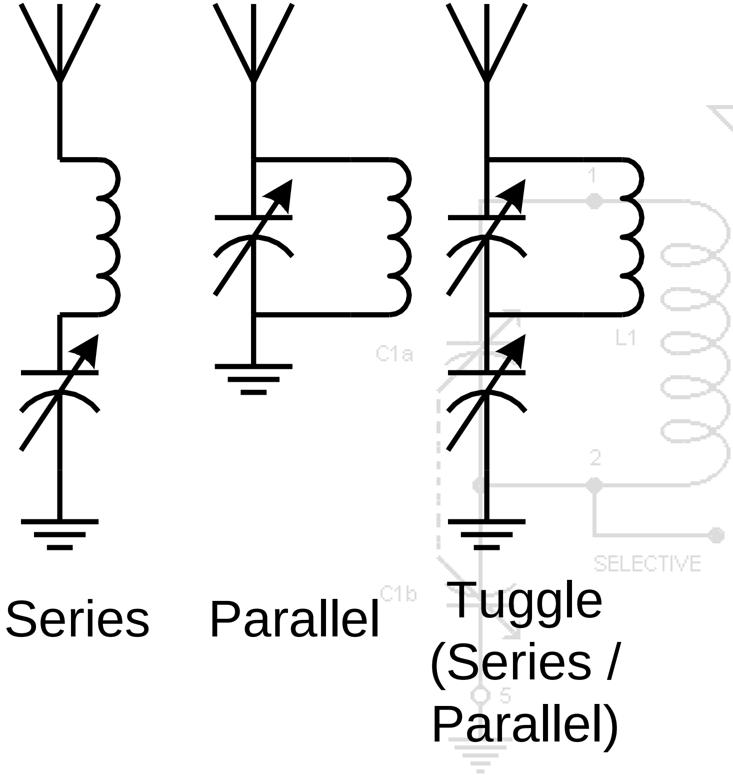
Series
Parallel
Tuggle
(Series /
Parallel)
C1a
C1b
L1
L2
SELECTIVE
BROAD
1
2
3
4
5

- To extract the most power, you need to match the resistance of the load (detector and earphone) to the impedance of the source (antenna)
- The function of the tuner is to match the radiation resistance of the antenna to the impedance of the detector, and to provide selectivity.

Tuner
-Loose Coupler

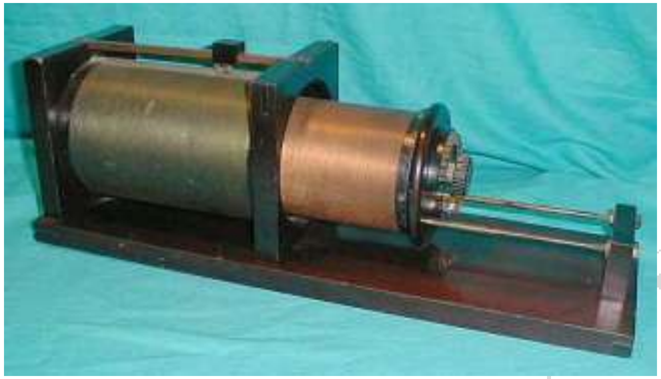

www.crystalradio.net

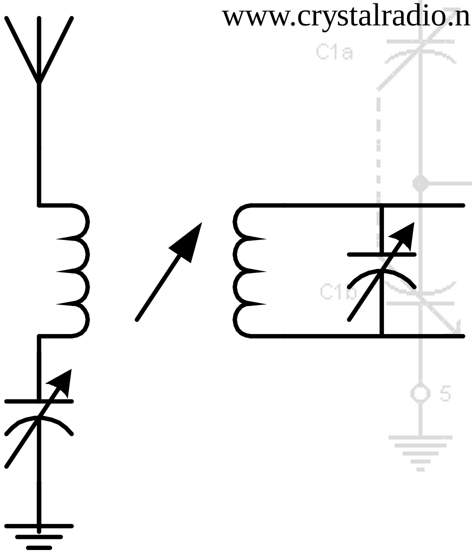

- The loose coupler is simply two coils (each generally resonated with a capacitor) lightly coupled to each other.
- One coil is connected to the antenna, the other is connected to the detector.
- Even cheap capacitors tend to be quite good, but high Q inductor fabrication is an “art”.

Loose Coupler

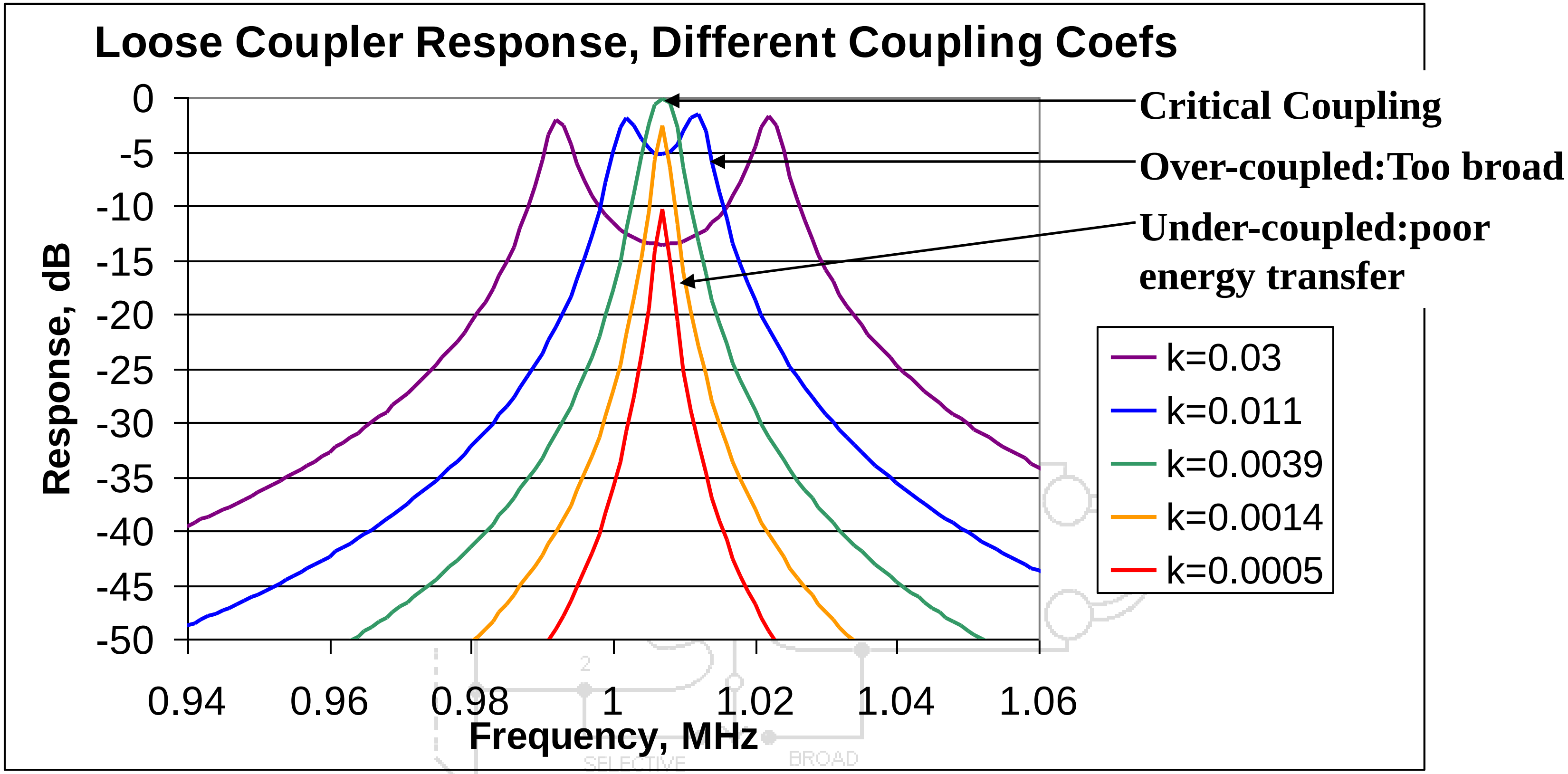

Critical Coupling

Over-coupled:Too broad

Under-coupled:poor energy transfer

Simulation of 1000pF, 1.0 ohm antenna, 100uH primary coil and ~330pF cap, coupled to a 100uH secondary coil, 250pF cap, 100k load.

- There is an optimum amount of coupling between the coils. (=1/Q). This is VERY LIGHT coupling. Under 1% of the magnetic field is coupled.

Coil "Q" = Quality

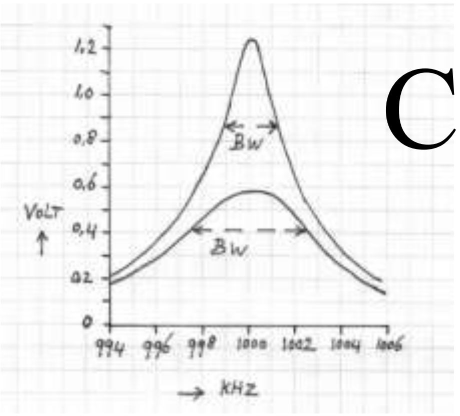

www.crystal-radio.eu

$$Q = \frac{\text{Frequency}}{\text{Bandwidth}}$$

Station Frequency (kHz)	Minimum Q
540	54
1000	100
1600	160

Assume Q=160

Station Frequency (kHz)	Audio Bandwidth (kHz)
540	1.7
1000	3.1
1600	5.0

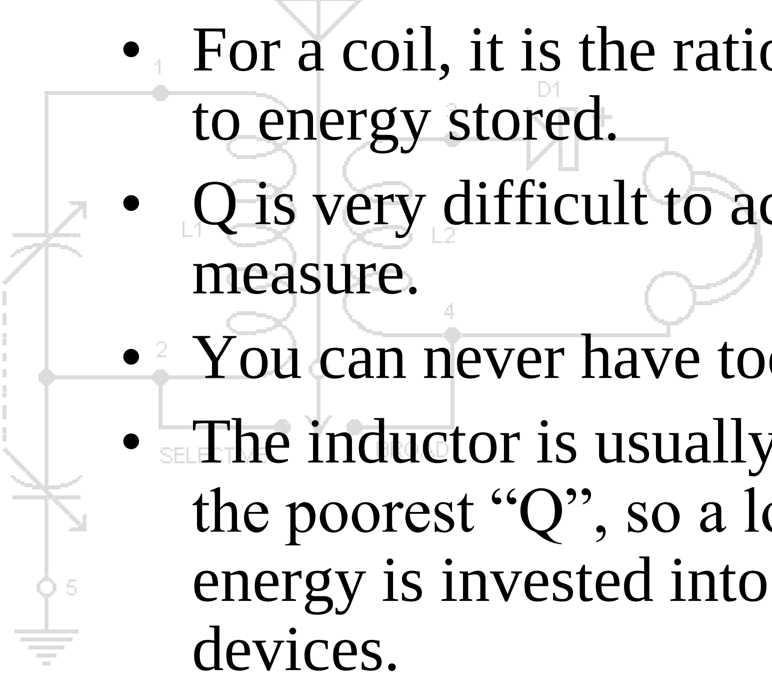

- For a tuned circuit "Q" is the ratio of the center frequency to the bandwidth.
- For a coil, it is the ratio of energy lost to energy stored.
- Q is very difficult to accurately measure.
- You can never have too much Q!
- The inductor is usually the part with the poorest "Q", so a lot of creative energy is invested into optimum devices.

Coil Types

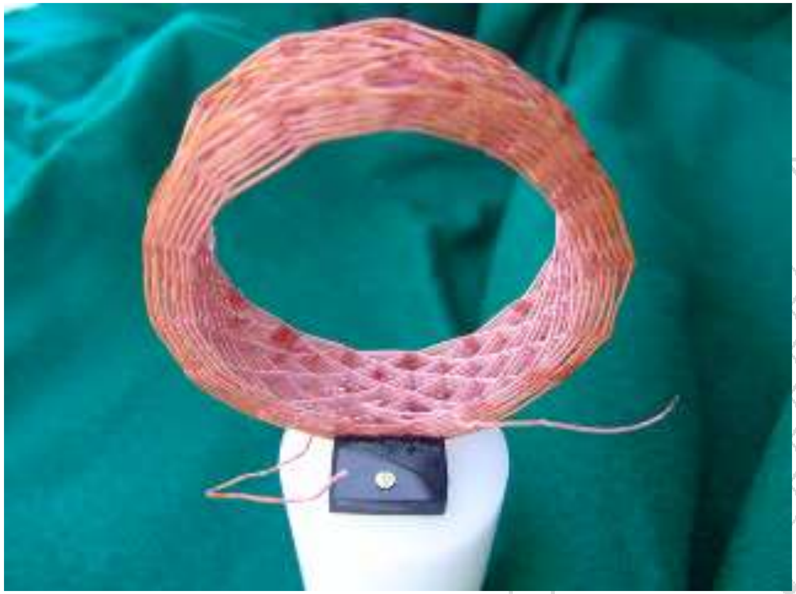

www.wynterarchtops.com

- The highest “Q” coils seem to be air core.
- Many different variations
  - Spider web,
  - Basket weave
  - Honey Comb
- All attempt to reduce capacitance and current crowding to increase Q

Coil Winding Jig

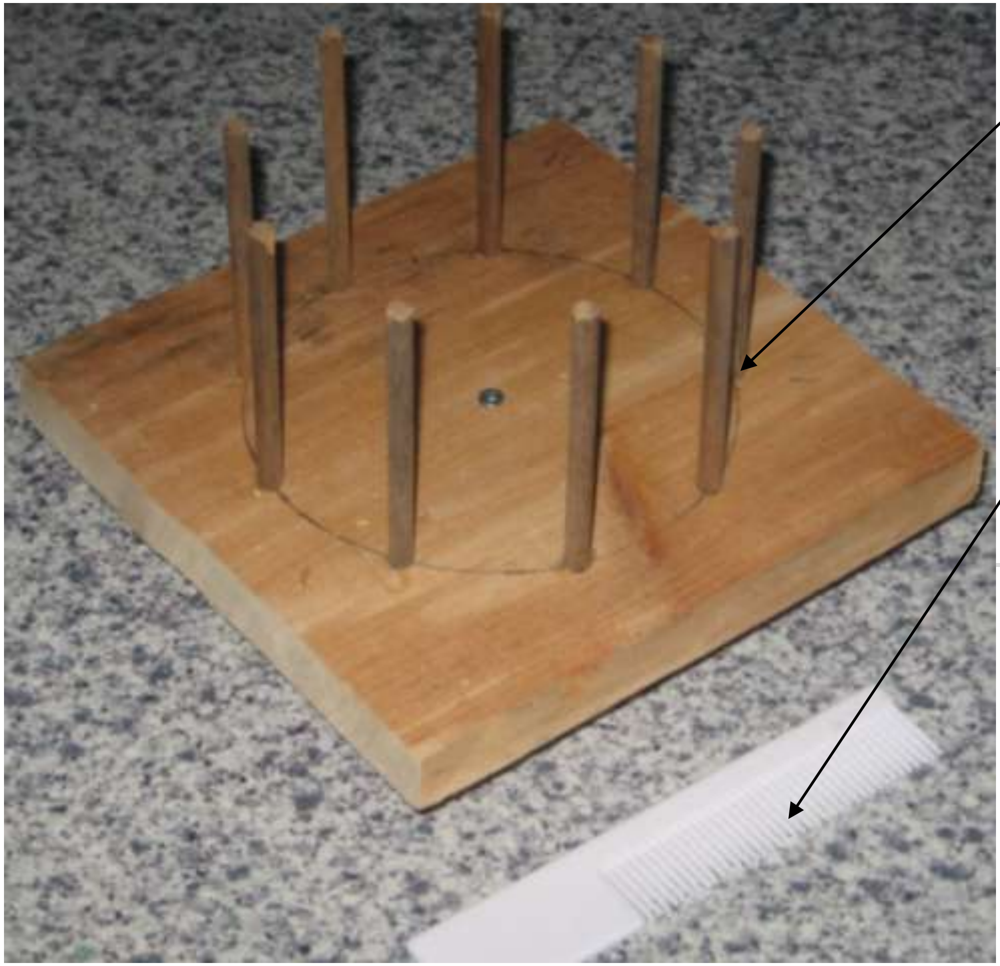

Drinking straws fit over ¼” dowels.

Comb helps space wires

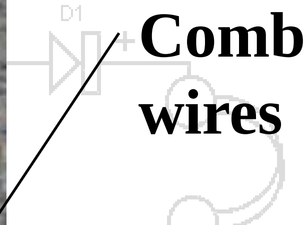

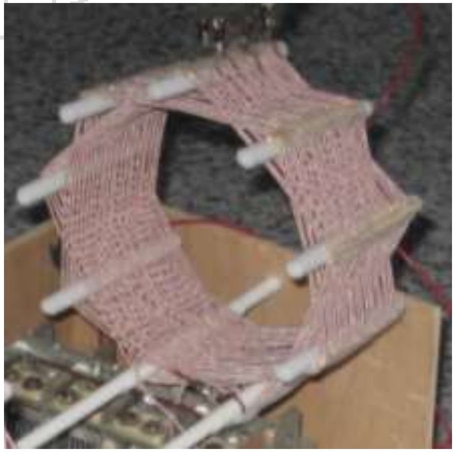

Litz Wire

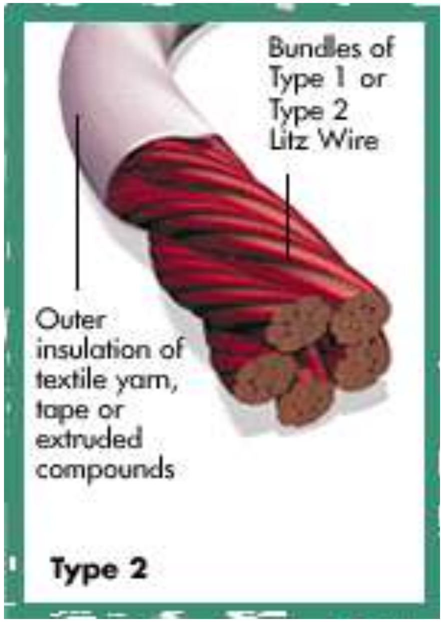
Bundles of Type 1 or Type 2 Litz Wire
Outer insulation of textile yarn, tape or extruded compounds
Type 2

www.newenglandwire.com

- Best wire for high “Q” coils is Litz wire.
- Litz is derived from the German word “Litzendraht” meaning woven wire.
- Consists of many strands of parallel connected, individually insulated wire woven together in a regular pattern.
- Each wire alternates between the middle and the outside of the bundle.
- Each wire forced to carry about the same current, minimizing skin effect (the tendency for current to flow along the outside surface of a wire), and loss.
- The holy grail of litz wire is made up of 420 to 660 individual strands of 46 AWG wire all twisted together to make a 16-18 AWG wire.
- One comparison: A basket wound coil with solid copper wire (~200uH) has a Q of 230 at 1MHz. With Litz wire, it has a Q of over 500!

www.crystalradio.net

Detector

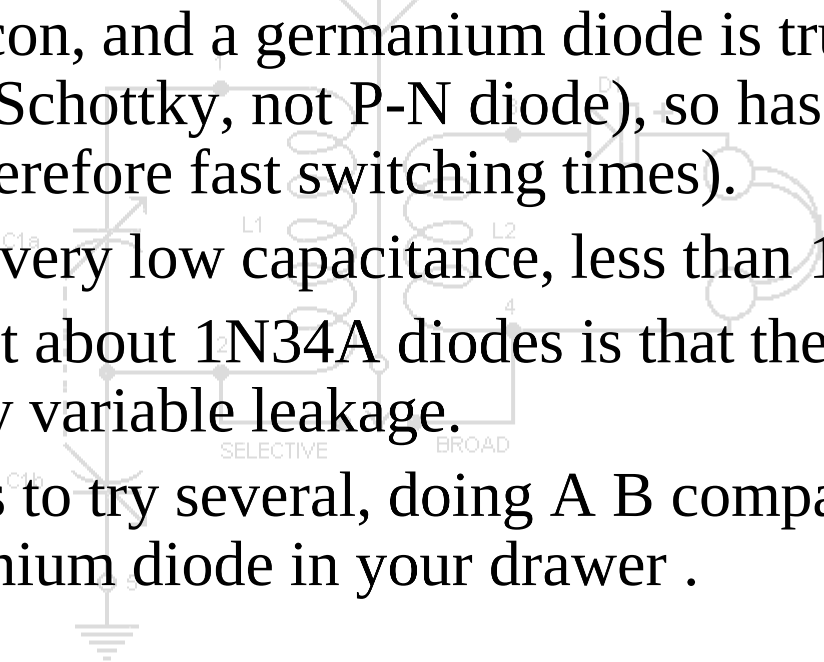

- In my experience, nothing beats a good 1N34A germanium diode. (Still widely available)
- Germanium is good because its barrier height (turn-on voltage) is lower than silicon, and a germanium diode is truly a point-contact diode (Schottky, not P-N diode), so has low charge storage (and therefore fast switching times).
- Specified with very low capacitance, less than 1.0pF.
- Main complaint about 1N34A diodes is that they tend to have high and highly variable leakage.
- Best solution is to try several, doing A B comparisons and select the best germanium diode in your drawer .

Detector Impedance

- Detectors are most efficient when driven with *high voltages*.
  - Detectors are “square law” devices (at low power), output voltage is proportional to square of input voltage.
- Since we have a fixed amount of input *power*, we need to increase voltage and decrease current applied to the detector (Power=Voltage x Current); in other words increase the detector *impedance*.
- At detector input: use a loose coupler.
- At detector output: use a 100kohm microphone transformer to match into the headphones.

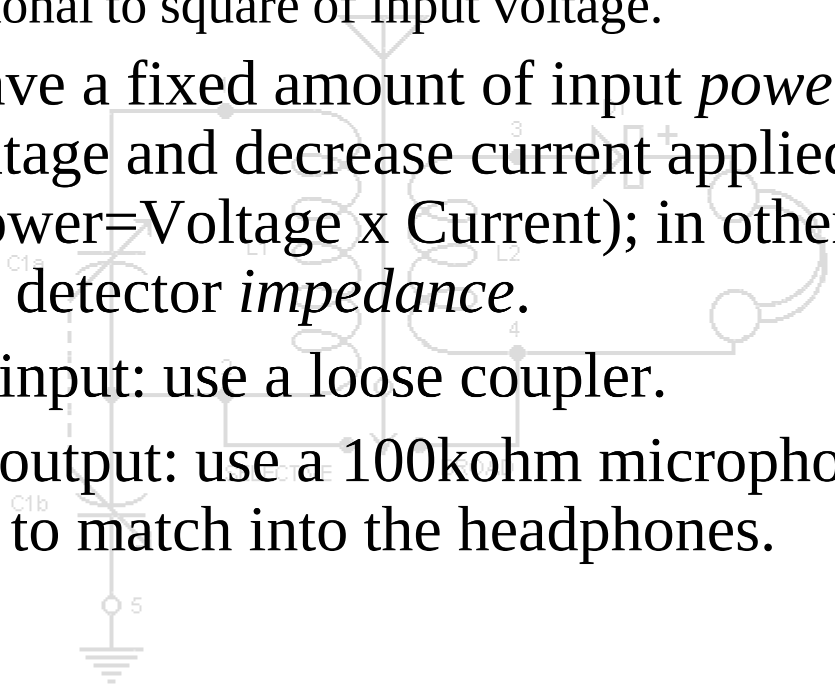

Germanium Diodes

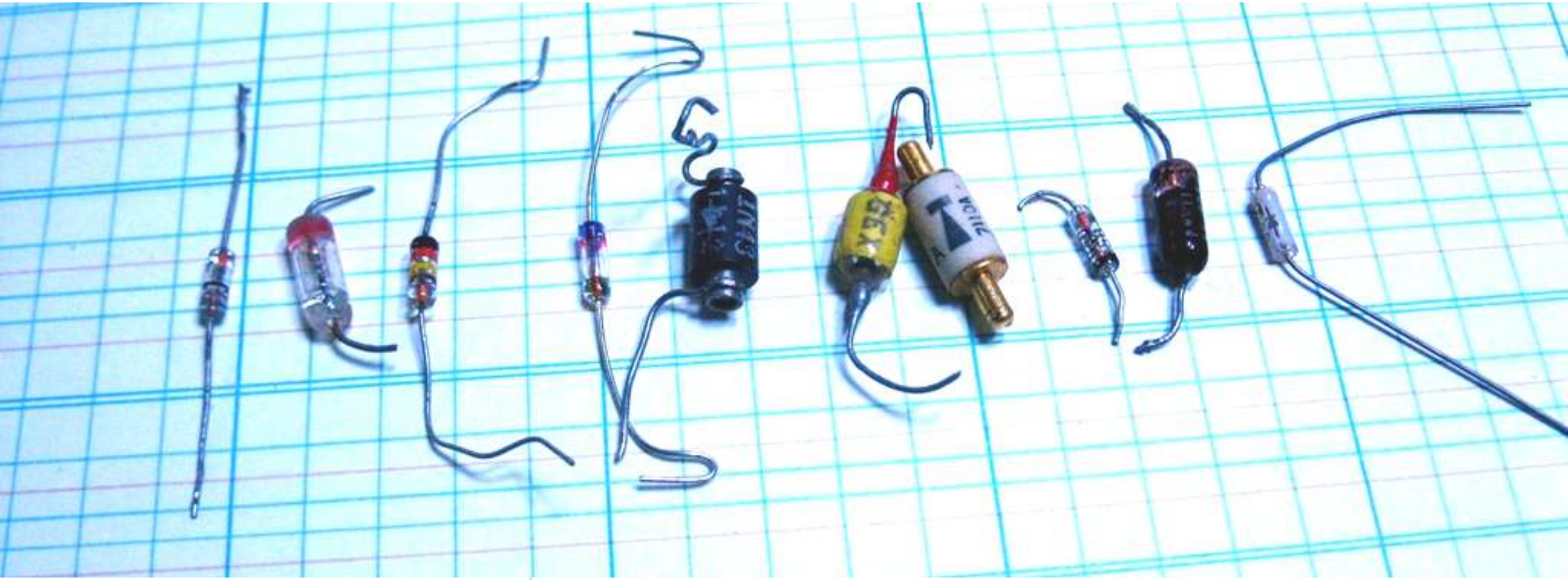

Detectors: I-V Curves

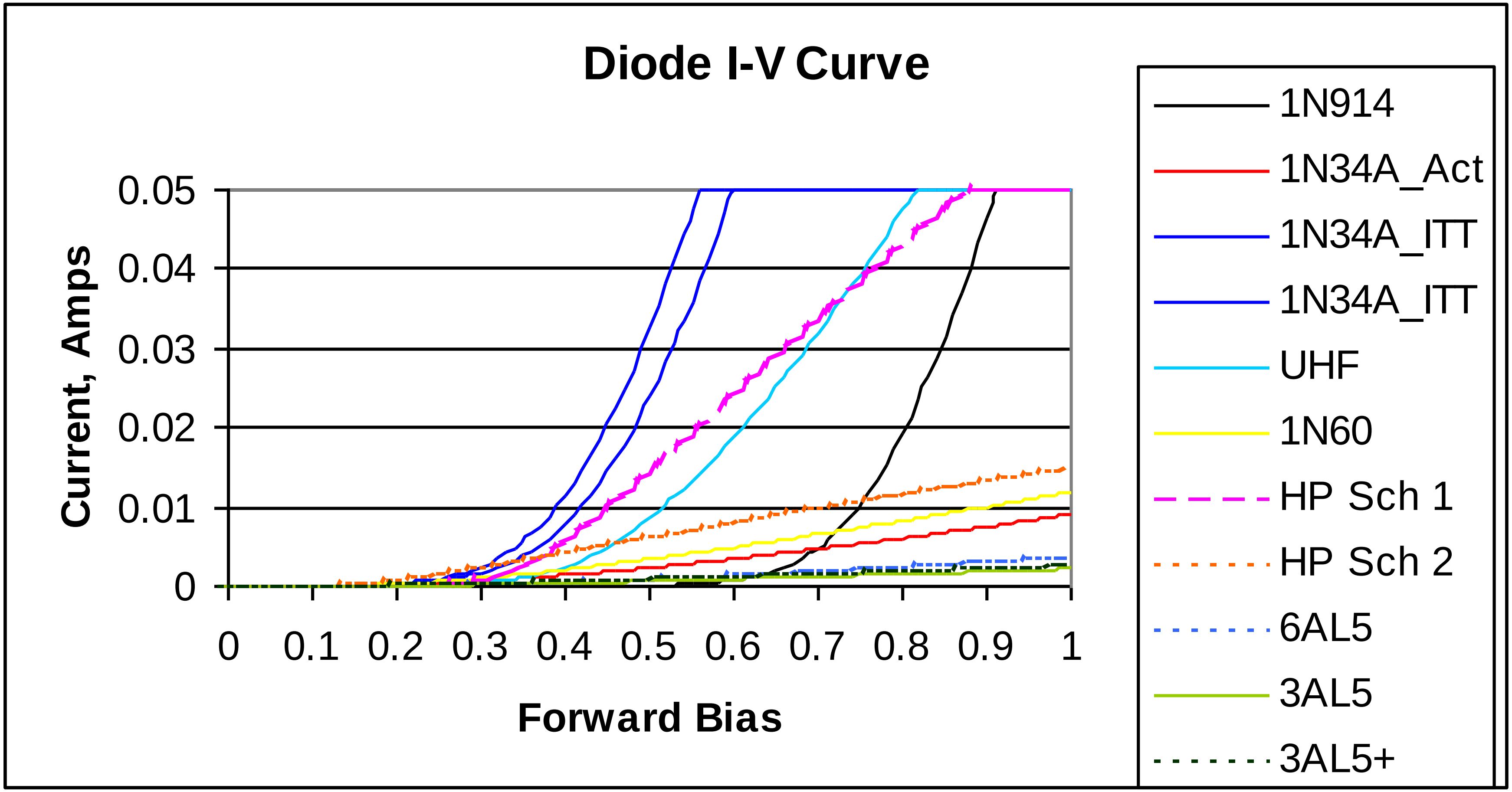

Detectors: I-V Curves

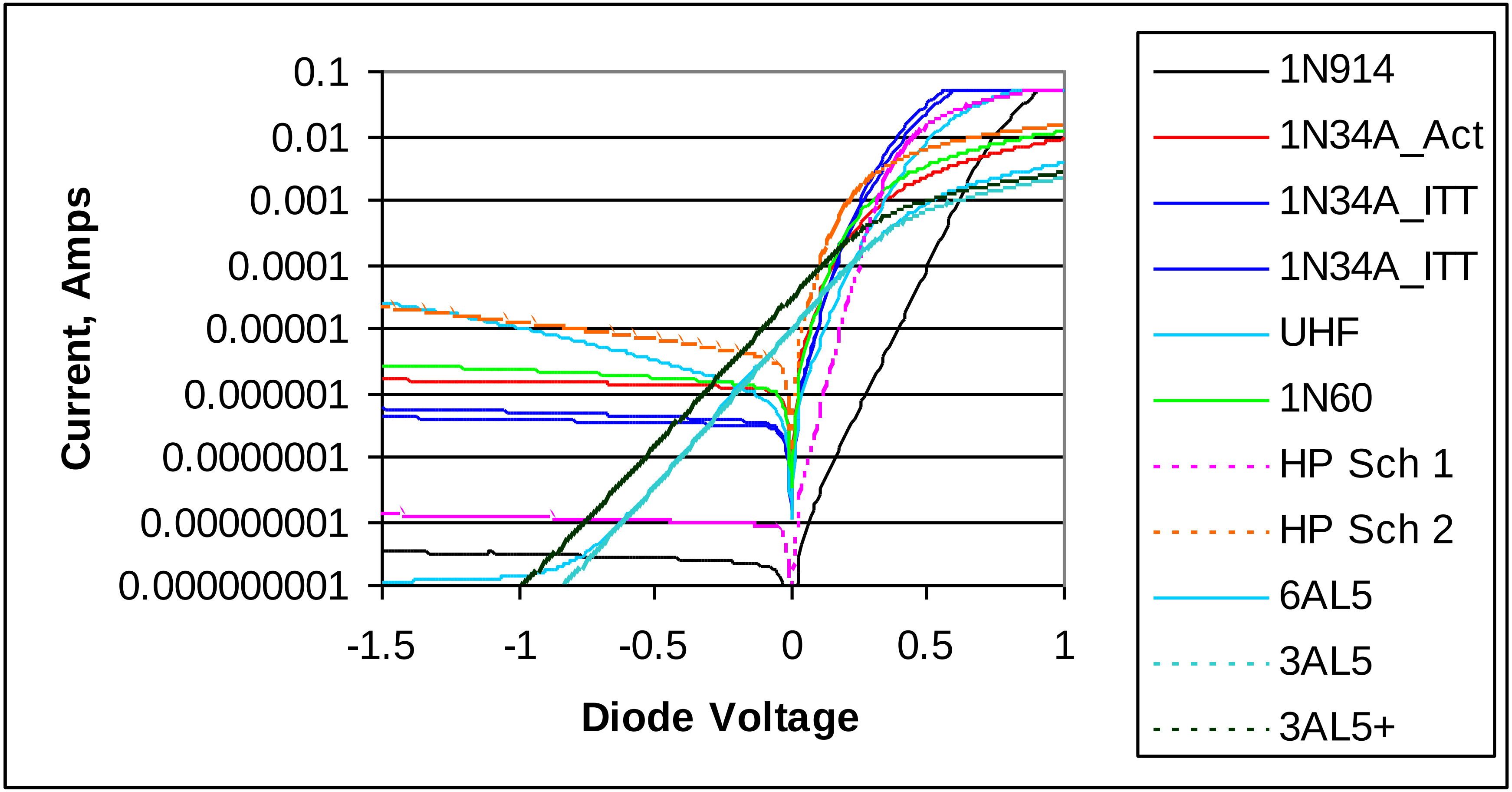

Mineral Detectors

Lead Sulfide, PbS (Galena), Iron Pyrite, Zincite, and other minerals have been used to make detectors (surprisingly good).

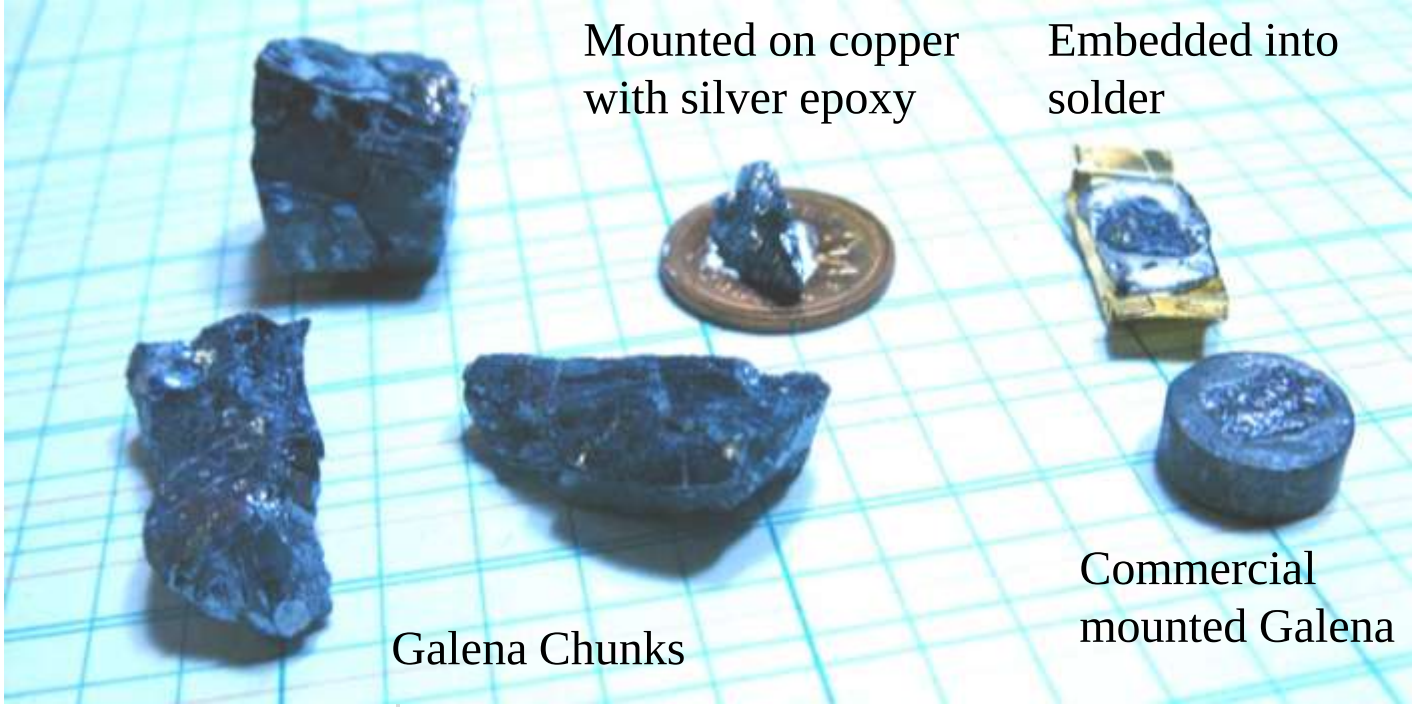

Mounted on copper with silver epoxy

Embedded into solder

Galena Chunks

Commercial mounted Galena

Galena Detector

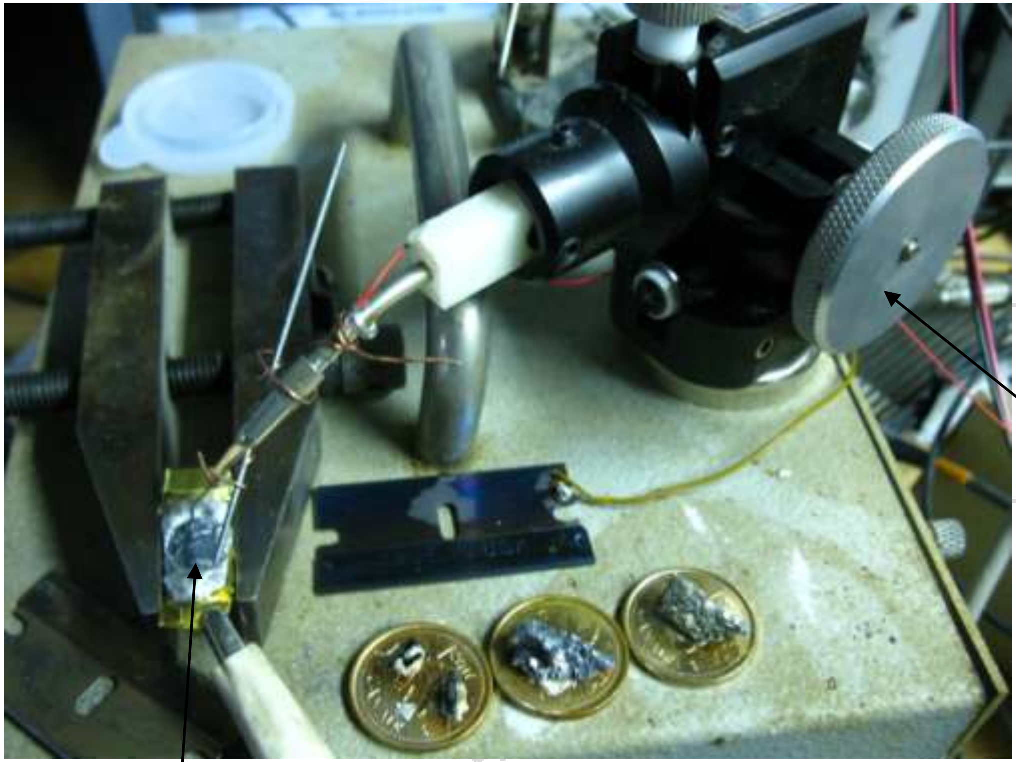

X-Y-Z manipulator

Galena mounted in solder

Audio Transducer

www.crystalradio.net

- Best transducers (headphones) for crystal radio operation are “Sound Powered” headphones or “Deck Talkers”.
- Developed in WWII for the navy to allow communications between the observers and gunners without the need for external power source.
- So sensitive that the pressure of the talkers voice in the microphone is sufficient to drive the headphones of the listener.

Audio Transducer

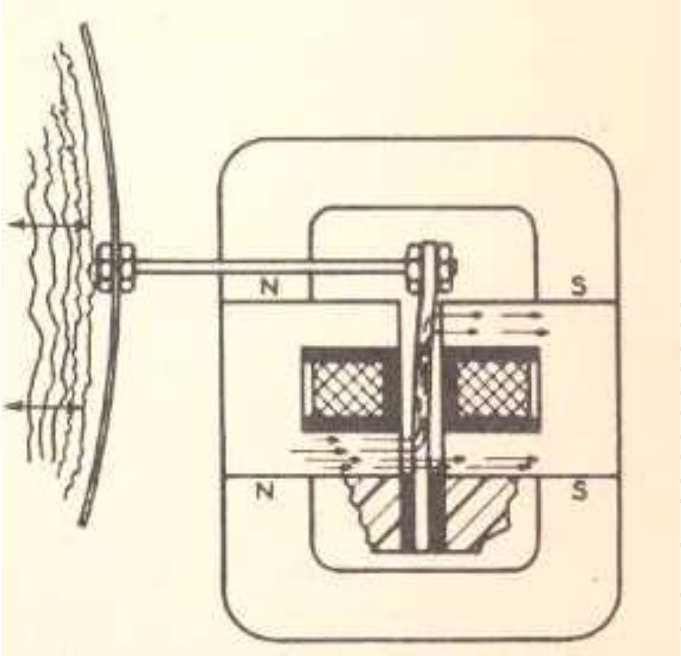

www.crystalradio.net

- An elaborate mechanism balanced armature system) give the sound powered headphones their sensitivity.
- Impedance typically around 1000 ohms, much too low for direct use in a crystal set. An impedance matching transformer is essential.
- Microphone transformers are excellent choices for matching the impedance of the diode to the headphones.

Final Results

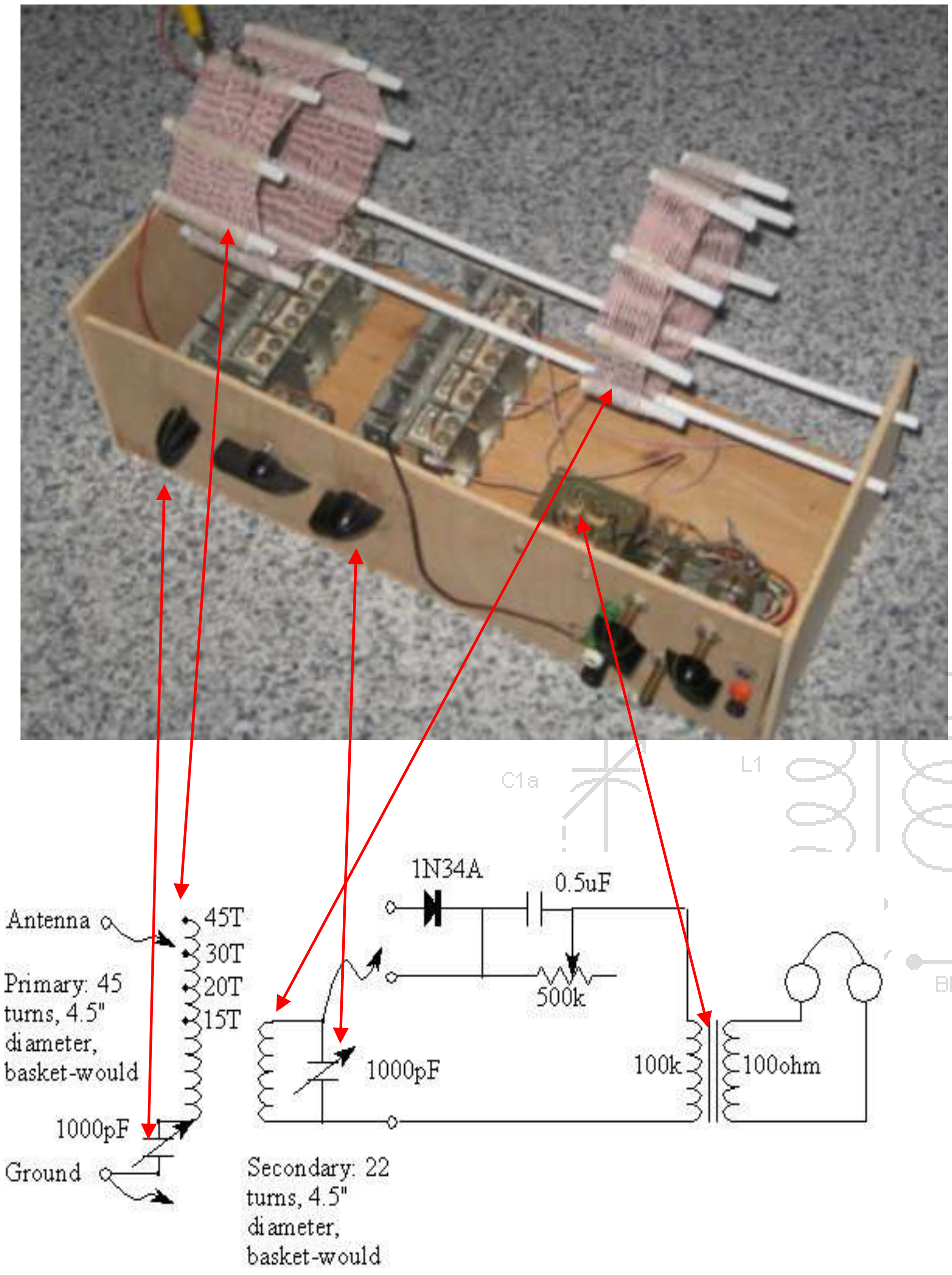
Antenna
45T
30T
20T
15T
Primary: 45 turns, 4.5" diameter, basket-would
1000pF
Ground
1N34A
0.5uF
500k
1000pF
100k
100ohm
Secondary: 22 turns, 4.5" diameter, basket-would
C1a
L1
L2
BROAD

- Litz wire basket wound with taps for experimentation.
- Series antenna tuner above 650kHz. Below 650kHz, capacitor has to be placed in parallel with the inductor
- Ceramic insulated variable capacitors for maximum Q
- R-C allows DC to build up, reduces detector loading and reduces distortion on local stations.

Clear Channel Stations

- Most of the distant stations we receive are 50,000 Watt “clear channel” stations.
- A clear channel station is a high power American station that shares its frequency with very few other stations
- The realization that a station nearly 1000 miles away can be heard in a completely passive crystal radio is amazing

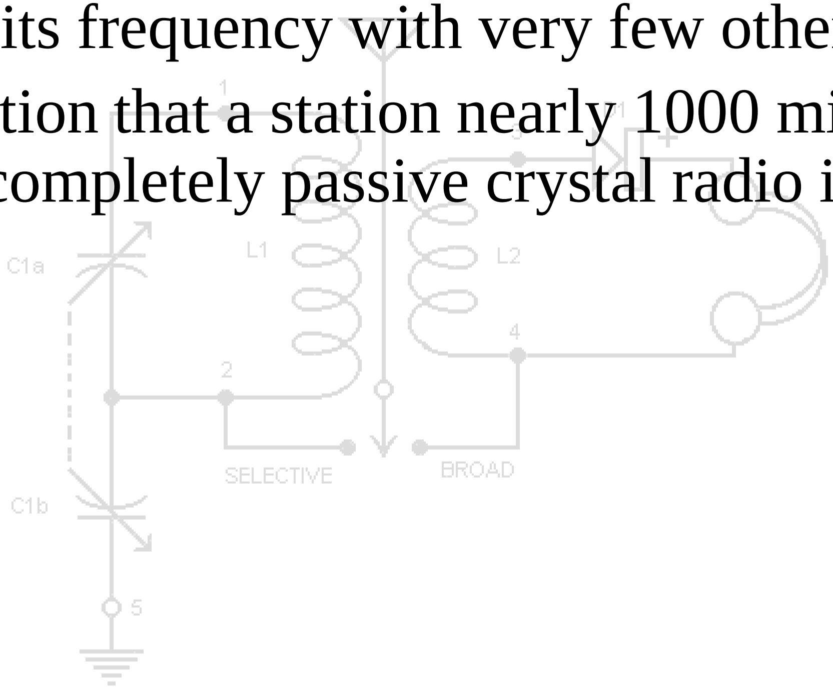

Clear Channel Stations

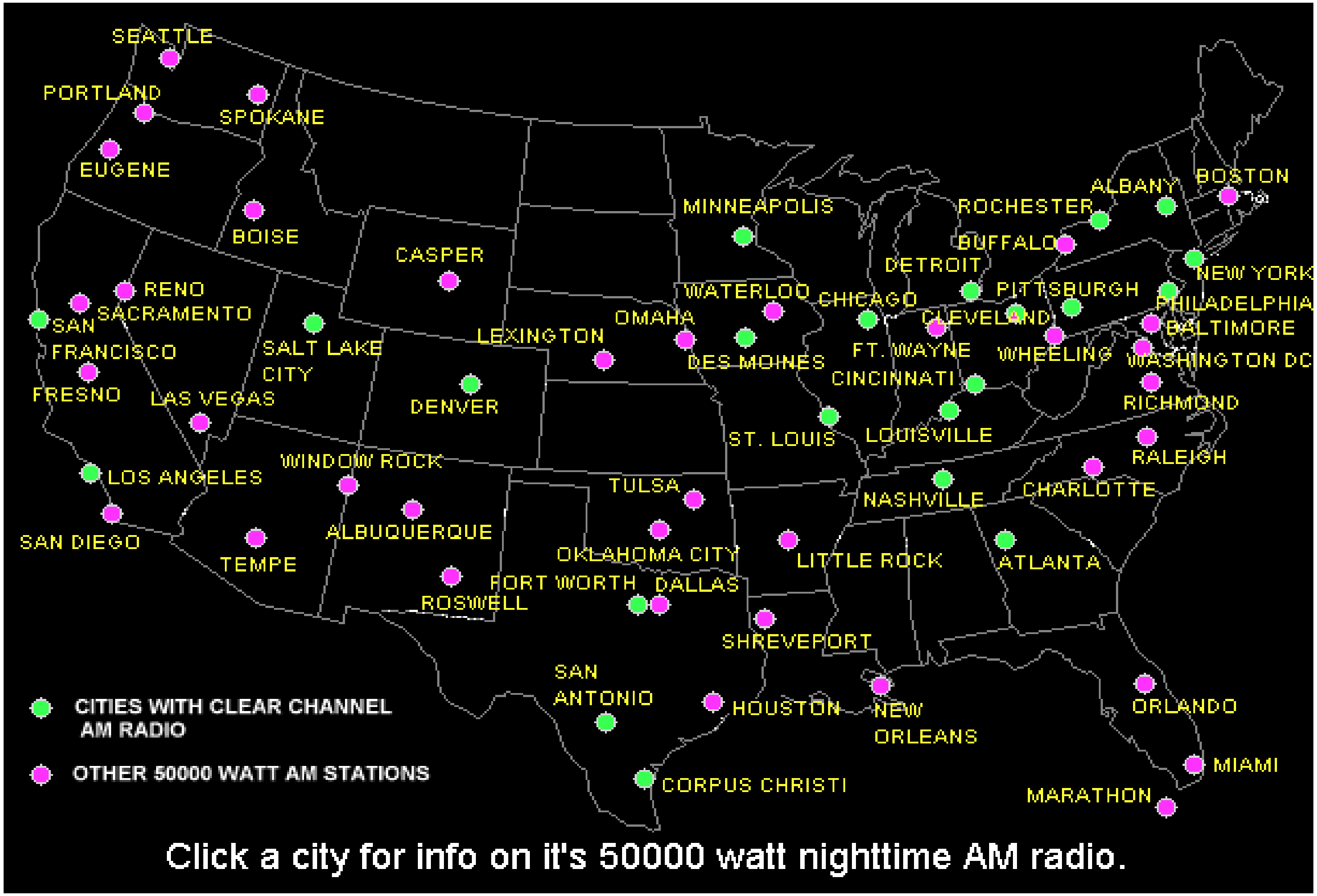
SEATTLE
PORTLAND
SPOKANE
EUGENE
BOISE
CASPER
MINNEAPOLIS
ROCHESTER
ALBANY
BOSTON
BUFFALO
DETROIT
NEW YORK
RENO
SACRAMENTO
SAN FRANCISCO
WATERLOO
CHICAGO
PITTSBURGH
PHILADELPHIA
OMAHA
CLEVELAND
BALTIMORE
SALT LAKE CITY
LEXINGTON
DES MOINES
FT. WAYNE
WHEELING
WASHINGTON DC
FRESNO
LAS VEGAS
DENVER
CINCINNATI
RICHMOND
ST. LOUIS
LOUISVILLE
RALEIGH
WINDOW ROCK
LOS ANGELES
TULSA
CHARLOTTE
NASHVILLE
SAN DIEGO
ALBUQUERQUE
OKLAHOMA CITY
TEMPE
LITTLE ROCK
ATLANTA
FORT WORTH
DALLAS
ROSWELL
SHREVEPORT
SAN ANTONIO
HOUSTON
NEW ORLEANS
ORLANDO
CORPUS CHRISTI
MIAMI
MARATHON
● CITIES WITH CLEAR CHANNEL AM RADIO
● OTHER 50000 WATT AM STATIONS
Click a city for info on it's 50000 watt nighttime AM radio.

- http://www.fortunecity.com/tinpan/bluenote/706/namrp/amradio.htm#anchor3

Stations Logged

- 640 CFYI, Toronto
- 660 WFAN, New York City
- 680 CFTR, Toronto
- 690, Montreal
- 700 WLW, Cincinati
- 720 WGN, Chicago
- 740, Toronto
- 760 WJR, Detroit
- 770 WABC, New York City
- 780 WBBM, Chicago
- 800 CJAD, Montreal
- 810 WGY, Schenectady
- 840 WHAS, Louisville
- 880 WCBS, New York City
- 920 WHJJ, Providence
- 940 CINW, Montreal
- 990 The Team, Montreal
- 1000 WMVP, Chicago
- 1010 CFRB, Toronto
- 1020 KDKA Pittsburgh
- 1030 WBZ, Boston
- 1060 KYW, Philadelphia
- 1080 WTIC, Hartford
- 1500 WTOP Washington
- 1520 WWKB, Buffalo
- 1560 WQEW, New York Citiy

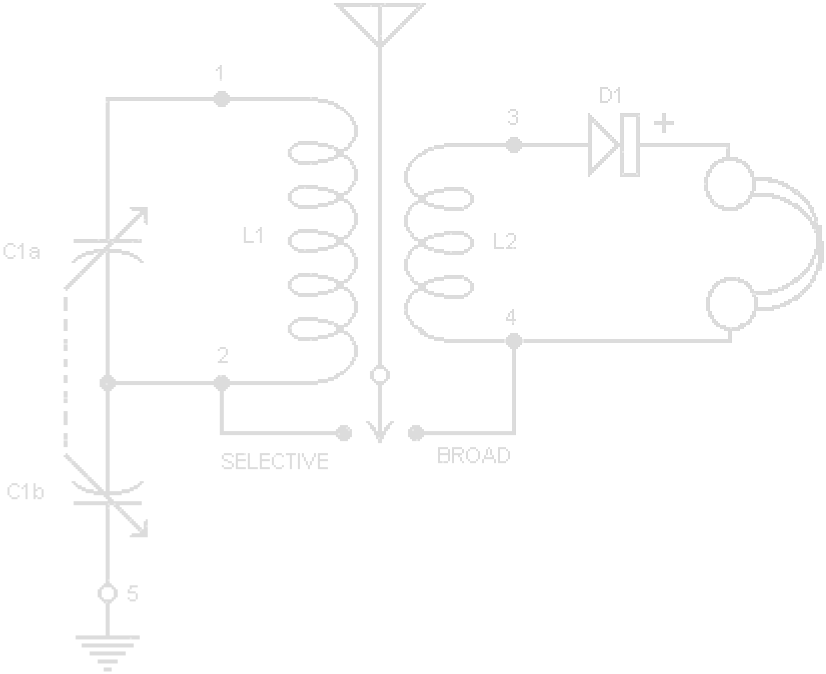

Top Performers

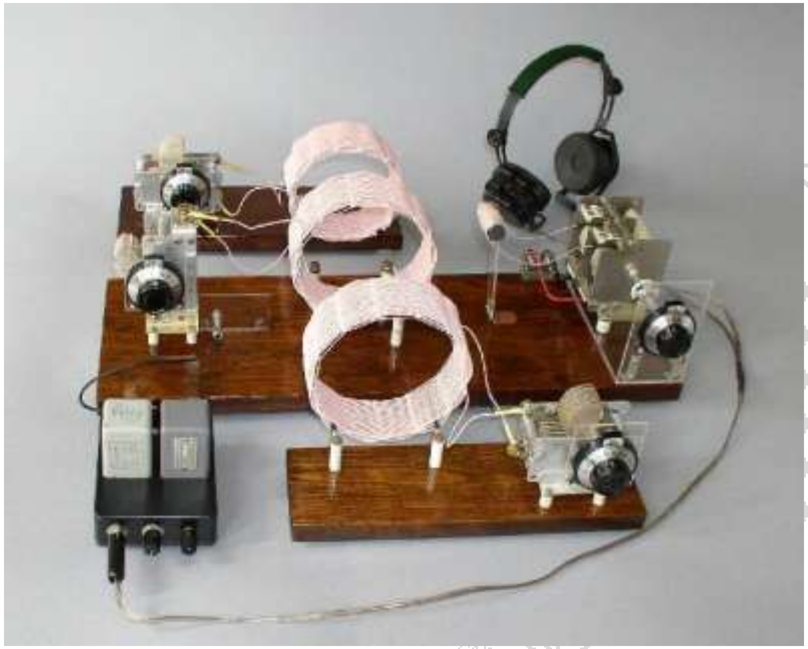

www.crystalradio.net

- Mike Tuggle of Hawaii is one of the top builders of crystal sets.
- His set, the Lyonodyne-17 has heard stations in Cuba from his home location in Hawaii!
- There are lots of web resources available. This presentation describes my efforts.

Conclusions

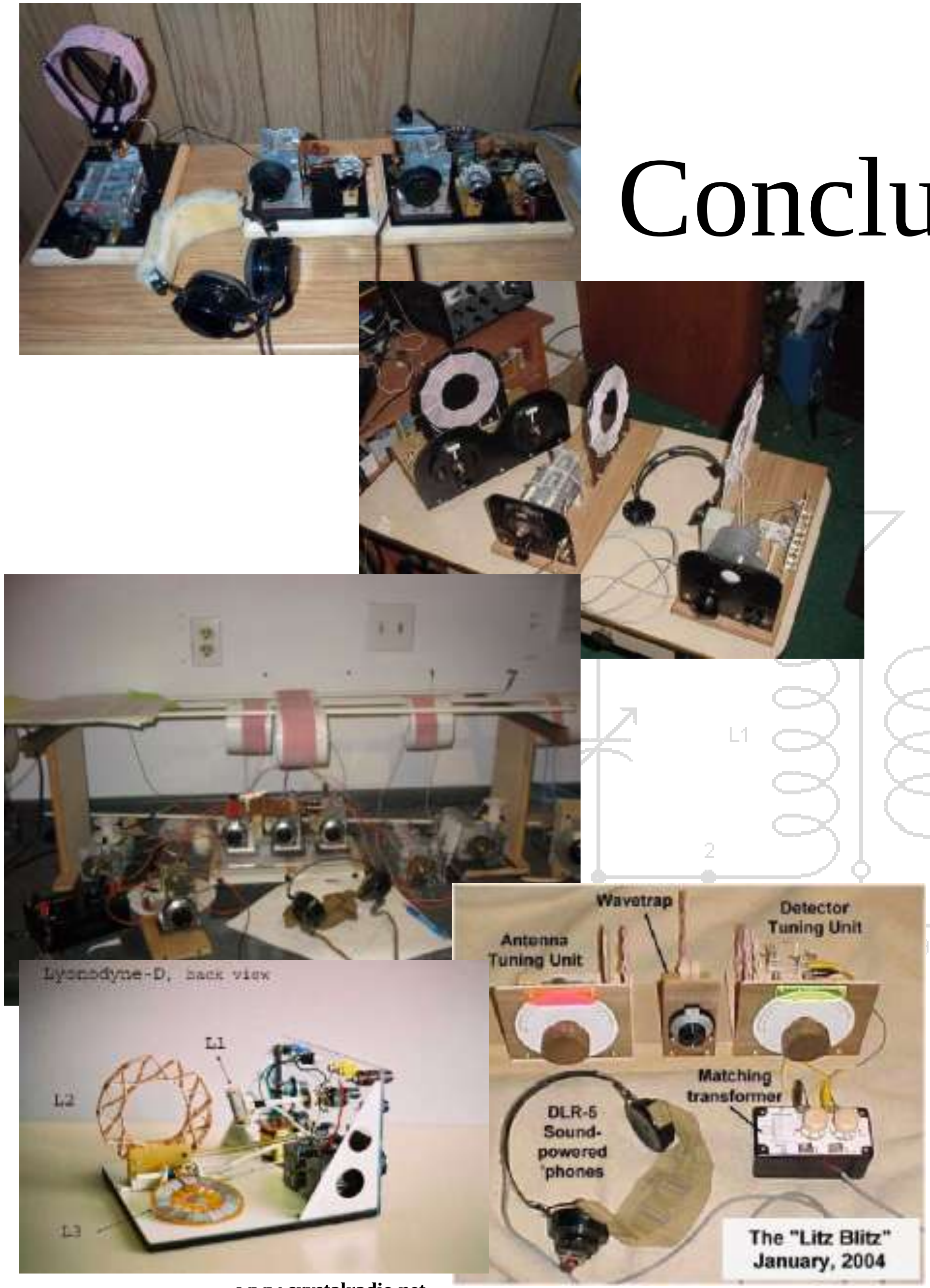
Lyonodyne-D, back view
L1
L2
L3
Wavetrap
Antenna Tuning Unit
Detector Tuning Unit
Matching transformer
DLR-5 Sound-powered 'phones
The "Litz Blitz" January, 2004

www.crystalradio.net

- Crystal sets appear to be simple, but attention must be paid to all the details.
- If you truly understand a crystal radio, you have a good foundation to RF engineering in general.

High Performance Crystal Radios

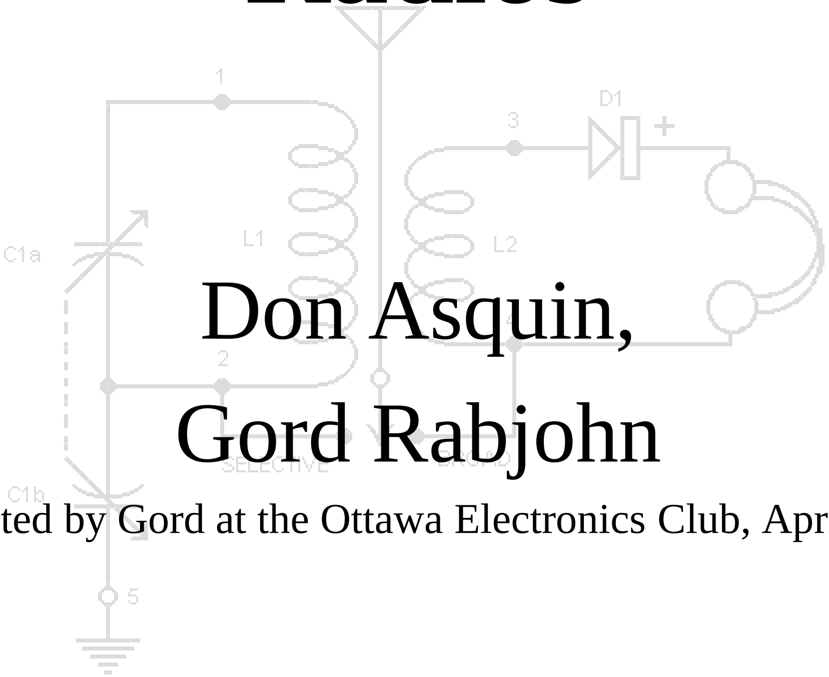
1
D1
3
+
C1a
L1
L2
2
SELECTIVE
C1b
5

Don Asquin,

Gord Rabjohn

(Presented by Gord at the Ottawa Electronics Club, April 2012)

Childhood History

www.crystalradio.net

www.crystalradio.net

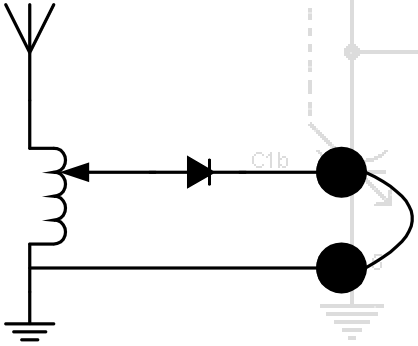

- Like many kids, I had frustrating childhood experiences with crystal radios!
- Simple crystal sets couldn't separate multiple AM radio stations.
- Only local stations were received, and everything was received all at once.

Rediscovering History

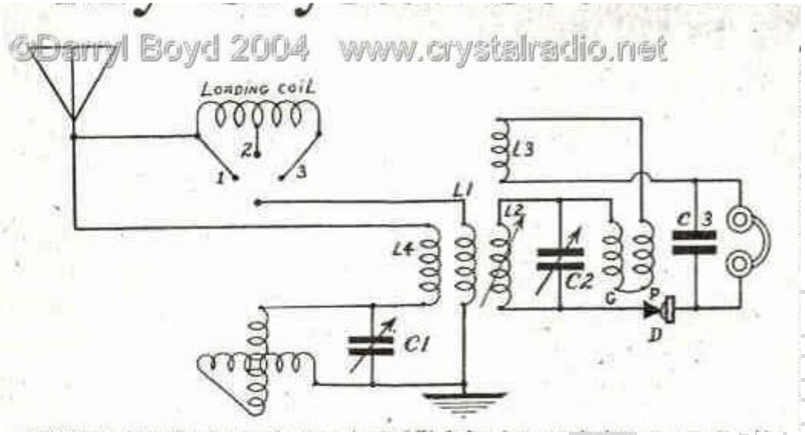
©Darryl Boyd 2004 www.crystalradio.net
LOADING COIL
1
2
3
L1
L2
L3
L4
C1
C2
C3
G
P
D

www.crystalradio.net

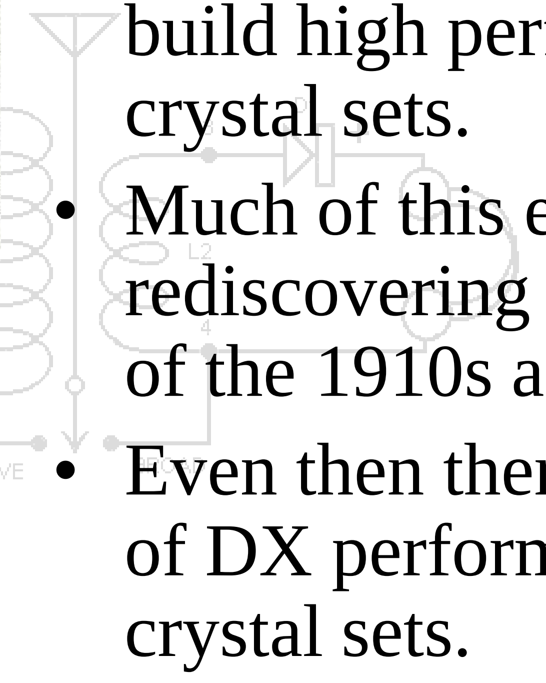
C1a
C1b
2
4
5
L2
SELECTIVE

- Recently, my friend Don Asquin and I decided to build high performance crystal sets.
- Much of this exploration is rediscovering the knowledge of the 1910s and early 1920s.
- Even then there were claims of DX performance with crystal sets.

How does a basic crystal set work?

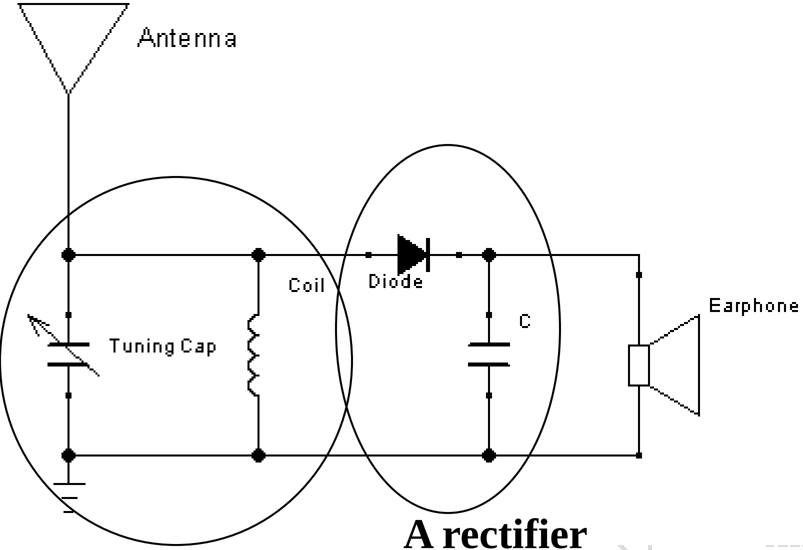
Antenna
Coil
Diode
Tuning Cap
C
Earphone

A rectifier

A filter

At low frequency, the inductor shorts the signal out. At high frequency the capacitor shorts the signal out. In-between, there is a magic point called "resonance"

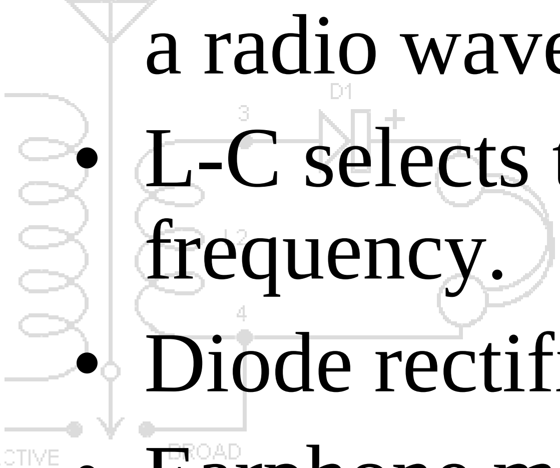

- Antenna couples to the (usually) electric field in a radio wave.
- L-C selects the frequency.
- Diode rectifies it.
- Earphone makes it audible

AM modulated signal

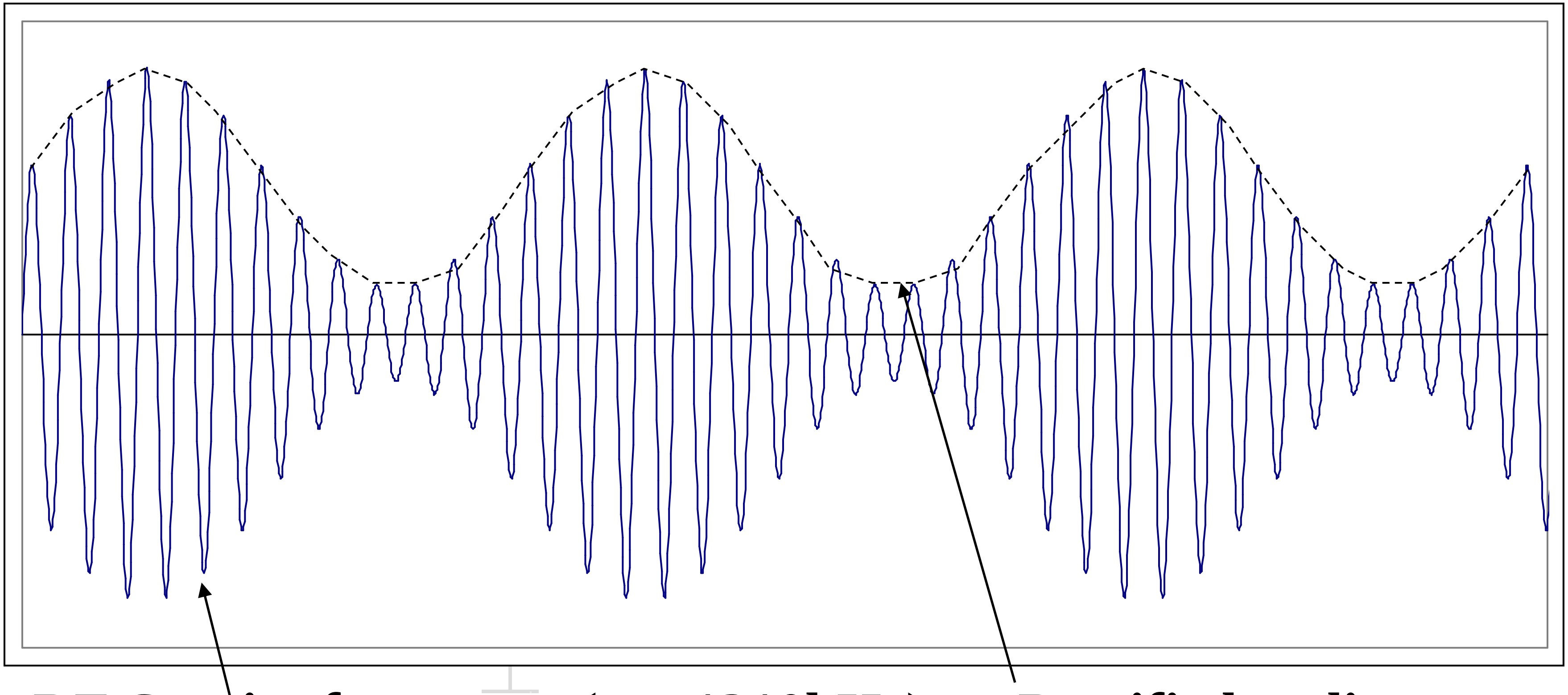

RF Carrier frequency (say 1310kHz): the filter selects only this frequency.

Rectified audio output, smoothed by the capacitor

The Four Key Elements

ANTENNA
Federal Jr.
www.crystalradio.net

www.crystalradio.net

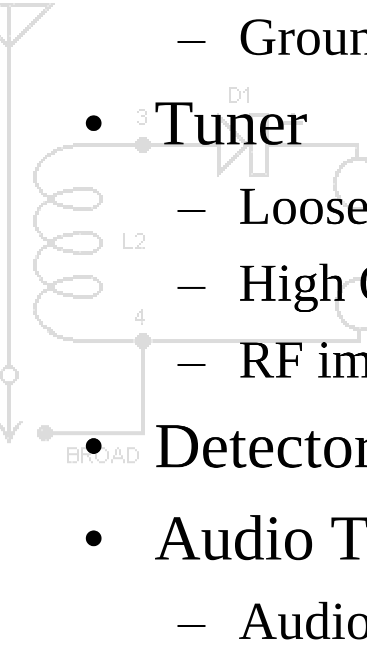
D1
3
L2
4
SELECTIVE
BROAD
C1b
5

- Antenna
  - Antenna
  - Ground
- Tuner
  - Loose coupler
  - High Q coils
  - RF impedance matching
- Detector
- Audio Transducer
  - Audio Impedance matching

Antenna

- A long exterior antenna is crucial.
- Why? It’s not just about area… “watts per acre”. It’s also about giving the antenna a better “radiation impedance”.
  - The energy from an antenna, can be modeled as a voltage source in series with a source resistance (or a current source in parallel with a source conductance), and the antenna capacitance (or inductance).
  - Radiation resistance is the resistance of this apparent source of energy in an antenna.
  - The tuned circuit matches this impedance to the load (detector) for optimum energy transfer.
  - A small antenna will tend to look like a “difficult to match” impedance….

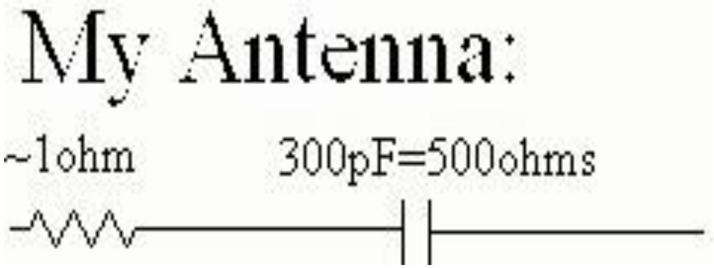

⏚ Ground ⏚

- The ground provides the completion of the current path for the radio signals.
- You can always attach a wire to a cold water pipe, but the best ground remains a copper plated rod pounded into the ground.
- Home Depot sells these rods for grounding electrical systems.
- If installed near a water tap, the ground can be kept moist and conductive.

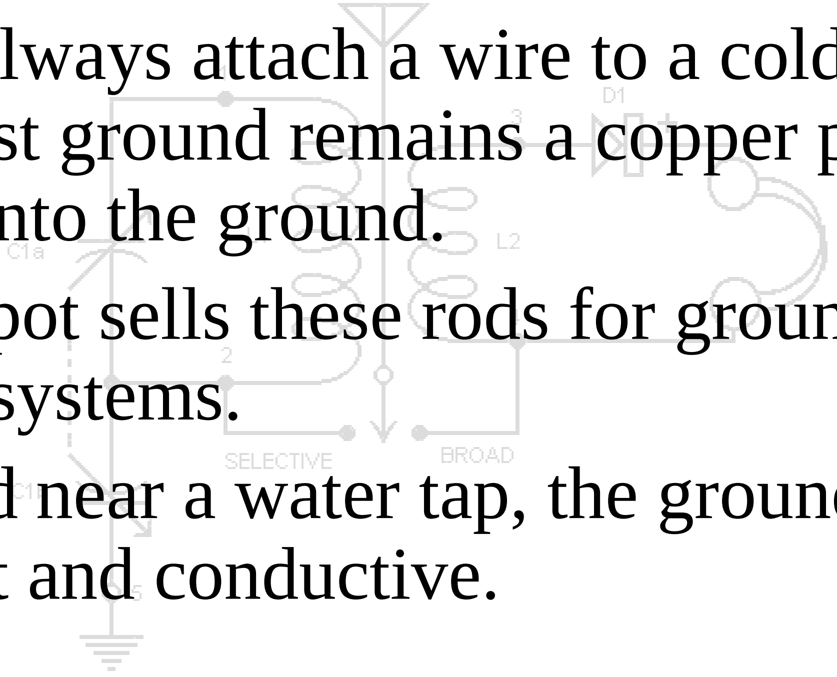

Tuner
-Antenna Tuner

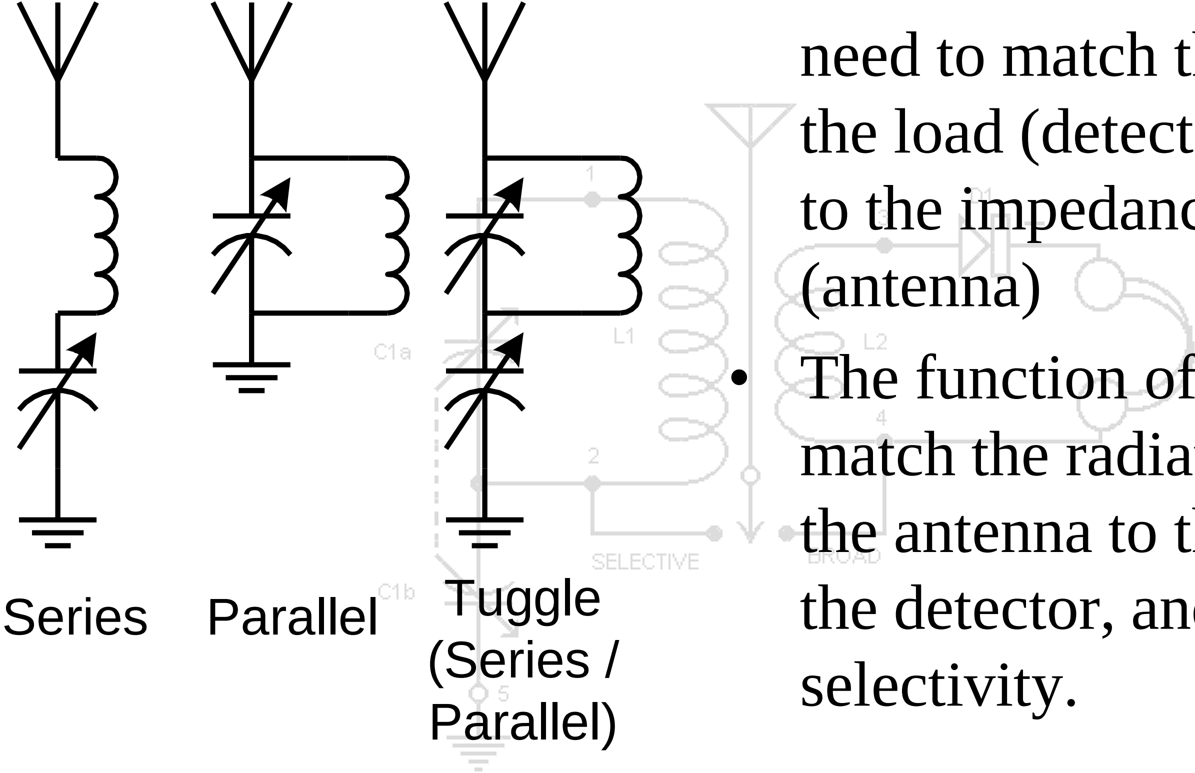
1
C1a
L1
2
SELECTIVE
C1b
3
D1
L2
4
BROAD
5

Series

Parallel

Tuggle (Series / Parallel)

- To extract the most power, you need to match the resistance of the load (detector and earphone) to the impedance of the source (antenna)
- The function of the tuner is to match the radiation resistance of the antenna to the impedance of the detector, and to provide selectivity.

Tuner
-Loose Coupler

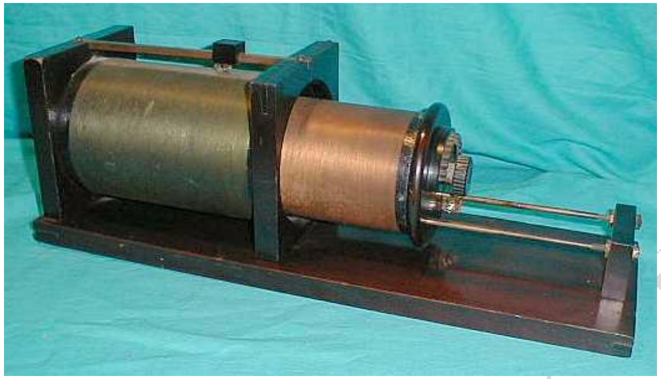

www.crystalradio.net

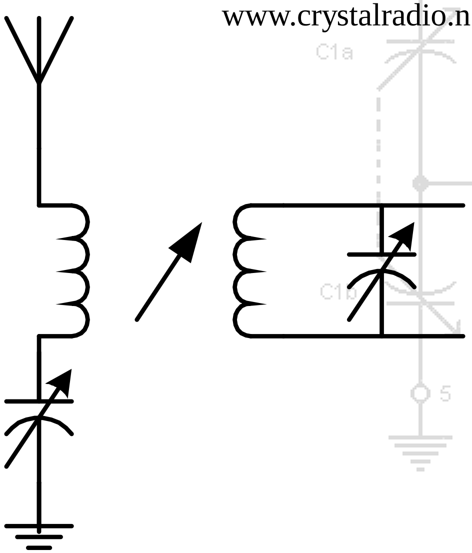

- The loose coupler is simply two coils (each generally resonated with a capacitor) lightly coupled to each other.
- One coil is connected to the antenna, the other is connected to the detector.
- Even cheap capacitors tend to be quite good, but high Q inductor fabrication is an “art”.

Loose Coupler

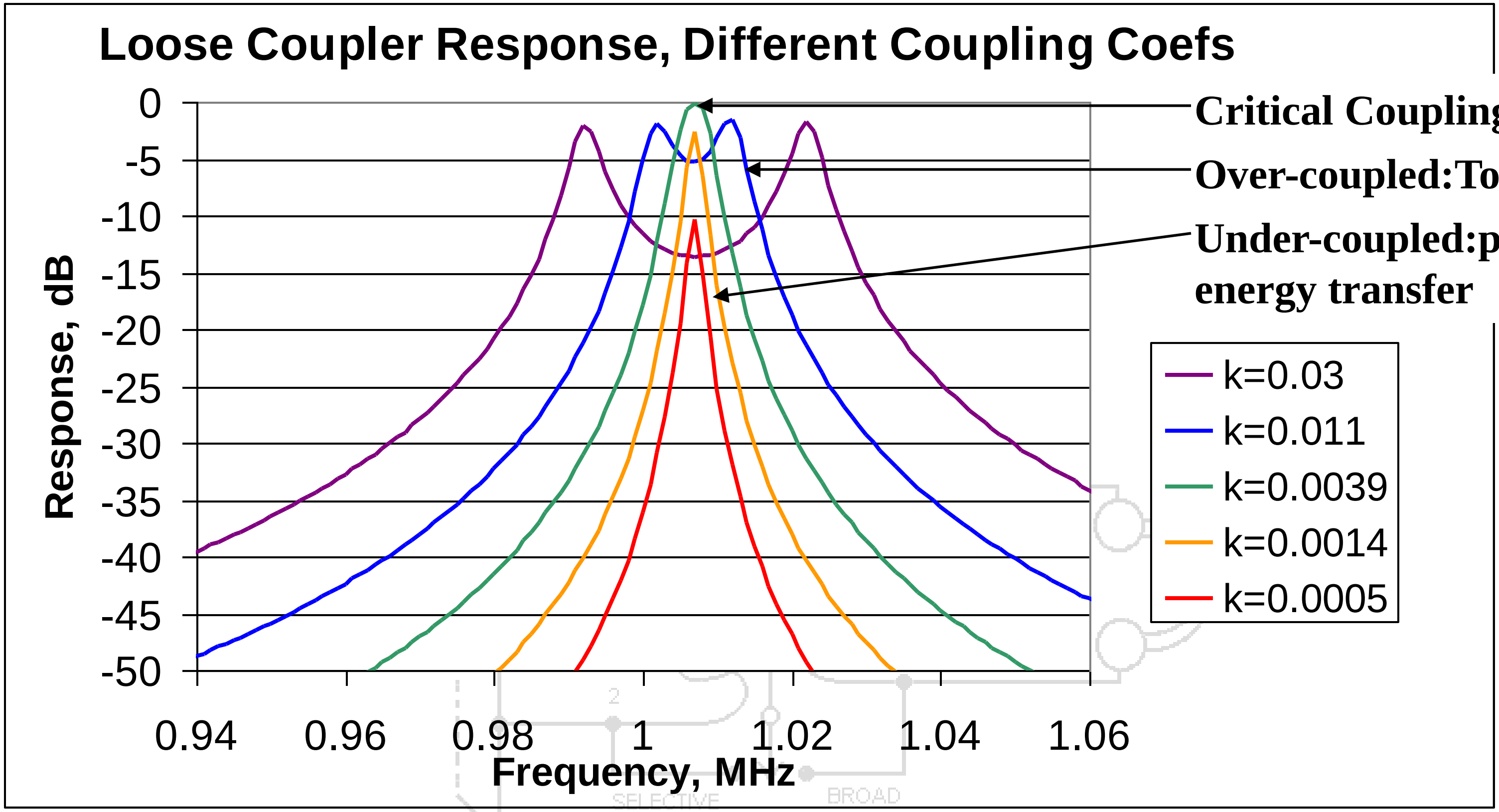

Critical Coupling

Over-coupled:Too broad

Under-coupled:poor energy transfer

Simulation of 1000pF, 1.0 ohm antenna, 100uH primary coil and ~330pF cap, coupled to a 100uH secondary coil, 250pF cap, 100k load.

- There is an optimum amount of coupling between the coils. (=1/Q). This is VERY LIGHT coupling. Under 1% of the magnetic field is coupled.

Coil “Q” = Quality

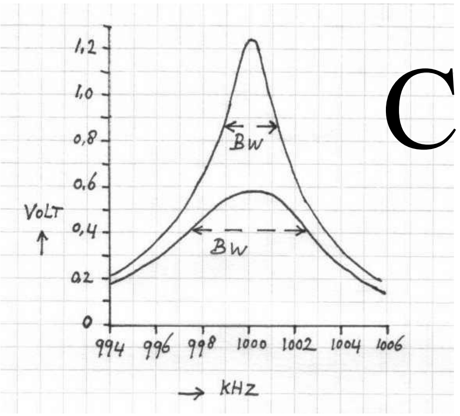

www.crystal-radio.eu

$$Q = \frac{\text{Frequency}}{\text{Bandwidth}}$$

Station Frequency (kHz)	Minimum Q
540	54
1000	100
1600	160

Assume Q=160

Station Frequency (kHz)	Audio Bandwidth (kHz)
540	1.7
1000	3.1
1600	5.0

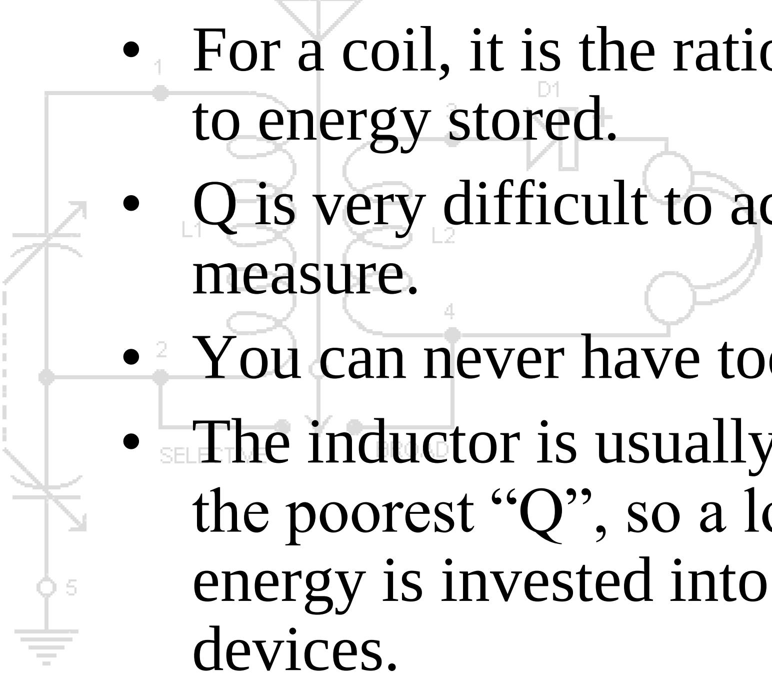

- For a tuned circuit “Q” is the ratio of the center frequency to the bandwidth.
- For a coil, it is the ratio of energy lost to energy stored.
- Q is very difficult to accurately measure.
- You can never have too much Q!
- The inductor is usually the part with the poorest “Q”, so a lot of creative energy is invested into optimum devices.

Coil Types

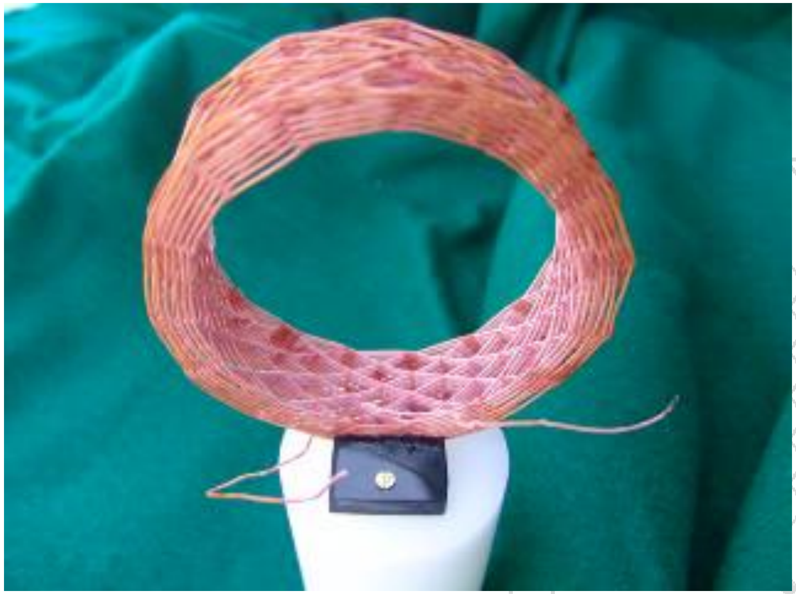

www.wynterarchtops.com

- The highest “Q” coils seem to be air core.
- Many different variations
  - Spider web,
  - Basket weave
  - Honey Comb
- All attempt to reduce capacitance and current crowding to increase Q

Coil Winding Jig

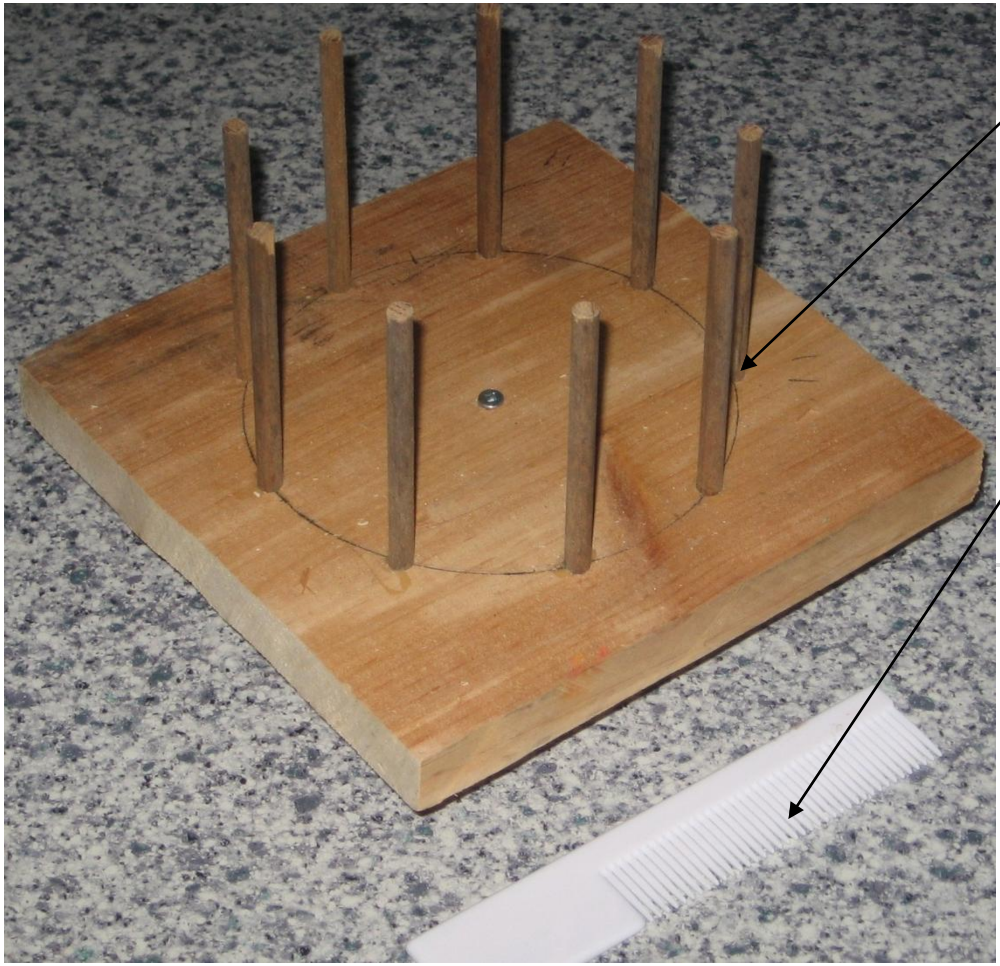

Drinking straws fit over ¼” dowels.

Comb helps space wires

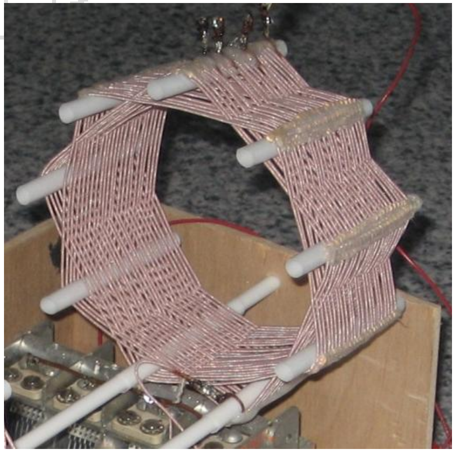

Litz Wire

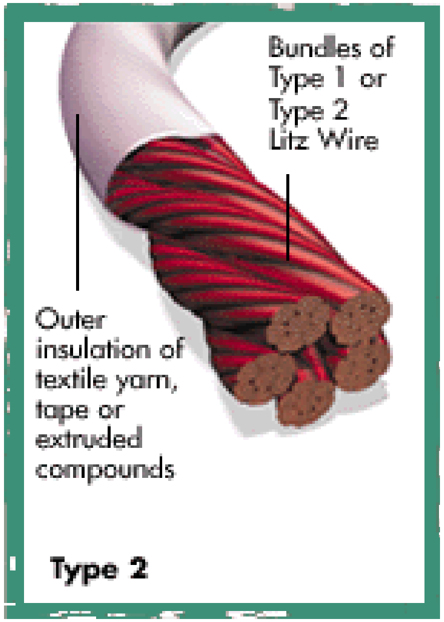
Bundles of Type 1 or Type 2 Litz Wire
Outer insulation of textile yarn, tape or extruded compounds
Type 2

www.newenglandwire.com

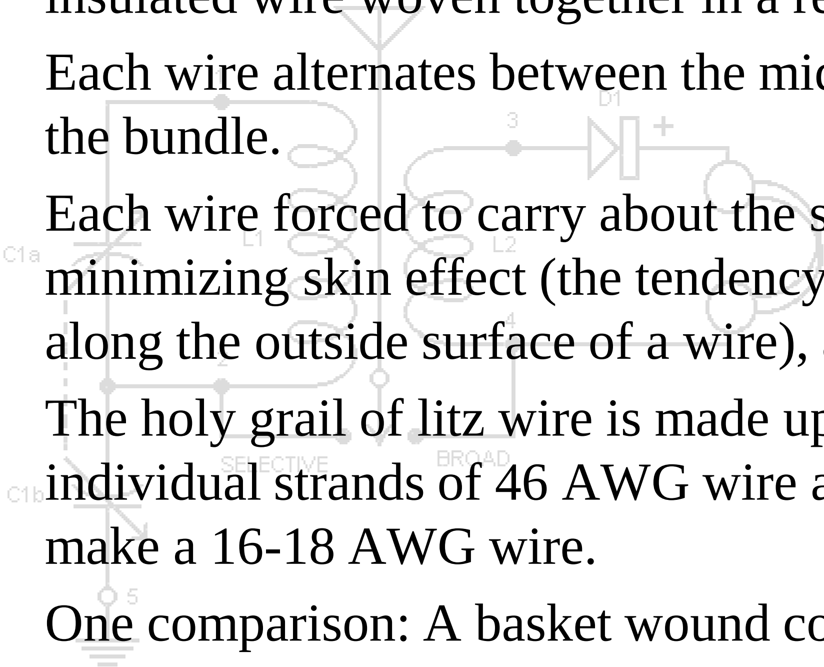
D1
3
+
C1a
L1
L2
4
SELECTIVE
BROAD
C1b
5

- Best wire for high “Q” coils is Litz wire.
- Litz is derived from the German word “Litzendraht” meaning woven wire.
- Consists of many strands of parallel connected, individually insulated wire woven together in a regular pattern.
- Each wire alternates between the middle and the outside of the bundle.
- Each wire forced to carry about the same current, minimizing skin effect (the tendency for current to flow along the outside surface of a wire), and loss.
- The holy grail of litz wire is made up of 420 to 660 individual strands of 46 AWG wire all twisted together to make a 16-18 AWG wire.
- One comparison: A basket wound coil with solid copper wire (~200uH) has a Q of 230 at 1MHz. With Litz wire, it has a Q of over 500!

www.crystalradio.net

Detector

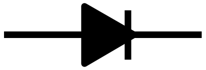

1N34A

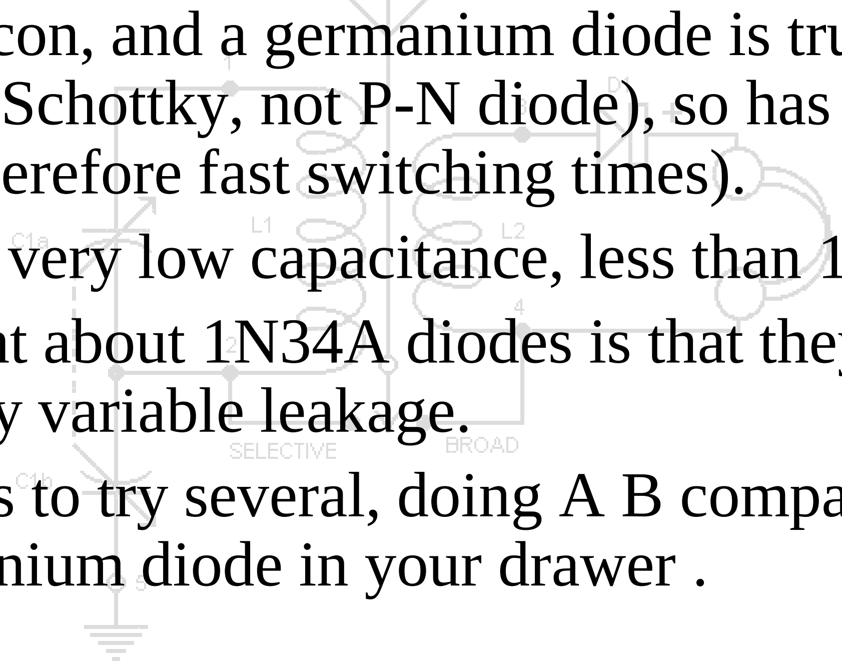

- In my experience, nothing beats a good 1N34A germanium diode. (Still widely available)
- Germanium is good because its barrier height (turn-on voltage) is lower than silicon, and a germanium diode is truly a point-contact diode (Schottky, not P-N diode), so has low charge storage (and therefore fast switching times).
- Specified with very low capacitance, less than 1.0pF.
- Main complaint about 1N34A diodes is that they tend to have high and highly variable leakage.
- Best solution is to try several, doing A B comparisons and select the best germanium diode in your drawer .

Detector Impedance

- Detectors are most efficient when driven with *high voltages*.
  - Detectors are "square law" devices (at low power), output voltage is proportional to square of input voltage.
- Since we have a fixed amount of input *power*, we need to increase voltage and decrease current applied to the detector (Power=Voltage x Current); in other words increase the detector *impedance*.
- At detector input: use a loose coupler.
- At detector output: use a 100kohm microphone transformer to match into the headphones.

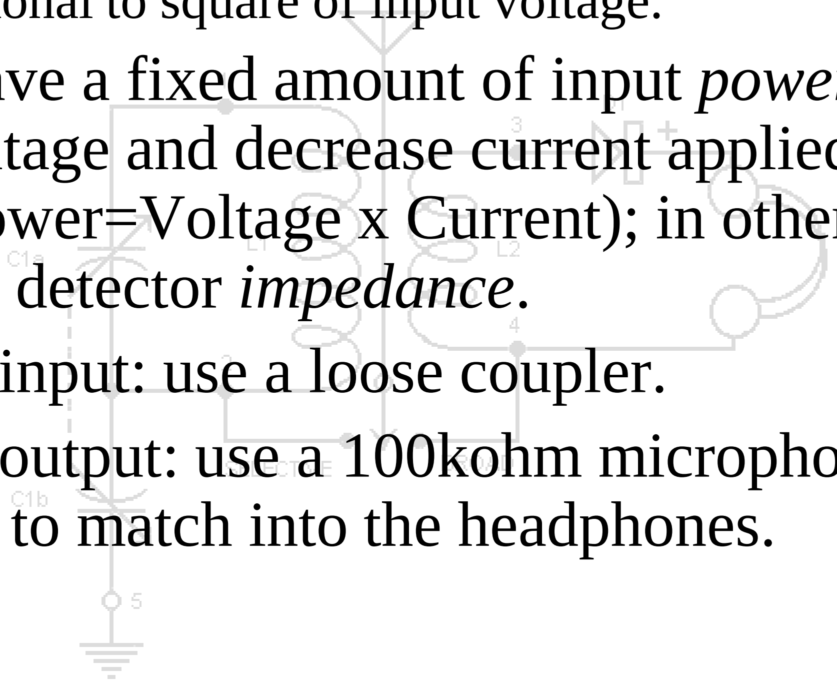

Germanium Diodes

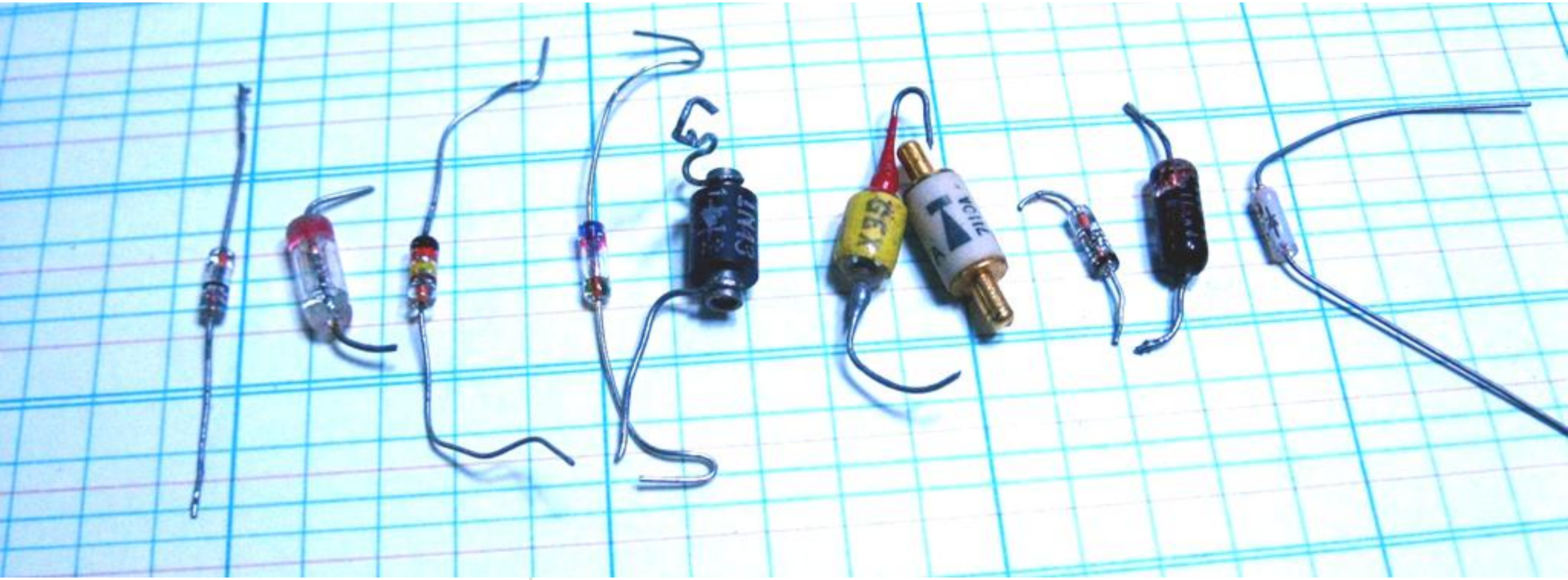

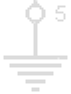

Detectors: I-V Curves

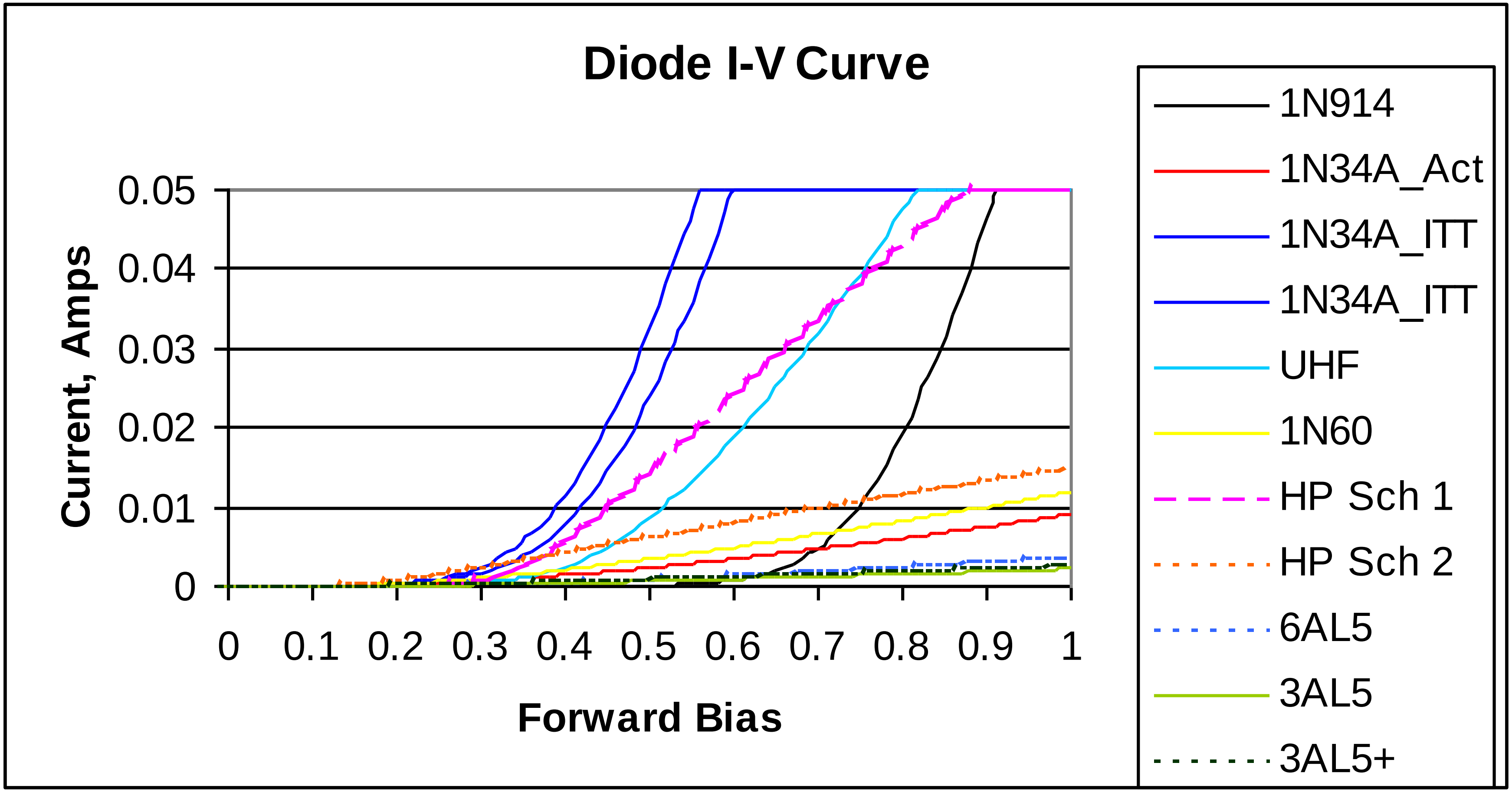

Detectors: I-V Curves

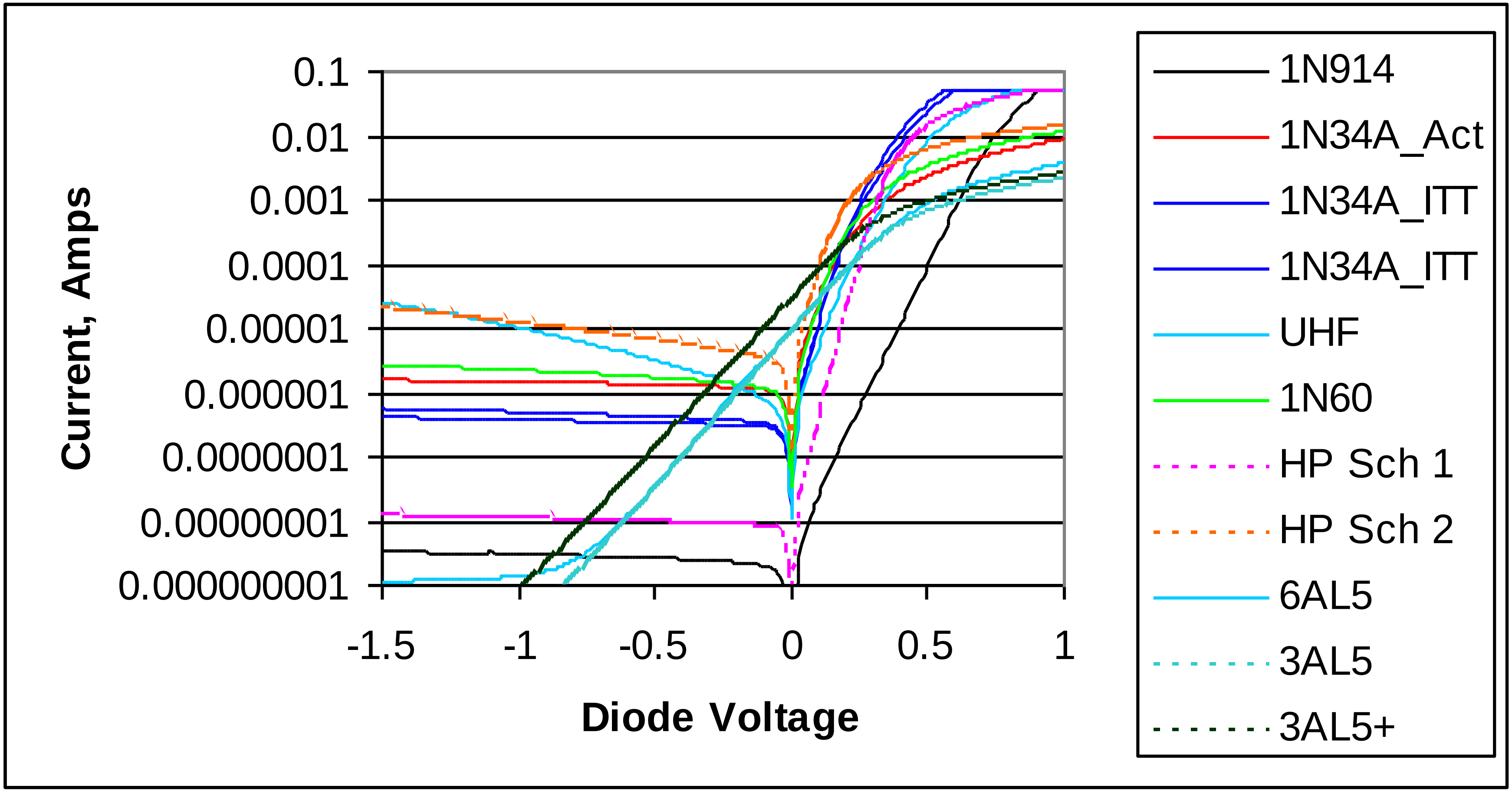

Mineral Detectors

Lead Sulfide, PbS (Galena), Iron Pyrite, Zincite, and other minerals have been used to make detectors (surprisingly good).

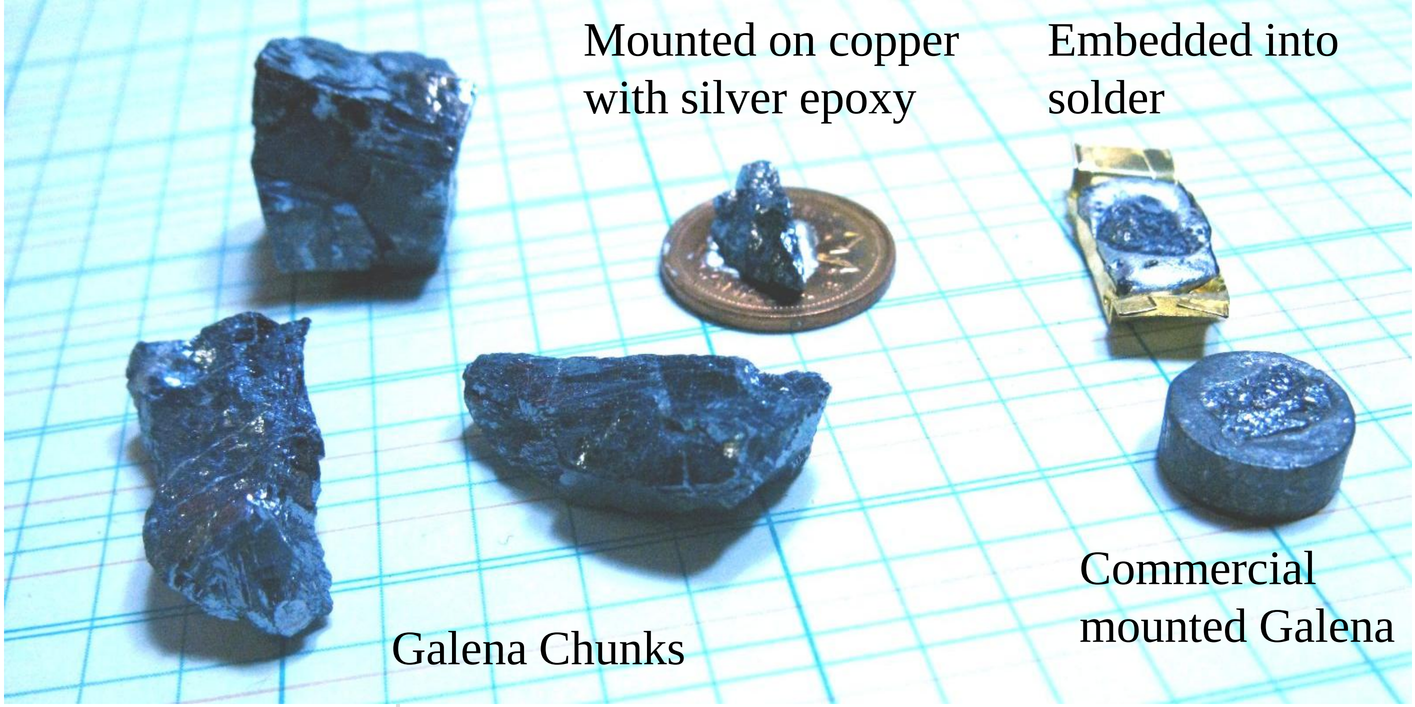
Mounted on copper with silver epoxy
Embedded into solder
Commercial mounted Galena
Galena Chunks

Galena Detector

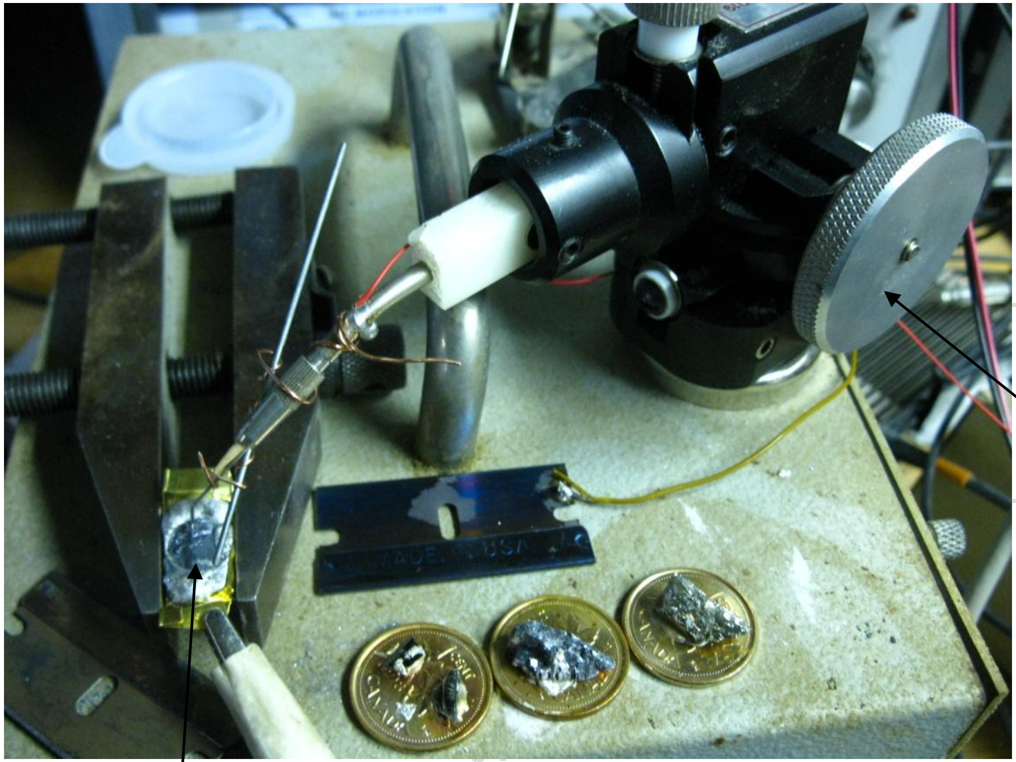

X-Y-Z manipulator

Galena mounted in solder

Audio Transducer

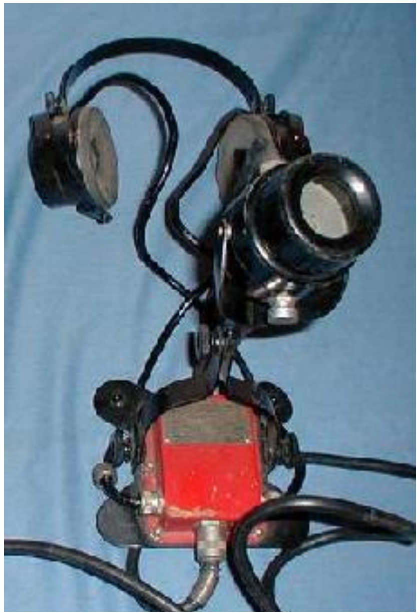

www.crystalradio.net

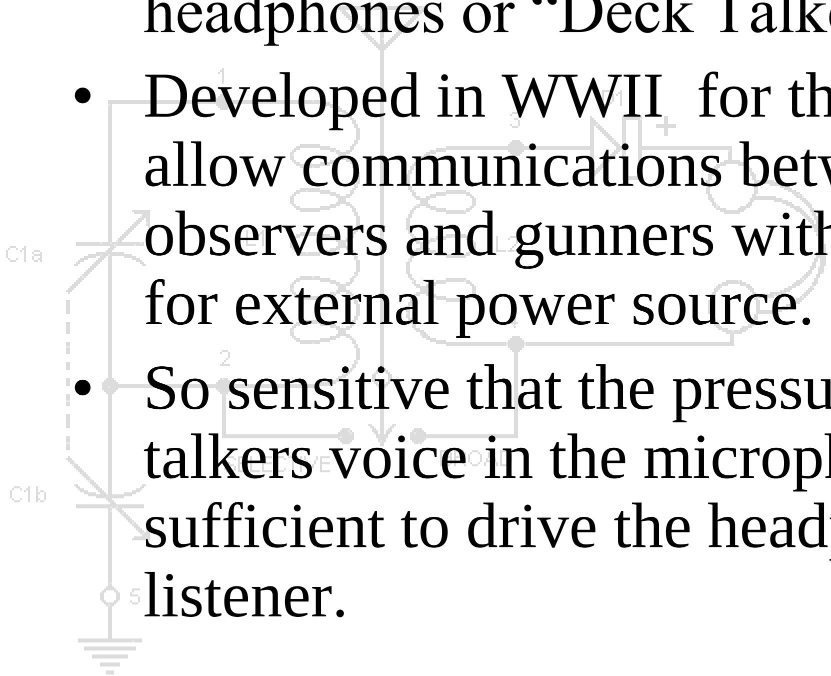

- Best transducers (headphones) for crystal radio operation are “Sound Powered” headphones or “Deck Talkers”.
- Developed in WWII for the navy to allow communications between the observers and gunners without the need for external power source.
- So sensitive that the pressure of the talkers voice in the microphone is sufficient to drive the headphones of the listener.

Audio Transducer

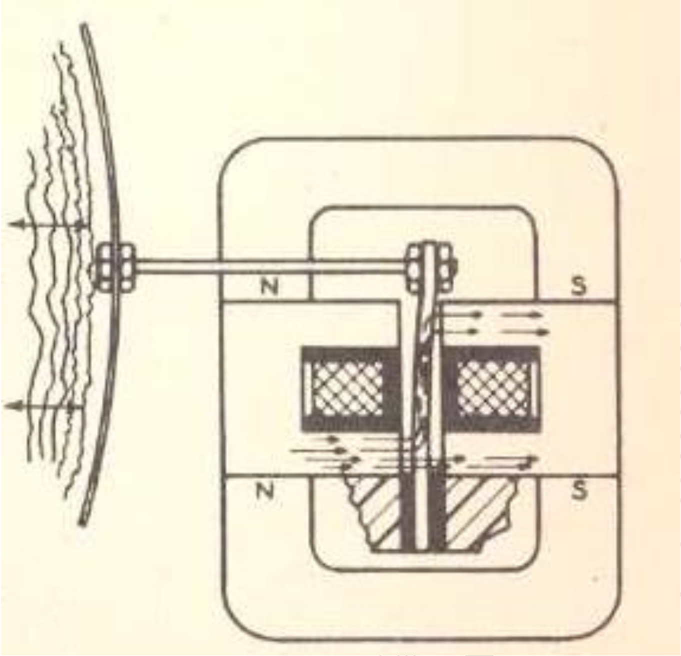

www.crystalradio.net

- An elaborate mechanism balanced armature system) give the sound powered headphones their sensitivity.
- Impedance typically around 1000 ohms, much too low for direct use in a crystal set. An impedance matching transformer is essential.
- Microphone transformers are excellent choices for matching the impedance of the diode to the headphones.

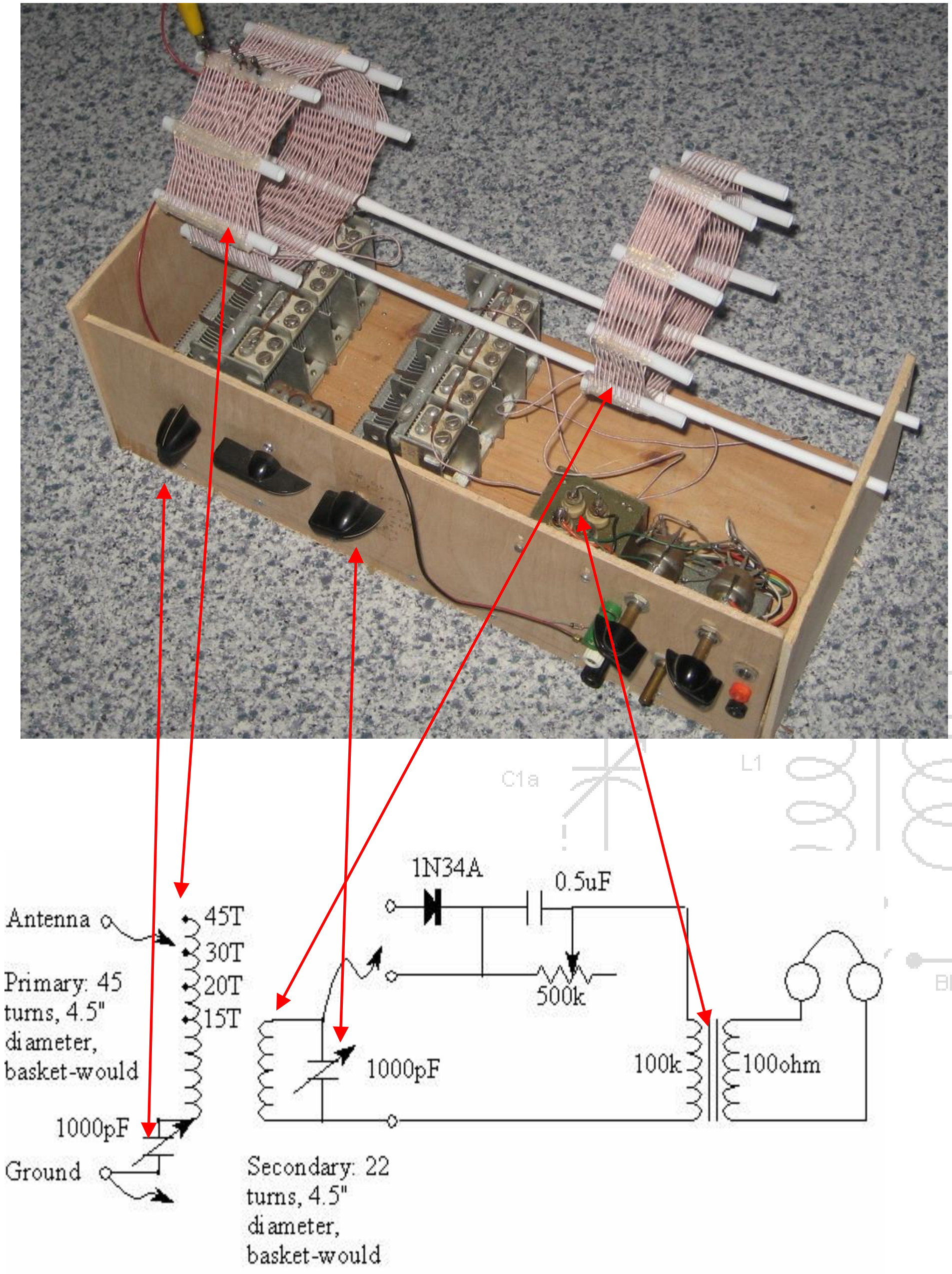
Antenna
45T
30T
20T
15T
Primary: 45 turns, 4.5" diameter, basket-would
1000pF
Ground
Secondary: 22 turns, 4.5" diameter, basket-would
1000pF
1N34A
0.5uF
500k
100k
100ohm
C1a
L1

Final Results

- Litz wire basket wound with taps for experimentation.
- Series antenna tuner above 650kHz. Below 650kHz, capacitor has to be placed in parallel with the inductor
- Ceramic insulated variable capacitors for maximum Q
- R-C allows DC to build up, reduces detector loading and reduces distortion on local stations.

Clear Channel Stations

- Most of the distant stations we receive are 50,000 Watt "clear channel" stations.
- A clear channel station is a high power American station that shares its frequency with very few other stations
- The realization that a station nearly 1000 miles away can be heard in a completely passive crystal radio is amazing

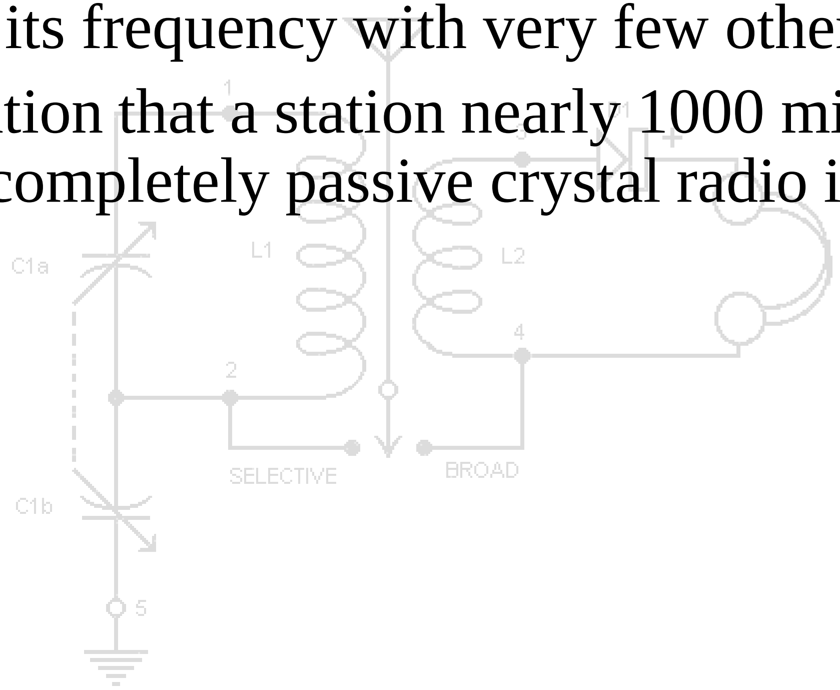

Clear Channel Stations

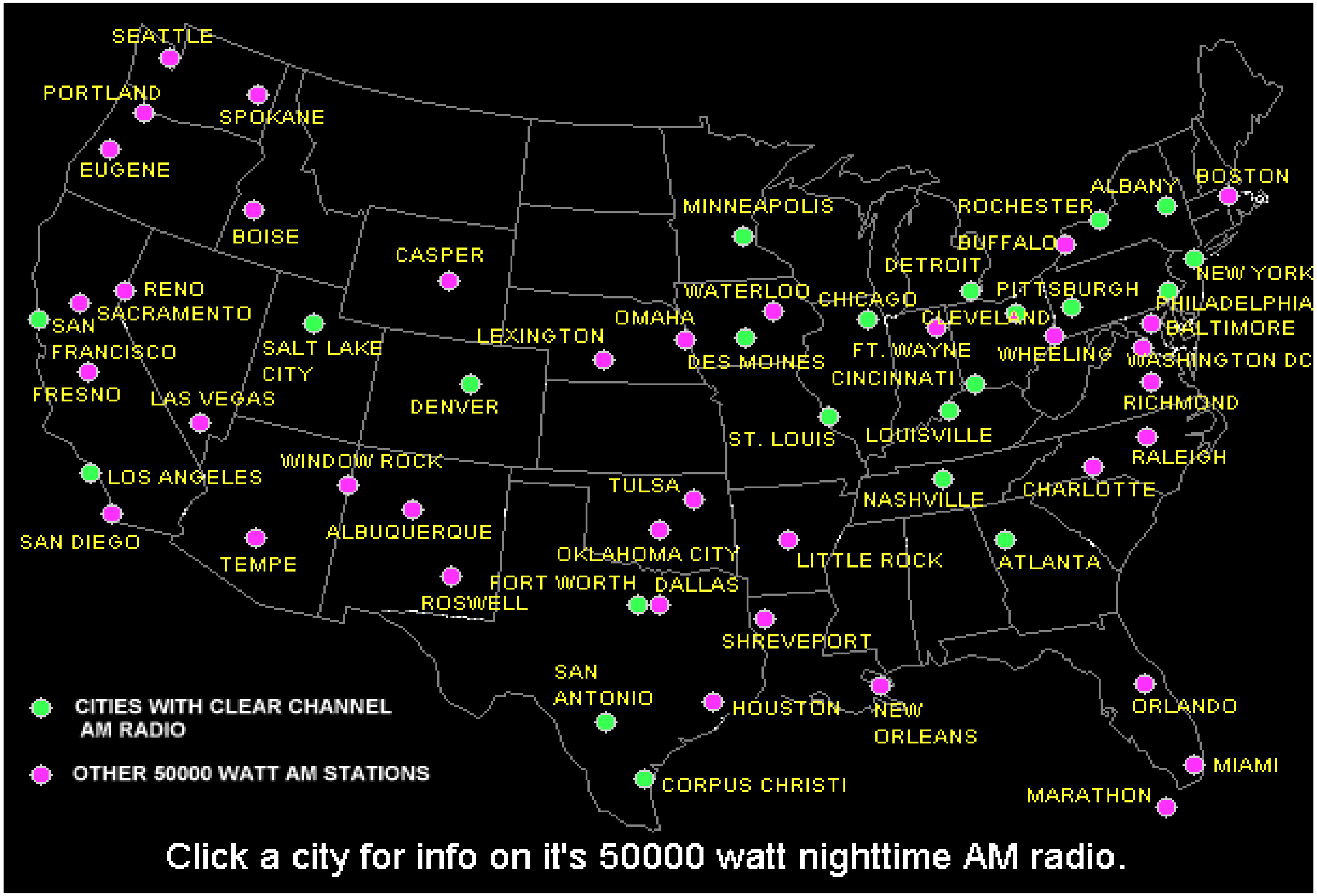
SEATTLE
PORTLAND
SPOKANE
EUGENE
BOISE
CASPER
MINNEAPOLIS
ROCHESTER
ALBANY
BOSTON
BUFFALO
DETROIT
NEW YORK
RENO
SACRAMENTO
SAN FRANCISCO
SALT LAKE CITY
LEXINGTON
OMAHA
WATERLOO
CHICAGO
PITTSBURGH
PHILADELPHIA
CLEVELAND
BALTIMORE
FT. WAYNE
WHEELING
WASHINGTON DC
DES MOINES
CINCINNATI
FRESNO
LAS VEGAS
DENVER
RICHMOND
ST. LOUIS
LOUISVILLE
WINDOW ROCK
LOS ANGELES
TULSA
RALEIGH
NASHVILLE
CHARLOTTE
SAN DIEGO
TEMPE
ALBUQUERQUE
OKLAHOMA CITY
LITTLE ROCK
ATLANTA
FORT WORTH
DALLAS
ROSWELL
SHREVEPORT
SAN ANTONIO
HOUSTON
NEW ORLEANS
ORLANDO
CORPUS CHRISTI
MIAMI
MARATHON
● CITIES WITH CLEAR CHANNEL AM RADIO
● OTHER 50000 WATT AM STATIONS
Click a city for info on it's 50000 watt nighttime AM radio.

- http://www.fortunecity.com/tinpan/bluenote/706/namrp/amradio.htm#anchor3

Stations Logged

- 640 CFYI, Toronto
- 660 WFAN, New York City
- 680 CFTR, Toronto
- 690, Montreal
- 700 WLW, Cincinati
- 720 WGN, Chicago
- 740, Toronto
- 760 WJR, Detroit
- 770 WABC, New York City
- 780 WBBM, Chicago
- 800 CJAD, Montreal
- 810 WGY, Schenectady
- 840 WHAS, Louisville
- 880 WCBS, New York City
- 920 WHJJ, Providence
- 940 CINW, Montreal
- 990 The Team, Montreal
- 1000 WMVP, Chicago
- 1010 CFRB, Toronto
- 1020 KDKA Pittsburgh
- 1030 WBZ, Boston
- 1060 KYW, Philadelphia
- 1080 WTIC, Hartford
- 1500 WTOP Washington
- 1520 WWKB, Buffalo
- 1560 WQEW, New York Citiy

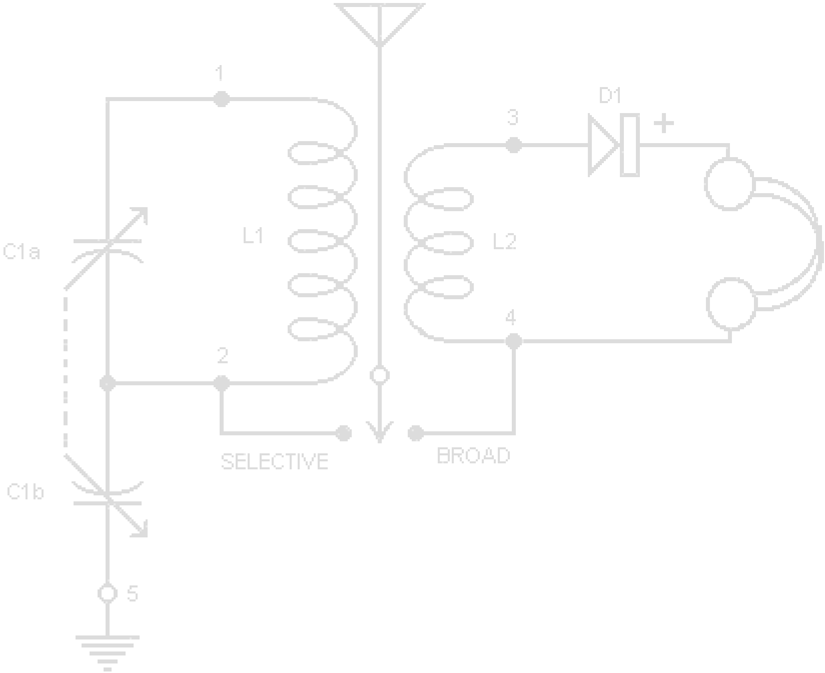

Top Performers

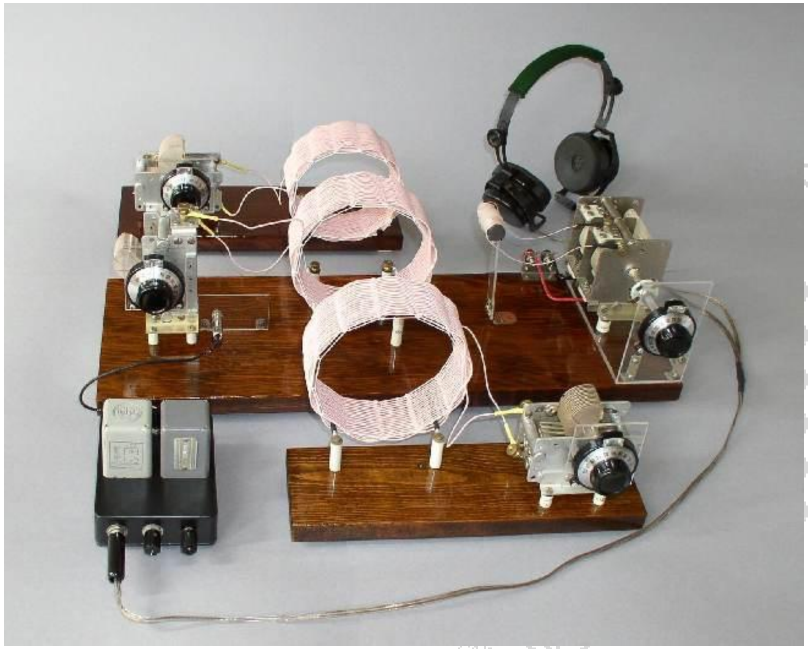

www.crystalradio.net

- Mike Tuggle of Hawaii is one of the top builders of crystal sets.
- His set, the Lyonodyne-17 has heard stations in Cuba from his home location in Hawaii!
- There are lots of web resources available. This presentation describes my efforts.

Conclusions

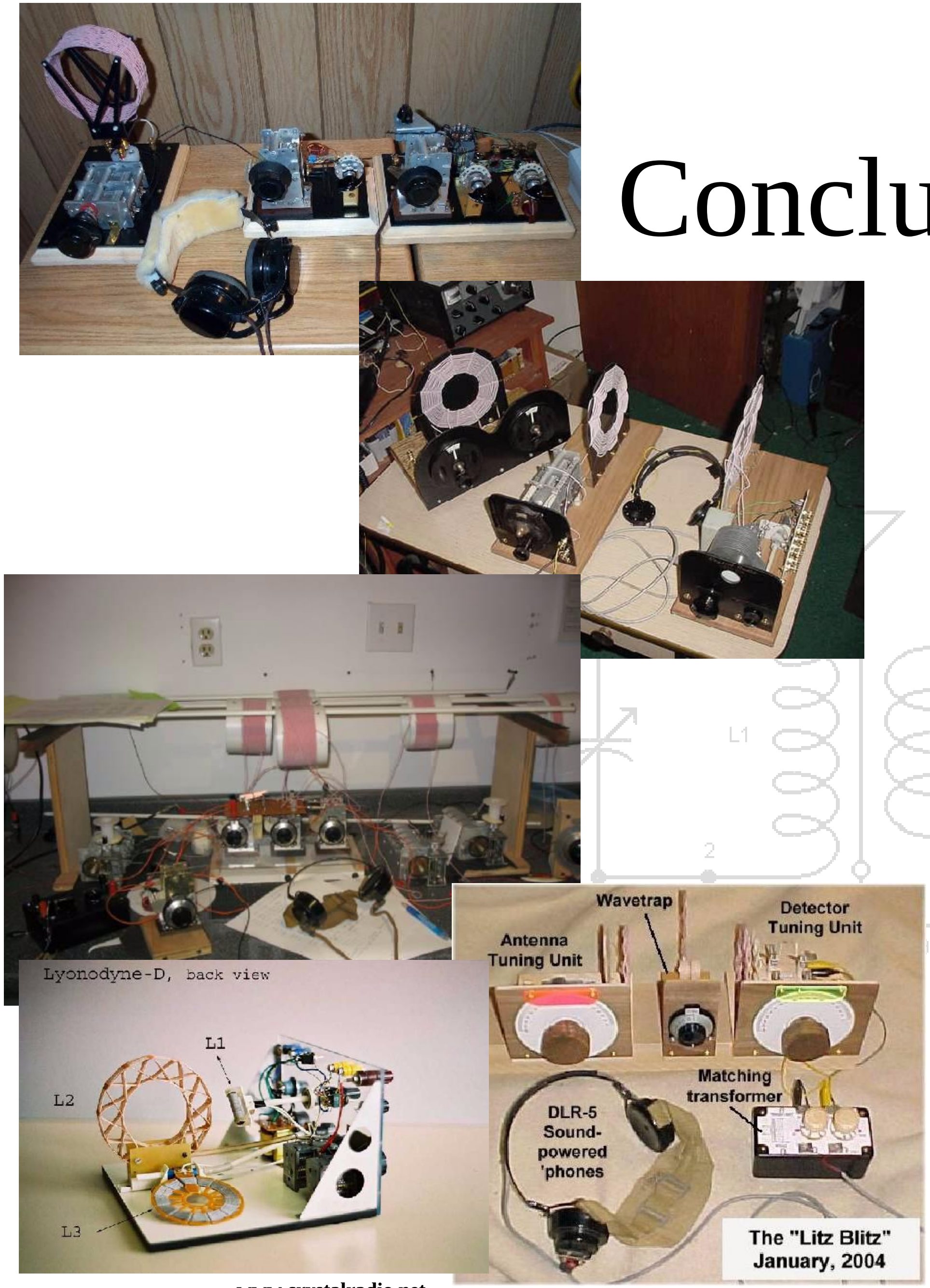
Lyonodyne-D, back view
L1
L2
L3
Wavetrap
Antenna Tuning Unit
Detector Tuning Unit
Matching transformer
DLR-5 Sound-powered 'phones
The "Litz Blitz" January, 2004

www.crystalradio.net

- Crystal sets appear to be simple, but attention must be paid to all the details.
- If you truly understand a crystal radio, you have a good foundation to RF engineering in general.

Homebrew Galena Crystal Set

Enjoying a 1920's radio experience

Ever since my youth, I have been attracted to crystal radios I had seen pictured in old books. There is a special appeal about a device, however primitive, that can pull in radio voices and music through the air from many miles away without batteries or electrical power of any kind, just an antenna and a ground.

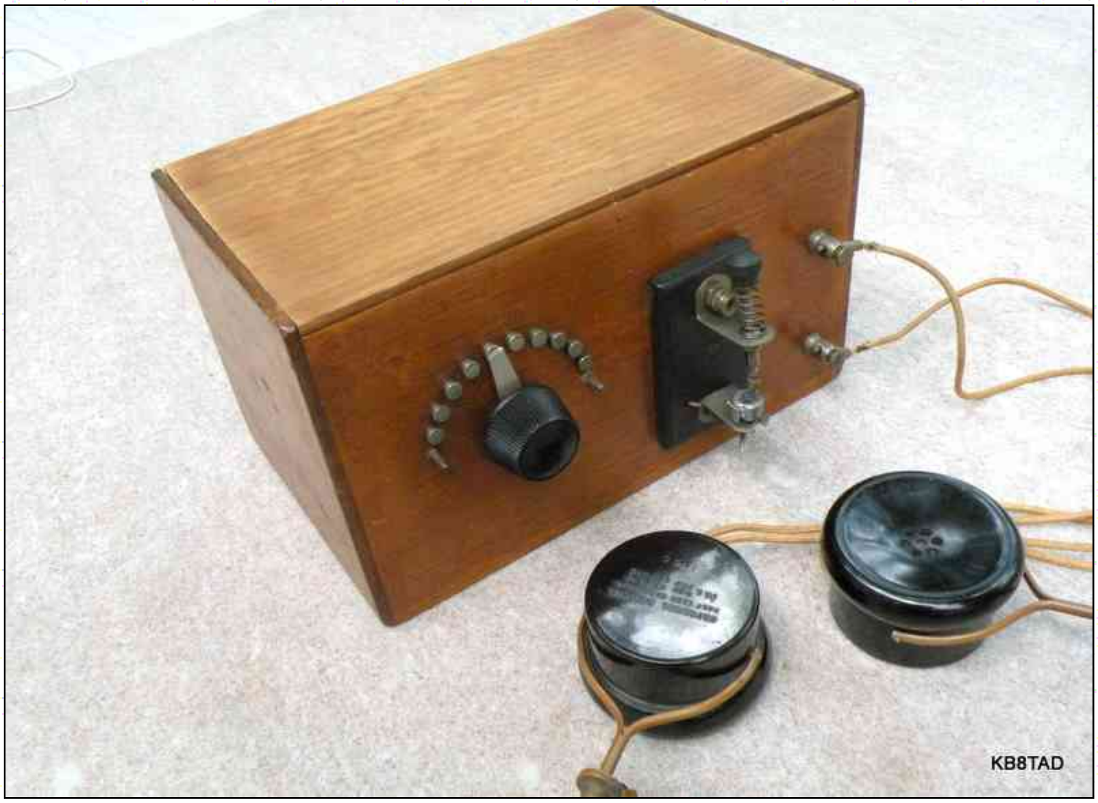

That explains my immediate attraction to this little homebrew crystal radio with galena detector found at a swap meet. It was built into a deep cigar box using a heavy paper roll as a coil form. The coil form measures 4 1/4 inches long with 3 3/4 inches outside diameter. The coil itself consists of 99 turns of solid double-cotton covered wire tapped every 9 turns. Those taps are fed to ten smooth-head screws that make up the tap selector switch contacts.

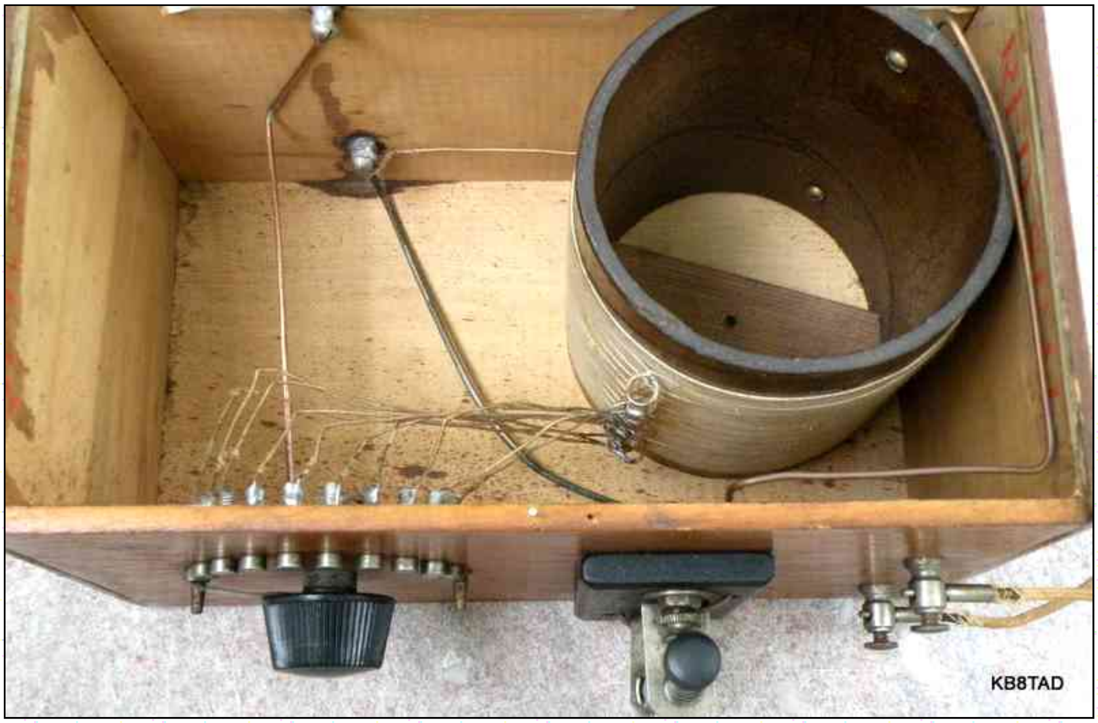

Repairs

As with all older electronics, the first repair is a good cleaning, making sure the solder joints are in good shape, and applying a bit of deoxit to all moving electrical contacts, in this case the ten-position tap switch and the crystal holder.

The top of the cigar box had been removed and apparently lost over the years. The cover had been fastened with a small nail which was still on the box. I made a top cover from another old cigar box lid.

The galena crystal was potted in Wood's metal that had separated into two pieces. Quick work with a soldering iron solved that problem. I cleaned the galena surface with rubbing alcohol as recommended in older articles. I then measured forward and back resistance ratios of the galena-and-cats-whisker diode. While there is a signifcant difference between forward and back resistance readings on an ohmmeter, the ratio was nowhere close to that of a modern germanium diode, thus limiting the efficiency of the crystal.

A 1920s experience

I hooked up the set to my 80 meter dipole antenna and listened to the two local AM stations with adequate volume using my equally old but very efficient Baldwin headphones. I could find sweet spots with the "cat's whisker" on the galena crystal that worked reasonably well. However, connecting a 1N34A germanium diode in series with the headphone lead to the galena crystal made a major difference. I could see why the introduction of fixed germanium diodes quickly ended the era of the galena and cat's whisker detector.

Crystal set schematic showing similar set with 7 position switch and coil taps
(Source: The Radio Amateur's Hand Book by A. Frederick Collins, 1923)

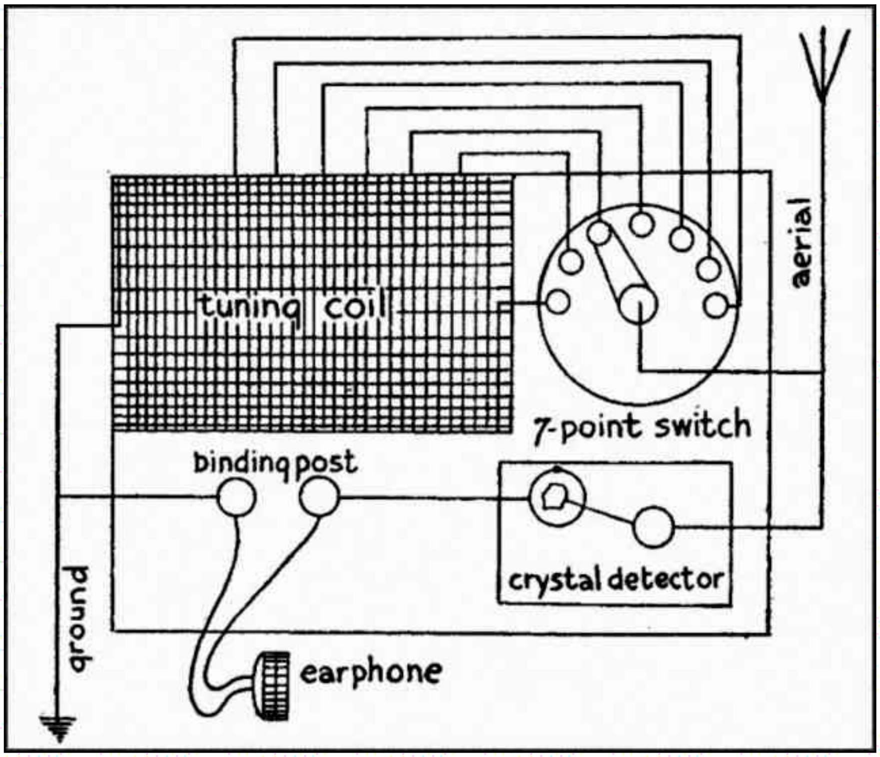

The ten-position tap switch enabled good separation in tuning of the two local stations which are 370 KHz apart. At night I was able to also comfortably listen to a some distant stations with strong signals.
I also tried a very long wire antenna but found the volume to be about the same as with the 80 meter dipole.
Living the experience of a 1920s listener was the fun part. My hat is off in tribute to the quality work of the unknown builder of this very nice little set.

Information sources
The internet has lots of good sources on building crystal radios with modern parts. Here are a few:
"Build A Crystal Radio" from **The Stay Tuned Website**
"Simple Crystal Radio" from **techlib.com**
"Building a crystal radio out of household items" from **scitoys**
"Quaker Oat Box Radio Project" from **The Xtal Set Society**

8-24-10;

An Altec Lansing A-324A Amplifier was the previous item on the bench.

Go back to the BA Pix Homepage.

(https://www.facebook.com/IonPowerGroup/) (https://twitter.com/ionpowergroup) in (https://www.linkedin.com/company/ion-power-group-llc)

(https://www.youtube.com/user/IonPowerGroup)

Ion Power Group (https://ionpowergroup.com/) (https://ionpowergroup.com/)
(https://ionpowergroup.com/) Electricity from Atmospheric Ions

How
Ion Harvesting
generates
clean electricity
on Earth

Ion Harvesting Technology – the process of utilizing patented carbon nanomaterial to harvest the electric charge of naturally occurring ions in the atmosphere offers a way to produce clean supplemental high voltage electricity anywhere on Earth – day and night.

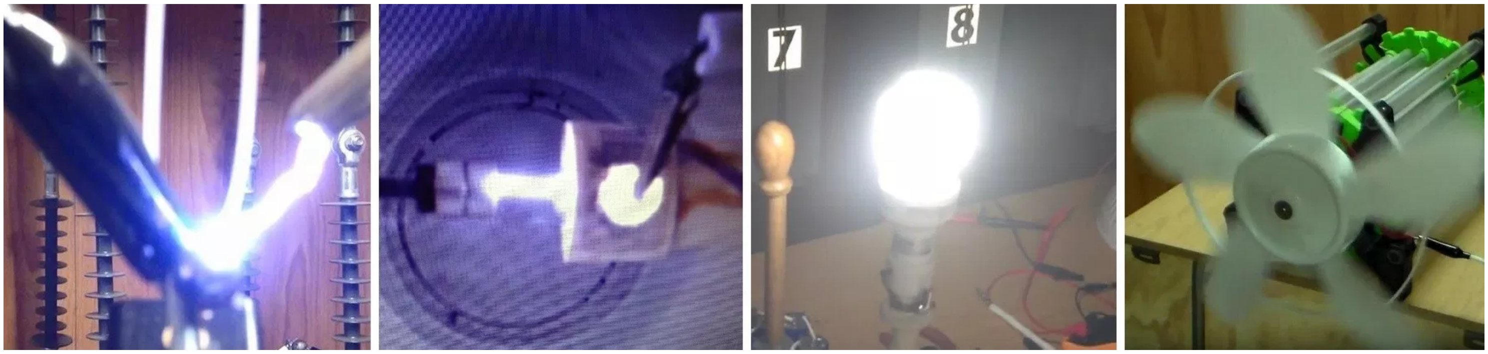

Experiments demonstrate that high voltage electricity harvested from airborne ions can power lights, motors or produce hydrogen & oxygen gas via water electrolysis or place a charge on lithium-iron (LiFePO4) batteries and be converted to 120VAC/60hZ ‘house style’ current. See ‘Concept Clips’ for videos.

Abstract of Economic Viability Report by Dr. Philip Metzger Planetary Scientist/Physicist PhD and NASA Kennedy Space Center Scientist/Engineer of the Year (blue text) Full Report is 16 pages

Dr. Philip Metzger
Planetary Scientist - Physicist PhD
NASA Kennedy Space Center
Scientist/Engineer of the Year

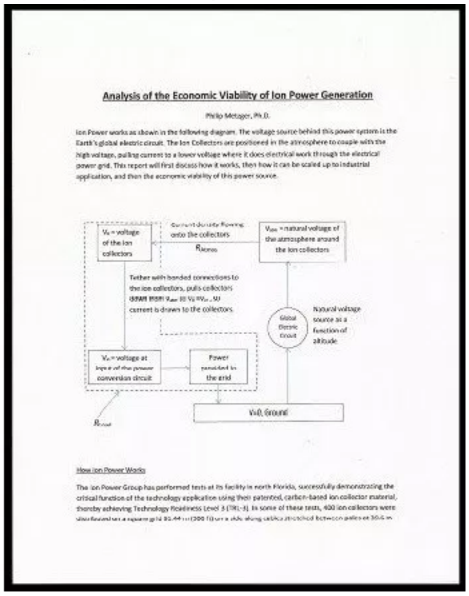

Analysis of the Economic Viability of Ion Power Generation

Philip Metzger, Ph.D.

Ion Power works as shown in the following diagram. The voltage source behind this power system is the Earth's global electric circuit. The Ion Collectors are positioned in the atmosphere to couple with the high voltage, pulling current to a lower voltage where it does electrical work through the electrical power grid. This report will first discuss how it works, then how it can be scaled up to industrial application, and then the economic viability of this power source.

Vc = voltage of the ion collectors
Current density flowing onto the collectors
Rionos
Vatm = natural voltage of the atmosphere around the ion collectors
Tether with bonded connections to the ion collectors, pulls collectors down from Vatm to Vc ≪Vatm, so current is drawn to the collectors
Global Electric Circuit
Natural voltage source as a function of altitude
Vin = voltage at input of the power conversion circuit
Power provided to the grid
V=0, Ground
Rcircuit

How Ion Power Works

The Ion Power Group has performed tests at its facility in north Florida, successfully demonstrating the critical function of the technology application using their patented, carbon-based ion collector material, thereby achieving Technology Readiness Level 3 (TRL-3). In some of these tests, 400 ion collectors were

Click for Full Report

“Physics-based modeling shows that an (https://ionpowergroup.com/wp-content/uploads/2017/11/Economic-model-study-Rev-B-authored-by-Dr.-Phil-Metzger-PhD.pdf)electron cloud forms around the ion collector material, and this electron cloud increases conductivity of the air around the collector. This causes electrical current to flow, driven by the atmospheric voltage, so that power is generated at the ground. The rate of this current is governed by complex physics in the boundary layer of the collector material, doubtless because of its very high surface area. Ionization can occur on this collector at extremely low voltages because of the sharp, needle-like protrusions on its surface. This understanding of the physics suggests that power generation can be scaled as surface area times voltage. This was incorporated into a simulation of the solar cycle with random storms, and power generation of photovoltaics (PV) and Ion Power were simulated as feeding into a smart grid. Because Ion Power collects the most energy at complementary times compared to PV, the energy storage requirements on the smart grid is greatly reduced. This causes the cost of electricity also to be greatly reduced compared to photovoltaics alone, resulting in energy costs for the overall smart grid of about 8.25 cents per kWh, which is less expensive than any other energy source other than some wind farms (which are only slightly less expensive and have limited geographic applicability). This analysis uses extremely conservative assumptions, so actual energy costs may be much less. **The estimated value to the US economy is between $62 billion and $150 billion (annually)** with proportionately large value to the global economy. Thus, it is recommended that further development of this energy source begin immediately."

Galactic Cosmic Rays (GCRs) arriving from deep space bombard the Earth's atmosphere around-the-clock creating trillions of naturally occurring electrically charged ions throughout the atmosphere. Ion Power Group utilizes patented carbon nanomaterials to couple to some of these electrically charged airborne ions resulting in the harvesting of clean renewable high voltage electricity day and night.

Galactic Cosmic Rays arriving from deep space impact electrically neutral oxygen and nitrogen atoms in the Earth’s atmosphere producing a particle shower (or particle cascade) creating electrically charged ions in the atmosphere day and night.

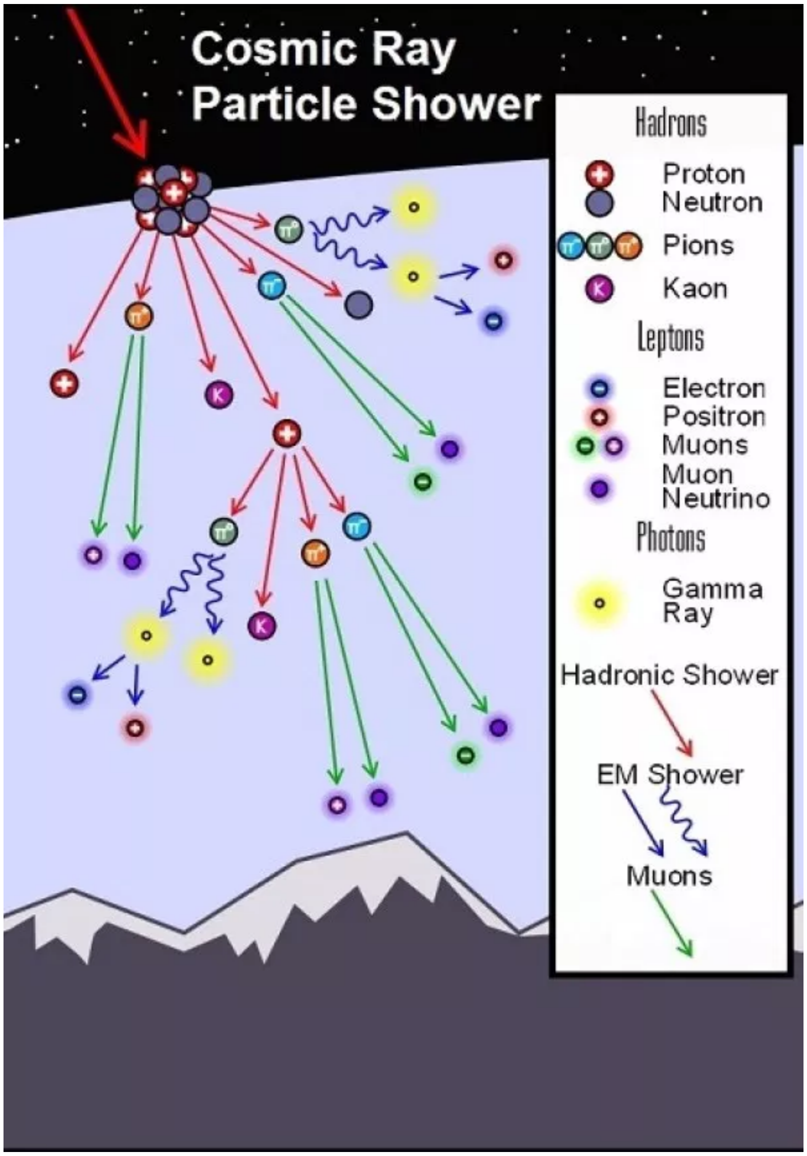
Cosmic Ray
Particle Shower
Hadrons
Proton
Neutron
Pions
Kaon
Leptons
Electron
Positron
Muons
Muon
Neutrino
Photons
Gamma
Ray
Hadronic Shower
EM Shower
Muons

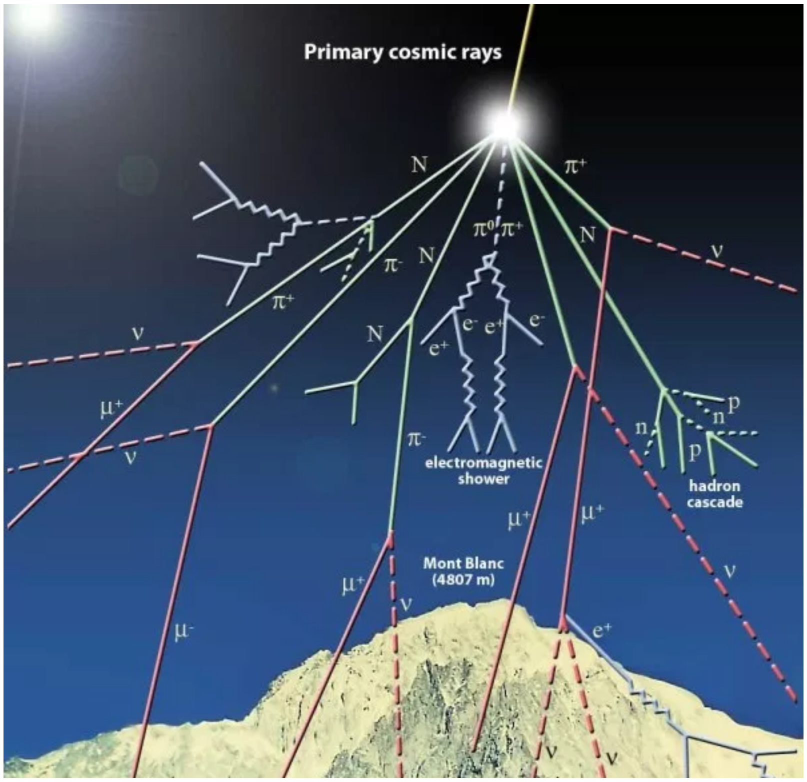
Primary cosmic rays
N
π+
π0 π+
π- N
N
ν
π+
N
e- e+ e-
e+
μ+
ν
π-
electromagnetic
shower
p
n
n
p
hadron
cascade
μ+
μ+
ν
Mont Blanc
(4807 m)
μ+
ν
μ-
e+
ν
ν

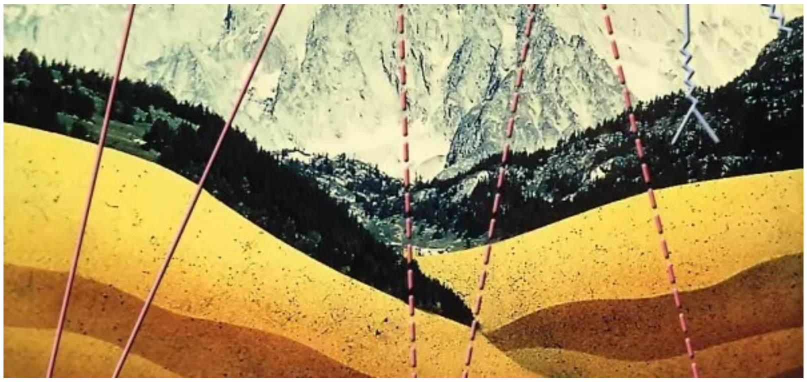

Very high-energy cosmic rays can generate huge showers of up to 10 billion secondary particles or more. The energies of the most energetic ultra-high-energy cosmic rays (UHECRs) have been observed to approach 3 × 1020 eV,[7] about 40 million times the energy of particles accelerated by the Large Hadron Collider. Each of the trillions of atmospheric ions generated throughout the Earth's atmosphere act as a charge carrier conveying a high voltage electric charge. The production of atmospheric ions is a natural process that occurs day and night across the entire Earth.

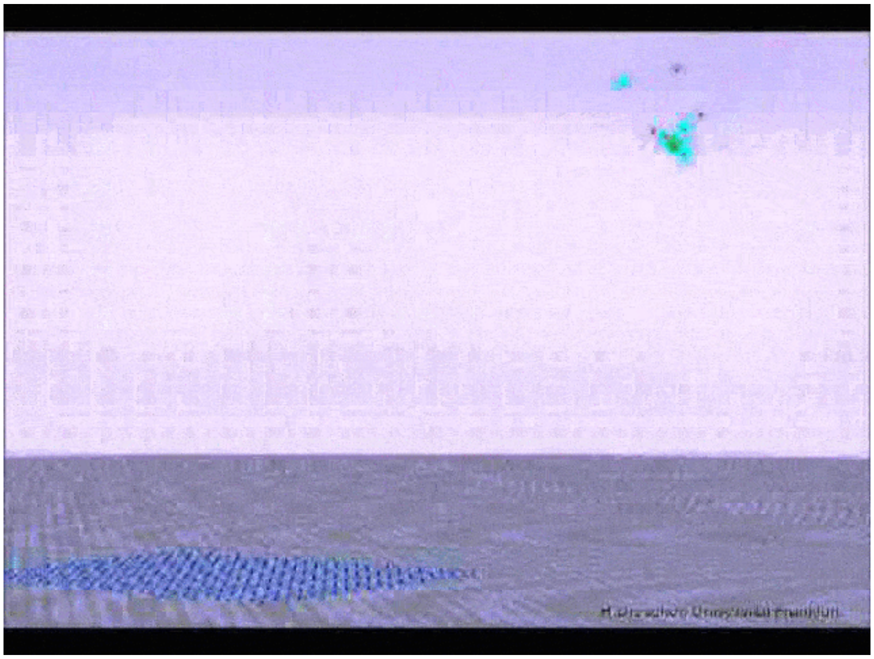

Watch the below video clip to see visual proof of cosmic rays bombarding the Earth's atmosphere – particularly the beginning and near the end of the clip.

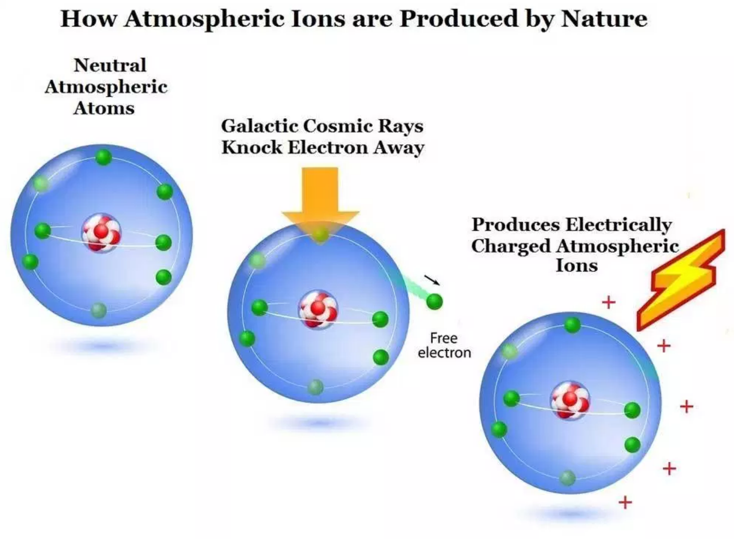
How Atmospheric Ions are Produced by Nature
Neutral Atmospheric Atoms
Galactic Cosmic Rays Knock Electron Away
Produces Electrically Charged Atmospheric Ions
Free electron
+
+
+
+
+

Atmospheric ions are produced by natural processes

Electrically charged ions are present in the Earth’s atmosphere day and night all around the world. Galactic Cosmic Rays arriving from deep space (which are actually microscopic particles traveling at relativistic speed) plus constant atmospheric electrical activity (thunderstorms) around the planet, plus Radon gas seeping from the Earth’s crust in certain geographic areas combine to replenish the supply of atmospheric ions. Atmospheric ions are a natural self-renewing energy resource that cannot be depleted.

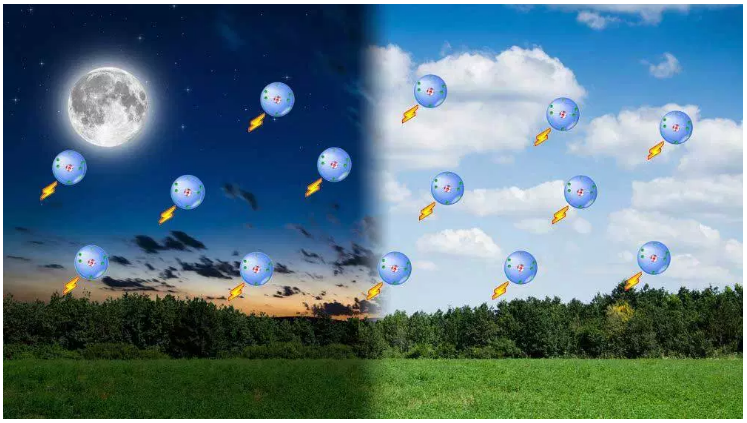

Atmospheric ions are present day and night

In addition to constant GCR activity, at any moment about 2,000 thunderstorms are occurring around the Earth. On average, across the world there are 4,320,000 lightning discharges each 24 hour period (3000 per minute). This tremendous amount of high voltage electrical energy, together with GCRs bombarding the atmosphere and Radon gas seeping from the Earth's crust in certain areas, combine to actively replenish the supply of electrically charged atmospheric ions around the world.

Ion Harvesting Technology harvests electricity day and night and even during bad weather

Ion Power Group's ion collectors, when provided with adequate altitude, can harvest the electric charge of near-Earth ions to produce clean renewable high-voltage electricity day and night and during storms. Ion Harvesting Technology should not be confused with other techniques that attempt to capture lightning bolts or harvest manmade radio frequency (RF) energy.

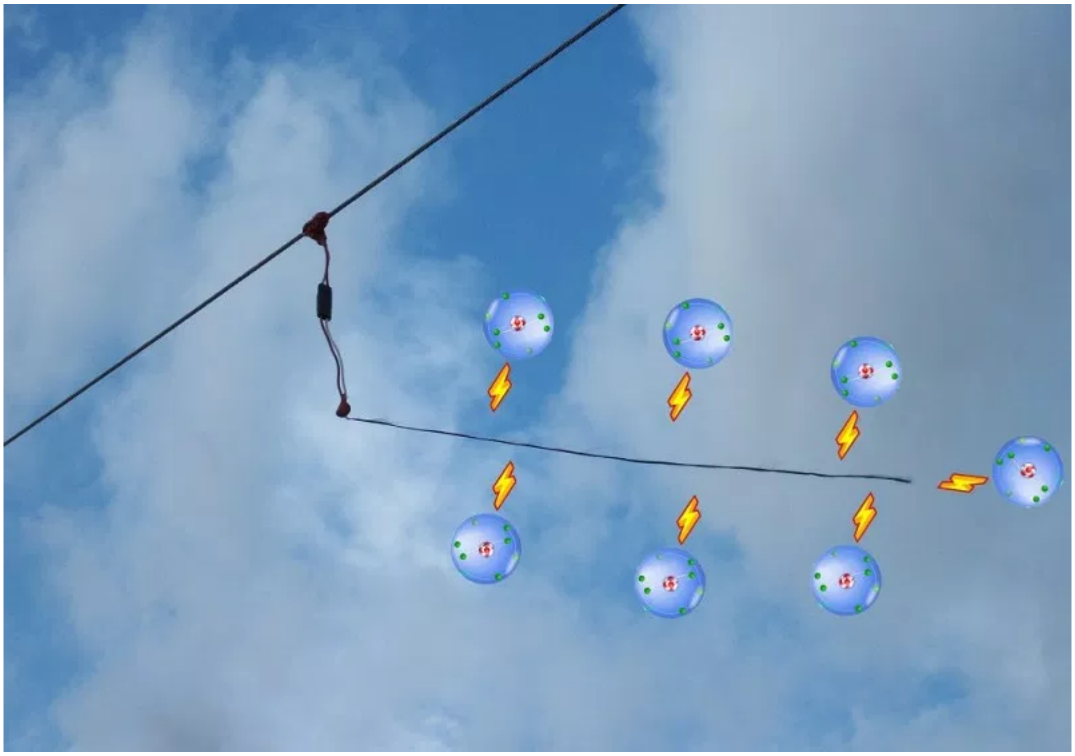

(https://i1.wp.com/ionpowergroup.com/wp-content/uploads/2016/09/Ion-Collector-harvesting-ions-PLAIN.jpg?ssl=1)

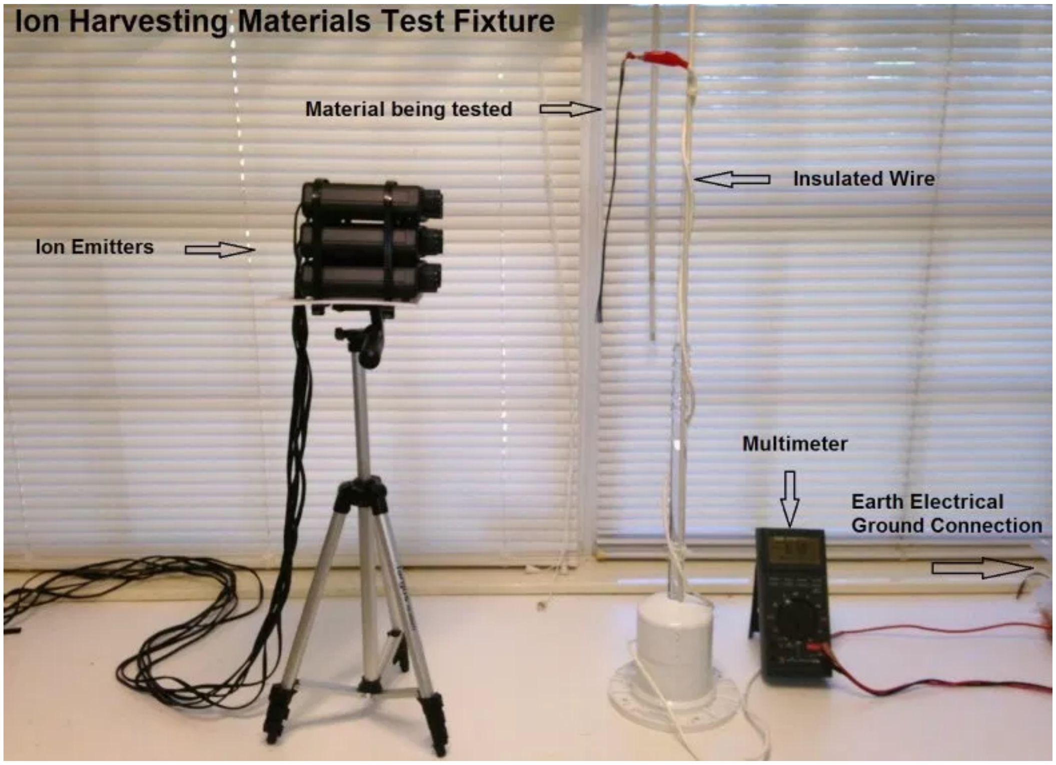
Ion Harvesting Materials Test Fixture
Material being tested
Insulated Wire
Ion Emitters
Multimeter
Earth Electrical Ground Connection

(https://i1.wp.com/ionpowergroup.com/wp-content/uploads/2016/10/Ion-Harvesting-Table-Top-Assembly-annotated.jpg?ssl=1)

Ion collectors can be composed of carbon/graphite/graphene, while not metallic, are electrically conductive, flexible and durable. Electric fields are known to coalesce at conductive points known as the “Corona Effect”. The solution of Poisson’s equation to determine electric field intensity around the collectors shows us that the highest electric fields will be around sharp points; the sharper the point, the greater the electric field. Ion collectors offer millions of microscopic electrically conductive points and protrusions which readily couple to the electric fields of near-Earth ions transferring high-voltage electricity from the atmospheric ions to the ion collectors.

$$\nabla^2 \varphi = -\frac{\rho}{\varepsilon}.$$

Ground based circuitry maintains the ion collector material at a voltage considerably lower than the surrounding atmospheric ions thereby employing the electrostatic attraction principle causing nearby ions to migrate to the ion collectors. High voltage wire connected to the ion collectors convey the harvested high voltage DC electricity to a storage medium or electrical load to produce work. Measurements have shown that ion collectors made of graphite are more efficient at harvesting electricity from airborne ions than all other tested metals. In some tests, by as much as 55x greater efficiency. Some of the tested metals include highly conductive silver, copper, and aluminum wire – aluminum spheres – copper and aluminum sheets.

The below graphs show voltage and current coupling characteristics comparing metal in one test (indicated as red and green) to carbon/graphite (indicated as purple).

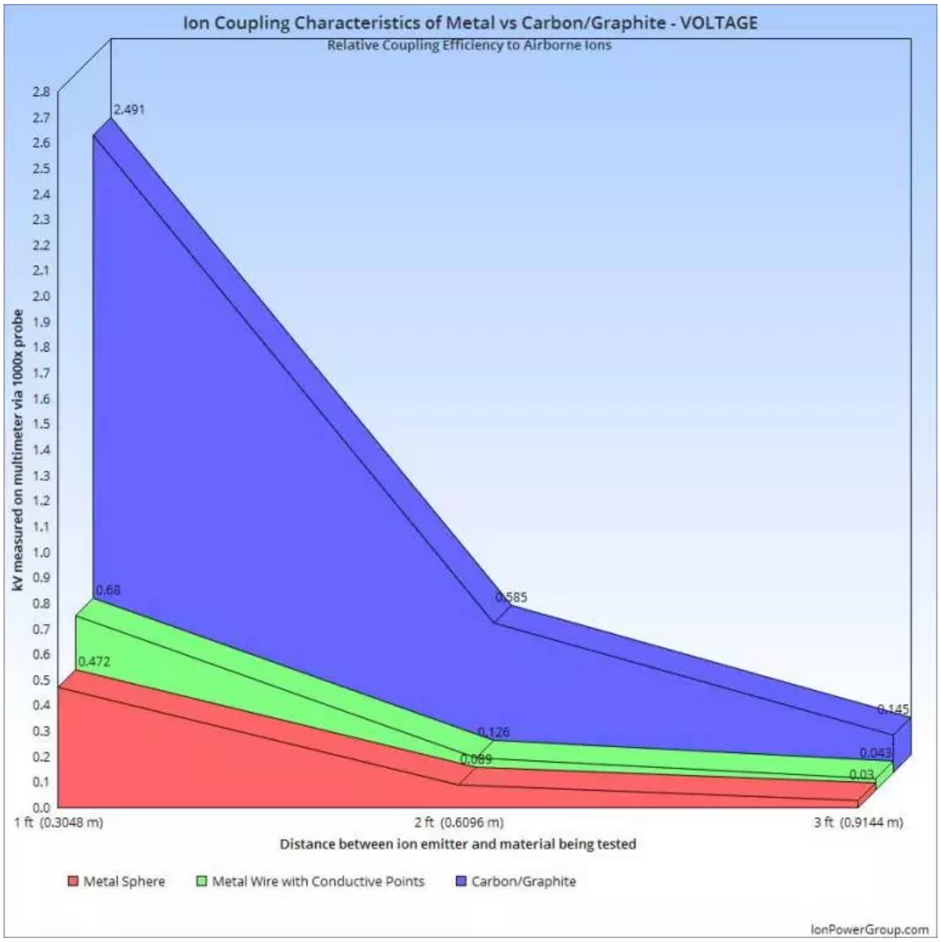

(/wp-content/uploads/2016/07/Ion-Coupling-Characteristics-VOLTAGE-1.jpg)

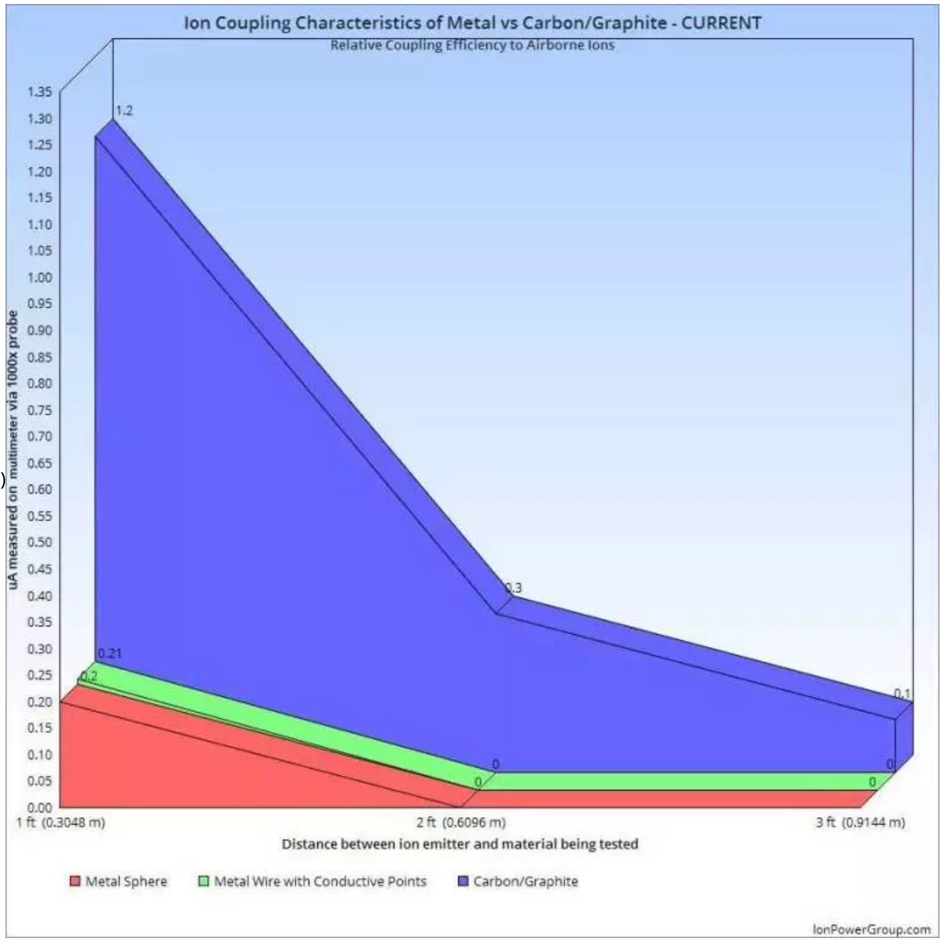

(/wp-content/uploads/2016/07/Ion-Coupling-Characteristics-CURRENT.jpg)

Below is a picture of carbon, a key element in ion collectors, taken at the microscopic level by an electron microscope.

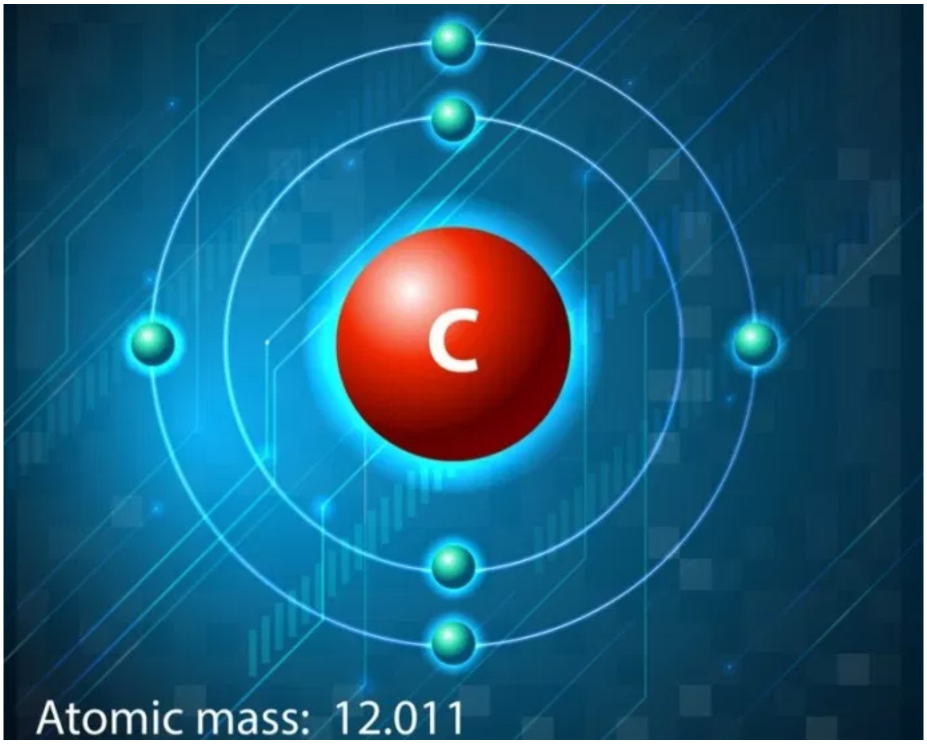
C
Atomic mass: 12.011

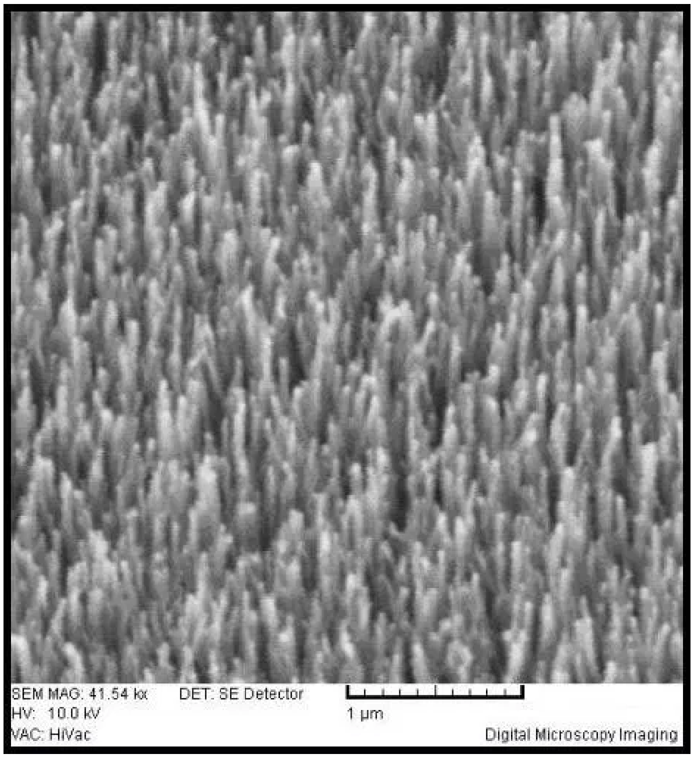
SEM MAG: 41.54 kx
DET: SE Detector
1 μm
HV: 10.0 kV
VAC: HiVac
Digital Microscopy Imaging

Electricity is produced by providing sufficient altitude to the ultra lightweight ion collector through the use of long-duration aerial platforms including tethered balloon, kite or tower. Atmospheric voltage increases at the rate of approximately 90-150vdc per meter of altitude above ground level or sea level and many orders of magnitude greater during disturbed weather. Experiments have demonstrated that the electric charge of atmospheric ions can be harvested by ion collectors and stored in capacitor banks, super-capacitors, lithium-iron (LiFePO4) batteries, directed to a water electrolyzer to produce hydrogen and oxygen gas or converted to 120VAC/60hZ ‘house style’ output power. An equally important role of atmospheric ions is that ions cause the atmosphere to be electrically conductive allowing ion collectors to harvest more electricity during disturbed weather or increased altitude above ground level. The atmosphere’s electrical conductivity increases with altitude. For example, when a storm or active cloud formation is nearby, the harvested DC power can skyrocket many orders of magnitude by reason of the conductivity of the atmosphere. This is often referred to as a peak ion period or disturbed weather.

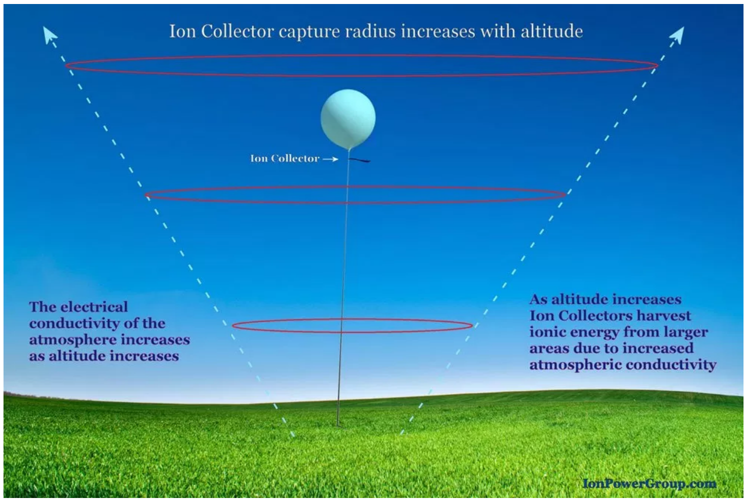
Ion Collector capture radius increases with altitude
Ion Collector →
The electrical conductivity of the atmosphere increases as altitude increases
As altitude increases Ion Collectors harvest ionic energy from larger areas due to increased atmospheric conductivity
IonPowerGroup.com

Atmospheric voltage increases about 90-150vdc per meter of altitude

Additionally, the Triboelectric Effect plays a role in how much electricity is harvested. Geographic areas with Radon gas can tend to have greater ionic airborne densities closer to the ground due to the gas leaking up from the Earth. To a degree not yet defined, man-made air pollution can also raise the level of airborne ion densities. Cities with polluted air may prove to enjoy an extra benefit from ion harvesting. Galactic Cosmic Rays (which are actually particles) travel at enormous speed through outer space and impact the Earth’s atmosphere. Many of the GCRs are attenuated by the atmosphere, but some are energetic enough to penetrate into the atmosphere where they can impact electrically neutral oxygen and nitrogen atoms with such intense kinetic energy that a particle cascade is produced that creates electrically imbalanced atoms (ions) all along a downward path allowing some of the ions to cascade nearer to the Earth’s surface. The imbalanced atoms (ions) created all along the particle cascade can extend down nearer to the Earth’s surface where ion collectors harvest the electrical field potential of the ions to receive and store usable high voltage DC electricity. Since the ions effectively cause the air to be electrically conductive, they create an electrical path for storms occurring in other locations of the globe to contribute to the Earth's overall ambient atmospheric electrical voltage. Ion Power Group's remote test site (est. 2005) located on a 300 acre Florida pasture, uses tall poles (130 feet) to elevate ion collectors for research purposes. A wire conveys the harvested energy down to the remote research lab. The act of harvesting the electric charge from atmospheric ions harmlessly reverts the ions back to their original state of neutral atoms. The former ions, now neutral atoms, are once again free to migrate through the atmosphere – available to be converted back into electrically charged ions by natural processes. Ion harvesting results in the generation of clean, natural, renewable energy that will not pollute Earth.

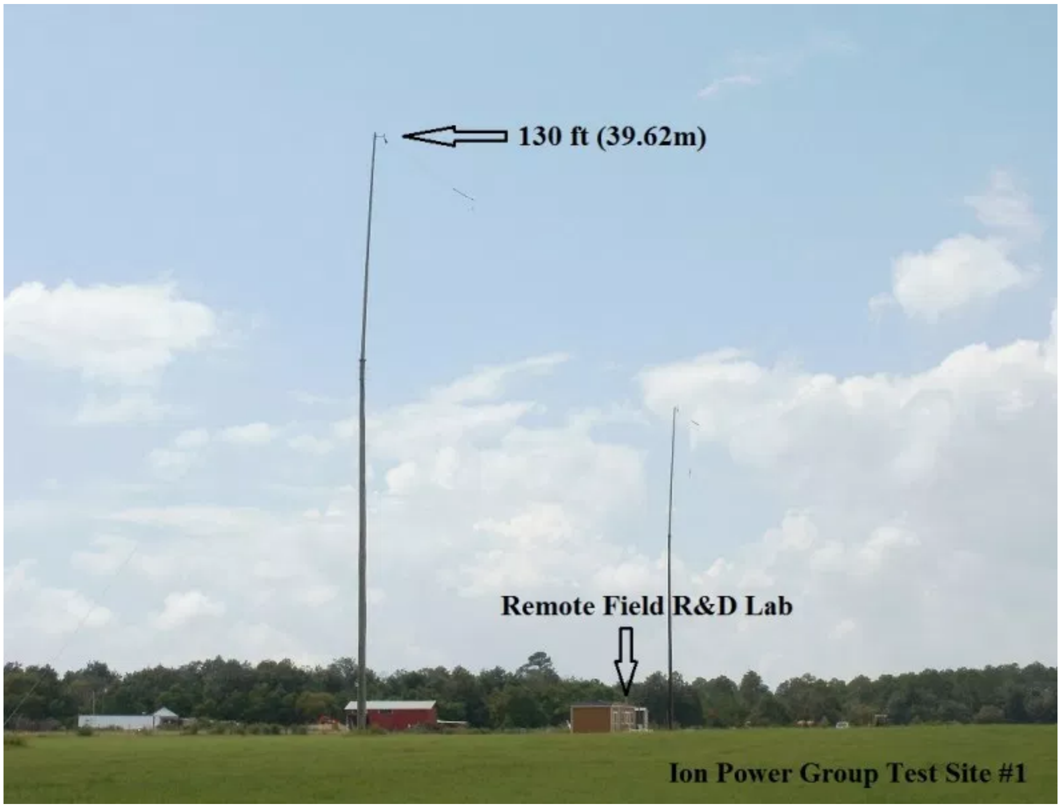
130 ft (39.62m)
Remote Field R&D Lab
Ion Power Group Test Site #1

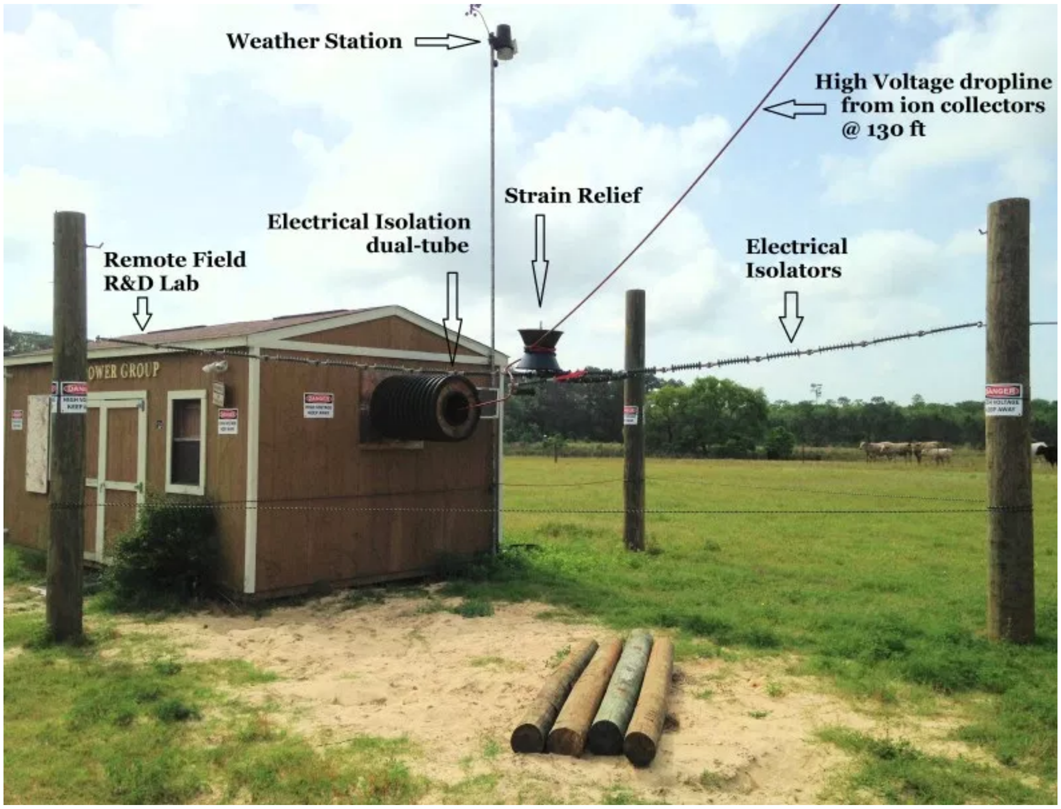
Weather Station
High Voltage dropline from ion collectors @ 130 ft
Strain Relief
Electrical Isolation dual-tube
Remote Field R&D Lab
Electrical Isolators

(https://i0.wp.com/ionpowergroup.com/wp-content/uploads/2016/12/Dropline-RD-room-FINAL.jpg?ssl=1)

The peak electrical power generated in real-time by Ion Power Group's proof-of-concept prototype during peak ion periods with the ion collectors at 130 feet above ground level is 1,236 watts (1.236kW). Derived from 41,200 volts @ 30.ma directly measured through a 1000x probe during peak ion periods.

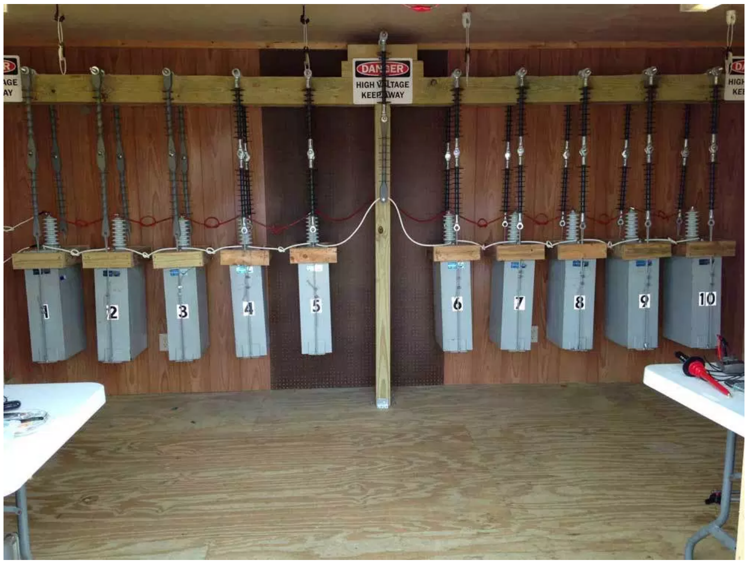

(https://i0.wp.com/ionpowergroup.com/wp-content/uploads/2016/09/Capacitor-Bank-Opti.jpg?ssl=1)

Bank of 10 General Electric high voltage capacitors at the Florida research and development test site

The above General Electric capacitor bank, rated at 75.uf, has repeatedly been charged up to >29,000.vdc (29.kV) equating to >31,000 Joules (31.kJ) of energy by harvesting the electric charge of airborne ions in real-time during peak ion periods with the ion collectors at 130 feet above ground level. The stored energy has been used for demonstrations including powering 32 feet of florescent lights, motors and producing hydrogen gas from water.

Hydrogen Gas produced from Water: Ion Power Group has demonstrated that it is possible to harvest atmospheric ions to generate electricity that can be utilized in electrolysis to produce hydrogen gas and oxygen gas from water (in proof-of-concept quantities so far). Hydrogen gas is used by next generation, non-polluting, hydrogen powered cars produced by worldwide automobile manufacturers including the world’s largest automaker – Toyota. Ion Power Group has been awarded a patent for a novel approach to producing hydrogen gas from water utilizing power obtained from ion harvesting.

Electricity harvested from Airborne Ions + Water = Hydrogen Gas

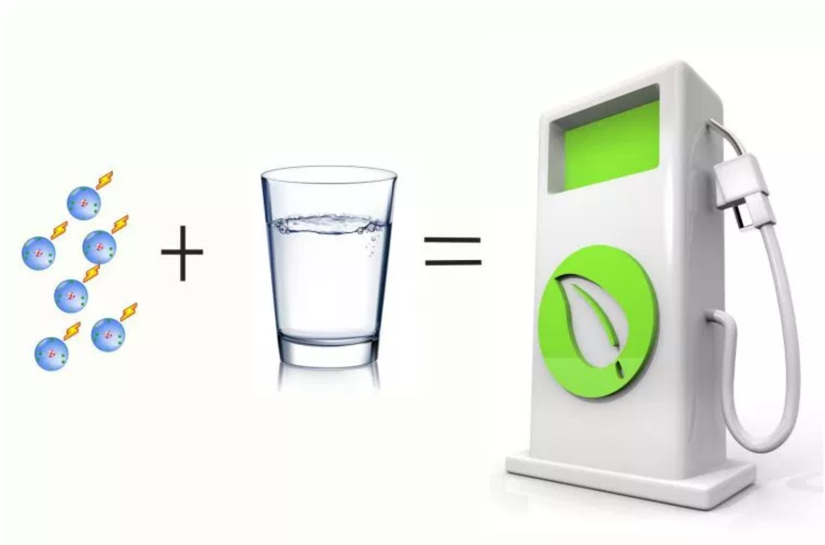

Ion Power Group has been awarded patents for installing ion collectors on automobiles to generate electricity including through the Triboelectric Effect. Preliminary field tests in which a truck, equipped with Ion Collectors, demonstrated that the concept can generate >500 volts when the vehicle is in motion. Initial tests are encouraging, however, further development is required in order to realize automobiles that generate their own supplemental electricity (or onboard hydrogen gas from water) through use of Ion Collectors.

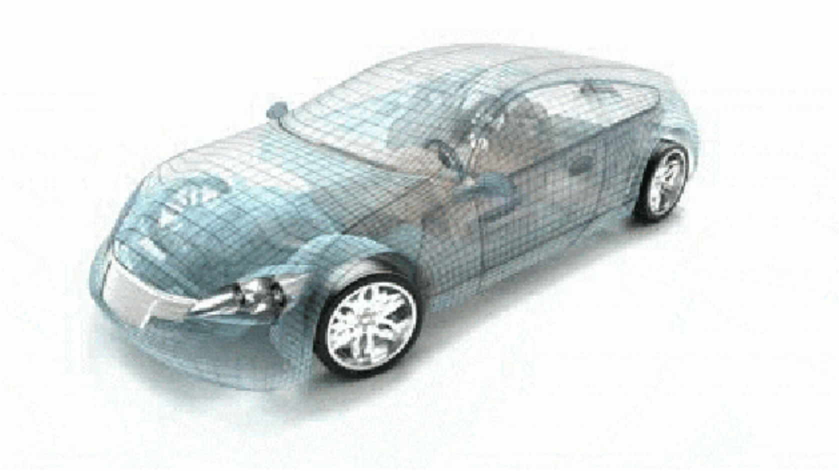

Illustrations (below) of proposed experimental ion power plants using long-duration aerostats (aerostats or balloons) or kites or towers to provide altitude to the ion collectors to harvest ions for electricity, day and night. Future Ion Harvesting Technology embodiments that use long-duration aerostats (balloons) to provide great altitude to ion collectors offer a theoretical energy producing duty-cycle approaching 99%.

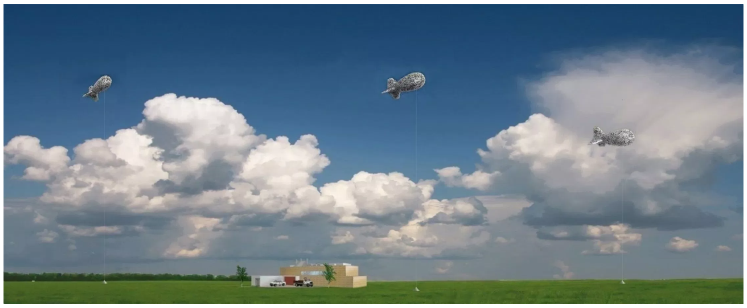

Tethered aerostats covered with patented carbon/graphite nanofiber to harvest clean electricity day and night from atmospheric ions. The harvested electricity is conveyed down the conductive tether for use on the ground.

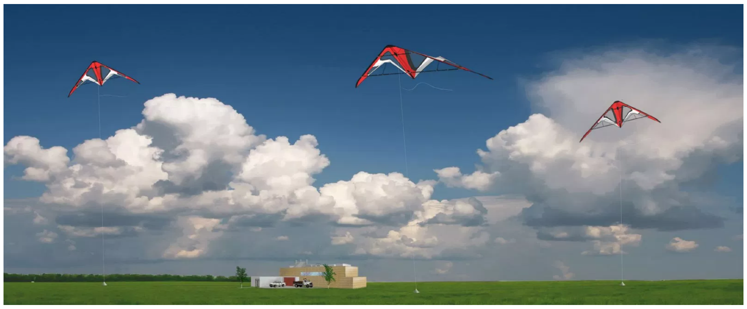

Kites used to provide altitude to patented ‘strip’ carbon/graphite nanofiber to harvest clean electricity day and night from atmospheric ions. The harvested electricity is conveyed down the conductive tether for use on the ground.

Long duration weather balloons used to provide altitude to ion collectors made of patented carbon/graphite nanofiber to harvest clean electricity day and night from atmospheric ions. The harvested electricity is conveyed down the conductive tether for use on the ground.

Theoretical vs Real-World:

There are a number of technical reports and peer reviewed journal articles available for viewing on the internet that do an excellent job of characterizing the Earth’s Electric Circuit addressing fair weather (the load of the circuit) as well as perturbed conditions (the generator part). These reports tend to model the Earth’s Fair Weather Circuit as a ‘whole’ averaging the energy produced by thousands of storms over the entire atmosphere. One might conclude from such reports that the atmosphere offers a maximum of 2 pA/m2 of electrical current. Ion Power Group has oftentimes measured significantly greater power during peak ion periods – many watts of harvested power – as great as 1,236 watts (41,200vdc @ 30ma) with ion collectors at 130 feet altitude. What accounts for the very wide discrepancy between theoretical models presented by some reports available on the internet and Ion Power Group’s real-world measurements? **Answer:** Localized disturbed weather can greatly increase the electrical component of the *local* atmosphere thereby significantly increasing the level of electricity available to ion harvesting by many orders of magnitude. In addition, for estimating the available power, the voltage increase with altitude, air conductivity (e.g. with respect to Radon emissions from the surface), and characteristics of the instrumentation (conductivity of the material, insulation) need to be considered. Because of these complexities, there are hardly any realistic studies or even measurements on the available power locally outside of research conducted by Ion Power Group. Also, the total power in the electric circuit is still an open research question. Most researchers who have published reports about the Fair Weather Circuit intentionally do not focus on the local effects of disturbed weather or Radon gas components. With regard to harvesting electricity from the atmosphere, the absence of modeling the beneficial effect of localized disturbed weather and the effects of Radon gas may lead readers to a conclusion that drastically underestimates the *actual* electrical power available to be harvested during localized disturbed weather. During real-world testing, Ion Power Group has repeatedly demonstrated that localized disturbed weather greatly increases the electrical power output of ion harvesting technology, many thousands of times greater than 2 pA/m2 when using carbon/graphite nanomaterials in the harvesting process. The solution to increasing electrical power is to provide greater altitude to the carbon nanomaterial ion collector, via long duration tethered balloon, thereby increasing the harvested voltage. The electrical conductivity and voltage of the atmosphere increases with altitude by approximately 90-150vdc per meter, therefore, the effective capture-area increases as the Ion Collector is raised higher in altitude (see the above image titled “Ion Collector Capture Radius Increases with Altitude”). Voltage (and sometimes current, depending e.g. on orography) have been observed to increase as an ion collector is raised in altitude. However, assuming a situation in which the atmospheric current stays constant while voltage increases (due to the ion collector being raised to a higher altitude) the result is a net power increase as shown in this formula.

P=VxI. Therefore, if I is constant, as V increases, P increases proportionately

Ion Harvesting Technology addresses the problem of low atmospheric current during fair weather conditions by increasing the altitude of ion collectors thereby increasing overall harvested power. It is noteworthy that none of the reports presently available via the internet address the harnessing of atmospheric electricity for power production based on using more efficient carbon nanomaterials such as Graphite and Graphene. Therefore, present reports do not reflect the most recent advancements in technology for harvesting atmospheric electricity. For the reasons stated herein, previous reports available prior to Ion Power Group’s research (while most are very good when read in context) should not be considered an accurate representation of the *actual* electrical power available to advanced Ion Harvesting Technology because the reports do not model available power based on the complex current flow considering the local atmospheric conductivity, potential distribution, and how utilizing carbon (https://ionpowergroup.com/wp-content/uploads/2018/04/Aerostat-Deployed-v2.0-with-output-cable-v2.5.jpg) nanomaterials (https://ionpowergroup.com/wp-content/uploads/2018/03/Aerostat-High-Wind-Design-v3.0.jpg) such (https://ionpowergroup.com/wp-content/uploads/2018/03/Tether-images-with-border-v3.0-100.jpg) as (https://ionpowergroup.com/wp-content/uploads/2018/03/Aerostats-WWII-montauge-v1.0.jpg) Graphite (https://ionpowergroup.com/wp-content/uploads/2018/04/Trailer-Cost-per-Watt.jpg) and Graphene alter (https://ionpowergroup.com/wp-content/uploads/2018/03/Ion-Harvesting-Power-Estimator-v6.0.jpg) the (https://ionpowergroup.com/wp-content/uploads/2018/02/Electrical-Power-Differential-Estimator-v5.0.jpg) process.

Ion Power Group is a member of the PEW Clean Energy Business Network

Copyright © 2018, Ion Power Group LLC

How Ion Harvesting can generate clean electricity on Mars

/>

DOCUMENTARY REPORT ABOUT IONPOWERGROUP COM V1 5 fi...

Highly energetic **G**alactic **C**osmic **R**ays (GCRs) arriving from deep space impact Mars' atmosphere on a consistent basis creating electrically charged ions throughout the atmosphere.

Carbon/graphite nanomaterials have been shown to be effective at harvesting electricity from atmospheric ions on Earth. Likewise, it is anticipated that carbon nanomaterials elevated into the Martian atmosphere can also harvest electricity from Mars' ion-rich atmosphere day and night and particularly during dust storms.

On Earth, experiments have demonstrated that high voltage electricity harvested from airborne ions can power lights, motors or produce hydrogen & oxygen gas via water electrolysis or place a charge on lithium-iron (LiFePO4) batteries and be converted to 120VAC/60hZ 'house style' current. Mars has an ion-rich atmosphere believed to be even more advantageous for Ion Harvesting Technology. See 'Concept Clips' for videos.

GCRs penetrate deeply into Mars' thin atmosphere as depicted below.

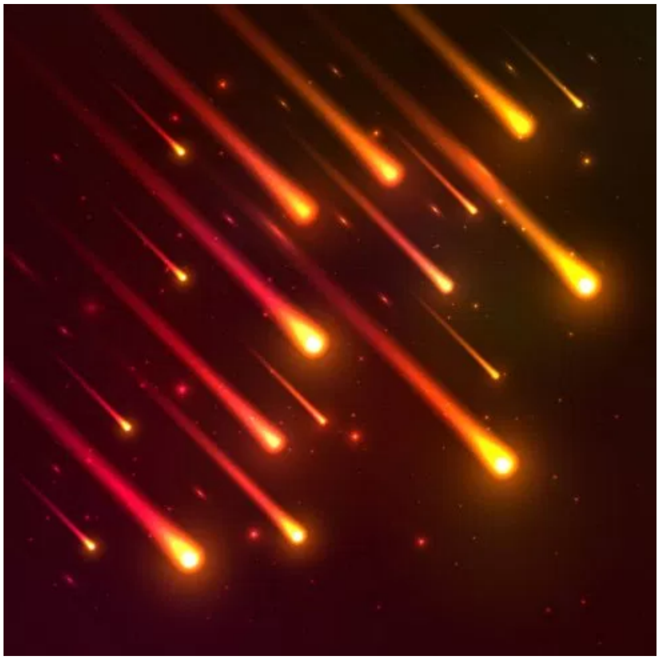

On February 10th 2016, Senior Scientist of Atmospheric Physics Dr. Andreas J .G. Baumgaertner PhD published a technical report titled ‘Power to Mars’. To read the report, click the below Power To Mars report image or read it on Zenodo by clicking here (https://zenodo.org/record/1045600#.Wgm41Bte6Uk). The report recommends that ion harvesting technology be combined with solar panels to provide reliable power day and night for future Mars missions. With a theoretical power generation duty-cycle approaching 99% ion harvesting technology is designed to continue to produce electricity during conditions that cause solar panels to fail such as nighttime and dust storms.

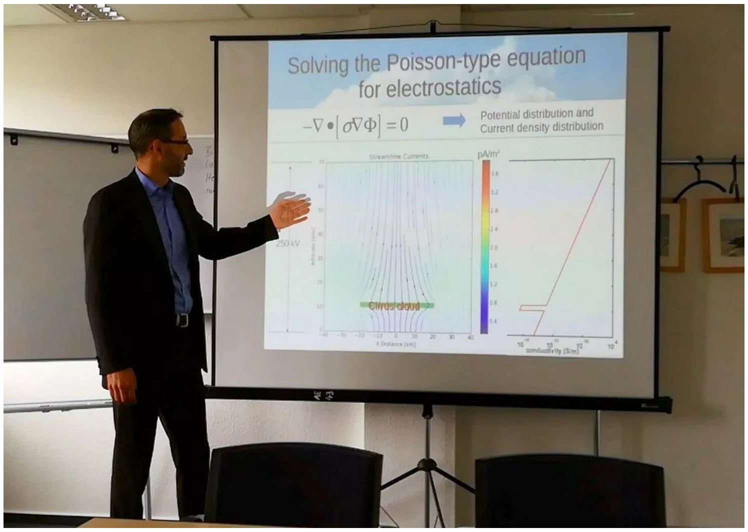

Solving the Poisson-type equation for electrostatics

$-\nabla \bullet [\sigma \nabla \Phi] = 0$ ➡ Potential distribution and Current density distribution

Dr. Andreas J. G. Baumgaertner PhD • Atmospheric Physicist, Senior Scientist – German Aerospace Center • Author of POWER TO MARS

(https://zenodo.org/record/1045600#.Wgm41Bte6Uk)

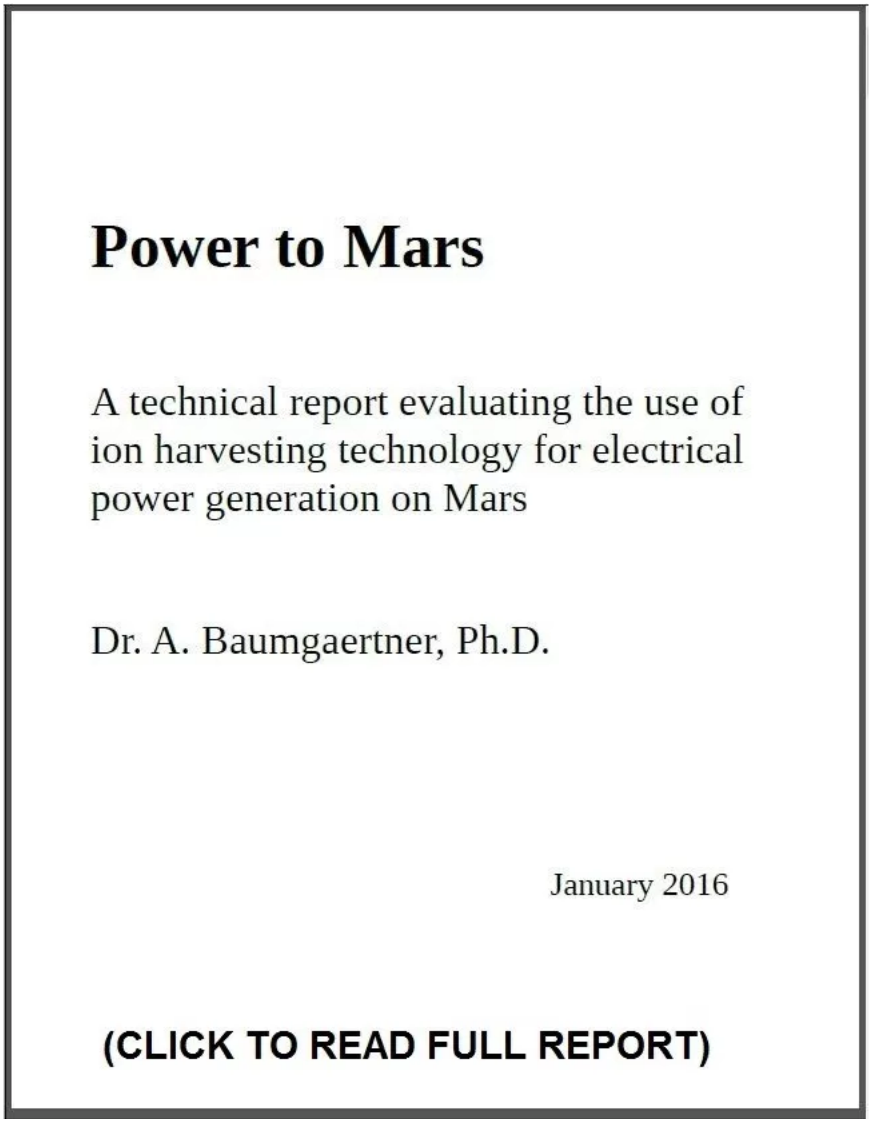

Power to Mars

A technical report evaluating the use of ion harvesting technology for electrical power generation on Mars

Dr. A. Baumgaertner, Ph.D.

January 2016

(CLICK TO READ FULL REPORT)

(https://zenodo.org/record/1045600#.Wgm41Bte6Uk) (https://zenodo.org/record/45877#.WP58tJU2yUk)

GCRs (which are actually microscopic particles traveling at enormous velocity) arriving from deep space impact neutral atoms in the Martian atmosphere knocking electron(s) away producing electrically charged atoms known as ions. Each of the billions of microscopic atmospheric ions acts as a charge carrier conveying a high voltage electric charge. The production of atmospheric ions is a natural process occurring day and night across the entire planet of Mars.

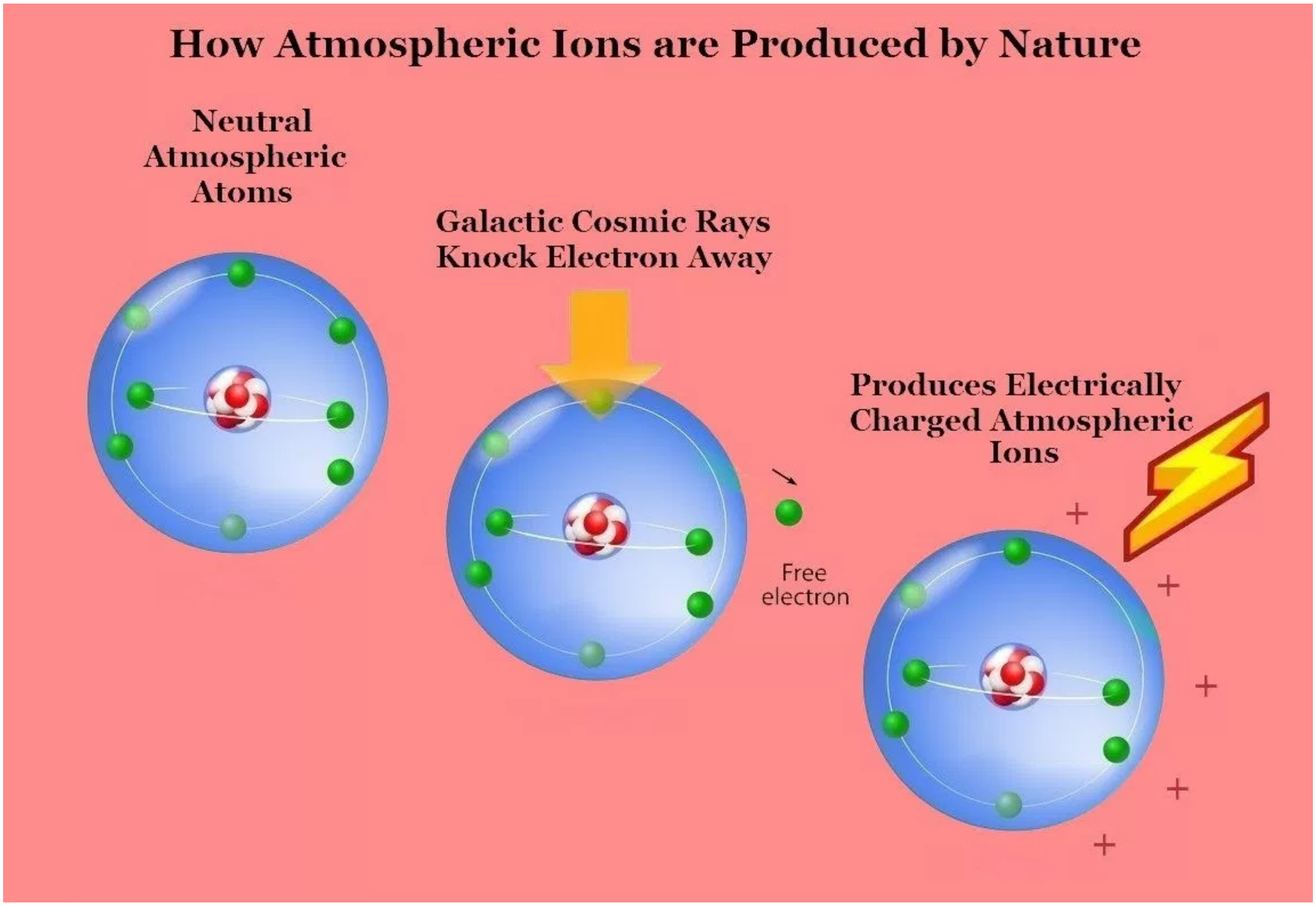
How Atmospheric Ions are Produced by Nature
Neutral Atmospheric Atoms
Galactic Cosmic Rays Knock Electron Away
Free electron
Produces Electrically Charged Atmospheric Ions
+
+
+
+
+

Dust storms occurring on Mars generate strong electrical fields that propagate throughout the Martian atmosphere conveyed by the ions constantly present in the atmosphere. Dust storms can cover large regions of the planet blotting out the sun and can persist for days or weeks. Because the ion-filled Martian atmosphere is electrically conductive, electricity generated by remote dust storms elsewhere on the planet is permitted to propagate throughout the Martian atmosphere. NASA Scientist, Michael Smith, who works at the NASA Goddard Space Flight Center told Space.com "***The dust is electrostatic, like foam peanuts.*"** In an interview with NewScientist.com, Brian Jackson with the Physics Department at Boise State University says **"*It's possible that all of the dust grains clattering together in these storms could produce a lot of electricity…*"** interview seen here (https://www.newscientist.com/article/2143217-marss-surface-hosts-millions-of-towering-dust-devils-every-day/).

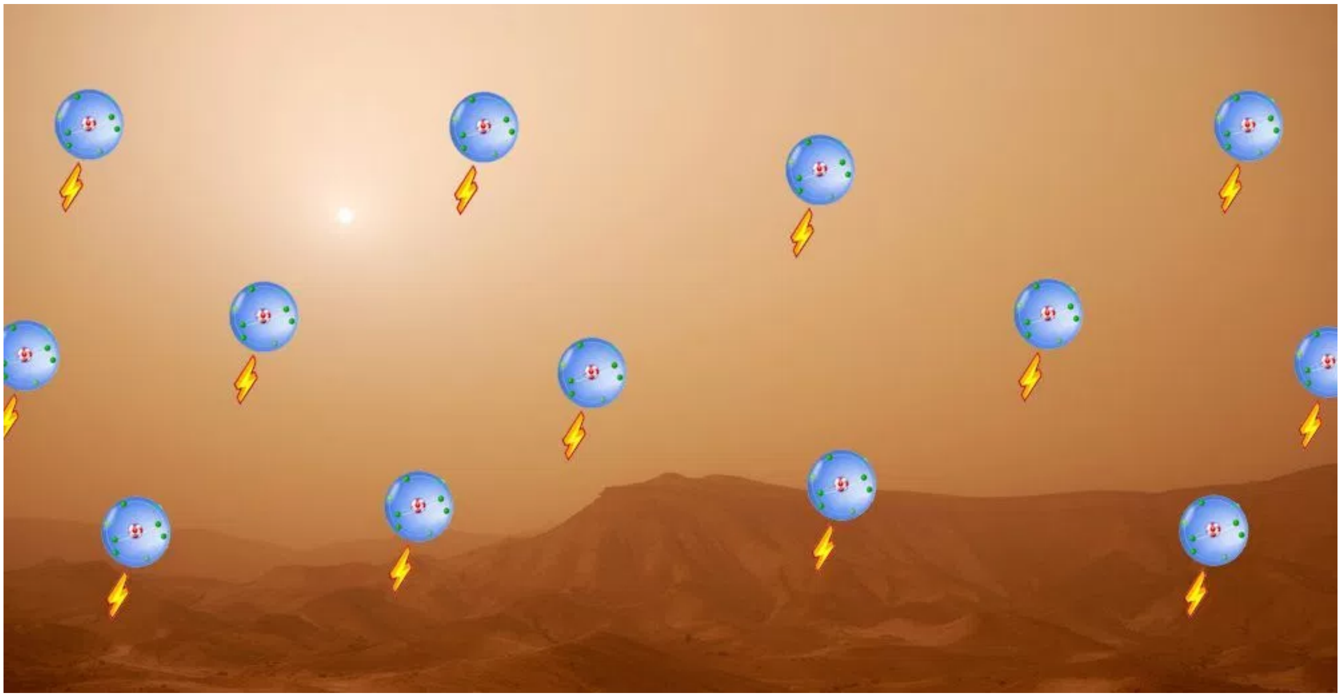

Ion Power Group’s ion collectors, elevated in the Martian atmosphere by tall poles or specially designed aerial platforms (balloons), are designed to harvest the electric charge conveyed by near-Mars ions to produce clean renewable high-voltage electricity day and night and during dust storms as depicted below.

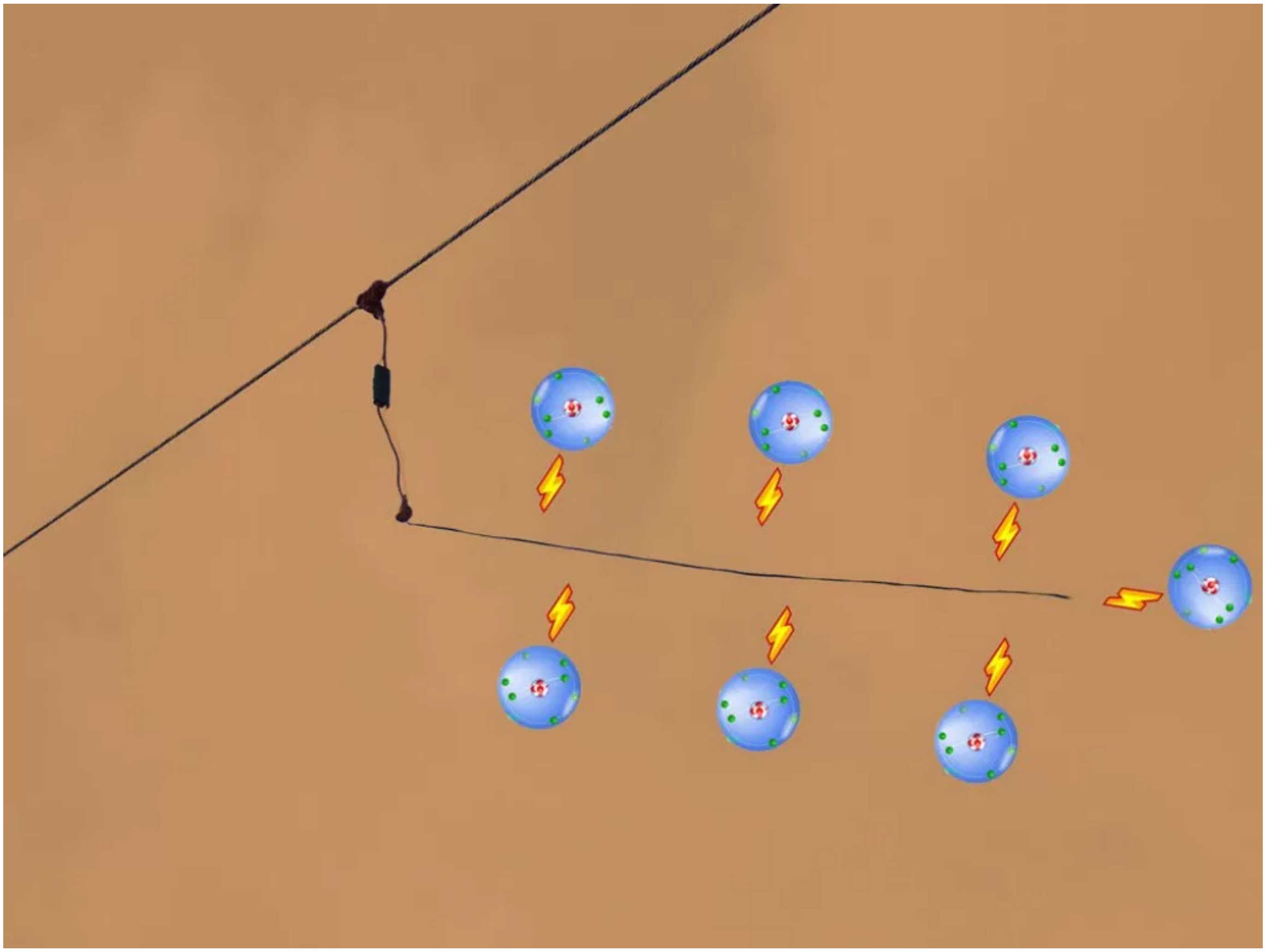

Ion collectors composed of Graphite and/or Graphene, while not metalic, are electrically conductive, flexible and durable. Electric fields are attracted to conductive points, known as the 'Corona Effect'. Ion collectors offer millions of microscopic electrically conductive points and protrusions which readily couple to the electric fields of nearby ions transferring high voltage electricity from the atmospheric ions to the ion collectors. Measurements have shown that ion collectors made of graphite are more efficient at harvesting electricity from airborne ions than tested metals. In some tests, by as much as 55x greater efficiency. Some of the tested metals include highly conductive silver, copper, and aluminum wire – aluminum spheres – copper and aluminum sheets.

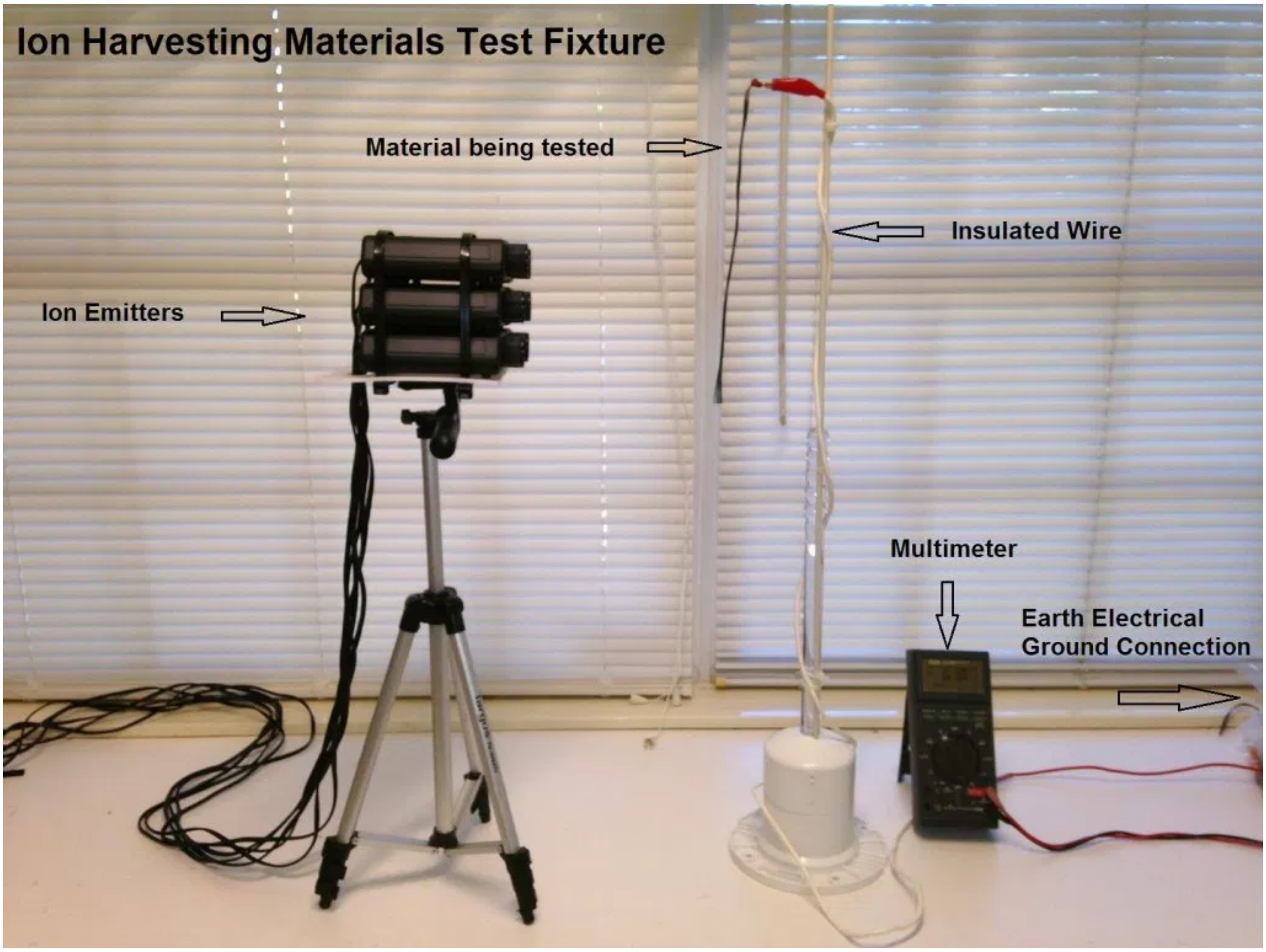
Ion Harvesting Materials Test Fixture
Material being tested
Insulated Wire
Ion Emitters
Multimeter
Earth Electrical Ground Connection

Below is a picture of carbon, a key component of ion collectors, taken at the microscopic level by an electron microscope.

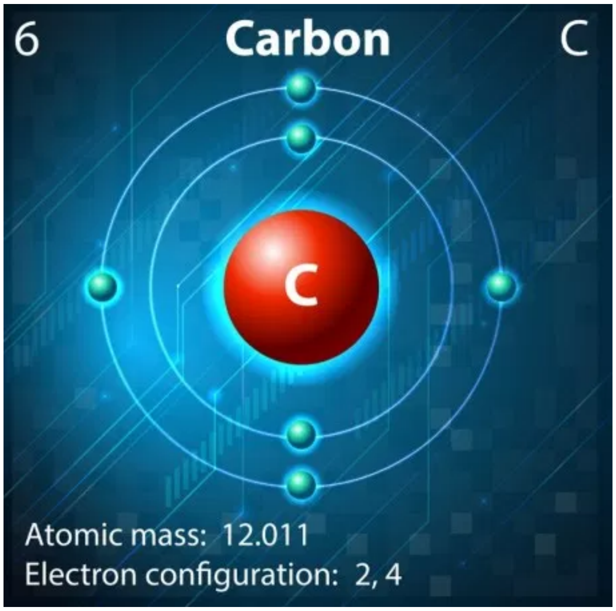
6
Carbon
C
C
Atomic mass: 12.011
Electron configuration: 2, 4

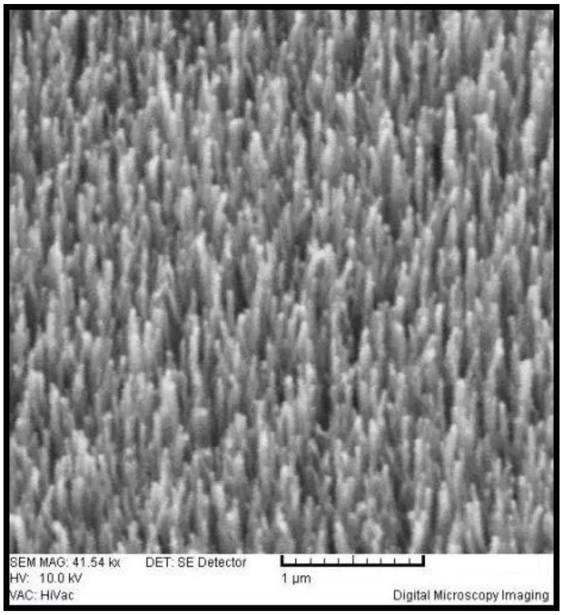
SEM MAG: 41.54 kx
DET: SE Detector
HV: 10.0 kV
VAC: HiVac
1 μm
Digital Microscopy Imaging

Ion Power Group's internationally patented ion collectors, elevated between two poles in the below depiction, are specially designed to couple to the energetic electrically charged ions in the Martian atmosphere. Ion collectors are designed to harvest the high voltage DC electric charge from naturally occurring atmospheric ions day and night and during dust storms and are not dependent on sun or wind. Ground based circuitry has been designed to maintain the ion collector material at a voltage considerably lower than the surrounding atmospheric ions thereby employing the electrostatic attraction principle causing nearby ions to migrate to the ion collectors. Wires will convey the harvested electricity to the base camp where it is routed to a storage medium (such as super capacitors, graphene batteries, lithium-ion batteries, hydrogen gas fuel cells) to be used for life support, light, heat, the production of oxygen for breathing and rocket fuel (hydrogen) from Martian water, communications, fisheries, agriculture and transportation.

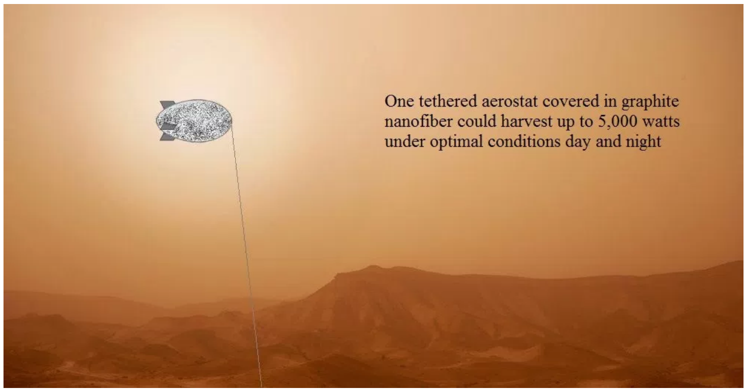
One tethered aerostat covered in graphite
nanofiber could harvest up to 5,000 watts
under optimal conditions day and night

Artists rendering of a tethered aerostat providing altitude for graphite nanofibers to harvest electricity from ions in the Martian atmosphere day and night. The harvested electricity is conveyed down the conductive tether for use on the ground.

Below is a genuine image of Mars' surface. Superimposed is an artist's image of Ion Power Group's ion collectors (thin black material dangling between the two poles) providing electrical power to a conceptual Mars base camp similar to those planned by governments, private and corporate aerospace organizations. Long duration aerial platforms (tethered balloons) may also be used to provide altitude to lightweight ion collectors.

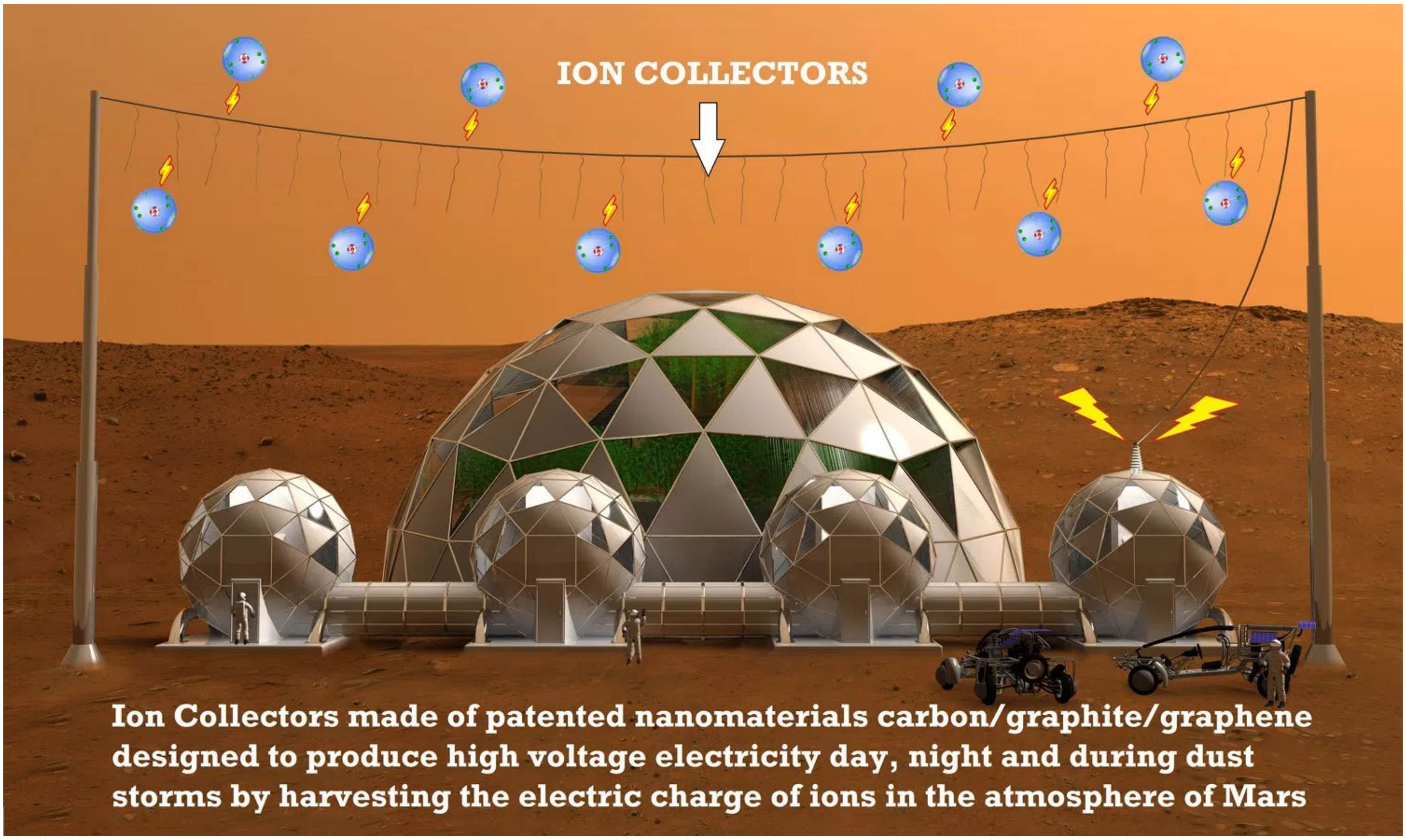
ION COLLECTORS
Ion Collectors made of patented nanomaterials carbon/graphite/graphene designed to produce high voltage electricity day, night and during dust storms by harvesting the electric charge of ions in the atmosphere of Mars

"Solar power… cannot be used at night and during dust storms. It has been shown (in the Power to Mars technical report) that ion power and solar power can be combined to provide a reliable and safe power generation concept." **PhD Atmospheric Physicist and Senior Scientist with the German Aerospace Center, Dr. Andreas J .G. Baumgaertner**

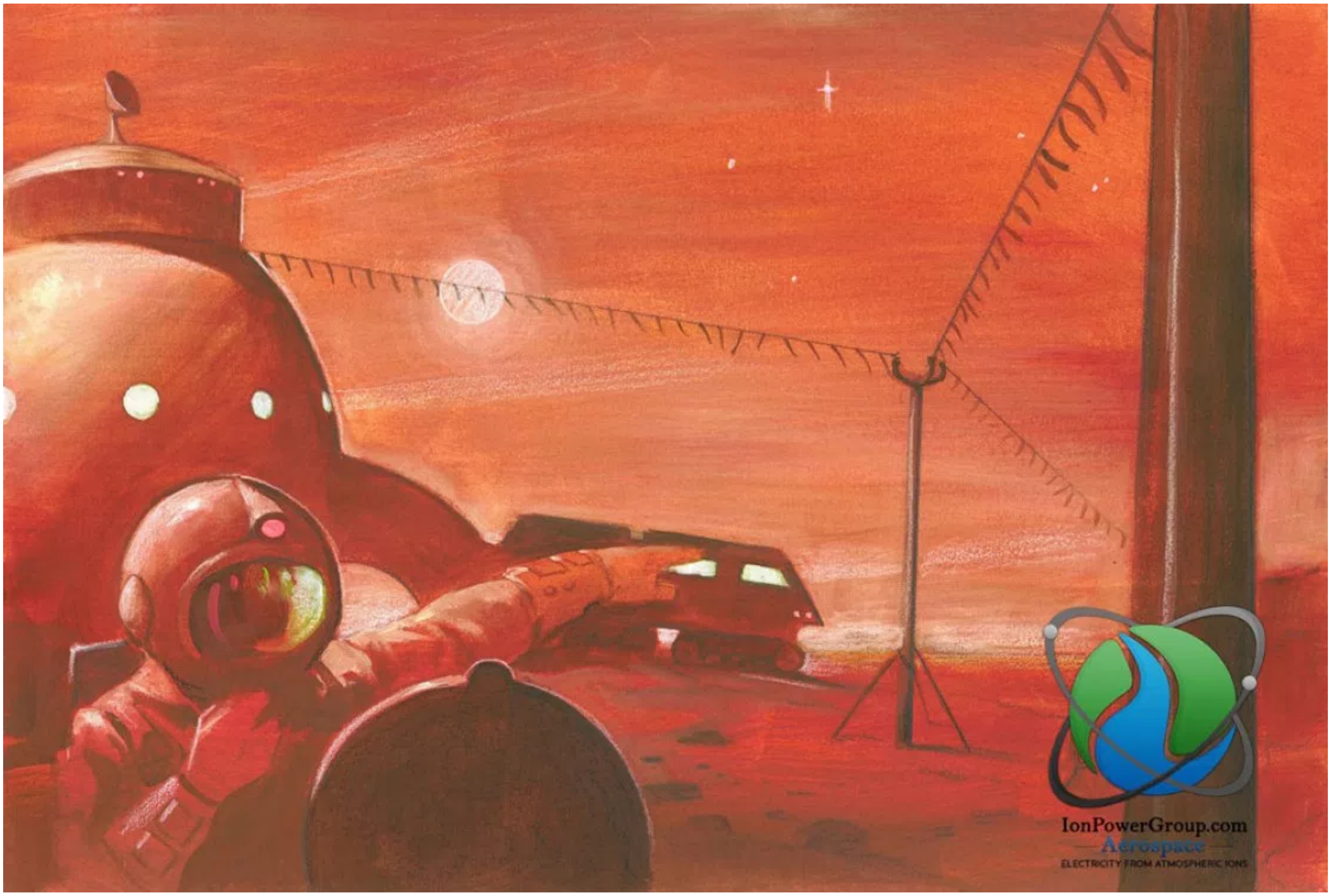

The conditions on Mars including a weak magnetic field, frequent dust storms around the planet and constant bombardment of the Martian atmosphere by GCRs provide an ideal electrical environment for Ion Power Group technology to harvest the electric charge of naturally occurring atmospheric ions to produce reliable, environmentally friendly, electrical power without reliance on sunlight, wind, nuclear or fossil fuels. The highly ionized Martian atmosphere will provide an extremely fertile electrical environment for Ion Power Group's technology to generate clean renewable pollution-free electrical energy day and night and during dust storms. The ability of ion harvesting to generate electricity day/night/dust storms will prove valuable to power robots, rovers, drones, vehicles, living quarters, lights, communications, entertainment equipment, heat to melt Martian ice for drinking, trenchers for extracting water from soil, water for agriculture and fisheries as well as producing hydrogen gas to use for rocket fuel and oxygen for breathing through electrolysis of Martian water.

Nearly 100 countries have some level of space program. A growing number of governments, as well as many private and corporate organizations, are developing plans to send robots and hardware to be followed by people to colonize Mars in the foreseeable future. All of these campaigns will require a reliable source of electrical power on the surface of Mars.

Additional information regarding Ion Harvesting Technology’s Principle-of-Operation can be found on the page titled “How it works on Earth. (/how-it-works-on-earth/)

The survival of humans on Mars requires that the settlers have access to reliable electrical power production, day and night and during the lengthy dust-storms known to occur on Mars. Ion Power Group believes that ion harvesting is the long term solution.

Ion Power Group is a member of the PEW Clean Energy Business Network

Copyright © 2017, Ion Power Group LLC

Strange Apparatus

a gallery of interesting and unusual things

Home | Crookes Radiometer | Stormglass | Crystal Set | Theremin | Rum | Guitar | Contact | Links

Crystal Set

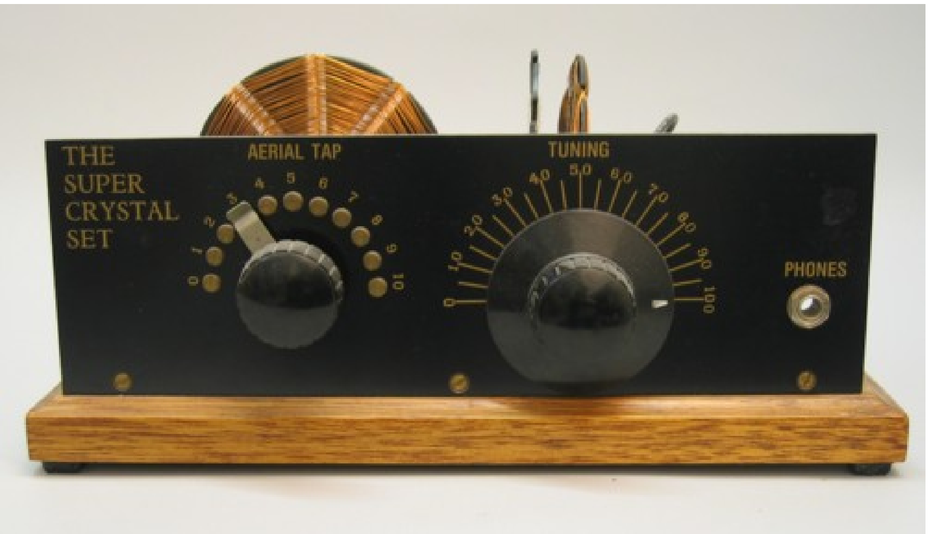

Many years ago I purchased a crystal set kit from Orpheus Radio in Ballarat. The design of the set was based on a 1920's naval radio and featured high-Q spider-web inductors, a variometer, mica capacitors and a rotary inductance selector switch. It is one of the best DX crystal sets I have ever owned. The selectivity and sensitivity are excellent and with this set I have been able to receive many stations from around Australia and south-east Asia. Unfortunately, Orpheus Radio doesn't seem to exist any more so for the crystal set enthusiast wanting to build this excellent radio apparatus, here is a description of that set. For the purist, the germanium diode can be replaced with a cat's whisker detector to produce a totally authentic piece.

Headphones

Before the construction of a crystal set is contemplated, a pair of high quality high-impedance headphones should be obtained. The performance of the crystal set is highly dependant upon the quality of the phones. It is possible (and I have done so) to modify a set of standard 600 ohm headphones to produce an excellent pair of high-impedance phones. The modifications include complete disassembly of the phones and tediously re-winding the coils. This is not a job for the feint-hearted but if your sense of adventure extends this far then the results can be very satisfying.

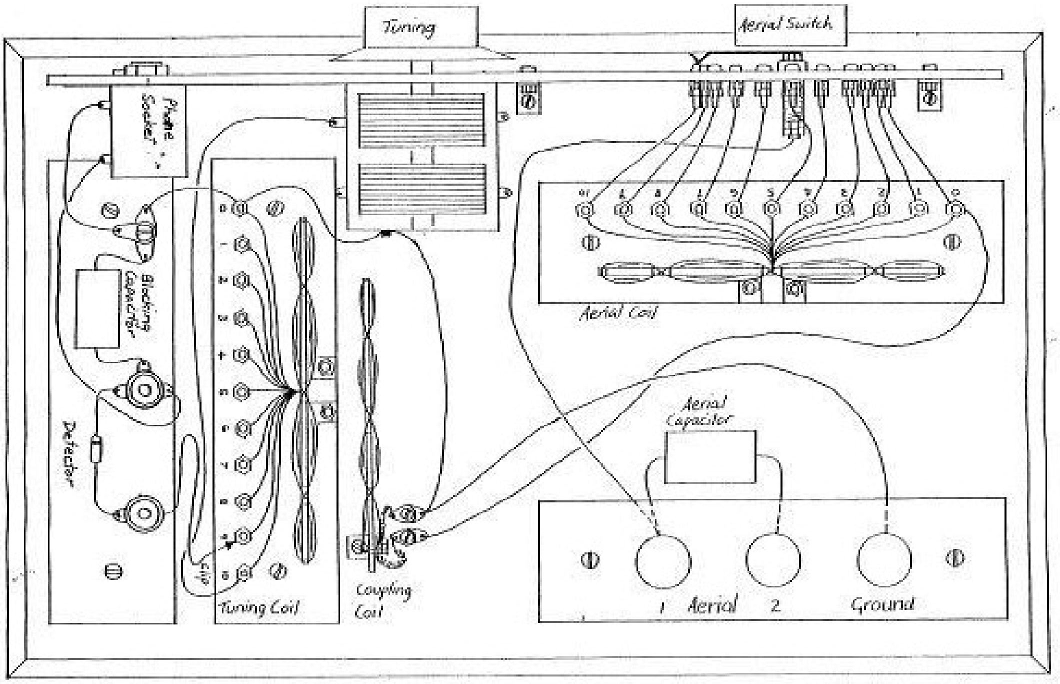
Tuning
Aerial Switch
Phone Socket
Blocking Capacitor
Detector
Clip
Tuning Coil
Coupling Coil
Aerial Coil
Aerial Capacitor
1 Aerial 2
Ground

A sketch showing the layout of the crystal set

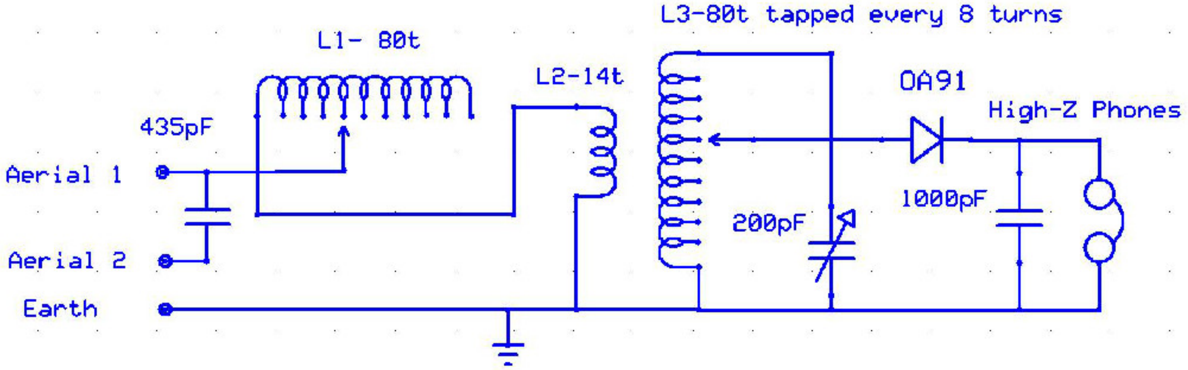
L1- 80t
L2-14t
L3-80t tapped every 8 turns
OA91
High-Z Phones
435pF
Aerial 1
Aerial 2
Earth
200pF
1000pF

Schematic of the crystal set

Circuit Description

The crystal set incorporates an aerial tuning circuit that is quite separate from the main tuning. Altering the aerial switch selects an appropriate tap on L1 to bring the aerial circuit in to tune. The 435pF aerial capacitor C1 will enable the circuit to tune a long aerial. In some cases, it may necessary to reduce the value of this capacitor to enable a very long aerial to be brought into tune.

Following the aerial circuit is the variometer coil L2, used to control the amount of RF coupling to the main tuned circuit. As L2 is pivoted away from the tuning coil L3, the amount of interaction (or mutual inductance) between these coils is reduced. In this way the amount of energy drawn from the aerial circuit is minimized, maintaining sharp tuning of the set.

The main tuning circuit consists of L3 and the tuning capacitor C2. This circuit is energised via the aforementioned variometer and some of its energy is drawn off to feed the detector circuit consisting of D1 and C3. As with the aerial circuit, it is important that the energy drawn from the tuned circuit is minimised. This is achieved by the tap selection on L3. For a given detector arrangement, an optimum tap is selected using the flying lead and alligator clip.

Construction

The coils are wound in the spider-web style and due to the low inter-winding capacitance of this winding arrangement, a coil of very high Q results. The details of the former and winding format are illustrated in the following sketches.

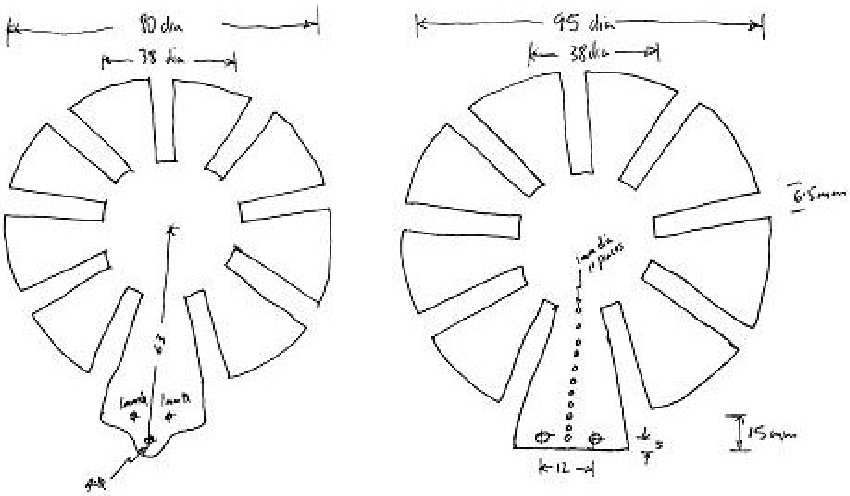

Spider-web coil formers, variometer on the left and aerial and tuning coils on the right

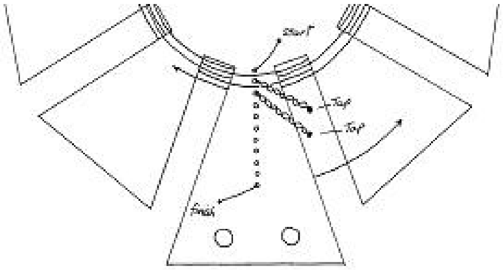

Coil winding details - aerial loading coil (L1) and tuning coil (L3) - 80 turns of 26g B&S enamelled wire with a tap every 8 turns. variometer coil (L2) - 14 turns

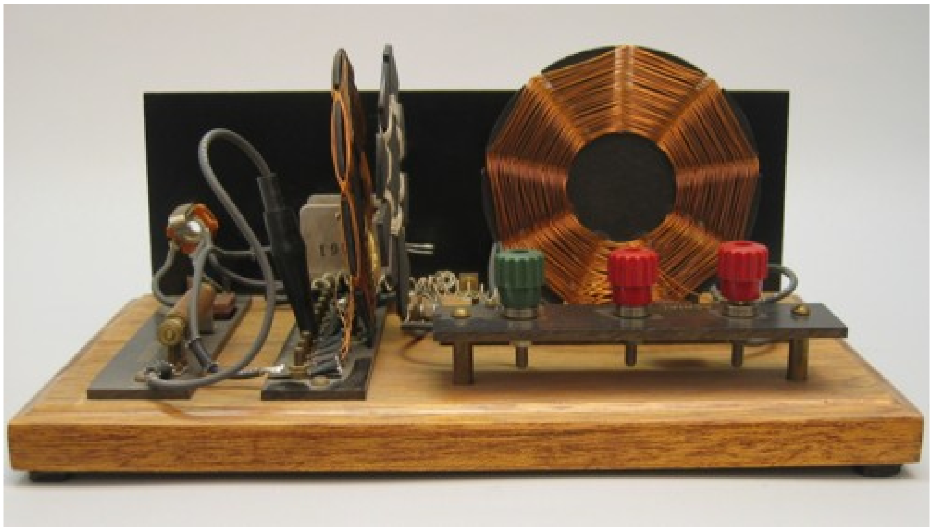

Rear view of the crystal set

Aerial coil taps

The coupling coil (variometer) is pivoted at its mounting bracket to allow the amount of RF coupling to be adjusted by the operator. For best selectivity, it should be adjusted for the least amount of coupling possible. Increased coupling will provide increased volume but the selectivity of the set will deteriorate markedly.

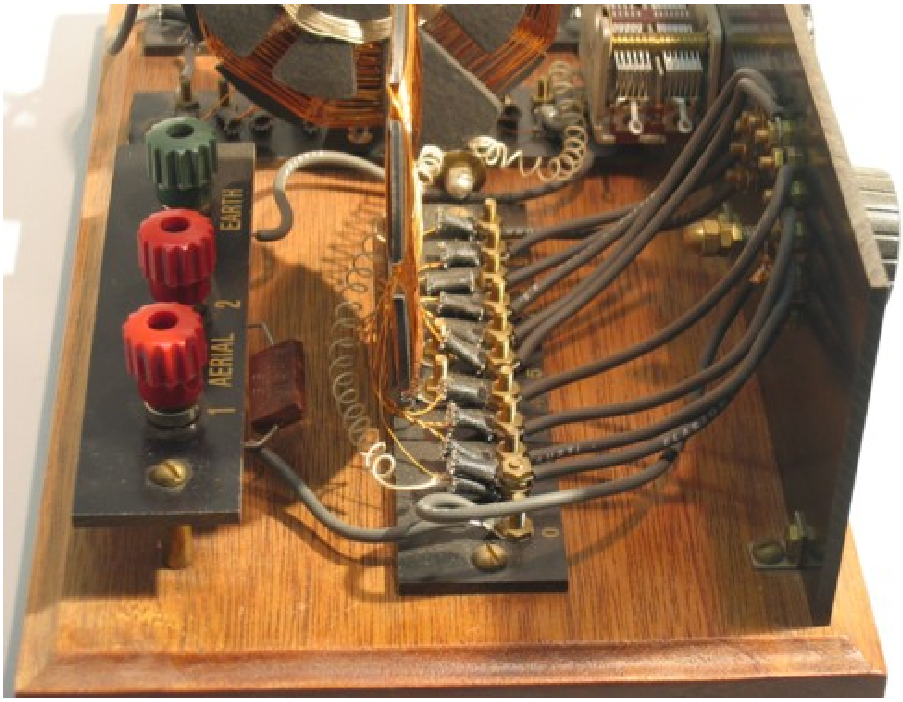

Aerial coil taps

Operation

The set requires a substantial aerial and earth system for best performance. It can be a little tricky to operate as several separate tuning operations are required. One operation selects the appropriate station by turning the tuning knob and the other tunes the aerial circuit by adjusting the aerial taps. Additionally, the variometer and detector tap will need to be adjusted and the correct aerial terminal selected (terminal 1 for short aerials and 2 for long ones). All of this makes for an interesting and challenging pastime. A log book is useful for recording the various adjustments so that an interesting station can be revisited at a later date.

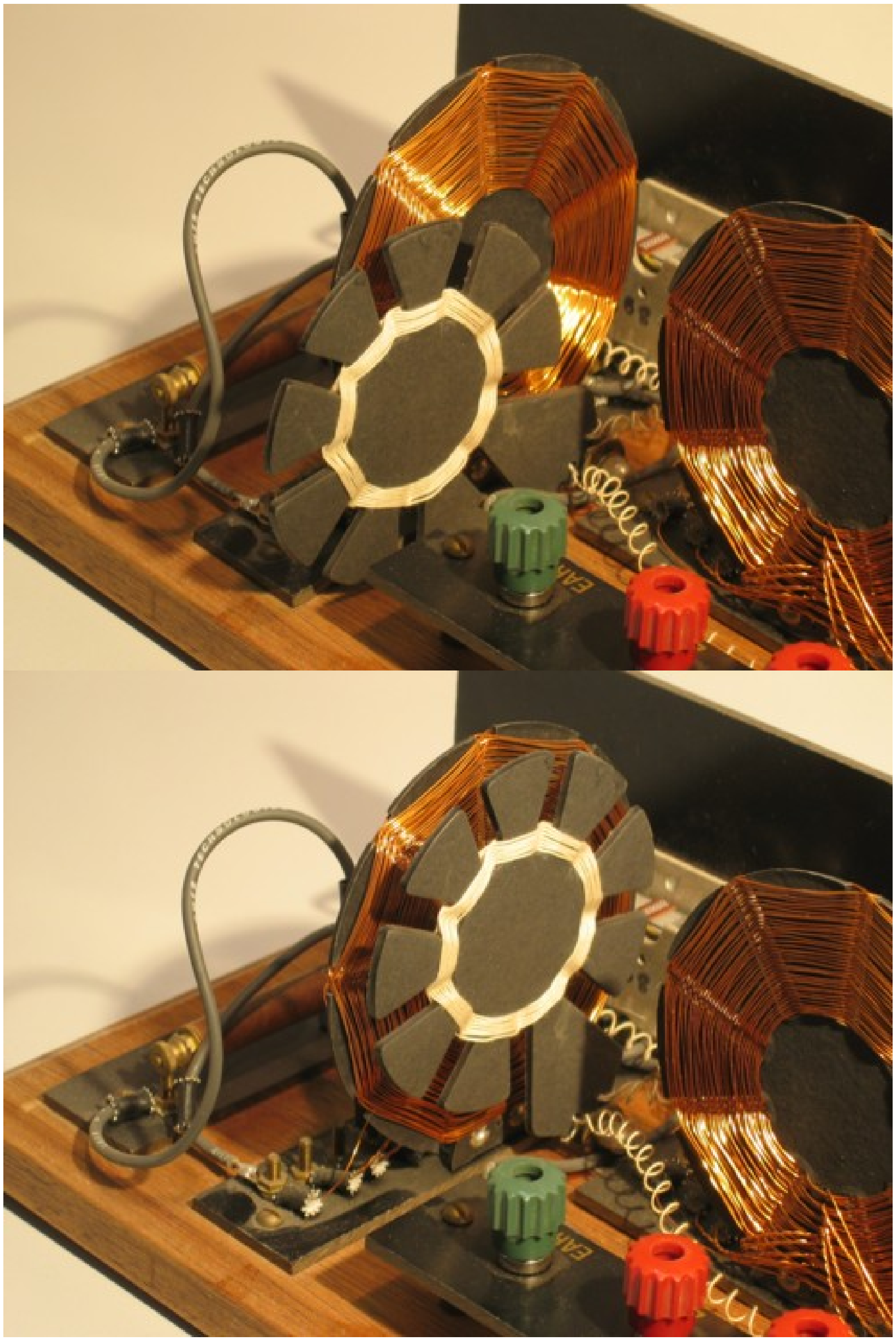

The variometer set for minimum and maximum coupling

To operate the set, connect the aerial and earth, plug in the high-impedance headphones, set the aerial tap to 0 and set the variometer for maximum coupling. It should now be possible to find a station using the tuning control (C2). Once a station is found, reduce the variometer coupling as much as possible while still hearing the station. Now adjust the aerial tap to find a peak. When this is found it should be possible to further reduce the variometer coupling and re-tune for best signal. The detector circuit tap should be set as low as possible to provide the best selectivity of the set.

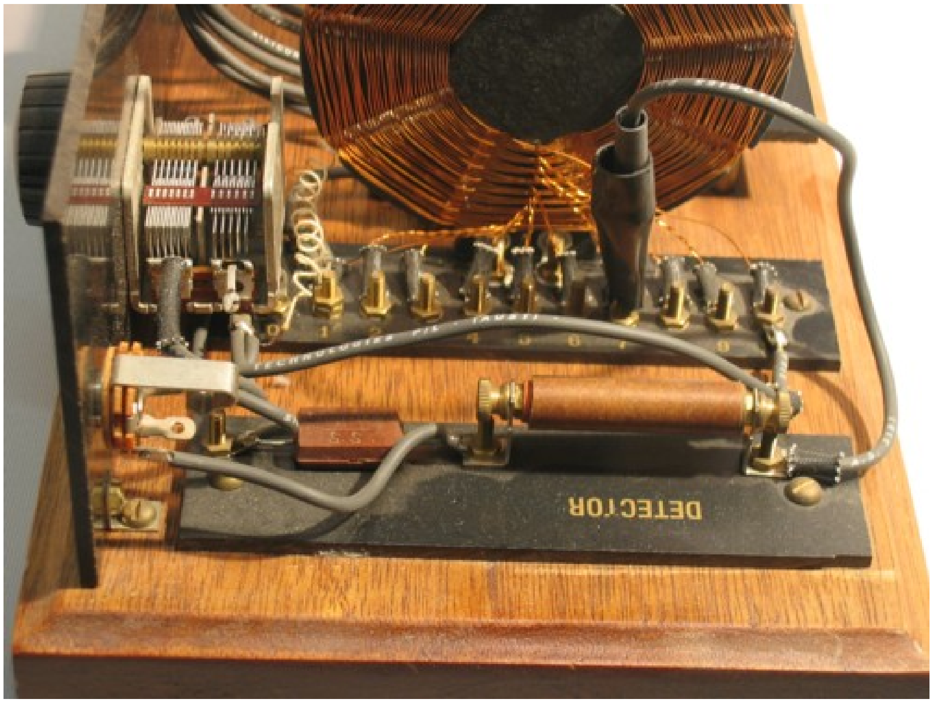

Detector set to tap no. 7

With careful operation of the controls it should be possible to select even quite weak stations from amongst the stronger signals. With a skilled operator, the performance of this crystal radio is quite remarkable.

A Variation on the Design

When cleaning out my shed the other day I came across a set that I had constructed many years ago. It is a variation on the Orpheus design with a small beehive trimmer capacitor used to control the coupling between the tuned circuits in place of the variometer. Also, a different approach to the coil winding design is used.

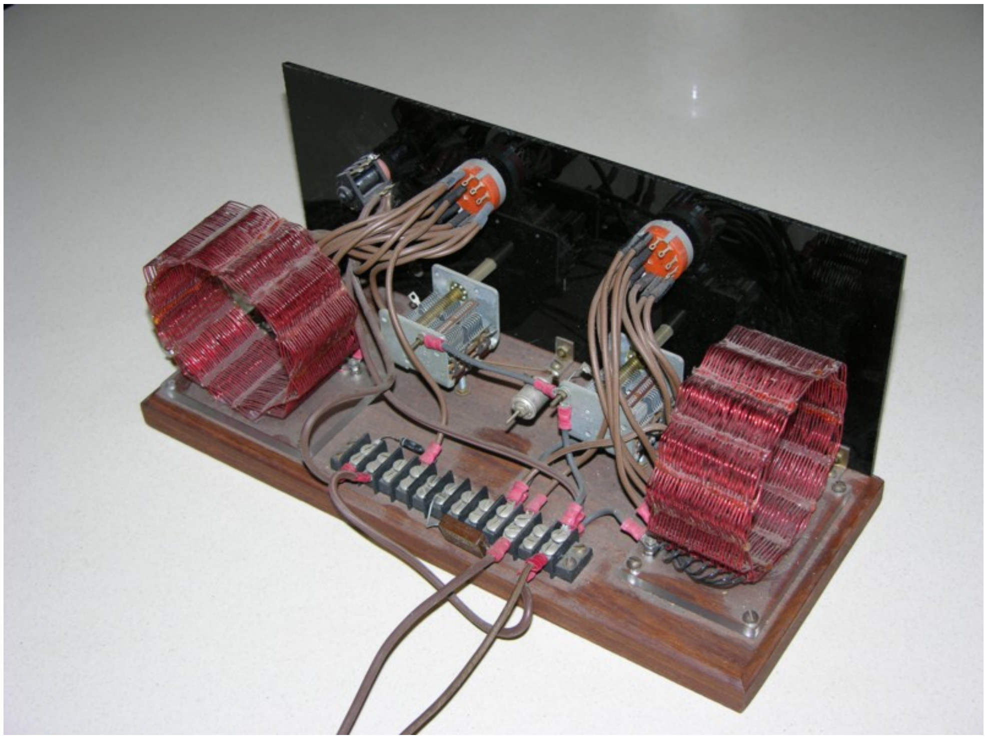

Rear View

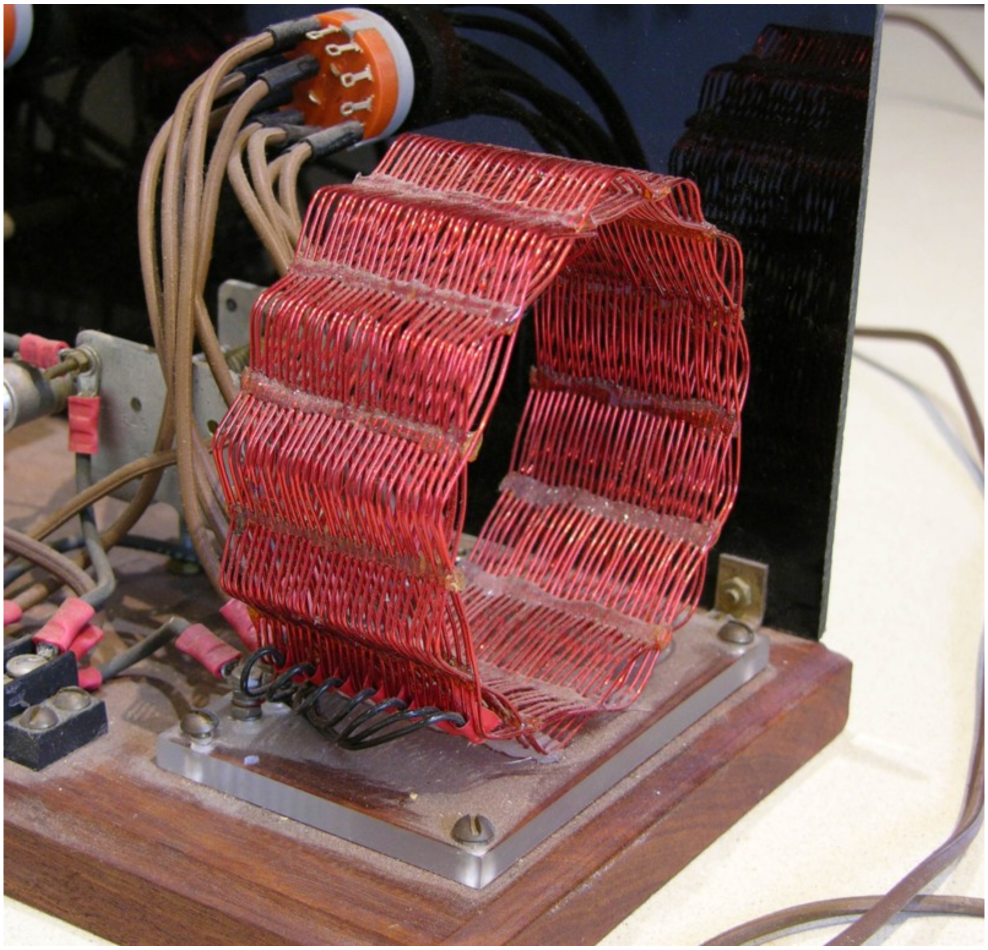

Aerial Coil - Note the alternative method used to reduce inter-winding capacitance.

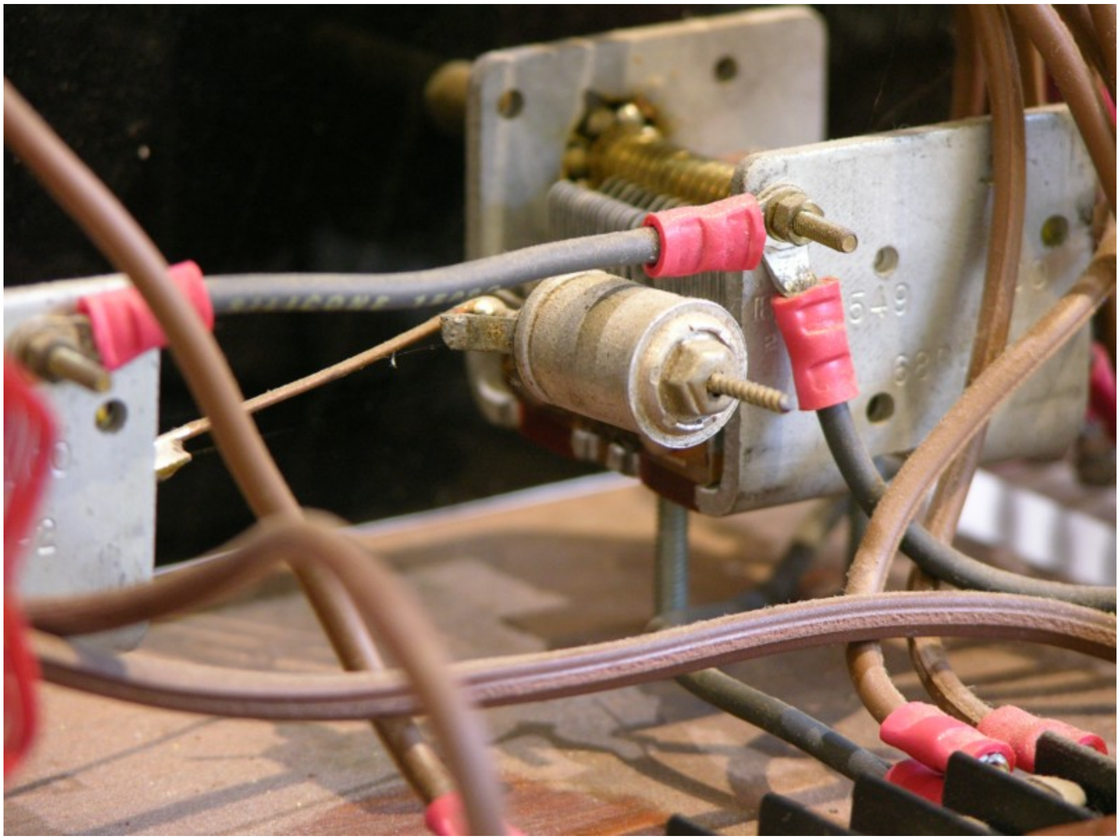

The small trimmer capacitor is used to couple the tuned circuits. Like the design using a variometer, it is adjusted to optimise the trade-off between sensitivity and selectivity.

How to make a batteryless (crystal set) radio

July 21, 2012

CRYSTAL SET

The simplest radio receiver, known as a Crystal Set, consists of nothing more than a coil, tuning capacitor, diode detector, and a pair of earphones. A typical circuit diagram for a Crystal Set Radio is given below where inductor or coil L1 is tuned by variable capacitor VC1 to the transmitter frequency. Diode D1 demodulates the signal, which is fed straight to the earphones. There is no amplification.

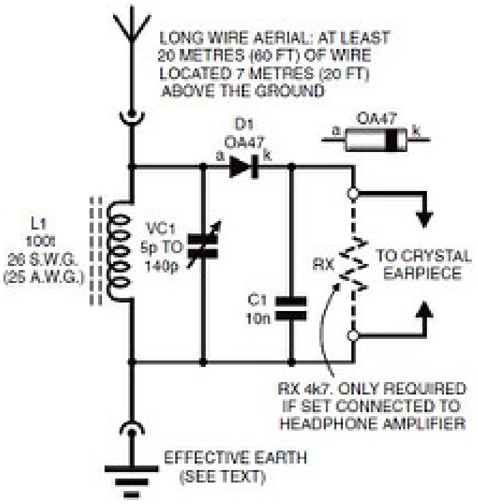
LONG WIRE AERIAL: AT LEAST
20 METRES (60 FT) OF WIRE
LOCATED 7 METRES (20 FT)
ABOVE THE GROUND
D1
OA47
a
k
OA47
a
k
L1
100t
26 S.W.G.
(25 A.W.G.)
VC1
5p TO
140p
RX
TO CRYSTAL
EARPIECE
C1
10n
RX 4k7. ONLY REQUIRED
IF SET CONNECTED TO
HEADPHONE AMPLIFIER
EFFECTIVE EARTH
(SEE TEXT)

CLICK HERE FOR BIGGER PHOTO

** You can use 1N34 Germanium diode in place of OA47*

A long (at least 20 metres), high (17 metres or more) aerial and a good earth (a buried biscuit tin or a metre of copper pipe driven into damp ground) are required in order to ensure audible headphone reception. The earphones originally used with these receivers had an impedance of around 4000 ohms and were very sensitive (and heavy and uncomfortable). They are no longer available, but a crystal earpiece, which relies on the piezoelectric effect, will give acceptable results.

Low impedance ”Walkman” type earphones are NOT suitable.

Component details:

Resistor- RX-4.7k, 0.25 W- only required if set is connect to audio amplifier

Capacitors- C1- 10nF disc ceramic VC1- 5p to 140p, polythene dielectric variable capacitor.

Semiconductors: D1: OA47 or 1N34- Germanium Diode

Miscellaneous: L1- Ferrite Rod, 100mm(4 inch)x9mm/10mm dia., with coil.

Crystal earpiece and jack socket to suit; plastic control knob; plastic insulated flexible cable for aerial wire, downlead and earth connection, 30 meters minimum; buried biscuit tin or 1 meter of copper pipe for earth system; 50gm reel of 26SWG enamelled copper wire, for tuning coil; card and glue for coil former; multistrand connecting wire; crocodile clips or terminals for aerial and earth lead connection; solder, etc.

READ MY EXPERIENCES- HOW EASILY I MADE A BATTERYLESS RADIO

DRAWBACKS

Quite apart from the absence of amplification, two factors seriously limit the performance of crystal receivers. Germanium diodes become increasingly reluctant to conduct as the applied voltage falls below 0.2V, and this makes the receiver insensitive to weak signals. Silicon diodes have a threshold of around 0.6V, and are, therefore, unsuitable for circuits of this kind.

** Silicon diodes like 1N4001 are not suitable*

The earphone loading imposes heavy damping on the tuned circuit, hence, reduces its ability to separate signals. With such low selectivity insensitivity can be a blessing, and crystal sets are normally only capable of receiving a single, strong transmission on the long and medium wavebands. They will sometimes receive more than one if a shortwave coil is fitted.

The aerial and diode can be connected to tappings on the tuning coil in order to reduce damping, but the improvement in selectivity is usually at the expense of audio output. When valves cost a week’s wages and had to be powered by large dry batteries and lead/acid accumulators, the construction of simple receivers of this kind could be justified. With high performance transistors now costing only a few pence or cents, crystal sets are now regarded as ‘nostalgic pieces”. Some readers may, however, wish to build one out of curiosity, or for the novelty of having a receiver that **does not require a power supply.**

CIRCUIT DETAILS

Ferrite loop aerial L1 and polythene dielectric variable capacitor VC1 form the tuned circuit. Point contact germanium diode D1 (1N34 or OA47) demodulates the signal; capacitor C1 bypasses residual r.f. (radio frequency) to earth and also exhibits a reservoir action, enabling the a.f. (audio frequency) output to approach its peak value. The recovered audio signal is fed directly to a crystal earpiece. Signal voltages introduced in the ferrite loop aerial by the radiated magnetic field are much too much to produce an output from the detector, and the component is used here simply as a tuning coil. The ferrite core does, however, reduce the number of turns required for the coil winding, thereby reducing its resistance and increasing its audio quality.

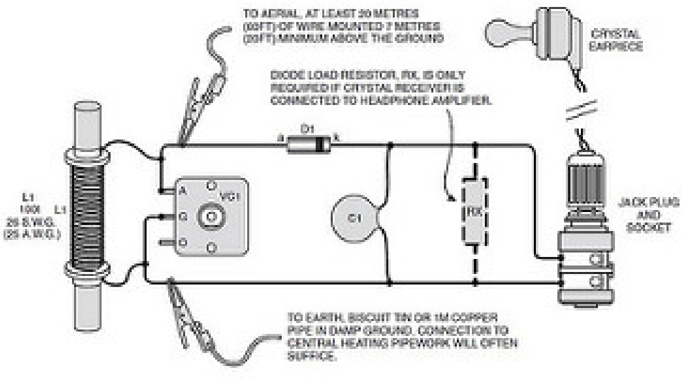
TO AERIAL, AT LEAST 20 METRES (66FT) OF WIRE MOUNTED 7 METRES (23FT) MINIMUM ABOVE THE GROUND
CRYSTAL EARPIECE
DIODE LOAD RESISTOR, RX, IS ONLY REQUIRED IF CRYSTAL RECEIVER IS CONNECTED TO HEADPHONE AMPLIFIER.
D1
a
k
A
VC1
G
O
L1
100t
L1
26 S.W.G.
(25 A.W.G.)
C1
RX
JACK PLUG AND SOCKET
TO EARTH, BISCUIT TIN OR 1M COPPER PIPE IN DAMP GROUND. CONNECTION TO CENTRAL HEATING PIPEWORK WILL OFTEN SUFFICE.

GET BIGGER PICTURE

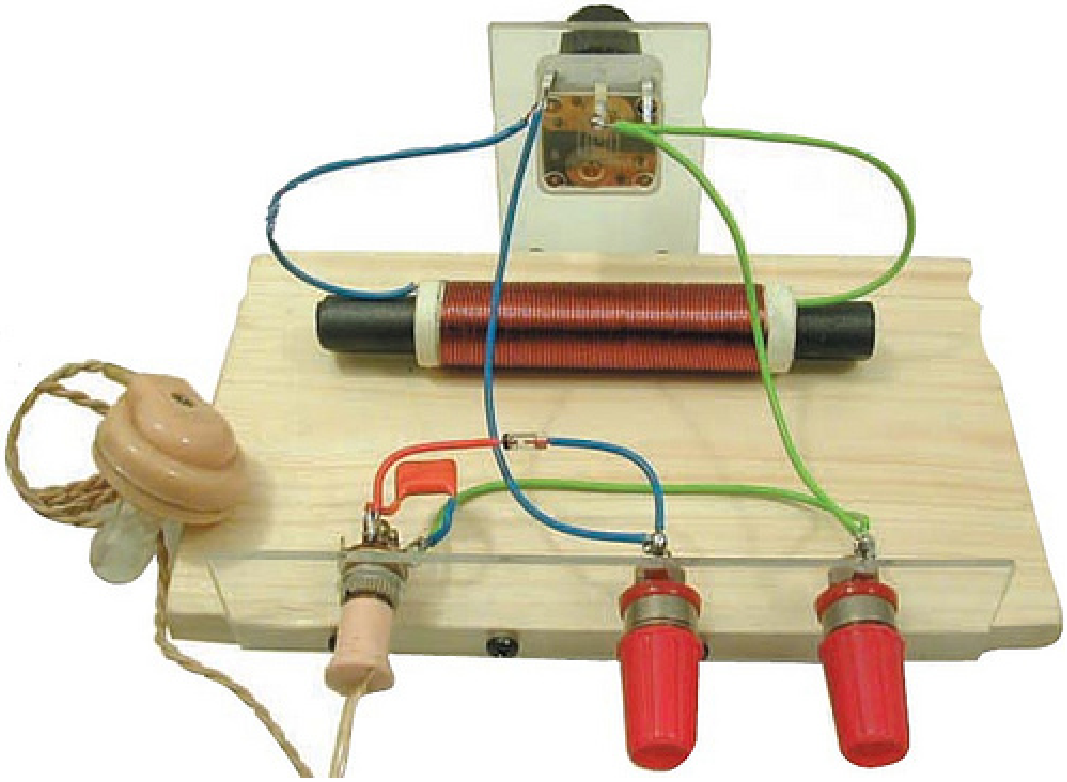

COIL DETAILS

Full construction and winding details for the ferrite tuning/ aerial coil L1 are shown in figure.

The coil is made from 26 SWG(25 AWG) enamelled copper wire, close wound on a cardboard former.

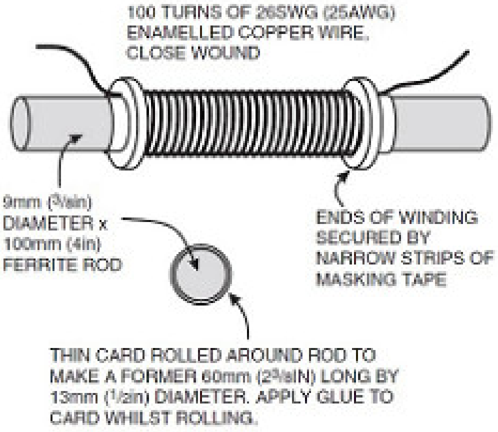
100 TURNS OF 26SWG (25AWG) ENAMELLED COPPER WIRE, CLOSE WOUND
9mm (3/8in) DIAMETER x 100mm (4in) FERRITE ROD
ENDS OF WINDING SECURED BY NARROW STRIPS OF MASKING TAPE
THIN CARD ROLLED AROUND ROD TO MAKE A FORMER 60mm (2 3/8IN) LONG BY 13mm (1/2in) DIAMETER. APPLY GLUE TO CARD WHILST ROLLING.

The r.f. bypass capacitor C1 can in practice, be omitted with no noticeable reduction in performance. However, if the set is to be connected to either the headphone amplifier or speaker amplifier decribed next month, the component together with diode, load resistor, RX, must be included.

AMPLIFICATION Audio frequency amplification, after the diode detector will permit the use of low impedance Walkman type earphones or even loud speaker operation. It will do nothing, however, to overcome the diode's insensitivity to weak signals. For this we must have radio frequency amplification of the signals picked up by the aerial before they reach the detector(The standard circuit for a transistor portable receiver has three stages of radio frequency amplification ahead of the diode).

For amplification, you can use a simple audio amplifier using LM386.

READ MY EXPERIENCES ON MAKING A BATTERYLESS RADIO

WATCH THESE VIDEOS:

BuildCircuit proudly recommends EasyEDA for circuit and PCB design

EasyEDA: A Powerful Free Circuit, Simulation & PCB Design Tool

Register now to use it for free. No Need to download. Lots of resources and Step by step tutorials

310-12 SR

F462A OR 4,151,431

United States Patent [19]

Johnson

[11] **4,151,431**

[45] **Apr. 24, 1979**

[54] **PERMANENT MAGNET MOTOR**

[76] Inventor: **Howard R. Johnson,** 3300 Mt. Hope Rd., Grass Lake, Mich. 49240

[21] Appl. No.: **422,306**

[22] Filed: **Dec. 6, 1973**

[51] **Int. Cl.**^{2} **H02K 41/00;** H02N 11/00

[52] **U.S. Cl.** **310/12;** 310/152

[58] **Field of Search** .... 24/DIG. 9; 415/DIG. 2; 46/236; 273/118 A, 119 A, 120 A, 121 A, 122 A, 123 A, 124, 125 A, 126 A, 130 A, 131 A, 131 AD, 134 A, 135 A, 136 B, 137 AE, 138 A

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,074,153	2/1978	Baker et al.	310/12

Primary Examiner—Donovan F. Duggan
Attorney, Agent, or Firm—Beaman & Beaman

[57] **ABSTRACT**

The invention is directed to the method of utilizing the unpaired electron spins in ferro magnetic and other materials as a source of magnetic fields for producing power without any electron flow as occurs in normal conductors, and to permanent magnet motors for utilizing this method to produce a power source. In the practice of the invention the unpaired electron spins occurring within permanent magnets are utilized to produce a motive power source solely through the superconducting characteristics of a permanent magnet and the magnetic flux created by the magnets are controlled and concentrated to orient the magnetic forces generated in such a manner to do useful continuous work, such as the displacement of a rotor with respect to a stator. The timing and orientation of magnetic forces at the rotor and stator components produced by permanent magnets to produce a motor is accomplished with the proper geometrical relationship of these components.

28 Claims, 10 Drawing Figures

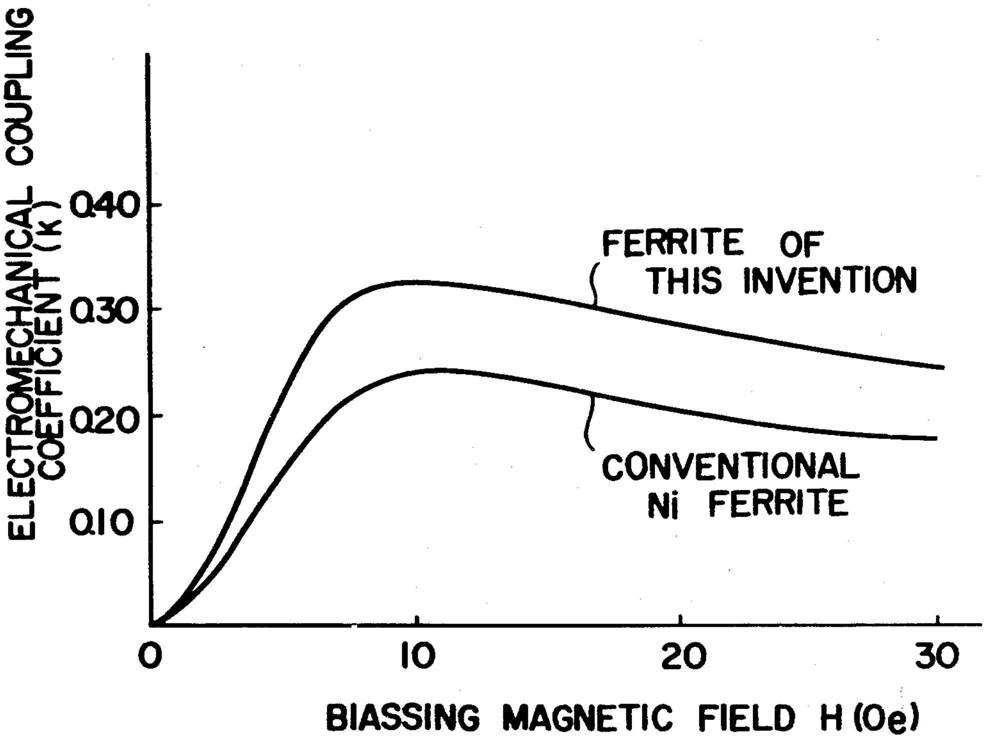

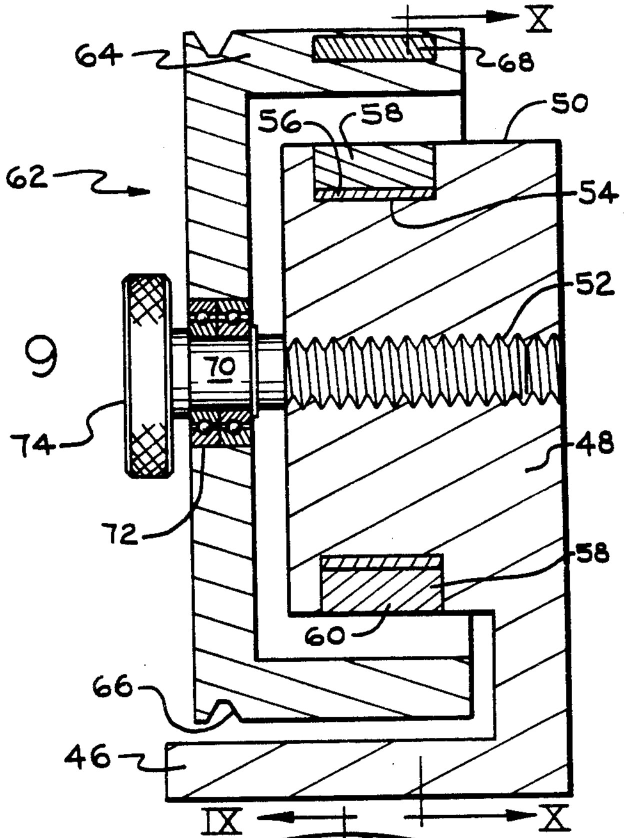
X
64
68
50
56
58
62
54
52
70
74
48
72
58
60
66
46
IX
X

FIG. 9

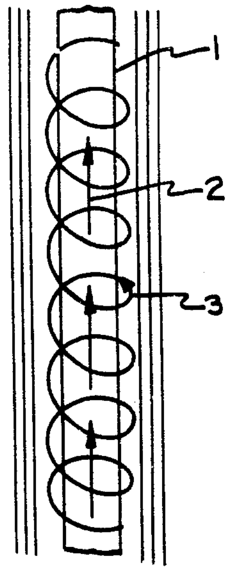

FIG. 1

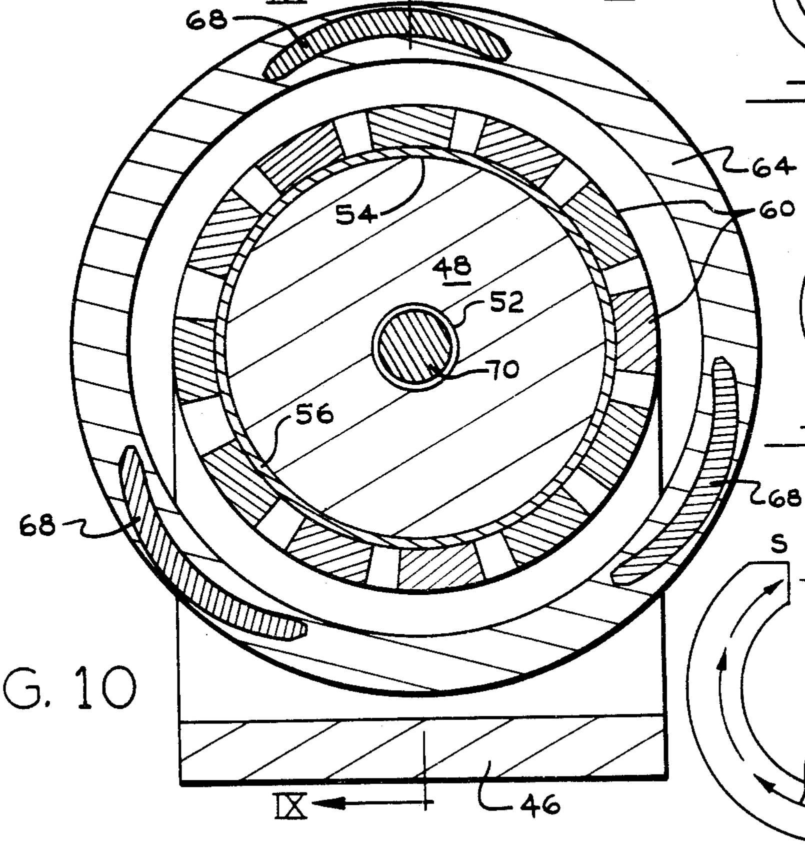
68
64
54
60
48
52
70
56
68
68
46
IX

FIG. 10

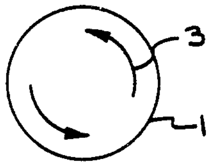

FIG. 2

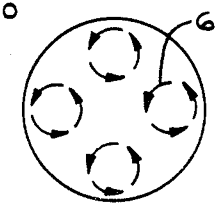

FIG. 4

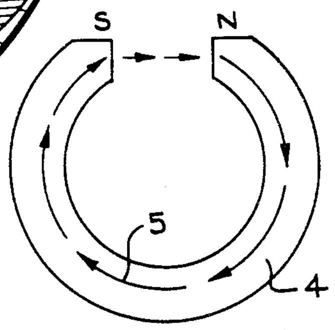

FIG. 3

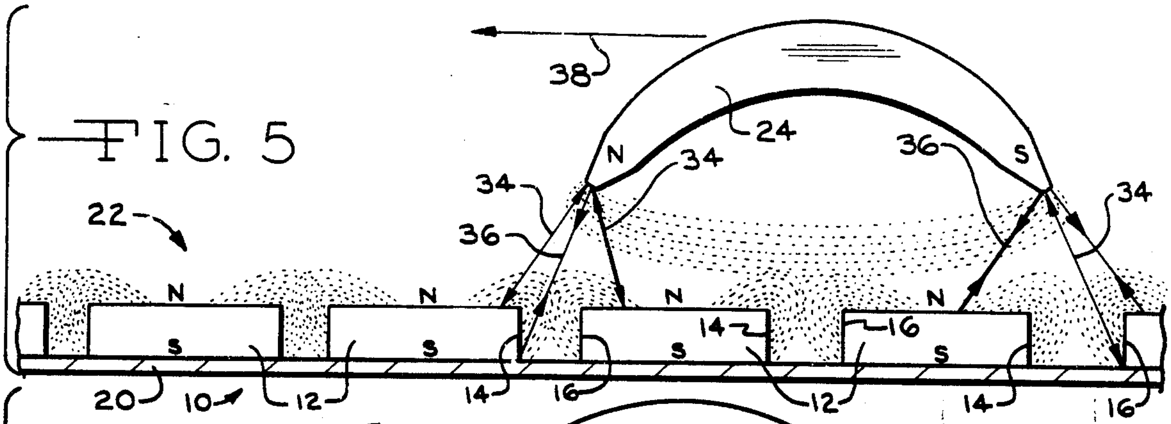
FIG. 5
38
24
N
S
34
34
36
34
36
22
N
N
N
N
S
S
S
S
14
16
20
10
12
12
14
16
12
14
16

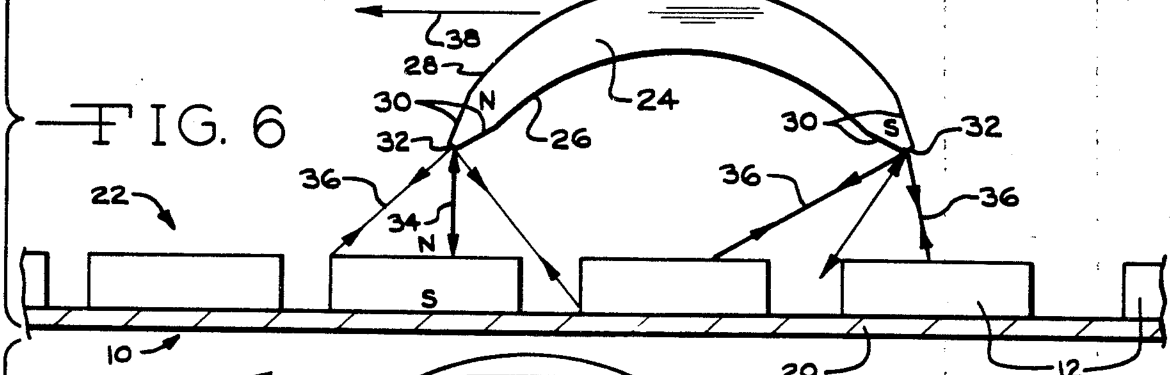
FIG. 6
38
28
24
30
N
26
30
S
32
32
36
36
36
34
22
N
S
10
20
12

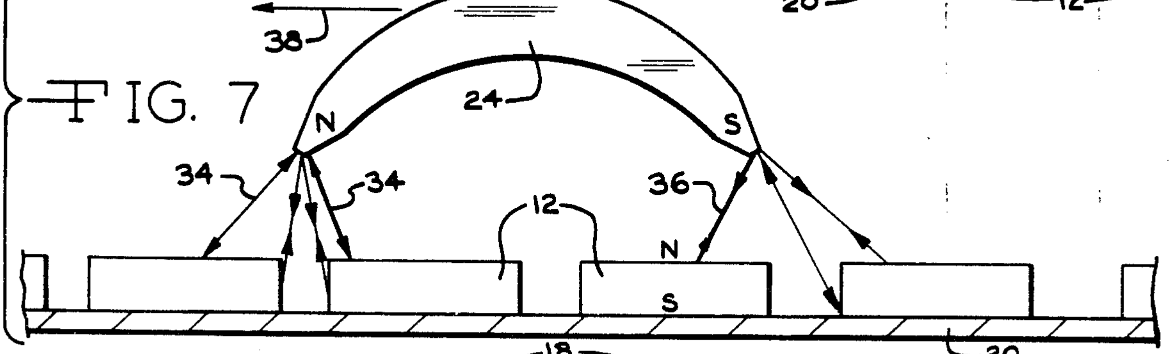
FIG. 7
38
24
N
S
34
34
36
12
N
S
20

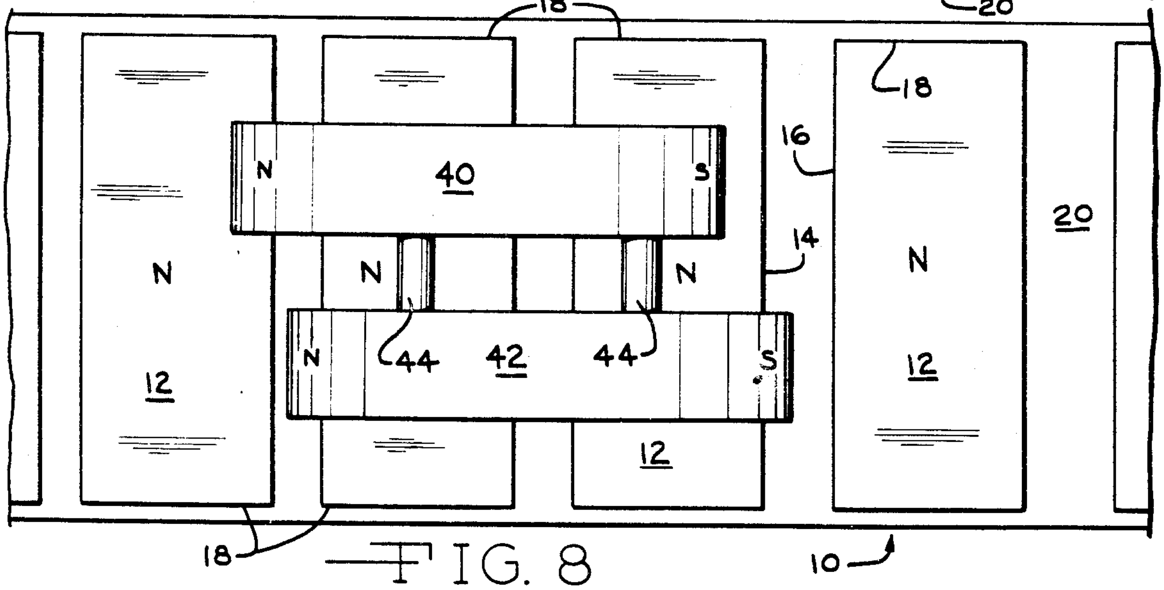
FIG. 8
18
18
16
N
40
S
20
14
N
N
N
N
N
44
42
44
S
12
12
12
18
10

PERMANENT MAGNET MOTOR

FIELD OF THE INVENTION

The invention pertains to the field of permanent magnet motor devices solely using the magnetic fields created thereby to product motive power.

BACKGROUND OF THE INVENTION

Conventional electric motors employ magnetic forces to produce either rotative or linear motion. Electric motors operate on the principle that when a conductor is located in a magnetic field which carries current a magnetic force is exerted upon it.

Normally, in a conventional electric motor, the rotor, or stator, or both, are so wired that magnetic fields created by electromagnetics may employ attraction, repulsion, or both types of magnetic forces, to impose a force upon the armature to cause rotation, or to cause the armature to be displaced in a linear path. Conventional electric motors may employ permanent magnets either in the armature or stator components, but in the art heretofore known the use of permanent magnets in either the stator or armature require the creation of an electromagnetic field to act upon the field produced by the permanent magnets, and switching means are employed to control the energization of the electromagnets and the orientation of the magnetic fields, to produce the motive power.

It is my belief that the full potential of magnetic forces existing in permanent magnets has not been recognized or utilized because of incomplete information and theory with respect to the atomic motion occurring within a permanent magnet. It is my belief that a presently unnamed atomic particle is associated with the electron movement of a superconducting electromagnet and the lossless current flow of Amperian currents in permanent magnets. The unpaired electron flow is similar in both situations. This small particle is believed to be opposite in charge and to be located at right angles to the moving electron, and the particle would be very small as to penetrate all known elements, in their various states as well as their known compounds, unless they have unpaired electrons which capture these particles as they endeavor to pass therethrough.

Ferro electrons differ from those of most elements in that they are unpaired, and being unpaired they spin around the nucleus in such a way that they respond to magnetic fields as well as creating one themselves. If they were paired, their magnetic fields would cancel out. However, being unpaired they create a measurable magnetic field if their spins have been oriented in one direction. The spins are at right angles to their magnetic fields.

In niobium superconductors at a critical state, the magnetic lines of force cease to be at right angles. This change must be due to establishing the required conditions for unpaired electronic spins instead of electron flow in the conductor, and the fact that very powerful electromagnets that can be formed with superconductors illustrates the tremendous advantage of producing the magnetic field by unpaired electron spins rather than conventional electron flow.

In a superconducting metal, wherein the electrical resistance becomes greater in the metal than the proton resistance, the flow turns to electron spins and the positive particles flow parallel in the metal in the manner occurring in a permanent magnet where a powerful flow of magnetic positive particles or magnetic flux causes the unpaired electrons to spin at right angles. Under cryogenic superconduction conditions the freezing of the crystals in place makes it possible for the spins to continue, and in a permanent magnet the grain orientation of the magnetized material results in the spins permitting them to continue and for the flux to flow parallel to the metal.

In a superconductor, at first the electron is flowing and the positive particle is spinning; later, when critical, the reverse occurs, i.e., the electron is spinning and the positive particle is flowing at right angles. These positive particles will thread or work their way through the electron spins present in the metal.

In a sense, a permanent magnet may be considered the only room temperature superconductor. It is a superconductor because the electron flow does not cease, and this electron flow can be made to do work because of the magnetic field it supplies. Previously, this source of power has not been used because it was not possible to modify the electron flow to accomplish the switching functions of the magnetic field. Such switching functions are common in a conventional electric motor where electrical current is employed to align the much greater electron current in the iron pole pieces and concentrate the magnetic field at the proper places to give the thrust necessary to move the motor armature. In a conventional electric motor, switching is accomplished by the use of brushes, commutators, alternating current, or other known means.

In order to accomplish the switching function in a permanent magnet motor, it is necessary to shield the magnetic leakage so that it will not appear as too great a loss factor at the wrong places. The best method to accomplish this is to use the superconductor of magnetic flux and concentrate it to the place where it will be the most effective. Timing and switching can be achieved in a permanent magnet motor by concentrating the flux and using the proper geometry of the motor rotor and stator to make most effective use of the magnetic fields generated by the electron spins. By the proper combination of materials, geometry and magnetic concentration, it is possible to achieve a mechanical advantage of high ratio, greater than 100 to 1, capable of producing a continuous motive force.

To my knowledge, previous work done with permanent magnets, and motive devices utilizing permanent magnets, have not achieved the result desired in the practice of the inventive concept, and it is with the proper combination of materials, geometry and magnetic concentration that the presence of the magnetic spins within a permanent magnet may be utilized as a motive force.

SUMMARY OF THE INVENTION

It is an object of the invention to utilize the magnetic spinning phenomenon of unpaired electrons occurring in ferro magnetic material to produce the movement of a mass in a unidirectional manner as to permit a motor to be driven solely by magnetic forces as occurring within permanent magnets. In the practice of the inventive concepts, motors of either linear or rotative types may be produced.

It is an object of the invention to provide the proper combination of materials, geometry and magnetic concentration to utilize the force generated by unpaired electron spins existing in permanent magnets to power a motor. Whether the motor constitutes a linear embodiment, or a rotary embodiment, in each instance the "stator" may consist of a plurality of permanent magnets fixed relative to each other in space relationship to define a track, linear in form in the linear embodiment, and circular in form in the rotary embodiment. An armature magnet is located in spaced relationship to such track defined by the stator magnets wherein an air gap exists therebetween. The length of the armature magnet is defined by poles of opposite polarity, and the length of the armature magnet is disposed relative to the track defined by the stator magnets in the direction of the path of movement of the armature magnet as displaced by the magnetic forces.

The stator magnets are so mounted that poles of like polarity are disposed toward the armature magnet and as the armature magnet has poles which are both attracted to and repelled by the adjacent pole of the stator magnets, both attraction and repulsion forces act upon the armature magnet to produce the relative displacement between the armature and stator magnets.

The continuing motive force producing displacement between the armature and stator magnets results from the relationship of the length of the armature magnet in the direction of its path of movement as related to the dimension of the stator magnets, and the spacing therebetween, in the direction of the path of armature magnet movement. This ratio of magnet and magnet spacings, and with an acceptable air gap spacing between the stator and armature magnets, will produce a resultant force upon the armature magnet which displaces the armature magnet across the stator magnet along its path of movement.

In the practice of the invention movement of the armature magnet relative to the stator magnets results from a combination of attraction and repulsion forces existing between the stator and armature magnets. By concentrating the magnetic fields of the stator and armature magnets the motive force imposed upon the armature magnet is intensified, and in the disclosed embodiments such magnetic field concentration means are disclosed.

The disclosed magnetic field concentrating means comprise a plate of high magnetic field permeability disposed adjacent one side of the stator magnets in substantial engagement therewith. This high permeability material is thus disposed adjacent poles of like polarity of the stator magnets. The magnetic field of the armature magnet may be concentrated and directionally oriented by bowing the armature magnet, and the magnetic field may further be concentrated by shaping the pole ends of the armature magnet to concentrate the magnet field at a relatively limited surface at the armature magnet pole ends.

Preferably, a plurality of armature magnets are used which are staggered with respect to each other in the direction of armature magnet movement. Such an offsetting or staggering of the armature magnets distributes the impulses of force imposed upon the armature magnets and results in a smoother application of forces to the armature magnet producing a smoother and more uniform movement of the armature component.

In the rotary embodiment of the permanent magnet motor of the invention the stator magnets are arranged in a circle, and the armature magnets rotate about the stator magnets. Means are disclosed for producing relative axial displacement between the stator and armature magnets to adjust the axial alignment thereof, and thereby regulate the magnitude of the magnetic forces being imposed upon the armature magnets. In this manner the speed of rotation of the rotary embodiment may be regulated.

BRIEF DESCRIPTION OF THE DRAWINGS

The aforementioned objects and advantages of the invention will be appreciated from the following description and accompanying drawings wherein:

FIG. **1** is a schematic view of electron flow in a superconductor indicating the unpaired electron spins,

FIG. **2** is a cross-sectional view of a superconductor under a critical state illustrating the electron spins,

FIG. **3** is a view of a permanent magnet illustrating the flux movement therethrough,

FIG. **4** is a cross-sectional view illustrating the diameter of the magnet of FIG. **3**,

FIG. **5** is an elevational representation of a linear motor embodiment of the permanent magnet motor of the invention illustrating one position of the armature magnet relative to the stator magnets, and indicating the magnetic forces imposed upon the armature magnet,

FIG. **6** is a view similar to FIG. **5** illustrating displacement of the armature magnet relative to the stator magnets, and the influence of magnetic forces thereon at this location,

FIG. **7** is a further elevational view similar to FIGS. **5** and **6** illustrating further displacement of the armature magnet to the left, and the influence of the magnetic forces thereon,

FIG. **8** is a top plan view of a linear embodiment of the inventive concept illustrating a pair of armature magnets in linked relationship disposed above the stator magnets,

FIG. **9** is a diametrical, elevational, sectional view of a rotary motor embodiment in accord with the invention as taken along section IX—IX of FIG. **10**, and

FIG. **10** is an elevational view of the rotary motor embodiment as taken along section X—X of FIG. **9**.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

In order to better understand the theory of the inventive concept, reference is made to FIGS. **1** through **4**. In FIG. **1** a superconductor **1** is illustrated having a positive particle flow as represented by arrow **2**, the unpaired electrons of the ferrous conducting material **1** spin at right angles to the proton flow in the conductor as represented by the spiral line and arrow **3**. In accord with the theory of the invention the spinning of the ferrous unpaired electrons results from the atomic structure of ferrous materials and this spinning atomic particle is believed to be opposite in charge and located at right angles to the moving electrons. It is assumed to be very small in size capable of penetrating other elements and their compounds unless they have unpaired electrons which capture these particles as they endeavor to pass therethrough.

The lack of electrical resistance of conductors at a critical superconductor state has long been recognized, and superconductors have been utilized to produce very high magnetic flux density electromagnets. FIG. **2** represents a cross section of a critical superconductor and the electron spins are indicated by the arrows **3**.

A permanent magnet may be considered a superconductor as the electron flow therein does not cease, and is without resistance, and unpaired electric spinning particles exist which, in the practice of the invention, are utilized to produce motor force. FIG. **3** illustrates a horseshoe shaped permanent magnet at 4 and the magnetic flux therethrough is indicated by arrows 5, the magnetic flow being from the south pole to the north pole and through the magnetic material. The accumulated electron spins occurring about the diameter of the magnet 5 are represented at 6 in FIG. 4, and the spinning electron particles spin at right angles in the iron as the flux travels through the magnet material.

By utilizing the electron spinning theory of ferrous material electrons, it is possible with the proper ferromagnetic materials, geometry and magnetic concentration to utilize the spinning electrons to produce a motive force in a continuous direction, thereby resulting in a motor capable of doing work.

It is appreciated that the embodiments of motors utilizing the concepts of the invention may take many forms, and in the illustrated forms the basic relationships of components are illustrated in order to disclose the inventive concepts and principles.

The relationships of the plurality of magnets defining the stator 10 are best appreciated from FIGS. 5 through 8. The stator magnets 12 are preferably of a rectangular configuration, FIG. 8, and so magnetized that the poles exist at the large surfaces of the magnets, as will be appreciated from the N (North) and S (South) designations. The stator magnets include side edges 14 and 16 and end edges 18. The stator magnets are mounted upon a supporting plate 20, which is preferably of a metal material having a high permeability to magnetic fields and magnetic flux such as that available under the trademark Netic CoNetic sold by the Perfection Mica Company of Chicago, Illinois. Thus, the plate 20 will be disposed toward the south pole of the stator magnets 12, and preferably in direct engagement therewith, although a bonding material may be interposed between the magnets and the plate in order to accurately locate and fix the magnets on the plate, and position the stator magnets with respect to each other.

Preferably, the spacing between the stator magnets 12 slightly differs between adjacent stator magnets as such a variation in spacing varies the forces being imposed upon the armature magnet at its ends, at any given time, and thus results in a smoother movement of the armature magnet relative to the stator magnets. Thus, the stator magnets so positioned relative to each other define a track 22 having a longitudinal direction left to right as viewed in FIGS. 5 through 8.

In FIGS. 5 through 7 only a single armature magnet 24 is disclosed, while in FIG. 8 a pair of armature magnets are shown. For purposes of understanding the concepts of the invention the description herein will be limited to the use of single armature magnet as shown in FIGS. 5 through 7.

The armature magnet is of an elongated configuration wherein the length extends from left to right, FIG. 5, and may be of a rectangular transverse cross-sectional shape. For magnetic field concentrating and orientation purposes the magnet 24 is formed in an arcuate bowed configuration as defined by concave surfaces 26 and convex surfaces 28, and the poles are defined at the ends of the magnet as will be appreciated from FIG. 5. For further magnetic field concentrating purposes the ends of the armature magnet are shaped by beveled surfaces 30 to minimize the cross-sectional area at the magnet ends at 32, and the magnetic flux existing between the poles of the armature magnet are as indicated by the light dotted lines. In like manner the magnetic fields of the stator magnets 12 are indicated by the light dotted lines.

The armature magnet 24 is maintained in a spaced relationship above the stator track 22. This spacing may be accomplished by mounting the armature magnet upon a slide, guide or track located above the stator magnets, or the armature magnet could be mounted upon a wheeled vehicle carriage or slide supported upon a nonmagnetic surface or guideway disposed between the stator magnets and the armature magnet. To clarify the illustration, the means for supporting the armature magnet 24 is not illustrated and such means form no part of invention, and it is to be understood that the means supporting the armature magnet prevents the armature magnet from moving away from the stator magnets, or moving closer thereto, but permits free movement of the armature magnet to the left or right in a direction parallel to the track 22 defined by the stator magnets.

It will be noted that the length of the armature magnet 24 is slightly greater than the width of two of the stator magnets 12 and the spacing therebetween. The magnetic forces acting upon the armature magnet when in the position of FIG. 5 will be repulsion forces 34 due to the proximity of like polarity forces and attraction forces at 36 because of the opposite polarity of the south pole of the armature magnet, and the north pole field of the sector magnets. The relative strength of this force is represented by the thickness of the force line.

The resultant of the force vectors imposed upon the armature magnet as shown in FIG. 5 produce a primary force vector 38 toward the left, FIG. 5, displacing the armature magnet 24 toward the left. In FIG. 6 the magnetic forces acting upon the armature magnet are represented by the same reference numerals as in FIG. 5. While the forces 34 constitute repulsion forces tending to move the north pole of the armature magnet away from the stator magnets, the attraction forces imposed upon the south pole of the armature magnet and some of the repulsion forces, tend to move the armature magnet further to the left, and as the resultant force 38 continues to be toward the left the armature magnet continues to be forced to the left.

FIG. 7 represents further displacement of the armature magnet 24 to the left with respect to the position of FIG. 6, and the magnetic forces acting thereon are represented by the same reference numerals as in FIGS. 5 and 6, and the stator magnet will continue to move to the left, and such movement continues the length of the track 22 defined by the stator magnets 12.

Upon the armature magnet being reversed such that the north pole is positioned at the right as viewed in FIG. 5, and the south pole is positioned at the left, the direction of movement of the armature magnet relative to the stator magnets is toward the right, and the theory of movement is identical to that described above.

In FIG. 8 a plurality of armature magnets 40 and 42 are illustrated which are connected by links 44. The armature magnets are of a shape and configuration identical to that of the embodiment of FIG. 5, but the magnets are staggered with respect to each other in the direction of magnet movement, i.e., the direction of the track 22 defined by the stator magnets 12. By so staggering a plurality of armature magnets a smoother movement of the interconnected armature magnets is produced as compared when using a single armature magnet as there is variation in the forces acting upon each armature magnet as it moves above the track 22 due to the change in magnetic forces imposed thereon. The use of several armature magnets tends to "smooth out" the application of forces imposed upon linked armature magnets, resulting in a smoother movement of the armature magnet assembly. Of course, any number of armature magnets may be interconnected, limited only by the width of the stator magnet track 22.

In FIGS. 9 and 10 a rotary embodiment embracing the inventive concepts is illustrated. In this embodiment the principle of operation is identical to that described above, but the orientation of the stator and armature magnets is such that rotation of the armature magnets is produced about an axis, rather than a linear movement being achieved.

In FIGS. 9 and 10 a base is represented at **46** serving as a support for a stator member **48**. The stator member **48** is made of a nonmagnetic material, such as synthetic plastic, aluminum, or the like. The stator includes a cylindrical surface **50** having an axis, and a threaded bore **52** is concentrically defined in the stator. The stator includes an annular groove **54** receiving an annular sleeve **56** of high magnetic field permeability material such as Netic Co-Netic and a plurality of stator magnets **58** are affixed upon the sleeve **56** in spaced circumferential relationship as will be apparent in FIG. **10**. Preferably, the stator magnets **58** are formed with converging radial sides as to be of a wedge configuration having a curved inner surface engaging sleeve **56**, and a convex outer pole surface **60**.

The armature **62**, in the illustrated embodiment, is of a dished configuration having a radial web portion, and an axially extending portion **64**. The armature **62** is formed of a nonmagnetic material, and an annular belt receiving groove **66** is defined therein for receiving a belt for transmitting power from the armature to a generator, or other power consuming device. Three armature magnets **68** are mounted on the armature portion **64**, and such magnets are of a configuration similar to the armature magnet configuration of FIGS. **5** through **7**. The magnets **68** are staggered with respect to each other in a circumferential direction wherein the magnets are not disposed as 120° circumferential relationships to each other. Rather, a slight angular staggering of the armature magnets is desirable to "smooth out" the magnetic forces being imposed upon the armature as a result of the magnetic forces being simultaneously imposed upon each of the armature magnets. The staggering of the armature magnets **68** in a circumferential direction produces the same effect as the staggering of the armature magnets **40** and **42** as shown in FIG. **8**.

The armature **62** is mounted upon a threaded shaft **70** by antifriction bearings **72**, and the shaft **70** is threaded into the stator threaded bore **52**, and may be rotated by the knob **74**. In this manner rotation of the knob **74**, and shaft **70**, axially displaces the armature **62** with respect to the stator magnets **58**, and such axial displacement will very the magnitude of the magnetic forces imposed upon the armature magnets **68** by the stator magnets thereby controlling the speed of rotation of the armature.

As will be noted from FIGS. 4–7 and 9 and 10, an air gap exists between the armature magnet or magnets and the stator magnets and the dimension of this spacing, effects the magnitude of the forces imposed upon the armature magnet or magnets. If the distance between the armature magents, and the stator magnets is reduced the forces imposed upon the armature magnets by the stator magnets are increased, and the resultant force vector tending to displace the armature magnets in their path of movement increases. However, the decreasing of the spacing between the armature and stator magnets creates a "pulsation" in the movement of the armature magnets which is objectionable, but can be, to some extent, minimized by using a plurality of armature magnets. The increasing of the distance between the armature and stator magnets reduces the pulsation tendency of the armature magnet, but also reduces the magnitude of the magnetic forces imposed upon the armature magnets. Thus, the most effective spacing between the armature magnets. Thus, the most effective spacing between the armature and stator magnets is that spacing which produces the maximum force vector in the direction of armature magnet movement, with a minimum creation of objectionable pulsation.

In the disclosed embodiments the high permeability plate **20** and sleeve **56** are disclosed for concentrating the magnetic field of the stator magnets, and the armature magnets are bowed and have shaped ends for magnetic field concentration purposes. While such magnetic field concentration means result in higher forces imposed upon the armature magnets for given magnet intensities, it is not intended that the inventive concepts be limited to the use of such magnetic field concentrating means.

As will be appreciated from the above description of the invention, the movement of the armature magnet or magnets resultsfrom the described relationship of components. The length of the armature magnets as related to the width of the stator magnets and spacing therebetween, the dimension of the air gap and the configuration of the magnetic field, combined, produce the desired result and motion. The inventive concepts may be practiced even though these relationships may be varied within limits not yet defined and the invention is intended to encompass all dimensional relationships which achieve the desired goal of armature movement. By way of example, with respect to FIGS. 4–7, the following dimensions were used in an operating prototype:

The length of armature magnet **24** is 3⅛", the stator magnets **12** are 1" wide, ¼" thick and 4" long and grain oriented. The air gap between the poles of the armature magnet and the stator magnets is approximately 1½" and the spacing between the stator magnets is approximately ½" inch.

In effect, the stator magnets define a magnetic field track of a single polarity transversely interrupted at spaced locations by the magnetic fields produced by the lines of force existing between the poles of the stator magnets and the unidirectional force exerted on the armature magnet is a result of the repulsion and attraction forces existing as the armature magnet traverses this magnetic field track.

It is to be understood that the inventive concept embraces an arrangement wherein the armature magnet component is stationary and the stator assembly is supported for movement and constitutes the moving component, and other variations of the inventive concept will be apparent to those skilled in the art without departing from the scope thereof. As used herein the term "track" is intended to include both linear and circular arrangements of the static magnets, and the "direction" or "length" of the track is that direction parallel or concentric to the intended direction of armature magnet movement.

I claim:

1. A permanent magnet motor comprising, in combination, a stator track defining a track direction and having first and second sides and composed of a plurality of track permanent magnets each having first and second poles of opposite polarity, said magnets being disposed in side-by-side relationship having a spacing between adjacent magnets and like poles defining said track sides, an elongated armature permanent magnet located on one of said track sides for relative movement thereto and in spaced relationship to said track side wherein an air gap exists between said armature magnet and said track magnets, said armature magnet having first and second poles of opposite polarity located at the opposite ends of said armature magnet defining the length thereof, the length of said armature magnet being disposed in a direction in general alignment with the direction of said track, the spacing of said armature magnet poles from said track associated side and the length of said armature magnet as related to the width and spacing of said track magnets in the direction of said track being such as to impose a continuous force on said armature magnet in said general direction of said track.

2. In a permanent magnet motor as in claim 1 wherein the spacing between said poles of said armature magnet and the adjacent stator track side are substantially equal.

3. In a permanent magnet motor as in claim 1 wherein the spacing between adjacent track magnets varies.

4. In a permanent magnet motor as in claim 1 wherein a plurality of armature magnets are disposed on a common side of said stator track, said armature magnets being mechanically interconnected.

5. In a permanent magnet motor as in claim 4 wherein said armature magnets are staggered with respect to each other in the direction of said track.

6. In a permanent magnet motor as in claim 1 wherein magnetic field concentrating means are associated with said track magnets.

7. In a permanent magnet motor as in claim 6 wherein said field concentrating means comprises a sheet of magnetic material of high magnetic field permeability engaging side and pole of said track magnets opposite to that side and pole disposed toward said armature magnet.

8. In a permanent magnet motor as in claim 1 wherein said armature magnet is of an arcuate configuration in its longitudinal direction bowed toward said track, said said armature magnet having ends shaped to concentrate the magnetic field at said ends.

9. In a permanent magnet motor as in claim 1 wherein said stator track is of a generally linear configuration, and means supporting said armature magnet relative to said track for generally linear movement of said armature magnet.

10. In a permanent magnet motor as in claim 1 wherein said stator track magnets define a circle having an axis, an armature rotatably mounted with respect to said track and concentric and coaxial thereto, said armature magnet being mounted upon said armature.

11. In a permanent magnet motor as in claim 10, means axially adjusting said armature relative to said track whereby the axial relationship of said armature magnet and said stator magnets may be varied to adjust the rate of rotation of said armature.

12. In a permanent magnet motor as in claim 10 wherein a plurality of armature magnets are mounted on said armature.

13. In a permanent magnet motor as in claim 12 wherein said armature magnets are circumferentially nonuniformily spaced on said armature.

14. A permanent magnet motor comprising, in combination, a stator comprising a plurality of circumferentially spaced stator permanent magnets having poles of opposite polarity, said magnets being arranged to substantially define a circle having an axis, the poles of said magnets facing in a radial direction with respect to said axis and poles of the same polarity facing away from said axis and the poles of opposite polarity facing toward said axis, an armature mounted for rotation about said axis and diposed adjacent said stator, at least one armature permanent magnet having poles of opposite polarity mounted on said armature and in radial spaced relationship to said circle of stator magnets, said armature magnet poles extending in the circumferential direction of armature rotation, the spacing of said armature magnet poles from said stator magnets and the circumferential length of said armature magnet and the spacing of said stator magnets being such as to impose a continuing circumferential force on said armature magnet to rotate said armature.

15. In a permanent magnet motor as in claim 14 wherein a plurality of armature magnets are mounted upon said armature.

16. In a permanent magnet motor as in claim 14 wherein said armature magnets are asymmetrically circumferentially spaced on said armature.

17. In a permanent magnet motor as in claim 14 wherein the poles of said armature magnet are shaped to concentrate the magnetic field thereof.

18. In a permanent magnet motor as in claim 14, magnetic field concentrating means associated with said stator magnets concentrating the magnetic fields thereof at the spacings between adjacent stator magnets.

19. In a permanent magnet motor as in claim 18 wherein said magnet field concentrating means comprises an annular ring of high magnetic field permeability material concentric with said axis and in substantial engagement with poles of like polarity of said stator magnets.

20. In a permanent magnet motor as in claim 14 wherein said armature magnet is of an arcuate bowed configuration in the direction of said poles thereof defining a concave side and a convex side, said concave side being disposed toward said axis, and said poles of said armature magnet being shaped to concentrate the magnetic field between said poles thereof.

21. In a permanent magnet motor as in claim 14, means for axially displacing said stator and armature relative to each other to adjust the axial alignment of said stator and armature magnets.

22. The method of producing a unidirectional motive force by permanent magnets using a plurality of spaced stator permanent magnets having opposite polarity poles defining a track having a predetermined direction, and an armature magnet having a length defined by poles of opposite polarity movably mounted for movement relative to the track in the direction thereof, and of a predetermined length determined by the width and dimensions of said stator magnets comprising forming a magnetic field track by said stator magnets having a magnetic field of common polarity interrupted at spaced locations in a direction transverse to the direction of said magnetic field track by magnetic fields created by magnetic lines of force existing between the

poles of the stator magnets and positioning the armature magnet in spaced relation to said magnetic field track longitudinally related to the direction of the magnetic field track such a distance that the repulsion and attraction forces imposed on the armature magnet by said magnetic field track imposes a continuing unidirectional force on the armature magnet in the direction of the magnetic field track.

23. The method of producing a unidirectional motive force as in claim **22** including concentrating the magnetic fields created by magnetic lines of force between the poles of the stator magnets.

24. The method of producing a unidirectional motive force as in claim **22** including concentrating the magnetic field existing between the poles of the armature magnet.

25. The method of producing a unidirectional motive force as in claim **22** including concentrating the magnetic fields created by magnetic lines of force between the poles of the stator magnets and concentrating the magnetic field existing between the poles of the armature magnet.

26. The method of producing a motive force by permanent magnets wherein the unpaired electron spinning particles existing within a permanent magnet are utilized for producing a motive force comprising forming a stator magnetic field track by means of at least one permanent magnet, producing an armature magnetic field by means of a permament magnet and shaping and locating said magnetic fields in such a manner as to produce relative continuous unidirectional motion between said stator and armature field producing magnets.

27. The method of producing a motive force by permanent magnets as in claim **26** wherein said stator magnetic field is substantially of a single polarity.

28. The method of producing a motive force by permanent magnets as in claim **26** including concentrating the magnetic field of said stator field track and armature magnetic field.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 4,151,431

DATED : April 24, 1979

INVENTOR(S) : Howard R. Johnson

It is certified that error appears in the above—identified patent and that said Letters Patent is hereby corrected as shown below:

The title page should be deleted to insert the attached title page therefor.

Column 12, claims 26, 27 and 28 should be deleted inclusively.

Signed and Sealed this

Twenty-sixth Day of *August 1980*

[SEAL]

Attest:

SIDNEY A. DIAMOND

Attesting Officer *Commissioner of Patents and Trademarks*

United States Patent [19]

Johnson

[11] **4,151,431**

[45] **Apr. 24, 1979**

[54] PERMANENT MAGNET MOTOR

[76] Inventor: **Howard R. Johnson,** 3300 Mt. Hope Rd., Grass Lake, Mich. 49240

[21] Appl. No.: **422,306**

[22] Filed: **Dec. 6, 1973**

[51] **Int. Cl.**^{2} **H02K 41/00;** H02N 11/00

[52] **U.S. Cl.** **310/12;** 310/152

[58] **Field of Search** .... 24/DIG. 9; 415/DIG. 2; 46/236; 273/118 A, 119 A, 120 A, 121 A, 122 A, 123 A, 124, 125 A, 126 A, 130 A, 131 A, 131 AD, 134 A, 135 A, 136 B, 137 AE, 138 A

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,074,153	2/1978	Baker et al.	310/12

Primary Examiner—Donovan F. Duggan
Attorney, Agent, or Firm—Beaman & Beaman

[57] ABSTRACT

The invention is directed to the method of utilizing the unpaired electron spins in ferro magnetic and other materials as a source of magnetic fields for producing power without any electron flow as occurs in normal conductors, and to permanent magnet motors for utilizing this method to produce a power source. In the practice of the invention the unpaired electron spins occurring within permanent magnets are utilized to produce a motive power source solely through the superconducting characteristics of a permanent magnet and the magnetic flux created by the magnets are controlled and concentrated to orient the magnetic forces generated in such a manner to do useful continuous work, such as the displacement of a rotor with respect to a stator. The timing and orientation of magnetic forces at the rotor and stator components produced by permanent magnets to produce a motor is accomplished with the proper geometrical relationship of these components.

25 Claims, 10 Drawing Figures

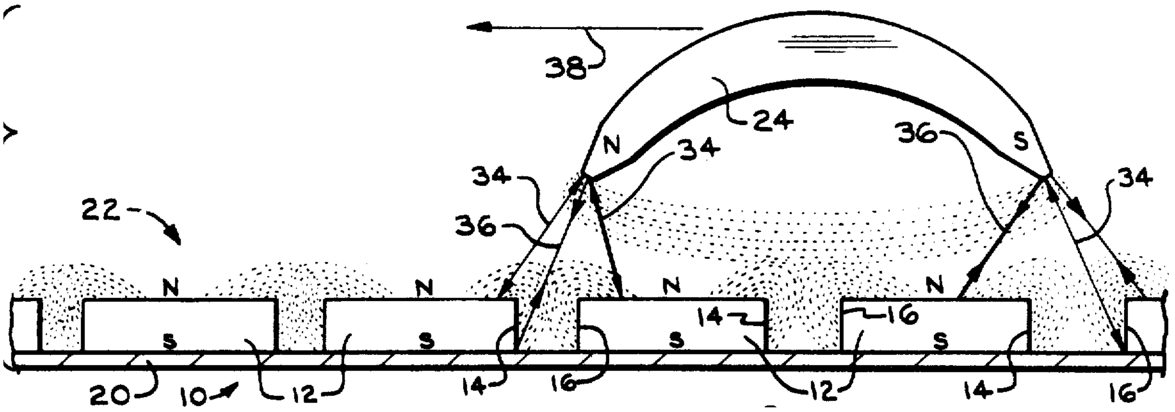

rexresearch.com

Howard JOHNSON

Magnet Motor

Permanent Magnet Motor
Looking for Permanent Magnet Motor? See Permanent Magnet Motor.
GrovesIndustrial.com

How to make Electricity
$197 DIY kit has energy co's execs screaming for a ban on its sale.
www.Power-4-Homes.com

Linear Synchronous Motors
Design and manufacture of LSMs Mining, defence, renewable energy
www.texchange.co.uk

Buckyballs Magnetic Toy
Get the Gadget on Everyone's List. Best prices on this year's hot toy!
www.getbuckyballs.com

Alternators/Generators
Permanent Magnet Alternators / Generators for aircraft, ATVs, etc…
www.compositemotors.com

Permanent Magnet Motor
We have great resources for Permanent Magnet Motor
maziro.com

Advanced Motion Control
Intelligent Servo Drives For AC, DC and Step Motors
www.technosoftmotion.com

Flexible Magnet Material
Flexible Magnet and rubber magnet Best Magnets Source for all kinds r
www.magnets-flexible.com

Magnets
Hunting For Magnets? Browse Our Magnets Directory.
kellyscornerstore.com

Science & Mechanics: "Amazing Magnet-Powered Motor" (Spring 1980)
H. Johnson & W. Harrison, Jr:" The Permanent Magnet Motor"
USP # 4,151,431: Permanent Magnet Motor
USP # 4,877,983: Magnetic Force Generating Method & Apparatus
USP # 5,402,021: Magnetic Propulsion System
Tom Bearden/Karl Bergmann Comments
Blueprints:
http://www.greaterthings.com/news/freeenergy/directory/howard_johnson_motor/blueprints/index.html

Lab Memorandum, CTEC, Inc (1996) (MS Word.doc, 1 MB)

Howard R. Johnson

Science & Mechanics **(Spring 1980) Cover Illustration**

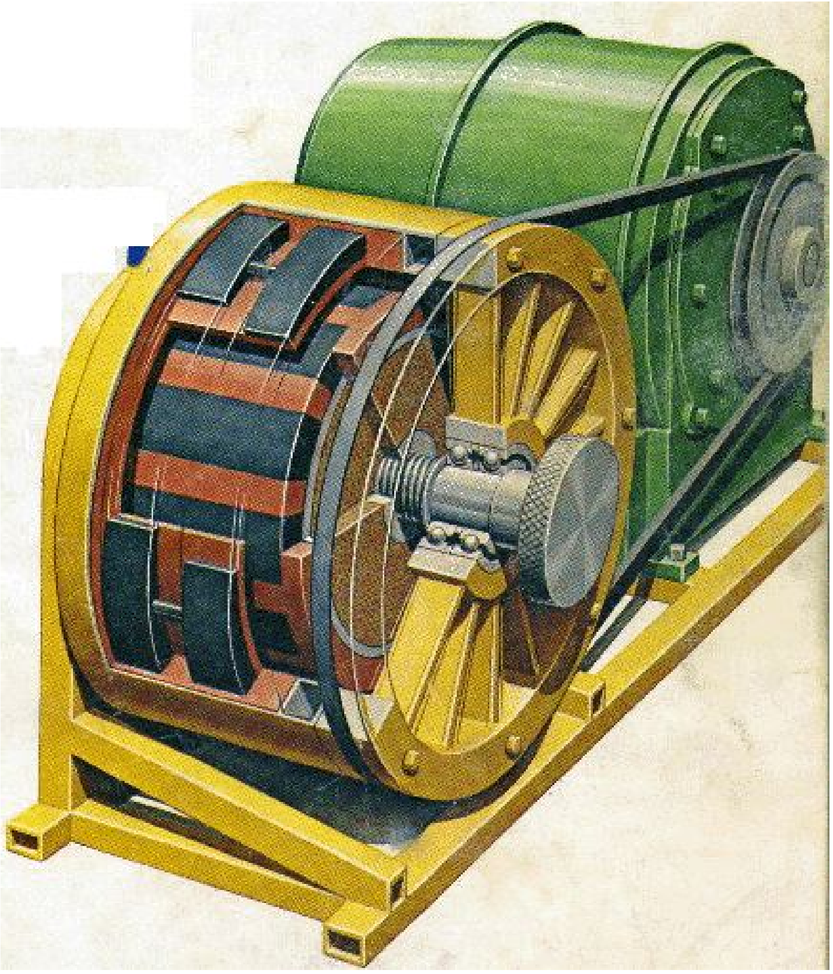

Figure 1

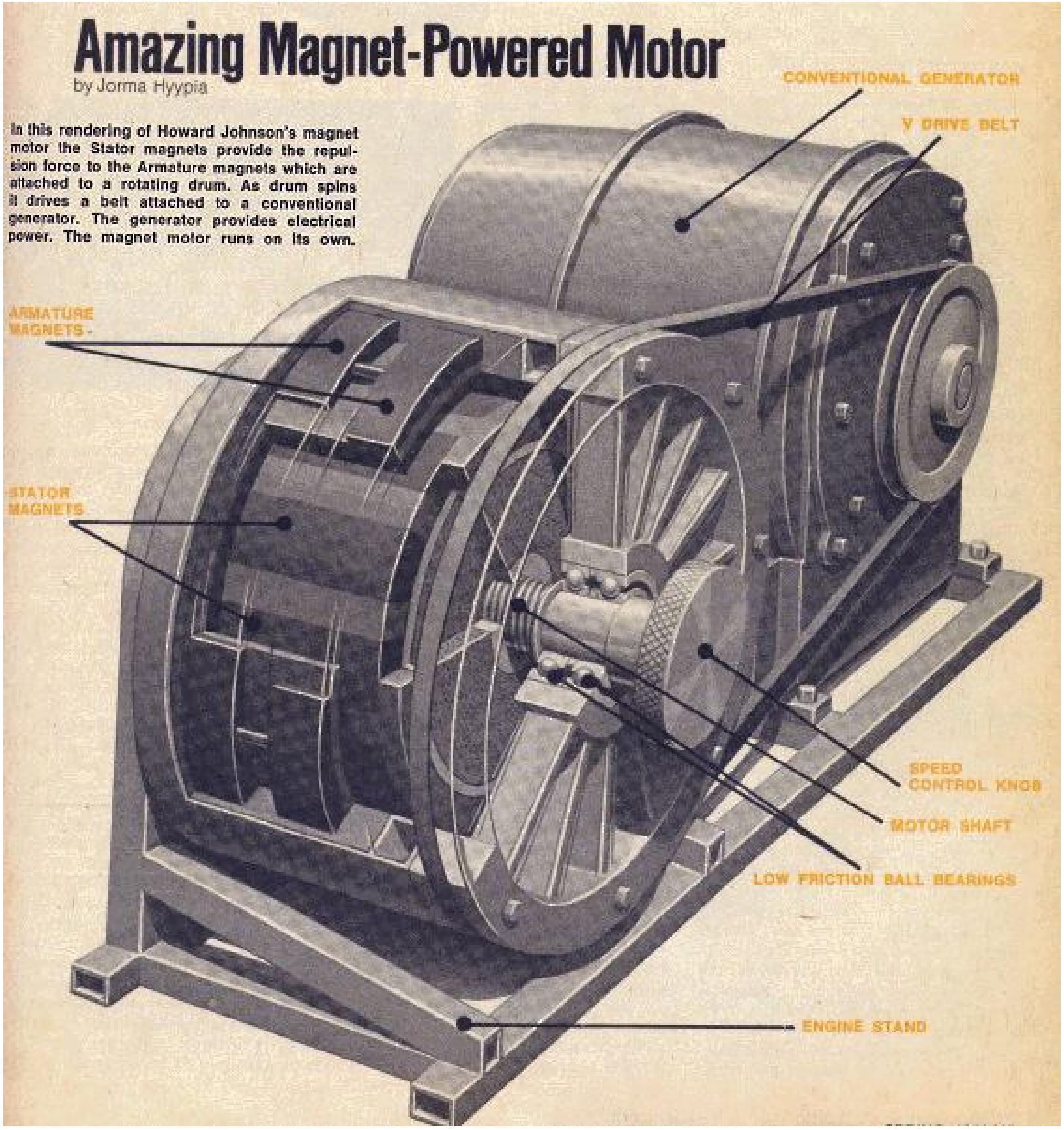

Amazing Magnet-Powered Motor

by Jorma Hyypia

In this rendering of Howard Johnson's magnet motor the Stator magnets provide the repulsion force to the Armature magnets which are attached to a rotating drum. As drum spins it drives a belt attached to a conventional generator. The generator provides electrical power. The magnet motor runs on its own.

Science & Mechanics (Spring 1980)

"Amazing Magnet-Powered Motor"

by
Jorma Hyypia

"We don't grant patents on perpetual motion machines," said the examiners at the U.S. Patent Office. "It won't work because it violates the law of Conservation of Energy," said one physicist after another. But because, inventor Howard Johnson is not the sort of man to be intimidated by such seemingly authoritative pronouncements, he now owns U.S. Patent No. 4,151,431 which describes how it is possible to generate motive power, as in a motor, using only the energy contained in the atoms of permanent magnets. That's right. Johnson has discovered how to build motors that run without an input of electricity or any other kind of external energy!

The monumental nature of the invention is obvious, especially in a world facing an alarming, escalating energy shortage. Yet inventor Johnson is not rushing to peddle his creation as the end-all solution to world- wide energy problems. He has more important work to do. First, there's the need to refine his laboratory prototypes into workable practical devices -in particular a 5,000-watt electric power generator already in the building. His second and perhaps more difficult major challenge: persuade a host of skeptics that his ideas are indeed practical.

Johnson, who has been coping with disbelievers for decades, can be very persuasive in a face-to-face encounter because he can not do more than merely theorize; he can demonstrate working models that unquestionably create motion using only permanent magnets. When this writer was urged by the editor of Science & Mechanics to make a thousand mile pilgrimage to Blacksburg, Virginia, to meet with the inventor, he went there as an "open-minded skeptic" and as a former research Scientist determined not to be fooled. Within two days, this former skeptic had become a believer. Here's why.

Doing the Unthinkable

Howard Johnson refuses to view the "laws" of science as somehow sacred, so doing the unthinkable and succeeding is second nature to him. If a particular law gets in the way, he sees no harm in going around it for a while to see if there's something on the other side. Johnson explains the persistent opposition he experiences from the established scientific community this way: "Physics is a measurement science and physicists are especially determined to protect the &lsquo;Law&rsquo; of Conservation of Energy. Thus the physicists become game wardens who tell us what laws' we can't violate. In this case they don't even know what the game is. But they are so scared that I and my associates are going to violate some of these laws, that they have to get to the pass to head us off!"

The critics say Johnson offers a "free lunch" solution to energy problems, and that there can be no such thing. Johnson demurs, reminding repeatedly that he has never suggested that his invention provides something for nothing. He also points gut that no one talks about a "free lunch" when discussing extraction of enormous amounts of atomic power by means of nuclear reactors and atom bombs. In his mind, it's much the same thing.

Johnson is the first to admit he doesn't actually know where the power be has tapped derives. But he postulates that the energy may be associated with spinning electrons, perhaps in the form of a "presently unnamed atomic particle." How do other physicists react to Johnson's suggestion that there may be an atomic particle so far overlooked by nuclear physicists? Says Johnson: "I guess it&rsquo;s fair to say that most of them are revolted." On the other hand, a few converted scientists, including some who are associated with large and prestigious research laboratories, are intrigued enough to suggest that there should be a hunt for the answer, be it a "particle" or some other as yet unsuspected characteristic of atomic structure.

This article is prefaced with the foregoing brief summary of the ongoing controversy so that, in fairness to the inventor, we might all view his claims with open minds, even if it means temporary setting aside of cherished scientific concepts until more complete explanations are forthcoming. The main question to be answered here and now is this: Does Johnson permanent magnet motor work?

Before providing the answer, we need to face up to another question that undoubtedly nags in the minds of many readers: Is Johnson a bona fide researcher, or merely a "garage mechanic" mad inventor? As the following brief summary suggests, the inventor's credentials appear to be impeccable. Following seven years of college and university training, Johnson worked on atomic energy projects at Oak Ridge, did magnetics research for Burroughs company, and served as scientific consultant to Lukens Steel. He has participated in the development of medical electrical products, including injection devices. For the military he invented a ceramic muffler that makes a portable motor generator silent at 50 feet; this has been in production for the past 18 years. His contributions to the motor industry include: a hysteresis brake; non-locking brake materials for anti- skid application, new methods of curing brake linings; and a method of dissolving asbestos fibers. He has also worked on silencers for small motors, a super charger, and has perfected a 92-pole no-brush generator to go in the wheel of Lincoln automobiles as a skid control; that last item reduced the cost to one-eighth of the cost of an earlier design by utilizing metal-filled plastics for the armature and field. In all, Johnson is connected with more than 30 patents in the fields of chemistry and physics.

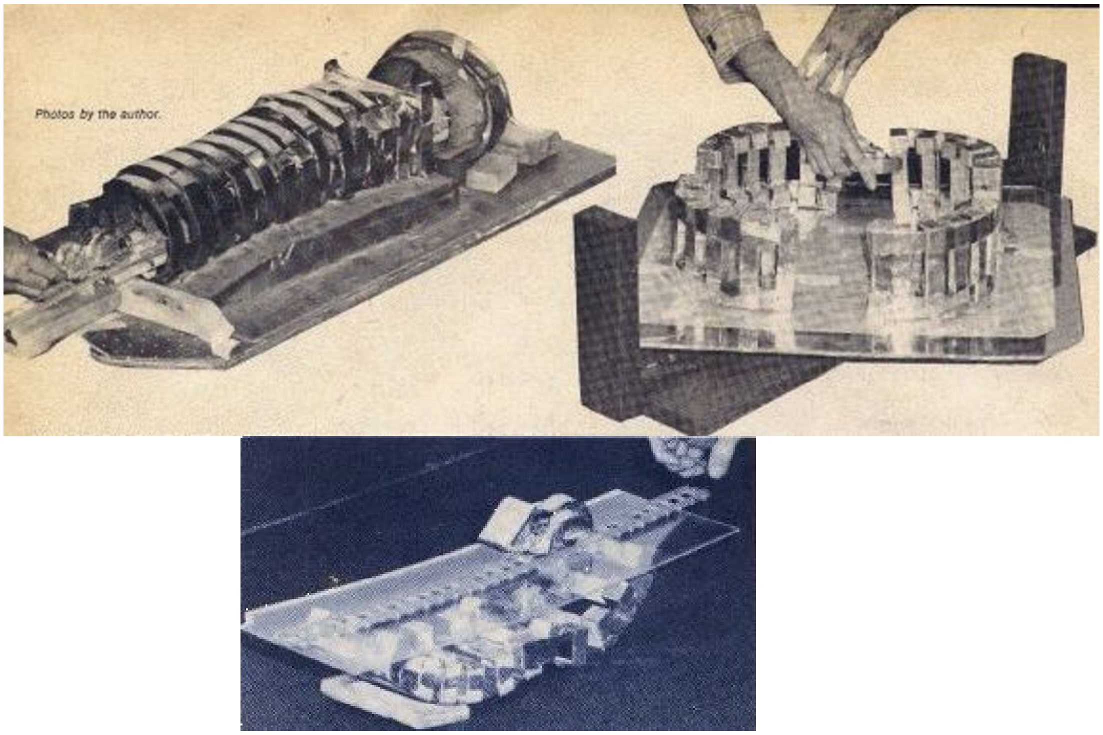
Photos by the author.

Figures 2, 3 & 4: Magnet Motor Models ~ Pictured here are three of the inventor's early models. Top left is a linear motor which propels a magnetic vehicle at high speeds through a series of rings. Top right is rotary motor upon which the prototype will be built. The 8-ounce manget, hand held to the large ring weighing 40 pounds, provides enough force to spin the entire assembly. In the third assembly above, the vehicle is propelled, in either direction, by the force of the large magnets arranged below tracks.

Sticky Tape Scientist

Despite his impressive credentials, this amiable and unpretentious inventor likes to characterize himself as a "Sticky tape" scientist. He sees no virtue in wasting time building fancy, elaborate equipment when more simple assemblies serve as well to test new ideas. The prototype devices shown in the photographs in this article were assembled with sticky tape and aluminum foil, the later material being used mainly to keep individual, permanent magnets packaged together so that they do not fly apart.

Perhaps the best way to describe what these three gadgets do is by reciting this writer's personal experiences during the interview demonstration. That way I will not merely be telling what the inventor says they do, but I will reveal what happened when I tried the experiments myself. When we start talking about how and why the things work as they do, well have to rely on the inventor&rsquo;s explanations.

The first item consists of more than a dozen foil-wrapped magnets assembled to form a broad arc. Each magnet is extended upward slightly at each end to form a low U-shape, the better to concentrate magnetic fields where they are needed. The overall curvature of the mass of magnets apparently has no particular significance except to show that the distance between these stator magnets and the moving vehicle is not critical. A transparent plastic sheet atop this magnet assembly supports a length of plastic model railroad track. The vehicle, basically a model railroad flatcar, supports a foil-wrapped pair of curved magnets, plus some sort of weight, in some cases merely a rock. The weight is needed to keep the vehicle down on the track, against the powerful magnetic forces that would otherwise push it askew. That 'is all there is to the construction of this representation of a "linear motor."

I was prepared to develop eye strain in an effort to detect some sort of motion in the vehicle. I need not have been concerned. The moment the inventor let go of the vehicle be carefully placed at one end of the track, it accelerated and literally zipped from one end to the other and flew onto the floor! Wow!

I tried the experiment myself, and could feel the powerful magnetic forces at work as I placed the vehicle on the track. I gently eased the vehicle to the critical starting point, taking great care not to exert any kind of forward push, even inadvertently. I let go, Zip! It was on the floor again, at the other end of the track. Knowing that I would be asked if the track might have had a slant, I reversed the vehicle and started it from the opposite end of the track. It worked just as effectively in the reverse direction. In fact, the vehicle can even navigate a respectable upgrade. In light of these tests, and considering the remarkable speed of the vehicle, you can discount any notion that this was a simple "coasting" effect.

Incidentally, the photograph shows the vehicle about half ways along the track. It was "frozen" there by the electronic flash used to make the picture; there is no way of "posing" the vehicle in that position short of tying it down.

The second device has the U-shaped magnets standing on end in a rough circular arrangement oddly reminiscent of England's Stonehenge. This assembly is mounted on a transparent plastic sheet supported on a plywood panel pivoted, underneath, on a free turning wheel obtained from a skateboard. As instructed, I eased the 8-ounce focusing magnet into the ring of larger magnets, keeping it at least four inches away from the ring. The 40 pound magnet assembly immediately began to turn and accelerated to a very respectable rotating speed which it maintained for as long as the focusing magnet was held in the magnetic field. When the focusing magnet was reversed, the large assembly turned in the opposite direction.

Since this assembly is clearly a crude sort of motor, there's no doubt that it is indeed possible to construct a motor powered solely by permanent magnets.

The third assembly, which looks like the bones of some prehistoric sea creature, consists of a tunnel constructed of rubber magnet material that can be easily bent to form rings. This was one of the demonstration models Johnson took to the U.S. Patent Office during his appeal proceedings. Normally the patent examiners spend only a few minutes with each patent applicant, but played with Johnson&rsquo;s devices for the better part of an hour. As the inventor was leaving, he overheard one sideline observer remark: "How would you like to follow that act?!"

It took Johnson about six years of legal hassling to finally obtain his patent, and he has been congratulated for his ultimate victory over patent office bureaucracy as well as for his inventiveness. One sign that he left the patent office more than a little shaken by the experience was the inclusion of diagrammatic material in the printed patent that does not belong there. So if you look up the patent, pay no attention to the "ferrite" graph on the first page; it belongs in some other patent!

The tunnel device of course worked very well in the inventor's office during my visit although Johnson observed that the rubber magnets are perhaps a thousand times weaker than the cobalt samarium magnets used the other assemblies. There's just one big problem with the more powerful magnets: they cost too much. According to the inventor, the magnets used to construct the Stonehenge rotating model are collectively worth more than one thousand dollars. But there's no need to depend solely on mass-production economies to bring the cost down to competitive levels. Johnson and U.S. Magnets and Alloy Co. are in the process of developing alternative, relatively low cost magnetic materials that perform very well.

How do they work?

The drawing that shows a curved "arcuate" armature magnet in three successive positions over a line of fixed stator magnets provides at least highly simplified insights into the theory of permanent magnet motive power generation. Johnson says curved magnets with sharp leading and trailing edges are important because they focus and concentrate the magnetic energy much more effectively than do blunt-end magnets. These arcuate magnets are made slightly longer than the lengths of two stator magnets plus the intervening space, in Johnson's setups about 3-1/8 inches long.

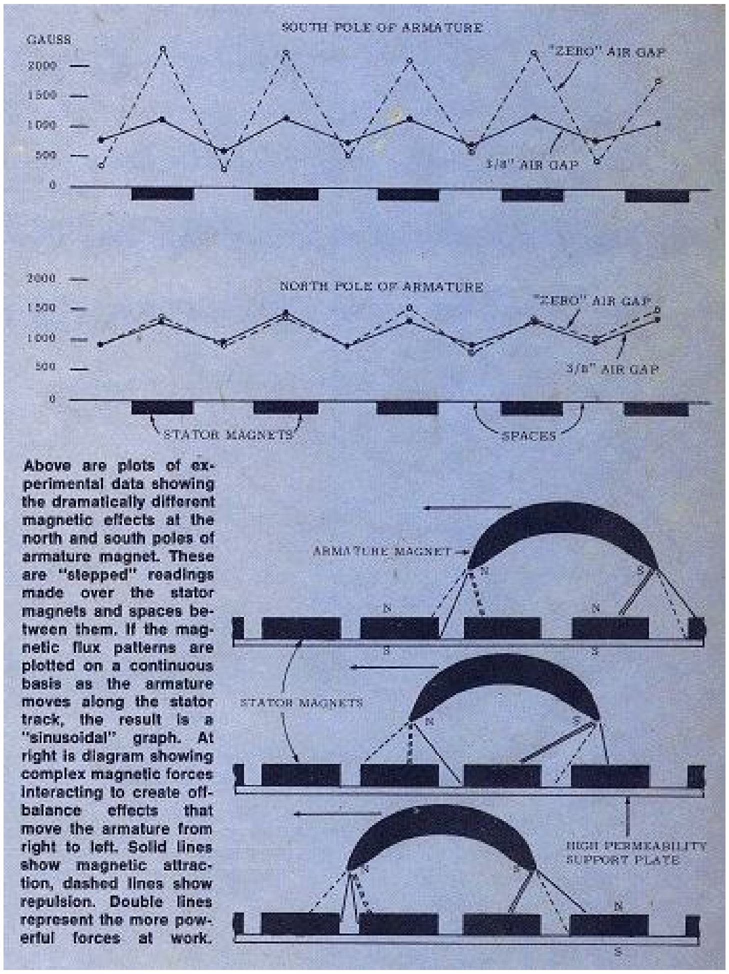

Above are plots of experimental data showing the dramatically different magnetic effects at the north and south poles of armature magnet. These are "stepped" readings made over the stator magnets and spaces between them. If the magnetic flux patterns are plotted on a continuous basis as the armature moves along the stator track, the result is a "sinusoidal" graph. At right is diagram showing complex magnetic forces interacting to create off-balance effects that move the armature from right to left. Solid lines show magnetic attraction, dashed lines show repulsion. Double lines represent the more powerful forces at work.

Note that the stator magnets all have their North faces upward, and that they are resting on a high magnetic permeability support plate that helps concentrate the force fields. The best gap between the end poles of the armature magnet and the stator magnets appears to be about 3/8 inch.

As the armature north pole passes over a magnet, it is repelled by the stator north pole; and there's an attraction when the north pole is passing over a space between the stator magnets. The exact opposite is of course true with respect to the armature South pole. It is attracted when passing over a stator magnet, repelled when passing over a space.

The various magnetic forces that come into play are extremely complex, but the drawing shows some of the fundamental relationships. Solid lines represent attraction forces, dashed lines represent repulsion forces, and double lines in each case indicate the more dominant forces.

As the top drawing indicates, the leading (N) pole of the armature is repelled by the north poles of the two adjacent magnets. But, at the indicated position of the armature magnet, these two repulsive forces .(which obviously work against each other), are not identical; the stronger of the two forces (double dashed line) overpowers the other force and tends to move the armature to the left. This left movement is enhanced by the attraction force between the armature north pole and the stator south pole at the bottom of the space between the stator magnets.

But that's not all! Let's see what is happening simultaneously at the other end (S) of the armature magnet. The length of this magnet (about 3-1/8 inches) is chosen, in relation to the pairs of stator in magnets plus the space between them, so that once again the attraction/repulsion forces work to move the armature magnet to the left. In this case the armature pole (S) is attracted by the north surfaces of the adjacent stator magnets but, because of the critical armature dimensioning, more strongly by the magnet (double solid line) that tends to "pull" the armature to the left. It overpowers the lesser "drag" effect of the stator magnet to the right. Here also there is the added advantage of, in this case, repulsion force between the south pole of the armature and the south pole in the space between the stator magnets.

The importance of correct dimensioning of the armature magnet cannot be over-emphasized. If it is either too long or too short, it could achieve an undesirable equilibrium condition that would stall movement. The objective is to optimize all force conditions to develop the greatest possible off-balance condition, but always' in the same direction as the armature magnet moves along the row of stator magnets. However, if the armature is rotated 180 degrees and started at the opposite end of the track, it would behave in exactly the same manner except that it would, in this example, move from left to right. Also note that once the armature is in motion, it has momentum that helps carry it into the sphere of influence of the next pair of magnets where it gets another push and pull, and additional momentum.

Complex Forces

Some very complex magnetic forces are obviously at play in this deceptively simple magnetic system, and at this time it is impossible to develop a mathematical model of what actually occurs. However, computer analysis of the system, conducted by Professor William Harrison and his associates at Virginia Polytechnic Institute (Blacksburg, VA), provide vital feedback information that greatly helps in the effort to optimize these complex forces to achieve the most efficient possible operating design.

As Professor Harrison points out, in addition to the obvious interaction between the two poles of the armature magnet and the stator magnets, many other interactions are in play. The stator magnets affect each other and the support plate. Magnet distances and their strengths vary despite best efforts of manufacturers to exercise quality controls. In the assembly of the working model, there are inevitable differences between horizontal and vertical air spaces. All these interrelated factors must be optimized, which is why computer analysis in this refinement stage is vital. It's a kind of information feedback system. As changes are made in the physical design, fast dynamic measurements are made to see whether the expected results have actually been achieved. The 'new computer data is then used to develop new changes in the design of the experimental model. And so on, and on.

That very different magnetic conditions exist at the two ends of the armature is shown by the actual experimental data displayed in the table and associated graph. To obtain this information, the researchers first passed the probe of an instrument used to measure magnetic field strengths over the stator magnets and the intervening spaces. We shall call this the "Zero" level although there is a very tiny gap between the probe and the tops of the stator magnets. These measurements in effect indicate what each pole of the armature magnet "sees" below as it passes over. the stator magnets.

Next the probe is moved to a position just beneath one of the armature poles, at the top of the 3/8-inch armature-to-stator air gap. Another set of magnetic flux measurements is made. The procedure is repeated with the probe positioned just beneath the other armature pole.

Now "Instinct" might suggest, and correctly so, that the flux measurements at the top and bottom of the air gap will differ. But if "instinct" also suggests that these differences are pretty much the same at the two armature pole positions, you would be very much in error!

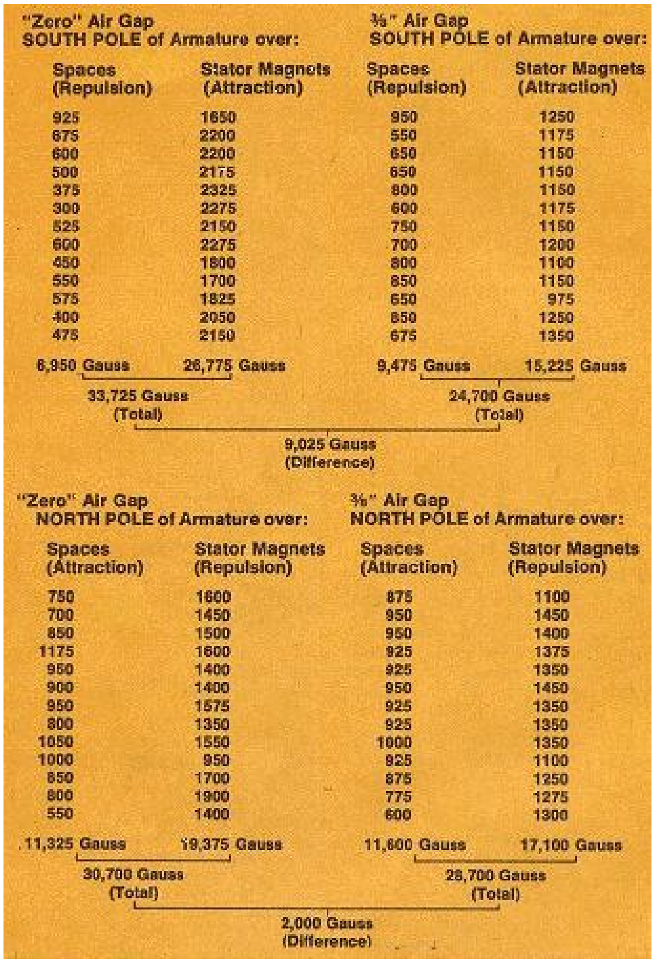

"Zero" Air Gap SOUTH POLE of Armature over: | ⅜" Air Gap SOUTH POLE of Armature over:

Spaces (Repulsion)	Stator Magnets (Attraction)	Spaces (Repulsion)	Stator Magnets (Attraction)
925	1650	950	1250
675	2200	550	1175
600	2200	650	1150
500	2175	650	1150
375	2325	800	1150
300	2275	600	1175
525	2150	750	1150
600	2275	700	1200
450	1800	800	1100
550	1700	850	1150
575	1825	650	975
400	2050	850	1250
475	2150	675	1350
6,950 Gauss	26,775 Gauss	9,475 Gauss	15,225 Gauss

33,725 Gauss (Total) | 24,700 Gauss (Total)

9,025 Gauss (Difference)

"Zero" Air Gap NORTH POLE of Armature over: | ⅜" Air Gap NORTH POLE of Armature over:

Spaces (Attraction)	Stator Magnets (Repulsion)	Spaces (Attraction)	Stator Magnets (Repulsion)
750	1600	875	1100
700	1450	950	1450
850	1500	950	1400
1175	1600	925	1375
950	1400	925	1350
900	1400	950	1450
950	1575	925	1350
800	1350	925	1350
1050	1550	1000	1350
1000	950	925	1100
850	1700	875	1250
800	1900	775	1275
550	1400	600	1300
11,325 Gauss	19,375 Gauss	11,600 Gauss	17,100 Gauss

30,700 Gauss (Total) | 28,700 Gauss (Total)

2,000 Gauss (Difference)

First study the two tables that show actual flux density measurements. Note that in this particular experiment the total magnetic flux amounted to 30,700 Gauss (the unit of magnetic strength) when the probe was held at the "Zero" level under the north pole of the magnet, and a total of 28,700 Gauss when the probe was moved to the top of the 3/8-inch air gap. The difference between these total 'measurements is 2,000 Gauss.

Similar readings made at the air gap between the south pole of the armature and the stator magnets indicates a total flux at "Zero" level of 33,725 Gauss, and 24,700 Gauss at the top of the air gap. This time the difference is a much larger 9,025 Gauss, or four and one half times greater than for the north pole! Clearly, the magnetic force conditions are far from identical at the two ends of the armature magnet.

The middle five pairs of figures from each table hive been plotted in graphic form to make these differences more obvious. In the top "South Pole" graph the dashed line connects, the "Zero" level readings made over the stator magnets and over the intervening air spaces. Points along the solid line indicate comparable readings made with the probe just beneath the armature south pole. It is easy to see that there is an average 43% reduction of the attraction between the armature and stator magnets created by the air gap. Equally true, but perhaps not so obvious, is the fact that there is an average 36% increase of repulsion when the south pole of the armature passes over the spaces between the stator magnets. The percentage increase only seems smaller because it applies to a much smaller "Zero" level value.

The second graph shows that the changes are much less dramatic at the north pole of the armature. In this case there's an average 11.7% decrease of attraction over the spaces, and a 2.4% increase, of repulsion when the armature north pole passes over the stator magnets.

As you study the data, be sure to note that the columns are labeled differently. In the case of the north pole data, the stator magnet areas repulse the armature north pole while the spaces between the stator magnets attract. The conditions are exactly the opposite for the south pole of the armature magnet. When the south pole passes over a magnet, there is strong attraction; when it passes over a space, there is repulsion.

The Ultimate Motor

A motor based on Johnson's findings would be of extremely simple design compared to conventional motors. As shown in the diagrams developed from Johnson&rsquo;s patent literature, the stator/base unit would contain a ring of spaced magnets backed by a high magnetic permeability sleeve. Three arcuate armature magnets would be mounted in the armature which has a belt groove for power transmission. The armature is supported on ball bearings on a shaft that either screws or slides into the stator unit. Speed control and start/stop action would be achieved by the simple means of moving the armature toward and away from the stator section.

There is a noticeable pulsing action in the simple prototype units that may be undesirable in a practical motor. The movement can be smoothed, the inventor believes, by simply using two or more staggered armature magnets as shown in another drawing.

What&rsquo;s Ahead?

For inventor Howard Johnson and his permanent magnet power source there's bound to be plenty of controversy, certainly, but also progress. A 5000 watt electric generator powered by a permanent magnet motor is already on the way, and Johnson has firm licensing agreements with at least four companies at this writing.

Will we see permanent magnet motors in automobiles in the near future? Johnson wants nothing to do with Detroit at this time because, as he puts it: "It&rsquo;s too emotional - we&rsquo;d get smashed into the earth!" The inventor is equally reluctant to make predictions about other applications as well, mainly because he just wants time to perfect his ideas and, hopefully, get the scientific establishment to at least consider his unorthodox ideas with a more open mind.

For example, Johnson argues that the magnetic forces in a permanent magnet represent superconductance that is akin to phenomena normally associated only with extremely cold superconducting systems. He argues that a magnet is a room temperature superconducting system because the electron flow does not cease, and because this electron flow can be made to do work. And for those who pooh- pooh the idea that permanent magnets do work, Johnson has an answer: "You come along with a magnet and pick up a piece of iron, then some physicist says you didn't do any work because you used that magnet. But you moved a mass through a distance.

Right? That's work that requires energy. Or you can hold one magnet in the air indefinitely by positioning it over another magnet with like poles facing. The physicist will argue that because it involves magnetic repulsion, no work is done. Yet if you support the same object with air, they will agree in a minute that work is done!"

There's no doubt in Johnson's mind that he has succeeded in extracting usable energy from the atoms of permanent magnets. But does that imply that the electron spins and associated phenomena that he thinks provide this power will eventually be used up? Johnson makes no pretense of knowing the answer: I didn't start the electron spins, and I don't know any way to stop them - do you? They may eventually stop, but that is not my problem."

Johnson still has many practical problems to solve to perfect his invention. But his greater challenge may be to win general acceptance of his ideas by an obviously nervous scientific community in which many physicists remain compulsive about defending the law of Conservation of Energy without ever wondering whether that "law" really needs defending.

The dilemma facing Johnson is not really his dilemma but rather that of other scientists who have observed his prototypes. The devices obviously do work. But the textbooks say it shouldn&rsquo;t work. And all that Johnson is really saying to the scientific community is this: here is a phenomenon which seems to contradict some of our traditional beliefs. For all our sakes let&rsquo;s not dismiss it outright but take the time to understand the complex forces at work here.

The Permanent Magnet Motor

by
Howard R. Johnson
&
William P. Harrison, Jr.

(Engineering Fundamentals Division, Virginia Polytechnic Institute and State University, Blacksburg, VA)

Presented at the UNITAR (UN Institute for Training & Research) Conference on Long-Term Energy Resource, Montreal, Canada, November 26-December 7, 1979.

I. Introductory remarks (by Mr. Johnson)

Today when energy is so expensive, it is not hard to drum up interst for most any avenue that offers a breath of hope or a way of escape, but this was not necessarily so in 1942. We were somewhat satisfied and convinced that we had the main sources of energy in view. So it took a pure act of faith to try to develop a new un-named source.

It took faith to spend time on it. It took faith to spend money on it. And it took faith to consider facing the opposition later when I made my work known and faced all the *status quo* people.

So, in 1942 using the Bohr model of the atom, and knowing that unpaired electron spins created a permanent magnet dipole, I kept wondering why we couldn't use these fields to drive something. I was sure that the magnetic effect of the spins was similar enough to the field of a current in a wire to do the same thing. I had no knowledge of electron spins stopping and knew no method that I could exert to stop them, so I decided to try to work out a method to use them.

At the same time there were no good hard magnetic materials that I knew of, materials that could be opposed with strong magnetic fields and not be demagnetized enough to damage them. Not only that, they would not give the thrust that I desired.

Having a chemical background, I thought it would be nice to use the best magnetic materials I could find in combination with an interstitial material that was highly diamagnetic to force the electron spin to stay in place.

The U.S. Navy later made such a compound using bismuth and good magnetic materials, but the internal coercive forces were so great that this strong magnet would fall apart if not encased in glass. It was also expensive.

So I kept checking magnetic materials while I worked on designs that I thought should be implemented. It was a quiet, sometimes lonely job over the years, for I didn't share my plans with my associates. My self-imposed security would not permit it, and I knew of few people who would be interested anyway.

In the fifties, as ceramic magnets became better and harder, and long-field metal magnets appeared on the scene, I began to freeze some designs and to have magnets custom made to fit them.

It was about this time that I mentioned the fact that just as I believed electron spins made permanent magnets, I also believed that they were responsible for the 60° angles in the structure of snowflakes giving the six-spoked wheel, the six-sided spokes, etc. The dean of the school where I was teaching said, "Maybe so" and ask me if I knew that snowflakes were mentioned in the Bible as being important. I told him, "No, I didn't know that," but I looked it up. It said: "Hast thou entered into the treasures of the snow? Or hast thou seen the treasures of the hail? Which I have reserved against the time of trouble, against the day of battle and war."

My comment was, "Well, maybe this is more important than I thought." So I went ahead and worked on it another ten years.

I went to the Library of Congress and looked up snowflakes. I found a wonderful book there by Dr. Bentley of New Hampshire. He has spend many years making these studies, and he had learned a lot, as well as turning out one of the world's most beautiful books. He had found that snowflakes have gas pockets oriented on 60° angles and that the gas has a higher percentage of oxygen than air. That's one reason why snow water rusts so well. This higher concentration of oxygen also interested me because oxygen is more attracted to a magnetic field than other gases.

Finally, using the best ceramic magnets I could find and the best metal magnets, I worked out a scheme for a linear motor. The stator would be laid out as if it were unwound from around a motor. The parts of the armature would ride just above the stator and have the same beveled angular orientations I have just mentioned.

Dies were made for the curved armature magnets, and an order was placed for these shapes, despite the objections of magnet manufacturers who said it was a bad design. They didn't know what it was for, but they were sure it was a bad design. They wanted to make horseshoe magnets. They even begged me to content myself with half an order. I did not agree --- and once again you have that little matter of faith; faith to try to implement a new theory; faith to spend your own limited funds when you have a a family and other financial responsibilities staring you in the face; faith to buck the recognized authorities and manufacturers in the field; faith to believe that your work is good and that some day, despite all the hazards, you will apply for and receive patent rights in your own country and perhaps throughout the rest of the world; and finally, faith that you can resist being smashed into dust by industrial giants and/or being robbed by others who know only how to steal.

Believe it or not, my first motor assembly showed about two pounds of thrust. The little toy car on which I fastened the armature magnets for support ran in both directions over the stator, showing that the focusing and timing of the interactions was not too bad.

This was the first light at the end of a rather dark tunnel I had been traveling for many years. I breathed a real sign of relief as my young son played with this "new toy," and was able to operate it as easily as I could.

After much testing of linear and circular designs, and looking for an attorney for years suited to securing a patent on the new theoretical work, I was led to Dunkan Beaman of Beaman & Beamon in Jackson, Michigan. It took some time to prepare the patent. The attorney built some models himself to check certain parameters. Finally, we entered the case in the patent office expecting a lot of opposition. We were correct. We got it. But again, faith saved the day as we battled for many years to gain a rather complete victory.

Now the work requires different kinds of faith: faith in those who have taken cut licenses and who will license; faith to continue the research to replace scarce materials in the magnets; and faith that this work will continue to progress and that it will eventually fulfill its goal.

For a number of reasons, the permanent magnet motor has not received much consideration. In fact, nothing too radical has been done since Faraday took some very crude materials and showed the world that it was possible to make a motor. This work of his largely influenced the thinking of Clerk Maxwell and others who followed.

Today, the two greatest obstacles to using a permanent magnet motor are, first, the belief that it violates the conservation of energy law; and, secondly, that the magnetic fields of attraction and repulsion decrease according to the inverse square law then the air gap is increased.

In fact, both contentions are quite wrong because they are based on wrong considerations.

The permanent magnet is a long time energy source. This has been shown for many years in the rating of magnets as high or low energy sources for many applications over long usage.

A loudspeaker composed entirely of electromagnets would be unreal in size and energy consumption. Yet, despite examples of this type, many hesitate to apply the same principles to motors and extend them even further by using permanent magnets for both the stator and armature.

The elements of all electric and permanent magnet motors are similar. A field imbalance must be created, the fields must be focused and timed, and magnetic leakage must be controlled.

In the wound motor, brushes and contact rings give the timing, the size and shape of the wound fields and poles gives the focusing, and the motor case and kind of iron used help to limit the leakage.

In our permanent magnet motors the timing is built into the motors by the size, shape, and spacing of the magnets in the stator and armature. The focusing is controlled by the shape of the magnets, pole length, and the width of the air gap. This air gap, through which magnets oppose and attract each other, is a rare phenomenon. Usually when a magnetic air gap is increased, the field decreases inversely as the square.

When the air gap of the permanent magnet motor is increased, a curious but definite change takes place. There is a large decrease in the reading at south pole of the armature and an increase in the reading at the north pole. Thus, a Hall-effect sensing probe will give a higher gauss reading at the north pole and a decreasing count at the south pole. This helps explain why the thrust is better with a larger air gap than a smaller one. The attracting field is minimized and will not produce a locking force, while the repulsion of the crescent magnet is great enough to generate a thrust vector component that will drive the armature.

As I tried to explain in the patent, I believe that the permanent magnet is the first room temperature super conductor. In fact, I believe that super conductors are simply large wound magnets. The current in a super conductor is not initiated by a strong emf, such as a battery, but is instead actually induced into existence by a magnetic field. Then, in order to determine how much current may be flowing in the super conductor coil, we measure its magnetic field. This appears to be something like going out the door and coming back in the window.

Another rather unique feature of super conductors is the fact that their magnetic lines of force experience a change in direction. No longer do these lines flow at right angles to the conductor, but they now exist parallel to the conductor. Theoretically, the heavy conductor currents exist in the fine filaments of niobium within each small wire of niobium tin from which such super conductors are made. Isn't it interesting that the finer the wire the less the resistance until eventually there is no resistance at all?

II. Theoretical Analysis (presented by William P. Harrison, Jr.)

1. Introduction

Despite the fact that the linear version of the permanent magnet motor (Johnson, 1979) may appear conceptually simple (see Fig. 1), the complex interactions of the fields alone place it in a class with other quite sophisticated motive systems.

Figure 1: Partial Front and Plan Views of a Linear Model of the Howard Johnson Permanent Magnet Motor

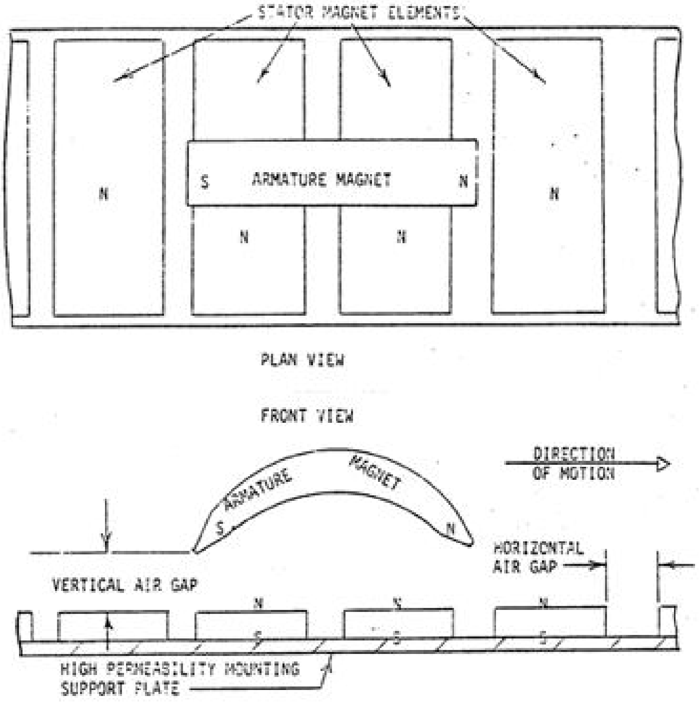

Many parameters play an important part in making possible the successful design of a permanent magnet motor. A number of these variables relate directly to the geometry of the system and its components. Mathematical models for both the linear and circular versions of Mr. Johnson's motors are presently under development, and include such controllable parameters as stator-to-armature air gap, stator element air gap spacing, armature pole length, stator magnet dimensions, magnet material variations, magnetic permeability and geometry of backing metals, and multiple armature couplings, to mention only a few. However, much of the early work involved quit simple mathematical investigations, and even at this level some remarkable revelations resulted. Also, as often is true with simple models, considerable insight into the mechanisms that might prove predominant was gained. Therefore, it is our intention to share with you some of those early analytical investigations and findings.

Even though Coulomb's Law, embodying the inverse square relationship as it does, may yet prove suspect, it nevertheless provides an exceedingly simple yet viable form upon which to base an elementary model of the linear version of the permanent magnet motor. Describing the interaction between two magnetic monopoles, Coulomb's Law in vector form is recalled as

(1) $$\bar{f} = \frac{\overline{MM'}}{ur^2}$$

where M and M' are the pole strengths (positive if north, negative if south), *u* [mu] is the permeability of the medium in which the poles are located, *r* is the straight-line separation distance between the two poles, and *f* [*f* with line over top] is the vector of force (see Fig. 2) acting at each pole (positive in magnitude for repulsion and negative for attraction).

Figure 2: Coulomb's Law

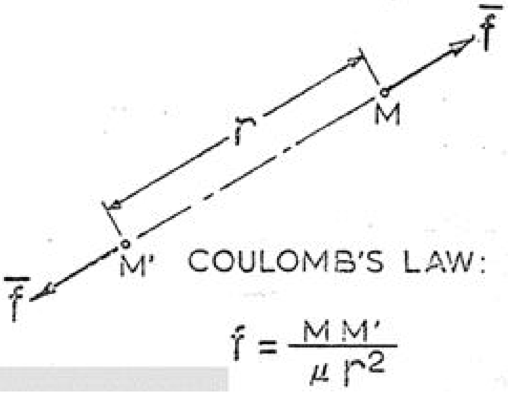

The vector nature of Eq. (1), the fact that f's line of action is colinear with the straight-line distance *r* between poles, its superposition properties when applied to multiple poles, and its restriction to static systems fixes in space are all well known conditions on Eq. (1). We will use the superposition property of Eq. (1) to extend its application to a spatial domain containing many more poles than the two shown in Fig. 2. However, Eq. (1) will first be resolved into scalar components so that analytical expressions can be more easily developed.

Our analysis will be two-dimensional and coplanar, restricted to the vertical x-y plane. It should be noted that the horizontal stator "track" of H.R. Johnson's linear model comprises a plurality of flat magnets, rectangular in cross section, each having an aspect ratio (length-to-thickness ratio) of 16. This high value contributes to the two-dimensional nature of the model and helps to minimize and effects in the z direction. Thus there is some justification for a two-dimensional analysis, at least in the case of the linear model we are considering here.

Figure 3: Positional Locations of Two Opposing North monopoles in X-Y Space ~

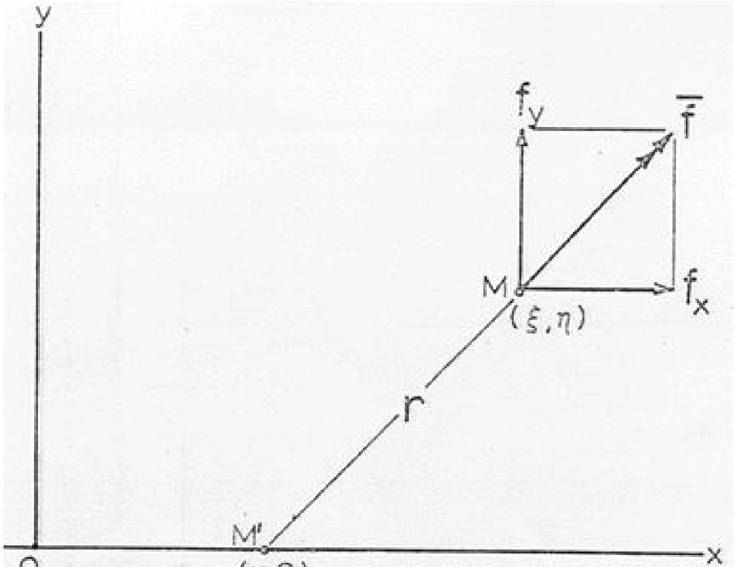

As shown in Fig. 3, we consider first a north pole of strength M located at coordinates (*E* [epsilon], *n* [nu]) with a second north pole of strength M', located on the x-axis at (x,0). Force *f*, acting on the monopole at (*E*,*n*), when resolved into its horizontal and vertical components yields, respectively,

$$f_x = \frac{MM'(\xi - y)}{\mu[(\xi - x)^2 + \eta^2]^{3/2}}$$

(2)

and

$$f_y = \frac{MM'\eta}{\mu[(\xi - x)^2 + \eta^2]^{3/2}}$$

(3)

2. The Attractive Sheet

Figure 4: Spatial Orientation of Thin, Magnetized Sheet having high aspect ratio and with S side face up

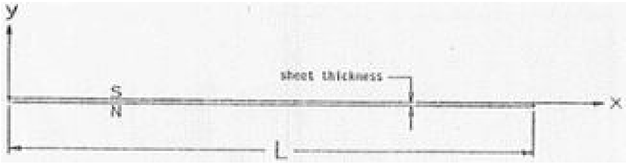

To illustrate some of the assumptions and extensions of Coulomb&rsquo;s Law that will be made, the simple example of a magnetic sheet lying along the x-axis will be considered first (see Fig. 4). The sheet, of finite length L, is a permanent magnet magnetized across its y-direction thickness and having a high aspect ratio (to eliminate z-direction edge effects). The south-pole face will be oriented up, with north facing downward on the underside of the sheet. Underside effects will be ignored as though the sheet represented a continuous distribution of only south monopoles along the x-axis. To incorporate such distributions into Eq. (1) we replace M&rsquo; with the differential dM&rsquo; and introduce the function B(x) so that

(4) dM&rsquo; = B(x) dx

Then the magnitude of the total force vector, F, acting on an isolated north monopole of strength M situated somewhere within the upper half of the x-y plane, becomes

$$F = \frac{ML}{\mu}\int_0 \frac{B(x)}{r^2}\,dx$$

(5)

where x is the ratio x/L. Assuming that the magnetic density along the sheet can be represented by the southern constant *-B*, and neglecting end effects at x = 0 and x = L, Eq (5) reduces to

$$F = -\mathcal{F}\int_0^1 \frac{1}{\rho^2}\, dx$$

(6)

where

$$\mathcal{F} \equiv \left|\frac{MM'}{\mu L^2}\right|$$

(7)

the strength parameter M&rsquo; having been determined by integrating Eq (4) over the sheet length L, and *p* is the ratio r/L.

Figure 5: North Monopole Positioned Symmetrically above the center of a magnetized, attracting sheet

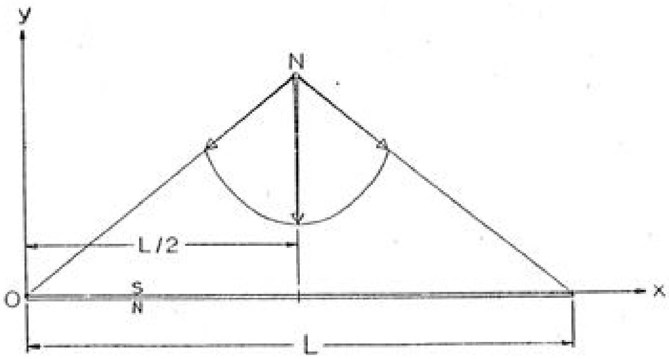

Figure 6: Force Imbalance Acting on a North Monopole above a mangetized sheet tending to restore the pole to sheet center

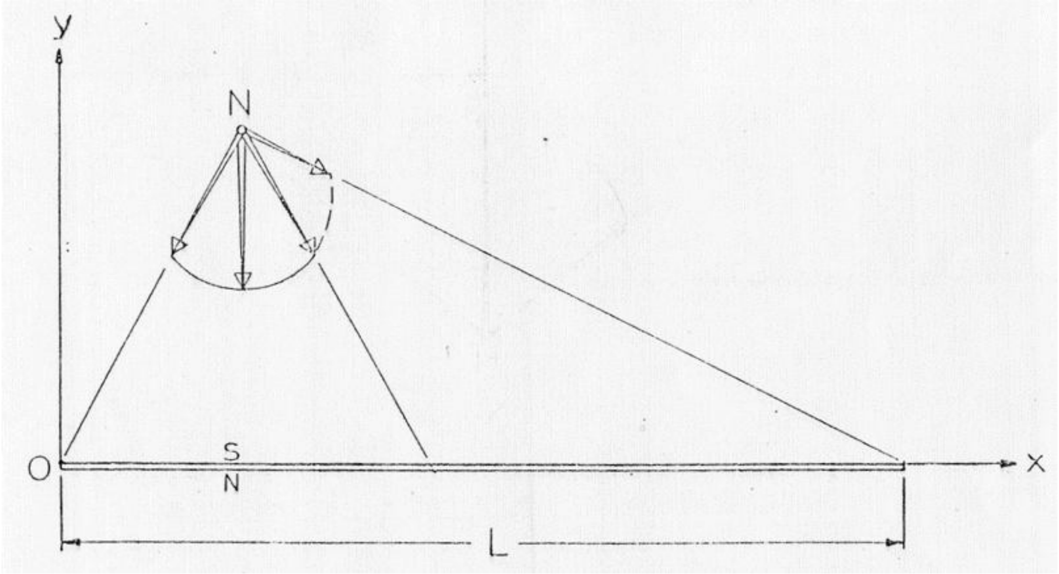

If the north monopole is placed directly above the center of the sheet, at coordinates (*E, n*), with *E* = L/2 and the vertical air-gap separation distance n taken as arbitrary, the symmetrical distribution of incremental force vectors acting at (*E,n*) will appear as shown in Fig. 5. Note that a shift of the north monopole to the left results in a force imbalance which tends to pull the pole back to the right, as shown in Fig. 6. So considering now only the x-component of F, similar to Eq (2) we write

(8) $$F_X = \mathcal{F}\int_0^1 \frac{X - x}{[(X - x)^2 + Y^2]^{3/2}}\, dx$$

where X and Y are the dimensionless ratios

(9) $$X \equiv \xi/L$$

and

(10) $$Y \equiv \eta/L$$

For any fixed position (X,Y) of the north monopole in the upper half plane, Eq (8) can be integrated to give

(11) $$\frac{F_X}{\mathcal{F}} = \frac{1}{\sqrt{X^2 + Y^2}} - \frac{1}{\sqrt{(X - 1)^2 + Y^2}}$$

Figure 7: X-Direction Distribution of the X-Component of attractive force exerted on a north monopole by a thin, magnetized sheet

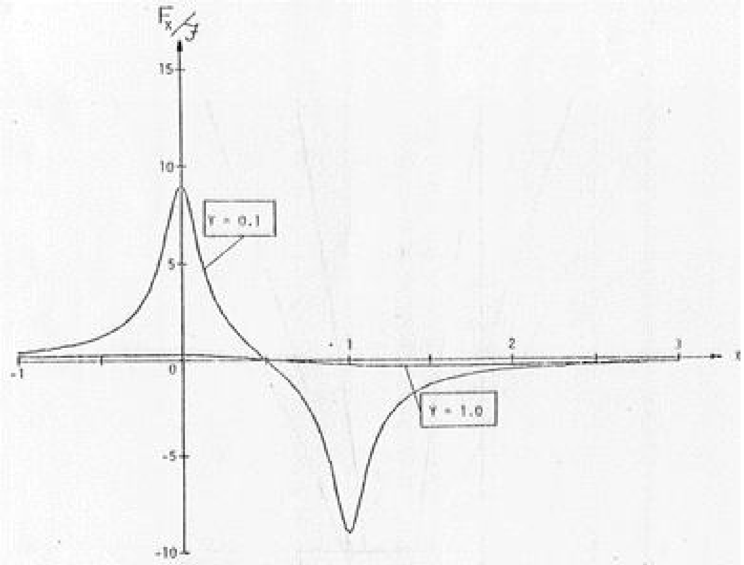

This ratio is shown in Fig (7) as a continuous function of X locations with Y treated parametrically. The Y = 1 curve represents the field influence on the north monopole situated at a constant air-gap separation (*n* = L) quite some vertical distance above the sheet; whereas at Y = 0.1 the monopole is located much closer to the x axis. Reversal of the force component through its zero value at mid-sheet (X = ½) is clearly shown.

In order to trace some trajectories through this field, we now observe that the y-component of force *F* will be

$$\frac{F_Y}{\mathcal{J}} = -\frac{1}{Y}\left[\frac{X}{\sqrt{X^2 + Y^2}} - \frac{X - 1}{\sqrt{(X - 1)^2 + Y^2}}\right] .$$

(12)

This function is shown in Fig (8) with a Y value of 0.20

Figure 8: [Missing]

In dimensionless form the equations of motion for trajectory paths of the monopole above the sheet in planar X-Y space become

(13) $$\frac{d^2X}{d\tau^2} = E(r_X)_{mag}$$

and

(14) $$\frac{d^2Y}{d\tau^2} = E(r_Y)_{mag} - G$$

where

(15) $$\tau \equiv \frac{t}{T}$$

(16) $$G \equiv \frac{gT^2}{L}$$

and

(17) $$E \equiv \frac{G}{(W/\mathcal{F})}$$

In these expression *t* is real time and *T* is simply a time constant chosen arbitrarily. As previously noted, *L* is the length of the sheet; whereas, *g* is the gravitational acceleration constant and *W* is the downward weight force of the moving monopole above the sheet. For magnetic force terms $(r_x)_{mag}$ and $(r_y)_{mag}$ we substitute directly Eq (11) and Eq (12), respectively.

Several of the trajectories resulting from the integration of Eq (13) and Eq (14) are shown in Fig.9. They all exhibit the expected behavior. As already implied in the discussion of Fig. 7, the function $(r_x)_{mag}$ given by Eq (11) has a stable point of equilibrium at X = ½ and therefore drives the free-falling monopole towards the sheet center, regardless of the initial drop-point location. The function $(r_y)_{mag}$ from Eq (12) is equally persuasive in pulling the monopole down towards the sheet itself, and manifests that attraction quite pervasively throughout the integration of Eq (14), even when the *G* term may be omitted (as it was in the trajectories of Fig. 9). Actually, the computer integration procedure will not carry the monopole all the way to surface contact with the sheet at Y = 0 because of the infinite condition which exists there as reflected by Eq (12). Thus, tailings of these trajectories (Fig 9) have been completed by manually overriding the plotter.

Figure 9: Trajectories of a North Monopole in an attractive field generated by the thin, magnetized sheet lying in the X-interval 0-1

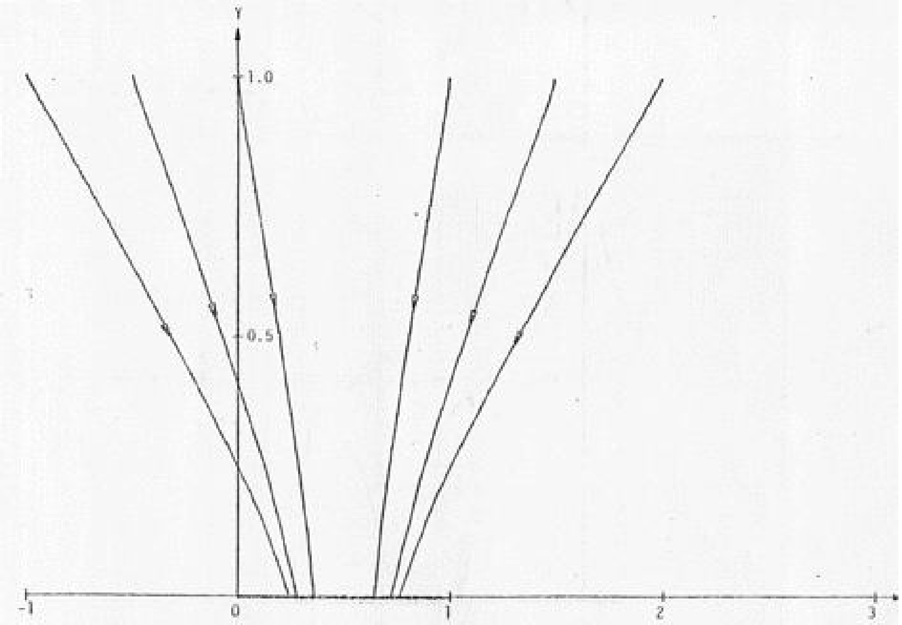

As we would anticipate in working with this type of central field, where *B* in Eq (4) is a simple constant, the field is conservative with curl of F vanishing. Also, the reverse symmetry of $(r_x)_{mag}$ about X = ½, as seen in Fig. 7, confirms that the energy integral for this function will vanish without any appropriate limit pairs of X.

3. The Repulsive Sheet

By substituting +*B* instead of -*B* for *B* in Eq (4), the sheet of length L lying along the x-axis becomes repulsive, with its northern face directed upward, opposing the north monopole above it at location (*E,n*). Of course the sign in Eq (6) becomes positive and the functions $(r_x)_{mag}$ and $(r_y)_{mag}$ reverse their behavior accordingly, as illustrated in Fig. 10. Again $(r_x)_{mag}$ will have an equilibrium point at X = ½, but now it is destabilizing. As a consequence, resulting trajectories for the north monopole are much more interesting in this case than they were with the attractive sheet. Several paths are shown in Fig 11 with different values used for the W/Ɉ trajectory in Eq (17). Parameter G was included, and in each example the trajectories commenced at (0.9, 0.2) with zero initial velocity.

Figure 10: X-Direction Distributions of $(r_x)_{mag}$ and $(r_y)_{mag}$ for the repulsive field of a thin, magnetized sheet acting on a moving north monopole

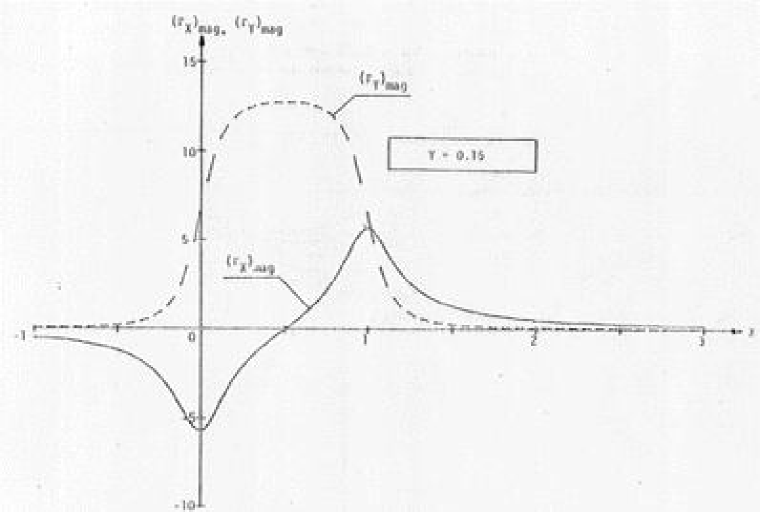

Figure 11: Trajectories of a North Monopole in a repulsive field generated by a thin, magnetized sheet lying in the X-interval 0-1

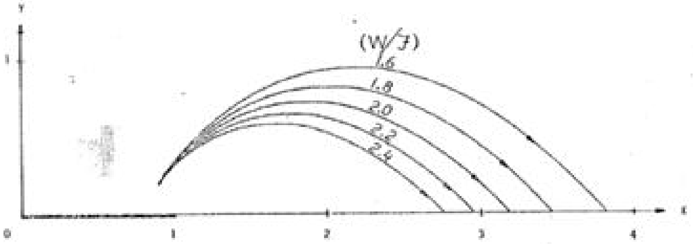

The attractive and repulsive sheet results are easily demonstrated since rubberized flexible sheet magnets are commercially available, such as those sold by the Permag Corp. of Jamaica, NY. It may also be interesting to note that with slight modifications this first simple analytical sheet model can be used to gain some insight into operation of the so-called "magnetic Wankel" reported on by Scott (1979).

Figure 12: Pole Strength Influence Factor, M', as a cosine function of linear displacement distance, x

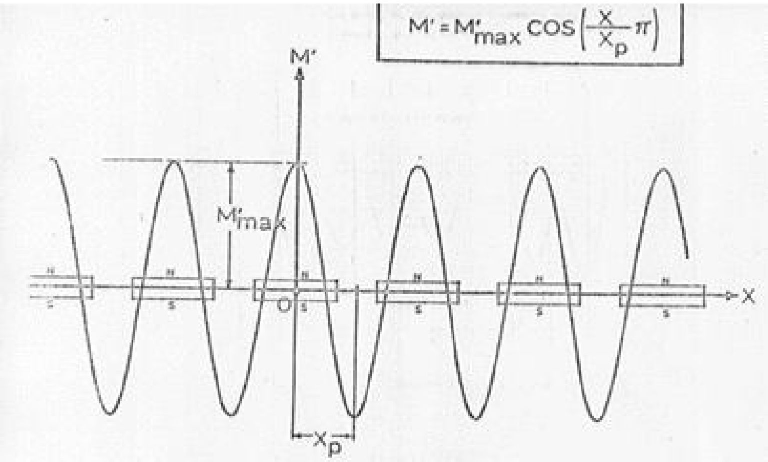

Figure 13: Experimentally Detemined Magnetic Flux Density, B, along a linear model of the Johnson permanent magnet motor

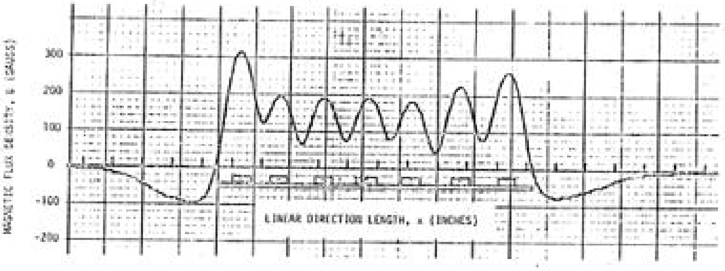

4. The Sinusoidal Model

The first paper (Harrison, 1979) relating, indirectly, to any mathematical analysis of the permanent magnet motor adopted a cosine function (Fig 12) to simulate the distribution of influence parameter M&rsquo; generated by the flat stator track of Mr Johnson&rsquo;s linear model. An experimentally determined distribution, shown in Fig 13, was obtained by moving a Hall-effect probe over the stator track of one of Mr Johnson&rsquo;s early linear models having seven flat ceramic magnet elements. The figure shown was produced by a plotter connected directly to the monitor computer controlling positioning of the Hall probe and processing its output signal. Ordinate values on the graph are magnetic flux density in gauss measured relative to a predetermined background value. These direct-reading experimental results suggest that the function

$$B(x) = \beta \cos(2\pi x/x_p) \tag{18}$$

substituted into Eq (4) should prove interesting to pursue as a more challenging test of what might be gleaned from this simple Coulomb model we have been discussing. It should be noted that one of the important differences between the function (18) and that shown in Fig 12 is that in Eq (18) the period length parameter xp is double that shown in Figure 12.

Using Eq (18), the total force magnitude expression Eq (5) becomes

$$(\Gamma_X)_{mag} \equiv \frac{F_X}{\mathcal{J}} = \int_0^1 \frac{(X - x)\cos(2\pi x/x_p)}{[(X - x)^2 + Y^2]^{3/2}}\, dx \tag{19}$$

where a total track length distance of L has been used to form the dimensionless ratios p = r/L, x = x/L, and xp = xp/L. Also, if Eq (7) is used for $\mathcal{J}$ in Eq (19), then in that expression one must substitute the product BL for M&rsquo;.

Now we plan to hold Y constant while investigating linear motion of the monopole along this track in the X-direction only. So we need consider only the X-component of $\underline{F}$ from Eq (19) which yields

$$(\Gamma_X)_{mag} \equiv \frac{F_X}{\mathcal{J}} = \int_0^1 \frac{(X - x)\cos(2\pi x/x_p)}{[(X - x)^2 + Y^2]^{3/2}}\, dx \tag{20}$$

Figure 14: Oscillatory Path of a North Monopole restrained to x-direction motion over a three-element linear stator assembly

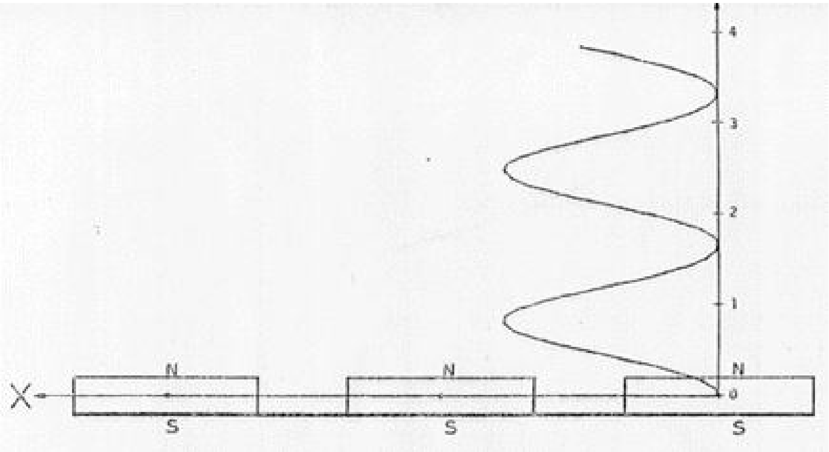

With this expression substituted into Eq (13), integration becomes straightforward and yields the typical oscillatory type of trajectory path shown in Fig 14. As Mr Johnson has brought out, the focusing armature magnet of his linear model will start at either end of the stator track simply by insuring that the north end of this bipoled crescent is leading the south (see Fig. 1). So, in Fig. 14, we are showing the X-direction motion from right to left instead of from left to right as in our previous examples. Also, by simply rotating the figure clockwise through 90 degrees, it becomes easy to follow the behavior of dimensionless velocity, V_X, in Fig 11, since Vx is defined as

$$V_X = \frac{dX}{d\tau} \tag{21}$$

It will be noted in Fig 14 that the north monopole has been allowed to self-start its motion at the origin with V_X initially zero.

We now discuss out final adjustment which proved to be an exciting revelation at the time it was first investigated several months ago. Johnson (1979, col. 5, line 39) states that the horizontal air-gap spacing between the magnet elements which the stator track comprises should vary slightly from normal in order to smooth out movement of the armature. Introducing this type of variation into a two-dimensional model, provided the charge is nonuniform, would certainly transform the field from conservative to nonconservative. It should by now be apparent that only a nonconservative model has any chance at all of even partially explaining the phenomena of the permanent magnet motor.

With these thoughts in mind, an attempt was made to drive the armature monopole of Fig 14 on to the second stator magnet and beyond by varying the horizontal gap parameter xp during the integration process (i.e., during the motion). The results are shown in Fig 15. It was found that through small variations in *xp* in Eq (20), as the monopole advanced along its trajectory path from one X position to another, sufficient control over the moving pole could be exercised to carry it over the full length of the stator and beyond.

Figure 15: Continuus Path of a North Monopole restrained to x-direction motion shown traversing a linear stator assembly comprised of seven permanent magnet elements

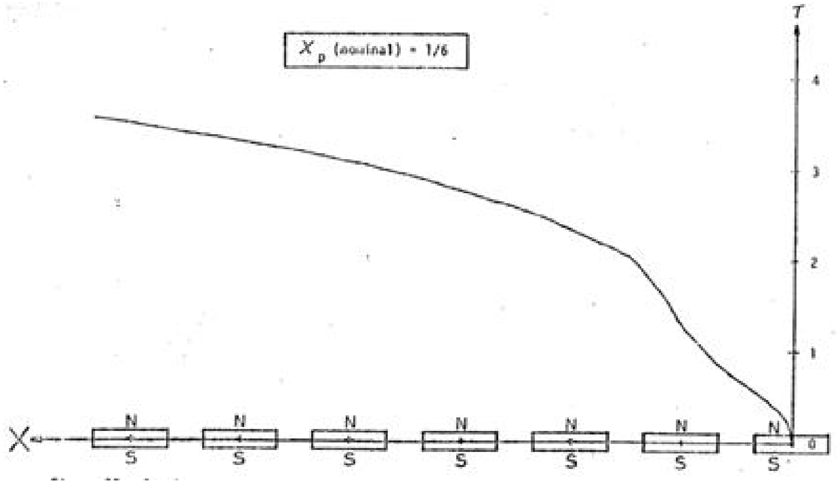

III. References

Harrison, William P., Jr.: "A Solution for the Optimal Gap of a Monopole Element Moving in a Sinusoidally Distributed Magnetic Field", paper presented to the Engineering Section, Virginia Academy of Science, 57th Annual Meeting, Richmond VA, May 8-11, 1979.

Johnson, Howard R: US Patent # 4,151,431 (April 24, 1979), "Permanent Magnet Motor".

Scott, David, "Magnetic &lsquo;Wankel&rsquo; for Electric Cars", *Popular Science*, p. 80, June 1979.

US Patent # 4,151,431

Permanent Magnet Motor

(April 24, 1979)

Howard R. Johnson

Abstract --- The invention is directed to the method of utilizing the unpaired electron spins in ferro magnetic and other materials as a source of magnetic fields for producing power without any electron flow as occurs in normal conductors, and to permanent magnet motors for utilizing this method to produce a power source. In the practice of the invention the unpaired electron spins occurring within permanent magnets are utilized to produce a motive power source solely through the superconducting characteristics of a permanent magnet and the magnetic flux created by the magnets are controlled and concentrated to orient the magnetic forces generated in such a manner to do useful continuous work, such as the displacement of a rotor with respect to a stator. The timing and orientation of magnetic forces at the rotor and stator components produced by permanent magnets to produce a motor is accomplished with the proper geometrical relationship of these components.

Inventors: Johnson; Howard R. (3300 Mt. Hope Rd., Grass Lake, MI 49240)
Appl. No.: 422306 ~ Filed: December 6, 1973

Current U.S. Class: 310/12; 310/152; 415/10; 415/916; 416/3; 505/877
Intern'l Class: H02K 041/00; H02N 011/00
Field of Search: 24/DIG. 9 415/DIG. 2 46/236 ;134 A;135 A;136 B;137 AE;138 A 273/118 A,119 A,120 A,121 A,122 A,123 A,124,125 A, 126 A,130 A,131 A,131 AD

References Cited:

U.S. Patent Documents
4,074,153 (Feb., 1978) Baker, et al. 310/12.

Description

FIELD OF THE INVENTION

The invention pertains to the field of permanent magnet motor devices solely using the magnetic fields created thereby to product motive power.

BACKGROUND OF THE INVENTION

Conventional electric motors employ magnetic forces to produce either rotative or linear motion. Electric motors operate on the principle that when a conductor is located in a magnetic field which carries current a magnetic force is exerted upon it.

Normally, in a conventional electric motor, the rotor, or stator, or both, are so wired that magnetic fields created by electromagnetics may employ attraction, repulsion, or both types of magnetic forces, to impose a force upon the armature to cause rotation, or to cause the armature to be displaced in a linear path. Conventional electric motors may employ permanent magnets either in the armature or stator components, but in the art heretofore known the use of permanent magnets in either the stator or armature require the creation of an electromagnetic field to act upon the field produced by the permanent magnets, and switching means are employed to control the energization of the electromagnets and the orientation of the magnetic fields, to produce the motive power.

It is my belief that the full potential of magnetic forces existing in permanent magnets has not been recognized or utilized because of incomplete information and theory with respect to the atomic motion occurring within a permanent magnet. It is my belief that a presently unnamed atomic particle is associated with the electron movement of a superconducting electromagnet and the lossless current flow of Amperian currents in permanent magnets. The unpaired electron flow is similar in both situations. This small particle is believed to be opposite in charge and to be located at right angles to the moving electron, and the particle would be very small as to penetrate all known elements, in their various states as well as their known compounds, unless they have unpaired electrons which capture these particles as they endeavor to pass therethrough.

Ferro electrons differ from those of most elements in that they are unpaired, and being unpaired they spin around the nucleus in such a way that they respond to magnetic fields as well as creating one themselves. If they were paired, their magnetic fields would cancel out. However, being unpaired they create a measurable magnetic field if their spins have been oriented in one direction. The spins are at right angles to their magnetic fields.

In niobium superconductors at a critical state, the magnetic lines of force cease to be at right angles. This change must be due to establishing the required conditions for unpaired electronic spins instead of electron flow in the conductor, and the fact that very powerful electromagnets that can be formed with superconductors illustrates the tremendous advantage of producing the magnetic field by unpaired electron spins rather than conventional electron flow.

In a superconducting metal, wherein the electrical resistance becomes greater in the metal than the proton resistance, the flow turns to electron spins and the positive particles flow parallel in the metal in the manner occurring in a permanent magnet where a powerful flow of magnetic positive particles or magnetic flux causes the unpaired electrons to spin at right angles. Under cryogenic superconduction conditions the freezing of the crystals in place makes it possible for the spins to continue, and in a permanent magnet the grain orientation of the magnetized material results in the spins permitting them to continue and for the flux to flow parallel to the metal.

In a superconductor, at first the electron is flowing and the positive particle is spinning; later, when critical, the reverse occurs, i.e., the electron is spinning and the positive particle is flowing at right angles. These positive particles will thread or work their way through the electron spins present in the metal.

In a sense, a permanent magnet may be considered the only room temperature superconductor. It is a superconductor because the electron flow does not cease, and this electron flow can be made to do work because of the magnetic field it supplies.

Previously, this source of power has not been used because it was not possible to modify the electron flow to accomplish the switching functions of the magnetic field. Such switching functions are common in a conventional electric motor where electrical current is employed to align the much greater electron current in the iron pole pieces and concentrate the magnetic field at the proper places to give the thrust necessary to move the motor armature. In a conventional electric motor, switching is accomplished by the use of brushes, commutators, alternating current, or other known means.

In order to accomplish the switching function in a permanent magnet motor, it is necessary to shield the magnetic leakage so that it will not appear as too great a loss factor at the wrong places. The best method to accomplish this is to use the superconductor of magnetic flux and concentrate it to the place where it will be the most effective. Timing and switching can be achieved in a permanent magnet motor by concentrating the flux and using the proper geometry of the motor rotor and stator to make most effective use of the magnetic fields generated by the electron spins. By the proper combination of materials, geometry and magnetic concentration, it is possible to achieve a mechanical advantage of high ratio, greater than 100 to 1, capable of producing a continuous motive force.

To my knowledge, previous work done with permanent magnets, and motive devices utilizing permanent magnets, have not achieved the result desired in the practice of the inventive concept, and it is with the proper combination of materials, geometry and magnetic concentration that the presence of the magnetic spins within a permanent magnet may be utilized as a motive force.

SUMMARY OF THE INVENTION

It is an object of the invention to utilize the magnetic spinning phenomenon of unpaired electrons occurring in ferro magnetic material to produce the movement of a mass in a unidirectional manner as to permit a motor to be driven solely by magnetic forces as occurring within permanent magnets. In the practice of the inventive concepts, motors of either linear or rotative types may be produced.

It is an object of the invention to provide the proper combination of materials, geometry and magnetic concentration to utilize the force generated by unpaired electron spins existing in permanent magnets to power a motor. Whether the motor constitutes a linear embodiment, or a rotary embodiment, in each instance the "stator" may consist of a plurality of permanent magnets fixed relative to each other in space relationship to define a track, linear in form in the linear embodiment, and circular in form in the rotary embodiment. An armature magnet is located in spaced relationship to such track defined by the stator magnets wherein an air gap exists therebetween. The length of the armature magnet is defined by poles of opposite polarity, and the length of the armature magnet is disposed relative to the track defined by the stator magnets in the direction of the path of movement of the armature magnet as displaced by the magnetic forces.

The stator magnets are so mounted that poles of like polarity are disposed toward the armature magnet and as the armature magnet has poles which are both attracted to and repelled by the adjacent pole of the stator magnets, both attraction and repulsion forces act upon the armature magnet to produce the relative displacement between the armature and stator magnets.

The continuing motive force producing displacement between the armature and stator magnets results from the relationship of the length of the armature magnet in the direction of its path of movement as related to the dimension of the stator magnets, and the spacing therebetween, in the direction of the path of armature magnet movement. This ratio of magnet and magnet spacings, and with an acceptable air gap spacing between the stator and armature magnets, will produce a resultant force upon the armature magnet which displaces the armature magnet across the stator magnet along its path of movement.

In the practice of the invention movement of the armature magnet relative to the stator magnets results from a combination of attraction and repulsion forces existing between the stator and armature magnets. By concentrating the magnetic fields of the stator and armature magnets the motive force imposed upon the armature magnet is intensified, and in the disclosed embodiments such magnetic field concentration means are disclosed.

The disclosed magnetic field concentrating means comprise a plate of high magnetic field permeability disposed adjacent one side of the stator magnets in substantial engagement therewith. This high permeability material is thus disposed adjacent poles of like polarity of the stator magnets. The magnetic field of the armature magnet may be concentrated and directionally oriented by bowing the armature magnet, and the magnetic field may further be concentrated by shaping the pole ends of the armature magnet to concentrate the magnet field at a relatively limited surface at the armature magnet pole ends.

Preferably, a plurality of armature magnets are used which are staggered with respect to each other in the direction of armature magnet movement. Such an offsetting or staggering of the armature magnets distributes the impulses of force imposed upon the armature magnets and results in a smoother application of forces to the armature magnet producing a smoother and more uniform movement of the armature component.

In the rotary embodiment of the permanent magnet motor of the invention the stator magnets are arranged in a circle, and the armature magnets rotate about the stator magnets. Means are disclosed for producing relative axial displacement between the stator and armature magnets to adjust the axial alignment thereof, and thereby regulate the magnitude of the magnetic forces being imposed upon the armature magnets. In this manner the speed of rotation of the rotary embodiment may be regulated.

BRIEF DESCRIPTION OF THE DRAWINGS

The aforementioned objects and advantages of the invention will be appreciated from the following description and accompanying drawings wherein:

FIG. 1 is a schematic view of electron flow in a superconductor indicating the unpaired electron spins,

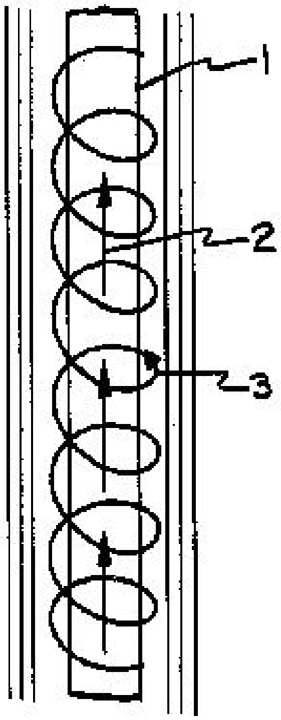

FIG. 1

FIG. 2 is a cross-sectional view of a superconductor under a critical state illustrating the electron spins,

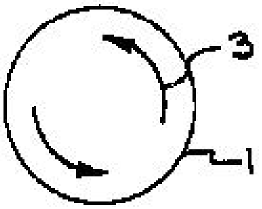

FIG. 2

FIG. 3 is a view of a permanent magnet illustrating the flux movement therethrough,

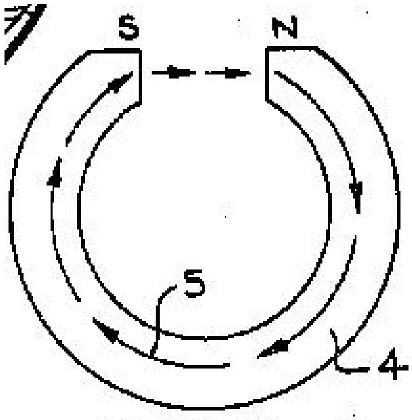

FIG. 3

FIG. 4 is a cross-sectional view illustrating the diameter of the magnet of FIG. 3,

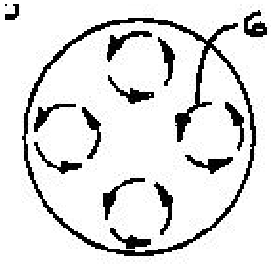

FIG. 4

FIG. 5 is an elevational representation of a linear motor embodiment of the permanent magnet motor of the invention illustrating one position of the armature magnet relative to the stator magnets, and indicating the magnetic forces imposed upon the armature magnet,

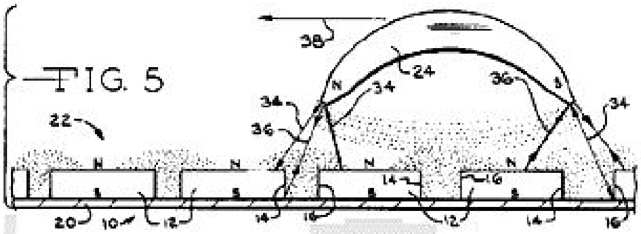
38
24
N
S
34
34
36
34
36
36
34
22
N
N
N
N
S
S
S
S
14
16
20
10
12
14
16
12
14
16

FIG. 5

FIG. 6 is a view similar to FIG. 5 illustrating displacement of the armature magnet relative to the stator magnets, and the influence of magnetic forces thereon at this location,

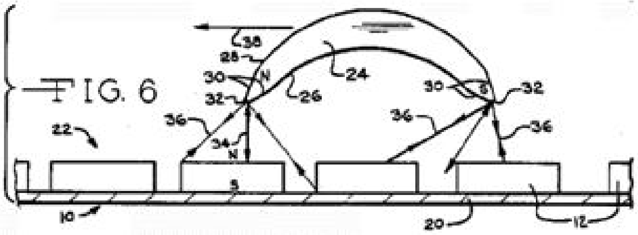
38
28
30
N
32
24
26
30
S
32
36
36
36
34
N
22
S
10
20
12

FIG. 6

FIG. 7 is a further elevational view similar to FIGS. 5 and 6 illustrating further displacement of the armature magnet to the left, and the influence of the magnetic forces thereon,

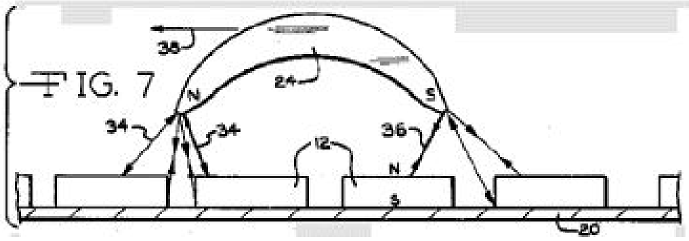
38
24
N
S
34
34
36
12
N
S
20

FIG. 7

FIG. 8 is a top plan view of a linear embodiment of the inventive concept illustrating a pair of armature magnets in linked relationship disposed above the stator magnets,

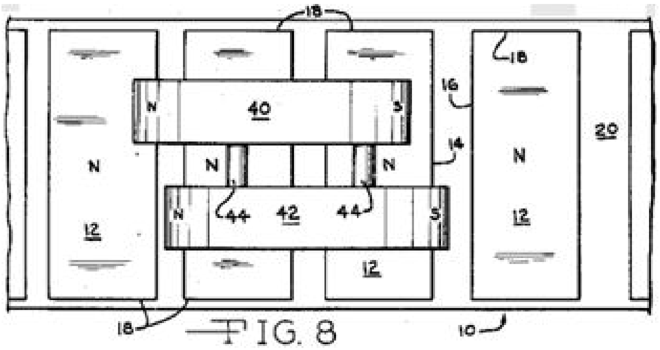
18
18
16
N
40
S
20
14
N
N
N
N
N
44
42
44
S
12
12
12
18
10

FIG. 8

FIG. 9 is a diametrical, elevational, sectional view of a rotary motor embodiment in accord with the invention as taken along section IX--IX of FIG. 10, and

FIG. 9

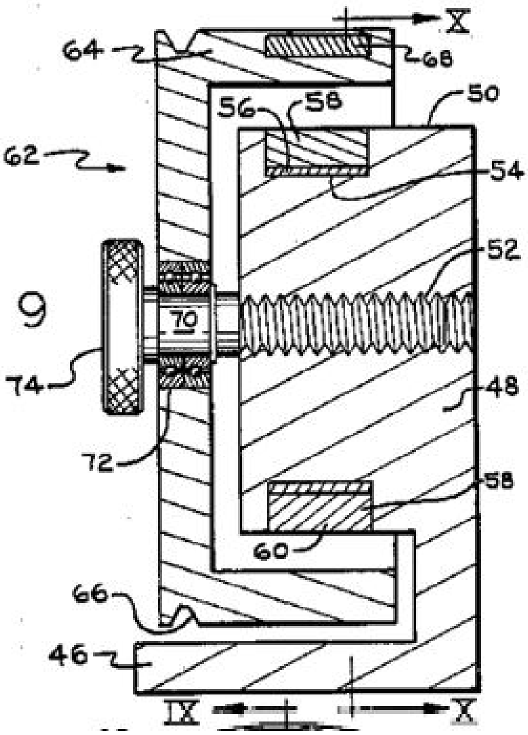
X
64
68
50
56
58
62
54
52
70
74
48
72
58
60
66
46
IX
X

FIG. 10 is an elevational view of the rotary motor embodiment as taken along section X--X of FIG. 9.

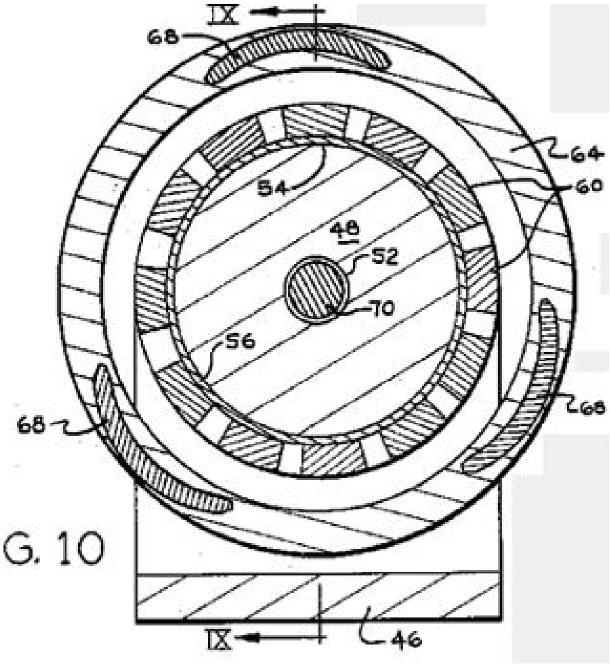

FIG. 10

DESCRIPTION OF THE PREFERRED EMBODIMENTS

In order to better understand the theory of the inventive concept, reference is made to FIGS. 1 through 4. In FIG. 1 a superconductor 1 is illustrated having a positive particle flow as represented by arrow 2, the unpaired electrons of the ferrous conducting material 1 spin at right angles to the proton flow in the conductor as represented by the spiral line and arrow 3. In accord with the theory of the invention the spinning of the ferrous unpaired electrons results from the atomic structure of ferrous materials and this spinning atomic particle is believed to be opposite in charge and located at right angles to the moving electrons. It is assumed to be very small in size capable of penetrating other elements and their compounds unless they have unpaired electrons which capture these particles as they endeavor to pass therethrough.

The lack of electrical resistance of conductors at a critical superconductor state has long been recognized, and superconductors have been utilized to produce very high magnetic flux density electromagnets. FIG. 2 represents a cross section of a critical superconductor and the electron spins are indicated by the arrows 3.

A permanent magnet may be considered a superconductor as the electron flow therein does not cease, and is without resistance, and unpaired electric spinning particles exist which, in the practice of the invention, are utilized to produce motor force. FIG. 3 illustrates a horseshoe shaped permanent magnet at 4 and the magnetic flux therethrough is indicated by arrows 5, the magnetic flow being from the south pole to the north pole and through the magnetic material. The accumulated electron spins occurring about the diameter of the magnet 5 are represented at 6 in FIG. 4, and the spinning electron particles spin at right angles in the iron as the flux travels through the magnet material.

By utilizing the electron spinning theory of ferrous material electrons, it is possible with the proper ferromagnetic materials, geometry and magnetic concentration to utilize the spinning electrons to produce a motive force in a continuous direction, thereby resulting in a motor capable of doing work.

It is appreciated that the embodiments of motors utilizing the concepts of the invention may take many forms, and in the illustrated forms the basic relationships of components are illustrated in order to disclose the inventive concepts and principles.

The relationships of the plurality of magnets defining the stator 10 are best appreciated from FIGS. 5 through 8. The stator magnets 12 are preferably of a rectangular configuration, FIG. 8, and so magnetized that the poles exist at the large surfaces of the magnets, as will be appreciated from the N (North) and S (South) designations. The stator magnets include side edges 14 and 16 and end edges 18. The stator magnets are mounted upon a supporting plate 20, which is preferably of a metal material having a high permeability to magnetic fields and magnetic flux such as that available under the trademark Netic CoNetic sold by the Perfection Mica Company of Chicago, Illinois. Thus, the plate 20 will be disposed toward the south pole of the stator magnets 12, and preferably in direct engagement therewith, although a bonding material may be interposed between the magnets and the plate in order to accurately locate and fix the magnets on the plate, and position the stator magnets with respect to each other.

Preferably, the spacing between the stator magnets 12 slightly differs between adjacent stator magnets as such a variation in spacing varies the forces being imposed upon the armature magnet at its ends, at any given time, and thus results in a smoother movement of the armature magnet relative to the stator magnets. Thus, the stator magnets so positioned relative to each other define a track 22 having a longitudinal direction left to right as viewed in FIGS. 5 through 8.

In FIGS. 5 through 7 only a single armature magnet 24 is disclosed, while in FIG. 8 a pair of armature magnets are shown. For purposes of understanding the concepts of the invention the description herein will be limited to the use of single armature magnet as shown in FIGS. 5 through 7.

The armature magnet is of an elongated configuration wherein the length extends from left to right, FIG. 5, and may be of a rectangular transverse cross-sectional shape. For magnetic field concentrating and orientation purposes the magnet 24 is formed in an arcuate bowed configuration as defined by concave surfaces 26 and convex surfaces 28, and the poles are defined at the ends of the magnet as will be appreciated from FIG. 5. For further magnetic field concentrating purposes the ends of the armature magnet are shaped by beveled surfaces 30 to minimize the cross-sectional area at the magnet ends at 32, and the magnetic flux existing between the poles of the armature magnet are as indicated by the light dotted lines. In like manner the magnetic fields of the stator magnets 12 are indicated by the light dotted lines.

The armature magnet 24 is maintained in a spaced relationship above the stator track 22. This spacing may be accomplished by mounting the armature magnet upon a slide, guide or track located above the stator magnets, or the armature magnet could be mounted upon a wheeled vehicle carriage or slide supported upon a nonmagnetic surface or guideway disposed between the stator magnets and the armature magnet. To clarify the illustration, the means for supporting the armature magnet 24 is not illustrated and such means form no part of invention, and it is to be understood that the means supporting the armature magnet prevents the armature magnet from moving away from the stator magnets, or moving closer thereto, but permits free movement of the armature magnet to the left or right in a direction parallel to the track 22 defined by the stator magnets.

It will be noted that the length of the armature magnet 24 is slightly greater than the width of two of the stator magnets 12 and the spacing therebetween. The magnetic forces acting upon the armature magnet when in the position of FIG. 5 will be repulsion forces 34 due to the proximity of like polarity forces and attraction forces at 36 because of the opposite polarity of the south pole of the armature magnet, and the north pole field of the sector magnets. The relative strength of this force is represented by the thickness of the force line.

The resultant of the force vectors imposed upon the armature magnet as shown in FIG. 5 produce a primary force vector 38 toward the left, FIG. 5, displacing the armature magnet 24 toward the left. In FIG. 6 the magnetic forces acting upon the armature magnet are represented by the same reference numerals as in FIG. 5. While the forces 34 constitute repulsion forces tending to move the north pole of the armature magnet away from the stator magnets, the attraction forces imposed upon the south pole of the armature magnet and some of the repulsion forces, tend to move the armature magnet further to the left, and as the resultant force 38 continues to be toward the left the armature magnet continues to be forced to the left.

FIG. 7 represents further displacement of the armature magnet 24 to the left with respect to the position of FIG. 6, and the magnetic forces acting thereon are represented by the same reference numerals as in FIGS. 5 and 6, and the stator magnet will continue to move to the left, and such movement continues the length of the track 22 defined by the stator magnets 12.

Upon the armature magnet being reversed such that the north pole is positioned at the right as viewed in FIG. 5, and the south pole is positioned at the left, the direction of movement of the armature magnet relative to the stator magnets is toward the right, and the theory of movement is identical to that described above.

In FIG. 8 a plurality of armature magnets 40 and 42 are illustrated which are connected by links 44. The armature magnets are of a shape and configuration identical to that of the embodiment of FIG. 5, but the magnets are staggered with respect to each other in the direction of magnet movement, i.e., the direction of the track 22 defined by the stator magnets 12. By so staggering a plurality of armature magnets a smoother movement of the interconnected armature magnets is produced as compared when using a single armature magnet as there is variation in the forces acting upon each armature magnet as it moves above the track 22 due to the change in magnetic forces imposed thereon. The use of several armature magnets tends to "smooth out" the application of forces imposed upon linked armature magnets, resulting in a smoother movement of the armature magnet assembly. Of course, any number of armature magnets may be interconnected, limited only by the width of the stator magnet track 22.

In FIGS. 9 and 10 a rotary embodiment embracing the inventive concepts is illustrated. In this embodiment the principle of operation is identical to that described above, but the orientation of the stator and armature magnets is such that rotation of the armature magnets is produced about an axis, rather than a linear movement being achieved.

In FIGS. 9 and 10 a base is represented at 46 serving as a support for a stator member 48. The stator member 48 is made of a nonmagnetic material, such as synthetic plastic, aluminum, or the like. The stator includes a cylindrical surface 50 having an axis, and a threaded bore 52 is concentrically defined in the stator. The stator includes an annular groove 54 receiving an annular sleeve 56 of high magnetic field permeability material such as Netic Co-Netic and a plurality of stator magnets 58 are affixed upon the sleeve 56 in spaced circumferential relationship as will be apparent in FIG. 10. Preferably, the stator magnets 58 are formed with converging radial sides as to be of a wedge configuration having a curved inner surface engaging sleeve 56, and a convex outer pole surface 60.

The armature 62, in the illustrated embodiment, is of a dished configuration having a radial web portion, and an axially extending portion 64. The armature 62 is formed of a nonmagnetic material, and an annular belt receiving groove 66 is defined therein for receiving a belt for transmitting power from the armature to a generator, or other power consuming device. Three armature magnets 68 are mounted on the armature portion 64, and such magnets are of a configuration similar to the armature magnet configuration of FIGS. 5 through 7. The magnets 68 are staggered with respect to each other in a circumferential direction wherein the magnets are not disposed as 120.degree. circumferential relationships to each other. Rather, a slight angular staggering of the armature magnets is desirable to "smooth out" the magnetic forces being imposed upon the armature as a result of the magnetic forces being simultaneously imposed upon each of the armature magnets. The staggering of the armature magnets 68 in a circumferential direction produces the same effect as the staggering of the armature magnets 40 and 42 as shown in FIG. 8.

The armature 62 is mounted upon a threaded shaft 70 by antifriction bearings 72, and the shaft 70 is threaded into the stator threaded bore 52, and may be rotated by the knob 74. In this manner rotation of the knob 74, and shaft 70, axially displaces the armature 62 with respect to the stator magnets 58, and such axial displacement will very the magnitude of the magnetic forces imposed upon the armature magnets 68 by the stator magnets thereby controlling the speed of rotation of the armature.

As will be noted from FIGS. 4-7 and 9 and 10, an air gap exists between the armature magnet or magnets and the stator magnets and the dimension of this spacing, effects the magnitude of the forces imposed upon the armature magnet or magnets. If the distance between the armature magents, and the stator magnets is reduced the forces imposed upon the armature magnets by the stator magnets are increased, and the resultant force vector tending to displace the armature magnets in their path of movement increases. However, the decreasing of the spacing between the armature and stator magnets creates a "pulsation" in the movement of the armature magnets which is objectionable, but can be, to some extent, minimized by using a plurality of armature magnets. The increasing of the distance between the armature and stator magnets reduces the pulsation tendency of the armature magnet, but also reduces the magnitude of the magnetic forces imposed upon the armature magnets. Thus, the most effective spacing between the armature magnets. Thus, the most effective spacing between the armature and stator magnets is that spacing which produces the maximum force vector in the direction of armature magnet movement, with a minimum creation of objectionable pulsation.

In the disclosed embodiments the high permeability plate 20 and sleeve 56 are disclosed for concentrating the magnetic field of the stator magnets, and the armature magnets are bowed and have shaped ends for magnetic field concentration purposes. While such magnetic field concentration means result in higher forces imposed upon the armature magnets for given magnet intensities, it is not intended that the inventive concepts be limited to the use of such magnetic field concentrating means.

As will be appreciated from the above description of the invention, the movement of the armature magnet or magnets resultsfrom the described relationship of components. The length of the armature magnets as related to the width of the stator magnets and spacing therebetween, the dimension of the air gap and the configuration of the magnetic field, combined, produce the desired result and motion. The inventive concepts may be practiced even though these relationships may be varied within limits not yet defined and the invention is intended to encompass all dimensional relationships which achieve the desired goal of armature movement. By way of example, with respect to FIGS. 4-7, the following dimensions were used in an operating prototype:

The length of armature magnet 24 is 31/8", the stator magnets 12 are 1" wide, 1/4" thick and 4" long and grain oriented. The air gap between the poles of the armature magnet and the stator magnets is approximately 11/2" and the spacing between the stator magnets is approximately 1/2" inch.

In effect, the stator magnets define a magnetic field track of a single polarity transversely interrupted at spaced locations by the magnetic fields produced by the lines of force existing between the poles of the stator magnets and the unidirectional force exerted on the armature magnet is a result of the repulsion and attraction forces existing as the armature magnet traverses this magnetic field track.

It is to be understood that the inventive concept embraces an arrangement wherein the armature magnet component is stationary and the stator assembly is supported for movement and constitutes the moving component, and other variations of the inventive concept will be apparent to those skilled in the art without departing from the scope thereof. As used herein the term "track" is intended to include both linear and circular arrangements of the static magnets, and the "direction" or "length" of the track is that direction parallel or concentric to the intended direction of armature magnet movement.

United States Patent 4,877,983

Magnetic Force Generating Method & Apparatus

Howard R. Johnson

(October 31, 1989)

Abstract --- A permanent magnet armature is magnetically propelled along a guided path by interaction with the field within a flux zone limited on either side of the path by an arrangement of permanent stator magnets.

Inventors: Johnson; Howard R. (Box 199, 314 N. Main, Blacksburg, VA 24060)
Appl. No.: 799618 ~ Filed: November 19, 1985

Current U.S. Class: 310/12; 310/152 ~ Intern'l Class: H02K 041/00
Field of Search: 310/152,12,46

References Cited:
U.S. Patent Documents

USP # 4,074,153 (Feb., 1978) Baker, et al. (Cl. 310/12).
USP # 4,151,431 (Apr., 1979) Johnson (Cl. 310/12).

Primary Examiner: Skudy; R. ~ Attorney, Agent or Firm: Fleit, Jacobson, Cohn, Price, Holman & Stern

Claims

What is claimed as new is as follows:

1. In combination with a movable armature, means for guiding movement of the armature along a predetermined path and a permanent armature magnet having magnetic poles of opposite polarity spaced from each other along said path to establish a magnetic field of limited extent movable with the armature and magnetic stator means for establishing a stationary magnetic flux zone along said path, the improvement comprising flux emitting surfaces of one polarity mounted on the stator means on opposite sides of said path for limiting said flux zone through which said path extends and means mounting the permanent armature magnet on the armature with the poles thereof orientated relative to said flux emitting surfaces on the stator means for unidirectionally propelling the armature along said path through the limited zone in response to magnetic interaction between the movable magnetic field and the limited flux zone, said magnetic stator means including a plurality of magnetic gate assemblies fixedly spaced from each other along said path and respectively establishing stationary magnetic fields, each of said gate assemblies including a plurality of interconnected bar magnets substantially bordering said limited flux zone exposing pole faces of opposite polarity in parallel spaced planes intersected by said path, and magnetic means connected to said interconnected bar magnets exposing one of the flux emitting surfaces of said one polarity perpendicular to said parallel planes for magnetic interaction of the stationary magnetic fields.

2. The combination of claim 1 wherein said armature magnet is curved between end faces at which said poles of opposite polarity are located, the end faces being orientated by the mounting means in converging relation to each other toward the guiding means.

3. In combination with a movable armature, means for guiding movement of the armature along a predetermined path and a permanent armature magnet having magnetic poles of opposite polarity spaced from each other along said path to establish a magnetic field of limited extent movable with the armature and magnetic stator means for establishing a stationary magnetic flux zone along said path, the improvement comprising flux emitting surfaces of one polarity mounted on the stator means on opposite sides of said path for limiting said flux zone through which said path extends and means mounting the permanent armature magnet on the armature with the poles faces thereof orientated relative to said flux emitting surfaces on the stator means for unidirectionally propelling the armature along said path through the limited zone in response to magnetic interaction between the movable magnetic field and the limited flux zone, said armature magnet being curved between end faces at which said poles of opposite polarity are located, the end faces being orientated by the mounting means in converging relation to each other toward the guiding means.

4. In combination with a movable armature, means for guiding movement of the armature along a predetermined path and a permanent armature magnet having

magnetic poles of opposite polarity spaced from each other along said path to establish a magnetic field of limited extent movable with the armature and magnetic stator means for establishing a stationary magnetic flux zone along said path, the improvement comprising flux emitting surfaces of one polarity mounted on the stator means on opposite sides of said path for limiting said flux zone through which said path extends and means mounting the permanent armature magnet on the armature with the poles thereof orientated relative to said flux emitting surfaces on the stator means for unidirectionally propelling the armature along said path through the limited zone in response to magnetic interaction between the movable magnetic field and the limited flux zone, said magnetic stator means including a pair of permanent magnet assemblies having continuous, confronting pole faces of said one polarity bordering said limited zone, each of said assemblies having means for varying magnetic field intensity in the flux zone along said path, and a second armature magnet connected to the first mentioned armature magnet in mirror image relation thereto.

5. The combination of claim 4 wherein said armature magnet is curved between end faces at which said poles of opposite polarity are located, the end faces being orientated by the mounting means in a plane parallel to said path.

6. In combination with a movable armature, means for guiding movement of the armature along a predetermined path and a permanent armature magnet mounted on the armature having magnetic poles of opposite polarity spaced from each other along said path, the improvement comprising a plurality of permanent magnet gate assemblies mounted in spaced relation to each other along said path establishing interacting stationary magnetic fields along said path, each of said assemblies including stator magnets interconnected in surrounding relation to said path and having pole faces of opposite polarity aligned with parallel planes intersected by said path and magnetic means fixed to the pole faces aligned with one of the parallel planes for interaction of the armature magnet with said stationary magnetic fields for unidirectional propulsion of the armature along said path, said magnetic means being an annular magnet having a radially inner pole surface of one polarity enclosing a magnetic flux zone through which said path extends.

Description

BACKGROUND OF THE INVENTION

This invention relates in general to the use of permanent magnets to generate unidirectional propelling forces.

The generation of unidirectional propelling forces by permanent magnets is already known and recognized in U.S. Pat. Nos. 4,151,431 and 4,215,330 to Johnson and Hartmen, respectively, by way of example. According to applicant's own prior Pat. No. 4,151,431, such forces are generated by magnetic interaction between a curved magnet bar of an armature guided for movement along a circular path and an arrangement of spaced stator magnets having pole faces of one polarity facing the armature on one side thereof parallel to the path of movement.

It is therefore an important object of the present invention to provide certain improved stator arrangements of permanent magnets interacting with a permanent magnet armature for unidirectional propulsion thereof in a novel manner believed to be more efficient.

SUMMARY OF THE INVENTION

In accordance with the present invention, the armature magnet is guided along a path through a magnetic flux zone limited on opposite sides of the path by an arrangement of magnetic pole surfaces of one polarity on stator magnets. According to one embodiment, the flux zone is formed by spaced gate assemblies of magnets having exposed pole faces of one polarity in a plane perpendicular to the armature path from which a magnetic field extends to the opposite pole faces and a ring magnet fixed to such opposite pole faces of the other polarity, with a radially inner pole surface of the same polarity producing a magnetic field perpendicular to the first mentioned field to their opposite radially outer pole surfaces.

According to another embodiment, the flux zone is formed between continuous confronting pole surfaces of one polarity on stator magnets arranged to produce a magnetic field of varying intensity along the armature path.

In yet another embodiment, at least two curved bar magnets are interconnected to form the armature with two pairs of pole faces spaced along the armature path.

These together with other objects and advantages which will become subsequently apparent reside in the details of construction and operation as more fully hereinafter described and claimed, reference being had to the accompanying drawings forming a part hereof, wherein like numerals refer to like parts throughout.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a somewhat schematic side elevational view showing an installation of the present invention in accordance with one embodiment, with parts broken away and shown in section.

FIG. 1

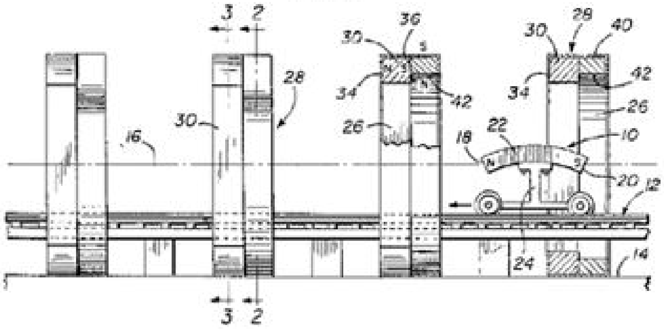

FIG. 2 is a transverse sectional view taken substantially through a plane indicated by section line 2--2 in FIG. 1.

FIG. 2

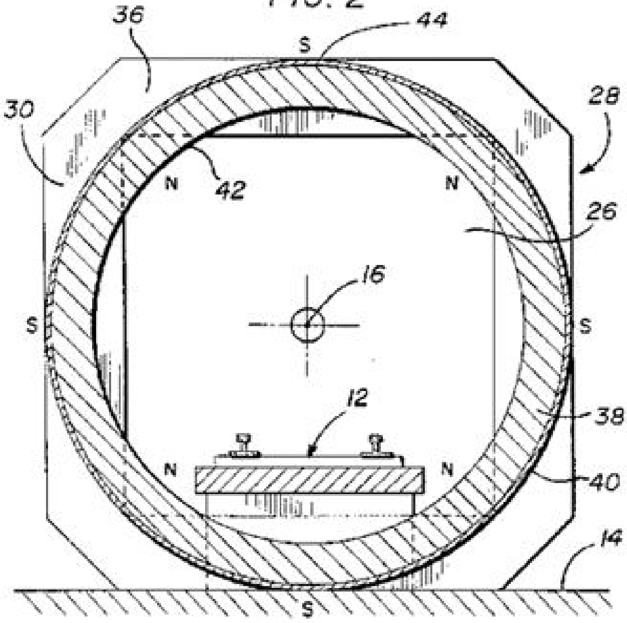
36
44
S
30
28
N
42
N
26
16
S
S
12
38
N
N
40
14
S

FIG. 3 is an enlarged partial sectional view taken substantially through a plane indicated by section line 3--3 in FIG. 1.

FIG. 3

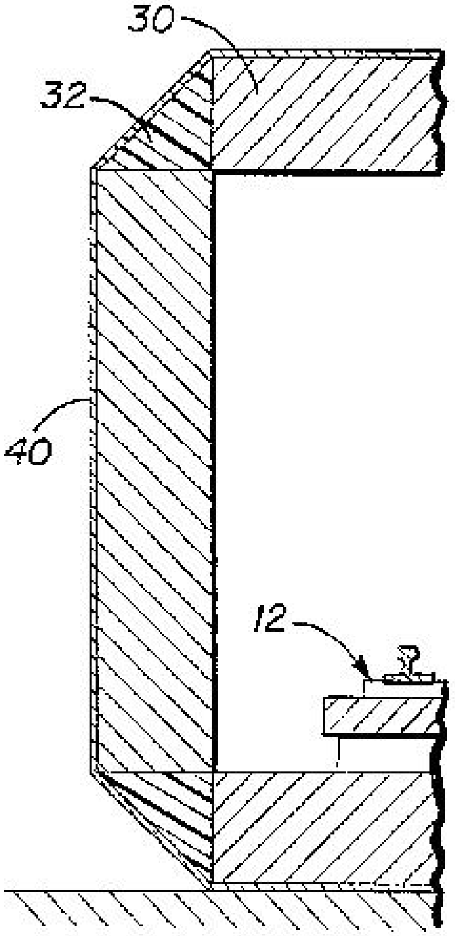

FIG. 4 is a top plan view of an installation in accordance with another embodiment of the invention.

FIG. 4

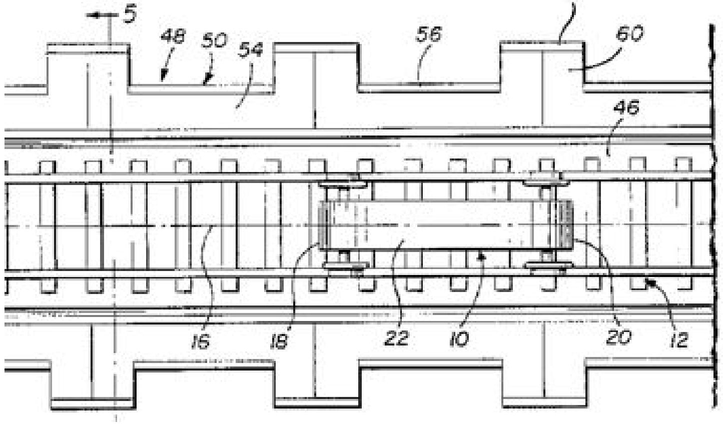
5
48
50
54
56
58
60
46
16
18
22
10
20
12

FIG. 5 is a sectional view taken substantially through a plane indicated by section line 5--5 in FIG. 4.

FIG. 5

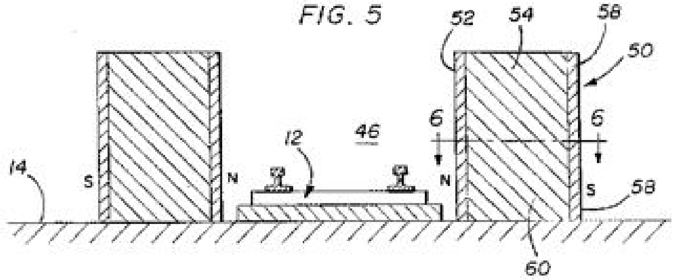
52
54
58
50
6
6
12
46
14
S
N
N
S
58
60

FIG. 6 is a sectional view taken substantially through a plane indicated by section line 5--5 in FIG. 5.

FIG. 6

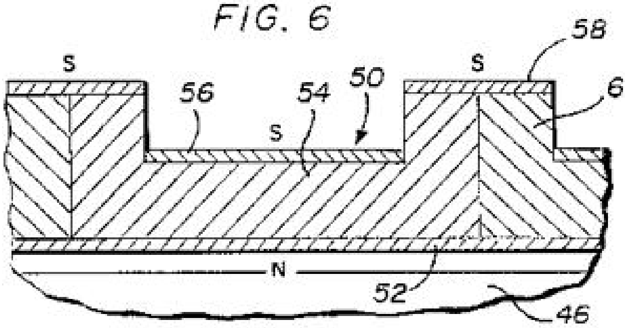
58
S
S
60
56
54
50
S
N
52
46

FIG. 7 is a simplified side view through the flux zone shown in FIGS. 4, 5 and 6 with the armature bar magnet positioned therein.

FIG. 7

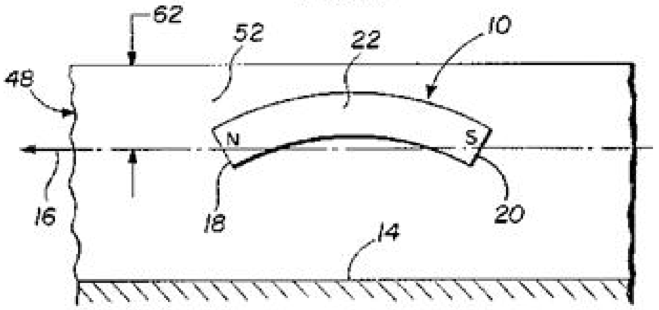
62
52
22
10
48
N
S
16
18
20
14

FIG. 8 is a top plan view of an installation in accordance with yet another embodiment.

FIG. 8

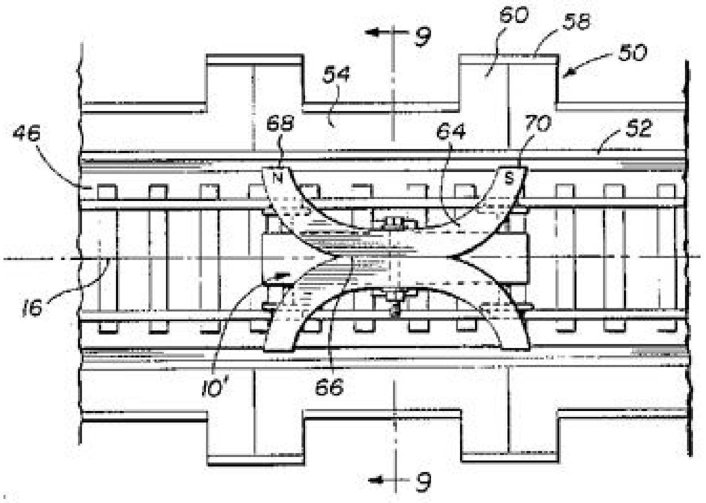

FIG. 9 is an enlarged partial sectional view through a plane indicated by section line 9--9 in FIG. 8.

FIG. 9

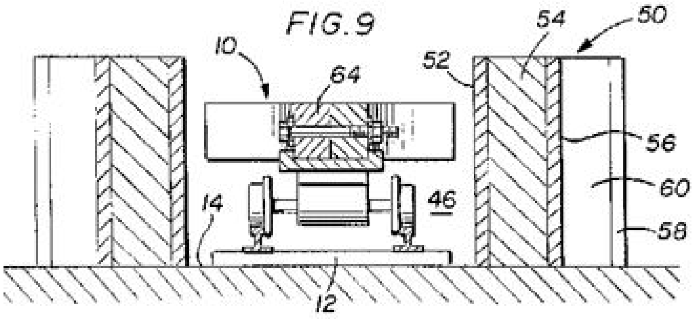

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Referring now to the drawings in detail, FIG. 1 illustrates one embodiment of the invention in which a magnetic armature generally referred to by reference numeral 10 is unidirectionally propelled along a predetermined path established by a motion guiding track 12 fixed to a frame or support 14. The path is represented by a line 16 extending through pole faces 18 and 20 of opposite polarity at the longitudinal ends of a curved armature bar magnet 22. The armature 10 in the illustrated example includes a wheeled vehicle mount 24 to which the armature magnet 22 is fixedly secured with the pole faces 18 and 20 converging toward the guiding track 12. The pole faces 18 and 20 are furthermore orientated so that the magnetic field extending between pole faces 18 and 20 is movable therewith within a flux zone 26 limited in surrounding relation to the guided path at spaced locations by stator gate assemblies 28 formed by permanent magnets fixed to the frame support 14.

Each of the stator gate assemblies 28 as shown in FIGS. 1-3, includes four bar magnets 30 interconnected at corners by non-magnetic elements 32, such as triangular wooden blocks as more clearly seen in FIG. 3, to form a rectangular enclosure in surrounding relation to the track 12. Pole faces 34 and 36 between which a stationary magentic field extends are formed on the bar magnets substantially aligned with parallel spaced planes in perpendicular intersecting relation to the path line 16. The pole face

34 of one polarity (north) is effective through its magnetic field to magnetically interact with the magnetic field of the armature magnet 22 causing unidirectional propulsion of the armature 10 as actually observed during tests. Such magnetic interaction is obviously influenced by the pole face 36 of opposite polarity (south) abutting and fixed to an annular or circular ring magnet 38 magnets 30. The interconnected and 38 may be held in assembled relation by an outer skin or sheathing 40 as shown in FIG. 3.

The ring magnet 38 has a radially inner pole surface 42 of the same polarity (north) as that of the pole faces 34 to interact with the other pole faces 36 as aforementioned, to the exclusion of the radially outer pole surface 44. The obvious effect of said arrangement is to exert a net magnetic force on the armature magnet 22 causing the observed continuous, unidirectional propulsion thereof through the gate assemblies 28. Such assemblies 28 are spaced apart distance dependent on the magnetic field intensity or strength of the permanent magnets 30 and 38 which dictate the effective axial extent of the aforementioned magnetic fields associated with the assemblies 28 and the armature magnet 22.

FIGS. 4-7 illustrate another embodiment of the invention utilizing the same type of movable armature 10 guided along a predetermined path by a frame mounted track 12 extending through a flux zone 46 established by another type of permanent magnet stator arrangement, generally referred to by reference numeral 48. The stator 48 includes a pair of permanent magnet assemblies 50 extending in parallel spaced relation to each other on opposite sides of the armature path established by the track 12. Each assembly 50 is a mirror image of the other so as to expose continuous confronting pole surfaces formed by a magnetic layer material 52 such as Neodynium, mounted on interconnected ceramic bodies 54. The confronting pole surfaces of the magnetic layers 52 are of like polarity (north), opposite to the polarity of the pole surface of magnetic layer sections 56 and 58 made of Samarium Cobalt, for example, and carried on the ceramic bodies 54. The bodies 54 have transversely extending flange portions 60 at the abutting ends so as to mount the layer sections sections 58 laterally outwardly of layer sections 56 as more clearly seen in FIGS. 4 and 6 to thereby vary the magnetic field intensity along the guided armature path within the limited flux zone 46 in which the magnetic fields of the stator assembly 48 interact with the magnetic field of bar magnet 22.

The curved armature magnet 22 is orientated within the flux zone 46 between the confronting pole surfaces on 52 as depicted in FIG. 7, with the pole faces 18 and 20 converging toward the track 12 as previously described in connection with FIGS. 1-3. However, it was found that maximum propelling thrust is produced by optimum location of the path line 16 through the pole faces 18 and 20 a distance 62 closer to the upper edge of surface layer 52 than the lower edge on the frame support 14.

FIGS. 8 and 9 illustrate yet another embodiment of the invention involving the same type of permanent magnet stator arrangement 50 as described with respect to FIGS. 4-7. a modified form of armature 10' is featured in FIGS. 8 and 9, including two curved armature magnets 64 that are mirror images of each other with respect to an intermediate abutting portion 66. The magnets 64 are interconnected at the abutting portion 66 in alignment with a plane containing the path line 16 centrally between the confronting pole surfaces on 52. The end pole faces 68 and 70 for each magnet 64, are aligned with a plane in parallel spaced relation between the path line 16 and the pole surface on 52. With the number of pole faces thereby doubled for the armature, a higher and more efficient propelling thrust may be achieved.

The foregoing is considered as illustrative only of the principles of the invention. Further, since numerous modifications and changes will readily occur to those skilled in the art, it is not desired to limit the invention to the exact construction and operation shown and described, and accordingly, all suitable modifications and equivalents may be restorted to, falling within the scope of the invention.

US Patent # 5,402,021

Magnetic Propulsion System

Howard R. Johnson

(March 28, 1995)

Abstract --- A magnetic propulsion system including a plurality of specifically arranged permanent magnets and a magnetic vehicle propelled thereby along a path defined by the permanent magnets. The magnetic vehicle which is to be propelled includes a rigidly attached armature comprising several curved magnets. The propulsion system further includes two parallel walls of permanent magnets arranged so as to define the lateral sides of a vehicle path. Preferably, the walls are identical to one another except that the polarities of the magnets which define one wall are opposite from the polarities of the corresponding magnets in the opposite wall. A first wall, for example, includes a series of generally rectangular magnets, each magnet arranged with a North-to-South axis pointing longitudinally down the wall in the intended direction of vehicle travel. Each of the rectangular magnets is separated from the next successive rectangular magnet by a thinner magnet, which thinner magnet is arranged with its North-to-South axis pointing laterally toward the opposite wall and therefore perpendicular with respect to the North-to-South axis of the rectangular magnets. The opposite (or second) wall includes the same general arrangement of magnets, except that the North-to-South axis for each of the generally rectangular magnets is in a direction opposite from the direction of vehicle travel and the North-to-South axis of the thinner magnets points away from the first wall. In addition, the propulsion system includes several spin accelerators.

Inventors: Johnson; Howard R. (1440 Harding Rd., Blacksburg, VA 24060)
Appl. No.: 064930 ~ Filed: May 24, 1993

Current U.S. Class: 310/12; 198/619; 310/152
Intern'l Class: B65G 035/06; H02K 041/00
Field of Search: 310/12,152,46 198/619,805

References Cited [Referenced By]

U.S. Patent Documents:
4,151,431 ~ Aug., 1979 ~ Johnson (Cl. 310/12).
4,215,330 ~ Jun., 1980 ~ Hartman (335/306).
4,877,983 ~ Oct., 1989 ~ Johnson (310/12).

Other References:
Advances in Permanent Magnetism, pp. 44-57 date unknown.
Scientific American, Jan. 1989, pp. 90-97.

Introduction to Magnetic Materials, pp. 129-135 date unknown.
Applications of Magnetism, pp. 42-47 date unknown.

Description

FIELD OF THE INVENTION

The present invention relates to a magnetic propulsion system including a plurality of specifically arranged permanent magnets and a magnetic vehicle propelled thereby along a path defined by the permanent magnets.

BACKGROUND OF THE INVENTION

The generation of unidirectional propelling forces by permanent magnets is already known and recognized in U.S. Pat. Nos. 4,151,431 and 4,877,983 to Johnson, and U.S. Pat. No. 4,215,330 to Hartmen, by way of example. According to applicant's first patent (U.S. Pat. No. 4,151,431), such forces are generated by magnetic interaction between a curved magnet bar of an armature guided for movement along a circular path and an arrangement of spaced stator magnets having pole faces of one polarity facing the armature on one side thereof parallel to the path of movement.

According to applicant's second patent (U.S. Pat. No. 4,877,983), the armature magnet is mounted on a vehicle and guided along a path through a magnetic flux zone limited on opposite sides of the path by an arrangement of magnetic pole surfaces of one polarity on stator magnets. According to one embodiment of the second patent, the flux zone is formed by spaced gate assemblies of magnets having exposed pole faces of one polarity in a plane perpendicular to the armature path from which a magnetic field extends to the opposite pole faces and a ring magnet fixed to such opposite pole faces of the other polarity, with a radially inner pole surface of the same polarity producing a magnetic field perpendicular to the first mentioned field to their opposite radially outer pole surfaces. Several other embodiments are illustrated including variations in the armature structure and in the stator structure; however, all of the embodiments teach use of an annular stator assembly.

SUMMARY OF THE INVENTION

It is therefore an object of the present invention to provide an improved magnetic propulsion system having a plurality of permanent magnets and a magnetic vehicle propelled thereby along a path defined by the permanent magnets, wherein the permanent magnets need not encircle the path of the magnetic vehicle.

In order to achieve this and other objects, the present invention comprises two parallel walls of permanent magnets arranged so as to define the lateral sides of a vehicle path. The walls are identical to one another except that the polarities of the magnets which define one wall are opposite from the polarities of the corresponding magnets in the opposite wall.

A first wall, for example, includes a series of generally rectangular magnets, each magnet arranged with a North-to-South axis pointing longitudinally down the wall in the intended direction of vehicle travel. Each of the rectangular magnets is separated from the next successive rectangular magnet by a thinner magnet, which thinner magnet is arranged with its North-to-South axis pointing laterally toward the opposite wall and therefore perpendicular with respect to the North-to-South axis of the rectangular magnets.

The pole-to-pole length of each thinner magnet is preferably no more than half the width of the generally rectangular magnets. Accordingly, a gap on the inside surface of the wall is defined by the presence of each thinner magnet.

The opposite (or second) wall includes the same general arrangement of magnets, except that the North-to-South axis for each of the generally rectangular magnets is in a direction opposite from the direction of vehicle travel and the North-to-South axis of the thinner magnets points away from the first wall.

In addition, the propulsion system of the present invention includes several spin accelerators for crowding the magnetic fields at predetermined positions along the length of the walls. This crowding of the magnetic fields serves to intensify the fields and causes the vehicle's armature to be accelerated faster than would otherwise be the case without the spin accelerators.

The spin accelerators project laterally outward from each of the walls at predetermined positions along the longitudinal length of each wall. Each spin accelerator comprises a generally rectangular permanent magnet which is preferably identical to that of the first and second walls. Each spin accelerator further includes a shorter magnet having a smaller pole-to-pole length than that of the generally rectangular magnet and a wedge separating the generally rectangular magnet of the spin accelerator from the shorter magnet.

The orientation of the generally rectangular magnet in the spin accelerator is determined by which pole of the wall's thinner magnet is facing outwardly. The rectangular magnet's orientation is such that face-to-face contact is established between opposite poles of the generally rectangular magnet in the spin accelerator and the thinner magnet in the wall. Accordingly, the North-to-South axis of the generally rectangular magnet in the spin accelerator points in the same direction as the North-to-South axis of the thinner magnet in the wall. The shorter magnet in the spin accelerator is likewise arranged with its North-to-South axis pointing in the same general direction as that of the thinner magnet in the wall; but here, an acute angular tilt away from the North-to-South axis of the thinner magnet is established by the wedge. In particular, the angle of the wedge determines the acute angle which exists between the North-to-South axis of the shorter magnet in the spin accelerator and the North-to-South axis of the thinner magnet in the wall.

The magnetic vehicle which is to be propelled by the instant propulsion system includes a rigidly attached armature comprising several curved magnets. Each curved magnet is arranged on the vehicle such that its North-to-South axis is parallel with respect to that of the other curved magnets. In particular, the North-to-South axes of all the curved magnets point in the same direction as the North-to-South axes of the thinner magnets in each wall. The vehicle itself is preferably a wheeled vehicle mounted on a track; however, it is understood that other vehicle structures will suffice so long as the vehicle is maintained between the walls of the propulsion system.

In operation, the magnetic fields created by the two walls exert a propelling force on the armature of the vehicle in the desired direction of travel. Since the armature of the vehicle is rigidly attached to the vehicle, the vehicle itself begins to accelerate and is hence set in motion by the propulsion system.

Preferably, the curved magnets of the vehicle armature are "Alnico 8" magnets tipped with neodymium magnets. The magnets which constitute the walls and spin accelerators are preferably made of neodymium and ceramic material, except for the thinner magnets. The thinner magnets are preferably made of rubber or plastic, and each can comprise a plurality of magnetic rubber or plastic layers.

Although the present invention has been described with regard to generally rectangular magnets, it is understood that other permanent magnet shapes will suffice, including but not limited to generally cylindrical shapes.

BRIEF DESCRIPTION OF THE DRAWING

FIG. 1 is a schematic plan view of a magnetic propulsion system in accordance with a preferred embodiment of the present invention.

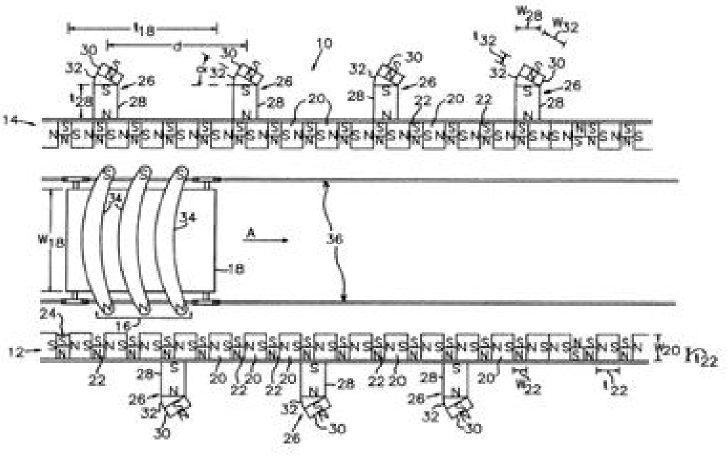

FIG. 1

DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT

With reference to FIG. 1, a preferred embodiment of the inventive magnetic propulsion system and vehicle propelled thereby will now be described.

FIG. 1 schematically illustrates a propulsion system 10 comprising two parallel magnetic walls 12,14 which are stationary, and an armature 16 rigidly attached to a vehicle 18. The two parallel walls 12,14 are formed from several permanent magnets arranged so as to define the lateral sides of a vehicle path. The desired direction of vehicle travel is indicated by an arrow A in FIG. 1. The two walls 12,14 are identical to one another except that the polarities of the magnets which define one wall 12 are opposite from the polarities of the corresponding magnets in the opposite wall 14. A first wall 12, for example, includes a series of generally rectangular magnets 20, each magnet arranged with a North-to-South axis pointing longitudinally down the wall in the intended direction of vehicle travel (indicated by arrow A). Each of the magnets 20 preferably comprises a ceramic magnet with a neodymium north pole. In addition, each of the generally rectangular magnets 20 is separated from the next successive rectangular magnet 20 by a thinner magnet 22. The thinner magnets 22 are arranged with their North-to-South axes pointing laterally toward the opposite wall 14 and therefore perpendicular with respect to the North-to-South axis of the rectangular magnets 20. Each thinner magnet 22 is preferably made from rubber or plastic permanently magnetic material. Also, the pole-to-pole length of each thinner magnet 22 is preferably no more than half the width of the generally rectangular magnets 20.

Consequently, a gap 24 on the inside surface of the wall 12 is defined by the presence of each thinner magnet 22.

The opposite (or second) wall 14 includes the same general arrangement of magnets 20,22, except that the North-to-South axis for each of the generally rectangular magnets 20 points in a direction opposite from the direction of vehicle travel, while the North-to-South axes of the thinner magnets 22 point away from the first wall 12.

By arranging the thinner magnets 22 between the generally rectangular magnets 20 in the foregoing manner, there is a pole shading effect on the magnets 20 of the walls 12,14.

In addition, the propulsion system 10 of the preferred embodiment includes several spin accelerators 26 for crowding the magnetic fields at predetermined positions along the length of the walls 12,14. This crowding of the magnetic fields serves to intensify the fields and causes the vehicle's armature to be accelerated faster than would otherwise be the case without the spin accelerators.

The spin accelerators 26 project laterally outward from each of the walls 12,14 at predetermined positions along the longitudinal length of each wall 12,14. According to the preferred embodiment, the spin accelerators 26 are positioned along the walls 12,14 at every other thinner magnet 22 (as is shown in the middle of FIG. 1). Each spin accelerator 26 comprises a generally rectangular permanent magnet 28 which is preferably identical or very similar to that of the first and second walls 12,14. Each spin accelerator 26 further includes a shorter magnet 30 having a smaller pole-to-pole length than that of the generally rectangular magnet 28 and a wedge 32 separating the generally rectangular magnet 28 of the spin accelerator 26 from the shorter magnet 30. The orientation of the generally rectangular magnet 28 in the spin accelerator 26 is determined by which pole of the wall's thinner magnet 22 is facing outwardly. The rectangular magnet's orientation is such that face-to-face contact is established between opposite poles of the generally rectangular magnet 28 in the spin accelerator 26 and the thinner magnet 22 in the wall 12,14. Accordingly, the North-to-South axis of the generally rectangular magnet 28 in the spin accelerator 26 points in the same direction as the North-to-South axis of the thinner magnet 22 in the wall 12,14. The shorter magnet 30 in the spin accelerator 26 is likewise arranged with its North-to-South axis pointing in the same general direction as that of the thinner magnet 22 in the wall 12,14; but here, an acute angular tilt away from the North-to-South axis of the thinner magnet 22 is established by the wedge 32. In particular, the angle .alpha. of the wedge determines the acute angle which exists between the North-to-South axis of the shorter magnet 30 and the North-to-South axis of the thinner magnet 22 in the wall 12,14. The shorter magnet 30 preferably consists of neodymium.

The magnetic vehicle 18 which is to be propelled by the instant propulsion system 10 includes a rigidly attached armature 16 comprising several curved magnets 34. Each curved magnet 34 is arranged on the vehicle 18 such that its North-to-South axis is parallel with respect to that of the other curved magnets 34. In particular, the North-to-South axes of all the curved magnets 34 point in the same direction as the North-to-South axes of the thinner magnets 22 in each wall 12,14. The vehicle 18 itself, according to the preferred embodiment, is a wheeled vehicle mounted on a track 36. It is understood, however, that other vehicle structures will suffice so long as the vehicle is maintained between the walls 12,14 of the propulsion system 10.

In operation, when the vehicle 18 is positioned as is shown in FIG. 1, the magnetic fields created by the two walls 12,14 exert a propelling force on the armature 16 of the vehicle 18 in the desired direction of travel (arrow A). Since the armature 16 is rigidly attached to the vehicle 18, the vehicle 18 itself begins to accelerate and hence is set in motion by the propulsion system 10.

Furthermore, since the spin accelerators 26 serve to crowd and thereby intensify the magnetic fields at predetermined positions along the walls 12,14, the acceleration of the vehicle is enhanced as the vehicle passes these predetermined positions.

The spin accelerators 26 can be reversed in order to lessen their effectiveness at crowding the magnetic fields. Reversing of the spin accelerators 26 can be accomplished by rotating the spin accelerators 26 so that the shorter magnets 30 tilt away from the intended direction of vehicle travel, rather than in the direction of travel as is the case for the illustrated embodiment.

Preferably, the curved magnets 34 of the vehicle armature 16 are "Alnico 8" magnets tipped with neodymium magnets, while the wedges 32 comprise wood or similar material and an angle .alpha. of 45 to 90 degrees.

The width w.sub.20, height, and pole-to-pole length 1.sub.20 of the generally rectangular magnets 20 in each wall 12,13 are 0.75 inches to 1.25 inches, 3.75 to 4.25 inches, and 1.25 inches to 1.75 inches, respectively. The width w.sub.22, height, and pole-to-pole length 1.sub.22 of the thinner magnets 22 in the walls 12,14 are 1 inch to 1.5 inches, 3.75 inches to 4.25 inches, and no more than one half the width w.sub.20 of the generally rectangular magnets, respectively. In the spin accelerators 26, the width w.sub.28, height, and pole-to-pole length 1.sub.28 of the generally rectangular magnets 28 are 1.125 to 1.625 inches, 3.75 to 4.25 inches, and 0.875 inches to 1.375 inches, respectively, while the width w.sub.32, height, and pole-to-pole length 1.sub.32 of the shorter magnets 30 are 0.75 inch to 1.25 inches, 3.75 inches to 4.25 inches, and 0.125 inch to 0.375 inch, respectively.

Preferably, the distance separating the walls 12,14, is such that each wall 12,14 is 0.5 inch to 1.25 inches away from the tips of the armature magnets 34, both walls 12,14 being equidistant from the tips of the armature 16. Also, the curved magnets 34 of the armature 16 are preferably 0.375 inch to 0.625 inch apart from one another.

Testing of the foregoing prototype propulsion system resulted in the vehicle moving 2 feet in one second.

Although the present invention has been described with reference to a preferred embodiment, it is understood that various modifications to this embodiment will become subsequently apparent to those having ordinary skill in the art. In this regard, the scope of the invention is limited only by the claims appended hereto, and not by the illustrated embodiment.

Tom Bearden Comments
(Cribbed from www.greaterthings.com)

From: "Karl" <krlbrgmnn@mailandnews.com>
To: "Sterling D. Allan, PerenTech.com" <sterlingda@perentech.com>
Sent: Thursday, December 19, 2002 2:01 PM

Subject: Truth about Howard Johnson and his motor

Sterling,

Just a quick note to let you know I spoke to Tom Bearden today and, regarding the statement on your webpage by your "anonymous source" who said "Howard Johnson never was able to get the rotary version to work. He was only able to get the linear version working, and there were some questions about its viability.", Tom said such a person doesn't know what they are talking about. I'd say Tom is in the best position to know, since he and Howard have been friends for decades and Howard has personally brought a functional, rotary PMM to Tom's house, and they've played with it for hours. Next to Howard himself, I'd consider Tom to be the best resource on the subject of HJ's PMM.

He also said that Howard is still working on getting another functioning unit constructed (he's had numerous setbacks over the years after his working unit was vandalized by thieves that broke into Howard's shop and stole only the magnets off that model, leaving many $K worth of other material nearby untouched). Howard is still plugging away every day, at over 70 years old, but his wife's health problems are keeping him busy. However, he has gotten hooked up with a respectable, honest businessman who is financing him and Howard also now has some younger help to do the grunt work that is becoming tougher for him. All in all, Tom hopes to see some real progress being made for Howard in the next year or so.

Tom mentioned that there are a few critical things that anyone who wants to successfully build a working HJ PMM needs to know, some of which are probably obvious to the more experienced builders:

1) The most critical element is the precise machining of the magnets. Joe Q. Public, with his diamond saw, cutting his own magnets by hand has little chance of succeeding in shaping the magnets to the aerospace-critical specs that are the minimum necessary, much less duplicating that feat several times for each necessary part.

2) Alignment of the parts is highly important too. One slight misalignment and the motor will not run continuously.

3) All magnets aren't the same. This application requires expensive, extremely high quality magnets. Howard, in order to get them cheaper, was buying them in $50K lots from China, where high quality and a lower price can be had.

I also spoke with a gentleman named Gary Hanson who had been in touch with Howard years ago when Gary was trying to build Howard's motor. Howard told him, as I think most researchers on this know now, that the motor *can* be build directly from the patent but that there needs to be 5 or 6 armtures and not just the one that is shown as an illustration in the patent. Just want to make sure any newbies understand this.

I hope some of this has helped clear up any misunderstandings for those of us interested in building a functional HJ PMM. My opinion is that, if you want to do it right, go to the horse's mouth (Howard) or else as close as you can possibly get (Tom). Hopefully, doing so will enable us to make this motor a reality.

Regards, Karl

Tom Bearden: ***Chasing the Wild Dragon*** **[**Excerpts from www.cheniere.org]

Howard Johnson is also a respected colleague, whom I very much admire. (See Figure 23). Howard has continued to work quietly and patiently upon his patented permanent magnet motor, [note 31] including patenting various magnetic gates, etc. that are necessary to make such a motor work in a rotary configuration. [note 32] Howard employs a two-particle theory of magnetism; i.e., each magnetic flux line is envisioned as having a particle traveling from the north pole to the south pole, and also a particle traveling from the south pole to the north pole. The particles are spinning; the forward-time particle spins in one direction, and the antiparticle spins in the other direction. Howard then slightly separates the two particle flows. [note 33] In other words, Johnson splits the flux lines themselves, into two different pieces. When so separated, the component lines are now curls, so their paths curve. The paths of the two "curl particles" are different; one curls in one direction and the other curls in the opposite direction. Further, a predominance of one form of curl particle gives a "time-forward" aspect, while a predominance of the other form of curl particle gives a "time-reversed" aspect. Johnson is thus able to employ a deeper kind of magnetism than the textbooks presently contain. He demonstrates that a "spin-altered" magnetic assembly exhibiting (to a compass or other such detector) a north polarity can attract another unaltered magnetic assembly exhibiting a north polarity. In short, he can make a north pole attract a north pole. We will give you further insight into Johnson's two-particle theory in a future article. [note 34] We will also explain how and why the physicists missed that antiparticle in the magnetic field's flux lines, and thereby failed to advance the theory of magnetism to a deeper level. Make no mistake, one day when the new theory is done, Johnson may well be awarded a Nobel Prize for his epochal discovery of a deeper structure of magnetism.

Footnotes:

[32] I personally saw and closely examined one demonstration rotary Johnson permanent magnet motor some years ago, and toyed with it for about one hour. It would definitely self- rotate as long as you wished to permit it to turn. It was not a powerful device at all, but just a small laboratory "proof of principle" prototype. It had cost Johnson an enormous amount of time, labor, and optimization to get the critical adjustment required of his two magnet assemblies. But the device had no power source other than the permanent magnet assemblies themselves. Johnson's nonlinear rotor and stator magnets interacted with each other in a manner to break local symmetry. So his machine was an open system and therefore a permissible overunity device; it was not a perpetuum mobile.

[33] As I have pointed out repeatedly in the past, photons also carry time, not just energy. We have previously shown the process and the photon interaction mechanism that creates the flow of time itself; we will discuss this mechanism in the future. So when Johnson separates the particles and antiparticles, not only does he partially separate them according to spin, but he also alters the local character of time flow during which the resulting magnetic field forces must appear. In other words, he accomplishes a partial separation of time-forward and time- reversed polar interactions. A south pole is just a time- reversed magnetic north pole, in the first place! So a north pole of a bar magnet that is slightly time reversed on one side will

partially act on that side just like a south pole. On the other side it will continue to act like a normal north pole. By partially time-reversing (phase conjugating) one side of the north magnetic pole piece, Johnson makes that side look and act like a south pole. that way Howard is able to create two north poles, one on a stator and the other on a rotor, and time-reverse part of one face of the stator's north magnetic pole-piece. Therefore when the proper sides of the stator and rotor north poles are facing, they attract each other, contrary to the conventional textbook. The two poles then repel each other normally as soon as the north rotor poles passes the north stator pole. Hence Johnson can make a surrounding north pole stator assembly "draw in" an approaching north pole rotor assembly, and then kick it on out the other side, because he has broken the local magnetic symmetry. In short, Johnson's magnetic gate can provide a legitimate component of unidirectional magnetic thrust, which means that he can indeed make a rotary permanent motor. Simply put, this "partially separating the spin particles," and thereby partially phase conjugating one face of a magnet, is what Johnson calls a "gate," and this is the patented secret by which his magnet assemblies can be made self-powering. The entire process is still very meticulous, and assembly and adjustments are extremely critical. With Johnson's blessings we hope to shed more light on this subject in coming articles.

[34] For a basic article on rotary permanent magnet motors, see T.E. Bearden, "On Rotary Permanent Magnet Motors and 'Free' Energy," *Raum & Zeit*, 1(3): 43-53 (Aug.-Sep. 1989).

Your Support Maintains this Service. Please Invest in the mind-boggling Rex Research Website CD -- Always The Best Investment ... Everything on this Website & More (the Bonus Files CD) -- Only $13, Postpaid Anywhere ! Secure Transaction via PayPal: All Major Credit Cards Accepted:

Buy Now

No Credit ? Don't like PayPal ? Send a Check/Money Order to:

Rex Research, PO Box 19250, Jean,
NV 89019 USA

rexresearch.com

From
Cheniere Press:

The secret world of Magnets

By the father of spintronics,
Howard Johnson

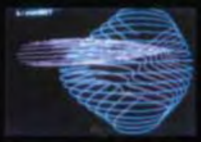

DEDICATION

This book is lovingly dedicated in memory of Dr. Gerhard H. Beyer, Distinguished University Professor at Virginia Polytechnic Institute and State University, Blacksburg, Virginia. He was former Head of Department of Chemical Engineering; Fellow in the Chemical Engineering Society and Active in the Nuclear Engineering Society and the Society of Professional Engineers. He enjoyed teaching and advanced research, including the discoveries and development of this book.

Howard Johnson first became interested in magnetics while doing some graduate work at Vanderbilt University. Several patents later he was joined by Jerry Beyer, a senior scientist in Chemical Engineering at V.P.I., and Steve Davis, an electrical engineer and leading computer man. Together they broke some of the magnetic code which they present here just as they found it.

"There is a God in heaven that revealeth secrets, and maketh known to the king Nebuchadnezzar what shall be in the latter days."

Daniel 2:1-28

"Hast thou entered into the treasures of the snow?". . .

"Which I have reserved against the time of trouble, against the day of battle and war? "

Job 38:22,23

CONTENTS

Page
IV. Description of book (IV), Howard Johnson
1. History of knowledge of magnets, Dr. Beyer
Description of magnet, conventional vs. discovery;
3. i.e., simple illustration of direction of lines of force
Computer illustration, B&W=> single plane only
4,5. Explanation of 2 Partical Theory;
Application; 92 pole generator
Application: Wire conducting D.C. current
Simple illustration(s)
Computer B&W graphics
Computer color illustration(s)
6,16. Spins are vortices
Simple illustration(s)
Computer graphics B&W illustration(s)
Colorado (alongside)
Layers (inside-outside)
17,20. Attraction and repulsion (within a magnet, and between map
Computer graphics showing curved magnet
Computer color illustration(s)
21,22. Corner spins
Simple illustrations and pictures
Computer color illustration(s)

23,27 Gate
Simple illustrations and pictures
Computer color illustrations
28. Making Use of the Time-Asymmetric Qualities of Permanent Magnets
29. Mapping methods, Howard R. Johnson
Bibliography
Introducing MAGWIND

WHO DISCOVERED MAGNETS?

BY DR. GERHARD H. BEYER

We'll never really know — it happened such a long time ago . ..

Maybe someone picked up a piece of "magnesian" rock on an Aegean coast and noticed the piece of lodestone was peculiar. It attracted a piece of iron, and could change the properties of the iron when the iron was rubbed with the rock.

Thales — who lived in Greece about 600 B.C. — studied attractive forces associated with magnets, and a resin called "amber." That started the long history of magnetism and electricity that is still being added to today.

It may have been that some Chinese used magnetic stones which pointed northward to find their way through the Gobi Desert many centuries ago.

The use of a magnetized needle floated on a cork, that has developed into the compass we know today, was a great boon to explorers and markedly changed our world.

More recently, the discoveries of new materials — such as ferrites and rare earth magnets — are likely to change our world again.

Have you ever wondered about:

How magnets work?
Why some elements are magnetic and others aren't?
How a magnet manages to change things without touching them?

This book may suggest at least partial answers to some of these questions. But most likely there will still be more questions than answers, for there are many things still to be discovered about magnets.

More work needs to be done. Maybe YOU will do it if you get inerested in magnets. That's one of the reasons for this book.

Way back in 1734, a Swedish scientist named Swedenborg showed the difference between magnetized iron and unmagnetized iron. And since then, we've discovered a lot of new materials and new techniques. Today there are better sensors for making measurements, and there are computers to help in recording, analyzing, and displaying them.

Another reason for this book is to tell you about these new materials and techniques and to show you some magnetic patterns no one else has ever seen.

The Direction of Magnetic Fields

For generations, physics students and others have been taught about magnets with iron filings.

It has been the popular belief of almost anyone with a common knowledge of magnetics that the pattern made by the filings represents the form and the movement of the magnetic fields. The following is an illustration of the popular view, using a simple bar magnet:

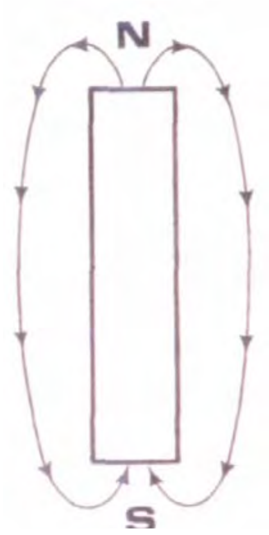

The over simplification of magnetic field, showing its movement from the north pole of the magnet to the south pole.

Today, however, it is quite evident that filings do not show magnetic fields as they are, but that they show what little pieces of magnets do in magnetic fields. The two are about as much alike as a Venetian blind and a blind Venetian.

The pieces of iron become little magnets that attract to each other, and are not free moving particles in the magnetic field, and cannot act as a dye to show where the fields are and what they look like.

These lines of force, that is, the magnetic fields, are much more complex than most minds would ever conceive. The concept that is about to be introduced here has been verified through much research, and will be demonstrated by experiments throughout the book.

This is what the direction of the lines of force really looks like, demonstrated with a cubical magnet having the top face for the north pole and the bottom face for the south pole:

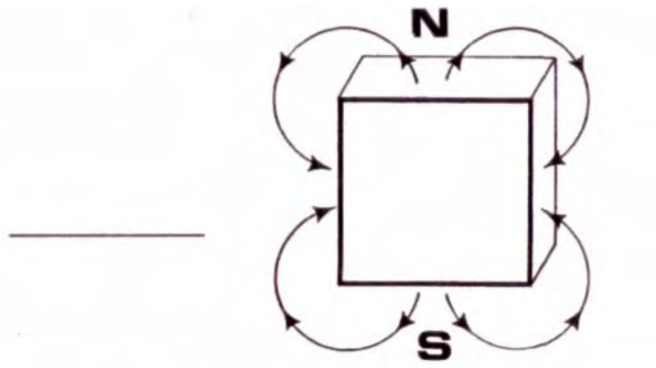

As opposed to this:

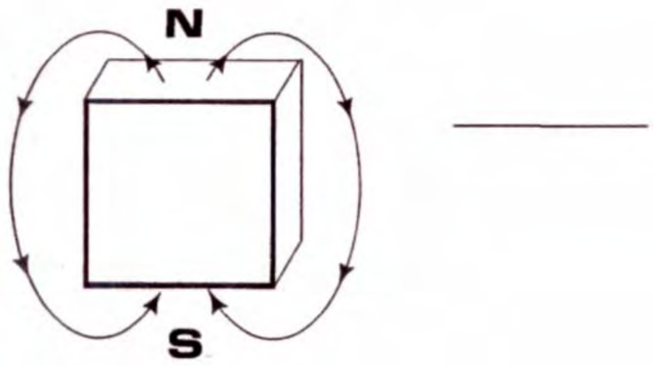

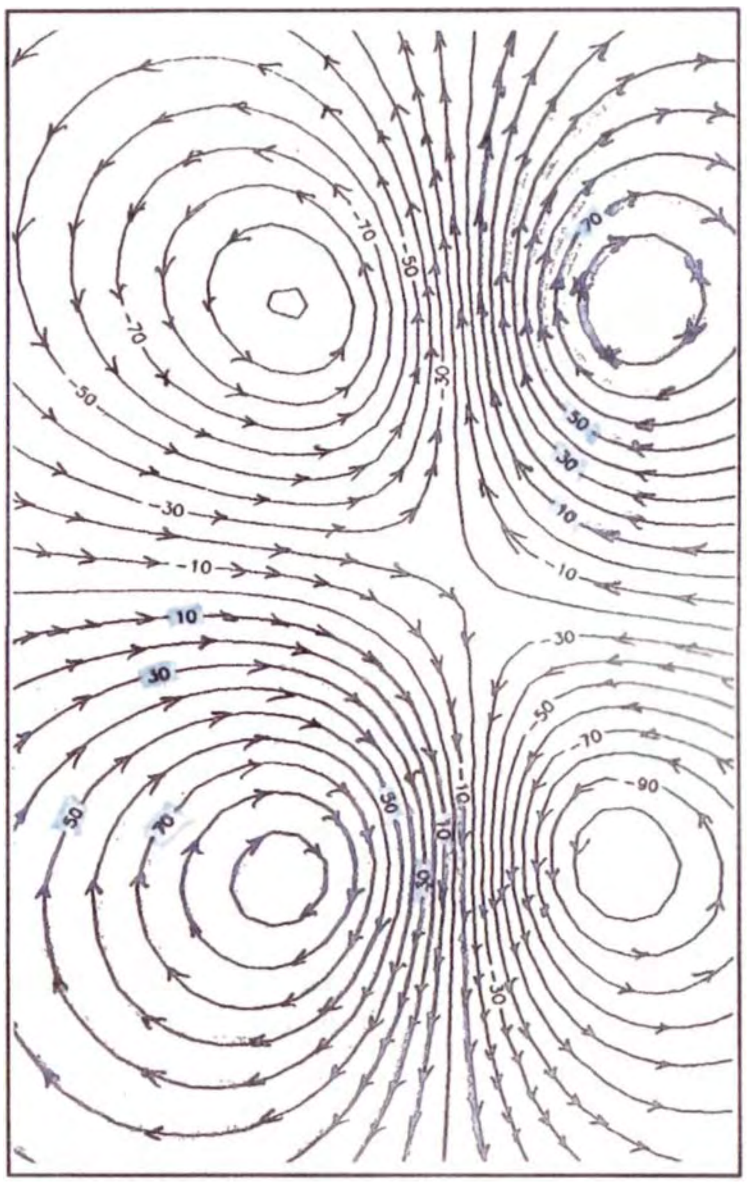

This actual graphic mapping of a magnet shows its lines of force by measuring the intensity of the magnetic field every 1/16" at each point on a grid, covering the entire magnet, as well as some of the field in the area around the magnet. (Sec page # 29 for description of method.) This measurement of the strength of the magnetic field is rated in gauss.

Upon careful examination of the illustration on page 3, you will notice that the lines of force leaving either pole are going in opposite directions.

For this to be possible, you must have two completely different lines of force which distinguish the north pole from the south pole, the difference being the direction of the lines of force. This brings us to the theory in which this work is based:

The lines of force of which a magnetic field consists are the track of a particle.

But, reason tells us, that if the illustration be true, and the lines of force are the track of a particle, then since there arc two lines of force, then there must be two different particles.

The knowledge of the existence of two particles came about by the design of a generator. As a result of DC. current being sent in one direction through a magnetizer around the rotor to be magnetized, alternating north/south poles are laid down. In illustrations:

The magnetizer, designed by George Eyerly, is assembled.

MAGNETIZING
A 92 POLE PERMANENT
MAGNET GENERATOR ROTOR

92 alternating north/south poles appear on the rotor. It is now ready to generate.

The preceding process uses the two particle principle, laying down lines going in opposite directions around a current carrying wire.

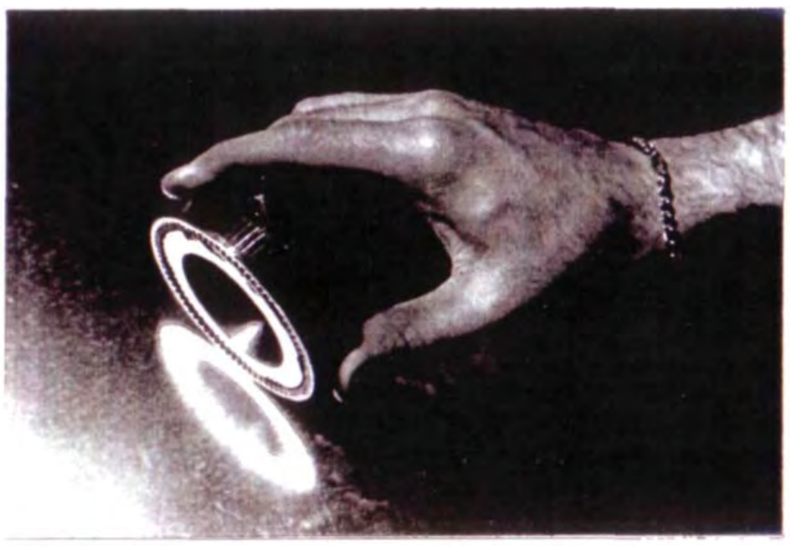

The rotor is put inside and a single wrap of conductor is placed around it.

A magnetizing current is released through the copper and . . .

This is made possible in keeping with the principle that, around the wire conducting current, these two opposing particles orbit in opposite directions. Illustration:

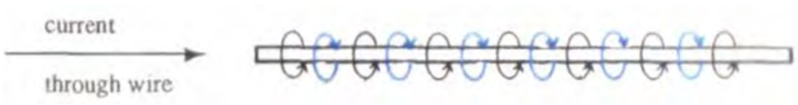
current
through wire

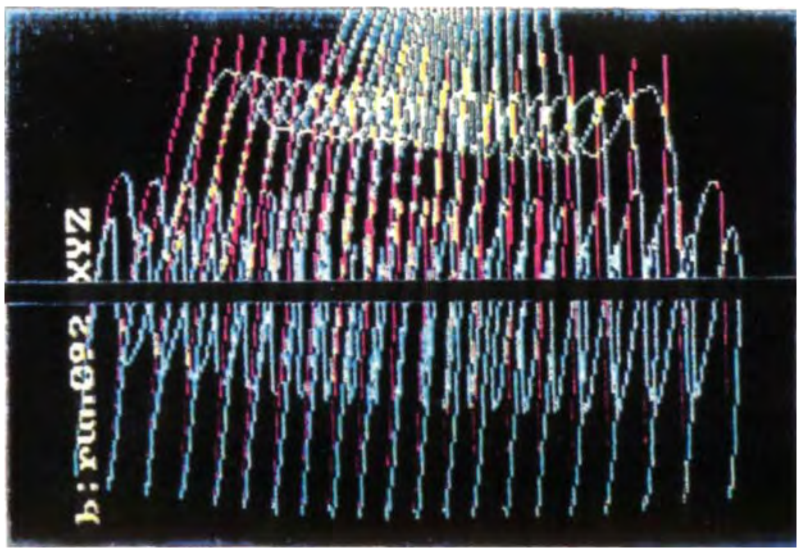

Computer color illustration.

In the permanent magnet, we have the same two spins in opposite directions. We do not know what makes them behave that way, but we do believe the record of our excellent monitoring and recording equipment.

Lines of Force are Spins forming Vortices

One of the most most amazingly illustrative and thoroughly innovative concepts in the area of magnetic field structure has been the discovery of vortices caused by the path of the particles which make up the lines of force. Notice the previously used illustration:

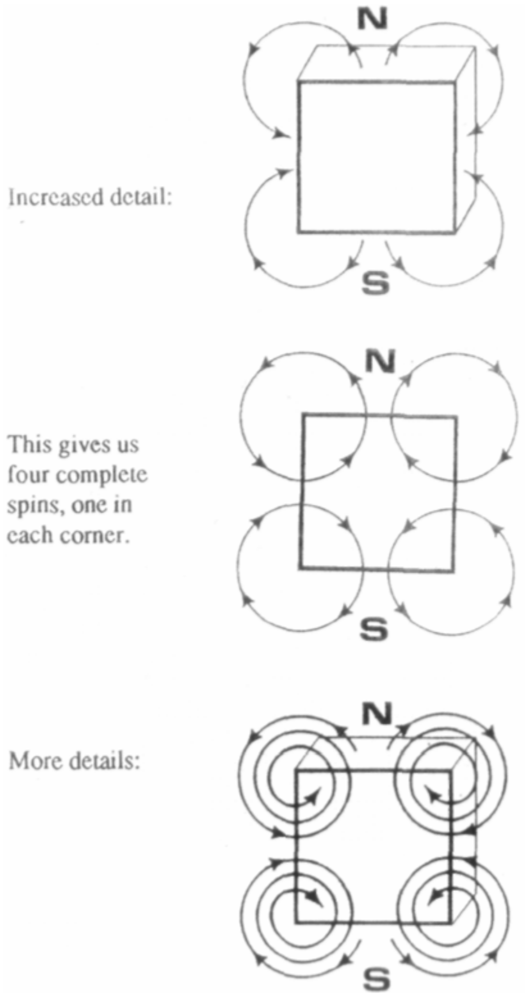
N
Increased detail:
S
N
This gives us four complete spins, one in each corner.
S
N
More details:
S

THE DOUBLE VORTEX WITH THE SPINS ALONGSIDE

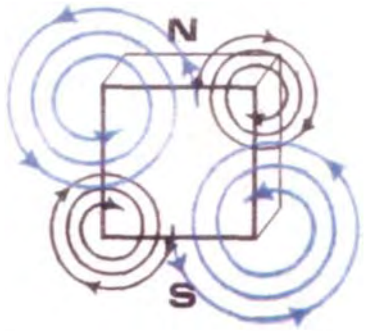

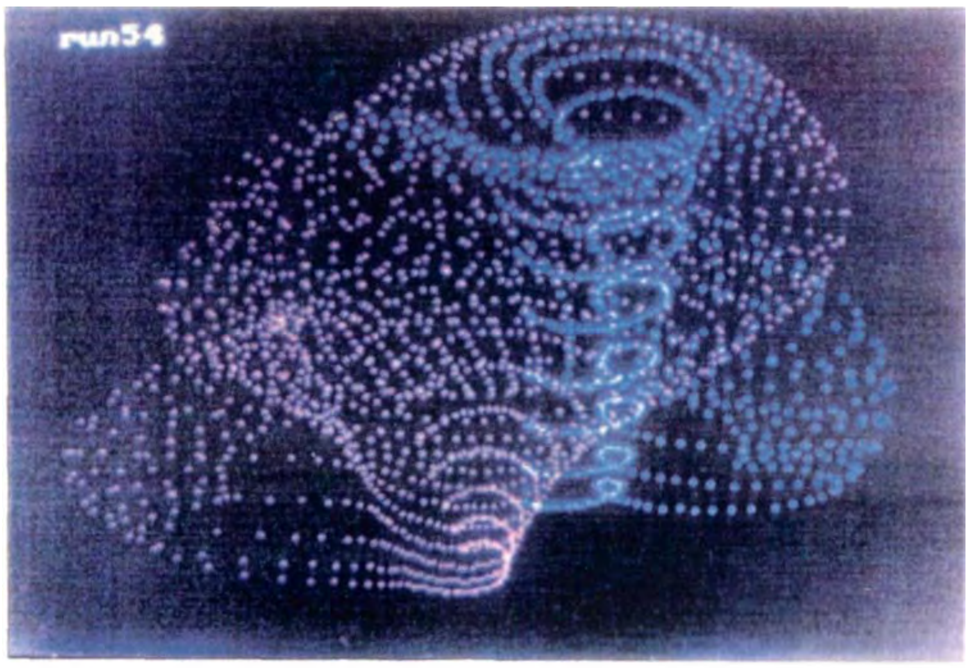

Here is the picture which recorded the discovery of the Double-Vortex

Noticing the last illustration, it is evident that the "whirlwind" or "tornado" effect is present and that there are two vortices present at each "pole".

An interesting and important piece of information, though, is that these vortices are not all the same, as is shown in previous illustration for clarity. Notice the distribution of the spins:

The Double-Vortex is highly significant in many ways, but the point to be reckoned with here is that both particles exist at both poles. Therefore, there is an element of both the "north" and the "south" in each pole. The north element (vortex) is dominant, and has proven to be the stronger vortex with higher gauss ratings.

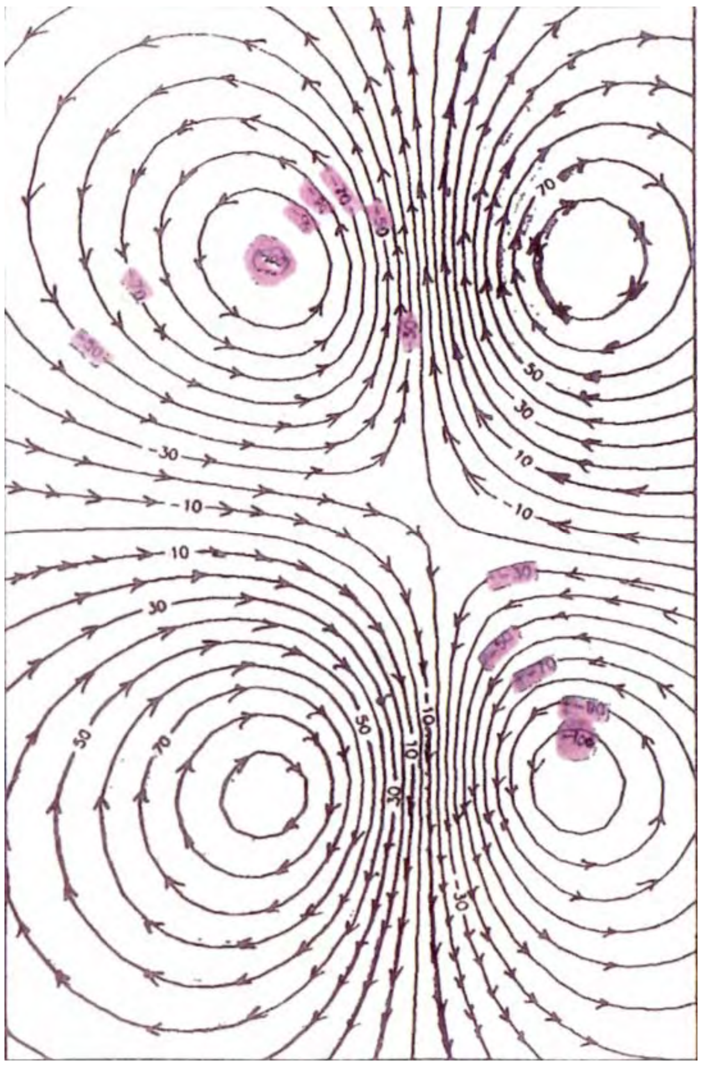
70
50
30
-30
-10
10
30
50
70
10
30
-30
-50
-70
-90
-100

Since the stronger north element (vortex) exists in both poles, you are sure to ask what the deciding factor is that distinguishes the north pole from the south pole. The same illustration just used shows that the north pole is the one with the weakest south element (vortex). This means the other pole must be south.

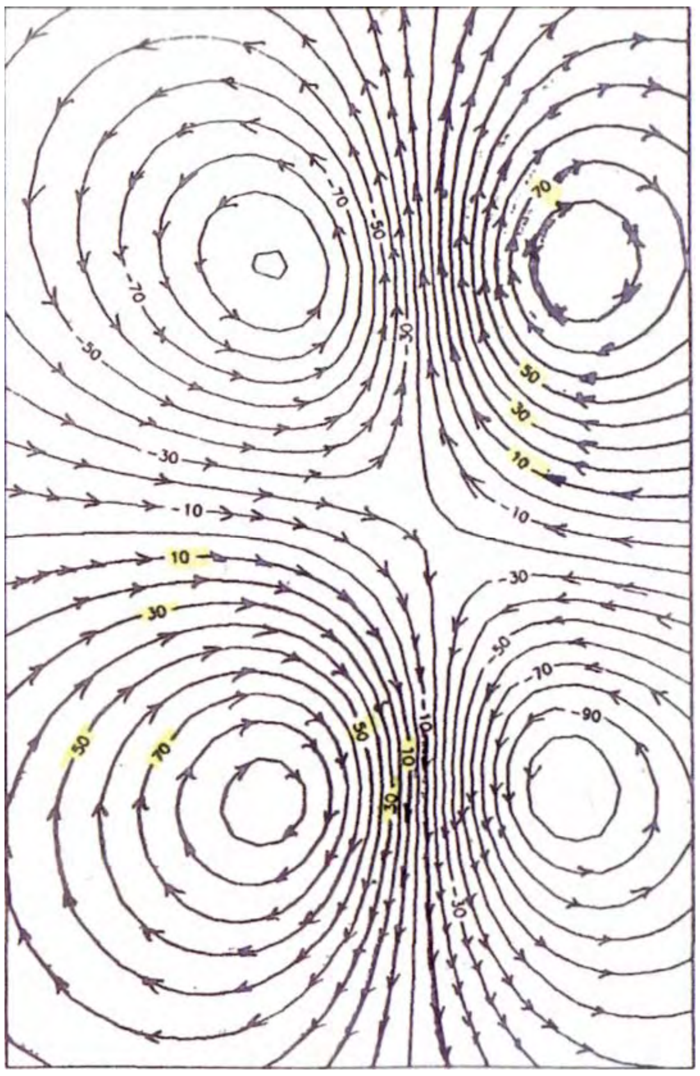
-70
-50
-30
-10
70
50
30
10
-30
-50
-70
-90
50
70
10
30

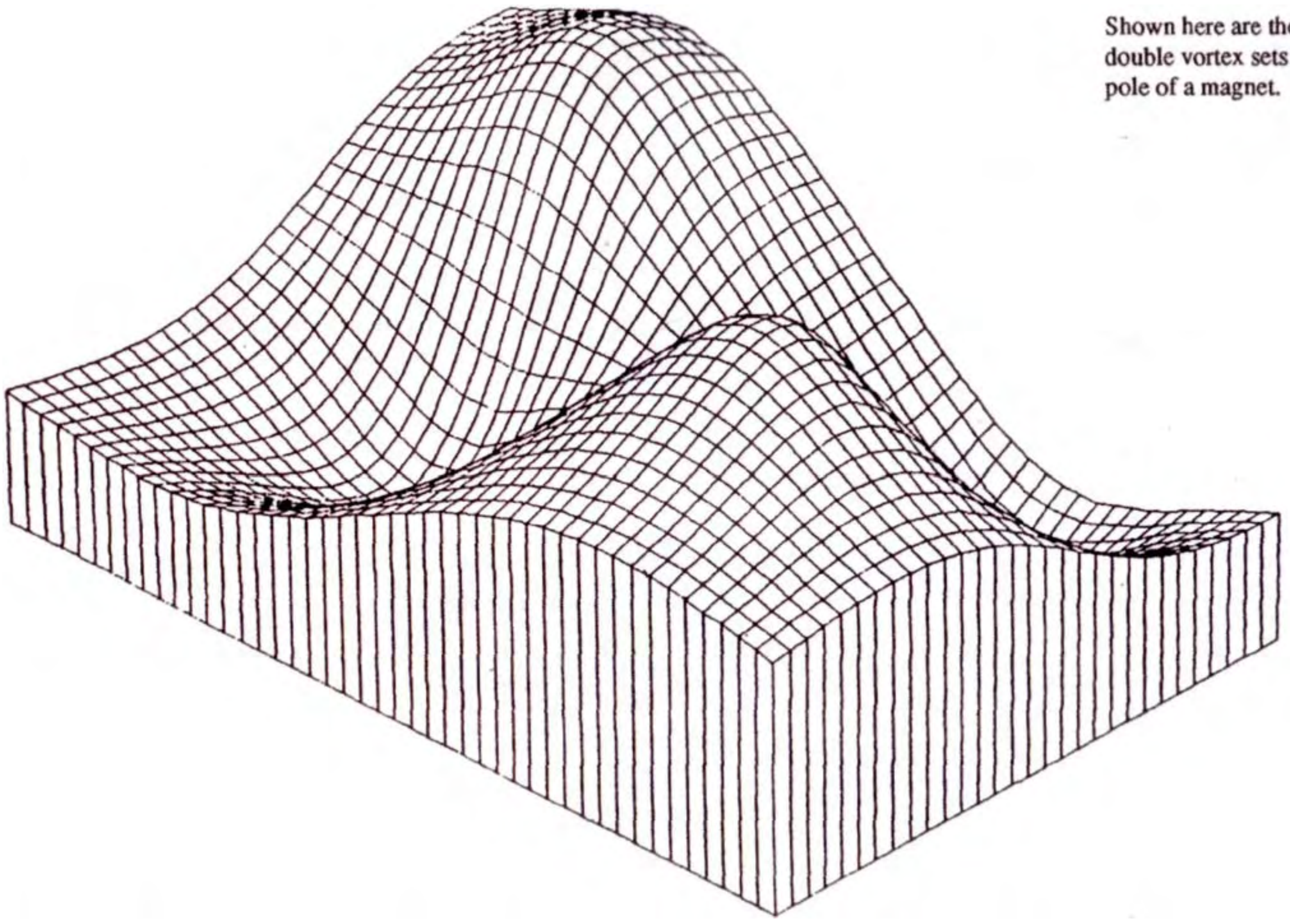

Shown here are the double vortex sets at each pole of a magnet.

This is a topographical map of the fields at the end of a square ceramic bar magnet magnitized through its thickness.

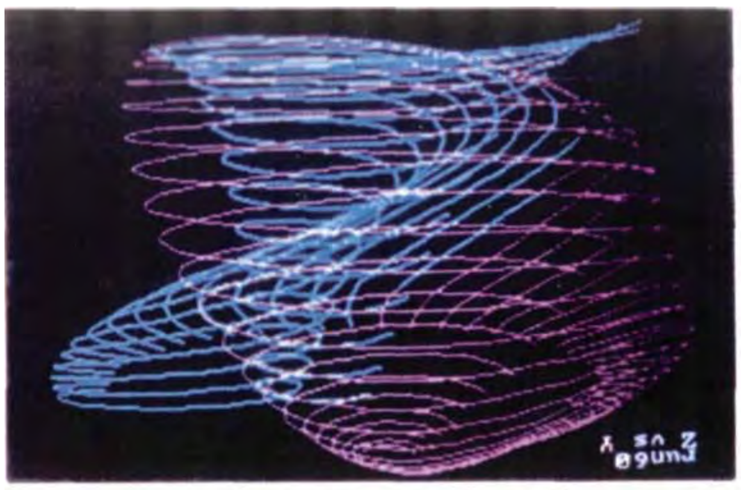

The
Double
Vortex:

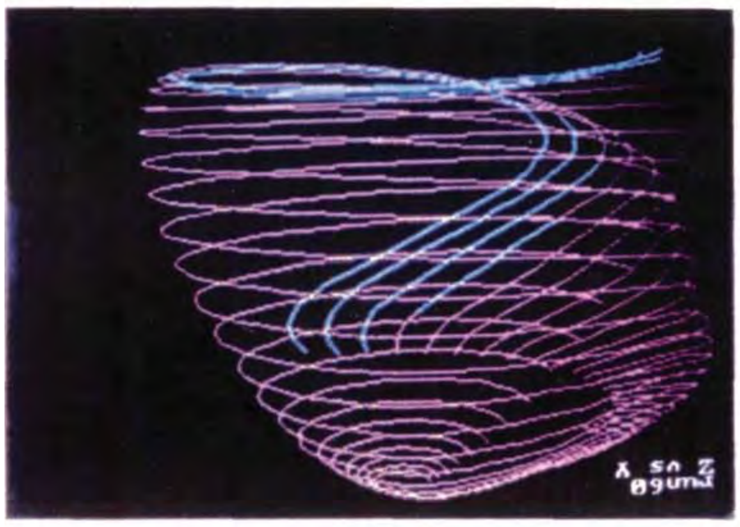

This vortex
is the
north element.

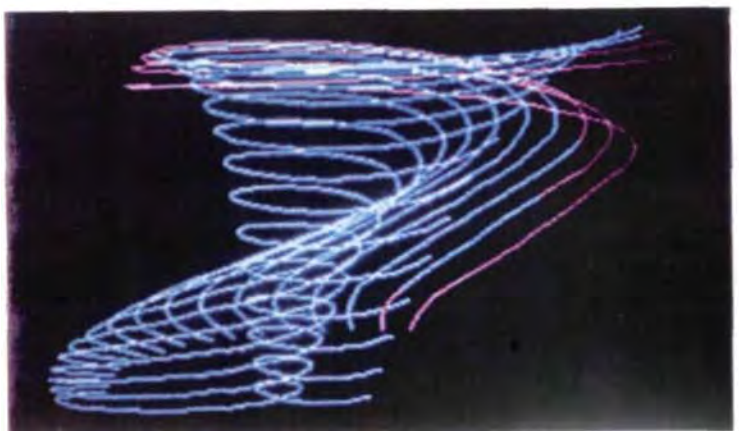

This vortex
is the
south element.

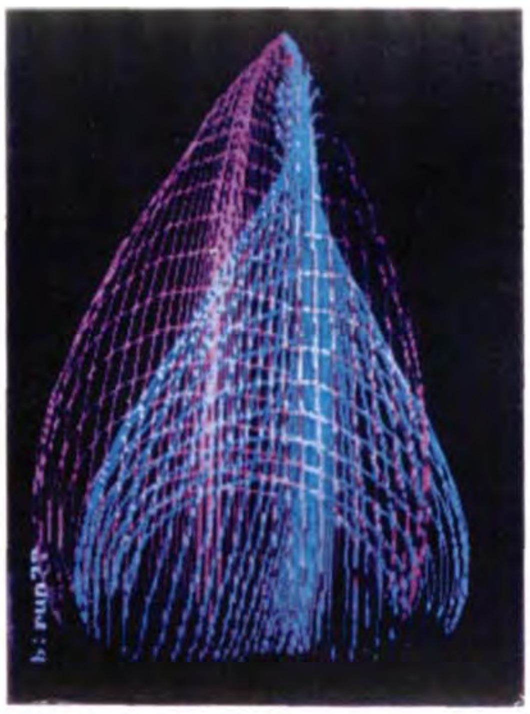

The Double Vortex of a different magnet
mapped at a slightly different angle.

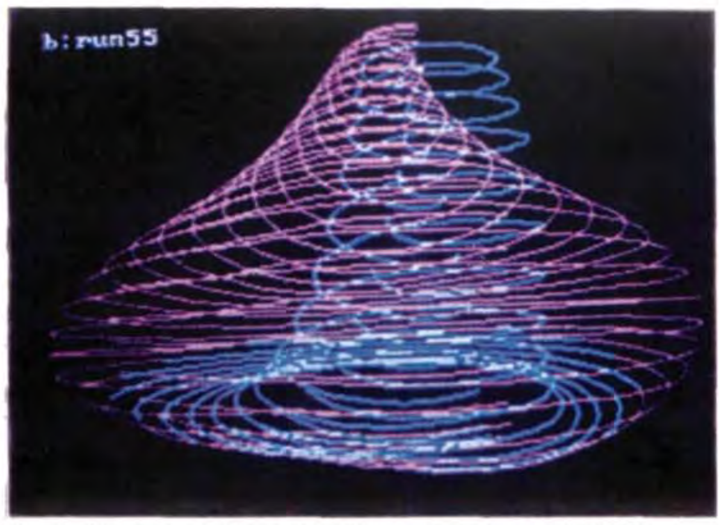

Slightly different materials result in different formations of the fields.

When dealing with Double Vortices, different arrangements of magnets can be used to manipulate the form in which a Double Vortex shows up.

In a different experiment, in which layers of different kinds of magnets are used, the manipulation of the strengths of the different layers produced the formation of a vortex within a vortex. Notice illustrations and descriptions:

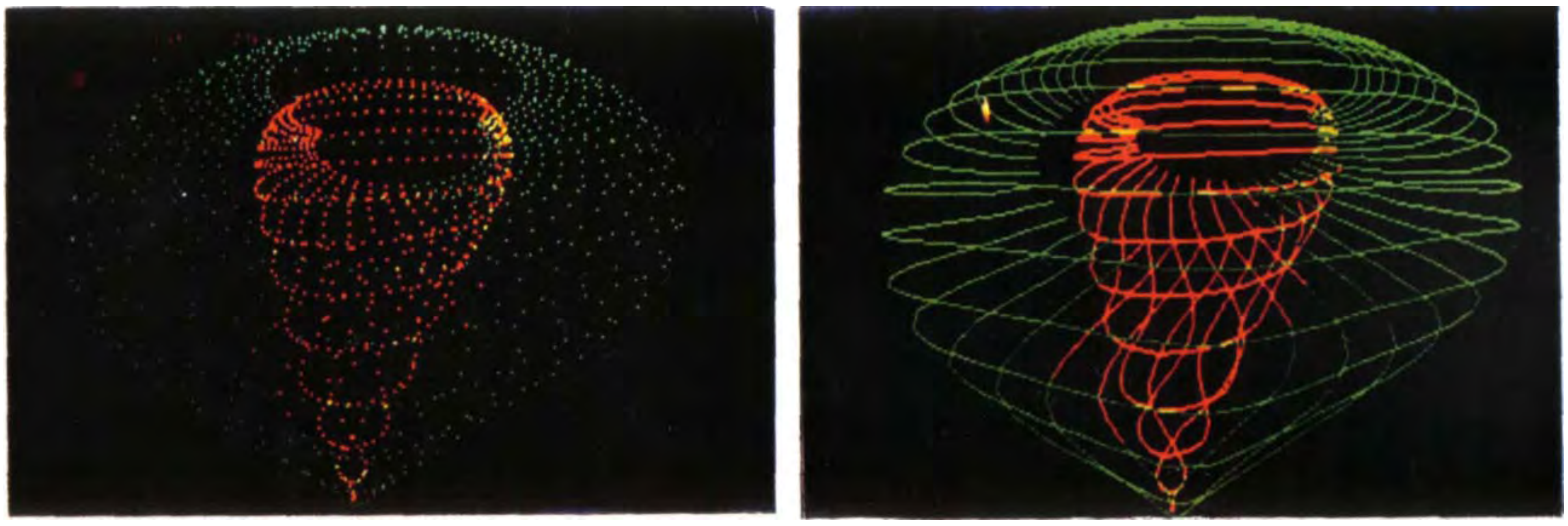

The 3-D mapping showing the tracks of the
particles in a particular "vortex in a vortex"

The following three pictures show the vortex in a vortex (a), the "south" vortex (b), and the "north" vortex (c).

(a)

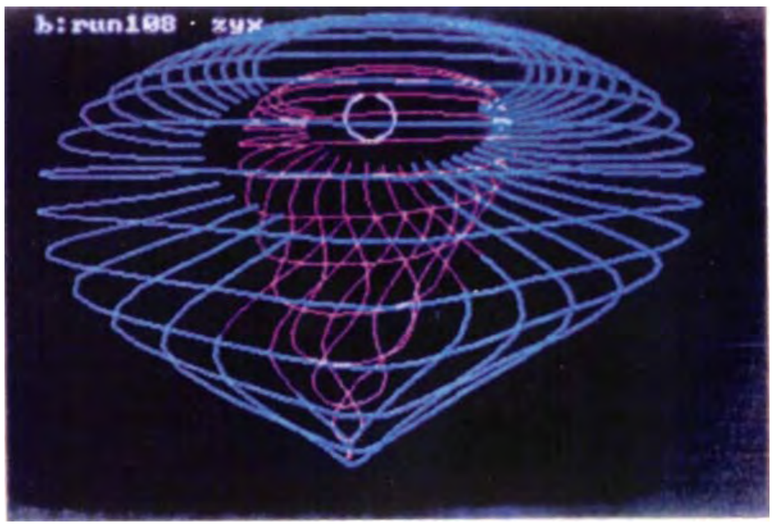
b:run108 · zyx

(b)

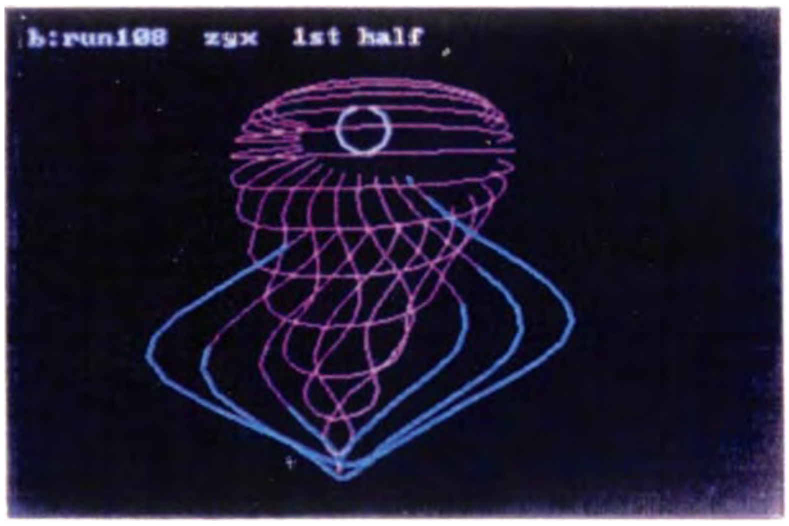
b:run108 zyx 1st half

(c)

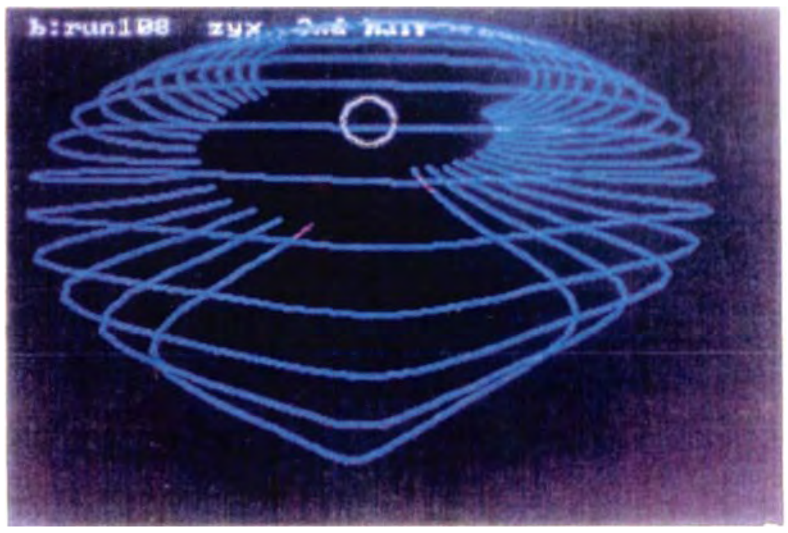
b:run108 zyx 2nd half

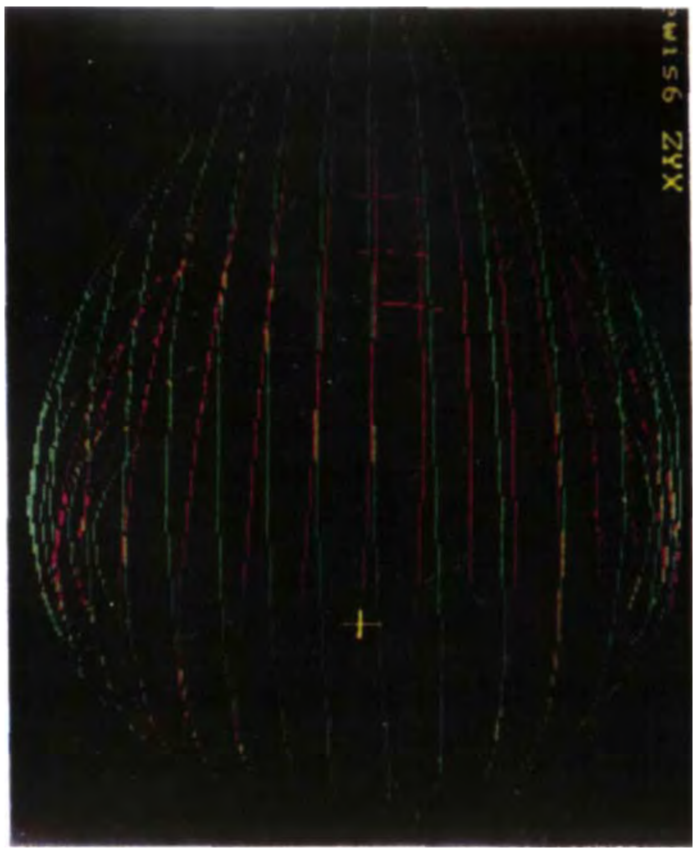

The vortex within another vortex is formed by the combination of three different magnets. The fields shown exist immediately above them when they are layered like a sandwich and standing on edge. This magnetic sandwich is composed of a ceramic magnet, neodymium magnets, and magnetic rubber or vinyl (similar to that on the door of your refrigerator).

The computer is also used to register the percentages of the two particles that make up the two vortices. (See "Mapping of Magnetic Fields" on page 29.) These percentages are important in determining the momentum of the magnetic field. These two populations are distinguishable in the recording process because the different particles are going in opposite directions.

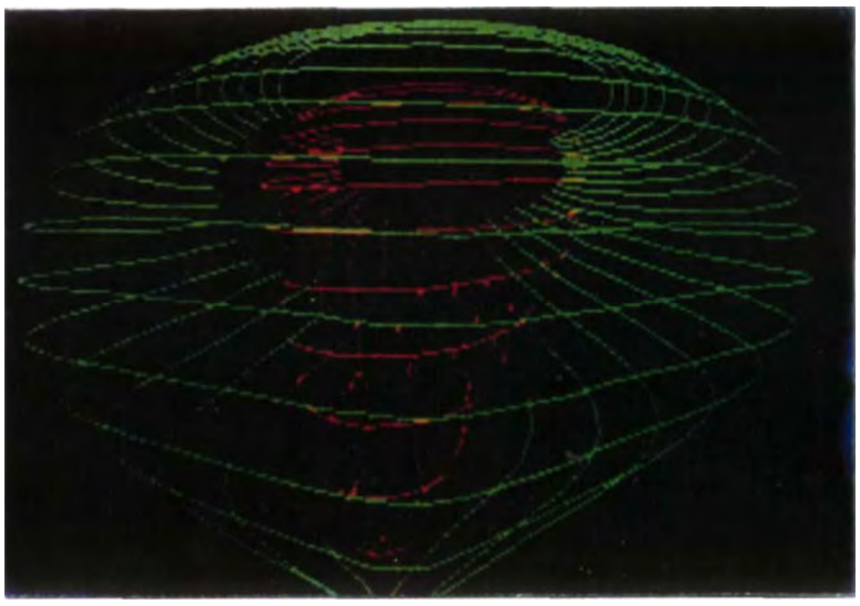

Notice the 3-D effect that the mapping produces.

The Double Vortex in a different magnet has a different form, as is shown here.

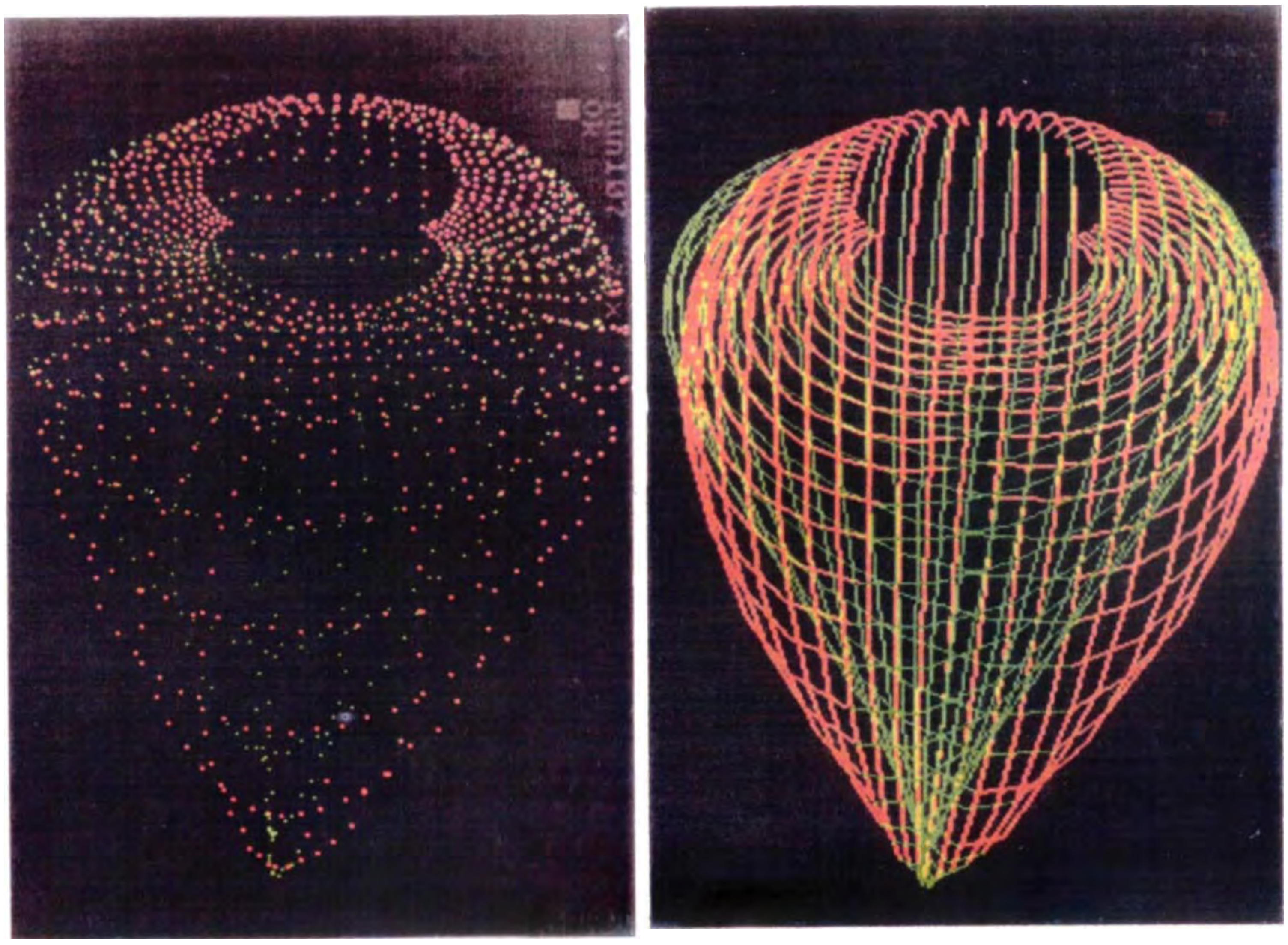

The following is a theory that may help to explain the various conditions of the Double Vortices:

Since the Double Vortices can be arranged so that they are in different relationships to each other (i.e., alongside or within each other) their relationship to each other determines, or may determine the momentum of the field.

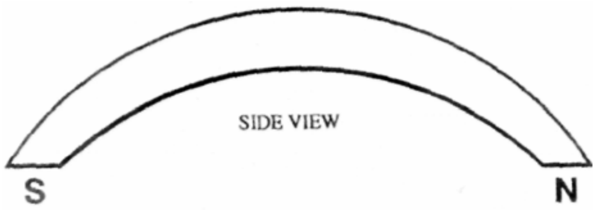

Here is a graphic computer printout of the plotting picturing the above.

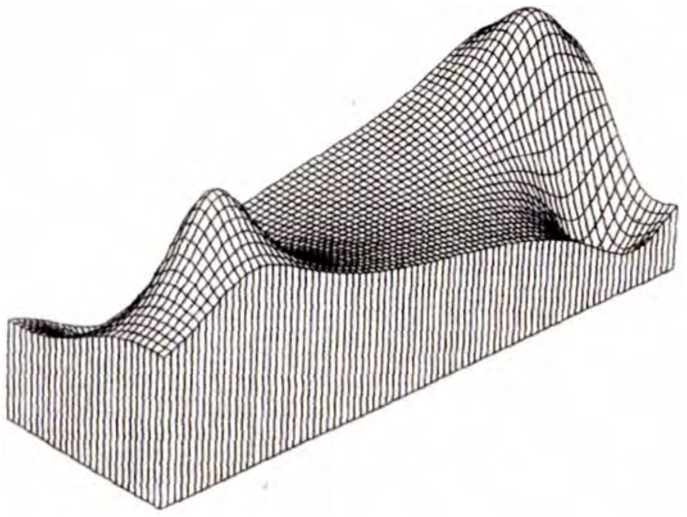

The different axes show the Double Vortices at either pole.

Case in point: Maybe the vortex in a vortex demonstrates the apex of unity and concentration of the field, giving a single pole the most direct thrust possible.

A magnet that clearly depicts the two vortices at each pole is the "banana" shaped curved magnet. The magnet:

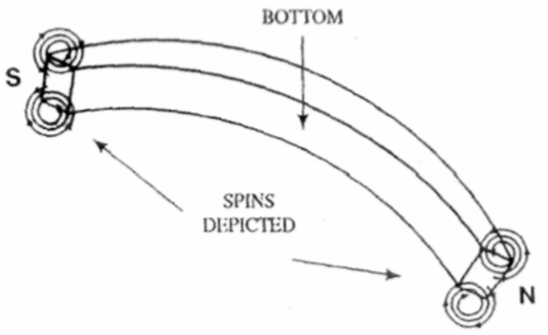

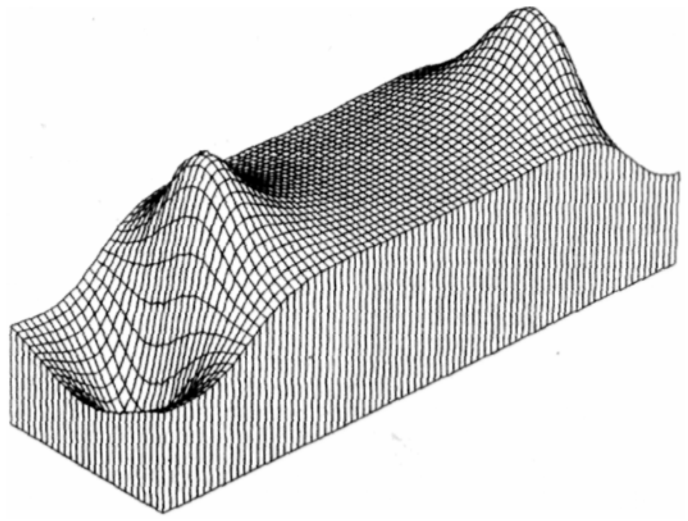

ATTRACTION AND REPULSION

To this point, the discussions and descriptions have dealt with single magnets, or single magnet arrangements and their fields. Now, we will present interactions between magnets, and show what really happens in attraction and repulsion.

Taking a ceramic magnet magnetized through the thickness we mount a curved metallic magnet over it and monitor the reacting fields in a one-half inch air gap. Study it carefully - the result may not be what you were expecting.
Notice first what happens in attraction;

We are all familiar with the pull of one magnet toward another. But, the mechanism is not visible, even if we use iron filings. What wc need to see is the activity of atomic particles that constitute the magnetic fields.

Our mapping operation shows these particles pairing off as the unlike fields merge.
Examine the illustration:

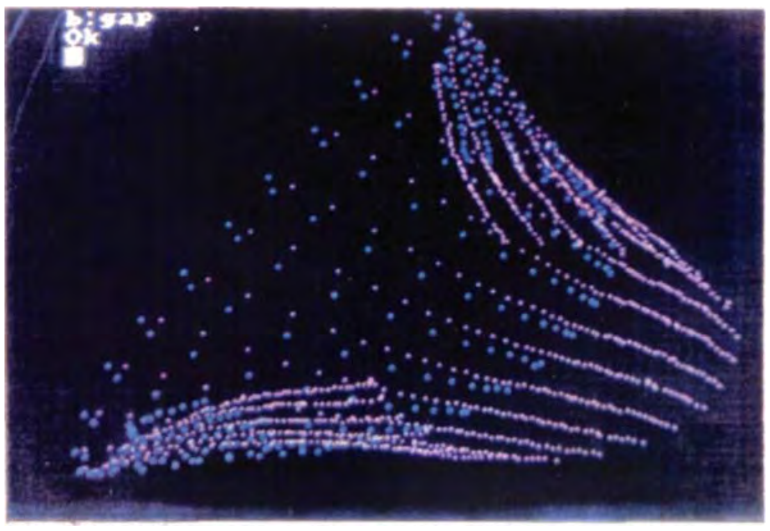

Then, our topographical program snows that the gauss count (the strength of the lines of force) at the attracting end has been reduced, because the pairing of a large part of the particle populations.

The repulsion of like poles represents particle activity which is quite different from attraction.

The particles react with each other as they form two vortices that spin in the same direction. There is no reduction in the gauss count, which registers about three times as high as it does at the attracting end.

Illustration:

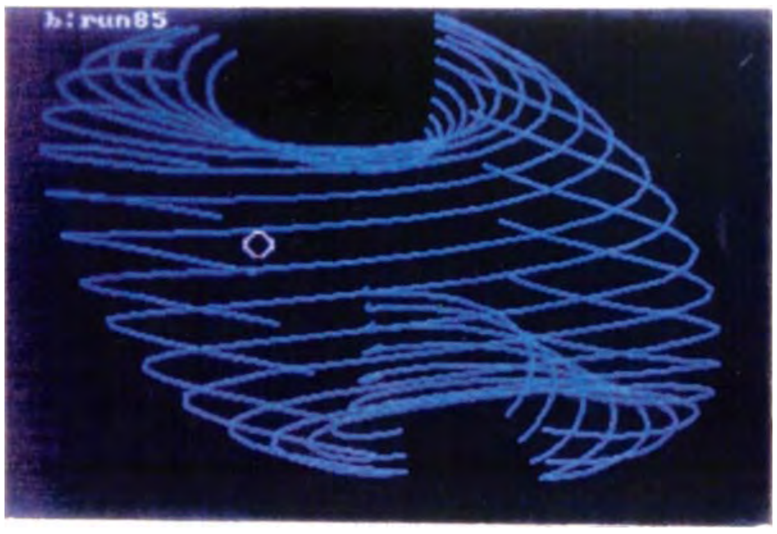

The magnets used in the previous two illustrations, and the one that will follow, appear like this:

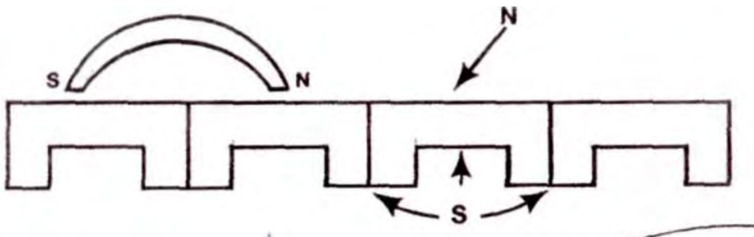

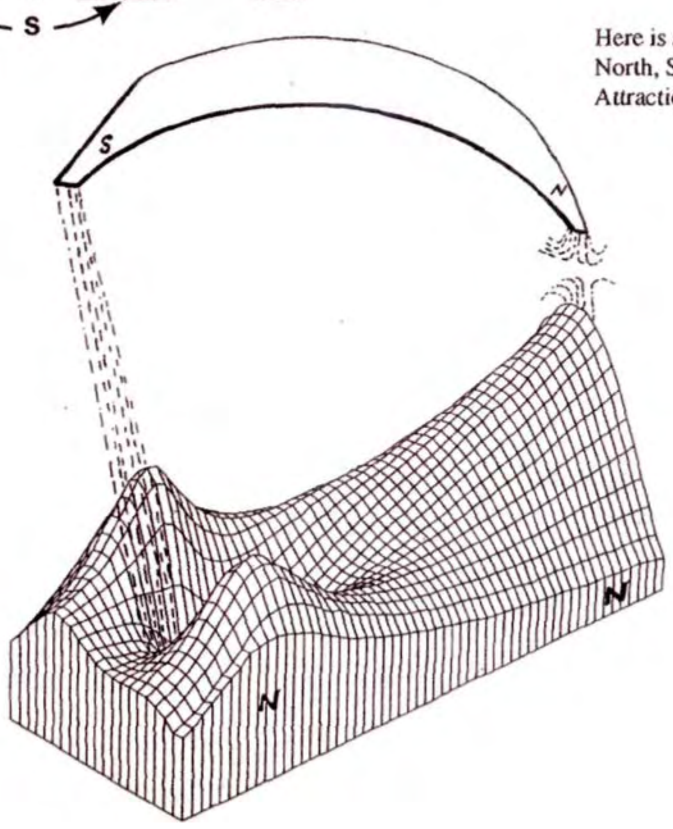

Here is a computer printout:
North, South Pole Differences
Attraction (L) Repulsion (R)

Notice how that (in the previous illustration) the south pole of the curved magnet bores a hole in the field below it. This beautifully illustrates the principles just started on attraction. Also, notice the repulsion, as it does not diminish the strength of the magnet mapped.

ATTRACTION and REPULSION of VORTICES WITHIN A MAGNET

This is a very unique area of interest. Notice the following topographical printout:

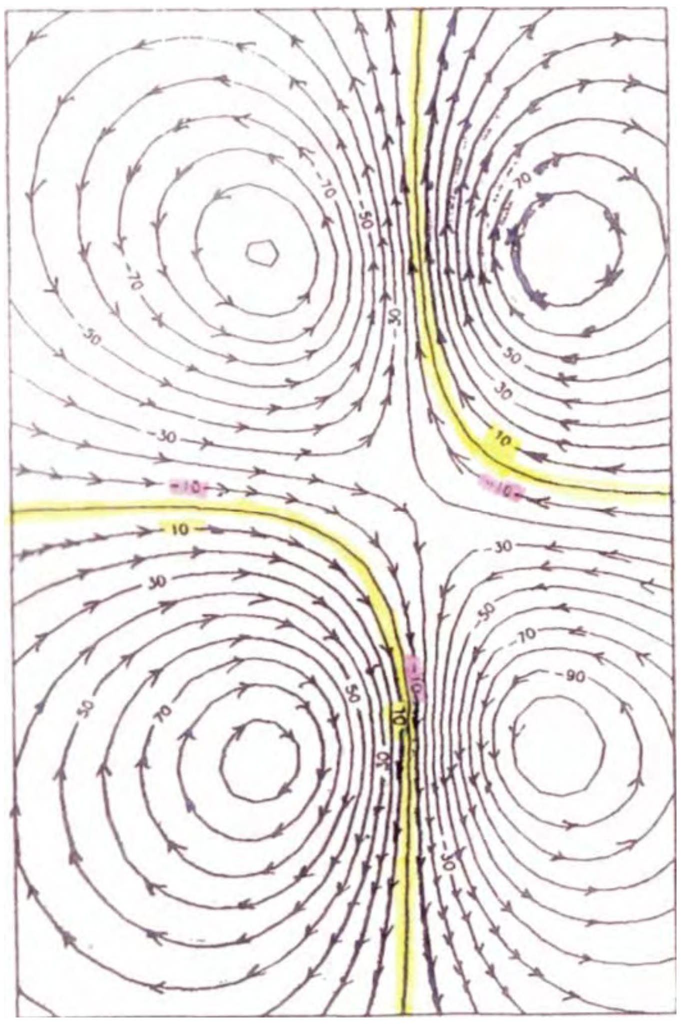

II you look carefully, you will seec that the vortices are separated by zerolines or dead space. The reason is the direction that the vortices spin. Illustration:

Lines of force going in the same direction REPEL

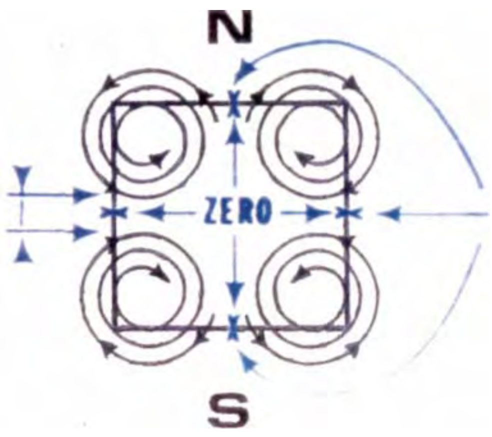

The vortices repulsion of each other causes spaces void of lines of force

Each vortex repels those next to it. Why? Magnetic lines of force going in the same direction repel. Notice that as the lines leave the poles, they are going the same direction, and therefore repel. And also, as they enter the sides, they enter going the same direction and repel each other. This leaves you a fine line in between vortices on the center of the magnet with no lines of force.

Another thing that is very interesting, though, is the fact that vortex spins in opposite corners (in the case of the stronger north element) attract each other. They can form a bond of continuous spins from corner to corner.

Notice the following illustration:

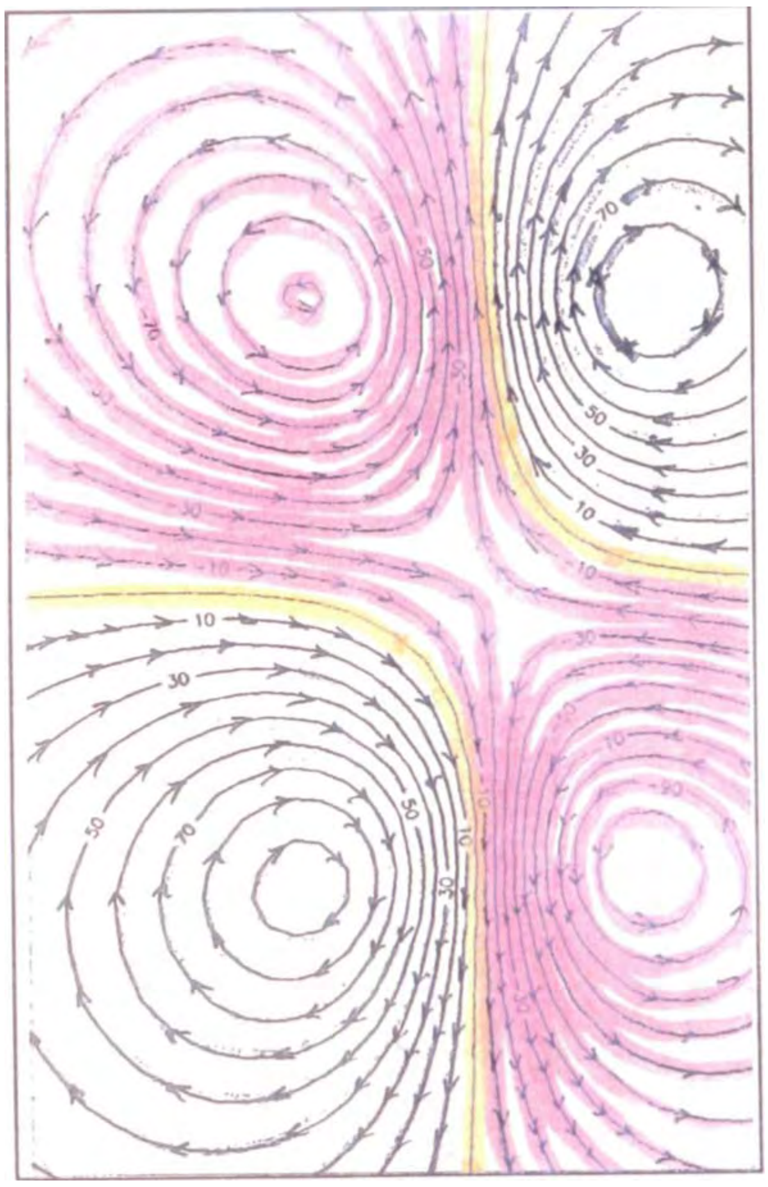
10
30
50
70
70
50
30
10
10
30
10
10
30
50
70
30
50
10
30
10
10
90
50

The evident bond of continuous spins from corner to comer that shows the linkage of the two north elements.

CORNER SPINS

Using the spins (vortex) of an individual corner of a magnet.

We now begin to discuss the arrangements of magnets designed for the purpose of doing work. The work is achieved by interactions between magnetic structures that cause one to drive the other.

The following structure uses a series of magnets with only one corner exposed so that the spins (vortex) of that corner only are (is) used to interact with the spins of a curved magnet, which is to be driven. Illustration:

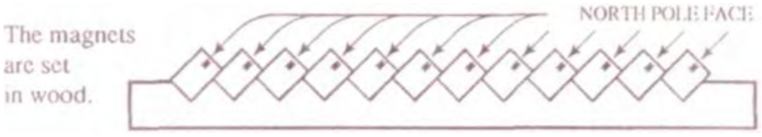
The magnets are set in wood.
NORTH POLE FACE

Actual photographs:

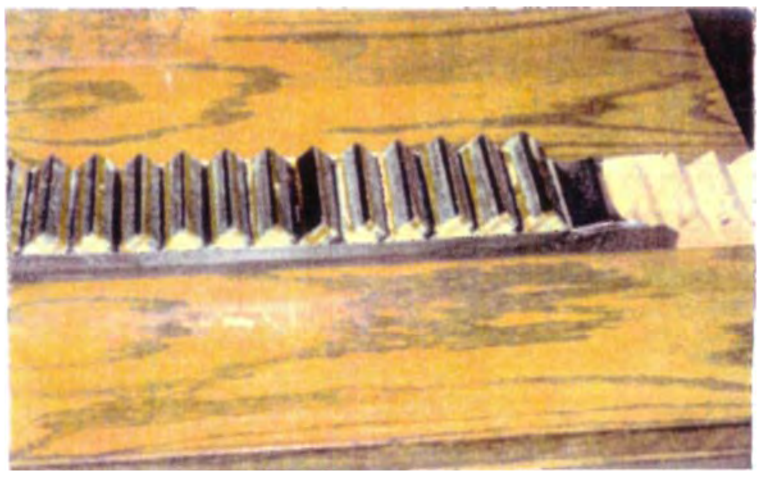

This picture shows the magnets in discussion in the foreground, and the mapping device in the background. (The 3-axis probe can be seen extended into the mapping area.)

HERE IS A LARGER DEPICTION, SHOWING SPIN DETAIL.

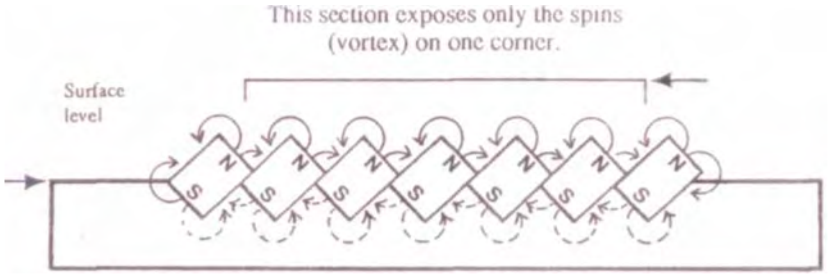
This section exposes only the spins (vortex) on one corner.
Surface level
N
S

Notice that, within the structure, the only spin (or vortex) that is exposed, and affects anything above the magnets, is the one at the uppermost corner, the other north pole vortex is "shorted out", and the south pole vortices are below the structure.

Therefore, with this structure, and a curved magnet placed above it —

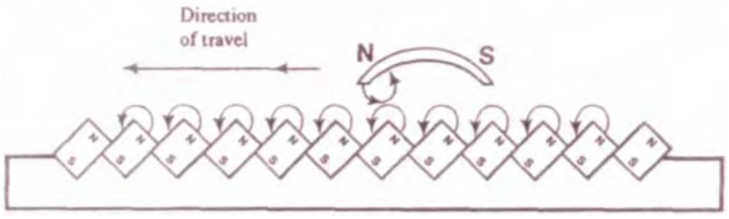

... the interacting spins, going in opposite directions, drive the curved magnet forward. This arrangement of the magnets greatly enhances the driving movement normally due to the right pulsing caused by simultaneous repulsion and attraction.

The pictures made by computer mapping show us that these comer spins tie knots in the lines of force, or make loops.

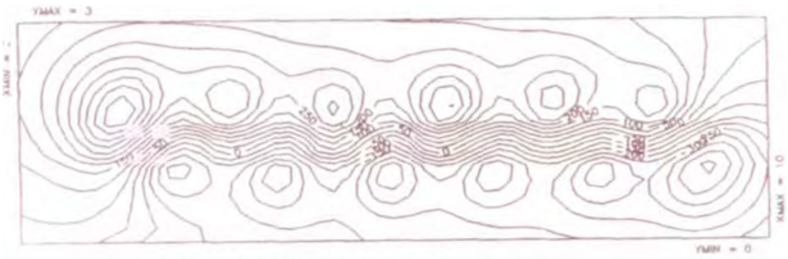

JAWS1 TRACK — RUN149 x axis

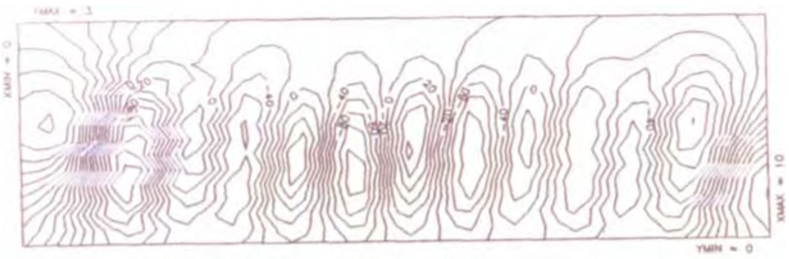

JAWS1 TRACK — RUN149 z axis

Here is how these spins register in this formation:

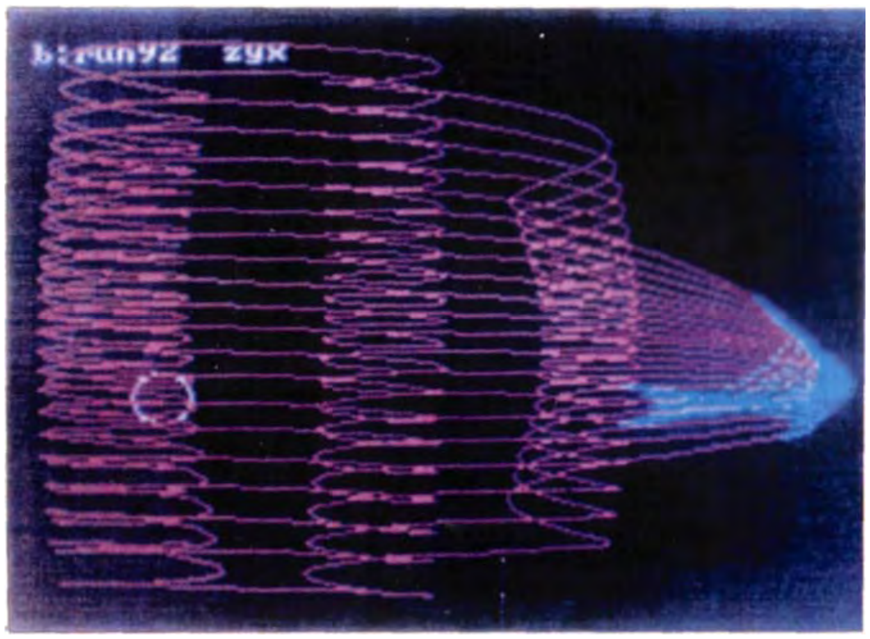

This is just one of the many ways that the magnetic field can be appropriated and used.

THE MAGNETIC GATE

One of the most radical new concepts due to the knowledge of the four spins (vortices) is the magnetic gate. (This is also an application designed for the use of doing work.)

To anyone with a knowledge of physics, the thing about a "gate" that is fully radical is the fact that, in this case, a north magnetic field attracts a north pole. It will reject an approaching south pole.

This result, not anticipated by the physics books, shows how the spins in a magnet and the fields outside can be controlled. In this case, one set of spins and its field is shorted out and the other set takes over. The second set is going in a direction that will provide attraction despite the fact that the compass will register it as an opposing field.

The gate is a complex arrangement of magnets. Its face is a square of four ceramic magnets magnetized through the thickness, with the north pole being the face for the whole square.

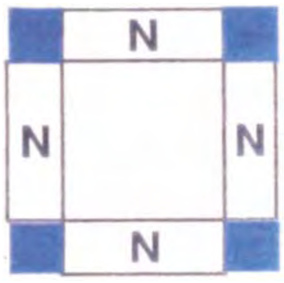

The back side of the gate is a ring of rubber magnet with the north pole on the inside face. This ring attracts to the back of the square. Illustration:

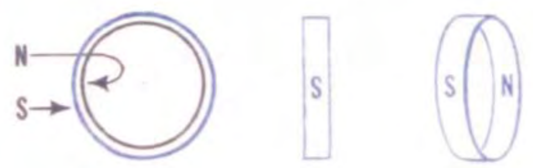

Here is a cross-section of the gate:

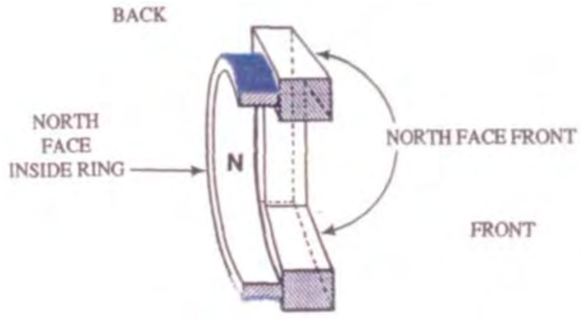
BACK
NORTH
FACE
INSIDE RING
N
NORTH FACE FRONT
FRONT

Another cross-section, to show the spins"

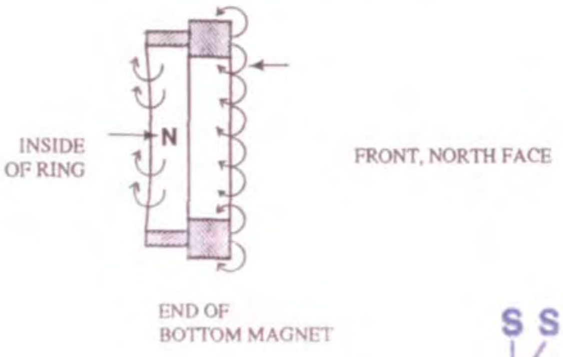
INSIDE
OF RING
N
FRONT, NORTH FACE
END OF
BOTTOM MAGNET

Only the front spins operate on the square; the spins on the back of the square, and the front of the ring, are shorted out.

The south spin on the back of the square goes directly into the north face inside the ring behind.

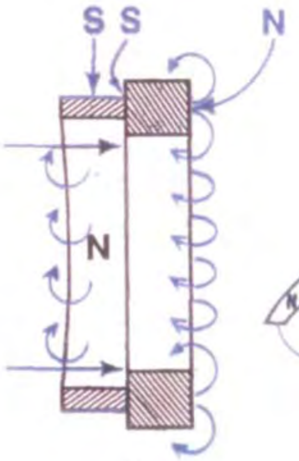

The other spin of the ring's north inside is shorted out into the south back of the square.

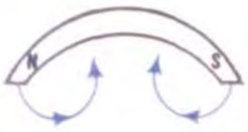

Since there are no back spins going the same direction as the leading north spin of the curved magnet, there is no repulsion. Because the spins of the gate are in the **opposite** direction of the lead north spin, there is **ATTRACTION of two north poles**, and therefore, the spins drive the curved magnet through the gate.

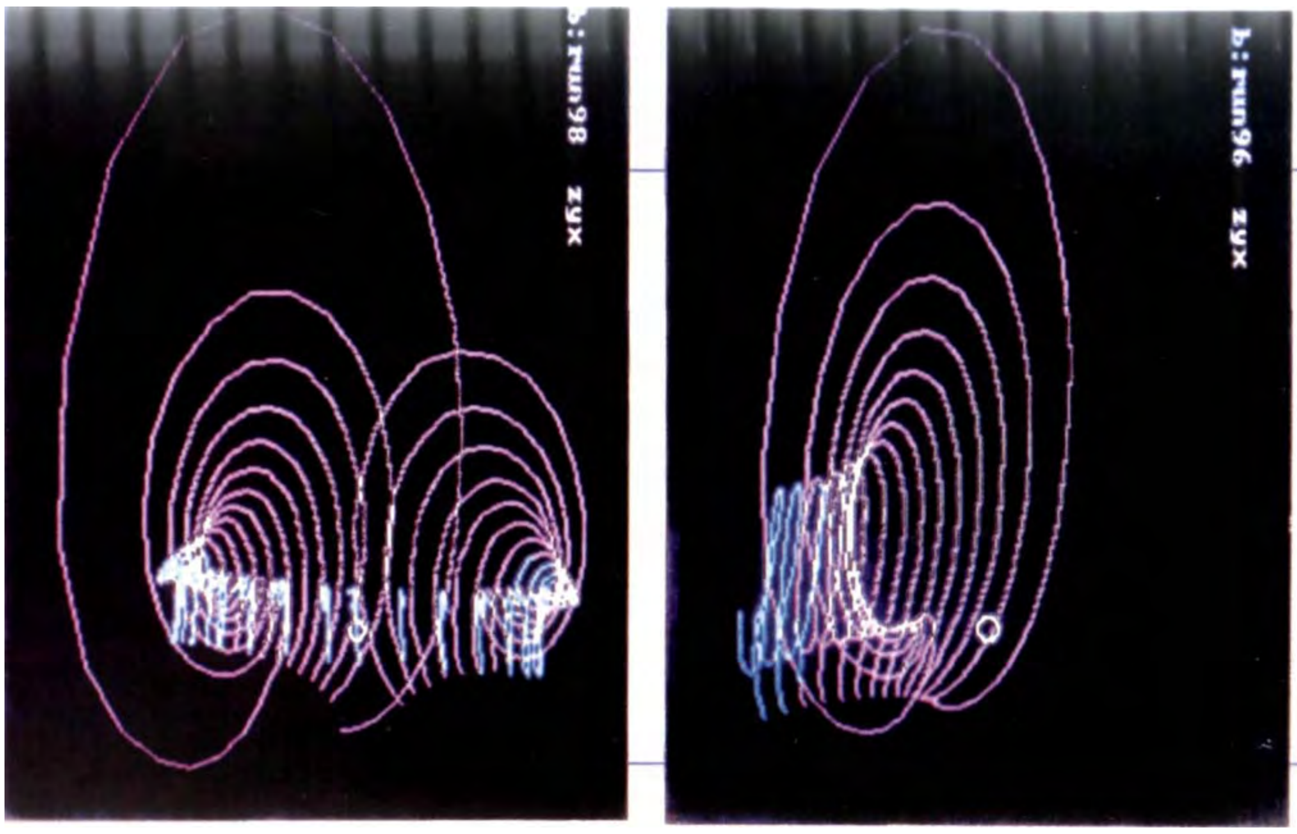
b:run98 zyx
b:run96 zyx

The computer beautifully pictures the spins in the sides of the gate that propel the curved metallic magnet through:

This picture is the interaction of the curved metallic magnet as it enters into the gate:

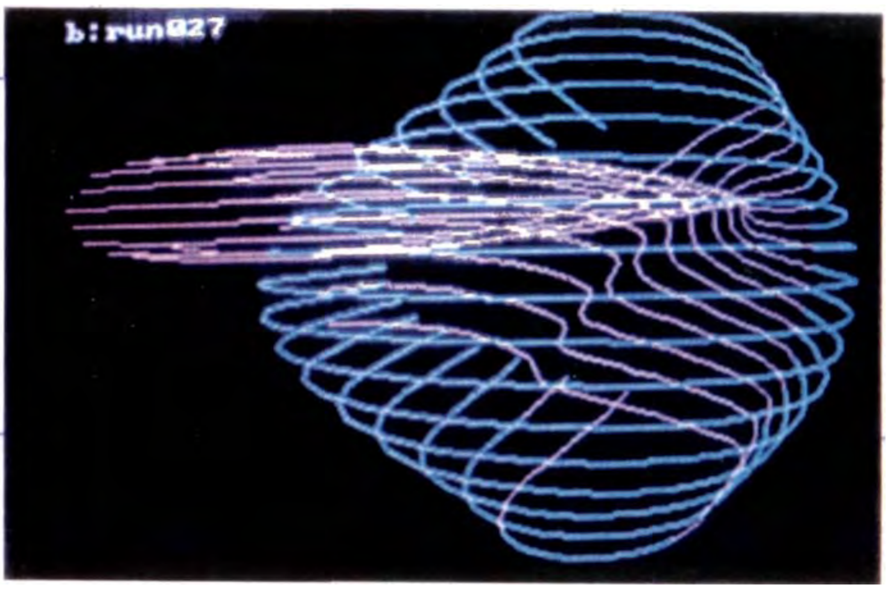

MORE ABOUT THE MAGNETIC GATE

This is a picture of a topographical model of the magnetic gate. The vortices on the sides, which enable the north magnet field to attract another north pole, arc the result of shorting out one set of spins in order to use the other set. The second set is going in a direction that will provide attraction despite the fact that the compass will register it as an opposing field. (Note: The gate's north magnetic field will reject an approaching south pole.)

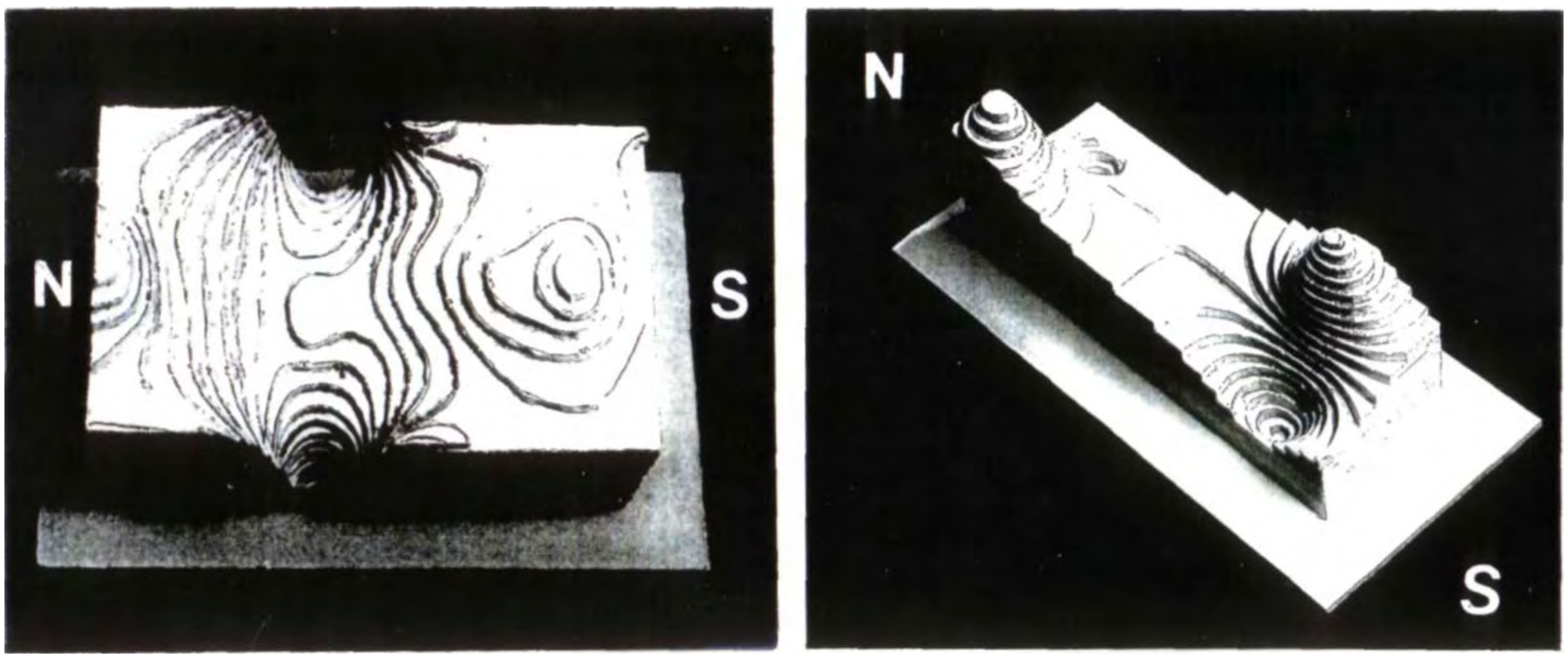

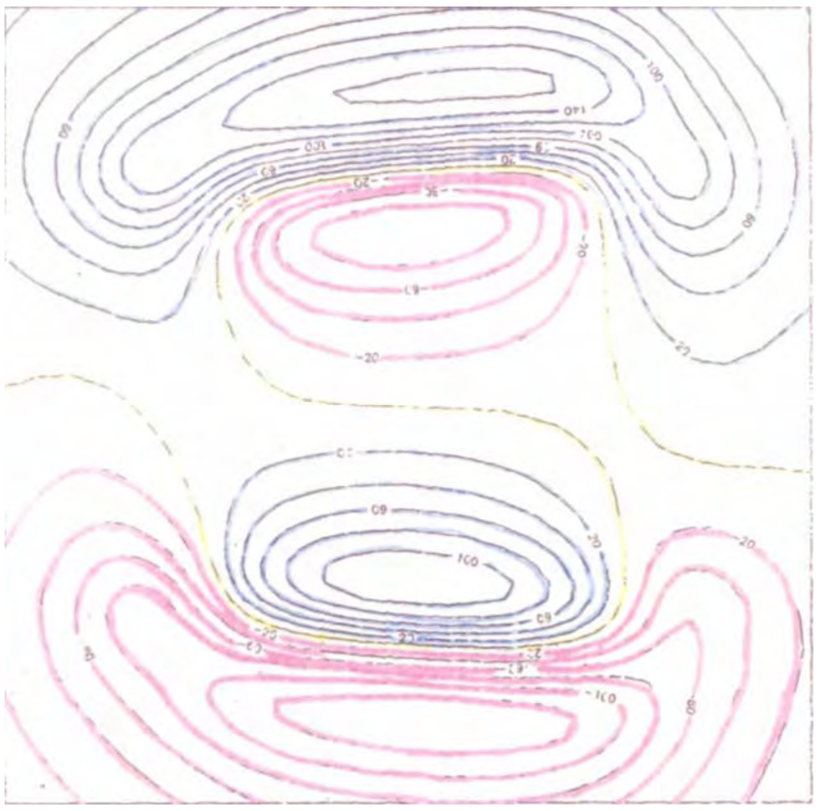

FRONT GATE FACE (north)

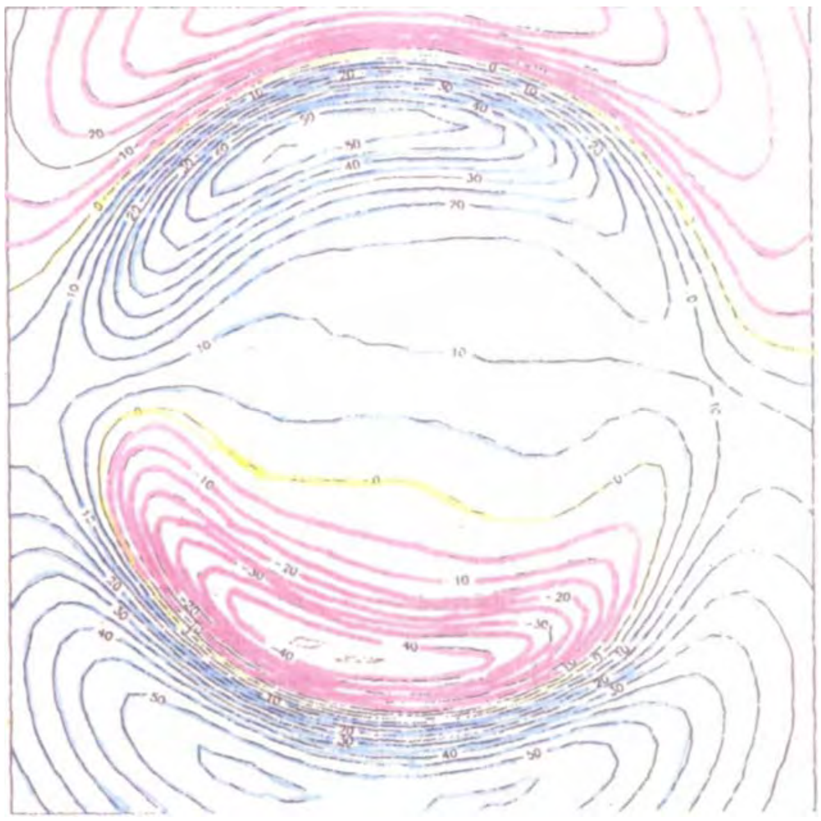

REAR GATE FACE (south)

MAKING USE of the
TIME-ASYMMETRIC QUALITIES of PERMANENT MAGNETS

The following excerpts are from a paper by H. Zocher and C. Torok which concerns research done by these two men. The location where this work look place was Laboratorio da Producao Mineral, Ministerio da Agricultura, Rio de Janeiro, Brasil. E. P. Wigner communicated this Brazilian work to the Proceedings of the national Academy of Science on April 15, 1953.[1]

According to Zocher and Torok, the circularity of conductivity is a time-asymmetric property. Processes of conduction are not only lime-asymmetric, they are irreversible. One direction of time corresponds to the probable course, the other is improbable according to the second law of thermodynamics. Potential differences are time-symmetric. The circular asymmetry consists in the difference of resistance against the clockwise and counterclockwise currents, being probable or not. These currents are naturally time-asymmetric. Hence, a crystal structure, which is time-symmetric, cannot cause acircularly asymmetric conductivity. Such a property can be found only in a system with time-asymmetric circularity, with a mechanical or electrical rotation, with Coriolis forces or magnetic fields. The Hall effect corresponds indeed to circular electric conductivity, the RighiLeduc effect to a circular thermal conductivity, both produced by a magnetic field.[2]

The consideration of space-time asymmetry may prove to be useful outside the realm of crystal physics. The process of splitting a magnetic dipole into two free magnetic charges is seen to be impossible if one considers the space-time asymmetry involved in this process. Present-day literature is ambiguous concerning this point. Thus Dirac[3] discusses the reason why the separation of electric charges is so much easier than that of magnetic charges and Ehrenhaft believes to have succeeded in obtaining free magnetic charges.[4]

According to the given concept, a magnetic moment, and therefore a spin as well, always corresponds to a real circulating movement and cannot be considered as an intrinsic property without the character of movement. This reemphasizes the fact that static structures in three-dimensional space arc not adequate to represent the physical bodies and that the space-time relations are inevitable necessities.[5]

We shall speak of time symmetry if the time inversion has no influence upon the sign of the quality to be reversed.[6]

Asymmetry is much more than the counterpart of symmetry. Asymmetry indicates the existence of characteristic differences, whereas symmetry discards characteristic features. Certain physical phenomena are intrinsically related to certain types of asymmetry, whereas certain symmetry elements may exist without necessarily being related to those effects.[7] "C'est la dissimetric, qui cree le phenomene," stated P. Curie.[8]

THE MAPPING OF MAGNETIC FIELDS

To map a magnet's field the sensor must be moved between readings in a regular pattern. Two servomotors advance the Hall Effect sensor after each reading is taken to collect the 6000 needed data points.

To map a 3" X 7" area, taking readings every 0.1", requires over 2000 readings. This must be done three times to measure the X, Y, and Z components of the magnet's fields.

Each component is unique. X, Y, and Z can be looked at seperately, or combined as "magnitude," the square root of the sum of the squares of X, Y, and Z.

Topographic maps can be made of these magnetic fields, just as they can be made to show land contours.

Of particular interest are the null lines where the field changes sign or direction.

Complex assemblies of magnets can be designed to shield, focus, and distort magnetic fields for various purposes.

BIBLIOGRAPHY

[1] National Academy of Science. Proceedings of National Academy of Science, vol. 39 (n. p., 15 April 1953), p. 681.

[2] Ibid., p. 684.

[3] P. A. M. Dirac, Proc. Roy Soc., A133,60 (1931); Phys. Rev., 74,817 (1948), cited by National Academy of Science, Proceedings of National Academy of Science, vol. 39 (n. p., 15 April 1953), p. 685.

[4] National Academy of Science, p. 685.

[5] Ibid., p. 686.

[6] Ibid., p. 681.

[7] Ibid., p. 682.

[8] P. Curie, "Oeuvres," Paris, 1908, p. 127, cited by National Academy of Science, Proceedings of National Academy of Science, vol. 39 (n. p., 15 April 1953), p. 682.

Literature Cited

Curie, P. "Oeuvres." Paris, 1908. Cited by National Academy of Science. Proceedings of National Academy of Science. Vol. 39: n.p., 15 April 1953.

Dirac, P. A.M. Proc. Roy. Soc., A133,60: 1931; Phys. Rev., 74,817: 1948. Cited by National Academy of Science. Proceedings of National Academy of Science. Vol.39: n. p., 15 April 1953.

National Academy of Science. Proceedings of National Academy of Science. Vol.39: n. p., 15 April 1953.

References

Richard P. Feynman, Robert B. Leighton, and Matthew Sands, "A Fields," The Feynman Lectures on Physics. Addison-Wesley, 1963, Vol. 2.

"Seeing is Believing," Elektro-Elektroteknisk Tidsskrift Bd. 95, nr. 12,24 Juni 1982: 20-55.

U.S. Patent, 4151431,24 April 1979, H. R. Johnson.

National Research Laboratory. Annual Report 1985. pp. 166-7.

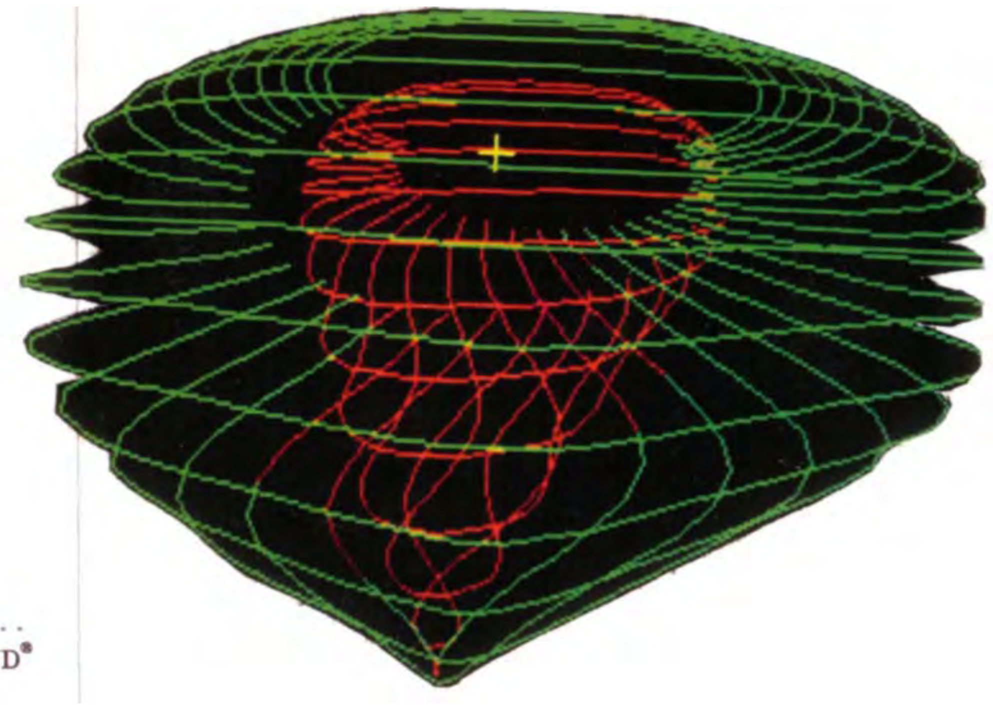
INTRODUCING . . .
WHIRLWIND®
Logo

The Discovery of the Double Vortex

After predicting for years the presence of a vortex in the fields of permanent magnets, Steve Davis and I were working late one night with our three axis gauss meters and a new computer, mapping magnetic fields. I was starting to go home when he announced, "I don't know what I am doing, but I have something here that looks pretty linear."

He proceeded to bring up on the screen, in living color, the forming of a double vortex. Not only was the double vortex there, but we could see as it formed, the opposite spins in such a perfect way. We knew that this had to be the beginning of something new and mighty important. The question now was, "How do we use it to the greatest advantage? How do we explain its importance to the patent structure that we have been developing for many years?"

We resorted to the libraries and studied many months to see what others had done. The results showed a great desert in this area.

Researchers, it seemed, had been content with the ancient iron filings as a mapping tool and had not used twentieth century methods to see what could be seen. The field had simply been ignored.

We ran a picture on the cover of National Laboratory but only the magazine seemed to sense the importance of it.

We have used our mapping methods to show the fields around a conductor, to show how a new generator works, to describe the thrust of permanent magnet units, and to explain our versatile gate discovery.

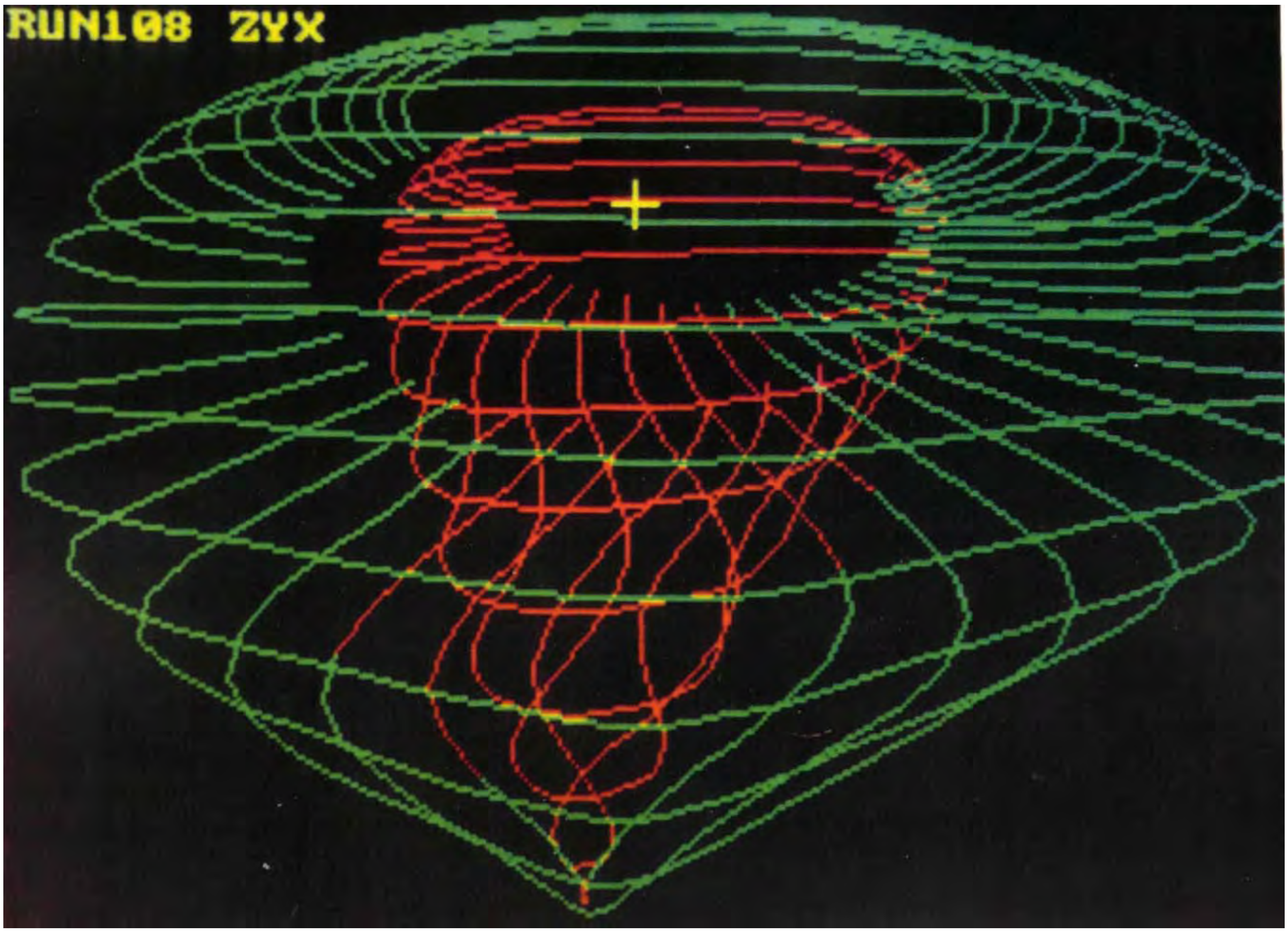
RUN108 ZYX

Amperian Currents

in Permanent Magnets

We are familiar with AC & DC electricity but we are not so familiar with Amperian currents in magnetic material. Yet, Ampere told us much about them over 150 years ago.

Today our magnetic materials are much better due to the rare earth usage. Fields of 35,000,000 Gauss Oersteds are available.

These Amperian currents are tightly wound in the material. They are firmly anchored so that they can not ordinarily be reversed. Thus, they should be available for many years of usage.

Three kinds of Amperian currents we have observed are:

1. The double vortex where opposite spins are found along side each other.
2. The double vortex where one vortex is inside the other.
3. The third form is the flat vortex.

Dr. Feynman recorded finding some of these in his Vol. 2 Physics lectures 37 - (12 - 13).

The interaction of the momentum of these currents is the basis of our patent work for the last number of years.

"Ampere was the first investigator to propose that the magnetism observed in permanent magnets is caused by tiny electric currents circulating within the molecules of magnetic material." Scientific American Jan. 89.

"Magnetism - more specifically ferromagnetism - in a material is associated with cooperative interactions between individual atoms tending to align the magnetic moments of those atoms parallel. The magnetic moment of an atom arises from the orbital and spin angular moments of its electrons. Only some elements have unpaired electrons - hence magnetic moments - and even fewer show the cooperative interaction necessary for ferromagnetism.

"A permanent magnet (PM) is a piece of a material that has stored within it magnetic energy - by alignment of magnetic moments - supplied by an electric field during the initial process of magnetization. The magnet retains this energy indefinitely - it is permanent. The material can be a metallic element, a metallic alloy, or even an oxide.

"There is a growing tendency to replace electromagnets by PMs because of major improvements in PM properties. The increasing cost of energy and the trend towards miniaturization are other reasons. The samarium-cobalt series of magnetic materials until recently provided the strongest PMs known. However, recent discoveries with alloy systems based on iron and neodymium promise even better performance.

"The two basic parameters used to define properties are the remanence, Br, and the coercivity, Hc, the vertical and horizontal axes respectively on the Lanthology diagram. The remanence arises from the cooperative alignment of magnetic moments. The coercivity measures the resistance to demagnetization of the material; a high value is essential in devices where the magnet will be subject to strong demagnetizing fields such as in motors.

"The coercivity depends not only on the underlying crystal structure but also on the microstructure of the material, on the domain morphology within the bulk magnet.

"Demagnetization is resisted when a large energy is needed to reverse the aligned magnetic moments within a crystallite. In some crystal structures certain directions - determined by the orbital moment of the minor component together with crystal-field effects and exchange interactions - provide an exceptional resistance. A great deal of energy will be required to reorient the magnetic moments from one easy direction to another. The crystal is said to have a high magnet-crystalline anisotropy.

"The role microstructure plays in giving high coercivity is related to the existence of domains - regions of common direction of magnetization - within a practical PM material. The movement of domain walls - separating domains - must be inhibited. This is often done by incorporation of another phase within the material. Much of the art of producing PMs lies in this microstructure control.

"The Neodymium-Iron systems, the latest and most powerful permanent magnets, seemingly provide the best yet approaches to the two mechanisms outlined above for high coercivity. They are being intensively studied by many research groups." "Neodymium, Iron, a Pinch of Boron and Permanent Magnets" Union Molycorp.

Visualizing Magnetic Fields

by

H.R. Johnson, S.M. Davis, and G.H. Beyer
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061

ABSTRACT

In unrestricted 3-space, what patterns would be seen if iron filings were not constrained by gravity to lie in a single plane? We have recorded magnetic field patterns and displayed them in a variety of ways: as 3-D surfaces; as contour maps; and as plots of two of the field components, with the third component's sign conferring color to the tip of the field vector. Our hope is that others may also find such patterns helpful in understanding magnetic phenomena.

One of Faraday's many contributions was his concept of force as traced by iron filings aligned on a glass plate above the poles of a magnet. However, the pattern is perhaps somewhat misleading since the iron filings can move only in the plane of the plate. In unrestricted 3-space, what patterns would be seen if the filings were not constrained by gravity to lie in a single plane, but could show the true direction of the magnetic vectors?

We measure the components of our magnetic field using three mutually-perpendicular indium arsenide semiconductors. Their Hall effect voltages are amplified, digitized, and recorded on a disk by an IBM personal computer. The probe containing the three semiconductors is accurately positioned by two servo motors which progressively scan the area above the poles in a series of small steps. The thousands of data points are then displayed in various ways.

What are the typical data and displays for a 6-inch Alnico magnet, shaped like a banana, having Nd-Fe-B pole pieces, each 1/8 x 5/16 x 1 inches?

The menu shown in Figure A summarizes such data. The area surveyed, 1/2 inch above the pole pieces, was 8.0 inches long by 2.0 inches wide. The sensor was moved in 1/10 inch steps to obtain 1701 data points for each component of the magnetic field. A Bell 620 gaussmeter amplified the voltages from the Hall effect sensors.
The signals were digitized by a Metrabyte interface board and recorded on a floppy disk. These data then served as input to our display programs.

Three-dimensional surface plots for the X (vertical), Y (horizontal), and Z (longitudinal) directions are shown in Figures B, C, and D, respectively. A typical contour, or topographic map, is shown in Figure E.

What good are these magnetic data? Are the 90 minutes necessary to obtain 5103 measurements well spent? We think that three distinct accomplishments are represented here.

First, we have learned to cope with the complexity of data acquisition using a personal computer, independent of dedicated, expensive mainframe time on a larger computer. The personal computer has made such activity affordable and efficient.

Second, we have learned to make contour plots and models which show the location of zero lines, where the field changes sign, or direction. Contour plots show also precipitous field changes where they occur, where the lines are closely spaced. Such data should prove helpful in the design of magnetic assemblies that can do unique tasks. By shading poles so they present an unsymmetrical structure when viewed relative to other magnetic components, unbalanced forces can be generated, causing motion in a preferred direction. Magnetic gates can be designed, and the optimum orientation of interacting magnets can be studied.

Third, we have made magnetic measurements that are accurate, reproducible, and easily stored for further review using a wide variety of display programs.

Using this method we have secured the following pictures:

1. The north and south poles of a curved magnet.
2. The picture of the fields around a current carrying wire.
3. The magnetic bullet formed in a permanent magnet railgun operation.

H. R. Johnson, Director of the Permanent Magnet Research Institute,
Box 199, Blacksburg, VA 24060.

S. M. Davis, Electrical Engineer and Consultant.

Dr. G. H. Beyer, Distinguished University Professor of Chemical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061. *Recently deceased.

Address all inquiries to H. R. Johnson at the above address.

FIGURE A

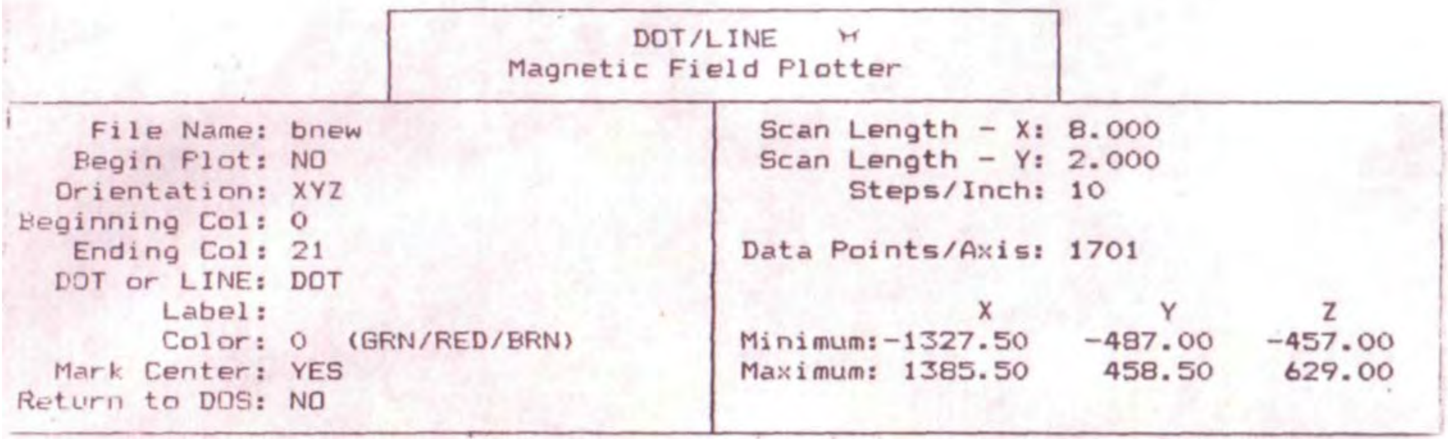

```
                         DOT/LINE
                   Magnetic Field Plotter

   File Name: bnew              Scan Length - X: 8.000
  Begin Plot: NO                Scan Length - Y: 2.000
 Orientation: XYZ                    Steps/Inch: 10
Beginning Col: 0
  Ending Col: 21               Data Points/Axis: 1701
 DOT or LINE: DOT
       Label:                          X          Y          Z
       Color: 0  (GRN/RED/BRN)  Minimum:-1327.50   -487.00   -457.00
 Mark Center: YES               Maximum: 1385.50    458.50    629.00
Return to DOS: NO
```

FIGURE B

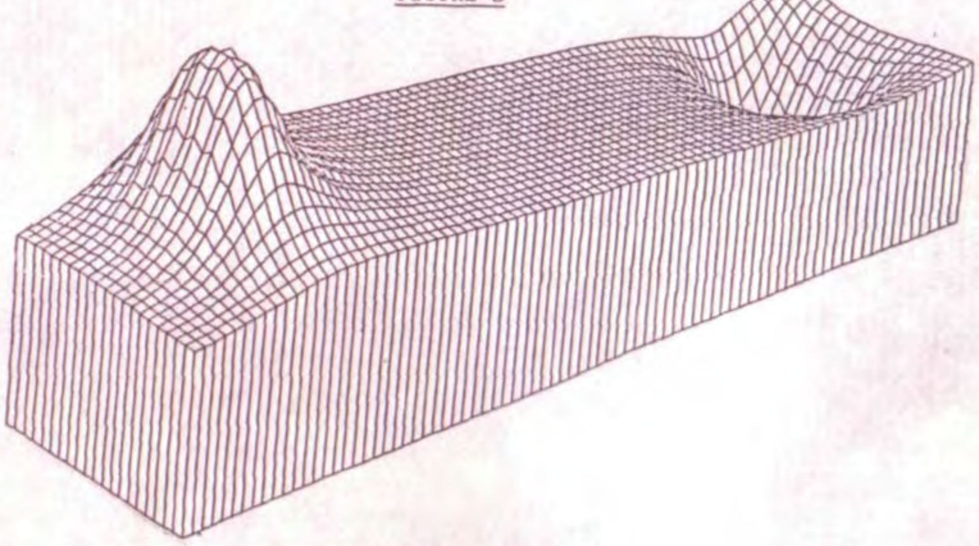

FIGURE C

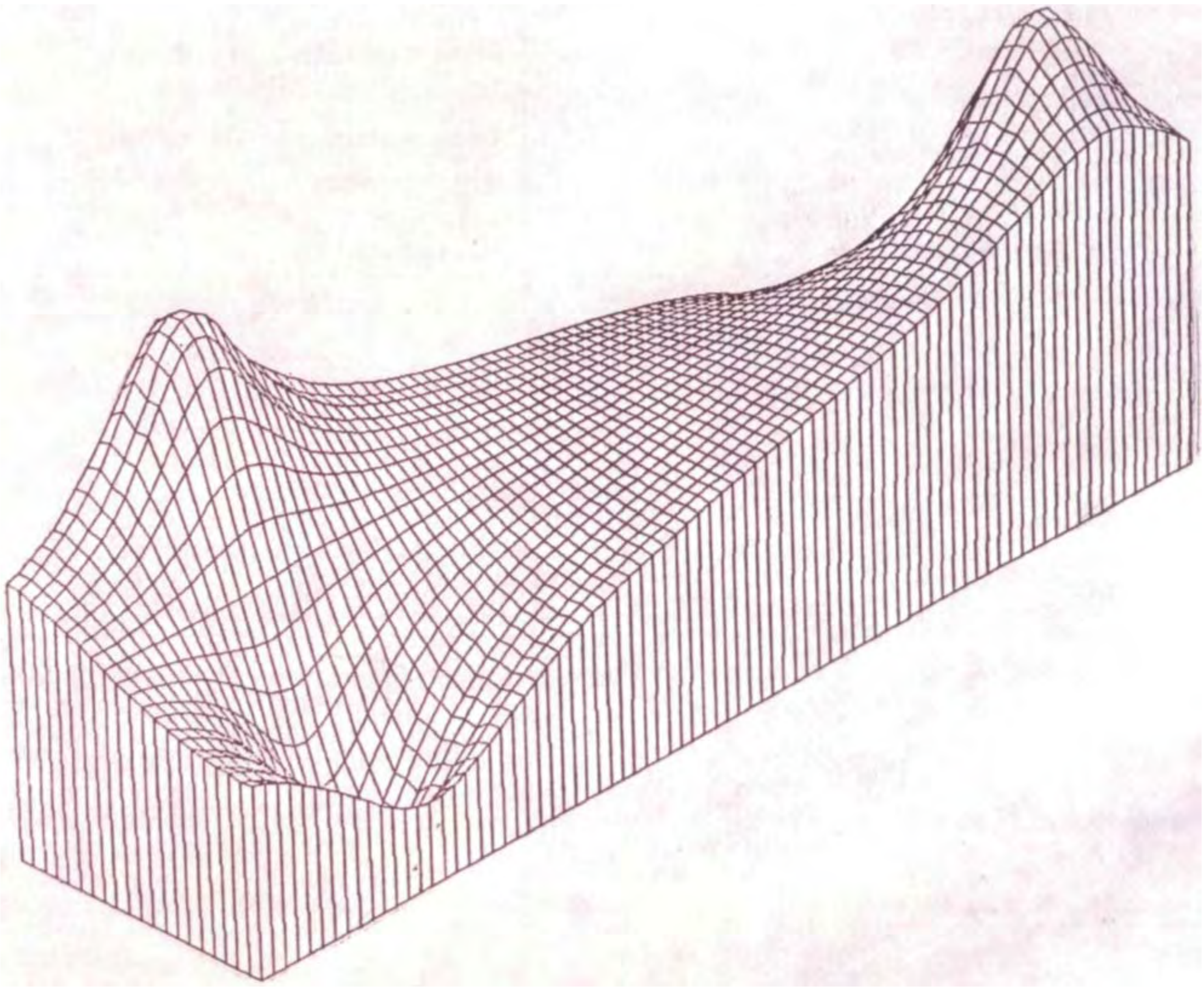

BNEW y axis

FIGURE D

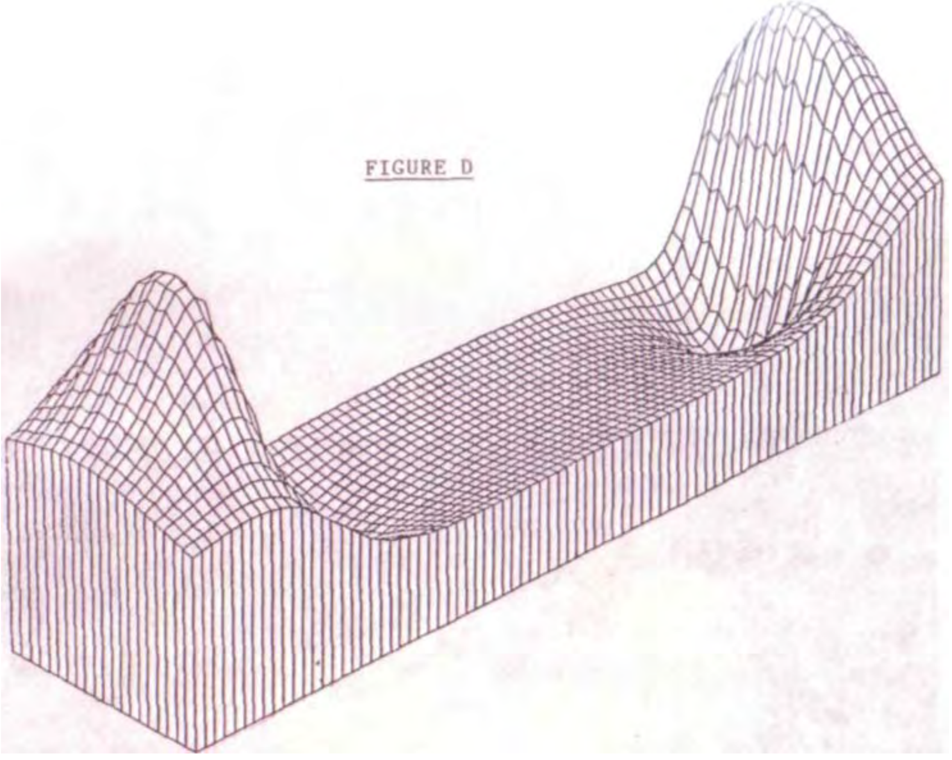

- BNEW z axis

FIGURE E

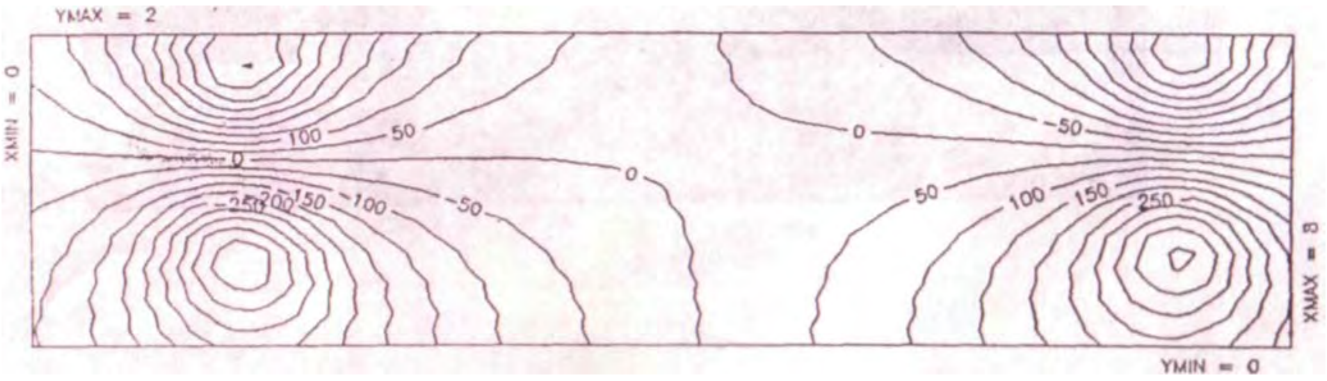

- BNEW y axis

1. THE NORTH AND SOUTH POLES OF A CURVED MAGNET:

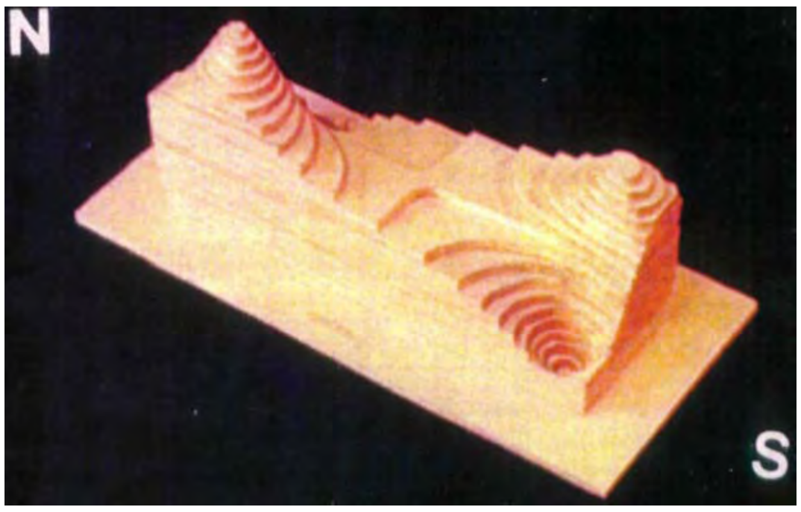

2. THE FIELDS AROUND A CURRENT CARRYING WIRE (D.C.):

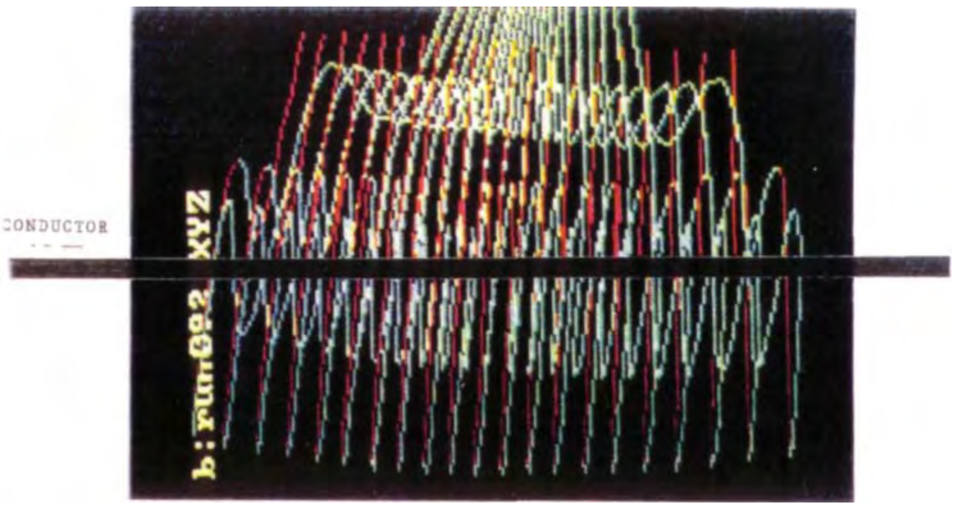
CONDUCTOR
b:run092 XYZ

3. THE MAGNETIC BULLET FORMED IN A PERMANENT MAGNET RAILGUN OPERATION:

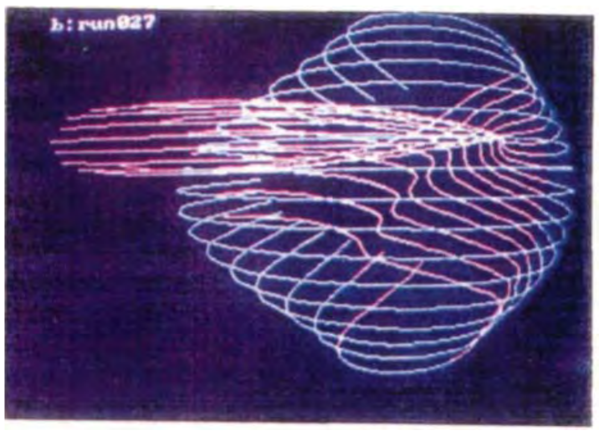

The Dynamo of Faraday and

The Piezomagnetic Dynamo

If a current is sent through a coil of wire between the poles of a large magnet, the coil will rotate. This is due to the directional magnetic field of the coil and the directional field of the magnet creating a squeeze effect in one direction.

The same action can be created by the reaction of two permanent magnet fields. A further refinement is to have the permanent magnet fields to attract one another until they are in a position where the fields are squeezed and the movement in the same direction is accelerated. This we call the piezomagnetic effect.

The secondary effect is also described as due to the quantum-mechanical exchange forces. Due to the squeeze effect, the spins become parallel instead of anti-parallel as they were in the original attraction. This molecular crowding is believed by some very good theorists to create an exchange force 1,000 times greater than
the purely attractive magnetic forces, we have noted this force for many years but did not have the explanation originally given by Heisenburg.

One thing we did notice was the fact that in a very strong field that a light armature did not run much faster than a unit that weighed four times that much.

Using the above method, we have constructed a strong shaded pole to attract a strong magnetic field into this intense magnetic field. Passing thru, this unit has developed effective thrust to accelerate the vehicle down a level track.

Carrying more than its own weight, the repeated actions of these cycling magnetic pressures have shown what a piezomagnetic effect can do.

It is a renewable energy source. Our maps of the fields show the forming of a composite magnetic field around the armature. This we have termed, the magnetic bullet. It is shown on the front cover of this book. We have entitled this activity THE PIEZOMAGNETIC RAILCAR because it is similar to the energy released by squeezing a piezoelectric crystal.

The crystal used does not get tired when used constantly; neither do the magnets, They will continue to generate the same amount of thrust as the car moves from section to section.

You have been reporting a number of defects in our society these days. How would you like to consider something completely new, something that is positive, not a tired rework of a current bit of knowledge or philosophy.

Applying our latest science to something considered old and established has shown how little we know about it. We have seen the vast sea of ignorance in which we swim. It has shown how the life long work of one individual can penetrate the layers of false concepts, learned superstitions, and degreed security blankets.

Not taking for granted some textbook conclusion, we have used direct methods of measuring that reveal big gaps in our storehouse of knowledge. Using this along with the belief that enough clues will enable you to catch a criminal, we have found that enough information lets you solve age old problems.

As we struggle to conserve and not to contaminate, it has given great comfort to find that we can do both. The goal is achievable.

Our new approaches are documented in three patent applications. Two of these have been granted and the third is pending.

As shown in our new book, we have found a way to shade the poles of permanent magnets, a way to make north poles attract north poles and reject south poles, a way to increase spin intensity and magnetic thrust, and a way to use the transfer forces to accelerate masses.

The resulting motor does not destroy its energy source. It is a reusable atomic source that is available world wide.

In biology, the discovery of the double helix and the genetic code has provided a tremendous fund of information.

In our case we have discovered a similar situation where we have the double vortex that occurs in all permanent magnet configurations.

The exploration of the genetic code by biologists with extensive government funding has begun.

We have just started the exploration of the double vortex and the magnetic code, but the fallout thus far has been extremely rewarding.

Note in the pictures, the mapped spins, the pictures of north poles attracting north poles and rejecting south poles, and the thrust developed by these units without any outside energy contribution.

United States Patent [19]

Stocklin

[11] **Patent Number:** **4,858,612**

[45] **Date of Patent:** **Aug. 22, 1989**

---

[54] **HEARING DEVICE**

[76] Inventor: **Philip L. Stocklin,** P.O. Box 2111, Satellite Beach, Fla. 32937

[21] Appl. No.: **562,742**

[22] Filed: **Dec. 19, 1983**

[51] **Int. Cl.⁴** **A61N 1/36**
[52] **U.S. Cl.** .... **128/422;** 178/419 S
[58] **Field of Search** .... 128/419 R, 419 S, 422, 128/653, 771, 732, 741, 746, 791, 804; 340/407

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,490,458	1/1970	Allison	128/421
3,751,605	8/1973	Michelson	128/1 R
3,951,134	4/1976	Malech	128/131
4,428,377	1/1984	Zollner et al.	128/419 R

FOREIGN PATENT DOCUMENTS

893311	2/1972	Canada	128/422
2811120	9/1978	Fed. Rep. of Germany	128/419 R
591196	1/1978	U.S.S.R.	128/419 R

OTHER PUBLICATIONS

Gerkin, G., "Electroencephalography & Clinical Neurophysiology", vol. 135, No. 6, Dec. 1973, pp. 652-653.
Frye et al., "Science", vol. 181, Jul. 27, 1973, pp. 356-358.
Bise, William, "Low Power Radio-Frequency and Microwave Effects on Human Electroencephalogram and Behavior", Physiol. Chem. & Physics 10 (1978).

Primary Examiner—William E. Kamm
Attorney, Agent, or Firm—Wegner & Bretschneider

[57] **ABSTRACT**

A method and apparatus for simulation of hearing in mammals by introduction of a plurality of microwaves into the region of the auditory cortex is shown and described. A microphone is used to transform sound signals into electrical signals which are in turn analyzed and processed to provide controls for generating a plurality of microwave signals at different frequencies. The multifrequency microwaves are then applied to the brain in the region of the auditory cortex. By this method sounds are perceived by the mammal which are representative of the original sound received by the microphone.

29 Claims, 7 Drawing Sheets

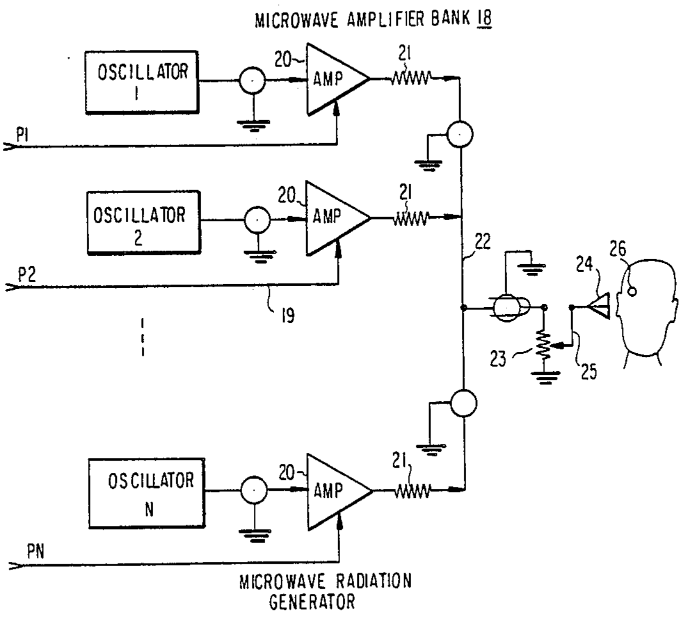

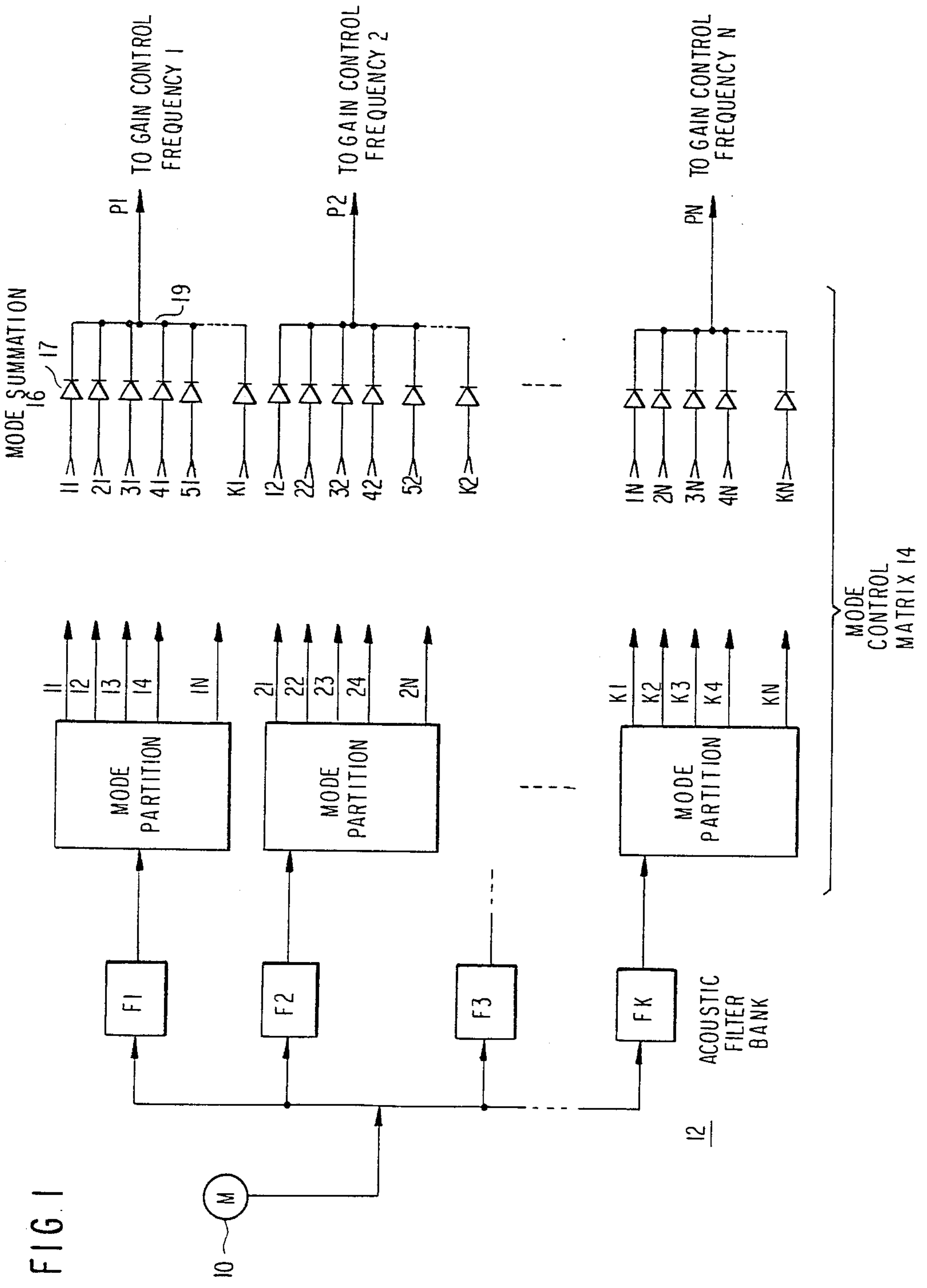

FIG. 1

M
10
ACOUSTIC FILTER BANK
12
F1
F2
F3
FK
MODE PARTITION
MODE PARTITION
MODE PARTITION
11
12
13
14
1N
21
22
23
24
2N
K1
K2
K3
K4
KN
MODE CONTROL MATRIX 14
MODE SUMMATION
16
17
19
11
21
31
41
51
K1
12
22
32
42
52
K2
1N
2N
3N
4N
KN
P1
P2
PN
TO GAIN CONTROL FREQUENCY 1
TO GAIN CONTROL FREQUENCY 2
TO GAIN CONTROL FREQUENCY N

FIG. 2

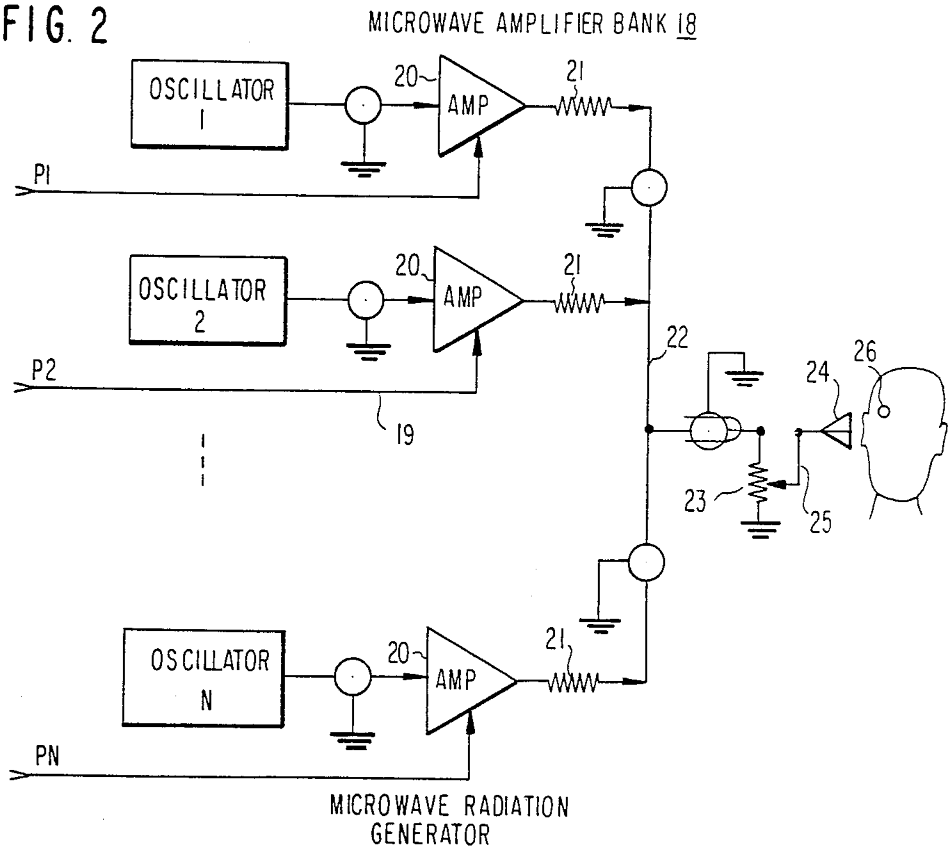
MICROWAVE AMPLIFIER BANK 18
OSCILLATOR 1
20
AMP
21
P1
OSCILLATOR 2
20
AMP
21
P2
19
22
24
26
23
25
OSCILLATOR N
20
AMP
21
PN
MICROWAVE RADIATION GENERATOR

FIG. 2a

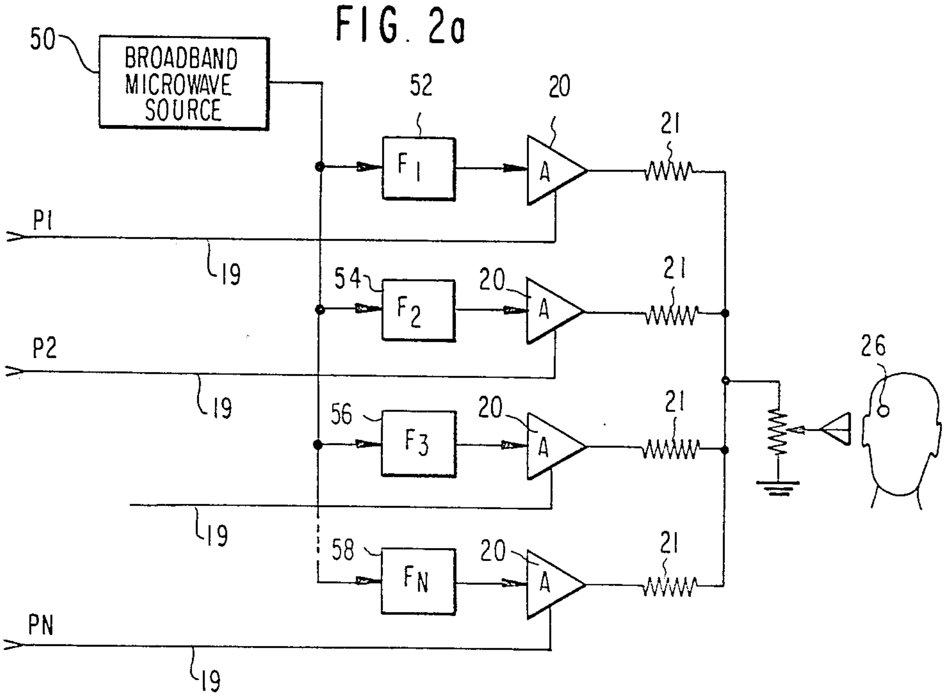
BROADBAND MICROWAVE SOURCE
F1
A
P1
F2
A
P2
F3
A
FN
A
PN

FIG. 5

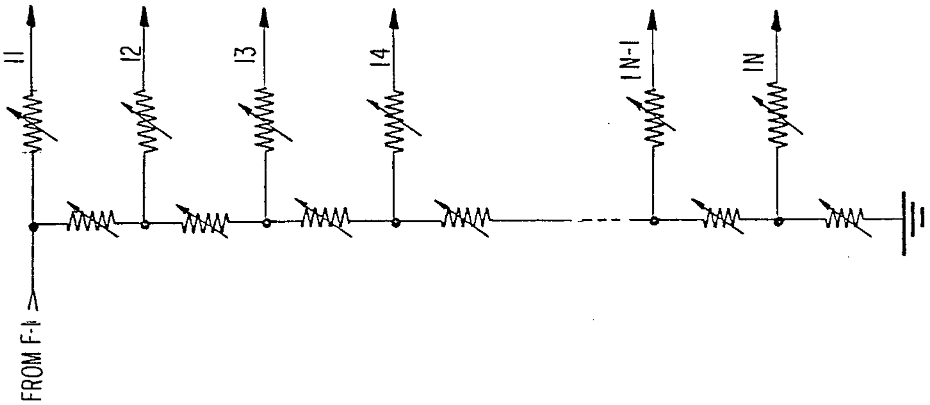
FROM F-1
I1
I2
I3
I4
IN-1
IN

FIG. 4

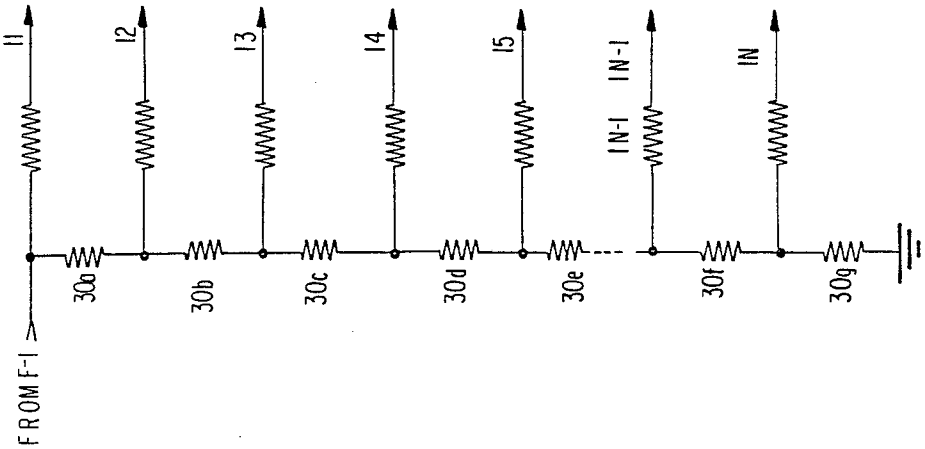
FROM F-1
I1
I2
I3
I4
I5
IN-1 IN-1
IN-1
IN
30a
30b
30c
30d
30e
30f
30g

FIG. 3

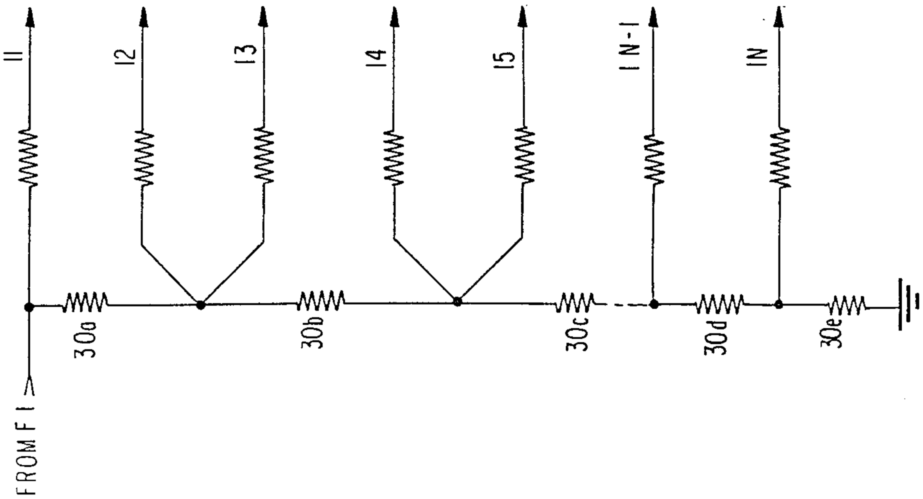
FROM F1
I1
I2
I3
I4
I5
IN-1
IN
30a
30b
30c
30d
30e

FIG. 6

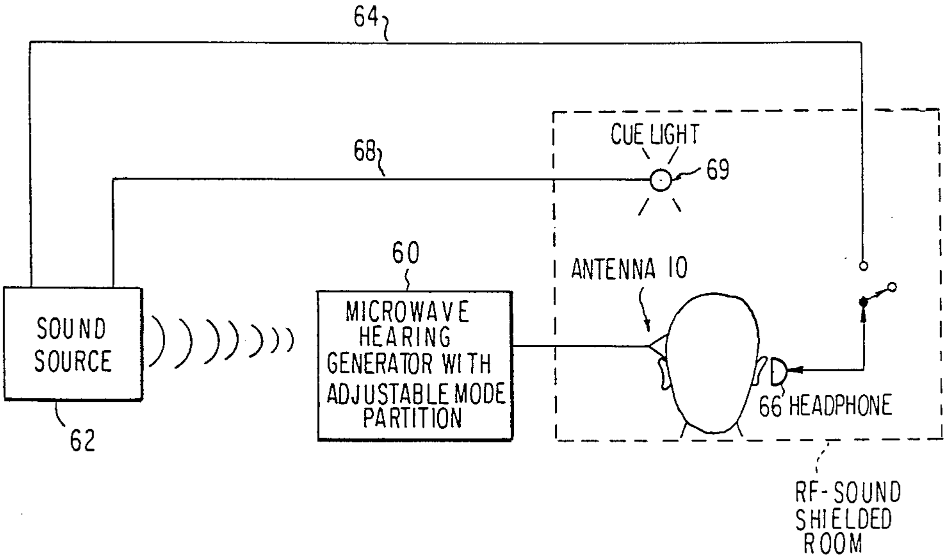
64
68
CUE LIGHT
69
60
ANTENNA 10
SOUND SOURCE
MICROWAVE HEARING GENERATOR WITH ADJUSTABLE MODE PARTITION
62
66 HEADPHONE
RF-SOUND SHIELDED ROOM

FIG. 7

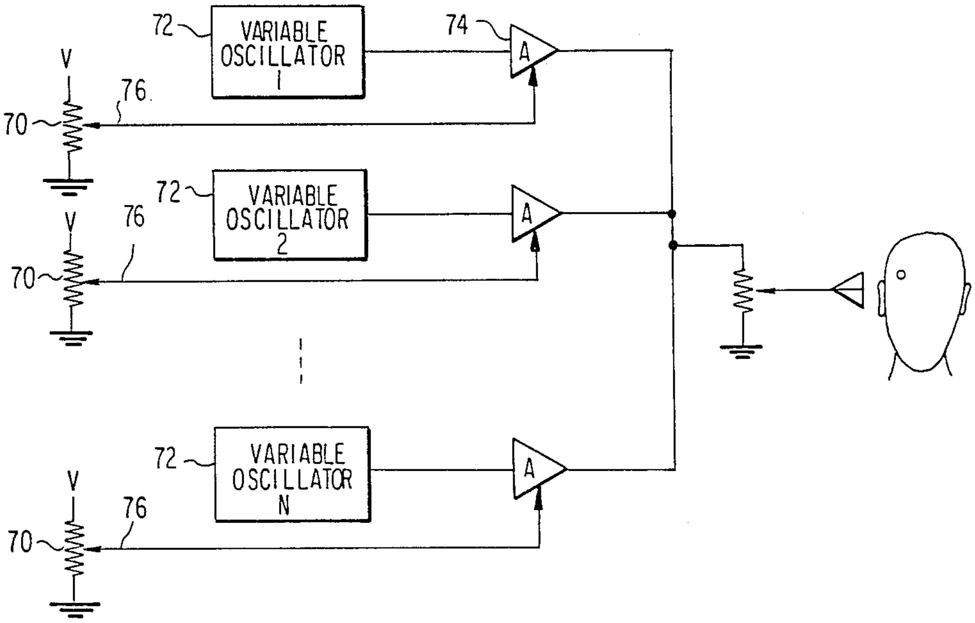
72
VARIABLE OSCILLATOR 1
74
A
V
76
70
72
VARIABLE OSCILLATOR 2
A
V
76
70
72
VARIABLE OSCILLATOR N
A
V
76
70

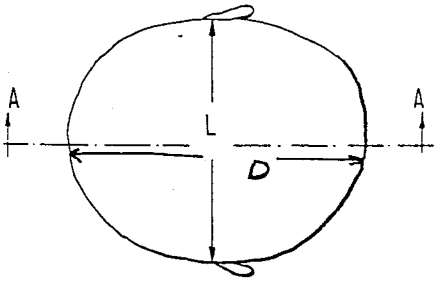

FIG.8

FIG.9

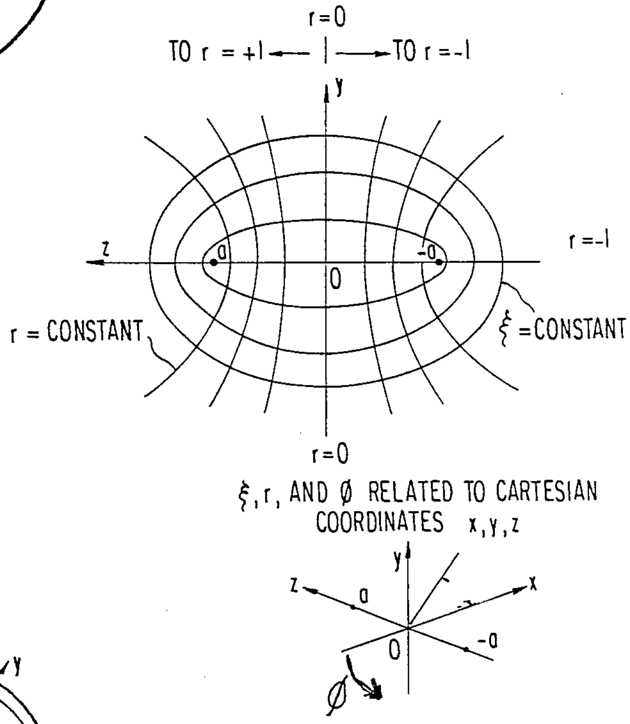
r=0
TO r = +1
TO r = -1
y
z
a
-a
0
r = -1
r = CONSTANT
ξ = CONSTANT
r=0
ξ, r, AND Ø RELATED TO CARTESIAN COORDINATES x, y, z
y
z
a
x
0
-a
Ø

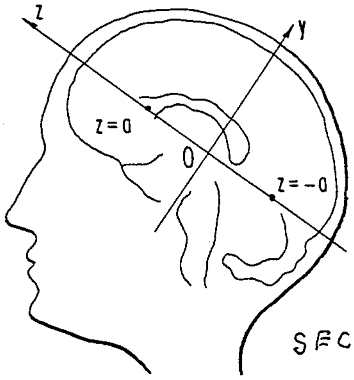

FIG.10

SECTION A-A

TRANSFORMATION EQUATIONS

$x = a(\xi^2 - 1)^{1/2} (1 - r^2)^{1/2} \cos \phi$

$y = -a(\xi^2 - 1)^{1/2} (1 - r^2)^{1/2} \sin \phi$

$z = a \xi r$

$1 \leq \xi \leq$

$-1 \leq \phi \leq +1$

$0 \leq \phi \leq 2\pi$

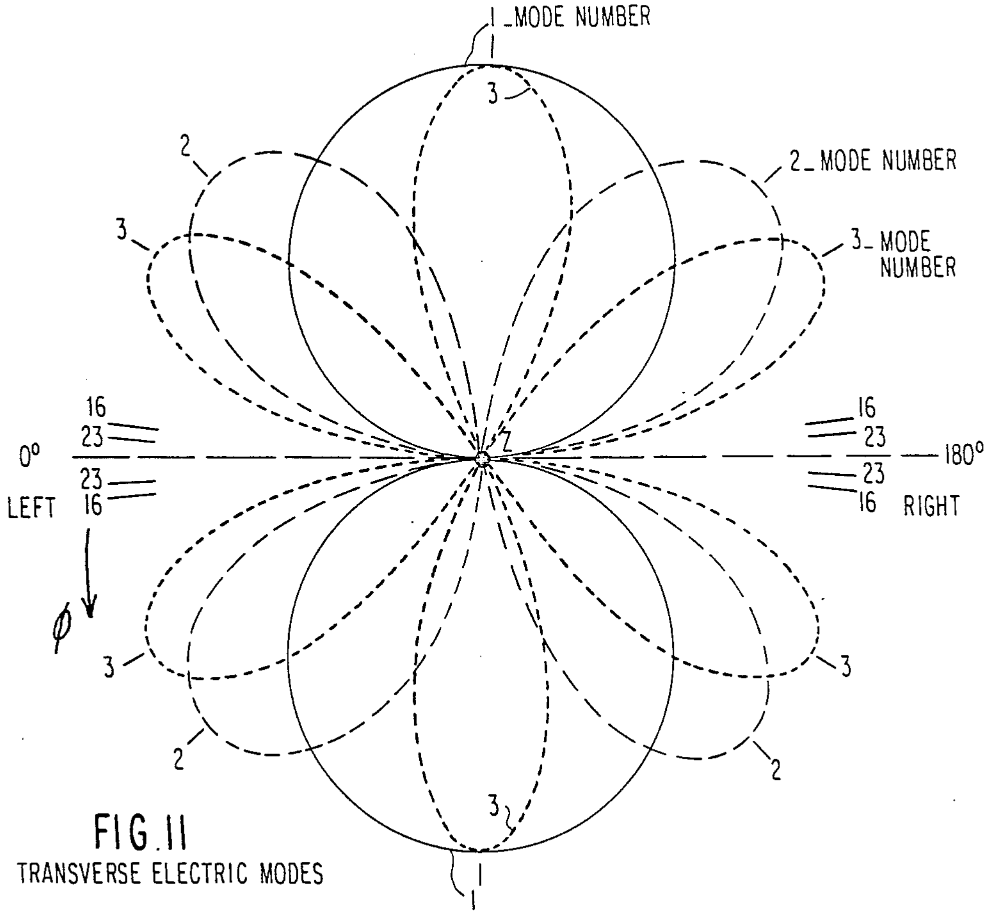
1_MODE NUMBER
1
3
2
2_MODE NUMBER
3
3_MODE NUMBER
16
23
0°
23
LEFT 16
φ
16
23
180°
23
16 RIGHT
Z
3
3
2
2
3
1
1

FIG.11
TRANSVERSE ELECTRIC MODES

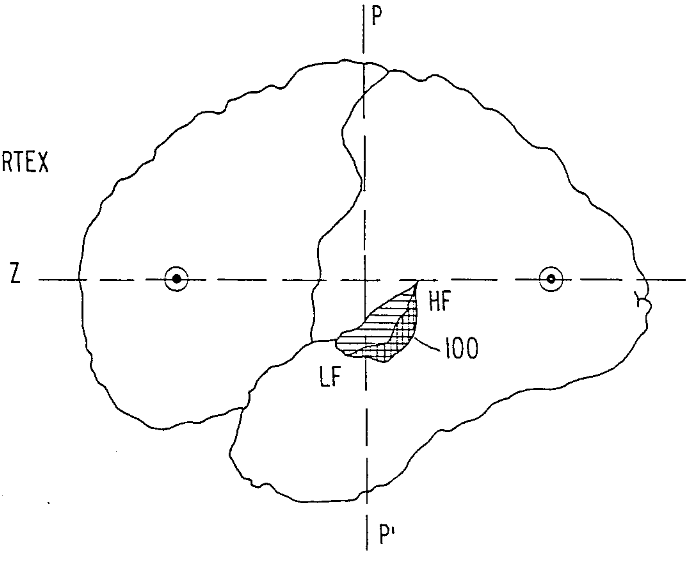
P
Z
HF
100
LF
P'

FIG.12
PRIMARY AUDITORY CORTEX

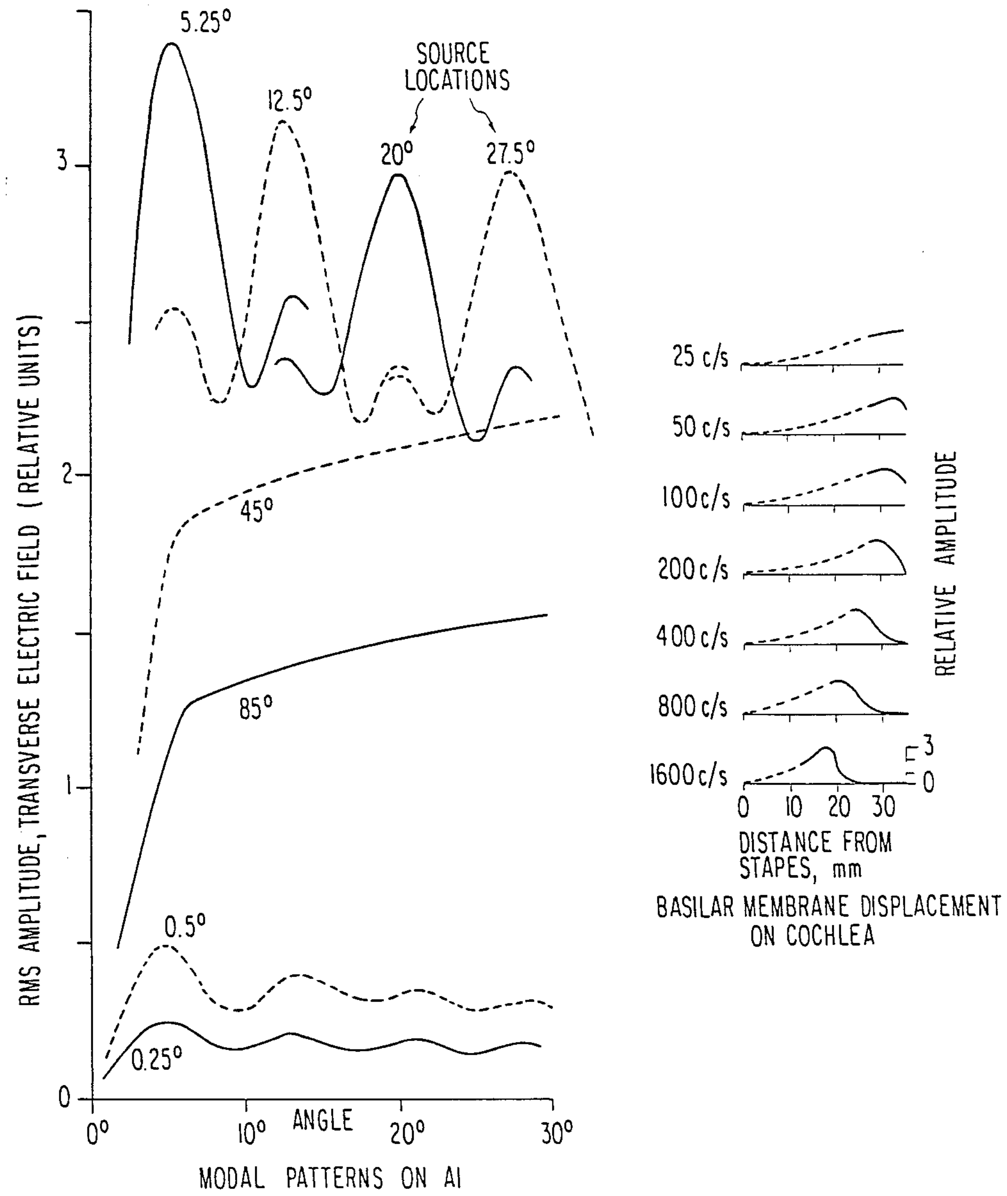
RMS AMPLITUDE, TRANSVERSE ELECTRIC FIELD (RELATIVE UNITS)
0
1
2
3
5.25°
12.5°
20°
27.5°
SOURCE LOCATIONS
45°
85°
0.5°
0.25°
0°
10°
ANGLE
20°
30°
MODAL PATTERNS ON AI
25 c/s
50 c/s
100c/s
200c/s
400c/s
800c/s
1600c/s
3
0
RELATIVE AMPLITUDE
0
10
20
30
DISTANCE FROM STAPES, mm
BASILAR MEMBRANE DISPLACEMENT ON COCHLEA

FIG.13

HEARING DEVICE

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates to devices for aiding of hearing in mammals. The invention is based upon the perception of sounds which is experienced in the brain when the brain is subjected to certain microwave radiation signals.

2. Description of the Prior Art

In prior art hearing devices for human beings, it is well known to amplify sounds to be heard and to apply the amplified sound signal to the ear of the person wearing the hearing aid. Hearing devices of this type are however limited to hearing disfunctions where there is no damage to the auditory nerve or to the auditory cortex. In the prior art, if there is damage to the auditory cortex or the auditory nerve, it cannot be corrected by the use of a hearing aid.

During World War II, individuals in the radiation path of certain radar installations observed clicks and buzzing sounds in response to the microwave radiation. It was through this early observation that it became known to the art that microwaves could cause a direct perception of sound within a human brain. These buzzing or clicking sounds however were not meaningful, and were not perception of sounds which could otherwise be heard by the receiver. This type of microwave radiation was not representative of any intelligible sound to be perceived. In such radar installations, there was never a sound which was generated which resulted in subsequent generation of microwave signals representative of that sound.

Since the early perception of buzzing and clicking, further research has been conducted into the microwave reaction of the brain. In an article entitled "Possible Microwave Mechanisms of the Mammalian Nervous System" by Philip L. Stocklin and Brain F. Stocklin, published in the TIT Journal of Life Sciences, Tower International Technomedical Institute, Inc. P.O. Box 4594, Philadelphia, Pa. (1979) there is disclosed a hypothesis that the mammalian brain generates and uses electro magnetic waves in the lower microwave frequency region as an integral part of the functioning of the central and peripheral nervous systems. This analysis is based primarily upon the potential energy of a protein integral in the neural membrane.

In an article by W. Bise entitled "Low Power Radio-Frequency and Microwave Effects On Human Electroencephalogram and Behavior", Physiol. Chemistry Phys. 10, 387 (1978), it is reported that there are significant effects upon the alert human EEG during radiation by low intensity CW microwave electromagnetic energy. Bise observed significant repeatable EEG effects for a subject during radiation at specific microwave frequencies.

SUMMARY OF THE INVENTION

Results of theoretical analysis of the physics of brain tissue and the brain/skull cavity, combined with experimentally-determined electromagnetic properties of mammalian brain tissue, indicate the physical necessity for the existence of electromagnetic standing waves, called modes in the living mammalian brain. The mode characteristics may be determined by two geometric properties of the brain; these are the cephalic index of the brain (its shape in prolate spheroidal coordinates)

and the semifocal distance of the brain (a measure of its size). It was concluded that estimation of brain cephalic index and semifocal distance using external skull measurements on subjects permits estimation of the subject's characteristic mode frequencies, which in turn will permit a mode by mode treatment of the data to simulate hearing.

This invention provides for sound perception by individuals who have impaired hearing resulting from ear damage, auditory nerve damage, and damage to the auditory cortex. This invention provides for simulation of microwave radiation which is normally produced by the auditory cortex. The simulated brain waves are introduced into the region of the auditory cortex and provide for perceived sounds on the part of the subject.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** shows the acoustic filter bank and mode control matrix portions of the hearing device of this invention.

FIG. **2** shows the microwave generation and antenna portion of the hearing device of this invention.

FIG. **3** shows a typical voltage divider network which may be used to provide mode partition.

FIG. **4** shows another voltage divider device which may be used to provide mode partition.

FIG. **5** shows a voltage divider to be used as a mode partition wherein each of the resistors is variable in order to provide adjustment of the voltage outputs.

FIG. **6** shows a modified hearing device which includes adjustable mode partitioning, and which is used to provide initial calibration of the hearing device.

FIG. **7** shows a group of variable oscillators and variable gain controls which are used to determine hearing characteristics of a particular subject.

FIG. **8** shows a top view of a human skull showing the lateral dimension.

FIG. **9** shows the relationship of the prolate spherical coordinate system to the cartesian system.

FIG. **10** shows a side view of a skull showing the medial plane of the head, section A—A.

FIG. **11** shows a plot of the transverse electric field amplitude versus primary mode number M.

FIG. **12** shows a left side view of the brain and auditory cortex.

FIG. **13** shows the total modal field versus angle for source location.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

This invention is based upon observations of the physical mechanism the mammalian brain uses to perceive acoustic vibrations. This observation is based in part upon neuro anatomical and other experimental evidence which relates to microwave brain stimulation and the perception of sounds.

It is has been observed that monochromatic acoustic stimuli (acoustic tones, or single tones) of different frequencies uniquely stimulate different regions of the cochlea. It has also been observed that there is a corresponding one to one relationship between the frequency of a monochromatic acoustic stimulus and the region of the auditory cortex neurally stimulated by the cochlear nerve under the physiologically normal conditions (tonotopicity).

It is has been observed that for an acoustic tone of a frequency which is at the lower end of the entire acoustical range perceivable by a person, that a thin lateral region ("Line") parallel to the medial axis of the brain and toward the inferior portion of the primary auditory cortex is stimulated. For an acoustic tone whose frequency is toward the high end of the entire perceivable acoustic range, a thin lateral region parallel to the medial axis and toward the superior portion of the primary auditory cortex is stimulated.

Neural stimulation results in the generation of a broad band of microwave photons by the change in rotational energy state of protons integral to the neuron membrane of the auditory cortex. The physical size and shape of the brain/skull cavity, together with the (semiconductor) properties (conductivity and dielectric constant) of the brain tissue provide an electromagnetic resonant cavity. Specific single frequencies are constructively reinforced so that a number of standing electromagnetic waves, each at its own single electromagnetic frequency in the microwave frequency region, are generated in the brain. Each such standing electromagnetic wave is called a characteristic mode of the brain/skull cavity.

Analysis in terms of prolate spheroidal wave functions indicates that transverse electric field components of these modes have maxima in the region of the auditory cortex. This analysis further shows that transverse electric field possess a variation of amplitude with angle in the angular plane (along the vertical dimension of the auditory cortex) and that is dependent only upon the primary mode number.

The auditory cortex in the normally functioning mammalian brain is a source of microwave modes. The auditory cortex generates these modes in accordance with the neural stimulation of the auditory cortex by the cochlear nerve. Mode weighting for any one acoustic tone stimulus is given by the amplitude of each mode along the line region of the auditory cortex which is neurally stimulated by that acoustic tone stimulus. A listing of mode weighting versus frequency of acoustic stimulus is called the mode matrix.

In this invention, the functions of the ear, the cochlear nerve, and the auditory cortex are simulated. Microwaves simulating the mode matrix are inserted directly into the region of the auditory cortex. By this insertion of simulated microwave modes, the normal operation of the entire natural hearing mechanism is simulated.

Referring now to FIG. 1 and FIG. 2 there is shown an apparatus which provides for induced perception of sound into a mammalian brain. This hearing device includes a microphone **10** which receives sounds, an acoustic filter bank **12** which separates the signals from the microphone into component frequencies, and a mode control matrix **14** which generates the mode signals which are used to control the intensity of microwave radiations which are injected into the skull cavity in the region of the auditory cortex.

The acoustic filter bank **12** consists of a bank of acoustic filters F1 through Fk which span the audible acoustic spectrum. These filters may be built from standard resistance, inductance, and capacitance components in accordance with well established practice. In the preferred embodiment there are 24 filters which correspond to the observed critical bandwidths of the human ear. In this preferred embodiment a typical list of filter parameters is given by Table 1 below:

TABLE I

Filter No.	Center Frequency (Hz)	Bandwidth (Hz)
1	50	less than 100
2	150	100
3	250	100
4	350	100
5	450	110
6	570	120
7	700	140
8	840	150
9	1,000	160
10	1,170	190
11	1,370	210
12	1,600	240
13	1,850	280
14	2,150	320
15	2,500	380
16	2,900	450
17	3,400	550
18	4,000	700
19	4,800	900
20	5,800	1,100
21	7,000	1,300
22	8,500	1,800
23	10,500	2,500
24	13,500	3,500

The rectifier outputs one through K are feed to K mode partition devices. The mode partitioning devices each have N outputs wherein N is the number of microwave oscillators used to generate the microwave radiation. The outputs **1** through N of each mode partition device is applied respectively to the inputs of each gain controlled amplifier of the microwave radiation generator. The function of the mode control matrix **14** is the control of the microwave amplifiers in the microwave amplifier bank **18**. In the preferred embodiment thus will be 24 outputs and 24 microwave frequency oscillators.

Connected to each microwave amplifier gain control line is a mode simulation device **16** which receives weighted mode signals from the mode partition devices **14**. Each mode simulation device consists of one through k lines and diodes **17** which are each connected to summing junction **19**. The diodes **17** provide for isolation from one mode partition device to the next. The diodes **17** prevent signals from one mode partition device from returning to the other mode partition devices which are also connected to the same summing junction of the mode summation device **16**. The diodes also serve a second function which is the rectification of the signals received from the acoustic filter bank by way of the mode partition devices. In this way each mode partition device output is rectified to produce a varying DC voltage with major frequency components of the order of 15 milliseconds or less. The voltage at the summation junction **19** is thus a slowly varying DC voltage.

The example mode partition devices are shown in greater detail in FIGS. **3**, **4**, and **5**. The mode partition devices are merely resistance networks which produce **1** through N output voltages which are predetermined divisions of the input signal from the acoustic filter associated with the mode partition device. FIG. **3** shows a mode partitioning device wherein several outputs are associated with each series resistor **30**. In the embodiment depicted in FIG. **4** there is an output associated with each series resistor only, and thus there are N series resistors, or the same number of series resistors as there are outputs. The values of the resistors in the mode partition resistor network are determined in accordance with the magnitudes of the frequency component from the acoustic filter bank **12** which is required at the summation point **19** or the gain control line for amplifiers **20**.

The microwave amplifier bank **18** consists of a plurality of microwave oscillators **1** through N each of which is connected to an amplifier **20**. Since the amplifiers **20** are gain controlled by the signals at summation junction **19**, the magnitude of the microwave output is controlled by the mode control matrix outputs F1 through F_n. In the preferred embodiment there are 24 amplifiers.

The leads from the microwave oscillators **1** through N to the amplifiers **20** are shielded to prevent cross talk from one oscillator to the next, and to prevent stray signals from reaching the user of the hearing device. The output impedance of amplifiers **20** should be 1000 ohms and this is indicated by resistor **21**. The outputs of amplifiers **20** are all connected to a summing junction **22**. The summing junction **22** is connected to a summing impedance **23** which is approximately 50 ohms. The relatively high amplifier output impedance **21** as compared to the relatively low summing impedance **23** provides minimization of cross talk between the amplifiers. Since the amplitude of the microwave signal needed at the antenna **24** is relatively small, there is no need to match the antenna and summing junction impedances to the amplifier **20** output impedances. Efficiency of the amplifiers **20** is not critical.

Level control of the signal at antenna **24** is controlled by pick off **25** which is connected to the summing impedance **23**. In this manner, the signal at antenna **24** can be varied from 0 (ground) to a value which is acceptable to the individual.

The antenna **24** is placed next to the subject's head and in the region of the subject's auditory cortex **26**. By placement of the antenna **24** in the region of the auditory cortex **26**, the microwave field which is generated simulates the microwave field which would be generated if the acoustic sounds were perceived with normal hearing and the auditory cortex was functioning normally.

In FIG. **2A** there is shown a second embodiment of the microwave radiation and generator portion of the hearing device. In this embodiment a broad band microwave source **50** generates microwave signals which are feed to filters **52** through **58** which select from the broad band radiation particular frequencies to be transmitted to the person. As in FIG. **2**, the amplifiers **20** receive signals on lines **19** from the mode control matrix. The signals on lines **19** provide the gain control for amplifiers **20**.

In FIG. **6** there is shown a modified microwave hearing generator **60** which includes a mode partition resistor divider network as depicted in FIG. **5**. Each of the mode partition voltage divider networks in this embodiment are individually adjustable for all of the resistances in the resistance network. FIG. **5** depicts a voltage division system wherein adjustment of the voltage partition resistors is provided for.

In FIG. **6**, the sound source **62** generates audible sounds which are received by the microphone of the microwave hearing generator **60**. In accordance with the operation described with respect to FIGS. **1** and **2**, microwave signals are generated at the antenna **10** in accordance with the redistribution provided by the mode control matrix as set forth in FIG. **5**.

The sound source **62** also produces a signal on line **64** which is received by a head phone **66**. The apparatus depicted in FIG. **6** is used to calibrate or fit a microwave hearing generator to a particular individual. Once the hearing generator is adjusted to the particular individual by adjustment of the variable resistors in the adjustable mode partition portion of the hearing generator, a second generator may be built using fixed value resistors in accordance with the adjusted values achieved in fitting the device to the particular subject. The sound produced by headphone **66** should be the same as a sound from the sound source **62** which is received by the microphone **10** in the microwave hearing generator **60**. In this way, the subject can make comparisons between the perceived sound from the hearing generator **60**, and the sound which is heard from headphone **66**. Sound source **62** also produces a signal on **68** which is feed to cue light **69**. Cue light **69** comes on whenever a sound is emitted from sound source **62** to the microwave generator **60**. In this manner, if the subject hears nothing, he will still be informed that a sound has been omitted and hence that he is indeed perceiving no sound from the microwave hearing generator **60**.

In FIG. **7** there is shown a modified microwave hearing generator which may be used to determine a subject's microwave mode frequencies. In this device, the acoustic filter bank and the mode control matrix have been removed and replaced by voltage level signal generated by potentiometers **70**. Also included are a plurality of variable frequency oscillators **72** which feed microwave amplifiers **74** which are gain controlled from the signal generated by potentiometers **70** and pick off arm **76**.

This modified microwave hearing generator is used to provide signals using one oscillator at a time. When an oscillator is turned on, the frequency is varied about the estimated value until a maximum acoustic perception by the subject is perceived. This perception however may consist of a buzzing or hissing sound rather than a tone because only one microwave frequency is being received. The first test of perception is to determine the subject's lowest modal frequency for audition ($M=1$). Once this modal frequency is obtained, the process is repeated for several higher modal frequencies and continued until no maximum acoustic perception occurs.

Another method of determination of a subject's modal frequencies is through anatomical estimation. This procedure is by measurement of the subject's cephalic index and the lateral dimensions of the skull. In this method, the shape is determined in prolate spheroidal coordinance.

Purely anatomical estimation of subject's modal frequencies is performed by first measuring the maximum lateral dimension (breadth) L FIG. **8**, of the subject's head together with the maximum dimension D (anterior to posterior) in the medial plane of the subject's head. D is the distance along Z axis as shown in FIG. **10**. The ratio L/D, called in anthropology the cephalic index, is monotonically related to the boundary value ξ_o defining the ellipsoidal surface approximating the interface between the brain and the skull in the prolate spheroidal coordinate system. ξ_o defines the shape of this interface; ξ_o and D together give an estimate of a, the semi-focal distance of the defining ellipsoid. Using ξ_o and a, together with known values of the conductivity and dielectric constants of brain tissue, those wavelengths are found for which the radial component of the electric field satisfies the boundary condition that it is zero at ξ_o.

These wavelengths are the wavelengths associated with the standing waves or modes; the corresponding frequencies are found by dividing the phase velocity of microwaves in brain tissue by each of the wavelengths.

A subject's microwave modal frequencies may also be determined by observing the effect of external microwave radiation upon the EEG. The frequency of the M equal 1 mode may then be used as a base point to estimate all other modal frequencies.

A typical example of such an estimation is where the subject is laterally irradiated with a monochromatic microwave field simultaneous with EEG measurement and the microwave frequency altered until a significant change occurs in the EEG, the lowest such frequency causing a significant EEG change is found. This is identified as the frequency of the $M=1$ mode, the lowest mode of importance in auditory perception. The purely anatomical estimation procedure (FIGS. **8**, **9**, **10**) is then performed and the ratio of each modal frequency to the $M=1$ modal frequency obtained. These ratios together with the experimentally-determined $M=1$ frequency are then used to estimate the frequencies of the mode numbers higher than 1. The prolate spheroidal coordinate system is shown in FIG. **9**. Along the lateral plane containing the x and y coordinates of FIG. **9**, the prolate spheroidal coordinate variable ϕ (angle) lies FIGS. **9** and **10**. Plots of the transverse electric field amplitude versus primary mode number m are shown in FIG. **11**. The equation is

$$E_{transverse}(m, \phi) = E_0 \sin(m\ \phi)$$

The "elevation view" FIG. **12**, of the brain from the left side, shows the primary auditory cortex **10**. The isotone lines and the high frequency region are toward the top of **100** and the low frequency region toward the bottom of **100**.

The formula I, set forth below is the formula for combining modes from an iso-tone line at $\phi=\phi j$ being excited to obtain the total modal field at some other angular location ϕ. For this formula, if we let $J=1$ (just one iso-tone single frequency acoustic stimulus line), then it can be shown that ALL modes (in general) must be used for any ONE tone.

FORMULA I
RMS TRANSVERSE ELECTRIC FIELD IN ANGULAR PLANE, $f(0)$

$$f(0) = \left[\sum_{m=1}^{M} \Bigg(\Bigg\{ \sin(m0) \cdot \sum_{j=1}^{J} e^{-(0-0j)/\Delta 0m} \sin(m0_j) \Bigg) \Bigg\}^{2} \right]^{\frac{1}{2}}$$

ϕ=ANGLE (0° LATERAL)
ϕ_j=LOCATION OF j-TH SOURCE (TOTAL NUMBER J)
$\Delta\phi_m$=ATTENUATION LENGTH (IN ANGLE) OF m-TH MODE
m=PRIMARY MODE NUMBER (HIGHEST MODE M)

FIG. **13** shows the resulting total modal field versus angle ϕ for source location ϕ at 5.25°, 12.5°, etc. With reference to the set of curves at the left top of this figure. A spacing of approximately 7.25° in ϕ corresponds to a tonal difference of about 1 octave. This conclusion is based on the side-lobes of pattern coming from $\phi=5.25°$, etc. The total filed (value on y-axis) falls considerably below the top curves for source locations well below 5.25° (toward the high acoustic stimulus end) and also as the source of frequency goes well above 30° (low frequency end). ϕ is plotted positive downward from 0° at lateral location as indicates in FIG. **11**.

Resistor weightings are obtained from the $|\sin(m[\phi-\phi j])|$, Formula I. The scale between acoustic frequency and ϕ must be set or estimated from experiment. Approximately $5.25 \pm 1°$ corresponds to a tonal stimulus at about 2 kHz (the most sensitive region of the ear) since this source location gives the highest electric field amplitude.

The apparatus of FIG. **7** may also be used to determine values for a hearing device which are required for a particular subject. Once the modal frequencies have been estimated, the device of FIG. **7** which includes variable microwave oscillators may be used to determine values for the oscillators which match the subject, and to determine resistance values associated with the mode partition devices of the mode control matrix.

In FIG. **7** manual control of the amplifier gain is achieved by potentiometers **76**. In this manner the amplifier gains are varied about the estimated settings for an acoustic tone stimulus in the region of two thousand Hertz (2 kHz) until maximum acoustic perception and a purest tone are achieved together. The term purest tone may also be described as the most pleasing acoustic perception by the subject. This process may be repeated at selected frequencies above and below 2 kHz. The selected frequencies correspond to regions of other acoustic filter center frequencies of the subject. When modal frequency (oscillator frequency) and gain set values (setting a potentiometer **76**) are noted, it is then possible to calculate fixed oscillator frequencies and control resistor values for the adjusted hearing device for this particular subject.

In the event the subject has no prior acoustic experience, that is deaf from birth, estimated resistor values must be used. Also, a complex acoustic stimulation test including language articulation and pairs of harmonically related tones may be developed to maximize the match of the hearing device parameters for those of this particular subject.

Typical components for use in this invention include commercially available high fidelity microphones which have a range of 50 Hz to 15 kHz with plus or minus 3 dB variation.

The audio filters to be used with the acoustic filter bank **12** are constructed in a conventional manner, and have Q values of about 6. The filters may also be designed with 3 dB down points (½ the bandwidth away from the center frequency) occurring at adjacent center frequency locations.

The diodes **17** in the mode control matrix which provide isolation between the mode partition circuits are commercially available diodes in the audio range.

The microwave oscillators **1** through N and the microwave amplifiers **20** are constructed with available microwave transistors which can be configured either as oscillators or amplifiers. Examples of the transistors are GaAsFET field effect transistors by Hewlitt Packard known as the HFET series or silicone bipolar transistors by Hewlitt Packard known as the HXTR series.

All the cable between the oscillators, the microwave amplifiers, and the antenna should be constructed with either single or double shielded coaxial cable.

The antenna **24** for directing microwave signals to the audio cortex **26** should be approximately the size of the auditory cortex. A typical size would be one and one half CM high and one half to one CM wide. The antenna as shown is located over the left auditory cortex, but the right may also be used. Since the characteristic impedance of the brain tissue at these microwave frequencies is close to 50 ohms, efficient transmission by commercially available standard 50 ohm coax is possible.

The invention has been described in reference to the preferred embodiments. It is, however, to be understood that other advantages, features, and embodiments may be within the scope of this invention as defined in the appended claims.

What is claimed is:

1. A sound perception device for providing induced perception of sound into a mammalian brain comprising in combination:

means for generating microwave radiation which is representative of a sound to be perceived, said means for generating including means for generating a simultaneous plurality of microwave radiation frequencies and means for adjusting the amplitude of said microwave radiation frequencies in accordance with the sound to be perceived; and

antenna means located in the region of the auditory cortex of said mammalian brain for transmitting said microwave energy into the auditory cortex region of said brain.

2. A hearing device for perception of sounds comprising in combination:

means for generating a signal representative of sounds;

means for analyzing said signal representative of said sounds having an output;

means for generating a plurality of microwave signals having different frequencies having a input connected to said output of said means for analyzing said signals, having an output;

means for applying said plurality of microwave signals to the head of a subject, and

whereby the subject perceives sounds which are representative of said sounds.

3. The apparatus in accordance with claim 2 wherein said means for generating a signal is a microphone for detecting sound waves.

4. The apparatus in accordance with claim 2 wherein said means for applying said plurality of microwave signals is an antenna.

5. The apparatus in accordance with claim 4 wherein said antenna is placed in the region of the auditory cortex of the subject.

6. The apparatus in accordance with claim 2 wherein the subject is a human being.

7. The apparatus in accordance with claim 2 wherein said means for analyzing said signal comprises:

an acoustic filter bank for dividing said sounds into a plurality of component frequencies; and

a mode control matrix means for providing control signals which are weighted in accordance with said plurality of component frequencies, having an output connected to said means for generating a plurality of microwave signal inputs.

8. The apparatus in accordance with claim 7 wherein said acoustic filter bank includes a plurality of audio frequency filters.

9. The apparatus in accordance with claim 8 wherein said audio frequency filters provide a plurality of output frequencies having amplitudes which are a function of said signal representative of sounds.

10. The apparatus in accordance with claim 9 wherein said amplitudes are the weighted in accordance with transform function of the signal representative of sounds.

11. The apparatus in accordance with claim 7 wherein said mode control matrix device includes a voltage divider connected to each of said plurality of said audio frequency filters.

12. The apparatus in accordance with claim 11 wherein each of said voltage dividers has a plurality of outputs which are connected in circuit to said means for generating a plurality of microwave signals.

13. The apparatus in accordance with claim 2 wherein said means for generating a plurality of microwave signals comprises a plurality of microwave generators each having a different frequency and means for controlling the output amplitude of each of said generators.

14. The apparatus in accordance with claims 2 wherein said means for generating a plurality of microwave signals comprises a broad band microwave source and a plurality of filters.

15. The apparatus in accordance with claim 13 wherein said generators each comprise a microwave signal source and a gain controlled microwave amplifier.

16. The apparatus in accordance with claim 13 wherein said means for analyzing output is connected to said means for controlling microwave amplifier output amplitudes.

17. The apparatus in accordance with claim 13 wherein analyzing includes K audio frequency filters.

18. The apparatus in accordance with claim 17 wherein there are N microwave generators.

19. The apparatus in accordance with claim 18 including a mode partitioning means which provides N outputs for each of said K audio frequency filters.

20. The apparatus in accordance with claim 19 wherein said N amplifiers each have K inputs from said mode partitioning means.

21. The apparatus in accordance with claim 20 wherein said N amplifiers have K inputs less the mode partitioning means outputs which are so small that they may be omitted.

22. The apparatus in accordance with claim 20 wherein said mode partitioning output device outputs each include a diode connected to each microwave amplifier gain control to provide isolation between all outputs.

23. The apparatus in accordance with claim 20 wherein said K audio frequency filters are chosen to correspond to the critical bandwidths of the human ear.

24. The apparatus in accordance with claim 20 wherein said N microwave generators are each adjustable in frequency output.

25. The apparatus in accordance with claim 18 wherein the frequency of each N microwave generators is determined by anatomical estimation.

26. The apparatus in accordance with claim 18 wherein the frequency of the lowest frequency microwave generator is chosen by determination of the effect of external microwave generation on the EEG of the subject.

27. The apparatus in accordance with claim 18 wherein the frequency of each of said N microwave generators corresponds to the subject's microwave modal frequencies.

28. The apparatus in accordance with claim **27** wherein the subject's modal frequencies are determined by measurement of the subject's cephalic index and the lateral dimensions of the skull.

29. The apparatus in accordance with claim **28** wherein the subject's lowest modal frequency is determined by varying the frequency of the lowest frequency microwave generator about the estimated value until a maximum acoustic perception is obtained by the subject.

* * * * *

rexresearch.com

Immortality

Leonard Orr's Rebirthing Technique, & 20 patented methods claiming immortality ...

Leonard ORR : Immortality --

https://www.youtube.com/watch?v=1g60sx-idWM
Leonard Orr on Immortality -- Interview with Alan Steinfeld

https://www.youtube.com/watch?v=1FjG5c1X-o0
The Yoga of Physical Immortality, Immortal Yogis and Babaji by Leonard Orr

https://www.youtube.com/watch?v=5fAX8y5yl2w
On Physical Immortality by Leonard Orr

https://www.youtube.com/watch?v=AErUwYDXUmw
Forever Young: Russian scientists reveal secret of eternal life (RT Documentary)

http://rebirthingbreathwork.com/

Physical Immortality – Two Powerful Words

by

Leonard Orr

A mind expanded by a new idea can never return to its former size. Two common words in everyone's vocabulary when combined into one idea become one of the most powerful ideas in the universe. Physical is a very common word. Immortality is also a common word. When you combine them, you get the idea of physical immortality. I've been exposing people to the idea of physical immortality for over 30 years.

Physical immortality means victory over physical death. It means body and mind mastery. Our Spirit is already an immortal master. This simple thought of physical immortality makes us realize that although death may be popular, it is not inevitable. Physical death is optional. The whole purpose of Jesus Christ was to teach us this. But Jesus was not the first nor the last human to conquer death. The Hebrew Bible is about five main immortal humans: Elijah, Moses, Melchizadek, and Enoch. In the beginning of Bible history, everyone lived 500 to 1000 years. India has its immortal yogis. Every continent has its favorite immortals, including native Americans. I've made this one of my favorite research projects for 37 years. Why do Christians die? Why? If Jesus died to save us from sin and death?

The founders and heroes and heroines of most religions who conquered death do not have a local address on planet Earth. When I learned about the immortal yogis who do, I was so excited I couldn't wait to meet them. Since my first trip to India in 1977, I have met over 8 immortals who each have maintained their body for at least 300 years. People have to be at least 300 to qualify for my research. I know of many more that I haven't had a chance to meet yet.

I've met several who are over 2000. These are yogis who have maintained a local address on planet Earth for over 2000 years. I've written books about two of them: Babaji and Bhartriji. People often tell me that after my lecture or book in which the idea of physical immortality got their attention, they couldn't stop thinking about it for several years. It makes so much sense. How could people be blinded by the death is inevitable belief system for so long?

After teaching the truth about physical immortality for over 30 years, I've concluded that the reason people don't like to know about it is because it I involves a little spiritual work and personal changes. Physical immortality is the natural result of a mind filled with wisdom and aliveness.

The three basic causes of death are:

1. ignorance

2. emotional energy pollution

3. poor diet.

Therefore, to master death you have to do three things:

1. Master the idea of physical immortality, which involves living from Spirit, our natural divinity.

2. Unravel our own death urge, which we received from our parents and encourage all the people in our environment to do the same. Learn how to keep our energy body & emotional mind clean and balanced through spiritual purification practices.

3. Learn the power of nutrition and fasting and the other simple and pleasurable habits that give us body mastery. Becoming a vegetarian is basic!

Mastering spiritual practices is less difficult than living a miserable life that leads to the grave. In fact, true spiritual practices are the ultimate pleasures and enrich all the pleasures of living. As I studied the lives of my immortal yogi friends, I began to realize that becoming an immortal yogi is not difficult; it just takes a different focus. People die because they kill themselves.

The greatest enemy of mankind has been and always will be ignorance. And the most evil form of ignorance is self-righteousness – thinking we know enough. Self-righteous people are murdering their own natural divinity and are guilty of murdering the natural divinity of everyone around them. It is the sin of eating of the fruit of the tree of knowledge and naturally yields the personal doom of death. The Bible says, "God is not willing for any to perish, but that all should come to a knowledge of repentance." Repentance means to change our mind. Repentance is neither difficult nor unpleasurable. In fact, it is exciting, inspiring, and produces pleasure, bliss, and ecstasy. If God made the path to eternal life difficult, He/She would be a sadist Who has doomed us to eternal misery. The only thing difficult about it is giving up your addiction to misery. Changing gives us the immense satisfaction of accomplishment.

Simple spiritual purification practices like bathing twice per day, conscious energy breathing, fire purification, an exercise system, a conscious diet, loving relationships, and mastering a mantra are the basic common denominators of the lifestyle of all the immortal masters I have met. These spiritual practices with earth, air, water, and fire can clean the mind faster than the mind can clean the mind. They are healing magic. In the process of mastering these, we must develop a balance between solitude and service.

I've written many books about physical immortality and the practical principles of conquering death and I recommend that you read them all and use my tapes repetitively if you are interested in personal mastery. But every year I try to put this simple truth into one page, in a clearer and easier way for people to understand.

Anyone can read or have read to them one page and have their mind expanded to receive the joy that this great idea brings. It unshackles your creativity and releases enormous reserves of energy and intelligence.

Is physical immortality realistic? Yes, if we honor the following. The three basic secrets to physical immortality are:

1. to stay tuned into God – Infinite Energy, our Source – all the time. Have pure thoughts and Life.

2. to heal the death urge. Healing the death urge can take us through the most ugly emotions of suicidal and homicidal thoughts – we have to survive the illusions in our own mind and go for wisdom and bliss.

3. to keep our energy body clean and balanced with the simple and pleasurable spiritual purification practices with mind, earth, air, water, fire, and love.

When we eat food we have to clean the dishes, etc. When our home gets dirty, we have to clean it. We also have to wash our clothes. When our energy body (also called emotional body) gets dirty through absorbing emotional energy pollution as we participate with people or in our self-healing process, we have to clean ourselves emotionally and spiritually. It is a daily process. Spiritual cleanliness is Energy and Peace.

If we stay ahead of the pollution – cleansing process, we become more alive every day. If we lose the spiritual purification game, we become more dead every day. Victory or victim?

One of the indicators is the guru belly phenomena. The solar plexus chakra is the social chakra and emotional energy pollution can be collected physically in the form of fat around the navel – it is sometimes called psychic fat.

Another indicator is stiffness. An over loaded body gets stiffer and stiffer every day and every year until we become a stiff – and are buried. When we absorb other people's death urge, we can get depressed, feel suicidal, become prone to anger, be grouchy, or have violent dreams. Victim consciousness in all its forms is a sign of death urge.

I'll finish this page with an idea which is a million times greater than physical immortality, and is the source of personal mastery or misery and death. It is this: 'Energy becomes what it thinks about'. Your personal energy is the Eternal Spirit – God Within – your natural divinity.

This Infinite Eternal Energy is helpless and powerless without thought and yet it is the power of thought. Your unlimited potential is only potential until it has been activated and directed by your thoughts. Energy is directed by thought.

Infinite Power directed by thoughts of weakness can produce infinite weakness. Infinite power directed by wholesome and healthy thoughts produce infinite wholesomeness, health, and safety – beyond safety – peace and pleasure. Thoughts of death produce death. Thoughts of Eternal Life produce physical immortality and body mastery. The body is Energy materialized by thought; not passive thinking, but active thinking; thought in motion. Your physical body and emotional body have been built up by millions of thoughts in action. To be a master, you now have to catch the thoughts and change them with the kind of actions that produce the results you like. We are all scientists and our body is the laboratory. Various nutrients have been created by thinking beings to produce certain effects in the body. Some of these cause and effect relationships are simple and some are quite complex. And simple fasting is the direct route to the purity of Infinite Energy. However, it is also possible to starve to death for spiritual purposes. We have to learn balance and common sense. More people die from overeating today than from starvation.

Repentance is fun! Changing our minds is the great, eternal adventure of all intelligent divine beings.

Playing in your Spirit Energy, your Divine Mind, and your Infinitely Beautiful Body is what Eternal Life is all about.

Have fun!

Franz BARDON : Hermetic Magic & Quabbalah

Six of the genii introduced by the Hermetic adept Franz Bardon teach methods of extreme life extension :

Earth Zone : Faluna (4* Taurus) ... Kirek (9* Virgo) ... Siges (7* Aquarius) ...

Mercury Zone : Lekabel (31) ... Rehael (39) ... Mehiel (64)

Patents for Immortality

JP2013240722
ASTRAL PROJECTION APPLIED IMMORTALITY DEVICE 6 AS MIND TRANSFER (MIND UPLOADING)

Inventor:KAJISA ISAO
[0001]

The present invention relates to a dedicated human humanoid cyborg robot with the spirit of the transfer-body application immortality device (spirit upload).
[0002]

Conventional, not the technology transfer spirit (spirit upload), out-of-body application immortality device technology had but spirit transfer of its application on the (spirit upload) and special remembrance in motion, humanoid humanoid cyborg robot that becomes the power source It was not.
[0003]

"Nihon Keizai Shimbun electronic edition" August 5, 2013, "the Nihon Keizai Shimbun electronic edition" January 2, 2013 "separate Nikkei Science perceived phantom" October 15, 2010, 114 pp -117 The Lancet 2001 year December 15, pp. 2039 2045 "Nature" September 19, 2002, 270 pages from 269 pages "free encyclopedia Wikipedia" mind uploading
[0004]

Conventional, extracorporeal withdrawal application technology related to immortality device had but spirit transfer of its application (spirit upload) and dedicated praying to the movement, and to realize it by the latest technology because there was no human humanoid cyborg robot that becomes the power source.
[0005]

The infrared sensing to humanoid humanoid cyborg robot to move in-body and consciousness energy (spirit) the spirit and transfer (spirit upload) praying with ketamine taking Hemishinku technology or right brain right corner gyrus electrode stimulation and right brain auditory cortex near the electrode stimulation and anesthetic can all flesh human similar behavior in that you have loaded fixed that vessel, it is assumed that even as a power source for the human humanoid cyborg robot.
[0006]

Surely enables action by the same storage consciousness and when of the-body can pray immortality of the body in human humanoid cyborg robot that moves and body with ketamine dose of anesthetic, that the spirit transfer (spirit upload) and a power source It uses a heat current of consciousness energy (spirit).
[0007]

Below, it described in the series logically sentence when the embodiment of the present invention. First, it is possible to reliably-body with ketamine taking Hemishinku technology or right brain right corner gyrus electrode stimulation or right brain auditory cortex near the electrode stimulation or anesthetic, out consciousness energy (spirit) to the outside of the body, around the senility and the sick and the elderly, In infrared sensor because the above-body was conscious energy (spirit) whole body moves in the electrical signals of the brain enters Prepare a humanoid humanoid cyborg robot, consciousness energy to the human humanoid cyborg robot (spirit) can be sensed by the infrared the humanoid by inducing the humanoid cyborg robot fixed (spirit transfer, spirit upload) and, the conscious energy (spirit) and power also becomes a source of power humanoid humanoid cyborg robot, from the portion corresponding to the brain of the consciousness energy all even after it moved from the human body to the human humanoid cyborg robot that have in mind the human body similar to this consciousness energy in an electric signal entering the humanoid humanoid cyborg robot of the brain (mental) and can do the same thing as human beings, power source it becomes a power supply, in-body dedicated human humanoid cyborg robot, mind uploading with the feature as more-body application immortality device as a (mental upload) 6.
[0008]

It is carried out as described above in the form and non-patent literature.
[0009]

Enables the realization of immortality by-body dedicated human humanoid cyborg robot as a spirit transfer (spirit upload), the availability of the industry in an aging society is I is understood to be high.

JP2013233468
ASTRAL PROJECTION-APPLIED IMMORTALITY DEVICE AS MEDICAL DEVICE

Inventor:KAJISA ISAO

[0001]

The present invention relates to a dedicated human humanoid cyborg robot-body application immortality device.
[0002]

Traditionally, there were technology related to ketamine dose-body application immortality preparation of human iPS rejuvenation body system drugs move in remembrance of its dedicated, there was no human humanoid cyborg robot that becomes the power source.
[0003]

"Nihon Keizai Shimbun electronic edition" August 5, 2013, "the Nihon Keizai Shimbun electronic edition" January 2, 2013 "separate Nikkei Science perceived phantom" October 15, 2010, 114 pp -117 The Lancet 2001 year December 15, pp. 2045 from page 2039 "Nature" September 19, 2002, 270 pages from 269 pages
[0004]

In the past, technology was related to ketamine dose-body application immortality preparation of human iPS rejuvenation body system drugs move in remembrance of the only, to be implemented by state-of-the-art technology because there was no human humanoid cyborg robot that becomes the power source.
[0005]

body human same by Hemishinku technology or right brain right corner gyrus to electrode stimulation and was placed in the right brain auditory cortex near the electrode stimulation and infrared sensor to humanoid humanoid cyborg robot to move in remembrance consciousness energy that was out of body with ketamine dose of anesthetic fixed behavior can be any, it is assumed that also a power source of humanoid humanoid cyborg robots.
[0006]

Surely enables action by the same storage consciousness and when of the-body can pray immortality of the body in human humanoid cyborg robot that moves and body with ketamine dose of anesthetic, to use the heat current of the consciousness energy to a power source .
[0007]

Below, it described in the series logically sentence when the embodiment of the present invention. First, Hemishinku technology or can be reliably-body with ketamine dose of right brain right corner gyrus electrode stimulation or right brain auditory cortex near the electrode stimulation or anesthetic, issued a consciousness energy to the outside of the body, around the senility and the sick and the elderly, said of body a humanoid humanoid cyborg robot is prepared consciousness energy that was systemic is in electrical signals of the brain move to enter, consciousness energy to the human humanoid cyborg robot was induced in the human humanoid cyborg robot with infrared sensor because it can sense the infrared fixed Te, the conscious energy and power also becomes a source of power humanoid humanoid cyborg robot, the human body similar to this consciousness energy in electrical signals of the brain went into the humanoid humanoid cyborg robot from the portion corresponding to the brain of the consciousness energy all even after he moved from the human body to the human humanoid cyborg robot of have in mind that it can do the same thing as human beings, become the power source of power, medical equipment with in-body dedicated human humanoid cyborg robot, the as features more -body application immortality device 5 as.

[0008]

It is carried out as described above in the form and non-patent literature.
[0009]

It enables the realization of immortality by-body dedicated human humanoid cyborg robot as medical equipment, industrial applicability in the aging society I is understood to be high.

JP2013231088
AGED ELDERLY PERSON HUMAN iPS 6 FACTOR COATING INITIALIZATION TELOMERE REJUVENATION
UTILIZING ELDERLY PERSON'S BRAIN PHOTOELECTRODE STIMULATION WHOLE MEMORY RECALL ORAL IMMORTALITY DRUG

Inventor: KAJISA ISAO

[0001]
The present invention relates mainly elderly old man case in improvement of human iPS6 factor cell initialization old man brain total storage resurrection use oral immortality drugs 3.

[0002]
Conventional, technical elderly for the elderly on human iPS cell initialization 6 factor use oral immortality drugs was small.

[0003]
"Mainichi" October 11, 2012, page 1.
"Mainichi" December 4, 2012, page 12.
"Nature online version of" March 2012 22 date, "Yomiuri Shimbun" November 5, 2011

[0004]
Conventional, technical elderly for the elderly on human iPS cell initialization 6 factor use oral immortality drugs thereby improved achieved through state-of-the-art technology because it was less.

[0005]
The elderly old man human iPS cell initialization 6 factor to perform the resurrection coating or basket put the elderly brain total storage put into plasmid in the hippocampus and cerebral light electrode stimulation.

[0006]
It kept the body of immortality.

[0007]
The following, it will be described in the series in the sentence when the embodiment of the present invention.
Elderly elderly person iPS6 factor a is initialized Oct3 / 4 needed to, Sox2, Klf4, L-Myc, Lin28, prepared Nanog, coating these six factors as taking into plasmid not destroyed by stomach acid and heat and or, is placed in a "cage" of a protein, conventionally, although the plasmid was placed in the blood, which was a simplification in that drinking from the mouth by the oral expression, the whole cells in elderly nursing Thus It is initialized, telomeres also rejuvenated as it becomes oral immortality drug return to its original length but in the initialization of the brain cell memory also revived either, the hippocampus and cerebral in the brain light stimulus or the electrode stimulation

to all storage revived old man brain also becomes younger, more oral immortality drugs 6 resurrection elderly elderly person iPS6 factor coating initialization telomere rejuvenation use old man brain light electrode stimulation total storage with features such as.
[0008]

It is carried out as described above in the form and non-patent literature.
[0009]

The elderly elderly person iPS6 factor coating initialization telomere rejuvenation use old man brain light electrode stimulation total storage resurrection oral immortality drugs 6 in the advance is aging made a big progress in improved techniques of human iPS cell initialization 3 gene use oral immortality drugs the industrial applicability of the demand of this medicine is understood to be high.

JP2013226466
IMMORTALITY DEVICE APPLYING ASTRAL PROJECTION AS MEDICAL DEVICE

Inventor: KAJISA ISAO

[No description or copy of patent]

JP2013151554
ORAL DRUG FOR ETERNAL YOUTH AND IMMORTALITY PREVENTING AGING INFLAMMATION BASED ON TELOMERE REJUVENATION CAUSED BY REDIFFERENTIATED T CELL OF HUMAN IPS

[0001]
The present invention relates to a T cell re-differentiation telomere rejuvenation aging inflammation stop oral immortality drug 4 of human iPS.

[0002]
Traditionally, technology related to human iPS cell initialization use oral immortality drugs was small.

[0003]
"Mainichi" January 4, 2013, page 3.
"1000 life" July 25, 2012, Hayakawa Jonathan Weiner.
"Aging could heal," March 20, 2013, Shueisha Shinsho, Makoto Goto.

[0004]
Traditionally, technology related to human iPS cell initialization use oral immortality drugs can be improved achieved through state-of-the-art technology because it was less.

[0005]
After T-cell initialization of human iPS re-differentiated to lengthen the telomere, a coating or basket put oral eliminate the garbage of aging inflammation stop and cells.

[0006]
It kept the body of immortality.

[0007]
The following, it will be described in the series in the sentence when the embodiment of the present invention.
Re-differentiation of T cells of the human iPS it shows signs of long stretched rejuvenation telomeres.
Aging is to eliminate the waste of the cells of the seven causes of cure outdated inflammation stop that aging in the re-differentiation T cells because the inflammation that in the garbage of somatic cell aging in this re-differentiation T cells.
The iPS cells are coated with a rejuvenation re-differentiation T cells so that it is not destroyed by stomach acid and heat so that growth, or is placed in a "basket" of the protein.
Conventionally, although the regenerated T cells were placed in the cells, which was a simplified that drinking from the mouth by the oral type.
In this way, the whole cells are initialized to eliminate waste of cells, most of the blood is to achieve immortality by feeding the regenerated T cells are in the T-cells into the somatic cells from the blood.
Rejuvenation T cell re-differentiation telomere of human iPS with features such as more aging inflammation stop oral immortality drugs 4.

[0008]
It is performed as in non-patent document mentioned above.

[0009]
Human iPS of T cell regeneration telomere rejuvenation aging inflammation stop oral immortality drug 4 on the industry demand for this drug in the advance is aging made a big progress in improved techniques of human iPS cell initialization 3 gene use oral immortality drugs The availability of it is understood to be high.

LOL..

CA2813395

ULTIMATE CYBER SECURITY SWITCH, 9/11-EQUIVALENT EVENT-PROOF SWITCH, AND OTHERS

Inventor: WU WEIMIN [CA]

Ever since the inception of internet, hackers from different countries have been successfully defeating many governments and more than 70 corporations around the world in cyber warfare: as, the USA, Canada, Taiwan, South Korea, Japan, and many others according to the news broadcasted by CNN, CBC, etc. All their helpless and unstoppable defeats only root in their lack of one of my inventions titled: Ultimate Cyber Security Switch. Only this invention is qualified to defeat all individual and government-sponsored hackers by preventing and stopping them from hacking into any cyberspace that is protected by this invention. Therefore, this invention can save all these losing countries billions if not trillions of US dollars each and every year of losses from intellectual properties, military intelligence, economic intelligence, financial information, etc. And since many ages ago, people have been mining gold with wash plant, panning, etc hoping and fantasizing becoming rich. But existing sluice and riffle systems are simply too inefficient with a lot of gold loss through the sluice and riffle systems. Hence, I have invented a series of revolutionary gold mining equipments and processes to expedite gold production yet minimize loss. Moreover, the Law of No Largest, or Smallest, or Any Other Absolute Limit established by me states that throughout the universe nothing can be the smallest or the largest. There is never either upper or lower physical limit to anything throughout the universe. There is perpetually something smaller than the smallest particle that can ever be found; there likewise is perpetual something larger than the largest expanse that can ever be found. Therefore, Higgs Boson (or God particle) sought in the Hadron Collider in Switzerland, or any other so-called particle can never ever be the smallest particle because there perpetually is something even smaller than the smallest particle that can ever be found according to this Law. The Law of Time Delusion established by me states that in fact time is but a delusion because there is no reality in any time. It is only because time is but a delusion that other animals live in perpetual non relativity with no concept or sense of time. And it is also only because time is but a delusion that the selfsame number of hours or any other periods of time to the selfsame person are felt and perceived completely different whenever and if he or she falls asleep and whenever and if he or she fells anxious. And it is also only because time is but a delusion different people can agree on an arbitrarily set or assigned time. Albert Einstein and a plethora of other scientists are absolutely wrong about the nature of time. Because time is but a delusion, any attempt to seek any form of time travel is foolishness, failure, and waste of life in the making. Time travel, such as Einstein's Rosen Bridge, wormhole, Higgs singlet, time travel portal, or any other form of time manipulation is absolutely impossible though people can always so fantasize, just as a lot of people would always waste lives fantasizing immortality though it is impossible. It is likely foolish for anyone to believe that Nazi Bell could somehow manipulate time by spinning red mercury, etc because time is but a delusion. The Law of Inevitable Randomness establish by me states that only by involving randomness in one way or another rather than regulated or calibrated actions and/or processes any ultimate state can be attained. For example, in order to polish any mirror to its ultimate resolution possible, random polishing process must be implemented. To expand this law to other area, according this law Albert Einstein's Spooking Communication Theory is also a fraud because it is but random coincidence by which two distant objects sometimes exhibit entanglements, same behaviors, etc, not any dictating law or regularity. But Einstein had mistaken the random coincidence between two distance objects for the act of law or regularity. Furthermore, ultimately on the one hand anything must encompass an element of randomness; on the other hand, only randomness can cause anything or state to achieve the culmination of that thing, much as only hand-polished or random-action-polished lens can attain the ultimate precision of their size. And then the Law of Uniqueness likewise established by me states that, at the level tending to either infinite large or small and no matter how adamantly a person desires otherwise, everything throughout the universe must be and can only be unique; therefore, there cannot be another planet or any other heavenly body like the Earth which sustains life. Therefore, there has never been, and never is, and will never be any life anywhere throughout the entire universe. The Law of Uncertainty also established by me states that there perpetually exists an element of uncertainty in all human affairs (especially whenever and if great interest becomes at stake), any phenomenon, and any existence. For example, albeit Popes have always done their best to conceal their secrets from their donors for decades, these laws had recently dictated a unique butler of one of the Popes to have just divulged only few unimportant Vatican secrets against all their ardent desires to hide the truth from the donors, supporters, and followers. Ever since the split second on such a birthday when the human species had emerged onto Earth, people have killed and harmed one another to establish empires, wealth, power, predominance, social strata, fame, successes, etc for selfish aggrandizement; but none of such lifetime have been wisely spent because all conquerors and/or winners have died, much as all ordinary people and losers do. Therefore, all their cunning, brutal, and hell-for-leather rampages, campaigns, maneuvers, intrigues, duplicities, chicaneries, deceptions, etc eventually have been but foolish dissipations of lifespan in vain at the pursuing for any thing but that which can last. And furthermore, billions of innocent people have already been defrauded, deceived, brainwashed, victimized, and/or inflicted in respect to the authentic and genuine truth and true reality of universe containing life, mind, and other existences. Even as recent as December 2012, delusional faiths in the country of India of multiple dysfunctional faiths had driven many Indian men into gang-raping and/or murdering women in their country in the most uncivilized manner. All people's all uncivilized conducts and inner sufferings can only ultimately root in delusional faiths because only their deceptive faiths, beliefs, and hence mindsets drive them into their behaviors and then foolish ways of life. Their faiths are delusional and hence they are dysfunctional because the ultimate faith always functions. All conventional faiths/belief systems have always been predicated on the diffusions of lies, frauds, deceits, misinformation, deceptions, and brainwashing that are devised to batten on, defraud, and abuse their parishioners, other followers, and their innocent children. Had these inventions not come to fruition, and so long as the human race can still endure the living conditions on this planet, people will continue enduring countless faith predators that will continue defrauding, brainwashing, victimizing billions of faith victims, as since the emergence of humankind on this planet they have always battened on their faith victims throughout the world and even have sexually exploited/molested the innocent children of their faith victims in none other than the places of worship. The places of worship are also sometimes used as their bases for other crime: as, some faith predators have actually been charged for selling child pornography, etc in none other than the places of worship. Therefore, these inventions are of unprecedented and historic significance, practicality, and value since they qualify to prevent future billions of innocent human beings from continuously being defrauded, brainwashed, victimized, and/or harmed by faith predators with their delusional faiths, and they revolutionize and upgrade the vision, sphere, and capability of the humanity. My qualification cannot manifest it clearer to me that the effective ways and strategies to prevent or reduce the sufferings of innocent lives from continuing being misled, brainwashed, victimized, and/or harmed are to live the correct faiththe correct belief system. As a member of true integrity and righteousness of the human species and we are stumbling across one another on the universal yet against-our-will-journey of aging towards inevitable death instincts, I am behooved and honor-bound to provide the humankind with these inventions to upgrade human knowledge so as to dispel the worldwide spiritual plague and precipices of delusional faiths, misled beliefs, misled perceptions, misled preconceptions, misled information, misled knowledge, misled science, etc that have been deluding, victimizing, and/or afflicting the humanity since its emergency tens of thousands of years ago. Only in so doing, said deceits, frauds, whoppers, delusions, etc against humankind can be so lifted and dispelled that mankind will no longer suffer in ignorance, delusion, deceit, pseudo-science, etc. Only the correct knowledge, insight, and faith about mind, life, and universe can be qualified, on all terms for anyone, to harness and/or manifestto the ultimate maximathe human potential of the mind and/or the body; otherwise, all talents, knowledge, knowing, faculties, etc are perpetually limited and confined within the spheres of apparent reality without ever becoming qualified to be expanded to, upgraded into, or evolved into true reality transcending apparent reality. Only by living true reality guided by the ultimate faith (the truly correct and hence functioning faith/belief system) anyone can become qualified for life to enjoy mind freedom including emotional freedom, etc and omniscience... &c &c

---

CN103127604
Anti-aging method through inorganic substance content reduction

Inventor: WEN DONG

The invention provides a method for achieving immortality and even achieving rejuvenescence by humans. The method is mainly composed of a human aging theory and a part about how to preventing human aging. The method is characterized in that when a human body sweats by a great amount, edible rice vinegar is sprayed all over the human body to enable pores all over the body to absorb the rice vinegar into the human body. After entering the human body, the rice vinegar can prevent most inorganic salts in the human body from generating precipitates and even dissolve inorganic salts which are already precipitated and include calcium oxalate, calcium phosphate, calcium carbonate and the like in the body. The specific weight of the inorganic salts in the human body in the human body is reduced through dissolving of parts of the inorganic salts in the human body and exhausting the parts of the inorganic salts out of the human body so as to prevent aging. The method of spraying rice vinegar on the human body has the advantages of being small in toxic and side effects of rice vinegar, low in cost, favorable for popularization, easy to operate, free of training, strong in controllability, and capable of increasing and decreasing the spraying times and the concentration of the rice vinegar for adjustment according to needs.

[0001]

Biological theory of aging and how to prevent biological aging process of the present invention patent relates.

[0002]
In the scientific community there are two theories of biological aging: one for congenital doctrine, they think the gene is a key factor in the decision to biological aging, so they stepped up to crack the genetic code.
Another theory that biological aging postnatal environment is a key factor in the decision.
I support the latter theory, following talk about my reasons.

[0003]
Many biological phenomena acquired environmental biological explanation is the main factor in determining the biological aging, but with the gene theory is difficult to make sense.
Illustration: ? Scientists have done experiments with rats, scientists fed mice by strictly controlling the amount of food, can significantly improve the lives of mice; ? millennium water lily seeds can still grow after germination millennium; ? the nature of the worker bees and the queen are the same Eggs females, almost similar gene, and because they eat different food, their life can vary several times as much, eat royal jelly queen bee, worker bees eat honey; ? biosphere has immortality phenomena, such as scientists are Research by the division of reproductive sector insects and other genes in the same situation, you can go on unlimited split.

[0004]
By summarizing many biological phenomena, we find that the metabolism of biological aging is caused by the main reason, does not participate in the metabolism of the organism does not occur aging, such as water lily seeds did not sprout before aging.
We believe that the organisms absorb nutrients from the outside world during the same time inhale a possible biological aging substance, it is an increase in the accumulation of such substances, resulting in a biological aging and even death.
Death from natural biological aging and biological aging should not be reversed.
I infer that this substance constitutes organism must matter, and according to the organism childhood ? adult ? elderly, this substance is gradually increased.
We then constitute organisms substances into three categories: 1, organics; 2, water; 3, inorganics, in line with the above conditions only inorganic.
Anatomy of an organism we can prove that all organisms inorganic content is based on biological, childhood ? adult ? elderly gradually increased, and this phenomenon in nature is impossible to reverse.

[0005]
Here we focus on to explain why the increase in the accumulation of inorganic substances, can cause biological aging and even death.
Both soluble and insoluble inorganic fraction non soluble salts in the body is difficult to gradually increase, can gradually accumulate in the body increase is mainly insoluble inorganic salts.
And yet one inorganic salt from a soluble to an insoluble.
We can no longer participate in life activities, and it is on the rise in vivo, inevitably resulting in a decline of vitality.
Suppose the same total body weight, increased body salts, the most straightforward is caused by the body of water and organic content share declined, and the vitality of life depends on the organic matter, organic matter decline in the proportion, vitality will decline.

[0006]
Here I vessels in the human body as an example Details: Patrons are inseparable from human cells and blood vessels, if the intake of food from outside the gradual deposition of inorganic salts in the body vessel wall, will inevitably result in smaller diameter vessels, resulting in the ability to pump blood vessels decreases, the supply of such cells that can be obtained will decrease, caused by aging

cells due to lack of support.
And then to human bones as an example, the following figure taken from biology textbooks

[0008]
We know that the bone marrow is where blood.
Bone marrow is the organic matter from the table, we can see very intuitive, as people age, bone mineral content increased, the proportion of organic matter decline, the body's hematopoietic function decline, hematopoietic capacity decreased, and causing cell received support decline, so that cell senescence.

[0009]
Based on the above reasoning, I believe that biological life depends on its in vivo accumulation of inorganic increased speed, long slow accumulation of biological life, rapid accumulation of short biological life.

[0010]
As the above theory is correct, we inorganic body by blocking an increase in vivo or in people insoluble salts into soluble salts from the body, reducing the proportion of body salts, humans will be possible to achieve longevity not old, even rejuvenated.
Here is the experiment conducted to prove that I assumed that I was 20 years old body mineral content as a percentage ratio of the weight value, and now I am 35 years old, if the proportion of inorganic let my body fall to 20-year-old value, then I will be able to reproduce the 20-year-old face.

[0011]
In order to make my body proportion of inorganic fall, it began in 2006 a large number of oral consumption of vinegar, hydrochloric acid, experiments carried out for almost a year, due to the mouth, stomach, liver, kidney and are subject to different degrees of damage, teeth can hardly bite hard, experiment was interrupted, after years of conditioning.
I also began at the end of this year, another experiment in vivo method of inorganic salts dissolved.
The in vitro method into the liquid, that is, after the body a lot of sweat, edible vinegar sprayed on the skin surface, absorbed through the skin into the body, several times a day.
Carried out for nearly four months, the experimental results began to appear, performance 1, the original bald hair out, hair began to grow new small, 2-3 cm long now; 2, skin wrinkles begin to disappear, the skin begins to become more tender.
3, has white hair began to darken; 4, the bones become more elastic, the whole body feels lighter, more power becomes abundant.

US2013035375
Gene re-cycling

Inventor: MATHEW ZACHARIAH

Through ecological re-cycling, plants take up the nutrients from the cadavers. But it is not understood whether the plant roots can take the nutrients and the genes of the dead plants and cadavers, to develop into a transgenic plants. Biotechnology has successfully developed transgenic plants and animals. Recently, gene sequencing studies, have shown that animals carry foreign genes making them as natural transgenic animals. Reason for the presence of transgenes not investigated. Further studies will enable us to learn more about the existence of transgenic flora and fauna and the role of gene re-cyling. Corroborating the Nature's hidden biological phenomena, like ethnicity, cyclic occurrence of diseases, re-appearance of extinct animals, Biblical citations etc., with modern literature on transgenic plants/animals, human genomic studies. gene sequencing, etc. it is hypothesized that genes of flora and fauna are recycled from generation to generation through immortal DNA. It will change the present concept of organism's mortality to immortality.

[0001] Current literature on the cutting edge Biotechnology experimental findings, genetic engineering, gene sequencing and gene resequencing studies have enabled to hypothesize, the theory of “Gene re-cycling”, that genes of every flora and fauna are recycled from generation to generation through the transgenic plants, humans, animals, birds or fishes.

[0002] Gene re-cycling theory will reveal the scientific truth behind the natural biological phenomena occurring in our planet.

[0003] As per the natural phenomenon occurring and the modern literature cited below, plant's roots can attract the genes from the humus of other organisms in the soil/water which will then become an integral part of the plant chromosome.

[0004] Such foreign genes from other species of organisms, after integrating it into the plant chromosome, will remain inactive and enter into a latent period. Genes, after entering into a homologous host and under appropriate environmental conditions will become active again and multiply with the gene expression of the original homologous host.

SUMMARY

[0005] This is a new theory invented/hypothesized. It has great utility for the following purposes:
1. Knowledge of disease outbreak/infection caused by the genes of pathogenic organisms transported from soil and water via transgenic plants/fishes to humans and animals, can help to take appropriate disease control measures,
2. Revealing the root cause of biosphere ethnicity due to genes in the food from Transgenic plants and animals,
3. Understanding the effect of expression of genes in the food on physiological and genetic changes in humans, animals and plants.

Revised Experimental Proof

Introduction:

[0009] In the earlier Provisional Application dated Sep. 30, 2008, an experiment was planned to conduct the Gene sequencing and re-sequencing of human genome to prove the Gene Recycling and development of ethnicity. On my preliminary enquiry with gene sequencing scientists, it was found out that human genome has more than 1 million base pairs and the cost of sequencing and re-sequencing of human genome will be very high, since sequencing each base pair will be around ¢5 per base pair.

[0010] Hence it was decided to plan the experiment of gene re-cycling, with beef cattle (meat) having less number of genome base pair (roughly 22,000 genes). Sequencing of cow's Genome has been studied and is available in the Genome bank.

[0011] Current literature: In the AARP bulletin today (http://bulletin_aarp.org/your/health/policy/articles/scientists _unravel_genome_of_the_cow.ht.8/82009 it was reported that “Of cow's 22,000 genes versions of at least 14,000 have counterparts in other mammals. Cows appear to have 1000 genes that they share with dogs and rodents (Brown, and Lewin 2009)”. It shows that 1000 latent genes of heterologus species (dogs and rodents) that are not found in people, might have entered the cow's genome through gene recycling, developing the animal into a transgenic cow. This transgenic cow, hidden in her roughly 22000 genes are hints of how natural selection/gene re-cycling, sculpted a bovine, with full gene expression over the past 60 million years (Brown, D and Lewin, H. A. 2009). It shows that

1000 latent genes of heterologus species (dogs and rodents) that are not found in people, might have entered the cow's genome through gene re-cycling, developing the animal into a transgenic cow. This transgenic cow, hidden in her roughly 22000 genes are hints of how natural selection/gene re-cycling, sculpted a bovine, with full gene expression over the past 60 million years (Brown and Lewin, 2009)". This scientific report indicate that natural biological Gene re-cycling results in the development of transgenic organism, creating new species of a unicellular/multicellular organism (extinct animals and birds) through Gene re-cycling take place with long periods of time, depending upon the biology of the organism.

[0012] The above scientific literature fully proves/support the theory of Gene Re-cycling, as transgenic animals are being developed naturally in the development of new life on Earth. When the genes, recycle through inter species and intra species, it makes the genes immortal with or without mutation. As time goes many such articles may be published proving gene re-cycling is not a Hypothesis, but a proven naturally existing scientific truth. This made me to submit a Non-provisional Patent Application.

[0013] Meanwhile I have also planned an experiment, which may not be necessary as per the scientific proof given in the paragraph [0008 and 0009] given above to produce transgenic cereal plant by planting the cereal seeds in soil buried with beef meat (putrefied). After conducting the squencing and re-sequencing studies of the beef cattle genes and that of cereals, the transgenic genotype of the cereal will be decided.

[0014] After confirming the transgenic nature of the cereal it will be fed to the mice/rabbits and observe those experimental animal for the transgenes from the beef cattle. The result will be published in scientific journal/scientific conference.

[0015] In the light of the above a Non-Provisional Application on "Gene re-cycling" as given below, has been submitted as a continuation of Utility Provisional application No. 61/194,637 Dt. Sep. 30, 2008 for Patent registration.

[0016] Plants take up the nutrients from the dead and putrefied flora (humus) & fauna, recycle it for the nutrition and growth of humans, plants and animals, and thus maintain ecological re-cycling system. Similarly the herbivores and carnivores animals also play their own role in the ecological recycling process. However more scientific data is required to know, whether the plants can take not only the nutrients but also the genes/viruses/bacteria from the infected materials, dead animals, humans, birds or plants in the soil & water and recycle those genes, generation after generation making the biosphere as immortal. Hence it is hypothesized that genes of every flora and fauna are recycled from generation to generation through the plants, humans, animals, birds and fishes (living in land, sky and water).

[0017] When the "Gene Recycling" hypothesis is proved the concept of mortality of living beings will be changed to immortality. This is a postmodern century of interrogation. "Gene Recycling" hypothesis can provide full or partial answer for many of the Mother nature's biological phenomenon we now see in the world.

[0018] Based on the modern scientific literature and technology, the theory of "Gene Recycling", can provide a strong support to the natural unknown biological phenomena occurring in our planet. It is in the hidden plan of the Creator for making the biosphere/life on this planet as immortal. As per the natural phenomenon occurring and the modern literature cited below, plant's roots can attract the genes from the humus of other organisms in the soil/water which will then become an integral part of the plant chromosome. Such foreign genes from other species of organisms, after integrating it into the plant chromosome will remain inactive and enter into a latent period. Genes, after entering into a homologous host or under appropriate environmental conditions will become active again and multiply with the gene expression of the original host. The scientific literature given below provide s a strong support to the “Gene Recycling” hypothesis.

Natural Biological Phenomena:

[0000]
1. Prevalence of Ethnic Population and Evolutionary Conservatism.

[0020] All organisms on Earth appear to be unrelated in the evolutionary process of different species of plants, birds, animals, mammals and humans. But a detailed study of the molecular biology of cells, has shown that basic molecular structure of all living organisms are conserved to a great extent and all have the same chemical structure such as amino acids, sugars, fatty acids and nucleotides. Besides the above, other resembling factors are that the genetic information of all multi cellular organisms, stored in the genome, is carried by DNA and is transported, from generation to generations. The hereditary information is stored in all the cells in the DNA as chemical code, arranged in linear fashion (Koonin, 2002).

[0021] We can also notice the family resemblances among genes coding for proteins. Genes and portions of genes get duplicated to develop into new diverse genes through mutation of the parent genes. Usually one cell expresses only a fraction of its genes. Changes in the environment due to signals from other cells can influence the pattern of genes they express. The gene expression by different sets of genes produces different types of cells in a multicellular organism. Thus a small group of genes may multiply to develop into higher plant or animal species with large number of genes. A broad phylogenic picture provides a rough idea of the relation ship between the species on our planet Earth (Li, W-H and Graur, 1991).

1. (a) “Gene Recycling” Hypothesis Claim:

[0023] If the roots of the plants are taking organic and Inorganic nutrients along with the genes from the human and animal cadavers, then the plant and its parts or fruits will also contain these human, animal genes, which will remain as tent genes. The plant/its parts or fruits containing such foreign (human/animal) latent genes, when consumed by other people in the locality or town, the genes may enter their gastro intestinal system and reproductive system (gonads). It may then remain in the system as latent genes. These latent foreign genes which have entered into the gonads may pass it on to their progeny and become active and provide gene expression (dominant genes) in the next generation of people in the locality/town. This will result in ethnicity (common characters related to place of origin, culture, speech, food habits etc.) in those local people.

1. (b) Supporting literature:
1. (b) 1. Transgenic plants:

[0026] Genetic engineering studies have revealed that a mutant gene can be inserted into the chromosome of a plant cell and it becomes a part of the plant genome. In the laboratory experiment it has been shown that, when the tobacco leaf discs incubated in a selection media with genetically engineered Agrobacteria, the injured cells at the edge of the plant leaf disc, released substances that attract the Agrobacteria which carry a recombinant plasmid with a selectable marker and desired transgene, inject its bacterial DNA into the plant cells. The plant cells which take up the appropriate bacterial DNA, survive and proliferate to express the selectable marker gene, and develop into a callus. When these calluses are placed in special medium, it will induce the production of shoots. These then develop into adult plants carrying transgenes (Davey et al., 1989 & Walden and Schell, 1990).
1. (b) 1.1. Transgenic Plant Experiment Supports the Hypothesis:

[0028] Similar to the experiment stated above if an injured cell from the plant leaf is attracting the DNA from the bacteria; as per the hypothesis, under natural environmental conditions, the injured cells of plant roots can also attract/absorb/take up the gene(s) along with the nutrients from the dead and putrefying flora (humus) and fauna in the soil/water. When such a foreign gene from other species is recognized by a plant cell, it will efficiently transfer a DNA strand of human/animal/bacterial/viral gene to the plant cell. It then becomes the part of plant chromosome, survive to grow into a transgenic plant carrying such a foreign gene.

1 (b) 1. 2. Possibility 1:

[0030] The transgenic plants (or its product) carrying foreign gene when consumed by species homologous to the gene, it can develop to express the phenotypic and genotypic characters of the gene. In other cases it may remain as latent gene and express it only when it enters into a host homologous to the gene.

1. (b) 1. 3. Possibility 2:

[0032] The transgenic plants (or its product) carrying foreign gene when consumed by species heterologous to the gene, it may remain as a latent gene and will pass it on to its progeny and continue to remain as latent gene in the progeny.

1. (b) 2. Transgenic animals:

[0034] In the case of mouse, foreign DNA can be randomly integrated into the genome. When fragments of DNA are introduced into cells, they are rapidly joined end to end by intra cellular enzymes and form long tandem arrays. These are then randomly integrated into a chromosome. When original gene(s) of an animal, is thus altered permanently, such an animal is known as transgenic animal. The genes which are responsible for altering are called transgenes (Boyd and Samid, 1993 & Palmiter and Brinster, 1986).

1. (b) 3. Role of Gene Regulatory Proteins:

[0036] Genome (whole set of genetic information belonging to a cell/organism carried by DNA) in its DNA sequenquence carries the information to express many thousands of different proteins and RNA molecules. Usually cells express only a fraction of its genes. The gene expression by different sets of genes produces different types of cells in a multicellular organism.

[0037] Changes in the environment such as signals from other cells can influence the pattern of genes they express. Even though the expression of a gene is generally regulated, the initiation of RNA transcription will be an important control point for most genes. The transcription of the gene in the DNA is controlled by the regulatory region at which it starts. These regulatory regions function like tiny microprocessors. Thus these regions act like switches by responding to the various signals, which they interpret and integrate the neighboring gene to switch on or off. Bacterial Genetic study conducted in 1950 gave the first evidence on the existence of “gene regulatory proteins”. These gene regulatory proteins, turns specific sets of genes on or off, the production of proteins. These are called “repressors” as they shut off the production of proteins. These repressors can be (1) viral encoded gene regulatory protein or (2) bacterial gene regulatory protein.

1. (b) 3. 1. Viral Encoded Gene Regulatory Protein

[0039] The bacteriophage lambda repressor, is a gene regulatory protein encoded by a bacterial virus. It shuts off the viral genes that code for the protein components of new virus particles. Thus it enables the viral genome to remain as silent occupier in the viral chromosome and multiplies along with the bacterial growth. Lambda repressor is one of such regulatory proteins which are encoded by bacteriophage lambda. This lambda repressor shuts off the viral genes that code for the protein components of new virus particles, and there by makes the virus genome inactive in the bacterial chromosome. But the mutant bacteriophage lambda cannot shut off the specific set of genes. With the help of DNA sequencing and DNA foot print studies it has been found that these regulatory proteins are bound to specific DNA sequences close to the genes which they regulate (Gilbert and Muller-Hill, 1967; Jacob and Monod, 1961; Kaiser, 1957 & Ptashne, 1967).

1 (b) 3. 2. Bacterial Gene Regulatory Protein:

[0041] The lac repressor is the first bacterial gene regulatory protein to be recognized. These generally respond when there is a nutritional requirement and they respond by shutting off genes encoding specific sets of metabolic enzymes when they are not necessary. Thus when sugar is absent in the medium, lac repressor turns off the production of the proteins responsible for lactose metabolism, (Jacob and Monod, 1961 & Gilbert and Muller-Hill, 1967).

1. (b) 4. Growth Factors Regulate the Protein Synthesis:

[0043] The growth factors regulate the protein synthesis and stimulate the cell/proliferation. Most growth factors also stimulate cell growth. (Loughlin and Fallon, 1992 and Purves, 1988).

1. (b) 5. Cells Enter Into Non Growing State—“G Zero” in the Absence of Growth Factors:

[0045] Mammalian cells stopped growth between mitosis and S phase, when they are deprived of serum which contains highly specific proteins called Growth Factors. In the absence of serum these, cells will delay division by entering a specialized non growing state called “G zero” (Pardee, 1974 & 1989).

1. (b) 6. Cell Senescence:

[0047] Cells from mammalian tissue cultures can be propagated only for a certain number of cell divisions (50 divisions for human cell cultures), after wards they cease to divide and die. This process is called cell senescence. But some cells derived from rodent tissue cultures escape senescence and divide indefinitely to produce cell lines (Todaro and Green, 1963).

1. (c) Hypothesis Linked to the Above Cited Literature:

[0049] Similar to the reference cited above the mammalian gene(s) taken up by the roots of the plants can remain as silent occupier, in the plant/heterogenous species, as the gene regulatory proteins switch off the foreign genes, or the cells may cease to divide due to cell senescence. It will only multiply and provide the original specific gene expression, . with phenotypic and genotypic characters, if it enters a homologous host and under suitable internal or external environment.

[0050] When the root of the plant takes up the (dominant) genes of other species from the soil, these foreign (to the plant) genes will lie in a latent state as lambda bacteriophage is integrated into E. Coli. These genes foreign to the plant may then enter into other species (humans/animals) through the consumption of the plant products, containing genes of other species and will be passed on to a different species (humans/animals/birds). The gene regulatory proteins can switch on or off specific sets of genes, depending upon the homologous host cell in which the gene is integrated. Parmley et al. (2007) recently reported that a secondary function of some pieces of DNA has the capacity to hold back the function of production of the protein, by keeping a check and slowing the process of associated genes, becoming ideal protein machines.

[0051] Experiments proposed: Experiment 1. 1. Study to find out whether the plant roots take up the human genes from the soil and get integrated into plant chromosomes.

Step 1. Objective:

[0052] To find out whether the plant roots will take up the human genes through its root and get it integrated into the plant chromosome.

[0053] Method: Conduct a DNA test of the person who is willing to participate in this study, using the commercially available ISO approved test kit. “The test will have the 16 DNA genetic markers including the 13 markers used by the FBI and other law enforcement agencies for human DNA testing. The 13 markers are referred to as the Combined DNA Index System (CODIS). All of these 13 markers are used only in the determination of human identity” (GTL, 2008). Conduct the experiment planting the seeds of paddy, wheat or carrot, in fresh soil containing only chemical fertilizer and organic nutrients from placenta, blood and or urine sourced from the above single person with known DNA profile.

Step 2.

[0054] At the fruit bearing age of the plant, the parts of the plant like leaves, stem, fruits etc. will be collected and send for gene testing/mapping, to the Plant Genome Mapping Laboratory (PGML), University of Georgia. If the results show that the human genome is integrated into the plant chromosome and present as latent human gene in the seeds of paddy, wheat or carrot, the following experiment will be conducted.

Experiment 1. 2. 1.

[0055] Study to find out whether the latent human transgenes in the plants are transmitted to animals through consumption of transgenic plant products (underground stem/fruits).

Objective:

[0056] To find out whether the underground stem/fruit of the transgenic plants carrying the human genes in the plant chromosome will be present as latent gene in the blood/gonads of mice/rabbits, when the transgenic plant product (rice, wheat carrot) are consumed by mice/rabbits.

Method: Experiment 1. 2. 1. 1. (Transgenic Wheat Plant):

[0057] Take one group of six (3 male and 3 female) Swiss albino mice of 3 weeks old. Keep them in separate cage with tag No. Collect the blood and conduct DNA test for base line data. Feed each mouse with wheat (harvested from the experiment 1.1, containing the human genes as integrated part of the plant chromosome) and water for 3 months. Collect duplicate sample of blood from each mouse at month interval and conduct DNA test to find out the presence of human latent gene.

[0000] Experiment 1. 2. 1. 2. (Transgenic Rice plant):

[0058] Take second group of six (3 male & 3 female) white mice of 3 weeks old. Keep them in separate cage with tag No. Collect the blood and conduct DNA test for base line data. Feed each mouse with rice (harvested from the experiment 1.1—containing the human genes as integrated part of the plant chromosome) and water for 3 months. Collect duplicate sample of blood from each mouse at month interval and conduct DNA test to find out the presence of human latent gene.

[0000] Experiment 1. 2. 1. 3. (Transgenic carrot):

[0059] Provide one group of six (3 male & 3 female) rabbits of one month old, each with carrot (harvested from the experiment 1.1—containing the human genes as integrated part of the plant chromosome) for 3 months. Keep them in separate cage with tag No. Collect the blood and conduct DNA test for base line data. Feed each mouse with carrot (harvested from the experiment 1.1) and water for 3 months. Collect duplicate sample of blood from each rabbit at month interval and conduct DNA test to find out the presence of human latent gene.

Experiment 1. 3.

[0060] Study to find out whether the human latent genes noticed in the blood of transgenic mice/rabbits will be transferred to their offspring.

Experiment 1. 3. 1.

[0061] At the end of the 3 months experiments, the animals whose blood showed positive for human latent gene will be grouped (separate for each food) together to breed and produce offspring. After weaning, test the blood of the offspring for human latent gene. Then euthanize the parent animals to harvest the gonads (testes and ovary) and send them for DNA testing, to confirm human latent genes in the gonads.

CONCLUSION

[0062] Finding human latent gene in the mice/rabbits and their progeny will be a proof of the “Gene Re-cycling”.

[0000] II. Natural phenomena:
Cyclic development of viral diseases.

[0064] Present scientific view is that the pathogenic bacteria/viruses exist in animals or birds as they become carriers of bacteria/viruses (host—parasite balance) without producing disease. Under adverse environmental conditions the disease outbreak occurs, due to the break down of immunity in those carrier animals/birds causing the cyclic development of disease outbreaks. Friedman (1992) reported that viral chromosomes can integrate into host chromosomes. The viruses which are integrated into the plant chromosome need not multiply immediately, but goes into a latent state and remain as inactive in the plant cells as provirus. It will come active only under suitable conditions sometimes after passing through several hosts/generations.

[0065] Hypothesis: The viruses/bacteria in the dead bodies of the humans/animals or the human/animal excreta, will be taken up by the roots of the plants grown in the soil in that area, get integrated into the plant chromosomes. The cyclic outbreak of certain viral diseases can then be attributed to the release of the viruses integrated into the plant chromosome, when such plants or its part/fruit are consumed by susceptible (immune compromised) humans or animals, resulting in an infection in the people/animals leading to fresh outbreak of the viral/bacterial infection in that area. The delay in the outbreak may be due to the time the virus/bacteria takes to travel from the roots to fruit/plant parts and the chance occurrence of a susceptible person/animal, eating the affected plant part. It is also important to note that certain plants like coconut trees/jack fruit trees etc. will take 5-8 years to bear fruits.

II. Supporting Literature:

[0000]
1. Viral Chromosomes Integrate Into Host Chromosomes:

[0067] The viral chromosomes after entering a host cell, necessarily need not multiply to produce more viral progeny. Because the genomes of these viruses enter into a latent state and become inactive in the host cell without producing any progeny. This viral latency was found out when uninfected bacteria, on exposure to UV light was able to induce the production of progeny of bacteriophages. Subsequently it was shown that these uninfected bacteria (lysogenic bacteria) carry in their chromosome, a complete viral (bacteriophage) chromosome lying in a dormant state. These types of host integrated viral chromosomes are called “provirus”. Bacteriophages which can integrate their DNA into bacterial chromosomes are called as temperate bacteriophages. Bacteriophage lambda is a temperate bacteriophage and when it infects a suitable E. coli host cell, it can grow either by lytic or lysogenic pathway. In the lytic path way lambda multiplies to several progeny particles and they will be released when the E. coli is lysed. In the lysogenic pathway, the integrated viral DNA replicates along with the host chromosome and produce new integrated viral DNA copies, resulting in the development of a lysogenic bacterium. This lysogenic cell multiplies in the normal way. But when it is exposed to any environmental insult/changes like exposure to UV light, or ionizing radiation, the cell becomes fragile/weakened leading to the release of the integrated provirus from the cell chromosome. Then it will lead a normal cycle of viral multiplication. It shows that the integrated provirus will not perish along with its damaged host cell, but can escape to infect other E. coli Campbell (1993).

[0068] Hypothesis: As per the literature cited above the mammalian/human gene(s) taken by the roots of the plants can integrate itself into plant chromosome and will remain as an integrated gene. The integrated foreign gene will be released from the plant chromosome, when the plant products are consumed by humans/animals etc. or cause any physical or chemical injury to the plant cell due to any environmental changes like extreme heat or cold or change in the nutritional status of the plant. Then the gene released from the plant chromosome will then follow the lytic or lysogenic path way and will lead to the escape of the released gene to other species. When the gene enters a heterologous unrelated species of organism, the gene regulatory proteins will switch off the foreign genes entered into unrelated host. If the gene released from the plant cell entered into a new suitable homologous species, related to the gene's original host, the gene regulatory proteins will be switched on.

[0069] To explain it further, the human gene integrated into the plant cell chromosome, when released by the consumption of humans, the gene(s) will enter a new cell in the human body, it will be expressed providing the phenotypic and genotypic characters of the gene remained in the plant chromosome as integrated gene.

[0000] III. Natural phenomena: Transmission of Prions (PrP Gene) Through Food and Hereditary:

[0070] Feeding “Renderers meat” with sheep offal, contaminated with Scrapie (PrPsc) produce BSE infection/mad cow disease in beef cattle. Consumption of beef infected with BSE/TSE prion (aberrant protein) will produce vCJD infection in humans. The causative agent of BSE/TSE is believed to be a small protein infectious particle (prion) which resist inactivation by the methods capable of destroying nucleic acids. The recent discovery that only a protein can cause and transmit the disease is shocking to the molecular biologists and virologists.

[0071] Hypothesis: As prions are transmitted through food and hereditary, the genes integrated into the plant chromosome when consumed by the animal or humans, will also produce gene expression in the animals/humans, if the gene is homologous or continue to remain as latent gene as in the case of prions (production of a slow latent infection).

III. 1. Supporting Literature:

[0072] Feeding “Renderers meat” with sheep offal, contaminated with ovine prion protein (PrP) gene in scrapie (Shiquan, et al. (2002) produce BSE infection/mad cow disease in beef cattle. Consumption of beef infected with BSE/TSE prion (aberrant protein) will produce vCJD infection in humans. Humans can be infected by Prions either as acquired through diet and surgical procedures or through “apparent hereditary Mendelian transmission where it is an autosomal and dominant trait” (Alan Cann, 2005). Prions are responsible for a slow developing disease due to the latent infection. It can be compared to the latent gene/host integrated viral chromosomes or provirus/bacteriophage lambda (Campbell, 1993).

IV. Natural Phenomena/Mythology:

[0073] Re-incarnation myth becomes a scientific reality due to modern scientific biotechnology success in cloning, and Transgenesis.

[0074] In India there are so many stories of re-incarnation/re-birth of a person in another part of India or world. Re-cycling of genes is probable explanation for the myth of re-incarnation. It is possible to explain hypothetically that the dominant genes of “X” person's cadaver integrated into the soil, when taken up by the root of the plant, will remain as latent genes in the plant part/fruit. If the plant part is then taken up by another person and entered into the gonad of another person, the phenotypic and genotypic character of the dominant gene of the “X” person will be expressed in the progeny as rebirth/reincarnation of the “X” person.

[0075] Re-incarnation story give strong support to the Gene Recycling Hypothesis.

[0076] Gene re-cycling hypothesis can be the scientific/nature's explanation for the Hindu mythological stories on people with animal face/head and human body or female body with fish tail (Mermaid) manifested as nature's freak, due to incomplete expression of human gene(s), integrated into animal chromosome (transgenic animal) or vice versa. Traditionally many people attribute such nature's freak is endowed with God's power Cherubim is a typical example for transgenic person with incomplete expression of genes from other species like animals & birds. It is also possible that the human gene is carried by another animal as shown in the movie of the dog “Fluke”.

V. Natural Phenomena:

[0000]
Reappearance of the Extinct Animals and Birds:

[0078] Cell theory of 19thcentury stated that all living organisms possess cells and the new cells are developed only from the existing cells. Thus all Eukaryotic or Prokaryotic unicellular or multi cellular organisms are developed through the evolution from an original population of primitive “cells”.

[0079] Hypothesis: New cells will be developed by recycling the genes of the previous generation of living organisms existed earlier on earth. The well-known saying “Life comes after death” should be a scientific fact. The animals and birds once lived on earth after extinction may come back again to life through the recycling of the genes deposited in the humus and remained as integral part of the soil. The plants grown on such soil with the genes of the extinct animals and birds, may take up the genes through its roots and can remain in the leaves and fruits. When animals/birds of other species eat the leaves/fruits, the genes of extinct animals/birds, may enter the body of the new host and by entering the ova/sperm may re produce the life of the extinct species. Recently it has been reported that Black footed Ferret which was extinct has been found recently.

VI. Biblical References: Genes in the Trees, Stones and Bones:

VI. A. Genes in the Tree—Tree of Life: Garden of Eden:Bible (1973). Genesis Ch. 2, Verse 8 and 9.

[0080] “8. And the Lord God planted a garden toward the east, in Eden; and there He placed the man whom he had formed.

[0081] “9. And out of the ground the Lord God caused to grow every tree that is pleasing to the sight and good for food; the tree of life also in the midst of the garden, and the tree of the knowledge of good and evil”.

VI. B. Forbidden Fruit: Does the Fruit From the Tree in the Garden of Eden Contained Gene to Know Good and Evil?

[0082] Bible (1973). Genesis Ch. 3. verse 2-5.
“2 And the woman said to the serpent, “From the fruit of the trees of the garden, we may eat;

[0083] “3. but from the fruit of the tree which is in the middle of the garden, God has said, ‘You shall not eat from it or touch it lest you die!

[0084] “4. And the serpent said to the woman, You surely shall not die!

[0085] “5. For God knows that in the day you eat from it your eyes, will be opened and “you will be like God, knowing good and evil”.

[0000] VI. C. Gene expression—Tree of Life: Disobedience and Fall of Man and Woman from God's Grace: Bible, 1973: Genesis Ch. 3, verse 22-24.

[0086] “22 Then the lord God said, “Behold, the man has become like one of Us, knowing good and evil; and now, lest he stretch out his hand, and take also from the tree of life, and eat and live forever”—

[0087] “23 therefore the Lord God sent him out from the Garden of Eden, to cultivate the ground from which he was taken.

[0088] “24 So He drove the man out; and at the east of the garden of Eden He stationed the cherubim and the flaming sword which turned every direction, to guard the way to the tree of life”.

[0089] Hypothesis: The above Biblical passage gives us the insight that God intended that man and woman to lead an eternal life by eating the fruit of “the Tree of Life”. But since man and woman disobeyed God and sinned, He threw them out of the Garden so that they will not have eternal life, simply by eating the fruit. Instead they (we) have to toil in the soil and go back into the soil, so that our gene will integrate with the soil and will regenerate again into humans with the gene integrated in the trees/plants/fruits rooted in the same soil. Thus each tree by taking the human genes from the soil with their roots, become the symbolic of the tree of life of the Garden of Eden.

[0090] Thus through the recycling of genes through the plants/tree of life, humans have eternal life and will have several opportunity to become self purified to receive the

[0091] Lord Jesus when He comes 2<nd >time. It can be presumed that each civilization on this earth is a re-incarnation of the previous civilization. If each person on this earth is a new entry into this planet, God must have created several billions or trillions of people, but through recycling of genes most probably, we are all born repeatedly on this earth (re-incarnation/rebirth?).

VI. D. God Can Raise Up Children From Stones:

[0092] Bible (1973). New Testament; Matthew Ch. 3, verse. 9 “and do not suppose that you can say to yourselves, We have Abraham for our father; “for I say to you, that God is able from these stones to raise up children to Abraham”. We cannot simply ignore this as a statement by John the Baptist out of his anger or frustration. If we read it in between the lines, then we can understand or presume, that the soil and the stones may contain the genes of the human beings, hence John the Baptist knowingly or unknowingly said that God can raise the people from the stones/soil.

[0093] Hypothesis: If we understand that every flora and fauna seen on the earth are mortal and become part of the earth after their life span, then we should also know that new life on earth starts not only by recycling the nutrients but also by genes of the dead flora and fauna.

[0094] The well known fact: "LIFE COMES AFTER DEATH", will be supportive of the hypothesis that genes of every flora and fauna are recycled from generation to generation.

[0000] VI. E. Ezekiel's Prophesy of Life from Dry Bones:
Bible (1973). Ezekiel—Book of Ezekiel: Ch. 37. verse 1-14,

[0096] "The hand of the lord was upon me, and He brought me out by the Spirit of the Lord and set me down in the middle of the valley; and it was full of bones.

[0097] "2 And He caused me to pass among them round about, and behold, there were very many on the surface of the valley; and lo, they were very dry.

[0098] "3 And He said to me, "Son of man, can these bones live?" And I answered, "0 Lord God, Thou knowest."

[0099] "4 Again He said to me "Prophecy over these bones and say to them, O dry bones, hear the word of the Lord."

[0100] "5 Thus says the Lord God to these bones, Behold I will cause breath (spirit) to enter you that you may come to life.

[0101] "6 And I will put sinews on you, make flesh grow back on you, cover you with skin, and put breath in you that you may come alive; and you will know that I am the Lord."

[0102] "7 So prophesied as I was commanded and as I prophesied, there was a noise, and behold a rattling; and the bones came together, bone to its bone.

[0103] "8 And I looked and behold, sinews were on them and flesh grew, and skin covered then: but there was no breath in them.

[0104] "9 Then He said to me "Prophecy to the breath, prophecy son of man, and say to the breath, "Thus says the Lord God, "Come from the four winds, O breath and breathe on these slain, that they come to life."

[0105] "10 So I prophesied as He commanded me, and the breath came into them, and they came to life, and stood on their feet, an exceedingly great army.

[0106] "11 Then He said to me "Son of man, these bones are the whole house of Israel; behold they say, "Our bones are dried up and our hope has perished. We are completely cut off.

[0107] "12 Therefore prophecy and say to them "Thus says the Lord God, "Behold I will open your graves and cause you to come up out of your graves, My people; and I will bring you unto the land of Israel.

[0108] "13 " Then you will know that I am the Lord, when I have opened your graves and caused you to come up out of your graves, My people".

[0109] "14 "And I will put My Spirit within you and you will come to life, and I will place you on your own land. Then you will know that I the Lord have spoken and done it, "declares the Lord."

SUMMARY

[0110] By applying modern science and technology knowledge, it is hypothesized that “Genes of every flora and fauna are recycled from generation to generation through the plants, humans, birds and animals living in land and water”. Gene Recycling plays the most crucial role in the continued propagation of flora and fauna on our planet. Through gene recycling, biosphere is made to immortal existence. This is a postmodern century of interrogation. Gene Recycling Hypothesis can provide full or partial answer for many of the nature's phenomenon we now see in the world.

REFERENCES

[0111] Alberts, B.; Bray, D.; Lewis, J.; Raft M.; Roberts, K, and Watson, J. D. (1994). Molecular Biology of THE CELL. 3rd Ed. Pub. By Garland Publishing Inc. 717 Fifth Avenue New York, N.Y. 10022.

[0112] Alberts, B.; Alexander, J.; Julian, L.; Martin, R.; Keith, R. and Peter,W. (2002). Molecular Biology of THE CELL 4thEd. Pub. By Garland Science, a member of the Taylor & Francis Group, 29 West 35thStreet, New York, N.Y. 1001-2299.

[0113] Bible, (1973). Genesis-In: Chapter 2. verse 8 and 9. p. 3; New American Standard Bible. Reference Edition; Foundation Press Publications Box 277, La Habra Calif. 90631.

[0114] Bible, (1973). Genesis-In: Chapter 3. verse 22-24. p. 5. New American Standard Bible. Reference Edition; Foundation Press Publications Box 277, La Habra Calif. 90631.

[0115] Bible, (1973). Ezekiel—Book of Ezekiel.In: Chapter 37. verse 1-14, p. 1213. New American Standard Bible. Reference Edition; Foundation Press Publications Box 277, La Habra Calif. 90631.

[0116] Bible, (1973). New Testament. Matthew. In: Chapter 3. verse. 9. p. 3 New American Standard Bible. Reference Edition; Foundation Press Publications Box 277, La Habra Calif. 90631.

[0117] Boyd, A. L. and Samid, D. (1993). Review: Molecular biology of transgenic animals. J. Anim. Sci. 71(S3): 1-9.

[0118] Brown, D. and *Levine, H. A. (2009) “Scientists Unravel Genome of the Cow” Washington Post Staff Writer: Friday, Apr. 24, 2009. And in the AARP Bulletin: (http://bulletin aarp.org/yourhealth/policy/articles/scientists _unravel_genome_of_the_cow.ht.8/8200); *Head of the Institute of Genomic Biology, Uni. Illinois, Urbana-Champaign.

[0119] Campbell, A. M. (1993). “Thirty years ago in genetics: prophage insertion into bacterial chromosomes”. Genetics 133 (3): 433-438.

[0120] Cann, A. J. (2005). In: Chapter 8. Principles of Molecular Virology (Standard. Edition) Fourth Edition (Kindle Edition) Academic Press. (cited In: http://www.microbiologybytes.com/virology/prions.html

[0121] Davey, M. R.; Rech, E. L and Mulligan, B. J. (1989). Direct DNA tranfer to plant cells. Plant Mol. Biol., 133: 273-285.

[0122] Friedman, D. I. (1992). Interaction between bacteriphage lambda and Escherichia coli host Curr. Opin. Genet. Dev., 2: 727-738.

[0123] Gilbert, W, and Muller-Hill, B. (1967). The lac operator is DNA. Proc. Natl. Acad. Sci., USA, 58: 2415-2421.

[0124] GTL-(GeneticTestingLaboratoriesInc.)2008. https://www.gtldna.net/whysocheap.html

[0125] Jacob, F. and Monod, J. (1961).Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol., 3:318-356.

[0126] Kaiser, A. D. (1957). Mutations in a temperate bacteriophage affecting its ability to lysogenize, E. coli. Virology, 3: 42-61.

[0127] Koonin, E. (2002). Cells and Genomes. (cited In: p. 3 Alberts et al., 2002).

[0128] Li W-H. and Graur, D. (1991). "Sequence studies reveal resemblances between several families with Homologous genes": Fundamentals of Molecular Evolution Sunderland, M. A. Sinauer Associates. (cited In: p. 281 "Molecular Biology of THE CELL" 3rd Ed. Alberts et al., 1994).

[0129] Loughlin, J. H. and Fallon,. J. H. (1992). "Neurotrophic Growth Factors". San Diego, Calif. Academic Press. "Cell Growth and cell division can be independently regulated". (Cited in p. 894 "Molecular Biology of THE CELL" 3rd ed. Alberts et al., 1994).

[0130] Palmiter, R. D. and Brinster, R. L. (1986). Germ line transformation of mice. Anim. Rev. Genet., 20: 465-499.

[0131] Pardee, A. B. (1974). A restriction point for control of normal animal cell proliferation. Proc. Natl. Acad. Sci. USA, 71: 1286-1290. Pardee, A. B. (1989). G1 Events and regulation of cell proliferation. Science, 246: 603-608.

[0132] Parmley, J. L. and Urrutia, A. O. (2007). In: "Putting breaks on evolution—Layered information holds back protein evolution" by Helen Pearson, newa@nature.com 5 Feb. 2007, doi:10. 1038/news 070205-3). By: Potrzebowski, J., Kaessmann, H. & Hurst, L. D. Plos. Biol., 5 e 14.

[0133] Ptashne, M. (1967). Specific binding of the ? phage repressor to ? DNA. Nature, 214: 232-234.

[0134] Purves, D. (1988). Body and Brain: A Trophic Theory of Neural Connections. Cambridge: Harvard University Press. "Cell Growth and cell division can be independently regulated". (cited In: p. 894 "Molecular Biology of THE CELL 3rd Ed. Alberts et al., 1994).

[0135] Raff, R. A. and Kaufman, T. C. (1983). Embryos, Genes Evolution. New York: Macmillan, (cited In: Alberts et al p. 32,33 & 34. 1994).

[0136] Reed Wickner, (2008). "Prion Diseases" Microbiology Bytes, Internet updated Jan. 14, 2008.

[0137] Ruby, S. W. and Abelson, J. (1991). Pre mRNA splicing in yeast. Trends Genet., 7: 79-85.

[0138] Shiquan, W.; Noelle, E. C.; Janice, M. M.; Tracy, L. S.; Alma, M.; Diane, L. S.; Warren, C. F.; Gilbert, R. H.; Ronald, C. E.; Thomas, D. B.; Jonathan, E. B.; Jay, W. C.; William, D. T. & Michael, R. M. (2002). Polymorphic distribution of the ovine prion protein (PrP) gene in scrapie infected sheep flocks in which embryo transfer was used to circumvent the transmission of scrapie. J. Theriogenol. 67:p. 1865-1875. http://wwwtheriojournal.com/articlesgenol. 67:p. 18651875. http://wwwtheriojournal.com/article/S0093-691X(02)00673-8/abstract.

[0139] Todaro, G. I. and Green, H. (1963). Quantitative studies of the growth of mouse embryo cells in culture and their development into established cell lines. J. Cell Biol., 17:299-313.

[0140] Walden, R. and Schell, J. (1990). Techniques in plant molecular biology—progress and problems". Eur. J. Biochem., 192: 563-576.

US2012136403
MEDICALLY PERPETUAL PERSISTENCE OF ESSENTIAL HUMAN LIFE

Inventor: CHOUDURY SPANDAN
Also published as: WO2009147634 // GB2475181

This Patent Application presents a biomedical mechanism of feasibly effecting tangible human immortality. The applicant believes that this technique is the only feasible and meaningful means of effecting perpetual persistence of human life. It forgoes all modern day approaches in the world up to now that target medical mechanisms to extend the life span of the body, and instead adopts a fundamentally novel innovation by way of enabling the biomedical transfer of the cognition that essentially constitutes a human to a different body prior to the death of the original body, whereby the transfer is NOT a copy of (and hence what would otherwise NOT be) the essential original human but IS the essential original human itself.

FIELD OF INVENTION

[0003] The field of invention is biomedical engineering. It primarily, but not only, involves neuroscience and electrical engineering. A novel, and possibly the ONLY feasible, means of effective persistence of human life for perpetuity is presented.

BACKGROUND OF INVENTION

[0004] As indicated above, the applicant deems it futile of anyone to attempt to keep the human body alive perpetually—THAT WILL NEVER WORK. EVER, although it will be possible to increase the relative duration of the longevity of the body. It is a reasonable expectation that regardless of the state of research even millions of years from now, each body WILL die some time, either from old age, or external events (including diseases caused by external events, accidents etc). That, is because given the virtually unlimited count of interacting variables involved inside and outside a body to keep it alive, it will forever be impossible to successfully manage them all for eternity.

[0005] This Patent Application presents the applicant's novel invention on a practical mechanism of essential persistence of life for perpetuity.

[0006] The applicant sees two perspectives to human immortality—

A. Immortality only from the perspective of people that knew the original human, but not from the perspective of the original human.
B. Genuine immortality from the perspective of the original human.

[0009] The applicant infers that while it will in principle be technically feasible soon to transfer the cognitive and sufficiently important individual-defining non-cognitive contents of the human brain to a new brain (or another equivalent entity) whereby the people who knew the dead human would perceive the new entity containing the transferred data as essentially having simulated the original human, the key flaw in that approach is that the original human being will NOT feel as having continued life in the new entity—the original human's sense of self identity “I” would not be transferred into the new entity. Instead the second human's sense of self identity would be of a new “I” that would be a reasonable COPY of the original human's identity, and there could in principle be an unlimited count of such copies across multiple entities. That, however, does not constitute genuine immortality. In short, if the original human was “Mary” or “James” the second entity would still say “I am Mary” or “I am James”, but the original Mary or James would not accept the respective second entities as them, both when the original Mary or James are still alive and after their body-death. For purposes of this Patent Application, the original human will be referred to as “Mary-I”, while the second human per “A” would be “Mary-II-A”, and that per “B” “Mary-II-B”. This Patent Application pertains to the creation of Mary-II-B.

SUMMARY OF INVENTION

[0010] This Patent Application is on the second perspective (“B.”) listed in the previous section, whereby the original human Mary-I would actually recognize herself as having been transferred into the new entity, with that entity being preferably a regular human brain attached to a regular human body.

[0011] There will be a common period of time prior to Mary-I's body-death when Mary-I and Mary-II-B would both be simultaneously alive, and while still alive Mary-I will be able to confirm her being able to recognize the (foreign) memory and cognition of the second brain and sense input from the second body. Mary-I will in effect be able to use not only her own brain's resources (i.e. her brain hardware and data) and input from the senses of her own body, but also be able to use the resources (i.e. brain hardware and data) of the second brain and receive input from the senses of the second body. Then, per the mechanism described in this Patent Application, Mary-I will willfully “transfer” her memory and consciousness to the second body. Of course, the people around the original human (Mary-I) would deem the second human (Mary-II-B) as essentially being even more of Mary-I, than would be the case with Mary-II-A.

SUMMARY OF DIAGRAMS

[0012] FIG. 1—Demonstrates a sample neural network of 8 brains, including Mary-I's original, that forms an integrated, single, logical, augmented center of consciousness. As indicated, upon subsequent physical delinking, ordinarily at most only one of the 7 other brains can manifest Mary-II-B, while each of the remaining 6 can at most manifest Mary-II-A. Arrows indicate direction of flow of information, hence the direction of consciousness transport prospect, albeit the invention recommends dual-directional rather than unidirectional arrows in each case. Of course, the at-most-single Mary-II-B brain is only in as much as that only a single, logical, (augmented or diminished) integrated center of consciousness can be created to control one brain or a set of linked brains—and so it is optionally possible to have Mary-II-B to be created across multiple brains that are combined to form a single, integrated, logical center of consciousness. There will still ever be only ONE Mary-II-B, regardless of whether or not she is created in one brain or across multiple brains. The latter would not be recommended because of the logistical inconvenience that would be associated with the need to have multiple brains constantly communicate towards maintaining that logical, integrated center of consciousness.

[0013] FIG. 2—Demonstrates a sample neural network of 8 brains, including Mary-I's original, that forms an integrated, single, logical, augmented center of consciousness, whereby, upon subsequent physical delinking, while only one (ordinarily the maximum count permissible per this invention's inference) of the 7 other brains manifest Mary-II-B, only 3 of those 7 other brains manifest Mary-II-A and the remaining choose to retain their respective original identities. Arrows indicate direction of flow of information, hence the direction of consciousness transport prospect, albeit the invention recommends dual-directional rather than unidirectional arrows in each case. Again, the at-most-single Mary-II-B brain is only in as much as that only a single, logical, (augmented or diminished) integrated center of consciousness can be created to control one brain or a set of linked brains—and so it optionally is possible to have Mary-II-B to be created across multiple brains that are combined to form a single, integrated, logical center of consciousness. There will still ever be only ONE Mary-II-B, regardless of whether or not she is created in one brain or across multiple brains. The latter would not be recommended because of the logistical inconvenience that would be associated with the need to have multiple brains constantly communicate towards maintaining that logical, integrated center of consciousness.

[0014] FIG. 3—Part of hippocampal circuit pertinent to IBC

[0015] FIG. 4—Part of Papez circuit pertinent to IBC

[0016] For descriptive legends on one or more of the figures see subsection “Descriptive legends on figures” towards the end of the section “DESCRIPTION OF INVENTION”

DESCRIPTION OF INVENTION

[0017] What is self identity and how is that relevant to perpetual persistence of life?

[0018] In the context of this invention it would be relevant to understand precisely what constitutes one's being aware of oneself and hence one's existence. While the views presented in this section are entirely the applicant's own, they do NOT by themselves constitute the invention itself, but instead justify/prove/validate the invention.

[0019] The first step towards that understanding would be to analyze if a human brain soon after conception were in principle medically isolated from all possible forms of input (to the brain) for decades afterwards while still being artificially provided all forms of nutrition to keep it medically alive and cause it to physiologically mature as brains would normally grow would at all have a consciousness and consequently have a sense of self-awareness. In other words, could a human brain develop consciousness when right from its inception it ceased to have absolutely any auditory, visual, olfactory, taste or tactile or absolutely any other form of input—the latter including a sense of pain, pressure, vibration etc. (which essentially are appropriate combined or singular renditions of the fundamental five senses anyway)? Regardless of whether that brain would still have medically “awake” and “sleep” cycles, including cycles of dreams, what could that brain base its consciousness on?

[0020] To answer that, one could ask oneself how one recognizes oneself cognitively immediately upon waking up from sleep. That awareness is from independent, corroborating recent (and sometimes past) memories of incidences, in addition to corroborating new input from the five senses with memories already in the brain of similar prior input (e.g. the visual/auditory input on surroundings being either similar to or different from what is already in memory, etc). Hence there need be some memory for self awareness. Memory is accepted as valid memory (and not imagination) with a high degree of conviction when it can be corroborated with other memory—and therefore the reality of memory is that each experience typically gets laid down in long term memory as a set of independent memories that effectively corroborate (and therefore reinforce/validate) each other—e.g., the memory of having simply sat in a chair somewhere could be recalled from various separate sitting-/time-/surroundings-/etc. angles, hence when recalling that experience one would essentially corroborate across multiple memories of that experience and therefore reasonably assess that one did indeed sit on that chair. In contrast to that, if all one recalled from that experience was just One Single Static memory of sitting on the chair, one would fail to adequately corroborate that memory with other memories of the same experience, and hence one would reasonably question the accuracy of that memory. The fact that most incidences are perceived across a span of time, however short, enables a memory of each such incident to be laid down as a set of corroborating memories, and of course the more intense an experience the greater would be the information content and count of such independent, corroborating memories. It could be noted that the count of memories of any experience would also increase with the intensity of that experience, even if the core incident behind that experience was fleeting. That is because the intensity of the experience would enable the individual to accept input (to be mentally tagged to that experience) on that experience for an appropriate period even after its occurrence. For example, if a person accidentally suffers a cut to a finger from a very momentary lapse of concentration while, say, cutting a piece of fruit to eat, the consequences of that cut would substantially add to the memory content of that experience, which would all be appropriately considered later for independent mental lookup and corroboration towards ascertaining whether one did indeed cut one's finger. As the brain continually lays new memories appropriate to the perceived significance of an experience, the longer an experience persists, and the more significant an experience is assessed to be at and shortly following the time it occurs, the more separate (yet linked, as they relate to the same event/experience) memories of it would be laid down that can later be collectively accessed for a corroborative validation on if the recalled past event/experience's having actually occurred. Each of those memories is built with various layers of the “knowledge framework of rules” (with basic rules at the lowest layers, and at the upper layers are more complex rules created using the relatively lower layers' basic rules and other rules of typical complexity not more that itself) that one gradually first develops at early childhood and continues developing on subsequently for one's entire lifetime (forming more and more complex rules). As particular complex rules are used more often, they themselves get directly interpreted to major extents without requiring referring the lower rules for interpretation, therefore increasing the speed of processing of related complex experiences. In short, self awareness requires memory, and for memory to form there have to be experiences, which are impossible in the absence of external input. It is important to recognize that if there is even a tiny amount of any input given to the brain at any time (e.g. in the case of any individual who might be born regular, but soon looses all physical senses within a day of birth), the brain would take that minimum information that was available for a day and naturally proceed to build up on that towards continually generating newer information databases—and in that case the brain WOULD be conscious. That build up of new information in that stated particular brain would largely be abstract of course as additional external input would be absent that would otherwise have enabled the buildup of a framework of a pool of rules adequately reflective of “reality”, but the brain would still develop an altered consciousness “in a world of its own”. It is significant to note that that state of abstraction is NOT truly unreal per se, but it is abstract in the context of the “real” world that regular humans live in. The human brain is designed to accept external input, however minimal, and, based on its significance, appropriately process the information contained in that input towards not only optimizing the subsequent application of the perceived rule(s) contained in that input, but also develop more complex rules premised on the basic rule(s) as voluntary and deemed necessary choices in anticipation of a need to apply such complex rules.

[0021] Therefore in the complete absence of all forms of external input, the applicant infers that it is logical that that brain would fail to have a meaningful form of consciousness.

[0022] There would essentially be no meaningful functional difference between its states of wakefulness and sleep, even if the brain goes through biological sleep and wakefulness cycles anyway.

[0023] The above conclusion is based on the hypothetical premise of a complete absence of external input into the neurons of the brain since the neurons' conception. In reality, that scenario is virtually impossible, because at the least the neurons Would be able to sense medical variations of the state of the brain itself as “internal” input that would be processed and developed upon towards an abstract perception of consciousness. In fact, even artificially rendering permanent unconsciousness to the brain since inception wouldn't make possible that state devoid of all input, since as long as the neurons are alive they would still automatically collect information from its surroundings, including, from within the brain itself, as well as from within itself. Of course the quality of consciousness would vary depending on the quality of input that it is based on.

[0024] However, the above conclusion is relevant to establishing the fact that being aware of the self and therefore being aware that one is alive mandates that the brain receive external input over a period at some point (at least in the past), as is enumerated further below. The mind of a regular human associates its sense of "I" with a virtually limitless count of memories that it can validate each of either entirely from other memories already in the brain related to the same event/entity/experience and/or with new input from the five senses. For example, one of those associations with one's self identity could be—"I recall having placed my finger on a cube of ice in the freezer compartment of my kitchen refrigerator a week ago, and my finger felt very cold" and "I recall placing my hand on a block of ice in the supermarket ten years ago and my finger did feel cold" and "Let me go touch snow ice that has collected outdoors in my garden this winter morning and let me try to experience if it indeed feels cold to the touch". Each of those two memories correctly corroborate with each other affirming the individual's belief in the validity of those memories and therefore in the individual's identity of self. This is further reinforced by subsequently determining that snow ice feels the same to the touch. When the brain is totally relaxed, yet awake —that bandwidth of the sense of "I" tends to narrow (in the sense that there's a lower count of memories of the self active uppermost in the mind). When there's total amnesia of past memories and a complete loss of the ability to lay down new long term memories, one loses one's sense of identity completely, yet one can still function for basic tasks with short term memory (which do not contribute to one's self identity). A further degeneration of brain abilities would ultimately lead towards a gradual to a complete loss of even the ability to function.

[0025] A highly relaxed, awake human mind is of course not always entirely aware of its identity at any of the uppermost predominant layers of consciousness. Such periods typically are no longer than a short few seconds. That state is different from when the mind is so focused in other deliberations that one's awareness of one's own identity ceases to be a priority consideration for the duration of that focus, which also typically is no more than a few seconds. However, in a matter of seconds, in either case, one's identity comes actively back in one's mind at the appropriate upper level(s) of awareness. Of course since the levels of awareness in one's mind are made up of various layers consciousness (most conscious at the highest and least conscious at the lowest), self awareness is manifested only in the appropriate layers, commensurate with the significance of the need to be self aware (per the ongoing deliberations in one's mind, and where relevant the external input as well). It is to be noted that even while one might not be consciously aware of the more profound aspects of one's identity during the stated periods of intense alternative focus, one still retains a form of self awareness in more simple parameters—by comparing incoming external input with short- and long term memory of the same input. For example, that awareness continuity could be in the form of a breeze blowing across one's face before and then, or some background sound persistent before and then, et al.

[0026] Therefore, what a human cares for in as much as being alive is to be able to CONSCIOUSLY refer back—and therefore naturally corroborate across—as many of those memories as is possible, in the process not only consciously and subconsciously affirming the individual's self identity but laying new memories (in the form of plans, etc) towards feeling positive about one's self identity in the long run on the whole. The latter is of course achieved in a complex combination of learnt behaviors (outside the scope of this Patent Application) that ultimately release the "feel good" neurotransmitter et al chemicals in the brain (e.g.

[0027] (A). learning that society currently affirms that one must love and take good care of one's child and therefore one must do good to one's child, based on which genuine love for the child develops (albeit that love also develops in the absence of society based on other aspects related to memory in the human mind), therefore doing good to one's child or seeing the child happy releases those feel-good chemicals in the brain.

[0028] (B). learning that society encourages success, hence planning for and attaining success releases those feel-good chemicals in the brain.

[0029] Therefore, that self identity (and not the brain hardware associated with it) is what constitutes being a particular human, and it is that self identity that the invention described in this Patent Application focuses on perpetually transferring sequentially, towards life's persistence in perpetuity. To be noted is that brain hardware is of course critical for that identity processing to be done, and subtle variations in the hardware of the brain would cause subtle changes in the perception of one's self-identity even with everything else remaining the same—for example if a pair of (genetically identical) clones were kept in the same environment since inception, they would NOT develop to have the exact identical memory banks or personality. However, the applicant infers that most brains' hardware is acceptably identical within reasonable limits in the context of this processing (distinguishing a human brain from, say, a dog's brain or—as an extreme incompatible example—a latest modern day supercomputer), rendering it possible that most brains be acceptably compatible to process that transferred self-identity in as much as their being able to reasonably manifest the original human's self awareness and traits.

[0030] For feasibility, this invention uses the brain's existing abilities and properties towards that transfer, as elaborated subsequently in this Patent Application.

[0031] The quality of human consciousness containing (by definition) one's self identity, even for the same individual, is plastic. There are multiple, non-discrete qualitative levels of increasing or decreasing consciousness. In all cases when one can reasonably identify oneself, one ultimately normally has a positive experience with one's being alive and awake, hence it would be any of these states of consciousness (where one can meaningfully identify oneself) that would need to be transferred in the context of immortality. However, when there is no sense of self identity at all (when, for example, there's total amnesia of past events associated with a complete inability to lay any new memory) even while the ability to otherwise fully function with short term memory alone is retained, the state would not be meaningful to transfer towards immortality because there is no self identity in such a case.

Key Feasibility Controls

[0032] At the core of this invention are the following controls that render this invention feasible—

[0033] (i.) The applicant's inference that what constitutes being any particular human is Not tied to any one physical brain, but the awareness of memory content and new external input towards direct and indirect self-recognition (i.e. consciousness of the self) and learned reactive rules (i.e. personality, which to a major extent is essentially a manifestation of consciousness of the self) (more importantly the former, because human personality can be modified/learnt later but critical information needs to be retained if self-identifying memories have to be passed on) that can each be Consciously (and Otherwise, primarily the former)—While the Original Brain is Alive—Rendered Familiar with a New Brain's Hardware and its Information Content, Integrated for Functioning (While Still Keeping the Original Information Separately Identifiable) with the New Information Content and Hardware of the Second Brain Along with the Original Brain Hardware. This Process can be Extended to Beyond a Set of Just Two Brains. Consequent to that and the other controls listed in this section the Applicant's Inference that any One Original Brain is able to Consciously (and Otherwise, primarily the former), Following the Stated Integration with One or More Additional Brains (& the Information Contained Therein), Choose One of Those Brains to Base its Consciousness (and therefore personality too) On. That is Based on the Fact that Each Individual Human Brain Constitutes a Reasonable Minimal Hardware that is able to Support Adequate Human Consciousness by Itself, Albeit the Quality of that Consciousness is Expected to be Slightly Inferior to the Integrated Consciousness Attained with Multiple Brains.

[0034] (ii.) The applicant's inference that there cannot be more than one integrated center of consciousness (that, by definition, acknowledges complete self identity) for any living entity at any one time. There can be multiple separate integrated centers of consciousness, each of which could in principle separately take complete control of one's entire core consciousness at different points of time, but only one can he the sole integrated center of consciousness at any one time. To rephrase, There can be at Most be One Logical (and hence PHYSICAL (which crucially doesn't entirely HAVE to be at one location—just that the constituent physical components need to not be physically delinked)) Center of Consciousness for any Living Entity at any One Point of Time. To rephrase, in the context of the human, when a set of integrated brains (manifesting an integrated center of consciousness) are physically delinked, the consciousness of that set can at best be transferred to AT MOST one human brain's center of consciousness (or—if chosen—an integrated, augmented logical center of consciousness formed of interlinked centers of consciousness of multiple regular brains) (and at worst to no human brain at all—the integrated consciousness, would just be destroyed), in as much as a human brain is considered the minimum (but NOT the maximum) physical hardware capable of adequately hosting a comprehensive center of human consciousness that (i.e. the consciousness) can be fully independent of reliance on any other hardware (i.e. any other brain), as is the case with every known regular human brain today. This for example, is to contrast the regular human brain from that of a primate in much as the latter being incapable of physically hosting a human-quality center of consciousness. Again, that is not to state that consciousness WILL be transferred to one brain even if not desired, but only that AT MOST one center of consciousness (spanning one or multiple brains) can have the integrated consciousness transferred to (if applicable/necessary at a slightly inferior quality, but adequate and comparable to the quality of consciousness in any regular human brain). Again, it is important to note that in principle it is 100% possible that if so elected an integrated center of consciousness can be transferred to span across multiple human brains whereby the multiple brains would either together host a SINGLE integrated center of consciousness or independently host separate centers of consciousness only one of which would take full control at any one point of time, however such would render this invention less practicable and less useful. because of the then need for perpetual communication and cooperation among multiple brains that typically would also imply the requirement of the same count of independent, perpetually collaborating humans.

[0035] (iii.) The already publicly recognized ability of the brain to attempt to functionally optimally (i.e. as best as is possible given the tissues and other aspects of the scenario) communicate with neurons not part of that brain, and with the quality of communication augmented in a technology enabled environment. This has been independently evidenced in various attempted-regenerative/“curative” procedures involving the cerebrospinal system.

[0036] (iv.) The brain's already publicly recognized ability to still function optimally (as best as is possible) without the original entirety of the brain, on account of information content multi-duplication and redundancy across the brain, brain hardware redundancy and brain hardware function flexibility. This is evidenced in cases of traumatic brain injury or surgical removal of brain components.

[0037] (v.) The (human) brain's already recognized ability to learn to do just about anything, with and/or without conscious will. There is ho dearth of evidence on the brain's remarkable learning abilities.

Specific Steps in Invention

Step 1—Link Brains to Communicate

[0038] [Key Feasibility Control Applied—(iii.)] (Ref. “Key Feasibility Controls”)

[0039] Mary-I is to be linked to the new brain.

[0040] The intention is that subsequently, per Step 2, Mary-I's brain naturally integrates (enabled by appropriate, minimal external technology support (Ref. “Technology supported interbrain communication”)), as far as optimal logical functioning, with the brain of the second individual towards jointly forming an augmented logical neural network, essentially aimed at rendering the two brains capable of functioning as one, while still maintaining their separate identities at will.

[0041] In order to be meaningful (for example for protection from transmissible terminal illnesses like hepatitis, cancer tissue, HIV etc) this integration will Not mandate a direct tissue connection, but instead can even be achieved with wires, or even entirely wirelessly.

[0000] Step 2—Learn to Use Other Brain, Towards Forming (Together with Mary-I's Own Brain) an Integrated, (Virtual) Single, Augmented, Logical Brain Containing an Integrated, Augmented, Logical Center of Consciousness
[Key Feasibility Controls Applied—(i.), (ii.), (iii.), (v.)] (Ref. “Key Feasibility Controls”)

[0042] This is the most critical component of this invention, whereby each of the two connected brains would, using a combination of the appropriate, minimal external technology support and optimal mental effort by way of using the brains' natural ability to learn and optimally integrate with new hardware, gradually recognize the hardware and information content of the other brain, towards forming a single, augmented, logical brain with a single, augmented, integrated logical center of consciousness.

[0043] That single, augmented, integrated logical center of consciousness will include, in whole or in (typically the most) part, the logical center of consciousness of each individual brain, and (typically) an additional consciousness component originally present in neither of the two original centers of consciousness. That additional consciousness component may preferably be reduced or eliminated for the integrated, augmented consciousness to maintain acceptable fidelity to the logical integration of the original set of consciousnesses.

[0044] The aim of this step is to establish one logical center of consciousness (while still consciously and otherwise retaining the separate identities) for the combined brain at any one point of time, such that it can be switched on or off at will (by the combined single logical center of consciousness, with (importantly—so that the rights and will of each original human are never compromised at any point) each original separate logical center of consciousness retaining full overriding control on its own share of original physical resources and original information content) to return back to the state of the two separate logical centers of consciousness. In short, the combined logical center of consciousness will happen only if both original separate centers of consciousness agree, while never losing awareness of and control on their separate original identities (i.e. on the separate original logical centers of consciousness), and should one of those identities wish to exit the combo it could. The combo would essentially remain an optimally integrated sum of two separately distinguishable parts that could choose to either separate or work together—with each maintaining the critical ability to recognize and influence the other. So there is an element of trust involved, necessitating a correct selection of the two original human candidates, and it would be meaningful that with that selection correctly done the logic of mutual- and overall good would largely prevent infraction of rights by the combined logical center of consciousness of its separate always identifiable component original centers of consciousness: in fact, it would also be easy for each brain's consciousness to willfully deny detrimental “takeover” against its will by another brain's consciousness, much though not exactly the same way for example the Autonomic Nervous System in any standalone individual optimally controls that individual's heart rate, digestion etc without direct conscious control by that individual's consciousness albeit the individual would still be appropriately aware of the individual's heart rate, digestion, etc. In other words, the consciousness combo in the combined logical center of consciousness would remain appropriately aware of each component center of consciousness and be able to control that component center of consciousness to an extent appropriately less than the extent of its awareness of it. As an additional option, that awareness by the integrated center of consciousness of each component center of consciousness and consequently of the human body originally attached to each respective component center of consciousness can be but does not have to be 100%—in other words it could be but doesn't have to be that sections of a component center of consciousness and consequently of the original human body attached to that component center of consciousness remain completely out of bounds of the integrated center of consciousness, not just in Step 2 but even at the conclusion of Step 3. That one logical center of consciousness could be physically located at either brain, or in both combined—of course ideally the former for portability logistics so that Mary-I's original brain hardware could be completely disregarded at will (if so chosen by the stated combined single logical center of consciousness, (pre-) approved by Mary-I).

[0045] After the two brains have been linked to communicate so that the process of logical functional integration could begin, the next step would be for Mary-I to consciously (and sub-/unconsciously) familiarize with the second brain's hardware's peculiarities and with the information content therein, and consequently "learn" to cogitate using "both" her existing brain as well as the second brain's hardware resources and information content. This process could last several hours up to multiple days (in sessions), in the earlier phases of implementation of this invention, and would expectedly very significantly speed up with improvements in supporting technology later. The stated processes of logical integration and learning by both the two brains would proceed at the same time. The neurons of the brains would know how to optimally integrate and function jointly, as long as a potent communication environment (outlined separately in this Patent Application) is provided. At the end of the process, the brains would be deemed to have provisionally been functionally integrated towards optimal logical operations as one (virtual) single, augmented brain. That virtual single brain would have two separate physical centers of consciousness of course (that would, at mutual consensus as part of the process of learning, switch off or on alternately and simultaneously as part of that learning), and Mary-I's brain and the other brain would each learn to recognize the logical consciousness center of the other brain, and as part of the familiarization process grant alternating permission for trial control of consciousness and data (i.e. retrieving and adding memory, processing new input from body etc.) by the other brain's center of consciousness. Essentially Mary-I's sense of self identity would now include recognizing new data and brain behavior that were not originally hers and which she would reasonably "tag" as such (i.e. she would not confuse new information in the second brain with her own original information in her brain, instead she would tag new information as not originally hers and simply learn to deal with that information while maintaining her identity in the integrated logical combo). As part of the integration, she would also learn to recognize and work with the other brain's different physical components and data therein towards manifesting her own original personality, to the extent that her original personality would need to be reasonably exhibited at will following the transfer. While a significant portion of the learning will be in conscious states, part of the learning will be in sub-/unconscious states as well. This process of "learning" towards a logically integrated single (virtual) brain with a single (virtual) logically integrated center of consciousness made up of the always identifiable separate original two logical centers of consciousness (with two physical centers of consciousness that will not always be either simultaneously awake or simultaneously asleep, by mutual consensus) where each brain learns to recognize and work with the other brain's physical components and data therein towards manifesting one's original self-identity and personality and other adequately significant aspects of oneself using one's own and the other brain's hardware, and identifies and tags the other brain's information content as not being one's own original memory, and with the combined single logical center of consciousness maintaining the ability to be switched on or off at will (by the combined single logical center of consciousness itself (or by one/more individual component original logical center(s) of consciousness where applicable), with (importantly—so that the rights and will of each original human are never compromised at any point) each original separate logical center of consciousness retaining full overriding control on its own share of original physical resources and original information content) to therefore enable a return to the original state of two separate logical centers of consciousness (if so desired), is the first significant step of this process. Post transfer, the traits of either Mary-II-B or the original second individual would be manifested at will whenever chosen. In effect, the “better” traits would ordinarily manifest (again, largely at will)—for example if Mary-I was level headed and the second individual originally was less so, then the combo would be level headed at will, and if Mary-I was poor in Math while the second individual was not then the combo would not have to be poor in Math.

[0046] The centromedian nucleus, which is part of the ILN (Intra-Laminar Nucleus) of the thalamus, is already recognized in the medical community today as an important center of consciousness in the human brain. Since, as part of the “learning” phase, the two separate brains would need aim to function as one with preferably hut not necessarily one physical center of consciousness (towards attaining the post-learning-completed requirement of having one logical center of consciousness for the combo at any one point of time, albeit with the combo's never losing awareness of the two separate identities), the applicant recommends enabling neural networking that allows the two centromedian nuclei and other separate components that all jointly constitute the two integrated centers of consciousness in the two brains to both remain functional (to allow for both the independent brains to learn to recognize and “learn” each other and consequently enable both operate jointly as one center of consciousness and also separately as two separate centers of consciousness that each brain would be able to sense both of), as well as alternatively temporarily disabling the integrated center of consciousness in each brain. For part of the process, both brains' centers of consciousness would optionally need to be simultaneously disabled (naturally as sleep, and artificially with technology support) to allow natural, enhanced with technology support, optimized information dissemination from the other brain into each brain. During the process, information from the second brain would be naturally, enhanced with technology support, optimally integrated with Mary-I's brain, and vice versa, resulting in Mary-I's being able to consciously sense and deliberate with memories in the second brain and sense and process input from the second body. The second brain would be able to do the same with memory in Mary-I's brain and input from Mary-I's body.

[0047] Mary-I and the other individual could be geographically located separately anywhere (albeit in the earlier incidences of implementation of the mechanism in this Patent Application they should be in proximity), and at the conclusion of the process of that optimal functional integration, Mary-I and the second human would each have a voluntary (i.e. it can be switched on and off at will), augmented cognitive ability primarily Solely (i.e. in the most part) using the physical resources of the other brain. The centers of consciousness of the two separate brains would also be able to at will operate as one or separately. At that stage, either Mary-I or the other individual, could be relocated to the other brain.

[0048] The applicant emphasizes that in this invention dual-directional transfer is more recommended than unidirectional, towards a higher degree of potency of the process.

[0000] Step 3—Relocate to Second Brain and Physically Disassociate from Original Brain

[Key Feasibility Controls Applied—(i.), (ii.), (iii.), (iv.), (v.). The notable inclusion of (iv.) is relevant for reasons of the reduction (by way of disassociation from the original brain (hence from the center of consciousness therein)) of the augmented, integrated, logical brain (hence of the augmented, integrated, logical center of consciousness) of Step 2] (Ref. “Key feasibility controls”)

[0049] The sequence of the two operations in this Step 3 surmised in the step title above is important—relocation at will happens first, physical disassociation with the original brain happens next.

[0050] With the two brains having already been functionally integrated in Step 2 towards the creation of one logical, augmented, integrated single center of consciousness (that will also have, to appropriate extents, retained awareness of and control on the separate original identities), it would be necessary here to get the two brains' centers of consciousness suitably collaborating towards confirming with Mary-I that she can acceptably sense an integrated consciousness using the centers of consciousness and memory in both the brains and input from both the bodies and then checking that the same responses are consciously made representing Mary-I from the second individual. Next, Mary-I would consciously (and otherwise, with external technology support) actively “relocate” in a highly focused, deliberated and conscious fashion to the augmented, integrated, logical center of consciousness and voluntarily (by focused deliberation or otherwise, with external technology support) disable (or render less relevant) her own brain's center of consciousness thereby gradually relinquishing control of self identity from her own brain and allowing self identity to be processed from the augmented, integrated, logical center of consciousness (which by then would essentially comprise of, in whole or in (typically the most) part, the logical center of consciousness of the second brain, and (typically) an additional consciousness component (that, as indicated elsewhere in this Patent Application, may be reduced or eliminated) originally present in neither of the two original centers of consciousness) such that when her original brain is then disassociated her consciousness (and hence by definition her self-identity) would all the while have remained active in the other brain, she becoming Mary-II-B. Implementing the process while conscious per the critical procedural preference stated above would render the process reasonably safe in the sense that Mary-I would definitively know that she has transferred to the second brain prior to her death.

Step 4—Repeat Steps 1 to 3 for Essential Immortality

[0051] Mary-II-B can repeat the three steps 1-3 at an appropriate time before the physical death of her new body. This cycle of 4 steps is continued, to render Mary-I effectively immortal.

Other Considerations

[0052] To be noted here is that memory/personality information in the sections of the brain that are not relevant to conscious cognition (e.g. breathing, swallowing, etc) will be given a lower priority for transfer. The reason for that is that while those components do have a minor contribution to self identity, that contribution is relatively minimal. Most of these lesser significant to self-identity inadequately cognitive body functions would be directly taken up by the host body's already existing capabilities.

Fair Scope of Invention

[0053] Without limiting itself to (i.e. irrespective of) any specific support hardware, this Patent Application covers the four-step process earlier detailed, even when more than two brains are involved. The latter is possible, per discussions in this Patent Application, if for example Mary-I chooses to link to multiple (and not just one) other brains towards forming an integrated brain with an integrated consciousness center, prior to delinking from them all into one brain (or—not recommended—multiple brains, towards an integrated, augmented center of consciousness) of her choice to form Mary-II-B.

[0054] In short, this Patent Application encompasses not only all morphs/variations/renditions/forms/flavors of the combination of the three steps (Step 1, Step 2 and Step 3) with and without Step 4—enumerated in the section “Specific steps in invention”—but also (all morphs/variations/renditions/forms/flavors) of the following core sequence of three steps Step A, Step B and Step C, with and without Step D (that together constitute an adequate, accurate superset to the steps in “Specific steps in invention”), in the premise of, but not limited by, the applicant's comprehensive logic, details, options and variations elaborated in the entirety of this Patent Application. As is self evident, all applications of existing or future technologies in the context of all morphs/variations/renditions/forms/flavors of the sequence of three steps—Step A, Step B and Step C—with and without Step D, are included in the scope of this Patent Application. Notably, Step B and Step C by themselves or together in any sequence, whether or not preceded by

Step A and succeeded by Step D are also included in the scope of this Patent Application. Step A by itself is covered in the Patent Application only to the extent that it is relevant to medically perpetual persistence of essential human life as detailed in this Patent Application, as well as to the form of connection outlined in (and therefore may be lawfully expanded or extrapolated on as being included in this Patent Application's scope) Step 1 of “Specific steps in invention”. Step A and elsewhere in this Patent Application and to the application of interbrain communicating technologies described in “Technology supported interbrain communication” and elsewhere in this Patent Application. In extension, of “Specific steps in invention” Step 2 and Step 3 by themselves or together in any sequence, whether or not preceded by Step 1 and succeeded by Step 4 are also included in the scope of this Patent Application. Step 1 by itself is covered only to the extent that it is relevant to medically perpetual persistence of essential human life as detailed in this Patent Application as well as to the form of connection outlined in (and therefore may be lawfully expanded or extrapolated on as being included in this Patent Application's scope) that Step 1 of “Specific steps in invention” and elsewhere in this Patent Application and to the application of interbrain communicating technologies described in “Technology supported interbrain communication” and elsewhere in this Patent Application. However the technologies themselves, to the extent not discussed in this Patent Application and to the extent not applied towards all morphs/variations/renditions/forms/flavors of one or more of the three steps below (Step A [except when Step A is excluded from the Patent Application's scope as described above]. Step B and Step C, in any sequence, with or without Step D) would be outside the scope of this Patent Application. Again, the novelty and hence the scope of this Patent Application is not limited to the application of the supporting technologies towards the steps below and (application) to the steps in “Specific steps in invention”, but importantly includes the steps themselves. The novelty of the steps is MORE significant than the novelty of the application of the technologies themselves towards manifesting those steps. However, the fact that the invention entails the use of technology as a necessary aspect of each of the four steps contributes to enabling the invention's lawful patentability.

[0000] A. Mary-I's brain is linked to one other brain, or multiple other brains.

[Key Feasibility Control Applied, Appropriately Extrapolated—(iii.)] (Ref. “Key Feasibility Controls”)

B. Mary-I learns (using a combination of (various levels of) conscious-, various levels of subconscious-, and (various levels of) unconscious-initiatives with technology support) to use other brain(s), towards forming one integrated, augmented brain (or a combination of communicating integrated, augmented brain systems) containing one integrated, augmented, logical center of consciousness (or a combination of communicating, integrated, augmented, logical centers of consciousness, only one of which would have authority over the set of participating brains in that particular augmented, integrated, logical center of consciousness, at any one time). Each integrated, augmented, logical center of consciousness would comprise of in whole or in (typically the most) part, the logical center of consciousness (hence include an appropriate fraction of the consciousness component therein of the associated brain) of each individual constituent brain, and (typically) an additional minor consciousness component (that may preferably be reduced or eliminated for the integrated, augmented consciousness to maintain acceptable fidelity to the logical integration of the original set of consciousnesses) originally associated with none of the constituent brains. Each integrated, augmented, logical center of consciousness will be physically located across one or up to all of the individual single original centers of consciousness in each constituent brains—preferably (for upcoming (per Step C) Mary-II-B's logistical convenience) in exactly one of the constituent brains' centers of consciousness—and only one integrated, augmented, logical center of consciousness may control a constituent brain at any one point of time, even though that brain may be part of multiple separate integrated, augmented, logical centers of consciousness. As part of that learning Mary-I would consciously, subconsciously and unconsciously adequately recognize herself (inclusive of all, most or an acceptable extent of her key original memories) using the (and each, when multiple) integrated, augmented, logical center(s) of consciousness—hence the logical center(s) of consciousness in the other brain(s). That recognition would alternatively be, at Mary-I's conscious/subconscious/unconscious initiative/will, using some/all the centers of consciousness (i.e. at will, including her original own and the other(s) and separately using only the other(s) and not her original own). As indicated at the top of "B", at the end of this step will be formed one integrated, augmented brain (or a combination of communicating integrated, augmented brain systems) including Mary-I's brain, containing one integrated, augmented, logical center of consciousness (or a combination of communicating, integrated, augmented, logical centers of consciousness, only one of which would have authority over the set of participating brains in that particular augmented, integrated, logical center of consciousness, at any one time).

[Key Feasibility Controls Applied, Appropriately Extrapolated—(i.), (ii.), (iii.), (v.)] (Ref. "Key Feasibility Controls")

C. Mary-I chooses a logical center of consciousness if multiple individual logical centers of consciousness have been created in step B., "relocates" to it, logically disassociates from Mary-I's original brain (i.e. from the original center of consciousness of Mary-I) (Note—while a physical disassociation from Mary-I's brain is not mandatory, that will typically ultimately follow), and therefore manifests Mary-II-B within that chosen logical center of consciousness. The logical center of consciousness would ideally be located physically in just one of those other (i.e. non-Mary-I) brains—hence in the center of consciousness therein that one brain—for enhanced logistical convenience of Mary-II-B (but could in principle be located in more than one of those brains as well, as long as they're logically linked forming one integrated, augmented, center of consciousness). At the and of that transfer Mary-I's brain will be logically, and in practice often physically, disassociated from the integrated, augmented, logical center of consciousness that Mary-II-B has relocated to.

[Key Feasibility Controls Applied, Appropriately Extrapolated—(i.), (ii.), (iii.), (iv.), (v.). The notable inclusion of (iv.) is relevant for reasons of the reduction (by way of disassociation from the original brain (hence from the center of consciousness therein)) of the augmented, integrated, logical brain (hence of the augmented, integrated, logical center of consciousness) of Step B] (Ref. "Key feasibility controls")

D. For essential immortality of Mary-I, Mary-II-B will repeat steps A, B and C at an appropriate time prior to the physical death of Mary-II-B's (new) body. The cycle of 4 steps is repeated for as long as Mary-I is wished to be kept alive.

[0055] To clarify scope, while the manifestation of Mary-II-B is complexly covered within the scope of this Patent Application, the manifestation of Mary-II-A in itself is not covered in the scope of this Patent Application as long the scope of this Patent Application detailed in the rest of this section ("Fair scope of invention") is otherwise rigorously applied—to rephrase, the use of morphs/variations/renditions/forms/flavors of the stated four-step process towards the creation of Mary-II-A is covered in this Patent Application's scope.

Proof of Feasibility/Tangibility/Concreteness/Originality/Innovation, Hence Patentability

[0056] The sections "Key feasibility controls" and "What is self identity and how is that relevant to perpetual persistence of life?" outline the primary bases of feasibility. Additional feasibility considerations are in the sections "Recommendations", "Deliberations & clarifications" and "Technology enabled interbrain communication". "Specific steps in invention" and "Fair Scope of invention" make explicit application of the bases of feasibility surmised in "Key feasibility controls".

[0057] The proof of tangibility is of course in the perpetual continuity of life this invention achieves.

[0058] The proof of concreteness is in the specific four steps listed in the sections “Fair Scope of invention” and “Specific steps in invention”.

[0059] The proof of originality and innovation is of course in the very fact that feasible, tangible and concrete perpetual persistence of essential human life constitutes this invention, because perpetual persistence of human life has never been possible yet in any other invention by humankind.

[0060] While the novelty of the overall four steps of the invention is substantially more significant than the novelty of the application of the technologies themselves as part of the invention towards manifesting those steps, the integration of such appropriate technologies as a necessary aspect of each of the four steps of the invention, therefore and otherwise rendering the overall invention original, innovative, feasible, tangible and concrete, contributes to enabling the invention's lawful patentability. Additional details on feasibility could be provided upon request—reflecting that it is appropriate to neither limit the applicable technologies nor immediately build a physical prototype because:

Many different technology options are currently feasible for this invention, with and/or without the relevant follow up insignificant (and/or optionally significant), time spanning, observation and analysis, whereby the immediate patentability of this invention remains entirely undiminished, and is in fact augmented, by such follow up—and other technical considerations.

A complete physical prototype would be feasible if sufficient funding, time and direct and/or at least indirect public will (for example, but not limited to, in terms of a commercial end user market for the invention) are confirmed.

The immediate building of a physical prototype is as unnecessary to affirming the validity (i.e. originality, novelty, feasibility, tangibility and concreteness)—and hence the patentability—of this invention as would be, for example, for any other entirely unrelated invention (not in the scope of this Patent Application) whose originality, novelty, feasibility, tangibility and concreteness are entirely self evident, yet building a physical prototype would require the availability of funds, time and/or at least indirect public will.

Recommendations

[0064] It is important to note that as part of the applicant's considerations on quick implementation feasibility the applicant emphasizes on the following—

Recommends that consciousness be preferably transferred across existing brains of reasonable maturity. Of course the more similar the second brain is to Mary-I's original brain (if largely “regular”), the better would the quality of Mary-II-B be—and if Mary-I's original brain is less than regular (e.g. from a congenital- or injury induced hardware shortcoming) then the target brain to transfer to should ideally be regular. Of course the younger the second individual would be, generally the better, in order to increase the likelihood of minimizing the frequency of such transfers (including, but not limited to, by delaying the onset of normal “old age” post transfer). If inter-brain communication is aggressively tuned per the direction outlined in this Patent Application, the applicant infers that the creation of Mary-II-B will begin being possible any time within 8 or 10 years.

Includes the possibility of rendering consciousness transfer across clones, which in principle would be the ideal form of consciousness transfer. That is because the technology to create clones is available today, and it would be supportive of this invention to create a young Mary-II-B when Mary-I has already reached a relatively senior age.
Strictly excludes fancy considerations like transferring to a computer. That is not practical simply

because the inherent hardware design of the modern day computer vastly reduces the likelihood of the hardware being able to sufficiently multi-process with extremely low consumption of energy to render the computer portable and otherwise feasible. In short, the modern day computer is completely counter to hosting cognition even remotely as optimally as human cognition in the human brain—the best modern day supercomputer does unacceptably more work than the human brain would for the same job.

Deliberations & Clarifications

[0068] Important considerations in this invention are the following—

a. The same brain and body resources cannot be set to be simultaneously controlled by two separate logical consciousness centers. At most only one logical center of consciousness can take control at any one time. In extension, there can at one point be only one (integrated) logical (i.e. constituent components not physically delinked) center of consciousness as far as self identity is concerned. Refer “f.” below.
b. It is not necessary for every bit of information in Mary-I's brain to be transferred to the second human. That is because human memory and personality are both plastic and are changeable. As long as a reasonable amount of information is transferred that would enable Mary-II-B to acceptably recognize herself, that would constitute a good start. Innate human cognitive capabilities would then naturally do the needful to attempt to build up a new memory bank and personality using that information. With the passage of time as newer and faster information transfer techniques are developed, Mary-II-B's brain (excluding prior information already in the second brain) would more closely represent Mary-I's brain at the time of transfer.

c. The technology to clone any human organ is already available today. Therefore, if the center of consciousness (namely the centromedian nucleus and allied components) for any human brain were cloned and connected in parallel to the same endpoints in the rest of that specific brain (namely sending nerve fibers to the cerebral cortex, putamen, caudate, collateral fibers etc. and receiving nerve fibers from the spinothalamic tract, cerebral cortex, superior colliculus, reticular formation, globus pallidus, vestibular nuclei etc.), would the sense of “I” be adequately realized in the cloned center of consciousness such that the two parallel centers of consciousness could operate in the following ways?
1. As one combined augmented logical center of consciousness.

2. As two separate logical centers of consciousness where each center remains aware of the other center and complete control of the resources (and hence of one's consciousness, and hence, by definition, of one's self identity) by any one center at one time is taken by voluntary joint consensus.

[0074] The applicant infers that the answer to both is “Yes”.

[0075] As an extension, “1,” would be possible differently if the cloned center of consciousness were not just connected at the ends with the original but entirely merged with the original.

d. If, in addition to the center of consciousness, the rest of the entire brain were cloned too, such that initially each component still remained functionally part of the corresponding original brain such that the entire brain functioned as a “super” brain with one consciousness, and then if the two sets of components were separated to form two separate exactly similar brains, would the original single sense of consciousnesses be copied in both brains and if so how would it be in relation to the original single sense of identity?

[0077] Obviously, with no physical connection between the two brains, the hitherto similar manifestations of the same consciousness would thenceforth develop into two distinct flavors of consciousnesses in separate ways from the point of separation. The original sense of self identity could remain in any one of the two brains, but never in both the physically disconnected brains at the same time. To rephrase, Mary-II-B can be manifested in at most one of the two physically disassociated brains.

e. Mary-I should select the individual to transfer to well in advance of her anticipated death, and effect transfer of all meaningful long term memory in advance. Following that, Mary-I should periodically transfer further long term memory since accumulated. Ideally this whole process should be undertaken with at least two geographically apart individuals to minimize the risk or not ultimately having any individual available to transfer to just prior to Mary-I's death.

f. Could Mary-I be transferred to multiple third party brains? The applicant infers that the answer is “NO” as far as Mary-II-B, but ‘yes’ in the context if Mary-II-A. There HAS to be at most one single logical (i.e. constituent physical components not delinked) center of consciousness where one can acceptably-self-identifiably transfer to at any one point of time. Therefore, while it is possible for multiple brains to function together as a whole (as earlier detailed), there HAS to be only one logical center of consciousness for them all as far as acceptable self-identification. That integrated logical center of consciousness doesn't have to necessarily physically be entirely in any one of the linked multiple brains, but suitably distributed across them all—non-functional individually but functional as a whole. Therefore when the physical link between the brains is broken—the actual transfer (towards Mary-II-B) would AT BEST be in only one brain—the other physically delinked brains can AT BEST have COPIES (towards Mary-II-A, that, as earlier elaborated, would not be acceptably-self-identifiable) of that original identity where SEPARATE, independent development would proceed from that point of physical disassociation. AT WORST, the transfer wouldn't happen at all, and the original individuals would just be left with new memories that they'd recognize as not theirs (since the memories would not involve them, but primarily Mary-I, and secondarily some of the other people in the set). What would NOT HAPPEN is Mary-II-B's manifestation in more than one separate brains.

g. Could, following a transfer, Mary-I separately sense herself being present in both her original brain and in the second brain that is no longer connected to her own still alive brain? This is an extension of sub-query “d.” and “f.” above, and the answer is “No”—she can be transferred as Mary-II-B to another brain while still alive, as long as she volunteers to let her own brain's center of consciousness give up control of her consciousness to the second brain's center of consciousness.

h. In the context of a brain that has been devoid of all forms of external input to it since inception, would that brain have medically wake and sleep cycles, and if so would the brain have dreams in those sleep cycles? Extending upon the behavior of a human baby in the womb, it can be reasonably presumed that the brain probably would have sleep and wakefulness cycles, simply because that property appears to be genetically so entrenched in the system that it would manifest itself regardless of input. As far as dreams—the applicant infers that an entirely-input-devoid brain does not logically appear to be capable of any cogitation because there would be no basic rules framework in the brain's memory database that can be used to cogitate with—and since cognition appears to be logically impossible, dreams in the way we know it appear to be impossible as well. In other words, there would potentially be no meaningful difference between that brain's state of wakefulness and sleep/dreams. However, this inference is not directly related to this invention.

Technology Enabled Interbrain Communication

[0082] The applicant infers that it is feasible to have two separate brains (learn to) communicate ((at least) initially with active conscious will, to enable the brains learn to recognize the signals and hence the communications mechanism) in a micro-focused electromagnetic communications environment of multiple simultaneous signals of the appropriate strengths and frequencies focused at equivalent brain components in the two brains, as briefly expounded in the following paragraph. Although neurotransmitters could potentially be used to render the process more potent, the applicant deems that essentially unnecessary given that the electromagnetic signals received would anyway enable the natural generation of the appropriate neurotransmitters as the brains deem necessary to process those signals. Essentially, the detected output signals from the relevant components of one brain would need to be appropriately modulated and micro-focused on the appropriate input components of the other brain. That modulated input would be in the form of ultra-low voltage signals at compatible frequencies and strengths (and duration for transfer effectiveness (because transfer will not always be real time, but suitably time lagged mandated by the complexity (i.e. relevant ranges of frequency and strength mapped to physiological zones) of the “sending” brain's emanated signals, the fuzzy logic feedback learning based, identification of the “receiving” brain's internal physiological locations, modulated receiving frequency and strength)). It needs be noted that each of the two brains would alternately be the sending and receiving brain, for purposes of familiarization of the data contents, initially towards the objective of creating the integrated brain system with one center of consciousness, then for Mary-I to transfer onto the other brain towards creating Mary-II-B (and both brains will have already previously agreed to this transfer so that transfer would typically not need to be consequent to a new decision after the creation of the integrated brain). For speedier and greater practical effectiveness, brain-physiologically-functionally meaningful ranges of signals detected in the “sending” brain should be modulated at meaningfully comparable strengths, frequencies and durations at the “receiving” brain at the latter's corresponding physiologically-functional reasonably equivalent micro locations, over electromagnetic waves at least at the “receiving” brain and preferably also detected as electromagnetic waves in the “sending” brain (since electromagnetic waves would enable optimizing (the process of) interbrain signal interactivity over a meaningfully variable range of modulated strengths, frequencies and relevant physiological locations at the “receiving” brain in the context of the counterparts of the “sending” brain). Extreme accuracy of physiological locations will not be important (only) at the commencement of communications, just reasonable micro-focus on output/input brain components, as signal attributes can subsequently be tuned from the feedback of the participating individuals as well as automatically directly from the neuronal responses at the receiving brain, and furthermore the brains would learn to recognize the signals and optimally adjust for the best sensing and interpretation of the signals. Operations research, numerical analysis and stochastic techniques, coupled with neural networks would optionally enable identifying the participant set specific fine tuned communication signal parameters, including, and of course not limited to, the component micro 3D coordinates internal to the corresponding brains. The logistical flexibility of using wireless electromagnetic waves at both the sending and receiving brains as elaborated in this section would enable the aforementioned computation premised search-and-find-for-optimality micro focus to be extended to if necessary the individual neurons, even to the level of the individual dendrite and axon (and allied) connections per neuron, but in the most part that would not be necessary because (a) there is sufficient neuronal redundancy of distribution of the same bit of information and (b) the equivalent location neuron counts across different brains would be substantially different so a super-micro-mapping would be inapplicable. The micro-focus process can even be expanded upon at the end (i.e. after appropriate focus maximization with electromagnetic devices) using atomic- and/or molecular particle based sensing devices, but that too would not be necessary in the most part. Search and find would essentially need to be targeted on the maximum feedback from participants and more importantly on the maximum automated neural responses at the receiving brain corresponding to the signal received from the output counterpart, at the sending brain, upon commencing the process at the approximate focused locations of known input/output counterparts, and honing in with the stated mathematically optimized techniques premised on the nature of non-optimized (only to reduce the search time—search without mathematical optimization would also work but would take longer, potentially infinitely longer in a few cases) searches until the maximal feedback from the participants and more importantly the maximal neuronal responses at the receiving brain is obtained that would indicate a successful link for that specific micro-component input output pair at the corresponding brains. Again, the data transfer needs preferably be in both directions (either simultaneously and/or alternately) so the sending brain receives real time feedback from the receiving brain and adjust suitably. Furthermore, and separately, and as indicated elsewhere in this Patent Application, from a brain learning and managing perspective, the data transfer needs preferably be in both directions (again either simultaneously and/or alternately).

[0083] As also indicated elsewhere in this Patent Application, as part of the connection process, the two linked brains are to be alternately (and sometimes jointly) rendered asleep or otherwise non-willed to allow each brain to learn to be aware of and function with the memory banks and consciousness components of the other brain.

[0084] While electrodes could be useful for physiological zone counterpart mapping at the brains, they aren't necessary and are not recommended for this invention, because of the need to participant-specific-variably-micro-“search”-and-“find” optimized input/output brain coordinates, which would be best possible with micro coordinate focused electromagnetic signals—that can (unlike with electrodes) be easily refocused anywhere—of the relevant frequencies, strengths and durations. Hyper fine tuned (per the above techniques) parallel connections of the consciousness- and memory bank relevant corresponding components (supported by the pertinent functional/communications accessories of course) in the two brains would be an additional consideration. Focus should primarily be on the regular communications channels internal to the brain (i.e. generally excluding the external input path components, for speed) so that key input information is laid down in a way that the receiving brain deems optimal, rather than using attempted “forced” communications via channels not ordinarily used by the human brain which could result in the information not being stored properly.

[0085] To the extent necessary, a chemical connection with neurotransmitters, as well as a direct, neuronal connection could and should also optionally be considered, as a catalyst to manifesting and optimizing inter-brain communication.

[0086] Although memory (and hence also the additional allied) data transfer is from a memory subcomponent buildup (i.e. bottom up) perspective (whereby the memory subcomponents are acquired with equal or more priority as/than are the higher memories that those subcomponents make), alternative consideration would be to allow the sending brain to consciously (and separately with altered, typically diminished, and at times raised levels of consciousness) “load” (i.e. become aware of to a degree of clarity that may be perceived by the (other) brain connected to this brain) relevant memories from its data bank that that brain deems relevant to the transfer and then enable only those memories to be transferred to the other brain. The advantage of this process would be that information would be relatively easily recognized by the receiving brain. The shortcoming would be that the process would be slow, often not include much of the memory (because of the sending brain's need to selectively load a memory first), and needs to be appropriately set up to not compromise or otherwise hinder creating the integrated, augmented, logical center of consciousness —i.e. the center of consciousness of the receiving brain must be suitably tuned to not simply consider the perceived (by it—the receiving brain) memory uploaded by (and at) the sending brain akin to just seeing new images and store that information (in it—the receiving brain) without, attempting to integrate with the sending brain.

[0087] Again, the key here for feasibility would be not to reinvent the wheel, but provide a supporting environment (that can be easily optimized per feedback from the two awake brains) conducive to enabling the brains' neurons to do what is natural to them—optimally integrate and communicate.

[0088] Any interbrain communication already patented, or otherwise not patentable, today is excluded from the scope of originality of this section, but such technologies are firmly denied the right to be applied to any aspect described as being within the scope of this Patent Application (Ref. “Fair scope of invention” and “Specific steps in invention”) without permission from this inventor.

[0089] Albeit already adequately clear above in this section appropriate additional technical elaborations on the above connection technologies may be provided with the non-Patent Application.

SPECIAL ADVANTAGES OF INVENTION

[0090] This approach is in principle a guarantee to perpetual continuity of life, provided of course that a transfer is done prior to Mary-I's death. Even terminal physical ailments in Mary-I would not be impediments to that continuity, as long as Mary-I's brain is still reasonably functional. In fact even mental illness too would be naturally corrected as long as the transfer is done to a brain without mental illness.

[0091] This invention pertains not to the speed of the process or emphasize on specific transfer mechanisms, but is the process itself. With time, as the transfer techniques speed up, the procedure can be applied for emergency circumstances when the Mary-I assesses a high likelihood of imminent death, for example when in an airplane that is about to crash, etc. In such scenarios, Mary-I would immediately wirelessly functionally integrate with the other brain located elsewhere outside the risk zone, rendering her brain and the new brain a single functional entity and “transfer” over.

[0092] The invention can subsequently be easily extended to primates or even animals of lower cognitive abilities.

What this Patent Application Does not Cover

[0093] This Patent Application does not cover the supporting technologies by themselves (that, when applied in the context of this Patent Application, would enable the brains to communicate) other than the supporting technologies that have been described in this Patent Application as being original—in particular in the section “Technology enabled interbrain communication”. Unless the applicant subsequently acquires patents on those technologies themselves, other inventors are free to develop and patent those supporting technologies to the extent permissible by law and use them for purposes other than that described in this Patent Application. However, all applications of current and/or future technology/-ies towards the following are obviously included within the lawful scope of this Patent Application: the scope of manifestation of perpetual persistence of human life as defined in the subsection “Fair scope of invention”. Although as earlier affirmed, the entirety of this Patent Application contains the applicant's own views (except of course the already medically affirmed components of consciousness of the brain comprising of the centromedian nucleus and the allied, and when implicitly or explicitly declared otherwise in this Patent Application) this Patent-Application also does not cover any aspect of the justification/validation/proof of this invention that might already have been previously patented by others in any other context, or is not patentable.

[0094] See “Fair scope of invention” for details on the scope of inclusion of just Step A (“Fair scope of invention”) or Step 1 (“Specific steps in invention”) by themselves in this Patent Application.

[0095] The coverage scope as detailed in “Fair scope of invention” will not be superseded by this section or any other section of this Patent Application.

[0096] Proof of enablement in principle of direct communication between human brains (termed “interbrain communication (IBC)”) towards unlimited extension to human lifespan

[0097] This section (extending for the remainder of the “DESCRIPTION OF INVENTION” of this patent application) (hereinafter referred to as “this section” or equivalent) presents the technical enablement of directly linking two human brains to establish viable direct communication between them, termed “interbrain communication (IBC)” by this inventor (and applicant), in the context of the original patent application filed by the same inventor (and applicant) on unlimited extension to human lifespan. The contents of this section elaborate on aspects of that original invention (of title

"Medically perpetual persistence of essential human life"), that (elaborations and aspects) are either expressly or deemed reasonably implicitly included in the original patent application, hence this section's contents are deemed appropriately protected by patent law towards that same prophetic invention as well as towards the same original patent application and therefore also towards this patent application with the original priority. The IBC mechanism and by and large the overall invention also extend to primates and other animals.

Keywords

[0098] IBC (interbrain communication), dendritic spines, corpus callosum, anterior commissure, event related potential (ERP), artificial synapses, stem cells, telencephalon, diencephalon, thalamus (in particular its centromedian nucleus), hippocampus, amygdala. Independent Component Analysis (ICA), (Multi-way/Kernel) Principal Component Analysis (PCA), Parallel Factor Analysis (PARFAC), signal decomposition, theta (4-7 Hz) and alpha (8-12 Hz) frequencies, neuron-to-glial-cell communication, protocadherin, CREB, NR3A NMDAR, entorhinal cortex, basal ganglia, precuneus, MEF2C, channelrhodopsins, synapse strengthening proteins, etc.

BACKGROUND

[0099] The original patent application for the invention that this section primarily aims at explaining the reasoned feasibility of the IBC portion of is PCT International Application No. PCT/IB2009/052355 filed with RO/IB on 3 Jun. 2009 (hereinafter referred to as "original patent application" or equivalent for the remainder of the "DESCRIPTION OF INVENTION" of this patent application (i.e. this section)) that incorporates by reference the entirety of USPTO Provisional Patent Application No. 61058253 filed on 3 Jun. 2008.

[0100] Not only would the invention enable human lifespan to be extended without limit for a healthy individual—it would in most cases even also enable the same for someone who is ordinarily deemed terminally ill or someone who has severe physical disability.

OBJECTIVE

[0101] This section highlights, and explains the reasoned feasibility of, some of the deliberated core technology approaches towards practically achieving IBC towards the stated perpetual persistence of human life objectives of the cited invention. For the reader's maximum comprehension of the most significant core aspects of the aforesaid IBC technologies, technical terms and non-essential technical details have been largely excluded from this section.

Governing Premises of Invention

[0102] As indicated in the original patent application, the invention's key objective is for perpetual persistence of human life to happen neither by attempting in futile to perpetually extend the life of the human body (which will never succeed) nor by attempting to merely create a biological "clone" of the entire human body or of the memory contents, as even theoretically exact biological clones would still be mere copies of, and not, the original and therefore (clones) would have the critical shortcoming of NOT being able to allow the original human deem the clone as herself/himself (i.e. the biological clone will NOT have the same, but will instead have a mere copy of, the sense of "I" that the original human will have) notwithstanding that that original human and others around that clone would correctly identify the clone as an accurate copy of that individual; therefore the original human would remain as mortal as ever with the aforementioned futile attempts at human lifespan extension regardless of whether or not her/his biological clone is created. Instead the invention aims to enable perpetual persistence of human life by transferring the brain's "self identity" (i.e. "consciousness" of the self, inclusive of consciousness-, subconscious- and unconscious awareness components) and allied essential and non-essential memory data (with that "self identity" and memory data, et al together constituting the same, not merely a copy of the, sense of “I”) to another brain, and in time from that second brain to yet another brain and so on in perpetuity. As indicated above, the phrase “consciousness” of the self purports to include not just conscious awareness of the self, but self awareness at the sub- and unconscious levels as well. One of the fundamental inferences, and therefore one among the core proofs/justifications/validations and governing premises and key feasibility controls, of/in the invention, is that any individual's specific self identity—therefore the specific “sense of I” originally belonging to that individual—could never be merely copied to be “active” while the original self identity is already still active, but can in principle be only transferred to (i.e. be rendered “active” in) only one logical center of consciousness (e.g. in one brain (or equivalent)). Another of the fundamental inferences of/in, and one of the key feasibility controls and governing premises of, the invention is that such “self identity” data is never tied to any specific brain (or other) hardware—in other words any brain is no more than a provisional holding/containing environment for, and is not itself (and is not uniquely necessary to hold/contain), the self identity of the individual who has that brain.

[0103] Another fundamental governing premise and a key feasibility control is that the entire original contents of any human brain are not necessary to uniquely and adequately define (i.e. uniquely identify the external manifestations of the functioning of) that brain—and therefore on extension (uniquely identify) the human who has that brain. Therefore, the entirety of a brain's conscious-, subconscious- and unconscious memory data would not need to be transferred to the brain of another human for adequately “transferring” the self identify of the particular human who has/had that (first) brain. Explained simply, just as when a human loses a limb (or surgically adds new limb(s), or grows massive quantities of new cells across the body as anyway naturally happens to extremely significant extents to every healthy human across the lifetime of that human) the remainder (or the partly or fully upgraded augmented rendition) of the human still remains uniquely and reasonably adequately identifiable (both to that specific individual's own self and to others) as being no other than that same original individual and appropriately functioning with/or without added prosthetics, and similarly just as when portions of a human's brain are lost to TBI (traumatic brain injury) or to surgery (or massive quantities of new brain cells are added to the brain as naturally happens to very significant extents to every healthy human brain across the lifetime of that brain) the remainder (or the updated partly or fully augmented rendition) of that human brain can still remain uniquely and reasonably adequately identifiable (both to that specific individual's own self and to others) as belonging to (i.e. manifesting) that same original individual and appropriately functioning without the individual's losing self identity (except in rare situations when conscious self identity is lost when the brain is too damaged, yet even then aspects of the sub-/unconscious identity and other functions of that brain often still remain sub-/unconsciously identifiable to the self, and consciously to others) with or without corrective/other alterations to the brain, the quality of the success of self identity transfer, hence the degree of effective perpetual persistence of life—i.e. the quality of effective biological immortality—of that specific human, would essentially be a function of the extent of the memories, primarily those related to conscious cognition and secondarily also including those that are not (namely sub- and unconscious memory data), defining that human that are transferred.

[0104] This premise is further justified by the fact that albeit every human's conscious and subconscious (and to relevant extents unconscious) memory contents are substantially different from the corresponding renditions 1 minute or 30 days or 1 year or 10-/20-/50 years ago by way of considerable alterations—primarily additive, secondarily modifying and/or subtractive—to the conscious-, subconscious- and unconscious memory content, each of the varying renditions of that same human's overall memory content essentially fully retains her/his identifying (both to that specific individual's own self and to others) uniqueness and to appropriate extents her/his adequacy of functioning.

[0105] Restating the above further, even when a huge volume of new memory content is added, and in most cases when much of the memory content is modified or lost, the person is able to consciously/subconsciously/unconsciously recognize herself/himself as being the same original person, as too do others around that person.

[0106] While the conscious aspects of memory (i.e. declarative memory (comprising of episodic and semantic memory)) would constitute the most prominent parameters uniquely defining any particular human's self identity, the sub- and unconscious memory components, learnt (i.e. procedural memory) and otherwise (i.e. genetically pre-existing sub-/unconscious “memories”, i.e. genetically pre-programmed neural attributes), not entirely or even at all related to conscious behavior, also contribute to uniquely defining that human's self identity—for example by way or the manner the human walks, laughs or otherwise interacts with society, other unique behavior or mannerisms of that individual, and even less significant, involuntary aspects as how that particular human breathes, etc. In other words, the identifying signature; of not just the brain's conscious and sub-/unconscious “memories” but even of the peripheral nervous system, including both its somatic and autonomic components—the latter's both parasympathetic and sympathetic categories—contribute to that human's self identity, although to a considerably lesser extent than the conscious memories. Therefore while IBC on conscious memory components would be the overwhelming core aspect of this invention, IBC on sub-/unconscious memories would also have a role, and to a very minor extent even IBC relevant to the peripheral nervous system.

[0107] It is important to note that the objectives of the IBC would include both to “transfer” (as in copy) memory data as well as to “transfer” (as in move) self identity. See the four steps listed in the original patent application (Ref. section “Specific steps in invention” on pages 10-16 of that original patent application) body for details.

[0108] For purposes of this invention, the areas of the brain that would be primarily, but not only, focused on towards enabling direct IBC with would be the memory and cognition sites in the telencephalon and diencephalon of the prosencephalon.

Feasibility of Invention

[0109] The original patent application cited in this section lists a total of five (5) key bases of feasibility of the invention (Ref. section “Key feasibility controls” on pages 8-10 of that, patent application). Those details are not comprehensively reiterated in this section; some of them have been (e.g. in section “Governing premises of invention” of this section, and in this section).

[0110] The feasibility of IBC itself has been covered at length in this section.

[0111] Albeit, owing to the complexity of the core innovations constituting the invention, the latter's proof of feasibility cannot be comprehensively expressed in simple terms (the entirety, not mere portions, of the original patent application requires the reader's meticulous, competent comprehension towards her/his comprehensive technical understanding of why and how the invention is feasible inclusive, and because of, its innovations), surmised below is the proof of the invention's feasibility in stripped down, simple, non-comprehensive terms constituting another legal proof of the invention's feasibility without detailing on the innovation aspects of the invention (other than naturally reflecting the cogently compelling innovation quantum extant in the following core approach itself towards enabling perpetual persistence of human life):

[0112] Brain neurons are already naturally highly versatile to essentially unlimited extents in terms of their ability to directly link up to (using newly generated intermediate neuronal pathways and/or existing neuronal pathways), and therefore communicate with, other already existing appropriate neurons in that brain, so there's no reason why, with the right, “reasonable” technology “aid” (i.e. with IBC), the neurons of one brain wouldn't naturally be able to directly link up and communicate with appropriate neurons of another brain. Such forms of neuronal communication are also already noted for example with transplanted organs/limbs.

[0113] Every healthy human brain naturally grows, at the appropriate time points in its lifetime (i.e. ranging from a very young age from before birth up to an elderly age), massive quantities of new neurons and integrate such neurons into that (regularly upgrading) brain causing almost every anatomical area of the brain to grow physically larger than it was before (as happens when the brain is still young) or (at least) be modified even when not physically larger (as happens both in young and older ages), allowing the original memory content and new memory content and (other) brain functions to continue to be appropriately integrated across that being-continually-physically-upgraded brain (with all earlier and newer memory data accessible as usual), and the brain still continues to have a single “logical” (as opposed to physical, as various physical locations in that brain are necessarily involved) center of consciousness—i.e. a well integrated single self identity continues to manifest (where the sub-components of self identity from the integrated memory data are all duly integrated, yet are still appropriately duly separately present for easy natural separate access willfully or otherwise). It is also scientifically well recognized that the extent and efficiency of the process of the overall memory encoding and retrieval and other capabilities of any brain can generally be substantially increased by deliberate natural effort/exercises in the brain. Therefore there's no reason why, with IBC and some deliberate natural effort/exercises in two brains, the same cannot happen with those two different brains—forming one integrated, single, “logical” brain (as opposed to one physical brain, since two physical brains are involved) with integrated memory data (i.e. combined memory data from the two brains, yet all memory data individually accessible post integration) and an integrated, single logical center of consciousness (and a well integrated single self identity formed of the two separate self identities (per the memory data) in the original two separate brains, yet those two separate component self-identities still appropriately duly separately present for easy natural separate access willfully or otherwise).

[0114] Because of the typically considerable multi-redundancy of both data storage and functionality within any healthy human brain, including the ability of the human brain to adjust to new data and develop functions anew when necessary, the healthy human brain is reasonably capable of appropriately adequately functioning even when portions of any of its functional/anatomical areas are naturally discarded or even suitably removed surgically. Therefore there's no reason why the integrated, single, “logical” brain of the above paragraph cannot have the first physical brain physically removed from it (by terminating the IBC following the conclusion of the action surmised in the above paragraph) with no loss to the integrated memory data and the integrated single self identity in the second brain. As indicated in the above paragraph the integrated single self identity and integrated memory data will already include memory and self identity from the first brain (integrated with, yet accessible separate from, the second brain's original memory and self identity), per this invention's objective.

[0115] The memory data contents of any healthy brain are being constantly increased (and to a certain extent modified or lost) across the lifetime of the brain; yet there is no loss to self identity (even if normal multiple (i.e. varied flavors of) “personalities” are present, which is a very common normal phenomenon for most humans). Therefore there's no reason why the integrated memory data in the second brain post-IBC-termination of the above paragraph would in any way compromise the integrated single self identity (duly composed of the two separate original self identities, each individually accessible at the integrated will of that integrated single self identity). As indicated above in the earlier two paragraphs the integrated single self identity in the second brain will have self identity from the first brain along with the self identity that the second brain originally had.

Feasibility of Interbrain Communication (IBC)

[0116] The feasibility of IBC is analyzed in this section under two sub-sections, one on the specific areas of the brain to focus IBC on, and the other on the specific technology particulars to use for such IBC. Some readers might prefer to review the IBC technology subsection before reviewing the subsection on brain areas.

Specific Areas of the Brain to Focus IBC On

[0117] The essential approach in the invention is a 1:1 component-mapped IBC simultaneously for all IBC-relevant components of count n in the two brains, where n<N, N being the count of all (i.e. inclusive of non-IBC-relevant) corresponding components in each of two brains. Additional considerations are to simultaneously, spanning all n components, have grouped m:m, m:M or M:m IBC, where m<M<n<N with the groups being functionally relevant, to allow greater efficiency yet ensuring that IBC data integrity would remain as good as in the 1:1 architecture because any irrelevant input signal (from the other brain) received over IBC at any member subcomponent of any group of m or M components would be automatically recognized and discarded by the brain.

[0118] The brain areas listed in this subsection include a majority, but not all, of the core areas to focus IBC on. Certain areas of the brain are mentioned more than once appropriate to their functional relevance to IBC. The areas are not listed in any particular sequence, other than by and large listing the IBC relevant components in the prosencephalon (constituted by the telencephalon and diencephalon, not necessarily covered in that order), mesencephalon and rhombencephalon in that order.

[0119] Before getting into IBC details per the above ordering it may be noted that the prefrontal cortex, amygdala, hippocampus, precuneus and other areas of the prosencephalon play significant roles in declarative memory; hence those regions would be highly relevant to IBC. The basal ganglia and the cerebellum have roles in procedural memory and hence are IBC relevant, albeit not to the extent of the brain components associated with declarative memory.

[0120] The neocortex (neopallium), hippocampus and amygdala of the telencephalon and the thalamus, hypothalamus, subthalamus and epithalamus of the diencephalon are among brain areas to be importantly considered for IBC. The neopallium subcomponents frontal cortex (specifically including and necessarily not limited to its prefrontal cortex), parietal cortex, anterior cingulate and parts of another component of the telencephalon—the basal ganglia—would be of high IBC significance.

[0121] Sections of the limbic system, namely the thalamus (including but not limited to the centromedian nucleus), amygdala, hippocampus, entorhinal cortex, limbic cortex, anterior thalamic nuclei, mamillary body, fornicate gyrus (inclusive of, but not limited to, its cingulate gyrus and the inferior parahippocampal gyrus), fornix, dentate gyrus, orbitofrontal cortex, and its other constituents would be among IBC priorities. It might be noted that there is not a universally accepted defined set of constituents of the limbic system.

[0122] The amygdala, for its role in modulating memory particularly but not limited to in the context of emotions, would have its IBC priority in this order: cortical nucleus, central nucleus, medial nucleus, basolateral nucleus.

[0123] Dentate gyrae, deemed part of the hippocampal formation, would be IBC relevant because neurogenesis occurs here allowing the formation of new memories, for example for the possible distinction between multiple memories of similar incidents.

[0124] The thalamus forms thalamo-cortico-thalamic circuits that are deemed critical to consciousness, and its role towards appropriate inter-brain memory data interchange with IBC towards individuality manifestation for perpetual persistence of human life would be vital. Among the most pertinent areas of the thalamus would be the centromedian nucleus (part of the intralaminar nucleus), also known as the centrum medianum (Cm-Pf/CM). The CM is known to have both excitatory (e.g. over acetylcholine) and inhibitory (e.g. over Gamma-Aminobutyric acid (GABA)) effect on the, corpus striatum. The medial dorsal nucleus is part of the thalamus and is connected to the limbic system and to the prefrontal cortex, and it has key roles in memory, abstract cognition, planning et al and will be part of IBC considerations.

[0125] The hippocampus is among the most important components relevant to memory formation, and would be a key IBC factor. It is believed that with the passage of time long term memories are formed outside the hippocampus. The key circuitry for the hippocampus involves receiving input from the relatively superficial layers (EC2, EC3) of the entorhinal cortex (EC) and sending output to the deeper layers of the EC. The cornus ammonis 3 (CA3) area of the hippocampus receives input from EC2 as well as from the dentate gyrus (DG, deemed part of the hippocampal formation as indicated previously in this section), while the cornus ammonis 1 (CA1) receives input from EC3 as well as from cornus ammonis 3 (CA3) (through the fibers of the “Schaffer collaterals”). The perforant pathway connects the entorhinal cortex to all the CA components of the hippocampus, as well as to both the subiculum and the dentate gyrus areas of the hippocampus. CA1 also receives input from the nucleus reuniens of the thalamus. Output from the CA1 goes to the deeper layers of the EC, either directly or via the subiculum (which is part of the hippocampus). This is represented in the circuit diagram in FIG. 3.

[0126] It is deemed that areas of the neopallium (a.k.a. neocortex/cerebral cortex/isocortex) and the hippocampus synchronize information and functioning in the theta frequency for various cognitive operations, including but not limited to for working memory activities.

[0127] Therefore the theta frequency would be an IBC synchronization basis for those components across the two brains.

[0128] As has been mentioned in greater detail in the IBC technology section of this section, this invention deems it meaningful to tap directly into the signals at inter-hemispheric neuronal pathways for IBC towards easier access into the respective hemispheres. Important among such IBC relevant inter-hemispheric neuronal connections are the corpus callosum, the anterior commissure and the hippocampal commissure (also known as the commissure of fornix).

[0129] The parahippocampal gyrus surrounds the hippocampus and is one among gyri that are part of the temporal lobe of the cerebral cortex (neopallium), and it is relevant to encoding and retrieving memory, hence pertinent to IBC. The parahippocampal cortex is a specific area combining the (medial) fusiform face area (FFA) and part of the (posterior) parahippocampal gyrus. The FFA complements the function of the PPA (which is part of the parahippocampal gyrus) in that the former is functional in the context of recognizing faces, while the latter is functional for scene recognition.

[0130] As with the entire above portion of this subsection, unless otherwise mentioned, all the components listed below in the remainder of this subsection would play a role in IBC.

[0131] Included in the (anterior) parahippocampal gyrus are both the entorhinal and perihinal cortices, both of which merit IBC. Forming the primary input to the hippocampus the entorhinal cortex (EC) (Brodmann areas 28 and 34) preprocesses the signals input to the hippocampus (for familiarity/recognition). Data connection with the EC is with the hippocampus, the prefrontal cortex, the perihineal cortex, etc.

[0132] The fornix of brain is a route for signals to the septal nuclei and mammillary bodies from the hippocampus, and deemed to be relevant to memory, hence is compatible to IBC considerations.

[0133] The mammillary bodies are material to IBC in that, among other functions, they route signals to the thalamus (via the mammillo-thalamic tract) from the hippocampus and amygdala, and are believed to have a role in recognition memory and smell memory.

[0134] The Papez circuit (FIG. 4)—a key circuit of the limbic system deemed significant to long term declarative memory encoding and (deemed by some) to control emotion—surmises some of the overall key connections of relevance to encoding memory, and hence its components are IBC pertinent.

[0135] The orbitofrontal cortex consists of Brodmann areas 11, 47 and 10 in humans, and is involved in decision making. It merits IBC considerations.

[0136] The medial temporal lobe (MTL) overall is thought to be associated with declarative memory, awareness and learning and would be relevant to IBC. Old memories are distinguished from new by the MTL regardless of consciousness state. Along the same lines, learning is facilitated by the MTL regardless of conscious awareness.

[0137] The basal ganglia is (are) highly germane to IBC, including but not limited to in the context of procedural memory. It is part of the telencephalon and is primarily comprised of the subthalamic nucleus, paleostriatum (a.k.a. globus pallidum), striatum (a.k.a. striate nucleus) and the substantia nigra. The caudate nucleus, which is part of the striatum, has learning and memory function. The putamen, which too is part of the striatum, has a role in various forms of learning, in particular but not limited to category learning, and although the exact functional specialization of the putamen is unknown the fact that it is connected to key brain areas and therefore and otherwise is deemed important to learning renders it notably IBC relevant to the project. The subthalamic nuclei, which are part of the basal ganglia, will also be specifically considered for IBC.

[0138] The primary motor cortex, which is part of the frontal lobe of the cerebral cortex, works along with the pre-motor cortex (Brodmann area 6) towards planning and making decisions. The supplementary motor cortex is relevant to memory and learning. The primary-, pre- and supplementary motor cortices will all be part of the IBC scope because the motor skills that they relate to would constitute part of the self identity signature.

[0139] The superior frontal gyrus—a major component of the cerebral cortex and an important IBC focus—has a key role in self awareness. Importantly, the inferior frontal gyrus is associated with verbal skills (Brodmann areas 44 and 45 constitute Broca's area for syntactic/semantic and phonological processing) and on selective decision making, hence it has IBC relevance vis-à-vis speech and decision signatures. While the superior- and inferior frontal gyri would be relatively important IBC candidates of the project, the middle frontal gyrus would be less so. The fusiform gyrus, by use of its FFA (Fusiform Face Area), would be a notable IBC candidate, for the reason of its playing an important role in recognizing faces and other objects.

[0140] Several among Brodmann areas associated with the temporal lobe would be appropriately assessed for IBC considerations. For example, BA20 (deemed to be associated with recognition memory), BA34 (also part of the entorhinal area), etc are relevant in that context.

[0141] The precuneus of the parietal lobe would have IBC relevance on account of its role in episodic memory, consciousness and self-awareness.

[0142] Among the other areas of the parietal lobe, the primary somatosensory cortex (S1, comprised of BA3-1) and the posterior parietal cortex would be low-to-moderately meaningful candidates for IBC on account of the individual identifying signatures they enable in the context of processing sensory and allied stimuli. The former is significant vis-à-vis sensing touch and other stimuli, and the latter enables planning by appropriately integrating motor and sensory functions. The postcentral gyms of the parietal lobe has functions that have some similarity to those of the posterior parietal gyrus, and would be a moderate IBC candidate for the project.

[0143] The insular cortex has several significant roles, including awareness of the body's internals, awareness of the self and control of emotions, control of motor functions, etc. and it would be an IBC candidate.

[0144] The cingulate cortex would be of IBC relevance, and appropriate of its segments—subgenual area 25 (BA25) (plays a role in the hippocampus pertinent to memory), anterior cingulate (essential to working memory), retrosplenial cortex (RSC) (relevant to memory)—would be given due IBC consideration. Both the parahippocampal place area (PPA) and the retrosplenial cortex are critical to spatial navigation, with the former focusing more on spatially interpreting the immediate local visual data while the latter on placing that data in the overall spatial context. The dorsal anterior cingulate cortex (ACC) effects cognition, while the ventral ACC is pertinent to emotion—so while both would be IBC candidates the former would be relatively more significant.

[0145] Brodmann areas 41, 42 and 22p (essentially comprising the primary auditory cortex and Wernicke's area) of the superior temporal gyrus would be IBC considered. The inferior- and middle temporal gyri would be relevant to some extent in as much as their visual processing (i.e. visual memory) signatures.

[0146] The mesocortical-, mesolimbic- and nigrostriatal dopamine pathways would be IBC contemplated.

[0147] The Nucleus Basalis of Meynert (NBM) of the basal forebrain is deemed relevant to attention, and hence would be material to IBC.

[0148] The temporal neocortex would be looked at for IBC, being where long term memory of the initial temporary memory is formed at the hippocampus.

[0149] The mesencephalon (mid brain)'s tectum has functions related to auditory (inferior colliculus of the tectum) and visual information processing (superior colliculus of the tectum), allowing it to carry information processing signatures (as different individuals would uniquely differently process the same visual/auditory information) that would uniquely identify a specific individual, rendering the mesencephalon moderately relevant to IBC. Much of the mesencephalon other than the tectum is the cerebral peduncle, which is often defined as being inclusive of the midbrain pretectum, tegmentum, and the crus cerebri.

[0150] The metencephalon and myelencephalon—of the rhombencephalon—would not be a priority for IBC but they would not be entirely excluded, because the sub- and un-conscious memory signatures they carry are unique to identifying specific humans. Of these two, the pons and cerebellum in the metencephalon have reasonable IBC relevance. Listed below are some of the functional areas where such signatures exist. The pons routes certain sensory data between the cerebellum and the cerebrum, and the pontine tegmentum of the pons has a role in REM sleep (and therefore in consciousness). The density of P-(ponto-geniculate-occipital)-waves in the pontine tegmentum increases following memory training activity, hence the inference that the P-waves are, and hence the pons is, relevant to learning. The parabrachial nucleus (PbN) has a role in taste sensation, including for signals received from the nucleus of the solitary tract (NTS). The pons also has a role in the synchrony between respiration and cardiac activity characteristics, thereby indirectly constituting one among brain centers associated with involuntary identifying signatures pertinent to IBC.

[0151] The cerebellum carries memory signatures by virtue of its important role in procedural memory, hence is relevant to IBC, in that it, among others, contributes to motor control with perception of sensory data, by enabling a coordinated integration of data exchange between the motor cortex of the cerebrum and spinocerebellar tract. The cerebellum in particular the neocerebellum—also has cognitive functions including but not limited to attention, of some IBC pertinence.

IBC Technology

[0152] i. Core Approach

[0153] For purposes of IBC quality and efficiency, every effort would be taken to refrain from reinventing "the wheel" from ground up, but rather by letting the brains maximally naturally communicate and acquire data content by themselves from each other in a signaling technology environment.

[0154] IBC would be an optimized functional and physical (i.e. anatomical) integrative process. In other words, IBC will be undertaken based primarily on anatomically equivalent locations of the two brains, the selection of which (locations) would be suitably optimized or modified based on functional equivalence, as precise functional equivalence might not always be identifiable by only anatomical or only functional equivalence alone. Towards that integrative process, to identify the approximate functionally equivalent locations, the two brains would be subject to a very wide range of categories of external stimuli (e.g. static and video images, language- and non-language audio and video (examples of language related video would be written text, video of people speaking, visually symbolic language), simple through complex logic processing scenarios that would evoke both regular-data-memory and logical-ability-(i.e. unique logic signature)-memory, olfactory input, tactile input (including but not limited to when combined with language), other memory evoking and allied memory relevant external input, etc) and internal stimuli (i.e. primarily electrical and electromagnetic (and optionally biochemical) stimuli applied internally directly into locations of anatomical relevance (including but not limited to information already publicly known on elaborate input/output sites of various neuronal circuits in the human brain) coupled with the aforementioned earlier-stage location conclusions consequent to the external stimuli) and the corresponding integrated locations and the respective brainwave frequencies would be noted. Those integrated locations in the two brains would then be identified with increasing accuracy in several stages by iterating the above process multiple times. Then IBC would be established between the two brains at those locations at (or otherwise compatible/synchronized to) the appropriate wave frequencies (including but not limited to in synchronized theta (4-7 Hz) and alpha (8-12 Hz) frequency ranges, given that those frequencies are deemed highly relevant to memory encoding and retrieval), strengths and durations—including but not limited to as a variable sequence of "signal bursts" of electromagnetic and/or electrical signals, while noting that IBC would (also) be appropriately directed to the respective input/output neural circuit areas—i.e. the output of a component of one brain fed to the input of the equivalent component in the other brain (instead of and/or in parallel addition to naturally feeding that output to the equivalent component in the same brain itself as well) and vice versa. The beta (12-30 Hz) and gamma (>30 Hz) frequencies would also be used for IBC, both when those are the base frequencies of the corresponding brainwaves at those locations as well as otherwise for augmenting IBC efficiency and quality. The categories of synchronizing the IBC signal with the brains that would be considered would be in-phase synchronization where there would he no phase shift and the frequencies match, out-of-phase synchronization where there would be a phase shift that could be greater than zero and less than half of the phase, anti-phase synchronization where the frequencies would be the same but there would be a half-phase shift, only-frequency-synchronization where the phases could be anything but the frequencies match, partial synchronization where in a large group of neurons some (typically the majority) would operate in synchrony but the rest wouldn't and the whole group would still maintain the overall synchrony, coincidental synchronization where there would be no uniform frequency or phase but nevertheless the probability would remain high that specific events (e.g. overall group of maximum or minimum spikes, etc) of the operations would match the corresponding events of the other, virtual modulated synchronization where the frequencies would be in integral multiples and therefore in a virtual phase. The reason for the need to consider all the above frequencies and phases is that two brains will not necessarily operate in the exact same frequencies and phases even at the earlier mentioned considerably precisely identified functionally equivalent sites: hence optimal frequency-phase-intensity-duration combination(s) corresponding to various functionally equivalent sites across the two brains would need to be identified specific to those two brains.

[0155] As mentioned above, external stimulus alone would not be adequate to evoking a sufficiently large domain of memory. The reason for that is that large chunks of memory would not have been originally created based on external stimulus alone, because much of such memory would have been formed consequent to logically variably internally processing (often involving a huge array and layers of internal logical processing and memory accesses) complex-external-input with already existing internal conscious-, subconscious- and unconscious memory data, hence such memory couldn't ordinarily be invoked by external input alone. Another reason for the need for direct internally applied memory retrieval triggers is that even for memories that have been formed primarily (Note: ordinarily it cannot be only) based on external input alone it would not be always practicable to select, all possible relevant combinations of external input as memory retrieval triggers to retrieve those memories, simply because such external input combination possibilities are too huge and too varied for different people. Accordingly, some of the memory would need to be accessed by direct internal stimulus to (i.e. internal signaling at—typically but not always necessarily electrical or electromagnetic) specific functional locations in the brain itself.

[0156] The synchrony in the partial and coincidental synchronization categories, among others, would be determined specifics of with appropriate numerical analysis, Fourier transform, and other processing.

[0157] Neuronal synchronization between the two brains would be across a massive count of neurons in each at the same time, over artificial chemical and/or electrical “synapses” located between each end of the virtual electromagnetic and/or electrical (and/or chemical) communications (for IBC) channel and the group of neurons at the two brains. All that would essentially be necessary, with relative technical ease, is to electrically detect the real (i.e. neuronal) action potential at the final (group of) neuron(s) in the pre-synaptic end(s) in one brain at the point it is connected to one (outbound) end of the IBC channel and transmit that electrical voltage—not at all necessarily at the neuron-typical +70 mV but at an appropriate (other) voltage—signal to the artificial synapse(s) at the other (inbound) end of the IBC channel connecting the other brain, which would result in the artificial synapse(s) at that inbound end of the IBC channel releasing the appropriate neurotransmitters (for an artificial chemical synapse) causing the triggering of, or simply electrically (for an artificial electrical synapse) directly triggering, an action potential in the first (group of) neuron(s) connected to the IBC channel at that receiving end, which should result in the natural chemical and/or electrical synaptic communication down the IBC link from one brain to the other, and back in a similar process.

[0158] These artificial synapses would be created to function at the appropriate oscillation frequencies pertinent to the categories of the pre- and post-synaptic neurons in the IBC link. Artificial synapses would be complimented by appropriate natural synapses that would (naturally) develop during IBC. It is to be noted that in the context of IBC these natural synapses are those that newly develop between neurons at either of the two brains to facilitate IBC per the brains' natural tendency to transfer, receive and assimilate information. While largely, other than the artificial and natural long-/short term synapses, the remainder of the two brains at the two ends would not need to be physiologically modified too much, some IBC effectiveness- and efficiency enhancing techniques, in part listed below in the remainder of this section, would be considered.

[0159] The corpus callosum is deemed in this invention an optional and/or complementing location for IBC connections. The reason for that is that because the corpus callosum already appropriately reaches into the depths of the two hemispheres of each brain, accessing those hemispherical depths would be appropriately relatively easier by directly accessing the corpus callosum itself. The specific connection points on the corpus callosum to enable IBC at would be identified in a mechanism using external and internal input to the two brains and determining which neuronal areas of the corpus callosum activate and integrating that information and the corpus callosum's already known anatomical (including circuitry) information with the corresponding activation and anatomical information for the remainder of the brain. That overall mechanism would be similar to that for identifying IBC connection points in the remainder of the brain as described in this section.

[0160] The anterior commissure which links the two cerebral hemispheres, albeit to an overall extent considerably lower than that for the corpus callosum, would be another area to focus IBC at, with proceedings as with the corpus callosum. Focus on the anterior commissure would be primarily, but not limited to, in the context of the temporal lobes and the amygdala pair, with those being among the primary brain components relevant to IBC that they connect across the two hemispheres.

[0161] In a manner similar to that for the corpus callosum and the anterior commissure, the hippocampal commissure (a.k.a. commissure of fornix) is important in as much as the hippocampus, because it in effect connects the hippocampus across the two hemispheres.

[0162] For IBC between equivalent components in the two brains it would not be necessary for those equivalent corresponding components to be linked in entirety or even to a highly major extent. When the entirety of two equivalent components in the two brains are not directly linked, the communication bandwidth available to the aforementioned directly linked subset of components would be used by the entirety of those equivalent components to communicate with, wherein data in the non-linked portions of the equivalent components of the brains would pass through the linked portions. In fact extreme accuracy in identifying IBC-relevant equivalent components/locations in the brains would not be necessary—reasonable accuracy would suffice—the equivalent components of each brain would know to filter in and filter out relevant and irrelevant signals from the corresponding equivalent and non-equivalent sections respectively of the other brain towards IBC. In fact irrelevant signals exchanged between even equivalent components across the brains would be similarly naturally discarded. However, the efficiency and to a certain extent the integrity (i.e. inclusive of quality) of information transferred (i.e. the extent to which relevant data is not excluded from transfer, and corruption is minimized of pre- and/or post-transferred data) over IBC would be increased with greater accuracy in connecting equivalent locations. A communicating “band” of reasonably high bandwidth, typically using an appropriately large count of virtual synapses, between reasonably equivalent brain components should be sufficient for IBC. Data would be exchanged across those bands between the equivalent areas of the two brains, somewhat akin to, but less extensively than, how the corpus callosum assists the two hemispheres of a brain to communicate. Just as any healthy brain continually naturally learns how to optimally use the corpus callosum depending on the content and other circumstances of the inter-hemispheric data transfer, the two brains linked for IBC would also naturally learn—with the appropriate technology assistance in the directions outlined in this section and in the original patent application—how to optimize the use of that IBC channel. In other words, IBC would happen, albeit inefficiently arid inadequately, towards the invention's objective of perpetual persistence of human Life, even with a relatively rudimentary IBC communications channel. Yet the more advanced the IBC mechanism per the directions outlined in this section and in the original patent application, the more efficient and adequate IBC would be towards the invention's objective of perpetual persistence of human life. The novelty/tangibility/etc—hence the patentability—of the invention is not only in the technology of the IBC mechanism towards that objective, but also significantly in the very fact that perpetual persistence of human life can at all be plausibly feasibly attained with IBC regardless of howsoever rudimentary (i.e. nevertheless always constituting the invention's use of the technology “tool” IBC) or sophisticated that IBC mechanism is.

[0163] Illustratively expanding with an example on the point of patent law on the logic of patentability of any invention, an ordinary nail in itself is not novel/tangible/etc as an invention—hence not patentable—any longer, but even if that ordinary nail with nothing else could be innovatively used for any purpose that newly achieves an extraordinarily significant, meaningful human objective even if a significant quantity of natural human biology constitutes a major proportion (hypothetically, even if as high as or higher than 99.99% as long as that proportion is less than 100%) of the overall technology process of attaining that objective, the use of that simple nail to at all enable that objective would constitute a patentable invention.

[0164] Various categories of event related potential (ERP)—namely potentials measured in the brains consequent to perception, cognition et al (as part of the earlier discussed overall process of external and internal input towards determining equivalent locations in the two brains for IBC)—would be used to identify the IBC relevant functionally equivalent locations in the two brains. Such ERPs would include, but wouldn't be limited to, N100, N400, P300 (P3), SPS (P600), ELAN (early left anterior negativity), late positive potential (LPP), etc. LPP is deemed relevant in the context of managing emotions, and is therefore deemed pertinent in the emotional memory signature transfer context of IBC. ERP P200 (P2) is deemed significant to the process of intent, and therefore and otherwise in memory. N270, N430 are deemed pertinent to conflict identification and processing. P3a and P3b are cognitively important in that while the former is associated with setting attention to something of adequate newness the latter occurs when unlikely events related to the task already being processed occurs.

[0165] Since EEGs contain information from several ERPs and other signals that are superimposed, thereby muffling the independent ERPs, it would be necessary to suitably process the EEGs with Independent Component Analysis (ICA), (Multi-way/Kernel) Principal Component Analysis (PCA) (including the Karhunen-Loève theorem rendition). Parallel Factor Analysis (PARFAC) a.k.a. Canonical Decomposition, Tucker Decomposition, Adomain Decomposition Method (ADM) for solving partial and ordinary non-linear differential equations using non-numerical techniques. Domain Decomposition Methods (DDM), Cholesky Decomposition, Largest Lyapunov Exponent (LLE), Hurst Exponent, Higher Order Spectra (HOS), et al.

[0000] ii. Optimization

[0166] Signal enhancing techniques need be suitably applied to render IBC faster and more adequate. Techniques surmised below would be among those that would be considered towards that objective.

[0167] Embryonic and adult stem cells as well as progenitor cells enabling the growth of new neurons would enhance IBC connectivity.

[0168] The CREB (cAMP response element binding) protein, deemed relevant in the formation of long term memories, would be assessed for optimized use at both the sending and receiving brains in the project. The memory augmenting effect of pregnenolone sulfate injected into the limbic system would be a factor considered towards expediting IBC. The memory consolidation effect of down-regulating NR3A NMDAR(NMDA (N-Methyl-D-Aspartate) receptor)s would be assessed, in particular, but not limited to, because it is understood that NR3A—a type of NMDAR—causes resistance to the formation of long term memories by way of reducing synaptic plasticity, (reducing) synaptic counts and (reducing) synaptic densities. PolyADP-ribosylation would be looked at in terms of maximizing long term memory formation. The interaction between NMDA receptors and D1 receptors would also be studied in the context of long term memory formation. Also would be assessed the effect (on long term memory) of the induction of AMPA receptors into synapses. Further assessment would be on the relevance (to working memory) of hippocampus dependent, configuration-relational (CR) synchrony in the thereabouts of the theta frequency across the temporal and occipital lobes and non-CR (non-hippocampal-dependent) maintenance across the parietal and frontal lobes, and the effect that alpha oscillations have on LTP. The effect of dopamine on LTP would also be considered. The effect of CAM kinase II on hippocampal LTP would be assessed. It is deemed that theta rhythms and interactions between theta and alpha rhythms are critical to encoding and recalling memory. Therefore substantial focus would be on theta and alpha synchronization for optimized IBC.

[0169] The significance of the MEF2C (MADS box transcription enhancer factor 2, polypeptide C a.k.a. Myocyte-specific enhancer factor 2C) protein to learning in the context of IBC would be investigated. In particular, the deemed reduction of memory (and therefore learning) abilities in specific areas of the brain, including but not limited to the hippocampus, (despite a possible consequent) associated increase in the synaptic densities, by under-expressing (i.e. reducing) MEF2C would be looked into.

[0170] Dendritic spines are deemed to be extremely significant to memory. In part because they're based on actin, they are highly plastic—i.e. they change their shape and size (and count) quickly, in particular the former (i.e. shape) which can change in seconds. The key types of receptors that dendritic spines express are receptors for glutamate and receptors for BDNF (Brain-derived Neurotrophic Factor). Examples of glutamate receptors are NMDAR, AMPAR, etc. Dendritic spines are deemed to be highly significant to cognition and short term memory as well, not just long term memory. Accordingly, a brain's ability to communicate effectively and fast would be directly dependent on the efficacy of its use of dendritic spines. The efficiency (i.e. speed) and extent of IBC would be significantly a function of the optimal production and use of dendritic spines in both brains (Note: not necessarily only the use of existing dendritic spines, but the use of dendritic spines specifically created to enable IBC be more efficient and effective).

[0171] The role of specific neuropeptides towards augmented neuronal interconnectivity would be analyzed, including but not limited to

Galanin: Of the three receptor classes GALR1, GALR2 and GALR3 (all of the G Protein Coupled Receptor (GPCR) class), significant would be GALR3 which is deemed relevant to memory and cognition. Galanin coexists with the neurotransmitter norepinephrine.
Neuropeptide Y: This neuropeptide coexists with the neurotransmitter norepinephrine (noradrenalin), GABA, epinephrine (adrenalin), and it is deemed to play a role in memory.
Substance P: This coexists with the neurotransmitters acetylcholine and serotonin (5-Hydroxytryptamine). It is specifically associated with pain reception. Artificially induced pain, at various locations, to various degrees and for varying durations, would be considered as an IBC enhancing factor for this project.

[0175] As indicated earlier in this section, the optimal production and use of dendritic spines would be critical to optimal IBC, so that huge volumes of memory (and related) data are transferred fast and without compromising the integrity of that data. Of high relevance in this context are the spine apparatus (endoplasmic reticulum) found inside some dendritic spines and the synaptopodin protein that constitutes an important part of the spinal apparatus. The factors that affect the morphology/morphogenesis of dendritic spines would be suitably used towards the production and use of dendritic spines for IBC. Such factors include, but aren't limited to, receptors, proteins for scaffolding, cytoplasmic kinases, enzymes, special proteins and chemicals, and electrical, magnetic and other stimulants, et al. For example, in addition to the AMPA receptor, NMDA receptor etc the SIGMA1 receptor (present at the endoplasmic reticulum) that encodes calcium signaling is deemed to be relevant to dendritic spine creation; enzymes, amino acids and tumor suppressor proteins like Scribble modulates the expression of dendritic spines; melatonin affects spine density; the GluN class (including but not necessarily limited to the GluN2B class) of NMDA receptors are deemed to be highly relevant to affecting dendritic spines in the hippocampus and cortex and thereby affect learning—optimally increasing GluN2B for example could augment IBC efficiency; protein kinase (a.k.a. phosphotransferase, which are enzyme(s) that function to transfer phosphate group(s)) that rely on Calmodulin (calcium binding protein) are also deemed important for dendritic spines in the context learning and therefore to IBC; etc. The effect of three of four classes of the oligopeptide angiotensin—namely classes II, III and IV—would be looked into in as much as enhancing adequate growth of dendritic spines and therefore of the efficacy of IBC. The extent to which the

ABI family of genes is able to affect actin polymerization and therefore (its receptors) affect dendritic spine formation would be a consideration in the project. Estrogen would be looked into in as much as its ability to optimize the morphology of dendritic spines, and therefore its use in IBC.

[0176] The molecular mechanism of storing memory would be appropriately IBC relevant. An overall mechanism of interest is the strengthening of synapses by creating proteins of the likes of Lypla, Cam Kinase, et al. The creation of those proteins happens by the relevant RNA that is otherwise blocked (from creating those proteins) with wrapping proteins, which are degenerated or broken down when new signals arrive freeing up the RNA to create the strengthening proteins. The use of this and allied mechanisms would improve IBC.

[0177] The absence of the chloride channel CIC-2 in nerve cells has been found to make them more excitable, hence in principle more capable of either transmitting or inhibiting information (the latter when such neurons are “inhibitory”), rendering that a possible means of controlling IBC.

[0178] The extent of glutamate released and reabsorbed (post signaling) would be a factor to be considered on how effectively neurons communicate with each other in the IBC context.

[0179] Advanced non-rigid ultra-thin electrodes in the 100 to 500 micrometer thickness range that are better able to mould to the human brain and transmit information much better than common rigid electrodes have already been developed, and these would if necessary assist in IBC.

[0180] Special DNA segments, called “enhancer regions”, are known to cause the release of specific RNA called “enhancer RNA” (eRNA) to release proteins that would augment the creation of synapses across neurons post stimulation, thereby enabling learning, making this of relevance to enhancing IBC.

[0181] The significance of “mirror neurons” that fire both when a human performs a task and the human watches another human perform a task, would be considered in the IBC context. The limbic striatum has been determined to be significant to implicit learning, and hence would be given due focus vis-à-vis IBC.

[0182] The effects of the hunger hormone ghrelin, produced in the stomach and pancreas, on memory and learning (by influencing the hippocampus) is of some relevance to IBC. Ghrelin is known to augment dendritic spine formation in the hippocampus and (thereby and/or otherwise) affect long term potentiation (LTP). IBC learning could be expedited and rendered more effective with the optimal administration of ghrelin.

[0183] The effect of neuron-to-glial-cell communication is pertinent, and would be specifically analyzed, in the context of enhancing neuron-to-neuron communication from the IBC perspective. The mechanisms of such neuron-to-glial-cell communication would include but not be limited to the neuron's synaptic and non-synaptic releases of specialized signaling chemicals, as well as other chemicals (e.g. neurotransmitters, ions, etc released into the glia) that ultimately boost neuron-to-neuron communication. Of particular interest would include, but not be limited to, the use of the calcium ions in the 3-way communication between glial cells and post- and pre-synaptic neurons. The glial cells of the greatest significance in this context are the oligodendrocytes and astrocytes of the CNS. Astrocytes are known to be able to regulate synaptic plasticity, not only as small groups on a single synapse but as larger communicating groups of astrocytes that combined affect a network of neurons. Astrocytes are known to be able to generate Ca<2+> signaling in response to neuronal activation, and its activity can be mediated by suitable agonists and antagonists, among others, thereby enabling an element of synaptic transmission control for use in IBC. Of secondary significance would be the ependymal cells and radial glia of the CNS. Schwann cells of the PNS, and to some extent even the role of microglia would be of some significance. Of added pertinence would be the role of P2X- and P2Y purinoceptors in the communication between glial and neuron cells using nucleotides, where gliotransmitter ATP plays a role. In addition, of interest would be determining if ephrin ligands and eph receptors could be optimally utilized towards neuron-glia communication, for enhancing IBC.

[0184] Fluorescence Resonance Energy Transfer (FRET) and/or allied technologies would be considered for IBC towards live observation of chemical synapse operation where necessary.

[0185] Protocadherin gamma-C3 (PCDHGC3), Protocadherin gamma-A11 (PCDHGA11), Protocadherin gamma-B7 (PCDHGB7), Protocadherin gamma-A12 (PCDHGA12), Protocadherin beta-13 (PCDHB13). Protocadherin beta-9 (PCDHB9), Protocadherin beta-16 (PCDHB16). Protocadherin beta-11 (PCDHB11), Protocadherin beta-14 (PCDHB14), Protocadherin beta-10 (PCDHB10). Protocadherin beta-3 (PCDHB3), Protocadherin beta-2 (PCDHB2), Protocadherin beta-4 (PCDHB4), Protocadherin beta-15 (PCDHB15). Protocadherin beta-7 (PCDHB7), Protocadherin beta-12 (PCDHB12), Protocadherin beta-5 (PCDHB5) would be investigated for relevance to neuron-to-neuron communication for IBC.

[0186] The effect of L-serine and other amino acids on aspects of neuronal growth, hence neuronal communication, would be IBC pertinent.

[0187] The use of light sensitive ion channels called Channelrhodopsins (Chr) towards selectively switching off areas of the brain as part of discerning activation for data transmission would be looked into in the IBC context. ChR2 could be among the more promising in that context, with the other options being ChR1 and VChR1—all of them being sensitive to the cations K<+>, Na<+>, H<+ and Ca><++>. Similarly haldosporin (NpHR) is a light sensitive ion channel that is sensitive to only anion (i.e. negative) chloride ions. ChR2 is sensitive to blue light (maximum absorption at wavelength 480 nm), while VChR1 is sensitive to yellow and could potentially trigger neurons around 580 nm. The use of the Arch and Mac genes, to switch off neurons when exposed to yellow and blue lights respectively, is another effective option, whereby those genes could be loaded onto viruses and inserted into neurons without affecting their operations, and when those neurons are exposed to blue or yellow light energy producing proteins in the cells are inhibited, resulting in the overall temporary deactivation of the neuron, constituting an IBC managing mechanism.

The Reversibility and Multi-Candidature Capabilities of Self Identity Transfer

[0188] With the feasibility of IBC already logically established across the above and earlier sections of this section, this section is not directly on IBC's feasibility but instead on some additional highly significant features of the invention, in the context of IBC.

[0189] Here's the gist, followed by summary elaborations:

[0190] Self identity transfer would be reversible, as long as the first human's body is still alive. That reversibility would be possible both at the end of Step 2 (when the self identity of the first human would be provisionally present in both brains (i.e. when the entire "transfer" of self identity (i.e. rendering self identity present in the second brain along with in the first brain and then removing it from the first brain) will have yet not happened, only "transfer" in the sense of its provisional presence in both brains will have happened)) or if elected at the end of Step 3 (when the self identity of the first human would be present in the second brain only).

[0191] The memory data that would be copied to the second brain as part, of Step 2 could be separately copied to the brains of multiple people as a separate set of Steps 1 & 2 for each, without undertaking Step 3 for any of them. The first person's self identity would remain not fully transferred (i.e. not rendered first present in the second brain along with in the first brain and then removed from the first brain) until Step 3 is completed with any one of them, so there'd be no loss of self identity from the first brain regardless of undertaking Step 2 for several different memory-recipient brains. Because Step 3 wouldn't mandatorily need to be undertaken immediately after Step 2, this would enable a choice of candidates to transfer to over Step 3 at the later time when the decision to transfer self identity over Step 3 is taken. That choice would be meaningful because external circumstances might render some of the memory recipient brains (i.e. those for which Step 1 and Step 2 will have been done) subsequently unavailable/unsuitable when Step 3 is decided.

[0192] As may be expressly inferred from the IBC details in this section, the actual IBC link would not need to be over physical wires, and could be entirely over wireless if so chosen.

[0193] It may be importantly noted that IBC proceedings in the invention would first be undertaken to comprehensively complete memory transfer activities (i.e. undertaken primarily as part of Step 2 of the four steps of the invention)—a process that would take time and therefore would need to be commenced adequately in advance of the actual physical body-death of the first human and (as an added procedural bonus) wouldn't require continuous IBC connectivity (i.e. it could be done in several separate sequential IBC sessions/phases) since memory data once transferred and stored in the second brain wouldn't ordinarily need to be retransferred—and only the core self-identity “transfer” (i.e. Step 3 of the invention) (that would not necessitate much memory data (re)transfer, optionally other than for some new latest and/or earlier memory data if so elected) would be left as the final, separate, IBC activity to be undertaken within a short span of time typically not long before the actual physical body-death of the first human. Because this step would be separate, there would be no need for the two human brains to remain IBC (re)connected until just before the point of that quick self identify transfer, thereby allowing the humans complete freedom to go about their own separate lives following conclusion of the memory data transfer of Step 2, until just before the decision is taken for self identity to be transferred from the first human to the second. Appropriate of the already previously created artificial and natural synapses and other hardware connectivity “ports” created for IBC would then be quickly reused to reestablish IBC towards that final step, to transfer self identity. Also, importantly, because largely no new artificial/natural synapse or other “port” hardware would need to be created for Step 3 of the invention, and since that final act (per Step 3) of self identity “transfer” would essentially constitute enabling access to already-transferred-and-stored-memory at the second brain, temporary simultaneous presence in both brains of the intended-to-be transferred self identity (i.e. the logical unification of the two logical centers of consciousness in the two brains into a single one), and then “switching off” the first brain's component of self identity, the process of self identity transfer would be reversible as long as the first body is physically alive.

[0194] Furthermore, multiple candidates for self identity transfer could be readied in advance (by completing extensive memory transfer to each) to reasonably ensure the availability of at least, one candidate when self identity transfer is finally elected.

[0195] These reversibility and multi-candidature aspects of self identity transfer would be of enormously significant practical benefit, as they would allow anticipatory full transfer (or simultaneous presence in two brains in preparation for full transfer) of self identity (i.e. in reasonable anticipation of possible physical death of the first body, for example if the first human is at the end stages of any terminal illness and therefore the body is likely to die in the imminent/near future, or if that first human is on an airplane that develops critical engine failure en route and appears unlikely to be able to land safely, so self identity is quickly simultaneously established in the second brain (i.e. simultaneous to that self identity also remaining in that human's own brain) over wireless IBC (to prevent disease transmission in the first example (of the terminally ill person), and wireless IBC would anyway be logistically mandatory in the second example (of the airline passenger)) for any one of multiple candidates who is healthy and who is then safely on the ground or elsewhere not on that particular airplane, and that self identity is either deliberately not fully transferred—if ever—until at the last, moment to maximally wait to complete Step 3 to be sure on the likelihood of any impending physical body-death or transferred anyway without waiting further, completing Step 3), and then if so elected reverse the process later while still in the simultaneous-self-identity-in-two-brains stage (i.e. with Step 3 in progress but not completed) or even at a post transfer stage (i.e. with Step 3 completed) should the anticipated physical death of the body doesn't actually happen (e.g. if the person in the above first example somehow recovers from that terminal illness, or if the airplane in the above second example is able to land safely). As indicated above, even if self identity is fully transferred, as long as the first body is still alive self identity can be transferred back in a mechanism similar to but significantly less cumbersome than how the identity was transferred out in the first place.

The Minimalism Option of Self Identity Transfer

[0196] An optional, overall much faster but less comprehensive, form of perpetual persistence of life with this invention using IBC would be to in large part forego the extensive non-essential memory transfer component, and undertake primarily just the self identity transfer component (inclusive of the fundamental, core essential memory absolutely mandatory for self identity transfer) over IBC. Post self identity transfer the individual would still be consciously able to recognize herself/himself in the second brain and also retain the minimal core sub-/unconscious memory (e.g. behavioral) traits and “adjust” into a modified, yet always acceptable (see the final two items of “Feasibility of invention” earlier in this section), rendition of herself/himself with selections of the already preexisting memory data in the second brain, albeit that second brain would lack most of the non-essential memory data that was present in the first brain.

CONCLUSION

[0197] IBC is logically reasonably feasible substantially exceeding the thresholds necessary for medically perpetual persistence of essential human life. That feasibility can be practically demonstrated by implementing the enablement principles outlined in the original patent application and in this section, in an R & D process that would take its due resources and time.

DESCRIPTIVE LEGENDS ON FIGURES

[0198] FIG. 1

[0199] Towards creating the integrated, augmented, single, logical center of consciousness, each of the 7 brains within the dashed box has in this example been shown to be linked only to Mary-I's brain. However, in reality some links could also be chosen to be among themselves (i.e. contained within the dashed box) and not all need to be directly linked to Mary-I. Ultimately, upon delinking them all, ordinarily at most one brain could have Mary-II-B, and the rest Mary-II-A.

[0200] FIG. 2

[0201] In this example, the 7 brains in the dashed box are selectively directionally interlinked among themselves and with Mary-I's brain, towards the formation of the integrated, single, logical, augmented center of consciousness. Note that the directions indicate information transport (hence consciousness transport prospect) direction. Again, upon delinking, ordinarily at most ONE Mary-II-B brain may be formed, and at the rest at most Mary-II-A. Some can choose to retain solely their original identities.

JP2012050872

PREPARATION 2 FOR ETERNAL YOUTH AND IMMORTALITY WITH OUT-OF-THE-BODY EXPERIENCE APPLICATION BY TAKING KETAMINE...

Inventor:KAJISA ISAO

[0001]

The present invention relates to a ketamine dose-body application protein memory material transplant immortality Production Method 2 of human iPS rejuvenation body system drugs.
[0002]

Traditionally, there were technology related to ketamine dose-body application immortality preparation of human iPS rejuvenation body system drugs had no memory material transplantation.
[0003]

"Nikkei Science" January 2008 issue, 21 pages from 20 pages, 76 pages from 68 pages "Mainichi Shimbun" November 21, 2007, page 1 and page 3.
"Separate Nikkei Science perceived phantom" October 15, 2010, "Opening the door of iPS cells regenerative medicine" 114 pp -117 October 13, 2009, 202 pages from the 197-page "Mu" in November 2011 Pat, 85 pages from 60 pages
[0004]

Conventionally, to be achieved through state-of-the-art technology, since there were technology related to ketamine dose-body application immortality preparation of human iPS rejuvenation body system drugs had no memory material transplant.
[0005]

Hemishinku technology or right brain right corner gyrus electrode stimulation and right brain auditory cortex near the electrode stimulation and a video camera system and drug awareness energy that was out of body with ketamine taken by fixing put in infrared sensor in human iPS rejuvenation body body man with similar action is assumed to be all.
In addition it wants to perfect in protein storage material transplant.
[0006]

Body and memory consciousness of iPS immortality becomes possible in reliably-body can be stored substance portable ketamine dose of medicine.
[0007]

Below, it described in the series logically sentence when the embodiment of the present invention. First, two video camera back in Hemishinku technology or right brain right corner gyrus electrode stimulation or right brain auditory cortex near the electrode stimulation or special headset, the right side of the camera to the left of the camera image to the right eye to the left eye, from the body and at the same time stroking the chest of body can be immediately, the two consciousness energy can-body to a different location in a separate experiment ready-body.

Also I can surely-body with ketamine dose of medicine.

In human iPS technology to create a whole body such as organs and nerve to resemble me of young days is allowed to rejuvenation to make the outside of the cyborg.

The human iPS body to consciousness energy to realize the infrared sensor in the rejuvenation in the human iPS flesh and induction to the fixed all can do the same thing as human beings this new universal cells in their skin cells immortality because it can sense the infrared .

The retrovirus culture and cancer gene I prevent from becoming cancer in ultraviolet B wave or vitamin D.

Note To move the storage to young flesh long-term memory in the hippocampus of the brain transplant or aged brain is injected into young flesh of the brain that were made since the memory material is a kind of protein in iPS to extract the proteins of this total storage to memory also I want to transplant.

Human iPS rejuvenation ketamine dose-body application protein memory material of the body system drugs transplant immortality process 2 with the feature as more.

JP2011121988

PROTEIN BEVERAGE ORAL DRUG 2 FOR IMMORTALITY OF LIFE USING HUMAN IPS CELL INITIALIZATION 4 GENE

Inventor: KAJISA ISAO

PROBLEM TO BE SOLVED: To provide a drug which completely achieves immortality of life by recent scientific application. ;SOLUTION: A human iPS cell initialization 4 gene is introduced into a human cultured cell to produce a protein for promoting a change to an iPS cell. The protein is extracted to permeate through the cell of the whole body of an aged man and an adult and to initialize each aged cell as to return it to a young cell. Therefore, immortality of life are achieved by orally drinking, while preventing injury to a brain and a human body, a protein beverage oral drug 2 for immortality of life using the human iPS cell initialization 4 gene

[0001]
The present invention relates to a human iPS cell initialization 4 gene utilization protein beverage oral immortality drugs.

[0002]
Traditionally, there was no technology for human iPS cell initialization 4 gene utilization protein beverage oral immortality drugs.

[0003]
"Nikkei Science" January 2008 issue, 76 pages from 68 pages.
"Newton" February 2008, 75 pages from 70 pages.
"Asahi Shimbun" May 30, 2009, page 9, evening

[0004]
Conventionally, to be achieved through state-of-the-art technology because there was no technology for human iPS cell initialization 4 gene utilization protein beverage oral immortality drugs.

[0005]
Human iPS cell initialization 4 gene to make a protein to promote changes to the introduced iPS cells to cell cultures of human.
It was taken out and drink in a manner harmless to the brain or body is for oral to return until the initialization of the cell aged spreads to cells throughout the body of the elderly human adult young cells.

[0006]
It kept the body of immortality.

[0007]
Below, it described in the series logically sentence when the embodiment of the present invention.
First, let me make a protein that encourages a change in the human iPS cell initialization 4 gene to

introducing iPS cells in cell cultures of human.
From the cultured cells of the entire body of the elderly adult human extract was taken, including the protein cells in the oral order to return until the young cells the initialization of each cell that was aging spreads ride in blood as a protein the brain and the human body harm drink out as no. Human iPS cells initialization 4 gene utilization protein beverage oral immortality drugs 2 with the feature as more.

[0008]
It is carried out as described above.

US2011137782

SYSTEM AND METHOD OF RECORDING, SHARING, AND STORING DATA AND MATERIALS FOR SELECTIVE RETRIEVAL

A method, a process, and a location for a customer to store life-time memories, feelings, emotions, and consciousness, or any other mental activity in a permanent database, based on a financing plan, aided by a collection of personal physical artifacts and personal biological data and materials in permanent physical storage, to share personal and communal information and physical collection with others, to build virtual and permanent personal and collective identities, and to prepare for the possible and probable human immortality where evocation of personal and communal consciousness and information is required when technology avails itself to make our consciousness alive again, either in biological form, in electronic form, in electro-biological form, or in any other possible form, so that our individual consciousnesses will live on both as separate entities and possibly form a unity with one another, in a state of community or singularity, thus achieving immortality.

BACKGROUND OF THE INVENTION

Field of Invention

[0002] This invention relates to a system and method of recording, sharing, and storing data and materials for selective retrieval.

BACKGROUND OF THE INVENTION

[0003] While much debate goes on about the reality and future of human immortality, the latest thinking not only points to its possibility, but almost inevitable probability. To understand the probability of human immortality in one form or another, let's start with the discovery and assertion of Francis Crick, the co-discoverer of DNA, who says this about human consciousness (or the “soul”) in his seminal book The Astonishing Hypothesis: The Scientific Search for the Soul:

[0004] The astonishing Hypothesis is that “You,” your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules.

[0005] Nothing more hopeful has ever come out of human existence than expressed in the quote above. What Francis Crick has pointed out is that, as humans, we will no longer need to rely on the existing 3 biological form we have taken to prolong our existence. Why? If our consciousness can be understood to be “the behavior of a vast assembly of nerve cells and their associated molecules,” then who is to say that this assembly of nerve cells and their associated molecules cannot be recreated or duplicated and then maintained to sustain our consciousness forever, thus enabling us to live forever? Who is to say that this behavior of a vast assembly of nerve cells and molecules cannot be achieved and maintained in some electronic or electro-biological form?

[0006] Ray Kurzweil, of Kurzweil Industries, has authored numerous books on the future development of the human race and on how technology will soon enable us to live forever. To his credit, he is widely known to have predicted the fall of Communism in Eastern Europe in his hooks; and there are many other predictions by him that have come truth. Bill Gates declares Ray Kurzweil to be the best person I know at predicting the future of artificial intelligence."

[0007] Here is the prediction of Ray Kurzweil about how technology will usher in human immortality in his best-selling book The Singularity Is Near:

[0008] . . . within several decades information-based technologies will encompass all human knowledge and proficiency, ultimately including the pattern-recognition power, problem-solving skills, and emotional and moral intelligence of the human brain itself.

[0009] . . . We will gain power over our fates. Our immortality will be in our own hands. We will be able to live as long as we want . . . . We will fully understand human thinking and will vastly extend and expand its reach. By the end of this century, the non-biological portion of our intelligence will be trillions of trillions of times more powerful than unaided human intelligence.

[0010] [When we reach the state of Singularity,] it will represent: the culmination of the merger of our biological thinking and existence with our technology, resulting in a world that is still human but that transcends our biological mots. There will be no distinction, post-Singularity, between human and machine or between physical and virtual reality. If you wonder what will remain unequivocally human in such a world, it's simply this quality: ours is the species that inherently seeks to extend its physical and mental reach beyond current limitations.

[0011] Essentially, these two experts, along with numerous others, have pointed to one unavoidable course of human development: our biological existence will soon be augmented or replaced by a non-biological or electro-biological existence, in one way or another, which may well usher in the age of immortality.

SUMMARY OF THE INVENTION

[0012] Essentially, Immordata is a method, a process, and a venue for our customers to store their electronic and biological data, as well as personal physical artifacts and biological data, for the purpose of creating an immortal consciousness, to share personal and communal information with others, to build virtual and permanent personal and collective identities, and to prepare for the possible and probable human immortality where evocation of personal and communal consciousness and information is required.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] FIG. 1 is a schematic diagram illustrating an exemplary system for data gathering, storing, and sharing;

[0014] FIG. 2 is an exemplary website for the system interface between customers and data;

[0015] FIG. 3 is an exemplary website interface where customers can to in to access the customer's account;

[0016] FIG. 4 is an exemplary website interface where customers may choose a data category to enter or retrieve the customers' data;

[0017] FIG. 5 is an exemplary website interface where customers may designate the category in which the customers may enter data;

[0018] FIG. 6 is an exemplary website interface where customer may record, enter, and retrieve audio data;

[0019] FIG. 7 is an exemplary website interface where customers may record, upload, and retrieve visual data;

[0020] FIG. 8 is an exemplary website interface where customers may record, upload, and retrieve other electronic data;

[0021] FIG. 9 is an exemplary website interface where customers may store and retrieve personal physical artifacts;

[0022] FIG. 10 is an exemplary website interface where customers may store and retrieve personal biological data and materials;

[0023] FIG. 11 is an exemplary website interface where customers may create various communities of any kind and size; and

[0024] FIG. 12 is a flowchart: detailing the method and system in which customers create accounts and communities, store data, share data, and retrieve data.

DETAILED DESCRIPTION

[0025] Here is the detailed description of how the Immordata will work, referring to the drawings, wherein like reference numbers refer to like components, a selective storage and retrieval system is shown at 100 in FIG. 1. The selective storage and retrieval system includes a network 102 operatively connected to a database 103. One or more users 101, such as a User A 101A, a User B 101B, a User C 101C, and a User N 101N, are in communication via a network with a network engine 102. The network engine 102 may include a website, one or more servers operating the networking website, or any other device or application capable of providing networking services via the network 102. The users 101 can access the services provided by the network engine via the network 102. Optionally, one or more of the users 101 may be able to access the network engine 103 directly.

[0026] The users 101 can access the network engine 102 via a computing device, such as a laptop or desktop computer, a cellular telephone, a personal digital assistant (PDA), a set top box, and so forth. The users 101 can view data about social network members via the network engine 102, enter data about themselves and possibly others, join network groups, and so forth. The users 101 are typically members of a networking website associated with the network engine 102, for example, and thus include the network members.

The Immordata Process, Method, and Other Elements:

[0027] 1. Immordata customers enter the website by using the process detailed below:
1.1. When customers enter the Immordata website, Immordata's home page will greet them, as shown in FIG. 2.
1.2. Immordata customers will sign in by using the sign-in website set up for the customers, where customers will need to use a User Name or an Email Address and also a User Password in order to enter the website, as shown in FIG. 3.
2. The Immordata Website: Immordata will establish the website for data storage, data sharing, and the building of Immordata individuals, families, communities, cities, states, ethnicities, countries,

the United Nations, and possibly the entire immordata humanity:
2.2. Web portals will be established as portals for customers to enter 1201, store, update, and share their written, audio, video, and all other electronic and personal data via the web interface, as shown in FIG. 12 at 1205. The data will be segmented into three distinct categories as shown in FIG. 12 at 1204: 1) one for public sharing, so others may have access to various published data, 2) one for sharing only within circles of family, friends, or any other community, and 3) one for private documentation that will only be viewed by each individual who enters the data and can be brought up as permitted by the executors of Immordata in the future for eventual establishment of individual immortality, as authorized by the individual in legal agreements.
2.3. Data submitted for publication will be subjected to data administrators' censorship as in FIG. 12 at 1220 and 1222. Once the submitted data passes censorship, the data will be published either to the general Internet public 1214 or to communities 1221 lithe submitted data does not pass censorship, the data will be permanently deleted 1218 and 1229 and the customer will be notified 1210.
2.4. Customers will be reminded to store their data permanently 1211 and customers can then submit data for permanent storage as in FIG. 12 at 1213 and 1215, but customers must sign up for financing plan, as shown at 1216, 1217 and 1223. If the financing plan is set up, then data is stored permanently 1213 and 1215
2.5. The websites will also enable our customers to build their communities 1226 as specified below, be they families, communities, races, ethnicities, countries, or all of humanity. As shown in FIG. 12, customers will be able to share their personal information to build communities of their own as they see fit at 1202. The data for these communities can be permanently stored or not, depending on the agreements between Immordata and these communities.
2.6. The websites will also direct our customers to store their Personal Physical Artifacts in storage centers across the globe in secure, locked-down, fire-proof, and quake-resistant facilities and the websites will also direct customers to store biological materials in biological storage in Immordata biological storage centers across the globe in secure lockdown, fire-proof, and quake-resistant facilities as shown in FIG. 12 at 1203 and 1225. Customers will be required to set up a financing plan 1224. If the customer does not have a financing plan set up, the customer will be rejected and notified 1227; if the customer has a financing plan, then the customer will be able to store physical artifacts and biological materials 1230 and the customer will be notified of the storage 1228.
3. Immordata process and method: This is the process and method for customers to store their daily activities, memories, feelings, emotions, and reflections, etc. Furthermore, Immordata customers will be able to save their personal physical artifacts and biological data as instructed by Immordata. The process and methods are listed below and shown in FIG. 4:

[0028] Data Category 1: Diaries Eternal

[0029] Data Category 2: Audios Eternal

[0030] Data Category 3: Visuals Eternal

[0031] Data Category 4: Other Personal Electronic Data

[0032] Data Category 5: Personal Artifacts Eternal

[0033] Data Category 6: Biological Data Eternal

3.1. Immortal Individuals:

[0034] 3.1.1. Immortal Written Entries: this is probably the best, the most direct, and the most intimate way to record one's consciousness for eternity. Customers' written entries of their memories and feelings, completed daily or otherwise, shall be stored permanently, to be used by our customers to build their own eternal consciousness and to build immortal communities of various kinds and sizes. As shown below, the contents of the immortal written entries will be categorized into three: 1)

Public, 2) Access Only for Families and Friends and Communities, and 3) Private. This function and the categorization will define all other data below to be stored with Immordata as shown in FIG. 5. The descriptions of the three categories will be further defined later in this patent application.

3.1.2. Immortal Audios: This shall be the voice recordings of our customers or other sounds and voices; the data is less direct, but it has dimensions unavailable in written words; it also records our biological vocal dimension and enriches the contents of the recorded consciousness. The contents of this kind of data will also have three categories: Public, or Access Only for Families and Friends and Communities, or Private. The audio recordings shall be stored permanently, to be used by our customers to build their own eternal consciousness and to build immortal communities of various kinds and sizes. Immordata customer will enter, modify, or retrieve their data by using the Immordata website as shown in FIG. 6.

3.1.3. Immortal Visuals: These are pictures or videos of customers; this data adds an even richer form of recording our consciousness, especially when we are recorded to discuss our memories, feelings, emotions, and reflections, or when they record other activities and observations. It also adds physical dimensions unavailable in audio and written recordings. The contents of this kind of data will also have three categories: Public, or Access Only for Families and Friends and Communities, or Private. These data elements are valuable in case our immortality may require sonic physical dimensions to establish our immortality. The visual recordings shall be stored permanently to be used by our customers to build their own eternal consciousness and to build immortal communities of various kinds and sizes. Immordata customer will enter, modify, or retrieve their data by using the Immordata website as shown in FIG. 7.

3.1.4. Other Electronic immortal Data: Customer may record any other electronic data they deem necessary to record their consciousness, such as medical records, publications of relevant subjects or events, or whatever electronic data. This data will enrich recordings of customers' consciousness. The contents of this kind of data will also have three categories: Public, or Access Only for Families and Friends and Communities, or Private. These recordings shall be stored permanently to be used by customers to build their own eternal consciousness and to build immortal communities of various kinds and sizes. Immordata customer will enter, modify, or retrieve their data by using the Immordata website as shown in FIG. 8.

3.1.5. Immortal Personal Artifacts: This will be a collection of personal physical artifacts, in limited spaces, to aid in the revival of personal consciousness of our customers. These personal physical artifacts, be they scrapbooks, old photos, letters, small collections, samples of any materials things with personal connections, will be stored permanently for customers, to be used by our customers to build their own eternal consciousness and to build immortal communities of various kinds and sizes. Immordata customer will enter, modify, or retrieve their data by using the Immordata website as shown in FIG. 9.
3.1.6. Biological Data Eternal: This will be a collection of personal biological data, to aid in the revival of personal consciousness of our customers. This personal biological data, whatever form it may take (DNA, tissue, MRI, etc.), will be stored permanently for customers, to be used by our customers to build their own eternal consciousness and to build immortal communities of various kinds and sizes. Immordata customer will be able to update their biological information continuously. The process for biological data storage is shown in FIG. 10.

3.2. Immortal Families: The data for the Immordata Families can be permanently stored or temporarily stored, based on the agreements between Immordata and the families:

3.2.1. Family Trees: Families can use the Immordata website and process to build their family trees, to share their activities, news, thoughts, photos, videos, physical artifacts, or any other such information as specified in the content categories later in this application. Access to the data can be categorized to allow different access permissions, such as public or family only.

3.2.2. Family Rooms: Families can use the Immordata website and process to share information on their family trees, to conduct live gathering on the web and record their activities, thoughts, feelings, voices, and all other information as specified in the content: categories later in this application. Families may also use the Family Rooms to record their birthdays, anniversaries, special events, etc., and Immordata will provide the service to send reminders to specified customers for those dates and events. Families may also use the Family Rooms to plan travels, events, and gatherings; and then they may use the Family Rooms to upload event photos, videos, audio recordings, and descriptions, etc., to be share among family members. Access to the data can be categorized to allow different access permissions, such as public or family only.

3.9.3. Family Eternal Child-Raising: As children, up to a certain age, cannot record their activities, families can use the Immordata website and process to record the activities of their children, to be shared with the children when they are old enough to record their own activities, memories, feelings, thoughts, knowledge, etc.

3.2.4 Family Reunions: Families can use the Immordata website and process for electronic family reunions, via web casts, live videos, or conference calls, on the web and then record their activities, thoughts, feelings, voices, and all other information as specified in the content categories later in this application. Access to the data can be categorized to allow different access permissions, such as public or family only.

3.2.5. Family Medical and Genetic Records: Families or individuals can use the Immordata website and process to record medical and genetic information, which will be accessible only to the individuals themselves or to others who are given the access by the collective family or their doctors who are granted the access by the collective family decisions.

3.2.6. Family Living Will: Families can use the Immordata website and process to record their will. This data will be accessible only to the family members who are given the access by the collective family or their lawyers who are granted the access by the collective family decisions or the government as mandated by the applicable law under the appropriate and applicable jurisdiction.

3.2.7. Families Immortal: Families can use the Immordata website and process to record and store immordata for the entire family and record their activities, thoughts, feelings, voices, and all other information as specified in the content categories later in this application. Access to the data can be categorized to allow different access permissions, such as public or family only.

3.3. Immortal Communities:

[0035] 3.3.1. Immortal Hoods: Individuals can use the Immordata website and process to build their immortal communities, be they neighborhoods, groups and communities of whatever specialty, such as artists, engineers, writers, inventors, etc., to share their products, activities, news, thoughts, photos, videos, physical artifacts, or any other such information as specified in the content categories later in this application and record their activities, thoughts, feelings, voices, and all other information as specified in the content categories later in this application. They may also use the immordata website to plan events and make announcements. Access to the data can be categorized to allow different access permissions, such as public or community only.

3.3.2. Immortal Cities: Individuals or cities can use the Immordata website and process to build their immortal cities, to share their products, activities, news, thoughts, photos, videos, physical artifacts, or any other such information as specified in the content categories later in this application and record their activities, thoughts, feelings, voices, and all other information as specified in the content categories later in this application. They may also use the Immordata website to plan events and make announcements. Access to the data can be categorized to allow different access

permissions, such as public or community only.
3.3.3. Immortal States: Individuals or states can use the Immordata website and process to build their immortal states or provinces or prefectures or whatever administrative regions, to share their products, activities, news, thoughts, photos, videos, physical artifacts, or any other such information as specified in the content categories later in this application and record their activities, thoughts, feelings, voices, and all other information as specified in the content categories later in this application. They may also use the Immordata website to plan events, conduct business, and make announcements. Access to the data can be categorized to allow different access permissions, such as public or specific community only.

3. Immortal Universities:

[0036] 3.4.1. As one of the purposes of Immordata is to establish a world in which learning will be for eternity for all interactive immordata individuals, it is fit and proper that Immortal Universities are to be established to benefit Immordata individuals. The purpose here may not be a diploma, but, instead, the Immordata individuals will matriculate into better beings all around, transcending the biological limitations.

3.4.2. Existing colleges and universities will set up their Immordata Colleges, Institutes, Universities, or any learning institutions of all kinds at the Immordata website to offer their education, for a fee, and to become a part of the Immordata world.

3.4.3. They may also use the Immordata website to plan events and make announcements. Access to the data can be categorized to allow different access permissions, such as public or community only.

3.4.4. Immordata will have the option to create its own Immordata University or any other higher education institution and offer its higher education services to Immordata customers and individuals.

3.5. Immortal Races and Ethnicities:

[0037] 3.5.1. Immortal races: Individuals or collective races can use the Immordata website a id process to build their racial communities, to share their cultural data, activities, news, thoughts, photos, videos, physical artifacts, or any other such information as specified in the content categories later in this application and record their activities, thoughts, feelings, voices, and all other information as specified in the content categories later in this application. They may also use the immordata website to plan events, conduct their business, and make announcements. Access to the data can be categorized to allow different access permissions, such as, public or community only.

3.5.2, Immortal ethnicities individuals or collective ethnicities can use immordata website and process to build their ethnic communities, to share their cultural data, activities, news, thoughts, photos, videos, physical artifacts, or any other such information as specified in the content categories later in this application and record their activities, thoughts, feelings, voices, and all other information as specified in the content categories later in this application. They may also use the immordata website to plan events, conduct their business, and make announcements. Access to the data can be categorized to allow different access permissions, such as public or community only.

3.6. Immortal Countries:

[0038] 3.6.1. Immortal countries: Sovereign countries or states can use the Immordata website and process to build their country or state communities, to share their data, activities, news, thoughts, photos, videos, physical artifacts, or any other such information as specified in the content categories later in this application and record their activities, thoughts, feelings, voices, visuals, and all other information as specified in the content categories later in this application. They may also use the Immordata website to plan events, conduct their business, and make announcements. Access to the data can be categorized to allow different access permissions, such as public or community only.

3.6.2. Immortal United Nations: The United Nations can use the Immordata website and process to share their data, activities, news, thoughts, photos, videos, physical artifacts, or any other such information as specified in the content categories later in this application and record their activities, thoughts, feelings, voices, and all other information as specified in the content categories later in this application. They may also use the Immordata website to plan events and make announcements. Access to the data can be categorized to allow different access permissions, such as public or community only.

3.7. Immortal Humanity:

[0039] 3.7.1. This is far-sighted, but if the Singularity occurs, as predicted by Ray Kurzweil, the entire Singularity may use the Immordata website and process to share their data, activities, news, thoughts, knowledge, photos, videos, physical artifacts, or any other such information as specified in the content categories later in this application and record their activities, thoughts, feelings, visuals, voices, and all other information as specified in the content categories later in this application.

3.7.2. Were that to happen, the Singularity may use the Immordata website and its process for whatever purposes, but within the confines of the law at the time when the Singularity occurs.

[0040] In FIG. 11, there is a preliminary webpage that enables Immordata customers to form their Immortal Families, Communities, Neighborhoods, Groups, Cities, Countries, Ethnicities. States/Provinces/Prefectures, United Nations, or whatever community, or the Singularity, as shown in FIG. 11.

[0000] In FIG. 12, a high-level flow chart delineates the process that governs the operations of Immordata process.

4. The Data Base Storage: Data storages will be established to permanently and securely store our customers' data resulting from our customers' writings, recordings, and collections.

4.1. Electronic data will be stored at the secure sites. Security will be of the uttermost importance, so the data storage will be established in the most secure locations.

4.1.1. Legal agreements will be established with customers on how data is to be shared or protected, and on how the eventual data will be used to establish individual immortality by using the data of their consciousness. The latest technology is available to store electronic data for thousands of years, and future technology will avail itself to preserve the data permanently.

4.1.2. Data Integrity: Immordata will establish a process in which the data will be stored permanently; Immordata will make sure Immordata data will forever be securely stored and will be safe to retrieve by ensuring all technological updates for the data storage.

4.2. Physical collections of Immordata customers will be stored in secured, fire-proof, water-proof, quake-resistant, and damage-proof storage facilities, to aid in the eventual evocation or retrieval of the consciousness data, in order to establish individual immortality by using their data of their consciousness. Security will be of the uttermost importance, so the data storage will be established in the most secure location.

4.3. Biological data: In case the future requires biological data, permanent DNA data and other biological data will be stored to enable customers to store their biological data, in case the eventual building of individual immortality needs the biological data. At the suitable time, Immordata may choose to form its own biological data storage company, so our customers will trust that their biological data is secure with Immordata, like the other data they have storied with Immordata,

4.4. Financing Plan:

[0041] 4.4.1. A financing plan will be established as an option for Immordata customers to enable the permanent data storage, until such a time as it is viable to transfer the Immordata from storage to a non-biological or altered biological form to let it: live permanently. The financial cost for permanent storage of all this data will be born by the principal amount in the form of an endowment and the resulting interest, as established by the financing plan.

4.4.2. Some of the Immordata costs will be borne by the income from advertising revenues as part of the Immordata is made public for viewing, discussion, and exchange.

5. Evocation of Immordata to Build Immortal Individuals, Families, or Communities of Varying Scales

[0042] 5.1. immordata will select an executive committee of the best scientists, community leaders, and prominent citizens in the world and form the Immordata Congress, whose responsibility is to assess the maturity of the technology to build and sustain immortal individuals in the future. The structure and decision-making process shall be determined by Immordata.
5.2. The decision of the Immordata Congress will be the decision Immordata will follow to evoke the private and public data to build immortal individuals.
5.3. The process to execute this decision will be outline by the Immordata Board of Directors and its executive officers, to be detailed in its company charter.
5.4. Immordata shall abide by all laws in the community in which it operates for all its operations and activities, without exception.

Further Definition of Immordata Contents:

[0043] This section is to detail some of the contents of the immordata, but not exclusively, so the patent's utility and its ease of application can be easily understood.

[0000] 1. Life Story: Below is a potential but not exhaustive list of events, things, people, places, and any other information, that: the Immordata customers may record, update, and store:
1.1. Memories of events at different stages in life:
1.1.1. early childhood
1.1.2. mid childhood
1.1.3. late childhood
1.1.4. teenage years
1.1.5. beginning of adulthood
1.1.6, college years
1.1.7. early professional years
1.1.8. middle professional years
1.1.9. middle age
1.1.10. conclusion of work life
1.1.11. retirement
1.1.12. any other events, real or imaginary or in dreams
1.2. People you know:
1.2.1. family members
1.2.2. friends
1.2.3. colleagues

1.2.4. neighbors
1.2.5. acquaintances
1.2.5. others you have met and known
1.2.7. any other people, real or imaginary or in dreams
1.3. Accomplishments of your life:
1.3.1. tasks you have accomplished with pride
1.3.2, plans for accomplishing future tasks
1.3.3. awards, prizes, certificates, diplomas, etc.
1.3.4. any and all accomplishments
1.3.5. things you wish you could accomplish
1.4. Knowledge you have:
1.4.1. things you have learned in school
1.4.2. things you have learned for your jobs
1.4.3. things you have learned about life's daily activities
1.4.4. things you have learned about the world
1.4.5, trivia
1.4.6. any other knowledge
1.5. Places where you have lived:
1.5.1. birth place
1.5.2. cities or villages or whatever place you have lived in
1.5.3, places where you have worked
1.6. Places you have visited:
1.6.1. places you have visited
1.6.2. recollections of those places that you have visited
1.7. Places you have dreamt about:
1.7.1. dreamscape: places you have dreamt up in sleep
1.7.2, places you have conjured up in daydreaming
1.7.3. imaginary communities or places you have conjured up creatively
1.7.4. any other place that has come to your mind or conscience
1.8. Places you wish to visit:
1.8.1. places you have read about and wish to visit
1.8.2. places you have seen visually but wish to visit in person
1.8.3. any other place you wish to visit

2. Feelings, Emotions, Thoughts, or any other mental activity or state: Immordata customers can use the Immordata website and databases to build their consciousness; that is, everyone can continuously make entries under each feeling, emotion, and thought in the categories (though not exclusive) listed below. This will enrich the volume of the consciousness, making sure that your consciousness is entirely you, beautifully you, and glorious you, rich, plentiful, detailed, and fully dimensioned. Furthermore, all the rich experiences, emotions, wisdom, and knowledge, etc., will build a collective human consciousness that is more interactive than otherwise possible. The vision is that, as an aggregate whole, this collective human consciousness, as supported by the individual consciousness, will evolve into something far greater than anything that we have thus far witnessed in human civilization. And, if we have the far-reaching vision of billions of years from now, we will need this collective human consciousness and knowledge to jump out of this universe, as this universe is, based on scientific consensus, to collapse in one way or another. If that is to happen, the most probable way to escape this universe is the electronic collective consciousness that Immordata and its customers and users will be able to start building.

2.1. Below is a list of possible but not exhaustive feelings, memories, emotions, thoughts, etc., which an Immordata customer can record and store:
2.1.1. happiness
2.1.2. joy
2.1.3. sympathy
2.1.4. understanding of humor
2.1.5. feeling of humor

2.1.6. empathy
2.1.7. ponderings
2.1.8. outrage
2.1.9. emotional discoveries
2.1.10. hatred
2.1.11. arrogance
2.1.12. regrets
2.1.13. pride
2.1.14, shame
2.1.15. disappointment
2.1.16. bias
2.1.17. resentment
2.1.18. envy
2.1.19. jealousy
2.1.20. yearning
2.1.21. dreams
2.1.22. wishes
2.1.23. desires
2.1.24. urges
2.1.25, anxiety
2.1.26, concerns
2.1.27. worries
2.1.28. peeves
2.1.29, love
2.1.30. bitterness
2.1.31, pain
2.1.32. suffering
2.1.33. frustration
2.1.34. depression
2.1.35. trivial and useless thoughts
2.1.36. musings
2.1.37. arrogance
2.1.38, snobbery
2.1.39. boredom
2.1.40. sweetness
2.1.41. bliss
2.1.42. exaltation
2.1.43. tedium
2.1.44. orgasm
2.1.45. sexual sensations
2.1.46, comfort
2.1.47. tactile touch and sensations
2.1.48. feel of different materials
2.1.49. athletic movements and reflexes
2.1.50, moral judgment
2.1.51, social responsibilities
2.1.52. communal duties and obligations
2.1.53. what you deem as the norm of society
2.1.54. what: to you is not acceptable social behavior
2.1.55. what you wish to know
2.1.56, what you wish to improve
2.1.57. what you would change about the world
2.1.58. how to build consensus and the importance of it
2.1.59. political opinions

2.1.60. uneasiness
2.1.61. fainting feeling
2.1.62, queasiness
2.1.63. madness
2.1.64. drunken feeling
2.1.65, elation
2.1.66, sadness
2.1.67. bewilderment
2.1.68, panic
2.1.69. sensations of heat, cold, chill, dryness, humidness, etc.
2.1.70. curiosity
2.1.71. embarrassment
2.1.72. hole-in-the-ground moments
2.1.73. earliest memories
2.1.74. betrayal
2.1.75. stubbornness
2.1.76. vengeance
2.1.77. all tastes
2.1.78. cowardice
2.1.79. hole-in-one feeling
2.1.80. being discriminated
2.1.81, being celebrated
2.1.82. any other feeling, emotion, knowledge, wisdom, thought, or mental activity that is “You”—specific to each individual Immordata customer
2.2. At the same time, Immordata customers can keep an Immordata diary, making daily entries and going back to the entries to make updates. This will be convenient and will help Immordata customers to build a habit of daily entries, which will enable the Immordata customers to detail their daily activities, feeling, emotions, reflections, etc., on a daily basis, thus making the collection of Immordata detailed, complete, rich, and comprehensive.

3. Building a Future:

[0044] 3.1. Community Forming: Using existing data, Immordata customers can build their communities of various kinds and levels, such as families, neighborhoods, cities, counties, states, provinces, prefectures, cantons, countries, or even the virtual United Nations.
3.2. Explorations of New Frontiers: Using existing data, Immordata customers can explore new potentials that are not yet thought of by the greatest theorists. But the data stored and operated by Immordata will build a base for them to spring from and provide a venue for such explorations.

The Different Levels of Data Storage:

[0045] Immordata data will be categorized into three different databases, as shown in FIG. 12: public data, limited-access data, and private data (1204), to be detailed below. The purpose of three different levels of data is that some Immordata data can be shared with the public; some may be shared only with families, friends, or authorized legal entities; but some will be private to the individuals only, and, as such, should be use only to be evoked at such a time as deemed appropriate by Immordata and the Immordata Congress to build individual immortality when such technology avails itself.

[0000] 1. The Public Immordata Data: Immordata customers will store data here that, is made public to varying groups:
1.1. Data public to everyone: This is the kind of data the Immordata customers will choose to be made public to anyone browsing the Immordata public information areas.
1.2. Data to be shared only with varying circles: This is the kind of data the Immordata customers

will choose to be made available only to people he/she will give consent to, such as friends, families, special groups, clubs, organizations, or whomever.
2. The Confidential Immordata Data for Families, Friends, Groups, and Communities of Their Choice: Immordata customers will store data here that will be shared with families and friends or authorized legal entities, such as doctors and lawyers, who have access to it. The data will not be available to the general public, but will be shared among Immordata customers who are granted the access by the Immordata customer, so that data can be shared within limited circles. The access can be granted or revoked by the Immordata customer who has created the data, but Immordata and the Immordata Congress will have the ultimate decision on whether the access should be granted, based on the power it is granted by the concurrent law.
3. The Private Immordata Data: Immordata customers will store data here that is guarded and will not be shared with anyone, as in the case of lawyer-client or physician-patient privileges, to be only used to build immortality for Immordata customers. Even when such information is used to build immortality for Immordata customers, this information will be used only for the immortal individual, not shared with anyone else in any situation for any purpose. It shall never be made public forever. This database is the crux of the trust between Immordata and its customers, without which Immordata cannot function or survive as a viable venue, method, or process to store our customers' data for purposes of immortality.

How It Benefits Mankind

[0046] The benefits of the present invention are many:

1. Learning Forever

[0047] 1.1. In the electronic form this Immordata existence will enable every consciousness or soul to learn forever, when the technology makes it possible for the immortal consciousness or soul to be live again.
1.2. The assumption is that all human short comings of the mind are rooted in the fact that the biological and neurological mind was constrained and confined within the biological boundaries of the human brain. A fully developed mind, in a form other than biological, will transcend its biological roots and rise above them to great heights of enlightenment, understanding, capabilities, and unlimited achievements.

2. Full Utilization of Latent Human Intelligence:

[0048] 2.1. We know much knowledge, wisdom, and creativity pass with the passing of a great mind or any mind. The recording or storage of these great minds could prevent such passing or loss. Just imagine, how many more ideas were still floating in Einstein's mind when he passed? And how many patent examiners like Einstein have passed without: writing down his or her great ideas? What would have happened if Leonardo de Vinci had written down his helicopter idea in more detail before his passing? Could the modern helicopter have been invented centuries earlier? Just imagine all the great ideas and minds of the billions of humans that have not been made known. How do we know those hidden or private ideas will not transform humanity entirely? What if suddenly we can harvest all those ideas and communicate all those ideas to each other for further advances? Could the human civilization see quantum leaps forward as a result?
2.2. I think it is a sure bet that, with the establishment of Immordata and later the evocation of Immordata to form immortal individuals and communal human immortality, this sea of knowledge, the vast collection of ideas, the immense wisdom, the endless scientific and technological know-how, and all that the human minds possess will make the human race not only a better race, but also a far more powerful race, enabling it: to achieve things one cannot even possibly imagine at this point in time.

3. Next Stage of Human Development

[0049] 3.1. Humans will probably no longer form prejudices related to physical attributes because the ultimate existence is in the realm of bio-electronics or some other advanced form.
3.2. It is possible that all souls will be connected in some form, thus forming a collective consciousness, which, as a part of the collective whole, is far more powerful, potent, productive, and beneficial than otherwise.
3.3. The probable immortal world in a form other than biological, such as the Singularity, also has a higher level or survivability, if we are to survive the disintegration of the universe in, say, 30 billion years, or in any length of time in which our universe will last.
3.4. This immortal world will enable souls or consciousness to continue to learn.
3.5. Because of this continuous learning, the souls and consciousnesses will evolve into higher beings with higher moral standards and better understanding, thus improving the souls and consciousnesses. Because of all that development, evil spirits and debased souls will transform to become better souls and consciousnesses.
3.6. Souls and consciousness will then command vast reservoirs of knowledge, thus being able to make better decisions, make new discoveries, improve the world tremendously, and eventually even find ways to survive the collapse of this universe and transport into another universe, were that to happen, to continue the human race and the human consciousness.

US2009062677

Method of Recording and Saving of Human Soul for Human Immortality and Installation for it

Inventor:BOLONKIN ALEXANDER ALEXANDROVICH

The inventor shows that a human soul is only the information in human brain. He offers a method for re-writing the human brain on electronic chips. This method allows for the modeling of a human soul in order to achieve man immortality. This method does not damage the brain but works to extend and enhance it. This method of writing and saving of Human Soul includes: writing of views which person see in during his life by a micro video recorder to a portable memory; writing of sounds which person speaks and hears by microphone; writing of person physical conditions by micro sensors; writing of time and dates; writing of person body position and so on. An Installation utilized the Method is comprising devices: micro video camera located at man head, microphone, computer (chip) for data processing, portable memory located at an person body and connected to the portable memory; portable recorder for writing of life film located at the man body and connected to the computer and to the portable memory; micro sensors for writing of person physical and environment conditions; a clock for writing of time, the micro sensors for writing of men body positions; navigation system, long time memory high capability storage and so on.

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0005] The inventor shows that human soul is only the information in human brain. He offers a new method for re-writing the human brain on electronic chips. This method allows for the modeling of a human soul in order to achieve immortality. This method does not damage the brain but works to extend and enhance it.

[0006] 2. Description of the Related Art

[0007] Immortality is the most cherished dream and the biggest wish of any person. People seldom think about it while they are still young, healthy, and full of energy. But when they get some incurable disease or become old, then there is no bigger wish for them than to live longer, put off the inevitable end. And no matter what heavenly existence in the after-life is promised to them by religion, the vast majority of people want to stay and enjoy life here, on Earth, as long as possible. The inventor shows a real immortality can be only electronically.

Medical Science and the Issue of Immortality

[0008] A great many of doctors and scientists are currently working on the problems of health and longevity. Substantial means are spent on it, about 15-25% of all human labor and resources. There are certain achievements in this direction: we have created wonderful medications (e.g. antibiotics); conquered many diseases; learnt to transplant human organs; created an artificial heart, kidneys, lungs, limbs; learnt to apply physiological solutions directly into the blood stream, and to saturate blood with oxygen. We have gotten inside the most sacred organ—the human brain, even inside its cells. We can record their signals, we can agitate some parts of the brain by electric stimuli inducing a patient to experience certain sensations, images, and hallucinations.

[0009] We can attribute the fact that the average life span has increased two times in the last two hundred years to the achievements of modern medicine.

[0010] But can medical science solve the problem of immortality? Evidently, it cannot. It cannot do that in principle. This is a dead-end direction in science. Maximum it can achieve is increasing the average life expectancy another 5-10 years. An average person will be expected to live 80 years instead of 70. But what kind of person will it be? A very old one, capable of only existing and consuming, whose medical and personal care will demand huge funds.

[0011] The proportion of the elderly and retirees has increased steeply in the last 20-30 years and continues to grow depleting the pension funds and pressuring the younger generation to support them. So it is hard to say whether the modern success of medicine is a blessing or a curse from the point of view of the entire humankind, even though it is definitely a blessing from the point of view of a separate individual.

[0012] Humanity as a whole, as a civilization, needs active, able to work and creative members, generating material wealth and moving forward technology and science, not the elderly retirees with their numerous ailments and a huge army of those tending to them. It dreams not of the immortality of an old person, but of the immortality of youthfulness, activity, creativity, enjoying life.

[0013] Now there are signs of a breakthrough, but not in the direction the humankind has been working on all along, since the times of the first sorcerers to modern-day highly-educated doctors. Striving to prolong his biological existence, man has been chiseling, so to speak, at the endless stone wall. All he has been able to accomplish is only a dent in that wall—increased life expectancy, conquering some diseases, relieving suffering. As a payoff, the humanity has received a huge army of pensioners and retirees and gigantic expenditure on their upkeep.

[0014] Of course, one can continue chiseling at the dent in the wall further on, make it somewhat bigger, aggravating side effects. But we are already approaching the biological limit, when the cause of death and feeblemindedness is not a certain disease which can be conquered, but general deterioration of the entire organism, its decay on the cellular level, when the cells stop to divide. A live cell is a very complex biological formation. In its nucleus it has DNA—biological molecules consisting of tens of thousands of atoms connected between themselves with very fragile molecular links. Suffice it to say, that temperature fluctuation of only a few degrees can ruin these links. That is why a human organism maintains a certain temperature—36.7 C. Raising this temperature only 2-3 degrees causes pain, and 5-7 degrees leads to death. Maintaining the existence of human cells also presents a big problem for humanity involving food, shelter, clothes and ecologically clean environment.

[0015] Nevertheless, human cells cannot exist eternally even under ideal conditions. This follows from the atomic-molecular theory. Atoms of biological molecules permanently oscillate and interact with each other. According to the theory of probability, sooner or later the impulses of adjacent atoms influencing the given atom, add up, and the atom acquires enough speed to break loose from its atomic chain, or at least to transfer into the adjacent position (physicists say that the impulse received by the atom has surpassed the energy threshold which retains the atom in its particular place in the molecular chain). It also means that the cell containing this atom has been damaged and cannot any longer function normally. Thus, for example, we get cancer cells which cannot fulfill their designated functions any more and begin to proliferate abnormally fast and ruin human organs.

[0016] This process accelerates manifold when a person has been exposed to a strong electromagnetic radiation, for instance, Roentgen or X-rays, a high-frequency electric current or radioactive materials.

[0017] Actually, the process of deforming of the hereditary DNA molecule under the influence of weak cosmic rays can take place from time to time, leading sometimes to birth defects, or it may turn out to be useful for the survival properties. And this plays a positive role for a particular species of plants or animals contributing to their adaptability to the changed environment and their survival as a species. But for a particular individual such aberration is a tragedy as a rule, since the overwhelming majority of such cases are birth defects, with only few cases of useful mutations. And human society in general is suspicious of people who are radically different in their looks or abilities.

BRIEF SUMMARY OF THE INVENTION

[0018] The inventor shows that a human soul is only the information in human brain. He offers a method for re-writing the human brain on electronic chips. This method allows for the modeling of a human soul in order to achieve immortality. This method does not damage the brain but works to extend and enhance it.

[0019] This method of writing and saving of Human Soul includes: writing of views which person see in during his life by micro video recorder to portable memory; writing of sounds which person speaks and hears by microphone; writing of person physical conditions by micro sensors; writing of time and dates; writing of person body position, etc.

[0020] An Installation used the Method comprising devices: portable memory located at an person body; computer (chip) for data processing located at the person body and connected to the portable memory; micro camcorder (micro video camera) for writing of life film located at the person body and connected to the computer and to the portable memory; microphone for writing all sounds, which person speaks and hears, located at the person body and connected to said computer and to said portable memory; micro sensors for writing of person physical conditions located at the person body and connected to the computer and to the portable memory; micro clock for writing of time located at the person body and connected to the computer and to the portable memory; micro sensors for writing of person body positions located at the person body and connected to the computer and to the portable memory; long time memory high capability storage.

[0021] Writing of human soul is absolutely new idea of modeling, re-writing, and saving of human soul and using it for person immortality. No any materials about this method and installation.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[0022] FIG. 1. Principal scheme of the installation (equipment) of writing human soul (full information about human life) for man immortality. The cells contain:

1) micro devices (micro video camera, microphone, computer (chip), portable recorder, portable power source, portable memory, etc.);
2) micro sensors and devices for writing of brain impulses, oscillation, fluctuation and a state of man health and person physical conditions (for example: sensors of blood press, heart palpitation, skin resistance, perspiration, skin temperature, respiration activity, mouth water, chill, rigor-like tremor, catarrh, muscle stress, blood mixture, pain, brain oscillation and so on);
3) micro sensors for writing of body positions (for example: sensors of motion, acceleration, distance, angle, mutual position, and so on);
4) devices for measure of outer environment conditions about man such as air temperature, pressure, humidity, gas composition, smells, wind, illumination, radiation, and so on;
5) devices of a portable navigation system for measure of a person position at Earth surface (for example, GPS navigation system);
6) emergency 911 communication;
7) portable power source (for example: chemical, solar, nuclear batteries, and so on);

[0030] FIG. 2. Re-writing the information from the portable memory to the high capability long-term storage,

[0031] FIG. 3. Location of video camera at human head (forehead).

DETAILED DESCRIPTION OF THE INNOVATION

1. Brief Description of Previous Works by the Inventor

[0032] In a series of articles (see referenced list at the end) the inventor shows that the purpose of Nature is to create Super Intelligence (SI). With its ability to understand the Universe, advanced entities with SI Power will be able to survive major cataclysms. There is the Law of Increasing Complexity (in opposition to the Entropy Law—increasing chaos). This Law created biological intelligence (people). Human have since became a sovereign entity on the Earth and in Nature above all other creatures.

2. Electronic Immortality. Advantages of Electronic Existence

[0033] In the noted works the author has shown that the problem of immortality can be solved only by changing the biological human into an artificial form. Such an immortal person made of chips and super-solid material (the E-man, as was called in his articles) will have incredible advantages in comparison to conventional people. An E-man will need no food, no dwelling, no air, no sleep, no rest, and no ecologically pure environment. His brain will work from radio-isotopic batteries (which will work for decades) and muscles that will work on small nuclear engines. Such a being will be able to travel into space and walk on the sea floor with no aqualungs. He will change his face and figure. He will have super-human strength and communicate easily over long distances to gain vast amounts of knowledge in seconds (by re-writing his brain). His mental abilities and capacities will increase millions of times. It will be possible for such a person to travel huge distances at the speed of light. The information of one person like this could be transported to other planets with a laser beam and then placed in a new body.

[0034] Such people will not be awkward robots as in the movies. An artificial person will have the opportunity to choose his or her face, body and skin. It will also be possible for them to reproduce and then avoid any period of adolescence including the need for education. It will be impossible to destroy this entity with any kind of weapons, since it will be possible to copy the information of their minds and than keep such information backed up in separate distant locations. As was written in the science fiction book, “The Price of Immortality”, by Igor Getmansky (Moscow, Publish House ECSMO, 2003, Russian) an artificial person will have all of these super-human abilities.

3. What are Men and Intelligent Beings

[0035] All intelligent creatures have two main components: 1. Information about their environment, about their experience of interacting with nature, people, society (soul) and 2. Capsule (shell), where this information is located (biological brain, body). The capsule supports existence and stores information and programs for all of its operations. The capsule also allows the creature to acquire different sensory information (eyes, ear, nose, tongue and touch) and it moves to different locations in order to interact with the environment.

[0036] The main component of an intelligent being is information (soul). The experiences and knowledge accumulated in the soul allows the entity to interact more efficiently in nature in order to survive. If the being has more information and better operational programs (ability to find good solutions), then it is more likely thrive.

[0037] For an intelligent being to save its soul it must solve the problem of individual immortality. Currently man creates a soul for himself by acquiring knowledge from parents, educational systems, employment and life experiences. When he dies, most knowledge is lost except for a very small part which is left through works, children and apprentices. Billions of people have lived on Earth, however, we know comparatively little about ancient history. Only after the invention of written language did people have the capacity to easily save knowledge and pass it on to the next generation.

[0038] As discussed earlier, the biological storage (human brain) of our soul (information) is unreliable. The brain is difficult to maintain and requires food, lodging, clothes, a good environment and education, etc. To support the brain and body, humans spend about 99% of their time and energy, and eventually what knowledge is gained is taken to the grave in death.

[0039] There is only one solution to this problem—re-write all of the brain information (our soul) in more strongly based storage. We must also give the soul the possibility to acquire and manipulate information from the world. This means we must give sensors to the soul so it may have communication and contact with people and other intelligent beings. We must give the soul a mobile system (for example, legs), systems for working (hands), etc. thus giving the soul a new body in which to LIVE.

[0040] The reader may ask—these ideas seem interesting, but how does one we re-write a human soul to live within a new carrier, for example, in electronic chips?

4. The Main Problem With Electronic Immortality—Re-Writing Brain Information (Soul) to Electronic Chips is That it's Impossible to do This With Current Technology

[0041] At present scientists are working to solve this problem. They know that the brain has about 15 billion neurons, and every neuron has about ten connections to neighboring neurons. Neurons gain signals from neighboring neurons, produce signals and then send these signals to others neurons. As a result, humans are able to think and find solutions. On the bases of this way of thinking, humans can come to solutions without exact data. (Concepts of brain were described in my previous articles. For example, see “Locate God in Computer-Internet Networks” or “Science, Soul, Heaven and Supreme Mind”. See also my articles on the Internet and references at end of this article.).

[0042] Scientists are learning how to take individual neurons on micro-electrodes and record their impulses. The ideas of scientists are very simple—study how single neurons and small neuronal network work and then model them by computer. They hypothesize that if we can model 15 billion neurons in a computer they will learn how the brain works, and then they will have Artificial

Intelligence equaling the human brain.

[0043] In my previous work I show this as a dead-end direction for Human Immortality. It's true that we'll create an Artificial Intelligence (AI) that will be more powerful than the human mind. However, it will be HIS AI, and a NEW entity altogether. Our purpose is focused on preserving the CONCRETE PERSON now (more exactly—his SOUL) in a new body in order to achieve immortality.

[0044] Why is it impossible to directly write the information of the human brain onto a chip? Because the human brain is constantly changing and neurons permanently change their states. Imagine you want to record the state of a working computer chip. The chip has millions of logical elements which change their state millions of times per second. It is obvious that if you write in series (one after other) the current state of the chip (it is impossible to instantly write ALL states of the chip's elements). To instantly write all neurons one would need to insert a microelectrode into EVERY neuron, this would destroy the human brain before the writing was complete.

[0045] I offered another method for the solution of the Main Problem of Immortality.

5. Modeling of Soul for a Concrete Person

[0046] As said, straight re-writing of a human mind (human soul) to chips is very complex. Straight re-writing is not possible in the near future. All scientific works studying the work of human brains at the present time are useless for the main problem of immortality. They are also unworkable for the problem of artificial intelligence (AI) in the near term, because the brain solves problems by way of general estimations. AI solves problems based on more exact computation and logical data.

[0047] To solve the Main Problem of Immortality (MPI) the author offers a method of "MODELLING SOUL" of a concrete person. This method does not require interventions into the brain of a given person. This method may be applied IMMEDIATELY at the present time. But an accurate modeling is needed depending on the modeling period.

[0048] Before describing this method, let us analyze the human soul and what components are important for each person and his environment. All information in the human brain (soul) may be separated in two unequal groups: 1. the Memory (permanent knowledge) about the person's life (all that has been seen, heard, made, felt, people which he has met, his (her) behaviors, opinions, wishes, dreams, programs of activity, etc.), environment, and 2. Methods of processing this information, i.e. producing new solutions and new behaviors based on this knowledge.

[0049] The first part (knowledge) is very large. It fills most of the memory and remains relatively constant (you remember your life, history and you can only fill it by what was in the past). The second part (methods for deciding, producing solutions based in your knowledge) is relatively small and constantly changing because of new information, facts and life experiences.

[0050] However, the most important part of a human soul can be written without any problem now. Industry is producing cheap micro-video recorders as small as a penny, microphones at grain size, and micro-sensors for vital signs (breathing, palpitation, blood pressure, skin resistance, perspiration, movement of body parts, etc.). These measurements allow for easy recording of not only the physical state, but of his moral state (joy, pleasure, grief, trouble, anxiety, nervousness, etc). For example, lie detectors are able to define not only the state of a man, but also the truth of his words. Now we can measure and record brain commands and we can produce small cards with four gigabytes of memory.

[0051] It would be easy to attach a video recorder and microphone to a man's forehead and then attach sensors to the body and record all that he sees, hears, speaks, his feelings, reactions, and activity. And then re-write this information into a personal hard drive (long-term memory of high capacity storage) at the end of each day. As a result, there is a record of the most important part our soul—history of life, feelings, environment, behaviors and actions. This would be more detailed than what is captured by the real man, because the humans forget many facts, feelings, emotions, and personal interactions. The electronic memory would not forget anything in the past. It would not forget any person or what they were doing.

[0052] But what about the second smaller part of the human soul—producing solutions based on personal knowledge—perhaps asks the meticulous reader.

[0053] This could be restored by using past information from the real man in similar situations. Moreover, an electronic man could analyze more factors and data in order to throw-out and exclude actions and emotions that happened under bad conditions. The electronic man (named E-being in my previous works) would have a gigantic knowledge base and could in a matter of second (write to his brain) produce the right answer, much faster than his biological prototype. That means he would not have the need for the second smaller part of memory.

[0054] Considering the environment and friends, the following is an important part of a man's soul: his relationship with parents, children, family, kin, friends, known people, partners and enemies. This part of his soul will be preserved more completely than even his prototype. Temporary factors will not influence his relationship with his enemy and friends as would happen with his former prototype.

[0055] There is one problem which may be troubling for some: if we were to record every part of a person's life, how do we keep intimate moments a secret? There are (will be) ways to protect private information which could be adapted from current usage, for example, the use of a password (known only by you). Also there may be some moments you choose not to record information or decide to delete the information from memory.

[0056] The offered system may become an excellent tool for defense again lies and false accusations. You may give the password in one given moment of your life, which proves your alibi or absence from the accusations.

[0057] Some people want to have better memory. Video takes 95% of storage capacity, sound takes 4% and the rest takes 1%. In usual situations, video can record only separate pictures, sound only when it appears. This type of recording practice decreases the necessary memory by tens of times. But every 1.5-2 years chip storage capacity doubles. There are systems which will compress the information and then may select to record the most important information (as is done in the human brain). During your life, the possibility to record all information will be available for all people. This type of recording apparatus will be widely available and inexpensive. It's possible now. The most advanced video recorder or DVD writes more information than a CD.

[0058] This solution (recording of human souls) is possible and must be solved quickly. By mass production (large productions) the apparatus will become inexpensive. The price will drop to about $300-1,000. If we work quickly we can begin recording and then more fully save our souls. The best solution is to begin recording in children when they become aware of “I”. But middle and older people should not delay. Unrecorded life periods may be restored by pictures, memories, notes, diaries and documents. Soul recovery will only be partial but it's better than nothing.

[0059] These records will also be useful in your daily life. You can restore recorded parts of your life, images of people, relatives, and then analyze and examine your actions for improvement.

6. Disadvantages of Biological Men and Biological Society

[0060] People understand Darwin's law, “survival of the fittest”. For a single person, this law is the struggle for his/her personal existence (life, well-being, satisfaction of requirements, pride, etc.). In a completely biological world built on Darwinian law the strongest survives and reaches his goal. Though they may be intelligent, humans are members of the animal world. They operate as any other animal in accordance with animal instincts of self-preservation. If one is poor, at first he struggles for food (currently half of world's population is starving), dwelling, and better living conditions. When one reaches material well-being, he may struggle for money, job promotion, reputation, renown, power, attractive women (men), and so on. Most people consider their activities (include official work) in only one way—what will I receive from it? Only a small number of people are concerned with the idea of sacrificing themselves to the well-being (seldom giving up their life) of society at large.

[0061] As a result, we see human history as a continuation of wars, dictatorships, and repression of people by power. Dictators kill all dissidents and opponents. Most people try to discriminate against opponents and play dirty against their enemy. There are murders, rapes, violence, robbery, underhand actions, fraud, and lying at all levels of society especially in lesser developed countries. Each person only cares for himself and his family and does not care how his actions effect other people or society.

[0062] Democratic countries try to cultivate a more civilized society. They create laws, courts, and have police. Dictator regimes, on the other hand, make only the law they want. I could give thousands of examples to verify this concept. But hundreds of millions of people are killed by war, aggressive campaigns, repressions, genocides, and thousands of criminals in the everyday world are a good illustration of this.

[0063] The human brain allows us to reach great success in science and technology. However, as a biological heritage, struggling for his INDIVIDUAL existence in a bloody, dangerous world, humans spend much of their resources on mutual extermination of intelligent beings. Moreover, humans have created ever powerful weapons (for example, nuclear and hydrogen bombs), which could wipe out humanity. In time, existence may depend on the volition of one man—perhaps the dictator of a nuclear state.

[0064] The second significant drawback to the biological body—is that it spends 99.99% of its effort and resources simply to support existence. Such as food, lodgings, clothing, sex, entertainment, relaxation, environment, ecological compatibility. Only a very small part is uses for scientific development and new ideas and technology. The reader may see something wrong here.

[0065] States use a parentage of their revenue for research into science and technology. This percent is used NOT for NEW ideas, but is used to commercialize modern processes All research is included in the state budget under the name, “Science and New Technology”. But much of this research has little relation to real new scientific progress. Even it the US, states spend only a small part of the assigned money on new science because state officers do not understand the research. People, organizations, and companies fight for a piece of the pie. Geniuses are rare and usually don't have the capacity to move forward because they must promote and pay for new ideas from their own empty packets.

[0066] Yet, science and technology has seen success. Most advancement (90%) was made recently in the 20th century, when governments started to finance a few scientific projects (compared with the millions of years of human existence). However, our current knowledge and new technologies are far from what we will eventually have. The first government of an industrialized country to understand and realize the leading role of new science and innovation will become powerful.

7. Electronic Society

[0067] The electronic society will be a society of clever electronic beings (or E-being, as they named in my articles). Most of the reasons and stimulus which incite men to crime, will be absent in E-beings. E-beings will not need food, shelter, sex, money, or ecology, which are the main factors in crime. E-people will not have intense infatuations or be distracted by behaviors, because they will have vast knowledge about the open electronic society. Their main work will be in science, innovations, and technologies. They will save their mental capacity for the production of chips and bodies, scientific devices, experimental equipments, space ships and space station, etc. They will need a number of robots, which do not need a big brain. It is likely they will award these robots better minds and memory. It is also likely that E-man will unite in a common distributed hyper-brain, which will become a sovereign of the Universe (God).

[0068] Nature is infinite and the development of a Super Brain (God) will not be limited. On the other hand, biological people will have limited mental capabilities. It will be difficult for them to image and predict the development and activity of Super beings, which we will generate.

[0069] Many, especially religious people, object because they say electronic beings will not have human senses such as love, sympathy, kindness, humanism, altruism, and the capacity to make mistakes, etc. E-beings are not people. Look back at human history. Human history shows that kindness played a very small role in human life. All human history is the history of human vices and human blood: struggle for power, authority, impact, money, riches, territory, and states. All human history is filled with fraud, underhanded actions, and trickery. Ordinary people were only playthings, flock of sheep for the tyrants and dictators.

[0070] Some people object that with an electronic face humans will loss the joy of sex, alcohol, narcotics, appreciation of art, beauty, nature, etc. My answer to this question is in my article “Science, Soul, Heaven, and Supreme Mind” (http://Bolonkin.narod.ru). The brief answer is that electronic humans will enjoy all this in a virtual world or virtual paradise. Time will run millions of times faster in the virtual World. E-man will spend a few seconds of real time and live millions of years in the paradise. He will enjoy any delight imaginable, include sex with any beautiful women (or handsome men), feel the emotions of any commander, leader, criminal, or even a dog.

8. Lot (Fortune) of Humanity

[0071] Biological humanity will be gradually transformed to electronic beings. Old people, when their biological bodies can not support their brains, will continue their existing in electronic bodies after death. They will become young, handsome, robust, and. Fertility in biological men will decrease. Birth-rates are less than death-rates in many civilized countries now (for example in France). Population growth is mainly supported by emigration from lesser developed countries. When education levels increase, birth-rates will fall.

[0072] For a time, biological and electronic people will exist together. However the distance between their capabilities will increase very quickly. Electronic people will reproduce (multiple) by coping, learn instantly, and will not need food or dwellings. They will work full days in any condition such as in space or on the ocean floor. They will gain new knowledge in a short time. They will pass this knowledge on to others who do not have enough time. The distance between biological and artificial intellects will reach a wide margin so that biological people will not understand anything about new science as monkeys do not understand multiplication now even after much explanation.

[0073] It is obvious, clever people will see that there will be a huge difference between the mental abilities of biological and electronic entities. They will try to transfer into electronic form and the ratio between biological and electronic entity will quickly change in electronic favor. A small number of outliers will continue to live in their biological body in special enclaves. They will not have industrial power or higher education and will begin to degrade.

[0074] Naysayers may promote laws against transferring into an electronic man (as cloning is forbidden now in some states). However, who would renounce immortality for themselves, especially while they are young and healthy? One may denounce immorality as blasphemy, but when your (parents, wife, husband, children) die, especially if you are near death yourself, one comes to understand that life is extremely important. The possibility to live forever, to gain knowledge that improves life, will also allow one to become a sovereign force in the Universe.

[0075] We summarize all said in the following steps and devices:

Method

[0000]
1. A method of writing and saving of Human Soul for man immortality comprising of steps:
(a) positioning a micro video camera at a man head;
(b) positioning a portable recorder at a man body;
(c) positioning a portable memory at said man body;
(d) positioning a microphone at said man body;
(e) positioning a portable computer at said man body;
(f) positioning a portable power source at said man body;
(g) connecting said micro video camera, said microphone to said micro computer, said portable recorder, said power source and said portable memory;
(h) programming said computer for writing needed data from said micro video camera, microphone whereby said portable recorder to said portable memory;
(i) writing of views which said man see in during his life whereby said micro video camera, said portable computer, said portable recorder to said portable memory;
(j) writing of sounds which person speaks and hears in during his life whereby said microphone, said portable computer, said portable recorder to said portable memory;
(k) writing of time and dates whereby said portable computer, said portable recorder to said portable memory;
(l) connecting said portable memory to a stationary long term high capability memory;
(m)periodical rewriting an information from said portable memory to said stationary long term high capability personal memory;
(n) read-protecting said information.
Note: The simplest method can contain the parts of these steps, for example, writing only video and sound.
2. The Method of writing of Human Soul for man immortality as recited above comprising at least one of the following additional steps:
(a) writing of brain impulses, oscillation, and a state of a man health and person physical conditions whereby micro sensors, said computer, said recorder to said portable memory;
(b) writing of a person body position whereby micro sensors, said computer, said recorder to said portable memory;
(c) writing of outer, environmental conditions around said man whereby micro sensors, said computer, said recorder to said portable memory;
(d) writing person position at Earth surface whereby portable navigation system, said computer, said recorder to said portable memory;
(e) permanently writing said conditions and information to said portable memory;
(f) permanently writing said condition and information to said portable memory except slipping time;
(g) periodically writing said condition and information to said portable memory;
(h) writing said condition and information only in selected time to said portable memory;
(i) data processing before writing to said portable memory;
(j) data compressing before writing to said portable memory;
(k) data coding of said information;

(l) transferring of said information to stationary receiver.
3. The Method of writing of Human Soul for man immortality as recited above comprising at least one of the following additional steps:
(a) writing of said man physical conditions includes at least one of the following: opening of man eyes, body temperature, blood press, heart palpitation, skin resistance, perspiration, skin temperature, respiration activity, mouth water, chill, rigor-like tremor, catarrh, muscle stress, blood mixture, pain;
(b) writing of said outer conditions includes at least one of the following: air temperature, pressure, humidity, gas composition, smells, wind, illumination, radiation;
(c) computer analyzing of said man physical and outer conditions;
(d) informing about emergency situation said man and an outer authority;
(e) long time saving whereby said high capability long-term memory.

Installation (System)

[0000]
4. An Installation utilized the Method of writing and saving of Human Soul for human immortality comprising devices:
(a) micro video camera located at man head for writing of views which said man see in during his life;
(b) microphone located at said man body for writing all sounds which person speaks and hears in during his life;
(c) portable computer for data processing located at said man body and connected to said micro video camera and said microphone;
(d) portable clock for writing of time located at man body and connected to said computer;
(e) portable recorder located at man body and connected to said computer, said micro video camera and said microphone;
(f) portable memory located at an said man body and connected to said recorder;
(g) portable source located at said man body and connected to said computer, said micro video camera, said microphone and said memory;
(h) stationary high capability long-term memory for rewriting of information from said portable memory;
(i) communications located at said man body and connected said micro video camera, said microphone, said computer, said recorder, said memory, said power source in a common net.
Note: The simplest instillations can contain the parts of these devices, for example, devices for writing video and sound.
5. The Installation recited above comprising at least one of the following devices:
(a) micro sensors for writing of brain impulses, oscillation, fluctuation and a state of a man health and man physical conditions located at said person body and connected to said computer and to said portable recorder;
(b) micro sensors for writing of man physical conditions located at said person body and connected to said computer and to said portable recorder;
(c) micro sensors for writing of a state of a man health and person body positions located at said person body and connected to said computer and to said portable recorder;
(d) devices for measure of outer environment conditions located at man body and connected to said computer;
(e) devices of a portable navigation system for measure of a person position at Earth surface located at said man body and connected to said computer and said recorder;
(f) devices for emergency communications located at man body and connected to said computer;
(g) portable power source located at said person body and connected to said micro sensors, said devices, said computer, and said portable memory.
6. The Installation recited above comprising at least one of the following devices and features:
(a) said micro video camera is located in man spectacles or on head decoration;
(b) said computer is programmed chip for writing of said data in selected time, data processing and

re-writing them to said high capability long-term data storage;
(c) memory card used as said portable memory;
(d) said physical condition sensors are at list one of the following: blood press, heart palpitation, skin resistance, perspiration, skin temperature, respiration activity, mouth water, chill, rigor-like tremor, catarrh, muscle stress, blood mixture, open eyes, pain;
(e) said body position sensors are motion, acceleration, distance, angle, mutual position;
(f) said devices for measuring of the outer environmental conditions includes at least one of the following values: air temperature, pressure, humidity, gas composition, smells, wind, illumination, radiation;
(g) said power source is at least one of the following: chemical, solar, nuclear batteries;
(h) said high capability long-term date storage is recorder used CD, DVD, type devices, memory carts.

US2003152907
Process of love

Inventor:KNAUER GEORGE

This invention combines God's gift of reincarnation with love resulting in immortality.This invention combines God's gift of reincarnation with love resulting in immortality

that they are complete, up-to-date or fit for specific purposes.

SUMMARY OF THE INVENTION

[0001] This invention resulted from my combining God's gift of reincarnation with love.

[0002] In Chapter 4 of the London translation of the Holy Bible, paragraph 8 says "He that loveth not knowth not God; for God is love." and paragraph 16 says "And we have known and believed the love that God hath to us. God is love; and he that dwelleth in love dwelleth in God, and God in him."

[0003] The scope and extent of my invention is defined in the following claim:

[0004] 1. The process of combining God's gift of reincarnation with love resulting in immortality.

US2004005535
Process of reincarnation

Inventor: KNAUER GEORGE

SUMMARY OF THE INVENTION

[0001] This invention resulted from my combining Einstein's Theory of Relativity and Newton's Second Law of Physics.

[0002] Reincarnation is defined in Webster's Third New Inernational Dictionary as "rebith". Thus my invention is a process of rebirth or in other words immortality.

JP2008178692
IMMORTALITY DEVICE 3, ADOPTING REJUVENATION METHOD WITH APPLICATION OF-OUT-OF-BODY EXPERIENCE...

Inventor: KAJISA ISAO

PROBLEM TO BE SOLVED: To provide an immortality device. ;SOLUTION: The immortality device includes processes as follows: (a) Conscious energy is placed to be out of body by a Hemi-Sync technology, right brain right angular gyrus electrode stimulation, or special video-camera; (b) Organs and nerves are generated from the pluripotent stem cells of adult human fibroblasts, and an outer body is generated from the pluripotent stem cells of the adult human fibroblasts is formed on a cyborg's bones made of a protein so that an external cyborg could be formed like oneself of its youth; (c) The conscious energy is lead to and fixed in the cyborg body generated from the pluripotent stem cells of the human adult by an infrared sensor.

[0001]

The present invention relates to a human adult skin universal cell-based protein made cyborg rejuvenation method-body application immortality device 3.
[0002]

Traditionally, there was no technology for human adult skin universal cell-based protein made cyborg rejuvenation method-body application immortality device 3.
[0003]

"Nikkei Science" January 2008 issue, 21 pages from 20 pages, 76 pages from 68 pages "Mainichi Shimbun" November 21, 2007, page 1 and page 3.
[0004]

Conventionally, to be achieved through state-of-the-art technology because there was no technology for human adult skin universal cell-based protein made cyborg rejuvenation method-body application immortality device 3.
[0005]

Body human similar behavior plus the artificial intelligence by the conscious energy that was out of body in Hemishinku technology or right brain right corner gyrus electrode stimulation and a video camera system to be fixed and put in the infrared sensor to a human adult skin universal cell-based protein made cyborg and it shall be all.
[0006]

Of immortality with human consciousness body also becomes possible to cyborg close endlessly in human beings.
[0007]

Below, it described in the series logically sentence when the embodiment of the present invention. First, two video camera back in Hemishinku technology or right brain right corner gyrus electrode stimulation or special headset, the left side of the camera image to the right eye to the left eye the right side of the camera, from the body and at the same time stroking the chest-body can immediately, another The two consciousness energy can-body to another location in the experiment ready-body.
The gist of the protein made cyborg made from human adult skin cells, such as universal organs and nerve from the human adult skin universal cells resemble me of young days is allowed to rejuvenation to make the outside of the cyborg.

This universal that human adult skin universal cell-based protein made cyborg body to consciousness energy can be the same thing all human beings are fixed by inducing in the human adult skin universal cell-based protein made cyborg body in the infrared sensor because it can sense the infrared We realize the rejuvenation immortality in their own skin cells in the cell.
The retrovirus culture and cancer gene I prevent from becoming cancer in ultraviolet B wave or vitamin D.
Human adult skin universal cell-based protein made cyborg is guided by the above as logical rejuvenation method-body application immortality device 3.
[0008]

It is carried out as described above.
[0009]

the industrial applicability of the demand of this device in human adult skin cells universal use and aging made even greater progress of body applied technology by the infrared sensor to the protein made cyborg advances is understood to be high.

JP2007135597
IMMORTALITY APPARATUS PRODUCED BY APPLYING PROTEINACEOUS CYBORG-TYPE OUT-OF-BODY EXPERIENCE 2

Inventor:KAJISA ISAO

PROBLEM TO BE SOLVED: To produce a complete proteinaceous cyborg for an immortality device produced by applying out-of-body experience. ;SOLUTION: Consciousness energy isolated from the body by a Hemi-sync technique or right brain right angular gyrus is introduced into a proteinaceous cyborg to make the cyborg to be closest possible to human being and produce an immortal cyborg.

[0001]

The present invention relates to a protein made cyborg method-body application immortality device.
[0002]

Traditionally, protein made cyborg method-body application immortality device technology was not.
[0003]

Conventionally, to be achieved through state-of-the-art technology because the protein made cyborg method-body application immortality device technology did not.
[0004]

Body Human Similar behavior plus artificial intelligence by putting the cyborg and fixed that made consciousness energy and of body at Hemishinku technical or right brain right corner gyrus electrodes stimulated protein is assumed to be all.
[0005]

Of immortality with human consciousness body also becomes possible to cyborg close endlessly in human beings.
[0006]

Below, it described in the series logically sentence when the embodiment of the present invention.

First, it is only in conscious energy that was out of body in Hemishinku technology or right brain right corner gyrus electrode stimulation.
It makes a raw material of the protein material in the nitrogen in the air.
Same cyborg body with it and protein human and make their organs and nerves, or the like.
Fixed to Put the consciousness energy that was outside the body removed from the body human in the cyborg.
Artificial intelligence plus body also becomes Cyborg close endlessly in human human similar behavior with the human consciousness of the strongest of immortality in this consciousness energy, it is assumed that it is possible to all.
Protein made cyborg method-body application immortality device 2 is guided by the above as logic.
[0007]

It is carried out as described above.
[0008]

Artificial intelligence also the industrial applicability of the demand of this device in the advance is aging achieved even greater progress cyborg been trying to exceed the human technology is understood to be high.

JP2006055651
FUTURE REJUVENATION DEVICE APPLYING OUT-OF BODY EXPERIENCE OF CONSCIOUSNESS AND BODY SUSPENDED ANIMATION STATE

Inventor:KAJISA ISAO

PROBLEM TO BE SOLVED: To realize immortality by applying out-of body experience of consciousness and a suspended animation state. ;SOLUTION: The out-of body experience of consciousness is carried out by electrode stimulation of right brain right angular gyrus or Hemi-Sync (R) technique to bring the body of a person approaching his/her life into suspended animation state by exposing the body to hydrogen sulfide and repairing the body to rejuvenation by clone technique, or the like, of future science and then returning the external consciousness to the body. Such procedures are repeated to realize the immortality.

[0001]

The present invention relates to a rejuvenation resuscitation to the future in-body awareness and physical state of suspended animation.
[0002]

Conventionally, the right corner gyrus of the electrode stimulation and Hemishinku technology capable of body has been discovered invention was not its application.
(For example, see Non-Patent Document 1.).
Further, although there is a corpse refrigeration technology it was not a state of suspended animation technology.
[0003]

"Ne~icha", September 19, 2002, 270 pages from the 269-page "Nikkei Science" September 2005, 23 pages from 16 pages
[0004]

Traditionally, there was no application of extracorporeal withdrawal and the state of suspended animation technology.

[0005]

Life consciousness that is outside the body withdrawal by returning the consciousness into the body is allowed to resuscitation after you rejuvenation repair the body in cloning technology such as the future of science is in suspended animation by exposing non-lethal dose in the body of hydrogen sulfide before life Matase in vitro me extend.
[0006]

Physically body to return many times that it was resuscitated-body consciousness from the rejuvenated state of suspended animation also made it possible to immortality by repeating.
[0007]

Below, it described in the series logically sentence when the embodiment of the present invention.
Consciousness by electrode stimulation or Hemishinku technology the right corner gyrus of the right cerebral cortex leaves the body.
Using a glass airtight container and exposing the body to hydrogen sulfide in non-lethal amount.
The extent and the rate of reduction of the human body temperature and metabolic rate shows the correlation between hydrogen sulfide concentration in the vessel.
Normally, the suspended animation-like state similar to hibernation in mammals that do not hibernate (human beings included) can be induced by hydrogen sulfide.
And at the same time resuscitation When the human body has been repaired to return the consciousness to the human body.
Life becomes the resuscitation revival can immortality also possible as the human body that have been repaired younger clone technology, etc. by future science in the state of suspended animation in order to extend the close of human life.
[0008]

It is carried out as described above.
[0009]

Western after death in each country, such as dead bodies from the frozen saved as a second method no pain without euthanasia and death with dignity of the baby-boom because it is state of suspended animation without dying not only the methods but rely on future science generation of life INDUSTRIAL APPLICABILITY it is understood that high in Japan.

JP2006021053
EXTERNAL DEPARTURE APPLIED IMMORTALITY APPARATUS

Inventor:KAJISA ISAO

PROBLEM TO BE SOLVED: To provide an apparatus to apply external departure consciousness and to realize immortality. ;SOLUTION: External departure consciousness energy is followed in an 11-dimensional space and regarded and interpreted as graviton. The frequency of a gravitational wave is guided and fixed in a humanoid cyborg with its graviton, allows the same behavior as human beings, and realizes immortality.

[0001]

The present invention relates to immortality that fixed in the cyborg to induce out of body consciousness.
[0002]

Conventionally, the right corner gyrus of the electrode stimulation and Hemishinku technology capable of body has been discovered invention was not its application.
(For example, see Non-Patent Document 1.).
[0003]

"Nature", September 19, 2002, 270 pages from the 269-page "The Lancet" December 15, 2001, 2045 pp. 2039
[0004]

Conventionally, there has been found the invention to be able of body was not immortality technique by applying it.
[0005]

It interpreted mainly to be aware that the graviton energy fraction was out of body together with the graviton the frequency of the gravitational wave is chasing in the gravity wave in 11-dimensional space in which it is going to induce the cyborg can be the same behavior as human beings in it.
[0006]

The physically-body consciousness made it possible to immortality by putting the cyborg.
[0007]

Below, it described in the series logically sentence when the embodiment of the present invention.
Consciousness by electrode stimulation or Hemishinku technology the right corner gyrus of the right cerebral cortex leaves the body.
It is the use of the 11-dimensional cosmology from going Planck scale that was rounded from one-dimensional of the six-dimensional minute it will be interpreted as a graviton that can pass through the six-dimensional from a one-dimensional (graviton).
Subject to further awareness brain waves can be interpreted as a thing that can survive in the flat is the same as the subject to be aware of the graviton from where it is known from the example of the near-death experience.
It is carried out in accordance with the frequency of the graviton and tunes gravitational wave gravity waves can pass through the 11-dimensional universe in to communicate with the graviton of the entity that this consciousness.
Further to realize the awareness of the possible human similar movement to the cyborg by inducing fixed to consciousness for the human type in the cyborg in the gravitational wave immortality.
[0008]

It is carried out as described above.
[0009]

Western postmortem in each country, such as is understood to have high industrial applicability as a second method not only the process but relies on future scientific and stored frozen bodies.

Your Support Maintains this Service --

BUY

The ***Rex Research Civilization Kit***

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

US 20120299559A1

(19) **United States**

(12) **Patent Application Publication**

McCowen

(10) **Pub. No.: US 2012/0299559 A1**

(43) **Pub. Date: Nov. 29, 2012**

(54) **ENERGY COLLECTION**

(76) Inventor: **Clint McCowen**, Navarre, FL (US)

(21) Appl. No.: **13/569,133**

(22) Filed: **Aug. 7, 2012**

Related U.S. Application Data

(63) Continuation-in-part of application No. 12/255,130, filed on Oct. 21, 2008, which is a continuation of application No. 11/358,264, filed on Feb. 21, 2006, now Pat. No. 7,439,712.

Publication Classification

(51) **Int. Cl.**
H02N 1/00 (2006.01)

(52) **U.S. Cl.** .... **322/2 A**

(57) **ABSTRACT**

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support wire system supported by poles. The support wire system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but carbon and graphite are preferred. Diodes may be used to restrict the backflow or loss of energy.

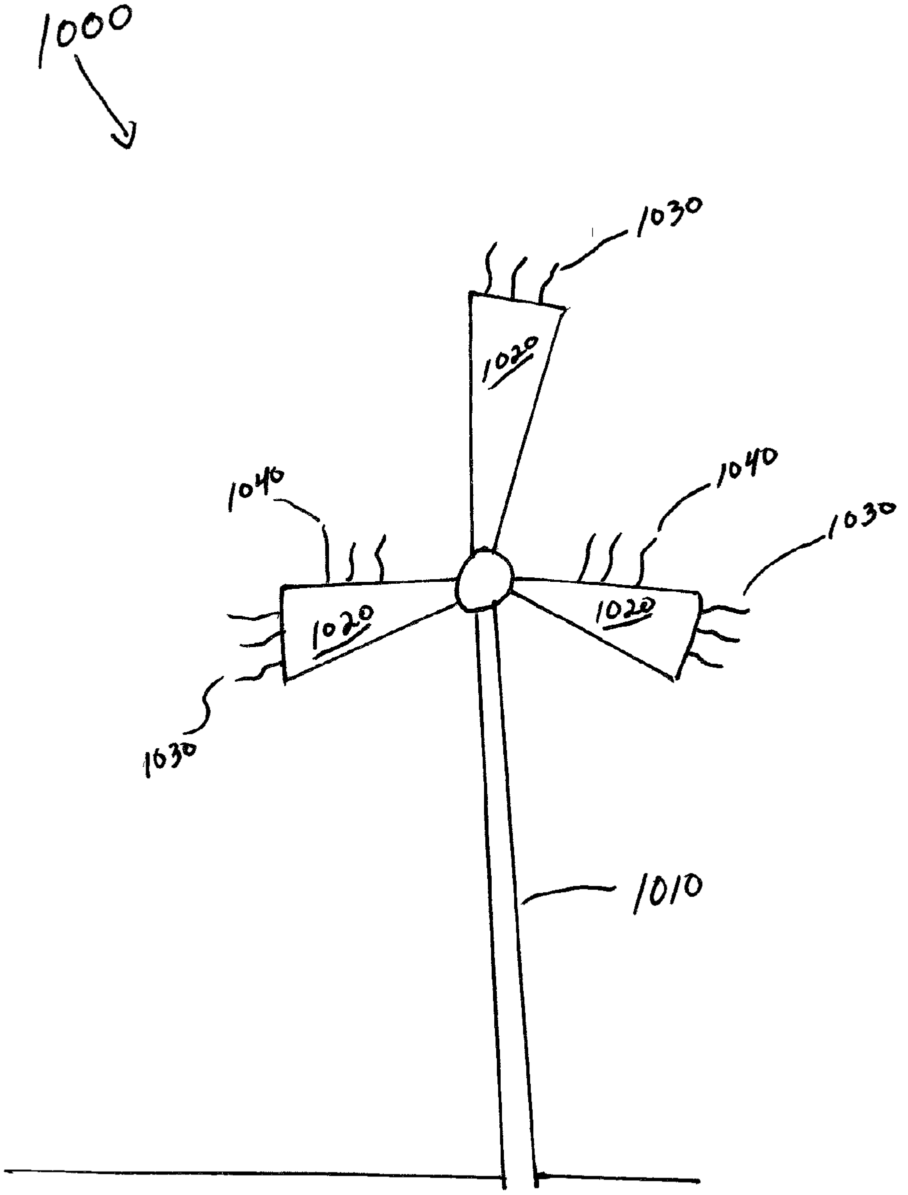

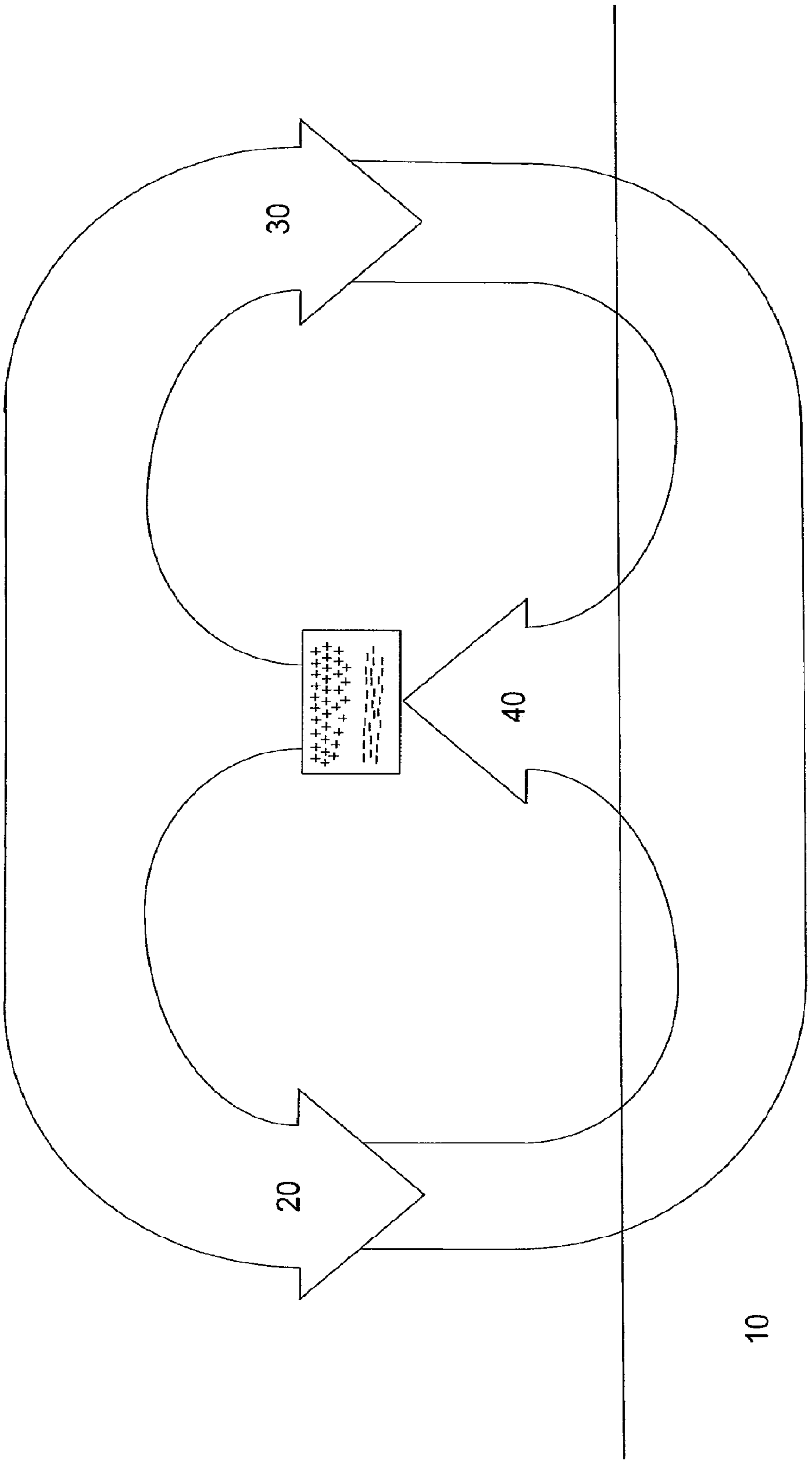

FIGURE 1

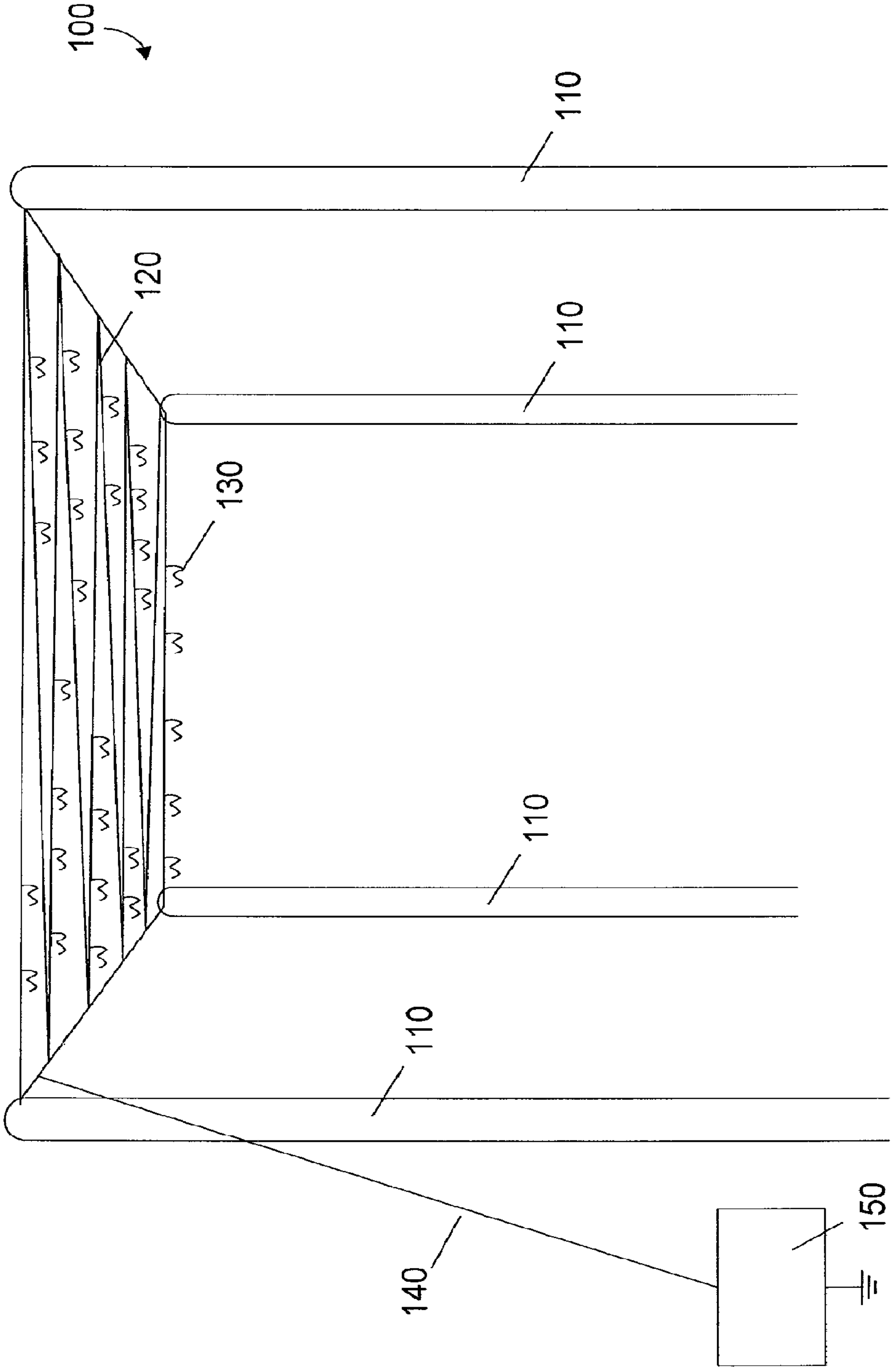
100
110
110
110
110
120
130
140
150

FIGURE 2

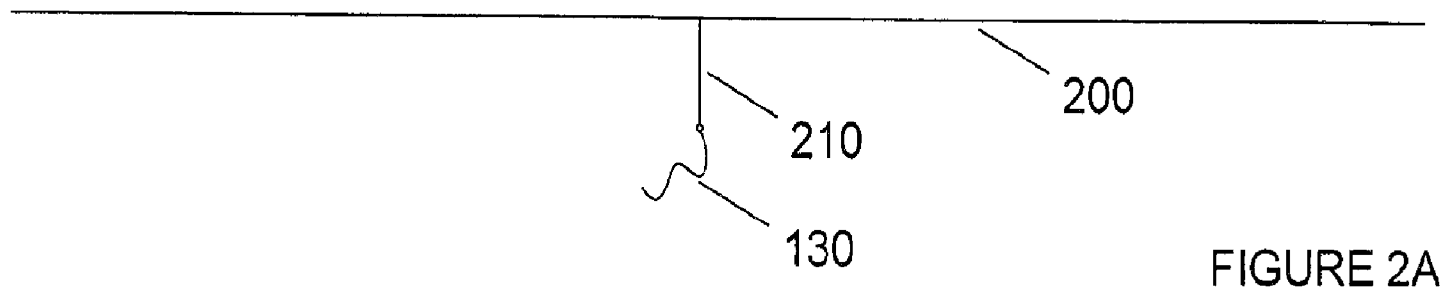

FIGURE 2A

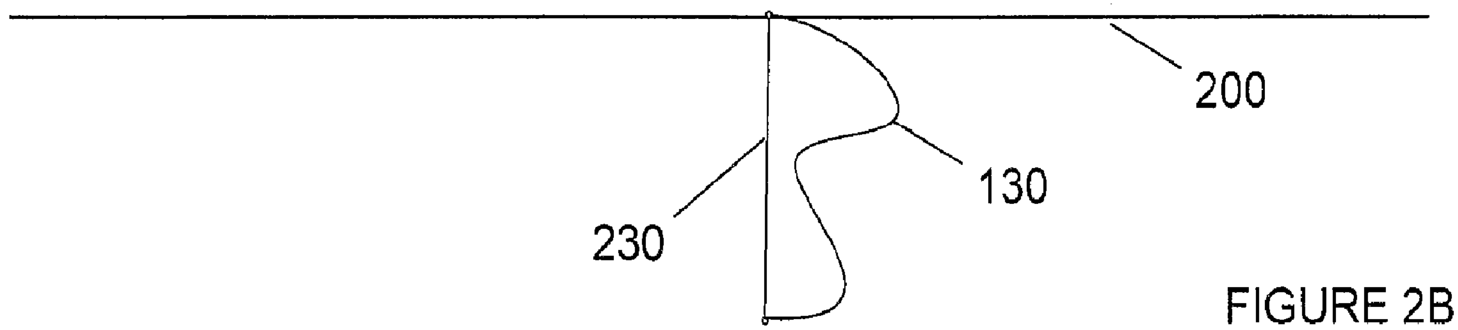

FIGURE 2B

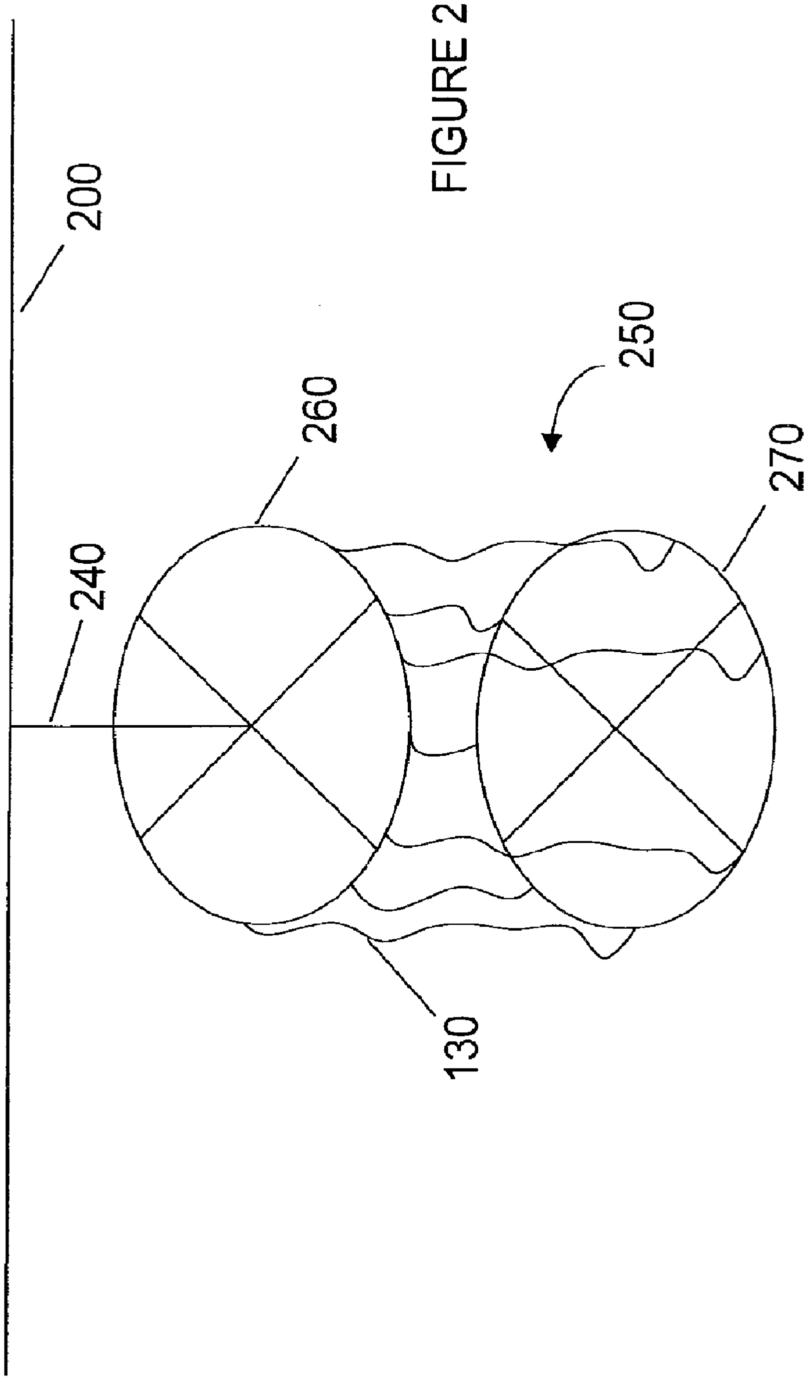
200
240
260
250
270
130

FIGURE 2C

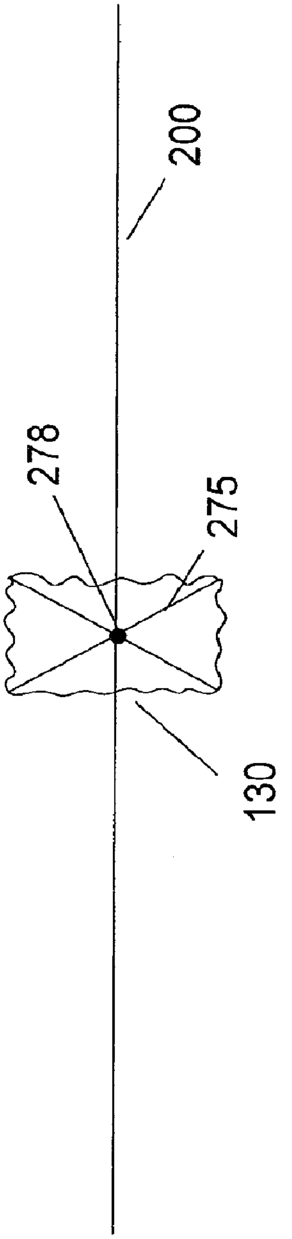
200
278
275
130

FIGURE 2D

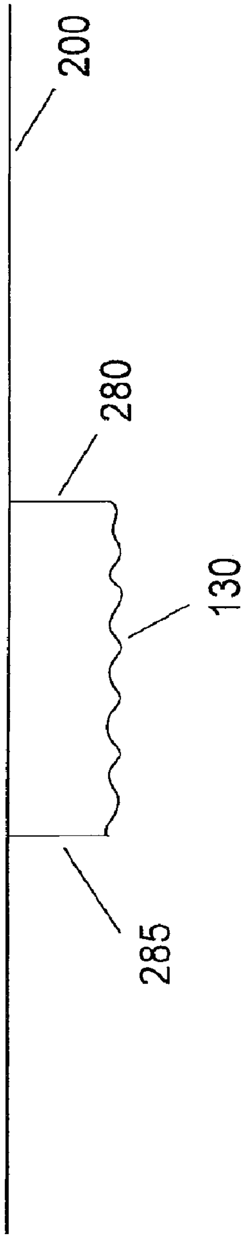
200
280
130
285

FIGURE 2E

FIGURE 2F

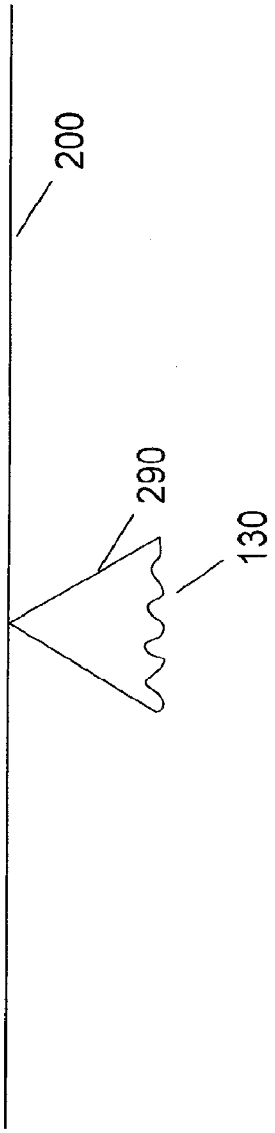

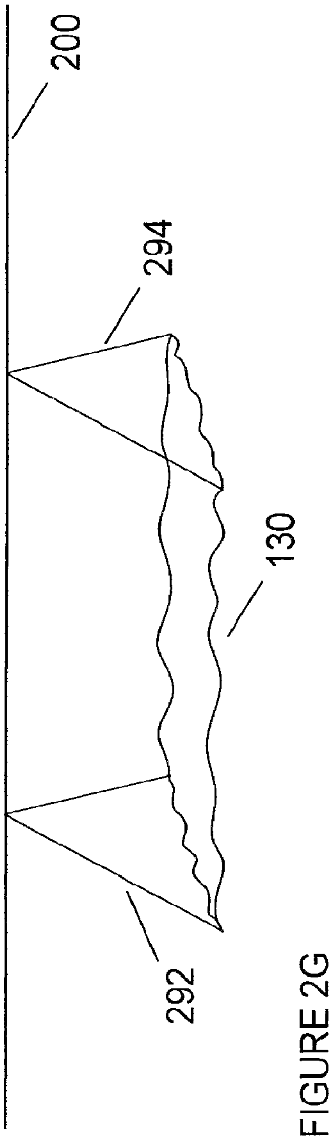
200
294
130
292

FIGURE 2G

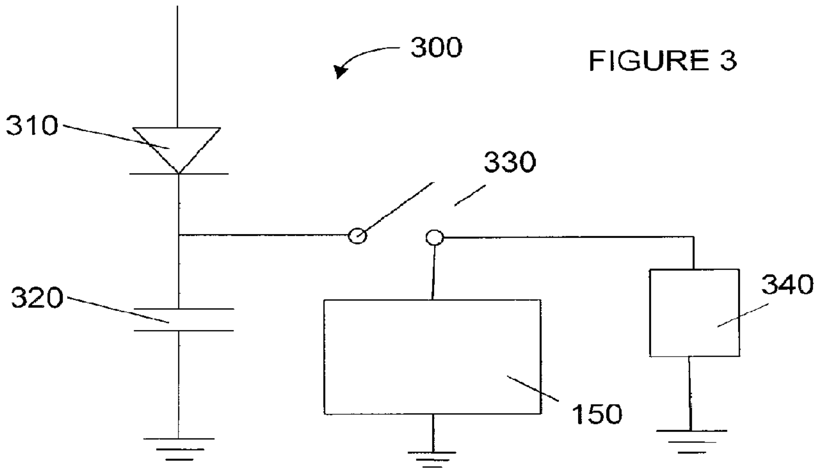
300
310
330
320
340
150

FIGURE 3

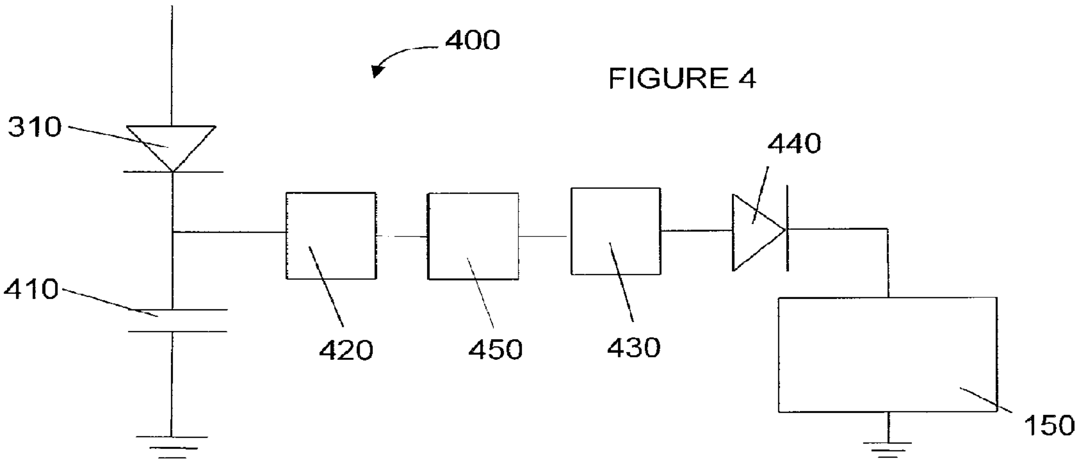
400
310
440
410
420
450
430
150

FIGURE 4

FIGURE 5

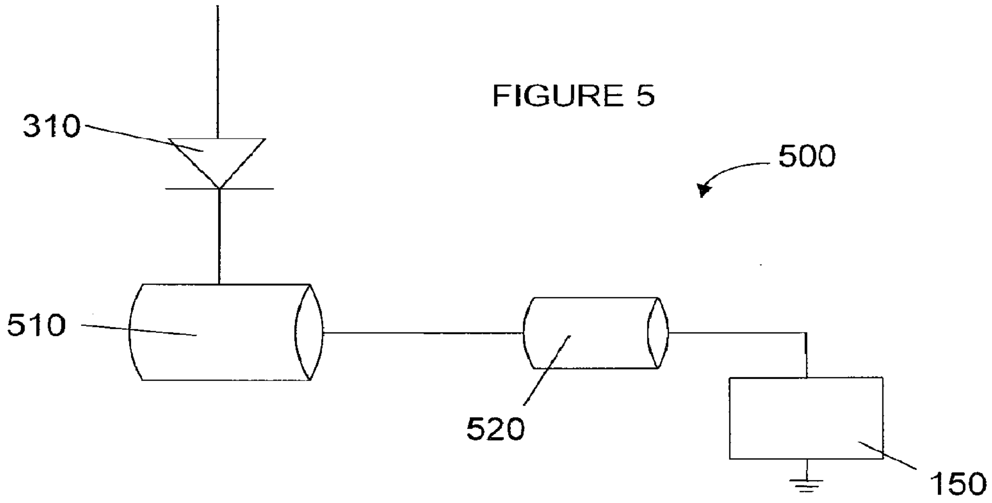
310
500
510
520
150

FIGURE 6

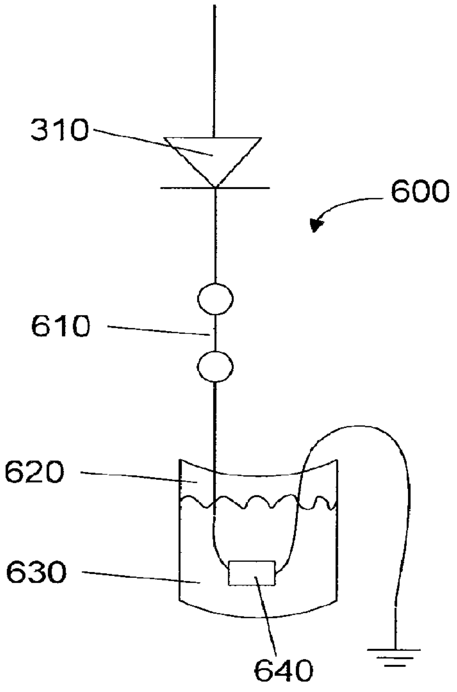
310
600
610
620
630
640

FIGURE 7

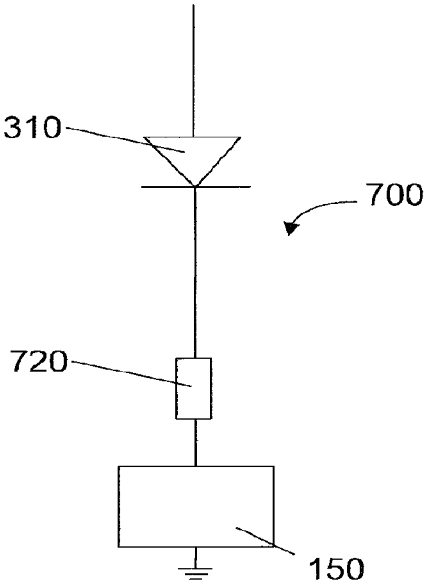
310
700
720
150

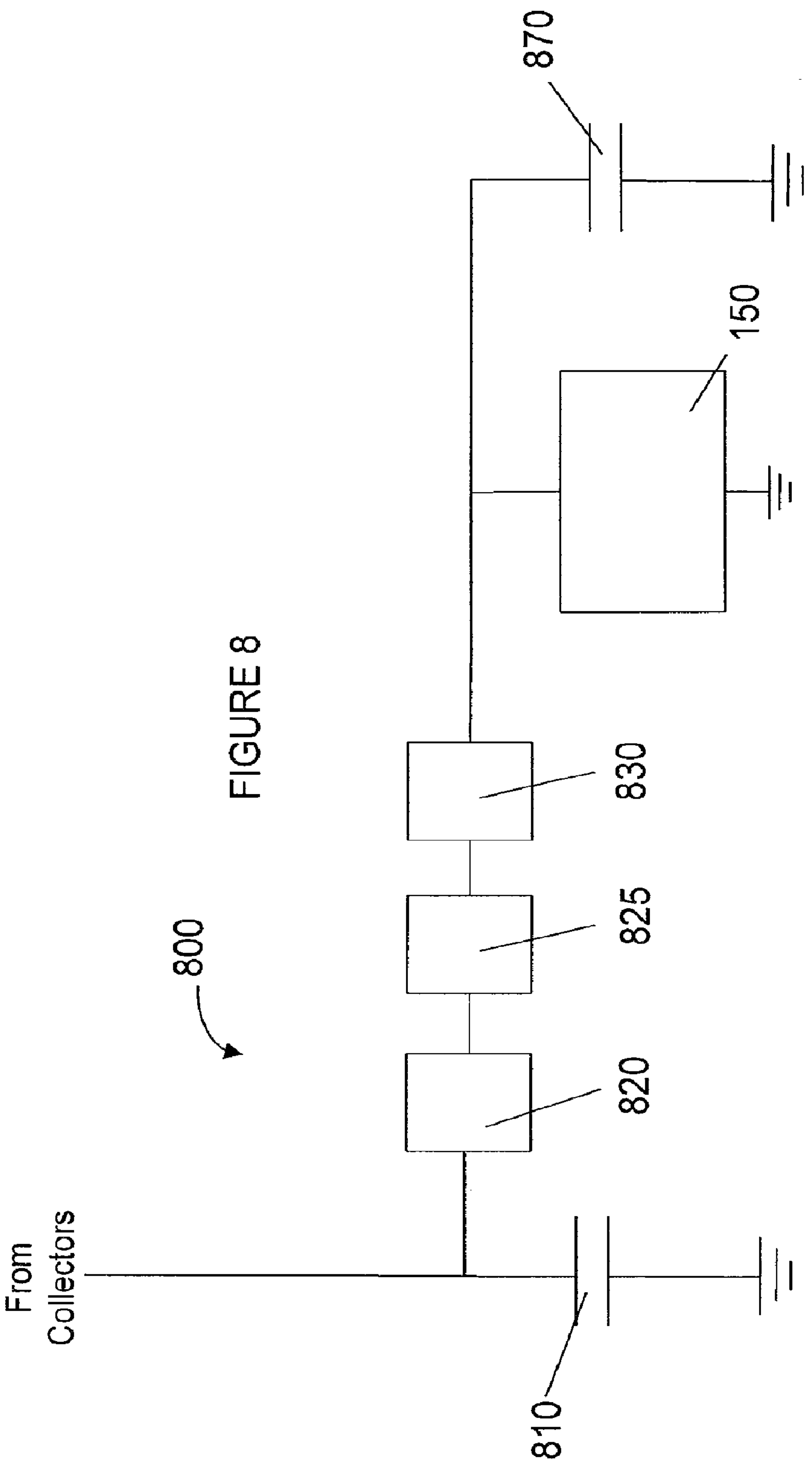
From Collectors
810
800
820
825
830
150
870

FIGURE 8

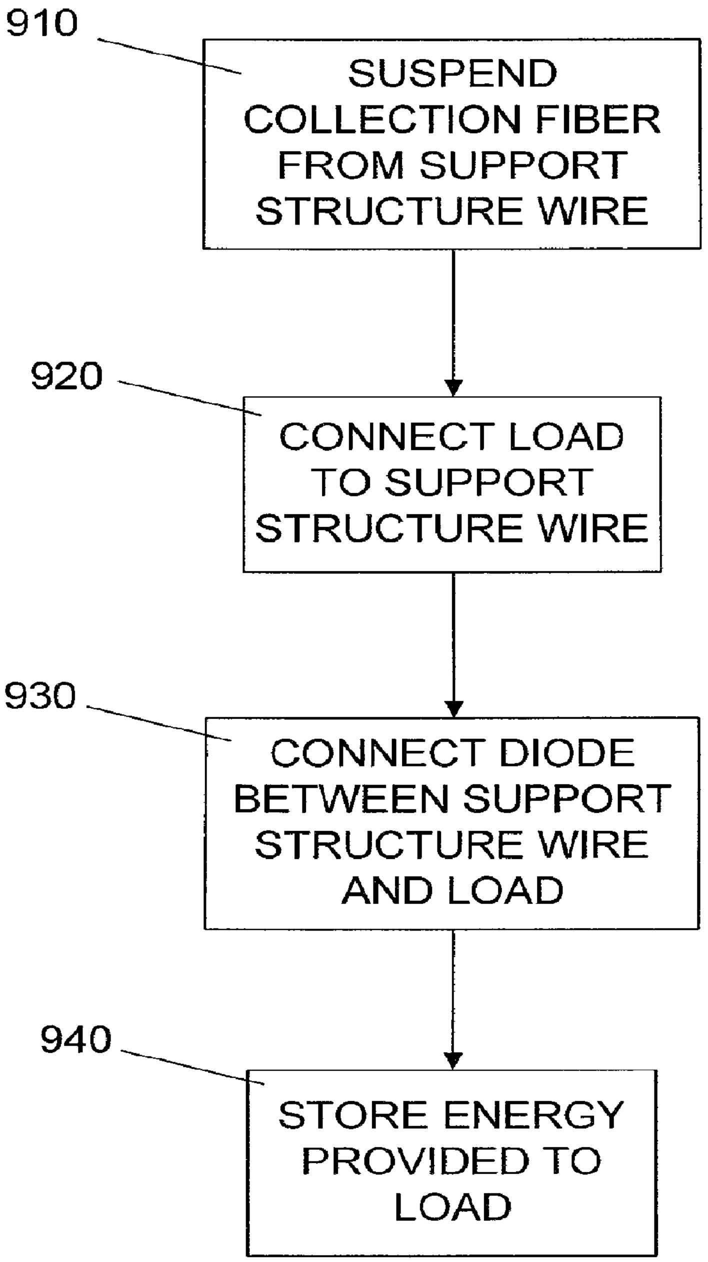
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

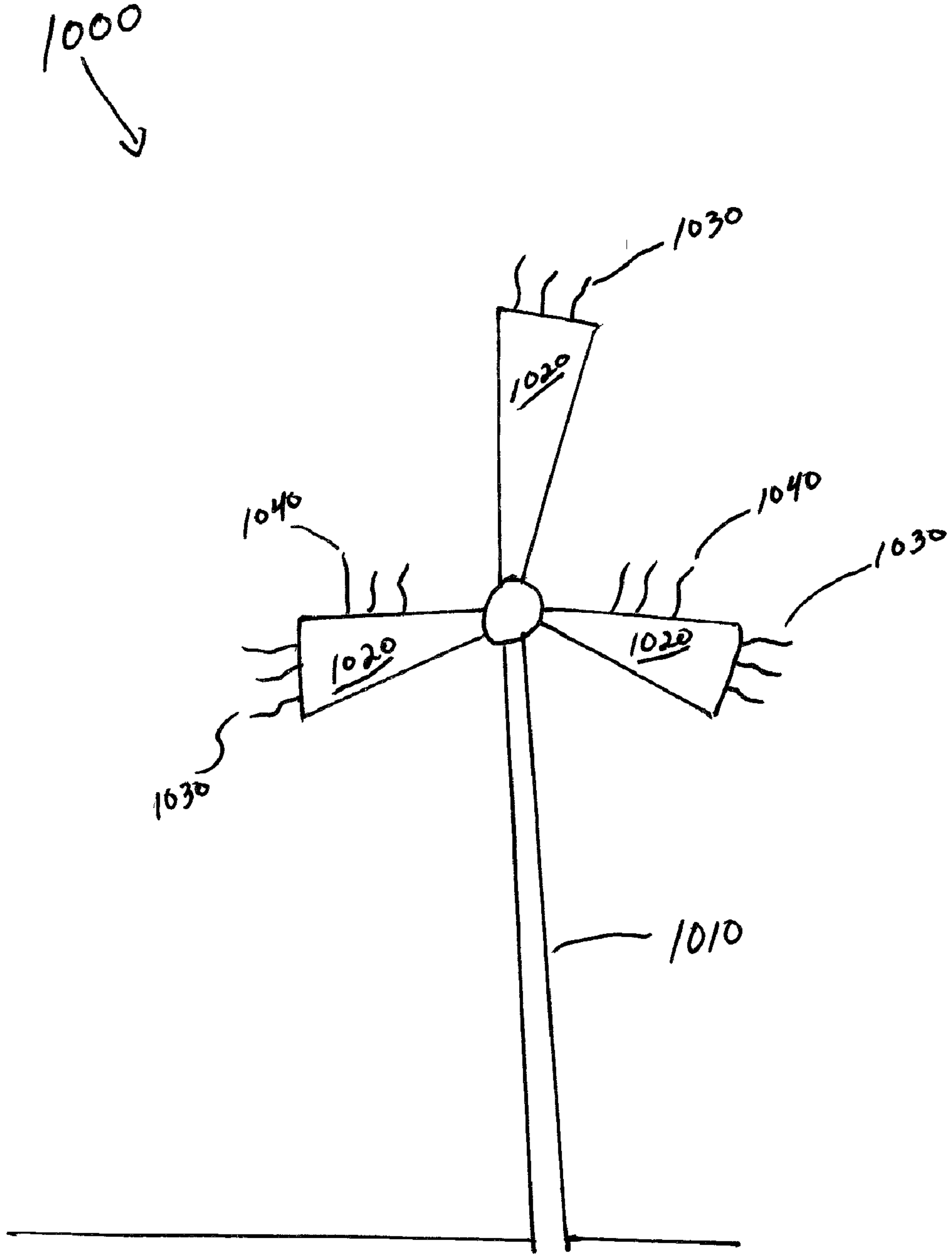
1000
1030
1020
1040
1040
1030
1020
1020
1030
1010

FIG. 10

ENERGY COLLECTION

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation in part application of U.S. patent application Ser. No. 12/255,130 filed on Oct. 21, 2008, which is a continuation application of U.S. patent application Ser. No. 11/358,264, filed on Feb. 21, 2006, which are both incorporated by reference herein.

TECHNICAL FIELD

[0002] The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

[0003] The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

[0004] In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

[0005] Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

[0006] Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

[0007] While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

[0008] Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure wire elevated above a ground level, at least one collection fiber electrically connected to the support structure wire; a load electrically connected to the support structure wire; and a diode electrically connected between the load and at least one collection fiber.

[0009] Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection fiber from a support structure wire elevated above ground level, the fiber electrically connected to the support structure wire; providing a load with an electrical connection to the support structure wire to draw current; and providing a diode electrically connected between the collection fiber and the load.

[0010] Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

[0012] FIG. **1** is a circuit diagram of a weather energy circuit.

[0013] FIG. **2** is a perspective view of an exemplary embodiment of many energy collectors elevated above ground by a structure.

[0014] FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

[0015] FIG. **2**B is a side view of an exemplary embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

[0016] FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

[0017] FIG. **2**D is a side view of an exemplary embodiment of a support structure for multiple energy collection fibers.

[0018] FIG. **2**E is a side view of a support structure for an energy collection fiber.

[0019] FIG. **2**F is a side view of an exemplary embodiment of a support structure for an energy collection fiber.

[0020] FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

[0021] FIG. **3** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

[0022] FIG. **4** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

[0023] FIG. **5** is a circuit diagram of an exemplary embodiment of an energy collection circuit for powering a generator and motor.

[0024] FIG. **6** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

[0025] FIG. **7** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

[0026] FIG. **8** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy.

[0027] FIG. **9** is a flow diagram of an exemplary embodiment of collecting energy with a collection fiber.

[0028] FIG. **10** is a system diagram of an exemplary embodiment of a windmill with energy collectors.

DETAILED DESCRIPTION

[0029] Electric charges on conductors reside entirely on the external surface of the conductors, and tend to concentrate more around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An exemplary embodiment of this disclosure takes advantage of this property, among others, to collect and use the energy generated by an electric field in the atmosphere. Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An exemplary embodiment utilizes carbon or graphite fibers for static electricity collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including aluminum or steel, but most notably, copper. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

[0030] An exemplary embodiment of the collection fibers as collection device **130** includes graphite or carbon fibers. Graphite and carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, or graphite fibers are examples that demonstrate exemplary collection ability.

[0031] In at least one exemplary embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than 100 volts per meter under some conditions. When support wire **120** is suspended in the air at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

[0032] A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An exemplary embodiment of the collection device comprises the diode with no collection fiber. A preferred embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restricts energy that may be collected through one fiber from escaping through another fiber.

[0033] Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2**A-**2**G using a collection fiber embodiment. FIG. **2**A presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**.

[0034] Likewise, FIG. **2**B shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

[0035] FIG. **2**C presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

[0036] FIG. **2**D presents another exemplary embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

[0037] FIG. **2**E provides another exemplary embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

[0038] FIG. **2**F presents another exemplary embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

[0039] FIG. **2**G provides two supports as provided in FIG. **2**F such that at least two support members **292**, **294** may be connected to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and between multiple support structures. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**.

[0040] FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load. Capacitor **320** maybe connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

[0041] In an exemplary embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel. Non-limiting examples of spark gap **420** include mercury-reed switches and mercury-wetted reed switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground. The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

[0042] FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

[0043] FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes maybe placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches and mercury-wetted reed switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel.

[0044] FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit. Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

[0045] FIG. **8** presents exemplary circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches and mercury-wetted reed switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

[0046] FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

[0047] A windmill is an engine powered by the energy of wind to produce alternative forms of energy. They may, for example, be implemented as small tower mounted wind engines used to pump water on farms. The modern wind power machines used for generating electricity are more properly called wind turbines. Common applications of windmills are grain milling, water pumping, threshing, and saw mills. Over the ages, windmills have evolved into more sophisticated and efficient wind-powered water pumps and electric power generators. In an example embodiment, as provided in FIG. **10**, windmill tower **1000** of suitable height and/or propeller **1020** of windmill tower **1000** may be equipped with energy collecting fibers **1030**, **1040**. Collecting fibers **1030**, **1040** may turn windmill **1000** into a power producing asset even when there is not enough wind to turn propellers **1020**. During periods when there is enough wind to turn propellers **1020**, collecting fibers **1030**, **1040** may supplement/boost the amount of energy the windmill produces.

[0048] Windmill **1000**, properly equipped with ion collectors **1030**, **1040**, such as non-limiting example carbon fibers, can produce electricity: 1) by virtue of providing altitude to the carbon fiber to harvest ions, and 2) while the propeller is turning, by virtue of wind blowing over the carbon fiber producing electricity, among other reasons, via the triboelectric effect (however, it is also possible for the triboelectric effect to occur, producing electricity, in winds too weak to turn the propeller).

[0049] There are at least two ways that energy collectors may be employed on or in a windmill propeller to harvest energy. Propellers **1020** may be equipped with energy collectors **1030**, **1040** attached to, or supported by, propeller **1020** with wires (or metal embedded in, or on propeller **1020**) electrically connecting energy collectors **1030**, **1040** to a load or power conversion circuit. There may be a requirement to electrically isolate energy collectors **1030**, **1040**, which is added to propeller **1020**, from electrical ground, so that the energy collected does not short to ground through propeller **1020** itself or through support tower **1010**, but rather is conveyed to the load or power conversion circuit. Energy collectors may be connected to the end of propellers **1020** such as collectors **1030**. Alternatively, energy collectors may be connected to the sides of propellers **1020** such as collectors **1040**.

[0050] Alternatively, propeller **1020** may be constructed of carbon fiber or other suitable material, with wires (or the structural metal supporting propeller **1020** may be used) electrically connecting to a load or power conversion circuit. In the case of propeller **1020** itself being constructed of carbon fiber, for example, the fiber may be 'rough finished' in selected areas so that the fiber is "fuzzy." For example, small portions of it may protrude into the air as a means of enhancing collection efficiency. The fuzzy parts of collectors **1030**, **1040** may do much of the collecting. There may be a requirement to electrically isolate carbon fiber propeller **1020** from electrical ground, so that the energy it collects does not short to ground through metal support tower **1010**, but rather is conveyed to the load or power conversion circuit. Diodes may be implemented within the circuit to prevent the backflow of energy, although diodes may not be necessary in some applications.

[0051] In an alternative embodiment, windmill **1000** may be used as a base on which to secure an even higher extension tower to support the energy collectors and/or horizontal supports extending out from tower **1010** to support the energy collectors. Electrical energy may be generated via ion collection due to altitude and also when a breeze or wind blows over the collectors supported by tower **1010**.

[0052] Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

[0053] It should be emphasized that the above-described embodiments of the present disclosure, particularly, any "preferred" embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A method of collecting energy comprising:

suspending at least one collection device with a high density of points per area of the device from a windmill, the at least one collection device electrically connected to the windmill; and

providing a load with an electrical connection to the at least one collection device to draw current.

2. The method of claim **1**, wherein the collection device collects energy by triboelectric effect.

3. The method of claim **1**, wherein the collection device comprises a collection fiber.

4. The method of claim **1**, wherein the collection device comprises a diode and a collection fiber and the diode is electrically connected between the collection fiber and the load.

5. The method of claim **1**, further comprising storing energy provided to the load.

6. The method of claim **5**, wherein storing energy provided to the load comprises storing energy in a capacitor or an inductor.

7. The method of claim **3**, wherein the collection fiber comprises carbon fiber or graphite fiber.

8. A system of energy collection comprising:

a windmill;

at least one collection device with a high density of points per area of the device electrically connected to the windmill; and

a load electrically connected to the at least one collection device.

9. The system of claim **8**, wherein the collection device collects energy by triboelectric effect.

10. The system of claim **8**, wherein the collection device comprises a collection fiber.

11. The system of claim **8**, wherein the collection device comprises a collection fiber and a diode electrically connected between the load and the collection fiber.

12. The system of claim **9**, wherein the diode is elevated relative to the ground level.

13. The system of claim **10**, wherein the collection fiber comprises a carbon fiber or a graphite fiber.

14. The system of claim **8**, further comprising a diode electrically connected between the at least one collection device and the support structure.

15. The system of claim **8**, further comprising: a switch connected in series between the at least one collection device and the load; and a capacitor connected in parallel with the switch and the load.

16. The system of claim **15**, wherein the switch comprises an interrupter connected between the plurality at least one collection device, and wherein the interrupter comprises at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

17. The system of claim **8**, further comprising: a motor for providing power, the motor connected between the at least one collection device and the load; and a generator powered by the motor.

18. The system of claim **8**, further comprising a fuel cell between the support structure and the load.

19. The system of claim **18**, wherein the fuel cell produces hydrogen and oxygen.

20. A system of collecting energy comprising:

means for suspending at least one collection device with a high density of points per area of the device from a windmill structure, the at least one collection device electrically connected to the means for suspending;

means for inducing current flow, the means for inducing current flow electrically connected to the means for suspending; and

means for restricting the backflow of charge carriers, the means for restricting the backflow of charge carriers electrically connected between the at least one collection device and the means for inducing current flow.

* * * * *

Sept. 7, 1965 L. G. SMITH 3,205,381

IONOSPHERIC BATTERY

Filed March 9, 1962

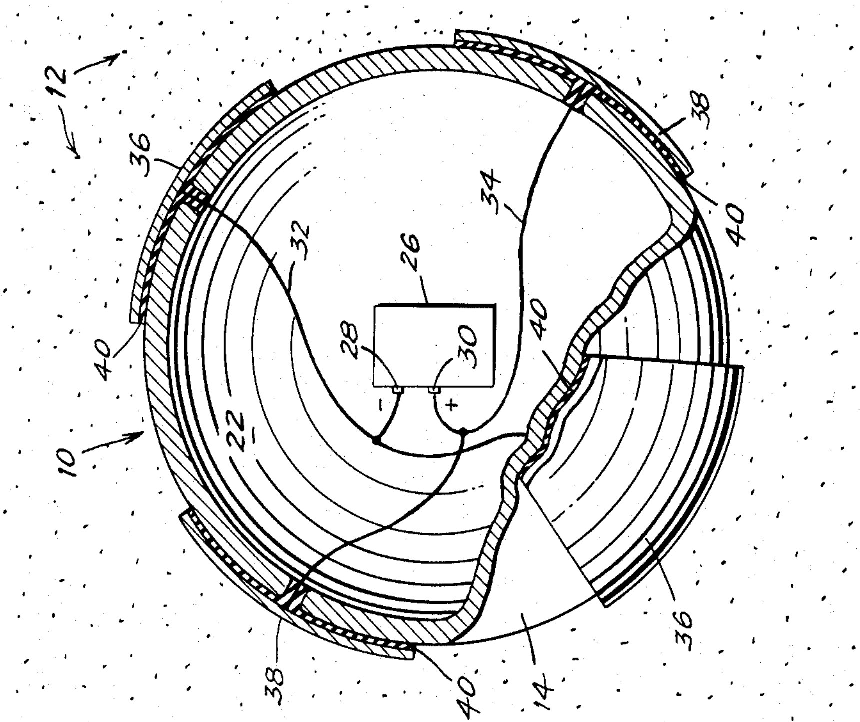
12
36
38
34
40
32
26
28
30
40
22
10
40
38
40
14
36
-
+

Fig. 2

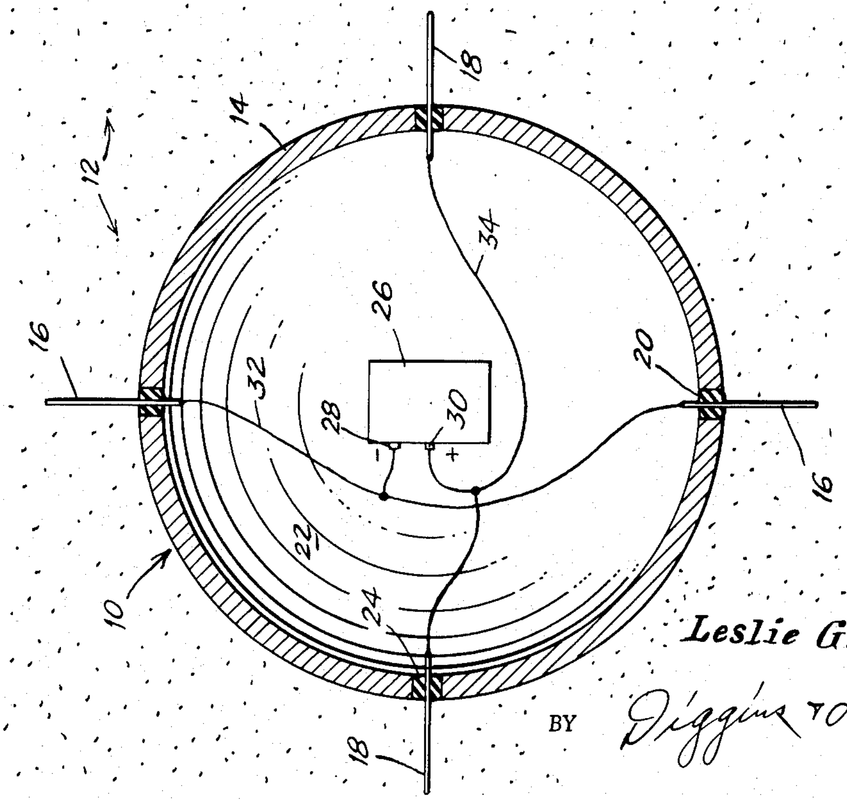
18
14
12
34
16
26
32
28
30
20
16
22
10
24
18
-
+

Fig. 1

INVENTOR

Leslie G. Smith,

BY Diggins & Le Blanc

ATTORNEYS

3,205,381
IONOSPHERIC BATTERY

Leslie G. Smith, Waltham, Mass., assignor, by mesne assignments, to the United States of America as represented by the Administrator of the National Aeronautics and Space Administration

Filed Mar. 9, 1962, Ser. No. 178,721
11 Claims. (Cl. 310—5)

This invention relates to a method and apparatus for tapping the enormous reservoir of energy existing in the ionosphere and more particularly for furnishing electrical power to space vehicles.

Within recent years various space probes and space vehicles such as satellites have been launched by man in an effort to learn more about the earth and the solar system within which he lives. Each of these various space craft has employed numerous electrical circuits which are utilized to detect and transmit back to earth certain information ascertained by the craft as it moves through space. In view of the presence of such electrical circuits, some source of electrical power must be made available whereby these circuits can be energized. In the past there have been three main sources of electrical power carried by space vehicles for energizing these circuits, namely, (1) conventional storage batteries, (2) solar batteries and (3) a combination of the conventional storage and solar batteries. Although these various electrical power sources have functioned with some degree of success each has certain serious limitations which cannot be readily overcome.

In the first instance, the conventional storage battery may be of either the wet or dry type but, due to its superior storage capacity, the wet cell type is used nearly exclusively in space vehicles. This additional storage capacity of the wet cell battery is of great importance since the useful life of the vehicle is usually dependent upon the proper operation of its electrical circuits which are, in turn, directly dependent upon the capacity and life of the available power source. Use in space vehicles of a storage battery that utilizes a liquid electrolyte presents several problems that have not been completely solved, however, nor is there any apparent solution to some of these problems. For example, to increase the output capacity of the battery and thereby extend the useful life of the satellite a greater amount of electrolyte must be carried by the battery. But to increase the quantity of electrolyte carried by the battery results in both an increase in size and weight of the power source and space vehicle which cannot, in most instances, be tolerated. Furthermore, the electrolyte may not function correctly if it is subjected to the extreme conditions of space such as, for example, the changes in temperature that are normally encountered. Thus, to insure the proper operation of the wet cell battery involved in powering a space vehicle and to prevent any leakage that might occur should the battery be damaged, some shielding or insulating of the battery is usually involved which, needless to say, adds materially to the cost and weight of the vehicle.

To overcome the space, weight and leakage problems encountered when storage batteries of the conventional wet cell type are used, space engineers have turned to the recently developed solar cell or battery which converts the light rays received from the sun directly into electrical energy. While these solar batteries do not require any type of electroylte and are, therefore, relatively light in comparison to the wet cell storage battery, they are expensive and their inherent principle of operation presents one problem which cannot be overcome, namely, the inability of the solar cell to produce power while the space vehicle is shaded from the sun such as occurs when a satellite is in the shadow cast by the earth.

In an effort to overcome this loss of power which occurs each time a satellite passes into the shadow cast by the earth or some other heavenly body, a storage battery of the wet cell type was added to the vehicle. While the satellite is in the sunlight the solar battery powers the various electrical circuits and charges the storage battery so that, once the sunlight is no longer available, the storage battery can be switched on to furnish the necessary electrical power to the vehicle. While this combined system functions satisfactorily, there is added to the already existing problems of weight, size and electrolyte leakage the problem of incorporating a battery charging and switch-over system between the solar and storage batteries. The addition of this type of system to the power source is undesirable since it adds unduly to the complexity, sensitivity and cost of the space vehicle.

According to the present invention, it has been found that some of the foregoing disadvantages of space vehicle power sources can be overcome by employing an ionospheric type battery which utilizes the ionosphere or other ionized medium existing in outer space as a supplemental source of energy. Since power is a scarce commodity essential for the operation of the space vehicle and for the proper functioning of its experiments, any means for supplying additional power while adding only slightly to the weight of the vehicle, increases the efficiency and life of the space vehicle. Although it has been known for some time that the ionosphere contains an enormous reservoir of energy in the form of excited and ionized atoms and molecules, no practical method has heretofore been found whereby this energy could be utilized as a source of power for earth satellites even though much thought has been given to this problem. The operation of the present invention is based upon the discovery that the ionosphere can be utilized as a part of supplemental power source for the load circuit in a space vehicle when said vehicle is equipped with electrodes having different work functions which, in turn, are electrically connected through the load circuit.

Accordingly, a primary object of this invention is to provide a method of producing electrical power from ionized media existing in nature.

Another object of this invention is to provide a simple, highly reliable supplemental power source for space vehicles.

Yet another object of this invention is to provide an ionospheric battery for use by satellites.

Still another object of this invention is to provide an ionospheric battery which utilizes the ionosphere as an electrolyte.

A further object of this invention is to provide a supplemental power source for space vehicles that will operate when shielded from the sun.

Yet a further object of this invention is to provide a power source for a satellite that is light and inexpensive to produce.

Another object of this invention is to provide a satellite with various shaped electrodes constructed of metals having different contact potentials.

These and further objects and advantages of this invention will become more apparent upon reference to the following description, claims and the appended drawings wherein:

FIGURE 1 is a cross sectional view of an orbiting satellite equipped with a power source constructed in accordance with one embodiment of the present invention; and

FIGURE 2 is a view in partial section of an orbiting satellite equipped with a power source constructed in accordance with a second embodiment of the present invention.

With continuing reference to the accompanying drawings wherein like reference numerals designate similar parts throughout the various views, and with initial attention directed to the embodiment of FIGURE 1, reference numeral **10** is used to generally designate a space satellite which is orbiting or otherwise passing through an ionized layer or medium such as, for example, the ionosphere which surrounds the earth. As is well known, the ionosphere contains an enormous number of excited and ionized atoms and molecules which are generated or produced when the gasses surrounding the earth are subjected to bombardment by high energy particles and rays such as, for example, ultra-violet or cosmic rays that originate throughout the solar system. Although these various ionized particles created by this bombardment are invisible to the eye they are represented pictorially by the various dots **12** for purposes of clarity.

The satellite **10** consists of an envelope or outer skin surface **14** which may be constructed of metal, ceramic or any other like material which is suitable for space vehicles of this type. As seen in FIGURE 1, a plurality of electrically conductive electrodes **16–18**, which are shown in the form of rods and are constructed of materials having different contact potentials (work functions), are mounted in a spaced relationship in the medium **12** through apertures **20** in the envelope **14** of the satellite. In order to prevent a leak from occurring between the normally pressurized interior **22** of the satellite and the ionosphere, which is a near vacuum, and to insulate the electrodes in case a metal is used for the envelope **14**, vacuum tight insulator feed-through elements **24** are fitted around the electrodes **16–18** and are securely mounted within the aperture **20**. These feed-through elements may be of any suitable type so long as they are constructed to withstand the changes in environmental conditions encountered in space without losing their insulative and sealing qualities.

Carried within the cavity or interior **22** of the satellite **10** is a load circuit **26** which may be, for example, a transmitter or any other electrical device that requires a source of electrical power for its proper operation. The load circuit **26** is mounted to the interior wall of the satellite by suitable means (not shown) and has its positive and negative power input terminals **28–30** connected to the various electrodes **16–18** respectively, by suitable electrical conductors **32–34**.

The overall principle and operation of the ionospheric battery in relationship to the satellite **10** can be explained substantially in the following manner. While the satellite remains on earth prior to its being hurled into or through an ionized medium, the electrodes **16–18** are at substantially the same potential since very few ionized particles are present. Thus, no current will flow through the load circuit **26**. However, once the satelllite is launched and it enters an ionized medium a potential difference will be established between the electrodes **16–18** and an electrical current will begin to flow through the load circuit **26**. This difference in electrical potential that exists between the electrodes **16–18** while the satellite is in an ionized medium is attributed to the fact that the electrodes **16** are constructed of a material having a different contact potential or work function from that of the electrodes **18**. This difference in contact potential between the electrodes **16–18** will result in the free electrons of the ionized medium being collected on one group of electrodes, for example the electrodes **16**, while the positive ions produced when the gas is ionized will be collected on the other electrodes **18**. With the electrodes **16** charged to a negative potential in relationship to electrodes **18**, a predetermined amount of current will flow through the load circuit **26** and can thus be utilized to power an electrical circuit such as the transmitter mentioned above.

The exact material from which the two types of electrodes **16** and **18** are constructed will depend upon several factors such as, for example, the amount of output power desired, the material readily available for constructing the electrodes, the susceptibility of the material to damage or destruction when exposed to the environmental conditions existing in space, and various other considerations that will be readily apparent to a worker in the battery field.

While the electrodes **16–18** used in the ionospheric battery of FIGURE 1 have been described as rods or whip-like elements it is to be realized that the electrodes may take other shapes. By way of example, a plurality of flat, plate-like electrodes **36–38** are shown mounted on a satellite in FIGURE 2. These electrodes are constructed of materials having different contact potentials and are insulated from the outer surface **14** of the satellite when this is necessary by an insulative layer **40** of any suitable material having the desired insulative properties under the severe conditions existing in space. The use of these plate electrodes **36–38** has the advantage of increasing the surface area exposed to the ionized medium thus permitting a larger output current to be produced without the necessity of using excessively long electrodes as might be the case if the rod electrodes of FIGURE 1 were used.

The output current of the battery is dependent upon the degree to which the medium is ionized and to the extent that the electrode surface area is exposed to the medium. An increase in either results in an increase in the efficiency of the battery. From an analysis of various experiments it is expected that a current of about two microamps per cm.^{2} of electrode surface at a potential of one volt would be produced if the satellite is orbiting in the ionosphere at an altitude of approximately 350 km. and electrodes of gold and platinum are used.

It will also be readily apparent that the envelope of the satellite itself could be constructed in sectional form and of materials having different contact potential thus eliminating the necessity of attaching different electrodes to the satellite or the slight increase in weight necessarily involved. Of course the various sections of material making up the envelope of the satellite would have to be insulated from one another to prevent electrical shorting from occurring.

It will be apparent from the foregoing that the power source of this invention is extremely light, rugged and involves a minimum of cost to construct. By using the vast natural reservoir of energy existing in the form of ionized gases for producing the electrical power necessary to energize the various electrical circuits of a space vehicle a practically indefinite source of electrical power is made available to the vehicle. An ionospheric battery constructed in accordance with this invention is also much more reliable in operation than other heretofore known power sources since no liquid electrolyte is involved which can leak, deteriorate or otherwise become spent, nor is the operation of the battery dependent upon its position in relationship to a source of light.

The invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. The present embodiments are therefore to be considered in all respect as illustrative and not restrictive, the scope of the invention being indicated by the appended claims rather than by the foregoing description, and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced therein.

What is claimed and desired to be secured by United States Letters Patent is:

1. A space vehicle equipped with a generator of electrical energy operable in the ionosphere as a supplemental source of potential for a load thereof, said generator comprising at least two electrode means on said vehicle and positioned to contact said ionosphere when said vehicle is in said ionosphere, said two electrode means each having different contact potentials, one of said two electrode means being gold and the other being

platinum, and electrical connector means connecting said two electrode means to said load.

2. Apparatus for generating electrical energy from the ionosphere comprising, a pair of spaced electrode means having different contact potentials positioned to contact said ionosphere so that a direct-current electric potential is established between said electrode means, load circuit means, and means electrically connecting said load circuit means to said electrode means whereby said load circuit means is powered by the direct-current electrical potential existing between said electrodes.

3. Apparatus for generating electrical energy from the ionosphere comprising, carrier means, a first electrode having a first contact potential mounted on said carrier means, a second electrode having a second contact potential spaced from said first electrode and mounted on said carrier means, said first and second electrodes being positioned on said carrier means so that they will be in contact with said ionosphere when said apparatus is placed in the ionosphere so that a direct-current electric potential is established between said first and second electrodes, and an electrical load circuit connected between said first and second electrodes, said load circuit being powered by the direct-current potential produced by said first and second electrodes.

4. Apparatus according to claim 3 wherein said first and second electrodes extend away from said carrier means.

5. Apparatus according to claim 3 wherein said first and second electrodes form a portion of said carrier means.

6. In a space vehicle adapted to pass through ionized areas in space, a power source for said vehicle while it is passing through said ionized areas comprising, a plurality of electrode means located on the outer surface of said vehicle, said electrode means being constructed of materials having different contact potentials, and means connecting said electrode means together through a load circuit whereby said circuit receives a source of direct-current as said space vehicle passes through said ionized areas.

7. In a space vehicle according to claim 6 wherein said electrode means are rod shaped.

8. In a space vehicle according to claim 6 wherein said electrode means are plate shaped.

9. A satellite comprising an envelope, electrical load means carried in said envelope, apertures extending through said envelope at spaced points, electrode means having different contact potentials, said electrode means passing through said apertures, means in said aperture for insulating said electrode means from said envelope, and electrical conductor means connected between said electrode means and said electrical load means whereby said electrical load means is energized by the direct-current electric potential established when said electrode means passes through an ionized medium.

10. A space satellite comprising an envelope, an electrical load circuit carried in said envelope, a power supply for said electrical load circuit comprising a first electrode having a first contact potential connected to said electrical load circuit and extending through said envelope and a second electrode having a second contact potential connected to said electrical load circuit and extending through said envelope, said first and second electrodes being insulated from one another so that said electrical load circuit is powered by the direct-current electric potential established by said first and second electrodes coming in contact with an ionized medium.

11. The method of powering electrical circuit means carried by a space vehicle comprising the steps of positioning first and second electrode means having different work functions on said vehicle, passing said vehicle through the ionosphere so that said first and second electrode means are in contact therewith and produce a direct-current potential between them, and connecting said first and second electrodes to said electrical circuit means by means of electrical connector means, whereby said direct-current potential is made available for operating said circuit.

References Cited by the Examiner

UNITED STATES PATENTS

414,943	11/89	Dewey	310—5 X
911,260	2/09	Pennock	310—5 X
2,303,321	12/42	Bennett	175—264
2,445,305	7/48	Hochgesaug	250—83.6
2,510,397	6/50	Hansell	171—330
2,802,113	8/57	Ohmart	250—83.6
2,864,012	12/58	Thomas	310—3
2,900,535	8/59	Thomas	310—3
3,021,472	2/62	Hernqvist	322—2

LLOYD McCOLLUM, *Primary Examiner.*

JOHN H. MACK, *Examiner.*

Nov. 3, 1959 D. G. TIPOTSCH 2,911,532

ION COLLECTOR FOR MASS SPECTROMETRY

Filed March 26, 1956 2 Sheets-Sheet 1

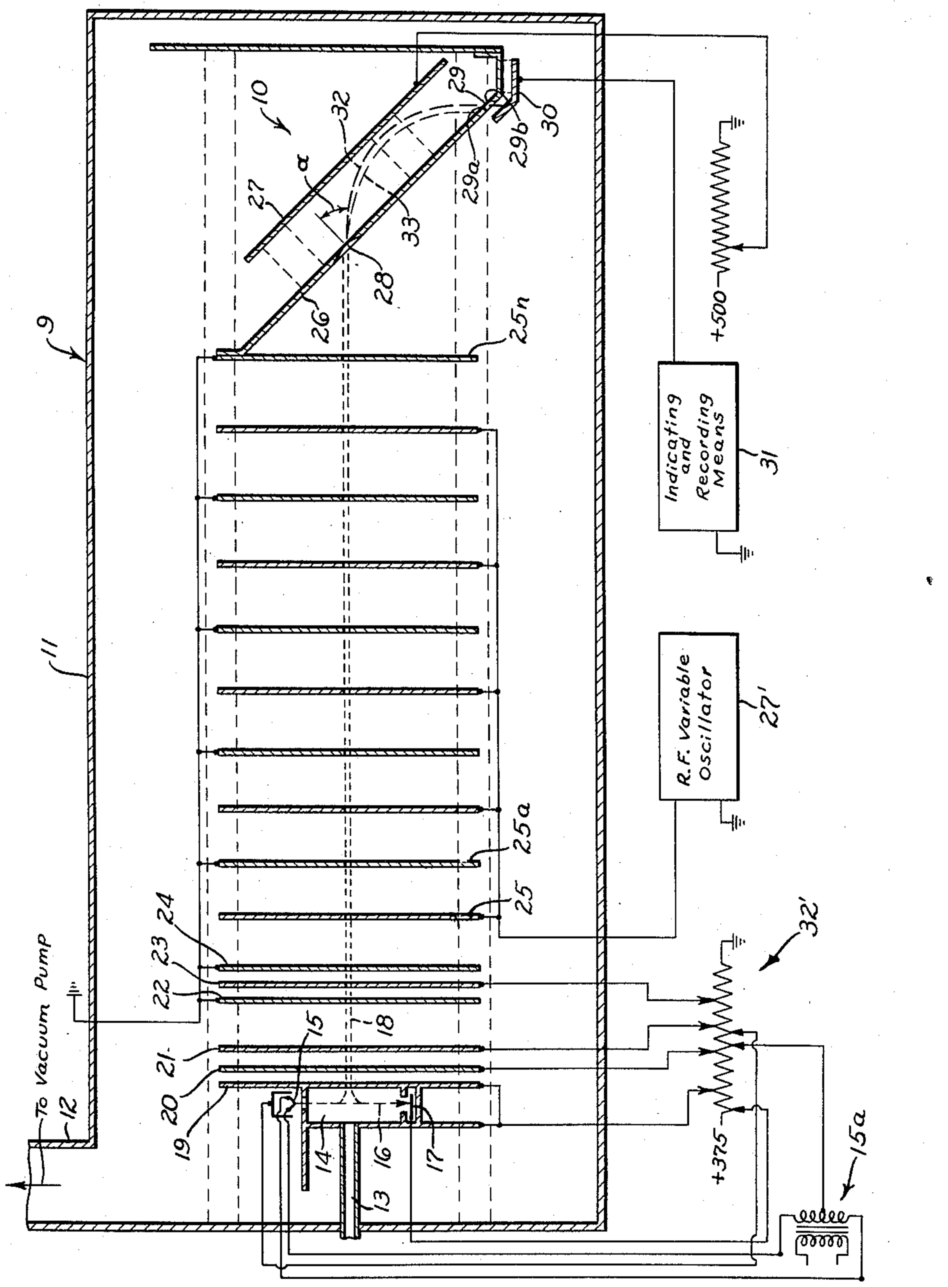
9
10
11
12
To Vacuum Pump
13
14
15
15a
16
17
18
19
20
21
22
23
24
25
25a
25n
26
27
α
28
29
29a
29b
30
32
33
+500
Indicating and Recording Means
31
R.F. Variable Oscillator
27'
32'
+375

Fig. 1.

INVENTOR.
DONALD G. TIPOTSCH

BY HIS ATTORNEYS.
HARRIS, KIECH, FOSTER & HARRIS

Nov. 3, 1959 D. G. TIPOTSCH 2,911,532

ION COLLECTOR FOR MASS SPECTROMETRY

Filed March 26, 1956 2 Sheets-Sheet 2

Fig. 2.

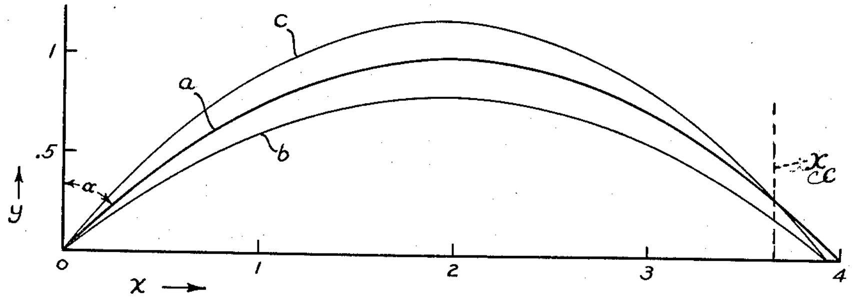

Fig. 3.

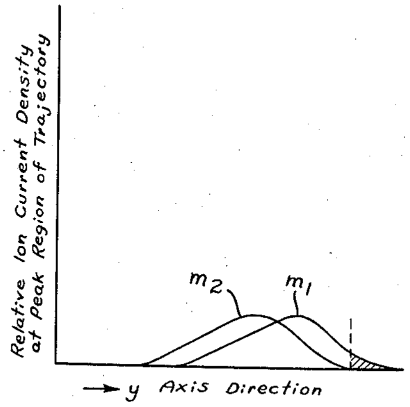

Fig. 4.

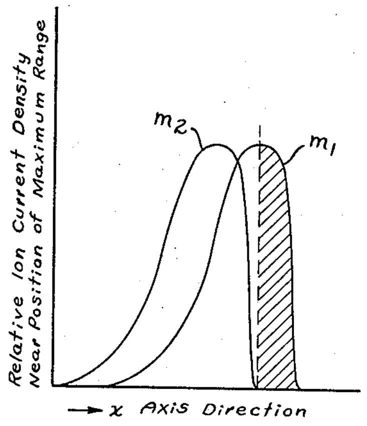

INVENTOR.
DONALD G. TIPOTSCH

BY HIS ATTORNEYS.
HARRIS, KIECH, FOSTER & HARRIS

United States Patent Office

2,911,532

Patented Nov. 3, 1959

2,911,532

ION COLLECTOR FOR MASS SPECTROMETRY

Donald G. Tipotsch, Fullerton, Calif., assignor to Beckman Instruments, Inc., Fullerton, Calif., a corporation of California

Application March 26, 1956, Serial No. 573,874

5 Claims. (Cl. 250—41.9)

The present invention relates in general to mass spectrometry and, more particularly, to an ion collector for radio frequency mass spectrometers, which may, for example, be of the type disclosed and claimed in the copending applications of George H. Hare and David R. Margetts, Serial No. 370,582, filed July 27, 1953, and Serial No. 378,756, filed September 8, 1953, now U.S. Patent 2,769,093 granted October 30, 1956.

As general background, mass spectrometers of this general type include: an ion source; ion analyzer means for selectively varying the kinetic energies of the ions emitted by the source in accordance with their respective masses, or more accurately, in accordance with their respective charge-to-mass ratios; and selective ion collector means. The analyzer provides selected ions of a selected charge-to-mass ratio with an optimum kinetic energy level while providing nonselected ions of other charge-to-mass ratios with other kinetic energy levels differing from the optimum. The ion collector separates the selected ions from the non selected ions on the basis of the energy level differences produced by the analyzer means, the ion collector providing a collecting electrode, or collecting electrode means, for collecting the selected ions. The optimum kinetic energy level with which the selected or preferred ions are provided may be either a maximum or a minimum, the ion collector of the present invention having particular utility with an instrument which maximizes the kinetic energy of the preferred ions, although it is useful also with one which minimizes the kinetic energy thereof. Further, while the ion collector of the invention is particularly applicable to a radio frequency mass spectrometer, and is considered with particular reference to a radio frequency instrument herein, it will be understood that it may be utilized in any instrument which selectively alters the kinetic energy level of ions of a preferred mass, or preferred mass-to-charge ratio, relative to nonpreferred ions of other charge-to-mass ratios.

Basically, the operation of an ion collector of the type to which the present invention relates involves projecting the ion beam emerging from the analyzer means into a substantially uniform electrostatic deflecting field at an angle of substantially 45° to the equipotential planes of such field and at an initial position in a plane of low field potential so that the ions traverse planes of successively higher field potentials. The ions constituting the beam accordingly execute parabolic trajectories within the deflecting field space varying in altitude and range in accordance with the respective entrant kinetic energies thereof. An ion having a high entrant energy level tends to attain both a higher altitude and a greater range than one of low entrant energy. This variation in trajectory makes it possible to separate ions having a selected entrant kinetic energy from those ions having different entrant kinetic energies, the selected entrant energy level corresponding to the preferred charge-to-mass ratio.

In the ion collectors of this general character which are described in the aforementioned copending applications, the ions are separated on the basis of the maximum or peak altitudes they attain in the electrostatic deflecting field. For example, in one prior ion collector, used in an instrument which imparts the maximum energy level to the selected ions, these ions are collected at the crest of their parabolic trajectory by means of a collecting electrode which intercepts only the ions attaining the highest peak altitude, but which does not intercept ions having lower maximum or peak altitudes. On the other hand, if the analyzer means of the instrument is such that it imparts the minimum energy to the selected ions, an electrode is provided in the region of maximum trajectory altitude for intercepting all ions attaining higher altitudes than that of the selected ions, the latter attaining only a minimum peak altitude and clearing the mentioned intercepting electrode to impinge on a separate, collecting electrode.

The foregoing prior ion collectors take advantage of a focusing action which causes parallel, equal-energy rays of the ion beam to converge within a narrow band at the crest of the trajectory. The lateral separation of such parallel rays results from the finite width of the ion entrance aperture through which ions are admitted into the electrostatic deflecting field. However, it has been found that while this focusing action compensates effectively for lateral displacements of the selected ions in the ion beam, it does not compensate for slight deviations in the angle at which the selected ions, or any single-energy ions, enter the deflecting field. In fact, the effect of such entrant angle deviations on resolution, and on the useful ion current obtainable at the collecting electrode, may be much more serious than the effect of lateral spacing due to the width of the entrance aperture. Thus, in an instrument wherein the maximum energy is imparted to the selected ions, it is possible for ions of lower energy than that of the preferred ions to impinge on the collecting electrode if they enter the field at an angle of less than 45° with respect to a line perpendicular to the equipotential planes of the field. Similarly, in an instrument which imparts the minimum energy to the preferred ions, ions of an energy higher than the selected ions, but entering the deflecting field at an angle of more than 45°, may miss the intercepting electrode adapted to block the nonpreferred ions and may impinge together with the preferred ions on the collecting electrode. In either event, mass resolution is seriously impaired. In some instances (in instruments wherein selected ions attain maximum energy) it has been necessary in order to fully resolve adjoining masses above mass 40, for example, to reduce the deflecting potential and accept only the 10% of the available ions of the selected mass which attain the maximum trajectory altitude and which are free from interfering ions of the adjoining mass. Obviously, this drastically reduces the ion current available at the collecting electrode.

Another disadvantage of prior ion collectors which operate on the basis of ion trajectory altitudes is that when a mass spectrum is scanned, i.e., when the instrument is made responsive to successive masses as by varying the resonant frequency in radio frequency instruments, the ions of successive masses may enter the ion collector at a continuously varying angle, the reasons for this phenomenon being only incompletely understood at the present time. Thus, when collecting on the basis of variations in trajectory altitude with ion mass, the proportion of the preferred ion beam collected varies somewhat with mass. This would not be objectionable except for the fact that it is desirable to vary the deflecting potential in order to change the resolution, e.g., in order to collect a larger portion of the available current where the mass resolution is inherently high, as at low mass numbers. When this is done, it is found that the fractionating pattern, or distribution of ion fragments of various masses for a given simple molecule, changes also. For maximum operating convenience however, this should remain independent of resolution.

A primary object of the present invention is to eliminate the foregoing and various other disadvantages of prior ion collectors, irrespective of the type of mass spectrometer with which the ion collector of the invention is used.

More particularly, an important object of the present invention is to provide an ion collector which separates the selected ions from the nonselected ions on the basis of the differences in the range of the trajectories which the ions execute, instead of on the basis of the differences in the altitudes thereof.

In other words, an important object of the present invention is to separate the selected and non-selected ions at the ends of their trajectories within the deflection space, instead of at the crests thereof. As will be shown hereinafter, variations in trajectory range with variations in entrant angle are much smaller than variations in trajectory altitude with such angular variations. This permits the ion collector of the invention to achieve superior resolution and increased ion current, these being important features of the invention.

Another object of the invention is to provide an ion collector which includes electrode means defining a substantially uniform electrostatic deflecting field, means for admitting ions into such field in a beam at a mean inclination of substantially 45° to the equipotential planes of the field and at an initial position in a plane of low potential, whereby ions of different masses execute different trajectory ranges in returning to terminal positions in a plane of substantially the same low field potential, means adjacent the terminal position of the trajectory range executed by the selected ions for intercepting ions deviating therefrom in range by more than a predetermined small amount, and collecting electrode means adjacent such intercepting means for collecting selected ions which are not intercepted by the intercepting means.

A further object is to provide an ion collector which includes two spaced, parallel electrodes disposed at an angle of substantially 45° to the ion beam, one of these electrodes being maintained at a lower potential than the other to provide a substantially uniform electrostatic deflecting field therebetween. A related object is to provide the lower-potential electrode with an ion entrance aperture and to provide it with an ion intercepting edge spaced from the ion entrance aperture, such ion intercepting edge preferably being formed as an edge of an ion exit aperture. Still another object is to provide an ion collecting electrode external to the space between the deflecting electrodes and so positioned with respect to the ion exit aperture that ions not intercepted by the ion intercepting edge of such deflecting electrode impinge on the collecting electrode as a beam of substantially selected ions only.

The foregoing objects, advantages, features and new results of the present invention, together with various other objects, advantages, features and new results thereof which will be evident to those skilled in this art in the light of the present disclosure, may be attained with the exemplary embodiment of the invention illustrated in the accompanying drawings and described in detail hereinafter. Referring to the drawings:

Fig. 1 is a semidiagrammatic, longitudinal sectional view of a mass spectrometer which incorporates the ion collector of the invention;

Fig. 2 is a graph comparing variations in ion trajectory altitude and ion trajectory range with variations in entrant angle into an electrostatic deflecting field;

Fig. 3 is a graph illustrating the distribution of ion current density in the region of the ion trajectory peaks in the electrostatic deflecting field; and

Fig. 4 is a graph illustrating the variation in ion current density near the position of maximum ion trajectory range in the electrostatic deflecting field.

An instrument 9 embodying an ion collector 10 of the present invention is illustrated in Fig. 1 of the drawing, this instrument being similar in principle to that described in said applicaton Ser. No. 370,582. The means for ionizing the sample, for imparting different energies to the ions of different mass, and for collecting a selected ion species are enclosed in a vacuum-tight housing 11 connected to a vacuum pumping system, not shown, via a duct 12. More specifically, the sample, in the form of a gas or vapor, is introduced into an ion chamber 14 at a suitable pressure through an inlet tube 13. A filament 15 powered by a transformer 15*a* provides an electron stream 16 which is directed through the ion chamber 14 and collected by a positively charged electrode 17. A portion of the sample material is ionized under bombardment by the electron stream. Two sets of perforated electrodes, 19 to 21, and 22 to 24, connected to points of suitable potential on a power supply 32′, draw the ions out of the ion chamber, giving them an initial acceleration and collimating them into a narrow beam 18. The perforations in these and all subsequent electrodes may either be holes or, preferably, elongated slits. Thus, Fig. 1 may be interpreted as showing slits which are oriented perpendicularly to the plane of the drawing. In this case, however, it would be preferable to reorient the electron beam so that it too is perpendicular to the plane of the drawing.

The ion chamber 14, the electron beam forming means 15, 17 and the electrodes 19 to 24 constitute what may be termed the ion source. This is followed by a so-called analyzer means or analyzer section which comprises an array of perforated plates, 25, 25*a*, . . . 25*n*, connected alternately to one terminal of a variable radio frequency oscillator 27′ and an oscillator return point. e.g., ground, as shown. For a given accelerating potential at the ion source, and any given frequency and amplitude of the radio frequency oscillator 27′, only particles of a particular mass, or, strictly speaking, of a particular charge-to-mass ratio, will pass through the analyzer plates in resonance with the radio frequency signal. Ions of this selected mass are accelerated in phase with the alternating voltage through each of the successive analyzer stages and thus acquire a maximum kinetic energy, whereas particles of different mass, higher or lower, fall out of phase and either do not emerge from the analyzer section or, at best, emerge with energies less than that of the preferred particle.

The ion collector 10 of the invention comprises two parallel electrodes or plates 26 and 27 inclined at an angle of substantially 45° to the entrant ion beam 18. The base plate 26 is at a low potential, for example, ground, and may be connected to the last analyzer plate 25*n* as shown. The plate 27 is at a relatively high potential, e.g., about one-half the ion beam energy in electron volts, or about 500 volts. Thus, there is provided between the plates 26 and 27 a substantially uniform electrostatic field having a plane of minimum field potential at the base plate 26. The ions from the analyzer section enter the space between the plates 26 and 27 through a relatively narrow entrance slit 28 in the plate 26, each ion traversing a parabolic trajectory and returning to the base plate 26, where ions of the predetermined charge-to-mass ratio pass through an ion-selective exit slit 29 in the plate 26 positioned at the point of maximum range. Particles traversing the slit 29 impinge on a collecting electrode or electrode means 30 at or near ground potential. Ions of lower or higher mass, and therefore of lower energy, fall short of the slit 29 and impinge on an ion intercepting means formed by the base plate 26 adjacent an edge 29*a* of the slit 29. The numerals 32 and 33 represent, respectively, the paths of preferred and nonpreferred ions. An indicating and/or recording means 31, connected to the electrode 30, indicates and/or records the relative abundance of the selected ions.

The ion collector 10 operates on the principle that, for ions of any given energy, those which enter at 45° attain the maximum range. Particles of the same given energy, but of an initial entrant angle either larger or smaller than 45°, return to the base plate **26** within a shorter distance. However, the variation in range, for varying entrant angle, is extremely small compared with the corresponding variation in the maximum attained altitude. Accordingly, by collecting at the position of maximum range, the overlapping of the paths of ions of adjacent masses due to deviations in the entrant angle, when more than one ion species enters the collector, is greatly reduced and resolution and useful ion current are corrpesondingly increased.

The foregoing principle is illustrated in Fig. 2. The origin, lower left, corresponds to a point on the entrance slit **28**, and the base line $y=0$ corresponds to the surface of the base plate **26**. The curve a is the trajectory of a particle entering at $\alpha=45°$, and the curves b and c are the trajectories for particles of the same energy deviating in either direction by 0.1 radian, or about 5.7°. This is a relatively large angular deviation and is shown for purposes of illustration only. Taking the peak trajectory value of y for $\alpha=45°$ as unity, the range, or maximum value of x, is 4. For particles entrant at any angle α,

$$x=v_o t \sin \alpha,$$

and

$$y=\tfrac{1}{2} at^2+v_o t \cos \alpha,$$

where v_o is the entrant velocity, t is time and a is the acceleration, a being equal to $-Ee/m$ where E is the deflecting field strength, and e/m the ionic charge-to-mass ratio.

Eliminating t, expressing a in terms of the deflection field and v_o in terms of the effective total voltage V accelerating the ion before entering the collector **10**, the following equation results:

$$y=\frac{-Ex^2}{4V \sin^2 \alpha}+\frac{x}{\tan \alpha}$$

It is found that y has a maximum value, or altitude, at the peak of the trajectory given by

$$y_{\max}=\frac{V \cos^2 \alpha}{E}$$

and that $x_{\max}$, or maximum range, is

$$x_{\max}=\frac{4V}{E} \sin \alpha \cos \alpha$$

The trajectory for a value of α less than 45° crosses the trajectory for $\alpha=45°$ at a value of x given by

$$x_c=\frac{-4V}{E}\left(\frac{\cos \alpha-\sin \alpha}{2 \sin \alpha-\csc \alpha}\right)$$

The incremental changes of maximum altitude and range with change of initial angle are respectively given by

$$\frac{dy_{\max}}{d\alpha}=\frac{-V}{E} \sin 2\alpha$$

and

$$\frac{dx_{\max}}{d\alpha}=\frac{4V}{E} \cos 2\alpha$$

It is noted that at $\alpha=45°$, $dx_{\max}/d\alpha$ is zero, i.e., x is a maximum, and that for even relatively large deviations from $\alpha=45°$, only very small changes in $x_{\max}$ may be expected. On the other hand, $dy_{\max}/d\alpha$ has a finite value at $\alpha=45°$, and for given deviations of α, the deviaation in $y_{\max}$ may be many time that of $x_{\max}$. This is borne out in Fig. 2, where, for the entrant angle deviation of ±0.1 radian, the total deviation in $y_{\max}$ is 0.396, and the total deviation of $x_{\max}$ is only 0.08, a ratio of approximately five to one. Taking smaller angles of deviation, closer to those encountered in practice, the ratio of the total deviation of $y_{\max}$ to that of $x_{\max}$ is rapidly increased. Thus, at a deviation of ±1°, total deviations in $y_{\max}$ and $x_{\max}$ are 0.068 and 0.0024, respectively, a ratio of approximately 28 to 1.

It will be noted that for any pair of rays deviating in entrant angle symmetrically from 45°, the deviation in $y_{\max}$ is symmetrical, and the range, or $x_{\max}$, values are identical. Thus the curves b and c of Fig. 2 have a common terminus on the x axis.

Other conditions being equal, the collector **10** of the present invention makes possible, in a typical instrument of the type described, an increase in the ion signal current by a factor of five or more, or for the same current, a substantial increase in resolving power. The reason for the current gain is explained by the schematic curves of Figs. 3 and 4. In Fig. 3, the curves m_1 and m_2 show the ion current density distribution, attributable to each of two adjoining masses, along the "vertical" or y direction, in the vicinity of the trajectory peaks, where prior ion collectors have operated. The preferred mass is represented by m_1 and, for simplification, the relative total abundance of the masses is assumed equal. The long "skirts" on these curves are largely the effect of ions entering the deflection space at angles which deviate in either direction from 45°. To obtain a signal for the ion m_1 which is substantially independent of the ion m_2, only that portion of the m_1 current may be collected which falls to the right of the dotted line and is indicated by the shaded area. Thus, because the right-hand skirt of m_2 overlaps a major portion of the range of m_1, only a small portion of m_1 may be collected to give a useful signal.

Fig. 4 shows the improvement achieved in the present invention by collecting in the region of maximum range. Since the distribution for each ion species tails off primarily in one direction only, i.e., to the left as shown in this figure, a much larger portion of the available current of the preferred ion, again shown by the shaded area, may be collected without interference by the adjoining ion. The leading (right-hand) edges of the curves of Fig. 4 are not perfectly vertical because particles of a given mass entering the collector **10** at 45° vary slightly in energy, and because of the particular shape of the energy distribution curve for such particles.

The effect of the invention is that aberration due to angular deviation is no longer a dominant factor in restricting attainable resolution or ion current. Instead, this effect is made at least comparable to, and usually smaller than, other effects which limit performance, such as energy separation of adjacent mass peaks, energy spread for a given mass, collector slit-width effects, etc.

Referring again to Fig. 1, it will be seen that additional advantages result from the fact that the collecting or sensing electrode **30** is now adjacent the deflection plate, **26**, which is at the lower potential, i.e., at or near ground potential, rather than near the high voltage plate **27**. For example, transients formerly appearing when the deflection voltage was changed, due to capacitive coupling between the deflector plate and the sensing electrode, are substantially eliminated. The fractionating pattern becomes substantially independent of the deflection voltage, because of the relative insensitivity to variations in the entrant angle. Moreover, the electrostatic deflection field in the vicinity of the sensing electrode **30** is now substantially free of distortion, and this further improves resolution. Additionally, the problem of electrical leakage between the relatively high-potential deflection electrode and the low-potential sensing electrode is substantially eliminated, since the sensing electrode now adjoins the deflection plate of lower potential. Finally, the lower potential deflection plate **26** may, if desired, be made about 20 volts negative with respect to the collector electrode **30**. This permits the plate **26** to act as a suppressor for secondary electrons originating at the collector electrode **30** by returning them to this electrode.

Because the present invention increases the ion current available at the sensing electrode, it is possible, if desired, to reduce the width of the entrance slit **28** appreciably, and thereby reduce the lateral spread in initial ion position. A suitable entrance slit width may be 0.015 inch. The collector or exit slit **29** is made somewhat larger, e.g., 0.040 inch, so that the beam of preferred ions may still impinge on the collector electrode **30** despite any residual angular deviation effect and the small energy spread for preferred ions entering at 45°. The resolution, in general, of the collector **10** is a function of the ratio of the entrance slit width to the distance between the slits **28** and **29**.

It will be apparent that the significant functional portion of the exit slit **29** is the inner edge **29*a*** thereof. For example, the base plate **26** could terminate at **29*a***, and the outer edge, **29*b***, of the slit could be omitted. In general, however, it is desirable to use an actual slit and thereby maintain a more nearly undisturbed deflection field in the exit slit region.

It will be apparent also that in a mass spectrometer where the analyzer section decelerates the selected ions instead of accelerating them, the collector of the invention may also be useful in certain cases. The exit slit or exit slit edge of the collector in this case is positioned to select ions of minimum, rather than maximum, energy.

Although I have disclosed an exemplary embodiment of my invention herein for purposes of illustration, it will be understood that various changes, modifications and substitutions may be incorporated in such embodiment without departing from the spirit of the invention as defined by the claims hereinafter appearing.

I claim as my invention:

1. In an ion collector for an ion sorting device comprising an ion source and analyzer means for imparting maximum energy to selected ions of a particular charge-to-mass ratio emitted from said source relative to non-selected ions emitted therefrom; the combination of: means providing a substantially uniform electrostatic deflecting field; means for admitting a beam of ions from said analyzer means to said field at a mean inclination of substantially 45° to the equipotential planes of said field and at an initial position in a plane of low field potential, the ions of said beam executing varying trajectories in said field and returning to terminal positions in a plane of substantially the same low field potential, said selected ions attaining a particular, maximum range greater than the range executed by said non-selected ions; means adjacent the terminal position of said maximum range for intercepting ions executing ranges shorter than said maximum by more than a predetermined small amount; and collecting electrode means adjacent said intercepting means for collecting substantially only said selected ions.

2. An ion collector as defined in claim 1 wherein said means providing said deflecting field comprises a pair of spaced, parallel electrodes and means for maintaining said electrodes at different electrode potentials, respectively, said ion admitting means including an ion entrance aperture in that electrode of said electrode pair which is at the lower of said electrode potentials.

3. An ion collector as defined in claim 2 wherein said ion intercepting means comprises an ion intercepting edge of said lower-potential electrode.

4. An ion collector as defined in claim 3 wherein said ion intercepting edge comprises an edge of an ion exit aperture located in said lower-potential electrode and spaced from said ion entrance aperture.

5. An ion collector as defined in claim 4 wherein said collecting electrode means is located external to the space defined by said electrodes and adjacent to said ion exit aperture.

References Cited in the file of this patent

UNITED STATES PATENTS

2,633,539	Altar	Mar. 31, 1953
2,660,677	Nier	Nov. 24, 1953
2,769,093	Hare et al.	Oct. 30, 1956

OTHER REFERENCES

Furry et al.: Physics, published 1952 by Blakiston Co., pp. 12–14 relied on.

(https://ionpowergroup.com/)

Ion Power Group

(https://ionpowergroup.com/) The Sun's Electricity Harvested Day & Night

00:00
01:24
Sun powered Global Electric Circuit
* How Ion Power Works *
Click to Play
www.ionpowergroup.com
00:07 / 01:23

The Sun provides the Earth’s atmosphere with 122 quadrillion watts of radiation daily (122,000,000,000,000,000 watts) sustaining a permanently charged electric field that propagates throughout the atmosphere known as the Earth’s Global Electric Circuit.

The Global Electric Circuit surrounds the

TOP

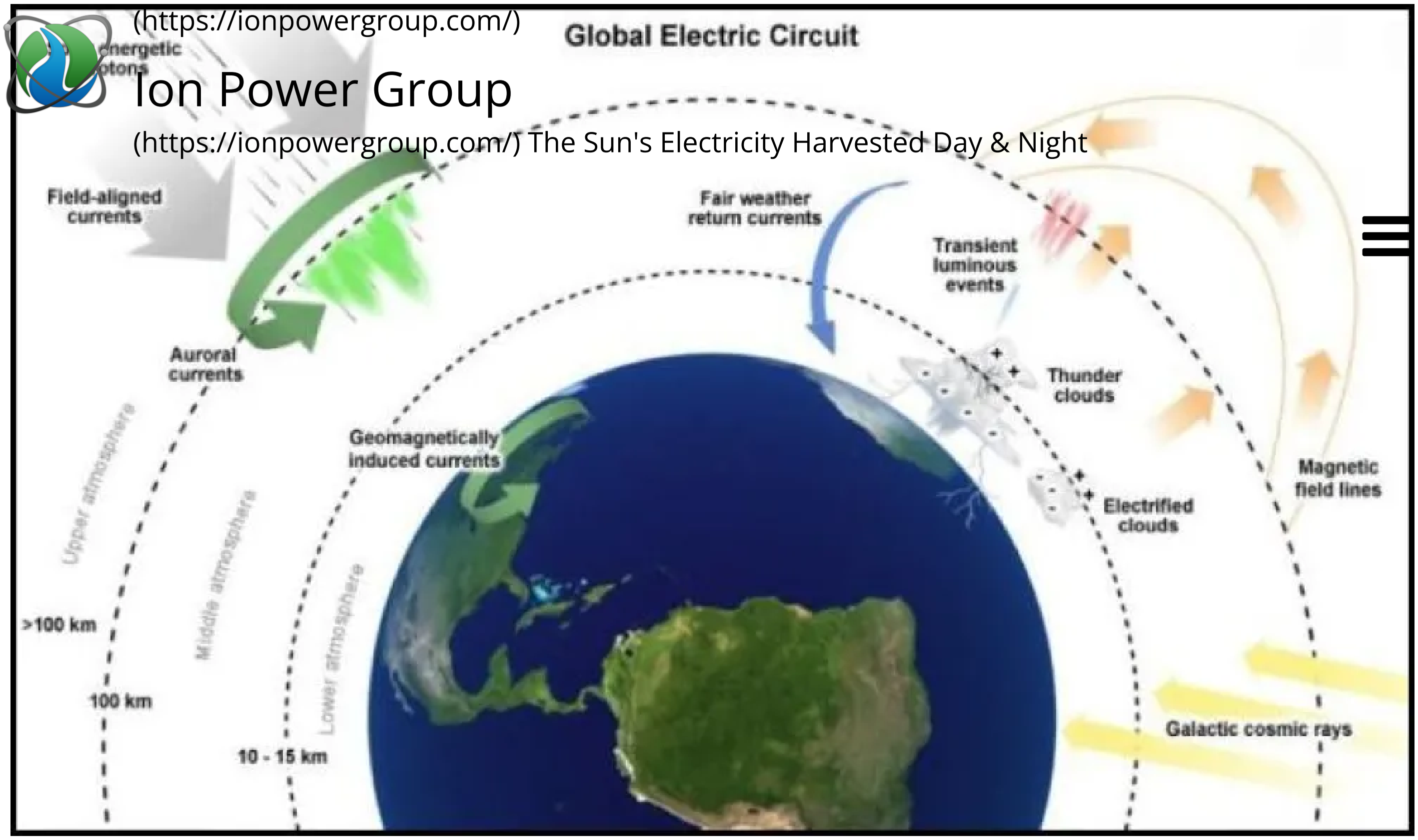
Global Electric Circuit
nergetic
otons
Field-aligned currents
Fair weather return currents
Transient luminous events
Auroral currents
Thunder clouds
Geomagnetically induced currents
Electrified clouds
Magnetic field lines
Upper atmosphere
Middle atmosphere
Lower atmosphere
>100 km
100 km
10 - 15 km
Galactic cosmic rays

Earth in a blanket of high voltage electricity.

TOP

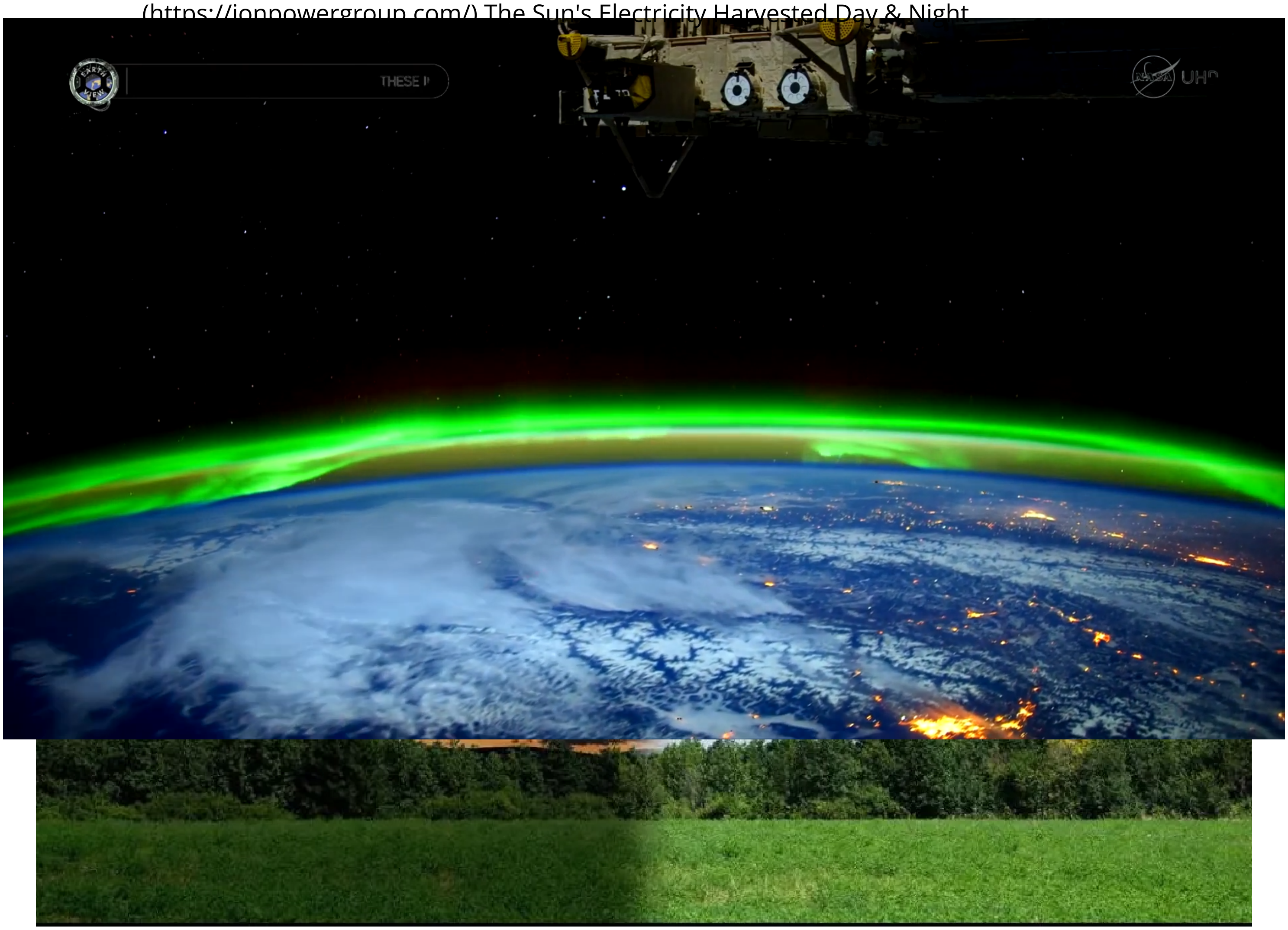

Trillions of naturally renewable electrically charged ions are continuously produced in the atmosphere, around the clock, day & night, powered by the Sun’s enormous energy.

A revolutionary breakthrough in nanomaterial research demonstrates that

TOP

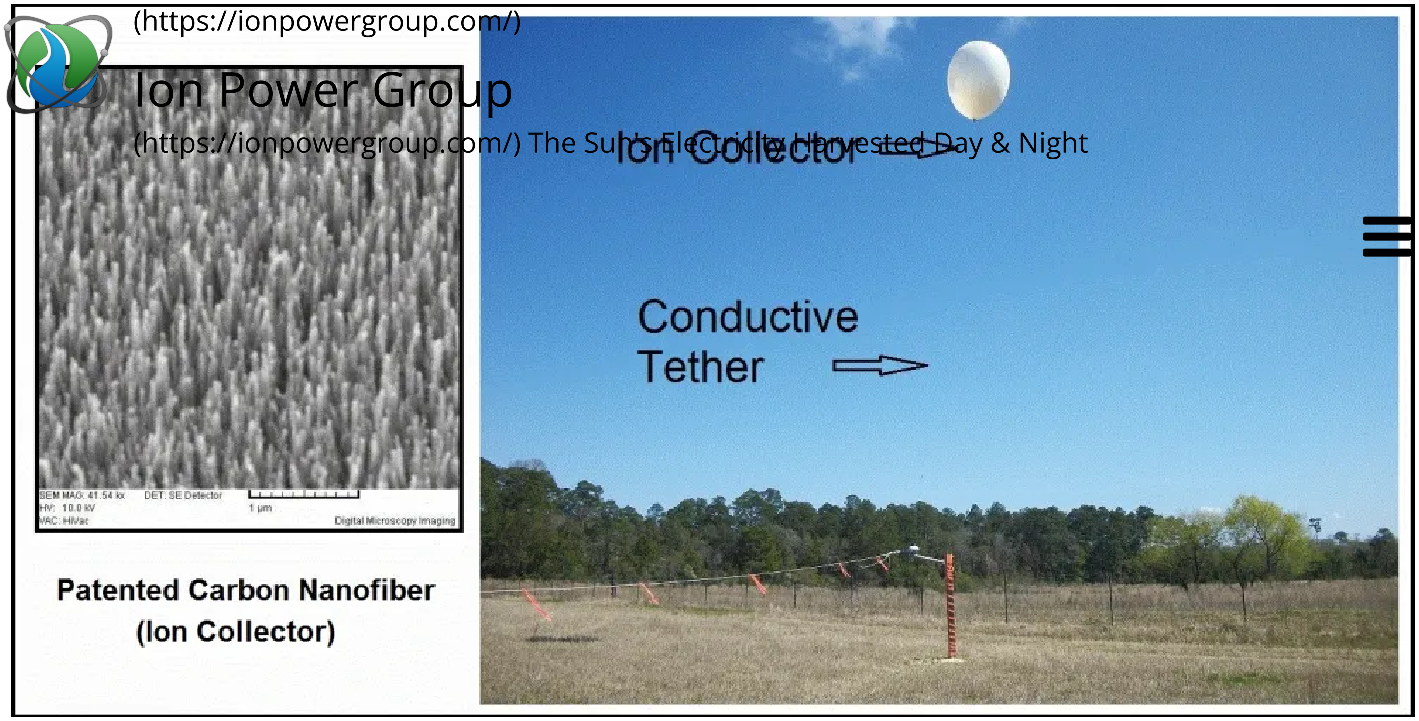
SEM MAG: 41.54 kx
DET: SE Detector
HV: 10.0 kV
VAC: HiVac
1 µm
Digital Microscopy Imaging
Patented Carbon Nanofiber
(Ion Collector)
Ion Collector
Conductive
Tether

exposing a special type of ultra-lightweight carbon nanomaterial to atmospheric ions harvests high-voltage electricity, day, night, and during storms.

“Physics-based modeling shows that an (https://secureservercdn.net/198.71.233.128/9a0.99e.myftpuploa content/uploads/2017/11/Economic-model-study-Rev-B-authored-by-Dr.-Phil-Metzger-PhD.pdf)electron cloud forms around the ion collector material, and this electron cloud increases conductivity of the air around the collector. This causes electrical current to flow, driven by the atmospheric voltage, so that power is generated at the ground. The rate of this current is governed by complex physics in the boundary layer of the collector material, doubtless because of its very high surface area. Ionization can occur on this collector at extremely low voltages because of the sharp, needle-like protrusions on its surface. This understanding of the physics suggests that power generation can be scaled as surface area

TOP

Dr. Philip Metzger
Planetary Scientist - Physicist PhD
NASA Kennedy Space Center
Scientist/Engineer of the Year

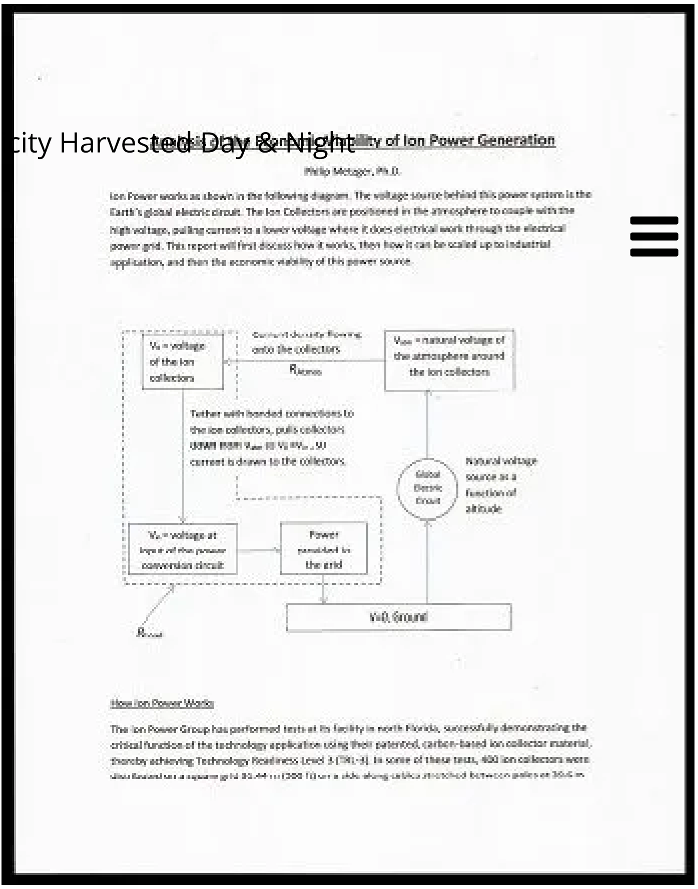

Philip Metzger, Ph.D.

Ion Power works as shown in the following diagram. The voltage source behind this power system is the Earth's global electric circuit. The Ion Collectors are positioned in the atmosphere to couple with the high voltage, pulling current to a lower voltage where it does electrical work through the electrical power grid. This report will first discuss how it works, then how it can be scaled up to industrial application, and then the economic viability of this power source.

Current density flowing onto the collectors
Global Electric Circuit
Natural voltage source as a function of altitude
Power provided to the grid
V=0, Ground

How Ion Power Works

The Ion Power Group has performed tests at its facility in north Florida, successfully demonstrating the critical function of the technology application using their patented, carbon-based ion collector material, thereby achieving Technology Readiness Level 3 (TRL-3). In some of these tests, 400 ion collectors were

Click for Full Report

times voltage. This was incorporated into a simulation of the solar cycle with random storms, and power generation of photovoltaics (PV) and Ion Power were simulated as feeding into a smart grid. Because Ion Power collects the most energy at complementary times compared to PV, the energy storage requirements on the smart grid is greatly reduced. This causes the cost of electricity also to be greatly reduced compared to photovoltaics alone, resulting in energy costs for the overall smart grid of about 8.25 cents per kWh, which is less expensive than any other energy source other than some wind farms (which are only slightly less expensive and have limited geographic applicability). This analysis uses extremely conservative assumptions, so actual energy costs may be much less. **The estimated value to the US economy is between $62 billion and $150 billion (annually)** with proportionately large value to the global economy.

TOP

Proof-of-concept tests have successfully demonstrated the powering of different devices by electricity harvested from atmospheric ions, including...

WARNING
RoHS

BATTERIES

TELEVISIONS

COMPUTERS

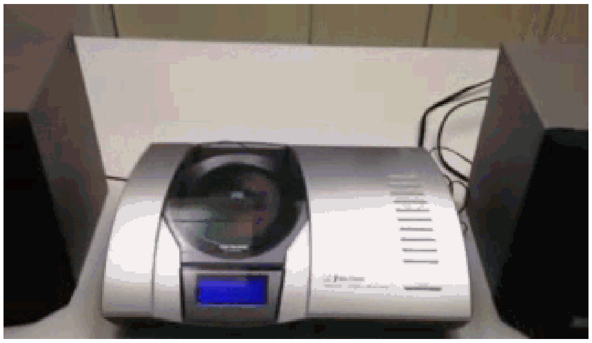

MUSIC SYSTEMS

POWER TOOLS

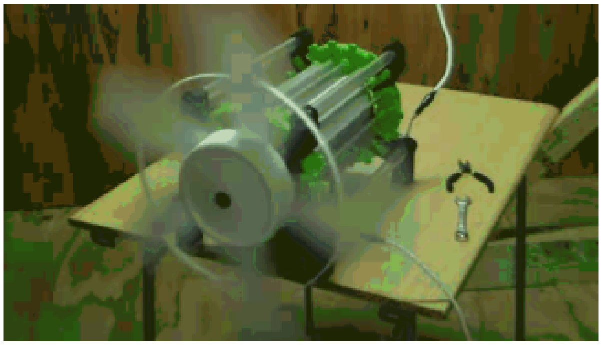

FANS

TOP

Powered by the Sun’s radiation about 2,000 thunderstorms are occurring around the Earth at any given moment. On average, across the world there are 4,320,000 lightning discharges each 24 hour period (3000 per minute). This tremendous amount of high voltage electrical energy, together with GCRs bombarding the atmosphere and Radon gas seeping from the Earth’s crust in certain areas, combine to help replenish the supply of electrically charged atmospheric ions around the world.

TOP

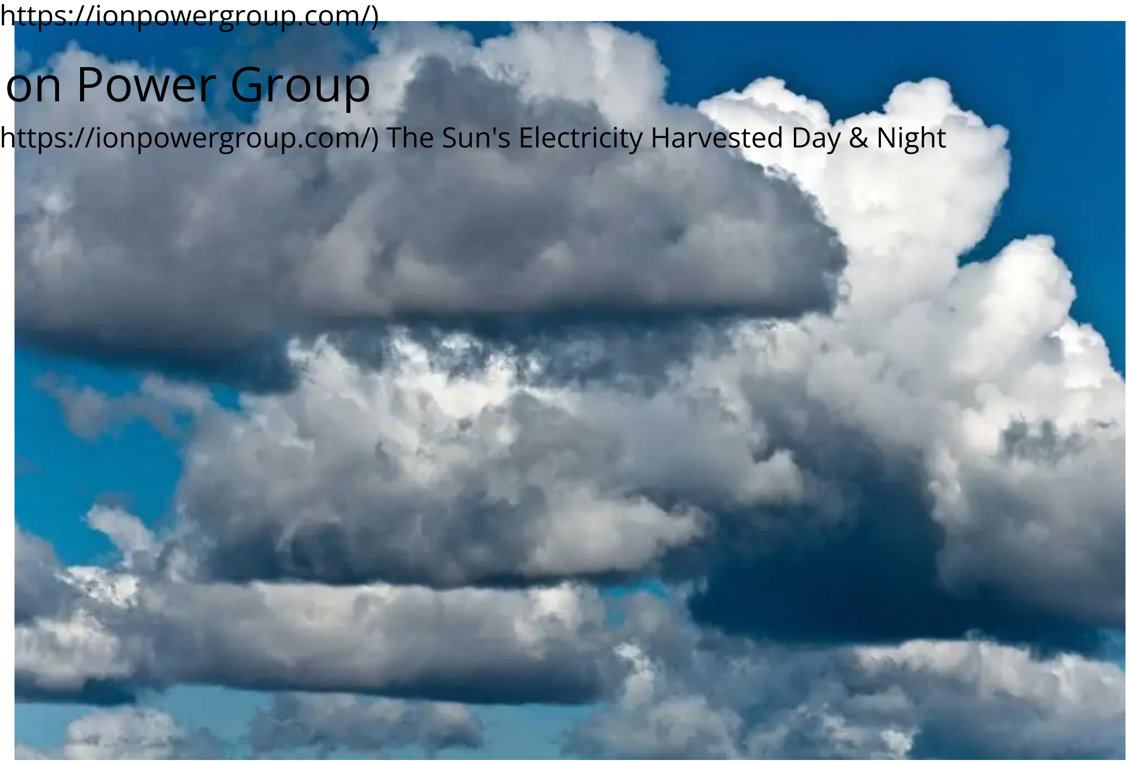

Ion Harvesting Technology harvests electricity day and night and even during bad weather

Ion Power Group's ion collectors, when provided with adequate altitude, can harvest the electric charge of near-Earth ions to produce clean renewable high-voltage electricity day and night and during storms. Ion Harvesting Technology should not be confused with other techniques that attempt to capture lightning bolts or harvest manmade radio frequency (RF) energy.

TOP

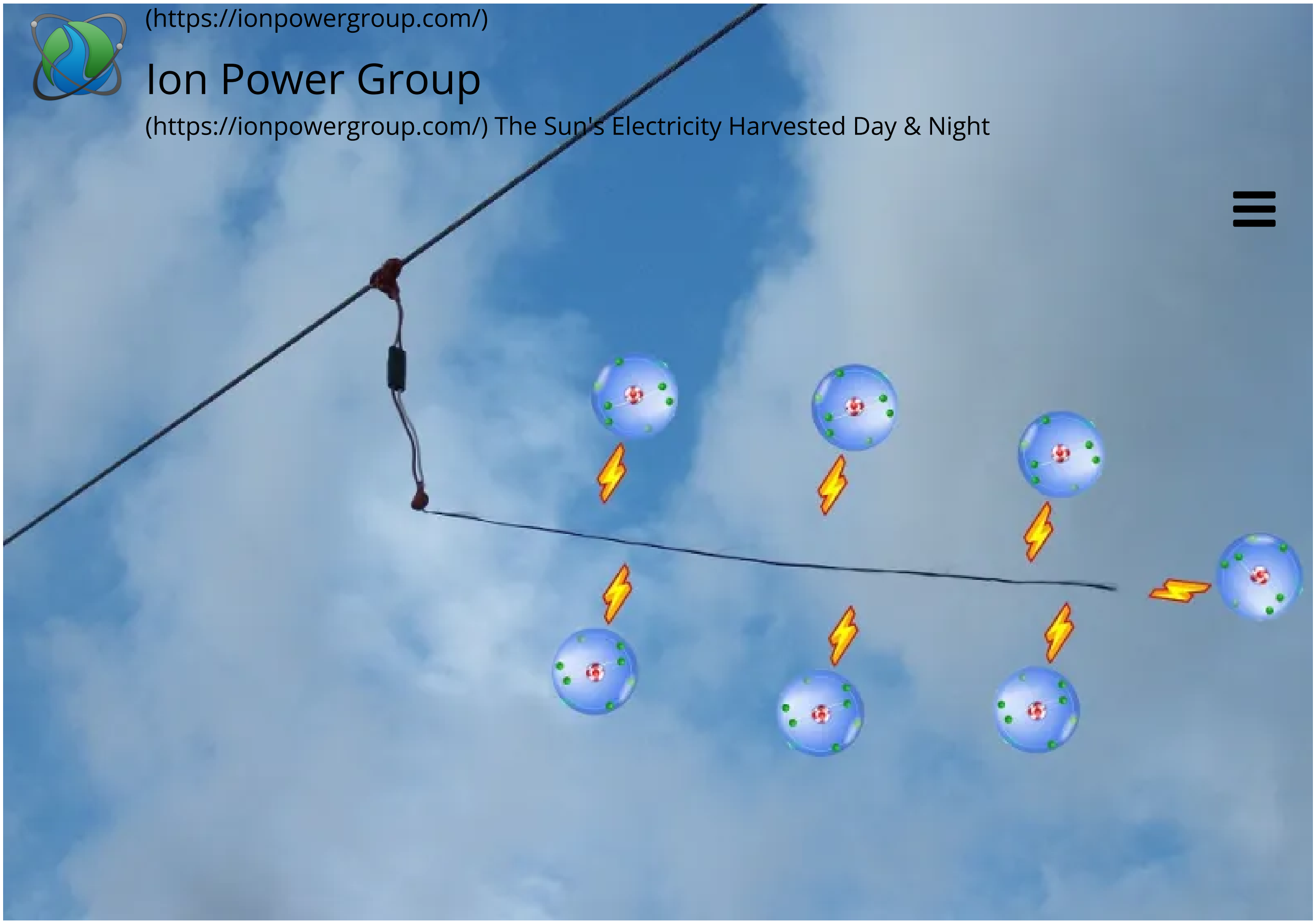

(https://i0.wp.com/ionpowergroup.com/wp-content/uploads/2016/09/Ion-Collector-harvesting-ions-PLAIN.jpg?ssl=1)

TOP

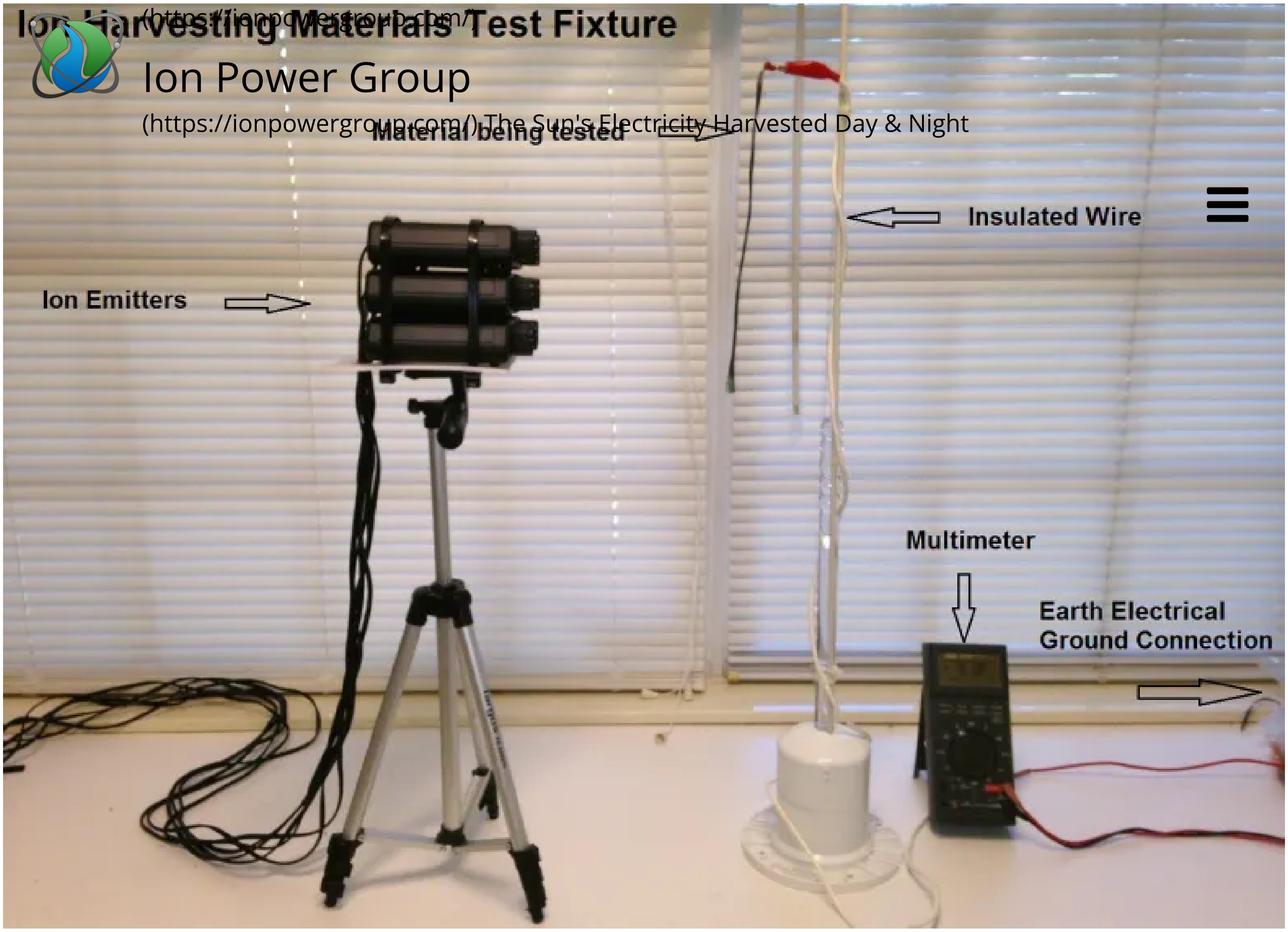
Ion Harvesting Materials Test Fixture
Material being tested
Insulated Wire
Ion Emitters
Multimeter
Earth Electrical Ground Connection

(https://i0.wp.com/ionpowergroup.com/wp-content/uploads/2016/10/Ion-Harvesting-Table-Top-Assembly-annotated.jpg?ssl=1)

Ion collectors can be composed of carbon/graphite/graphene, while not metallic, are electrically conductive, flexible and durable. Electric fields are known to coalesce at conductive points known as the "Corona Effect". The solution of Poisson's equation to determine electric field intensity around the collectors shows us that the highest electric fields will be around sharp points; the sharper the point, the greater the electric field. Ion collectors offer millions of microscopic electrically conductive points

TOP

and protrusions which readily couple to the electric fields of near-Earth ions transferring high-voltage electricity from the atmospheric ions to the ion collectors.

$$\nabla^2\varphi = -\frac{\rho}{\varepsilon}.$$

Ground based circuitry maintains the ion collector material at a voltage considerably lower than the surrounding atmospheric ions thereby employing the electrostatic attraction principle causing nearby ions to migrate to the ion collectors. High voltage wire connected to the ion collectors convey the harvested high voltage DC electricity to a storage medium or electrical load to produce work. Measurements have shown that ion collectors made of graphite are more efficient at harvesting electricity from airborne ions than all other tested metals. In some tests, by as much as 55x greater efficiency. Some of the tested metals include highly conductive silver, copper, and aluminum wire – aluminum spheres – copper and aluminum sheets.

The below graphs show voltage and current coupling characteristics comparing metal in one test (indicated as red and green) to carbon/graphite (indicated as purple).

TOP

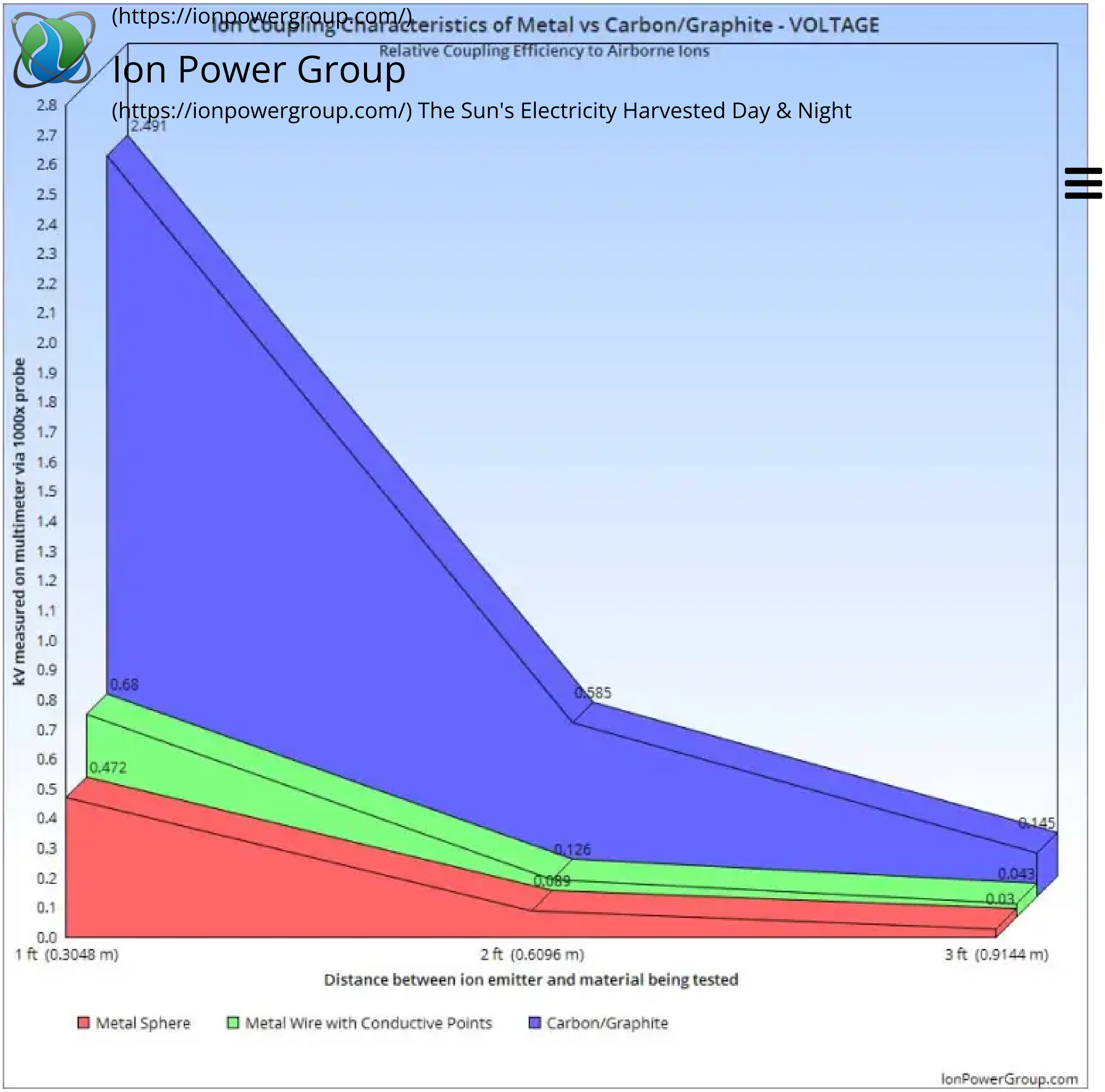

(/wp-content/uploads/2016/07/Ion-Coupling-Characteristics-VOLTAGE-1.jpg)

TOP

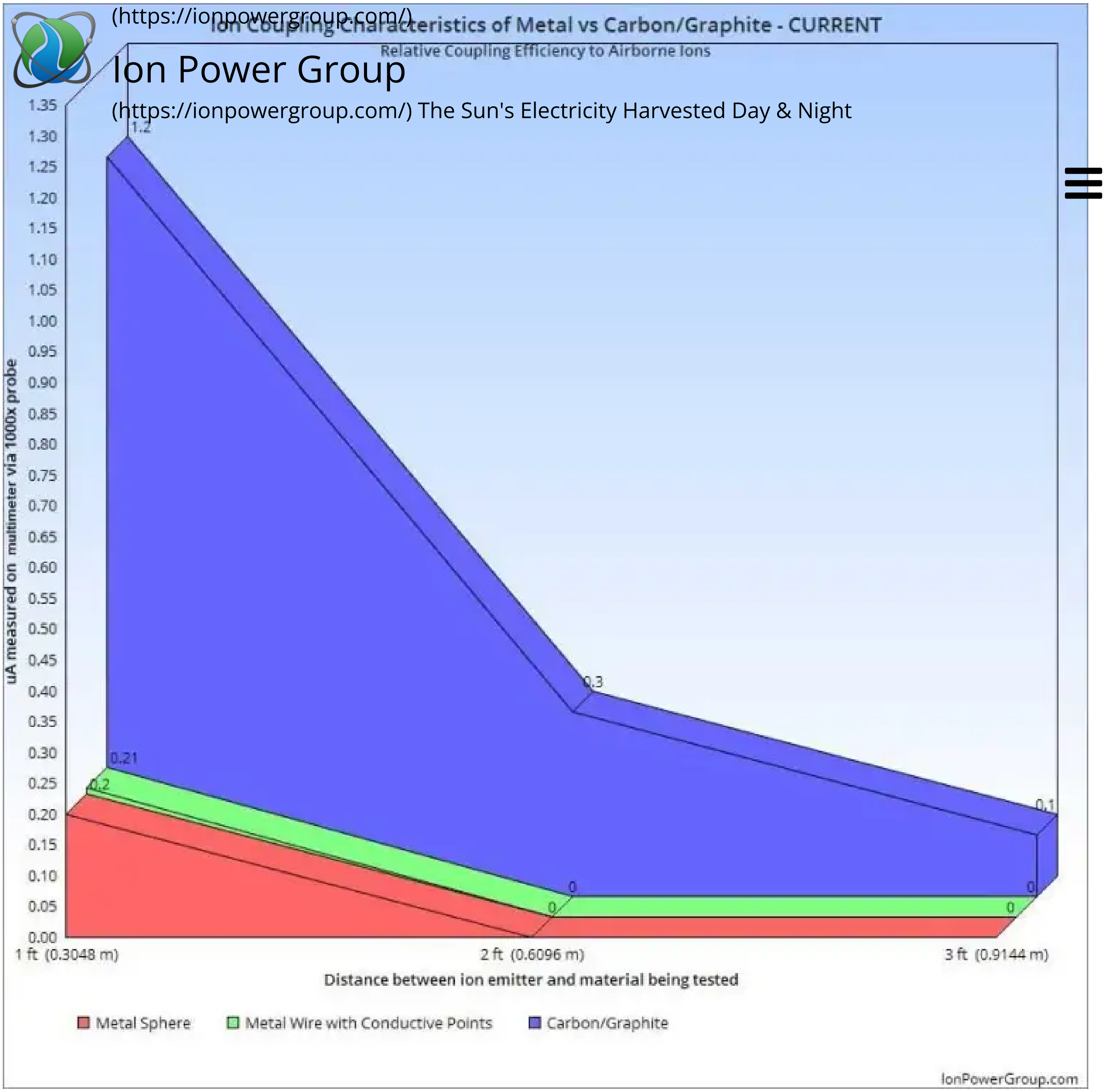

(/wp-content/uploads/2016/07/Ion-Coupling-Characteristics-CURRENT.jpg)

Electricity is produced by providing sufficient altitude to the ultra lightweight ion collector through the use of long-duration aerial

TOP

platforms including tethered balloon, kite or tower. Atmospheric voltage increases at the rate of approximately 90-150vdc per meter of altitude above ground level or sea level and many orders of magnitude greater during disturbed weather. Experiments have

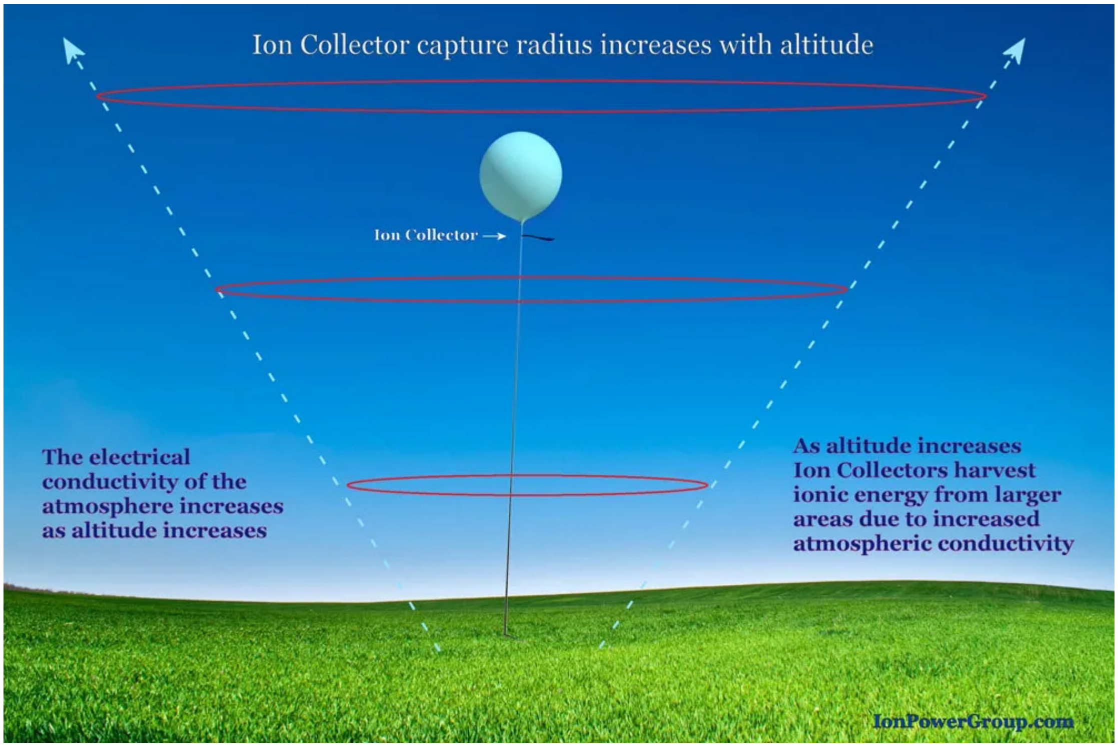
Ion Collector capture radius increases with altitude
Ion Collector →
The electrical conductivity of the atmosphere increases as altitude increases
As altitude increases Ion Collectors harvest ionic energy from larger areas due to increased atmospheric conductivity
IonPowerGroup.com

Atmospheric voltage increases about 90-150vdc per meter of altitude

TOP

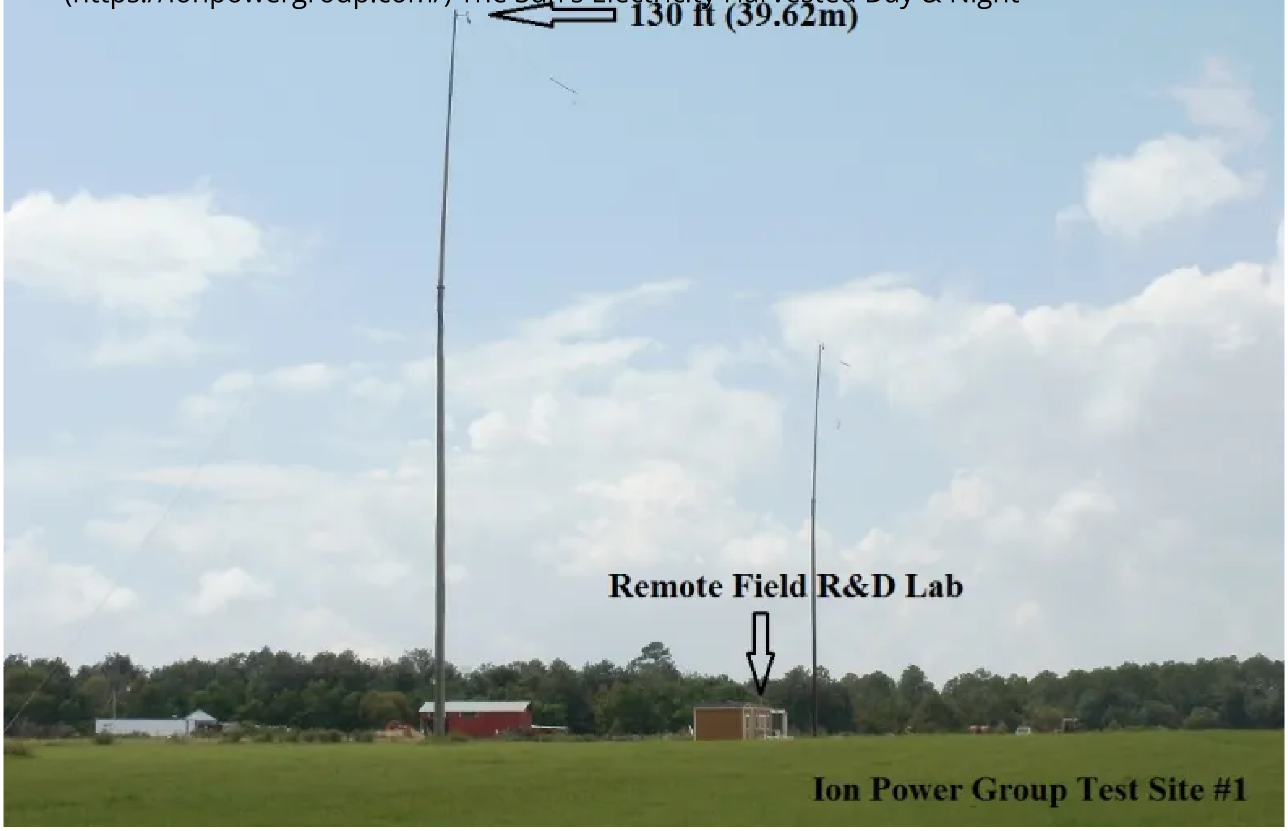
130 ft (39.62m)
Remote Field R&D Lab
Ion Power Group Test Site #1

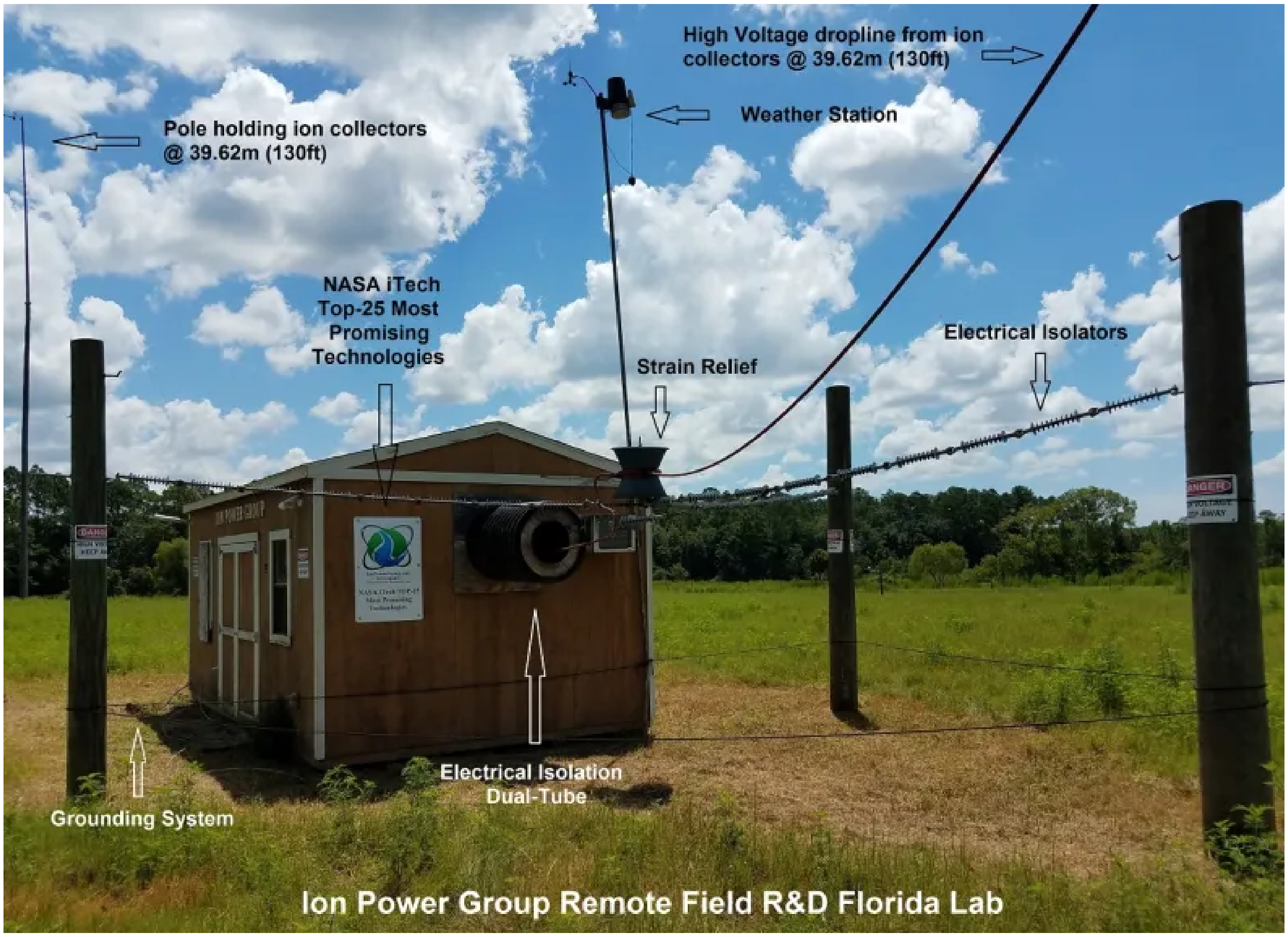
High Voltage dropline from ion collectors @ 39.62m (130ft)
Weather Station
Pole holding ion collectors @ 39.62m (130ft)
NASA iTech Top-25 Most Promising Technologies
Strain Relief
Electrical Isolators
Grounding System
Electrical Isolation Dual-Tube
Ion Power Group Remote Field R&D Florida Lab

TOP

IonPowerGroup.com
Aerospace
ELECTRICITY FROM ATMOSPHERIC IONS
NASA iTech TOP-25
Most Promising
Technologies

(https://i0.wp.com/ionpowergroup.com/wp-content/uploads/2018/08/Ion-Power-Group-Test-Site-08-05-2018-100.jpg?ssl=1)

TOP

TOP

Electricity harvested from Airborne Ions + Water = Hydrogen Gas

TOP

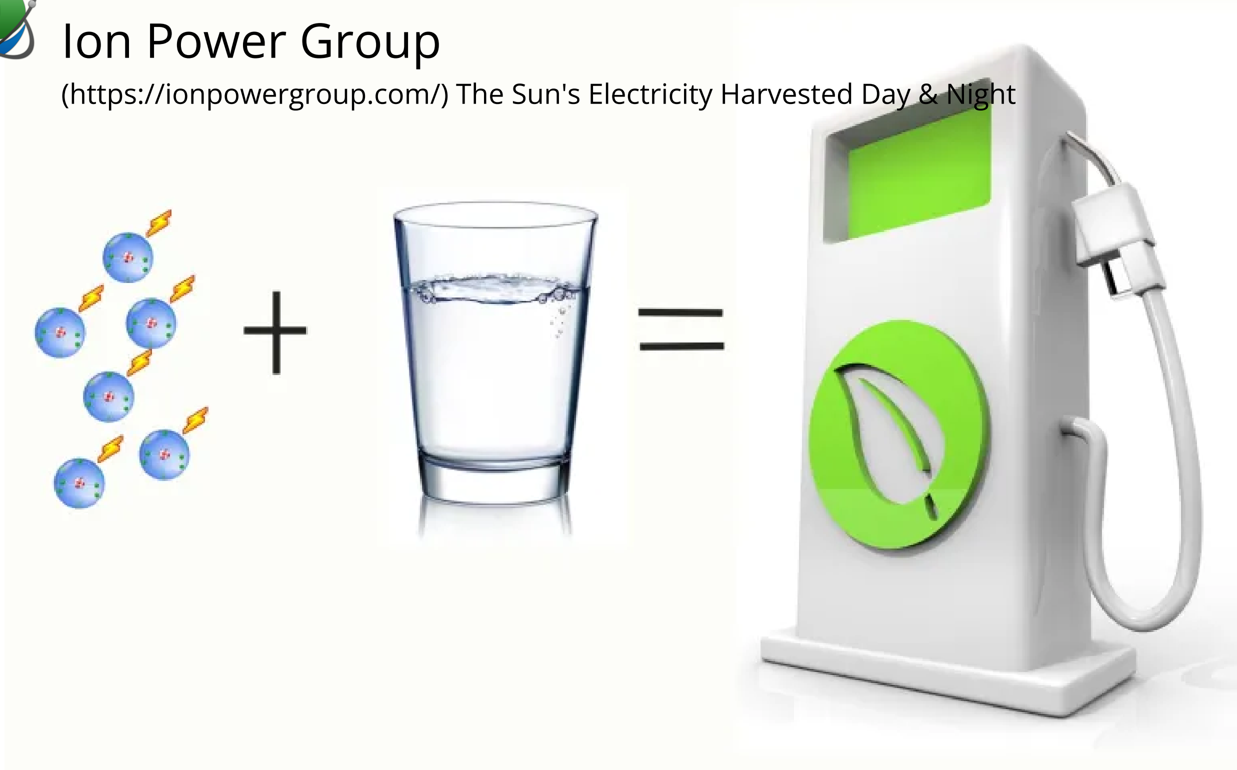

TOP

Ion Power Group has been awarded patents for installing ion collectors on automobiles to generate electricity including through the Triboelectric Effect. Preliminary field tests in which a truck, equipped with Ion Collectors, demonstrated that the concept can generate >500 volts when the vehicle is in motion. Initial tests are encouraging, however, further development is required in order to realize automobiles that generate their own supplemental electricity (or onboard hydrogen gas from water) through use of Ion Collectors.

TOP

Illustrations (below) of proposed experimental ion power plants using long-duration aerostats (aerostats or balloons) or kites or towers to provide altitude to the ion collectors to harvest ions for electricity, day and night. Future Ion Harvesting Technology embodiments that use long-duration aerostats (balloons) to provide great altitude to ion collectors offer a theoretical energy producing duty-cycle approaching 99%.

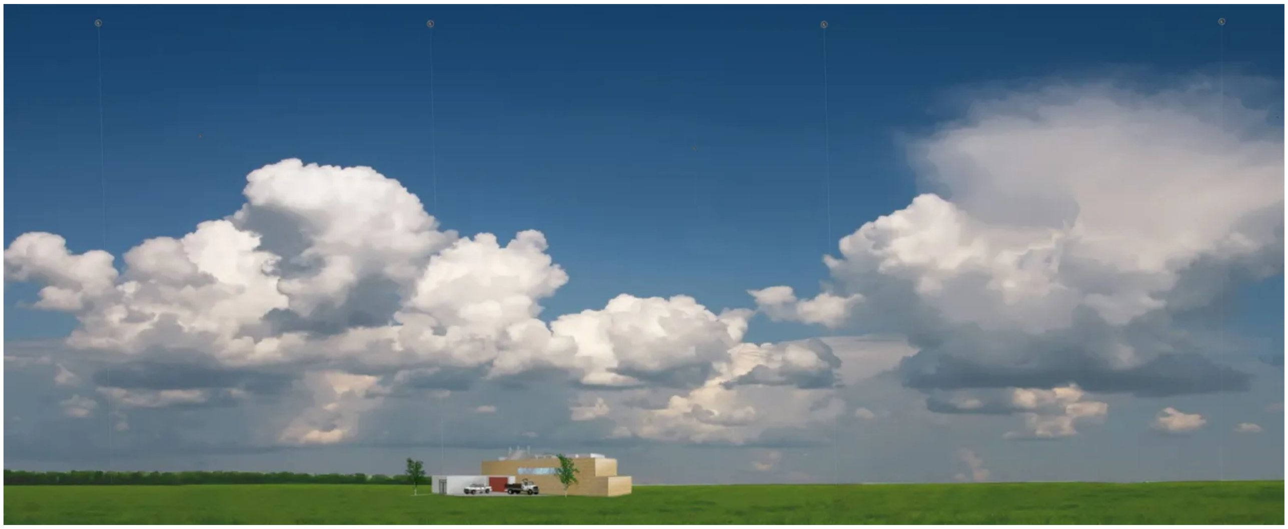

(https://i0.wp.com/ionpowergroup.com/wp-content/uploads/2020/03/Ion-Harvesting-Farm-ORIGINAL1.2-scaled.jpg?ssl=1)

TOP

(https://ionpowergroup.com/)

Ion Power Group

(https://ionpowergroup.com/) The Sun's Electricity Harvested Day & Night

Highly energetic **G**alactic **C**osmic **R**ays (GCRs) arriving from deep space impact Mars' atmosphere on a consistent basis creating electrically charged ions throughout the atmosphere. Carbon/graphite nanomaterials have been shown to be effective at harvesting electricity from atmospheric ions on Earth. Likewise, it is anticipated that carbon nanomaterials elevated into the Martian atmosphere can also harvest electricity from Mars' ion-rich atmosphere day and night and particularly during dust storms.

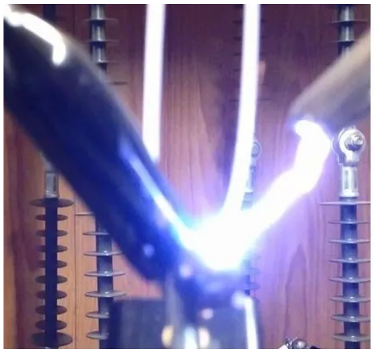

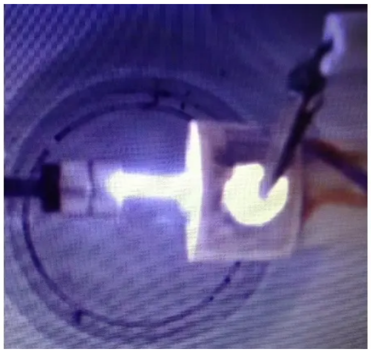

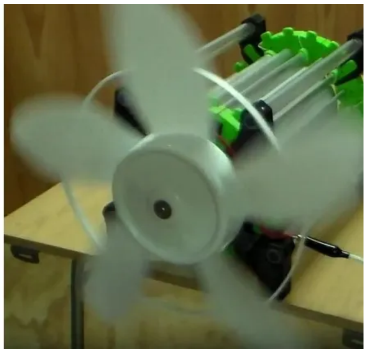

↑ TOP

On Earth, experiments have demonstrated that high voltage electricity harvested from airborne ions can power lights, motors or produce hydrogen & oxygen gas via water electrolysis or place a charge on lithium-iron (LiFePO4) batteries and be converted to 120VAC/60hZ 'house style' current. Mars has an ion-rich atmosphere believed to be even more advantageous for Ion Harvesting Technology. See 'Concept Clips' for videos.

Ion Harvesting Technology appears in FORBES

TOP

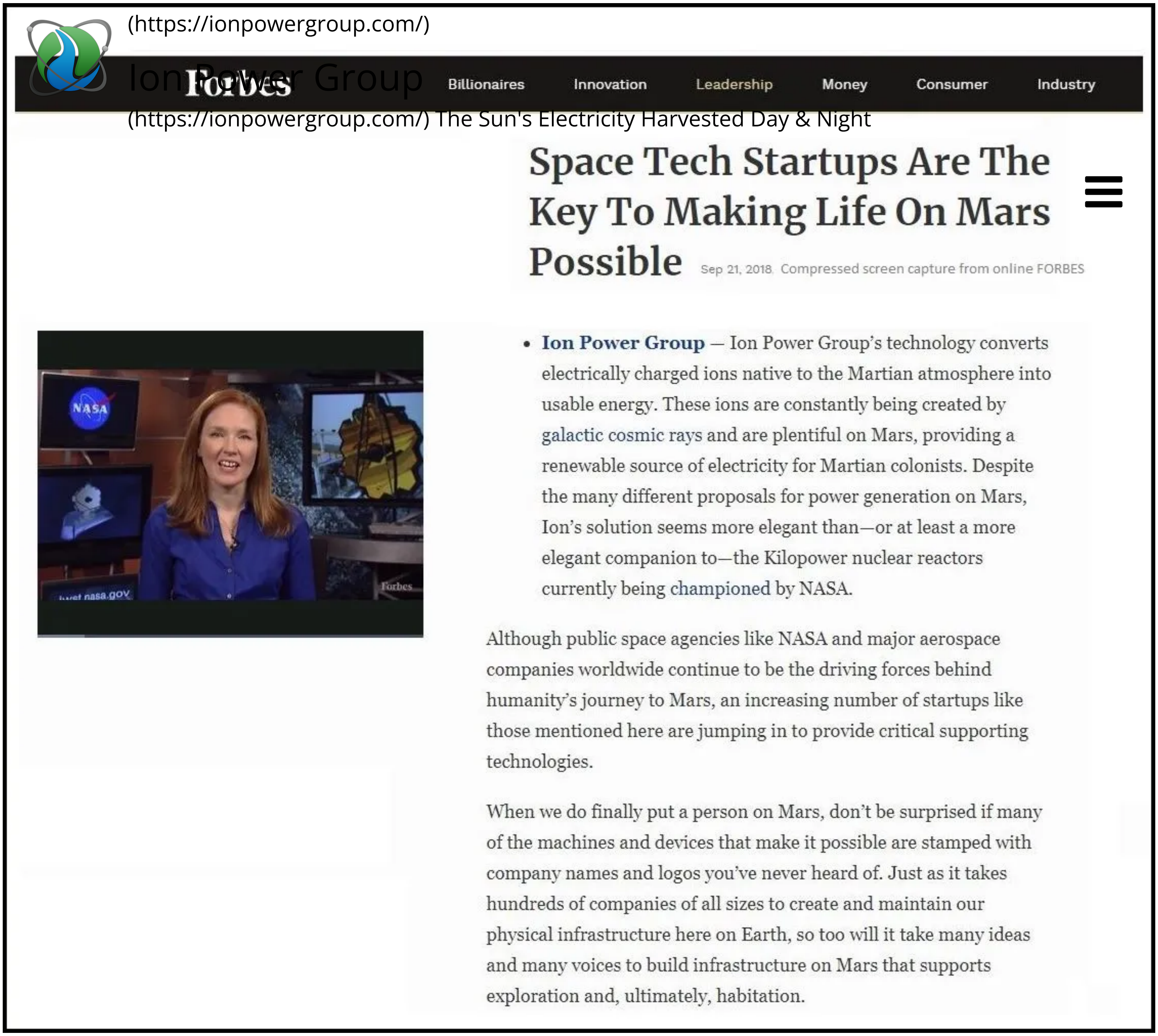

Space Tech Startups Are The Key To Making Life On Mars Possible

Sep 21, 2018. Compressed screen capture from online FORBES

- **Ion Power Group** — Ion Power Group's technology converts electrically charged ions native to the Martian atmosphere into usable energy. These ions are constantly being created by galactic cosmic rays and are plentiful on Mars, providing a renewable source of electricity for Martian colonists. Despite the many different proposals for power generation on Mars, Ion's solution seems more elegant than—or at least a more elegant companion to—the Kilopower nuclear reactors currently being championed by NASA.

Although public space agencies like NASA and major aerospace companies worldwide continue to be the driving forces behind humanity's journey to Mars, an increasing number of startups like those mentioned here are jumping in to provide critical supporting technologies.

When we do finally put a person on Mars, don't be surprised if many of the machines and devices that make it possible are stamped with company names and logos you've never heard of. Just as it takes hundreds of companies of all sizes to create and maintain our physical infrastructure here on Earth, so too will it take many ideas and many voices to build infrastructure on Mars that supports exploration and, ultimately, habitation.

(https://i0.wp.com/ionpowergroup.com/wp-content/uploads/2018/12/FORBES-final-framed.jpg?ssl=1)

GCRs penetrate deeply into Mars' thin atmosphere as depicted below.

TOP

TOP

On February 10th 2016, Senior Scientist of Atmospheric Physics Dr. Andreas J .G. Baumgaertner PhD published a technical report titled ‘Power to Mars’. To read the report, click the below Power To Mars report image or read it on Zenodo by clicking here (https://zenodo.org/record/1407206#.XNUfNRRKiUk). The report recommends that ion harvesting technology be combined with solar panels to provide reliable power day and night for future Mars missions. With a theoretical power generation duty-cycle approaching 99% ion harvesting technology is

TOP

designed to continue to produce electricity during conditions that cause solar panels to fail such as nighttime and dust storms.

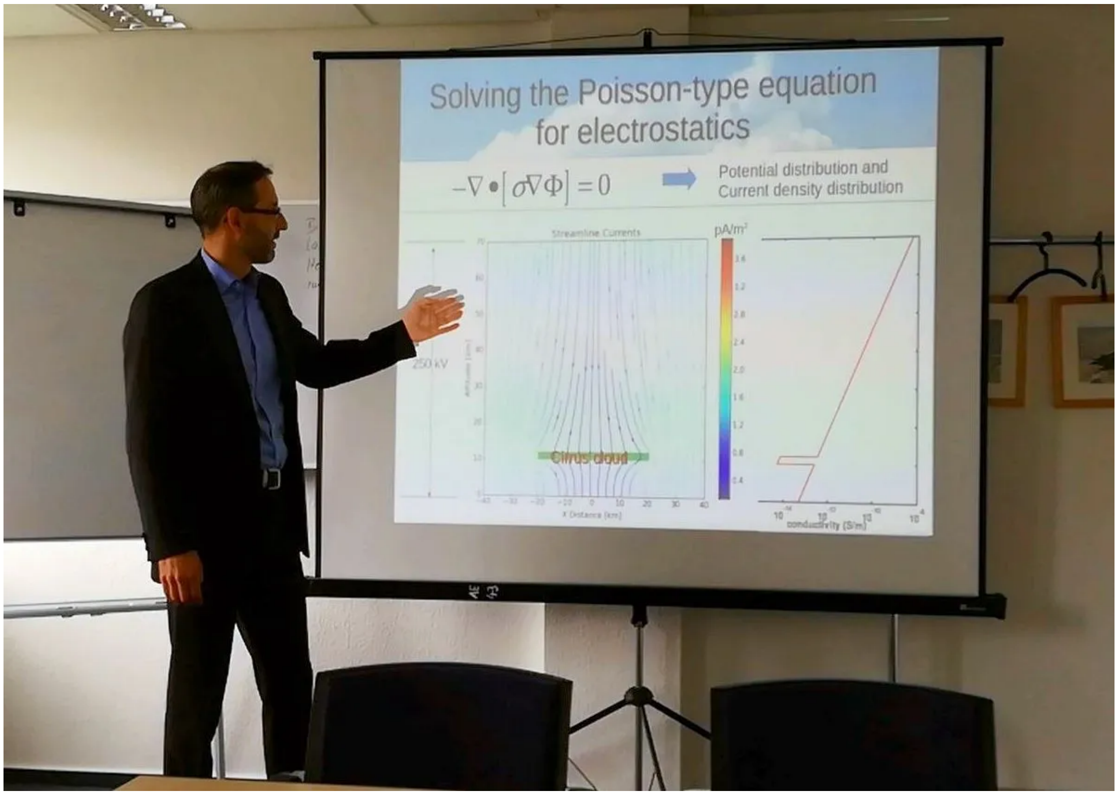
Solving the Poisson-type equation for electrostatics
$-\nabla \bullet [\sigma \nabla \Phi] = 0$
Potential distribution and Current density distribution
Streamline Currents
pA/m²
250 kV
Cirrus cloud
X Distance (km)
conductivity (S/m)

Dr. Andreas J. G. Baumgaertner PhD • Atmospheric Physicist, Senior Scientist – German Aerospace Center • Author of POWER TO MARS

(https://zenodo.org/record/1045600#.Wgm41Bte6Uk)

TOP

(https://ionpowergroup.com/)

Ion Power Group

(https://ionpowergroup.com/) The Sun's Electricity Harvested Day & Night

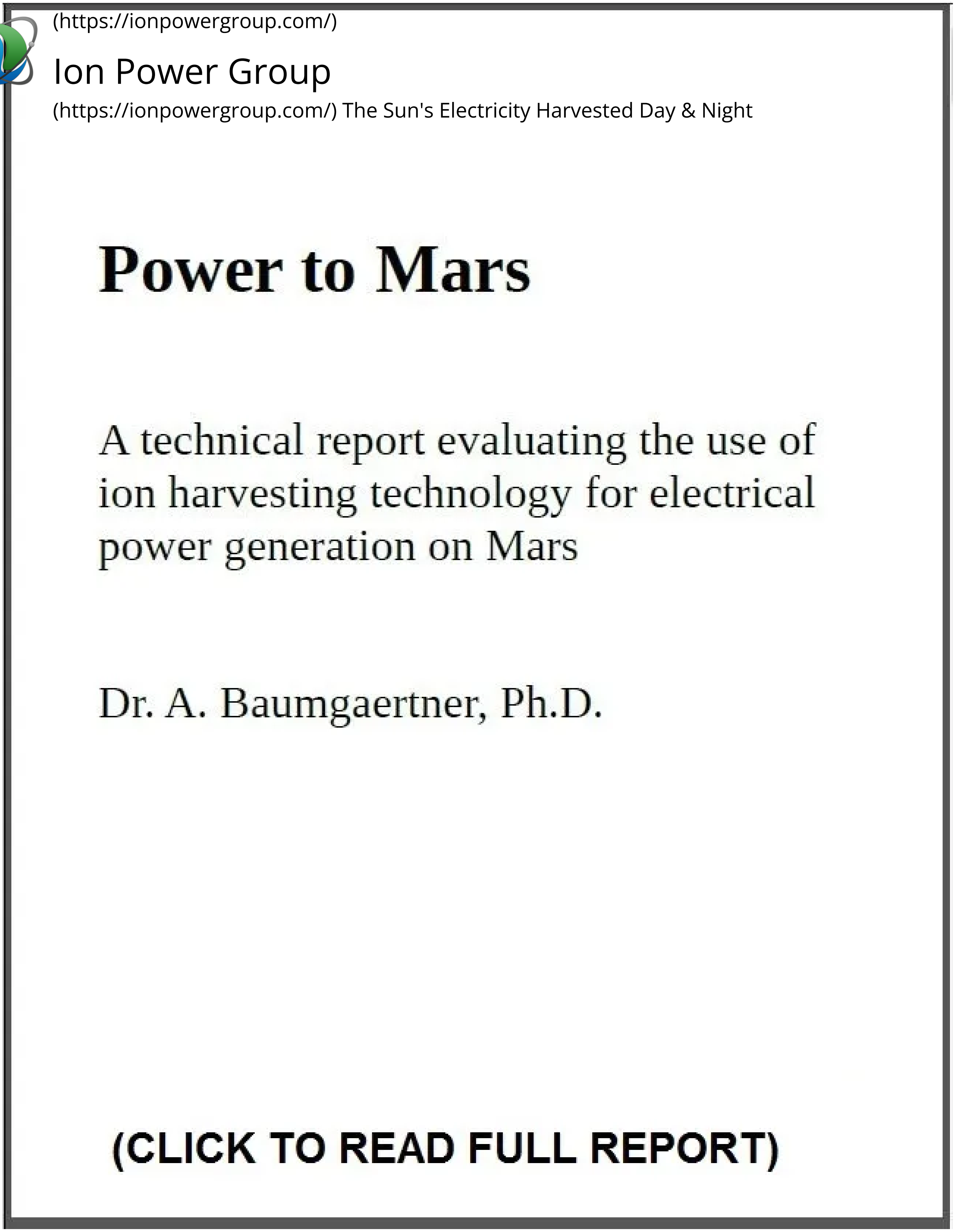

Power to Mars

A technical report evaluating the use of ion harvesting technology for electrical power generation on Mars

Dr. A. Baumgaertner, Ph.D.

(CLICK TO READ FULL REPORT)

(https://zenodo.org/record/1407206#.XQvB6hZKiUl) (https://zenodo.org/record/45877#.WP58tJU2yUk)

↑ TOP

UPDATE

NASA iTech
2018 Cycle III
TOP 25 SEMIFINALIST

Click for more information (https://secureservercdn.net/198.71.233.128/9a0.99e.myftpu content/uploads/2018/08/NASA-iTech-Selects-Top-25-Semifinalists-in-Energy-Competition-_-NASA-1.pdf)

“Space is a more than $350 billion economy, Jonas says, but the industry has room to expand to $1.1 trillion by 2040”. Source: Bloomberg – Morgan Stanley analyst Adam Jonas. (https://www.bloomberg.com/news/articles/201 08-23/space-investors-urged-to-get-their-heads-in-the-clouds)

TOP

GCRs (which are actually microscopic particles traveling at enormous velocity) arriving from deep space impact neutral atoms in the Martian atmosphere knocking electron(s) away producing electrically charged atoms known as ions. Each of the billions of microscopic atmospheric ions acts as a charge carrier conveying a high voltage electric charge. The production of atmospheric ions is a natural process occurring day and night across the entire planet of Mars.

↑ TOP

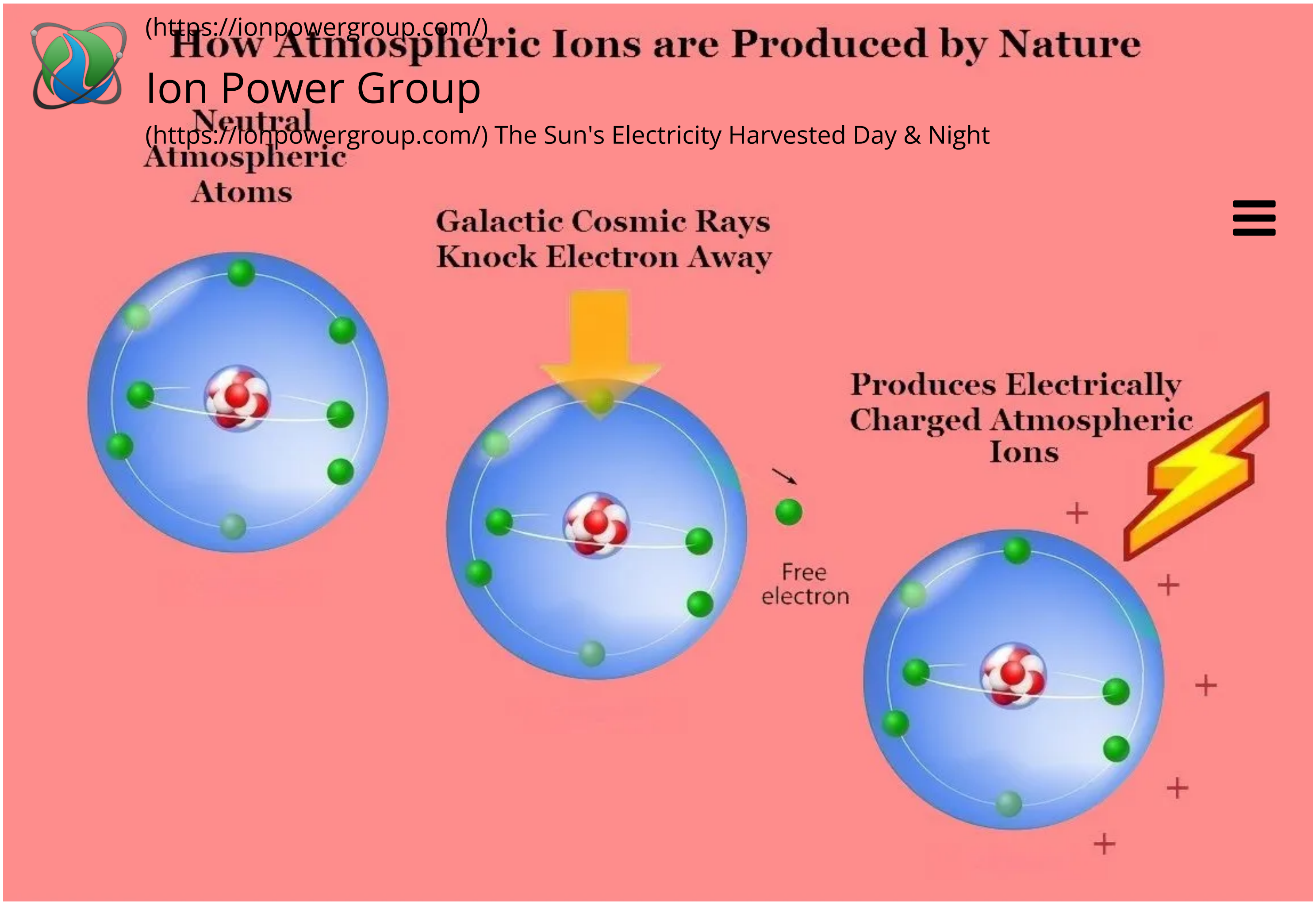
How Atmospheric Ions are Produced by Nature
Neutral Atmospheric Atoms
Galactic Cosmic Rays Knock Electron Away
Produces Electrically Charged Atmospheric Ions
Free electron
+
+
+
+
+

Dust storms occurring on Mars generate strong electrical fields that propagate throughout the Martian atmosphere conveyed by the ions constantly present in the atmosphere. Dust storms can cover large regions of the planet blotting out the sun and can persist for days or weeks. Because the ion-filled Martian atmosphere is electrically conductive, electricity generated by remote dust storms elsewhere on the planet is permitted to propagate throughout the Martian atmosphere. NASA Scientist, Michael Smith, who works at the NASA Goddard Space Flight Center told Space.com "***The dust is electrostatic, like foam peanuts.***" In an interview with NewScientist.com, Brian Jackson with the Physics

TOP

Department at Boise State University says ***“It’s possible that all of the dust grains clattering together in these storms could produce a lot of electricity…”*** interview seen here (https://www.newscientist.com/article/2143217-marss-surface-hosts-millions-of-towering-dust-devils-every-day/).

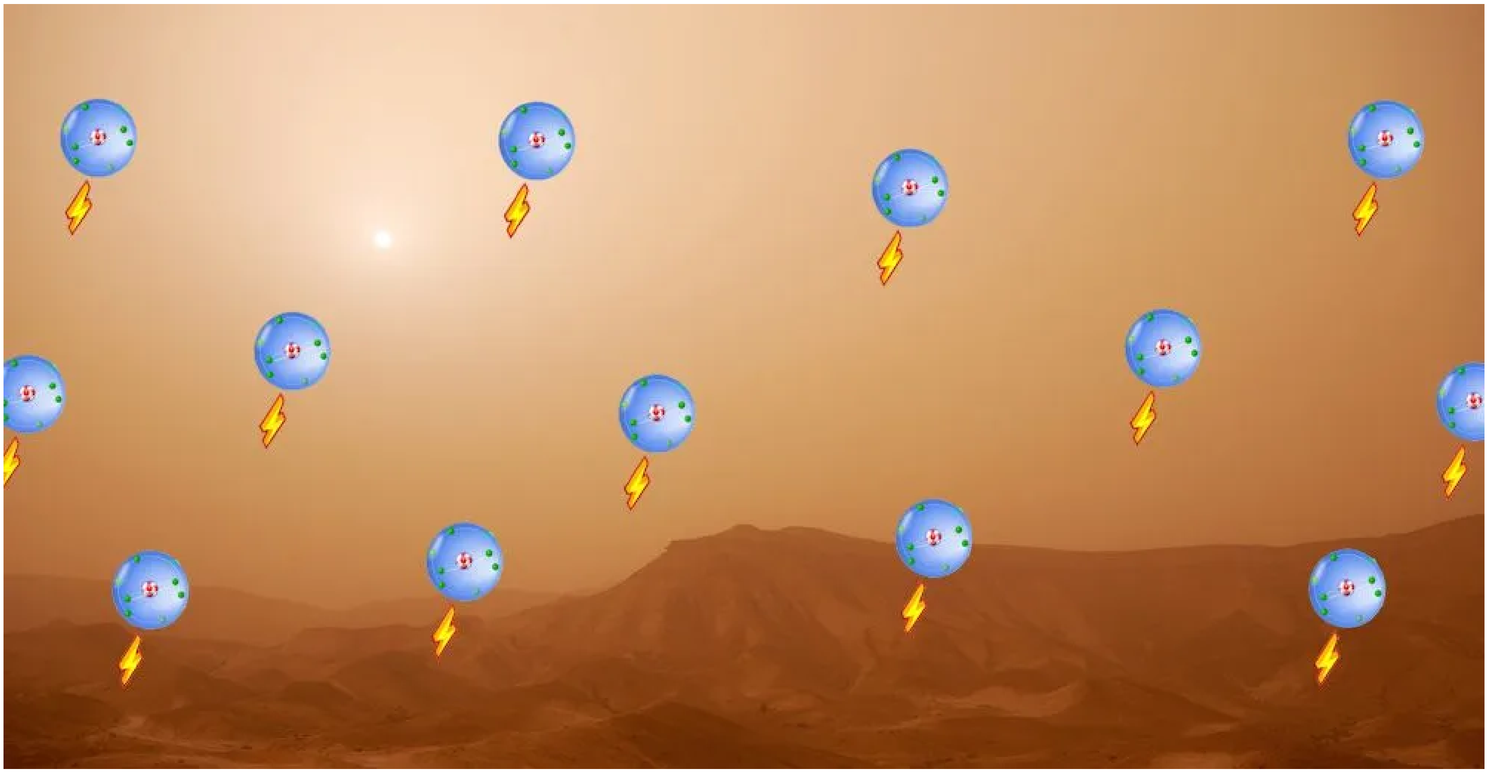

Ion Power Group’s ion collectors, elevated in the Martian atmosphere by tall poles or specially designed aerial platforms (balloons), are designed to harvest the electric charge conveyed by near-Mars ions to produce clean renewable high-voltage electricity day and night and during dust storms as depicted below.

↑ TOP

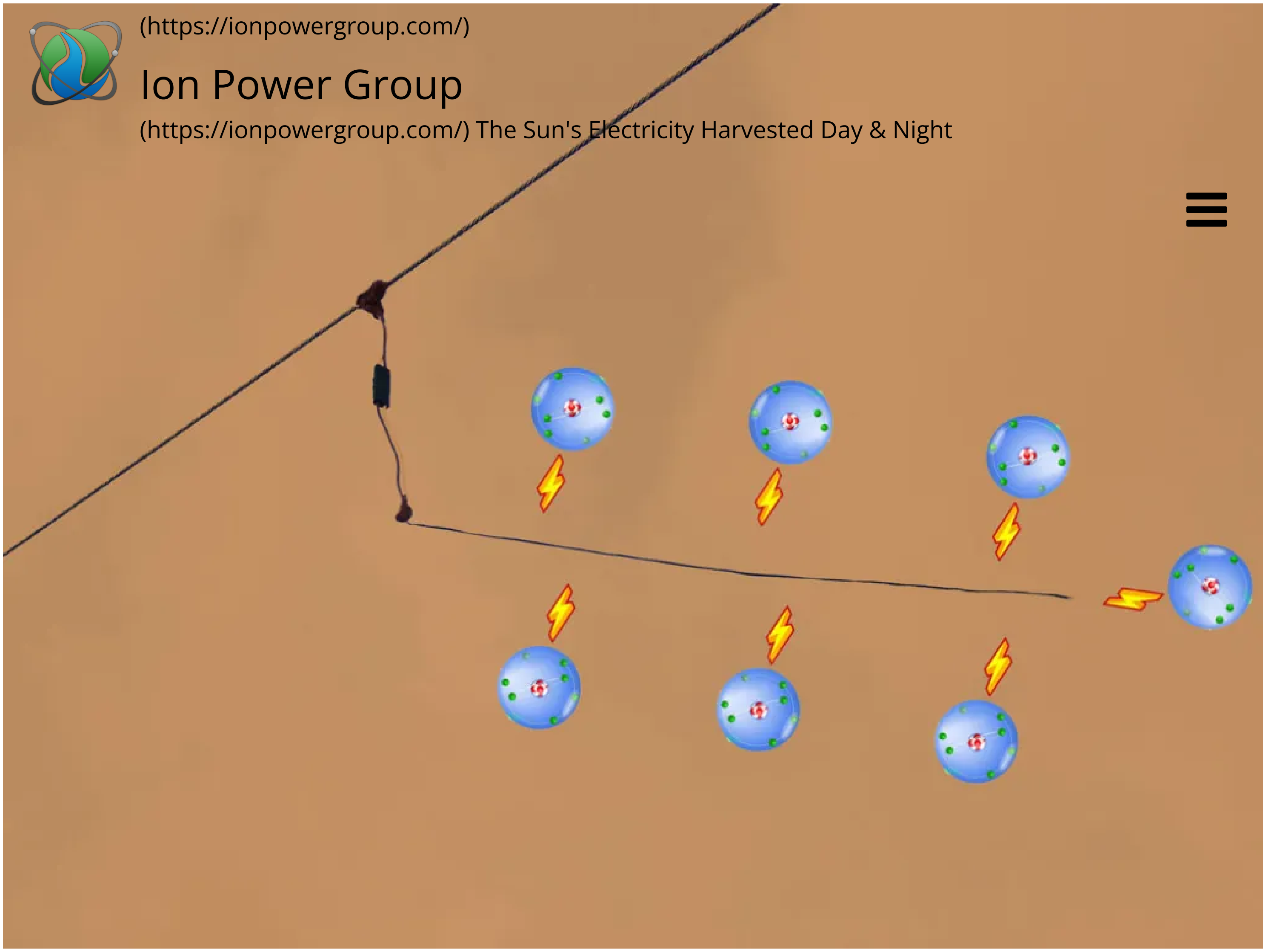

Ion collectors composed of Graphite and/or Graphene, while not metalic, are electrically conductive, flexible and durable. Electric fields are attracted to conductive points, known as the ‘Corona Effect’. Ion collectors offer millions of microscopic electrically conductive points and protrusions which readily couple to the electric fields of nearby ions transferring high voltage electricity from the atmospheric ions to the ion collectors. Measurements have shown that ion collectors made of graphite are more efficient at harvesting electricity from airborne ions than tested metals. In some tests, by as much as

TOP

55x greater efficiency. Some of the tested metals include highly conductive silver, copper, and aluminum wire – aluminum spheres – copper and aluminum sheets.

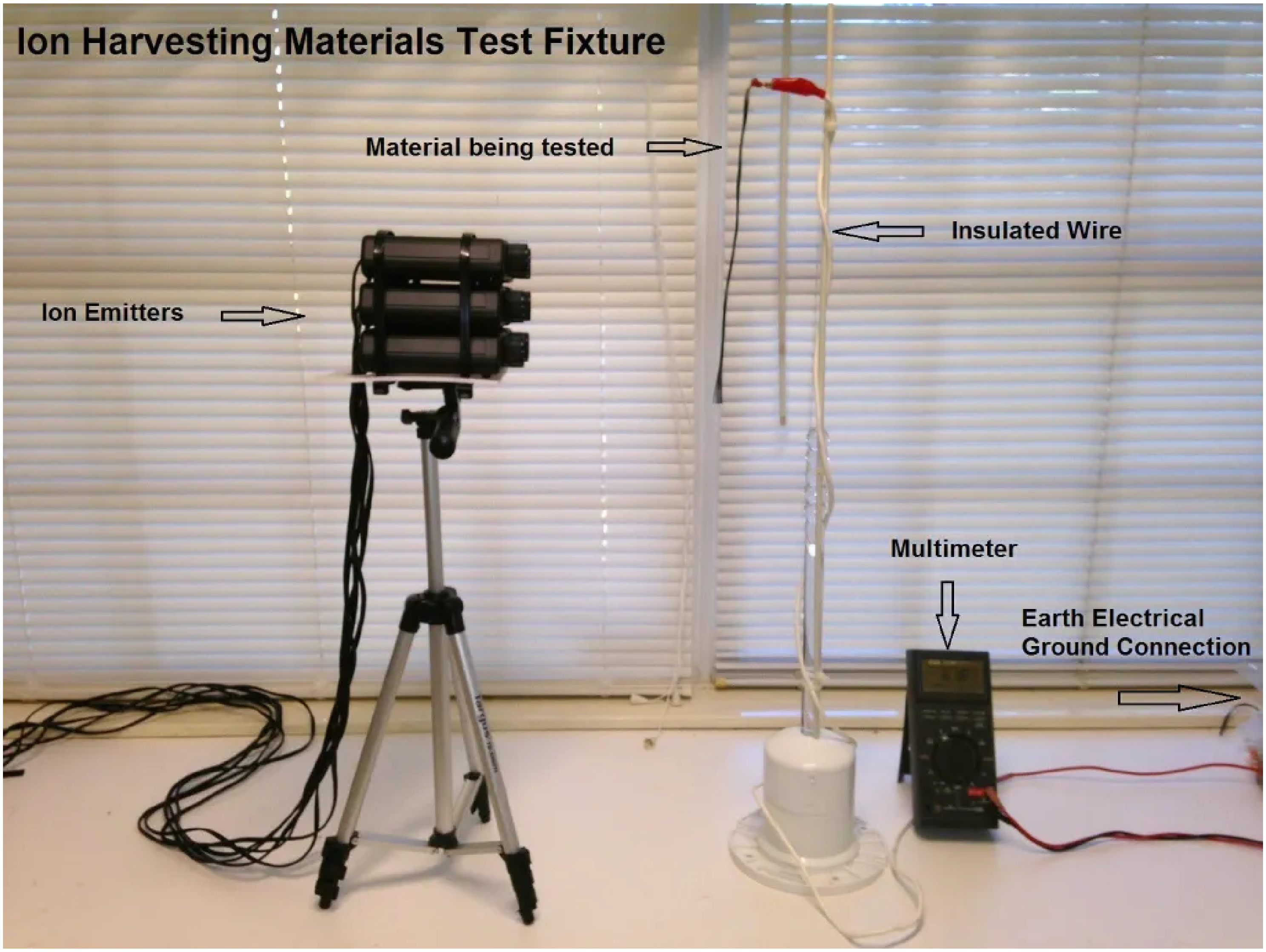
Ion Harvesting Materials Test Fixture
Material being tested
Insulated Wire
Ion Emitters
Multimeter
Earth Electrical Ground Connection

Below is a picture of carbon, a key component of ion collectors, taken at the microscopic level by an electron microscope.

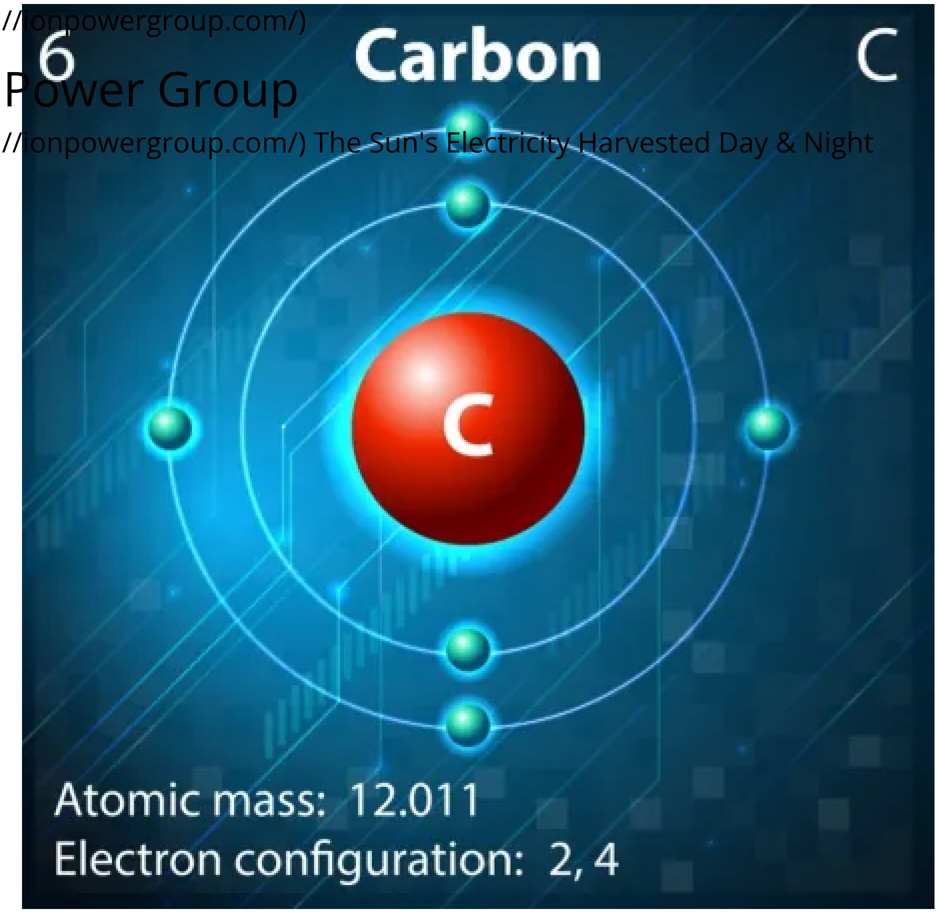
6
Carbon
C
C
Atomic mass: 12.011
Electron configuration: 2, 4

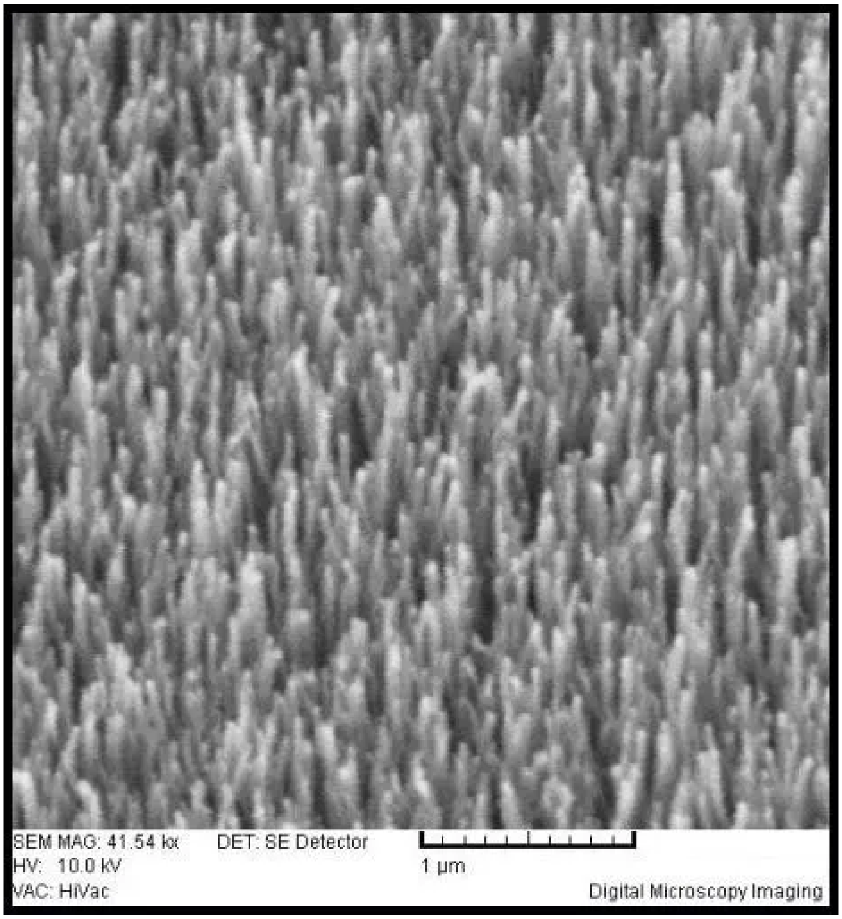
SEM MAG: 41.54 kx
DET: SE Detector
HV: 10.0 kV
VAC: HiVac
1 µm
Digital Microscopy Imaging

TOP

Ion Power Group's internationally patented ion collectors, elevated between two poles in the below depiction, are specially designed to couple to the energetic electrically charged ions in the Martian atmosphere. Ion collectors are designed to harvest the high voltage DC electric charge from naturally occurring atmospheric ions day and night and during dust storms and are not dependent on sun or wind. Ground based circuitry has been designed to maintain the ion collector material at a voltage considerably lower than the surrounding atmospheric ions thereby employing the electrostatic attraction principle causing nearby ions to migrate to the ion collectors. Wires will convey the harvested electricity to the base camp where it is routed to a storage medium (such as super capacitors, graphene batteries, lithium-ion batteries, hydrogen gas fuel cells) to be used for life support, light, heat, the production of oxygen for breathing and rocket fuel (hydrogen) from Martian water, communications, fisheries, agriculture and transportation.

TOP

(https://ionpowergroup.com/)

Ion Power Group

(https://ionpowergroup.com/) The Sun's Electricity Harvested Day & Night

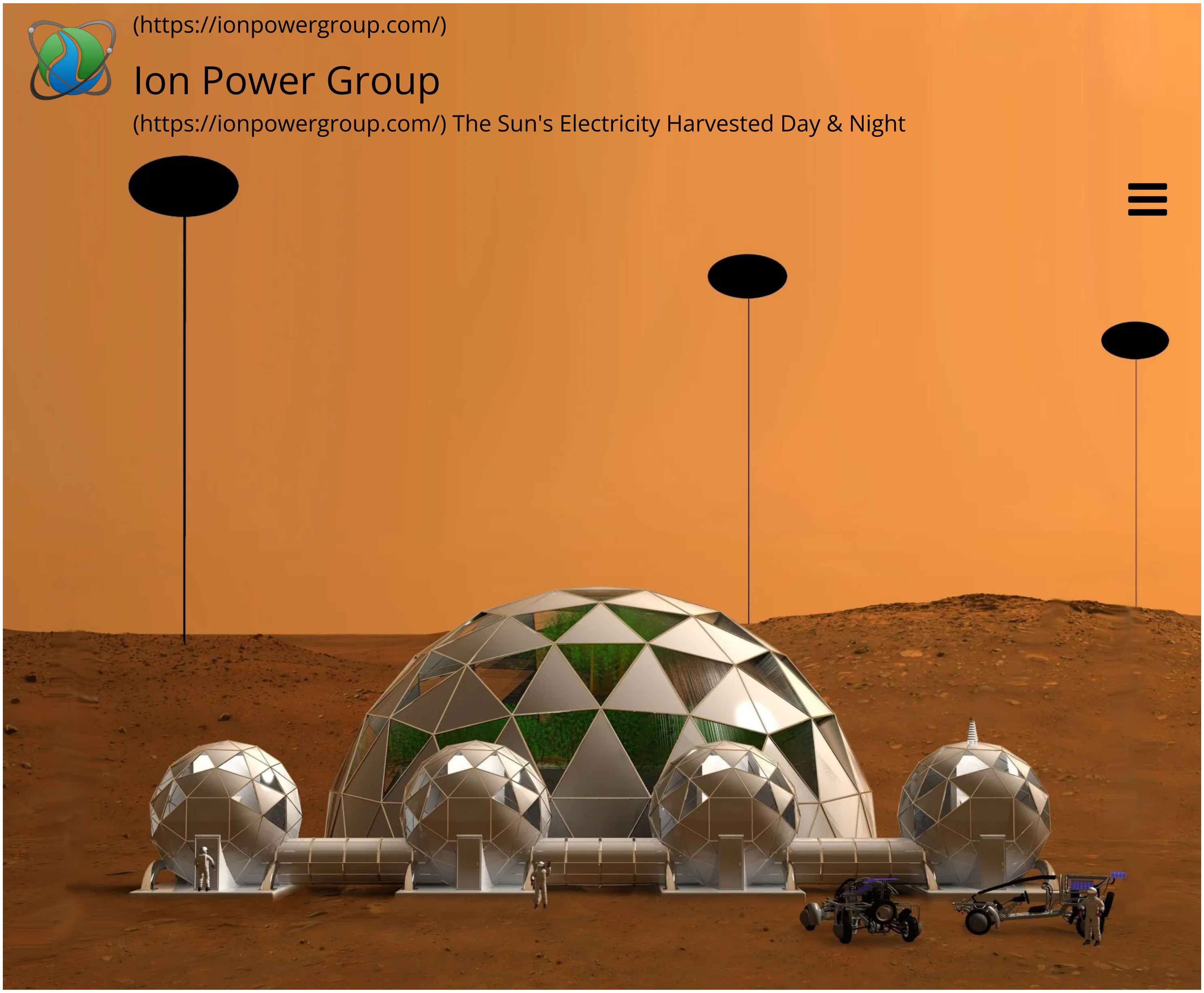

Three tethered aerostats (specially sized for the Martian atmosphere) equipped with Ion Harvesting Technology could generate up to 60kW of electricity on Mars.

TOP

"Solar power… cannot be used at night and during dust storms. It has been shown (in the Power to Mars technical report) that ion power and solar power can be combined to provide a reliable and safe power generation concept." **PhD Atmospheric Physicist and Senior Scientist with the German Aerospace Center, Dr. Andreas J .G. Baumgaertner**

The conditions on Mars including a weak magnetic field, frequent dust storms around the planet and constant bombardment of the Martian atmosphere by GCRs provide an ideal electrical environment for Ion Power Group technology to harvest the electric charge of naturally occurring atmospheric ions to produce reliable, environmentally friendly, electrical power without reliance on sunlight, wind, nuclear or fossil fuels. The highly ionized Martian atmosphere will provide an extremely fertile electrical environment for Ion Power Group's technology to generate clean renewable pollution-free electrical energy day and night and during dust storms. The ability of ion harvesting to generate electricity day/night/dust storms will prove valuable to power robots, rovers, drones, vehicles, living quarters, lights, communications, entertainment equipment, heat to melt Martian ice for drinking, trenchers for extracting

↑ TOP

water from soil, water for agriculture and fisheries as well as producing hydrogen gas to use for rocket fuel and oxygen for breathing through electrolysis of Martian water.

Nearly 100 countries have some level of space program. A growing number of governments, as well as many private and corporate organizations, are developing plans to send robots and hardware to be followed by people to colonize Mars in the foreseeable future. All of these campaigns will require a reliable source of electrical power on the surface of Mars.

Additional information regarding Ion Harvesting Technology’s Principle-of-Operation can be found on the page titled “How it works on Earth. (/how-it-works-on-earth/)

↑ TOP

(https://ionpowergroup.com/)

Ion Power Group

(https://ionpowergroup.com/) The Sun's Electricity Harvested Day & Night

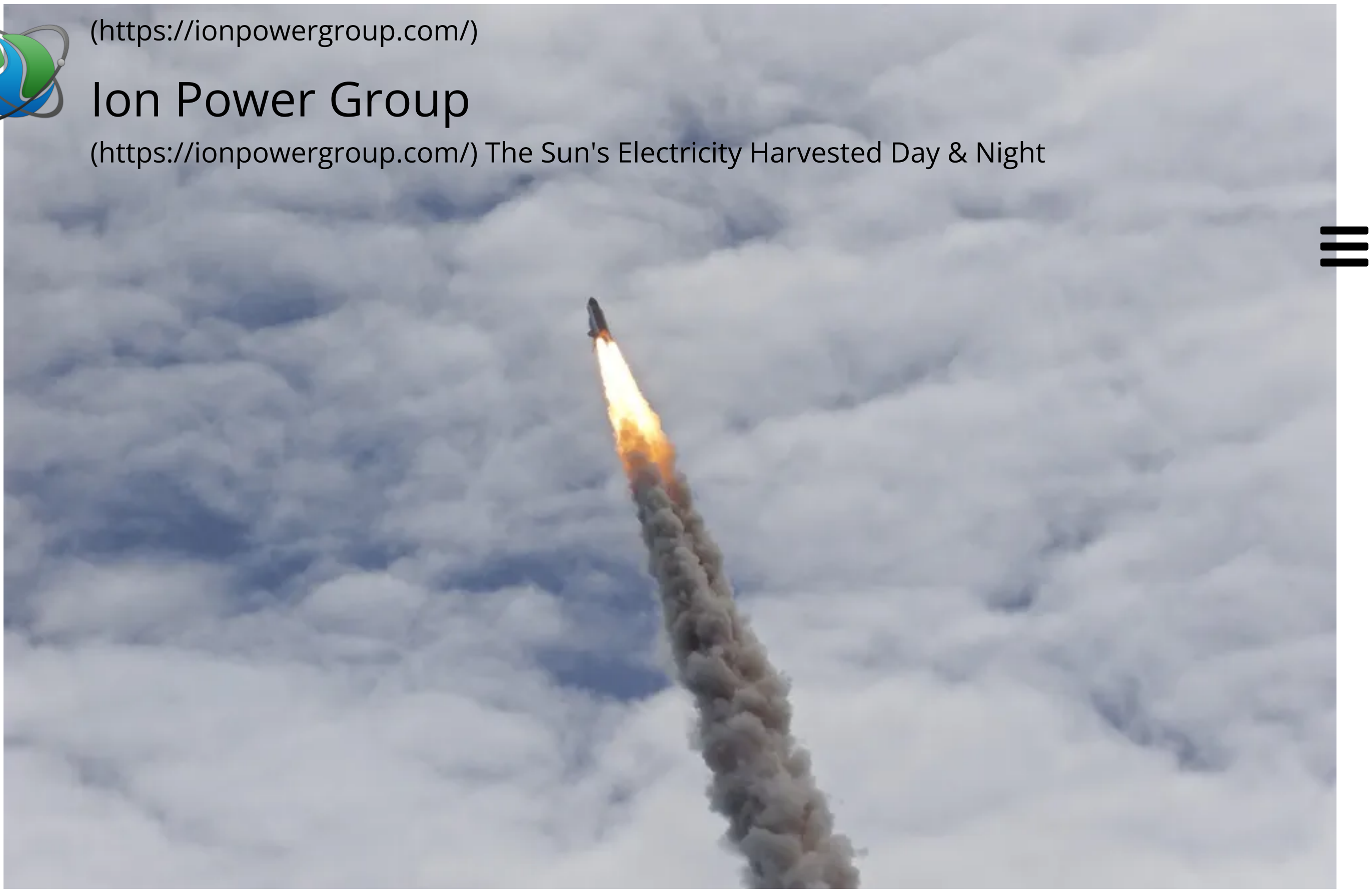

TOP

The survival of humans on Mars requires that the settlers have access to reliable electrical power production, day and night and during the lengthy dust-storms known to occur on Mars. Ion Power Group believes that ion harvesting is the long term solution.

Ion Power Group is a member of the PEW Clean Energy Business Network

Copyright © 2020, Ion Power Group LLC

US 20160248345A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2016/0248345 A1**

McCowen (43) **Pub. Date: Aug. 25, 2016**

(54) **ENERGY COLLECTION**

(71) Applicant: **Ion Power Group LLC**, Navarre, FL (US)

(72) Inventor: **Clint McCowen**, Navarre, FL (US)

(73) Assignee: **Ion Power Group LLC**, Navarre, FL (US)

(21) Appl. No.: **15/145,794**

(22) Filed: **May 3, 2016**

Related U.S. Application Data

(63) Continuation of application No. 14/454,308, filed on Aug. 7, 2014, now Pat. No. 9,331,603.

Publication Classification

(51) **Int. Cl.**

H02N 1/00	(2006.01)
H02N 1/04	(2006.01)

(52) **U.S. Cl.**

CPC . *H02N 1/002* (2013.01); *H02N 1/04* (2013.01)

(57) **ABSTRACT**

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support system. The support system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but graphene, carbon and graphite are preferred. Diodes may be used to restrict the backflow or loss of energy.

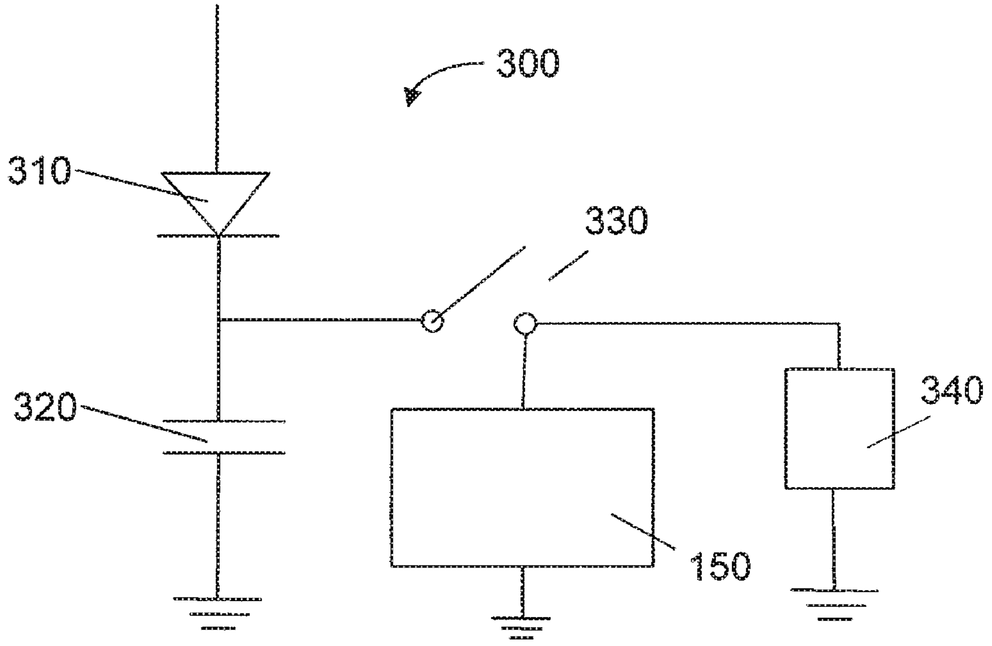

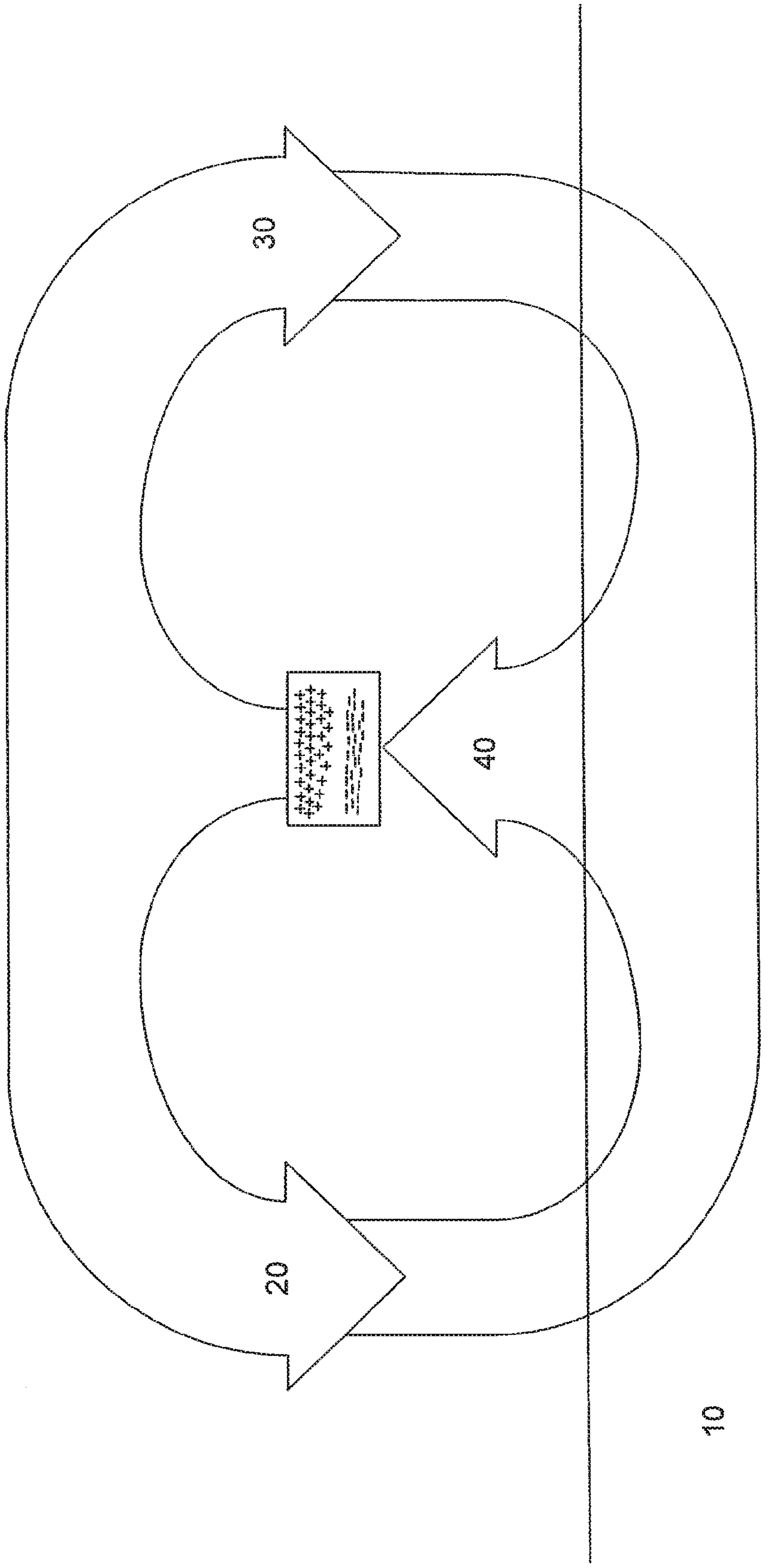

FIGURE 1

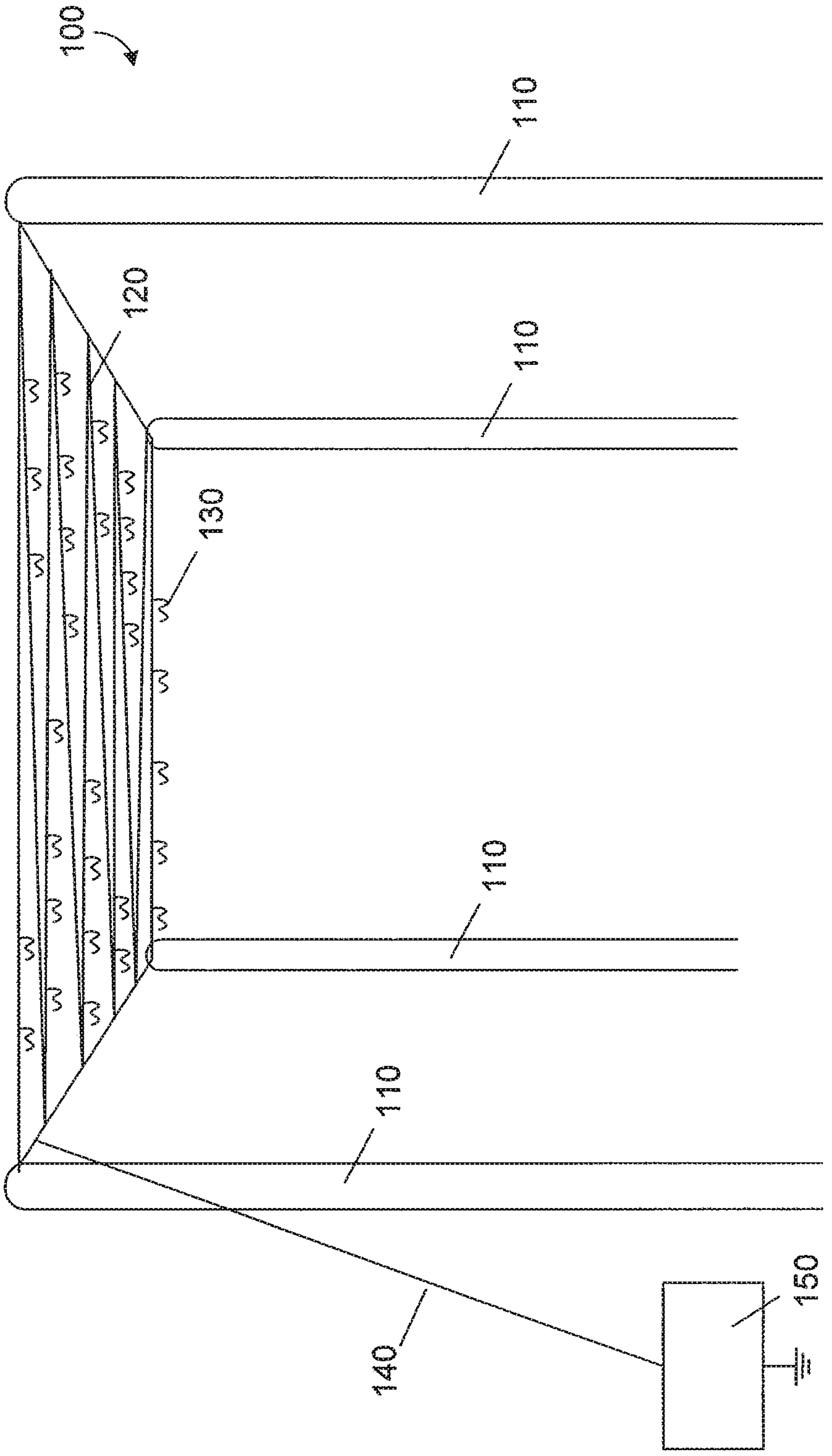
100
110
110
110
110
120
130
140
150

FIGURE 2

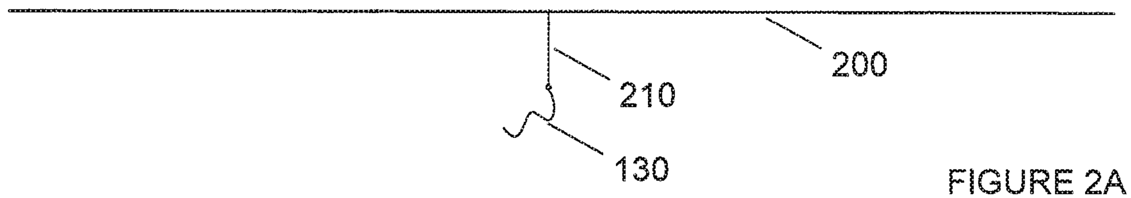

FIGURE 2A

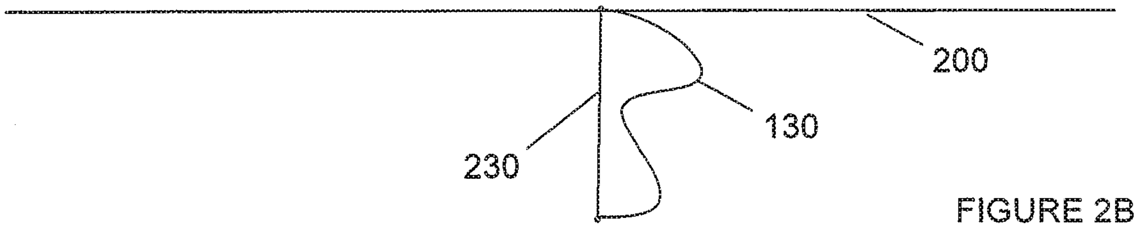

FIGURE 2B

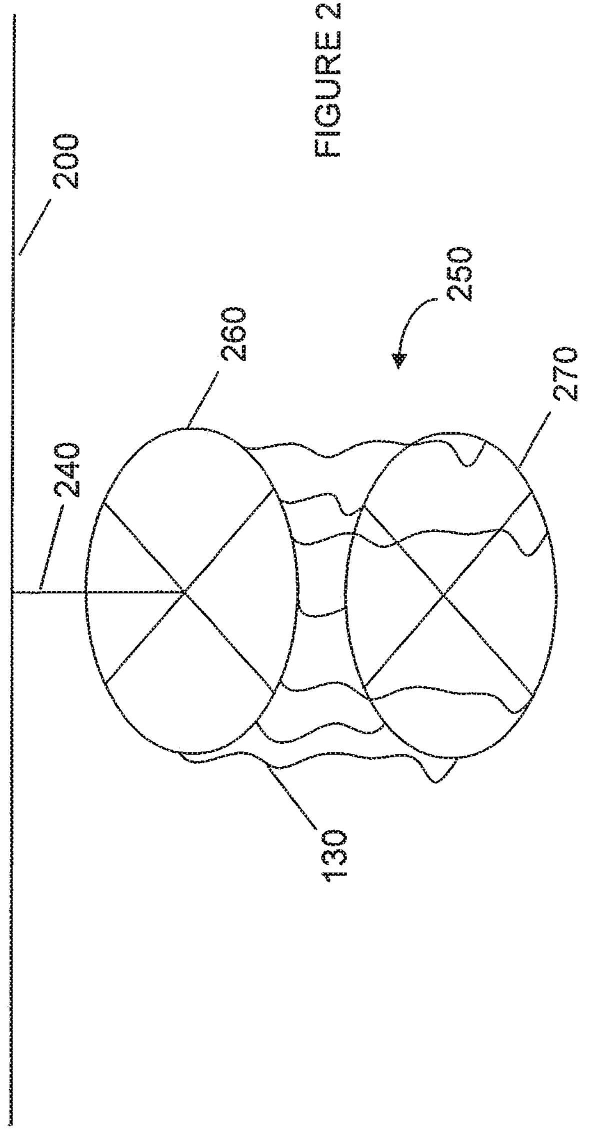
200
240
260
250
270
130

FIGURE 2C

FIGURE 2D

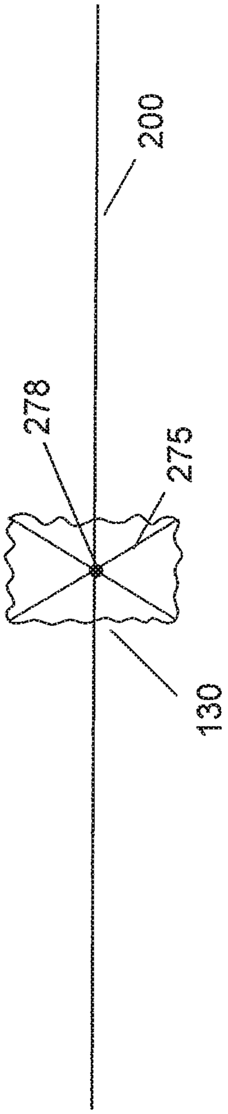
200
278
275
130

FIGURE 2E

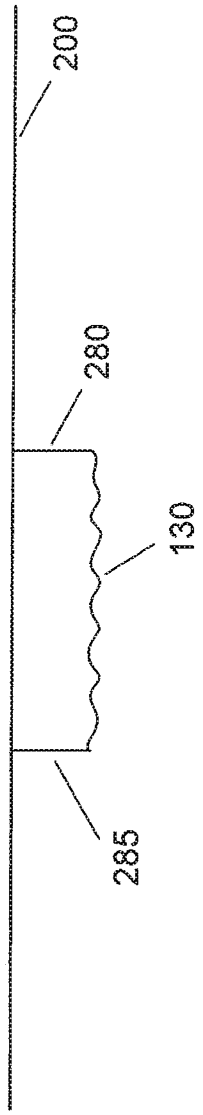
200
280
130
285

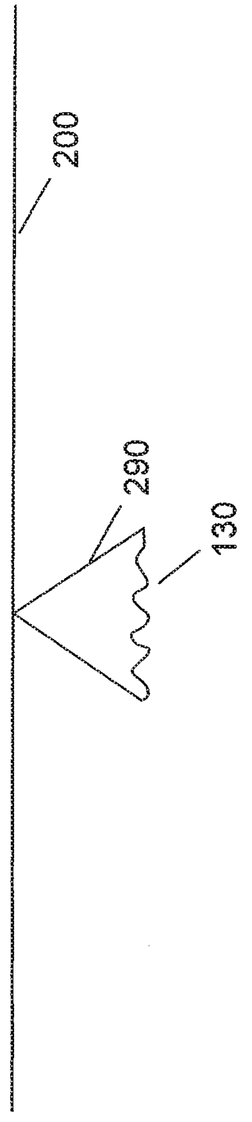

FIGURE 2F

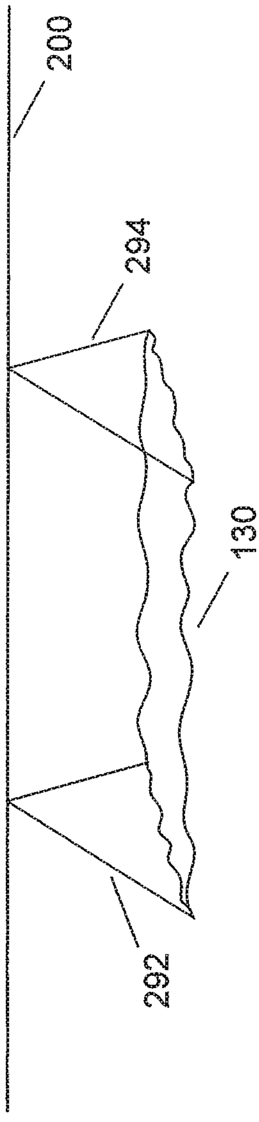
200
294
130
292

FIGURE 2G

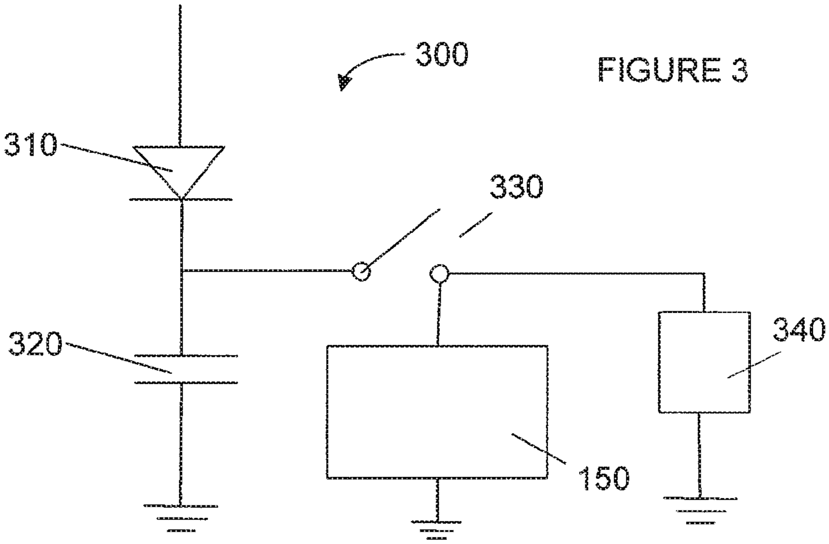
300
310
330
320
340
150

FIGURE 3

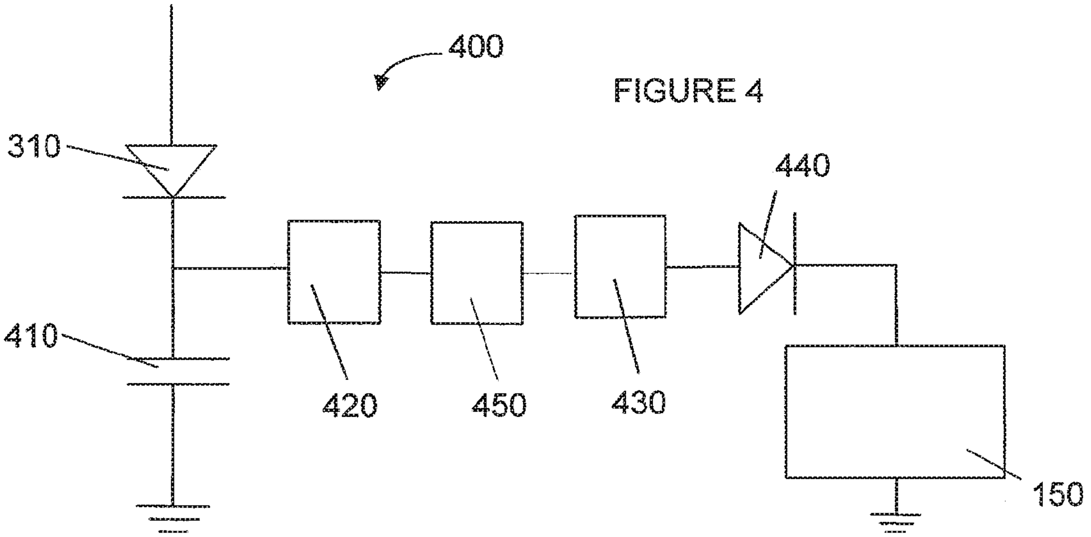
400
310
440
410
420
450
430
150

FIGURE 4

FIGURE 5

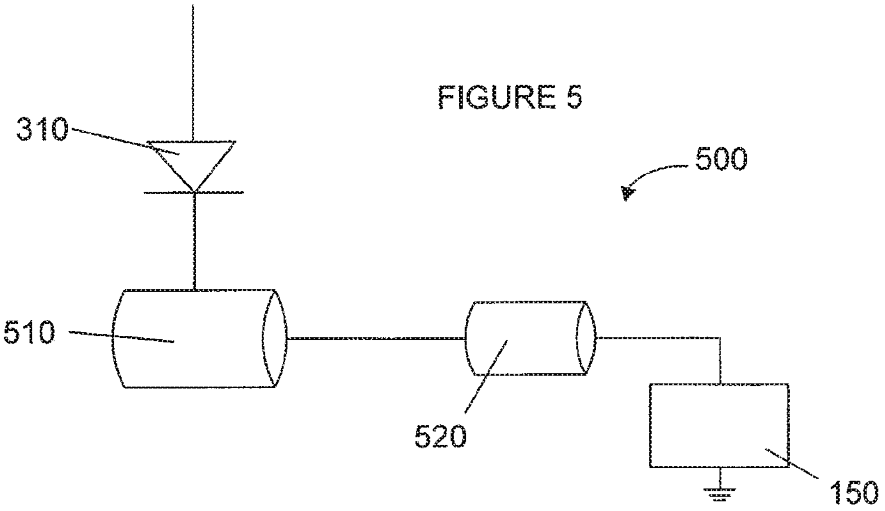
310
500
510
520
150

FIGURE 6

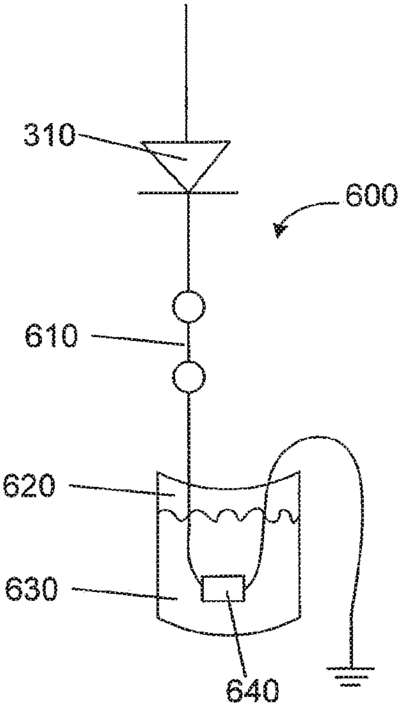
310
600
610
620
630
640

FIGURE 7

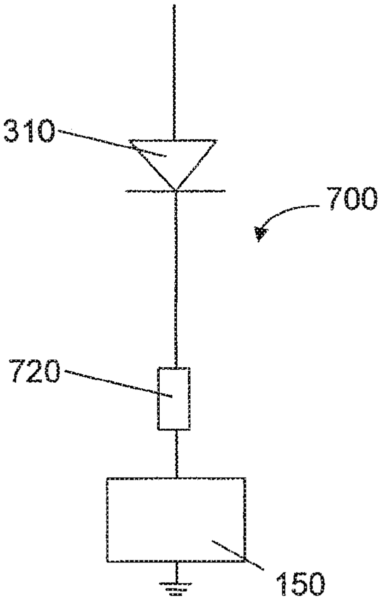
310
700
720
150

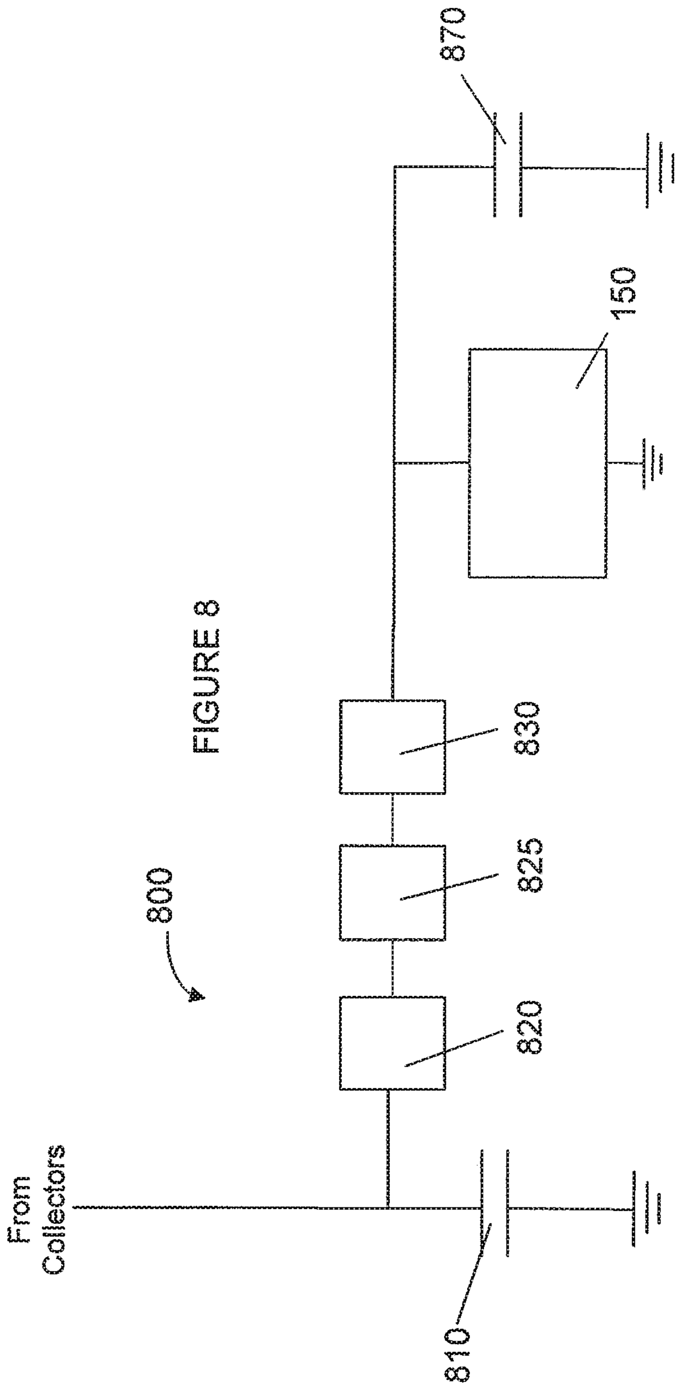

FIGURE 8

800
From Collectors
810
820
825
830
150
870

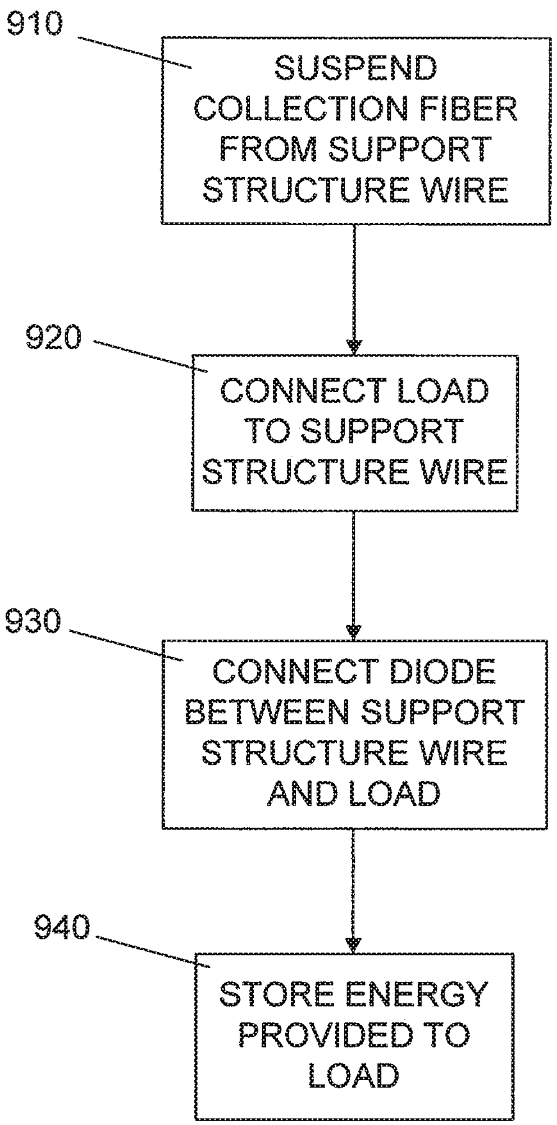
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

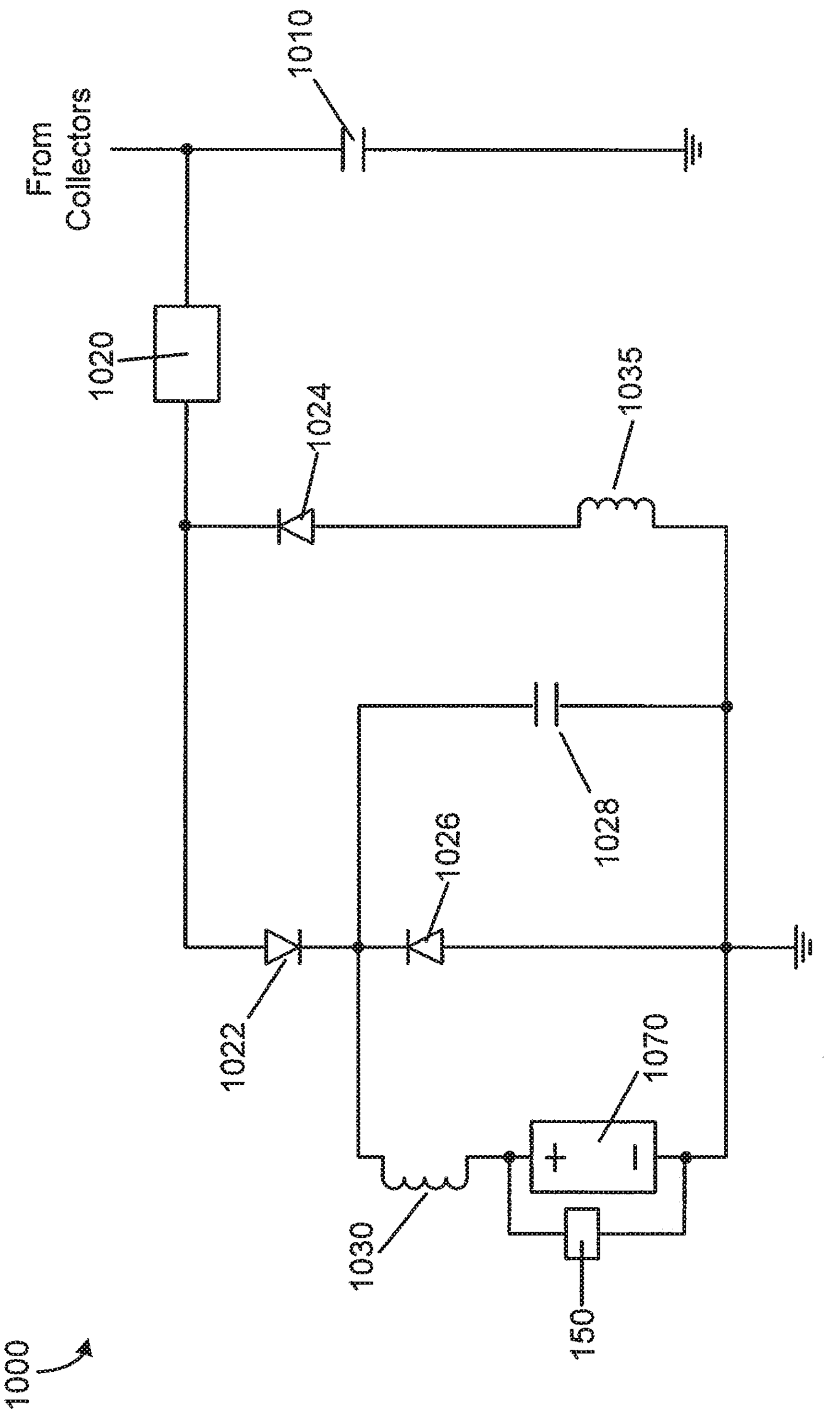
From Collectors
+
−

FIGURE 10

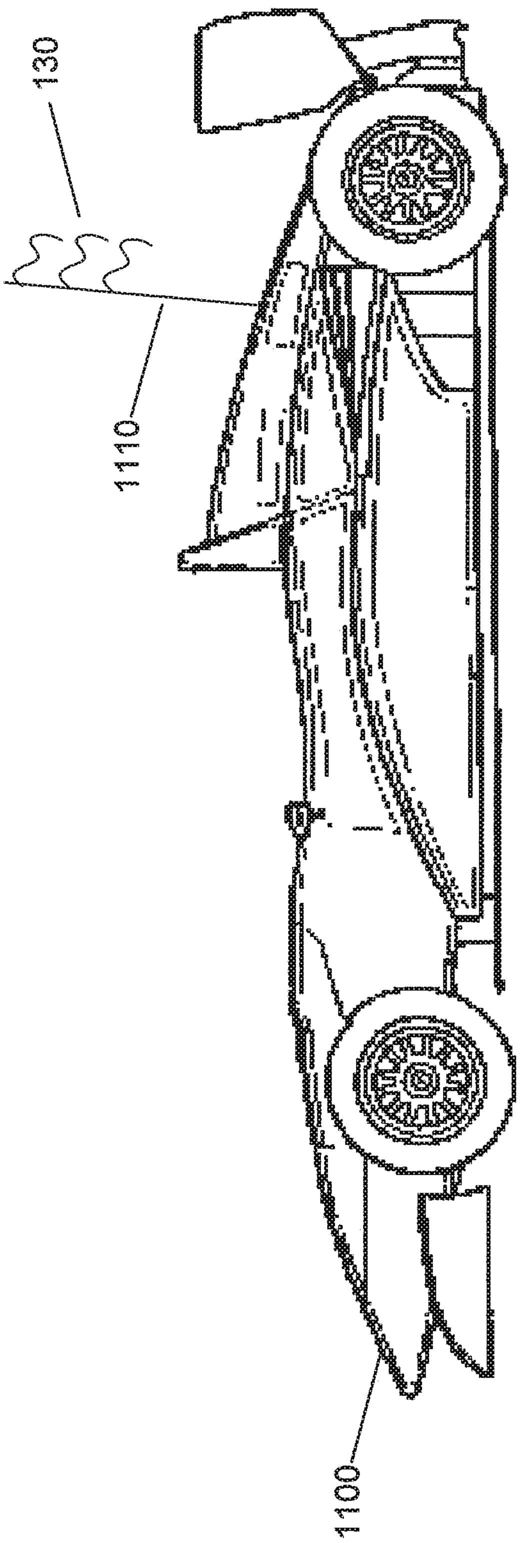
130
1110
1100

FIGURE 11

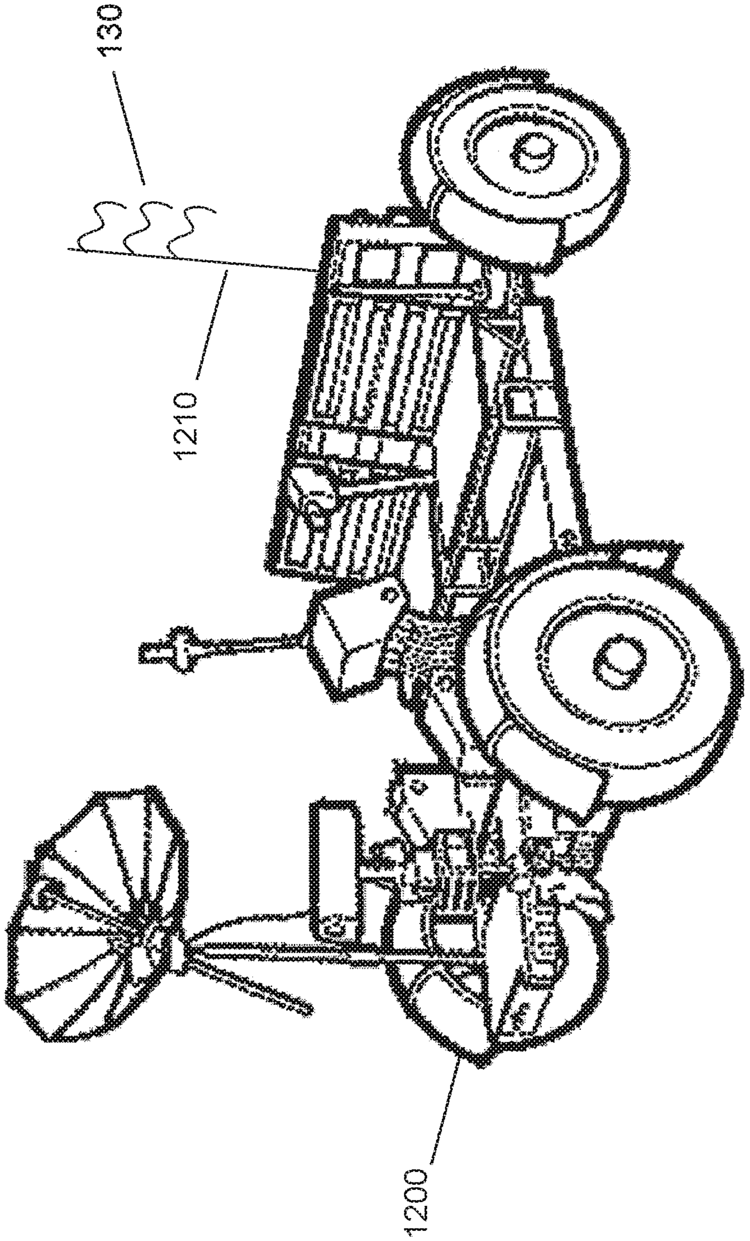
130
1210
1200

FIGURE 12

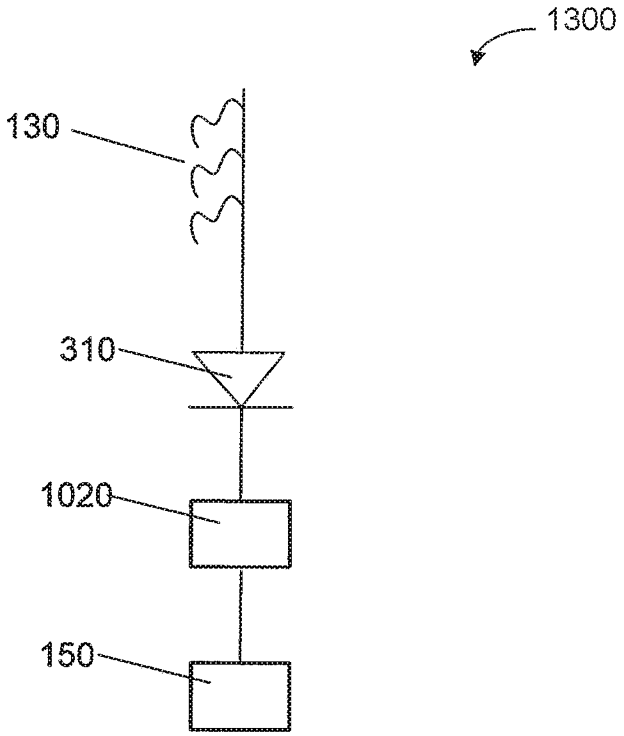
1300
130
310
1020
150

FIGURE 13

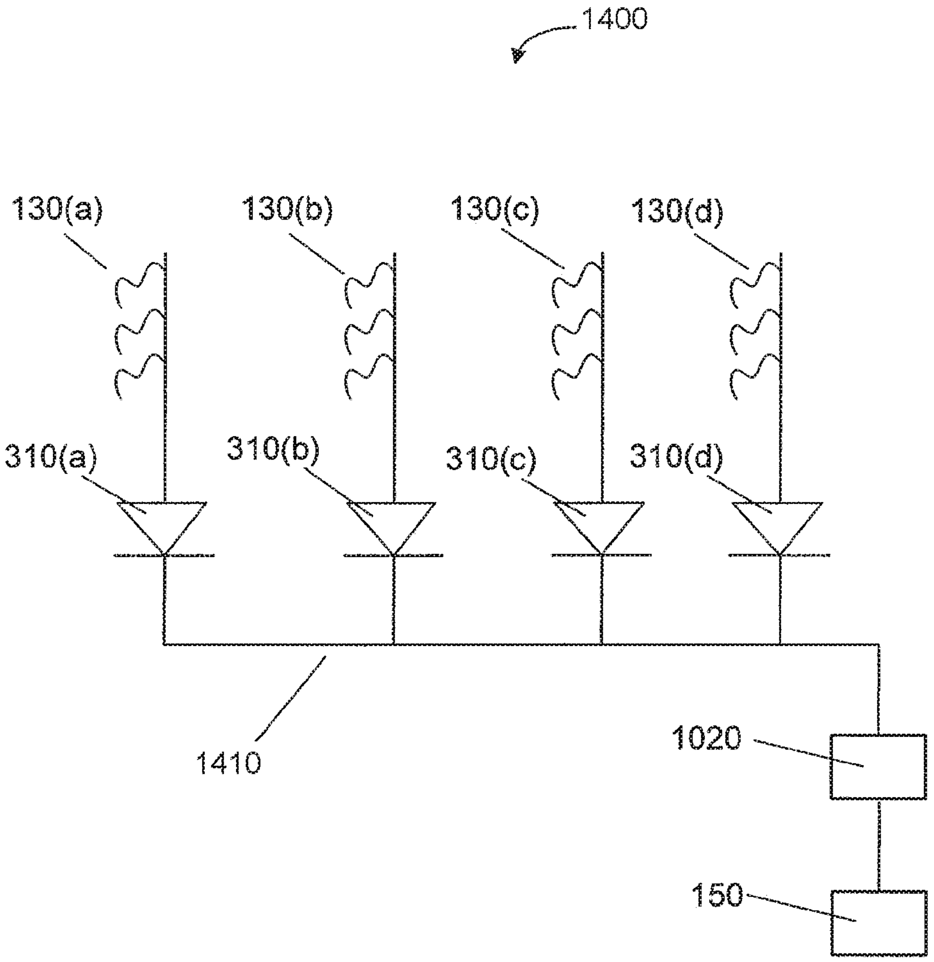
1400
130(a)
130(b)
130(c)
130(d)
310(a)
310(b)
310(c)
310(d)
1410
1020
150

FIGURE 14

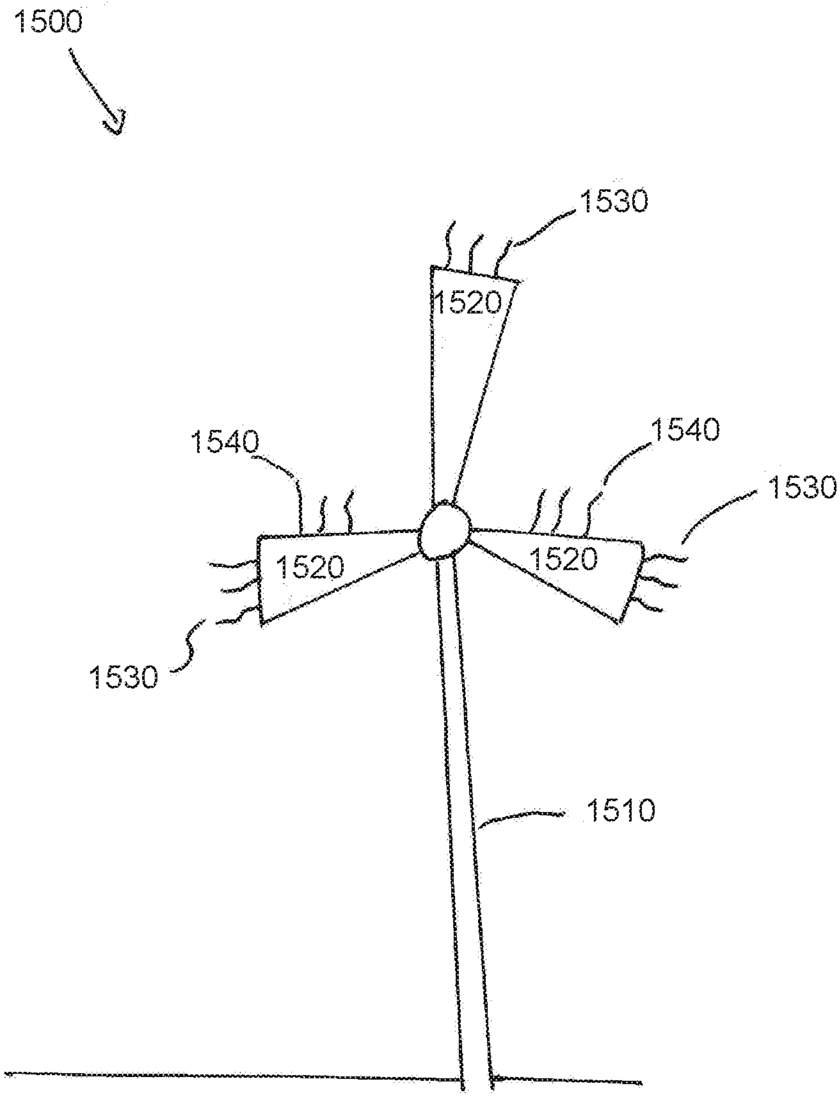
1500
1530
1520
1540
1540
1530
1520
1520
1530
1510

FIGURE 15

ENERGY COLLECTION

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application is a continuation of U.S. utility patent application Ser. No. 14/454,308, filed on Aug. 7, 2014, which is incorporated by reference herein.

TECHNICAL FIELD

[0002] The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

[0003] The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

[0004] In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

[0005] Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

[0006] Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

[0007] While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

[0008] Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure, the support structure comprising at least one of an airplane, drone, blimp, balloon, kite, satellite, train, motorcycle, bike, skateboard, scooter, hovercraft, electronic device, electronic device case, billboard, cell tower, radio tower, camera tower, flag pole, telescopic pole, light pole, utility pole, water tower, building, sky scraper, coliseum, roof top, solar panel and a fixed or mobile structure exceeding 1 inch in height above ground or sea level; at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment electrically connected to the support structure; and a load electrically connected to the at least one collection device.

[0009] Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment from a support structure, the at least one collection device electrically connected to the support structure, the support structure comprising at least one of an airplane, drone, blimp, balloon, kite, satellite, train, motorcycle, bike, skateboard, scooter, hovercraft, electronic device, electronic device case, billboard, cell tower, radio tower, camera tower, flag pole, telescopic pole, light pole, utility pole, water tower, building, sky scraper, coliseum, roof top, solar panel and a fixed or mobile structure exceeding 1 inch in height above ground or sea level; and providing a load with an electrical connection to the at least one collection device to draw current.

[0010] Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

[0012] FIG. **1** is a circuit diagram of a weather energy circuit.

[0013] FIG. **2** is a perspective view of an example embodiment of many energy collectors elevated above ground by a structure.

[0014] FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

[0015] FIG. **2**B is a side view of an example embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

[0016] FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

[0017] FIG. **2**D is a side view of an example embodiment of a support structure for multiple energy collection fibers.

[0018] FIG. **2**E is a side view of a support structure for an energy collection fiber.

[0019] FIG. **2**F is a side view of an example embodiment of a support structure for an energy collection fiber.

[0020] FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

[0021] FIG. **3** is a circuit diagram of an example embodiment of a circuit for the collection of energy.

[0022] FIG. **4** is a circuit diagram of an example embodiment of a circuit for the collection of energy.

[0023] FIG. **5** is a circuit diagram of an example embodiment of an energy collection circuit for powering a generator and motor.

[0024] FIG. **6** is a circuit diagram of an example embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

[0025] FIG. **7** is a circuit diagram of an example embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

[0026] FIG. **8** is a circuit diagram of an example embodiment of a circuit for collecting energy.

[0027] FIG. **9** is a flow diagram of an example embodiment of collecting energy with a collection fiber.

[0028] FIG. **10** is a circuit diagram of an example embodiment of a circuit for collecting energy from a dual polarity source.

[0029] FIG. **11** is a system diagram of an example embodiment of an energy collection system connected to an automobile vehicle.

[0030] FIG. **12** is a system diagram of an example embodiment of an energy collection system connected to a lunar rover vehicle.

[0031] FIG. **13** is a system diagram of an example embodiment of an energy collection system comprising collection devices with a diode.

[0032] FIG. **14** is a system diagram of an example embodiment of an energy collection system comprising multiple legs of the system of FIG. **13**.

[0033] FIG. **15** is a system diagram of an example embodiment of a windmill with energy collectors.

DETAILED DESCRIPTION

[0034] Electric charges on conductors reside entirely on the external surface of the conductors, and tend to concentrate more around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An example embodiment of this disclosure takes advantage of this property, among others, to collect and use the energy generated by an electric field in the atmosphere. Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An example embodiment utilizes carbon or graphite fibers for static electricity collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including aluminum or steel, but most notably, copper. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

[0035] An example embodiment of the collection fibers as collection device **130** includes graphite or carbon fibers. Graphite and carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, or graphite fibers are examples that demonstrate collection ability.

[0036] In at least one example embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than **100** volts per meter under some conditions. When support wire **120** is suspended in the air at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

[0037] A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An example embodiment of the collection device comprises the diode with no collection fiber. A preferred embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restricts energy that may be collected through one fiber from escaping through another fiber.

[0038] Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2**A-**2**G using a collection fiber embodiment. FIG. **2**A presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**.

[0039] Likewise, FIG. **2**B shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

[0040] FIG. 2C presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

[0041] FIG. 2D presents another example embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

[0042] FIG. 2E provides another example embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

[0043] FIG. 2F presents another example embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

[0044] FIG. 2G provides two supports as provided in FIG. 2F such that at least two support members **292**, **294** may be connected to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and between multiple support structures. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**.

[0045] FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load. Capacitor **320** maybe connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

[0046] In an example embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel. Non-limiting examples of spark gap **420** include mercury-reed switches and mercury-wetted reed switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground. The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

[0047] FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

[0048] FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes maybe placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches and mercury-wetted reed switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel.

[0049] FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit. Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

[0050] FIG. **8** presents example circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820**

may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches and mercury-wetted reed switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

[0051] FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

[0052] FIG. **10** presents circuit **1000** as an example embodiment for the collection of energy from a dual polarity source. This may be used, for example, to collect atmospheric energy that reverses in polarity compared with the ground. Such polarity reversal has been discovered as occurring occasionally on Earth during, for example, thunderstorms and bad weather, but has also been observed during good weather. Such polarity reversal may occur on other planetary bodies, including Mars and Venus, as well. Energy polarity on other planets, in deep space, or on other heavenly bodies, may be predominantly negative or predominantly positive. Collector fibers (**130** from FIG. **2**), which may comprise graphene, silicene, and/or other like materials, are capable of collecting positive energy and/or negative energy, and circuit **1000** is capable of processing positive and/or negative energy, providing an output which is always positive. Circuit **1000** may collect energy from one or more collection devices (**130** from FIG. **2**). Charging capacitor **1010** may be used to store a charge until the voltage at spark gap **1020** achieves the spark voltage. Capacitor **1010** is optional.

[0053] A plurality of diodes may be placed in a plurality of positions in circuit **1000**. The voltage from capacitor **1010** may be used to charge spark gap **1020** to a sufficient voltage. Spark gap **1020** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **1020** include mercury-reed switches, mercury-wetted reed switches, open-gap spark gaps, and electronic switches. When spark gap **1020** arcs, energy will arc from an emitting end of spark gap **1020** to a receiving end of spark gap **1020**. The output of spark gap **1020** is electrically connected to the anode of diode **1022** and the cathode of diode **1024**. The cathode of diode **1022** is electrically connected to the cathode of diode **1026** and inductor **1030**. Inductor **1030** may be a fixed value inductor or a variable inductor. The anode of diode **1026** is electrically connected to ground. Capacitor **1028** is electrically connected between ground and the junction of the cathodes of diode **1022** and diode **1026**. Inductor **1035** is electrically connected between ground and the anode of diode **1024**. Inductor **1035** may be a fixed value inductor or a variable inductor. Capacitor **1070**, the anode of diode **1026**, inductor **1035**, and load **1050** are electrically connected to ground. Capacitor **1070** may be placed in parallel with load **150**.

[0054] FIGS. **11** and **12** provide example embodiments of vehicle **1110**, which utilizes electricity, the vehicle employing systems of energy collection provided herein. Vehicle **1100** in FIG. **11** is shown as an automobile vehicle, but could be any means of locomotion that utilizes electricity, including a car, a train, a motorcycle, a boat, an airplane, robotic rovers, space craft, etc. Vehicle **1200** in FIG. **12** is shown as a lunar rover vehicle. In FIGS. **11** and **12**, support rod **1110**, **1210** is electrically connected to an electrical system in vehicle **1100**, **1200**. Energy collectors **130**, which may comprise graphene, silicene, and/or other like materials, are electrically connected to support rod **1110**, **1210** and may be used to supply energy to electrical circuits within the vehicle. A non-limiting use includes a top-off charge for a battery system, on-board hydrogen production, and/or assisting in the same. Energy collectors **130** may be used to augment the efficiency of the locomotion that utilizes electrical energy as well.

[0055] FIG. **13** provides an example embodiment of energy collection system **1200** in which diode **310** is used to isolate collection devices **130** from spark gap **1020** and load **150**. Collection devices **130** may comprise graphite, carbon fibers, carbon/carbon fibers, graphene, silicene, and/or other like materials, or a mixture thereof.

[0056] FIG. **14** provides an example embodiment of energy collection system **1400** in which a plurality of energy collection systems, such as that provided in FIG. **13**, are combined. Each leg consisting of collection devices **130**, which may comprise graphene, silicene, and/or other like materials, and diode **310** are connected in parallel with other legs, each leg electrically connected to trunk wire **1410**. The legs could also be connected serially. Trunk wire **1410** is electrically connected to a collection circuit, which may comprise load **150** and spark gap **1020** in any configuration that has been previously discussed. Each leg may comprise one or more collection devices **130** and at least one diode electrically connected between the collection devices and the collection circuit. Although three collection devices **130** are shown on each leg, any number of collection devices may be used. Although four legs are shown, any number of legs may be used.

[0057] FIG. **15** presents a system diagram of an example embodiment of a windmill with energy collectors, which may comprise graphene, silicene, and/or other like materials in an example embodiment. A windmill is an engine powered by the energy of wind to produce alternative forms of energy. They may, for example, be implemented as small tower mounted wind engines used to pump water on farms. The modern wind power machines used for generating electricity are more properly called wind turbines. Common applications of windmills are grain milling, water pumping, threshing, and saw mills. Over the ages, windmills have evolved into more sophisticated and efficient wind-powered water pumps and electric power generators. In an example embodiment, as provided in FIG. **10**, windmill tower **1500** of suitable height and/or propeller **1520** of windmill tower **1500** may be equipped with energy collecting fibers **1530**, **1540**, which may comprise graphene, silicene, and/or other like materials in an example embodiment. Collecting fibers **1530**, **1540** may turn windmill **1500** into a power producing asset even when there is not enough wind to turn propellers **1520**. During periods when there is enough wind to turn propellers **1520**, collecting fibers **1530**, **1540** may supplement/boost the amount of energy the windmill produces.

[0058] A windmill is an engine powered by the energy of wind to produce alternative forms of energy. They may, for example, be implemented as small tower mounted wind engines used to pump water on farms. The modern wind power machines used for generating electricity are more properly called wind turbines. Common applications of windmills are grain milling, water pumping, threshing, and saw mills. Over the ages, windmills have evolved into more sophisticated and efficient wind-powered water pumps and electric power generators. In an example embodiment, as provided in FIG. **10**, windmill tower **1000** of suitable height and/or propeller **1020** of windmill tower **1000** may be equipped with energy collecting fibers **1030**, **1040**. Collecting fibers **1030**, **1040** may turn windmill **1000** into a power producing asset even when there is not enough wind to turn propellers **1020**. During periods when there is enough wind to turn propellers **1020**, collecting fibers **1030**, **1040** may supplement/boost the amount of energy the windmill produces.

[0059] Windmill **1500**, properly equipped with ion collectors **1530**, **1540**, such as a non-limiting example of fibers with graphene, silicene, and/or other like materials, can produce electricity: 1) by virtue of providing altitude to the fiber to harvest ions, and 2) while the propeller is turning, by virtue of wind blowing over the fiber producing electricity, among other reasons, via the triboelectric effect (however, it is also possible for the triboelectric effect to occur, producing electricity, in winds too weak to turn the propeller).

[0060] There are at least two ways that energy collectors may be employed on or in a windmill propeller to harvest energy. Propellers **1520** may be equipped with energy collectors **1530**, **1540** attached to, or supported by, propeller **1520** with wires (or metal embedded in, or on propeller **1520**) electrically connecting energy collectors **1530**, **1540**, which may comprise graphene, silicene, and/or other like materials, to a load or power conversion circuit. There may be a requirement to electrically isolate energy collectors **1530**, **1540**, which are added to propeller **1520**, from electrical ground, so that the energy collected does not short to ground through propeller **1520** itself or through support tower **1510**, but rather is conveyed to the load or power conversion circuit. Energy collectors may be connected to the end of propellers **1520** such as collectors **1530**. Alternatively, energy collectors may be connected to the sides of propellers **1520** such as collectors **1540**.

[0061] Alternatively, propeller **1520** may be constructed of carbon fiber or other suitable material, with wires (or the structural metal supporting propeller **1520** may be used) electrically connecting to a load or power conversion circuit. In the case of propeller **1520** itself being constructed of carbon fiber, for example, the fiber may be ‘rough finished’ in selected areas so that the fiber is “fuzzy.” For example, small portions of it may protrude into the air as a means of enhancing collection efficiency. The fuzzy parts of collectors **1530**, **1540** may do much of the collecting. There may be a requirement to electrically isolate carbon fiber propeller **1520** from electrical ground, so that the energy it collects does not short to ground through metal support tower **1510**, but rather is conveyed to the load or power conversion circuit. Diodes may be implemented within the circuit to prevent the backflow of energy, although diodes may not be necessary in some applications.

[0062] In an alternative embodiment, windmill **1500** may be used as a base on which to secure an even higher extension tower to support the energy collectors and/or horizontal supports extending out from tower **1510** to support the energy collectors. Electrical energy may be generated via ion collection due to altitude and also when a breeze or wind blows over the collectors supported by tower **1510**.

[0063] In alternative embodiments to windmill **1500**, other non-limiting example support structures include airplanes, drones, blimps, balloons, kites, satellites, cars, boats, trucks, (including automobile and other transportation conveyance tires), trains, motorcycles, bikes, skateboards, scooters, hovercraft (automobiles and conveyance of any kind), billboards, cell towers, radio towers, camera towers, flag poles, towers of any kind including telescopic, light poles, utility poles, water towers, buildings, sky scrapers, coliseums, roof tops, solar panel and all fixed or mobile structures exceeding 1 inch in height above ground or sea level.

[0064] An example embodiment of a support structure may also include cell phones and other electronic devices and their cases, including cases containing rechargeable batteries. For example, someone may set her cell phone or other electronic device or battery pack on the window ledge of a tall apartment building to help charge it. Other example support structures may include space stations, moon and Mars structures, rockets, planetary rovers and drones including robots and artificial intelligence entities.

[0065] Under some conditions, ambient voltage may be found to be 180-400 volts at around 6 ft, with low current. With the new generation of low current devices being developed, a hat containing ion harvesting material may provide enough charge, or supplemental charge, collected over time to help power low current devices such as future cell phones, tracking devices, GPS, audio devices, smart glasses, etc. Clothes may also be included as examples of support structures. Friction of the ion collection material (such as non-limiting examples of carbon, graphite, silicene and graphene) against unlike materials, such as wool, polyester, cotton, etc (used in clothes) may cause a voltage to be generated when rubbed together. Additionally, wind passing over the ion collection material has been demonstrated to generate voltage, even at low altitude. In an additional example embodiment, embedding collection devices into automobile tires (for example, in a particular pattern) could generate collectible voltage.

[0066] Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

[0067] It should be emphasized that the above-described embodiments of the present disclosure, particularly, any “preferred” embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

1. A method of collecting energy comprising:

suspending at least one graphene collection device, electrically connected to the support structure, the support structure comprising at least one of an airplane, drone, blimp, balloon, kite, satellite, train, motorcycle, bike, skateboard, scooter, hovercraft, electronic device, electronic device case, billboard, cell tower, radio tower, camera tower, flag pole, telescopic pole, light pole, utility pole, water tower, building, sky scraper, coliseum, roof top, solar panel and a fixed or mobile structure exceeding 1 inch in height above ground or sea level; and

providing a load with an electrical connection to the at least one collection device to draw current.

2. The method of claim **1**, wherein the graphene collection device collects energy by triboelectric effect.

3. The method of claim **1**, wherein the graphene collection device comprises a graphene mesh.

4. The method of claim **1**, wherein the graphene collection device comprises a diode and a collection fiber and the diode is electrically connected between the collection fiber and the load.

5. The method of claim **1**, further comprising storing energy provided to the load.

6. The method of claim **5**, wherein storing energy provided to the load comprises storing energy in a capacitor or an inductor.

7. The method of claim **3**, further comprising a fuel cell for producing hydrogen from the collected energy.

8. A system of energy collection comprising:

a support structure, the support structure comprising at least one of an airplane, drone, blimp, balloon, kite, satellite, train, motorcycle, bike, skateboard, scooter, hovercraft, electronic device, electronic device case, billboard, cell tower, radio tower, camera tower, flag pole, telescopic pole, light pole, utility pole, water tower, building, sky scraper, coliseum, roof top, solar panel and a fixed or mobile structure exceeding 1 inch in height above ground or sea level;

at least one graphene collection device electrically connected to the support structure; and

a load electrically connected to the at least one collection device.

9. The system of claim **8**, wherein the graphene collection device collects energy by triboelectric effect.

10. The system of claim **8**, wherein the graphene collection device comprises a graphene mesh.

11. The system of claim **8**, wherein the graphene collection device comprises a collection fiber and a diode electrically connected between the load and the collection fiber.

12. The system of claim **11**, wherein the diode is elevated relative to the ground level.

13. The system of claim **10**, wherein the support structure comprises a vehicle.

14. The system of claim **8**, further comprising a diode electrically connected between the at least one graphene collection device and the support structure.

15. The system of claim **8**, further comprising: a switch connected in series between the at least one graphene collection device and the load; and a capacitor connected in parallel with the switch and the load.

16. The system of claim **15**, wherein the switch comprises an interrupter connected between the load and at least one graphene collection device, and wherein the interrupter comprises at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

17. The system of claim **8**, further comprising: a motor for providing power, the motor connected between the at least one graphene collection device and the load; and a generator powered by the motor.

18. The system of claim **8**, further comprising a fuel cell between the support structure and the load.

19. The system of claim **18**, wherein the fuel cell produces hydrogen and oxygen.

20. A system of collecting energy comprising:

means for suspending at least one graphene collection device, the at least one graphene collection device electrically connected to the means for suspending; and

means for inducing current flow, the means for inducing current flow electrically connected to the graphene collection device.

* * * * *

US008686575B2

(12) **United States Patent**
McCowen

(10) **Patent No.: US 8,686,575 B2**
(45) **Date of Patent: Apr. 1, 2014**

(54) **ENERGY COLLECTION**

(75) Inventor: **Clint McCowen**, Navarre, FL (US)

(73) Assignee: **Ion Power Group, LLC**, Navarre, FL (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **12/255,130**

(22) Filed: **Oct. 21, 2008**

(65) **Prior Publication Data**

US 2009/0040680 A1 Feb. 12, 2009

Related U.S. Application Data

(63) Continuation of application No. 11/358,264, filed on Feb. 21, 2006, now Pat. No. 7,439,712.

(51) **Int. Cl.**

F02B 63/04	(2006.01)
F03G 7/08	(2006.01)
H02K 7/18	(2006.01)
H02N 1/00	(2006.01)

(52) **U.S. Cl.**
USPC .... **290/1 R**; 322/2 A

(58) **Field of Classification Search**
USPC .... 290/1 R; 322/2 A; 126/660; 60/641.8, 60/641.11, 641.12, 659, 698; 244/158.2
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427	A		5/1901	Palencsar	307/149
911,260	A		2/1909	Pennock	307/149
1,014,719	A		1/1912	Pennock	307/149
2,473,819	A	*	6/1949	Pittman	315/36
3,532,959	A	*	10/1970	Erickson	320/166
3,780,722	A	*	12/1973	Swet	126/680
3,946,227	A		3/1976	Bingham	250/281
4,079,225	A	*	3/1978	Warner	219/110
4,092,250	A		5/1978	Sano et al.	210/500.29
4,104,696	A		8/1978	Cochran, Jr.	361/229
4,146,800	A	*	3/1979	Gregory et al.	290/44
4,180,056	A	*	12/1979	Schnabel et al.	126/660
4,201,197	A	*	5/1980	Dismer	126/600
4,206,396	A	*	6/1980	Marks	322/2 A
4,224,496	A	*	9/1980	Riordan et al.	219/110
4,307,936	A	*	12/1981	Ochiai	385/120
4,314,192	A	*	2/1982	Caro	322/2 R
4,346,478	A	*	8/1982	Sichling	398/107
4,425,905	A	*	1/1984	Mori	126/578
4,433,248	A	*	2/1984	Marks	290/44
RE31,678	E	*	9/1984	Ochiai	385/115

(Continued)

FOREIGN PATENT DOCUMENTS

WO	2007098341	8/2007

OTHER PUBLICATIONS

"Emissivity Values for Common Materials & Non-Metals", Raytek—A Fluke Company, http://www.raytek.com/Raytek/en-r0/IREducation/EmissivityNonMetals.htm, 1999.*

Primary Examiner — Pedro J Cuevas
(74) *Attorney, Agent, or Firm* — Benjamin A. Balser; Balser & Grell IP Law

(57) **ABSTRACT**

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support wire system supported by poles. The support wire system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but carbon and graphite are preferred. Diodes may be used to restrict the backflow or loss of energy.

25 Claims, 10 Drawing Sheets

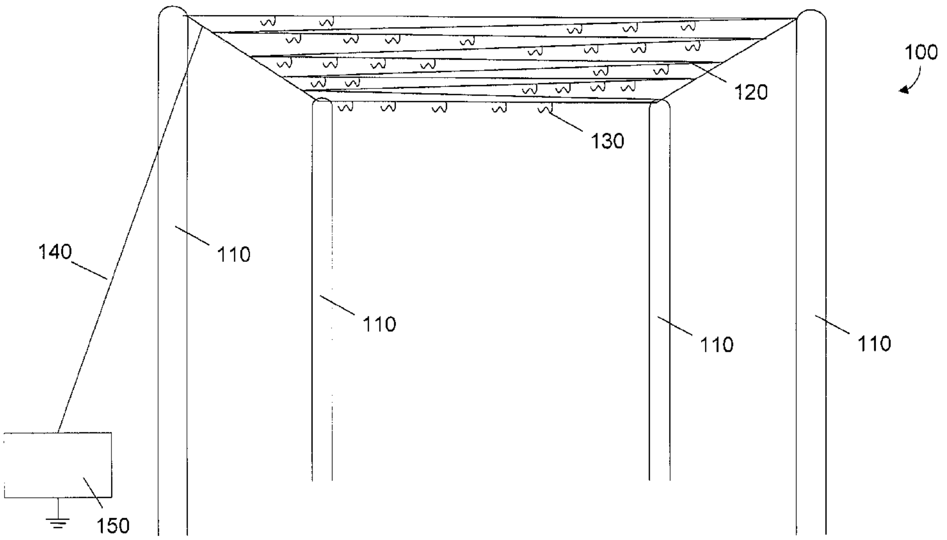

(56) References Cited

U.S. PATENT DOCUMENTS

4,483,311	A *	11/1984	Whitaker	126/602
4,489,269	A *	12/1984	Edling et al.	322/2 R
4,512,335	A *	4/1985	Mori	126/678
4,653,223	A *	3/1987	Mori	47/17
4,653,472	A *	3/1987	Mori	126/683
4,676,226	A *	6/1987	Mori	126/680
4,809,675	A *	3/1989	Mori	126/578
4,852,454	A *	8/1989	Batchelder	89/1.11
4,943,125	A *	7/1990	Laundre' et al.	385/35
4,988,159	A *	1/1991	Turner et al.	385/33
5,047,892	A	9/1991	Sakata et al.	361/231
5,114,101	A *	5/1992	Stern et al.	244/172.8
5,145,257	A *	9/1992	Bryant et al.	374/131
5,379,103	A *	1/1995	Zigler	356/73
5,851,309	A *	12/1998	Kousa	136/248
5,942,806	A *	8/1999	Veliadis	290/1 R
6,025,591	A	2/2000	Taylor et al.	250/292
6,038,363	A *	3/2000	Slater et al.	385/147
6,116,544	A *	9/2000	Forward et al.	244/158.2
6,173,922	B1 *	1/2001	Hoyt et al.	244/158.2
6,226,440	B1 *	5/2001	Lyons	385/119
6,419,191	B1 *	7/2002	Hoyt et al.	244/158.2
6,425,391	B1 *	7/2002	Davoren et al.	126/683
6,735,830	B1	5/2004	Merciel	250/324
6,765,212	B2 *	7/2004	Goetz et al.	250/339.11
6,846,447	B2 *	1/2005	Mbachu et al.	264/406
6,853,447	B2 *	2/2005	Goetz	356/237.1
6,894,772	B2 *	5/2005	Goetz et al.	356/237.1
6,925,852	B2 *	8/2005	Susko	73/23.2
6,974,110	B2	12/2005	Grandics	244/171.5
7,109,597	B1 *	9/2006	Bose	290/43
7,128,867	B2 *	10/2006	Mbachu et al.	264/406
7,231,809	B2 *	6/2007	Susko	73/31.05
7,259,475	B2 *	8/2007	Hong et al.	307/64
7,439,712	B2 *	10/2008	McCowen	322/2 A
7,449,668	B2 *	11/2008	Schutten et al.	250/214 SW
7,478,712	B2 *	1/2009	McCowen	191/45 R
7,532,819	B1 *	5/2009	Triebes et al.	398/121
2002/0026933	A1 *	3/2002	Gottlieb	126/628
2002/0108892	A1 *	8/2002	Goetz et al.	209/576
2002/0109094	A1 *	8/2002	Goetz et al.	250/339.11
2002/0109835	A1 *	8/2002	Goetz	356/237.1
2003/0107869	A1	6/2003	Pan et al.	361/306.2
2003/0125630	A1 *	7/2003	Furnish	600/461
2003/0193319	A1 *	10/2003	Wood et al.	322/2 A
2004/0011925	A1 *	1/2004	Grandics	244/62
2004/0083793	A1 *	5/2004	Susko	73/31.05
2004/0094853	A1 *	5/2004	Mbachu et al.	264/40.1
2004/0160711	A1	8/2004	Stumberger	361/20
2005/0101024	A1 *	5/2005	Mbachu et al.	436/85
2005/0136311	A1 *	6/2005	Ueda et al.	429/30
2005/0270525	A1 *	12/2005	Susko	356/326
2007/0107765	A1 *	5/2007	Schutten et al.	136/243
2007/0195481	A1 *	8/2007	McCowen	361/233
2007/0218323	A1 *	9/2007	Sudo et al.	429/13
2007/0223171	A1	9/2007	Guy et al.	361/126
2007/0273206	A1 *	11/2007	McCowen	307/9.1
2007/0297892	A1 *	12/2007	Kildegaard	415/1
2008/0145714	A1 *	6/2008	Kagami	429/13
2009/0107842	A1	4/2009	Park et al.	204/554
2009/0114495	A1 *	5/2009	McCowen	191/45 R
2009/0216292	A1 *	8/2009	Pless et al.	607/33
2010/0090562	A1 *	4/2010	McCowen	310/309
2010/0090563	A1 *	4/2010	McCowen	310/309
2012/0164555	A1 *	6/2012	Yamazaki et al.	429/482
2012/0299559	A1 *	11/2012	McCowen	322/2 A
2013/0276776	A1 *	10/2013	Te Kamp et al.	126/660

* cited by examiner

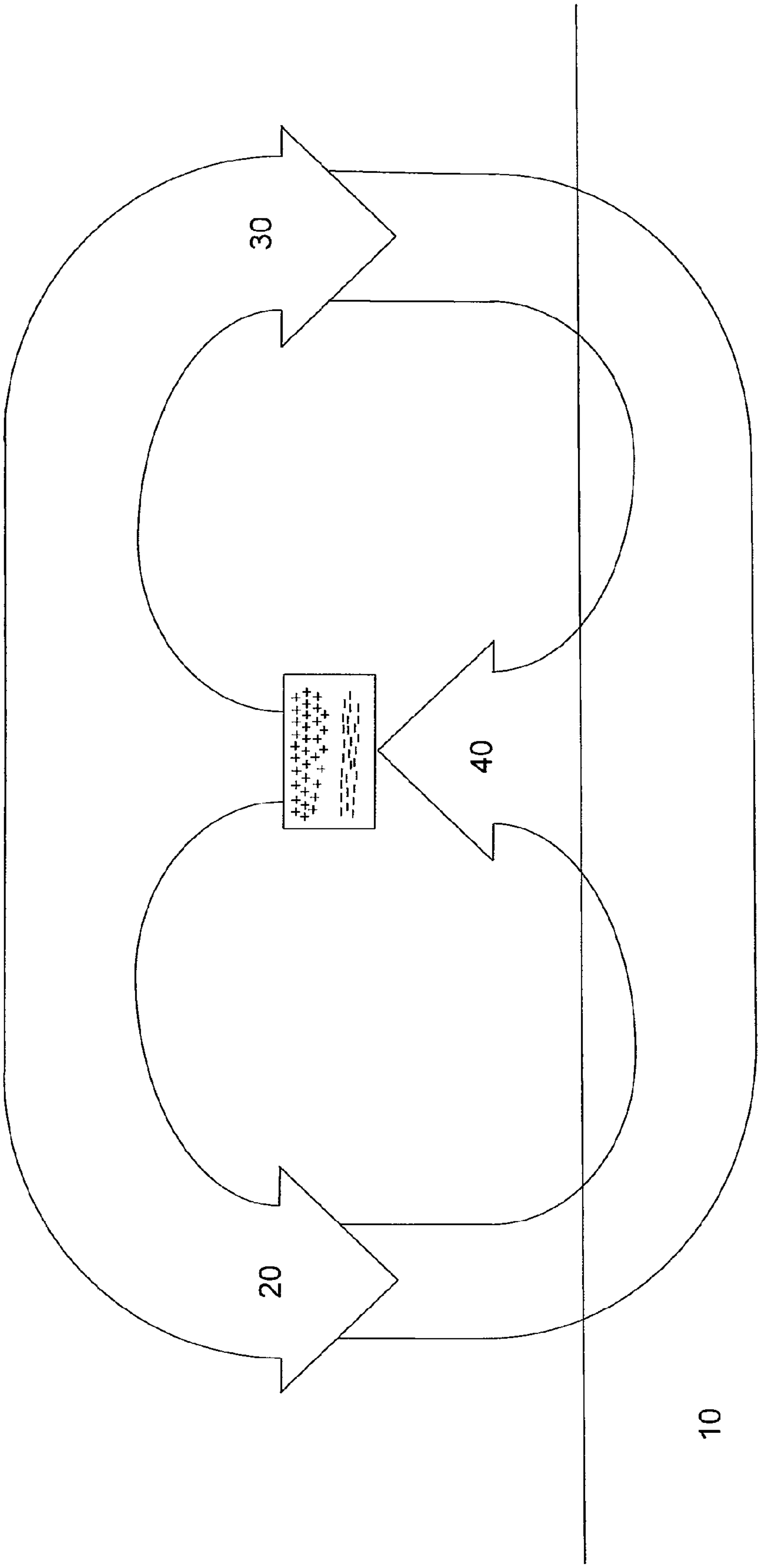

FIGURE 1

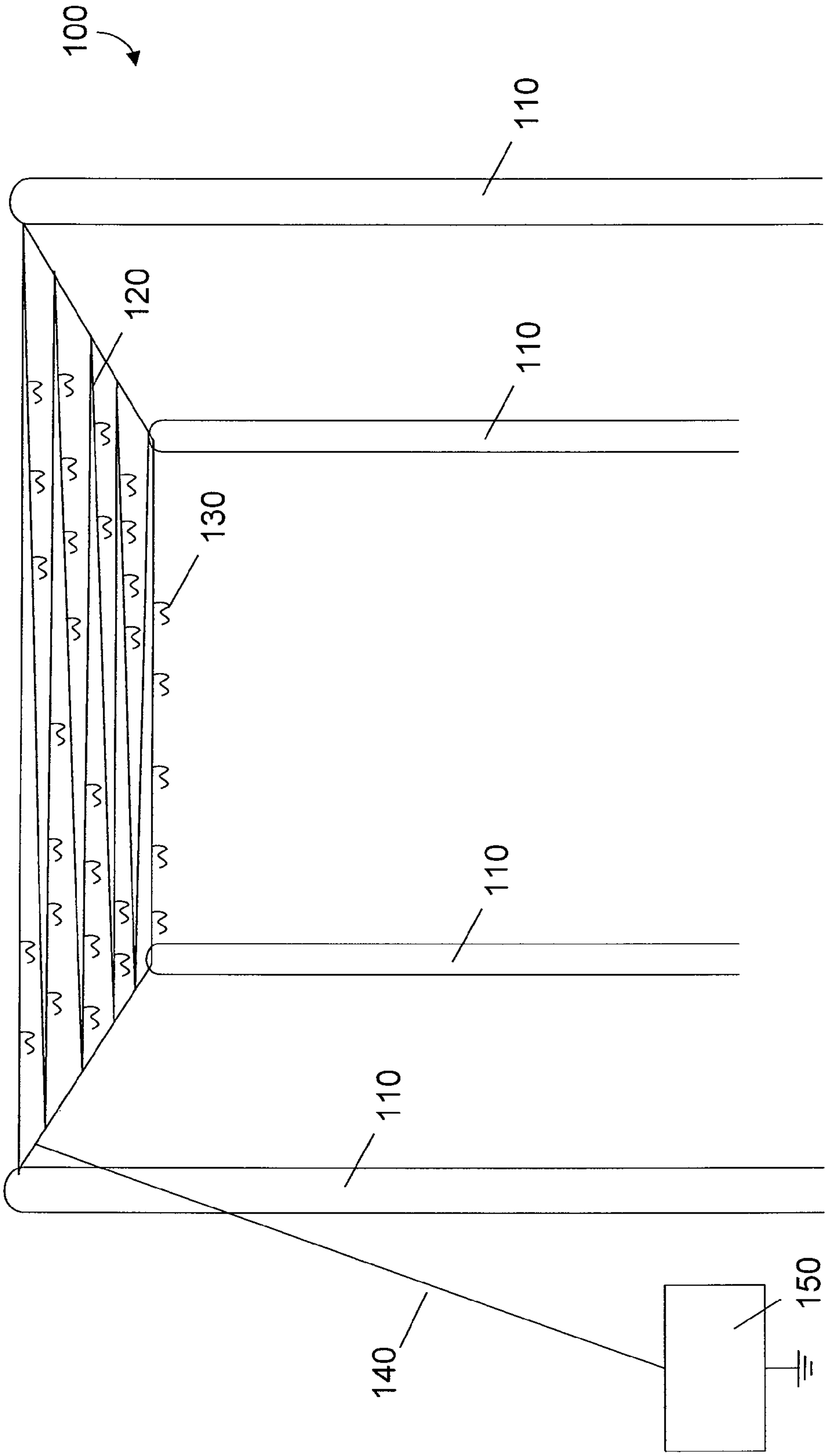
100
110
110
120
130
110
110
140
150

FIGURE 2

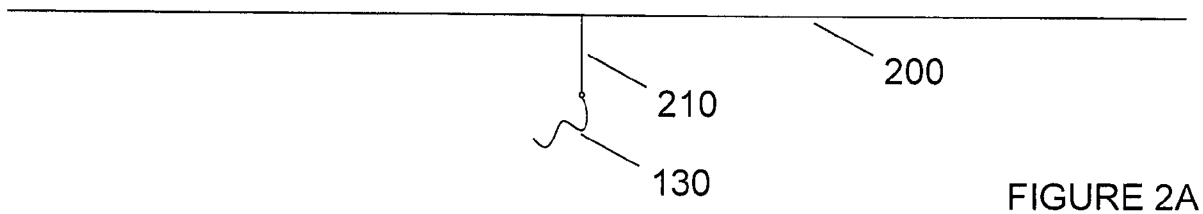

FIGURE 2A

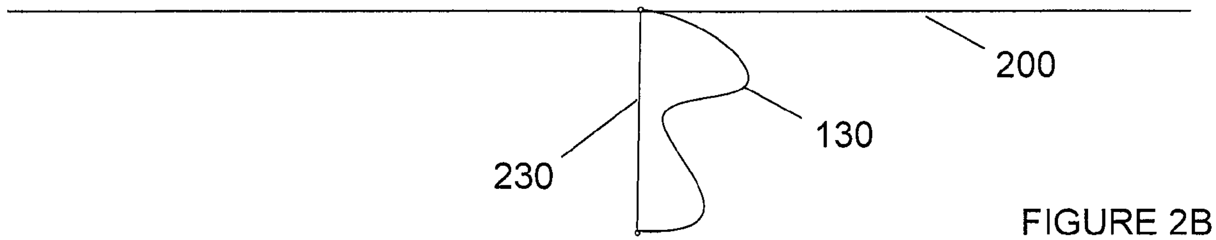

FIGURE 2B

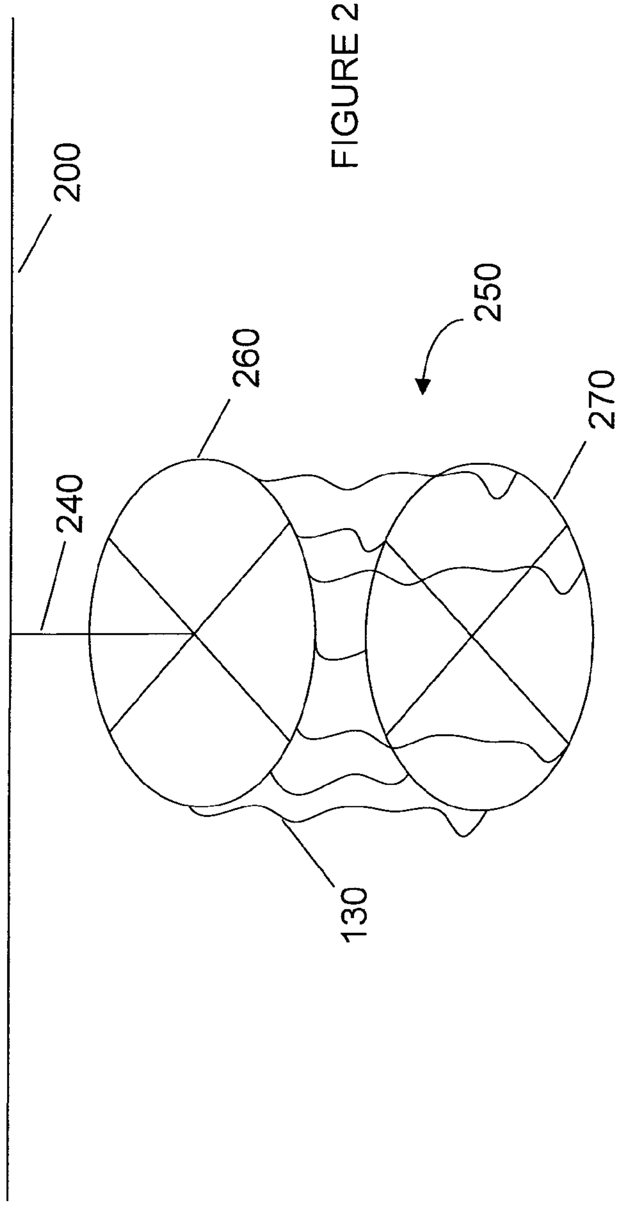
200
250
260
240
270
130

FIGURE 2C

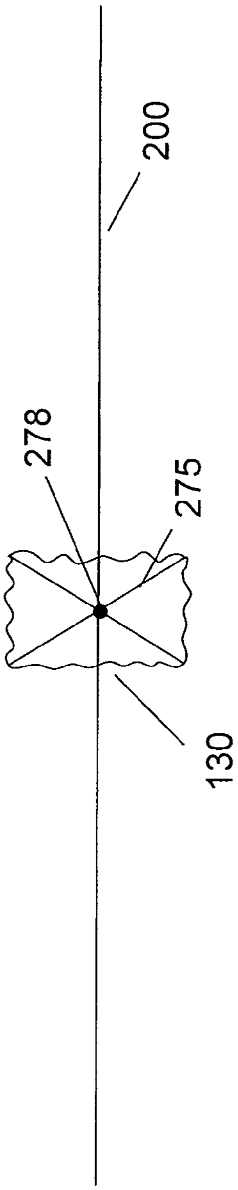
278
275
200
130

FIGURE 2D

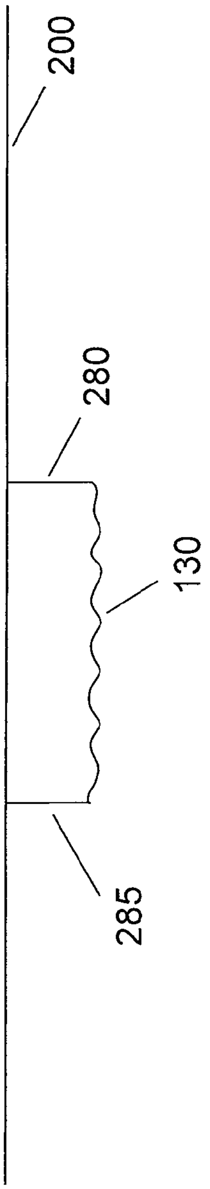
200
280
130
285

FIGURE 2E

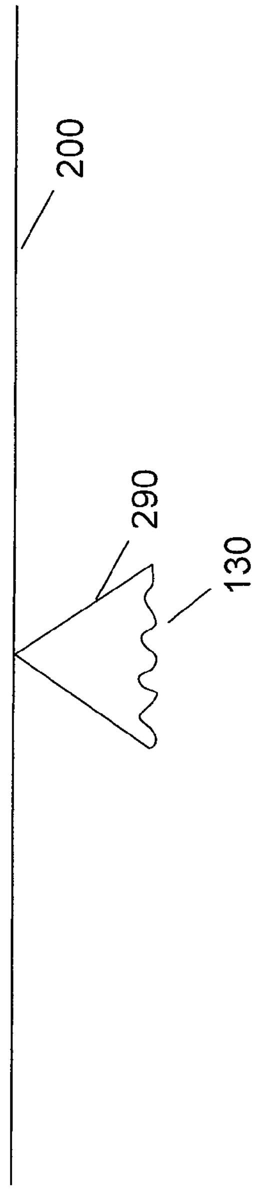

FIGURE 2F

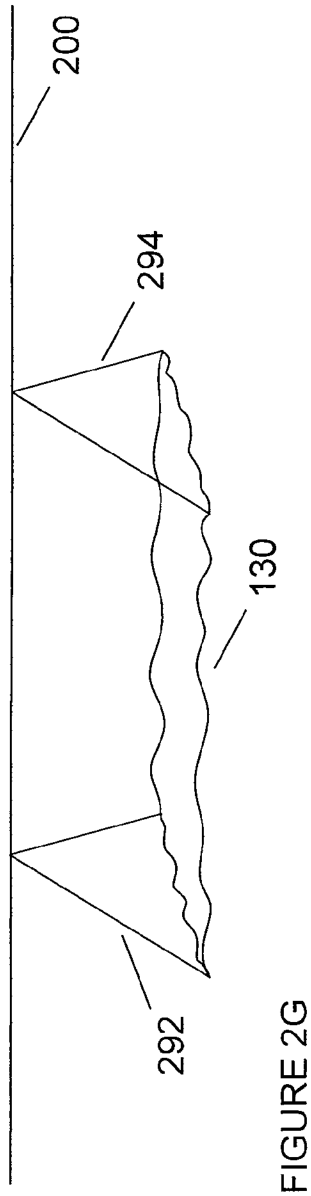
200
294
130
292

FIGURE 2G

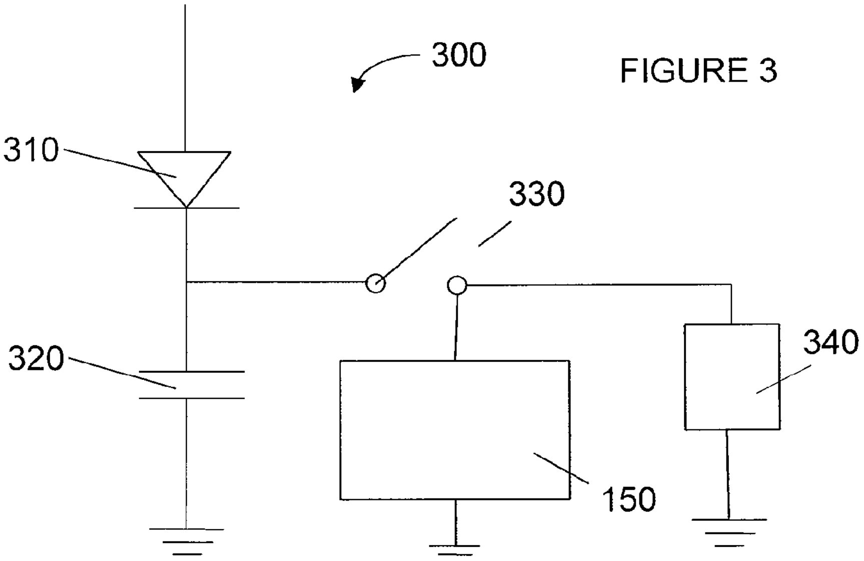
300
310
330
320
340
150

FIGURE 3

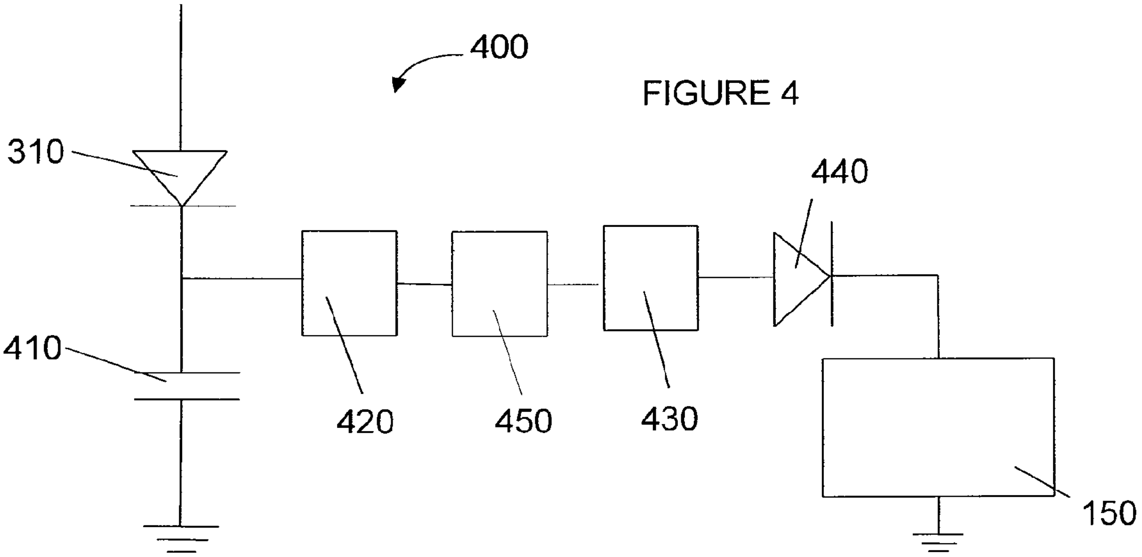
400
310
440
410
420
450
430
150

FIGURE 4

FIGURE 5

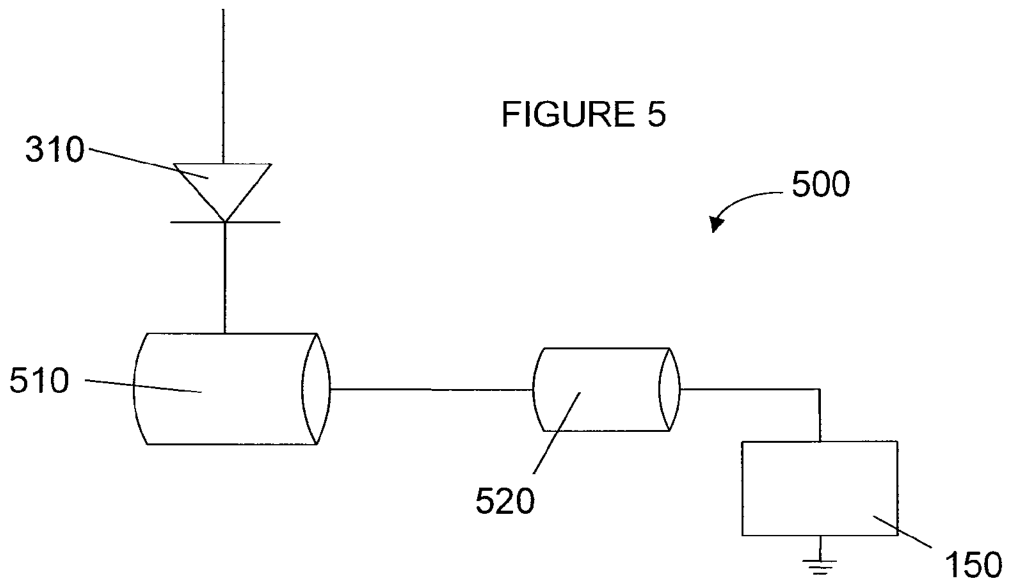
310
500
510
520
150

FIGURE 6

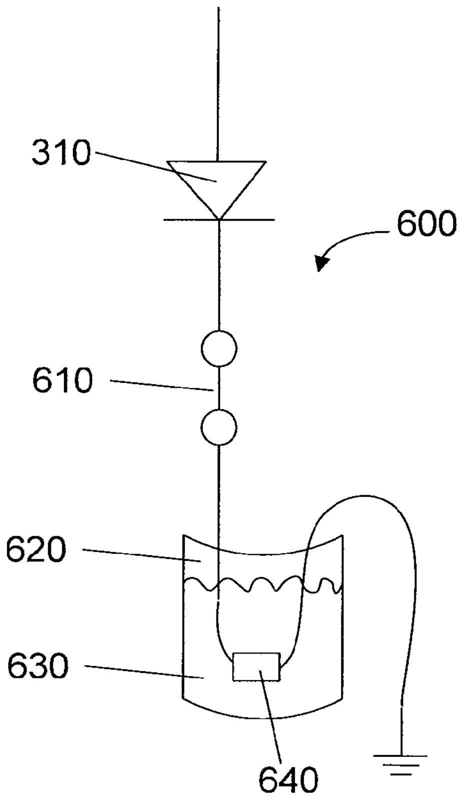
310
600
610
620
630
640

FIGURE 7

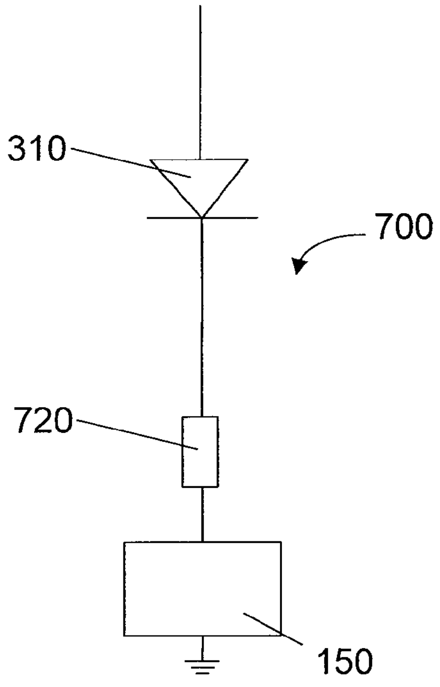
310
700
720
150

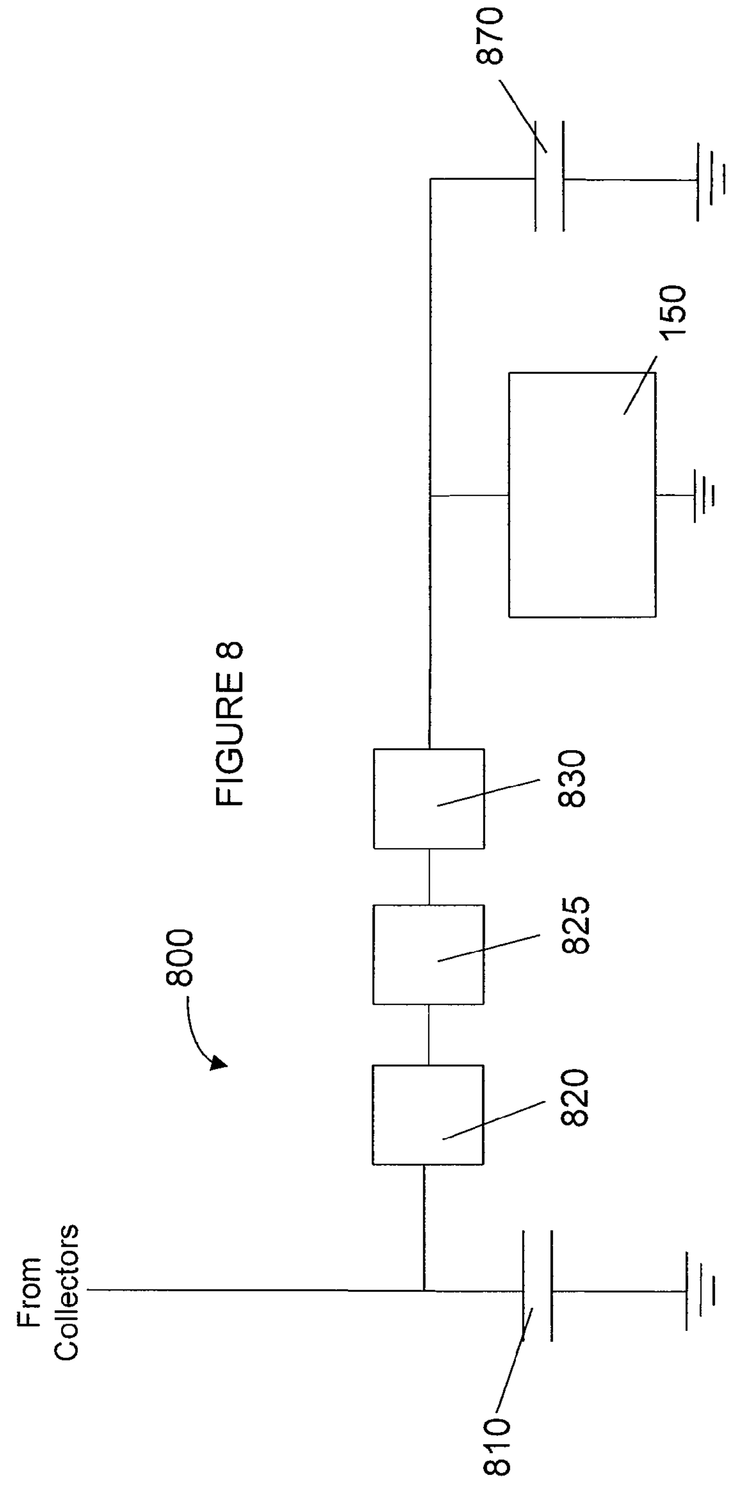
From Collectors
810
820
825
830
800
150
870

FIGURE 8

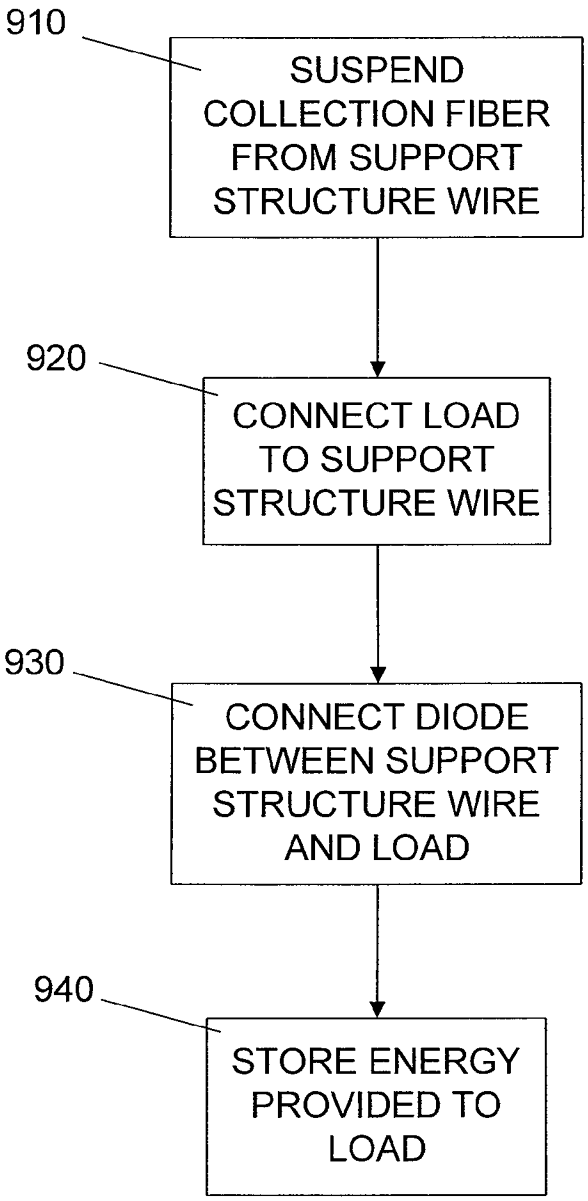
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

ENERGY COLLECTION

CROSS REFERENCE TO RELATED APPLICATION

This application is a continuation application of U.S. patent application Ser. No. 11/358,264, filed on Feb. 21, 2006, which is incorporated by reference herein.

TECHNICAL FIELD

The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure wire elevated above a ground level, at least one collection fiber electrically connected to the support structure wire; a load electrically connected to the support structure wire; and a diode electrically connected between the load and at least one collection fiber.

Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection fiber from a support structure wire elevated above ground level, the fiber electrically connected to the support structure wire; providing a load with an electrical connection to the support structure wire to draw current; and providing a diode electrically connected between the collection fiber and the load.

Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. **1** is a circuit diagram of a weather energy circuit.

FIG. **2** is a perspective view of an exemplary embodiment of many energy collectors elevated above ground by a structure.

FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

FIG. **2**B is a side view of an exemplary embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

FIG. **2**D is a side view of an exemplary embodiment of a support structure for multiple energy collection fibers.

FIG. **2**E is a side view of a support structure for an energy collection fiber.

FIG. **2**F is a side view of an exemplary embodiment of a support structure for an energy collection fiber.

FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

FIG. **3** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **4** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **5** is a circuit diagram of an exemplary embodiment of an energy collection circuit for powering a generator and motor.

FIG. **6** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

FIG. **7** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

FIG. **8** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy.

FIG. **9** is a flow diagram of an exemplary embodiment of collecting energy with a collection fiber.

DETAILED DESCRIPTION

Electric charges on conductors reside entirely on the external surface of the conductors, and tend to concentrate more

around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An exemplary embodiment of this disclosure takes advantage of this property, among others, to collect and use the energy generated by an electric field in the atmosphere. Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An exemplary embodiment utilizes carbon or graphite fibers for static electricity collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including aluminum or steel, but most notably, copper. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

An exemplary embodiment of the collection fibers as collection device **130** includes graphite or carbon fibers. Graphite and carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, or graphite fibers are examples that demonstrate exemplary collection ability.

In at least one exemplary embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than 100 volts per meter under some conditions. When support wire **120** is suspended in the air at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An exemplary embodiment of the collection device comprises the diode with no collection fiber. A preferred embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restricts energy that may be collected through one fiber from escaping through another fiber.

Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2A-2G** using a collection fiber embodiment. FIG. **2A** presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**.

Likewise, FIG. **2B** shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

FIG. **2C** presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

FIG. **2D** presents another exemplary embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2E** provides another exemplary embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

FIG. **2F** presents another exemplary embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2G** provides two supports as provided in FIG. **2F** such that at least two support members **292**, **294** may be connected

to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and between multiple support structures. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**.

FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load. Capacitor **320** may be connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

In an exemplary embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel. Non-limiting examples of spark gap **420** include mercury-reed switches and mercury-wetted reed switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground. The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes may be placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches and mercury-wetted reed switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel.

FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit. Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

FIG. **8** presents exemplary circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches and mercury-wetted reed switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

It should be emphasized that the above-described embodiments of the present disclosure, particularly, any "preferred" embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A method of collecting energy comprising:

suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to the support structure;

providing a load with an electrical connection to the at least one collection device to passively draw current; and

powering a fuel cell between the support structure and the load, the powering performed with energy collected from the at least one collection device.

2. The method of claim **1**, wherein the collection device comprises a diode.

3. The method of claim **1**, wherein the collection device comprises a collection fiber.

4. The method of claim **1**, wherein the collection device comprises a diode and a collection fiber and the diode is electrically connected between the collection fiber and the load.

5. The method of claim **1**, further comprising storing energy provided to the load.

6. The method of claim **5**, wherein storing energy provided to the load comprises storing energy in a capacitor or an inductor.

7. The method of claim **3**, wherein the collection fiber comprises carbon fiber or graphite fiber.

8. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the collection device electrically connected to and suspended from the support structure, the at least one collection device configured to passively draw current;

a load electrically connected to the at least one collection device; and

a diode electrically connected between the at least one collection device and the support structure.

9. The system of claim **8**, wherein the collection device comprises a diode.

10. The system of claim **8**, wherein the collection device comprises a collection fiber.

11. The system of claim **8**, wherein the collection device comprises a collection fiber and a diode electrically connected between the load and the collection fiber.

12. The system of claim **9**, wherein the diode is elevated relative to the ground level.

13. The system of claim **10**, wherein the collection fiber comprises a carbon fiber or a graphite fiber.

14. The system of claim **8**, further comprising:

a switch connected in series between the at least one collection device and the load; and

a capacitor connected in parallel with the switch and the load.

15. The system of claim **14**, wherein the switch comprises an interrupter connected between the at least one collection device and the load, and wherein the interrupter comprises at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

16. The system of claim **15**, further comprising a transformer connected between the interrupter and the load.

17. The system of claim **8**, further comprising:

a motor for providing power, the motor connected between the at least one collection device and the load; and

a generator powered by the motor.

18. The system of claim **8**, further comprising a fuel cell between the support structure and the load.

19. The system of claim **18**, wherein the fuel cell produces hydrogen and oxygen.

20. A system of collecting energy comprising:

means for suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to the means for suspending;

means for passively inducing current flow, the means for inducing current flow electrically connected to the means for suspending; and

means for restricting the backflow of charge carriers, the means for restricting the backflow of charge carriers electrically connected between the at least one collection device and the means for inducing current flow.

21. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the collection device electrically connected to and suspended from the support structure, the at least one collection device configured to passively draw current;

a load electrically connected to the at least one collection device; and

a switch connected in series between the at least one collection device and the load, the switch comprising an interrupter connected between the at least one collection device and the load, and wherein the interrupter comprises at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

22. The system of claim **21**, further comprising a transformer connected between the interrupter and the load.

23. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to and suspended from the support structure, the collection device configured to passively draw current;

a load electrically connected to the at least one collection device;

a motor for providing power, the motor connected between the at least one collection device and the load; and

a generator powered by the motor.

24. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to and suspended from the support structure the at least one collection device configured to passively draw current;

a load electrically connected to the at least one collection device; and

a fuel cell between the support structure and the load.

25. The system of claim **24**, wherein the fuel cell produces hydrogen and oxygen.

* * * * *

US008686575B2

(12) United States Patent
McCowen

(10) **Patent No.: US 8,686,575 B2**
(45) **Date of Patent: Apr. 1, 2014**

(54) **ENERGY COLLECTION**

(75) Inventor: **Clint McCowen**, Navarre, FL (US)

(73) Assignee: **Ion Power Group, LLC**, Navarre, FL (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **12/255,130**

(22) Filed: **Oct. 21, 2008**

(65) **Prior Publication Data**

US 2009/0040680 A1 Feb. 12, 2009

Related U.S. Application Data

(63) Continuation of application No. 11/358,264, filed on Feb. 21, 2006, now Pat. No. 7,439,712.

(51) **Int. Cl.**

F02B 63/04	(2006.01)
F03G 7/08	(2006.01)
H02K 7/18	(2006.01)
H02N 1/00	(2006.01)

(52) **U.S. Cl.**
USPC .... **290/1 R**; 322/2 A

(58) **Field of Classification Search**
USPC .... 290/1 R; 322/2 A; 126/660; 60/641.8, 60/641.11, 641.12, 659, 698; 244/158.2
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427 A		5/1901	Palencsar	307/149
911,260 A		2/1909	Pennock	307/149
1,014,719 A		1/1912	Pennock	307/149
2,473,819 A	*	6/1949	Pittman	315/36
3,532,959 A	*	10/1970	Erickson	320/166
3,780,722 A	*	12/1973	Swet	126/680
3,946,227 A		3/1976	Bingham	250/281
4,079,225 A	*	3/1978	Warner	219/110
4,092,250 A		5/1978	Sano et al.	210/500.29
4,104,696 A		8/1978	Cochran, Jr.	361/229
4,146,800 A	*	3/1979	Gregory et al.	290/44
4,180,056 A	*	12/1979	Schnabel et al.	126/660
4,201,197 A	*	5/1980	Dismer	126/600
4,206,396 A	*	6/1980	Marks	322/2 A
4,224,496 A	*	9/1980	Riordan et al.	219/110
4,307,936 A	*	12/1981	Ochiai	385/120
4,314,192 A	*	2/1982	Caro	322/2 R
4,346,478 A	*	8/1982	Sichling	398/107
4,425,905 A	*	1/1984	Mori	126/578
4,433,248 A	*	2/1984	Marks	290/44
RE31,678 E	*	9/1984	Ochiai	385/115

(Continued)

FOREIGN PATENT DOCUMENTS

WO	2007098341	8/2007

OTHER PUBLICATIONS

"Emissivity Values for Common Materials & Non-Metals", Raytek—A Fluke Company, http://www.raytek.com/Raytek/en-r0/IREducation/EmissivityNonMetals.htm, 1999.*

Primary Examiner — Pedro J Cuevas
(74) *Attorney, Agent, or Firm* — Benjamin A. Balser; Balser & Grell IP Law

(57) **ABSTRACT**

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support wire system supported by poles. The support wire system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but carbon and graphite are preferred. Diodes may be used to restrict the backflow or loss of energy.

25 Claims, 10 Drawing Sheets

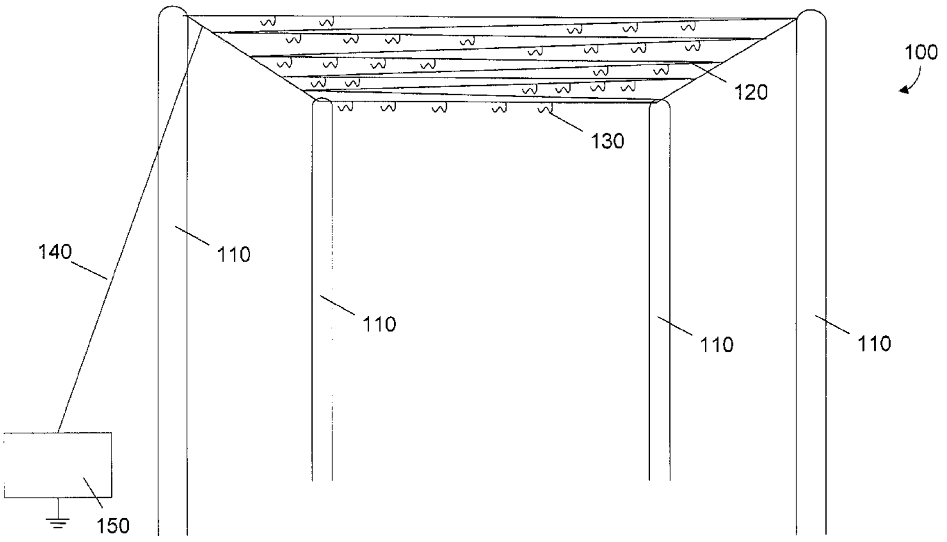

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,483,311	A *	11/1984	Whitaker	126/602
4,489,269	A *	12/1984	Edling et al.	322/2 R
4,512,335	A *	4/1985	Mori	126/678
4,653,223	A *	3/1987	Mori	47/17
4,653,472	A *	3/1987	Mori	126/683
4,676,226	A *	6/1987	Mori	126/680
4,809,675	A *	3/1989	Mori	126/578
4,852,454	A *	8/1989	Batchelder	89/1.11
4,943,125	A *	7/1990	Laundre' et al.	385/35
4,988,159	A *	1/1991	Turner et al.	385/33
5,047,892	A	9/1991	Sakata et al.	361/231
5,114,101	A *	5/1992	Stern et al.	244/172.8
5,145,257	A *	9/1992	Bryant et al.	374/131
5,379,103	A *	1/1995	Zigler	356/73
5,851,309	A *	12/1998	Kousa	136/248
5,942,806	A *	8/1999	Veliadis	290/1 R
6,025,591	A	2/2000	Taylor et al.	250/292
6,038,363	A *	3/2000	Slater et al.	385/147
6,116,544	A *	9/2000	Forward et al.	244/158.2
6,173,922	B1 *	1/2001	Hoyt et al.	244/158.2
6,226,440	B1 *	5/2001	Lyons	385/119
6,419,191	B1 *	7/2002	Hoyt et al.	244/158.2
6,425,391	B1 *	7/2002	Davoren et al.	126/683
6,735,830	B1	5/2004	Merciel	250/324
6,765,212	B2 *	7/2004	Goetz et al.	250/339.11
6,846,447	B2 *	1/2005	Mbachu et al.	264/406
6,853,447	B2 *	2/2005	Goetz	356/237.1
6,894,772	B2 *	5/2005	Goetz et al.	356/237.1
6,925,852	B2 *	8/2005	Susko	73/23.2
6,974,110	B2	12/2005	Grandics	244/171.5
7,109,597	B1 *	9/2006	Bose	290/43
7,128,867	B2 *	10/2006	Mbachu et al.	264/406
7,231,809	B2 *	6/2007	Susko	73/31.05
7,259,475	B2 *	8/2007	Hong et al.	307/64
7,439,712	B2 *	10/2008	McCowen	322/2 A
7,449,668	B2 *	11/2008	Schutten et al.	250/214 SW
7,478,712	B2 *	1/2009	McCowen	191/45 R
7,532,819	B1 *	5/2009	Triebes et al.	398/121
2002/0026933	A1 *	3/2002	Gottlieb	126/628
2002/0108892	A1 *	8/2002	Goetz et al.	209/576
2002/0109094	A1 *	8/2002	Goetz et al.	250/339.11
2002/0109835	A1 *	8/2002	Goetz	356/237.1
2003/0107869	A1	6/2003	Pan et al.	361/306.2
2003/0125630	A1 *	7/2003	Furnish	600/461
2003/0193319	A1 *	10/2003	Wood et al.	322/2 A
2004/0011925	A1 *	1/2004	Grandics	244/62
2004/0083793	A1 *	5/2004	Susko	73/31.05
2004/0094853	A1 *	5/2004	Mbachu et al.	264/40.1
2004/0160711	A1	8/2004	Stumberger	361/20
2005/0101024	A1 *	5/2005	Mbachu et al.	436/85
2005/0136311	A1 *	6/2005	Ueda et al.	429/30
2005/0270525	A1 *	12/2005	Susko	356/326
2007/0107765	A1 *	5/2007	Schutten et al.	136/243
2007/0195481	A1 *	8/2007	McCowen	361/233
2007/0218323	A1 *	9/2007	Sudo et al.	429/13
2007/0223171	A1	9/2007	Guy et al.	361/126
2007/0273206	A1 *	11/2007	McCowen	307/9.1
2007/0297892	A1 *	12/2007	Kildegaard	415/1
2008/0145714	A1 *	6/2008	Kagami	429/13
2009/0107842	A1	4/2009	Park et al.	204/554
2009/0114495	A1 *	5/2009	McCowen	191/45 R
2009/0216292	A1 *	8/2009	Pless et al.	607/33
2010/0090562	A1 *	4/2010	McCowen	310/309
2010/0090563	A1 *	4/2010	McCowen	310/309
2012/0164555	A1 *	6/2012	Yamazaki et al.	429/482
2012/0299559	A1 *	11/2012	McCowen	322/2 A
2013/0276776	A1 *	10/2013	Te Kamp et al.	126/660

* cited by examiner

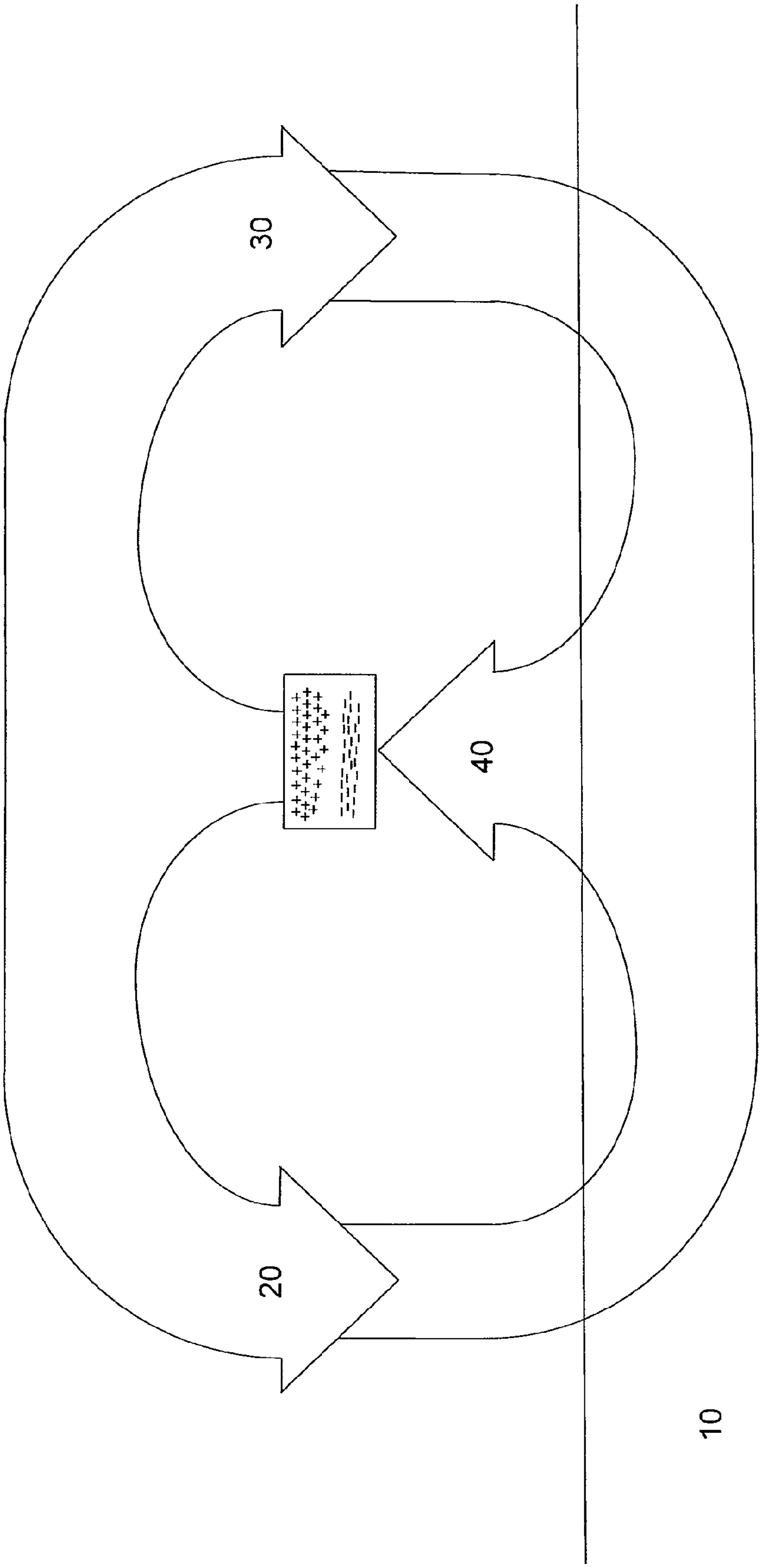

FIGURE 1

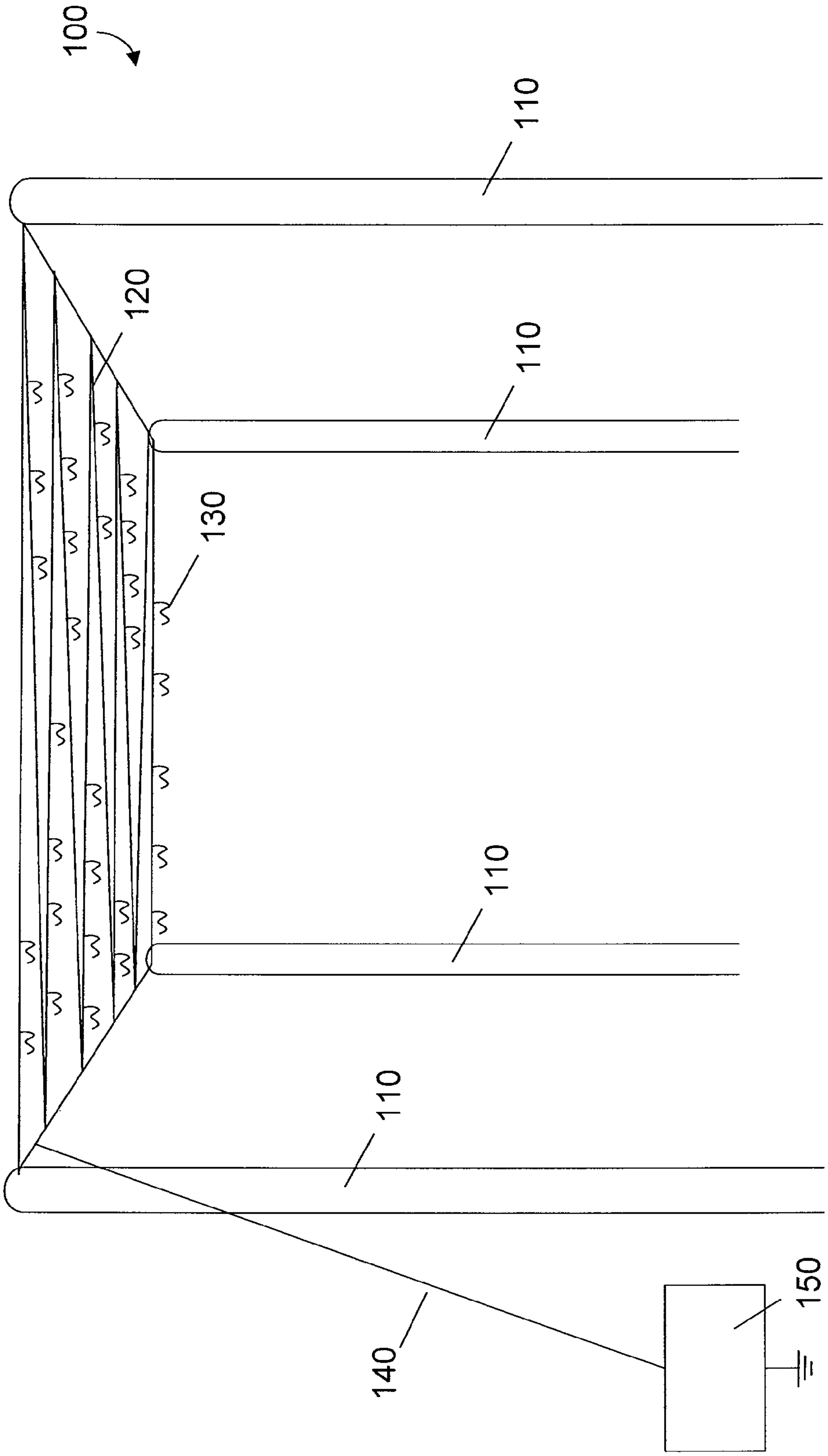
100
110
110
110
110
120
130
140
150

FIGURE 2

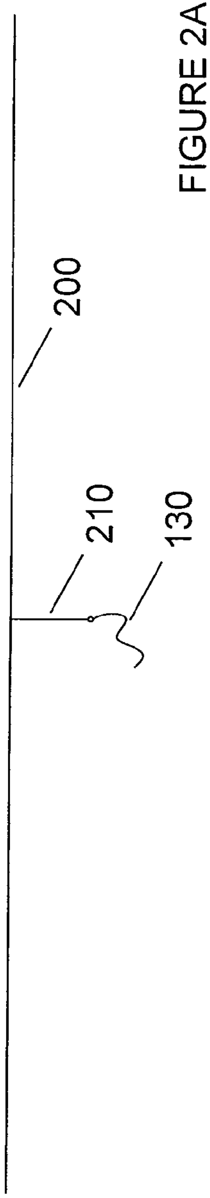

FIGURE 2A

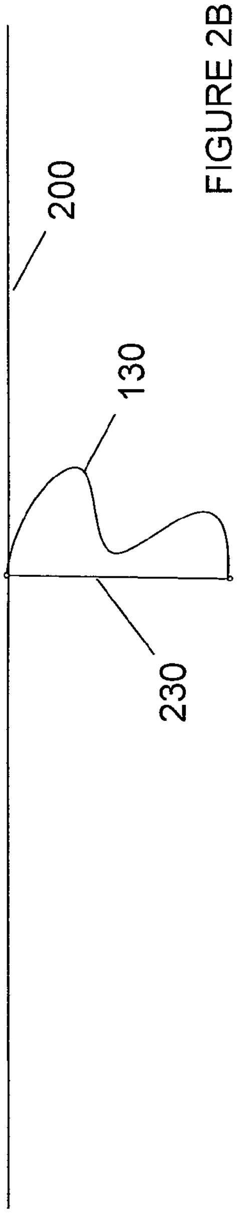

FIGURE 2B

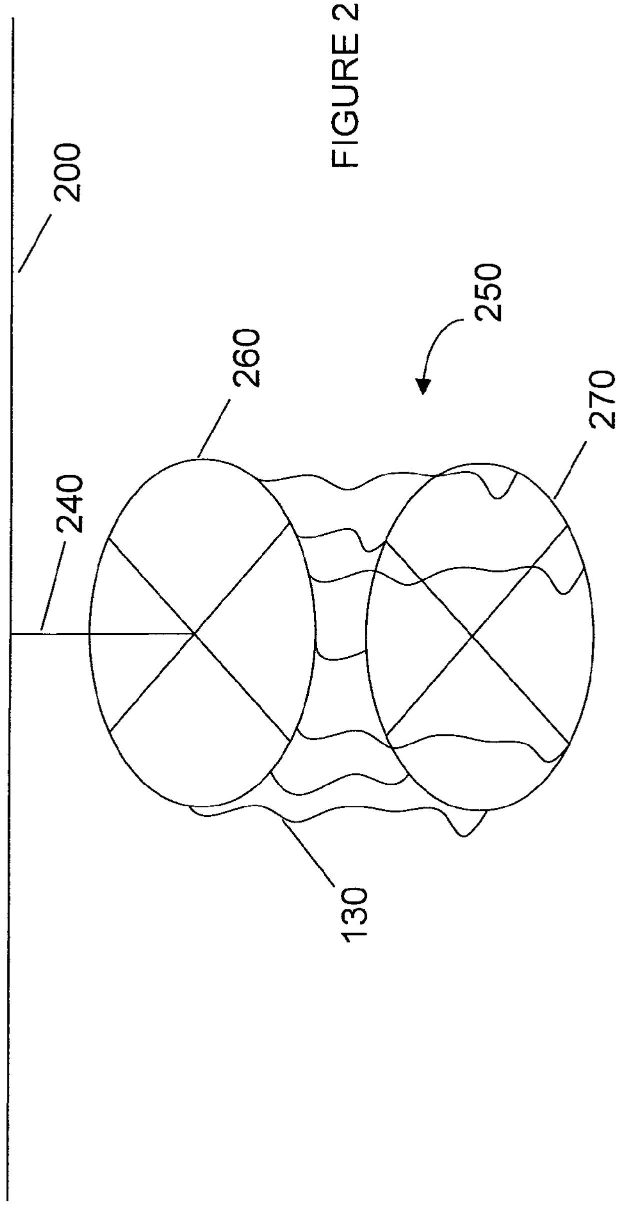
200
250
260
270
240
130

FIGURE 2C

FIGURE 2D

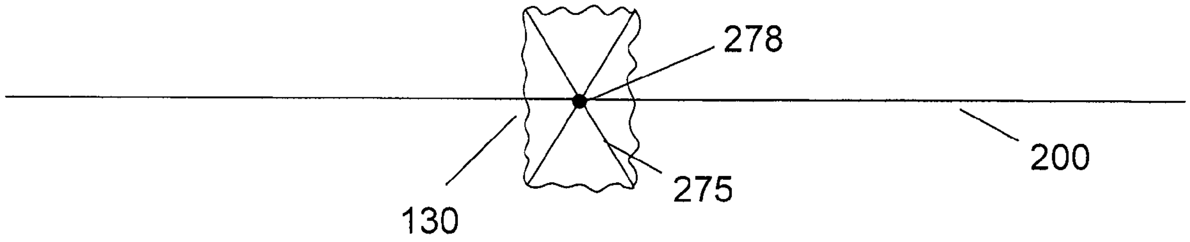
278
200
275
130

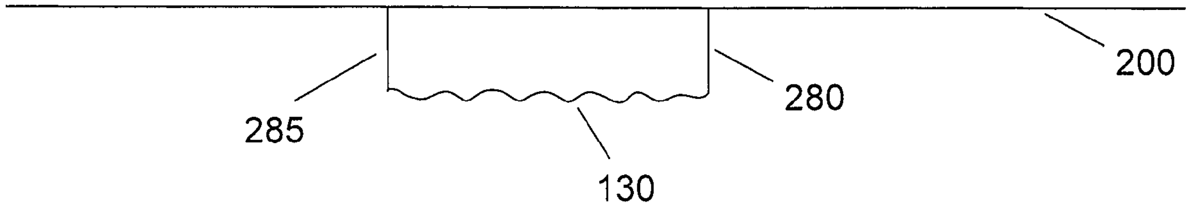
200
280
285
130

FIGURE 2E

FIGURE 2F

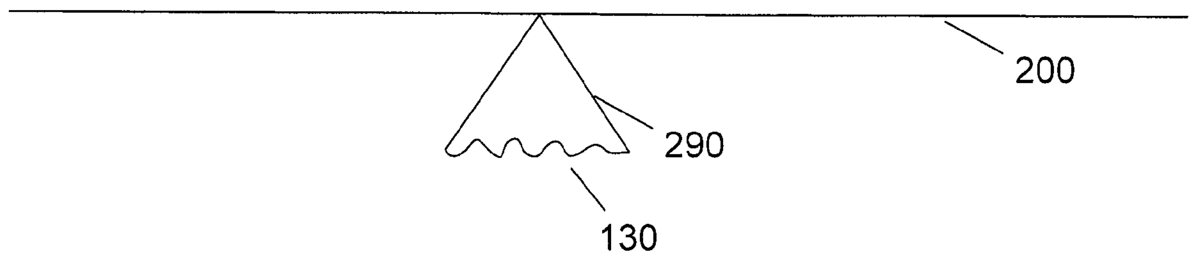

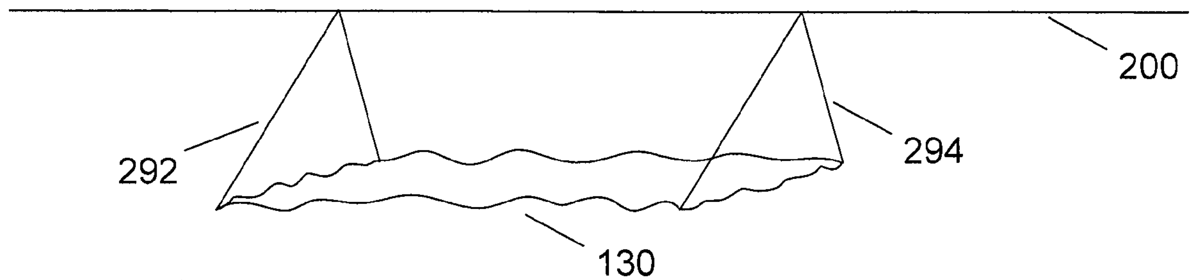
200
292
294
130

FIGURE 2G

FIGURE 3

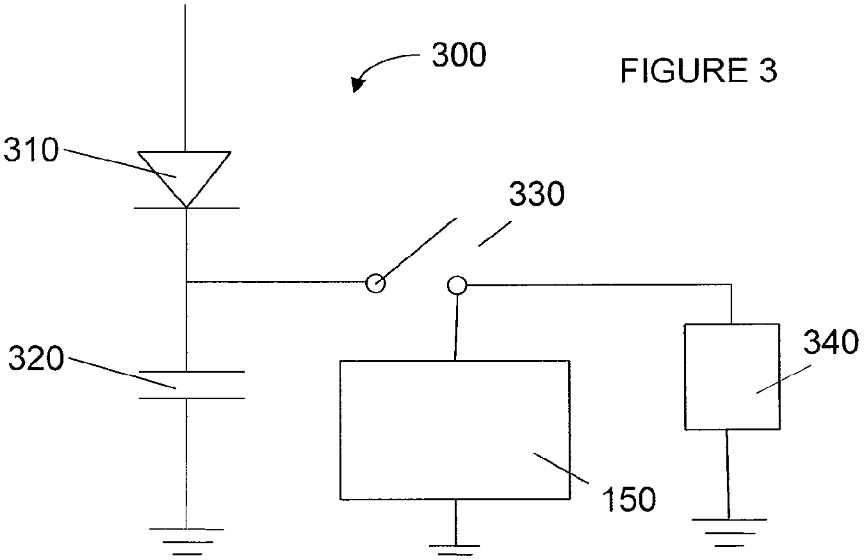
300
310
330
320
340
150

FIGURE 4

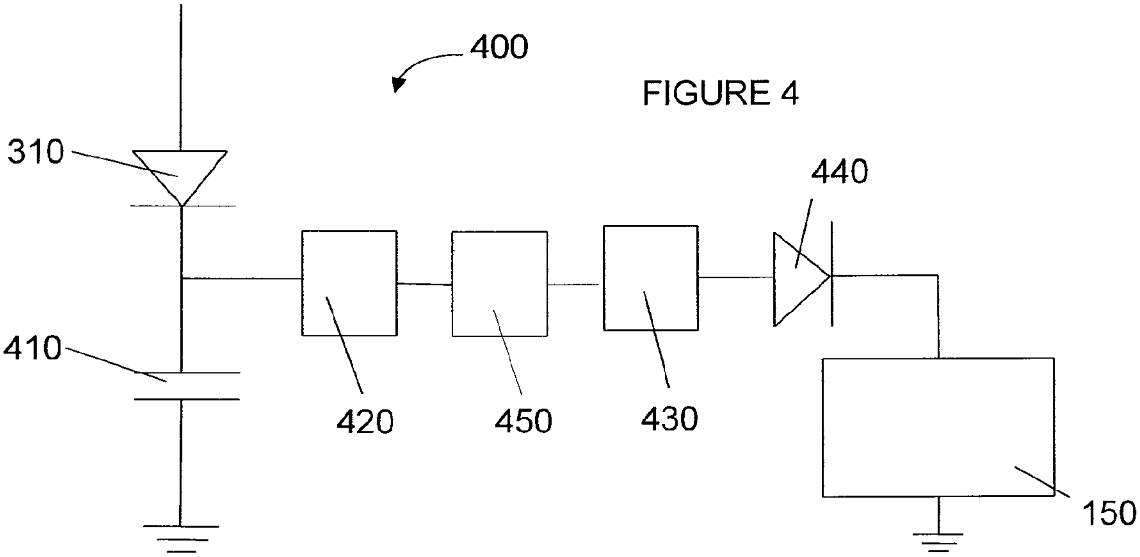
400
310
440
410
420
450
430
150

FIGURE 5

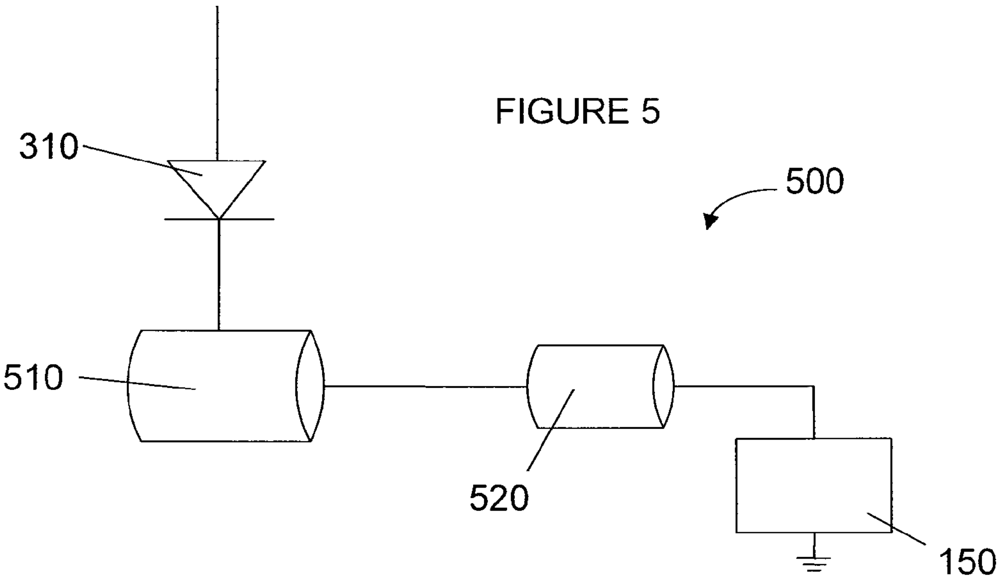
310
500
510
520
150

FIGURE 6

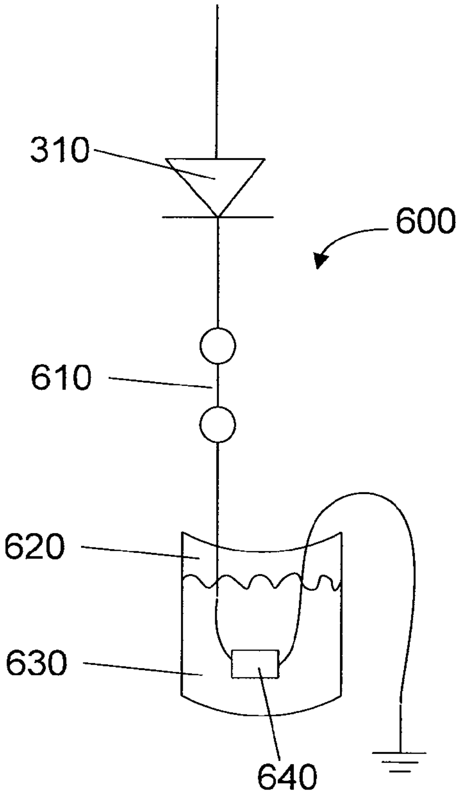
310
600
610
620
630
640

FIGURE 7

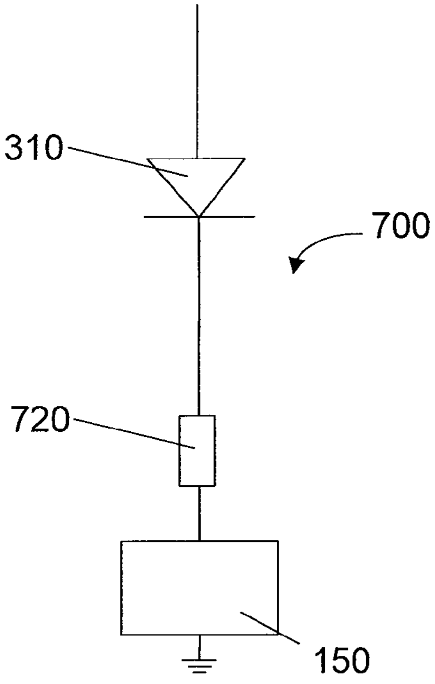
310
700
720
150

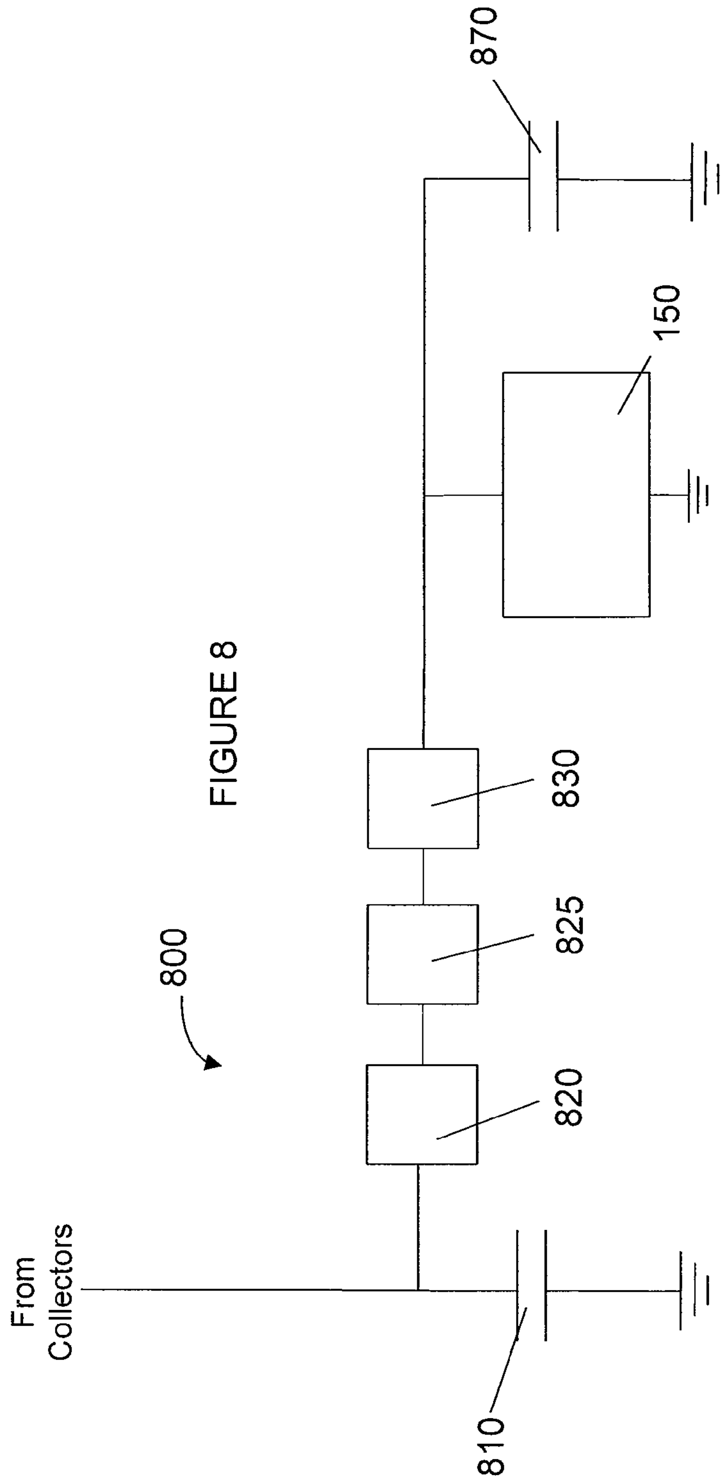
870
150
830
825
820
800
From
Collectors
810

FIGURE 8

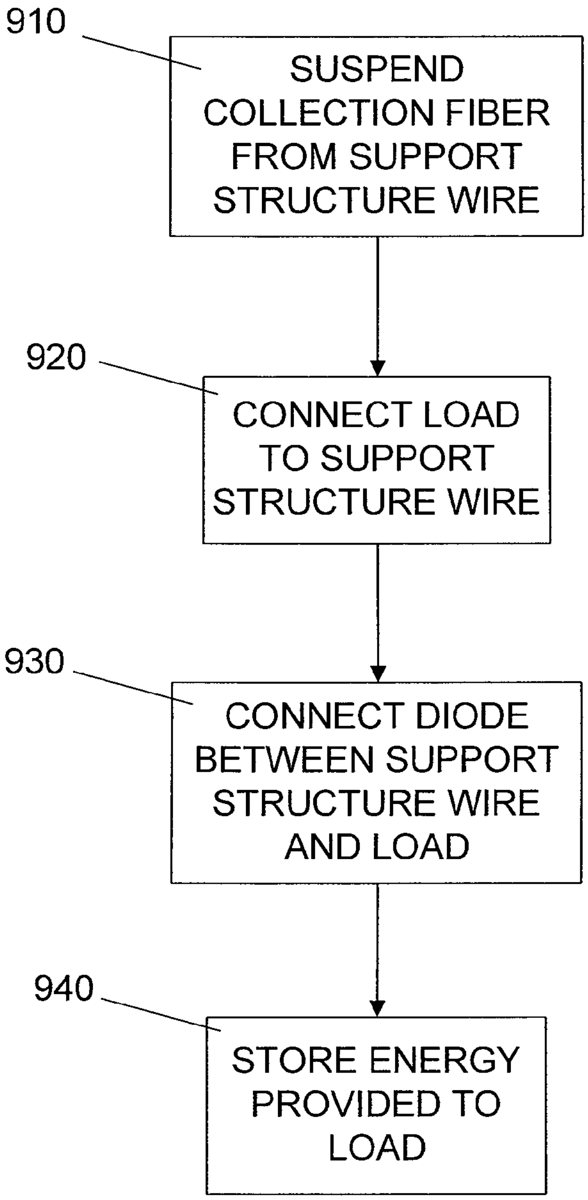
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

ENERGY COLLECTION

CROSS REFERENCE TO RELATED APPLICATION

This application is a continuation application of U.S. patent application Ser. No. 11/358,264, filed on Feb. 21, 2006, which is incorporated by reference herein.

TECHNICAL FIELD

The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure wire elevated above a ground level, at least one collection fiber electrically connected to the support structure wire; a load electrically connected to the support structure wire; and a diode electrically connected between the load and at least one collection fiber.

Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection fiber from a support structure wire elevated above ground level, the fiber electrically connected to the support structure wire; providing a load with an electrical connection to the support structure wire to draw current; and providing a diode electrically connected between the collection fiber and the load.

Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. **1** is a circuit diagram of a weather energy circuit.

FIG. **2** is a perspective view of an exemplary embodiment of many energy collectors elevated above ground by a structure.

FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

FIG. **2**B is a side view of an exemplary embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

FIG. **2**D is a side view of an exemplary embodiment of a support structure for multiple energy collection fibers.

FIG. **2**E is a side view of a support structure for an energy collection fiber.

FIG. **2**F is a side view of an exemplary embodiment of a support structure for an energy collection fiber.

FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

FIG. **3** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **4** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **5** is a circuit diagram of an exemplary embodiment of an energy collection circuit for powering a generator and motor.

FIG. **6** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

FIG. **7** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

FIG. **8** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy.

FIG. **9** is a flow diagram of an exemplary embodiment of collecting energy with a collection fiber.

DETAILED DESCRIPTION

Electric charges on conductors reside entirely on the external surface of the conductors, and tend to concentrate more

around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An exemplary embodiment of this disclosure takes advantage of this property, among others, to collect and use the energy generated by an electric field in the atmosphere. Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An exemplary embodiment utilizes carbon or graphite fibers for static electricity collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including aluminum or steel, but most notably, copper. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

An exemplary embodiment of the collection fibers as collection device **130** includes graphite or carbon fibers. Graphite and carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, or graphite fibers are examples that demonstrate exemplary collection ability.

In at least one exemplary embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than 100 volts per meter under some conditions. When support wire **120** is suspended in the air at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An exemplary embodiment of the collection device comprises the diode with no collection fiber. A preferred embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restricts energy that may be collected through one fiber from escaping through another fiber.

Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2A-2G** using a collection fiber embodiment. FIG. **2A** presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**.

Likewise, FIG. **2B** shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

FIG. **2C** presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

FIG. **2D** presents another exemplary embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2E** provides another exemplary embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

FIG. **2F** presents another exemplary embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2G** provides two supports as provided in FIG. **2F** such that at least two support members **292**, **294** may be connected

to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and between multiple support structures. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**.

FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load. Capacitor **320** may be connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

In an exemplary embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel. Non-limiting examples of spark gap **420** include mercury-reed switches and mercury-wetted reed switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground. The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes may be placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches and mercury-wetted reed switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel.

FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit. Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

FIG. **8** presents exemplary circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches and mercury-wetted reed switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

It should be emphasized that the above-described embodiments of the present disclosure, particularly, any "preferred" embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A method of collecting energy comprising:

suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to the support structure;

providing a load with an electrical connection to the at least one collection device to passively draw current; and

powering a fuel cell between the support structure and the load, the powering performed with energy collected from the at least one collection device.

2. The method of claim **1**, wherein the collection device comprises a diode.

3. The method of claim **1**, wherein the collection device comprises a collection fiber.

4. The method of claim **1**, wherein the collection device comprises a diode and a collection fiber and the diode is electrically connected between the collection fiber and the load.

5. The method of claim **1**, further comprising storing energy provided to the load.

6. The method of claim **5**, wherein storing energy provided to the load comprises storing energy in a capacitor or an inductor.

7. The method of claim **3**, wherein the collection fiber comprises carbon fiber or graphite fiber.

8. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the collection device electrically connected to and suspended from the support structure, the at least one collection device configured to passively draw current;

a load electrically connected to the at least one collection device; and

a diode electrically connected between the at least one collection device and the support structure.

9. The system of claim **8**, wherein the collection device comprises a diode.

10. The system of claim **8**, wherein the collection device comprises a collection fiber.

11. The system of claim **8**, wherein the collection device comprises a collection fiber and a diode electrically connected between the load and the collection fiber.

12. The system of claim **9**, wherein the diode is elevated relative to the ground level.

13. The system of claim **10**, wherein the collection fiber comprises a carbon fiber or a graphite fiber.

14. The system of claim **8**, further comprising:

a switch connected in series between the at least one collection device and the load; and

a capacitor connected in parallel with the switch and the load.

15. The system of claim **14**, wherein the switch comprises an interrupter connected between the at least one collection device and the load, and wherein the interrupter comprises at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

16. The system of claim **15**, further comprising a transformer connected between the interrupter and the load.

17. The system of claim **8**, further comprising:

a motor for providing power, the motor connected between the at least one collection device and the load; and

a generator powered by the motor.

18. The system of claim **8**, further comprising a fuel cell between the support structure and the load.

19. The system of claim **18**, wherein the fuel cell produces hydrogen and oxygen.

20. A system of collecting energy comprising:

means for suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to the means for suspending;

means for passively inducing current flow, the means for inducing current flow electrically connected to the means for suspending; and

means for restricting the backflow of charge carriers, the means for restricting the backflow of charge carriers electrically connected between the at least one collection device and the means for inducing current flow.

21. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the collection device electrically connected to and suspended from the support structure, the at least one collection device configured to passively draw current;

a load electrically connected to the at least one collection device; and

a switch connected in series between the at least one collection device and the load, the switch comprising an interrupter connected between the at least one collection device and the load, and wherein the interrupter comprises at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

22. The system of claim **21**, further comprising a transformer connected between the interrupter and the load.

23. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to and suspended from the support structure, the collection device configured to passively draw current;

a load electrically connected to the at least one collection device;

a motor for providing power, the motor connected between the at least one collection device and the load; and

a generator powered by the motor.

24. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to and suspended from the support structure the at least one collection device configured to passively draw current;

a load electrically connected to the at least one collection device; and

a fuel cell between the support structure and the load.

25. The system of claim **24**, wherein the fuel cell produces hydrogen and oxygen.

* * * * *

455-619 AU 233 EX
FIP8106 OR 685,958

55

No. 685,958. Patented Nov. 5, 1901.

N. TESLA.

METHOD OF UTILIZING RADIANT ENERGY.

(Application filed Mar. 21, 1901.)

(No Model.)

Electric steping motor energized by corpusculor energy from sun

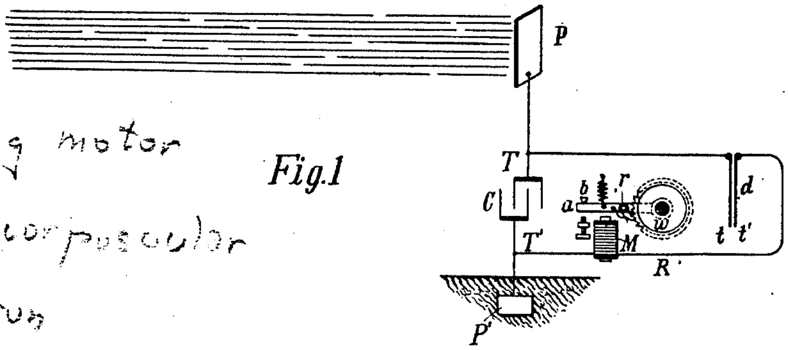

Fig.1

No refs.

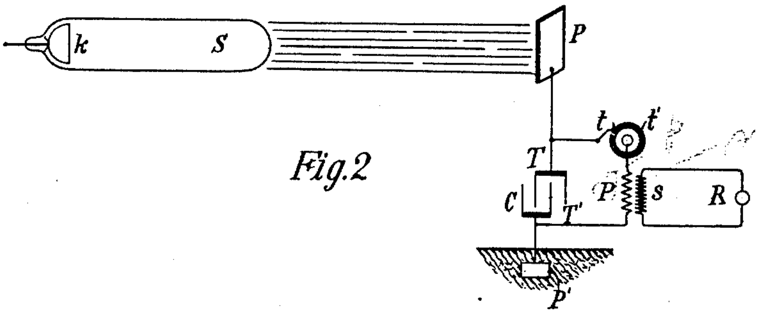

Fig.2

Witnesses:
Raphaël Netter
M. Lawson Dyer

Nikola Tesla, Inventor
by Kerr, Page & Cooper Attys

UNITED STATES PATENT OFFICE.

NIKOLA TESLA, OF NEW YORK, N. Y.

METHOD OF UTILIZING RADIANT ENERGY.

SPECIFICATION forming part of Letters Patent No. 685,958, dated November 5, 1901.

Application filed March 21, 1901. Serial No. 52,154. (No model.)

To all whom it may concern:

Be it known that I, NIKOLA TESLA, a citizen of the United States, residing at the borough of Manhattan, in the city, county, and State of New York, have invented certain new and useful Improvements in Methods of Utilizing Radiant Energy, of which the following is a specification, reference being had to the drawings accompanying and forming a part of the same.

It is well known that certain radiations—such as those of ultra-violet light, cathodic, Roentgen rays, or the like—possess the property of charging and discharging conductors of electricity, the discharge being particularly noticeable when the conductor upon which the rays impinge is negatively electrified. These radiations are generally considered to be ether vibrations of extremely small wave lengths, and in explanation of the phenomena noted it has been assumed by some authorities that they ionize or render conducting the atmosphere through which they are propagated. My own experiments and observations, however, lead me to conclusions more in accord with the theory heretofore advanced by me that sources of such radiant energy throw off with great velocity minute particles of matter which are strongly electrified, and therefore capable of charging an electrical conductor, or even if not so may at any rate discharge an electrified conductor either by carrying off bodily its charge or otherwise.

My present application is based upon a discovery which I have made that when rays or radiations of the above kind are permitted to fall upon an insulated conducting body connected to one of the terminals of a condenser, while the other terminal of the same is made by independent means to receive or to carry away electricity, a current flows into the condenser so long as the insulated body is exposed to the rays, and under the conditions hereinafter specified an indefinite accumulation of electrical energy in the condenser takes place. This energy after a suitable time interval, during which the rays are allowed to act, may manifest itself in a powerful discharge, which may be utilized for the operation or control of mechanical or electrical devices or rendered useful in many other ways.

In applying my discovery I provide a condenser, preferably of considerable electrostatic capacity, and connect one of its terminals to an insulated metal plate or other conducting body exposed to the rays or streams of radiant matter. It is very important, particularly in view of the fact that electrical energy is generally supplied at a very slow rate to the condenser, to construct the same with the greatest care. I use by preference the best quality of mica as dielectric, taking every possible precaution in insulating the armatures, so that the instrument may withstand great electrical pressures without leaking and may leave no perceptible electrification when discharging instantaneously. In practice I have found that the best results are obtained with condensers treated in the manner described in a patent granted to me February 23, 1897, No. 577,671. Obviously the above precautions should be the more rigorously observed the slower the rate of charging and the smaller the time interval during which the energy is allowed to accumulate in the condenser. The insulated plate or conducting body should present as large a surface as practicable to the rays or streams of matter, I having ascertained that the amount of energy conveyed to it per unit of time is under otherwise identical conditions proportionate to the area exposed, or nearly so. Furthermore, the surface should be clean and preferably highly polished or amalgamated. The second terminal or armature of the condenser may be connected to one of the poles of a battery or other source of electricity or to any conducting body or object whatever of such properties or so conditioned that by its means electricity of the required sign will be supplied to the terminal. A simple way of supplying positive or negative electricity to the terminal is to connect the same either to an insulated conductor, supported at some height in the atmosphere, or to a grounded conductor, the former, as is well known, furnishing positive and the latter negative electricity. As the rays or supposed streams of matter generally convey a positive charge to the first condenser-terminal, which is connect-

262 RADIANT ENERGY.

EXAMINER

ed to the plate or conductor above mentioned, I usually connect the second terminal of the condenser to the ground, this being the most convenient way of obtaining negative electricity, dispensing with the necessity of providing an artificial source. In order to utilize for any useful purpose the energy accumulated in the condenser, I furthermore connect to the terminals of the same a circuit including an instrument or apparatus which it is desired to operate and another instrument or device for alternately closing and opening the circuit. This latter may be any form of circuit-controller, with fixed or movable parts or electrodes, which may be actuated either by the stored energy or by independent means.

The rays or radiations which are to be utilized for the operation of the apparatus above described in general terms may be derived from a natural source, as the sun, or may be artificially produced by such means, for example, as an arc-lamp, a Roentgen tube, and the like, and they may be employed for a great variety of useful purposes.

My discovery will be more fully understood from the following detailed description and annexed drawings, to which reference is now made, and in which—

Figure 1 is a diagram showing typical forms of the devices or elements as arranged and connected in applying the method for the operation of a mechanical contrivance or instrument solely by the energy stored; and Fig. 2 is a diagrammatical representation of a modified arrangement suitable for special purposes, with a circuit-controller actuated by independent means.

Referring to Fig. 1, C is the condenser, P the insulated plate or conducting body, which is exposed to the rays, and P′ another plate or conductor, all being joined in series, as shown. The terminals T T′ of the condenser are also connected to a circuit including a receiver R, which is to be operated, and a circuit-controlling device *d*, which in this case is composed of two very thin conducting-plates *t t′*, placed in close proximity and very mobile, either by reason of extreme flexibility or owing to the charater of their support. To improve their action, they should be inclosed in a receptacle from which the air may be exhausted. The receiver R is shown as consisting of an electromagnet M, a movable armature *a*, a retractile spring *b*, and a ratchet-wheel *w*, provided with a spring-pawl *r*, which is pivoted to armature *a*, as illustrated. The apparatus being arranged as shown, it will be found that when the radiations of the sun or of any other source capable of producing the effects before described fall upon the plate P an accumulation of electrical energy in the condenser C will result. This phenomenon, I believe, is best explained as follows: The sun as well as other sources of radiant energy throw off minute particles of matter positively electrified, which, impinging upon the plate P, communicate an electrical charge to the same. The opposite terminal of the condenser being connected to the ground, which may be considered as a vast reservoir of negative electricity, a feeble current flows continuously into the condenser, and inasmuch as these supposed particles are of an inconceivably small radius or curvature, and consequently charged to a relatively very high potential, this charging of the condenser may continue, as I have found in practice, almost indefinitely, even to the point of rupturing the dielectric. Obviously whatever circuit-controller be employed it should operate to close the circuit in which it is included when the potential in the condenser has reached the desired magnitude. Thus in Fig. 2 when the electrical pressure at the terminals T T′ rises to a certain predetermined value the plates *t t′*, attracting each other, close the circuit connected to the terminals. This permits a flow of current which energizes the magnet M, causing it to draw down the armature *a* and impart a partial rotation to the ratchet-wheel *w*. As the current ceases the armature is retracted by the spring *b* without, however, moving the wheel *w*. With the stoppage of the current the plates *t t′* cease to be attracted and separate, thus restoring the circuit to its original condition.

Many useful applications of this method of utilizing the radiations emanating from the sun or other source and many ways of carrying out the same will at once suggest themselves from the above description. By way of illustration a modified arrangement is shown in Fig. 2, in which the source S of radiant energy is a special form of Roentgen tube devised by me having but one terminal *k*, generally of aluminium, in the form of half a sphere with a plain polished surface on the front side, from which the streams are thrown off. It may be excited by attaching it to one of the terminals of any generator of sufficiently-high electromotive force; but whatever apparatus be used it is important that the tube be exhausted to a high degree, as otherwise it might prove entirely ineffective. The working or discharge circuit connected to the terminals T T′ of the condenser includes in this case the primary *p* of a transformer and a circuit-controller comprising a fixed terminal or brush *t* and a movable terminal *t′* in the shape of a wheel with conducting and insulating segments which may be rotated at an arbitrary speed by any suitable means. In inductive relation to the primary wire or coil *p* is a secondary *s*, usually of a much greater number of turns, to the ends of which is connected a receiver R. The terminals of the condenser being connected as indicated, one to an insulated plate P and the other to a grounded plate P′, when the tube S is excited rays or streams of matter are emitted from the same, which convey a positive charge to the plate P and condenser-terminal T, while terminal T′ is continuously receiving negative electricity from the plate

P′. This, as before explained, results in an accumulation of electrical energy in the condenser, which goes on as long as the circuit including the primary p is interrupted. Whenever the circuit is closed, owing to the rotation of the terminal t', the stored energy is discharged through the primary p, this giving rise in the secondary s to induced currents which operate the receiver R.

It is clear from what has been stated above that if the terminal T′ is connected to a plate supplying positive instead of negative electricity the rays should convey negative electricity to plate P. The source S may be any form of Roentgen or Lenard tube; but it is obvious from the theory of action that in order to be very effective the electrical impulses exciting it should be wholly or at least preponderatingly of one sign. If ordinary symmetrical alternating currents are employed, provision should be made for allowing the rays to fall upon the plate P only during those periods when they are productive of the desired result. Evidently if the radiations of the source be stopped or intercepted or their intensity varied in any manner, as by periodically interrupting or rythmically varying the current exciting the source, there will be corresponding changes in the action upon the receiver R, and thus signals may be transmitted and many other useful effects produced. Furthermore, it will be understood that any form of circuit-closer which will respond to or be set in operation when a predetermined amount of energy is stored in the condenser may be used in lieu of the device specifically described with reference to Fig. 1, and also that the special details of construction and arrangement of the several parts of the apparatus may be very greatly varied without departure from the invention.

Having described my invention, what I claim is—

1. The method of utilizing radiant energy, which consists in charging one of the armatures of a condenser by rays or radiations, and the other armature by independent means, and discharging the condenser through a suitable receiver, as set forth.

2. The method of utilizing radiant energy, which consists in simultaneously charging a condenser by means of rays or radiations and an independent source of electrical energy, and discharging the condenser through a suitable receiver, as set forth.

3. The method of utilizing radiant energy, which consists in charging one of the armatures of a condenser by rays or radiations, and the other by independent means, controlling the action or effect of said rays or radiations and discharging the condenser through a suitable receiver, as set forth.

4. The method of utilizing radiant energy, which consists in charging one of the armatures of a condenser by rays or radiations and the other by independent means, varying the intensity of the said rays or radiations and periodically discharging the condenser through a suitable receiver, as set forth.

5. The method of utilizing radiant energy, which consists in directing upon an elevated conductor, connected to one of the armatures of a condenser, rays or radiations capable of positively electrifying the same, carrying off electricity from the other armature by connecting the same with the ground, and discharging the accumulated energy through a suitable receiver, as set forth.

6. The method of utilizing radiant energy, which consists in charging one of the armatures of a condenser by rays or radiations, and the other by independent means, and effecting by the automatic discharge of the accumulated energy the operation or control of a suitable receiver, as set forth.

NIKOLA TESLA.

Witnesses:
M. LAWSON DYER,
RICHARD DONOVAN.

US007439712B2

(12) United States Patent
McCowen

(10) **Patent No.:** **US 7,439,712 B2**
(45) **Date of Patent:** **Oct. 21, 2008**

(54) **ENERGY COLLECTION**

(76) Inventor: **Clint McCowen**, 1902 Rue La Fontaine, Navarre, FL (US) 32566

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 509 days.

(21) Appl. No.: **11/358,264**

(22) Filed: **Feb. 21, 2006**

(65) **Prior Publication Data**

US 2007/0195481 A1 Aug. 23, 2007

(51) **Int. Cl.**
H02N 1/00 (2006.01)

(52) **U.S. Cl.** .... **322/2 A**; 307/149; 361/225; 361/117; 174/2; 320/101

(58) **Field of Classification Search** .... 361/220, 361/117, 212, 233, 221, 222, 225; 322/2 A, 322/2 R; 307/149; 174/2, 3; 320/101
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427	A *	5/1901	Palencsar	307/149
911,260	A *	2/1909	Pennock	307/149
1,014,719	A *	1/1912	Pennock	307/149
4,104,696	A	8/1978	Cochran, Jr.	
5,047,892	A	9/1991	Sakata et al.	
6,735,830	B1	5/2004	Merciel	
6,974,110	B2	12/2005	Grandics	
2003/0107869	A1	6/2003	Pan et al.	
2004/0160711	A1	8/2004	Stumberger	
2004/0212945	A1	10/2004	Sprenger et al.	

* cited by examiner

Primary Examiner—Mark T Le
(74) *Attorney, Agent, or Firm*—Thomas, Kayden, Horstemeyer & Risley, LLP

(57) **ABSTRACT**

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support wire system supported by poles. The support wire system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but carbon and graphite are preferred. Diodes may be used to restrict the backflow or loss of energy.

28 Claims, 10 Drawing Sheets

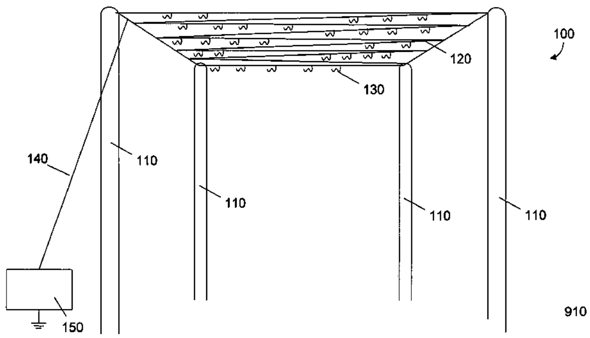
100
120
130
140
110
110
110
110
110
150

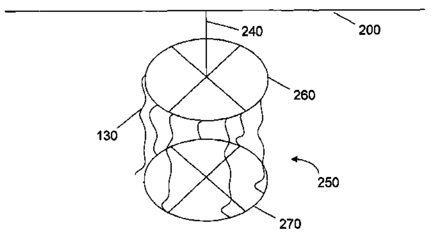
240
200
260
130
250
270

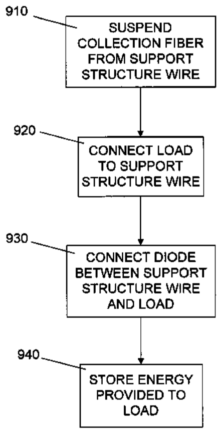
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

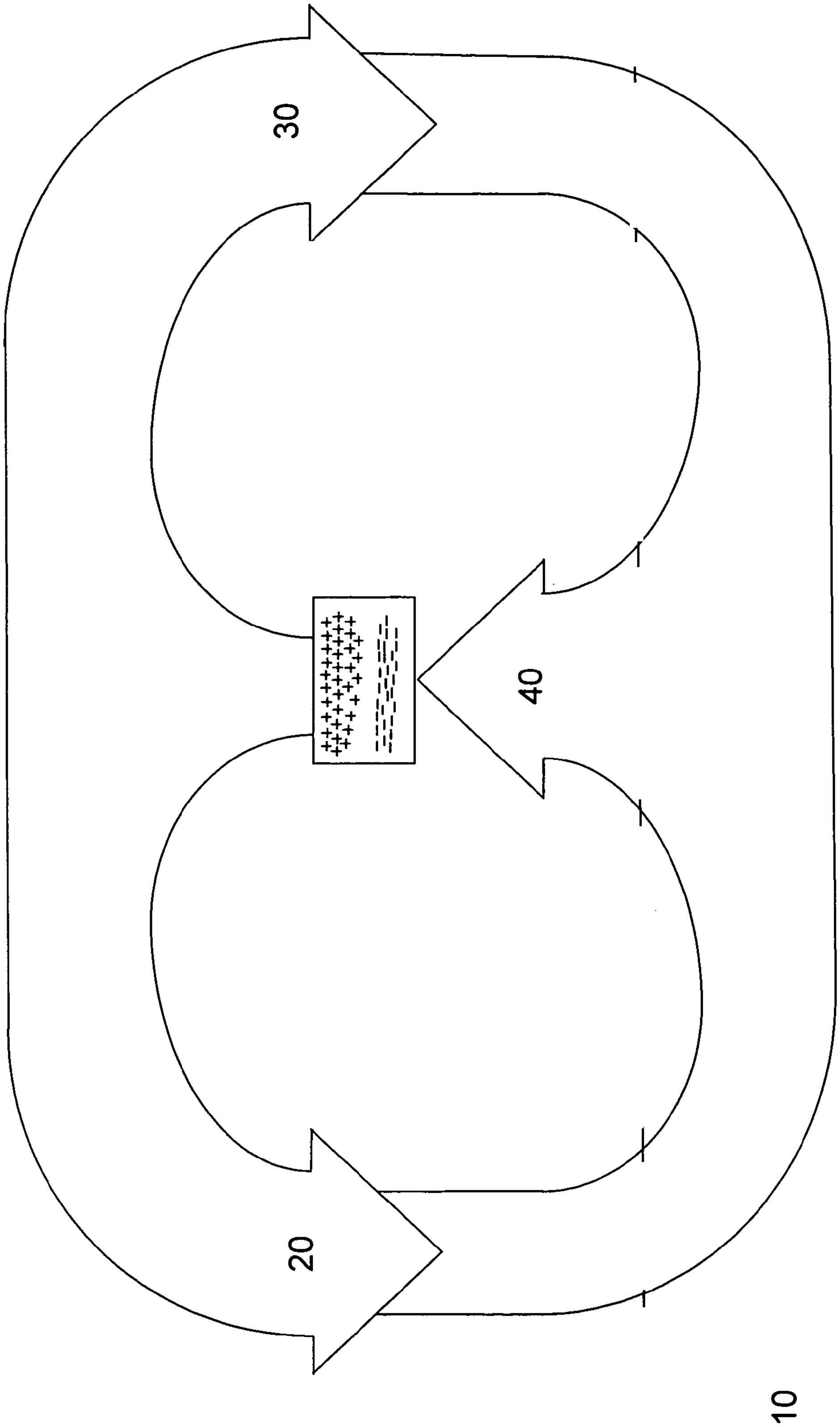

FIGURE 1

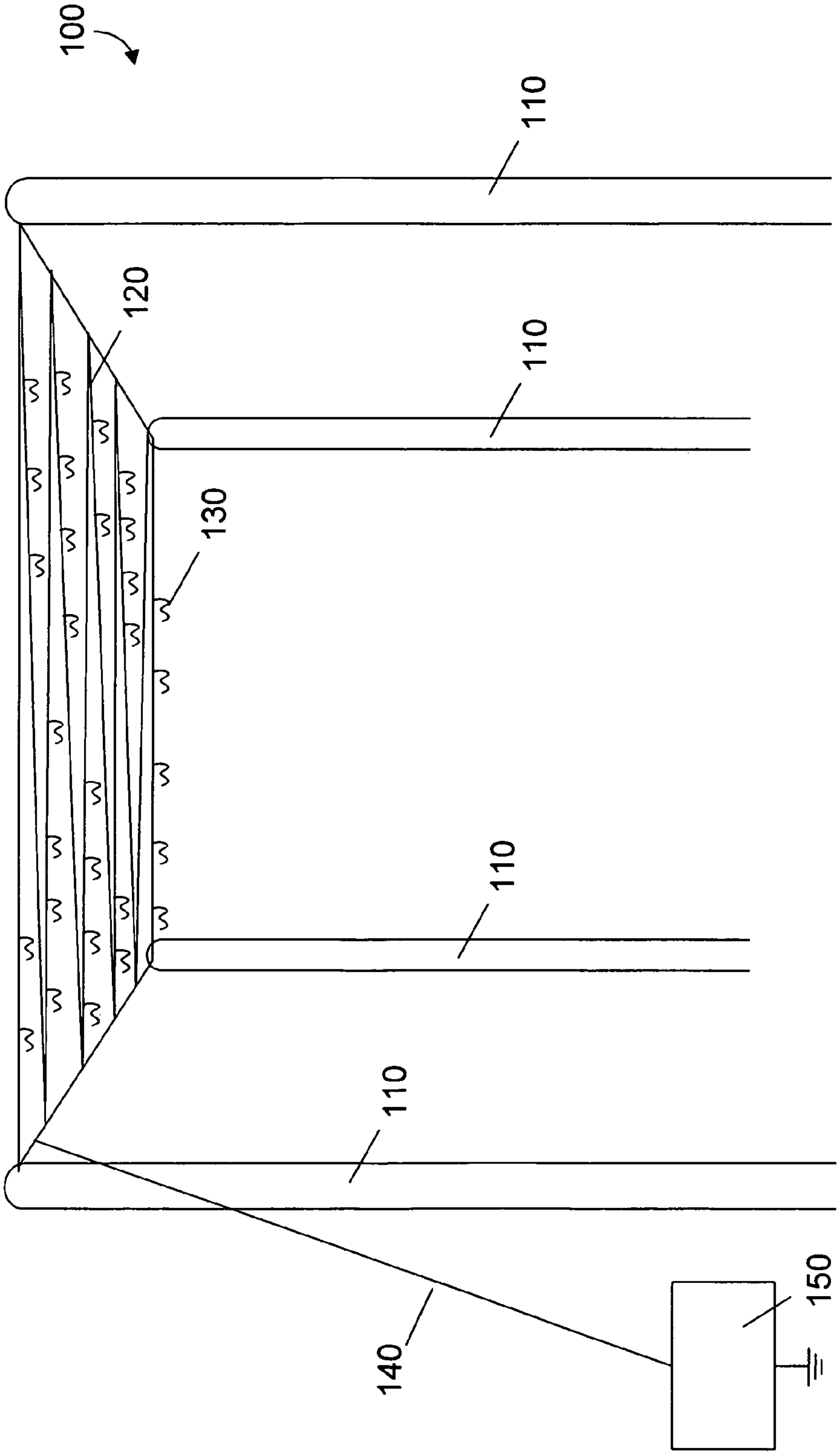
100
110
110
120
130
110
110
140
150

FIGURE 2

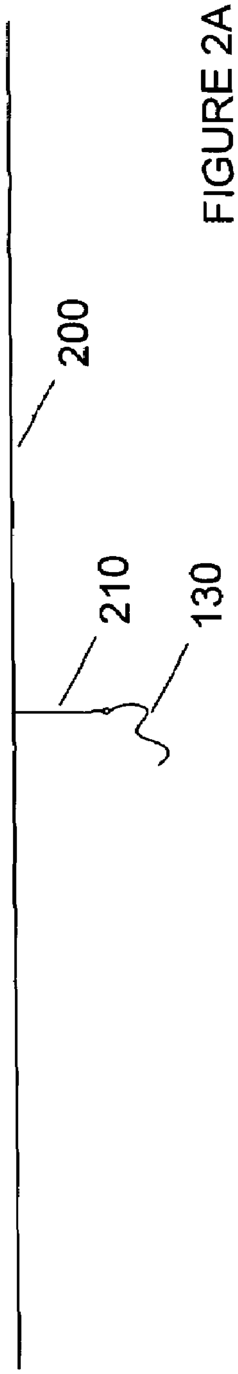

FIGURE 2A

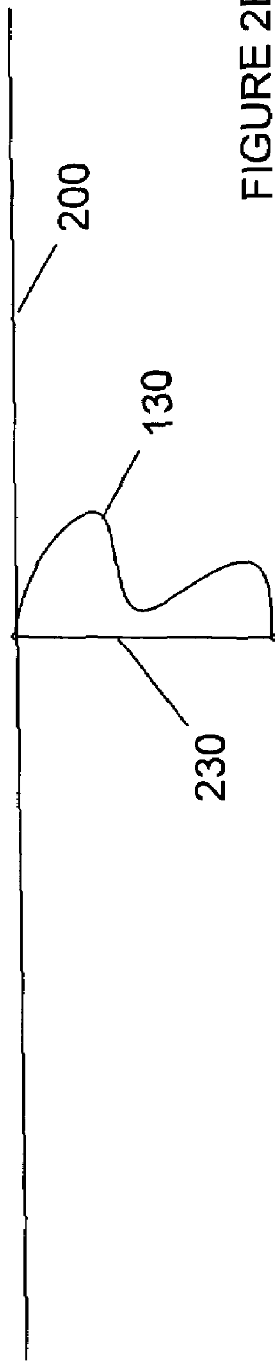

FIGURE 2B

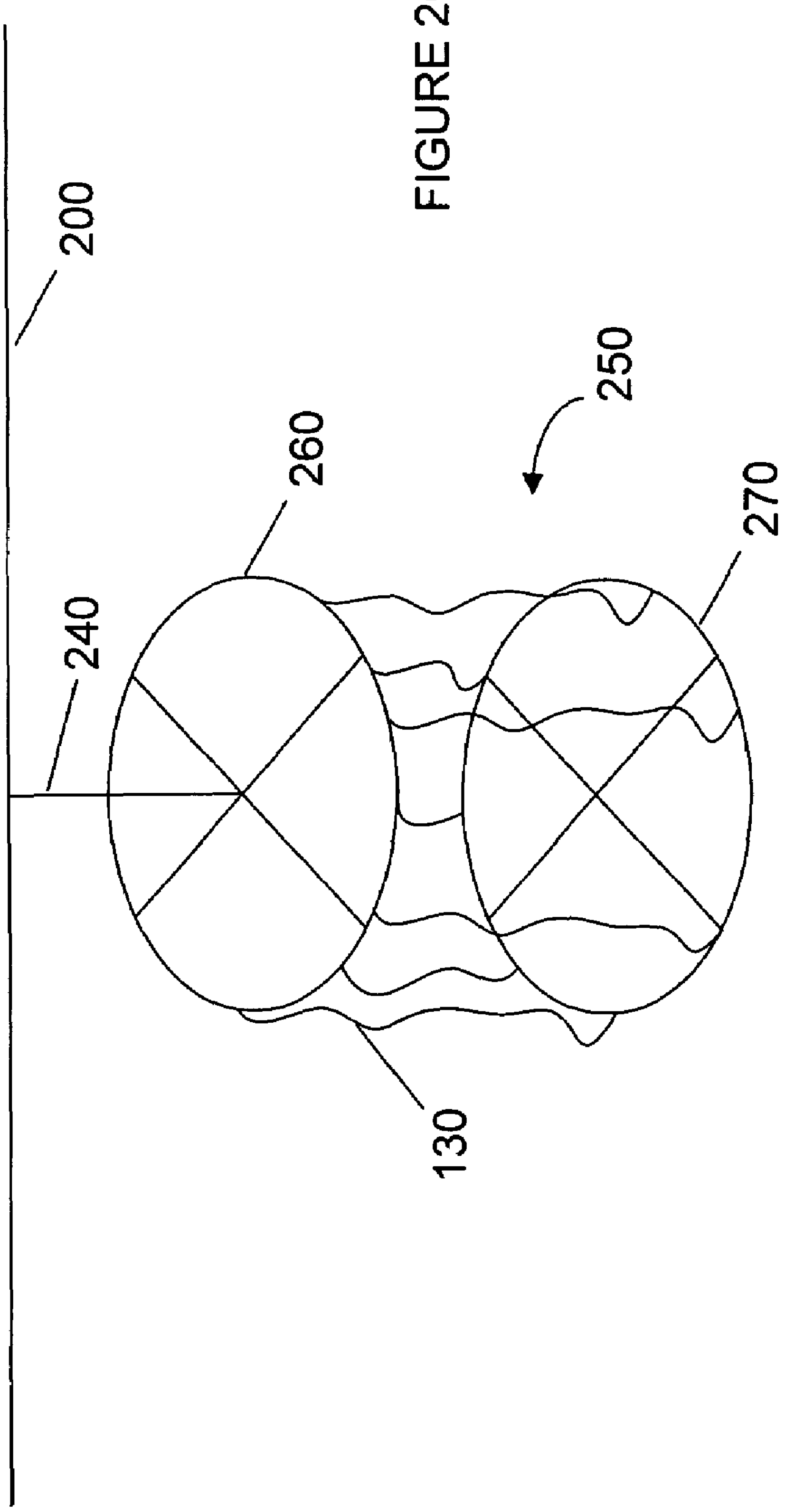
200
240
260
250
270
130

FIGURE 2C

FIGURE 2D

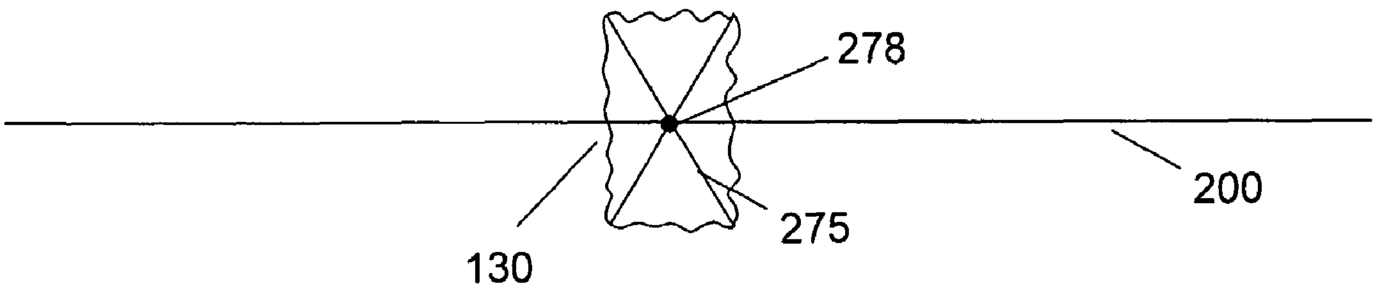
278
200
275
130

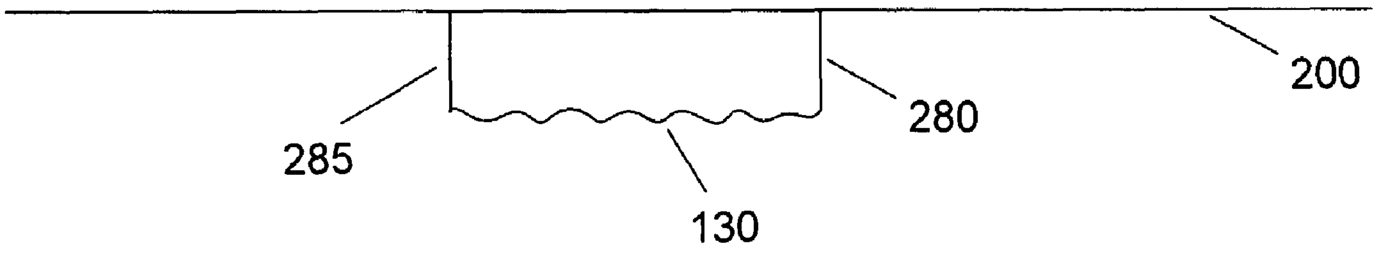
200
280
285
130

FIGURE 2E

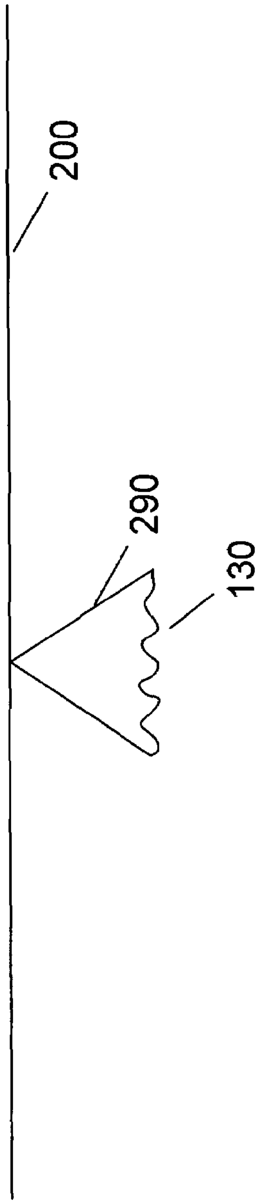

FIGURE 2F

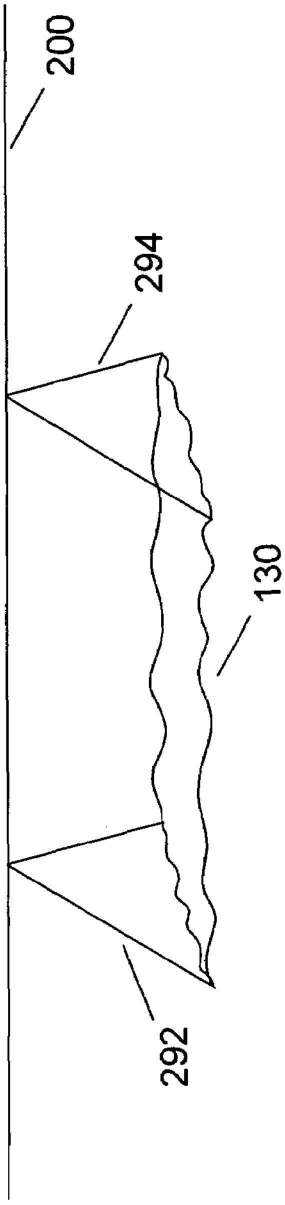
200
294
130
292

FIGURE 2G

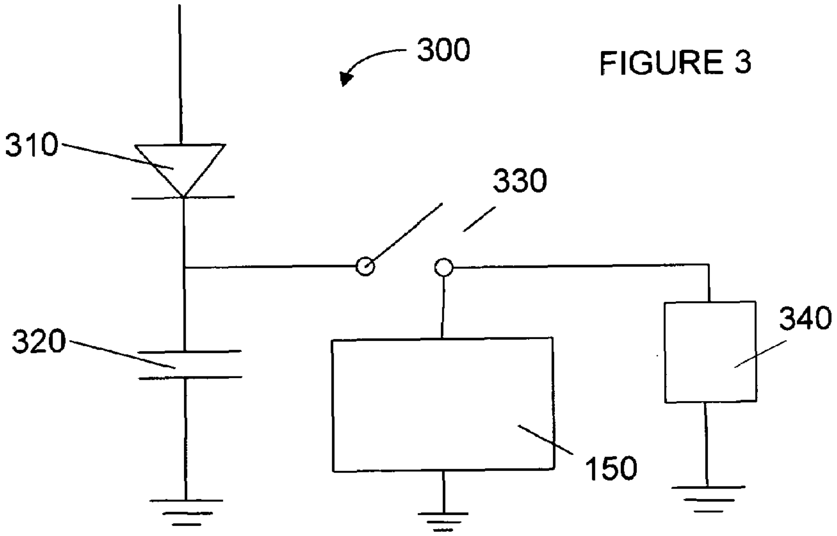
300
310
330
320
340
150

FIGURE 3

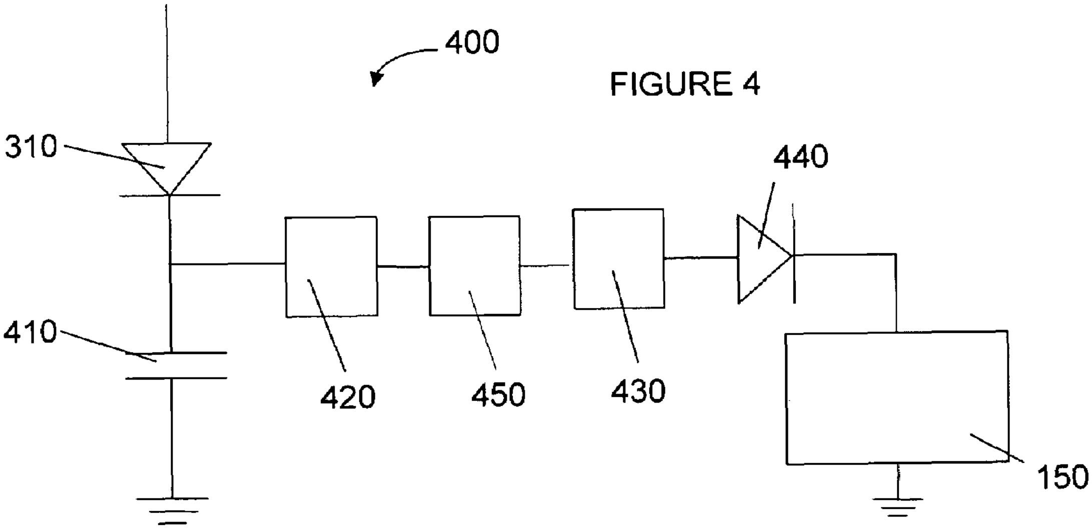
400
310
440
410
420
450
430
150

FIGURE 4

FIGURE 5

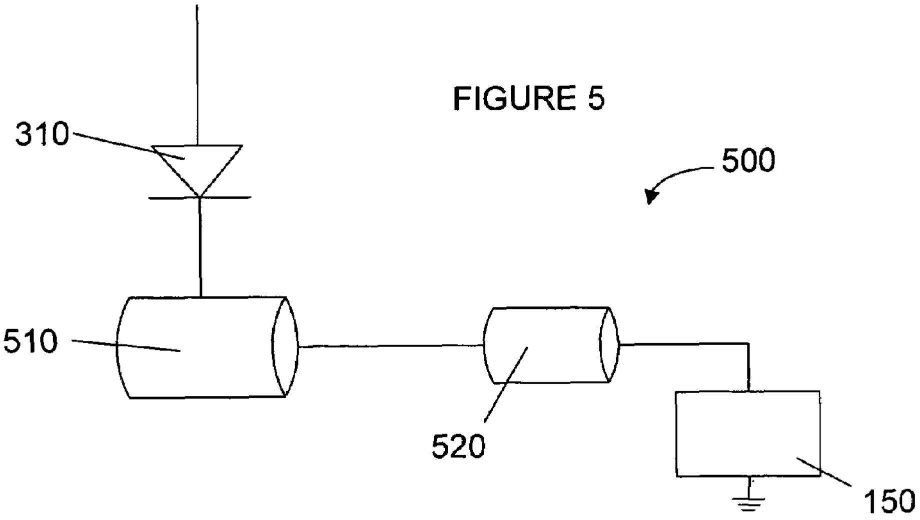
310
500
510
520
150

FIGURE 6

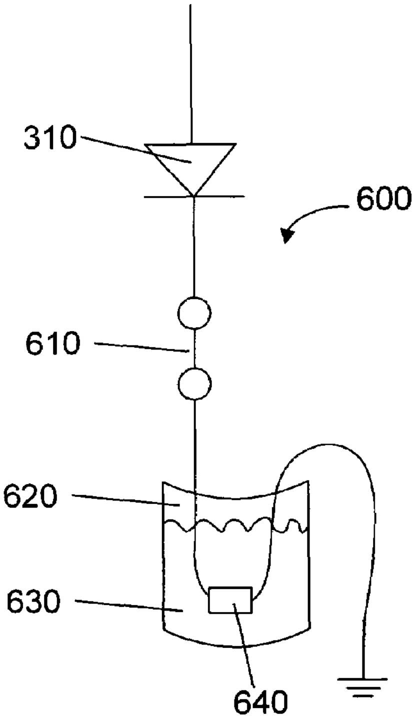
310
600
610
620
630
640

FIGURE 7

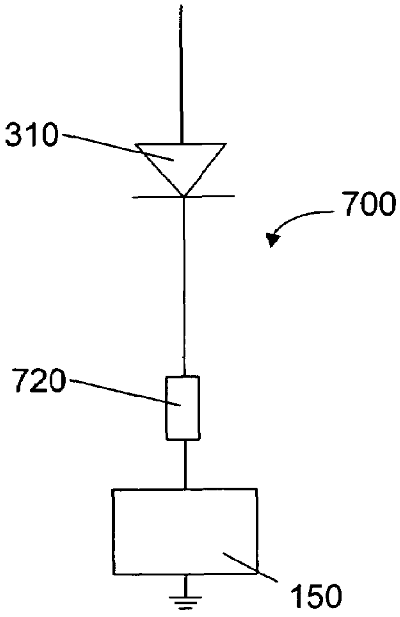
310
700
720
150

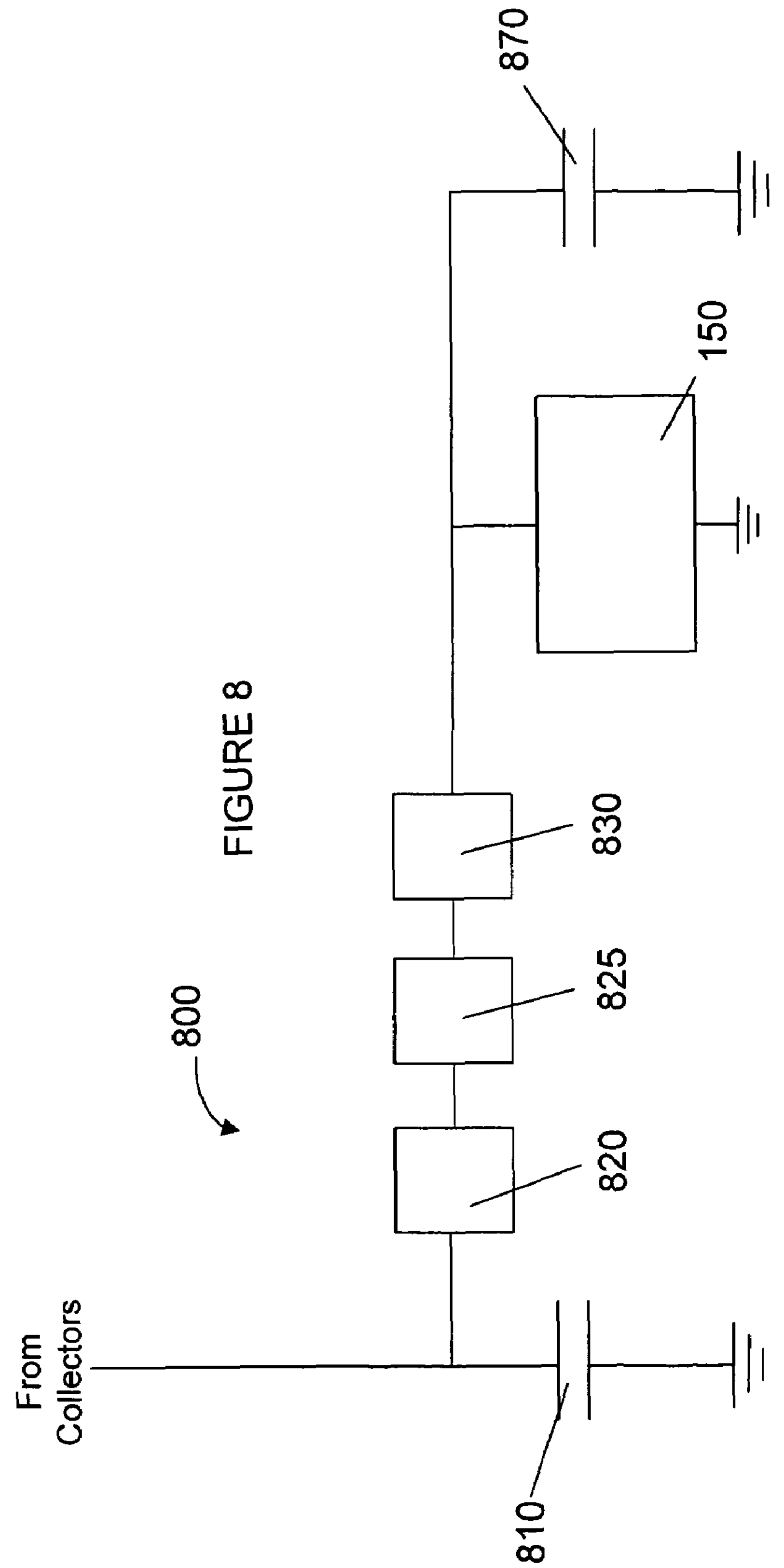
From Collectors
810
820
825
830
800
150
870

FIGURE 8

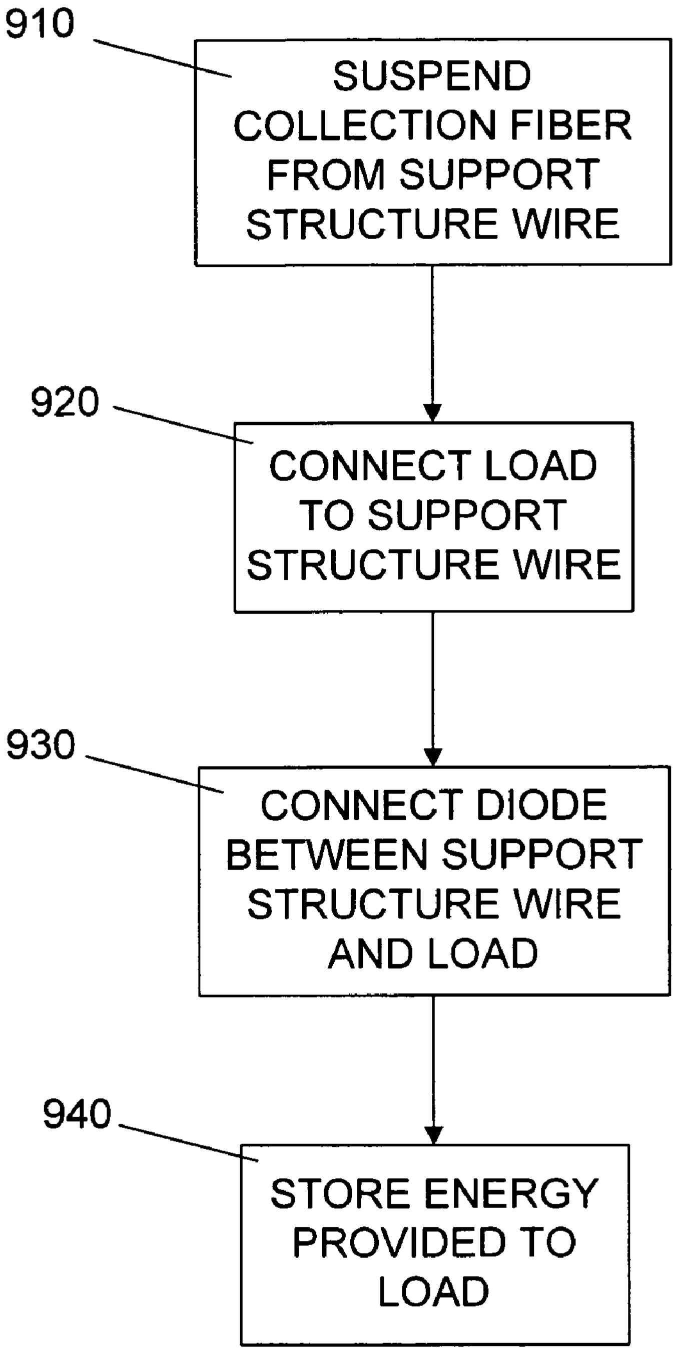
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

ENERGY COLLECTION

TECHNICAL FIELD

The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure wire elevated above a ground level, at least one collection fiber electrically connected to the support structure wire; a load electrically connected to the support structure wire; and a diode electrically connected between the load and at least one collection fiber.

Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection fiber from a support structure wire elevated above ground level, the fiber electrically connected to the support structure wire; providing a load with an electrical connection to the support structure wire to draw current; and providing a diode electrically connected between the collection fiber and the load.

Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. **1** is a circuit diagram of a weather energy circuit.

FIG. **2** is a perspective view of an exemplary embodiment of many energy collectors elevated above ground by a structure.

FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

FIG. **2**B is a side view of an exemplary embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

FIG. **2**D is a side view of an exemplary embodiment of a support structure for multiple energy collection fibers.

FIG. **2**E is a side view of a support structure for an energy collection fiber.

FIG. **2**F is a side view of an exemplary embodiment of a support structure for an energy collection fiber.

FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

FIG. **3** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **4** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **5** is a circuit diagram of an exemplary embodiment of an energy collection circuit for powering a generator and motor.

FIG. **6** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

FIG. **7** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

FIG. **8** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy.

FIG. **9** is a flow diagram of an exemplary embodiment of collecting energy with a collection fiber.

DETAILED DESCRIPTION

Electric charges on conductors reside entirely on the external surface of the conductors, and tend to concentrate more around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An exemplary embodiment of this disclosure takes advantage of this

property, among others, to collect and use the energy generated by an electric field in the atmosphere. Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An exemplary embodiment utilizes carbon or graphite fibers for static electricity collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including aluminum or steel, but most notably, copper. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

An exemplary embodiment of the collection fibers as collection device **130** includes graphite or carbon fibers. Graphite and carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, or graphite fibers are examples that demonstrate exemplary collection ability.

In at least one exemplary embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than **100** volts per meter under some conditions. When support wire **120** is suspended in the air at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An exemplary embodiment of the collection device comprises the diode with no collection fiber. A preferred embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restricts energy that may be collected through one fiber from escaping through another fiber.

Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2A-2G** using a collection fiber embodiment. FIG. **2A** presents support wire **200** with connecting member **210** for collection device **130**.

Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**.

Likewise, FIG. **2B** shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

FIG. **2C** presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

FIG. **2D** presents another exemplary embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2E** provides another exemplary embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

FIG. **2F** presents another exemplary embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2G** provides two supports as provided in FIG. **2F** such that at least two support members **292**, **294** may be connected to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and between multiple support structures. A plurality of diodes may be

placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**.

FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load.

Capacitor **320** may be connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

In an exemplary embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel. Non-limiting examples of spark gap **420** include mercury-reed switches and mercury-wetted reed switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground.

The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes may be placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches and mercury-wetted reed switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel.

FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit. Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

FIG. **8** presents exemplary circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches and mercury-wetted reed switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

It should be emphasized that the above-described embodiments of the present disclosure, particularly, any "preferred" embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A method of collecting energy comprising:

suspending at least one collection device from a plurality of support structure wires elevated above a ground level, the collection device including a carbon or graphite fiber that collects electrical enemy from the air and is electrically connected to the support structure wire; and

providing a load with an electrical connection to the plurality of support structure wires to draw current.

2. The method of claim **1**, wherein the collection device comprises a diode.

3. The method of claim **1**, wherein the collection device comprises a diode electrically connected between the fiber and the load.

4. The method of claim **1**, further comprising storing energy provided to the load.

5. The method of claim **4**, wherein storing energy provided to the load comprises storing energy in a capacitor or an inductor.

6. The method of claim **2**, wherein the diode is elevated relative to the ground level.

7. A system of energy collection comprising:

a plurality of support structure wires elevated above ground level; at least one collection device including at least one carbon or graphite fiber that collects electrical enemy from the air and is electrically connected to the plurality of support structure wires; and

a load electrically connected to the plurality of support structure wires.

8. The system of claim **7**, wherein the collection device comprises a diode.

9. The system of claim **7**, wherein the collection device comprises a diode electrically connected between the load and the fiber.

10. The system of claim **8**, wherein the diode is elevated relative to the ground level.

11. The system of claim **7**, further comprising a diode electrically connected between the at least one collection device and the plurality of support structure wires.

12. The system of claim **7**, wherein the fiber comprises a first end and a second opposing end, and wherein the system further comprises a support structure connected to both ends of the fiber.

13. The system of claim **7**, comprising:

said at least one fiber including a plurality of collection fibers; a support frame having a top side and a bottom side; and a conducting connecting wire between the plurality of support structure wires and the support frame, wherein one end of each of a plurality of collection fibers is connected to the top side of the support frame, and the opposing end of each of a plurality of collection fibers is connected to the bottom side of the support frame.

14. The system of claim **7**, further comprising a rigid structure, the rigid structure comprising multiple supports extending outward from a single point of the support structure, wherein each end of the fiber is connected to an end of the multiple supports.

15. The system of claim **7**, further comprising a rigid structure, the rigid structure comprising multiple supports extending outward from multiple points of the support structure, wherein each end of the fiber is connected to an end of the multiple supports.

16. The system of claim **14**, further comprising a plurality of rigid structures wherein said at least one fiber includes collection fibers connected between a plurality of the rigid structures.

17. The system of claim **7**, further comprising:

a switch connected in series between the plurality of support structure wires and the load; and

a capacitor connected in parallel with the switch and the load.

18. The system of claim **17**, wherein the switch comprises an interrupter connected between the plurality of support structure wires and the load.

19. The system of claim **18**, wherein the interrupter comprises one of a fluorescent tube, a neon bulb, an AC light, or a spark gap.

20. The system of claim **18**, further comprising a transformer connected between the interrupter and the load.

21. The system of claim **7**, further comprising:

a motor for providing power the motor connected between the plurality of support structure wires and the load; and

a generator powered by the motor.

22. The system of claim **7**, wherein the load comprises a spark gap in a container of fluid, and the load is used to produce a chemical reaction.

23. The system of claim **22**, wherein the fluid comprises water and the chemical reaction comprises the production of hydrogen and oxygen.

24. The system of claim **7**, further comprising a fuel cell between the plurality of support structure wires and the load.

25. The system of claim **7**, wherein the load comprises a fuel cell.

26. The system of claim **24**, wherein the fuel cell produces hydrogen and oxygen.

27. The system of claim **24**, further comprising a diode connected between the plurality of support structure wires and the fuel cell.

28. A system of collecting energy comprising:

means for suspending carbon or graphite collection fibers, the means elevated above a ground level, the collection fibers collecting electrical enemy from the air and electrically connected to the means for suspending collection fibers;

means for inducing current flow, the means for inducing current flow electrically connected to the means for suspending collection fibers; and

means for restricting the backflow of charge carriers, the means for restricting the backflow of charge carriers electrically connected between the collection fibers and the means for inducing current flow.

* * * * *

US007478712B2

(12) **United States Patent**
McCowen

(10) **Patent No.: US 7,478,712 B2**
(45) **Date of Patent: Jan. 20, 2009**

(54) ENERGY COLLECTION

(76) Inventor: **Clint McCowen**, 1902 Rue La Fontaine, Navarre, FL (US) 32566

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **11/839,112**

(22) Filed: **Aug. 15, 2007**

(65) **Prior Publication Data**

US 2007/0273206 A1 Nov. 29, 2007

Related U.S. Application Data

(63) Continuation-in-part of application No. 11/358,264, filed on Feb. 21, 2006.

(51) **Int. Cl.**
H02N 1/00 (2006.01)

(52) **U.S. Cl.** .... **191/45 R**; 307/149; 361/225; 361/117; 174/2; 320/101; 322/2 A; 191/1 R

(58) **Field of Classification Search** .... 361/220, 361/117, 212, 233, 221, 222, 225; 322/2 A, 322/2 R; 307/149; 174/2, 3; 320/101; 191/2, 191/1 R, 45 R

See application file for complete search history.

(56) References Cited

U.S. PATENT DOCUMENTS

674,427	A *	5/1901	Palencsar	307/149
911,260	A *	2/1909	Pennock	307/149
1,014,719	A *	1/1912	Pennock	307/149
4,104,696	A	8/1978	Cochran, Jr.	
5,047,892	A	9/1991	Sakata et al.	
6,735,830	B1	5/2004	Merciel	
6,974,110	B2	12/2005	Grandics	
2003/0107869	A1	6/2003	Pan et al.	
2004/0160711	A1	8/2004	Stumberger	
2004/0212945	A1	10/2004	Sprenger et al.	

* cited by examiner

Primary Examiner—Mark T Le
(74) *Attorney, Agent, or Firm*—Thomas, Kayden, Horstemeyer & Risley, LLP

(57) ABSTRACT

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support wire system supported by poles. The support wire system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but carbon, graphite, carbon/carbon, or a mixture thereof are preferred. Diodes may be used to restrict the backflow or loss of energy.

18 Claims, 15 Drawing Sheets

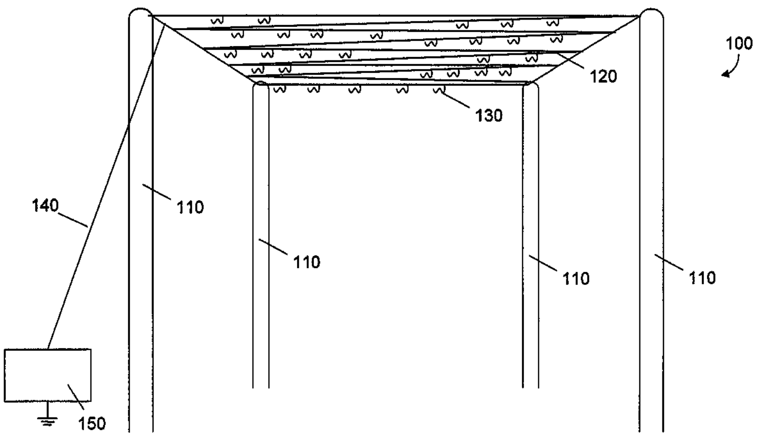
100
120
130
140
110
110
110
110
110
150

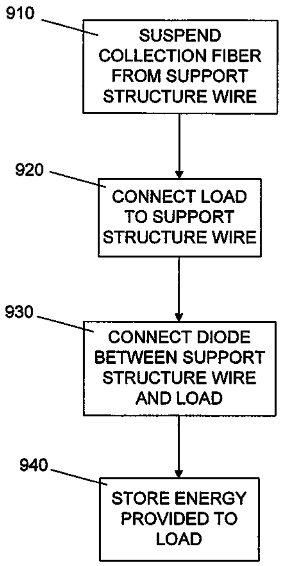
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

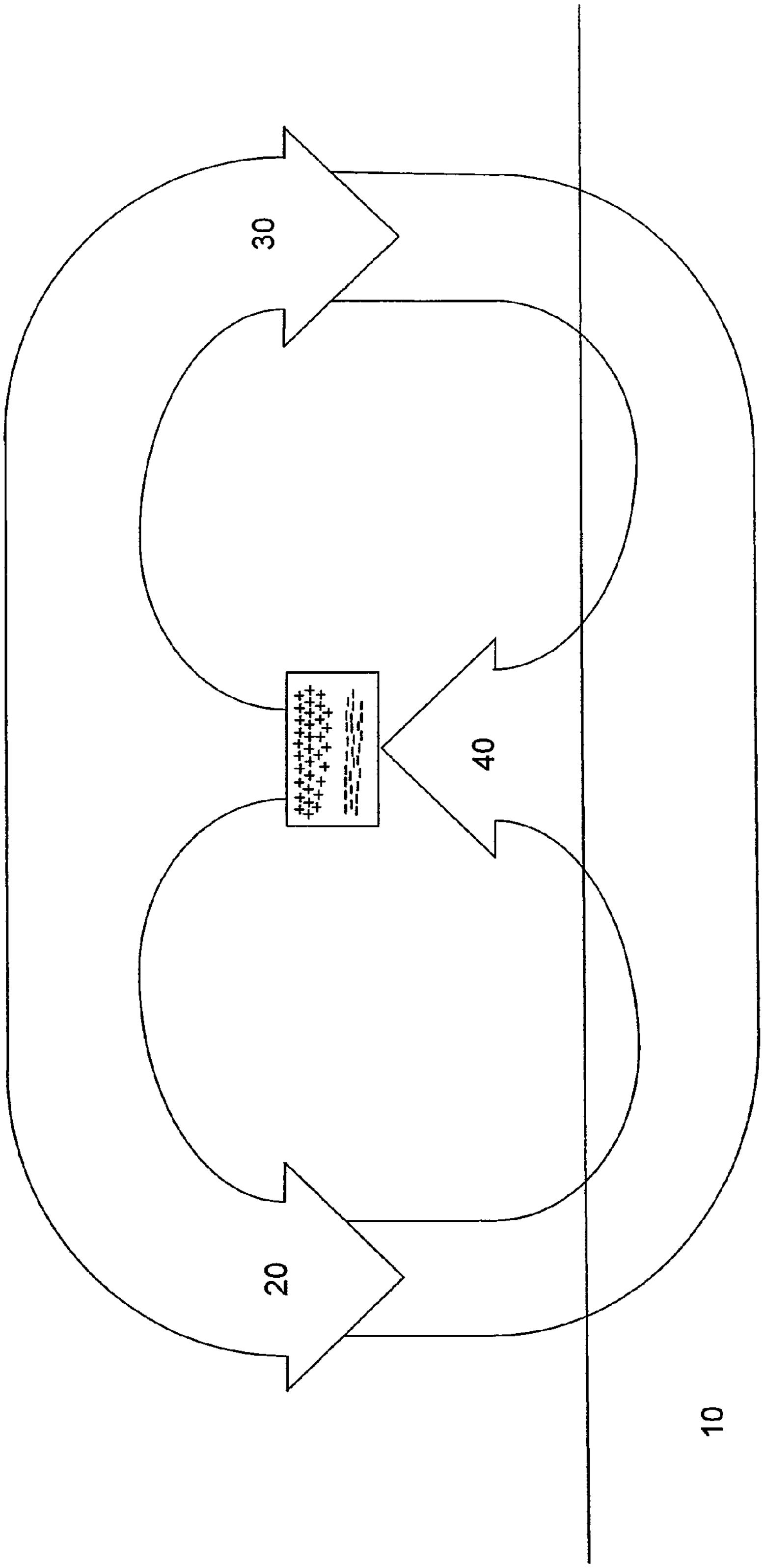

FIGURE 1

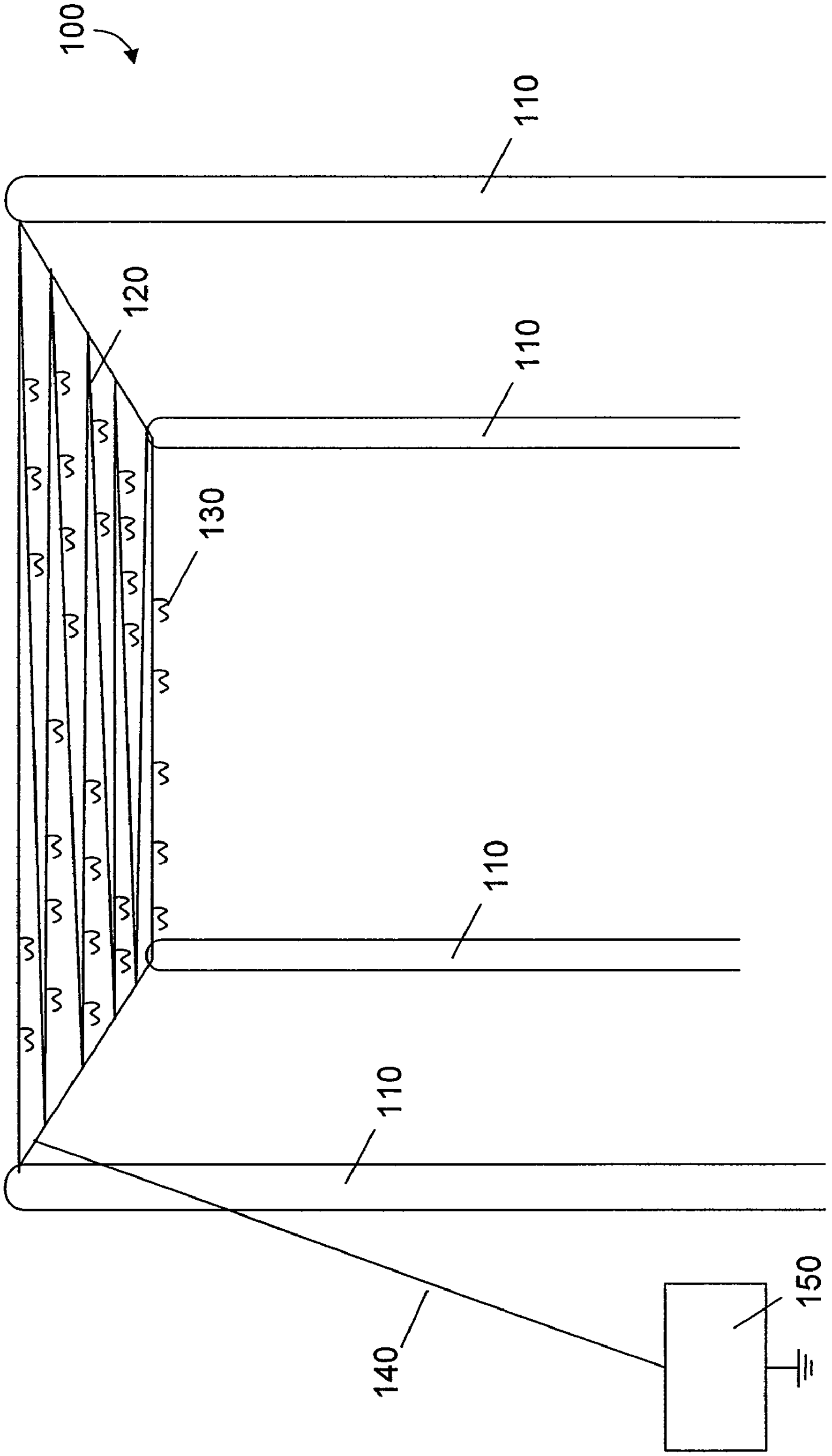
100
110
120
110
130
110
110
140
150

FIGURE 2

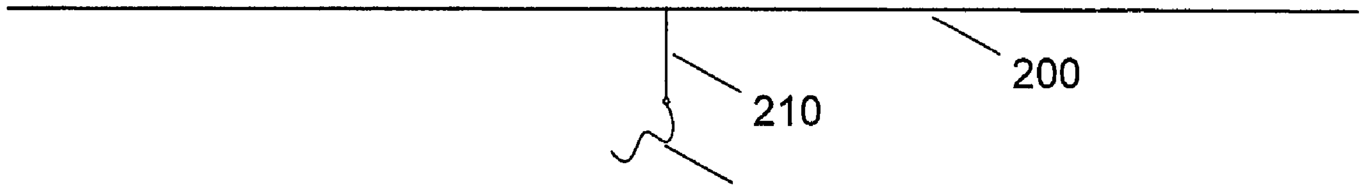

FIGURE 2A

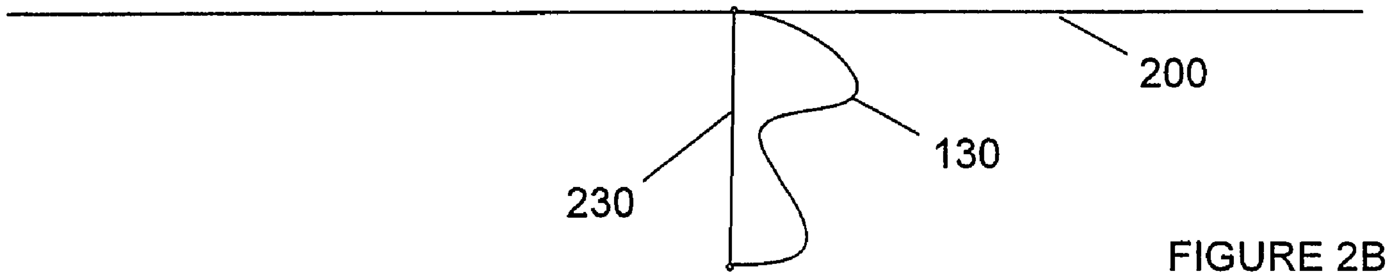

FIGURE 2B

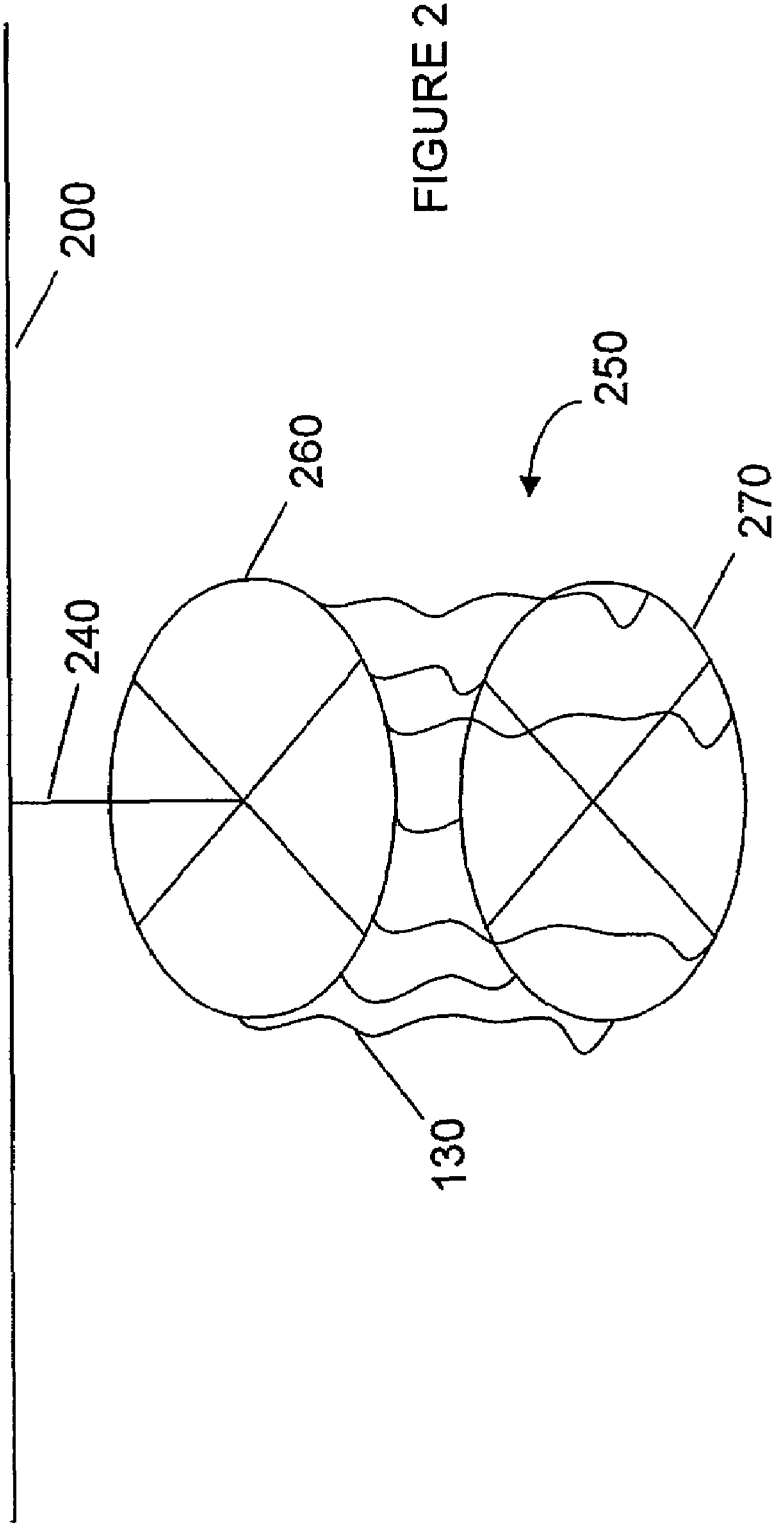
200
250
260
270
240
130

FIGURE 2C

FIGURE 2D

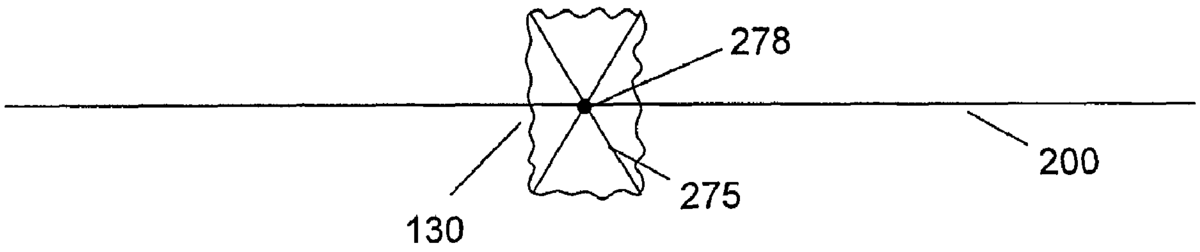
278
200
275
130

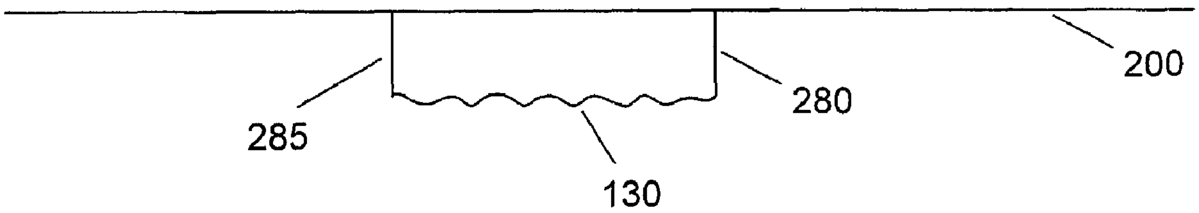
200
280
285
130

FIGURE 2E

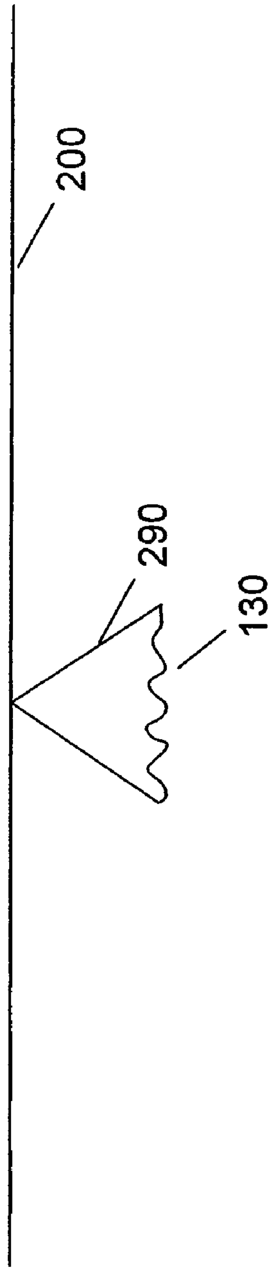

FIGURE 2F

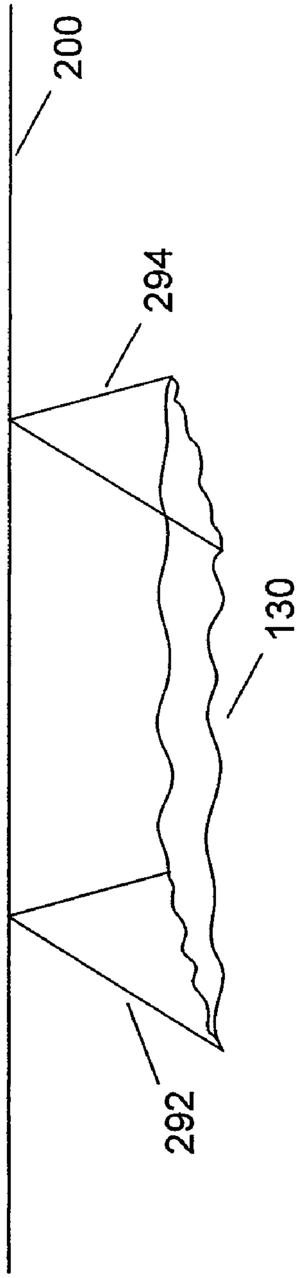
200
294
130
292

FIGURE 2G

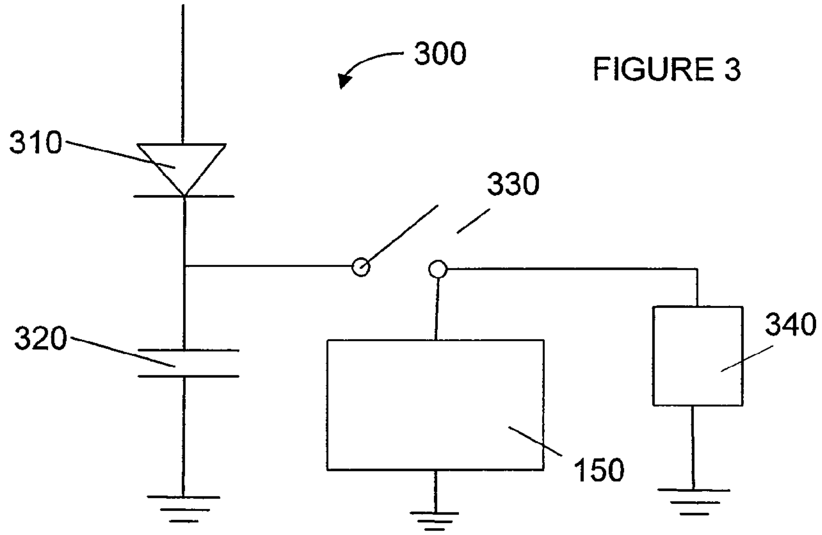
300
310
330
320
340
150

FIGURE 3

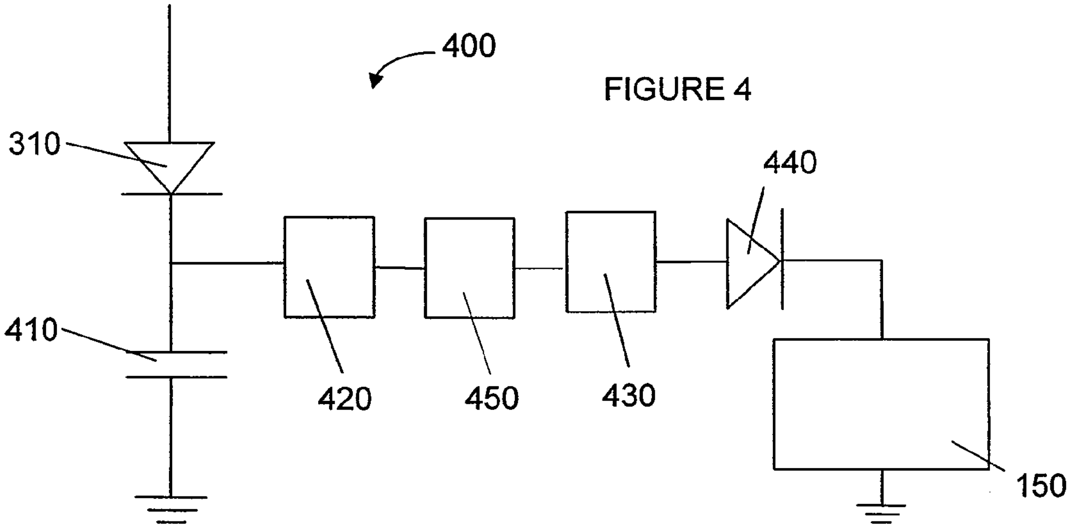
400
310
440
410
420
450
430
150

FIGURE 4

FIGURE 5

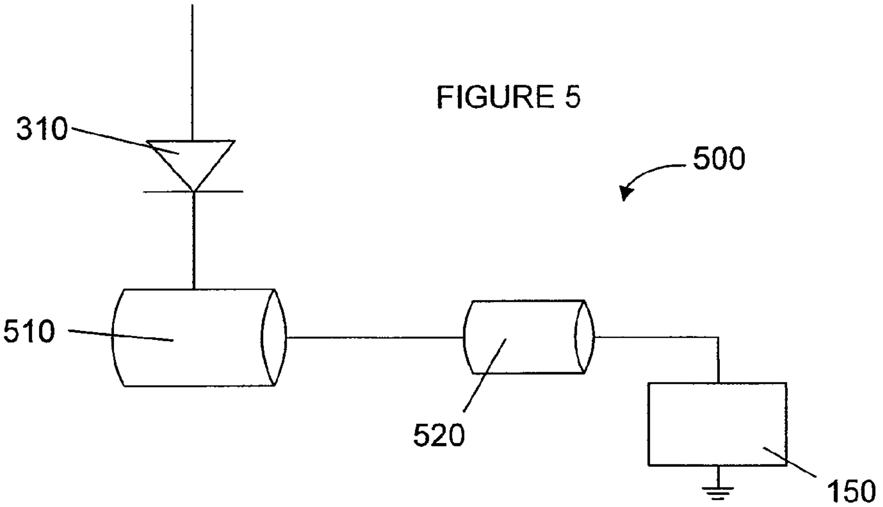
310
500
510
520
150

FIGURE 6

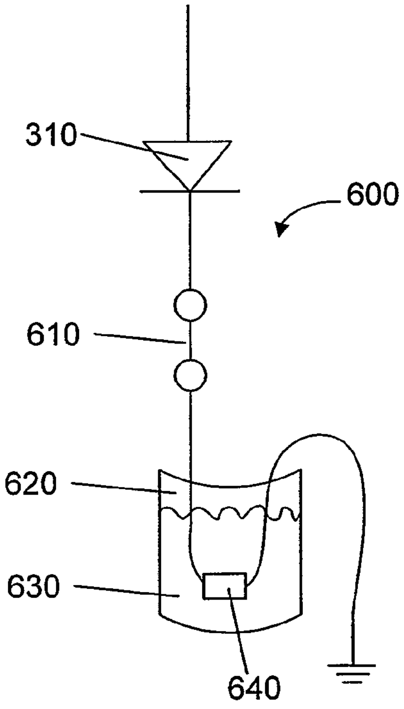
310
600
610
620
630
640

FIGURE 7

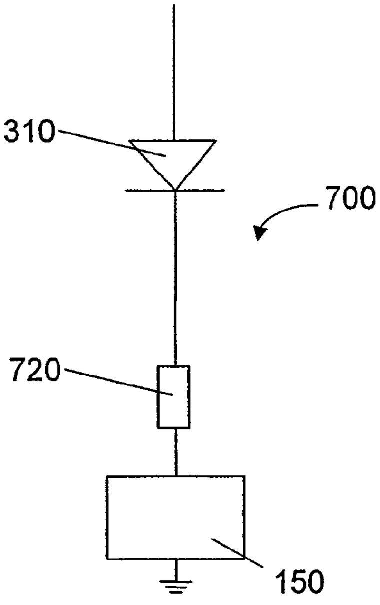
310
700
720
150

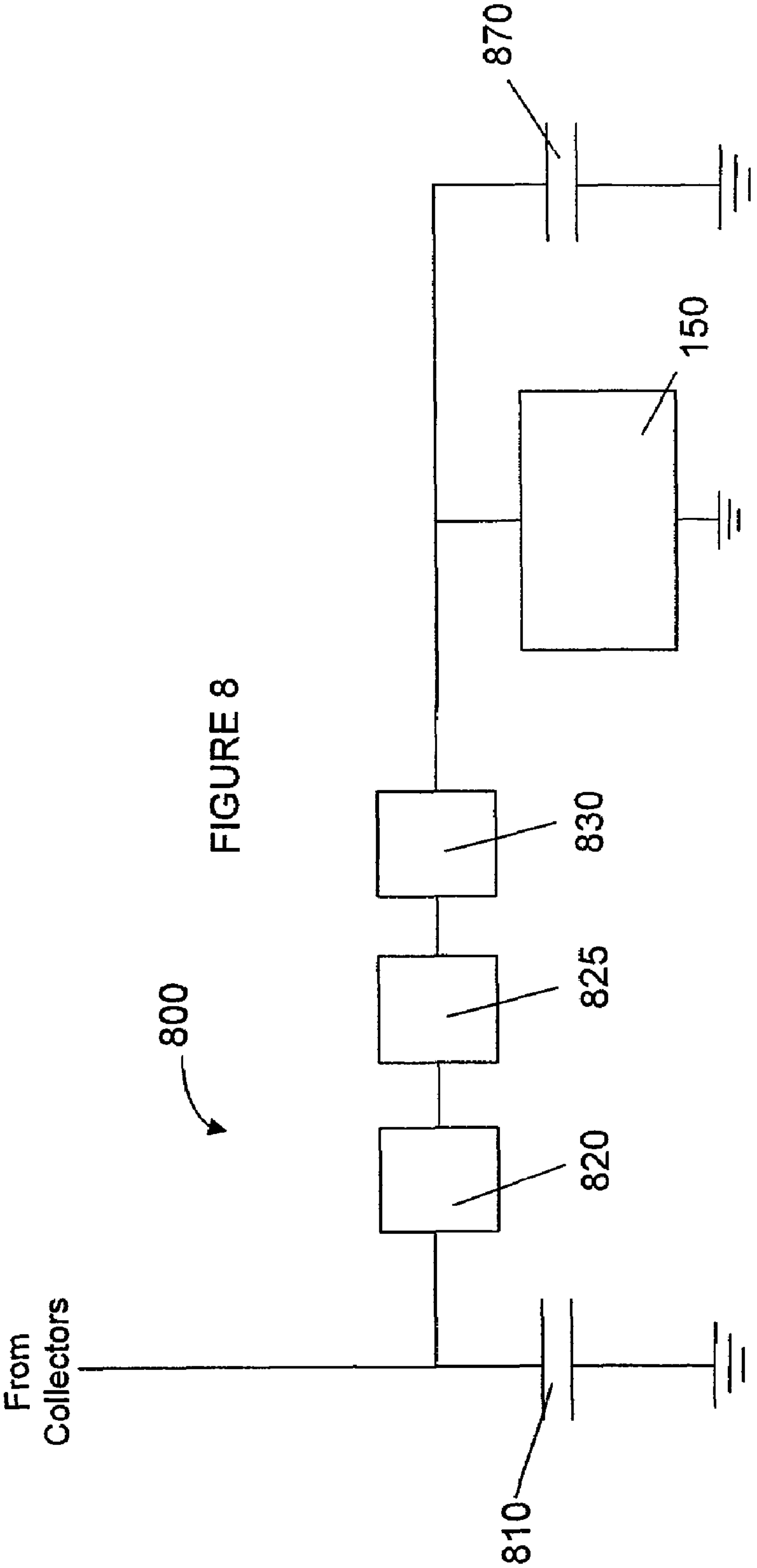
800
From Collectors
810
820
825
830
150
870

FIGURE 8

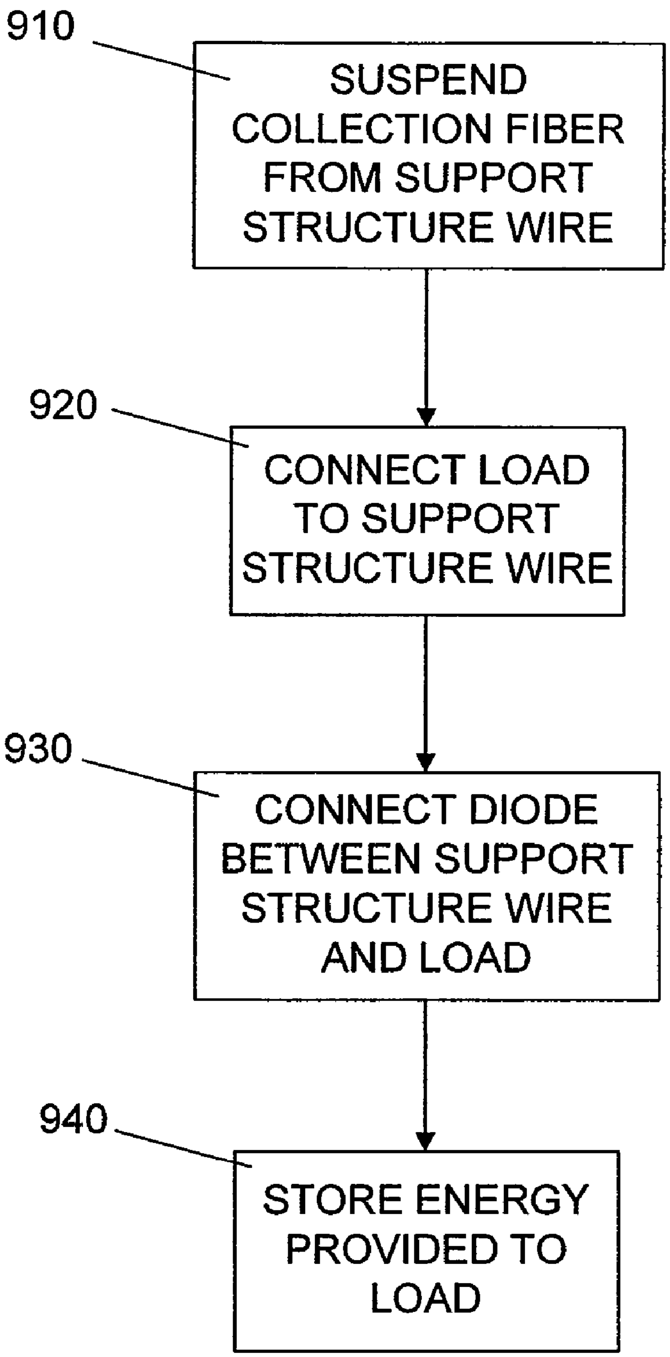
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

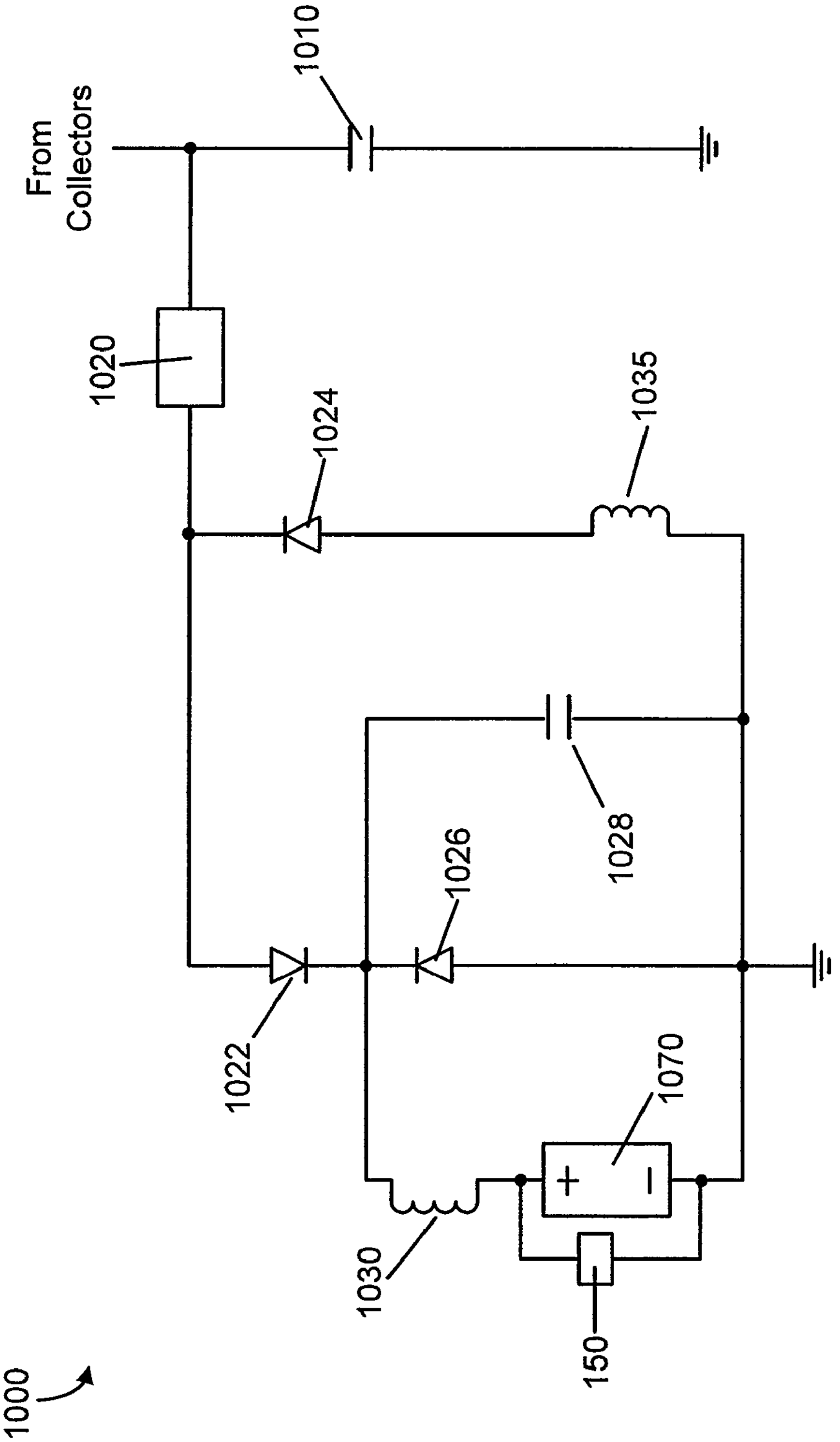
From Collectors
+
-

FIGURE 10

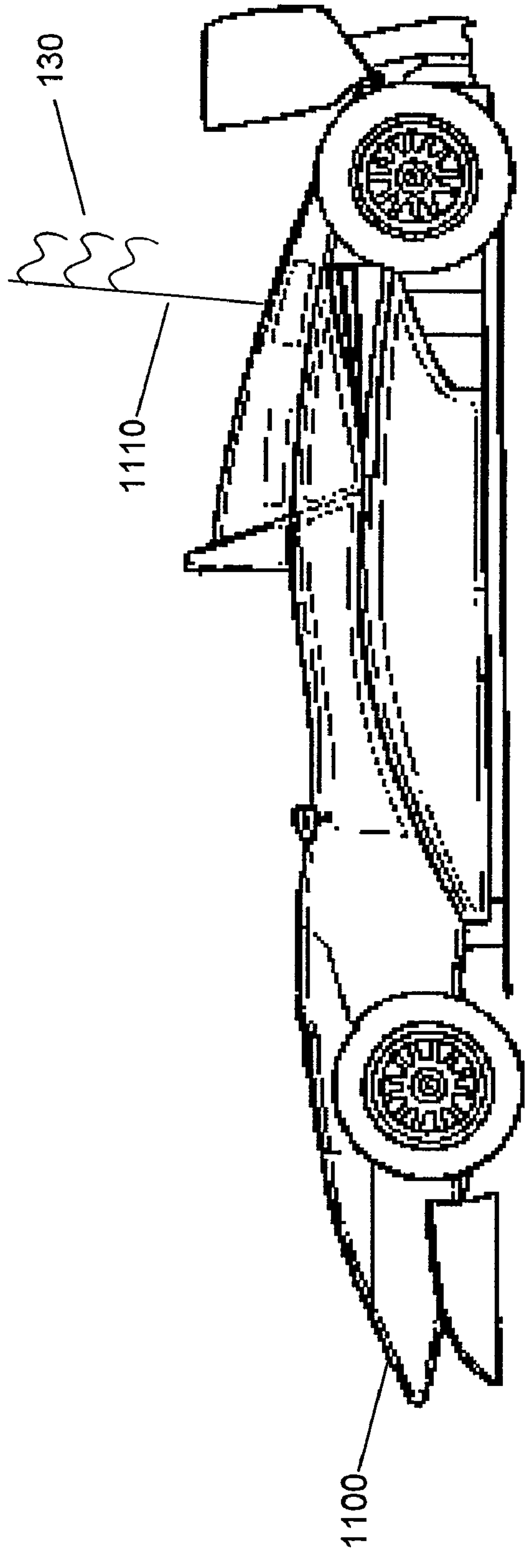
130
1110
1100

FIGURE 11

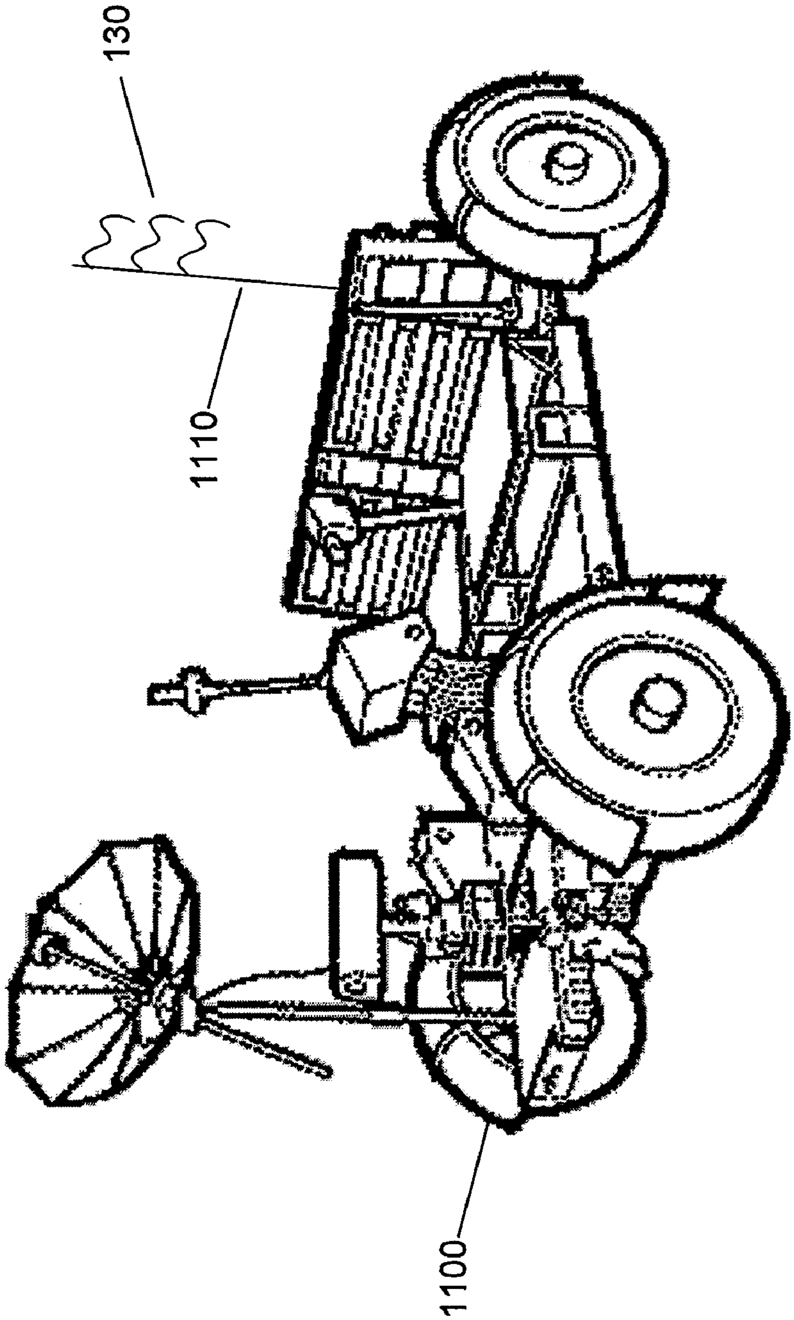
130
1110
1100

FIGURE 12

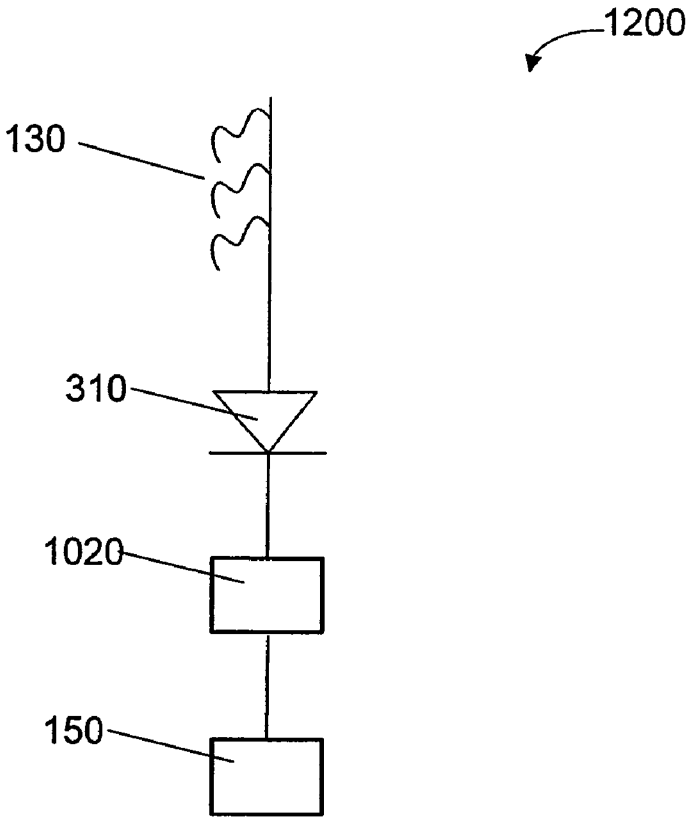
1200
130
310
1020
150

FIGURE 13

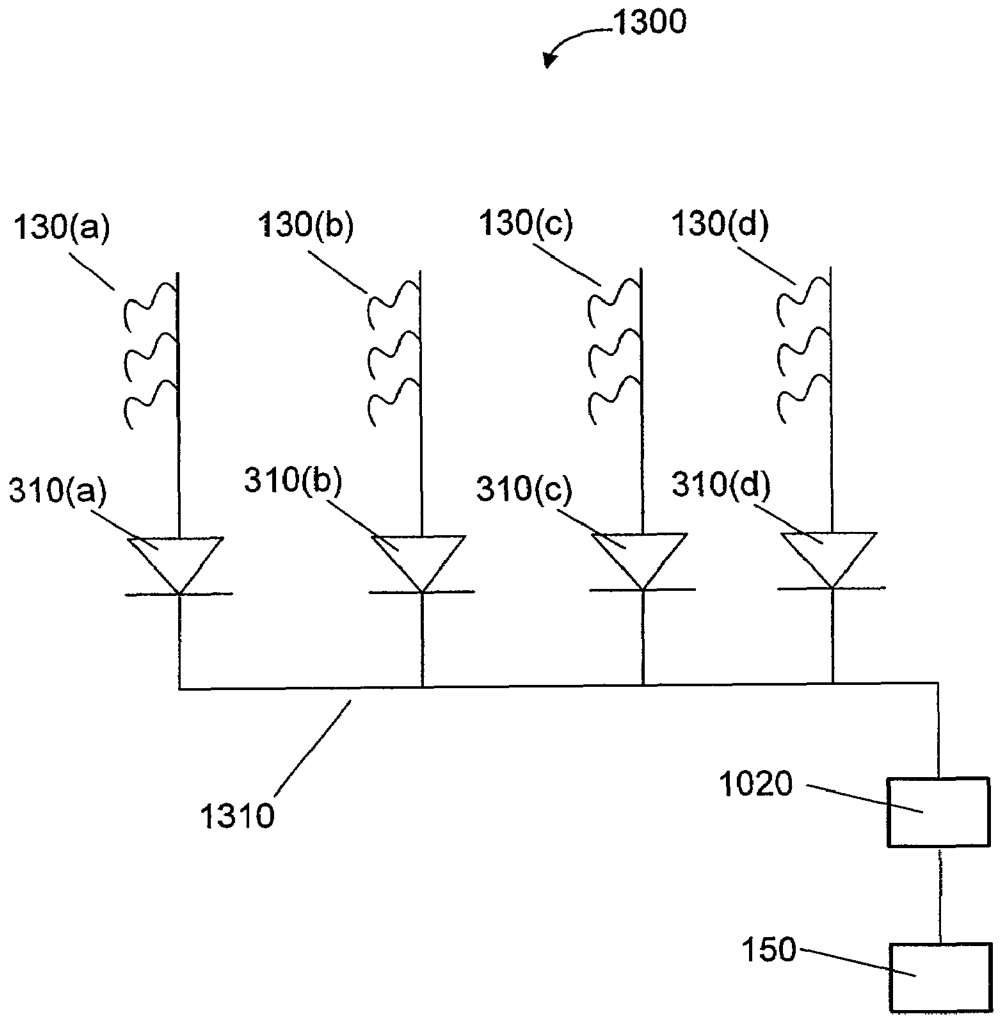
1300
130(a)
130(b)
130(c)
130(d)
310(a)
310(b)
310(c)
310(d)
1310
1020
150

FIGURE 14

ENERGY COLLECTION

CROSS REFERENCE TO RELATED APPLICATION

This application is a continuation-in-part application of U.S. patent application Ser. No. 11/358,264,filed on Feb. 21, 2006,which is incorporated by reference herein.

TECHNICAL FIELD

The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air or other gases. On Earth, clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the Earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the Earth's atmosphere or above the surface or in the atmosphere of other planets may include cosmic rays penetrating and interacting with the planet's surface and/or atmosphere, and also the creation and migration of electrically charged ions, as well as other effects yet to be fully studied.

Some of the ionization near the surface of a planet or in the lower atmosphere is caused by airborne radioactive substances. In most places on Earth, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

Alpha-active materials are also responsible for atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

While there is a large amount of usable energy available in the atmosphere of Earth, as well as above the surface or in the atmosphere of other bodies in space, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure wire elevated above a ground level, at least one collection fiber electrically connected to the support structure wire; a load electrically connected to the support structure wire; and a diode electrically connected between the load and at least one collection fiber.

Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection fiber from a support structure wire elevated above ground level, the fiber electrically connected to the support structure wire; providing a load with an electrical connection to the support structure wire to draw current; and providing a diode electrically connected between the collection fiber and the load.

Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. **1** is a circuit diagram of a weather energy circuit.

FIG. **2** is a perspective view of an example embodiment of many energy collectors elevated above ground by a structure.

FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

FIG. **2**B is a side view of an example embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

FIG. **2**D is a side view of an example embodiment of a support structure for multiple energy collection fibers.

FIG. **2**E is a side view of a support structure for an energy collection fiber.

FIG. **2**F is a side view of an example embodiment of a support structure for an energy collection fiber.

FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

FIG. **3** is a circuit diagram of an example embodiment of a circuit for the collection of energy.

FIG. **4** is a circuit diagram of an example embodiment of a circuit for the collection of energy.

FIG. **5** is a circuit diagram of an example embodiment of an energy collection circuit for powering a generator and motor.

FIG. **6** is a circuit diagram of an example embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

FIG. **7** is a circuit diagram of an example embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

FIG. **8** is a circuit diagram of an example embodiment of a circuit for collecting energy.

FIG. **9** is a flow diagram of an example embodiment of collecting energy with a collection fiber.

FIG. **10** is a circuit diagram of an example embodiment of a circuit for collecting energy from a dual polarity source.

FIG. **11** is a system diagram of an example embodiment of an energy collection system connected to an automobile vehicle.

FIG. **12** is a system diagram of an example embodiment of an energy collection system connected to a lunar rover vehicle.

FIG. **13** is a system diagram of an example embodiment of an energy collection system comprising collection devices with a diode.

FIG. **14** is a system diagram of an example embodiment of an energy collection system comprising multiple legs of the system of FIG. **13**.

DETAILED DESCRIPTION

Methods for collecting ambient energy have been known since the early 1900's. However, due to low efficiency, those methods did not become commercially feasible. One of the primary advantages of the present disclosure over past arts is the efficiency achievable by utilizing the collection fibers disclosed herein to collect or harvest said ambient energy. Another primary advantage over past arts is the dual-polarity circuit (FIG. **10**) which makes use of both positive and negative ambient energy collected by the collection fibers to convert said energy into a steady positive electrical output.

Electric charges on conductors may reside entirely on the external surface of the conductors, and tend to concentrate more around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An example embodiment of this disclosure takes advantage of this property, among others, to collect and use the energy generated by an electric field in the atmosphere or above the surface of a planet or heavenly body. "Energy" is not limited herein to energy generated by an electric field in the atmosphere. Energy as used herein may include any event, particle, or wave, including, but not limited to photons, protons, neutrons, electrons, neutrinos, leptons, muons, antimuons, mesons, baryons, positrons, bosons, gluons, gravitons, fermions, hadrons, ionic energy, gamma rays, cosmic rays, solar wind, quarks, energy generated by the triboelectric effect, convection lightning, and many other forms of energy known to one of ordinary skill in the art. These forms of energy, as well as others, may be present not only on Earth, but also on the moon of the Earth, on Mars, on Venus, and on other heavenly bodies and even in the relative vacuum of deep space.

Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An example embodiment utilizes carbon fibers, graphite fibers, carbon/carbon fibers, or a mixture thereof for energy collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including, but not limited to, aluminum, steel, copper, and super-conductive material. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire.

Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

An example embodiment of the collection fibers as collection device **130** includes graphite, carbon fibers, carbon/carbon fibers, or a mixture thereof. Graphite, carbon fibers, and carbon/carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity, energetic particles, or waves may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, graphite fibers, carbon/carbon fibers, or a mixture thereof are examples that demonstrate example collection ability.

In at least one example embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than 100-300 volts per meter under some conditions on Earth. When support wire **120** is suspended at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An example embodiment of the collection device comprises the diode with no collection fiber. Another embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restrict energy that may be collected through one fiber from escaping through another fiber.

Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2A-2G** using a collection fiber embodiment. FIG. **2A** presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. One embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**. In an alternative embodiment, a diode is not placed between collection device **130** and connection member **210**.

Likewise, FIG. **2B** shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting.

A plurality of diodes may be placed at any position on the support structure wire. One embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**. In an alternative embodiment, a diode is not placed between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

FIG. **2**C presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. One embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**. In an alternative embodiment, a diode is not placed at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

FIG. **2**D presents another example embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. One embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**. In an alternative embodiment, a diode is not placed at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2**E provides another example embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. One embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**. In an alternative embodiment, one or more diodes are not placed at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

FIG. **2**F presents another example embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. One embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**. In an alternative embodiment, a diode is not placed at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2**G provides two supports as provided in FIG. **2**F such that at least two support members **292**, **294** may be connected to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and/or between multiple support structures. A plurality of diodes may be placed at any position on the support structure. One embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**. In an alternative embodiment, one or more diodes are not placed at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load. Capacitor **320** may be connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

In an example embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **420** include mercury-reed switches, mercury-wetted reed switches, open-gap spark gaps, and electronic switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground. The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes may be placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches, mercury-wetted reed switches, open-gap spark gaps, and electronic switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel or as components for breathable air, among other uses.

FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit.

Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

FIG. **8** presents example circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches, mercury-wetted reed switches, open-gap spark gaps, and electronic switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

FIG. **10** presents circuit **1000** as an example embodiment for the collection of energy from a dual polarity source. This may be used, for example, to collect atmospheric energy that reverses in polarity compared with the ground. Such polarity reversal has been discovered as occurring occasionally on Earth during, for example, thunderstorms and bad weather, but has also been observed during good weather. Such polarity reversal may occur on other planetary bodies, including Mars and Venus, as well. Energy polarity on other planets, in deep space, or on other heavenly bodies, may be predominantly negative or predominantly positive. Collector fibers (**130** from FIG. **2**) are capable of collecting positive energy and/or negative energy, and circuit **1000** is capable of processing positive and/or negative energy, providing an output which is always positive. Circuit **1000** may collect energy from one or more collection devices (**130** from FIG. **2**). Charging capacitor **1010** may be used to store a charge until the voltage at spark gap **1020** achieves the spark voltage. Capacitor **1010** is optional.

A plurality of diodes may be placed in a plurality of positions in circuit **1000**. The voltage from capacitor **1010** may be used to charge spark gap **1020** to a sufficient voltage. Spark gap **1020** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **1020** include mercury-reed switches, mercury-wetted reed switches, open-gap spark gaps, and electronic switches. When spark gap **1020** arcs, energy will arc from an emitting end of spark gap **1020** to a receiving end of spark gap **1020**. The output of spark gap **1020** is electrically connected to the anode of diode **1022** and the cathode of diode **1024**. The cathode of diode **1022** is electrically connected to the cathode of diode **1026** and inductor **1030**. Inductor **1030** may be a fixed value inductor or a variable inductor. The anode of diode **1026** is electrically connected to ground. Capacitor **1028** is electrically connected between ground and the junction of the cathodes of diode **1022** and diode **1026**. Inductor **1035** is electrically connected between ground and the anode of diode **1024**. Inductor **1035** may be a fixed value inductor or a variable inductor. Capacitor **1070**, the anode of diode **1026**, inductor **1035**, and load **1050** are electrically connected to ground. Capacitor **1070** may be placed in parallel with load **150**.

FIGS. **11** and **12** provide example embodiments of vehicle **1110**, which utilizes electricity, the vehicle employing systems of energy collection provided herein. Vehicle **1100** in FIG. **11** is shown as an automobile vehicle, but could be any means of locomotion that utilizes electricity, including a car, a train, a motorcycle, a boat, an airplane, robotic rovers, space craft, etc. Vehicle **1100** in FIG. **12** is shown as a lunar rover vehicle. In FIGS. **11** and **12**, support rod **1110** is electrically connected to an electrical system in vehicle **1100**. Energy collectors **130** are electrically connected to support rod **1110** and may be used to supply energy to electrical circuits within the vehicle. A non-limiting use includes a top-off charge for a battery system, on-board hydrogen production, and/or assisting in the same. Energy collectors **130** may be used to augment the efficiency of the locomotion that utilizes electrical energy as well.

FIG. **13** provides an example embodiment of energy collection system **1200** in which diode **310** is used to isolate collection devices **130** from spark gap **1020** and load **150**. Collection devices **130** may comprise graphite, carbon fibers, carbon/carbon fibers, or a mixture thereof.

FIG. **14** provides an example embodiment of energy collection system **1300** in which a plurality of energy collection systems, such as that provided in FIG. **13**, are combined. Each leg consisting of collection devices **130** and diode **310** are connected in parallel with other legs, each leg electrically connected to trunk wire **1310**. The legs could also be connected serially. Trunk wire **1310** is electrically connected to a collection circuit, which may comprise load **150** and spark gap **1020** in any configuration that has been previously discussed. Each leg may comprise one or more collection devices **130** and at least one diode electrically connected between the collection devices and the collection circuit. Although three collection devices **130** are shown on each leg, any number of collection devices may be used. Although four legs are shown, any number of legs may be used.

Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

It should be emphasized that the above-described embodiments of the present disclosure, particularly, any "preferred" embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A system of energy collection comprising: a plurality of support structure wires elevated above ground level; at least one collection device including at least one carbon or graphite fiber that collects electrical energy from the air and is electrically connected to the plurality of support structure wires; a load electrically connected to the plurality of support structure wires; and a dual polarity collection circuit electrically connected between the collection device and the load.

2. The system of claim **1**, wherein the collection device comprises a diode.

3. The system of claim **1**, wherein the collection device comprises:

a diode electrically connected between the load and the fiber.

4. The system of claim **2**, wherein the diode is elevated relative to the ground level.

5. The system of claim **1**, wherein the fiber comprises graphite, carbon fibers, carbon/carbon fibers, or a mixture thereof.

6. The system of claim **1**, further comprising a diode electrically connected between the at least one collection device and the plurality of support structure wires.

7. The system of claim **1**, wherein the dual polarity collection circuit comprises a plurality of diodes, at least one inductor, and at least one capacitor.

8. The system of claim **7**, wherein the capacitor is electrically connected in parallel with the load.

9. A vehicle employing a system of energy collection comprising: a means of locomotion; a support structure electrically connected to the vehicle; at least one collection device including at least one carbon or graphite fiber that collects electrical energy from the air and is electrically connected to the support structure; and a load electrically connected to the support structure.

10. The structure of claim **9**, wherein the collection device comprises a diode.

11. The structure of claim **9**, wherein the collection device comprises: a diode electrically connected between the load and the fiber.

12. The structure of claim **10**, wherein the diode is elevated relative to the vehicle.

13. The structure of claim **9**, wherein the fiber comprises graphite, carbon fibers, carbon/carbon fibers, or a mixture thereof.

14. The structure of claim **9**, further comprising a diode electrically connected between the at least one collection device and the support structure.

15. A system of collecting energy comprising:

a plurality of collection legs, each of the collection legs comprising:

at least one collection device including at least one carbon or graphite fiber that collects electrical energy from the air; and

a diode; and

a load electrically connected to ground.

16. The system of claim **15**, wherein the fiber comprises a surface having projections.

17. The system of claim **15**, further comprising a trunk line electrically connected between the plurality of collection legs and the load.

18. The system of claim **17**, further comprising a spark gap device electrically connected between the plurality of collection legs and the load.

* * * * *

US 20160043661A1

(19) **United States**

(12) **Patent Application Publication**

McCowen

(10) **Pub. No.: US 2016/0043661 A1**

(43) **Pub. Date: Feb. 11, 2016**

(54) **ENERGY COLLECTION**

(71) Applicant: **Ion Power Group LLC**, Navarre, FL (US)

(72) Inventor: **Clint McCowen**, Navarre, FL (US)

(73) Assignee: **ION POWER GROUP LLC**, Navarre, FL (US)

(21) Appl. No.: **14/454,308**

(22) Filed: **Aug. 7, 2014**

Publication Classification

(51) **Int. Cl.**
H02N 1/04 (2006.01)

(52) **U.S. Cl.**
CPC ***H02N 1/04*** (2013.01)

(57) **ABSTRACT**

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support system. The support system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but graphene, carbon and graphite are preferred. Diodes may be used to restrict the backflow or loss of energy.

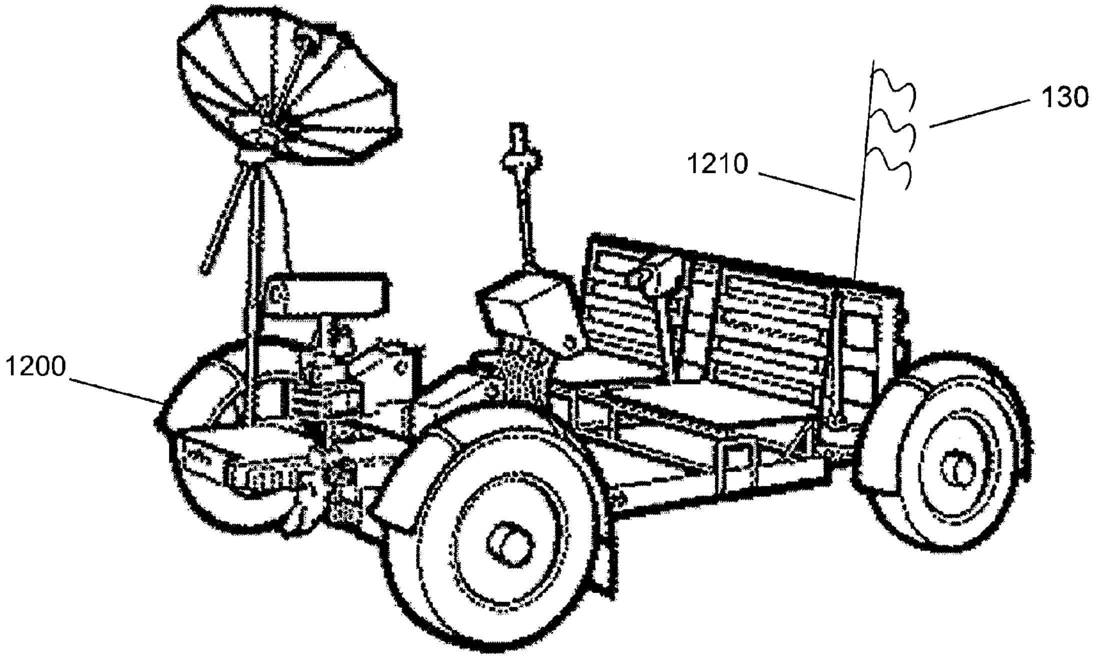

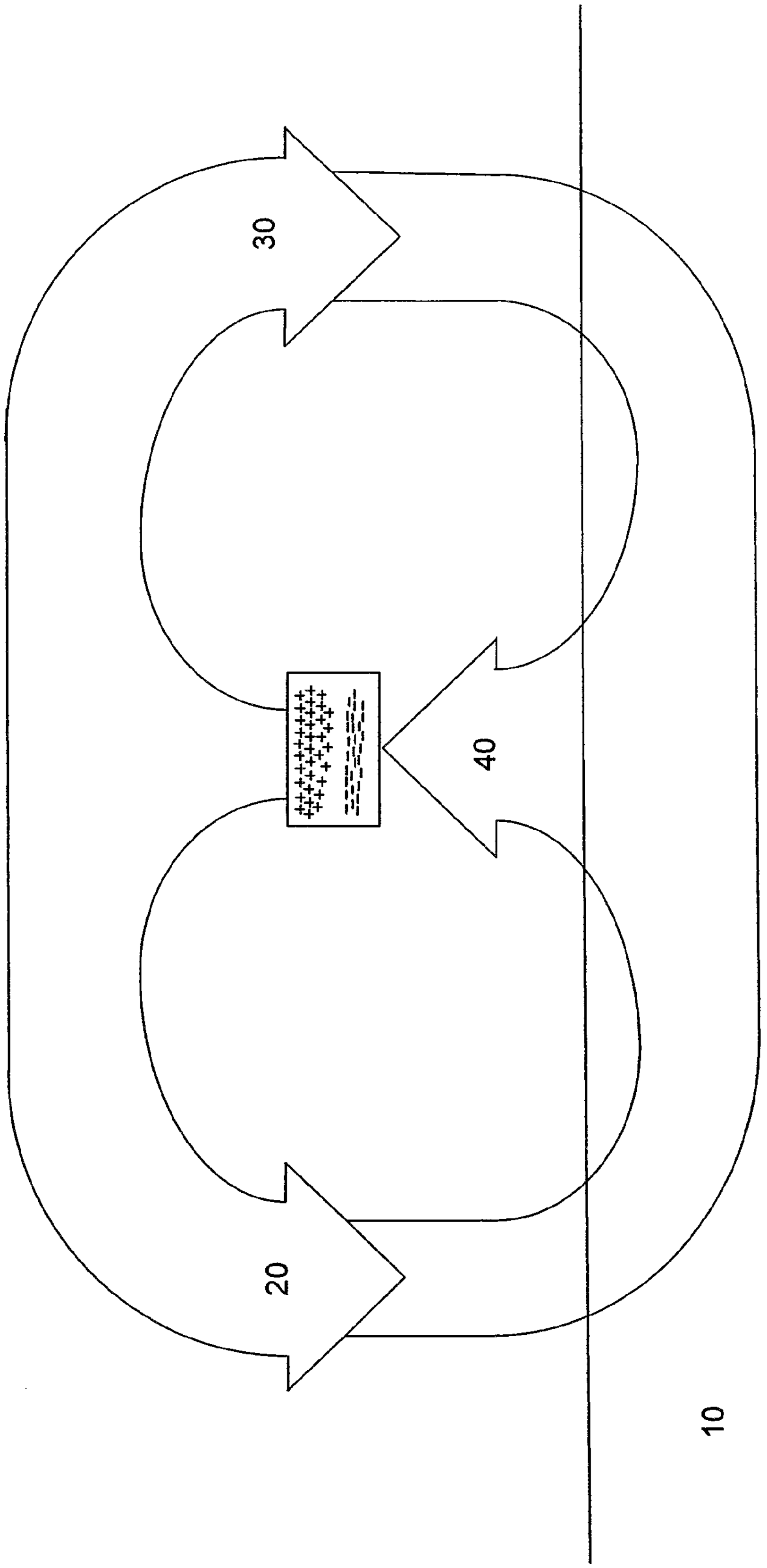

FIGURE 1

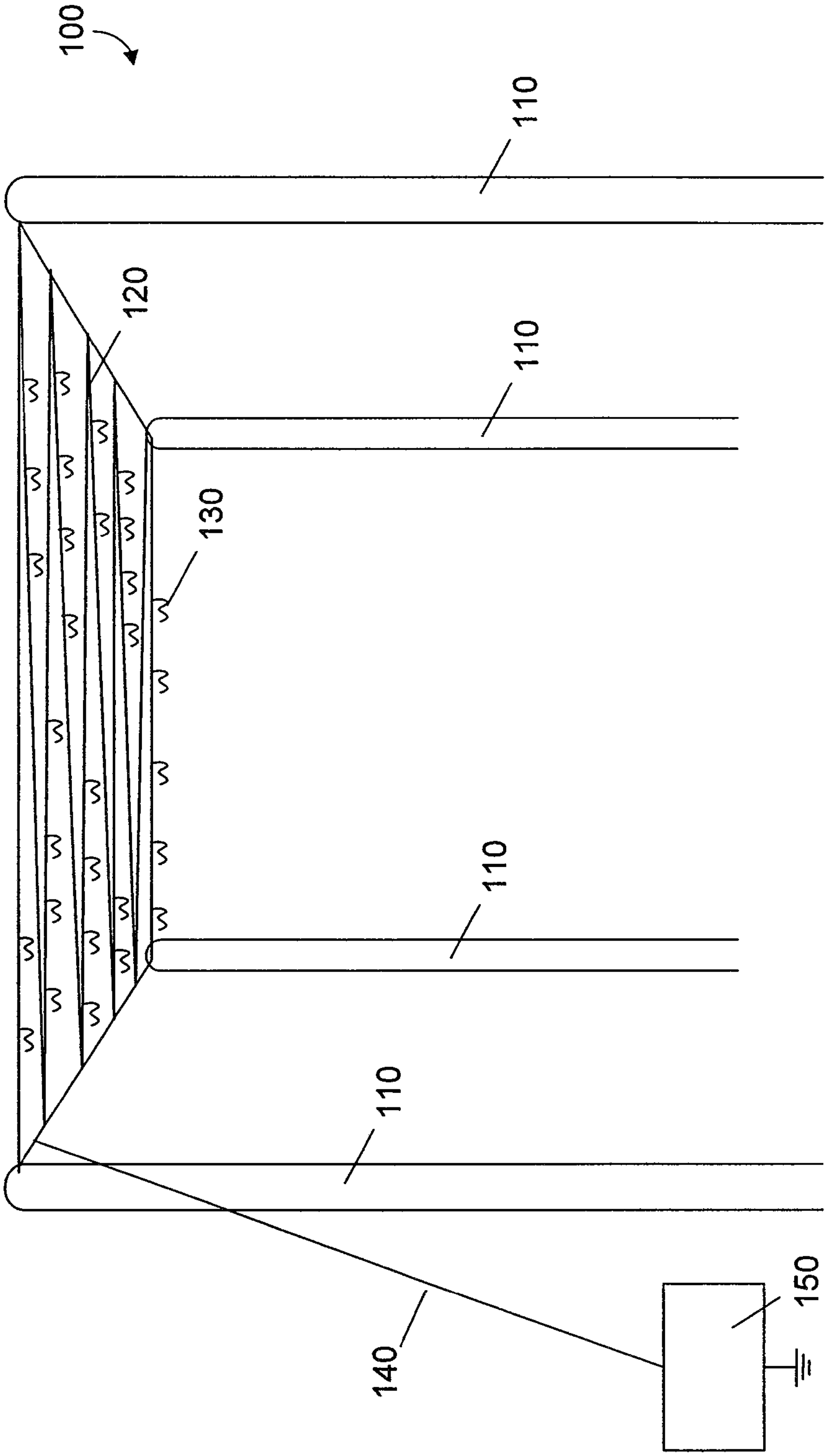
100
110
110
120
130
110
110
140
150

FIGURE 2

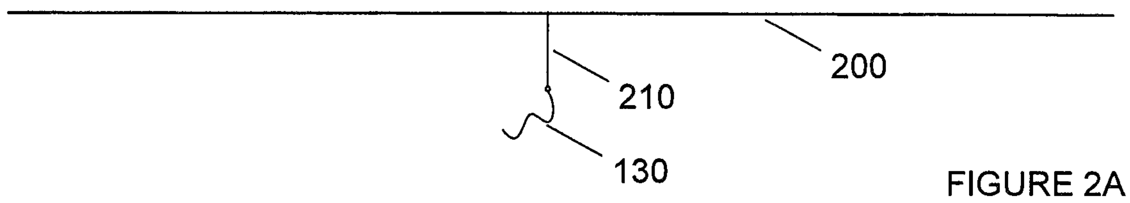

FIGURE 2A

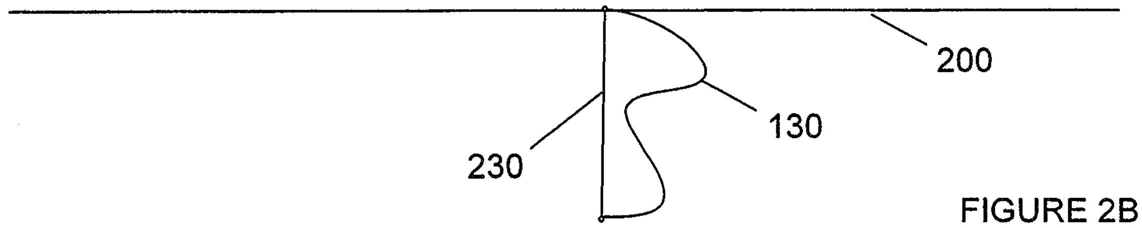

FIGURE 2B

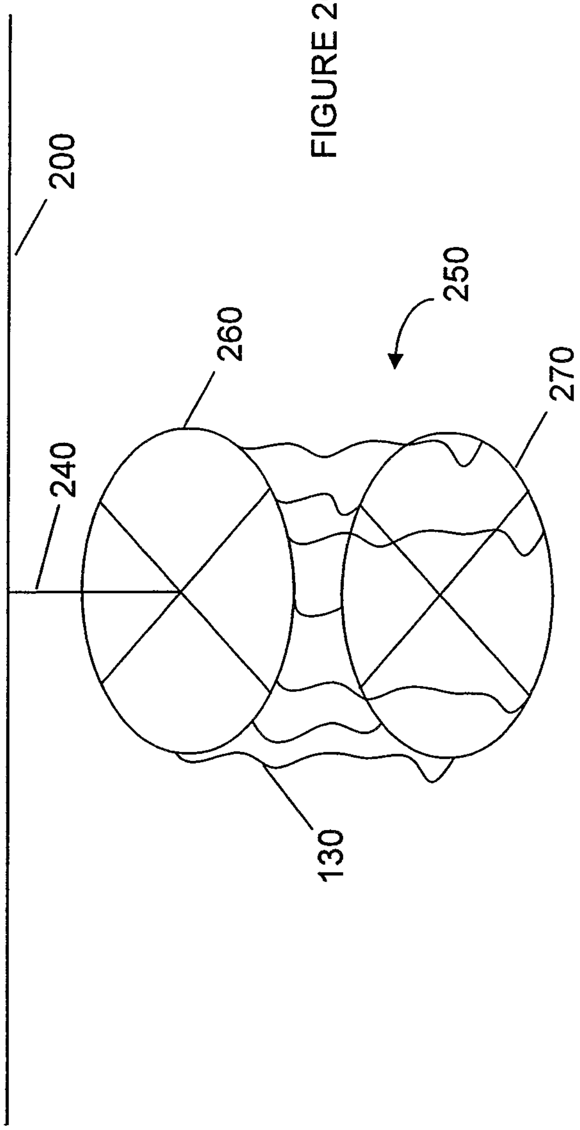
200
240
260
250
270
130

FIGURE 2C

FIGURE 2D

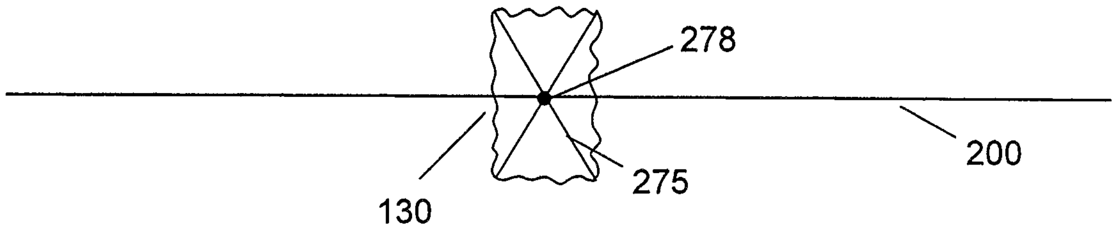
278
200
275
130

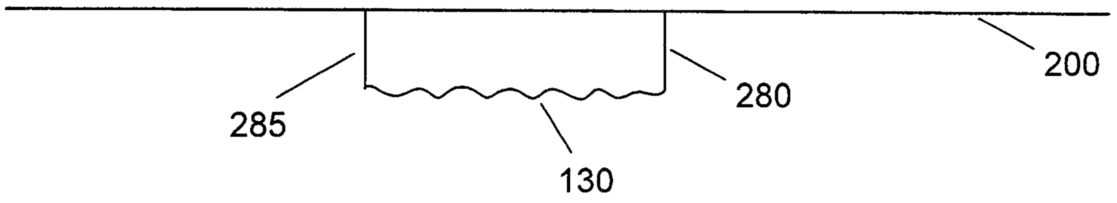
200
285
280
130

FIGURE 2E

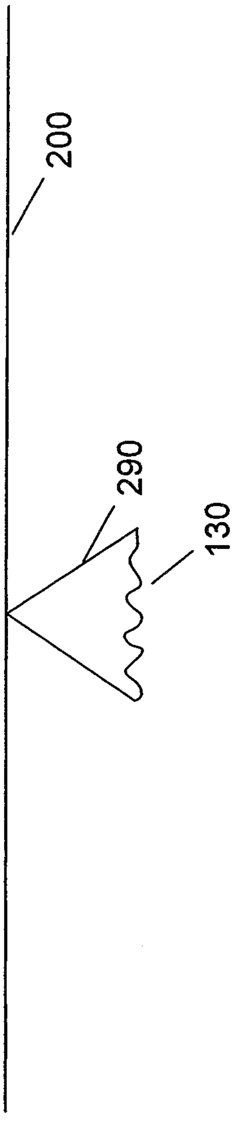

FIGURE 2F

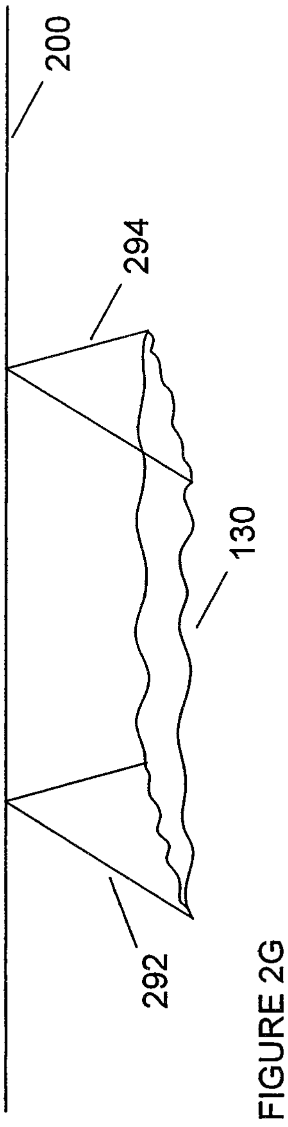
200
294
130
292

FIGURE 2G

FIGURE 3

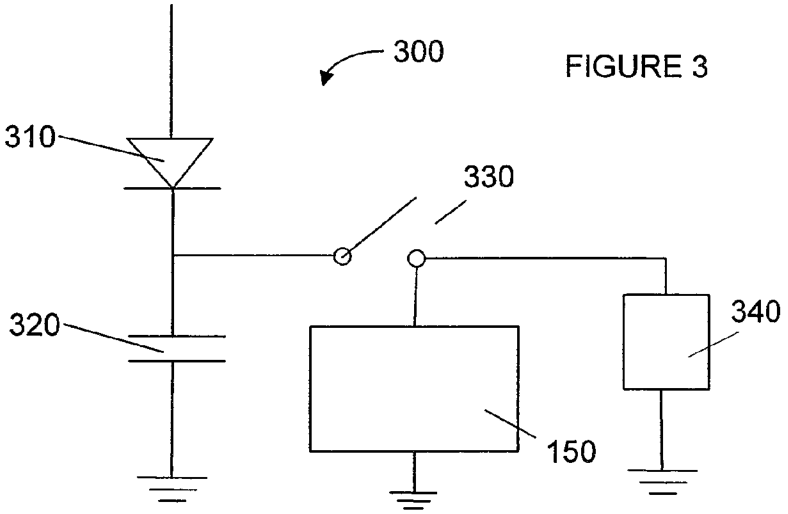
300
310
330
320
340
150

FIGURE 4

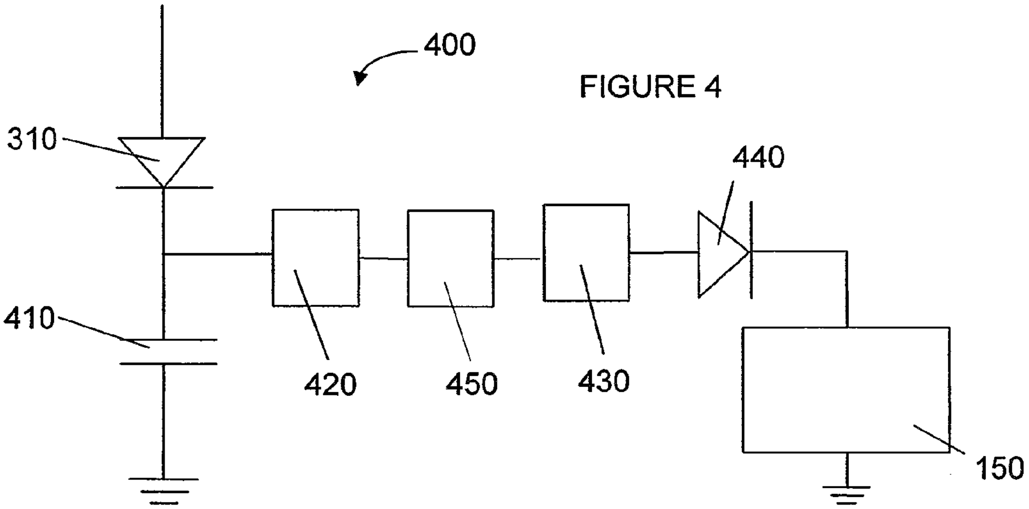
400
310
440
410
420
450
430
150

FIGURE 5

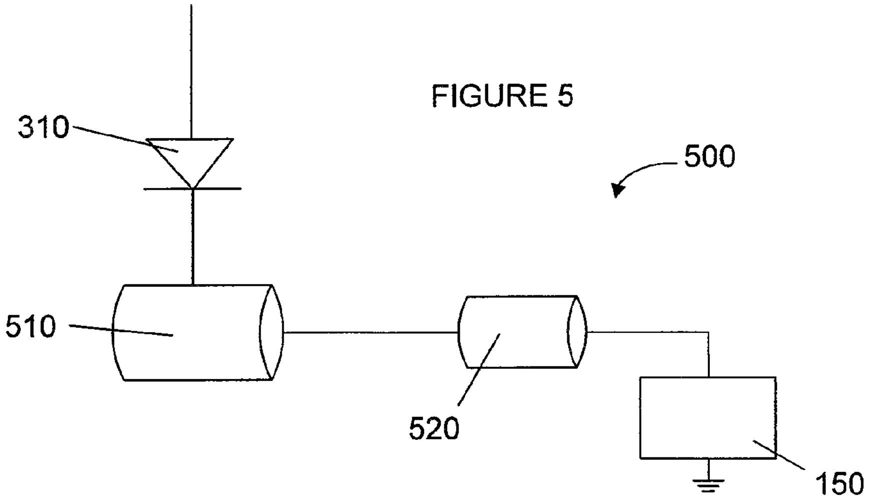
310
500
510
520
150

FIGURE 6

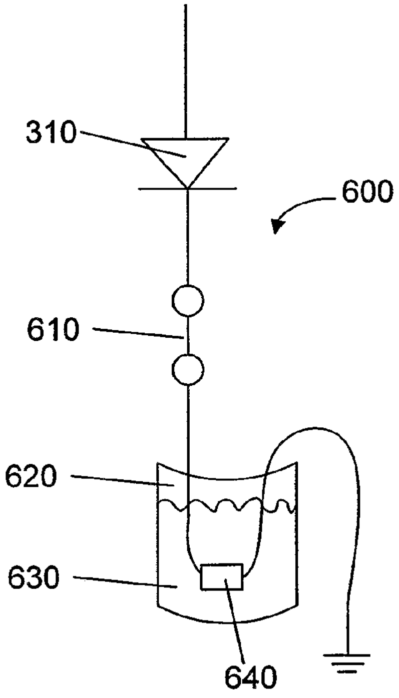
310
600
610
620
630
640

FIGURE 7

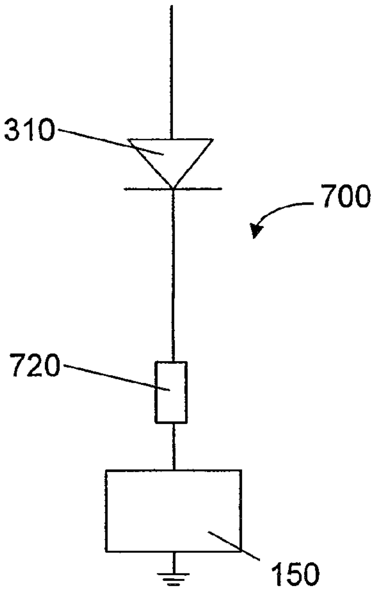
310
700
720
150

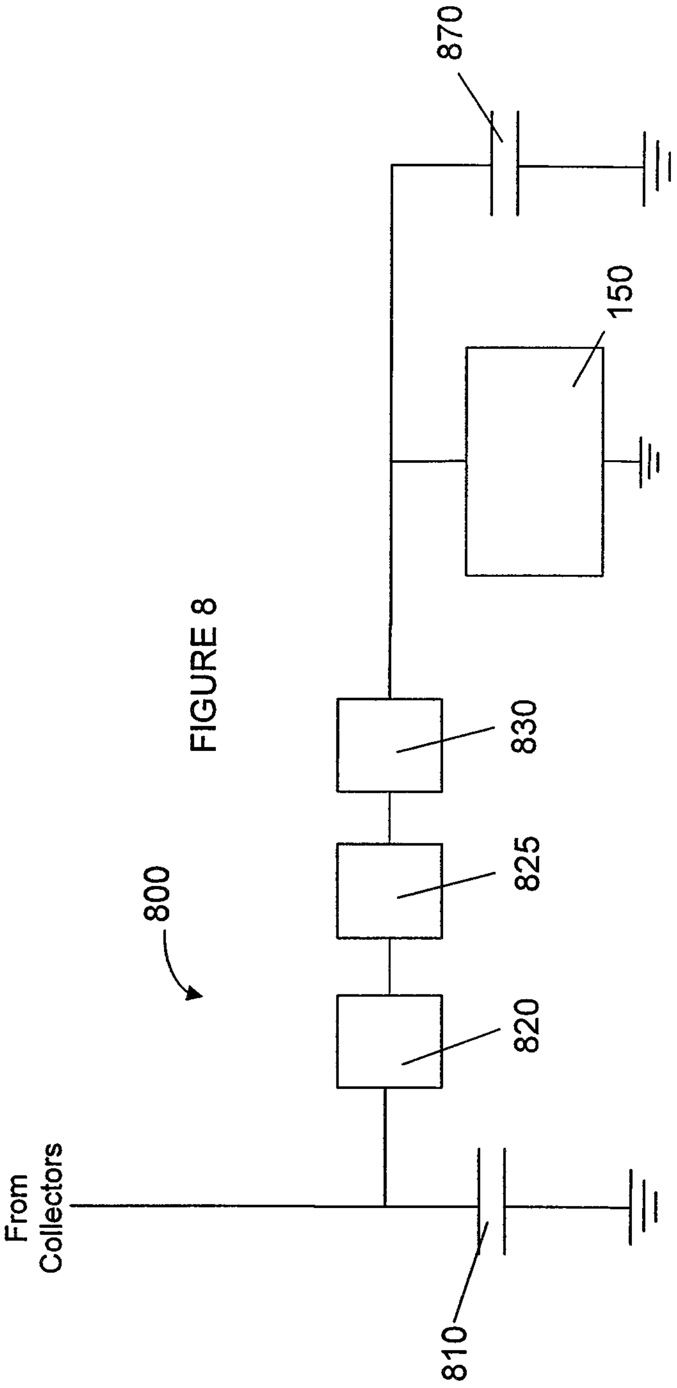
800
From Collectors
810
820
825
830
150
870

FIGURE 8

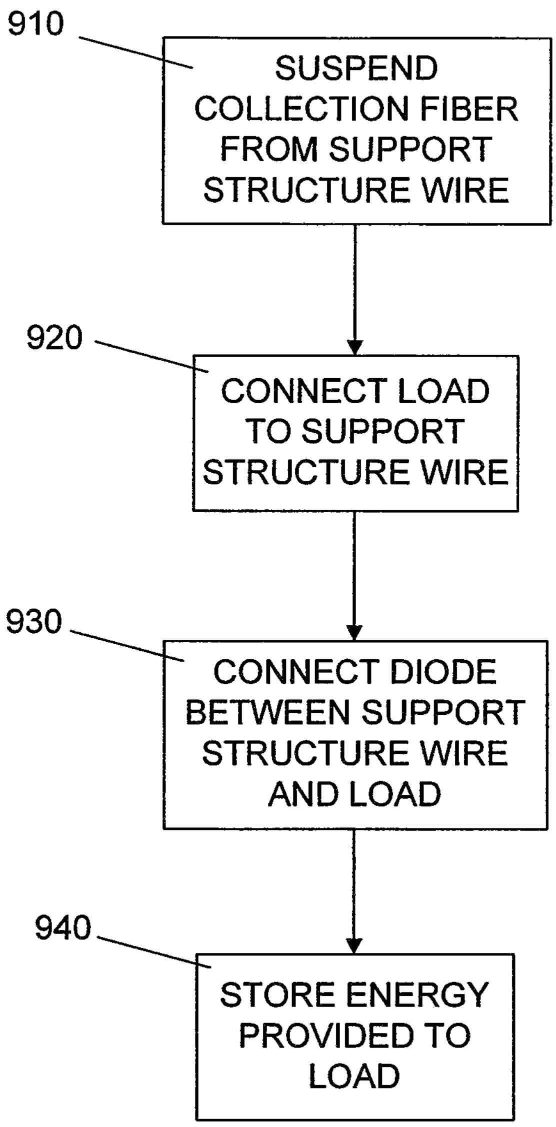
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

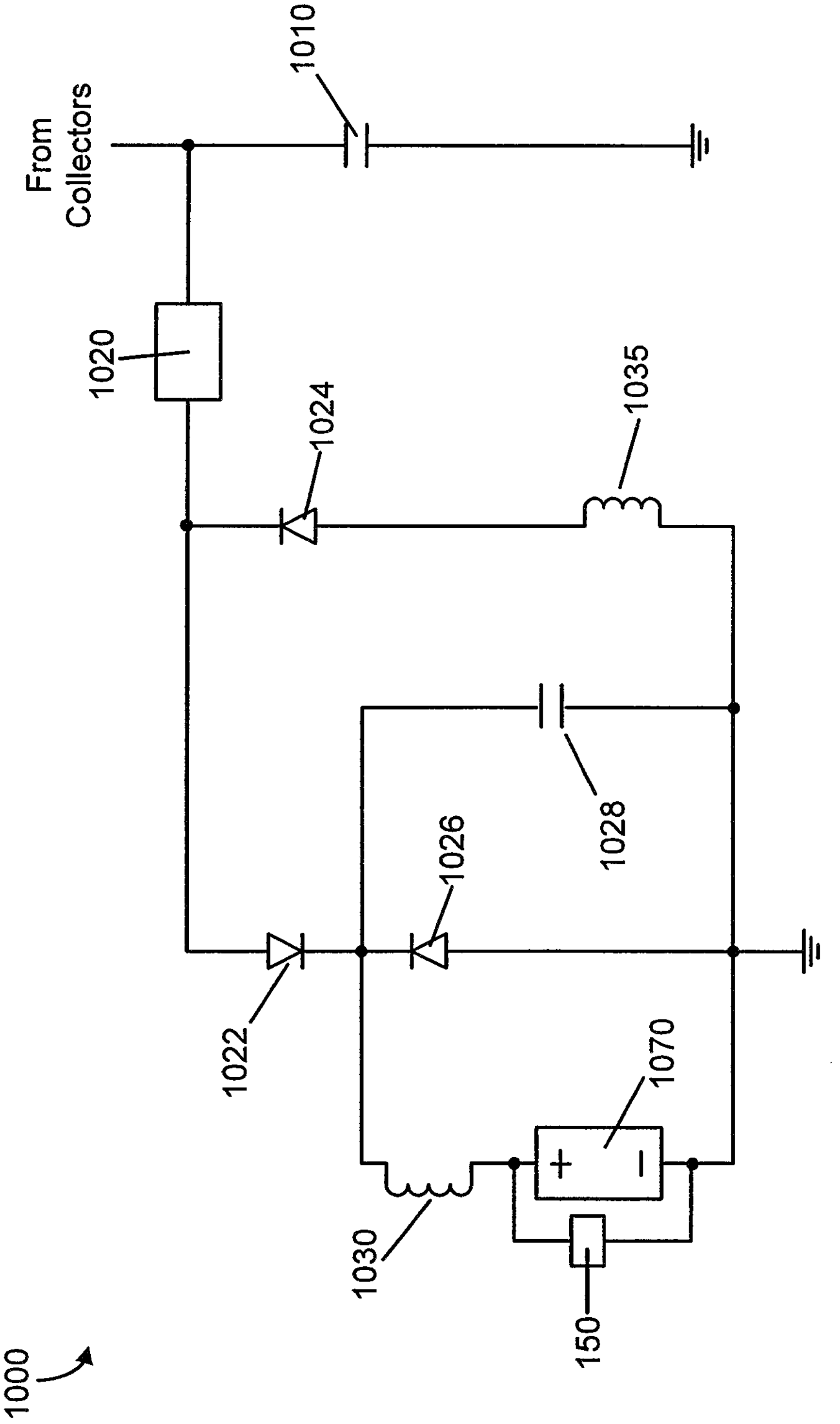
From Collectors
+
-

FIGURE 10

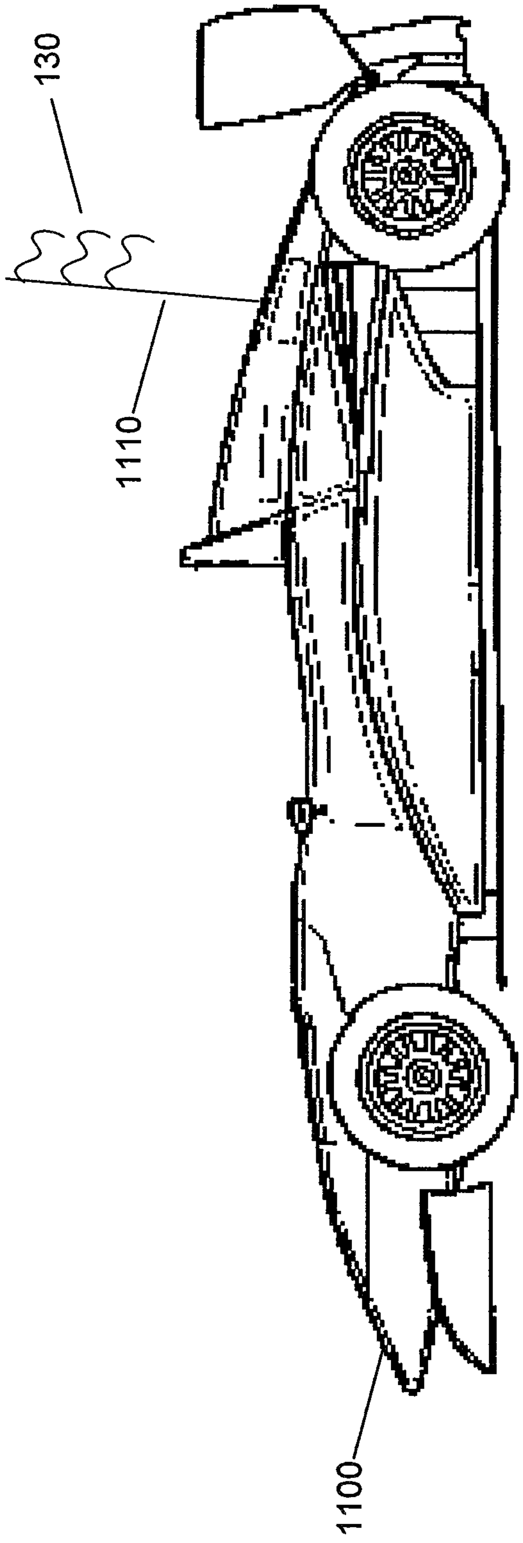
130
1110
1100

FIGURE 11

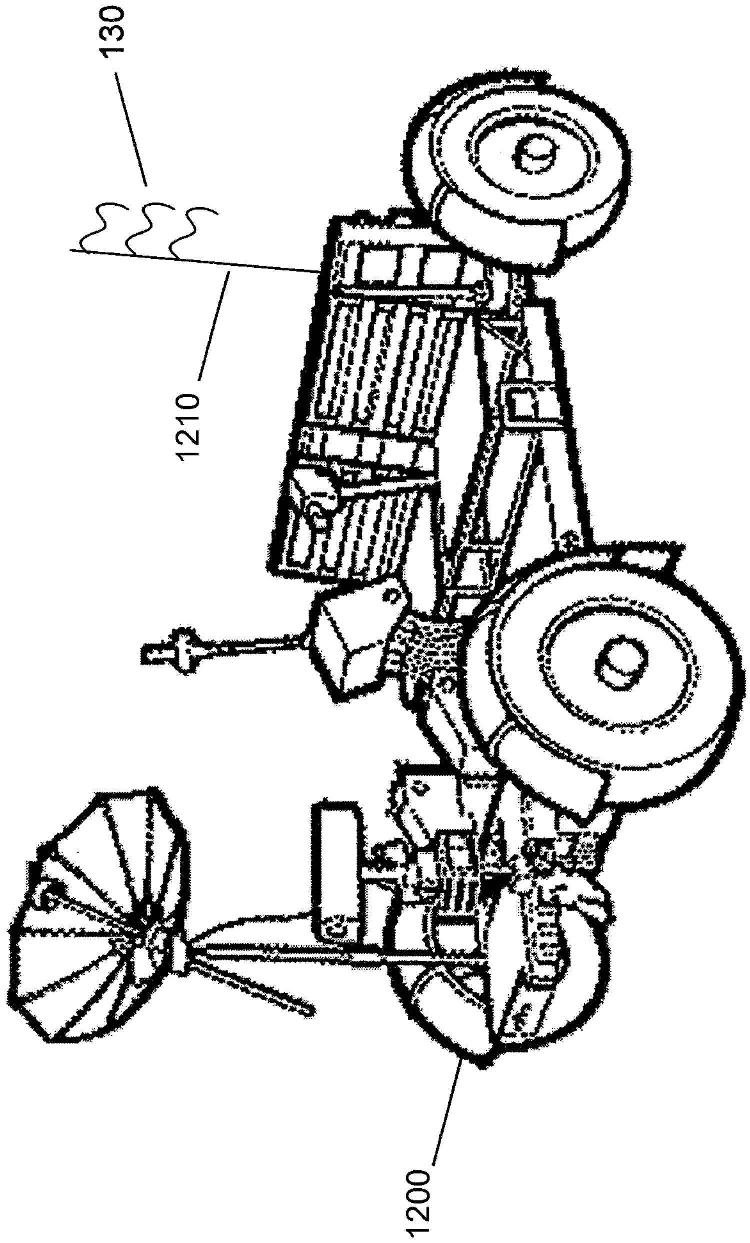
130
1210
1200

FIGURE 12

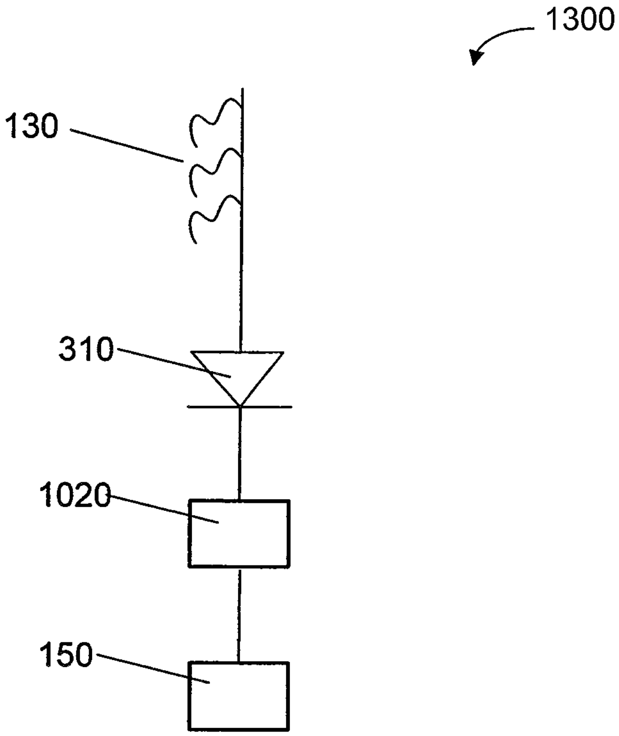
1300
130
310
1020
150

FIGURE 13

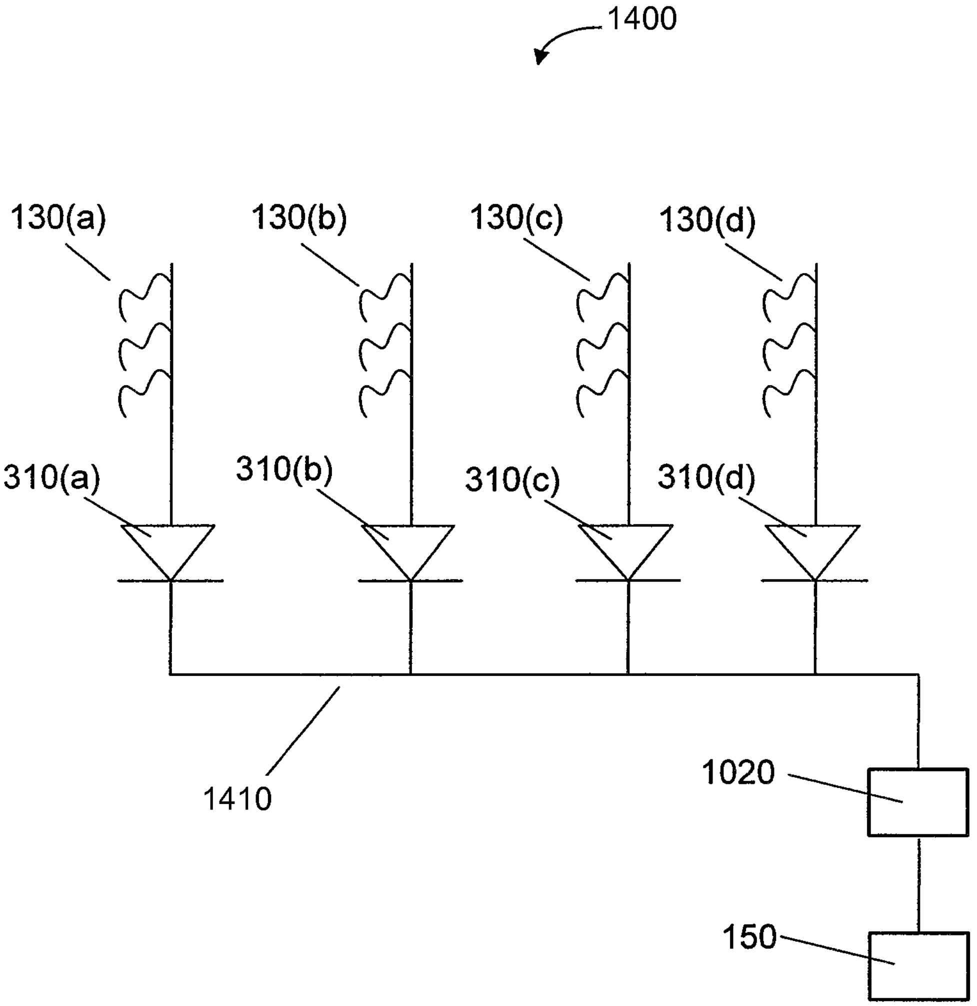
1400
130(a)
130(b)
130(c)
130(d)
310(a)
310(b)
310(c)
310(d)
1410
1020
150

FIGURE 14

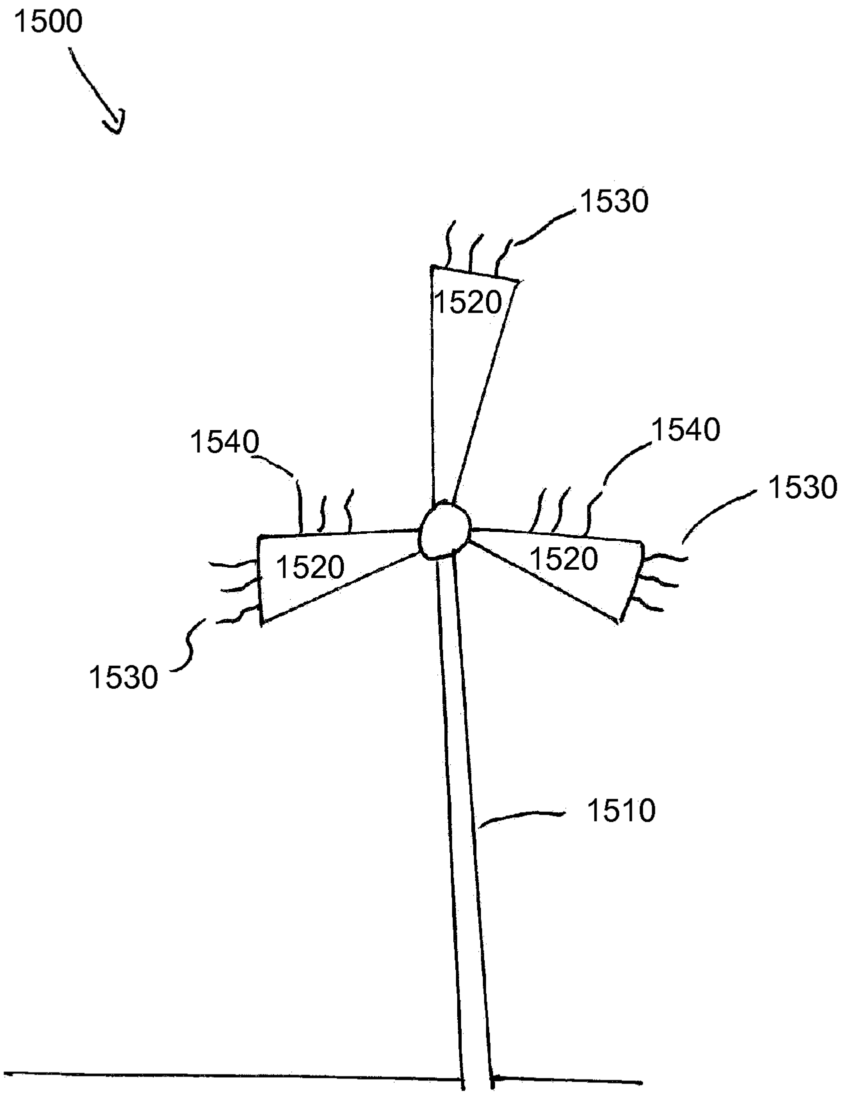
1500
1530
1520
1540
1540
1530
1520
1520
1530
1510

FIGURE 15

ENERGY COLLECTION

TECHNICAL FIELD

[0001] The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

[0002] The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

[0003] In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

[0004] Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

[0005] Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

[0006] While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

[0007] Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure, the support structure comprising at least one of an airplane, drone, blimp, balloon, kite, satellite, train, motorcycle, bike, skateboard, scooter, hovercraft, electronic device, electronic device case, billboard, cell tower, radio tower, camera tower, flag pole, telescopic pole, light pole, utility pole, water tower, building, sky scraper, coliseum, roof top, solar panel and a fixed or mobile structure exceeding 1 inch in height above ground or sea level; at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment electrically connected to the support structure; and a load electrically connected to the at least one collection device.

[0008] Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment from a support structure, the at least one collection device electrically connected to the support structure, the support structure comprising at least one of an airplane, drone, blimp, balloon, kite, satellite, train, motorcycle, bike, skateboard, scooter, hovercraft, electronic device, electronic device case, billboard, cell tower, radio tower, camera tower, flag pole, telescopic pole, light pole, utility pole, water tower, building, sky scraper, coliseum, roof top, solar panel and a fixed or mobile structure exceeding 1 inch in height above ground or sea level; and providing a load with an electrical connection to the at least one collection device to draw current.

[0009] Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

[0011] FIG. **1** is a circuit diagram of a weather energy circuit.

[0012] FIG. **2** is a perspective view of an example embodiment of many energy collectors elevated above ground by a structure.

[0013] FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

[0014] FIG. **2**B is a side view of an example embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

[0015] FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

[0016] FIG. **2**D is a side view of an example embodiment of a support structure for multiple energy collection fibers.

[0017] FIG. **2**E is a side view of a support structure for an energy collection fiber.

[0018] FIG. **2**F is a side view of an example embodiment of a support structure for an energy collection fiber.

[0019] FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

[0020] FIG. **3** is a circuit diagram of an example embodiment of a circuit for the collection of energy.

[0021] FIG. **4** is a circuit diagram of an example embodiment of a circuit for the collection of energy.

[0022] FIG. **5** is a circuit diagram of an example embodiment of an energy collection circuit for powering a generator and motor.

[0023] FIG. **6** is a circuit diagram of an example embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

[0024] FIG. **7** is a circuit diagram of an example embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

[0025] FIG. **8** is a circuit diagram of an example embodiment of a circuit for collecting energy.

[0026] FIG. **9** is a flow diagram of an example embodiment of collecting energy with a collection fiber.

[0027] FIG. **10** is a circuit diagram of an example embodiment of a circuit for collecting energy from a dual polarity source.

[0028] FIG. **11** is a system diagram of an example embodiment of an energy collection system connected to an automobile vehicle.

[0029] FIG. **12** is a system diagram of an example embodiment of an energy collection system connected to a lunar rover vehicle.

[0030] FIG. **13** is a system diagram of an example embodiment of an energy collection system comprising collection devices with a diode.

[0031] FIG. **14** is a system diagram of an example embodiment of an energy collection system comprising multiple legs of the system of FIG. **13**.

[0032] FIG. **15** is a system diagram of an example embodiment of a windmill with energy collectors.

DETAILED DESCRIPTION

[0033] Electric charges on conductors reside entirely on the external surface of the conductors, and tend to concentrate more around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An example embodiment of this disclosure takes advantage of this property, among others, to collect and use the energy generated by an electric field in the atmosphere. Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An example embodiment utilizes carbon or graphite fibers for static electricity collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including aluminum or steel, but most notably, copper. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

[0034] An example embodiment of the collection fibers as collection device **130** includes graphite or carbon fibers. Graphite and carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, or graphite fibers are examples that demonstrate collection ability.

[0035] In at least one example embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than 100 volts per meter under some conditions. When support wire **120** is suspended in the air at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

[0036] A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An example embodiment of the collection device comprises the diode with no collection fiber. A preferred embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restricts energy that may be collected through one fiber from escaping through another fiber.

[0037] Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2**A-**2**G using a collection fiber embodiment. FIG. **2**A presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**.

[0038] Likewise, FIG. **2**B shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

[0039] FIG. 2C presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

[0040] FIG. 2D presents another example embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

[0041] FIG. 2E provides another example embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

[0042] FIG. 2F presents another example embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

[0043] FIG. 2G provides two supports as provided in FIG. 2F such that at least two support members **292**, **294** may be connected to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and between multiple support structures. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**.

[0044] FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load. Capacitor **320** maybe connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

[0045] In an example embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel. Non-limiting examples of spark gap **420** include mercury-reed switches and mercury-wetted reed switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground. The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

[0046] FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

[0047] FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes maybe placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches and mercury-wetted reed switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel.

[0048] FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit. Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

[0049] FIG. **8** presents example circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches and mercury-wetted reed switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820**

to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

[0050] FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

[0051] FIG. **10** presents circuit **1000** as an example embodiment for the collection of energy from a dual polarity source. This may be used, for example, to collect atmospheric energy that reverses in polarity compared with the ground. Such polarity reversal has been discovered as occurring occasionally on Earth during, for example, thunderstorms and bad weather, but has also been observed during good weather. Such polarity reversal may occur on other planetary bodies, including Mars and Venus, as well. Energy polarity on other planets, in deep space, or on other heavenly bodies, may be predominantly negative or predominantly positive. Collector fibers (**130** from FIG. **2**), which may comprise graphene, silicene, and/or other like materials, are capable of collecting positive energy and/or negative energy, and circuit **1000** is capable of processing positive and/or negative energy, providing an output which is always positive. Circuit **1000** may collect energy from one or more collection devices (**130** from FIG. **2**). Charging capacitor **1010** may be used to store a charge until the voltage at spark gap **1020** achieves the spark voltage. Capacitor **1010** is optional.

[0052] A plurality of diodes may be placed in a plurality of positions in circuit **1000**. The voltage from capacitor **1010** may be used to charge spark gap **1020** to a sufficient voltage. Spark gap **1020** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **1020** include mercury-reed switches, mercury-wetted reed switches, open-gap spark gaps, and electronic switches. When spark gap **1020** arcs, energy will arc from an emitting end of spark gap **1020** to a receiving end of spark gap **1020**. The output of spark gap **1020** is electrically connected to the anode of diode **1022** and the cathode of diode **1024**. The cathode of diode **1022** is electrically connected to the cathode of diode **1026** and inductor **1030**. Inductor **1030** may be a fixed value inductor or a variable inductor. The anode of diode **1026** is electrically connected to ground. Capacitor **1028** is electrically connected between ground and the junction of the cathodes of diode **1022** and diode **1026**. Inductor **1035** is electrically connected between ground and the anode of diode **1024**. Inductor **1035** may be a fixed value inductor or a variable inductor. Capacitor **1070**, the anode of diode **1026**, inductor **1035**, and load **1050** are electrically connected to ground. Capacitor **1070** may be placed in parallel with load **150**.

[0053] FIGS. **11** and **12** provide example embodiments of vehicle **1110**, which utilizes electricity, the vehicle employing systems of energy collection provided herein. Vehicle **1100** in FIG. **11** is shown as an automobile vehicle, but could be any means of locomotion that utilizes electricity, including a car, a train, a motorcycle, a boat, an airplane, robotic rovers, space craft, etc. Vehicle **1200** in FIG. **12** is shown as a lunar rover vehicle. In FIGS. **11** and **12**, support rod **1110**, **1210** is electrically connected to an electrical system in vehicle **1100**, **1200**. Energy collectors **130**, which may comprise graphene, silicene, and/or other like materials, are electrically connected to support rod **1110**, **1210** and may be used to supply energy to electrical circuits within the vehicle. A non-limiting use includes a top-off charge for a battery system, on-board hydrogen production, and/or assisting in the same. Energy collectors **130** may be used to augment the efficiency of the locomotion that utilizes electrical energy as well.

[0054] FIG. **13** provides an example embodiment of energy collection system **1200** in which diode **310** is used to isolate collection devices **130** from spark gap **1020** and load **150**. Collection devices **130** may comprise graphite, carbon fibers, carbon/carbon fibers, graphene, silicene, and/or other like materials, or a mixture thereof.

[0055] FIG. **14** provides an example embodiment of energy collection system **1400** in which a plurality of energy collection systems, such as that provided in FIG. **13**, are combined. Each leg consisting of collection devices **130**, which may comprise graphene, silicene, and/or other like materials, and diode **310** are connected in parallel with other legs, each leg electrically connected to trunk wire **1410**. The legs could also be connected serially. Trunk wire **1410** is electrically connected to a collection circuit, which may comprise load **150** and spark gap **1020** in any configuration that has been previously discussed. Each leg may comprise one or more collection devices **130** and at least one diode electrically connected between the collection devices and the collection circuit. Although three collection devices **130** are shown on each leg, any number of collection devices may be used. Although four legs are shown, any number of legs may be used.

[0056] FIG. **15** presents a system diagram of an example embodiment of a windmill with energy collectors, which may comprise graphene, silicene, and/or other like materials in an example embodiment. A windmill is an engine powered by the energy of wind to produce alternative forms of energy. They may, for example, be implemented as small tower mounted wind engines used to pump water on farms. The modern wind power machines used for generating electricity are more properly called wind turbines. Common applications of windmills are grain milling, water pumping, threshing, and saw mills. Over the ages, windmills have evolved into more sophisticated and efficient wind-powered water pumps and electric power generators. In an example embodiment, as provided in FIG. **10**, windmill tower **1500** of suitable height and/or propeller **1520** of windmill tower **1500** may be equipped with energy collecting fibers **1530**, **1540**, which may comprise graphene, silicene, and/or other like materials in an example embodiment. Collecting fibers **1530**, **1540** may turn windmill **1500** into a power producing asset even when there is not enough wind to turn propellers **1520**. During periods when there is enough wind to turn propellers **1520**, collecting fibers **1530**, **1540** may supplement/boost the amount of energy the windmill produces.

[0057] A windmill is an engine powered by the energy of wind to produce alternative forms of energy. They may, for example, be implemented as small tower mounted wind engines used to pump water on farms. The modern wind power machines used for generating electricity are more properly called wind turbines. Common applications of windmills are grain milling, water pumping, threshing, and saw mills. Over the ages, windmills have evolved into more sophisticated and efficient wind-powered water pumps and electric power generators. In an example embodiment, as provided in FIG. **10**, windmill tower **1000** of suitable height and/or propeller **1020** of windmill tower **1000** may be equipped with energy collecting fibers **1030**, **1040**. Collecting fibers **1030**, **1040** may turn windmill **1000** into a power producing asset even when there is not enough wind to turn propellers **1020**. During periods when there is enough wind to turn propellers **1020**, collecting fibers **1030**, **1040** may supplement/boost the amount of energy the windmill produces.

[0058] Windmill **1500**, properly equipped with ion collectors **1530**, **1540**, such as a non-limiting example of fibers with graphene, silicene, and/or other like materials, can produce electricity: 1) by virtue of providing altitude to the fiber to harvest ions, and 2) while the propeller is turning, by virtue of wind blowing over the fiber producing electricity, among other reasons, via the triboelectric effect (however, it is also possible for the triboelectric effect to occur, producing electricity, in winds too weak to turn the propeller).

[0059] There are at least two ways that energy collectors may be employed on or in a windmill propeller to harvest energy. Propellers **1520** may be equipped with energy collectors **1530**, **1540** attached to, or supported by, propeller **1520** with wires (or metal embedded in, or on propeller **1520**) electrically connecting energy collectors **1530**, **1540**, which may comprise graphene, silicene, and/or other like materials, to a load or power conversion circuit. There may be a requirement to electrically isolate energy collectors **1530**, **1540**, which are added to propeller **1520**, from electrical ground, so that the energy collected does not short to ground through propeller **1520** itself or through support tower **1510**, but rather is conveyed to the load or power conversion circuit. Energy collectors may be connected to the end of propellers **1520** such as collectors **1530**. Alternatively, energy collectors may be connected to the sides of propellers **1520** such as collectors **1540**.

[0060] Alternatively, propeller **1520** may be constructed of carbon fiber or other suitable material, with wires (or the structural metal supporting propeller **1520** may be used) electrically connecting to a load or power conversion circuit. In the case of propeller **1520** itself being constructed of carbon fiber, for example, the fiber may be ‘rough finished’ in selected areas so that the fiber is “fuzzy.” For example, small portions of it may protrude into the air as a means of enhancing collection efficiency. The fuzzy parts of collectors **1530**, **1540** may do much of the collecting. There may be a requirement to electrically isolate carbon fiber propeller **1520** from electrical ground, so that the energy it collects does not short to ground through metal support tower **1510**, but rather is conveyed to the load or power conversion circuit. Diodes may be implemented within the circuit to prevent the backflow of energy, although diodes may not be necessary in some applications.

[0061] In an alternative embodiment, windmill **1500** may be used as a base on which to secure an even higher extension tower to support the energy collectors and/or horizontal supports extending out from tower **1510** to support the energy collectors. Electrical energy may be generated via ion collection due to altitude and also when a breeze or wind blows over the collectors supported by tower **1510**.

[0062] In alternative embodiments to windmill **1500**, other non-limiting example support structures include airplanes, drones, blimps, balloons, kites, satellites, cars, boats, trucks, (including automobile and other transportation conveyance tires), trains, motorcycles, bikes, skateboards, scooters, hovercraft (automobiles and conveyance of any kind), billboards, cell towers, radio towers, camera towers, flag poles, towers of any kind including telescopic, light poles, utility poles, water towers, buildings, sky scrapers, coliseums, roof tops, solar panel and all fixed or mobile structures exceeding 1 inch in height above ground or sea level.

[0063] An example embodiment of a support structure may also include cell phones and other electronic devices and their cases, including cases containing rechargeable batteries. For example, someone may set her cell phone or other electronic device or battery pack on the window ledge of a tall apartment building to help charge it. Other example support structures may include space stations, moon and Mars structures, rockets, planetary rovers and drones including robots and artificial intelligence entities.

[0064] Under some conditions, ambient voltage may be found to be 180-400 volts at around 6 ft, with low current. With the new generation of low current devices being developed, a hat containing ion harvesting material may provide enough charge, or supplemental charge, collected over time to help power low current devices such as future cell phones, tracking devices, GPS, audio devices, smart glasses, etc. Clothes may also be included as examples of support structures. Friction of the ion collection material (such as non-limiting examples of carbon, graphite, silicene and graphene) against unlike materials, such as wool, polyester, cotton, etc (used in clothes) may cause a voltage to be generated when rubbed together. Additionally, wind passing over the ion collection material has been demonstrated to generate voltage, even at low altitude. In an additional example embodiment, embedding collection devices into automobile tires (for example, in a particular pattern) could generate collectible voltage.

[0065] Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

[0066] It should be emphasized that the above-described embodiments of the present disclosure, particularly, any “preferred” embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A method of collecting energy comprising:

suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment from a support structure, the at least one collection device electrically connected to the support structure, the support structure comprising at least one of an airplane, drone, blimp, balloon, kite, satellite, train, motorcycle, bike, skateboard, scooter, hovercraft, electronic device, electronic device case, billboard, cell tower, radio tower, camera tower, flag pole, telescopic pole, light pole, utility pole, water tower, building, sky scraper, coliseum, roof top, solar panel and a fixed or mobile structure exceeding **1** inch in height above ground or sea level; and

providing a load with an electrical connection to the at least one collection device to draw current.

2. The method of claim **1**, wherein the collection device collects energy by triboelectric effect.

3. The method of claim **1**, wherein the collection device comprises a collection fiber.

4. The method of claim **1**, wherein the collection device comprises a diode and a collection fiber and the diode is electrically connected between the collection fiber and the load.

5. The method of claim **1**, further comprising storing energy provided to the load.

6. The method of claim **5**, wherein storing energy provided to the load comprises storing energy in a capacitor or an inductor.

7. The method of claim **3**, wherein the collection fiber comprises at least one of carbon, graphite, silicene and graphene.

8. A system of energy collection comprising:

a support structure, the support structure comprising at least one of an airplane, drone, blimp, balloon, kite, satellite, train, motorcycle, bike, skateboard, scooter, hovercraft, electronic device, electronic device case, billboard, cell tower, radio tower, camera tower, flag pole, telescopic pole, light pole, utility pole, water tower, building, sky scraper, coliseum, roof top, solar panel and a fixed or mobile structure exceeding 1 inch in height above ground or sea level;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment electrically connected to the support structure; and

a load electrically connected to the at least one collection device.

9. The system of claim **8**, wherein the collection device collects energy by triboelectric effect.

10. The system of claim **8**, wherein the collection device comprises a collection fiber.

11. The system of claim **8**, wherein the collection device comprises a collection fiber and a diode electrically connected between the load and the collection fiber.

12. The system of claim **11**, wherein the diode is elevated relative to the ground level.

13. The system of claim **10**, wherein the collection fiber comprises at least one of carbon, graphite, silicene, or graphene.

14. The system of claim **8**, further comprising a diode electrically connected between the at least one collection device and the support structure.

15. The system of claim **8**, further comprising: a switch connected in series between the at least one collection device and the load; and a capacitor connected in parallel with the switch and the load.

16. The system of claim **15**, wherein the switch comprises an interrupter connected between the load and at least one collection device, and wherein the interrupter comprises at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

17. The system of claim **8**, further comprising: a motor for providing power, the motor connected between the at least one collection device and the load; and a generator powered by the motor.

18. The system of claim **8**, further comprising a fuel cell between the support structure and the load.

19. The system of claim **18**, wherein the fuel cell produces hydrogen and oxygen.

20. A system of collecting energy comprising:

means for suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to the means for suspending, the means for suspending comprising at least one of an airplane, drone, blimp, balloon, kite, satellite, train, motorcycle, bike, skateboard, scooter, hovercraft, electronic device, electronic device case, billboard, cell tower, radio tower, camera tower, flag pole, telescopic pole, light pole, utility pole, water tower, building, sky scraper, coliseum, roof top, solar panel and a fixed or mobile structure exceeding 1 inch in height above ground or sea level;

means for inducing current flow, the means for inducing current flow electrically connected to the means for suspending; and

means for restricting the backflow of charge carriers, the means for restricting the backflow of charge carriers electrically connected between the at least one collection device and the means for inducing current flow.

* * * * *

Boise State University
ScholarWorks

Electrical and Computer Engineering Faculty Publications and Presentations

Department of Electrical and Computer Engineering

1-1-2016

Energy Efficient Switched Parasitic Array Antenna for 5G Networks and IoT

Ahmed Kausar
Boise State University

Hani Mehrpouyan
Boise State University

Mathini Sellathurai
Boise State University

Rongrong Qian
Heriot Watt University

Shafaq Kausar
National University of Sciences and Technology

© 2016 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. doi: 10.1109/LAPC.2016.7807569

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at *2016 Loughborough Antennas & Propagation Conference (LAPC)*, published by IEEE. Copyright restrictions may apply. doi: 10.1109/LAPC.2016.7807569

Energy Efficient Switched Parasitic Array Antenna for 5G Networks and IoT

Ahmed Kausar*, Hani Mehrpouyan *, Mathini Sellathurai†, Rongrong Qian†, and Shafaq Kausar‡
*Department of Electrical and Computer Engineering
Boise State University, Boise, USA
Email: ahmedkausar@boisestate.edu, hanimehrpouyan@boisestate.edu
† School of Engineering and Physical Sciences
Heriot Watt University, Edinburgh, U.K
Email: m.sellathurai@hw.ac.uk, rq31@hw.ac.uk
‡Department of Electrical Engineering
National University of Sciences and Technology, Islamabad, Pakistan
Email: shafaqkausar@nust.edu.pk

***Abstract*—This paper includes design and implementation result of an adaptive beam forming antenna for upcoming 5G and Internet of Things (IoT). Switched parasitic array antennas are low cost, small sized and compact circular array antennas that steer beam in a desired direction by variation in switching pattern of parasitic elements. The proposed antenna design has an active center element, which is surrounded by several symmetrically placed parasitic elements. The designed antenna has a gain of 8 dB and is capable of 360 degrees beam steering in steps of 60 degrees each. Simulations are validated with results of the fabricated antenna. Antenna beam is steered by controlling parasitic elements. Future application of Electronically Steerable Parasitic Array Radiator (ESPAR) antennas and switched parasitic array antennas in next generation communication networks and methods for reducing size of the antenna are also highlighted.**

***Index Terms*—Internet of Things, 5G, circular array, reconfigurable, smart antenna, adaptive beamforming, Switched-Parasitic Array**

I. INTRODUCTION

Smart antennas are capable of adaptive beam forming in accordance with the environment. Over the last three decades wireless communication has made rapid progress due to better modulation schemes and better error correction schemes. Wireless communication systems are replacing the wired ones. Antenna is an integral part of wireless devices. To meet the ever-increasing demand of higher data rate, smart antennas capable of adaptive beam forming are introduced as a potential solution to increasing data rate and improving signal-to-noise ratio (SNR) in upcoming 5G and Internet of Things (IoT) [1]. By using smart antennas Equivalent Isotropically Radiated power (EIRP) can be increased. EIRP denotes product of the transmit power (P_t) and transmit antenna gain (G_t), i.e.

$$EIRP = P_t \times G_t \tag{1}$$

From (1), it is observed that by using a smart antenna G_t increases, which in turn leads to the increase in the EIRP. The beam of an antenna can be steered by means of mechanical movement of a directional antenna or by electronic beam steering. Electronic beam steering is resilient and provides more accurate approximation of the desired signal source. Moreover electronic beam steering antenna does not involve moving parts. Therefore, there is less wear and tear on the antenna as compared to mechanical beam steering. Generally electronic beam forming is achieved by using phase shifters coupled with an array of antenna elements. The phase of an individual phase shifter is varied such that there is net constructive interference in the intended direction of the antenna beam and destructive interference in all other directions [2]. Hence, electronic beam steering is preferred over mechanical beam steering.

Electronically Steerable Parasitic Array Radiator (ESPAR) antennas use mutual coupling between antenna elements to steer the beams instead of phase shifters. Design of parastic array radiator antennas is discussed in [3], [4] in which there is a center active element surrounded by a circular array of passive elements, where each passive element is loaded with a specific reactance. ESPAR antennas are smaller in size as compared to phase array antennas because elements have to be placed at a quarter of wavelength or less distance from each other for effective mutual coupling. Hence, ESPAR antennas are cost effective and space efficient. In such antennas we have only one active element, therefore, only one feed is required. Consequently, ESPAR antennas have lower losses (due to single feed) as compared to phase array antennas [3].

Previously designed antenna Electronically Steerable Parasitic Raidator(ESPAR) antennas have all cylindrical element, including the feed element [6], [7]. In this paper we have proposed design of seven element ESPAR antenna with conical center(active element)[1]. Length of conical element is optimized using optimetrics in HFSS. Our proposed antenna has higher bandwidth as compared to previously designed ESPAR antennas. [6], [7]. In multiple-input-multiple-output (MIMO)

---

[1] This work was supported by National Science Foundation (NSF) on Enhancing Access to Radio Spectrum (EARS)

978-1-5090-0783-7/16/$31.00 ©2016 IEEE

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at *2016 Loughborough Antennas & Propagation Conference (LAPC)*, published by IEEE. Copyright restrictions may apply. doi: 10.1109/LAPC.2016.7807569

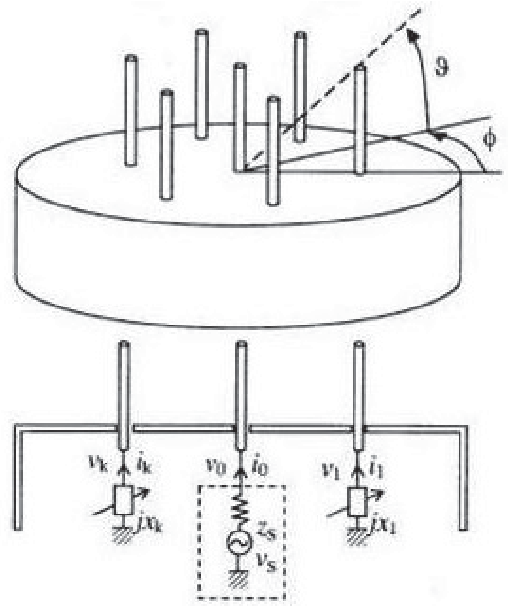

Fig. 1. Seven Element switched parasitic array antenna, ϕ is azimuth angle and θ is elevation angle.

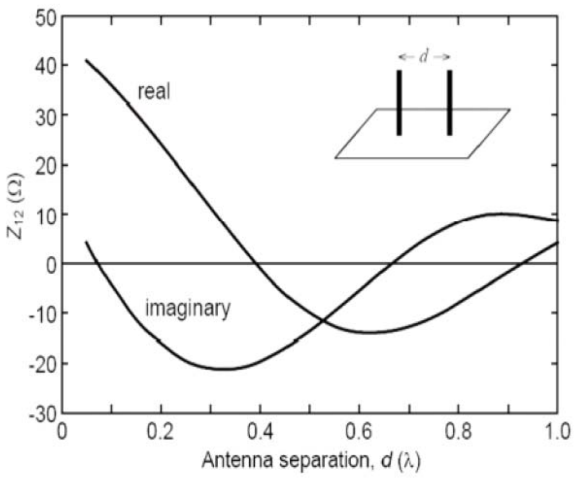

Fig. 2. Z_{21} versus separation distance.

systems there are multiple transmit and and receive antennas. Each receive antenna receives signal over different propagation channels yielding diversity gain. The MIMO technique can increase system throughput and reduce propagation losses. Similarly diversity gain can also be achieved by using smart antennas. In particular, we can transmit the same signal in multiple directions through directional beams, such that they experience different propagation channels [4]. Our proposed antenna is more suited for MIMO applications compared to previously designed model because of its higher gain and improved bandwidth efficiency.

This paper focuses on the design of a seven element switched parasitic array antenna. The antenna is first simulated and then it is fabricated. Simulation results are validated with the hardware results. Previously done work on ESPAR and switched array antennas involve cylindrical monopole elements mounted on a ground plain [8] due to cylindrical nature bandwidth of active monopole is confined, subsequently bandwidth of switched array and ESPAR antennas is limited. Instead on cylindrical monopole element we have used tapered conical monopole. Due to tapered design bandwidth is enhanced.

II. SWITCHED PARASITIC ARRAY ANTENNA CONCEPT

The conceptual overview of the switched parasitic array antenna is given in Fig. 1. We have first simulated the antenna with seven monopole elements. There is one element at center which is surrounded by a circular array. Feed is applied to the center element and it is termed as active element, whereas elements in the circular array are parasitic elements. Parasitic array antennas do not involve phase shifters, so that the manufacturing cost is less as compared to the phase array antennas. Each element is separated from other element by a quarter of wavelength, thus mutual coupling phenomena comes into account and there is a shift in phase of current in each element because of this mutual coupling phenomenon. Since switched parasitic array antennas offer adaptive beamforming solution with compact size and low cost, they are likely to be used in future communication devices to increase data-rates [6], [7]. Parasitic array antennas have vast potential for usage in laptops, cell phones, WLANs etc.

A. *Mutual Coupling*

Mutual Coupling is basically the interaction between antenna elements in an array. Since parasitic elements are placed closer to each other (at a distance of $\lambda/4$) there will be mutual coupling.This mutual coupling phenomenon is responsible for a shift in phase of current induced in each element [8].The current induced by one antenna element in another antenna element [9] is given by (3)

$$I_2 = \frac{Z_{21}}{Z_{22}} \times I_0 \tag{2}$$

$$Z_{21} = \left|\frac{V_2}{I_1}\right|_{I_2=0} \tag{3}$$

$$Z_{21} = \left|\frac{V_2}{I_2}\right|_{I_1=0} \tag{4}$$

Here Z_{22} is constant as it is open-circuit output impedance. Z_{21} is open-circuit transfer impedenace from port 2 to port 1 and it can be approximated by graph shown in Fig. 2. From equation (3), beam can be steered in the desired direction by varying the phase of current I_2 [10].

978-1-5090-0783-7/16/$31.00 ©2016 IEEE

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at *2016 Loughborough Antennas & Propagation Conference (LAPC)*, published by IEEE. Copyright restrictions may apply. doi: 10.1109/LAPC.2016.7807569

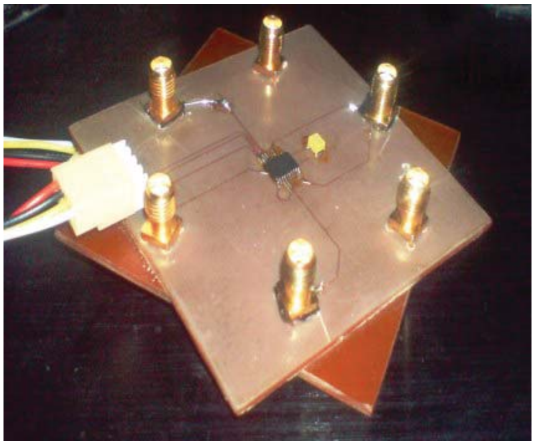

Fig. 3. RF PCB with Analog devices ADG 904 RF switches.

TABLE I
DESIGN PARAMETERS

Parameters	Values
Ground Skirt length	lambda/4
Monople Length	lambda/4
Frequency	2.45 GHz
Ground Radius	lambda/2
Monopole Radius	lambda/200
Ground Thickness	3 mm

B. *Beam Steering Mechanism*

In this work the designed antenna is capable of 360^0 beam steering in steps of 60^0. The beam steering is achieved by using two sets of 4:1 Analog devices multiplier switch ADG 904. Fig.3 shows RF PCB designed for beam steering. Beam is steered towards an element which is open. The open element behaves as a director and shorted elements act as reflectors. In order to achieve a maxima at 0^0, we make the element 1 open by using RF switch IC while other elements (i.e., elements 2,3,4,5 and 6) are shorted. This mechanism is similar to Yagi-Uda antenna where elements with shorter lengths in front of feed element act as directors and elements at rear of feed element acts as reflectors [11]. In Yagi-Uda antenna feed element is often a dipole or folded dipole.

III. PARASITIC ARRAY ANTENNA DESIGN

Design parameters and mechanical design of antenna is included in this section.

A. *Design Parameters*

Antenna design parameters are tabulated and the results are optimized using antenna simulation tools. Ground is skirted, since skirted ground provides mechanically sound model for PCB assembly to mount at back of antenna [12]. Skirt is $\lambda/4$ in length similarly each monopole element is lambda/4 in length and $\lambda/200$ is its radius.Table I shows the design parameters for the seven-element switched parasitic array antenna.

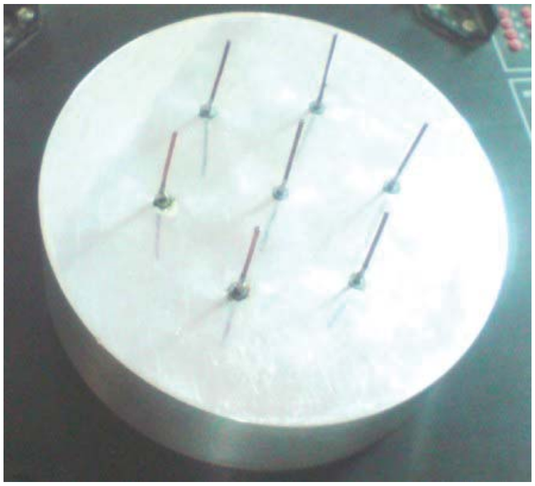

Fig. 4. Designed seven element parasitic switched array antenna.

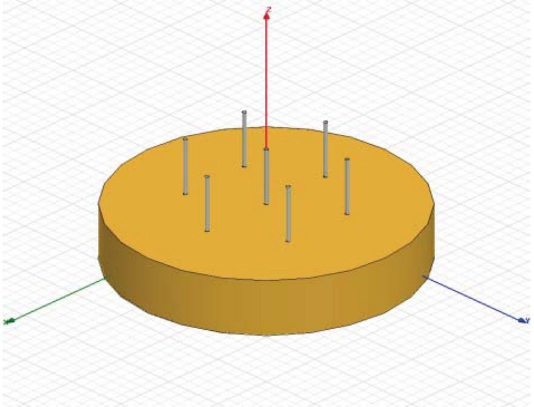

Fig. 5. Simulated antenna model.

B. *Mechanical Design*

Monopole elements are made by cutting copper wire of desired gauge i.e lambda/200 in radius [13] and these elements are soldered on pin of SMA connectors.Aluminum is used as ground plane. Electromagnetic waves are reflected by the ground. Instead of using seven dipole elements we have used seven monopole elements mounted on the ground plane [14]. Ground is made hollow in order to create space for mounting control circuitry [15]. Fig. 4 shows designed hardware.

IV. ANTENNA SIMULATION AND PLOTS

The antenna is first simulated and then simulated results are compared with the measured results of fabricated antenna. Simulated results were in line with the hardware results and 8 dB gain was observed in simulated results as well as in fabricated antenna. In simulation each of the parasitic element is loaded with lumped port excitation, simulating opening and shorting [16] of monopole elements. Fig. 5 shows simulated model.

978-1-5090-0783-7/16/$31.00 ©2016 IEEE

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at *2016 Loughborough Antennas & Propagation Conference (LAPC)*, published by IEEE. Copyright restrictions may apply. doi: 10.1109/LAPC.2016.7807569

TABLE II
MAXIMA PATTERN

Azimuth Angle	Element 1	Element 2	Element 3	Element 4	Element 5	Element 6
0 degree	Open	Short	Short	Short	Short	Short
30 degree	Open	Open	Short	Short	Short	Short

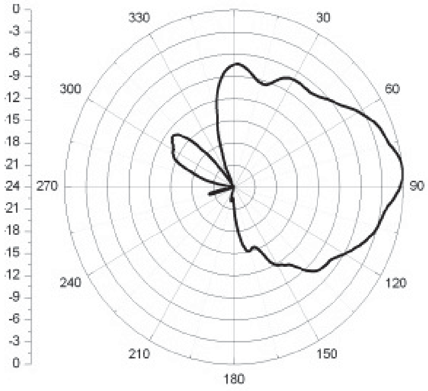

Fig. 6. Anechoic chamber results for $\theta=90^0$ (Beam in elevated because we are using monopole elements with finite ground).

Two dimension polar plots depicts the beam steering in the direction of given elevation angle, theta (θ) and azimuth angle, phi (ϕ). Anechoic chamber results for θ=90 and ϕ=0 and 30 degrees are shown in Fig. 7 and 8 respectively. Using large ground plane angle of elevation is reduced closer to 90^0. Fig. 6 shows anechoic chamber result for θ=90 one can observe small elevation form desired x-y axis because of ground skirt [17].In simulations elevation angle of 85^0 is observed. Fig. 6(a) shows antenna simulated plot for elevation angle θ and Fig. 6(b) shows anechoic chamber results.Anechoic chamber plot of Fig. 7 shows maxima at $\phi=0^0$, in order to get maxima at 0^0 elements from element number 2-6 are shorted and the element 1 is opened. Similarly Anechoic chamber plot of Fig. 8 shows maxima at $\phi=30^0$, in order to get maxima at 30^0 elements from element number 3-6 are shorted and element 1 and element 2 are open.

Traditional ESPAR antenna and switched parasiric array antenna have all cylindrical elements, including the active element [3], [4]. We have made center element conical instead of cylindrical. Due to conical nature of active element response our frequency range and improved. In order to validate the gain enhancement both models were simulated and results were compared. Fig 9. (a) shows conventional design and Fig. 9(b) shows simulated antenna with conical center element. Fig. 10 shows S_{11} plot for cylindrical active element. Considering -8_{dB} as cut off point bandwidth range from 2.358 GHz to 2.51 GHz (0.152 GHz). Fig. 11 shows $S-11$ plot for centre conical element, considering 8_{dB} as cut off point bandwidth ranges from 2.36 GHz to 2.555 GHz (0.195 GHz). Therefore my making center element conical we have improved bandwidth as compared to traditional switch array parasitic antennas. Low profile design for switched parasitic array antennas in also highlighted in [18].

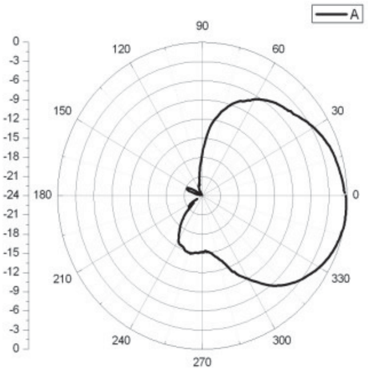

Fig. 7. Anechoic chamber results for $\phi=0^0$.

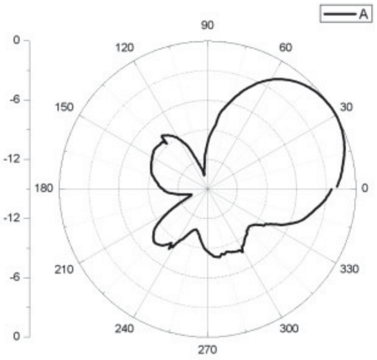

Fig. 8. Anechoic chamber results for $\phi=30^0$.

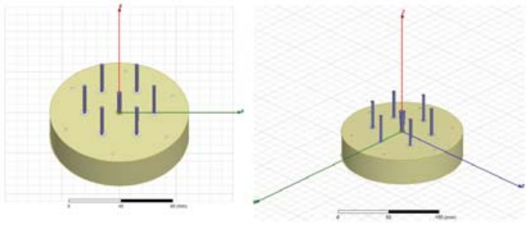

Fig. 9. (a) Conventional ESPAR antenna model with cylindrical Center Element, Fig. 9 (b) Designed ESPAR antenna with conical element for Bandwidth Enhancement.

V. FUTURE WORK AND CONCLUSION

5G networks will be designed for device to device, device to human and human to human interactions. High data rates are required to meet demands of 5G systems. One of proposed bands for 5G is in mm-Waves range i.e from 30GHz to 60GHz for such high frequencies path losses are considerable and beam steering antennas are way forward to overcome path

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at *2016 Loughborough Antennas & Propagation Conference (LAPC)*, published by IEEE. Copyright restrictions may apply. doi: 10.1109/LAPC.2016.7807569

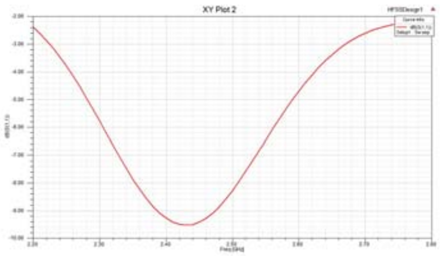

Fig. 10. S_{11} plot for conventional switched array antenna.

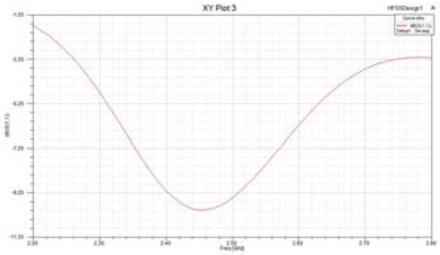

Fig. 11. S_{11} plot for center conical element switched array antenna.

losses [19]. ESPAR and switched parasitic array antennas provide a cost effective way of incorporating adaptive beamforming in future 5G and internet of Things (IoT) devices [20]. However for the efficient use of ESPAR antennas in mobile terminals the size of the antenna system must be reduced, for this further research on dielectric switched parasitic array antennas is underway. Seven element switched parasitic array antenna is designed that is capable of dynamic beam steering. Gain of 8dB is achieved practically. Simulated results are in line with the anechoic chamber results. Designed antenna increase the channel capacity by improving signal to noise ratio. Gain of 8dB was achieved in specified directions of elevation angle ,theta (θ) and azimuth angle, phi (ϕ). Antenna is optimized in terms of interference reduction for its use in wireless ad hoc networks, operating at frequency of 2.45 GHz. Bandwidth of designed antenna is 450 MHz and beam is steered in 360^0.

ACKNOWLEDGEMENT

Thanks to EPSRC Project EP/M014126/1, Large Scale Antenna Systems Made Practical: Advanced Signal Processing for Compact Deployments [LSAS-SP] and National Science Foundation (NSF), Enhancing Access to Radio Spectrum (EARS) program for funding.

REFERENCES

[1] M. Rzymowski, P. Woznica, and L. Kulas, "Single-anchor indoor localization using espar antenna," *IEEE Antennas and Wireless Propagation Letters*, vol. 15, pp. 1183–1186, 2016.

[2] L. Zhou, F. A. Khan, T. Ratnarajah, and C. B. Papadias, "Achieving arbitrary signals transmission using a single radio frequency chain," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 4865–4878, Dec 2015.

[3] V. I. Barousis and C. B. Papadias, "Arbitrary precoding with single-fed parasitic arrays: Closed-form expressions and design guidelines," *IEEE Wireless Communications Letters*, vol. 3, no. 2, pp. 229–232, April 2014.

[4] M. R. Islam and M. Ali, "Elevation plane beam scanning of a novel parasitic array radiator antenna for 1900 mhz mobile handheld terminals," *IEEE Transactions on Antennas and Propagation*, vol. 58, no. 10, pp. 3344–3352, Oct 2010.

[5] S. Kausar, H. U. Rahman, T. Hassan, and A. Kausar, "Miniaturization of espar antenna using folded monopoles and conical central element," in *Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET), 2015 International Conference on*. IEEE, 2015, pp. 87–91.

[6] H. T. Liu, S. Gao, and T. H. Loh, "Electrically small and low cost smart antenna for wireless communication," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 3, pp. 1540–1549, March 2012.

[7] T. H. Loh, H. Liu, and S. Gao, "A cost-effective direct magnitude measurement methodology for smart antennas," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 4, pp. 2043–2050, April 2013.

[8] F. Khan, K. Ali, A. Kausar, and S. Kausar, "Energy generation capacity analysis of a canal based hydro project," in *2013 Fifth International Conference on Computational Intelligence, Modelling and Simulation*, Sept 2013, pp. 277–281.

[9] A. Kausar, M. A. Cheema, S. Kausar, H. Rehman, and T. Hassan, "Espar antenna system designing & simulation," in *Proceedings of the 2014 First International Conference on Systems Informatics, Modelling and Simulation*, ser. SIMS '14. IEEE Computer Society, 2014, pp. 193–196. [Online]. Available: http://dx.doi.org/10.1109/SIMS.2014.37

[10] S. Akkar and A. Gharsallah, "Reactance domains unitary music algorithms based on real-valued orthogonal decomposition for electronically steerable parasitic array radiator antennas," *IET Microwaves, Antennas Propagation*, vol. 6, no. 2, pp. 223–230, January 2012.

[11] S. Kausar, H. u. Rahman, A. Kausar, and T. Hassan, "Espar antenna system for dynamic tracking of active targets," in *Modelling Symposium (EMS), 2013 European*, Nov 2013, pp. 533–535.

[12] A. Umar Kausar, F. Aslam Khan, T. Hassan, A. Nasser, and S. Kausar, "Adaptive espar antenna system modelling & design." *International Journal of Simulation–Systems, Science & Technology*, vol. 14, no. 4, 2013.

[13] D. v. B. M. Trindade, C. Mller, M. C. F. D. Castro, and F. C. C. D. Castro, "Metamaterials applied to espar antenna for mutual coupling reduction," *IEEE Antennas and Wireless Propagation Letters*, vol. 14, pp. 430–433, 2015.

[14] S. Kausar, Ahmed Kausar, "Smart adaptive beam forming antenna for interference cancellation," in *Wireless Days (WD), 2013 IFIP*. IEEE, 2013, pp. 1–3.

[15] A. Kausar, S. Kausar, and T. Hassan, "Smart adaptive beam forming antenna for interference minimization," in *Future Generation Communication Technology (FGCT), 2013 Second International Conference on*. IEEE, 2013, pp. 6–9.

[16] F. Aslam, E. Ulhaq, A. Nasser, and A. Umar, "Intelligent modeling scheme for detection of line losses in power distribution system," in *Computer Modelling and Simulation (UKSim), 2013 UKSim 15th International Conference on*. IEEE, 2013, pp. 218–223.

[17] R. Qian, M. Sellathurai, and D. Wilcox, "A study on mvdr beamforming applied to an espar antenna," *IEEE Signal Processing Letters*, vol. 22, no. 1, pp. 67–70, Jan 2015.

[18] A. Pal, A. Mehta, D. Mirshekar-Syahkal, and H. Nakano, "A low-profile switched-beam dual-band capacitively coupled square loop antenna," in *Antennas and Propagation Conference (LAPC), 2013 Loughborough*, Nov 2013, pp. 563–566.

[19] A. Kausar, S. Kausar, T. Hassan *et al.*, "Interference cancellation using adaptive beam forming for 3 d beam steering." *International Journal of Simulation–Systems, Science & Technology*, vol. 14, no. 1, 2013.

[20] M. R. Nikkhah, P. Loghmannia, J. Rashed-Mohassel, and A. A. Kishk, "Theory of espar design with their implementation in large arrays," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 6, pp. 3359–3364, June 2014.

[21] T. Hassan, A. Kausar, H. Umair, and M. A. Anis, "Gain optimization of a seven element espar antenna using quasi-newton method," in *Microwave technology & computational electromagnetics (ICMTCE), 2011 IEEE international conference on*. IEEE, 2011, pp. 293–296.

12/9/85 XR 4,628,299 30 - 29

United States Patent [19]

Tate et al.

[11] **Patent Number: 4,628,299**

[45] **Date of Patent: Dec. 9, 1986**

[54] **SEISMIC WARNING SYSTEM USING RF ENERGY MONITOR**

[75] Inventors: **Joseph B. Tate**, Sausalito; **David E. Brown**, Mill Valley, both of Calif.

[73] Assignee: **David Pressman**, San Francisco, Calif. ; a part interest

[21] Appl. No.: **695,632**

[22] Filed: **Jan. 28, 1985**

[51] **Int. Cl.^{4}** **G08B 21/00**

[52] **U.S. Cl.** **340/540**; 340/600; 340/690

[58] **Field of Search** 340/540, 600, 690

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,214,238	7/1980	Adams et al.	340/540
4,364,033	12/1982	Tsay	340/540

Primary Examiner—Glen R. Swann, III

Attorney, Agent, or Firm—David Pressman

[57] **ABSTRACT**

The ambient broadband radio frequency field strength from broadcast stations is monitored (FIG. **4**) by periodic sampling (**50, 52**). A warning indication is provided if the field strength drops significantly. Drops in such field strength have been correlated empirically with the occurrence of seismic activity, usually several days later. Thus the indication serves as an early warning of an impending earthquake. In one preferred embodiment, a broadband, horizontal, very long monopole antenna (**40**) was connected to a rectifying and smoothing circuit (FIG. **3**) to provide a dc output proportional to the ambient rf field. This voltage is digitized (**50**), and using a suitably-programmed computer (**52**), the digital version of the field strength signal is sampled once per minute (**78**). A cumulative or running average of the minute samples is calculated (**80**) and held. Once per hour the latest running average is stored (**84**) and a standard deviation (SD) of the last 24 hourly stored running averages is calculated (**88**). If the SD exceeds a predetermined value, 0.3 in one embodiment, an alarm is triggered (**92**). The use of the SD eliminates the effect of day-to-day changes in the amounts of the variations of the ambient field strength, due to changes in tides and other factors. Once per day the samples are written (**96**) to a permanent storage file and a continuous plot of the field strength is also made (**14**). Preferably the alarm is triggered only if another detector also provides an indication (FIG. **6**), thereby to eliminate the effect of machine error.

20 Claims, 6 Drawing Figures

Ambient RF Level vs Time from Before to Occurrence of Quake

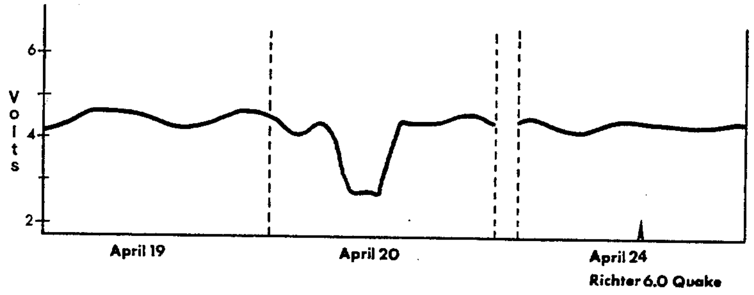

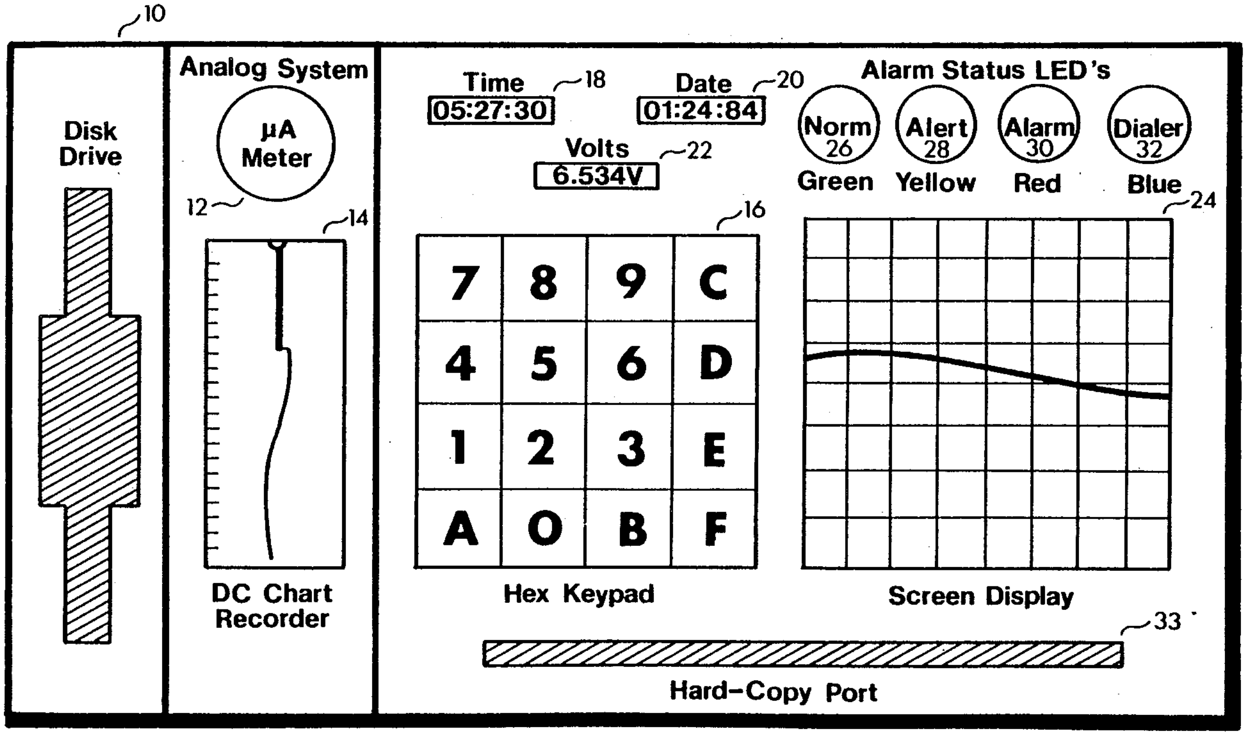
10
Disk Drive
Analog System
µA Meter
12
14
DC Chart Recorder
Time
05:27:30
18
Date
01:24:84
20
Volts
6.534V
22
16
7 8 9 C
4 5 6 D
1 2 3 E
A O B F
Hex Keypad
Alarm Status LED's
Norm 26
Green
Alert 28
Yellow
Alarm 30
Red
Dialer 32
Blue
24
Screen Display
33
Hard-Copy Port

FIG 1

Ambient RF Level vs Time from Before to Occurrence of Quake

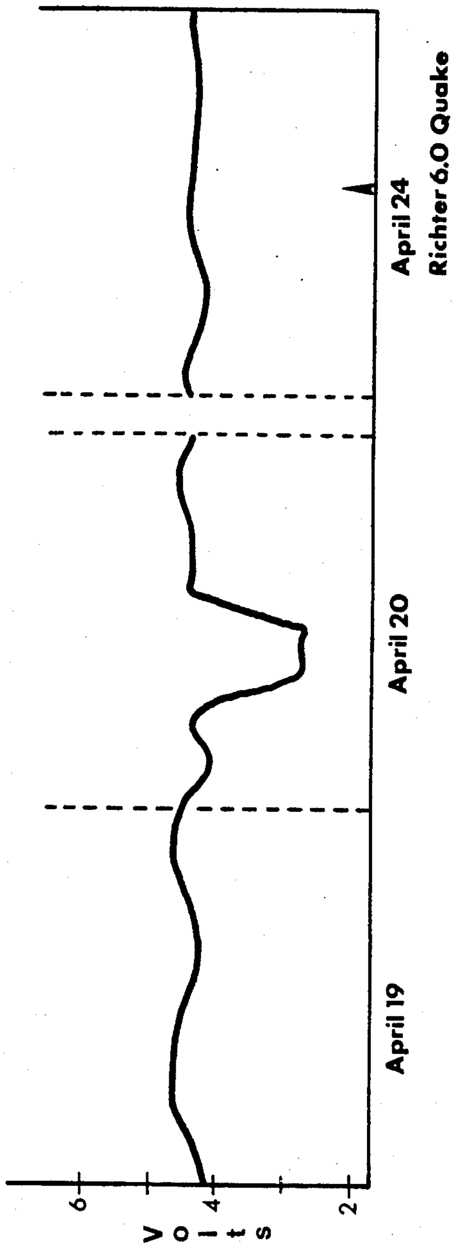

FIG 2

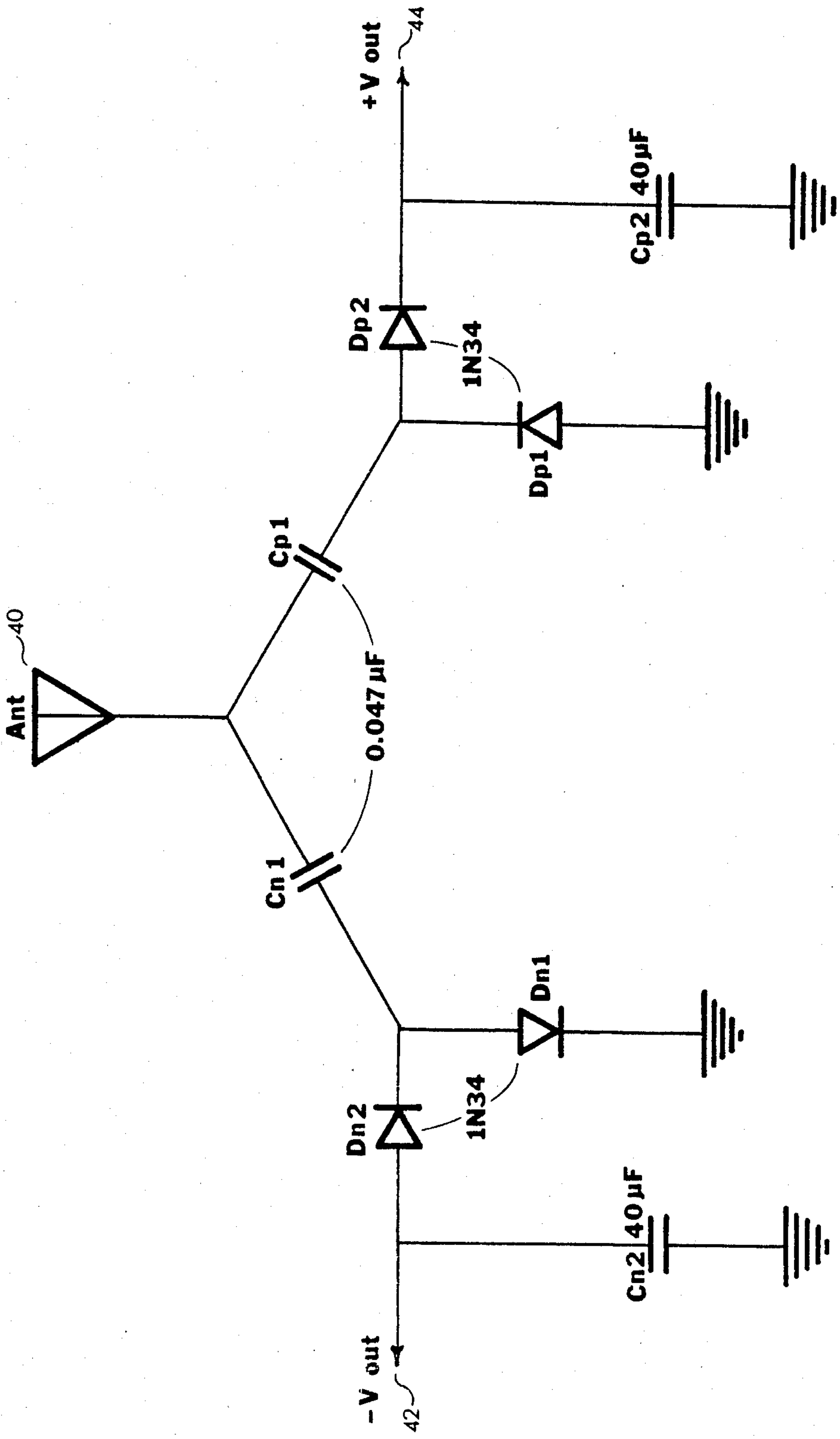
+V out
44
Cp2
40µF
Dp2
1N34
Dp1
Cp1
0.047µF
Ant
40
Cn1
Dn1
1N34
Dn2
Cn2
40µF
–V out
42

FIG 3

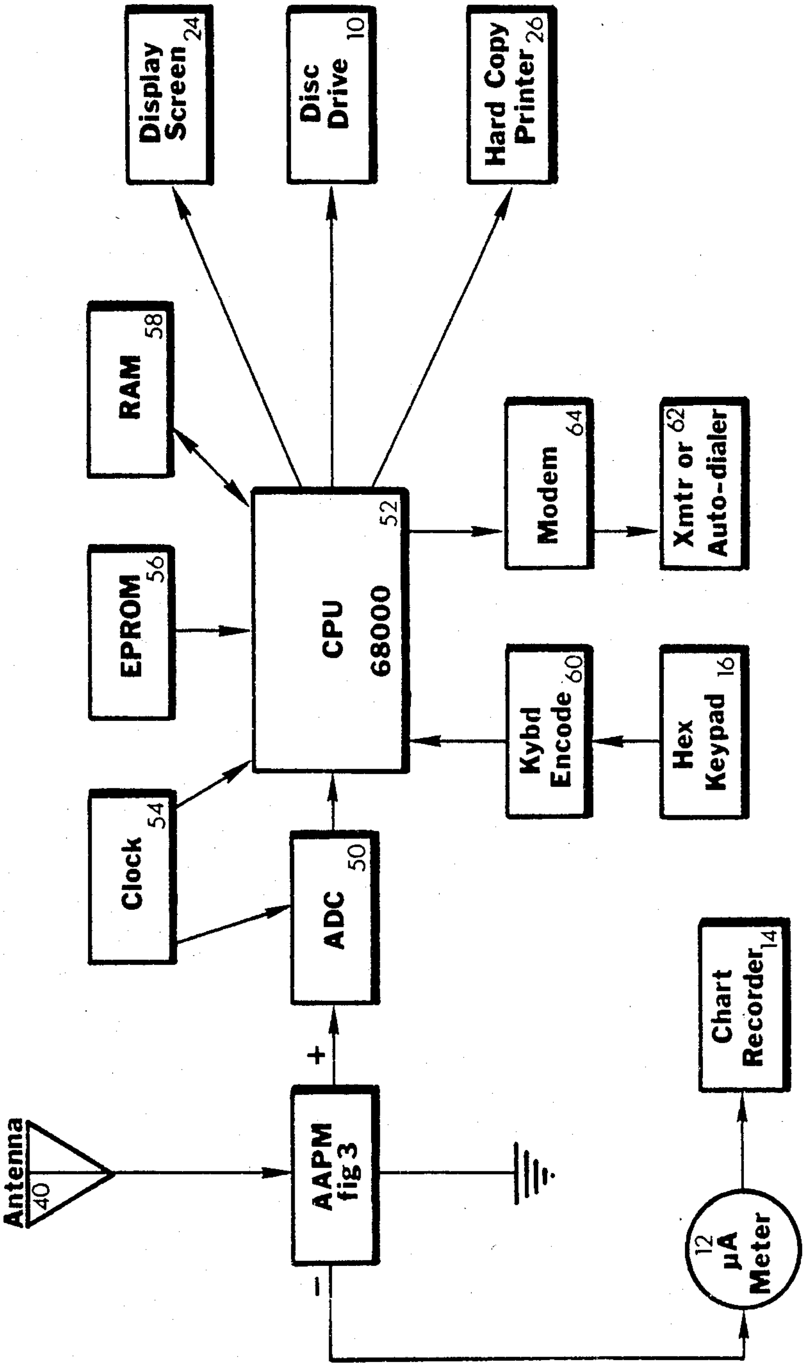
Antenna
40
AAPM
fig 3
+
-
ADC
50
Clock
54
EPROM
56
RAM
58
CPU
68000
52
Display Screen
24
Disc Drive
10
Hard Copy Printer
26
Modem
64
Xmtr or Auto-dialer
62
Kybd Encode
60
Hex Keypad
16
12
μA Meter
Chart Recorder
14

FIG 4

FIG 5

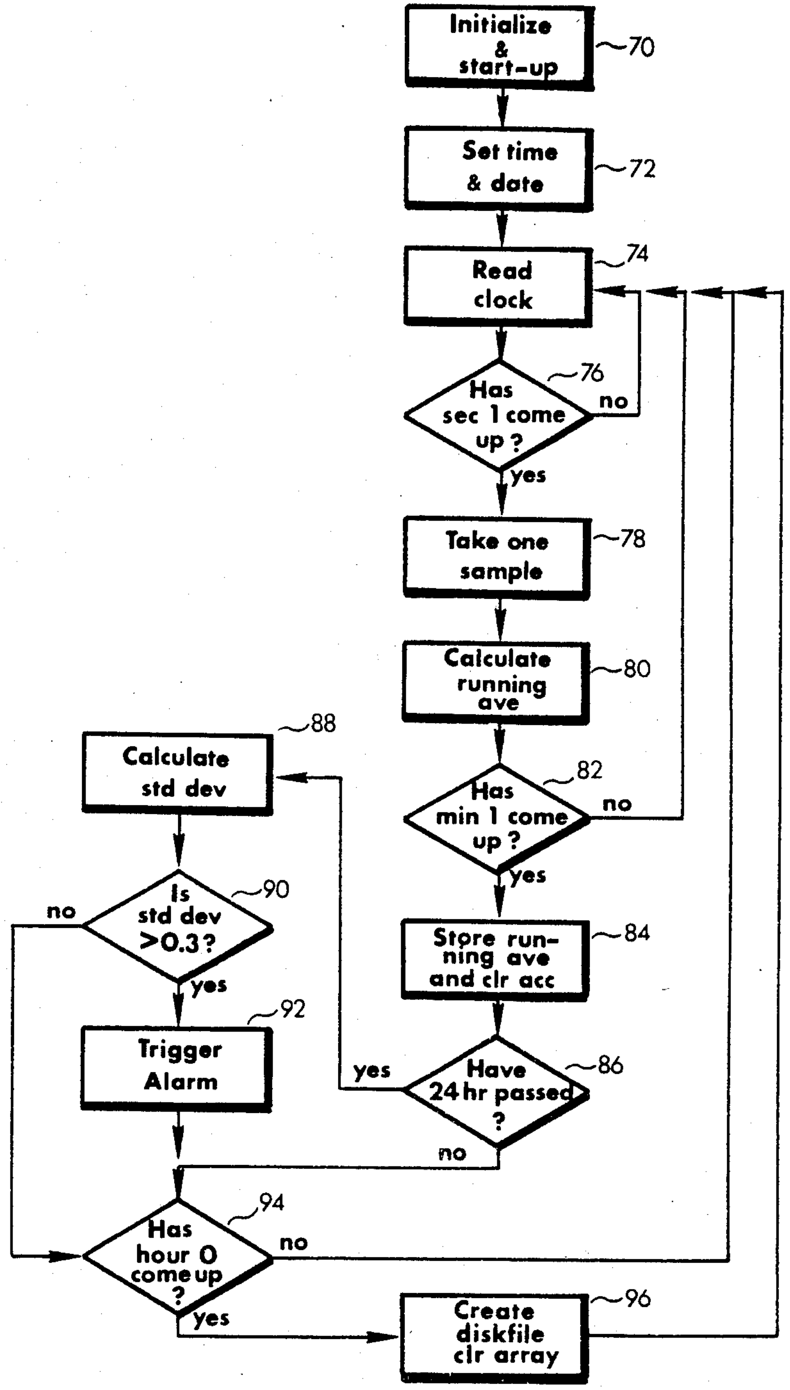
Initialize & start-up
70
Set time & date
72
Read clock
74
Has sec 1 come up ?
76
no
yes
Take one sample
78
Calculate running ave
80
Has min 1 come up ?
82
no
yes
Store running ave and clr acc
84
Have 24hr passed ?
86
yes
no
Calculate std dev
88
Is std dev >0.3?
90
no
yes
Trigger Alarm
92
Has hour 0 come up ?
94
no
yes
Create diskfile clr array
96

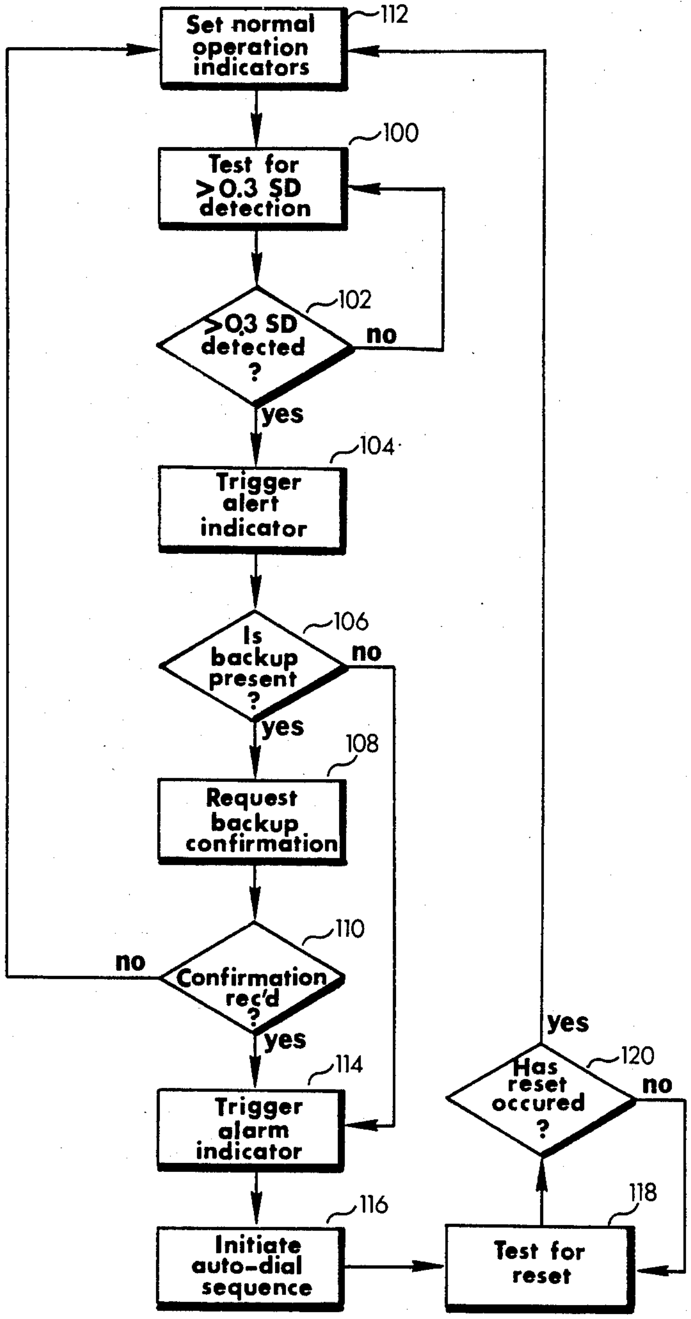
Set normal operation indicators
112
Test for >0.3 SD detection
100
>0.3 SD detected ?
102
no
yes
Trigger alert indicator
104
Is backup present ?
106
no
yes
Request backup confirmation
108
Confirmation rec'd ?
110
no
yes
Trigger alarm indicator
114
Initiate auto-dial sequence
116
Test for reset
118
Has reset occured ?
120
yes
no

FIG 6

SEISMIC WARNING SYSTEM USING RF ENERGY MONITOR

BACKGROUND—FIELD OF INVENTION

This invention relates to the prediction of the fugure occurrence of seismic activity, particularly to the advance notification of earthquakes through the monitoring of ambient radio frequency (rf) energy.

BACKGROUND—DESCRIPTION OF PRIOR ART

Heretofore, insofar as we are aware, seismology, the science of earthquakes, has not been able to make any near-term predictions of earthquakes.

While scientists have known that certain animals may have had some sort of advance knowledge of quakes, due to the fact that they exhibited peculiar behavior before quakes, and not at other times, this behavior has not been consistent and reliable enough to be of practical use.

Also, while scientists have also been able to predict thunderstorms in advance by monitoring the ambient electrostatic field (see, e.g., U.S. Pat. No. 3,611,365 to Husbyorg and Scuka, 1968; 3,790,884 to Kohl, 1974; and 4,095,221 to Slocum, 1978), they have not been aware of any corresponding system for earthquake prediction.

Scientists have been able actually to detect earthquakes during their occurrence by monitoring air pressure variations (e.g., as described in U.S. Pat. No. 4,126,203 to Miller, 1978) and by monitoring the earth's physical movement by seismographs but, again, science has not been aware of any system for short-term advance detection or prediction of quakes.

Due to the devastating effects of quakes to property, life, and limb, public and governmental authorities would derive great benefit from any system which could provide short-time advance notification of great earthquakes. As it is now, except for aftershocks, which seismologists know will occur after any large quake, all great and small quakes occur without warning. Because people in the vicinity of such quakes are unprepared, they often are in places of great vulnerability, such as beside or inside collapsible buildings, so that severe and human injury usually occurs during a quake. Also, property itself is left vulnerable, e.g., by leaving automobiles in or near collapsible buildings, leaving gas and electricity connected such that disruption of these facilities causes fires, and leaving other valuable property in vulnerable areas. If advance notification of a large quake could be provided to the public and civil authorities, people and valuable property could be evacuated and protected, thereby preventing deaths, injuries, and greatly reducing property damage. Further, advance notification of quakes would eliminate the severe psychological trauma which often affects large segments of the populace due to the surprise occurrence of quakes.

OBJECTS AND ADVANTAGES

Accordingly several objects and advantages of the invention are to provide a reliable and effective method of earthquake prediction, to provide a method of preventing death, injuries, and reducing property damage in earthquakes, and to provide a method of reducing the psychological trauma which often accompanies quakes due to their surprise occurrence. Additional objects are to provide such a system which is easy to use, economical, reliable, and portable. Further objects will become apparent from a consideration of the ensuing description, taken in conjunction with the accompanying drawings.

BACKGROUND—THEORY OF INVENTION

The following is a discussion of the background theory of the invention. While we believe it to be technically accurate, we do not wish to be limited by this theory since the operability of the invention has been empirically verified, as will be apparent from the later discussion.

We have recently worked work with the reception and utilization of broadband radio-frequency reception, e.g., for low-power utilization applications, as discussed in the copending application Ser. No. 06/539,223 of Joseph B. Tate, filed Oct. 6, 1983. While doing this work, we have noted that the antenna's output voltage fluctuated with time due to certain, known causes.

First, we noted that the higher we placed an antenna above the ground, the the greater the output signal it provided. We have observed this by raising the physical height of an antenna and observing an increase in power output, and also by observing variations in the output of a fixed antenna near a body of ocean water as a function of the tides: the antenna's output was greatest at low tide and lowest at high tide. We believe that the the change in water level, which serves as a ground plane, effectively lowers or raises the height of the antenna above the ground.

We also noted that the antenna's output was affected by solar flares to a limited extent; these caused the antenna to produce a higher output voltage during their occurrence. We believe this phenomena is caused by an increase in the level of ambient ionization due to the flares.

Further, we noted that the antenna's output dropped at certain irregular times; at first we would not attribute any cause to these drops. However investigation enabled us to correlate these drops with the subsequent occurrence of seismic activity. We found that the magnitude of the drop was proportional to the size of the subsequent earthquake.

Certain phenomena have been discovered to precede earthquakes. These include an anomalous uplift of the ground, changes in the electrical conductivity of rock, changes in the isotopic composition of deep well water, changes in the nature of small earthquake activity (e.g., bunching of small foreshocks), anomalous ground tilt or strain changes, changes in physical properties, such as porosity, electrical conductivity, and elastic velocity in the hypocentral region. Earthquake, McGraw-Hill Encyclopedia of Science And Technology, 1960; Earth by F. Press, W. H. Freeman & Co., 1974.

Phenomena associated with rocks have attracted much recent attention. Wm. Brace of the Mass. Inst. of Technology has found that when rocks were squeezed or compressed, just before they fractured, they tended to develop hairline cracks, swell or dilate (dilatancy), become more porous and electrically conductive, and transmitted high frequency seismic-like waves more slowly. Two of Brace's former students, Amos Nur of Stanford University and Christopher Scholz of Lamont-Doherty furthered Brace's work, connecting the dilatancy theory with seismic P-wave velocity shifts and rock resistivity changes as a precursor for earthquakes. See. e.g., Brace, Orange, and Madden, J. Geophys Res., 70(22), 5669, 1965; A. Nur, Bull. Seis. Soc. of Amer., V 62, Nr. 5, pp. 1217–1222, 1972 Oct.; Earthquake by B. Walker, Time-Life Books, 1982.

Based upon the above background, we have developed a theory as to the cause of this drop in antenna output as a precursor or predictor of earthquakes. We believe that before a quake occurrs, the pressure within underground rock bodies temporarily increases greatly, causing the rocks to dilate and become conductive, in accordance with the works of Brace, Nur, and Scholz. This increase in conductivity effectively raises the ground plane, thereby causing the antenna's output to decrease temporarily.

Thus before the occurrence of a quake, the underground pressure increases greatly temporarily, causing underground rock bodies to swell and become more conductive, thereby raising the ground plane, which in turn causes the voltaic output of nearby antennas to drop.

We accordingly constructed an apparatus to automatically monitor antenna output and provide a suitable indication if the output level dropped significantly. The indication was calibrated empirically after much experimentation so as to filter ouo the effects of solar- and tide-caused variations. We did this by arranging the apparatus so that an output indication was provided only if the antenna output dropped a predetermined degree beyond its average level; we utilized statistical filtering techniques to accomplish this.

DRAWINGS

FIG. 1 shows the front panel of a Seismic Early Warning (SEW) apparatus according to the invention.

FIG. 2 is a plot of voltage (representing ambient rf level) v. time as measured by the apparatus of FIG. 1.

FIG. 3 is a schematic diagram of an ambient power module circuit (used in the SEW apparatus) for producing a DC output voltage proportional to the ambient rf energy

FIG. 4 is a block diagram of a computer in the apparatus of FIG. 1.

FIG. 5 is a flowchart which depicts the operation of the the SEW system.

FIG. 6 is a flowchart which depicts the operation of an optional alarm trigger system useable with the SEW apparatus.

FIG. 1—SEISMIC EARLY WARNING APPARATUS

In accordance with the invention, a seismic early warning apparatus is provided as shown in FIG. 1. The apparatus consists of a housing containing a general purpose computer (not shown), a disc drive **10**, an analog system comprising a microampere meter **12** arranged to monitor direct current (which is proportional to the ambient rf energy), and a direct current strip chart recorder **14** arranged to provide a continuous indication of the current antenna output, which will be called the ambient power level. A hexidecimal keypad **16** is provided to enter data, such as time, for entering programs and changes and for operating the system according to preset codes. The time, date, and voltaic level of the antenna's output are continuously indicated by digital readouts **18**, **20**, and **22**, respectively. A screen display **24** is provided to display graphic and alphanumeric information of the current status of the apparatus and previous data records.

Lastly the apparatus includes four status indicating lamps, which preferably are LEDs (light-emitting diodes) as follows: A green LED **26** indicates that the system is on and functioning normally. A yellow LED **28** indicates that the system has detected an event, namely the occurrence of a drop in ambient power below the preset level, which would be the prediction of an impending earthquake. A red LED **30** is provided as backup confirmation of the occurrence of the event; LED **30** is illuminated when a duplicate receiving system also detects an event. A blue LED **32** indicates initiation of operation of an automatic telephone dialer within the system, which has been preprogrammed to dial a predetermined number and provide a warning in the event of an occurrence of an alarm condition. Lastly the apparatus includes a hard copy output port **33** for providing printed graphic and numeric outputs of all system data.

FIG. 2—AMBIENT RF LEVEL V. TIME BEFORE QUAKE

FIG. **2** illustrates a reproduction of an actual plot of a voltage as a function of time, which voltage was proportional to the ambient RF (radio frequency) level, from the period from before to after a relatively large earthquake. This plot, which is typical of many we have observed before a quake, was made by deriving the voltage with a 30-meter, long-wire monopole antenna (not shown) which was mounted horizontally and which extended over San Francisco (Richardson) Bay easterly from Sausalito, California, 9 meters above sea level. The antenna thus intercepted and converted to an RF voltage the ambient RF energy, mainly from local (San Francisco area) AM radio stations. We rectified and filtered the output of the antenna using one-half of the circuit of FIG. **3** (described below) to provide a DC voltage which was plotted on a conventional ink-on-paper plotter. Note that on the section of the chart for Apr. 19 (1984), which begins at time 0:00 (midnight) and ends at 24:00, the voltage or ambient RF power level at the antenna increased and fell and then increased slightly in the 24-hour period. This wavelike variation typically occurs on a daily basis and is caused by tides: the peaks occurring at low tide when the effective ground plane provided by the water drops and the troughs occurring at high tide when the ground plane rises.

On Apr. 20, from about 8:00 to about 12:00, a sharp and constant-level dip in the ambient rf power occurred, as indicated. The magnitude of this pronounced dip is far greater than the normal tide-caused variations, as is its beginning and ending slope.

Thereafter, from Apr. 20 to Apr. 23, the plot (not shown) continued unremarkably, albeit with a slight variation from normal.

The same occurred on Apr. 24, with the plot actually being generally similar to a normal day. However at 13:15 on Apr. 24, as indicated, a large, Richter magnitude 6.0 quake occured near Hollister, Calif., about 340 km away from the antenna. No change in the plot occurred at this time.

Correlation of this quake with the plot's marked dip of Apr. 20 was made by the repeated observation of dozens of similar dips and subsequent quakes. Pronounced dips were always followed by a quake several days later. Thus we have empirically established causal and theoretical connections between pronounced dips of the type shown and the occurrence of subsequent seismic activity.

FIG. 3—AMBIENT POWER MODULE

The circuit of FIG. **3** is used to convert the ambient RF energy to a direct voltage which can be used and handled by data processing equipment. Designated an ambient power module (APM), it is connected to an antenna **40**, preferably a broadband monopole antenna of the type described in the preceeding section. The distal end of the antenna is free and its proximal end is connected to the circuit via two capacitors **Cp1** and **Cn1**, each being in series with the signal line for coupling and each having a value of 0.047 microfarad. Taking the left or negative side of the circuit first, it comprises two rectifiers (diodes) Dn1 and Dn2 (1N34 type) and a filter capacitor Cn2 (40 microfarads). Rectifier Dn1 is connected in parallel to the signal path and rectifier Dn2 is connected in series, in the wellknown voltage multiplier arrangement. Capacitor Cn2 is connected in parallel across the output of the APM to smooth the rectified output. The right or positive side of the circuit is similar, except for the polarity of the diodes.

In operation, an RF voltage is developed across antenna **40**; this voltage is voltage multiplied by the two rectifiers on each side of the circuit. The resultant direct voltages are smoothed or filtered by capacitors Cn1 and Cp2 and are supplied to output terminals **42** and **44**. A positive version of this direct voltage is plotted in FIG. **2**, as described above.

FIG. 4—BLOCK DIAGRAM OF COMPUTER

A computer for performing the monitoring and alarm functions of the invention and which is provided within the apparatus of FIG. **1** is shown in FIG. **4**. The computer receives the positive voltage from the APM (FIG. **3**) and processes this, providing an alarm if the voltage dips a predetermined amount from its recent average value.

The computer comprises an analog to digital converter (ADC) **50** which is arranged to convert the positive DC voltage from the AAPM to digital form, preferably in the form of a parallel signal at the output of ADC **50**. The digitized voltage from ADC **50** is supplied to a central processing unit **52**, which is a type 68000 microprocessor or computer on a chip. CPU **52** and ADC **50** are clocked by a clock **54** in conventional fashion.

CPU **52** operates on instructions from a program contained in an electricallyprogrammed read only memory (EPROM), the program being listed later. CPU **52** temporarily stores data in a read and write memory (RAM) **58**. CPU **52** also supplies output data to display screen **24**, disc drive **10**, and hard-copy printer **26'**, each of which was already described in conjunction with FIG. **1**.

CPU **52** can receive input data manually from hexidecimal keypad **16** (see FIG. **1**) via a keyboard encoder **60**.

CPU **52** can supply an alarm output to a radio transmitter or automatic telephone dialer **62** via a modem (modulator-demodulator) **64** for connecting the CPU to a phone (not shown).

As also indicated in FIG. **4**, the negative output of the AAPM of FIG. **3** is connected to ammeter **12** and chart recorder **14**.

FIG. 5—FLOWCHART OF SEISMIC EARLY WARNING SYSTEM

In operation, the system of FIG. **4** operates under control of the program in EPROM **56** in accordance with the flowchart of FIG. **5** as follows:

Startup: Blocks **70** and **72**: An initialization and startup sequence is first initiated when the machine is turned on, as indicated by block **70**; this sets all registers and counters to zero. The time and data are then set manually (using EPROM **56**), as indicated by block **72**.

Clock Reading: Blocks **74** and **76**: Next, under automatic program control, the machine reads the elapsed time on its clock display register, as indicated by block **74**. If the "seconds" register does not indicate the number one (#1), the machine continues to read the clock, as indicated by the "no" output of decision block **76**.

Minute Sample: Block **78**: When second #1 appears, as it will once per minute, the decision in block **76** will be "yes", so that the machine will take one sample of the rectified, smoothed, and digitized version of the antenna's output, i.e., the output of ADC **50** of FIG. **4**, as indicated in block **78**. This sample will be taken once per minute, i.e., whenever second #1 is displayed.

Running Average: Block **80**: Next, as indicated by block **80**, a running average of the samples taken in block **78** is calculated. This is done by accumulating the samples to keep a running total of their values, counting the number of samples accumulated, and dividing the running total by the latest number of samples each time a new sample is taken.

Store Hourly Average: Blocks **82** and **84**: Next, as indicated in block **82**, a test is made to see if the time display register indicates that minute number one (#1) has come up. If not, the decision is "no" and the clock is read again (block **74**). If the decision is "yes", as it will be once per hour, the running average in the accumulator will be stored (block **84**) and the accumulator will be cleared or reset to zero.

One Day Test: Block **86** ("No" decision) and Block **94**: Next the machine makes a test to see if 24 hours have passed. If not, the machine will not be able to make any valid statistical determinations. Thus it must run at least 24 hours before being operative. Assuming the decision in block **86** is negative (24 hours have not yet elapsed) another test is made (block **94**) to see if hour zero is indicated, which will occur once per day. If hour zero is not indicated, (decision in block **94** is negative), the clock will be read again (block **74**) in the usual loop.

Calculate SD: Block **86** ("Yes") and Block **88**: If a full day has elapsed, so that valid statistics can be calculated ("yes" from block **86**), the standard deviation (SD) of the last **24** hourly averages is calculated, as indicated in block **88**. This is done once per hour. The calculation is made using the usual SD formula

$$SDDEV = SQR([sum(x-X)^2]/n)$$

where SDDEV=SD; SQR=the square root; sum=the sum of; x=the individual hourly averages; X=the mean of the hourly averages; and n=the number of individual hourly averages. Essentially the SD is calculated by taking the mean of all of the hourly averages, taking the difference or deviation of each hourly average from the mean, squaring each deviation, taking the mean of the squared deviations, and then taking the square root of the mean of the squared deviations.

Evaluate SD: Block **90**: The SD is then evaluated to see if it is greater than 0.3. This value has been empirically determined to be the level at which a the present apparatus will provide a reasonably positive indication that an earthquake will occur, while neglecting the effects of non-seismic-caused variations. If the SD is less than 0.3, (a "no" output from block **90**), this indicates that the last hourly average was not greatly different from the average of the last 24 hourly samples, so that no alarm need be indicated. I.e., the antenna's output did not drop significantly to indicate an impending earthquake. Thereupon the program moves to block **94**, where a test is made for the existence of hour zero, as described. If, however the SD exceds 0.3 ("yes" output of block **90**), this indicates that the antenna's output has dropped significantly so as to affect the last hourly average, thereby to indicate an impending earthquake.

Alarm: Block **92**: In response to the Yes output of block **92**, an alarm is triggered (block **94**). The alarm may be a bell, the dialing of a telephone to a location where personnel are present if the apparatus is placed at a remote or non-manned location, or the initiation of the further program of the Flowchart of FIG. **6**, the alarm trigger sequence. To eliminate the possibility of equipment failure and to provide confirmation from another apparatus at another location, we prefer to provide an alarm only upon confirmation from another apparatus, as discussed in the description of FIG. **6** below.

Make Record: Block **94** ("Yes") and Block **96**: If hour zero is being displayed when the operation of block **94** is performed, which occurs once per day at midnight, the operation of block **96** will be performed, i.e., the data in the registers will be stored to disc to create a permanent record and the registers will be cleared to create new data for the next day. However the previous 24 hourly averages are still stored at all times so that a valid SD can be calculated and tested every hour. After the operation of block **96**, the clock is read again in accordance with the regular program (block **74**).

FIG. 6—ALARM TRIGGER FLOWCHART

The sequence of FIG. **6** is performed when the alarm is triggered in block **92** of FIG. **5** as an optional, but preferred backup confirmation of an impending earthquake. The operations in the backup confirmation system will be described briefly.

Beginning with blocks **100** and **102**, the system is continually tested (hourly) for the occurrence of a SD of the hourly averages of greater than 0.3. If the SD is greater than 0.3, the alert indicator (**28** of FIG. **1**) is triggered (block **104**) and the program initiates a test (block **106**) to see if a backup apparatus (not shown) is present. If so (yes output of block **106**) the backup apparatus is also checked (blocks **108** and **110**). If the backup does not indicate an excess SD, the indicators are reset to normal (block **112**), but if backup confirmation is received, the alarm indicator (**30** of FIG. **1**) is triggered per block **114** and a preprogrammed telephone number is dialed and indicator **32** is lit (block **116**).

After the alarm condition is manually checked and the system is reset, the output of block **120** will be a "yes" and the system will be reset to normal (block **112**). If a valid alarm condition is indicated and confirmed, civil authorities will have time (usually several days) to notify the populace, evacuate the area, or take any other needed precautions, depending on the size of the impending quake as indicated by the size of the standard deviation.

PROGRAMS:

The attached computer programs will perform the calculations and operations above described. These programs are written in the BASIC programming language. Program "RECVOLT.AL" runs continuously and writes the information to disc every 24 hours. Program "GRASTAT.*" is manually run; it reads data from the disc and plots it on the screen or printer, as desired.

While the above description contains many specifications, these should not be construed as limitations on the scope of the invention, but merely as an exemplification of one preferred embodiment thereof. Many other variations are possible. For example, the programming language can be changed, or the calculations and operations can be performed with hard-wired conventional circuitry in lieu of a programmed computer. More than two corroboration receivers can be used, and these can be placed at various locations. In lieu of testing the antenna's output reception of the area's AM stations, a special, decicated transmitter with a special, dedicated frequency and a specially-tuned matching receiver can be used to avoid depedence on stations which are not under the control of the earthquake prediction system and its personnel. The transmitter and the receiver should be spaced apart geographically, preferably by at least several km, so that the ground plane conduction phenemon can operate. Also the transmitted signal can be a specially-coded or modulated signal, or it can be an auxiliary signal of a regular transmitter, e.g., a SSB or SCA signal, together with a matching receiver. In lieu of a test for an excess SD, the apparatus can be arranged to test for a predetermined drop in the value of the antenna output from its immediately previous value or its average value over a predetermined period, such as an hour or day, or for a drop having greater than a predetermined slope. Accordingly the full scope of the invention should be determined by the appended claims and their legal equivalents, and not by the examples given.

"RECVOLT.AL"

8299
Issue of 12·9

```
5 ? "}"
6 HO=0:U=1:G=1
7 DIM P$(100),A$(1):A$=" "
100 DIM B$(3)
110 DIM Z$(15)
115 DIM HAR(24):DIM HA(24):DIM HAL(25)
117 FOR I=1 TO 25:HAL(I)=0:NEXT I
120 POSITION 5.5:? "RECORDING VOLTMETER":POSITION 5,7:? "COPYRIGHT 1984":POSITIO
N 5,9:? "JOE TATE"
125 FOR D=1 TO 1000:NEXT D:? "}"
150 ? "DO YOU WANT TO SEE A FILE?"
```

```
160 INPUT B$
170 IF B$="YES" THEN GOTO 12040
180 IF B$="NO" THEN GOTO 200
200 N=1:J=0:AVG=0:A=0:K=0
205 DIM D$(3),M$(8),DATE$(15),X$(3)
220 FOR R=1 TO 24:HA(R)=0:NEXT R
280 ? "}":POKE 752,1
420 ? "IF DATE IS CORRECT PRESS C"
425 ? "PRESS N FOR NEXT MONTH"
430 READ M$:K=K+1:IF K>13 THEN K=1
432 IF M$<>"RES" THEN 435
433 RESTORE :GOTO 430
435 POSITION 6,12:? M$
440 INPUT X$
445 IF X$="C" THEN 450
448 IF X$="N" THEN 430
449 GOTO 435
450 ? "}"
455 ? "ENTER TODAY'S DATE"
458 INPUT D
459 D$=STR$(D)
460 ? M$;" ";D$
505 ? "DO YOU WANT TO SET THE TIME?";:INPUT X$
510 IF X$="YES" THEN 520
515 GOTO 1050
520 ? "}"
525 ? "HOURS";:INPUT H
530 ? "MINUTES";:INPUT M
535 ? "SECONDS";:INPUT S
540 ? "HIT START TO BEGIN THE TIME"
545 IF PEEK(53279)<>6 THEN 545
550 ? "}"
1010 REM *****PUT CURRENTTIME IN HARWARE REGISTER*****
1011 GRAPHICS 2
1015 POKE 18,0:POKE 19,0:POKE 20,0
1020 T=H*60^3+M*60^2+S*60
1025 POKE 18,INT(T/(256*256))
1030 T=T-(256*256)*(INT(T/(256*256)))
1035 POKE 19,INT(T/256)
1040 T=T-256*(INT(T/256))
1045 POKE 20,INT(T)
1050 ? "}"
1055 TIME=PEEK(20)+PEEK(19)*256+PEEK(18)*256*256
1065 IF TIME>5184000 THEN 1265
1070 TIME=INT(TIME/60+0.5)
1075 SEC=TIME-60*(INT(TIME/60))
1076 R=PEEK(53279):IF R=5 THEN SOUND 0,0,0,0:SOUND 1,0,0,0:U=1
1077 POKE 752,1:IF R=3 THEN G=2:GOTO 12040
1079 IF U=1 THEN POSITION 3,3:? #6;"E.T. ";HO;" HRS.  "
1080 POSITION 14,8:? #6;SEC;":S. "
1082 IF SEC=7 AND G=1 THEN GOSUB 15200:GOTO 14020
1083 IF SEC=8 AND G=0 AND HOURS<>0 THEN GOSUB 13000:GOSUB 13100:FOR D=1 TO 7500:
NEXT D:GOSUB 15200
1085 IF SEC>1 THEN N=1
1086 IF SEC=1 AND N=1 THEN 2020
1095 TIME=INT((TIME-SEC)/60)
1200 MIN=TIME-60*(INT(TIME/60))
1205 HOURS=INT((TIME-MIN)/60)
1210 IF SEC>=60 THEN 1220
1213 IF MIN=1 AND SEC=1 AND HO>24 THEN GOTO 15000
1215 GOTO 1230
1220 MIN=INT(SEC/60)+MIN
1225 SEC=SEC-60*(INT(SEC/60))
1230 IF MIN>=60 THEN 1240
1240 HOURS=INT(MIN/60)+HOURS
1245 MIN=MIN-60*(INT(MIN/60))
1252 IF HOURS>23 THEN HOURS=0
1255 GOTO 1055
1260 ? INT(HOURS+0.5);":";INT(MIN+0.5);":";INT(SEC+0.5):GOTO 1055
1262 ? "LINE1262"
1265 TIME=PEEK(18)*256*256+PEEK(19)*256+PEEK(20)
1270 TIME=TIME-5184000*(INT(TIME/5184000))
1275 POKE 18,INT(TIME/(256*256))
```

```
1280 TIME=TIME-(256*256)*INT(TIME/(256*256))
1285 POKE 19,INT(TIME/256)
1290 TIME=TIME-256*(INT(TIME*256))
1295 IF HOURS>23 THEN H=0:M=0:S=0:GOTO 1015
1300 GOTO 1055
2020 REM *****SAMPLE ROUTINE*****
2022 N=2
2025 LET X=PEEK(625)
2030 IF X>40 THEN POSITION 1,14:? "VOLTAGE BELOW RANGE(LESS THAN 1.8V)"
2035 IF X<3 THEN POSITION 1,10:? "VOLTAGE ABOVE RANGE(ABOVE 7V)"
2040 IF X>26<40 THEN LET V=(40-X)*0.01+1.85
2045 IF X>22<26 THEN LET V=(26-X)*0.062+2
2050 IF X>20<22 THEN LET V=(22-X)*0.125+2.25
2055 IF X>16<20 THEN LET V=(20-X)*0.125+2.5
2060 IF X>13<16 THEN LET V=(16-X)*0.166+3
2065 IF X>11<13 THEN LET V=(13-X)*0.225+3.5
2070 IF X>10<11 THEN LET V=(11-X)*0.4+3.95
2075 IF X>5<10 THEN LET V=(10-X)*0.15+3.25
2080 IF X>8<10 THEN LET V=(10-X)*0.325+4.35
2085 IF X>6<8 THEN LET V=(8-X)*0.5+5.1
2090 IF X>3<6 THEN LET V=(6-X)*0.33+5.9
2095 POKE 77,0
3000 GOSUB 15200
3500 J=J+1
3510 AVG=(A+V)/J
3512 LET A=A+V
3520 ? "SAMPLE #";J;"  RUN AVG ";AVG;M$;D$
3600 HA(HOURS)=AVG
3610 IF MIN<>0 THEN 1055
3612 HAL(25)=AVG
3613 FOR I=1 TO 24:HAL(I)=HAL(I+1):NEXT I
3615 IF MIN=0 THEN A=0:HO=HO+1
3625 J=0
4020 IF HOURS<>0 THEN 1055
4048 REM ****
4050 REM ** CREATE FILENAME-DISKWRITE****************
4052 REM ****
4053 FOR Q=0 TO 23:HAR(Q)=HA(Q):NEXT Q
4055 DATE$(LEN(DATE$)+1)=M$
4056 DATE$(LEN(DATE$)+1)=D$
4060 OPEN #1,8,0,DATE$
4065 FOR Q=0 TO 23
4070 HAR=HAR(Q)
4072 PRINT #1,HAR:NEXT Q
4075 CLOSE #1
4080 DATE$=""
5000 REM ******DATE ROLLOVER********
5010 IF K=4 OR K=6 OR K=9 OR K=11 THEN DMAX=30:GOTO 5025
5015 IF K=2 THEN DMAX=28:GOTO 5025
5020 DMAX=31
5025 D=VAL(D$)
5030 D=D+1
5035 IF D>DMAX THEN D=1:READ M$:K=K+1
5040 D$=STR$(D)
5050 GOTO 1055
10000 DATA D:JAN.,D:FEB.,D:MAR.,D:APR.,D:MAY.,D:JUN.,D:JUL.,D:AUG.,D:SEP.,D:OCT.
,D:NOV.,D:DEC.
10010 DATA RES
12000 REM READ OUT ROUTINE********
12010 ? "DO YOU WANT TO READ FILE?"
12020 INPUT B$
12030 IF B$="YES" THEN G=2:GOTO 12040
12040 ? "}":GRAPHICS 0:? "ENTER FILE SPEC PLEASE"
12050 INPUT Z$
12055 GOSUB 13000
12056 PLOT 29,75
12057 ? Z$
12060 OPEN #1,4,0,Z$
12070 FOR Q=0 TO 23
12075 FOR X=29 TO 149 STEP 5
12080 LET HAR(Q)=HAR
12085 IF X=149 THEN 12130
12090 INPUT #1,HAR
```

```
12098 LET Y=75-14*(HAR-2):Y=INT(Y)
12100 IF HAR<2 THEN Y=79
12120 DRAWTO X,Y:NEXT X:NEXT Q
12130 CLOSE #1
12140 ? "DO YOU WANT TO SEE ANOTHER FILE?"
12150 INPUT B$
12160 IF B$="YES" THEN 12040
12170 IF B$="NO" THEN ? "}":? "FINISHED"
12175 FOR D=1 TO 200:NEXT D:? "}":GRAPHICS 2:POSITION 3,1:? #6;"SDDEV ";SDDEV:PO
SITION 15,1:? #6;"      "
12177 GOTO 1055
13000 TRAP 13000:? "}":GRAPHICS 7
13010 COLOR 1
13015 PLOT 16,0
13020 DRAWTO 16,79:DRAWTO 159,79
13030 IF G=2 THEN ? "HOUR 0      6       12      18      24"
13035 IF G=0 THEN ? "HOUR  ";HOURS-24;"      ";HOURS-18;"      ";HOURS-12;"      ";
HOURS-6;"      ";HOURS:? "TODAY"
13036 G=1
13040 PLOT 32,4:DRAWTO 36,12:DRAWTO 40,4
13045 PLOT 51,4:DRAWTO 49,6:DRAWTO 49,10:DRAWTO 51,12:DRAWTO 54,12:DRAWTO 57,10:
DRAWTO 57,6:DRAWTO 54,4
13046 DRAWTO 51,4
13050 PLOT 67,4:DRAWTO 67,12:DRAWTO 73,12
13055 PLOT 80,4:DRAWTO 88,4:PLOT 84,4:DRAWTO 84,12
13060 PLOT 96,9:DRAWTO 96,10:DRAWTO 98,12:DRAWTO 101,12:DRAWTO 103,10:DRAWTO 103
,9:DRAWTO 96,6:DRAWTO 96,5
13065 DRAWTO 99,4:DRAWTO 101,4:DRAWTO 103,5:DRAWTO 103,6
13070 PLOT 14,5:DRAWTO 18,5
13072 PLOT 14,19:DRAWTO 18,19
13074 PLOT 14,33:DRAWTO 18,33
13076 PLOT 14,47:DRAWTO 18,47
13078 PLOT 14,61:DRAWTO 18,61
13080 PLOT 14,75:DRAWTO 18,75
13082 PLOT 9,74:DRAWTO 10,73:DRAWTO 11,73:DRAWTO 12,74:DRAWTO 9,77:DRAWTO 12,77
13084 PLOT 9,45:DRAWTO 9,47:DRAWTO 12,47:PLOT 11,45:DRAWTO 11,49
13086 PLOT 12,18:DRAWTO 11,17:DRAWTO 10,17:DRAWTO 9,18:DRAWTO 9,21:DRAWTO 10,22:
DRAWTO 11,22:DRAWTO 12,21
13088 DRAWTO 11,20:DRAWTO 10,20
13089 TRAP 40000
13090 RETURN
13100 FOR Q=1 TO 24:LET X=29+5*(Q-1):LET HAL=HAL(Q)
13110 LET Y=75-14*(HAL-2):Y=INT(Y)
13120 IF HAL<2 THEN Y=79
13122 IF X=29 THEN PLOT X,Y
13125 IF X>29 THEN DRAWTO X,Y
13127 FOR D=1 TO 20:NEXT D:NEXT Q
13130 RETURN
14020 G=0:FOR S=1 TO 7:IF PEEK(53279)=3 THEN 12040
14021 IF S=1 THEN 14120
14022 IF S=2 THEN 14150
14023 IF S=3 THEN 14160
14024 IF S=4 THEN 14170
14025 IF S=5 THEN 14180
14026 IF S=6 THEN 14190
14027 IF S=7 THEN POSITION 1,5:? #6;".                    ":GOTO 1055
14035 LM=LEN(P$)
14040 POSITION 1,5:? #6;"                    "
14060 TRAP 100:FOR I=1 TO 19:POSITION 20-I,5:? #6;P$(1,I):FOR D=1 TO 8:NEXT D:NE
XT I
14070 TRAP 80:FOR I=2 TO LM-19:POSITION 0,5:? #6;P$(I,19+I):FOR D=1 TO 8:NEXT D:
NEXT I
14080 FOR I=0 TO 5:POSITION 0,5:? #6;P$((LM-18)+I,LM);A$:FOR D=1 TO 8:NEXT D:NEX
T I
14090 POSITION 0,5:? #6;A$:GOTO 14200
14100 FOR I=19-LM TO 1 STEP -1:POSITION I,5:? #6;P$(1,LM):A$:FOR D=1 TO 8:NEXT D
:NEXT I
14110 FOR I=1 TO LM:POSITION 0,5:? #6;P$(I,LM):FOR D=1 TO 8:NEXT D:NEXT I:GOTO 1
4090
14120 P$="THIS ATARI COMPUTER IS BEING USED AS A RECORDING VOLTMETER":GOTO 14035
14150 P$="VOLTAGE TO BE MEASURED AND RECORDED IS BROUGHT IN THROUGH THE JOYSTICK
 PORT":GOTO 14035
14160 P$="AN ANTENNA ON THE ROOF OF THE BUILDING PROVIDES A SOURCE OF FLUCTUATIN
```

```
G VOLTAGE":GOTO 14035
14170 P$="A RECORD OF THE VARIATIONS IS SENT TO THE DISK DRIVE":GOTO 14035
14180 P$="THIS ALLOWS YOU TO SEE GRAPHS OF HOURLY VOLTAGE CHANGES FOR PREVIOUS D
AYS":GOTO 14035
14190 P$="TO SEE PREVIOUS RECORDS PRESS OPTION UNTIL SCREEN CHANGES":GOTO 14035
14200 NEXT S
15000 REM ---CALC MEAN & STD DEV---
15005 SUM=0:SDDEV=0:DFF=0
15010 FOR I=1 TO 24
15020 SUM=SUM+HAL(I):NEXT I
15030 MEAN=SUM/24
15040 FOR I=1 TO 24
15050 DFF=DFF+(HAL(I)-MEAN)^2:NEXT I
15060 SDDEV=SQR(DFF/(23))
15065 POSITION 3,1:? #6;"SDDEV ";SDDEV:POSITION 15,1:? #6;"      "
15070 IF SDDEV>0.5 THEN GOSUB 15100
15080 GOTO 1055
15100 REM -----ALARM TONE-----
15105 POSITION 1,3:? #6;"A.T. ";HOURS;" ";M$;D$:U=0
15110 FOR TONE=1 TO 30
15120 SOUND 0,120,10,10
15125 SOUND 1,85,10,10
15130 FOR D=1 TO 20:NEXT D
15140 SOUND 0,0,0,0
15145 SOUND 1,0,0,0
15150 FOR D=1 TO 10:NEXT D
15160 NEXT TONE
15165 SOUND 0,120,10,10
15166 SOUND 1,85,10,10
15170 RETURN
15200 GRAPHICS 2:POSITION 3,1:? #6;"SDDEV ";SDDEV:POSITION 15,1:? #6;"       "
15210 POSITION 3,3:? #6;"E.T. ";HO;" HRS.  "
15215 POSITION 5,7:? #6;V;" VOLTS  "
15220 POSITION 1,8:? #6;HOURS;":HRS ";MIN;":M. "
15225 ? "SAMPLE #";J;" RUN AVG ";AVG:M$;D$
15230 RETURN
```

"GRASTAT.*"

```
90 DIM HAR(24)
95 DIM Z$(10)
100 GRAPHICS 0:CLOSE #1:OPEN #1,4,0,"K:":POKE 752,1:SETCOLOR 2,0,0
110 DIM SA(300)
120 REM GENERAL INSTRUCTIONS
130 POSITION 15,6:? "STATISTICS":POSITION 18,10:? "FOR":POSITION 11,14:? "NON-ST
ATISTICIANS"
140 FOR K=1 TO 300:SA(K)=0:NEXT K
150 ? ")":POSITION 14,3:? "THIS PROGRAM":POSITION 13,4:? "CALCULATES THE":POSITI
ON 12,5:? "FOLLOWING VALUES"
160 POSITION 9,6:? "FROM THE DATA YOU INPUT:"
170 POSITION 14,9:? "1. MEAN":POSITION 14,11:? "2. STANDARD":POSITION 17,12:? "D
EVIATION"
180 POSITION 14,14:? "4. RANGE"
190 POSITION 9,20:? "PRESS 'C' TO CONTINUE"
200 GET #1,A:IF A<>67 THEN 200
210 REM REQUEST INSTRUCTIONS
220 SUM=0:MEAN=0:DFF=0:SDDEV=0:RG=0
230 ? "}":POSITION 5,12:? "Do you need instructions (y/n)?"
240 GOSUB 1350
250 IF A=89 THEN GOSUB 830
260 REM DATA ENTRY
265 CLOSE #1
270 ? "}":POSITION 9,11:? "Enter Filedate     ":INPUT Z$:N=24
280 IF N>300 OR N<1 THEN FOR I=28 TO 38:POSITION I,11:? CHR$(32):NEXT I:GOTO 270
290 FOR Q=1 TO 2
300 OPEN #1,4,0,Z$
310 FOR I=1 TO 24
320 LET HAR(I)=HAR
330 INPUT #1;HAR
331 NEXT I:CLOSE #1:NEXT Q
335 FOR I=1 TO 24:SA(I)=HAR(I):NEXT I
336 SA(I)=SA
340 FOR J=22 TO 38:POSITION J,12:? CHR$(32):NEXT J
341 OPEN #1,4,0,"K:"
```

```
350 REM ERROR CORRECTION REQUEST
360 ? "}":POSITION 3,12:? "Wish to make any corrections (y/n)?"
370 GOSUB 1350
380 IF A=89 THEN GOTO 1120
390 REM CALCULATION OF MEAN AND STD. DEVIATION
400 ? "}":POSITION 15,12:? "PLEASE WAIT":POSITION 6,14:? "STATISTICS BEING CALCU
LATED"
410 FOR I=1 TO N
420 SUM=SUM+SA(I):NEXT I
430 MEAN=SUM/N
440 FOR I=1 TO N
450 DFF=DFF+(SA(I)-MEAN)^2:NEXT I
460 SDDEV=SQR(DFF/(N-1))
470 REM SORTING THE DATA
480 FL=0
490 FOR I=1 TO N-1
500 IF SA(I)<=SA(I+1) THEN 550
510 Q=SA(I)
520 SA(I)=SA(I+1)
530 SA(I+1)=Q
540 FL=1
550 NEXT I
560 IF FL=1 THEN 480
570 REM CALCULATION OF RANGE
580 RG=SA(N)-SA(1)
590 LR=SA(1):HR=SA(N)
600 REM CALCULATION OF MEDIAN
610 IF N/2<>INT(N/2) THEN 650
620 IF SA(N/2)=SA(N/2+1) THEN MDD=SA(N/2)
630 IF SA(N/2)<>SA(N/2+1) THEN MDD=(SA(N/2)+SA(N/2+1))/2
640 GOTO 670
650 MDD=SA(INT(N/2+1))
660 REM PRINT RESULT TO SCREEN
670 ? "}":POSITION 10,2:? "CALCULATION RESULTS":POSITION 9,4:? "*****  ";Z$;" **
***"
680 POSITION 4,6:? "SAMPLE SIZE:":POSITION 26,6:? N
690 POSITION 4,8:? "MEAN(X BAR)":POSITION 26,8:? INT(MEAN*10000+0.5)/10000
700 POSITION 4,10:? "STD. DEVIATION":POSITION 26,10:? INT(SDDEV*10000+0.5)/10000
710 POSITION 4,12:? "MEDIAN":POSITION 26,12:? MDD
720 POSITION 4,14:? "RANGE":POSITION 26,14:? RG
730 POSITION 4,16:? "LOWEST SAMPLE VALUE":POSITION 26,16:? LR
740 POSITION 4,18:? "HIGHEST SAMPLE VALUE":POSITION 26,18:? HR
750 POSITION 13,22:? "PRESS ANY KEY"
760 GET #1,A
770 REM REQUEST TO CONTINUE OR END
775 GOSUB 1400
780 ? "}":POSITION 4,12:? "Wish to Process more data (y/n)?"
790 GOSUB 1350
800 IF A=78 THEN GRAPHICS 0:END
810 FOR I=1 TO N:SA(I)=0:NEXT I:GOTO 220
820 REM INSTRUCTION SUBROUTINE
830 ? "}":POSITION 5,5:? "The maximum number of enties is    300, while the mini
mun number is 2."
840 POSITION 5,9:? "The mean is the arithmatic average of the numbers you enter.
"
850 POSITION 5,13:? "THE STANDARD DEVIATION is a measure of how widely your numb
ers spread      from the average."
860 POSITION 10,21:? "Press a key for more."
870 GET #1,A
880 ? "}":POSITION 6,3:? "Since the values you enter tend to form a bell curve(N
ormal Dist.), the Std. Deviation"
890 POSITION 17,5:? " is a measure of the     area under the bell curve."
900 POSITION 3,9:? "No of Std.Dev.(+/-)":POSITION 3,10:? "----------------------
-"
910 POSITION 29,9:? "% Area":POSITION 29,10:? "-------"
920 POSITION 4,12:? "(+/-) 1 Std. Dev.":POSITION 31,12:? "68.3"
930 POSITION 4,14:? "(+/-) 2 Std. Dev.":POSITION 31,14:? "95.5"
940 POSITION 4,16:? "(+/-) 3 Std. Dev.":POSITION 31,16:? "99.7"
950 POSITION 4,18:? "(+/-) 4 Std. Dev.":POSITION 31,18:? "99.9"
960 POSITION 10,21:? "Press a key for more"
970 GET #1,A
980 ? "}":POSITION 4,5:? "The MEDIAN is the value at the      mid-Point of your d
ata."
```

```
990 POSITION 4,10:? "The RANGE is the difference between your lowest data value"
1000 POSITION 25,11:? "and the highest. The range is a quick-and-dirty"
1010 ? "estimate of the spread. The standard  deviation is more reliable tha the
    spread."
1020 POSITION 10,20:? "Press 'S' to start"
1030 GET #1,A:IF A<>83 THEN 1030
1040 RETURN
1050 REM DISPLAY CORRECTION OPTIONS
1060 ? "}":POSITION 1,23:? "C=CHANGE DATA   N=NEXT TABLE    Q=QUIT":RETURN
1070 GET #1,A:IF A<>67 AND A<>78 AND A<>81 THEN 1070
1080 IF A=78 THEN 1280
1090 IF A=81 THEN 390
1100 ? "}"
1110 REM ERROR CORRECTION SUBROUTINE
1120 ? "}":POSITION 3,12:? "Remember incorrect sample # (y/n)?"
1130 GOSUB 1350
1140 IF A=78 THEN 1230
1150 ? "}":POSITION 7,9:? "What is the sample #";:INPUT EN
1160 IF EN>N OR EN<1 OR EN<>INT(EN) THEN 1150
1170 POSITION 7,11:? "Sample ";EN:POSITION 22,11:? "Value ";SA(EN)
1180 POSITION 7,13:? "Enter your new value":POSITION 7,14:INPUT C:SA(EN)=C:HAR(E
N)=C
1190 POSITION 7,19:? "Any more changes (y/n)?"
1200 GOSUB 1350
1210 IF A=89 THEN 1120
1220 GOTO 400
1230 GOSUB 1060:POSITION 5,2:? "These are the first ten values:"
1240 POSITION 11,5:? "ENTRY":POSITION 22,5:? " VALUE"
1250 FOR K=1 TO 10
1260 POSITION 12,K+7:? K:POSITION 24,K+7:? SA(K):NEXT K
1270 GOTO 1070
1280 POSITION 5,2:? "These are the next ten values:      ":IF K<=300 THEN GOSUB 1
340
1290 CT=8:FOR K=K TO K+9
1300 IF K>300 THEN K=K+9:NEXT K:GOTO 400
1310 POSITION 12,CT:? K:POSITION 24,CT:? SA(K)
1320 CT=CT+1:NEXT K
1330 GOTO 1070
1340 FOR J=1 TO 10:POSITION 12,J+7:? "    ":POSITION 24,J+7:? "                ":
NEXT J:RETURN
1350 GET #1,A:IF A<>89 AND A<>78 THEN 1350
1360 RETURN
1400 ? "DO YOU WISH TO PRINT RESULTS?"
1410 GET #1,A:IF A<>89 AND A<>78 THEN 1410
1420 IF A=78 THEN RETURN
1430 LPRINT :LPRINT :LPRINT "24 HOUR VOLTAGE STATISTICS FOR ";Z$:LPRINT
1440 LPRINT "SAMPLE SIZE...............";N:LPRINT
1445 LPRINT "MEAN(X BAR)...............";INT(MEAN*10000+0.5)/10000:LPRINT
1450 LPRINT "STD.DEVIATION.............";INT(SDDEV*10000+0.5)/10000:LPRINT
1455 LPRINT "MEDIAN....................";MDD:LPRINT
1460 LPRINT "RANGE.....................";RG;:LPRINT
1465 LPRINT "LOWEST SAMPLE VALUE.......";LR:LPRINT
1470 LPRINT "HIGEST SAMPLE VALUE.......";HR:LPRINT
1480 ? "DO YOU WANT TO PRINT GRAPHICS?"
1490 GET #1,A:IF A<>89 AND A<>78 THEN 1490
1495 IF A=78 THEN RETURN
1500 GOTO 11960
11960 DIM B$(3)
11990 DIM D$(3),M$(8),DATE$(15),X$(5)
12000 REM ***READ OUT ROUTINE*******
12055 GOSUB 13000
12057 CLOSE #1
12060 X=58
12062 U=HAR(1)
12065 FOR I=1 TO 23:HAR(I)=HAR(I+1):NEXT I
12066 HAR(24)=U
12070 FOR I=1 TO 24
12080 LET HAR=HAR(I)
12090 LET Y=150-28*(HAR-2):Y=INT(Y)
12100 IF HAR<2 THEN Y=158
12115 IF X=58 THEN PLOT 58,Y
12120 IF X>58 THEN DRAWTO X,Y
12122 X=X+10
12125 NEXT I
```

```
12130 CLOSE #1
12132 GOSUB 14000:GOTO 13990
12140 ? "DO YOU WANT TO SEE ANOTHER FILE?"
12150 INPUT B$
12160 IF B$="YES" THEN 12040
12170 IF B$="NO" THEN ? "}":? "FINISHED"
12175 FOR D=1 TO 1000:NEXT D:GOTO 12000
13000 ? "}":GRAPHICS 8+16
13010 COLOR 1
13015 PLOT 32,0
13020 DRAWTO 32,158:DRAWTO 318,158
13040 PLOT 64,8:DRAWTO 72,24:DRAWTO 80,8
13045 PLOT 102,8:DRAWTO 98,12:DRAWTO 98,20:DRAWTO 102,24:DRAWTO 108,24:DRAWTO 11
4,20:DRAWTO 114,12:DRAWTO 108,8
13046 DRAWTO 102,8
13050 PLOT 134,8:DRAWTO 134,24:DRAWTO 146,24
13055 PLOT 160,8:DRAWTO 176,8:PLOT 168,8:DRAWTO 168,24
13060 PLOT 192,18:DRAWTO 192,20:DRAWTO 196,24:DRAWTO 202,24:DRAWTO 206,20:DRAWTO
 206,18:DRAWTO 192,12:DRAWTO 192,10
13065 DRAWTO 198,8:DRAWTO 202,8:DRAWTO 206,10:DRAWTO 206,12
13070 PLOT 28,10:DRAWTO 36,10
13072 PLOT 28,38:DRAWTO 36,38
13074 PLOT 28,66:DRAWTO 36,66
13076 PLOT 28,94:DRAWTO 36,94
13078 PLOT 28,122:DRAWTO 36,122
13080 PLOT 28,150:DRAWTO 36,150
13082 PLOT 18,148:DRAWTO 20,146:DRAWTO 22,146:DRAWTO 24,148:DRAWTO 18,154:DRAWTO
 24,154
13084 PLOT 18,90:DRAWTO 18,94:DRAWTO 24,94:PLOT 22,90:DRAWTO 22,98
13086 PLOT 24,36:DRAWTO 22,34:DRAWTO 20,34:DRAWTO 18,36:DRAWTO 18,42:DRAWTO 20,4
4:DRAWTO 22,44:DRAWTO 24,42
13088 DRAWTO 22,40:DRAWTO 20,40
13090 PLOT 16,165:DRAWTO 16,169:PLOT 16,167:DRAWTO 19,167:PLOT 19,165:DRAWTO 19,
169:PLOT 24,165
13092 DRAWTO 23,166:DRAWTO 23,168:DRAWTO 24,169:DRAWTO 25,169:DRAWTO 26,168:DRAW
TO 26,166:DRAWTO 25,165
13094 DRAWTO 24,165
13096 PLOT 29,165:DRAWTO 29,168:DRAWTO 30,169:DRAWTO 31,169:DRAWTO 32,168:DRAWTO
 32,165
13098 PLOT 36,165:DRAWTO 36,169:PLOT 37,166:DRAWTO 38,165:DRAWTO 39,165
13100 PLOT 45,165:DRAWTO 43,165:DRAWTO 42,166:DRAWTO 43,167:DRAWTO 44,167:DRAWTO
 45,168:DRAWTO 44,169
13102 DRAWTO 42,169
13104 PLOT 57,165:DRAWTO 56,166:DRAWTO 56,168:DRAWTO 57,169:DRAWTO 58,169:DRAWTO
 59,168:DRAWTO 59,166
13106 DRAWTO 58,165:DRAWTO 57,165
13108 PLOT 107,165:DRAWTO 105,165:DRAWTO 104,166:DRAWTO 104,168:DRAWTO 105,169:D
RAWTO 106,169:DRAWTO 107,168
13110 DRAWTO 106,167:DRAWTO 105,167
13112 PLOT 163,165:DRAWTO 163,169
13114 PLOT 166,166:DRAWTO 167,165:DRAWTO 168,165:DRAWTO 169,166:DRAWTO 166,169:D
RAWTO 169,169
13116 PLOT 220,165:DRAWTO 220,169
13118 PLOT 224,165:DRAWTO 223,166:DRAWTO 224,167:DRAWTO 223,168:DRAWTO 224,169:D
RAWTO 225,169:DRAWTO 226,168
13120 DRAWTO 225,167:DRAWTO 226,166:DRAWTO 225,165
13122 PLOT 275,166:DRAWTO 276,165:DRAWTO 277,165:DRAWTO 278,166:DRAWTO 275,169:D
RAWTO 278,169
13124 PLOT 281,165:DRAWTO 281,167:DRAWTO 284,167:PLOT 283,165:DRAWTO 283,169
13500 RETURN
13990 GOSUB 15000
14000 TRAP 14030:RESTORE :READ N:DIM P$(N)
14020 FOR I=0 TO N-1
14030 READ J
14040 POKE ADR(P$)+I,J
14050 NEXT I
14065 READ I
14067 RETURN
14989 REM
14990 REM
14991 REM *** DUMP PICTURE ON SCREEN
14992 REM
14993 REM ***
```

```
14994 REM
14995 REM ***
14996 REM
14997 REM ***
14998 REM
14999 REM
15000 X=USR(ADR(P$),INV)
15010 IF X THEN GRAPHICS 0:PRINT "DUMP ERROR -":END
15020 RETURN
15030 REM
19990 REM
19991 REM
19992 REM
19993 REM
19994 REM
19995 REM
19996 REM
19997 REM
20000 DATA 566
20010 DATA 169,96,133,7,32,7,0,186,202,189,0,1,56,233,6,133
20020 DATA 204,232,189,0,1,233,0,133,205,160,195,177,204,201,240,144
20030 DATA 123,165,204,24,105,97,133,206,165,205,105,0,133,207,160,0
20040 DATA 177,206,200,17,206,240,101,136,177,206,24,101,204,133,208,200
20050 DATA 177,206,101,205,133,209,140,2,6,160,1,177,208,24,101,204
20060 DATA 145,208,200,177,208,41,15,101,205,145,208,172,2,6,200,208
20070 DATA 207,193,0,203,0,209,0,225,0,230,0,248,0,255,0,68
20080 DATA 1,91,1,114,1,119,1,129,1,145,1,149,1,155,1,193
20090 DATA 1,205,1,211,1,217,1,10,2,42,2,0,0,170,240,5
20100 DATA 104,104,202,208,251,169,132,133,212,134,213,96,104,201,1,208
20110 DATA 236,104,141,7,6,104,13,7,6,240,4,169,0,240,2,169
20120 DATA 255,141,3,6,186,142,0,6,169,0,133,212,133,213,141,5
20130 DATA 6,32,122,241,162,48,169,3,157,66,3,173,43,242,157,68
20140 DATA 3,173,44,242,157,69,3,169,8,157,74,3,169,0,157,75
20150 DATA 3,32,17,242,169,10,32,230,241,166,88,164,89,134,204,132
20160 DATA 205,169,28,141,7,6,160,0,185,48,242,201,255,240,6,32
20170 DATA 230,241,200,208,243,160,0,165,204,72,165,205,72,162,0,177
20180 DATA 204,157,9,6,165,204,24,105,40,133,204,144,2,230,205,232
20190 DATA 224,8,144,235,138,72,152,72,169,8,141,8,6,162,0,138
20200 DATA 30,9,6,106,232,224,8,144,247,77,3,6,174,7,6,202
20210 DATA 208,2,41,7,32,162,241,206,8,6,208,225,104,168,104,170
20220 DATA 104,133,205,104,133,204,200,192,40,144,172,32,132,241,165,204
20230 DATA 24,105,24,133,204,165,205,105,1,133,205,206,7,6,208,134
20240 DATA 169,15,32,230,241,169,155,32,230,241,162,48,169,12,157,66
20250 DATA 3,76,17,242,173,5,6,240,21,72,169,0,141,5,6,169
20260 DATA 28,32,230,241,104,32,230,241,173,6,6,32,230,241,169,10
20270 DATA 208,68,9,128,174,5,6,240,52,205,6,6,208,8,238,5
20280 DATA 6,208,50,206,5,6,72,173,5,6,201,3,176,13,173,6
20290 DATA 6,32,230,241,206,5,6,208,245,240,17,169,28,32,230,241
20300 DATA 173,5,6,32,230,241,173,6,6,32,230,241,104,141,6,6
20310 DATA 169,1,141,5,6,96,140,2,6,141,4,6,162,48,169,11
20320 DATA 157,66,3,169,1,157,72,3,169,0,157,73,3,157,75,3
20330 DATA 169,4,157,68,3,169,6,157,69,3,32,17,242,172,2,6
20340 DATA 96,32,86,228,162,48,189,67,3,16,245,133,212,174,0,6
20350 DATA 154,162,48,169,12,157,66,3,76,86,228,45,242,80,58,155
20360 DATA 8,27,16,0,74,255,63093
25000 LPRINT :LPRINT :LPRINT "          ";Z$
```

We claim:

1. A method for providing an early warning of the future occurrence of an earthquake, comprising the following steps:

(a) measuring the field strength of at least one broadcast radio frequency signal at a location separated from the place of transmission of said signal, and

(b) providing a humanly-sensible indication if the strength of said signal decreases beyond a predetermined amount from a previous value thereof.

2. The method of claim 1 wherein said measuring is done on a broadband basis so as to measure the strength of a plurality of radio signals.

3. The method of claim 1 wherein said measuring is performed by rectifying and filtering said radio frequency to provide a direct current voltage and wherein said indication is provided in response to a predetermined drop in the value of said voltage.

4. The method of claim 1 wherein said indication is provided if the strength of said signal decreases a predetermined amount from an average value thereof.

5. The method of claim 4 wherein the value of said signal is periodically sampled for a predetermined period, a standard deviation of the resultant samples is calculated, and said indication is provided if said standard deviation exceeds a predetermined value.

6. The method of claim 4 wherein said signal is sampled once per minute, the resultant minute samples are averaged each hour, and the resultant hourly averages are tested to determine if the latest hourly average has

deviated from previous hourly averages beyond a predetermined amount.

7. The method of claim 1 wherein said indication is provided only if at least wo separated receivers detect said predetermined drop in the strength of said signal.

8. The method of claim 1 wherein a visible record of said signal's field strength is plotted as it is measured.

9. The method of claim 1 wherein the strength of a plurality of broadcast signals are measured by rectifying and filtering said signals to provide a direct current voltage, said direct current voltage is periodically sampled, the resultant samples are averaged periodically to provide periodic averages, the standard deviation of said periodic averages is calculated, and said indication is provided if the value of said standard deviation exceeds a predetermined value.

10. A system for providing an early warning of the future occurrence of an earthquake, comprising:

(a) means for measuring the field strength of a least one broadcast radiofrequency signal, said means being arranged to measure said field strength at a location separated from the place of transmission of said signal, and

(b) means responsive to the measured strength of said signal for providing a humanly-sensible indication if said measured field strength decreases beyond a predetermined amount from an average value thereof.

11. The system of claim 10 wherein said means for measuring comprises a broadband receiver for measuring the strength of a plurality of radio signals.

12. The system of claim 10 wherein said means for measuring comprises means for rectifying and filtering said radio frequency signal to provide a direct current voltage and wherein means for providing said indication is arranged to do so in response to a predetermined drop in the value of said voltage.

13. The system of claim 11 wherein said means for providing said indication is arranged to do so if the strength of said signal decreases a predetermined amount from an average value thereof.

14. The system of claim 13 wherein said means for measuring comprises means for periodically sampling the value of said radio signal for a predetermined period, and wherein said means for providing said indication is arranged to calculate the standard deviation of the resultant samples and to provide said indication if said standard deviation exceeds a predetermined value.

15. The system of claim 13 wherein said means for periodically sampling said signal is arranged to take a sample once per minute and to average the resultant minute samples each hour, and wherein said means for providing said indication is arranged to do so by testing the resultant hourly averages to determine if the latest hourly average has deviated from previous hourly averages beyond a predetermined amount.

16. The system of claim 10 wherein said means for providing said indication is arranged to do so only in response to the detection of a predetermined drop in the strength of said signal by two separated receivers.

17. The system of claim 10 further including means for making a visible record of said signal's field strength as it is measured.

18. The system of claim 10 wherein said means for measuring includes means for measuring the strength of a plurality of broadcast signals by rectifying and filtering said signals to provide a direct current voltage, means for periodically sampling said direct current voltage, means for averaging the resultant samples periodically to provide periodic averages, and wherein said means for providing said indication includes means for calculating the standard deviation of said periodic averages, and providing said indication if the value of said standard deviation exceeds a predetermined value.

19. The system of claim 10 wherein said means for measuring and providing said indication comprises a broadband receiver arranged to provide a direct current output voltage, an analog to digital converter, and a programmed computer arranged to receive the outpt of said converter.

20. The system of claim 19 further including means for periodically storing received field strength values and providing a visible plot of the continuous value of said field strength.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 4,628,299
DATED : 1986 Dec 9
INVENTOR(S) : J. B. Tate and D. E. Brown

It is certified that error appears in the above-identified patent and that said Letters Patent are hereby corrected as shown below:

Col. 1, line 6, change "fugure" to --future--.
Col. 3, line 24, change "ouo" to --out--.
Col. 8, line 25, change "specifications" to --"specificities--.
Col 8, near bottom, delete handwriting at top of program listing reading "8299 Issue of 12-9".

Signed and Sealed this

Tenth Day of March, 1987

Attest:

DONALD J. QUIGG

Attesting Officer *Commissioner of Patents and Trademarks*

rexresearch.com

John BEDINI

Scalar Beam Generator

Ad
ORACLE NETSUITE
LEARN HOW TO CRUSH THE FIVE BARRIERS TO GROWTH
Claim guide now
The #1 Cloud ERP

fashionstock
shutterstock

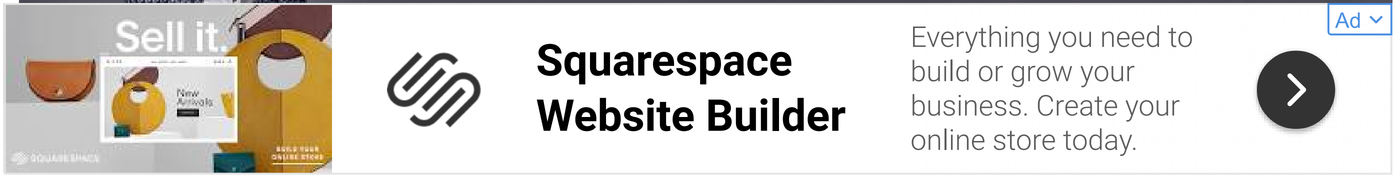
Ad
Sell it.
Squarespace Website Builder
Everything you need to build or grow your business. Create your online store today.

---

http://amasci.com/freerng/bedini.html

John Bedini's "Scalar Beam" device

Here's a suggestion for "scalar" experiments from a conversation with John Bedini. Mr. Bedini encourages everyone to try this experiment, but warns us that this device is patent applied for, so you should only build a single unit for your own use.

```
   ________
  |\       \
  | \       \
  |\ \       \
  | \ \________\ S
   \  \ |       |     Obtain two Radio Shack ceramic magnets and
    \  \|________| N    glue their north pole faces together.
      \ |       | N
       \|________|
              S
```

```
     ________        Wind the magnets with about 50 turns
    |\  \\\ \      of #30 magnet wire.  Wire gauge is not
    | \ \\\\ \     critical.
```

```
|\  \  \\\\  \
| \  \ _\\\\__\
 \  \ |  |||| |
   \  \|__||||_|
     \ |  |||| |
       \|__||||_|
          \ |      ________
          | |    [ small, ]
          |  -----[ noisy  ]----------o
          |      [_motor__]         6v to 12v power supply
          |
          |______________________________o
```

The brush noise from the DC motor provides a pulse signal to the coil, which modulates the 'colliding' field pattern of the magnets and creates interesting scalar effects within a narrow pencil-beam pattern which extends from each face of the magnet out to a few inches.

```
   ________
|\   \\\  \
| \  \\\\  \
|\  \  \\\\  \
<<<<<<<<<<<\  \ _\\\\__\>>>>>>>>>>>>>>>>>>>>> scalar effect comes from the
<<<<<<<<<<<<\  |  |||| |>>>>>>>>>>>>>>>>>>>>> joint between magnet faces
   \  \|__||||_|
     \ |  |||| |
       \|__||||_|
          \ |
          | |
          | |
          | |
```

Mr. Bedini suggests these experiments:

Purchase two identical music CDs. Listen to both to verify that they are identical. Now let the "scalar beam" play all over the surface of one of the CDs for about one minute. You may want to build a simple rotating platform to make this process more convenient. Now play the two CDs and compare them again. Hear any difference? (Note, this process is patent pending, so do not use it for any other purpose except to demonstrate the reality of the effect)

Connect a small probe-coil to an oscilloscope, then move it around in the beam and observe the waveforms.

Taste some wine, then put it in a small airtight container and place it against the magnet face for a few (minutes? hours?) Taste it again. Improvements? Try it with and without the power supply connected to verify that any changes are caused by the scalar beam and by just the magnetic field.

FROM : Some tests I intend to try (but as yet have not!):

Place various foodstuffs in the beam then compare flavor with untreated samples.

Grow two collections of plants, water one with normal water, water the other with water that's been treated by several minutes??hours?? exposure to the beam. As a control, use water which was held nearby identical magnets but without the coil.

Aim the beam directly at a plant for many days, compare it with another untreated plant as a control. (Shield the magnet, or place a similar magnet-block near the control plant.)

Sprout two groups of seeds, one treated and one untreated, and look for differences in number, health, growth rate, etc., between the two groups.

Measure the growth of the tip of a plant stem by using a tiny lever, mirror, and laser beam. Graph the growth rate, then treat the plant with the scalar beam and look for changes in the growth rate. (Note that this method can also be used to observe plants' realtime response to numerous stimuli both conventional and "weird." Fertilizer? Light? Music? Magnetism? Pyramids? Good/Bad thoughts?)

Observe microscopic lifeforms in pond water, then expose them to the beam and see if their behavior changes while it is operating. Or, expose the water to the beam for several minutes??hours??, then compare the number and activity of lifeforms in the water with an untreated bottle. Or, compare the effects of adding treated or untreated water to the slide under the microscope.

Use an opamp buffer and an audio amplifier to listen to the noise output of a capacitor which is shielded in a thick copper box, (or does a resistor or transistor work better?) then aim the beam at the box and listen for signals, or monitor changes to component values. See: http://amasci.com/freenrg/grav3.txt, Hodowanec's capacitor-based gravity detector, for more info.

Apparatus and method for reducing electronic relaxation noise present information recording medium

US5487057

A method and apparatus is provided for reducing relaxation noise in a conducting medium. The device is fabricated by affixing two magnets at like, repelling poles; wrapping said magnets with a coil of wire in an orientation orthogonal to the interface between the joined magnets and the like poles of the magnets to form a magnetic unit; connecting said coil to a motor means, an electrical power supply means and a switch means; attaching a spindle to said motor means. The spindle receives said conducting medium. The apparatus can be in a housing. The conducting medium is placed on said spindle. By activating the device, a modulated magnetic electromagnetic field is created simultaneous to the spinning of the conducting medium. The information recording medium rotates through the modulated electromagnetic field, thereby reducing the relaxation noise.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The invention relates to the general field of playback of recorded conducting media or conductors, more specifically, an apparatus and method that reduces electronic relaxation noise, also known as "linger noise", that exists in information recording medium.

2. Description of Related Art and Information

The utility of the present invention is based on fundamental laws of nuclear and electronic physics at the electron level, particularly with respect to electron gas relaxation phenomena in conductors immediately preceding the initiation of normal current flow per ohms law.

For current to flow through a medium conductor or a conducting plate (conductor) that involves electromagnetic (EM) effects in recording and playback, and in transmission of signals, and thus to provide signals through normal circuitry, conducting plates, etc., a finite time is required before the flow is established, and before stability, referred to as a "stable signal" or stable flow pattern, can be established. This time delay is referred to as electron gas relaxation time.

The relaxation phenomenon involves at least three major stages, all of which must be substantially completed before the relaxation time is actually complete, and coherent signals or patterns are being transmitted in the conductor. Further, the relaxation time involves damped oscillations.

The first stage in the relaxation time is the relaxation of the electrical charge density. Electrons initially distributed throughout the conductor are "excited" by taking on excess energy in the form of potential gradients across them. They are now substantially involved in working their way to the surface, since most of the current flow must occur on the surface of the conductor, not in its interior. If the total charge is not zero, a very highly non-linear first phase results, and the division between first and second stages is substantially blurred.

A simplifying assumption is made, that the initial charge is zero, so that the compensated charge fluctuations may be described by linear equations. Then the relaxation of electrical charge density can be assumed to be mostly independent of the initial conditions and of the size and shape of the conductor.

Treating the Drude electron gas model as applicable, and assuming each electron is independent of the rest, damped harmonic oscillator equations result for the relaxation. This yields a short relaxation time for stage 1. This erroneously short relaxation time basically results from assuming that the electrons move rather independently and are de-coupled. With coupling remaining as is almost always the case, a combination of stage 1 and stage 2 actually applies immediately, and sometimes an extra combination of stage 3 as well Much "noise" known as small field perturbations in random directions, is present during phase 1. This is because of the random motion of the electrons in all directions, as the addition of the excess energy to them essentially results in an increase in the average electron's kinetic energy and increases the violence and frequency of their collisions with the lattice and with each other.

The second stage is the expulsion of the electric and magnetic fields to the exterior of the conductor, and the expulsion of the excited electrons as axial currents (on the average) to the surface. This electron movement during the second stage is still immersed in the midst of a great collision/violence among the electrons in the electron gas in the conductor. Accordingly, there are erratic electric and magnetic fields from the erratically moving charges, whose violent movements are, in fact, expelling these fields from the conductor and in all directions in the conductor. Consequently, there is much noise, distortion and scattering going on in this stage.

In the third stage, the electrons reach the skin of the conductor, resulting in a marked decrease in collision frequency and violence. The relaxation process terminates with the slower ohmic and radiative damping of the surface currents.

As an end result of the three relaxation stages, a sinusoidal charge disturbance is formed and propagated in the conductor with a phase velocity Vp of roughly 1.times.10@8 cm/sec, or roughly 10@6 meters per second. Further, this disturbance is extremely noisy. It is also moving far slower than the speed of light, hence "lingers" in a circuit after each and every stimulation, directly adding noise to slightly succeeding "signal stimuli." The "signal" moves down the conductor at nearly C, the velocity of light in space.

The net result in a continual digital process is that the 3-stage relaxation phenomena continually generates lingering noise signals in a circuit. This noise is continually produced both when signals are initially recorded on a conductor, increasing the noise in the signal actually recorded, and again when the conducting medium is later optically re-stimulated to detect the recorded "signal and noise". Examples of conducting media are audio compact disc, CD-ROM (read only memory), video laser disc, photo-CD and photgraphic film. The net result is that continual relaxation noise is added to both the recording and playback stages of any recording process utilizing such a conducting medium. Relaxation noise is also added to any signal translated down or in a conductor and through circuits.

In searching prior art that disclosed magnetic and/or modulated EM devices designed for processing or clarifying conducting media for the purpose of reducing relaxation noise, no references were located that either disclosed or anticipated the present invention.

SUMMARY OF THE INVENTION

The device is a housed magnetic unit that, when electrically activated by a switch and motor means, creates a modulated EM beam or field that dampens and reduces electron relaxation or linger noise in a recorded conducting medium as the conducting medium passes through the EM field.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a perspective view of a device that produces a permanent magnet magnetic field.

FIG. 2 is a perspective view of the outside of a housing for the apparatus.

FIG. 3 is an inside view of the bottom half of the housing and components.

FIG. 4 is a perspective view of the orientation of the conducting medium (compact disc), to a molded spindle.

FIG. 5 is an inside view of the top half of the housing and components.

FIG. 6 is a side view of the housing and components.

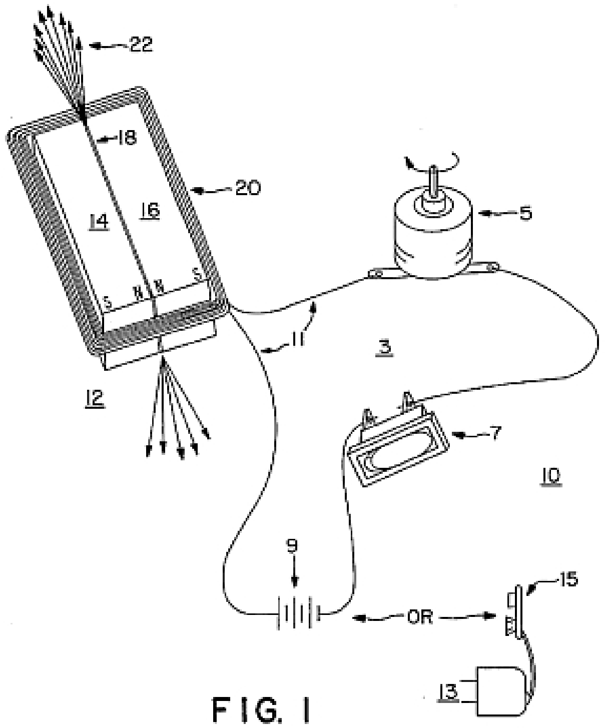

FIG. 1

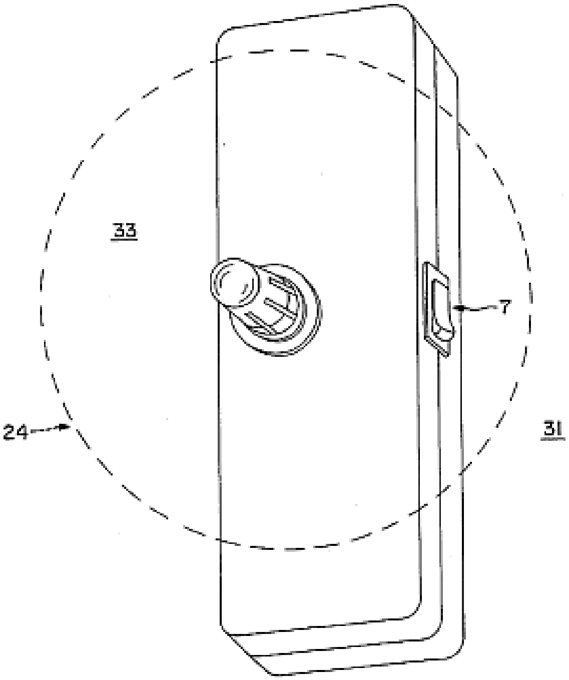

FIG. 2

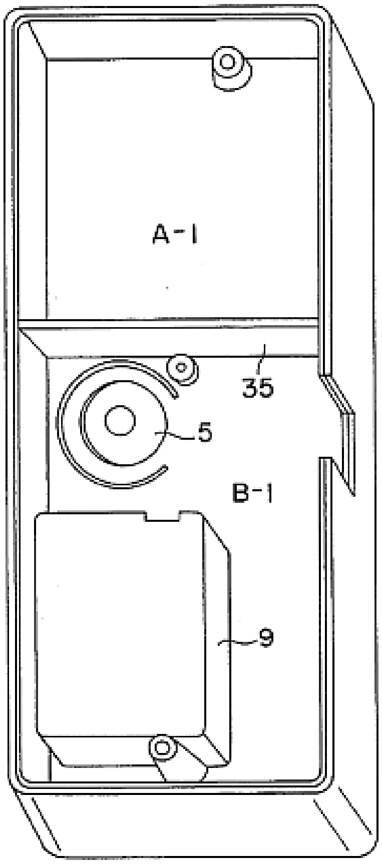

FIG. 3

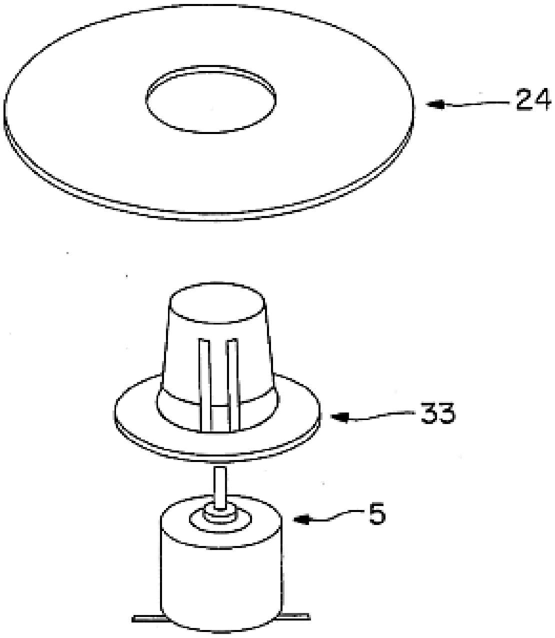

FIG. 4

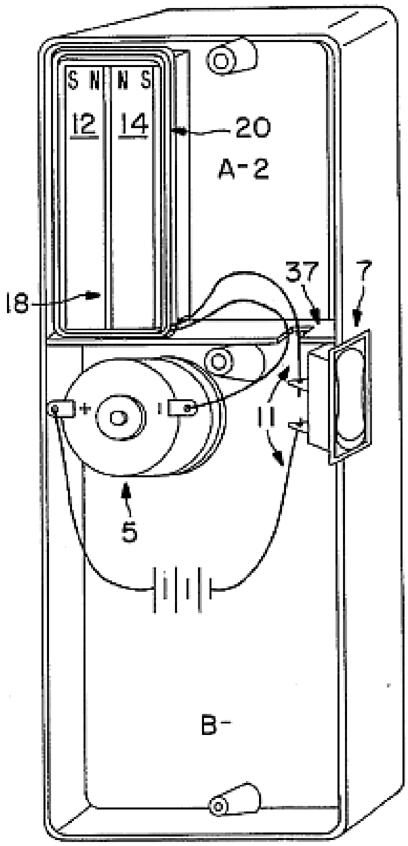
S N
N S
12
14
20
A-2
18
37
7
+
-
11
5
B-

FIG. 5

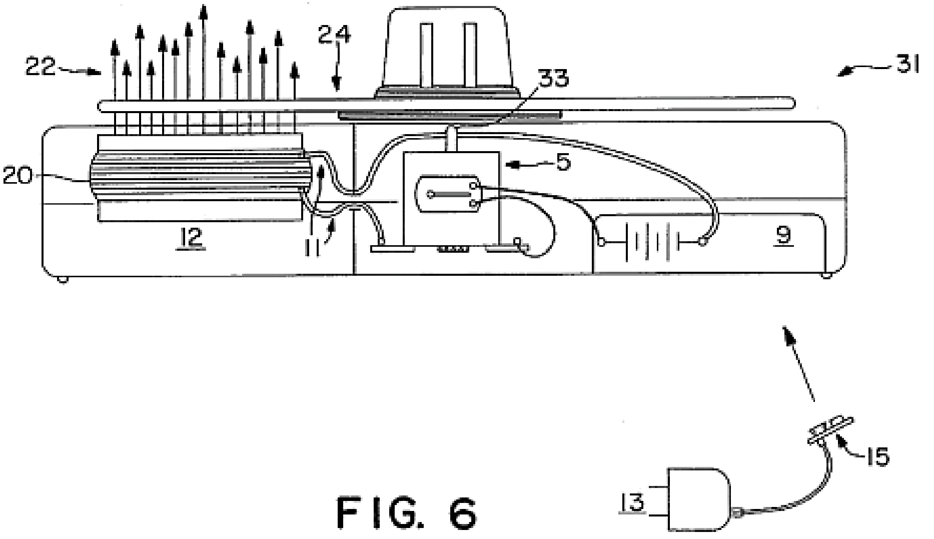

FIG. 6

DETAILED DESCRIPTION OF THE INVENTION

The method and apparatus or device of the present invention is discussed hereinafter in terms of 3-stage relaxation noise actually introduced to all signal conducting media by the recording process, and again introduced in the playback from such recorded conducting media.

First, the sharp distinction between the first and second relaxation stages is removed. The moment the charges present in the electron gas in a conductor are excited with excess energy from the leading part of the incoming signal stimulus, excess fields exist on the electrons, directly combining with the E- and B-fields already in existence. Consequently, stages 1 and 2 are considered combined. In other words, even in stage one, a strong component of the excited fields is attempting to axially eject or expel electrons from the conductor, while other strong fields oriented within the conductor are affecting the entire electron gas, increasing its collision frequency and noise. Axially, there is less collision damping, so on the average an axial movement toward the surface results. Axial expulsion is a radial effect.

Next is the effect of an additional magnetic field from an external magnetic pole, which is what the present invention provides. The magnetic pole, for example, may be fairly localized in the case of recording signals and replaying signals, or it may be distributed in a "cover" or "blanket" covering the outside surface of the conductor. In the case of flat conducting plates, the blanket may cover both sides.

When any field element attempts to emerge from the conductor, or can be decomposed into longitudinal and axial components where the axial component attempts to emerge, it is directly coupled to one or more axially (on the average) moving electrons within the plate, which constitutes the source or sources of the fields. The source electrons are also attempting to move to the surface themselves. However, the strength of the electron-associated magnetic field will be altered, affected and resisted by the externally introduced modulated EM field of the present invention.

The net effect of the magnetic field, as produced by the present invention, is to induce damping and "magnetic braking" of the electrons and damping of their fields. In addition, a moving electron in a magnetic field is forced to turn away from its path, again increasing damping. In the externally introduced modulated EM field of the present device, the axially (on the average) moving electron is forced to develop an electric field gradient at right angles to its (on the average) axial movement, and "back against" the "linger noise" stimulus. In short, the electron develops a "back EMF" from all of the other electrons in the gas, which resist its movement and damp it. This dampens the electron's axial component, consequently its movement and concomitant relaxation "stage 1" and "stage 2" noise, resulting in a materially reduced relaxation noise.

The result of the use of the present invention, is that as the electrons continue to violently alter their directions due to collisions, they are continually damped by the modulated EM field created by the device. Consequently, the erratic movement of the electrons are constrained, their velocities are lowered, their collision frequency is diminished, and relaxation noise is reduced, compared to the situation where no external EM field is present. The net result is a reduction of relaxation or linger noise resulting in clarifying of audio and/or visual signals from the conductor.

When the external modulated EM field of the present invention crosses the relaxation excitation E-fields of the conductor, longitudinal components on the electrons, in either direction, are produced which contribute to the overall damping. The fundamental principle is that, as the electrons increase their kinetic energy during stage 1 and stage 2 of the relaxation time, that increase is being resisted by a magnetic braking/damping effect, and the axial ejection and noise components that continually develop are damped by the introduced EM field.

This results in a reduction in the production of both immediate and linger noise. It is the mirror symmetry of the "magnetic brake" effect. The excited electrons try to move faster, in effect creating "eddy currents" in the conductor. Moving in the fixed magnet's modulated EM field, these "eddy currents" exert a magnetic repulsion force upon the magnetic pole, which in turn exerts a braking effect back upon the "eddy currents" themselves. This latter effect constitutes noise reduction in the conductor.

The end total result of the use of the device of the present device is:

1. Reduction of the velocity and kinetic energy of the electrons during coupled phase 1 and phase 2;

2. introduction of substantial additional damping upon the harmonic damped-oscillation relaxation noise that results in phases 1, 2 and 3,; and

3. substantial reduction of the overall relaxation noise in the system, in the resulting conducting medium, and in the resulting playback from the conducting medium.

The device directly reduces the relaxation noise in all conducting media which involve EM effects in recording and playback, and in transmission of signals. In any operating system, whether recording or playback as such is being performed or not, the present device directly reduces the continuously-forming "lingering relaxation noise" from signals that have previously occurred, and that are actually slower-moving charge disturbances constituting relaxation noise in the system's operation. The relaxation noise is also a power loss from the signal. Reduction of the signal power being lost in the system constitutes effective enhancement of the noise-free power transmitted by the system.

Application of the device increases the signal-to-noise ratio by dramatically reducing the lingering relaxation noise that is continually produced in: 1) systems operation; 2) the conducting medium; and 3) detection, amplification, and playback from the conducting medium. Media conductors or conducting plates that the apparatus may be applied to are audio compact discs, CD-ROM discs, photographic compact discs, video laser discs, motion picture and still-camera film of all types, x-ray film, optical media, and other medium which involve EM effects in recording and playback and in the transmission of signals through normal circuitry and conducting plates. Video, audio and digital tapes also benefit from the present invention but conducting tape must be treated with the device before recorded upon.

The systems may be any EM system such as electrical, electro-optical, magnetic, and magneto-optical systems. The continuously-formed lingering relaxation noise is a direct part of the so-called "standard noise" in the system.

When the device is applied to a conductor, such as an audio compact disc (CD) that normally produces playback distortion as a result of relaxation noise, the distortion is reduced and there is an improvement in sound quality with improved clarity and sharpness of the music or other such sounds. When the apparatus is applied to a medium that has a visual expression, such as a CD-ROM, Photo-CD, video laser disc or photogaphic film, there is improved visual effects with greater clarity, sharpness and color to the resultant pictures and photographs.

One embodiment of the invention is a device comprised of a pair of permanent magnets in an orientation in which their like poles, i.e., north or south, are in a facing and repulsing relationship. A coil of wire is then wound around the pair of fixed magnets in an orientation orthogonal to both the interface between the facing magnets and their north and south poles.

A motor means and electric supply means is provided for creating a pulsed current characterized by sharp voltage spikes to the coil of wire creating a modulated EM field while also producing a means for spinning the conducting medium when placed on a spindle that is driven by the motor means. A far greater effect is given by the brushes of the motor means when DC current is applied at a constant rate. At the same time, the motor means provides a more constant means of maintaining optimum rotation of the spindle thus exposing the conductor to the constant effect of the EM field.

The molded spindle allows for improved centering and stability during rotation of the conducting medium. The conducting medium is oriented to the device so that its plane is parallel to a plane running through the north pole orientation of the magnets. As the conducting medium passes through the modulated EM field, the electrons of the conductor are effected by the field, thus reducing the relaxation noise.

The device is housed in a container, such as a molded plastic housing, to maintain magnetic unit placement and to allow for greater penetration of the modulate magnetic field by the magnetic unit and with greater effect on the conductor.

One application of the device of the present invention is in clarifying sounds emitting from an audio compact disc. Compact discs are comprised of an aluminum surface encased in a plastic coating. The disc itself is composed of a very thin piece of stamped aluminum, which is the conducting medium, encased in plastic. The thin aluminum layer is the active ingredient that contains the recorded data.

A second application is in the clarification of compact discs such as video laser discs, photo-CDs and CD-ROMs. Processing of visual compact discs produces a clearer, sharper, brighter and more colorful image both on the computer screen and in a photo image.

A third application is in the clarifying of photographic films of all types. Exposed film that has been clarified before developing the film produces a consistantly clearer image than film that is unclarified.

Referring first to FIG. 1, 10 is a device comprised of a magnetic unit 12 which is further comprised of a pair of permanent magnets 14 and 16 orientated with like poles in a facing, contacting and repelling relationship and fused with an adhesive means 18 forming a seam with joined magnets 14 and 16 producing a shaped magnetic field 22. Magnets 14 and 16 are oriented with their north poles in an abutting relationship, although the magnets may be abutted using their south poles.

FIG. 1 further shows a coil of wire 20 wound around magnets 14 and 16 of the magnetic unit's out facing poles with coil 20 having an orientation perpendicular to the plane of the interface between magnets 14 and 16. Coil 20 may consist of 150 to 300 turns of a #34 enamel-insulated high temperature wire, with 250 turns preferable. Magnetic field 22 is shown emitting above adhesive seam 18 of magnetic unit 12.

Also shown in FIG. 1 is circuitry 3 for providing an electrical supply to coil 20 of magnetic unit 12 used in apparatus 10. Circuitry 3 is comprised of a DC motor 5, a controlling switch 7, either a 9 volt DC battery or a DC power supply 9 and connecting wire 11 to coil 20. DC motor 5 also provides a means for rotating spindle 33 seen in FIG. 2. In the alternative, electrical power may be provided by a 120 volt AC power supply 13 which is reduced to 9 volts by passing through a 9 volt AC to DC convertor or power step-down unit 15.

FIG. 2 is a perspective view of a housing 31 which contains apparatus 10 and its components. Seen emerging from the top of housing 31 is spindle 33 for the placement and orientation of conducting medium 24. Spindle 33 emerges from the top side of housing 31 for ease of placement of the conducting medium. Closure of switch 7 creates a modulated EM field 22 simultaneous to the spinning of conducting medium 24. As the conducting medium spins, it passes through field 22 thus

producing the desired effect on the conductor.

FIG. 3 is an inside view of the bottom half of housing 31 showing compartments, A-1 and B-1 separated by a divider 35. Compartment A-1 is adapted to fit over compartment A-2 of the top half of housing 31, seen in FIG. 5. Compartment B-1 contains a 9 volt DC battery or DC power supply 9, space for switch 7 and space for motor 5.

FIG. 4 is a perspective view showing the orientation of conducting medium 24 to spindle 33 and motor 5.

FIG. 5 is an inside view of the top half of housing 31 showing compartments A-2 and B-2 separated by divider 37. Compartment A-2 contains magnetic unit 12 with connecting wire 11 from coil 20. Wire 11 passes through a recess in divider 37 and into compartment B-2 to connect to motor 5 and switch 7. Compartment B-2 contains motor 5 and a side wall of B-2 contains a recess for switch 7. Beneath motor 5 and on the roof of the top half of housing 31 is an opening for the emergence of spindle 33.

FIG. 6 is a side view of housing 31 showing the relationship of the various elements to each other. Magnetic unit 12 is located in compartment A-2 of housing 31. Connecting wires 11 connect coil 20 to a 9 volt battery or AC to DC convertor 9 and motor 5. Spindle 33 is connected to motor 5 and emerges through the top of housing 31 to accept conducting medium 24. Conducting medium 24 is shown in close proximity to magnetic unit 12 and oriented in a plane perpendicular to the plane of intersection of magnets 14 and 16 so that when apparatus 10 is activated, the rotating conductor 24 will pass through the modulated EM field 22.

EXAMPLE I

An analysis of an audio compact disc was carried out with the test completed using a Schlumberger 1510 Audio Spectrum Analyzer to display real-time fast fourier transform (FFT) spectra of the sound, a Tektronix 465B dual-trace oscilloscope to display phase differences between left and right stereo channels as well as waveform shapes, and a Leader LMV-185A 2 channel AC Millivoltmeter to simultaneously read the amplitude of the left and right stereo channels. The spectrum analyzer has a frequency range of up to 25 khz, a dynamic range of 70 db, and an absolute range of 100 db using attenuators, and an accuracy of 0.1 db and 1/256 of the frequency range selected. All measurements were made using the 25 khz range. The oscilloscope has an upper frequency range of 100 mhz. The test CD was played on the Denon DCD-1500 II compact disk player connected directly to the above instrumentation.

The test CD, CDP 7 46446 2, DIDX 2249 Stereo, UK: CD-PCS 7078, manufactured by Capitol Records, otherwise known as the Beatles: Abbey Road album originally recorded in 1969 and digitally re-mastered in 1987 for CD, was subjected to the above equipment and technique. Spectral measurements at a 00:24 interval was made on track #7, a selection known as "Here Comes the Sun".

Track #7, at a 24 second interval, is exhibited in Example I as a series of bar graphs in diagrams 1-8. The bar graphs simultaneously show the relative loudness in decibels "before" and "after" clarifying of the device. Each pair of bars represents the relative loudness of frequencies between 100 hertz and 25600 hertz of Track #7 time 24 seconds of the audio CD analyzed.

Diagram 1 shows the entire spectrum from 100 to 25600 hertz in 8 frequency bands as labeled at the bottom of the diagram. It is noted that each frequency range except for 200 to 400 hertz is about 1 or 2 decibels louder after processing by the device.

Diagrams 2-8 display the same data in more detail showing a larger number of smaller frequency ranges in intervals of only 100 hertz each. Because frequency ranges are smaller there are more variations in the data, however, the readings show the trend of being 1 to 2 decibels louder after processing.

It is noted that each of the spectral differences as well as the average spectrum difference follow a profile demonstrates the effect that clarifying by the instant device has on the CD.

EXAMPLE II

An analysis of an audio compact disc was carried out by first recording "The Best of Mike Olfield--Elements", track "Portsmouth" without clarifying the CD, on a hard drive using a WAV file of Voyetra's WINDAT software with a sound card connected to a CD-ROM's audio output. The recording was in mono at a sample rate of 44100.

Following the first recording, the CD was then clarified, using the device, by placing the CD in the modulated electormagnetic field of the present device for 10 seconds on each side followed by again recording the CD using the above technique.

Using the WINDAT program, the "before" clarifying and "after" clarifying recordings were trimmed so that application of the device could be easily compared with non-application. Ten second and 0.25 second excerpts were used. The ten second excerpts were measured from the point where sound began to come off of the CD. The 0.25 second excerpts were taken from a consistent point within the body of the recording.

The sound files were converted into graphics using two Shareware programs: Screen Thief and Blaster Master. The recordings were loaded into Blaster Master individually and the graphical representation of each recording was captured using Screen Thief. The graphics of the sound files were then laid down and compared using Deluxe Paint II Enhanced.

The graphical representation of the sounds, in Diagrams 9-11, is based on amplitude and time. The left side of the graphic is the beginning of the recording, and right side is the end. The beginning of the graphic is defined as the point where sound began showing up on the graphic.

Diagram 9 illustrate a 10 second comparison of the unclarified with the clarified CD. The waveform represents amplitude. It is noted that waveform A, the unclarified sample, is taller than waveform B, the clarified sample.

Diagram 10 is of the same recording but compares exclusively unclarified, partial waveform C, unique to waveform A, and clarified partial waveform D, unique to waveform B. That is, all of the points in the unclarified graph, waveform A, that correspond to points in the clarified graph, waveform B, have been erased resulting in waveform C, and all of the points in the clarified graph, waveform B, corresponding to the unclarified graph, waveform A, have been erased, resulting in waveform D. This technique leaves a graph showing the data that is unique to the recording in question. Thus, in Diagram 10, the graph labelled "Exclusively Clarified", the visible lines show data that is unique to the clarified recording, and the "Exclusively Unclarified" graph shows sound or noise that was no longer present after clarification.

Diagram 11 illustrates a 0.25 second comparison of the unclarified CD, waveform E, with the clarified CD, waveform G. Waveform F is the exclusively clarified and shows what is unique to waveform G. A graphic of an exclusively unclarified waveform was eliminated because interest was in the clarified waveform and what is unique to it.

Three criteria were used in analyzing the graphs of Diagrams 9-11: Response Time, Noise level and Density. Response Time refers to "when the sound starts", a key factor because the CD excerpt recorded was timed from the moment that sound began recording off of the CD. The position of the remaining material recorded is dependent on the accuracy of the starting point. In the ten second excerpt of Diagrams 9 and 10, there were no measurable time discrepancies, as the graphs matched up as represented in the Diagrams with the peaks in the sound occurring in the same places throughout the sample.

The 0.25 second excerpts showed a lag time between unclarified waveform E with clarified waveform G. As many of the peaks of waveforms E and G as possible were matched-up and then compared to how much "extra" sound was left at the beginning. By comparing the length of the sound to the length of the whole sample, it was determined that the unclarified 0.25 second excerpt was lagging behind the clarified excerpt by about 0.01 second. A small discrepancy at the starting point would not be visible on the ten second scale, but on the 0.25 second scale is noticeable.

In terms of Noise Level, the unclarified sample was "noisier" in both excerpts, ten second and 0.25. By noise level is meant, that overall, the unclarified sample consistently had several peaks where the volume line went further than the corresponding volume line in the clarified sample. This lower level in the clarified sample is the result of a reduction of "noise" level in the clarified sample. Extraneous noise makes the overall signal noisier and less clear. The clarified sample has an improved and smoothed out dynamic because of a reduction in noise level.

Density is defined as the solidity of the volume or amplitude lines rather than the range of the lines. Comparison of the volume lines of waveform A, unclarified, with waveform B, clarified, shows waveform B to have a much more consistent volume level than waveform A. Increased Density gives greater consistency to the CD, producing a much fuller sound.

EXAMPLE III

In Example III, the identical technique used in Example II was duplicated. An analysis of the audio compact disc, "The Cross of Changes" by Enigma, track "Return to Innocence" was carried out by first recording, without clarifying the CD, on a hard drive using a WAV file of Voyetra's WINDAT software with a sound card connected to a CD-ROM's audio output. The recording was in mono at a sample rate of 44100. Following the first recording, the compact disc was then clarified, using the device, by allowing the CD to spin for 10 seconds on each side followed by again recording the CD using the above technique. The analysis technique as described in Example II was then completed.

Diagram 12 illustrates a ten second excerpt comparison of waveform A, unclarified, with waveform D, clarified. Waveform B is exclusively unclarified, and unique to waveform A. Waveform C is exclusively clarified and is what is unique to waveform D of the sound track. Waveform A, the unclarified sample, is significantly different than waveform D, the clarified sample.

Diagram 13 illustrates a 0.25 second comparison of the same recording of unclarified waveform E with clarified waveform G of the CD. It also shows exclusively clarified waveform F, demonstrating the unique difference between waveforms E and G.

Response Time, Loudness and Density were analyzed. In the 0.25 graphic of Diagram 13, the Response Time of waveform G is slightly ahead of waveform E, indicating that sound from the recording began sooner than in waveform E. The amplitude lines of waveform G are slightly reduced overall in comparison with waveform E, resulting in a reduction of Noise Level. The Density of waveform G is much greater than in waveform E, especially in the last one-third segment of Diagram 13.

EXAMPLE IV

A Photo-CD was analyzed both "before" and "after" passing the CD through the modulated eletromagnetic field of the device to determine specific changes in the retrievable data from the CD. The method and steps used to derive and quantify the effect that clarifying had on the conductor was determined using the following method and steps:

1. A demonstration Photo-CD sampler, produced by Kodak, was down loaded to the hard drive of an Apple Centris 650, 25 HRZ, 6800040 motherboard and a new file created and properly designated as "before". The photo-CD had not been clarified by the device. Images were transferred to a commercial graphics software program, PhotoShop, developed by Adobe Software.

2. The Photo-CD was removed from the ROM and treated with the present device by spinning the CD on each side for 15 seconds. The CD was then down loaded to the hard drive and a second new file created and properly designated as "after". The images created were transferred to the PhotoShop graphics software program with file separation maintained at all times.

3. Print number 3 of the Photo-CD list, contained in the "before" or unclarifyied, was brought to the screen and a text translator was used to convert the picture to post script computer language. Picture number 3 represented a photo image that measured 21/2.times.13/4 inches, using 132 LPI, 72 DPI and utilized approximately 250K of memory. Following the text translation process to convert the small image to post script, it was then saved to a new file. Following the conversion, there were 360 pages of post script computer language created to describe the image number 3.

4. The identical process as described above was used to create a post script file for the image following use of the device on the photo-CD. The "after" or clarifying effect resulted in 348 pages of post script computer language.

5. The first 16 pages of post script language was specific to language to program the printer. The balance of the pages in both cases was post script language to describe the specific image that was converted. The first 4 pages of both files were then converted into text files using Wordperfect 6.0. The process of converting the post script files to text files was for the specific purpose of utilizing a specialized software program that could compare the text language of both files. The process of converting the first 4 pages of both post script files to text files took 25 minutes and resulted in the generation of 101 pages of new text language. The ratio of pages of text converted from post script is approximately 1 to 25. If the entire post script files (354 pages average) were converted to text file using the same ratio, it would require approximately 8,850 pages of text file to fully describe a single 21/2.times.13/4 inch colored image.

6. The software used to simultaneously compare the text in the two files, DocuComp II, version 1.05, is a product of Advance software, Inc.

7. The first 4 pages produced a total of 101 pages of text file. When placed into the DocuComp file, it resulted in the following summary of information changes: The 4 pages represent slightly more than 1% of the total data available from each image. If the entire 708 pages were converted and analyzed, it would take considerable memory capacity and about 37 hours (4 pages of each file takes 25 minutes) of computer running time to analyze 708 pages of post script computer language.

8. Four pages from each file generated a total of 56 material changes, on the average, resulting in over 5,000 changes for the entire image when "before" clarifying was compared to "after" clarifying.

9. The "before" unclarified file, resulted in 53,336 bytes of information The "after" clarified file version, generated 50,521 bytes of information. The comparison represents a difference in information retrieved equal to 5.7% less for the clarified file.

Clarifying the Photo-CD with the device realigns the data into a tighter format by reducing superfluous data. Each Pixel has a "halo" or influence zone. Exposing the pixel to a modulated EM field reduces the halo by suppressing electron activity or relaxation noise thus causing the pixel to be sharper with improved clarity. The photo-CD manufacturing process can produce residual color around each individual pixel referred to as pixel residual or bleeding. When a pixel, represented by post script and text script, is sharper, it does not require as much information to accurately describe or define it. When a pixel halo or influence zone is reduced, it requires less information to describe the data. In Example IV, 5.7% less information was required following clarifying by the device, at the same time improving the image to describe the same segment of the picture.

The analysis process of converting each image to post script and then to text is accurate. Changes created by treating a photo-CD with the present device can be demonstrated by a computer comparison of the information generated from both images.

Diagram 14 is a Summary of Post Script Changes with a list of differences between the "before" and "after" application of the device to the photo-CD.

EXAMPLE V

Using the identical technique and steps as described in Example IV, steps 1-9, a second Photo-CD, Nature's Way, published by Gazelle Technologies and produced by Kodak, was tested "before" and "after" application of the present device. In this Example, the original bytes numbered 74,151 before clarifying with the device. After clarifying, the number of bytes were reduced to 70,214 resulting in a 5.6% byte reduction.

Diagram 15 is a Summary of Post Script Changes with a list of differences between the "before" and "after" application of the device to the photo-CD.

Diagrams 16 and 17 are black and white, "before" and "after", copies of color pictures of the photo image of the photo-CD, demonstrating the effect that clarifying the conducting medium has on the comparable results.

John BEDINI's Patents

Device and method for utilizing a monopole motor to create back EMF to charge batteries
US6545444

Analog vector processor and method for producing a binaural signal
US5748745

Circuits and related methods for charging a battery
US7990110

Monaural to binaural audio processor
US4555795

Anti-copy system
US4644422

Device and method of a back EMF permanent electromagnetic motor generator
US6392370

Apparatus and method for reducing electronic relaxation noise present information recording medium
US5487057

MONAURAL TO BINAURAL AUDIO PROCESSOR
WO8400661

Device and method for pulse charging a battery and for driving other devices with a pulse
US6677730

ANTI-COPY SYSTEM
AU1827083

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

THE LAW OF THE SQUARES.

LUNIC ENTERPRISES LTD
BRITISH SECTION
(MANNED FLIGHT DIVISION)
SWALLOW COMMAND H.Q.
FLIGHT CONTROL CENTRE
MORTIMER
BERKSHIRE
ENGLAND
SWA
MK II

THE IMPOSSIBLE MADE POSSIBLE BY FOOLS LIKE SEARL

TOMORROW'S TRANSPORT AND ENERGY SYSTEMS.

SEARL INTERNATIONAL SPACE RESEARCH CONSORTIUM.

BOOK 4B.

SERIAL NO:000001

THE TEAM PHOTO 1968.

INTRODUCTION TO THE AUTHOR, INVENTOR.

INTRODUCTION TO JOHN THOMAS

PHOTO OF THE MRS & MR SHERWOOD WITH PROF. SEARL

PHOTO OF PROF. SEARL WITH MAG DEMO.

PHOTO OF PROF. SEARL WITH MAG. DEMO.

INTRODUCTION 1989.
    PHOTO OF PAPER MODEL CUTOUT

INTRODUCTION 1989.
    PHOTO. LUIS JARILLO WITH PROF. SEARL

INTODUCTION 1989.
    PHOTO SUSAN

INTRODUCTION 1989.
    PHOTO DOMONIC

INTRODUCTION 1989.
    PHOTOS RITA AND JOY

COMPONENTS FOR EVALUATION 1993.

BRAINTEASERS ANSWERS TO BOOK 4A.

IMPOSSIBLE MADE POSSIBLE 1959

IMPOSSIBLE MADE POSSIBLE 1968.
    PHOTO DEMO 1 UNDER CONSTRUCTION

MY MEDICAL KNOWLEDGE 1993.

REALITY OR FANTASY 1978.
    PICTURE OF ONE OF SEARL'S TRAIN UNITS

SEARL'S RAIL CONTROL SYSTEM 1968.

MY KNOWLEDGE 1946

MY KNOWLEDGE 1946
    Take a break 1947.

COMPONENTS FOR EVALUATION 1993.

REALITY OR FANTASY CHOCOLATE 1956.

REALITY OR FANTASY CHOCOLATE 1956
Do not hand me that shit!

REALITY OR FANTASY 1995.
PHOTO JOHN THOMAS WITH PROF. SEARL

REALITY OR FANTASY 1957.
PHOTO NATURE STUDY
THE HOMO SAPIENS STRUCTURE

REALITY OR FANASY 1957.
PHOTO FEMALE BUM.

REALITY OR FANTASY 1957.
PHOTO HOMO SAPIENS POSTERIOR SKELETON

REALITY OR FANTASY 1957.
PHOTO HOMO SAPIENS ANTERIOR SKELELETON

FACTS 1993.

FACTS 1946.

FACTS.
Fact 27th February 1874 coal cutting by machinery

FACTS.
Fact 12th September 1941 Technical ability wasted

FACTS.
Fact 24th January 1936 Kipling, The Engineer’s Poet

FACTS.
Brainteaser No. 1.

FACTS.
11th November 1993
11th November 1993

FACTS.
October 1993
October 1993

FACTS.
Brainteaser No. 2.

DREAM 1 STATES.

PERIODIC TABLE OF THE ELEMENTS 1968.

PERIODIC PROPERTIES OF THE ELEMENTS 1968.

DREAM 1 STATES.

THE SEARL EFFECT.
A German article.

HOMO SAPIENS ATITUDES 1963.

SEARL UNDERSTANDS FLIGHT 1963.
Skin friction

NEWSLETTER NO. 27. DATE 22.06.1987

STARSHIP EZEKIEL MK.V. AIR LAW 1968. .

GENERAL STATEMENTS 1993.

QUESTION TIME 1946.
15 Questions to answer

FINAL REPORT 1956.

I, Prof. John Roy Robert Searl hereby release this book 4B to the public as unclassified matter, upon this date of 25th December 1993.

PROF. JOHN ROY ROBERT SEARL

PRESIDENT
CONSULTANT ENGINEER.

SEARL INTERNATIONAL SPACE RESEARCH CONSORTIUM
DATE RELEASED DECEMBER 25TH, 1993.

Tomorrow’s Energy & Transportation Systems.

THE LAW OF THE SQUARES

This introduction, which shows that work out of star ship Ezekiel MK V, took three people to hold it out to film it. At this date, John Thomas in Rochester New York holds it at his home.

JOHN A. THOMAS Jr.

Vice President, D.I.S.C.

1 Mr. Thomas has a technical background in electricity and process control.

2 He has studied such varied topics as UFO's and Ancient Technology that has been lost or forgotten.

3 He has carried on the efforts of Mr. William Sherwood, formerly Professor Searl's contact and helpmate in the ***National Space Research Consortium. (NSRC).***

4 He first contacted Professor Searl in 1990 and since that, time has printed and distributed his books on "***The Law of the Squares***."

5 John has been a student of Professor Searl and has come to an understanding of "***the Square Technology***" which is the basis for the ***Searl Effect Generator (SEG)*** and the ***Inverse-G-Vehicle (IGV)***

6 John is presently Vice President of the ***Direct International Science Consortium*** (***D.I.S.C***.) and is Technical Representative for Professor Searl.

7 Prof. Searl promised John that he could ride in a small hut pulled behind the ***IGV.***

1 This is the man who owns the technology termed the **SET**.

2 The technology sits in his brain; therefore it cannot be own by anyone else.

3 What he states is that he is prepared to form a group to develop this technology.

4 That statement means that his dream at this time is in the world of fantasy

5 But that dream can and has been defined by structure and function

6 Therefore, it is legal for him to state that he has a concept for research and development, and no other person living or dead has the right to that statement.

7 It is known that others make belief that they own it, how can them when they have no idea how to do it.

7 The concept herein: shall be defined as being a ring and a roller set.

8 The purpose of that set is to demonstrate proof of a phenomenon herein termed the '***Searl Effect'***

9 Proof of the '***Searl Effect'*** being accepted when the roller operates non-stop for 3 days or greater.

10 The term '***operate***' refers to that a roller; without any form of assistance commences to rotate around a ring, and continues to do so; without any form of assistance from any other source other then itself.

11 This definition defines the concept for research and development.

1 Up to this moment in time in which I had been living at No. 13 Blackburn the Following people were involved.

Luis Jarillo found John Searl and the accommodation at No. 13 Blackburn.

2 That paper work shown on the first page is the scale model of ***Starship Ezekiel MK.V***.

3 Three people against wall just yards are holding it from ***No. 13 Blackburn*** protects against the wind.

4 I myself work out the structure than cut it out to check that my figures were correct.

5 That cut out is held in John Thomas hands in Rochester. USA.

6 These are those who were encouraging me to return to this work

Miss Susanah N. Mutua.

7 First picture is Luis who found me as a lonely down hearted person who had been destroyed by his own family.

8 He found the home in which that picture was taken

9 The next picture of Susan is that person who gave me her home to live and work so that all humankind could benefit in the end.

10 This next picture is Dominic Mutua son of Susan who helped me from time to time with the heavy lifting work.

11 That is basic the main story, from the time my first wife destroyed my life and work.

12 Susan became my whole life; I work many hours to recover this work under terrible pain conditions.

13 All this based upon faith that this time I would succeed to the market place and Susan would become my wife.

14 Faith that here at number 13 Blackburn would be my home, and never again would I be lonely and homeless.

15 Only time will tell the truth what happen to this work.

16 This move to number 13 Blackburn cost me thousands of pound loss.

17 Equipment again was needed to be replaced so I could undertake this work.

18 Susan went out of her way to help me upon that issue, where no one else made the offer.

19 Latter another big problem occurred which would have ended this work, but John Thomas did rush in to save the day.

20 Teams form and then disappear.

21 WHY?

22 The answer is quite simple. A person informs us that they will have millions in a month's time and wish to invest a million or more to develop this engine.

23 The group start to work hard to get ready for full speed ahead as soon as the funds arrive.

24 Alas, as always, no funds came, and it is only natural that the hard work put into this without payment cannot continue.

25 Therefore, the team slowly breaks up, as it had no purpose to continue if no product could be produced.

26 This class of work clicks into millions of £'s or $'s to develop it to a point in which it could be mass-produced.

27 Without that funding there is almost nothing I can do, except to write my life story, which is the Law of the Squares.

28 That is also greatly expensive to do.

29 Here are two more ladies who helped me in different ways.

30 The first was Rita a lifetime friend of Susan; she helped me with the book called the dream that came true.

31 The young lady named Joy came into my life right from the moment of her birth, at this age she was communicating with me as a full-grown adult.

32 She amazed me with her questions, which clearly showed me that she was reasoning every thing, which she was being taught at school.

33 She would give me her feelings on the subject and then ask me what I think about it.

34 Her reasoning was sound, her judgement was perfect, in reference to adult women and men who come here to talk, and she was far more intelligent. I enjoyed listening to her reasoning and judgement.

35 By the time this photo was taken, she was capable of correctly and quickly works out the answers to extremely large sums, which her mother Phyllis always complained to me that I should not give her such large sums to do.

36 However, she loved them, because she loved me so much that she put her heart and soul behind her learning with me.

37 This little woman gave me the driving force, not only to make Susan’s home lovely but also to speed up the work on this technology, and to record what I did and what was happening around me at that time.

38 In return, I gave her love and the power to learn, because she knew I would tell her the truth to all her questions.

39 Not only that, but I would listen to all her problems without getting nasty about what she had said.

40 It was a bond which is rarely seen between an adult and a child, a bond so great that it became her driving force to learn

41 If this bond is ever broken, the power to learn will also be lost for good.

42 If a child loses a bond it will result in the lost of ability to learn.

43 These books, which I hope I shall live to write will present what has happened to me and how I have responded to each situation involved.

44 However, when I am dead that may be all which is left to show the effort which I went to for support to create a new technology which would solve all of our present day problems.

45 Yet 80 % of the world’s population has no idea that there are any problems, because they fully believe without question that if there were any problems their government would solve that problem.

46 Though most people may not believe that there is any problem to worry about, but the very few who do know, many of which go to the extreme state and start to present insane activity, in place of logic thinking.

47 Like for example, Aliens will be arriving in the year 2000 to collect all you UFO nut cases and take you to a place of safety.

48 Maybe the backside of the Moon or better still on top of a mountain on Mars.

49 They are in communications with Aliens, but do not show you their transmitter or receiver so you can order one from your local shop.

50 Others are constructing large airships, arming members, and teaching them how to defend themselves and kill. This is the crap, which this planet is loaded with sadly to say but the insanity level is increasing at a much faster rate than ever before from where I sitting.

COMPONENTS FOR EVALUATION 1993.

1 ***OCTOBER 4th, 1993***, work has re-started upon my technology, now is the time to evaluate the latest components for selecting the best option for the various functions, which will be required within this technology.

2 I know there is just one person who states that people claim that I write bullshit in my books to blind readers with technology.

3 I know these people buy these UFO mags and videos and believe absolutely what they state.

4 To those people, the truth can never be accepted.

5 I do not intend to try to convince them, in **FACT,** the **TRUTH** is, and I have no intentions whatsoever of trying to convince anyone as far as it goes.

6 What I am offering to the world is a **CONCEPT FOR DEVELOPMENT** and the **FACTS**, which I can present in support of such a concept, what I actually know, and how to apply that knowledge to such a concept.

7 I can propose what components should be considered to meet each function part thereof, to arrive at the best option for success.

8 Therefore, everything stated within these books are more precise than that of the bible will ever be in **FACTS**.

9 Far more **TRUTHFUL** than the bible is.

10 It is a **BIBLE** of absolute **FACTS**, as I know them.

11 Only advancement in science and technology can actually update these books.

12 That is why there is no end to this story because it is forever extending its bonds in science and technology.

13 So let me introduce you to the latest problems, which must be solved for this technology, which has just re-started.

14 The hydraulic connection between an agricultural implement and a tractor as shown below is now an easier task with these Pioneer multi-couple.

15 Now such a concept could very well be applied to the slender disc project for the flight cell function.

16 In the above figure, that system connects up to seven hydraulic lines and six electrical lines in a single operation.

17 Well done!

18 Two heavy-duty die cast aluminium housings hold the quick couplings during connection.

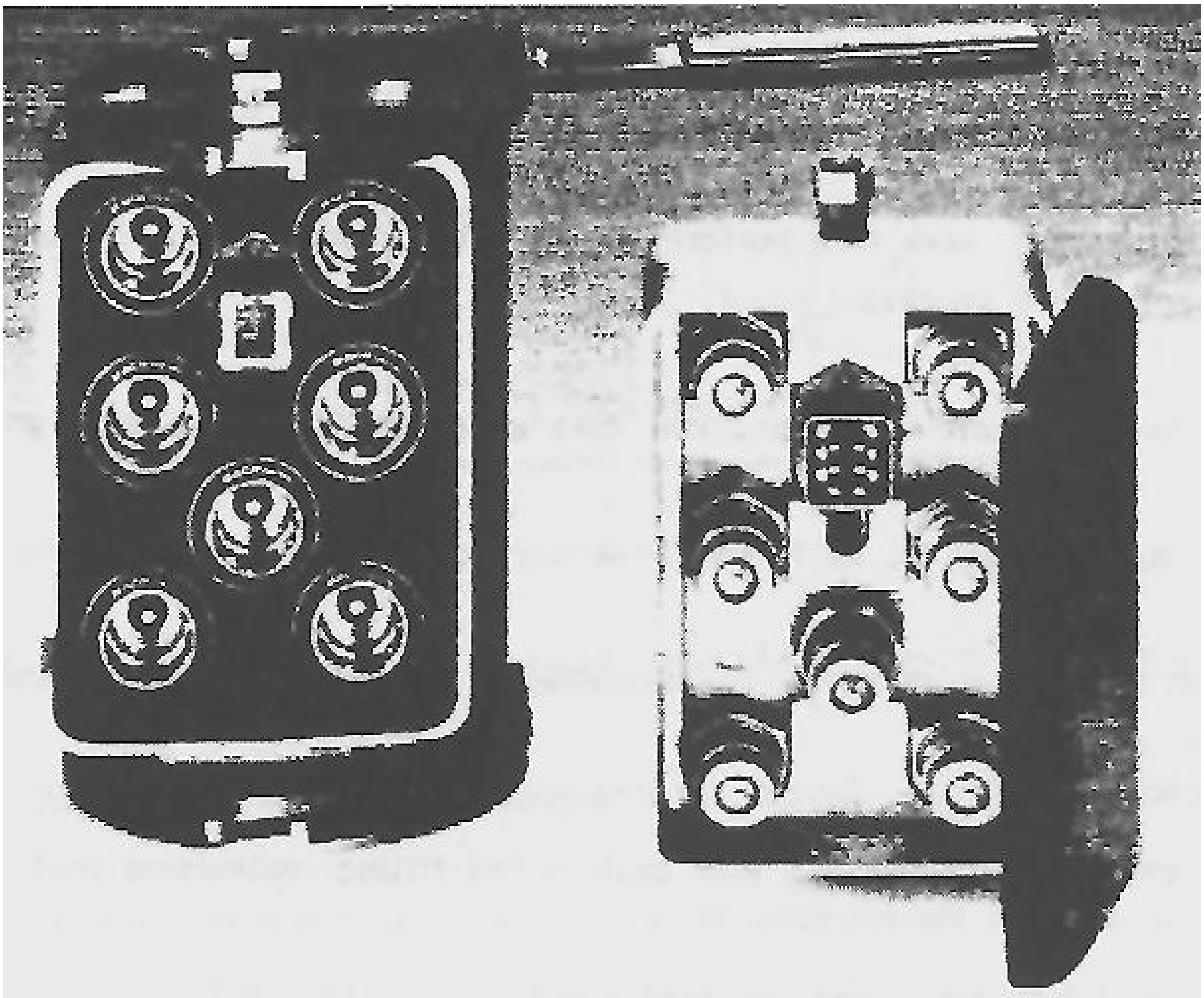

MULTI-COUPLER SYSTEM

19 In the above figure, the quick couplings used conform to ISO standard 7241 series A.

20 It allows simultaneous connection of all hydraulic and electrical lines and offers many advantages:

- ***(1) Service pressure up to 250 bar***
- ***(2) Compact, lightweight, multi-coupler with a strong construction***
- ***(3) Simple, effective and fast connection even under severe service conditions***
- ***(4) Housing and cover for total dust and moisture protection and long service life***
- ***(5) Safe connection, even under pressure***
- ***(6) Impossible to connect incorrectly***
- ***(7) Developed with the user for the user***
- ***(8) Compatible with any equipment already fitted with standard ISO A quick***

COMPONENTS FOR EVALUATION 1993.

Couplings

21 This multi-coupler system may also be used for materials handling and public works equipment, machine tools and other industrial equipment.

22 That is just one point to think about, remember that there are 64 flight cells, which must be able to function by more than just one option, there must be at least three different options of operation of these flight cells.

23 So think upon it!

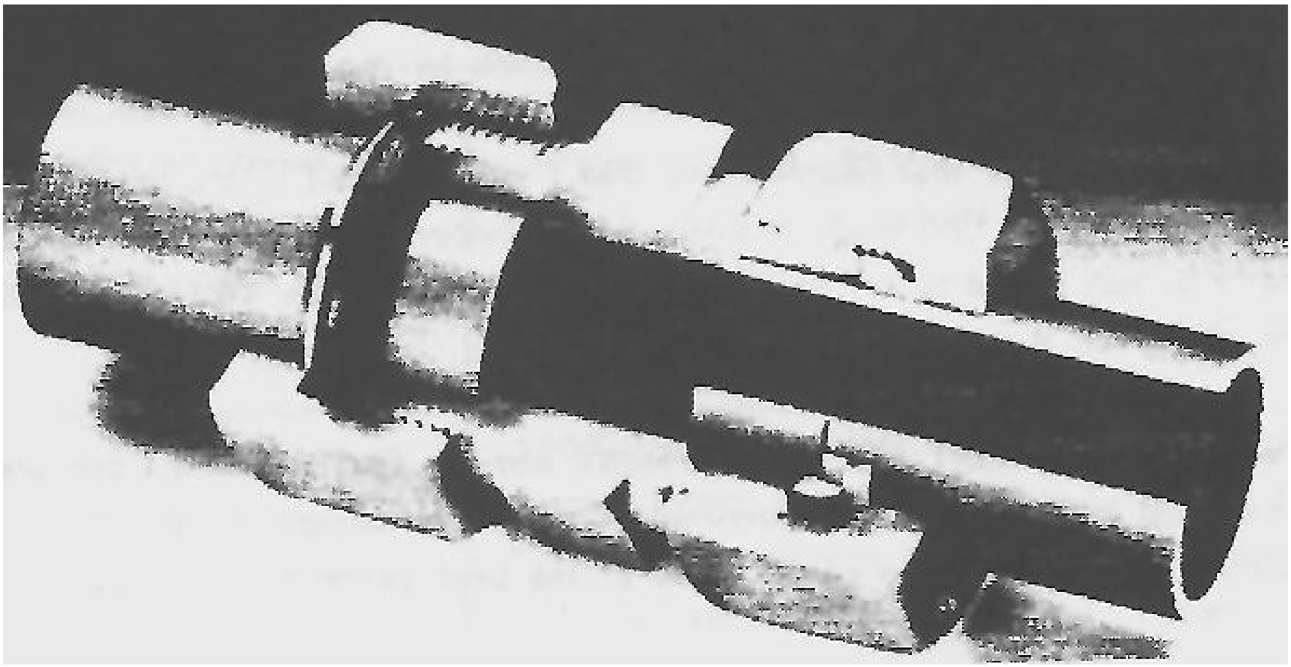

EO-2 DRY TECHNOLOGY.

24 EO-2 is a new generation of leak free fittings, and forms part of a “Searl’s ***Dry Technology***” programme for the ***INVERSE-G-VEHICLE***, for a safe, clean, leak free sealing even under the most severe conditions.

25 Conforming to DIN 2353 standard, the EO-2 bite type fitting has been completely re-engineered to feature vastly simplified assembly characteristics.

26 EO-2 incorporates soft sealing for secure, leak free performance, the number of parts has been reduced for easier handling, and stocking, which will be very important point.

27 Simple one stage assembly to block, without lubrication, cuts assembly time, two very good points to consider.

28 Assembly by torque is possible, as is full assembly by electrohydraulic tooling.

29 The nitrile rubber bonded seal on the ferrule gives all the advantages of soft sealed joints, including tolerance of surface imperfections, dirt, misalignment, etc.

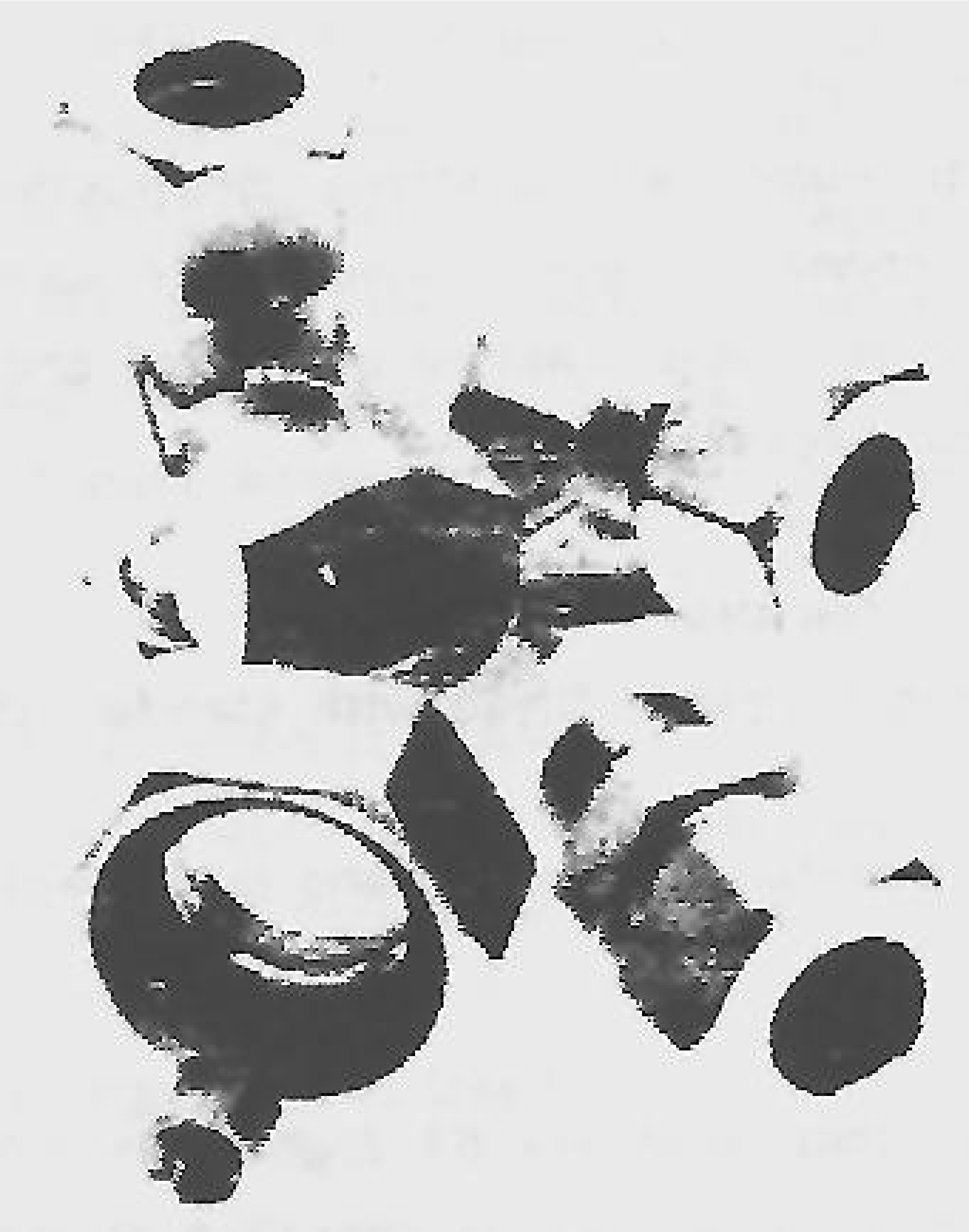

EO-2 Anew generation of fittings with soft sealing, allowing leaks to be prevented more effectively than ever before

30 The sealing lip blocks the only possible leak path in the fitting, optimising performance.

31 An integral bite ring gives the mechanical joint onto the tube.

32 To make the use of EO-2 even easier, the soft seal ferrule is factory fitted into the body nut to operate as a single unit, making it impossible to lose ferrules or assemble them the wrong way round.

33 Another unique advantage is that the nut stays where it is placed on the tube, instead of sliding to the lowest point or off the end, as no doubt that many of my readers can well appreciate the ruddy nuisance it is when you are trying to fit a pipe

In a hurry, to find that the nut is not where you want it.

34 I know what you say even if it is only mentally, when you are up 50 or more feet in the air, and that nut is lying on the floor.

35 You need not be ashamed of that fact, there are many more members of the same Club, you are not alone.

36

EO-2 can solve problems for applications where assembly difficulties have been experienced, boy they can state that again, or leaks have led to costly rework, cleaning up, or warranty charges.

37 FAULT-TOLERANT CMOS MULTIPLEXERS OFFER "BEST BUY" PROTECTION

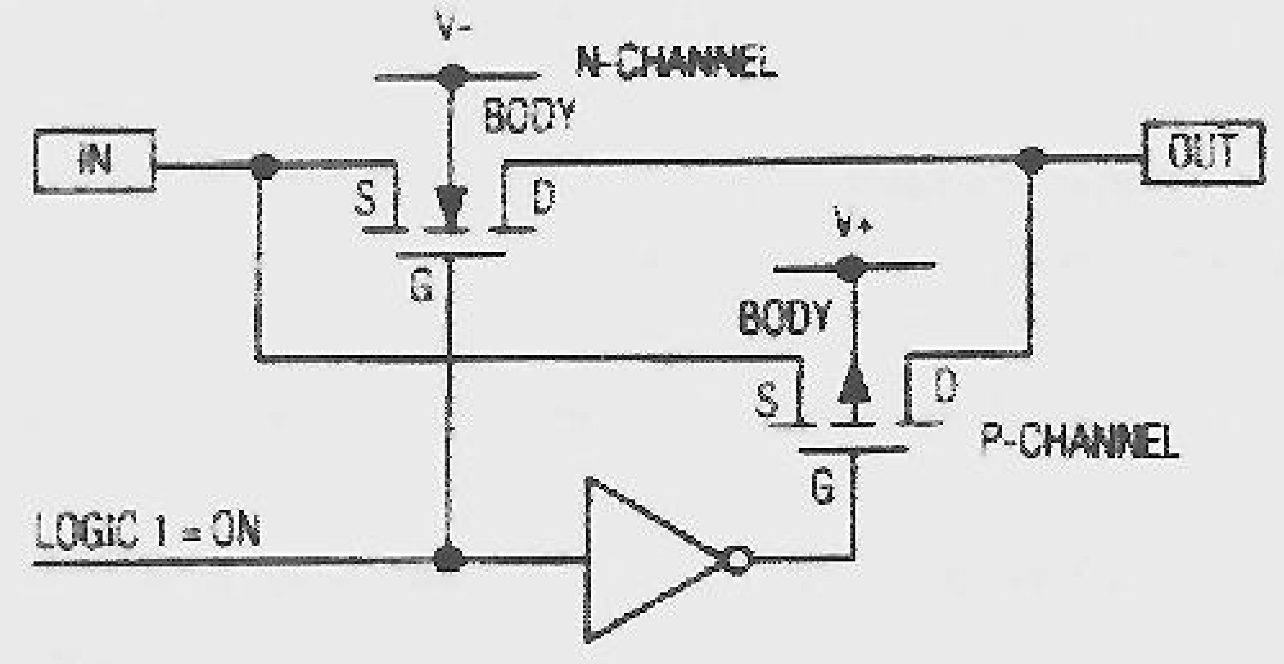

THE TRADITIONAL CMOS ANALOG SWITCH IS A TRANSMISSION GATE

38 All analog multiplexes act as the solid-state equivalents of digitally controlled rotary switches.

39 However, multiplexes at the interface between an electronic system and its environment have an additional function - they act as insurance policies against malfunction and damage.

40 As a designer, I may control the system, but input multiplexes contend with everything outside the system - poor installation, careless operators, and electrical noise.

41 They must operate correctly in the presence of ground loops, electrical interference from motors and engines, and unintentional inputs such as 240 VAC.

42 In my following discussion, a particular design (the serial MOSFET switch) comes to mind, as one of the devices which emerges as the most economical choice in

overcoming these problems.

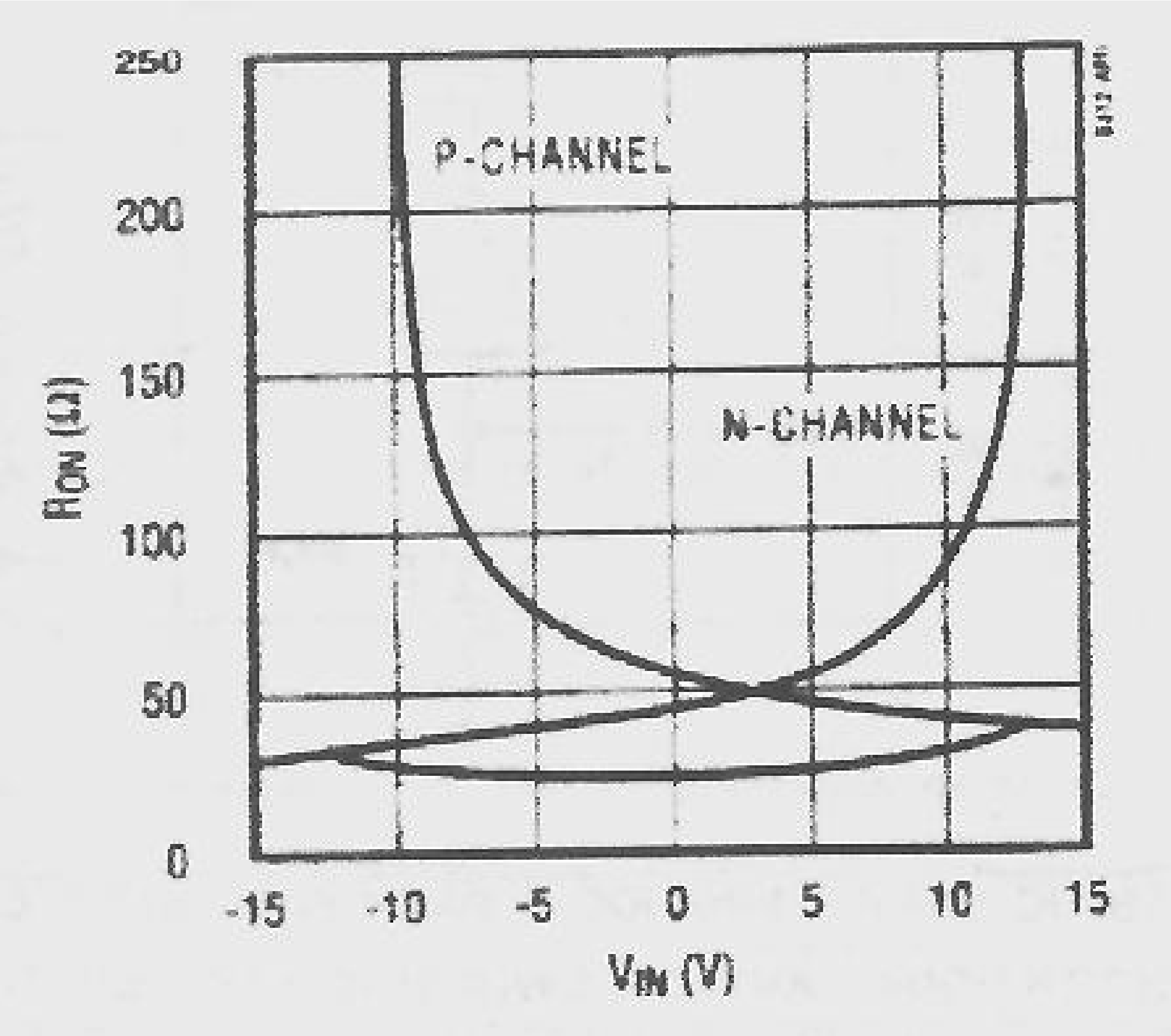

WHOSE ON-RESISTANCE VS. SIGNAL VOLTAGE CHARACTERISTIC EXHIBITS A DOUBLE HUMP AS SHOWN

43 Other designs offer over voltage protection alone, but only, at least to my knowledge, the serial-MOSFET approach combines overvoltage and fault tolerance without the need for external components.

44 The switches in common multiplexers have been designed the same way for nearly twenty years: each consists of an n-channel and p-channel MOSFET connected in parallel on a SILICON substrate, and driven with opposite-polarity gate-drive voltages as shown in my figure on page 11.

45 This connection provides a symmetrical signal path through the parallel source-to-drain resistances, producing a characteristic double hump in the curve of on-resistance vs. input voltage as shown in my second figure above.

46 Many designs minimise this effect by driving the body connection of the n-channel MOSFET with signal voltage.

47 The presence of each device polarity guarantees that at least one of the two MOSFETs will conduct for any input voltage between the supply rails.

48 Thus, the multiplexers can handle any signal level that falls between the rails.

49 A multiplexer switch ceases to be a switch, however, when signal voltage exceeds either supply rail.

50 Each switch includes two parasitic diodes, intrinsic to the MOSFET source and drain structures, which provide current paths to the rails, at this precise time.

51 I cannot put in the figure that I want, the drawing is far too large to scan in here, so hopefully I can find a way of reducing the size of it and add it at the back of the book, if I don't forget it.

52 Both diodes are reverse-biased during normal operation, but any signal excursion beyond the rails applies forward bias to one of the diodes, clamping the signal at 600 mV beyond the rail.

53 Because the diodes are present when power is removed, they also clamp (at ± 600 mV) when the rails are at zero volts.

54 Parasitic diodes provide a useful clamping function, but they also introduce problems.

55 Excessive current in the diodes can cause overheating and damage in the signal source as well as the multiplexer (figure 3)

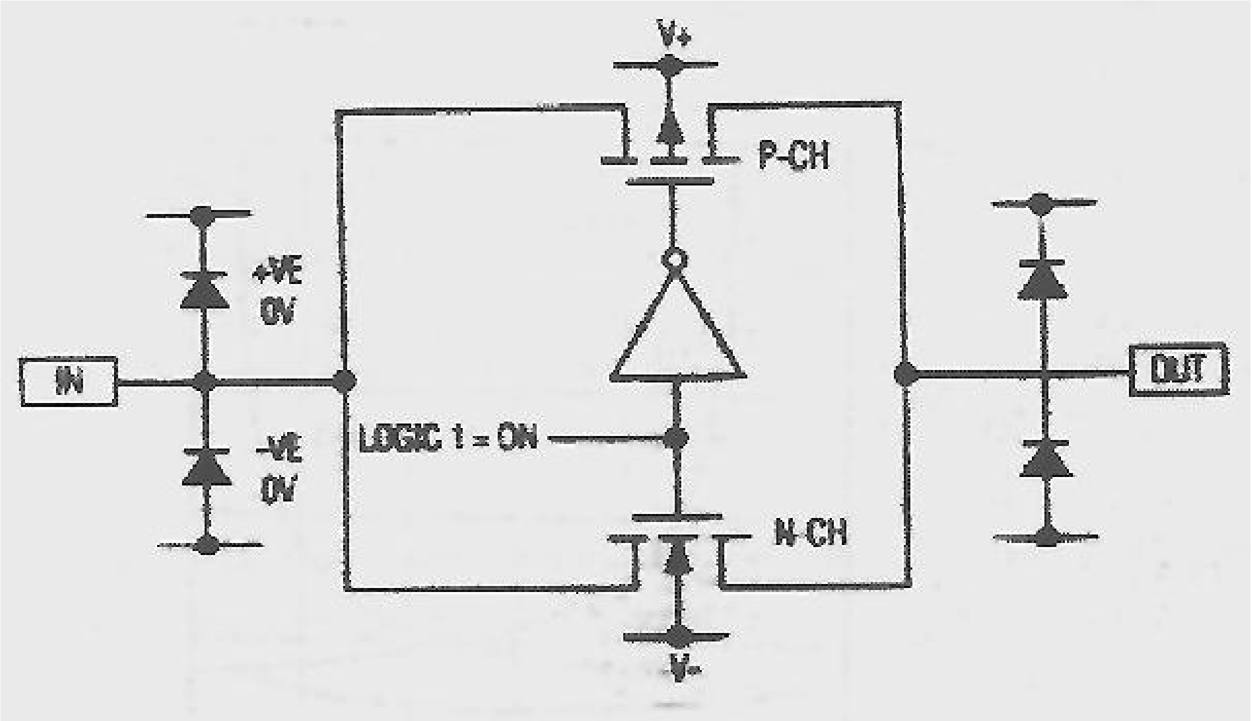

FIGURE 3. PARASITIC DIODES PROVIDE A PATH FOR FAULT CURRENT WHEN A CONVENTIONAL ANALOG SWITCH IS EXPOSED TO OVERVOLTAGE

56 Somewhat lower levels of current (below that of overheating and damage) can still cause latch up in the multiplexer.

57 And once it crosses a diode junction, the fault current becomes a flow of injected minority carriers that "spray" into the SILICON substrate.

58 Collected by other switching devices, this current can induce an error voltage in every channel.

59 Turning on parasitic diode clamps the multiplexer output to one supply rail - an action that can damage external circuits connected to that output.

60 The cause of damage may not be obvious, but an output transient (to the rail) caused by momentary over voltage at the multiplexer can destroy an A/D converter's input, or cause differential overload and long settling times in an op amp.

61 Several design measures offer protection for a CMOS multiplexer and its associated external circuits.

62 These measures include connecting a resistor in series with each channel input, connecting diode-resistor networks to control the fault effects, and choosing a multiplexer whose architecture and process technology provide fault-tolerant properties.

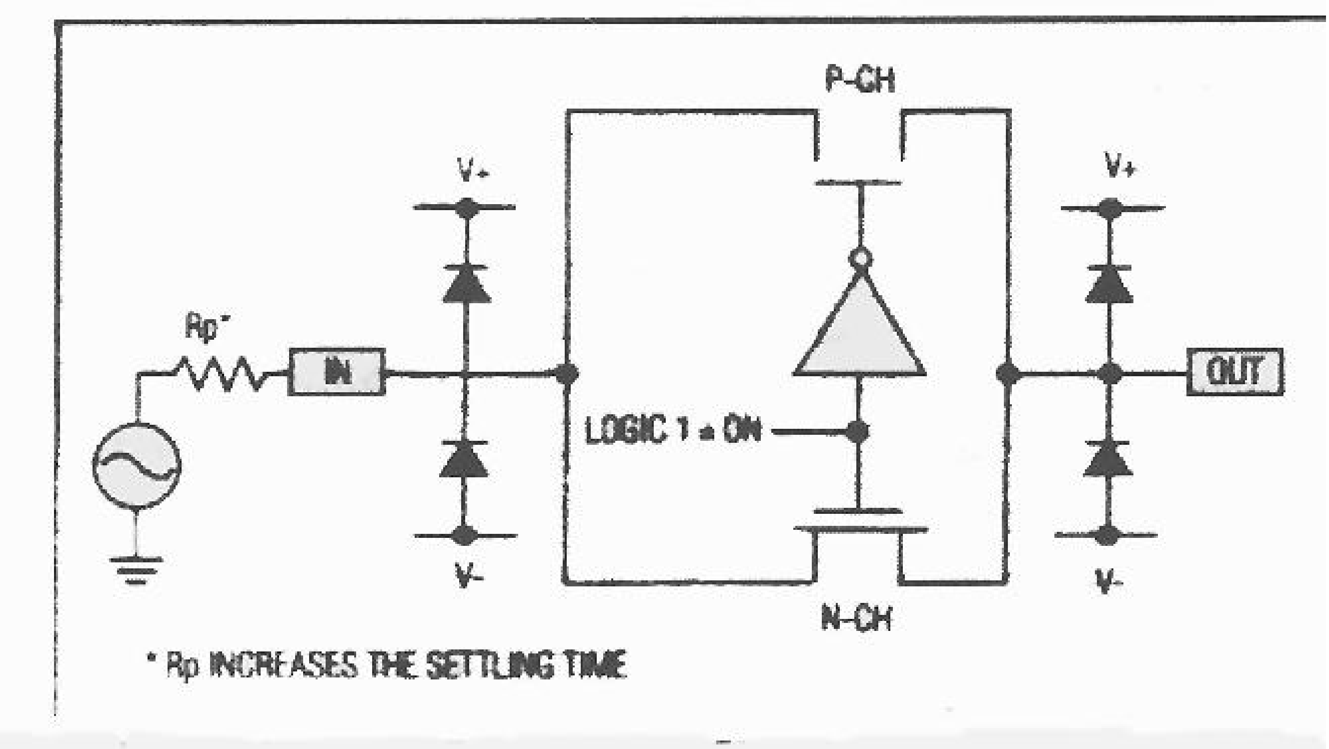

FIGURE 4. Adding a series resistor to the switch of figure 4 limits fault current, But it also adds to the switch resistance and lengthens settling time

63 The simplest form of protection adds series resistors that work in conjunction with the internal protection diodes (Figure 4)

64 The resistor values are critical because they present a trade-off: larger values give more protection but introduce greater signal errors.

65 Leakage current into the multiplexer also flows through the series resistors, causing an error voltage that worsens with temperature (the leakage doubles for each 8°C increase above ambient).

66 Lowering the resistor values can reduce error to an acceptable level, but the lower value may allow too much diode current, threatening latch-up in the multiplexer.

67 As a rule, unless otherwise specified in the data sheet section on absolute maximum ratings, you should limit the diode current to 20 mA continuous or 40 mA peaks.

68 Low leakage currents can offset this drawback of large protection resistors.

69 New, ultra-low-leakage multiplexers such as (MAX328 and MAX329) have extended the design limits for series-resistor protection over those for earlier-generation multiplexers.

70 The new devices low leakage (±1 pA at 25°C; ± 20 nA at 125°C) allows very high-valued protection resistors.

71 Resistors of 150 kΩ, for example, admit fault currents of only 1 mA while withstanding ± 150 V input.

72 At ± 1500 V, they admit fault currents of only ± 10 mA.

73 The resistors produce only ± 3 mV of additional error at 125°C.

74 Note that in the next figure at ± 1500 V protection resistors require 15 W ratings for continuous duty.

75 However, in most applications you can scale this thermal rating considerably because the over voltage has a much lower duty cycle.

76 External resistors thus offer flexibility - you can choose different resistor values for different channels in the same device, and scale their power ratings as required.

77 Integrated resistors, on the other hand, are constrained by their package power rating; this rating may limit the number of channels that can withstand over voltage at the same time.

78 The series-resistor approach protects the multiplexer, but it does not prevent corruption of signals in the selected channel.

79 These signals are at the mercy of over voltage in any of the unselected channels.

80 But the direct cause isn't overvoltage ; it's fault current (the injected minority carriers mentioned earlier) flowing into the substrate via one or more protection diodes.

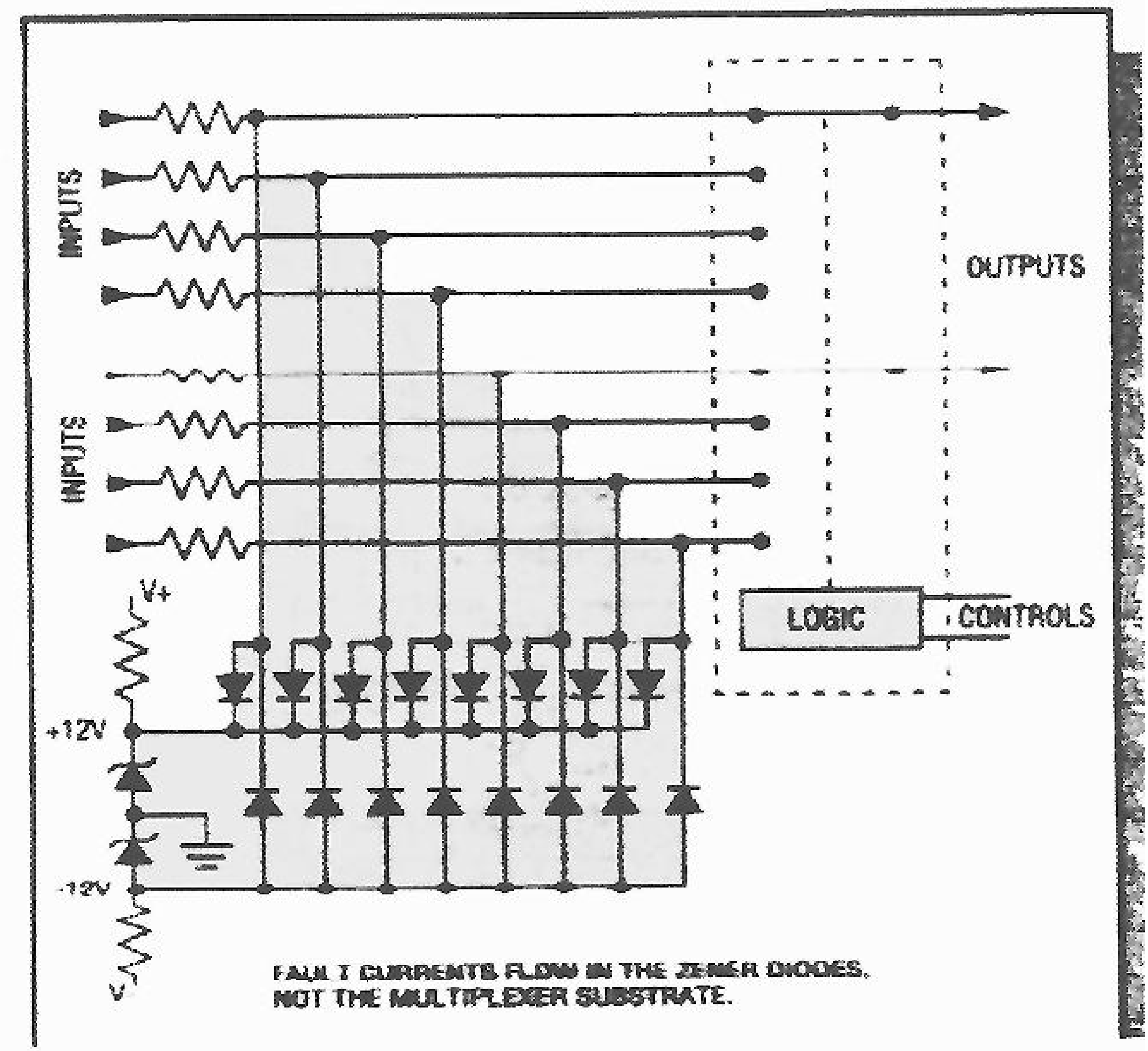
INPUTS
INPUTS
OUTPUTS
LOGIC
CONTROLS
V+
+12V
-12V
V-
FAULT CURRENTS FLOW IN THE ZENER DIODES,
NOT THE MULTIPLEXER SUBSTRATE.

FIGURE 5 Fault protections for a conventional multiplexer entails current limiting resistors. Two zener diodes for a bipolar clamp voltage network and dual clamp diodes for each channel.

81 Eliminate that substrate current and you eliminate the gross signal errors.

82 One way to handle the fault current is to divert it into an external network (FIGURE 5 above).

83 Two zener diodes produce ± 12 V clamps levels, centred within the multiplexer's ± 15 V supply rails.

84 Then, instead of flowing through an internal protective diode, the fault current due to over voltage on any channel flows through one of the two external protection diodes for that channel.

85 Though it offers excellent protection, this technique requires a large number of

external components.

86 And, the external diodes produce additional leakage current that precludes use of the high-valued series resistors discussed earlier.

87 The external components represent extra board space, not to mention the cost of purchase, test, through-hole assembly, and inventory.

88 I think that a better solution is to integrate this protection with the multiplexer, on a single chip.

89 Fault-tolerant multiplexer require no external components, yet are capable of withstanding high levels of over voltage without corresponding high levels of fault current.

90 They achieve this protection with an internal design that is quite different from that of conventional multiplexers.

91 Each switch in a fault-tolerant multiplexer is actually a series connection of three MOSFETs, in order n-channel/p-channel/n-channel (FIGURE 6a) below.

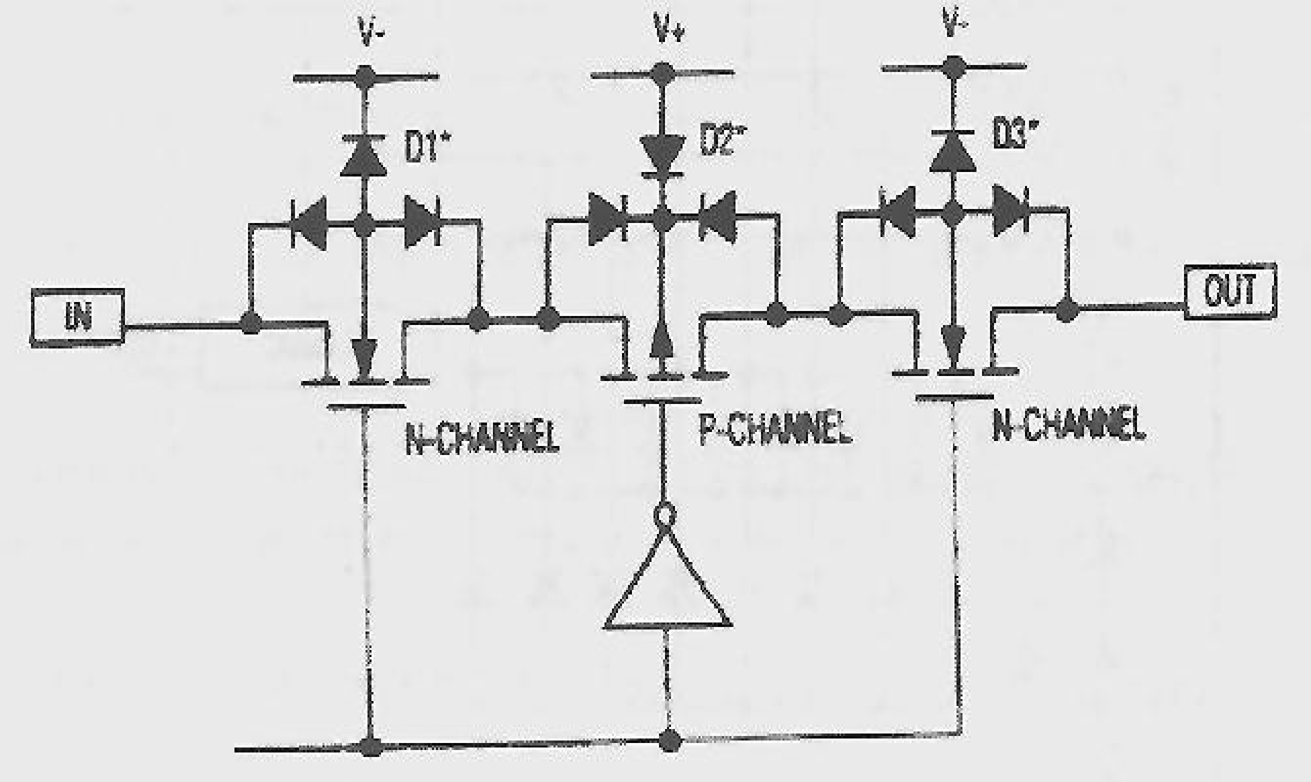

FIGURE 6a. The 3 MOSFET switch element of a fault-tolerant multiplexer

92 Internally generated drive voltages turn the switch on by simultaneously driving the n-channel gates to the positive rail and the p-channel gate to the negative rail.

93 The switch then remains on for as long as the analogue signal (which modulates the gate-source voltages) remains within limits set by the n- and p-channel gate-source thresholds.

94 Typical gate-source thresholds are 1.5 V for n-channel devices and 3 V for p-channel devices.

95 Therefore, with ± 15 V supplies the thresholds confine a multiplexer's input signals to the range - 12 V to 13.5 V.

96 Because one of the three MOSFETs in a switch begins to turn off as the signal exceeds either limit, the switch on-resistance versus input voltage assumes a characteristic "bathtub-shaped" curve FIGURE 6b, below.

97 The resulting high impedance in the off state is very convenient: the switch is off, blocking the over voltage, and fault current is virtually zeroed.

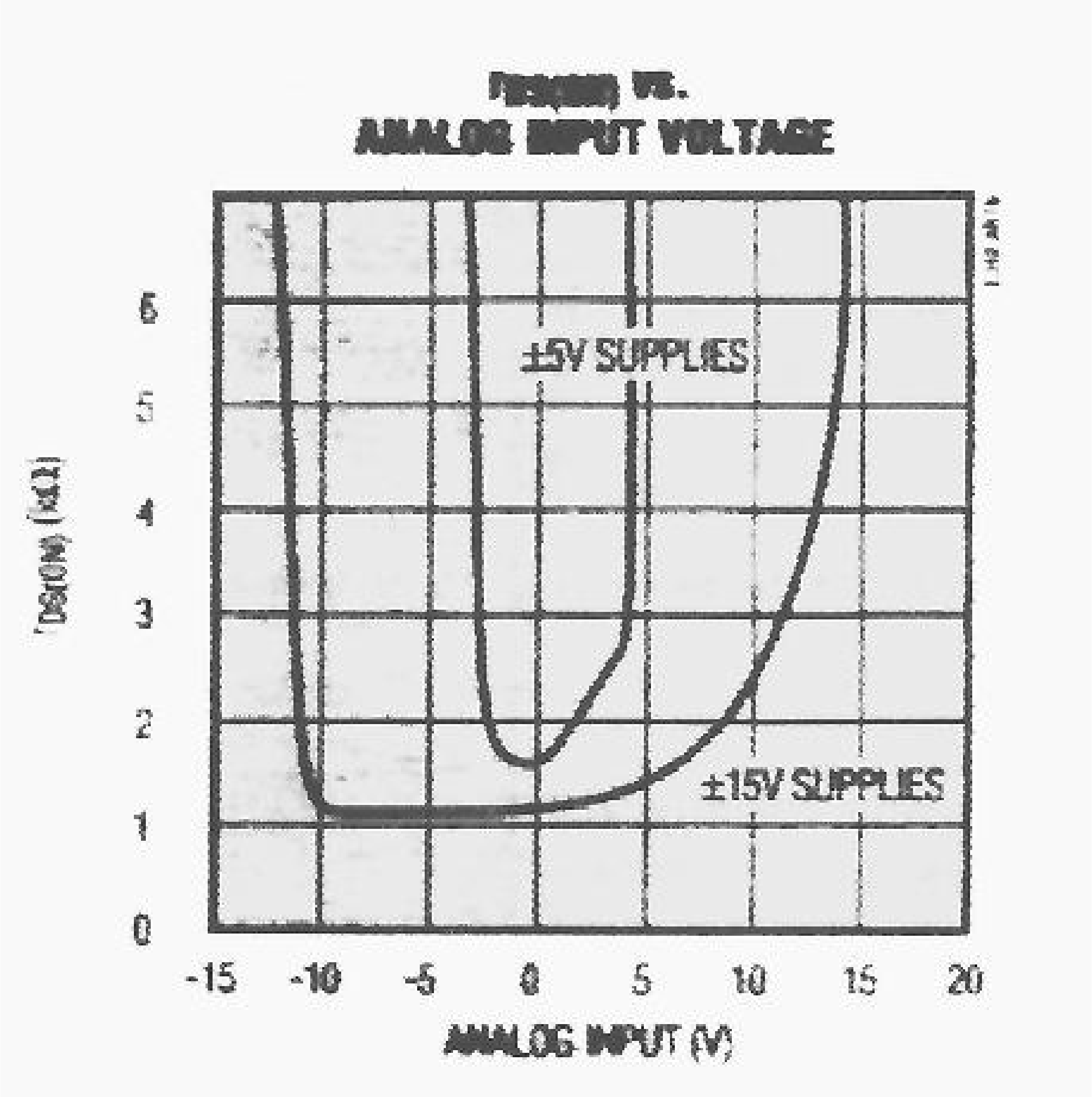

FIGURE 6B. **Has an on resistance characteristic shaped like a bathtub.**

COMPONENTS FOR EVALUATION 1993.

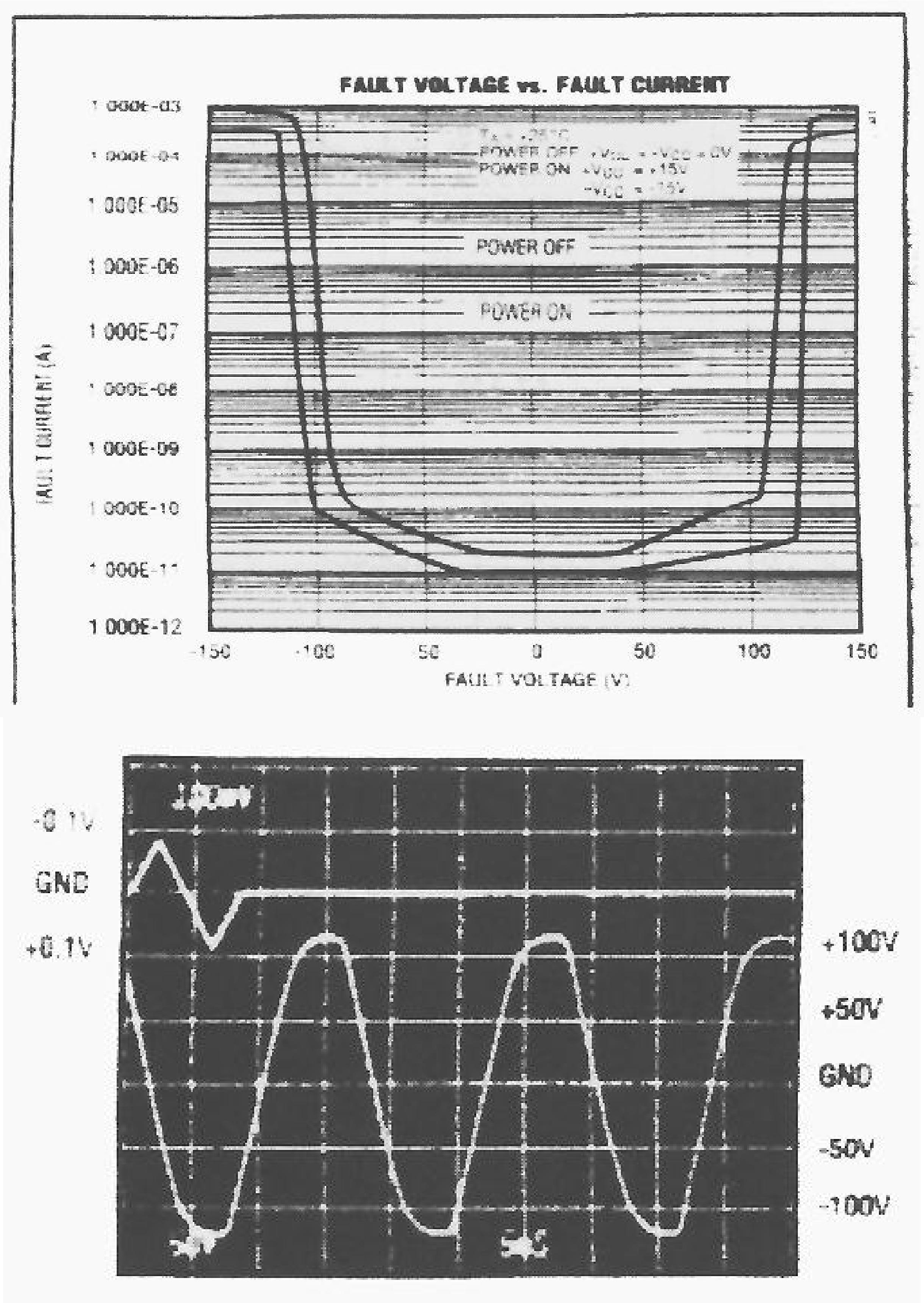

FIGURE 7. Report on next page.

COMPONENTS FOR EVALUATION 1993.

FIGURE 7. The onset of avalanche (fault) current in a MAX388 multiplexer defines a Fault-tolerant region of approximately ± 100 V.

A 0 V signal in the selected channel (top trace) is unaffected by ± 100 V applied to an off channel.

---

98 Substrate (fault) current flows only as a result of avalanche, which occurs when the over voltage exceeds a limit, set by the MOSFET's geometry and doping levels.

99 Below the avalanche limit, signals in the selected channel remain unaffected because the over voltage produces no substrate current

100 The series-connected switch also turns off when power is removed.

101 This behaviour simplifies the design of redundant systems; because multiplexers connect to common signal lines can be powered down without loading the lines.

102 Dr. Edwards please note that all these wonderful things quoted here were conceived in the world of fantasy, but their birth was delivered in the world of reality.

103 Yes Dr. Edwards I am here to teach you as well as every one else, so please listen very careful to what I state, and take note.

104 To my knowledge, Maxim offers several series-structure multiplexers.

105 The MAX358 and MAX359 devices (1-of-8 and dual 1-of-4) withstand over voltages to ± 35 V, and the MAX378 and MAX379 are similar devices that withstand over voltages to ± 75 V.

106 MAX368 and MAX369 add latched address inputs to the basic 35 V tolerant models, and the new MAX388 and MAX389 are latched models that withstand ± 100 V.

107 The non-latched devices are pin compatible with industry-standard multiplexers DG508/509: the latched devices are pin compatible with the latched models DG528/529.

108 As shown earlier in FIGURE 5, conventional multiplexers such as the DG508/509 require numerous external components to duplicate the fault-tolerant capabilities mentioned in this report.

109 MAX3XX devices not only save the cost and board area associated with external components, they offer capabilities not available in discrete-component circuits.

110 MAX3XX devices always turn off when over voltage is applied, but the switches of FIGURE 5, remain on in the presence of over voltage whether power is applied or not; this is only a tiny part of my thinking process for this development work.

I never knock NASA product, in effort to push mine - that is not cricket!

COMPONENTS FOR EVALUATION 1993.

111 The following (TABLE 1) shows the cost advantages of fault-tolerant multiplexers vs. the non-protected DG508/509 alternatives.

TABLE 1. NONLATCHED MULTIPLEXERS

	DG528/ DG529	MAX368/ MAX369	MAX388/ MAX389
Inherent O/V protection	NONE	± 35 V	± 100 V
Resale cost, 1 K pcs	$1.78	$3.25	$3.75
External component cost And assembly	$2.15*	0.00	0.00
TOTAL SYSTEM COST	$3.93	$3.25	$3.75

* = 28 External components (x) .03 (+) 28 (x) 2 leads (x) .007

The cost of the:

(1) 10 resistors

(2) 16 diodes

(3) 2 zeners

112 In FIGURE 5, is about $0.92, based on the purchases of 1 K pieces or more.

113 Plated-through holes are about $0.007 each, and the assembly cost for axial-leaded parts is about $0.03 per component.

114 The resulting total for multi-component protection exceeds that of the simpler, single IC protection offered by Maxim.

115 External components also require five times more board area than the IC alone.

116 The cost in FIGURE 5's, circuit is even greater if you consider the cost of troubleshooting, reliability effects, and other hidden expenses.

117 All of TABLE 1's ICs have an Absolute Maximum rating of 44 V between the V+ and V- terminals, but increasing die sizes from left to right necessarily escalate the 1 K-pc prices.

118 I had to leave out what would have been FIGURE 2, as part of it was far too large to include there. I have now reduced its size to half; it means that the contents do not stand out as clear as planned.

119 I hope that you will obtain enough **FACTS** from it to understand what it shows.

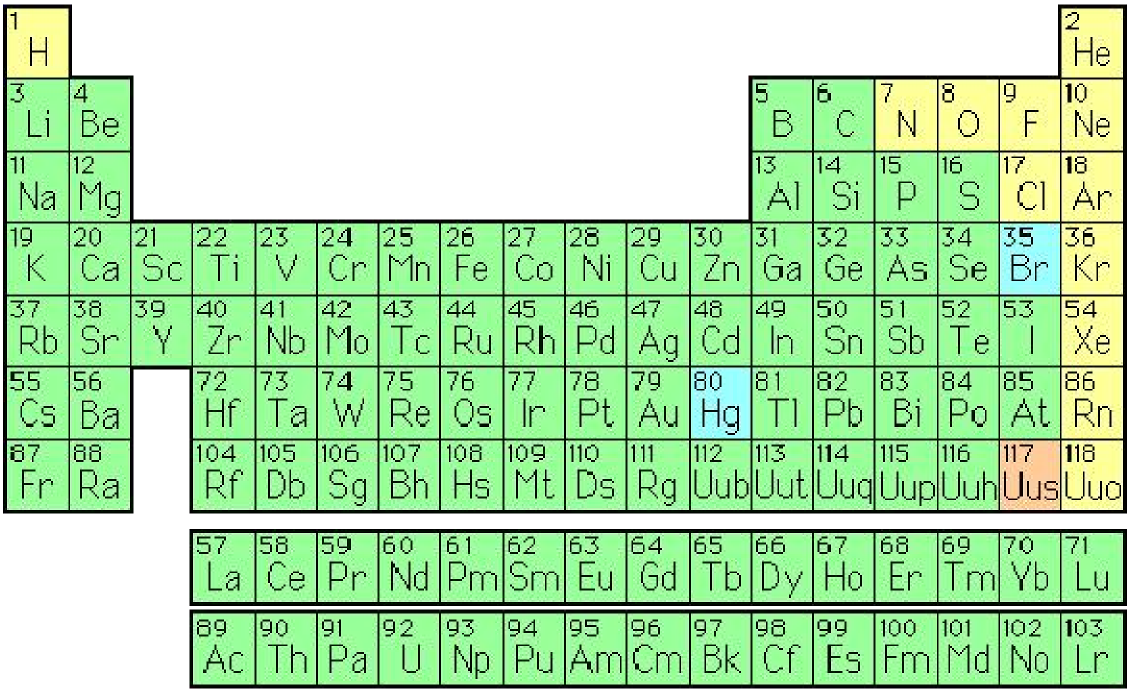

1 H																	2 He
3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
11 Na	12 Mg											13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
55 Cs	56 Ba		72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
87 Fr	88 Ra		104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Uub	113 Uut	114 Uuq	115 Uup	116 Uuh	117 Uus	118 Uuo
			57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu
			89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr

Mortimer-Berkshire-England.

LOCATION : *Headquarters-Mortimer-Berkshire-England.*
DIVISION : *Manned Flight.*
SEMINAR : *Star Ship Ezekiel MK V Concept.*
LECTURER : *John Roy Robert Searl.*
STATUS : *Head of R&D Human Studies.*

Star Ship Ezekiel MK V agreed concept:

Should be based upon my dreams one and two contents, that is to use the law of the squares in the concept design.

First, let us look at the squares, what are we talking about in reality?

8	30	34	3	5	31
35	1	28	9	26	12
18	17	15	21	27	13
19	20	16	22	10	24
25	11	14	23	36	2
6	32	4	33	7	29

22	88	100	7	13	91
103	1	82	25	76	34
52	49	43	61	79	37
55	58	46	64	28	70
73	31	40	67	106	4
16	94	10	97	18	85

40	150	170	15	25	155
175	5	140	45	130	60
90	85	75	105	135	65
95	100	80	110	50	120
125	55	70	115	180	10
30	160	20	165	35	145

37	147	167	12	22	152
172	2	137	42	127	57
87	82	72	102	132	62
92	97	77	107	47	117
122	52	67	112	177	7
27	157	17	162	32	142

You should see that I am using square six; you can also see that it is constructed from 36 cells, each cell holds a value.

I use square six for flight requirements, why do I use square six and not some other form of mathematics.

Son that is a good question, you see I have no knowledge on mathematics, as I had no formal education, like you because I have no normal hearing.

It was impossible for me to understand what my teacher was saying, I did nether the less understood what her hand was saying to my checks of my bum that she was going to make it red and hot – and she did to that issue there is no argument.

They fail to try to find out why a child cannot learn, do not question if they are deaf – they assume they hear.

Have selected the square that you are going to employ you then have to select the elements, which is important issue that the dream one say go square four, which is absolutely correct but only for land base use, unfortunate not powerful enough for flight work. Looking at dream one, I cannot help seeing that always there are eight squares, not only that I am always in square 2.

This is where you actually use your brain, ***WHY***?

Because that is what it is actually for, to be used, and I had no option but to use mine, that is why I conceived the Searl Effect concept.

Here was a dream, which kept re-occurring surely, there must be some logic reason for it. Now if we take two away from eight we get six, you agree, strange that dream run for six years.

That two covers many issues that are actually involved, I shall be involved in two completely different fields of technology:

1. ***Land base energy requirements.***
2. ***Flight base requirements.***

That dream also stating that you have to use four layers, three of which must be true elements plus one, which is not a true element but a molecule structure.

A concept is being born, and it is from such a chart, which as displayed at the heading of this document that I used within this technology.

Now I am standing in square two: there are two states involved here up and return back, thus I accept that I shall be involved in two different domains.

One, which is down representing ground systems; and other one up; representing flight. Thus cell two and cell four plus cell eight are the key issues as cell three you have to hop over and not into it.

I multiply these together and obtain:

2 x 4 = 8   8 x 8 = 64

Therefore, eight must have some importance, and likewise 64 – this is where my brain does the thinking and not me.

You can guess that eight relates to the periodic table, the 64 relates to the flight side, just for this moment I shall concentrate upon that issue eight.
It relates to the electrons, which I have already explained within this book the behaviour of the atom in its construct of the electrons.
That the four layers move from the outside layer the density goes up as you go towards the centre, except for the molecule layer, which controls the flow of electrons from the rare earth material.
Slowly a concept for research and study is in the development state
With all new idea, it is a case of probing in the dark until some light appears, as it will do if you probe deep enough, as I had to do.
At first, it looks impossible and often it is only a small issue to be corrected then its all go.

Dream one state there are two prime states in the universe.

1. ***Positive and negative***
2. ***Electrons and positrons***
3. ***Magnetic north and magnetic south***
4. ***Stationary plates and moving parts the roller s***
5. ***Conductors and insulators***
6. ***Superconductivity and non superconductivity***
7. ***Yes and no***
8. ***Left and right***
9. ***Hot and cold***
10. ***Particle and wave***
11. ***Male and female***
12. ***Current and magnetism***

Prof. John R.R. Searl
Inventor
of the 21. Century

These are all important issues, because they are all part of reality, and my world is reality without question.

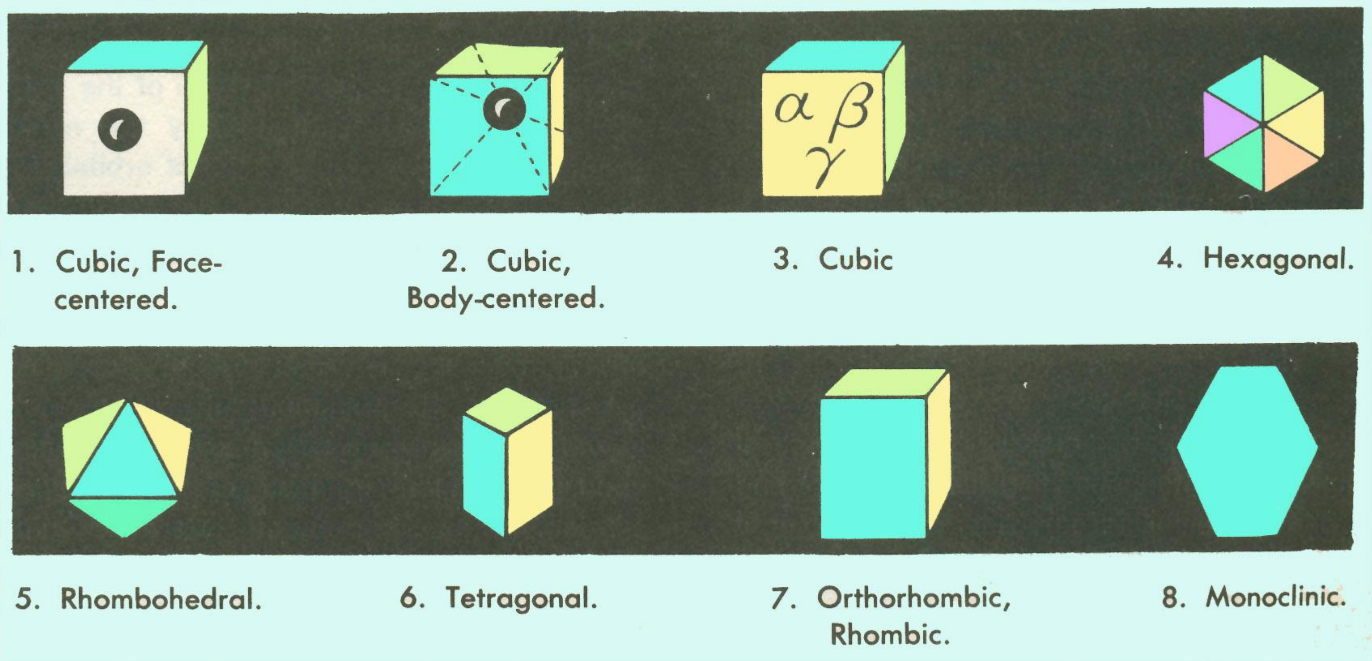

1. Cubic, Face-centered.
2. Cubic, Body-centered.
3. Cubic
4. Hexagonal.
5. Rhombohedral.
6. Tetragonal.
7. Orthorhombic, Rhombic.
8. Monoclinic.

Here are eight conventional diagrams of crystals forms.

Which is strange that this hopscotch patch had its system based on eight squares, in the following manner; from bottom to top?

Squares one, two and three in a vertical line then square four and five in horizontal mode followed by square six in the vertical mode ending with square seven and eight in the horizontal mode.

Square one is outside of the function as it represents that level one of all squares structures is its DNA upon which that square system is constructed.

Squares are no different to all the animals; a DNA system controls the construction upon how each species or square created by.

Maybe I have given you a bit too much to digest, after all it is my world born into it and have lived it all my life, therefore, to me it is easy for me to follow, as a yellow brick road would be.

It is an old saying when they is something unknown to be explain it is wiser to give it in small bite at a time so that it may be digested slowly by those who wish to learn and understand, agree most people can learn the problem remains will they understand that which they have learnt.

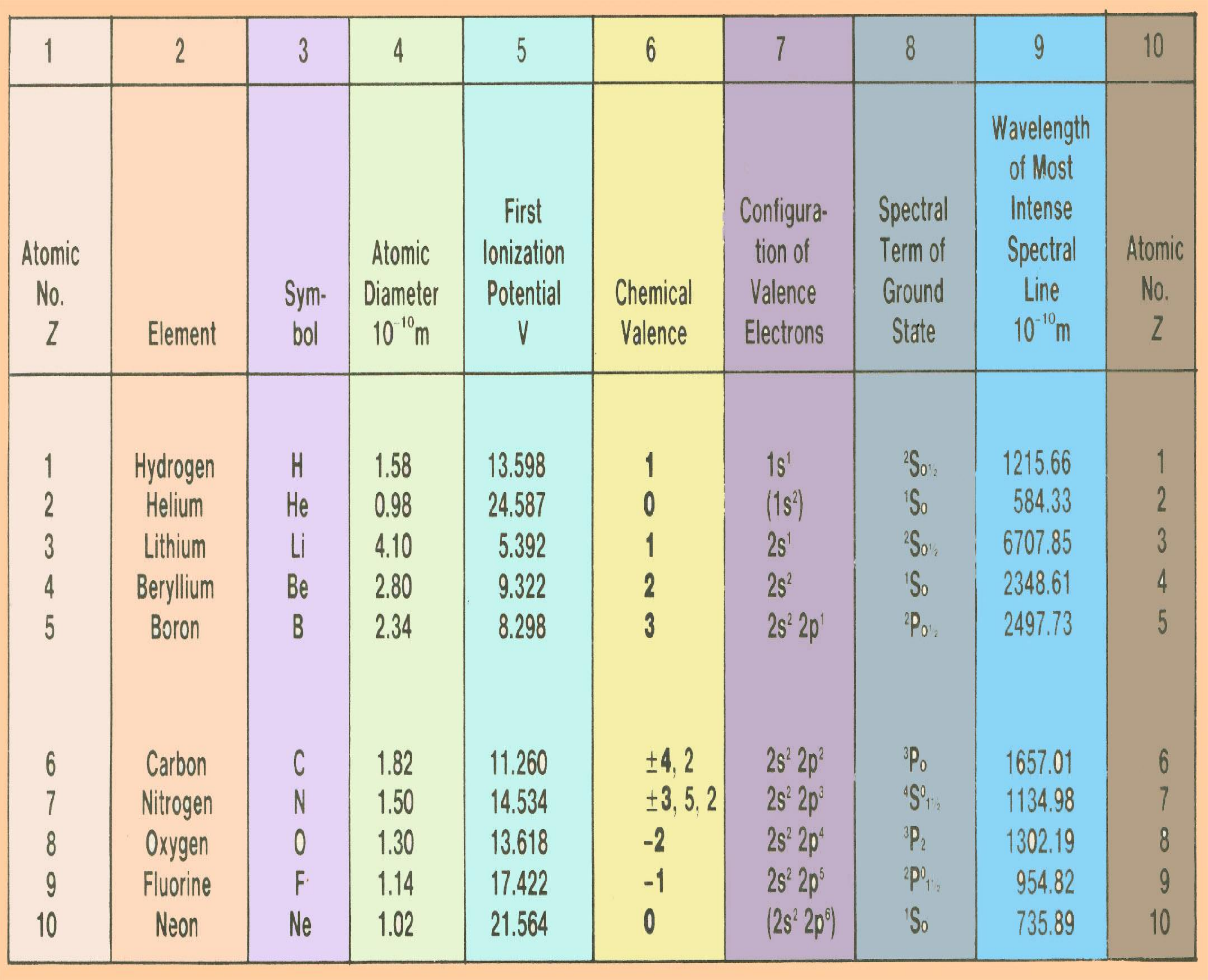

1	2	3	4	5	6	7	8	9	10
Atomic No. Z	Element	Sym- bol	Atomic Diameter 10^{-10}m	First Ionization Potential V	Chemical Valence	Configura- tion of Valence Electrons	Spectral Term of Ground State	Wavelength of Most Intense Spectral Line 10^{-10}m	Atomic No. Z
1	Hydrogen	H	1.58	13.598	1	$1s^1$	$^2S_{0 1/2}$	1215.66	1
2	Helium	He	0.98	24.587	0	$(1s^2)$	$^1S_0$	584.33	2
3	Lithium	Li	4.10	5.392	1	$2s^1$	$^2S_{0 1/2}$	6707.85	3
4	Beryllium	Be	2.80	9.322	2	$2s^2$	$^1S_0$	2348.61	4
5	Boron	B	2.34	8.298	3	$2s^2\ 2p^1$	$^2P_{0 1/2}$	2497.73	5
6	Carbon	C	1.82	11.260	±4, 2	$2s^2\ 2p^2$	$^3P_0$	1657.01	6
7	Nitrogen	N	1.50	14.534	±3, 5, 2	$2s^2\ 2p^3$	$^4S^0_{1 1/2}$	1134.98	7
8	Oxygen	O	1.30	13.618	-2	$2s^2\ 2p^4$	$^3P_2$	1302.19	8
9	Fluorine	F	1.14	17.422	-1	$2s^2\ 2p^5$	$^2P^0_{1 1/2}$	954.82	9
10	Neon	Ne	1.02	21.564	0	$(2s^2\ 2p^6)$	$^1S_0$	735.89	10

This is my world of reality, the world of tomorrow's technology, the ***Gyro-Flywheel-High Energy Density-Mechanical-magnetic device*** the ***Searl Effect Generator (S.E.G)*** and the ***Inverse-Gravity-Vehicle (I.G.V),*** you might not yet got it, but you should had done so, except for evil minds whose ignorance and greed have blocked you from having it. Unfortunate there is no more dangerous combination than those who steal in ignorance for personal gain

TABLE 1:

Electronic Properties of individual Atoms:

Columns 1 and 10 contain the atom number Z, equal to the number of elementary positive charges borne by the nucleus and also to the number of extra-nuclear electrons in the neutral atom.

Column 2 contains the name of the element or atom.

Column 3 contains the symbol for the element or atom.

Column 4 contains the atomic diameter as an expectation value from quantum mechanical average-energy-of-configuration Hartree-Fock computations for free atoms – atomic data 4, the unit is 10^{-10}m.

=0.0000000001 m WOW!

What do you think about that Flowerbower?

You poor creature of misfortune: that is beyond your brain capabilities to understand because that is reality.

Column5 contains the ionization potential in volts:
The ionization energy is the ionization potential multiplied by the elementary charge e, and has the same numerical value as the ionization potential when the latter is expressed in volts and the former in electron volts:

(1 eV≅ 1.60219 x 10^{-19} J)

That is: 1.60219 x 0.0000000000000000001 J

How are you coping Flowebower, guess there is a great amount of hot air rushing out of that rear end of yours practice makes perfect ready for YouTube attack again?
I feel that I should halt here to let you recover from this reading, as there is such a lot to show what I knew so long ago now.

Though I cannot no longer study at the same rate due to lack of hard cash not being available, at least I try to continue to learn about reality, which also includes you.
This document released by authority of:

Square 3
83 89 80 90
92 78 87 85
86 84 93 79
81 91 82
Square 4
elite

Prof. J. R. R. Searl Head of research and development.
Manned Flight – human studies.

Electric Express Locomotive, modelled on the class BB10000 of the Société Nationale des Chemins de Ferrancais (SNCF, or French National Railways Co.)
Finished in green
This was one of a hundred different trains that I used to display on special show days which were stolen while I was on Business in Canada by my family and replace to be stolen again by then while I was away.
They sold them for next to nothing in cash.

Cost me £8.00 plus postage and package.

IMPOSSIBLE MADE POSSIBLE BY A FOOL NAMED SEARL AND GROUP 1968.

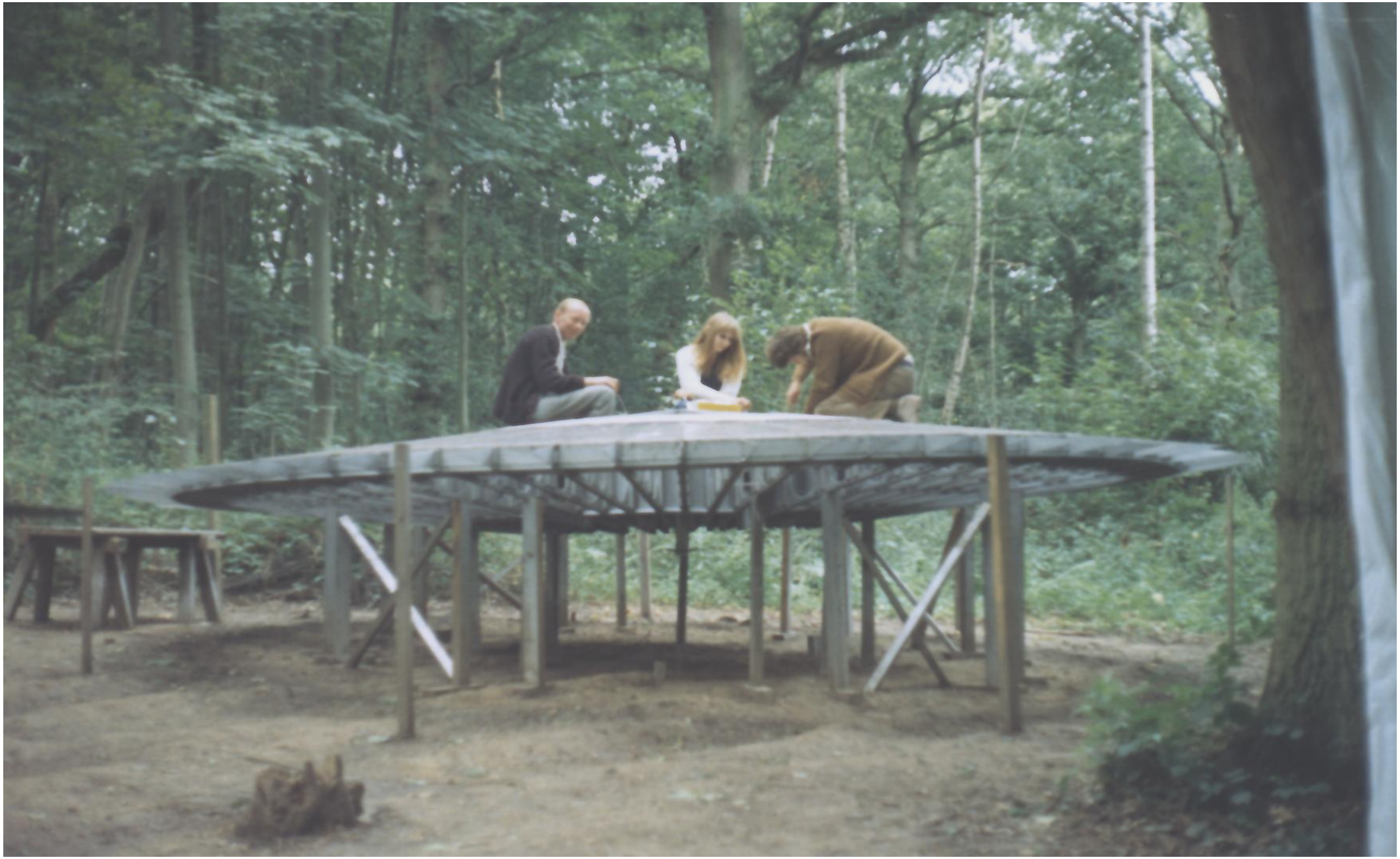

MY MEDICAL KNOWLEDGE 1993.

JANUARY 1993

1 **STARSHIP EZEKIEL MK.V**. is designed to have its own hospital, and this also apply to a housing complex upon that design.

2 But in 1968, its equipment design was very bulky and heavy, so now 25 years later, what is new that can be use to reduce both the size and weight of that equipment which is needed.

3 There is a subject matter upon the human body which is really never spoken about, yet to many it is a hell of a problem, one which I must make allowances for in the design of **STARSHIP EZEKIEL MK.V**. or even a housing complex upon the same structural concept.

4 The subject again is sensors, but in this case it is a chemical fibre sensor, which is the subject of this discussion.

5 This discussion is based upon a subject matter which one do not talk about at the dinner table, yet it is involved with your dinner one way or another.

6 It is a new optical sensor for monitoring bile in the stomach which is not just more accurate than existing techniques, it also works around the clock with little disruption to the patient's daily life.

7 Which of cause is great stuff for all of us who suffer from any bile related problems.

8 Chemical sensors based on optical fibres are advancing faster in biomedicine than those, which depend on competing technologies.

9 And fibre-based chemical sensors are developing more rapidly in biomedicine than in any other field.

BILITEC 2000
mark

THE COMPLETE BILE SENSOR, WHICH MONITORS BILIRUBIN.

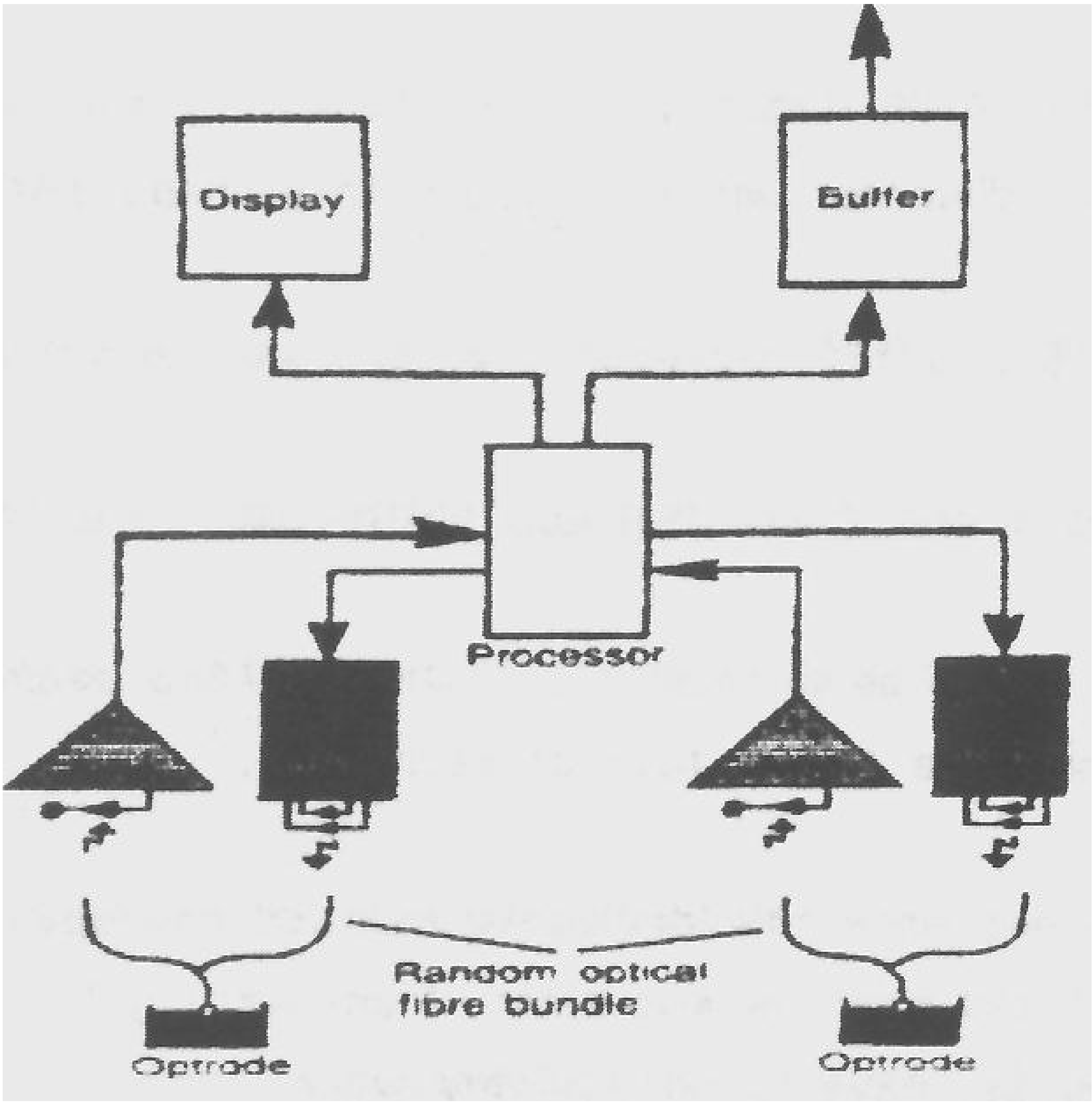

FIGURE 1b. IT'S OPTOELECTRONIC SYSTEM.

10 I regret that these last two pictures are missing, and I have had to scan them from book 4, which was a photocopy of a photocopy, so as to maintain the same data.

11 Wherever I can I have replaced those pictures or drawings with new ones, but some take too long to re-draw them.

12 But I hope that it will give some indication of what I am talking about.

13 Some scientists might disagree with these two claims which were made by Francesco Baldini, who researches chemical sensors at the Istituto di Ricerca sulle Onde Elettromagnetiche (IROE), Florence, Italy.

14 Yes Mr. Anders Heerfordt, Dr. Edwards and Gunnar Sandberg, I bet that none of

you knew that I know about that place or that person.

15 I also bet that you did not know that I knew the following **FACT** that an IROE sensor, which continuously monitors bile in the oesophagus (for those who do not know what oesophagus refers to, don't worry I am here to tell you, its the canal which extends from the pharynx to the stomach.

16 To my knowledge it is about nine inches long, at least that is what I understood back there in 1950.

17 You are happy now that you know what I am talking about; you are not, why?

18 You are confused as to what the term pharynx relate, o don't worry, I am here to tell you. It’s the muscular pouch lined with mucous membrane situated at the back of the mouth.

19 It leads into the oesophagus, and also communicates with the nose through the posterior nares, with the ears through the Eustachian tubes, and with the larynx.

20 Now you are happy?

21 You are not, you don't understand the terms used, well I am sorry I have to leave that for another book, or otherwise the subject of this discussion will never appear within this book.

22 This sensor functions by an optical fibre probe inserted through the patient's nose, has just gone on sale.

23 At this moment in time, I do not know what it cost.

24 Data from the optical transducer or optrode on the end of the fibre are collected in a small battery-powered unit, which is light enough to hang off the patient's belt.

25 The sensor takes readings every 4, 8 or 16 seconds, which it displays or records for analysis.

26 Buttons on the unit allow the patient to record pain, meals and periods of rest.

27 The Italian researchers have just started laboratory tests of a similar sensor to monitor acidity levels in the stomach.

28 They are also working on prototypes of fibre sensors that measure the pH of blood, oxygen levels in liquid or gas streams, and others, which, in combination with enzyme reagents, could analyse biological compounds.

29 By now you should all know that I am determined to see that every functional unit within the concept of **STARSHIP EZEKIEL MK.V**. or in any similar housing compound will be of the absolute accurate and fast technology that science can offer.

Well you so-call experts what a disgrace you are to your profession with your insults of crap.

30 Baldini told the Biomedical Optics conference in Hungary that fibre sensors are succeeding in biomedicine because they lack electrical contacts, they are tiny and they come in virtually any shape.

31 These qualities, which I appreciate, added to fibre's inherent safety, make the technology "competitive with traditional sensors in both chemical diagnostics and multi-test kits and allow, in some cases, performances which are unique."

32 As I stated in 1968, that biomedicine's requirements on **STARSHIP EZEKIEL MK.V**. demand for quality and safety drives innovation by forcing designers to think carefully and to base their sensors on a "thorough spectrophotometric investigation".

33 As I stated in 1968, spectrophotometry leads to a "better understanding of chromophores, immobilised or not, which act as an optical transducer and optimise the optrode's optical performance", and this is also what Baldini stated.

34 "If the goal of the sensor is in vivo measurements, then particular attention must be paid to the biocompatibility of the optrode."

35 There are three basic elements to each sensor :

(1) An optoelectronic system

(2) An optical fibre

(3) An optrode on the end of the fibre

36 Some optrodes carry chromophores, substances that have unique absorption spectra in the presence of specific chemicals.

37 Other optrodes depend on the absorption spectra of target chemicals such as bilirubin, which has an absorption peak at 452 nm.

38 Bilirubin is the main pigment in bile, which can cause disease if it enters the stomach from the duodenum.

39 You look confused at a much higher level, let me deconfuse you, Duodenum (du-o-de'-num) the first nine or ten inches of the small intestine, from the pyloric opening of the stomach to the jejunum.

40 The pancreatic and common bile ducts open into it.

41 Now you are happy, you are not, sorry, I must continue with my discussion, or otherwise I will never get it done.

42 A typical sensor has two identical fibre channels to monitor a pair of sites simultaneously.

43 Two diodes in each channel couple a reference wavelength and a signal

Wavelength (matched to the target chemical or the chromophore in the optrode) into one arm of a forked fibre.

44 The forked probe is a randomly mixed bundle of polymer optical fibres in a biocompatible black polyurethane catheter.

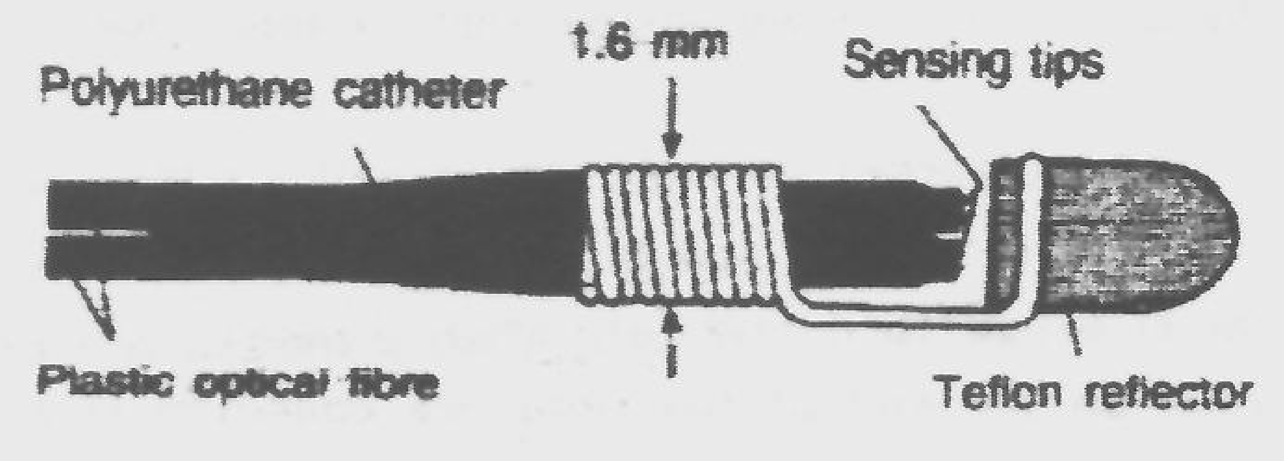

FIGURE 2. Open optrode for pH measurements with bromophenol blue

45 Light from the two sources passes to the optrode and then reflects back along the fibre and out of the second arm onto a photodetector, an amplifier and a processor.

46 The processor drives the two light sources and passes collected data to an internal buffer for transfer to an external computer.

47 Baldini has developed several designs for optrodes.

48 One is based on chromophores immobilised on glass microspheres in a stainless steel tube, which has perforations or slits in its wall.

49 Each of these "windows" is smaller than the spheres but large enough to allow liquids to enter.

50 An early optrode to measure acidity levels in the stomach was based on a tube 15 mm long by 0.8 mm in diameter, perforated with 70 µm holes and containing a thymol blue chromophore immobilised on glass.

51 Test measurements show that the prototype is sensitive to 0.007 pH to 4 pH, but the response time of several minutes is too long, for my requirements for **STARSHIP EZEKIEL MK.V**. Programme.

52 A narrower tube containing more holes would improve response times but the viscosity of gastric juices in vivo would reduce them again.

53 Although the perforated optrode was abandoned as a gastric monitor, a modified design with slits instead of holes measures the acidity of blood to 0.02 pH from 6.8 to 8 pH, with response times of about one minute for each 0.25 pH step.

54 Well that is getting better.

55 However, can we improve on that timing period?

56 The chromophore is phenol red immobilised on polymer spheres and the sensor's sources emit at 570 nm for the signal and 660 nm for the reference.

57 Tubular optrodes are compact and simple.

58 They also suffer from slow leakage of the phenal red and, unless degassed, air bubbles reduce their typical response time of 30 seconds.

59 IROE has decided, therefore, that the design of the optrode is not suited to in vivo monitoring of acidity levels in blood.

60 Another optrode, designed to detect OXYGEN, is based on an organic metallic compound adsorbed onto a plate.

61 Such compounds, bis (histidinato) cobalt (II) in this case, change their absorbency spectra in the presence of OXYGEN.

62 A layer of SILICONE protects the compound, which is fixed to one wall of a flow cell.

63 The SILICONE allows the compound to react with liquids while protecting it from physical damage.

64 The 470 nm signal and 620 nm reference beams shine through a quartz window in the aluminium cell and reflect off the compound.

65 The optrode is accurate to 0.3 vol. percentage in the first 0 to 10-vol. percentage of OXYGEN in an OXYGEN -ARGON mixture.

66 In addition, is slightly less accurate at higher concentrations of OXYGEN.

67 Response time is about two minutes, depending on the thickness of the SILICONE layer.

68 One of IROE's most promising optrode designs exposes bare ends, or chromophore-carrying spheres fixed on the ends of fibres, to fluids such as gastric juices.

69 This open optrode design was used for the bile sensor, which has just gone on sale, and a version of the design is under test as an acid sensor.

70 At the end of a 500 µm diameter polymer fibre is placed against a heated layer of glass microspheres, the polymer softens and up to 50 spheres carrying about 5 Ng of chromophore penetrate the tip.

71 When the polymer cools, it hardens and fixes the spheres in place.

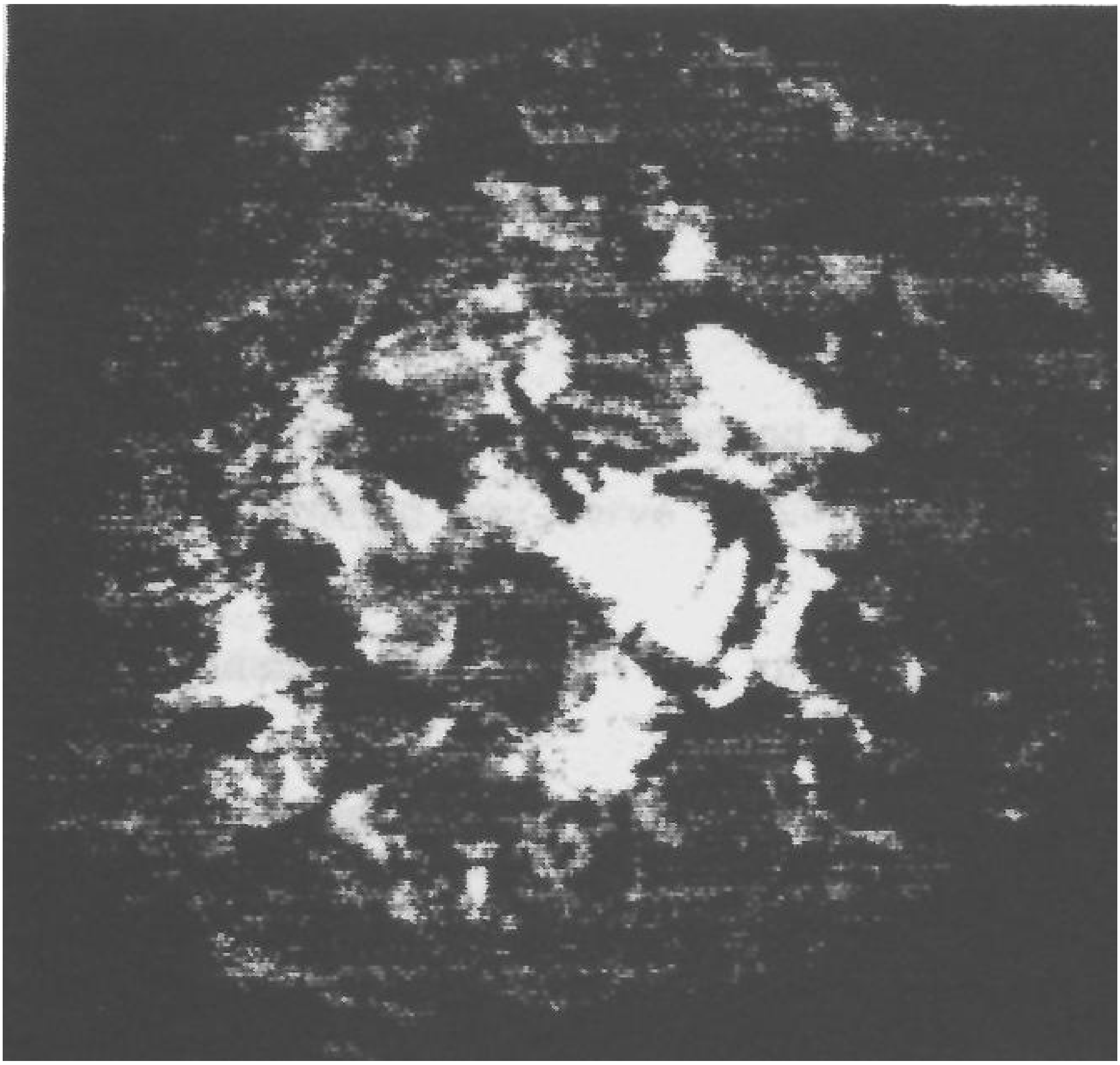

FIGURE 3. Glass spheres carrying a chromophore fixed to a 500 µm Diameter polymer optical fibre.

72 A Teflon reflector mounted by a wire in front of the bare or bead-carrying fibre tips improves the coupling of the signal and reference beams to create an optrode, which is open to gastric juices.

73 Its open structure allows liquids in the stomach to remove contaminants, such as mucus collected by the optrode as it is inserted through the nose and oesophagus, and so ensures a fast response from the sensor.

74 Measurements in vitro proved that an open optrode with bare fibre tips detects bilirubin to an accuracy of four µmol compared with the 0 to 100 µmol range expected in the stomach.

75 Francesco Baldini reports that, based on over 100 in vivo measurements, bare fibre tips are better than conventional medical techniques for detecting bile.

76 For example, the sensor can determine the moment when bile enters the stomach from the duodenum in conditions where conventional measurements do not even detect a change in acidity.

77 Other techniques, such as the injection of a radioactive marker, which is rapidly eliminated by the liver and bile ducts, record the presence of bile in the stomach.

78 However, radioactive markers, unlike the new sensor, cannot give information about the concentration of bile.

79 An open optrode, with spheres carrying bromophenol blue fixed to its fibre tips, has just begun in vivo trials to monitor acid levels inside the stomachs of patients.

80 The trials are in collaboration with Prof. Paolo Bechi of Clinica Chirurgica III at the University of Florence.

81 Two sources shine 595-nm signal and 870 nm reference beams down its 0.5-mm diameter fibres.

82 So far, the in vitro trials show that the optrode is accurate to ó0.05 pH in the range of 3 to 7 pH, with a response time of a few seconds.

83 Now we are talking my kind of language that is more like what I wish to hear.

84 Baldini reports that hysteresis is greater in acid detection with the open designs than with the perforated stainless steel tube optrode.

85 "At the moment we can't give a sure interpretation of this because the only difference is the thickness of the beads," he says.

86 However, the effect of hysteresis with the open optrode is consistent enough to be removed by software.

87 The other good news about new acidity sensor is that its sensitivity is not affected by temperature and it can be stored in dry conditions for long periods.

88 Baldini now predicts the device will join the bile sensor on the market in about 18 months. I will be watching out for it.

89 This is only one subject matter where sensors will play a vital part of the **STARSHIP EZEKIEL MK.V. PROJECT**, other sensors will be discussed within these books.

90 I will change this point to another point, which often appears.

91 QUESTION

Over the last two years, I have noticed that my penis becomes curved when I get an erection, I have seen my doctor, who has diagnosed **PEYRONIE'S DISEASE**.

MY MEDICAL KNOWLEDGE 1993.

I am 22 and a virgin - and about to go away with my girlfriend, with whom I have a very caring and deep relationship.

I am sure that we will make love and I am worried that I will not be able to manage it.

Can you give me any advice?

92 ANSWER

My first piece of advice is not to worry.

PEYRONIE'S DISEASE, named after a French surgeon, is caused by harmless hard tissue, which pulls the penis to one side or another when you get an erection.

It rarely affects your ability to make love and is best forgotten about - unless it is painful or causes severe deformity, in which case your GP may want to refer you to a specialist.

I can tell you from actual experience that the pain can be beyond a limit, which would allow you to even think of trying to make love.

However, it was not **PEYRONIE'S DISEASE**, which was the cause of my problem; it was the lack of lovemaking.

Not all men have straight penises when erect.

I have seen some unusual sights in my career, one case that was a perfect
Cork screw from base to tip, very, few of which ever interfere with
Love making.

I must admit I do not know how the chap above managed, as it would need three times more space to accommodate it.

Indeed some men are proud of their individuality!

Enjoy your holiday.

93 In 1968, Peyronie's disease was unknown to me, so at that time I would not had known of that problem.

94 Strange as it may seem to one, as I reflect upon this matter, that during my training on the hospital ward, I did not know of such problems. Then let us face it, it is not a normal practice for men to erect themselves, just to show the medical staff what they got as a weapon.

95 To some people they may feel that these discussions are not important to the **STARSHIP EZEKIEL MK.V. PROJECT**, but that is where you go wrong.

96 The designer needs every bit of knowledge across the whole of humankind requirements.

7155
BALTIMORE AND OHIO

THE IMPOSSIBLE MADE POSSIBLE BY FOOLS LIKE SEARL.

1 This is but one of the trains Searl operated on his system using the *S.E.G*. Technology

2 An articulated steam locomotive with tender of the Baltimore & Ohio Company, which again shows proof that the impossible can often become possible.

3 It has a wheel arrangement 2-8-8-0; somehow dream one has again popped in to say hello!

4 Projected by the USRA – United States Railway Administration, the loco group ED-5 of the B & O was built between 1919 and 1920 by the Baldwin locomotive company.

5 Well Dr. Edwards I bet you never knew that I knew that, but the question is did you know that?

6 They had the Baker distribution system and were equipped with a Vanderbilt type tender.

7 The B & O used the articulated locos to haul freight trains on the uneven lines of the Cumberland and on the severe ruling gradients of Crandberry, Cheat River, Newburg.

REALITY OR FANTASY 1978.

Searl’s rail system, which became public, shows locos operating on the **S.E.G**. principal.

8 Very powerful motor driving the two articulated motor groups through a longitudinal shaft and pendular transmission with universal joints.

9 Metal frame in spite of its considerable length this model has an exceptional smooth running and could negotiate curves down to 40 CMS. Radius thanks to the two articulated motor groups.

10 This model is extremely detailed and fully operating valve gear system

11 Length of this model was 40 CMS.

12 Of course, I can carry on and on about this train and its history, but the sole purpose of this article is to answer those idiots who claim by their expertise of bullshit that Searl never had anything.

13 You the reader have to make up your mind if Searl did have or not.

14 Was there such a model available, which he could, had brought matching the claims shown here?

15 Searl does not confirm that today it is still possible to buy a copy of this train, as he has been out of action for some time.

16 His sons, either sold at extreme stupid prices, as they had no knowledge of the cost of those engines stole all of his rail system, from the first to the second one.

17 Stupid people who have no common sense use this unfortunate situation to expose how cleaver they are in bullshitting.

18 That is the only reason they are experts.

19 They are sick in the mind, and represent the reason why this planet is in such a mess.

20 No other animal upon this planet does so much damage. This is a fact!

21 This is the man who gives his whole life to study the problems of this planet; his answer is to transfer all Homo sapiens to Mars. Then this planet will be a paradise.

Prof. John Roy Robert Searl Head of R&D Human Studies.
Medical Division.

SEARL'S RAIL CONTROL SYSTEM 1968.

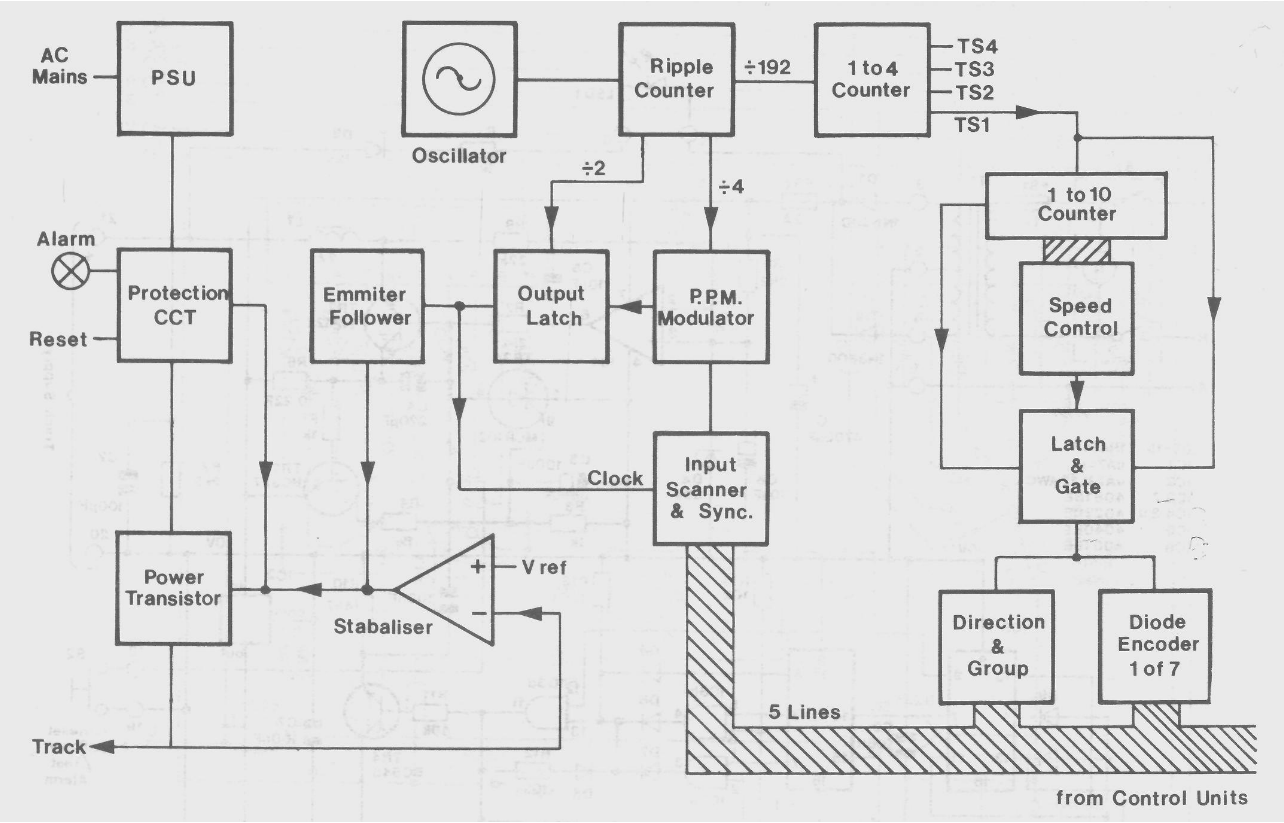
AC Mains
PSU
Oscillator
Ripple Counter
÷192
1 to 4 Counter
TS4
TS3
TS2
TS1
÷2
÷4
1 to 10 Counter
Alarm
Reset
Protection CCT
Emmiter Follower
Output Latch
P.P.M. Modulator
Speed Control
Latch & Gate
Clock
Input Scanner & Sync.
Power Transistor
Stabaliser
V ref
Direction & Group
Diode Encoder 1 of 7
5 Lines
Track
from Control Units

1 Yes, this gives you the insight how things progressed through time

2 It is a time-related project based upon cost and components availability.

3 The above diagram shows the common and control boards, bear in mind that at this time much of the controls were operated by a person name Searl.

4 Below is the receiver circuit.

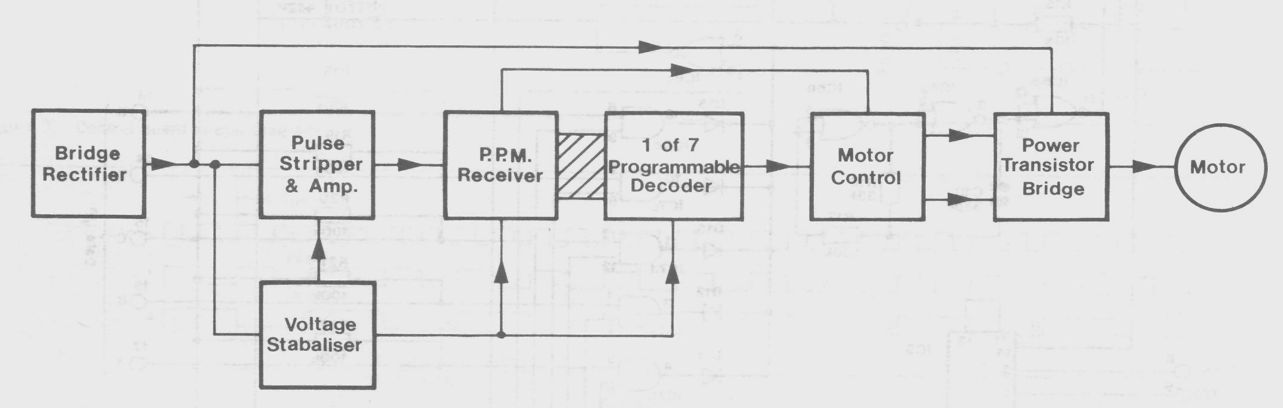
Bridge Rectifier
Pulse Stripper & Amp.
P.P.M. Receiver
1 of 7 Programmable Decoder
Motor Control
Power Transistor Bridge
Motor
Voltage Stabaliser

5 It was a great step forward from the masses of levels, which had been in use prior to this insertion to the system.

6 I may have already referred to this before; I feel that I should extend this information in details, because it could still be used if ever I could afford to replace that loss.

SEARL'S RAIL CONTROL SYSTEM 1968.

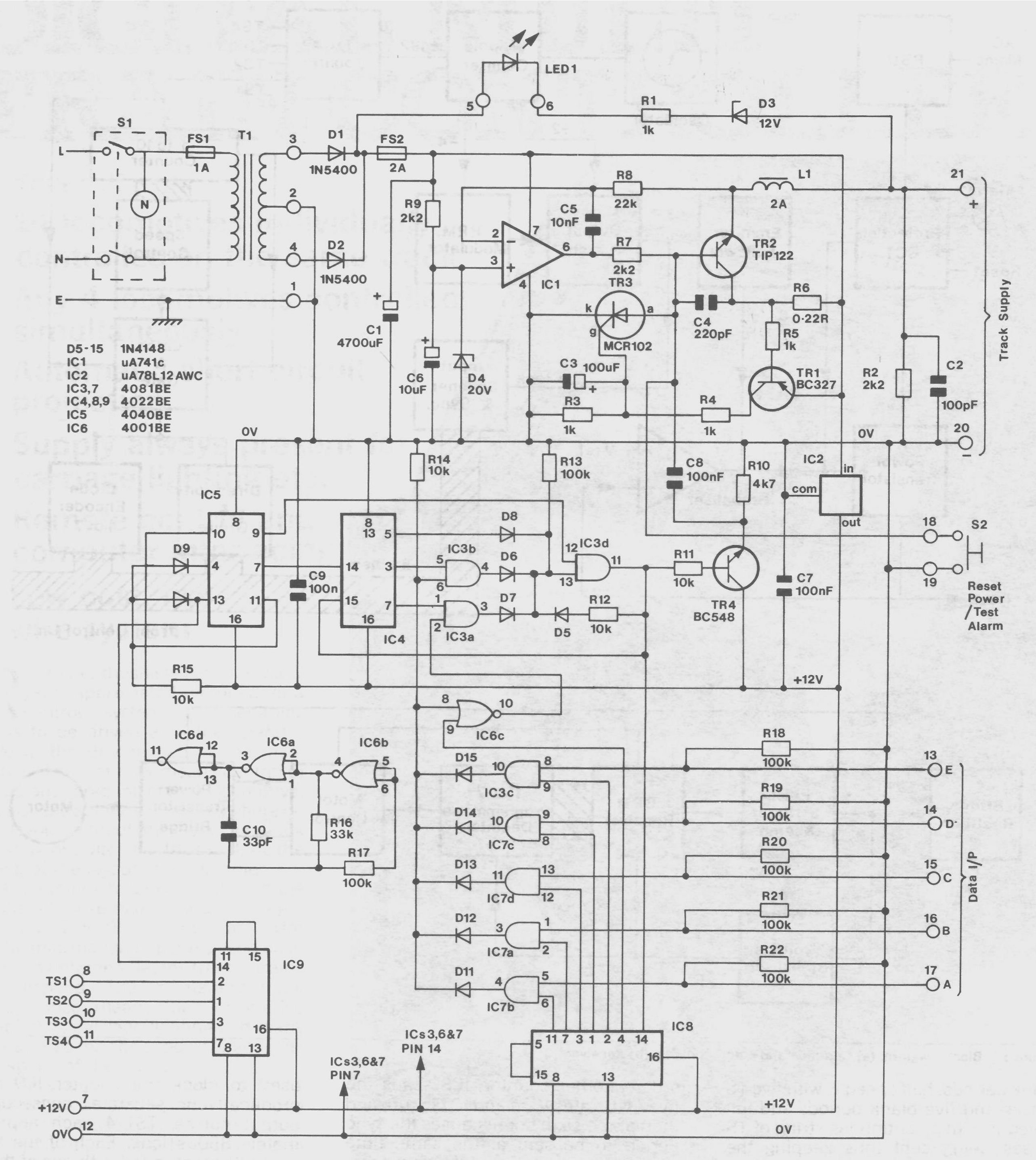
LED1
5
6
R1
1k
D3
12V
S1
FS1
1A
T1
3
D1
1N5400
FS2
2A
L
N
2
4
D2
1N5400
E
1
R9
2k2
R8
22k
L1
2A
21
+
C5
10nF
7
2
3
6
4
IC1
R7
2k2
TR3
TR2
TIP122
R6
0·22R
k
a
g
MCR102
C4
220pF
R5
1k
C1
4700uF
C3
100uF
TR1
BC327
R2
2k2
C2
100pF
C6
10uF
D4
20V
R3
1k
R4
1k
Track Supply
D5-15 1N4148
IC1 uA741c
IC2 uA78L12 AWC
IC3,7 4081BE
IC4,8,9 4022BE
IC5 4040BE
IC6 4001BE
0V
20
−
R14
10k
R13
100k
C8
100nF
R10
4k7
IC2
in
com
out
IC5
D8
IC3b
D6
12
IC3d
R11
10k
18
S2
19
Reset
Power
/Test
Alarm
D9
C9
100n
14
15
16
D7
IC3a
D5
R12
10k
TR4
BC548
C7
100nF
IC4
R15
10k
+12V
IC6c
IC6d
IC6a
IC6b
R18
100k
13
E
D15
IC3c
R19
100k
D
D14
IC7c
C10
33pF
R16
33k
R17
100k
R20
100k
C
D13
IC7d
Data I/P
R21
100k
B
D12
IC7a
IC9
R22
100k
17
A
D11
IC7b
TS1
TS2
TS3
TS4
IC8
ICs3,6&7
PIN 14
ICs3,6&7
PIN7
+12V
0V
+12V
0V

Common board PSU circuit diagram.

7 If you remember from my past books and newsletters that I wanted, because of the increased number of trains to be able to operate them faster in the effort to demonstrate that you can operate a real system by the same methods.

8 Up till 1968 the trains could be controlled manually, by masses of switches.

9 Unfortunate that did not represent much of a real business by rail.

SEARL'S RAIL CONTROL SYSTEM 1968.

10 Here is the view of the completed control PCB, remember there is not just one off there are many of them.

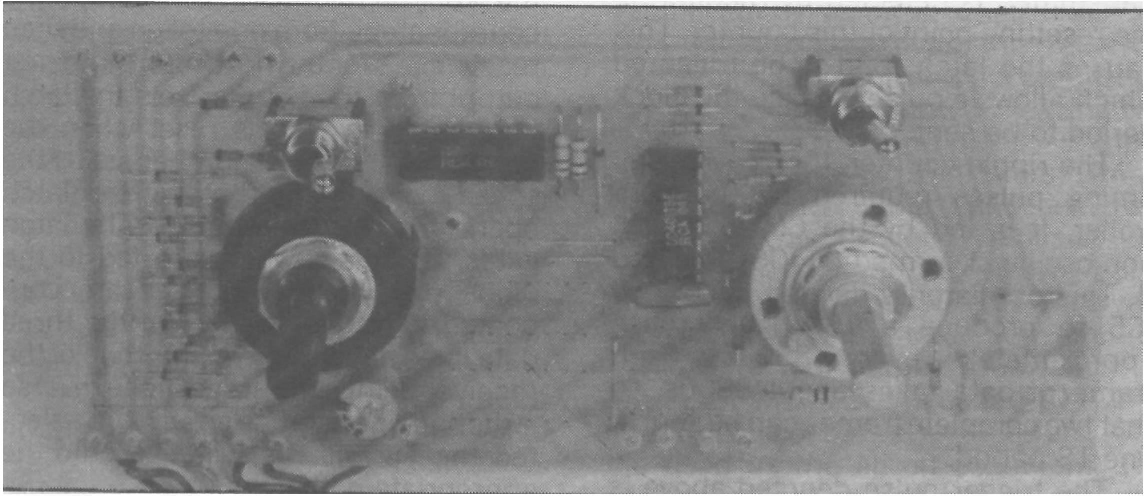

Completed control PCB, sorry it has not come out very clear.

11 Bear in mind at this time the operation is still controlled by manpower.

12 However, we can now speed up the system; therefore, more trains can operate.

13 As this is a demonstration of a real time business rail system, we shall need to operate far more trains without accident.

14 A model system must be able to perform the same operation requirements as that of a real business system

15 Searl through his model rail system must bring out of the world of fantasy into the world of reality that which sits in his mind.

16 Until he has modelled it, no one can really appreciate what he states possible.

17 Thus in 1968 he has to change much of his rail system, which is rather elementary stuff Dear John to something more complex, yet still logic in operation.

18 That calls for a lot of research for the components to be able to bring out of Searl's mind that which you cannot see. Therefore, you too can see what he states.

19 Searl appreciated that if he went digital in thinking and construction followed suit; he would be able to have multi-train control.

20 Up to 1968 Searl had just 14 locomotives operating over the same track.

21 One can appreciate that 14 trains running over the same track by the class of switching supplied at that time, is time consuming, and thus a slow process.

22 Not only children like model rail layouts, so doe's adults, maybe much more than children do.

PROF. JOHN ROY ROBERT SEARL

Author, Inventor, President, S.I.S.R.C.

Prof. Searl is the inventor of the **SEARL EFFECT GENERATOR, S.E.G.**, which is its own prime mover and generates electricity. He has written a series of books entitled “The Law of the Squares.”

Prof. Searl looks at things differently from most people in that he equates everything to “The Law of the Squares.”He is only interested in the Truth in all things. He has written his books to dispel all the miss-information and lies that have been written about him and his work.

He is setting the record straight. These books tell of his life experiences as well as the technology that he has developed. After reading these books, you may look at things differently.

He is currently President and Consultant Engineer of the ***Searl International Space Research Consortium (S.I.S.R.C.)*** Prof. Searl is truly trying to bring us into a better future through this method of thinking and the technology that it has produced. Sorry this is the end of chapter 17. May the power be with you always, take care; I am watching you watching me watching you the future looks bright turn up the light and be seen.

THE LAW OF THE SQUARES.

By Prof. John Roy Robert Searl.

The law of the squares states that all things within this universe have two prime states. They are opposites of which one will be greater in value than the other.

The law of the squares also states that there is a division between these two states.

Throughout my books you are shown hundreds of these samples. The one I shall discuss here is in reference to the animal termed Homo sapiens – that is correct, it represents you and me.

Let us look at the law of the squares statements:

1 There will be the non experts that representing the higher value.
2 There will be experts that representing the lower value.
3 Between these two states is a very small dividing line of inventors.

Thus the law of the squares is up held.

4 The non experts live in the today, not the yesterday nor the tomorrow.
5 The expert does not live in the today but rather in the 1,000 to 5 years of the past.
6 The inventor lives in the tomorrow, for he or she creates the future that is to be.

Without inventors there would be no tomorrows but a continuation of the today, thus never changing in any manner except for the change of seasons.

Inventors are labelled as living in a world of fantasy, while the so called experts claim to live in a world of reality; ironically, that very claim proves that their world came from the world of fantasy by those who suffered and struggled to create the present reality that these experts claim as their right.

Before any one can conceive a concept to generate some new invention a number of triggers have to be in place.
.

1. The brain must be free from all conventional brainwashing so that the sub-conscious can freely communicate to the conscious mind. Thus between these two states the data stored can be made active by the correct input of our body's sensors.

2 There must be information stored in the brain either at the time of its conception or gained after birth at the expense of time.

3 Once the correct triggers have been stimulated so that the conscious or sub-conscious mind can start searching and assembling a dream or a thought wave, data makes its way to the conscious mind of the owner and impresses the owner with the possibility of creating a new project.

4 It is extremely unlikely that you can invent because the brain that creates does so based on information that was invented already by brains of the past; even as far back as the day man first appeared upon planet earth.

5 My work is based on the success of hundreds of inventors before me; my work has just extended their work. They invented the materials, tools and equipment that make it possible for me to extend their work.

6 Within the laws of nature there is nothing impossible except that the state of your mind makes it so. And it is your mind which has and still does delay progress for a better world then we have today.

7 In my case; the sub-conscience assembled an extremely simple game which strangely contains the mathematics which is required to achieve this technology, without that input there is no S.E.G. or I.G.V. Anyone making such claims is talking out their rear end, what a shame to waste such energy.

8 It also gives much information on science and technology, which was a blessing as I had no need to go to school to learn it.

9 The device is basically a linear motor riding on a magnetic bearing; in reality there are three such units in one body.

10 Its characteristics are similar to that of an autotransformer; again three such devices in one body.

11 From the inner unit which generates its own output that feeds into the second unit that is also generating its own, but much greater output, is therefore boosted vastly by the input from the first unit. The second unit output is then feed into the third unit which also is generating a much larger output is again vastly increased by the input from the second unit.

12 Every operation of the S.E.G. follows Newton's laws precisely, these experts' claims that it is not possible are just protecting their image of importance. Such has always been the case since this animal called Homo sapiens arrive upon the planet earth and will always exist until the end of time.

13 Not only is the problem to design and test the specialised equipment to undertake this research; but one of the real major problems is that there are far too many of this species that suffer from an obsessive compulsive disorder that functions on ignorance and greed. There is no more dangerous combination that has so far managed to delay this technology from reaching its marketplace, and would had done so back in 1968 and again in 2003 if it was not for the effects relating to this disease.

14 Today for the fifth time efforts to recreate this technology are moving forward but cost is the greatest problem which increases each year that it takes to reach the marketplace. There is yet another major problem for delay. This is that the man involved has to live and therefore has to maintain a job by which he can survive and find from his pocket that cost which is involved. Time is related to availability of hard cash and spare time to work on it. It also requires absolute dedication and devotion and faith for its success and this man has all three requirements for such a success.

15 This world is dying from pollution factors which this technology offers a solution that no other technology can offer. Yet everyone ignores its possibilities because the power man of today would loose profit by such a technology being available on the marketplace.

16 Yet it would slowly halt this hurrican forming problem now greatly being experienced along the U.S.A. coast line, which will eventually reach places like Canada and even the UK if we allow this global warming to increase.

17 Is it humane for life and property to be destroyed in the name of ignorance and greed?

18 You may wonder if the S.E.G. output is Direct Current or Alternating Current. It is an Alternating Current at very high voltage that has to be passed through a step down transformer for the voltage required like all other power systems available today; current is proportional to the voltage.

19 Rare earths play an important part in the success of the S.E.G. operation. The one that has become the keystone of my work is Nd because everyone wants to follow the same one that I used in 1946. It is far from the best but it is the first choice for cost and availability.

20 One major problem is that it looses electrons to the air. As such it will oxidize in air. This can be prevented by the following means:

1. Keep it in argon while not in use which is the better choice; or in oil which is the choice that I do not prefer.
2. Or mix 1% of another rare earth which does not loose electrons to the air that will block the Nd from loosing its electrons to the air.
3. As you move up the element chart there are better options to use. Also it makes the S.E.G. smaller for the same output.

21 To experts it's impossible as they had stated about the cycle, motor car, Train, Plane, Electricity, gas, telephone, and television and space travel plus everything else made or done by man, in reality they were confirming that it could be done, and it was only done by those who lived in the world of fantasy, for it is only such people who are capable of free thinking by which progress in advancing man's knowledge can be achieved.

22 Everything made by man came out of the world of fantasy which existed there before man ever appeared on earth including all inventions of the future have been here since time began. But before any such invention can be brought forth from the world of fantasy into the world of reality some ingenious person must tune in to it; that action requires a set of triggers in a precise order of events and then and only then will such products become available to mankind.

23 Though the S.E.G. is a marvellous device it can only reduce the pollution problem which helps to reduce global warming that will give trees and animals a chance to survive. It will require everyone to help to clean up this planet as we made the mess due to our lack of knowledge and understanding and replant trees which in our ignorance have been cut down. These are urgent requirements to block the creation of hurricanes and tornados.

24 It takes about 30 years before a tree can fully function in its duty cycle in reducing this global warming, clearly the need is present today to start a massive replacement task in the effort to balance nature to our advantage, otherwise it will be to our disadvantage.

25 If you really care for your children's health and well fare then you must act today and help clean up these problems, To assist me to do my best to help nature to recover from our wrongs which we have done and are still doing.

26 Listen to our children who have no tomorrows if we don't care; our ignorance and greed is killing this planet. If we don't stop our destruction of this planet there will be no tomorrows for anyone. There will be just a resting place for our remains.

AS-93JS
113 pgs.

ANTIGRAVITY:

The Dream Made Reality

The Story of John R. R. Searl

By: John A. Thomas Jr.

Copyright © 1993 by John A. Thomas Jr.

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without permission in writing from the publisher. Inquiries should be addressed to John A. Thomas Jr., 373 Rock Beach Road, Rochester, New York, 14617-1316.

ISBN 1 898827 99 0
Issue Date November 11, 1993

Published By: Direct International Science Consortium
13 Blackburn, Low Strand, Grahame Park Estate, London, NW95NG, England
373 Rock Beach Road, Rochester, New York, 14617-1316, U.S.A.

ANTIGRAVITY: THE DREAM MADE REALITY
THE STORY OF JOHN R. R. SEARL
BY: JOHN A. THOMAS, JR.

This book contains information that was received by me from John R. R. Searl about his life and inventions. Searl has had an antigravity device since 1946. He also developed a generator of electricity that uses no input power as we know it.

There has been much written about Searl and his work but most of it is misinformation or guesswork. Many people have tried to duplicate his work but none have succeeded. All information in this book is correct and directly from the source, Prof. John R. R. Searl himself. This book is the story of a man who is trying to move the world into the future... and has been for years.

Published By:

John A. Thomas Jr.
373 Rock Beach Rd.
Rochester, New York
14617-1316
Phone: (716) 467-2694
Fax: (716) 338-2663

John A. Thomas Jr.
Author

Mr. Thomas has a technical background in electricity and process control. He has studied such varied topics as UFO's and Ancient Technology that has been lost or forgotten.

He has carried on the efforts of Mr. William Sherwood, formerly Professor Searl's contact and help mate in the National Space Research Consortium (NSRC).

He first contacted Professor Searl in 1990 and since that time has printed and distributed his books on "The Law of the Squares".

John has been a student of Professor Searl and has come to an understanding of "the square technology" which is the basis for the Searl Effect Generator (SEG) and the Inverse-G-Vehicle (IGV). This book is a reflection of that learning process.

John is presently Vice President of the Direct International Science Consortium (DISC) and is Technical Representative for Professor Searl.

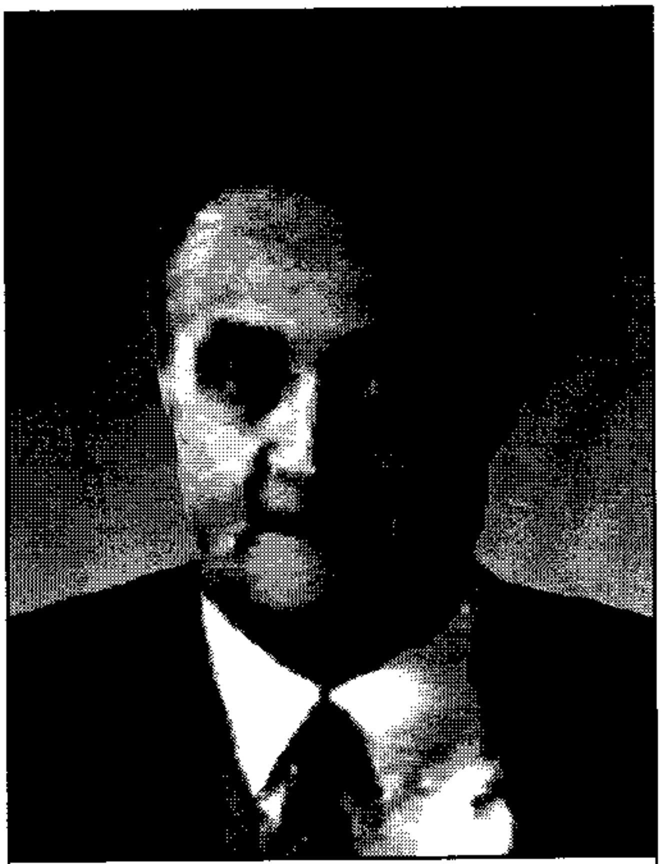

Prof. John R. R. Searl
Inventor

When I first called Prof. Searl I was apprehensive as to how I would be received. I was delighted to find him very receptive and eager to talk about his technology. We became good friends in the years to follow. I have never met such a sincere and honest individual.

Prof. Searl looks at things differently from most people in that he equates everything to "The Law of the Squares". He is only interested in the Truth in all things. He has written his books to dispel all the mis-information and lies that have been written about him and his work. He is setting the record straight. These books tell of his life experiences as well as the technology that he has developed. After reading these books you may look at things differently, as I do now.

He is currently the President and Consultant Engineer of the Direct International Science Consortium (DISC).

Prof. Searl is truly trying to bring us into a better future through this method of thinking and the technology that it has produced.

INDEX

CHAPTER 1:
Birth and Childhood
First Dreams and How They Were Received
First Jobs

CHAPTER 2:
The Law of the Squares

CHAPTER 3:
First SEG Made
SEG's Made
First Units That Went Airborne
First Magnets Made
Description of SEG Plates And Rollers
SEG is a Prime Mover
Description of Electron Action in SEG
Searl's Flying Discs
Inertia and Gravity Inside the Flying Craft
What Makes the Inverse-G-Vehicle Fly?
Application of Theory

CHAPTER 4:
Photos, Articles and Television Coverage
Demonstrations of the SEG
The Searl Effect
First SEG Electric Generator
First Generator Trial
Domestic Power Plant
Flying Discs
Starter Generator For P-11
DEMO 1
U.S. Demonstration of Flying Craft
Flying Craft Size and Use
Publicity
Patent Taken Out
Explosion
Uses for the SEG

CHAPTER 5:
NSRSC
Arrest and Imprisonment
Court Battle
Life After Prison
Family Breakup
Thumped the Power Company & How SEG Works
Destruction of a Lifetimes Work

CHAPTER 6:
The Road Back
Meeting Susan
Starting Over on Research-date
Writing Books

CHAPTER 7:
The Future of Man
The Pole Shift
Meteorite Coming
Government & Big Business Blocking Progress
Do We Really Want Progress?
Prof. John R.R. Searl's Books

AS-93 JS1

THE LAW
OF THE
SQUARES

TRUTH TABLE
THE SECRET OF THE SEARL TECHNOLOGY

By Prof. J.R.R. Searl, M. Inst. Pi.

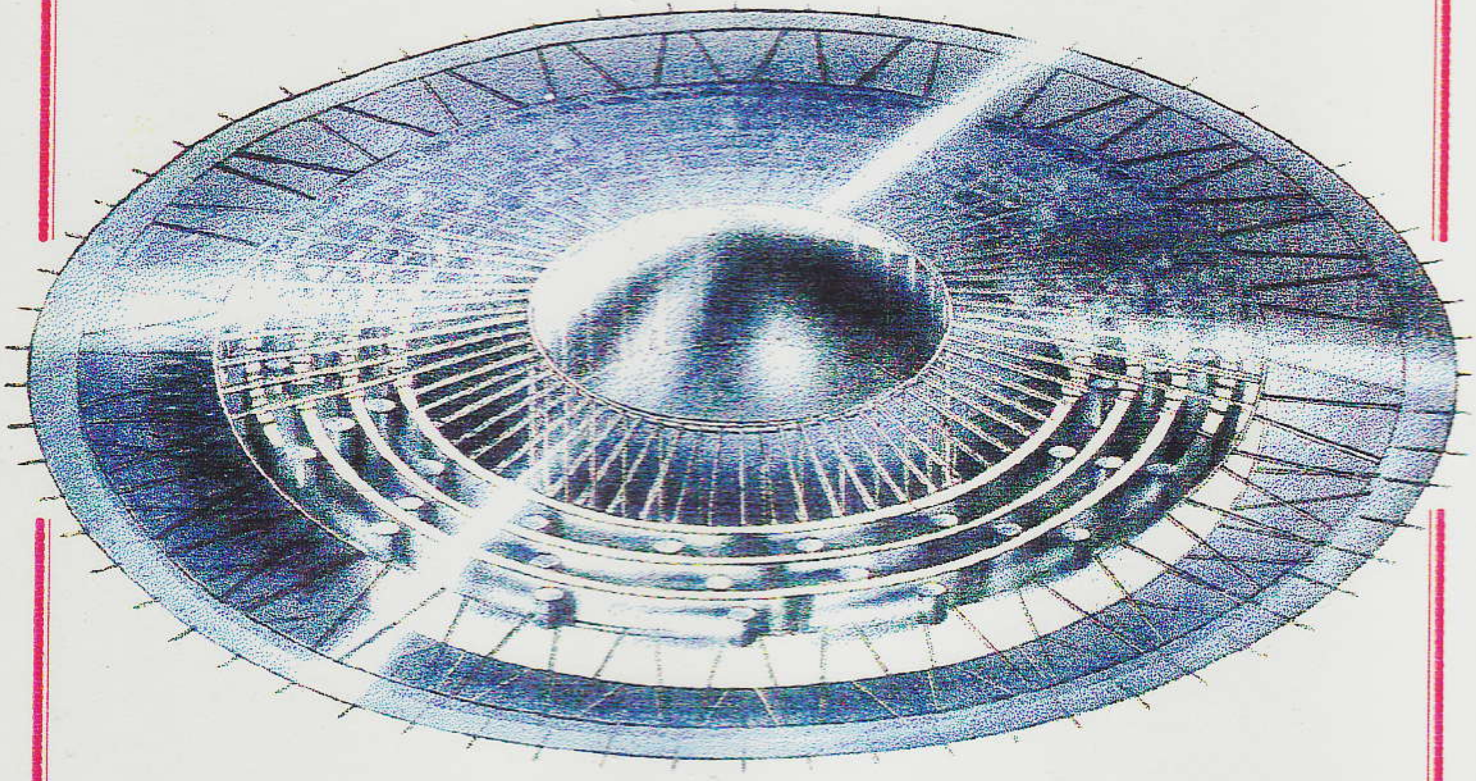

TOMORROW'S TRANSPORT SYSTEM

Book 1

CONTENTS.

COPYRIGHT PAGE		
ORIGINAL FRONT COVER		
LETTER	GUINNESS BOOK OF RECORDS	
LETTER	THE C.W.DANIEL COMPANY LTD	
ORIGINAL INSIDE COVER		
EDITOR'S INTRODUCTION REPORT	ABOUT THE AUTHOR	
GLOSSARY		
BOOK ONE	UNDERSTANDING THE SQUARES	1
THE LAW OF THE SQUARES	TRUTH TABLE - THE NO GO'S - 1 TO 4619	
THE LAW OF THE SQUARES	TRUTH TABLE - WILL GO - 1 TO 4619	23
DISCUSSION	BY SEARL	82
FACTS	NATURE AND THE LAW OF THE SQUARES	83
VALUES WHICH WILL NOT SHARE WITH ANY OTHER SQUARE		88
VALUES WHICH 3^2 WILL SHARE WITH 4^2		89
VALUES WHICH 3^2 WILL SHARE WITH 5^2		90
VALUES WHICH 3^2 WILL SHARE WITH 4^2 AND 5^2		90
VALUES WHICH 3^2 WILL SHARE WITH 6^2		90
VALUES WHICH 3^2 WILL SHARE WITH 5^2 AND 6^2		96
VALUES WHICH 3^2 WILL SHARE WITH 7^2		96
VALUES WHICH 3^2 WILL SHARE WITH 4^2 AND 7^2		96
VALUES WHICH 3^2 WILL SHARE WITH 5^2 AND 7^2		97
VALUES WHICH 3^2 WILL SHARE WITH 6^2 AND 7^2		97
VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2 AND 7^2		97
VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2 AND 7^2		97
VALUES WHICH 4^2 WILL NOT SHARE WITH ANY OTHER SQUARE		99
VALUES WHICH 5^2 WILL NOT SHARE WITH ANY OTHER SQUARE		102
VALUES WHICH 7^2 WILL NOT SHARE WITH ANY OTHER SQUARE		105
VALUES WHICH 8^2 WILL NOT SHARE WITH ANY OTHER SQUARE		106
VALUES WHICH 11^2 WILL NOT SHARE WITH ANY OTHER SQUARE		111
VALUES WHICH 13^2 WILL NOT SHARE WITH ANY OTHER SQUARE		114
VALUES WHICH 16^2 WILL NOT SHARE WITH ANY OTHER SQUARE		119

VALUES WHICH 17² WILL NOT SHARE WITH ANY OTHER SQUARE 120

VALUES WHICH 19² WILL NOT SHARE WITH ANY OTHER SQUARE 125

I have now started to release all past reports, statements made by me since 1946 up to the present time, this includes the data released within newsletters upon my work. Lectures etc.

These books contain the story of my life and those who were or are involved, they show letters connected with me upon this matter. Note that I alone can hold the copyright upon my life, what I state or do, no one else can do so.

I have given the rights to RAUM & ZEIT and NEXUS to write upon my life or work as and when they feel the readers would be interested in such matter.

I have accepted Mr.John A.Thomas Jr, as my represntive, and he has been given the rights to write any article he wishes upon this matter in the effort to push this technology right to the market place.

Now I am quite aware that many people since 1984 have written upon my work but have never paid me a penny for that right and many have not even asked me for the right to write upon my work. Slowly I have received copies of their articles from good people who have brought the matter to my notice. Many of these article I do not know who have written them, most contain details which are from the world of dreams. Therefore they wrote without my consent and rubbish, it is my duty to all those thousands who have helped me to correct this position NOW, that is precisely what I am doing.

If any one is stupid enough to make claims that I have copied they rights they must be prepared to be torn to pieces in a manner that will make newspapers have a hayday. I am a tiger, and boy, if I get you in court, I will give the court the best laugh of their life. I will rip you to pieces, I will bounce you so hard that you will never stop stuttering for a very long time.

I have so far shut my eyes to those writers who have written upon my work or claims, but the time has arrive that I must stand up and be counted, that is my humman right, now at last I am taking that right to be heard as a FACT, now I shall clean up the shit that has been thrown around the world. A number of you want like it, but that is precisely what I intended it is time that you received some punishment for your wrongs.

Any letters from you will appear in the next book released with whatever action that I have taken in relation to that letter. So it is far wiser that you don't send any such stupid letter to me. I mean to expose those who make statements with intent to turn people off from helping me. You should not be that stupid, so don't cry to others if you see that I am making fun of you, you asked for it, and I am certain that I will give it to you.

While I write and speak the truth, no man will stop me, that is a FACT, a point which you should accept immediately. If you think that I will not hear about the bullshit you are passing around, you are very much wrong it will come to me, that is a fact, and to prove, it will appear within these books for all to see that I do know about it.

The information contain within these books are precise at the date which it is listed, all information released has been released by me unless its an article by some other person upon me, whose name is quoted, or person unknown at this time. If their name become known to me I will quote it in the next book released. No doubt some one will know who wrote that article. Any article written about me will appear within these books, because that happens to be my copyright, my life not yours. You have not paid me any value for that right. I don't care if you think you got right, you have not, and in court I shall demand a heavy price for failure to secure that right. Therefore, unless you have a lot money available to pay me, just shut up, and keep quite, so I don't hear about it.

I have no intentions to create trouble for any one, unless they go out their way looking for it. My whole intentions, as it has always been, is to try to unite all those who work upon ideas of cleaner energy systems, no lecture have I given where I have not pressed that point calling all inventors to group together, to work together to success. The choice is yours.

These books are released upon the authority of

Prof J.R.R. Searl

PROF. JOHN ROY ROBERT SEARL
CONSULTANT ENGINEER.
DATED 16.07.1993

ISBN 1 898827 00 1

AS-93JS1A
161 pgs.

The Law of the Squares

By Prof. John R. R. Searl

Book 1A

Copyright © 1993 by Prof. John R. R. Searl

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without permission in writing from the publisher. Inquiries should be addressed to John A. Thomas Jr., 373 Rock Beach Road, Rochester, New York, 14617-1316.

ISBN 1 898827 05 2
Issue Date July 25, 1993

Published By: Direct International Science Consortium
13 Blackburn, Low Strand, Grahame Park Estate, London, NW95NG, England
373 Rock Beach Road, Rochester, New York, 14617-1316, U.S.A.

COPYRIGHT PAGE

ORIGINAL FRONT COVER

FORWARD I

LETTER MIRROR PUBLICATIONS

DISCUSSION MY ME 1

SEARL DESIGNED HOME 6

LETTER ATTENTION FARMERS DISCUSSION 7

EXTEND MY DISCUSSION OF BOOK 1 10

FACT DISCUSSION SCIENTIFIC 14

TELEPHONE CONVERSATION BETWEEN THE SOLICITOR FOR DAVID STRILING AND GUNNER SANDBERG 16

MY DISCUSSION UPON THAT TELEPHONE CONVERSATION 18a

S.E.G. DESIGN DRAWN BY DAVID STIRLING COMPUTER PROGRAMMER 19

S.E.G. COMPUTER CALCULATIONS DONE BY PAUL H.HUDZ, COMPUTONE SYSTEMS 06.12.1986. 20

S.E.G. COMPUTER CALCULATIONS DONE BY PAUL H.HUDZ, COMPUTONE SYSTEMS 06.12.1986. 21

S.E.G. COMPUTER CALCULATIONS DONE BY PAUL H.HUDZ, COMPUTONE SYSTEMS 08.12.1988. 22

ONE OF THE FAMOUS LETTER WHICH INVENTORS CAN GET 23

MY REPLY TO THAT LETTER 24

TABLE GROUP 1's SHELLS VALUE CONTENTS IN BOOK 1. 26

TABLE GROUP 2's SHELLS VALUE CONTENTS IN BOOK 1. 27

POEM WHICH I HAD ONLY HEARD ONCE IN MY CHILDHOOD DAYS. WHICH EXPLAINS ME TO A TREAT 30

TABLE SQUARES ARE FORM BY SQUARES IN BOOK 1. 31

TABLE CORNER VALUES OF EACH SQUARE BOOK 1. 32

TABLE VALUES OF BUILDING BLOCKS BOOK 1. 32

VALUES WHICH 3^2 WILL NOT SHARE WITH ANY OTHER SQUARES 33

VALUES WHICH 4^2 WILL NOT SHARE WITH ANY OTHER SQUARES 33

VALUES WHICH 5^2 WILL NOT SHARE WITH ANY OTHER SQUARES 35

VALUES WHICH 7^2 WILL NOT SHARE WITH ANY OTHER SQUARES 35

VALUES WHICH 8^2 WILL NOT SHARE WITH ANY OTHER SQUARES 36

VALUES WHICH 11^2 WILL NOT SHARE WITH ANY OTHER SQUARES 37

VALUES WHICH 13^2 WILL NOT SHARE WITH ANY OTHER SQUARES 37

VALUES WHICH 16^2 WILL NOT SHARE WITH ANY OTHER SQUARE 38

VALUES WHICH 17^2 WILL NOT SHARE WITH ANY OTHER SQUARE 38

VALUES WHICH 19^2 WILL NOT SHARE WITH ANY OTHER SQUARE 38

VALUES WHICH 3^2 WILL SHARE WITH 4^2 39

VALUES WHICH 3^2 WILL SHARE WITH 5^2 39

VALUES WHICH 3^2 WILL SHARE WITH 6^2 39

VALUES WHICH 3^2 WILL SHARE WITH 7^2 42

VALUES WHICH 3^2 WILL SHARE WITH 8^2 42

VALUES WHICH 3^2 WILL SHARE WITH 9^2 43

VALUES WHICH 3^2 WILL SHARE WITH 11^2 43

VALUES WHICH 3^2 WILL SHARE WITH 13^2 43

VALUES WHICH 3^2 WILL SHARE WITH 16^2 43

VALUES WHICH 3^2 WILL SHARE WITH 17^2 43

VALUES WHICH 3^2 WILL SHARE WITH 19^2 43

VALUES WHICH 4^2 WILL SHARE WITH 5^2 44

VALUES WHICH 4^2 WILL SHARE WITH 7^2 46

VALUES WHICH 4^2 WILL SHARE WITH 11^2 46

VALUES WHICH 4^2 WILL SHARE WITH 13^2 47

VALUES WHICH 4^2 WILL SHARE WITH 17^2 47

VALUES WHICH 4^2 WILL SHARE WITH 19^2 48

VALUES WHICH 5^2 WILL SHARE WITH 7^2 48

VALUES WHICH 5^2 WILL SHARE WITH 8^2 48

VALUES WHICH 5^2 WILL SHARE WITH 10^2 49

VALUES WHICH 5^2 WILL SHARE WITH 11^2 51

VALUES WHICH 5^2 WILL SHARE WITH 13^2 51

VALUES WHICH 5^2 WILL SHARE WITH 16^2 51

VALUES WHICH 5^2 WILL SHARE WITH 17^2 51

VALUES WHICH 7^2 WILL SHARE WITH 8^2 52

VALUES WHICH 7^2 WILL SHARE WITH 11^2 53

VALUES WHICH 7^2 WILL SHARE WITH 13^2 53

VALUES WHICH 7^2 WILL SHARE WITH 14^2 53

VALUES WHICH 7^2 WILL SHARE WITH 16^2 54

VALUES WHICH 7^2 WILL SHARE WITH 17^2 54

VALUES WHICH 7^2 WILL SHARE WITH 19^2 54

VALUES WHICH 8^2 WILL SHARE WITH 11^2 55

VALUES WHICH 8^2 WILL SHARE WITH 13^2 55

VALUES WHICH 8^2 WILL SHARE WITH 17^2 55

VALUES WHICH 8^2 WILL SHARE WITH 19^2 55

VALUES WHICH 11^2 WILL SHARE WITH 13^2 57

VALUES WHICH 11^2 WILL SHARE WITH 16^2 57

VALUES WHICH 11^2 WILL SHARE WITH 17^2 57

VALUES WHICH 11^2 WILL SHARE WITH 19^2 57

VALUES WHICH 13^2 WILL SHARE WITH 16^2 58

VALUES WHICH 13^2 WILL SHARE WITH 17^2 58

VALUES WHICH 13^2 WILL SHARE WITH 19^2 58

OPTION TABLE FOR SQUARES LISTED IN BOOK 1. 59

VALUES WHICH 16^2 WILL SHARE WITH 17^2 59

VALUES WHICH 16^2 WILL SHARE WITH 19^2 59

VALUES WHICH 3^2 WILL SHARE WITH 4^2 AND 5^2 60

VALUES WHICH 3^2 WILL SHARE WITH 4^2 AND 7^2 60

VALUES WHICH 3^2 WILL SHARE WITH 4^2 AND 9^2 60

VALUES WHICH 3^2 WILL SHARE WITH 4^2 AND 11^2 60

VALUES WHICH 3^2 WILL SHARE WITH 4^2 AND 12^2 60

GREEN HOUSE EFFECT 62

VALUES WHICH 3^2 WILL SHARE WITH 5^2 AND 6^2 67

VALUES WHICH 3^2 WILL SHARE WITH 5^2 AND 7^2 67

VALUES WHICH 3^2 WILL SHARE WITH 5^2 AND 8^2 67

VALUES WHICH 3^2 WILL SHARE WITH 5^2 AND 9^2 67

VALUES WHICH 3^2 WILL SHARE WITH 5^2 AND 13^2 67

VALUES WHICH 3^2 WILL SHARE WITH 5^2 AND 15^2 67

VALUES WHICH 3^2 WILL SHARE WITH 6^2 AND 7^2 68

VALUES WHICH 3^2 WILL SHARE WITH 6^2 AND 9^2 68

VALUES WHICH 3^2 WILL SHARE WITH 6^2 AND 11^2 69

VALUES WHICH 3^2 WILL SHARE WITH 6^2 AND 13^2 69

VALUES WHICH 3^2 WILL SHARE WITH 6^2 AND 17^2 69

VALUES WHICH 3^2 WILL SHARE WITH 6^2 AND 19^2 69

VALUES WHICH 3^2 WILL SHARE WITH 7^2 AND 8^2 70

VALUES WHICH 3^2 WILL SHARE WITH 7^2 AND 9^2 70

VALUES WHICH 3^2 WILL SHARE WITH 7^2 AND 11^2 70

VALUES WHICH 3^2 WILL SHARE WITH 7^2 AND 13^2 70

VALUES WHICH 3^2 WILL SHARE WITH 7^2 AND 16^2 70

VALUES WHICH 3^2 WILL SHARE WITH 8^2 AND 9^2 71

VALUES WHICH 3^2 WILL SHARE WITH 8^2 AND 11^2 71

VALUES WHICH 3^2 WILL SHARE WITH 8^2 AND 13^2 71

VALUES WHICH 3^2 WILL SHARE WITH 8^2 AND 17^2 71

VALUES WHICH 3^2 WILL SHARE WITH 8^2 AND 19^2 71

VALUES WHICH 3^2 WILL SHARE WITH 9^2 AND 11^2 72

VALUES WHICH 3^2 WILL SHARE WITH 9^2 AND 13^2 72

VALUES WHICH 3^2 WILL SHARE WITH 9^2 AND 16^2 72

VALUES WHICH 3^2 WILL SHARE WITH 9^2 AND 17^2 72

AN INVENTOR SHOULD ASK 'WILL THE GOVERNMENT HELP ?' 72

VALUES WHICH 3^2 WILL SHARE WITH 11^2 AND 16^2 76

VALUES WHICH 3^2 WILL SHARE WITH 13^2 AND 16^2 76

SHAPE MEMORY MATERIALS 77

VALUES WHICH 4^2 WILL SHARE WITH 5^2 AND 7^2 80

VALUES WHICH 4^2 WILL SHARE WITH 5^2 AND 11^2 80

VALUES WHICH 4^2 WILL SHARE WITH 5^2 AND 13^2 80

VALUES WHICH 4^2 WILL SHARE WITH 5^2 AND 17^2 80

VALUES WHICH 4^2 WILL SHARE WITH 5^2 AND 19^2 80

VALUES WHICH 4^2 WILL SHARE WITH 5^2 AND 20^2 80

VALUES WHICH 4^2 WILL SHARE WITH 7^2 AND 11^2 81

VALUES WHICH 4^2 WILL SHARE WITH 7^2 AND 13^2 81

VALUES WHICH 4^2 WILL SHARE WITH 7^2 AND 17^2 81

VALUES WHICH 4^2 WILL SHARE WITH 11^2 AND 13^2 82

VALUES WHICH 4^2 WILL SHARE WITH 11^2 AND 17^2 82

VALUES WHICH 4^2 WILL SHARE WITH 11^2 AND 19^2 82

VALUES WHICH 5^2 WILL SHARE WITH 7^2 AND 8^2 84

VALUES WHICH 5^2 WILL SHARE WITH 7^2 AND 10^2 84

VALUES WHICH 5^2 WILL SHARE WITH 8^2 AND 11^2 84

VALUES WHICH 5^2 WILL SHARE WITH 8^2 AND 13^2 84

VALUES WHICH 5^2 WILL SHARE WITH 10^2 AND 11^2 85

VALUES WHICH 5^2 WILL SHARE WITH 10^2 AND 13^2 85

VALUES WHICH 5^2 WILL SHARE WITH 10^2 AND 17^2 85

VALUES WHICH 5^2 WILL SHARE WITH 10^2 AND 19^2 85

VALUES WHICH 5^2 WILL SHARE WITH 11^2 AND 16^2 86

VALUES WHICH 5^2 WILL SHARE WITH 13^2 AND 16^2 86

VALUES WHICH 5^2 WILL SHARE WITH 16^2 AND 19^2 87

VALUES WHICH 7^2 WILL SHARE WITH 8^2 AND 11^2 88

VALUES WHICH 7^2 WILL SHARE WITH 8^2 AND 13^2 88

VALUES WHICH 7^2 WILL SHARE WITH 8^2 AND 17^2 88

VALUES WHICH 7^2 WILL SHARE WITH 8^2 AND 19^2 88

VALUES WHICH 7^2 WILL SHARE WITH 11^2 AND 14^2 89

VALUES WHICH 7^2 WILL SHARE WITH 13^2 AND 14^2 89

VALUES WHICH 7^2 WILL SHARE WITH 14^2 AND 17^2 89

VALUES WHICH 7^2 WILL SHARE WITH 14^2 AND 19^2 89

SCIENCE IS STRANGER THAN FICTION LETTER BY DR.GEORGE WHITE 26.07.1973. 91

VALUES WHICH 11^2 WILL SHARE WITH 17^2 AND 19^2 93

VALUES WHICH 13^2 WILL SHARE WITH 17^2 AND 19^2 93

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2 AND 12^2 94

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 7^2 AND 9^2 94

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 7^2 AND 12^2 94

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 9^2 AND 12^2 95

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 11^2 AND 12^2 97

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 12^2 AND 13^2 97

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 12^2 AND 17^2 97

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 12^2 AND 19^2 98

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2 AND 7^2 98

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2 AND 9^2 98

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2 AND 10^2 98

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 7^2 AND 8^2 99

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 7^2 AND 15^2 99

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 8^2 AND 9^2 99

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 8^2 AND 11^2 99

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 8^2 AND 15^2 99

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 9^2 AND 15^2 100

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 11^2 AND 15^2 100

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 13^2 AND 15^2 100

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 15^2 AND 16^2 100

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 15^2 AND 17^2 100

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 15^2 AND 19^2 101

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 7^2 AND 9^2 101

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 7^2 AND 14^2 101

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 9^2 AND 11^2 102

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 9^2 AND 13^2 102

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 9^2 AND 17^2 102

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 9^2 AND 18^2 102

VALUES WHICH 3^2 WILL SHARE WITH 7^2, 8^2 AND 9^2 103

VALUES WHICH 3^2 WILL SHARE WITH 7^2, 8^2 AND 11^2 103

VALUES WHICH 3^2 WILL SHARE WITH 7^2, 8^2 AND 13^2 103

VALUES WHICH 3^2 WILL SHARE WITH 8^2, 9^2 AND 13^2 104

VALUES WHICH 3^2 WILL SHARE WITH 8^2, 11^2 AND 13^2 104

VALUES WHICH 3^2 WILL SHARE WITH 8^2, 13^2 AND 17^2 104

VALUES WHICH 3^2 WILL SHARE WITH 8^2, 17^2 AND 19^2 104

VALUES WHICH 3^2 WILL SHARE WITH 9^2, 11^2 AND 16^2 104

VALUES WHICH 3^2 WILL SHARE WITH 9^2, 13^2 AND 16^2 104

VALUES WHICH 3^2 WILL SHARE WITH 9^2, 16^2 AND 17^2 104

VALUES WHICH 4^2 WILL SHARE WITH 5^2, 11^2 AND 13^2 105

VALUES WHICH 4^2 WILL SHARE WITH 5^2, 11^2 AND 20^2 105

VALUES WHICH 4^2 WILL SHARE WITH 5^2, 13^2 AND 20^2 105

VALUES WHICH 4^2 WILL SHARE WITH 5^2, 17^2 AND 20^2 105

VALUES WHICH 4^2 WILL SHARE WITH 5^2, 19^2 AND 20^2 105

VALUES WHICH 4^2 WILL SHARE WITH 7^2, 11^2 AND 13^2 105

VALUES WHICH 5^2 WILL SHARE WITH 7^2, 10^2 AND 14^2 106

VALUES WHICH 5^2 WILL SHARE WITH 7^2, 11^2 AND 16^2 106

VALUES WHICH 5^2 WILL SHARE WITH 7^2, 13^2 AND 16^2 106

VALUES WHICH 5^2 WILL SHARE WITH 8^2, 11^2 AND 13^2 106

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 12^2 AND 15^2 107

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 7^2, 9^2 AND 12^2 107

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 7^2, 11^2 AND 12^2 107

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 9^2, 12^2 AND 17^2 108

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 7^2 AND 10^2 108

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 9^2 AND 10^2 108

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 10^2 AND 11^2 108

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 10^2 AND 15^2 108

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 7^2, 8^2 AND 9^2 109

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 7^2, 8^2 AND 15^2 109

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 7^2, 9^2 AND 14^2 110

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 7^2, 13^2 AND 14^2 110

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 9^2, 11^2 AND 13^2 110

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 9^2, 11^2 AND 18^2 110

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 9^2, 13^2 AND 18^2 110

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 9^2, 17^2 AND 18^2 111

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 9^2, 18^2 AND 19^2 111

VALUES WHICH 3^2 WILL SHARE WITH 7^2, 8^2, 9^2 AND 11^2 111

VALUES WHICH 3^2 WILL SHARE WITH 7^2, 8^2, 9^2 AND 13^2 111

VALUES WHICH 3^2 WILL SHARE WITH 7^2, 8^2, 9^2 AND 17^2 111

VALUES WHICH 4^2 WILL SHARE WITH 5^2, 7^2, 13^2 AND 20^2 111

MY ARTICLE ON THE IC555 112

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 9^2, 11^2 AND 12^2 114

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 9^2, 12^2 AND 13^2 114

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 9^2, 12^2 AND 15^2 114

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 11^2, 12^2 AND 15^2 114

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 12^2, 13^2 AND 15^2 114

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 12^2, 15^2 AND 17^2 114

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 12^2, 15^2 AND 20^2 115

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 7^2, 9^2, 11^2 AND 12^2 115

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 7^2, 9^2, 12^2 AND 13^2 115

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 9^2, 11^2, 12^2 AND 13^2 115

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 9^2, 11^2, 12^2 AND 17^2 115

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 9^2, 11^2, 12^2 AND 19^2 115

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 9^2, 12^2, 13^2 AND 17^2 115

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 9^2, 12^2, 13^2 AND 19^2 116

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 7^2, 8^2 AND 10^2 116

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 7^2, 10^2 AND 11^2 116

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 9^2, 10^2 AND 15^2 116

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 10^2, 11^2 AND 15^2 116

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 10^2, 13^2 AND 15^2 117

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 10^2, 15^2 AND 17^2 117

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 7^2, 8^2, 9^2 AND 15^2 117

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 7^2, 9^2, 15^2 AND 16^2 117

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 8^2, 9^2, 11^2 AND 15^2 117

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 8^2, 9^2, 13^2 AND 15^2 117

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 8^2, 9^2, 15^2 AND 17^2 117

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 8^2, 9^2, 15^2 AND 19^2 118

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 9^2, 11^2, 15^2 AND 16^2 118

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 7^2, 9^2, 11^2 AND 14^2 118

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 7^2, 9^2, 13^2 AND 14^2 118

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 7^2, 9^2, 14^2 AND 18^2 118

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 7^2, 11^2, 13^2 AND 14^2 118

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 7^2, 11^2, 14^2 AND 17^2 118

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 7^2, 11^2, 14^2 AND 19^2 119

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 9^2, 11^2, 13^2 AND 18^2 119

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 7^2, 9^2, 12^2 AND 15^2 119

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 7^2, 11^2, 12^2 AND 15^2 119

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 7^2, 12^2, 13^2 AND 15^2 119

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 7^2, 12^2, 15^2 AND 17^2 119

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 9^2, 11^2, 12^2 AND 15^2 119

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 9^2, 12^2, 13^2 AND 15^2 120

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 9^2, 12^2, 15^2 AND 20^2 120

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 7^2, 9^2, 10^2 AND 14^2 120

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 7^2, 10^2, 13^2 AND 14^2 120

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 7^2, 10^2, 14^2 AND 15^2 120

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 9^2, 10^2, 11^2 AND 15^2 120

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 9^2, 10^2, 13^2 AND 15^2 121

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 9^2, 10^2, 15^2 AND 18^2 121

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 10^2, 11^2, 13^2 AND 15^2 121

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 10^2, 11^2, 15^2 AND 17^2 121

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 10^2, 13^2, 15^2 AND 17^2 121

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 10^2, 13^2, 15^2 AND 19^2 121

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 7^2, 9^2, 14^2, 17^2 AND 18^2 122

VALUES WHICH 3^2 WILL SHARE WITH 6^2, 7^2, 9^2, 14^2, 18^2 AND 19^2 122

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 7^2, 9^2, 12^2, 15^2 AND 20^2 #122

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 7^2, 12^2, 15^2, 19^2 AND 20^2 #122

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 9^2, 12^2, 15^2, 17^2 AND 20^2 #122

VALUES WHICH 3^2 WILL SHARE WITH 4^2, 5^2, 11^2, 12^2, 13^2, 15^2 AND 20^2 #122

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 7^2, 9^2, 10^2, 14^2 AND 15^2 #122

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 9^2, 10^2, 11^2, 15^2 AND 18^2 #123

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 9^2, 10^2, 13^2, 15^2 AND 18^2 #123

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 9^2, 10^2, 15^2, 17^2 AND 18^2 #123

VALUES WHICH 3^2 WILL SHARE WITH 5^2, 6^2, 9^2, 10^2, 15^2, 18^2 AND 19^2 #123

DISCUSSION ON THE SQUARES 124

COMPANY ZANUSSI 137

MY KNOWLEDGE POWER SWITCHES 138

FLUORIDATION 142

TRIG CONVERTER FOR INDUCTANCE MEASUREMENT 142

LETTERS (2 OFF) GERMAN TV FILM TRANSMIT MAY 20TH 1971 144

LETTER LAKER AIRWAYS LIMITED 03.03.1981 145

LETTERS DUKE OF EDINBURGH 12.12.1963 146
UNIVERSITY OF LONDON 19.07.1973
TECHNOLOGY REPORTS CENTRE 09.07.1973

THE WHATELEY ROAD SOLAR PROJECT 147

Please note that the contents within this book is precisely correct to the best of my knowledge. The contents are related to my work, directly or indirectly. All information is vital to my future achievements. All statements are made by me unless it is stated by some other person/s.

My release of data started as newsletters and lectures from 1946, now all this information of the past, present and future, is being woven together into a book format and released. It deals with my life and those who are involved. My work, my education and knowledge and how I employ that knowledge.

My life cannot be copyrighted by anyone except myself. Therefore, I have appointed Mr.John A Thomas Jr. as my representative with premission to write articles upon my work in the effort to speed this work to the market place.

I am quite aware that people have written upon my life without my permission, who have not paid me one penny for that right. I have shut my eyes to that offense, and will continue to do so, unless they wish to make any rows about me quoting their article and pointing out what is wrong with it...THAT IS MY RIGHT...NOT YOURS.

Any such offensive letters being received will appear in my very next release with what ever action I have taken. I mean to clean the shit up that has gone out upon me, and I will. If any one is stupid enough to write me or others, statements of claims which are unfounded, watch out, I am a tiger and I will tear you to pieces before the public. It is my RIGHT and DUTY to all those who have helped me over the years, to stand up and be counted, and that is precisely what I am doing.

I have given authority to both **RAUM & ZEIT** and **NEXUS** to write upon my work whenever they feel that the public would be interested upon such matters.

I have given authority to societies who have given me time to lecture there, in the interest to bring more people to one place, to write upon my work in the manner to which they feel would interest their readers.

Today, to my knowledge, I am the only person, who within my lectures, plead for all inventors to join together as a group to develop tomorrow's technology. This way progress would be speeded up, thus the media will stop mocking you. But if you like being mocked, then just remain an individual.

This set of books have been released upon the authority of

Prof. J.R.R. Searl

PROF.JOHN ROY ROBERT SEARL
CONSULTANT ENGINEER
DATE.25.07.1993

ISBN 1 898827 05 2

A8-93JS1B
185 pgs.

The Law of the Squares

By Prof. John R. R. Searl

Book 1B

Copyright © 1993 by Prof. John R. R. Searl

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without permission in writing from the publisher.

Inquiries should be addressed to John A. Thomas Jr., 373 Rock Beach Road, Rochester, New York, 14617-1316.

ISBN 1 898827 10 9
Issue Date August 10, 1993

Published By: Direct International Science Consortium
13 Blackburn, Low Strand, Grahame Park Estate, London, NW95NG, England
373 Rock Beach Road, Rochester, New York, 14617-1316, U.S.A.

CONTENTS.

FRONT COVER SHOWS THE DRAWING BY PETER BRUNDT		
COPYRIGHT PAGE		
THE OLD FRONT COVER PAGE		
MY INTRODUCTION		i
GOOD NEWS	19.07.1990	iii
INTRODUCTION	MAIN ARTICLE	1
REST FRAMES	3^2 NOT SHARING - BOOK 1	2
REST FRAMES	4^2 NOT SHARING - BOOK 1	8
REST FRAMES	5^2 NOT SHARING - BOOK 1	13
DISCUSSION	KILLER BUGS	14
SEARL DESIGNED HOUSE	NO.2 DRAWING DONE BY A GERMAN LADY	16
BIOSENSOR		20
REST FRAMES	7^2 NOT SHARING - BOOK 1	23
LETTER	GUINNESS BOOK OF RECORDS 22.01.1988	25
LETTER	THE C.W.DANIEL 09.04.1990	26
LETTER SENT OUT WITH MY BOOK TO PUBLISHERS		28
LETTER	DOVER PUBLICATIONS,INC. 10.05.1990	29
LETTER	LETTER SENT TO PRINCE CHARLES BY PARTY OF 3. 04.05.1990	30
LETTER	REPLY PRINCE OF WALES 19.06.1990	32
LETTERS	REPLY DUKE OF EDINBURGH 25.09.1986 REPLY RT.HON.JOHN SMITH Q.C. M.P.	33
LETTERS	REPLY DUKE OF YORK 30.09.1986 REPLY DEPARTMENT OF TRANSPORT 08.10.1986	34
LETTER	REPLY PRINCE OF WALES 06.10.1986	35
LETTERS	REPLY THE QUEEN 21.06.1977 REPLY PRESIDENT CARTER 13.09.1977	36
LETTERS	REPLY MALCOLM BRUCE M.P. 01.10.1986 REPLY ROBERT MACLENNAN M.P. 02.02.1987	37
LETTERS	REPLY STEPHEN ROSS M.P. 01.10.1986 REPLY JOHN PRESCOTT M.P. 22.10.1986	38
LETTERS	REPLY RT.HON.STAN ORME M.P. 17.10.1986 REPLY RT.HON.DAVID STEEL M.P. 17.10.1986	39
LETTER	REPLY MISS S.KNIGHT AROUND OCT.1986	41
LETTER	REPLY FRIENDS OF THE EARTH 25.05.1988	42

CONTENTS.

REST FRAMES 8^2 NOT SHARING - BOOK 1 43

DISCUSSION VARIOUS POINTS 46

BIBLE FACTS LIFE TABLE IN ONE YEAR PERIOD 50

FACT NOAH ECT, THE TRUTH 51

S.E.G. FACT DRAWING BY PETER BRUNDT 53

S.E.G. MATERIALS & EQUIPMENT APRIL 4th, 1989 - DAVID STERLING - CANARY ISLANDS 54

PERIODIC TABLE OF THE ELEMENTS 55

PERIODIC PROPERTIES OF THE ELEMENTS 56

DATA OF THE S.E.G. 31.12.1988 BY PAUL HUDZ WHILE I WAS IN CALIFORNIA 57

DATA OF THE S.E.G. 31.12.1988 BY PAUL HUDZ WHILE I WAS IN CALIFORNIA 58

DATA OF THE S.E.G. 31.12.1988 BY PAUL HUDZ WHILE I WAS IN CALIFORNIA 59

DATA OF THE S.E.G. 31.12.1988 BY PAUL HUDZ WHILE I WAS IN CALIFORNIA 60

RARE EARTH MAGNETS PRODUCED BY POWDER METALLURGY MAINTAINS STEADY GROWTH 61

CHART OF THE ELEMENTS 62

FIELDBUS: PROGRESS TOWARDS THE INTERNATIONAL STANDARD 63

SEARL EFFECT GENERATION LTD - DAVID STERLING 07.04.1989 -
WHILE I WAS IN THE CANARY ISLANDS - CONTRACT 64

PICTURE OF THE INVERSE-G-VEHICLE ONE OF A NUMBER DONE BY TREVOR ROBERTS - 1968 76

PRESS ADVERT BY VERA COMER. NEW ZEALAND ABOUT 1972 ? 77
"FLYING SAUCERS" ARE REAL !

PRESS WRITE UP BY VERA COMER. NEW ZEALAND ABOUT 1972 ?
LOOK OUT,THE DISCS ARE COMING 78

LETTERS REPLY INVENTIONS UNIT MINISTRY OF DEFENCE 19.05.1972 79
REPLY INVENTIONS UNIT MINISTRY OF DEFENCE 17.05.1972

LETTER REPLY DEPARTMENT OF AVIATION BANGKOK 12. 25.05.1978 80

LETTER REPLY DEPARTMENT OF CIVIL AVIATION SINGAPORE 04.03.1978 81

LETTER REPLY CIVIL AVIATION DEPARTMENT HONG KONG -21.06.1978 82

LETTER REPLY DEPARTMENT OF INDUSTRY LONDON 14.06.1978 83

HOW CORROSION OCCURS 84

ARTICLE RELEASED BY ME SOME YEARS AGO UPON THE S.E.G. 85

DISCUSSION ON MAGNETISM A VERY OLD REPORT WHICH WAS RELEASED BY ME YEARS AGO. 89

MY MEDICAL KNOWLEDGE 1968 QUESTIONS AND ANSWERS 94

REST FRAMES 11^2 NOT SHARING - BOOK 1 96

DISCUSSION EC RESEARCH PROGRAM 100
JULY 1990 NEW GLASS FIBRE MATERIAL

REST FRAME 13^2 NOT SHARING - BOOK 1 101

MARCH 1990 NATIONAL POWER AND POWERGEN GAS 102

SOLAR HEATING FREE ENERGY 103

REST FRAME 16^2 NOT SHARING - BOOK 1 107
FACT 01.03.1990 BA COMPANIES

SURGE ARRESTORS PROTECTION WITHOUT RADIOACTIVITY 109

REST FRAME 17^2 NOT SHARING - BOOK 1 112

REST FRAME 19^2 NOT SHARING - BOOK 1 113
FACT 26.07.1990 NATIONAL POWER

DATA AlGainP LASER DIODE 114

MAN-MACHINE COMMUNICATIONS STUDIES 117

COMPUTER DISCUSSION 128

26.07.1990 WATER DISCUSSION 123

TOTAL SQUARES SHARING REST FRAME IN BOOK 1 124

POWER SUPPLIES DISCUSSION 133

KEEPING FIT DISCUSSION 134

GIVE UP SMOKING DISCUSSION 135
THE CHANNEL TUNNEL - RISK OF FIRE DISCUSSION

LITHIUM ALLOY IRON SULPHIDE BATTERY DISCUSSION 136

REPRODUCTION OF DISCUSSION IN NEWSLETTER NO.39. DATE 21.10.1987 137

SPACECRAFT POWER DISCUSSION 139

VOLTAIC CELLS DISCUSSION 157

LETTER REPLY OMNI PUBLICATIONS INTERNATIONAL LTD. 17.07.1990 161

LETTER MY LETTER TO THE BANK OF ENGLAND 26.07.1990 162

LETTER REPLY BANK OF ENGLAND 02.08.1990 163

RELATIVE MOTION DISCUSSION 164

LIST OF THE ELEMENTS AS I KNEW THEM IN 1946 175

LETTER REPLY ROYAL MINT 10.08.1990 176

DATA STARSHIP EZEKIEL MK.V. RELEASED DATE 01.01.1984 UPDATE 177

02.08.1990 EVENTS OF THE GULF DISCUSSION 179

02.08.1990 COMPANY TAKES OVER ANOTHER COMPANY 180

02.08.1990 RUMOURS. HANSON'S INTEREST BUYING POWERGEN 180
MY MEDICAL KNOWLEDGE ON HEALTH

DATA INFRA-RED DETECTORS 181
FACT 02.08.1912 GREAT EASTERN JUBILEE
FACT 01.07.1862 GREAT EASTERN RAILWAY COMPANY

02.08.1980 HOLD BACK THE WAVES 182
SEXUAL HARASSMENT DISCUSSION

MY MEDICAL KNOWLEDGE DISCUSSION 183

All contents within this book are correct at the time they were released, and cover my life, my knowledge and how I employ that knowledge. These books cover articles which I have released over the years from 1946 to the present date as newsletters or lectures given by me, which is now being issued in book format.

As I am the only person who can copyright my life, and my life's work, therefore, there is nothing which I can released which relates to any other person copyright, as they do not have one upon me or my work. Any person/s making such claims are in fact in breach of my copyright.

I am quite aware that there have been, and will be those writers who write upon me or my work, which have had no authority from me to do so. Nor have they paid me any value for the rights to release such information. I have shut my eyes to this offense, and will continue to do so, unless they are stupid enough to make claimes that I have stolen their copyright's, then I have a duty to all those who have help me to stand up and be counted. You will find that I am a tiger, and a powerful one at that, and will tear you to pices in public.

TAKE WARNING : any stupid letters on such subjects shall be released in the very next book that I release with whatever action that I have taken in response to that letter. for all to see. I have a task to undertake, and let no one be mistaken upon that point, that I am not capable of undertaing it. The task to create clean cheap energy, and better transportation systems, and better homes for all mankind is a must, it can be done, and it will be done. I shall take whatever measures are deemed necessary to get this technology to the market place.

What ever I write within these books are matters which are related to me or my work, directly or indirectly, are of my own statements unless stated as being an article by another person/s known or unknown. Any statements which are claimed by anyone as there's are simly talking out of their arsehole. If they have made such statements they did so without my consent, therefore they are in breach of my HUMAN RIGHTS.

Therefore, I have appointed Mr.John A.Thomas Jr, as my legal representative. who has the authority to write reports upon my work. In the effort to speed this technology to the market place.

I have given authority to RAUM & ZEIT and to NEXUS to release articles whenever they feel that the public would be interested. Authority has also been granted to Societies where ever I lecture to write upon my work or lecture given. In the effort to unite inventors and researches together.

To my knowledge. I am the only Inventor who at my lectures plead with all inventors to join together in an effort to speed up new products to the market place. For as long as you remain as an individual the media will laugh at your claims. United we can work together, thus the media will have to re-think on what they say.

New cleaner power and transportation systems are a must, we all see things different, we all think different, but the time has arrived where we must try to think the same, see the same, work together as one large family, for our childrens sake if not for our own sake of survival.

Yours for a cleaner better world for all mankind.

This book has been released upon the authority of

Prof. J. R.R. Searl

PROF. JOHN ROY ROBERT SEARL
CONSULTANT ENGINEER.
DATE. 10.08.1993.

ISBN 1 898827 10 9

AS-93JS2
324 pgs.

The Law of the Squares

By Prof. John R. R. Searl

Book 2

Copyright © 1993 by Prof. John R. R. Searl

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without permission in writing from the publisher.

Inquiries should be addressed to John A. Thomas Jr., 373 Rock Beach Road, Rochester, New York, 14617-1316.

ISBN 1 898827 15 X
Issue Date September 16, 1993

Published By: Direct International Science Consortium
13 Blackburn, Low Strand, Grahame Park Estate, London, NW95NG, England
373 Rock Beach Road, Rochester, New York, 14617-1316, U.S.A.

FORWARD i

TRUTH TABLE FIRST BLOCK - NO GO VALUES FROM 4621 TO 5312 1

TRUTH TABLE SECOND BLOCK - NO GO VALUES FROM 5323 TO 6067 3

TRUTH TABLE THIRD BLOCK - NO GO VALUES FROM 6073 TO 6899 6

TRUTH TABLE FOURTH BLOCK - NO GO VALUES FROM 6901 TO 7793 9

TRUTH TABLE FIFTH BLOCK - NO GO VALUES FROM 7801 TO 8773 13
WORKED SAMPLE 3^2

PRESENTING THE LAST LEVEL OF EACH SQUARE CONTAINED WITHIN THIS BOOK 17
4^2 GROUP 2 LEVEL 2187 2 SHELLS
5^2 GROUP 1 LEVEL 1743 3 SHELLS
6^2 GROUP 3 LEVEL 1445 3 SHELLS
7^2 GROUP 1 LEVEL 1230 4 SHELLS
8^2 GROUP 2 LEVEL 1066 4 SHELLS
9^2 GROUP 1 LEVEL 935 5 SHELLS

PRESENTING THE LAST LEVEL OF EACH SQUARE CONTAINED WITHIN THIS BOOK 18
10^2 GROUP 3 LEVEL 828 5 SHELLS
11^2 GROUP 1 LEVEL 738 6 SHELLS
12^2 GROUP 2 LEVEL 660 6 SHELLS
13^2 GROUP 1 LEVEL 591 7 SHELLS

PRESENTING THE LAST LEVEL OF EACH SQUARE CONTAINED WITHIN THIS BOOK 19
14^2 GROUP 3 LEVEL 530 7 SHELLS
15^2 GROUP 1 LEVEL 473 8 SHELLS
16^2 GROUP 2 LEVEL 421 8 SHELLS

PRESENTING THE LAST LEVEL OF EACH SQUARE CONTAINED WITHIN THIS BOOK 20
17^2 GROUP 1 LEVEL 373 9 SHELLS
18^2 GROUP 3 LEVEL 326 9 SHELLS

PRESENTING THE LAST LEVEL OF EACH SQUARE CONTAINED WITHIN THIS BOOK 21
19^2 GROUP 1 LEVEL 282 10 SHELLS
20^2 GROUP 2 LEVEL 240 10 SHELLS

PRESENTING THE LAST LEVEL OF EACH SQUARE CONTAINED WITHIN THIS BOOK 22
21^2 GROUP 1 LEVEL 198 11 SHELLS
22^2 GROUP 3 LEVEL 158 11 SHELLS

PRESENTING THE LAST LEVEL OF EACH SQUARE CONTAINED WITHIN THIS BOOK 23
23^2 GROUP 1 LEVEL 118 12 SHELLS

PRESENTING THE LAST LEVEL OF EACH SQUARE CONTAINED WITHIN THIS BOOK 24
24^2 GROUP 1 LEVEL 79 12 SHELLS

PRESENTING THE LAST LEVEL OF EACH SQUARE CONTAINED WITHIN THIS BOOK 25
25^2 GROUP 1 LEVEL 39 13 SHELLS

TRUTH TABLE BLOCK 1 - WILL GO VALUES - 4620 TO 5310 26

TRUTH TABLE BLOCK 2 - WILL GO VALUES - 5313 TO 6071 43

TRUTH TABLE BLOCK 3 - WILL GO VALUES - 6072 TO 6898 61

TRUTH TABLE BLOCK 4 - WILL GO VALUES - 6900 TO 7799 81

CONTENTS.

BASIC KNOWLEDGE ON HEALTH MATTERS SEX 102
DRINK

BASIC KNOWLEDGE ON HEALTH MATTERS HEART ATTACK 103
DIVERTICULAR DISEASE OF THE COLON
SEVERE ECZENA
JOHN BARDEEN INVENTOR OF THE TRANSISTOR

BASIC KNOWLEDGE ON HEALTH ELECTRICAL/MAGNETIC FEILDS DANGER 104

FACTS 07.02.1890 THE STANLEY CYCLE EXHIBITION 106

FACTS 07.02.1991 USE OF CONDOMS TECHNOLOGY 107
07.02.1991 CARBON FIBRE COMPOSITES

FACTS 07.02.1991 DUCTILE CAST-IRON PARTS 108

FACTS BRAIN TEASER 109
07.02.1991 CRISIS IN THE GULF

FACTS 07.02.1989 SOLAR POWER CELLS 110
02.02.1963 STATEMENT MADE BY ME

TRUTH TABLE BLOCK 5 - WILL GO VALUES - 7800 TO 8774 111

FACTS 13.03.1991 LASER DIODE 134

QUESTIONS 7 TRUE OR FALSE 138

SEARL'S KNOWLEDGE 139

SEARL'S MAGNETIC KNOWLEDGE DISCUSSION 179

FACTS 04.04.1962 STEEL 213

FACT 10.04.1991 DIODE LASER 214

SEARL'S KNOWLEDGE 1946 ATOM 235

FACT DATA APRIL 1991 238

LATEST NEWS TODAY 23.04.1991 247

DATA 21.03.1991 248

DATA 21.03.1991 ON THE ENERGY FRONT 249

DATA 21.03.1991 SAFETY TYRES 251

DATA 25.04.1902 IDIOSYNCRASY 252
JANUARY 1991 SICKLE CELL

DATA 06.12.1929 LONDON TRAFFIC 254
06.12.1990 LONDON TRAFFIC
18.04.1952 MULTIPLE STATION TRANSFER

DATA JANUARY 1991 UK PROBLEMS 256

DATA 14.03.1947 ROCKET PROPULSION 258
28.02.1890 FORTH BRIDGE

CADMIUM 259

DATA 22.11.1935 ENGINEERS 261

DATA 21.02.1890 WELDING 262
29.11.1878 NETTLES
17.05.1935 LOCOMOTIVE

DATA 15.03.1907 ALCOHOL ENGINES 263
29.03.1990 AEROSOL SPRAY

DATA 29.03.1946 YOUNG SCIENTISTS 264
OCTOBER 1989 STM
OCTOBER 1989 EIFFEL TOWER

DATA 22.02.1889 CYCLE TRADE 265
14.02.1913 OILFIELDS

DATA 14.02.1991 BATTERIES 266
08.03.1929 BAIRD TELEVISION

DATA 04.10.1901 SUBURBAN TRAINS 267
17.05.1990 MAGNET

DATA 17.05.1929 STEEL CARTEL 268
20.09.1851 PADDLE WHEELS
20.09.1990 STEEL MOLE
13.12.1867 FIRE

DATA 27.09.1990 SILICON 269
27.09.1889 BRIDGE EUROPE

DATA 08.11.1878 SULPHURIC IMPURITIES 270
25.10.1963 SIR HARRY RICARDO
22.03.1918 WOMEN (WORK)

DATA 02.05.1991 MOBILE TELEPHONE 271
02.05.1991 GALLIUM ARSENIDE IC
02.05.1991 GALLIUM ARSENIDE DEVICES
02.05.1991 HEAD-UP DISPLAY

DATA 01.05.1991 F-22 272
01.05.1991 AMSTRAD LEGAL ACTION
01.05.1991 ICL

DATA 02.05.1991 AIR MARKET 273
01.05.1991 80486SX BLACK MARKET
APRIL 1991 GLOBETECH
01.11.1990 SOLAR-POWER CARS

DATA 13.09.1935 LAND SPEED RECORD 274
12.04.1895 STEAM ENGINES BILL
16.08.1889 WAVE MOTOR
16.08.1990 RAMS

DATA 02.05.1930 FIRST TELEPHONE UK/AUSTRALIA 275
02.05.1991 BRIAN COLLINS

DATA 21.06.1907 SMOKE POLLUTION 276
21.06.1990 DE-ICING AIRCRAFT

CONTENTS.

DATA 14.06.1929 FLYING GROWTH 277
14.06.1990 SPACE STATION FREEDOM

DATA 06.09.1929 QUESTION OF WAGES 278
30.08.1990 SENSOR

DATA 18.04.1990 IRAQI TUBES 279
11.10.1878 GARLIC FRON GRAIN

DATA 05.05.1991 MY DETAILS 280

DATA 25.01.1990 MAGNETICALLY POWERED SHIPS 281
25.01.1924 HELICOPTER

DATA 18.10.1935 12-CYLINDER ENGINE 282
18.10.1990 PROMETHEUS

DATA 18.10.1990 KENYA 283
18.10.1990 GLOBAL WARNING
18.10.1990 NOx PROTOCOL
18.10.1968 SEARL STATES
07.03.1947 MARCONI
10.05.1990 FBR AT DOUNREAY
10.05.1990 GOVERNMENT GRANTS

DATA 05.04.1889 CAMERA 284
05.04.1990 GALLIUM ARSENIDE
04.04.1991 ANTI-TRUST LAWSUIT
04.04.1991 IRAQIS

DATA 04.04.1991 CHARGING BULL 285
04.04.1991 MOBILE TELEPHONE
04.04.1991 PORTABLE PHONES

FULL PAGE DRAWING DONE BY J.P.ROOS. 20th, SEPTEMBER 1970 286
(I WONDER IF HE IS STILL ALIVE ?)

FULL PAGE DRAWING DONE BY J.P.ROOS. 20th, SEPTEMBER 1970 287

DATA 20.09.1970 MY STATEMENT 289
04.04.1991 FIRE EXTINGUISHERS

DATA 04.04.1991 BATTERY 289
01.11.1918 OIL

DATA 01.11.1990 AIR POLLUTION 290
MARCH 1991 PROBLEM

DATA MARCH 1991 PROBLEM 291
10.05.1946 SEARL STATES

DATA 10.05.1990 CHROMIUM PLATING 292
08.05.1991 DFMA

DATA 02.05.1991 UNDERGROUND LINE 293
03.05.1991 POLYARIDE

DATA 02.03.1991 TEMPGUARD 294

DATA 02.05.1991 NYLON 66 295
02.05.1991 EPOXY RESIN

02.05.1991 RIM-CAST 6HI
09.05.1862 KITE
09.05.1991 IRAQI
DATA 09.05.1991 SPACE FLIGHT 296
JOHN THOMAS Jr. QUESTION 1 8^2 NOT SYMMETRICAL 298
JOHN THOMAS Jr. QUESTION 2 12^2 CORRECTED 301
JOHN THOMAS Jr QUESTIONS PLUS MY ANSWERS 303
FIRST NEWSPAPER ARTICLE 'THE NEWBURY WEEKLY NEWS MARCH 19th,1936, PAGE 7. 308
SECOND ARTICLE NEWBURY WEEKLY NEWS.23rd.APRIL 1936. PAGE 7. 310
JOHN THOMAS Jr. MORE QUESTIONS MY ANSWERS 311
BRAIN TEASER 316
16.05.1991 BT PIGEON PROBLEMS
16.05.1991 MP DEMANDED
16.05.1890 TARIFF
FACTS 16.05.1991 COOLING MIRRORS 317
16.05.1991 SIR FRANK WHITTLE
DATA 16.05.1991 CHERNOBYL 320
16.05.1991 HELEN SHARMAN
DATA 17.05.1991 5 MISSING AIRCRAFT 321
MAY 1991 WRAPPED SPRING CLUTCHES
DATA MAY 1991 KENSEY CATHETER 322
DATA 18.05.1991 HELEN SHARMAN ORBIT 324

This set of books are released so that readers can understand the truth about myself JOHN ROY ROBERT SEARL, my life, my knowledge, my skills, how I employ that knowledge, and what others have stated about me, and what is wrong with that statement. There is no copyright held by any one upon my life, except that of myself. I and I alone hold the full copyright upon my life. All this information upon my work has been known around the world since 1956. I have given lectures so often over those years. Therefore anything written by any person other than those which I have declared within these books, are in fact in breach of my copyright. Please take note. This is a public statement.

John Thomas Jr. of the U.S.A. has been accepted by me as my representative, he has been given rights to write upon my life from the mass of data which he holds. More will be supplied him in due time.

Others whom have been given the rights are RAUM & ZEIT plus NEXUS and all places or societies where I have been booked to lecture throughout the world. I shall continue to release the TRUTH at all cost. It is my life and not yours.

These books are released by the authority of

Prof J.R.R. Searl

PROF.JOHN ROY ROBERT SEARL
CONSULTANT ENGINEER.
DATED.16.09.1993.

ISBN 1 898827 15 X

AS-93JS3
360 pgs.

The Law of the Squares

By Prof. John R. R. Searl

Book 3

Copyright © 1993 by Prof. John R. R. Searl

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without permission in writing from the publisher.
Inquiries should be addressed to John A. Thomas Jr., 373 Rock Beach Road, Rochester, New York, 14617-1316.

ISBN 1 898827 20 6
Issue Date September 20, 1993

Published By: Direct International Science Consortium
13 Blackburn, Low Strand, Grahame Park Estate, London, NW95NG, England
373 Rock Beach Road, Rochester, New York, 14617-1316, U.S.A.

CONTENTS. PAGE.

FACTS ABOUT J.R.R.SEARL i

EDUCATION ii
LECTURES

FACTS ABOUT SEARL'S FAMILY iv

MAIL FACTS vi

INTRODUCTION 1

SCIENTIFIC INFORMATION (GLASS FACTS) 2

ELECTRIC SPACECRAFT JOURNAL, ARTICLES, CORRECTIONS 10

RECENT CANADA TOUR 24

POEM..THE GREAT MOWEGEE 30

BASIC FACTS 32

PAST STATEMENTS MADE BY ME. 33
LETTER FROM MR.VOS OF HOLLAND, 1970

MR.ANDERS HEERFORDT ARTICLE 35

MR.JOHN A.THOMAS LETTER TO ELECTRIC SPACECRAFT JOURNAL 39

ELECRIC SPACECRAFT JOURNAL REPLY TO JOHN A.THOMAS 41

MY STATEMENTS OF THE PAST 42

A STATEMENT BY THE REV.G.H.NICHOLSON 45
ATHOL PARK'S REPORT TO VERA COMER, 1978

FACTS 51

SOVIET MAN IN SPACE 65

STARSHIP EZEKIEL CRITERIA 74

REV.GEORGE NICHOLSON BOOKLET..THE EPIC STORY OF FREE ENERGY 98

STATEMENTS MADE BY SEARL 106

PAUL G.J.VAN DER REEKE LETTER TO VERA COMER 109

MRS.FERGUSON'S REPORT TO MRS COMER 110

THE LAW OF THE SQUARES (BLOCK 24) 111

MY POINTS OF VIEW 149

STARSHIP EZEKIEL MK.V. 152

FLIGHT TIME LIMITATIONS 165

PAST STATEMENTS MADE BY SEARL 172

LETTERS 173

CONTENTS. PAGE.

MY PAST STATEMENTS 179

SQUARES BONDING 181

SQUARES PAIRING 184

FACTS 192

3 SQUARES SHARING 197

FACTS 203

3 SQUARES SHARING 204

STATEMENTS MADE BY SEARL 210

4 SQUARES SHARING 213

FACTS 218

FACTS AS I KNOW THEM 219

4 SQUARES SHARING 220

FACTS AS I KNOW THEM 221

4 SQUARES SHARING 222

FACTS AS I KNOW THEM 224

4 SQUARES SHARING 225

A POEM TO THINK ABOUT 228

5 SQUARES SHARING 229

BRAIN TEASER 2 231

5 SQUARES SHARING 232

STATEMENTS MADE BY SEARL 234

BRAIN TEASER 3 236

6 SQUARES SHARING 237

FACTS QUOTED BY SEARL 239

6 SQUARES SHARING 240
FACTS

7 SQUARES SHARING 241

8 SQUARES SHARING 243
FACTS

STATEMENTS MADE BY SEARL 246

10 SQUARES SHARING 247

CONTENTS. PAGE.

THE DREAM THAT CAME TRUE 248

MR. SHERWOOD'S REPORT 286

LETTERS 288

S.E.G. REQUIREMENTS 290

FACTS 296

MY KNOWLEDGE IN 1946..MAGNETISM 304

MENSA LETTER 309

MY KNOWLEDGE IN 1946..HYDRAULICS 311

FACTS...FLYING 321

MY KNOWLEDGE IN 1946 336

MY KNOWLEDGE TODAY 342
ANSWERS TO BRAIN TEASERS

MEDICAL FACTS AS I KNOW THEM...STONES. 342

FACTS AS I KNOW THEM 347

ATHOL PARKS DRAWINGS TO HIS ARTICLE 352

DREAM 1, DETAILS 353

INFORMATION 355

SUMMING UP 357

THESE BOOKS ARE OFFERING A CONCEPT FOR RESEARCH AND DEVELOPMENT. AND NOT A PROJECT FOR SALE.

The contents in this set of books are the views or understanding of JOHN ROY ROBERT SEARL, unless it is a report by another person, in relation to JOHN ROY ROBERT SEARL, in which he is not responsible for its contents, but the person who wrote it.

You may view these statements in a different light, you are invited to submit your views, which Searl is only too pleased to study them. Only by exchange of views can progress be made, Searl appreciate this point, and welcome all views which will lead to success.

PLEASE NOTE: I HAVE ACCEPTED MR.JOHN A. THOMAS Jr. AS MY REPRESENTATIVE. as such, he has been given the authority to write upon my technology in the effort to push this technology to the market place.

These books have been released upon the authority of

PROF.JOHN ROY ROBERT SEARL
CONSULTANT ENGINEER
DIRECT INTERNATIONAL SCIENCE CONSORTIUM.
DATE. 20.09.1993
ISBN 1 898827 20 6

AS-93JS4
413 pgs.

The Law of the Squares

By Prof. John R. R. Searl

Book 4

Copyright © 1993 by Prof. John R. R. Searl

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without permission in writing from the publisher. Inquiries should be addressed to John A. Thomas Jr., 373 Rock Beach Road, Rochester, New York, 14617-1316.

ISBN 1 898827 25 7
Issue Date December 16, 1993

Published By: Direct International Science Consortium
13 Blackburn, Low Strand, Grahame Park Estate, London, NW95NG, England
373 Rock Beach Road, Rochester, New York, 14617-1316, U.S.A.

BOOK 4.

CONTENTS 1.

		PAGE
MY REPORT JANUARY 1st. 1968.		1
MY REPORT NOVEMBER 15th. 1987		2
FACTS AS I VIEWED THEM 1946	(x-rays)	3
STATEMENTS MADE BY ME 1946	(The 'Sachsen' and other Zeppelins)	11
MY KNOWLEDGE 1946	(Brain Teaser 1)	12
MY ATMOSPHERE KNOWLEDGE 1946		13
REPORT RELEASED BY ME, APRIL 13th, 1961	(Y.A.Gagarin)	15
MY INFORMATION STUDY JULY 1993		25
MY ELECTRICAL KNOWLEDGE 1946	(13 questions 1946)	27
MY STATEMENT ON STARSHIP EZEKIEL MK.V. 1968		28
MY KNOWLEDGE 1946		29
MY STATEMENTS 1968		37
MY KNOWLEDGE 1968	(flight)	43
MY KNOWLEDGE 1968	(brain teaser 2)	46
MY KNOWLEDGE 1993	(facts)	47
MY ELECTRICAL KNOWLEDGE 1946	(magnetism)	48
MY ELECTRICAL KNOWLEDGE 1946	(brain teaser 3)	55
FACTS		56
FACTS AS I KNOW THEM	(medical facts as I know them 1968)	58
MY MEDICAL KNOWLEDGE 1968		59
MY MEDICAL KNOWLEDGE 1968	(brain teaser 4)	64
MY CHEMICAL REACTION KNOWLEDGE 1946		65
MY CHEMICAL REACTION KNOWLEDGE 1946	(brain teaser 5)	70
JOHN WALKER ARTICLE 1993	VORTEX PROPULSION	71
FACT 30, JULY 1869	ARTIFICAL OCEAN FACT 11 JULY 1947 100 YEARS OF BRITISH CHEMISTRY	73
ARTICLES UPON ME	THE SEARL EFFECT AND RELATED UFO PHENOMENA	74
FACTS	17th,August 1969 National Space and UFO Convention	78
MY MEDICAL KNOWLEDGE 1968		80
FACTS		81

BOOK 4.

CONTENTS 2.

FROM IDEA TO MARKET PLACE 1968 86
GLOBAL PROPERTY PROTECTION 1993 91
THE SEARL EFFECT DESIGN 1992 92
LETTERS Russell Hobbs 94
PAST STATEMENTS MADE BY ME 96
LETTERS 102
STATEMENTS MADE BY ME 1946 103
COMMERCIAL PILOT'S LICENCE AND INSTRUMENT RATING 1977 (test) 104
MY KNOWLEDGE 1972 electrical propulsion of space vehicles 106
MY KNOWLEDGE 1972 (brain teaser 6) 124
FACTS Liberty Bell 7 125
STARSHIP EZEKIEL MK.V. CRITERIA 127
WIRING REGULATIONS 1968 137
COMPONENTS FOR EVALUATION 1993 143
MY KNOWLEDGE 1946 152
COMPONENTS FOR EVALUATION 1993 161
TECHNOLOGY 1993 162
MY MEDICAL KNOWLEDGE 1993 166
FACTS 1993 171
FACTS 1946 173
FACTS 184
STATEMENTS MADE BY ME 1964 191
PAST STATEMENTS MADE BY ME 192
FACTS 1993 199
FACTS 214
MY MEDICAL KNOWLEDGE 1964 215
FACTS AS I KNEW THEM 1964 229
STARSHIP EZEKIEL MK.V. AIR LAW 1968 239
MY ELECTRICAL KNOWLEDGE 1946 answers to my questions on page 27 258
MY ELECTRICAL KNOWLEDGE 1946 questions 259
FACTS fact November 1993 287

S.E.G. PATENT U.S.A. 291
S.E.G. PATENT U.S.A. my letter to the Kenya High Commission reply 293
LETTERS reply 294
FACTS AS I SEE THEM 1993 Mars 295
ANSWERS TO QUESTIONS 300
FACTS AS I SEE THEM 1993 302
TEST EQUIPMENT/WIRING REQUIREMENTS 1993 308
WIRING REQUIREMENTS 1993 312
FACTS 323
FACTS letter from BBC in the Midlands angular kinetic energy 325
FACTS Potential energy 326
A MATTER OF ANTIGRAVITY 332
SUPERCONDUCTIVITY 352
FACTS 363
SENSORS 364
DREAM 1 STATES reference ohms law colour plate 372
DREAM 1 STATES COLOUR PLATE OHMS LAW 373
DREAM 1 STATES PLATE PERIODIC TABLE OF THE ELEMENTS. 1 375
DREAM 1 STATES PLATE PERIODIC TABLE OF THE PROPERTIES OF THE ELEMENTS.2. 377
DREAM 1 STATES PLATE CONVENTIONAL DIAGRAMS OF CRYSTALS FORMS 379
DREAM 1 STATES DER SEARL-EFFECT GENERATOR in German 381
DREAM 1 STATES PLATE SANDBERG'S REPORT UPON WHAT HE SAW OF MY MAGNETIC DEMONSTRATION 15-08-1982 384
DREAM 1 STATES PLATE S.E.G PLATE DESIGN.DAVID STIRLING 385
DREAM 1 STATES PLATE S.E.G. PLATE DESIGN. DAVID STIRLING 387
DREAM 1 STATES PLATE S.E.G. LAYUP DESIGN. DAVID STIRLING 389
DREAM 1 STATES PLATE ROLLER ACTION. PETER BRUNDT 391
DREAM 1 STATES PLATE ROLLER ACTION. PETER BRUNDT 393
DREAM 1 STATES PLATE ROLLER MAGNETIC/ELECTRON ACTION PETER BRUNDT 395
DREAM 1 STATES PLATE S.E.G. LAYUP. PETER BRUNDT 397

BOOK 4. CONTENTS 4.

DREAM 1 STATES PLATE S.E.G. LAYUP. DAVID STIRLING 399

DREAM 1 STATES PLATE STARSHIP EZEKIEL MK.V. THILO MüLLER 401

DREAM 1 STATES PLATE INVERSE-G-VEHICLE FLIGHT CELLS CONTROL
PETER BRUNDT 403

DREAM 1 STATES PLATE SEARL DESIGNED HOME. PLANUNGSBüRO 405

DREAM 1 STATES PLATE TOP MAGNETIC DRIVE SYSTEM. DEVON TASSEN 407
MODEL DISC..JOHN HOCKNELL

DREAM 1 STATES PLATE INVERSE-G-VEHICLE MODEL
J.R.R.SEARL TALKING TO A REPORTER 409

DREAM 1 STATES PLATE J.R.R.SEARL LECTURING TO REPORTERS
FIRST EVER PRESS PICTURE ARTICLE OUT approx.1967 411

The contents of this book are absolutely correct, unless an article is quoted as being an article of another person, they are my reports as seen through my eyes. based upon my knowledge and education.

I appreciate that some of the readers will have different views, if so write me your views or ideas, let me hear them, by keeping them to yourself achieves nothing. Airing them, might come up with something interesting.

I have within these books shown views of others. as they have presented them, right or wrong. they have a view. I repect that view, as their rights.

NOTE : QUESTION

WHAT IS THE ANSWER TO 4^2 ?

You know that answer, do you ? You think you do. Or you do know it ? I doubt if you do know the full answer, unless of course you have read and studied my books. You see, I know what answer you will give, which is 16, for that I would give you just 1 point out of 10 points for trying.

Each square has four main forces. is not symmetric. but only nearly symmetric, for if it was fully symmetric, the four forces would equal zero, there would therefore be no forces acting upon the roller. But the squares are only nearly symmetric thus the roller experiences a force acting upon it. So it ruddy well tries to move out of the way, and it does.

Please take note that **MR.JOHN A.THOMAS** Jr. has been accepted as my representative and has been given the authority to write upon my work, in the direct effort to push this technology to the market place.

These books have been released by the authority of

Prof. J. R. R. Searl

PROF. J.R.R.SEARL
CONSULTANT ENGINEER.
DIRECT INTERNATIONAL SCIENCE CONSORTIUM.U.K.
DATE. 16.12.1993

ISBN 1 898827 25 7

1. Minimize physical size. Since the coil is the largest single piece (other than the AA battery) it makes sense to align the diameter of the coil to the battery and pack everything else as tightly inside as I can.
2. Minimize component count. Most Joule Thief examples on the internet (including the top picture on the Wikipedia page) soldered the legs of the individual components together. No circuit board needed.
3. Friendly to hand soldering. There are some ready-made Joule Thief circuits for sale on the internet using surface mount components and a circuit board. I wanted something I can build by hand and maybe use as a soldering teaching project to be shared on the internet.

After a few iterations, I have something I'm happy to share with the world. This is purely about the mechanical assembly – the electronic schematic is identical to the one in the Wikipedia article linked at the top of this post.

An overview in words:

- The resistor for the NPN transistor base is installed between the collector and emitter. The resistor acts as physical separation in order to avoid a short-circuit.
- The transistor and LED are pointing in opposite directions, allowing their pins to point towards each other and soldered together. The aforementioned resistor keeps the LED anode and cathode separate.
- The transistor is stuffed into the middle of the coil, utilizing the center volume.

The build sequence in pictures:

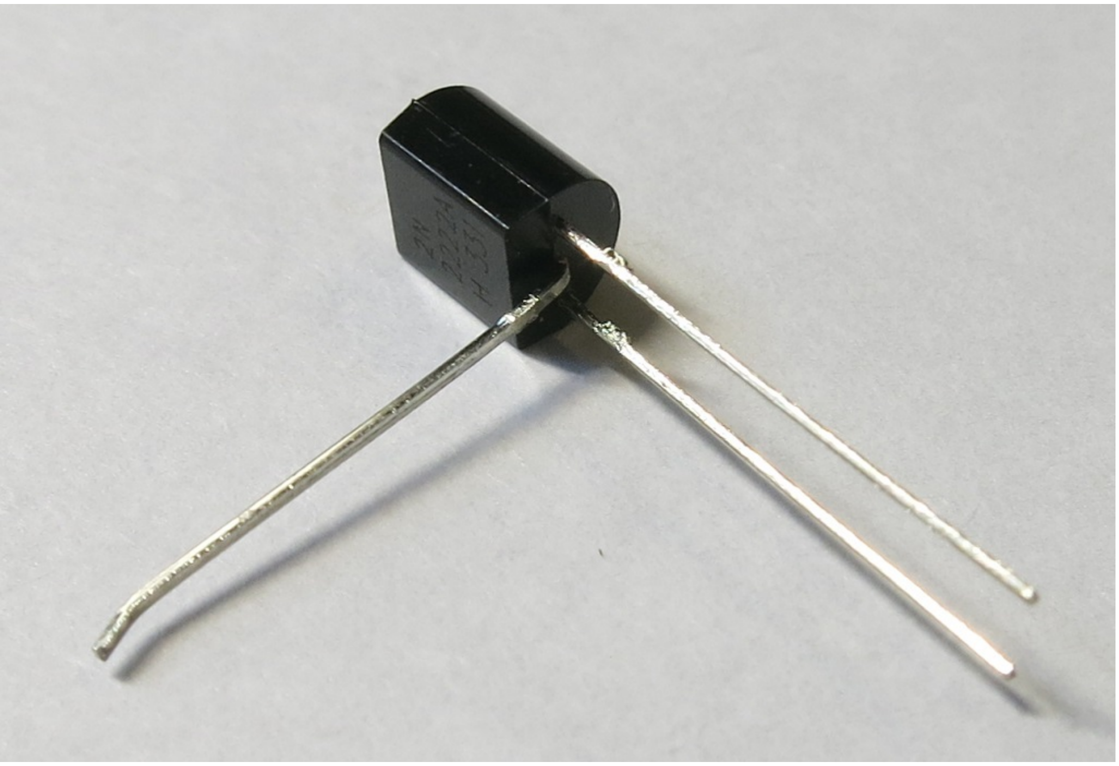

Transistor with the base bent in preparation for resistor installation.

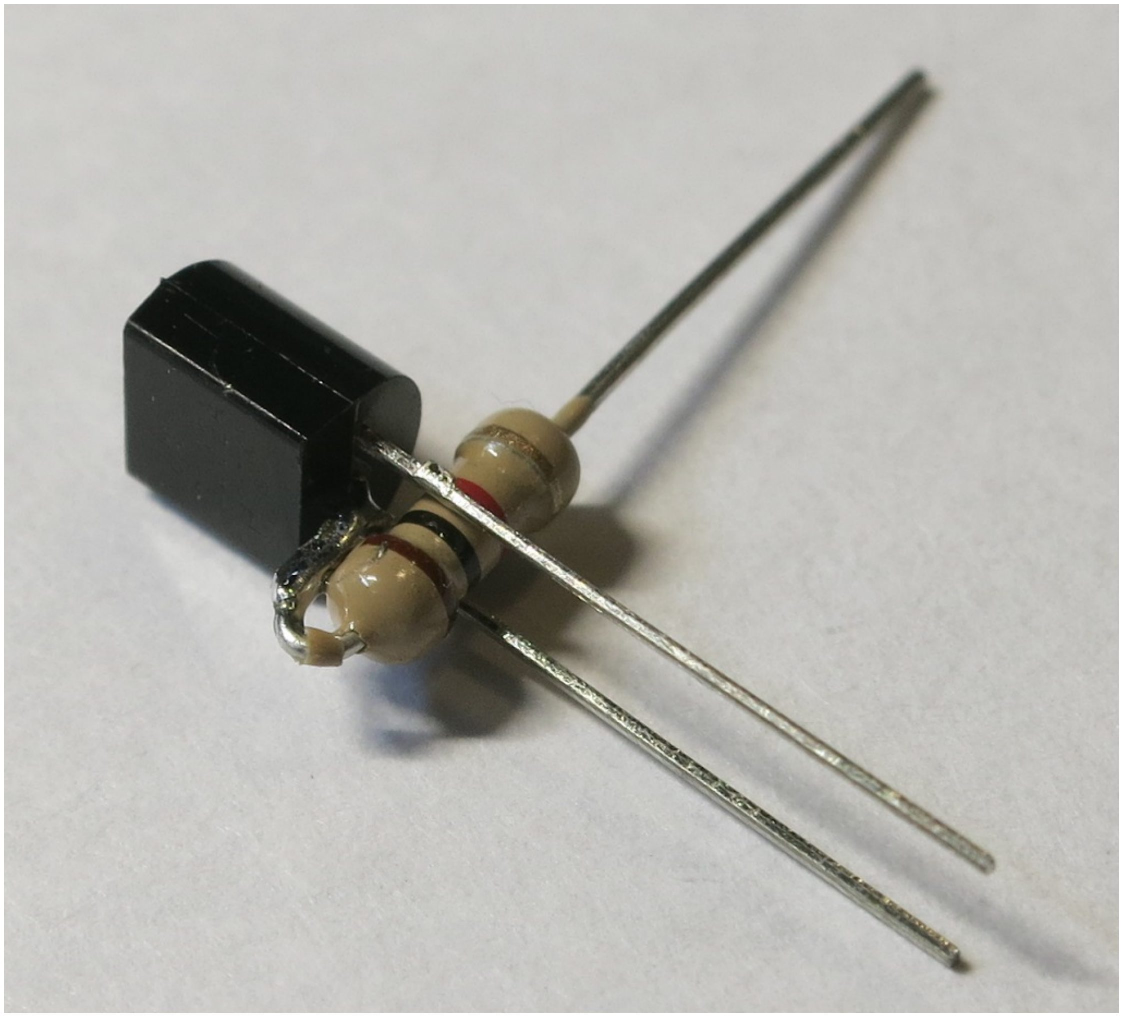

The 1K Ohm resistor is installed on the base, between collector and emitter.

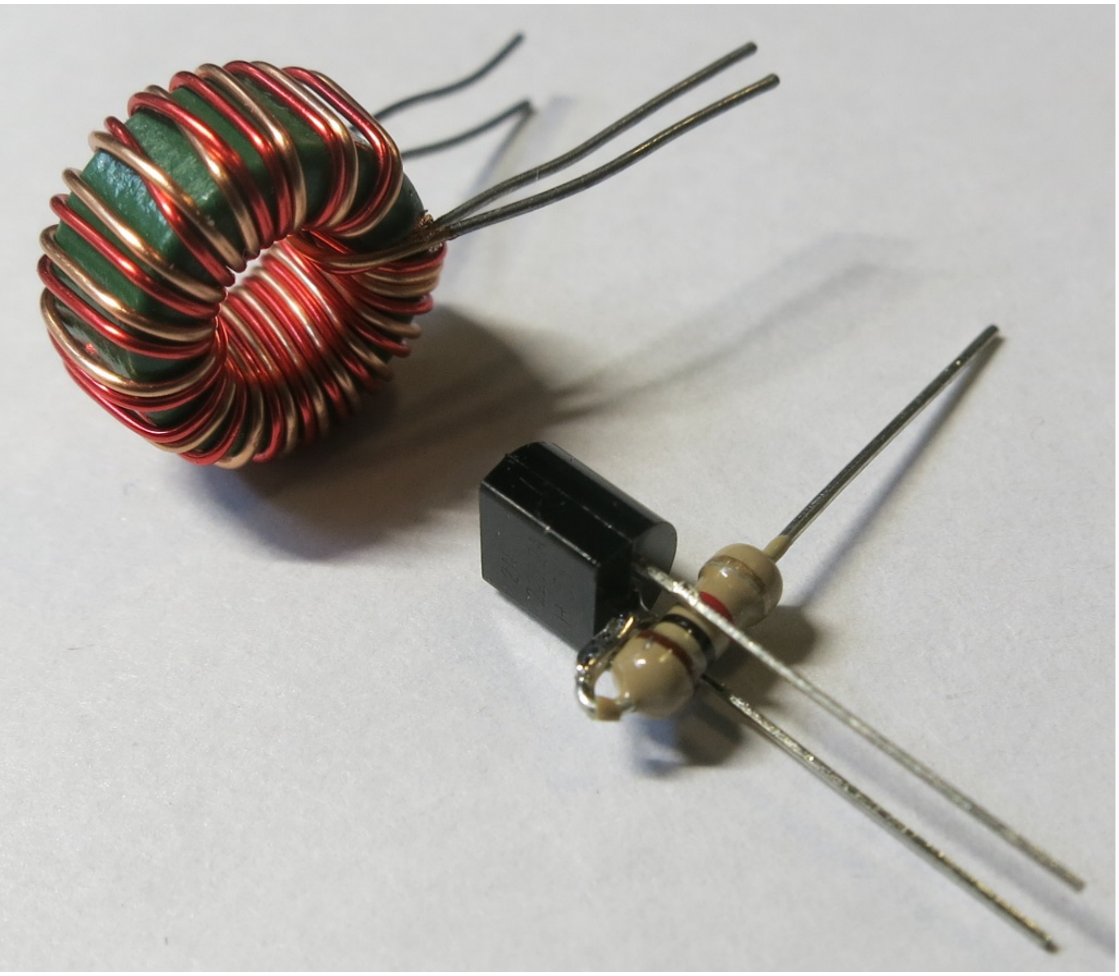

The coil has two wires wound together. One end of this dual-coil is facing the camera, the other end facing away. Since we need to wire up the coil in opposite directions, we'll bend one wire of the front pair towards the back, and the opposite back side wire to the front.

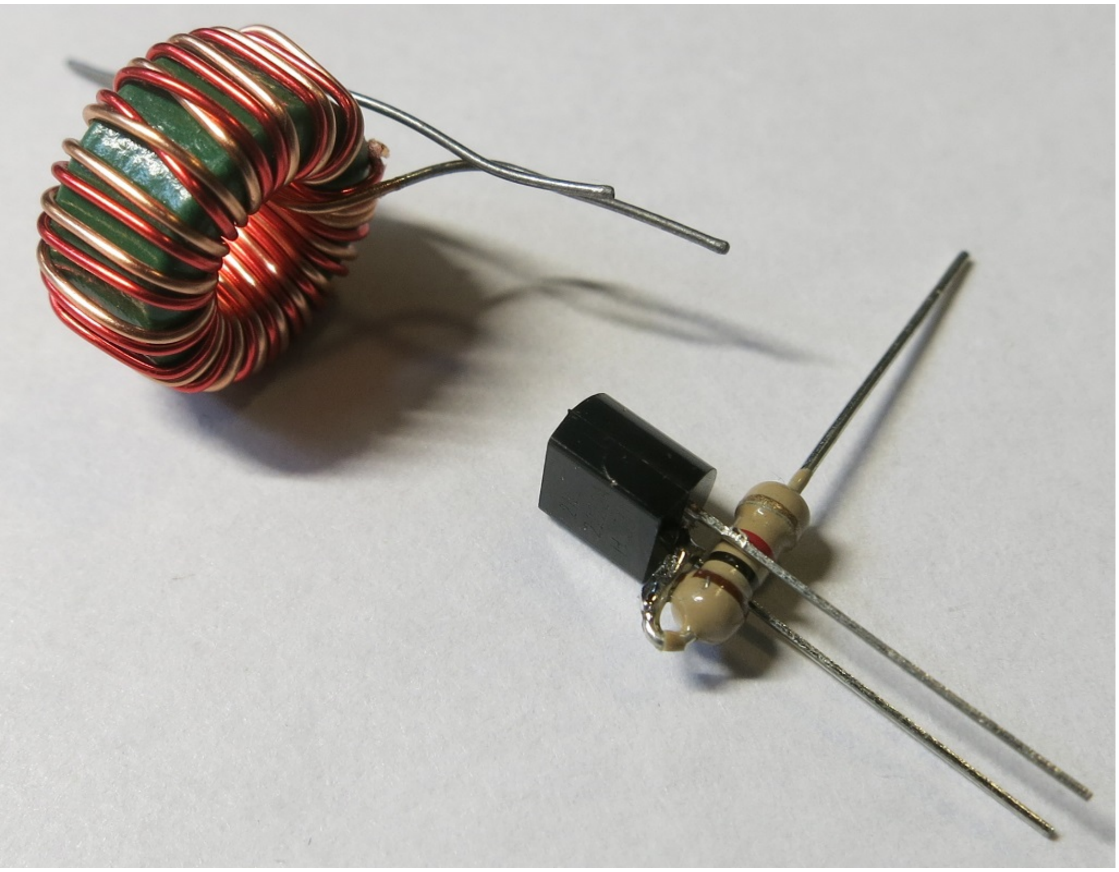

The two wires now facing away from the camera are soldered together to become the positive terminal of the circuit. One of the two wires facing the front will be soldered to the resistor, and the other to the emitter.

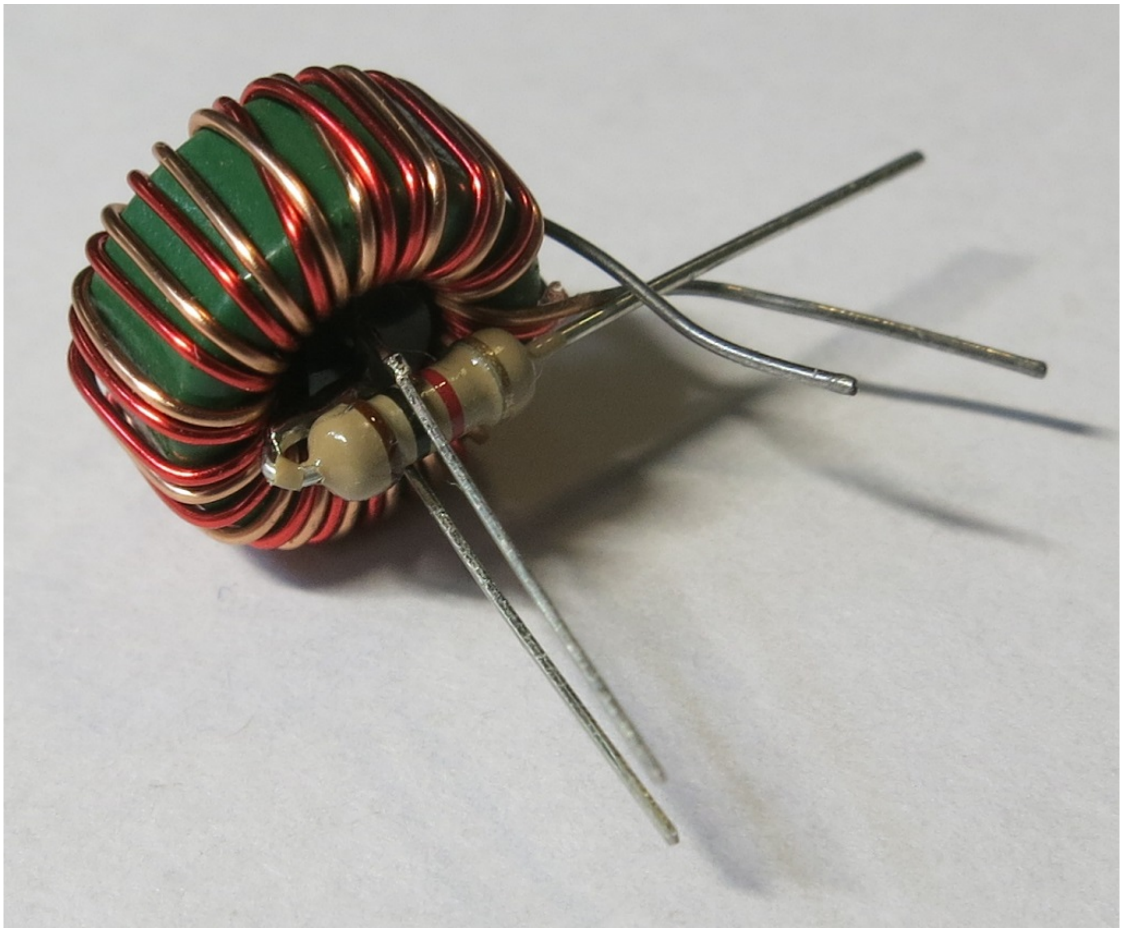

Transistor in the center of the coil. Now the coil wires can be soldered.

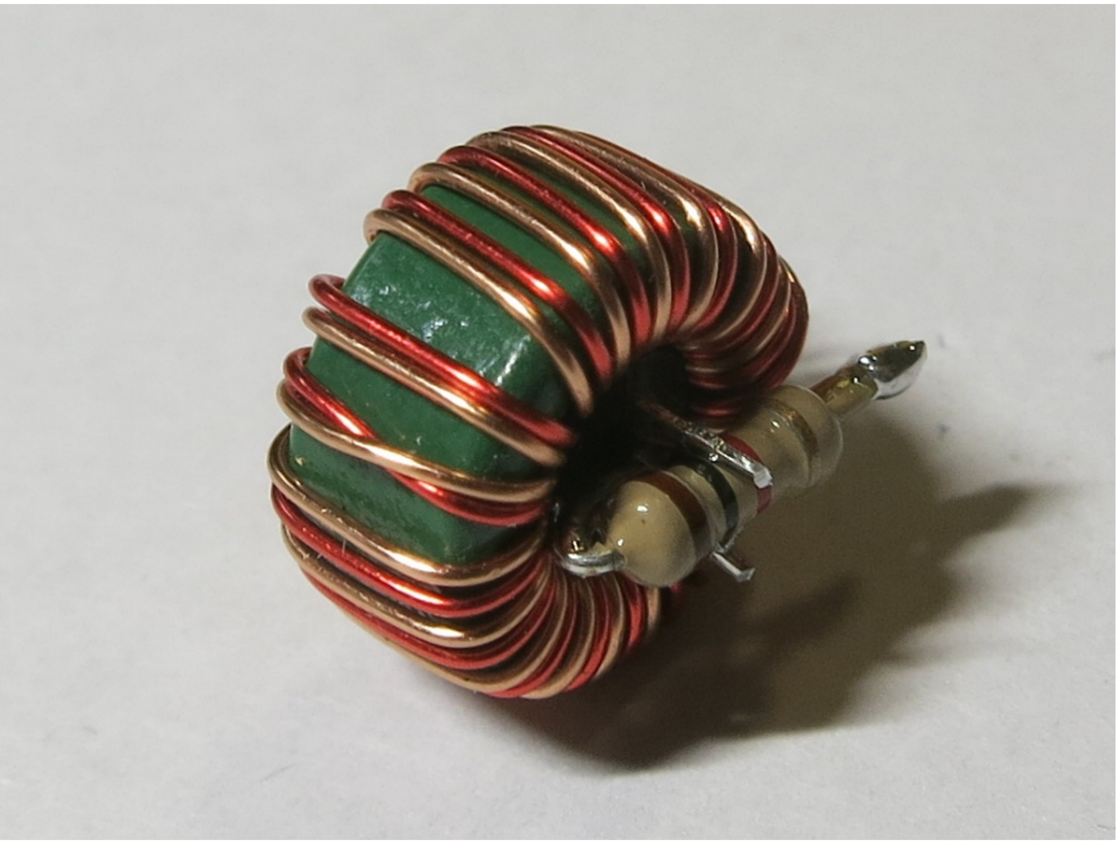

Resistor soldered to one coil wire, all the others have been trimmed short in preparation for attaching the LED.

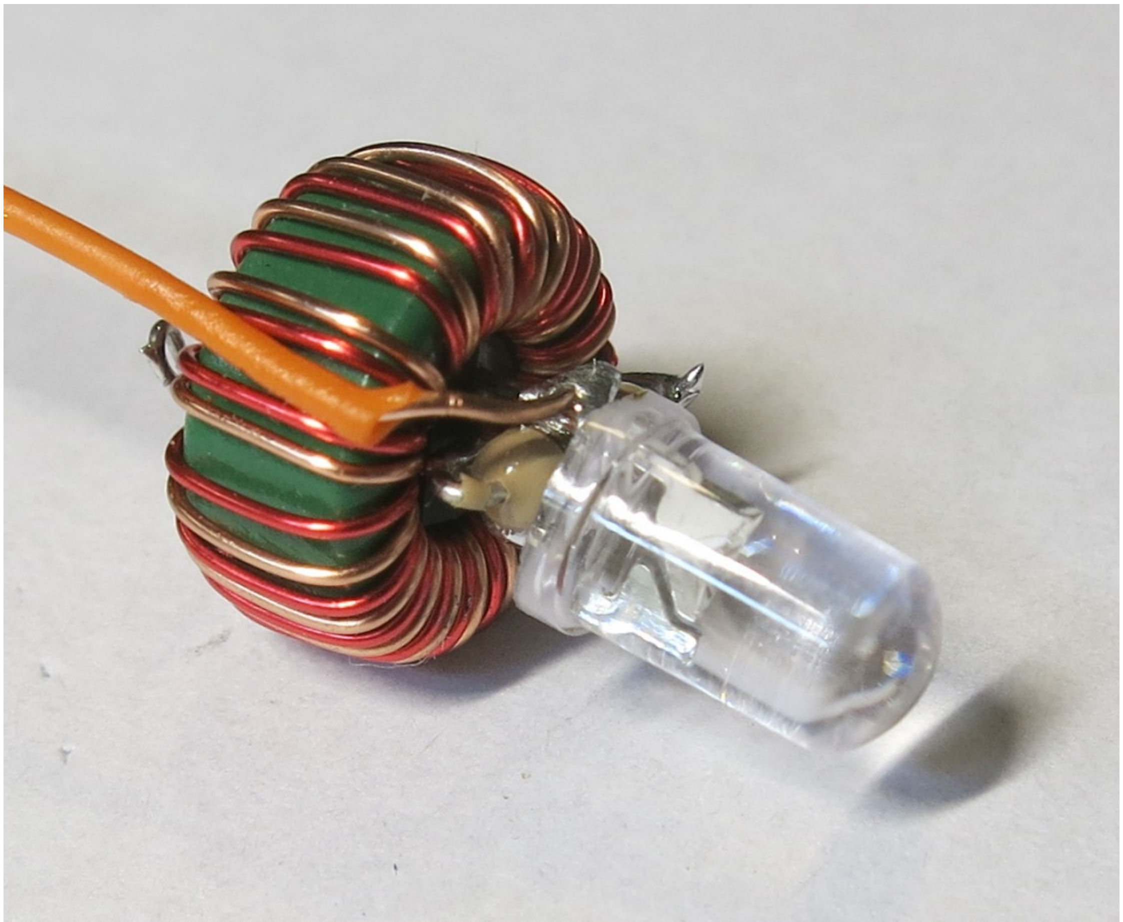

LED is soldered to the circuit, as is a wire to act as negative (return) wire.

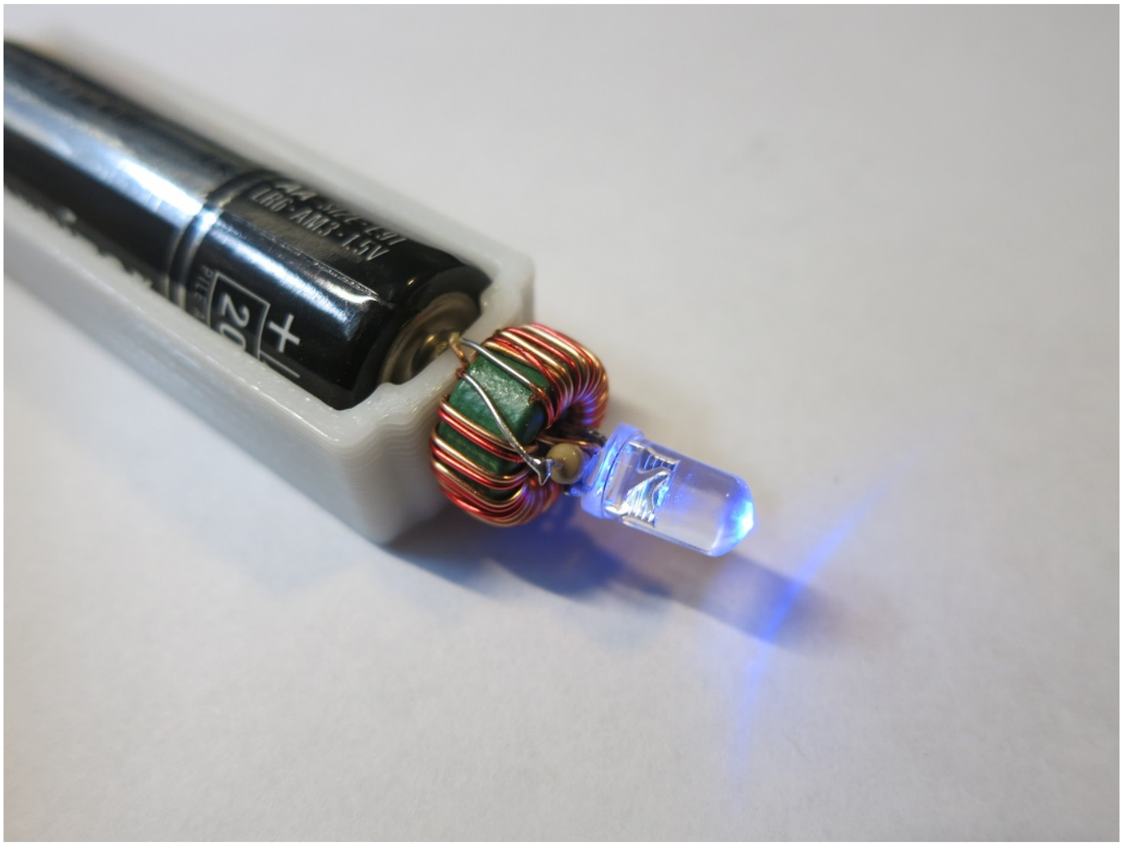

Install the whole assembly in front of a 3D-printed AA battery tray: Let there be light!

2017-10-192017-10-19 / **ROGER CHENG**

A 3D-Printed Enclosure to Take My LED Project On The Go

For the **Connect Week (https://connectpasadena.com)** event put on by **Innovate Pasadena (http://www.innovatepasadena.org/)**, the **Hackaday LA (https://www.meetup.com/Hackaday-Los-Angeles/)** group is hosting the "**Bring-A-Hack (https://www.meetup.com/Hackaday-Los-Angeles/events/243694093/)**" event where attendees are encouraged to bring projects (in any stage of completion) for show and discussion. Since I've been building my LTC-4627JR driver board as a learning project, I wanted to bring it in for show-and-tell.

Now I could just bring the assembled circuit board and pass it around as an inert object, but what fun would that be? I wanted to bring in something that shows it doing something, and provide some way for people to interact with the whole contraption. Looking at my parts on hand, it seemed easiest to rebuild my thermometer test project. I can have a simple Python program run on the Raspberry Pi, reading temperature from the Tux-Lab Si7021 breakout board, and sending it out to my display. That makes 3 circuit boards, plus they'll need portable power. I will enlist my Amazon purchases: the 3-cell lithium ion battery pack protected by a S-8254A IC, and the MP1584 buck converter to translate the battery pack's power into Raspberry Pi friendly voltage.

Making a Joule Thief Charger Circuit with Steven Chiverton

Posted by hitman

Continued from the previous article

Here is the joule thief used to test my button cells out after I successfully recharged them.

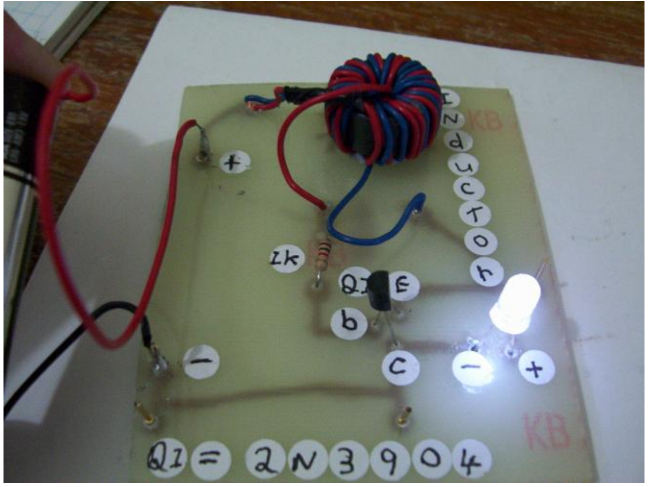

Here's my battery charger joule thief all installed into a circuit box.

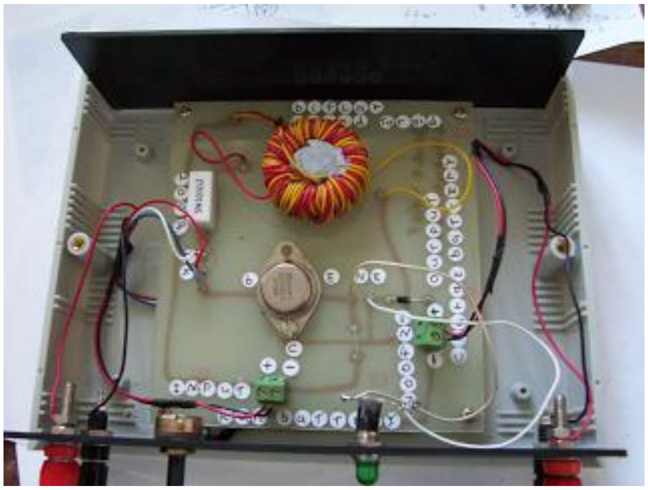

Below is the front view, notice the 12 volts green light at the front panel that is modified to house a neon globe inside it.

While looking through my electronics parts draws I found another 7 or 8 more button batteries and some had charge still on them and some I recharged successfully and 2 couldn’t be charged anyhow a few small ones took a little zapping to get them to take a charge so maybe if they are got some sulphate in them if not then why did it take so long to get them to except a charge.?

Yesterday I found another button battery cell and after a few seconds charging it started to get warm so I stopped it before it got to hot otherwise it would explode
Now I have 24 button cell batteries.

Below is a picture of the bright white with pink flash of a photo flash capacitor neon circuit test The outputs of my battery charger joule thief run to the 330 volts photo flash capacitor, and across its leads I placed a neon bulb, loosely so every so long you hear a spark jump but cant see it unless the neon leg is getting pulled in by the high voltage attraction then a short time after that the flash occurs, so if the neon is rated at 90 volts then the cap charge may be the same till it charges up more I filmed this then put it onto my pc and then I had to slow it down so much to capture the super fast flash then pause it then take a picture of it to load back onto my pc.

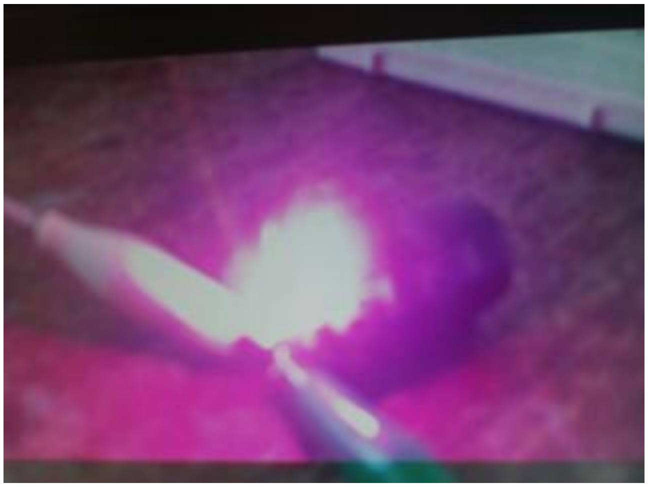

Here's what it looks like when it aint flashing

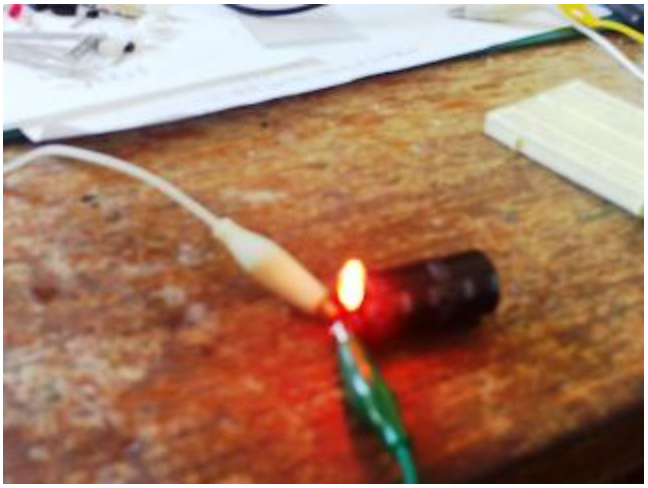

ive soldered one of the neon legs to one of the high voltage 330 volts photo flash capacitor legs and the other neon leg I had it just nearly touching the other capacitor terminal but I didn't get the arc and flash after it like I got before .
next test was running one of my jumper leads to a folded peace of aluminum foil and back to the hv capacitor and the other lead I ran to the other hv cap leg to a pencil and when the neon at the 330 volts 2200uf photo flash cap come on I gently touched the lead pencil tip to the foil edge where its folded and I got flashes etc and tiny holes melted through it and when I tried to unfold the foil it was stuck like it was spot welded , and there sure was this horrible smell made by spot welding aluminum foil.

Next experiment was running the leads from the capacitor to 2x 316 grade stainless steel bolts in a jar of tap water to create any electrolysis effects and I got very fine misty vapours come up to the top of the water and no bubbles and even when the bolts slipped together I was still getting it and even when they are apart to, and the further apart the bolts are the lower the frequency of sound is emitted from the battery charger joule thief and the closer the bolt electrodes are the higher the frequency of the sound the circuit emits gets ,

My watch battery I charged before has stopped I guess the charge didn't last long enough, I charged it on the 24th may 2010 and it lasted till today at 8past 8 , either in the morning or last night , so it lasted a week only im gona try re charging it again . Also today at the shops I went to the place they do watches and shoes and there was thousands of button cells inside this Perspex box even large ones next to it , but no one was there to answer my questions about them .
Ive recharged my watch battery again today but I was charging it the wrong way around so I got it right after that now my watch is running again so ill see how long its lasts for this time. Today is June the 1st 2010 so lets see if it lasts a week, also if it lasts longer then it may be the reverse charging that contributed to the extra running time before I charged it right

I scored a packet of free button cell batteries including large lithium cells and a few small 12 volts pencil batteries for free from 2 places in Armadale that also replace watch batteries, they have thousands of them and I counted what they gave me in total I had 85 batteries mostly button cells and ive successfully recharged them all except some lithium large button cells so also my 25 batteries I had it makes 110 batteries in total including the few large lithium cells that couldn't be charged or revived at all some of the batteries still had power in them to .

Half the button cell batteries couldn't maintain there charge and went flat overnight
So I recharged them again so ill soon check to see if they hold there charge this time also I dismantled a lithium battery it has a thin white membrane that reacts with water and foams up and also a thicker black membrane with what looks like very thin silver wire looking cross patterns on its sides and when I dismantled that battery it sure stunk inside it to
Today Tuesday June 8th 2010 I decided to use my 4 rechargeable batteries from my camera to power my radiant joule thief battery charger, to restore my 12 volts 7.2 amps an hour 20 hour sla battery, and so after an hour or so I used a pair of tweezers to short out the battery terminal and I got some sparks where I didn't before that, so I left the joule thief battery charger running all day today and now when I plug in the battery charger the red charging led light doesn't go to green after a short charge that's trickle charge, when green its still on red so im looking forward to how it goes when fully charged then in the green when its charged so if I get plenty sparks when I short the battery terminals then I can say the battery has been revived .

My 12 volts sla battery appears to still have a tiny bit of sulphate on its plates as its voltage dropped 0.7 volts over night so when the battery is low and needs topping up ill put it back on the joule thief circuit.

archive.today
webpage capture
Saved from http://www.rexresearch.com/guillot/guillot.htm search
history
5 May 2015 11:19:54 UTC
All snapshots from host www.rexresearch.com
Webpage Screenshot
share download .zip report error or abuse

rexresearch.com

Jules GUILLOT

Atmospheric Electric Generator

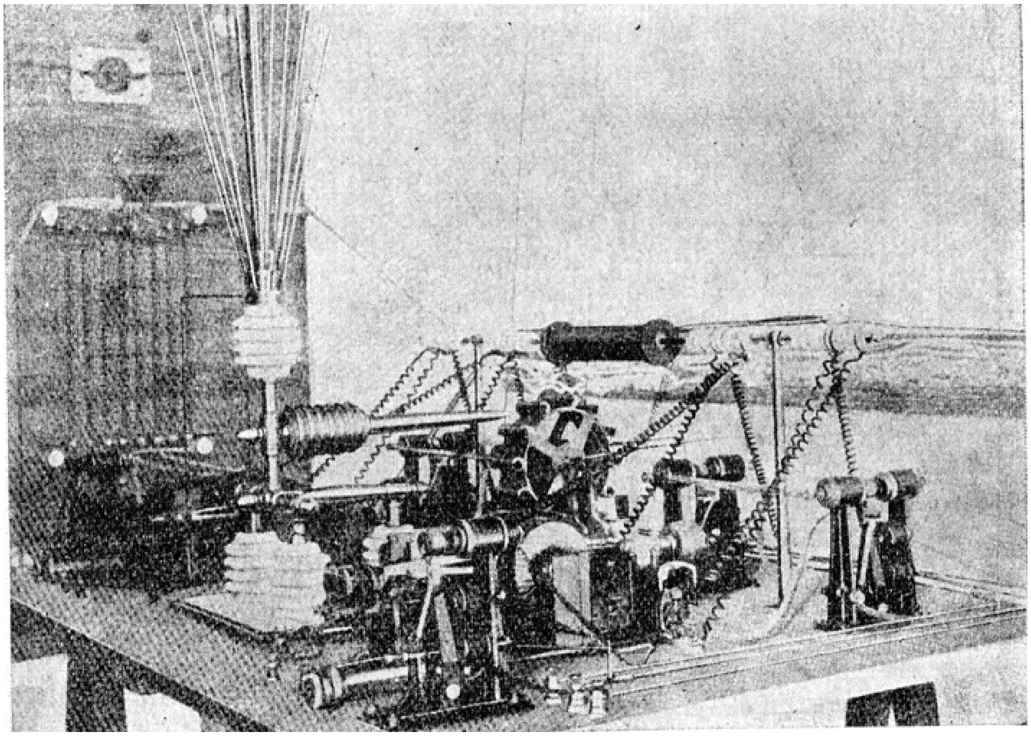

Shared by L. Cozma (Thanks !) --

The Guillot device generated about 2.5-3 Kilowatts with antenna height of ~ 20 meters. Power depends on the total collector surface and height of the vertical antenna. The apparatus in the photo produced ~300 watts with a collector 2 meters tall.

"Practical Utilization of Atmospheric Electricity"

by

Gillbert Darida

in *The Invention Encyclopedia*, pp. 204-207 (1930, Geo. Constantinescu, Ed.)

"The Earth has its own negative electricity, in the soil, and the atmosphere which surrounds the Earth is charged with positive electricity. The electric potential (the voltage) increases with the altitude, so we can say that the electric force is proportional to the atmospheric altitude (after Franklin, Quetelet, Lord Kelvin, Mascart, Joubert and other scientists).

The recent observations demonstrate that the air at 6000-7000 meters in altitude, is very highly charged with positive electricity, which could be explained by the friction between the external photosphere and the upper atmosphere of Earth, which rotates at a speed of more than 100,000 Kilometer per hour.

In that way, the Earth works like an electrostatic generator with electric charging by influence --- the upper atmosphere is positively charged by influence and the Earth crust obtains the negative polarity.

Between the two environments, the air and the soil, and inside the low atmosphere, in conditions of good weather, there are sbout 800 positive ions and also, 680 negative ion (and electrons) in just one square-centimeter of normal ionised air.

The Earth behaves like an huge electric armature negatively charged, which repels the electrons and attracts the positive ions. That positive ions' attraction determinates an electric current, also called "convective current". That's like an invisible continous bombardment, subject to daily and seasonal variations, which could be aproximated at **3 x 10 ^ -16 Amperes per square-centimeter,** and that is a total value of 1500 Amperes for the entire surface of the Earth.
The question is --- how does this current always maintains the same direction ? We can suppose that the natural radioactive emissions of the soil is responsable for this. We also know that radioactive emissions of the Earth works usually near the soil surface, and that explains the ionisation phenomenona inside caves.

The ionisation of the low atmosphere could also be the effect of the radioactive emissions of the Earth, especially when the X radiation works. Also, the Earth atmosphere is ionisated by the external radiation proceeding from the Sun and from the space environment, especially the action of ultraviolet radiation and the electron fascicles emitted by the Sun surface, at the temperature of 6000 deg. Celsius.

The capture of atmosphere electricity has been used in France, with aerial cables mounted on the Mont Blanc, and also in Germany --- with conductive cables carried by the captive baloons.

The atmosphere electricity collect system invented by eng. Jules Guillot is most ingenious and it relies on "the **electric siphon**" [*m.n. --- the discharging devices or spark-gap used today from the protection of aerial electric cables against the atmospheric electricity ; Jules Guillot has thought to recuperate and to use that electricity*].

His method consisted in the direct "pumping" of the atmosphere electricity using a collecting device which had two antennae and several collecting rods.

One antenna is **vertical** and it has a lot of **rods** scattered like an opened **fan**, with the tips against the **zenith**, for collecting the **negative** electricity which comes from the air ; the **horizontal** antenna is orientated against the **South** and its role is to collect the **positive** electricity.

The air electricity seems to have the double-polarity, as we can see at two electrised clouds when between them apears the electric discharge. The air could have different electric charging and the ionisation processus of atmosphere is very heterogenous. The inventor Guillot used **two separated and insulated armatures with the positive armature against the South** (more precisely, against the **Equator**...) and the **negative** armature against the **zenith**.

We can see in the scheme :

1,- the device P is the "**collector apparatus**" having **15-20 meters height** ; it also has a lot of **rods** which are scattered **fan**-like, all this on top of an **iron pillar** ; and also an **horizontal antenna** S oriented against the **South** ; the two antennae are disposed at **90 deg. angle** forming the two armatures where the electrical tension must apear. As the **height** of the pillar **increases**, the **voltage** increases more and more ;

2,- a system of **lightning-rods** (p1, p2, p3) for protection against discharges of atmospheric electricity during storms;

3,- the regulator system (**voltage regulator**, etc) R ;

4,- the electric **resistors** (Cn) for absorbing the parasite currents ;

5,- the "electric siphons" which have the role of making a magnetic field as a good environment for the transport of the ionized fluid, absorbing the electricity from the air ; the E1 and E2 are their regulator devices ;

6,- an **auxilliary continous current power supply** which send the current against an special **excitation coil**, used as starting device ; it works only when the system is starting.
This collector device works very good for lighting and heating. The device made by Guillot had
20 meters height on the total surface of **3 square-meters**. Also, J.Guillot used and electrical transformer for the industrial utilization of this "collector" --- as power supply for industrial electric engines."

French Patent # 551,882

Apparatus for Capture of Electric Currents in the Atmosphere

16 April 1923

Considering the terrestrial globe like the inductor of a dynamo, where the extraterrestrial ether is the inductor of electric currents circulating in the atmosphere; the apparatus of the invention described here selects through the device described below two perfectly distinct currents and eliminates others.

The device includes:

1, The air sensor; 2, a series of lightning rods; 3, vacuum; 4, an array of resistances; 5, vacuum; 6, witnesses.

1. An aerial sensor mounted on a pole is composed of **magnetic steel** fixed and sealed by a porcelain **insulator**, and surrounded a base by a **bronze ring** which are screwed **32 points of soft iron**, all forming antenna a pole. level beside it and isolated from the first, forming the 2nd pole, is fixed a sharp point of magnetic steel fixed in the horizontal position and directed towards the south and **movable to an angle of 45** is sealed in a porcelain insulator. This point is also circled a **ring of copper** notches figure. 1.

2. A series of 6 lightning surge arrestors at corners between each pole and the earth and the various gauges.

3. A first **regulator** form of a device similar absolutely similar to the aerial amtenna, but the two poles are superposed and opposed to the vertical point, and a **copper disk connects to ground**. Opposite the horizontal point, a **ring of tin** to which are welded tubes alternatively **16 tubes** composed each composed of **copper-lead** and **iron-lead** . The ring connects to the **ground** as shown in figure. 2.

A second regulator form of **automatic breakers** balanced on 3 poles and two poles also equally balanced.

4. A panel of **resistances** composed of **wire mesh** in bunches from different sections of **glass tubes** containing **copper** dust, **coal** and flowers of **sulfur**.

Vacuum cleaners consists of a **wooden box** on each pole, containing a **porcelain vase** in which **layers** isolated with **mica** is made up of **mercury, tin, coal, copper** and **sulfur**, all contaied in a **copper tube**.

6. The witnesses are comprised of ordinary incandescent lamps.

Resume

By the point at the zenith and the point to the south we channel two currents forming the two poles. We also protect from lightning. It regularizes the flow by a regulator and similar devices by a controlling each devise of adequate strength of the current harmful nature of these devices do not have loads. The refined current is conducted by ordinary copper wires**.**

French Patent # 565,395

Combined Apparatus for Capture of Atmospheric Electric Currents with Immediate Implementation

25 January 1924

We know that earth with its constitution, its rotation and movement in space, provides the electricity in the atmosphere. The electrical currents escape into space or largely accumulate towards the equator, as a result of the greater periphery of the globe.

In the atmosphere, there are two clearly distinct poles, i.e., clouds that can be electrically positive or negative; everything in nature shows this, otherwise it would be difficult to explain the lightning that occurs between the clouds and which are none other than contrary cloud electrical charges discharged by too close proximity.

The invention relates to a set of devices capable of capturing the atmospheric electricity.

So far, all searches made for this purpose has been to capture the atmosphere, i.e., that brought together into one system to capture two poles and opposes effect has been to destroy or cancel each other leaving the amount as the difference of the strongest to weakest.

It is therefore easy to understand why, considering the two poles of air as about equal, it is almost impossible to measure some potential with the land, the highest on the lowest remaining which can be positive or negative, copper being the strongest of one or other of these polarities.

It is on this basis that all research until now have been made, and that is why we must consider that all the time, between the two poles captured in the atmosphere and not in any mixture of systems uptake, it is possible to obtain tension and intensity, a considerable and measurable power captured separately between the two poles and remaining isolated from one another until their utilization .

In accordance with this invention, this collection is made by attraction of primary electrical currents in teh atmosphere, either positively or negatively charged, by an antenna that has two points absolutely isolated from one another, where one points to the zenith to attract negative electricity, and the other turned south, toward the equator to draws in the masses grouped in this area.

This shows and proves the existence of atmospheric electrical currents, and that it is possible to capture them with specially designed equipment primarily acting as regulators, as the persistent difficulty encountered so far in realizing their capture lies with large variations of tension in which the current atmosphere is present in space, for each of the poles, and an overload could inevitably fatal.

The patent in its present form of presentation does not cover the regulators or devices that are anticipated in the system, because they can be designed in different ways and give the same result, but covers the application and grouping of equipment operating as automatic relays with variable influence to limit current tension to that chosen for utilization.

These devices have a role associated with an overflow reservoir outside carrying excess liquid; these regulators will divert the excess voltage flows to ground.

We can still absorb these regulators having filters because they have the sole purpose of diverting the post being used , currents that are not yet known but likely that we will call for simplicity, abnormal currents. [sic...]

The patent also covers systems for regulating currents, applied to each of the poles because it is recognized that the shape and tension of the currents of the two poles are not equivalent.

The method of capture is by special antenna has directed two points in the atmosphere, as will be stated after this, with Boot ecoulement prior to the current atmosphere is also of great importance.

Finally, the current atmospheric reception is adequate even with the installation of a post, which can be expected at any location without the need for special altitude, as has been attempted in previous efforts.

We refer to the attached drawings:.

Figure 1 represents in elevation and side view an antenna that only has two points, to overcome an elevation ;

Figure 2 is the end of one of the peaks, the horizontal:

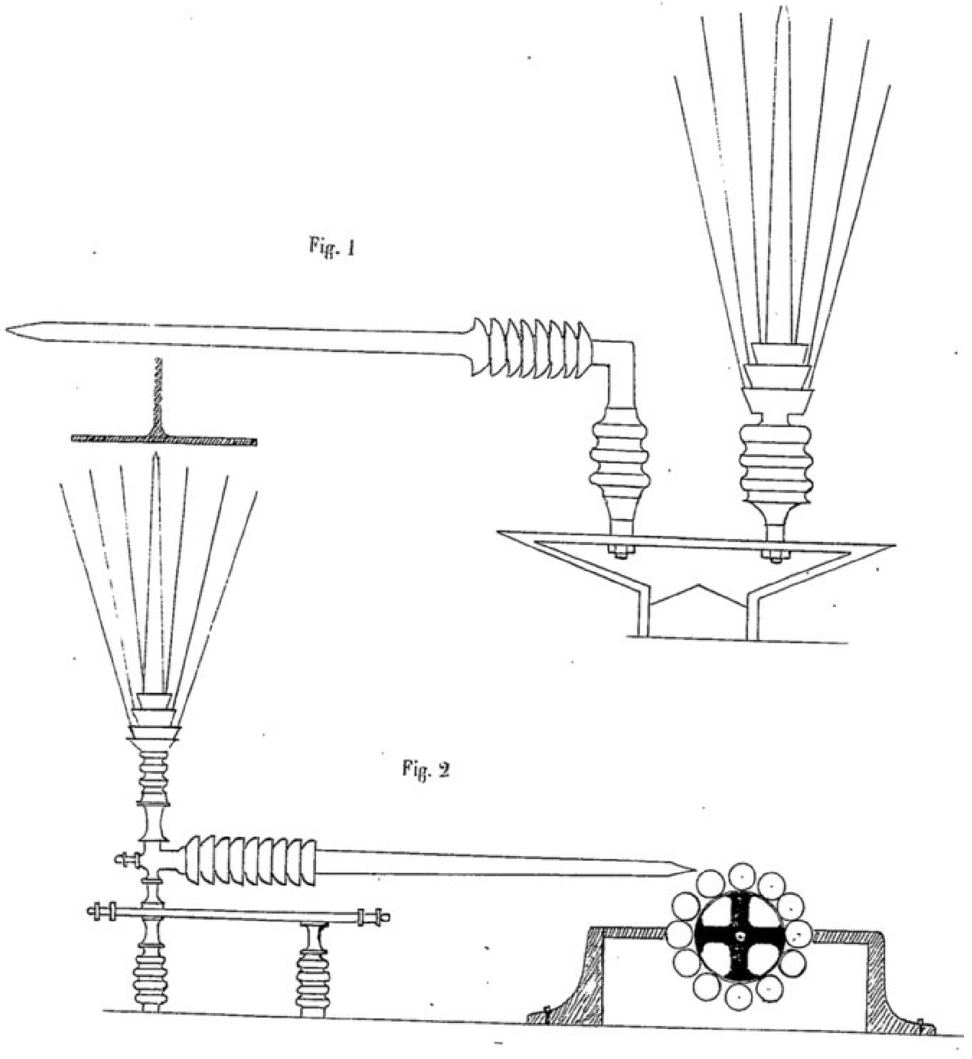

Fig. 1

Fig. 2

Figure 3 shows schematically the table on which are mounted various devices or accessories that act as regulators of current tensions.

Fig. 3.

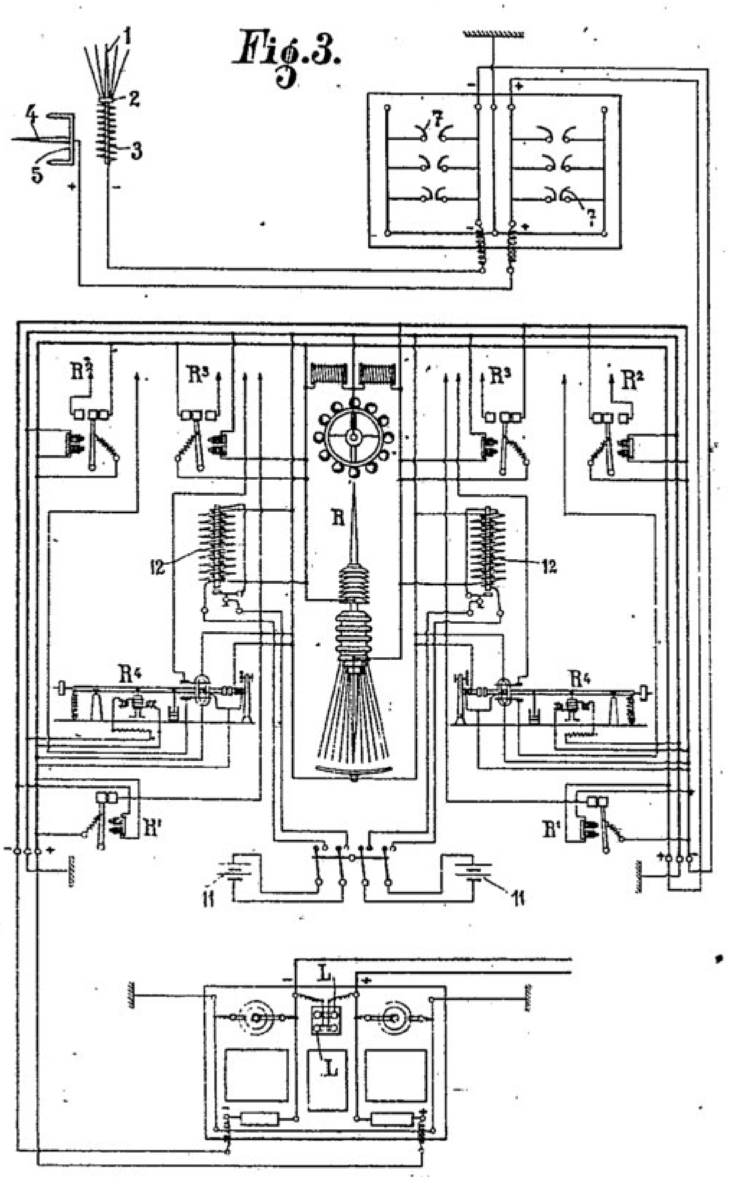
1
2
3
4
5
7
R²
R³
R
12
R⁴
R¹
11
L

Figure 4 is a scheme of surge arrestors, and

Figure 5 a table of resistances.

Fig. 4.

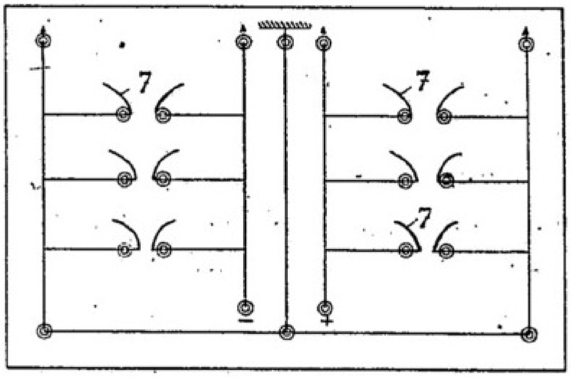

Fig. 5.

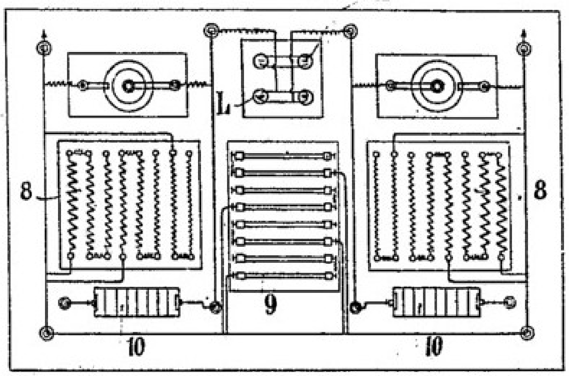

The necessary arrangements to capture atmospheric current form is as follows:

1. The **antenna** in figure 1 is formed with two peaks: the 1 is **steel**, supported by a **bronze ring** 2, surrounded by a **spiral form** 3; this point is oriented for attracting the **zenith** layers of **negative electricity** and the other edge 4 contains **copper coils** 6, and is oriented **south** to the **equator** to draw in the masses grouped in this area. To use this antenna it suffices to be raised on a mast or **pole** to a height a little higher than that of surrounding houses.

2. On the table **lightning arrestors**, contained 4, forms of 7 **copper horns** used to absorb large atmospheric discharges by certain time, as it would be dangerous to allow movement in the devices regulators.

3. On a table of **two** symmetrical systems **regulators**, Figure 3, one for each pole, to obtain and allow a regularization of large excesses in the current tensions, by automatically absorbing surges before passing over the surfacea chosen for the desired use. These regulators are represented on the drawing by the references R, R1, R2, R3, R4.

4. In the **resistance** panel in communication with regulators, these resistances are formed as **spiral coils or screens** 8, **tubes 9 of glass filled with fine coal dust , aluminum powder, sulfur and fine copper dust**. Finally, **two** special 10 **processors** and constitutes **layers** of metal forming an **absorbant** for harmful or abnormal currents.

5. As a source of energy formed **auxiliary batteries** 11 and **Ruhmkorff coils** 12 to get into the antenna sufficient attraction by means of an adjustment of intensity, which once it begins, allows the constant flow of atmospheric current in the system.

In considering Figure 3, which represents a sort of an installation scheme of the device, the reception of atmospheric electrical current will, as set out above, producing a priming of these currents. This boot is made in launching the system which is double to answer each of the poles, the current batteries 11 through 12 and reels going to the antenna.

This boot will continue until the disposal of electrical current atmosphere is evident by the one or more table lamps L of resistance shown in Figure 5.

Once the lights illuminate it is advisable to charge the batteries to boot. Natural flow and constant currents will be captured by the regulators , Figure 3; they will regularize in tension and in amperage and then directed to use in table lamps.

It has been said above that regulators R, etc., are intended and designed to automatically eliminate earth surge currents capture and abnormal currents that cannot yet be level but that these regulators could be replaced by devices with the same function.

There is therefore no need to describe in detail and especially the claims. We only specify that these regulators must be proper protection for the purposes described above.

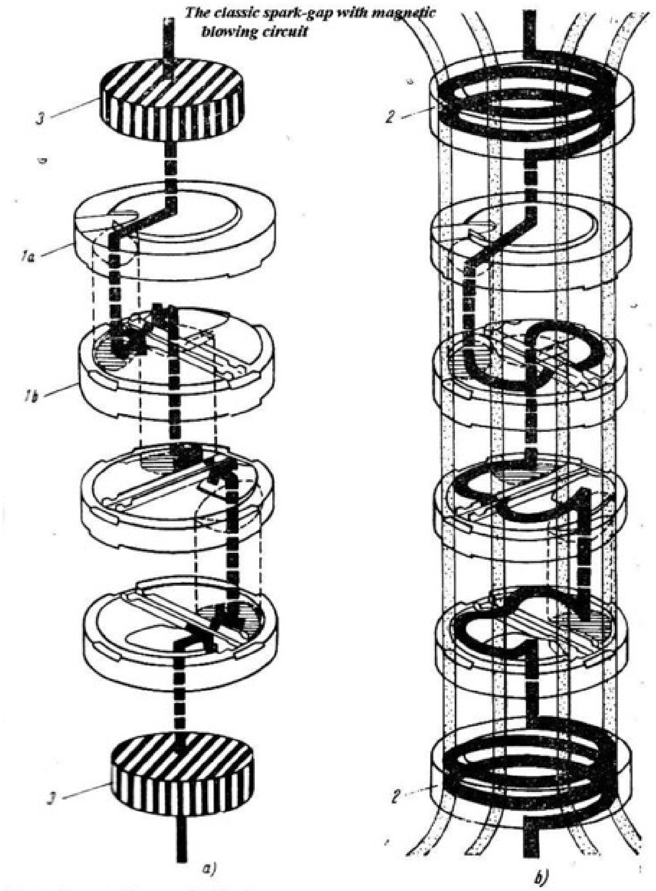

The spark-gap with magnetic blowing :
a,- the initial stage ; the impulse current send in soil through the shunt-resistences 3, because the coils 2 have an impedance too big for the impulse of 10/20 µs ;
the spark-gap 1 has two elements disc-shaped 1a and 1b, the inferior disc have one electrode and the upper disc the other electrode of the spark-gap circuit.
b,- the second stage ; the accompanying current discharge - after the impulse current ;
this accompanying current pass through the coils 2 which produce the magnetic field with the role to extend the electric arc into the slit between insulator discs 1a and 1b ; the thickness of this slit is smaller and smaller in accord with the extension of the electric arc, in this manner producing the magnifing of its electric resistance and finally, the electric arc extinction.

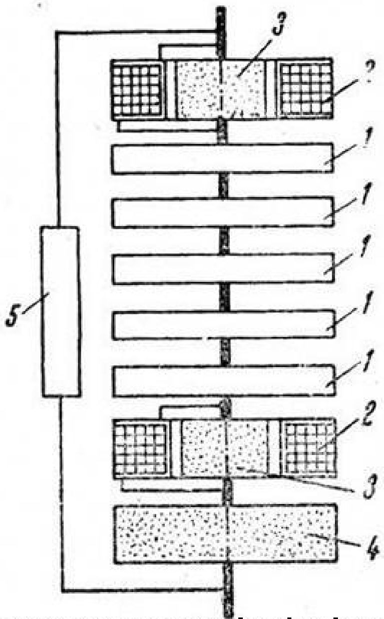

1,- the spark-gap individual devices which work in the manner to extend the electric arc of the accompanying electric current ;
2,- the coils which produce the magnetic field for the magnetic blowing ;
3,- the electric shunt-resistances in parallel connection on the blowing coils, with the role of reducing the current inside this coils
4,- the disc with unlinear resistance ;
5,- electrical resistance for the homogenous distribution of the electric field in all the elements of the spark-gap.

In Jules Guillot scheme, that kind of device could be used for leading the current.

Spark-gap element with circuit for the magnetic blowing of the electric fascicle

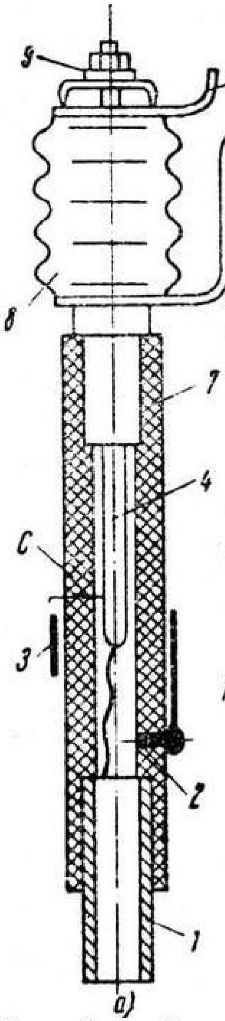

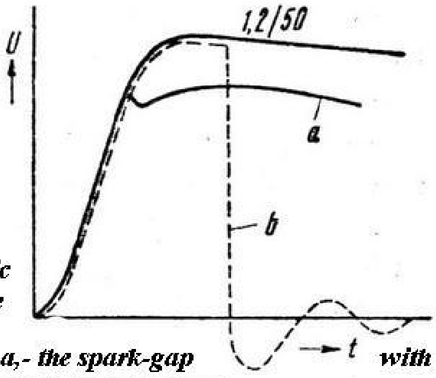

a,- the spark-gap with variable resistance ;
b,- the tube spak-gap ;
t - the time

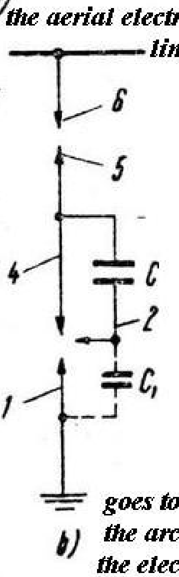

The tube spark-gap has the aerial spark gap with the electrodes 5 and 6 disposed at the insulator 8 extremities ; and one arc-over spark-gap device between the rod 4 and the electrode 1.

At start, the first spark apear in the aerial spark-gap and then, the electric dicharge goes to the gas-producing tube 7, where it makes the arc-over. The local pressure in the gas allow the electric arc extinction, after managing the over electric tension into soil. The capacity C which is created between the external ring 3 and the rod 4 is bigger than the Capacity C_1 between 1 an 2 in that manner, when the overtension apears, the auxiliary spark-gap 1, 2 will start working and then, the main spark-gap 1, 4. In this circuit, the electrode 9 is connected at the aerial electric line and the electrode 1 at the Earth.

In the image above, it is an classic tube spark-gap used for the protection of aerial electric cables. The atmospheric electricity is discharged in soil and lost... The Guillot's inventions used that kind of devices for leading and absorbing of the atmospheric electricity.

DARIDA, Gillbert : "Practical Utilization of Atmospheric Electricity";
(in *The Invention Encyclopedia,* pp. 204-207 ; 1930, Geo. Constantinescu, Ed.) --

Pământul ca și celelalte astre, posedă o electricitate proprie negativă în interior, iar aerul care-l înconjoară este pozitiv. Potențialul sau tensiunea electrică a aerului se mărește pe măsură ce ne ridicăm în înălțime, astfel că putem spune că intensitatea electrică este proporțională cu înălțimea (după lucrările remarcabile ale lui Franklin, Quetelet, Lord Kelvin, Mascart, Joubert și a altor mari savanți fizicieni).

Observațiile făcute recent, au arătat că aerul regiunilor înalte dela 6000—7000 m este foarte încărcat cu electricitate pozitivă, a cărei explicație se datorește frecării fotosferei sau atmosferei terestre, prin învârtirea pământului în jurul soarelui cu mai mult de o sută de mii de km pe oră. Această frecare directă în fluidul eteric ambiant, face ca aerul să se încarce cu electricitate pozitivă prin influență, iar pământul să se încarce negativ.

Intre cele două medii, fluidic (aerul) și (pământul) materie, știința ne spune, că în partea de jos a atmosferei, pe timp frumos un centimetru cub de aer ionizat, conține circa 800 particule cu ioni pozitivi și 680 ioni negativi (electroni).

Pământul se comportă ca un enorm conductor încărcat negativ, respingând electroni și atrăgând ionii pozitivi. Această atracție a ionilor pozitivi determină un curent electric, zis și curent de convențiune. E un fel de bombardament invizibil supus variațiunilor zilei și a sezoanelor, putându-l evalua aproximativ la o densitate mijlocie remarcabilă, dela 3×10^{16} sau a 30 quatrilionime de amper pe cm cub, ceea ce dă pentru întreagă suprafață a pământului un curent formidabil, de 1500 de amperi.

Se pune întrebarea cum un astfel de curent poate să-și întreție mereu același sens? Aceasta este enigma pe care ne-o punem, studiind curenții naturali ai atmosferei noastre, presupunem că poate să fie o legătură a acțiunii emanațiunilor radioactive. Aceste emanațiuni fiind din gaze radioactive grele, se găsesc de obiceiu în mare abundență aproape de pământ ceea ce ne explică marea ionizare observată în general în grote și caverne.

Ionizarea atmosferică poate de asemenea fi produsă în parte, de acțiunea razelor X, foarte pătrunzătoare și de diferite substanțe radioactive închise în subsolul pământului. De asemenea poate să mai fie, ținând seama de acțiunea «Calei Lactee» și de ionizarea luminei solare, foarte bogată în radiațiuni ultraviolete precum și a electronilor emiși direct din radierea căldurii solare de circa 6000 de grade.

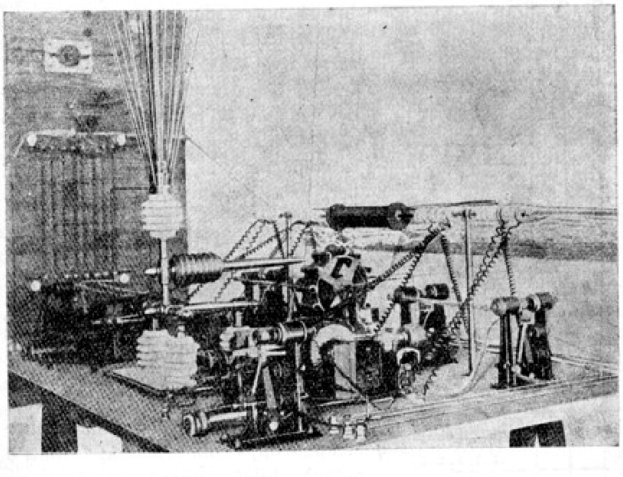

Fig. 108

Captarea electricității atmosferice a fost utilizată în Franța prin cabluri aeriene montate pe Mont Blank, iar în Germania prin cabluri susținute de baloane captive.

Sunt mai multe sisteme, vom da pe scurt pe cel mai ingenios procedeu, inventat de ing. Jules Guillot, bazat pe sifonul electric. Procedeul consistă de a pompa direct din atmosferă, electricitate

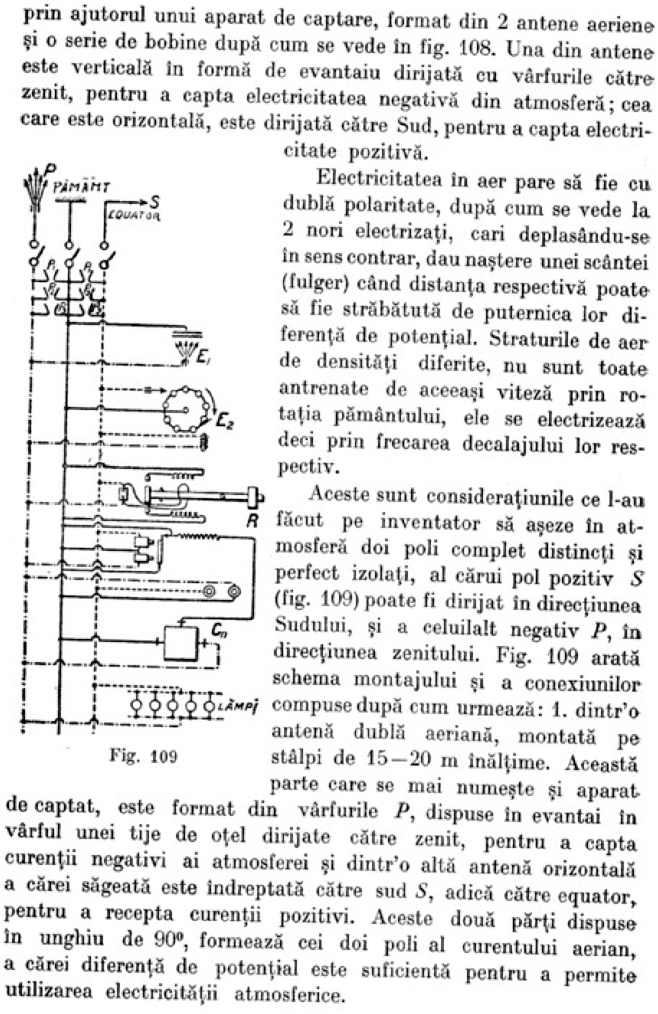

prin ajutorul unui aparat de captare, format din 2 antene aeriene şi o serie de bobine după cum se vede în fig. 108. Una din antene este verticală în formă de evantaiu dirijată cu vârfurile către zenit, pentru a capta electricitatea negativă din atmosferă; cea care este orizontală, este dirijată către Sud, pentru a capta electricitate pozitivă.

P
PĂMÂNT
S
ECUATOR
E_1
E_2
R
C_n
LĂMPI

Fig. 109

Electricitatea în aer pare să fie cu dublă polaritate, după cum se vede la 2 nori electrizaţi, cari deplasându-se în sens contrar, dau naştere unei scântei (fulger) când distanţa respectivă poate să fie străbătută de puternica lor diferenţă de potenţial. Straturile de aer de densităţi diferite, nu sunt toate antrenate de aceeaşi viteză prin rotaţia pământului, ele se electrizează deci prin frecarea decalajului lor respectiv.

Aceste sunt consideraţiunile ce l-au făcut pe inventator să aşeze în atmosferă doi poli complet distincţi şi perfect izolaţi, al cărui pol pozitiv *S* (fig. 109) poate fi dirijat în direcţiunea Sudului, şi a celuilalt negativ *P*, în direcţiunea zenitului. Fig. 109 arată schema montajului şi a conexiunilor compuse după cum urmează: 1. dintr'o antenă dublă aeriană, montată pe stâlpi de 15—20 m înălţime. Această parte care se mai numeşte şi aparat de captat, este format din vârfurile *P*, dispuse în evantai în vârful unei tije de oţel dirijate către zenit, pentru a capta curenţii negativi ai atmosferei şi dintr'o altă antenă orizontală a cărei săgeată este îndreptată către sud *S*, adică către equator, pentru a recepta curenţii pozitivi. Aceste două părţi dispuse în unghiu de 90°, formează cei doi poli al curentului aerian, a cărei diferenţă de potenţial este suficientă pentru a permite utilizarea electricităţii atmosferice.

French Patent # 551,882

Appareil Capteur de Courants Electriques dans L'Atmosphere

BREVET D'INVENTION.

XII. — Instruments de précision, électricité.

5. — Production de l'électricité, moteurs électriques.

N° 551.882

Appareil capteur de courants électriques dans l'atmosphère.

M. Jules GUILLOT résidant en France (Tarn-et-Garonne).

Demandé le 27 septembre 1921, à 15^{h} 20^{m}, à Paris.
Délivré le 15 janvier 1923. — Publié le 16 avril 1923.
[Brevet d'invention dont la délivrance a été ajournée en exécution de l'art. 11 § 7 de la loi du 5 juillet 1844 modifiée par la loi du 7 avril 1902.]

Considérant le globe terrestre comme l'induit d'une dynamo, dont l'éther supraterrestre est l'inducteur, des courants électriques circulent dans l'intervalle c'est-à-dire dans la couche atmosphérique; l'inventeur capte au moyen de l'appareil décrit ci-dessous deux de ces courants parfaitement distincts et élimine les autres.

L'appareil comprend :

1° Le capteur aérien; 2° une série de parafoudres; 3° les régulateurs; 4° un tableau de résistances; 5° les aspirateurs; 6° les témoins.

1° Le capteur aérien monté sur un poteau est composé d'une pointe d'acier aimanté fixée par scellement sur un isolateur porcelaine et cerclée à sa base par une bague de bronze sur laquelle sont vissées 32 aiguilles de fer doux, le tout formant l'antenne d'un pôle. A côté et isolé de la première, formant le 2e pôle, est fixée une pointe d'acier aimanté dans la position horizontale et dirigée vers le sud dont la base coudée à 45° est scellée sur un isolateur porcelaine. Cette pointe est également cerclée d'une bague de cuivre à encoches fig. 1.

2° Une série de 6 parafoudres à cornes entre chaque pôle et la terre et à écartements variés.

3° Un premier régulateur formé d'un appareil absolument semblable au capteur aérien, mais ayant les deux pôles superposés, et, opposé à la pointe verticale, un disque de cuivre relié à la terre. Opposé à la pointe horizontale, une roue à jante d'étain sur laquelle sont soudés alternativement 16 tubes composés eux-mêmes de deux métaux soudés : cuivre-plomb et fer-plomb. Cette roue est reliée à la terre par son pivot fig. 2.

Un deuxième régulateur formé d'appareils à déclanchement automatique au nombre de 3 par pôles et deux basculateurs également par pôles.

4° Un tableau de résistances composé de fils de maillechort en boudins de différentes sections, de tubes en verre contenant des poussières de cuivre, du charbon de cornue et de la fleur de soufre.

5° Les aspirateurs composés d'une boîte en bois sur chaque pôle, contenant un vase de porcelaine dans lequel par couches superposées et isolées au mica est placé du mercure, de l'étain, du charbon, du cuivre et du soufre, le tout traversé par une tige de cuivre.

6° Les témoins composés de lampes à incandescence ordinaires.

RÉSUMÉ.

Par la pointe au zénith et par la pointe au sud nous canalisons deux courants formant les deux pôles. On protège l'ensemble par des parafoudres. On régularise le débit par un régulateur semblable et par des appareils à déclanchement commandant chacun des résistances appropriées de la nature du courant nuisible dont ces appareils sont chargés. Les courants épurés sont conduits aux témoins par des fils de cuivre ordinaire.

Jules GUILLOT,
Moissac (Tarn-et-Garonne).

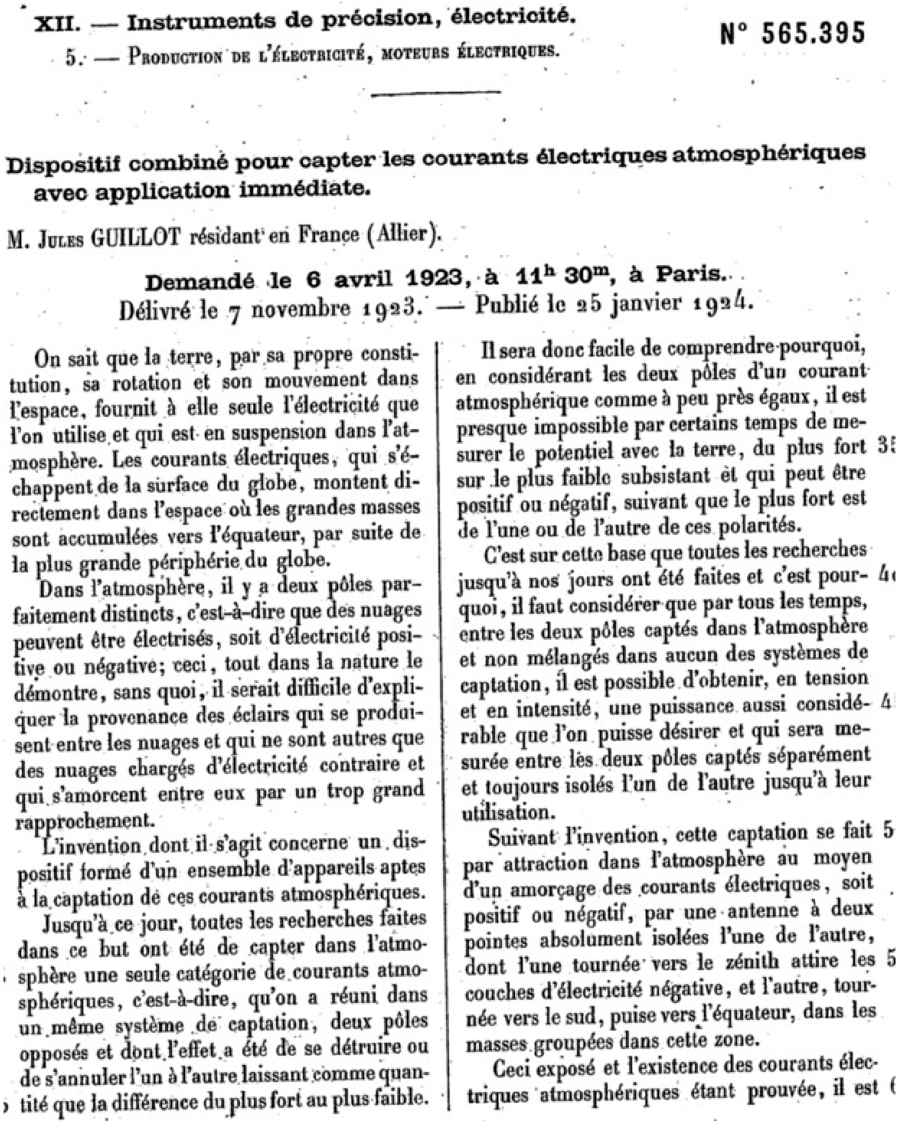

XII. — Instruments de précision, électricité.

5. — Production de l'électricité, moteurs électriques.

N° 565.395

Dispositif combiné pour capter les courants électriques atmosphériques avec application immédiate.

M. Jules GUILLOT résidant en France (Allier).

Demandé le 6 avril 1923, à 11h 30m, à Paris.
Délivré le 7 novembre 1923. — Publié le 25 janvier 1924.

On sait que la terre, par sa propre constitution, sa rotation et son mouvement dans l'espace, fournit à elle seule l'électricité que l'on utilise et qui est en suspension dans l'atmosphère. Les courants électriques, qui s'échappent de la surface du globe, montent directement dans l'espace où les grandes masses sont accumulées vers l'équateur, par suite de la plus grande périphérie du globe.

Dans l'atmosphère, il y a deux pôles parfaitement distincts, c'est-à-dire que des nuages peuvent être électrisés, soit d'électricité positive ou négative; ceci, tout dans la nature le démontre, sans quoi, il serait difficile d'expliquer la provenance des éclairs qui se produisent entre les nuages et qui ne sont autres que des nuages chargés d'électricité contraire et qui s'amorcent entre eux par un trop grand rapprochement.

L'invention dont il s'agit concerne un dispositif formé d'un ensemble d'appareils aptes à la captation de ces courants atmosphériques.

Jusqu'à ce jour, toutes les recherches faites dans ce but ont été de capter dans l'atmosphère une seule catégorie de courants atmosphériques, c'est-à-dire, qu'on a réuni dans un même système de captation, deux pôles opposés et dont l'effet a été de se détruire ou de s'annuler l'un à l'autre laissant comme quantité que la différence du plus fort au plus faible.

Il sera donc facile de comprendre pourquoi, en considérant les deux pôles d'un courant atmosphérique comme à peu près égaux, il est presque impossible par certains temps de mesurer le potentiel avec la terre, du plus fort sur le plus faible subsistant et qui peut être positif ou négatif, suivant que le plus fort est de l'une ou de l'autre de ces polarités.

C'est sur cette base que toutes les recherches jusqu'à nos jours ont été faites et c'est pourquoi, il faut considérer que par tous les temps, entre les deux pôles captés dans l'atmosphère et non mélangés dans aucun des systèmes de captation, il est possible d'obtenir, en tension et en intensité, une puissance aussi considérable que l'on puisse désirer et qui sera mesurée entre les deux pôles captés séparément et toujours isolés l'un de l'autre jusqu'à leur utilisation.

Suivant l'invention, cette captation se fait par attraction dans l'atmosphère au moyen d'un amorçage des courants électriques, soit positif ou négatif, par une antenne à deux pointes absolument isolées l'une de l'autre, dont l'une tournée vers le zénith attire les couches d'électricité négative, et l'autre, tournée vers le sud, puise vers l'équateur, dans les masses groupées dans cette zone.

Ceci exposé et l'existence des courants électriques atmosphériques étant prouvée, il est

possible de les capter à la condition de disposer d'appareils spéciaux étudiés surtout pour jouer le rôle de régulateurs, attendu que la difficulté persistante rencontrée jusqu'alors pour réaliser la captation incombe aux grandes variations de tensions sous lesquelles les courants atmosphériques se présentent dans l'espace, pour chacun des pôles, et dont l'écoulement dans un poste de captation serait inévitablement funeste.

Le présent brevet sous sa forme de présentation doit donc couvrir, non pas le ou les dispositifs régulateurs qui sont prévus dans le système, car ils peuvent être conçus d'une façon toute différente et donner le même résultat, mais sur l'application raisonnée et le groupement d'appareils fonctionnant automatiquement et intervenant chacun comme relais d'influence variable pour limiter la tension des courants captés dans l'atmosphère à celle choisie pour utilisation.

Ces appareils ont un rôle assez connexe avec celui de trop-pleins sur un réservoir qui acheminent à l'extérieur l'excès de liquide; ici ces régulateurs détournent l'excès de tension des courants à la terre.

On peut encore assimiler ces régulateurs à des filtres car ils ont pour but de détourner du poste d'utilisation, des courants qui ne sont pas encore de nature connue mais que pour simplicité on appellera courants parasites ou anormaux.

Le brevet portera aussi sur le système de régulation des courants, appliqué à chacun des pôles, car il est reconnu que la forme et la tension des courants aux deux pôles ne sont pas équivalentes.

Le mode de captation par antenne spéciale à deux pointes orientées dans l'atmosphère, comme il sera indiqué ci-après, avec amorçage préalable pour l'écoulement des courants atmosphériques est aussi d'une grande importance.

Enfin la réception des courants atmosphériques se fait à l'endroit même de l'installation du poste; celui-ci peut être prévu en un lieu quelconque sans nécessiter d'altitude particulière, comme cela a été tenté dans des essais préconçus.

Sur les dessins ci-annexés auxquels on se réfère

La fig. 1 représente en élévation et en coupe partielle l'antenne seule, à deux pointes, surmontant un poteau d'élévation;

La fig. 2 est le plan de l'une des pointes, celle horizontale:

La fig. 3 montre schématiquement le tableau sur lequel sont montés les divers appareils ou accessoires qui jouent le rôle de régulateurs des tensions de courants.

La fig. 4 est un schéma de parafoudre, et

La fig. 5 un tableau de résistances.

Le dispositif nécessaire à la captation des courants atmosphériques est formé comme suit :

1° D'une antenne, fig. 1, formée de deux pointes dont l'une 1 est en acier, supportée par une bague de bronze 2 entourée d'une spirale 3 formant self; cette pointe est tournée vers le zénith pour attirer les couches d'électricité négative; l'autre pointe 4 comporte un cadre 5 de huit branches supportant des spirales de cuivre 6 et est tournée vers le sud pour puiser vers l'équateur dans les masses groupées dans cette zone. Pour l'utilisation de cette antenne il suffira qu'elle soit portée par un mât ou poteau de hauteur un peu supérieure à celle des habitations environnantes.

2° D'un tableau de parafoudre, fig. 4, formé de cornes de cuivre 7 servant à absorber les tros grosses décharges atmosphériques que par certains temps, il serait dangereux de laisser circuler dans les appareils régulateurs.

3° D'une table de deux systèmes symétriques de régulateurs, fig. 3, un pour chaque pôle, qui permet d'obtenir une régularisation assez grande dans les tensions des courants en absorbant automatiquement les surtensions dépassant la tension choisie pour l'utilisation désirée. Ces régulateurs sont représentés sur le dessin par les références R, R^1, R^2, R^3, R^4.

4° D'un tableau de résistance en communication avec les régulateurs, ces résistances sont formées de spirales de maillechor 8, de tubes 9 de verre remplis de poussière fine de charbon de cornue, de poudre d'aluminium, de soufre et de poussière fine de cuivre; enfin, deux appareils spéciaux 10 formant transformateurs et constitués de couches de métaux absorbant certains courants nuisibles ou anormaux.

5° D'une source auxiliaire d'énergie formée de piles 11 et de bobines Ruhmkorff 12 permettant d'obtenir dans l'antenne une attraction suffisante au moyen d'un dispositif de

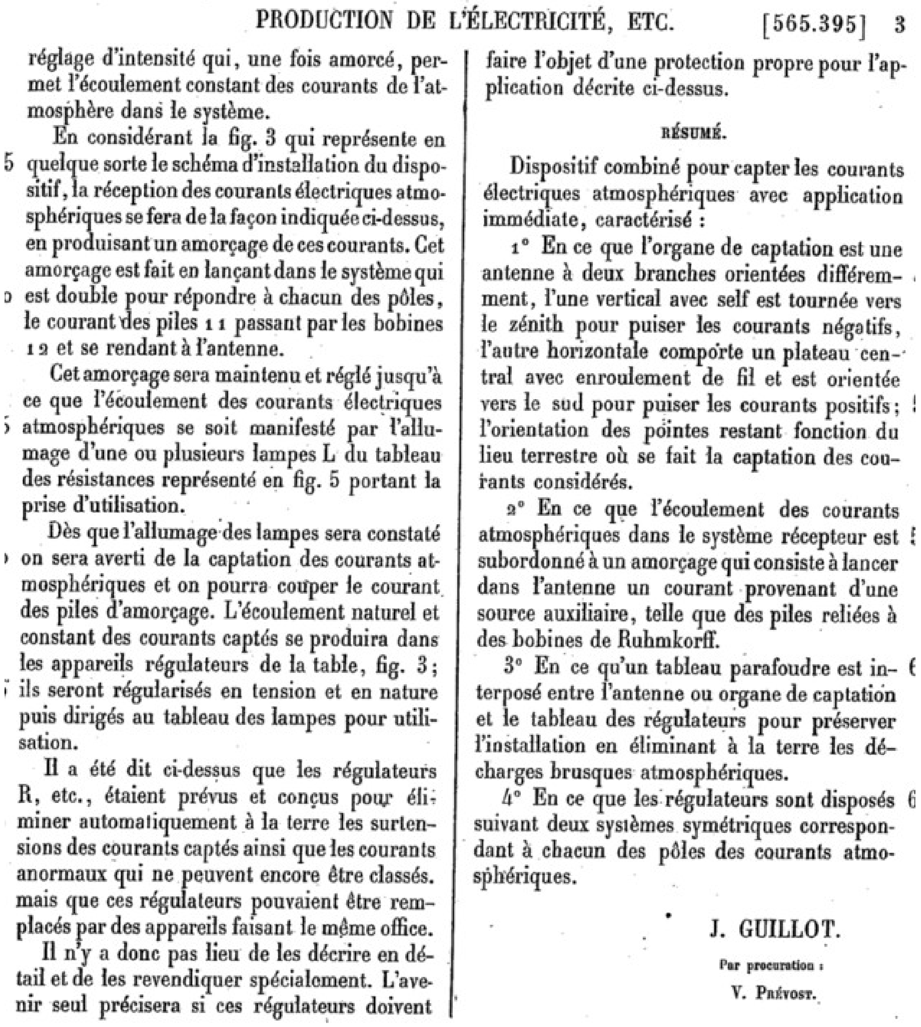

réglage d'intensité qui, une fois amorcé, permet l'écoulement constant des courants de l'atmosphère dans le système.

En considérant la fig. 3 qui représente en quelque sorte le schéma d'installation du dispositif, la réception des courants électriques atmosphériques se fera de la façon indiquée ci-dessus, en produisant un amorçage de ces courants. Cet amorçage est fait en lançant dans le système qui est double pour répondre à chacun des pôles, le courant des piles 11 passant par les bobines 12 et se rendant à l'antenne.

Cet amorçage sera maintenu et réglé jusqu'à ce que l'écoulement des courants électriques atmosphériques se soit manifesté par l'allumage d'une ou plusieurs lampes L du tableau des résistances représenté en fig. 5 portant la prise d'utilisation.

Dès que l'allumage des lampes sera constaté on sera averti de la captation des courants atmosphériques et on pourra couper le courant des piles d'amorçage. L'écoulement naturel et constant des courants captés se produira dans les appareils régulateurs de la table, fig. 3; ils seront régularisés en tension et en nature puis dirigés au tableau des lampes pour utilisation.

Il a été dit ci-dessus que les régulateurs R, etc., étaient prévus et conçus pour éliminer automatiquement à la terre les surtensions des courants captés ainsi que les courants anormaux qui ne peuvent encore être classés. mais que ces régulateurs pouvaient être remplacés par des appareils faisant le même office. Il n'y a donc pas lieu de les décrire en détail et de les revendiquer spécialement. L'avenir seul précisera si ces régulateurs doivent faire l'objet d'une protection propre pour l'application décrite ci-dessus.

RÉSUMÉ.

Dispositif combiné pour capter les courants électriques atmosphériques avec application immédiate, caractérisé :

1° En ce que l'organe de captation est une antenne à deux branches orientées différemment, l'une vertical avec self est tournée vers le zénith pour puiser les courants négatifs, l'autre horizontale comporte un plateau central avec enroulement de fil et est orientée vers le sud pour puiser les courants positifs; l'orientation des pointes restant fonction du lieu terrestre où se fait la captation des courants considérés.

2° En ce que l'écoulement des courants atmosphériques dans le système récepteur est subordonné à un amorçage qui consiste à lancer dans l'antenne un courant provenant d'une source auxiliaire, telle que des piles reliées à des bobines de Ruhmkorff.

3° En ce qu'un tableau parafoudre est interposé entre l'antenne ou organe de captation et le tableau des régulateurs pour préserver l'installation en éliminant à la terre les décharges brusques atmosphériques.

4° En ce que les régulateurs sont disposés suivant deux systèmes symétriques correspondant à chacun des pôles des courants atmosphériques.

J. GUILLOT.

Par procuration :

V. Prévost.

Parts List

Steel Rods, magnetic, 2
Insulator, porcelain
Ring, bronze
Rods, soft iron, 32
Ring, copper
Surge Protector
Disk, copper
Ring, tin
Tubes, steel
Rx Cu-Pb, Fe-Pb
Breakers, automatic
Resistances : Wire Mesh, Glass Tube. Cu, C, S powder
Vacuum Cleaners : Box, Jar, Mica &
Tube, copper;
Rx : Hg, Sn, C, Cu, S
Battery
Ignition Coil

Your Support Maintains this Service -- and Your Survival ...

The Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Everything @ rexresearch.com, plus the Bonus Files CD !
<< $13, Postpaid Anywhere >> from :

Rex Research, POB 19250, Jean, NV 89019 USA

BUILD A FREE WEBSITE OF YOUR OWN ON TRIPOD. Click here to find a Relay For Life near you.

"CRYSTAL RADIOS AND SPARK GAP TRANSMITTERS"

"Built to order by, AIRWAVE EXPLORER LABS"

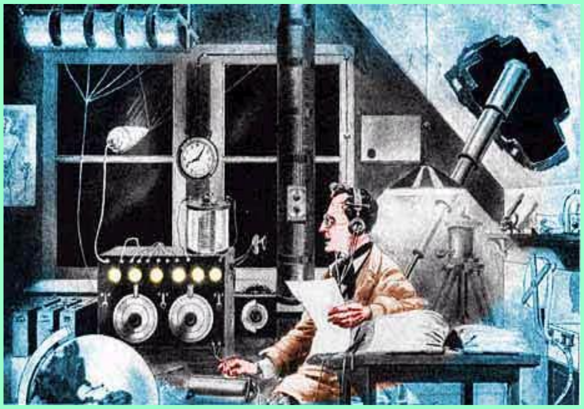

These are a sample of some of the Crystal Radios I build. My Father and Uncle had a shop in Jackson, Michigan from 1919 until 1929. They designed and built custom Crystal Radios to sell. I am attempting to relive those Golden Times & Days of Radio by following in their footsteps. I have been a Ham Radio Operator for over 65 years now. This has given me the experience to hand build these radios. I am trying to follow the schematics and designs my Father used. I use as many vintage parts as I can. I try to always use old tongue and groove, box joint, finger joint and dove tail boxes. I test and "Pot" my own crystals. I can build you a radio per your specifications or mine. Using your box or mine. These spark gap transmitters are meant to be for display. Not to be connected to an antenna or ground and operated. It is illegal to operate a spark gap transmitter. I accept, PAY PAL, checks, or Money Orders. So light up that old oil lamp, put on your headphones and travel back in time to the Golden Age of Radio.

"The "Sky Pilot", is a very sensitive Crystal Radio. Cabinet has hinged top. and has vintage parts and an S-meter. SOLD-SOLD"

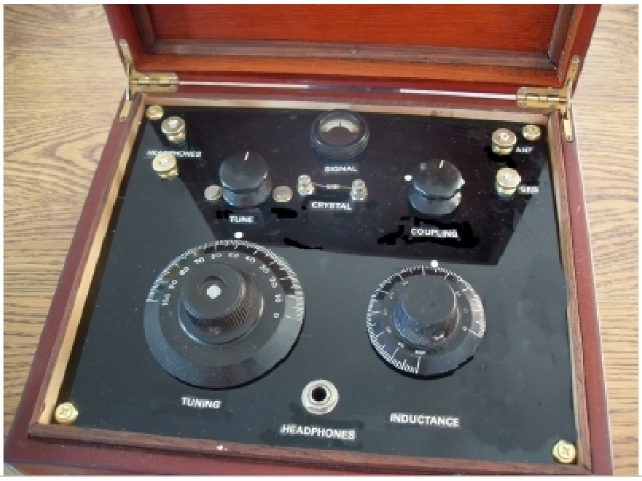
HEADPHONES
SIGNAL
ANT
GND
TUNE
CRYSTAL
COUPLING
TUNING
HEADPHONES
INDUCTANCE

$289.00 + Shipping

"The "Little Boy", is a very sensitive Crystal Radio. Cabinet has hinged top. and has vintage parts and a sliding inductor above tune knob for coverage below and above the Broadcast Band. Price is $80 including shipping and insurance, email---k8cix@aol.com "

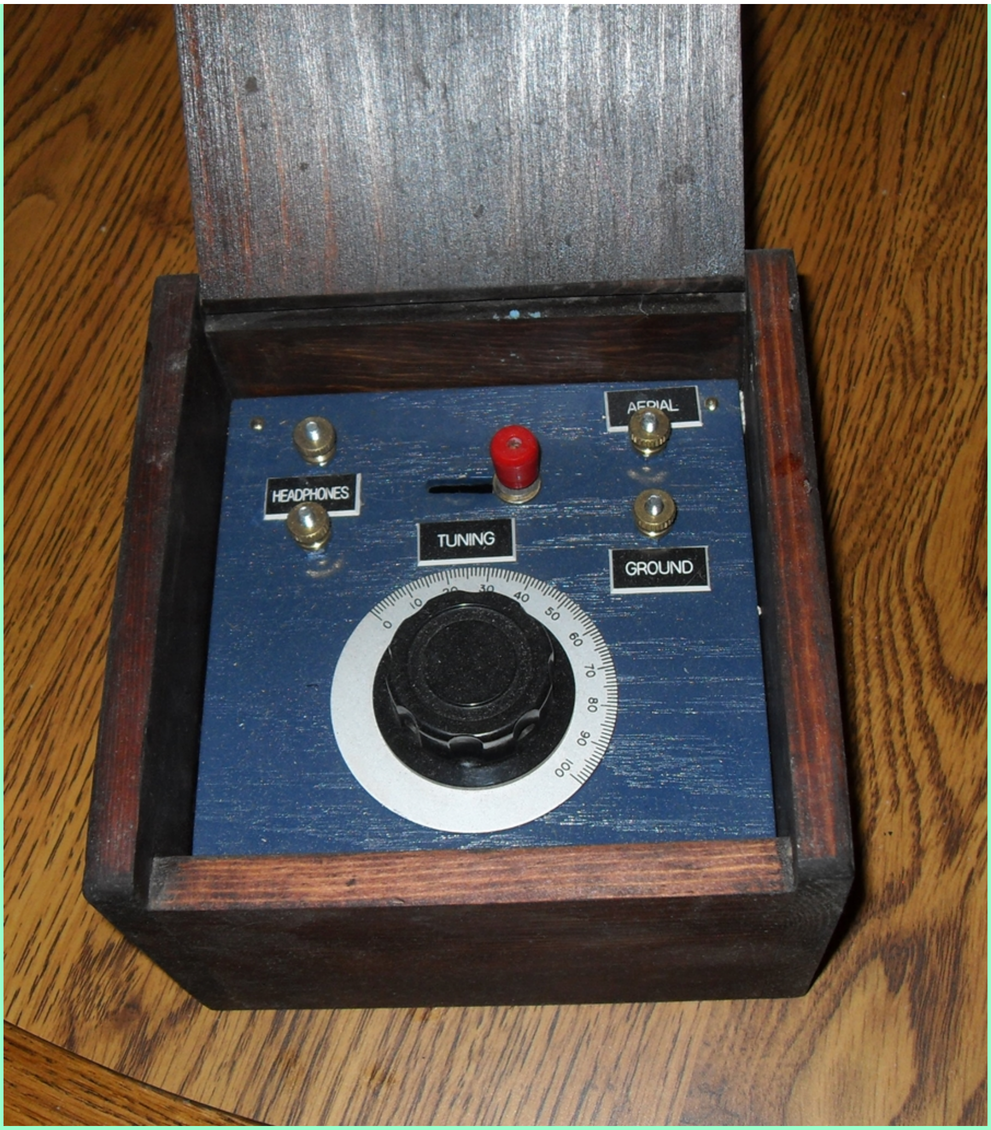
AERIAL
HEADPHONES
TUNING
GROUND
0
10
20
30
40
50
60
70
80
90
100

email for price, k8cix@aol.com

Wave Raker, Super Crystal Radio
email for price, k8cix@aol.com

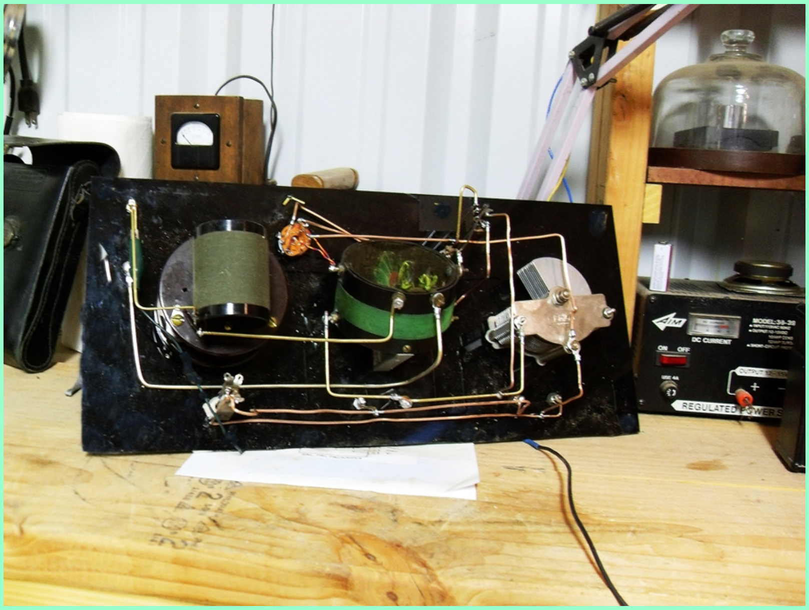

"The "Wave Raker". This is a super sensitive and selective Crystal Radio, special build. Vintage parts, 1 variometer, 1 variocoupler. Vintage tuning condenser, binding posts, knobs and buss wiring throughout. Has 2 detector posts on panel for either crystal diode or outboard detector, included. Beautiful radio, beautiful panel and gorgeous vintage cabinet. Cabinet has hinged top. Radio has antenna tuner and wavetrap. $365.00 includes shipping and insurance. k8cix@aol.com "

"Night Scanner", Crystal Radio, built for John, N3EHL."

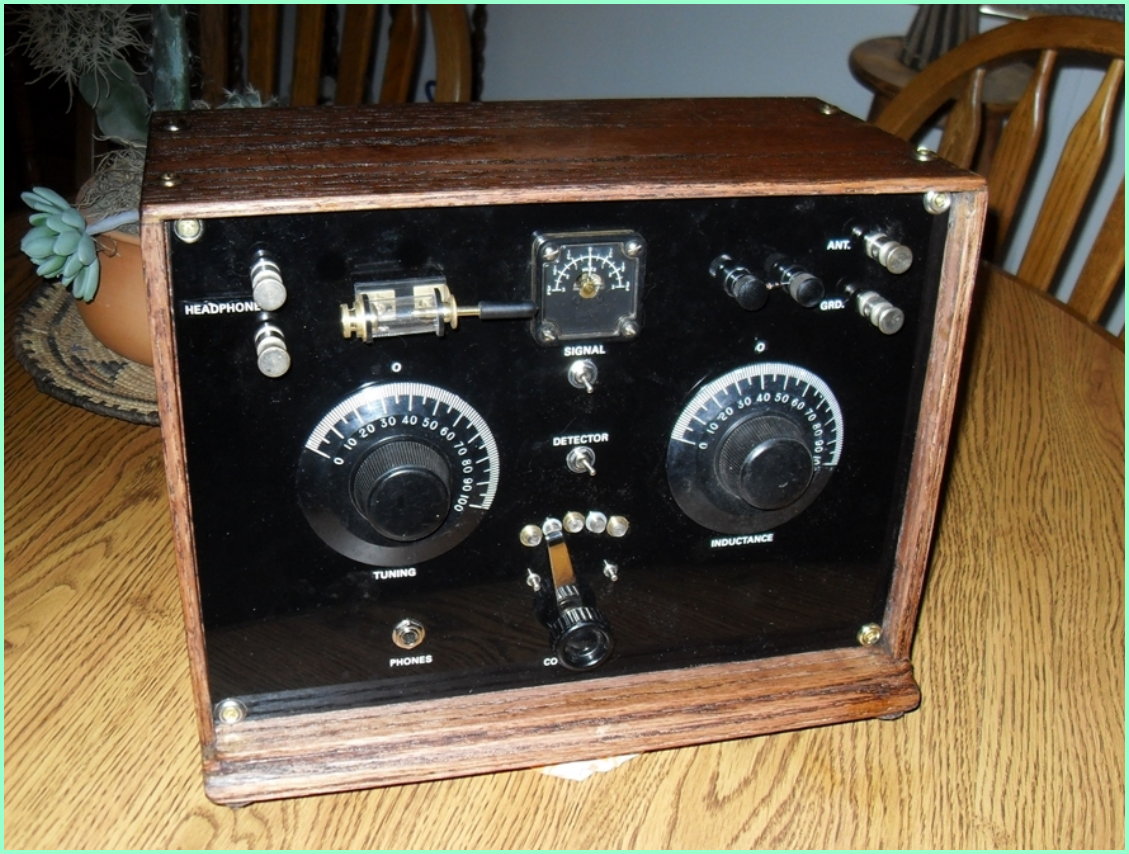

"New, 50KV Spark Gap Transmitter for sale Demo only, no ant. or grd. is allowed by FCC. Sold to ALEX"

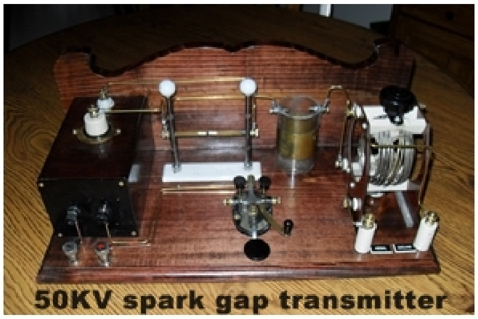
50KV spark gap transmitter

leyden jar condenser

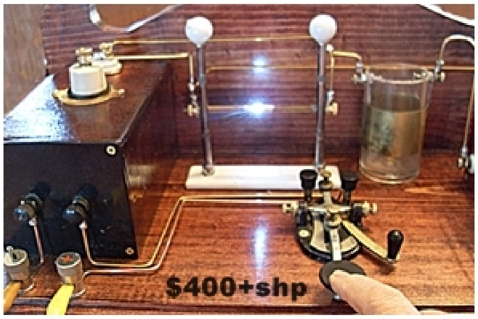

"Dodd Vintage Crystal Radio, built for Jim, N0XE."

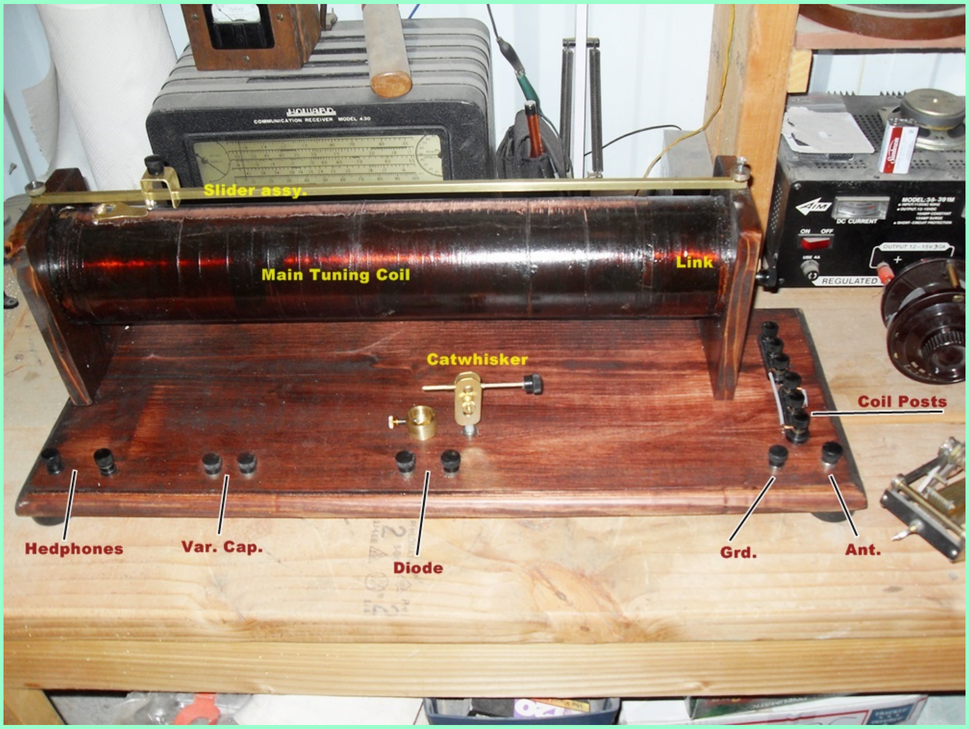
Slider assy.
Main Tuning Coil
Link
Catwhisker
Coil Posts
Hedphones
Var. Cap.
Diode
Grd.
Ant.

"201A vintage 1 tube amp., built for Jim, N0XE."

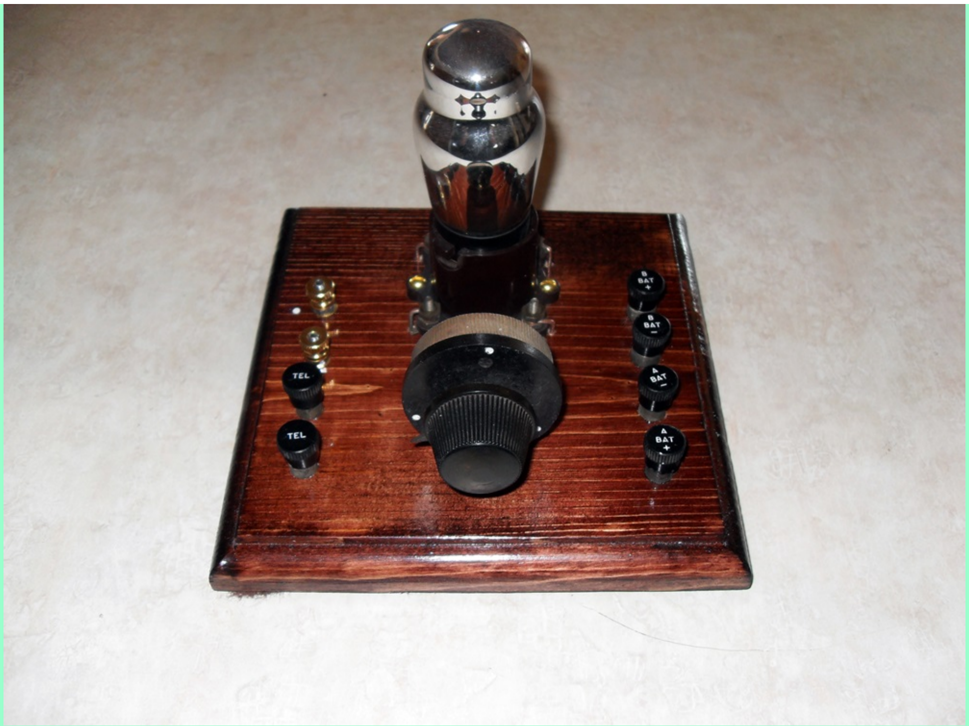
TEL
TEL
B BAT +
B BAT -
A BAT -
A BAT +

"Cohere Receiver built for Jim in New York."

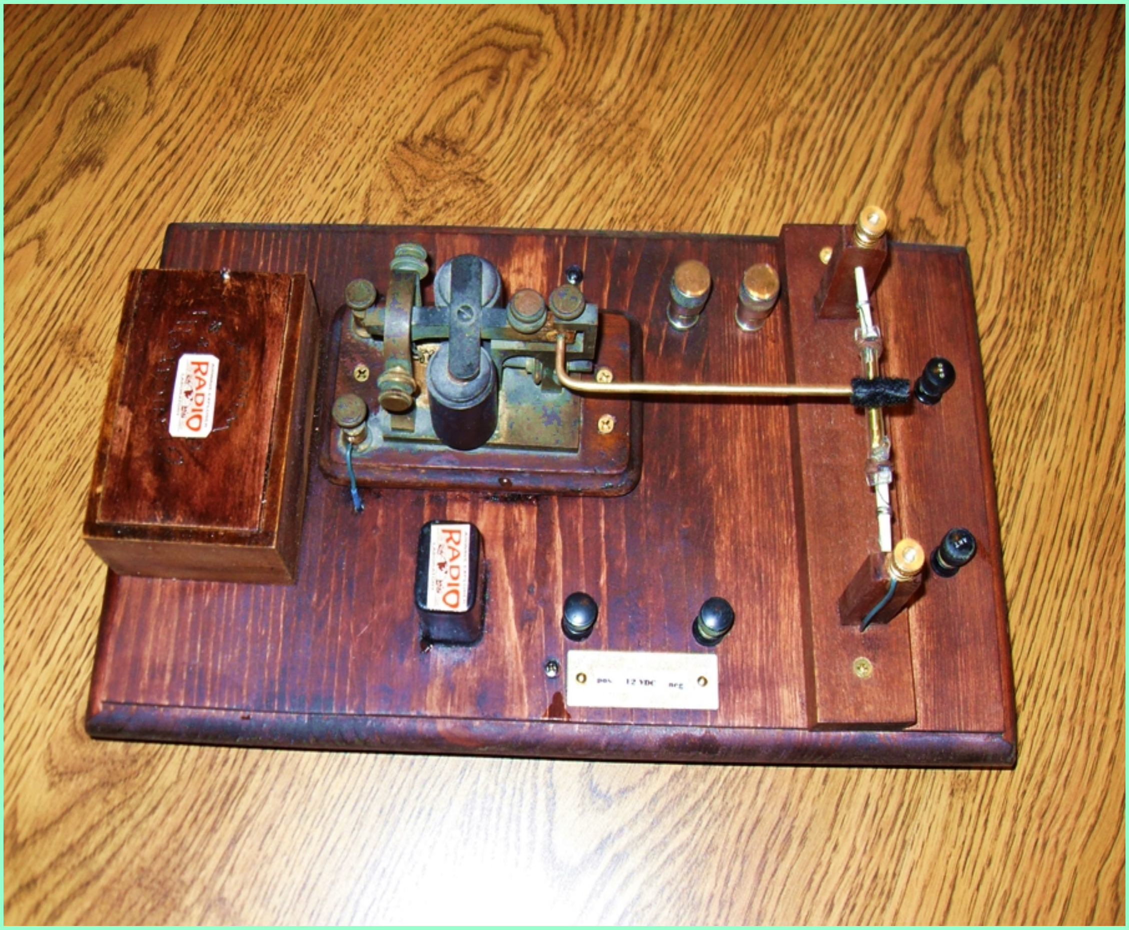
RADIO
RADIO
pos 12 VDC neg

"Ducretet Crystal Radio, French, reproduction, ordered by Neil K8RX"

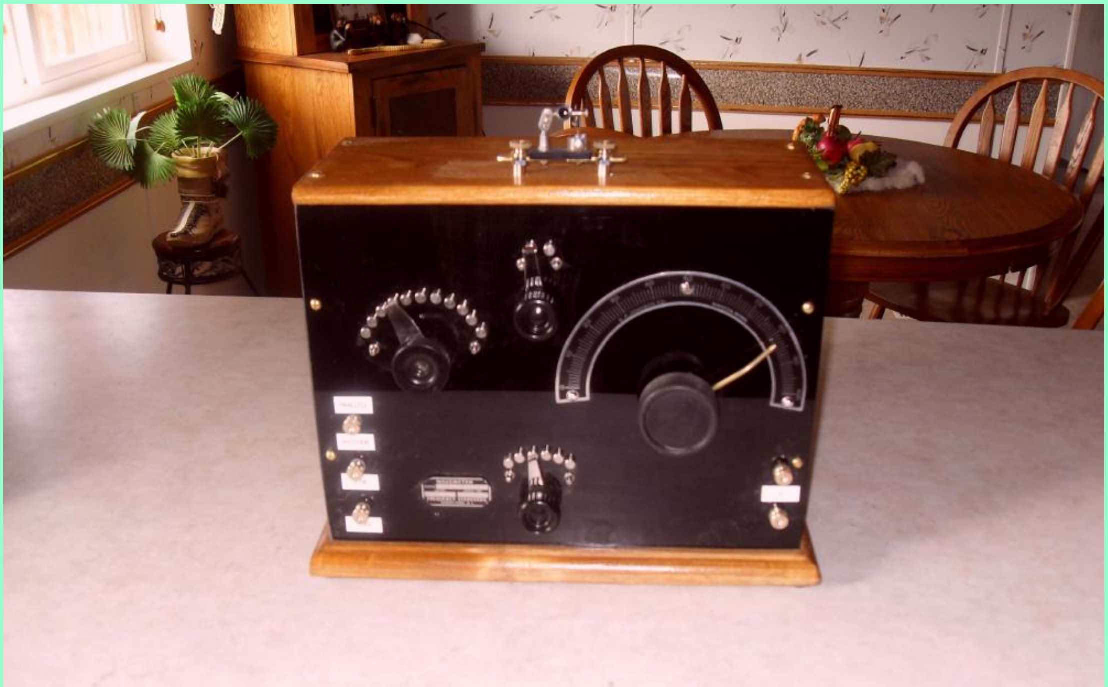

"Crystal Radio going to John in Australia"

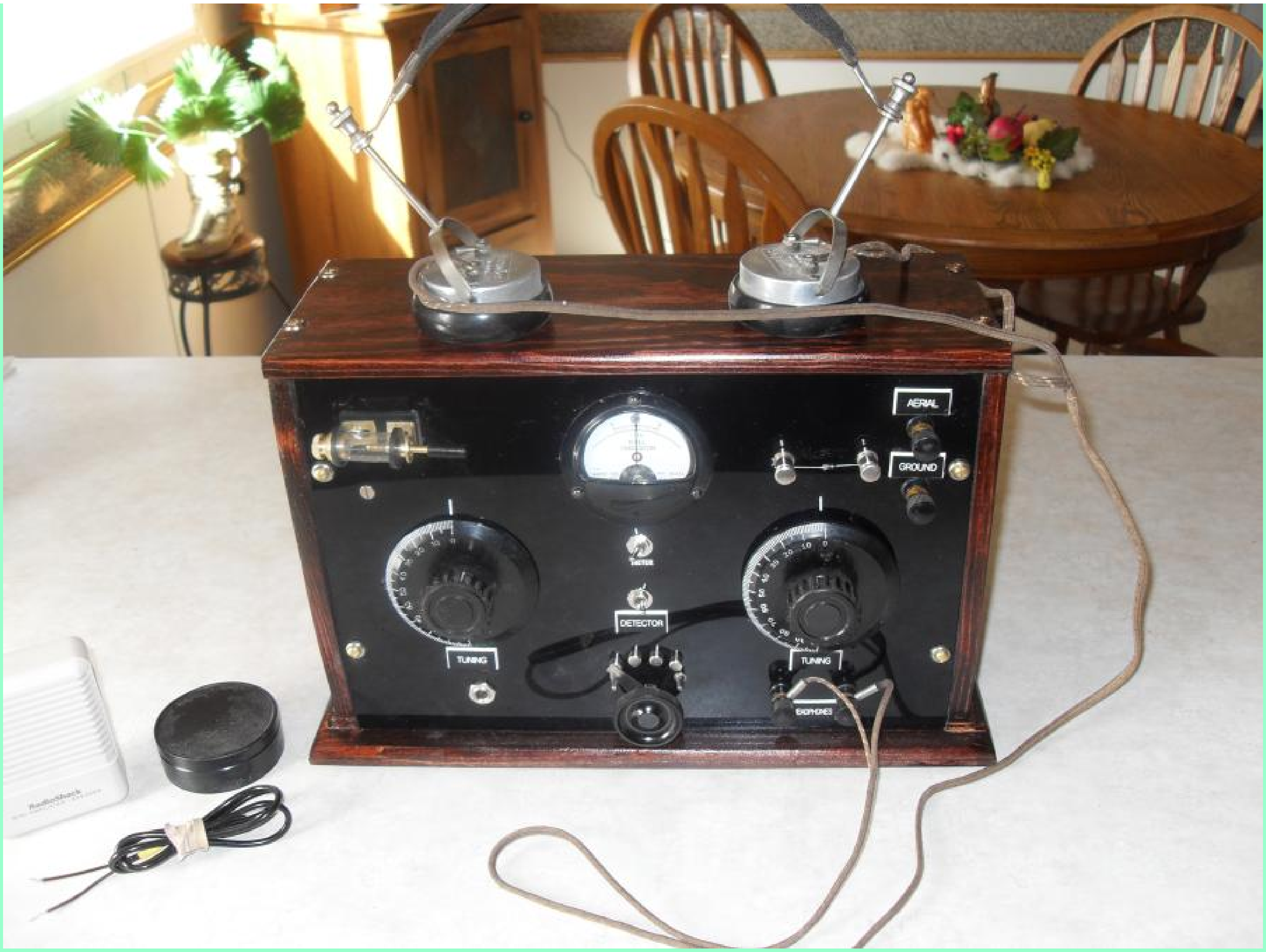
AERIAL
GROUND
METER
DETECTOR
TUNING
TUNING
EARPHONES

AIRWAVE EXPLORER NEW LAB & WORKSHOP

Example of a Custom Built Spark Gap Transmitter"

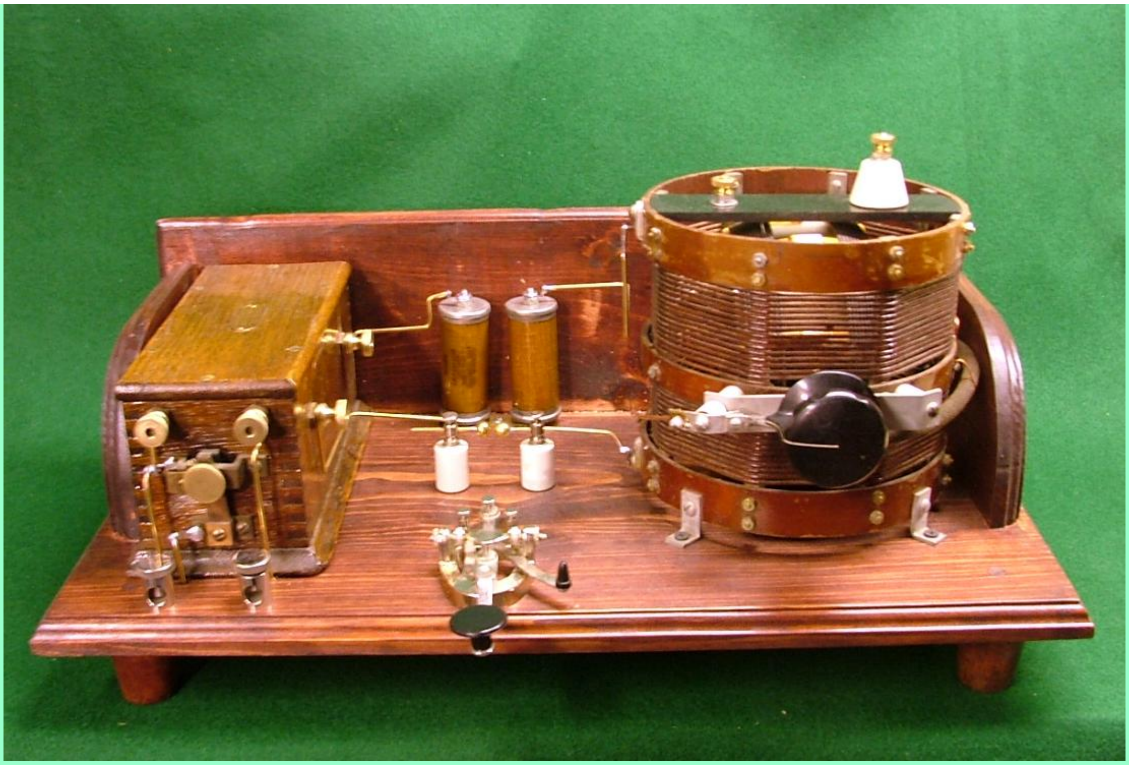

ROTARY SPARK GAP, "SOLD TO ANDY"

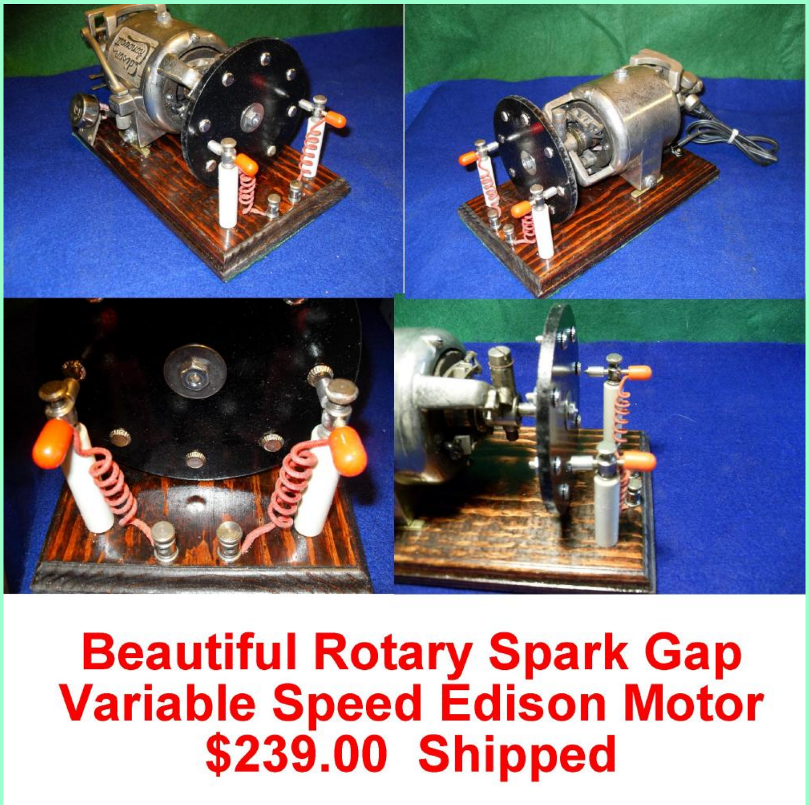
Beautiful Rotary Spark Gap
Variable Speed Edison Motor
$239.00 Shipped

"SOLD--SOLD!

CRYSTAL RADIO, "THE STAR" Excellent crystal radio with catwhisker and fixed detector. Hi-Z and Lo-Z coupling of link to tank. SOLD--SOLD.

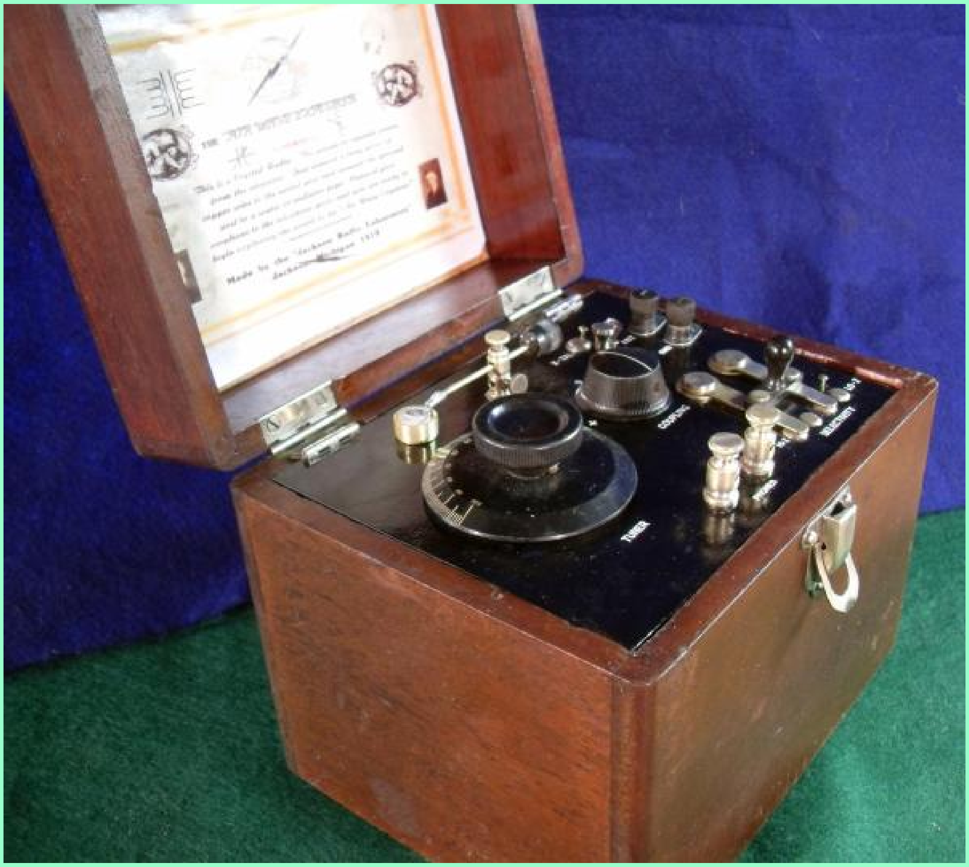
COUPLING
TUNER
SELECTIVITY

"SOLD,SOLD!

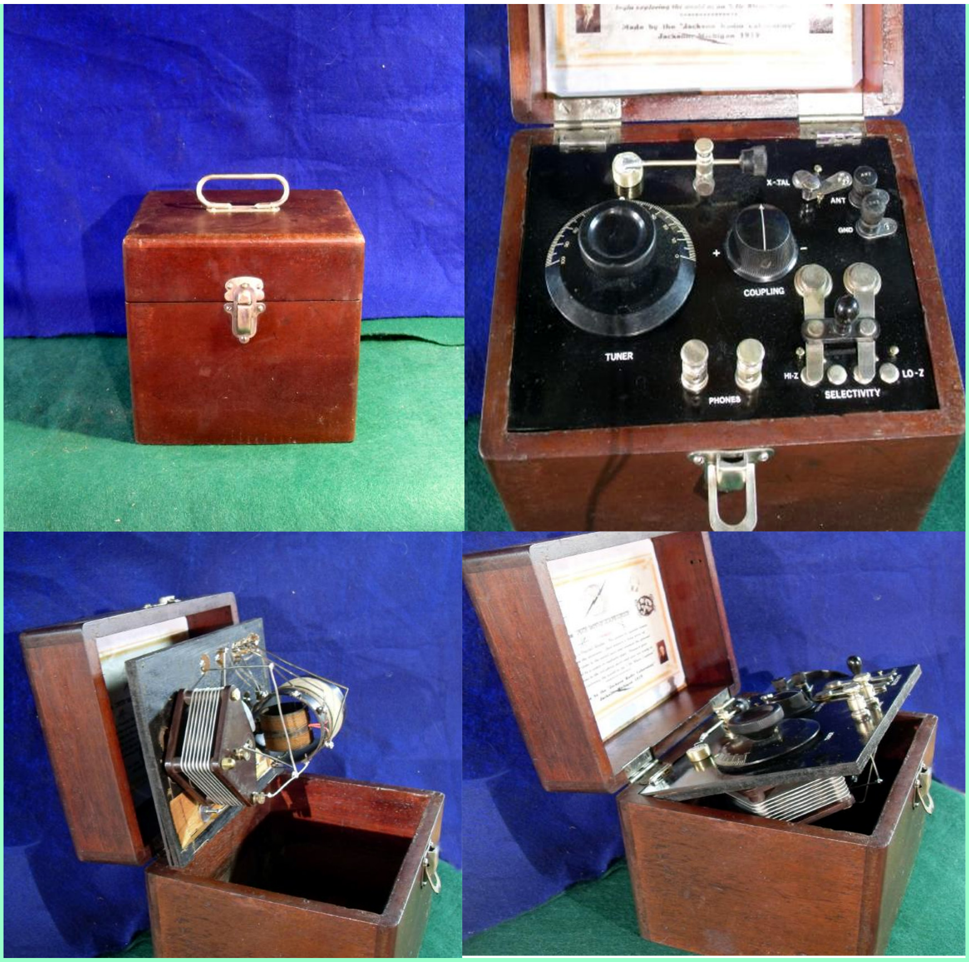
X-TAL
ANT
GND
COUPLING
TUNER
PHONES
HI-Z
SELECTIVITY
LO-Z

COHERE, SOLD.

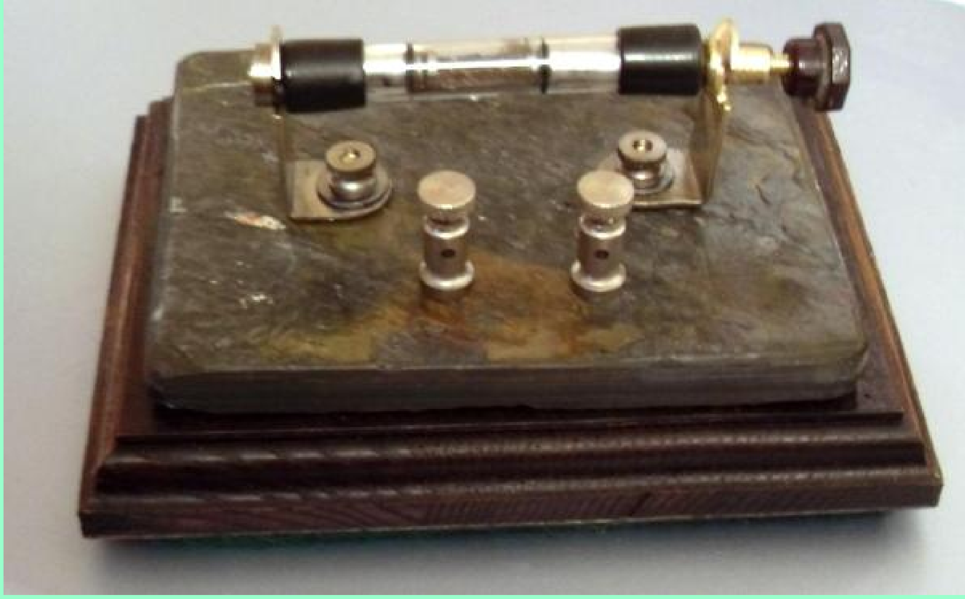

Latest addition to the Crystal Radio Receivers, Sold.

"The MAGIC BOX, Crystal Radio, Covers AM Broadcast Band" Sold to Grover K7TP

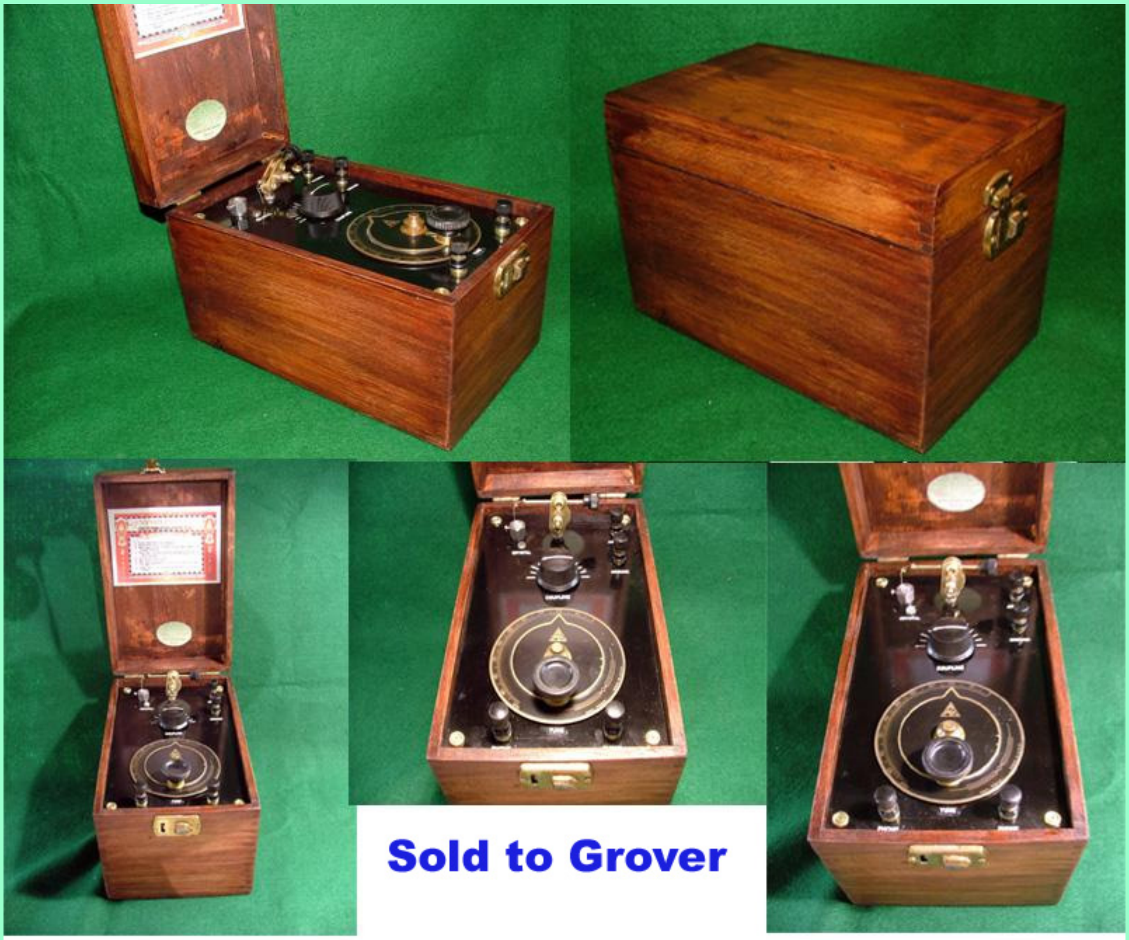

"Excellent, sensitive & selective Crystal Radio"

This radio uses a variable coupled "Variocoupler", vintage ca. 1918. Also two detectors, one catwhisker detector and one fixed detector. Complete instructions and an aerial primer is included. A vintage vernier dial for tuning.

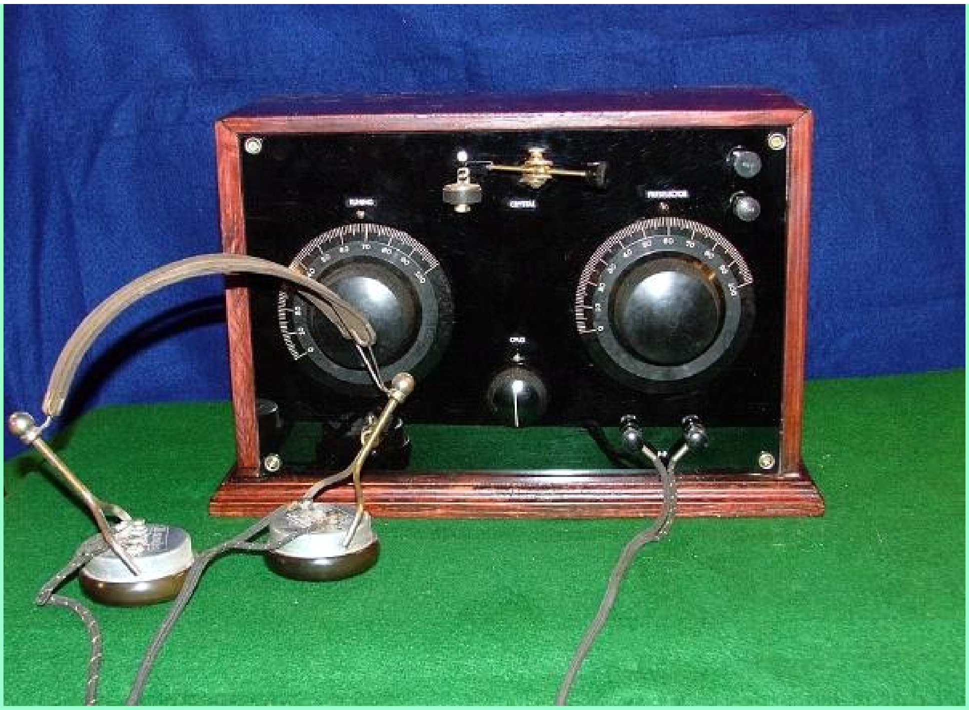

"Custom Radio going to George in Moscow, Russia"

"Custom Radio going to Jan, N5YPK"

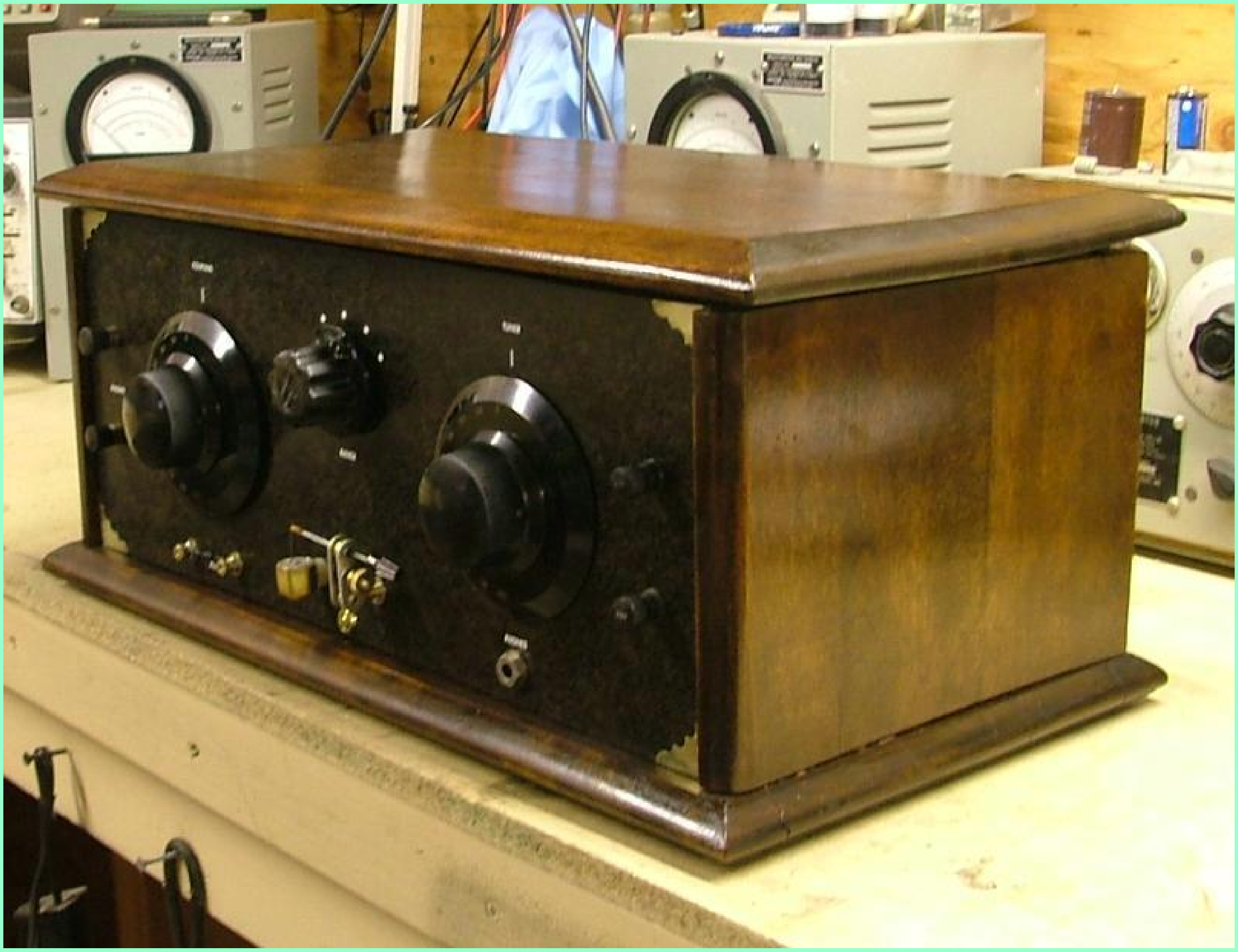

"MAGNETIC DETECTOR--(SOLD)"

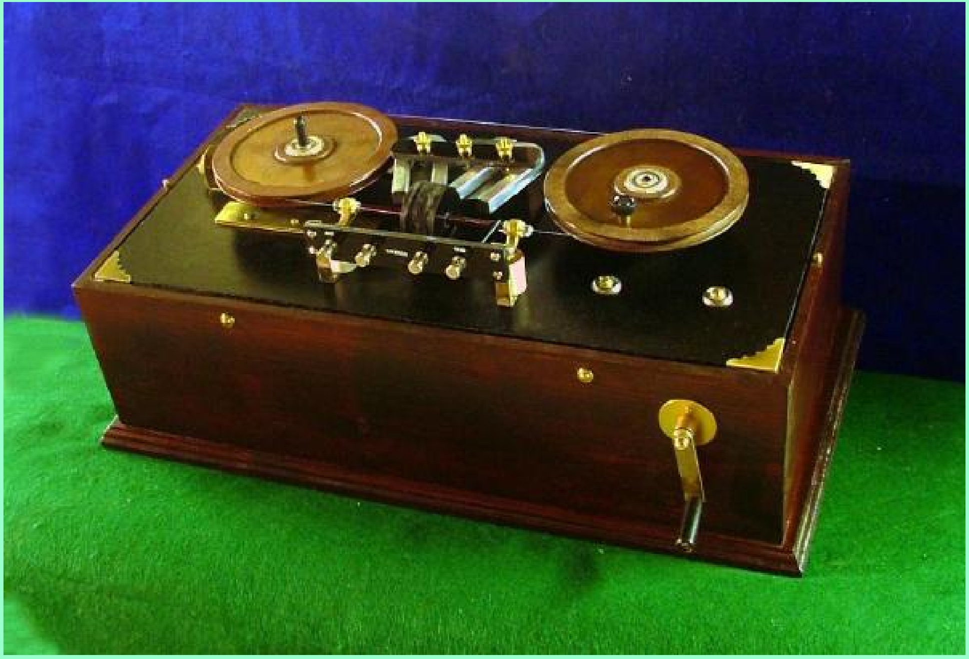

"RADIOS THAT ARE SOLD"

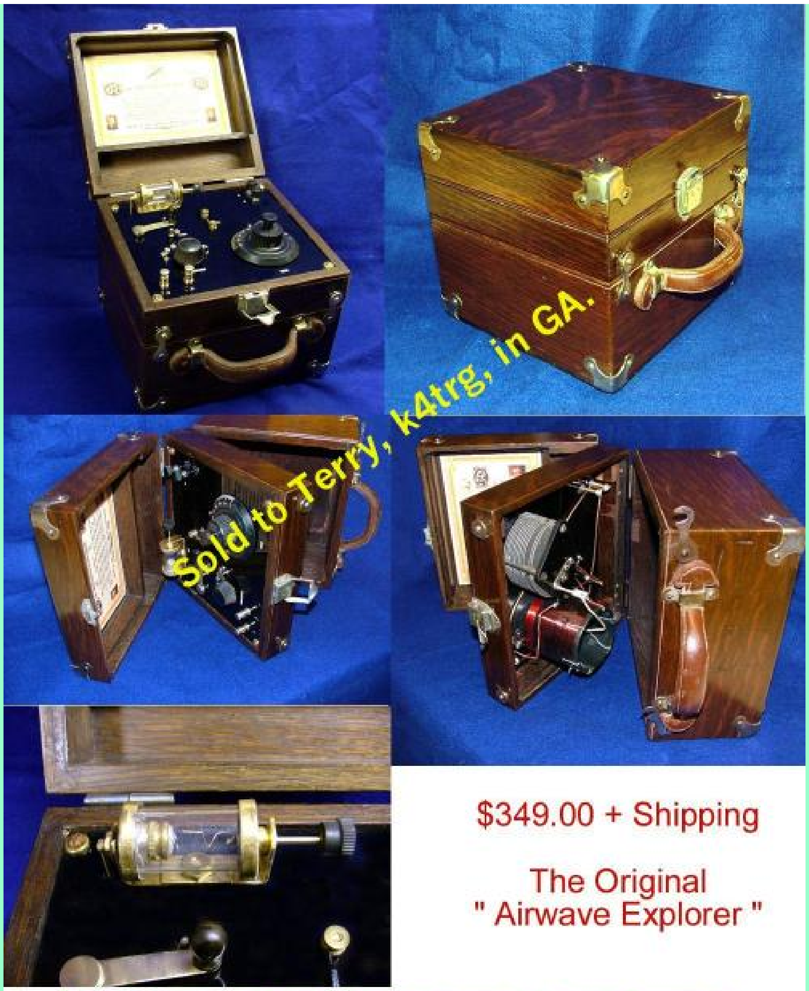

$349.00 + Shipping

The Original
" Airwave Explorer "

Exact duplicate of my Fathers personal Crystal Receiver,
used with the Spark Gap Transmitter, below.
Link coupled, enclosed brass Crystal Detector.
Regular and bandspread tuning capacitor.
All original vintage parts used in the construction.
The only variation from the original circuit is the addition
of a diode detector selected by a 2 position Post Switch
on the panel. Excellent and very Hot performing Crystal Radio.

X-TAL
S-METER
TUNE
TRIM
PHASE
OUTPUT

Custom Crystal Radio
going to Jose in Spain

WAVEMETER
CRYSTALS
BCST
SW
ANT.
GND
PHONE

AIRWAVE EXPLORER LABS
"Come to the party but bring your own headphones!"

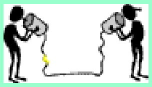

[Additional radios](#)

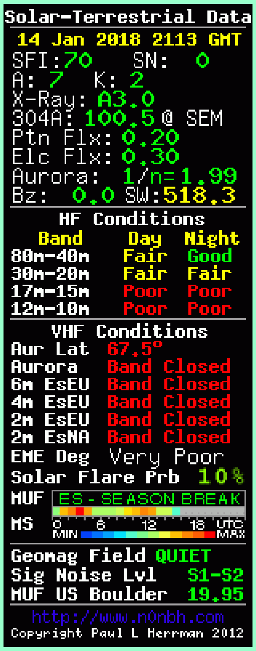

Solar-Terrestrial Data
14 Jan 2018 2113 GMT
SFI: 70 SN: 0
A: 7 K: 2
X-Ray: A3.0
304A: 100.5 @ SEM
Ptn Flx: 0.20
Elc Flx: 0.30
Aurora: 1/n=1.99
Bz: 0.0 SW: 518.3

HF Conditions

Band	Day	Night
80m-40m	Fair	Good
30m-20m	Fair	Fair
17m-15m	Poor	Poor
12m-10m	Poor	Poor

VHF Conditions

Aur Lat	67.5°
Aurora	Band Closed
6m EsEU	Band Closed
4m EsEU	Band Closed
2m EsEU	Band Closed
2m EsNA	Band Closed
EME Deg	Very Poor
Solar Flare Prb	10%

MUF ES - SEASON BREAK
MS 0 6 12 18 UTC MIN MAX
Geomag Field QUIET
Sig Noise Lvl S1-S2
MUF US Boulder 19.95
http://www.n0nbh.com
Copyright Paul L Herrman 2012

[Click Here to Send Me an Email](#) Send Us Mail

Email me with questions or requests:

"k8cix@aol.com"

+ −

REVOLVERMAPS

39363

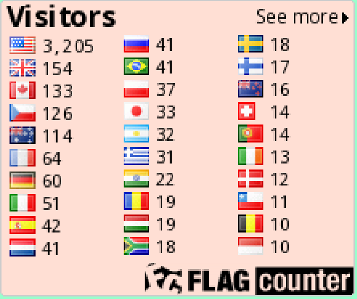

Visitors See more▸

🇺🇸 3,205	🇷🇺 41	🇸🇪 18
🇬🇧 154	🇧🇷 41	🇫🇮 17
🇨🇦 133	🇵🇱 37	🇳🇿 16
🇨🇿 126	🇯🇵 33	🇨🇭 14
🇦🇺 114	🇦🇷 32	🇵🇹 14
🇫🇷 64	🇬🇷 31	🇮🇪 13
🇩🇪 60	🇮🇳 22	🇩🇰 12
🇮🇹 51	🇷🇴 19	🇨🇱 11
🇪🇸 42	🇭🇺 19	🇧🇪 10
🇳🇱 41	🇿🇦 18	🇮🇩 10

FLAG counter

CANCER DOESN'T
COME WITH
INSTRUCTIONS.
LIVESTRONG

BUILD A FREE WEBSITE
OF YOUR OWN ON
TRIPOD

American
Diabetes

diabetes

US 20090294300A1

(19) **United States**

(12) **Patent Application Publication**

Kanzius

(10) **Pub. No.: US 2009/0294300 A1**

(43) **Pub. Date: Dec. 3, 2009**

(54) RF SYSTEMS AND METHODS FOR PROCESSING SALT WATER

(75) Inventor: **John Kanzius**, Erie, PA (US); **Mary Ann Kanzius**, legal representative, Erie, PA (US)

Correspondence Address:
CALFEE HALTER & GRISWOLD, LLP
800 SUPERIOR AVENUE, SUITE 1400
CLEVELAND, OH 44114 (US)

(73) Assignee: **KC ENERGY, LLC**, Erie, PA (US)

(21) Appl. No.: **12/398,646**

(22) Filed: **Mar. 5, 2009**

Related U.S. Application Data

(60) Provisional application No. 61/033,962, filed on Mar. 5, 2008.

Publication Classification

(51) **Int. Cl.**
C02F 1/469 (2006.01)

(52) **U.S. Cl.** **205/687**

(57) ABSTRACT

Systems and methods for processing salt water and/or solutions containing salt water with RF energy. Exemplary systems and methods may use RF energy to combust salt water, to produce hydrogen from salt water or solutions containing salt water, to volatilize a secondary fuel present in solutions containing salt water, to produce and combust hydrogen obtained from salt water or solutions containing salt water, to volatilize and combust secondary fuel sources present in solutions containing salt water, to desalinate seawater, and/or to carry out the electrolysis of water are presented. An exemplary system may comprise a reservoir for containing a salt water solution or salt water mixture; a reaction chamber having an inlet; a feed line operatively connecting the reservoir to the inlet of the reaction chamber; an RF transmitter having an RF generator in circuit communication with a transmission head, the RF generator capable of generating an RF signal absorbable by the salt water solution or the salt water mixture having a frequency for transmission via the transmission head; and an RF receiver; wherein the reaction chamber is positioned such that some of the salt water solution or salt water mixture is positioned between the RF transmission head and the RF receiver.

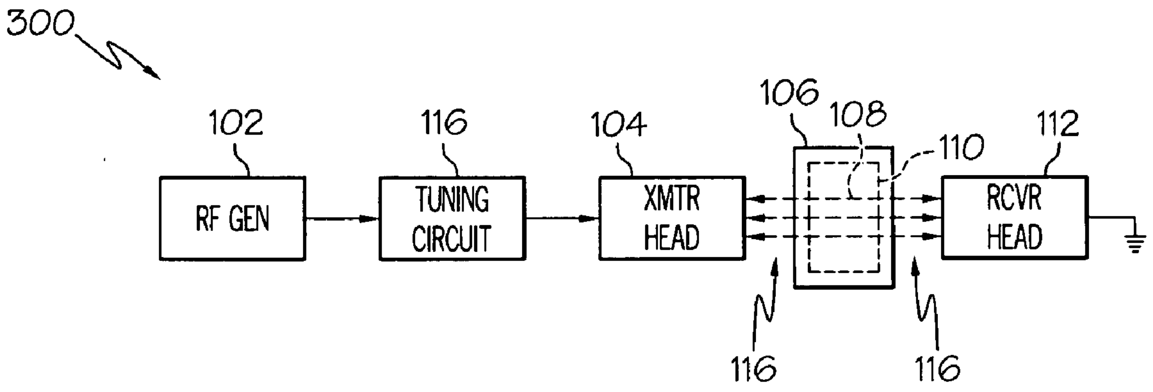

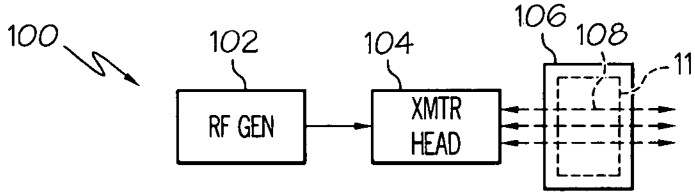
100
102
RF GEN
104
XMTR HEAD
106
108
110

FIG. 1

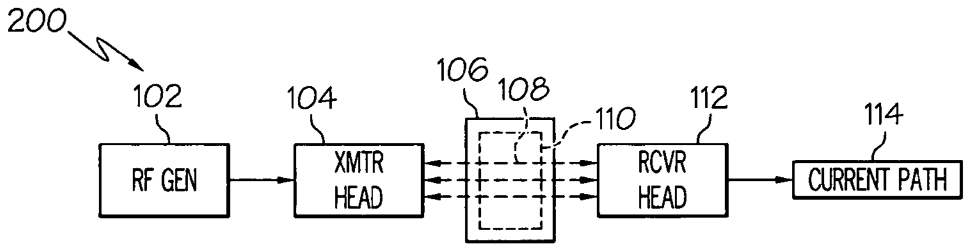
200
102
RF GEN
104
XMTR HEAD
106
108
110
112
RCVR HEAD
114
CURRENT PATH

FIG. 2

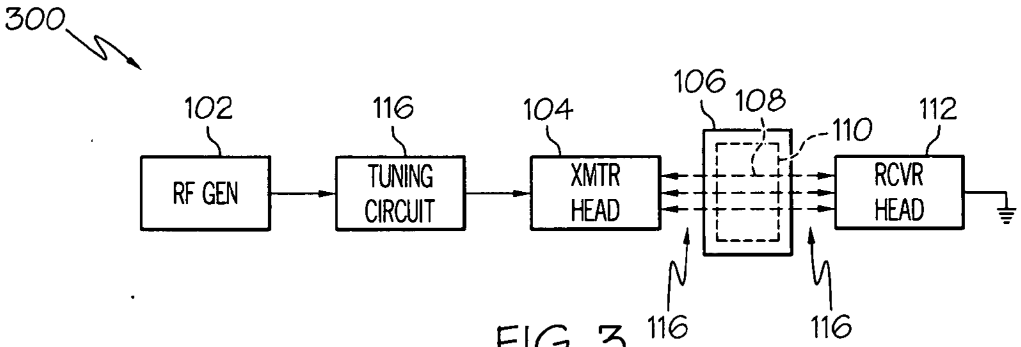
300
102
RF GEN
116
TUNING CIRCUIT
104
XMTR HEAD
106
108
110
112
RCVR HEAD
116
116

FIG. 3

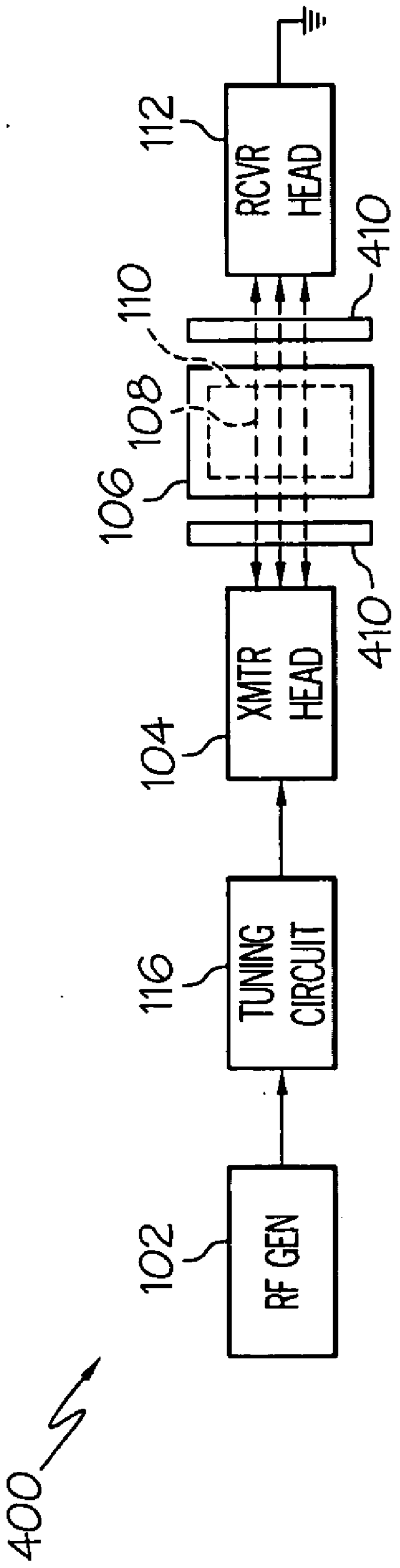
RF GEN
TUNING CIRCUIT
XMTR HEAD
RCVR HEAD

FIG. 4

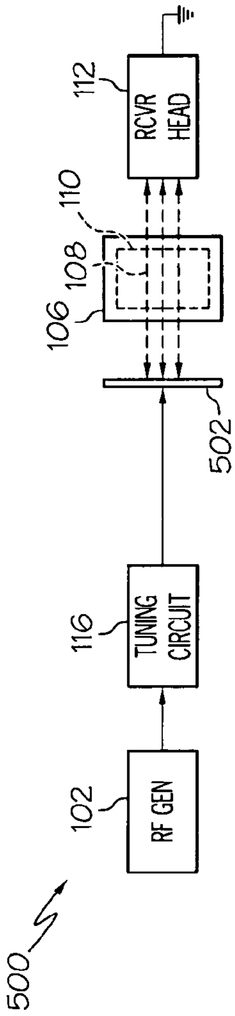
500
102
RF GEN
116
TUNING CIRCUIT
502
106
108
110
112
RCVR HEAD

FIG. 5

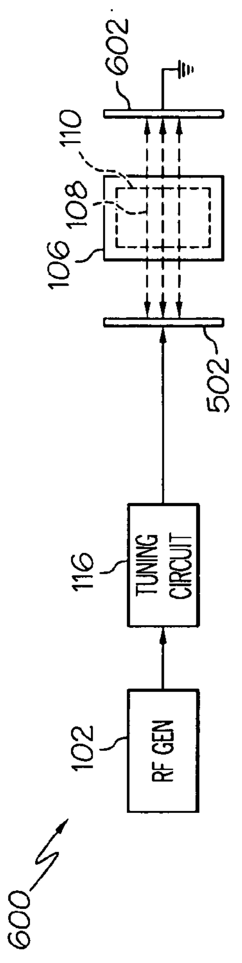
600
102
RF GEN
116
TUNING CIRCUIT
502
106
108
110
602

FIG. 6

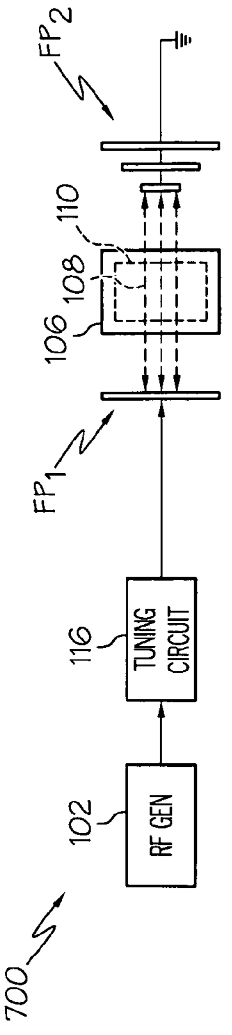
700
102
RF GEN
116
TUNING CIRCUIT
FP_1
106
108
110
FP_2

FIG. 7

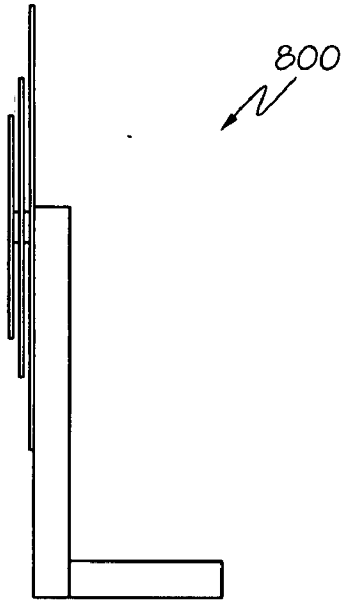

FIG. 8A

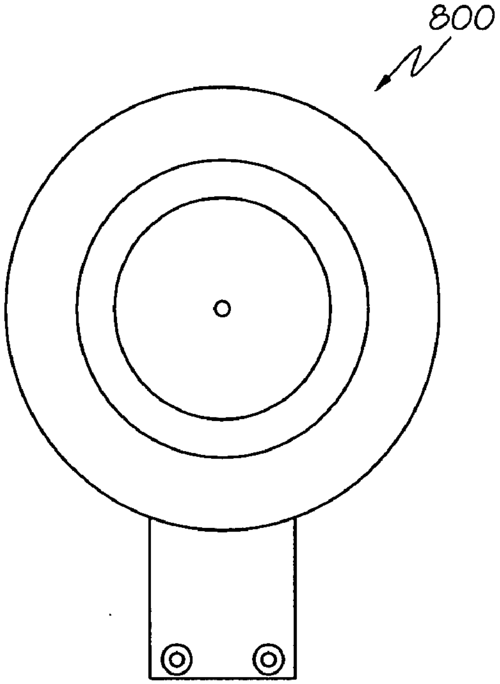

FIG. 8B

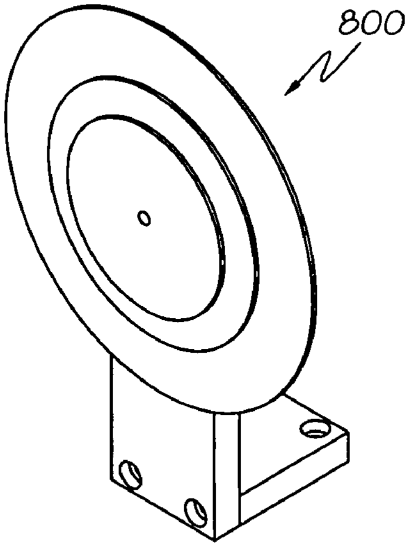

FIG. 8C

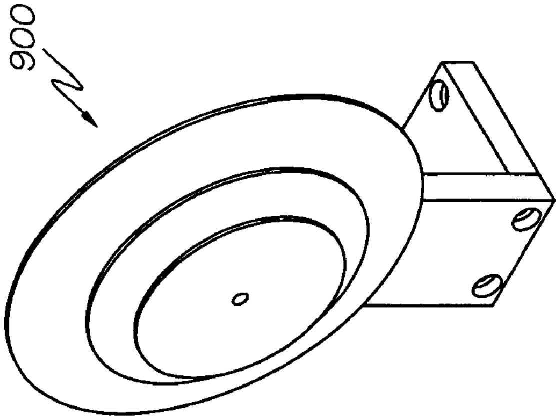

FIG. 9C

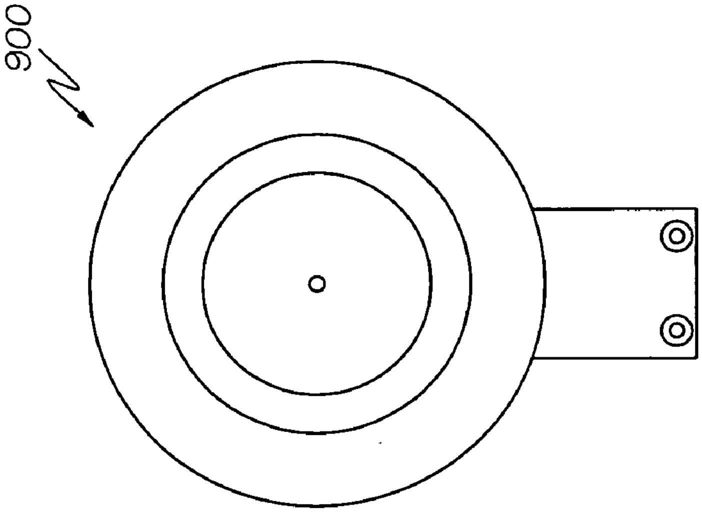

FIG. 9B

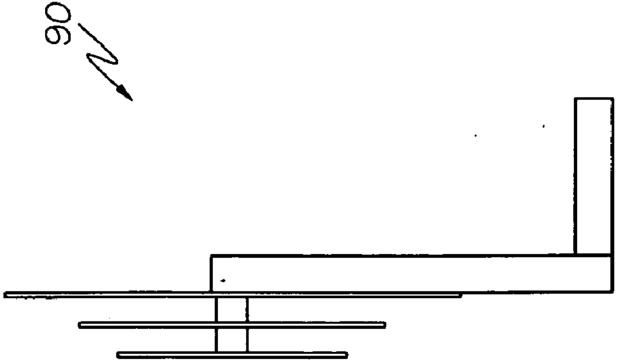

FIG. 9A

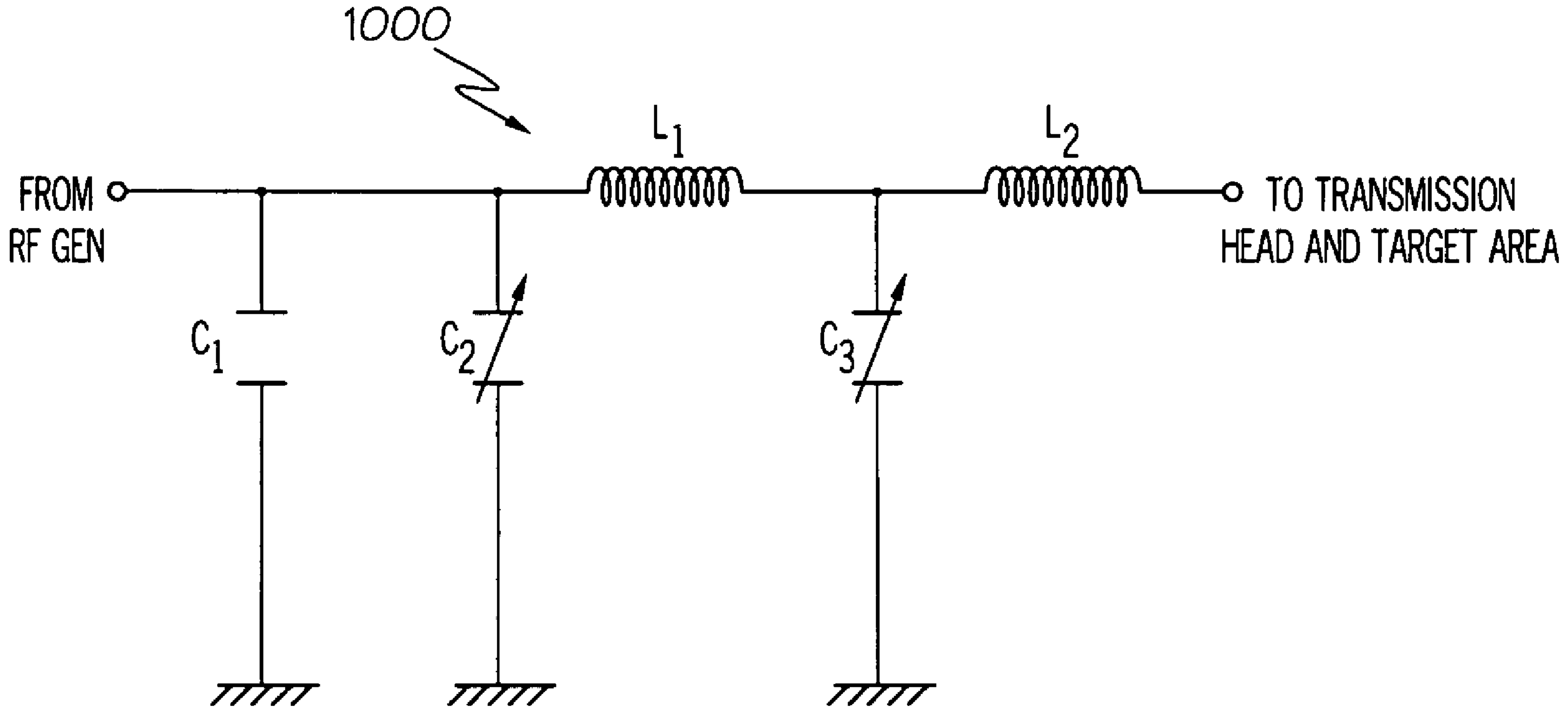
1000
FROM
RF GEN
L_1
L_2
TO TRANSMISSION
HEAD AND TARGET AREA
C_1
C_2
C_3

FIG. 10

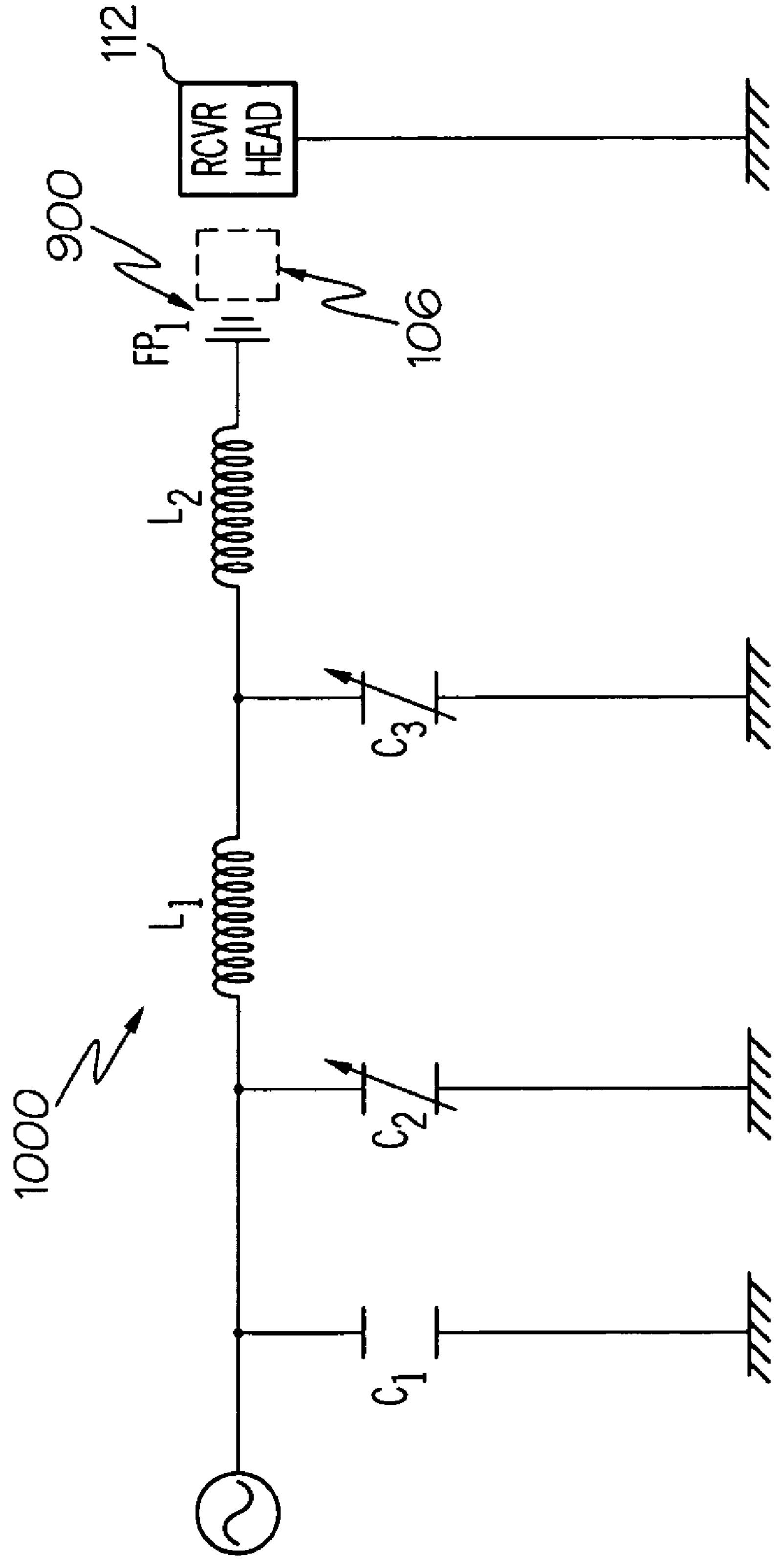
112
RCVR HEAD
900
FP_1
106
L_2
C_3
L_1
1000
C_2
C_1

FIG. 11

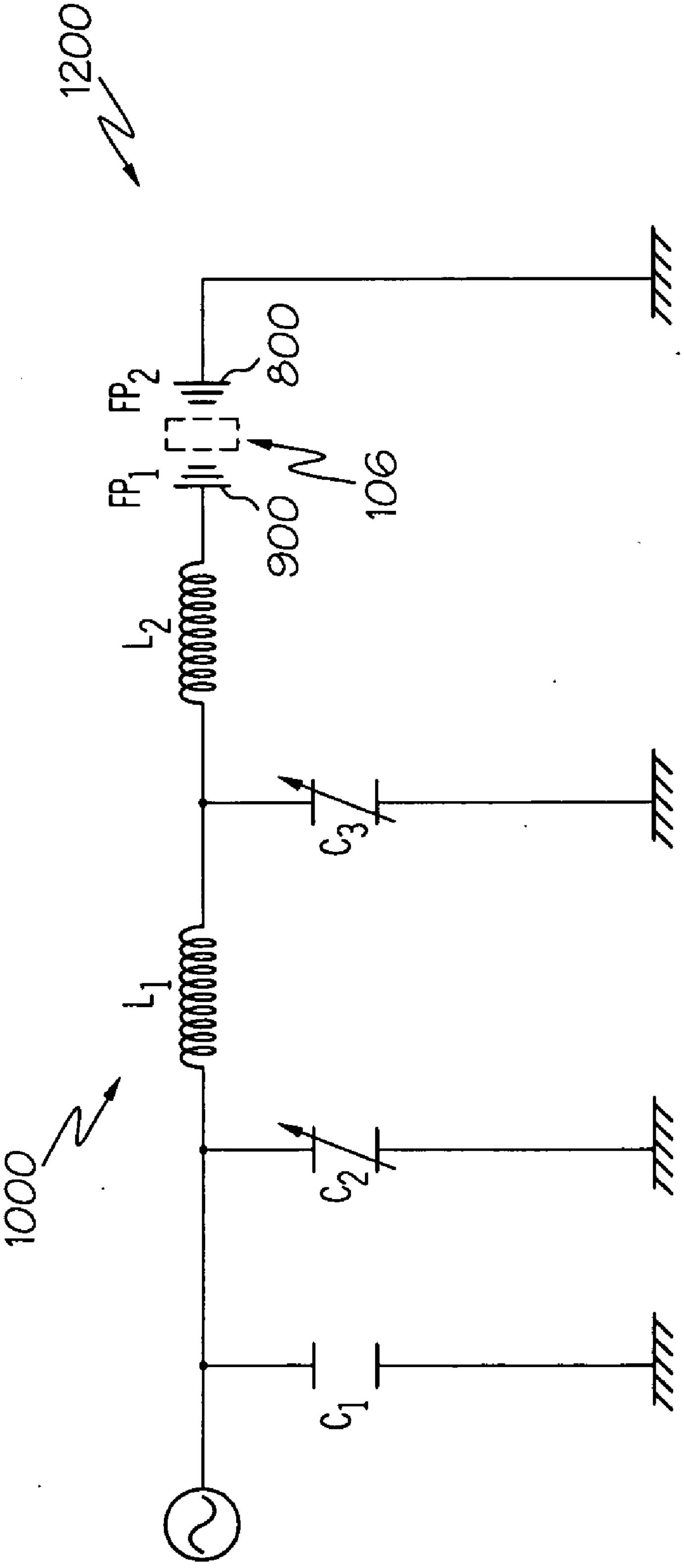
1200
FP_2
800
FP_1
106
900
L_2
C_3
L_1
1000
C_2
C_1

FIG. 12

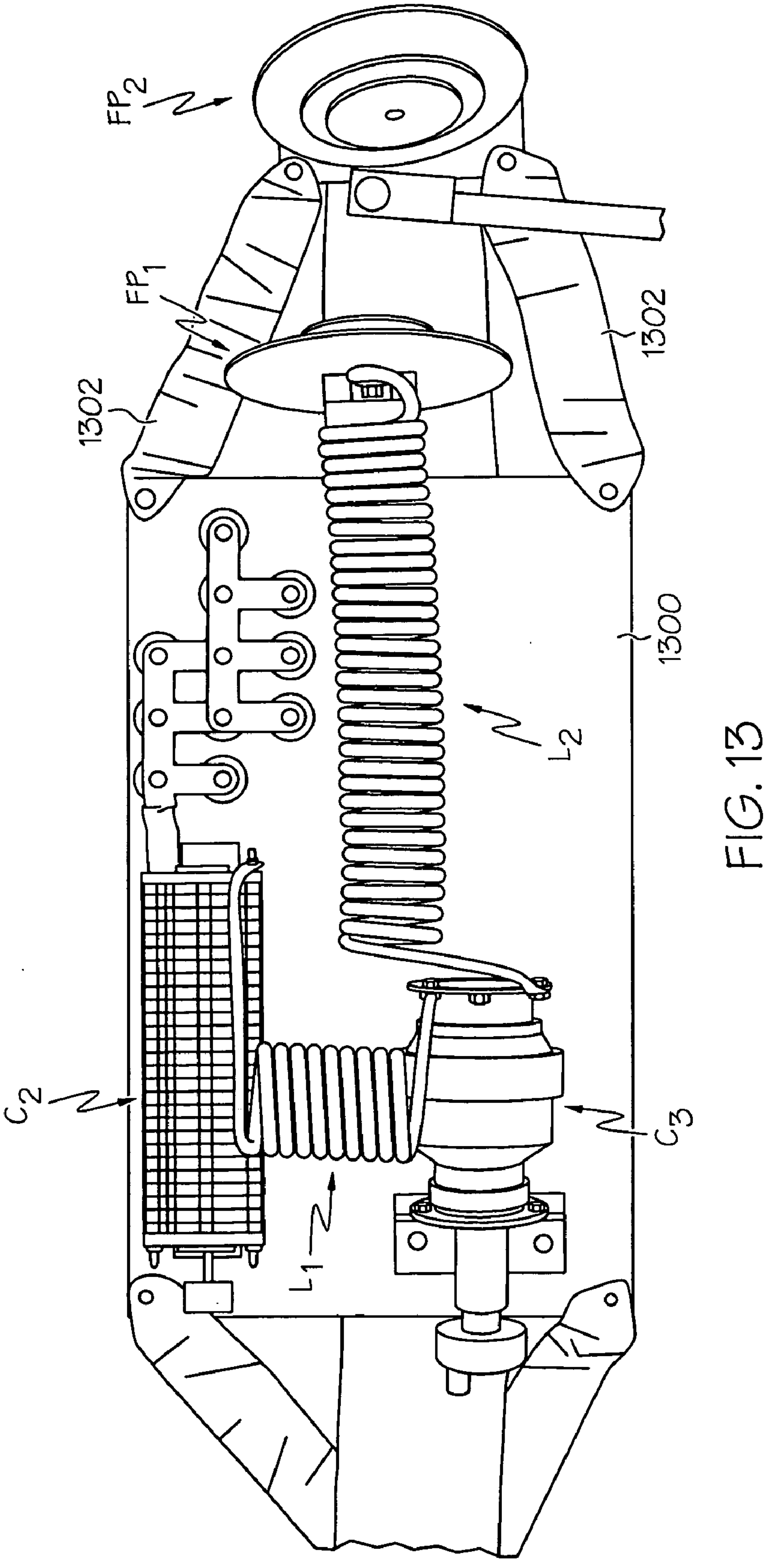
FP_2
FP_1
1302
1302
1300
L_2
C_2
C_3
L_1

FIG. 13

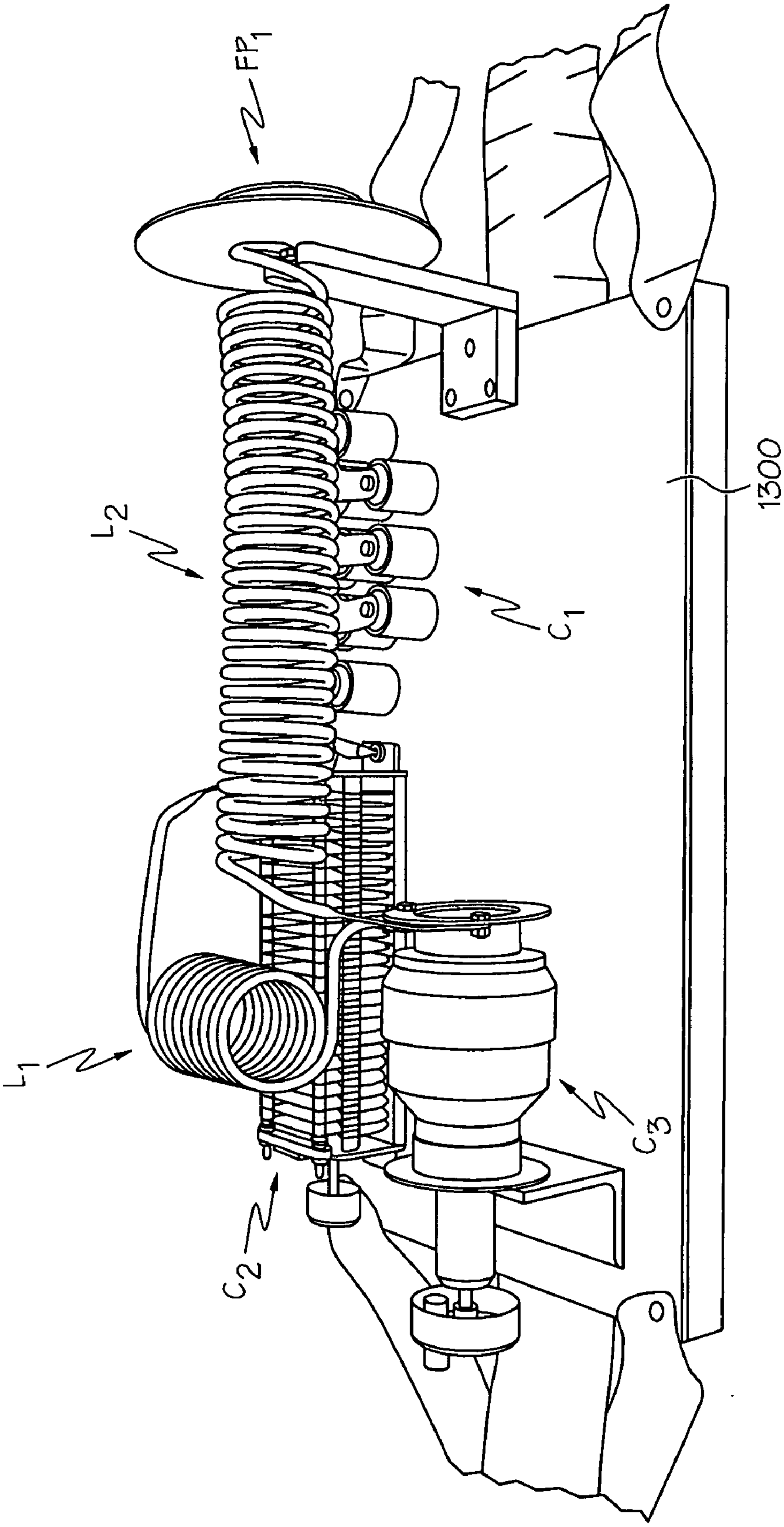
FP1
L2
C1
1300
L1
C3
C2

FIG. 14

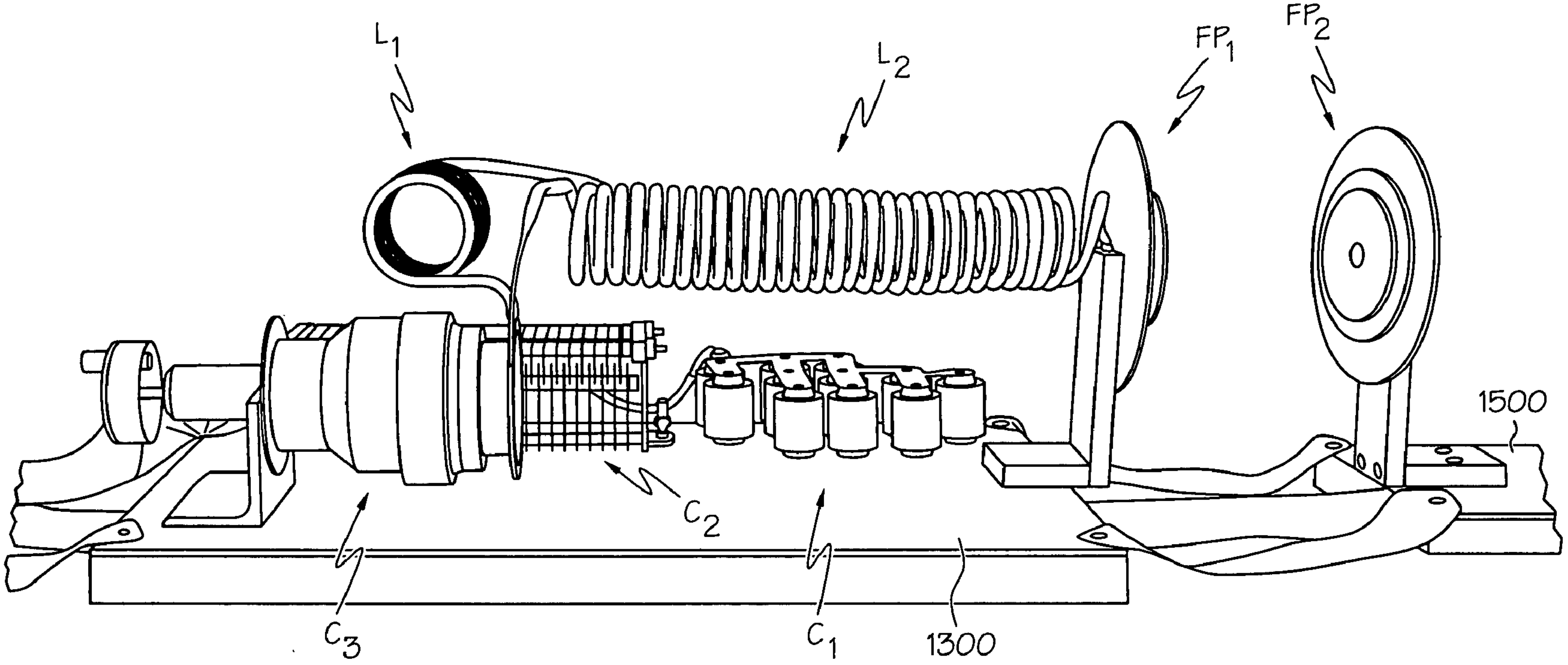
L_1
L_2
FP_1
FP_2
1500
C_2
C_3
C_1
1300

FIG. 15

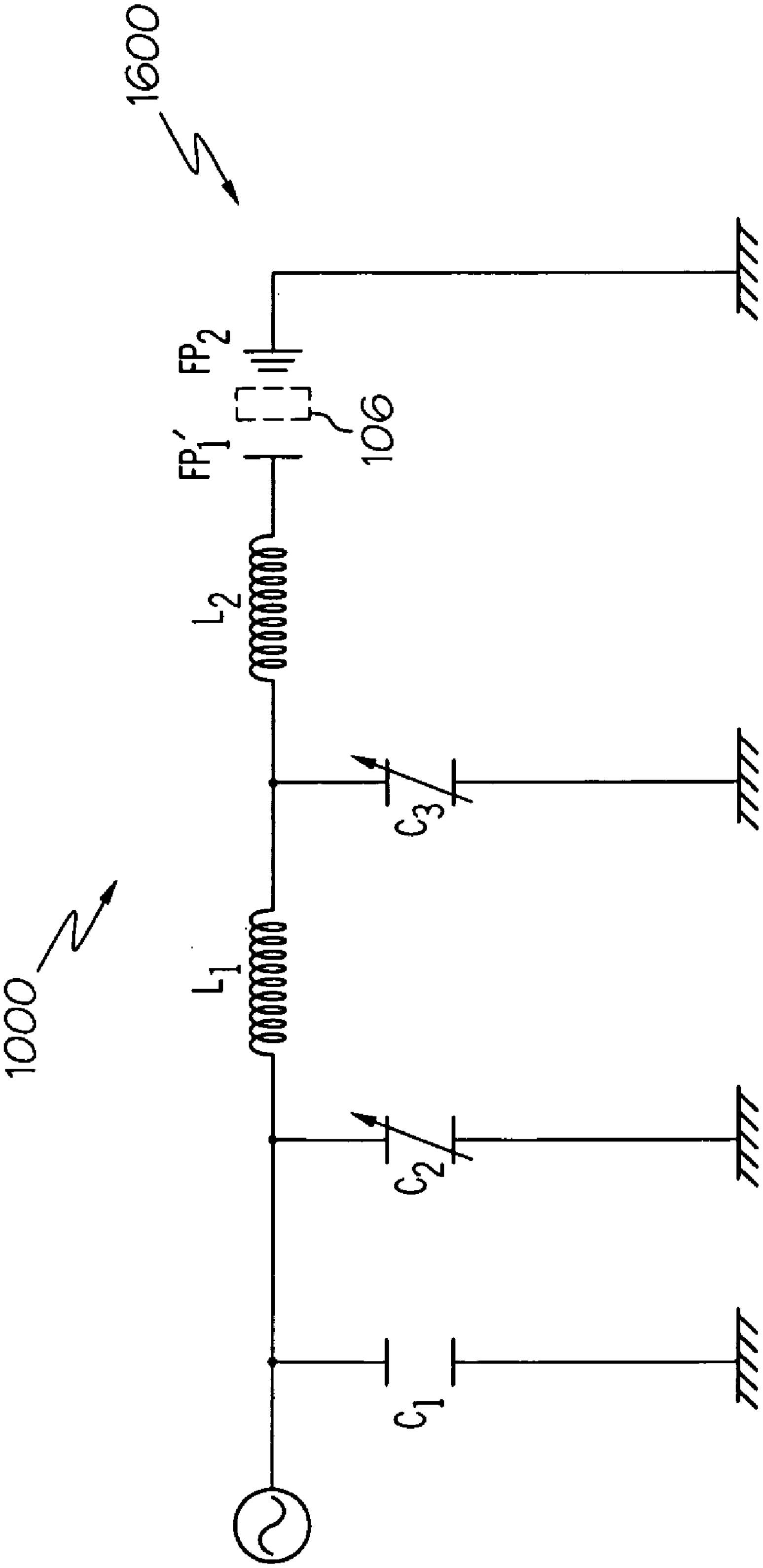
1600
FP_2
FP_1'
106
L_2
C_3
L_1
1000
C_2
C_1

FIG. 16

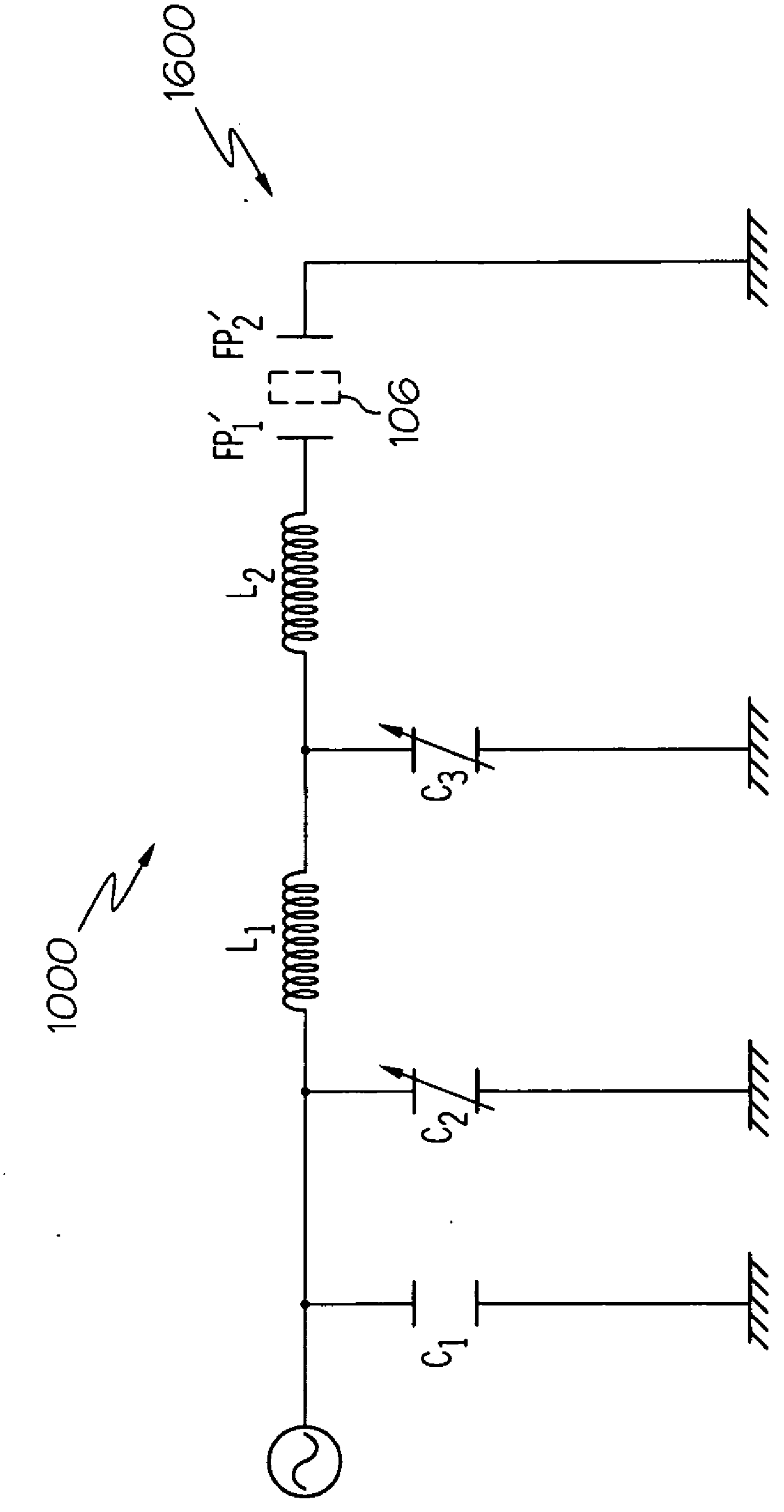
1600
FP'$_2$
FP'$_1$
106
L$_2$
C$_3$
1000
L$_1$
C$_2$
C$_1$

FIG. 16a

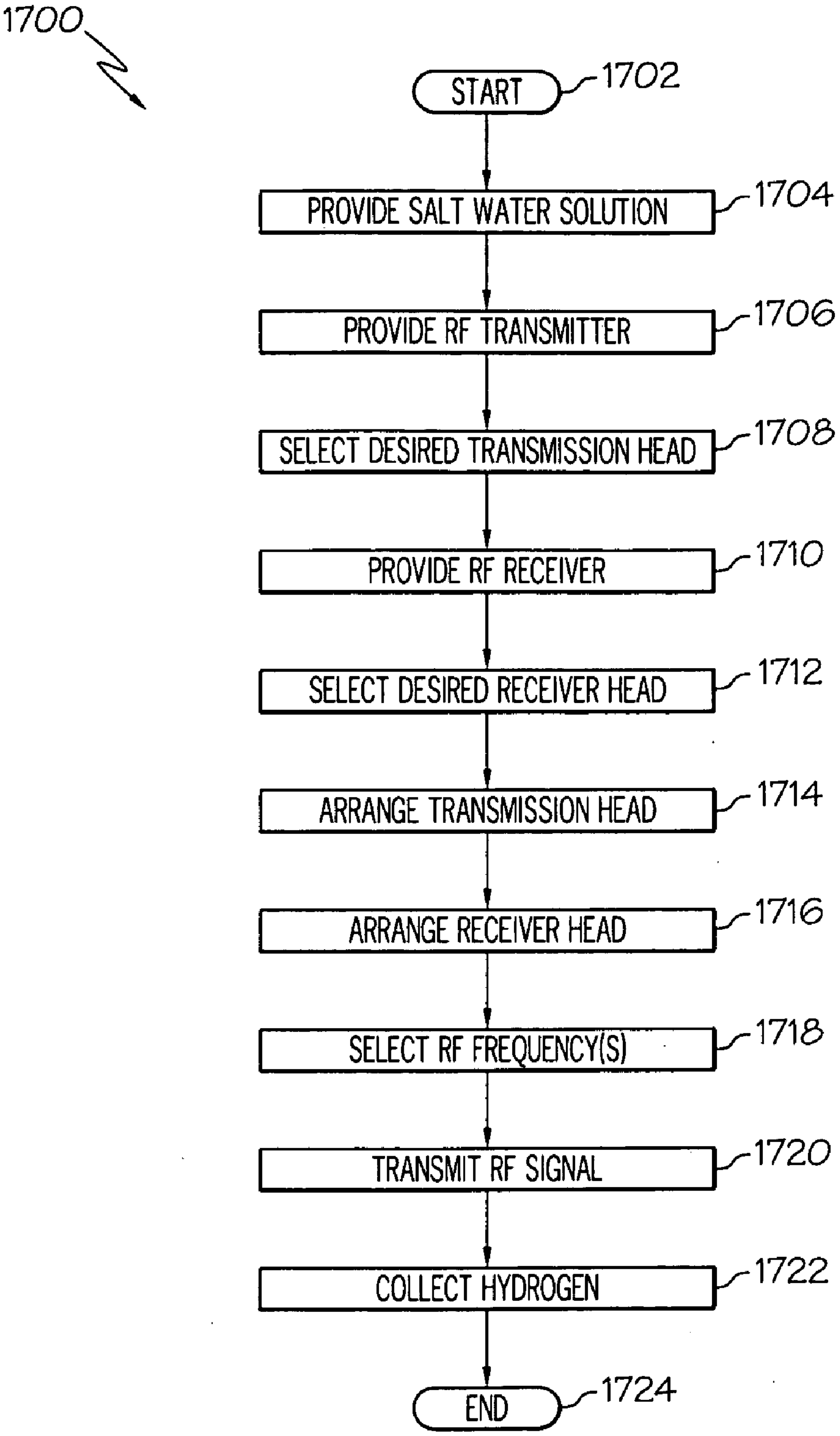
1700
START
1702
PROVIDE SALT WATER SOLUTION
1704
PROVIDE RF TRANSMITTER
1706
SELECT DESIRED TRANSMISSION HEAD
1708
PROVIDE RF RECEIVER
1710
SELECT DESIRED RECEIVER HEAD
1712
ARRANGE TRANSMISSION HEAD
1714
ARRANGE RECEIVER HEAD
1716
SELECT RF FREQUENCY(S)
1718
TRANSMIT RF SIGNAL
1720
COLLECT HYDROGEN
1722
END
1724

FIG. 17

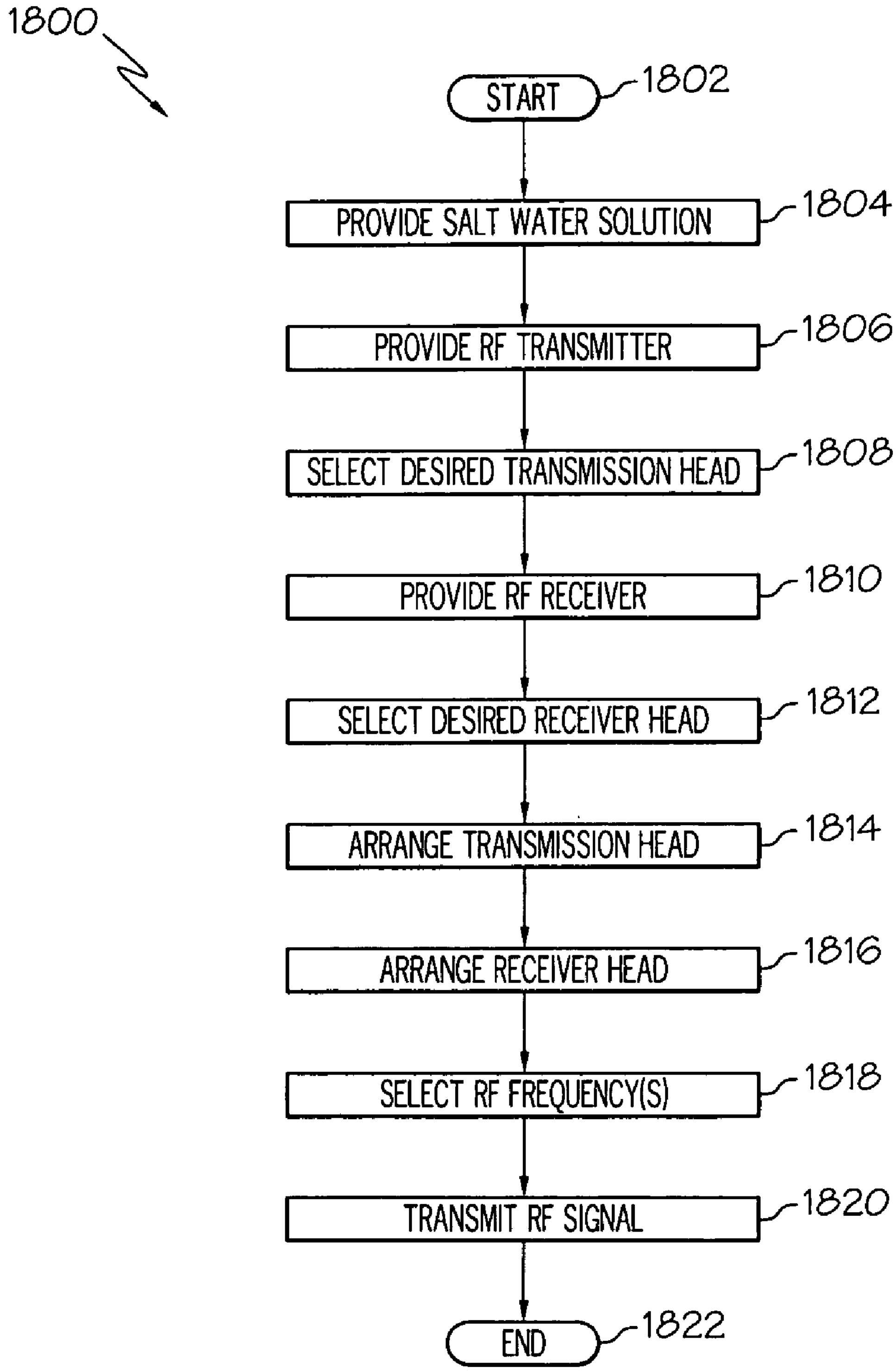
1800
START
1802
PROVIDE SALT WATER SOLUTION
1804
PROVIDE RF TRANSMITTER
1806
SELECT DESIRED TRANSMISSION HEAD
1808
PROVIDE RF RECEIVER
1810
SELECT DESIRED RECEIVER HEAD
1812
ARRANGE TRANSMISSION HEAD
1814
ARRANGE RECEIVER HEAD
1816
SELECT RF FREQUENCY(S)
1818
TRANSMIT RF SIGNAL
1820
END
1822

FIG. 18a

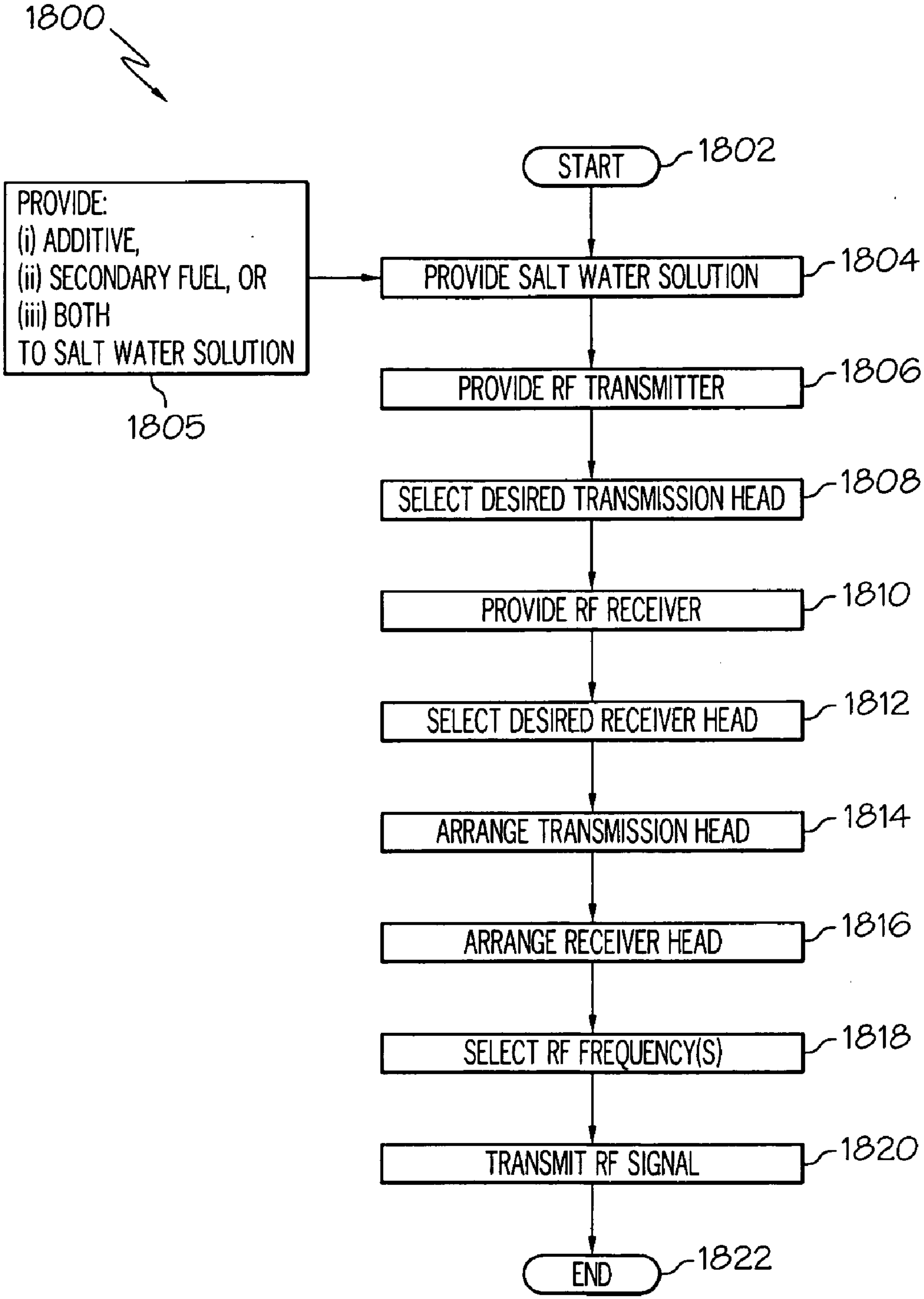
1800
START
1802
PROVIDE:
(i) ADDITIVE,
(ii) SECONDARY FUEL, OR
(iii) BOTH
TO SALT WATER SOLUTION
1805
PROVIDE SALT WATER SOLUTION
1804
PROVIDE RF TRANSMITTER
1806
SELECT DESIRED TRANSMISSION HEAD
1808
PROVIDE RF RECEIVER
1810
SELECT DESIRED RECEIVER HEAD
1812
ARRANGE TRANSMISSION HEAD
1814
ARRANGE RECEIVER HEAD
1816
SELECT RF FREQUENCY(S)
1818
TRANSMIT RF SIGNAL
1820
END
1822

FIG. 18b

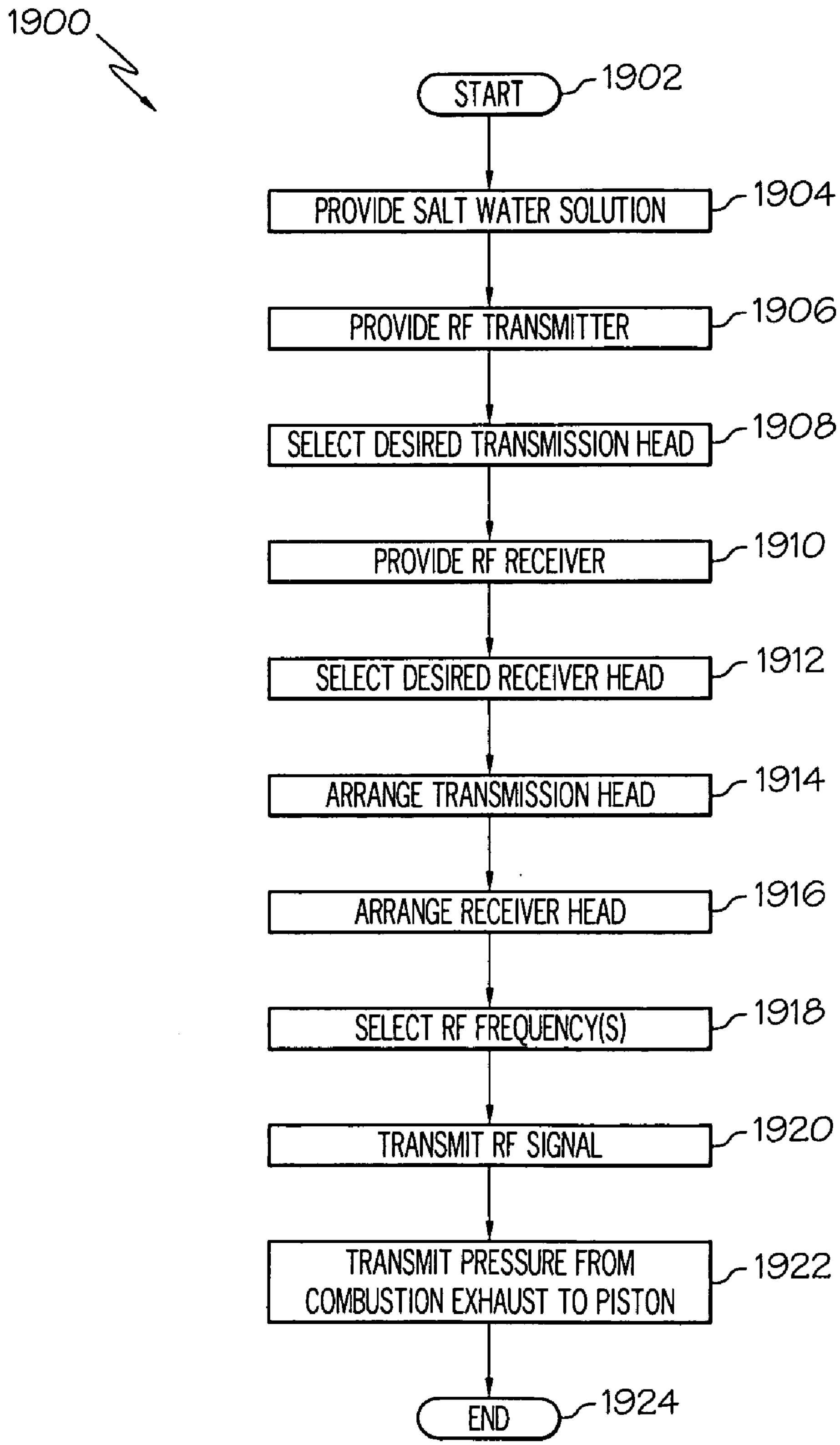
1900
START
1902
PROVIDE SALT WATER SOLUTION
1904
PROVIDE RF TRANSMITTER
1906
SELECT DESIRED TRANSMISSION HEAD
1908
PROVIDE RF RECEIVER
1910
SELECT DESIRED RECEIVER HEAD
1912
ARRANGE TRANSMISSION HEAD
1914
ARRANGE RECEIVER HEAD
1916
SELECT RF FREQUENCY(S)
1918
TRANSMIT RF SIGNAL
1920
TRANSMIT PRESSURE FROM COMBUSTION EXHAUST TO PISTON
1922
END
1924

FIG. 19a

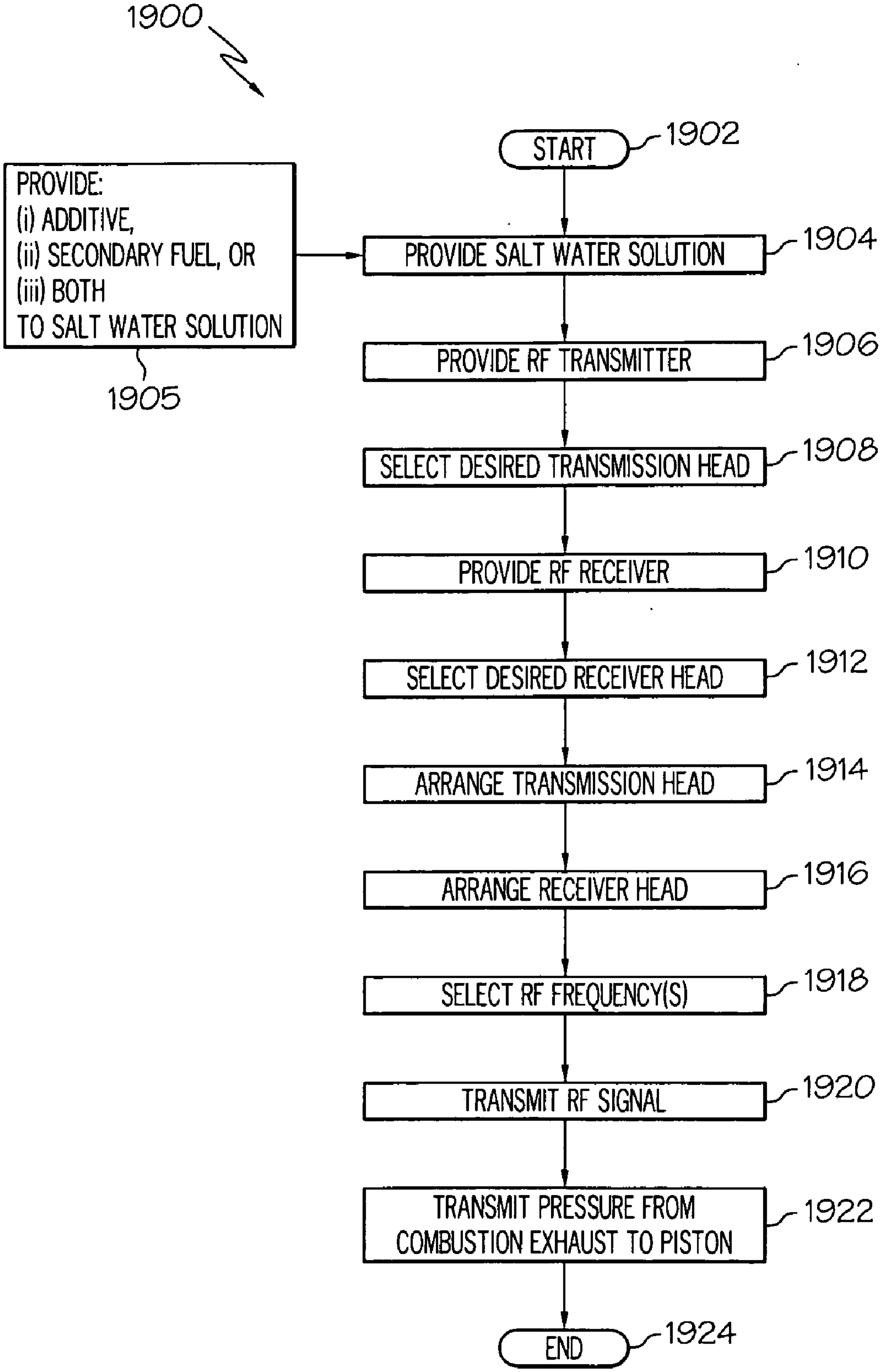
1900
START
1902
PROVIDE:
(i) ADDITIVE,
(ii) SECONDARY FUEL, OR
(iii) BOTH
TO SALT WATER SOLUTION
1905
PROVIDE SALT WATER SOLUTION
1904
PROVIDE RF TRANSMITTER
1906
SELECT DESIRED TRANSMISSION HEAD
1908
PROVIDE RF RECEIVER
1910
SELECT DESIRED RECEIVER HEAD
1912
ARRANGE TRANSMISSION HEAD
1914
ARRANGE RECEIVER HEAD
1916
SELECT RF FREQUENCY(S)
1918
TRANSMIT RF SIGNAL
1920
TRANSMIT PRESSURE FROM COMBUSTION EXHAUST TO PISTON
1922
END
1924

FIG. 19b

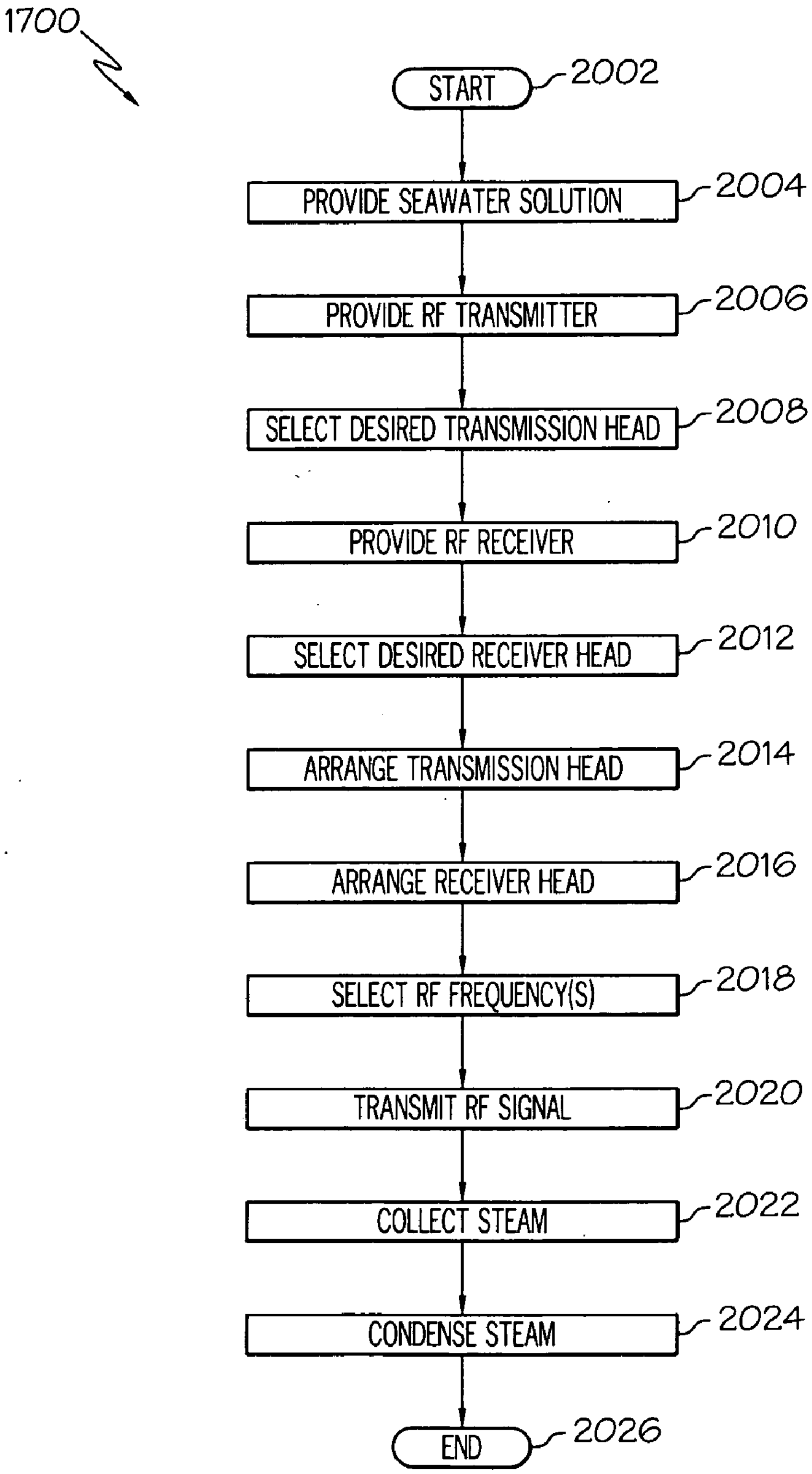
1700
START
2002
PROVIDE SEAWATER SOLUTION
2004
PROVIDE RF TRANSMITTER
2006
SELECT DESIRED TRANSMISSION HEAD
2008
PROVIDE RF RECEIVER
2010
SELECT DESIRED RECEIVER HEAD
2012
ARRANGE TRANSMISSION HEAD
2014
ARRANGE RECEIVER HEAD
2016
SELECT RF FREQUENCY(S)
2018
TRANSMIT RF SIGNAL
2020
COLLECT STEAM
2022
CONDENSE STEAM
2024
END
2026

FIG. 20

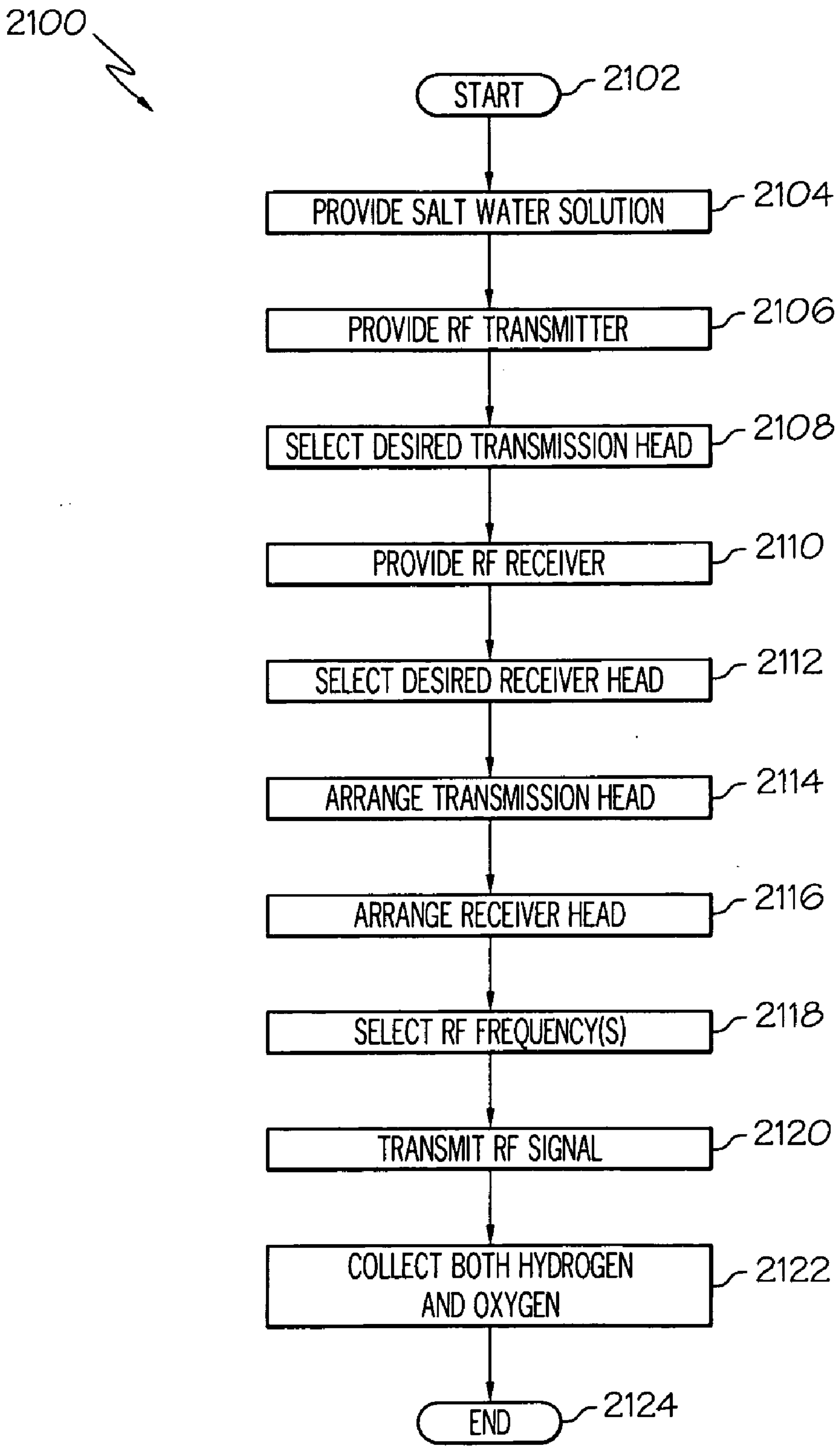
2100
START
2102
PROVIDE SALT WATER SOLUTION
2104
PROVIDE RF TRANSMITTER
2106
SELECT DESIRED TRANSMISSION HEAD
2108
PROVIDE RF RECEIVER
2110
SELECT DESIRED RECEIVER HEAD
2112
ARRANGE TRANSMISSION HEAD
2114
ARRANGE RECEIVER HEAD
2116
SELECT RF FREQUENCY(S)
2118
TRANSMIT RF SIGNAL
2120
COLLECT BOTH HYDROGEN AND OXYGEN
2122
END
2124

FIG. 21

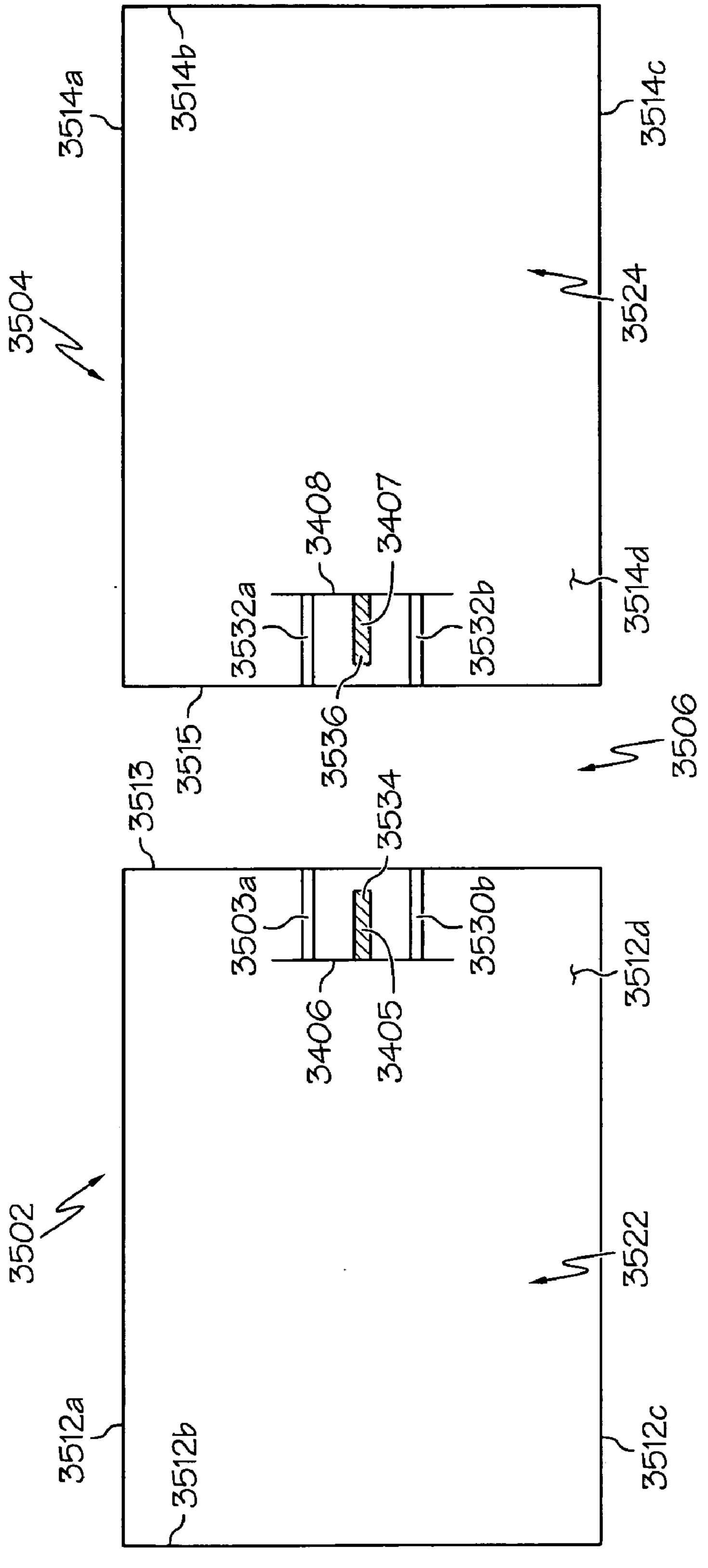

FIG. 22

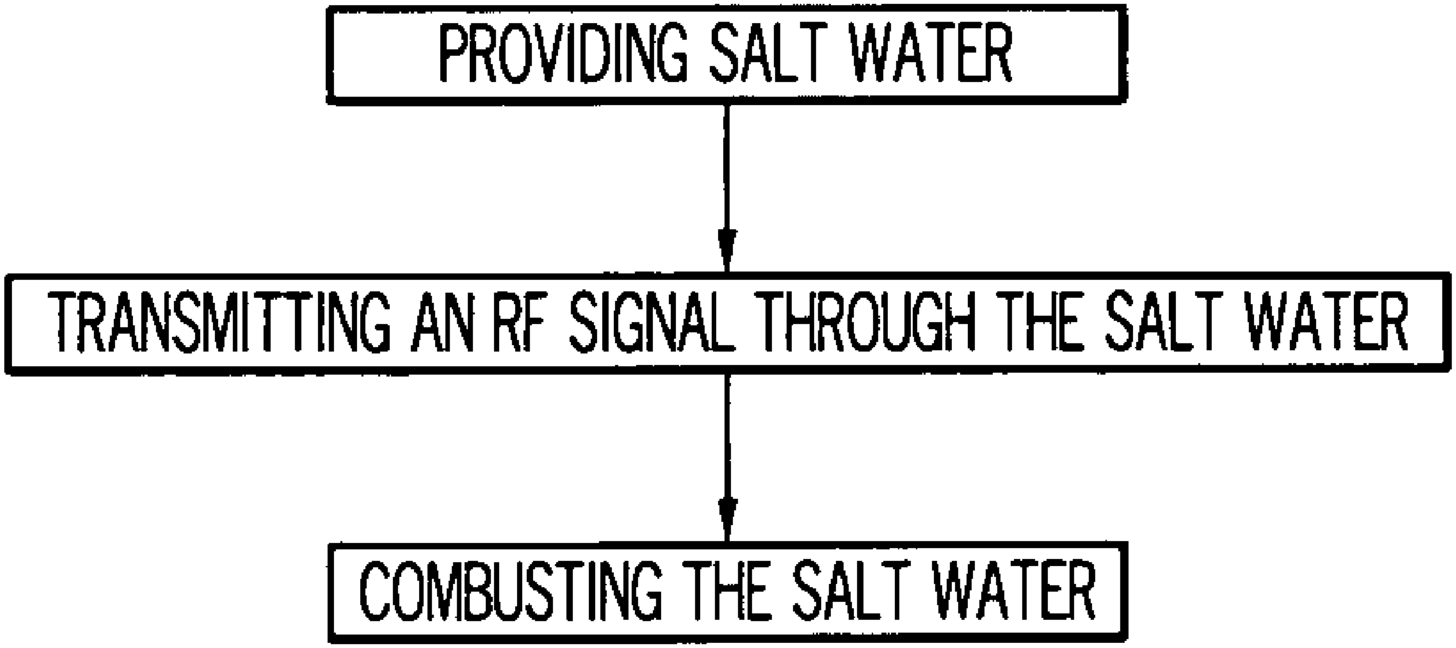
PROVIDING SALT WATER
TRANSMITTING AN RF SIGNAL THROUGH THE SALT WATER
COMBUSTING THE SALT WATER

FIG. 23

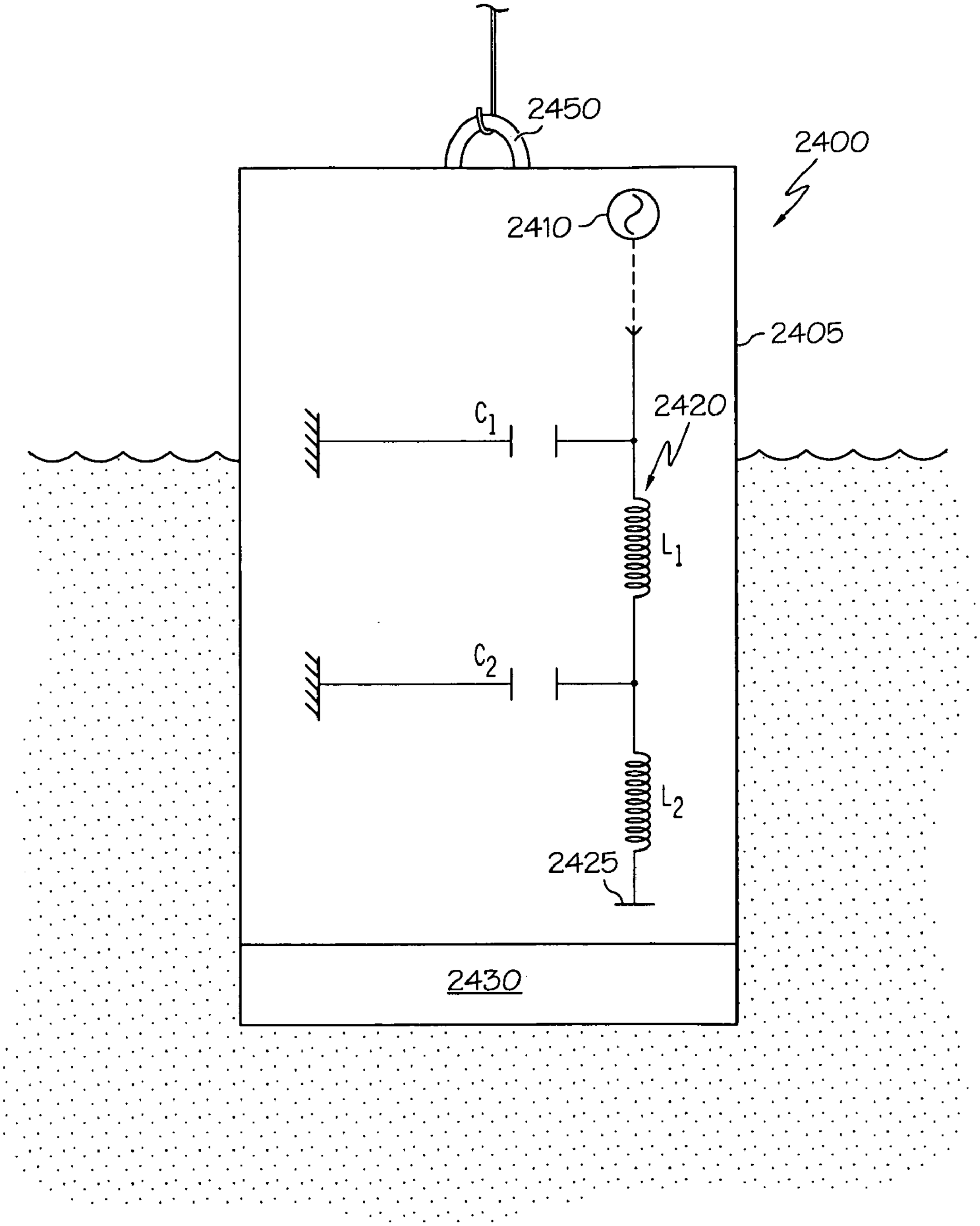
2450
2400
2410
2405
2420
C_1
L_1
C_2
L_2
2425
2430

FIG. 24

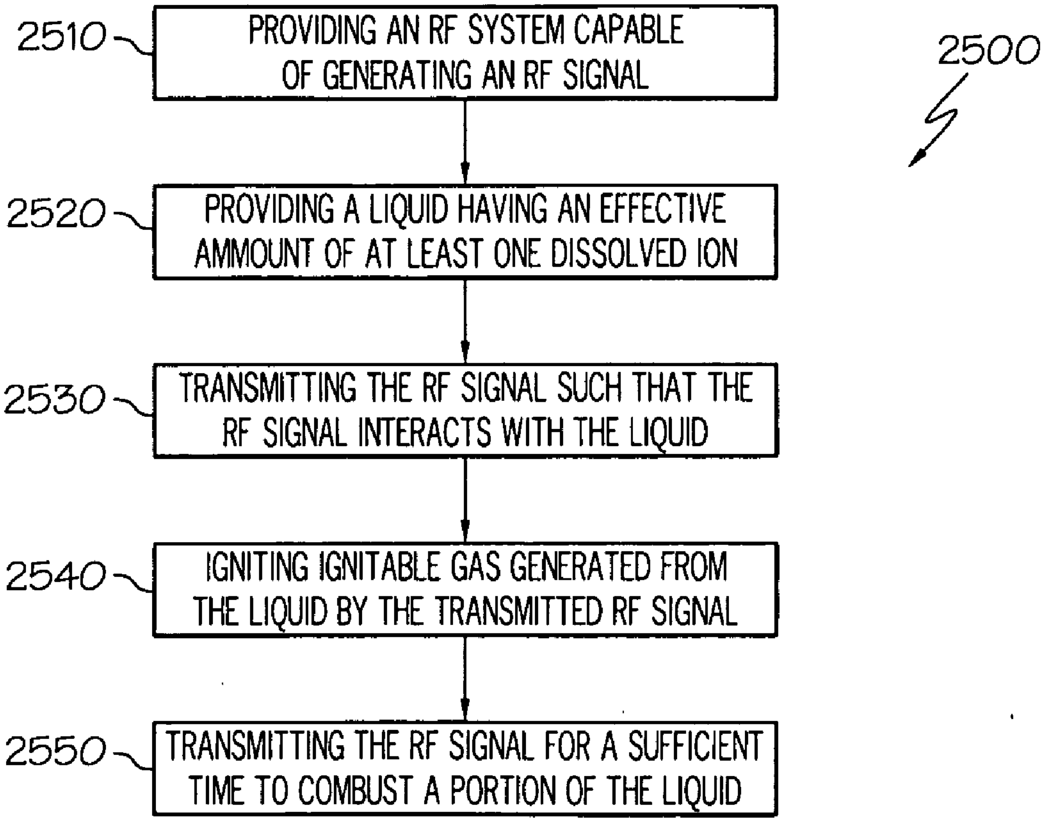
2500
2510
PROVIDING AN RF SYSTEM CAPABLE OF GENERATING AN RF SIGNAL
2520
PROVIDING A LIQUID HAVING AN EFFECTIVE AMMOUNT OF AT LEAST ONE DISSOLVED ION
2530
TRANSMITTING THE RF SIGNAL SUCH THAT THE RF SIGNAL INTERACTS WITH THE LIQUID
2540
IGNITING IGNITABLE GAS GENERATED FROM THE LIQUID BY THE TRANSMITTED RF SIGNAL
2550
TRANSMITTING THE RF SIGNAL FOR A SUFFICIENT TIME TO COMBUST A PORTION OF THE LIQUID

FIG. 25

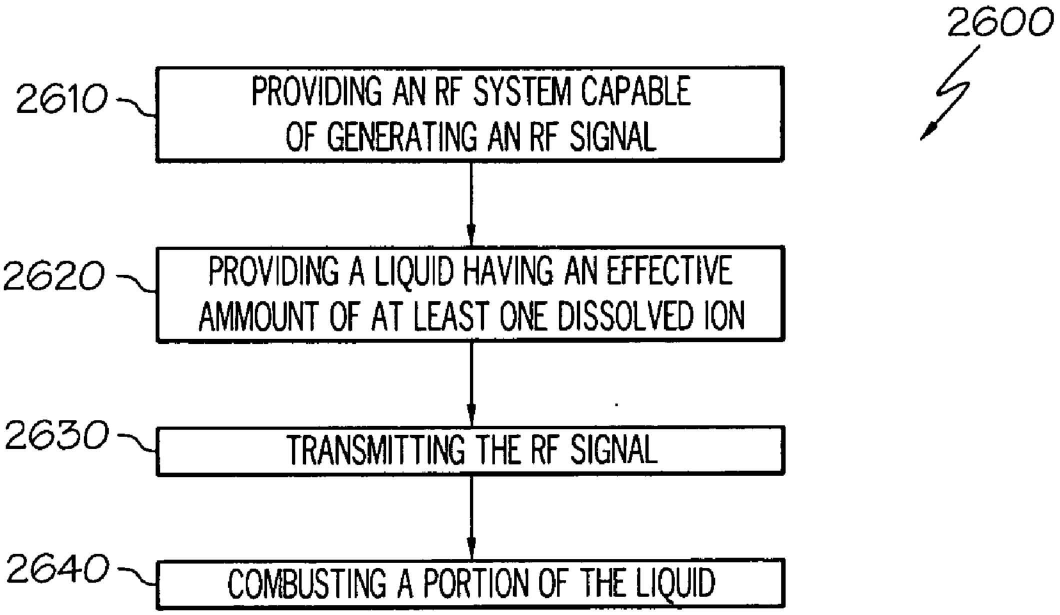
2600
2610
PROVIDING AN RF SYSTEM CAPABLE OF GENERATING AN RF SIGNAL
2620
PROVIDING A LIQUID HAVING AN EFFECTIVE AMMOUNT OF AT LEAST ONE DISSOLVED ION
2630
TRANSMITTING THE RF SIGNAL
2640
COMBUSTING A PORTION OF THE LIQUID

FIG. 26

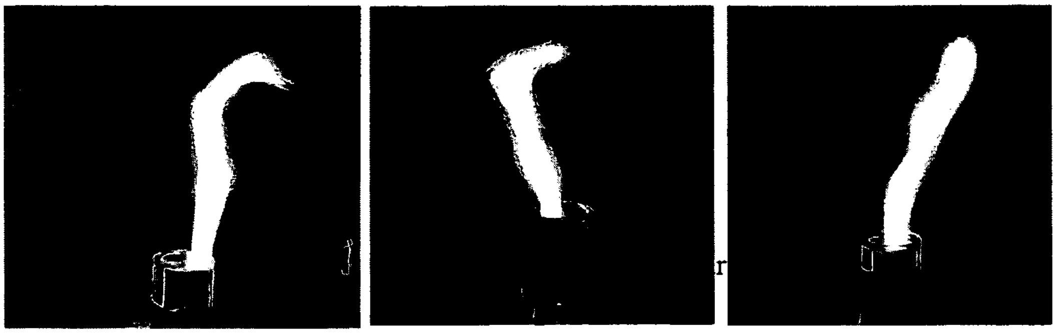

NaCl NaBr NaF

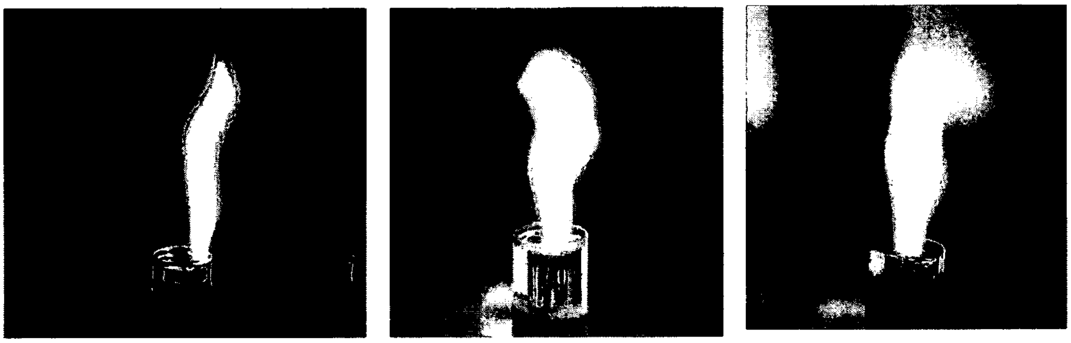

NaNO3 NaH2PO4 Na2SiO3

Figure 27

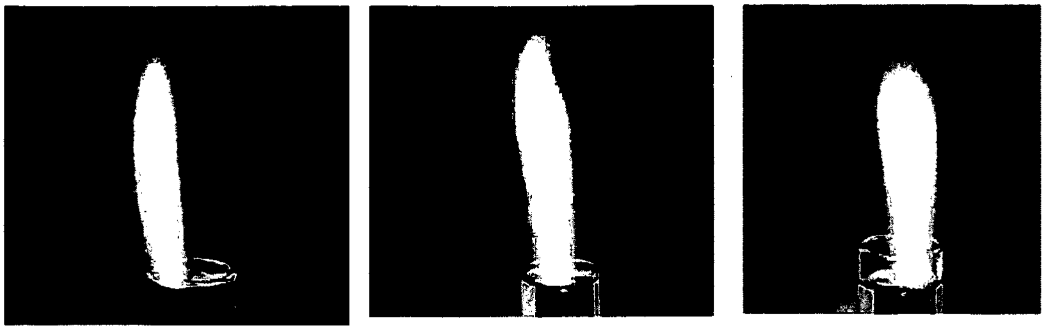

$Cr(NO_3)_3$ $Fe(NO_3)_2$ $Co(NO_3)_2$

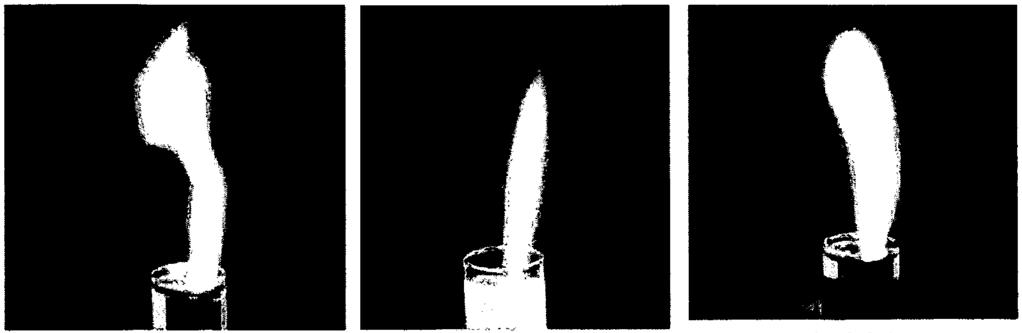

$Ni(NO_3)_2$ $Ni(OH)_2$ $CuSO_4$

Figure 28

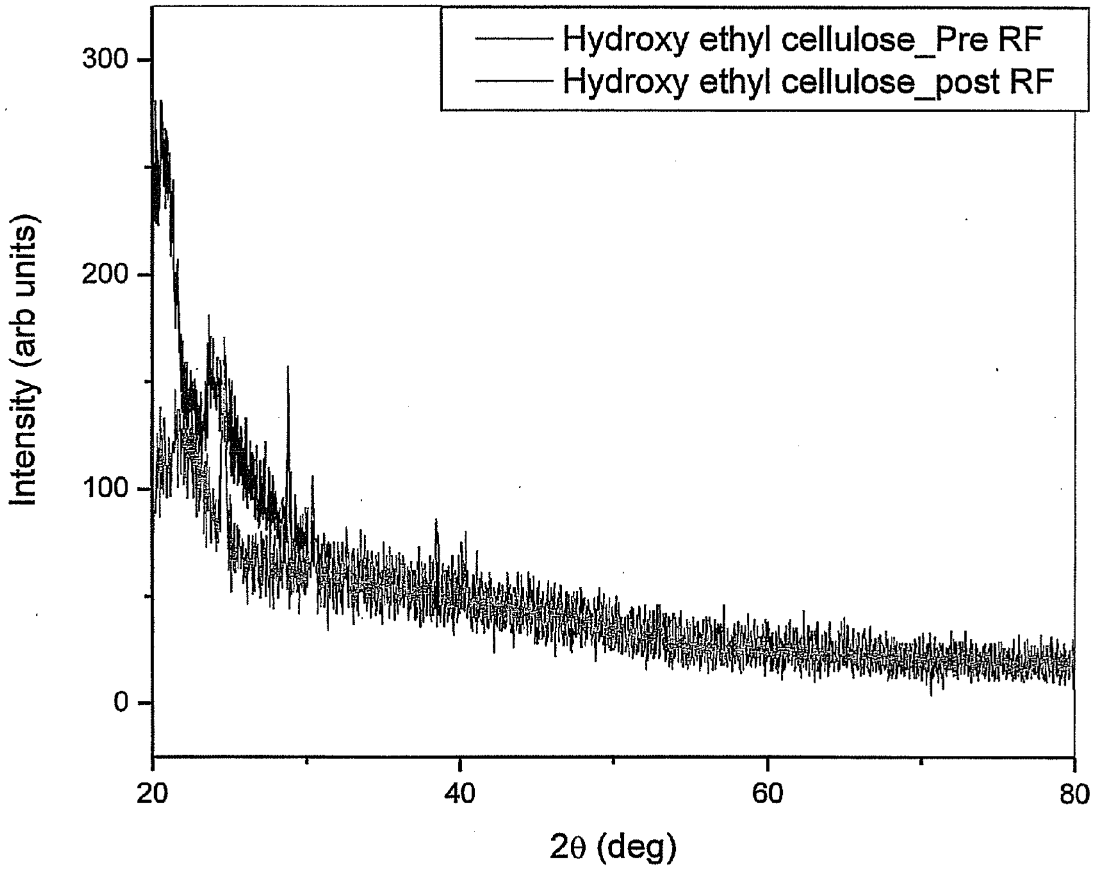
Hydroxy ethyl cellulose_Pre RF
Hydroxy ethyl cellulose_post RF
300
200
100
0
Intensity (arb units)
20
40
60
80
2θ (deg)

Figure 29

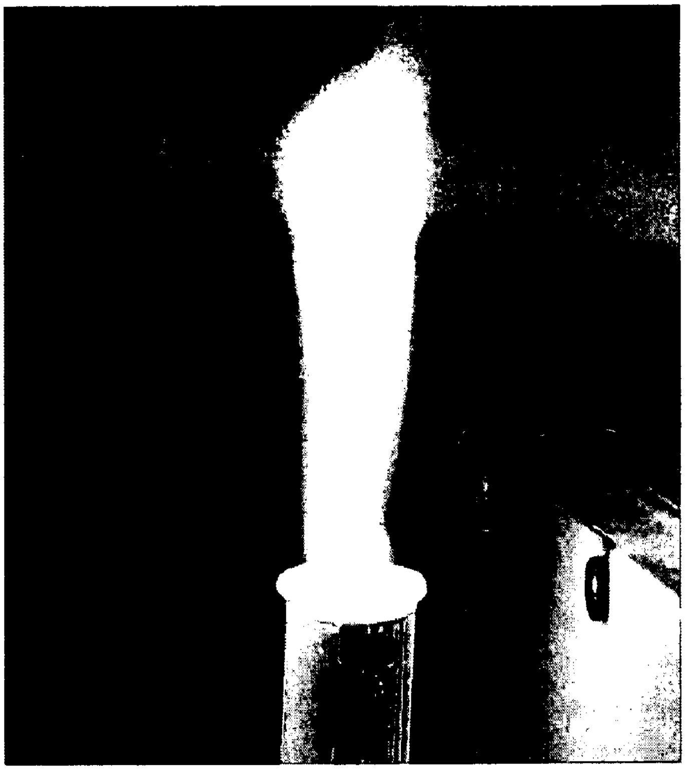

Figure 30 - 'burning' of Phosphoric acid in RF

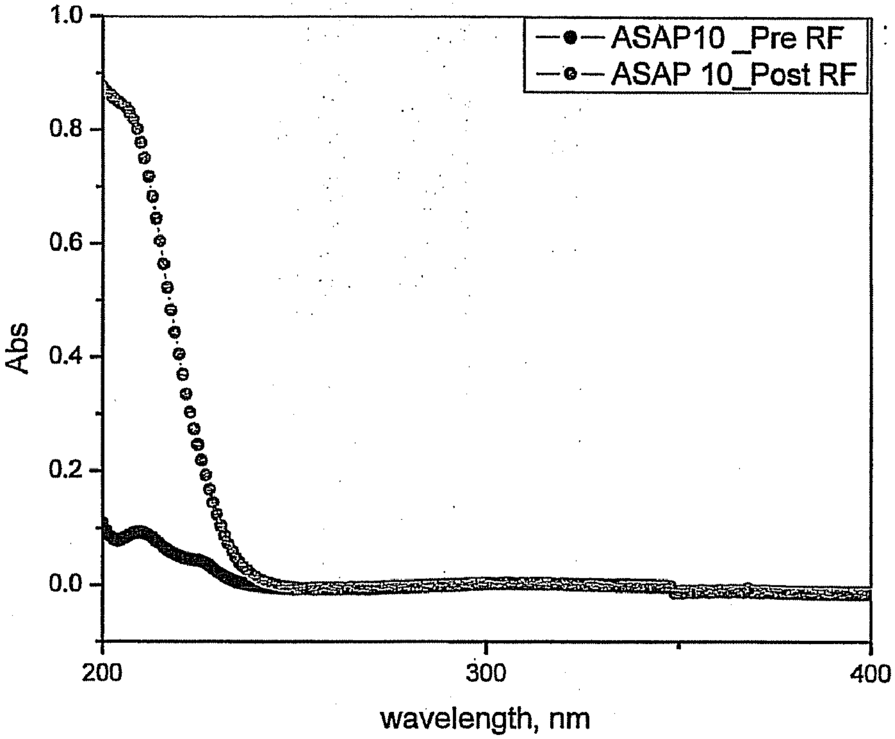

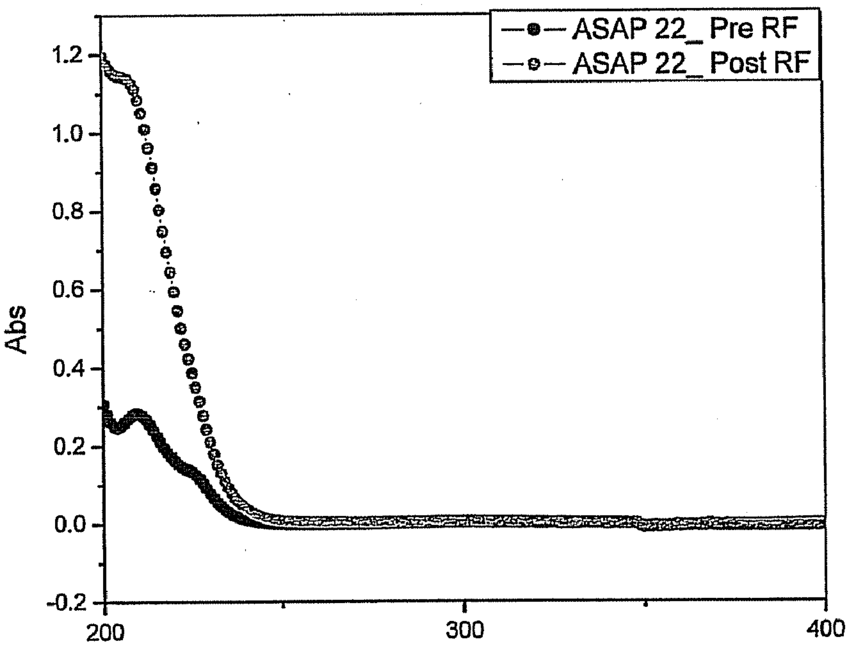

Figure 31: UV_VIS spectra of colloidal silver samples

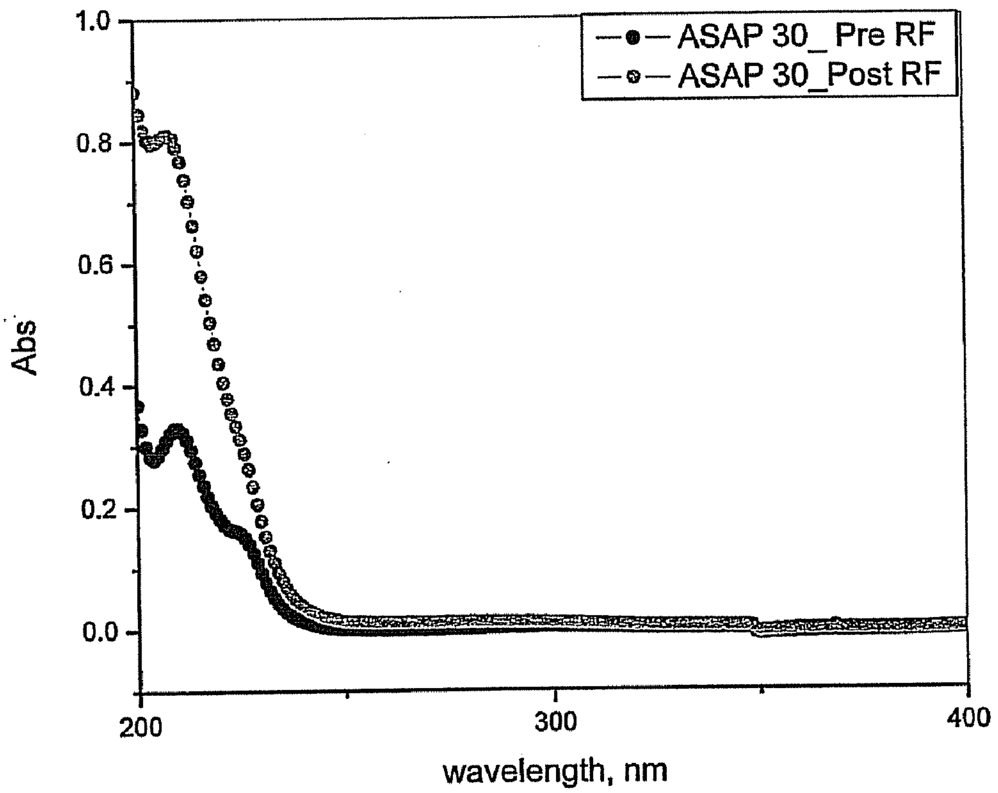

Figure 31 (con't): UV_VIS spectra of colloidal silver samples

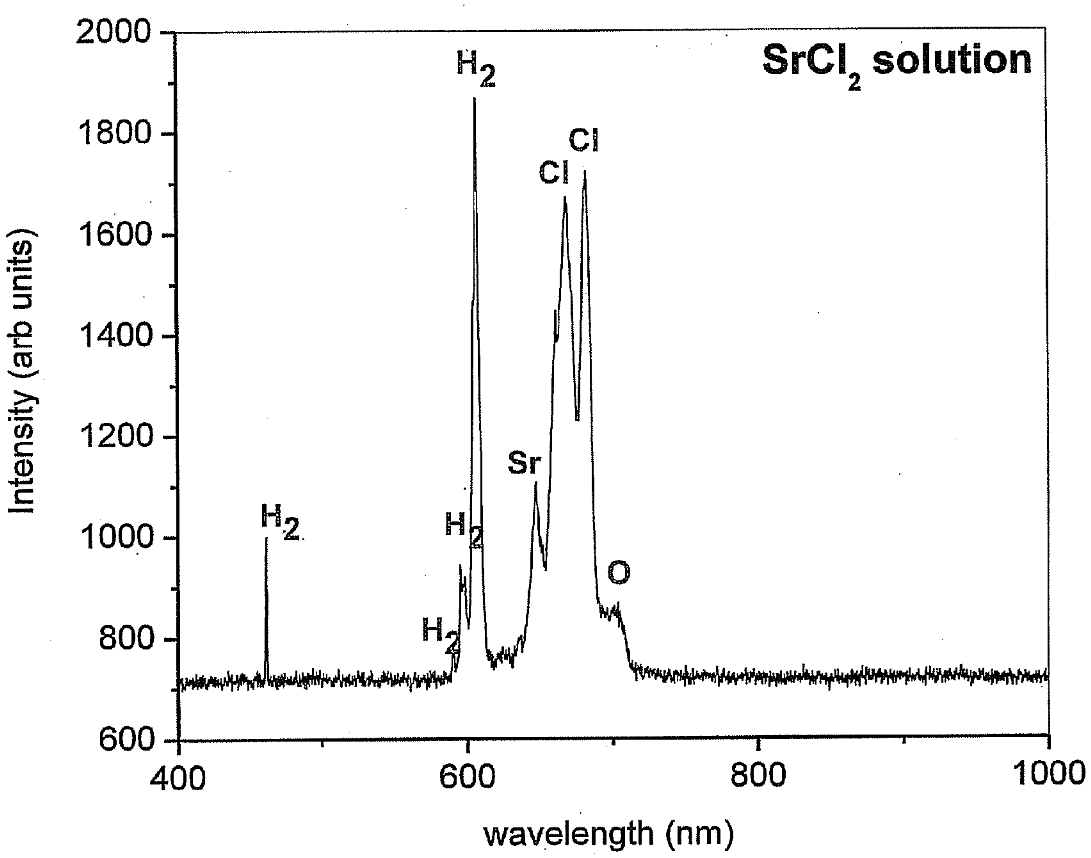

Figure 32: Emission spectra of SrCl2 in DI water indicative of evolution of H2 gas

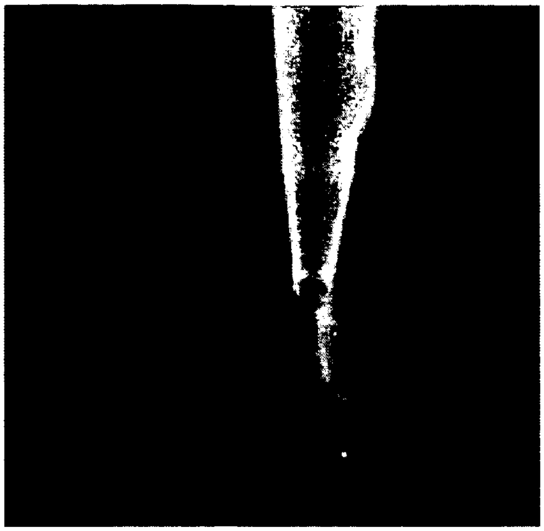

Figure 33: Self ignition of salt water droplet in the RF field

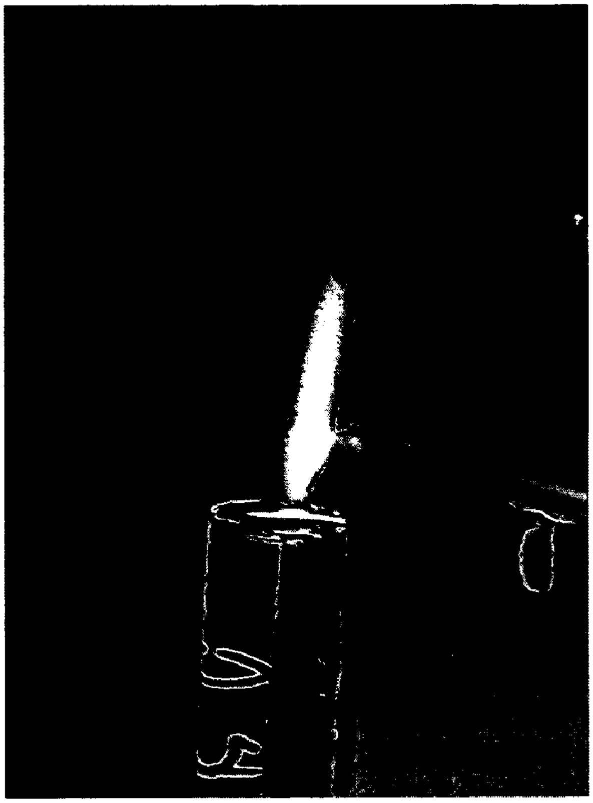

Figure 34: microwave - structured water from Steve Sedlmayr

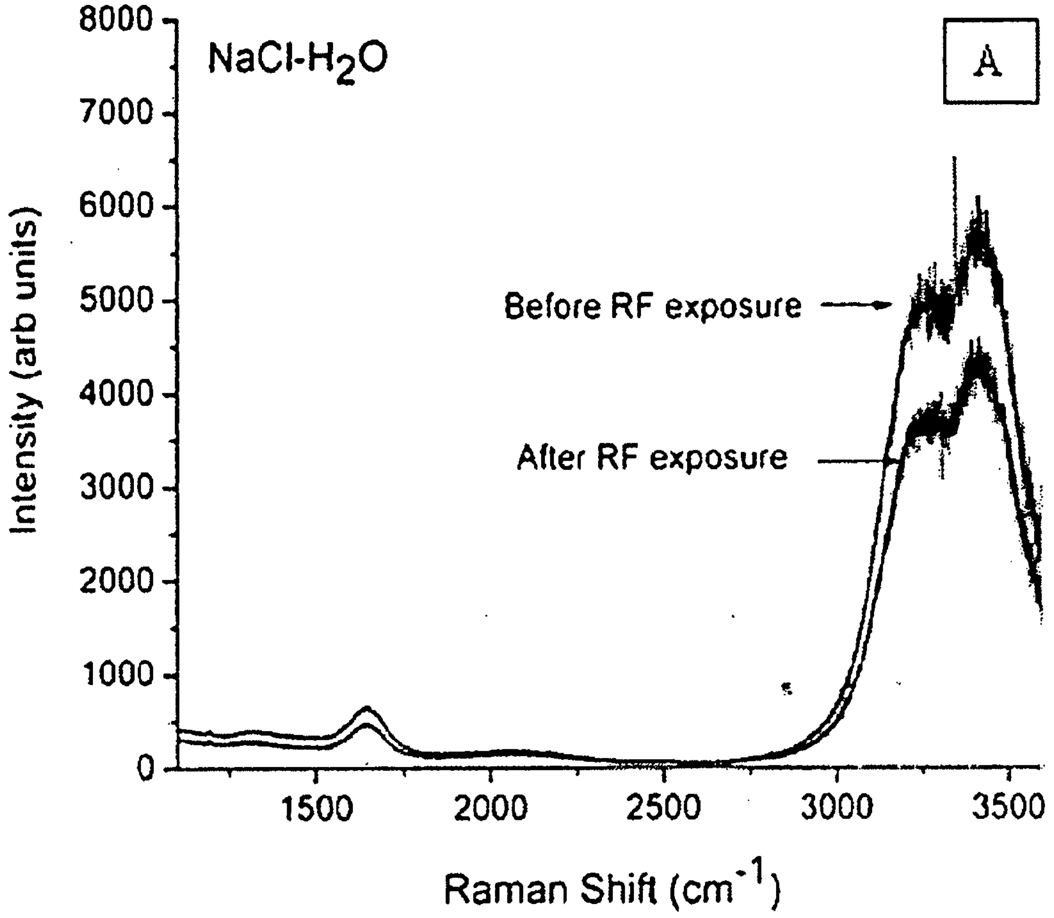
NaCl-H$_2$O
A
Before RF exposure
After RF exposure
Intensity (arb units)
Raman Shift (cm^{-1})
8000
7000
6000
5000
4000
3000
2000
1000
0
1500
2000
2500
3000
3500

Figure 35

RF SYSTEMS AND METHODS FOR PROCESSING SALT WATER

RELATED CASES

[0001] This application claims priority to, and any other benefit of, U.S. Provisional Patent Application No. 61/033,962, filed Mar. 5, 2008 and entitled "RF SYSTEMS AND METHODS FOR PROCESSING MATERIALS," which is herein incorporated by reference. This case is related to U.S. patent application Ser. No. 11/939,225 of Kanzius, filed Nov. 13, 2007, and entitled RF SYSTEMS AND METHODS FOR PROCESSING SALT WATER, (Attorney Docket 32403/04004); PCT Patent Application Serial No. PCT/U.S.07/84541 of Kanzius and Roy, filed Nov. 13, 2007, and entitled RF SYSTEMS AND METHODS FOR PROCESSING SALT WATER, (Attorney Docket 32403/04003); U.S. Provisional Patent Application Ser. No. 60/865,530 of Kanzius, filed Nov. 13, 2006, and entitled RF SYSTEM AND METHODS FOR PROCESSING SALT WATER (Attorney Docket 30064/04004) ("the '530 application"); U.S. Provisional Patent Application Ser. No. 60/938,613. of Kanzius, filed May 17, 2007, and entitled RF SYSTEM AND METHODS FOR PROCESSING SALT WATER II (Attorney Docket 30064/04008) ("the '613 application"); U.S. Provisional Patent Application Ser. No. 60/953,829 of Kanzius, filed Aug. 3, 2007, entitled RF SYSTEM AND METHODS FOR PROCESSING SALT WATER III (Attorney Docket 30064/04009); U.S. patent application Ser. No. 12/036,731 of Kanzius, filed on Feb. 25, 2008, and entitled FIELD GENERATOR FOR TARGETED CELL ABLATION (Attorney Docket 30274/04034); and U.S. Provisional Patent Application Ser. No. 60/915,345 of Kanzius, filed on May 1, 2007, and entitled FIELD GENERATOR FOR TARGETED CELL ABLATION (Attorney Docket 30274/04036), the entire disclosures of which, including all appendices, diagrams, figures, and photographs of which, are hereby incorporated by reference in their entireties.

FIELD OF THE INVENTION

[0002] The present invention relates to systems and methods for processing water utilizing radio frequency (RF) energy, such as, for example, RF systems and methods for combustion of salt water and/or solutions containing salt water, RF systems and methods for desalinating seawater, RF systems and methods for heating seawater, salt water, and/or solutions containing salt water, RF systems and methods for generating steam, RF systems and methods for volatilizing secondary fuels, RF systems and methods for the electrolysis of salt water and salt water mixtures, RF systems and methods for producing hydrogen from salt water and/or solutions containing salt water, RF systems and methods for combustion of volatiles produced from solutions containing salt water, and/or RF systems and methods for combustion of hydrogen produced from salt water and/or solutions containing salt water.

BACKGROUND OF THE INVENTION

[0003] Hydrogen gas is combustible and is therefore a potentially viable fuel source particularly for use in internal combustion engines. Water can be a source of hydrogen gas and unlike crude oil, which is used to produce gasoline, water and particularly seawater has an advantage over crude oil in that it is present on earth in great abundance. Furthermore, the burning of hydrogen produces water, an environmentally clean byproduct. Many other volatile organic compounds, such as ethanol for example, are also combustible and so they too are potentially viable fuel sources for use in internal combustion engines. Likewise, ethanol has an advantage over crude oil in that ethanol can be synthesized from fermentation of corn, sugar cane or other agricultural products and it is therefore a renewable resource, while by contrast crude oil is not.

BRIEF DESCRIPTION OF THE DRAWINGS

[0004] FIGS. **1-7** are high-level block diagrams of exemplary RF systems for RF processing of salt water and/or solutions containing salt water, such as combusting salt water or solutions containing salt water, generating steam from salt water, producing and collecting hydrogen from salt water or solutions containing salt water, and desalinating seawater;

[0005] FIGS. **8**A-**8**C, **9**A-**9**C are various views of exemplary RF transmission and RF reception heads;

[0006] FIGS. **10-12**, **16**, and **16***a* are schematic diagrams of exemplary RF circuits for exemplary RF systems for RF processing of salt water and/or solutions containing salt water, such as combusting salt water or solutions containing salt water, generating steam from salt water, producing and collecting hydrogen from salt water or solutions containing salt water, and desalinating seawater;

[0007] FIGS. **13-15** are top, top/side perspective, and side views of an exemplary RF coupling circuit for exemplary RF systems for RF processing of salt water and/or solutions containing salt water, such as combusting salt water or solutions containing salt water, generating steam from salt water, producing and collecting hydrogen from salt water or solutions containing salt water, and desalinating seawater;

[0008] FIG. **17** is a medium-level flowchart of an exemplary embodiment of an RF methodology for producing and collecting hydrogen gas from salt water and solutions containing salt water;

[0009] FIGS. **18**(*a*) and **18**(*b*) are medium level flow charts of exemplary embodiments of an RF methodology for producing and combusting hydrogen gas from salt water and for producing and combusting hydrogen gas and producing and combusting other volatiles from solutions containing salt water;

[0010] FIGS. **19**(*a*) and **19**(*b*) are medium level flow charts of exemplary embodiments of an RF methodology for producing and combusting hydrogen gas from salt water and for producing and combusting hydrogen gas and producing and combusting other volatiles from solutions containing salt water, and transferring the chemical energy generated by the combustion of the hydrogen gas and other volatiles into mechanical energy capable of moving a piston;

[0011] FIG. **20** is a medium level flow chart of an exemplary embodiment of an RF methodology for desalinating seawater;

[0012] FIG. **21** is a medium level flow chart of an exemplary embodiment of an RF methodology for carrying out the electrolysis of water;

[0013] FIG. **22** is a schematic illustration showing exemplary transmission and reception enclosures with their top walls removed;

[0014] FIG. **23** is a high-level flowchart showing an exemplary method of combusting salt water and solutions containing salt water with RF energy;

[0015] FIG. **24** is a schematic illustration showing an exemplary sealed transmission enclosure which may be suitable for lowering into the ground; and

[0016] FIGS. **25**-**26** are medium level flowcharts of exemplary embodiments of an RF methodology for combusting gas generated from a liquid by a transmitted RF signal.

[0017] FIGS. **27**, **28**, **30** and **34** are photographs showing combustion of the gas product generated by the electromagnetic dissociation of water in accordance with this invention.

[0018] FIGS. **29**, **31**, **32** and **35** are spectra of selected samples subjected to RF irradiation according to this invention.

[0019] FIG. **33** is a photograph showing that a salt water droplet self ignites when dropped into an RF field according to the technology of this invention.

SUMMARY

[0020] Systems are presented for using RF energy to combust salt water and/or various solutions containing salt water, to produce hydrogen from salt water, to produce volatiles from solutions containing salt water, to desalinate seawater, and/or to carry out the electrolysis of water. An exemplary system may comprise a reservoir for containing salt water that is a mixture comprising water and salt, the salt water having an effective amount of salt dissolved in the water; a reaction chamber having an inlet and an outlet; a feed line operatively connecting the reservoir to the inlet of the reaction chamber; an RF transmitter having an RF generator in circuit communication with a transmission head, the RF generator capable of generating an RF signal at least partially absorbable by the salt water having at least one frequency for transmission via the transmission head; and an RF receiver; wherein the reaction chamber is positioned such that at least a portion of the reaction chamber is between the RF transmission head and the RF receiver. Other exemplary systems may comprise a reservoir for containing a solution that is a mixture of water and salt and optionally containing (i) at least one additive, or (ii) at least one secondary fuel, or (iii) mixtures thereof.

[0021] Similarly, methods are presented for using RF energy to combust salt water and solutions containing salt water, to desalinate seawater, to produce hydrogen from salt water and solutions containing salt water, and/or to carry out the electrolysis of salt water. An exemplary method may comprise providing salt water comprising a mixture of water and at least one salt; or a salt water solution comprising a mixture of water and at least one salt and optionally containing (i) at least one additive, or (ii) at least one secondary fuel, or (iii) mixtures thereof; the salt water or salt water solution having an effective amount of the salt dissolved in the water; providing an RF transmitter having an RF generator in circuit communication with a transmission head, the RF generator capable of generating an RF signal at least partially absorbable by the salt water or salt water component of the solution containing salt water and having at least one frequency for transmission via the transmission head; arranging the transmission head near the salt water or solution containing salt water such that the RF signal transmitted via the transmission head interacts with at least some of the salt water; and transmitting the RF signal via the transmission head for a time sufficient to combust the salt water or to heat the solution containing salt water to volatilize and to combust a secondary fuel source that may be optionally present. If hydrogen gas is created from the salt water or the solution containing salt water by the RF signal, the RF signal may also be transmitted via the transmission head sufficient to combust the hydrogen gas so produced.

DETAILED DESCRIPTION

[0022] In the accompanying drawings which are incorporated in and constitute a part of the specification, exemplary embodiments of the invention are illustrated, which, together with a general description of the invention given above, and the detailed description given below, serve to example principles of the invention.

General Terms

[0023] "Additive" as used herein is a chemical compound having solubility, miscibility, or compatibility with various solutions of salt water (including sea water, salt water, or solutions containing salt water and optionally containing at least one secondary fuel) that furthermore is capable of altering the responsiveness of the various solutions of salt water to stimulation by RF energy.

[0024] "Circuit communication" as used herein is used to indicate a communicative relationship between devices. Direct electrical, optical, and electromagnetic connections and indirect electrical, optical, and electromagnetic connections are examples of circuit communication. Two devices are in circuit communication if a signal from one is received by the other, regardless of whether the signal is modified by some other device. For example, two devices separated by one or more of the following—transformers, optoisolators, digital or analog buffers, analog integrators, other electronic circuitry, fiber optic transceivers, or even satellites—are in circuit communication if a signal from one reaches the other, even though the signal is modified by the intermediate device (s). As a final example, two devices not directly connected to each other (e.g. keyboard and memory), but both capable of interfacing with a third device, (e.g., a CPU), are in circuit communication.

[0025] "Combustion" as used herein indicates a process that rapidly produces heat and light (perhaps caused by a rapid chemical change and with or without "burning" or "oxidation" in the classic sense). Salt water and solutions containing salt water respond to RF energy in many of the various systems and methods taught herein with rapid heating and rapid generation of light, which may be visible, UV, IR, etc. This is considered "combustion" herein, even though it may or may not be "burning" in the classic sense. "Combustion" herein also is used to indicate more typical incendiary "combustion," i.e., the process of burning in which a rapid chemical change occurs that produces heat and light, which includes burning in the classical sense of the products produced from salt water reacting with RF. For example, when hydrogen is combusted or burned in air the hydrogen is chemically oxidized into water and undergoes such a rapid reaction that a flame is produced and the water is discharged in the form of steam.

[0026] "Desalinate" as used herein is used to indicate the process of removing salt and other chemicals from water. For example, when desalination of seawater is carried out through heating, e.g., boiling, steam is produced and collected. When the collected steam is subsequently condensed back into a liquid, pure water is obtained free of any salt or minerals. "Electrolysis" as used herein is used to indicate the process of applying energy to water in order to decompose the water into its constituent elements hydrogen and oxygen. Energy can be applied in the form of either electrical energy, as for example in the application of an electric current, or in the form of heat energy.

[0027] "Operatively connected" or "operatively connecting" as used herein is used to indicate that a functional connection (e.g., a mechanical or physical connection or an electrical or optical or electromagnetic or magnetic connection) exists between the components of a system.

[0028] "Salt water" as used herein is used to indicate a mixture comprising water and salt, the salt water having an effective amount of salt dissolved in the water. "Solution containing salt water" and "salt water solutions" are used interchangeably and as used herein indicate a mixture comprising salt water and optionally containing one or more of the following: (i) at least one additive, (ii) at least one secondary fuel, or (iii) mixtures of both. Hence, a solution containing salt water may comprise only salt water. "Salt water mixture" as used herein is used to indicate a mixture containing salt water that is used in conducting electrolysis with the various systems and methods taught herein.

[0029] "Secondary fuel" as used herein is used to indicate combustible organic compounds that can be made volatile and that have solubility, miscibility, or compatibility with various salt water solutions (including salt water, sea water, or salt water solutions containing salt water and optionally containing at least one additive). As used herein, a secondary fuel may be the only substance that is combusting; thus, use of the term secondary fuel does not necessary require that there is a primary fuel also combusting. Salt and salt solutions may be used to increase the combustion of secondary fuels without the salt or salt solution also combusting.

Systems

[0030] Referring to the drawings and to FIGS. **1**-**16**A, various different views of exemplary systems and system components are shown. It is believed that these systems and components may be used with virtually all the various RF absorption enhancers and virtually all the various methods discussed herein.

[0031] The exemplary systems of FIGS. **1**-**4** include an RF generator **102** in circuit communication with a transmission head **104** for transmitting through a reaction chamber **106** an RF signal **108** generated by the RF generator **102** and transmitted by the transmitter head **104**. The reaction chamber **106** may be open or closed, depending on the specific application. The reaction chamber may be, for example, a vessel or a cylinder with an associated piston.

FIG. 1

[0032] Referring to FIG. **1**, there is shown a first exemplary embodiment of an RF system **100** that uses an RF signal **108** to process solutions containing salt water **110** in the reaction chamber **106**. For example, the RF signal **108** may combust the solution containing salt water **110**. As another example, the RF signal **108** may heat the solution containing salt water **110** for further processing, e.g., steam collection and condensing to desalinate a solution containing salt water **110**. As yet another example, the RF signal **108** may produce hydrogen from the solution containing salt water **110** or the RF signal may heat the solution containing salt water and volatilize any secondary fuel that may be optionally contained in the solution. The hydrogen produced as well as any volatilized secondary fuel optionally present may be collected as a gas and stored for various uses, e.g., stored for use as a fuel. Alternative, the hydrogen or any volatilized secondary fuel or both may be combusted in the reaction chamber **106**. Exemplary system **100** comprises an RF generator **102** in circuit communication with a transmission head **104**. A reaction chamber **106** is positioned such that at least a portion of the reaction chamber **106** is RF coupled to the transmission head **104**. In exemplary system **100**, the RF generator **102** communicates an RF signal for transmission to the transmission head **104**. The RF signal **108** transmitted by the transmission head **104** passes through at least a portion of the reaction chamber **106**. A solution containing salt water (and also a solution optionally containing (i) at least one additive, (ii) at least one secondary fuel, or (iii) mixtures thereof) **110** contained within the reaction chamber **106** is positioned such that the solution containing salt water **110** (and in particular the salt water component of the solution) absorbs at least some of the RF signal **108**. Optionally, the RF generator **102** may be controlled adjusting the frequency and/or power and/or envelope, etc. of the generated RF signal and/or may have a mode in which an RF signal at a predetermined frequency and power are transmitted via transmission head **104**. In addition, optionally, the RF generator **102** provides an RF signal **108** with variable amplitudes, pulsed amplitudes, multiple frequencies, etc.

[0033] The solution containing salt water **110** absorbs energy as the RF signal **108** travels through the reaction chamber **106**. The more energy that is absorbed by the salt water component of the solution containing salt water **110** the higher the temperature increase in the area which leads to water decomposition and hydrogen production, and in instances where the solution containing salt water **110** also contains a secondary fuel, this may also lead to volatization and to combustion of the secondary fuel instead of or in addition to decomposition of the salt water and hydrogen production. As even more energy is absorbed by the salt water component of the solution containing salt water **110**, combustion of the hydrogen that is being produced eventually occurs. The rate of energy absorption by the solution containing salt water **110** can be increased by increasing the RF signal **108** strength, which increases the amount of energy traveling through the reaction chamber **106**. Other means of increasing the rate of energy absorption may include but are not limited to concentrating the signal on a localized area of the solution containing salt water **110**, or further mixing with the solution containing salt water at least one additive that is appropriately selected from various chemical species to be capable of altering the rate of energy absorption of the solution containing salt water **110** and as a result may be able to increase the rate of energy absorption by the solution containing salt water **110**. Examples of additives that it is believed may be useful in this regard include surfactants, chemical species that form azeotropic mixtures with water, and chemical species that alter the freezing point of water.

FIGS. 2-4

[0034] As shown in FIGS. **2**-**4**, exemplary systems may also include a receiver head **112** and an associated current path **114** to permit the RF signal **108** to be coupled through the reaction chamber **106**. The systems **200**, **300**, **400** also use an RF signal **108** to process solutions **110** in the reaction chamber **106**. For example, the RF signal **108** may combust the solution containing salt water **110**. As another example, the

RF signal **108** may heat the salt water component of the solution containing salt water **110** in preparation for further processing (e.g.: in instances where the solution containing salt water **110** is salt water alone, steam collection and condensing to desalinate the salt water; in instances where the solution containing salt water contains a secondary fuel, the volatization of the secondary fuel). As yet another example, the RF signal **108** may produce hydrogen from or may volatilize a secondary fuel contained within the solution containing salt water **110** and the hydrogen or the volatilized secondary fuel or both may be collected as a gas and stored for various uses, e.g., stored for use as a fuel. In the alternative, the hydrogen produced or the volatilized secondary fuel or both may be combusted in the reaction chamber **106**.

[0035] Referring to FIG. **2**, the exemplary system **200** has a transmission head **104** and receiver head **112** arranged proximate to and on either side at least a portion of the reaction chamber **106**. This allows at least a portion of the solution containing salt water **110** in the reaction chamber **106** to be exposed to the RF signal **108** transmitted by the transmission head **104**. Some portion of the RF system may be tuned so that the receiver head **112** receives at least a portion of the RF signal **108** transmitted via the transmission head **104**. As a result, the receiver head **112** receives the RF signal **108** that is transmitted via the transmission head **104**.

[0036] The heads **104**, **112** may each or both have associated tuning circuitry such as pi-networks or tunable pi-networks, to increase throughput and generate a voltage in the area of the reaction chamber **106** and in the solution containing salt water salt **110** contained within. Thus, as shown in FIG. **3**, the transmission head **104** may have an associated tuning circuit **116** in circuit communication between the RF generator **102** and the transmission head **104**. Additionally, or in the alternative, as shown in FIG. **3**, the current path **114** may comprise the receiver head **112** being grounded.

[0037] Referring to FIG. **3**, the transmission head **104** and receiver head **112** may be insulated from direct contact with the reaction chamber **106**. The transmission head **104** and receiver head **112** may be insulated by means of an air gap **118**. An optional means of insulating the transmission head **104** and receiver head **112** from the reaction chamber **106** is shown in FIG. **4**. The exemplary system **400** includes inserting an insulating layer or material **410** such as, for example, Teflon® between the heads **104**, **112** and the reaction chamber **106**. Other optional means include providing an insulation area on the heads **104**, **112**, and allowing the heads to be put in direct contact with the reaction chamber **106**. The transmission head **104** and the receiver head **112**, described in more detail below, may include one or more plates of electrically conductive material.

[0038] One optional method of inducing a higher temperature in the solution containing salt water **110** includes using a receiver head **112** that is larger than the transmission head **104** (although it was earlier believed that a smaller head would concentrate the RF to enhance RF heating, a larger reception head was found to generate a higher temperature, perhaps because of the use of a high-Q resonant circuit described in more detail below). For example, a single 6" circular copper plate may be used on the Tx side and a single square 9.5" copper plate may be used on the Rx side. Optionally, an RF absorption enhancer may be added to the solution containing salt water **110**. An RF absorption enhancer is any means or method of increasing the tendency of the solution containing salt water **110** to absorb more energy from the RF signal that the salt water component of the solution containing salt water would otherwise absorb. Suitable RF absorption enhancers include, for example, suspended particles of electrically conductive material, such as metals, e.g., iron, various combination of metals, e.g., iron and other metals, or magnetic particles. The many types of RF absorption enhancers are discussed in greater detail below.

[0039] The RF generator **102** may be any suitable RF signal generator, generating an RF signal at any one or more of the RF frequencies or frequency ranges discussed herein. The RF signal **108** generated by the RF generator **102** and transmitted by the transmission head **104** may have a fundamental frequency in the HF range or the VHF range or an RF signal at some other fundamental frequency. The RF signal **108** may be a signal having one or more fundamental frequencies in the range(s) of 1-2 MHz, and/or 2-3 MHz, and/or 3-4 MHz, and/or 4-5 MHz, and/or 5-6 MHz, and/or 6-7 MHz, and/or 7-8 MHz, and/or 8-9 MHz, and/or 9-10 MHz, and/or 10-11 MHz, and/or 11-12 MHz, or 12-13 MHz, or 13-14 MHz, or 14-15 MHz. The RF signal **108** may have a fundamental frequency at 13.56 MHz. The RF generator **102** may be an ENI Model No. OEM-12B (Part No. OEM-12B-07) RF generator, which is marked with U.S. Pat. No. 5,323,329 and is known to be used to generate a 13.56 MHz RF signal for etching systems. Among other things, the ENI OEM-12B RF generator has an RF power on/off switch to switch a high-power (0-1250 Watt) RF signal, has an RF power output adjust to adjust the power of the signal generated, and has an RF power meter to measure the power of the RF signal being generated that can be switched to select either forward or reverse power metering. The power meter in reverse mode can be used to calibrate a tuning circuit, as explained above, by adjusting any variable components of the tuning circuit until minimum power is reflected back to the power meter (minimum VSWR). The ENI OEM-12B RF generator may be cooled by a Thermo Neslab Merlin Series M33 recirculating process chiller. A at 13.56 MHz RF signal from the ENI OEM-12B RF generator having a power of about 800-1000 Watts will combust salt water. In the alternative, the RF generator may be a commercial transmitter, e.g., the transmitter portion of a YAESU brand FT-1000MP Mark-V transceiver. An RF signal can be generated at about 13.56 MHz (one of the FCC-authorized frequencies for ISM equipment) by the transmitter portion of a YAESU brand FT-1000MP Mark-V transceiver by clipping certain blocking components as known to those skilled in the art. The RF generator and transmission head may have associated antenna tuner circuitry (not shown) in circuit communication therewith or integral therewith, e.g., automatic or manual antenna tuner circuitry, to adjust to the impedance of transmission head and the reaction chamber (and a receiver, if any). The transmitter portion of a YAESU brand FT-1000MP Mark-V transceiver has such integral antenna tuner circuitry (pressing a "Tune" button causes the unit to automatically adjust to the load presented to the RF generator portion). The RF generator and transmission head may have associated antenna tuner circuitry (not shown) in circuit communication therewith or integral therewith, e.g., automatic or manual antenna tuner circuitry, to adjust to the combined impedance of the reaction chamber and the receiver and compensate for changes therein. The transmitter portion of a YAESU brand FT-1000MP Mark-V transceiver has such integral antenna tuner circuitry. Various configurations for the transmission head and reception head are possible, as exemplified herein.

FIGS. 5-6

[0040] The transmission head **104** may be any of a number of different transmitter head configurations, such as an electrically conductive plate having a coaxial coil in circuit communication therewith. In the alternative, as exemplified by FIG. **5**, the transmission head **104** may comprise (or consist of) an electrically conductive plate **502** (e.g., a 6" diameter, flat, planar plate made of 0.020" stainless steel) without a corresponding coil. The transmission plate **502** may be circular and may be sized depending on the size of the target area and the desired voltage field generated by the plate. Similarly, as exemplified by FIG. **6**, the receiver head **112** may comprise (or consist of) an electrically conductive plate **602** (e.g., a 6" diameter, flat, planar plate made of 0.020" stainless steel) without a corresponding coil. The reception plate **602** may be circular and may be sized depending on the size of the target area and the desired voltage field generated by the plate. The reception plate **602** may be sized substantially smaller or substantially larger than the transmission plate **502** to change the field generated in the reaction chamber **106** by the coupled RF signal **108**. In the alternative, either the reception plate **602** or the transmission plate **502** (which includes both of them) may be parabolic plates with their convex side facing the target area (not shown). The plates may be made of copper (e.g., 0.090" copper plate) instead of stainless steel.

FIG. 7-9

[0041] In the alternative, the transmission head **104** or receiver head **112** may each or both be comprised of a series of spaced, stacked electrically conductive plates. The spaced, stacked electrically conductive plates may be coaxial, circular plates and may have sequentially decreasing diameters. FIG. **7** shows an exemplary system **700** wherein the receiver head **112** comprising spaced, stacked, electrically conductive, coaxial, and circular plates that have sequentially decreasing diameters. The plates of exemplary receiver head **800** may be constructed as described in FIGS. **8**A-**8**C (e.g., sized as shown with an Aluminum base) and may be insulated from each other as described in FIGS. **8**A-**8**C. The plates may be made of copper (e.g., 0.090" copper plate) instead of stainless steel.

[0042] Similarly, the transmission head **104** may comprise a series of spaced, stacked electrically conductive plates. The spaced, stacked electrically conductive plates may be coaxial, circular plates and may have sequentially decreasing diameters. FIGS. **9**A-**9**C show an exemplary transmission head **900** comprising spaced, stacked, electrically conductive, coaxial, and circular plates that have sequentially decreasing diameters. The plates of exemplary transmission head **900** may be constructed as described in FIGS. **9**A-**9**C (e.g., sized as shown with a Teflon base) and may be insulated from each other as described in FIGS. **9**A-**9**C. In the alternative, plates of exemplary receiver head **800** and/or the plates of exemplary transmission head **900** may be in circuit communication with each other, e.g., directly electrically coupled in their spaced configuration with electrically conductive fasteners. The plates may be made of copper (e.g., 0.090" copper plate) instead of stainless steel. A transmission head **900** with electrically insulated plates may be used with a receiver head **800** with electrically connected plates, and vice versa.

FIGS. 10-16

[0043] The tuning circuit **116** may be in circuit communication between the RF generator **102** and the transmission head **104** and may comprise and pi-network or a tunable pi-network. An exemplary tuning circuit **1000** is shown in FIG. **10** formed with components listed in that figure. Exemplary component values for FIGS. **10**-**16***a* are shown in Table A. Tuning circuit **1000** may be connected between an RF generator **102** and a transmission head **104**. Thus, as shown in FIG. **11** an exemplary system may include an ENI OEM-12B RF generator in circuit communication with exemplary tuning circuit **1000**, which is in circuit communication with exemplary transmission head **900** to generate an RF signal **108** through the reaction chamber **106** by coupling the RF signal **108** to a receiver head **112**. The receiver head **112** may be the same as exemplary receiver head **800**, as shown in the exemplary system of FIG. **11**.

[0044] The exemplary implementation of the exemplary tuning circuit **1000** used in FIGS. **10**-**15** appears to show a voltage gain of about 15-to-1 with respect to the voltage of the RF signal generated by the ENI RF generator. Thus exemplary tuning circuit **1000** may be considered to be a voltage step up transformer. Voltages of the larger plate of the transmission head have been estimated to be in excess of 40,000 volts per inch. Accordingly, some or all of the transmission head and/or the receiving head may be sealed, enclosed in an enclosure, or otherwise encapsulated in an insulating material.

[0045] FIGS. **13**-**15** show different views of an exemplary implementation of portions of the exemplary system of FIG. **12**. As shown in those figures, in implementing the exemplary tuning circuit **1000** used in FIGS. **10**-**12**, the larger inductor L_2 may be positioned with its longitudinal axis substantially coaxial with the central axis of plates of transmission head FP_1, and the central axis of the small inductor L_1 may be substantially perpendicular to the longitudinal axis of the larger inductor L_2. Other components may be used to implement tuning circuit **1000** instead of the exemplary components listed on FIGS. **10**-**12**. For example, the smaller inductor L_1 may be silver-coated or may be made of 12 turns of 5⁄16" copper tubing (or more turns of larger diameter copper tubing) for increased current carrying capacity (smaller inductor L_1 can get relatively hot in exemplary embodiments), and the capacitor C_1 may be made from thirteen (13) 100 pF capacitors instead of eleven (11) for a 1300 pF capacitor C_1. As another example, the plates in the heads may be made of copper (e.g., made from 0.090" copper plate) instead of stainless steel. In the exemplary implementation shown in FIGS. **13**-**15**, a region of the target area slightly closer to the transmission head (about 60/40 distance ratio) heats slightly more than dead center between the two heads. The grounded portion of the components of FIGS. **10**-**15** may be mounted to a copper sheet **1300** or other suitable conducting sheet, and the conducting stand of reception head FP_2 may be mounted on a copper sheet **1500** or other suitable conducting sheet, as shown in FIG. **15**. The grounded plates **1300**, **1500** may be connected by one or more copper straps **1302**.

FIG. 16

[0046] FIG. **16** shows another exemplary system **1600** that is the same as system **1200** (shown in FIGS. **8**A-**8**C, **9**A-**9**C, **12**-**15** and as described above), except the transmission head FP_1' has a single 6" plate, the one 6" circular plate of transmission head FP_1, and the three **6**" and 4" and 3" plates of receiver head FP_2 are made from 0.090" thick copper, capacitor C_1 is 1300 pF instead of 1100 pF, and the smaller inductor L_1 is silver-coated and made of 12 turns of 5⁄16" copper tubing. FIG. **16***a* shows another exemplary system **1600** that is the same as system **1600** except that the receiver head FP_2' has a single 6" circular plate. The transmitting portion and the receiving portion may be enclosed in one or more suitable enclosures, e.g., enclosures **3502**, **3504** in FIG. **22**. Open circuit voltage readings at the transmission head of exemplary physical embodiments have taken. Open circuit voltages of the RF field at 100 W of transmitted power have been measured with a broadband oscilloscope at about 6000 volts (e.g., about 5800 V) peak-to-peak amplitude, which rises to about 22,000 volts at 1000 W of transmitted power (FIG. **16**A in the configuration of FIGS. **13-15**). Additionally, it is believed that in these exemplary systems the voltage and current are not in phase (e.g., out of phase by a certain phase angle). Additionally, perhaps improved RF heating efficiency and/or RF transmission efficiency may be realized by changing the phase relationship between the voltage and current to a predetermined phase angle or real-time determined (or optimal) phase angle. In addition, the Q of exemplary physical embodiments have been estimated using bandwidth (S9 or 3 dB point) in excess of 250 (e.g., 250-290) (FIG. **16**A in the configuration of FIGS. **13-15**). As should be apparent, the RF heating using these exemplary embodiments is significantly different than inductive heating (even substantially different from inductive heating at similar frequencies).

[0047] As shown in FIG. **22**, the circuits may be mounted in two enclosures: a transmission enclosure **3502** and a reception enclosure **3504**, with a reaction chamber **3506** there between. Exemplary transmission enclosure **3502** has grounded metallic walls **3512** on all sides except the side **3513** facing the reception enclosure **3504** (only four such grounded walls **3512***a*-**3512***d* of five such walls **3512** of exemplary transmission enclosure **3502** are shown; the top grounded wall has been removed). Similarly, exemplary reception enclosure **3504** has grounded metallic walls **3514** on all sides except the side **3515** facing the transmission enclosure **3502** (only four such grounded walls **3514***a*-**3514***d* of five such walls **3514** of exemplary reception enclosure **3504** are shown; the top grounded wall has been removed). The grounded walls **3512** of transmission enclosure **3502** are in circuit communication with the grounded walls **3514** of reception enclosure **3504**. Facing walls **3513** and **3515** may be made from TEFLON or another suitable electrical insulator. Transmission enclosure **3502** and/or reception enclosure **3504** may be movably mounted to permit variable spacing between the transmission head and the reception head to accommodate create differently-sized reaction chambers **3506**. Facing walls **3513** and **3515** may have associated openings (not shown) to which various racks and other structures can be connected to support a body part or other target structure between the transmission head and the reception head. Dispersive pads (not shown) may be provided for direct grounding of the target or capacitive grounding of the target structure, which grounding pads may be connected to the grounded walls **3512**, **3514** (such direct or capacitive grounding pads may be help smaller target structures absorb relatively higher levels of RF and heat better). The transmission side components **3522** may be mounted inside exemplary transmission enclosure **3502** and the reception side components **3524** may be mounted inside exemplary reception enclosure **3504**. Exemplary transmission enclosure **3502** and reception enclosure **3504** both may be cooled with temperature-sensing fans that turn on responsive to the heat inside the enclosures **3502**, **3504** reaching a predetermined thermal level. Exemplary transmission enclosure **3502** and reception enclosure **3504** also have a plurality of pass-through connectors, e.g., permitting the RF signal to pass from the RF signal generator into the exemplary transmission enclosure **3502** (perhaps via a power meter) and permitting the received signal to pass outside exemplary reception enclosure **3504** to a power meter and back inside reception enclosure **3504**. In this exemplary embodiment, the enclosures **3502**, **3504** may be moved to vary the spacing between the distal, adjacent ends of the heads from about two inches to a foot or more apart. Various other embodiments may have different ranges of spacing between the distal, adjacent ends of the heads, e.g., from about 2" to about 20" or more apart or from about 2" to about 40" or more apart.

[0048] Each such enclosure may have grounded (e.g., aluminum) walls with a grounded (e.g., copper) base plate, except for the walls proximate the transmission head FP_1' and the reception head FP_2., which may be made from an electrical insulator such as ceramic or TEFLON brand PTFE, e.g., TEFLON brand virgin grade electrical grade PTFE, or another insulator. The walls may be grounded to the copper plate using copper straps and, if a plurality of enclosures are used, the enclosures may have copper strap between then to ground the enclosures together. A long standard fluorescent light bulb can be used to confirm effective grounding (e.g., by turning on the RF signal and repeatedly placing the light bulb proximate the transmission head to illuminate the bulb and then moving the bulb to locations around the enclosure watching for the light bulb to cease illumination, which confirms acceptable grounding). The grounded walls may have a layer of electrical insulator on the inside thereof, such as ceramic or TEFLON brand PTFE, e.g., TEFLON brand virgin grade electrical grade PTFE, or another insulator.

[0049] The exemplary systems of FIGS. **12-16** are believed to generate a very high voltage field in the target area, which very high voltage field can be used to heat many different types of RF absorbing particles as part of RF absorption enhancers in connection with the various methods taught herein. For example, the exemplary systems of FIGS. **12-16** are believed to be capable of heating and combusting salt water solutions in connection with the various methods taught herein.

[0050] FIG. **24** illustrates an exemplary transmission arrangement **2400** that is adapted for at least partial submersion in a liquid. The enclosure includes a sealed circuit housing **2405** in which is enclosed a tuning circuit **2420** and a transmission head **2425**. The tuning circuit receives an RF signal from an RF generator **2410** that may be enclosed in the enclosure as shown or located outside of the enclosure **2405**. An insulated region **2430**, e.g., an air pocket or pocket of another gas, is disposed between the transmission head **2425** and the enclosure **2405**. The enclosure may also include a mounting means, such as a hook or loop **2450**, that is used to mechanically couple the enclosure to a cable or other similar mechanism for lowering the enclosure into a hole or confined treatment area, e.g., with a winch or crane (not shown) or other means for mowering. If the RF generator **2410** is located outside the sealed enclosure **2405**, an insulated electrical conductor (not shown) may be provided to place the circuit **2420** in circuit communication with the RF generator. During construction, air from the portion of the enclosure **2405** surrounding the coupling circuit may be evacuated and the enclosure **2405** filled with an inert gas, such as nitrogen or xenon and then sealed. The coupling circuit may be tunable or not (e.g., pre-tuned), and may be the same as any of the coupling circuits shown or described herein, with virtually any of the transmission heads shown herein. If the coupling circuit portion of the enclosure **2405** is filled with an inert gas, it is believed that much higher powered RF signals may be coupled using the various coupling circuits disclosed herein, e.g., FIGS. **13-15** or FIG. **16***a*. In the alternative, if the coupling circuit portion of the enclosure **2405** is filled with an inert gas, it is believed that significantly smaller coupling circuits may be used vis-á-vis the exemplary coupling circuit of FIGS. **13-15**, because smaller components may be used (by increasing the voltage break down of the coupled components within the enclosure). If the coupling circuit is tunable, such tuning may be accomplished using remotely controllable tunable components, e.g., variable capacitors having stepper motors configured to change the value of the capacitor, or with remote cables to remotely mechanically change the value of the capacitor. Thus, a control unit remove from the enclosure (not shown) may be used to send electrical signals to tune the circuit to reduce or remove reflected power or a user may mechanically remotely tune the circuit to reduce or remove reflected power. Although a grounded reception head (not shown) may be used in this configuration (e.g., also mounted to the enclosure and configured to permit water to flow between the transmission and reception heads or between the insulated region and the reception head) it is believed that it may be possible to tune the circuit without a reception head per se, using the target water as a receiver and a current path (as a sort of grounded reception head).

Methods

[0051] Solutions containing salt water and that optionally contain (i) at least one additive, or (ii) at least one secondary fuel, or (iii) mixtures thereof may be combusted using RF signals by passing a high-voltage RF signal through the solution containing salt water. In a general sense, the methods may be characterized by providing a solution containing salt water and that may optionally contain (i) at least one additive, or (ii) at least one secondary fuel, or (iii) mixtures thereof and passing an RF signal through the solution containing salt water to combust the solution containing salt water (FIG. **23**). Alternatively, in a general sense the methods may be characterized as methods for adding salt to enhance the heating of water or other liquids. Salt water has been combusted using an exemplary system that included a circuit implementation of the circuit of FIG. **16** being used to transmit an RF signal through the salt water to combust the salt water. A solution of OCEANIC brand Natural Sea Salt Mix having a specific gravity of about 1.026 g/cm^3 was used. A 13.56 MHz RF signal from an ENI OEM-12B RF generator having a power of about 800-1000 Watts (e.g., about 900 Watts) was used to combust the salt water.

FIG. 17

[0052] FIG. **17** illustrates a high level exemplary methodology **1700** for producing hydrogen from salt water or from solutions containing salt water.

[0053] The methodology begins at block **1702**. At block **1704** the salt water is provided. The salt water comprises water and at least one salt wherein an effective amount of salt is dissolved in the water. In certain embodiments salt is added to water or other liquids to enhance heating. Optionally, a solution containing salt water may be used that contains salt water and (i) at least one additive, or (ii) at least one secondary fuel, or (iii) mixtures thereof. The salt can be any type of useful salt which is water soluble. Several examples of useful salts are described in greater detail below. An effective amount of salt is the amount of salt necessary to absorb sufficient energy output from the RF signal such that salt water or a solution containing salt water undergoes decomposition to generate hydrogen. OCEANIC brand Natural Sea Salt Mix may be used to approximate the composition of naturally occurring seawater having an effective amount of salt, and that may be used further as either salt water or as the salt water component in a solution containing salt water that is used in the systems and methods discussed and shown herein. Such approximations of naturally occurring seawater may have a specific gravity of about 1.02 g/cm^3 to 1.03 g/cm^3, e.g., between about 1.020-1.024 or about 28-32 PPT, as read off of a hydrometer. As an approximation of naturally occurring seawater, a mixture of water with the above-identified sea salt having a specific gravity of about 1.026 g/cm^3 (as measured with a refractometer) was used in exemplary systems and methods. In the alternative, it is believed that actual seawater may be used in the systems and methods discussed and shown herein.

[0054] It is contemplated that a reservoir of salt water or a solution containing salt water could be made beforehand and stored in a tank such that it would be available upon demand. For example, the storage tank could be connected to the reaction chamber by means of a feed tube. In this manner, a supply of the previously prepared salt water or solution could be pumped from the storage tank into the reaction chamber via the feed tube; wherein the feed tube has one end connected to the storage tank and the other end connected to an inlet present on the reaction chamber. Again, it is believed that ordinary sea water may be used.

[0055] At block **1706** an RF transmitter is provided. The RF transmitter may be any type of RF transmitter generating a suitable RF signal. RF transmitter may be a variable frequency RF transmitter. Optionally, the RF transmitter is also multi-frequency transmitter capable of providing multiple-frequency RF signals. Optionally the RF transmitter is capable of transmitting RF signals with variable amplitudes or pulsed amplitudes. One or more of a variety of different shapes and sizes of transmission and reception heads may be provided.

[0056] The transmission head may be selected at block **1708**. The selection of the transmission head may be based in part on the type of RF transmitter provided. Other factors, such as, for example, the depth, size and shape of the general target area, or specific target area to be treated, and the number of frequencies transmitted may also be used in determining the selection of the transmission head.

[0057] The RF receiver is provided at block **1710**. The RF receiver may be tuned to the frequency(s) of the RF transmitter. At block **1712**, the desired receiver head may be selected. Similarly to the selection of the transmission head, the receiver head may be selected to fit the desired characteristics of the particular application. For example, a receiver head that is larger than the transmission head can be selected to concentrate the RF signal on a specific area in the reaction chamber (although it was earlier believed that a smaller head would concentrate the RF to enhance RF heating, a larger reception head was found to generate a higher temperature). For example, a single 6" circular copper plate may be used on the Tx side and a single square 9.5" copper plate may be used on the Rx side. In this manner, selection of various sizes and shapes of the receiver heads allow for optimal concentration of the RF signal in the salt water mixture.

[0058] At block **1714** the transmission head is arranged. Arrangement of the transmission head is accomplished by, for example, placing the transmission head proximate to and on one side of the reaction chamber. At block **1716** the receiver head is arranged. Arrangement of the receiver head is similarly accomplished by, for example, placing the receiver head proximate to and on the other side of the reaction chamber so that an RF signal transmitted via the transmission head to the receiver head will pass through the reaction chamber and be absorbed by the salt water or the salt water component of the solution containing salt water. The transmission head and reception heads are insulated from direct contact with the reaction chamber. The heads may be insulated from the reaction chamber by means of an air gap. Optionally, the heads may be insulated from the target area by means of another insulating material.

[0059] The RF frequency(s) may be selected at block **1718**. In addition to selecting the desired RF frequency(s) at block **1718**, the transmission time or duration may also be selected. The duration time is set to, for example, a specified length of time, or set to raise the temperature of at least a portion of the salt water or the solution containing salt water to a desired temperature/temperature range, or set to a desired change in temperature. In addition, optionally, other modifications of the RF signal may be selected at this time, such as, for example, amplitude, pulsed amplitude, an on/off pulse rate of the RF signal, a variable RF signal where the frequency of the RF signal varies over a set time period or in relation to set temperatures, ranges or changes in temperatures.

[0060] At block **1720** the RF signal is transmitted from the transmission head to the receiver head. The RF signal passes through the reaction chamber and is absorbed by the salt water or the salt water component of the solution containing salt water that is contained within the reaction chamber. Absorption of the RF energy results in decomposition of the salt water or the salt water component of the solution containing salt water to generate hydrogen.

[0061] At block **1722** the hydrogen produced by decomposition of a salt water or solution containing salt water is collected. Hydrogen may be collected by any means. An example of a means for collecting hydrogen would be to utilize a vacuum or pump apparatus to remove the hydrogen gas as it is produced and to then retain the hydrogen in a location physically separated from the reaction chamber. For example, such a vacuum or pump apparatus could have one end attached to an outlet present on the reaction chamber and the other end attached to a gas storage container. It is contemplated that the gas storage container may be fitted with valves, as for example a one way valve, such that gas could enter or be pumped into the tank but then the gas could not leave the tank.

[0062] The methodology may end at block **1724** and may be ended after a predetermined time interval and/in response to a determination that a desired amount of hydrogen production has been achieved. The method may be performed once or repeatedly, or continuously, or periodically, or intermittently.

FIGS. **18**(*a*) and **18**(*b*)

[0063] FIG. **18**(*a*) illustrates a high level exemplary methodology **1800** for producing hydrogen from salt water and subsequently for the combustion of the hydrogen produced. FIG. **18**(*b*) illustrates a high level exemplary methodology **1800** for (i) sufficiently heating a solution containing salt water that may optionally contain a secondary fuel in order to volatilize and combust the secondary fuel; or (ii) decomposing the salt water component of the solution containing salt water to generate hydrogen and to subsequently combust the hydrogen produced; or (iii) both.

[0064] The methodology for both FIGS. **18**(*a*) and **18**(*b*) begins at block **1802**. At block **1804** either salt water or a solution containing salt water is provided. In FIG. **18**(*a*) the salt water comprises water and at least one salt, wherein an effective amount of salt is dissolved in the water. In certain embodiments salt is added to water or other liquids to enhance heating. In FIG. **18**(*b*) the salt water solution comprises the salt water of FIG. **18**(*a*) and optionally: (i) at least one additive, or (ii) at least one secondary fuel source, or (iii) mixtures thereof. The salt used in FIGS. **18**(*a*)-(*b*) can be any type of useful salt which is water soluble. Several examples of useful salts are described in greater detail below. An effective amount of salt is the amount of salt necessary to allow surrounding water to absorb sufficient energy output from the RF signal such that it undergoes decomposition to generate hydrogen, or the amount of salt necessary to allow surrounding water to absorb sufficient energy output from the RF signal such that it undergoes sufficient heating to volatilize and combust any secondary fuel source optionally present. OCEANIC brand Natural Sea Salt Mix may be used to approximate the composition of naturally occurring seawater having an effective amount of salt and that may be used further as the salt water component of the salt water containing solution in the systems and methods discussed and shown herein. Such approximations of naturally occurring seawater may have a specific gravity of about 1.02 g/cm^3 to 1.03 g/cm^3, e.g., between about 1.020-1.024 or about 28-32 PPT, as read off of a hydrometer. As an approximation of naturally occurring seawater, a mixture of water with the above-identified sea salt having a specific gravity of about 1.026 g/cm^3 (as measured with a refractometer) was used in exemplary systems and methods. In the alternative, it is believed that actual seawater may be used in the systems and methods discussed and shown herein.

[0065] It is contemplated that a reservoir of salt water or a solution containing salt water could be made beforehand and stored in a tank such that it would be available upon demand. For example, the storage tank could be connected to the reaction chamber by means of a feed tube. In this manner, a supply of the salt water or the salt water containing solution previously prepared could be pumped from the storage tank into the reaction chamber via the feed tube; wherein the feed tube has one end connected to the storage tank and the other end connected to an inlet present on the reaction chamber.

[0066] At block **1806** an RF transmitter is provided. The RF transmitter may be any type of RF transmitter generating a suitable RF signal. RF transmitter may be a variable frequency RF transmitter. Optionally, the RF transmitter may also be a multi-frequency transmitter capable of providing multiple-frequency RF signals. Still yet, optionally the RF transmitter may be capable of transmitting RF signals with variable amplitudes or pulsed amplitudes. A variety of different shapes and sizes of transmission and reception heads may be provided.

[0067] The transmission head may be selected at block **1808**. The selection of the transmission head may be based in part on the type of RF transmitter provided. Other factors, such as, for example, the depth, size and shape of the general target area, or specific target area to be treated, and the number of frequencies transmitted may also be used in determining the selection of the transmission head.

[0068] The RF receiver is provided at block **1810**. The RF receiver may be tuned to the frequency(s) of the RF transmitter. At block **1812**, the desired receiver head may be selected. Similarly to the selection of the transmission head, the receiver head may be selected to fit the desired characteristics of the particular application. For example, a receiver head that is larger than the transmission head can be selected to concentrate the RF signal on a specific area in the reaction chamber (although it was earlier believed that a smaller head would concentrate the RF to enhance RF heating, a larger reception head was found to generate a higher temperature). Various sizes and shapes of the receiver heads allow for optimal concentration of the RF signal in the salt water and solutions containing salt water.

[0069] At block **1814** the transmission head is arranged. Arrangement of the transmission head is accomplished by, for example, placing the transmission head proximate to and on one side of the reaction chamber. At block **1816** the receiver head is arranged. Arrangement of the receiver head is similarly accomplished by, for example, placing the receiver head proximate to and on the other side of the reaction chamber so that an RF signal transmitted via the transmission head to the receiver head will pass through the reaction chamber and be absorbed by the salt water or the salt water component of a solution containing salt water. The transmission head and reception heads are insulated from direct contact with the reaction chamber. The heads may be insulated from the reaction chamber by means of an air gap. Optionally, the heads may be insulated from the target area by means of another insulating material.

[0070] The RF frequency(s) may be selected at block **1818**. In addition to selecting the desired RF frequency(s) at block **1818**, the transmission time or duration may also be selected. The duration time is set to, for example, a specified length of time, or set to raise the temperature of at least a portion of the salt water or the solution containing salt water to a desired temperature/temperature range, or set to a desired change in temperature. In addition, optionally, other modifications of the RF signal may be selected at this time, such as, for example, amplitude, pulsed amplitude, an on/off pulse rate of the RF signal, a variable RF signal where the frequency of the RF signal varies over a set time period or in relation to set temperatures, ranges or changes in temperatures.

[0071] At block **1820** the RF signal is transmitted from the transmission head to the receiver head. The RF signal passes through the reaction chamber and is absorbed by the salt water or the salt water component of the solution containing salt water that is present within the reaction chamber. In FIG. **18**(*a*), absorption of the RF energy initially results in decomposition of the salt water to produce hydrogen, while still further absorption of the RF energy eventually leads to the combustion of the hydrogen produced by the decomposition of the salt water. In FIG. **18**(*b*), absorption of the RF energy initially results in (i) sufficiently heating the solution containing salt water in order to volatilize and to combust any secondary fuel that may be optionally present; or (ii) decomposition of the salt water component of the solution containing salt water to generate hydrogen; or (iii) both.

[0072] The methodology may end at block **1822** and may be ended after a predetermined time interval and/in response to a determination that a desired amount of hydrogen production and hydrogen combustion, or alternatively a desired amount of volatilization and combustion of the secondary fuel that may be optionally present is achieved. The method may be performed once or repeatedly, or continuously, or periodically, or intermittently.

FIGS. **19**(*a*) and **19**(*b*)

[0073] FIG. **19**(*a*) illustrates a high level exemplary methodology **1900** for producing hydrogen from salt water, for the combustion of the hydrogen produced, and for the subsequent conversion of this chemical energy into mechanical energy that moves a piston. FIG. **19**(*b*) illustrates a high level exemplary methodology **1900** for (i) sufficiently heating a solution containing salt water that may optionally contain a secondary fuel in order to volatilize and combust the secondary fuel; or (ii) decomposing the salt water component of the solution containing salt water to generate hydrogen and to subsequently combust the volatilized secondary fuel source or the hydrogen produced; or (iii) both; and for the subsequent conversion of the chemical energy that combustion releases into mechanical energy that moves a piston.

[0074] The methodology for both FIGS. **19**(*a*) and **19**(*b*) begins at block **1902**. At block **1904** either salt water or a solution containing salt water is provided. In FIG. **19**(*a*) the salt water comprises water and at least one salt wherein an effective amount of salt is dissolved in the water. In certain embodiments salt is added to water or other liquids to enhance heating. In FIG. **19**(*b*) the solution containing salt water comprises the salt water from FIG. **19**(*a*) and optionally (i) at least one additive, or (ii) at least one secondary fuel, or (iii) mixtures thereof. The salt can be any type of useful salt which is water soluble. Several examples of useful salts are described in greater detail below. An effective amount of salt is the amount of salt necessary to allow surrounding water to absorb sufficient energy output from the RF signal such that it undergoes decomposition to generate hydrogen, or the amount of salt necessary to allow surrounding water to absorb sufficient energy output from the RF signal such that it undergoes sufficient heating to volatilize and combust any secondary fuel source optionally present. OCEANIC brand Natural Sea Salt Mix may be used to approximate the composition of naturally occurring seawater having an effective amount of salt and that may be used further as the salt water component of the solutions containing salt water that are used in the systems and methods discussed and shown herein. Such approximations of naturally occurring seawater may have a specific gravity of about 1.02 g/cm^3 to 1.03 g/cm^3, e.g., between about 1.020-1.024 or about 28-32 PPT, as read off of a hydrometer. As an approximation of naturally occurring seawater, a mixture of water with the above-identified sea salt having a specific gravity of about 1.026 g/cm^3 (as measured with a refractometer) was used in exemplary systems and methods. In the alternative, it is believed that actual seawater may be used in the systems and methods discussed and shown herein.

[0075] It is contemplated that a reservoir of the salt water or a solution containing salt water could be made beforehand and stored in a tank such that it would be available upon demand. For example, the storage tank could be connected to the reaction chamber by means of a feed tube. In this manner, a supply of the salt water or the solution containing salt water previously prepared could be pumped from the storage tank into the reaction chamber via the feed tube; wherein the feed tube has one end connected to the storage tank and the other end connected to an inlet present on the reaction chamber. Alternatively, it is contemplated that a spray nozzle could be attached onto the end of the feed tube leading into the inlet present on the reaction chamber. In this arrangement it is believed that the salt water or the solution containing salt water could be introduced into the reaction chamber in the form of a mist or spray.

[0076] At block **1906** an RF transmitter is provided. The RF transmitter may be any type of RF transmitter generating a suitable RF signal. RF transmitter may be a variable frequency RF transmitter. Optionally, the RF transmitter may also be a multi-frequency transmitter capable of providing multiple-frequency RF signals. Still yet, optionally the RF transmitter may be capable of transmitting RF signals with variable amplitudes or pulsed amplitudes. A variety of different shapes and sizes of transmission and reception heads may be provided.

[0077] The transmission head may be selected at block **1908**. The selection of the transmission head may be based in part on the type of RF transmitter provided. Other factors, such as, for example, the depth, size and shape of the general target area, or specific target area to be treated, and the number of frequencies transmitted may also be used in determining the selection of the transmission head.

[0078] The RF receiver is provided at block **1910**. The RF receiver may be tuned to the frequency(s) of the RF transmitter. At block **1812**, the desired receiver head may be selected. Similarly to the selection of the transmission head, the receiver head is may be selected to fit the desired characteristics of the particular application. For example, a receiver head that is larger than the transmission head can be selected to concentrate the RF signal on a specific area in the reaction chamber (although it was earlier believed that a smaller head would concentrate the RF to enhance RF heating, a larger reception head was found to generate a higher temperature). Various sizes and shapes of the receiver heads allow for optimal concentration of the RF signal in the salt water and solution containing salt water.

[0079] At block **1914** the transmission head is arranged. Arrangement of the transmission head is accomplished by, for example, placing the transmission head proximate to and on one side of the reaction chamber. At block **1916** the receiver head is arranged. Arrangement of the receiver head is similarly accomplished by, for example, placing the receiver head proximate to and on the other side of the reaction chamber so that an RF signal transmitted via the transmission head to the receiver head will pass through the reaction chamber and be absorbed by the salt water or the salt water component of a solution containing salt water. The transmission head and receiving heads are insulated from direct contact with the reaction chamber. The heads may be insulated from the reaction chamber by means of an air gap. Optionally, the heads are insulated from the target area by means of another insulating material.

[0080] The RF frequency(s) may be selected at block **1918**. In addition to selecting the desired RF frequency(s) at block **1918**, the transmission time or duration may also be selected. The duration time is set to, for example, a specified length of time, or set to raise the temperature of at least a portion of the salt water or salt water solution to a desired temperature/temperature range, or set to a desired change in temperature. In addition, optionally, other modifications of the RF signal may be selected at this time, such as, for example, amplitude, pulsed amplitude, an on/off pulse rate of the RF signal, a variable RF signal where the frequency of the RF signal varies over a set time period or in relation to set temperatures, ranges or changes in temperatures.

[0081] At block **1920** the RF signal is transmitted from the transmission head to the receiver head. The RF signal passes through the reaction chamber and is absorbed by the salt water or the salt water component of the salt water containing solution present within the reaction chamber. In FIG. **19**(*a*), absorption of the RF energy initially results in decomposition of the salt water to produce hydrogen, while still further absorption of the RF energy eventually leads to the combustion of the hydrogen produced by the decomposition of the salt water. In FIG. **19**(*b*), absorption of the RF energy initially results in (i) sufficiently heating the solution containing salt water in order to volatilize and to combust any secondary fuel that may be optionally present; or (ii) decomposition of the salt water component of the aqueous solution to generate hydrogen; or (iii) both.

[0082] Alternatively, it is contemplated that an ignition source, for example a spark plug, could be attached to the reaction chamber. This ignition source would also be in circuit communication with a current source, such as for example a battery. The arrangement contemplated here would provide for a current going to the ignition source to be switched on and off when desired. This would result in generation of an ignition event, as for example with a spark plug a spark would be produced, on demand. It is believed that this ignition event would cause the combustion of the hydrogen that had been produced by the decomposition of the salt water, or would cause the combustion of either the hydrogen or any volatilized secondary fuel or both that is produced by RF treatment of a solution containing salt water in the reaction chamber.

[0083] At block **1922**, the energy generated from the combustion of hydrogen, which is produced from the decomposition of the salt water (or more generally, the energy generated from either (i) combustion of the hydrogen produced from decomposition of the salt water, or (ii) the volatilization and combustion of any secondary fuel that may be optionally present in a solution containing salt water, or (iii) both), is transmitted to a piston in order to perform mechanical work. In any event, the combustion of either the hydrogen or any secondary fuel or both generates hot exhaust gases including steam. These hot exhaust gases expand and in doing so create an increase in pressure. It is contemplated that the head of a piston could be attached to the outlet present on the reaction chamber and the other end of piston attached to a lever arm. As expanding exhaust gases push against the piston head, the lever arm is moved transforming the chemical energy of expanding exhaust gases into mechanical energy and into the performance of mechanical work.

[0084] It is further contemplated that this piston arrangement could be utilized together with the spray nozzle and ignition source described above, to allow one to convert chemical energy into mechanical energy and subsequently into the performance of mechanical work, on demand. For example, this method could be used in such an arrangement in order to power an internal combustion engine. It is further contemplated that one example of how this method together with the appropriate system could be utilized, would be in providing an engine that would be fueled by salt water or various solutions containing salt water, or even directly by seawater taken from the ocean without further purification, rather than requiring gasoline or other water incompatible hydrocarbon fuels to operate. Specifically, it is contemplated that this engine could be provided in an appropriate size and in a manner such that it could be used to power an automobile or other form of motorized vehicle.

[0085] The methodology may end at block **1924** and may be ended after a predetermined time interval and/in response to a determination that a desired amount of hydrogen production and hydrogen combustion, or alternatively that a desired amount of volatilization and combustion of any secondary fuel source that is optionally present has been achieved. The method may be performed once or repeatedly, or continuously, or periodically, or intermittently.

FIG. 20

[0086] FIG. **20** illustrates a high level exemplary methodology **2000** for desalinating seawater.

[0087] The methodology begins at block **2002**. At block **2004** seawater is provided. Any manner of seawater from any ocean or of any concentration or salinity would suffice. Furthermore, it is contemplated that the seawater could be taken from the source in its natural occurring form and used directly without the need for any further purification or processing. Examples of several sources for seawater are described below. It is also contemplated that an amount of seawater could be stored in a reservoir or storage tank such that it would be available to fill the reaction chamber upon demand. For example, the storage tank could be connected to the reaction chamber by means of a feed tube. In this manner, a supply of seawater could be pumped from the storage tank into the reaction chamber via the feed tube; wherein the feed tube has one end connected to the storage tank and the other end connected to an inlet present on the reaction chamber.

[0088] At block **2006** an RF transmitter is provided. The RF transmitter may be any type of RF transmitter generating a suitable RF signal. RF transmitter may be a variable frequency RF transmitter. Optionally, the RF transmitter may also be a multi-frequency transmitter capable of providing multiple-frequency RF signals. Still yet, optionally the RF transmitter may be capable of transmitting RF signals with variable amplitudes or pulsed amplitudes. A variety of different shapes and sizes of transmission and reception heads are provided.

[0089] The transmission head may be selected at block **2008**. The selection of the transmission head may be based in part on the type of RF transmitter provided. Other factors, such as, for example, the depth, size and shape of the general target area, or specific target area to be treated, and the number of frequencies transmitted may also be used in determining the selection of the transmission head.

[0090] The RF receiver is provided at block **2010**. The RF receiver may be tuned to the frequency(s) of the RF transmitter. At block **2012**, the desired receiver head may be selected. Similarly to the selection of the transmission head, the receiver head may be selected to fit the desired characteristics of the particular application. For example, a receiver head that is larger than the transmission head can be selected to concentrate the RF signal on a specific area in the reaction chamber (although it was earlier believed that a smaller head would concentrate the RF to enhance RF heating, a larger reception head was found to generate a higher temperature). Various sizes and shapes of the receiver heads allow for optimal concentration of the RF signal in the seawater.

[0091] At block **2014** the transmission head is arranged. Arrangement of the transmission head is accomplished by, for example, placing the transmission head proximate to and on one side of the reaction chamber. At block **2016** the receiver head is arranged. Arrangement of the receiver head is similarly accomplished by, for example, placing the receiver head proximate to and on the other side of the reaction chamber so that an RF signal transmitted via the transmission head to the receiver head will pass through the reaction chamber and be absorbed by the seawater. The transmission head and reception heads are insulated from direct contact with the reaction chamber. The heads may be insulated from the reaction chamber by means of an air gap. Optionally, the heads may be insulated from the target area by means of another insulating material.

[0092] The RF frequency(s) may be selected at block **2018**. In addition to selecting the desired RF frequency(s) at block **2018**, the transmission time or duration may also be selected. The duration time is set to, for example, a specified length of time, or set to raise the temperature of at least a portion of the seawater to boiling. In addition, optionally, other modifications of the RF signal are selected at this time, such as, for example, amplitude, pulsed amplitude, an on/off pulse rate of the RF signal, a variable RF signal where the frequency of the RF signal varies over a set time period or in relation to set temperatures, ranges or changes in temperatures or desired phase transitions.

[0093] At block **2020** the RF signal is transmitted from the transmission head to the receiver head. The RF signal passes through the reaction chamber and is absorbed by the seawater contained within the reaction chamber. Absorption of the RF energy results in heating of the seawater causing the seawater to undergo a phase change and produce steam. The steam produced would be free of any salt, minerals, or any other nonvolatile impurities initially present in the seawater.

[0094] At block **2022** the steam produced by heating the seawater to boiling is collected. At block **2024** the collected steam is condensed to form purified water. The steam may be collected by any means. An example of a means for collecting and condensing steam would be to utilize a the natural tendency of hot gases, such as steam, to rise. For example, it is contemplated that an exhaust pipe having one end attached to the outlet present in the reaction chamber and positioned to be directly above the reaction chamber could conduct the steam, as it is produced, away from the reaction chamber. It is further contemplated that the other end of the exhaust pipe could be attached to a remotely positioned tank and that this tank would functioned as a condenser such that, upon entering the tank, the steam would cool and convert phases from steam into water. As a result, it is believed that purified water would be condensed and collect in such a condenser tank. It is contemplated that, optionally, the condenser tank could be externally cooled in order to facilitate the rate of condensation of the steam.

[0095] The methodology may end at block **2026** and may be ended after a predetermined time interval and/in response to a determination that a desired amount of steam production and desalination has been achieved. The method may be performed once or repeatedly, or continuously, or periodically, or intermittently.

FIG. 21

[0096] FIG. **21** illustrates a high level exemplary methodology **2100** of carrying out the electrolysis of water.

[0097] The methodology begins at block **2102**. At block **2104** a salt water mixture is provided. The salt water mixture comprises water and at least one salt wherein an effective amount of salt is dissolved in the water. The salt should be water soluble and, in order to effectively form both hydrogen and oxygen gases, the salt should be selected such that the corresponding cation of the salt has a lower standard electrode potential than H^+ and the corresponding anion of the salt has a higher standard electrode potential than OH^-. A more detailed description of various salts and their effective amounts which are useful in this regard is given below.

[0098] At block **2106** an RF transmitter is provided. The RF transmitter may be any type of RF transmitter generating a suitable RF signal. RF transmitter may be a variable frequency RF transmitter. Optionally, the RF transmitter may also be a multi-frequency transmitter capable of providing multiple-frequency RF signals. Still yet, optionally the RF transmitter may be capable of transmitting RF signals with variable amplitudes or pulsed amplitudes. A variety of different shapes and sizes of transmission and reception heads may be provided.

[0099] The transmission head may be selected at block **2108**. The selection of the transmission head may be based in part on the type of RF transmitter provided. Other factors, such as, for example, the depth, size and shape of the general target area, or specific target area to be treated, and the number of frequencies transmitted may also be used in determining the selection of the transmission head.

[0100] The RF receiver is provided at block **2110**. The RF receiver may be tuned to the frequency(s) of the RF transmitter. At block **2112**, the desired receiver head may be selected. Similarly to the selection of the transmission head, the receiver head may be selected to fit the desired characteristics of the particular application. For example, a receiver head that is larger than the transmission head can be selected to concentrate the RF signal on a specific area in the reaction chamber (although it was earlier believed that a smaller head would concentrate the RF to enhance RF heating, a larger reception head was found to generate a higher temperature). Various sizes and shapes of the receiver heads allow for optimal concentration of the RF signal in the salt water mixture.

[0101] At block **2114** the transmission head is arranged. Arrangement of the transmission head is accomplished by, for example, placing the transmission head proximate to and on one side of the reaction chamber. At block **2116** the receiver head is arranged. Arrangement of the receiver head is similarly accomplished by, for example, placing the receiver head proximate to and on the other side of the reaction chamber so that an RF signal transmitted via the transmission head to the receiver head will pass through the reaction chamber and be absorbed by the salt water mixture. The transmission head and reception heads are insulated from direct contact with the reaction chamber. The heads may be insulated from the reaction chamber by means of an air gap. Optionally, the heads are insulated from the target area by means of another insulating material.

[0102] The RF frequency(s) may be selected at block **2118**. In addition to selecting the desired RF frequency(s) at block **2118**, the transmission time or duration may also be selected. The duration time is set to, for example, a specified length of time, or set to raise the temperature of at least a portion of the salt water mixture to a desired temperature/temperature range, or set to a desired change in temperature. In addition, optionally, other modifications of the RF signal are selected at this time, such as, for example, amplitude, pulsed amplitude, an on/off pulse rate of the RF signal, a variable RF signal where the frequency of the RF signal varies over a set time period or in relation to set temperatures, ranges or changes in temperatures.

[0103] At block **2120** the RF signal is transmitted from the transmission head to the receiver head. The RF signal passes through the reaction chamber and is absorbed by the salt water mixture contained within the reaction chamber. Absorption of the RF energy results in decomposition of the salt water mixture to produce hydrogen and oxygen.

[0104] At block **2122** both the hydrogen and oxygen produced by decomposition of the salt water mixture is collected. Means for collecting and separating the hydrogen and oxygen produced by the electrolysis of the salt water mixture will be known to those skilled in the art. Such techniques may include using two evacuated, gas collection bells that are nested within one another; where the opening to the innermost gas collection bell is covered with a semi-permeable membrane. The semi-permeable membrane may be made from a material that has a greater permeability to hydrogen gas than it does to oxygen gas. In this regard, as the mixture of hydrogen and oxygen gases are directed using a series of tubes and valves towards the two gas collection bells nested within one another, only hydrogen gas would be able to effectively pass through the membrane covering the innermost gas collection bell. As such, the hydrogen gas would become concentrated in the innermost gas collection bell, while the oxygen gas would become concentrated in the outermost gas collection bell. In this manner, it is believed that the hydrogen gas could be isolated and collected separately from the oxygen gas.

[0105] The methodology ends at block **2124** and may be ended after a predetermined time interval and/in response to a determination that a desired amount of hydrogen production has been achieved.

FIG. **25**

[0106] FIG. **25** illustrates a high level exemplary methodology **2500** of carrying out the combustion of a liquid. The methodology begins at block **2510**. At block **2510** an RF system is provided that is capable of generating an RF signal. The RF system may include an RF generator, transmitter and transmission head and be of the type described above such that it is capable of generating an ignitable gas from sea water in an open container proximate to the transmission head. At block **2520** a liquid is provided that includes an effective amount of at least one ion dissolved in the liquid for generation of an ignitable gas by the RF signal. At block **2530** the RF signal is transmitted such that it interacts with at least some of the liquid. At block **2540** the ignitable gas generated from the liquid by the RF signal is ignited. At block **2550** the methodology ends and may be ended after a predetermined time interval and/in response to a determination that a portion of the liquid has been combusted.

FIG. **26**

[0107] FIG. **26** illustrates a high level exemplary methodology **2600** of carrying out the combustion of a liquid. The methodology begins at block **2610**. At block **2610** an RF system is provided that is capable of generating an RF signal. The RF system may include an RF generator, transmitter, and transmission head and be of the type described above such that it is capable of generating an ignitable gas from sea water in an open container proximate to the transmission head. At block **2620** a liquid is provided that includes an effective amount of at least one ion dissolved in the liquid for generation of an ignitable gas by the RF signal. At block **2630** the RF signal is transmitted and at block **2640** a portion of the liquid is combusted.

[0108] Additional methods are contemplated using the systems described herein where a frequency for operation of the RF signal may be selected such that the frequency is the same as, or overlaps (either partially or completely)—or has harmonics that are the same as or overlaps—specific RF frequencies that are capable of stimulating or exciting any of the various energy levels of various ions, e.g., any of the various metal species that comprise the salts that are dissolved in the salt water solutions. One having ordinary skill in the art will understand how to determine and to measure RF frequencies that stimulate or excite various energy levels for various metal species. In this regard and based on empirical testing, we believe that 13.56 MHz stimulates and/or excites Na ions better than any other ions herein so tested. As such, it is believed that useful embodiments of the methods described herein may therefore also include (i) selecting an RF signal having a preferred frequency, (ii) selecting a metal salt comprising a metal species capable of being stimulated or excited by the preferred frequency selected (or a harmonic thereof), (iii) transmitting the RF signal having the preferred frequency through or to an aqueous solution of the metal salt for a sufficient time in order to stimulate or excite the metal species present in the aqueous solution to generate heat. Alternatively, methods may also include (i) selecting a salt comprising a preferred metal species, (ii) selecting an RF signal having a frequency (or a harmonic thereof) capable of stimulating or exciting the preferred metal species, (iii) transmitting the RF signal having the frequency to or through an aqueous solution of the metal salt comprising the preferred metal species for a sufficient time to generate heat.

[0109] Additional methods are contemplated using the systems described herein where the RF signal may be used to process clays and soils to heat and sterilize the clays and soils, to directly generate hydrogen from the clays and soils, and for remediation of the clays or soils by removing or extracting organic contaminants and wastes. It is contemplated, as above, that a frequency for operation of the RF signal may be selected such that the frequency (or a harmonic thereof) is the same as or overlaps with (either partially or completely) specific RF frequencies capable of stimulating or exciting any of the various energy levels of any of the various metal species comprising metal salts or metal compounds that are dissolved or distributed within the soils. Since soils often contain moisture or the metal species present in the soils and clays have water molecules coordinated to them, it is therefore believed that the systems and methods described herein could be used to heat and process such metal-containing soils. As such, we believe the RF signal could be used (in any of the various manners herein described for treatment of salt water solutions) to produced heat and/or steam and/or hydrogen and oxygen free radicals in-situ within various soils, and in particular in clays and clay containing soils. The heat and/or the steam and/or the hydrogen and oxygen free radical produced from the water molecules present in the soil would treat the surrounding soil, in particular the heat and/or the free radicals generated would perhaps sterilize the soil, killing any animal, vegetable or microbial life that may also be present. It is further contemplated that steam produced in-situ in this manner may also be used to volatize and extract any hydrocarbon pollutants that may also be present in the soils and clays. As such, it is contemplated that soils of contaminated commercial residential and industrial sites, hazardous waste dump sites, gas stations, etc. could be remediated using the systems and methods described herein. One skilled in the art will understand how the RF systems and methods described herein could be coupled with known extraction and remediation processes and methods for in-situ treatment of contaminated soils. Exemplary hydrocarbon contaminants that could be extracted or removed would include but are not limited to organic solvents, oil and oil byproducts, insecticides, and polychlorinated biphenyls. Similarly, it is contemplated that clathrates, zeolites, and other materials containing or having various metal species adsorbed to their surfaces or in there structures and containing either moisture or water molecules coordinated to the metal species present may be processed and heated in similar manners as has been described herein for soils and clays.

[0110] In accordance with the systems and methods of the present invention previously described, further embodiments are contemplated of an RF system for selective disinfection of surfaces and materials is provided. The system includes an RF transmitter having an RF generator and a transmission head, and an RF receiver having a resonant circuit and a reception head. When the transmission and reception heads are arranged proximate to and on either side of a surface or material and an RF signal is transmitted from the transmission head, through the surface or material, to the reception head, at least a portion of the surface or material is disinfected without direct contact of the heads to the surface or material. It is contemplated, as above, that a frequency for operation of the RF signal may be selected such that the frequency (or harmonic thereof) is the same as or overlaps with (either partially or completely) specific RF frequencies that are capable of stimulating or exciting any of the various energy levels of any of the various metal species or metal salts or metal compounds that may, for example, be present within various targeted microbes, bacteria, or viruses. Since environments where microbes, bacteria, and viruses are found also often contain moisture, we therefore believe that the systems and methods described herein could be used to disinfect surfaces and materials through selectively heating and destroying various targeted microbes, bacteria, and viruses that are present on the surfaces or materials to be disinfected. The RF signal would be applied for a sufficient time to locally heat and destroy any targeted microbes, bacteria, and viruses that contain metals (metals that are either coordinated by water molecules or in an environment containing moisture) that are stimulated or excited by the RF signal having the particular frequency so selected.

[0111] In accordance with the systems and methods of the present invention previously described, further embodiments are contemplated of an RF system for affecting a change in the germination and growth of plant life is provided. The system includes an RF transmitter having an RF generator and a transmission head, and an RF receiver having a resonant circuit and a reception head. When the transmission and reception heads are arranged proximate to and on either side of a seed or a plant and an RF signal is transmitted from the transmission head, through the seed or plant, to the reception head, at least a portion of the seed or plant is processed without direct contact of the heads to the seed or plant. For example, a seed may be placed in a brackish environment or a plant may be watered with brine solution and natural biological processes such as osmotic pumping mechanisms may be taken advantage of in order to create a seed or plant having an internal environment with an increased salt concentration. We believe that any of the systems or methods described herein may be used to then expose the so prepared seed or plant to an RF signal, wherein the RF signal would affect a change in the rate of germination of the seed or affect a change in the rate of growth of the plant. We believed that that a frequency for operation of the RF signal may be selected such that the frequency (or harmonic thereof) is the same as or overlaps with (either partially or completely) specific RF frequencies that are capable of either increasing or decreasing the rates of seed germination and plant growth in order to affect such a change in the germination and growth of plant life.

[0112] In accordance with the systems and methods of the present invention previously described, further embodiments contemplating RF systems and methods for processing a fluid are provided. Processing a fluid includes but is not limited to heating and/or combusting the fluid. Fluids can be processed whether or not they contain any of the useful salts or ions (either cations or anions) herein described. An exemplary fluid in this regard includes but is not limited to water that is extracted from oil wells and that is contaminated with oil residues and/or other hydrocarbon contaminants. Methods for processing (including heating and/or combusting) a fluid involve using any of the systems previously described and (i) providing a fluid to be processed (including heating and/or combusting the fluid), (ii) adding an effective amount of salt to the fluid (e.g., by adding solid salt or by adding a salt solution), and (iii) passing RF through the fluid containing an effective amount of salt to process the fluid. In general, useful systems may include an RF transmitter having an RF generator and a transmission head, and an RF receiver having a resonant circuit and a reception head. When the transmission and reception heads are arranged proximate to and on either side of the fluid having an effective amount of salt added to it an RF signal is transmitted from the transmission head, through the fluid containing the salt, to the reception head, and at least a portion of the fluid is processed. Processing in this regard may include heating the fluid and/or combusting the fluid and in such situations salt is added to enhance heating of the fluid.

Salt Water, Salt Water Solutions, and Salt Water Mixtures

[0113] Ordinary and naturally occurring seawater may be used. Generally, a salt which is useful as the salt water or in the solution containing salt water or in the salt water mixtures employed in these systems and methods disclosed herein include any salt which has solubility in water. For example, NaCl is a useful salt because NaCl is very soluble in water. Other useful salts may include salts that have as their cation any element in cationic form, which may selected from the group consisting of Li^+, Na^+, K^+, Rb+, Cs^+, Be^{2+}, Mg^{2+}, Ca^{2+}, Ba^{2+}, Sr^{2+}, Mn^{2+}, Fe^{2+}, Fe^{3+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, Ag^+, Au^+, B^{3+}, Al^{3+}, Ga^{3+}, In^{3+} and that have as the anion any element in anionic form that is selected from the group consisting of Cl^-, Br^-, I^-, borate, citrate, nitrate, phosphate, sulfate, carbonate, and hydroxide. The salt used in the systems and methods disclosed herein can be used as either a pure salt, the salt made from one type of cation and one type of anion that are those cations and anions listed above; or it can be a salt mixture, made from more than one type salt, made from one or more types of cations and/or one or more types of anions that are those cations and anions listed above. Again, ordinary and naturally occurring seawater may be used.

[0114] Another useful salt water (or salt water component of either solutions containing salt water or salt water mixtures) for use in the systems and methods disclosed herein is seawater. This includes all types of seawater, including water taken from any of the oceans or other naturally salty bodies of water found on the earth. Using seawater as disclosed herein includes using seawater in its natural occurring form, that is, seawater which is taken from the ocean and used directly without any further processing or purification.

[0115] Another useful salt water or salt water solution for use in the systems and methods disclosed herein is brine water. Brine water may be water extracted from the ground (ground water) and includes water that is taken from water wells and oil wells. Using brine water as disclosed herein includes using brine water that has been further processed or treated (for example, by addition of salt, e.g., adding solid salt or a salt solution) or that is in its naturally occurring form and used directly without any further processing or purification.

[0116] OCEANIC brand Natural Sea Salt Mix may be used to approximate naturally occurring seawater having an effective amount of salt and used as the salt water or salt water component of solutions containing salt water and salt water mixtures employed in the systems and methods discussed and shown herein. Such approximations of naturally occurring seawater may have a specific gravity of about 1.02 g/cm^3 to 1.03 g/cm^3, e.g., between about 1.020-1.024 or about 28-32 PPT, as read off of a hydrometer. A mixture of the above-identified sea salt mix having a specific gravity of about 1.026 g/cm^3 (as measured with a refractometer) was used in exemplary systems and methods. In the alternative, it is believed that actual seawater may be used in the systems and methods discussed and shown herein. The precise amount of salt in salt water or in the salt water component of the solutions containing salt water and salt water mixtures used and contemplated herein may vary from specific application to specific application.

[0117] In order to form both hydrogen and oxygen gas, salts capable of forming salt water mixtures that are useful for use in the electrolysis systems and electrolysis methods disclosed herein, should be water soluble salts and also should have a cation and an anion selected such that the cation has a lower standard electrode potential than H^+ and the anion has a greater standard electrode potential than OH^-. For example, the following cations have lower standard electrode potential than H^+ and are therefore suitable for use as electrolyte cations: Li^+, Rb^+, K^+, Cs^+, Ba^{2+}, Sr^{2+}, Ca^{2+}, Na^+, and Mg^{2+}. For example, a useful anion would be SO_4^{2-}, because it has a greater standard electrode potential than OH^- and is very difficult to oxidize. It is contemplated that Na_2SO_4 would be a useful salt for use with the electrolysis systems and methods disclosed here within because it is a water soluble salt that is composed of a cation (Na^+) that has a lower standard electrode potential than H^+ and an anion (SO_4^{2-}) that has a greater standard electrode potential than OH^-.

Additive

[0118] As previously indicated, as used herein an additive may be an organic, organometallic, or inorganic chemical compound having solubility, miscibility, or compatibility with salt water and solutions containing salt water and salt water mixtures (including seawater or solutions containing salt water and optionally containing at least one secondary fuel) and that is capable of altering the response of the salt water, various solutions containing salt water, and salt water mixtures in response to stimulation by RF energy. Both molecular and polymeric species are contemplated as being useful additives. It is further believe that useful amounts of additive include solutions containing salt water where the additive is present as at least one minor component in the solution containing salt water. Embodiments contemplated in this regard would include solutions containing salt water and having from about 0.001 to about 10.0 weight % additive, and more preferably from about 0.001 to about 1.0 weight % additive, and even more preferably from about 0.001 to about 0.1 weight % additive.

[0119] It is contemplated that a salt water solution or salt water mixture containing an additive will respond differently to RF stimulation versus comparable salt water solution or salt water mixture that does not contain any additive. We believe that the response of a salt water solution or salt water mixture to RF energy may be altered in a variety of ways. For example, an alteration in RF response that an additive may have may include but is not limited to increasing or decreasing the rate at which a solution or mixture containing the additive either heats, combusts, or both upon exposure to a fixed amount or flux of RF energy; exhibiting a desired temperature change or level of combustion of a salt water solution containing an additive with exposure to a larger or a smaller amount of RF energy; and decreasing the surface tension of a salt water solution containing an additive such that combustion of the salt water solution or mixture occurs upon application of an RF field without any need for externally perturbing the surface of the salt water solution. Surfactants, including soaps and detergents, are embodiments of useful additives in this regard since they are known to lower the surface tension of aqueous solutions. Furthermore, we believe that water soluble organic compounds that can lower the heat capacity of an aqueous solution or that can change the freezing point of water or that can form azeotropic mixtures with water would also be useful additives in this regard.

Secondary Fuels

[0120] As previously indicated, as used herein a secondary fuel may be any combustible organic compound that has solubility, miscibility, or compatibility with salt water or various solutions containing salt water or salt water mixtures (including seawater, salt water or solutions containing salt water that optionally contain at least one additive). It is believe that a useful amount of secondary fuel includes solutions containing salt water were the secondary fuel is present as the minor component. Alternatively, it is also believe that a useful amount of secondary fuel includes solutions containing salt or salt water were the secondary fuel is present as the major component. In this regard, embodiments are contemplated of salt water solutions containing from about 0.01 to about 99.99 weight % of at least one alternative fuel, and preferably from about 1.0 to about 99.0 weight % of at least one alternative fuel, and more preferably from about 10 to about 90 weight % of at least one alternative fuel, and even more preferably from about 30 to about 70 weight % of at least one alternative fuel, and even more preferably from about 40 to about 60 weight % of at least one alternative fuel.

[0121] It is contemplated that exposure to RF energy of a salt water solution containing at least one secondary fuel, wherein the secondary fuel is the minor constituent, may result in an enhancement or in a boost in performance in terms of the combustibility of the salt water solution versus the results obtained by a comparable salt water solution that does not contain any secondary fuel. Alternatively, it is also contemplated that exposure to RF energy of a salt water solution containing at least one secondary fuel, where the secondary fuel is the major constituent of the mix, allows RF energy to be used to combust the secondary fuel even though the secondary fuel itself may be RF inert. Without intending to be bound by theory, we believe that the secondary fuel may be useful as either the minor or the major component in a salt water solution because the salt water component of the salt water solution is stimulated by the RF signal and absorbs energy. As such, absorption of RF energy by the salt water component causes the temperature of the salt water solution to increase to the point where secondary fuel present in any amount volatilizes and becomes more capability of combusting in the presence of a spark, flame, or any other incendiary source. In this regard, methanol, ethanol, and iso-propanol are useful as secondary fuels because they are combustible organic solvents and are soluble with or have chemical compatibility with water. Furthermore, we believe that many additional organic solvents and compounds, which may have both volatility and solubility or miscibility with aqueous solutions, would also be useful as secondary fuels in this regard. For example, we contemplate that n-propanol, acetone, formaldehyde, acetic acid, and formic acid may also be useful secondary fuels.

RF Absorption Enhancers

[0122] Salt water, solutions containing salt water, and salt water mixtures may be processed using RF as-is. In the alternative, it is also believed that RF absorption enhancers may be added to the salt water, solutions containing salt water, and salt water mixtures prior to processing with RF to enhance the effects of the RF energy on the salt water, e.g., enhanced heating, enhanced, combustion, enhanced desalination, etc. The RF absorption enhancers may be particles made from RF absorbing materials that absorb one or more frequencies of an RF electromagnetic signal substantially more than other materials. This may permit the RF signal to heat salt water (or any solution containing salt water or salt water mixture) containing RF absorbing enhancers substantially more than it would salt water (or salt water solution or salt water mixture) that does not contain additional RF absorption enhancers.

[0123] Exemplary RF absorption enhancers include particles of electrically conductive material, such as silver, gold, copper, magnesium, iron, any of the other metals, and/or magnetic particles, or various combinations and permutations of gold, iron, any of the other metals, and/or magnetic particles. Examples of other RF absorption enhancers include: metal tubules (such as silver or gold nanotubes or silver or gold microtubes, which may be water-soluble), particles made of piezoelectric crystal (natural or synthetic), particles made of synthetic materials, particles made of biologic materials, robotic particles, particles made of man made applied materials, like organically modified silica (ORMOSIL) nanoparticles. Examples of yet other RF absorption enhancers that may be useful include RF absorbing carbon molecules and compounds: fullerenes (any of a class of closed hollow aromatic carbon compounds that are made up of twelve pentagonal and differing numbers of hexagonal faces), carbon nanotubes, other molecules or compounds having one or more graphene layers, and other RF-absorbing carbon molecules and compounds e.g., C60 (also known as a "buckyball" or a "buckminsterfullerene"), C70, C76, C84, buckytubes (single-walled carbon nanotubes, SWNTs), multi-walled carbon nanotubes (MWNTs), and other nano-sized or micro-sized carbon cage molecules and compounds. Such carbon-based particles may be in water-soluble form. Such carbon-based particles may have metal atoms (e.g., nickel atoms) integral therewith, which may affect their ability to absorb RF energy and heat in response thereto. Any of the foregoing (and subsequently listed) particles may be sized as so-called "nanoparticles" (microscopic particles whose size is measured in nanometers, e.g., 1-1000 nm) or sized as so-called "microparticles" (microscopic particles whose size is measured in micrometers, e.g., 1-1000 μm).

[0124] Additionally, RF absorbing carbon molecules and compounds may be fabricated as RF absorption enhancers to be particles with non-linear I-V characteristics (rectifying characteristics) and/or capacitance. Such non-linear I-V characteristics may result from, for example, nanotubes with a portion doped (e.g., by modulation doping) with a material giving n-type semiconducting properties adjacent a portion doped with p-type semiconducting properties to form a nanotube having an integral rectifying p-n junction. In the alternative, nanotubes can be fabricated with an integral Schottky barrier. In either case, it may be helpful to use nanotubes having at least two conducting regions with a rectifying region therebetween. Accordingly, rectifying circuits for RF absorbing particles for RF absorption enhancers may be fabricated from RF absorbing carbon molecules and compounds having non-linear I-V characteristics.

[0125] Any of the RF absorption enhancers described herein may be used alone or in virtually any combination of and/or permutation of any of the particle or particles described herein. For example, it may be beneficial to use a plurality of different RF absorbing particles described herein for purposes of tuning the reaction kinetics of the various methods herein described. Accordingly, virtually any combination or permutation of RF absorption enhancers may be used in virtually any combination of and/or permutation of any RF absorbing particle or particles described herein to create RF absorption enhancers for use in accordance with the teachings herein.

[0126] Of the RF absorption enhancers mentioned herein, some may be suitable for a 13.56 MHz RF signal, e.g., silver nanoparticles, gold nanoparticles, copper nanoparticles, magnesium nanoparticles, aqueous solutions of any of the metal sulfates mentioned herein, and RF absorbing carbon molecules and compounds. RF absorption enhancers using these RF absorbing particles are also expected to be effective at slightly higher frequencies, such as those having a frequency on the order of the second or third harmonics of 13.56 MHz.

RF Signal

[0127] The RF signals may have a frequency corresponding to a selected parameter of an RF enhancer, e.g., 13.56 MHz, 27.12 MHz, 915 MHz, 1.2 GHz. Several RF frequencies have been allocated for industrial, scientific, and medical (ISM) equipment, e.g.: 6.78 MHz±15.0 kHz; 13.56 MHz±7.0 kHz; 27.12 MHz±163.0 kHz; 40.68 MHz±20.0 kHz; 915 MHz±13.0 MHz; 2450 MHz±50.0 MHz. See Part 18 of Title 47 of the Code of Federal Regulations. These and other frequencies of the same orders of magnitude may be used in virtually any of the systems and methods discussed herein, depending on which RF absorbing particles are used. For example, RF signals having a fundamental frequency of about 700 MHz or less might be suitable for many of the systems and methods described herein. RF signals having a fundamental frequency in the high frequency (HF) range (3-30 MHz) of the RF range might be suitable for many of the systems and methods described herein. Similarly, RF signals having a fundamental frequency in the very high frequency (VHF) range (30-300 MHz) of the RF range might also be suitable for many of the systems and methods described herein. Of course, RF signals at any fundamental frequency may also have harmonic components that are multiples of the fundamental frequency of frequencies. Also, RF signals at any fundamental frequencies or periodic multiples of such fundamental frequencies that are harmonics of a fundamental frequency may be selected such that the frequency is the same as or has overlap with (either partially or completely) specific RF frequencies capable of stimulating or exciting any of the various electron energy levels of any of the various metal species that comprise the salts that are dissolved in the salt water solutions. For example, based on empirical testing we believe that an RF signal with a frequency of 13.56 MHz stimulates and/or excites Na ions better than any other ions herein so tested.

[0128] Additionally, in any of the embodiments discussed herein, the RF signal used may be a pulsed, modulated FM RF signal, or a pulse fixed frequency signal. A pulsed signal may permit a relatively higher peak-power level (e.g., a single "burst" pulse at 1000 Watts or more, or a 1000 Watt signal having a duty cycle of about 10% to about 25%) and may create higher local temperatures at RF absorption enhancer particles. Such pulsed signals may have any of various characteristics. For example, the RF pulse may be a square wave, or may be a sine wave, or may have a sharp rise time with an extended ringing effect at base line, or may have a slow rise time and a fast decay, etc. Pulsed RF signals (and other shaped RF signals) may produce very localized temperatures that are higher for a length of time on the order of about a millisecond or longer. For example, a short 5 kilowatt RF pulse of less than a second, e.g., on the order of microseconds (e.g., 3-4 microseconds) may be sufficient to raise the temperature of the mixture sufficiently to achieve the desired effect, e.g., combustion of the salt water, desalination, heating, creation of hydrogen gas, etc.

[0129] As discussed herein, the RF energy directed toward the salt water (or any solution containing salt water, or salt water mixture) may be RF energy having a very high field strength and may also be coupled through the portion of the reaction chamber with coupling heads having a very high Q (e.g., a Q on the order of 250 or more). A pulsed RF signal with a relatively higher power may be effective to quickly heat the salt water, etc., such as a pulse of BF or VHF RF energy (e.g., 27.12 MHz).

Rate of Combustion

[0130] Salt water combusts relatively quickly in a test tube using a 600 Watt 13.56 MHz RF signal. For example, sea water—natural or artificial—combusts in a test tube on the order of about 1 ml per minute initially and later combusts on the order of about 1 ml per every 30 seconds as a substantial amount of water has been combusted from the test tube. In some cases, less salt permits better combustion than more salt. For example, a mixture of 99.5% ethanol and 0.5% salt solution combusts much better (faster) than a 50/50 mixture of ethanol and salt solution (see examples below). As another example, sea water from the Gulf of Mexico combusted at about 2-3 ml per 90 second period at about 1000 watts, using either a 10 ml or 100 ml test tube, with the upper surface of the sea water in the RF field.

COMPARATIVE EXAMPLES

Series 1: Experiments with Ocean Water

[0131] It was previously demonstrated that salt water made from sea salt mix will combust using the RF system described in the '530 application. It has been confirmed that ocean water will combust using the ENI RF generator using the coupling circuit of FIGS. 46-49 of U.S. Provisional Patent Application Ser. No. 60/915,345, filed on May 1, 2007, and entitled Field Generator For Targeted Cell Ablation (Attorney Docket 30274/04036) ("the '345 application"), the entire disclosure of which is hereby incorporated by reference in its entirety, with a 6" silver coated circular copper Tx head (single plate) and a 9.5" silver coated square copper Rx head (single plate).

[0132] It is believed that the RF field that combusts salt water is substantially the same as the field discussed in the '345 application (see FIG. **53**—end of that application). (It is also believed that Ocean water will combust with the other head configurations discussed in the '530 application, as well.)

[0133] With respect to the combustion of ocean water, water from the Gulf of Mexico having the following characteristics was capable of being rapidly combusted with the above-described RF system (a 10 ml sample was analyzed prior to any combustion):

Parameter	Date of Analysis	Results	Units	Reporting Limit	Method
Bromide	May 10, 2007	57.0	mg/l	0.5	300.0
Calcium	May 11, 2007	970.0	mg/l	0.05	6010
Chloride	May 10, 2007	18552.0	mg/l	1.0	300.0
Fluoride	May 10, 2007	BRL	mg/l	0.1	300.0
Magnesium	May 11, 2007	1600.0	mg/l	0.5	6010
pH	May 9, 2007	8.02	s.u.@23.8 C.	0.01	EPA 150.1
Potassium	May 11, 2007	770.0	mg/l	0.1	6010
Sodium	May 16, 2007	12000.0	mg/l	1.0	6010
Sulfate	May 10, 2007	2633.0	mg/l	1.0	300.0

[0134] In this example, combusted ocean water differed from uncombusted ocean water in the concentration of most of these components increases, while the concentration of calcium decreases. Two 10 ml samples of the above water from the Gulf of Mexico were combusted down to 5 ml each and combined, and the resulting 10 ml of combusted ocean water was analyzed to reveal the following:

Bromide	May 10, 2007	57.0	mg/l	0.5	300.0
Calcium	May 11, 2007	730.0	mg/l	0.05	6010
Chloride	May 10, 2007	20316.0	mg/l	1.0	300.0
Fluoride	May 10, 2007	BRL	mg/l	0.1	300.0
Magnosium	May 11, 2007	1900.0	mg/l	0.5	6010
pH	May 9, 2007	8.55	s.u.@22.8 C.	0.01	EPA 150.1
Potassium	May 11, 2007	880.0	mg/l	0.1	6010
Sodium	May 16, 2007	17000.0	mg/l	1.0	6010
Sulfate	May 10, 2007	3036.0	mg/l	1.0	300.0

[0135] A white residue forms on the inside of the test tube after combustion of salt water. The calcium may be part of that residue.

[0136] Using the above-described RF system, salt water will combust, as will solutions of HCl and NaCl. Distilled water will boil in the RF field, but will not combust. Adding additional sea salt mix (e.g., OCEANIC brand Natural Sea Salt Mix) to ocean water causes the rate of combustion to increase. Adding sea salt mix sufficient to approximately triple the sodium of ocean water causes a dramatic increase in the rate of combustion of the resulting salt water mixture. Thus, the methods herein may be modified by including the additional step of adding additional ions to the sea water prior to combustion.

[0137] Salt water (ocean water and/or salt water made from OCEANIC brand Natural Sea Salt Mix) will begin to combust in the above-described RF system at RF wattages of about 250 Watts and salt water will continue to combust at lower wattages, e.g., about 200 Watts, after igniting. Salt water may begin to combust spontaneously at higher temperatures, or may require some sort of igniter (e.g., a drop of salt water dropped through the RF field, which combusts and ignites the other salt water in the field). Additionally, some sort of wick (e.g., a piece of paper towel) extending above the surface of the salt water in the field will greatly increase the tendency of salt water in the RF field to spontaneously ignite. Filling the test tube to the brim with salt water and then adding a couple more drops of salt water facilitates ignition.

[0138] Using a setup with about 5.5" spacing between Tx plate and Rx plate, and the test tube being about 2" from the Tx plate at about the top of the Tx plate, and applying RF to the salt water, the products produced from exposure of salt water to RF energy burn. The temperature of the burning products of salt water exposed to RF energy has been measured as high as about 1700° C. using a FLIR Systems ThermaCAM P65 thermometer with ThermaCAM Quick View V2.0 Software, which measures temperatures up to 1700° C. (it is believed that the salt water is combusting at a higher temperature). Surprisingly, the temperature of the salt water in the test tube remains relatively low (e.g., less than 45° C.) while the salt water is combusting.

[0139] Without intending to be bound by this description, it is believed that the special RF field generated by the above-described RF system causes hydrogen in salt water to separate from oxygen, and then the hydrogen is burned in the presence of the released oxygen and the oxygen in the surrounding air.

[0140] Heat from RF-induced combustion of salt water may be used in any of the traditional methods of gathering and using heat, e.g., a heat exchanger, a Stirling Engine, a turbine system, etc.

[0141] Additionally, multiple Tx and Rx heads may be used at one or more frequencies.

Series 2: Experiments with Salt Water and Solutions with Additives and Secondary Fuels

[0142] For all the Series 2 examples described below, a circuit implementation of FIG. **16** was used to transmit the RF signal through the exemplary solutions to yield the various results. Unless otherwise indicated, for all examples a 13.56 MHz RF signal from an ENI OEM-12B RF generator having a variable power output of up to about 1000 Watts was applied for thirty seconds to the reaction chamber, which in these instances consisted of a glass test tube (in which the various exemplary solutions were placed) connected to a support arm that was positioned such that the test tube was suspended between the transmission head (one plate) and the reception head (three plates). Unless otherwise indicated, the salt water solutions used in carrying out the various examples included Gulf of Mexico salt water, Brine salt water extracted from an oil well (located in Erie, Pa.), and a 3.5 wt % stock solution of OCEANIC brand Natural Sea Salt Mix having a specific gravity of about 1.026 g/cm^3. For all examples containing ethanol, denatured, Apple Products® brand ethanol was used.

Salt Water

[0143] A first 100 mL sample containing salt water was placed in a test tube and the test tube was then attached to a support arm and positioned between the transmission head and receiver head of the RF apparatus (described above). The temperature of the salt water was measured using a fiber optic thermometer. A 13.56 MHz RF signal at about 300 Watts was then applied for about 30 seconds, after which the temperature was again measured using a fiber optic thermometer. Starting temperature=24.0° C.; Ending temperature=25.9° C.

[0144] A second 100 mL sample of salt water was placed in a test tube and the test tube was then attached to a support arm and positioned between the transmission head and receiver head of the RF apparatus (described above). The temperature of the salt water was measured using a fiber optic thermometer. A 13.56 MHz RF signal at about 600 Watts was then applied and, as soon as the RF signal was applied, combustion of the salt water was initiated by momentarily placing an ordinary steel screwdriver in contact with the lip of the test tube. The screw driver was removed and the RF signal was left on for about 30 seconds as combustion of the salt water continued. After about 30 seconds, the RF signal was turned off and the combustion of the salt water ceased. The temperature of the salt water sample was then measured using a fiber optic thermometer at both the top part of the test tube and the bottom part of the test tube. Starting temperature=20.5° C.; Ending temperature (Top)=66.0° C.; Ending temperature (Bottom)=28.0° C.

[0145] A third 100 mL sample of salt water was placed in a test tube and the test tube was then attached to a support arm and positioned between the transmission head and receiver head of the RF apparatus (described above). However, the salt water used here contained 1 mL of stock salt water diluted to 100 mL with distilled water to give a 0.0035% salt water solution. A 13.56 MHz RF signal at about 600 Watts was then applied for about 30 seconds, after which the temperature was again measured using a fiber optic thermometer. Unlike the second sample of salt water, combustion of this third sample of salt water could not be initiated by placing an ordinary steel screwdriver in contact with the lip of the test tube. Starting temperature=26.6° C.; Ending temperature=75.5° C.

Salt Water+Carbonate and/or CO_2 (as the “Additive”)

[0146] Carbon dioxide may be useful as an additive, as may other additives that produce carbon dioxide. Photographs **9-11** of the incorporated material show the combustion of ground water—here a sample of brine water collected from an oil well (located in Erie, Pa.), while photograph **12** of the incorporated material shows the combustion of a sample of brine water obtained from the Gulf of Mexico. We have observed that the brine water obtained from the Gulf of Mexico combusts in a less sporadic manner than brine water collected from the oil well located in Erie, Pa. Without intending to be bound by theory, we believe high levels of carbonate salts present in the brine water collected from the oil well located in Erie, Pa., that is not present in the brine water collected from the Gulf of Mexico, effects the combustibility of the brine water collected from the oil well located in Erie, Pa. We further believe that, as the brine water collected from the oil well located in Erie, Pa. combusts carbonate salts that are present release carbon dioxide into the sample which acts to suppress or limit further combustion of the brine water as the RF signal is applied. Therefore, additional embodiments are contemplate wherein additives capable of inhibiting combustion or that are combustion suppressants may be added to any of the various salt water solutions herein disclosed in order to control or hinder the rate of salt water combustion or limit the amount of overall combustion.

Salt Water+Surfactant (as the “Additive”)

[0147] A 100 mL sample of salt water that also contained 1 metric drop (about 0.05 mL) of an ordinary hand soap (Liquid Nature Antibacterial Hand Soap) was placed in a test tube and the test tube was then attached to a support arm and positioned between the transmission head and receiver head of the RF apparatus (described above). A 13.56 MHz RF signal at about 600 Watts was then applied to the sample and as soon as the RF signal was applied, combustion of the salt water sample was initiated immediately. No external perturbation of the test tube (by a screwdriver, a drop of salt water, use of a wick or otherwise) was required. The RF signal was repeatedly switched on and off; each time the RF signal was switched on the salt water sample immediately began combusting, while each time the RF signal was switched off the salt water sample immediately ceased combusting.

Salt Water+Ethanol (as the “Secondary Fuel”)

[0148] A first 100 mL sample containing a mixture of 50 mL of ethanol and 50 mL of salt water was placed in a test tube and the test tube was then attached to a support arm and positioned between the transmission head and receiver head of the RF apparatus (described above). A 13.56 MHz RF signal at several hundred Watts was then applied to the sample and, as soon as the RF signal was applied, combustion of the sample was initiated by momentarily placing an ordinary steel screwdriver in contact with the lip of the test tube. Once the RF signal was turned off the combustion of the sample ceased. Surprisingly, in the absence of any applied RF signal combustion of the sample could not be initiated even when an open flame was used to attempt initiation of combustion.

[0149] A second 100 mL sample containing a mixture of 99.5 mL of ethanol and 0.5 mL of salt water was placed in a test tube and the test tube was then attached to a support arm and positioned between the transmission head and receiver head of the RF apparatus (described above). The temperature of the salt water was measured using a fiber optic thermometer. A 13.56 MHz RF signal at several hundred Watts was then applied for about 15 seconds, after which the temperature was again measured using a fiber optic thermometer. Starting temperature=26.6° C.; Ending temperature=62.0° C. This example shows that an effective amount of salt (e.g., solid salt or a salt solution) can be added to enhance heating of liquids.

[0150] A third 100 mL sample containing a mixture of 99.5 mL of ethanol and 0.5 mL of salt water was placed in a test tube and the test tube was then attached to a support arm and positioned between the transmission head and receiver head of the RF apparatus (described above). The temperature of the salt water was measured using a fiber optic thermometer. A 13.56 MHz RF signal at several hundred Watts was then applied and, as soon as the RF signal was applied, combustion of the sample was initiated by momentarily placing an ordinary steel screwdriver in contact with the lip of the test tube. Combustion of the sample was highly energetic and resulted in a very large flame as compared to RF combustion of a stock solution of salt water that did not contain any ethanol. The screw driver was removed and the RF signal was left on for 15 seconds as energetic combustion of the sample continued. Combustion was so energetic that some of the sample solution bubbled out of the test tube and onto the laboratory floor when it continued to combust. After about 15 seconds, the RF signal was turned off. However, combustion of the sample did not cease and the sample had to be extinguished using a fire extinguisher.

Control **1**: Distilled Water

[0151] A 100 mL sample containing distilled water was placed in a test tube and the test tube was then attached to a support arm and positioned between the transmission head and receiver head of the RF apparatus (described above). The temperature of the distilled water was measured using a fiber optic thermometer. A 13.56 MHz RF signal at about 300 Watts was then applied for about 30 seconds, after which the temperature was again measured using a fiber optic thermometer. Starting temperature=24.0° C.; Ending temperature=24.8° C.

Control **2**: Tap Water

[0152] A 100 mL sample containing ordinary tap water was placed in a test tube and the test tube was then attached to a support arm and positioned between the transmission head and receiver head of the RF apparatus (described above). The temperature of the ordinary tap water was measured using a fiber optic thermometer. A 13.56 MHz RF signal at about 300 Watts was then applied for about 30 seconds, after which the temperature was again measured using a fiber optic thermometer. Starting temperature=23.7° C.; Ending temperature=47.8° C.

Control **3**: 100% Ethanol

[0153] A 100 mL sample containing ethanol was placed in a test tube and the test tube was then attached to a support arm and positioned between the transmission head and receiver head of the RF apparatus (described above). The temperature of the ethanol was measured using a fiber optic thermometer. A 13.56 MHz RF signal at several hundred Watts was then applied for about 15 seconds, after which the temperature was again measured using a fiber optic thermometer. Starting temperature=25.0° C.; Ending temperature=30.0° C.

Dissociation of Water

[0154] In his earlier patent documents, inventor John Kanzius describes a process in which liquid water is dissociated into hydrogen and oxygen using an external RF field. For convenience, this process is referred to in the following paragraphs as the "electromagnetic dissociation of water." Normally, the electromagnetic radiation is in the form of an RF field generated between a transmission head and a reception head external to the water being dissociated. These earlier patent Kanzius documents also describe various systems and methods for electromagnetic dissociation of water with an external RF field generated by any of various RF systems which may be referred to as Kanzius RF transmitters or Kanzius RF generators.

[0155] In conventional electrolysis of water, a direct electrical current in passed through water between two electrodes immersed in the water. At the negatively charged electrode, the cathode, water is reduced according to the following half cell reaction:

$$2H_2O+2e^-\rightarrow H_2+2OH^-,$$

[0156] thereby liberating hydrogen gas. At the positively charged electrode, or anode, water is oxidized according to the following half cell reaction:

$$2H_2O\rightarrow O_2+4H^++4e^-,$$

[0157] thereby liberating oxygen gas.

[0158] In the Kanzius electromagnetic dissociation of water, it is believed that the same reactions are occurring, i.e., the dissociation of water into hydrogen and oxygen. However, in the electromagnetic dissociation of water, these reactions occur in response to applied electromagnetic radiation only. Thus, the electromagnetic dissociation of water differs from conventional electrolysis of water in that, in the electromagnetic dissociation of water, alternating electrical energy is applied without electrodes as opposed to applying direct electrical energy with electrodes. Similarly, the electromagnetic dissociation of water differs from other techniques for the direct application of electrical energy such as microwave heating in that, in the electromagnetic dissociation of water, the applied electrical energy causes water in the liquid phase to dissociate into its component parts without substantial heating. In contrast, microwave heating causes substantial heating and eventual evaporation, and is not known to cause direct dissociation of water. The electromagnetic dissociation of water also differs from known techniques for the thermal decomposition (thermolysis) of water which occur in the vapor phase at very high temperatures, with the energy needed for causing the reaction to occur being supplied by conventional heating (e.g., conduction or convection) as opposed to irradiation with electromagnetic radiation. Note, also, that in the Kanzius electromagnetic dissociation of water (as well as the other RF treatments described in this disclosure), no additional energy input such as applied audio signal or an external magnetic field need be applied, and indeed is not normally used.

Electrolytes

[0159] Electrolysis of pure water is impossible to carry out, as a practical matter, because the electrical conductivity of pure water is too low. Therefore an electrolyte, i.e., a substance capable of increasing the electrical conductivity of the water, is added. Ionizable salts are most commonly used for this purpose, the salt chosen normally having a cation with a lower standard electrode potential than H^+ and an anion with a higher standard electrode potential than OH^-.

[0160] In the electromagnetic dissociation of water, the same additives useful for facilitating conventional electrolysis of water can be included in the water being dissociated here. Thus, any additive capable of increasing the electrical conductivity of water can be added. Many of these additives are described above and in John Kanzius' prior patent documents. For example, this additive may be an ionizable salt having a cation selected from Groups I, II and III of the Periodic Table, as described above, more commonly Groups IA, IIA and IIIA. Essentially any anion capable of producing a soluble salt from such cations can be used, examples of which include chlorides, bromides, fluorides, other halides, nitrates, sulfates, acetates, carbonates, phosphates, carboxylates, sulfates, sulfonates, anions of other organic acids (e.g., acrylates, methacrylates, any other C_3-C_{12} organic acid, etc.) can be used. Specific examples include NaCl, NaBr, NaF, Na2SO4, NaNO3, Na2(PO4), Na-acetate etc.

[0161] Similarly, soluble metal salts and complexes of transition metals such as Cr, Mo, W, Cu, Co, Ni, Fe, Cd, etc. can also be used.

[0162] In addition to ionizable salts, other materials can be used as an electrolyte for fostering the electromagnetic dissociation of water, provided they are used in appropriate amounts. For example, as illustrated in the following working examples, aqueous colloidal dispersions of silver particles containing up to about 200 ppm silver readily combusted when irradiated with electromagnetic radiation in accordance with this invention, at least under the particular conditions tried in these examples. Similar colloidal dispersions containing more than about 200 ppm silver did not readily combust, which is believed due to the fact the silver in these colloids had settled out of dispersion as confirmed by visual inspection.

[0163] In addition to silver, it is believed that dispersions of other electrically conductive particles will also foster the electromagnetic dissociation of water in accordance with this invention-again, provided they are used in appropriate amounts. Examples of materials which can be used to form such electrically conductive particles include metals, e.g., Cu, Ag, Zn, semi-conductive materials such as TiO_2, ZrO_2, etc., carbon in various forms such as graphite, activated carbon black, etc., and so forth.

[0164] In addition to ionizable salts and electrically conductive colloids, still other materials can be used to increase the electrical conductivity of the water being electromagnetically dissociated according to this invention. For example, hydroxyethyl cellulose when added to water in an amount sufficient to produce a gel has been found to foster this electromagnetic dissociation, at least under the conditions used in the following working examples. In addition, various other oxygen and nitrogen-containing organic compounds such as oxalic acid, L-alanine, polyvinyl alcohol, ethyl acetate, ethylene glycol, glycerol, methanol, ethanol and propanol have also been found to foster this electromagnetic dissociation, provided they are used in suitable amounts under appropriate conditions.

[0165] In contrast, similar oxygen and nitrogen-containing organic compounds such as formic acid, acetic acid, lauric acid and salicylic acid were found not to foster this electromagnetic dissociation, at least under these same conditions. In the same way, many inorganic materials which do not readily ionize were also found not to foster this electromagnetic dissociation, at least under these same conditions. Examples include boric acid, MS Type 5 Zeolite, quartz, Na-montmorillonite, Ca-montmorillonite, CabOSil (untreated fumed sub-micronometer silica), Plaster of Paris, Kaolinite and Hydrotalcite. All of this merely indicates that, in order to function as an effective electrolyte, an additive needs to have a combination of properties including sufficient solubility in water, effect on the hydrogen/oxygen bond strength of water and other factors so that a not-insignificant increase in the electrical conductivity of water is achieved when the additive is added.

[0166] The concentration of the electrolyte included in the water to be electromagnetically dissociated according to this invention is not critical and essentially any amount can be used so long as the electrical conductivity of the solution or dispersion obtained is increased. For ionizable salts, typical amounts include enough to produce solutions which are at least about 0.2 molar up to saturation (for example, for sodium chloride, aqueous solutions of from about 0.3% to about 3% to about 30% combusted, as shown in the working examples, below). For colloidal dispersions of electrically conductive particles, amounts as low as 10 ppm can be used, while concentrations on the order of 20 to 800 ppm, or more especially 30 to 500 ppm, are more typical. For hydroxy ethyl cellulose and its hydroxy C_1-C_{12} alkyl analogs, essentially the same amounts can be used although amounts suitable for creating a gel are desirable. Greater concentrations of these additives can also be used, since these additives will still impart their electrical conductivity-increasing effect even if present in greater than saturation amounts.

[0167] In this connection, it is believed that this aspect of the invention in which water is electromagnetically dissociated will occur in any composition containing free water (i.e., water which is not chemically combined) even if substantial amounts of other ingredients are also present. So, for example, even moist powders can be electromagnetically dissociated, it is believed, since immersion of electrodes in the water being dissociated is unnecessary.

[0168] Similarly, it is also believed that concentration of the electrolyte to be included in the water to be electromagnetically dissociated in particular embodiments of this invention also depends on the intensity of the electromagnetic radiation used for the irradiation. As can be seen from the following working examples, some samples showed little or no evidence of dissociation (as reflected by the generation of a combustible gas) when irradiated under the standard conditions used in these examples (600 watts) but did show significant evidence of dissociation when irradiated at higher power levels, e.g., 800 watts or 1000 watts. Accordingly, it should be appreciated that the concentration of the electrolyte to be used in particular embodiments of the electromagnetic dissociation process of this invention depends not only the identity of the particular electrolyte employed but also on the particular irradiations conditions employed as well. These particulars—e.g., the power(s) (and wavelength(s)) of RF signal, the location in the field, and the amount of time of each RF signal transmission, and number of times, needed for effective electromagnetic dissociation of water, or to obtain a desired change result or change in properties of the substances disclosed herein, such as combustion or a shift in Raman spectrum or change in XRD data—can be easily determined by those of ordinary skill in the art by routine experimentation using the teachings in the earlier Kanzius patent documents, and the teachings herein.

[0169] Still other aqueous materials that can be electromagnetically dissociated in accordance with this aspect of the invention are the so-called "structured waters," i.e., microwave distilled deionized water, occasionally referred to as "Sedlmayr" water. See the following patent documents: U.S. Pat. No. 7,119,312 B2; WO 08020912 A2; US 2007/0095823 A1; US 2006/0289502 A1; US 2006/0006172 A1 and US 2006/0006171 A1.

Non-Aqueous Materials

[0170] In another application of this invention, "non-aqueous" materials are heated by the application of radio frequency electromagnetic radiation, or "RF energy," i.e., electromagnetic radiation having a lower frequency and a higher wavelength than microwave energy. In this context, "non-aqueous" means compositions in which the water content is so low that electromagnetic dissociation of water will not occur to any significant degree when the composition is irradiated with the RF energy.

[0171] Microwave radiation is commonly accepted to have a wave length of less than 1 meter (<1 M) to greater than 1 millimeter (>1 mm) and a corresponding frequency of 300 MHz (3×10^8 Hz) to 300 GHz (3×10^{11} Hz). Electromagnetic radiation having wave lengths shorter than 1 millimeter and corresponding frequencies greater than 300 GHz is generally opaque to the atmosphere until the atmosphere becomes transparent again in the infrared and optical window frequency ranges. Electromagnetic radiation having wave lengths longer than 1 meter and corresponding frequencies less than 300 MHz is generally regarded as "radio frequency" electromagnetic radiation or "RF energy." In accordance with this aspect of the invention, RF energy is used to heat a wide variety of different non-aqueous materials. RF signals having a fundamental frequency of about 250 MHz or less, 200 MHz or less, 150 MHz or less, 100 MHz or less, or even 50 MHz or less, are even more interesting. RF signals having a fundamental frequency in the high frequency (HF) range (3-30 MHz) of the RF range are especially interesting, since living tissue is virtually transparent to RF signals in this range. RF signals in the VHF range (30-300 MHz) are also of interest. Most commonly, the frequencies of such signals will not be below about 2 MHz, although frequencies in the medium frequency (MF) range (3 MHz-300 kHz) and even the low frequency (LF) range (300 kHz-30 kHz) are contemplated. It will be appreciated that these frequencies can also be used in connection with the electromagnetic dissociation of water as described above.

[0172] The various frequencies for the Kanzius RF transmitter and the associated methods can be characterized in terms of electron volts (eV): e.g., 3 MHz RF energy may be characterized as photons of 1.24×10^{-8} eV; 13.56 MHz RF energy may be characterized as photons of 5.6×10^{-8} eV; 30 MHz RF energy may be characterized as photons of 1.24×10^{-7} eV; and 300 MHz RF energy may be characterized as photons of 1.24×10^{-6} eV. Thus, the Kanzius RF transmitter may be further characterized as an RF transmitter (e.g., any of the configurations and/or characterizations herein such as those directly shown herein and those incorporated herein by reference) that generates low-energy photons (e.g., any one or more of the following: photons on the order of 10^{-8} eV, or less than 10^{-7} eV, or less than 10^{-6} eV, or less than 10^{-5} eV, or 5.6×10^{-8} eV, or $5\frac{1}{2}\times10^{-8}$ eV, or 1.24×10^{-8}–1.24×10^{-7} eV, or 1.24×10^{-8}–1.24×10^{-6} eV, or 1.24×10^{-8}–1.24×10^{-6} eV, or "about" any of these) that are capable of electromagnetic dissociation of water in various aqueous solutions and/or aqueous suspensions and/or aqueous pastes (and the other non-aqueous substances described herein) without requiring any additional energy input such as an external audio signal or an external magnetic field or current via electrodes. This is, indeed, remarkable since the H—OH bond energy is 5.2 eV.

[0173] One type of non-aqueous material that can be heated in this way are organic compounds (liquids, solids and gases), particularly those substituted with oxygen, nitrogen and/or sulfur. For example, various oxygen-containing organic liquids such as organic acids, esters, ketones, aldehydes, alcohols, glycols, carbonates, epoxides, furans, and derivatives, etc. can be heated in this way. Indeed, it is believed that any organic compound containing an —OH radical can be heated in this way. Similarly, various nitrogen containing organic liquids such as amides, amines, amino acids, as well as DNA, RNA and other "nucleic acid related compounds." In this context, "nucleic acid related compound" means ribonucleic acid compounds ("RNA") and deoxyribonucleic acid compounds ("DNA"), as well as the nucleotides, nucleosides and heterocyclic amine bases which are the nitrogen-containing precursors of or decomposition products of these RNA and DNA compounds, and mixtures of these compounds. See, for example, pages 1107-1149 of Organic Chemistry, 3rd Edition, by John MaMummy, Brooks/Cole Publishing Company, 1992; Lehninger, Biochemistry, Second Edition, Worth Publishers, Inc., pp. 729-747, © 1975; as well as King et al., Chemistry of Nucleic Acids, Version May 11, 2002, available on the web at http://www.med.unibs.it/~marchesi/nucleic.html.

[0174] Naturally occurring cellulosic materials such as plant leaves, bark, wood, etc. can also be heated. In addition, various carbonaceous materials such as coal, shale oil shale, tar sands, etc., can also be heated.

[0175] Another type of non-aqueous material that can be heated in accordance with this invention are inorganic solid materials. For example essentially any mineral, especially those containing combined oxygen, can be heated in this way. Examples include various clays, zeolites, silicas, aluminas and aluminosilicates, e.g., brucite, gibbsite, quartz, bauxite, etc. In addition to naturally occurring minerals, essentially any type of refined inorganic compound (i.e., a compound produced by purifying a naturally occurring mineral) as well as man made inorganic compounds, can be heated to a greater or lesser extent by the RF heating technology of this invention. For example, salts, oxides, bases, oxide complexes such as transition metal complexes, and so forth can be heated.

[0176] Another interesting phenomenon achieved by the present invention, at least in some embodiments, is a chemical and/or phase change in the non-aqueous material being irradiated as reflected by a change in its X-ray diffraction pattern. In some instances, such as in the case of hydroxyethyl cellulose both in solid form and contained in water, analysis by Raman Spectroscopy shows that an actual chemical change in the sample occurred, as reflected by a change in the characteristic peaks of its X-ray diffraction pattern. In other instances, only the relative intensities of certain peaks changed, thereby indicating a change in state (e.g., a change in crystal structure) but not a change in chemical identity. Due to the preliminary nature of this work, no conclusions can be drawn at this time other than noting that, at least in some instances, the application of RF energy in accordance with this aspect of the invention is sufficiently intense and powerful to effect a change in state of the non-aqueous materials being treated.

WORKING EXAMPLES

[0177] In order to more thoroughly describe this invention, the following working examples are provided:

Part 1: Liquid Samples

[0178] Liquid samples were suspended in a test tube of standard size and volume in the RF field positioned off-center as shown in the photographs (exact measurement of the position was not determined and there was some play in the exact location, however it was always closer to the coupler). Because the goal was to learn what would or would not combust, little attention has been placed on threshold RF power values, and RF power values were not recorded for a number of samples. Even if the RF values were recorded, those values generally reflect the standard instrument configuration at which the sample was tested and so all that can be ascertained from their values is that Sample X did or did not combust well at that power. In general, if certain samples failed to combust, the power was increased until the threshold for combustion was attained.

[0179] Unless otherwise noted, all flames self-extinguished upon turning off the RF field. Those that sustained a flame with the field off were snuffed out with a flame-proof mitt.

[0180] Unless otherwise noted, all sample solutions in water were prepared with the de-ionized (DI) water. Most solutions were tested at 3% concentration (as specified in the table). Dilute acid solutions were prepared using 8 drops of the pure acid in the test tube to which DI water was added to fill the tube. All samples were run in standard Pyrex test tubes (tt) unless otherwise noted. Teflon (PTFE) test tubes and Quartz (Q) test tubes were used as noted. The following results were obtained:

TABLE 1

Liquid Samples

Sample Description, holder	RF power	Observations [C = combusts]
GROUP 1 SALTS- THE SODIUM SERIES		
0.3% NaCl in DI water		C,
3% NaCl in DI water		C,
3% NaCl in DI water, PTFE tt		C,
3% NaCl in DI water, Q tt		C,
3% NaCl in DI water, pyrex tt	600 W	C,
30% NaCl in DI water		C
3% NaF in DI water	600 W	C
3% NaBr in DI water	600 W	C
3% NaNO3 in DI water	600 W	C, cracked tt
3% Na2SiO3 in DI water	600 W	C
3% NaH2PO4 in DI water	550 W	C, cracked tt
<3% NaCl in D2O, Q tt		C
3% Na-acetate, in DI water, pyrex tt		C
3% Na2(CO3)	600 W	Combusts and flame dies down after a short while due to possible release of CO2
3% Na-oxalate	600 W	C
3% Na2(SO4)	600 W	C
GROUP 1 SALTS		
3% KCl in DI water	600 W	C
2% KCl in DI water		C
1% KCl in DI water		C
3% KBr in DI water	600 W	C
3% KNO3 in DI water	600 W	C
3% K2(CO3) in DI water	600 W	Combusts and flame dies down after a short while due to possible release of CO2
3% LiCl in DI water	600 W	C
3% CsCl in DI water	600 W	C
GROUP 2 SALTS		
3% MgCl2 in DI water	600 W	C
3% Mg(NO3)2 in DI water	600 W	C
3% CaCl2 in DI water	600 W	C
3% BaCl2 in DI water	600 W	C
3% Ba(CO3)2	600 W	C

TABLE 1-continued

	Liquid Samples	
Sample Description, holder	RF power	Observations [C = combusts]
3% Sr(NO3)2 in DI water	600 W	C
3% SrCl2 in DI water	600 W	C
3% Ca(CO3)2 in DI water	800 W	C, but had trouble igniting at lower power
3% CaH4(PO4)2 in DI water	600 W	C
GROUP 3, 4, 5 (representatives)		
3% Al(NO3)3 in DI water	600 W	C
3% Pb(NO3)2 in DI water	600 W	C
3% Bi(NO3)3 in DI water	600 W	C
TRANSITION METAL SALTS		
3% Cr(NO3)3 in DI water	600 W	C
3% Fe(NO3)2 in DI water	600 W	C
3% Co(NO3)2 in DI water	600 W	C, cracked tt
3% Ni(NO3)2 in DI water	600 W	C, cracked tt
3% Ni(OH)2 in DI water	600 W	C, but sparingly soluble in water
3% CuSO4 in DI water	600 W	C
3% La(NO3)3 in DI water	600 W	C
3% Ag(NO3) in DI water	600 W	C
3% Cd(CH3COO)2 in DI water	600 W	C
INORGANIC ACIDS		
Dilute HCl	600 W	C
Dilute H2SO4	600 W	C
Dilute HNO3 (3 drops in DIw)	600 W	C
H3PO4 in pyrex tt	600 W	C
3% H2SiO3 in DI water	800 W	C
H2BO3 (Boric acid) solid		No Combustion
3% H2O2		C
STRUCTURED WATERS		
Yuri Kronn's Spring Water ("Rejuvenation Energy in Spring Water" commercially-available from Cedar Canyon Artesian Well, Forest Grove, OR)	n/a	No combustion
Yuri Kronn's Mineral Water ("Rejuvenation Energy In Trace Minerals" commercially available from Cedar Canyon Artesian Well, Forest Grove, OR)	n/a	C
Sedlmayr Machine3-A1 water (from a machine in SedlMayr U.S. Pat. No. 7,119,312)	n/a	C
Sedlmayr Machine3-new water (from a machine in SedlMayr U.S. Pat. No. 7,119,312)	n/a	C
Pure DI water (control)		No combustion
Tap water from PSU MRL	Up to 800	No combustion, but getting ready to boil
MISCELLANEOUS		
Quartz solution	600 W	No combustion
Carbon black in 3% NaCl	600 W	C, but sparked and crackled quite a bit
OTHER INORGANICS		
3% NaOH		C
<3% NH4OH	Up to 800	No combustion
3% NH4(CO3)	600 W	C
3% NH4Cl	600 W	C
ORGANICS		
D-Glucose		C, weakly at higher RF power
Sucrose		
Formic Acid		No combustion, but boiled over
Acetic Acid, in mini tt holder		Inconclusive. No combustion, but realized mini-holder may have been located too far up and out of the active field area; NaCl in same holder and location did not ignite

TABLE 1-continued

	Liquid Samples	
Sample Description, holder	RF power	Observations [C = combusts]
Acetic Acid		No combustion
Citric acid diammonium salt		*see solid samples data below*
Laurie Acid		*see solid samples data below*
Oxalic Acid		C, cracked tt
Salicylic Acid solution	600 W	No combustion - expanded up and nearly oozed out of the tt
2-Methoxycrotonic acid		*see solid samples data below*
3% L-Alanine in DI water	800 W	C, but had trouble igniting at lower power; faint flame
Poly-vinyl Alcohol solution	1200 W	C
Poly-vinyl alcohol, PTFE tt		No combustion but boiled over after field was turned off
Ethylene Acetate		C; sustains burn after field is off
Ethylene Glycol		No combustion - though it almost ignites
Ethylene Glycol	1000 W	C
Glycerol		C after increasing RF power; sustained flame after field turned off
Methanol	1000 W	C, just barely . . . low thin flame
Ethanol	0 W	EtOH combustion while adding drops of 3% NaCl
2-Propanol		C
Acetone	600 W+	C, Intensity of flame increases with RF power increase
n-decane	~1000 W	No combustion
Dodecane	~1000 W	No combustion
LAYERED INORGANICS		
CabOSil		C
CabOSil, PTFE tt		C
COLLOIDAL SILVERS		
ABL's start water*	800 W	C
ABL 10 ppm Ag	800 W	C
ABL 22 ppm Ag	800 W	C
ABL 32 ppm Ag	800 W	C
ABL 200 ppm Ag	High power	C but flame does not sustain
ABL 200 ppm Ag	800 W	No combustion, but boiled over, No combustion, Note that sample bottle showed silver settled to the bottom and hence the sample may not be considered as a colloid hence inconclusive data.
ABL 400 ppm Ag	800 W	No combustion, Note that sample bottle showed silver settled to the bottom and hence the sample may not be considered as a colloid hence inconclusive data.
ABL 200 ppm Ag, PTFE tt		C, but not sustained . . . then flame went out and it boiled over
ABL 410 ppm Ag, PTFE tt		No combustion, but overflowed from nearly boiling and turned brownish
GR water w 20 ppm Ag	1000 W	C
GR water w 32 ppm Ag	1000 W	C
GR water w Zinc colloid	1000 W	C

*"ABL" refers to reverse osmosis water from American Biotech Laboratories
**GR" refers to reverse osmosis water from General Resonance Corp.

Part 2: Solid Samples

[0181] Irradiation of the solid samples identified in the following Table 2 with Code A was done with the sample contained in a mini-holder having 5 ml volume capacity. After a series of runs on the first few samples, it was determined that this arrangement was not optimal, and so we conclude that most of this data is less than reliable. Nonetheless, it is reported here for the sake of completeness.

[0182] Irradiation of the solid samples identified in the following Table 2 with Code B was done using a standard ceramic crucible to hold the powder and pellet samples. A temperature sensor was dipped into the samples to register their temperature changes. All of the samples were positioned and run identically at 600 watts RF power for 2 minutes each, and their temperature rises were recorded. Sample mass varied—most of the time the crucible was filled ¾ of the way, but there were a few samples for which there was only a very small amount, so data was taken on what was available. Nonetheless, this data is believed to be more reliable, as the object of these experiments was only to discern gross effects.

[0183] Additional structural data from Raman spectroscopy and x-ray diffraction analysis of select samples is presented in the Attachment hereto. The following results were obtained

TABLE 2

Solid Samples

Sample Description, holder	Code	RF power	Observations [C = combusts]
INORGANICS			
Mg(OH)2	A		No combustion
Ca(OH)2	A		No combustion
Al(OH)3	A		No combustion. T went upto 170 F.
AlOOH. *x*H2O	A		No combustion. Temperature up with field on; field off at T = 122 F.
ORGANICS			
D-Glucose (anhydrous) dry	A		No combustion. Slow temperature rise; T = 77 F. at field off
Sucrose	A		No combustion. Temperature rose slowly with the field on: T = 76 F. at field off
Citric acid diammonium salt	A		No combustion. Slow temperature rise; T = 74 F. at field off
Na-oxalate, mini holder	A		No combustion. VERY slow rise in temperature
L-2-Alanine	A		No combustion
Poly-vinyl alcohol	A		No combustion
LAYERED INORGANICS			
Na-Montmorillonite	A		No combustion.
Ca-Montmorillonite	A		No combustion.
CabOSil	A		No combustion.
Plaster of Paris	A		No combustion.
Kaolinite	A		No combustion.
Hydrotalcite	A		No combustion.
Al2O3	B		T rises upto 97.4° C. in 2 min
Y2O3	B		No significant temp rise
Hydroxyl ethyl cellulose	B		T rises upto 89.9° C. in 2 min
LaPO4: Ce, Tb	B		T, Peaked upto 210° C. in 1.30 min then drops to 180 C.
CuO	B		T rises upto 111° C. in 2 min
B4C	B		Slow increase up to 39.7° C.
Crystalline silica	B		Slow increase up to 28° C.
Cellulose from crashed leaf	B		Slow increase up to 42.2° C.
Lead Manganese Niobate pellet	B		Slow increase up to 27° C.
LiNbOS pellet	B		Slow increase up to 27.8° C.
Doped ZnS, EL blend	B		Slow increase up to 30.8° C.

[0184] Additional structural data from x-ray diffraction analyses of selected samples subjected to RF irradiation according to this invention, as well as photographs showing combustion of the gas product generated by the electromagnetic dissociation of water in accordance with this invention, are shown in FIGS. **27-36**.

[0185] FIG. **27** shows that water containing a variety of different sodium salts, when electromagnetically dissociated according to this invention, produces a gas product that readily combusts.

[0186] FIG. **28** shows the same result occurring with water containing a variety of different transition metal salts.

[0187] FIG. **29** shows an XRD pattern on hydroxyethyl cellulose irradiated with RF energy in accordance with this invention, both before and after irradiation.

[0188] FIG. **30** is a view similar to FIGS. **1** and **2** showing that water containing phosphoric acid also produces a combustible product gas when subjected to RF energy in accordance with this invention.

[0189] FIG. **31** illustrates UV-VIS spectra of a group of aqueous colloidal silver samples containing different amounts of silver, both before and after irradiation.

[0190] FIG. **32**, which is an emission spectra of $SrCl_2$ in DI water, indicates that combustible H2 gas has been produced.

[0191] FIG. **33** illustrates the self ignition of salt water droplets in an RF field of this invention. Water droplets containing many other electrolytes described above also self combusts in this way.

[0192] FIG. **34** shows that structured water produced according to U.S. Pat. No. 7,119,312 of Steve Sedlayer also readily combusts when subjected to RF energy in accordance with this invention.

[0193] FIG. **35** shows the Raman Spectra of an NaCl-containing sample irradiated with RF energy in accordance with this invention, both before and after irradiation.

[0194] These figures further illustrate that irradiation with RF energy in accordance with this invention produces a significant change in the materials being irradiated.

[0195] Although only a few embodiments have been described above, many changes and additions can be made without departing from the spirit and scope of the invention. All such embodiments are intended to be included within the scope of the present invention.

TABLE A

EXEMPLARY COMPONENT SPECIFICATIONS

Fig #	RF Gen	C_1	C_2	C_3	L_1	L_2	FP_1	FP_2
10	N/A	Capacitor 100 pF, 15 kVDC PN: HECHT57Y101MA QTY: 11	Air Variable Capacitor 36-249 Pf 4 kVDC Surplus Sales of Nebraska PN: 12-53	Vacuum Variable Capacitor 12-500 pF 15 kVDC, Comet PN CVIC-500TN/15	Coil ¼"ϕ Copper Tube 10 turns 2.4" ϕ OD	Coil ¼"ϕ Copper Tube 32 turns 2.2" ϕ OD	N/A	N/A
11	ENI- 240 V 10-1500 W	Same as FIG. 10	Same as FIG. 10	Same as FIG. 10	Same as FIG. 10	Same as FIG. 10	3 plates 6"ϕ . . . 0.020" S.S. 0.375" Teflon Spacer 4"ϕ 0.020" S.S 0.375" Teflon Spacer 3"ϕ . . . 0.020" S.S. Teflon Spacer	N/A
12	Same as FIG. 11	Same as FIG. 10	Same as FIG. 10	Same as FIG. 10	Same as FIG. 10	Same as FIG. 10	Same as FIG. 11	3 plates 6"ϕ . . . 0.020" S.S. 0.375" Teflon Spacer 4"ϕ 0.020" S.S 0.375" Teflon Spacer 3"ϕ . . . 0.020" S.S. Teflon Spacer
16	Same as FIG. 11	Same as FIG. 10, QTY 13	Same as FIG. 10	Same as FIG. 10	Same as FIG. 10	Same as FIG. 10	6" ϕ × 0.125" thk copper plate	Same as FIG. 12
16a	Same as FIG. 11	Same as FIG. 16	Same as FIG. 10	Same as FIG. 10	Same as FIG. 10	Same as FIG. 10	Same as FIG. 16	9.5" silver coated square 0.125" thick copper

1. The electromagnetic dissociation of water.

2. The process of claim **1**, wherein the water is electrically conductive.

3. The process of claim **2**, wherein the water contains an electrolyte.

4. The process of claim **3**, wherein the electrolyte is an ionizable salt.

5. The process of clam **3**, wherein the electrolyte is a dispersion of electrically conductive particles.

6. The process of claim **1**, wherein the water is irradiated with electromagnetic radiation having a frequency of less than 300 MHz.

7. The process of claim **6**, wherein the frequency is less than 200 MHz.

8. The process of claim **7**, wherein the frequency is about 3-30 MHz.

9. The process of claim **6**, wherein the water is irradiated with photons having an energy of 10^{-6} eV or less.

10. The process of claim **6**, wherein the water is irradiated with photons having an energy of 10^{-7} eV or less.

11. The process of claim **6**, wherein the water is irradiated with photons having an energy of 10^{-8} eV or less.

12. Heating a material by irradiating the material with RF energy having a frequency of less than 300 MHz.

13. The process of claim **12**, wherein the frequency is less than 200 MHz.

14. The process of claim **13**, wherein the frequency is about 3-30 MHz.

15. The process of claim **12**, wherein the material is irradiated with photons having an energy of 10^{-6} eV or less.

16. The process of claim **15**, wherein the material is irradiated with photons having an energy of 10^{-7} eV or less.

17. The process of claim **16**, wherein the material is irradiated with photons having an energy of 10^{-8} eV or less.

* * * * *

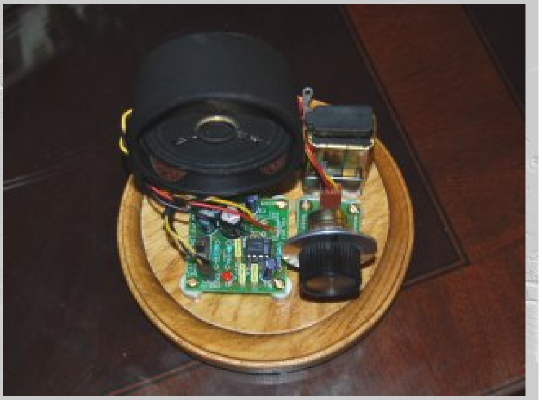

A breadboard Audio Amplifier

Back to Radio..

A little project stemming from my persistant difficulty in hearing. (Losing one's hearing is never fun to admit, but combined with crystal radio, well, its trouble!) I hope I do not offend any "purists" with this project. Still, I understand that many crystal radio owners keep a small amplifier too, for "those times" when, well, you know.. So, this amplifier project, its for, well, those times.. (I suspect Radio Shack would have discontinued their nifty Mini Amplifier years ago had it not been for the crystal radio community.)

My idea was to build my own unit rather than use a store-bought amplifier. Additionally, I wished to put my own stamp on it, make it unique. I started with a kit for the basic layout and electronics. I wanted to base things on the NI LM386 chip which is a good part and hard to break. I found several distributers of what appears to me as the exact same kit and so chose the cheapest $6 dollar kit. NI sells the LM386 for 34 cents a pop and would even provide a few "samples" but then I would be stuck laying things out and wireing.. Been there, done that. A printed circuit board is just SO much easier to use.

One of my inspirations for this project also stems from the nature of modern electronics in general, and the Radio Shack mini amplifier in partictular. Using my RS mini, I am always struck by the contrast between electronics new and old. With crystal radios one builds with passion, creating something more than a mere radio, but a work of art, a piece of furniture. Still, the technology itself is primitive dating back to the dawn of radio. When I look at the mini amplifier, it looks like a toy yet inside there is the most wonderful technology, solid state integrated circuitry. We take this advanced technology and slap it inside a cheap plastic box, an appliance. Naturally, I don't complain about the price, but for my project I wish for something above.

The solution struck me as obvious, give the same care and respect for the amplifier as I do the radio. I also sought to maintain a "vintage" visual aspect for the unit, breadboard layout, antique binding posts and a large adequate volume control knob, (the pitiful volume control wheel on my RS mini has already come apart once and threatens to break at any time). I wanted a nice lacquered base and sturdy speaker enclosure. The enclosure was the main challenge. I finally ended up with a heavy plactic goblet from a local thrift shop cut down to a tapered cylinder into which the speaker can be placed tightly. The rest of the set is modular with plug-in connections for the audio in, power, speaker, and volume control potentiometer. A question, do I add a brass knife switch for the on-off? I worry I am out of room on the base. Still pondering this.

I was initially worried that the audio input impedance may cause me trouble. I read so much about this, impedance matching, transformers, etc etc, almost a fetish really. I have subsequently learned that the LM386 input impedance is about 800 Megohms, the 10k pot being the limiting factor. So, with considerable relief, I concluded this ought not present any problems with the crystal radio circuits. I find the amplifier works very nicely with my radios. The sound is as good as the Radio Shack unit.

The amplifier certainly adds to the vintage "look" of my setup. Well, you be the judge, I hope I come close, please check out the photos...

From the front, wooden base, full display of components, big adequate volume control..

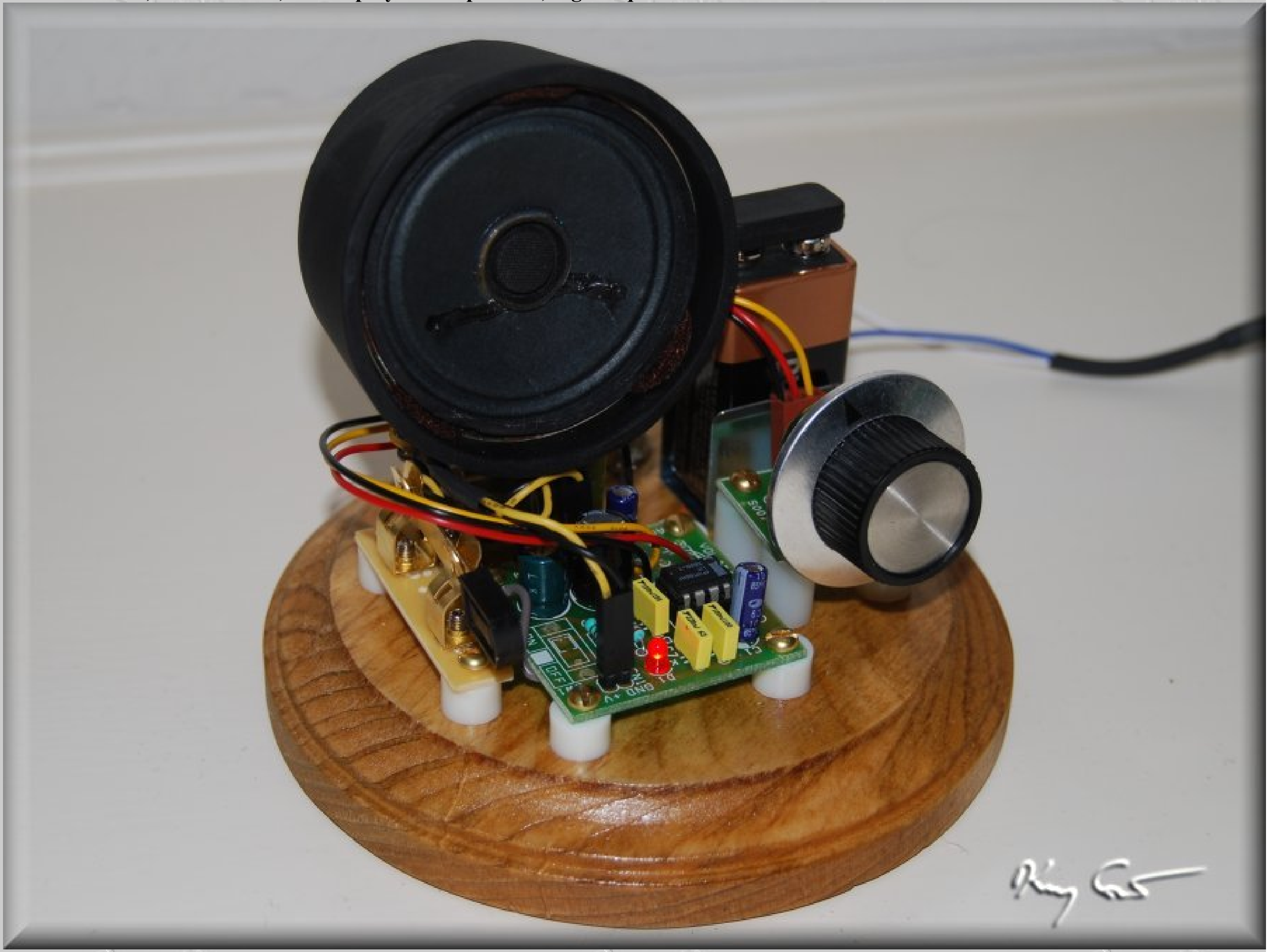

[illegible]

It even looks nice from behind, vintage binding posts, naturally..

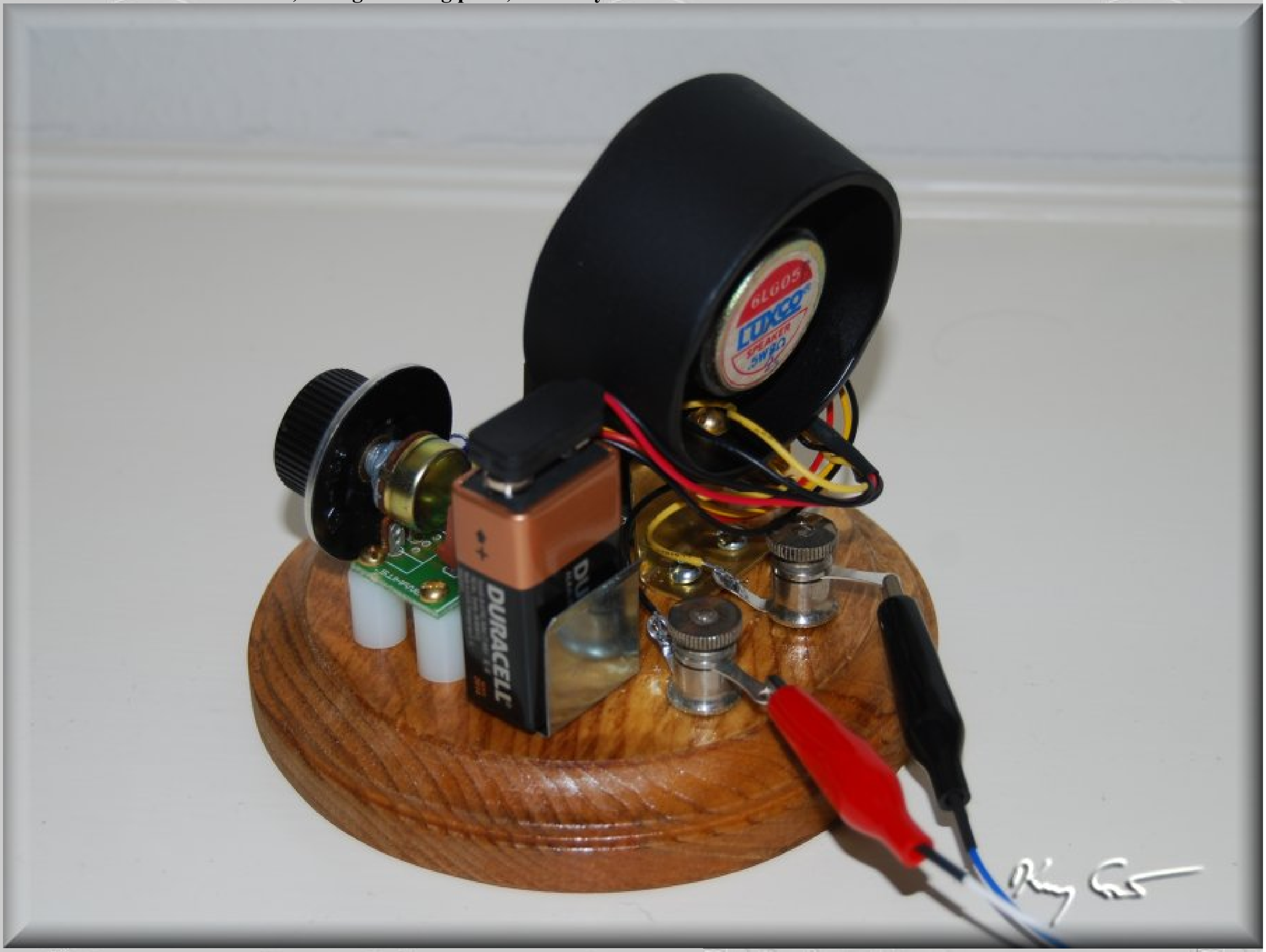

Schematic of the kit, (AnyKits.com) A005. The kit is configured with a maximum gain between 45 and 50, too high for headphones but too low really for listening with the speaker easily. You will note the location of R1 (in series with a 10uF capacitor) between pins 1 and 8 on the LM386. If you short the resistor leaving only the capacitor, then the gain will increase to the maximum of 200. If you do such a project, I recommend you download the LM386 datasheet from National Instruments.

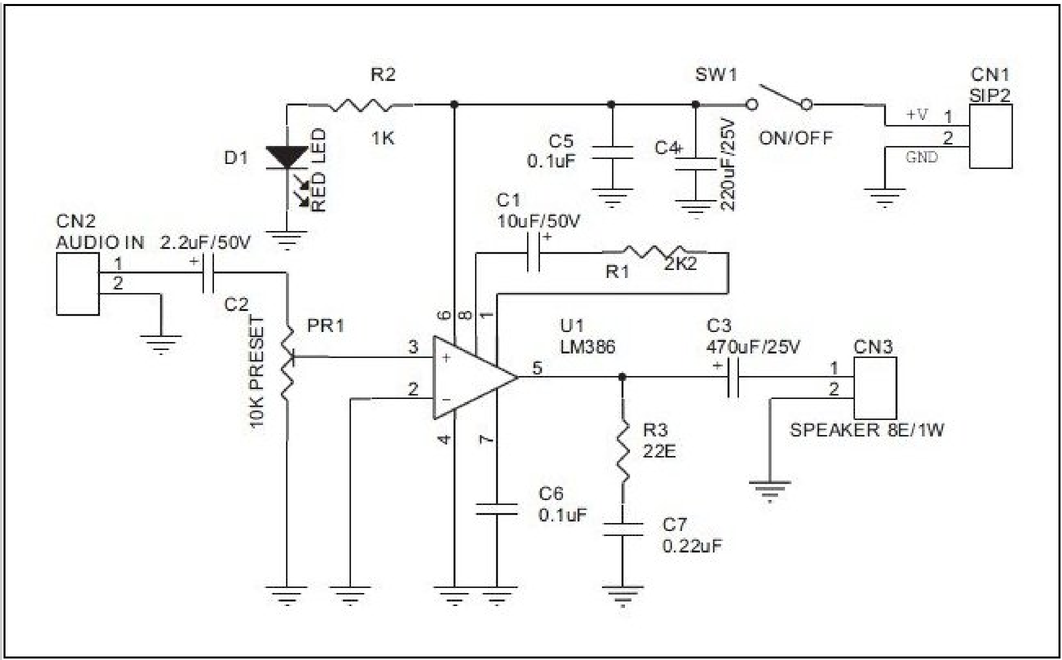
R2
1K
SW1
ON/OFF
CN1
SIP2
+V
GND
D1
RED LED
C5
0.1uF
C4
220uF/25V
C1
10uF/50V
R1
2K2
CN2
AUDIO IN
2.2uF/50V
C2
PR1
10K PRESET
U1
LM386
C3
470uF/25V
CN3
SPEAKER 8E/1W
R3
22E
C6
0.1uF
C7
0.22uF

[Return..](#)

[Back to Main....](#)

Thursday, 04-Jan-2018 16:25:31 CST

Comments and/or Correspondance to: kj-smith@sbcglobal.net

.::. Alsace .::. Life List .::. Ionosphere .::. Crystal Radio .::. Snapshots .::. Thoughts .::.
.::. Sitemap .::. Beer Labels .::. International Pix .::. Western US Pix .::. Pinhole Pix .::. Petrology .::.

Part II: Diode Test Data Report

Back to Lab..

Test Results

Over a period of several weeks in 2013 I ran I-V characterization measurements on a good number and variety of diodes, crystals and my two tubes to see how very thing stacks up. The following presents the resulting curves, diode photos, and some puzzling questions/conclusions I churned up in the process. These Characteristic Curves carry the voltage out to 2.0 Volts, well beyond any signal strength one might expect to receive with a crystal set. This is important to note because the curves themselves are dominated by the diode series resistance. After each discussion I then include my spreadsheet with the detail DX-level analyses that provide the critical information about diode "n", (Ideality factor) "Is", (Reverse Saturation current) and "Ro", (Zero-bias Diode resistance). These are the data the set designer requires.

I start off with the realization that when one orders diodes, it is by no means certain what one will get. This seems especially true for the supposedly ubiquitous 1N34A. Unless you see the part number actually on the diode, you probably need to test it to know what it really is. So, lets take a look!

Germanium Diodes

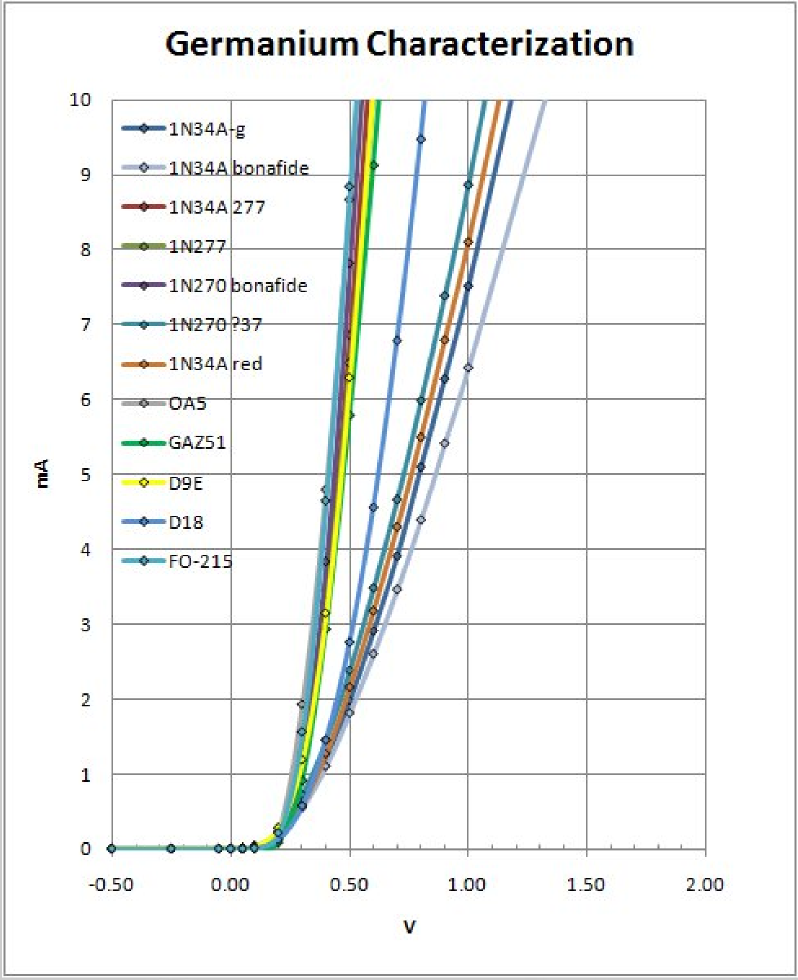

1N34A series

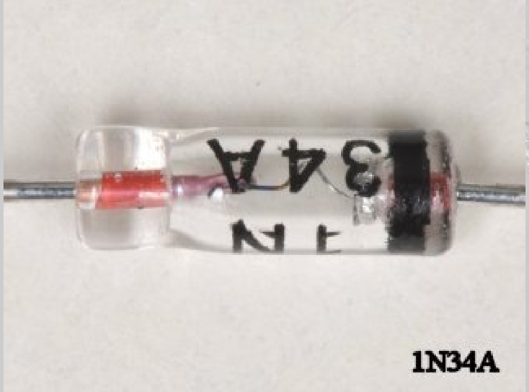

1N34A

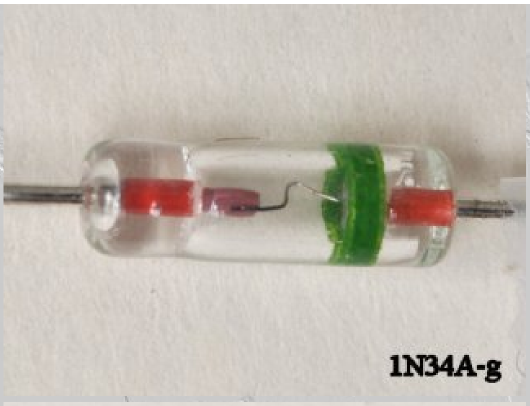

1N34A-g

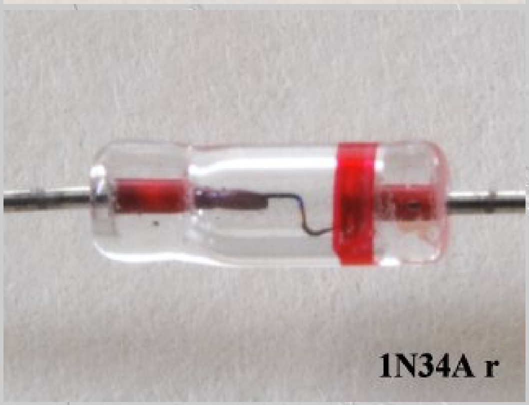

1N34A r

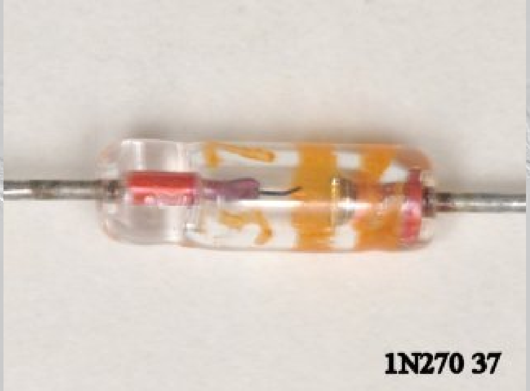

1N270 37

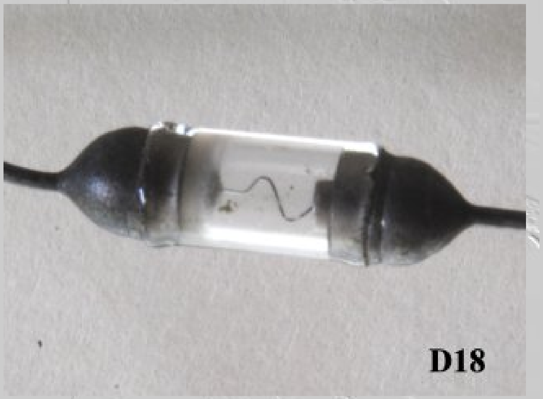

D18

1N270-277 series

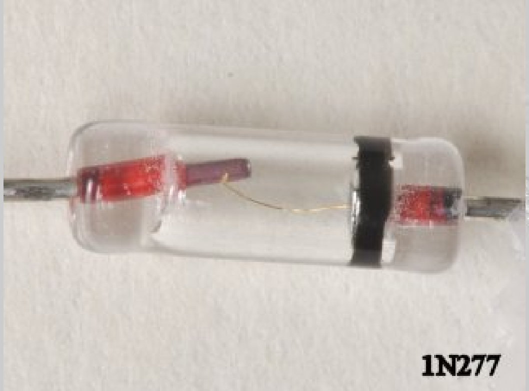

1N277

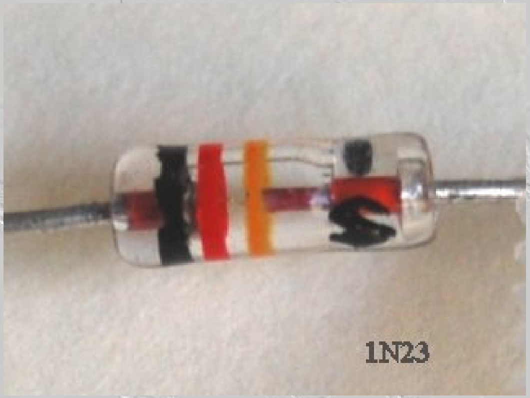

1N23

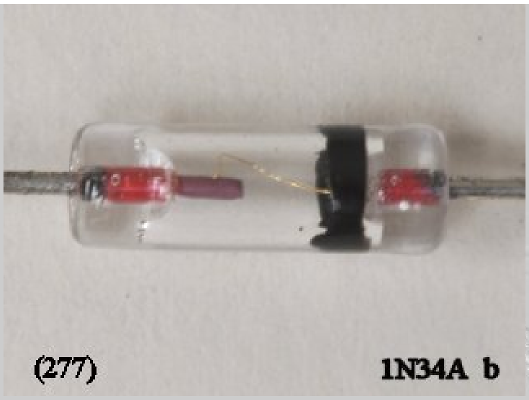

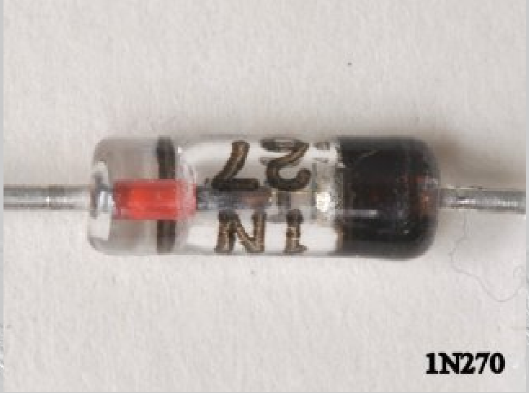

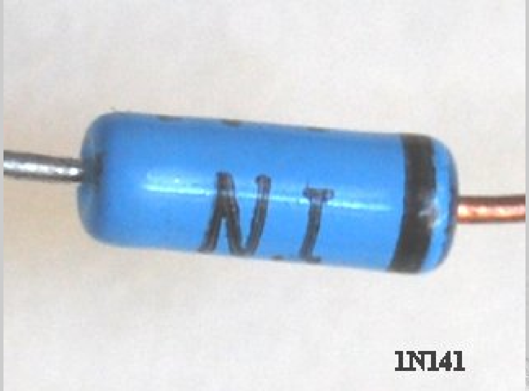

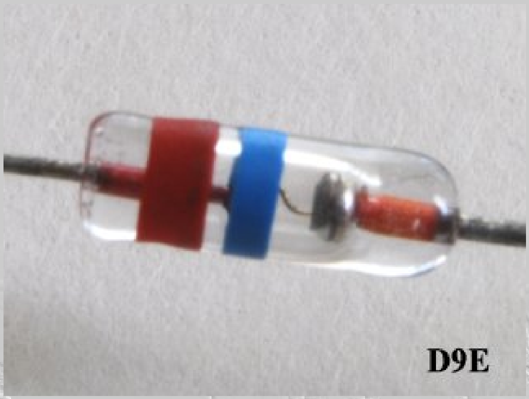

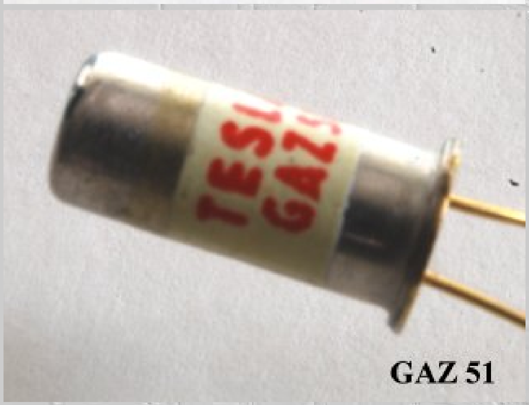
TESLA
GAZ 51

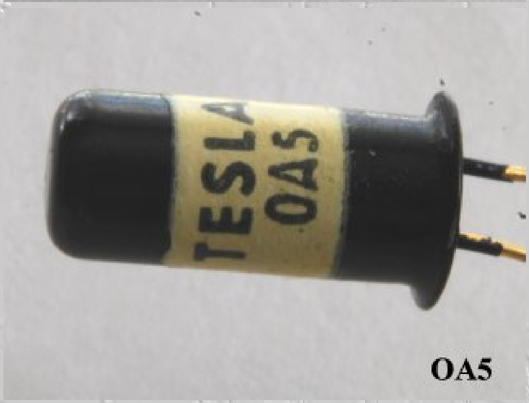
TESLA
OA5
OA5

ITT Diodes, Something special here.

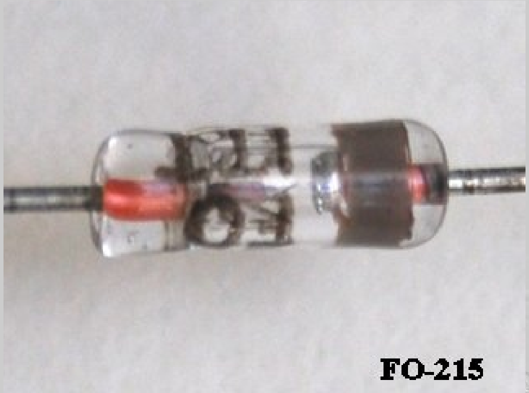

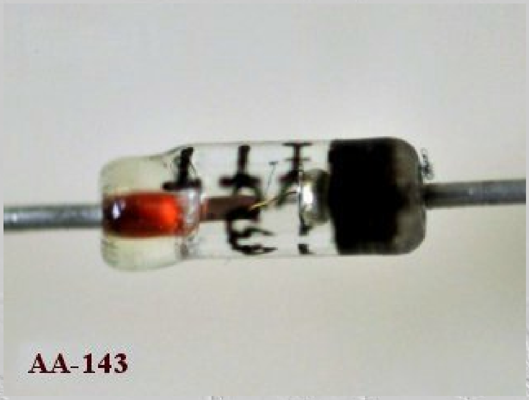

One oddball Ge diode, not sure what it is for..

Here we see a set of Germanium diodes and photos of the diodes under test. Immediately you should notice two sets of curves, what I call the "1N34A" set and the "1N270-277" set. In this analysis, the second set appears clearly superior in sensitivity and deservedly so. In the photos it is clear that one bunch of diodes I purchased as 1N34A were in fact better 1N277, the black band and thin gold contact wire as well as common curve unite these. "True" 1N34A's have a more robust contact wire. I also purchased a 1N270 and received pretty orange diodes with the number "37" stamped on it, a fairly robust contact wire and a curve that most closely resembles a 1N34A, go figure. Mercifully, I DID manage to get some bona-fide 1N34A's with the part number on the diode, as well as a single bona-fide ITT 1N270 diode with the part number. These are my base for comparison.
In addition to the standard Germanium Diodes tested above, I also have latched my hands on a few "other" germanium diodes for the fun of testing. These include two vintage Russian types, a D9E (in the 270 class) and a D-18 (transitional between 270 and 1N34A). Additionally, I acquired a few FUZZ diodes popular, I suspect, with the guitar gadget crowd. These are larger packages in metal cases so I cannot see how the internal contact is made other that they are stated to be gold bonded point contact type. Clearly these are highly sensitive diodes in my 270 class, the OA5 looking best of the lot. Interestingly, I recently purchased from good Mr. Peebles a few of his "Holy Grail" ITT FO-215 diodes. My measured characteristic for one of these is a dead lay down on the gold-bonded OA5. From the photos, these are in different packaging with the FO-215 in a traditional glass casing. What makes the FO-215 so great is the fact that its resistance Rd is an interesting 150k ohm or so which matches very well with typical moderate Q tank circuits. Interestingly, I find that an old Russian D18 diode has very similar characteristics to the FO-215 and so must also be placed into the "Holy Grail" class of germanium diode. Take your pick!

Spreadsheet for calculation of Germanium diode n and Is (modified from Mike Tuggle's spreadsheet)

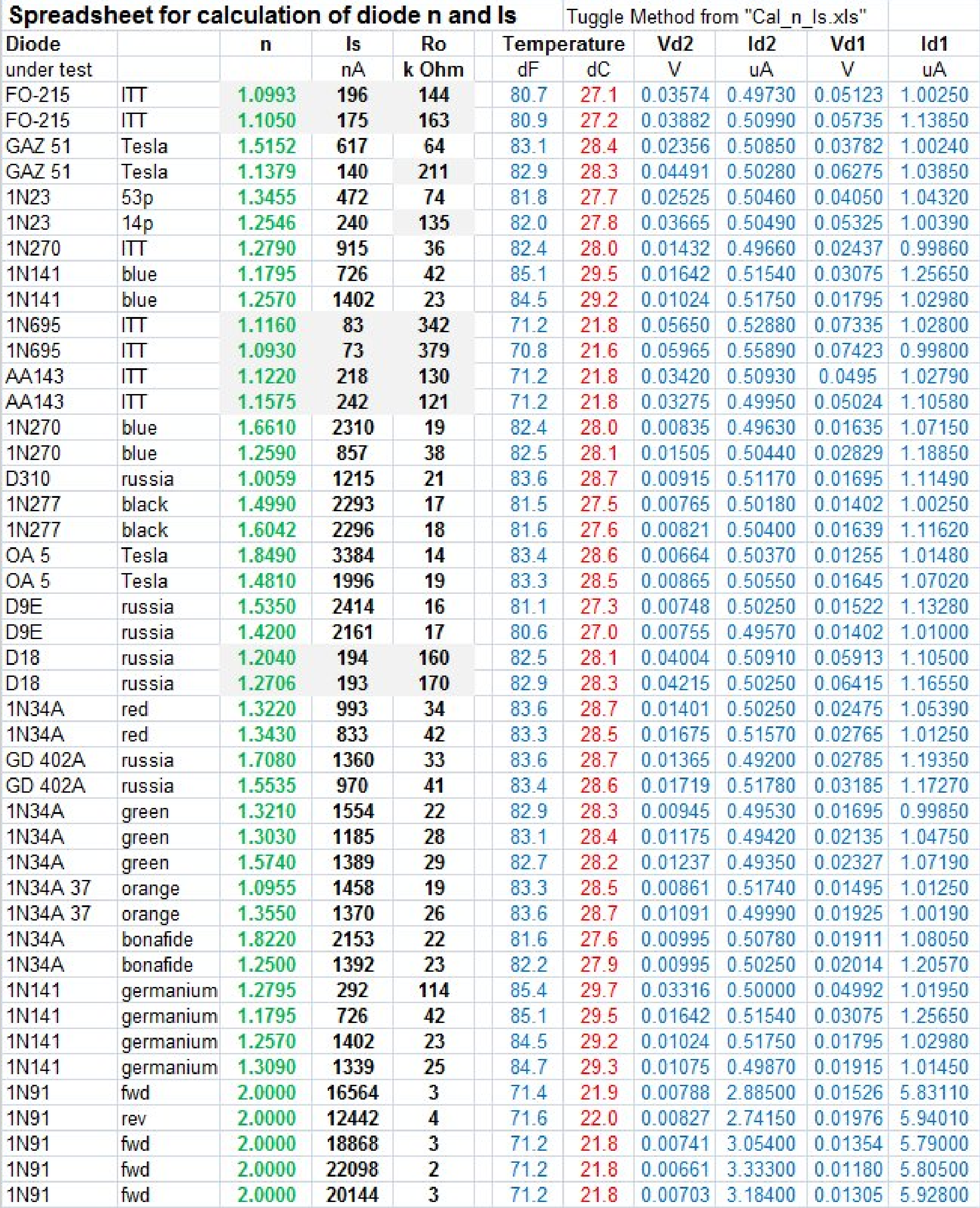

Spreadsheet for calculation of diode n and Is Tuggle Method from "Cal_n_Is.xls"

Diode		n	Is	Ro	Temperature		Vd2	Id2	Vd1	Id1
under test			nA	k Ohm	dF	dC	V	uA	V	uA
FO-215	ITT	1.0993	196	144	80.7	27.1	0.03574	0.49730	0.05123	1.00250
FO-215	ITT	1.1050	175	163	80.9	27.2	0.03882	0.50990	0.05735	1.13850
GAZ 51	Tesla	1.5152	617	64	83.1	28.4	0.02356	0.50850	0.03782	1.00240
GAZ 51	Tesla	1.1379	140	211	82.9	28.3	0.04491	0.50280	0.06275	1.03850
1N23	53p	1.3455	472	74	81.8	27.7	0.02525	0.50460	0.04050	1.04320
1N23	14p	1.2546	240	135	82.0	27.8	0.03665	0.50490	0.05325	1.00390
1N270	ITT	1.2790	915	36	82.4	28.0	0.01432	0.49660	0.02437	0.99860
1N141	blue	1.1795	726	42	85.1	29.5	0.01642	0.51540	0.03075	1.25650
1N141	blue	1.2570	1402	23	84.5	29.2	0.01024	0.51750	0.01795	1.02980
1N695	ITT	1.1160	83	342	71.2	21.8	0.05650	0.52880	0.07335	1.02800
1N695	ITT	1.0930	73	379	70.8	21.6	0.05965	0.55890	0.07423	0.99800
AA143	ITT	1.1220	218	130	71.2	21.8	0.03420	0.50930	0.0495	1.02790
AA143	ITT	1.1575	242	121	71.2	21.8	0.03275	0.49950	0.05024	1.10580
1N270	blue	1.6610	2310	19	82.4	28.0	0.00835	0.49630	0.01635	1.07150
1N270	blue	1.2590	857	38	82.5	28.1	0.01505	0.50440	0.02829	1.18850
D310	russia	1.0059	1215	21	83.6	28.7	0.00915	0.51170	0.01695	1.11490
1N277	black	1.4990	2293	17	81.5	27.5	0.00765	0.50180	0.01402	1.00250
1N277	black	1.6042	2296	18	81.6	27.6	0.00821	0.50400	0.01639	1.11620
OA 5	Tesla	1.8490	3384	14	83.4	28.6	0.00664	0.50370	0.01255	1.01480
OA 5	Tesla	1.4810	1996	19	83.3	28.5	0.00865	0.50550	0.01645	1.07020
D9E	russia	1.5350	2414	16	81.1	27.3	0.00748	0.50250	0.01522	1.13280
D9E	russia	1.4200	2161	17	80.6	27.0	0.00755	0.49570	0.01402	1.01000
D18	russia	1.2040	194	160	82.5	28.1	0.04004	0.50910	0.05913	1.10500
D18	russia	1.2706	193	170	82.9	28.3	0.04215	0.50250	0.06415	1.16550
1N34A	red	1.3220	993	34	83.6	28.7	0.01401	0.50250	0.02475	1.05390
1N34A	red	1.3430	833	42	83.3	28.5	0.01675	0.51570	0.02765	1.01250
GD 402A	russia	1.7080	1360	33	83.6	28.7	0.01365	0.49200	0.02785	1.19350
GD 402A	russia	1.5535	970	41	83.4	28.6	0.01719	0.51780	0.03185	1.17270
1N34A	green	1.3210	1554	22	82.9	28.3	0.00945	0.49530	0.01695	0.99850
1N34A	green	1.3030	1185	28	83.1	28.4	0.01175	0.49420	0.02135	1.04750
1N34A	green	1.5740	1389	29	82.7	28.2	0.01237	0.49350	0.02327	1.07190
1N34A 37	orange	1.0955	1458	19	83.3	28.5	0.00861	0.51740	0.01495	1.01250
1N34A 37	orange	1.3550	1370	26	83.6	28.7	0.01091	0.49990	0.01925	1.00190
1N34A	bonafide	1.8220	2153	22	81.6	27.6	0.00995	0.50780	0.01911	1.08050
1N34A	bonafide	1.2500	1392	23	82.2	27.9	0.00995	0.50250	0.02014	1.20570
1N141	germanium	1.2795	292	114	85.4	29.7	0.03316	0.50000	0.04992	1.01950
1N141	germanium	1.1795	726	42	85.1	29.5	0.01642	0.51540	0.03075	1.25650
1N141	germanium	1.2570	1402	23	84.5	29.2	0.01024	0.51750	0.01795	1.02980
1N141	germanium	1.3090	1339	25	84.7	29.3	0.01075	0.49870	0.01915	1.01450
1N91	fwd	2.0000	16564	3	71.4	21.9	0.00788	2.88500	0.01526	5.83110
1N91	rev	2.0000	12442	4	71.6	22.0	0.00827	2.74150	0.01976	5.94010
1N91	fwd	2.0000	18868	3	71.2	21.8	0.00741	3.05400	0.01354	5.79000
1N91	fwd	2.0000	22098	2	71.2	21.8	0.00661	3.33300	0.01180	5.80500
1N91	fwd	2.0000	20144	3	71.2	21.8	0.00703	3.18400	0.01305	5.92800

The above spreadsheet is based on measurements of the diodes shown above. For the determination of Is and n, I chose at random two examples from my collection of various diodes, (all germanium in this case) and measured via a modified version of the methodology outlined by Ben Tongue and Mike Tuggle. For this work I have chosen to set Id2 about 0.5uA and Id1 about 1.0uA and then read the needed voltage. This is the inverse of the normal method but the justification is that for any diode regardless of forward voltage drop the measurement is made at the same part of the LOG I vs V characteristic. This allows comparison between all diodes. I have included the actual room temperature in the calculations in order to eliminate this variable as a source of doubt or error. All the measured parameters as well as the determined results are presented in order to facilitate repeatability.

A final note on the plots above. You will have noticed that I sub-divide germanium diodes into a "1N34A" class and a "1N270-277" class. This is a result of looking at their I / V characteristics on a plot of 1 - 2V DC versus 0 - 10mA scale. At such a scale the slope of the characteristic for the 1N270 class remains much steeper than that for 1N34 type diodes. As this is not seen as a distinguishing feature at the small-signal scales looked at here, I judge the chief difference between the two types lies in the diode series resistance Rs. This impacts strong-signals but is of no concern for small-signal DX work. Just so you know.

Schottky Diodes

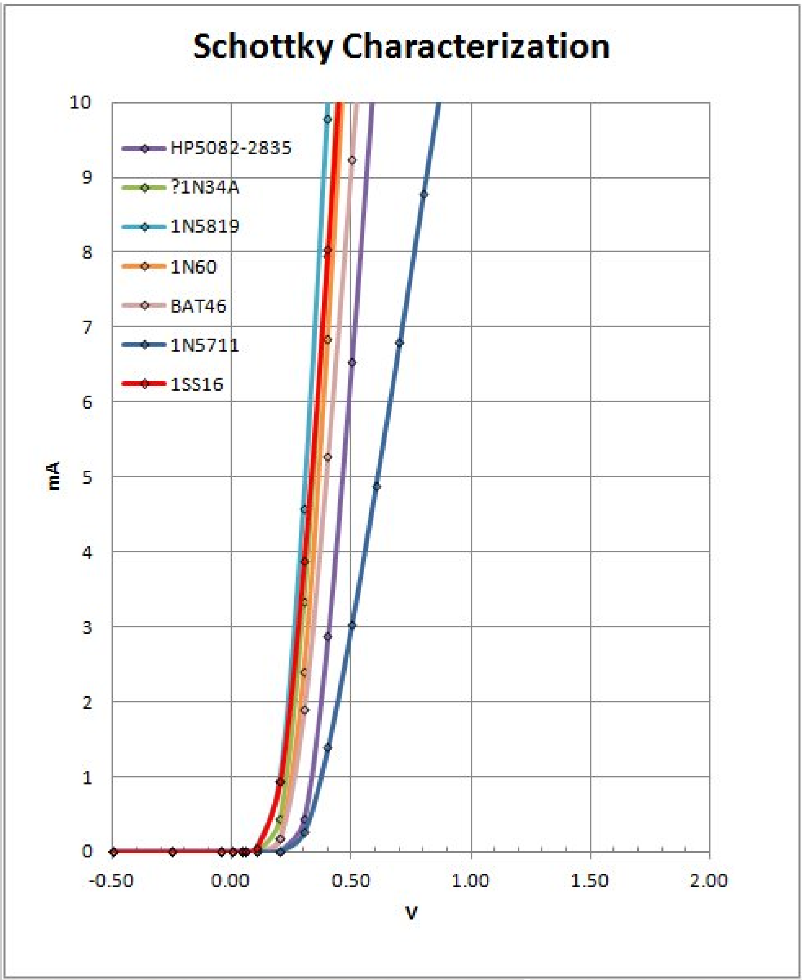

HP5082-2835

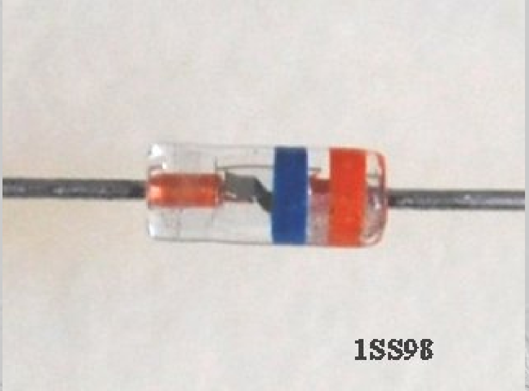
1SS98

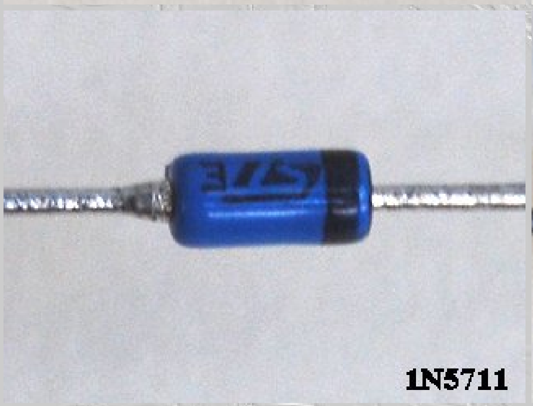
1N5711

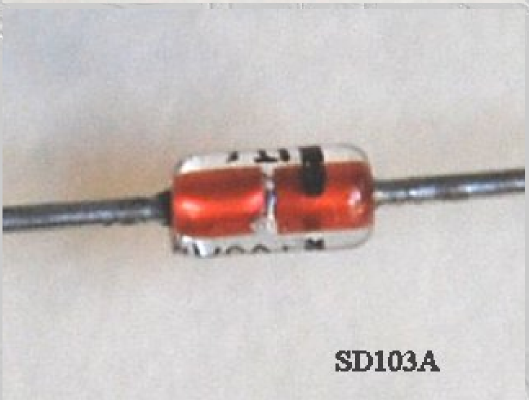
SD103A

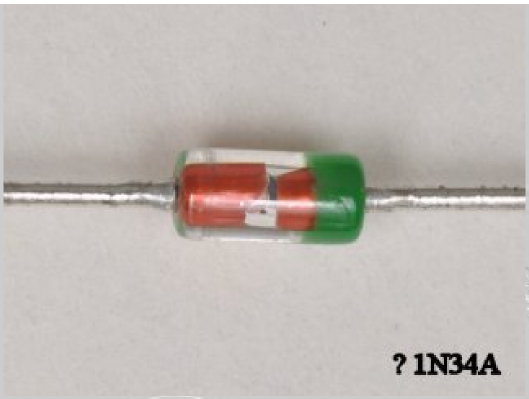
? 1N34A

1N60

BAT46

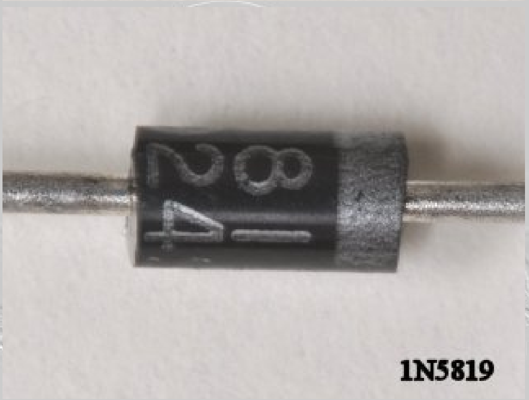
1N5819

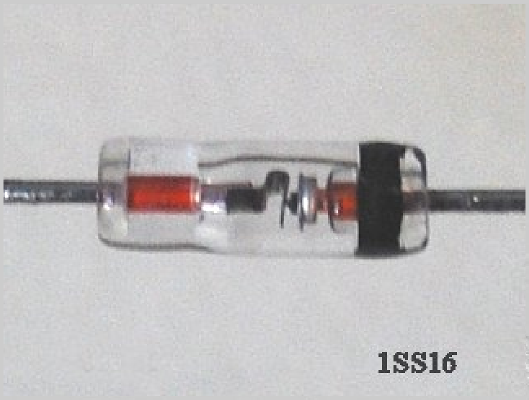
1SS16

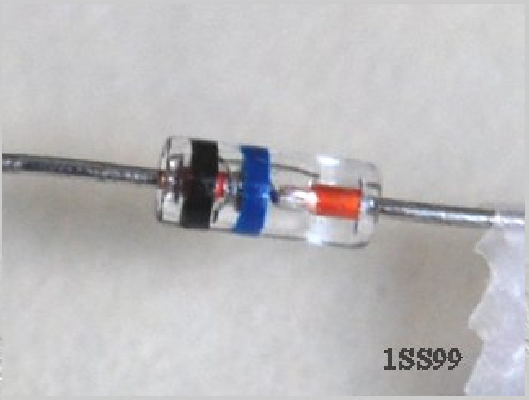
1SS99

Schottky diodes are very sensitive diodes that work excellently in crystal radios. Their construction and theory are different and I confess to not fully understanding these components. Still, from the characteristic curves, they are excellent! Again, note that you don't always get what you bargain for. Here I found what was supposed to be 1N34A's to be some sort of Schottky of unknown pedigree. The 1N5819, while having the lowest forward voltage drop, has a rather very high Junction Capacitance and will not perform well. Posts on the RaidoBoard Crystal Radio Forum however highly recommend the 1N5711 for crystal sets although the characteristic curve doesn't look that fabulous. Many web pages out there on Shottky diodes, I recommend you do your homework. Of all the schottky's, I note that Ben Tongue recommends most highly the HP5083-2835. The high resistance Ro makes them useful for DX sets with very very high Q tanks. Even so the diodes need to be paralleled with up to 4 or 5 diodes to match correctly the Rd with the tank Rp. I have found these somewhat hard to find and expensive, especially when one requires using several in parallel. I recently measured a few 1SS98 diodes and I discover them to have characteristics extremely similar to the HP and I feel they deserve more attention. Sadly, on ebay they seem to be just as difficult to find and expensive. No free ride!

Spreadsheet for calculation of Schottky diode n and Is (modified from Mike Tuggle's spreadsheet)

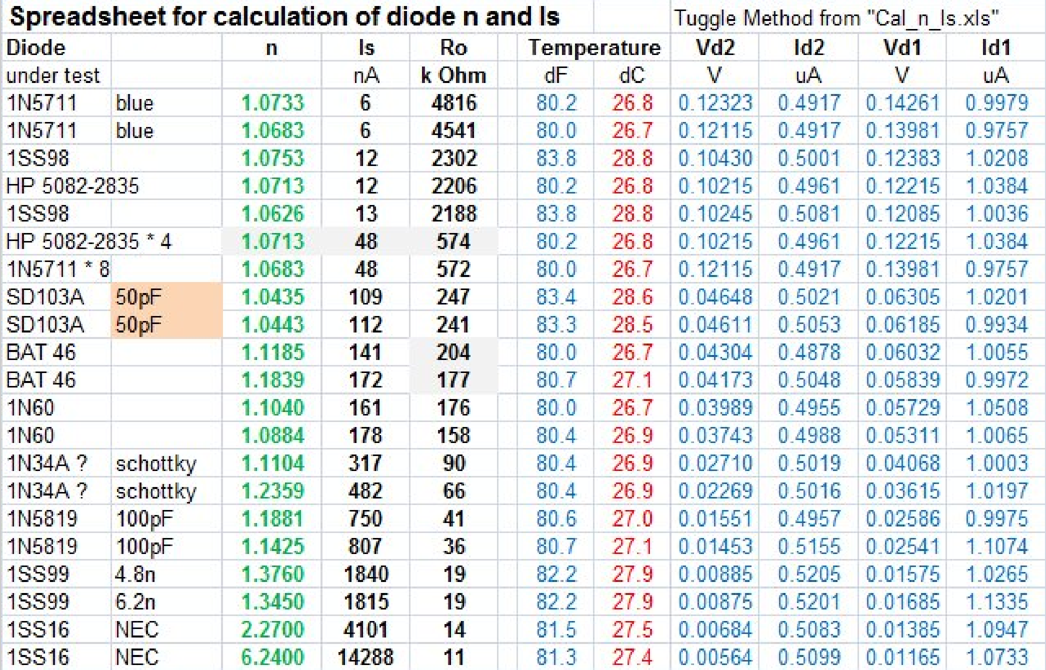

Spreadsheet for calculation of diode n and Is							Tuggle Method from "Cal_n_Is.xls"			
Diode		n	Is	Ro	Temperature		Vd2	Id2	Vd1	Id1
under test			nA	k Ohm	dF	dC	V	uA	V	uA
1N5711	blue	1.0733	6	4816	80.2	26.8	0.12323	0.4917	0.14261	0.9979
1N5711	blue	1.0683	6	4541	80.0	26.7	0.12115	0.4917	0.13981	0.9757
1SS98		1.0753	12	2302	83.8	28.8	0.10430	0.5001	0.12383	1.0208
HP 5082-2835		1.0713	12	2206	80.2	26.8	0.10215	0.4961	0.12215	1.0384
1SS98		1.0626	13	2188	83.8	28.8	0.10245	0.5081	0.12085	1.0036
HP 5082-2835 * 4		1.0713	48	574	80.2	26.8	0.10215	0.4961	0.12215	1.0384
1N5711 * 8		1.0683	48	572	80.0	26.7	0.12115	0.4917	0.13981	0.9757
SD103A	50pF	1.0435	109	247	83.4	28.6	0.04648	0.5021	0.06305	1.0201
SD103A	50pF	1.0443	112	241	83.3	28.5	0.04611	0.5053	0.06185	0.9934
BAT 46		1.1185	141	204	80.0	26.7	0.04304	0.4878	0.06032	1.0055
BAT 46		1.1839	172	177	80.7	27.1	0.04173	0.5048	0.05839	0.9972
1N60		1.1040	161	176	80.0	26.7	0.03989	0.4955	0.05729	1.0508
1N60		1.0884	178	158	80.4	26.9	0.03743	0.4988	0.05311	1.0065
1N34A ?	schottky	1.1104	317	90	80.4	26.9	0.02710	0.5019	0.04068	1.0003
1N34A ?	schottky	1.2359	482	66	80.4	26.9	0.02269	0.5016	0.03615	1.0197
1N5819	100pF	1.1881	750	41	80.6	27.0	0.01551	0.4957	0.02586	0.9975
1N5819	100pF	1.1425	807	36	80.7	27.1	0.01453	0.5155	0.02541	1.1074
1SS99	4.8n	1.3760	1840	19	82.2	27.9	0.00885	0.5205	0.01575	1.0265
1SS99	6.2n	1.3450	1815	19	82.2	27.9	0.00875	0.5201	0.01685	1.1335
1SS16	NEC	2.2700	4101	14	81.5	27.5	0.00684	0.5083	0.01385	1.0947
1SS16	NEC	6.2400	14288	11	81.3	27.4	0.00564	0.5099	0.01165	1.0733

The above spreadsheet is based on measurements of the diodes shown above. For the determination of Is and n, I chose at random two examples from my collection of various diodes, (all schottky in this case) and measured via a modified version of the methodology outlined by Ben Tongue and Mike Tuggle. For this work I have chosen to set Id2 about 0.5uA and Id1 about 1.0uA and then read the needed voltage. This is the inverse of the normal method but the justification is that for any diode regardless of forward voltage drop the measurement is made at the same part of the LOG I vs V characteristic. This allows comparison between all diodes. I have included the actual room temperature in the calculations in order to eliminate this variable as a source of doubt or error. All the measured parameters as well as the determined results are presented in order to facilitate repeatability.

Note in the above data that both the BAT-46 and the 1N60 diodes appear to have excellent properties for crystal radio work. In fact they have very similar parameters overall. Looking at the photos you should also note that the two types of diode look suspiciously exactly alike. The BAT 46 has it correct markings but the 1N60 is unmarked. You never know what you get!

Silicon Diodes

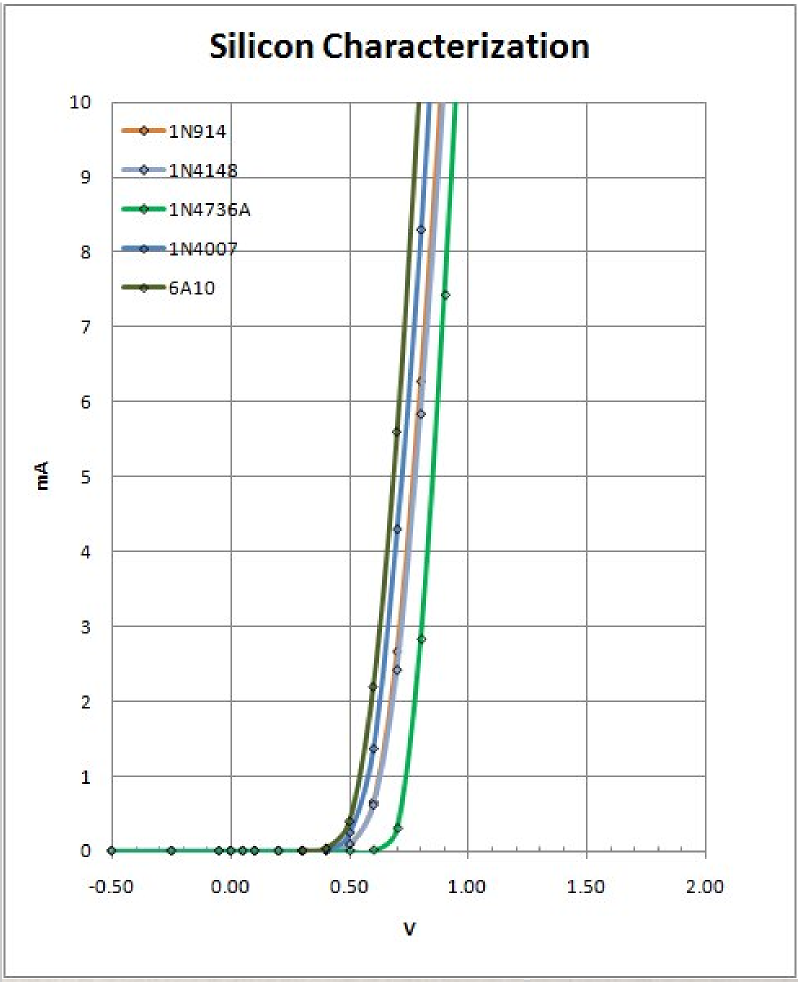

1N914

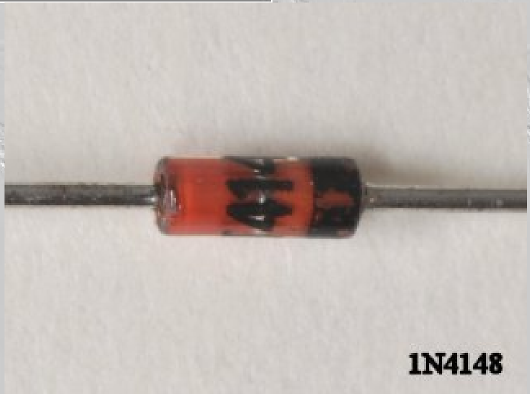
1N4148

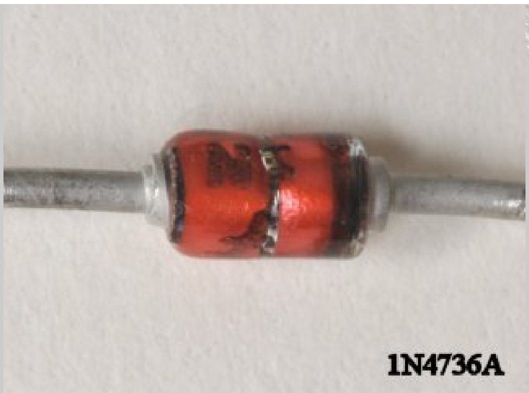
1N4736A

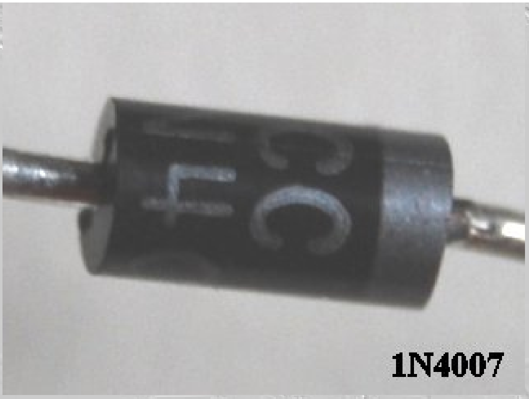
1N4007

6A10

Silicon diodes have good characteristics, but an unacceptably high forward voltage drop making them a very poor choice for crystal radio unless used with bias. The 1N4736A is a Zenner diode.

Spreadsheet for calculation of Silicon diode n and Is (modified from Mike Tuggle's spreadsheet)

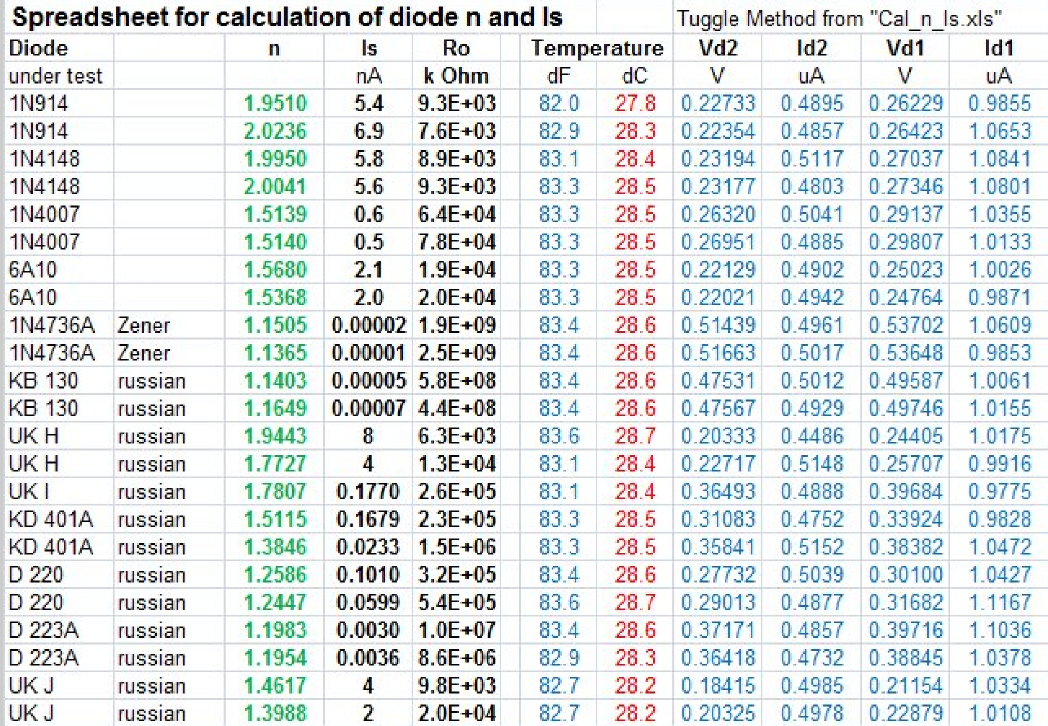

Spreadsheet for calculation of diode n and Is							Tuggle Method from "Cal_n_Is.xls"			
Diode		n	Is	Ro	Temperature		Vd2	Id2	Vd1	Id1
under test			nA	k Ohm	dF	dC	V	uA	V	uA
1N914		1.9510	5.4	9.3E+03	82.0	27.8	0.22733	0.4895	0.26229	0.9855
1N914		2.0236	6.9	7.6E+03	82.9	28.3	0.22354	0.4857	0.26423	1.0653
1N4148		1.9950	5.8	8.9E+03	83.1	28.4	0.23194	0.5117	0.27037	1.0841
1N4148		2.0041	5.6	9.3E+03	83.3	28.5	0.23177	0.4803	0.27346	1.0801
1N4007		1.5139	0.6	6.4E+04	83.3	28.5	0.26320	0.5041	0.29137	1.0355
1N4007		1.5140	0.5	7.8E+04	83.3	28.5	0.26951	0.4885	0.29807	1.0133
6A10		1.5680	2.1	1.9E+04	83.3	28.5	0.22129	0.4902	0.25023	1.0026
6A10		1.5368	2.0	2.0E+04	83.3	28.5	0.22021	0.4942	0.24764	0.9871
1N4736A	Zener	1.1505	0.00002	1.9E+09	83.4	28.6	0.51439	0.4961	0.53702	1.0609
1N4736A	Zener	1.1365	0.00001	2.5E+09	83.4	28.6	0.51663	0.5017	0.53648	0.9853
KB 130	russian	1.1403	0.00005	5.8E+08	83.4	28.6	0.47531	0.5012	0.49587	1.0061
KB 130	russian	1.1649	0.00007	4.4E+08	83.4	28.6	0.47567	0.4929	0.49746	1.0155
UK H	russian	1.9443	8	6.3E+03	83.6	28.7	0.20333	0.4486	0.24405	1.0175
UK H	russian	1.7727	4	1.3E+04	83.1	28.4	0.22717	0.5148	0.25707	0.9916
UK I	russian	1.7807	0.1770	2.6E+05	83.1	28.4	0.36493	0.4888	0.39684	0.9775
KD 401A	russian	1.5115	0.1679	2.3E+05	83.3	28.5	0.31083	0.4752	0.33924	0.9828
KD 401A	russian	1.3846	0.0233	1.5E+06	83.3	28.5	0.35841	0.5152	0.38382	1.0472
D 220	russian	1.2586	0.1010	3.2E+05	83.4	28.6	0.27732	0.5039	0.30100	1.0427
D 220	russian	1.2447	0.0599	5.4E+05	83.6	28.7	0.29013	0.4877	0.31682	1.1167
D 223A	russian	1.1983	0.0030	1.0E+07	83.4	28.6	0.37171	0.4857	0.39716	1.1036
D 223A	russian	1.1954	0.0036	8.6E+06	82.9	28.3	0.36418	0.4732	0.38845	1.0378
UK J	russian	1.4617	4	9.8E+03	82.7	28.2	0.18415	0.4985	0.21154	1.0334
UK J	russian	1.3988	2	2.0E+04	82.7	28.2	0.20325	0.4978	0.22879	1.0108

The above spreadsheet is based on measurements of the diodes shown above. For the determination of Is and n, I chose at random two examples from my collection of various diodes, (all silicon in this case) and measured with a modified methodology to that outlined by Ben Tongue and Mike Tuggle. In this case, with the radically different forward voltage drop of these diodes from Germanium or Schottky diodes, I have kept the Id values constant (about Id2=0.5uA and Id1=1.0uA) and varied Vd. I have included the actual room temperature in the calculations in order to eliminate this variable as a source of doubt or error. All the measured parameters as well as the determined results are presented in order to facilate repeatability.

Light Emitting Diodes

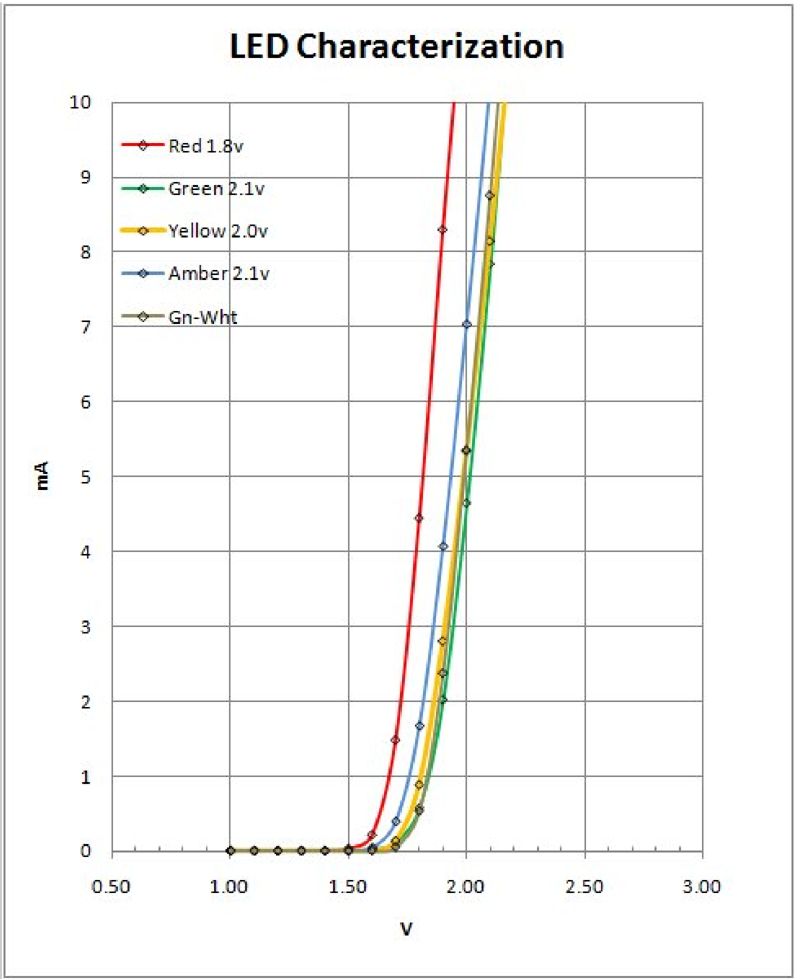
LED Characterization
10
9
8
7
6
5
4
3
2
1
0
mA
Red 1.8v
Green 2.1v
Yellow 2.0v
Amber 2.1v
Gn-Wht
0.50
1.00
1.50
2.00
2.50
3.00
V

Red 1.8v

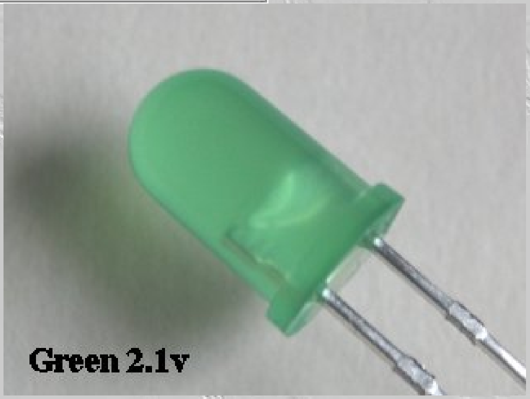
Green 2.1v

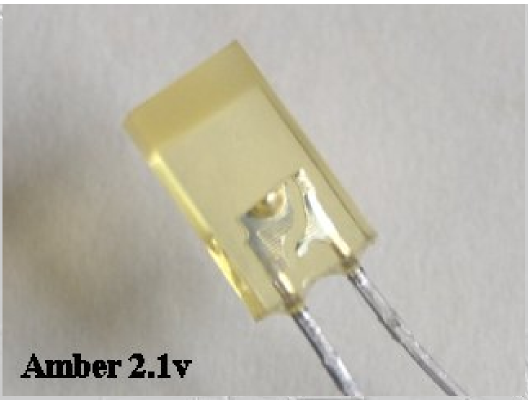

As long as I am measuring various and sundry diodes, I figure I ought to include that most ubiquitous of modern diode, the LED. Found everywhere, these diodes are rapidly becoming the low-energy light source of choice for many lighting applications. I have read occasionally of someone asking whether these ought to be useful for radio applications as well. To this question the answer is generally a resounding "No!". The turn-on voltage is waaay above any reasonable value expected to be delivered by an antenna to a crystal set. Still, this simple answer avoids the actual question, what in fact does the characteristic curve of a LED really look like? Where is the turn-on voltage with respect to the published junction voltage, (assuming you can find that).

To provide just such a look I visited my local electronics store and bought a small handful of LED's, most with the junction voltage listed and took them home to measure. Typical LED junction voltages seem to range from about 1.8v to 2.1v or more. The turn-on voltages look closer to 1.6v-1.7v. Anyone used to working with Carborundum crystals or silicon diodes will be used to biasing their rectifier to get good sensitivity. These LED's once on have a very sharp rise and with a proper bias should work quite well as detector diodes. As a bonus you will get a sweet glow as well. OK, not as cool as the glow of a vacuum diode, but certainly more sensitive!

Spreadsheet for calculation of Light Emitting Diode n and Is (modified from Mike Tuggle's spreadsheet)

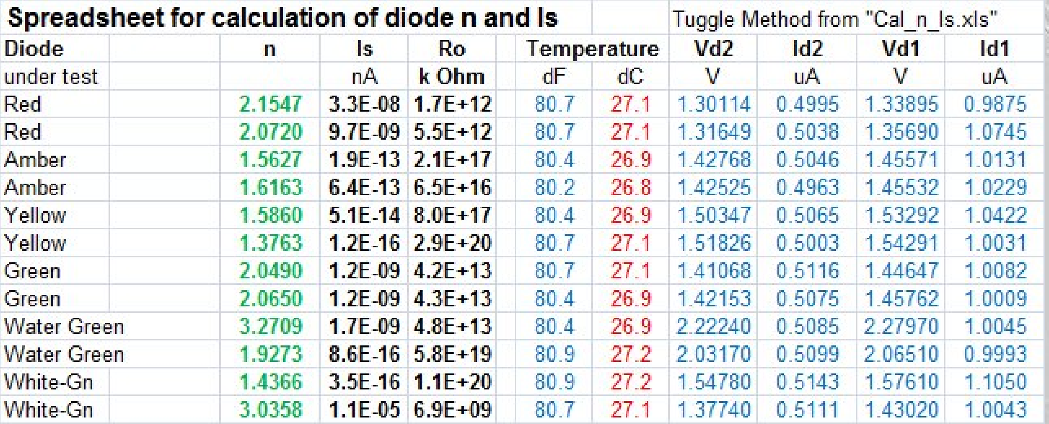

Spreadsheet for calculation of diode n and Is Tuggle Method from "Cal_n_Is.xls"

Diode	n	Is	Ro	Temperature		Vd2	Id2	Vd1	Id1
under test		nA	k Ohm	dF	dC	V	uA	V	uA
Red	2.1547	3.3E-08	1.7E+12	80.7	27.1	1.30114	0.4995	1.33895	0.9875
Red	2.0720	9.7E-09	5.5E+12	80.7	27.1	1.31649	0.5038	1.35690	1.0745
Amber	1.5627	1.9E-13	2.1E+17	80.4	26.9	1.42768	0.5046	1.45571	1.0131
Amber	1.6163	6.4E-13	6.5E+16	80.2	26.8	1.42525	0.4963	1.45532	1.0229
Yellow	1.5860	5.1E-14	8.0E+17	80.4	26.9	1.50347	0.5065	1.53292	1.0422
Yellow	1.3763	1.2E-16	2.9E+20	80.7	27.1	1.51826	0.5003	1.54291	1.0031
Green	2.0490	1.2E-09	4.2E+13	80.7	27.1	1.41068	0.5116	1.44647	1.0082
Green	2.0650	1.2E-09	4.3E+13	80.4	26.9	1.42153	0.5075	1.45762	1.0009
Water Green	3.2709	1.7E-09	4.8E+13	80.4	26.9	2.22240	0.5085	2.27970	1.0045
Water Green	1.9273	8.6E-16	5.8E+19	80.9	27.2	2.03170	0.5099	2.06510	0.9993
White-Gn	1.4366	3.5E-16	1.1E+20	80.9	27.2	1.54780	0.5143	1.57610	1.1050
White-Gn	3.0358	1.1E-05	6.9E+09	80.7	27.1	1.37740	0.5111	1.43020	1.0043

The above spreadsheet is based on measurements of the diodes shown above. For the determination of Is and n, I chose at random two examples from my collection of various diodes, (all led's in this case) and measured with a modified methodology to that outlined by Ben Tongue and Mike Tuggle. In this case, with the radically different forward voltage drop of these diodes from Germanium or Schottky diodes, I have kept the Id values constant (about Id2=0.5uA and Id1=1.0uA) and varied Vd. I have included the actual room temperature in the calculations in order to eliminate this variable as a source of doubt or error. All the measured parameters as well as the determined results are presented in order to facilitate repeatability.

Crystal Galena and Pyrite Diodes

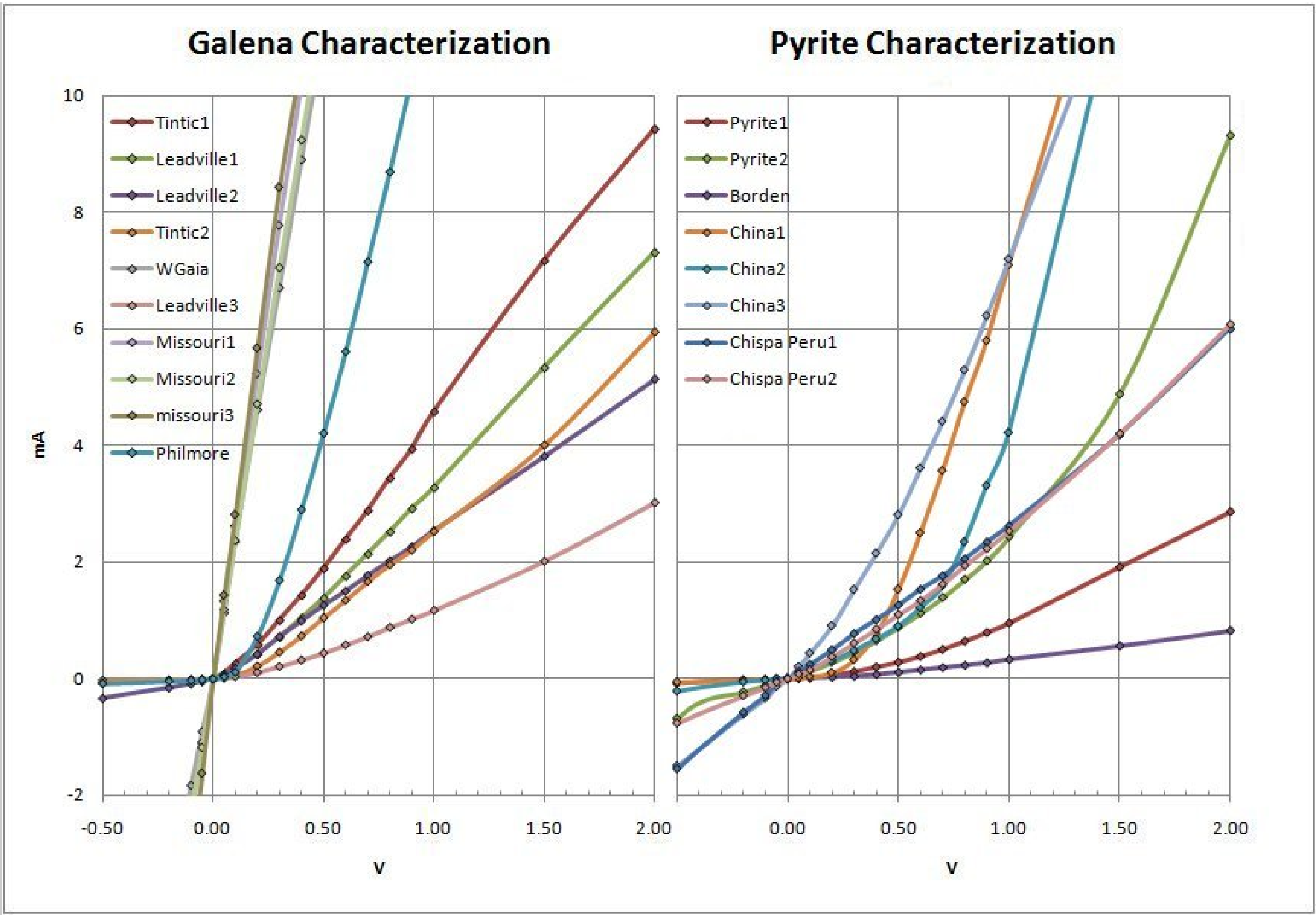

The above curves demonstrate the wide variation in properties and qualities that can be found in natural crystals of galena or pyrite, the two most common and best quality natural stones. In each crystal test I first poked around the crystal for some time to 1) determine a typical sensitivity for the crystal in question and 2) to locate the best hot spot with which to test. This turns out to be a non-trivial exercise on the diode test setup. In a crystal radio one need only listen for the loudest spot. With the test setup one needs test both the forward and reverse current in order to determine if the whisker is on a hot spot or not. Very tedious work! (In retrospect, if I were starting over making a test jig, I would definitely add a DPDT switch to readily change between forward and reverse current measurements. I'd probably toss in a rheostat for fine adjustments as well).

For many crystals there are limited number of possible hot spots, but these may be hot indeed. For most of my "Steel Galena" samples (Tintic Utah, or Leadville Colorado) there are numerous hot spots under virtually every place I touch the probe, but in general the sensitivity is good to so so. These crystals are very kind to work with in terms of finding spots and avoiding frustration. Mirror galena on the other hand may have quality hot spots, but any hot spots at all are rare and frustratingly difficult to locate. Here my Philmore detector crystal shines with an almost ideal "Galena" response. To chase down this rabbit I purchased some lovely mirror galena from Sweetwater Missouri. I broke off a few appropriate-size chunks to pot in woods metal and test. At first I was very excited with the high currents I was seeing at moderate voltages. Figured I had struck gold. When these crystals failed miserably to rectify anything in my radios, I re-measured things in both forward and reverse directions. These crystals obey Ohm's law and act like typical resistors, not suitable for radio work at all.

For my pyrite crystals the work has been especially tedious and frustrating. With one of the crystals one "hot spot" alternated, entirely on its own, between hot and bad while I was making the measurements. I would start over and over, sometimes getting interesting readings then suddenly it would drop to low values and I'd start over, back and forth. I present this data as best as I have measured, and I don't intend to go back! You see at least one of the crystals, my "China 1", (from a lead/silver mine in Hunan) gave a sweet classic-looking curve. More to the point, "ideal" curves for natural minerals are difficult to come by. Most crystals you use will be less than ideal. The good news is that, while listening to your set, poking about for a good spot is far easier than what I have gone through to produce these curves. Your ear will take care of you!

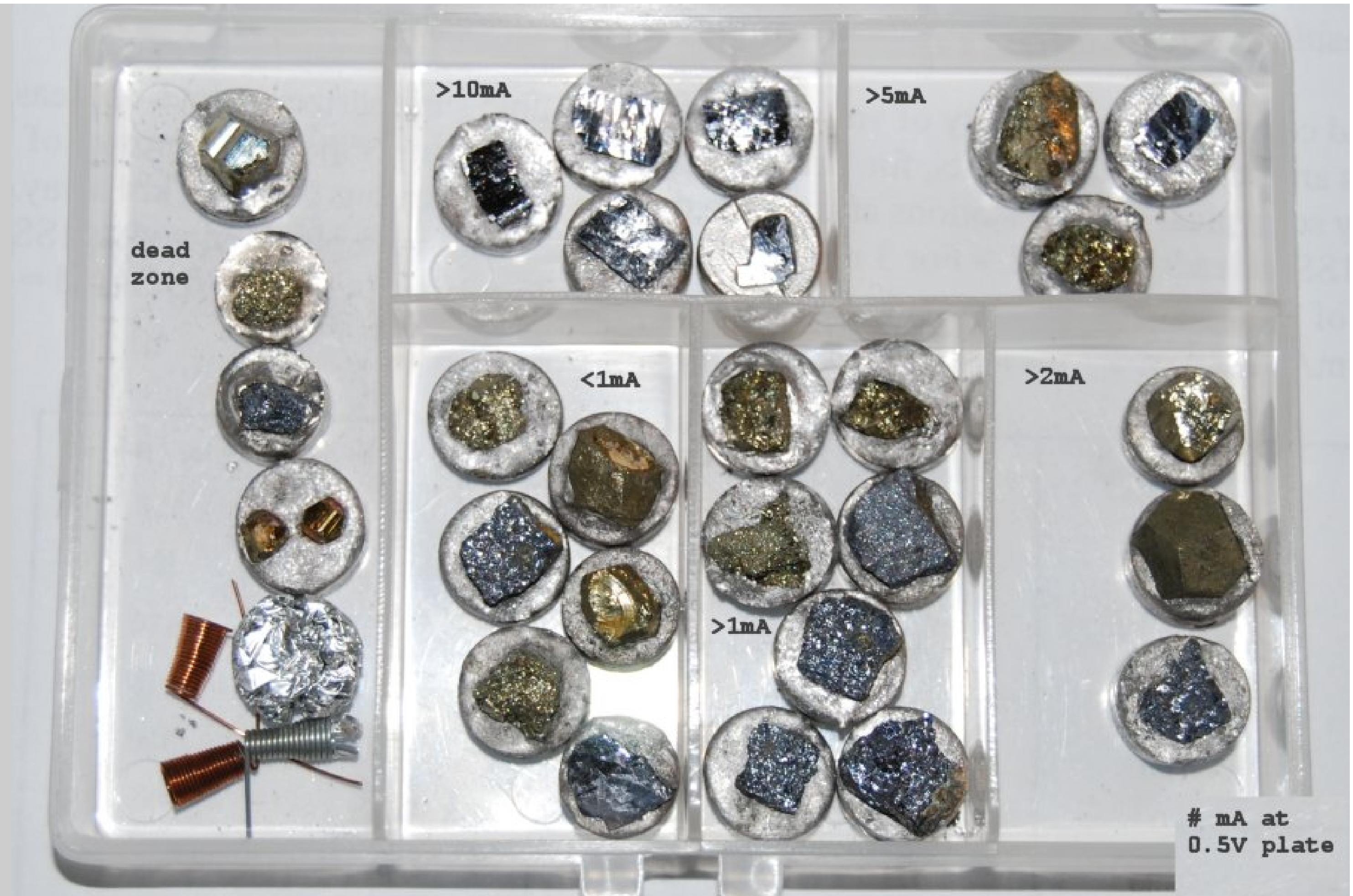

In the photo I indicate groupings based on an easy measure of performance. I note the current in milliamps for each set where the plate voltage is set at 0.5V. The greater the current the better are your chances to get a sensitive crystal, assuming Ohm's Law is not followed! Crystals in the "dead zone" on the left will have their woods metal re-melted for new detector crystals and the bad ones tossed, its tough love for crystals. I find that easily half the potted crystals I make are tossed this way, and only a few can be considered superlative.

My measurements and discussion on vacuum diodes has moved to my section "Diodes III" under the Lab Menu.

Return..

Back to Main....

Sunday, 09-Jul-2017 00:50:33 CDT

Comments and/or Correspondance to: kj-smith@sbcglobal.net

.::. Sitemap .::. International Pix .::. Western US Pix .::. Pinhole Pix .::. Crystal Radio .::.
.::. Snapshots .::. Thoughts .::. Beer Labels .::. Petrology .::. Life List .::. Ionosphere.::.

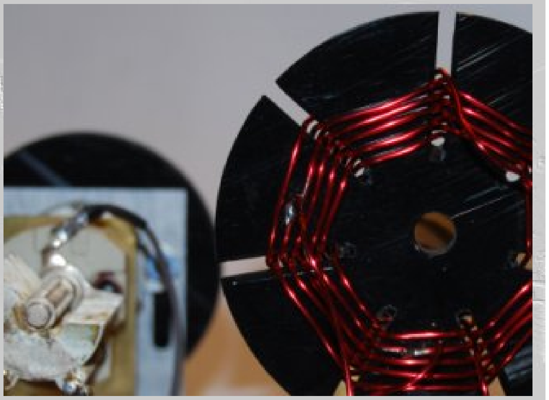

Global DX'ter

Back to Radio..

Finally, a shortwave set comes off the design bureau and out into the world. This is a type of set I have been planning and pondering on for some time. With many crystal radio design possibilites this will certainly not be the only SW set I plan, but it is my first effort, my chance to "do it" on shortwave. For a first effort I settled on a fairly simple design, a single-tuned rig with some degree of control, but not overly much. I took as my base a design by Rainer Steinfuehr on his Crystal Reciever World site. This design features a high Q spider coil with two taps, and adjustable capacitance on both the antenna and tuning circuits. It made a rather lovely small set indeed.

The specifics of the design were entirely my own and were based on the frequency range I wished to target and the specifications of the materials I had available for the project. In partictular, I had available small good-quality 10 - 115 pF variable caps for which my clever elder brother had lathed lovely 1/4 inch shaft extenders for use with standard tuning knobs. My target spectrum was about 5 to 18 MHZ as that is where the dominant international BC action takes place. With luck that would also allow receiving WWV at 10 and 15MHz as well. Digging into the design spreadsheets, it seemed an inductor with 7 to 9 mH would best deliver the goods. I made a spider form from a circlar cut of 1/8 inch Delrin plastic and wound 11 turns of 18awg magnet wire, carefully soldering taps at 3 turns and 8 turns from the ground (center) of the coil. I ended with 9 turns on the final set with testing giving me a 7.6 mH coil. The final schematic presented below.

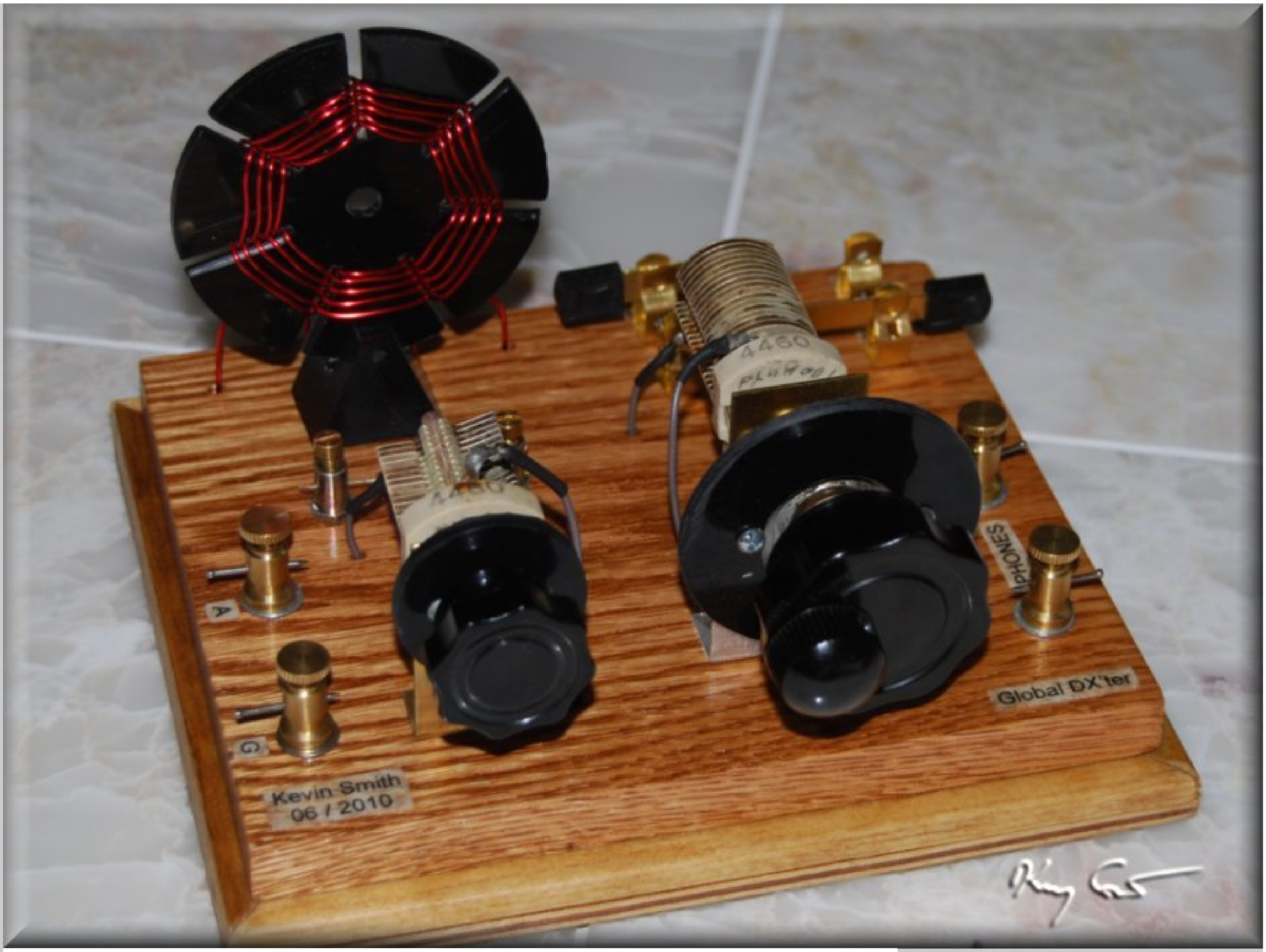
4460
A
G
PHONES
Global DX'ter
Kevin Smith
06 / 2010

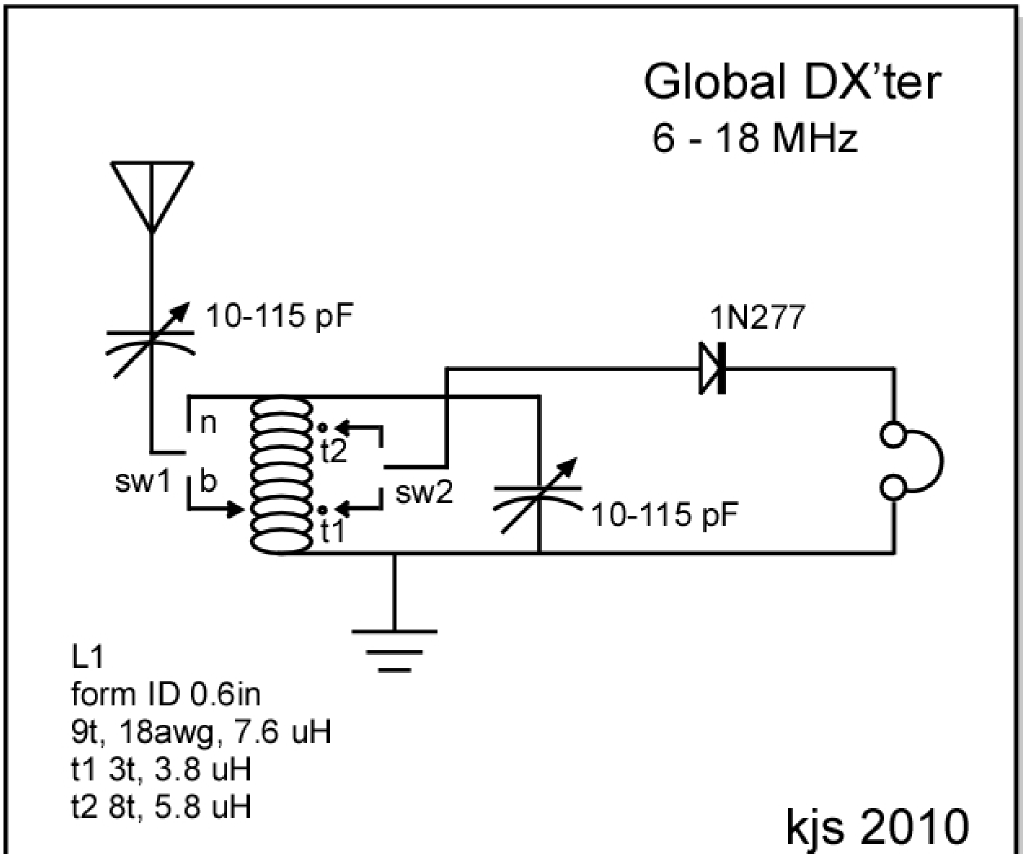
Global DX'ter
6 - 18 MHz
10-115 pF
1N277
n
t2
sw1
b
sw2
10-115 pF
t1
L1
form ID 0.6in
9t, 18awg, 7.6 uH
t1 3t, 3.8 uH
t2 8t, 5.8 uH
kjs 2010

Initial testing on the set was dismal. NO RECEPTION. There followed many hours of double checking all wireing, solder joints, fidelity to the schematic, etc. Everything checked out, no good. Time to seek professional help, which happily corresponded to my annual visit home. There my radio engineer brother took the radio and gave it his close attention, puzzled over it as had I, and concluded a defective part somewhere. This he soon found in the tuning capacitor which had had a connection slip off its set point. This easily corrected, the radio seemed ready for testing. Unfortunately he did not have an antenna handy (he works in uhf+ research stuff) so all I could do was hook the set to a signal generator and test its performance, summarized below.

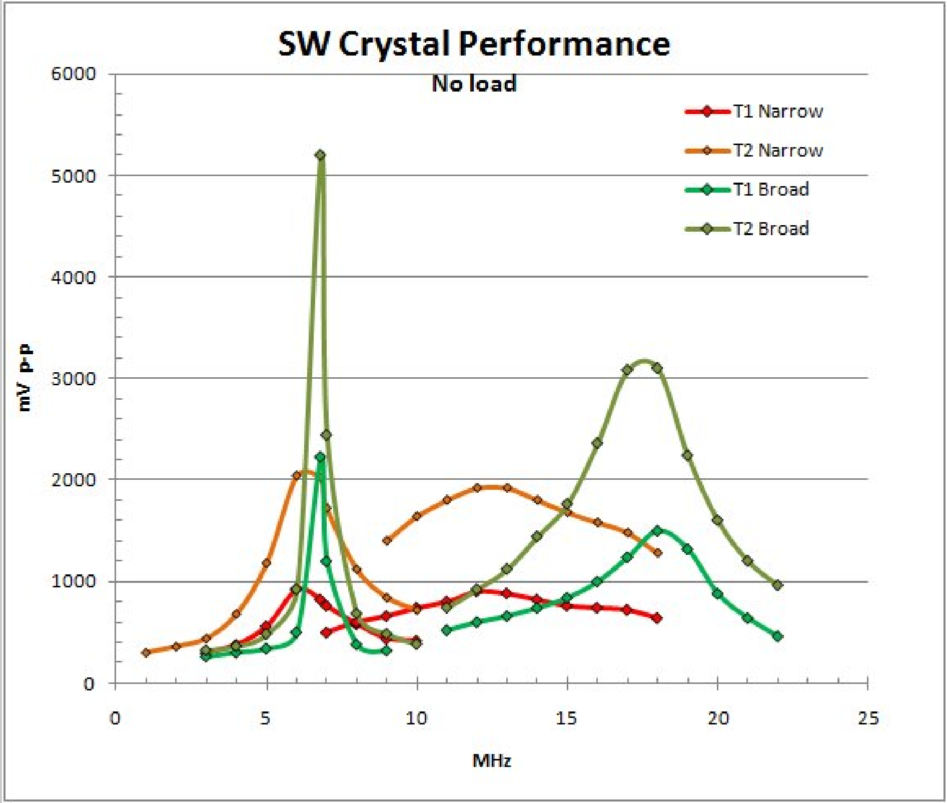

The plot above charts the received signal strength (p-p mV) versus input frequency for the four possible switch configurations (Detector on tap 1 or 2, Antenna on selective [top of coil] or broad [on tap t1]. These four conditions were tested with the tuning cap fully meshed (low frequency) and fully open (high frequency) to give an idea of the actual tuning range as compared to the calculated range. All measurements were made with no load on the output, and the antenna trim cap in the full meshed position. Other conditions not shown were also tested.

The results indicate that placing the antenna low on the coil, on the T1 position gives the best sensitivity and selectivity in tuning. The detector, a 1N277 germanium diode placed high on the coil at T2 also gives dramatic increase in sensitivity. What I found then, was a farily sensitive but, to be truthful, woefully un-selective set tuning from 6 to about 17-18 MHz. Still, not bad for a single-tuned shortwave rig.

Final thoughts. Upon returning to my home base, I quickly hooked the set to my longwire to make first impressions of actual reception. In midday I was able to pick up a spanish language format broadcast and two english language religous format stations and a brief encounter with WWV. At night the bands are flooded with competing stations, success! Note that all this with my handy dandy Radio Shack mini-amplifier. When I connected a crystal earphone and used my poor hearing, there was nothing as expected. One pity is the ongoing presence, even on this set, of KGOW 1.65MHz forever audible in the background. I consider the set a success as I was not expecting selectivity in a single-tuned rig. Still, with the bands dominated by religous and spanish broadcasts, I cant say it differs much from the local Houston AM fare. Oh well...

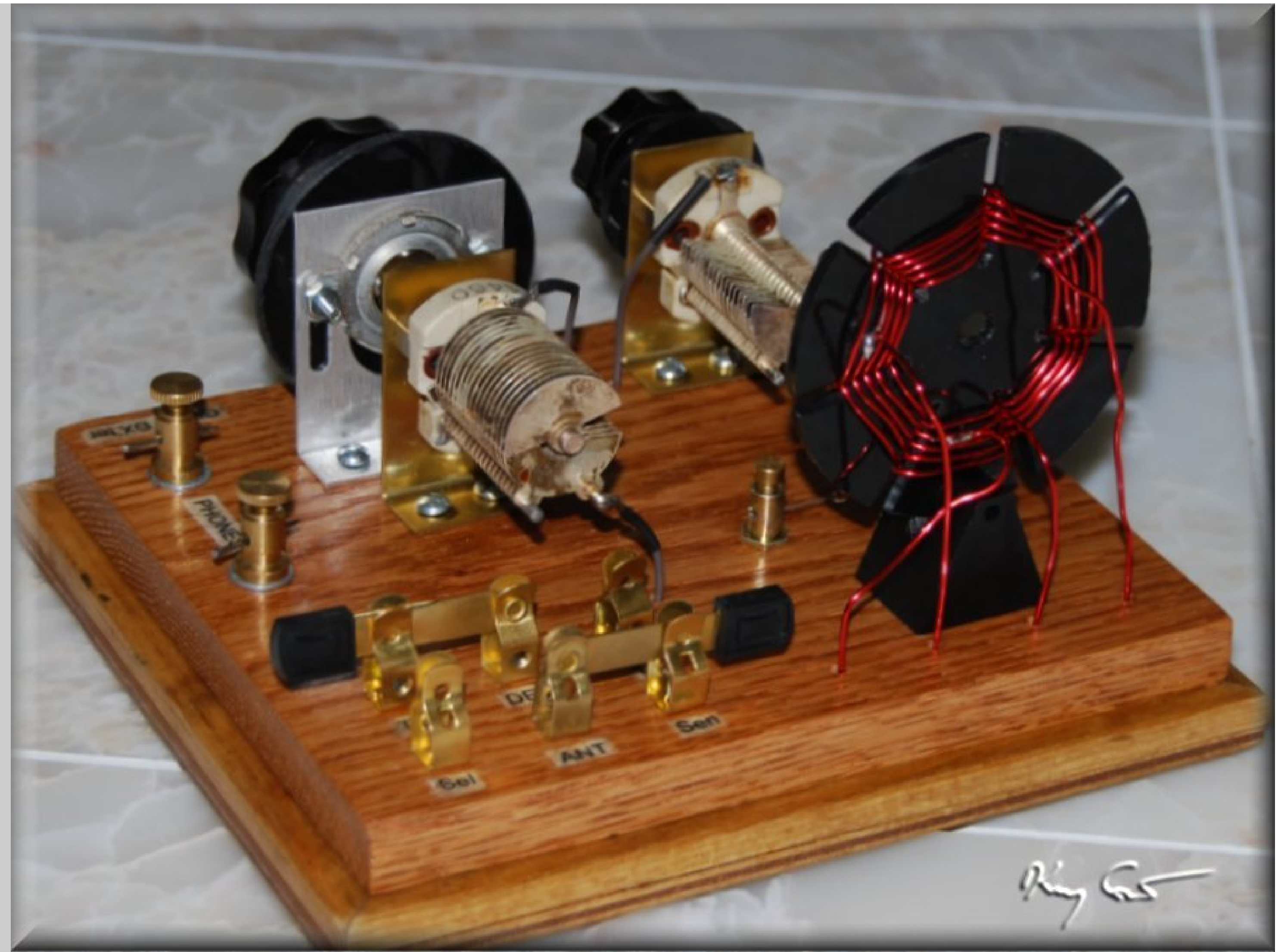
PHONE
Sel
ANT
Sen

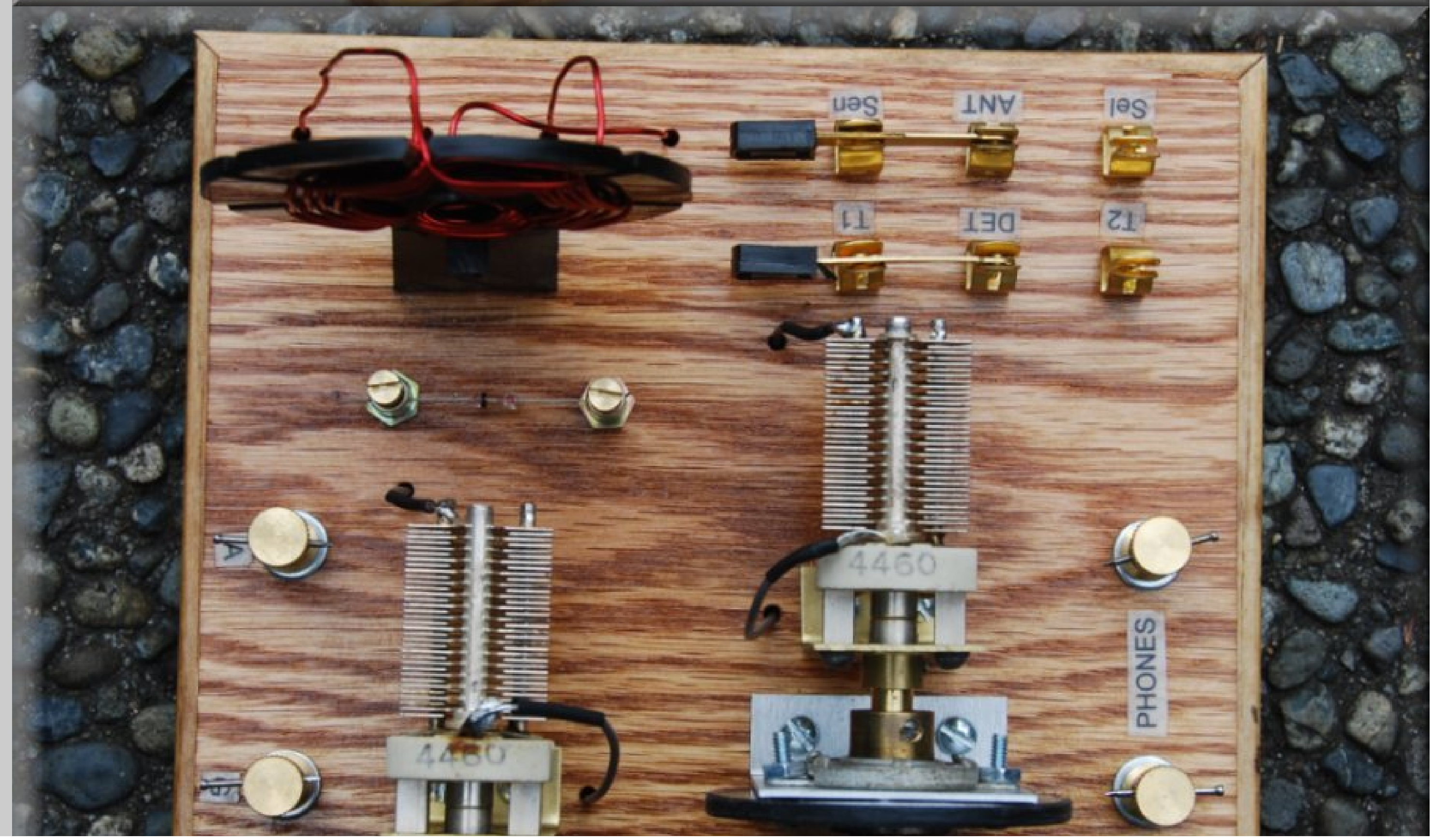
Sen
ANT
Sel
T1
DET
T2
A
4460
4460
PHONES

Kevin Smith
06 / 2010
Global DX'ter

[Return..](#)

[Back to Main....](#)

Monday, 19-Mar-2018 22:40:58 CDT

Comments and/or Correspondance to: kj-smith@sbcglobal.net

.::. Alsace .::. Life List .::. Ionosphere .::. Crystal Radio .::. Snapshots .::. Thoughts .::.
.::. Sitemap .::. Beer Labels .::. International Pix .::. Western US Pix .::. Pinhole Pix .::. Petrology .::.

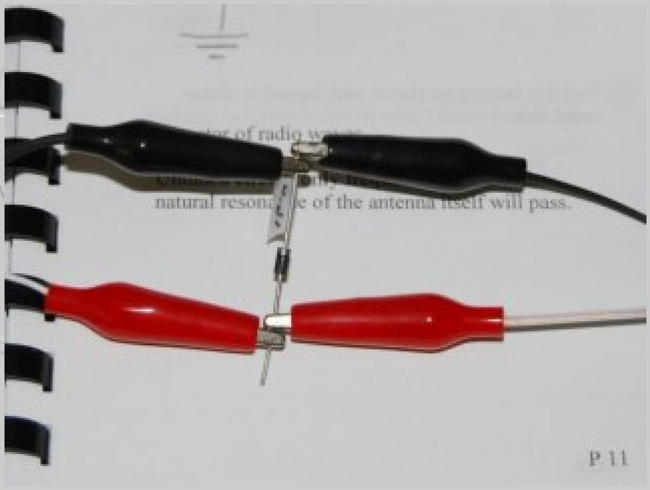

Sleeper #1

Back to Radio..

OK, this is a bit of a cheap shot for a radio, but.. it IS a radio!

It is always interesting to get down to bare essentials, to do the absolute minimum and still get a working instrument. It is not for nothing that MB Sleeper included a circuit diagram for this hookup, it is a functionning radio. This circuit is essentially a detector of radio broadcast energy. It has no tuneing functionality. Resonance is extremely broad although weakly peaking as determined by the impedence and capacitance of the antenna itself. In part my idea is to see if it is possible to determine this peak region.
I present the set photo with a 1N5819 Schottky diode but it works well with virtually any diode, only the sensitivity changes. I even tried a Zenner diode but had to attach the Radio Shack mini-amp and crank it up. Still, it worked ok. So, more to come..

As shown, the radio passes most loudly a local powerhouse spanish format station. It is quite sensitive as one can imagine. Listening closely and you can certainly hear the other stations on the broadcast band competing for attention. My question now, do I make a breadboard extravaganza of this set? On va voir...

Sleeper #1

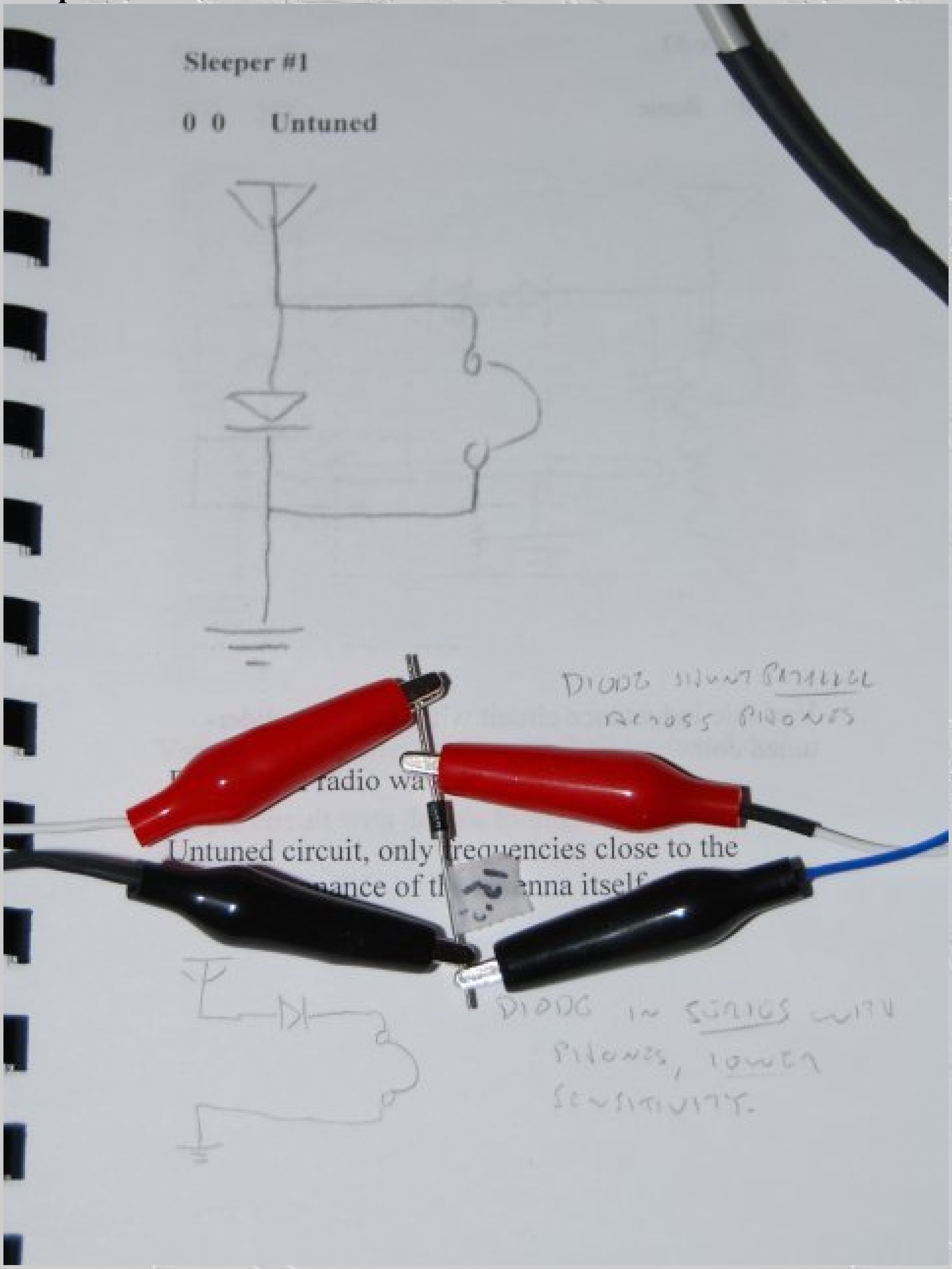
Sleeper #1
0 0 Untuned
radio wa
Untuned circuit, only frequencies close to the
nance of th enna itself
DIODE SHUNT PARALLEL ACROSS PHONES
DIODE IN SERIES WITH PHONES, LOWER SENSITIVITY.

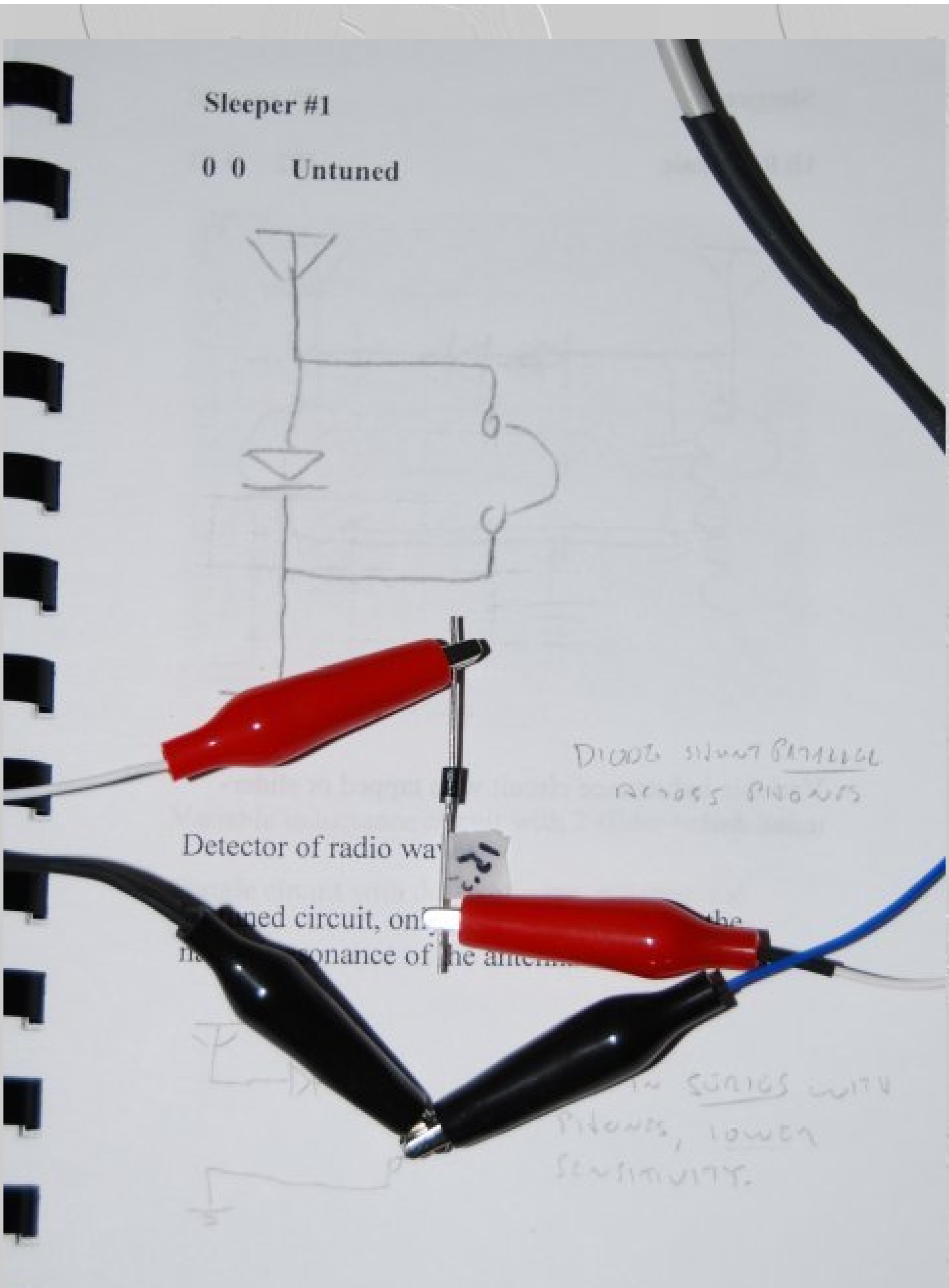
Sleeper #1
0 0 Untuned
DIODE SHUNT PARALLEL ACROSS PHONES
IN SERIES WITH PHONES, LOWER SENSITIVITY.

The hookup in its parallel and series configuragions. In either case I really cannot hear any signal with just a headphone or crystal earplug. With amplification I can get good signal strength with the parallel hookup giving slightly more sensitivity than the series hookup.

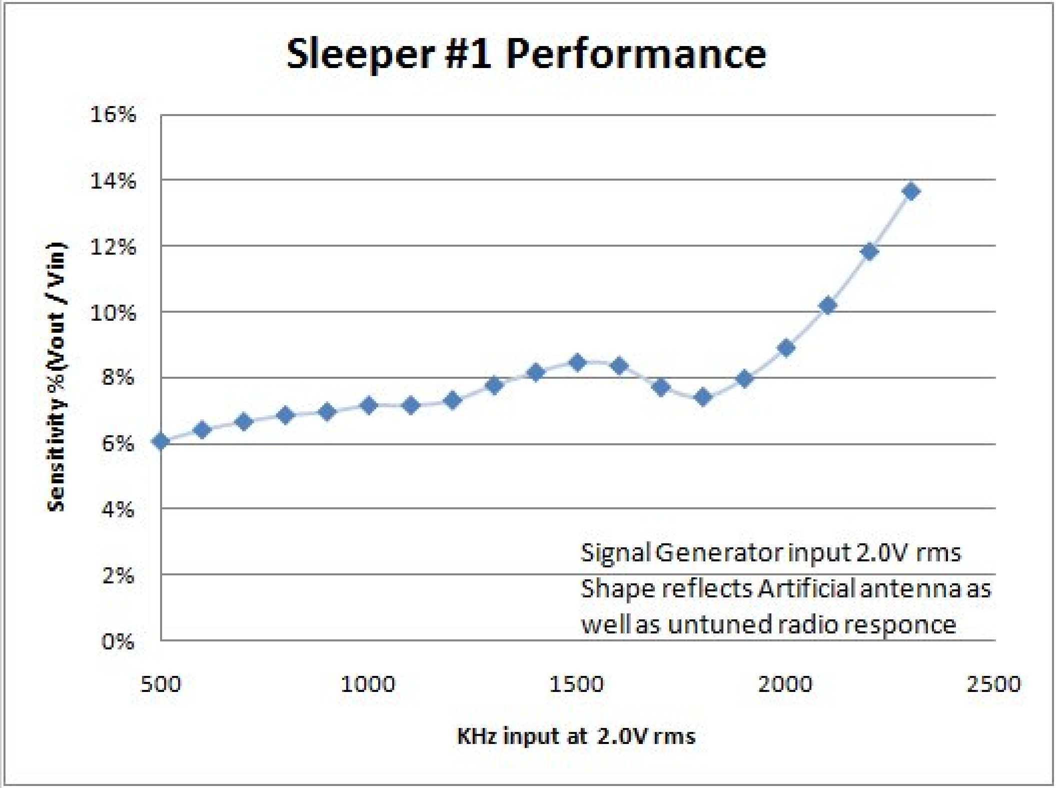

Hooking parallel configuration to my signal generator can give an idea of the performance of the radio across the spectrum. I measured the mV output in 100 KHz increments from 500 to 2300 KHz. Across the broadcast band the set has faitly flat responce with a slight tendancy to resonance about 1500 KHz and then increasing sensitivity from 1900 KHz on up. Keep in mind that the set is hooked to a dummy antenna rather than my actual long wire. Still this gives an idea of the overall untuned nature of the circuit.

When I hook in a series configuraiton I am unable to get a measurement, only 0V DC output. I am working on why this may be the case. Drives me nuts.

Return..

Back to Main....

Sunday, 09-Jul-2017 01:04:13 CDT

Comments and/or Correspondance to: kj-smith@sbcglobal.net

.::. Sitemap .::. International Pix .::. Western US Pix .::. Pinhole Pix .::. Crystal Radio .::.
.::. Snapshots .::. Thoughts .::. Beer Labels .::. Petrology .::. Life List .::. Ionosphere.::.

Wavetrap, a useful gadget

[Back to Radio..](#)

This little gadget is most handy for crunching those local behemoths that always seem to be directing their strongest signals right in your direction. My wave trap is a bit of a combination of 1) Osterhoudt's QRM coil, 2) a design I found online on Owen Pool's excellent web site, and 3) a bit of my own seeing as how I had some, but not all of the ingredients needed from both traps. That is, I used what I found in the junk box.

The trap is formed on a stiff cardboard paper core with 1 1/4 inch outside diameter. On this I drilled holes for winding the tank coil, 110 turns of #30 awg enameled wire. Both Poole and Osterhoudt recommend #32, but 30 is what I had. Over this I would a coupling coil, 15 turns space wound of #22 awg enameled wire. Tuning of the tank is completed with a 410 pF var cap.

Tuning is sharp and deep, it takes a bit of getting used to, but then, schezzam!!!.. problem station gone! I luv it.

Wave Trap

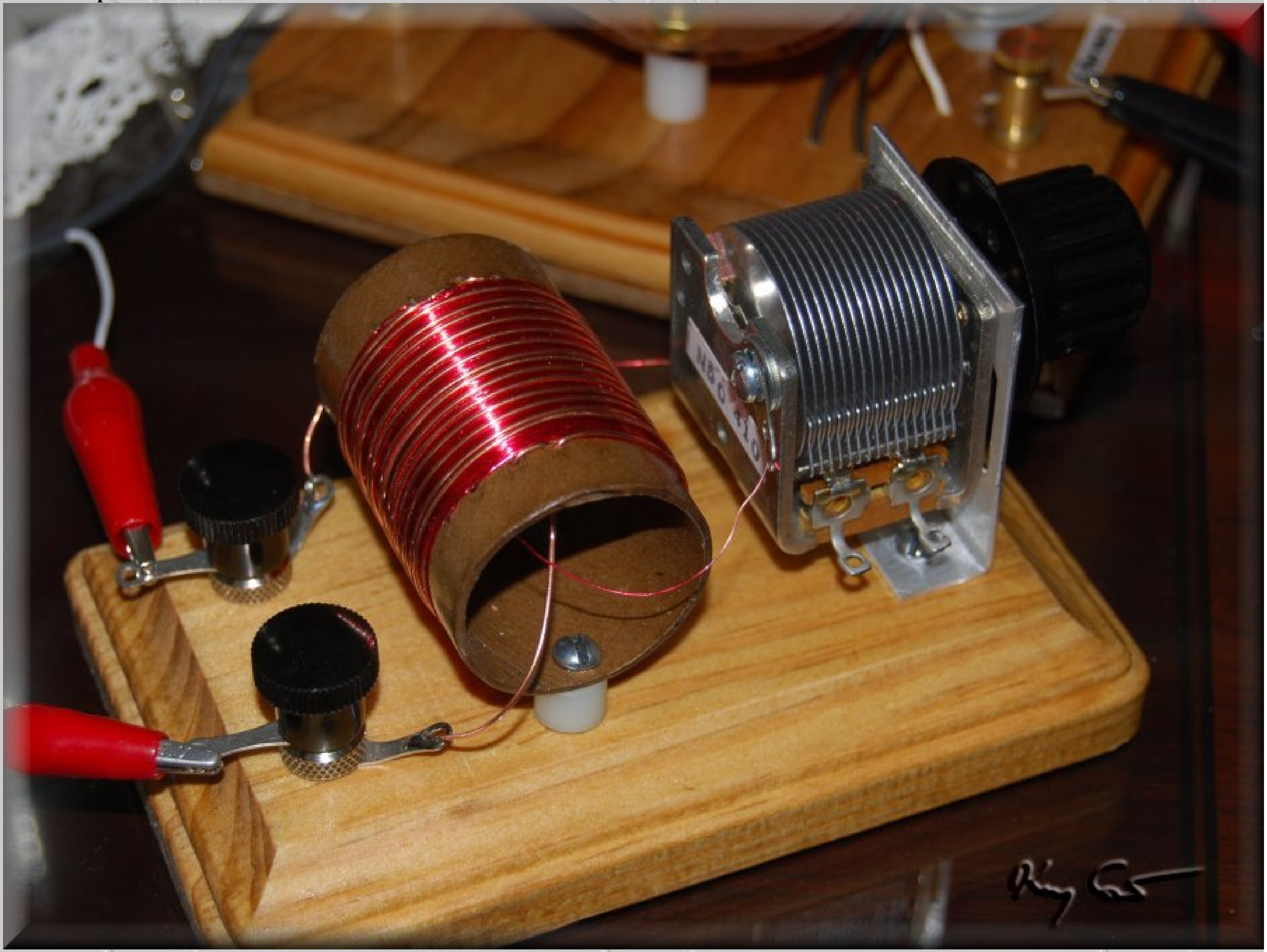

Return..

Back to Main....

Sunday, 09-Jul-2017 01:03:30 CDT

Comments and/or Correspondance to: kj-smith@sbcglobal.net

.::. Sitemap .::. International Pix .::. Western US Pix .::. Pinhole Pix .::. Crystal Radio .::.
.::. Snapshots .::. Thoughts .::. Beer Labels .::. Petrology .::. Life List .::. Ionosphere.::.

United States Patent [19] [11] **4,292,804**

Rogers, Sr. [45] **Oct. 6, 1981**

[54] METHOD AND APPARATUS FOR OPERATING AN ENGINE ON COMPRESSED GAS

[76] Inventor: **Leroy K. Rogers, Sr.**, #5 Capistrano Ct., Ft. Myers, Fla. 33908

[21] Appl. No.: **158,303**

[22] Filed: **Jun. 10, 1980**

[51] **Int. Cl.**^{3} **F15B 11/06**

[52] **U.S. Cl.** **60/407**; 91/187; 91/275

[58] **Field of Search** 60/407, 412; 91/187, 91/275, 364

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,881,399	5/1975	Sagi et al.	91/187 X
3,885,387	5/1975	Simington	60/407 X
4,018,050	4/1977	Murphy	60/412 X

Primary Examiner—Allen M. Ostrager
Attorney, Agent, or Firm—Burns, Doane, Swecker & Mathis

[57] ABSTRACT

The present invention relates to a method and apparatus for operating an engine having a cylinder and a piston reciprocable therein on compressed gas. The apparatus comprises a source of compressed gas connected to a distributor which distributes the compressed gas to the cylinder. A valve is provided to selectively admit compressed gas to the cylinder when the piston is in an approximately top dead center position. In one embodiment of the present invention the timing of the opening of the valve is advanced such that the compressed gas is admitted to the cylinder progressively further before the top dead center position of the piston as the speed of the engine increases. In a further embodiment of the present invention a valve actuator is provided which increases the length of time over which the valve remains open to admit compressed gas to the cylinder as the speed of the engine increases. A still further embodiment of the present invention relates to an apparatus for adapting a conventional internal combustion engine for operation on compressed gas.

22 Claims, 8 Drawing Figures

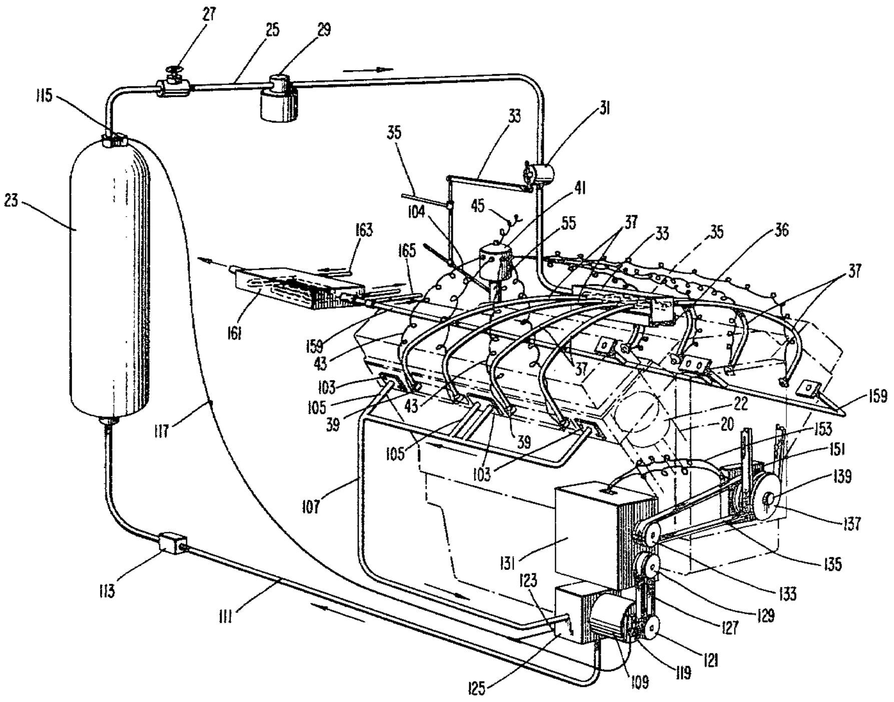

FIG. 1

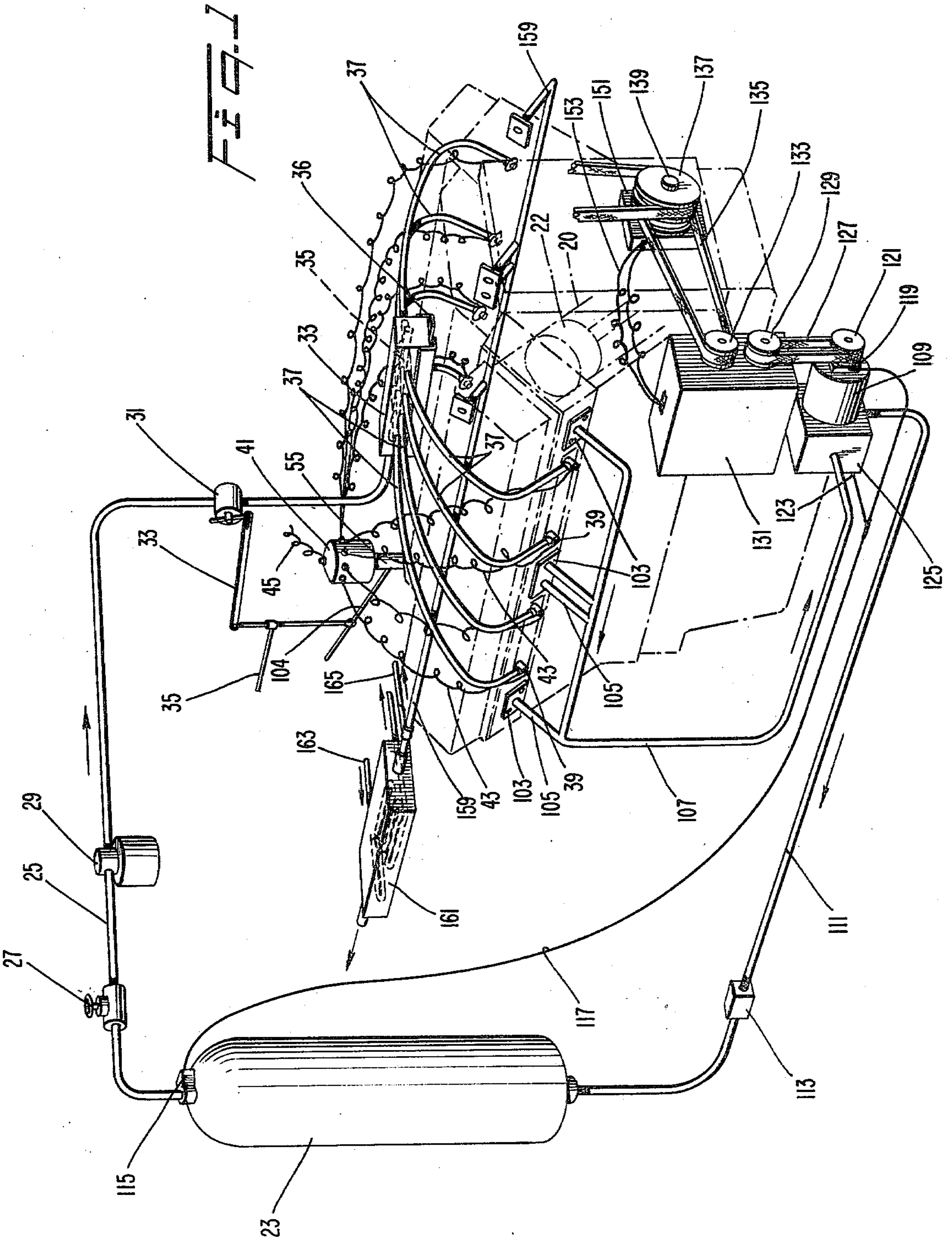

FIG. 6

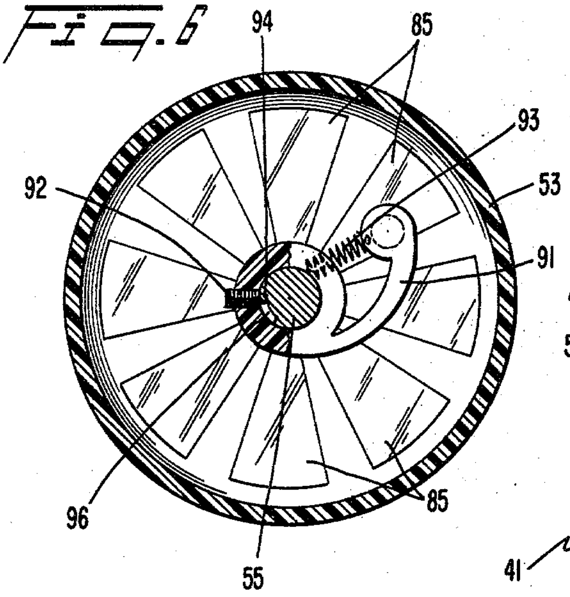
94
85
92
93
53
91
96
55
85

FIG. 7

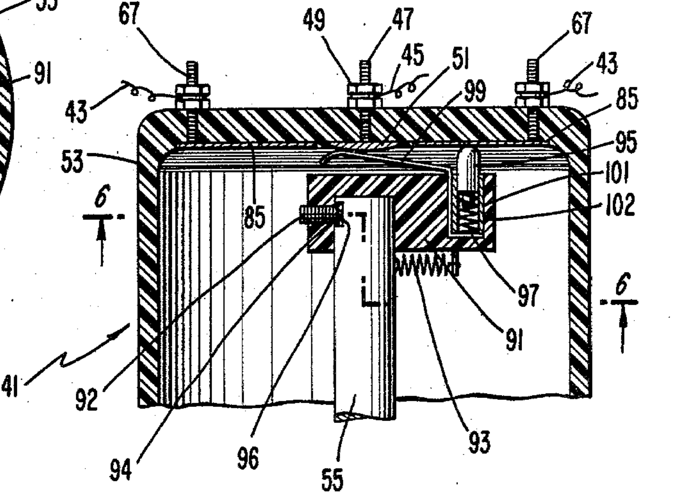
67
49
47
45
51
67
43
99
43
85
95
53
101
102
6
85
6
97
91
41
92
94
96
55
93

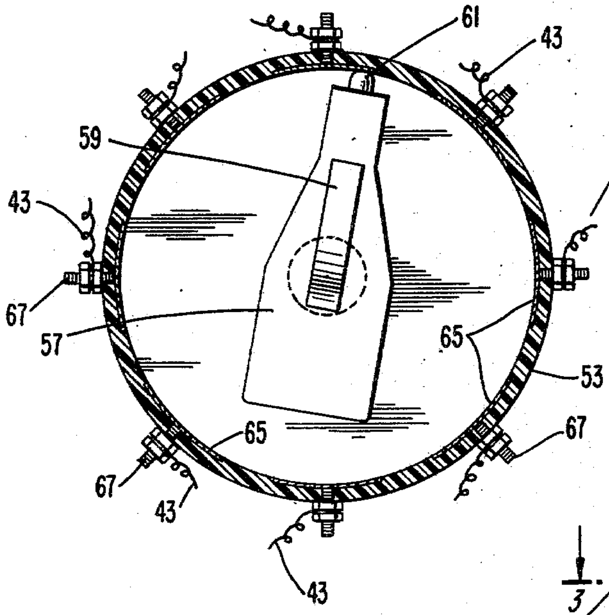
61
43
59
43
67
57
65
53
65
67
67
43
43

FIG. 3

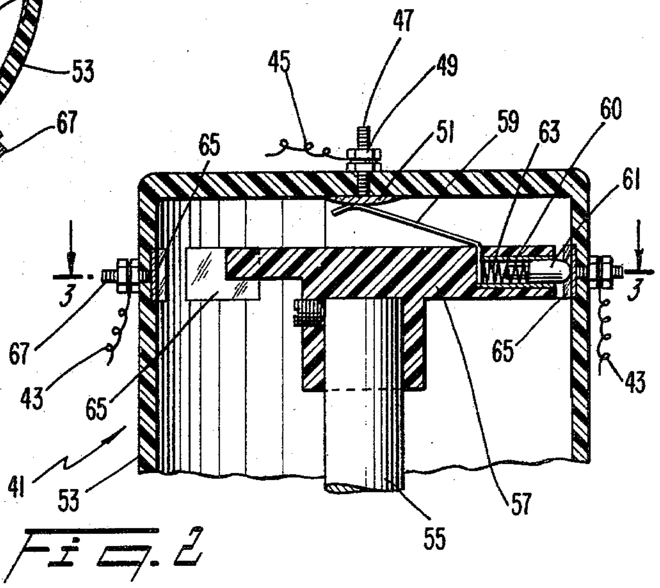
47
45
49
51
59
63
60
65
61
3
3
67
43
65
65
43
41
53
55
57

FIG. 2

Fig. 5

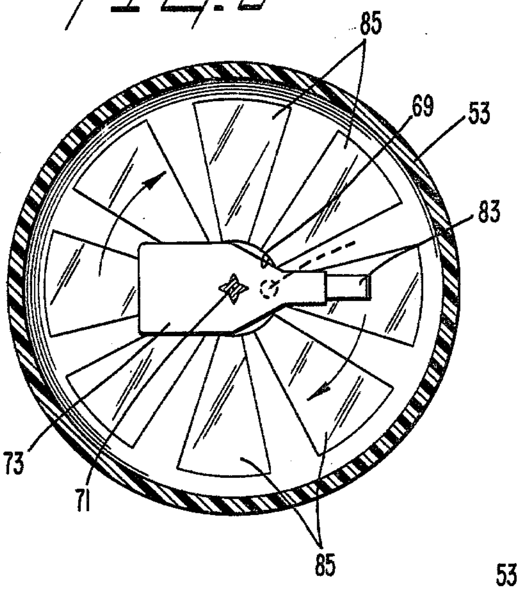
85
69
53
83
73
71
85

Fig. 4

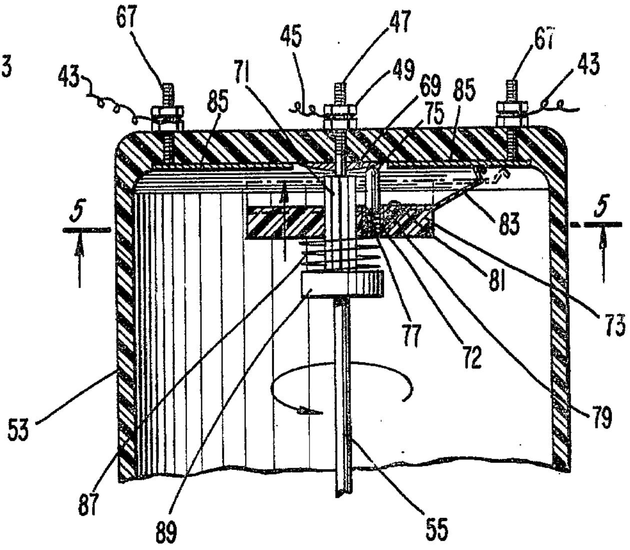
67
45
47
67
43
71
49
69
85
43
85
75
5
83
5
81
73
77
72
53
79
87
89
55

Fig. 6

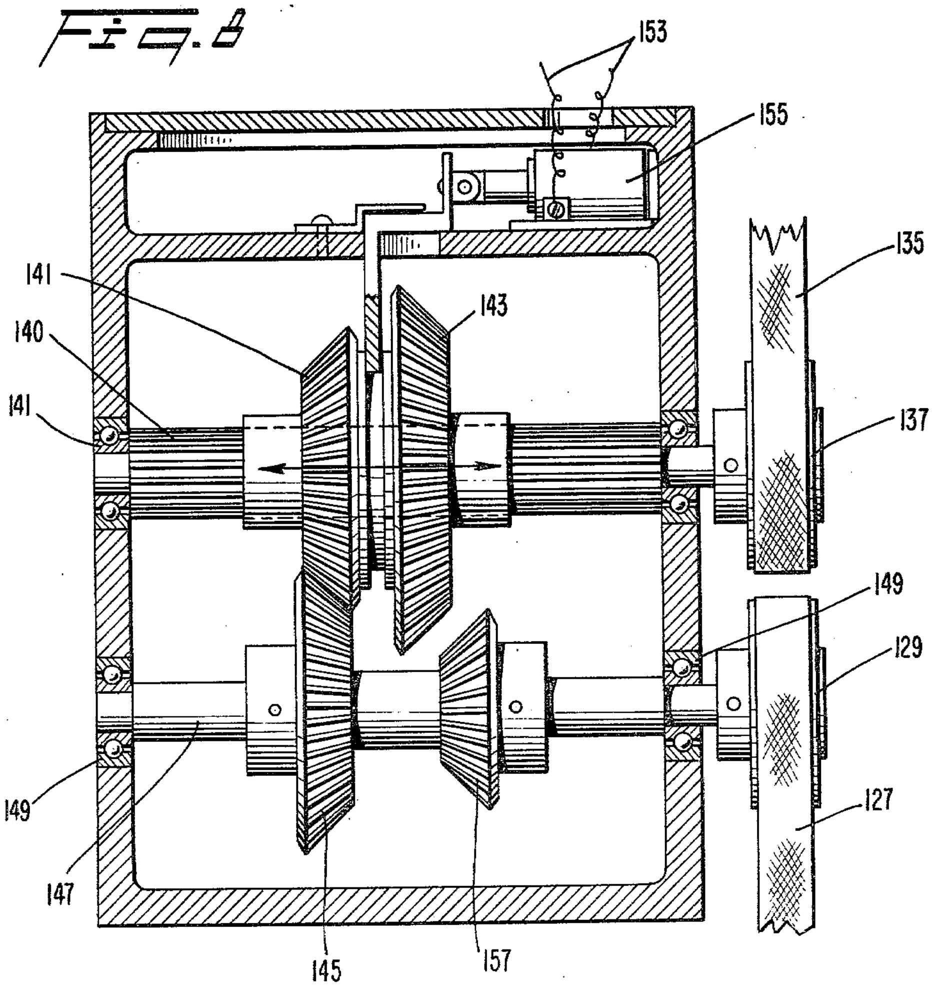
153
155
141
143
135
140
141
137
149
129
149
127
147
145
157

METHOD AND APPARATUS FOR OPERATING AN ENGINE ON COMPRESSED GAS

BACKGROUND AND SUMMARY OF THE PRESENT INVENTION

The present invention relates to a method and apparatus for operating an engine using a compressed gas as the motive fluid. More particularly, the present invention relates to a apparatus for adapting a pre-existing internal combustion engine for operation on a compressed gas.

Air pollution is one of the most serious problems facing the world today. One of the major contributors to air pollution is ordinary internal combustion engine which are used in most motor vehicles today. Various devices, including many items mandated by legislation, have been proposed in an attempt to limit the pollutants which an internal combustion engine exhausts to the air. However, most of these devices have met with limited success and are often both prohibitively expensive and complex. A clean alternative to the internal combustion engine is needed to power vehicles and other machinery.

A compressed gas, preferably air, would provide an ideal motive fluid for a engine since it would eliminate the usual pollutants exhausted from an internal combustion engine. An apparatus for converting an internal combustion engine for operation on compressed air is disclosed in U.S. Pat. No. 3,885,387 issued May 27, 1975 to Simington. The Simington patent discloses an apparatus including a source of compressed air and a rotating valve actuator which opens and closes a plurality of mechanical poppet valves. The valves deliver compressed air in timed sequence to the cylinders of an engine through adapters located in the spark plug holes. However, the output speed of an engine of this type is limited by the speed of the mechanical valves and the fact that the length of time over which each of the valves remains open cannot be varied as the speed of the engine increases.

Another apparatus for converting an internal combustion engine for operation on steam or compressed air is disclosed in U.S. Pat. No. 4,102,130 issued July 25, 1978 to Stricklin. The Stricklin patent discloses a device which changes the valve timing of a conventional four stroke engine such that the intake and exhaust valves open once for every revolution of the engine instead of once every other revolution of the engine. A reversing valve is provided which delivers live steam or compressed air to the intake valves and is subsequently reversed to allow the exhaust valves to deliver the expanded steam or air to the atmosphere. A reversing valve of this type however does not provide a reliable apparatus for varying the amount of motive fluid injected into the cylinders when it is desired to increase the speed of the engine. Further, a device of the type disclosed in the Stricklin patent requires the use of multiple reversing valves if the cylinders in a multi-cylinder engine were to be fired sequentially.

Therefore, it is an object of the present invention to provide a reliable method and apparatus for operating an engine or converting an engine for operation with a compressed gas.

A further object of the present invention is to provide a method and apparatus which is effective to deliver a constantly increasing amount of compressed gas to an engine as the speed of the engine increases.

A still further object of the present invention is to provide a method and apparatus which will operate an engine using compressed gas at a speed sufficient to drive a conventional automobile at highway speeds.

It is still a further object of the present invention to provide a method and apparatus which is readily adaptable to a standard internal combustion engine to convert the internal combustion engine for operation with a compressed gas.

Another object of the invention is to provide a method and apparatus which utilizes cool expanded gas, exhausted from a compressed gas engine, to operate an air conditioning unit and/or an oil cooler.

These and other objects are realized by a method and apparatus according to the present invention for operating an engine having at least one cylinder and a reciprricating piston therein using compressed gas as a motive fluid. The apparatus includes a source of compressed gas and a distributor connected with the source of the compressed gas for distributing the compressed gas to the at least one cylinder. A valve is provided for admitting the compressed gas to the cylinder when the piston is in approximately a top dead center position within the cylinder. An exhaust is provided for exhausting the expanded gas from the cylinder as the piston returns to approximately the top dead center position.

In a preferred embodiment of the present invention a device is provided for varying the duration of each engine cycle over which the valve remains open to admit compressed gas to the cylinder dependent upon the speed of the engine. In a further preferred embodiment of the present invention, an apparatus for advancing the timing of the opening of the valve is arranged to admit the compressed gas to the cylinder progressively further before the top dead center position of the piston as the speed of the engine increases.

Further features of the present invention include a valve for controlling the amount of compressed gas admitted to the distributor. Also, a portion of the gas which has been expanded in the cylinder and exhausted through the exhaust valve is delivered to a compressor to be recompressed and returned to the source of compressed gas. A gear train is selectively engagable to drive the compressor at different operating speed depending upon the pressure maintained at the source of compressed air and/or the speed of the engine. Still further, a second portion of the exhaust gas is used to cool a lubricating fluid for the engine or to operate an air conditioning unit.

In a preferred embodiment of the present invention, the valve for admitting compressed gas to the cylinder is electrically actuated. The device for varying the duration of each engine cycle over which the intake valve remains open as the speed of the engine increase comprises a rotating element whose effective length increases as the speed of the engine increases such that a first contact on the rotating element is electrically connected to a second contact for a longer period of each engine cycle. The second contact actuates the valve whereby the valve remains in an open position for a longer period of each engine cycle as the speed of the engine increases.

Still further features of the present invention include an adaptor plate for supporting the distributor above an intake manifold of a conventional internal combustion engine after a carburetor has been removed to allow air

to enter the cylinders of the engine through the intake manifold and conventional intake valves. Another adaptor plate is arranged over an exhaust passageway of the internal combustion engine to reduce the cross-sectional area of the exhaust passageway.

BRIEF DESCRIPTION OF THE DRAWINGS

Preferred embodiments of a method and apparatus for operating an engine according to the present invention will be described with reference to the accompanying drawings wherein like members bear like reference numerals and wherein:

FIG. 1 is a schematic representation of an apparatus according to the present invention arranged on an engine;

FIG. 2 is a side view of one embodiment of a valve actuator according to the present invention;

FIG. 3 is a cross-sectional view taken along the line 3—3 in FIG. 2;

FIG. 4 is a cross-sectional view of a second embodiment of a valve actuator according to the present invention;

FIG. 5 is a view taken along the line 5—5 in FIG. 4;

FIG. 6 is a cross-sectional view of a third embodiment of a valve actuator according to the present invention;

FIG. 7 is a view taken along the line 7—7 in FIG. 6;

FIG. 8 is a cross-sectional view of a gearing unit to drive a compressor according to the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

With reference to FIG. 1, an engine block **21** (shown in phantom) having two banks of cylinders with each bank including cylinders **20** having pistons **22** reciprocable therein (only one of which is shown in phantom) in a conventional manner. While the illustrated engine is a V-8 engine, it will be apparent that the present invention is applicable to an engine having any number of pistons and cylinders with the V-8 engine being utilized for illustration purposes only. A compressed gas tank **23** is provided to store a compressed gas at high pressure. It may also be desirable to include a small electric or gas compressor to provide compressed gas to supplement the compressed gas held in the tank **23**. In a preferred embodiment, the compressed gas is air which can be obtained from any suitable source.

A line **25** transports the gas withdrawn from the tank **23** when a conventional shut off valve **27** is open. In addition, a solenoid valve **29** preferably operated by a suitable key operated switch (not shown) for the engine is also arranged in the line **25**. In normal operation, the valve **27** is maintained open at all times with the solenoid valve **29** operating as a selective shut off valve to start and stop the engine **21** of the present invention.

A suitable regulating valve **31** is arranged downstream from the solenoid valve **29** and is connected by a linkage **33** to a throttle linkage **35** which is operator actuated by any suitable apparatus such as a foot pedal (not shown). The line **25** enters an end of a distributor **33** and is connected to an end of a pipe **35** which is closed at the other end. A plurality of holes, which are equal to the number of cylinders in the engine **21**, are provided on either side of the pipe **35** along the length of the pipe **35**.

When the present invention is used to adapt a conventional internal combustion engine for operation on compressed gas, an adaptor plate **36** is provided to support the distributor **33** in spaced relation from the usual intake opening in the intake manifold of the engine after a conventional carburetor has been removed. In this way, air is permitted to enter the internal combustion engine through the usual passageways and to be admitted to the cylinders through suitable intake valves (not shown). The adaptor plate **36** is secured to the engine block **21** and the distributor **33** by any suitable apparatus, e.g., bolts.

Each of the holes in the pipe **35** is connected in fluid-tight manner to a single line **37**. Each line **37** carries the compressed gas to a single cylinder **20**. In a preferred embodiment, each of the lines **37** is $\frac{1}{2}$ inch high pressure plastic tubing attached through suitable connectors to the distributor **33** and the pipe **35**. Each of the lines **37** is connected to a valve **39** which is secured in an opening provided near the top of each of the cylinders **20**. In the case of a conversion of a standard internal combustion engine, the valves **39** can be conveniently screwed into a tapped hole in the cylinder **20** typically provided for a spark plug of the internal combustion engine. In a preferred embodiment, the valves **39** are solenoid actuated valves in order to provide a fast and reliable opening and closing of the valves **39**.

Each of the valves **39** is energized by a valve actuator **41** through one of a plurality of wires **43**. The valve actuator **41** is driven by a shaft of the engine similar to the drive for a conventional distributor of an internal combustion engine. That is, a shaft **55** of the valve actuator **41** is driven in synchronism with the engine **21** at one half the speed of the engine **21**.

A first embodiment of the valve actuator **41** (FIGS. 2 and 3) receives electrical power through a wire **45** which is energized in a suitable manner by a battery, and a coil if necessary (not shown) as is conventional in an internal combustion engine. The wire **45** is attached to a central post **47** by a nut **49**. The post **47** is connected to a conducting plate **51** arranged within a housing **53** for the valve actuator **41**. Within the housing **53**, the shaft **55** has an insulating element **57** secured to an end of the shaft **55** for co-rotation therewith when the shaft **55** is driven by the engine **21**. A first end of a flexible contact **59** is continuously biased against the conducting plate **51** to receive electricity from the battery or another suitable source. A second end of the contact **59** is connected to a conducting sleeve **60** which is in constant contact with a spring biased contact **61** which is arranged within the sleeve **60**. The contact **61** is biased by a spring **63** which urges the contact **61** towards a side wall of the housing **53**.

With reference to FIG. 3, a plurality of contacts **65** are spaced from one another and are arranged around the periphery of the housing **53** at the same level as the spring biased contact **61**. Each contact **65** is electrically connected to a post **67** which extends outside of the housing **53**. The number of contacts **65** is equal to the number of cylinders in the engine **21**. One of the wires **43**, which actuate the valves **39**, is secured to each of the posts **67**.

In operation, as the shaft **55** rotates in synchronism with the engine **21**, the insulating element **57** rotates and electricity is ultimately delivered to successive ones of the contacts **65** and wires **43** through the spring biased contact **61** and the flexible contact **59**. In this way, each of the electrical valves **39** is actuated and opened in the proper timed sequence to admit compressed gas to each of the cylinders **20** to drive the pistons **22** therein on a downward stroke.

The embodiment illustrated in FIGS. 2 and 3 is effective to actuate each of the valves 39 to remain open for a long enough period of time to admit sufficient compressed gas to each of the cylinders 20 of the engine 21 to drive the engine 21. The length of each of the contacts 65 around the periphery of the housing 53 is sufficient to permit the speed of the engine to be increased when desired by the operator by moving the throttle linkage 35 which actuates the linkage 33 to further open the regulating valve 31 to admit more compressed gas from the tank 23 to the distributor 33. However, it has been found that the amount of air admitted by the valves 39 when using the first embodiment of the valve actuator 41 (FIGS. 2 and 3) is substantially more than required to operate the engine 21 at an idling speed. Therefore, it may be desirable to provide a valve actuator 41 which is capable of varying the duration of each engine cycle over which the solenoid valves 39 are actuated, i.e., remain open to admit compressed gas, as the speed of the engine 21 is varied.

A second embodiment of a valve actuator 41 which is capable of varying the duration of each engine cycle over which each of the valves 39 remains open to admit compressed gas to the cylinders 20 dependent upon the speed of the engine 21 will be described with reference to FIGS. 4 and 5 wherein members corresponding to those of FIGS. 2 and 3 bear like reference numerals. The wire 45 from the electrical source is secured to the post 47 by the nut 49. The post 47 has a annular contact ring 69 electrically connected to an end of the post 47 and arranged within the housing 53. The shaft 55 rotates at one half the speed of the engine as in the embodiment of FIGS. 2 and 3.

At an upper end of the shaft 55, a splined section 71 slidably receives an insulating member 73. The splined section 71 of the shaft 55 positively holds the insulating member 73 for co-rotation therewith but permits the insulating member 73 to slide axially along the length of the splined section 71. Near the shaft 55, a conductive sleeve 72 is arranged in a bore 81 in an upper surface of the insulating element 73 generally parallel to the splined section 71. A contact 75, biased towards the annular contact ring 69 by a spring 77, is arranged within the conductive sleeve 72 in contact therewith. The conductive sleeve 72 also contacts a conductor 79 at a base of the bore 81.

The conductor 79 extends to the upper surface of the insulating element 73 near an outer periphery of the insulating element 73 where the conductor 79 is electrically connected to a flexible contact 83. The flexible contact 83 selectively engages a plurality of radial contacts 85 arranged on an upper inside surface of the housing 53. A weak spring 87 arranged around the splined section 71 engages a stop member 89 secured on the shaft 55 and the insulating element 73 to slightly bias the insulating element 73 towards the upper inside surface of the housing 53 to ensure contact between the flexible contact 83 and the upper inside surface of the housing 53. As best seen in FIG. 5, the radial contacts 85 on the upper inside surface of the housing 53 are arranged generally in the form of radial spokes extending from the center of the housing 53 with the number of contacts being equal to the number of cylinders 20 in the engine 21. The number of degrees covered by each of the radial contacts 85 gradually increases as the distance from the center of the upper inside surface of the housing 53 increases.

In operation of the device of FIGS. 4 and 5, as the shaft 55 rotates, electricity flows along a path through the wire 45 down through post 47 to the annular contact member 69 which is in constant contact with the spring biased contact 75. The electrical current passes through the conductive sleeve 72 to the conductor 79 and then to the flexible contact 83. As the flexible contact 83 rotates along with the insulating member 73 and the shaft 55, the tip of the flexible contact 83 successively engages each of the radial contacts 85 on the upper inside of the housing 53. As the speed of the shaft 55 increases, the insulating member 73 and the flexible contact 83 attached thereto move upwardly along the splined section 71 of the shaft 55 due to the radial component of the splines in the direction of rotation under the influence of centrifugal force. As the insulating member 73 moves upwardly, the flexible contact 83 is bent such that the tip of the contact 83 extends further radially outwardly from the center of the housing 53 (as seen in phantom lines in FIG. 4). In other words, the effective length of the flexible contact 83 increases as the speed of the engine 21 increases.

As the flexible contact 83 is bent and the tip of the contact 83 moves outwardly, the tip remains in contact with each of the radial contacts 85 for a longer period of each engine cycle due to the increased angular width of the radial contacts with increasing distance from the center of the housing 53. In this way, the length of time over which each of the valves 39 remains open is increased as the speed of the engine is increased. Thus, a larger quantity of compressed gas or air is injected into the cylinders as the speed increases. Conversely, as the speed decreases and the insulating member 73 moves downwardly along the splined section 71, a minimum quantity of air is injected into the cylinder due to the shorter length of the individual radial contact 85 which is in contact with the flexible contact 83. In this way, the amount of compressed gas that is used during idling of the engine 21 is at a minimum whereas the amount of compressed gas which is required to increase the speed of the engine 21 to a level suitable to drive a vehicle on a highway is readily available.

With reference to FIGS. 6 and 7, a third embodiment of a valve actuator 41 according to the present invention includes an arcuate insulating element 91 having a first end pivotally secured by any suitable device such as screw 92 to the shaft 55 for co-rotation with the shaft 55. The screw 92 is screwed into a tapped hole in the insulating element 91 such that a tab 94 at an end of the screw 92 engages a groove 96 provided in the shaft 55. In this way, the insulating element 91 positively rotates with the shaft 55. However, as the shaft 55 rotates faster, a second end 98 of the insulating element 91 is permitted to pivot outwardly under the influence of centrifugal force because of the groove 96 provided in the shaft 55. A spring 93 connected between the second end 98 of the element 91 and the shaft 55 urges the second end of the element 91 towards the center of the housing 53.

A contact 99 similar to the contact 59 (FIG. 2) is arranged such that one end of the contact 99 is in constant contact with the conducting plate 51 located centrally within the housing 53. The other end of the contact 99 engages a conductive sleeve 101 arranged in bore 102. A contact element 95 is arranged in the conductive sleeve 101 in constant contact with the sleeve 101. The bore 102 is arranged generally parallel to the shaft 55 near the second end of the arcuate insulating

element **91**. The contact **95** is biased by a spring **97** towards the upper inside surface of the housing **53** for selective contact with each of the plurality of radial contacts **85** which increase in arc length towards the outer peripheral surface of the housing **53** (FIG. **6**).

In operation of the device of FIGS. **6** and **7**, as the shaft **55** rotates the arcuate insulating element **91** rotates with the shaft **55** and the second end **98** of the insulating element **91** tends to pivot about the shaft **55** due to centrifugal force. Thus, as the effective length of the contact **95** increases, i.e., as the arcuate insulating element **91** pivots further outwardly, the number of degrees of rotation over which the contact **95** is in contact with each of the radial contacts **85** on the upper inside surface of the housing **53** increases thereby permitting each of the valves **39** to remain open for a longer period of each engine cycle to admit more compressed gas to the respective cylinder **20** to further increase the speed of the engine **21**.

With reference to FIG. **1**, a mechanical advance linkage **104** which is connected to the throttle linkage **35**, advances the initiation of the opening of each valve **39** such that compressed gas is injected into the respective cylinder further before the piston **22** in the respective cylinder **20** reaches a top dead center position as the speed of the engine is increased by moving the throttle linkage **35**. The advance linkage **104** is similar to a conventional standard mechanical advance employed on an internal combustion engine. In other words, the linkage **104** varies the relationship between the angular positions of a point on the shaft **55** and a point on the housing **53** containing the contacts. Alternatively, a conventional vacuum advance could also be employed. By advancing the timing of the opening of the valves **39**, the speed of the engine can more easily be increased.

The operation of the engine cycle according to the present invention will now be described. The compressed gas injected into each cylinder of the engine **21** drives the respective piston **22** downward to drive a conventional crankshaft (not shown). The movement of the piston downwardly causes the compressed gas to expand rapidly and cool. As the piston **22** begins to move upwardly in the cylinder **20** a suitable exhaust valve (not shown) arranged to close an exhaust passageway is opened by any suitable apparatus. The expanded gas is then expelled through the exhaust passageway. As the piston **22** again begins to move downwardly a suitable intake valve opens to admit ambient air to the cylinder. The intake valve closes and the ambient air is compressed on the subsequent upward movement of the piston until the piston reaches approximately the top dead center position at which time the compressed gas is again injected into the cylinder **20** to drive the piston **22** downward and the cycle begins anew.

In the case of adapting a conventional internal combustion engine for operation on compressed gas, a plurality of plates **103** are preferably arranged over an end of the exhaust passageways in order to reduce the outlet size of the exhaust passageways of the conventional internal combustion engine. In the illustrated embodiment, a single plate having an opening in the center is bolted to the outside exhaust passageway on each bank of the V-8 engine while another single plate having two openings therein is arranged with one opening over each of the interior exhaust passageways on each bank of the V-8 engine. A line **105** is suitably attached to each of the adaptor places to carry the exhaust to an appropriate location. In a preferred embodiment, the exhaust lines **105** are 1½" plastic tubing.

In a preferred embodiment, the exhaust lines **105** of one bank of the V-8 engine are collected in a line **107** and fed to an inlet of a compressor **109**. The pressure of the exhaust gas emmanating from the engine **21** according to the present invention is approximately 25 p.s.i. In this way, the compressor **109** does not have to pull the exhaust into the compressor since the gas exhausted from the engine **21** is at a positive pressure. The positive pressure of the incoming fluid increases the efficiency and reduces wear on the compressor **109**. The exhaust gas is compressed in the compressor **109** and returned through a line **111** and a check valve **113** to the compressed gas storage tank **23**. The check valve **113** prevents the flow of compressed gas stored in the tank **23** back towards the compressor **109**.

A suitable pressure sensor **115** is arranged at an upper end of the tank **23** and sends a signal along a line **117** when the pressure exceeds a predetermined level and when the pressure drops below a predetermined level. The line **117** controls an electrically actuated clutch **119** disposed at a front end of the compressor **109**. The clutch **119** is operative to engage and disengage the compressor **109** from a drive pulley **121**. Also, the signal carried by the line **117** actuates a suitable valve **123** arranged on a compressor housing **125** to exhaust the air entering the compressor housing **125** from the line **107** when the clutch **119** has disengaged the compressor **109** from the drive pully **121**.

In a preferred embodiment, when the pressure is the tank **23** reaches approximately 600 p.s.i., the clutch **119** is disengaged and the compressor **109** is deactivated and the valve **123** is opened to exhaust the expanded gas delivered to the compressor **109** from the line **107** to the atmosphere. When the pressure within the tank **23** drops below approximately 500 p.s.i., the sensor **115** sends a signal to engage the clutch **119** and close the valve **123**, thereby operating the compressor **109** for supplying the tank **23** with compressed gas.

The pulley **121** which drives the compressor **109** through the clutch **119** is driven by a belt **127** which is driven by a pulley **129** which operates through a gear box **131**. With reference to FIGS. **1** and **8**, a second pulley **133** on the gear box is driven by a belt **135** from a pulley **137** arranged on a drive shaft **139** of the engine **21**. The pulley **137** drives a splined shaft **140** which has a first gear **141** and a second larger gear **143** arranged thereon for rotation with the splined shaft **140**. The splined shaft **140** permits axial movement of the gears **141** and **143** along the shaft **140**.

In normal operation (as seen in FIG. **8**), the first gear **141** engages a third gear **145** arranged on a shaft **147** which drives the pulley **129**. The shafts **140** and **147** are arranged in suitable bearings **149** arranged at each end thereof. When the speed of the engine **21** drops below a predetermined level, a suitable sensor **151** responsive to the speed of the drive shaft **139** of the engine **21** generates a signal which is transmitted through a line **153** to a solenoid actuator **155** arranged within the gear box **131**. The solenoid actuator **155** moves the first and second gears **141**, **143** axially along the splined shaft **140** to the right as seen in FIG. **8** such that the second, larger gear **143** engages a fourth smaller gear **157** which is arranged on the shaft **147**. The ratio of the second gear **143** to the fourth gear **157** is preferably approximately 3 to 1.

In this way, when the speed of the engine **21** drops below the predetermined level as sensed by the sensor **151** (which predetermined level is insufficient to drive the compressor **109** at a speed sufficient to generate the 500-600 pounds of pressure which is preferably in the tank **23**), the solenoid actuator **155** is energized to slide the gears **143**, **141** axially along the splined shaft **140** so that the second, larger gear **143** engages the fourth, smaller gear **157** to drive the pulley **129** and hence the compressor **109** at a higher rate of speed to generate the desired pressure. When the speed of the engine increases above the predetermined level, in a preferred embodiment approximately 1500 rpm, the solenoid actuator **155** is deactivated by the sensor **151** thereby moving the gears **143** and **141** to the left as seen in FIG. **8** such that the first gear **141** re-engages with the third gear **145** to effectuate a 1 to 1 ratio between the output shaft **139** of the engine **21** and the pulley **129**.

The other bank of the V-8 engine has its exhaust ports arranged with adapter plates **103** similar to those on the first bank. However, the exhaust from this bank of the engine **21** is not collected and circulated through the compressor **109**. In a preferred embodiment, a portion of the exhaust is collected in a line **159** and fed to an enlarged chamber **161**. A second fluid is fed through a line **163** into the chamber **161** to be cooled by the cool exhaust emmanating from the engine **21** in the line **159**. The second fluid in the line **163** may be either transmission fluid contained in a transmission associated with the engine **21** or a portion of the oil used to lubricate the engine **21**. A second portion of the exhaust from the second bank of the V-8 engine is removed from the line **159** in a line **165** and used as a working fluid in an air conditioning system or for any other suitable use.

It should be noted that the particular arrangement utilized for collecting and distributing the gas exhausted from the engine **21** would be determined by the use for which the engine is employed. In other words, it may be advantageous to rearrange the exhaust tubing such that a larger or smaller percentage of the exhaust is routed through the compressor **109**. It should also be noted that since the exhaust lines **105** are plastic tubing, a rearrangement of the lines for a different purpose is both simple and inexpensive.

In operation of the engine of the present invention, the engine **21** is started by energizing the solenoid valve **29** and any suitable starting device (not shown), e.g., a conventional electric starter as used on an internal combustion engine. Compressed gas from the full tank **23** flows through the line **25** and a variable amount of the compressed gas is admitted to the distributor **33** by controlling the regulator valve **31** through the linkage **33** and the operator actuated throttle linkage **35**. The compressed gas is distributed to each of the lines **37** which lead to the individual cylinders **20**. The compressed gas is admitted to each of the cylinders **20** in timed relationship to the position of the pistons within the cylinders by opening the valves **39** with the valve actuator **41**.

When it is desired to increase the speed of the engine, the operator moves the throttle linkage **35** which simultaneously admits a larger quantity of compressed gas to the distributor **33** from the tank **23** by further opening the regulator valve **31**. The timing of the valve actuator **41** is also advanced through the linkage **104**. Still further, as the speed of the engine **21** increases, the effective length of the rotating contact **83** (FIG. **4**) or **95** (FIG. **6**) increases thereby electrically contacting a wider portion of one of the stationary radial contacts **85** to cause each of the valves **39** to remain open for a longer period of each engine cycle to admit a larger quantity of compressed gas to each of the cylinders **20**.

As can be seen, the combination of the regulating valve **31**, the mechanical advance **104**, and the valve actuator **41**, combine to produce a compressed gas engine which is quickly and efficiently adaptable to various operating speeds. However, all three of the controls need not be employed simultaneously. For example, the mechanical advance **104** could be utilized without the benefit of one of the varying valve actuators **41** but the high speed operation of the engine may not be as efficient. By increasing the duration of each engine cycle over which each of the valves **39** remains open to admit compressed gas to each of the cylinders **20** as the speed increases, conservation of compressed gas during low speed operation and efficient high speed operation are both possible.

After the compressed gas admitted to the cylinder **20** has forced the piston **22** downwardly within the cylinder to drive the shaft **139** of the engine, the piston **22** moves upwardly within the cylinder **20** and forces the expanded gas out through a suitable exhaust valve (not shown) through the adapter plate **103** (if employed) and into the exhaust line **105**. The cool exhaust can then be collected in any suitable arrangement to be compressed and returned to the tank **23** or used for any desired purpose including use as a working fluid in an air conditioning system or as a coolant for oil.

When using the apparatus and method of the present invention to adapt a ordinary internal combustion engine for operation with compressed gas it can be seen that considerable savings in weight are achieved. For example, the ordinary cooling system including a radiator, fan, hoses, etc. can be eliminated since the compressed gas is cooled as it expands in the cylinder. In addition, there are no explosions within the cylinder to generate heat. Further reductions in weight are obtained by employing plastic tubing for the lines which carry the compressed gas between the distributor and the cylinders and for the exhaust lines. Once again, heavy tubing is not required since there is little or no heat generated by the engine of the present invention. In addition, the noise generated by an engine according to the present invention is considerably less than that generated by an ordinary internal combustion engine since there are no explosions taking place within the cylinders.

The principles of preferred embodiments of the present invention have been described in the foregoing specification. However, the invention which is intended to be protected is not to be construed as limited to the particular embodiments disclosed. The embodiments are to be regarded as illustrative rather than restrictive. Variations and changes may be made by others without departing from the spirit of the invention. Accordingly, it is expressly intended that all such variations and changes which fall within the spirit and the scope of the present invention as defined in the appended claims be embraced thereby.

What is claimed is:

1. An apparatus for operating an engine having at least one cylinder and a reciprocating piston therein comprising:

a source of compressed gas;

distributor means connected with the source of compressed gas for distributing the compressed gas to the at least one cylinder;

valve means for admitting the compressed gas to the at least one cylinder when the piston is in approximately a top dead center position within the cylinder;

altering means for increasing the duration of each engine cycle over which the valve means admits compressed gas to the at least one cylinder as the speed of the engine increases; and

exhaust means for exhausting gas as the piston subsequently approaches approximately the top dead center position.

2. The apparatus of claim **1** further comprising control means for controlling the amount of compressed gas admitted to the distributor means.

3. The apparatus of claim **1** wherein the valve means is a solenoid valve secured in an opening in the cylinder above the level of the piston at the top dead center position.

4. The apparatus of claims **1** or **2** further comprising means for advancing the timing of the valve means as the speed of the engine increases such that compressed gas is admitted progressively further before the top dead center position as the speed of the engine increases.

5. The apparatus of claim **4** wherein the means for advancing the timing comprises a mechanical linkage connected to an operator actuated accelerator linkage.

6. The apparatus of claim **1** wherein a portion of the gas exhausted through the exhaust means is compressed in a compressor driven by an output shaft of the engine and is returned to the source of compressed gas.

7. The apparatus of claim **1** wherein a portion of the gas exhausted through the exhaust means is used to cool transmission fluid for a transmission associated with the engine.

8. The apparatus of claim **1** wherein a portion of the gas exhausted through the exhaust means is used as a working fluid in an air conditioning system.

9. The apparatus of claim **6** further comprising first gearing means interposed between the output shaft of the engine and the compressor for increasing the speed at which the compressor is driven.

10. The apparatus of claim **6** further comprising clutch means attached to the compressor both for disengaging the compressor from the output shaft of the engine when a first predetermined pressure at the source of compressed gas is exceeded and for engaging the compressor with the output shaft of the engine when the pressure at the source of compressed gas drops below a second predetermined pressure.

11. The apparatus of claim **9** further comprising means for both disengaging the first gearing means when a predetermined speed of the engine is exceeded and engaging a second gearing means for driving the compressor at a speed slower than the first gearing means when the predetermined speed of the engine is exceeded.

12. The apparatus of claim **1** wherein the valve means is electrically actuated and wherein the altering means comprises:

a rotating member timed with the at least one cylinder and arranged within a housing;

first and second contacts arranged on a first end of the rotating member and on an inside surface of the housing, respectively;

means for increasing the distance of the first contact from the rotational axis of the rotating member as the speed of the engine increases such that the first contact moves radially outwardly within the housing; and

said second contact presenting a longer arc length to the first contact as the distance of the first contact from the rotational axis of the rotating member increases.

13. The apparatus of claim **12** wherein the rotating member comprises an arcuate arm and wherein the means for increasing the distance of the first contact comprises pivotally mounting a second end of the arcuate arm about the axis of rotation of the rotating member and spring means for biasing the first end of the arcuate arm towards a radially inward position whereby the first end of the arcuate arm pivots radially outwardly as the speed of the engine increases.

14. The apparatus of claim **12** wherein the rotating member is axially slidably received on a rotating shaft for co-rotation therewith, said shaft having splines with a radial component in the direction of rotation, and wherein the first contact comprises a flexible contact located on an upper surface of the rotating member, said flexible contact being biased against the inside surface of the housing which carries the second contacts whereby as the speed of the engine increases the rotating member is urged axially along the splined shaft towards the inside surface of the housing such that the flexible contact is forced radially outwardly along the inside surface.

15. The apparatus of claim **12** wherein the second contact comprises of radially extending conductor arranged on an upper inside surface of the housing, said conductor increasing in arc length as the conductor extends radially outwardly from a central portion of the housing.

16. An apparatus for adapting an internal combustion engine for operation with compressed gas, the internal combustion engine having at least one cylinder, a piston reciprocable within the at least one cylinder, intake and exhaust means disposed in the at least one cylinder, and a tapped hole in the at least one cylinder adapted to receive a spark plug, the apparatus comprising:

a source of compressed gas;

distributor means connected with the source of compressed gas for distributing the compressed gas to the at least one cylinder;

valve means arranged in the tapped hole for admitting the compressed gas to the at least one cylinder when the piston is in approximately a top dead center position within the cylinder; and

altering means for increasing the duration of each engine cycle over which the valve means remains open to admit the compressed gas as the speed of the engine increases.

17. An apparatus as in claim **16** further comprising first adapter plate means for supporting the distributor means above an intake manifold of the engine, which adaptor plate means allows ambient air to enter through the intake manifold.

18. The apparatus of claim **16** further comprising second adapter plate means for reducing the exit area of the exhaust means.

19. A method of operating an engine on compressed gas, said engine having at least one cylinder and a piston reciprocable therein comprising the steps of:

delivering compressed gas from a source to a distributor;

distributing the compressed gas to the at least one cylinder;

admitting compressed gas to the at least one cylinder through an intake valve when the piston is at approximately a top dad center position;

increasing the duration of each engine cycle over which compressed gas is admitted to the at least one cylinder as the engine speed increases; and

exhausting the remaining gas when the piston subsequently reaches approximately the top dead center position.

20. The method of claim **19** further comprising the step of controlling the amount of compressed gas which is delivered to the distributor.

21. The method of claim **19** further comprising the step of advancing the timing of the opening of the intake valve as the speed of the engine increases.

22. An apparatus for operating an engine having at least one cylinder and a piston reciprocable therein on compressed gas comprising:

a source of compressed gas;

distributor means connected with the source of compressed gas for distributing the compressed gas to the at least one cylinder;

electrically actuated valve means secured in an opening in the at least one cylinder for selectively admitting compressed gas to the at least one cylinder when the piston is in approximately a top dead center position; and

means for advancing the timing of the valve means as the speed of the engine increases whereby compressed gas is admitted progressively further before the top dead center position as the speed of the engine increases.

* * * * *

United States Patent [19]

Rogers, Sr.

[11] **4,292,804**

[45] **Oct. 6, 1981**

---

[54] **METHOD AND APPARATUS FOR OPERATING AN ENGINE ON COMPRESSED GAS**

[76] Inventor: **Leroy K. Rogers, Sr.**, #5 Capistrano Ct., Ft. Myers, Fla. 33908

[21] Appl. No.: **158,303**

[22] Filed: **Jun. 10, 1980**

[51] **Int. Cl.^{3}** **F15B 11/06**

[52] **U.S. Cl.** **60/407**; 91/187; 91/275

[58] **Field of Search** 60/407, 412; 91/187, 91/275, 364

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,881,399	5/1975	Sagi et al.	91/187 X
3,885,387	5/1975	Simington	60/407 X
4,018,050	4/1977	Murphy	60/412 X

Primary Examiner—Allen M. Ostrager
Attorney, Agent, or Firm—Burns, Doane, Swecker & Mathis

[57] **ABSTRACT**

The present invention relates to a method and apparatus for operating an engine having a cylinder and a piston reciprocable therein on compressed gas. The apparatus comprises a source of compressed gas connected to a distributor which distributes the compressed gas to the cylinder. A valve is provided to selectively admit compressed gas to the cylinder when the piston is in an approximately top dead center position. In one embodiment of the present invention the timing of the opening of the valve is advanced such that the compressed gas is admitted to the cylinder progressively further before the top dead center position of the piston as the speed of the engine increases. In a further embodiment of the present invention a valve actuator is provided which increases the length of time over which the valve remains open to admit compressed gas to the cylinder as the speed of the engine increases. A still further embodiment of the present invention relates to an apparatus for adapting a conventional internal combustion engine for operation on compressed gas.

22 Claims, 8 Drawing Figures

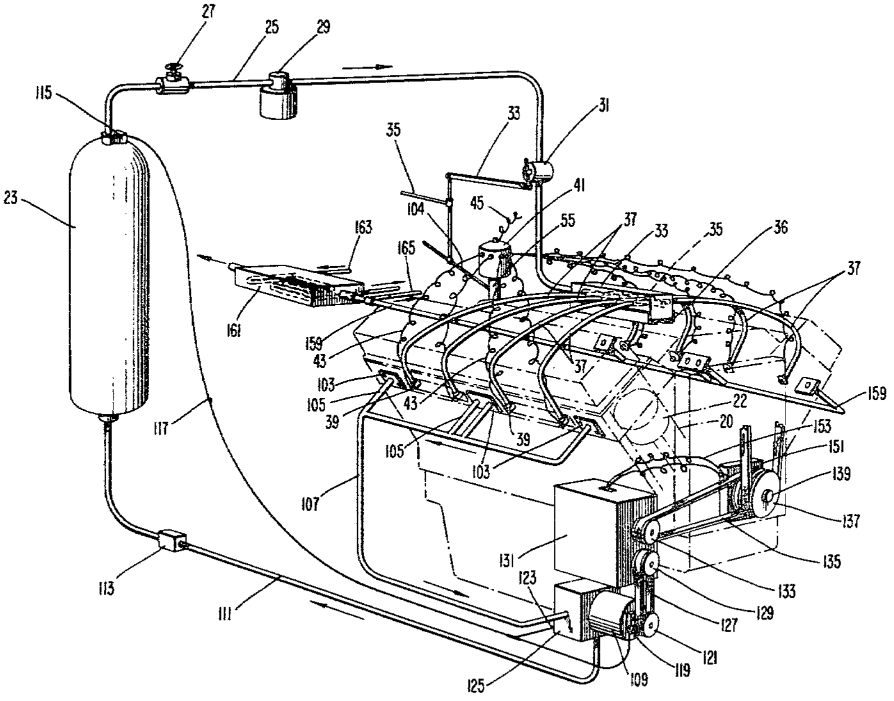

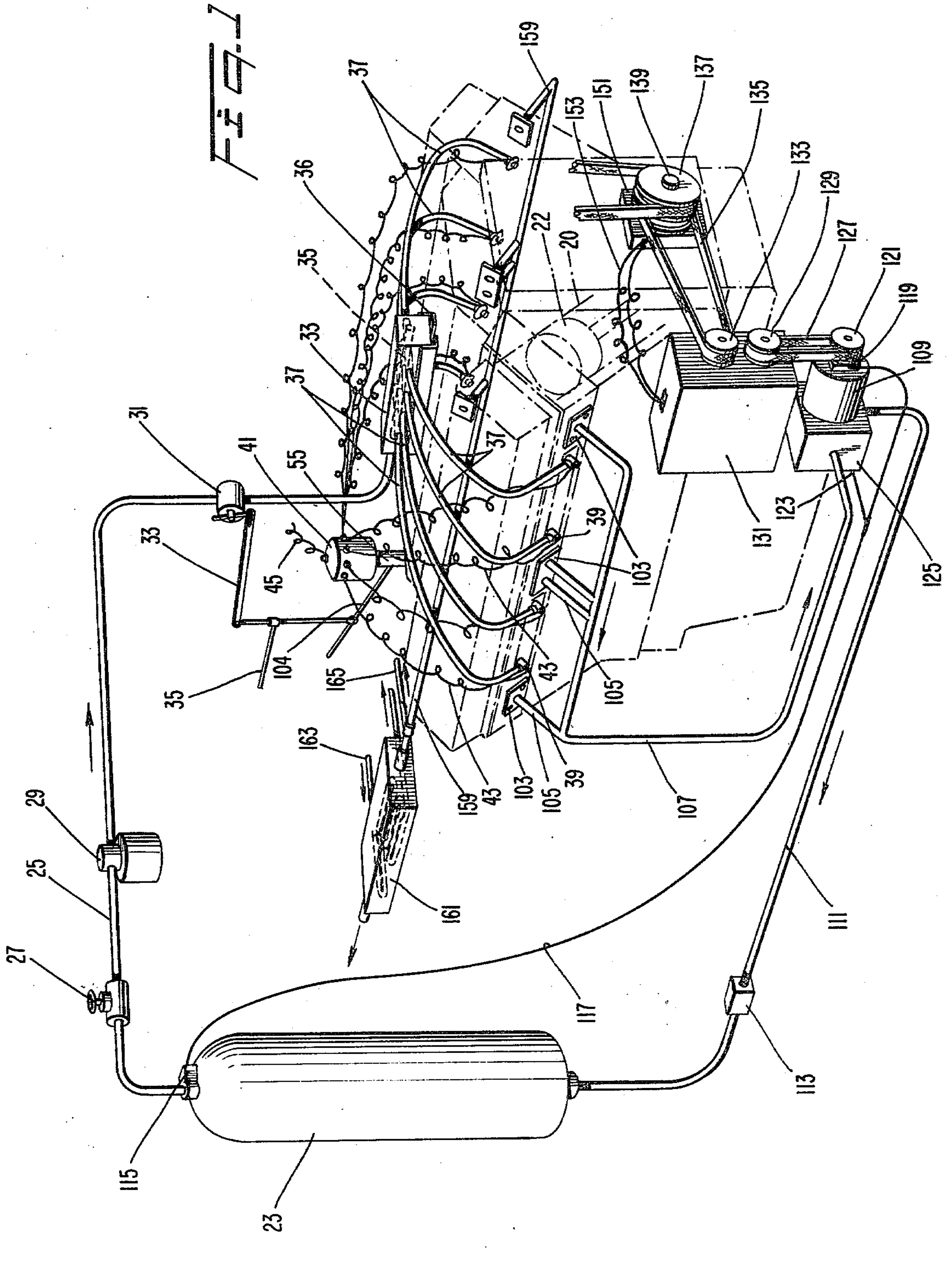

FIG. 1

FIG. 6

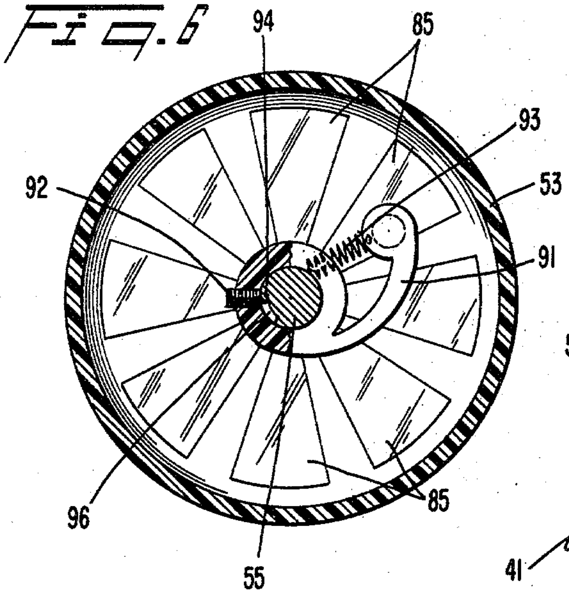
94
85
93
92
53
91
96
55
85

FIG. 7

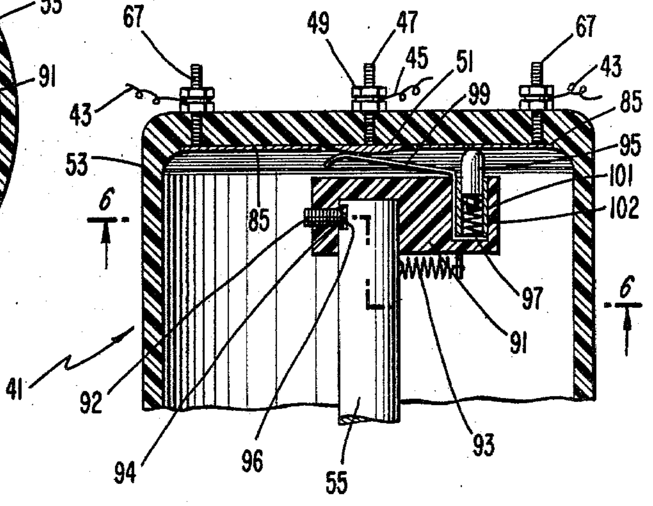
67
49
47
67
43
45
51
43
99
85
53
95
101
102
6
85
97
6
91
92
41
93
94
96
55

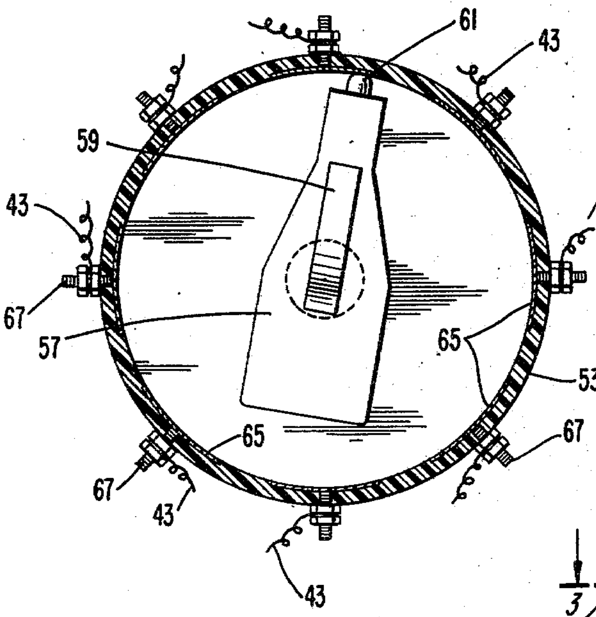
61
43
59
43
67
57
65
53
65
67
67
43
43

FIG. 3

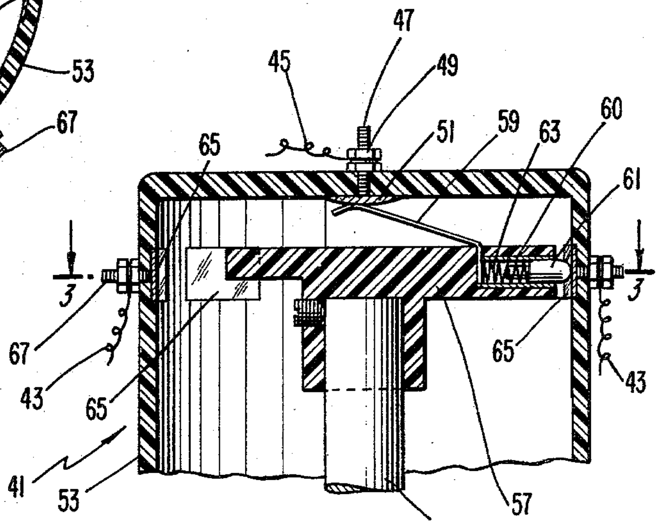
47
45
49
51
59
63
60
65
61
3
3
67
43
65
65
43
57
41
53
55

FIG. 2

Fig. 5

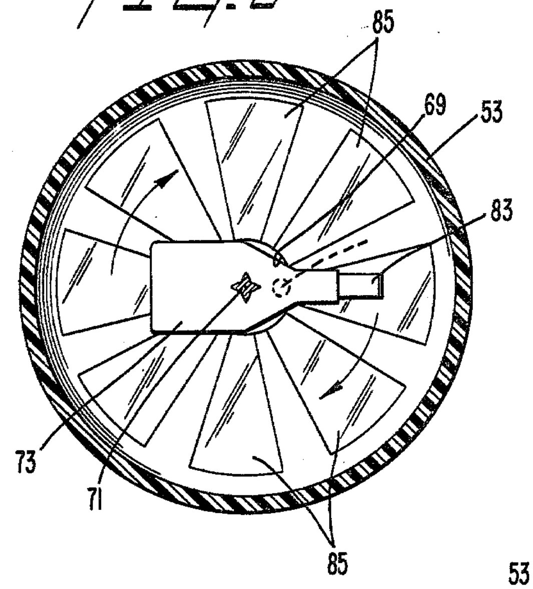
85
69
53
83
73
71
85

Fig. 4

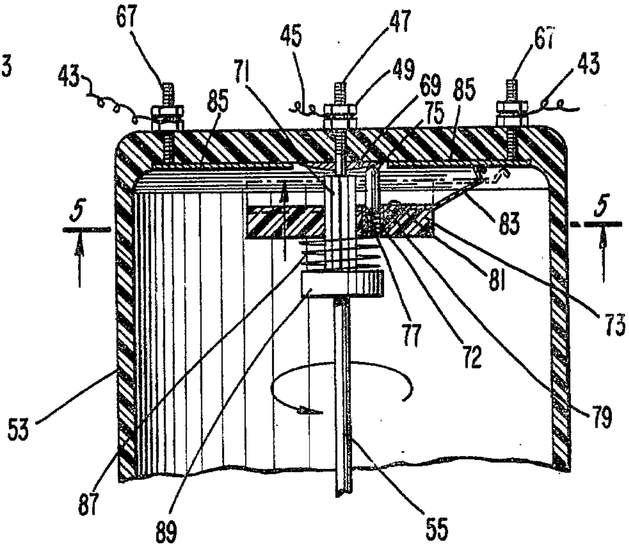
67
47
45
67
43
49
43
71
69
85
85
75
5
5
83
81
73
77
72
79
53
87
89
55

Fig. 6

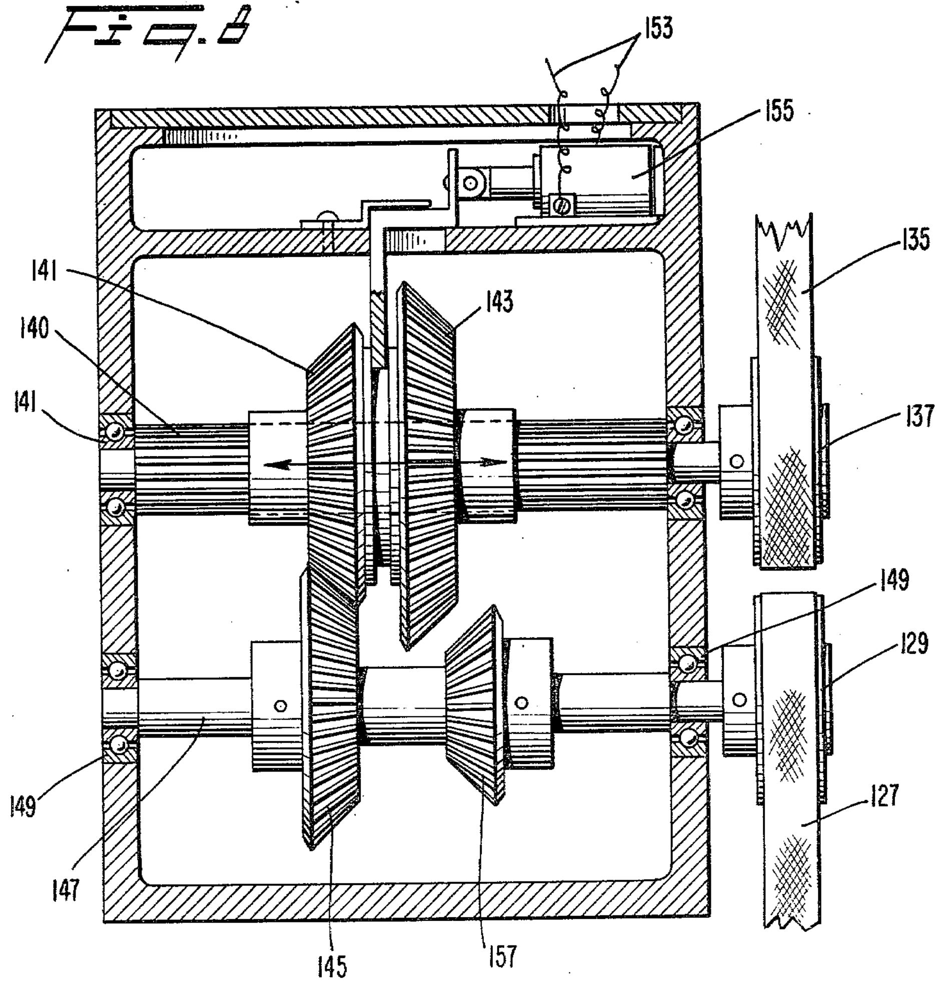
153
155
135
141
143
140
141
137
149
129
149
127
147
145
157

METHOD AND APPARATUS FOR OPERATING AN ENGINE ON COMPRESSED GAS

BACKGROUND AND SUMMARY OF THE PRESENT INVENTION

The present invention relates to a method and apparatus for operating an engine using a compressed gas as the motive fluid. More particularly, the present invention relates to a apparatus for adapting a pre-existing internal combustion engine for operation on a compressed gas.

Air pollution is one of the most serious problems facing the world today. One of the major contributors to air pollution is ordinary internal combustion engine which are used in most motor vehicles today. Various devices, including many items mandated by legislation, have been proposed in an attempt to limit the pollutants which an internal combustion engine exhausts to the air. However, most of these devices have met with limited success and are often both prohibitively expensive and complex. A clean alternative to the internal combustion engine is needed to power vehicles and other machinery.

A compressed gas, preferably air, would provide an ideal motive fluid for a engine since it would eliminate the usual pollutants exhausted from an internal combustion engine. An apparatus for converting an internal combustion engine for operation on compressed air is disclosed in U.S. Pat. No. 3,885,387 issued May 27, 1975 to Simington. The Simington patent discloses an apparatus including a source of compressed air and a rotating valve actuator which opens and closes a plurality of mechanical poppet valves. The valves deliver compressed air in timed sequence to the cylinders of an engine through adapters located in the spark plug holes. However, the output speed of an engine of this type is limited by the speed of the mechanical valves and the fact that the length of time over which each of the valves remains open cannot be varied as the speed of the engine increases.

Another apparatus for converting an internal combustion engine for operation on steam or compressed air is disclosed in U.S. Pat. No. 4,102,130 issued July 25, 1978 to Stricklin. The Stricklin patent discloses a device which changes the valve timing of a conventional four stroke engine such that the intake and exhaust valves open once for every revolution of the engine instead of once every other revolution of the engine. A reversing valve is provided which delivers live steam or compressed air to the intake valves and is subsequently reversed to allow the exhaust valves to deliver the expanded steam or air to the atmosphere. A reversing valve of this type however does not provide a reliable apparatus for varying the amount of motive fluid injected into the cylinders when it is desired to increase the speed of the engine. Further, a device of the type disclosed in the Stricklin patent requires the use of multiple reversing valves if the cylinders in a multi-cylinder engine were to be fired sequentially.

Therefore, it is an object of the present invention to provide a reliable method and apparatus for operating an engine or converting an engine for operation with a compressed gas.

A further object of the present invention is to provide a method and apparatus which is effective to deliver a

constantly increasing amount of compressed gas to an engine as the speed of the engine increases.

A still further object of the present invention is to provide a method and apparatus which will operate an engine using compressed gas at a speed sufficient to drive a conventional automobile at highway speeds.

It is still a further object of the present invention to provide a method and apparatus which is readily adaptable to a standard internal combustion engine to convert the internal combustion engine for operation with a compressed gas.

Another object of the invention is to provide a method and apparatus which utilizes cool expanded gas, exhausted from a compressed gas engine, to operate an air conditioning unit and/or an oil cooler.

These and other objects are realized by a method and apparatus according to the present invention for operating an engine having at least one cylinder and a reciprricating piston therein using compressed gas as a motive fluid. The apparatus includes a source of compressed gas and a distributor connected with the source of the compressed gas for distributing the compressed gas to the at least one cylinder. A valve is provided for admitting the compressed gas to the cylinder when the piston is in approximately a top dead center position within the cylinder. An exhaust is provided for exhausting the expanded gas from the cylinder as the piston returns to approximately the top dead center position.

In a preferred embodiment of the present invention a device is provided for varying the duration of each engine cycle over which the valve remains open to admit compressed gas to the cylinder dependent upon the speed of the engine. In a further preferred embodiment of the present invention, an apparatus for advancing the timing of the opening of the valve is arranged to admit the compressed gas to the cylinder progressively further before the top dead center position of the piston as the speed of the engine increases.

Further features of the present invention include a valve for controlling the amount of compressed gas admitted to the distributor. Also, a portion of the gas which has been expanded in the cylinder and exhausted through the exhaust valve is delivered to a compressor to be recompressed and returned to the source of compressed gas. A gear train is selectively engagable to drive the compressor at different operating speed depending upon the pressure maintained at the source of compressed air and/or the speed of the engine. Still further, a second portion of the exhaust gas is used to cool a lubricating fluid for the engine or to operate an air conditioning unit.

In a preferred embodiment of the present invention, the valve for admitting compressed gas to the cylinder is electrically actuated. The device for varying the duration of each engine cycle over which the intake valve remains open as the speed of the engine increase comprises a rotating element whose effective length increases as the speed of the engine increases such that a first contact on the rotating element is electrically connected to a second contact for a longer period of each engine cycle. The second contact actuates the valve whereby the valve remains in an open position for a longer period of each engine cycle as the speed of the engine increases.

Still further features of the present invention include an adaptor plate for supporting the distributor above an intake manifold of a conventional internal combustion engine after a carburetor has been removed to allow air

to enter the cylinders of the engine through the intake manifold and conventional intake valves. Another adaptor plate is arranged over an exhaust passageway of the internal combustion engine to reduce the cross-sectional area of the exhaust passageway.

BRIEF DESCRIPTION OF THE DRAWINGS

Preferred embodiments of a method and apparatus for operating an engine according to the present invention will be described with reference to the accompanying drawings wherein like members bear like reference numerals and wherein:

FIG. 1 is a schematic representation of an apparatus according to the present invention arranged on an engine;

FIG. 2 is a side view of one embodiment of a valve actuator according to the present invention;

FIG. 3 is a cross-sectional view taken along the line 3—3 in FIG. 2;

FIG. 4 is a cross-sectional view of a second embodiment of a valve actuator according to the present invention;

FIG. 5 is a view taken along the line 5—5 in FIG. 4;

FIG. 6 is a cross-sectional view of a third embodiment of a valve actuator according to the present invention;

FIG. 7 is a view taken along the line 7—7 in FIG. 6;

FIG. 8 is a cross-sectional view of a gearing unit to drive a compressor according to the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

With reference to FIG. 1, an engine block **21** (shown in phantom) having two banks of cylinders with each bank including cylinders **20** having pistons **22** reciprocable therein (only one of which is shown in phantom) in a conventional manner. While the illustrated engine is a V-8 engine, it will be apparent that the present invention is applicable to an engine having any number of pistons and cylinders with the V-8 engine being utilized for illustration purposes only. A compressed gas tank **23** is provided to store a compressed gas at high pressure. It may also be desirable to include a small electric or gas compressor to provide compressed gas to supplement the compressed gas held in the tank **23**. In a preferred embodiment, the compressed gas is air which can be obtained from any suitable source.

A line **25** transports the gas withdrawn from the tank **23** when a conventional shut off valve **27** is open. In addition, a solenoid valve **29** preferably operated by a suitable key operated switch (not shown) for the engine is also arranged in the line **25**. In normal operation, the valve **27** is maintained open at all times with the solenoid valve **29** operating as a selective shut off valve to start and stop the engine **21** of the present invention.

A suitable regulating valve **31** is arranged downstream from the solenoid valve **29** and is connected by a linkage **33** to a throttle linkage **35** which is operator actuated by any suitable apparatus such as a foot pedal (not shown). The line **25** enters an end of a distributor **33** and is connected to an end of a pipe **35** which is closed at the other end. A plurality of holes, which are equal to the number of cylinders in the engine **21**, are provided on either side of the pipe **35** along the length of the pipe **35**.

When the present invention is used to adapt a conventional internal combustion engine for operation on compressed gas, an adaptor plate **36** is provided to support the distributor **33** in spaced relation from the usual intake opening in the intake manifold of the engine after a conventional carburetor has been removed. In this way, air is permitted to enter the internal combustion engine through the usual passageways and to be admitted to the cylinders through suitable intake valves (not shown). The adaptor plate **36** is secured to the engine block **21** and the distributor **33** by any suitable apparatus, e.g., bolts.

Each of the holes in the pipe **35** is connected in fluid-tight manner to a single line **37**. Each line **37** carries the compressed gas to a single cylinder **20**. In a preferred embodiment, each of the lines **37** is $\frac{1}{2}$ inch high pressure plastic tubing attached through suitable connectors to the distributor **33** and the pipe **35**. Each of the lines **37** is connected to a valve **39** which is secured in an opening provided near the top of each of the cylinders **20**. In the case of a conversion of a standard internal combustion engine, the valves **39** can be conveniently screwed into a tapped hole in the cylinder **20** typically provided for a spark plug of the internal combustion engine. In a preferred embodiment, the valves **39** are solenoid actuated valves in order to provide a fast and reliable opening and closing of the valves **39**.

Each of the valves **39** is energized by a valve actuator **41** through one of a plurality of wires **43**. The valve actuator **41** is driven by a shaft of the engine similar to the drive for a conventional distributor of an internal combustion engine. That is, a shaft **55** of the valve actuator **41** is driven in synchronism with the engine **21** at one half the speed of the engine **21**.

A first embodiment of the valve actuator **41** (FIGS. **2** and **3**) receives electrical power through a wire **45** which is energized in a suitable manner by a battery, and a coil if necessary (not shown) as is conventional in an internal combustion engine. The wire **45** is attached to a central post **47** by a nut **49**. The post **47** is connected to a conducting plate **51** arranged within a housing **53** for the valve actuator **41**. Within the housing **53**, the shaft **55** has an insulating element **57** secured to an end of the shaft **55** for co-rotation therewith when the shaft **55** is driven by the engine **21**. A first end of a flexible contact **59** is continuously biased against the conducting plate **51** to receive electricity from the battery or another suitable source. A second end of the contact **59** is connected to a conducting sleeve **60** which is in constant contact with a spring biased contact **61** which is arranged within the sleeve **60**. The contact **61** is biased by a spring **63** which urges the contact **61** towards a side wall of the housing **53**.

With reference to FIG. 3, a plurality of contacts **65** are spaced from one another and are arranged around the periphery of the housing **53** at the same level as the spring biased contact **61**. Each contact **65** is electrically connected to a post **67** which extends outside of the housing **53**. The number of contacts **65** is equal to the number of cylinders in the engine **21**. One of the wires **43**, which actuate the valves **39**, is secured to each of the posts **67**.

In operation, as the shaft **55** rotates in synchronism with the engine **21**, the insulating element **57** rotates and electricity is ultimately delivered to successive ones of the contacts **65** and wires **43** through the spring biased contact **61** and the flexible contact **59**. In this way, each of the electrical valves **39** is actuated and opened in the proper timed sequence to admit compressed gas to each of the cylinders **20** to drive the pistons **22** therein on a downward stroke.

The embodiment illustrated in FIGS. **2** and **3** is effective to actuate each of the valves **39** to remain open for a long enough period of time to admit sufficient compressed gas to each of the cylinders **20** of the engine **21** to drive the engine **21**. The length of each of the contacts **65** around the periphery of the housing **53** is sufficient to permit the speed of the engine to be increased when desired by the operator by moving the throttle linkage **35** which actuates the linkage **33** to further open the regulating valve **31** to admit more compressed gas from the tank **23** to the distributor **33**. However, it has been found that the amount of air admitted by the valves **39** when using the first embodiment of the valve actuator **41** (FIGS. **2** and **3**) is substantially more than required to operate the engine **21** at an idling speed. Therefore, it may be desirable to provide a valve actuator **41** which is capable of varying the duration of each engine cycle over which the solenoid valves **39** are actuated, i.e., remain open to admit compressed gas, as the speed of the engine **21** is varied.

A second embodiment of a valve actuator **41** which is capable of varying the duration of each engine cycle over which each of the valves **39** remains open to admit compressed gas to the cylinders **20** dependent upon the speed of the engine **21** will be described with reference to FIGS. **4** and **5** wherein members corresponding to those of FIGS. **2** and **3** bear like reference numerals. The wire **45** from the electrical source is secured to the post **47** by the nut **49**. The post **47** has a annular contact ring **69** electrically connected to an end of the post **47** and arranged within the housing **53**. The shaft **55** rotates at one half the speed of the engine as in the embodiment of FIGS. **2** and **3**.

At an upper end of the shaft **55**, a splined section **71** slidably receives an insulating member **73**. The splined section **71** of the shaft **55** positively holds the insulating member **73** for co-rotation therewith but permits the insulating member **73** to slide axially along the length of the splined section **71**. Near the shaft **55**, a conductive sleeve **72** is arranged in a bore **81** in an upper surface of the insulating element **73** generally parallel to the splined section **71**. A contact **75**, biased towards the annular contact ring **69** by a spring **77**, is arranged within the conductive sleeve **72** in contact therewith. The conductive sleeve **72** also contacts a conductor **79** at a base of the bore **81**.

The conductor **79** extends to the upper surface of the insulating element **73** near an outer periphery of the insulating element **73** where the conductor **79** is electrically connected to a flexible contact **83**. The flexible contact **83** selectively engages a plurality of radial contacts **85** arranged on an upper inside surface of the housing **53**. A weak spring **87** arranged around the splined section **71** engages a stop member **89** secured on the shaft **55** and the insulating element **73** to slightly bias the insulating element **73** towards the upper inside surface of the housing **53** to ensure contact between the flexible contact **83** and the upper inside surface of the housing **53**. As best seen in FIG. **5**, the radial contacts **85** on the upper inside surface of the housing **53** are arranged generally in the form of radial spokes extending from the center of the housing **53** with the number of contacts being equal to the number of cylinders **20** in the engine **21**. The number of degrees covered by each of the radial contacts **85** gradually increases as the distance from the center of the upper inside surface of the housing **53** increases.

In operation of the device of FIGS. **4** and **5**, as the shaft **55** rotates, electricity flows along a path through the wire **45** down through post **47** to the annular contact member **69** which is in constant contact with the spring biased contact **75**. The electrical current passes through the conductive sleeve **72** to the conductor **79** and then to the flexible contact **83**. As the flexible contact **83** rotates along with the insulating member **73** and the shaft **55**, the tip of the flexible contact **83** successively engages each of the radial contacts **85** on the upper inside of the housing **53**. As the speed of the shaft **55** increases, the insulating member **73** and the flexible contact **83** attached thereto move upwardly along the splined section **71** of the shaft **55** due to the radial component of the splines in the direction of rotation under the influence of centrifugal force. As the insulating member **73** moves upwardly, the flexible contact **83** is bent such that the tip of the contact **83** extends further radially outwardly from the center of the housing **53** (as seen in phantom lines in FIG. **4**). In other words, the effective length of the flexible contact **83** increases as the speed of the engine **21** increases.

As the flexible contact **83** is bent and the tip of the contact **83** moves outwardly, the tip remains in contact with each of the radial contacts **85** for a longer period of each engine cycle due to the increased angular width of the radial contacts with increasing distance from the center of the housing **53**. In this way, the length of time over which each of the valves **39** remains open is increased as the speed of the engine is increased. Thus, a larger quantity of compressed gas or air is injected into the cylinders as the speed increases. Conversely, as the speed decreases and the insulating member **73** moves downwardly along the splined section **71**, a minimum quantity of air is injected into the cylinder due to the shorter length of the individual radial contact **85** which is in contact with the flexible contact **83**. In this way, the amount of compressed gas that is used during idling of the engine **21** is at a minimum whereas the amount of compressed gas which is required to increase the speed of the engine **21** to a level suitable to drive a vehicle on a highway is readily available.

With reference to FIGS. **6** and **7**, a third embodiment of a valve actuator **41** according to the present invention includes an arcuate insulating element **91** having a first end pivotally secured by any suitable device such as screw **92** to the shaft **55** for co-rotation with the shaft **55**. The screw **92** is screwed into a tapped hole in the insulating element **91** such that a tab **94** at an end of the screw **92** engages a groove **96** provided in the shaft **55**. In this way, the insulating element **91** positively rotates with the shaft **55**. However, as the shaft **55** rotates faster, a second end **98** of the insulating element **91** is permitted to pivot outwardly under the influence of centrifugal force because of the groove **96** provided in the shaft **55**. A spring **93** connected between the second end **98** of the element **91** and the shaft **55** urges the second end of the element **91** towards the center of the housing **53**.

A contact **99** similar to the contact **59** (FIG. **2**) is arranged such that one end of the contact **99** is in constant contact with the conducting plate **51** located centrally within the housing **53**. The other end of the contact **99** engages a conductive sleeve **101** arranged in bore **102**. A contact element **95** is arranged in the conductive sleeve **101** in constant contact with the sleeve **101**. The bore **102** is arranged generally parallel to the shaft **55** near the second end of the arcuate insulating

element **91**. The contact **95** is biased by a spring **97** towards the upper inside surface of the housing **53** for selective contact with each of the plurality of radial contacts **85** which increase in arc length towards the outer peripheral surface of the housing **53** (FIG. **6**).

In operation of the device of FIGS. **6** and **7**, as the shaft **55** rotates the arcuate insulating element **91** rotates with the shaft **55** and the second end **98** of the insulating element **91** tends to pivot about the shaft **55** due to centrifugal force. Thus, as the effective length of the contact **95** increases, i.e., as the arcuate insulating element **91** pivots further outwardly, the number of degrees of rotation over which the contact **95** is in contact with each of the radial contacts **85** on the upper inside surface of the housing **53** increases thereby permitting each of the valves **39** to remain open for a longer period of each engine cycle to admit more compressed gas to the respective cylinder **20** to further increase the speed of the engine **21**.

With reference to FIG. **1**, a mechanical advance linkage **104** which is connected to the throttle linkage **35**, advances the initiation of the opening of each valve **39** such that compressed gas is injected into the respective cylinder further before the piston **22** in the respective cylinder **20** reaches a top dead center position as the speed of the engine is increased by moving the throttle linkage **35**. The advance linkage **104** is similar to a conventional standard mechanical advance employed on an internal combustion engine. In other words, the linkage **104** varies the relationship between the angular positions of a point on the shaft **55** and a point on the housing **53** containing the contacts. Alternatively, a conventional vacuum advance could also be employed. By advancing the timing of the opening of the valves **39**, the speed of the engine can more easily be increased.

The operation of the engine cycle according to the present invention will now be described. The compressed gas injected into each cylinder of the engine **21** drives the respective piston **22** downward to drive a conventional crankshaft (not shown). The movement of the piston downwardly causes the compressed gas to expand rapidly and cool. As the piston **22** begins to move upwardly in the cylinder **20** a suitable exhaust valve (not shown) arranged to close an exhaust passageway is opened by any suitable apparatus. The expanded gas is then expelled through the exhaust passageway. As the piston **22** again begins to move downwardly a suitable intake valve opens to admit ambient air to the cylinder. The intake valve closes and the ambient air is compressed on the subsequent upward movement of the piston until the piston reaches approximately the top dead center position at which time the compressed gas is again injected into the cylinder **20** to drive the piston **22** downward and the cycle begins anew.

In the case of adapting a conventional internal combustion engine for operation on compressed gas, a plurality of plates **103** are preferably arranged over an end of the exhaust passageways in order to reduce the outlet size of the exhaust passageways of the conventional internal combustion engine. In the illustrated embodiment, a single plate having an opening in the center is bolted to the outside exhaust passageway on each bank of the V-8 engine while another single plate having two openings therein is arranged with one opening over each of the interior exhaust passageways on each bank of the V-8 engine. A line **105** is suitably attached to each of the adaptor places to carry the exhaust to an appropriate location. In a preferred embodiment, the exhaust lines **105** are 1½" plastic tubing.

In a preferred embodiment, the exhaust lines **105** of one bank of the V-8 engine are collected in a line **107** and fed to an inlet of a compressor **109**. The pressure of the exhaust gas emmanating from the engine **21** according to the present invention is approximately 25 p.s.i. In this way, the compressor **109** does not have to pull the exhaust into the compressor since the gas exhausted from the engine **21** is at a positive pressure. The positive pressure of the incoming fluid increases the efficiency and reduces wear on the compressor **109**. The exhaust gas is compressed in the compressor **109** and returned through a line **111** and a check valve **113** to the compressed gas storage tank **23**. The check valve **113** prevents the flow of compressed gas stored in the tank **23** back towards the compressor **109**.

A suitable pressure sensor **115** is arranged at an upper end of the tank **23** and sends a signal along a line **117** when the pressure exceeds a predetermined level and when the pressure drops below a predetermined level. The line **117** controls an electrically actuated clutch **119** disposed at a front end of the compressor **109**. The clutch **119** is operative to engage and disengage the compressor **109** from a drive pulley **121**. Also, the signal carried by the line **117** actuates a suitable valve **123** arranged on a compressor housing **125** to exhaust the air entering the compressor housing **125** from the line **107** when the clutch **119** has disengaged the compressor **109** from the drive pully **121**.

In a preferred embodiment, when the pressure is the tank **23** reaches approximately 600 p.s.i., the clutch **119** is disengaged and the compressor **109** is deactivated and the valve **123** is opened to exhaust the expanded gas delivered to the compressor **109** from the line **107** to the atmosphere. When the pressure within the tank **23** drops below approximately 500 p.s.i., the sensor **115** sends a signal to engage the clutch **119** and close the valve **123**, thereby operating the compressor **109** for supplying the tank **23** with compressed gas.

The pulley **121** which drives the compressor **109** through the clutch **119** is driven by a belt **127** which is driven by a pulley **129** which operates through a gear box **131**. With reference to FIGS. **1** and **8**, a second pulley **133** on the gear box is driven by a belt **135** from a pulley **137** arranged on a drive shaft **139** of the engine **21**. The pulley **137** drives a splined shaft **140** which has a first gear **141** and a second larger gear **143** arranged thereon for rotation with the splined shaft **140**. The splined shaft **140** permits axial movement of the gears **141** and **143** along the shaft **140**.

In normal operation (as seen in FIG. **8**), the first gear **141** engages a third gear **145** arranged on a shaft **147** which drives the pulley **129**. The shafts **140** and **147** are arranged in suitable bearings **149** arranged at each end thereof. When the speed of the engine **21** drops below a predetermined level, a suitable sensor **151** responsive to the speed of the drive shaft **139** of the engine **21** generates a signal which is transmitted through a line **153** to a solenoid actuator **155** arranged within the gear box **131**. The solenoid actuator **155** moves the first and second gears **141**, **143** axially along the splined shaft **140** to the right as seen in FIG. **8** such that the second, larger gear **143** engages a fourth smaller gear **157** which is arranged on the shaft **147**. The ratio of the second gear **143** to the fourth gear **157** is preferably approximately 3 to 1.

In this way, when the speed of the engine **21** drops below the predetermined level as sensed by the sensor **151** (which predetermined level is insufficient to drive the compressor **109** at a speed sufficient to generate the 500-600 pounds of pressure which is preferably in the tank **23**), the solenoid actuator **155** is energized to slide the gears **143**, **141** axially along the splined shaft **140** so that the second, larger gear **143** engages the fourth, smaller gear **157** to drive the pulley **129** and hence the compressor **109** at a higher rate of speed to generate the desired pressure. When the speed of the engine increases above the predetermined level, in a preferred embodiment approximately 1500 rpm, the solenoid actuator **155** is deactivated by the sensor **151** thereby moving the gears **143** and **141** to the left as seen in FIG. **8** such that the first gear **141** re-engages with the third gear **145** to effectuate a 1 to 1 ratio between the output shaft **139** of the engine **21** and the pulley **129**.

The other bank of the V-8 engine has its exhaust ports arranged with adapter plates **103** similar to those on the first bank. However, the exhaust from this bank of the engine **21** is not collected and circulated through the compressor **109**. In a preferred embodiment, a portion of the exhaust is collected in a line **159** and fed to an enlarged chamber **161**. A second fluid is fed through a line **163** into the chamber **161** to be cooled by the cool exhaust emmanating from the engine **21** in the line **159**. The second fluid in the line **163** may be either transmission fluid contained in a transmission associated with the engine **21** or a portion of the oil used to lubricate the engine **21**. A second portion of the exhaust from the second bank of the V-8 engine is removed from the line **159** in a line **165** and used as a working fluid in an air conditioning system or for any other suitable use.

It should be noted that the particular arrangement utilized for collecting and distributing the gas exhausted from the engine **21** would be determined by the use for which the engine is employed. In other words, it may be advantageous to rearrange the exhaust tubing such that a larger or smaller percentage of the exhaust is routed through the compressor **109**. It should also be noted that since the exhaust lines **105** are plastic tubing, a rearrangement of the lines for a different purpose is both simple and inexpensive.

In operation of the engine of the present invention, the engine **21** is started by energizing the solenoid valve **29** and any suitable starting device (not shown), e.g., a conventional electric starter as used on an internal combustion engine. Compressed gas from the full tank **23** flows through the line **25** and a variable amount of the compressed gas is admitted to the distributor **33** by controlling the regulator valve **31** through the linkage **33** and the operator actuated throttle linkage **35**. The compressed gas is distributed to each of the lines **37** which lead to the individual cylinders **20**. The compressed gas is admitted to each of the cylinders **20** in timed relationship to the position of the pistons within the cylinders by opening the valves **39** with the valve actuator **41**.

When it is desired to increase the speed of the engine, the operator moves the throttle linkage **35** which simultaneously admits a larger quantity of compressed gas to the distributor **33** from the tank **23** by further opening the regulator valve **31**. The timing of the valve actuator **41** is also advanced through the linkage **104**. Still further, as the speed of the engine **21** increases, the effective length of the rotating contact **83** (FIG. 4) or **95** (FIG. 6) increases thereby electrically contacting a wider portion of one of the stationary radial contacts **85** to cause each of the valves **39** to remain open for a longer period of each engine cycle to admit a larger quantity of compressed gas to each of the cylinders **20**.

As can be seen, the combination of the regulating valve **31**, the mechanical advance **104**, and the valve actuator **41**, combine to produce a compressed gas engine which is quickly and efficiently adaptable to various operating speeds. However, all three of the controls need not be employed simultaneously. For example, the mechanical advance **104** could be utilized without the benefit of one of the varying valve actuators **41** but the high speed operation of the engine may not be as efficient. By increasing the duration of each engine cycle over which each of the valves **39** remains open to admit compressed gas to each of the cylinders **20** as the speed increases, conservation of compressed gas during low speed operation and efficient high speed operation are both possible.

After the compressed gas admitted to the cylinder **20** has forced the piston **22** downwardly within the cylinder to drive the shaft **139** of the engine, the piston **22** moves upwardly within the cylinder **20** and forces the expanded gas out through a suitable exhaust valve (not shown) through the adapter plate **103** (if employed) and into the exhaust line **105**. The cool exhaust can then be collected in any suitable arrangement to be compressed and returned to the tank **23** or used for any desired purpose including use as a working fluid in an air conditioning system or as a coolant for oil.

When using the apparatus and method of the present invention to adapt a ordinary internal combustion engine for operation with compressed gas it can be seen that considerable savings in weight are achieved. For example, the ordinary cooling system including a radiator, fan, hoses, etc. can be eliminated since the compressed gas is cooled as it expands in the cylinder. In addition, there are no explosions within the cylinder to generate heat. Further reductions in weight are obtained by employing plastic tubing for the lines which carry the compressed gas between the distributor and the cylinders and for the exhaust lines. Once again, heavy tubing is not required since there is little or no heat generated by the engine of the present invention. In addition, the noise generated by an engine according to the present invention is considerably less than that generated by an ordinary internal combustion engine since there are no explosions taking place within the cylinders.

The principles of preferred embodiments of the present invention have been described in the foregoing specification. However, the invention which is intended to be protected is not to be construed as limited to the particular embodiments disclosed. The embodiments are to be regarded as illustrative rather than restrictive. Variations and changes may be made by others without departing from the spirit of the invention. Accordingly, it is expressly intended that all such variations and changes which fall within the spirit and the scope of the present invention as defined in the appended claims be embraced thereby.

What is claimed is:

1. An apparatus for operating an engine having at least one cylinder and a reciprocating piston therein comprising:

a source of compressed gas;

distributor means connected with the source of compressed gas for distributing the compressed gas to the at least one cylinder;

valve means for admitting the compressed gas to the at least one cylinder when the piston is in approximately a top dead center position within the cylinder;

altering means for increasing the duration of each engine cycle over which the valve means admits compressed gas to the at least one cylinder as the speed of the engine increases; and

exhaust means for exhausting gas as the piston subsequently approaches approximately the top dead center position.

2. The apparatus of claim 1 further comprising control means for controlling the amount of compressed gas admitted to the distributor means.

3. The apparatus of claim 1 wherein the valve means is a solenoid valve secured in an opening in the cylinder above the level of the piston at the top dead center position.

4. The apparatus of claims 1 or 2 further comprising means for advancing the timing of the valve means as the speed of the engine increases such that compressed gas is admitted progressively further before the top dead center position as the speed of the engine increases.

5. The apparatus of claim 4 wherein the means for advancing the timing comprises a mechanical linkage connected to an operator actuated accelerator linkage.

6. The apparatus of claim 1 wherein a portion of the gas exhausted through the exhaust means is compressed in a compressor driven by an output shaft of the engine and is returned to the source of compressed gas.

7. The apparatus of claim 1 wherein a portion of the gas exhausted through the exhaust means is used to cool transmission fluid for a transmission associated with the engine.

8. The apparatus of claim 1 wherein a portion of the gas exhausted through the exhaust means is used as a working fluid in an air conditioning system.

9. The apparatus of claim 6 further comprising first gearing means interposed between the output shaft of the engine and the compressor for increasing the speed at which the compressor is driven.

10. The apparatus of claim 6 further comprising clutch means attached to the compressor both for disengaging the compressor from the output shaft of the engine when a first predetermined pressure at the source of compressed gas is exceeded and for engaging the compressor with the output shaft of the engine when the pressure at the source of compressed gas drops below a second predetermined pressure.

11. The apparatus of claim 9 further comprising means for both disengaging the first gearing means when a predetermined speed of the engine is exceeded and engaging a second gearing means for driving the compressor at a speed slower than the first gearing means when the predetermined speed of the engine is exceeded.

12. The apparatus of claim 1 wherein the valve means is electrically actuated and wherein the altering means comprises:

a rotating member timed with the at least one cylinder and arranged within a housing;

first and second contacts arranged on a first end of the rotating member and on an inside surface of the housing, respectively;

means for increasing the distance of the first contact from the rotational axis of the rotating member as the speed of the engine increases such that the first contact moves radially outwardly within the housing; and

said second contact presenting a longer arc length to the first contact as the distance of the first contact from the rotational axis of the rotating member increases.

13. The apparatus of claim 12 wherein the rotating member comprises an arcuate arm and wherein the means for increasing the distance of the first contact comprises pivotally mounting a second end of the arcuate arm about the axis of rotation of the rotating member and spring means for biasing the first end of the arcuate arm towards a radially inward position whereby the first end of the arcuate arm pivots radially outwardly as the speed of the engine increases.

14. The apparatus of claim 12 wherein the rotating member is axially slidably received on a rotating shaft for co-rotation therewith, said shaft having splines with a radial component in the direction of rotation, and wherein the first contact comprises a flexible contact located on an upper surface of the rotating member, said flexible contact being biased against the inside surface of the housing which carries the second contacts whereby as the speed of the engine increases the rotating member is urged axially along the splined shaft towards the inside surface of the housing such that the flexible contact is forced radially outwardly along the inside surface.

15. The apparatus of claim 12 wherein the second contact comprises of radially extending conductor arranged on an upper inside surface of the housing, said conductor increasing in arc length as the conductor extends radially outwardly from a central portion of the housing.

16. An apparatus for adapting an internal combustion engine for operation with compressed gas, the internal combustion engine having at least one cylinder, a piston reciprocable within the at least one cylinder, intake and exhaust means disposed in the at least one cylinder, and a tapped hole in the at least one cylinder adapted to receive a spark plug, the apparatus comprising:

a source of compressed gas;

distributor means connected with the source of compressed gas for distributing the compressed gas to the at least one cylinder;

valve means arranged in the tapped hole for admitting the compressed gas to the at least one cylinder when the piston is in approximately a top dead center position within the cylinder; and

altering means for increasing the duration of each engine cycle over which the valve means remains open to admit the compressed gas as the speed of the engine increases.

17. An apparatus as in claim 16 further comprising first adapter plate means for supporting the distributor means above an intake manifold of the engine, which adaptor plate means allows ambient air to enter through the intake manifold.

18. The apparatus of claim 16 further comprising second adapter plate means for reducing the exit area of the exhaust means.

19. A method of operating an engine on compressed gas, said engine having at least one cylinder and a piston reciprocable therein comprising the steps of:

delivering compressed gas from a source to a distributor;

distributing the compressed gas to the at least one cylinder;

admitting compressed gas to the at least one cylinder through an intake valve when the piston is at approximately a top dad center position;

increasing the duration of each engine cycle over which compressed gas is admitted to the at least one cylinder as the engine speed increases; and

exhausting the remaining gas when the piston subsequently reaches approximately the top dead center position.

20. The method of claim **19** further comprising the step of controlling the amount of compressed gas which is delivered to the distributor.

21. The method of claim **19** further comprising the step of advancing the timing of the opening of the intake valve as the speed of the engine increases.

22. An apparatus for operating an engine having at least one cylinder and a piston reciprocable therein on compressed gas comprising:

a source of compressed gas;

distributor means connected with the source of compressed gas for distributing the compressed gas to the at least one cylinder;

electrically actuated valve means secured in an opening in the at least one cylinder for selectively admitting compressed gas to the at least one cylinder when the piston is in approximately a top dead center position; and

means for advancing the timing of the valve means as the speed of the engine increases whereby compressed gas is admitted progressively further before the top dead center position as the speed of the engine increases.

* * * * *

United States Patent [19]

Rogers, Sr.

[11] **4,292,804**

[45] **Oct. 6, 1981**

---

[54] **METHOD AND APPARATUS FOR OPERATING AN ENGINE ON COMPRESSED GAS**

[76] Inventor: **Leroy K. Rogers, Sr.,** #5 Capistrano Ct., Ft. Myers, Fla. 33908

[21] Appl. No.: **158,303**

[22] Filed: **Jun. 10, 1980**

[51] **Int. Cl.**^{3} **F15B 11/06**

[52] **U.S. Cl.** **60/407**; 91/187; 91/275

[58] **Field of Search** 60/407, 412; 91/187, 91/275, 364

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,881,399	5/1975	Sagi et al.	91/187 X
3,885,387	5/1975	Simington	60/407 X
4,018,050	4/1977	Murphy	60/412 X

Primary Examiner—Allen M. Ostrager

Attorney, Agent, or Firm—Burns, Doane, Swecker & Mathis

[57] **ABSTRACT**

The present invention relates to a method and apparatus for operating an engine having a cylinder and a piston reciprocable therein on compressed gas. The apparatus comprises a source of compressed gas connected to a distributor which distributes the compressed gas to the cylinder. A valve is provided to selectively admit compressed gas to the cylinder when the piston is in an approximately top dead center position. In one embodiment of the present invention the timing of the opening of the valve is advanced such that the compressed gas is admitted to the cylinder progressively further before the top dead center position of the piston as the speed of the engine increases. In a further embodiment of the present invention a valve actuator is provided which increases the length of time over which the valve remains open to admit compressed gas to the cylinder as the speed of the engine increases. A still further embodiment of the present invention relates to an apparatus for adapting a conventional internal combustion engine for operation on compressed gas.

22 Claims, 8 Drawing Figures

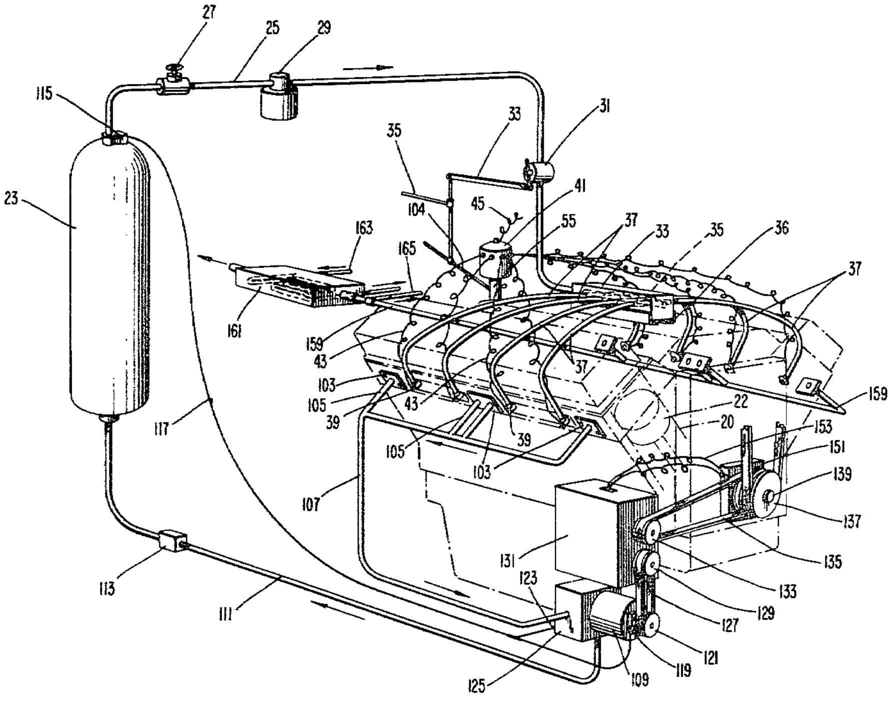

FIG. 1

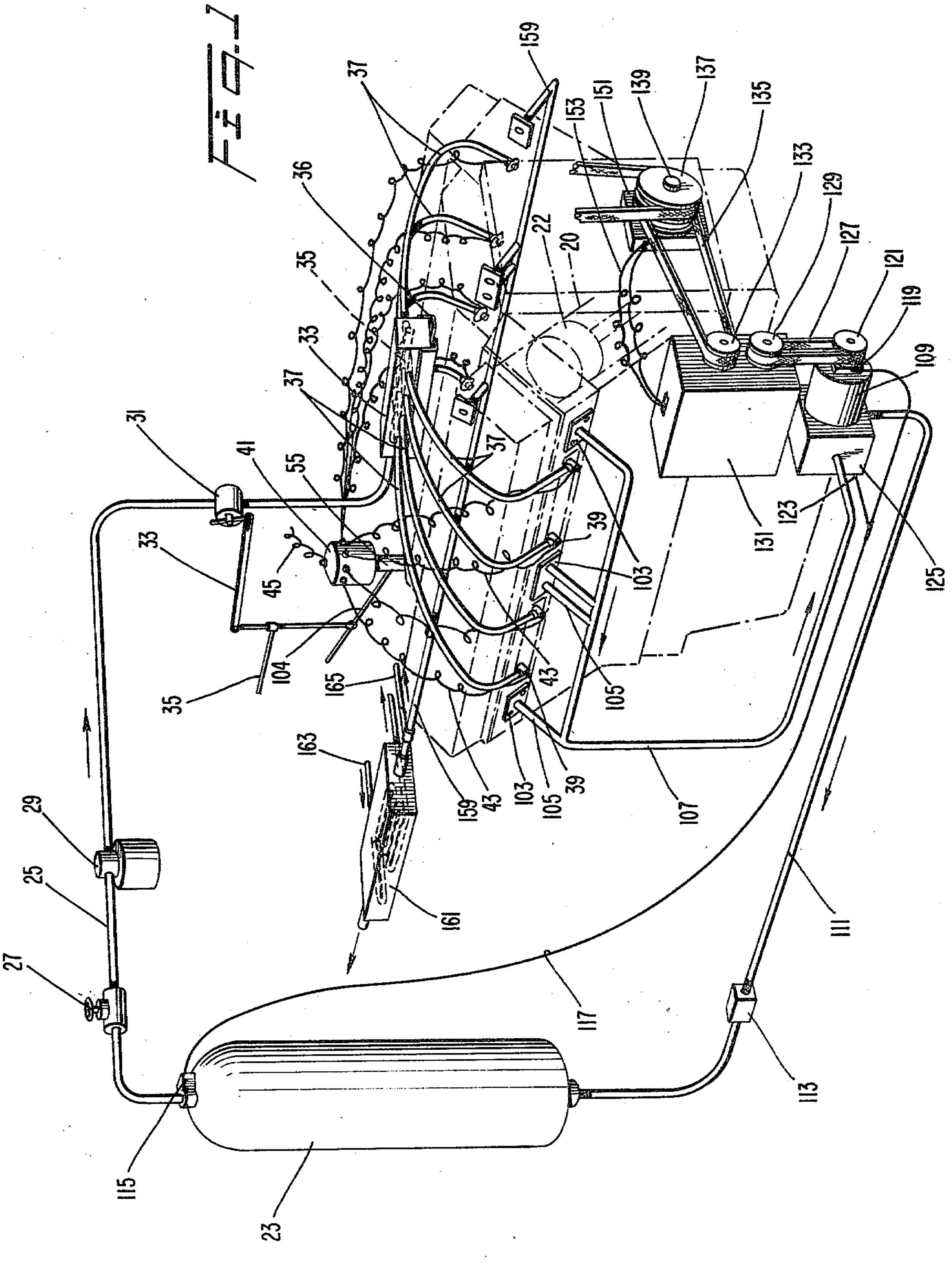

FIG. 6

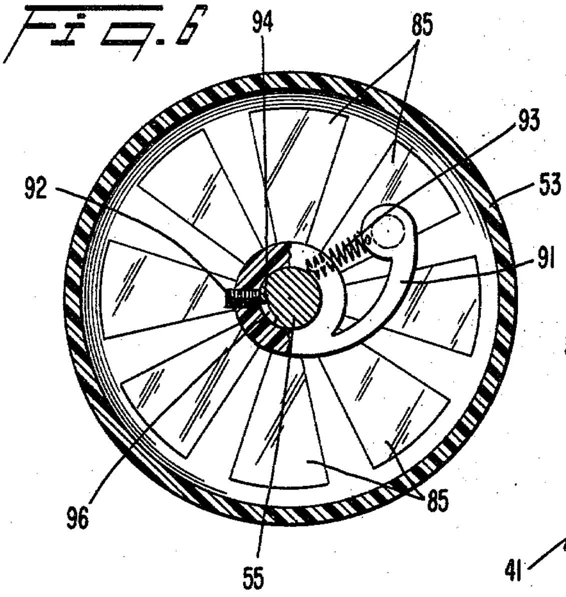
94
85
93
92
53
91
96
55
85

FIG. 7

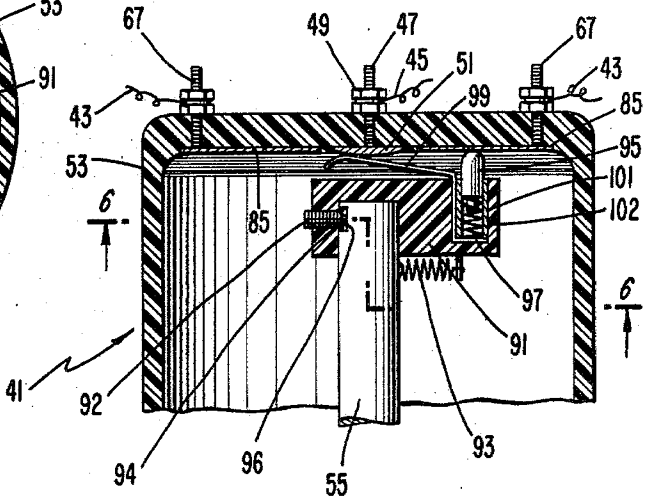
67
49
47
45
51
67
43
99
43
85
53
95
101
6
85
102
6
97
91
41
92
93
94
96
55
93

FIG. 3

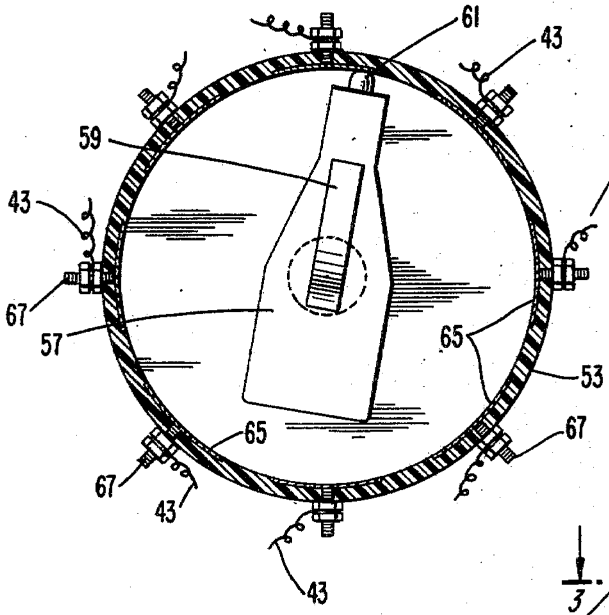
61
43
59
43
67
57
65
53
65
67
67
43
43

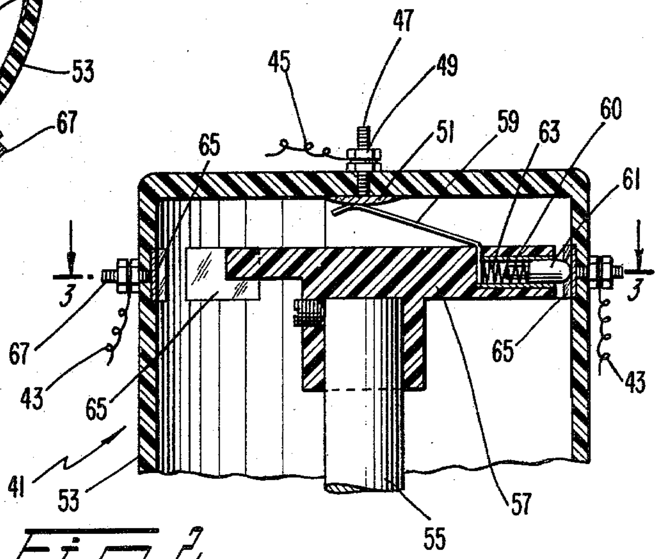
47
45
49
51
59
63
60
65
61
3
3
67
65
43
43
65
41
53
57
55

FIG. 2

Fig. 5

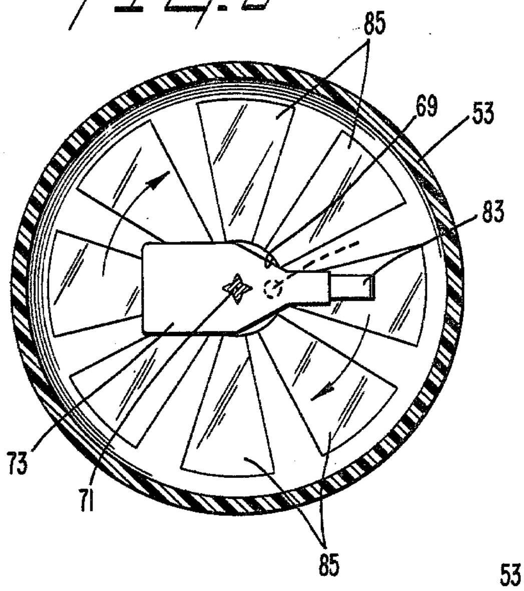
85
69
53
83
73
71
85

Fig. 4

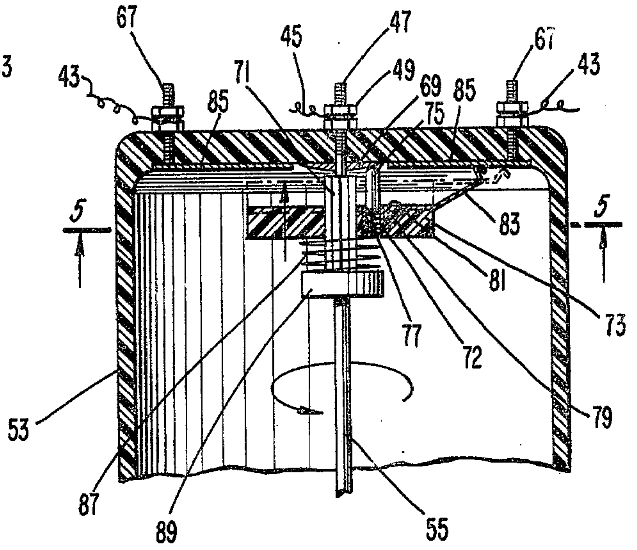
67
45
47
67
43
71
49
69
85
43
85
75
5
83
5
81
73
77
72
53
79
87
89
55

Fig. 6

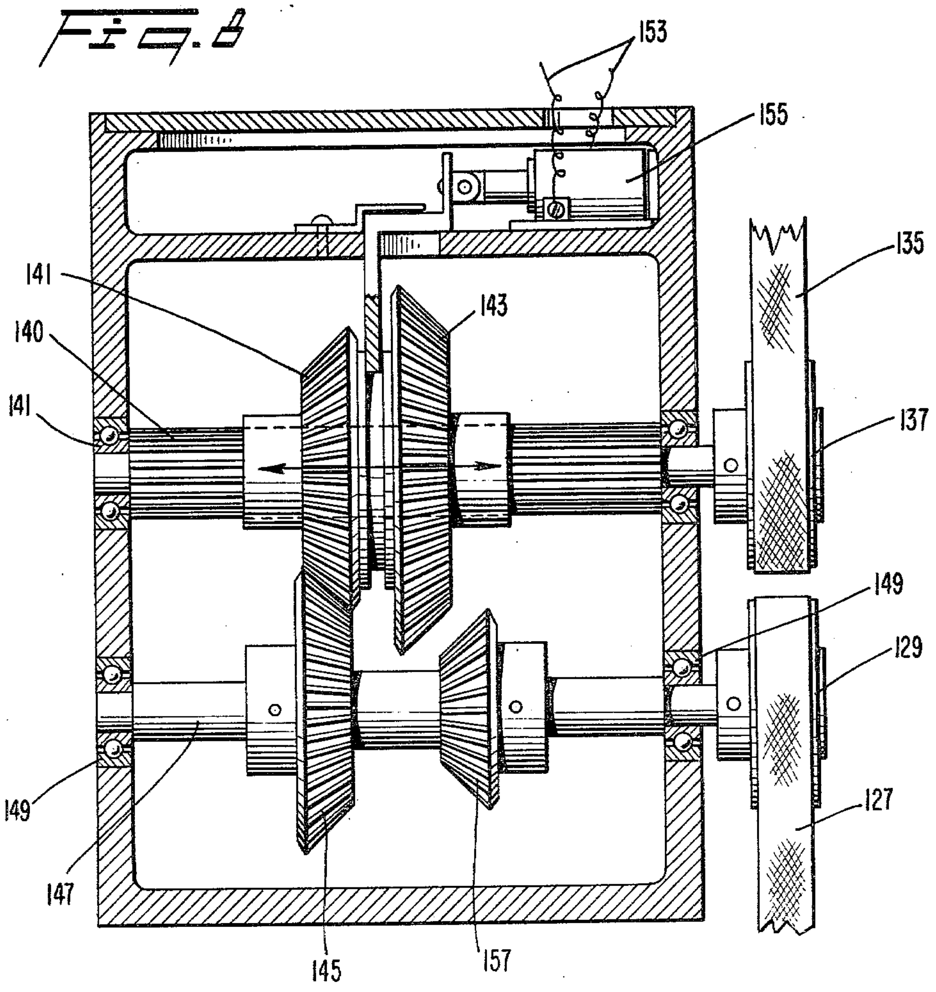
153
155
135
141
143
140
141
137
149
129
149
127
147
145
157

METHOD AND APPARATUS FOR OPERATING AN ENGINE ON COMPRESSED GAS

BACKGROUND AND SUMMARY OF THE PRESENT INVENTION

The present invention relates to a method and apparatus for operating an engine using a compressed gas as the motive fluid. More particularly, the present invention relates to a apparatus for adapting a pre-existing internal combustion engine for operation on a compressed gas.

Air pollution is one of the most serious problems facing the world today. One of the major contributors to air pollution is ordinary internal combustion engine which are used in most motor vehicles today. Various devices, including many items mandated by legislation, have been proposed in an attempt to limit the pollutants which an internal combustion engine exhausts to the air. However, most of these devices have met with limited success and are often both prohibitively expensive and complex. A clean alternative to the internal combustion engine is needed to power vehicles and other machinery.

A compressed gas, preferably air, would provide an ideal motive fluid for a engine since it would eliminate the usual pollutants exhausted from an internal combustion engine. An apparatus for converting an internal combustion engine for operation on compressed air is disclosed in U.S. Pat. No. 3,885,387 issued May 27, 1975 to Simington. The Simington patent discloses an apparatus including a source of compressed air and a rotating valve actuator which opens and closes a plurality of mechanical poppet valves. The valves deliver compressed air in timed sequence to the cylinders of an engine through adapters located in the spark plug holes. However, the output speed of an engine of this type is limited by the speed of the mechanical valves and the fact that the length of time over which each of the valves remains open cannot be varied as the speed of the engine increases.

Another apparatus for converting an internal combustion engine for operation on steam or compressed air is disclosed in U.S. Pat. No. 4,102,130 issued July 25, 1978 to Stricklin. The Stricklin patent discloses a device which changes the valve timing of a conventional four stroke engine such that the intake and exhaust valves open once for every revolution of the engine instead of once every other revolution of the engine. A reversing valve is provided which delivers live steam or compressed air to the intake valves and is subsequently reversed to allow the exhaust valves to deliver the expanded steam or air to the atmosphere. A reversing valve of this type however does not provide a reliable apparatus for varying the amount of motive fluid injected into the cylinders when it is desired to increase the speed of the engine. Further, a device of the type disclosed in the Stricklin patent requires the use of multiple reversing valves if the cylinders in a multi-cylinder engine were to be fired sequentially.

Therefore, it is an object of the present invention to provide a reliable method and apparatus for operating an engine or converting an engine for operation with a compressed gas.

A further object of the present invention is to provide a method and apparatus which is effective to deliver a constantly increasing amount of compressed gas to an engine as the speed of the engine increases.

A still further object of the present invention is to provide a method and apparatus which will operate an engine using compressed gas at a speed sufficient to drive a conventional automobile at highway speeds.

It is still a further object of the present invention to provide a method and apparatus which is readily adaptable to a standard internal combustion engine to convert the internal combustion engine for operation with a compressed gas.

Another object of the invention is to provide a method and apparatus which utilizes cool expanded gas, exhausted from a compressed gas engine, to operate an air conditioning unit and/or an oil cooler.

These and other objects are realized by a method and apparatus according to the present invention for operating an engine having at least one cylinder and a reciprocating piston therein using compressed gas as a motive fluid. The apparatus includes a source of compressed gas and a distributor connected with the source of the compressed gas for distributing the compressed gas to the at least one cylinder. A valve is provided for admitting the compressed gas to the cylinder when the piston is in approximately a top dead center position within the cylinder. An exhaust is provided for exhausting the expanded gas from the cylinder as the piston returns to approximately the top dead center position.

In a preferred embodiment of the present invention a device is provided for varying the duration of each engine cycle over which the valve remains open to admit compressed gas to the cylinder dependent upon the speed of the engine. In a further preferred embodiment of the present invention, an apparatus for advancing the timing of the opening of the valve is arranged to admit the compressed gas to the cylinder progressively further before the top dead center position of the piston as the speed of the engine increases.

Further features of the present invention include a valve for controlling the amount of compressed gas admitted to the distributor. Also, a portion of the gas which has been expanded in the cylinder and exhausted through the exhaust valve is delivered to a compressor to be recompressed and returned to the source of compressed gas. A gear train is selectively engagable to drive the compressor at different operating speed depending upon the pressure maintained at the source of compressed air and/or the speed of the engine. Still further, a second portion of the exhaust gas is used to cool a lubricating fluid for the engine or to operate an air conditioning unit.

In a preferred embodiment of the present invention, the valve for admitting compressed gas to the cylinder is electrically actuated. The device for varying the duration of each engine cycle over which the intake valve remains open as the speed of the engine increase comprises a rotating element whose effective length increases as the speed of the engine increases such that a first contact on the rotating element is electrically connected to a second contact for a longer period of each engine cycle. The second contact actuates the valve whereby the valve remains in an open position for a longer period of each engine cycle as the speed of the engine increases.

Still further features of the present invention include an adaptor plate for supporting the distributor above an intake manifold of a conventional internal combustion engine after a carburetor has been removed to allow air

to enter the cylinders of the engine through the intake manifold and conventional intake valves. Another adaptor plate is arranged over an exhaust passageway of the internal combustion engine to reduce the cross-sectional area of the exhaust passageway.

BRIEF DESCRIPTION OF THE DRAWINGS

Preferred embodiments of a method and apparatus for operating an engine according to the present invention will be described with reference to the accompanying drawings wherein like members bear like reference numerals and wherein:

FIG. 1 is a schematic representation of an apparatus according to the present invention arranged on an engine;

FIG. 2 is a side view of one embodiment of a valve actuator according to the present invention;

FIG. 3 is a cross-sectional view taken along the line 3—3 in FIG. 2;

FIG. 4 is a cross-sectional view of a second embodiment of a valve actuator according to the present invention;

FIG. 5 is a view taken along the line 5—5 in FIG. 4;

FIG. 6 is a cross-sectional view of a third embodiment of a valve actuator according to the present invention;

FIG. 7 is a view taken along the line 7—7 in FIG. 6;

FIG. 8 is a cross-sectional view of a gearing unit to drive a compressor according to the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

With reference to FIG. 1, an engine block **21** (shown in phantom) having two banks of cylinders with each bank including cylinders **20** having pistons **22** reciprocable therein (only one of which is shown in phantom) in a conventional manner. While the illustrated engine is a V-8 engine, it will be apparent that the present invention is applicable to an engine having any number of pistons and cylinders with the V-8 engine being utilized for illustration purposes only. A compressed gas tank **23** is provided to store a compressed gas at high pressure. It may also be desirable to include a small electric or gas compressor to provide compressed gas to supplement the compressed gas held in the tank **23**. In a preferred embodiment, the compressed gas is air which can be obtained from any suitable source.

A line **25** transports the gas withdrawn from the tank **23** when a conventional shut off valve **27** is open. In addition, a solenoid valve **29** preferably operated by a suitable key operated switch (not shown) for the engine is also arranged in the line **25**. In normal operation, the valve **27** is maintained open at all times with the solenoid valve **29** operating as a selective shut off valve to start and stop the engine **21** of the present invention.

A suitable regulating valve **31** is arranged downstream from the solenoid valve **29** and is connected by a linkage **33** to a throttle linkage **35** which is operator actuated by any suitable apparatus such as a foot pedal (not shown). The line **25** enters an end of a distributor **33** and is connected to an end of a pipe **35** which is closed at the other end. A plurality of holes, which are equal to the number of cylinders in the engine **21**, are provided on either side of the pipe **35** along the length of the pipe **35**.

When the present invention is used to adapt a conventional internal combustion engine for operation on compressed gas, an adaptor plate **36** is provided to support the distributor **33** in spaced relation from the usual intake opening in the intake manifold of the engine after a conventional carburetor has been removed. In this way, air is permitted to enter the internal combustion engine through the usual passageways and to be admitted to the cylinders through suitable intake valves (not shown). The adaptor plate **36** is secured to the engine block **21** and the distributor **33** by any suitable apparatus, e.g., bolts.

Each of the holes in the pipe **35** is connected in fluid-tight manner to a single line **37**. Each line **37** carries the compressed gas to a single cylinder **20**. In a preferred embodiment, each of the lines **37** is $\frac{1}{2}$ inch high pressure plastic tubing attached through suitable connectors to the distributor **33** and the pipe **35**. Each of the lines **37** is connected to a valve **39** which is secured in an opening provided near the top of each of the cylinders **20**. In the case of a conversion of a standard internal combustion engine, the valves **39** can be conveniently screwed into a tapped hole in the cylinder **20** typically provided for a spark plug of the internal combustion engine. In a preferred embodiment, the valves **39** are solenoid actuated valves in order to provide a fast and reliable opening and closing of the valves **39**.

Each of the valves **39** is energized by a valve actuator **41** through one of a plurality of wires **43**. The valve actuator **41** is driven by a shaft of the engine similar to the drive for a conventional distributor of an internal combustion engine. That is, a shaft **55** of the valve actuator **41** is driven in synchronism with the engine **21** at one half the speed of the engine **21**.

A first embodiment of the valve actuator **41** (FIGS. 2 and 3) receives electrical power through a wire **45** which is energized in a suitable manner by a battery, and a coil if necessary (not shown) as is conventional in an internal combustion engine. The wire **45** is attached to a central post **47** by a nut **49**. The post **47** is connected to a conducting plate **51** arranged within a housing **53** for the valve actuator **41**. Within the housing **53**, the shaft **55** has an insulating element **57** secured to an end of the shaft **55** for co-rotation therewith when the shaft **55** is driven by the engine **21**. A first end of a flexible contact **59** is continuously biased against the conducting plate **51** to receive electricity from the battery or another suitable source. A second end of the contact **59** is connected to a conducting sleeve **60** which is in constant contact with a spring biased contact **61** which is arranged within the sleeve **60**. The contact **61** is biased by a spring **63** which urges the contact **61** towards a side wall of the housing **53**.

With reference to FIG. 3, a plurality of contacts **65** are spaced from one another and are arranged around the periphery of the housing **53** at the same level as the spring biased contact **61**. Each contact **65** is electrically connected to a post **67** which extends outside of the housing **53**. The number of contacts **65** is equal to the number of cylinders in the engine **21**. One of the wires **43**, which actuate the valves **39**, is secured to each of the posts **67**.

In operation, as the shaft **55** rotates in synchronism with the engine **21**, the insulating element **57** rotates and electricity is ultimately delivered to successive ones of the contacts **65** and wires **43** through the spring biased contact **61** and the flexible contact **59**. In this way, each of the electrical valves **39** is actuated and opened in the proper timed sequence to admit compressed gas to each of the cylinders **20** to drive the pistons **22** therein on a downward stroke.

The embodiment illustrated in FIGS. 2 and 3 is effective to actuate each of the valves 39 to remain open for a long enough period of time to admit sufficient compressed gas to each of the cylinders 20 of the engine 21 to drive the engine 21. The length of each of the contacts 65 around the periphery of the housing 53 is sufficient to permit the speed of the engine to be increased when desired by the operator by moving the throttle linkage 35 which actuates the linkage 33 to further open the regulating valve 31 to admit more compressed gas from the tank 23 to the distributor 33. However, it has been found that the amount of air admitted by the valves 39 when using the first embodiment of the valve actuator 41 (FIGS. 2 and 3) is substantially more than required to operate the engine 21 at an idling speed. Therefore, it may be desirable to provide a valve actuator 41 which is capable of varying the duration of each engine cycle over which the solenoid valves 39 are actuated, i.e., remain open to admit compressed gas, as the speed of the engine 21 is varied.

A second embodiment of a valve actuator 41 which is capable of varying the duration of each engine cycle over which each of the valves 39 remains open to admit compressed gas to the cylinders 20 dependent upon the speed of the engine 21 will be described with reference to FIGS. 4 and 5 wherein members corresponding to those of FIGS. 2 and 3 bear like reference numerals. The wire 45 from the electrical source is secured to the post 47 by the nut 49. The post 47 has a annular contact ring 69 electrically connected to an end of the post 47 and arranged within the housing 53. The shaft 55 rotates at one half the speed of the engine as in the embodiment of FIGS. 2 and 3.

At an upper end of the shaft 55, a splined section 71 slidably receives an insulating member 73. The splined section 71 of the shaft 55 positively holds the insulating member 73 for co-rotation therewith but permits the insulating member 73 to slide axially along the length of the splined section 71. Near the shaft 55, a conductive sleeve 72 is arranged in a bore 81 in an upper surface of the insulating element 73 generally parallel to the splined section 71. A contact 75, biased towards the annular contact ring 69 by a spring 77, is arranged within the conductive sleeve 72 in contact therewith. The conductive sleeve 72 also contacts a conductor 79 at a base of the bore 81.

The conductor 79 extends to the upper surface of the insulating element 73 near an outer periphery of the insulating element 73 where the conductor 79 is electrically connected to a flexible contact 83. The flexible contact 83 selectively engages a plurality of radial contacts 85 arranged on an upper inside surface of the housing 53. A weak spring 87 arranged around the splined section 71 engages a stop member 89 secured on the shaft 55 and the insulating element 73 to slightly bias the insulating element 73 towards the upper inside surface of the housing 53 to ensure contact between the flexible contact 83 and the upper inside surface of the housing 53. As best seen in FIG. 5, the radial contacts 85 on the upper inside surface of the housing 53 are arranged generally in the form of radial spokes extending from the center of the housing 53 with the number of contacts being equal to the number of cylinders 20 in the engine 21. The number of degrees covered by each of the radial contacts 85 gradually increases as the distance from the center of the upper inside surface of the housing 53 increases.

In operation of the device of FIGS. 4 and 5, as the shaft 55 rotates, electricity flows along a path through the wire 45 down through post 47 to the annular contact member 69 which is in constant contact with the spring biased contact 75. The electrical current passes through the conductive sleeve 72 to the conductor 79 and then to the flexible contact 83. As the flexible contact 83 rotates along with the insulating member 73 and the shaft 55, the tip of the flexible contact 83 successively engages each of the radial contacts 85 on the upper inside of the housing 53. As the speed of the shaft 55 increases, the insulating member 73 and the flexible contact 83 attached thereto move upwardly along the splined section 71 of the shaft 55 due to the radial component of the splines in the direction of rotation under the influence of centrifugal force. As the insulating member 73 moves upwardly, the flexible contact 83 is bent such that the tip of the contact 83 extends further radially outwardly from the center of the housing 53 (as seen in phantom lines in FIG. 4). In other words, the effective length of the flexible contact 83 increases as the speed of the engine 21 increases.

As the flexible contact 83 is bent and the tip of the contact 83 moves outwardly, the tip remains in contact with each of the radial contacts 85 for a longer period of each engine cycle due to the increased angular width of the radial contacts with increasing distance from the center of the housing 53. In this way, the length of time over which each of the valves 39 remains open is increased as the speed of the engine is increased. Thus, a larger quantity of compressed gas or air is injected into the cylinders as the speed increases. Conversely, as the speed decreases and the insulating member 73 moves downwardly along the splined section 71, a minimum quantity of air is injected into the cylinder due to the shorter length of the individual radial contact 85 which is in contact with the flexible contact 83. In this way, the amount of compressed gas that is used during idling of the engine 21 is at a minimum whereas the amount of compressed gas which is required to increase the speed of the engine 21 to a level suitable to drive a vehicle on a highway is readily available.

With reference to FIGS. 6 and 7, a third embodiment of a valve actuator 41 according to the present invention includes an arcuate insulating element 91 having a first end pivotally secured by any suitable device such as screw 92 to the shaft 55 for co-rotation with the shaft 55. The screw 92 is screwed into a tapped hole in the insulating element 91 such that a tab 94 at an end of the screw 92 engages a groove 96 provided in the shaft 55. In this way, the insulating element 91 positively rotates with the shaft 55. However, as the shaft 55 rotates faster, a second end 98 of the insulating element 91 is permitted to pivot outwardly under the influence of centrifugal force because of the groove 96 provided in the shaft 55. A spring 93 connected between the second end 98 of the element 91 and the shaft 55 urges the second end of the element 91 towards the center of the housing 53.

A contact 99 similar to the contact 59 (FIG. 2) is arranged such that one end of the contact 99 is in constant contact with the conducting plate 51 located centrally within the housing 53. The other end of the contact 99 engages a conductive sleeve 101 arranged in bore 102. A contact element 95 is arranged in the conductive sleeve 101 in constant contact with the sleeve 101. The bore 102 is arranged generally parallel to the shaft 55 near the second end of the arcuate insulating

element **91**. The contact **95** is biased by a spring **97** towards the upper inside surface of the housing **53** for selective contact with each of the plurality of radial contacts **85** which increase in arc length towards the outer peripheral surface of the housing **53** (FIG. **6**).

In operation of the device of FIGS. **6** and **7**, as the shaft **55** rotates the arcuate insulating element **91** rotates with the shaft **55** and the second end **98** of the insulating element **91** tends to pivot about the shaft **55** due to centrifugal force. Thus, as the effective length of the contact **95** increases, i.e., as the arcuate insulating element **91** pivots further outwardly, the number of degrees of rotation over which the contact **95** is in contact with each of the radial contacts **85** on the upper inside surface of the housing **53** increases thereby permitting each of the valves **39** to remain open for a longer period of each engine cycle to admit more compressed gas to the respective cylinder **20** to further increase the speed of the engine **21**.

With reference to FIG. **1**, a mechanical advance linkage **104** which is connected to the throttle linkage **35**, advances the initiation of the opening of each valve **39** such that compressed gas is injected into the respective cylinder further before the piston **22** in the respective cylinder **20** reaches a top dead center position as the speed of the engine is increased by moving the throttle linkage **35**. The advance linkage **104** is similar to a conventional standard mechanical advance employed on an internal combustion engine. In other words, the linkage **104** varies the relationship between the angular positions of a point on the shaft **55** and a point on the housing **53** containing the contacts. Alternatively, a conventional vacuum advance could also be employed. By advancing the timing of the opening of the valves **39**, the speed of the engine can more easily be increased.

The operation of the engine cycle according to the present invention will now be described. The compressed gas injected into each cylinder of the engine **21** drives the respective piston **22** downward to drive a conventional crankshaft (not shown). The movement of the piston downwardly causes the compressed gas to expand rapidly and cool. As the piston **22** begins to move upwardly in the cylinder **20** a suitable exhaust valve (not shown) arranged to close an exhaust passageway is opened by any suitable apparatus. The expanded gas is then expelled through the exhaust passageway. As the piston **22** again begins to move downwardly a suitable intake valve opens to admit ambient air to the cylinder. The intake valve closes and the ambient air is compressed on the subsequent upward movement of the piston until the piston reaches approximately the top dead center position at which time the compressed gas is again injected into the cylinder **20** to drive the piston **22** downward and the cycle begins anew.

In the case of adapting a conventional internal combustion engine for operation on compressed gas, a plurality of plates **103** are preferably arranged over an end of the exhaust passageways in order to reduce the outlet size of the exhaust passageways of the conventional internal combustion engine. In the illustrated embodiment, a single plate having an opening in the center is bolted to the outside exhaust passageway on each bank of the V-8 engine while another single plate having two openings therein is arranged with one opening over each of the interior exhaust passageways on each bank of the V-8 engine. A line **105** is suitably attached to each of the adaptor places to carry the exhaust to an appropriate location. In a preferred embodiment, the exhaust lines **105** are $1\frac{1}{2}''$ plastic tubing.

In a preferred embodiment, the exhaust lines **105** of one bank of the V-8 engine are collected in a line **107** and fed to an inlet of a compressor **109**. The pressure of the exhaust gas emmanating from the engine **21** according to the present invention is approximately 25 p.s.i. In this way, the compressor **109** does not have to pull the exhaust into the compressor since the gas exhausted from the engine **21** is at a positive pressure. The positive pressure of the incoming fluid increases the efficiency and reduces wear on the compressor **109**. The exhaust gas is compressed in the compressor **109** and returned through a line **111** and a check valve **113** to the compressed gas storage tank **23**. The check valve **113** prevents the flow of compressed gas stored in the tank **23** back towards the compressor **109**.

A suitable pressure sensor **115** is arranged at an upper end of the tank **23** and sends a signal along a line **117** when the pressure exceeds a predetermined level and when the pressure drops below a predetermined level. The line **117** controls an electrically actuated clutch **119** disposed at a front end of the compressor **109**. The clutch **119** is operative to engage and disengage the compressor **109** from a drive pulley **121**. Also, the signal carried by the line **117** actuates a suitable valve **123** arranged on a compressor housing **125** to exhaust the air entering the compressor housing **125** from the line **107** when the clutch **119** has disengaged the compressor **109** from the drive pully **121**.

In a preferred embodiment, when the pressure is the tank **23** reaches approximately 600 p.s.i., the clutch **119** is disengaged and the compressor **109** is deactivated and the valve **123** is opened to exhaust the expanded gas delivered to the compressor **109** from the line **107** to the atmosphere. When the pressure within the tank **23** drops below approximately 500 p.s.i., the sensor **115** sends a signal to engage the clutch **119** and close the valve **123**, thereby operating the compressor **109** for supplying the tank **23** with compressed gas.

The pulley **121** which drives the compressor **109** through the clutch **119** is driven by a belt **127** which is driven by a pulley **129** which operates through a gear box **131**. With reference to FIGS. **1** and **8**, a second pulley **133** on the gear box is driven by a belt **135** from a pulley **137** arranged on a drive shaft **139** of the engine **21**. The pulley **137** drives a splined shaft **140** which has a first gear **141** and a second larger gear **143** arranged thereon for rotation with the splined shaft **140**. The splined shaft **140** permits axial movement of the gears **141** and **143** along the shaft **140**.

In normal operation (as seen in FIG. **8**), the first gear **141** engages a third gear **145** arranged on a shaft **147** which drives the pulley **129**. The shafts **140** and **147** are arranged in suitable bearings **149** arranged at each end thereof. When the speed of the engine **21** drops below a predetermined level, a suitable sensor **151** responsive to the speed of the drive shaft **139** of the engine **21** generates a signal which is transmitted through a line **153** to a solenoid actuator **155** arranged within the gear box **131**. The solenoid actuator **155** moves the first and second gears **141**, **143** axially along the splined shaft **140** to the right as seen in FIG. **8** such that the second, larger gear **143** engages a fourth smaller gear **157** which is arranged on the shaft **147**. The ratio of the second gear **143** to the fourth gear **157** is preferably approximately 3 to 1.

In this way, when the speed of the engine **21** drops below the predetermined level as sensed by the sensor **151** (which predetermined level is insufficient to drive the compressor **109** at a speed sufficient to generate the 500–600 pounds of pressure which is preferably in the tank **23**), the solenoid actuator **155** is energized to slide the gears **143**, **141** axially along the splined shaft **140** so that the second, larger gear **143** engages the fourth, smaller gear **157** to drive the pulley **129** and hence the compressor **109** at a higher rate of speed to generate the desired pressure. When the speed of the engine increases above the predetermined level, in a preferred embodiment approximately 1500 rpm, the solenoid actuator **155** is deactivated by the sensor **151** thereby moving the gears **143** and **141** to the left as seen in FIG. **8** such that the first gear **141** re-engages with the third gear **145** to effectuate a 1 to 1 ratio between the output shaft **139** of the engine **21** and the pulley **129**.

The other bank of the V-8 engine has its exhaust ports arranged with adapter plates **103** similar to those on the first bank. However, the exhaust from this bank of the engine **21** is not collected and circulated through the compressor **109**. In a preferred embodiment, a portion of the exhaust is collected in a line **159** and fed to an enlarged chamber **161**. A second fluid is fed through a line **163** into the chamber **161** to be cooled by the cool exhaust emmanating from the engine **21** in the line **159**. The second fluid in the line **163** may be either transmission fluid contained in a transmission associated with the engine **21** or a portion of the oil used to lubricate the engine **21**. A second portion of the exhaust from the second bank of the V-8 engine is removed from the line **159** in a line **165** and used as a working fluid in an air conditioning system or for any other suitable use.

It should be noted that the particular arrangement utilized for collecting and distributing the gas exhausted from the engine **21** would be determined by the use for which the engine is employed. In other words, it may be advantageous to rearrange the exhaust tubing such that a larger or smaller percentage of the exhaust is routed through the compressor **109**. It should also be noted that since the exhaust lines **105** are plastic tubing, a rearrangement of the lines for a different purpose is both simple and inexpensive.

In operation of the engine of the present invention, the engine **21** is started by energizing the solenoid valve **29** and any suitable starting device (not shown), e.g., a conventional electric starter as used on an internal combustion engine. Compressed gas from the full tank **23** flows through the line **25** and a variable amount of the compressed gas is admitted to the distributor **33** by controlling the regulator valve **31** through the linkage **33** and the operator actuated throttle linkage **35**. The compressed gas is distributed to each of the lines **37** which lead to the individual cylinders **20**. The compressed gas is admitted to each of the cylinders **20** in timed relationship to the position of the pistons within the cylinders by opening the valves **39** with the valve actuator **41**.

When it is desired to increase the speed of the engine, the operator moves the throttle linkage **35** which simultaneously admits a larger quantity of compressed gas to the distributor **33** from the tank **23** by further opening the regulator valve **31**. The timing of the valve actuator **41** is also advanced through the linkage **104**. Still further, as the speed of the engine **21** increases, the effective length of the rotating contact **83** (FIG. 4) or **95** (FIG. 6) increases thereby electrically contacting a wider portion of one of the stationary radial contacts **85** to cause each of the valves **39** to remain open for a longer period of each engine cycle to admit a larger quantity of compressed gas to each of the cylinders **20**.

As can be seen, the combination of the regulating valve **31**, the mechanical advance **104**, and the valve actuator **41**, combine to produce a compressed gas engine which is quickly and efficiently adaptable to various operating speeds. However, all three of the controls need not be employed simultaneously. For example, the mechanical advance **104** could be utilized without the benefit of one of the varying valve actuators **41** but the high speed operation of the engine may not be as efficient. By increasing the duration of each engine cycle over which each of the valves **39** remains open to admit compressed gas to each of the cylinders **20** as the speed increases, conservation of compressed gas during low speed operation and efficient high speed operation are both possible.

After the compressed gas admitted to the cylinder **20** has forced the piston **22** downwardly within the cylinder to drive the shaft **139** of the engine, the piston **22** moves upwardly within the cylinder **20** and forces the expanded gas out through a suitable exhaust valve (not shown) through the adapter plate **103** (if employed) and into the exhaust line **105**. The cool exhaust can then be collected in any suitable arrangement to be compressed and returned to the tank **23** or used for any desired purpose including use as a working fluid in an air conditioning system or as a coolant for oil.

When using the apparatus and method of the present invention to adapt a ordinary internal combustion engine for operation with compressed gas it can be seen that considerable savings in weight are achieved. For example, the ordinary cooling system including a radiator, fan, hoses, etc. can be eliminated since the compressed gas is cooled as it expands in the cylinder. In addition, there are no explosions within the cylinder to generate heat. Further reductions in weight are obtained by employing plastic tubing for the lines which carry the compressed gas between the distributor and the cylinders and for the exhaust lines. Once again, heavy tubing is not required since there is little or no heat generated by the engine of the present invention. In addition, the noise generated by an engine according to the present invention is considerably less than that generated by an ordinary internal combustion engine since there are no explosions taking place within the cylinders.

The principles of preferred embodiments of the present invention have been described in the foregoing specification. However, the invention which is intended to be protected is not to be construed as limited to the particular embodiments disclosed. The embodiments are to be regarded as illustrative rather than restrictive. Variations and changes may be made by others without departing from the spirit of the invention. Accordingly, it is expressly intended that all such variations and changes which fall within the spirit and the scope of the present invention as defined in the appended claims be embraced thereby.

What is claimed is:

1. An apparatus for operating an engine having at least one cylinder and a reciprocating piston therein comprising:

a source of compressed gas;

distributor means connected with the source of compressed gas for distributing the compressed gas to the at least one cylinder;

valve means for admitting the compressed gas to the at least one cylinder when the piston is in approximately a top dead center position within the cylinder;

altering means for increasing the duration of each engine cycle over which the valve means admits compressed gas to the at least one cylinder as the speed of the engine increases; and

exhaust means for exhausting gas as the piston subsequently approaches approximately the top dead center position.

2. The apparatus of claim **1** further comprising control means for controlling the amount of compressed gas admitted to the distributor means.

3. The apparatus of claim **1** wherein the valve means is a solenoid valve secured in an opening in the cylinder above the level of the piston at the top dead center position.

4. The apparatus of claims **1** or **2** further comprising means for advancing the timing of the valve means as the speed of the engine increases such that compressed gas is admitted progressively further before the top dead center position as the speed of the engine increases.

5. The apparatus of claim **4** wherein the means for advancing the timing comprises a mechanical linkage connected to an operator actuated accelerator linkage.

6. The apparatus of claim **1** wherein a portion of the gas exhausted through the exhaust means is compressed in a compressor driven by an output shaft of the engine and is returned to the source of compressed gas.

7. The apparatus of claim **1** wherein a portion of the gas exhausted through the exhaust means is used to cool transmission fluid for a transmission associated with the engine.

8. The apparatus of claim **1** wherein a portion of the gas exhausted through the exhaust means is used as a working fluid in an air conditioning system.

9. The apparatus of claim **6** further comprising first gearing means interposed between the output shaft of the engine and the compressor for increasing the speed at which the compressor is driven.

10. The apparatus of claim **6** further comprising clutch means attached to the compressor both for disengaging the compressor from the output shaft of the engine when a first predetermined pressure at the source of compressed gas is exceeded and for engaging the compressor with the output shaft of the engine when the pressure at the source of compressed gas drops below a second predetermined pressure.

11. The apparatus of claim **9** further comprising means for both disengaging the first gearing means when a predetermined speed of the engine is exceeded and engaging a second gearing means for driving the compressor at a speed slower than the first gearing means when the predetermined speed of the engine is exceeded.

12. The apparatus of claim **1** wherein the valve means is electrically actuated and wherein the altering means comprises:

a rotating member timed with the at least one cylinder and arranged within a housing;

first and second contacts arranged on a first end of the rotating member and on an inside surface of the housing, respectively;

means for increasing the distance of the first contact from the rotational axis of the rotating member as the speed of the engine increases such that the first contact moves radially outwardly within the housing; and

said second contact presenting a longer arc length to the first contact as the distance of the first contact from the rotational axis of the rotating member increases.

13. The apparatus of claim **12** wherein the rotating member comprises an arcuate arm and wherein the means for increasing the distance of the first contact comprises pivotally mounting a second end of the arcuate arm about the axis of rotation of the rotating member and spring means for biasing the first end of the arcuate arm towards a radially inward position whereby the first end of the arcuate arm pivots radially outwardly as the speed of the engine increases.

14. The apparatus of claim **12** wherein the rotating member is axially slidably received on a rotating shaft for co-rotation therewith, said shaft having splines with a radial component in the direction of rotation, and wherein the first contact comprises a flexible contact located on an upper surface of the rotating member, said flexible contact being biased against the inside surface of the housing which carries the second contacts whereby as the speed of the engine increases the rotating member is urged axially along the splined shaft towards the inside surface of the housing such that the flexible contact is forced radially outwardly along the inside surface.

15. The apparatus of claim **12** wherein the second contact comprises of radially extending conductor arranged on an upper inside surface of the housing, said conductor increasing in arc length as the conductor extends radially outwardly from a central portion of the housing.

16. An apparatus for adapting an internal combustion engine for operation with compressed gas, the internal combustion engine having at least one cylinder, a piston reciprocable within the at least one cylinder, intake and exhaust means disposed in the at least one cylinder, and a tapped hole in the at least one cylinder adapted to receive a spark plug, the apparatus comprising:

a source of compressed gas;

distributor means connected with the source of compressed gas for distributing the compressed gas to the at least one cylinder;

valve means arranged in the tapped hole for admitting the compressed gas to the at least one cylinder when the piston is in approximately a top dead center position within the cylinder; and

altering means for increasing the duration of each engine cycle over which the valve means remains open to admit the compressed gas as the speed of the engine increases.

17. An apparatus as in claim **16** further comprising first adapter plate means for supporting the distributor means above an intake manifold of the engine, which adaptor plate means allows ambient air to enter through the intake manifold.

18. The apparatus of claim **16** further comprising second adapter plate means for reducing the exit area of the exhaust means.

19. A method of operating an engine on compressed gas, said engine having at least one cylinder and a piston reciprocable therein comprising the steps of:

delivering compressed gas from a source to a distributor;

distributing the compressed gas to the at least one cylinder;

admitting compressed gas to the at least one cylinder through an intake valve when the piston is at approximately a top dad center position;

increasing the duration of each engine cycle over which compressed gas is admitted to the at least one cylinder as the engine speed increases; and

exhausting the remaining gas when the piston subsequently reaches approximately the top dead center position.

20. The method of claim **19** further comprising the step of controlling the amount of compressed gas which is delivered to the distributor.

21. The method of claim **19** further comprising the step of advancing the timing of the opening of the intake valve as the speed of the engine increases.

22. An apparatus for operating an engine having at least one cylinder and a piston reciprocable therein on compressed gas comprising:

a source of compressed gas;

distributor means connected with the source of compressed gas for distributing the compressed gas to the at least one cylinder;

electrically actuated valve means secured in an opening in the at least one cylinder for selectively admitting compressed gas to the at least one cylinder when the piston is in approximately a top dead center position; and

means for advancing the timing of the valve means as the speed of the engine increases whereby compressed gas is admitted progressively further before the top dead center position as the speed of the engine increases.

* * * * *

United States Patent [19]
Rogers, Sr.

[11] **Patent Number: 4,693,669**
[45] **Date of Patent: Sep. 15, 1987**

[54] **SUPERCHARGER FOR AUTOMOBILE ENGINES**

[76] Inventor: **Leroy K. Rogers, Sr.,** Rte. 13, P.O. Box 815-DD, Briarcliff Rd., Fort Myers, Fla. 33908

[21] Appl. No.: **717,652**

[22] Filed: **Mar. 29, 1985**

[51] **Int. Cl.⁴** **F04D 17/02;** F04D 29/40

[52] **U.S. Cl.** .... **415/143;** 415/199.6; 415/140; 415/219 C

[58] **Field of Search** .... 415/131, 132, 140, 142, 415/143, 198.1, 199.1, 199.4, 199.6, 207, 219 C

[56] **References Cited**

U.S. PATENT DOCUMENTS

759,584	5/1904	Wyland	415/199.6 X
1,585,218	5/1926	Watanabe	415/131 X
2,013,079	9/1935	Slocum	415/143 X
2,408,918	10/1946	Curtis	415/199.6 X
2,418,184	4/1947	McConaghy	415/199.6 X
2,570,081	10/1951	Szczeniowski	415/143 X
2,625,883	1/1953	Howser	415/131 X
2,681,645	6/1954	Hoke, Jr.	123/559
2,767,549	10/1956	Martin	415/140 X
2,933,884	4/1960	Foster	415/199.6 X
3,046,901	7/1962	Montgomery	415/143
3,143,103	8/1964	Zuhn	415/199.6 X
3,155,044	4/1960	Lattanzio et al.	415/142
3,370,417	2/1968	Koziara	123/559 X
3,692,420	9/1972	Mittelstaedt	415/62
3,775,023	11/1973	Davis et al.	415/143 X
3,918,828	11/1975	Kumm	415/61
3,918,852	11/1975	Carter	415/143 X
3,924,963	12/1975	Zerrer	415/143
3,935,847	2/1976	Gamell	123/119 CF
4,292,804	10/1981	Rogers, Sr.	60/407

FOREIGN PATENT DOCUMENTS

3012406	10/1981	Fed. Rep. of Germany .

Primary Examiner—Robert E. Garrett
Assistant Examiner—Joseph M. Pitko
Attorney, Agent, or Firm—Burns, Doane, Swecker & Mathis

[57] **ABSTRACT**

A supercharger for delivering supercharged air to an engine, comprising a shrouded axial compressor, a radial compressor which is located downstream of the axial compressor and a housing. The housing comprising four sections, including a section defining a highly convergent, frustoconical transition duct which favorably directs the discharge of the axial compressor to the inlet of the radial compressor and a hollow, highly convergent, exhaust cone section immediately downstream of the radial compressor which converges into the exhaust port of the supercharger. An annular flow deflector is provided for directing the discharge of the radial compressor into the exhaust cone.

7 Claims, 8 Drawing Figures

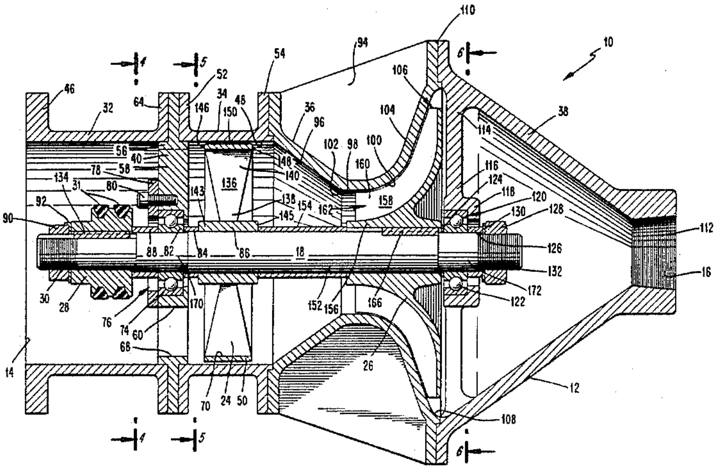
110
94
10
54
52
34
48
46
32
64
146
150
36
96
106
104
38
114
56
40
78
58
134
31
80
102
98
100
160
148
140
136
143
138
145
154
162
158
116
124
118
120
130
128
92
90
88
82
84
86
18
112
126
132
16
172
122
166
156
152
170
30
28
76
74
60
68
14
70
24
50
26
12
108
4
5
6

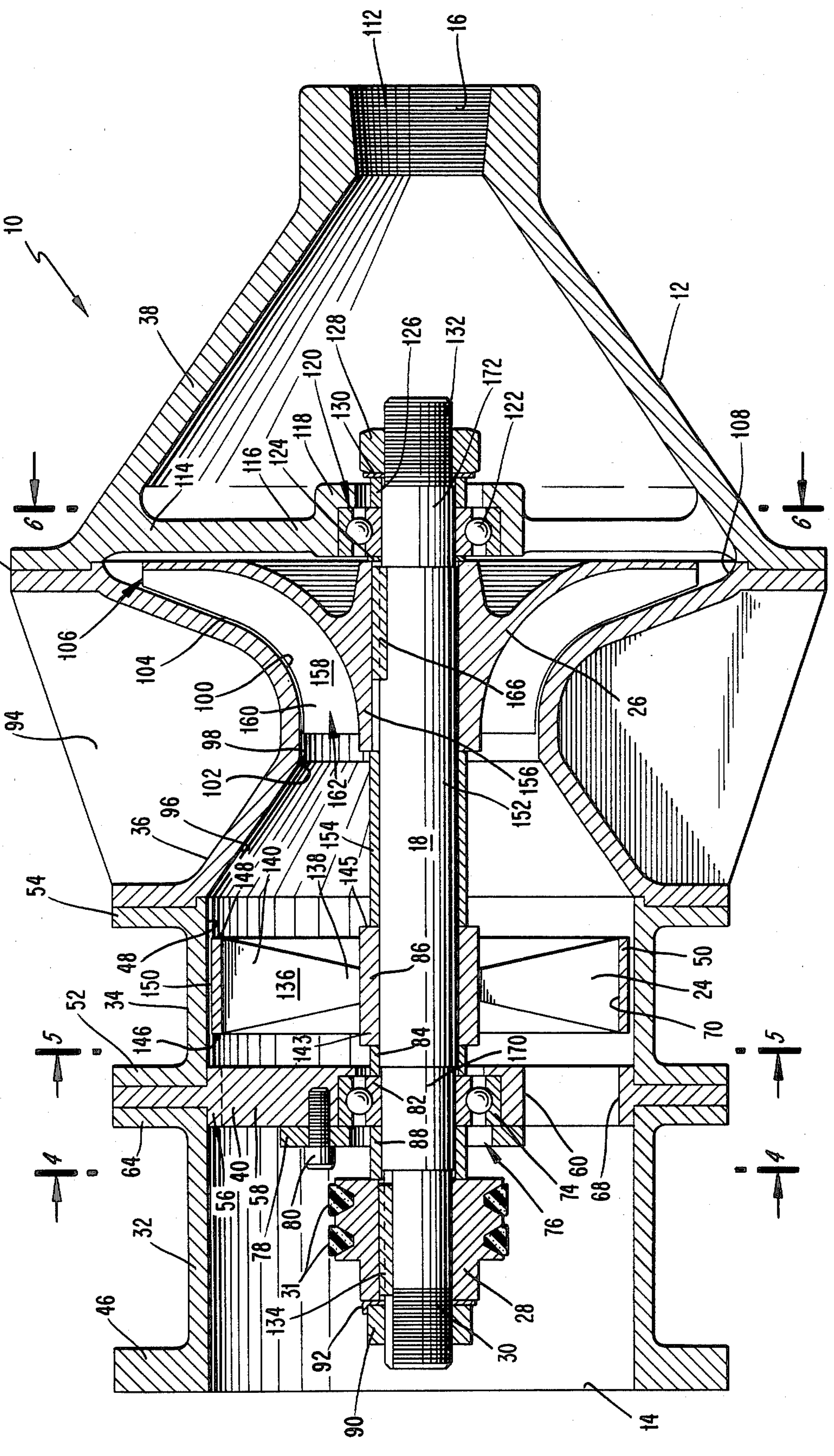

Fig-1

Fig. 2

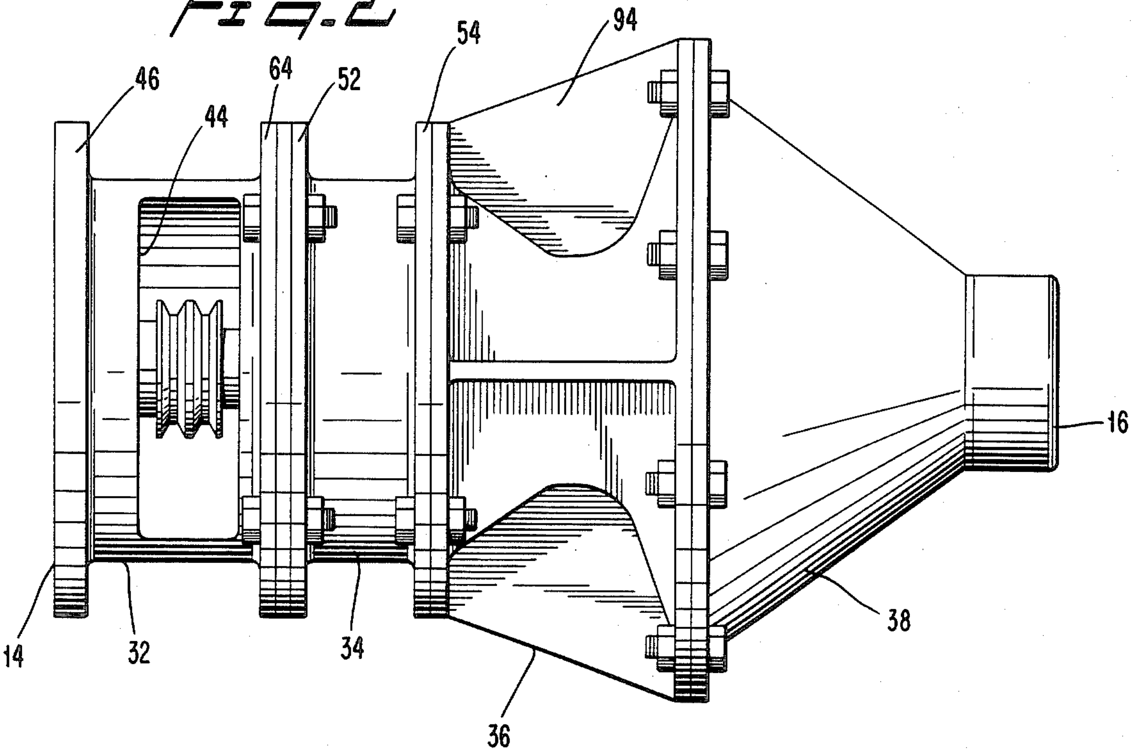
94
54
64
52
46
44
16
38
14
32
34
36

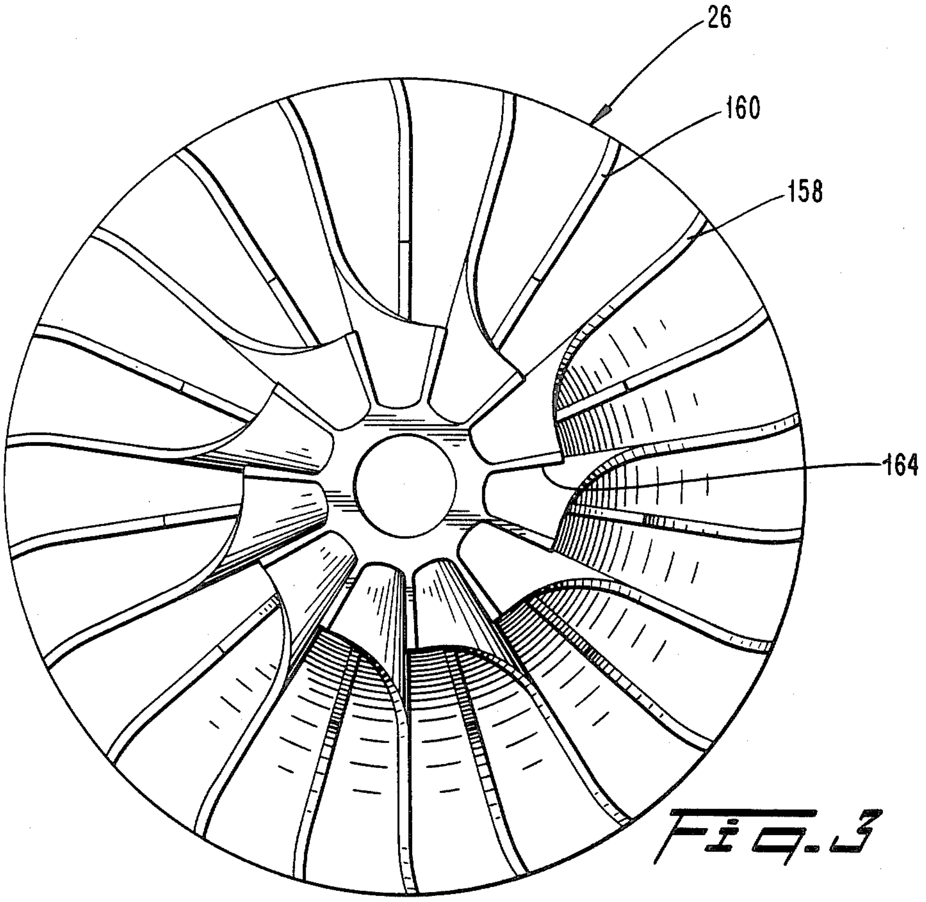
26
160
158
164

Fig. 3

Fig. 4

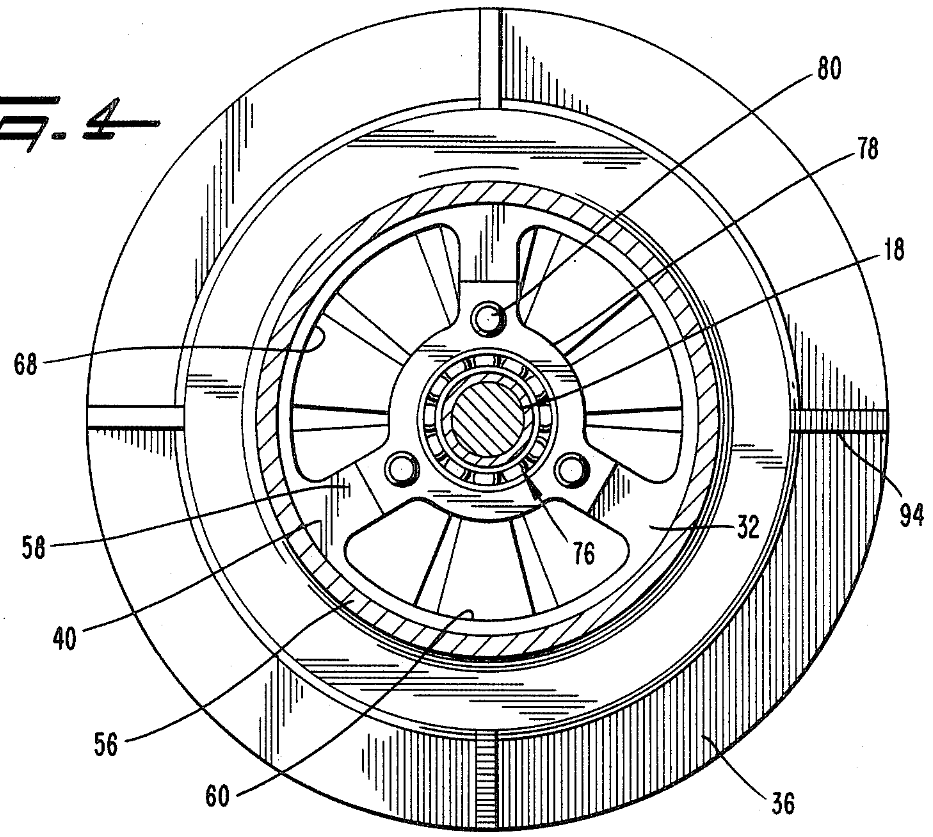

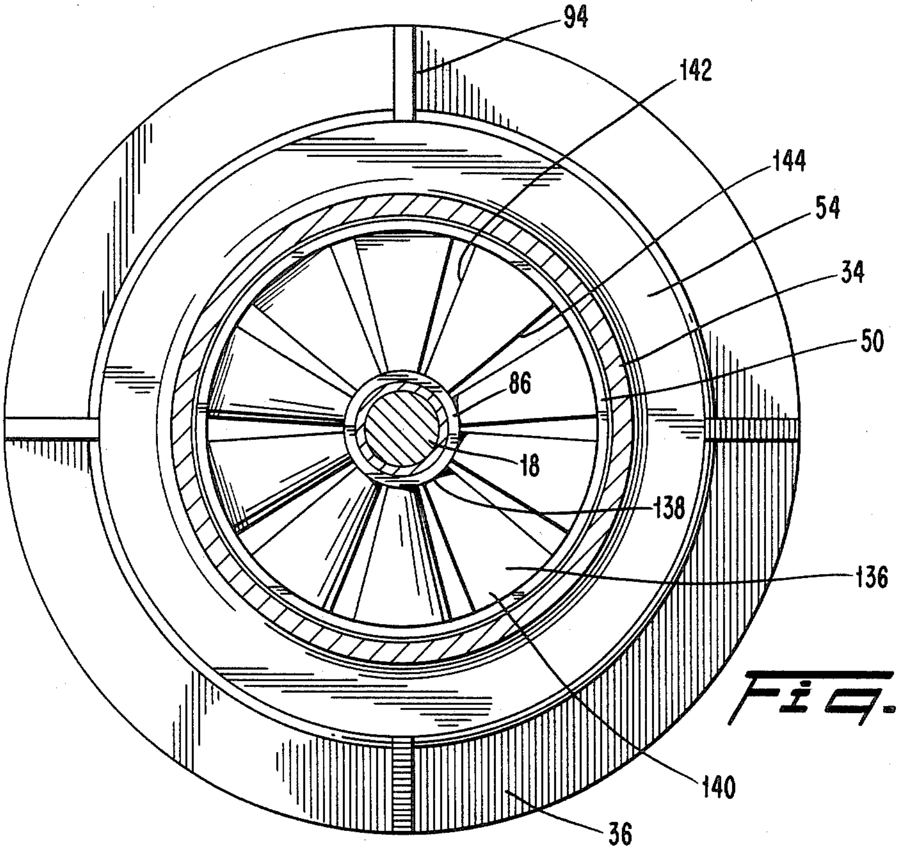

Fig. 5

Fig. 6

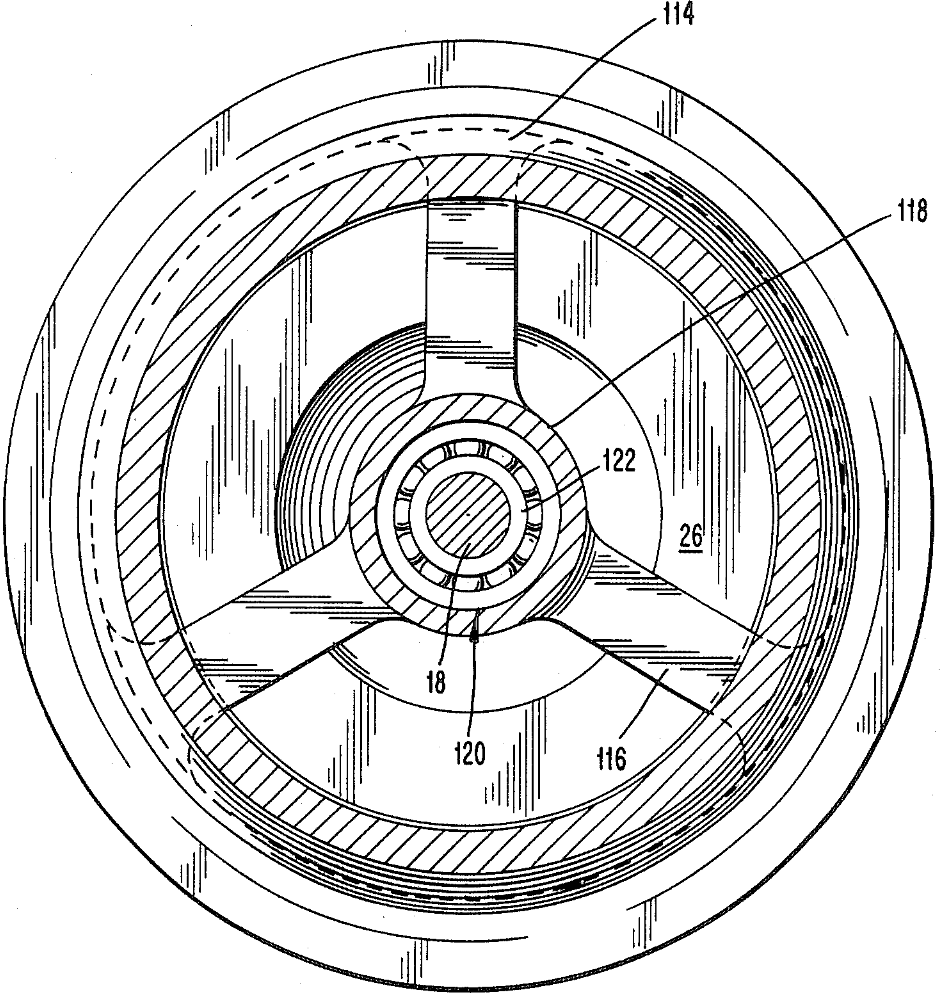
114
118
122
26
18
120
116

Fig. 7

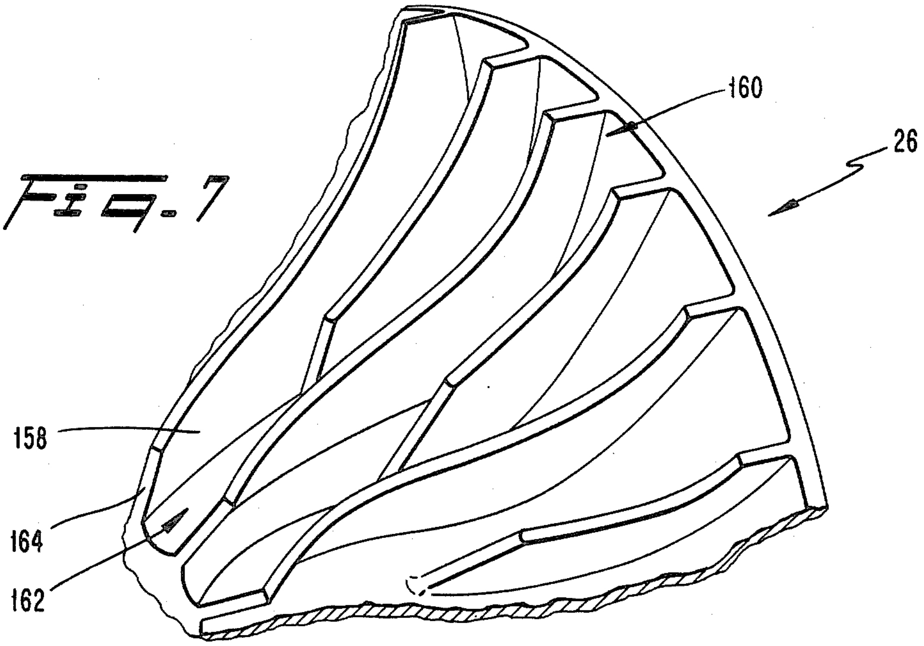
160
26
158
164
162

Fig. 6

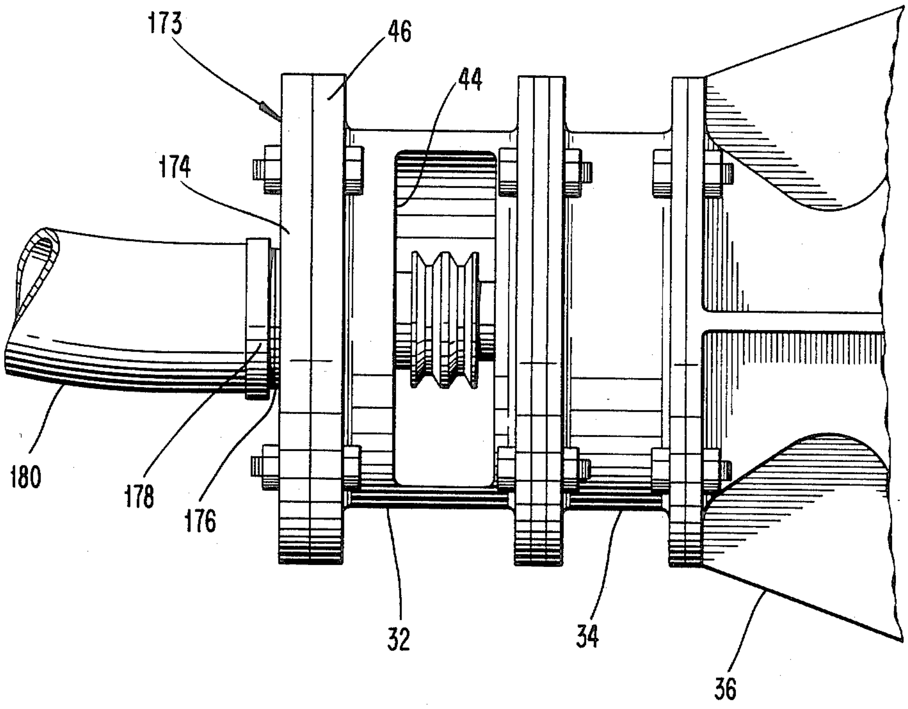
173
46
44
174
180
178
176
32
34
36

SUPERCHARGER FOR AUTOMOBILE ENGINES

FIELD OF THE INVENTION

The present invention relates to superchargers generally and more particularly to superchargers for automotive engines which include both an axial stage compressor and a centrifugal stage compressor.

BACKGROUND OF THE INVENTION

Superchargers impart additional pressure to the air or the air/fuel mixture of an engine so that the cylinders receive a greater weight per unit volume of air or air/fuel mixture than would otherwise be supplied. As a result, the volumetric efficiency and power output of the engine are improved.

According to prior practices, superchargers generally comprise a single airblower which forces air on an air/fuel mixture into the cylinders of an engine. Typically, the airblower is driven by a gear train which is connected to the crankshaft of the engine with a gear ratio of about 6 to 1. These prior types of superchargers have been used extensively in racing engines and radial aircraft engines. However, by reason of their high operating speeds and their gear trains, these superchargers have been considered too complicated, too heavy and too costly for use with mass production engines such as are found in automobiles and trucks.

Recently, some automobile manufacturers have been offering turbocharged engines which expand to exhaust gases of the engine through a turbine to drive a centrifugal compressor. Although turbochargers are advantageous in that the turbine can deliver large amounts of power to the compressor, their extreme operating speeds require special bearings, lubrication and maintenance. In addition, turbochargers require special ducting, such as by-pass arrangements, which only add to their cost and maintenance requirements. Consequently, turbochargers are only offered as expensive options in automobiles.

Further, there is current interest in a new type of automobile engine which operates from tanks of compressed gas to effect reciprocation of its pistons. An example of such an engine can be found in the U.S. Pat. No. 4,292,804 issued to the same inventor of the present invention. In the referenced patent, at least a portion of the partially expanded exhaust gas from the cylinders is directed to a compressor where it is recompressed and then returned to the storage tanks from whence it originally came. It would be desirable that at least some, if not all of the aforementioned recompression of the exhaust gas could be achieved with a belt-driven, rotary supercharger that is easily manufactured and maintained, yet is capable of providing ample recompression.

OBJECTS OF THE INVENTION

Accordingly, an object of the present invention is to provide a supercharger suitable for improving the performance of engines of automobiles, helicoptors or the like, which supercharger is inexpensive to produce and easy to maintain.

It is another object of the present invention to provide a supercharger which provides sufficient boost without resort to extreme operating speeds and accordingly avoids the costly complications associated with high speed operation.

It is yet another object of the present invention to provide a relatively compact and lightweight supercharger which is inexpensive to manufacture and maintain.

Another object of the present invention is to provide a belt-driven supercharger having a design which provides supercharging compression at relatively low operating speeds.

It is still another object of the present invention to provide a supercharger which can be quite readily disassembled and reassembled for purposes of low cost maintenance and repair.

Still another object of the present invention is to provide a supercharger which can be constructed from mass productible parts to thereby reduce the cost of its manufacture.

It is still another object of the present invention to provide a belt-driven supercharger which provides supercharging compression without resort to a larger number of compressor stages.

Yet another object of the present invention is to provide a rotary supercharger for a gas operated engine, which supercharger is easily manufactured and maintained, yet capable of providing ample recompression of the recirculating drive fluid.

SUMMARY OF THE INVENTION

These and other objects are achieved by the present invention which provides a supercharger comprising a housing having an inlet and an outlet, a shrouded axial compressor and a radial compressor rotatably mounted within the housing, a highly convergent shallow, frustoconical transition duct for favorably directing the discharge of the axial compressor to the inlet of the radial compressor.

In accordance with a further aspect of the invention, the above-described supercharger further comprises an exhaust cone at a location downstream of the radial compressor and a flow deflector for directing the discharge of the radial compressor to the exhaust cone.

In the preferred embodiment, the housing itself comprises four sections: a cylindrical front housing section which defines an axially directed inlet; a second, cylindrical ducting section enclosing the axial compressor; a rear housing section defining the transition duct as well as the inlet and casing for the radial compressor; and the exhaust cone section which defines at its terminis the outlet of the housing. For driving the compressor shaft, a double-tracked pulley wheel is secured to the forward end of the common shaft, which pulley wheel is adapted to receive one or more drive belts from the crank-shaft wheel of the engine. A lateral opening in the front housing section accommodates the connection with the drive belts.

With the disclosed arrangement, compression can be achieved for supercharging purposes without resort to a large number of compressor stages or high operating speeds. Additionally, the design of the disclosed supercharger avoids the need for guide vanes between the axial compressor and the radial compressor. The exhaust cone section also favorably avoids the build-up of back pressure against the radial compressor. The design is also very simple and therefore inexpensive to manufacture and maintain.

Other objects, advantages and novel features of the present invention will become apparent from the following detailed description of the invention when considered in conjunction with the accompanying drawing.

BRIEF DESCRIPTION OF THE DRAWINGS

A preferred embodiment of the present invention is described in greater detail with reference to the accompanying drawing wherein like elements bear like reference numerals, and wherein:

FIG. 1 is a cross-sectional side view of a supercharger constructed in accordance with the preferred embodiment of the present invention;

FIG. 2 is a side view of the supercharger of FIG. 1;

FIG. 3 is a frontal view of the impeller of the supercharger of FIG. 1;

FIG. 4 is a cross-sectional view taken along line 4—4 of FIG. 1;

FIG. 5 is a cross-sectional view taken along line 5—5 in FIG. 1;

FIG. 6 is a cross-sectional view taken along line 6—6 in FIG. 1;

FIG. 7 is a perspective view of a segment of the impeller of the supercharger of FIG. 1; and

FIG. 8 is a partial side view of the supercharger of FIG. 1 with an adaptor.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring to FIGS. **1** and **2**, a supercharger **10** is provided for supplying supercharged air to an automobile engine or the like, so that the engine receives a greater weight per unit volume of air or a fuel/air mixture than would be otherwise supplied. In accordance with a preferred embodiment of the present invention, the supercharger **10** comprises a housing **12** having an axially directed inlet **14** for receiving ambient air and an axially directed outlet **16** for delivering supercharged air to the intake of the automobile engine. Rotatably mounted within the housing **12** is a shaft **18** on which are secured an axial compressor **24** and a radial compressor **26**, the radial compressor **26** being positioned downstream of the axial compressor **24**. A pulley wheel **28** is secured to a forward end **30** of the shaft for receiving drive belts **31**, which belts drivingly connect the shaft **18** to a pulley wheel on the crankshaft of the engine (not shown). The drive belts **31** deliver torque to the shaft **18** as required for driving the compressors **24** and **26** of the supercharger **10**.

The housing **12** itself is constructed from four sections which are preferably bolted together at flanged connections in end-to-end relationship. These sections include a front housing section **32**, an axial compressor duct section **34**, a rear housing section **36** and an exhaust cone section **38**. The shaft **18** extends along the longitudinal axis of the housing **12**.

The front housing section **32** is a hollow cylinder which extends forward of a front bearing support **40**. The front housing section **32** encloses the forward end **30** of the shaft **18** and the associated pulley wheel **28**. At its forward end, the front housing section **32** defines the inlet **14** for receiving air from an external source (not shown). Referring particularly to FIG. **2**, the front housing section **32** includes a lateral opening **44** on one side in order to accommodate the connection of the drive belts **31** to the pulley wheel **28**. The front housing section **32** also includes a forward flange **46** for accommodating the connection of air filters, carburetors, air scoops or the like upstream of the supercharger **10** according to the particular engine layout.

It is to be understood that in the usual engine layout, the supercharger **10** receives air or a fuel/air mixture from an external source through its inlet **14**, compresses the air or fuel/air mixture and then delivers it to the intake of the engine.

Referring again to FIG. **1**, the pulley wheel **28** is interference-fitted upon the forward end **30** of shaft **18** and a key **134** is used to lock the pulley wheel **28** in place. The pulley wheel **28** is preferably a double-track design which is suitable for the attachment of twin drive belts, although a single-belt type pulley wheel would be adequate. The pulley wheel **28** is preferably sized such that the ratio of its diameter with respect to the diameter of the drive wheel of the engine's crankshaft provides an effective gearing ratio in the range of approximately two and one-half to four and one-half. Thusly at idle, when the automobile engine is running approximately 700 rpm, the supercharge **10** is running at approximately 2,400 rpm, and at cruise, when the engine is running in the range of 2,500, the supercharger **10** is preferably turning over in the range of 6,000 to 8,000 rpm. It is to be noted that although the diameter of the pulley wheel **28** may be substantially reduced in order to achieve a desired gearing ratio, the double-track wheel **28** presents a sufficient sum total of surface area to avoid slippage of the belts **31**.

The next adjacent section of housing **12** is the axial compressor duct **34** comprising a short cylinder which is coaxially disposed about the axial compressor **24**. Preferably, the axial compressor duct **34** is constructed from cast aluminum, with the interior surfaces **48** machined to assure uniform clearance between the duct **34** and a shroud **50** of the axial compressor **24**. As with other sections of the housing **12**, the axial compressor duct **34** is provided with flanges **52** and **54** for effecting connection to the adjacent housing sections. The axial compressor duct **34** guides air delivered from the front housing section **32** toward the axial compressor **24**.

Referring now to FIGS. **1** and **4**, a front bearing support **40** is interposed between the front housing section **32** and the axial compressor duct **34**. The front bearing support **40** includes an outer annulus **56** and three radially directed arms **58** between which arms are defined passages **60** for allowing air to pass through the bearing support **40**. The outer annulus **54** is secured by bolts connecting a rear flange **64** of the front housing section **32** and the flange **52** of the axial compressor duct **34**. By such arrangement, the front bearing support **40** is rigidly secured to the housing **12** such that loads and shocks to the shaft **18** can be transferred through the front bearing support **40** to the housing **12**.

In the preferred embodiment, the outer annulus **56** of the bearing support **40** extends into the region of the inlet **14** of the front housing section **32** such that its inner rim **68** coincides with the inner rim **70** of the shroud **50** of the axial compressor **24**. In this fashion the outer annulus **56** contributes to the guiding of the flow of air toward the axial compressor **24**.

An outer raceway **74** of the front roller bearing assembly **76** is secured between the front bearing support **40** and a bearing retainer plate **78**, which plate **78** is secured by removable bolts **80**. In the preferred embodiment, the front bearing assembly **76** is of the sealed, high speed type. A suitable commercially available bearing assembly is marketed under model Fafnir 405KDD. A lower raceway **82** of the front bearing assembly **76** is preferably secured to the shaft **18** with an interference fit. A spacer **84** is provided on one side of the lower raceway **82**, which spacer **84** also abuts a hub **86** of the axial compressor **24** to thereby place the axial compressor **24** a predetermined distance downstream of the bearing support **40**. Similarly, a spacer **88** is provided on the other side of the lower raceway **84**, which spacer **88** also abuts the pulley wheel **28** so as to space apart the pulley wheel **28** from the front bearing support **40** to assure sufficient clearance between same.

It is to be appreciated that the bearing retainer plate **78** allows ready access to the front bearing assembly **76** for purposes of maintenance or repair. To service the front bearing assembly **76**, a nut **90** and lock-washer **92** on the forward end **30** of the shaft **18** are loosened and removed together with the pulley wheel **28** and the spacer **88**. Then bolts **42** and the bearing retainer plate **76** are removed, leaving the whole bearing assembly **76** exposed for servicing and/or removal.

The rear housing section **36** is connected by bolts to the downstream end of the axial compressor duct **34**. Preferably, the rear housing section **36** is constructed from a single section of cast aluminum and includes external longitudinal ribs **94** for enhancing the structural rigidity of the rear housing section **34**. The walls of the rear housing section **36** define three elements of the supercharger **10**: a highly conical transition duct **96** which favorably directs the output of the axial compressor to an inlet **98** of the radial compressor **26**; the inlet **98** of the radial compressor **26**, itself; and a casing **100** for the radial compressor **26**.

The transition duct **96** is a hollow, frustoconical portion having a half-apex angle (from the generatrix to the axis of symmetry) of approximately 35°. The angle is selected such that the inlet to the radial compressor **26** is as close as possible to the outlet of the axial compressor **24** without causing undue back-pressure. In the preferred embodiment, the transition duct **96** begins a short distance downstream of the axial compressor **24** and ends at the beginning of the inlet **98** of the radial compressor **26**. The highly conical shape of the transition duct **96** is believed to roll-in the higher volume of air being discharged from the more radially outward portions of the axial compressor **24**. This rolling-in action is believed to promote a favorable flow regime at the inlet **98** of the radial compressor **26** such that the need for inlet guide vanes for the radial compressor **26** is avoided. It is also believed that the highly conical shape of the transition duct **96** affects upstream flow conditions at the axial compressor **24** such that its performance is improved. It has also been found that the need for a stator (or exit guide vane) for the axial compressor **24** is likewise avoided.

In essence, it is believed that the transition duct **96** performs the functions of the aforementioned exit vanes of axial compressors and inlet guide vanes of radial compressors, but without the pressure losses commonly associated with them. Because of the avoidance of these pressure losses and by reason of the expected improvement in the performance of the axial compressor, the supercharger **10** is able to impart a higher overall pressure ratio than would otherwise be achieved without the transition duct **96**. As a result, adequate compression is achieved at moderate operating speeds without resort to a bank of several axial compressors. It is to be understood, however, that when connecting the supercharger **10** to a relatively slowly recziprating diesel or a very large engine, it may be desirable to include two or more axial compressors in order to boost the supercharger's overall pressure ratio. In such case, practice of the present invention would include the placement of a transition duct as above described downstream of at least the last axial compressor.

At the inlet **98** of the radial compressor **26**, the walls of the rear housing **36** are cylindrical and coaxially disposed about the shaft **18**. It is to be noted that in the preferred embodiment, the surface transition **102** from the transition duct **96** to the inlet **98** is rounded-off.

The casing portion **100** of the rear housing section **36** closely follows the contour defined by blade edges **104** of the radial compressor **26** in a close, substantially sealing manner as is well known in the art of radial compressors. The casing portion **100** of the rear housing section **78** channels air between the rotating blades of the radial compressor **26** so that the blades can impart work to the passing air. The casing portion **100** also defines a discharge outlet **106** for the radial compressor **26**.

Just beyond the discharge outlet **106** of the radial compressor **26**, the interior surfaces of the rear housing section **36** begin to curve immediately inwardly to provide a transition into the next adjacent section of the housing **12**, the exhaust cone **38**. In this fashion, the interior surfaces at the rear-most portion of rear housing section **36** and those of the forward portion of the exhaust cone **92** define internally a flow deflector **108**. In the preferred embodiment, the flow deflector **108** is closely and concentrically disposed about the outlet **106** of the radial compressor **26** such that the air being discharged from the radial compressor **26** does not have the opportunity to diffuse significantly prior to its arrival at the annular flow deflector **108**. The annular flow deflector **108** directs the output of the radial compressor **26** into the exhaust cone **38** by providing a smooth surface transition from the interior of rear housing section **36** to the interior of the exhaust cone **38**.

The exhaust cone **38** is a highly convergent, hollow, frustoconical section placed immediately downstream of the radial compressor **26** for receiving the output of the radial compressor **26** from the annular flow deflector **108**. In the preferred embodiment, the exhaust cone **38** is a single section of cast aluminum which is joined to the downstream end of the rear housing section **36** at a flanged joint **110**. Preferably, the exhaust cone **92** converges according to a half-apex angle of approximately 35° and defines the exhaust port **16** at its terminus. Threading **112** at the exhaust port **16** accommodates the attachment of the appropriate external ducting (not shown) leading to the intake of the engine.

During operation of the supercharger **10**, the space enclosed by the exhaust cone **92** prevents the build up of an elevated back pressure which might otherwise arise and detract from the operation and efficiency of the radial compressor **26**. The enclosed space of the exhaust cone **92** is also of sufficient volume to absorb pulses and to average out unsteady flow conditions so to promote a smooth and continuous output from the supercharger **10**.

Referring now to FIGS. **1** and **6**, the exhaust cone **38** includes a rear bearing support **114** which comprises members **116** which extend radially inwardly from the outer walls of the exhaust cone **38**. At a radial inward location close to the shaft **18**, the members **116** converge to form a cupped annulus which serves as a housing **118** for the rear bearing assembly **120**. The housing **118** is open towards the rear face of the radial compressor **24** to facilitate disassembly of the supercharger **10**. The rear bearing assembly **120** is the same type and size as the front bearing assembly **76**. The inner race **122** of the bearing assembly **120** is set in place on the shaft **18** by spacers **124** and **126** in conjunction with a nut **128** and washer **130** on the rearward end **132** of the shaft **18**. In the preferred embodiment, the members **116** are integrally formed with the walls of the exhaust cone **38**.

Referring to FIGS. **1** and **5**, the axial compressor **24** upon rotation draws air through the inlet **14** and imparts an initial amount compression to the air as it forces the air into the transition duct **96** of the rear housing section **36**. In the preferred embodiment, the axial compressor **24** comprises a hub **86**, the shroud **50** and a series of ten (10) equally spaced, radially disposed blades **136**. Preferably, each blade **136** increases in cord from a root **138** to a tip **140** and includes a trailing edge **142** and a leading edge **144**, which edges are both slightly curved. The blades gradually increase in pitch from approximately 12° at the root **138** to approximately 36° at the tips **140**. However, the particular values of pitch and other geometrical aspects of the blades **136** might be varied in accordance with different operating speeds or other parameters as would be apparent to one skilled in the pertinent art and familiar with this disclosure.

The axial compressor **24** is preferably constructed from a single, cast aluminum section with the faces **143** and **145** of the hub **86** being machined for purposes of achieving accurate, axial positioning of the axial compressor **24** on the shaft **18** relative to the housing **12**. The faces **146** and **148** of the shroud **72** are also machined flat. Additionally, the outer periphery **150** of the shroud is machined to assure uniform clearance between the shroud and the adjacent interior surfaces **48** of the axial compressor duct **34**. The axial compressor **24** is preferably secured to the shaft **18** by an interference-fit onto a stepped portion **152** of the shaft **18**. The spacers **84** and **154** axially position the axial compressor **24** relative to the front bearing support **40** and the radial compressor **26**, respectively.

Dynamic balance test machines of the conventional type may be used to test the balance of the axial compressor **24** prior to its installation. If an imbalance is detected, material can be removed at the outer periphery **150** of the shroud **50** so as to achieve proper balance.

Referring now to FIGS. **1**, **3**, and **7**, the radial compressor **26** is constructed from a single section of cast aluminum and includes a hub **156** and curved blades **158**. Interposed between each pair of blades **158** are a second set of blades **160** which terminate short of the intake **162** of the radial compressor **26** so that the intake **162** is not crowded by both sets of blades. Accordingly, the radial compressor **26** features both a large total number of blades and an intake of relatively small diameter, which features enhance the performance of the compressor **26**. In the region of the intake **162**, the blades **158** present leading edges **164** and undergo a twist into the direction of rotation so as to prevent a favorable angle of attack at the intake **162**.

Preferably, the radial compressor **26** is positioned upon the stepped section **128** of the shaft **18** with an interference-fit and locked against rotational slippage by a key **166**. The spacer **124** assures clearance between the rear face of the radial compressor **26** and the rear bearing assembly **120**.

The shaft **18** is constructed from a hardened steel and is threaded at both ends **30** and **132** for receiving nuts **90** and **128**, respectively. In addition to the central stepped portion **152**, which receives the compressors **24** and **26**, the shaft **18** also features stepped portions **170** and **172** for receiving the front and rear bearing assemblies **76** and **120**, respectively. The stepped arrangement of the shaft **18** facilitates assembly and disassembly in that the stepped portion **152** of the greatest diameter is centrally located on the shaft **18** and all the stepped portions are greater than the diameter of the threading at ends **30** and **132**.

It is to be noted that the bearing supports **40** and **114** are in a fixed position relative to the housing **12** and that the compressors **24** and **26** are positioned between the bearing supports **22** and **40** by spacers **84**, **124** and **154**, which spacers have predetermined lengths. Consequently, the placement of the compressors **24** and **26** relative to the longitudinal axis of the housing **12** is fixed by the aforementioned spacers and not by the axial location of the shaft **18** relative to the housing **12**. It is also to be noted that the stepped portions **152**, **170** and **172** of the shaft **18** are each provided with extra lengths such that the respective components (the bearing assemblies and compressors) can each be situated over a relatively wide range of situses in the respective stepped portions. Thusly, the shaft **18** need not be accurately positioned along the longitudinal axis of the housing **12** in order to achieve proper assembly of the supercharger **10**. For instance, if nuts **90** and **128** had been tightened differently than as they appear in FIG. **1**, the shaft **18** might have been displaced slightly in the axial direction from where it is shown in FIG. **1**. However, the relative positioning of the various components on the shaft **18**, i.e., the pulley wheel **28**, the compressors **24** and **26** and the bearing assemblies **76** and **120**, would have remained the same relative to themselves and the housing **12**. This feature eases the process of manufacture and accordingly, reduces the costs of same. It also reduces the amount of labor required for reassembly after repair.

In operation, the supercharger **10** is suitably connected at its outlet **16** to an intake of an automobile engine, with the drive belts **31** from the crankshaft of the engine being attached to the pulley wheel **28** of the supercharger **10**. Then, as the engine is operated, torque is transferred by the drive belts **31** to the pulley wheel **28** for driving the compressors **24** and **26**. Upon rotation, the axial compressor **24** draws air through the inlet **14**, imparts an initial amount of compression to the air and discharges it into the transition duct **96** with a swirl. By reason of its design, the axial compressor **24** is believed to move a greater volume of air in the region of its blade tips **140** than at its more radially inward locations. Accordingly, there is a greater of mass of air situated at the outer annular region behind the axial compressor **24** than at the inner annular region. As the discharge from the axial compressor **24** is caused to leave the axial compressor duct **34**, the highly convergent, transition duct **96** is believed to cause the outer annulus of air which is discharged from the axial compressor **24** to roll-in. This action is believed to have two favorable results. First, the roll-in action causes a flow regime to be established at the inlet **98** of the radial compressor **26** such that the need for a guide vane is wholly avoided. Secondly, and of equal importance, the rolling-in action, in conjunction with the large volume of space enclosed by the transition duct **96**, is believed to favorably effect the performance of the axial compressor **24** such that a higher pressure ratio is obtained therefrom.

Since the overall pressure ratio of the supercharger **10** is the product of the pressure ratios of the two compressors, it can be seen that the aforementioned increase in performance of the axial compressor **24** results in a corresponding improvement in overall performance of the supercharger **10**. It is also to be noted that the deletion of inlet guide vanes for the radial compressor **26** and of exit vanes for the axial compressor **24** greatly simplifies the design of the rear housing section **36** and therefore provides savings in costs of manufacture. It also avoids the pressure losses associated with such guide vanes, which are often quite significant.

Upon leaving the transition duct **96**, the pre-swirled flow of air enters the inlet **98** of the radial compressor **26** and then into the compressor **26** itself. In passing through the radial compressor **26**, the air is turned and whirled such that the airflow is centrifugally discharged with a substantial radial velocity component, whereupon the resultant flow is abruptly turned by the annular flow deflector **108** and caused to enter the exhaust cone **38**. As previously explained, the large volume of space enclosed by the exhaust cone **38** induces flow conditions behind the radial compressor **26** such that elevated back pressures are avoided which might otherwise impair the performance of the radial compressor **26**. Pulses in the output of the radial compressor **26** are also moderated. The air is then delivered in a compressed state to the exhaust port **16** of the exhaust cone **38**. The supercharged air then flows down the appropriate intake system of the engine until it reaches the cylinder or cylinders of the engine.

With respect to the application of the supercharger **10** to air-tank powered engines, such as disclosed in U.S. Pat. No. 4,292,804, the supercharger **10** functions in the same manner as described above, but is connected to the engine differently. In the air tank powered engine, at least one of the exhaust manifolds of the engines delivers partially expanded air to a line connected to the inlet **14** of the supercharger **10**. Referring to FIG. **8**, in most of such applications, this line will be of a smaller diameter than the housing **12** at the inlet **14** of the supercharger **10**, such that an adaptor **173** is needed. The adaptor **173** comprises an annular plate **174** having a threaded aperture **176** sized to receive a mating, threaded end **178** of the aforementioned line **180**. The plate **174** is secured the flange **36** of the front housing section **32** by a plurality of bolts. Because the air coming from the line **180** is usually less than the full capacity of the supercharger, additional air is introduced through the lateral opening **44** along the side of the front housing section **32**. In this application, the opening **44** thusly serves as an air intake port as well as a means for accommodating the drive belts **31** and must therefore be sized upon the additional criteria that it not be so large as to upset the flow of the incoming air in the line **180**. Upon the passage of the air through the supercharger, the air is directed through the exhaust port **16** and into a suitable line connected thereto, which line may lead directly to the engine or to the storage tanks of the engine. If directed to the tanks, this recompressed air is utilized to supplement the required recharging of the storage tanks.

It is to be appreciated that savings in the cost of manufacturing the supercharger **10** are achieved by reason that the housing **12**, the bearing supports **40** and **114**, the axial compressor **24** and the radial compressor **26** are all constructed from cast aluminum parts and require only a minimum amount of machining. Moreover, the roller bearing assemblies **76** and **120** are commercially available components, and the supercharger **10** is easily assembled. These aspects further reduce the cost of manufacture and render the disclosed supercharger inexpensive to maintain and overhaul. More importantly, the supercharger **10**, despite its simple design, provides supercharging at relatively low operating speeds. With its lower operating speeds, the service life of the supercharger **10** is extended and the risk of it suffering mechanical failure is reduced. The need for special bearing designs and lubrication is also avoided. Accordingly, the supercharger **10** is highly suitable for mass production and for use in automobiles, trucks, helicoptors or the like.

It is to be understood that the present invention may be embodied in other specific forms without departing from the spirit or essential characteristics of the present invention. The preferred embodiments are therefore to be considered illustrative and not restrictive. The scope of the invention is set forth in the appended claims rather than by the foregoing descriptions and all changes or variations which fall within the meaning and range of the claims are therefore intended to be embraced therein.

What is claimed is:

1. A supercharger comprising:

a housing having a longitudinal axis and being constructed from sections, which sections are connected in end-to-end relationship with flanged joints, said sections including a front housing section defining an axially directed inlet, an axial compressor duct section for housing an axial compressor, a rear housing section downstream of said axial compressor duct section, and a hollow, highly convergent, frustoconical, exhaust cone section downstream of said rear housing section, a downstream portion of said rear housing section defining a cylindrical, axially directed inlet for a radial compressor, a casing for a radial compressor, and a substantially radially directed outlet for a radial compressor, said rear housing section further defining a highly convergent, hollow, frustoconical transition duct between a downstream end of said axial compressor duct section and said radial compressor inlet, said rear housing section having interior surfaces for defining a flow deflector, which flow deflector receives the output of said outlet for a radial compressor, which flow deflector provides a smooth surface transition from said rear housing section into said exhaust cone section, said exhaust cone section defining at a downstream end a coaxial outlet, said housing further including at least two bearing supports affixable within said housing according to predetermined locations along said longitudinal axis of the housing, said bearing mounts rotatably supporting a compressor shaft, which compressor shaft is positioned along said longitudinal axis;

sealed bearing assemblies removably implaced in said bearing supports, said sealed bearing assemblies receiving said compressor shaft;

a shrouded axial compressor located within said axial compressor duct and secured to said shaft to be rotatable therewith, said shrouded axial compressor drawing a flow from said inlet of the front housing section and imparting an initial compression to said flow;

a radial compressor located within said casing and secured to said shaft to be rotatable therewith, said casing being in a substantially sealing relationship with said radial compressor, said radial compressor including a hub, a first set of blades extending radially from said hub and having leading edges at an intake region of said radial compressor and a second set of blades extending radially from said hub and having leading edges downstream of said intake region, said radial compressor further compressing said flow;

at least one pulley wheel secured to said compressor shaft and adapted to receive drive belts;

spacers fitted upon said compressor shaft for axially positioning said axial compressor and said radial compressor relative to each other and relative to said bearing supports; and

means for securing said compressor shaft against axial displacement;

wherein said transition duct favorably directs the output of the axial compressor into said inlet of the radial compressor and said exhaust cone section encloses sufficient volume to moderate the output of said radial compressor.

2. The supercharger as claimed in claim **1** wherein said transition duct section and said exhaust cone section each have a half-apex angle of approximately 35°.

3. A supercharger comprising:

a housing having a longitudinal axis and being constructed from sections, which sections are connected in end-to-end relationship, said sections including a front housing section defining an axially directed inlet, an axial compressor duct section for housing an axial compressor, a rear housing section downstream of said axial compressor duct section and an exhaust section having a hollow, highly convergent, frustoconical, exhaust cone portion downstream of said rear housing section, a downstream portion of said rear housing section defining a cylindrical, axially directed inlet for a radial compressor, a casing for a radial compressor, and a substantially radially directed outlet for a radial compressor, said rear housing section further defining a highly convergent, hollow, frustoconical transition duct between a downstream end of said axial compressor duct section and said radial compressor inlet, said rear housing section having interior surfaces for defining a flow deflector, which flow deflector receives the output of said outlet for a radial compressor, which flow deflector provides a smooth surface transition from said rear housing section into said exhaust section, said exhaust cone portion defining at a downstream end a coaxial outlet, said housing further including at least two bearing supports affixable within said housing according to predetermined locations along said longitudinal axis of the housing, said bearing mounts rotatably supporting a compressor shaft, which compressor shaft is positioned along said longitudinal axis;

bearing assemblies removably implaced in said bearing supports, said bearing assemblies receiving said compressor shaft;

an axial compressor located within said axial compressor duct and secured to said shaft to be rotatable therewith, said axial compressor drawing a flow from said inlet of the front housing section and imparting an initial compression to said flow;

a radial compressor located within said casing and secured to said shaft to be rotatable therewith, said casing being in a substantially sealing relationship with said radial compressor, said radial compressor including a hub, a first set of blades extending radially from said hub and having leading edges at an intake region of said radial compressor and a second set of blades extending radially from said hub and having leading edges downstream of said intake region, said radial compressor further compressing said flow;

at least one pulley wheel secured to said compressor shaft and adapted to receive drive belts;

spacers fitted upon said compressor shaft for axially positioning said axial compressor and said radial compressor relative to each other; and

means for securing said compressor shaft against axial displacement;

wherein said transition duct favorably directs the output of the axial compressor into said inlet of the radial compressor and said exhaust cone section encloses sufficient volume to moderate the output of said radial compressor.

4. The supercharger as claimed in claim **3**, wherein said exhaust cone portion has a half-apex angle of approximately 35°.

5. The supercharger as claimed in claim **3**, wherein said transition duct has a half-apex angle of 35°.

6. The supercharger as claimed in claim **3**, wherein said transition duct is substantially coaxial with said inlet of the radial compressor.

7. The supercharger as claimed in claim **3**, wherein said axial compressor comprises a hub, a plurality of blades extending radially from said hub and a shroud connected to tips of said blades.

* * * * *

United States Patent [19]
Rogers, Sr.

[11] **4,292,804**
[45] **Oct. 6, 1981**

[54] METHOD AND APPARATUS FOR OPERATING AN ENGINE ON COMPRESSED GAS

[76] Inventor: **Leroy K. Rogers, Sr.,** #5 Capistrano Ct., Ft. Myers, Fla. 33908

[21] Appl. No.: **158,303**
[22] Filed: **Jun. 10, 1980**
[51] **Int. Cl.**^{3} **F15B 11/06**
[52] **U.S. Cl.** **60/407**; 91/187; 91/275
[58] **Field of Search** 60/407, 412; 91/187, 91/275, 364

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,881,399	5/1975	Sagi et al.	91/187 X
3,885,387	5/1975	Simington	60/407 X
4,018,050	4/1977	Murphy	60/412 X

Primary Examiner—Allen M. Ostrager
Attorney, Agent, or Firm—Burns, Doane, Swecker & Mathis

[57] ABSTRACT

The present invention relates to a method and apparatus for operating an engine having a cylinder and a piston reciprocable therein on compressed gas. The apparatus comprises a source of compressed gas connected to a distributor which distributes the compressed gas to the cylinder. A valve is provided to selectively admit compressed gas to the cylinder when the piston is in an approximately top dead center position. In one embodiment of the present invention the timing of the opening of the valve is advanced such that the compressed gas is admitted to the cylinder progressively further before the top dead center position of the piston as the speed of the engine increases. In a further embodiment of the present invention a valve actuator is provided which increases the length of time over which the valve remains open to admit compressed gas to the cylinder as the speed of the engine increases. A still further embodiment of the present invention relates to an apparatus for adapting a conventional internal combustion engine for operation on compressed gas.

22 Claims, 8 Drawing Figures

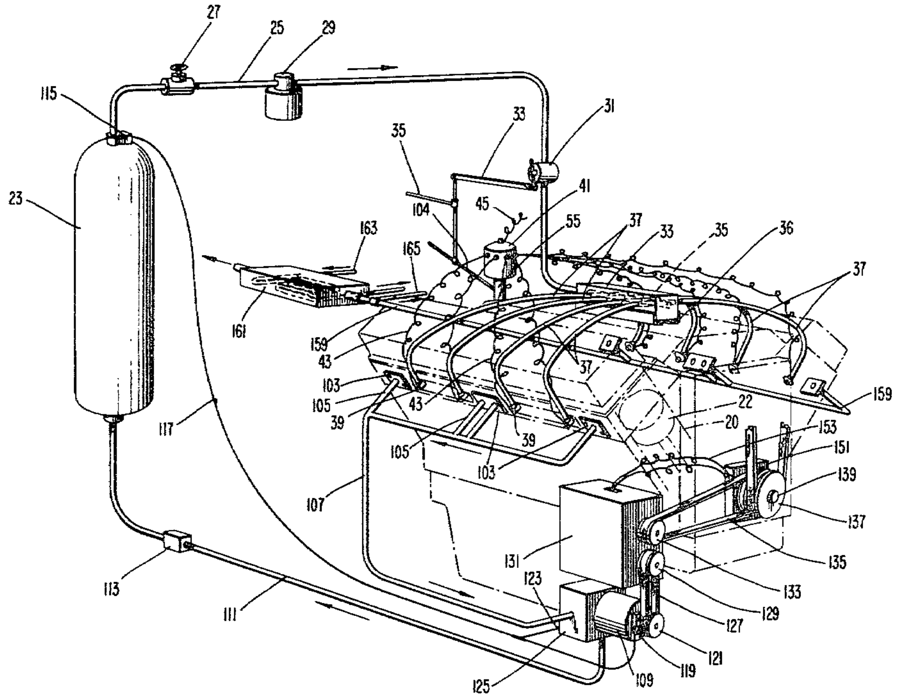

FIG. 1

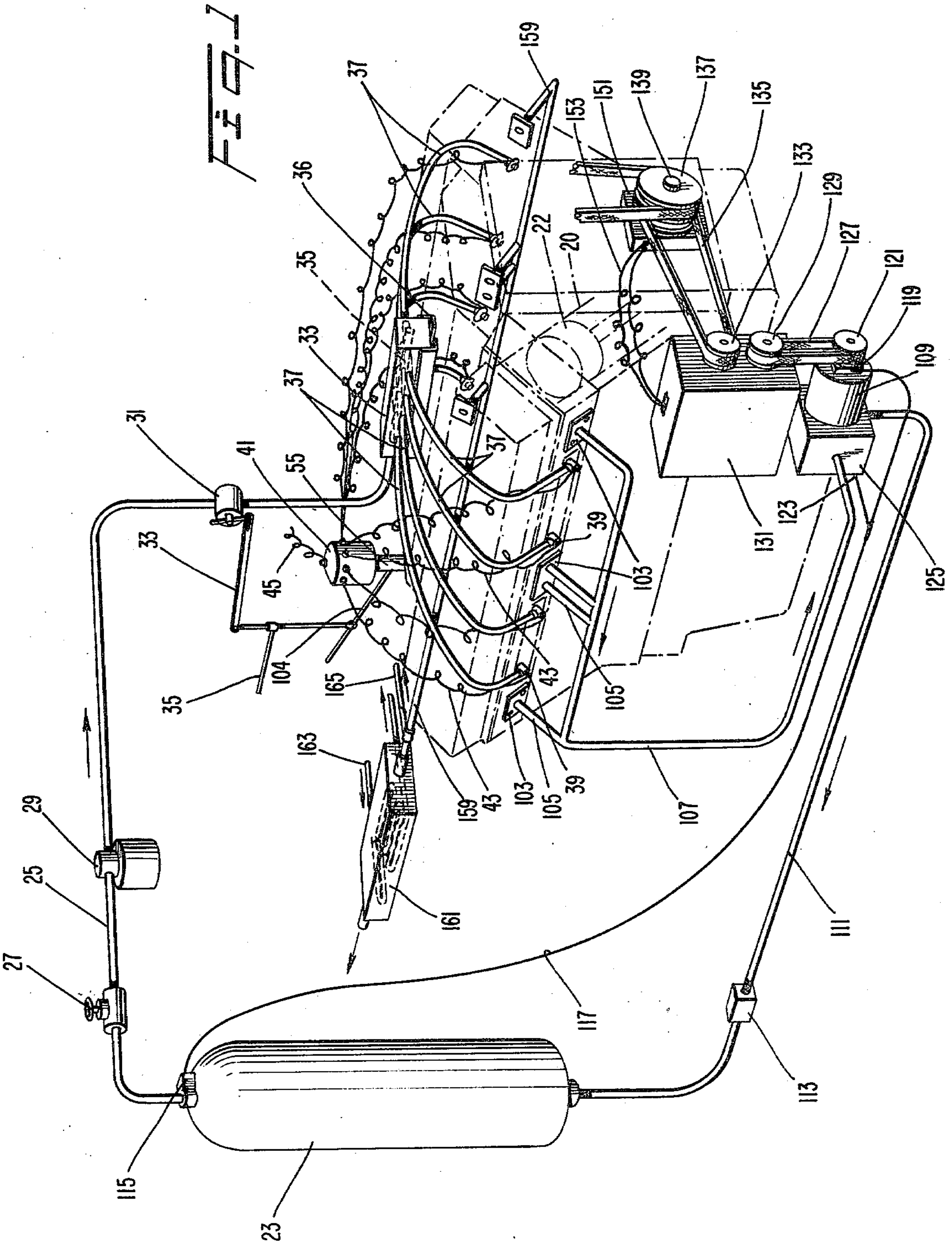

FIG. 6

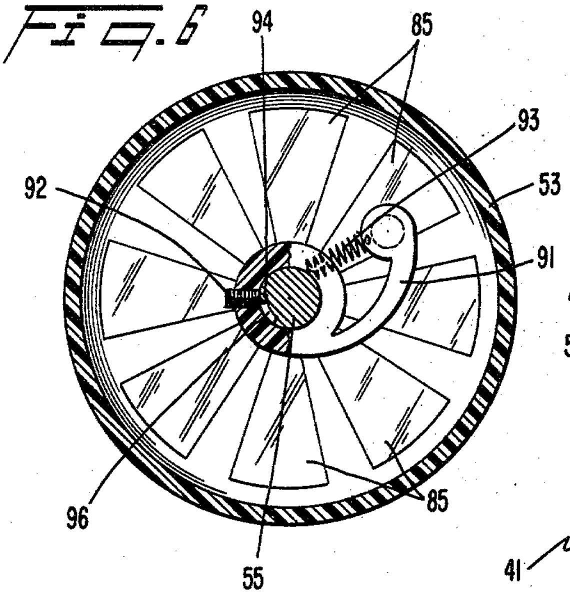
94
85
93
92
53
91
96
85
55

FIG. 7

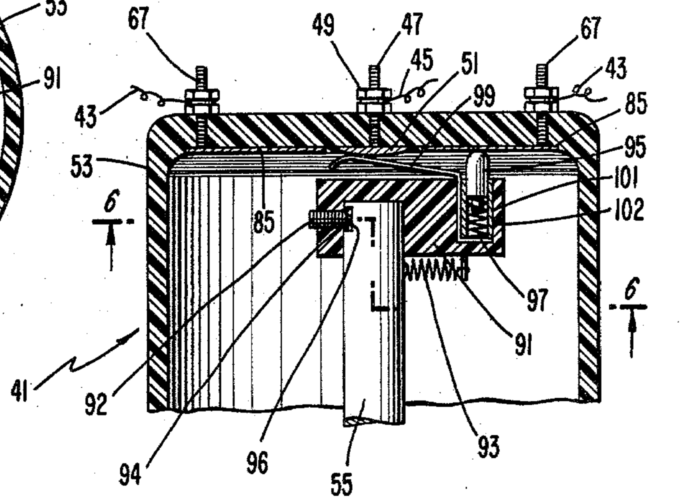
67
49
47
45
51
67
43
99
43
85
43
95
53
101
6
85
102
6
97
91
41
92
93
94
96
55

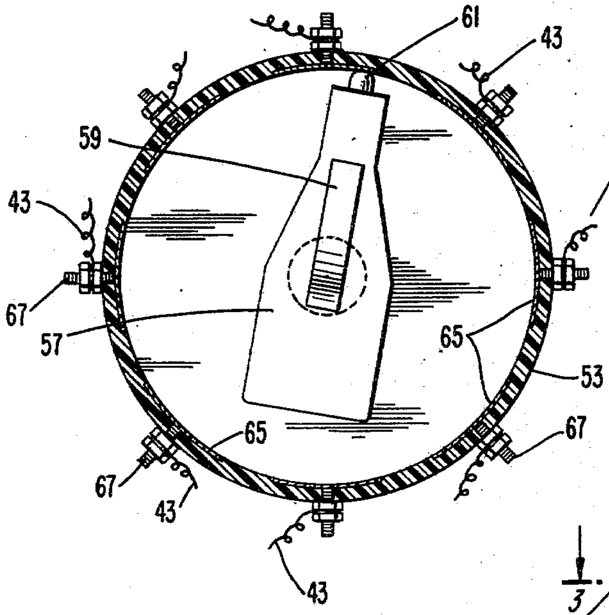
61
43
59
43
67
57
65
53
67
65
67
43
43

FIG. 3

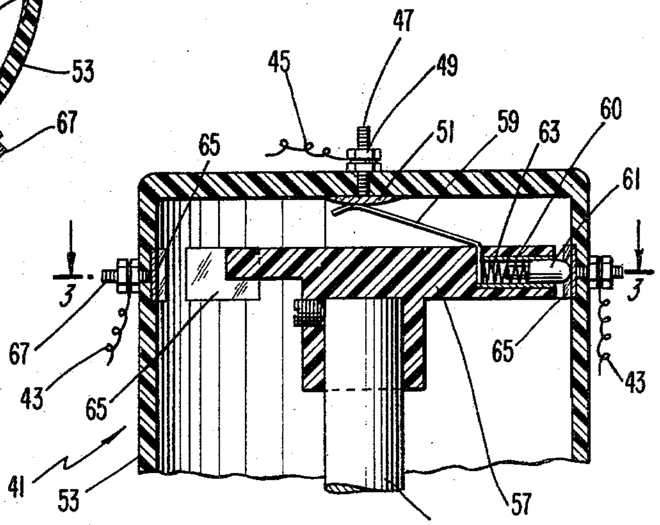
47
45
49
51
59
63
60
65
61
3
3
67
65
43
43
65
43
65
41
53
57
55

FIG. 2

Fig. 5

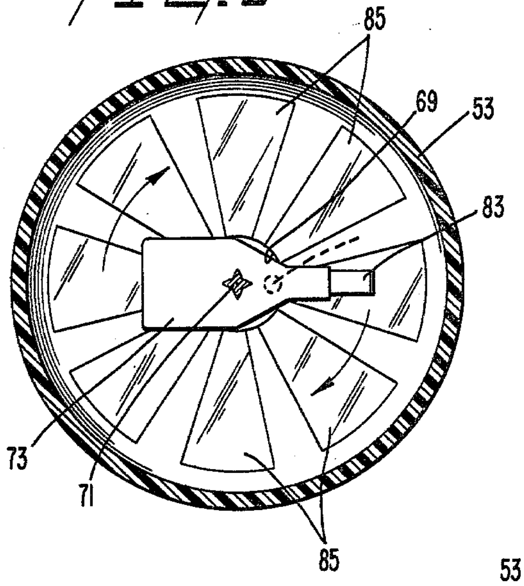
85
69
53
83
73
71
85

Fig. 4

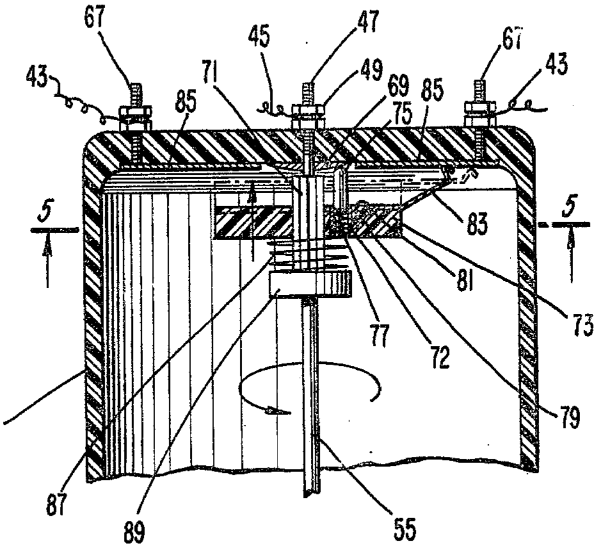
67
45
47
67
43
71
49
69
85
43
85
75
5
83
5
81
73
77
72
53
79
87
89
55

Fig. 6

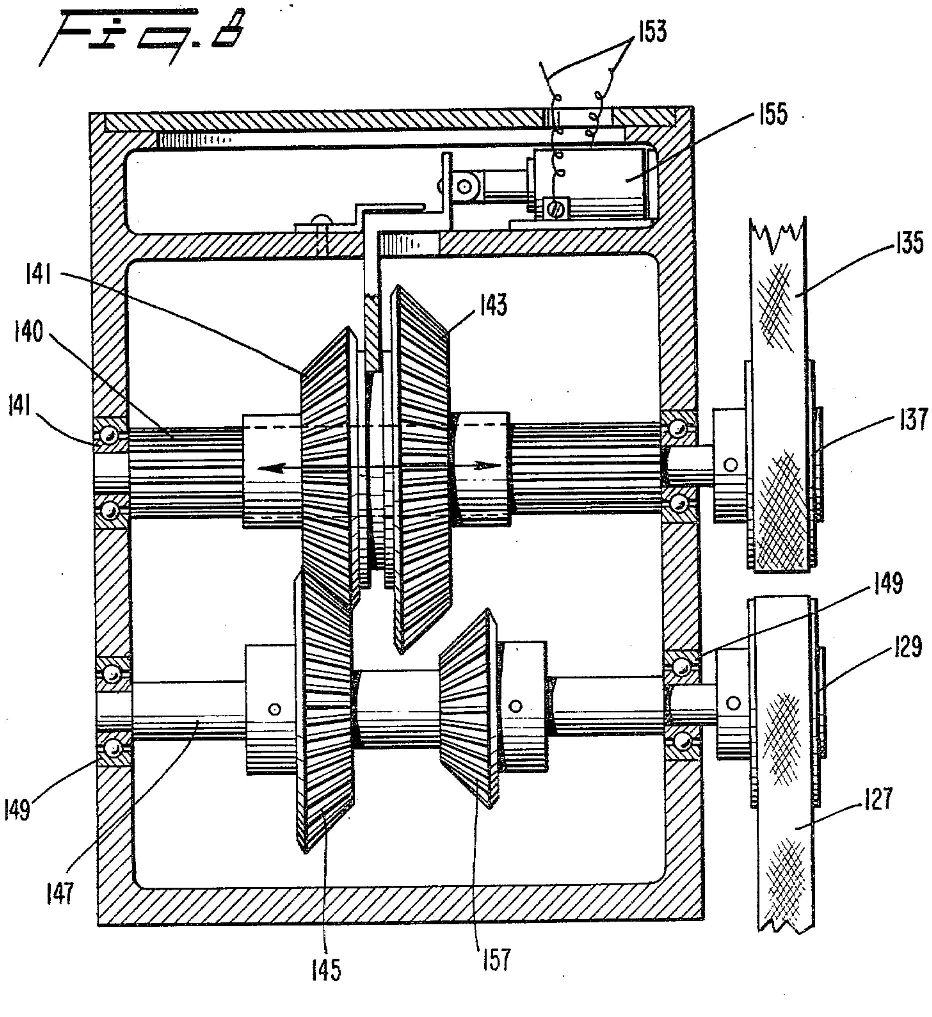
153
155
135
141
143
140
141
137
149
129
127
149
147
145
157

METHOD AND APPARATUS FOR OPERATING AN ENGINE ON COMPRESSED GAS

BACKGROUND AND SUMMARY OF THE PRESENT INVENTION

The present invention relates to a method and apparatus for operating an engine using a compressed gas as the motive fluid. More particularly, the present invention relates to a apparatus for adapting a pre-existing internal combustion engine for operation on a compressed gas.

Air pollution is one of the most serious problems facing the world today. One of the major contributors to air pollution is ordinary internal combustion engine which are used in most motor vehicles today. Various devices, including many items mandated by legislation, have been proposed in an attempt to limit the pollutants which an internal combustion engine exhausts to the air. However, most of these devices have met with limited success and are often both prohibitively expensive and complex. A clean alternative to the internal combustion engine is needed to power vehicles and other machinery.

A compressed gas, preferably air, would provide an ideal motive fluid for a engine since it would eliminate the usual pollutants exhausted from an internal combustion engine. An apparatus for converting an internal combustion engine for operation on compressed air is disclosed in U.S. Pat. No. 3,885,387 issued May 27, 1975 to Simington. The Simington patent discloses an apparatus including a source of compressed air and a rotating valve actuator which opens and closes a plurality of mechanical poppet valves. The valves deliver compressed air in timed sequence to the cylinders of an engine through adapters located in the spark plug holes. However, the output speed of an engine of this type is limited by the speed of the mechanical valves and the fact that the length of time over which each of the valves remains open cannot be varied as the speed of the engine increases.

Another apparatus for converting an internal combustion engine for operation on steam or compressed air is disclosed in U.S. Pat. No. 4,102,130 issued July 25, 1978 to Stricklin. The Stricklin patent discloses a device which changes the valve timing of a conventional four stroke engine such that the intake and exhaust valves open once for every revolution of the engine instead of once every other revolution of the engine. A reversing valve is provided which delivers live steam or compressed air to the intake valves and is subsequently reversed to allow the exhaust valves to deliver the expanded steam or air to the atmosphere. A reversing valve of this type however does not provide a reliable apparatus for varying the amount of motive fluid injected into the cylinders when it is desired to increase the speed of the engine. Further, a device of the type disclosed in the Stricklin patent requires the use of multiple reversing valves if the cylinders in a multi-cylinder engine were to be fired sequentially.

Therefore, it is an object of the present invention to provide a reliable method and apparatus for operating an engine or converting an engine for operation with a compressed gas.

A further object of the present invention is to provide a method and apparatus which is effective to deliver a constantly increasing amount of compressed gas to an engine as the speed of the engine increases.

A still further object of the present invention is to provide a method and apparatus which will operate an engine using compressed gas at a speed sufficient to drive a conventional automobile at highway speeds.

It is still a further object of the present invention to provide a method and apparatus which is readily adaptable to a standard internal combustion engine to convert the internal combustion engine for operation with a compressed gas.

Another object of the invention is to provide a method and apparatus which utilizes cool expanded gas, exhausted from a compressed gas engine, to operate an air conditioning unit and/or an oil cooler.

These and other objects are realized by a method and apparatus according to the present invention for operating an engine having at least one cylinder and a reciprицating piston therein using compressed gas as a motive fluid. The apparatus includes a source of compressed gas and a distributor connected with the source of the compressed gas for distributing the compressed gas to the at least one cylinder. A valve is provided for admitting the compressed gas to the cylinder when the piston is in approximately a top dead center position within the cylinder. An exhaust is provided for exhausting the expanded gas from the cylinder as the piston returns to approximately the top dead center position.

In a preferred embodiment of the present invention a device is provided for varying the duration of each engine cycle over which the valve remains open to admit compressed gas to the cylinder dependent upon the speed of the engine. In a further preferred embodiment of the present invention, an apparatus for advancing the timing of the opening of the valve is arranged to admit the compressed gas to the cylinder progressively further before the top dead center position of the piston as the speed of the engine increases.

Further features of the present invention include a valve for controlling the amount of compressed gas admitted to the distributor. Also, a portion of the gas which has been expanded in the cylinder and exhausted through the exhaust valve is delivered to a compressor to be recompressed and returned to the source of compressed gas. A gear train is selectively engagable to drive the compressor at different operating speed depending upon the pressure maintained at the source of compressed air and/or the speed of the engine. Still further, a second portion of the exhaust gas is used to cool a lubricating fluid for the engine or to operate an air conditioning unit.

In a preferred embodiment of the present invention, the valve for admitting compressed gas to the cylinder is electrically actuated. The device for varying the duration of each engine cycle over which the intake valve remains open as the speed of the engine increase comprises a rotating element whose effective length increases as the speed of the engine increases such that a first contact on the rotating element is electrically connected to a second contact for a longer period of each engine cycle. The second contact actuates the valve whereby the valve remains in an open position for a longer period of each engine cycle as the speed of the engine increases.

Still further features of the present invention include an adaptor plate for supporting the distributor above an intake manifold of a conventional internal combustion engine after a carburetor has been removed to allow air

to enter the cylinders of the engine through the intake manifold and conventional intake valves. Another adaptor plate is arranged over an exhaust passageway of the internal combustion engine to reduce the cross-sectional area of the exhaust passageway.

BRIEF DESCRIPTION OF THE DRAWINGS

Preferred embodiments of a method and apparatus for operating an engine according to the present invention will be described with reference to the accompanying drawings wherein like members bear like reference numerals and wherein:

FIG. 1 is a schematic representation of an apparatus according to the present invention arranged on an engine;

FIG. 2 is a side view of one embodiment of a valve actuator according to the present invention;

FIG. 3 is a cross-sectional view taken along the line 3—3 in FIG. 2;

FIG. 4 is a cross-sectional view of a second embodiment of a valve actuator according to the present invention;

FIG. 5 is a view taken along the line 5—5 in FIG. 4;

FIG. 6 is a cross-sectional view of a third embodiment of a valve actuator according to the present invention;

FIG. 7 is a view taken along the line 7—7 in FIG. 6;

FIG. 8 is a cross-sectional view of a gearing unit to drive a compressor according to the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

With reference to FIG. 1, an engine block **21** (shown in phantom) having two banks of cylinders with each bank including cylinders **20** having pistons **22** reciprocable therein (only one of which is shown in phantom) in a conventional manner. While the illustrated engine is a V-8 engine, it will be apparent that the present invention is applicable to an engine having any number of pistons and cylinders with the V-8 engine being utilized for illustration purposes only. A compressed gas tank **23** is provided to store a compressed gas at high pressure. It may also be desirable to include a small electric or gas compressor to provide compressed gas to supplement the compressed gas held in the tank **23**. In a preferred embodiment, the compressed gas is air which can be obtained from any suitable source.

A line **25** transports the gas withdrawn from the tank **23** when a conventional shut off valve **27** is open. In addition, a solenoid valve **29** preferably operated by a suitable key operated switch (not shown) for the engine is also arranged in the line **25**. In normal operation, the valve **27** is maintained open at all times with the solenoid valve **29** operating as a selective shut off valve to start and stop the engine **21** of the present invention.

A suitable regulating valve **31** is arranged downstream from the solenoid valve **29** and is connected by a linkage **33** to a throttle linkage **35** which is operator actuated by any suitable apparatus such as a foot pedal (not shown). The line **25** enters an end of a distributor **33** and is connected to an end of a pipe **35** which is closed at the other end. A plurality of holes, which are equal to the number of cylinders in the engine **21**, are provided on either side of the pipe **35** along the length of the pipe **35**.

When the present invention is used to adapt a conventional internal combustion engine for operation on compressed gas, an adaptor plate **36** is provided to support the distributor **33** in spaced relation from the usual intake opening in the intake manifold of the engine after a conventional carburetor has been removed. In this way, air is permitted to enter the internal combustion engine through the usual passageways and to be admitted to the cylinders through suitable intake valves (not shown). The adaptor plate **36** is secured to the engine block **21** and the distributor **33** by any suitable apparatus, e.g., bolts.

Each of the holes in the pipe **35** is connected in fluid-tight manner to a single line **37**. Each line **37** carries the compressed gas to a single cylinder **20**. In a preferred embodiment, each of the lines **37** is ½ inch high pressure plastic tubing attached through suitable connectors to the distributor **33** and the pipe **35**. Each of the lines **37** is connected to a valve **39** which is secured in an opening provided near the top of each of the cylinders **20**. In the case of a conversion of a standard internal combustion engine, the valves **39** can be conveniently screwed into a tapped hole in the cylinder **20** typically provided for a spark plug of the internal combustion engine. In a preferred embodiment, the valves **39** are solenoid actuated valves in order to provide a fast and reliable opening and closing of the valves **39**.

Each of the valves **39** is energized by a valve actuator **41** through one of a plurality of wires **43**. The valve actuator **41** is driven by a shaft of the engine similar to the drive for a conventional distributor of an internal combustion engine. That is, a shaft **55** of the valve actuator **41** is driven in synchronism with the engine **21** at one half the speed of the engine **21**.

A first embodiment of the valve actuator **41** (FIGS. 2 and 3) receives electrical power through a wire **45** which is energized in a suitable manner by a battery, and a coil if necessary (not shown) as is conventional in an internal combustion engine. The wire **45** is attached to a central post **47** by a nut **49**. The post **47** is connected to a conducting plate **51** arranged within a housing **53** for the valve actuator **41**. Within the housing **53**, the shaft **55** has an insulating element **57** secured to an end of the shaft **55** for co-rotation therewith when the shaft **55** is driven by the engine **21**. A first end of a flexible contact **59** is continuously biased against the conducting plate **51** to receive electricity from the battery or another suitable source. A second end of the contact **59** is connected to a conducting sleeve **60** which is in constant contact with a spring biased contact **61** which is arranged within the sleeve **60**. The contact **61** is biased by a spring **63** which urges the contact **61** towards a side wall of the housing **53**.

With reference to FIG. 3, a plurality of contacts **65** are spaced from one another and are arranged around the periphery of the housing **53** at the same level as the spring biased contact **61**. Each contact **65** is electrically connected to a post **67** which extends outside of the housing **53**. The number of contacts **65** is equal to the number of cylinders in the engine **21**. One of the wires **43**, which actuate the valves **39**, is secured to each of the posts **67**.

In operation, as the shaft **55** rotates in synchronism with the engine **21**, the insulating element **57** rotates and electricity is ultimately delivered to successive ones of the contacts **65** and wires **43** through the spring biased contact **61** and the flexible contact **59**. In this way, each of the electrical valves **39** is actuated and opened in the proper timed sequence to admit compressed gas to each of the cylinders **20** to drive the pistons **22** therein on a downward stroke.

The embodiment illustrated in FIGS. 2 and 3 is effective to actuate each of the valves 39 to remain open for a long enough period of time to admit sufficient compressed gas to each of the cylinders 20 of the engine 21 to drive the engine 21. The length of each of the contacts 65 around the periphery of the housing 53 is sufficient to permit the speed of the engine to be increased when desired by the operator by moving the throttle linkage 35 which actuates the linkage 33 to further open the regulating valve 31 to admit more compressed gas from the tank 23 to the distributor 33. However, it has been found that the amount of air admitted by the valves 39 when using the first embodiment of the valve actuator 41 (FIGS. 2 and 3) is substantially more than required to operate the engine 21 at an idling speed. Therefore, it may be desirable to provide a valve actuator 41 which is capable of varying the duration of each engine cycle over which the solenoid valves 39 are actuated, i.e., remain open to admit compressed gas, as the speed of the engine 21 is varied.

A second embodiment of a valve actuator 41 which is capable of varying the duration of each engine cycle over which each of the valves 39 remains open to admit compressed gas to the cylinders 20 dependent upon the speed of the engine 21 will be described with reference to FIGS. 4 and 5 wherein members corresponding to those of FIGS. 2 and 3 bear like reference numerals. The wire 45 from the electrical source is secured to the post 47 by the nut 49. The post 47 has a annular contact ring 69 electrically connected to an end of the post 47 and arranged within the housing 53. The shaft 55 rotates at one half the speed of the engine as in the embodiment of FIGS. 2 and 3.

At an upper end of the shaft 55, a splined section 71 slidably receives an insulating member 73. The splined section 71 of the shaft 55 positively holds the insulating member 73 for co-rotation therewith but permits the insulating member 73 to slide axially along the length of the splined section 71. Near the shaft 55, a conductive sleeve 72 is arranged in a bore 81 in an upper surface of the insulating element 73 generally parallel to the splined section 71. A contact 75, biased towards the annular contact ring 69 by a spring 77, is arranged within the conductive sleeve 72 in contact therewith. The conductive sleeve 72 also contacts a conductor 79 at a base of the bore 81.

The conductor 79 extends to the upper surface of the insulating element 73 near an outer periphery of the insulating element 73 where the conductor 79 is electrically connected to a flexible contact 83. The flexible contact 83 selectively engages a plurality of radial contacts 85 arranged on an upper inside surface of the housing 53. A weak spring 87 arranged around the splined section 71 engages a stop member 89 secured on the shaft 55 and the insulating element 73 to slightly bias the insulating element 73 towards the upper inside surface of the housing 53 to ensure contact between the flexible contact 83 and the upper inside surface of the housing 53. As best seen in FIG. 5, the radial contacts 85 on the upper inside surface of the housing 53 are arranged generally in the form of radial spokes extending from the center of the housing 53 with the number of contacts being equal to the number of cylinders 20 in the engine 21. The number of degrees covered by each of the radial contacts 85 gradually increases as the distance from the center of the upper inside surface of the housing 53 increases.

In operation of the device of FIGS. 4 and 5, as the shaft 55 rotates, electricity flows along a path through the wire 45 down through post 47 to the annular contact member 69 which is in constant contact with the spring biased contact 75. The electrical current passes through the conductive sleeve 72 to the conductor 79 and then to the flexible contact 83. As the flexible contact 83 rotates along with the insulating member 73 and the shaft 55, the tip of the flexible contact 83 successively engages each of the radial contacts 85 on the upper inside of the housing 53. As the speed of the shaft 55 increases, the insulating member 73 and the flexible contact 83 attached thereto move upwardly along the splined section 71 of the shaft 55 due to the radial component of the splines in the direction of rotation under the influence of centrifugal force. As the insulating member 73 moves upwardly, the flexible contact 83 is bent such that the tip of the contact 83 extends further radially outwardly from the center of the housing 53 (as seen in phantom lines in FIG. 4). In other words, the effective length of the flexible contact 83 increases as the speed of the engine 21 increases.

As the flexible contact 83 is bent and the tip of the contact 83 moves outwardly, the tip remains in contact with each of the radial contacts 85 for a longer period of each engine cycle due to the increased angular width of the radial contacts with increasing distance from the center of the housing 53. In this way, the length of time over which each of the valves 39 remains open is increased as the speed of the engine is increased. Thus, a larger quantity of compressed gas or air is injected into the cylinders as the speed increases. Conversely, as the speed decreases and the insulating member 73 moves downwardly along the splined section 71, a minimum quantity of air is injected into the cylinder due to the shorter length of the individual radial contact 85 which is in contact with the flexible contact 83. In this way, the amount of compressed gas that is used during idling of the engine 21 is at a minimum whereas the amount of compressed gas which is required to increase the speed of the engine 21 to a level suitable to drive a vehicle on a highway is readily available.

With reference to FIGS. 6 and 7, a third embodiment of a valve actuator 41 according to the present invention includes an arcuate insulating element 91 having a first end pivotally secured by any suitable device such as screw 92 to the shaft 55 for co-rotation with the shaft 55. The screw 92 is screwed into a tapped hole in the insulating element 91 such that a tab 94 at an end of the screw 92 engages a groove 96 provided in the shaft 55. In this way, the insulating element 91 positively rotates with the shaft 55. However, as the shaft 55 rotates faster, a second end 98 of the insulating element 91 is permitted to pivot outwardly under the influence of centrifugal force because of the groove 96 provided in the shaft 55. A spring 93 connected between the second end 98 of the element 91 and the shaft 55 urges the second end of the element 91 towards the center of the housing 53.

A contact 99 similar to the contact 59 (FIG. 2) is arranged such that one end of the contact 99 is in constant contact with the conducting plate 51 located centrally within the housing 53. The other end of the contact 99 engages a conductive sleeve 101 arranged in bore 102. A contact element 95 is arranged in the conductive sleeve 101 in constant contact with the sleeve 101. The bore 102 is arranged generally parallel to the shaft 55 near the second end of the arcuate insulating

element **91**. The contact **95** is biased by a spring **97** towards the upper inside surface of the housing **53** for selective contact with each of the plurality of radial contacts **85** which increase in arc length towards the outer peripheral surface of the housing **53** (FIG. **6**).

In operation of the device of FIGS. **6** and **7**, as the shaft **55** rotates the arcuate insulating element **91** rotates with the shaft **55** and the second end **98** of the insulating element **91** tends to pivot about the shaft **55** due to centrifugal force. Thus, as the effective length of the contact **95** increases, i.e., as the arcuate insulating element **91** pivots further outwardly, the number of degrees of rotation over which the contact **95** is in contact with each of the radial contacts **85** on the upper inside surface of the housing **53** increases thereby permitting each of the valves **39** to remain open for a longer period of each engine cycle to admit more compressed gas to the respective cylinder **20** to further increase the speed of the engine **21**.

With reference to FIG. **1**, a mechanical advance linkage **104** which is connected to the throttle linkage **35**, advances the initiation of the opening of each valve **39** such that compressed gas is injected into the respective cylinder further before the piston **22** in the respective cylinder **20** reaches a top dead center position as the speed of the engine is increased by moving the throttle linkage **35**. The advance linkage **104** is similar to a conventional standard mechanical advance employed on an internal combustion engine. In other words, the linkage **104** varies the relationship between the angular positions of a point on the shaft **55** and a point on the housing **53** containing the contacts. Alternatively, a conventional vacuum advance could also be employed. By advancing the timing of the opening of the valves **39**, the speed of the engine can more easily be increased.

The operation of the engine cycle according to the present invention will now be described. The compressed gas injected into each cylinder of the engine **21** drives the respective piston **22** downward to drive a conventional crankshaft (not shown). The movement of the piston downwardly causes the compressed gas to expand rapidly and cool. As the piston **22** begins to move upwardly in the cylinder **20** a suitable exhaust valve (not shown) arranged to close an exhaust passageway is opened by any suitable apparatus. The expanded gas is then expelled through the exhaust passageway. As the piston **22** again begins to move downwardly a suitable intake valve opens to admit ambient air to the cylinder. The intake valve closes and the ambient air is compressed on the subsequent upward movement of the piston until the piston reaches approximately the top dead center position at which time the compressed gas is again injected into the cylinder **20** to drive the piston **22** downward and the cycle begins anew.

In the case of adapting a conventional internal combustion engine for operation on compressed gas, a plurality of plates **103** are preferably arranged over an end of the exhaust passageways in order to reduce the outlet size of the exhaust passageways of the conventional internal combustion engine. In the illustrated embodiment, a single plate having an opening in the center is bolted to the outside exhaust passageway on each bank of the V-8 engine while another single plate having two openings therein is arranged with one opening over each of the interior exhaust passageways on each bank of the V-8 engine. A line **105** is suitably attached to each of the adaptor places to carry the exhaust to an appropriate location. In a preferred embodiment, the exhaust lines **105** are 1½" plastic tubing.

In a preferred embodiment, the exhaust lines **105** of one bank of the V-8 engine are collected in a line **107** and fed to an inlet of a compressor **109**. The pressure of the exhaust gas emmanating from the engine **21** according to the present invention is approximately 25 p.s.i. In this way, the compressor **109** does not have to pull the exhaust into the compressor since the gas exhausted from the engine **21** is at a positive pressure. The positive pressure of the incoming fluid increases the efficiency and reduces wear on the compressor **109**. The exhaust gas is compressed in the compressor **109** and returned through a line **111** and a check valve **113** to the compressed gas storage tank **23**. The check valve **113** prevents the flow of compressed gas stored in the tank **23** back towards the compressor **109**.

A suitable pressure sensor **115** is arranged at an upper end of the tank **23** and sends a signal along a line **117** when the pressure exceeds a predetermined level and when the pressure drops below a predetermined level. The line **117** controls an electrically actuated clutch **119** disposed at a front end of the compressor **109**. The clutch **119** is operative to engage and disengage the compressor **109** from a drive pulley **121**. Also, the signal carried by the line **117** actuates a suitable valve **123** arranged on a compressor housing **125** to exhaust the air entering the compressor housing **125** from the line **107** when the clutch **119** has disengaged the compressor **109** from the drive pully **121**.

In a preferred embodiment, when the pressure is the tank **23** reaches approximately 600 p.s.i., the clutch **119** is disengaged and the compressor **109** is deactivated and the valve **123** is opened to exhaust the expanded gas delivered to the compressor **109** from the line **107** to the atmosphere. When the pressure within the tank **23** drops below approximately 500 p.s.i., the sensor **115** sends a signal to engage the clutch **119** and close the valve **123**, thereby operating the compressor **109** for supplying the tank **23** with compressed gas.

The pulley **121** which drives the compressor **109** through the clutch **119** is driven by a belt **127** which is driven by a pulley **129** which operates through a gear box **131**. With reference to FIGS. **1** and **8**, a second pulley **133** on the gear box is driven by a belt **135** from a pulley **137** arranged on a drive shaft **139** of the engine **21**. The pulley **137** drives a splined shaft **140** which has a first gear **141** and a second larger gear **143** arranged thereon for rotation with the splined shaft **140**. The splined shaft **140** permits axial movement of the gears **141** and **143** along the shaft **140**.

In normal operation (as seen in FIG. **8**), the first gear **141** engages a third gear **145** arranged on a shaft **147** which drives the pulley **129**. The shafts **140** and **147** are arranged in suitable bearings **149** arranged at each end thereof. When the speed of the engine **21** drops below a predetermined level, a suitable sensor **151** responsive to the speed of the drive shaft **139** of the engine **21** generates a signal which is transmitted through a line **153** to a solenoid actuator **155** arranged within the gear box **131**. The solenoid actuator **155** moves the first and second gears **141**, **143** axially along the splined shaft **140** to the right as seen in FIG. **8** such that the second, larger gear **143** engages a fourth smaller gear **157** which is arranged on the shaft **147**. The ratio of the second gear **143** to the fourth gear **157** is preferably approximately 3 to 1.

In this way, when the speed of the engine **21** drops below the predetermined level as sensed by the sensor **151** (which predetermined level is insufficient to drive the compressor **109** at a speed sufficient to generate the 500-600 pounds of pressure which is preferably in the tank **23**), the solenoid actuator **155** is energized to slide the gears **143**, **141** axially along the splined shaft **140** so that the second, larger gear **143** engages the fourth, smaller gear **157** to drive the pulley **129** and hence the compressor **109** at a higher rate of speed to generate the desired pressure. When the speed of the engine increases above the predetermined level, in a preferred embodiment approximately 1500 rpm, the solenoid actuator **155** is deactivated by the sensor **151** thereby moving the gears **143** and **141** to the left as seen in FIG. **8** such that the first gear **141** re-engages with the third gear **145** to effectuate a 1 to 1 ratio between the output shaft **139** of the engine **21** and the pulley **129**.

The other bank of the V-8 engine has its exhaust ports arranged with adapter plates **103** similar to those on the first bank. However, the exhaust from this bank of the engine **21** is not collected and circulated through the compressor **109**. In a preferred embodiment, a portion of the exhaust is collected in a line **159** and fed to an enlarged chamber **161**. A second fluid is fed through a line **163** into the chamber **161** to be cooled by the cool exhaust emmanating from the engine **21** in the line **159**. The second fluid in the line **163** may be either transmission fluid contained in a transmission associated with the engine **21** or a portion of the oil used to lubricate the engine **21**. A second portion of the exhaust from the second bank of the V-8 engine is removed from the line **159** in a line **165** and used as a working fluid in an air conditioning system or for any other suitable use.

It should be noted that the particular arrangement utilized for collecting and distributing the gas exhausted from the engine **21** would be determined by the use for which the engine is employed. In other words, it may be advantageous to rearrange the exhaust tubing such that a larger or smaller percentage of the exhaust is routed through the compressor **109**. It should also be noted that since the exhaust lines **105** are plastic tubing, a rearrangement of the lines for a different purpose is both simple and inexpensive.

In operation of the engine of the present invention, the engine **21** is started by energizing the solenoid valve **29** and any suitable starting device (not shown), e.g., a conventional electric starter as used on an internal combustion engine. Compressed gas from the full tank **23** flows through the line **25** and a variable amount of the compressed gas is admitted to the distributor **33** by controlling the regulator valve **31** through the linkage **33** and the operator actuated throttle linkage **35**. The compressed gas is distributed to each of the lines **37** which lead to the individual cylinders **20**. The compressed gas is admitted to each of the cylinders **20** in timed relationship to the position of the pistons within the cylinders by opening the valves **39** with the valve actuator **41**.

When it is desired to increase the speed of the engine, the operator moves the throttle linkage **35** which simultaneously admits a larger quantity of compressed gas to the distributor **33** from the tank **23** by further opening the regulator valve **31**. The timing of the valve actuator **41** is also advanced through the linkage **104**. Still further, as the speed of the engine **21** increases, the effective length of the rotating contact **83** (FIG. 4) or **95** (FIG. 6) increases thereby electrically contacting a wider portion of one of the stationary radial contacts **85** to cause each of the valves **39** to remain open for a longer period of each engine cycle to admit a larger quantity of compressed gas to each of the cylinders **20**.

As can be seen, the combination of the regulating valve **31**, the mechanical advance **104**, and the valve actuator **41**, combine to produce a compressed gas engine which is quickly and efficiently adaptable to various operating speeds. However, all three of the controls need not be employed simultaneously. For example, the mechanical advance **104** could be utilized without the benefit of one of the varying valve actuators **41** but the high speed operation of the engine may not be as efficient. By increasing the duration of each engine cycle over which each of the valves **39** remains open to admit compressed gas to each of the cylinders **20** as the speed increases, conservation of compressed gas during low speed operation and efficient high speed operation are both possible.

After the compressed gas admitted to the cylinder **20** has forced the piston **22** downwardly within the cylinder to drive the shaft **139** of the engine, the piston **22** moves upwardly within the cylinder **20** and forces the expanded gas out through a suitable exhaust valve (not shown) through the adapter plate **103** (if employed) and into the exhaust line **105**. The cool exhaust can then be collected in any suitable arrangement to be compressed and returned to the tank **23** or used for any desired purpose including use as a working fluid in an air conditioning system or as a coolant for oil.

When using the apparatus and method of the present invention to adapt a ordinary internal combustion engine for operation with compressed gas it can be seen that considerable savings in weight are achieved. For example, the ordinary cooling system including a radiator, fan, hoses, etc. can be eliminated since the compressed gas is cooled as it expands in the cylinder. In addition, there are no explosions within the cylinder to generate heat. Further reductions in weight are obtained by employing plastic tubing for the lines which carry the compressed gas between the distributor and the cylinders and for the exhaust lines. Once again, heavy tubing is not required since there is little or no heat generated by the engine of the present invention. In addition, the noise generated by an engine according to the present invention is considerably less than that generated by an ordinary internal combustion engine since there are no explosions taking place within the cylinders.

The principles of preferred embodiments of the present invention have been described in the foregoing specification. However, the invention which is intended to be protected is not to be construed as limited to the particular embodiments disclosed. The embodiments are to be regarded as illustrative rather than restrictive. Variations and changes may be made by others without departing from the spirit of the invention. Accordingly, it is expressly intended that all such variations and changes which fall within the spirit and the scope of the present invention as defined in the appended claims be embraced thereby.

What is claimed is:

1. An apparatus for operating an engine having at least one cylinder and a reciprocating piston therein comprising:

a source of compressed gas;

distributor means connected with the source of compressed gas for distributing the compressed gas to the at least one cylinder;

valve means for admitting the compressed gas to the at least one cylinder when the piston is in approximately a top dead center position within the cylinder;

altering means for increasing the duration of each engine cycle over which the valve means admits compressed gas to the at least one cylinder as the speed of the engine increases; and

exhaust means for exhausting gas as the piston subsequently approaches approximately the top dead center position.

2. The apparatus of claim 1 further comprising control means for controlling the amount of compressed gas admitted to the distributor means.

3. The apparatus of claim 1 wherein the valve means is a solenoid valve secured in an opening in the cylinder above the level of the piston at the top dead center position.

4. The apparatus of claims 1 or 2 further comprising means for advancing the timing of the valve means as the speed of the engine increases such that compressed gas is admitted progressively further before the top dead center position as the speed of the engine increases.

5. The apparatus of claim 4 wherein the means for advancing the timing comprises a mechanical linkage connected to an operator actuated accelerator linkage.

6. The apparatus of claim 1 wherein a portion of the gas exhausted through the exhaust means is compressed in a compressor driven by an output shaft of the engine and is returned to the source of compressed gas.

7. The apparatus of claim 1 wherein a portion of the gas exhausted through the exhaust means is used to cool transmission fluid for a transmission associated with the engine.

8. The apparatus of claim 1 wherein a portion of the gas exhausted through the exhaust means is used as a working fluid in an air conditioning system.

9. The apparatus of claim 6 further comprising first gearing means interposed between the output shaft of the engine and the compressor for increasing the speed at which the compressor is driven.

10. The apparatus of claim 6 further comprising clutch means attached to the compressor both for disengaging the compressor from the output shaft of the engine when a first predetermined pressure at the source of compressed gas is exceeded and for engaging the compressor with the output shaft of the engine when the pressure at the source of compressed gas drops below a second predetermined pressure.

11. The apparatus of claim 9 further comprising means for both disengaging the first gearing means when a predetermined speed of the engine is exceeded and engaging a second gearing means for driving the compressor at a speed slower than the first gearing means when the predetermined speed of the engine is exceeded.

12. The apparatus of claim 1 wherein the valve means is electrically actuated and wherein the altering means comprises:

a rotating member timed with the at least one cylinder and arranged within a housing;

first and second contacts arranged on a first end of the rotating member and on an inside surface of the housing, respectively;

means for increasing the distance of the first contact from the rotational axis of the rotating member as the speed of the engine increases such that the first contact moves radially outwardly within the housing; and

said second contact presenting a longer arc length to the first contact as the distance of the first contact from the rotational axis of the rotating member increases.

13. The apparatus of claim 12 wherein the rotating member comprises an arcuate arm and wherein the means for increasing the distance of the first contact comprises pivotally mounting a second end of the arcuate arm about the axis of rotation of the rotating member and spring means for biasing the first end of the arcuate arm towards a radially inward position whereby the first end of the arcuate arm pivots radially outwardly as the speed of the engine increases.

14. The apparatus of claim 12 wherein the rotating member is axially slidably received on a rotating shaft for co-rotation therewith, said shaft having splines with a radial component in the direction of rotation, and wherein the first contact comprises a flexible contact located on an upper surface of the rotating member, said flexible contact being biased against the inside surface of the housing which carries the second contacts whereby as the speed of the engine increases the rotating member is urged axially along the splined shaft towards the inside surface of the housing such that the flexible contact is forced radially outwardly along the inside surface.

15. The apparatus of claim 12 wherein the second contact comprises of radially extending conductor arranged on an upper inside surface of the housing, said conductor increasing in arc length as the conductor extends radially outwardly from a central portion of the housing.

16. An apparatus for adapting an internal combustion engine for operation with compressed gas, the internal combustion engine having at least one cylinder, a piston reciprocable within the at least one cylinder, intake and exhaust means disposed in the at least one cylinder, and a tapped hole in the at least one cylinder adapted to receive a spark plug, the apparatus comprising:

a source of compressed gas;

distributor means connected with the source of compressed gas for distributing the compressed gas to the at least one cylinder;

valve means arranged in the tapped hole for admitting the compressed gas to the at least one cylinder when the piston is in approximately a top dead center position within the cylinder; and

altering means for increasing the duration of each engine cycle over which the valve means remains open to admit the compressed gas as the speed of the engine increases.

17. An apparatus as in claim 16 further comprising first adapter plate means for supporting the distributor means above an intake manifold of the engine, which adaptor plate means allows ambient air to enter through the intake manifold.

18. The apparatus of claim 16 further comprising second adapter plate means for reducing the exit area of the exhaust means.

19. A method of operating an engine on compressed gas, said engine having at least one cylinder and a piston reciprocable therein comprising the steps of:

- delivering compressed gas from a source to a distributor;
- distributing the compressed gas to the at least one cylinder;
- admitting compressed gas to the at least one cylinder through an intake valve when the piston is at approximately a top dad center position;
- increasing the duration of each engine cycle over which compressed gas is admitted to the at least one cylinder as the engine speed increases; and
- exhausting the remaining gas when the piston subsequently reaches approximately the top dead center position.

20. The method of claim **19** further comprising the step of controlling the amount of compressed gas which is delivered to the distributor.

21. The method of claim **19** further comprising the step of advancing the timing of the opening of the intake valve as the speed of the engine increases.

22. An apparatus for operating an engine having at least one cylinder and a piston reciprocable therein on compressed gas comprising:

- a source of compressed gas;
- distributor means connected with the source of compressed gas for distributing the compressed gas to the at least one cylinder;
- electrically actuated valve means secured in an opening in the at least one cylinder for selectively admitting compressed gas to the at least one cylinder when the piston is in approximately a top dead center position; and
- means for advancing the timing of the valve means as the speed of the engine increases whereby compressed gas is admitted progressively further before the top dead center position as the speed of the engine increases.

* * * * *

Antenna Tuning Unit Notes:

Back to Lab..

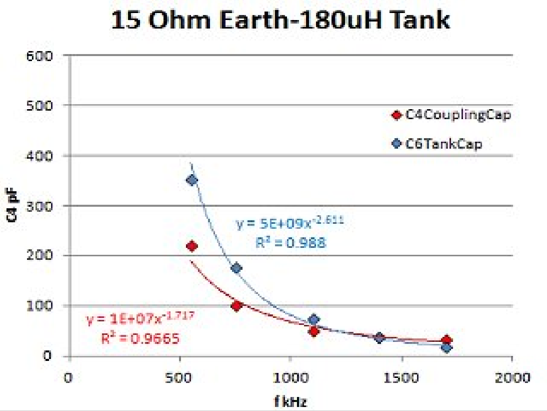

Introduction:

When designing my radios I always make extensive use of Mike Peebles and Dan Petersen's "Professor Coil" spreadsheet. This tool along with the many tutorials and explanations online have vastly eased the design work around homebrew coil winding. We can now model a coil with good accuracy prior to the actual job and thus better match the coil to the other components we plan to include. One quickly becomes aware that most coils for the broadcast band vary around a nominal 220 or so uH. More than this and most variable capacitors will have too much bottom-end capacitance to tune the top of the band. Less than this and you need caps with fairly high values when fully meshed, 500+pF or more.

All the above is pretty straight forward. On the other hand, when I have been reviewing double-tuned set designs, I often note that the coil used for the ATU (antenna tuning unit) will have an inductance quite a bit lower than the main tank coil. Professor Coil doesn't address this aspect of coil design and I have found little discussion online concerning ATU design. Mostly this seems to be dealt with in passing or as a digression when dealing with other subjects. The most useful web sites for addressing this aspect of crystal set design include Dick Kleijer's excellent work and Ramon Vargas's detailed analysis of the "Tuggle Front End", that's about it. Neither site discusses the ramifications of varying different parameters such as the Earth Resistance, Coil Inductance, etc although Kleijier's site includes a great calculator page to allow one to ask these questions. In my explorations of ATU design (primarily Tuggle) I have made extensive use of Kleijier'e page and for this I am deeply in his debt. My hat's off to Dick, thanks so much.

The following discussion presents the results of many models cranked through the design calculator. I wished to understand what factors play main roles and which have minor parts.

For a conventional detector-circuit coil one can pretty much control all the main factors, really only the inductance and capacitance of the circuit. The main and only surprise comes in the form of "stray capacitance" resulting from the spacing of the coil wires. This can readily be 1) guessed at and 2) minimized by good coil winding technique. For the antenna circuit by contrast many of the needed parameters, earth resistance, antenna capacitance and inductance, and antenna resistance in the forms of actual wire resistance and radiation resistance are, for most of us, unknown and only guessed at. The following figure illustrates the antenna and ATU with a list of the main "components" that need to be understood and/or modeled.

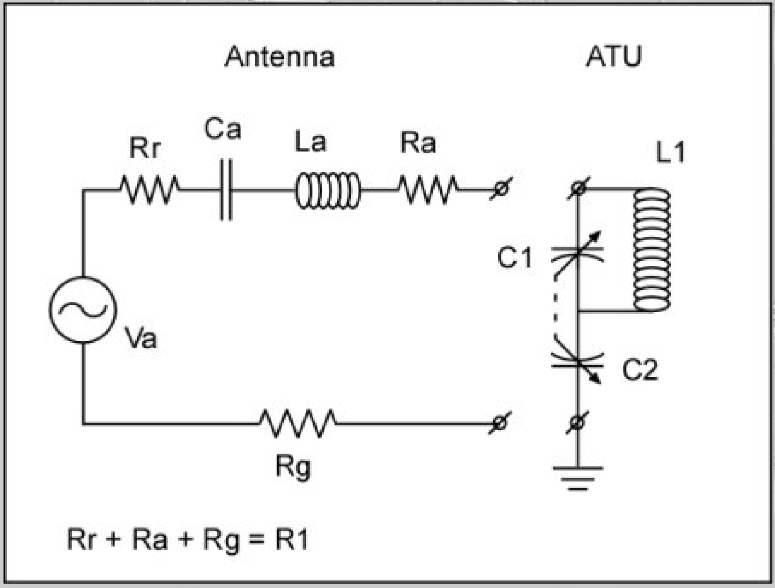

The Antenna consists of an AC voltage source (the signal of interest) Va, in series with some radiation resistance Rr, antenna capacitance Ca, antenna inductance La, antenna resistance Rr, and finally a ground resistance for the return path to complete the circuit. In addition the ATU consists of a coupling capacitance C2, and a tank with inductance L1 and capacitance C1. The capacitance and inductance of the ATU is needed to tune out the reactance of the various components of the antenna for which we do not have data. What to do?

Without an antenna analyzer (expensive) one can only estimate the many component values based on published antenna models. Ken Kuhn's engineering page has a number of nice models and a good discussion of the antenna parameters we will be trying to understand and use. For my personal setup my antenna consists of about 75' of 14awg wire averaging about 10-12' high and with another 25' of lead-in. For such an antenna Kuhn models a 30m antenna 3m high which approximates my situation closely enough. Wire resistance is negligible as is the radiation resistance which he shows to vary between some 0.1 to 1.5 ohms. Such an antenna is capacitive by nature and will have

about 220 - 375pF along with some small 20uH inductance, not too far off a standard "dummy" antenna. This pretty much leaves earth resistance Rg and the main unknown parameter.

For Rg one has the option of punting and taking the "Standard" value of 25 ohms. In my modeling I have found this parameter to be critical and highly sensitive. I do not recommend guessing here, it is recommended one go to the internet page of their state, county, or local government and search for reports on ground or soil resistivity (or conductivity). As this is an important agricultural and engineering parameter, it has been surveyed for most places and should be available with some effort. Effort well rewarded. Earth resistivity for the Texas Gulf Coast is a mercifully low 10 - 15 ohm meters, but for many regions this will not be the case.

Earth resistivity depends on a number of variables including the material, moisture, mineral salt, and temperature. A table of typical ranges in ohm meters for some different soils follows:

Loam	5 - 50 ohm meter
Clay	4 - 100
Sand/Gravel	50 - 1,000
Limestone	5 - 10,000
Sandstone	20 - 2,000
Granite	1,000 - 2,000
Slates	600 - 5,000

Moisture up to about 17% dramatically lowers resistivity, mineral salts are needed as pure water is an insulator, and as the temperature approaches freezing the resistivity also rises dramatically. All these factor into your estimation of Rg. The actual resistivity seen by the circuit depends on the earth resistivity and the type of grounding system you have installed. The more metal in the ground and deeper, the lower the resistance. Know thy earth! A map of USA soil resistivity follows:

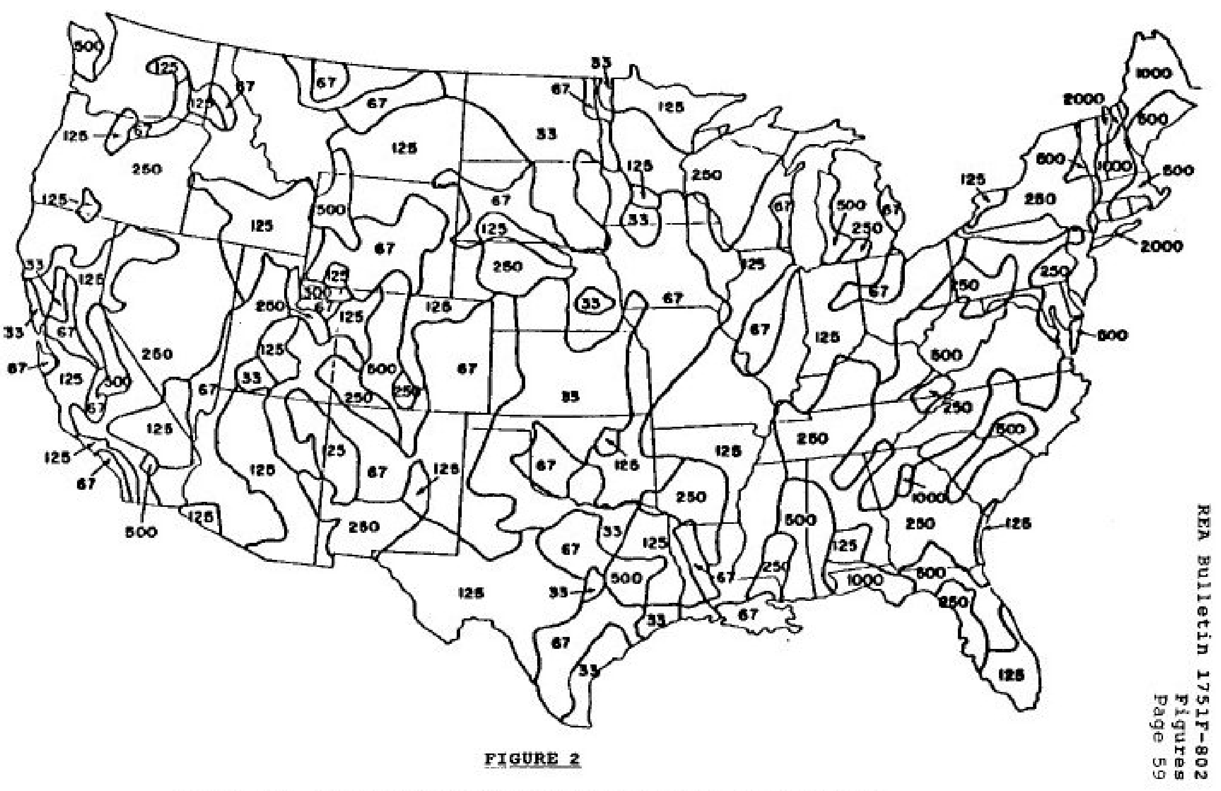
500
125
67
67
67
125
33
33
125
250
125
125
250
125
500
67
125
250
33
67
67
250
500
250
67
1000
2000
500
800
1000
800
2000
800
125
500
250
250
67
125
250
33
33
67
125
67
250
33
250
500
250
125
125
1000
250
125
500
125
67
250
125
67
500
33
67
33
500
125
250
800
250
125
1000
67
33
33
REA Bulletin 1751F-802
Figures
Page 59

FIGURE 2

Estimated Average Earth Resistivity in U.S. (ohm·meters)

Modeling:

All the above discussion is great but... What is important really? The following section presents the results of applying different parameters to get a feel for the sensitivity of it all. To begin with I need explain my base assumptions. All the models are based on Kleijer's calculation spreadsheet which requires inputs for the following parameters:

Frequency
Coil inductance
LC circuit Q unloaded
*** Complex impedance of antenna or**

Series Resistance (Rg+Ra+Rr)
Series Capacitance Ca
Series Inductance La

*** If the series values are given, the complex impedance is calculated.**

1) Frequency is chosen per your model, I have taken models at F = 550-1100-1700 kHz.
2) Coil inductance one has control over when winding. I have chosen base values that resonate with about 400pF tank capacitance, with some sensitivities.
3) Unloaded Q I do not have. For modeling I have taken Vargas's measures for a 4.5" coil wound with 660/46 Litz wire. I assume this will be about as good a coil as one can wind. I also made a sensitivity for lower Q. Q is a function of frequency and I have used the appropriate value of Q to match the modeled frequency.
4) Series Resistance is basically what the earth and ground system will deliver plus a small 1-2 ohm contribution from Ra and Rr. I have modeled four cases 10, 15, 25 and 50 ohms. The first two cases reflect my needs on the Gulf Coast with low resistivity soil. The 25 ohm case is a "typical" or "standard" ground. Finally, many will have much higher resistivity soils and you need to know just how difficult things can get!
5) Series capacitance I based for my specific antenna on the model of Ken Kuhn (Mathematical Model of Wire Antenna). The capacitance of a 30m antenna 3m high he calculated to range from 220pF and low frequencies to 375pF and high frequencies. I input the correct capacitance to match the frequency modeled.
6) Series inductance I just input 20 uH every time.

The data and plots for four different scenarios follow:

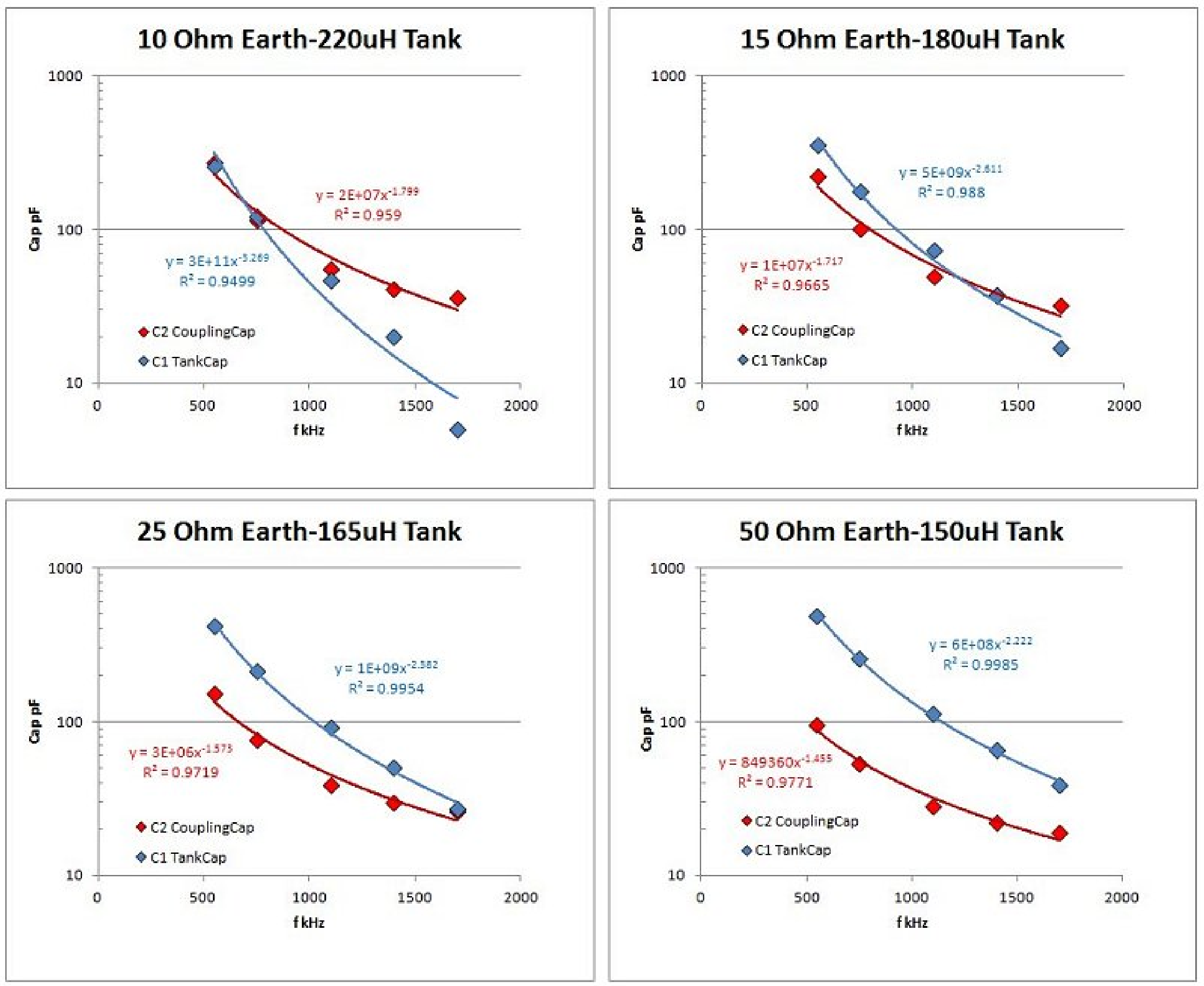

The above plots show the calculated capacitance versus frequency for four different models of Earth Resistivity / Tank Inductance. In each case I maintain the same log capacitance vs frequency scale for easy comparison. The cases modeled represent increasing earth resistance presented to the ATU from a very low 10 ohms to a fairly high 50 ohms. Actual earths

can go up to two orders of magnitude higher. I chose L1 inductances such that the maximum needed capacitance would approach 500pFs, easily found on many variable caps. With respect to the coupling cap (C2, red curve), as the earth resistivity increases, the needed capacitance declines. The tank cap value needs to resonate with the inductor and increases with smaller-value inductance. In searching for a 500pF cap max value, the choice of inductance needs to decrease with increasing earth resistance. Parameters other than Rg, L1 and C1-C2 have minor impact on the models.

Conclusions:

From the plots one can readily see that above 15 ohm earth the two variable capacitors on the ATU do not track well. As the earth resistivity increases, the worse the tracking. Most locations are not blessed with a low resistivity earth and this needs to be factored into the ATU design. There is little to be done, changing the design coil inductance will change the needed capacitance on BOTH C1 and C2, (larger L1 leads to smaller C1 and C2 for resonance and vice versa). A possible solution well worth trying is to use a dual-gang capacitor where one section has a different value than the other, the above plots give an easy way to decide the max values needed. Use the smaller gang on the coupling circuit and the larger on the tank. Experience tells us that ganging the capacitors on a "Tuggle" front end works well, but from the models I have to imagine one might squeeze a bit more performance by giving up the convenience of one-dial tuning on the ATU, especially where your earth has a fairly high resistance.

The best design methodology requires that you know first and critically your earth resistance. If this is just a guessed-at parameter then I strongly urge you to put down the Litz, set aside the silver-plated-ceramic-insulated caps, box the holy-grail diodes and go do some research on your local ground. All the time, expense, and effort on the greatest "state-of-the-art" crystal receiver will be wasted if the ground is not attended to. If the resistivity of your earth is high in ohm-meters, get more metal in the ground or consider a counter-poise. The following table from LEM Instruments shows the relation between earth resistivity in ohm meters and actual earth resistance in ohms as a function of the earthing method used.

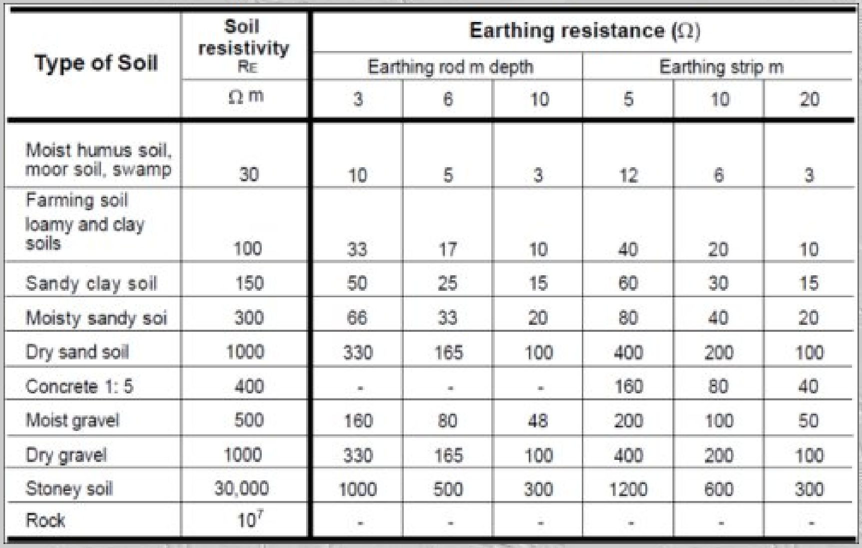

Type of Soil	Soil resistivity R_E	Earthing resistance (Ω)					
		Earthing rod m depth			Earthing strip m		
	Ω m	3	6	10	5	10	20
Moist humus soil, moor soil, swamp	30	10	5	3	12	6	3
Farming soil loamy and clay soils	100	33	17	10	40	20	10
Sandy clay soil	150	50	25	15	60	30	15
Moisty sandy soi	300	66	33	20	80	40	20
Dry sand soil	1000	330	165	100	400	200	100
Concrete 1: 5	400	-	-	-	160	80	40
Moist gravel	500	160	80	48	200	100	50
Dry gravel	1000	330	165	100	400	200	100
Stoney soil	30,000	1000	500	300	1200	600	300
Rock	10^7	-	-	-	-	-	-

Given just how little discussion there is on the web concerning antenna tuning and the impact of ground resistivity I hope this page can serve as a starting point for further research. Certainly there is more to be said, Crystal Radio, its a bottomless pit!

Kevin Smith
09/2011

Models and Data:

```
10 ohm / 220 uH  Lowest resistivity Earth Case sensitivities
Ra      Ca      La      L5      Qu      f       C4 cpl  C6 tank Za Impedance    LC Rp
Ohm     pF      uH      uH              kHz     pF      pF          Ohm         kOhm
10      220     20      220     475     550     442     228     10  -j  1246    361
10      300     20      220     400     1100    68      36      10  -j  344     608
10      375     20      220     200     1700    44      -3      10  -j  36      470

10      220     20      220     700     550     273     255     10  -j  1246    532
10      260     20      220     640     750     114     122     10  -j  721     663
10      300     20      220     585     1100    55      47      10  -j  344     890
10      340     20      220     440     1400    41      20      10  -j  158     851
```

10	375	20	220	300	1700	36	5	10 -j 36	705
10	220	20	200	700	550	304	287	10 -j 1246	484
10	300	20	200	585	1100	58	54	10 -j 344	807
10	375	20	200	300	1700	38	7	10 -j 36	641

15 ohm / 180 uH Low resistivity Earth Case sensitivities

Ra	Ca	La	L5	Qu	f	C4 cpl	C6 tank	Za Impedance	LC Rp
Ohm	pF	uH	uH		kHz	pF	pF	Ohm	kOhm
15	220	20	180	475	550	337	328	15 -j 1246	295
15	300	20	180	400	1100	61	63	15 -j 344	498
15	375	20	180	200	1700	40	10	15 -j 36	385
15	220	20	180	700	550	220	352	15 -j 1246	435
15	260	20	180	640	750	100	176	15 -j 721	543
15	300	20	180	585	1100	49	73	15 -j 344	728
15	340	20	180	440	1400	37	37	15 -j 158	697
15	375	20	180	300	1700	32	17	15 -j 36	577
15	220	20	130	700	550	313	511	15 -j 1246	314
15	300	20	130	585	1100	59	110	15 -j 344	526
15	375	20	130	300	1700	38	30	15 -j 36	417
15	220	20	205	700	550	195	302	15 -j 1246	496
15	300	20	205	585	1100	45	61	15 -j 344	829
15	375	20	205	300	1700	30	13	15 -j 36	657

25 ohm / 165 uH "Standard" resistivity Earth Case sensitivities

Ra	Ca	La	L5	Qu	f	C4 cpl	C6 tank	Za Impedance	LC Rp
Ohm	pF	uH	uH		kHz	pF	pF	Ohm	kOhm
25	220	20	165	475	550	213	396	25 -j 1246	271
25	300	20	165	400	1100	48	84	25 -j 344	456
25	375	20	165	200	1700	32	22	25 -j 36	352
25	220	20	150	700	550	164	462	25 -j 1246	363
25	300	20	150	585	1100	41	102	25 -j 344	606
25	375	20	150	300	1700	27	31	25 -j 36	481
25	220	20	165	700	550	151	416	25 -j 1246	399
25	260	20	165	640	750	76	213	25 -j 721	498
25	300	20	165	585	1100	39	91	25 -j 344	667
25	340	20	165	440	1400	30	50	25 -j 158	639
25	375	20	165	300	1700	26	27	25 -j 36	529
25	220	20	200	700	550	130	335	25 -j 1246	484
25	300	20	200	585	1100	35	72	25 -j 344	807
25	375	20	200	300	1700	24	20	25 -j 36	641

50 ohm / 150 uH High resistivity Earth Case sensitivities

Ra	Ca	La	L5	Qu	f	C4 cpl	C6 tank	Za Impedance	LC Rp
Ohm	pF	uH	uH		kHz	pF	pF	Ohm	kOhm
50	220	20	150	475	550	128	476	50 -j 1246	246
50	300	20	150	400	1100	34	108	50 -j 344	415
50	375	20	150	200	1700	24	35	50 -j 36	320
50	220	20	150	700	550	96	490	50 -j 1246	363
50	260	20	150	640	750	53	256	50 -j 721	452
50	300	20	150	585	1100	28	113	50 -j 344	606
50	340	20	150	440	1400	22	65	50 -j 158	581
50	375	20	150	300	1700	19	39	50 -j 36	481
50	220	20	125	700	550	110	595	50 -j 1246	302
50	300	20	125	585	1100	31	139	50 -j 344	505
50	375	20	125	300	1700	21	49	50 -j 36	401

Return..

Back to Main....

Friday, 05-Jan-2018 15:30:34 CST

Comments and/or Correspondance to: kj-smith@sbcglobal.net

.::. Alsace .::. Life List .::. Ionosphere .::. Crystal Radio .::. Snapshots .::. Thoughts .::.
.::. Sitemap .::. Beer Labels .::. International Pix .::. Western US Pix .::. Pinhole Pix .::. Petrology .::.

Crystal Radio Capacitors and ESR

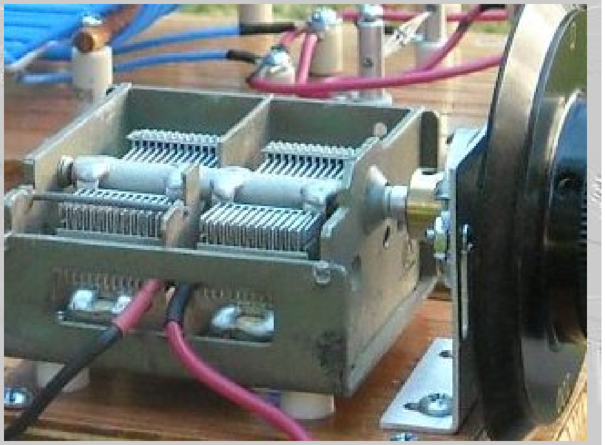

Back to Lab..

In my ongoing effort to understand crystal radios and the theory that makes them work I have recently began looking into all things capacitance. The tuning capacitor is, next the coil itself, the principal component of the set. Since few builders fabricate their own tuning caps I suppose one takes them a bit for granted. I have been reading on the radio forums some threads concerning radio Q and the subject of the variable cap Q seems to be getting more attention. So, what is the deal?

I set about my study by gathering what I could find on the web and at the close of this note I include a short reference list of useful pages from which I extracted data the ideas. I wish to extend special thanks to Ben Tongue who, as usual, has provided essential measurements and background in his Article 24. The plot below shows capacitor Q versus frequency for a small number of air-variable capacitors that I have in the box. It is a standard plot with log Q versus log frequency in MHz. On the plot, a bit messy I admit, I show data for six capacitors along with the Q versus f equation. You immediately see that for each capacitor the relation follows a power function and in all cases the function is good to R-squared at or above 99%. Additionally, I include the same data for two capacitors presented in Ben Tongue's excellent summary. It is also immediately obvious that not all caps are created equal!

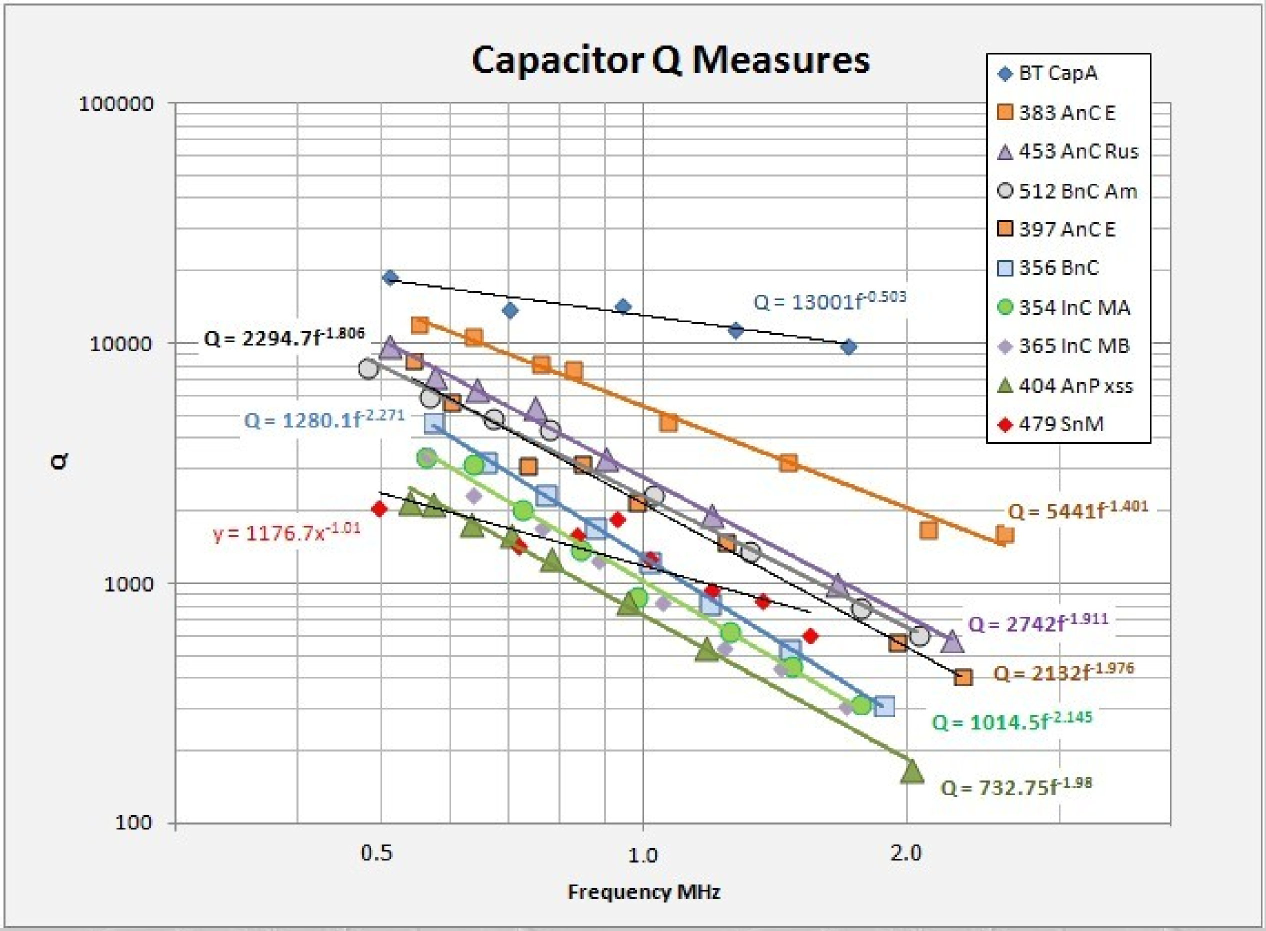

I have named the caps in the following manner, the highest tested RF capacitance, materials of interest (A = aluminum plates, B = brass plates; C = ceramic insulators, P = phenolic insulators); finally something to do with the origin. Photos of the caps at the bottom of this page.

Identification of the caps as follows (from highest to lowest Q):

BT CapA	Ben Tongue's CapA is the best "Holy Grail" of the lot
383 AnC E	A Russian (I presume) eBay cap from Estonia
453 AnC Rus	A Russian cap frequently offered on eBay
512 BnC Am	A brass straignt-line frequency cap AMSOO
397 AnC E	Another Russian (I presume) eBay cap from Estonia, larger capacitance
356 BnC	Another Lovely straight-line frequency brass cap
354 InC M	Chinese caps (A and B) offered from Ming on eBay
404 AnC xss	A typical Crystal Set Society cap
479 SnM	Switched set of individual Silver Mica caps, not an air-variable

Any discussion of component quality needs first to acknowledge that the fight to increase the component Q is in reality a fight to eliminate (not possible) or reduce as much as feasible the resistive losses associated with that component. It helps then to look not just at Q, but at how Q relates to the component series resistance Rs or ESR. There are two fundamental equations that describe this relation: 1) Q = cotan theta = 1/D. D is the dissipation factor and is an attribute of the dielectric material used in the construction of the capacitor. D = tan theta = ESR / Xc = 2pi f C * ESR. Solving for ESR we get: 2) ESR = D / 2pi f C. This equation tells us that capacitor losses expected due to resistance are dependant on both angular frequency and capacitance. It is important to remember that not all Q's are created equal. What we want is to keep the ESR to a minimum.

The plot below summarizes the same data only this time plotting log Q versus log Rs (ESR). Here most of the data sets fall into a narrow zone running from High-Q / low ESR in the upper left to Low-Q / high ESR in the lower right. Ben Tongue's Holy Grail cap sits above the pack with an overall narrower range of ESR from max to min capacitance. No surprises, good caps are low-loss. I have also plotted lines of Q versus ESR at 1MHz for threee capacitance values, 10, 100, and 1000 pF. These lines indicate that Q is highest at high values of capacitance independant of individual capacitor loss. Note that all the caps measured straddle the 100 pF line. For typical MW coils with about 230 uH or so, 100 pF resonates around 1MHz.

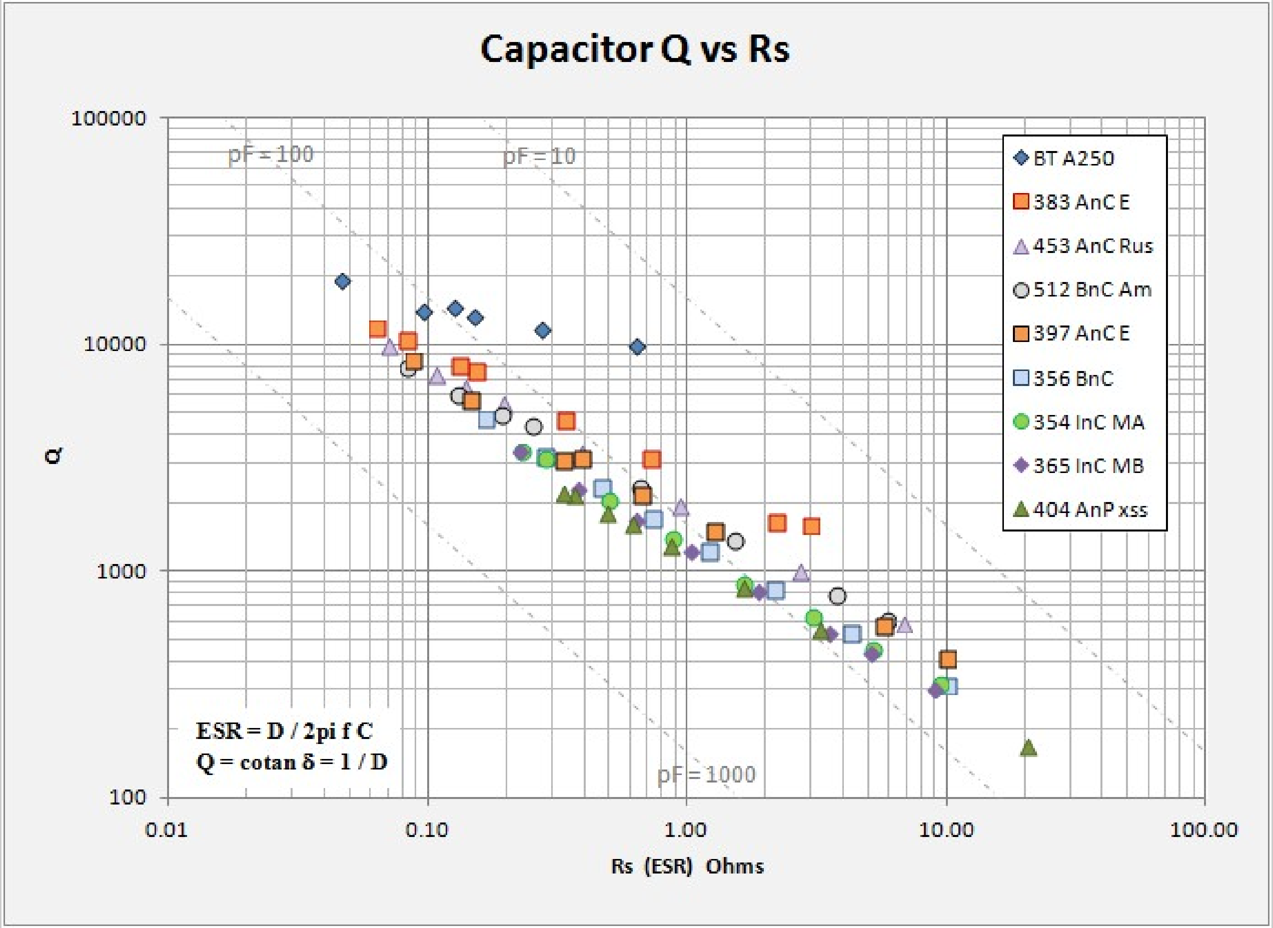

From the datasets above it can be seen that high-Q caps are high-Q across the band, but Q varies and is highest at high capacitance, (lower frequency) and vice versa. All sets show a direct power-law relation between Q and f or ESR

regardless of overall quality or type of capacitor. The data table from Ben Tongue (below) was found in his Article 24 and represent precision measuremnts by Bill Hebbert.

pF		f	Q		
cap	tank	MHz	Cap A	Cap B	
355	375	520	19000	4750	
170	190	730	14000	3230	
94	114	943	14400	1610	
81	101	1000	13148	1828	(my estimate)
40	60	1300	11600	870	
15	35	1710	9800	460	

for I = 250 uH and 20pF stray in tank

Some notes on my results. It should be immediately clear that I have yet to snag an actual "holy grail" cap such as the one evaluated by Ben Tongue. Ben's good CapA has ceramic insulated stators and silver-plated brass plates with silver-plated wiper contacts. Essential features of a good cap. Inexpensive caps often have lossy phenolic insulators, something easily noticeable before you buy, look for this. I have noticed that on some caps the stator is held by ceramic "buttons" rather than a large ceramic plate and I have to wonder how effective this is. I would say that ceramic insulation is far more available that silver plating on the plates and wipers.

At the bottom end of the quality profile is the poor but readily available XSS variable cap (404 AnP). This has aluminum plates, phenolic insulation and a single aluminum wiper. An OK basic cap, but one easily beat if you are willing to make an effort. Next up in terms of quality was a big disappointment for me. These are two Chinese caps I recently purchased on eBay (354 and 365 InC). They are large and heavy units with big ceramic insulators and I figured this ought to be good. Both caps have decidedly lower Q that my vintage brass straight-line frequency caps. I do not think these will find their way into any radios here.

In the mid-Q range I find one brass cap (512 BnC) and two Russian caps (453 AnC, and 397 AnC). These all have ceramic insulation but no silver plating. The brass cap insulators are a black material that I call ceramic as it does not appear phenolic, but frankly I am unsure what it is. The cap is pretty good though. The two Russian caps are commonly available on eBay and are of two different styles. One somewhat large with thick "gun metal" type casing and the other a small delicate unit with a plastic dust-cover. Both are pretty good caps.

Finally and best cap that I have in my box is another Russian cap (383 AnC). This cap is clearly an unused NOS unit with clean shiny parts and button ceramic insulators. This is a slightly smaller version of one of the previous caps (397 AnC). The moderate Q version was also clean and NOS, so I do not expect all such caps to be automatically good Q. It is a pity as I have seen additional of these small caps on eBay, though clearly in used condition so I suspect they will have lower Q.

From all the above I would conclude that caps having ceramic insulation alone is not nirvana. Caps need also to be clean and ultimately need silver plating to truly aspire to "holy grail" status. A good reference worth looking at is a thread on The Radio Board: Improving 'Q' of capacitors.
The post shows how merely taking an older cap and cleaning it up (no small task) and replacing the insulators with HDPE sheet (high-density polyethylene) can take an original Q = 664 (1MHz) and raise it to Q = 12248 (1MHz). This value is on par with Ben Tongue's "holy grail" cap (Q = 13148 @ 1MHz) and either it was a dim fine cap to start with or something not fully kosher with the measurements. Still, the point is that improvements can be significant.

Images of the tested caps, same scale.

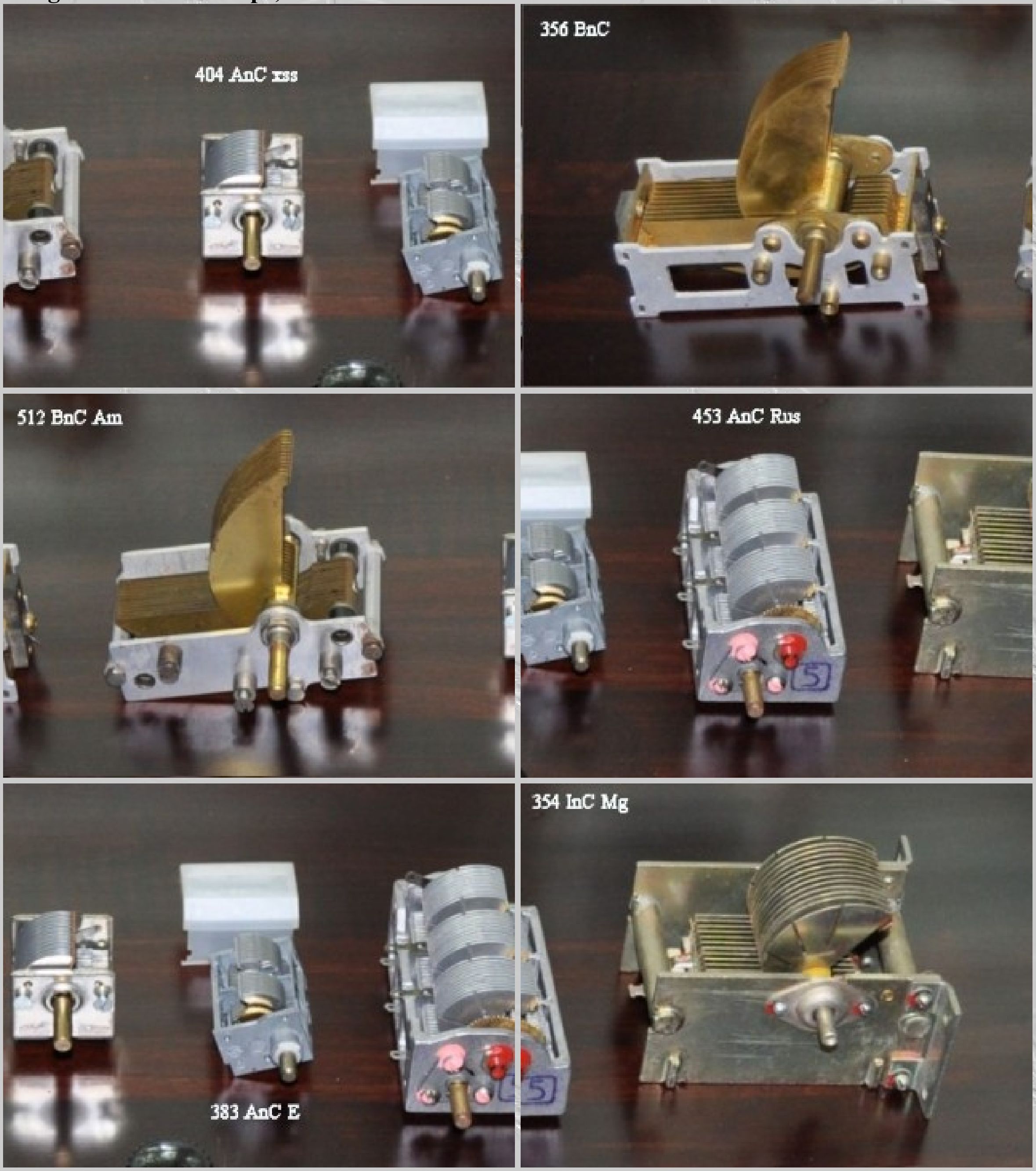

kjs (updated 6/2016)

References:

**AVX Basin Capacitor Formulas.
http://www.avx.com/docs/Catalogs/cbasic.pdf**

**Ben Tongue 1999, Sensitivity and selectivity issues in crystal radio sets including diode problems; measurements of the Q of variable and fixed capacitors, RF loss in slide switches and loss tangent of various dielectrics.
http://www.bentongue.com/xtalset/24Cmnts/24Cmnts.html**

**Wikipedia, Types of Capacitors, Ohmic Losses, Dissipation Factor and Q.
http://en.wikipedia.org/wiki/Types_of_capacitor#Ohmic_losses.2C_ESR.2C_dissipation_factor.2C_and_quality_factor**

Dr. Gary L. Johnson, 2001. Lossy Capacitors
http://www.eece.ksu.edu/~gjohnson/tcchap3.pdf

A short note concerning my Q measurements as compared to those of Bill Hebbert and presented by Ben Tongue in his Article 24. By some chance, both Ben's work as well as my own include a commonly available 365 pF capacitor from the XSS and elsewhere. While his cap and my own are not the same cap, they ought to be similar enough. Therefore it is worth to take a look ar the two datasets together.

In the plot below of Q vs ESR I show Ben's three measurement sets as well as my own measuremnts on the same type cap. Ben's A250 (holy grail) is included just for show. His measurements on CapB were made with two separate inductors, one with 250 uH and one with 180 uH. My standard inductor has 225 uH. On the plot I include a trend for Ben's 250 uH and my 225 uH data. It is clear that my measurements are systematically lower than his.

At this time I am uncertain as to whether my setup delivers lower Q measurements or if the partictular cap I measured was a bit more lossy than Ben's. This is a bit of ambiguity in my presentation that the reader ought to be aware of. Maybe my caps are better than presented? I do not know the answer but I include this plot just so you know.

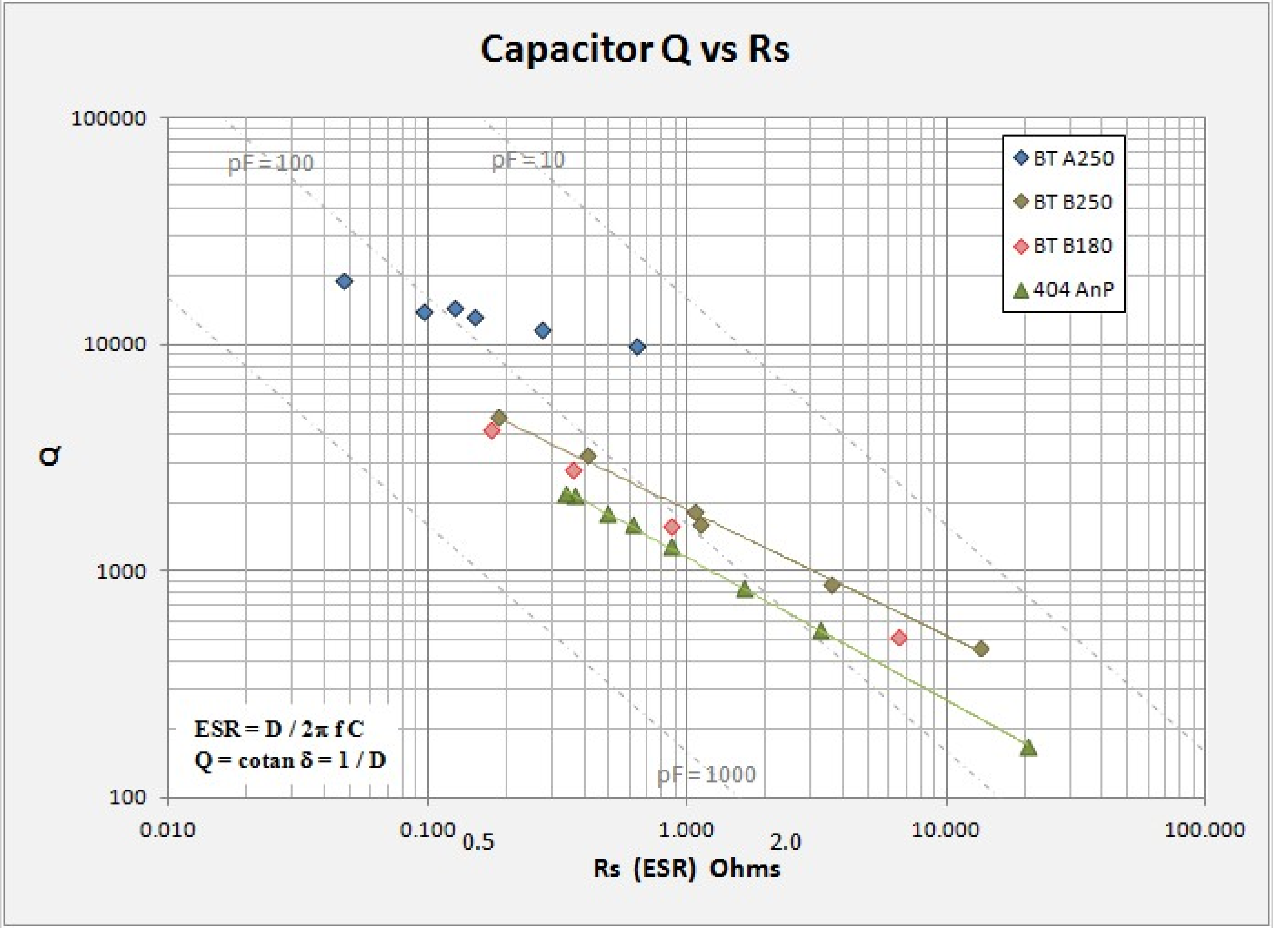

A note concerning Loss Tangent

When looking into capacitor losses one often runs into the term "loss tangent". For many our eyes just glaze over and we read on understanding only that it has something interesting to do with theory. In the interest of knowing the code I give a brief explanation.

With a perfect lossless capacitor in AC operation (sine wave) current leads the voltage by 90 degrees. Where losses occur the current will lead by slightly less than this. If you imagine these losses as a resistance in series with a perfect

capacitor, then this is called Effective Series Resistance, ESR and is valid in reference to a specific frequency: ESR = D / 2pi * f * C.

The lag angle between current and voltage is generally a fraction of a degree and is expressed as a positive value. The tangent of this lag angle is the "loss tangent" and is referred to more commonly as the Dissipation factor D. For example, polystyrene has a D of about 0.0001 for a lag angle of 0.006 degrees. Good dielectrics have very small loss tangents. Why do we express loss as D or loss tangent? Well, it related to the quality factor Q such that Q = 1/D.

Modeled Data from plots shown above:

f MHz	L uH	C pF	Qtank	Qcoil	Qcap	Rp Mohm	ESR ohm	D	Qvsf	R²
Brass n Ceramic								356 BnC	Q = 1280.1 * f^-2.271	0.9991
1.878	267	27	142	258	315	0.99	10.03	0.003179		
1.461	252	47	174	256	541	1.25	4.27	0.001849		
1.183	241	75	190	245	835	1.50	2.14	0.001197		
1.013	234	105	198	235	1240	1.85	1.20	0.000806		
0.875	229	144	199	225	1735	2.19	0.726	0.000576		
0.769	225	190	198	216	2358	2.56	0.461	0.000424		
0.655	221	268	192	205	3238	2.94	0.280	0.000309		
0.572	217	356	188	196	4713	3.69	0.166	0.000212		
Aluminum n Ceramic Estonia A								383 AnC E	Q = 5441 * f^-1.401	0.9901
2.587	295	13	196	223	1609	7.70	2.98	0.000622		
2.116	277	20	218	251	1671	6.15	2.20	0.000598		
1.460	251	47	237	256	3197	7.38	0.72	0.000313		
1.066	236	94	227	238	4676	7.40	0.34	0.000214		
0.827	227	163	215	221	7701	9.09	0.153	0.000130		
0.761	225	195	209	215	8135	8.74	0.132	0.000123		
0.636	220	285	199	203	10574	9.29	0.083	0.000095		
0.552	217	383	190	194	11912	8.96	0.063	0.000084		
Aluminum n Phenolic XSS								404 AnP xss	Q = 732.75 * f^-1.98	0.9922
2.019	273	23	101	255	169	0.58	20.52	0.005929		
1.176	241	76	169	245	547	0.97	3.25	0.001827		
0.957	232	119	181	231	845	1.18	1.65	0.001184		
0.783	226	183	186	217	1286	1.43	0.86	0.000778		
0.703	222	230	185	209	1604	1.58	0.613	0.000624		
0.632	220	288	182	202	1790	1.56	0.488	0.000559		
0.573	218	354	180	196	2149	1.68	0.365	0.000465		
0.539	216	404	176	192	2185	1.60	0.335	0.000458		
Aluminum n Ceramic Russia 3-gang								453 AnC Rus	Q = 2742 * f^-1.911	0.9966
2.248	282	18	173	245	588	2.34	6.76	0.001699		
1.661	259	35	206	259	1007	2.72	2.69	0.000993		
1.191	241	74	218	246	1933	3.49	0.93	0.000517		
0.899	230	136	212	227	3329	4.33	0.39	0.000300		
0.748	224	202	206	214	5435	5.73	0.194	0.000184		
0.642	220	279	197	203	6451	5.73	0.138	0.000155		
0.574	218	354	191	196	7358	5.77	0.107	0.000136		
0.510	215	453	185	189	9783	6.74	0.070	0.000102		
Brass n Ceramic Amsoo								512 BnC Am	Q = 2294.7 * f^-1.806	0.9915
2.057	274	22	179	253	607	2.15	5.84	0.001647		
1.763	263	31	195	259	787	2.29	3.70	0.001270		
1.318	246	59	212	251	1356	2.76	1.50	0.000737		
1.021	235	104	214	236	2332	3.51	0.65	0.000429		
0.776	225	187	206	216	4371	4.80	0.251	0.000229		
0.669	221	255	198	206	4866	4.53	0.191	0.000206		
0.564	217	367	189	195	6005	4.62	0.128	0.000167		
0.481	214	512	181	185	7834	5.07	0.083	0.000128		
High-stability Invar n Ceramic Ming A								354 InC MA	Q = 1014.5 * f^-2.145	0.9871
21.769	263	31	142	259	315	0.92	9.29	0.003173		
1.475	252	46	164	256	455	1.06	5.14	0.002199		
1.252	244	66	178	248	631	1.21	3.04	0.001585		

```
0.980   233     113     184     233     880     1.26    1.63    0.001136
0.845   228     156     192     222     1389    1.68    0.871   0.000720
0.724   223     216     192     211     2050    2.08    0.496   0.000488
0.637   220     284     190     203     3161    2.78    0.278   0.000316
0.562   217     369     184     195     3351    2.57    0.229   0.000298

Aluminum n Ceramic Estonia B                                    397 AnC E       Q = 2132 * f^-1.976     0.9875
2.313   284     17      152     242     411     1.70    10.06   0.002436
1.940   270     25      177     257     573     1.88    5.74    0.001745
1.239   243     68      213     248     1499    2.84    1.26    0.000667
0.978   233     114     210     233     2168    3.10    0.66    0.000461
0.847   228     155     208     222     3125    3.79    0.388   0.000320
0.733   224     211     199     212     3105    3.20    0.332   0.000322
0.598   218     324     192     199     5687    4.67    0.144   0.000176
0.543   216     397     188     193     8446    6.24    0.087   0.000118

High-stability Invar n Ceramic Ming B                           365 InC MB      Q = 927.01 * f^-2.170   0.9960
1.697   261     34      140     259     306     0.85    9.08    0.003269
1.434   250     49      162     255     441     0.99    5.12    0.002270
1.231   243     69      169     247     535     1.01    3.51    0.001868
1.045   236     98      184     237     824     1.28    1.88    0.001213
0.887   230     140     191     226     1244    1.59    1.029   0.000804
0.762   225     194     191     215     1685    1.81    0.639   0.000594
0.636   220     284     186     203     2338    2.06    0.376   0.000428
0.565   217     365     184     195     3386    2.61    0.228   0.000295
```

Return to Lab..

Back to Main....

Friday, 05-Jan-2018 15:30:39 CST

Comments and/or Correspondance to: kj-smith@sbcglobal.net

.::. Alsace .::. Life List .::. Ionosphere .::. Crystal Radio .::. Snapshots .::. Thoughts .::.
.::. Sitemap .::. Beer Labels .::. International Pix .::. Western US Pix .::. Pinhole Pix .::. Petrology .::.

Part I: Diode Test and Protocol

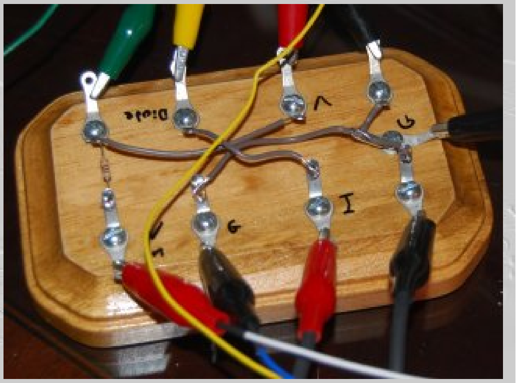

Back to Lab..

The following explains my first attempt to compare a measured diode characteristics with calculated plots based on the Shockley Equation.

I have been taking a serious look into diode modeling and measurement recently in an attempt to better understand these solid-state successors to the humble crystal and cat's whisker detector. There is much good information on diodes to be found in many excellent web sites of course, but there is nothing quite like actually making the measurements and working with them to get a good feel. This page reports some of my protocols and test setups which I have found to be useful.

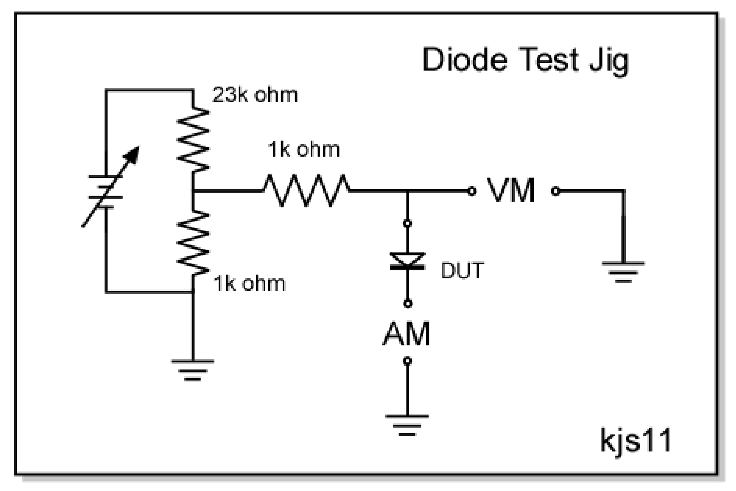

This is a very simple circuit allowing power input, attaching a diode for testing, and sampling both the curent and voltage across the diode under test, (DUT). I made a simple modification by adding a voltage divider, about 20:1 at the power input. The power supply can be connected either at the front across the divider, or after the divider in front of the 100k current limiting resistor for higher voltage readings. I get the input voltage from a modest 18V, 2A regulated DC power supply. This unit has both volt and amp controls with digital readouts, nice to use and doesn't take up much room on my desk. Here is a case where I recommend you look at getting a new unit. I looked long and hard on ebay, but found few bargins and few units with the same features. Still, the supply regulator was never designed to adjust the output to precise small values making it somewhat difficult/tedious to adjust. Adding a small rheostat to the test jig may aid for making fine adjustments on the voltage. Another piece of equipment you will need for this test is an amp meter. At first I used a nice small digital meter I found new on ebay for $15 shipping included. It reads amps to a lowest range of 200uA, more than enough precision for the task. For that price get two. Later as I got to making more precision measurements for the determination of diode n and Is I acquired a Keithley 195A bench volt/amp meter with very high input impedance. The DVM is, of course, my trusty Keithley workhorse.

I have received occasional criticism for the simple test jig with the suggestion that proper measurements cannot be taken with the jig as-is. I note that, while the test setup is simple, the instrumentation is high quality and performs the job well. I have had the opportunity to collaborate with Mike Tuggle, comparing our respective measurements. Even my early characteristic curves taken with the small amp meter compare well and provide a good view of the diode and determination of the diode Rs series resistance. The detail n and Is measurements presented here essentially match, though at somewhat less precision, those data from Mike. The reader should have some assurance that the data and measurements found on this page are vetted and reliable.

Easy protocol as follows:

1) Set up the test with the diode and meters clipped to the jig.
2) Adjust the power supply until the reading on the DVM meter is where you plan to measure. Adjust the voltage (DVM readout) in steps from 0 to 1.0 Volt in 0.1 V increments and record the current (Amps) in a spreadsheet.
3) Repeat at each tenth volt recording while the spreadsheet makes a sweet graph.

With this protocol you get a good look at the diode Current to Voltage (I/V) Characteristic out to about 1.5 V or slightly higher, well into the region where series resistance dominates the curve. This curve unfortunately does not communicate much about how well the diode will perform under small signal reception with a crystal radio. For this kind of information one needs to push the simple test jig described above somewhat further, down to low voltage / low current regions where DX level signals predominate. To understand your diode and to better match your diode with the set, the needed information concerns the diode resistance at the zero-crossing point, Ro, or Rd depending on the source of information you are using. The value of Ro cannot be measured directly but may be calculated from I/V measurements at very low currents.

Most probably the single best web article for the measurement of diode Is and n parameters is given by Ben Tongue in his Article 16. In this article he describes an interesting circuit for the measurement as well as protocols for making them. For my purposes I did not wish to make this a construction project and felt I might get along with good quality meters and my already-built "Diode Test Jig". Essentially Ben Tongue's method consists of making two precision measurements of Voltage and Current through a diode at small signal levels, essentially about 3 and 6 times Is (sufficiently low that the voltage drop across the series resistance Ro can be ignored). The measurements are then substituted into the Shockley equation Is*{exp[(qe/(n*k*T))*(V-I*Rs)]-1} and solved simultaneously for the two data pairs. Mike Tuggle has provided a nice excel spreadsheet to do the math. This spreadsheet, Cal_n_Is.xls forms the basis of my technique and I am thus indebted to both Ben Tongue and Mike Tuggle for making this project feasable.

In addition to simply calculating the main parameters Is and n, I also wished to measure enough points to plot a characteristic curve, and to compare the measured curve with one calculated directly from the Shockley equation. My first spreadsheet then combined Tongue/Tuggle calculation protocol and a graphical view of the match between measured and theoretical. I did make a few methodology modifications which I thought/hoped would allow added accuracy:

1) Using the full Shockly equation without simplifying for an assumed 25dC room temperature. While the 25dC assumption is generally good and probably within the measurement error (or perhaps not), including the actual measured temperature eliminates doubts of inaccuracies due to this parameter. With the power of PC's and spreadsheets there is no reason simplify the equation.
2) Used Vd = 0.04 and 0.05 V (adequately close to Ben Tongue's recommended 0.039 and 0.055 V) whenever possible. Note for some diodes with significantly different forward voltage drops, Si diodes in particular I had to use higher values for Vd.
3) Reporting: I decided that, as there is no unique solution to n and Is, each is dependent on the values Vd and Id, I feel that reporting both Vd and Id is necessary for repeatability. Naturally I also include the ambient temperature in the report as well, and
4) I included a calculation for Ro (or Rg if you are using Tongue's reference) as that is the goal of the exercise.
A screen shot shown below:

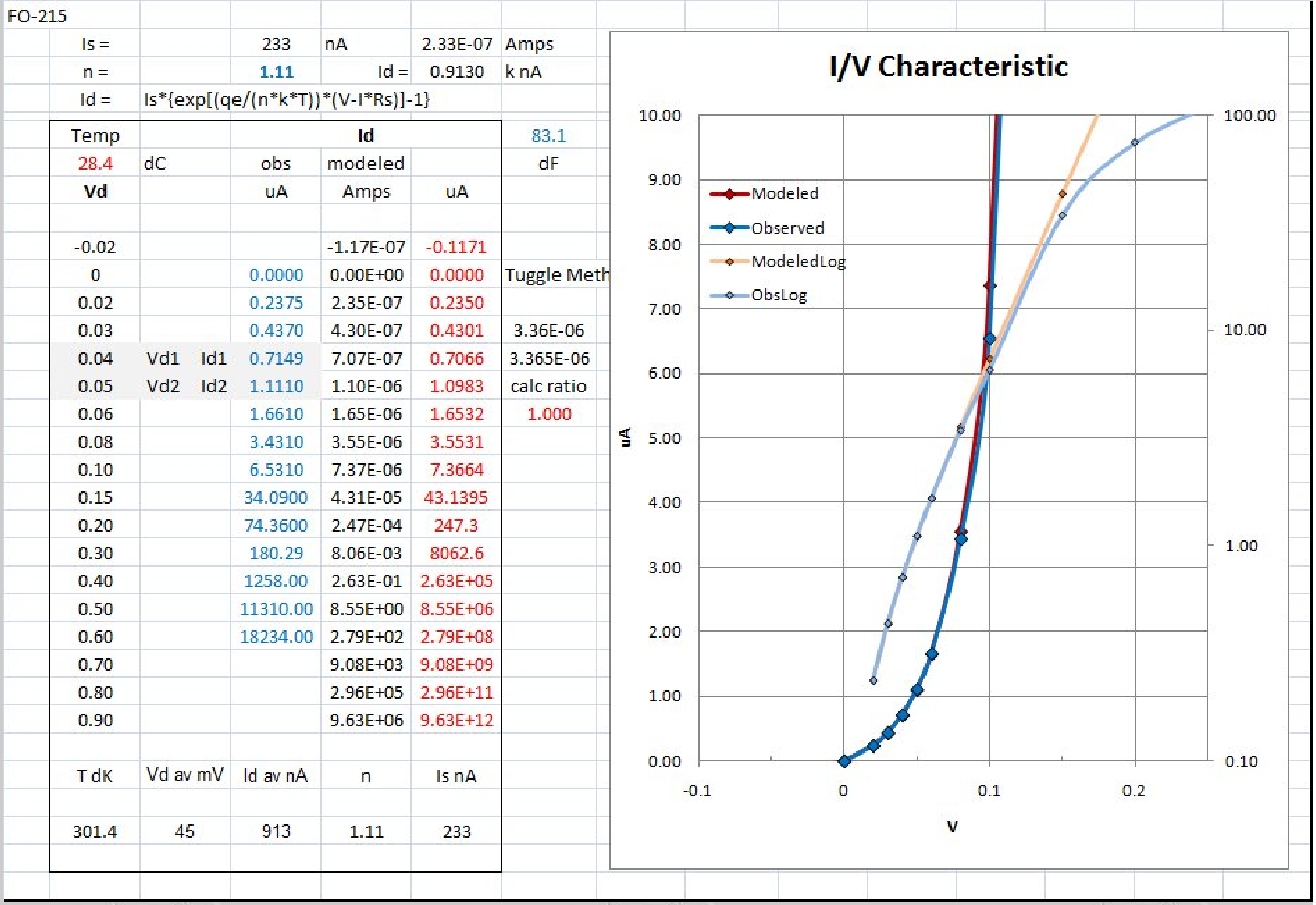

FO-215					
Is =		233	nA	2.33E-07	Amps
n =		1.11	Id =	0.9130	k nA
Id =	Is*{exp[(qe/(n*k*T))*(V-I*Rs)]-1}				
Temp			Id		83.1
28.4	dC	obs	modeled		dF
Vd		uA	Amps	uA	
-0.02			-1.17E-07	-0.1171	
0		0.0000	0.00E+00	0.0000	Tuggle Meth
0.02		0.2375	2.35E-07	0.2350	
0.03		0.4370	4.30E-07	0.4301	3.36E-06
0.04	Vd1 Id1	0.7149	7.07E-07	0.7066	3.365E-06
0.05	Vd2 Id2	1.1110	1.10E-06	1.0983	calc ratio
0.06		1.6610	1.65E-06	1.6532	1.000
0.08		3.4310	3.55E-06	3.5531	
0.10		6.5310	7.37E-06	7.3664	
0.15		34.0900	4.31E-05	43.1395	
0.20		74.3600	2.47E-04	247.3	
0.30		180.29	8.06E-03	8062.6	
0.40		1258.00	2.63E-01	2.63E+05	
0.50		11310.00	8.55E+00	8.55E+06	
0.60		18234.00	2.79E+02	2.79E+08	
0.70			9.08E+03	9.08E+09	
0.80			2.96E+05	2.96E+11	
0.90			9.63E+06	9.63E+12	
T dK	Vd av mV	Id av nA	n	Is nA	
301.4	45	913	1.11	233	

To use the sheet one need only adjust the voltage Vd to that shown in the first column and record the current Id. If only Vd1/Vd2 and Id1/Id2 are measured then the next step is to adjust the value for "n" until the match between the two equations is exact. I even provided a simple ratio calculation to easily test for a match.
While the math is good and results excellent, I quickly found a couple deficiencies in my method,
1) measurements require high precision and its nearly impossible to land EXACTLY at the voltage required. I needed only to get quite close, let the meters stabilize for 2 - 5 minutes, and then record BOTH Vd and Id.
2) the work is rather tedious and, for a large number of diodes it pays dividends to measure only the needed Vd and Id and let the plot aside. My second spreadsheet thus dispenses with the plot. It is with my second spreadsheet that all my data tables and results are posted.
3) better comparisons between diodes can be made if one targets specific Currents rather than Voltages, discussion on this below.

A view of my current data reporting spreadsheet showing ALL input parameters as well as determined values of n, Is, and Ro.
Data input fields in blue: (target Id1 close to 1.0 uA and Id2 close to 0.5 uA).
Calculated fields in red and black.
Adjustment field "n" in green: (adjust value of n until the ratio of the two calculations = exactly 1).

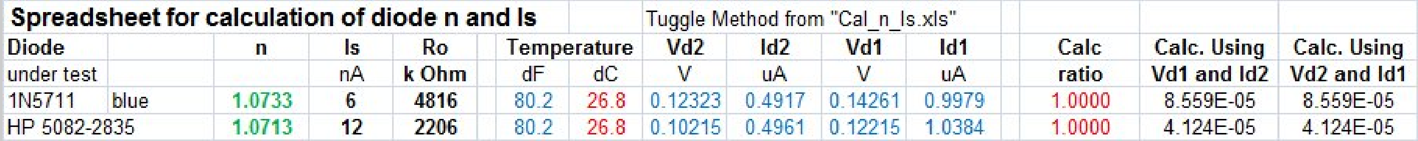

Spreadsheet for calculation of diode n and Is							Tuggle Method from "Cal_n_Is.xls"							
Diode		n	Is	Ro	Temperature		Vd2	Id2	Vd1	Id1		Calc	Calc. Using	Calc. Using
under test			nA	k Ohm	dF	dC	V	uA	V	uA		ratio	Vd1 and Id2	Vd2 and Id1
1N5711	blue	1.0733	6	4816	80.2	26.8	0.12323	0.4917	0.14261	0.9979		1.0000	8.559E-05	8.559E-05
HP 5082-2835		1.0713	12	2206	80.2	26.8	0.10215	0.4961	0.12215	1.0384		1.0000	4.124E-05	4.124E-05

(Note: the engineers in the crowd will have noticed in the above report than I am carrying a degree of precision not justified by the level of accuracy in my measurements. The final results of n, Is, and Ro should be rounded off to not more than three significant figures.)
A copy of my spreadsheet can be downloaded here: Cal_n_Is_Rd.xls

While measuring a good number of diodes, both Germanium, Schottky, Silicon, and even a few LED's, I noted that Tongue's recommended measurements at Vd = 0.039 and 0.055 V resulted in current reports spread out over more than two orders of magnitude. As the determination of Is and n is not unique but a function of Id, I felt uneasy by this method. While most Germanium diodes I have measured have a fairly narrow range of Forward Voltage drops, (Vf), that of Schottky's can range up to a tenth of a volt. This is the factor responsible for the huge range of measured Id values, see the following plot of I/V characteristics for some selected diodes:

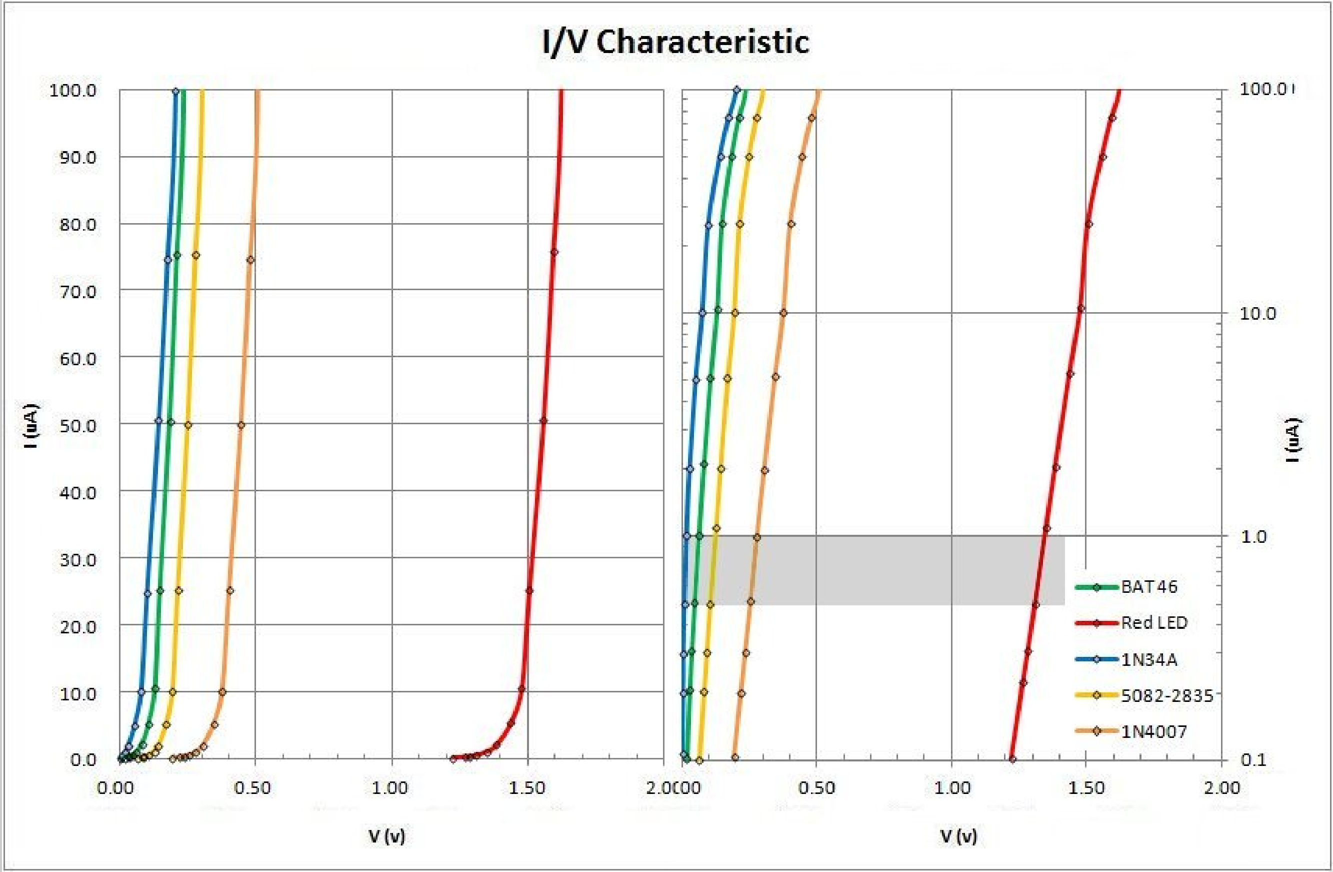

In order to get what I feel is a better comparison between diodes, I have decided to target not some pre-determined voltage, but rather target currents of Id = 0.5 and 1.0 uA. On the above plot on the right (with Log Id vs Vd scale) it will be clear that I am aiming at the same part of the characteristic curve regardless of Vf. Hopefully this will allow good comparison between diodes with rather different Vf, even as far as including Silicon and LED's in my mix. Still, on study of the plot above, I must note that the Germanium characteristic at 0.5 to 1.0 uA appears to be curving towards its zero-crossing. Perhaps targeting an amp range between 3.0 and 5.0 uA would have been better, still thinking here, on va voir....

Shockley equation Id = Is*{exp[(qe/(n*k*T))*(V-I*Rs)]-1}
where:
n = ideality factor
Is = Saturation current in Amps
Id = Diode Current in Amps
Vd = Diode Voltage
k = boltzmann = 1.38E-23 J/K
T = temp K = 300
K = dC + 273.15 Kelvin
qe = electron charge = 1.609E-19 cmb

Often simplified to:
Id = Is*(exp(Vd/(0.0256789*n))-1)

Diode Resistance Ro = VT * n / Is
Where:

T = 300K
VT = k*T/qe = 0.0256789
SO: Ro = k * T / qe * n / Is

Ebers-Moll equation Vf = m (kTq) ln [(If / Is) +1]

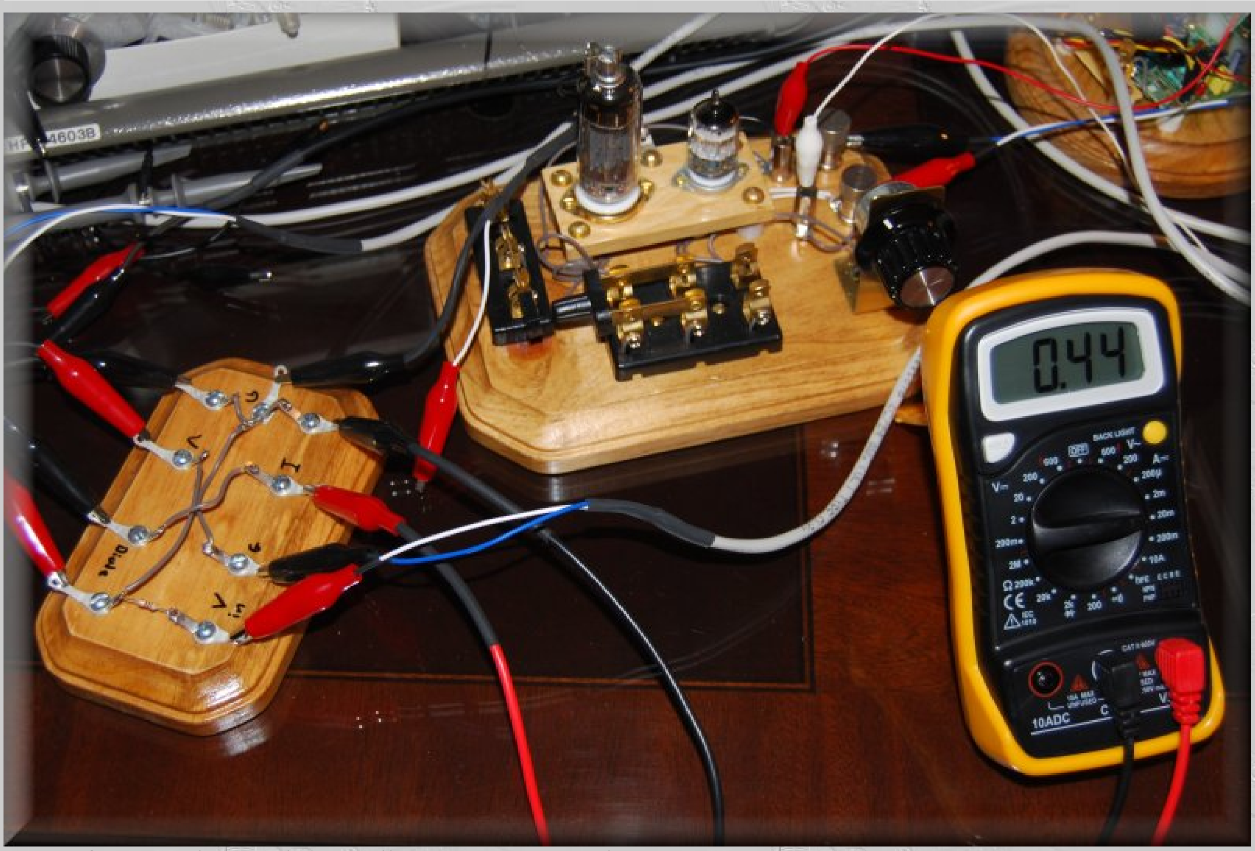

Photo of the lab with my diode test jig, amp meter, and diode vacuum tubes under test.

Return to Lab..

Back to Main....

Friday, 05-Jan-2018 15:30:42 CST

Comments and/or Correspondance to: kj-smith@sbcglobal.net

.::. Alsace .::. Life List .::. Ionosphere .::. Crystal Radio .::. Snapshots .::. Thoughts .::.
.::. Sitemap .::. Beer Labels .::. International Pix .::. Western US Pix .::. Pinhole Pix .::. Petrology .::.

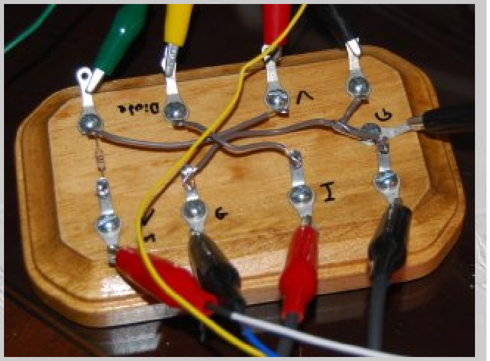

My radio "laboratory"

Back to Radio..

Having now acquired several crystal radios by purchase (two) or building my own (four), I begin to wonder how they stack up against each other and against other radios I read about and dream of on the net. It can be fairly common to hear radio performance described in qualitative terms, "high Q", "sharp tuning", "highly sensitive", etc. Without measured numbers to back them up, it is extremely difficult to judge the subjective notes of different radio users. It is inevitable I believe that in this hobby that one ends up getting together as much test equipment as possible, (read "affordable") in order to know more objectively about the radios one builds and/or uses. This page is the tale of my own effort to equip a small "laboratory" of sorts for just such purpose.

My first needs were related to my building projects and included meters for measuring capacitance and inductance. I have excellent-quality bench meters for other hobby projects including a Keithley 192 and 181 voltmeter with IEEE 488 computer interfaces. These are voltmeters only though, nothing else. My first purchase then was a capacitance/inductance meter. It worked great for capacitance, but the lowest range for inductance, 20mH unfortunately is useless. So, this meter was quickly followed by a stand-alone inductance meter ranging down to 200uH. I use these frequently when making or measuring coils, and measuring capacitors purchased on ebay, better to design a radio around a specific componant.

My next needs follow logically, what about the radios I build?, how do they do? For the non-electrical engineer, testing radios is a bit of a daunting task and this page is by no means intended to be a tutorial. I do give my experiences so that perhaps someone with the same idea may benefit and possibly avoid the mistakes I have certainly made. I describe my equipment and the protocols used in my testing, but I STRONGLY refer the reader to the following two pages for technical guidance and explanations, most of what I do came directly from here:
Dick Kleijer crystal-radio.eu http://www.crystal-radio.eu/enqmeting.htm
Gollum愀 Crystal Receiver World http://home.snafu.de/wumpus/gollum/testing.htm article by C.A. Lauter
I heartily express my appreciation to our European friends for making available on the web these excellent technical tutorials for crystal radio testing.

Testing and understanding your set from the Antenna at the front, the tank in the middle and on to the Diode at the back and then the Radio complete, the whole shebang!

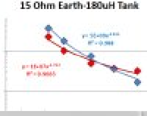

ANTENNA TUNING UNIT.. Design notes and Models, Effect of earth resistivity..

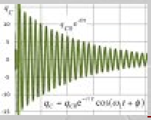

COIL Q.. An Analytical Approach to the measurement of coil Q.

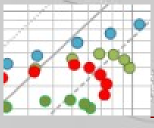

COIL Q Revisited.. Web research and personal measurements: Setting expectations for Q.

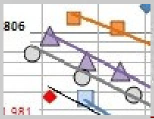

Crystal Radio Capacitors.. A look at capacitor Q and its relation to ESR (Equivelant Series Resistance).

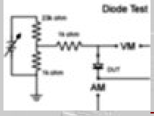

Diode I: Setup and Protocol.. Describes a useful protocol for determination of Diode Is, n, and Ro.

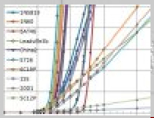

Diode II: Test Data Measurements .. Reports the measurements made under the above protocols for a large variety of diodes

Diode III: Vacuum Diodes.. Measurements and discussion on vacuum diode utility in crystal radio.

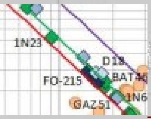

Diode IV: Discussion and Results.. A discussion of diode I/V characteristics, Ro and matching, and sundry ravings.

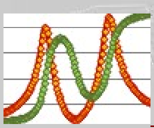

RESONANCE.. A look at some theory and resonance calculations for MW sets.

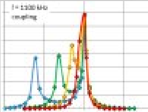

RADIO TEST.. Describes protocols and results of radio performance testing on my sets.

Conclusion

So, for an investment in a few key pieces of good equipment, yes at some non-trivial cost, you can appreciate the hobby so much more. I have an oscilloscope as well, but so far have not explored its value. I prefer the recording and spreadsheet methods as they give a better record of the results and allow graphing and easy comparison. Knowing how your sets perform, being able to test your diodes, knowing your coil inductance are all important aspects of knowing your radios. I certainly recommend this!
Ciao!

Return..

Back to Main....

Friday, 05-Jan-2018 15:24:42 CST

Comments and/or Correspondance to: kj-smith@sbcglobal.net

.::. Alsace .::. Life List .::. Ionosphere .::. Crystal Radio .::. Snapshots .::. Thoughts .::.
.::. Sitemap .::. Beer Labels .::. International Pix .::. Western US Pix .::. Pinhole Pix .::. Petrology .::.

Part III: Vacuum Diode Test Data Report

Back to Lab..

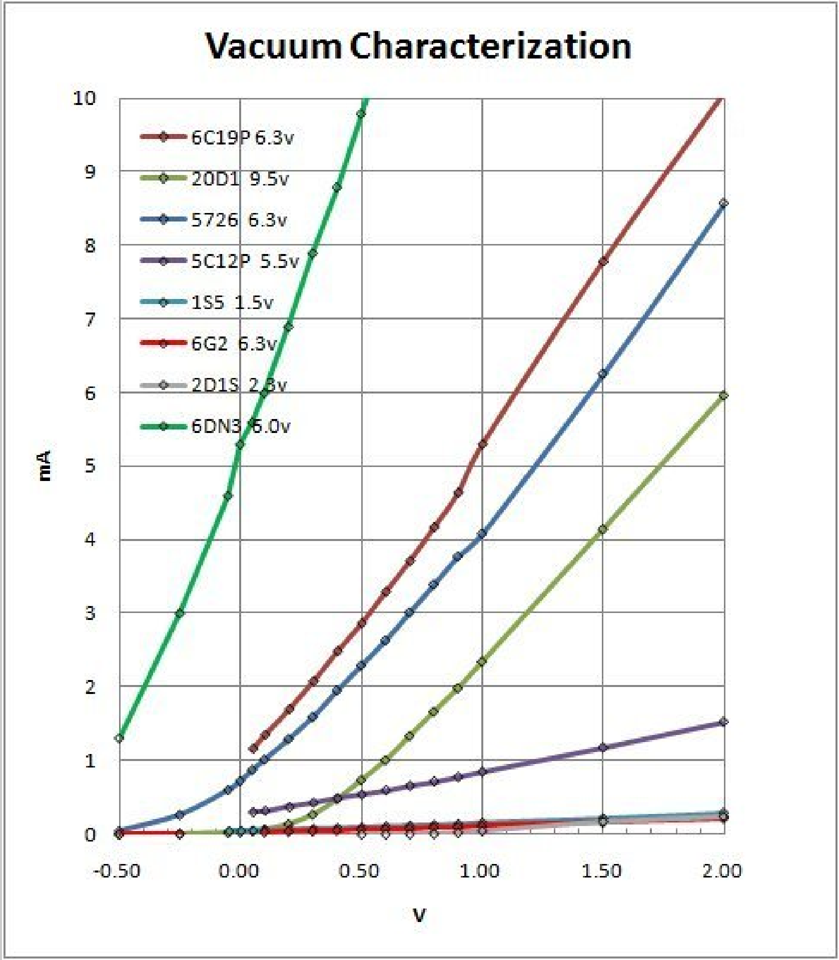

When I decided early on to construct a radio based on the vacuum diode (see my Fleming Radio section), I had to find a suitable tube for the project. That began with an extensive search through the datasheets checking vacuum diode properties and reviewing the characteristic curves for a large number of diode tubes. In the above figure I plot the characteristics as best I can on a common plot for comparison. I was hoping to find a good candidate with sensitive characteristics and low energy consumption. As you will find, such a beast did not exist. Following this research began a period ot purchase and testing.

I have tested seven different tube types including 6.3V rectifier diodes that take a lot of power to run and are not really suitable for battery use, a 9v dual diode tube (20D1), a pentode/diode tube (1S5) that is designed to run on a battery at 1.5V and 50mA, and a couple miscellaneous but cute tubes. In testing the 1S5, I found the filament never glowed incandescent at 1.5V and, on measuring, was barely sensitive to anything. I cannot imagine this tube would make much of a diode for crystal radio use. I tested different manufacture 1S5 tubes from two different suppliers. No dice. Finally, I tried to push the tube to operate at higher-than-specified voltages. At 4.3V the I-V curve was still very flat, but shifted slightly higher, crossing the Amp axis at about 0.4 mA, (see graph). At 5.5V one of my tubes gave up the ghost and I didnt do more. I pretty much rule the 1S5 out for my crystal radio work. Looks like i'll be needing the power supply when I get around to building/operating that Fleming Radio.

A question remains as to what in fact is different about these tubes to give such different characteristics. While the I/V characteristic of most vacuum diodes pretty much follow the Langmuir-Child Law, the steepness of the I/V characteristic is largely due to the geometry of the tube elements and the volume of electron space-charge between them. JB Calvert's Theory of Vacuum Tubes informs one that this is the property called Purveance in a vacuum diode. Measuring this requires plotting I raised to the 2/3 power against V and taking the slope of the best-fit line through the data. This slope raised to the 3/2 power is the perveance. The following chart illustrates graphically the relation between the tube characteristic and perveance. For the 5726 tube I plot the characteristic in dark blue for the tube heated with the design 6.3 volts. In light blue I have plotted the measurements and raised them to the 2/3 power. One sees immediately that the new curve is linear. The slope of the curve raised to the 3/2 power yields the perveance factor for the tube.
(Note in addition that when running the tube at a more sensitive point with 4 volts only on the heater the perveance is slightly less than spec.).

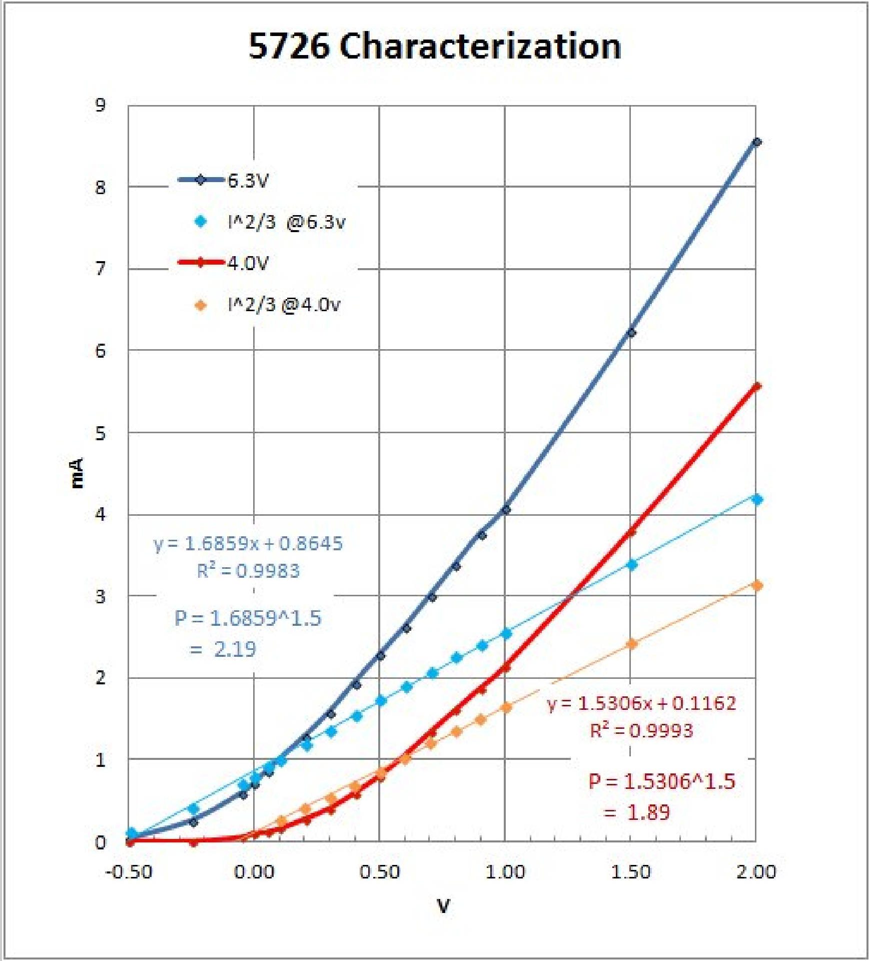

With my measurements calculating the perveance is practical and quickly done, results as follows:

Tube	Perveance		R2	
6C19P	2.54		0.997	
5726	2.19		0.998	
6DN3	5.40		0.997	
20D1	2.02	LC?	0.985	(may not follow Langmuir-Child)
5C12P	0.32		0.999	
1S5	0.063		0.999	
2D1S	0.155	LC?	0.975	(may not follow Langmuir-Child)
6G2	0.050		0.999	

The perveance numbers above range from 0.02 to over 2 mA/V^3/2 (with >2 being good for small-signal detection) and clearly shows my choice of the 6C19P to be an excellent one. I am surprised to see a huge 5.4 for the 6DN3 diode, a color television damping diode. It takes a Novar 9-pin socket and is a large tube so doing more with this tube may have to wait a bit.

Calvert's 2001 published data

Tube	Perveance
6AL5	2.42
6H6	0.50
7Y4	0.58
2X2A	0.017
6V3-A	2.3
6AX4GT	1.42
6AV6	0.085
1A3	0.075

The high-perveance diode tube types I tested show an interesting property in that their characteristic curves do not pass through the origin of the graph. These curves are wholly acceptable as crystal radio detectors although they will require power to operate. What I have noticed with such tubes is that when I turn them off after using them with a radio, on cooling they go through a period of much increased sensitivity (very loud) before fading to nothing. It is as though running them at the full 6.3V lowers their full potential as crystal radio rectifiers. The reason is in the high perveance of the tubes, the anode

is already proximate to the space charge before any plate voltage is applied. Effectively, while the plate (anode) "Zero" voltage is measured with respect to ground, the plate itself is still positive with respect to the cathodic space charge. Small currents will continue to flow even with a negative bias to the plate, (Contact potential).

In order to explore this idea, I ran a series of tests with the tubes running at lower operating voltages (effectively diminishing the space charge thus lowering the perveance) as shown:

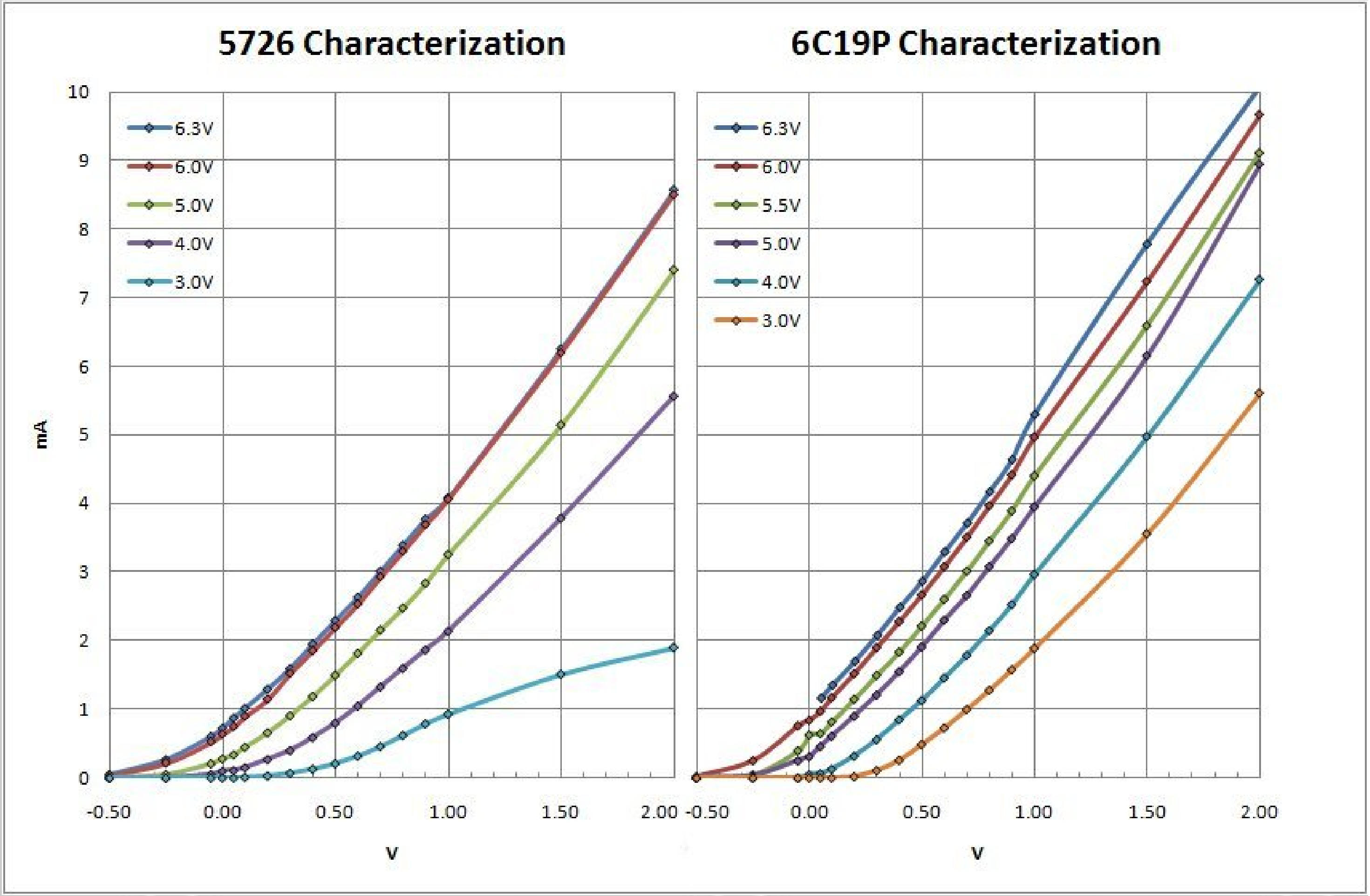

Here you see the impact of lower operating voltages. With the tube operating between 3 and 4 volts the characteristic curves begin to pass through the origin as in regular solid-state diodes and crystals. It is precisely in this voltage range that I found the radio sensitivity, as measured by loudness to my ear, to be most pronounced. The diode characteristic needs to pass through the origin of the I-V graph for rectification at highest sensitivity. This takes me to the 20D1 tube which operates at 9v but only 200mA. This will still take a power supply to use, but the good thing is that the characteristic curve passes through the origin. Running at its design parameters the tube looks a lot like the 6C19P at 4v.

A conceptual look at the impact of heater temperature on the space charge distribution in the annulus between cathode and anode. I start by drawing a qualitative plot of charge density versus distance from the cathode. This is not how such plots are done in the literature but it is a nice visualization. The distribution of charge is a statistical function of heater temperature and plate bias. For the zero-bias case I show qualitative distributions for three heater voltages, 3.0, 4.0, and 6.0 volts. On my chart of current/voltage I find no current flows with 3.0 volts on the heater. Current just begins to be measurable at 4.0V and at 6.0V I found a significant 0.83mA of current indicating that a good portion of the space charge is not impinging on the plate even with zero bias.
Biasing the plate positive will have the effect of "pulling" the space charge towards the plate increasing the current while a negative plate bias will have the opposite effect. With 6.0V heater it will take nearly a negative half volt before the current stops. This is death to sensitivity in a crystal set. Operating the heater at 4.0V provides the most efficient rectification.
Note also that high perveance is not merely a matter of placing the plate as close as possible to the cathode. This may cause an excessive zero-bias current to overcome. The surface-area of both cathode and plate are also critical with large areas providing high current transfer. Vacuum tubes such as pentodes with an included diode plate are poor rectifiers having very low perveance. This is due to the small plate stuck in with all the other hardware in the tube. Get a "fit for purpose" diode for best results.

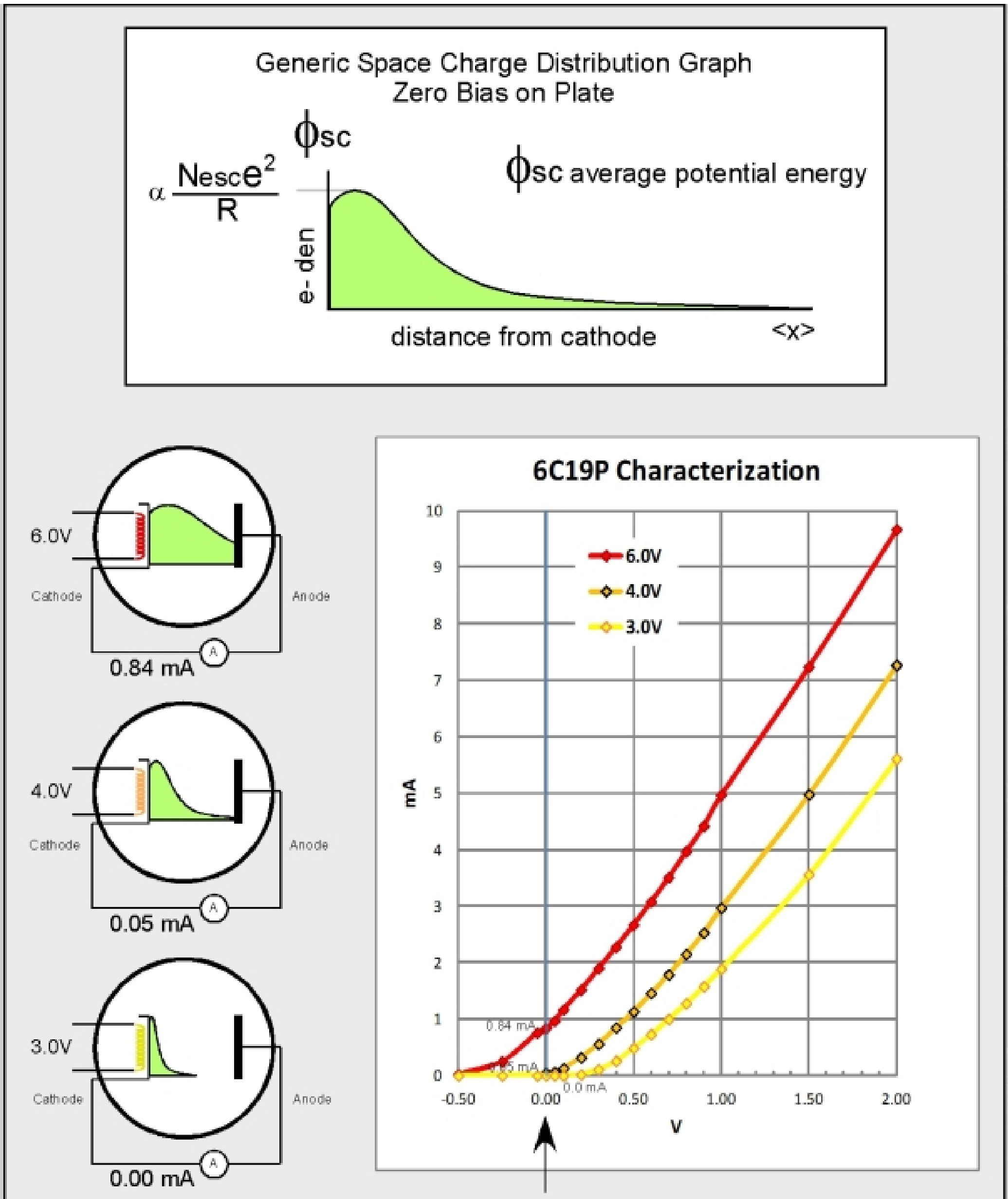
Generic Space Charge Distribution Graph
Zero Bias on Plate
ϕsc
α Nesce²/R
ϕsc average potential energy
e- den
distance from cathode
<x>
6.0V
Cathode
Anode
0.84 mA
A
4.0V
Cathode
Anode
0.05 mA
A
3.0V
Cathode
Anode
0.00 mA
A
6C19P Characterization
6.0V
4.0V
3.0V
mA
V
10
9
8
7
6
5
4
3
2
1
0
-0.50
0.00
0.50
1.00
1.50
2.00
0.84 mA
0.05 mA
0.0 mA

Current across diode at ZERO applied voltage

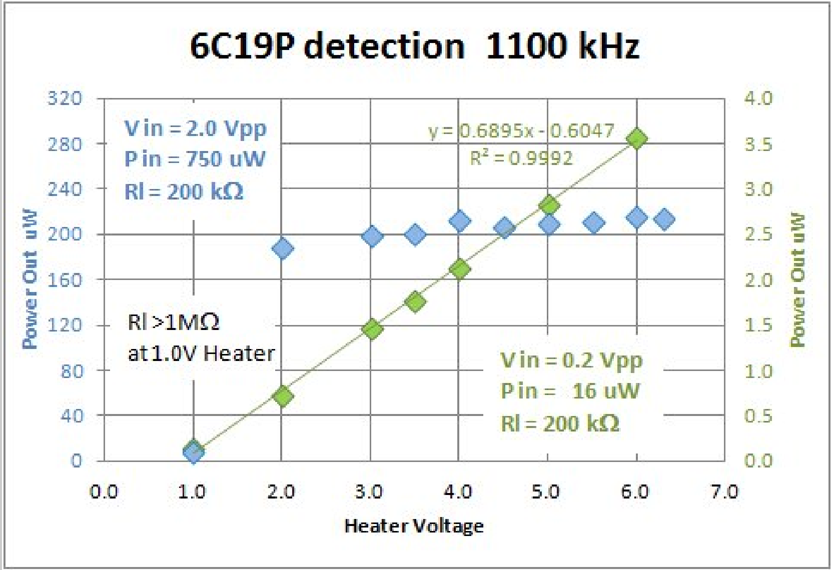

Note that while the above model as well as hearing (a kind of "hearing is beleiving" thing) supports the 4.0V heater voltage as a sweet spot for highest rectification efficiency, I have yet to be able to measure this. Using my tube-driven Telefunken set (see description on my main radio page) I made a series of output signal power measurements at various tube heater voltages. In this I found no relation between matched output power and the tube heater voltage at 20mV input (2.0Vpp on signal generator), and a monotonic increase in output power at 6mV input (0.2Vpp on the generator). I was seeking a peak output power at around 4.0V on the heater so this result is disturbing. I can hear it, but cannot measure it.

Portraits:

Return..

Back to Main....

Friday, 05-Jan-2018 15:30:54 CST

Comments and/or Correspondance to: kj-smith@sbcglobal.net

.::. Alsace .::. Life List .::. Ionosphere .::. Crystal Radio .::. Snapshots .::. Thoughts .::.
.::. Sitemap .::. Beer Labels .::. International Pix .::. Western US Pix .::. Pinhole Pix .::. Petrology .::.

US008045314B2

(12) **United States Patent**
Ibok

(10) **Patent No.: US 8,045,314 B2**
(45) **Date of Patent: Oct. 25, 2011**

(54) **METHOD OF ATMOSPHERIC DISCHARGE ENERGY CONVERSION, STORAGE AND DISTRIBUTION**

(75) Inventor: **Effiong Etukudo Ibok**, Sunnyvale, CA (US)

(73) Assignee: **The Travis Business Group, Inc.**, Sunnyvale, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **12/534,115**

(22) Filed: **Aug. 1, 2009**

(65) **Prior Publication Data**

US 2010/0220424 A1 Sep. 2, 2010

(51) **Int. Cl.**

H01T 23/00	(2006.01)
H02H 1/04	(2006.01)
H02H 1/00	(2006.01)

(52) **U.S. Cl.** **361/231**; 361/117; 361/126; 361/127; 361/128; 361/233

(58) **Field of Classification Search** 361/117, 361/126–128, 231, 233
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

911,260	A *	2/1909	Pennock	307/149
1,869,661	A *	8/1932	Bumbaugh	174/2
5,367,245	A *	11/1994	Mims	322/2 R
6,038,117	A *	3/2000	Dullni et al.	361/58
6,813,137	B2 *	11/2004	Matsuoka et al.	361/305
7,033,406	B2 *	4/2006	Weir et al.	29/623.5
2002/0101700	A1 *	8/2002	Betts	361/230
2006/0001068	A1 *	1/2006	Mosley et al.	257/306
2009/0109712	A1 *	4/2009	Calloway	363/32
2010/0007321	A1 *	1/2010	Ogram	323/299

FOREIGN PATENT DOCUMENTS

DE	4205521	A1 *	4/1993
JP	04208637	A *	7/1992
JP	08330093	A *	12/1996

OTHER PUBLICATIONS

English language abstract of Kobayashi (JP 04-208637).*
English language machine translation of Kimura (JP 08-330093).*
English language machine translation of Pengel (DE 4205521).*

* cited by examiner

Primary Examiner — Jared Fureman
Assistant Examiner — Scott Bauer
(74) *Attorney, Agent, or Firm* — Blakely, Sokoloff, Taylor & Zafman, LL

(57) **ABSTRACT**

A method of converting atmospheric electrical discharge to a useable form of energy by arresting, storage and retransmission of lightning induced electrical discharge is disclosed. The invention discloses methods of deploying this technology even in isolated locations where no electricity infrastructure exists. Additionally, the potential for achieving in excess of 1 GWe of electrical power supply at costs orders of magnitude lower than fossil fuel or solar is also disclosed. Isolated collection units are disclosed. A method of deployment of these devices on individual cars and recharge stations is disclosed which in effect unplugs these automotive and recharge stations from the grid. This capability significantly decreases their dependence on the electricity grid infrastructure, enabling a much more enhanced rollout capability for the industry. The concept of Energy Dams is disclosed. These are facilities with substantial electrical energy storage capacities with the capability of receiving energy feeds from various generation sources.

18 Claims, 21 Drawing Sheets

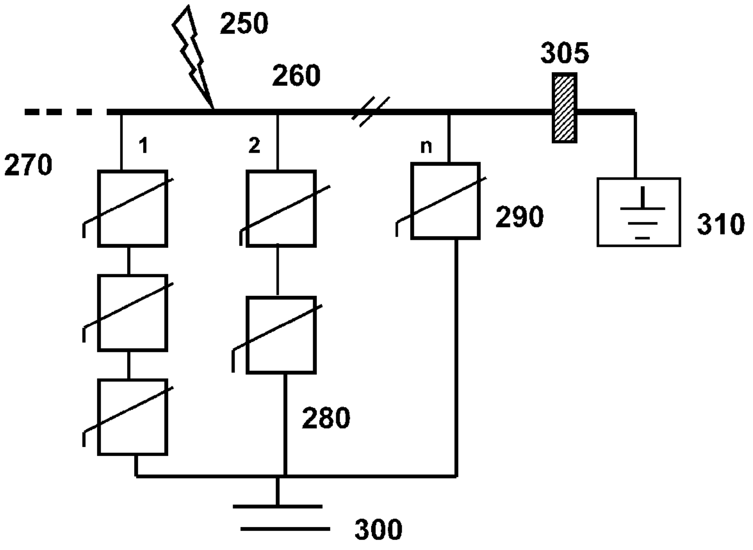

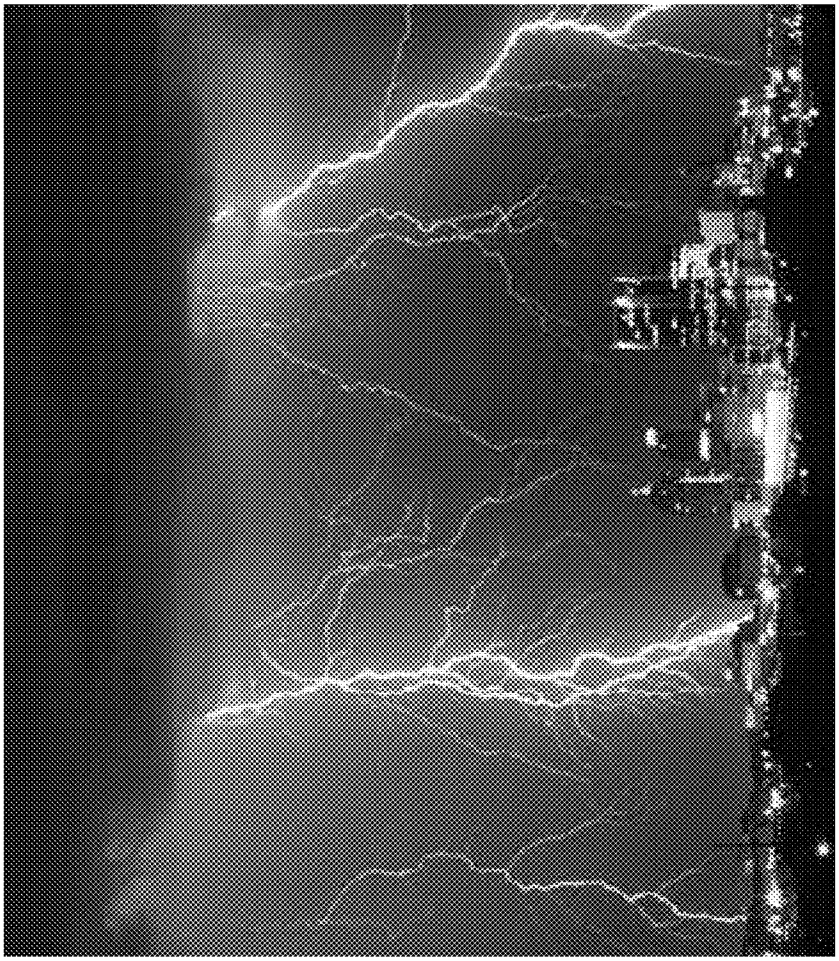

Fig. 1a

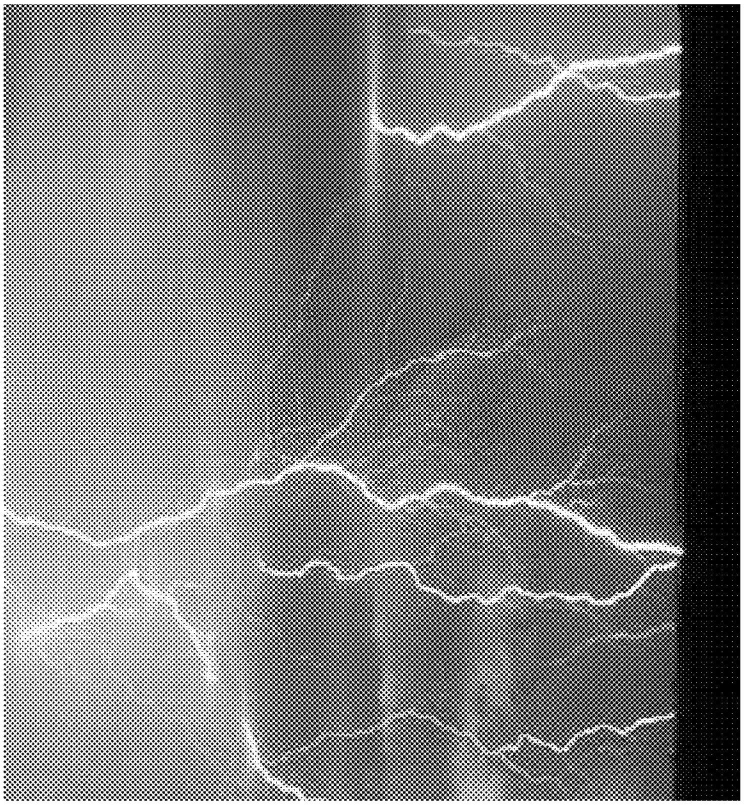

Fig. 1b

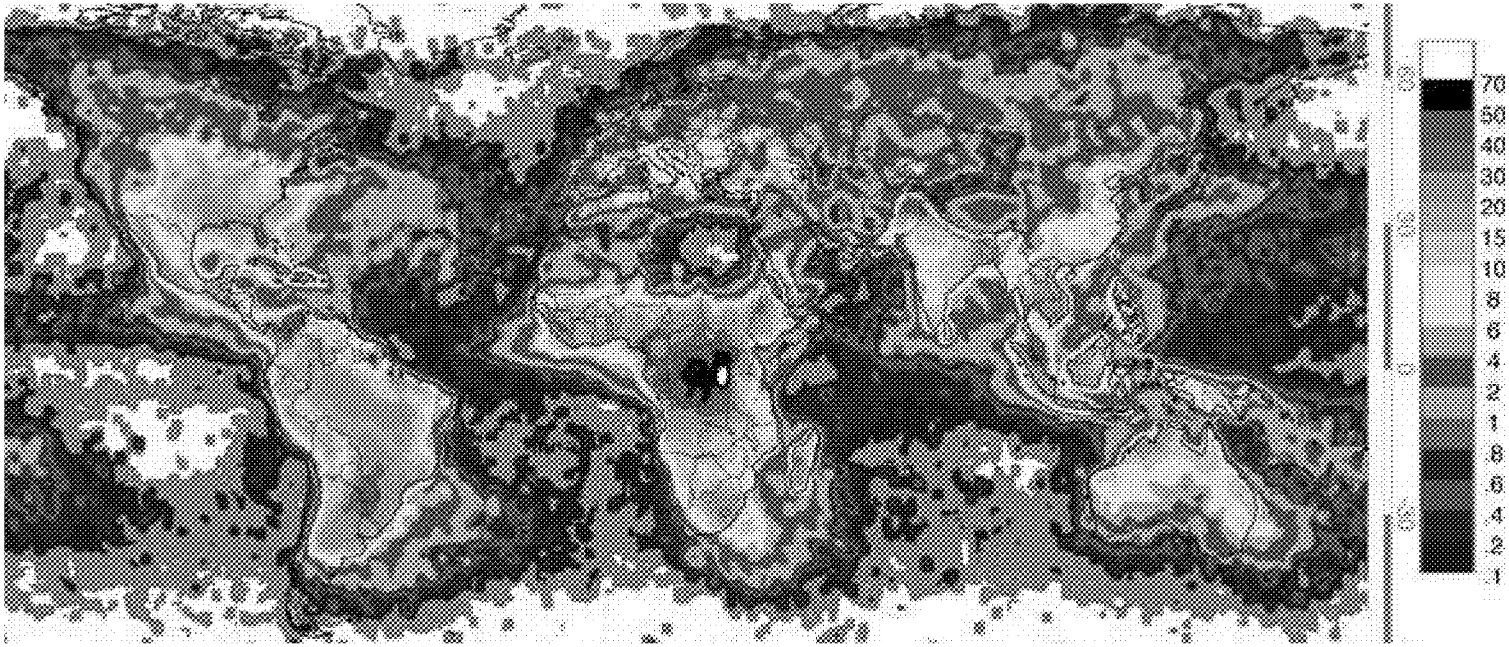
60
30
0
-30
70
50
40
30
20
15
10
8
6
4
2
1
.8
.6
.4
.2
.1

Global Distribution of Lightning Strikes: April 1995 – February, 2003. (Source: NASA)

Fig. 1c

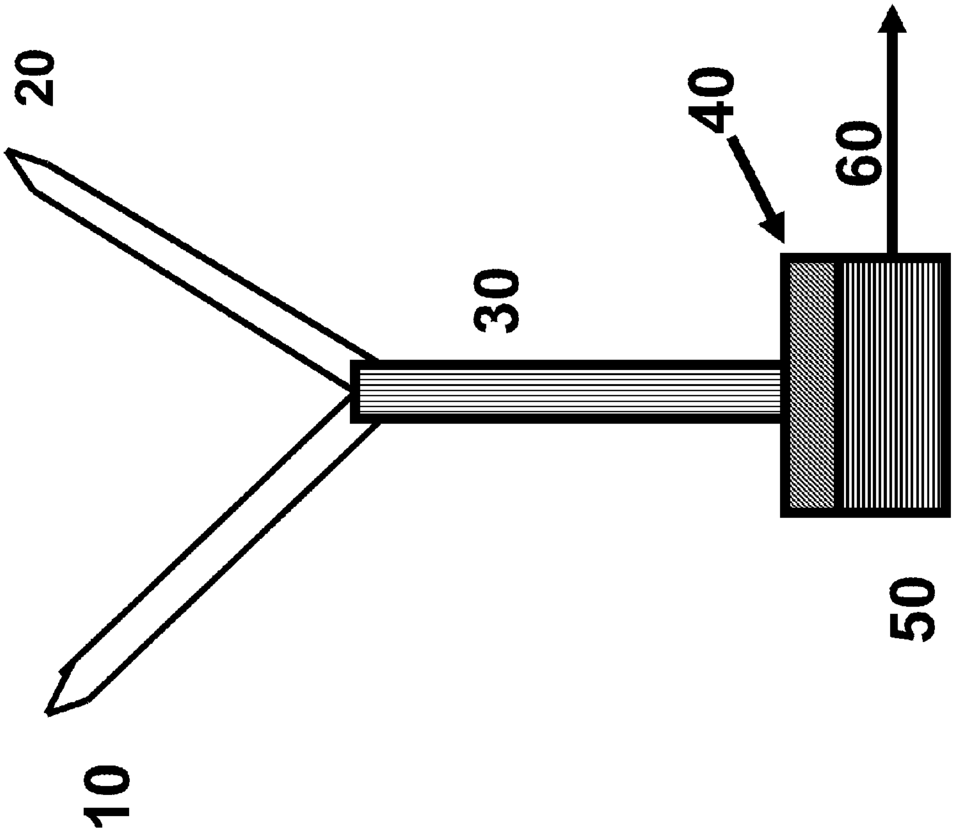
10
20
30
40
50
60

Fig. 2

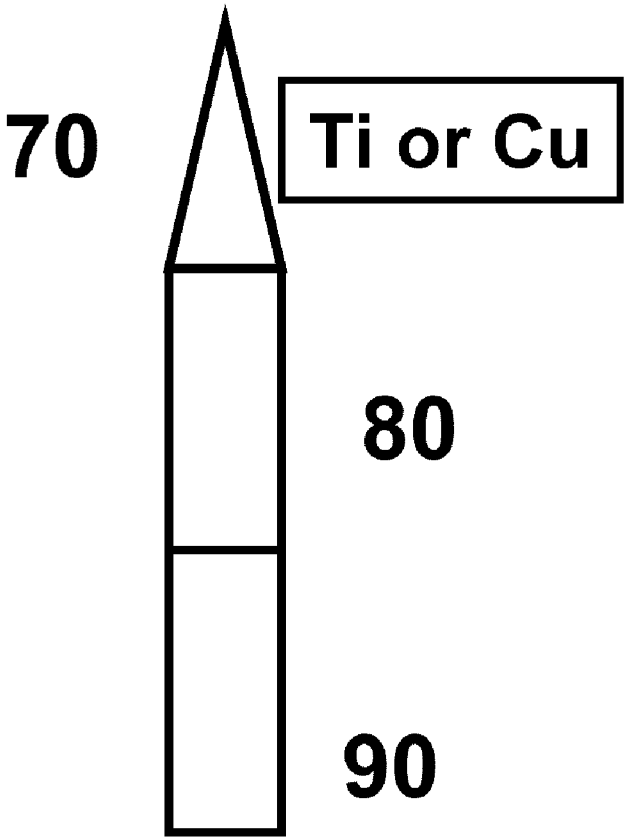
70
Ti or Cu
80
90

Fig. 3

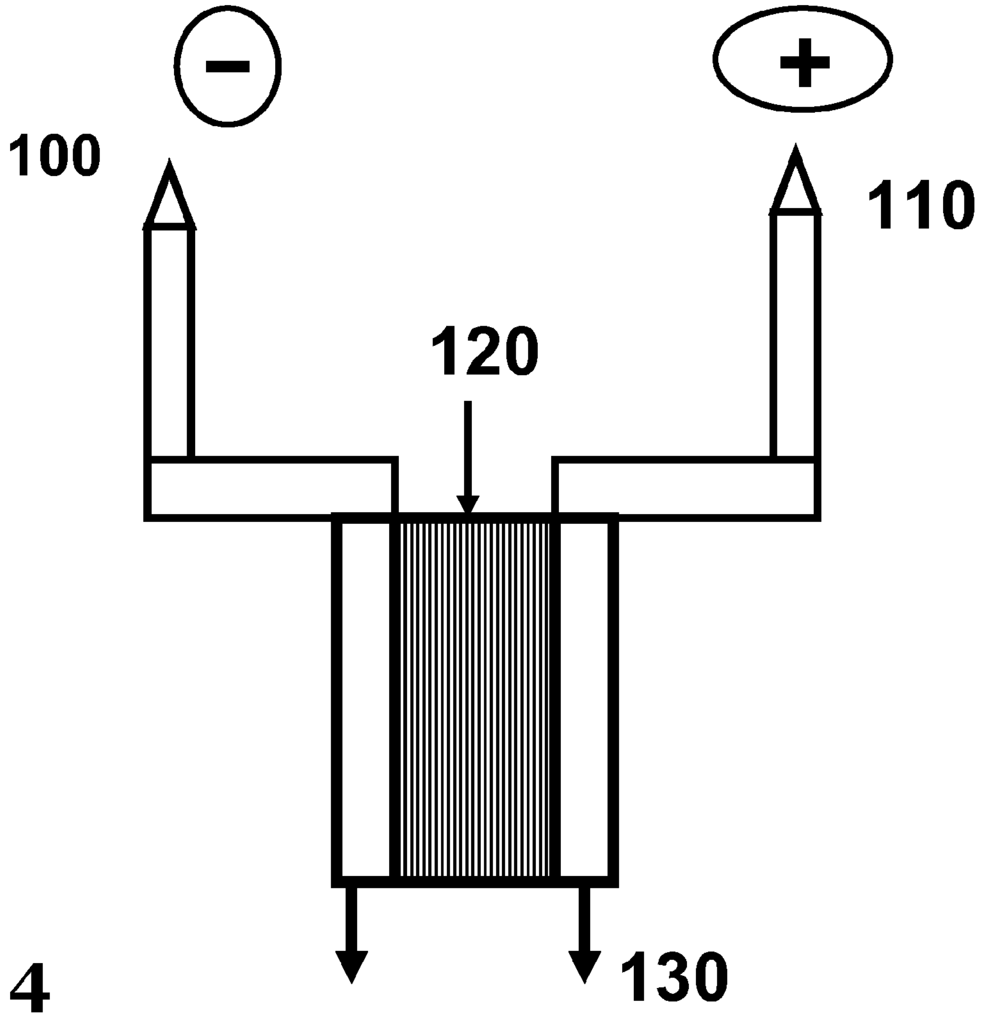
−
+
100
110
120
130

Fig. 4

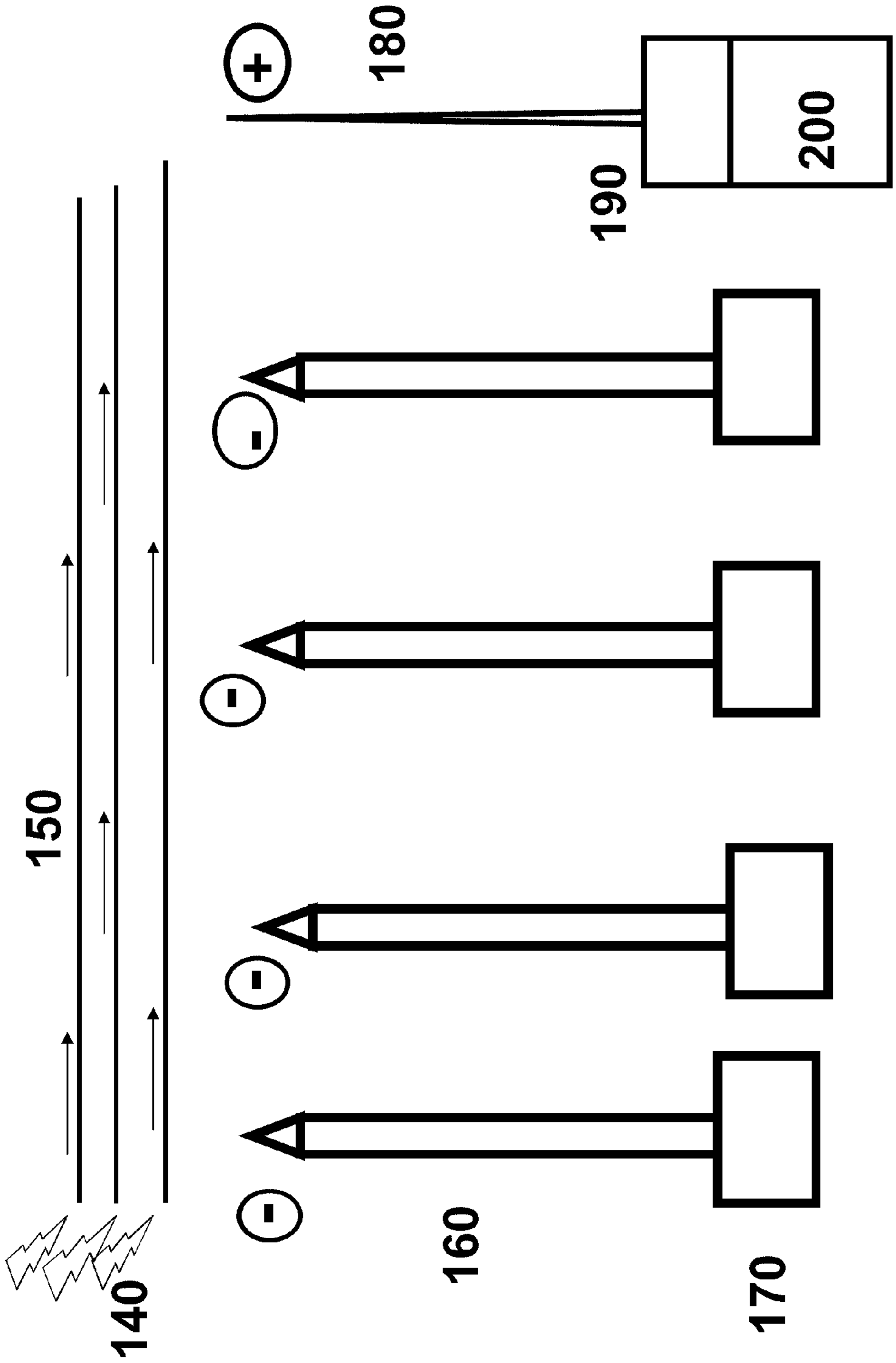
180
+
200
190
150
-
-
-
-
160
170
140

Fig. 5

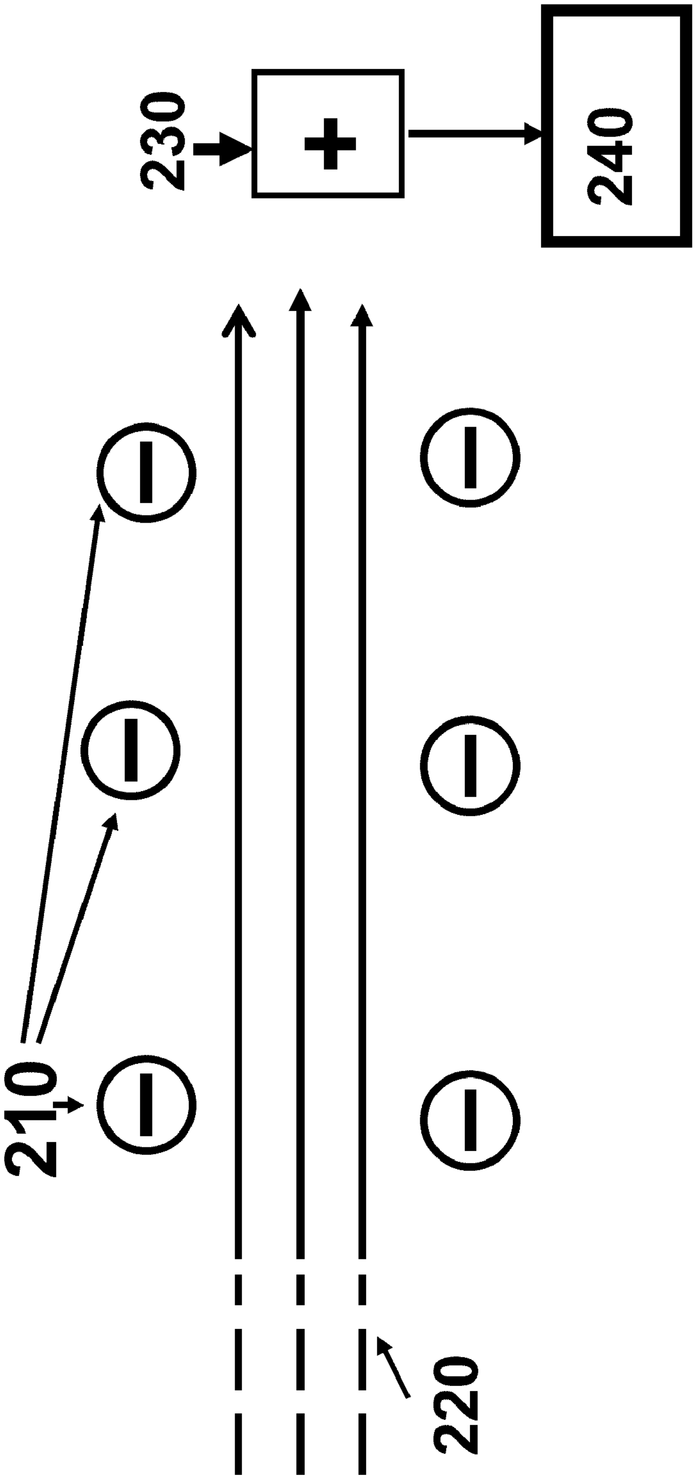
210
220
230
+
240

Fig. 6

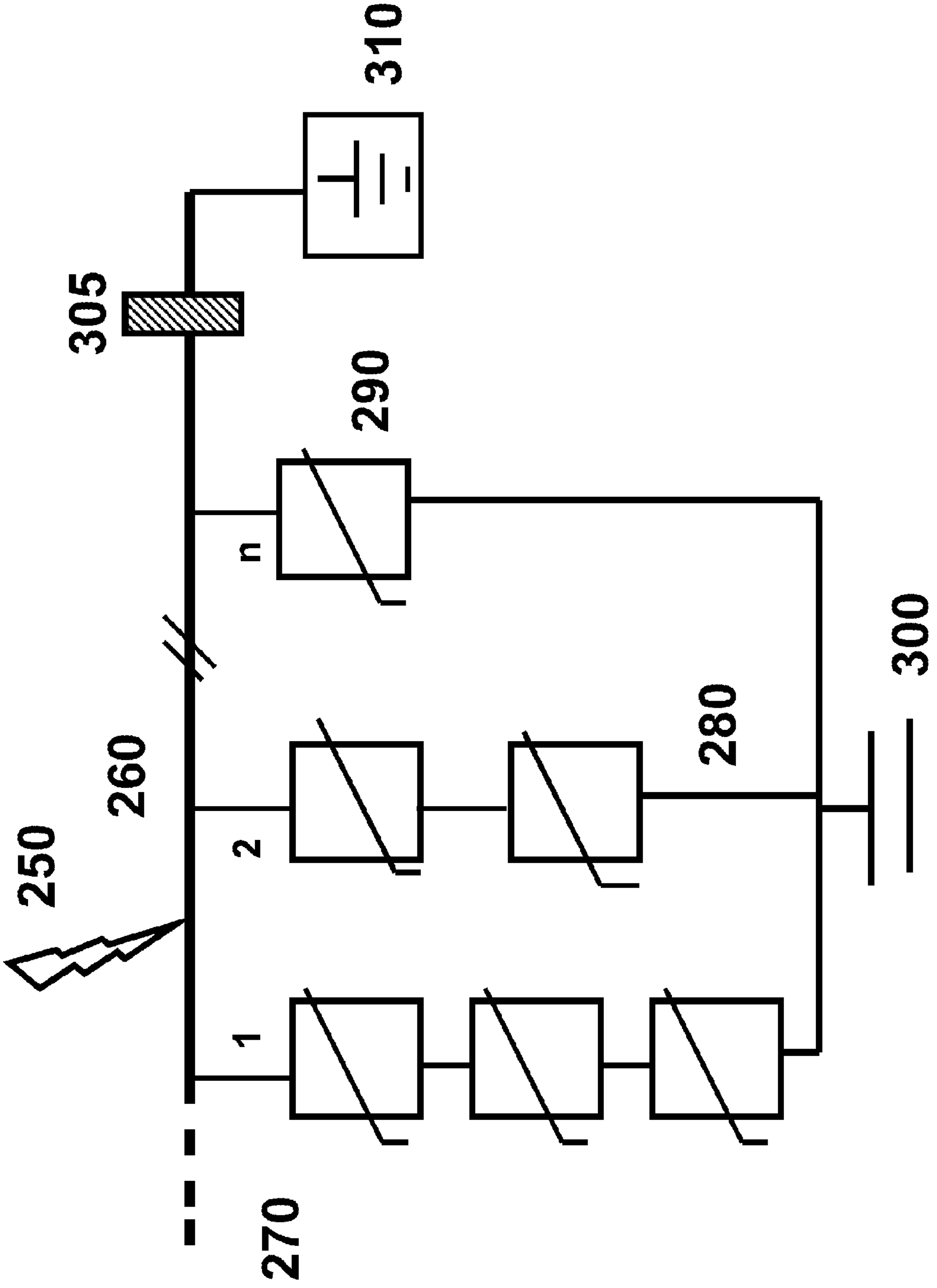
310
305
290
n
300
280
260
2
250
1
270

Fig. 7a

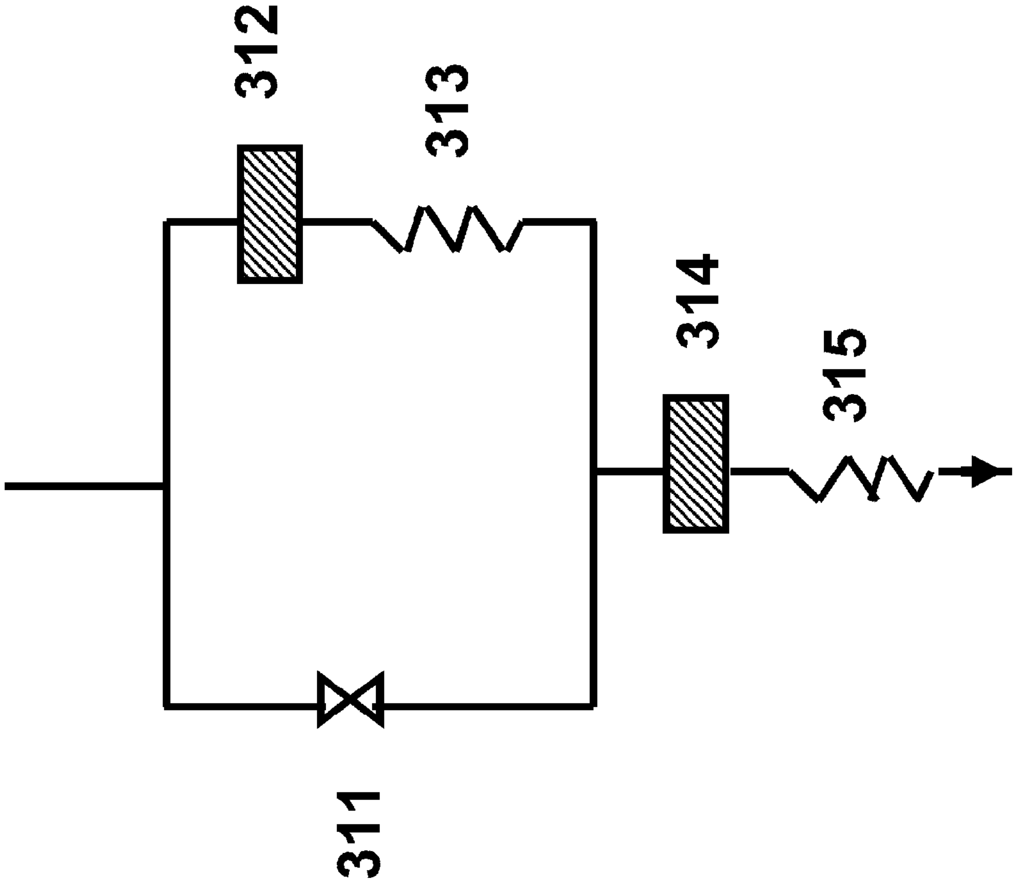
312
313
314
315
311

Fig. 7b

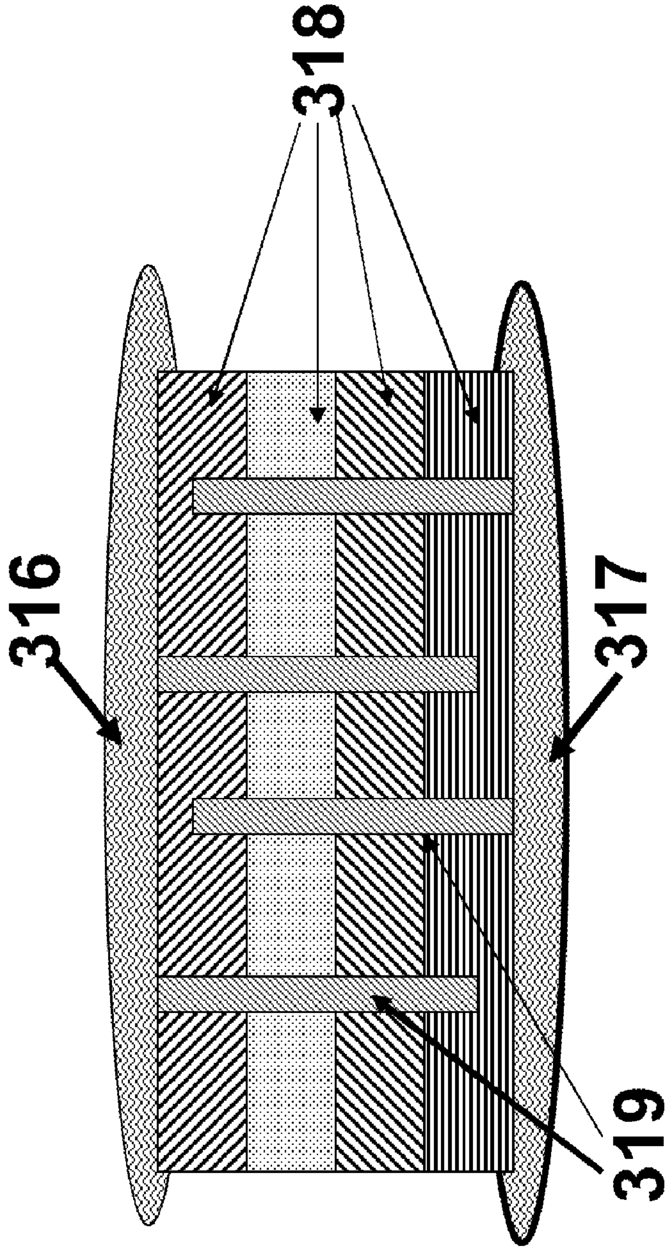
318
316
317
319

Fig 7c

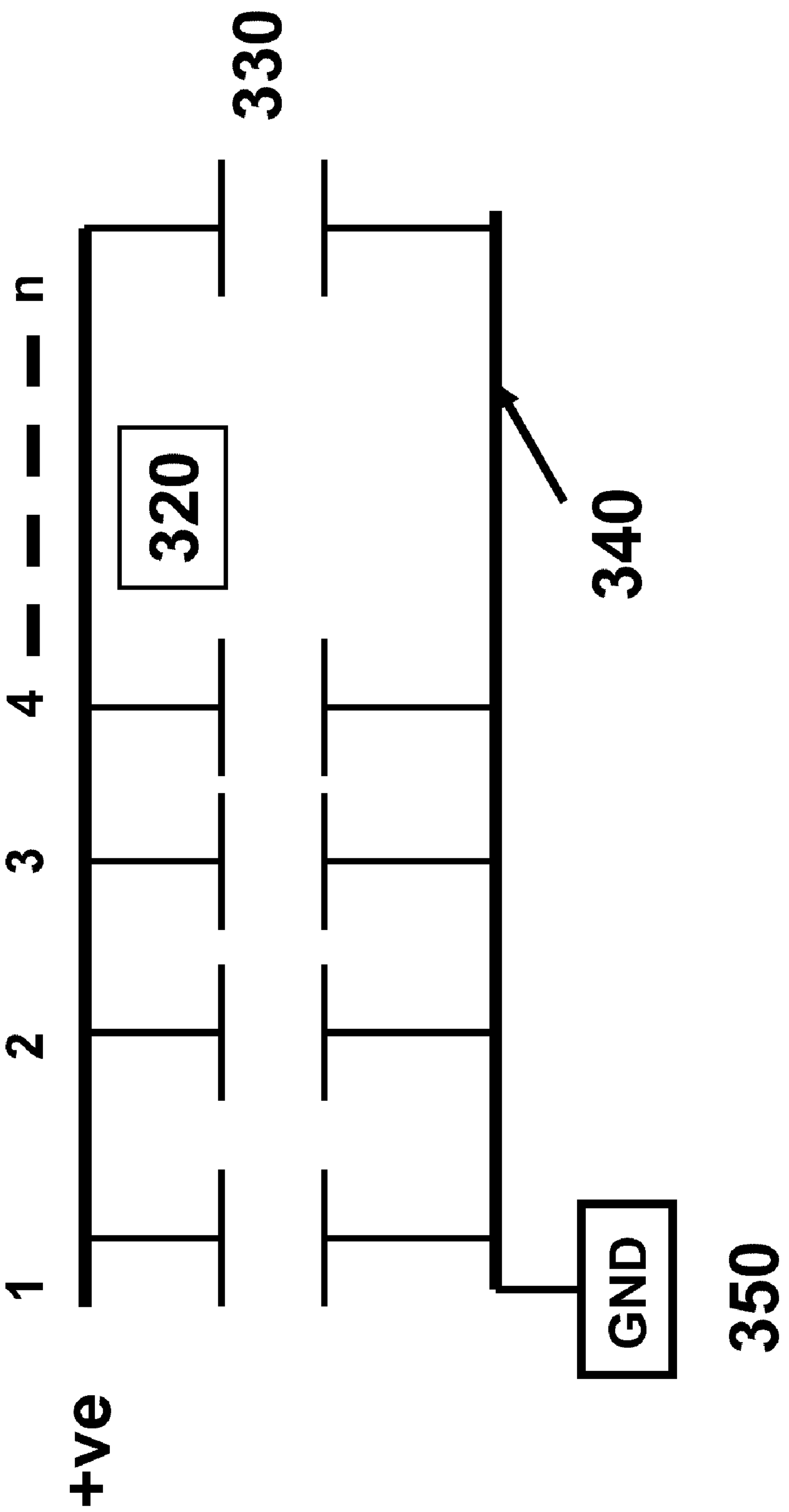
+ve
1
2
3
4
n
320
330
340
GND
350

Fig. 8a

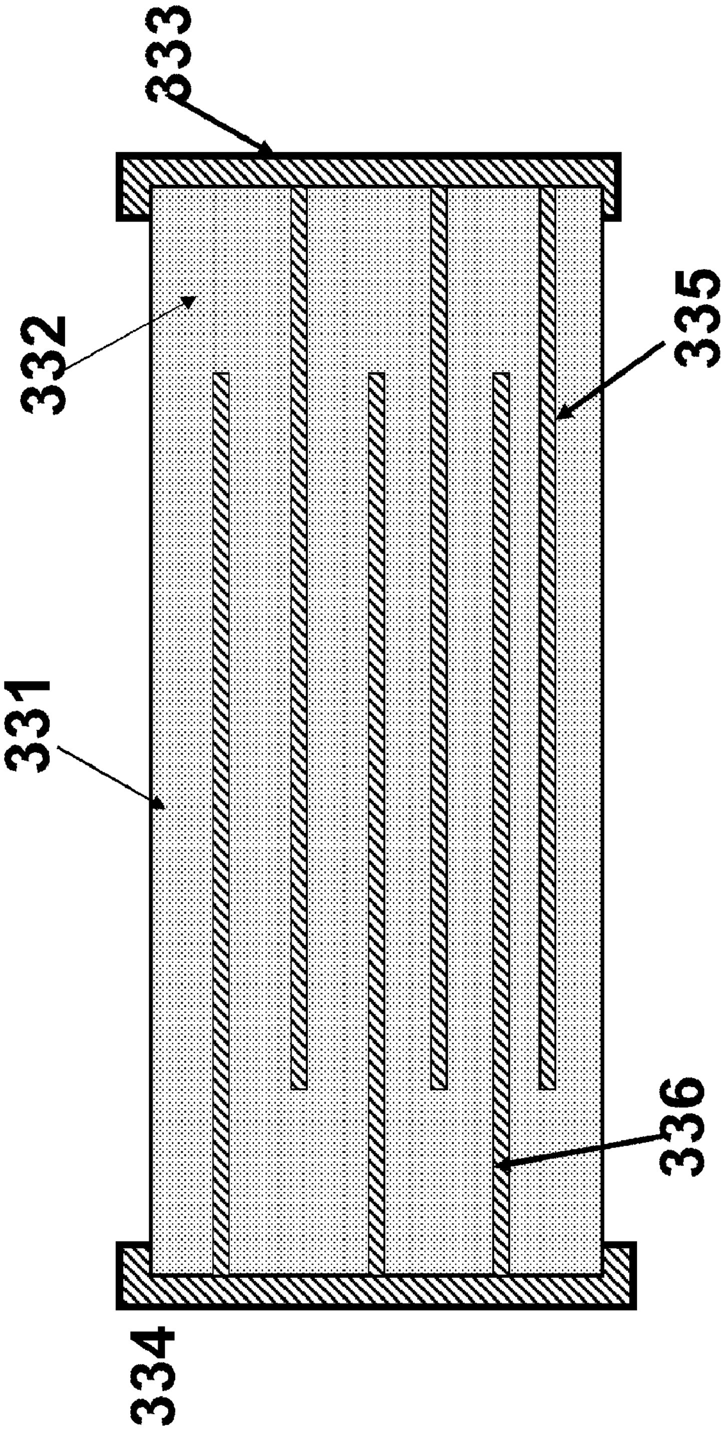
333
332
335
331
336
334

Fig. 8b

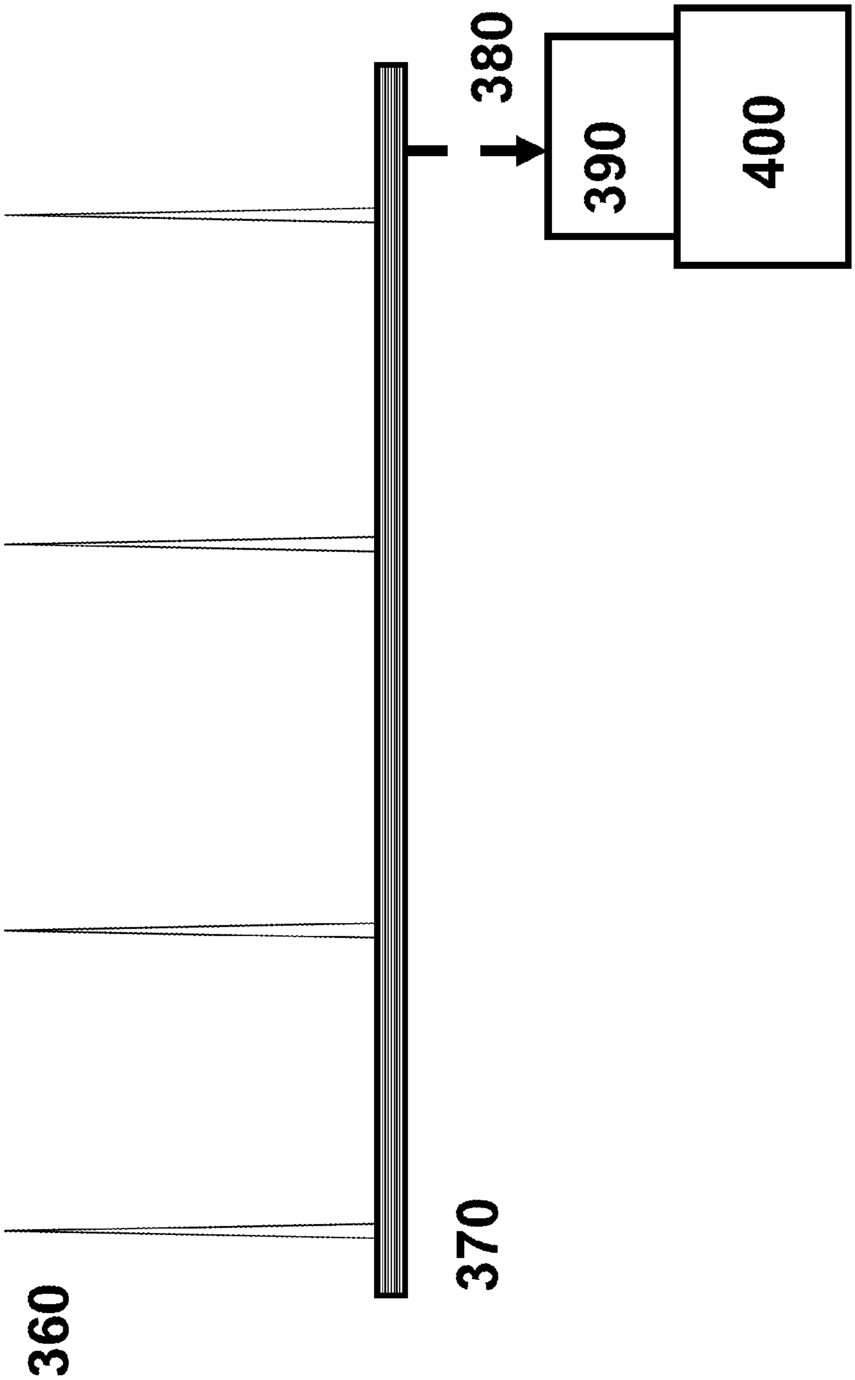
380
390
400
370
360

Fig. 9

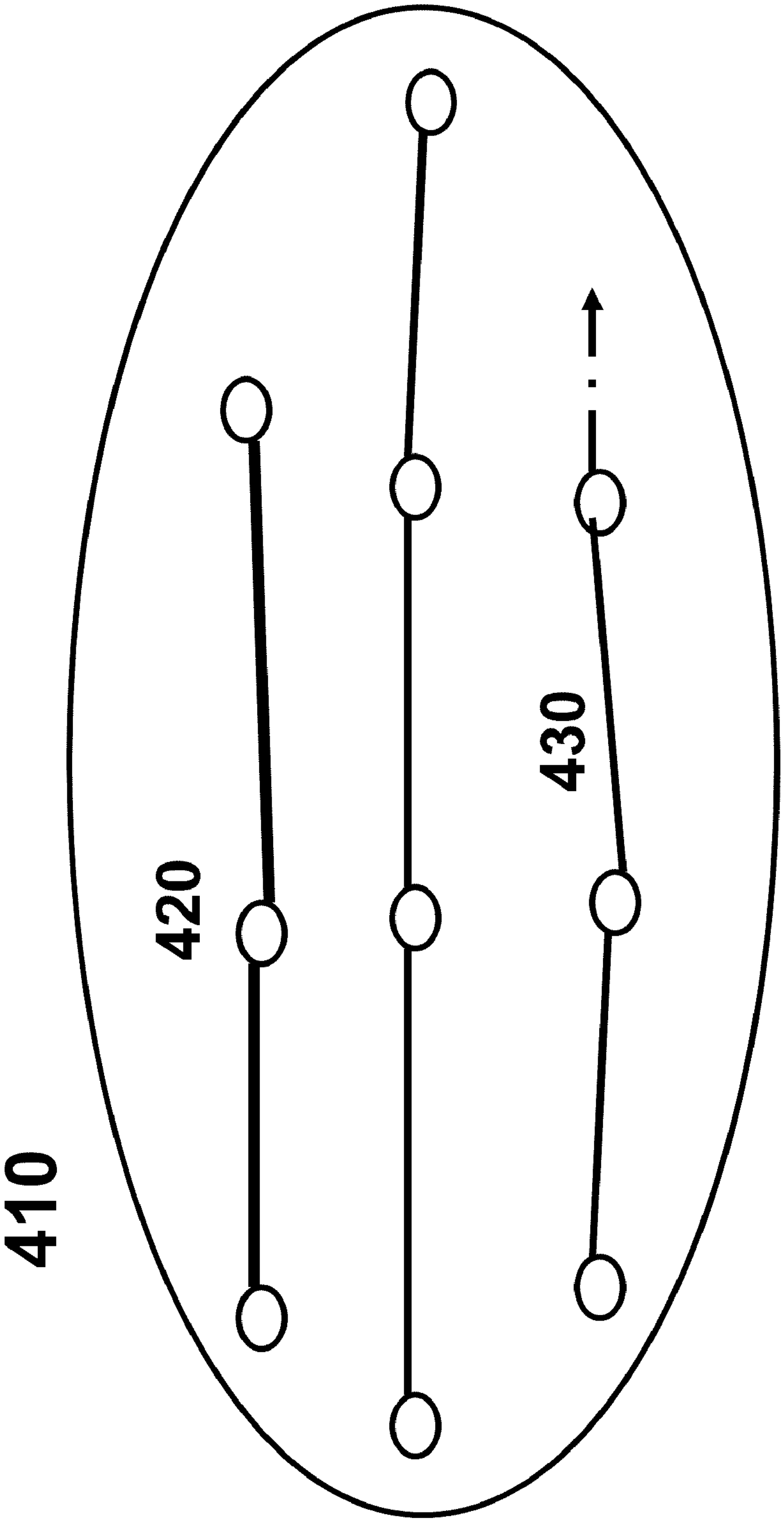

Fig. 10

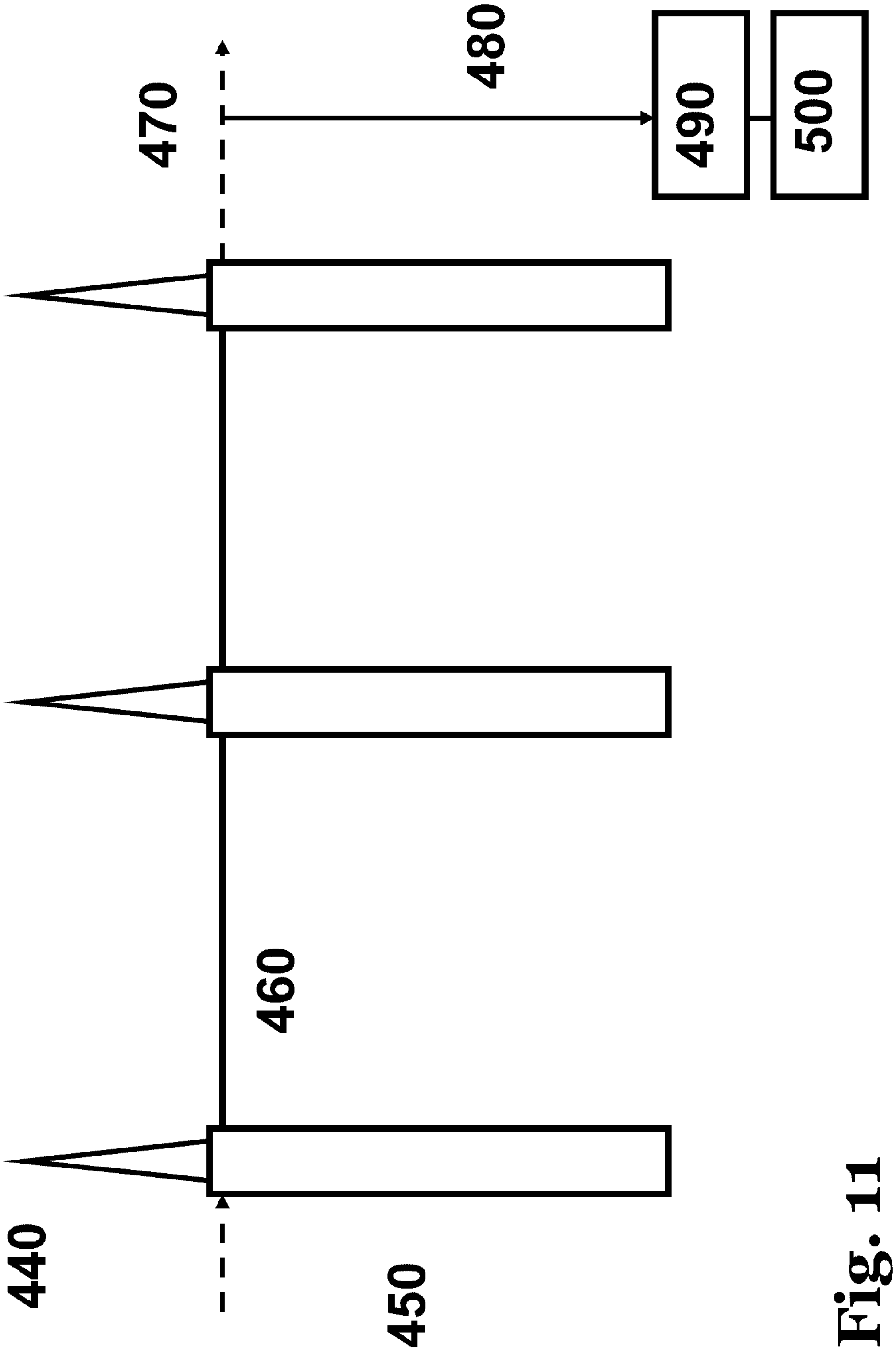
440
450
460
470
480
490
500

Fig. 11

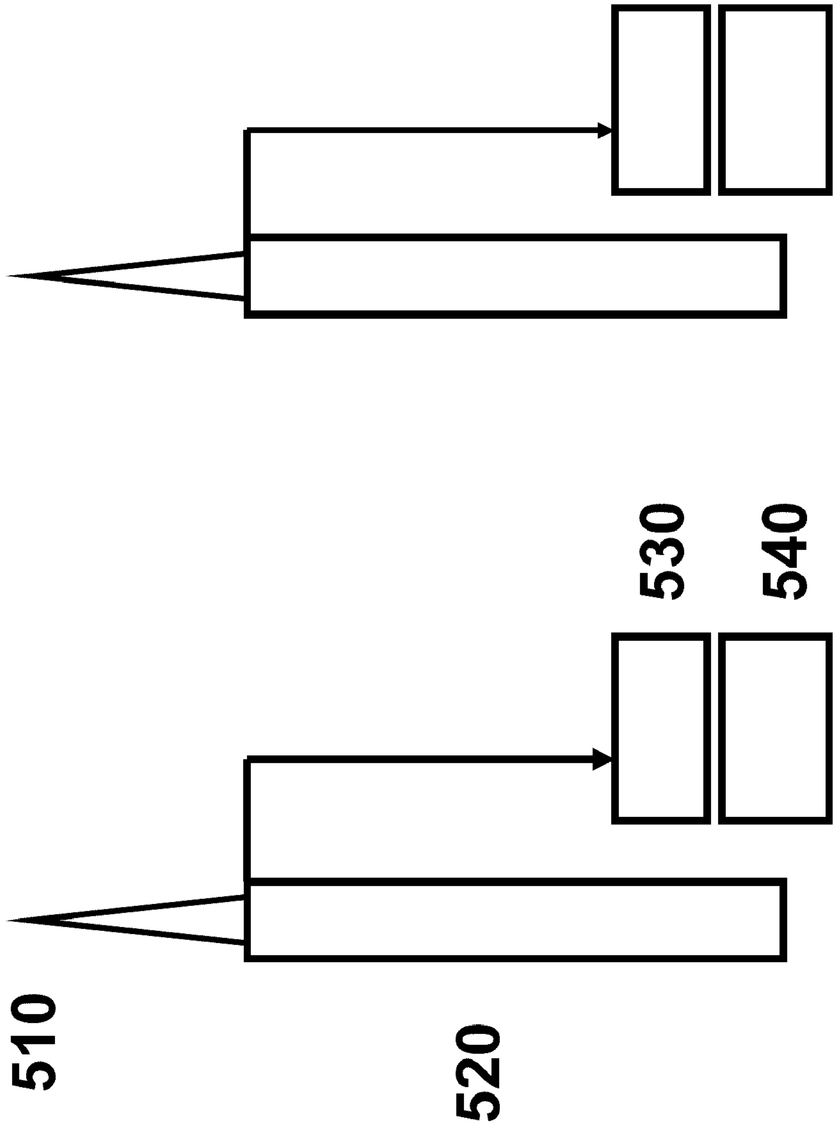
510
520
530
540

Fig. 12

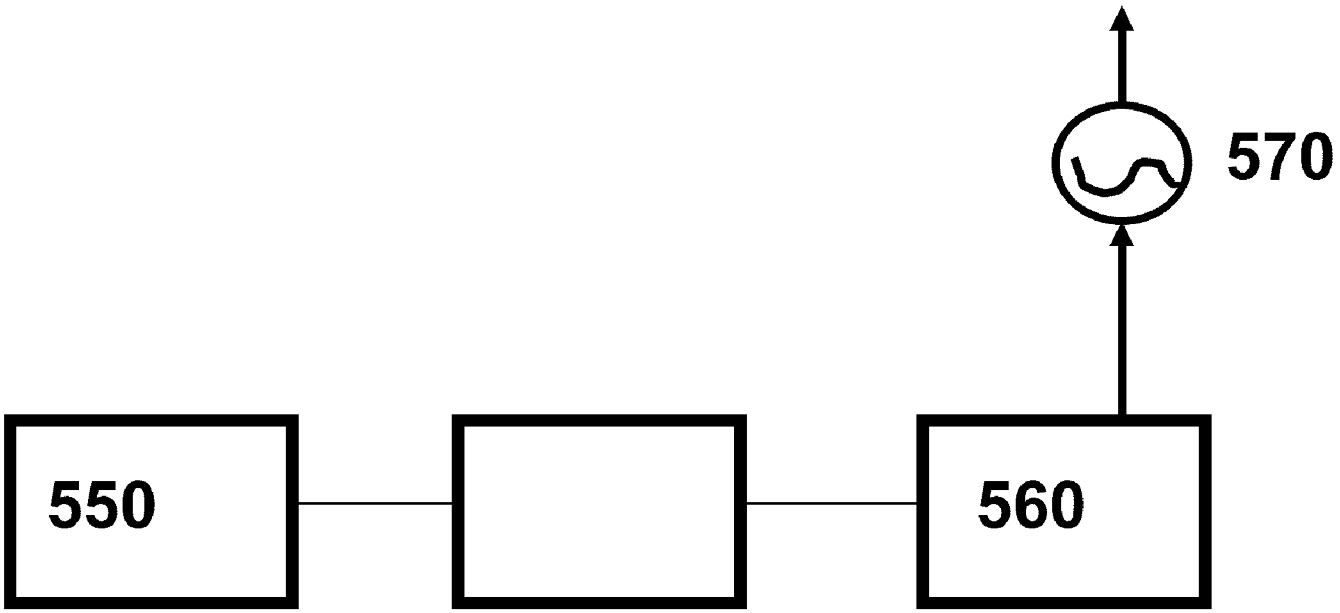

Fig. 13

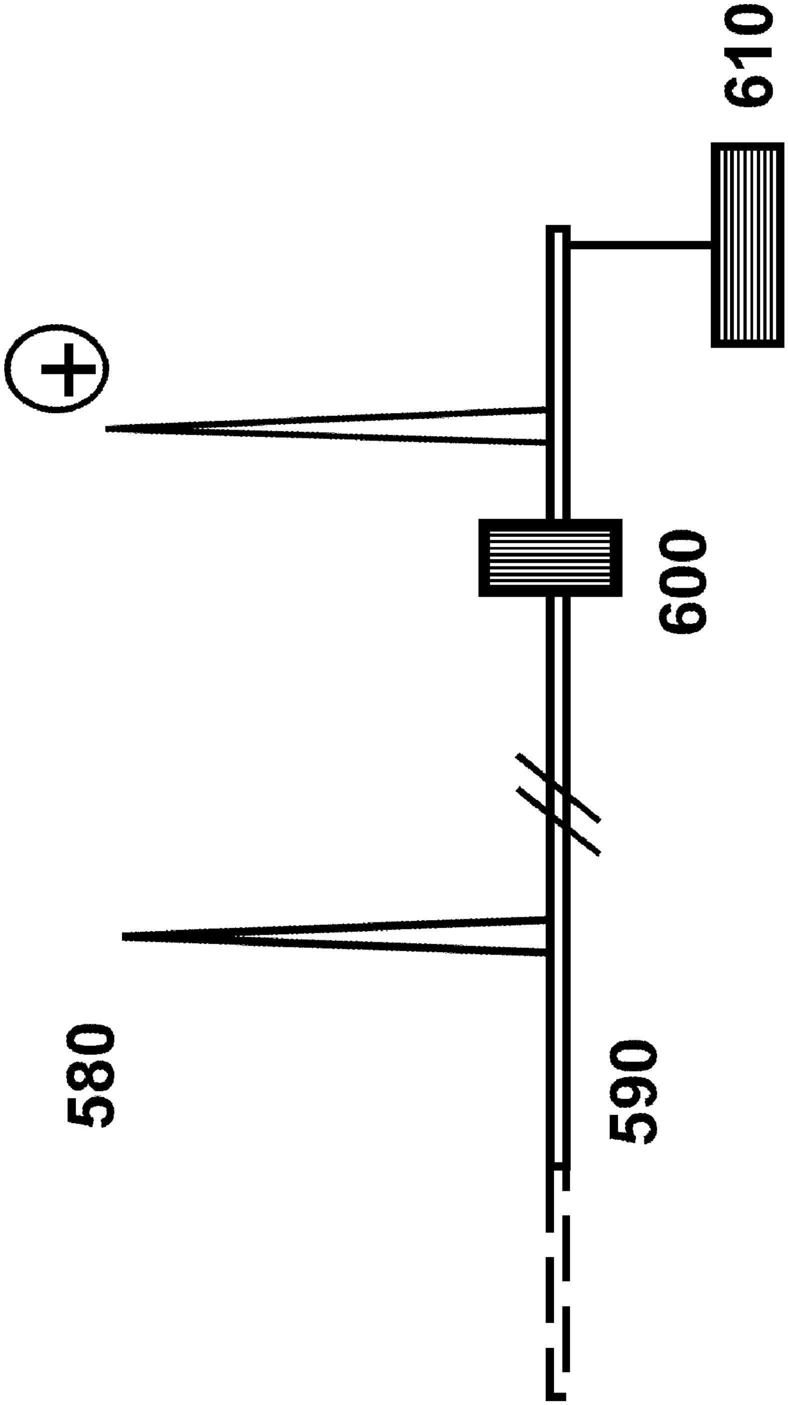
610
+
600
580
590

Fig. 14

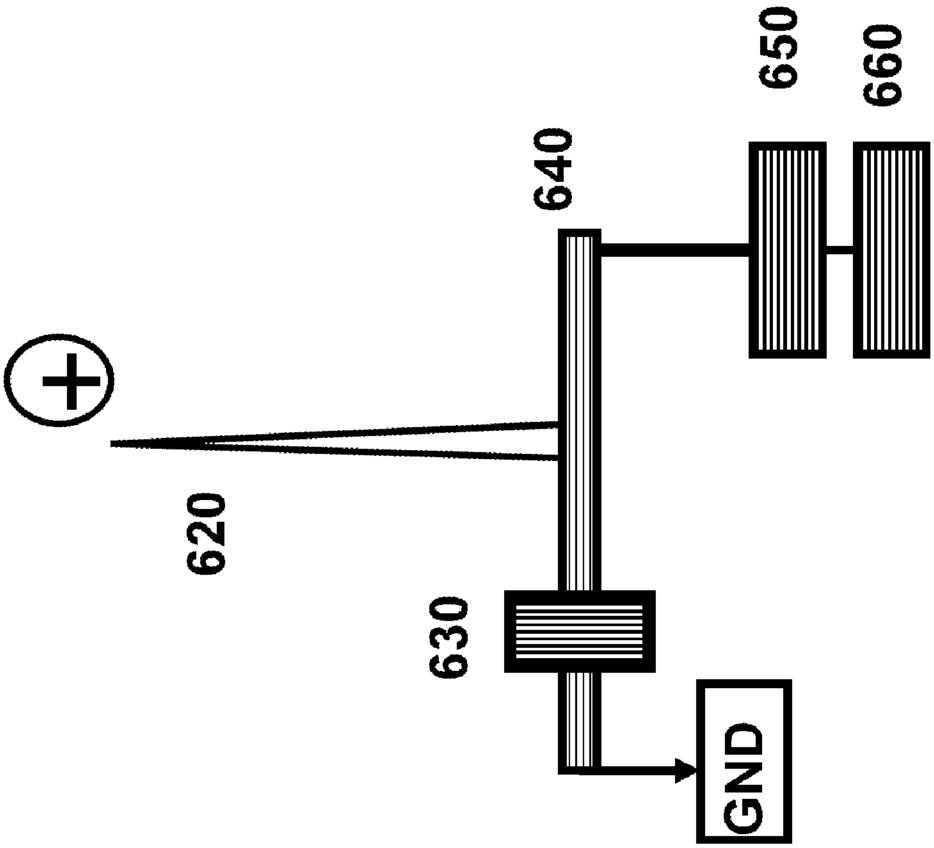
650
660
640
620
630
GND

Fig. 15

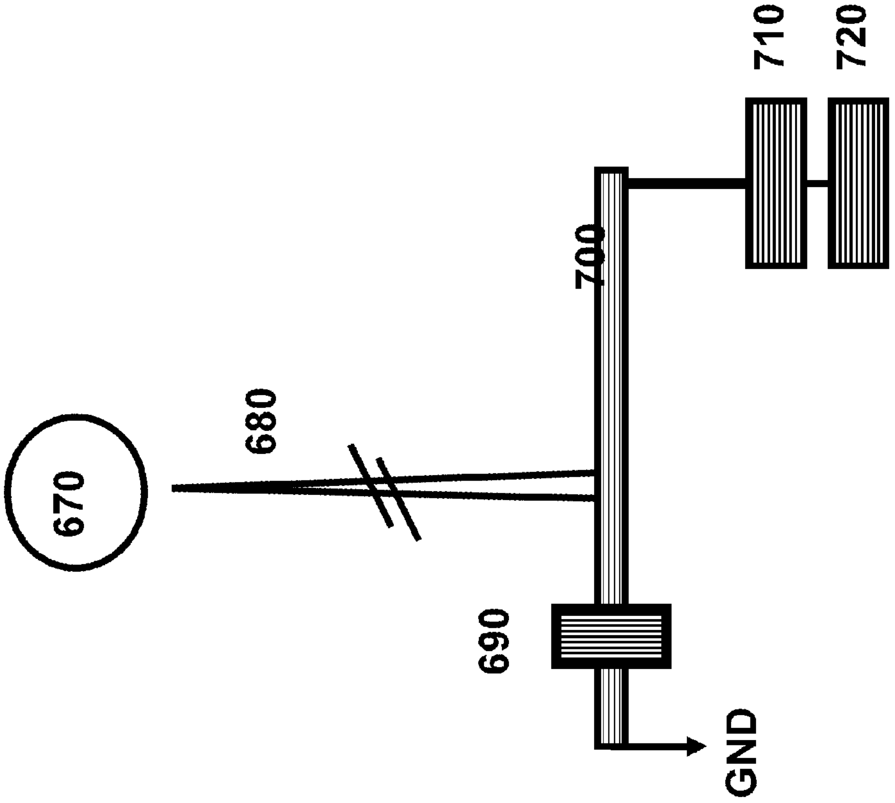
670
680
690
700
710
720
GND

Fig. 16

METHOD OF ATMOSPHERIC DISCHARGE ENERGY CONVERSION, STORAGE AND DISTRIBUTION

TECHNICAL FIELD

This invention relates to the field of electrical energy generation and storage in general and the conversion of atmospheric electrical discharge to a useable form of energy by arresting, storage and retransmission of lightning induced electrical discharge in particular.

BACKGROUND OF INVENTION

Lightning has always been a destructive force of nature. Causing death by electrocution and starting fires both in residential areas and in forests. In either case the loss of life and property is substantial. Lightning strikes have affected communication systems, and by uncontrolled arcing, electricity grids leading to loss of power with substantial economic loss

Lightning is the number one cause of forest and farm fires just in the United States and causes over 80 percent of all livestock losses due to accidents. Property losses run into billions of dollars annually and thousands of deaths and injury are reported yearly. Business losses from power and communications failure run into billions of dollars annually. Up to now, lightning strike damage has generally been controlled primarily by grounding.

A lightning discharge can have voltage of 10 MV to 100 MV and current of 1 kA to 300 kA. It packs so much voltage that it can leap a mile through the air and strike another object. Lightning can strike a building directly or leaping to it after striking a tree or other nearby object or by following a power line. It can strike any object that provides an easier path to the ground for it than the air. Because of the large network of lines, lightning can strike a utility pole at one point and be transmitted to any location in the network. Because of this, utility companies use multiple grounding and other protective devices per mile to ameliorate the lightning effects.

There has been a lot of activity in the alleviation of lightning damage but none in utilizing its energy potential.

U.S. Pat. No. 7,495,168 discloses a dipole lightning conductor which can gather electrical charges from the atmosphere during a thunderstorm, opposite to the polarity of the adjacent cloud on one side and induce a different polarity on the other end contacting the ground. When a sufficient potential difference is reached between the dipole and the earth, a dielectric breakdown occurs sending a large amount of earth charge to the thundercloud, inducing a thunderbolt.

In U.S. Pat. No. 6,012,330, Palmer discloses a method of inducing lightning strikes by shooting a stream of ionized water into a thundercloud and triggering electrical conduction through the ionized water column to the ground. In this invention, while ionized water is used, advantage is taken also of capillary wetting of conducting mesh cable at the tip of which is tethered a balloon of conductor coated Mylar. It is easier to float a balloon than to pump a column of water to any significant height.

In U.S. Pat. No. 6,320,119 Gumley discloses a capacitively coupled composite air terminal with sharp and rounded end components for voltage manipulation to minimize electric field ahead of streamers to minimize corona discharge.

Kato, in U.S. Pat. No. 5,280,335, discloses a compound lightning low voltage arrester with high energy tolerance for power and telephone lines comprising a spark gap is connected to a serial combination of Zinc Oxide arrestor and a non-inductive resistor. By cooperative execution between the spark gap and the Zinc Oxide arrester, the Zinc Oxide is chosen to be of a rating such that thermal breakdown does not occur before the spark gap ignition voltage is reached. The arrangement ensures reliable and predictable discharge and arrester survival even under multiple lightning strikes of close proximity.

Subbarao in U.S. Pat. No. 4,338,648, discloses a discharge counter and arrester current meter to determine arrester steady state current and the number of discharge events occurring during a thunder storm lightning arrester event.

Greenwald and Moses in U.S. Pat. No. 5,610,813 disclose a thunderstorm cell detection and mapping system for acquiring localized lightning strike information and identifying and locating active thunderstorm cells.

In U.S. Pat. No. 4,272,720, a method of differentiating between cloud-cloud and cloud-ground discharges is disclosed. Indicating discharge events with no accompanying lightning ground strike. This indicates that even in the absence of streamers, there is significant electrical discharge that can be converted and stored.

Weir and Nelson disclose in U.S. Pat. No. 7,033,406, a high energy density ceramic capacitor with energy capacity of 52 kW·h in 2005 cu·in, the equivalent of 1600 kW·h/m^3. Hansen in U.S. Pat. No. 6,078,494 discloses multilayer ceramic capacitors comprising doped barium-calcium-zirconium-titanate dielectric, the materials basis for U.S. Pat. No. 7,033,406. It is claimed to be characterized by high dielectric constant, high stability of capacitance value, long service life, low loss factor, high insulation resistance capacitance, low voltage dependence, and wide temperature range stability. Electrodes of base metal alloys from the group of Ni, Fe, Co or their alloys are claimed to be perhaps just as effective as noble metals containing gold, silver, platinum and palladium and may also contain Cr, Ti, Zr, V, Al, Zn, Cu, Sn, Pb, Mn, Mo and W.

Although the energy density capacity is not disclosed, it may be in the same range as U.S. Pat. No. 7,033,406 showing a path for storing the energy that can be extracted from lightning discharge.

U.S. Pat. No. 5,361,187, claims dielectric constant (k) of up to 19000 using similar group of dielectrics and processes. In U.S. Pat. No. 5,604,167 dielectric constant values of between 11,000 and 14,000 are disclosed by the same inventor with equivalent materials and processes.

Since these capacitors use standard powder processing technique, it can reasonably be expected that equivalent characteristics may be observed with naturally occurring doped silicates such as phyllosilicates for example, montmorillonite: $(Na,Ca)_{0.33}(Al,Mg)_2(Si_4O_{10})(OH)_2.nH_2O$, or cyclosilicates such as Benitoite—$BaTi(Si_3O_9)$ as base materials appropriately calcined and doped.

In U.S. Pat. No. 6,078,494 multilayer ceramic capacitors comprising doped barium-calcium-zirconium-titanate dielectric is disclosed. It is claimed to be characterized by high dielectric constant, high stability of capacitance value, long service life, low loss factor, high insulation resistance and low voltage dependence, and wide temperature range stability. Electrodes of base metal alloys from the group of Ni, Fe, Co or their alloys are claimed to be perhaps just as effective as noble metals containing gold, silver, platinum and palladium and may also contain Cr, Ti, Zr, V, Al, Zn, Cu, Sn, Pb, Mn, Mo and W. In U.S. Pat. No. 5,361,187, the inventors claim dielectric constants of up to 19000 using similar group of dielectrics and processes. In U.S. Pat. No. 5,604,167 dielectric constant values of between 11,000 and 14,000 is disclosed by the same inventors with equivalent materials and processes. Since these capacitors use standard powder processing technique, it can reasonably be expected that equivalent characteristics may be observed with naturally occurring doped silicates such as phyllosilicates for example, various grades of montmorillonite: $(Na,Ca)_{0.33}(Al,Mg)_2(Si_4O_{10})(OH)_2.nH_2O$, or cyclosilicates such as Benitoite—$BaTi(Si_3O_9)$ as base materials appropriately calcined and doped.

Although the technology to arrest lightning strikes has been around for at least a hundred years, this invention to harness and store that power has only become viable due to recent advances in super-capacitor technology which has made possible the fabrication of ultrahigh capacitance capacitors with very long lifetimes, high power, high energy densities and very fast charge and discharge rates.

Just as flood water, another destructive force of nature in times past was controlled by building dams, and from the dams eventually came hydroelectric power; and wind including thunderstorms has now been harnessed to generate electricity, this invention describes the technology to harness the power of atmospheric discharges including lightening storms to generate useable electrical energy.

This disclosure also introduces the concept of an Energy Dam. Facilities of substantial energy storage capacity capable of receiving energy feeds from various energy sources such as Solar, Wind, Thermal Power plants and Lightning. For example, during a thunderstorm, electricity generated by wind and lightning is collected and stored.

Lightning strikes every part of the globe but not uniformly. The regions with the highest historical concentration of lightning strikes are shown in FIG. **1***c*. These include Florida and the Gulf Coast in the Americas, the Equatorial Highlands of DRC, Rwanda and Burundi in Central Africa and the Monsoon Belt in Asia.

Except for the Americas, typically, these regions have very little electricity infrastructure. With the capability disclosed here, substantial reserves of electricity can be generated, stored and possibly traded.

In regions of epileptic electricity supply, disruptions in power do not necessarily translate into disruption in service. Since when power is generated, it is fed into these energy dams or local storage units for eventual use. Just as in water supply, the customer is not necessarily conscious of when the local water company is pumping water into storage tanks or local dams, in energy dams, the customer need not be conscious of whether or when electricity is being generated by the utility company because the power to his home is drawn not necessarily live from the grid but from the local energy storage units or energy dams.

BRIEF SUMMARY OF THE INVENTION

A method of harnessing atmospheric discharge and storing said energy in a useful form for subsequent use is presented.

Since the voltage of a lighting bolt can be as high as 100 MV, a methodology for safely channeling this discharge for storage is disclosed.

It is the also the object of this invention to present a storage medium in which an energy capacity of 1 MW·h in a volume of under 1 m^3 can be achieved.

It is also a further objective of this invention to show how lightening strikes on tall buildings can be harnessed to provide energy for the building either directly or fed into the grid.

It is also an objective of this invention to present a methodology for harnessing lightening, heretofore a source of power disruptions for electricity grids and converting it into a useable form of energy and feeding it back into the grid.

Since each electricity grid has a lightning suppression unit, a methodology for harnessing the impulse from these units from lightening strikes and feeding it back into the grid is disclosed.

a method of deploying a network of said storage units along a utility grid is also disclosed.

It is also the objective of this invention to disclose a method of modifying the storage units for single building deployment.

A method of deploying said storage units as vehicular charging centers even in locations not readily served by a utility grid is also disclosed.

A method of manipulating the direction and movement of the lightening bolt in open air locations and redirecting them to predetermined collection points is disclosed.

Lightning storms are some of the major causes of wild fires in the world. A method of redirecting the electrical discharges in any given space via a network of polarized and non-polarized collectors is presented. The discharges can be stored, fed into the grid or grounded.

A method for non-spontaneous discharge generation and storage is disclosed. And a method for adapting the methodology for electric automotive deployment is also disclosed.

BRIEF DESCRIPTION OF DRAWINGS

For understanding of the present invention reference is made of the accompanying drawings in the following DETAILED DESCRIPTION OF THE INVENTION. In the drawings:

FIG. **1***a*. Picture of lightning strike over a city.

FIG. **1***b*. Picture of lightning strike over open plain.

FIG. **1***c*: Global Distribution of Lightning Strikes: April 1995-February, 2003. (Source: NASA).

FIG. **2** shows a schematic of discharge collection and storage system.

FIG. **3** shows a schematic of collection unit showing metallic composition variation.

FIG. **4** Shows a Schematic of composite collector showing polarized ends.

FIG. **5** Schematic of collector network configuration with negatively polarized ends to manipulate and redirect lightning electrical discharge away from the earth for eventual collection at a predetermined location.

FIG. **6** Plan view of polarized collection network as shown in FIG. **5** acting to concentrate and modulate the electrical discharge and redirect such for collection at predetermined location.

FIG. **7***a*. Schematic showing a representation of a network of varistors arranged in rows and columns with the net effect of applying resistive arrestors in parallel and in series to effect voltage mitigation of atmospheric discharges prior to collection and storage.

FIG. **7***b*: Schematic of varistor of power mitigation unit showing cascade of ZnO arrestor in series with non-inductive resistor in parallel with spark gap.

FIG. **7***c*: Multilayer varistor with ZnO, ZnS and WO_3/WSi_2

FIG. **8***a*. Schematic showing cross section of a collection unit.

FIG. **8***b*. Schematic showing cross-section of multilayer ceramic capacitor of **8***a*.

FIG. **9** Schematic of collection unit network mesh in forested area with air frames deployed on each tree as collection of trees linked in a mesh network.

FIG. **10** Plan-view schematic of collection unit network mesh of FIG. **9**

FIG. **11** Schematic showing collection unit deployment on electricity transmission poles linked in a network and fed into the electricity grid.

FIG. **12** Schematic showing collection unit deployment on light poles or wireless towers.

FIG. **13** Schematic showing network of linked collection units and storage units on transmission lines, light poles or wireless towers fed into the grid.

FIG. **14**. Schematic of a system for non-spontaneous atmospheric discharge generation and storage.

FIG. **15** Schematic of atmospheric discharge collection and storage unit deployable on automotive units.

FIG. **16** Schematic of balloon-assisted discharge collection and storage unit.

DETAILED DESCRIPTION OF THE INVENTION

This invention describes a method of harnessing the electrical energy in atmospheric discharges, including, but not limited to lightning, storing said electrical energy and transmitting said energy for use. To this end, as an illustration, the preferred embodiments are represented in the drawings shown in FIGS. **2** to **15**.

Since lightning discharge voltages range from 10 MV to 100 MV with current of 10 kA to 100 kA, the technology challenge will be in mitigating the ultrahigh power which range from 100 GW to in excess of 10,000 GW. Although individual strikes last less than 100 μsec, thunderstorms usually last longer than 30 min. in any cloud formation and location. Indicating that the energy that can be mined from any location in a thunderstorm can run in excess of 50 GW·H. Enough energy for 500,000 homes consuming 1000 kW·h per month. At $100 per 1000 kWh, the revenue equivalent is $5 million per thunderstorm per location.

In solar energy generation, the cost of a 1 MWe rating PV panel is about $1 million. To achieve 1 GWe, the cost implications will be about $1 billion. In fossil fuel plants, the construction cost is approximately $1.3 million per MWe, putting the cost for a 1 GWe at ~$1.3 billion excluding fuel, carbon footprint, and other operating cost. Though thunderstorms are not continuous, with storage capability, the cost basis for an atmospheric discharge energy source will be several orders of magnitude lower than fossil fuel or solar of any architecture.

One embodiment of this invention is depicted in FIG. **2**. Lightning is collected at the air terminals **10**, **20** and transmitted through the conductor **30** to an arrestor **40** and on to the capacitor **50** where the electricity is stored. From which it can be extracted **60** at point of use or transmitted into the grid.

The degree of atmospheric discharge charge collection can be influenced by the metallurgical composition of the collectors. FIG. **3** illustrates the fabrication of the collector with different metals of varying conductivities and atmospheric and electrical stability. In this embodiment, metallurgical compositions for the various segments comprise Ti, Cu, Al. Other metals or their alloys can be used depending on the required properties and the locations of deployment.

In yet another embodiment of the invention as shown in FIG. **4**, a composite collector with opposite polarity air terminals **100**, **110** is used. The air terminals are isolated from each other by an insulator **120**. The negative polarity terminal serves to deflect and manipulate the charge to the positive terminal from which it is conducted to the storage unit **130**.

Another embodiment of the invention is illustrated in FIGS. **5** and **6**. In this representation, negative polarity collectors, **160** deflect the atmospheric discharge so that they do not reach the ground but are deflected for collection at a predetermined terminal **180** of positive polarity and thereafter storage **200**. A plan view representation of this embodiment is shown in FIG. **6**.

It is essential to mitigate the power of the lightning before storage. In one embodiment of this invention, a collection of varistors of various power capacities, linked in series **270**, **280** and **290** are connected in parallel in lines **1** to n, to achieve a desired capability as shown in FIG. **7***a*. The resultant power is stored in the capacitor **300**. An insulator or fuse **305** is placed at the end of the varistor chain to protect the capacitor **300** against over-current. In the event of overstressing, the dielectric or varistor is deformed and the excess current is grounded.

Each component of the power mitigation unit is any of the following configurations as shown in FIG. **7***b*: A ZnO varistor **312** connected in series with a non-inductive resistor **313** and in parallel with a spark gap **311**. The resultant current is further modulated through another ZnO unit **314** and resistor **315**. A cascade of this configuration could be adopted to achieve the desired properties.

Zinc Oxide has a relatively large band gap of ~3.3 eV at room temperature. The consequence of this large band gap includes, higher breakdown voltages, ability to sustain large electric field and high power and temperature operation. Because of this, ZnO is the primary component in varistors and in voltage suppression devices. Zinc Sulfide has a band gap of 3.5 eV for the cubic form and as high as 3.91 eV for the hexagonal form at 300K. While ZnO band gap can be tuned from between 3-4 eV by alloying with magnesium oxide or cadmium oxide, the co-processing of ZnO with ZnS of the hexagonal wurtzite form should provide electrical field and other operational advantages over ZnO alone.

Multilayers of ZnO and ZnS or as sandwich layers with each other with ZnO achieving 50-99.9 mol. % of each layer can be adopted to improve on the suppression properties of the single layer ZnO varistor as is shown in FIG. **7***c*. Said varistor comprising a ceramic base body **318** of multiple layers, terminal electrodes **316**, **317** and internal electrodes **319**.

Alloying of ZnO with metal oxides and nitrides of transition elements including W, Zr, Mo, Mg, Cr, and Ti may improve its band gap characteristics and improve its voltage response non-linearity. Tungsten, in the form of tungsten silicide has very unique crystallization characteristics, particularly in the presence of dopants such as phosphorus or arsenic. Upon post deposition anneal, the Si in the non-stoichiometric tungsten silicide (WSix) segregates to the grain boundaries and the film interfaces. The dopant atoms diffuse preferentially along the grain boundaries. When oxidized, it is the segregated silicon that is oxidized selectively. If the W is exposed to the oxidizing ambient of O_2/Ar or O_2/N_2 at temperatures in excess of 700° C., tungsten oxide WOx (x=3, or 1.5) is formed. It has no corresponding crystallographic orientation to the underlying tungsten silicide or metal, being in the form of an amorphous powder. The fabrication of multilayer varistors with ZnO and WSi_2/WO_3, SiO_2 should provide the flexibility of modulation of the effective grain sizes of the host material to within 0.5-10μ as the case may be. This precise grain size control provides corresponding modulation of the electrical properties including overvoltage suppression.

A schematic of the storage unit **320** is shown in FIG. **8***a*. A collection of capacitors **330** are connected in parallel to achieve a desired capacitance. A schematic of the cross-section of each capacitor unit is shown in FIG. **8***b*. The unit **331** with dielectric layers **332** is a multilayer capacitor comprising alternately stacked internal electrodes **335**, **336** connected to external electrodes **333**, **334**. The dielectric layers may be of uniform composition, or layers of different compositions optimized for flexibility in the modulation of desired electrical, physical and reliability properties.

Capacitance in excess of 1 kF/m^3 with energy storage capacity of 1.5 MW·h is reported as being possible by existing art. With compositional and manufacturing improvements disclosed here these benchmarks are expected to be surpassed making possible energy storage capacities of up to 1 GW·h. Sufficient energy for 1000 households consuming 1000 kW·h of electricity per month for one month.

Lightning storms are a major cause of wildfire. To address this, in yet another embodiment of this invention as represented in FIG. **9**, air terminals **360** are set up on a selection of trees and connected by a conductor **370** in a mesh array as shown in FIG. **10**. The impinging charge is transmitted through a conductor **380**, through a suppressor unit **390** and on to the storage unit **400**.

Utility transmission lines usually have lightning collectors along the length of the network. In this invention as demonstrated in FIG. **11** air terminals **440** deployed transmission poles **450** are linked by a polarized conductor **460**. At predetermined locations the impinging charge is conducted through a suppression unit **490** and thereafter into the storage unit **500**. The stored energy is periodically released into the grid.

In yet another embodiment of the invention, air terminals **510** are deployed on light poles **520**. The collected charged are modified and stored in a storage unit **540**.

In another embodiment, the air terminal/collection units are deployed on radio transmission towers. The charge could be stored individually at each tower location or linked in a network as shown in FIG. **13**. In either case, the stored energy be used at each individual location or fed into the grid **570**.

In the absence of a thundercloud, lightening flash can be generated by the modification of the natural bias between the cloud and ground. This is demonstrated in FIG. **14**, where a positively biased conductor **580** of height greater than nearest ground elevation or nearest structure. The increased elevation and bias facilitates cloud to ground discharge even when there is no active lightning storm. The incoming charge is isolated by an insulator **600** and stored in the storage unit **610**.

In yet another embodiment of this invention, a balloon of conductive material is tethered to a terminal via a mesh line of conducting material. The balloon captures cloud to cloud discharges which are transmitted down to storage. In yet another embodiment this tether line may be wetted by a conductive solution such as but not limited to salt solution to facilitate the discharge conduction.

Another embodiment of this invention is shown in FIG. **15**. In an isolated collection system as shown here, the air terminal **620** which may be biased or unbiased is connected to a collection/storage unit **650**, **660** respectively by conductor **640** and isolated by a dielectric or insulator **670**. Should the discharge exceed the capacity of the isolation, the isolation is bridged and the charge grounded. This isolated collection system is particularly useful where there is no utility grid, particularly in very remote and isolated locations.

This isolated collection capability enables deployment on individual houses and isolated farms. In electric automotive recharge stations, this capability enables the location of these facilities in any location even those not served by regular electricity utilities lines.

In isolated rural areas, the deployment of these isolated collection units enables the provision of electricity where electricity grid infrastructure is not available or practical. The storage units can be deployed in banks as the estimated anticipated usage dictates.

In another embodiment of this invention, the isolated collection unit is modified to be portable for deployment on automobiles. The charge in the storage unit is linked with the automobile's electrical power or storage unit. Alternatively, it can replace the power storage unit of the automobile for pure EV or hybrid cars. Additionally, it can unplug these cars from the grid. With the ambitious industry plans for electric or hybrid cars the infrastructure for charging these cars either in garages or charging stations is somewhat limited. The capability to both 'charge as you go' and unplug from the grid would expand the infrastructure significantly.

In yet another embodiment of this invention, the isolated collection unit is deployed on trains. This could be in commuter trains that directly fed from the grid, or long-haul trains to augment or replace the fossil fuel used.

It is understood that the presentation of these steps in this disclosure is not exhaustive. Only the preferred embodiments of the invention and but a few of the examples of its versatility are shown and described in the present disclosure. It should be readily apparent to those of ordinary skill in the art that the invention is capable of use in various other combinations, environments and applications and is capable of changes or modifications within the scope of the inventive concept as expressed herein. These changes and modifications may be made without departing from the spirit and scope of the invention as set forth in the appended claims.

The invention claimed is:

1. A method of converting an atmospheric electrical discharge into a

useable form of electrical energy comprising

deflecting the discharge to an air terminal via one or more separate air terminals of a polarity, the air terminal having an opposite polarity to the polarity of the separate air terminals;

arresting the discharge via the air terminal to a power mitigating (step down) unit via a conducting unit connected to the air terminal; wherein said power mitigating unit is not a transformer,

reducing voltage of the discharge via the power mitigating unit modulated according to a capacity of the power mitigating unit, wherein the discharge is grounded if in excess of the capacity, and wherein the power mitigating unit reduces the voltage of the discharge from approximately higher than 100 MV (mega volt) and 100 kA (kilo amperes) to below 20 kV (kilo volt), 10 kA; and

storing the discharge with the reduced voltage in an energy storage unit connected to the power mitigating unit, the storage unit comprising modules of high energy density capacitors and the storage unit discharging the electrical energy into a power grid or to an appliance.

2. The method of claim **1** wherein the discharge is caused by incident lightning and wherein the arrestor and mitigating unit comprise an array of self-healing varistors in series and parallel chosen to ensure total conversion of the discharge to a voltage/current level suitable for storage in said capacitors.

3. The method of claim **1**, where the power mitigation unit a configuration selected from a group consisting of:

(a) single ZnO varistor,

(b) Zinc Oxide arrester connected in series with a non-inductive resistor,

(c) Zinc Oxide arrester in series with a non-inductive resistor and in parallel with a spark gap, and

(d) a cascade of configurations as described in (c).

4. The method of claim **1** wherein the storage unit comprises modular array of high energy density multilayered doped ceramic capacitors with storage capacity of 10 kW·h (kilowatt hour) to 1 GW·h (gigawatt hour) and volume 0.1 to 1 m^3 and up to 1000 units.

5. The method of claim **1** in which the air terminal and the storage unit are configured in a mobile unit and deployed on mobile platform having a battery, wherein the mobile unit provides the electrical energy to power the battery without a need for recharging the battery at fixed charging locations.

6. The method of claim **1** wherein the storage unit is located in a charging station, and wherein the storage unit includes modular capacitors for charging an electric vehicle from the charging station.

7. The method of claim **1** wherein the air terminals are deployed on buildings, wherein a portion of the electrical energy stored is used by the buildings and wherein a portion of electrical energy is stored in a central energy storage facility connected to a plurality of storage units.

8. The method of claim **1** wherein the said storage units are located underground.

9. The method of claim **1** wherein the storage unit is receives energy feeds from more than one electricity generation sources, each source being selected from a group consisting of Thermal energy source, Hydro energy source, Solar energy source, Wind energy source and Lightning energy source.

10. The method of claim **4** wherein the ceramic capacitors are fabricated to provide modulation of effective grain sizes within a specified range.

11. The method of claim **4** wherein the multilayer ceramic capacitors comprises dielectric layers of varying compositions individually or as sandwich layers with each other containing Barium-Titanate silicates as host material doped with Zr, WSi_2/WO_3 to modulate grain growth to between 0.1 to 10 μm, said modulation providing corresponding control of electrical and reliability properties.

12. The method of claim **2** wherein the varistors comprise multilayers of ZnO and ZnS individually layered or as sandwich layers with each other with ZnO achieving 50-99.9% of each layer, said layers being alloyed metal oxides of transition elements to improve voltage response non-linearity.

13. A method to protect a position from impact of an electrical

discharge caused by lightning, the method comprises redirecting the electrical discharge from the position via an array of air terminals to an air terminal of a collection unit, said collection unit comprising a power mitigating unit and a storage unit and wherein the power mitigating unit is not a transformer, the air terminal of the collection unit biased with a first

polarity, each of the array of air terminals biased with a second polarity with a magnitude, the second polarity opposite to the first polarity, wherein the redirection is controlled by the magnitude and the second polarity; and

storing the electrical discharge in the collection unit located significantly removed from the impact position.

14. The method of claim **13** wherein the discharge redirection is via a biased conductor to the collection unit.

15. The method of claim **13** wherein the position is located in an area of a forest, wherein the array of air terminals are deployed on trees in the forest, wherein tips of the array of air terminals are positioned in close proximity to treetops of the trees, and wherein the discharge is fed into an electricity grid as electrical energy from the collection unit.

16. The method of claim **1** wherein the storage unit is connected to electrical power lines of the power grid and wherein the electrical energy is fed back into the power grid via the electrical power lines.

17. The method of claim **1** wherein the storage unit is located at a telecom tower, a light poles or other free standing structure with a need for operational energy and wherein the electrical energy from the storage unit is used for the operational energy.

18. A method of generation of atmospheric discharge, the method comprising:

modulating bias of a dipole air terminal conductor to capture charges from clouds to ground without occurrence of natural lightning, the conductor including a balloon, a terminal and a tether line connecting the balloon and the terminal, wherein the bias has a magnitude and a polarity controlled by the modulation relative to the ground; and wherein the tether line is a conductive cable soaked by an ionic liquid including salt solution to facilitate conduction of the charge, and

storing the charge in a collection unit comprising a power mitigation unit and a storage unit wherein said power mitigation unit is not a transformer, feeding said charge into an electricity grid or appliance.

* * * * *

A lightning flash to a 23-foot-tall European ash tree near Lugano, Switzerland. The tree suffered no external damage. Photograph was taken from a distance of about 200 feet by Richard E. Orville, State University of New York at Albany, using Kodachrome II 35 mm daylight film, a lens setting of f/5.6, and the shutter on time exposure. (Reproduced in color on front cover.) Additional details are found in Orville, R. E., Photograph of a Close Lightning Flash, *Science, 162,* 666-667, 1968. Lightning photography is discussed in Chapter 10; lightning damage to trees in Chapter 5.

All About
LIGHTNING

Martin A. Uman
Department of Electrical Engineering
University of Florida, Gainesville

Dover Publications, Inc., New York

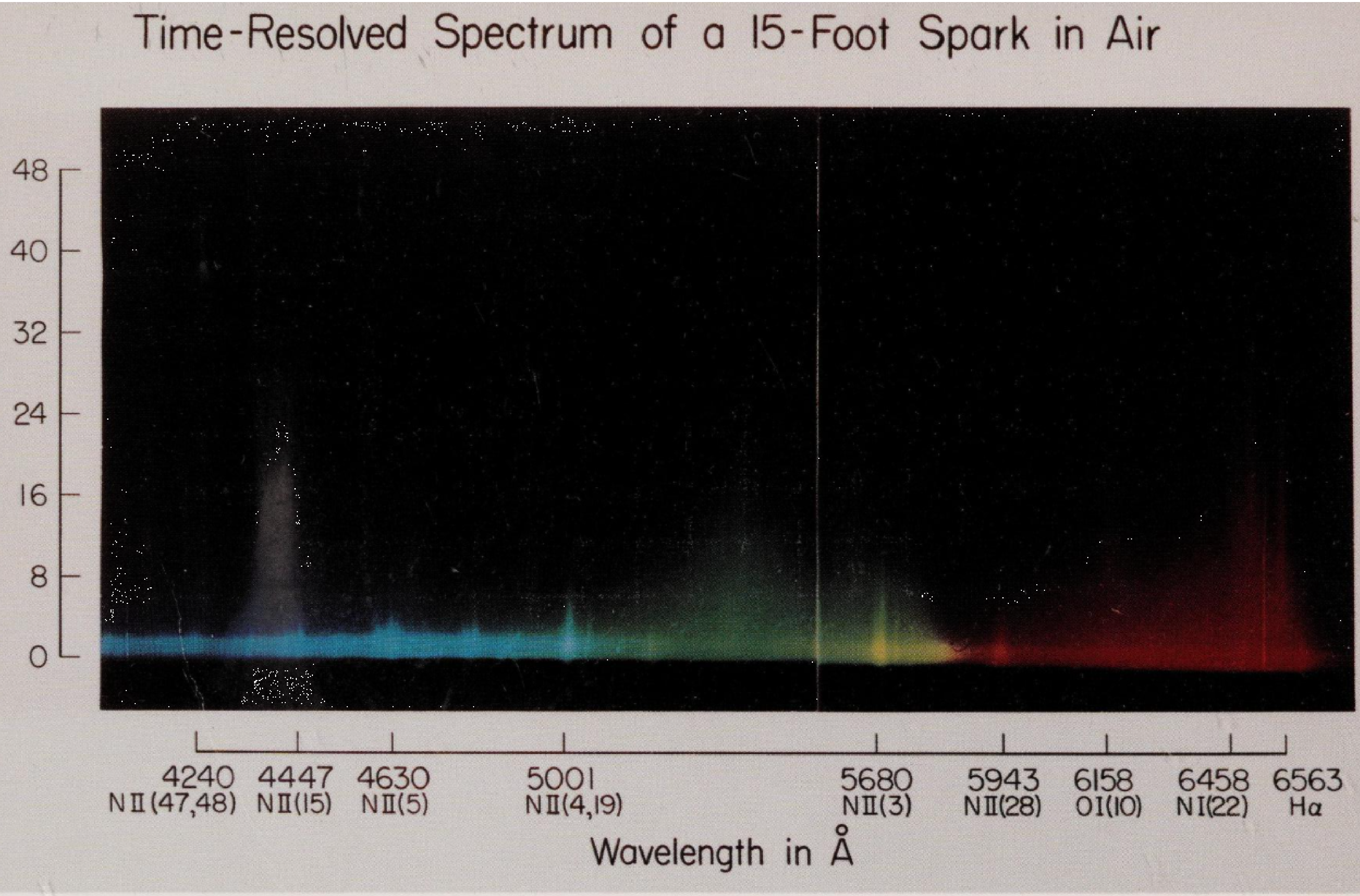

For caption, see page 99.

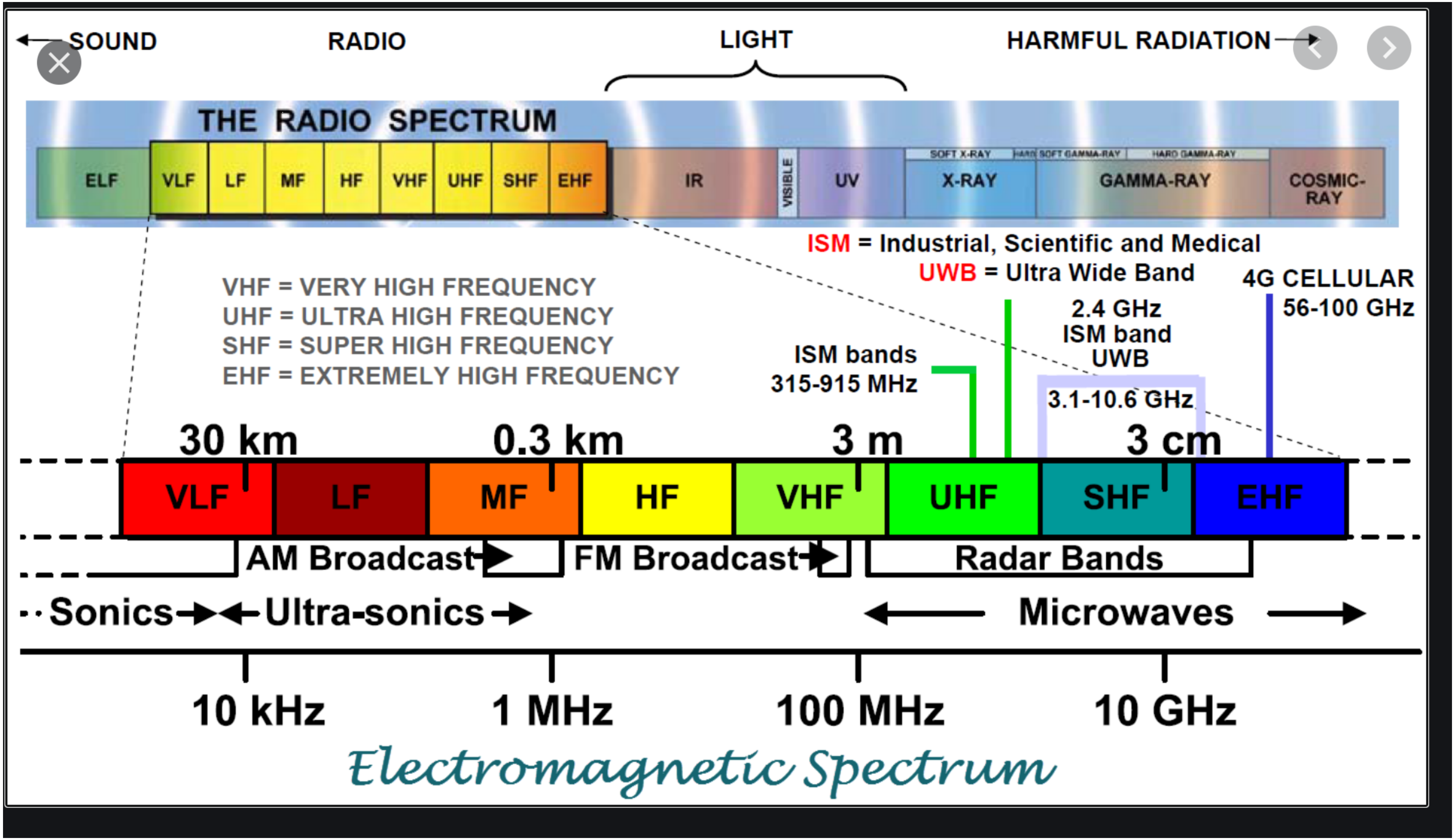
SOUND
RADIO
LIGHT
HARMFUL RADIATION
THE RADIO SPECTRUM
ELF
VLF
LF
MF
HF
VHF
UHF
SHF
EHF
IR
VISIBLE
UV
SOFT X-RAY
HARD
SOFT GAMMA-RAY
HARD GAMMA-RAY
X-RAY
GAMMA-RAY
COSMIC-RAY
ISM = Industrial, Scientific and Medical
UWB = Ultra Wide Band
4G CELLULAR
56-100 GHz
VHF = VERY HIGH FREQUENCY
UHF = ULTRA HIGH FREQUENCY
SHF = SUPER HIGH FREQUENCY
EHF = EXTREMELY HIGH FREQUENCY
2.4 GHz
ISM band
UWB
ISM bands
315-915 MHz
3.1-10.6 GHz
30 km
0.3 km
3 m
3 cm
VLF
LF
MF
HF
VHF
UHF
SHF
EHF
AM Broadcast
FM Broadcast
Radar Bands
Sonics
Ultra-sonics
Microwaves
10 kHz
1 MHz
100 MHz
10 GHz
Electromagnetic Spectrum

ELECTROMAGNETIC RADIATION SPECTRUM

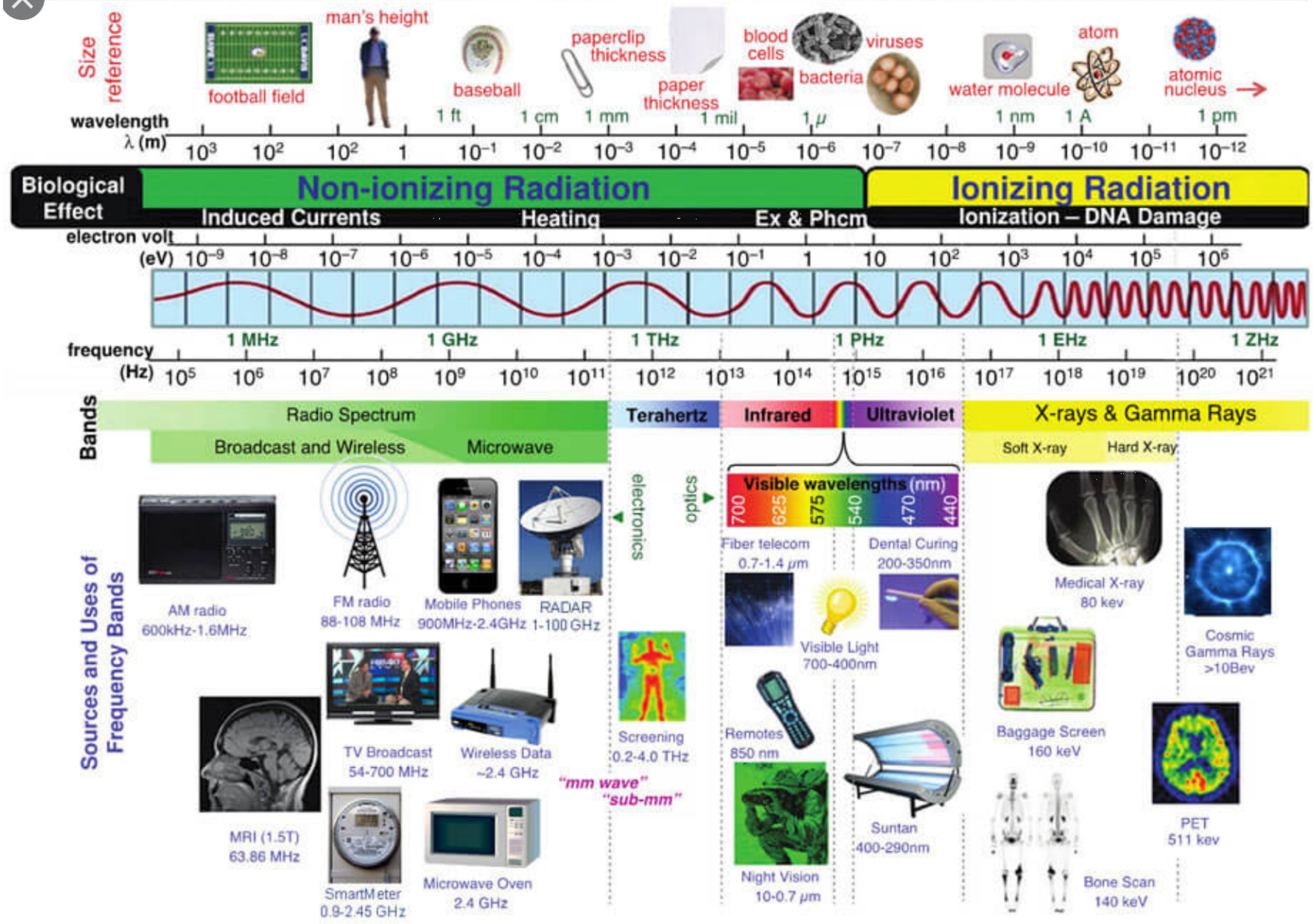
Size reference
football field
man's height
baseball
paperclip thickness
paper thickness
blood cells
bacteria
viruses
water molecule
atom
atomic nucleus
wavelength λ (m)
1 ft
1 cm
1 mm
1 mil
1 μ
1 nm
1 A
1 pm
10³ 10² 10² 1 10⁻¹ 10⁻² 10⁻³ 10⁻⁴ 10⁻⁵ 10⁻⁶ 10⁻⁷ 10⁻⁸ 10⁻⁹ 10⁻¹⁰ 10⁻¹¹ 10⁻¹²
Biological Effect
Non-ionizing Radiation
Induced Currents
Heating
Ex & Phcm
Ionizing Radiation
Ionization – DNA Damage
electron volt (eV)
10⁻⁹ 10⁻⁸ 10⁻⁷ 10⁻⁶ 10⁻⁵ 10⁻⁴ 10⁻³ 10⁻² 10⁻¹ 1 10 10² 10³ 10⁴ 10⁵ 10⁶
frequency (Hz)
1 MHz
1 GHz
1 THz
1 PHz
1 EHz
1 ZHz
10⁵ 10⁶ 10⁷ 10⁸ 10⁹ 10¹⁰ 10¹¹ 10¹² 10¹³ 10¹⁴ 10¹⁵ 10¹⁶ 10¹⁷ 10¹⁸ 10¹⁹ 10²⁰ 10²¹
Bands
Radio Spectrum
Broadcast and Wireless
Microwave
Terahertz
Infrared
Ultraviolet
X-rays & Gamma Rays
Soft X-ray
Hard X-ray
Visible wavelengths (nm)
700 625 575 540 470 440
electronics
optics
Sources and Uses of Frequency Bands
AM radio 600kHz-1.6MHz
FM radio 88-108 MHz
Mobile Phones 900MHz-2.4GHz
RADAR 1-100 GHz
MRI (1.5T) 63.86 MHz
TV Broadcast 54-700 MHz
Wireless Data ~2.4 GHz
SmartMeter 0.9-2.45 GHz
Microwave Oven 2.4 GHz
"mm wave"
"sub-mm"
Screening 0.2-4.0 THz
Fiber telecom 0.7-1.4 μm
Visible Light 700-400nm
Dental Curing 200-350nm
Remotes 850 nm
Night Vision 10-0.7 μm
Suntan 400-290nm
Medical X-ray 80 kev
Baggage Screen 160 keV
Bone Scan 140 keV
Cosmic Gamma Rays >10Bev
PET 511 kev

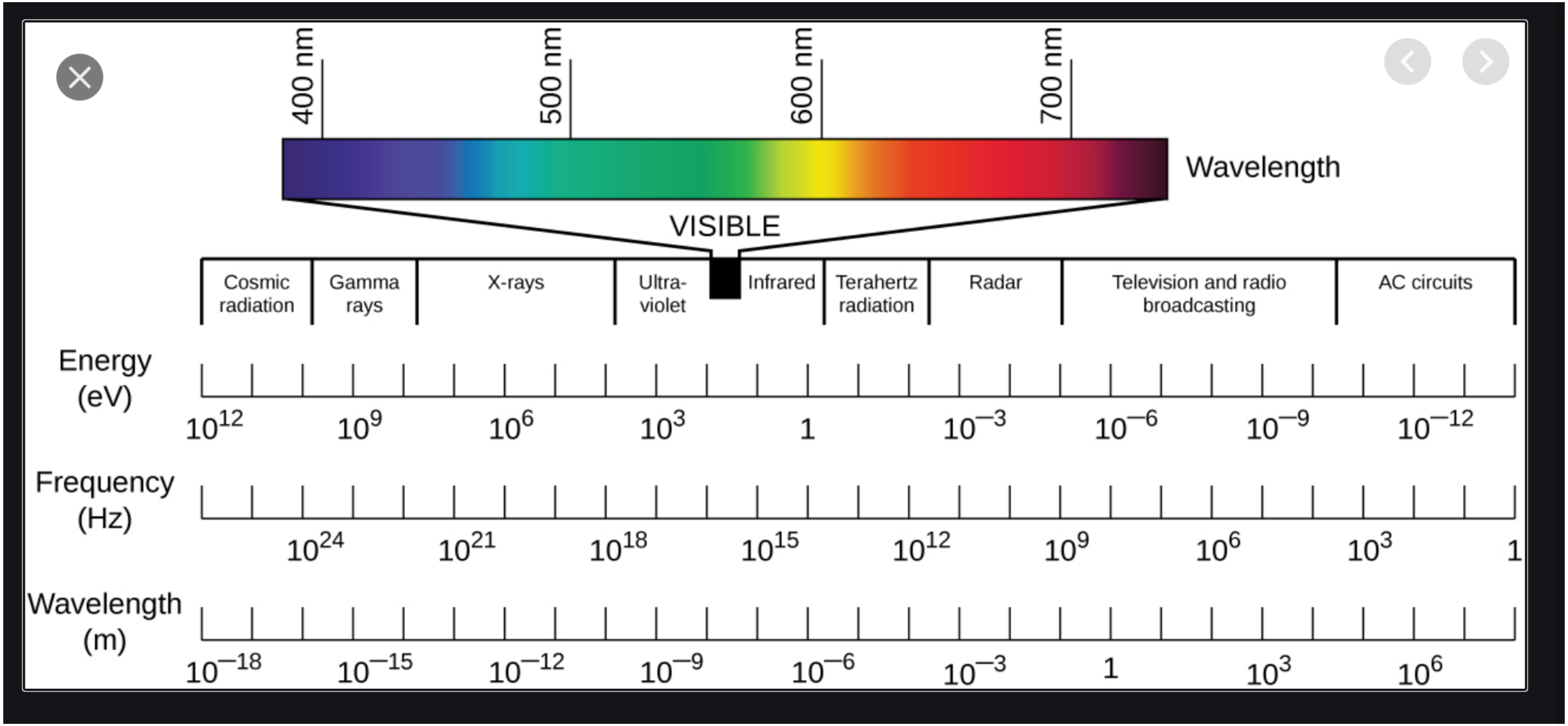
400 nm
500 nm
600 nm
700 nm
Wavelength
VISIBLE
Cosmic radiation
Gamma rays
X-rays
Ultra-violet
Infrared
Terahertz radiation
Radar
Television and radio broadcasting
AC circuits
Energy (eV)
10^{12}
10^{9}
10^{6}
10^{3}
1
10^{-3}
10^{-6}
10^{-9}
10^{-12}
Frequency (Hz)
10^{24}
10^{21}
10^{18}
10^{15}
10^{12}
10^{9}
10^{6}
10^{3}
1
Wavelength (m)
10^{-18}
10^{-15}
10^{-12}
10^{-9}
10^{-6}
10^{-3}
1
10^{3}
10^{6}

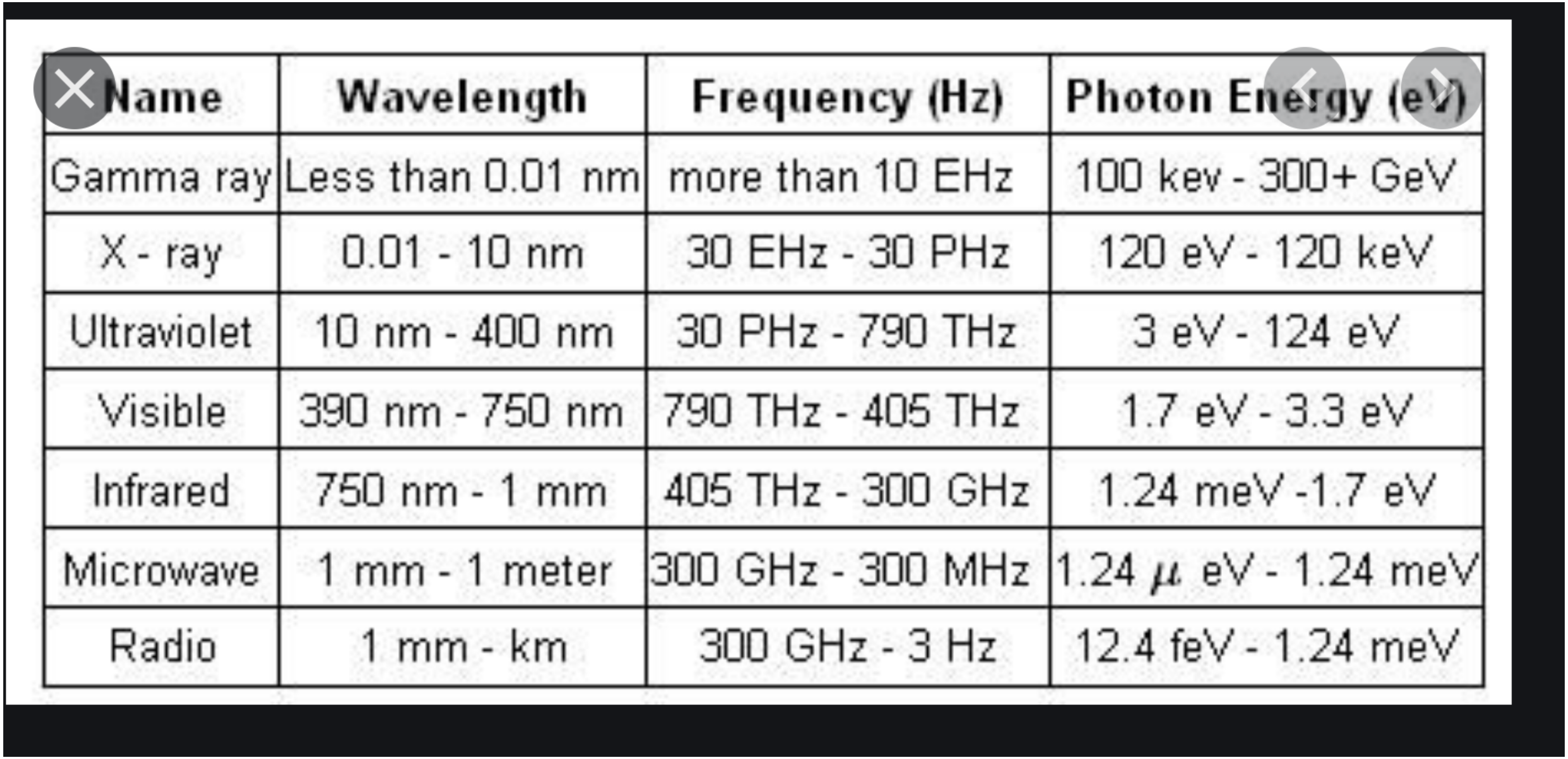

Name	Wavelength	Frequency (Hz)	Photon Energy (eV)
Gamma ray	Less than 0.01 nm	more than 10 EHz	100 kev - 300+ GeV
X - ray	0.01 - 10 nm	30 EHz - 30 PHz	120 eV - 120 keV
Ultraviolet	10 nm - 400 nm	30 PHz - 790 THz	3 eV - 124 eV
Visible	390 nm - 750 nm	790 THz - 405 THz	1.7 eV - 3.3 eV
Infrared	750 nm - 1 mm	405 THz - 300 GHz	1.24 meV -1.7 eV
Microwave	1 mm - 1 meter	300 GHz - 300 MHz	1.24 μ eV - 1.24 meV
Radio	1 mm - km	300 GHz - 3 Hz	12.4 feV - 1.24 meV

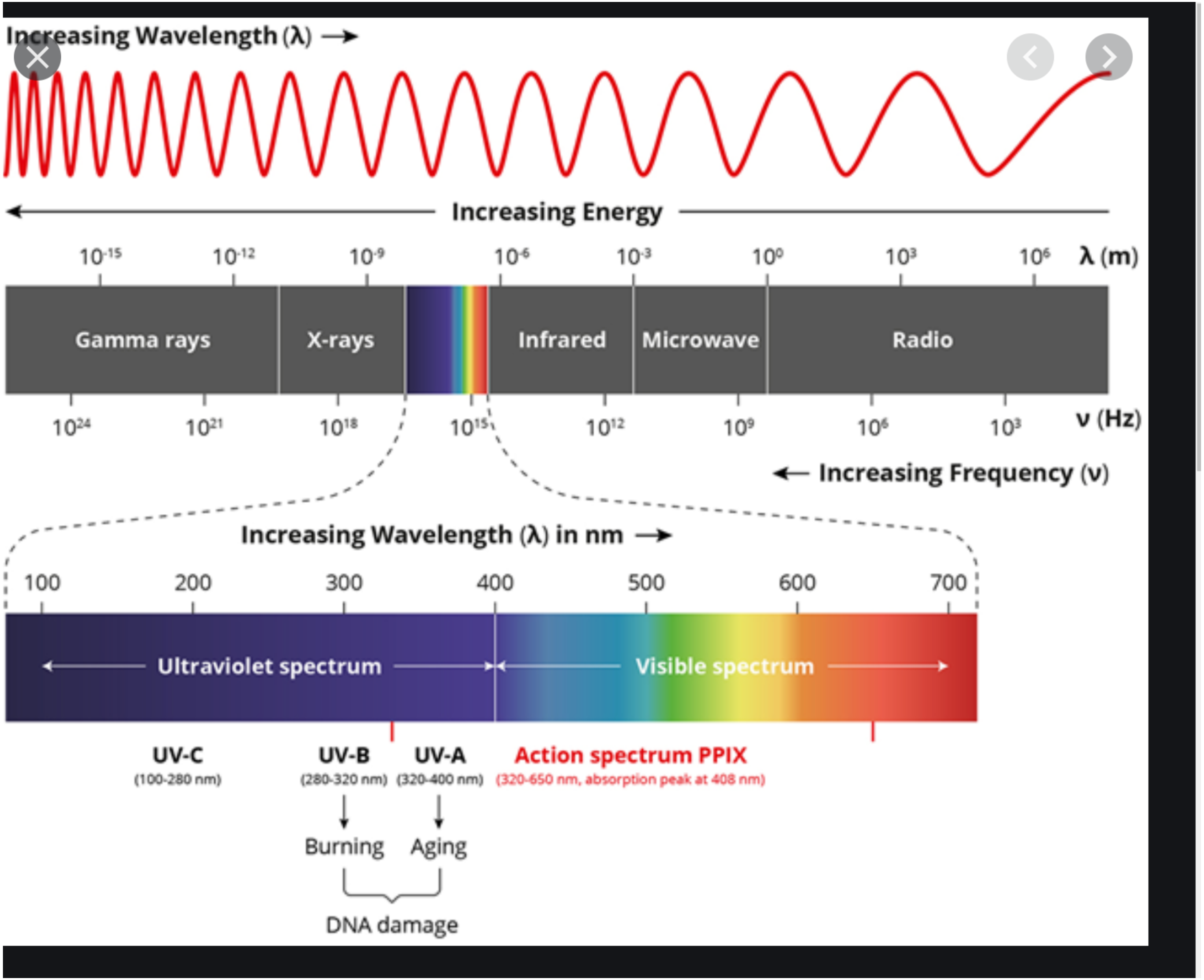
Increasing Wavelength (λ) →
Increasing Energy
10-15
10-12
10-9
10-6
10-3
10^0
10^3
10^6
λ (m)
Gamma rays
X-rays
Infrared
Microwave
Radio
10^24
10^21
10^18
10^15
10^12
10^9
10^6
10^3
ν (Hz)
← Increasing Frequency (ν)
Increasing Wavelength (λ) in nm →
100
200
300
400
500
600
700
Ultraviolet spectrum
Visible spectrum
UV-C
(100-280 nm)
UV-B
(280-320 nm)
UV-A
(320-400 nm)
Action spectrum PPIX
(320-650 nm, absorption peak at 408 nm)
Burning
Aging
DNA damage

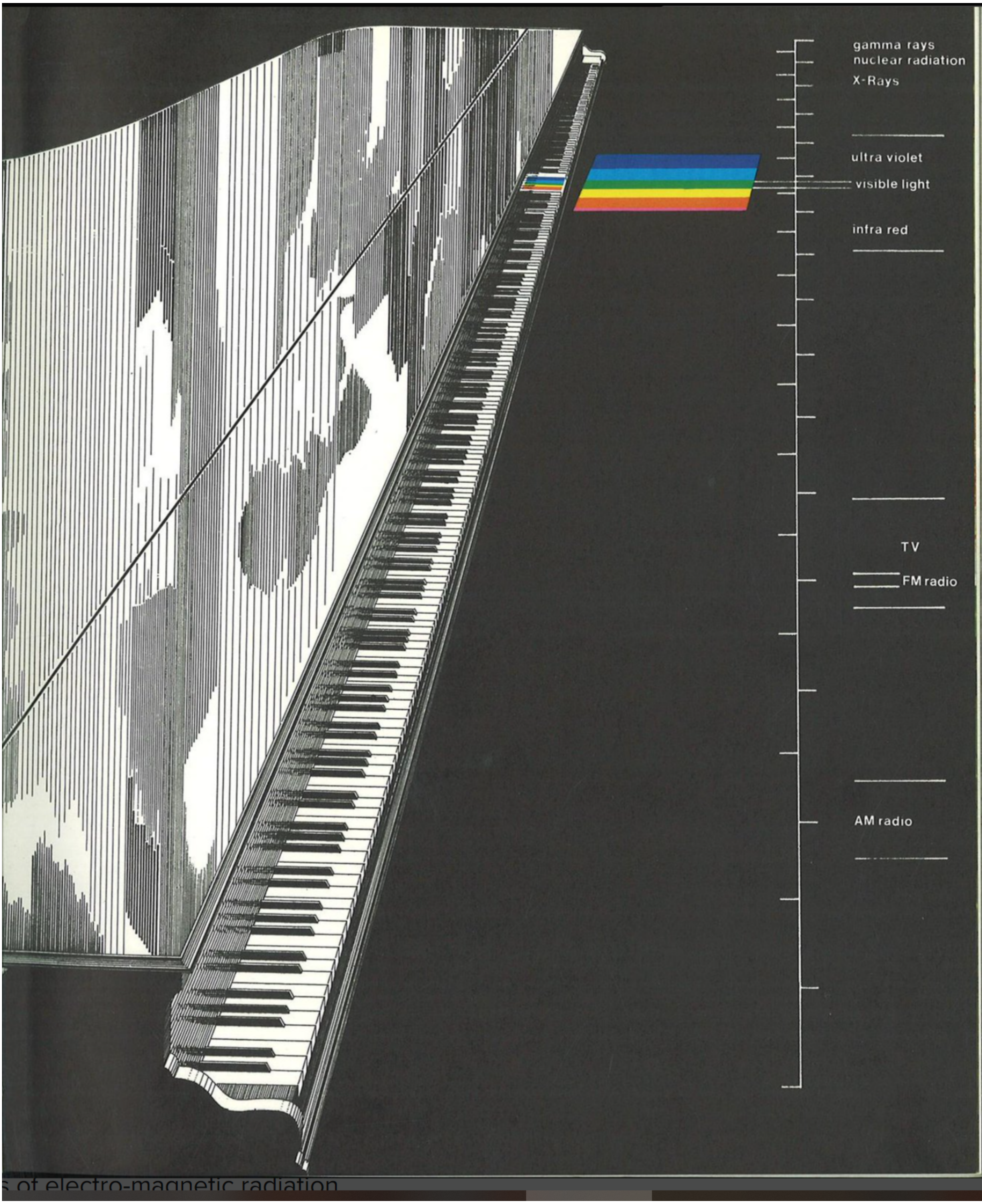
gamma rays
nuclear radiation
X-Rays
ultra violet
visible light
infra red
TV
FM radio
AM radio

s of electro-magnetic radiation

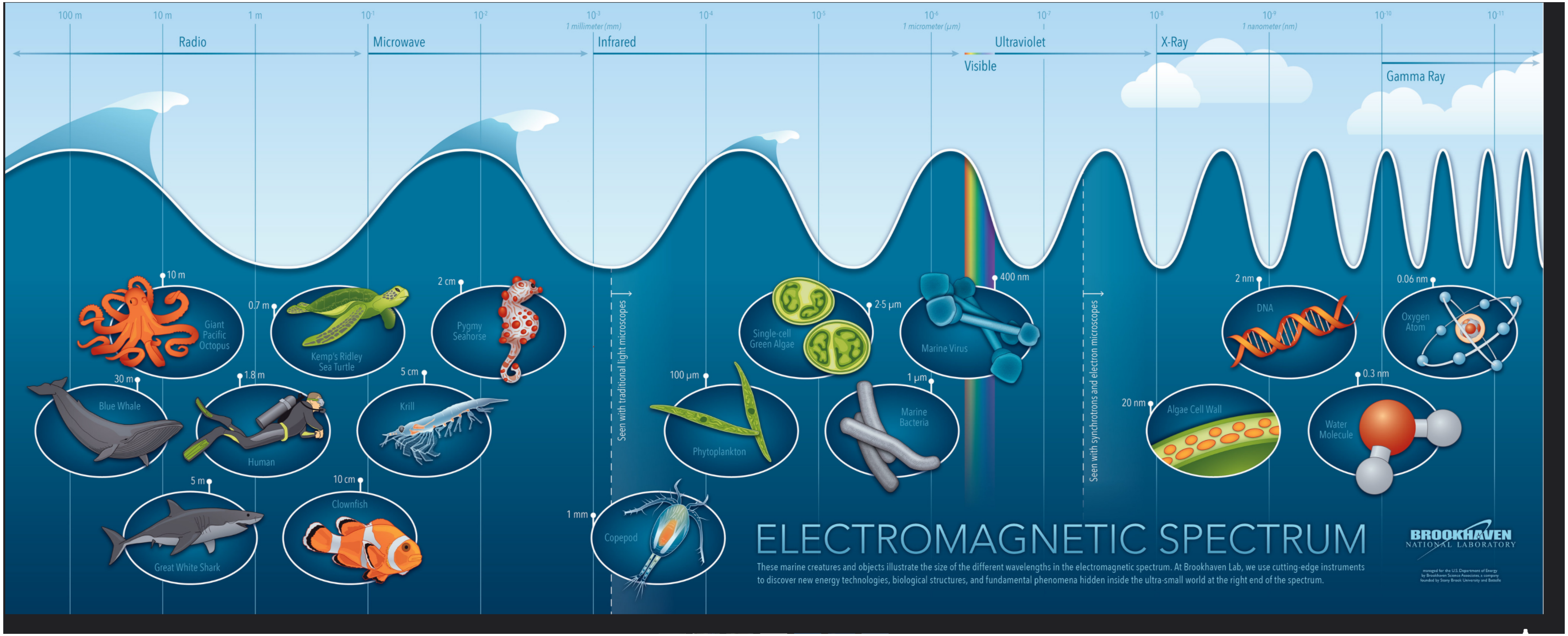
100 m
10 m
1 m
10^{-1}
10^{-2}
10^{-3}
1 millimeter (mm)
10^{-4}
10^{-5}
10^{-6}
1 micrometer (µm)
10^{-7}
10^{-8}
10^{-9}
1 nanometer (nm)
10^{-10}
10^{-11}
Radio
Microwave
Infrared
Visible
Ultraviolet
X-Ray
Gamma Ray
10 m
Giant Pacific Octopus
0.7 m
Kemp's Ridley Sea Turtle
2 cm
Pygmy Seahorse
30 m
Blue Whale
1.8 m
Human
5 cm
Krill
5 m
Great White Shark
10 cm
Clownfish
Seen with traditional light microscopes
1 mm
Copepod
100 µm
Phytoplankton
Single-cell Green Algae
2-5 µm
Marine Virus
400 nm
1 µm
Marine Bacteria
Seen with synchrotrons and electron microscopes
2 nm
DNA
20 nm
Algae Cell Wall
0.06 nm
Oxygen Atom
0.3 nm
Water Molecule

ELECTROMAGNETIC SPECTRUM

These marine creatures and objects illustrate the size of the different wavelengths in the electromagnetic spectrum. At Brookhaven Lab, we use cutting-edge instruments to discover new energy technologies, biological structures, and fundamental phenomena hidden inside the ultra-small world at the right end of the spectrum.

BROOKHAVEN
NATIONAL LABORATORY
managed for the U.S. Department of Energy
by Brookhaven Science Associates, a company
founded by Stony Brook University and Battelle

Microwaves

Microwaves are the highest-frequency electromagnetic waves that can be produced by currents in macroscopic circuits and devices. Microwave frequencies range from about 10^9Hz to nearly 10^{12}Hz. Their high frequencies correspond to short wavelengths compared with other radio waves—hence the name “microwave.” Microwaves also occur naturally as the cosmic background radiation left over from the origin of the universe. Along with other ranges of electromagnetic waves, they are part of the radiation that any object above absolute zero emits and absorbs because of **thermal agitation**, that is, from the thermal motion of its atoms and molecules.

Most satellite-transmitted information is carried on **microwaves**. **Radar** is a common application of microwaves. By detecting and timing microwave echoes, **radar** systems can determine the distance to objects as diverse as clouds, aircraft, or even the surface of Venus.

Microwaves of 2.45 GHz are commonly used in microwave ovens. The electrons in a water molecule tend to remain closer to the oxygen nucleus than the hydrogen nuclei (Figure). This creates two separated centers of equal and opposite charges, giving the molecule a dipole moment (see Electric Field). The oscillating electric field of the microwaves inside the oven exerts a torque that tends to align each molecule first in one direction and then in the other, with the motion of each molecule coupled to others around it. This pumps energy into the continual thermal motion of the water to heat the food. The plate under the food contains no water, and remains relatively unheated.

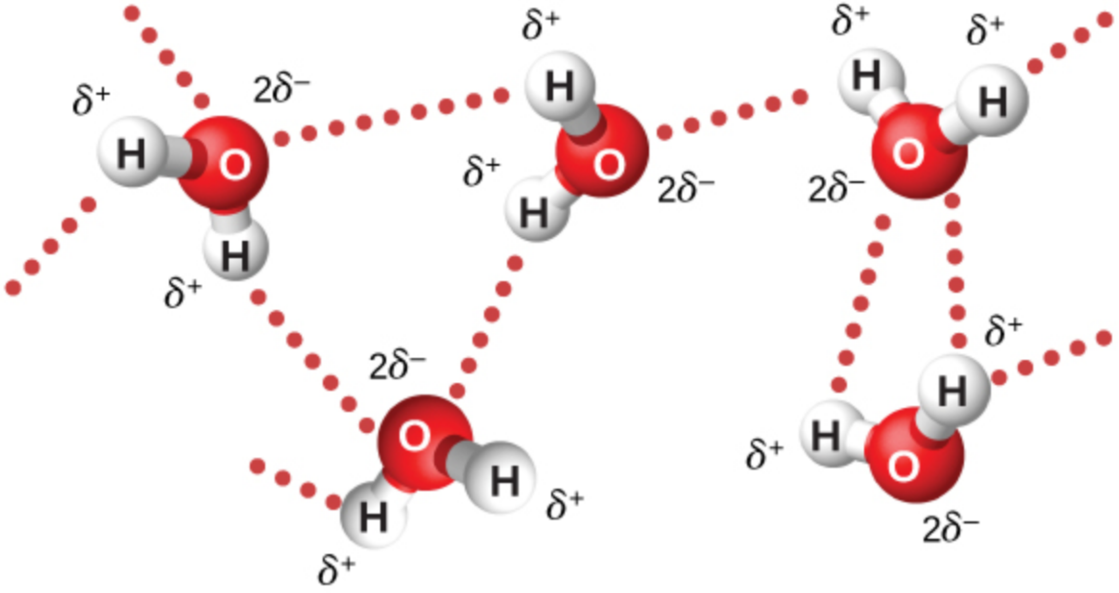

Figure 3. The oscillating electric field in a microwave oven exerts a torque on water molecules because of their dipole moment, and the torque reverses direction 4.90×10^9 times per second. Interactions between the molecules distributes the energy being pumped into them. The δ^+ and δ^- denote the charge distribution on the molecules.

Infrared Radiation

The boundary between the microwave and infrared regions of the electromagnetic spectrum is not well defined (see Figure). **Infrared radiation** is generally produced by thermal motion, and the vibration and rotation of atoms and molecules. Electronic transitions in atoms and molecules can also produce **infrared radiation**. About half of the solar energy arriving at Earth is in the infrared region, with most of the rest in the visible part of the spectrum. About 23% of the solar energy is absorbed in the atmosphere, about 48% is absorbed at Earth's surface, and about 29% is reflected back into space.[1]

The range of infrared frequencies extends up to the lower limit of visible light, just below red. In fact, infrared means "below red." Water molecules rotate and vibrate particularly well at infrared frequencies. Reconnaissance satellites can detect buildings, vehicles, and even individual humans by their infrared emissions, whose power radiation is proportional to the fourth power of the absolute temperature. More mundanely, we use infrared lamps, including those called *quartz heaters*, to preferentially warm us because we absorb infrared better than our surroundings.

The familiar handheld "remotes" for changing channels and settings on television sets often transmit their signal by modulating an infrared beam. If you try to use a TV remote without the infrared emitter being in direct line of sight with the infrared detector, you may find the television not responding. Some remotes use Bluetooth instead and reduce this annoyance.

Radio Waves

The term **radio waves** refers to electromagnetic radiation with wavelengths greater than about 0.1 m. Radio waves are commonly used for audio communications (i.e., for radios), but the term is used for electromagnetic waves in this range regardless of their application. Radio waves typically result from an alternating current in the wires of a broadcast antenna. They cover a very broad wavelength range and are divided into many subranges, including microwaves, electromagnetic waves used for AM and FM radio, cellular telephones, and TV signals.

There is no lowest frequency of radio waves, but ELF waves, or “extremely low frequency” are among the lowest frequencies commonly encountered, from 3 Hz to 3 kHz. The accelerating charge in the ac currents of electrical power lines produce electromagnetic waves in this range. ELF waves are able to penetrate sea water, which strongly absorbs electromagnetic waves of higher frequency, and therefore are useful for submarine communications.

In order to use an electromagnetic wave to transmit information, the amplitude, frequency, or phase of the wave is *modulated*, or varied in a controlled way that encodes the intended information into the wave. In AM radio transmission, the amplitude of the wave is modulated to mimic the vibrations of the sound being conveyed. Fourier’s theorem implies that the modulated AM wave amounts to a superposition of waves covering some narrow frequency range. Each AM station is assigned a specific carrier frequency that, by international agreement, is allowed to vary by $\pm 5\ \mathrm{kHz}$. In FM radio transmission, the frequency of the wave is modulated to carry this information, as illustrated in Figure, and the frequency of each station is allowed to use 100 kHz on each side of its carrier frequency. The electromagnetic wave produces a current in a receiving antenna, and the radio or television processes the signal to produce the sound and any image. The higher the frequency of the radio wave used to carry the data, the greater the detailed variation of the wave that can be carried by modulating it over each time unit, and the more data that can be transmitted per unit of time. The assigned frequencies for AM broadcasting are 540 to 1600 kHz, and for FM are 88 MHz to108 MHz.

Ultraviolet Radiation

Ultraviolet means “above violet.” The electromagnetic frequencies of **ultraviolet radiation (UV)** extend upward from violet, the highest-frequency visible light. The highest-frequency ultraviolet overlaps with the lowest-frequency X-rays. The wavelengths of ultraviolet extend from 400 nm down to about 10 nm at its highest frequencies. Ultraviolet is produced by atomic and molecular motions and electronic transitions.

UV radiation from the Sun is broadly subdivided into three wavelength ranges: UV-A (320–400 nm) is the lowest frequency, then UV-B (290–320 nm) and UV-C (220–290 nm). Most UV-B and all UV-C are absorbed by ozone (O_3) molecules in the upper atmosphere. Consequently, 99% of the solar UV radiation reaching Earth’s surface is UV-A.

Sunburn is caused by large exposures to UV-B and UV-C, and repeated exposure can increase the likelihood of skin cancer. The tanning response is a defense mechanism in which the body produces pigments in inert skin layers to reduce exposure of the living cells below.

As examined in a later chapter, the shorter the wavelength of light, the greater the energy change of an atom or molecule that absorbs the light in an electronic transition. This makes short-wavelength ultraviolet light damaging to living cells. It also explains why ultraviolet radiation is better able than visible light to cause some materials to glow, or *fluoresce*.

Besides the adverse effects of ultraviolet radiation, there are also benefits of exposure in nature and uses in technology. Vitamin D production in the skin results from exposure to UV-B radiation, generally from sunlight. Several studies suggest vitamin D deficiency is associated with the development of a range of cancers (prostate, breast, colon), as well as osteoporosis. Low-intensity ultraviolet has applications such as providing the energy to cause certain dyes to fluoresce and emit visible light, for example, in printed money to display hidden watermarks as counterfeit protection.

X-Rays

X-rays have wavelengths from about $10^{-8}\text{m to }10^{-12}\text{m}$. They have shorter wavelengths, and higher frequencies, than ultraviolet, so that the energy they transfer at an atomic level is greater. As a result, X-rays have adverse effects on living cells similar to those of ultraviolet radiation, but they are more penetrating. Cancer and genetic defects can be induced by X-rays. Because of their effect on rapidly dividing cells, X-rays can also be used to treat and even cure cancer.

The widest use of X-rays is for imaging objects that are opaque to visible light, such as the human body or aircraft parts. In humans, the risk of cell damage is weighed carefully against the benefit of the diagnostic information obtained.

Gamma Rays

Soon after nuclear radioactivity was first detected in 1896, it was found that at least three distinct types of radiation were being emitted, and these were designated as alpha, beta, and gamma rays. The most penetrating nuclear radiation, the **gamma ray (γ ray)**, was later found to be an extremely high-frequency electromagnetic wave.

The lower end of the γ- ray frequency range overlaps the upper end of the X-ray range. Gamma rays have characteristics identical to X-rays of the same frequency—they differ only in source. The name “gamma rays” is generally used for electromagnetic radiation emitted by a nucleus, while X-rays are generally produced by bombarding a target with energetic electrons in an X-ray tube. At higher frequencies, γ rays are more penetrating and more damaging to living tissue. They have many of the same uses as X-rays, including cancer therapy. Gamma radiation from radioactive materials is used in nuclear medicine.

Displacement current	$I_{\text{d}} = \varepsilon_0 \frac{d\Phi_{\text{E}}}{dt}$
Gauss's law	$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \frac{Q_{\text{in}}}{\varepsilon_0}$
Gauss's law for magnetism	$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} = 0$
Faraday's law	$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{s}} = -\frac{d\Phi_{\text{m}}}{dt}$
Ampère-Maxwell law	$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{s}} = \mu_0 I + \varepsilon_0 \mu_0 \frac{d\Phi_{\text{E}}}{dt}$
Wave equation for plane EM wave	$\frac{\partial^2 E_y}{\partial x^2} = \varepsilon_0 \mu_0 \frac{\partial^2 E_y}{\partial t^2}$
Speed of EM waves	$c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}$
Ratio of *E* field to *B* field in electromagnetic wave	$c = \frac{E}{B}$
Energy flux (Poynting) vector	$\vec{\mathbf{S}} = \frac{1}{\mu_0} \vec{\mathbf{E}} \times \vec{\mathbf{B}}$
Average intensity of an electromagnetic wave	$I = S_{\text{avg}} = \frac{c\varepsilon_0 E_0^2}{2} = \frac{cB_0^2}{2\mu_0} = \frac{E_0 B_0}{2\mu_0}$
Radiation pressure	$p = \begin{cases} I/c & \text{Perfect absorber} \\ 2I/c & \text{Perfect reflector} \end{cases}$

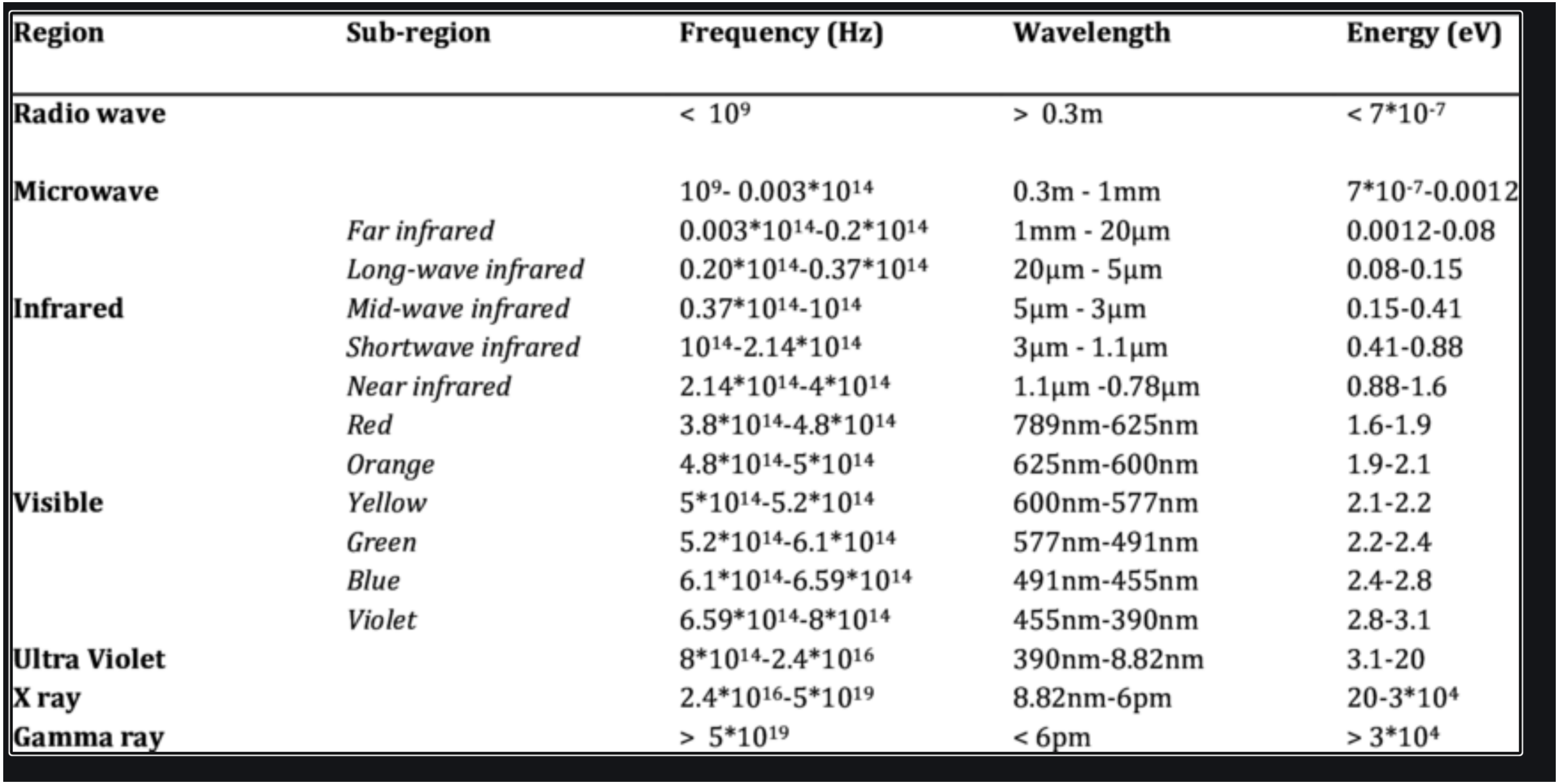

Region	Sub-region	Frequency (Hz)	Wavelength	Energy (eV)
Radio wave		$< 10^9$	> 0.3m	$< 7*10^{-7}$
Microwave		$10^9 - 0.003*10^{14}$	0.3m - 1mm	$7*10^{-7}-0.0012$
Infrared	*Far infrared*	$0.003*10^{14}-0.2*10^{14}$	1mm - 20µm	0.0012-0.08
	Long-wave infrared	$0.20*10^{14}-0.37*10^{14}$	20µm - 5µm	0.08-0.15
	Mid-wave infrared	$0.37*10^{14}-10^{14}$	5µm - 3µm	0.15-0.41
	Shortwave infrared	$10^{14}-2.14*10^{14}$	3µm - 1.1µm	0.41-0.88
	Near infrared	$2.14*10^{14}-4*10^{14}$	1.1µm -0.78µm	0.88-1.6
Visible	*Red*	$3.8*10^{14}-4.8*10^{14}$	789nm-625nm	1.6-1.9
	Orange	$4.8*10^{14}-5*10^{14}$	625nm-600nm	1.9-2.1
	Yellow	$5*10^{14}-5.2*10^{14}$	600nm-577nm	2.1-2.2
	Green	$5.2*10^{14}-6.1*10^{14}$	577nm-491nm	2.2-2.4
	Blue	$6.1*10^{14}-6.59*10^{14}$	491nm-455nm	2.4-2.8
	Violet	$6.59*10^{14}-8*10^{14}$	455nm-390nm	2.8-3.1
Ultra Violet		$8*10^{14}-2.4*10^{16}$	390nm-8.82nm	3.1-20
X ray		$2.4*10^{16}-5*10^{19}$	8.82nm-6pm	$20-3*10^4$
Gamma ray		$> 5*10^{19}$	< 6pm	$> 3*10^4$

Are radio waves completely different physical objects than gamma-rays? They are produced in different processes and are detected in different ways, but they are not fundamentally different. Radio waves, gamma-rays, visible light, and all the other parts of the electromagnetic spectrum are electromagnetic radiation.

Electromagnetic radiation can be described in terms of a stream of mass-less particles, called photons, each traveling in a wave-like pattern at the speed of light. Each photon contains a certain amount of energy. The different types of radiation are defined by the the amount of energy found in the photons. Radio waves have photons with low energies, microwave photons have a little more energy than radio waves, infrared photons have still more, then visible, ultraviolet, X-rays, and, the most energetic of all, gamma-rays.

Measuring electromagnetic radiation

Electromagnetic radiation can be expressed in terms of energy, wavelength, or frequency. Frequency is measured in cycles per second, or Hertz. Wavelength is measured in meters. Energy is measured in electron volts. Each of these three quantities for describing EM radiation are related to each other in a precise mathematical way. But why have three ways of describing things, each with a different set of physical units?

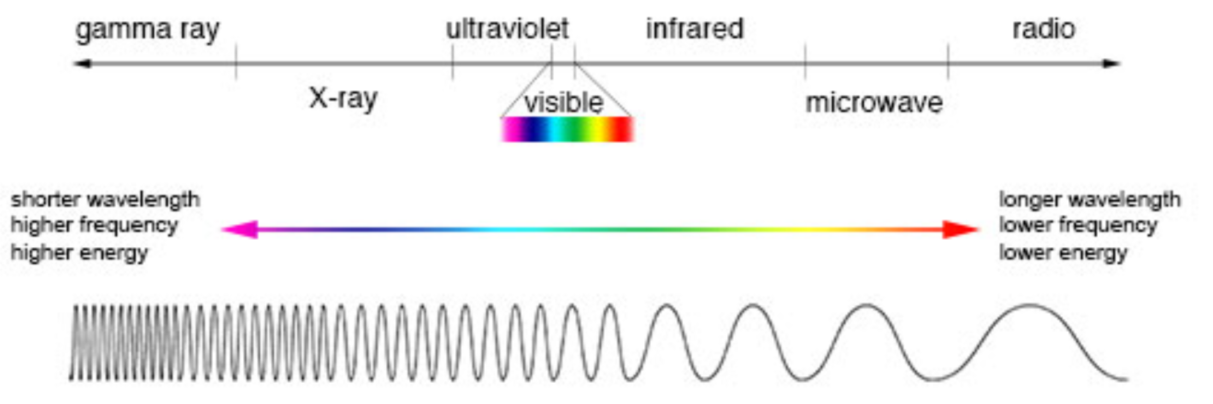
gamma ray
ultraviolet
infrared
radio
X-ray
visible
microwave
shorter wavelength
higher frequency
higher energy
longer wavelength
lower frequency
lower energy

Comparison of wavelength, frequency and energy for the electromagnetic spectrum. (Credit: NASA's Imagine the Universe)

The short answer is that scientists don't like to use numbers any bigger or smaller than they have to. It is much easier to say or write "two kilometers" than "two thousand meters." Generally, scientists use whatever units are easiest for the type of EM radiation they work with.

Astronomers who study radio waves tend to use wavelengths or frequencies. Most of the radio part of the EM spectrum falls in the range from about 1 cm to 1 km, which is 30 gigahertz (GHz) to 300 kilohertz (kHz) in frequencies. The radio is a very broad part of the EM spectrum.

Infrared and optical astronomers generally use wavelength. Infrared astronomers use microns (millionths of a meter) for wavelengths, so their part of the EM spectrum falls in the range of 1 to 100 microns. Optical astronomers use both angstroms (0.00000001 cm, or 10^{-8} cm) and nanometers (0.0000001 cm, or 10^{-7} cm). Using nanometers, violet, blue, green, yellow, orange, and red light have wavelengths between 400 and 700 nanometers. (This range is just a tiny part of the entire EM spectrum, so the light our eyes can see is just a little fraction of all the EM radiation around us.)

chapter one

Why Did Benjamin Franklin Fly the Kite?

Shuffle across a nylon rug on a dry winter day and your body acquires an excess electrical charge. That charge may be violently released by a spark jumping from your finger tip to a light switch or doorknob. Lightning is nothing more than a very long spark which discharges regions of excess electrical charge developed in thunderclouds. Frictional charging, that is charging by rubbing together certain dissimilar materials such as a shoe sole and the rug, has probably always been familiar to man. In 1746 Benjamin Franklin began his experiments in electricity. His experiments were made possible by (1) the frictional charging mechanism and (2) the fortuitous invention earlier that year of the Leyden jar – a primitive capacitor to store electrical charge.

Prior to Franklin's interest with electricity, a number of scientists had suggested that lightning might be an electrical phenomenon. Their concern, however, had gone no further than suggestion. In November 1749 Franklin wrote the following about the sparks (in his terminology, electrical fluid) he had studied.

> Electrical fluid agrees with lightning in these particulars. 1. Giving light. 2. Colour of the light. 3. Crooked direction. 4. Swift motion. 5. Being conducted by metals. 6. Crack or noise in exploding. 7. Subsisting in water or ice. 8. Rending bodies as it passes through. 9. Destroying animals. 10. Melting metals. 11. Firing inflammable substances. 12. Sulphureous smell. The electrical fluid is attracted by points. We do not know whether this property is in lightning. But since they agree in all particulars wherein we can already compare them, is it not possible they agree likewise in this? Let the experiment be made. (Ref. 1.1)

Franklin was the first to design an experiment to prove that lightning was electrical. In July 1750 he wrote:

> To determine the question whether the clouds that contain lightning are electrified or not, I would propose an experiment to be tried where it may be done conveniently. On the top of some high tower or steeple place a kind of sentry box . . . big enough to contain a man and an electrical stand [an insulator]. From the middle of the stand let an iron rod rise and pass bending out of the door, and then upright twenty or thirty feet, pointed very sharp at the end. If the electrical stand be kept clean and dry, a man standing on it when such clouds are passing low might be electrified and afford sparks, the rod drawing fire to him from the cloud. If any danger to the man should be apprehended (though I think there would be none), let him stand on the floor of his box and now and then bring near to the rod the loop of a wire that has one end fastened to the leads, he holding it by a wax handle; so the sparks, if the rod is electrified, will strike from the rod to the wire and not affect him. (Ref. 1.2)

His experiment and the results he expected to achieve are illustrated in Fig. 1.1. The aim was to show that the clouds were electrically charged, for if this were the case, it followed that lightning was also electrical. Franklin did not appreciate

the danger involved in his experiment. If the iron rod were directly struck by lightning, the experimenter would almost certainly be killed. Such was eventually to be the case as we shall see.

In France in May 1752 Thomas-Francois D'Alibard successfully performed Franklin's suggested experiment. Sparks were observed to jump from the iron rod during a thunderstorm. It was proved that thunderclouds contain electrical charge. Soon after, the experiment was successfully repeated in France again, in England, and in Belgium. In July 1753 G. W. Richmann, a Swedish physicist working in

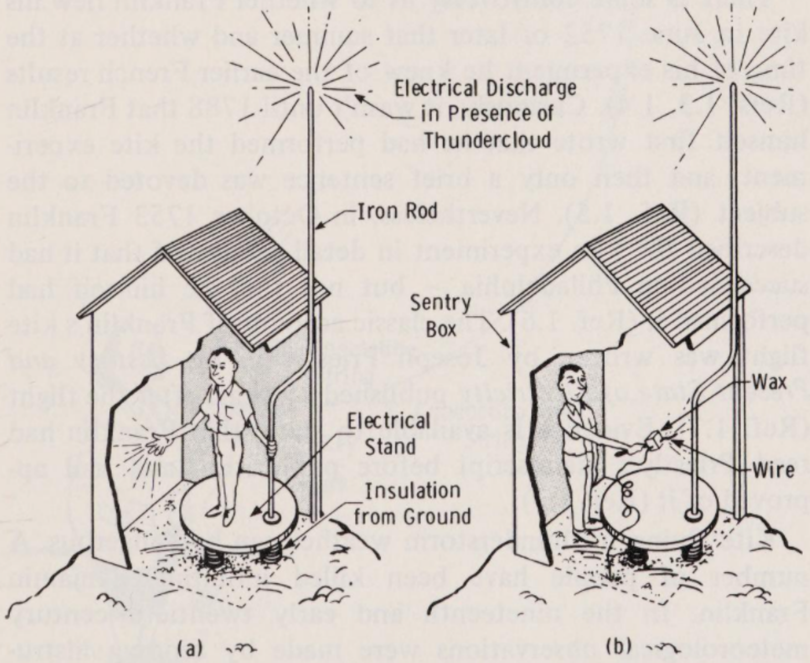

Fig. 1.1. Franklin's original experiment to show that thunderclouds are electrified. (a) Man on electrical stand holds iron rod with one hand and obtains an electrical discharge between the other hand and ground. (b) Man on ground draws sparks between iron rod and a grounded wire held by an insulating wax handle.

Russia, put up an experimental rod and was killed by a direct lightning strike.

Before Franklin himself got around to performing the experiment, he thought of a better way of proving his theory – an electrical kite. It was to take the place of the iron rod, since it could reach a greater elevation than the rod and could be flown anywhere. During a thunderstorm in 1752 Franklin flew the most famous kite in history. Sparks jumped from a key tied to the bottom of the kite string to the knuckles of his hand (Fig. 1.2). He had verified his theory, and had probably done so before he knew that D'Alibard had already obtained the same proof.

There is some controversy as to whether Franklin flew his kite in June 1752 or later that summer and whether at the time of his experiment he knew of the earlier French results (Refs. 1.3, 1.4). Curiously, it wasn't until 1788 that Franklin himself first wrote that he had performed the kite experiment; and then only a brief sentence was devoted to the subject (Ref. 1.5). Nevertheless, in October 1753 Franklin described the kite experiment in detail and stated that it had succeeded in Philadelphia – but not that he himself had performed it (Ref. 1.6). The classic account of Franklin's kite flight was written by Joseph Priestly in his *History and Present State of Electricity* published 15 years after the flight (Ref. 1.7). Evidence is available to show that Franklin had read Priestly's manuscript before publication and had approved of it (Ref. 1.3).

Kite flying in thunderstorm weather can be dangerous. A number of people have been killed imitating Benjamin Franklin. In the nineteenth and early twentieth century meteorological observations were made by sending instruments aloft on large box kites. In 1898, 17 U.S. Weather Bureau stations were equipped for daily kite ascensions. The kites used weighed 8 lb, carried 2 lb of instruments, and dragged as much as 20 or 30 lb of piano wire beneath them.

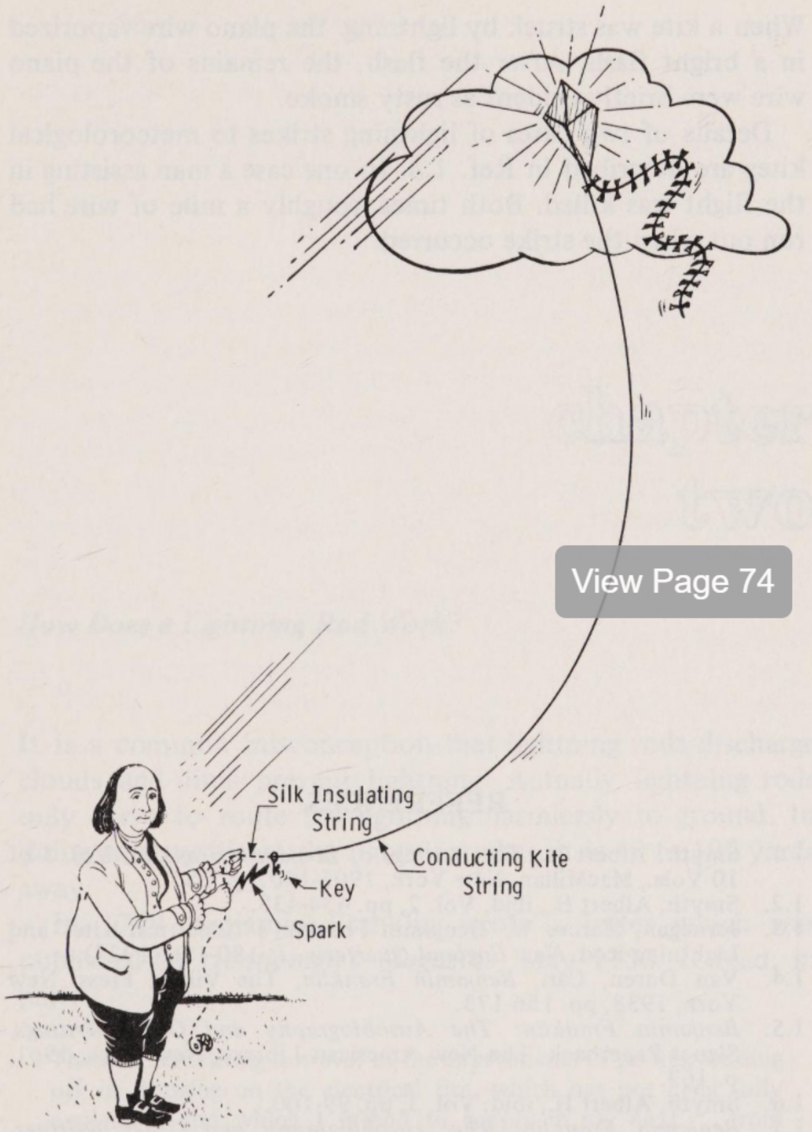

Fig. 1.2. Franklin's electrical kite experiment: Sparks jump from the electrified key at the end of the electrified kite string to Franklin's hand.

When a kite was struck by lightning the piano wire vaporized in a bright flash. After the flash, the remains of the piano wire were briefly evident as rusty smoke.

Details of two cases of lightning strikes to meteorological kites are described in Ref. 1.8. In one case a man assisting in the flight was killed. Both times, roughly a mile of wire had run out when the strike occurred.

REFERENCES

1.1. Smyth, Albert H., *The Writings of Benjamin Franklin*, Vol. 3 of 10 Vols., MacMillan, New York, 1905-1907, p. 255.
1.2. Smyth, Albert H., ibid, Vol. 2, pp. 434-438.
1.3. Jernegan, Marcus W., Benjamin Franklin's "Electrical Kite" and Lightning Rod, *New England Quarterly, 1*, 180-196 (1928).
1.4. Van Doren, Carl, *Benjamin Franklin*, The Viking Press, New York, 1938, pp. 156-173.
1.5. *Benjamin Franklin: The Autobiography and Other Writings*, Signet Paperback, The New American Library, New York, 1961, p. 167.
1.6. Smyth, Albert H., ibid, Vol. 3, pp. 99-100.
1.7. *Benjamin Franklin: The Autobiography and Other Writings*, Signet Paperback, The New American Library, New York, 1961, pp. 234-235 gives Priestly's account of Franklin's kite flight.
1.8. Brunk, I. W., Ben Franklin Was Lucky, *Weatherwise, 11*, 92-93, June 1958; Lightning Death During Kite Flight (or Radiosondes are Safer), *Weatherwise, 11*, 204-205, December 1958.

chapter two

How Does a Lightning Rod Work?

It is a common misconception that lightning rods discharge clouds and thus prevent lightning. Actually lightning rods only serve to route the lightning harmlessly to ground. In doing so they divert the lightning when it is 10 to 100 yards away.

In 1749 Benjamin Franklin wrote a letter which was published in *Gentlemen's Magazine*, May 1750. It read, in part,

> There is something however in the experiments of points, sending off or drawing on the electrical fire, which has not been fully explained, and which I intend to supply in my next . . . from what I have observed on experiments, I am of opinion that houses, ships, and even towers and churches may be eventually secured from the strokes of lightning by their means; for if instead of the round balls of wood or metal which are commonly

> placed on the tops of weathercocks, vanes, or spindles of churches, spires, or masts, there should be a rod of iron eight or ten feet in length, sharpened gradually to a point like a needle, and gilt to prevent rusting, or divided into a number of points, which would be better, the electrical fire would, I think, be drawn out of a cloud silently, before it could come near enough to strike.

This is Franklin's earliest recorded suggestion of the lightning rod. In the "experiments of points" he placed electrical charge on isolated conductors and then showed that the charge could be drained away (discharged) slowly and silently if a pointed and grounded (attached to ground) conductor were introduced into the vicinity. When the pointed conductor was brought too close to the charged conductor, the discharge occurred violently via an electric spark.

In the discussion in which he proposed the original experiment to determine if lightning were electrical (July 1750 – see Chapter 1), Franklin repeated his suggestion for protective lightning rods, adding that they should be grounded (i.e., that a wire should connect the lightning rod to the ground or, in the case of a ship, to the water).

Lightning rods were apparently first used for protective purposes in 1752 in France and later the same year in the United States (Refs. 2.1, 2.2).

Franklin originally thought – erroneously – that the lightning rod silently discharged the electric charge in a thundercloud and thereby prevented lightning. However, in 1755 he stated:

> I have mentioned in several of my letters, and except once, always in the *alternative, viz.,* that pointed rods erected on buildings, and communicating with the moist earth, would *either* prevent a stroke, or, if not prevented, would *conduct* it, so that the building should suffer no damage. (Ref. 2.3)

It is in the latter manner that lightning rods actually work. The charge flowing between a lightning rod and a thundercloud is much too small to discharge the thundercloud (Ref. 2.4). The rod diverts to itself a stroke on its way to earth but can do so only in the final part of the stroke's earthward trajectory. Diversion is achieved by the initiation of an electrical discharge (Fig. 2.1), a sort of traveling spark, which propagates from the rod, intercepts the downward-moving lightning, and provides a conducting path to the rod. Before the traveling spark is initiated, the downward-moving lightning is essentially uninfluenced by objects on the ground beneath it. The traveling spark is generally 10 to 100 yards long when it meets the lightning.

Any high object may initiate an upward-moving spark which attempts to reach the downward-moving lightning. It is therefore important that the lightning rod be the tallest object near the structure it protects, so that its traveling spark catches the lightning rather than a spark initiated by the chimney or a nearby tree.

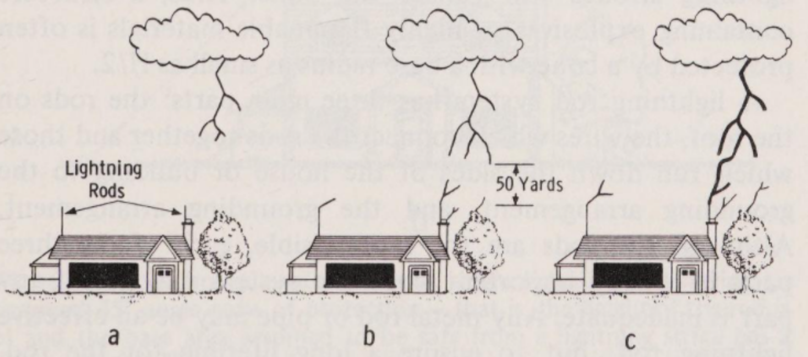

Fig. 2.1. Final stages of a strike to a lightning rod. The time interval between B and A and between C and B is about 1/5000 second. The lightning channel becomes very bright (view C) immediately after the upward-traveling spark connects with the downward-moving lightning. Drawings are not to scale.

Franklin refused to patent the lightning rod or otherwise to profit by its invention.

No lightning rod, however tall, can offer absolute protection; lightning has struck the Empire State Building 50 ft below the top (see Chapter 6). Nevertheless, lightning rod systems are effective if used properly, and many codes have been written to describe their use (Ref. 2.5). A single vertical rod will almost always protect objects within an imaginary cone (the "cone of protection") formed by all lines connecting the top of the rod at height H with a circle on the ground, beneath the rod, of radius between H and 2H. In Fig. 2.2a, the house has a single lightning rod with a cone of protection assumed to have a base radius H in accord with the British lightning code. To protect a large house it is more practical to use multiple lightning rods: in Fig. 2.2b, three cones of protection overlap to provide a large volume of protection without excessive height. The house in Fig. 2.1 is protected according to the U.S. lightning code which specifies a base radius of 2H for the cone of protection. The smaller the base radius, the greater the probability that no lightning strokes will violate the cone. Thus, a structure containing explosives or highly flammable materials is often protected by a cone with a base radius as small as H/2.

A lightning rod system has three main parts: the rods on the roof, the wires which connect the rods together and those which run down the sides of the house or building to the grounding arrangement, and the grounding arrangement. Although the rods are the most visible, each of the three parts is equally important since the system may fail if any part is inadequate. Any metal rod or pipe may be an effective lightning rod, but to ensure a long lifetime for the rod, corrosion-resistant metal such as copper, aluminum, or galvanized iron should be used. There is no evidence that a pointed rod is better than one with a ball on the top. A wide variety of lightning rod shapes can be seen on urban and rural structures.

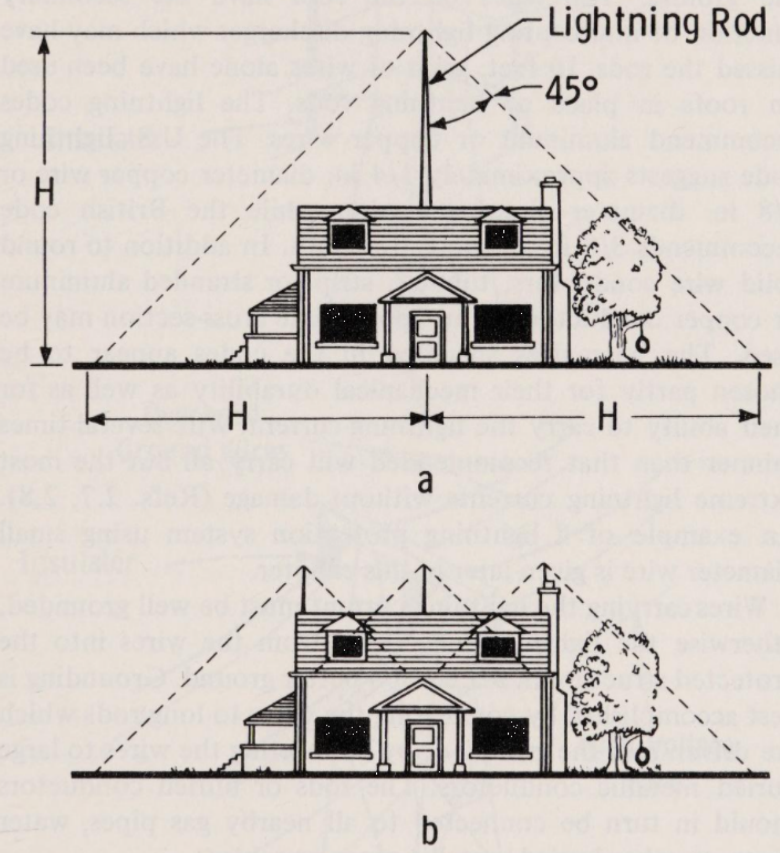

Fig. 2.2. (a) A house protected by a single lightning rod having an assumed 45°-angle cone of protection – that is, the height of the rod is H and the base area assumed to be safe from a lightning strike has a radius H. (b) The same house protected by a lightning rod system consisting of three smaller rods, each assumed to provide a 45°-angle cone of protection.

The primary function of the wires which link the lightning rods and those connecting the rods to the grounding arrangement is to carry the lightning current from the rods to the ground. The wires on the roof have the secondary function of intercepting lightning discharges which may have missed the rods. In fact, grids of wires alone have been used on roofs in place of lightning rods. The lightning codes recommend aluminum or copper wires. The U.S. lightning code suggests approximately 1/4 in. diameter copper wire or 3/8 in. diameter aluminum wire, while the British code recommends 3/8 in. diameter for both. In addition to round solid wire conductors, tubular, strip, or stranded aluminum or copper conductors of an appropriate cross-section may be used. The wire sizes specified in the codes appear to be chosen partly for their mechanical durability as well as for their ability to carry the lightning current. Wire several times thinner than that recommended will carry all but the most extreme lightning currents without damage (Refs. 2.7, 2.8). An example of a lightning protection system using small diameter wire is given later in this chapter.

Wires carrying the lightning current must be well grounded, otherwise the lightning may jump from the wires into the protected structure in search of a better ground. Grounding is best accomplished by connecting the wires to long rods which are driven into the ground or by connecting the wires to large buried metallic conductors. The rods or buried conductors should in turn be connected to all nearby gas pipes, water pipes, or other buried metallic pipes or cables.

It is sometimes imperative to keep the lightning current and possible attendant sparks from contacting any part of a protected structure – a typical case being a liquid-fuel storage vat in which flammable vapors are present. Here, the roof rods and wire conductors are often replaced by a system of wires suspended between tall towers arranged around the structure. A similar scheme is used to protect high voltage

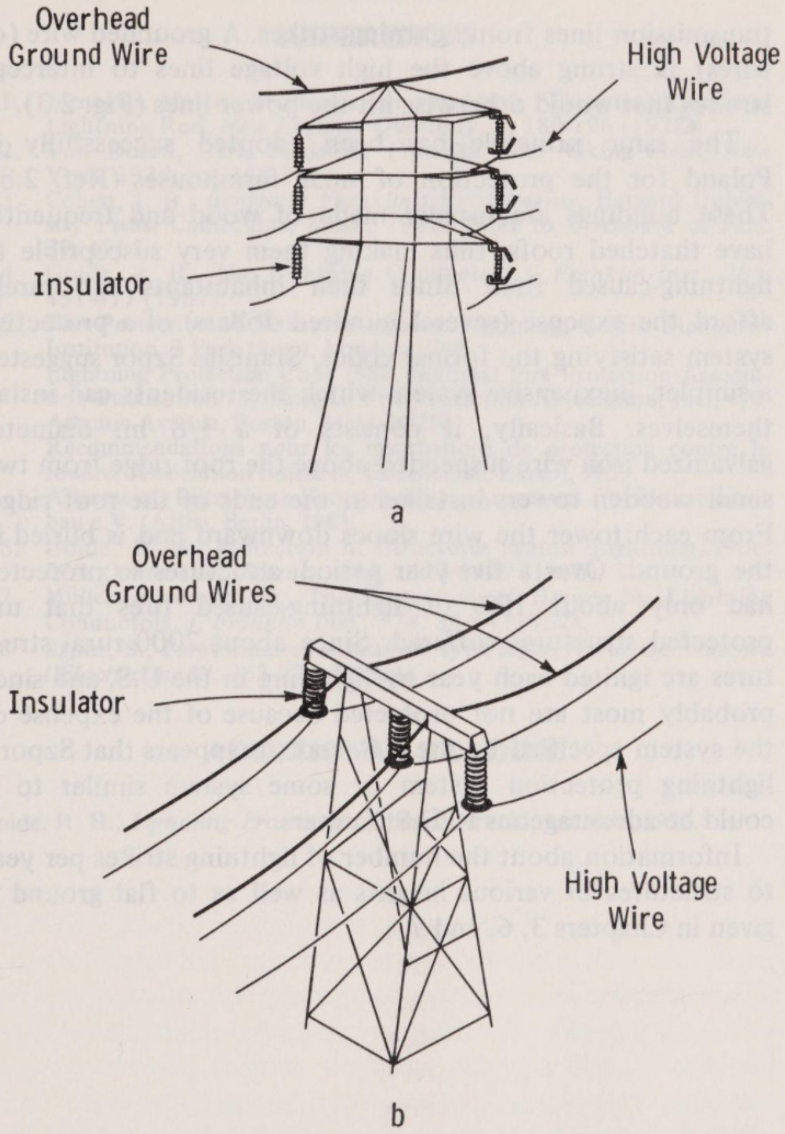

Fig. 2.3. Use of overhead ground wires to protect high voltage transmission lines from lightning. Each of the high voltage wires is supported by the bottom of an insulator. Ground wires are attached to the metal transmission line towers which are well grounded. Drawings are not to scale. (a) A typical 132,000 volt line. (b) A typical 220,000 volt line.

transmission lines from lightning strikes. A grounded wire (or wires) is strung above the high voltage lines to intercept strokes that would otherwise hit the power lines (Fig. 2.3).

The same principle has been adopted successfully in Poland for the protection of small farmhouses (Ref. 2.8). These buildings are usually made of wood and frequently have thatched roofs, thus making them very susceptible to lightning-caused fires. Since their inhabitants can rarely afford the expense (several hundred dollars) of a protective system satisfying the formal codes, Stanislas Szpor suggested a simpler, inexpensive system which the residents can install themselves. Basically, it consists of a 1/8 in. diameter galvanized iron wire suspended above the roof ridge from two small wooden towers installed at the ends of the roof ridge. From each tower the wire slopes downward and is buried in the ground. Over a five-year period, structures so protected had only about 10% of lightning-caused fires that unprotected structures suffered. Since about 2000 rural structures are ignited each year by lightning in the U.S. and since probably most are not protected because of the expense of the system specified by the U.S. code, it appears that Szpor's lightning protection system or some system similar to it could be advantageous to U.S. farmers.

Information about the number of lightning strikes per year to structures of various heights as well as to flat ground is given in Chapters 3, 6, and 7.

chapter three

How Many People Are Killed By Lightning Each Year?

What Should I Do If Caught Outdoors In A Thunderstorm?

Is It Safe To Talk On The Telephone During A Thunderstorm?

Should I Unplug Radios And TVs?

How Can I Help Someone Struck By Lightning?

Although no exact figures on lightning fatalities are available – there is no central agency to which lightning accidents must be reported – various studies (Refs. 3.1, 3.2, 3.3, 3.4) have placed the average annual number of deaths in the U.S. at 100, 137, 151, 300, and even 600. Whatever the exact number, lightning causes more direct deaths than any other weather phenomena. Snowstorms cause more indirect deaths (e.g., heart attacks from shoveling the driveway). Tornadoes cause about half as many deaths as lightning but inflict much more property damage. It is perhaps surprising that about two-thirds of the people involved in lightning accidents subsequently make a complete recovery. Most, if not all, of these survivors are probably not directly struck by lightning, but rather receive electrical shock from being in the close vicinity of a strike.

About 70% of all fatal lightning accidents involve one person, about 15% occur in groups of two, the remainder involves three or more deaths per lightning event. Between 75 and 85% of all lightning deaths and injuries are to men, probably because they are more likely to be out-of-doors. About 70% of all injuries and fatalities occur in the afternoon, about 20% between 6 p.m. and midnight (Ref. 3.1).

The largest single category of lightning deaths (12 to 15% of all fatalities) is composed of those unfortunate individuals who seek refuge under trees during thunderstorms and have their sheltering trees struck by lightning. Perhaps one-third of these are golfers. Two additional categories in which large numbers of lightning fatalities occur are designated "Open Water" and "Tractors", each claiming about 10% of the total fatalities. "Open Water" includes people fishing, swimming, on beaches, piers, and levees, in boats, and on water skis; "Tractors" includes people on, in, or near farm tractors or other implements, construction equipment, cars, and trucks (Refs. 3.1, 3.3).

Most lightning deaths occur outdoors. Over 30% of all lightning deaths involve people who work outdoors; over 25% of all deaths involve outdoor recreationists. A breakdown of deaths and injuries to farmers and ranchers and to outdoor recreationists for the 20-year period 1950 through 1969 is given in Table 3.1. The apparent increase from the 1950s to the 1960s in the number and percentage of deaths to outdoor recreationists is perhaps due to the increasing number of outdoor recreationists. Table 3.2 gives more detailed data on death and injury to outdoor recreationists, while Table 3.3 breaks down the "Open Water" category of Table 3.2.

In addition to humans, animals are killed by lightning. In 1968, lightning caused the death of 464 cattle (362 killed directly, the remainder by lightning-caused fires), 13 horses (11 killed directly), 42 hogs, 2 dogs, and 1 white ox resident in a museum pasture (Ref. 3.4).

flow in the metal skin of the car and not in the occupant. The lightning will ground itself by jumping from the car to the earth either through the air, along the surface of a tire, or through a tire. In the latter case, the tire will be destroyed. For maximum security, the car windows should be rolled up and wet (so that they conduct current along their surfaces), and the occupant should not touch any metal part of the car or the car radio.

Because modern commercial aircraft are essentially all metal, lightning currents seldom penetrate them. An airplane *on the ground* is a relatively safe refuge during a thunderstorm. What about planes in the air? Are aircraft in flight hit by lightning? If hit, do they continue to fly?

Airplanes in flight *are* struck by lightning. Between January 1965 and December 1966, the Federal Aviation Administration (FAA) collected reports of about 1000 lightning strikes to commercial aircraft. On the average, a given commercial airplane is struck by lightning once in every 5000 to 10,000 hours of flying time (Refs. 4.1, 4.2). As extreme examples of lightning activity, on one day in January 1969 four airliners were hit by lightning in separate incidents near Los Angeles (Ref. 4.3), and on September 26, 1964 a Boeing 727 in a holding pattern at Chicago was hit by lightning five separate times within a 20-minute period (Ref. 4.4).

Aircraft struck by lightning almost always continue to fly. Generally, lightning leaves pit marks or burn marks on the aircraft's metallic skin or burn or puncture holes through it. The FAA reports hole diameters up to 4 in., a common size being 1/2 in. (Ref. 4.1). Photographs of a lightning-damaged aircraft wing are shown in Figs. 4.1, 4.2, and 4.3. In addition to the generally minor damage caused to metallic aircraft parts, unprotected non-metallic parts may be more severely damaged. (Differences between lightning damage to metallic and non-metallic materials are discussed in the next chapter.)

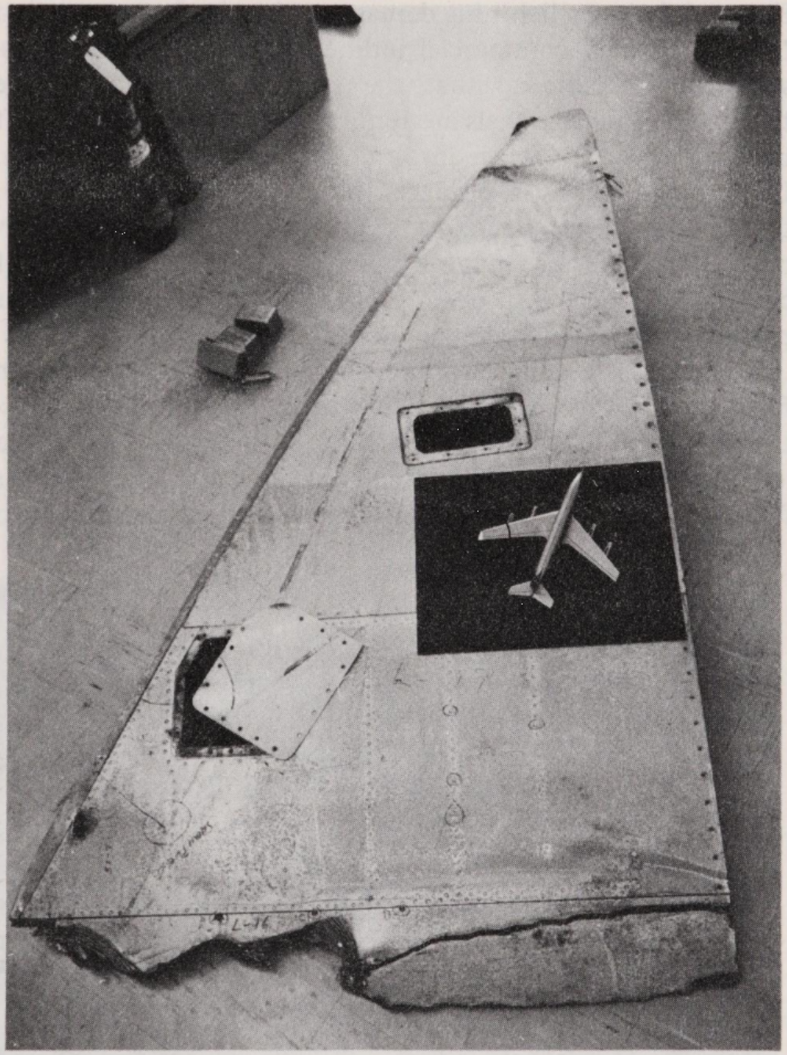

Fig. 4.1. Left wing tip of a Boeing 707 which exploded after being hit by lightning near Elkton, Maryland on December 8, 1963. On the model insert the white portion of the wing gives the size and orientation of the portion of the actual wing shown. A number of lightning-caused holes and considerable pitting are evident. (Courtesy, Bernard Vonnegut and Roger Cheng, State University of New York at Albany)

Figure 4.4 shows lightning damage to the plastic nose section or radome of a commercial jetliner. (The radome is necessarily non-metallic so that the plane's radar can operate through it – radar signals are reflected by metal.) The aircraft in Fig. 4.4 had its radar unit put out of commission. Lightning is clearly a potential hazard. On the other hand, thunderstorm turbulence is a greater hazard. Severe turbulence can cause structural failure of the aircraft or loss of control by the pilot. The problem of loss of control is particularly critical during take-off and landing.

Lightning is a danger to aircraft for several reasons. It can fuse, burn away, or otherwise destroy mechanical or electrical parts; it can temporarily blind the pilot causing him to lose control of the aircraft; and it can ignite the aircraft's fuel. Fuel ignition was the probable cause of the only two

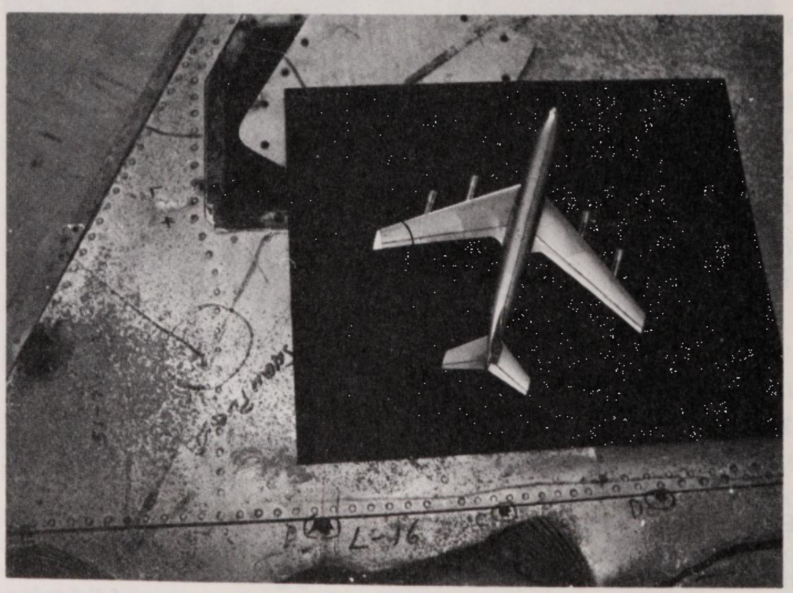

Fig. 4.2. Closer view of the Boeing 707 wing tip shown in Figure 4.1. Five lightning-caused holes are visible. (Courtesy, Bernard Vonnegut and Roger Cheng)

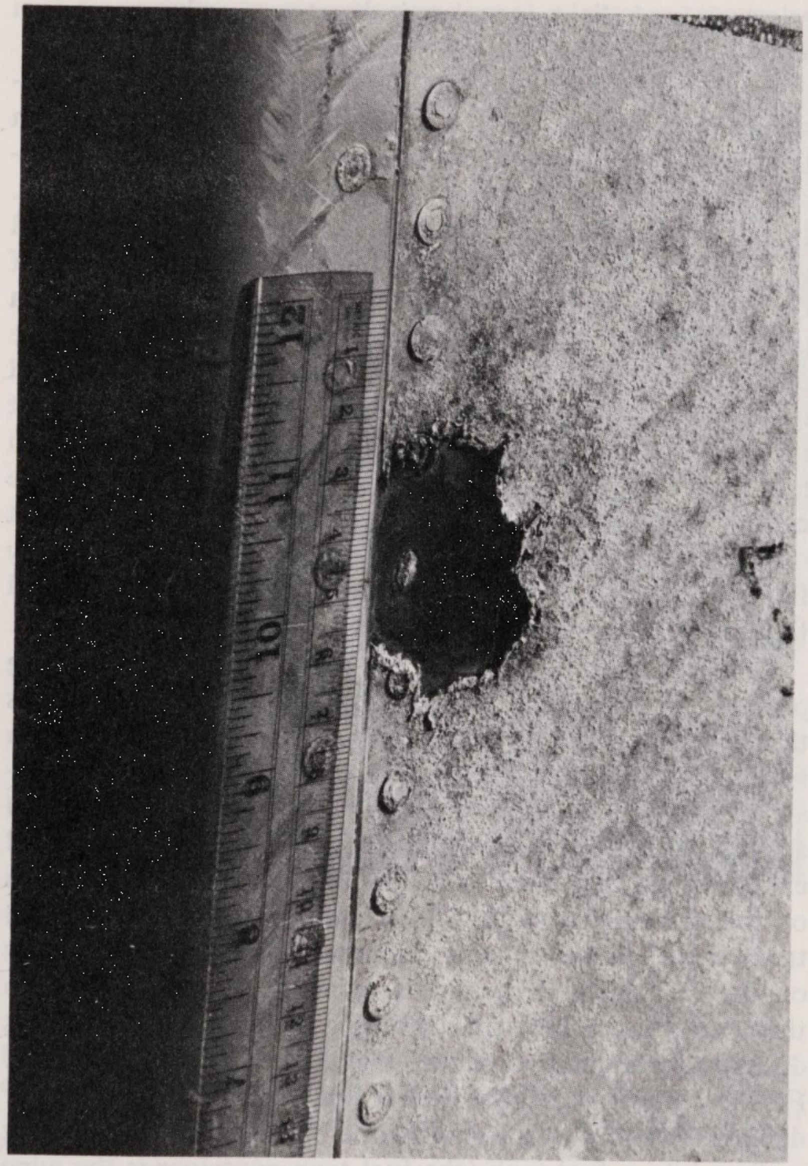

Fig. 4.3. Close-up of the major wing-tip hole and surrounding pitting shown at the left of Figures 4.1' and 4.2. (Courtesy, Bernard Vonnegut and Roger Cheng)

acknowledged lightning-related disasters involving modern commercial aircraft.

On June 26, 1959 near Milan, Italy, a Lockheed Constellation, a propeller-driven U.S. air carrier, was climbing in thunderstorm activity through about 11,000 ft when structural disintegration occurred. All persons aboard were killed. Investigation revealed that two of the fuel tanks had exploded. Fuel vent pipes extending off the trailing edges of the wings were exhausting air and fuel vapors (to maintain a pressure inside the fuel tank equal to the decreasing pressure outside, as the plane climbed) and were a likely point for fuel ignition. It is not clear whether lightning struck a fuel vent pipe or whether a traveling spark (similar to those which leave lightning rods; Chapter 2) was initiated at the pipe causing ignition. In either event, it is probable that some sort of electric spark ignited the fuel at the vent pipe, the burning fuel propagated into the number 6 and 7 fuel tanks, and those fuel tanks exploded (Ref. 4.1).

On December 8, 1963, a Boeing 707 was in a holding pattern at 5000 ft near Elkton, Maryland. There was thunderstorm activity in the area. Witnesses reported a lightning stroke near the aircraft concurrent with or immediately followed by explosion and burning of the 707. All aboard were killed. It was found that three fuel tanks had exploded and that there were lightning strike marks and holes on the left wing tip (Figs. 4.1, 4.2, 4.3). Exactly how the fuel tanks were ignited remains unknown. The fuel vent outlets on the 707 are underneath the wings and are recessed into them, so it is unlikely that lightning could strike the outlets directly or that they could be the source of traveling sparks. Further, when recovered, the vent outlets showed no signs of damage. Possibly a lightning stroke burned through the wing surface into a fuel tank (the fuel tank container is the wing skin). No evidence for this effect was found, although the explosion might have destroyed any pertinent evidence. The

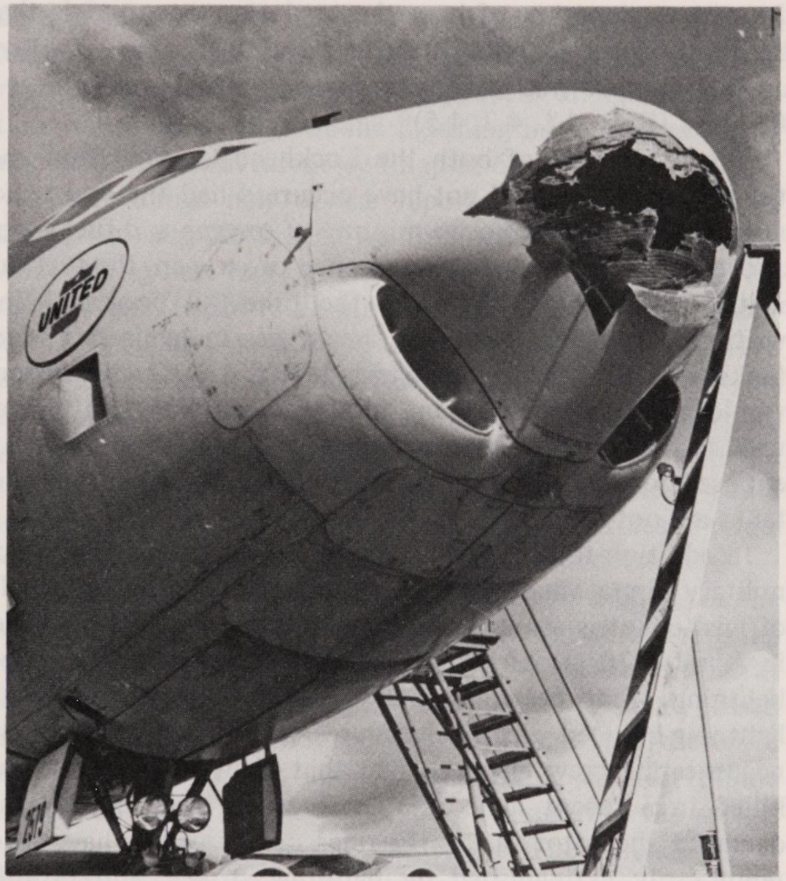

Fig. 4.4. Lightning damage to the radome of a commercial jetliner — one of four planes hit by lightning on one day in January 1969 in separate incidents near Los Angeles (Ref. 4.3). ("Los Angeles Times" photograph)

favored explanation of how the explosion occurred is that lightning caused sparks at some point where an explosive air-fuel mixture was present, perhaps in a fuel tank or near a fuel leak (Refs. 4.1, 4.2, 4.5).

The explosions of both the Lockheed Constellation and the Boeing 707 would not have occurred had the fuel tanks not contained an explosive mixture of oxygen and fuel vapor above the liquid fuel. In the absence of oxygen, the fuel can neither burn nor explode. At the time this book is being written, systems for replacing the oxygen in an aircraft's fuel tanks with nitrogen are being tested (Ref. 4.2). Such a system would make aircraft immune to explosions from lightning as well as from fires due to accidents in which the fuel is not exposed to the oxygen of the air (accidents in which the fuel tanks are not ruptured).

In addition to these two commercial aircraft, a number of military aircraft have also been destroyed by lightning. No official count is available, but it is known that, for example, in April 1970 an Air Force jet exploded after being hit by lightning, that there have been several recent cases of lightning-induced electrical failures on jet aircraft in flight necessitating crew ejection, and that eight servicemen were killed in a recent helicopter crash when lightning hit and damaged the rotor blades. Engines on many jets have been temporarily extinguished by lightning strikes, and in July 1969 a lightning strike to a jet resulted in two bombs being released.

When lightning strikes a plane, the plane substitutes for part of the lightning's air path. Typically lightning attaches to the plane's extremities, often entering at one wing tip and leaving by the top of the tail. When a lightning stroke is forming near a plane, the extremities of the plane send out traveling sparks to meet that lightning and cause it to pass through the plane, much as a lightning rod draws the lightning to itself. Thus a plane may attract lightning. In

addition, a plane may fly into an already existing lightning channel and become a part of the channel, or it may initiate a brand new lightning stroke. The process of lightning initiation is similar to that of the lightning initiation caused by tall conducting structures (see Chapter 6).

In only about half of the reported lightning strikes to commercial aircraft are active thunderstorms present (Ref. 4.4). The other half of the reported strikes occurs in precipitation (ice or rain) which is not related to a thunderstorm. Typically, in these cases no lightning is observed before the strike which hits the plane. There appears to be no difference between thunderstorm and non-thunderstorm lightning as regards damaging effects to an aircraft. Undoubtedly, thunderstorm-related lightning strikes would represent a much higher percentage of total strikes if pilots did not intentionally avoid flying through thunderstorms.

It is interesting to note that Apollo 12 (launched for the moon on November 14, 1969) was twice struck by lightning in the first minute after lift-off, once at 6000 ft and again at 13,000 ft. The clouds through which the 360 ft rocket passed had not previously produced any lightning. It is therefore likely that the rocket and its conducting exhaust plume initiated the two discharges. Apollo 12 survived the strikes although their effects were considerable. They included: (1) temporary disconnection of the fuel cells which powered the command module systems, (2) temporary disruption of the primary inertial guidance system used to put the spacecraft into and out of earth orbit and on its way toward the moon, and (3) permanent loss of a number of measurements of the skin temperature of the rocket and of the quantity of propulsion fuel remaining (Ref. 4.6).

From 1964 to 1966 a F-100F jet fighter aircraft, specially protected against the hazardous effects of lightning and equipped to photograph lightning and to measure its electrical current, pressure wave (see Chapter 12), and other properties, was deliberately flown through thunderstorms (Refs. 4.2, 4.7). The plane was struck by lightning 55 times and much valuable information was accumulated.

Ball lightning inside airplanes is discussed in Chapter 15.

REFERENCES

4.1. Conference on Fire Safety Measures for Aircraft Fuel Systems: Report of Conference, December 11-12, 1967, Federal Aviation Administration, Washington, D. C., 1967. Available from Defense Documentation Center as AD 672036.

4.2. Lightning and Static Electricity Conference, 3-5 December 1968, Part II. Conference Papers, May 1969, Technical Report AFAL-TR-68-290, Part II, Air Force Avionics Laboratory, Wright-Patterson Air Force Base, Ohio. Available from Defense Documentation Center as AD 693135.

4.3. *Los Angeles Times,* Final Edition, First Page, January 29, 1969.

4.4. Harrison, Henry T., United Air Line Turbojet Experience with Electrical Discharges, UAL Meteorological Circular No. 57, January 1, 1965.

4.5. Aircraft Accident Report, Boeing 707-121, N709PA Pan American World Airways, Inc., near Elkton, Maryland, December 8, 1963, Civil Aeronautics Board File No. 1-0015, February 25, 1965.

4.6. Godfrey, R., E. R. Mathews, and J. A. McDivitt, Analysis of Apollo 12 Lightning Incident, NASA MSC-01540, January 1970.

4.7. Petterson, B. J., and W. R. Wood, Measurements of Lightning Strikes to Aircraft, Report No. SC-M-67-549, Sandia Laboratory, Albuquerque, New Mexico, January 1968.

ADDITIONAL READING

Fisher, F. A., and J. A. Plumer, *Lightning Protection of Aircraft,* NASA Reference Publication 1008, 550 pages, October 1977. For sale by the National Technical Information Service, Springfield, Virginia 22161.

Clifford, D. W., Aircraft Mishap Experience from Atmospheric Electricity Hazards, Atmospheric Electricity—Aircraft Interaction, 1980, AGARD Lecture Series No. 110, Document No. AGARD-LS-110, for sale by the National Technical Information Service, Springfield, Virginia 22161.

chapter five

How Does Lightning Damage Trees And Buildings?

As discussed in Chapter 7, roughly 2000 thunderstorms are in progress over the earth's surface at a given time, and collectively they may produce as many as 100 cloud-to-ground discharges each second – or more than 8 million per day. If these discharges were evenly distributed over the earth, about six percent of them would strike in the world's forested lands (Ref. 5.1). Fortunately, most discharges striking in forests do not cause forest fires. Nevertheless, about 10,000 forest fires are ignited by lightning in the United States each year.

Lightning may strike a tree and leave it apparently unharmed (Ref. 5.2), or it may cause considerable structural damage without noticeable burning. The detailed effects of lightning on trees have been recently studied by Alan R. Taylor of the U.S. Forest Service (Refs. 5.1, 5.3, 5.4, 5.5, 5.6). He found that most trees struck are not killed. The majority recover from whatever lightning damage they have sustained, though many are weakened and ultimately succumb to attacks by insects and disease. Visible damage to tree trunks ranges from superficial bark flaking, to strip-like furrowing along the trunk, to almost total destruction. Damage of intermediate severity is shown in Fig. 5.1: the tree top has been shattered away and a spiral scar with a crack along its axis is wound around the trunk. A close-up of how typical lightning tree scars are formed is shown in Fig. 5.2. Figures 5.3 and 5.4 show lightning-damaged trees.

Taylor, in one study, examined 1000 lightning-damaged Douglas firs in western Montana. Most had shallow continuous scars a few inches wide along their trunks. About 20% had two or more scars, 10% had severed tops, and about 1% had been reduced to slabs and slivers. Most of the scars were spiral, a few were straight. The average scar extended along 80% of the tree height, but none extended to the very tips of the trees. Scars either reached to ground level or close to ground level. Often along the center line of the lightning scar was a crack which penetrated into the tree (Figs. 5.1 and 5.2), and, when wood was removed from a tree by lightning, it was usually ejected as two parallel slabs, separated along this crack. Sometimes, in place of the crack lightning left a narrow strip of shredded inner-bark fiber fixed in a smooth shallow groove about 1/16-in. wide (Fig. 5.5).

These damage characteristics are typical not only for Douglas firs but for most conifers (cone bearing trees, mostly evergreens) throughout the United States. The descriptions also appear valid for most rough-barked species of deciduous trees (trees which shed their leaves each year) such as the oaks. The relatively smooth barks of other deciduous trees – birches, for example – present quite different damage characteristics. The major difference is that the bark is not removed in narrow, uniform strips but is torn off in large, somewhat irregular patches or sheets (Fig. 5.6).

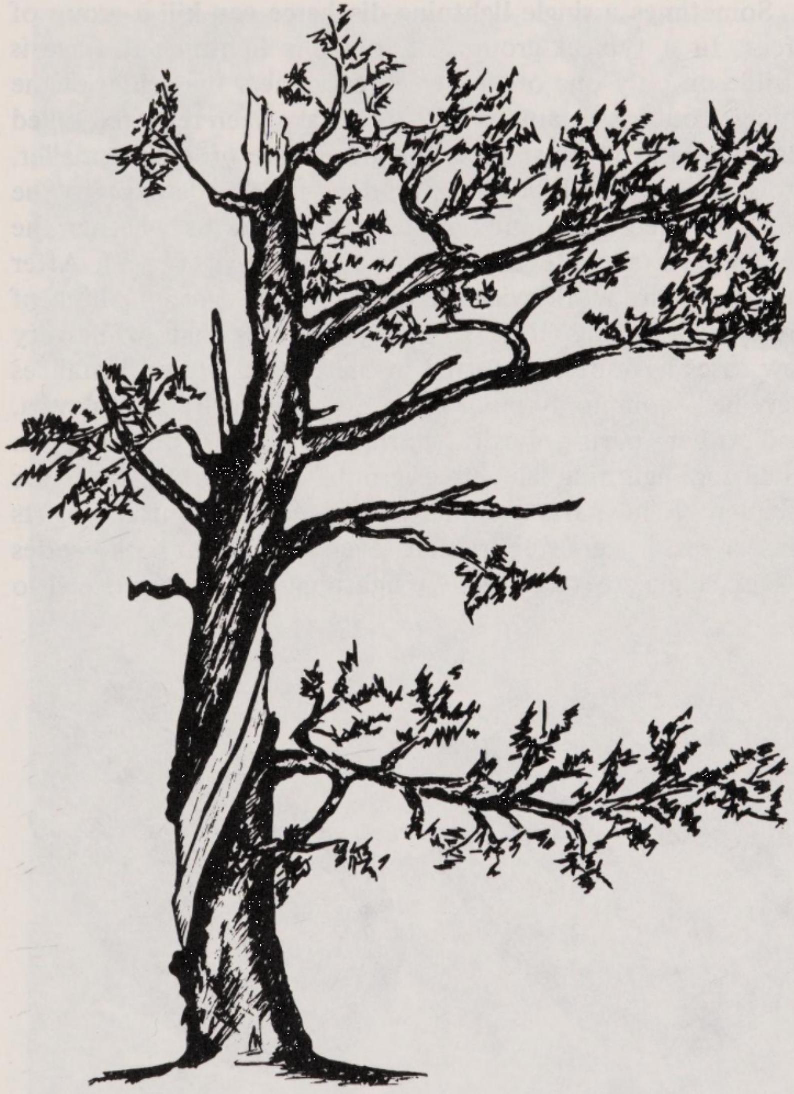

Fig. 5.1. The top of this Douglas fir was shattered away and a spiral lightning scar with a crack along its axis winds around the trunk. (Field sketch by A. R. Taylor)

Sometimes a single lightning discharge can kill a group of trees. In a typical group kill, obvious lightning damage is visible on only one or two trees, often near the center of the dying group. As many as 160 trees have been reported killed this way, but in most cases the groups are probably smaller. It is unclear whether lightning does unseen damage to the roots of trees surrounding the struck tree or whether the aerial parts are affected by the discharge (Ref. 5.7). After reviewing the world-wide literature on the group killing of trees by lightning (Ref. 5.1), Taylor notes that, while very few cases have been reported in the United States, instances have been commonly reported in Europe, Australia, Malaysia, and other parts of the world. A similar phenomenon involving lightning and tree groups is reported to occur frequently in parts of the United States. Entomologists (insect researchers) report that several species of bark beetles attack single trees damaged by lightning and then proceed to

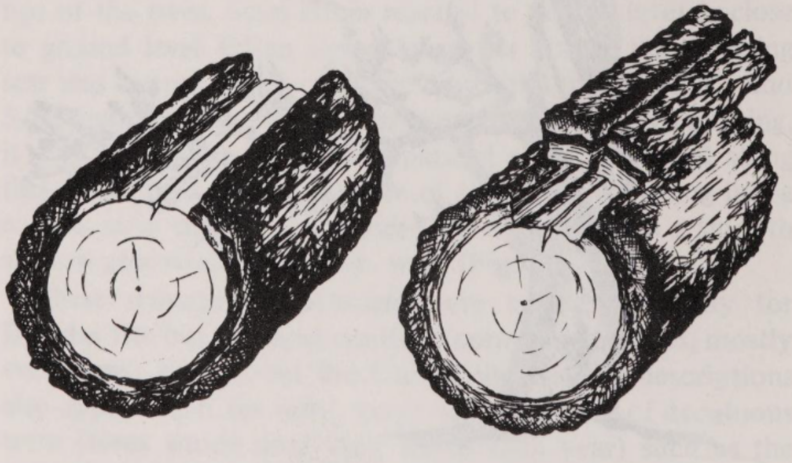

Fig. 5.2. Formation of typical lightning tree scars. Left, bark and wood loss; right, bark loss only. Bark and wood removed in this way are usually found in two slabs. The slabs have been separated along the crack at the centerline of the tree scar. (Sketch by A. R. Taylor)

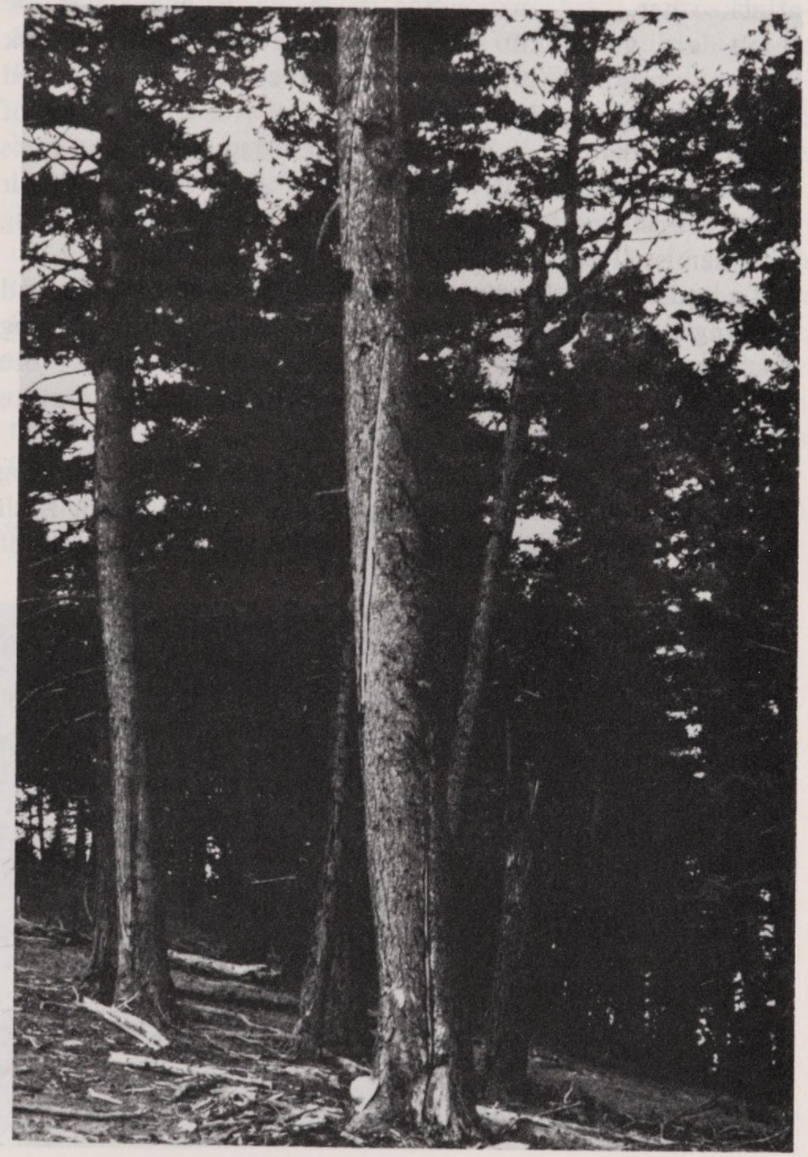

Fig. 5.3. A Douglas fir with three spiral lightning scars. Only bark was removed, exposing the outer wood. (U.S. Forest Service photo)

attack other trees surrounding the damaged one. Often the result is a group kill similar to those not involving bark beetles – so similar in fact that it has recently been suggested that lightning may be playing a hidden role in this type of group kill; that is, it may do unseen damage to the trees surrounding an obviously struck tree, thereby reducing their natural resistance to attack by the bark beetles and hastening their demise (Refs. 5.6, 5.8).

To understand the direct effects of lightning on trees and buildings, we must first take a closer look at the lightning discharge. In the course of this discussion we will define more specifically some of the words used in the technical literature to describe lightning.

A complete lightning discharge is called a *flash.* A lightning flash typically lasts a few tenths of a second. Each flash between cloud and ground is composed of a number of component strokes. The number is sometimes one (a single-stroke flash), most often is three or four (multiple-stroke flash), and may be as many as 20 or 30. Strokes are typically 40 or 50 thousandths of a second apart. Cloud-to-ground lightning often appears to flicker due to the fact that the lightning channel is dark between the component strokes. The longer the time between strokes, the greater the flickering effect. When strokes are very close together (e.g., when they are separated by only 20 thousandths of a second), little flickering is observed because the eye retains the image of the stroke during the "dark time". It is this slow response of the eye that makes movies and TV appear continuous rather than the series of individual pictures that comprise them. Movies that are run too slowly also appear to flicker.

The bright stroke channels we see (an example is shown in Fig. 2.1c) and the stroke currents which cause the light are

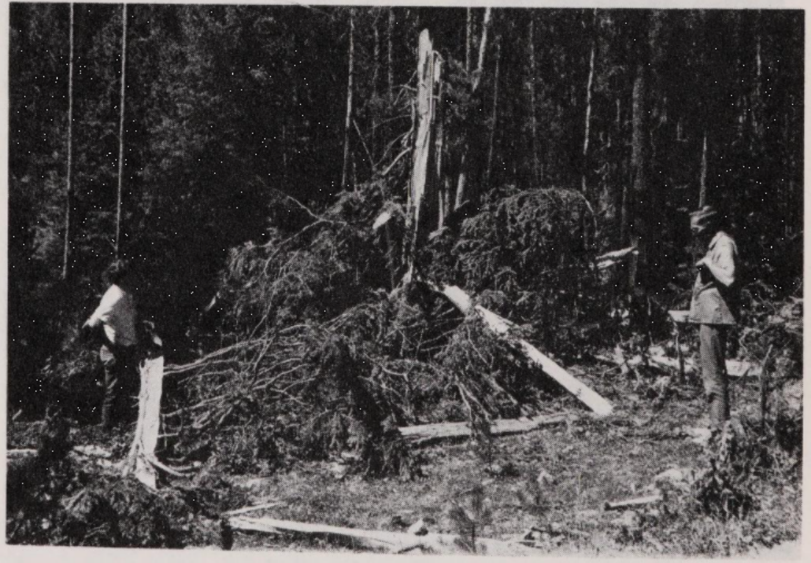

Fig. 5.4. Forest Service scientists examine the remains of a 70-ft fir demolished by lightning in western Montana. (U.S. Forest Service photo)

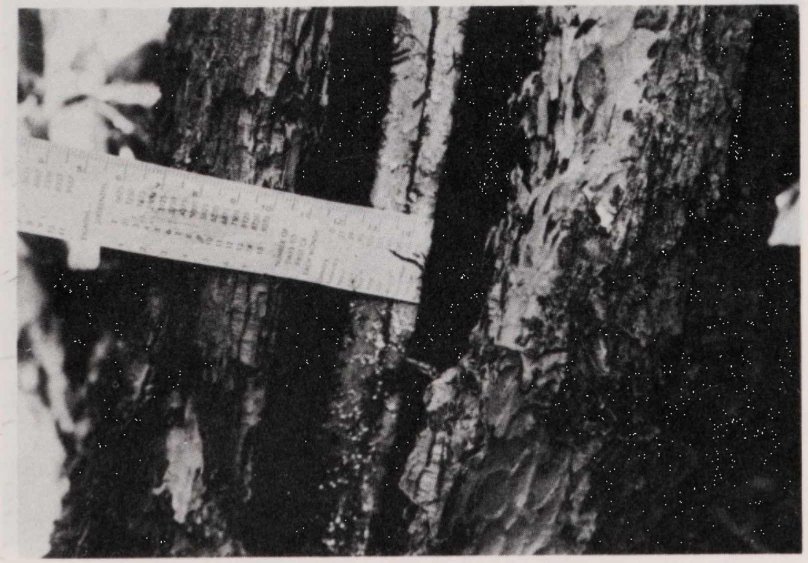

Fig. 5.5. Close-up of a lightning tree scar showing a narrow strip of inner-bark fiber along the centerline of a bark-depth furrow. (U.S. Forest Service photo)

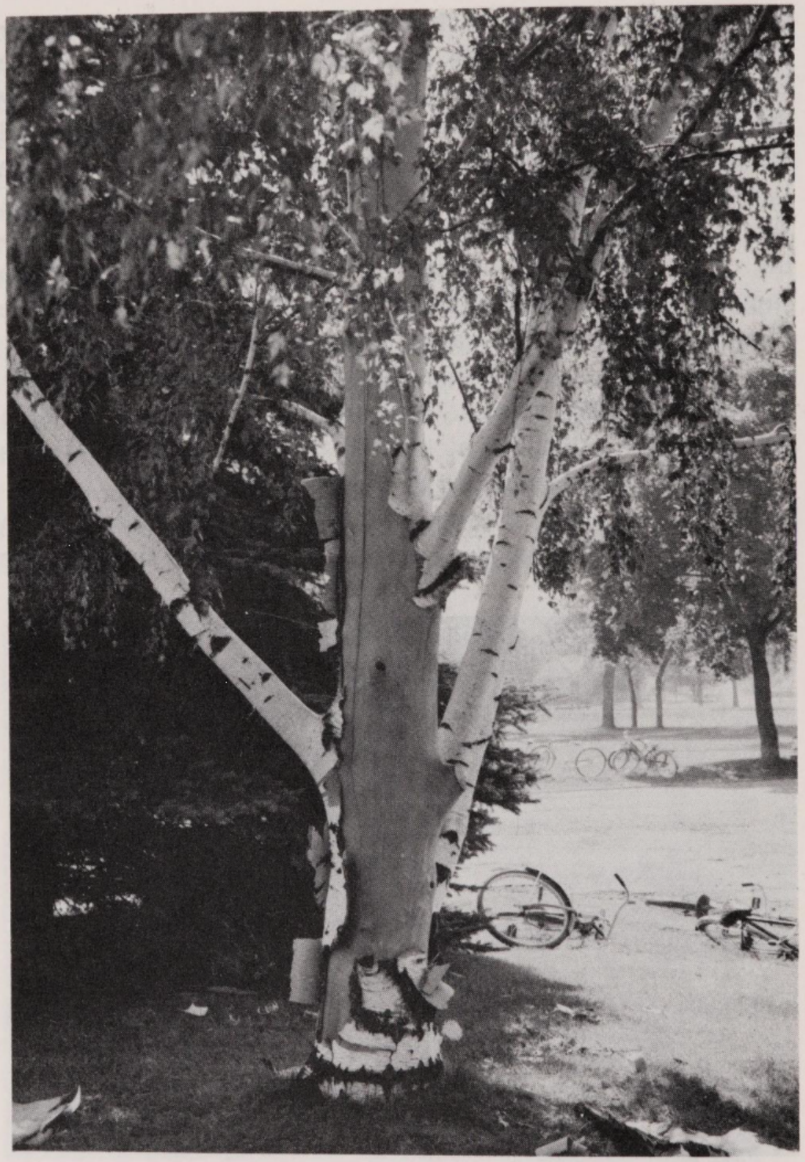

Fig. 5.6. Lightning removed bark from this paper birch in large, irregular patches. (U.S. Forest Service photo)

initiated by electrical discharges which move from cloud to ground (a discharge preceding the first stroke of a flash is shown in Fig. 2.1a,b). (Details of the lightning stroke formation are given in Chapters 6 and 9.)

Typical lightning currents measured at ground are shown in Fig. 5.7 (Ref. 5.9). Peak currents are generally 10,000 to 20,000 amps, but occasionally they range up to hundreds of thousands of amps. (The conventional household electric circuit will carry up to 15 amps. A 100 watt light bulb uses about 1 amp.) First strokes generally have larger currents than subsequent strokes. A lightning stroke reaches peak current in a few millionths of a second. The current then decreases, generally terminating in a thousandth of a second or so unless continuing current flows. Figure 5.7a shows a three-stroke flash with no continuing current. Figure 5.7b

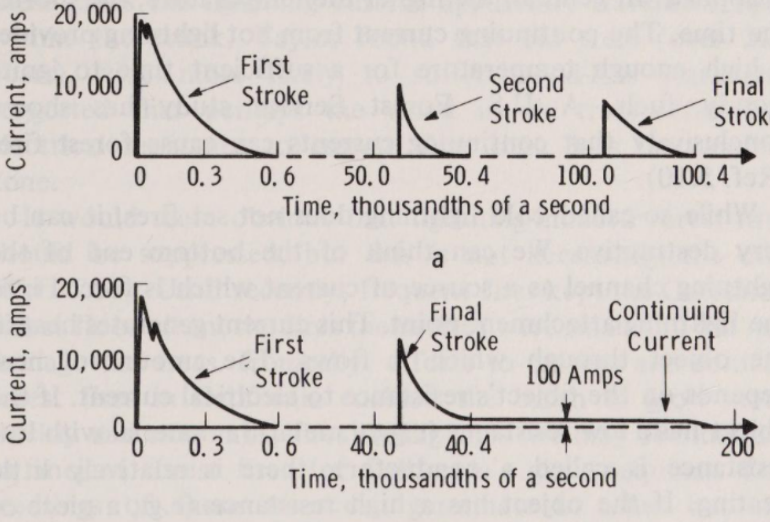

Fig. 5.7. (a) Current at ground due to a typical lightning flash without continuing current – so-called cold lightning. (b) Current at ground due to a typical lightning flash with continuing current – so-called hot lightning.

shows a two-stroke flash in which the final stroke is followed by continuing current. Continuing current is typically 100 amps and may last for one or two tenths of a second. About half of all flashes contain a continuing current following one of the component strokes.

Lightning flashes which contain continuing current are commonly called *hot lightning;* flashes which do not are called *cold lightning.* Hot lightning sets fires, cold lightning does not. The reason for the terminology is obvious. It is, nevertheless, misleading. The temperature in both hot and cold lightning is the same – 15,000 to 60,000°F (see Chapter 11). The difference is that it is maintained longer in hot lightning (tenths of a second vs. thousandths for cold lightning). To set fire to a piece of wood, the flame must contact the wood for a reasonable length of time: initiation of burning depends on the length of time a given temperature is applied. In general, the higher the temperature, the shorter the time. The continuing current from hot lightning provides a high enough temperature for a sufficient time to ignite woody fuels. A U.S. Forest Service study has shown conclusively that continuing currents can cause forest fires (Ref. 5.10).

While so-called cold lightning does not set fires, it can be very destructive. We can think of the bottom end of the lightning channel as a source of current which is forced into the lightning attachment point. This current generates heat in the object through which it flows. The amount of heat depends on the object's resistance to electrical current. If the object has a low resistance (e.g., a metal) (a material with low resistance is called a conductor), there is relatively little heating. If the object has a high resistance (e.g., a piece of wood or plastic), there is a great deal of heating although not necessarily enough to cause burning. The essentially instantaneous lightning current rise inside a high-resistance material causes rapid heating and consequent vaporization (i.e., conversion from solid to gas) of some of the internal material. As a result, a very high pressure is quickly generated within the material and, typically, this pressure blows the material apart.

For many of the tree scars examined by Taylor, the lightning apparently followed a path through the cambium (a thin layer of living cells between the inner bark and the wood) or through the moist inner bark tissue. Possibly these zones were chosen as the lightning's paths because they offered lower resistance than the outer bark or the wood. In any event, the pressure generated in the cambial region blew off a bark strip creating a scar as shown in Fig. 5.5. The pressure also could cause a split in the tree, but it is not clear how a strip of inner-bark fiber could be pasted along the centerline of the scar. In some of the trees examined by Taylor, wood as well as bark was blown out from the tree. In these cases the lightning current apparently traveled deeper within the trunk. Taylor found that old trees (over 200 years) were more likely to suffer wood-loss scars and suggested that perhaps the wood in old trees offers less electrical resistance to current flow than does the cambial zone.

It would seem obvious that lightning-induced forest fires should be suppressed, but this is not necessarily the case (Ref. 5.11). Until recently, frequent fires kept the California forest floor clean; the fires themselves were small and did not damage the trees. Ironically, efforts to prevent and contain forest fires in California enabled the brush to grow more thickly and now most fires are big ones. We may be indebted to ancient forest fires for California's giant sequoias. The seedlings of these trees can germinate in ashes but are suppressed under the thick layer of needles that might cover an unburned forest floor.

From all the foregoing discussion it is obvious that houses can be set on fire by lightning and that high-resistance material comprising houses can be blown apart. Besides directly setting fire to houses and other structures, lightning can indirectly set fires by causing the power company's current to jump through the air between electrical wires or from a wire to a grounded object. The resulting electrical discharge may have a considerably longer duration than the lightning flash which started it. The only known way to protect a house against a direct lightning strike is with a lightning rod system (see Chapter 2). Similarly, the only known way to protect a tree against lightning is to outfit it with a similar rod system. This would consist of metal wire or strip fixed to the tree from a well-established ground and running up the tree trunk to the top of the tree and out the major branches. Thus lightning striking the tree top or any major branch would be harmlessly carried to earth.

REFERENCES

5.1. Taylor, A. R., Lightning Effects on the Forest Complex, *Proc. 9th Annual Tall Timbers Fire Ecol. Conf.*, 127-150 (1969).
5.2. Orville, R. E., Photograph of a Close Lightning Flash, *Science*, *162*, 666-667 (1968). The photograph from this article appears on the frontispiece.
5.3. Taylor, A. R., Lightning Damage to Trees in Montana, *Weatherwise*, *17* (No. 2), 62-65 (1964).
5.4. Taylor, A. R., Diameter of Lightning as Indicated by Tree Scars, *J. Geophys. Res.*, *70*, 5693-5695 (1965).
5.5. Taylor, A. R., Tree-bole Ignition in Superimposed Lightning Scars, USDA Forest Serv. Res. Note INT-90, 4 pp. (1969).
5.6. Schmitz, R. F., and A. R. Taylor, An Instance of Lightning Damage and Infestation of Ponderosa Pines by the Pine Engraver Beetle in Montana, USDA Forest Serv. Res. Note INT-88, 8 pp. (1969).
5.7. Minko, G., Lightning in Radiata Pine Stands in Northeastern Victoria, *Australian Forester*, *30* (No. 4), 257-267 (1966).
5.8. Komarek, E. V., Sr., The Natural History of Lightning, *Proc. 3rd Annual Tall Timbers Fire Ecol. Conf.*, 139-186 (1964).
5.9. Uman, M. A., *Lightning*, McGraw-Hill Book Co., Inc., New York, 1969, pp. 114-137.
5.10. Fuquay, D. M., R. G. Baughman, A. R. Taylor, and R. G. Hawe, Characteristics of Seven Lightning Discharges That Caused Forest Fires, *J. Geophy. Res.*, *72*, 6371-6373 (1967).
5.11. Love, R. M., The Rangelands of the Western United States, *Sci. Am.*, *222*, 88-96, February 1970.

chapter six

Does Lightning "Never Strike Twice?"

Does It Always Strike The Tallest Object?

Much of what is known about lightning today has been discovered precisely because lightning *does* strike the same structure over and over again. To measure a given property of many lightning strokes (for example, the range of values for the peak current), it is necessary to position appropriate instrumentation close to many lightning strokes. The Empire State Building in New York City is struck by lightning an average of about 23 times per year. As many as 48 strikes have been recorded in one year, and during one thunderstorm eight strikes occurred within 24 minutes (Refs. 6.1, 6.2). The General Electric Company conducted a 10-year study (1935-1941, 1947-1949) of lightning at the Empire State Building. Various properties of the lightning current were measured as the current flowed into the building, and high-speed photographs of the formation of the lightning channel near the top of the building were obtained. A similar but more detailed study of lightning was begun on Mount San Salvatore near Lugano, Switzerland in 1942 and still continues at the time of the writing of this book. The instrumented towers near Lugano have recorded lightning current properties an average of about 40 times a year, with a maximum in excess of 100 times a year (Ref. 6.3). A sizable fraction of what we know about lightning currents and, in fact, about lightning in general was determined by the Empire State Building and Lugano studies.

The probability of a structure being struck by lightning depends on its height. For a structure projecting above a moderately flat area in a region of moderate thunderstorm activity, say Pennsylvania or New York (see Chapter 7), the expected number of strikes per year is roughly one for a 300 ft structure, three for a 600 ft structure, five for an 800 ft structure, ten for a 1000 ft structure, and twenty for a 1200 ft structure (Ref. 6.4). A 50 ft structure will be struck about once every four to six years (Refs. 6.2, 6.4), and a quarter acre of flat land (a large residential lot) will be struck about once every 100 years or more (see Chapter 7). Any structure, no matter what its size, may be struck by lightning. Sometimes, however, it is necessary to rebuild the structure before it can be struck again. The bell tower of St. Mark's in Venice was completely destroyed by lightning three times and severely damaged nine times in a period of about 400 years. In 1766 a lightning rod was installed on the tower and since then no further lightning damage has occurred (Refs. 6.5, 6.6).

It is interesting to note that as late as the 18th century church bells were rung in Europe as an effort to protect against lightning. The "ignorant" believed that the ringing of church bells dispersed the evil spirits; the "informed" believed (erroneously) that the sound produced by the bells broke the continuity of the lightning path. Medieval church bells are often inscribed *Fulgura frango* (I break the

lightning). Ringing church bells during thunderstorms proved very hazardous to the bell-ringers. In a period of 33 years, 386 church towers were struck, and 103 bell-ringers were killed at their ropes (Refs. 6.5, 6.6).

Since lightning preferentially strikes the taller structures, it is reasonable to ask whether tall structures merely act as lightning rods or whether they influence the lightning to a greater extent. The question is important since much of what we know about lightning comes from studies of lightning strikes to tall buildings. Scientists and engineers would like to be certain that lightning characteristics so obtained also apply to lightning which strikes small buildings, power lines, missiles, etc. The Empire State Building study was the first to show that most of the flashes to a very tall structure are initiated by the structure itself. In the Lugano study about 75% of the flashes observed were initiated by the towers. These structure-initiated flashes differ significantly from the usual cloud-to-ground flash. Flashes to very tall structures which are not structure-initiated are thought to resemble the usual lightning to ground, but there is still some dispute about this.

It is appropriate now to look at how lightning gets started. The usual flash between cloud and ground is initiated in the base of the cloud. The initiating discharge, the downward-traveling spark shown in Fig. 2.1a,b, is called the *stepped leader*. The stepped leader is a low-luminosity traveling spark which moves from the cloud to the ground in rapid steps about 50 yards long. Each step takes less than a millionth of a second, and the time between steps is about 50 millionths of a second. The stepped leader is not seen by the eye, but may be photographed with special cameras (see Chapter 16, Fig. 16.3). The visible lightning flash discussed in the previous chapter occurs when the stepped leader contacts the ground. Precisely what happens when the stepped leader reaches the ground is the subject of Chapter 9. What is important to note now is that the usual stepped leader starts from channel near the top of the building were obtained. A similar but more detailed study of lightning was begun on Mount San Salvatore near Lugano, Switzerland in 1942 and still continues at the time of the writing of this book. The instrumented towers near Lugano have recorded lightning current properties an average of about 40 times a year, with a maximum in excess of 100 times a year (Ref. 6.3). A sizable fraction of what we know about lightning currents and, in fact, about lightning in general was determined by the Empire State Building and Lugano studies.

The probability of a structure being struck by lightning depends on its height. For a structure projecting above a moderately flat area in a region of moderate thunderstorm activity, say Pennsylvania or New York (see Chapter 7), the expected number of strikes per year is roughly one for a 300 ft structure, three for a 600 ft structure, five for an 800 ft structure, ten for a 1000 ft structure, and twenty for a 1200 ft structure (Ref. 6.4). A 50 ft structure will be struck about once every four to six years (Refs. 6.2, 6.4), and a quarter acre of flat land (a large residential lot) will be struck about once every 100 years or more (see Chapter 7). Any structure, no matter what its size, may be struck by lightning. Sometimes, however, it is necessary to rebuild the structure before it can be struck again. The bell tower of St. Mark's in Venice was completely destroyed by lightning three times and severely damaged nine times in a period of about 400 years. In 1766 a lightning rod was installed on the tower and since then no further lightning damage has occurred (Refs. 6.5, 6.6).

It is interesting to note that as late as the 18th century church bells were rung in Europe as an effort to protect against lightning. The "ignorant" believed that the ringing of church bells dispersed the evil spirits; the "informed" believed (erroneously) that the sound produced by the bells broke the continuity of the lightning path. Medieval church bells are often inscribed *Fulgura frango* (I break the lightning). Ringing church bells during thunderstorms proved very hazardous to the bell-ringers. In a period of 33 years, 386 church towers were struck, and 103 bell-ringers were killed at their ropes (Refs. 6.5, 6.6).

Since lightning preferentially strikes the taller structures, it is reasonable to ask whether tall structures merely act as lightning rods or whether they influence the lightning to a greater extent. The question is important since much of what we know about lightning comes from studies of lightning strikes to tall buildings. Scientists and engineers would like to be certain that lightning characteristics so obtained also apply to lightning which strikes small buildings, power lines, missiles, etc. The Empire State Building study was the first to show that most of the flashes to a very tall structure are initiated by the structure itself. In the Lugano study about 75% of the flashes observed were initiated by the towers. These structure-initiated flashes differ significantly from the usual cloud-to-ground flash. Flashes to very tall structures which are not structure-initiated are thought to resemble the usual lightning to ground, but there is still some dispute about this.

It is appropriate now to look at how lightning gets started. The usual flash between cloud and ground is initiated in the base of the cloud. The initiating discharge, the downward-traveling spark shown in Fig. 2.1a,b, is called the *stepped leader*. The stepped leader is a low-luminosity traveling spark which moves from the cloud to the ground in rapid steps about 50 yards long. Each step takes less than a millionth of a second, and the time between steps is about 50 millionths of a second. The stepped leader is not seen by the eye, but may be photographed with special cameras (see Chapter 16, Fig. 16.3). The visible lightning flash discussed in the previous chapter occurs when the stepped leader contacts the ground. Precisely what happens when the stepped leader reaches the ground is the subject of Chapter 9. What is important to note now is that the usual stepped leader starts from the cloud without any "knowledge" of what buildings or geography are present below. In fact, it is thought, as suggested in Chapter 2, that the stepped leader is "unaware" of objects beneath it until it is some tens of yards from the eventual strike point. When "awareness" occurs, a traveling spark is initiated from the point to be struck and propagates upward to meet the downward-moving stepped leader, completing the path to ground. In this way the usual lightning flash to ground or to small structures is started. A small fraction of the strikes to very tall buildings also occurs in this way.

However, most lightning flashes to very tall buildings are the result of a reverse process. They are initiated by stepped leaders which start at the building top and propagate upward to the cloud. Because they propagate upward, they and the resulting visible discharge channel branch upward (Fig. 6.1). It is possible that upward-going leaders are just long versions of the leaders (previously called traveling sparks) which propagate upward some tens of yards from lightning rods. That is, the upward-moving stepped leaders from tall buildings may well contact downward-moving leaders beneath or within the cloud. On the other hand, the upward-moving stepped leaders may simply terminate in some region of cloud charge.

Lightning flashes initiated by tall buildings exhibit different current characteristics from those of the usual cloud-to-ground flash discussed in Chapter 5 and illustrated in Fig. 5.7. The current in a structure-initiated flash rises slowly, in hundredths or even tenths of a second, to a current peak of a few hundred amperes and flows for a few tenths of a second. Often the low-level continuous current is punctuated by one or more current peaks similar to those shown in Fig. 5.7 for strokes subsequent to the first. Sometimes the continuous current stops before a current peak occurs. These current peaks are generally larger than those which occur during the continuing current. More details of the characteristics of upward-moving lightning are given in Ref. 6.7.

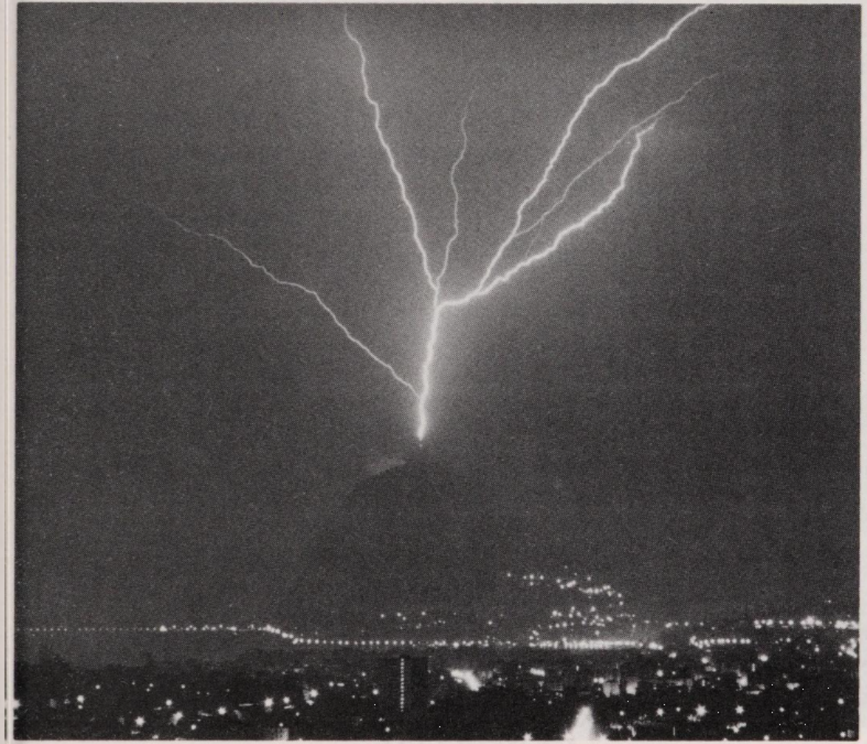

Fig. 6.1a. Lightning initiated by an upward-moving leader from a tower on Mt. San Salvatore near Lugano, Switzerland. The spot directly beneath the bottom of the lightning channel is a tower light. Upward-initiated lightning is branched upward in contrast to the downward branching of the usual cloud-to-ground lightning flash. (Courtesy, Richard E. Orville, State University of New York at Albany)

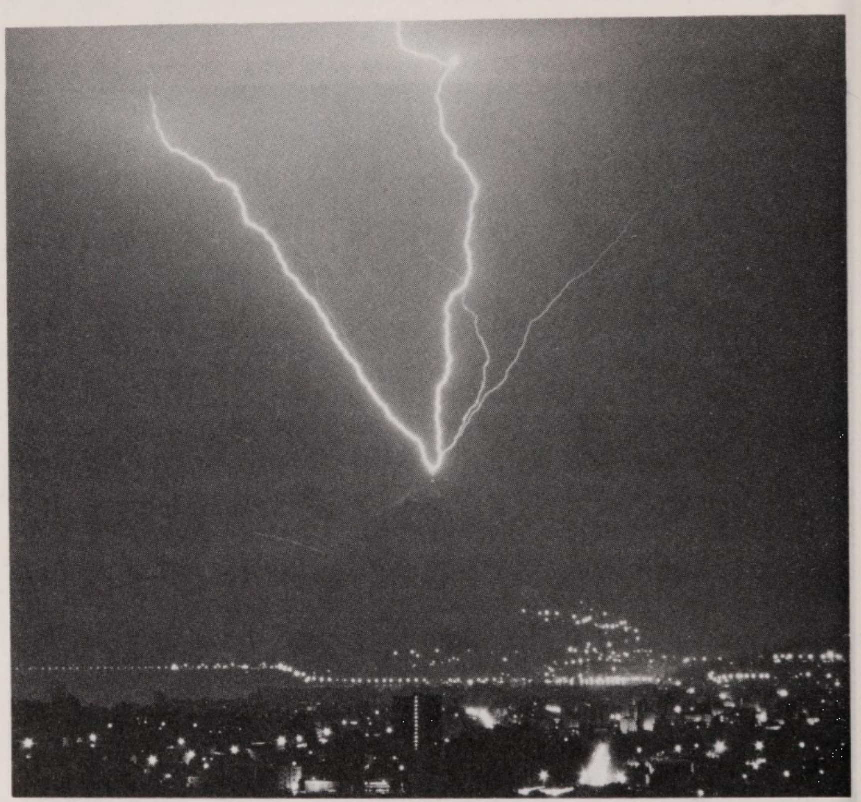

Fig. 6.1b. Lightning strikes twice! Another upward-going lightning from the San Salvatore tower. (Courtesy, Richard E. Orville)

It may appear logical from the content of this chapter that lightning should always "hit the tallest object". (Since the point of "strike" is determined by a leader which moves upward from the strike point, the strike point could just as well be called the "point of initiation".) Occasionally it doesn't! The Empire State Building has been struck 50 ft below the top (Ref. 6.2). The usual explanation for such anomalous behavior is the following: It is thought that the greater the electrical charge residing on the downward-moving stepped leader, the longer will be the connecting leader propagating upward from the building top. Thus, a relatively weak stepped leader can come closer to a building top without drawing an upward leader than can a relatively strong stepped leader. It is possible, therefore, that a weak stepped leader might "sneak" past the top of a building and only draw a leader when it had reached some lower level. Another explanation for the anomalous behavior could be that the obvious strike point has been kept from generating an upward-moving leader by a pocket of airborne charge, so-called *space charge.* Properly distributed regions of space charge could cause the lightning to strike almost anywhere. In any event, the cone of protection around the "highest point" (see Chapter 2) almost always describes the zone in which lightning will not strike.

REFERENCES

6.1. Hagenguth, J. H., and J. G. Anderson, Lightning to the Empire State Building – Part 3, *AIEE, 71,* pt. 3, 641-649 (1952).

6.2. Towne, H. M., Lightning, Its Behavior and What to Do About It, 1956, pamphlet published by United Lightning Protection Assn., Inc., Box 9, Onondaga, New York.

6.3. Berger, K., Novel Observations on Lightning Discharges: Results of Research on Mount San Salvatore, *J. Frankl. Inst., 283,* 478-525 (1967).

6.4. McCann, G. D., The Measurement of Lightning Currents in Direct Strokes, *Trans. AIEE, 63,* 1157-1164 (1944).

6.5. Hepburn, F., Man's Study of Lightning, *Science Progress, 44,* 635-646 (1956).

6.6. Schonland, B. F. J., *The Flight of Thunderbolts,* Second Edition, Clarendon Press, Oxford, 1964.

6.7. Uman, M. A., *Lightning,* Dover Publications, Inc., New York, 1984.

chapter seven

How Are Thunderstorms Formed?

Are There Locations With No Lightning?

How Many Thunderstorms Are In Progress In The World At One Time?

Roughly 2000 thunderstorms are in progress in the world at any one time (Ref. 7.1). A typical storm is thought to produce one to three cloud-to-ground flashes each minute; and thus, considering all thunderstorms in progress, there are perhaps 30 to 100 flashes to ground every second (Refs. 7.1, 7.2). The statistics on the number of thunderstorms in progress in the world at any one time were obtained in the 1920s by examining the records of meteorological stations located around the world. Where there were no meteorological stations, estimates were made of the thunderstorm activity. For this reason the number 2000 may be somewhat in error. Further, estimates of the number of cloud-to-ground flashes per minute for a typical storm are based primarily on data obtained at relatively few locations throughout the world and may also be in error. For example, very little is known about the lightning or other properties of storms which occur in the tropical regions of South America and South Africa. With the help of orbiting satellites, a more accurate determination of world-wide thunderstorm activity is possible and will probably be made in the near future.

A *thunderstorm-day* for a given location is defined as a day on which thunder is heard at that location, independent of whether one or many thunders were heard. Thunder cannot usually be heard if the lightning causing it is more than about 15 miles away (see Chapter 12). In Fig. 7.1 and 7.2, thunderstorm-day maps are shown for the U.S. and the world, respectively. Each curved line represents a fixed and stated number of thunderstorm-days *per year* and is drawn through all geographical locations that have that particular number of thunderstorm-days. One of the most active regions on earth is Java where thunder is heard 223 days per year. In the U.S., central Florida has the highest number of thunderstorm-days per year with 90. The lowest number occurs along the Pacific coast region of northern California, Oregon, and Washington where thunderstorms and lightning are rare.

A more pertinent statistic than the number of thunderstorm-days per year is the number of lightning strikes per square mile per year. While it is relatively easy to determine thunderstorm days (one needs merely to listen), it is a more difficult task to determine accurately the number of strikes per square mile. The strikes per area have been determined (1) from photographs taken sequentially over periods of several years, (2) from records of strikes to power lines, and (3) from electrical lightning counters which respond to the radio waves emitted by lightning. The combined results of several studies indicate that the number of flashes to ground per square mile per year is equal to between 0.05 and 0.8 times the number of thunderstorm-days per year (Refs. 7.3, 7.4, 7.5). For example, if you live in a region of the U.S. with 40 thunderstorm-days per year, you can expect between 2

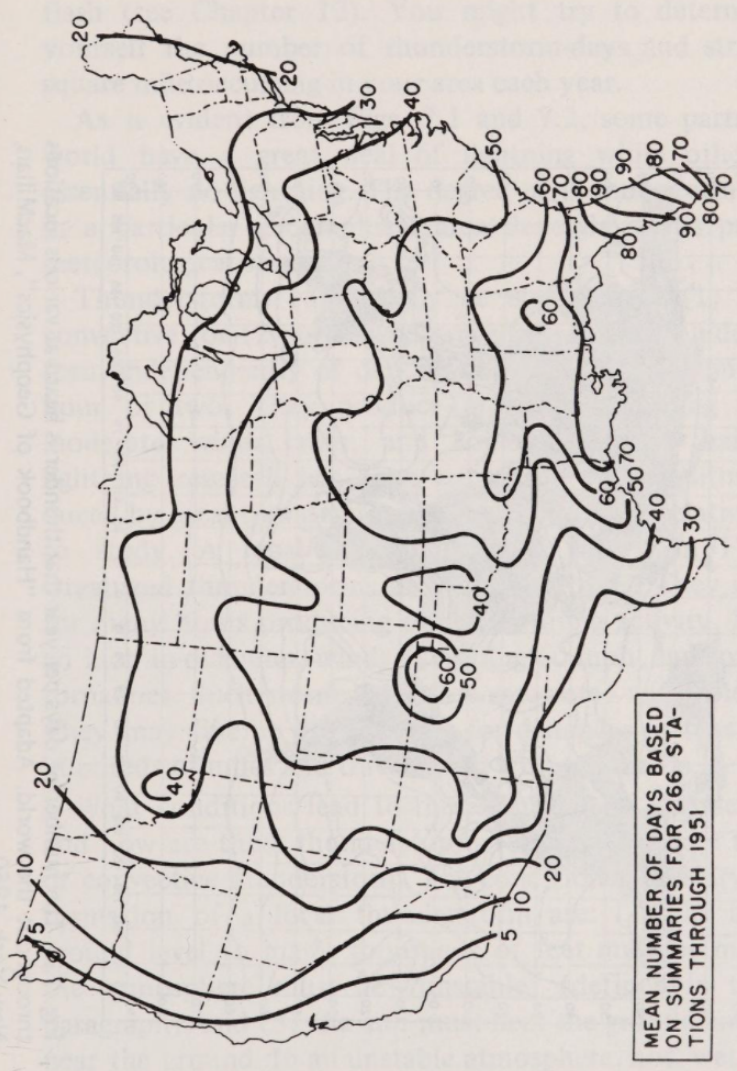
MEAN NUMBER OF DAYS BASED ON SUMMARIES FOR 266 STATIONS THROUGH 1951
20
20
30
40
50
60
70
80
90
90
80
70
80
90
80
70
60
60
70
50
40
30
40
60
50
40
20
10
5
20
10
5

Fig. 7.1. The number of days per year on which thunder is heard at various locations in the U.S.A. Adapted from "Mean Number of Thunderstorm Days in the United States", Technical Paper No. 19, Climatological Services Division, Weather Bureau, September 1952.

and 30 strikes per year in a square mile area surrounding your home. A lightning strike within a square mile area produces thunder which is heard less than about 3 seconds after the flash (see Chapter 12). You might try to determine for yourself the number of thunderstorm-days and strikes per square mile occurring in your area each year.

As is evident from Figs. 7.1 and 7.2, some parts of the world have a great deal of lightning while others have essentially no lightning. The degree of thunderstorm activity in a particular geographical area depends on its prevailing meteorological conditions.

Thunderstorms can usually be classified as (1) local or convective, or (2) organized traveling. Local thunderstorms form independently of one another and have lifetimes of an hour or two. They produce moderate lightning activity, moderate winds, rain, and sometimes small hail. Most lightning research has been concerned with lightning produced by local thunderstorms because they are relatively easy to study. A local thunderstorm is shown in Fig. 7.3. Organized thunderstorms are violent storms. They may last for many hours producing severe lightning activity, hail over an inch in diameter, winds exceeding 60 mph, and sometimes tornadoes. Such storms are often associated with cold fronts. They may line up along fronts for distances of tens or even hundreds of miles and travel along with the fronts.

What conditions lead to the formation of thunderstorms, and how are those thunderstorms formed? Consider the local or convective thunderstorm. The conditions necessary for the formation of a local thunderstorm are: (1) the air from ground level to many thousands of feet must be moist, (2) the atmosphere must be "unstable" (defined in the next paragraph), and (3) the sun must heat the ground and the air near the ground. In an unstable atmosphere, hot, wet air near the ground will rise to heights where the temperature is below freezing, eventually forming a thundercloud. Let's look now at what is meant by an unstable atmosphere.

If a volume or *parcel* of air is to rise in the atmosphere, it must be lighter than the air which surrounds it. Hot air balloons rise because hot air is less dense (lighter) than normal air. (Helium filled balloons rise because helium is less dense than air.) For a parcel of warm air to continue rising, it must always be at a higher temperature than the surrounding air. Temperature decreases with altitude in the atmosphere, so that the warm air parcel from the ground might be expected to rise continuously. This is not necessarily the case, however, since as the parcel rises it encounters decreasing air pressure. As a result, the parcel expands, and the act of expansion causes the parcel to cool. If the water vapor in the air parcel does not condense, then the cooling rate as the parcel rises will be about 5.5°F per 1000 ft. The air temperature must decrease faster than this if the parcel is to keep moving upward. The air temperature does not often decrease this rapidly much above the earth's surface. Fortunately for the thunderstorm, if the air parcel is moist, water vapor will condense as the parcel rises, and the process of condensation releases heat into the parcel. A moist parcel of air will cool at about 3.3°F per 1000 ft of ascent. If the atmospheric temperature decreases faster than 3.3°F per 1000 ft, the atmosphere is said to be "unstable". The parcel will continue to rise.

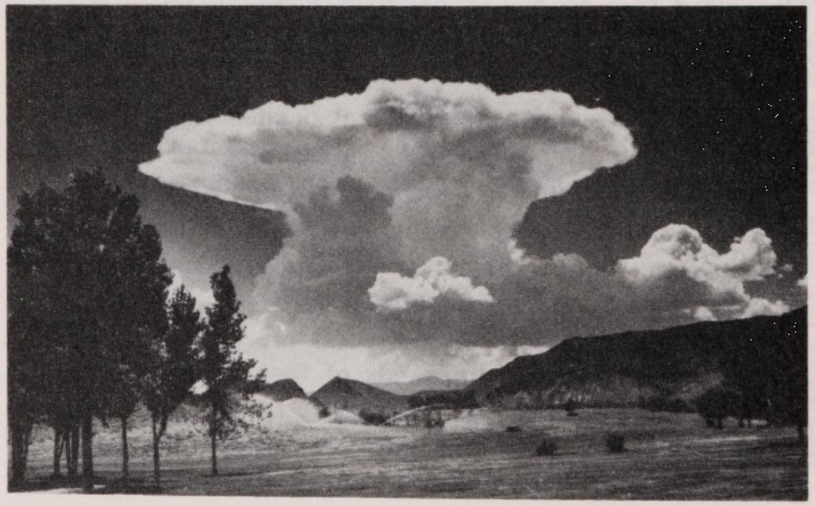

Fig. 7.3. A local or convective thunderstorm at Socorro, New Mexico. (Courtesy, Marx Brook, New Mexico Institute of Mines and Technology)

There are three stages in the life of a local or convective thundercloud: cumulus, mature, and dissipating. The cumulus stage begins when small, fluffy white, cumulus clouds form from rising air parcels. In time, many cumulus clouds combine to form a larger cloud, a cumulus congestus, perhaps a mile in diameter, with a well-defined top which rises at 1000 to 2000 ft per minute. The primary characteristic of the cumulus stage is that air motion throughout the cloud is upward. The cumulus stage lasts 10 to 15 minutes. As the cumulus congestus cloud grows, water drops and ice form within the cloud. Eventually the rain, hail, and snow within the cloud become sufficiently heavy that they can no longer be moved upward by the prevailing updrafts.

The formation of heavy precipitation signals the beginning of the mature stage. Precipitation begins to fall, dragging air downward with it. The cloud, now a cumulonimbus, contains both updrafts and downdrafts, and rain from the cloud reaches the ground. The top of a mature thundercloud may extend to 60,000 ft. It flattens out and assumes a characteristic anvil shape on reaching the stratosphere, the region of the atmosphere in which temperature is constant or increases with height. The mature stage of the storm lasts 15 to 30 minutes and is accompanied by considerable lightning activity. Finally the storm enters its dissipating stage. The violent updrafts and downdrafts decrease, and precipitation is less intense. The water droplets in the cloud evaporate, and the remainder of the cloud is blown away. The dissipating stage lasts about 30 minutes. The total lifetime of the convective thundercloud is roughly an hour.

A single convective thundercloud is technically referred to as a *cell.* Before one thundercloud cell has dissipated others may form. If the various cells form in close proximity, it may be impossible to distinguish between them from the ground. The result is a total storm of relatively long time duration and with relatively constant lightning activity – actually the combined result of the life cycles of many cells. An average local thunderstorm incorporates perhaps three thundercloud cells. Each is electrically active for about 15 to 30 minutes (during its mature stage) and, if the electrically active periods of the three cells do not overlap much, the storm produces appreciable lightning for about an hour and has a total duration of about two hours.

We look now at the traveling organized thunderstorms which produce severe weather. Typically, lines of severe thunderstorms are formed near the boundary of a moving cold front and stationary warm moist air. The cold front, a large mass of cold air, pushes beneath the warm moist air and forces it to rise. The rising air creates thunderstorms along the front. For the forming thunderstorms to become severe, the meteorological conditions must usually differ somewhat from those during the formation of local thunderstorms. Prior to the formation of severe thunderstorms there is often a warm dry layer of air aloft (and a region in which temperature increases with height) which tends to hold down the warm moist air near the earth's surface. The air near the ground becomes progressively warmer and more humid. When the cold front causes the warm air to rise, the ability of the dry air to hold down the wet is destroyed and thunderclouds are violently formed. It should be pointed out that sometimes severe thunderstorms occur under conditions that would normally lead to the generation of only local thunderstorms. There is some argument as to whether severe thunderstorms are composed of many individual cells each similar to a convective thundercloud cell or whether the structure of a severe storm is basically different. It appears now that some severe storms are composed of many individual cells each with a fairly well-defined life cycle while other severe storms do not fit this pattern at all.

One of the primary characteristics of the organized storms is that they can propagate themselves. There are several views of how the propagation process takes place. Common to most theories is the idea that cold air brought to ground by downdrafts through a given thundercloud spreads outward and forces warm moist air adjacent to that thundercloud upward generating a new thundercloud. It appears also that in order for severe storms to propagate effectively over long distances it is usually necessary for the winds aloft to be relatively strong and to increase with height. An excellent non-technical description of what is known about traveling thunderstorms, as well as what is known about local thunderstorms, is given in the paperback book *The Thunderstorm* written by L. J. Battan of the University of Arizona (see References).

chapter eight

Does Cloud Lightning Differ From Cloud-to-Ground Lightning? Which Is More Common?

Does Lightning Occur Only In Thunderstorms?

There are two principal types of lightning discharges – flashes which occur between the thundercloud and the earth (cloud-to-ground discharges) and flashes within the thundercloud (intracloud discharges). Other types of discharges such as cloud-to-cloud lightning and cloud-to-air lightning also occur but not very frequently. It is a widespread misconception that cloud-to-cloud lightning is common.

Although the thundercloud is the most common source of lightning, it is not the only one. Lightning occurs in snowstorms, in sandstorms, in non-thunderstorm rain and ice (see Chapter 4), in the ejected material above erupting volcanoes (Ref. 8.1; Fig. 8.1), near the fireballs created by nuclear explosions (Fig. 8.2), and apparently even out of the clear blue sky (giving rise to the expression "a bolt from the blue"). For lightning to occur, a region of the atmosphere must attain an electrical charge sufficiently large to cause electrical breakdown of the air. Some charging almost always occurs when any particulate matter (for example, dust or sand) is subjected to strong winds. (An obvious parallel to the frictional charging discussed in Chapter 1.) Within a typical thundercloud a turmoil of wind, water, and ice exists in the presence of a temperature which decreases with height. Small particles are carried upward by the wind; large particles move downward under the dominant influence of gravity. The various ascending and descending particles exhibit different velocities depending on their size. Particles that move at different velocities collide with one another, and out of these interactions, in ways not yet fully understood, emerge light,

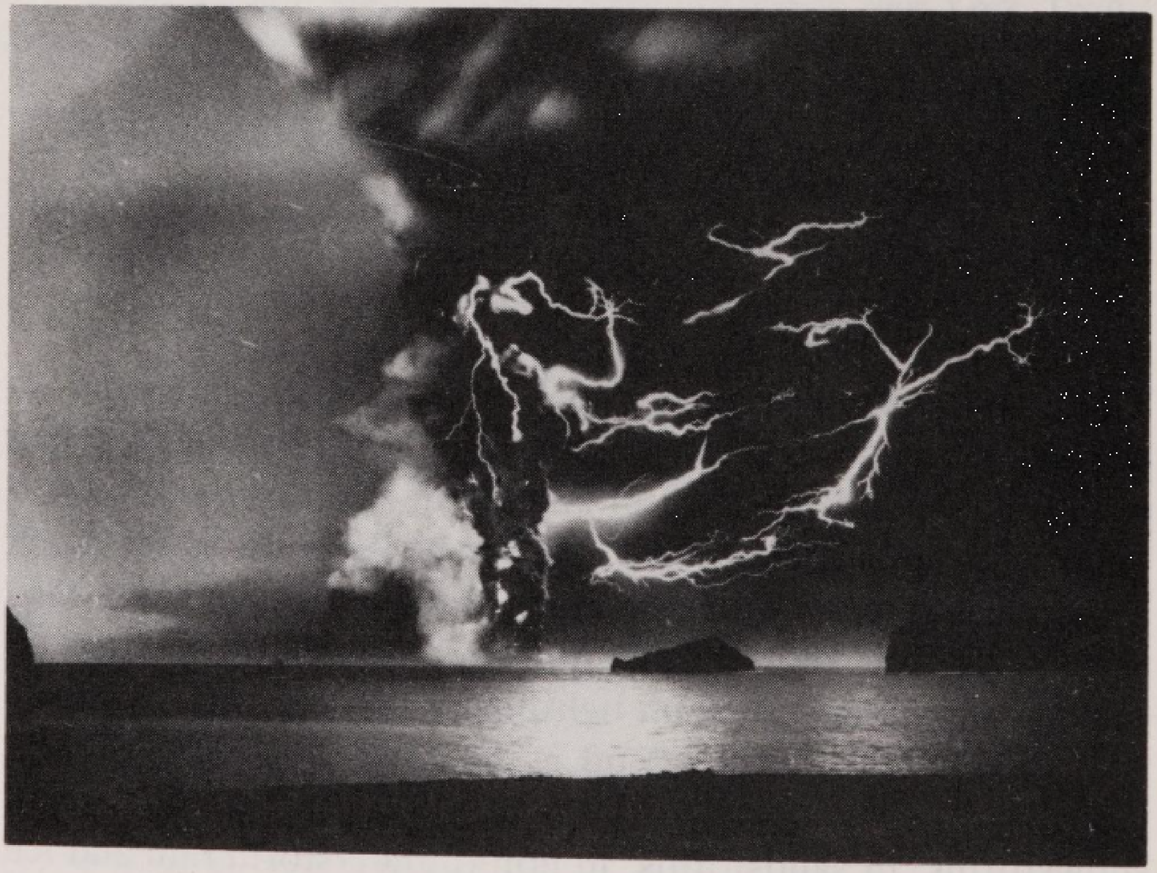

Fig. 8.1. Lightning in the volcano cloud over Surtsey, near Iceland, in December 1963. Volcanic eruptions were observed on November 14, 1963 off the southern coast of Iceland in water about 130 yards deep. Within 10 days an island over half a mile long and about 100 yards above sea level was formed. The island was named Surtsey by the Icelandic Government. (Courtesy, Sigurgeir Jónasson, Icelandair)

positively-charged particles which move upward and heavy, negatively-charged particles which move downward.

The probable distribution of cloud charge for a typical South African thundercloud cell is shown in Fig. 8.3 (Ref. 8.2). The upper part of the thundercloud carries a preponderance of positive (+) charge while the lower part carries a net negative (–) charge. The primary region of positive charge is referred to as the P-region, the primary region of

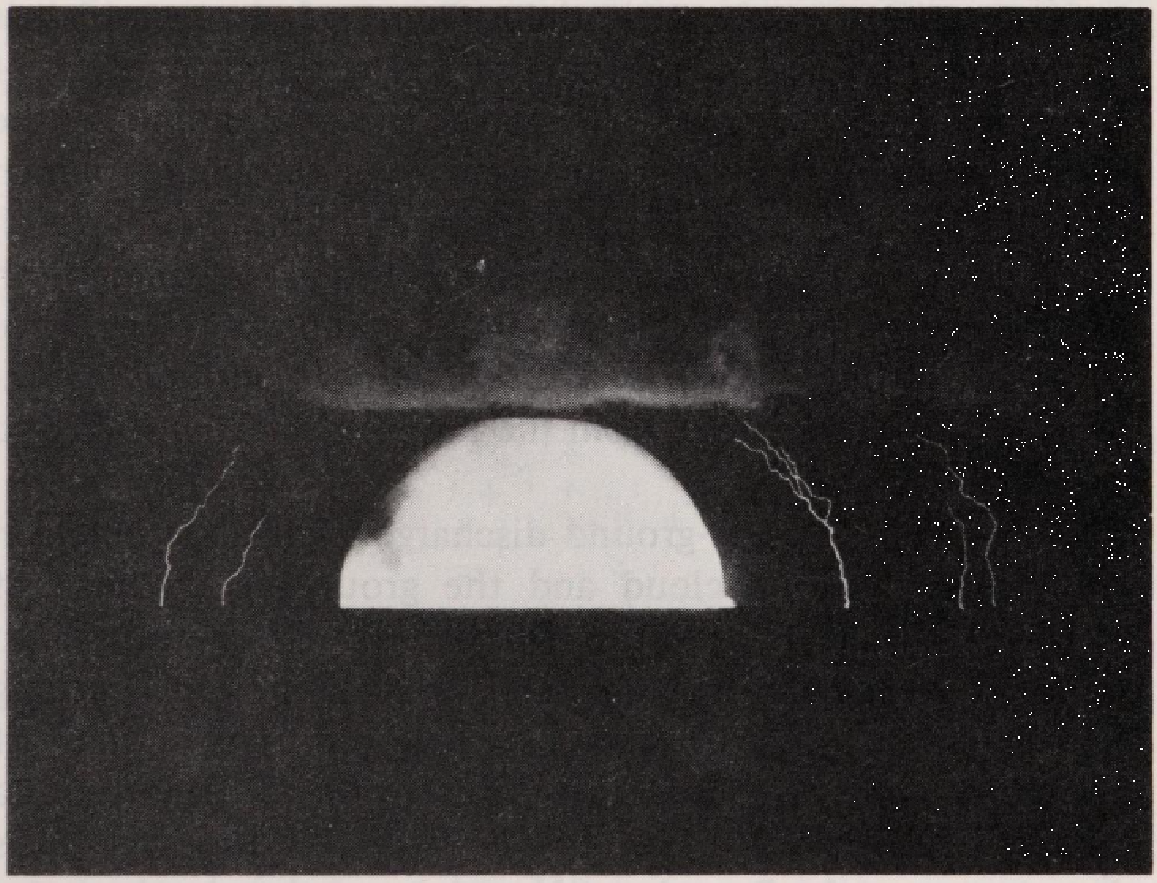

Fig. 8.2. Five lightning flashes induced by an experimental thermonuclear device exploded on October 31, 1952, at Eniwetok in the Pacific. Photograph is frame number 72 (detonation occurs in frame 1) of a 2000 frame-per-second movie taken from 20 miles away. The five flashes were initiated immediately after detonation probably from instrument bunkers projecting above sea level and propagated upward similar to typical flashes from tall buildings (see Chapter 6). All flashes lasted almost one-tenth of a second and apparently none contained subsequent strokes. The tops of the lightning channels bend towards the fireball. Fleecy cumulus clouds are visible with bases at 2000 ft. The clouds were not lightning producers. (Courtesy, U.S. Atomic Energy Commission)

negative charge the N-region. Recent research indicates that the N-region is not necessarily the vertical pillar of negative charge shown in Fig. 8.3, but rather may be a pillar which leans at an appreciable angle from the vertical (Ref. 8.3). In addition to the main cloud charges (P and N), there may be a small pocket of positive charge at the base of the thundercloud (the p-region). Exact values of the charges in the P-, N-, and p-regions are still a matter of controversy. Malan (Ref. 8.2) suggests P = +40 coulombs, N = -40 coulombs, p = +10 coulombs. (One coulomb is the amount of charge which is moved past a given cross-section of a wire when a current of 1 amp flows for 1 second.) Whatever the exact numbers, the charge in the cloud must be more than the charge which flows in the lightning channel since it is the cloud which provides the charge for the channel. A typical lightning flash to ground generally lowers about 25 coulombs of negative charge from the N-region of the cloud to ground. Occasionally discharges occur from the P-region or the p-region to ground.

The typical cloud-to-ground discharge takes place between the N-region of the cloud and the ground, a distance of roughly two miles. Often the cloud-to-ground lightning path occurs in clear (rain-free) air. The cloud-to-ground lightning flash is, as noted in Chapter 5, composed of a number of discrete strokes, the total time duration of the flash being a few tenths of a second. The typical intracloud discharge takes place between the P- and the N-regions of the cloud, over a distance of a mile or two. The discharge path is through ice and supercooled water (water below 32°F which is not frozen) since most of a typical thundercloud is at a temperature below the freezing level. While the total time duration, charge transfer, and length of an intracloud discharge are similar to that of a cloud-to-ground discharge, the discharge processes differ. This is so because of the different environments in which the two discharges occur and

because the cloud-to-ground discharge terminates on a conductor (the earth) while the intracloud discharge does not. Instead of being composed of a number of discrete current and light pulses (strokes), the typical intracloud discharge is fundamentally composed of a single slowly-moving spark or leader which bridges the gap between the N- and P-regions in a few tenths of a second. There is controversy as to whether the leader moves up from the N-region and carries negative charge or moves down from the P-region and carries positive charge. Each may in fact occur on different occasions. A low and continuous luminosity is

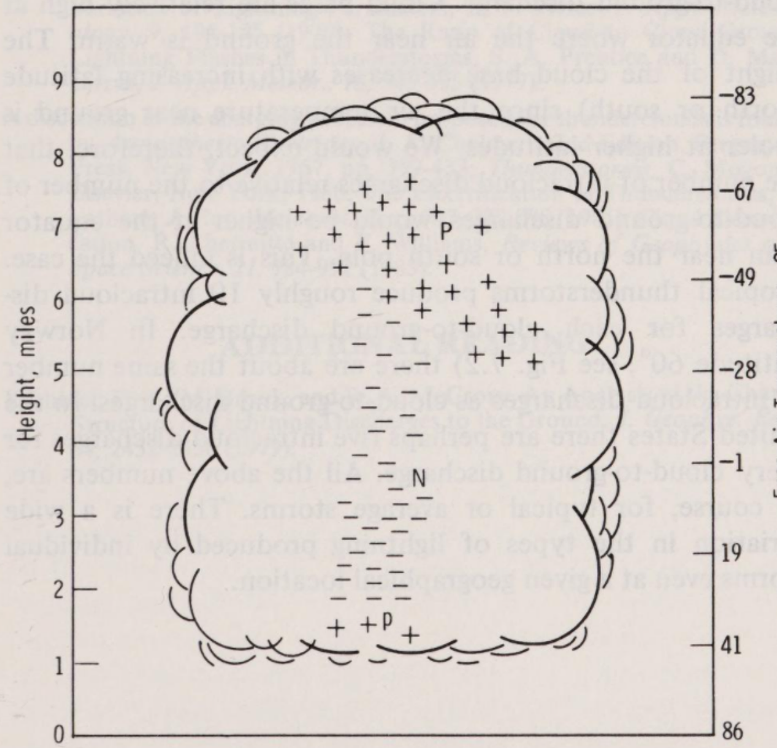

Fig. 8.3. Probable distribution of thundercloud charge according to D. J. Malan (Ref. 8.2).

observed in the cloud during the leader propagation. Superimposed on the continuous luminosity are a number of luminous pulses each lasting about a thousandth of a second. It is thought that these pulses, called K-changes, occur when the propagating leader contacts a pocket of charge opposite in polarity to that of the leader. Much more is known about cloud-to-ground discharges than about intracloud discharges: it is impossible to photograph adequately the intracloud discharge channel since it is obscured by the cloud.

The base of the N-region of a thundercloud is generally situated at the height at which freezing occurs (32°F). The closer the N-region is to ground, the more likely is a cloud-to-ground discharge. Cloud bases are relatively high at the equator where the air near the ground is warm. The height of the cloud base decreases with increasing latitude (north or south) since the air temperature near ground is cooler at higher latitudes. We would expect, therefore, that the number of intracloud discharges relative to the number of cloud-to-ground discharges would be higher at the equator than near the north or south pole. This is indeed the case. Tropical thunderstorms produce roughly 10 intracloud discharges for each cloud-to-ground discharge. In Norway (latitude 60°; see Fig. 7.2) there are about the same number of intracloud discharges as cloud-to-ground discharges. In the United States there are perhaps five intracloud discharges for every cloud-to-ground discharge. All the above numbers are, of course, for typical or average storms. There is a wide variation in the types of lightning produced by individual storms even at a given geographical location.

chapter nine

Does A Stroke Between Cloud And Ground Travel Upwards or Downwards?

The answer to the question "Does lightning between cloud and ground go upwards or downwards?" is that, in a sense, it does both. The usual lightning flash between cloud and ground (excluding the discharges initiated by tall structures discussed in Chapter 6) begins with a visually-undetected downward-moving traveling spark called the stepped leader (see Chapters 2 and 6). Since the lightning flash begins with a downward-moving discharge, lightning moves from the cloud to the ground. On the other hand, when the stepped leader reaches ground (or is contacted by an upward-moving discharge some tens of yards above the ground) the leader channel first becomes highly luminous at the ground and then at higher and higher altitudes. The bright, visible channel, or so-called *return stroke,* is formed from the ground up, and one could say, therefore, that visible lightning moves from the ground to the cloud.

It is thought by most lightning researchers that the usual cloud-to-ground discharge begins as a local discharge between the p-charge region in the cloud base and the N-charge region above it (Fig. 9.1, see also Fig. 8.3). This discharge frees *electrons* in the N-region previously immobilized by attachment to water or ice particles. (Electrons are fundamental particles which carry the smallest known unit of negative electrical charge.) Because of their small mass, free electrons are extremely mobile compared to air atoms or to charged ice or water particles. The free electrons overrun the p-region, neutralizing its small positive charge, and then continue their trip toward ground. The vehicle for moving the negative charge to earth is the stepped leader.

Exactly how the stepped leader works is not understood. What is known is that it moves from cloud to ground in rapid luminous steps about 50 yards long. In Figs. 9.1 and 9.2 the luminous steps appear as darkened tips on the less-luminous leader channel which extends upward into the cloud. Each leader step occurs in less than a millionth of a second. The time between steps is about 50 millionths of a second. Negative charge is continuously lowered from the N-region of the cloud into the leader channel. The average velocity of the stepped leader during its trip toward ground is about 75 miles per second with the result that the trip between cloud and ground takes about 20 thousandths of a second. A typical stepped leader has about 5 coulombs of negative charge distributed over its length when it is near ground. To establish this charge on the leader channel an average current of about 100 or 200 amperes must flow during the whole leader process. The pulsed currents which flow at the time of the leader steps probably have a peak current of about 1000 amperes. The luminous diameter of the stepped leader has been measured photographically to be between 1 and 10 yards. It is thought that most of the stepped-leader current flows down a narrow conducting core less than an inch in diameter at the center of the observed leader. The large photographed diameter is probably due to a luminous electrical corona surrounding the conducting core.

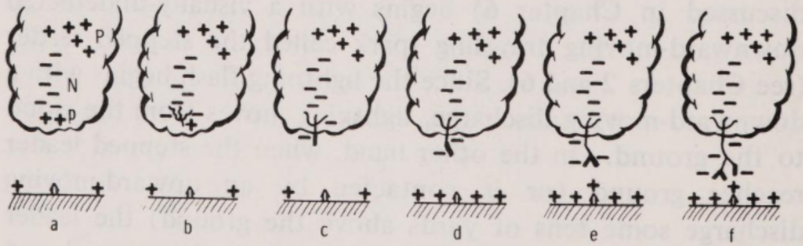

Fig. 9.1. Stepped leader initiation and propagation. (a) Cloud charge distribution just prior to p-N discharge. (b) p-N discharge. (c)-(f) Stepped leader moving toward ground in 50-yard steps. Time between steps is about 50 millionths of a second. Scale of drawing is distorted for illustrative purposes.

When the stepped leader is near ground, its relatively large negative charge induces large amounts of positive charge on the earth beneath it and especially on objects projecting above the earth's surface (Fig. 9.2). Since opposite charges attract each other, the large positive charge attempts to join the large negative charge, and in doing so initiates upward-going discharges (Figs. 2.1 and 9.2). One of these upward-

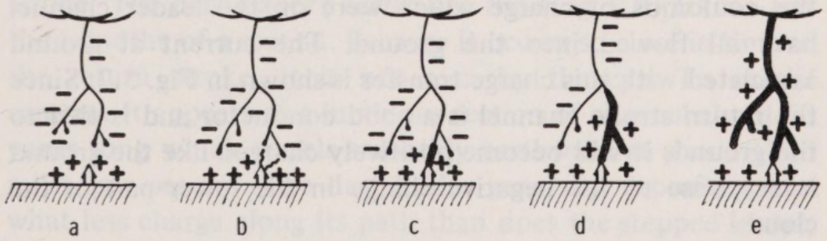

Fig. 9.2. Return stroke initiation and propagation. (a) Final stages of stepped leader descent. (b) Initiation of upward-moving discharges to meet downward-moving leader. (c)-(e) Return stroke propagation from ground to cloud. Return stroke propagation time is about 100 millionths of a second; propagation is continuous. Scale of drawing is distorted.

going discharges contacts the downward-moving leader and thereby determines the lightning strike-point. When the leader is attached to ground, negative charges at the bottom of the channel move violently to ground, causing large currents to flow at ground and causing the channel near ground to become very luminous. Since electrical signals (or any signals, for that matter) have a maximum speed of 186,000 miles per second – the speed of light – the leader channel above ground has no way of knowing for a short time that the leader bottom has touched ground and has become highly luminous. The channel luminosity, the return stroke, propagates continuously up the channel and out the channel branches at a velocity somewhere between 20,000 and 60,000 miles per second, as shown in Fig. 9.2. The trip between ground and cloud takes about 100 millionths of a second. As the return stroke luminosity moves upward so also does the region of high current. When the leader initially touches ground, copious numbers of electrons flow to ground from the channel base. As the return stroke moves upward, large numbers of electrons flow at greater and greater heights. Electrons at all points in the channel *always move downward* even though the region of high current and high luminosity moves upward. Eventually, in some thousandths of a second, the coulombs of charge which were on the leader channel have all flowed into the ground. The current at ground associated with this charge transfer is shown in Fig. 5.7. Since the return stroke channel is a good conductor and is tied to the ground, it will become positively charged like the ground in response to the negative charge in the lower part of the cloud.

It is the return stroke that produces the bright channel of high temperature (Chapter 11) that we see. The eye is not fast enough to resolve the propagation of the return stroke and it seems as if all points on the channel become bright simultaneously. The reason we do not visually detect the stepped leader preceding a first return stroke is because the eye cannot resolve the time between when the weakly luminous leader is formed and the bright return stroke illuminates the leader channel. The return stroke produces most of the thunder we hear (see Chapter 12).

After the stroke current has ceased to flow, the lightning flash may be ended, in which case the discharge is called a single-stroke flash. As noted in Chapter 5, most flashes contain three or four strokes, typically separated by gaps of 40 or 50 thousandths of a second. Strokes subsequent to the first are initiated only if additional charge is made available to the top of the previous stroke channel less than about 100 thousandths of a second after current has stopped flowing in the previous stroke. Additional charge can be made available to the channel top by the action of electrical discharges (so called *K-streamers* and *J-streamers*) which move upward from the top of the previous return stroke into higher areas of the N-charge region of the cloud (Fig. 9.3). When this additional charge is available, a continuous (as opposed to stepped) leader, known as a *dart leader*, moves down the defunct return stroke channel again depositing negative charge from the N-region along the channel length. The dart leader thus sets the stage for the second (or any subsequent) return stroke. The dart leader's earthward trip (Fig. 9.4) takes a few thousandths of a second. Because it occurs so close in time to the return stroke, it is not seen by the eye. To special cameras it appears as a luminous section of channel about 50 yards long which travels smoothly earthward at about 1000 miles per second. The dart leader generally deposits somewhat less charge along its path than does the stepped leader, with the result that subsequent return strokes generally lower less charge to ground and have smaller currents (Fig. 5.7) than first strokes.

The first stroke in a flash is usually strongly branched downward because the stepped leader is strongly branched

(Figs. 9.1 and 9.2). Dart leaders generally follow only the main channel of the previous stroke and hence subsequent strokes show little branching (Fig. 9.4).

The time between strokes which follow the same path can be tenths of a second if a continuing current (Chapter 5) flows in the channel between strokes. Apparently, the channel is ripe for a dart leader only after the continuing current has terminated.

A typical cloud-to-ground discharge lowers about 25 coulombs of negative charge from the N-region of the cloud to the earth. This charge is transferred in a few tenths of a second by the three or four component strokes and any continuing current which may flow. While the leader-return stroke process transfers charge to ground in two steps (charge is put on the leader channel and then is discharged to ground), the continuing current represents a relatively steady charge flow between the N-region and ground.

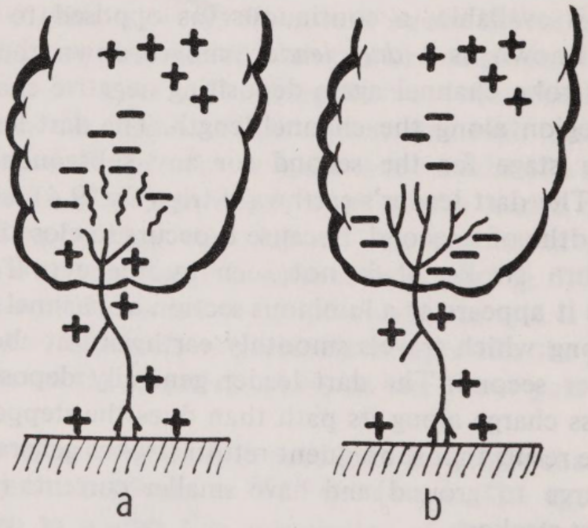

Fig. 9.3. K-streamers and J-streamers making more negative charge available to the channel top during the 50 thousandths of a second or so following the cessation of current flow in the first return stroke. Scale of drawing is distorted.

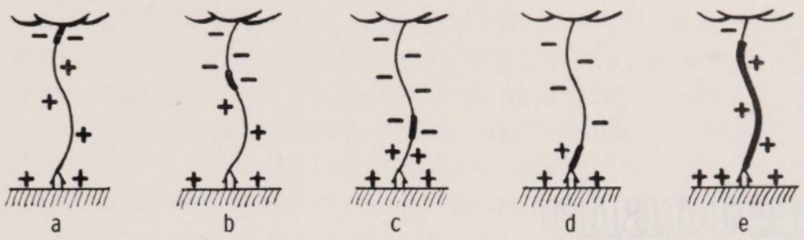

Fig. 9.4. Dart leader and subsequent return stroke. (a)-(c) Dart leader deposits negative charge on defunct first-stroke channel during its thousandth-of-a-second trip to ground. (d)-(e) Subsequent return stroke propagates from ground to cloud in about 100 millionths of a second. Scale of drawing is distorted.

Thus far in this chapter we have discussed the usual stepped leader which lowers negative charge between the cloud and the earth. Occasionally downward-moving stepped leaders are observed that lower positive charge. Currents due to the resulting "positive" return strokes have been measured directly during discharges to instrumented towers. Positive strokes are characterized by rates-of-rise of current at the ground roughly five times slower than those for typical negative strokes, and by charge transfers roughly three times greater than those of typical negative strokes. The maximum measured charge transfer due to a positive stroke is 300 coulombs. Positive strokes are probably initiated between the P-charge region of the cloud and ground when strong winds blow the cloud such that the P-region is brought relatively close to a mountain side or to the earth's surface. Positive discharges rarely consist of more than one stroke.

REFERENCES

The material contained in Chapter 9 is considered in more detail in the following three references presented in order of decreasing technical content: *Lightning*, M. A. Uman, Dover Publications, Inc., New York, 1984. *Physics of Lightning*, D. J. Malan, The English Universities Press Ltd., London, 1963. *The Flight of Thunderbolts*, B. F. J. Schonland, 2nd Edition, Clarendon Press, Oxford, 1964.

chapter ten

How Long And How Wide Is The Lightning Channel?

Why Is Lightning Zig-Zag?

How Can I Best Photograph Lightning?

Let's begin with "how long". Channels longer than 90 miles have been observed (Ref. 10.1). On the other hand, sparks of a few yards length have been seen in and around clouds, and probably even shorter "lightning" exists.

There are basically four ways of determining channel length. (1) The most straightforward way is by photographing lightning and then determining its length from the photograph. Since channels inside clouds cannot be photographed, the value of this technique for determining channel length is somewhat restricted. (2) Radar has been used to measure channel length. A radar set sends out electromagnetic pulses which are reflected back by certain objects (e.g., metallic airplanes). Since the pulses travel at 186,000 miles per second, measuring the elapsed time between the emission of a pulse and the reception of its reflection enables the distance to the reflector to be measured. An actual radar set has a TV-like screen on which appears images of the reflecting objects. A lightning channel is a good conductor and hence reflects radar pulses much as an airplane does. Radar wavelengths can be chosen to minimize reflection of radar pulses by the ice and water in clouds so that a good image of the lightning channel can be obtained on the radar screen. (3) Channel lengths can be determined from the duration of thunder. The minimum possible channel length associated with a thunder duration of T seconds is roughly T/5 miles (see Chapter 12). For example, if thunder lasts 20 seconds, the channel producing it has a length of about 4 miles or more. (4) Finally, there are a number of ways of determining lightning channel height within a cloud from electrical measurements made on the ground (Ref. 10.2). Charge motion in the lightning channel or in the cloud will induce voltages and currents in ground-based instruments enabling properties of the channel and cloud charge to be determined. For example, the vertical height of the J-streamers and K-streamers occurring in the cloud between strokes (Fig. 9.3) can be determined from electrical measurements made at ground.

Now that we know how channel lengths are measured, let's look at the results of the measurements. The average vertical stroke height is about 3 to 4 miles. The maximum vertical extent of the usual stroke to ground is the top of the N-charge region, a height of 6 miles or more. Each stroke in a multiple-stroke flash averages about a third of a mile longer than the preceding one. This is the case because, in order to obtain negative charge for a new stroke, J-streamers and K-streamers tap new areas of the N-charge region of the cloud (Figs. 8.3 and 9.3) during the time between strokes, and these tapped regions become part of the new stroke. The horizontal component of the lightning channel within a cloud or through several clouds may be larger than the vertical component. Horizontal sections of 5 to 10 miles within clouds are not uncommon. Radar studies have shown horizontal channels longer than 90 miles (Ref. 10.1). Long horizontal channels must feed on the charge of many thundercloud cells. In addition to cloud-to-ground discharges with long horizontal in-cloud components or similar long intracloud or cloud-to-cloud discharges which do not reach ground, long lightning discharges have occasionally been observed to occur from the tops of thunderclouds upward (Ref. 10.2). These vertical air discharges may be tens of miles in length.

When we speak of the length of lightning, we normally picture long sections of the channel as being fairly straight. Photographs show, however, that the channel is very tortuous on almost any scale. There are zigs and zags 100 yards long and, within these, other zigs and zags 10 yards long, and within these yet smaller zigs and zags (Refs. 10.3, 10.4, 10.5). Very detailed photographs have shown lightning channels that twist and bend on a distance scale measured in inches (Ref. 10.5). Figure 10.1 shows a section of a lightning channel with a height of about 1000 yards; Fig. 10.2 shows about 200 yards of channel; Fig. 10.3 about 10 yards. If we could grab a long section of the channel and pull on both ends to straighten it out, the channel section might well be two or more times longer than we initially thought.

Lightning diameter measurements have been made in two ways: (1) by examining the interaction between lightning and objects and (2) from photographs. Photographic measurements of lightning almost always overestimate the luminous diameter of the channel: the bright channel overexposes the film making the recorded image broader than it should be. It is extremely difficult to expose a lightning channel photograph properly. The best channel diameter photographs (Ref. 10.5) yield diameters of between about 2 and 7 in., and these values are probably overestimates. Lightning photography is discussed again at the end of this chapter.

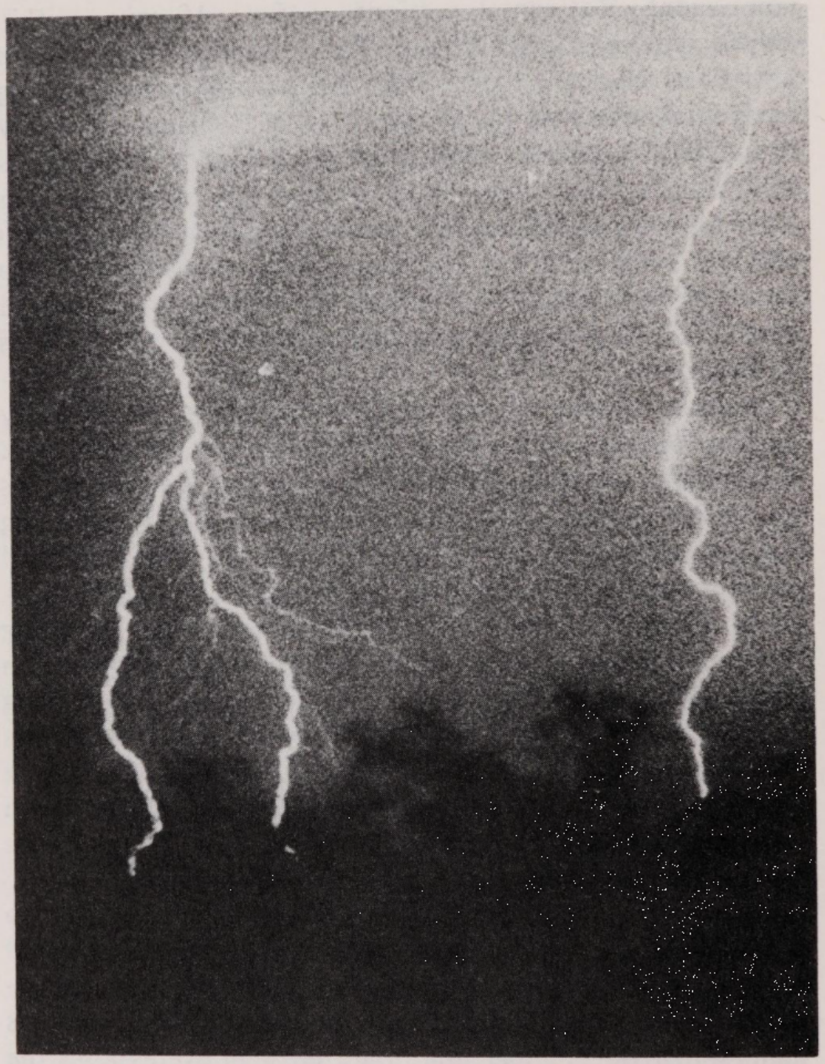

Fig. 10.1. About 1000 vertical yards of lightning channel near Tucson, Arizona (Ref. 10.4). Note the large scale channel tortuosity. (Courtesy, H. B. Garrett and A. A. Few, Rice University)

When lightning hits an object such as a tree or a rock, it leaves visible damage. In many cases the damage can be related to channel diameter. Taylor (Ref. 10.6; see also Chapter 5) found that the lightning-caused furrows that spiral along tree trunks are between about 0.6 and 5.0 in. wide. The lightning diameter, therefore, must be roughly this size or smaller. If it were much bigger, it is hard to see how it could create the furrows.

When lightning strikes sand or certain kinds of rocks, the channel heat melts the material along its path. When the melted material solidifies, the resulting *fulgurite* (from the Latin *fulgur*, meaning lightning) represents a permanent record of the lightning diameter and path. Fulgurites in dry

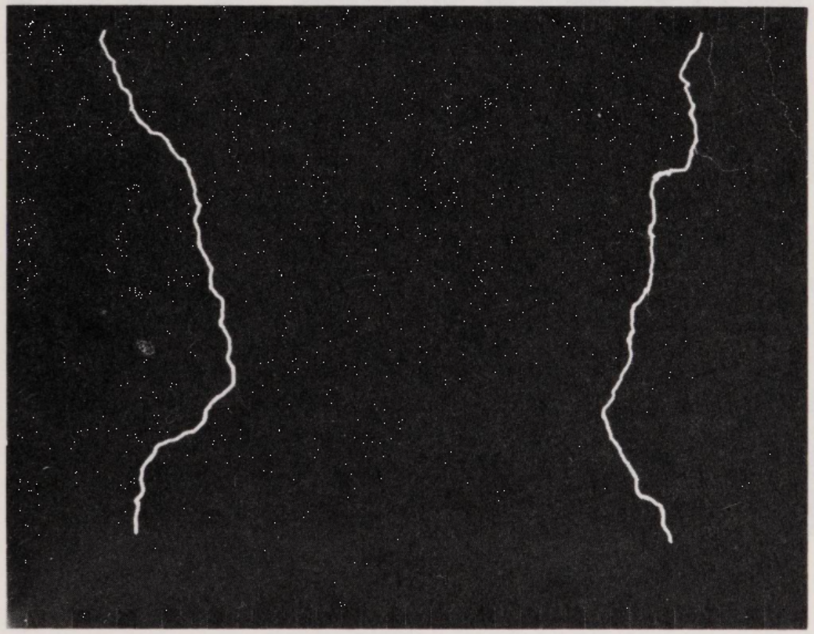

Fig. 10.2. Bottom portion of the split channel shown in Figure 10.1 (Ref. 10.4). Vertical height shown is about 200 yards. Tortuosity on a scale smaller than 10 yards is evident. (Courtesy, Leon E. Salanave, University of Arizona)

sand are generally long hollow tubes with corrugated glassy walls. An artificial fulgurite produced by an electrical discharge in the laboratory is shown in Fig. 10.4. Fulgurites have been traced downwards into sand as much as 20 yards. Diameters are usually 0.5 to 2 in. From the corrugated appearance of fulgurite walls, it is probable that the initial diameter collapses somewhat on cooling. "Fossil" fulgurites 250 million years old have been discovered (Ref. 10.7). The

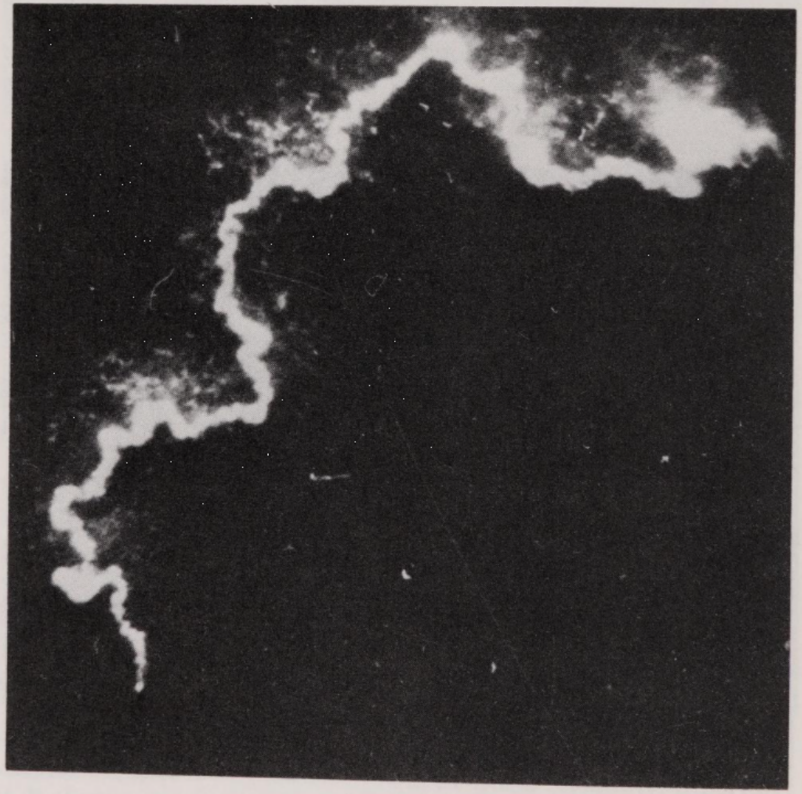

Fig. 10.3. One frame from a high-speed movie of lightning striking a TV tower near Tucson, Arizona (Ref. 10.5). About 10 yards of channel are shown and tortuosity on an inch scale is evident. (Courtesy, W. H. Evans, University of Arizona, and R. L. Walker, University of Florida, Cape Canaveral)

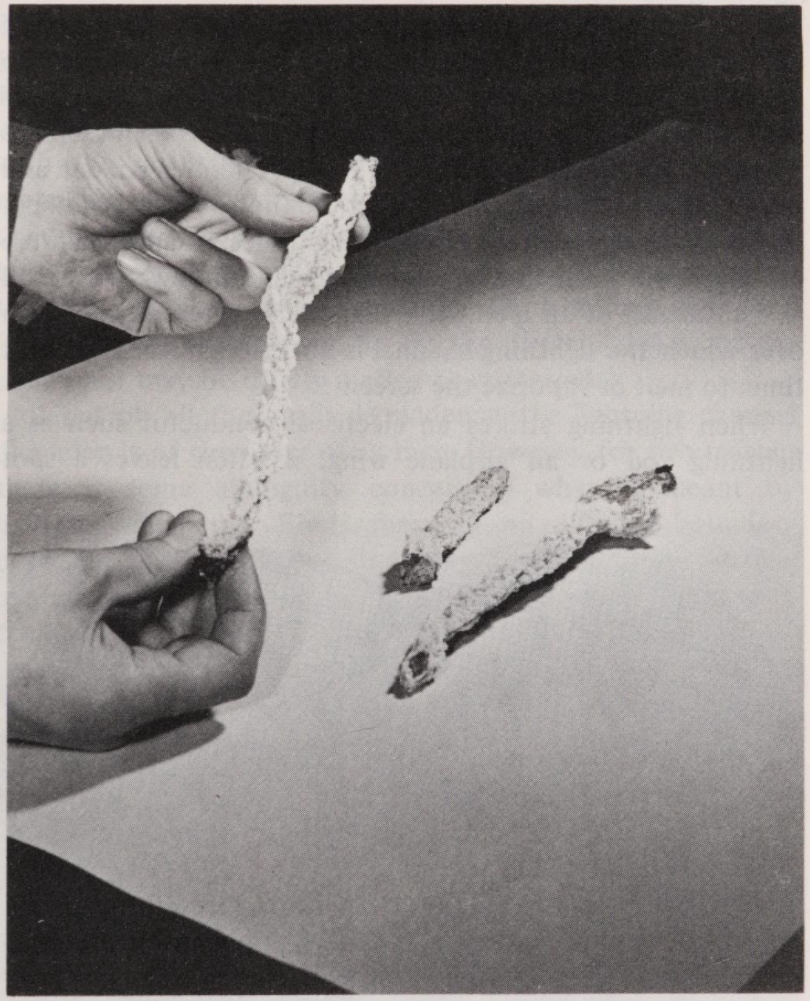

Fig. 10.4. Artificial fulgurites produced by P. L. Bellaschi, Westinghouse Electric Corporation, by allowing an electrical discharge to pass through quartz sand. The hollow tubes of solidified sand are close replicas of natural fulgurites.

fulgurite diameter probably does not differ much from the lightning diameter.

The author has measured lightning diameters by allowing lightning to pass through Fiberglas screens on the way to lightning rods and then measuring the size of the holes melted in the screen by the lightning (Ref. 10.8). The screens covered rods on the tops of towers near Tucson, Arizona and near Lugano, Switzerland. About 50 holes were obtained. Figure 10.5 shows typical lightning holes. Diameters were usually about an inch or smaller and sometimes were as small as a sixteenth of an inch. Hole diameters represent the width over which the lightning channel is hot enough for sufficient time to melt or vaporize the screen.

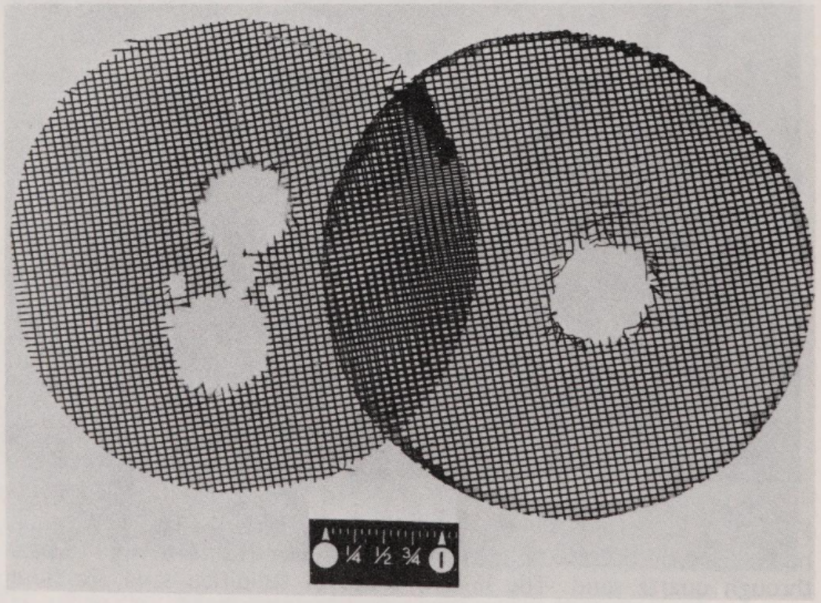

Fig. 10.5. Holes melted in two Fiberglas screens by lightning. At least four strokes passed through the screen on the left. One stroke passed through the screen on the right.

When lightning strikes an electrical conductor such as a lightning rod or an airplane wing, it often leaves a spot

(sometimes a raised lump, sometimes a depression) about an eighth of an inch wide. Examples of such spots on three copper disks are shown in Fig. 10.6. The size of a lightning spot is roughly the diameter of the lightning channel when it enters the conductor. At the contact point the lightning current must flow through a mixture of metal vapor and air, with the result that its diameter is smaller than in the metal-free air away from the conductor. Thus, spot sizes on conductors provide a lower diameter limit. It should be noted that a large lightning current flowing for a relatively long time may do considerably more damage to a conductor than just a spot. A hole may be burned (Figs. 4.1-4.3) or the conductor may be totally melted or vaporized.

Based on all the available evidence, the lightning channel diameter is of the order of an inch. However, for two reasons there is some ambiguity concerning what is meant by lightning diameter. First, there is no definite boundary between the stroke and the air surrounding the stroke; second, the diameter must change with time in response to the changing current.

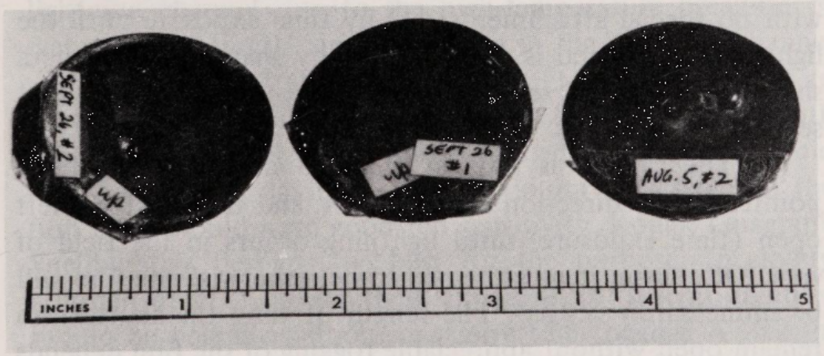

Fig. 10.6. Spots produced by lightning strikes to three copper disks which were placed on top of TV towers near Tucson, Arizona. The center disk was apparently hit by three lightning strokes, one of which raised a one-eighth inch lump. The disk on the right exhibits a copper lump apparently formed as the lightning channel moved back and forth across the disk.

Why is the lightning channel so tortuous? The answer is not known, but some reasonable guesses may be made. The larger-scale tortuosity in the channel (representing, say, tens of yards or more) is due to the fact that the stepped leader makes such an errant trip to ground. Why does it do this? Possibly various airborne regions of charge (space charge, see Chapter 6) divert the leader on its trip. More likely, the leader just doesn't know exactly where it wants to go, except that ultimately it wants to move downward. The smaller-scale tortuosity in the channel may be formed when the leader steps are formed, or may be formed by action of the magnetic forces associated with the return stroke current. As is probably becoming more and more apparent with each succeeding chapter in this book, there is an awful lot we still don't know about lightning.

Since we have discussed in this chapter the use of photography as a scientific tool to determine lightning channel parameters, it is appropriate to say a few words about how to take lightning photographs. There are basically two ways to photograph lightning using ordinary cameras with no special attachments: (1) by time exposure until the lightning occurs and is over, and (2) by snapping the camera shutter (set at 1/30 sec or slower) when the lightning is first seen.

(1) The camera is put on a tripod or other firm base, pointed in the direction of the storm, and the shutter is left open (time exposure) until lightning occurs in the field of view. Of course, it must be nighttime, otherwise the film will be immediately fogged. The length of time the shutter can remain open without appreciable fogging of the film depends on the ambient light level. Away from city lights with a film speed of ASA-100 (roughly the speed of Kodak Plus-X black and white film or Kodacolor-X color film) and an aperture opening of f/11, several minutes are possible. Camera focus should be set at infinity. Since there are several lightning strokes to ground during each minute of the electrically active period of a storm (Chapter 7), it is not difficult to obtain photographs. (2) If the lightning flash to be photographed consists of only one stroke, your reflexes will not be fast enough to fire the shutter in time to record the flash. However, if it is a multiple-stroke flash, as most flashes are, you will record a subsequent stroke or strokes. Since subsequent strokes occur at intervals of roughly 1/20 of a second, a shutter speed of 1/30 second or slower is necessary to avoid taking a picture of what goes on between strokes. Since subsequent strokes are generally unbranched (Chapter 9), the photographed lightning will be unbranched and will look somewhat peculiar. Again, for a film speed of ASA-100 try an aperture of f/11 with the focus set at infinity.

The two basic techniques for photographing lightning can be improved by using special equipment with ordinary cameras. (1) Time exposures of lightning in daylight are possible if the proper filter is used (Refs. 10.9, 10.10). Light from the sun is weak at a wavelength of 6563Å* (in the red) because hydrogen in the sun's outer layers absorbs that wavelength, the so-called $H\alpha$ line. Hydrogen is present in the lightning channel due to the high-temperature break-up of water vapor (H_2O). Since hydrogen in the lightning channel radiates at 6563 Å ($H\alpha$) (see Chapter 11), it follows that an $H\alpha$ filter that passes only light near this wavelength both allows the channel to be photographed and prevents rapid film fogging by daylight. For maximum effectiveness, the film used should have good red sensitivity. Kodak Linagraph Shellburst is perhaps the best for the purpose. In cloudy weather at f/11, a daytime exposure of several minutes should be possible without excessive film fogging. A 2x2 in. $H\alpha$ filter with a bandwidth of about 40 Å costs about $200. (2) Triggering the camera shutter can be accomplished with instruments that respond to the electrical signals generated in the cloud or by the stepped leader prior to the first return stroke (Refs. 10.11, 10.12). In this way the first stroke can be photographed. Essentially, the shutter is tripped by a solenoid coupled to a radio antenna and appropriate electronics.

---

*One Å (Ångstrom) equals one ten-billionth of a centimeter. There are 2.54 centimeters to the inch.

REFERENCES

10.1. Ligda, M., The Radar Observation of Lightning, *J. Atmospheric Terrest. Phys.*, *9*, 329-346 (1956).

10.2. Uman, M. A., *Lightning*, McGraw-Hill Book Company, New York, New York, 1969, pp. 2, 81-85.

10.3. Hill, R. D., Analysis of Irregular Paths of Lightning Channels, *J. Geophys. Res.*, *73*, 1897-1906 (1968).

10.4. Few, A. A., H. B. Garrett, M. A. Uman, and L. E. Salanave, Comments on Letter by W. W. Troutman 'Numerical Calculation of the Pressure Pulse from a Lightning Stroke', *J. Geophys. Res.*, *75*, 4192-4195 (1970).

10.5. Evans, W. H., and R. L. Walker, High Speed Photographs of Lightning at Close Range, *J. Geophys. Res.*, *68*, 4455-4461 (1963).

10.6. Taylor, A. R., Diameter of Lightning as Indicated by Tree Scars, *J. Geophys. Res.*, *70*, 5693-5695 (1965).

10.7. Harland, W. B., and J. L. F. Hacker, 'Fossil' Lightning Strikes 250 Million Years Ago, *Advancement of Science*,*22*, 663-671 (1966).

10.8. Uman, M. A., The Diameter of Lightning, *J. Geophys. Res.*, *69*, 583-585 (1964).

10.9. Salanave, L. E., and M. Brook, Lightning Photography and Counting in Daylight, Using H_α Emission, *J. Geophys. Res.*, *70*, 1285-1289 (1965).

10.10. Krider, E. P., Comment on Paper by Leon E. Salanave and Marx Brook, 'Lightning Photography and Counting in Daylight, Using H_α Emission,' *J. Geophys. Res.*, *71*, 675 (1966).

10.11. Schonland, B. F. J., and J. S. Elder, Anticipatory Triggering Devices for Lightning and Static Investigations, *J. Franklin Inst.*, *231*, 39-47 (1941).

10.12. Hawe, R. G., Electrostatic Trigger Used for Daylight Lightning Photography, *Photographic Science and Engineering*, *12*, 219-221 (1968).

chapter eleven

How Hot Is Lightning?

The lightning return stroke is more than four times hotter than the surface of the sun. The peak temperature of lightning is greater than 50,000°F. The surface of the sun has a temperature of about 11,000°F.

Exactly how hot the lightning return stroke channel gets is not known. The best peak temperature measurements made to date represent the average temperature existing during a time interval of a few millionths of a second. For time varying systems, average values are necessarily less than peak values. For example, the lightning temperature could be very high for a few billionths of a second and relatively low for the remainder of the millionths-of-a-second measuring time interval. The resultant average temperature would then be close to the low value. Lightning temperature has been determined by examining the characteristics of the light emitted by the channel.

The lightning channel is composed of very hot air. Air at room temperature is made up primarily of nitrogen and oxygen molecules. A nitrogen molecule consists of two joined nitrogen atoms; an oxygen molecule of two joined oxygen atoms. The nitrogen atom has a core or nucleus containing seven positively-charged particles (protons) and seven uncharged particles (neutrons). Orbiting the nucleus at relatively large radii are seven negatively-charged electrons. (The electrons circle the nucleus much as the earth and other planets circle the sun.) The oxygen atom consists of eight protons, eight neutrons, and eight electrons. As the temperature of air at atmospheric pressure is increased above room temperature to about 12,000°F, the heat causes the molecules to break apart into atoms. With a further increase in temperature to about 30,000°F each atom loses one of its electrons, and with higher temperatures still (say above 60,000°F), more electrons are lost. An atom which has lost an electron or electrons is called an *ion*. If one electron is lost, the atom is said to be singly ionized; if two, doubly ionized, and so on. The lightning channel is composed primarily of nitrogen atoms and ions, oxygen atoms and ions, and free electrons. How do we know this?

If sunlight is passed through a prism, the light is decomposed into its component colors or spectrum. (The rainbow is the sun's spectrum produced by water droplets.) Sunlight is composed of a range of colors varying in hue continuously from red to orange to yellow to green to blue to violet. Incandescent solids (e.g., tungsten filaments) or incandescent gases at very high pressure or occupying large volumes (e.g., the sun) produce continuous spectra; that is, light which includes all colors. Luminous gases at lower pressure or occupying small volumes (e.g., the lightning channel) produce spectra with "lines" at characteristic wavelengths. A given molecule, atom, or ion is capable of emitting radiation only at specific wavelengths which are different for different kinds of particles. Each molecule, atom, or ion has its own unique radiation "signature" by which it can be positively identified. For example, in the visible wavelength region a nitrogen atom radiates primarily one red spectral line while a nitrogen ion radiates primarily a number of blue lines. We can get a rough idea of the temperature of air by determining from the radiation emitted by the air whether it is predominantly composed of molecules, atoms, or ions. A more exact determination can be made by examining the intensities of the characteristic spectral lines from a given molecule, atom, or ion.

A spectrometer is an instrument that disperses the light from a source so that characteristics of the component spectrum can be measured. The dispersion is effected by either a prism or a diffraction grating. A drawing of a lightning spectrometer is shown in Fig. 11.1. Light from the lightning channel passes through the transmission diffraction grating and is split into its spectral components. These components as well as an undiffracted image of the lightning channel are focused by a lens onto the channel isolator slit assembly. For illustrative purposes, the lightning in Fig. 11.1 is assumed to contain only three spectral lines: a red, a green, and a blue. Note that each of the spectral lines has the shape of the lightning channel from which it came. If a stationary piece of photographic film were put in front of the channel isolator slit assembly, a photograph like that in Fig. 11.2 would be obtained. The photograph shown was taken by placing an inexpensive transmission grating† over the lens of a 35 mm camera while taking a time exposure (see Chapter 10) until lightning occurred. In Fig. 11.2, the red line to the far left (the H_α line discussed in the previous chapter) is emitted by hydrogen atoms present in the channel due to the break-up of water vapor (H_2O). Some water vapor is always present in the air; quite a lot may be present during a thunderstorm. The red line just to the right of the hydrogen line is emitted by nitrogen atoms. Most of the remaining lines are emitted by singly ionized nitrogen atoms.

---

†A page-size sheet of diffraction grating can be purchased for a few dollars. One supplier is Edmund Scientific Co., Barrington, N.J.

Spectra such as that in Fig. 11.2 yield no information about what is happening in the lightning channel as a function of time, since all the light of a given wavelength is recorded at the same place on the film no matter when it is emitted by the channel. With the help of Fig. 11.1 we see how to obtain time-resolved lightning spectra. The channel isolator slit allows the spectrum emitted by a short section of the lightning channel to fall on a piece of moving film at the back of the slit. Thus the spectrum is streaked out in time on the moving film. In the example shown, the green line appears at the earliest time followed by the blue and then the red lines. A time-resolved spectrum of a long spark in air is shown in Fig. 11.3 (Ref. 11.1). All existing time-resolved lightning spectra have been taken on black and white film since it permits much easier intensity measurements. Obtaining lightning spectra is too difficult a job to waste a spectrum by taking it on color film. The spark and lightning spectra differ primarily in that the lightning spectrum is stretched out three to five times longer than the spark spectrum (Ref. 11.1). For example, a nitrogen ion spectral line might last 5 millionths of a second in the spark, whereas for a typical lightning stroke it might last 20 millionths of a second.

From measurements of the intensities of the various spectral lines and use of appropriate theory, channel temperature can be determined as a function of time (Refs. 11.2, 11.3). A graph of typical lightning return stroke temperature vs. time is shown in Fig. 11.4. The peak temperature measured is about 55,000°F. The actual peak temperature could be higher as was previously noted. The temperature falls to about half of peak value in about 30 millionths of a

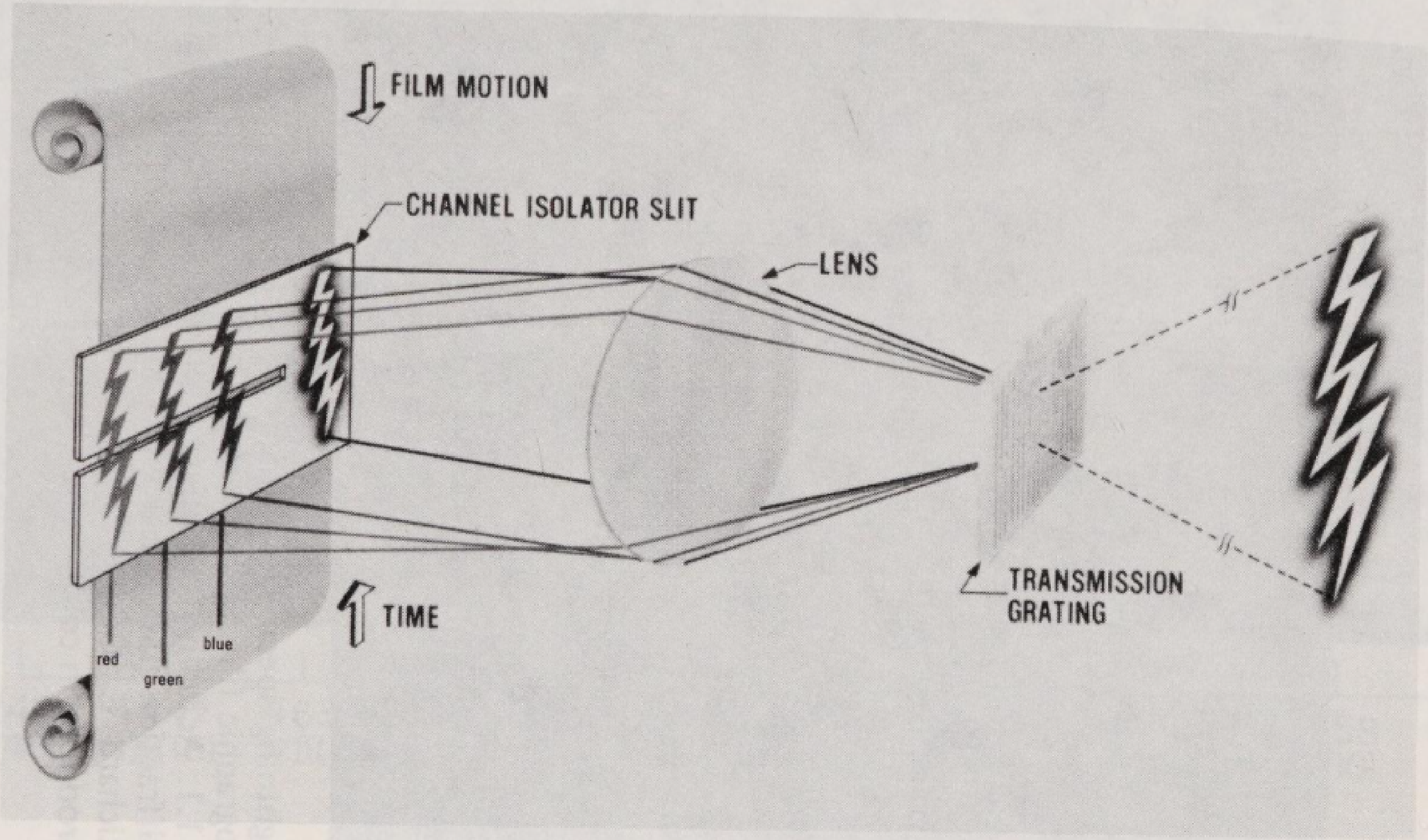
FILM MOTION
CHANNEL ISOLATOR SLIT
LENS
TRANSMISSION GRATING
TIME
red
green
blue

Fig. 11.1. A lightning spectrometer. Lightning image is broken up into its spectrum by transmission grating (similar to prism). Lens focuses various spectral images of lightning, as well as a "white light" image, on the channel isolator slit assembly. A spectrum from one section of the channel passes through the slit and is recorded on the moving film.

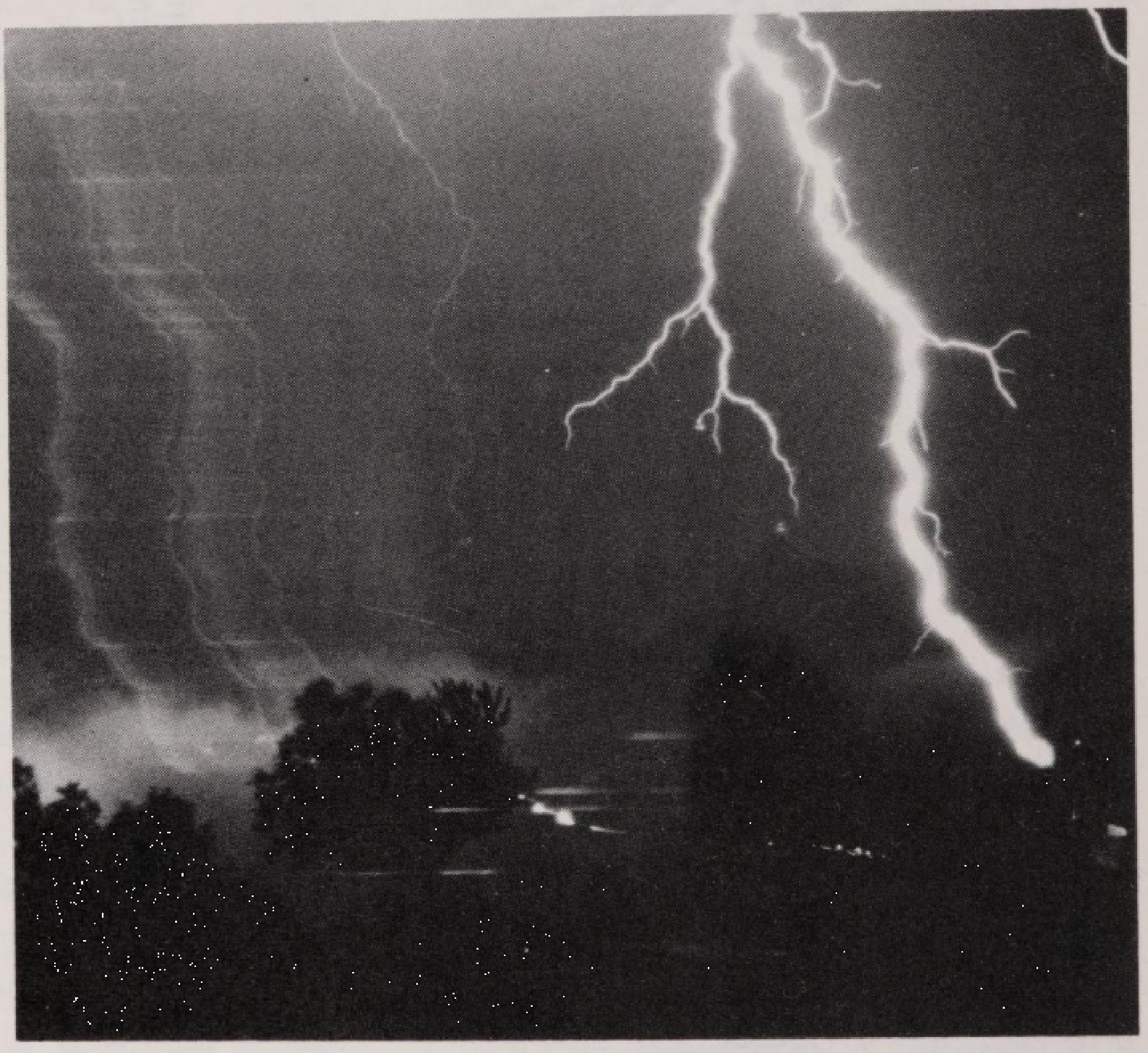

Fig. 11.2. Lightning spectrum as might be obtained by putting stationary photographic film in front of the channel isolator slit assembly of Fig. 11.1. Actually, the photograph was taken by placing an inexpensive diffraction grating in front of the lens of a 35 mm camera. (Courtesy, Richard E. Orville, State University of New York at Albany) (Reproduced in color on inside front cover)

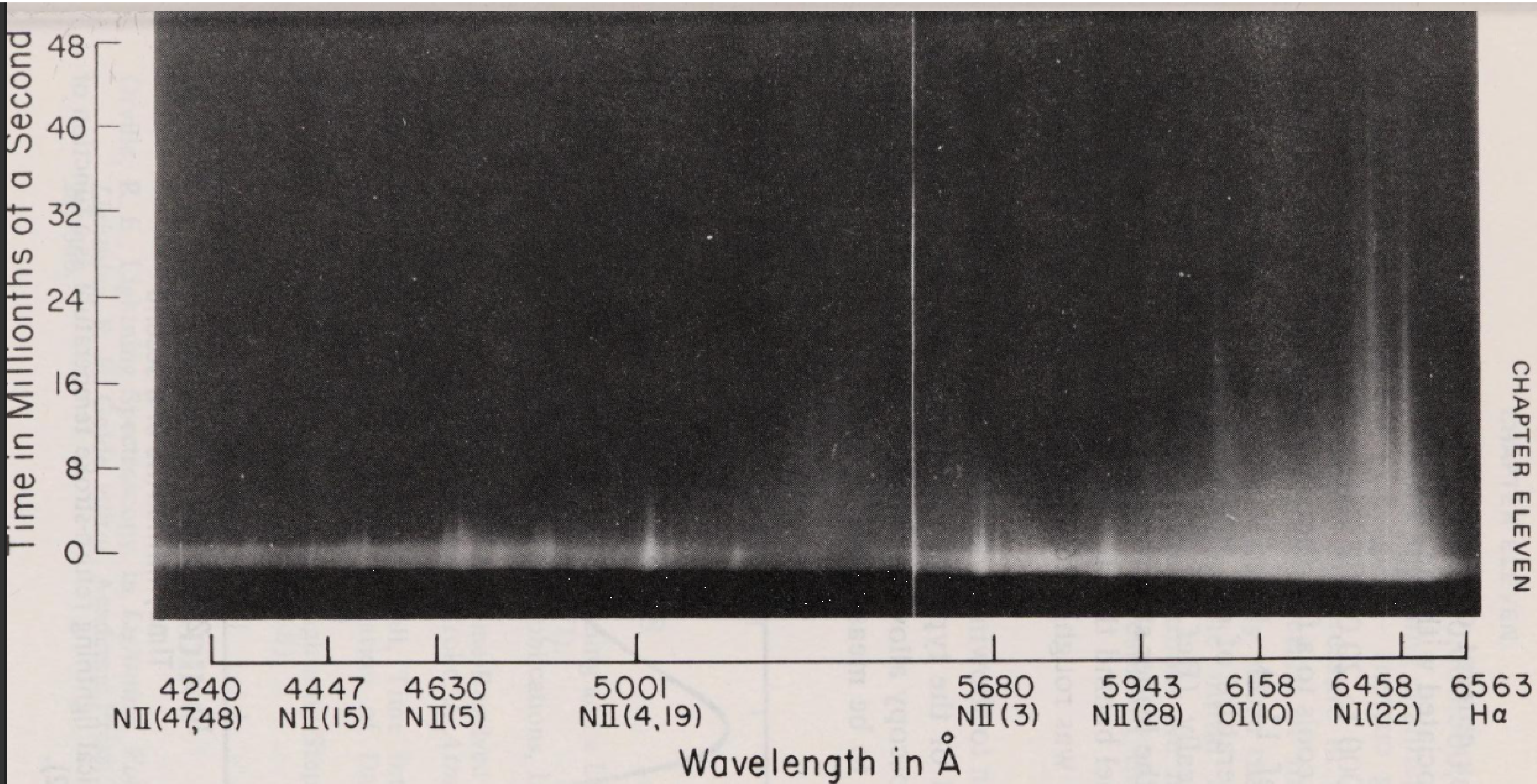

g. 11.3. Time-resolved spectrum of a 15 ft laboratory spark (Ref. 11.1) as obtained on the moving film of a ectrometer similar to that shown in Fig. 11.1. NI is neutral nitrogen, OI neutral oxygen, NII singly ionized nitrogen. ne numbers in parenthesis following the species designation are the multiplet numbers found in the technical erature. One Å (Ångstrom) equals one ten-billionth of a centimeter. There are 2.54 centimeters to the inch. Zero time ıs been arbitrarily set at the point where blue radiation first appears. (Reproduced in color on the inside back cover) second. For a current of about 100 amperes in the lightning channel (associated with the return stroke current tail or with a continuing current – Fig. 5.7) the temperature will be roughly 15,000 to 20,000°F. When the stroke current stops, the channel cools to a few thousand degrees in hundredths of a second (Ref. 11.4).

The temperature of one stepped-leader has been measured spectroscopically (Ref. 11.5). The peak temperature associated with the leader step was about 50,000°F. Temperature in the channel behind the step decreased with distance up the channel and was roughly 25,000°F five step lengths beyond the step.

In addition to allowing a determination of the channel temperature and of the types of particles in the channel, quantitative spectroscopy allows the number densities of all the channel particles to be measured. For example, it has been determined that there are over 1,000,000,000,000,000,000 free electrons per cubic inch in the lightning channel (Refs. 11.2, 11.3).

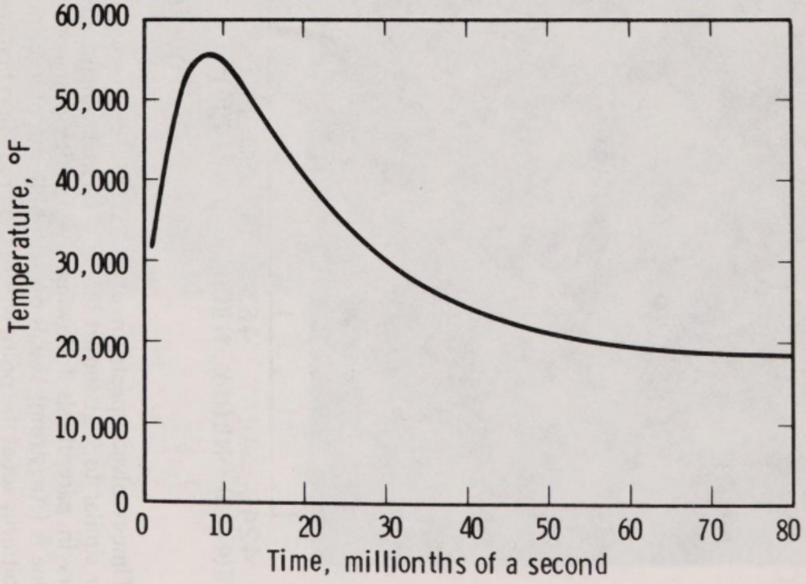

Fig. 11.4. Typical lightning return-stroke temperature as a function of time. (Ref. 11.2).

Also, spectroscopy permits the pressure in the lightning channel to be determined as a function of time. It is the initial high channel pressure (possibly more than 100 times atmospheric pressure) caused by the initial high channel temperature which produces the thunder we hear. We will consider channel pressure and the resultant thunder generation in the next chapter.

REFERENCES

11.1. Uman, M. A., Comparison of Lightning and a Long Laboratory Spark, *Proc. IEEE, 59*, 457-466 (1971).
11.2. Uman, M. A., *Lightning*, Dover Publications, Inc., New York, 1984, Chapter 5.
11.3. Orville, R. E., A High-Speed Time-Resolved Spectroscopic Study of the Lightning Return Stroke, *J. Atmosph. Sci., 25*, 827-856 (1968).
11.4. Uman, M. A., and R. E. Voshall, Time Interval Between Lightning Strokes and the Initiation of Dart Leaders, *J. Geophys. Res., 73*, 497-506 (1968).
11.5. Orville, R. E., Spectrum of the Lightning Stepped Leader, *J. Geophys. Res., 73*, 6999-7008 (1968).

ADDITIONAL READING

Orville, R. E., Lightning Spectroscopy, in *Lightning, Vol. 1, Physics of Lightning*, R. H. Golde, editor, Academic Press, N.Y., 1977, pp. 281-306.

chapter twelve

How Is Thunder Generated?

How Can It Be Used To Measure The Distance And Length Of The Lightning Channel?

Throughout recorded history there has been a succession of interesting views on the origin of thunder. Aristotle, in the fourth century B.C., believed that thunder was the noise created when "dry exhalation" forcibly ejected from a cooling cloud struck the surrounding clouds – "dry exhalation", according to Aristotle, being one of the constituents of air (Ref. 12.1). We know now that there is no such substance. Interestingly, Aristotle thought that the origin of thunder preceded that of lightning and that the latter was a "burning wind" produced after the impact of the "dry exhalation" on a surrounding cloud. Lucretius, in the first century B.C., wrote that thunder was produced by clouds banging together (Ref. 12.2). In 1637 Descartes, the mathematician after whom the cartesian coordinate system is named, suggested that thunder was an organ-pipe effect: that thunder was due to a resonance of the air between two clouds occurring when a higher cloud descended upon a lower cloud (Ref. 12.3).

All of these early theories erroneously associated thunder with clouds rather than with the lightning channel. Later speculation regarding the origin of thunder generally associated thunder with the lightning channel. For example, Benjamin Franklin wrote in 1749 (see Chapter 1) that lightning and laboratory sparks were similar in that they both gave a "crack or noise in exploding". The correct meaning for Franklin's "exploding" was not generally agreed upon until the twentieth century. In the mid-nineteenth century the accepted theory seems to have been that lightning produced a vacuum along its path, and thunder resulted from the subsequent motion of air into the vacuum. Two other theories that achieved some respectability involved (1) the chemical explosion of gaseous materials created by the lightning, and (2) the creation of steam explosions along the heated lightning path. One of the first adequate descriptions of the origin of thunder was published by M. Hirn in 1888 (Ref. 12.4).

The thunder we hear is, in fact, created in the following way: The return stroke rapidly deposits a large amount of energy along the leader channel. That channel is heated by the energy input to above 50,000° F (see Chapter 11). Heating of a short section (say 10 yards) of channel takes only millionths of a second and hence the channel section has no time to expand while it is being heated. Air heated from room temperature or from a leader temperature of a few thousand degrees to above 50,000° F without having time to expand attains a pressure considerably in excess of normal atmospheric pressure (1 atmosphere). The initial pressure of the return stroke channel is definitely in excess of 10 atmospheres and may be 100 atmospheres or more. The high pressure channel rapidly expands into the surrounding air (initially at atmospheric pressure) and compresses it. This disturbance of the air propagates outward in all directions. For the first ten yards or so it propagates as a *shock wave* (a major disturbance in the air which travels faster than the speed of sound) and after that as an ordinary sound wave (small compressions and expansions of the air density). The sound pulse from a short section of lightning channel lasts less than 0.1 seconds and travels at about 1090 ft/second (about 0.21 miles/second) at sea level. The thunder we hear, then, is the pressure variations induced in the air by the expansion of each part of the lightning channel (main channel and branches) due to its initial high pressure.

Since sound travels about a foot in a thousandth of a second and the return stroke heats the entire lightning channel in less than a thousandth of a second (see Chapter 9), for all practical purposes each point on the lightning channel can be considered to emit a pulse of sound (or a shock wave) at the same time. Because light emitted by the return stroke channel travels away from the channel at 186,000 miles/second (or a mile in about 5 millionths of a second), we see light almost simultaneously with channel formation, but the sound of the thunder from that channel frequently takes many seconds to reach us. The initial thunder heard comes from the point on the lightning channel nearest to the observer. The last sound (excluding echoes) comes from the point farthest away. In Fig. 12.1, a 2-mile long intracloud discharge is oriented along the line of sight. We assume (direct evidence is lacking) that sound generation by the intracloud discharge occurs essentially simultaneously at all points on the channel as it does with the ground discharge. From the time it takes for the first sound to arrive, we can establish the distance to the nearest point on the channel. The first sound, traveling about 0.21 miles/second, arrives from the nearest channel point in about 13.5 seconds; multiplying these two figures provides the distance to the

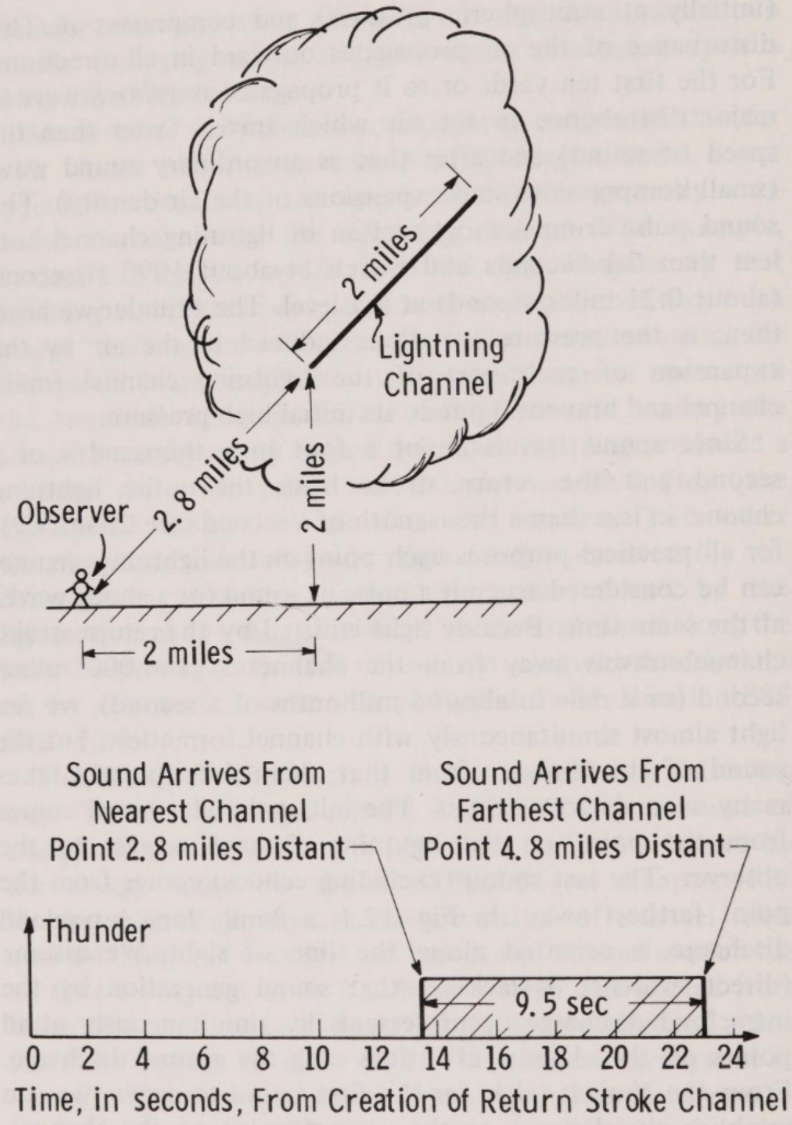
2 miles
Lightning
Channel
2.8 miles
2 miles
Observer
2 miles
Sound Arrives From
Nearest Channel
Point 2.8 miles Distant
Sound Arrives From
Farthest Channel
Point 4.8 miles Distant
Thunder
9.5 sec
0
2
4
6
8
10
12
14
16
18
20
22
24
Time, in seconds, From Creation of Return Stroke Channel

Fig. 12.1. Intracloud lightning flash and its thunder.

nearest point on the channel – 2.8 miles. (In other words, if the time until the thunder arrives is T, the distance to the channel section producing the thunder is roughly T/5, the formula given in Chapter 10.) Sound from the far end of the channel, 4.8 miles away, arrives about 23 seconds after the channel is created, ending the thunder. The duration of the thunder is about 9.5 seconds, from which we can establish the channel length (channel length in miles = 9.5 seconds x 0.21 miles/second). The result represents the actual channel length only if the channel is oriented along the observer's line of sight. Otherwise it is an underestimate of the true length (that is, the true length is longer than the calculated value) as we can see by reference to Fig. 12.1 and 12.2.

In Fig. 12.2, a 2-mile long channel to ground produces sound from its base which arrives at the observer 2 miles from the base in about 9.5 seconds (2 miles = 9.5 seconds x 0.21 miles/second). Sound from the top of the channel, 2.8 miles away, arrives in about 13.5 seconds and thus the total thunder duration is about 4.0 seconds. By measuring only the thunder duration we can conclude that the minimum channel length that could have produced the thunder was 4.0 seconds x 0.21 miles/second = 0.84 miles. The true channel length, 2 miles, is of course longer than the calculated minimum value. Sometimes actual thunder durations are as long as a minute (minimum channel lengths of over 10 miles).

To summarize, from a measurement of the time of arrival of the first sound of thunder we can determine the distance to the closest point of the lightning channel; and from the thunder's duration we can determine a minimum length for the channel. In both cases a distance is derived by multiplying the appropriate time interval by the speed of sound. This is the case even if the channel under consideration is not straight, as the reader may easily prove to himself. It is worth noting that a visible channel to ground may be much farther away than its thunder would indicate if the in-cloud part of the lightning channel passes overhead.

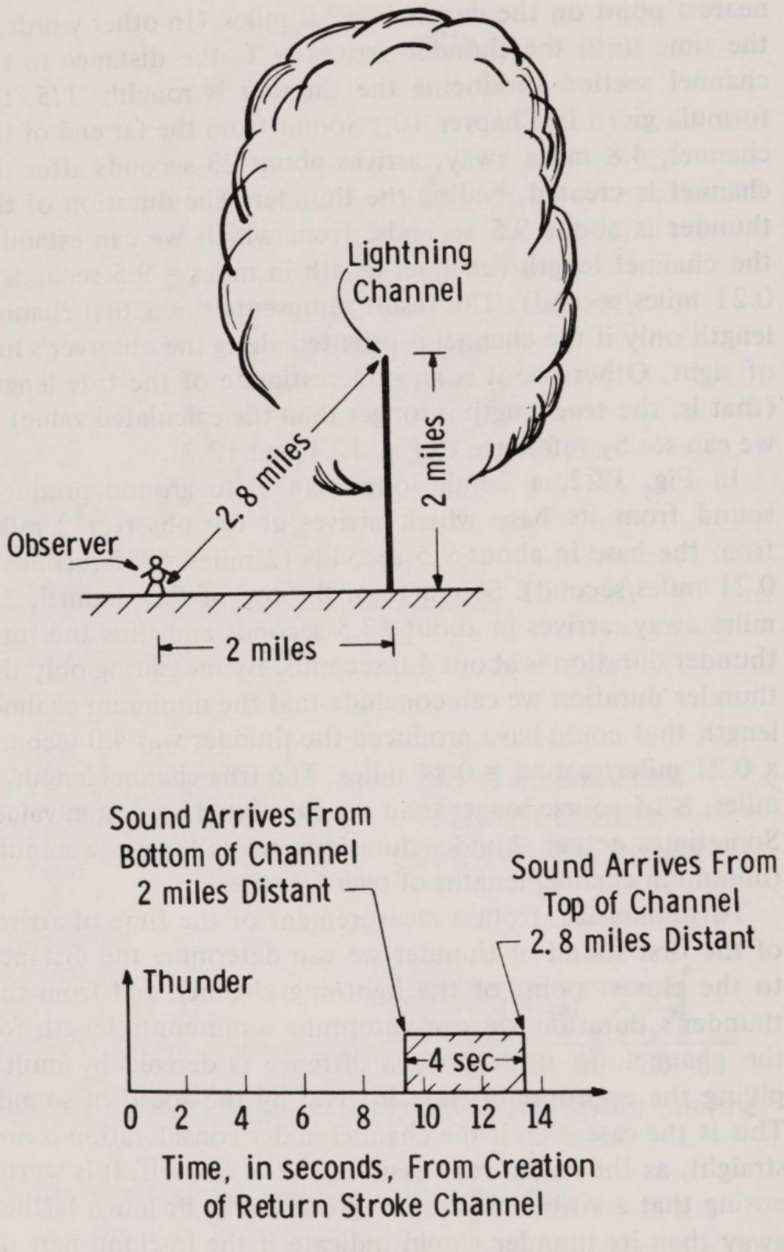
Lightning
Channel
2.8 miles
2 miles
Observer
2 miles
Sound Arrives From
Bottom of Channel
2 miles Distant
Sound Arrives From
Top of Channel
2.8 miles Distant
Thunder
4 sec
0
2
4
6
8
10
12
14
Time, in seconds, From Creation
of Return Stroke Channel

Fig. 12.2. Cloud-to-ground lightning flash and its thunder.

Thunder is generated by each of the return strokes in a multiple-stroke flash. Since the time between the first and last strokes may be a few tenths of a second, the thunder duration may be increased by a few tenths of a second over the time taken for sound to arrive from different ends of the channel. The error made in computing minimum channel lengths due to this effect is not important. It should be noted that the last stroke in a multiple-stroke discharge usually has the greatest length (see Chapter 10), and hence produces the longest lasting thunder. It is therefore the length of the final stroke that is measured.

In addition to thunder produced by return strokes, sound is also produced by stepped leaders and dart leaders. Leader thunder is relatively weak compared with that produced by return strokes.

Why does thunder sound like it does? Why does it rumble, roll, peal, and clap? Two factors must be considered: the intensity or loudness of the sound, and its pitch or frequency. We look first at the loudness. If a section of the tortuous lightning channel is perpendicular to the observer's line of sight, all points on that section will produce sound that arrives almost simultaneously at the observer, and the result is a high-intensity sound – a peal or a clap. If a channel section is along the observer's line of sight (as with the whole channel in Fig. 12.1), the sound arrival times from various points on the channel will be spread out, resulting in a low-intensity sound – a rumble or a roll. Channel tortuosity enhances the rumbling effect. Typical thunder consists of rumble and roll on which three or four peals or claps are superimposed.

Now to pitch. As we have seen, thunder generation is similar to an explosion. As everyone knows, small explosions (e.g., small firecrackers) make a high pitched sound while large explosions (e.g., large aerial fireworks or sticks of dynamite) make a low pitched sound. The larger the energy input, the lower the pitch. Thunder generally has a pitch, or frequency, of about 50 cycles per second. From measurements of thunder pitch and use of appropriate theory, scientists have determined the energy input per unit length of lightning stroke channel (100,000 to 1,000,000 watt-seconds per yard of channel length). Thunder pitch is influenced by two additional factors: First, the lower the air density for a given energy input to the lightning channel, the lower the thunder pitch; this means that thunder from a channel at high altitude should have a lower pitch than thunder from a channel near ground. Second, air filters out the higher pitches preferentially, so that the further away thunder is heard from its channel, the lower should be its pitch.

Thunder is seldom heard more than 15 miles away from a lightning channel. The reason is that, due to the effects of temperature and wind variation with height, sound waves generally curve upward in the air under and surrounding thunderstorms. Thus, the thunder from a lightning channel passes over the head of an observer if he is more than about 15 miles from the channel. If he climbed a tall enough ladder, he would hear the thunder.

Within a few hundred yards of a lightning strike, the observer hears one loud (terrifying) bang possibly preceded by hissing or clicking or something similar to cloth tearing. The bang is the sound wave from the channel base. No long-duration thunder is usually heard because, due to the upward curvature of the sound waves, the thunder from sections of the channel above the channel base pass over the observer's head. The various sounds reported to occur before the loud bang may be due to the downward-moving stepped leader, to upward-moving discharges attempting to meet the stepped leader, and/or to a general corona-discharge blanket at the ground.

REFERENCES

12.1. Aristotle (384-322 B.C.), *Meteorologica*, H. P. D. Lee, trans., Loeb Classical Library, Harvard University Press, Cambridge, Mass., 1951, pp. 223-225.

12.2. Lucretius, T. (98-55 B.C.), *On the Nature of Things*, book VI, H. A. J. Munro, trans., Great Books of the Western World, William Benton, Chicago, 1952, p. 81.

12.3. Haldane, E. S., *Life of René Descartes*, John Murray (Publishers), Ltd., London, 1905, p. 181.

12.4. Hirn, M., The Sound of Thunder, *Sci. Am.*, *59*, 201 (1888).

A general reference for this chapter and one which contains a more detailed bibliography of the early theories of thunder is: *Lightning*, M. A. Uman, Dover Publications, Inc., New York, 1984, Chapter 6.

ADDITIONAL READING

Few, A. A., Acoustic Radiations from Lightning, in *CRC Handbook of Atmospherics, Vol. 2*, H. Volland, editor, CRC Press, Boca Raton, Florida, 1982, pp. 257-290.

chapter thirteen

Does Lightning Occur Without Thunder? Thunder Without Lightning?

What Is Heat Lightning? Sheet Lightning?

Ask a lot of people the question "Did you have lightning in your vicinity yesterday?" and, assuming there was lightning, at least one person will invariably answer "No lightning, but a lot of thunder!" Possibly, this is just another way of saying that thunder was heard but lightning wasn't seen. More likely, the answer indicates that the person does not understand the relation between lightning and thunder.

We have seen in the previous chapter that the origin of thunder is the explosive expansion of the lightning channel following a rapid energy input to the channel. Since it is the lightning channel that creates the thunder, there can be no thunder without lightning. But can there be lightning without thunder? Strictly speaking, no. Every electrical discharge, when initiated, produces some noise. However, if we ask whether there is lightning whose thunder may be inaudible a relatively short distance from the channel, the answer is apparently yes. There have been lightning flashes reported to strike the Washington Monument without producing audible thunder, and flashes to the Empire State Building which were reported to have produced a sound described as being the same as if a paper bag full of water had been dropped a short distance to the floor (Ref. 13.1).

As noted in Chapter 6, lightning flashes to tall buildings are generally initiated by stepped leaders which move upward from the building tops. The current in a structure-initiated flash rises slowly to a peak current of a few hundred amperes and this low level of current typically flows for a few tenths of a second. Often this continuous current is punctuated by current peaks caused by return strokes whose leaders are initiated in the cloud and travel from cloud to building top down the active channel. Sometimes no return strokes occur and the flash consists of only the low-level current. It is the upward-initiated discharge with no return strokes that might be expected to generate very little sound. Its current is small and increases slowly compared to the current of a typical return stroke; the energy input is neither large nor rapid and thus cannot lead to the explosive channel expansion and typical thunder characteristic of the usual return stroke channel.

What about "heat lightning"? Why does it have no thunder? Before we answer this question, we had better define what we mean by "heat lightning". In common terminology, heat lightning is the name given to the illumination of distant clouds close to the horizon which occurs without thunder and in the absence of a visible lightning channel. Heat lightning apparently is so named because it occurs on hot nights. The fact that it occurs on hot, generally humid nights is due simply to the fact that hot humid air near the ground is an ingredient in the thunderstorm recipe (see Chapter 7). The nocturnal thunderstorms accused of generating heat lightning actually produce the usual intracloud and cloud-to-ground lightning flashes. The light from these flashes illuminates the clouds from within and by reflection so that the clouds are easily seen even when they are sufficiently far away (say 30 miles) that individual lightning channels cannot be discerned. Since thunder cannot generally be heard more than about 15 miles from a lightning channel (see Chapter 12), the distant heat lightning produces no audible thunder.

"Sheet lightning" is a close relative of heat lightning. Sheet lightning is the name given to the intracloud discharges which light up a large cloud area simultaneously, giving the impression of a sheet of light. Sheet lightning may or may not produce audible thunder, depending on its distance from the observer.

REFERENCES

13.1. McEachron, K. B., Lightning to the Empire State Building, *J. Franklin Institute, 227*, 149-217 (1939).

chapter fourteen

Do Gushes Of Rain Follow Thunder?

Observers of thunderstorms, from antiquity to the present, have noted that a heavy gush of rain often reaches ground a minute or two after a lightning flash and its accompanying thunder (Refs. 14.1, 14.2). Since precipitation is believed necessary to provide a thunderstorm with its electrical charge (see Chapter 8), it could be argued that a rain gush occurring in the cloud provides the charge separation that causes the lightning. On the other hand, it could just as well be argued that the lightning or its thunder modifies the character of the precipitation in the cloud in such a way as to cause a rain gush. Finally, it is also possible that both lightning and the rain gush are caused by some other effect or that it is only fortuitous that they appear to be related. Not until the 1960s did radar observations of the precipitation in clouds before and after lightning show that lightning and thunder precede the formation of the rain gush (Refs. 14.1, 14.2).

As long ago as 58 B.C., Lucretius wrote that gushes of rain were caused by thunder. Modern proponents of this currently unpopular view suggest that the compression and rarefaction waves associated with the thunder accelerate some drops relative to others causing collisions between the drops which would otherwise not have occurred. The collisions allow small drops to coalesce into bigger drops which are heavy enough to fall out of the cloud as rain. It is interesting to note that from radar measurements of clouds there is evidence to suggest that thunder can cause supercooled water droplets (water droplets below 32°F which are not frozen) in the clouds to turn to ice (Ref. 14.3). Further, supercooled water droplets can be turned to ice crystals in the laboratory by the sound from a toy pop gun or by a bursting balloon (Ref. 14.4). The freezing mechanism involved in both the laboratory and the cloud may be the lowering of the droplet temperature by the expansion of the air associated with the sound wave rarefaction.

Most investigators now believe that the lightning/rain gush relation is electrical rather than acoustic. As seen in Chapters 8 and 9, the top of the lightning channel is found in the N-region of the cloud. The N-charge resides on supercooled water and ice particles. Stepped and dart leaders are negatively charged, drawing their charge from the N-region. The return stroke drains the negative charge on the leader to the ground. After the return stroke has reached the top of the lightning channel, the whole channel, a good conductor tied to ground, becomes positively charged in response to the negative cloud charge above and around it. The leader-return stroke process results in a negatively-charged leader channel being replaced by a positively-charged channel in a time that is short compared to that in which any of the surrounding negative cloud charge can move appreciably (Fig. 9.3a). If electrical effects cause the rain gush, the charged cloud particles must coalesce in response to the rapid change in the sign of the charge on the lightning channel.

The mechanism for coalescence most closely examined is that proposed by Bernard Vonnegut and Charles Moore (Refs. 14.1, 14.2, 14.5). They postulate that the cloud droplets in the immediate vicinity of the newly-completed return stroke channel acquire a net positive charge from the positively-charged channel. Since similar charges repel one another, the positively charged droplets fly away from the channel, initially at perhaps tens of yards per second, colliding and coalescing with the surrounding negatively-charged droplets. The initially-stationary negative droplets are attracted toward the moving positive droplets by virtue of their opposite charge, further increasing the collision rate. The positive droplets may travel about 10 yards, increasing in size all the time. The end result is a group of large water drops surrounding the channel.

If the process just described is to produce a rain gush, it must be operative over a reasonably large volume of the cloud. Vonnegut and Moore suggest that this is the case because the lightning channel in the cloud has a tree-like structure. A drawing of this structure is shown in Fig. 9.3. Some idea of the extent of the structure may be inferred from electrical discharges in insulating materials (the cloud is essentially an insulator). In Fig. 14.1 a photograph of a plastic block in which a tree-like discharge has taken place is shown. The block was charged negatively by shooting high energy electrons into it. The tree-like discharge occurred when a grounded electrode was placed at the edge of the block. Unfortunately, there is at present no direct evidence regarding the extent of the lightning channel structure in the cloud, and we must await future lightning research to determine its actual structure.

Some of the gushes of precipitation that fall after lightning are in the form of hail (Ref. 14.2). When a rain gush forms by a coalescence process at temperatures below 32°F, the drops may freeze as they become large and then continue their growth as hail particles.

While most rain gushes are preceded by lightning, it does not follow that lightning invariably produces rain gushes. Whether there is a rain gush or not depends, among other things, on the liquid water content of the cloud. If this is too low, heavy rain cannot form.

Although it is possible for lightning to occur in the absence of appreciable precipitation, the two usually go hand in hand. Recently the amount of rainfall reaching the ground near Tucson, Arizona has been correlated with the number of cloud-to-ground discharges (Ref. 14.6). These results and those of earlier studies show that the greater the number of lightning discharges, the greater the rainfall. (The clouds which produce both the most rain and the most lightning are those of the greatest vertical extent.) From measurements of the amount of rainfall and the number of lightning flashes occurring near Tucson between 1 and 6 p.m. on 52 days, it was determined that an average of about a thousandth of an inch of rain arrives at the ground for each cloud-to-ground lightning stroke observed. On 23 days of heavy rain (exceeding 0.10 in.) there was an average of 175 cloud-to-ground lightning flashes per day. On 29 days of light rain (less than 0.01 in.) there was an average of only three flashes per day.

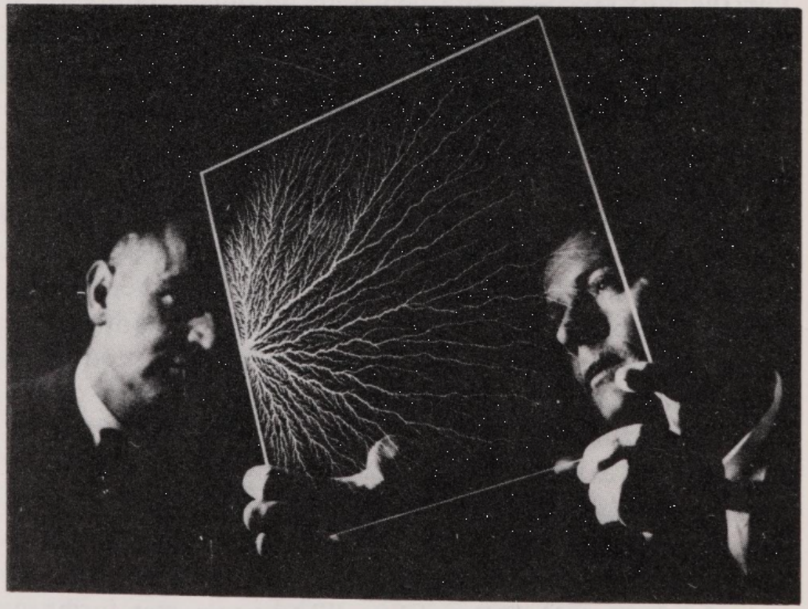

Fig. 14.1. Electrical breakdown in a plastic block. The tree-like structure which extends into the block may be similar to the lightning discharge structure extending from the top of the channel upward into the cloud. (Courtesy, Arthur D. Little, Inc.)

REFERENCES

14.1. Moore, C. B., B. Vonnegut, J. A. Machado, and H. J. Survilas, Radar Observations of Rain Gushes Following Overhead Lightning Strokes, *J. Geophys. Res.*, *67*, 207-220 (1962).

14.2. Moore, C. B., B. Vonnegut, E. A. Vrablik, and D. A. McCaig, Gushes of Rain and Hail after Lightning, *J. Atmosph. Sci.*, *21*, 646-665 (1964).

14.3. Vonnegut, B., and C. B. Moore, Nucleation of Ice Formation in Supercooled Clouds as the Result of Lightning, *J. Appl. Meteorology*, *4*, 640-642 (1965).

14.4 Vonnegut, B., Production of Ice Crystals by Adiabatic Expansion of Gas, *J. Appl. Phys.*, *19*, 959 (1948).

14.5. Vonnegut, B., and C. B. Moore, A Possible Effect of Lightning Discharge on Precipitation Formation Process, *Am. Geophys. Union Monograph Number 5*, 287-304 (1960).

14.6. Battan, L. J., Some Factors Governing Precipitation and Lightning from Convective Clouds, *J. Atmosph. Sci.*, *22*, 79-84 (1965).

ADDITIONAL READING

Piepgrass, M. V., and E. P. Krider, Lightning and Surface Rainfall During Florida Thunderstorms, *J. Geophys. Res.*, *87*, 11,193-11,201 (1982).

Szymanski, E. W., S. J. Szymanski, C. R. Holmes, and C. B. Moore, An Observation of Precipitation Echo Intensification Associated with Lightning, *J. Geophys. Res.*, *85*, 1951-1953 (1980).

Zrnic, D. S., W. D. Rust, and W. L. Taylor, Doppler Radar Echoes of Lightning and Precipitation at Vertical Incidence, *J. Geophys. Res.*, *87*, 7179-7191 (1982).

chapter fifteen

What is Ball Lightning?

Are UFOs and Ball Lightning Related?

Ball lightning (in French, *boules de feu* or *foudre sphérique;* in German, *Kugelblitz*) is the name given to the mobile luminous spheres which have been observed during thunderstorms. A typical ball lightning is about the size of an orange or a grapefruit and has a lifetime of a few seconds. A 19th century woodcut depicting the occurrence of a somewhat larger than average ball lightning is shown in Fig. 15.1. A number of compilations of eye-witness reports of ball lightning have been published (Refs. 15.1 to 15.6). Visual sightings are often accompanied by permanent material damage, sounds, and odors. Nevertheless several noted scientists have attributed ball lightning to persistence of vision (after seeing a bright lightning channel) or to other optical effects (Ref. 15.7).

In a letter to the editor of the London *Daily Mail* (Ref. 15.8) W. Morris described an unusual incident in which a ball lightning caused a tub of water to boil:

> ● During a thunderstorm I saw a large, red hot ball come down from the sky. It struck our house, cut the telephone wire, burnt the window frame, and then buried itself in a tub of water which was underneath.
>
> The water boiled for some minutes afterwards, but when it was cool enough for me to search I could find nothing in it.

As a researcher in the field of lightning, the writer has personally received over one hundred unsolicited eye-witness accounts of ball lightning. Here are five representative observations:

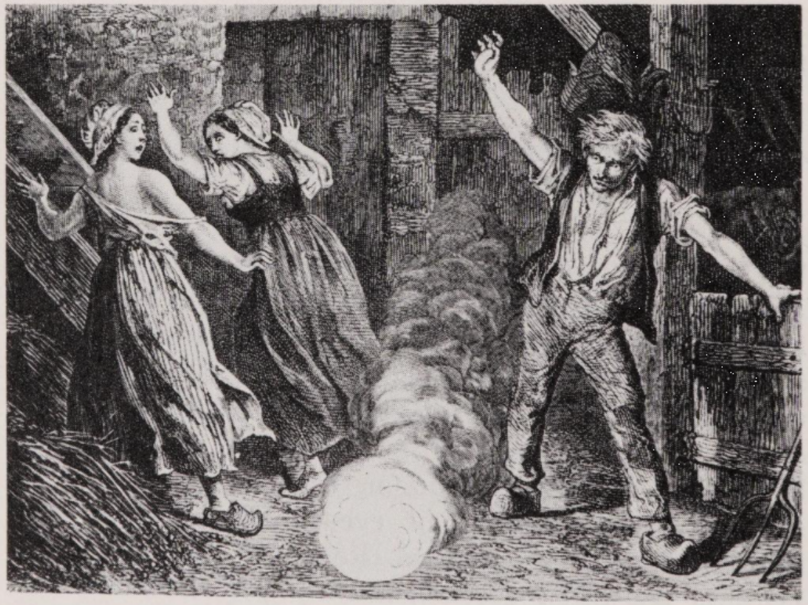

Fig. 15.1. Ball lightning in a 19th century woodcut. The original title, translated from French, reads "Ball lightning crossing a kitchen and a barn." Perhaps the ball lightning came down the chimney used to exhaust the cooking fires. How the young lady's blouse came to be in such a state of disarray is not known. (Courtesy, Burndy Library)

● I was visiting my uncle's farm in Eastern Norway where the country is very flat and the summers warm with cloud gatherings and thunderstorms in the middle of the day. One day when my cousin and I went shopping, a thunderstorm worse than usual broke out. The rain poured down, and we stayed in the open door of the shop, looking out. Suddenly we saw a glowing ball, the size of a small football [soccerball], passing the door at great speed a little above ground level. We had never heard about 'ball lightning' before, and wondered what it might be. When the rain stopped, we started on our way home and walked along a path leading through the forest. Then it started thundering again, and another ball, just like the first one, came rolling towards us. Now we got curious, but not very much – to 10-year-old children there are so many strange things to wonder about – so what is a rolling glowing ball more or less? Then my uncle came to meet us, because he thought we might be frightened by the heavy thundering, and the first thing he said was, 'Did you see the ball lightning?' Not till then did we realize that we had seen something strange and dangerous.

● While outdoors during a particularly violent thunderstorm such as are common there [Iowa], I heard a heavy rushing noise like an extra strong wind. This caused me to look about me and I saw a ball of fire, yellowish white about the size of a wash tub bouncing down the dirt road. The ball traveled a little faster than one could run. I did not see from where it came. I watched the ball travel about a city block, when it struck a small shed, maybe 10 feet by 12 feet in which a horse was stabled. The shed seemed to explode and the horse was killed.

● Near Murray's dairy, a property on the Queanbeyan road on the then outskirts of Canberra [Australia], and not far from what is now the industrial area of Fyshwick, I was riding along the right hand side of the road, just off the paving to prevent a possible slip and fall on the wet surface, and about 20 to 25 yards ahead, on the left hand side and also off the bitumen, a farm employee was leading a Shorthorn bull. There was very little, if any rain at that time. I can clearly recall that there was one of those periods of "quietness" that sometimes precedes a downpour. Just as I drew level with the bull there was one very loud bang or explosion and immediately down the white traffic line in the centre of the road appeared the fireball. It seemed to be about 6″ or 8″ off the ground, was about the size of a basketball like very golden butter in color, and had the appearance of being "spun" or "fuzzy", like silk threads or wool, as distinct from a "molten" liquid look. It did not sparkle – just a ball of fuzz. It came straight down say three of the white marker lines. These have about equal distance of unmarked road between them, so I should say it travelled about 18 to 20 feet. Then it simply disappeared. It did not break apart. There was no further noise like an explosion. It was there one moment and not there the next.

The whole thing would have been over in probably two or three seconds, before the horse had time to be startled.

The young men leading the bull cried out – in Pure Australian – "What the bloody hell was THAT?".

As there was another downpour shortly, we did not stop to discuss it.

- I was standing in the kitchen of my home in Omaha, Nebraska while a terrible thunderstorm was in progress. A sharp cracking noise caused me to look toward a window screen to my left. Then I saw a round, iridescent (mostly blue) object, baseball size, coming toward me. It curved over my head and went through the isinglass [mica] door of the kitchen range, striking the back of the oven and spattering into brilliant streamers. There was no sound and no effect on me except a tingle as it passed over my hair. Later examination showed a tiny hole with scorched edges in the screen and isinglass, and scorch-like marks on the back of the oven.

- The whole family was sitting at the supper table when a ball about 4 inches in diameter came and fluttered about the center of the table. It made a buzzing noise and was about 8 to 10 inches above the dishes. Its color was a mixture of blue and orange and may have had some red. It fluttered about 5 to 6 seconds. I could easily have grabbed it had I dared to do so. Then it exploded with a loud bang like a large fire cracker and gave off a smell somewhat like city cooking gas which lasted several minutes. After it exploded the table was left as before, no dishes broken or moved about.

A number of photographs purported to be of ball lightning have been published (Refs. 15.9 to 15.14). Most of them are time exposures in which the ball lightning appears as a meandering ribbon of light. Some of the photographs may well be of automobile or other moving lights, or of stationary lights photographed by a moving camera. It appears that no motion pictures of ball lightning have been taken.

A phenomenon very similar to, if not identical with, ball lightning has been reported to occur in submarines due to the discharge of a high current (about 150,000 amp D.C. from a 260 volt source) across a circuit breaker (Ref. 15.15). In addition, the writer has received a number of reports of ball-lightning-like phenomena being initiated in electrical equipment, both with and without lightning present. One such thunderstorm-associated incident was reported by a retired professor and engineering department chairman at a major state university:

- In 1914 I was working for the Interurban Co. and happened to be in the main power house during a very severe thunder storm. We had recently installed a 600 volt dc rotary converter which supplied power to both the city and interurban lines. I was working on the main switch board and watching a bit carefully as there were many breakers opening. I looked up just in time to see a ball of fire about 15 to 18 inches in diameter come off the converter commutator and float at a good speed down the length of the switch board, maybe three feet above my head. Some hundred feet from the converter it struck the ceiling of an office room where a cat was asleep. The ball splashed in all directions like water and the cat sprang high in the air. The cat squalled loudly but seemed to be uninjured. In both cases there was the strong odor of ozone, but due to the constant lightning I would not attribute it to the 'ball lightning.'

Ball lightning has been seen by 5 to 10% of the population. Although it is generally thought that ball lightning is a rare phenomenon, recent research has shown that the numbers of ball lightning observers are not much different from the numbers who have observed lightning impact points (Ref. 15.6). The implication is that ball lightning is usually created at or near the lightning channel and that an appreciable fraction of all cloud-to-ground lightning flashes may give birth to ball lightning. Since the balls generally last for only a few seconds, they cannot get too far from the mother channel. Thus ball lightning may well be common, but rarely seen.

Ball lightning and St. Elmo's fire are sometimes confused. St. Elmo's fire is a corona discharge induced at a pointed conducting object by thunderstorm charges. Like ball lightning, St. Elmo's fire may assume a spherical shape. Unlike ball lightning, it must remain attached to a conductor, although it may exhibit some motion along the conductor. Further, St. Elmo's fire may last much longer than the typical ball lightning.

From the many published ball lightning observations, it is possible to compile a list of ball lightning characteristics:

Occurrence: Most observations of ball lightning are made during thunderstorm activity. Most, but not all, of thunderstorm-related ball lightning appears almost simultaneously with a cloud-to-ground lightning discharge. These balls appear within a few yards of the ground. Sometimes ball lightning is reported to appear near ground in the absence of a lightning discharge. Lightning balls have also been observed to hang in mid-air above the ground and have been observed falling from a cloud toward the ground.

Appearance: Ball lightning is generally spherical, although other shapes have been reported. Balls are usually 4 to 8 in. in diameter, with reported diameters ranging from 1/2 in. to many feet. Various colors have been observed, the most common being red, orange, and yellow. Luminosity is generally not exceptional but the balls can be seen clearly in daylight. They are usually reported to maintain a relatively constant brightness and size, although changes in brightness and size are not uncommon.

Lifespan: Ball lightning generally lasts less than 5 seconds. In a small fraction of the reports a lifespan of over a minute is indicated.

Motion: Lightning balls usually move horizontally at a velocity of a few yards per second. They may also remain motionless in mid-air or may descend from a cloud toward the ground. They do not often rise, as would be the case if they were spheres of hot air at atmospheric pressure subject only to gravity. Many reports describe balls which appear to spin or rotate as they move. Sometimes they are reported to bounce, typically off the ground.

Heat, sound, and odor: Rarely is the sensation of heat reported although there are accounts of lightning balls which burned barns and melted wires. One report (Ref. 15.5) describes a ball lightning which hit a pond of water with a sound "as if putting a red hot piece of iron into the water." Sometimes a hissing sound is said to be emitted. Many observers report an accompanying odor usually described as sharp and repugnant, resembling ozone, burning sulphur, or nitric oxide.

Attraction to objects and enclosures: Reports often describe balls attached to, and moving along, metallic objects such as wire fences or telephone lines. Some or all these observations may refer to a type of St. Elmo's fire. Lightning balls often enter houses through screens or chimneys, and sometimes through glass window panes. They are also reported to originate within buildings, on occasion from telephones. Balls can exist in an all-metal enclosure such as the interior of an airplane. For instance:

> • I was at the controls of a KC-97 USAF tanker aircraft, heavily loaded with JP-4 fuel for offload to B-47 bombers. En route to the refueling rendezvous (Elko, Nevada vicinity) we were in the clouds at 18,000′. There was light precipitation, temp. was above freezing and there was no turbulence.
>
> I recall that St. Elmo's fire was dancing around the edges of the aircraft front windows. (This is a not too uncommon occurrence but may have some significance to you.) The crew was experienced in all phases of all-weather operation and not concerned or apprehensive about any portion of the mission to be accomplished.

As I was concentrating on the panel (no outside visual references were visible) a ball of yellow-white color approximately 18″ in diameter emerged through the windshield center panels and passed at a rate about that of a fast run between my left seat and the co-pilot's right seat, down the cabin passageway past the Navigator and Engineer. I had been struck by lightning 2 times through the years in previous flights and recall waiting for the explosion of the ball of light! I was unable to turn around and watch the progress of the ball as it proceeded to the rear of the Aircraft, as I was expecting the explosion with a full load of JP-4 fuel aboard and concentrated on flying the aircraft. After approximately 3 seconds of amazingly quiet reaction by the 4 crew members in the flight compartment, the Boom operator sitting in the rear of the aircraft called on the interphone in an excited voice describing a ball of fire that came rolling through the aft cargo compartment abeam the wings, then danced out over the right wing and rolled off into the night and clouds! No noise accompanied the arrival or departure of the phenomenon. (Ref. 15.16)

- This communication records the observation of ball lightning in unusual circumstances. I was seated near the front of the passenger cabin of an all-metal airliner (Eastern Airlines Flight EA 539) on a late night flight from New York to Washington. The aircraft encountered an electrical storm during which it was enveloped in a sudden bright and loud electrical discharge (0005 h EST, March 19, 1963). Some seconds after this a glowing sphere a little more than 20 cm [8 inches] in diameter emerged from the pilot's cabin and passed down the aisle of the aircraft approximately 50 cm [20 inches] from me, maintaining the same height and course for the whole distance over which it could be observed.

. . . the relative velocity of the ball to that of the containing aircraft was 1.5 ± 0.5 meters [or yards] per second . . . the object did not seem to radiate heat. . . . the optical output could be assessed as 5 to 10 watts and its color was blue-white. . . . the course was straight down the whole central aisle of the aircraft. (Ref. 15.17)

Demise: Ball lightning decays in one of two modes, either explosively or silently. Explosive decay occurs rapidly and is

accompanied by a loud noise. Silent decay can take place either rapidly or slowly. Most lightning balls apparently exhibit a rapid decay. After the ball has decayed, it is sometimes reported that a mist or residue remains. Occasionally a ball lightning has been observed to break up into two or more smaller balls.

Types: There may be more than one type of ball lightning. For example, the ones that attach to conductors may differ from the free-floating balls; the balls that appear near ground level may differ from those that hang high in the air or those that fall out of a cloud.

Unfortunately, there is at present no adequate theory of ball lightning. For example, no theory can account simultaneously for the degree of mobility, the constancy of light output, and for the fact that the ball does not rise. Despite numerous theoretical models proposed, the causal mechanisms remain unknown. All theories fall into one of two general classes: those in which the energy source for the ball is postulated to be outside the ball (externally powered ball lightning) and those in which the sustaining energy is postulated to be stored within the ball itself (internally powered ball lightning).

In the internally powered models there are essentially six subclasses: (1) Ball lightning is gas or air behaving in an "unusual" way. It has been suggested that the ball lightning is slowly burning gas, is the light produced by the slow recombination of unspecified ions existing in the ball, is due to chemical reactions involving dust, soot, etc., and so on. (2) Ball lightning is a sphere of heated air or heated air containing various impurities at atmospheric pressure – Ref. 15.18. (3) Ball lightning is a very high-density plasma (ionized gas) which exhibits properties characteristic of solid materials – Ref. 15.19. (4) Ball lightning is due to one of several suggested configurations of closed-loop current flow contained by its own magnetic field. It has been shown that plasma containment of this type is not possible in air – Ref. 15.20. (5) Ball lightning is due to some sort of air vortex (like a smoke ring) providing containment for luminous gases. (6) Ball lightning is a high frequency electromagnetic field contained within a thin spherical sheet of ionized air – Ref. 15.21.

In the theories which provide the ball with an external power source, three types of power sources have been suggested: (1) A high-frequency (hundreds of megacycles per second) electromagnetic field, (2) a steady current flow from cloud to ground, and (3) focused cosmic-ray particles. (1) M. Cerrillo (Ref. 15.22) and P. Kapitza (Ref. 15.23) proposed that focused radio frequency energy from the thundercloud could create and maintain a ball lightning. Radio waves with the energy necessary to effect this mechanism have never been observed in thunderstorms. (2) D. Finkelstein and J. Rubinstein (Ref. 15.20), the writer and C. Helstrom (Ref. 15.24), and J. Powell and Finkelstein (Ref. 15.25) have suggested that a steady current flowing from cloud-to-ground would contract in cross section at the ball (originally provided by part of the lightning channel) and that the increased energy input due to the constriction of current could maintain the ball. This type of theory cannot account for the existence of ball lightning inside structures, particularly inside metal structures. (3) V. Arabadzhi (Ref. 15.26) has suggested that radioactive cosmic-ray particles could be focused by the charges of the thundercloud so that the cosmic rays would create an air discharge at one point in space. This would appear to be very unlikely.

Are UFOs (Unidentified Flying Objects) and ball lightning related? There are all kinds of UFOs. Some are hoaxes; some are misinterpretations of well-understood physical phenomena like meteors, not known to the observer or seen under unusual circumstances; some, like re-entry vehicles, are unpublicized objects of our advanced technology; some are

imperfectly understood physical phenomena; some may be due to poorly understood psychological effects; and, finally, some may be spaceships (true flying saucers). Although there are many observations of flying-saucer-like objects reported by reliable people (Refs. 15.27, 15.28), incontrovertible proof that interplanetary spaceships exist – a captured or crashed saucer – has not been forthcoming. The UFO literature referenced (Refs. 15.27, 15.28) and the references contained in that literature make fascinating reading.

But let us return to our original question. A small percentage of the UFO reports are so similar to a certain class of ball lightning reports that they both must refer to the same "imperfectly understood physical phenomena." These particular lightning balls are much larger, much brighter, and much longer-lived than the typical balls. They are reported to be 10 to 20 ft in diameter, give the impression of being as bright as lightning, and may last a minute or more. When such objects appear immediately after lightning, it is clear that they should be called ball lightning. Sometimes, however, they appear near or in clouds or in snow without the apparent presence of lightning, and occasionally such objects are reported in seemingly clear air.

chapter sixteen

What is Ribbon Lightning? Bead Lightning?

Ribbon lightning is the name given to the optical illusion occurring when a cloud-to-ground lightning flash is moved sideways an appreciable distance by the wind during the time between the component strokes of the flash. Each stroke in the flash is then seen separated horizontally in space. To the eye, each identically-shaped stroke (ribbon) appears to occur simultaneously. A photograph of ribbon lightning is shown in Fig. 16.1.

The first solid evidence that lightning flashes are generally composed of a number of separate strokes was provided in the 1880s by photographs of ribbon lightning. By the early twentieth century several investigators had obtained ribbon-like photographs of normal (not wind-blown) lightning by moving their cameras as the flash was occurring. Each component stroke of the flash appeared at a different

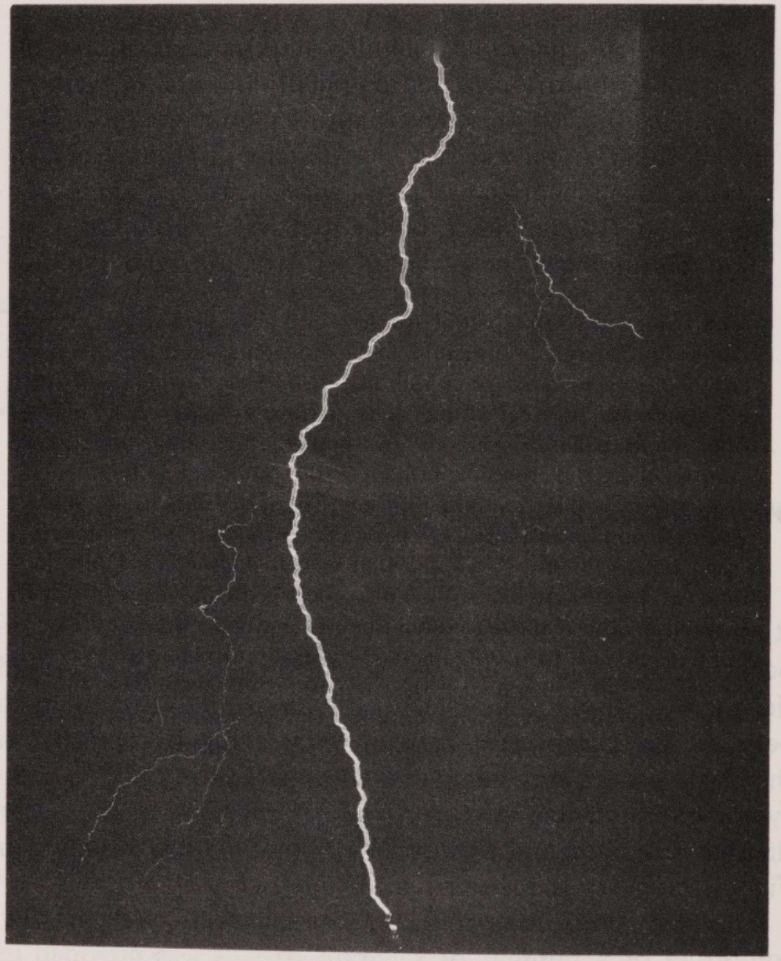

Fig. 16.1. Ribbon lightning near Tucson, Arizona photographed with a tripod-mounted camera. (Courtesy, George Marcek, Catalina High School, Tucson)

position on the film because the lens and film position had changed relative to the fixed lightning position during the time between strokes. The technique of moving (by hand or mechanically) a camera during a flash to time-resolve the features of the flash has yielded much valuable information about lightning. For example, in this way stepped and dart leaders were discovered and their velocities measured. Modern "streak" cameras which time-resolve various high-speed luminous events (including lightning) employ the same basic idea, relative motion between lens and film, as the hand-moved camera. The streak camera lens remains stationary while the motor-driven film moves at high speed horizontally behind the lens. A modern streak camera photograph of a 12-stroke lightning flash is shown in Fig. 16.2. A streak camera photograph of a stepped leader and the resultant return stroke is shown in Fig. 16.3. As is evident, the stepped leader channel is photographically dark between steps and elongates itself by a new step each time it lights up. A detailed description of the stepped leader is given in Chapter 9 and the stepped leader's movement towards ground is illustrated in Figs. 9.1 and 9.2.

Bead or chain lightning is a visually well-documented phenomenon in which the lightning channel to ground breaks up, or appears to break up, into luminous fragments generally reported to be some tens of yards long. The beads are reported to persist longer than the usual cloud-to-ground discharge channel.

No reliable still photograph of bead lightning has been published in the literature. Further, it is not clear that a still camera would record a beaded effect. If the channel is first continuous and then breaks up into beads, the image on the film will be a superposition of the continuous channel and the beads. The beads will only be apparent if the light output of the beads to which the film is sensitive is at least a small fraction of the light output of the channel prior to the formation of beads and if the film is not over-exposed.

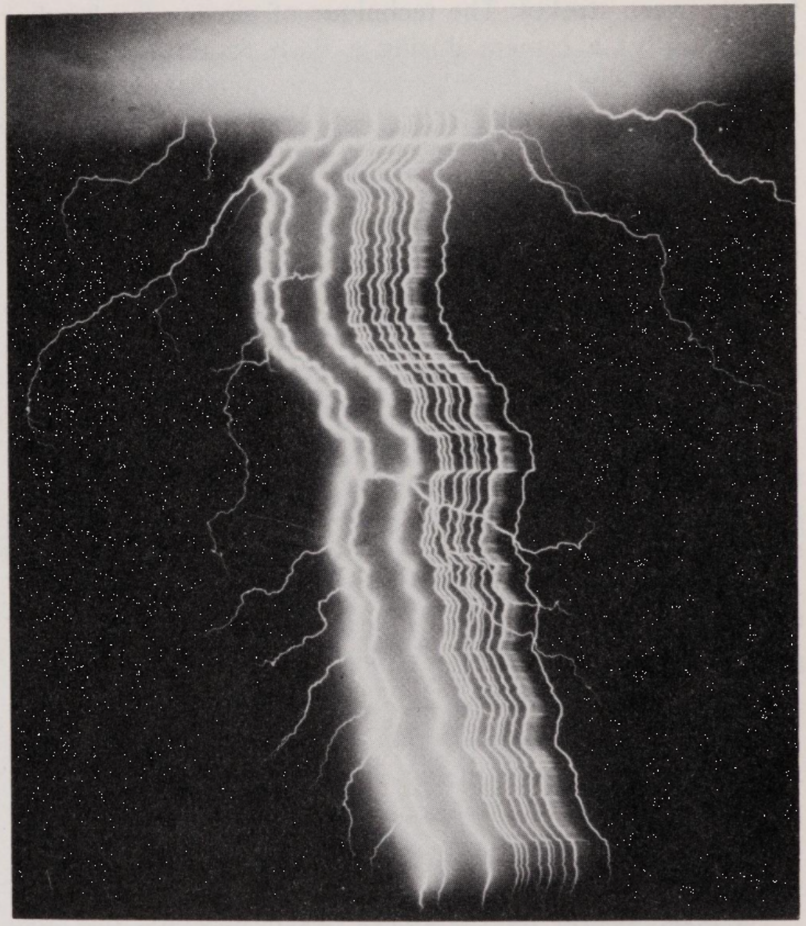

Fig. 16.2. Streak-camera photograph of a 12-stroke lightning flash taken near Socorro, New Mexico. The first stroke is on the left and is the only branched stroke (see Chapter 9). Increasing time goes from left to right. Continuing current (see Chapter 5), as evidenced by continuing luminosity, flows after the eleventh stroke. (Courtesy, Marx Brook, New Mexico, Institute of Mines and Technology)

To document bead lightning adequately, a movie is probably necessary. Figure 16.4 is a frame from one such movie taken by U. S. Navy personnel while photographing a plume of water rising from a depth charge explosion (Ref. 16.1). The lightning flash of Fig. 16.4 which struck the depth charge plume consisted of three strokes, the first lasting more than 0.5 second, the second about 0.1 second, and the last more than 0.1 second. Continuing current (see Chapter 5) was apparently involved in each stroke. The latter frames of each of the three strokes show a channel which has broken up into light and dark sections a few yards in length.

There are various theories of bead lightning and some of these are now considered:

(1) The visual appearance of bead lightning is caused by

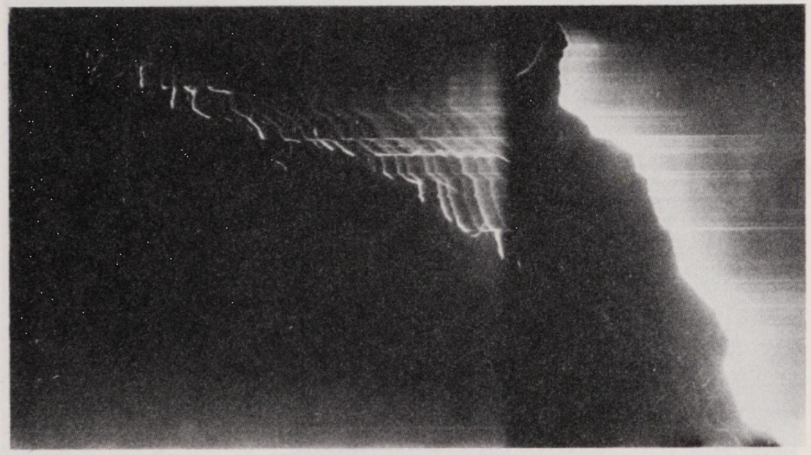

Fig. 16.3. Streak-camera photograph of a stepped leader and the resultant return stroke taken near Lugano, Switzerland. Increasing time goes from left to right; the time duration of the photograph is about 0.002 sec. The vertical height shown is about 400 yards. On the left side of the photograph the intensity of the leader is greatly enhanced. This enhancement was accomplished in printing the photograph in the darkroom. The right side of the photograph gives some idea of the relative brightness of the stepped leader compared to the return stroke. (Courtesy, Karl Berger, Eidg. Technische Hochschule, Zürich)

viewing portions of the lightning channel end-on. That is, if the lightning channel is coming toward or going away from an observer, the observer will see a greater length of the channel within a given viewing angle than would be the case were the channel perpendicular to his viewing direction. The greater length of channel appears to the eye as a normal channel of greater-than-normal brightness. To account for the beaded effect, the channel must periodically be slanted toward or away from the observer.

(2) Bead lightning is due to the periodic masking of a normal cloud-to-ground channel by clouds or rain.

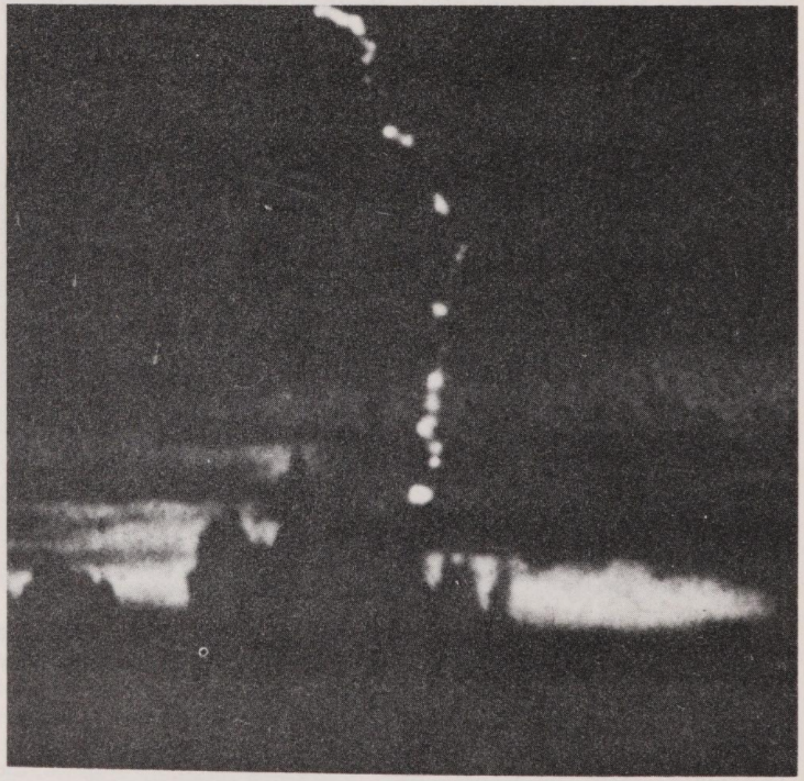

Fig. 16.4. Bead lightning as it appears on one frame of a motion picture showing a three-stroke flash striking the top of a depth charge plume. (Courtesy, U.S. Naval Ordnance Laboratory)

(3) A "magnetic pinch-effect" instability may distort the current-carrying channel into a "string of sausages" with the strong light emission coming from the "necked-off" regions of high-current density. To obtain a series of beads this way, the channel must somehow be conditioned periodically as a function of height. Possibly, the periodicity is just an accident of nature.

(4) Bead lightning is a series of ball lightnings (see Chapter 15).

(5) Large-radius sections of the lightning channel have long luminous lifetimes since channels of large radius take longer to cool than channels of small radius. To a first approximation the cooling time is proportional to the square of the radius. If the lightning channel radius were somehow periodically modulated as a function of height, then, as channel luminosity decayed, the channel would take on the appearance of a string of beads. Perhaps this modulation occurs accidentally when the channel consists of a large number of kinks or bends.

To account for the observed long persistence of the beads, those theories which require current flow may be invoked in conjunction with long continuing currents. Perhaps the observed long persistence is due to the eye rather than to the channel.

REFERENCES

16.1. Young, G. A., A Lightning Strike of an Underwater Explosion Plume, U. S. Naval Ordnance Lab., NOLTR 61-43, March 1962.

A more detailed discussion of bead lightning is given in: *Lightning,* M. A. Uman, Dover Publications, Inc., New York, 1984, Appendix C.

A comprehensive discussion of lightning photography including several streak-camera photographs and references to published photographs of ribbon lightning is given in *Lightning,* Chapter 2.

chapter seventeen

Has Lightning Any Practical Use?

Energy is neither created nor destroyed. Rather, it is converted from one form to another. For example, in a light bulb the input electrical energy is converted to light and heat. The energy content of the output light and heat equals that of the input electrical energy. No energy is lost. A charged thundercloud contains stored electrical energy derived from the energy of winds and particle motions and from the energy released by condensing water vapor (see Chapter 7). Some of the stored electrical energy of the cloud is transferred to lightning. We consider now the energy and power available to lightning and the possibility of its controlled use for the benefit of man.

Each cloud-to-ground lightning flash involves tens of coulombs of charge (see Chapter 8) at a voltage of between 100 million and a billion volts. The resultant electrical energy is about 1 to 10 billion watt-seconds per flash. If there are 100 flashes to ground each second over the whole world (see Chapter 7), a maximum value for the total electrical power input to worldwide cloud-to-ground lightning is 1,000 billion watts. By comparison, the total capacity of all electric power generators in the United States in 1970 was about 500 billion watts. It is evident, therefore, that the power and energy available to lightning is appreciable.

Unfortunately, no efficient method of tapping lightning's power and energy is presently available – for two reasons: (1) It is impractical to intercept (for example, with tall towers) any significant number of the world-wide cloud-to-ground flashes. (2) Most of the energy available to the lightning is converted along the lightning channel to thunder, heat, light, and radio waves, leaving only a fraction available at the channel base for immediate use or storage.

The first problem is the more serious. If its total energy were available, a *single* lightning flash would run an ordinary household light bulb for only a few months. It is the high rate of occurrence of world-wide lightning (100 per second) which provides for the high total power and energy levels. A 1000 ft tower in a region of moderate thunderstorm activity is struck by lightning about 10 times a year (see Chapter 6). It would take one hundred thousand such towers, each collecting all the energy from ten strokes per year, to equal the 100 million watts generated by a typical small power station. It is impractical to erect a sufficient number of towers to capture significant quantities of energy from lightning, and no more practical means of intercepting lightning has been proposed.

Ancient man considered lightning the ultimate weapon. Its use and control were generally attributed to the gods. The Norse god Thor hurled lightning from his chariot while riding across the sky. In ancient Greece, Zeus threw the thunderbolts; in Rome, Jupiter. The advent of our modern civilization has not eliminated man's interest in the use of lightning as a weapon. Military research in several countries has recently been concerned with the feasibility of directing lightning and ball lightning at specific targets. While technology is not capable of achieving these goals at present, it would not be surprising if they were attained in the not too distant future. In view of the awesome destructive power of modern weaponry, the military use of lightning or ball lightning would probably be more as a psychological than as a destructive weapon.

Nitrogen comprises about 80% of the atmosphere surrounding the earth, yet it cannot be used directly by the large majority of plants and animals until it is "fixed." "Fixed" nitrogen is nitrogen incorporated in chemical compounds necessary to the chemical processes of life, as opposed to the relatively inert form of nitrogen, the nitrogen molecule, found in the air. Fixing is accomplished by (1) special organisms in the soil and waters, (2) industrial processes, and (3) ionizing atmospheric processes including lightning. Nitrogen which has been fixed in the atmosphere is brought to earth in rain. It has been estimated that of the 100 million tons of nitrogen fixed per year, about 8% is atmospheric, about 30% industrial, and the remainder biological (Ref. 17.1). Estimates of this sort are difficult to make and subject to large errors. The role played by lightning in generating fixed nitrogen is frequently accorded prominence in the popular literature. We can, however, show that lightning is not as important in nitrogen fixing as is generally believed.

About 8 million tons of fixed nitrogen are thought to be brought to earth each year in rain. If we assume that (1) there are 500 lightning discharges (cloud-to-ground and intracloud) occurring each second (see Chapters 7 and 8), (2) each lightning channel is 2 in. in diameter and 4 miles long (see Chapter 10), (3) every nitrogen atom in every channel is "fixed" (a great overestimate), then the total amount of nitrogen fixed by the lightning channels is about 20,000 tons per year. Since 20,000 tons is a very small fraction of 8 million tons, we conclude on theoretical grounds that the lightning channel itself is not an important producer of fixed nitrogen. Experimental evidence to support this conclusion has been given by several investigators (Refs. 17.2 to 17.5) who show that the nitrate and ammonia content in precipitation is not correlated with the amount of lightning activity. In fact, there is roughly the same nitrate and ammonia content in precipitation in winter when there are few thunderstorms as in summer when there are many (Ref. 17.2). What, then, is the source of the atmospheric nitrogen compounds? The answer is not known, but there are two reasonable possibilities: (1) They are generated in the lower regions of the ionosphere (the ionized layer of atmosphere beginning about 40 miles above the earth) and diffuse downward, and/or (2) they are created by corona discharges in all types of precipitating clouds and along the ground under those clouds (Ref. 17.5). If possibility (2) is important, a close relative of lightning has contributed and continues to contribute significantly to the world's fixed nitrogen supply.

The lightning channel generates radio waves (sometimes called atmospherics or sferics) which have been used for practical purposes. Lightning is the strongest terrestrial source of electromagnetic noise in the radio band. The lightning radio noise is strongest near a frequency of 5 kilocycles per second and extends with decreasing intensity up and down the frequency scale as far as has been measured. Everyone is familiar with the lightning static produced on AM radios and the "snow" on TV pictures (particularly on the lower channel numbers) due to lightning radiation. Studies of the lightning electromagnetic noise have led to several practical applications. (1) Measurements made on lightning radio waves can be used to pinpoint the location of the thunderstorm area producing the lightning. Among those who use this information are meteorologists (who study as well as predict the weather) and aircraft pilots. Radiolocation techniques can spot thunderstorms thousands of miles away. (2) Measurements of the properties of lightning radio waves returned from the ionosphere and from ionized layers above the ionosphere have yielded valuable information about the numbers and kinds of charged particles in those regions. (3) Measurements of the properties of lightning radio waves as they propagate along the earth's surface have enabled the electrical resistivity of various portions of the earth to be determined. Similarly, lightning radio waves can be used in some kinds of geophysical prospecting (e.g., in the search for low resistivity ore lying far below the earth's surface).

REFERENCES

17.1. Delwiche, C. C., The Nitrogen Cycle, *Sci. American, 223* (No. 3), 137-146, September 1970.

17.2. Viemeister, P. E., Lightning and the Origin of Nitrates Found in Precipitation, *Jour. of Meteorology, 17*, 681-683 (1960)

17.3. Wetselaar, R., and J. T. Hutton, The Ionic Composition of Rainwater at Katherine, N. T., and its Part in the Cycling of Plant Nutrients, *Australian J. Agr. Res., 14*, 319-329 (1963).

17.4. Gambell, A. W., and D. W. Fisher, Occurrence of Sulfate and Nitrate in Rainfall, *J. Geophys. Res., 69*, 4203-4210 (1964).

17.5. Reiter, R., On the Causal Relation between Nitrogen-Oxygen Compounds in the Troposphere and Atmospheric Electricity, *Tellus, 22*, 1-135 (1970).

ADDITIONAL READING

Griffing, G. W., Ozone and Oxides of Nitrogen Production during Thunderstorms, *J. Geophys. Res., 82,* 943-950 (1977).

Chameides, W. L., D. H. Stedman, R. R. Dickerson, D. W. Rusch, and R. J. Cicerone, NO_x Production in Lightning, *J. Atmosph. Sci., 34,* 143-149 (1977).

Hill, R. D., R. G. Rinker, and A. Coucourinos, Nitrous Oxide Production by Lightning, *J. Geophys. Res., 89,* 1411-1421 (1984).

Levine, J. S., T. R. Augustsson, I. C. Anderson, and J. M. Hoell, Jr., Tropospheric Sources of NO_x: Lightning and Biology, *Atmospheric Environment, 18,* 1797-1804 (1984).

chapter eighteen

What Would Happen If Lightning Were Eliminated?

First, it is appropriate to point out that it may well have been lightning in the primordial soup covering the earth several billion years ago that produced the complex molecules from which life eventually evolved. Laboratory experiments have shown that electrical discharges in what is believed to be the constituents of the primordial atmosphere can create the necessary molecules. Thus, we may be indebted to lightning for the presence of life on earth.

Still, lightning is primarily troublesome rather than helpful, as the contents of this book attest, and except for some of the benefits mentioned in the previous chapter it might seem advantageous to eliminate it altogether. Exactly what would happen if lightning were eliminated is not known, but clearly the electrical balance of the atmosphere would have to change as we shall see in the following paragraphs. If and

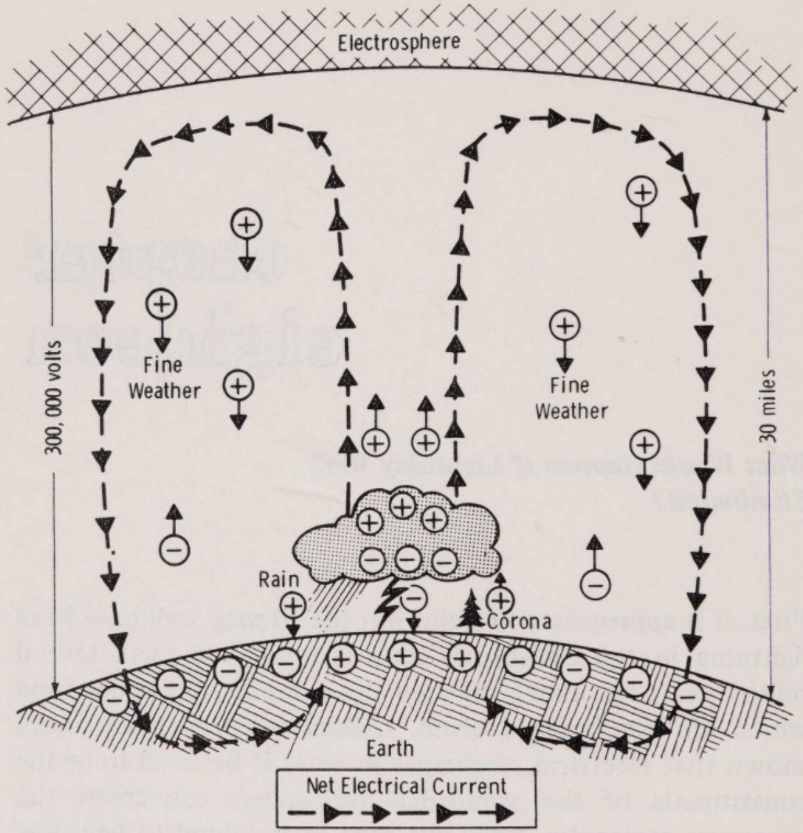

Fig. 18.1. Thunderstorms act as batteries to keep the earth charged negatively and the atmosphere charged positively. Atmospheric electrical currents flow downward in fine weather and upward in thunderstorms. Thunderstorms deliver charge to the earth by lightning, rain, and corona discharges.

how the change in electrical balance would change the weather is not known. Prudence, however, would dictate extreme caution in tampering with lightning.

The electrical resistivity of the atmosphere decreases with height. From the point of view of atmospheric electricity, the resistivity is sufficiently low at an altitude of about 30 miles that the voltage does not vary much above that point. The region beginning at about 30 miles and extending upward is called the *electrosphere*. (The ionosphere, which reflects lightning radio waves downward, begins, as noted in Chapter 17, at about 40 miles and is even a better conductor than the bottom of the electrosphere.) The voltage between the earth and the electrosphere in regions of fine weather is about 300,000 volts. To maintain this voltage the earth has a negative charge of about a million coulombs on its surface and an equal net positive charge is distributed throughout the atmosphere. Measurements have shown that the negative charge on the earth remains roughly constant with time. At first glance, this fact is difficult to understand since the charge on the earth is continuously leaking off into the conducting atmosphere. In fact, calculations show that, if the earth's charge were not being continuously re-supplied, the charge on the earth would disappear in less than an hour.

The earth is recharged by thunderstorms. Fig. 18.1 shows how the electrical balance of the atmosphere is maintained. Thunderstorms deliver a net negative charge to the earth as a result of the sum of the effects of the following processes: (1) negative charge carried from cloud to earth by lightning, (2) positive charge carried from cloud to ground by rain, and (3) positive charge carried upward (the equivalent of negative charge carried downward) through the air beneath and above a thunderstorm, the source of the positive charge being corona discharge off grass, trees, and other objects with sharp points on the ground beneath the thunderstorms. The total current flowing beneath all thunderstorms in progress

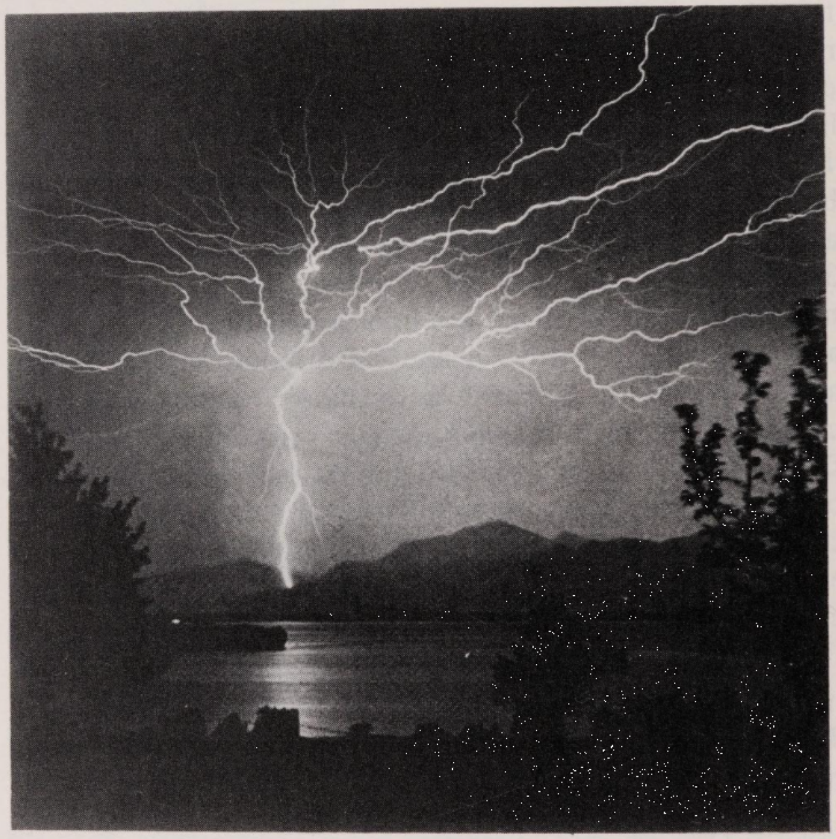

Fig. 18.2. Lightning over Lake Chiem in Bavaria, Germany. (Courtesy, Agfa-Gavaert AG)

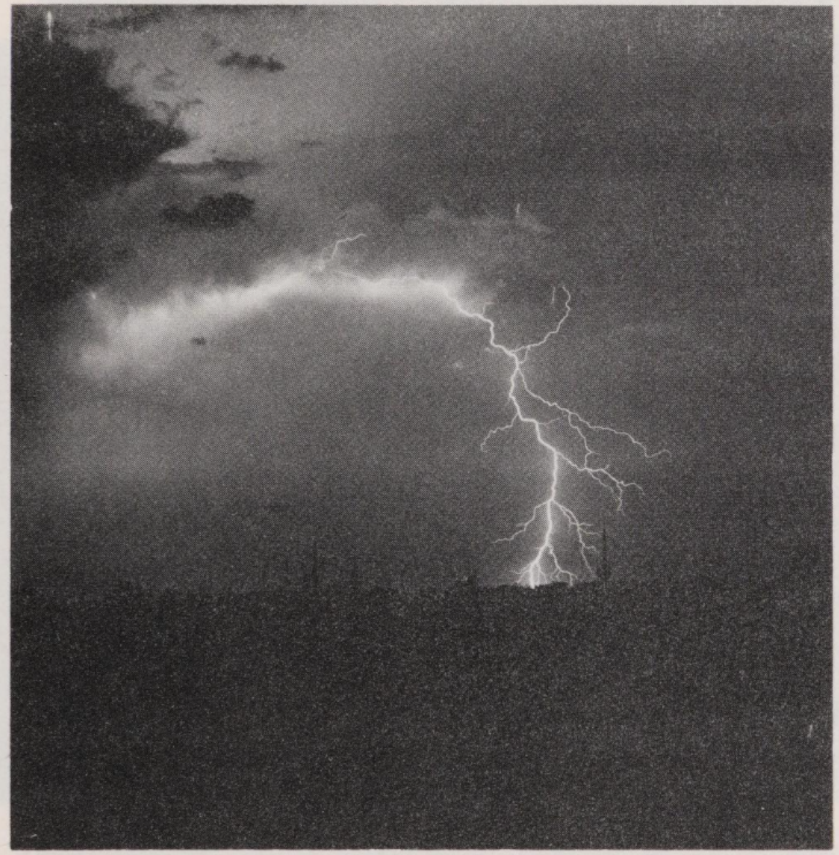

Fig. 18.3. Lightning over the Arizona desert. (Courtesy, J. Rodney Hastings, University of Arizona)

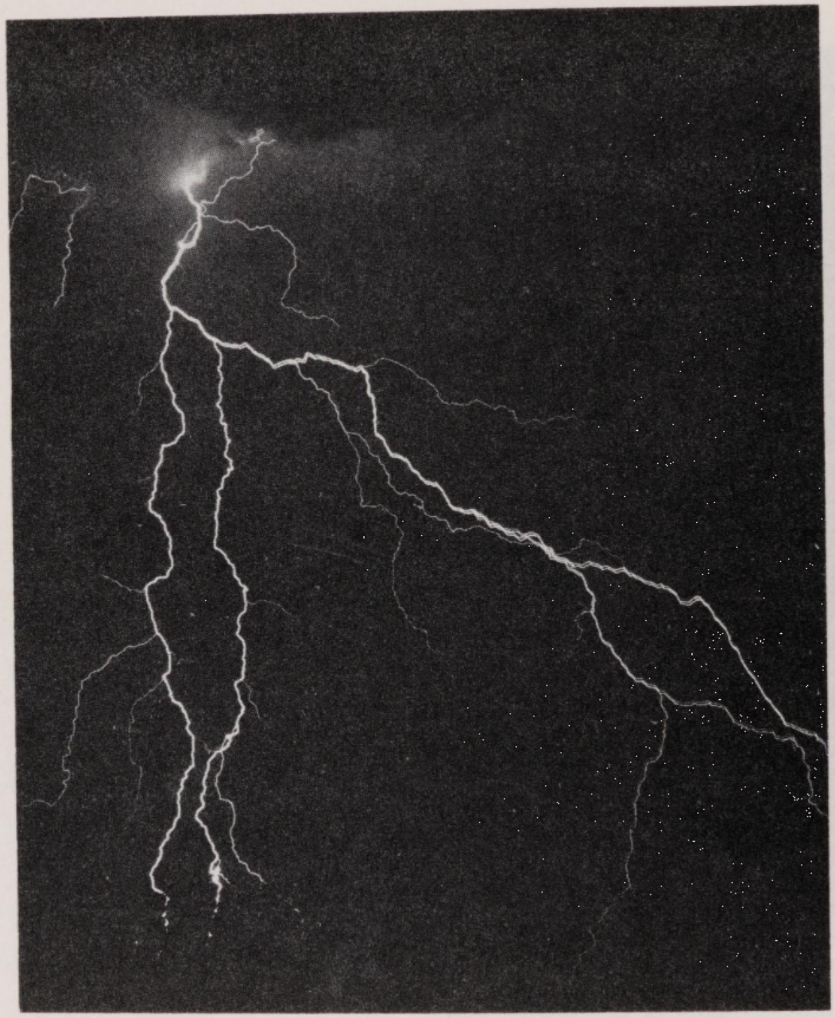

Fig. 18.4. A lightning flash exhibiting multiple ground points and the ribbon effect discussed in Chapter 16. (Courtesy, George Marcek, Catalina High School, Tucson)

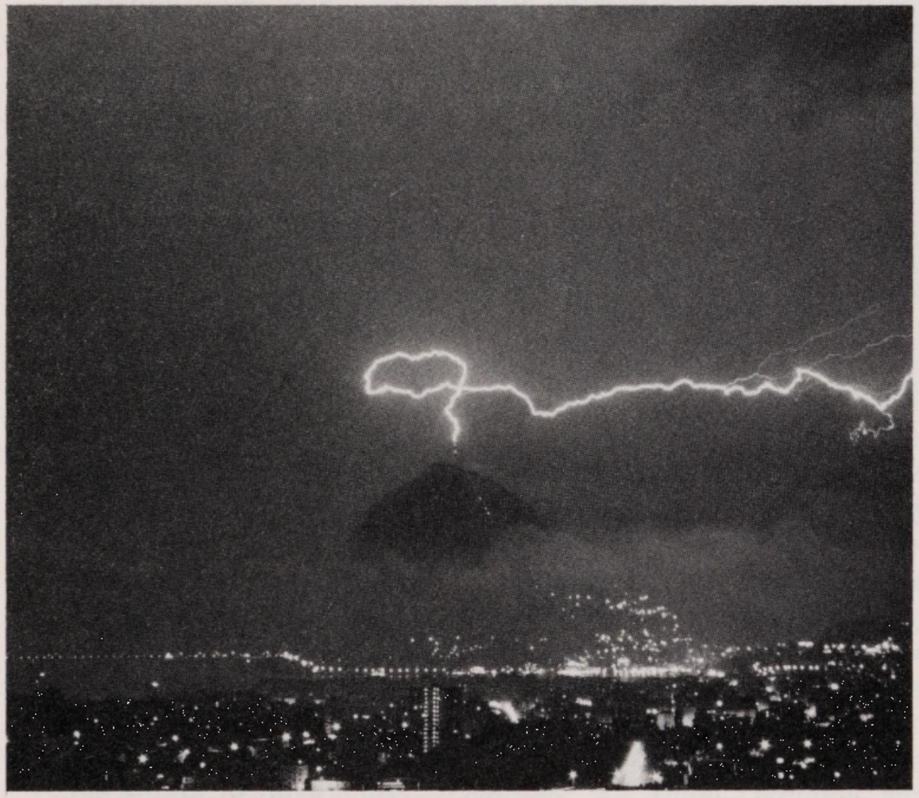

Fig. 18.5. Lightning initiated by an upward-moving leader from a tower on Mt. San Salvatore near Lugano, Switzerland. Photographs of other discharges to the tower are shown in Fig. 6.1a, b. (Courtesy, Richard E. Orville, State University of New York at Albany)

throughout the world at any given time is thought to be about 2000 amps, and is in such a direction as to charge the earth negatively. An approximately equal and opposite current flows in regions of fine weather. The result is that the net negative charge on the earth and the equal and opposite net positive charge in the atmosphere remain approximately constant.

Finally, if all lightning were eliminated, we earthly viewers would be deprived of one of the most spectacular visual displays that Nature has to offer. In its variety and brilliance lightning puts man-made fireworks to shame, as evidenced by Figs. 18.2 to 18.5. Watching lightning is fun; one of the aims of this book has been to make it more so. Many confirmed lightning watchers even enjoy listening to the variety of sounds that comprise thunder.

REFERENCES

The details of and references to the bulk of the material found in this chapter are to be found in *Atmospheric Electricity*, J. A. Chalmers, 2nd Edition, Pergamon Press, New York, 1967, pp. 33-35, 292-308.

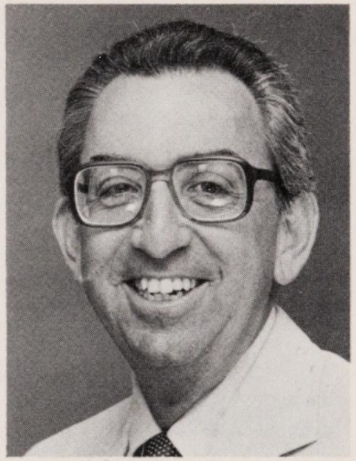

Martin A. Uman received the B.S.E. degree in electrical engineering in 1959 from Princeton University, where he was elected to Phi Beta Kappa. After receiving his Ph.D. from Princeton University in 1961, he took a post as Associate Professor of Electrical Engineering at the University of Arizona, where he became interested in lightning research, particularly in lightning spectroscopy. In 1965, Dr. Uman moved to the Westinghouse Research Laboratories in Pittsburgh, where he continued his lightning research and, in addition, studied the electrical, optical, and acoustic properties of long laboratory sparks. He has been Professor of Electrical Engineering at the University of Florida since January 1972, where his primary research interest has been in the electromagnetic radiation produced by lightning. In 1976, Dr. Uman and Dr. E. P. Krider, Professor of Atmospheric Sciences at the University of Arizona, formed Lightning Location and Protection, Inc. (LLP), a small company that specializes in consulting on unusual lightning problems and manufacturing lightning detection equipment. Dr. Uman served as president of LLP from 1976 until 1983. Dr. Uman is the author of three books—*Lightning, All About Lightning,* and *Introduction to Plasma Physics*—and approximately 85 journal articles, and has served as Associate Editor of the *Journal of Geophysical Research.* He is a member of the International Commission on Atmospheric Electricity, the American Geophysical Union, and the IEEE, and is a fellow of the American Meteorological Society.

ALL ABOUT LIGHTNING

Martin A. Uman

". . . contains a surprising wealth of well-organized, authoritative information about lightning."—Bulletin of Amer. Meteorological Society

Does lightning strike twice in the same place? How does a lightning rod work? What is ball lightning? How many thunderstorms are in progress in the world at any one time? Why does lightning zigzag? What is St. Elmo's Fire?

These and many more often-asked questions about lightning are answered in this fascinating and informative guide for the layman, presented in an easy-to-follow question-and-answer format. One of nature's most awesome phenomena, lightning has intrigued man since earliest times. In this book, a noted scientist and expert on lightning dispels many misconceptions while offering a wealth of scientific and technical information about the nature of lightning and its effects.

You'll discover how Benjamin Franklin proved that lightning was electrical, how to protect yourself from lightning, how to photograph lightning (it's not difficult), the possible relationship between ball lightning and UFOs, what to do for a person struck by lightning, the nature of sheet lightning, ribbon lightning, bead lightning and other variations, and much more. While the overall approach is nontechnical, Dr. Uman has incorporated scientific data in the answers in such a way that laymen will find the book a near-painless introduction to current scientific knowledge about lightning.

Simple, well-drawn diagrams illuminate the text, along with a selection of spectacular lightning photographs, including a remarkable image of 5 lightning bolts produced by the explosion of the first thermonuclear device. In addition, each chapter contains a list of references cited in the text which suggest further reading for anyone interested in finding out more about earth's dazzling atmospheric fireworks.

Unabridged, corrected Dover (1986) republication of *Understanding Lightning*, published by Bek Technical Publications, Inc., Carnegie, Pa., 1971. Preface. 16 black-and-white illustrations. 27 black-and-white photos. Color photos on covers. Bibliography. Index. 192pp. 5⅜ × 8½. Paperbound.

ALSO AVAILABLE

LIGHTNING, Martin A. Uman. 320pp. 5⅜ × 8¼. 64575-4 Pa. $8.95

Free Complete Dover Science Catalog available upon request.

For caption to front cover illustration, see frontispiece inside book.

ISBN 0-486-25237-X

9780486252377
2016-03-09 9:46
22

You can now help support WorldwideDX when you shop on Amazon at no additional cost to you! Simply follow this Shop on Amazon link first and a portion of any purchase is sent to WorldwideDX to help with site costs.

Loop Antenna Crystal Radio

C2
Sr. Member

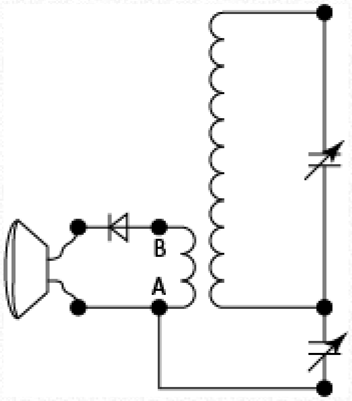

View attachment 15935 I built a fully integrated loop antenna crystal radio on a 23" diameter form (one slat of plastic/wood window blind). I like it because it is completely portable and does not require any additional antenna.

11 turns form the primary and 3 turns the secondary, small gauge phone line stranded copper, all close wound. All turns work out to about 1/2" wide. The primary currently measures 216 uH and I think the secondary was around 40 uH (I'll check this again as I think it was actually much lower, like 16 uH even).

I used a dual gang 15-400, 12-200 pF air variable capacitor. I tried connecting the 12-200 to either point A or B, or not at all. Point B gave about 2 uA more current in the secondary and Point A about 2 uA more again, so I left it at point A.

It tunes from a little over 600 kHz to a little over 1800 kHz. I'm able to pick up 10 stations and the antenna is highly directional, I can completely null the strongest station by turning the loop in-line to the station. The strongest station produces 18 uA in the secondary. What is interesting to note, however, is that I was using an RF generator to find the limits with its output into an open wire loosely wrapped around the antenna. It might be the generator output, but the 600 kHz resonance point produced much less current (~60 uA) than the top end.

I would like to lower its response by 100 kHz. Any suggestions? I was thinking of adding a turn or two to the primary, but given the note above, I'm not sure if something else would be more appropriate?

Automated Interference Hunting - With InterferenceAdvisor

Get automated interference hunting with InterferenceAdvisor.

viavisolutions.com

Jun 11, 2015 Share Link #1

1 | Trade-in your ham gear - Get a FlexRadio
Take any of your used ham gear and trade it in for a FlexRadio! flexradio.com

2 | Magnetic Loop Receive Antenna - Great Noise Fighter
Great Complement to Vertical and Wire Transmit Antennas. w6lvp.com

C2
Sr. Member

I added one more turn to the primary and that brought the frequency response from 540 to 1700 with increased sensitivity at the low end.

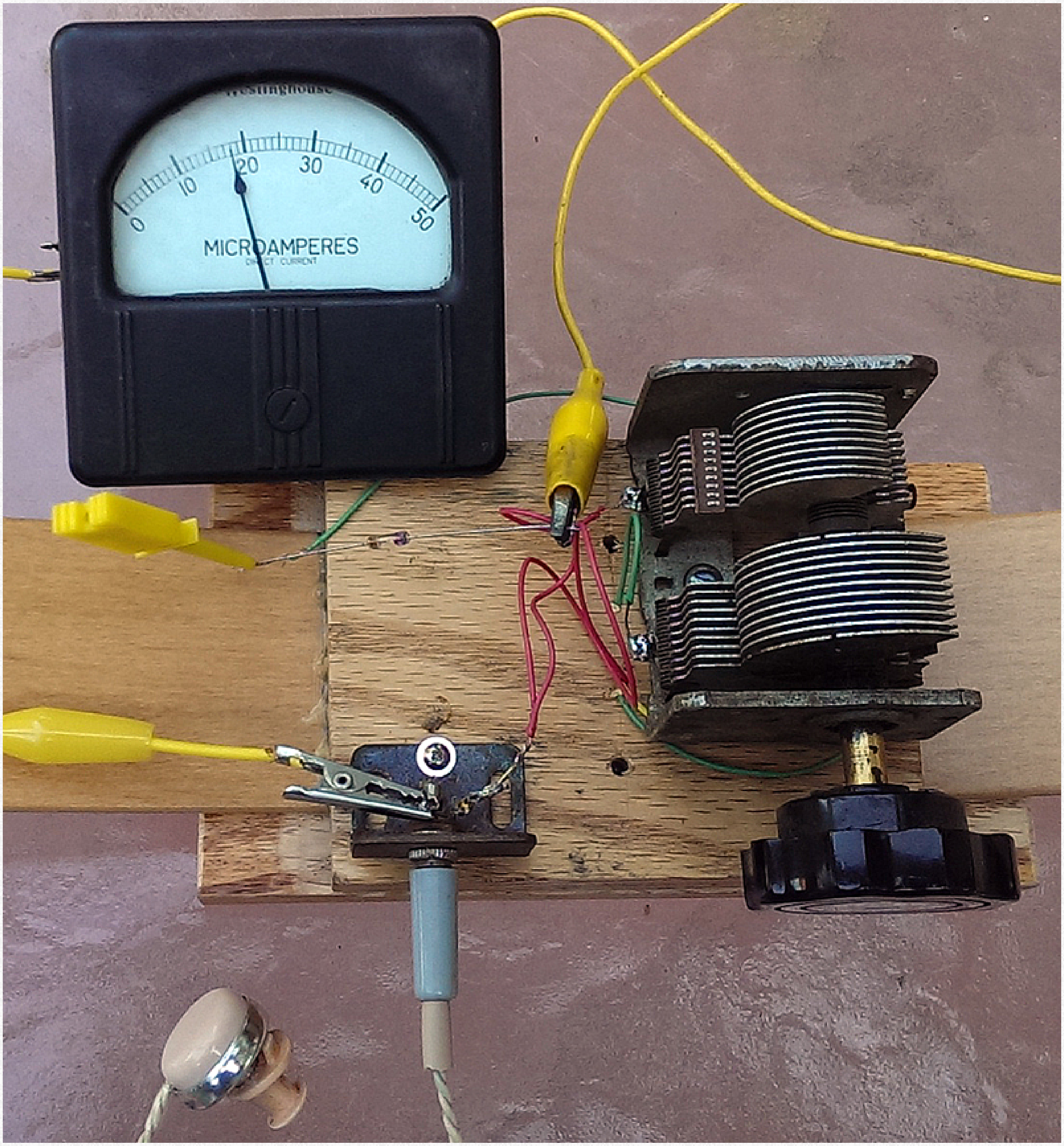
Westinghouse
0
10
20
30
40
50
MICROAMPERES
DIRECT CURRENT

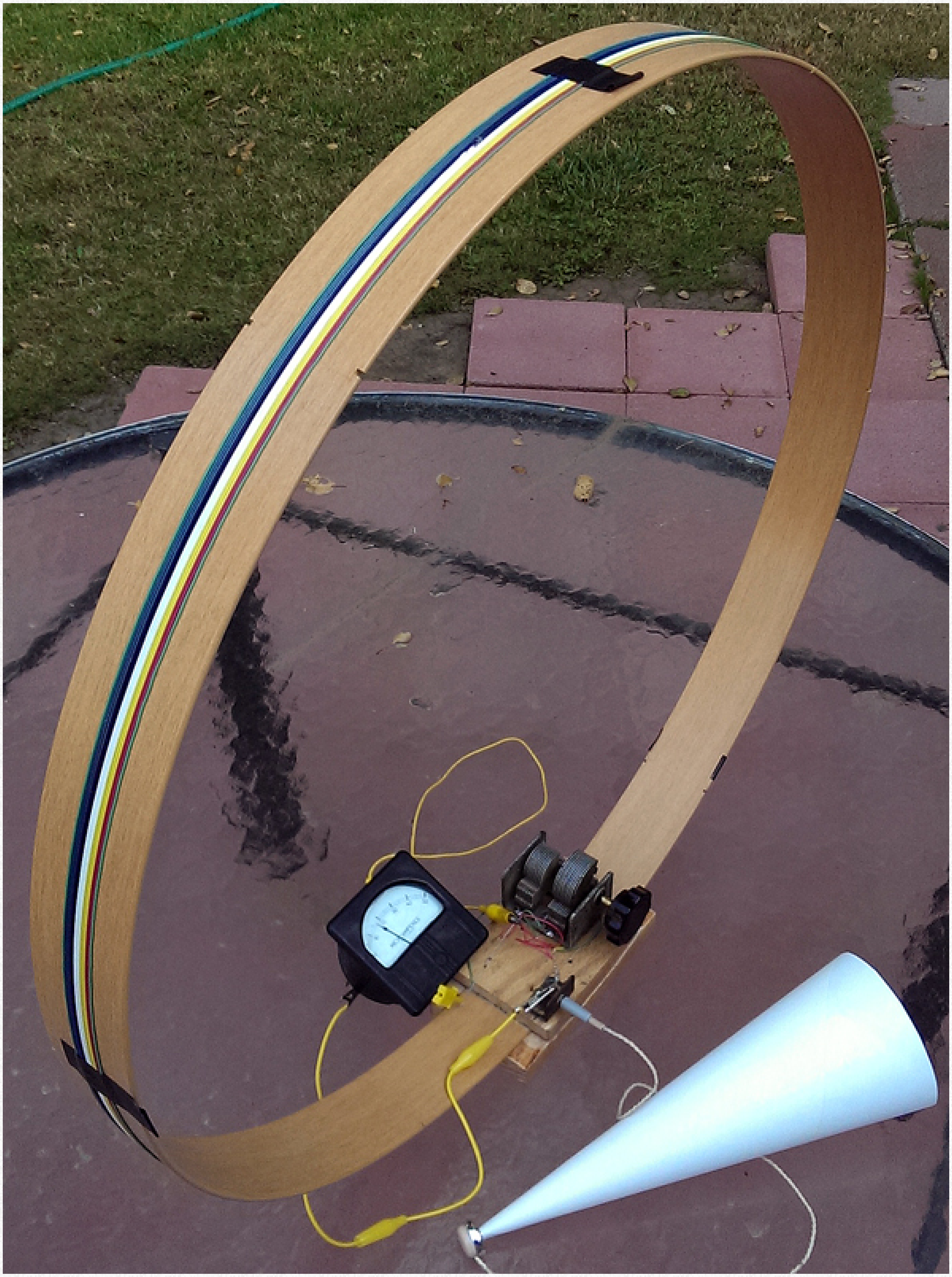

Jun 12, 2015

Share Link #2

The RadioBoard Forums

Quick links FAQ Register Login

Home ‹ **Board index** ‹ **Homemade Receivers** ‹ **Crystal Radios**

Loop-Stick Crystal Set

Post Reply | Search this topic… | 16 posts ‹ 1 **2**

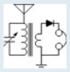

Dan McGillis

Mon Jan 26, 2009 12:52 pm

Hi all.

I used this little loop-stick radio in the just-ended ‘09 Crystal Radio DX contest. Wanted to see how it would do as a simple double-tuned rig.

Not too bad: 84 stations ID’d in about 13 hrs of listening time.
-------A big old 200’ antenna certainly helps a bit 🙂

This setup makes a nice progressive radio for the grandkids (& grandpa):

- start out with the basic cat’s whisker detector (thanks Mike),
- add a good FO-215 diode,
- add a Bogen+Benny for some audio matching,
- add an antenna tuner,

and you wind up with a decent little set.

Here’s a picture.

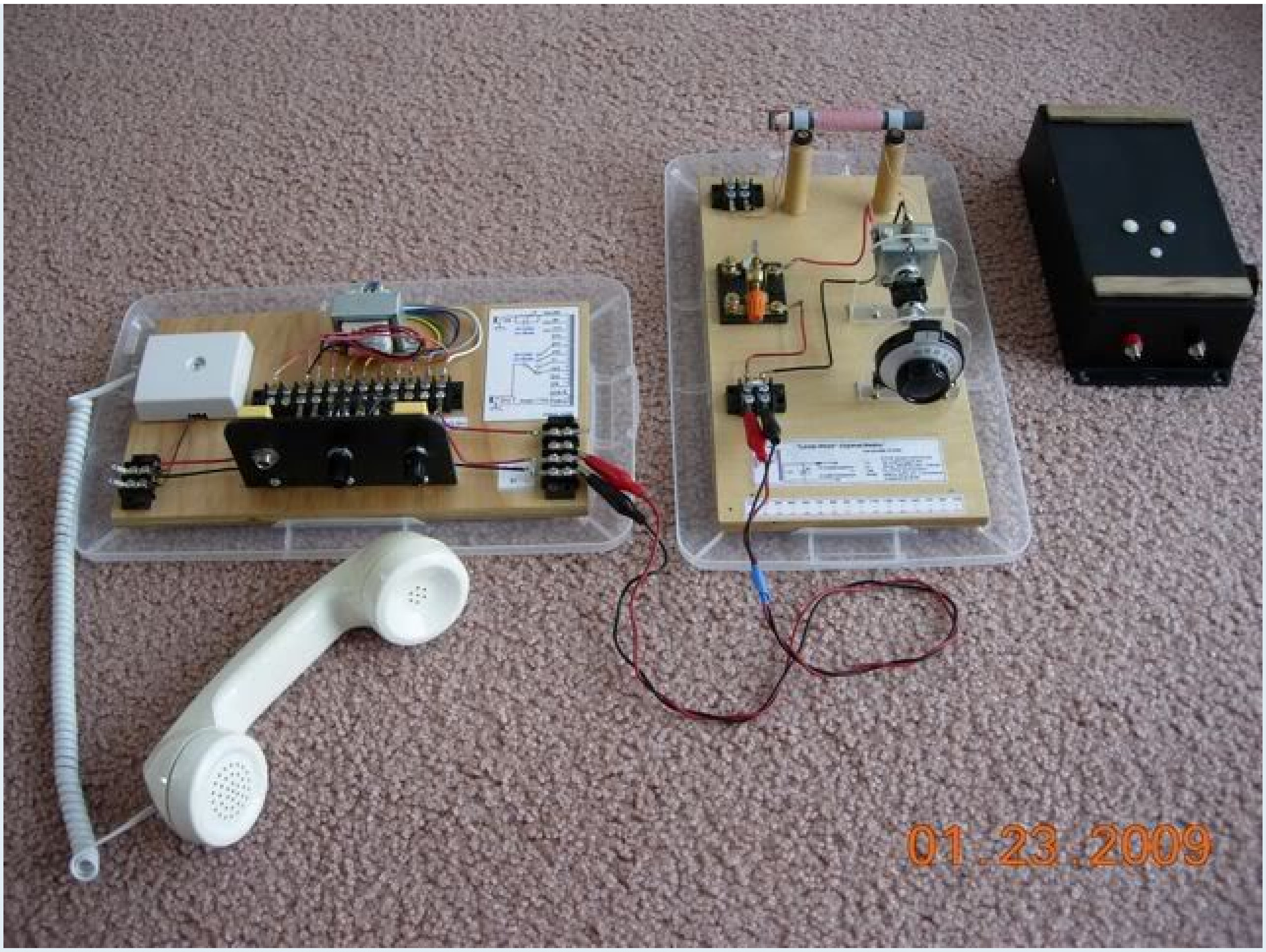

Fun stuff.

73, Dan

Post Reply

16 posts 1 2

Return to "Crystal Radios"

Jump to

Home < Board index The team Delete all board cookies All times are UTC

Powered by phpBB® Forum Software © phpBB Limited

Loop-Stick Crystal Set

Post Reply Search this topic… 16 posts 1 2

Dan McGillis

Loop-Stick Crystal Set

Fri Sep 26, 2008 2:25 pm

I wanted a simple crystal set that would highlight Mike Peebles' rock stand ('cause I really like it) and one that could serve as a nucleus for a project radio. This little "loop stick" radio seems to be just the ticket.

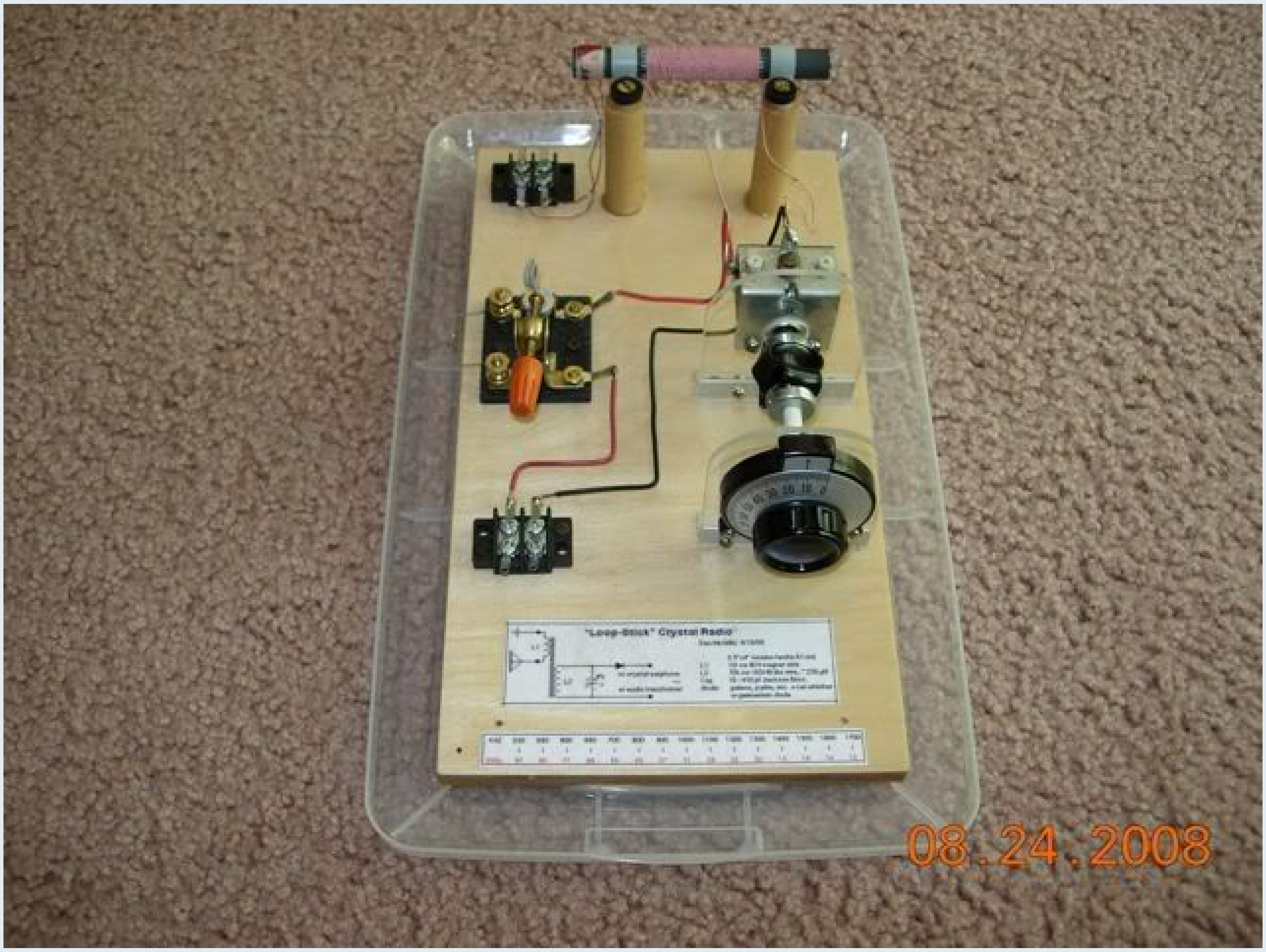

0.5"x4" Amidon ferrite-61 rod,
10t cw #24 magnet wire for the antenna coupling coil,
55t cw 165/46 litz (about 230 uH),
10-410 pf Jackson Bros. capacitor,
Peebles rock stand + crystal, or a diode.

The antenna-ground is connected at the top terminal block and the audio comes out the bottom terminal block.

It works pretty good by itself with a piezo earpiece. But it really perks-up when some bells & whistles are added. Like an antenna coupling capacitor, and/or a Bogen+Benny at the output, and/or a sensitive headset, and/or a loose coupled antenna tuner, etc., etc.. It's kind of fun to fiddle around with it.

A problem I've always had is DUST. I'm banished to the basement for crystal set work and the sets seem to accumulate a thick coating of dust in short order if they're not covered-up. But I wanted this little setup to be out in the open so the grandkids could see it & maybe increase their interest in radio. (It's hard to compete with video games!)

I found some $1.50 clear plastic storage boxes - with lids - at Bed Bath & Beyond that make nice dust covers. The radios' wooden base is screwed to the inside of the lid and the box part then sits on top. No dust on the parts and you can see the inner workings through the plastic. 🙂 The inside clearance is almost 4" so there's plenty of room in there - especially if ferrite-assisted coils are used.

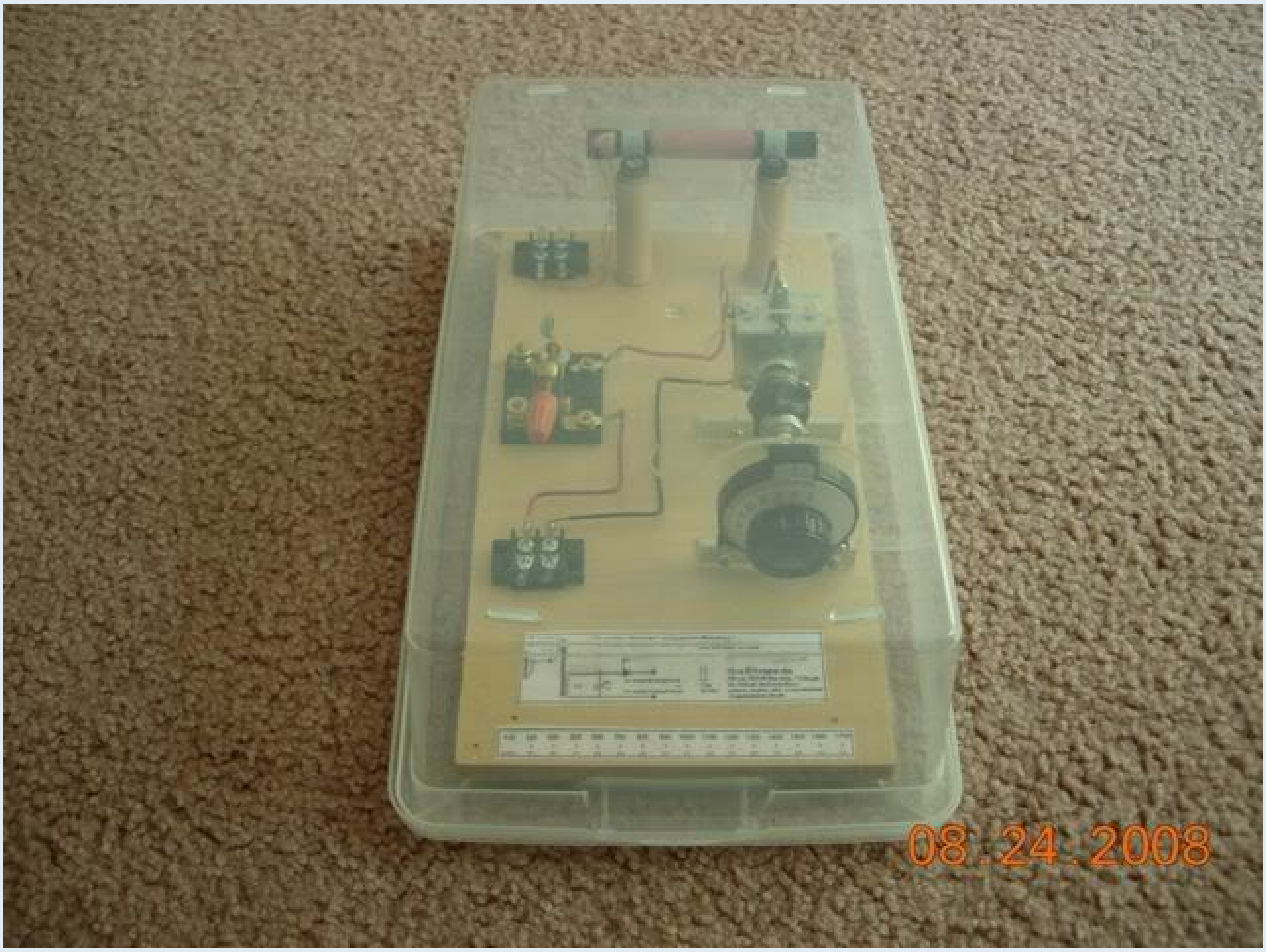

Not elegant, but it works.

73, Dan

OErjan

Fri Sep 26, 2008 3:38 pm

heh, great idea and a good looking set,
I store my competition rig (ATU, detector unit, Phones, ultimatch2...) in a similar box, just MUCH larger (22liters), no

wooden base screwed to the top, just store.
I also use the box as table for the rig in use (it is Poly Propylene so low loss at RF).

Last edited by OErjan on Fri Sep 26, 2008 3:43 pm, edited 1 time in total.

peeblesoriginals

Fri Sep 26, 2008 3:42 pm

Dan,

Looks great!

Mike

corne

Tue Sep 30, 2008 6:01 am

Hi Dan,

Nice set and I like the dust cover idea.

Corne

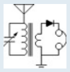

Dan McGillis

Tue Sep 30, 2008 2:42 pm

Thanks fellas.

And thanks for all the work you guys have posted -- I sure have learned a lot from you nice folks.

73, Dan

homebrew

Tue Sep 30, 2008 9:06 pm

Does anyone know why production loopstick antennas are multi layered?

Ham-er

loopstick layering

Tue Sep 30, 2008 11:53 pm

> **homebrew wrote:**
> Does anyone know why production loopstick antennas are multi layered?

Good question Homebrew,

My best guess is cost(maunfacturer being cheap). As most of the commercial "loopstick" are probably intended to go into cheap transistor superhet desing radio,s, the loopstick doesnt need the best Q or efficiency.
In these sets the selectivity is controlled by the IF stage, and the sensitivity by amplification either in the RF amp if it has one, or the converter transistor if not, and by the AGC.

Can anyone verify my guess?

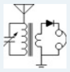

Dan McGillis

Tue Oct 14, 2008 12:45 am

To complete the story - I used the little loop-stick radio in the just finished ‘08 Sprint contest. It was single tuned (no antenna tuner) and no diode was used - only “crystal” detectors, primarily galena. Kyocera KBT-44SB-1A piezo elements made-up the headset. I did put a variable capacitor in the antenna line to vary the coupling.

The log is shown below:

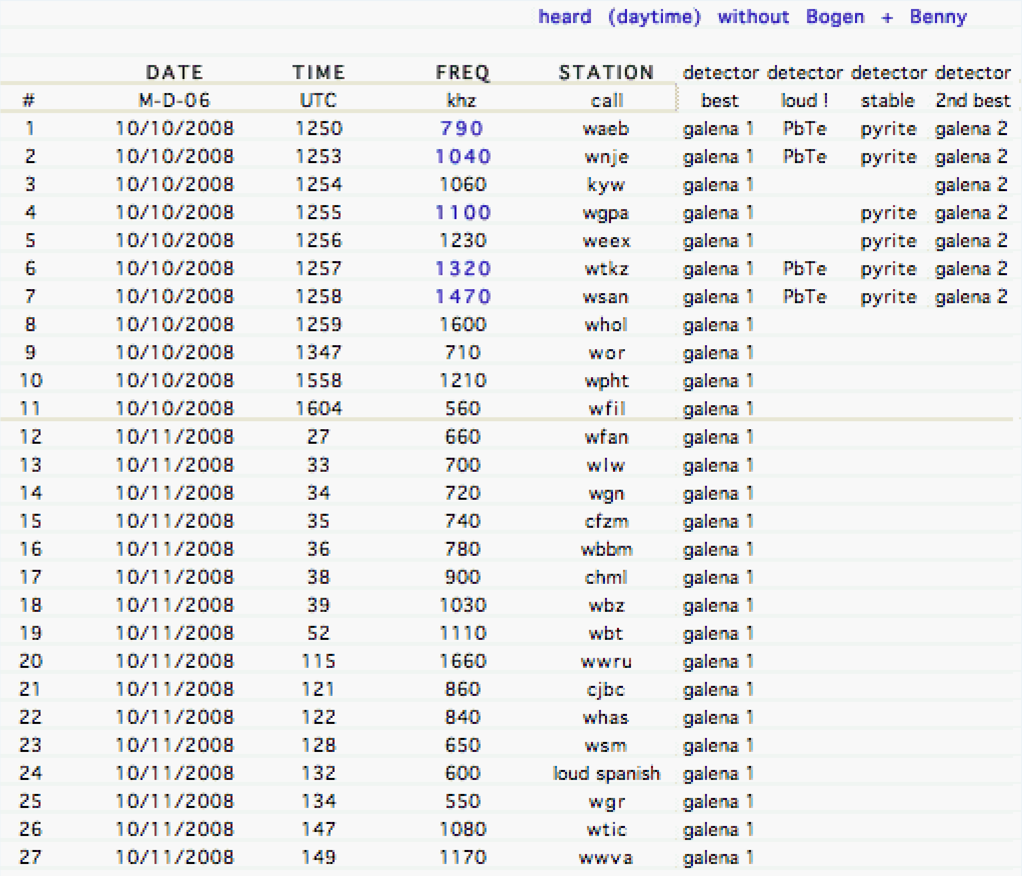

heard (daytime) without Bogen + Benny

	DATE	TIME	FREQ	STATION	detector	detector	detector	detector
#	M-D-06	UTC	khz	call	best	loud !	stable	2nd best
1	10/10/2008	1250	790	waeb	galena 1	PbTe	pyrite	galena 2
2	10/10/2008	1253	1040	wnje	galena 1	PbTe	pyrite	galena 2
3	10/10/2008	1254	1060	kyw	galena 1			galena 2
4	10/10/2008	1255	1100	wgpa	galena 1		pyrite	galena 2
5	10/10/2008	1256	1230	weex	galena 1		pyrite	galena 2
6	10/10/2008	1257	1320	wtkz	galena 1	PbTe	pyrite	galena 2
7	10/10/2008	1258	1470	wsan	galena 1	PbTe	pyrite	galena 2
8	10/10/2008	1259	1600	whol	galena 1			
9	10/10/2008	1347	710	wor	galena 1			
10	10/10/2008	1558	1210	wpht	galena 1			
11	10/10/2008	1604	560	wfil	galena 1			
12	10/11/2008	27	660	wfan	galena 1			
13	10/11/2008	33	700	wlw	galena 1			
14	10/11/2008	34	720	wgn	galena 1			
15	10/11/2008	35	740	cfzm	galena 1			
16	10/11/2008	36	780	wbbm	galena 1			
17	10/11/2008	38	900	chml	galena 1			
18	10/11/2008	39	1030	wbz	galena 1			
19	10/11/2008	52	1110	wbt	galena 1			
20	10/11/2008	115	1660	wwru	galena 1			
21	10/11/2008	121	860	cjbc	galena 1			
22	10/11/2008	122	840	whas	galena 1			
23	10/11/2008	128	650	wsm	galena 1			
24	10/11/2008	132	600	loud spanish	galena 1			
25	10/11/2008	134	550	wgr	galena 1			
26	10/11/2008	147	1080	wtic	galena 1			
27	10/11/2008	149	1170	wwva	galena 1			

Although I didn’t spend a lot of time listening, this simple set will bring-in a lot of stations.

The 2 close-by stations (10 km @ 3.6kW & 11 km @ 5kW) are VERY loud -- just right for a grandkids' first taste of crystal radio. Adding a coupling capacitor to the antenna line tamed the "flame-throwers" a bit and then more stations could be heard.

At night, DX stations were strong and in-the-clear. Easy,pleasant listening.

The detector stands' whisker adjustment was smooth and very stable -- even a grandpa could do it 😉 When a VERY light contact to the crystal was achieved, the set's selectivity improved greatly. A heavier hand produced broader signals. The effect was dramatic to hear.

The addition of a Bogen + Benny makes a worthwhile improvement -- as everyone here knows.

As-is, it's a kid-friendly, easy to operate, well behaved little set.

When a decent (high Rd) diode like an FO-215 replaces the crystal detector, and an antenna tuner coil - like the one at

viewtopic.php?t=1132

is loosely coupled to the set - there's DX all over the dial. Fun stuff.

73, Dan

FtForger

Nice results...

Tue Oct 14, 2008 5:29 pm

Nice results.

My only problem was that I was never able to find a link to the 08 contest for rules/entry info etc...

OErjan

Tue Oct 14, 2008 8:08 pm

here you are FT, sadly somewhat late.
http://crystalradio.us/crystalsprint/index.htm

foolsgolden

Sun Dec 21, 2008 3:13 pm

i like the incredibly long stick fr0m older portables winding my own
enamald wire round em and tere easy to mount my nostalgia compells
me to usehome brue air cores and vario couplers my next expariment
is a tribute to marconi hes the italian stalion god fater of radio.

gusnaz

Sun Dec 21, 2008 3:23 pm

> Not elegant, but it works.

Elegance is in the eye of the beholder!

I believe it to be elegant.

Gus

golfguru

Sun Dec 21, 2008 8:52 pm

> my next expariment
> is a tribute to marconi hes the italian stalion god fater of radio.

Sounds good Goldy, keep us posted.

exray

Sun Dec 21, 2008 11:45 pm

Thats really nice, Dan. I could use a cover like that over my whole shack 😃 I recently built the breadboard rcvr and audio amp that I was showing off here a few months ago. They already look somewhat "vintage" with the tarnished brass.

foolsgolden

Tue Dec 23, 2008 2:50 am

yah dont have camra yet till after holidays but think out side the ciggar box a varyo copler w tuner selector {tikler coyle} wt optional
det/amp/set three peace 😉 ciggar box S.E.T. amp is done with feed back gitar amp neatly hidden in ciggar box witth feed back
controle gain vol for 20 ohm all tgether new sp set.stil working on
other half took 1 ciggar left the rest @ nebors rpm shop for his cellibrations the amps not going to be entrd on contest but a
fools gold piyrite det mated with tuner will be.
feed back by the way on gitar amp. works well with tune n det. sets eavin after trans former and ceramic disc cap. ❓
just good gon rite idea i gess 😳 might turn old amp box to all together new set or rf osc or somthinng. radio bord is new

Loose Coupler Set

In spring of 2016, I decided to build a crystal set with my friends. After searching information on the net, I found a series of introductory videos shown in MICROWAVE1 channel on YouTube very interesting. So, I designed a set based on the loose coupler in the video and added antenna tuning and the detector circuit.

The sliding notch of the antenna coil was not well-made or smooth, but once its position is set, it works fine. The move of secondary coil is also not satisfactory, but I have no complain. You need to reach the slider and the moving coil behind the front panel, so it is a bit annoying. With S4, *DP3T, you can make connection of the coil and capacitor serial or parallel. Rotary switches, S5 and S6 are *SP6T to select taps. Capacitance of the variable capacitors can be added by S2 and S7.

L1 and L2 use 0.5mm polyurethan coated wire and are wound on PVC pipes. R1 should be determined according to the phone(s) you use, or you can make it variable.

Finding the best settings is not so easy but very interesting. You need to adjust L1, L2, C1, and C2 simultaneously, little by little. With an 18 m long wire antenna around my house and a shallow ground earth, maximum voltage at the phone was 50-80 mV when tuned to local stations (10 kW / 10 km). It cannot pick up stronger but distant stations, such as the one with 300 kW at 200 km away, even at night. I may need longer and higher antenna for that.

This is not a super high performance set. It has OK sensitivity and not so great selectivity (no QRM though), but it outputs decent level of sound and certainly is fun to play with. Please note that all the evaluations are very subjective as I have no way to compare other ones.

July 15, 2016

*DP3T: Double pole 3 throw, SP6T: Single pole 6 throw

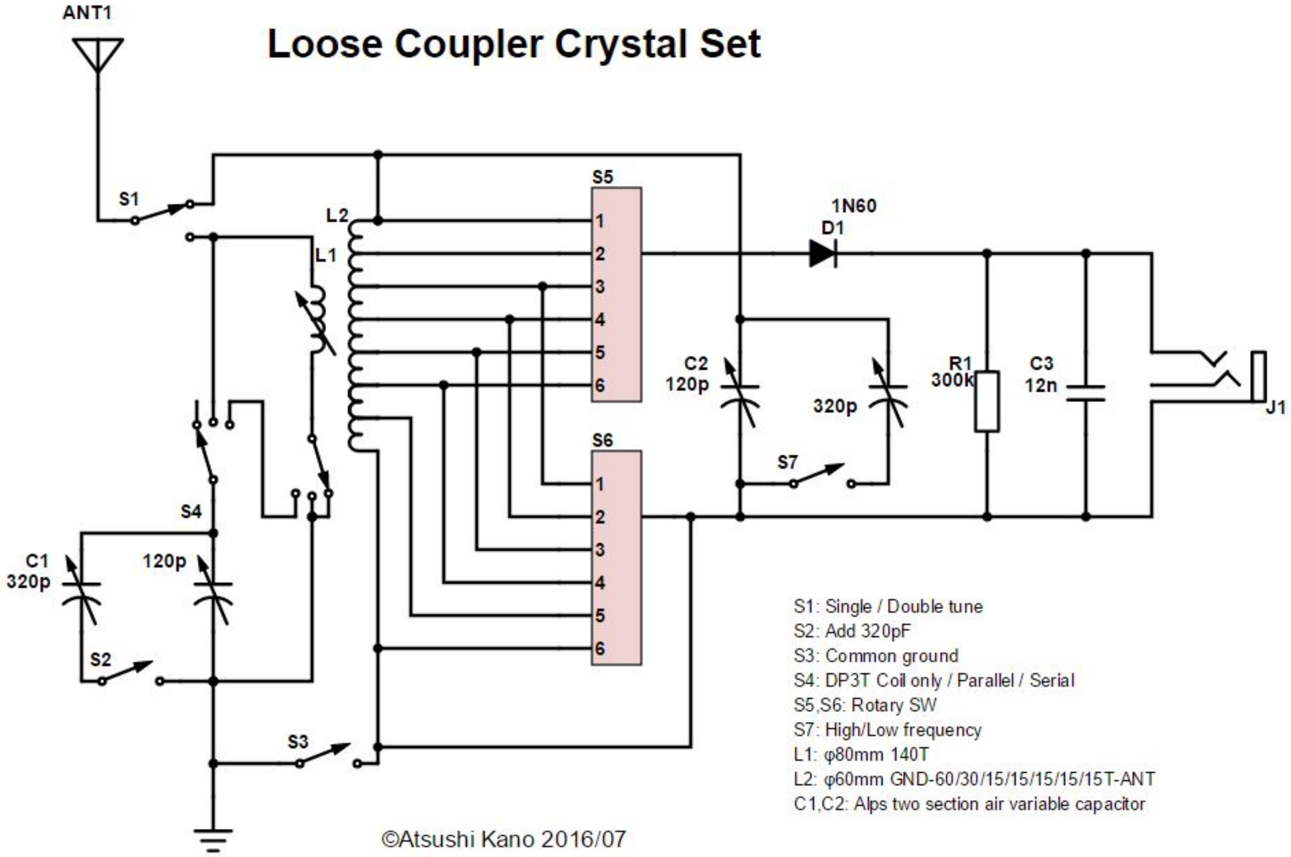

Loose Coupler Crystal Set

ANT1
S1
L2
L1
S5
1
2
3
4
5
6
S6
1
2
3
4
5
6
1N60
D1
C2
120p
320p
R1
300k
C3
12n
J1
S7
S4
C1
320p
120p
S2
S3

S1: Single / Double tune
S2: Add 320pF
S3: Common ground
S4: DP3T Coil only / Parallel / Serial
S5,S6: Rotary SW
S7: High/Low frequency
L1: φ80mm 140T
L2: φ60mm GND-60/30/15/15/15/15/15T-ANT
C1,C2: Alps two section air variable capacitor

©Atsushi Kano 2016/07

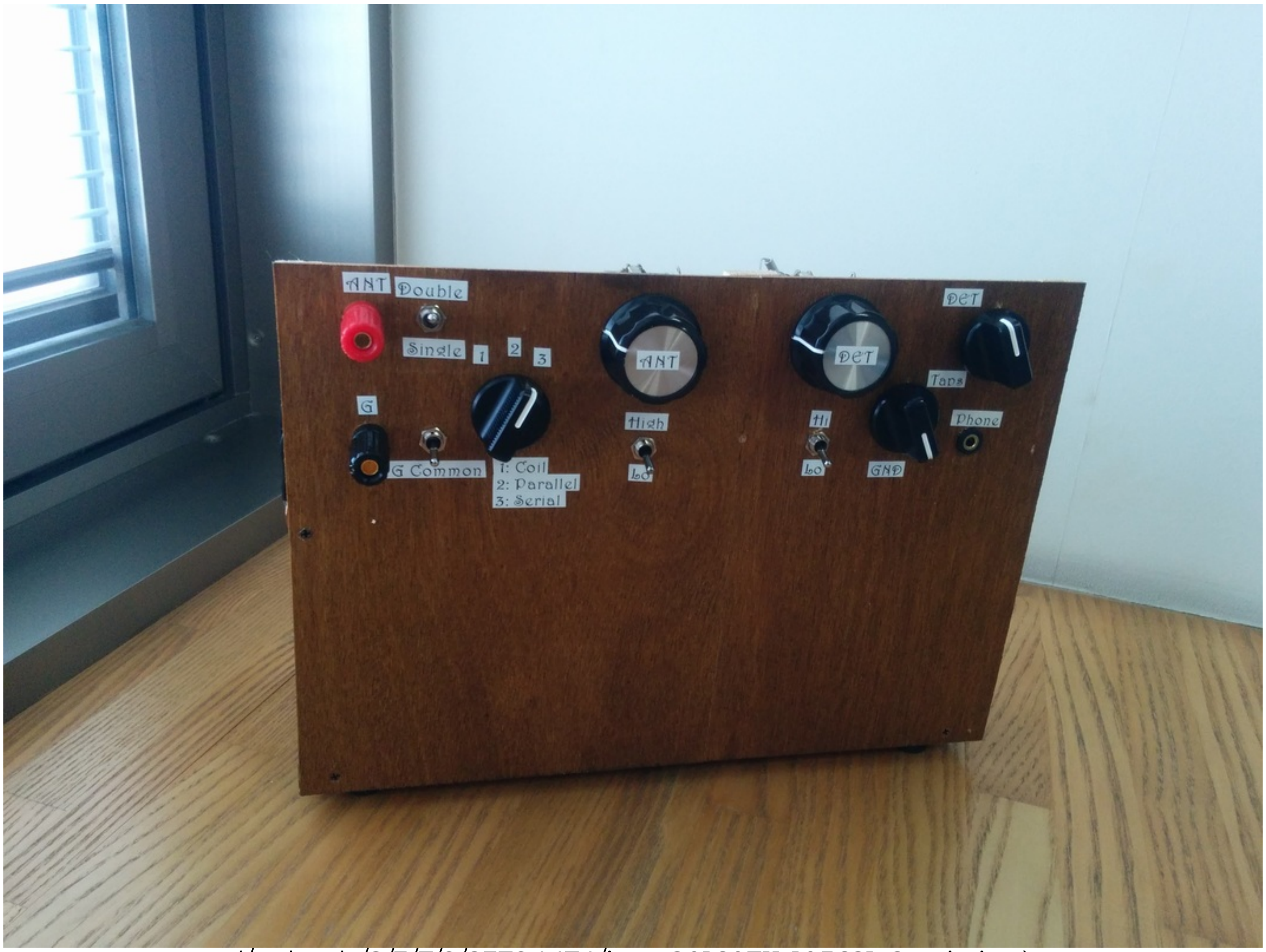
ANT
Double
Single
1
2
3
G
G Common
1: Coil
2: Parallel
3: Serial
ANT
High
Lo
DET
Hi
Lo
DET
Taps
GND
Phone

(/uploads/8/3/7/9/83794474/img-20160711-105621_2_orig.jpg)

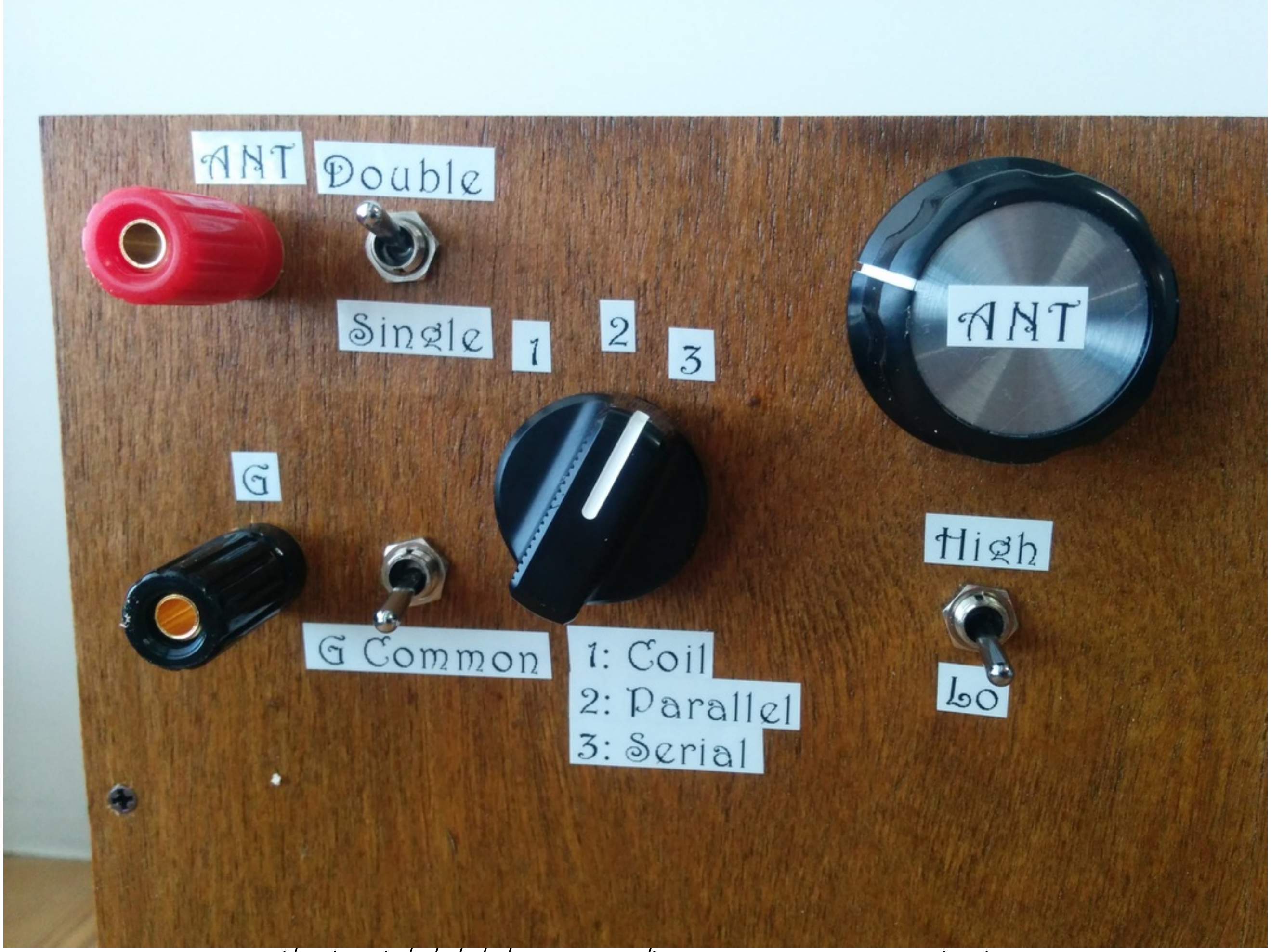
ANT
Double
Single
1
2
3
ANT
G
High
G Common
1: Coil
2: Parallel
3: Serial
Lo

(/uploads/8/3/7/9/83794474/img_20160711_105732.jpg)

rexresearch.com

Louis MAICHE

Wireless Telluric Telephone

NEW!
LinkRunner G2
SMART NETWORK TESTER
LEARN MORE
NETSCOUT
Guardians of the Connected World

Sure, it looks silly now -- but wait until an EMP destroys the modern grid -- then you'll want/need this proven simple, inexpensive alt-tech earth-radio -- directional, low energy, & untappable -- and anyone can build it...

Jerry Vassilatos : Vril Compendium

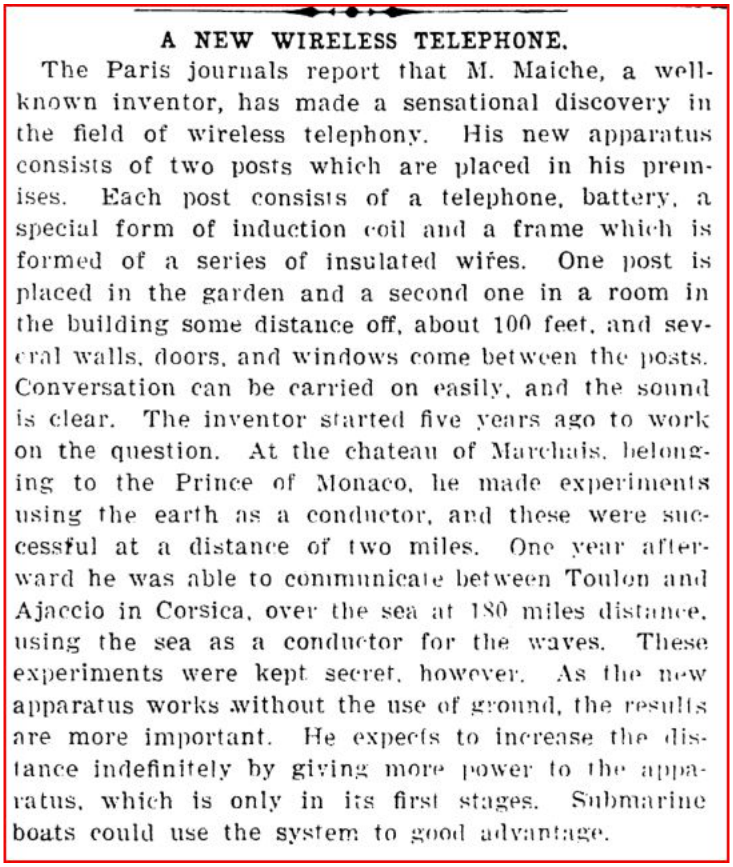

A NEW WIRELESS TELEPHONE.

The Paris journals report that M. Maiche, a well-known inventor, has made a sensational discovery in the field of wireless telephony. His new apparatus consists of two posts which are placed in his premises. Each post consists of a telephone, battery, a special form of induction coil and a frame which is formed of a series of insulated wires. One post is placed in the garden and a second one in a room in the building some distance off, about 100 feet, and several walls, doors, and windows come between the posts. Conversation can be carried on easily, and the sound is clear. The inventor started five years ago to work on the question. At the chateau of Marchais, belonging to the Prince of Monaco, he made experiments using the earth as a conductor, and these were successful at a distance of two miles. One year afterward he was able to communicate between Toulon and Ajaccio in Corsica, over the sea at 180 miles distance, using the sea as a conductor for the waves. These experiments were kept secret, however. As the new apparatus works without the use of ground, the results are more important. He expects to increase the distance indefinitely by giving more power to the apparatus, which is only in its first stages. Submarine boats could use the system to good advantage.

---

https://books.google.com/books?id=de5OAAAAYAAJ&pg=PA477&lpg=PA477&dq=Louis+MAICHE&source=bl&ots=0MwfiZciry&sig=ePs25QJg3djFbnx_ln0ivOIyIWg&hl=en&sa=X&v

The Christian Work and the Evangelist, Volume 82

Wireless Telephone of Louis Maiche, the French Scientist

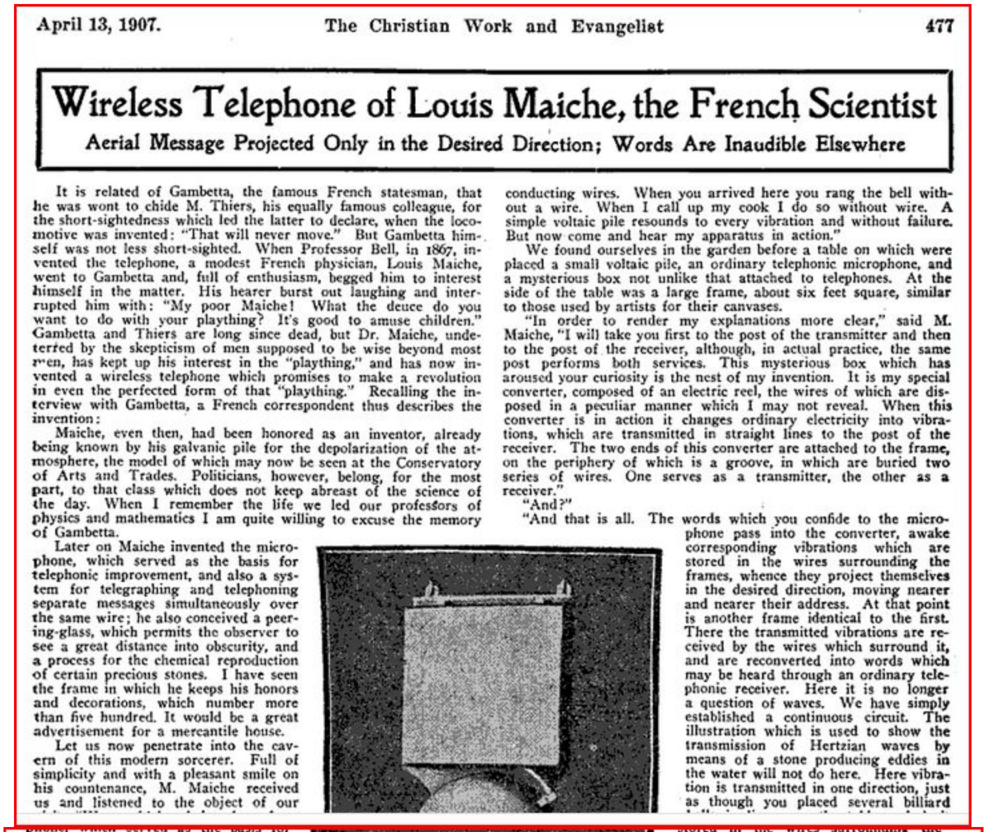

Wireless Telephone of Louis Maiche, the French Scientist

Aerial Message Projected Only in the Desired Direction; Words Are Inaudible Elsewhere

It is related of Gambetta, the famous French statesman, that he was wont to chide M. Thiers, his equally famous colleague, for the short-sightedness which led the latter to declare, when the locomotive was invented: "That will never move." But Gambetta himself was not less short-sighted. When Professor Bell, in 1867, invented the telephone, a modest French physician, Louis Maiche, went to Gambetta and, full of enthusiasm, begged him to interest himself in the matter. His hearer burst out laughing and interrupted him with: "My poor Maiche! What the deuce do you want to do with your plaything? It's good to amuse children." Gambetta and Thiers are long since dead, but Dr. Maiche, undeterred by the skepticism of men supposed to be wise beyond most men, has kept up his interest in the "plaything," and has now invented a wireless telephone which promises to make a revolution in even the perfected form of that "plaything." Recalling the interview with Gambetta, a French correspondent thus describes the invention:

Maiche, even then, had been honored as an inventor, already being known by his galvanic pile for the depolarization of the atmosphere, the model of which may now be seen at the Conservatory of Arts and Trades. Politicians, however, belong, for the most part, to that class which does not keep abreast of the science of the day. When I remember the life we led our professors of physics and mathematics I am quite willing to excuse the memory of Gambetta.

Later on Maiche invented the microphone, which served as the basis for telephonic improvement, and also a system for telegraphing and telephoning separate messages simultaneously over the same wire; he also conceived a peering-glass, which permits the observer to see a great distance into obscurity, and a process for the chemical reproduction of certain precious stones. I have seen the frame in which he keeps his honors and decorations, which number more than five hundred. It would be a great advertisement for a mercantile house.

Let us now penetrate into the cavern of this modern sorcerer. Full of simplicity and with a pleasant smile on his countenance, M. Maiche received us and listened to the object of our

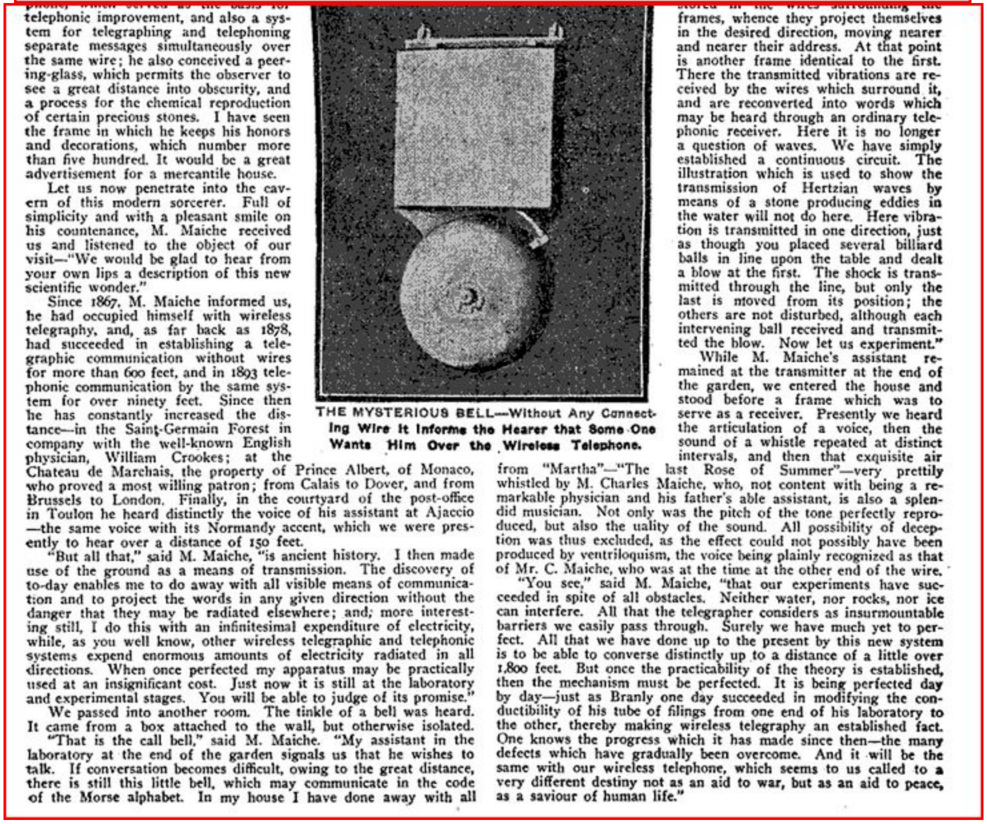

telephonic improvement, and also a system for telegraphing and telephoning separate messages simultaneously over the same wire; he also conceived a peering-glass, which permits the observer to see a great distance into obscurity, and a process for the chemical reproduction of certain precious stones. I have seen the frame in which he keeps his honors and decorations, which number more than five hundred. It would be a great advertisement for a mercantile house.

Let us now penetrate into the cavern of this modern sorcerer. Full of simplicity and with a pleasant smile on his countenance, M. Maiche received us and listened to the object of our

visit—"We would be glad to hear from your own lips a description of this new scientific wonder."

Since 1867, M. Maiche informed us, he had occupied himself with wireless telegraphy, and, as far back as 1878, had succeeded in establishing a telegraphic communication without wires for more than 600 feet, and in 1893 telephonic communication by the same system for over ninety feet. Since then he has constantly increased the distance—in the Saint-Germain Forest in company with the well-known English physician, William Crookes; at the Chateau de Marchais, the property of Prince Albert, of Monaco, who proved a most willing patron; from Calais to Dover, and from Brussels to London. Finally, in the courtyard of the post-office in Toulon he heard distinctly the voice of his assistant at Ajaccio—the same voice with its Normandy accent, which we were presently to hear over a distance of 150 feet.

"But all that," said M. Maiche, "is ancient history. I then made use of the ground as a means of transmission. The discovery of to-day enables me to do away with all visible means of communication and to project the words in any given direction without the danger that they may be radiated elsewhere; and, more interesting still, I do this with an infinitesimal expenditure of electricity, while, as you well know, other wireless telegraphic and telephonic systems expend enormous amounts of electricity radiated in all directions. When once perfected my apparatus may be practically used at an insignificant cost. Just now it is still at the laboratory and experimental stages. You will be able to judge of its promise."

We passed into another room. The tinkle of a bell was heard. It came from a box attached to the wall, but otherwise isolated.

"That is the call bell," said M. Maiche. "My assistant in the laboratory at the end of the garden signals us that he wishes to talk. If conversation becomes difficult, owing to the great distance, there is still this little bell, which may communicate in the code of the Morse alphabet. In my house I have done away with all conducting wires. When you arrived here you rang the bell without a wire. When I call up my cook I do so without wire. A simple voltaic pile resounds to every vibration and without failure. But now come and hear my apparatus in action."

We found ourselves in the garden before a table on which were placed a small voltaic pile, an ordinary telephonic microphone, and a mysterious box not unlike that attached to telephones. At the side of the table was a large frame, about six feet square, similar to those used by artists for their canvases.

"In order to render my explanations more clear," said M. Maiche, "I will take you first to the post of the transmitter and then to the post of the receiver, although, in actual practice, the same post performs both services. This mysterious box which has aroused your curiosity is the nest of my invention. It is my special converter, composed of an electric reel, the wires of which are disposed in a peculiar manner which I may not reveal. When this converter is in action it changes ordinary electricity into vibrations, which are transmitted in straight lines to the post of the receiver. The two ends of this converter are attached to the frame, on the periphery of which is a groove, in which are buried two series of wires. One serves as a transmitter, the other as a receiver."

"And?"

"And that is all. The words which you confide to the microphone pass into the converter, awake corresponding vibrations which are stored in the wires surrounding the frames, whence they project themselves in the desired direction, moving nearer and nearer their address. At that point is another frame identical to the first. There the transmitted vibrations are received by the wires which surround it, and are reconverted into words which may be heard through an ordinary telephonic receiver. Here it is no longer a question of waves. We have simply established a continuous circuit. The illustration which is used to show the transmission of Hertzian waves by means of a stone producing eddies in the water will not do here. Here vibration is transmitted in one direction, just as though you placed several billiard

frames, whence they project themselves in the desired direction, moving nearer and nearer their address. At that point is another frame identical to the first. There the transmitted vibrations are received by the wires which surround it, and are reconverted into words which may be heard through an ordinary telephonic receiver. Here it is no longer a question of waves. We have simply established a continuous circuit. The illustration which is used to show the transmission of Hertzian waves by means of a stone producing eddies in the water will not do here. Here vibration is transmitted in one direction, just as though you placed several billiard

balls in line upon the table and dealt a blow at the first. The shock is transmitted through the line, but only the last is moved from its position; the others are not disturbed, although each intervening ball received and transmitted the blow. Now let us experiment."

While M. Maiche's assistant remained at the transmitter at the end of the garden, we entered the house and stood before a frame which was to serve as a receiver. Presently we heard the articulation of a voice, then the sound of a whistle repeated at distinct intervals, and then that exquisite air from "Martha"—"The last Rose of Summer"—very prettily whistled by M. Charles Maiche, who, not content with being a remarkable physician and his father's able assistant, is also a splendid musician. Not only was the pitch of the tone perfectly reproduced, but also the uality of the sound. All possibility of deception was thus excluded, as the effect could not possibly have been produced by ventriloquism, the voice being plainly recognized as that of Mr. C. Maiche, who was at the time at the other end of the wire.

"You see," said M. Maiche, "that our experiments have succeeded in spite of all obstacles. Neither water, nor rocks, nor ice can interfere. All that the telegrapher considers as insurmountable barriers we easily pass through. Surely we have much yet to perfect. All that we have done up to the present by this new system is to be able to converse distinctly up to a distance of a little over 1,800 feet. But once the practicability of the theory is established, then the mechanism must be perfected. It is being perfected day by day—just as Branly one day succeeded in modifying the conductibility of his tube of filings from one end of his laboratory to the other, thereby making wireless telegraphy an established fact. One knows the progress which it has made since then—the many defects which have gradually been overcome. And it will be the same with our wireless telephone, which seems to us called to a very different destiny not as an aid to war, but as an aid to peace, as a saviour of human life."

THE MYSTERIOUS BELL—Without Any Connecting Wire It Informs the Hearer that Some One Wants Him Over the Wireless Telephone.

---

https://books.google.com/books?id=euDmAAAAMAAJ&pg=PA79&lpg=PA79&dq=Louis+MAICHE&source=bl&ots=1k4NAB8mnd&sig=NTExYinKMaUfIbhxtExc0kaKA-0&hl=en&sa=X&ved=0ahUKEwjgp8uehoHQAhVj0FQKHeFyDlk4ChDoAQgbMAA#v=onepage&q&f=false

Telephony, Volumes 3-4

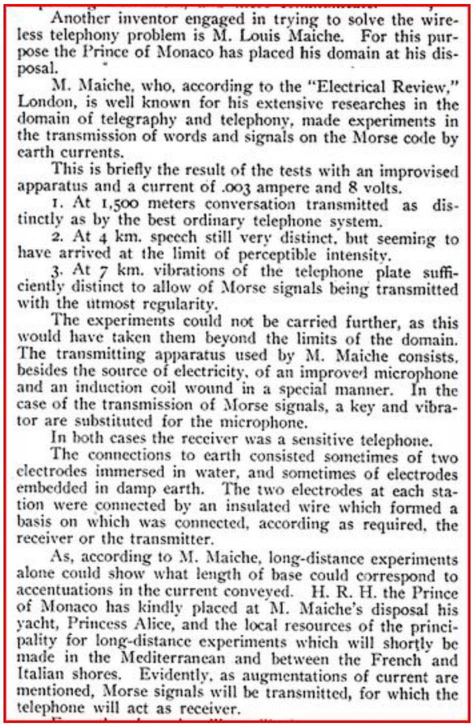

Another inventor engaged in trying to solve the wireless telephony problem is M. Louis Maiche. For this purpose the Prince of Monaco has placed his domain at his disposal.

M. Maiche, who, according to the "Electrical Review," London, is well known for his extensive researches in the domain of telegraphy and telephony, made experiments in the transmission of words and signals on the Morse code by earth currents.

This is briefly the result of the tests with an improvised apparatus and a current of .003 ampere and 8 volts.

1. At 1,500 meters conversation transmitted as distinctly as by the best ordinary telephone system.

2. At 4 km. speech still very distinct, but seeming to have arrived at the limit of perceptible intensity.

3. At 7 km. vibrations of the telephone plate sufficiently distinct to allow of Morse signals being transmitted with the utmost regularity.

The experiments could not be carried further, as this would have taken them beyond the limits of the domain. The transmitting apparatus used by M. Maiche consists, besides the source of electricity, of an improved microphone and an induction coil wound in a special manner. In the case of the transmission of Morse signals, a key and vibrator are substituted for the microphone.

In both cases the receiver was a sensitive telephone.

The connections to earth consisted sometimes of two electrodes immersed in water, and sometimes of electrodes embedded in damp earth. The two electrodes at each station were connected by an insulated wire which formed a basis on which was connected, according as required, the receiver or the transmitter.

As, according to M. Maiche, long-distance experiments alone could show what length of base could correspond to accentuations in the current conveyed. H. R. H. the Prince of Monaco has kindly placed at M. Maiche's disposal his yacht, Princess Alice, and the local resources of the principality for long-distance experiments which will shortly be made in the Mediterranean and between the French and Italian shores. Evidently, as augmentations of current are mentioned, Morse signals will be transmitted, for which the telephone will act as receiver.

---

http://www.telephones-anciens.org/l-maiche-electrophone-a-pupitre/

L.Maiche électrophone à pupitre

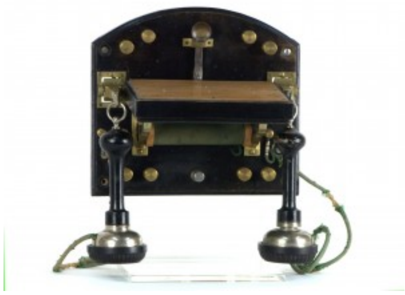

Appareil dit Electrophone de Maiche: nom donné par Louis Maiche a ses appareils téléphoniques. Celui-ci est équipé d'une paire d'écouteurs a manches Lensczewski et d'un microphone type Ader. Le modèle en photo doit être antérieur à 1890

---

http://www.lookandlearn.com/history-images/M240118/Maiches-wireless-telephone-system-1906?img=15874

Maiche's wireless telephone system, 1906

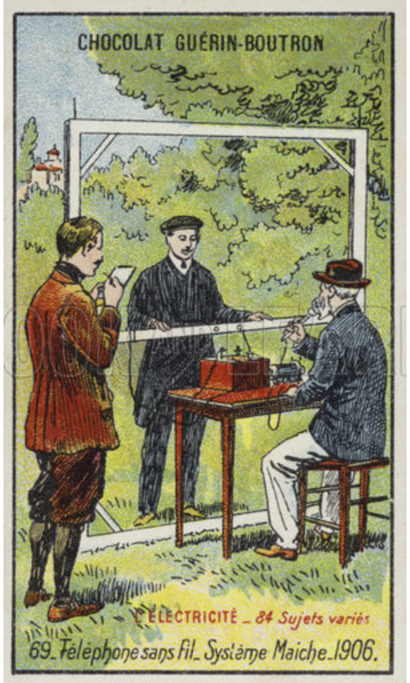
CHOCOLAT GUÉRIN-BOUTRON
L'ÉLECTRICITÉ _ 84 Sujets variés
69. Téléphone sans Fil_ Système Maiche_1906.

---

http://thethingsienjoy.blogspot.com/2013_04_21_archive.html

The world's first "mobile" phone (in 1906)

French inventor Louis Maiche and his wireless phone, arguably also the world's first mobile phone.
Image published in the Swedish weekly Hvar 8 Dag in September 1906 (Click to enlarge)

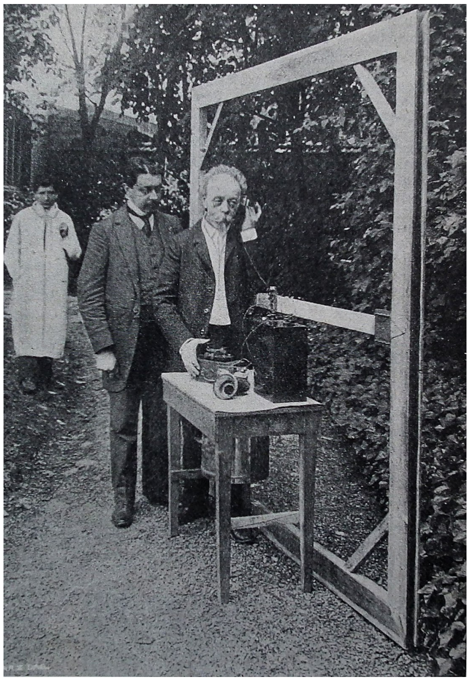

"M. Louis Maiche is the inventor of the wireless telephone and, like many scientists, is not so famous as he ought to be".

That is how an article in the Auckland Star on March 6, 1907, introduced the French inventor Louis Maiche, who also could be called the inventor of a mobile phone.

In the article M. Maiche describes how it all began:

"My first trials took place in 1867, but they were concerned more with wireless telegraphy, and it was not until 1893 that I exchanged a conversation through a wireless telephone at a distance of 30 yards."

Two months earlier, the New York Times had also published an article on M. Maiche and his wireless telephone invention. In the article the inventor refers to his successful experiments with telegraphy and then describes his work with the wireless telephone:

"I then made use of the ground as a means of transmission. The discovery of today enables me to do away with all visible means of communication and to project the words in any given direction without the danger that they may be radiated elsewhere; and, more interesting still, I do this with an infinitesimal expenditure of electricity, while, as you well know, other wireless telegraphic and telephone systems expend enormous amounts of electricity radiated in all directions."

"When once perfected my apparatus may be practically used at an insignificant cost. Just now it is still at the laboratory and experimental stages."

"You see", said M. Maiche,"that our experiments have succeeded in spite of all obstacles. Neither the water, nor rocks, nor ice can interfere. All that the telegrapher considers as insurmountable barriers we easily pass through. Surely we have much to perfect! All that we have done up to the present is to be able to converse distinctly up to a distance of a little over 1,800 feet. But once the practicality of the theory is established, then the mechanism must be perfected."

"It is being perfected day by day - just as Branly one day succeeded in modifying the conductibility of his tube fillings from one end of his laboratory to the other, thereby making wireless telegraphy an established fact. And it will the same with our wireless telephone,which seems to us called to a very different destiny, not as aid to war, but as an aid to peace, as a savior of human life."

In the Auckland Star article M. Maiche gives some examples of the future use of his wireless telephone:

"Imagine two steamers steering through a fog. With my little apparatus the captain can tell the direction the other ship is taking. In case of an accident to a submarine my apparatus would enable the crew in danger to communicate with the convoying tug without fearing the breaking of a line as in the case of a telephone buoy. Miners entombed after a disaster like that at Courrieres could communicate with the rescue party. Two army corps making a night attack could keep in constant touch without risk of interruption. The ordinary citizen could have it in his drawing room, and would no longer be dependent upon the vagaries of the telephone exchange."

Both articles also include some technical descriptions of the use of M. Maiche's new wireless phone. It is possible that the French inventor's theories were not as successful as he himself thought. Otherwise it is difficult to understand why this page (in French) is the only one with information about him that I have been able to locate. Still, his concept of wireless (and mobile) telephony should merit him a place in the hall of fame of modern communication technology .

Patents

Nouveau mode de transmission et de réception sans fil des signaux électriques
FR335990

1904-02-22

The object of the present invention is a new method of transmitting and receiving wireless electrical signals. The essential means on which is based the system is the production transmits position and uses the position receiver electricity manifested in a new form which may be referred to as "electrostatic vibration" and is obtained as will be explained further by modification of a spark inductor. One of distinctive characters of this new manifestation of electricity is that it can be transmitted in space without the aid antennas and ground at all distances, and, in the same form as sound, light and heat, that is to say without poles.

The inductor being communicated c with a power source, a battery, for example, and a special switch which will be described below, each interruption is to be born in the armature current giving a spark, and it is the latter that is transform an "electrostatic vibration." The armature out of the coil wire after the last winding outside, is for this purpose connected to a large metal surface which can be exposed to air. This surface may be constituted either by c a hollow or solid cylinder or by one or more sheets of iron, copper or any other metal usual more or less thick. The surface is preferably rounded at the corners or packed with a border all around to remove spikes, as in the conductive electrostatic machines. It should be about one square meter per watt of the current flowing in the inductor. The other end of the armature is connected to an adjusting screw of which the tip k of platinum or gold is next to a button, also in a gold or platinum, carried on the large surface.

FIGS. 1 and 2 of the accompanying drawings show two provisions which may be 5 h adopted for the preparation of the transmitter station.

FIGS. 3 and k show two provisions and receiver.

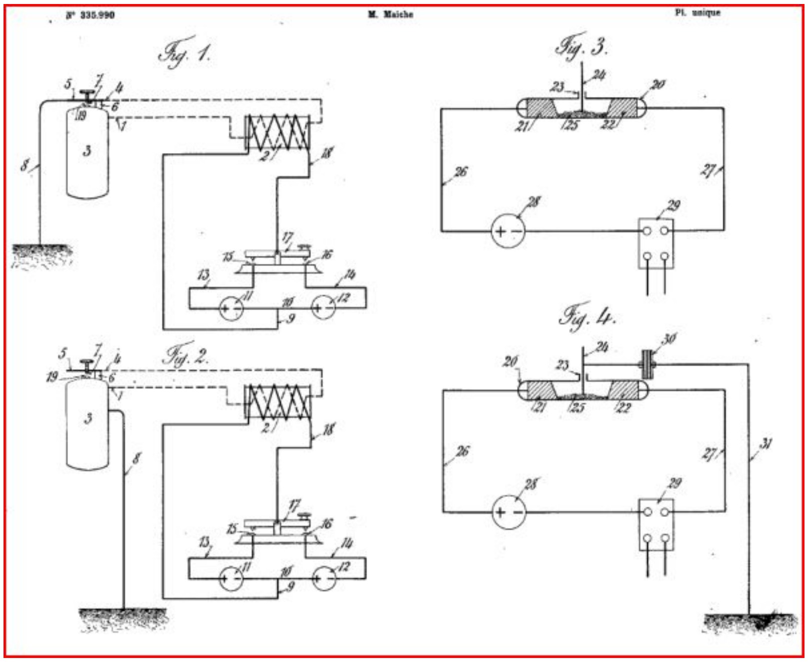

One end of the armature wire of the coil -j is connected to the cylinder 3 rounded at its ends, and the end k of the wire is connected to plate 5 mounted on the insulating support 6 which is fixed to the cylinder 3 . The plate 5 is traversed by a setscrew 55 provided beyond tip 7 made of gold or platinum. The end of the induced wire k is that by which begins rolling on a bobbin. Wire 8 communicates with the earth or the plate 5 (Fig. I) is the cylinder 3 (Fig. A).

Improvements in and relating to Electric Transformers.
GB190727157

1908-10-29

Comprises a large number of windings wound in the form of a cable a on a coreless frame b, and led out to terminals 1 ... 10 and, adapted to be connected in quantity or tension by means of a commutator so that three or more independent circuits may be made.

COMPLETE SPECIFICATION.

The present invention relates to electric transformers of the class described by Specifications Nos. N;821901 (Amberg), z1854 (Swinburne), and 4/I83 (Clerc) and it particularly concerns the arrangement of the windings.

The invention is represented in the accompanying drawings;-

Figs. 1 and 2, show a section and a front view respectively of the improved transformer.

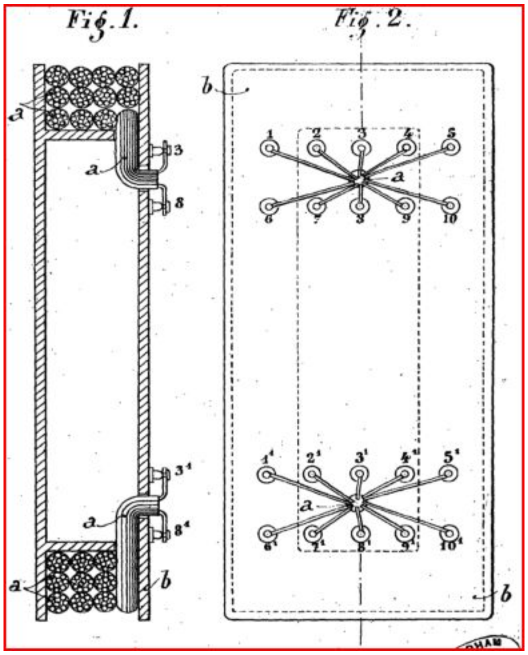

The transformer which has no magnetic core, is merely constituted by a convenient number of insulated wires, 10 for example, which are united to form a cable a which is wound around a frame b. The ends of the cable project through two holes of the frame and the ends of each wire are joined to terminals 1 to 10 and 11 to 101 respectively. In the drawings are shown 20 terminals, the first ten of which correspond to the beginnings of the 10 wires, and the other ten correspond to the ends of the same wires.

Thus, a winding of a certain number of identical wires is obtained which are all wound parallel and simultaneously and which may be connected by means of a commutator so as to form groups to 'be varied according to requirement; so that there are thus constituted three or more circuits which are entirely independent the one of the other and the length of which can be varied at will.

By conveniently joining the free ends of some of the wires, a primary winding of weak resistance is obtained with the wires of same joined for quantity, while one or more of the secondary windings are joined for tension, thus providing any required resistance.

The construction of the transformer by means of a great number of separate wires offers numerous advantages amongst which are the facility of connecting the separate wires at will to constitute the circuits; the arrangement enabling a supplementary circuit to be branched off at any point of the said circuits; and the reciprocal independence of these circuits with large number.

Improvements in and relating to Electric Transformers.
GB190721535

1908-05-28

Transformers for telegraphy and wireless telegraphy are made without a magnetic core, and of little depth but of large diameter. They have the primary winding e enclosed within the secondary winding c, g, but separated therefrom by insulation d, f. The windings may be formed in sections, and coupled up in any order by means of a commutator.

COMPLETE SPECIFICATION.

This invention relates to electric transformers the construction of which is improved so that currents of induction of very short periodicity are obtained.

The currents, the duration of which is extremely short, possess special properties which are particularly useful when applied in telegraphy and in wireless telephony.

Transformers without magnetic core and with the secondary winding wound around the primary, or with the primary winding enclosed within insulating material round wh,ich is wound the secondary winding so that the whole forms a cable, have been provided, and my invention is limited to the form and construction of the transformer as hereinafter set forth.

My improved electric transformer contains its primary wire within a cable of the secondary wires in the form of a single coil of little thickness but of very large dia,meter, said coil or cable being applied to a ring-shaped transformer body.

The accompanying drawing shows my improved electric transformer.

Fig. 1 is a section on line A-B of Fig. 2 and

Fig. 2 is a plan of the ring-shaped transformer.

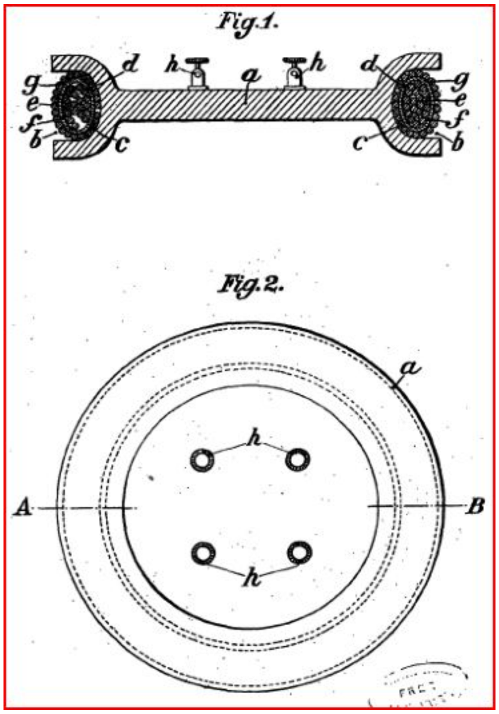

As will be seen from the drawings, the transformer a is of comparatively large diameter which may be several meters. In the drawing the transformer is represented of circular cross-section, it could however be of square or other cross-section. This transformer is very short because all the wires of the winding are located on its periphery in an appropriate hollow or groove b. For manufacturing the trans-former I place around the rim of this kind of wheel or dial: a, a first layer of secondary wire which is covered by a suitable insulator d and upon this insulator the primary wire e is wound which is covered by another layer of insulating material f. The two insulating layers d and f are connected in such a manner that the primary wires e are well insulated, and the coil is then finished by applying an outer winding of secondary wire g, the latter surrounding the primary in such a manner as to form a cable. In this manner a ring of large diameter is obtained which is composed of the secondary wires c, g which enclose a core of suitably insulated and wholly surrounded primary wires e and which ring is applied to a ring-shaped transformer-body having a peripheral groove of suitable cross-section.

The apparatus comprises four terminals h, two of which serve for fixing the ends of the secondary wires, the other two serving for fixing the ends of the primary winding.
There could be arranged several independent cables or rings in each of which the secondary wires entirely surround the primary wires and in such case each of the terminals would be replaced by a multiple commutator permitting to join up the different secondary wires as well as the different primary wires either for tension or for quantity according to the effect to be obtained.

Improvements in connection with Wireless Telegraphy and Telephony.
GB190721252

1907-11-21

The transmitting and receiving circuits at one station for an inductive system of telephony or telegraphy are arranged as shown. A coil a, consisting of a few turns of thick wire, is used alone in transmitting, but is connected in series with a coil b, consisting of a large number of turns of fine wire, for receiving; the circuit of the coil b is broken automatically, during transmission, by means of a Morse key c. The circuit of the coil a includes a battery n and a microphone o for telephony, or an interrupter for telegraphy, and the coil b is connected with a telephone h.

There are known systems for telegraphy and telephony without connecting wires which utilise, instead of Hertzian waves, the phenomena of distant induction by means of electric vibrations of very short duration, but great amplitude. With such systems, one and the same winding is used fur sending and for receiving; for sending it is placed in circuit with a source of electricity (battery, accumulator, etc.) and with a microphone or an interrupter destined to produce the vibrations in question. The same winding is placed in circuit with a .receiving telephone .for receiving the currents which arc induced at a. distance from the sending winding of the other station.

It is obvious that this single coil cannot equally well serve for the sending as for the receiving. To transmit well it ought to be made from thick and short wire, whilst for receiving well it ought to be of long and thin wire.

It would evidently be the simplest to provide each station with two different coils of which one serves as transmitter and the other as receiver, but lucre is another difficulty to be considered; for receiving, the long and thin coil has to be closed by the telephone and in that condition it would absorb nearly all the energy which ought to be sent out by the transmitter.

The present invention rotates to a special method of mounting the parts constituting a station in order to overcome the above mentioned difficulty; besides, the special method of mounting permits to join the two coils one behind the other for receiving whereby the receiving power of the station is considerably increased.

The accompanying drawing represents diagrammatically a station according to the present Invention.

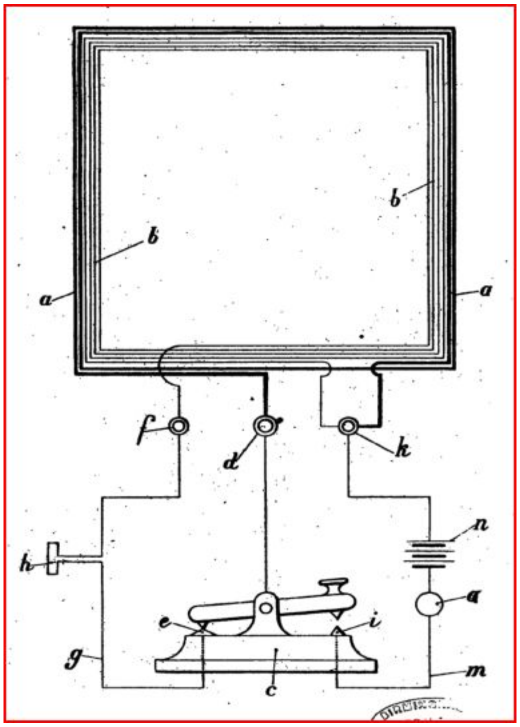

a is the inductor wire which serves fur transmitting; this wire is thick and relatively short. b is the induced winding for receiving the induced vibrations from the inductor winding of the transmitting station; this winding consists of thin wire of considerable length. c is a. Morse-manipulator the lever of which is connected through the contact let with one end of wire a, its connector e is through the contact f connected with one of the ends of wire; the wire g which connects the contact f with wire b comprises the telephonic receiver h ;the other connector i of the Morse apparatus is connected through terminal with the free ends of the two wires a and b which are joined at this point. The connecting wire m which serves for this purpose, comprises the battery n and the microphone o.

It is easily' understood that in the inoperative position of the Morse key (as represented in the drawings) the two coils and b form one circuit closed on the telephone h. This position is the receiving position. For transmitting, the key is depressed and brought in contact with the connector whereby the coil b is opened whilst the coil a is closed on the microphone and the battery.

It is to be understood, that the station described serves for the transmission of sounds. For the transmission of telegraphic signals, the microphone o is replaced by a suitable interruptor-vibrator and the apparatus is operated through the Morse key as usual.

Improvements in and relating to Inductive Telegraphy and Telephony.
GB190712529

1908-03-12

In a system of telegraphy or telephony by magnetic induction, the coils are placed vertically and the same coil is used for transmitting and receiving. The coil b is wound on a frame a of any suitable shape, which may be lined with iron. One terminal of the coil b is connected to a key f, one contact of which is connected through a microphone p and battery o to the other terminal d of the coil. A telephone m is connected across the terminals of the coil through another contact of the key f, and its circuit is normally completed for receiving telephonic or telegraphic messages. Telegraphy is effected by manipulating the kev f to complete the circuit of battery o ; to telephone, the key is depressed and the microphone p is utilized. A high-frequency alternator may be used instead of the battery o, and a screen composed of sheets of paper soaked in paraffin, varnish, or glue may be placed behind the frame.

The present invention relates to improvements in that class of magnetic induction telegraphy and telephony in which transmission is effected between similar coils of wire at the receiving and transmitting stations. The invention is characterized by the fact that senders and receivers.are reversible, the member for transmitting the waves being the, same as the member which receives the waves.

In the drawing, a diagram of a station according to the present system is represented.

Fig. 1.

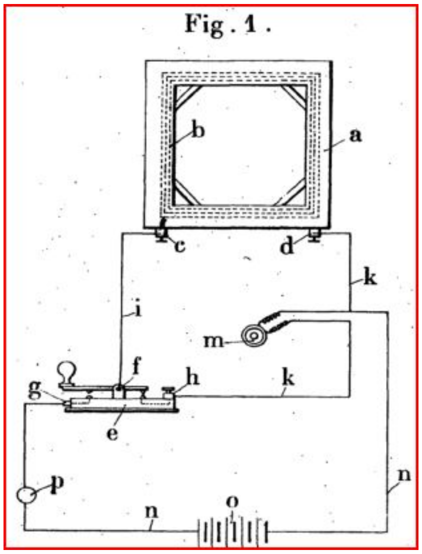

The transmitter and receiver for the electric waves consists of a. wooden frame (a) of fairly large dimensions; for example 4 meters square for a distance of 20 kilometers. In the drawing the frame is shown as being square but it may be of any other geometric shape having as center a.rectangle a lozenge, a circle, etc. The wall of the frame is provided with a groove (b) in which an insulated conducting wire is placed which is connected with the terminals (c) and (d). This wire is wound a. varying number of times around the frame so that it is of sufficient length according to the power of transmission required of the apparatus.

The frame (a) must be of sufficiently large dimensions to prevent the currents generated in opposite directions in the sides of the frame from tending to neutralise each other.

Each station comprises besides a Morse manipulator (e), the commutator of which is connected with one of the terminals (c) of frame (a) through the wire, its terminals (r) and (h) being connected with the other terminal (d) of frame (a).

An ordinary telephonic receiver (m) is inserted in, the circuit (k) which leads from the terminal (h.) of the manipulator to the frame, and in the other circuit (n), connecting the terminal (g) of the manipulator with. the frame, a battery (o) is inserted near the frame and , microphone (p) is inserted near the manipulator.

When the manipulator (c) is at rest (position shown in the drawing) the station is ready to receive the transmissions, the electric circuit (a, c, i, f, m, k, d, a) enclosing the frame and the telephone -being closed. One has only to hold the telephone to the ear to perceive all the currents which a.re induced in the frame (a).

The telephone (m) is put out of circuit when the commutator (f) is pressed down, whilst t.he circuit will be closed upon the frame (a.) and battery (o). The station has thus become 11 sender, the battery sending variable currents into the frame according to the nature of the transmission to be effected. To telegraph according to the Morse alphabet one manipulates the commutator (f) in the usual manner, the telephone (m) of the other station permitting to distinctly perceive the dots and dashes. The operator simply gives one contact for the dot and two contacts following in quick succession for a dash. In this case the microphone (p) is superfluous but it does not disturb. When however one wants to telephone, the contact (g) has to be kept closed and one speaks in front of or into the microphone (p); the telephone (in.) of the other station will reproduce the sounds.

The source of electricity (o) can be a battery or an electro-magnetic generator giving alternating currents of high frequency. The currents which are thus sent into the frame (a) of the sender station have no appreciable effect upon the telegraphic receivers which could be provided at this station; they even do not influence the magnet needle owing to their **illst,nntfmeousness.**

To establish the communication between two stations the two frames (a) have to be placed in parallel planes, perpendicular to the line which would connect their centers, which can be effected through various means. The action (if the one of the frames on the other is effected through any obstacles whatsoever, no earth connection or air connection being necessary.

The reciprocal action of the frames can be increased through lining the bottom of groove (b) for its entire length with soft sheet iron, the lining being continuous or interrupted. The frame will thus become a kind of electro-magnet, the proportion of which are just the reverse of those generally used.

To communicate at very great distances several frames (n,) are joined up, the wires being joined up in series or for tension.

The efficiency can be considerably increased and the direction of transmission localised in avoiding dispersion at the sides or behind the frame, by means of a thick screen composed of superposed sheets of paper soaked in paraffn ; the paper could also be impregnated with varnish, glue or any other equivalent material.

Improvements in and relating to Telegraphic and Telephonic Systems.
GB190126600

1902-11-06

Relates to the arrangements of circuits in telegraph and telephone systems. In Fig. 1, a microphone b is shown in a main line a, and at the receiving-station is connected to the middle of a split battery f, g, two receivers l, m being connected between battery and earth. When induction coils are employed, two separate primaries take the place of the receivers in the Figure, and the receiver employed is placed in a common secondary. In another modification, the transmitter is bridged across two parallel closed circuits, including primaries and batteries, the secondary forming the line circuit.

This invention relates to telegraphic and telephonic systems and has for its object certain improvements hereinafter more particularly referred to and finally pointed out in the claims.

As is well known, in order to prevent over-heating of' the receiving and transmitting apparatus in telephonic or telegraphic installations it is necessary to prevent very strong current passing over them. It may however, in certain cases, be necessary to provide currents of varying energy within rather large limits, and this invention has for its object to provide a simple device which will permit currents of any force being passed through telegraphic and telephonic apparatus without the slightest risk of deterioration to such apparatus.

In order that my invention may be readily understood and carried into effect I will now proceed to describe the same fully with reference to the accompanying drawings wherein are shewn diagrammatically several methods of carrying my. invention into practice.

In the drawings :

Figure 1 shows the principle of my improved arrangement, and

Figures 2. and 3 are modified arrangements.

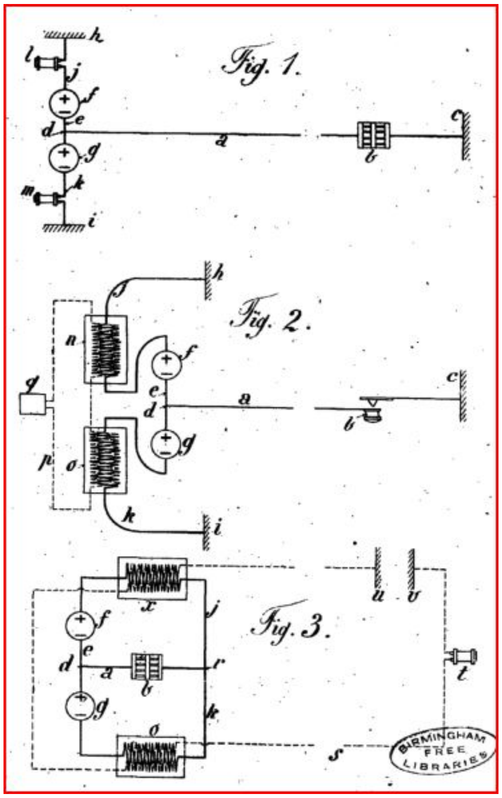

My improved system consists essentially of a line a of any convenient length in which is interpolated a microphone b or other appropriate transmitting apparatus. One end of this line is grounded at c and the other end is connected at d to.a wire e adapted to place into communication the opposite poles of the two sources of electricity of equal power, which for greater clearness sake I have represented as two single batteries f and g. These two batteries are respectively connected to earth at h and i. In the earth wires j and k are interpolated telegraphic or telephonic receivers l, m, of any convenient construction. In the example shown in the drawing the receivers and m are indicated as being telephonic receivers.

Having thus described the general arrangement of the system I will now explain the function of the various apparatus of the system.

Presuming that the wire a is non-existent and that the communication of the battery f and g with the earth has been established, a current passes from the positive pole of the battery f to earth through the wire then back .through the wire k, battery 9 and finally through the wire e to the negative pole of the battery f. Thus a closed circuit is established. Admitting now the wire a to be connected to the wire e at d and the transmitter b operated, the single current generated in the previous case will be replaced, by two currents which circulate in the same direction.

One of these currents starts from the positive pole of the battery f through the wire through earth from h to c through the transmitter b, wire a and wire e to the negative pole of the battery f.

The other current starts from the positive pole of the battery 9 through the 'wires e and a, the transmitter b, through earth from c to i and wire k to the negative pole of the battery g.

It will thus be seen that the wire a and transmitter b are simultaneously traversed by two equal currents in inverse directions.

In Figure 2 I have shown a telegraphic transmitter or key b but the general arrangement of the system is the the same as that previously described with the exception that the wires j and k form the primary conductors of the transformers n and o. The secondary conductor p, which is common to both transformers, leads,to the receiving apparatus q, for example a telegraphic receiving instrument.

Each contact ejected with the key b produces, as before explained a double derivation acting by subtraction of intensity upon the two transformers n and o, produces an induced current which is reproduced in totality in the receiving apparatus q. In the case of telephonic transmission the key must be replaced by a microphone as shown in Figure 1, while further the receiver q would also be a telephonic receiver.

By extending the same principle I can, as indicated in Figure 3, replace the two earth wires by two conductors and k joined together at r after having traversed the transformers x and o. The induced current may then be conveyed to the receiving instrument t over two lines, or as in the previous example over a line s and two earth connections u and v, placed more or less apart from each other.

It will be noticed that the microphone b forms part of the neutral conductors.

It acts exactly as in Figure 1. The differences of resistance produced by its vibrations influence the two coils x and o, and it is easy to understand that, whatsoever may be the intensity of the two electric sources f and g, the microphone is never exposed to the risk of becoming overheated for the reason that it is traversed by two equal currents in opposite directions.

It will thus be understood that the transmissions will be rendered perceptible in the receivers through the differences of intensity of the electric current; the differences being due to the differences of resistance of the neutral conductor which, in receiving two equal currents in opposite directions, is not much affected itself.

Improvements relating to the Transmission of Telegraphic and Telephonic Signals.
GB190117711

1901-02-19

The local circuits at each station are doubly earthed at e, f, j, k symmetrically about a straight line joining the stations. The earths may be formed of porous pots containing water in which the ends of the conductors are immersed, or deep ditches filled with gas coke may be employed. Insulating screens may be placed in deep ditches dug behind the earths.

This invention relates to an improved method of transmitting electrical telegraphic and telephonic signals from one point to another without the employment of wires connecting the two points.

In my improved system of transmission, the two poles of the transmitting station are put to earth and the two conductors, upon which the receiving apparatus is mounted are also both put to earth; the signals emitted by the transmitting apparatus are transmitted to the receiving station through the earth alone.

In order that my invention may be readily and clearly understood, I have represented in the accompanying drawing, by way of example only, a diagrammatic view of my device as a whole.

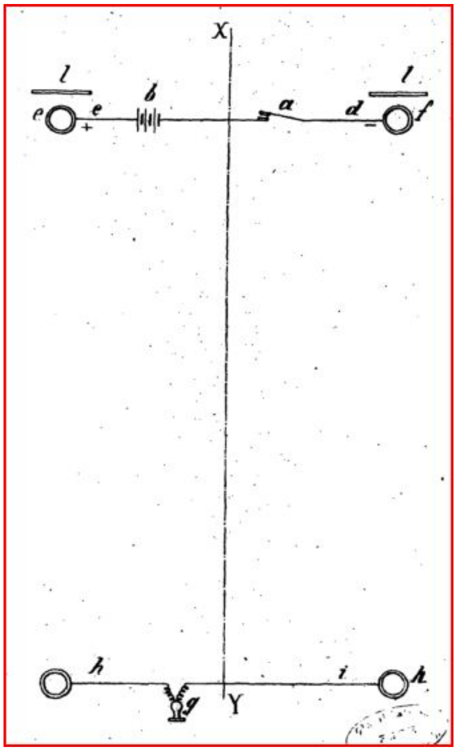

As shewn in the drawing, the transmitting station comprises a transmitter a one of the terminals of which is electrically connected to one of the poles of a source of electricity b whilst the other pole communicates with the earth; the second pole of the source of electricity is also run to earth and with this object the two conductors c d terminate in two porous pots e f embedded in the ground and filled with water.

The receiving station comprises a receiver 9 such as a telephone for example, the conductors h i which are put to earth by the intermediary of two porous pots j k similar to those of the transmitting station.

The porous pots of the receiving station should be arranged upon a line parallel with that which joins the two porous pots of the transmitting station, and they should be equidistant from an axis passing through the middle point of this line and at right angles to said axis.

The porous pots may advantageously be replaced by deep ditches sunk down to the humid earth, filled with gas coal or coke and containing at their centre a plate of carbon which is a good conductor, to which the transmission and receiving wires respectively are attached.

When contact is established by manipulation the transmitter a, the electric circuit passing through the source of electricity b, the conductor c, the porous pot e earth-connected to the other porous pot f the conductor d and the transmitter a, is closed. The current thus produced causes a change of condition in the ground ; all around the porous pots electric radiation is produced, the polarity of which is that of the corresponding conductor; its density decreases in proportion as it is further removed from the point of departure. This radiation is infinite with respect to space if nothing obstructs its passage.

The radiation extends so much the further according as the distance between the two porous pots of the transmitting station is greater.

These radiations, starting respectively from the porous pots e f thus create two zones one of which is positive and the other negative, the intersection of which constitutes % line absolutely neutral. It follows from this that the porous pots of the receiving station should be arranged upon either side of the neutral axis, because if these two earth connections were upon the same side they would both collect fluid of the same polarity, and in these conditions the receiver would not be traversed by any current.

In order to collect the maximum of current at the receiving station, the two porous receiving pots should be as nearly as possible equidistant from the neutral axis, as if one of the two porous pots is nearer than the other, the difference of separation of that which is the furthest removed is not only lost but is prejudical since it causes useless resistance without adding anything to the effect.

Experience has demonstrated that the disposition which furnishes the maximum of reception is that which consists in arranging the porous pots of the receiver at the same interval as separates the porous pots of the transmitting station.

In order to install my method of electrical transmission, it is first of all necessary to draw a straight line a y connecting the two stations and .then to produce upon each of these stations a line proceeding from this line x y at right angles, the said perpendicular lines being of equal length.

The transmitting and receiving appliances may be arranged at any convenient point on the conductor which connects the porous pots.

In order that the whole of the radiation emitted may be utilised at the receiving station, the propagation of this radiation is prevented (at the transmitting station upon the side opposite to that on which the receiving station is situated by burying in the ground an insulating screen l which may be of glass, tarred fabric, ebonite or the like, or by sinking a sufficiently deep ditch with or without an insulating screen.

Each time contact is made at the transmitter a, a sharp blow is audible at 'the receiving telephone g ; a second blow is heard when the circuit is broken.

It is obvious that there may be arranged both at the transmitting and receiving station transmitting and receiving appliances of the greatest possible sensibility, either for serving as relays for telegraphic purposes, or for directly receiving and recording the signals; for example an ink siphon mounted upon a galvanometric frame, or any other suitable device.

My method of transmission also permits of the transmission of speech; it is only necessary to replace the transmitter by an ordinary microphone interposed either directly upon the circuit, or arranged upon the primary of an induction coil, the secondary wire of which terminates at each extremity of the two porous pots.

It will of course be understood that in practice each station comprises both a transmitter and a receiver.

Improvements relating to the Transmission of Electric Currents
GB190108906

1901-04-30

Telegraph and telephone systems are arranged with the line-wire a connected to the center point of a battery c, c, the opposite poles of which are earthed and shunted by a normally closed circuit key g or microphone. The earth connections are made through non-polarizable cells f, f, containing, for example, two plates of copper immersed in copper-sulphate solution. The receiver is bridged between one earth and a switch h by which the transmitter is normally connected to line. The microphone may be inserted in the primary of an induction coil, the two secondaries of which oppose and are connected to the two earths and line. The two earth connections may be brought nearer together by the receiver and switch in the primary of the induction coil, so that this coil acts both to receive and transmit according to the position of the switch h. The microphone is arranged in the primary of a second induction coil.

This invention relates to improvements in the transmission of electric currents and is more particularly adapted for telegraph and telephone lines, whereby condensation in the cable is avoided, while at the same time "tapping" of the , line at any point thereof, for the purpose of ascertaining the purport of the message in transmission is rendered impossible.

My improved system of transmission comprises a single line wire between the two stations and branched at these two stations to the zero point of a circuit, the two extremities of which are connected to earth or the sea ; the current furnished at each station by a local source of electricity being sent through the circuit by means of any convenient transmitting apparatus, (such as a Morse key, microphone or the like) either direct or by induction.

Each station is furnished amongst other apparatus, with a receiving apparatus consisting of a key-instrument connected with the line-wire and permitting the current sent into the line from the transmitting station, to pass into a conductor leading direct to one of the wires running to earth and to which is connected the receiver proper.

In order to avoid the polarization of the earth-wires, each such wire is connected with a conductor leading to a copper or other metallic blade immersed in an impolarizable saline solution in which is disposed a second similar blade connected to earth or sea.

In the accompanying drawings :

Figure 1, is a diagram showing my improved system as applied to telegraphy

Figure 2, is a diagram showing the same applied to telephony.

Figure 3 shows the system of Figure 2 with induced current.'

Figure 4, shows a modified arrangement.

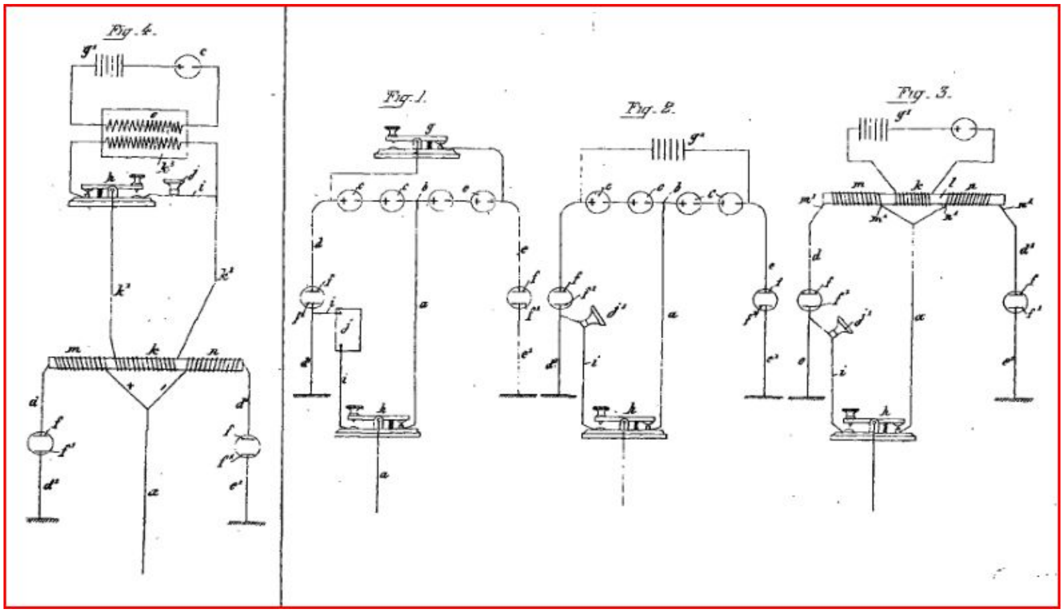

In all the figures the same letters of reference designate the same parts.

Referring to Figure 1 the two stations are connected by a single line-wire a, branched at the zero point b of the battery formed of the different elements c of which the end poles are connected to earth, or to sea, by a conductor d e lead- ing to a copper or other metallic plate f immersed in a liquid of good conductivity, such as for example a solution of a sulphate of copper, in which another copper plate f1 is placed connected to earth, or sea by a conductor d1 .

A transmitting apparatus g of any convenient construction, such as a Morse- key instrument, is connected with the two end poles of the battery in such a manner that in its position of rest it closes the circuit of the latter by establishing a short circuit so that no current passes, (or at most a very weak one) through the wire line a or through the earth-wires.

To send a current to the receiving station, the key 9 is depressed whereupon the short circuit is interrupted and the current is caused to pass simultaneously through the line-wire a and the earth-wires d d1 e e*. The line-wire is in reality traversed by two currents in opposite directions, so that it is in an apparently neutral state, and thus no interception of the messages can be effected for the, reason that if a receiver be interpolated, at any point of. the line, said receiver would receive two currents in opposite directions' which would neutralize each other.

Each station is furnished also with a receiving apparatus consisting of a commutator h connected with the line-wire a and permitting, when depressed, the current from the transmitter to pass into a wire i leading direct to one of the earth-wires (d1 for example) and in which is interpolated the receiving instrument j.

To receive signals sent from the transmitting station, it 'is necessary that the commutator lever h be kept depressed, so as to cause the current from the transmitting station to pass into the wire In connection with the receiving instrument.

'The connections may also be established in such a manner that when the commutator h is in its normal .position, it cuts the wire (and connects the wire with that part of the said wire which leads to the other station ; in this case for transmitting a signal it is necessary to depress the commutator lever h previously so as to re-establish the continuity of the line-wire and the operator then acts on the key-lever g as hereinbefore stated.

My arrangement is equally applicable to telephonic messages as will now be explained with reference to Figure 2, The transmitting instrument g is in this case replaced by a microphone g1 and the receiving instrument j is replaced by a telephone receiver j 1 the operation of the arrangement is exactly, the same as that above described.

The vibrations of the microphone cut the short circuit and the current from the local, source of electricity thereupon passes over the line-wire a and the earth-wires d d1 c el as previously stated.

The telephonic message is received in the same manner as before through the intervention of the commutator h which passes the current, sent from the transmitting station; to the telephonic receiving instrument jl.

The line-wire may also be operated by induction currents a3 is indicated at Figure 3.

The local circuit in which the transmitting apparatus (for instance, the microphone g1) is inserted comprises a primary winding k surrounding a central soft-iron core, or bundle of wires, l which is further provided with two secondary windings m m in opposite directions. The line wire a is connected to each of these secondary windings at the two extremities m1 n1 of opposite polarity and the opposite extremities ms p2 are connected with the earth-wires d d c set in the manner already explained.

When operating the transmitting apparatus or when speaking into the microphone 0 for example, induction currents are generated in the secondary windings m n and they then affect the receiving apparatus jl at the receiving station as already stated.

The primary winding and the two secondary windings may have the three wires wound in the same direction upon the soft iron core and the two extremities of the primary convolutions are connected with the local circuit in which is interpolated the transmitter while two extremities of opposite polarities of the secondary convolutions are connected to line and the two osiers to earth as previously explained.

In my improved transmitting arrangement the two earth wires at 'each station must be the more separated from each other the longer the distance is between the two stations so that the resistance of the earth between the said earth-wires will be sufficient to cause the single line-wire to be traversed by the greatest possible quantity of the current.

In order to reduce as far as possible the distance between the earth-wires the beginning of the single wire a and the wires rl rl' leading to earth are advantageously rm distance from the transmitting instrument and receiving instrument respectively at the telegraphic or telephonic station and thus permit the two earth connections of the two telegraphic or telephonic stations to be placed nearer each other.

The telegraphic or telephonic station is then divided and the transmitting instrument is connected with the magnet with its earth connections and line wire a ' by two conducting wires kl k2 connected with the ends of the convolutions k which generates induction currents in the secondary windings m n (Figure 4).

The current sent into the conducting wires k1 k2 and. passing through the convolutions k may itself be an induced current generated in the coil k3 when a current from the source of electricity c is passing through the. primary coil o through the action of the transmitting microphone gl for example.

In the case when' the distance of the telegraphic or .telephonic transmitting station from its earth-connections is not too excessive the current traversing the conductors k1 1.:2 instead of being an induction current may be furnished direct by the source of electricity of such station, and naturally the, transmitting instrument, (microphone or 'key), and the source of electricity are then placed direct in the circuit;1 k2.

An interrupter A interpositioned in the conductor kl permits of cutting the circuit in leading to the secondary coil k3 and of closing the circuit for the conductor i leading to the conductor k 2 in which is interpolated' the receiving instrument j.

Thus the coil k is either inductor or induction according to whether there is transmission or reception.

It is to be well understood that my system of transmission is entirely independent of the nature of the telegraphic or telephonic apparatus employed therewith, and that the source of electricity at the stations may be of any convenient kind, either derived from a battery, accumulators, dynamo, magneto-electric machine or the like.

COMPLETE SPECIFICATION.

This invention elates to improvements in the transmission of electric currents and is more particularly adapted for telegraph and telephone lines, whereby condensation in the cable is avoided, while at the same time "tapping" of the line at any, point thereof, for the purpose of ascertaining the purpose of the message in transmission is rendered impossible.

My improved system of transmission comprises a single line-wire between the two stations and branched at these two stations to the zero point of a circuit, the two extremities of which are connected to earth or the set; the current furnished at each station by a local source of electricity being sent through the circuit by means of any convenient transmitting apparatus, '(such as a Morse key .microphone or the like) either direct or by induction.

Each station is furnished, amongst other apparatus, with a receiving apparatus consisting of a key-instrument connected, with the line-wire and permitting the current sent into the line from the transmitting station, to pass into a conductor leading direct to one of the wires running to earth and to which is connected the receiver proper.

In order to avoid the polarization of the earth-wires, each such wire is connected with a conductor leading to a copper or other metallic blade immersed in an impolarizable saline solution in which is disposed a second similar blade connected to earth or sea..

In order that my invention may be clearly understood and readily carried into effect I will now proceed to describe the same more fully with reference to the drawings accompanying my Provisional Specification in which:.

Figure 1, is a diagram showing my improved system as applied to telegraphy.

Figure 2, is a diagram showing the same applied to telephony.

Figure 3 shows the system of Figure 2 with induced current.

Figure 4 shows a modified arrangement.

In all the figures the same letters of reference designate the same parts.

Referring to Figure 1 the two stations are connected by a single line-wire a branched at the zero point b of the battery formed of the different elements c of 'which 'the end poles are connected to earth, or to sea, by a conductor d e leading to a copper or other metallic plate f immersed in a liquid of good conductivity, such as for example a solution of sulphate of copper, in which another copper plate f1 is placed connected to earth, or sea by a. conductor d1 e1.

A transmitting apparatus q of any convenient construction, such as a Morse-key instrument, is connected with the two end poles of the battery in such a manner that in its position of rest it closes the circuit of the latter by establishing a short circuit so that no current passes, (or at most a very weak one) through the wire line a or through the earth-wires.

To send a current to the receiving station, the key g is depressed whereupon the short circuit is interrupted and the current is caused to pass simultaneously through the line-wire a and the earth-wires d d1 e el. The line wire is in reality traversed by two currents in opposite directions, so that it is in an apparently neutral state, and thus non interception of the messages can be effected for the reason that if a receiver be .interpolated at any point of the line, said receiver would receive two currents in opposite directions which would neutralize each other

Each station is furnished also with a receiving apparatus consisting of a commutator h connected with the line-wire a and permitting, when depressed. the current from the transmitter to pass into a wire i leading direct to one of the earth-wires (d1 for example) and in which is interpolated the receiving instrument j.

To receive signals sent from the transmitting station, it is necessary that the commutator lever h be kept depressed, so as to cause the current from the transmitting station to pass into the wire i in connection with the receiving instrument.

The connections may also be established in such a manner that when the commutator h is in its normal position, it cuts the wire' a and connects the wire i with that part of the said wire which leads to the other station; in this case for transmitting a signal it is necessary to depress the commutator lever h previously so as to re-establish the continuity of the line-wire and the operator then acts on the key-lever as hereinbefore stated.

My arrangement is equally applicable to telephonic messages as will now be explained with reference to Figure 2. The transmitting instrument g is in this case replaced by a microphone 1 and the receiving instrument j is replaced by a telephone receiver il the operation of the arrangement is exactly the same as that above described.

The vibrations of the microphone cut the short circuit and the current from the local source of electricity thereupon passes over the line-wire a and the earth-wires d d1e e1 as previously stated,

The telephonic message is received in the same manner as before through the intervention of the commutator h which passes the current sent from, the transmitting station to the telephonic receiving instrument j1.

The line wire may also be operated by induction currents as is indicated at Figure 3.

The local circuit in which the transmitting apparatus (for instance, the microphone g1). is inserted comprises a ,primary winding k surrounding a central soft iron core,, or bundle of wires, 1 which is further provided with two secondary windings m n in opposite directions. The line wire a is connected to each of these secondary windings at the two extremities in n1 of opposite polarity and the opposite extremities m2 n2 are connected with the earth wires clt c ex in the manner already explained.

When operating the transmitting apparatus or when speaking into the microphone for example, induction currents are generated in the secondary windings m n and they then affect the receiving apparatus l at the receiving station as already stated. The primary winding and the two secondary windings may have the three wires wound in the same direction upon the soft iron core and the two extremities of the primary convolutions are connected with the local circuit in which is interpolated the transmitter while two extremities of opposite polarities of the secondary convolutions are connected to line and the two others to earth as previously explained.

In my improved transmitting, arrangement the. two earth wires at each. station must be the more separated from each other the longer the distance is between the two stations so that the resistance of the earth between the said earth-wires will be sufficient to cause the single line-wire to be traversed by the greatest possible quantity of the current.

In order to reduce as far as possible the distance between the earth-wires the beginning of the single line wire a and the wires d d1 leading to earth are advantageously removed at a distance front the transmitting instrument and receiving instrument respectively at the telegraphic or telephonic station and thus permit the two earth connections of the two telegraphic or telephonic stations to be placed nearer each other.

The telegraphic or telephonic station is then divided and the transmitting instrument is connected with the magnet with its earth connections and line wire z by two conducting wires k' k2 connected with the ends of the convolutions which generates Induction currents in the secondary windings m n (Figure 4).

The current sent into the conducting wires k1 k2 .and, passing through the convolutions k may itself be an induced current generated in the coil k3 when a current from the source of electricity c is passing through the primary coil or through the action of the transmitting microphone f/I for example.

In the case when the distance of the telegraphic or telephonic transmitting station from its earth connections is not too excessive the current traversing the conductors k le2 instead of. being an induction current may be furnished direct by the) source, of electricity of such station, and naturally transmitting instrument (microphone or key) and the source of electricity are then placed direct in the circuit leI k-S.

An interrupter It interpolated in the conductor k1 permits of cutting the circuit in leading to the secondary coil k3 and of closing the circuit for in conductor i leading to the conductor 3 in which is interpolated the receiving instrument j1.

Thus the coil k is either inductor or induction according to whether there is transmission or reception.

It is to be well understood that my system of transmission is entirely independent of the nature of the telegraphic or telephonic apparatus employed therewith, and that the source of electricity at the stations may be of any convenient kind,

---

Improvements in Connections of Telephonic and Telegraphic Instruments to Conducting Lines.
GB189706773

1897-03-15

Telegraph and telephone lines are arranged so that any pair of lines bearing instruments may be used as a lead or return for another set of instruments without interference. Thus, as an example, four metallic circuits carry four separate sets of instruments individually but may have circuits A and B, C and D, used as other metallic circuits, or A and B may form a lead and C and D a return for a seventh metallic circuit. The Specification describes telephone instruments as used on the metallic circuits and telegraph or telephone instruments inserted in the earth return circuits. Fig. 1 shows a pair of lines L<1>, L<2> forming a metallic circuit and capable of use as the line with an earth return through L<3>, L<3>, the lines L<2> being connected to the loop at the centre of the secondary coil S', S<2> of the induction coil X. The receiving-instrument is shown in a bridge, such as Z. The positions of the transmitting T and receiving R instruments may be reversed. Each side of a metal loop may also be used as the line with an earth return.

This invention relates to the connecting of telephonic and telegraphic instruments to conducting lines in such a manner that several telephonic and telegraphic messages can be simultaneously transmitted without interfering with one another.

Fig. 1 of the accompanying drawings shews diagrammatically connections for a telephone and also for another telephone or telegraph to a double conducting line.

Fig. 2 shews a modification.

Fig. 3 shews how seven sets of the connections may be applied to four double conducting lines.

Fig. 4 shews the connections for a telephone and telegraph to a single line with earth return.

Fig..5 shews connections for telephoning over two telegraph lines one of them forming a metal return.

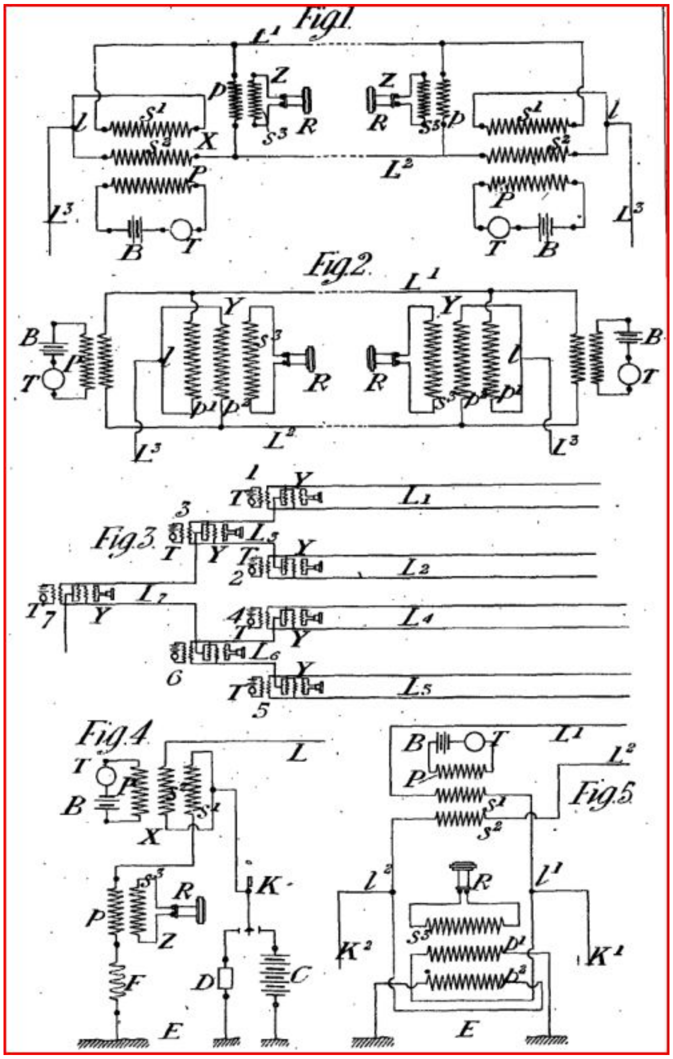

Referring first to Fig. 1 the two lines Ll L2 are at each end connected as shewn at X through two equal secondary coils 8 1 and s 2 wound with a primary coil P which is in circuit with a telephone transmitter T and local battery B.

The lines L1 L2 are so connected to the secondary coils 8 1 s 2 that the currents induced in the lines by the telephone transmitter are always in opposite directions.

The telephone receiver R, as shewn at Z, is in circuit with a secondary coil 8 3 wound with a primary coil p connecting the two lines L1 L2.

At a point l intermediate to the coils s1 s2 a third line L3 is connected, this line being employed for telephone or telegraph transmission. When either of the telephone transmitters T is employed, the induced currents pass through the primary coil p thus causing the receivers R to act. But should any current flow by L3 it must flow in opposite directions through 8 1 and s2, and therefore in the same direction along the lines L1 L2 so that it cannot influence the receivers R nor the transmitters T. At each end a line L3 may be connected as shewn by Y, Fig. 2, at l a point intermediate to two equal primary coils p1 p2 which connect the lines Ll LO and are wound with the secondary S3 of each telephone receiver R.

With connections thus arranged any current flowing in the same direction through the coils pl p 2 actuates the telephone receivers R, whereas a current from L3 flows in opposite directions through the coils pl p 2 and therefore is ineffective on the receivers R. By combining several ordinary transmitters T and several of the receiver arrangements marked Y in Fig. 2 with a number of double lines as shewn in Fig. 3, a number of messages can be sent over these lines without interfering with each other.

Thus referring to Fig. 3 the transmitter 1 works over the pair of lines L1 in the - usual way, one of these lines serving as metal return. In like manner the transmitter 2 works over the pair L2, transmitter 4 over the pair L 4 and transmitter 5 over the pair L 5. Again the transmitter 3 works through the pair of lines L 3 and the two pairs L 1 and L 2, one of these pairs serving as metal return.

In like manner the transmitter 6 works through the pair of lines L 6 and through the two pairs L 4 L 5. Finally the transmitter 7 works through the pairs of lines L 7 and through the two pairs L 3 and L 6 and the. four pairs L 1 L 2 L 4 and L 5, two of these pairs such as L 4 and L 5 serving as metal return.

Obviously this system may be duplicated, and reduplicated so that one transmitter can be made to work over 8 pairs or 16 pairs of lines or other multiple of 4.

As shewn in Fig. 4 a telephone and a telegraph are connected to one line L having earth return. In this case X and Z are the telephone arrangements so marked in Fig. 1, K is the key, C the battery and D the receiver for telegraphing in the connection of Z to earth there is an adjustable resistance F to balance that of the line. As shewn in Fig. 5 two sets of the telegraphic connections shewn in Fig. 4 and two lines Ll L2 which may be both to one place or each to a different place with earth return are combined with one telephone arrangement.

The transmitter T by its primary P induces currents in the secondaries $ 1 89 both in the same direction and therefore in opposite directions in the lines Ll L2. The currents being in the same direction in p1 and p2 act on the secondaries 1 3 of the receivers 1,. Currents from the telegraph gi dividing at 11 pass in opposite directions through sl and pl and therefore do not affect the telephonic receivers. And in like manner the currents from the telegraph K2 are ineffective on the receivers.

For the sake of clearness the primary and secondary coils have been shewn in the form of many lines side by side.

It is to be understood however that they are wound in each case all on one bobbin.

In Figs. 3,4 and 5 the arrangements are shewn at only one end of the conducting lines.

It is to be understood however that they are symmetrically repeated at the other end as shewn in Figs. 1 and 2.

COMPLETE SPECIFICATION.

This invention relates to the connecting of telephonic and telegraphic instruments to conducting lines in such a manner that several telephonic and telegraphic messages can be simultaneously transmitted without interfering with one other.

Figure 1 of the drawings accompanying my Provisional Specification shews diagrammatically connections for a telephone and also for another telephone or .telegraph to a double conducting line. Figure 2 shews a modification. Figure 3 shews how seven sets of the connections may be applied to four double conducting . lines: Figure 4 shews the connections for a telephone and telegraph to a single line with earth return. Figure 5 shews connections for telephoning over two telegraph lines one of them forming a metal return.

1 . Referring first to Figure 1 the two lines Ll L2 are at each end connected as shewn at X through two equal secondary coils s1 and 8 2 wound with a primary coil P which is in circuit with a telephone transmitter T and local battery B.

The lines Ll L 2 are so connected to the secondary coils 8 1 s'" that the currents induced in the lines by the telephone transmitter arc always in opposite directions.

The telephone receiver R, as shewn at Z, is in circuit with a secondary coil s3 wound with a primary coil p of high resistance connecting the two lines L1 L2.

At the point intermediate to the coils sr s 2 a third line L3 is connected, this line being employed for telephone or telegraph transmission. 'When either of tho telephone transmitters T is employed, the induced currents pass through tho primary coil p thus causing the receivers R to act. But should any current flow by L3 it must flow in opposite directions through sl and s2, and therefore in the same direction along the lines Ll L2 so that it cannot influence the receivers It nor the transmitters T. At each end a line L3 may be, connected as shewn by Y, Figure 2, at l a point intermediate to two equal primary coils p1 p2 which connect the lines L1 L2 and are wound with the secondary 8 3 of each telephone receiver R. With connections thus arranged any current flowing in the same direction through the coils yx actuates the telephone receivers R, whereas a current from L3 flows in opposite directions through the coils p1 p2 and therefore is ineffective on the receivers R. By combining several ordinary transmitters T and several of the receiver arrangements marked Y in Figure 2 with a number of double lines as shewn in Figure 3, a number of messages can be sent over these lines without interfering with each other.

Thus referring to Figure 3 the transmitter 1 works over the pair of lines L 1 in the usual way, one of these lines serving as metal return.. In like manner tho transmitter 2 works over the pair L 2, transmitter 4 over the pair L 4 and transmitter 5 over the pair L 5. Again the transmitter 3 works through the pair of lines L 3 and the two pairs L 1 and L 2, one of these pairs serving as metal return.

In like manner the transmitter ti works through the pair of lines JLA b and through the two pairs L 4 L 5. Finally the transmitter 7 works through the pairs of lines L 7 and through the two pairs L 3 and L 6 and the four pairs L 1 L 2 L 4 and L 5, two of these pairs such as L 4 and L 5 serving as metal return.

Obviously this system may be duplicated, and reduplicated so that) one transmitter can be made to work over 8 pairs or 16 pairs of lines or other multiple of 4.

As shewn in Figure 4 a telephone and a telegraph are connected to one line L having earth return. In this case X and Z are the telephone arrangements so marked in Figure 1, K is the key, C the battery and D the receiver for telegraphing.

In the connection of p to earth there is an adjustable resistance F to balance that of the line. As shewn in Figure 5 two sets of the telegraphic connections shewn in Figure 4 and two lines Ll L2 which may be both to one place or each to a different place with earth return are combined with one telephone arrangement.

The transmitter T by its primary P induces currents in the secondaries sl S2 both in the same direction and therefore in opposite directions in the lines L1 L2. The currents being in the same direction in pI and pS act on the secondaries i1 of the receivers It.

Currents from the telegraph K1 dividing at 1 1 pass in opposite directions through sl and I > I and therefore do not affect the telephonic receivers. And in like manner the currents from the telegraph K2 are ineffective on the receivers.

The connections of p1 and p2 to earth may have suitable resistances as shewn at F in Figure 4.

For the sake of clearness the primary and secondary coils have been shewn in the form of many lines side by side.

It is to be understood however that they are wound in each case all on one bobbin.

In Figures 3,4 and 5 the arrangements are shewn at only one end of the conducting lines.

It is to be understood however that they are symmetrically repeated at the other end as shewn in Figures 1 and 2.,

---

Perfectionnements aux bobines d'induction
FR382386

1908-02-05

The present invention relates to an improved induction coil in order to obtain induction currents in very short periods. These currents, whose durations are extremely short 5, enjoy special properties which find application mainly in telegraphy and wireless telephony.

This induction coil is characterized by the absence of the magnetic core and the winding mode of the inductor and the armature on a short length coil, but a very large diameter, the inductor coil being virtually drowned in May 1 the armature winding.

The appended drawing schematic represents for example an induction coil according to the present invention.

FIG. 1 is a section along line A-20B of FIG. 2 ;

FIG. 2 is a plan.

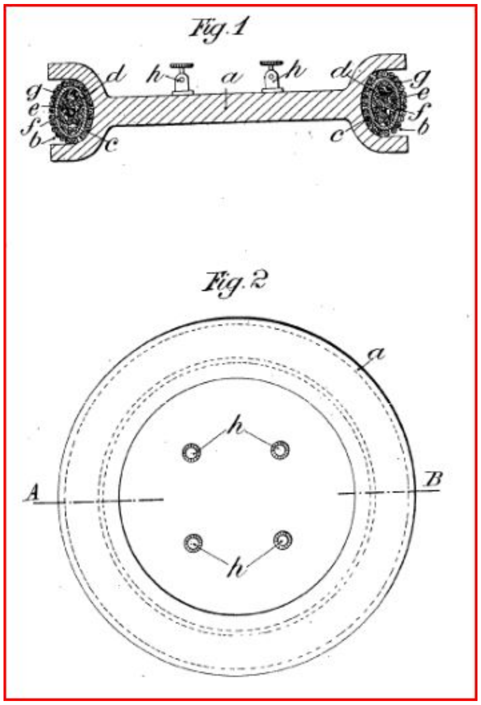

As shown in these drawings, this coil has a very large diameter which can be up to several meters; 26 is drawing the round, but it could equally be square or affect any other polygonal shape. This reel is very long because all winding son are housed on its outskirts in a suitable groove b; it will be understood by 3o begins winding on the rim of this kind of pulley a first layer of wire induced E and then covered with a suitable insulating and d are wound over this insulating layer the inductor wire E which is covered another 35 isolated layer; we **réuuïï uêUi** the insulating layers d and f, so as to isolate the inductors e and ends the coil by an external winding armature wire g. ko

Is carried out in this way a large-diameter ring compound induced son c, g, containing a central core of an adequately insulated son inductors.

The machine has four terminals 5 h which two are used to secure the ends of the armature winding and the other two ends of the inductor winding.

We can arrange several induced son and son more inductors; in this case, each terminal 5o h will be replaced by a multiple switch that will bring together the different armatures on, and the different inductors on, either voltage or the contrary in quantity, following the effects that one wishes to obtain.

Dispositif de transmission téléphonique
FR377785

1907-09-14

Telephone transmission system of mounting each station transmitter and or serial receivers and to bring the two stations by three son line, including 5o first two respectively together the poles of the two stations and the third is mounted derivation at a point of each station caught between the transmitters and the receiver or receivers, one or even two of these three son may be replaced by equivalent landings.

The present invention provides a way to increase the distance at which it is possible to call or to increase the intensity of sound transmitted at a given distance 5.

When installing a regular telephone line, it is customary to bring the two stations using two wires, one for the second round for the return; at-10 there is no fear induction from neighboring lines, one connecting the two stations by a thread alone and return through the ground at each station.

FIG. 1 of the accompanying drawing shows the 15 most simple provision of a regular line with two son; there is one at each station magnetic phone, a1 met by both son line c and d; when speaking in the phone, the phone is 30 a1 receiver and vice versa; this provision is entirely satisfactory from the standpoint of the actual transmission, but it has the obvious disadvantage that the two parties can not break interfere, each necessarily listen to while the other speaks .

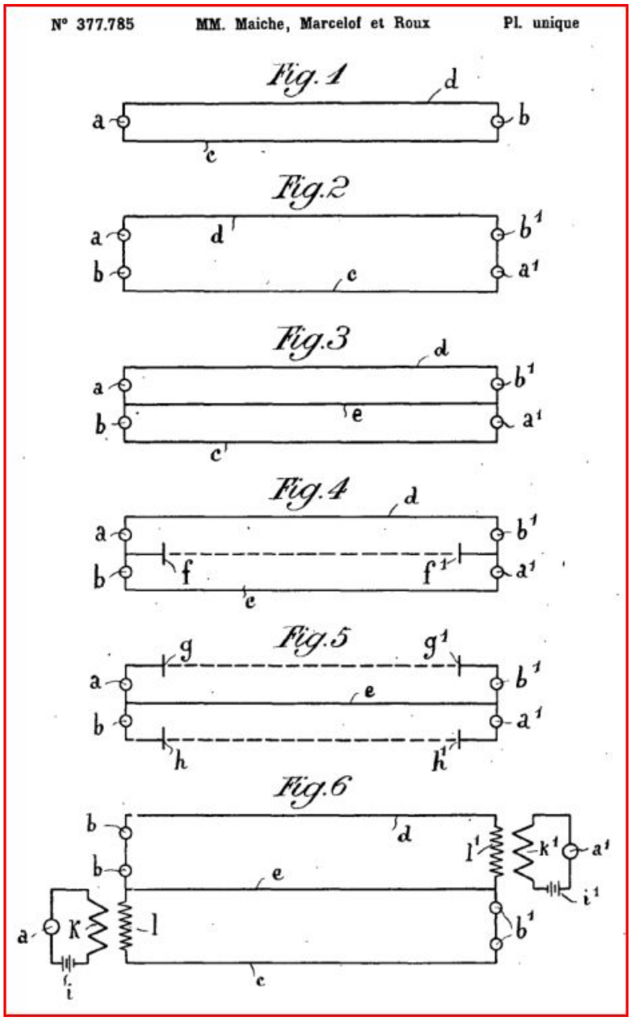

To overcome this disadvantage, it immediately offered to have each transmitter station a phone or a 1 and a receiver or telephone b bl, allowing 3 to the two interlocutors to talk while listening. This solution suffers a disadvantage from the standpoint of the transmission; in fact, as shown in Fig. 2, the four phone a, b, a}, bx 3 are mounted in series on the line cd, so that the sound transmitted by a transmitter has or a1 necessarily travels in addition interesting b1 or b receptor both other phones of the circuit. The telephone circuit thus unnecessarily h has a high resistance which reacts disadvantageously either on the possible length of the line cd, either the intensity of the sound transmitted to a given line length.

The present invention relates to a device easy positive h to remedy the drawbacks both of the two transmission systems according to Fig. 1 and 2 and make a new telephone transmission system providing both the benefits of the 5 two known systems.

As shown in Fig. 3, this invention is simply to dispose between the two stations a third e line wire terminating at each station between the telephone transmitter 5 tor a or a1 and the receiver phone b or b1. In these conditions the wall sent by "will be fully received by b1 without crossing unnecessarily b and a1 resistance and vice versa speech transmitted by a1 will be fully received by b.

The need for three son line c, d, e 5 is required when the line is around other transmissions; but when isolated, it can simplify the transmission as shown in Fig. k or 5.

According to FIG. h, the wire line drifted o-tion e is deleted and replaced by a ground f and f1 taken at each station between the transmitter and receiver.

In special cases, it is even possible to reduce the line to a single wire e in '5 with each station two landings g, h and # 1, h1), the receiver and transmitter are each disposed station in series between the two landings; under these conditions, the two conductors c and d are formed by o the ground and there is no longer a single line wire e.

In the above, to simplify the description and drawings, we assumed the exclusive use of magnetic phones; 5 Fig. 6 shows the circuit diagram in the case of induction phones. In this case each position i includes a battery, a transmitter microphone and primary k of an induction coil, mounted in series o on a closed circuit; the line circuit c, d comprises series headphones b, b and b1, b1 and side and ll of the coil. In this

If the invention is to attach the derivative e line wire between the two receivers b, b or i1, A1 on the one hand and the secondary or ll 35 on the other hand, as shown in FIG. 6.

The result of these provisions is summed up by the possibility of quadrupling of the intensity of the reception for a given length of line, or the length of the line Zio maintaining the same intensity.

Nouveau mode de transmission sans fils des courants télégraphiques et téléphoniques en employant comme conducteurs la terre et l'eau
FR320267

1902-12-05

To understand our invention, it is good to remember that in an ordinary telegraph line a thread, more or less isolated, from a point to another, communicates with earth at each end. A power source; and a transmitter unit are sandwiched between the starting line and earth. Upon arrival, a receiver also communicates with the line and the earth.

We know that each of the posts supporting the line passes a small amount of power that is lost in the ground. Large amount of current is now lost all **plus 1 large 5** that the distance between stations is itself larger and, consequently, the number of' poles is higher. Loss can become considerable, without preventing transmissions. It may even happen that the ao thread without being broken, drags on earth there for a certain length of son'parcours and continues to allow the current of a shift to another, albeit very weakly,

The experience proved us uninsulated conductor could even, for certain distances, be simply placed on the ground and that transmissions would nevertheless be possible.

We are able to continue, transmit signals is greatly increased if, at each end, the uninsulated wire touching the ground, met in a well-insulated wire that extends the line to a ground to the point where so are transmitters and receivers.

It is understood that the ground at the end of a secluded part forces the return current from sinking into the depths of the earth ho and that, consequently, there is an infinitely small closure point where the uninsulated wire touches the ground. In other words, the loss suffered in ground contact, the uninsulated wire, is even less than each end of uninsulated wire so is an insulated wire from the ground at a greater distance.

Hence, if the lead wire, instead of iron, consisting of a much less conductive material, coal, for example, everything would be the same way. We also recognized that even when iron, coal or other similar material, are replaced by placed water, salt if possible, and, consequently, by a stretch of sea or any other part of sea whose geographic provision lends itself to this kind of installation.

When it comes to cross an arm of the sea, the isolated line extension is placed preferably in the direction of the straight line through the two points to be connected.

In case one transmission on a role, isolated leagues will be perpendicular to the straight line joining the two stations.

Isolated lines will not be too long, and therefore, will provide not only coast to coast, but between two points inland. The thickness of the salt water layer being at even greater depths, only about the thickness of the land mass, the proportions remain between water and adequate the earth so that water plays exactly the same role that the uninsulated conductor touching the ground.

To make our evplications as clear as possible, we have shown diagram-matically, in the accompanying drawing, an installation in which our invention is applied.

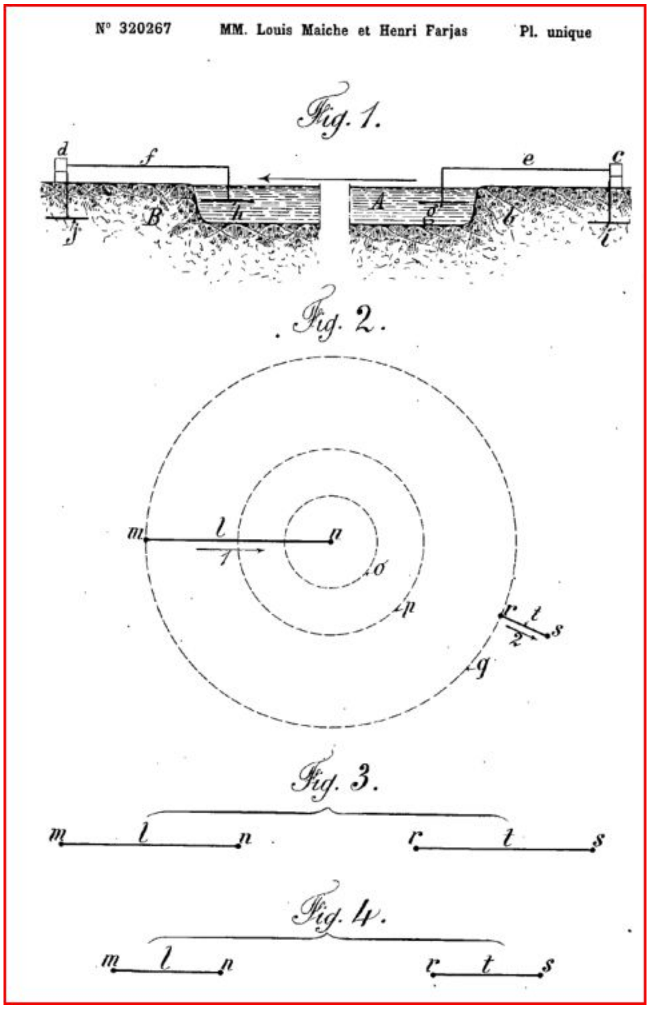

A is the sea (Fig. I) and 13 b are both shores, c and d both stations, e and f insulated, and the plates located in the water, i and it updates earth.
We see that the sea is an extension isolated son e and f, and it plays the role of an uninsulated conductor based on floor.

To make it clear how 3 o operates the propagation of electric currents, we assume an isolated line (Fig. A) within two earthed m and n. If the current is moving in the direction of the **llèche**, electricity arriving ashore in n flourishes spheres 3.) concentric o, p, q, etc., all points of each sphere simultaneously receiving the same amount of electricity if the propagation medium is uniform. Be grounded in r therefore receive the current along the earth m, r and if this land is connected by a line by t, s land to another, even very close to her, the current finding less resistance in the isolated line in the middle (Jun separates the two nets land, will follow this line, as shown in llèche a. It will be thus possible to establish a wireless communication between any point of the line / on which find transmitting devices, and any point on line T which will be **appareds** receptors.

To establish simultaneously on each of the leagues and t, a position transmitter and a receiver will require that the line is the same length as the line 1, J J as shown in Fig. 3.

If the medium is found to be twice as conductor between h and r between m and n, on the one hand, and between r and s on the other hand, the lines l and t may have as half the length of the distance between the points n and r (fig.û).

Of course, transmissions based on the principle that we have indicated, may be made by taking as con- bo different producers, not only the earth and the sea water, but the earth and air, land in this case, acting as a better conductor brought by the air; the insulated wire from the earth lead to the air at each station 7. tion to a more or less remote distance according to the distance that would take. It is obvious that the latter provision, the sea and fresh water advantageously replace the earth. 7°

In summary, we claim as our exclusive property, the new method, described above, transmission without telegraph and telephone currents, said transmission mode of making pass in a conductive medium the current emitted by the base station and to receive on arrival in this current an insulated wire of suitable length, which are sandwiched the receivers and resulting in a second medium whose specific conductivity is less than that of the first medium, but passes the total quantity of electricity because of its mass which is greater than that first medium, an insulated wire, on which the transmitting devices is also leading into the second medium, so that the first medium is an extension of two isolated lines, and the second medium acts as the return wire, the two isolated lines to be used both for shipping and for the reception.

Un nouveau mode de transmission et de réception des courants télégraphiques et téléphoniques sans autre conducteur que la terre ou l'eau

FR318007

1902-10-03

In an electrical circuit consisting of a power source and a metallic conductor whose ends are grounded, which closes the circuit, the earth acts as 5 driver not only a straight line between the two landings, but on a almost infinite extent; However, with a power density decreasing in proportion to the distance.

If parts of the ground are better conductors than others (because of the moisture, the saltiness of the water or any other cause) the current preferably passes through these parts describing curves or bends as does a river following the slopes of the terrain.

It follows that if, on part of the land used as return conductor, it artificially creates a better conductor line as the earth, it will be traversed by the current in preference to any other lower conductive part.

The effect is much more complete and can collect the current and note its existence when the best conductive line is insulated from earth and communicates with it by its ends. The current that would have crossed the land finding a wise not easier by the wire preferably follows and restores much faster electrical neutrality of the ground. It is isolated by cutting the line at any point that can be inserted in the course of the current instruments (telegraphic or telephonic receivers) that can detect its existence.

It is understood that if the isolated artificial line is a very good conductor, if its ends are connected to metal surfaces large enough in contact with the ground, the closing of the circuit will cause the ground to be in a state of electrical neutrality absolute.

Under these conditions, if a current electric is sent intermittently in the metal line, charge the earth, for each issue, a certain amount of electricity instantaneously discharge into the isolated artificial line.

Experience has shown that these phenomena can occur even when there is almost no gap between the two landings transmitter station or between two plates that collect the color rant crossing the line isolated the receiving station, to the condition that at each station the plates which lead to the two lands are carefully insulated from each other by means of a sufficiently thick insulating separation that no induction effect or condensation on one another can occur, and fairly large compared to the dimensions of the plates to prevent current from closing on itself too close to the to start. The current is so obliged to flourish "right and left of the landings, and the point where the positive and negative electric waves begin to reach is considerably related. I can further increase the desired effect by minimizing the current closure on itself through a number of insulating walls arranged below the plates to prevent **farm-i o ture** of the circuit by the depth of the earth, on the side opposite that in which the transmission must be received and, if necessary, above the plates.

In general, the plates will be placed vertically on either side of an insulating wall, their smaller dimensions being turned to the opposite position.

I can also use a wide insulating tube divided in two by a thick wall normal to isolated generators. The plates will be set back to back on either side of this partition in the interior of the tube and placed in communication with the outer conductor through insulated **son**.

The tube may be straight, or consists of two parts forming between them an angle properly calculated so that the circuit is closed by the earth is done by almost entirely in the **3 o** where the current is sent.

It is obvious that the installation is completed by transmitters and telegraph or telephone receivers that may be any, as the source of electricity.

I claim as my exclusive new property above described transmission and reception of telegraph and telephone currents without other driver as land or water, said system of separating the two landings each position by established insulating material to prevent the production of induction phenomena of condensation and the circuit closing on or by the depth of the earth, and force the air to flourish right and left, so that they can begin to get to and distant from the transmitter station, the two earthed of each position can thus without drawbacks, are moved towards one another until they fit on either side of the same insulating wall.

Nouvelle disposition de bobine d'induction
FR647 (E)
1903-02-18

Système de communication télégraphique et téléphonique

FR376100
1907-07-30

Mode de transmission de signaux télégraphiques
FR361733
1906-10-23

Nouveau moyen de transmission et de réception sans fil des courants électriques
FR321454
1903-01-10

Récepteur télégraphique pour toutes transmissions en général, avec ou sans fils
FR320467
1902-12-11

Nouveau récepteur des courants électriques
FR319195
1902-11-06

Nouveau moyen destiné à augmenter la sensibilité des microphones
FR319155
1902-11-06

Nouveau système de télégraphie et de téléphonie sans fil
FR318507
1902-10-17

Système de communication télégraphique et téléphonique
FR7464 (E)
1907-08-09

Dynamo - oder magnetelektrische Maschine
DE46075
1889-02-05

Transformateur téléphonique sans noyau magnétique
CH44013
1909-07-01

Wicklung von eisenfreien Transformatoren für Fernsprechzwecke.
AT40498
1910-01-10

Système de télégraphie et de téléphonie sans fil
FR392252
1908-11-21

Dispositif pour transmetteur et récepteur téléphonique
FR390346
1908-10-02

Other Patents

METHOD OF PRODUCING MOTIVE POWER
CA78259
1902-11-18

Appareil cinématographique
FR388334
1908-08-08

Perfectionnements aux papiers photographiques
FR400693
1909-08-04

Réducteur de pression pour tous fluides
FR361404
1906-06-22

Surchauffeur de vapeur pour locomotives
FR352810
1905-08-22

Apparatus for Heating Steam.
GB189801540
1898-01-19

An Improved Method for the Production of Motive Power by Heat and Fluid Pressure
GB190227627

1903-11-12

Improved Means of and Apparatus for Producing Motive Power
GB190226257
1901-11-29

An Improved Method of Producing Motive Power
GB190202489
1903-01-29

épuration bactériologique de l'eau
FR350264
1905-12-18

Dispositif générateur de vapeur
FR350080
1905-10-13

Perfectionnement aux moteurs à explosions
FR349940
1905-07-27

Mode de production de force motrice
FR349939
1905-07-27

Moyen de propulsion et de direction des navires
FR349935
1905-07-27

Nouveau mode d'application de la vapeur surchauffée
FR339220
1905-01-17

Dispositif pour la condensation de la vapeur d'échappement avec récupération de chaleur
FR337543
1904-04-13

Nouveaux verres d'optique
FR322234
1903-01-30

Improvements in Apparatus for Sterilising Water.
GB189705007
1898-02-19

An Improved Process for the Purification of Spirit.
GB189704089
1898-02-12

An Improved Process for Separating Heavy from Light Particles by Centrifugal Action.
GB189704087
1898-02-12

Improvements relating to the Purification of Water, and to Apparatus therefor.
GB189312941
1894-06-23

Improvements in Apparatus for Heating and Cooling Fluids by Fluids.
GB189706157
1898-03-0

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

---

rexresearch.com

Luc MONTAGNIER

DNA "Teleportatation"

Here's your local agent Dan Chao
FREE Quote!
GEICO Bay Area
Ad ✕

GEICO
LOCAL OFFICE
in Bay Area.
Contact us
Ad ✕

SEPHORA
OVERNIGHT SKIN SENSATIONS
SHOP NOW ▸
Ad ✕

Hypothesizes that that water formed a nano-structure & was able to emit an EM signal similar to the original DNA. The water nano-structures also were able to amplify into actual DNA strands.

See also : J. BENEVISTE / Digital Biology & Petr GAJAREV & V. POPONIN / DNA Reprogramming

---

http://www.pcworld.com/article/216767/dna_molecules_can_teleport_nobel_winner_says.html?tk=hp_new
Jan 16, 2011

DNA Molecules Can 'Teleport,' Nobel Winner Says
By John E Dunn, Techworld.com

A Nobel Prize winning biologist has ignited controversy after publishing details of an experiment in which a fragment of DNA appeared to 'teleport' or imprint itself between test tubes.

According to a team headed by Luc Montagnier, previously known for his work on HIV and AIDS, two test tubes, one of which contained a tiny piece of bacterial DNA, the other pure water, were surrounded by a weak electromagnetic field of 7Hz.

Eighteen hours later, after DNA amplification using a polymerase chain reaction, as if by magic the DNA was detectable in the test tube containing pure water.

Oddly, the original DNA sample had to be diluted many times over for the experiment to work, which might explain why the phenomenon has not been detected before, assuming that this is what has happened.

The phenomenon might be very loosely described as 'teleportation' except that the bases project or imprint themselves across space rather than simply moving from one place to another.

To be on the safe side, Montagnier then compared the results with controls in which the time limit was lowered, no electromagnetic field was present or was present but at lower frequencies, and in which both tubes contained pure water. On every one of these, he drew a blank.

The quantum effect - the imprinting of the DNA on the water - is not in itself the most contentious element of the experiment, so much as the relatively long timescales over which it appears to manifest itself. Quantum phenomena are assumed to show their faces in imperceptible fractions of a second and not seconds minutes and hours, and usually at very low temperatures approaching absolute zero.

Revealing a process through which biology might display the underlying 'quantumness' of nature at room temperature would be startling.

Montagnier's experiment will have to be repeated by others to have any hope of being taken seriously. So far, some scientists have been publically incredulous.

"It is hard to understand how the information can be stored within water over a timescale longer than picoseconds," said the Ruhr University in Bochum's Klaus Gerwert, quoted by New Scientist magazine, which broke the story (requires registration).

What does all of this mean? It could be that the propagation of life is able to make use of the quantum nature of reality to project itself in subtle ways, as has been hinted at in previous experiments. Alternatively, it could be that life itself is a complex projection of these quantum phenomena and utterly depends on them in ways not yet understood because they are incredibly hard to detect.

Speculatively, (and Montagnier doesn't directly suggest anything so unsubstantiated), it could also be the little-understood quantum properties of the water molecule and not just its more obvious chemical bonding properties that gives it such a central role in the bio-engineering of life-forms. Water might be a good medium in which DNA can copy itself using processes that hint at quantum entanglement and 'teleportation' (our term).

Montagnier's paper goes on to discuss the phenomenon he claims to have uncovered using 'quantum field theory' within the context of his personal interest, disease propagation.

---

http://arxiv.org/PS_cache/arxiv/pdf/1012/1012.5166v1.pdf

DNA Waves & Water

[PDF]

DNA waves and water

L. Montagnier[1,2], J. Aissa[2], E. Del Giudice[3], C. Lavallee[2], A. Tedeschi[4], and G. Vitiello[5]

[1] World Foundation for AIDS research and Prevention (UNESCO), Paris, France
[2] Nanetics Biotecnologies, S.A. 98 rue Albert Calmette, F78350 Jouy-en-Josas, France
[3] IIB, International Institute for Biophotonics, Neuss, Germany
[4] WHITE HB, Milano, Italy
[5] Dipartimento di Matematica e Informatica, Università di Salerno and
INFN, Gruppo Collegato Salerno, I-84100 Salerno, Italy

E-mail: `vitiello@sa.infn.it`

Abstract. Some bacterial and viral DNA sequences have been found to induce low frequency electromagnetic waves in high aqueous dilutions. This phenomenon appears to be triggered by the ambient electromagnetic background of very low frequency. We discuss this phenomenon in the framework of quantum field theory. A scheme able to account for the observations is proposed. The reported phenomenon could allow to develop highly sensitive detection systems for chronic bacterial and viral infections.

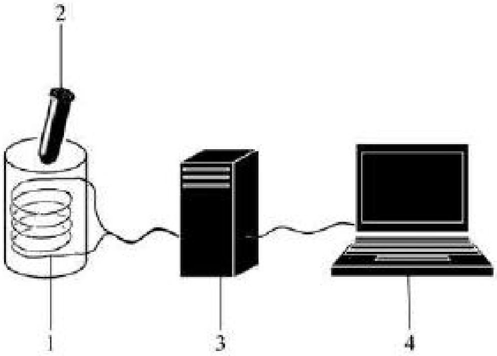

Figure 1. Device for the capture and analysis of em signals. (1) Coil made up of copper wire, impedance 300 Ohms. (2) Plastic stoppered tube containing 1 ml of the solution to be analyzed. (3) Amplifier. (4) Computer. From Ref. 1

Here is a brief summary of the laboratory observations, which are described in more detail in [1, 2, 3]:

1) Ultra Low Frequency Electromagnetic Waves (ULF $500-3000\ Hz$) were detected in certain dilutions of filtrates ($100\ nm$, $20\ nm$) from cultures of micro-organisms (virus, bacteria) or from the plasma of humans infected with the same agents (Fig. 2). Same results are obtained from their extracted DNA.

2) The electromagnetic signals (EMS) are not linearly correlated with the initial number of bacterial cells before their filtration. In one experiment the EMS were similar in a suspension of *E. coli* cells varying from 10^9 down to 10. It is an all or none phenomenon.

3) EMS are observed only in some high water dilutions of the filtrates. For example, from 10^{-9} to 10^{-18} dilutions in some preparations of *E. coli* filtrates.

4) In the case of *M. pirum*, an isolated single gene (adhesin, previously cloned and sequenced) could induce the EMS. As the gene was cloned in two fragments, each of the isolated fragments was able to generate EMS, suggesting that a short DNA sequence was sufficient to induce the signals. Similarly, a short HIV DNA sequence (104 base pair) is found to be sufficient to produce the EMS.

5) Some bacteria are not producing EMS: this is the case of probiotic bacteria such as *Lactobacillus* and also of some laboratory strains of *E. coli* used as cloning vector.

6) These studies have been extended to viruses, although not all the viral families have been explored. Similar EMS were detected from some exogenous retroviruses (HIV, FeLV), hepatitis viruses (HBV, HCV), and influenza A (in vitro cultures). In general, EMS are produced by 20 nm filtrates of viral suspensions or from the extracted DNA: a question remains for RNA viruses (HCV, influenza) as to whether the RNA from the mature viral particles is a source of EMS, or not. In the case of HIV, EMS are not produced by the RNA of viral particles, but rather are produced by the proviral DNA present in infected cells. In the case of bacteria, EMS are produced by 100 nm filtrates and not by 20 nm filtrates, indicating that the size of the structures producing EMS is ranging between 20 and 100 nm. This justifies the name of nanostructures. These studies are highly suggestive that one is dealing with nanostructures made of water. Highly purified water samples are used, although one cannot exclude the role of minimal traces of impurities. The EMS production by the nanostructures is resistant to: Rnase treatment, Dnase (while this will destroy the DNA at the origin of EMS), Protease (proteinase K), Detergent (SDS). However, they are sensitive to heat (over 70 $^\circ C$) and freezing (-80 $^\circ C$). This sensitivity is reduced when dealing with purified short DNA sequences. The technical conditions for EMS induction is summarized by the following list:

- Filtration: $\frac{450}{100}$ nm for bacterial DNA, $\frac{450}{20}$ nm for viral DNA
- High dilutions in water
- Mechanical agitation (Vortex) between each dilution
- Excitation by the electromagnetic background of extremely low frequency (ELF), starting very low at 7 Hz. The excitation is not induced when the system is shielded by a mu-metal cage.

The stimulation by the electromagnetic background of very low frequency is essential. The background is either produced from natural sources (the Schumann resonances [4] which start at 7.83 Hz) or from artificial sources.

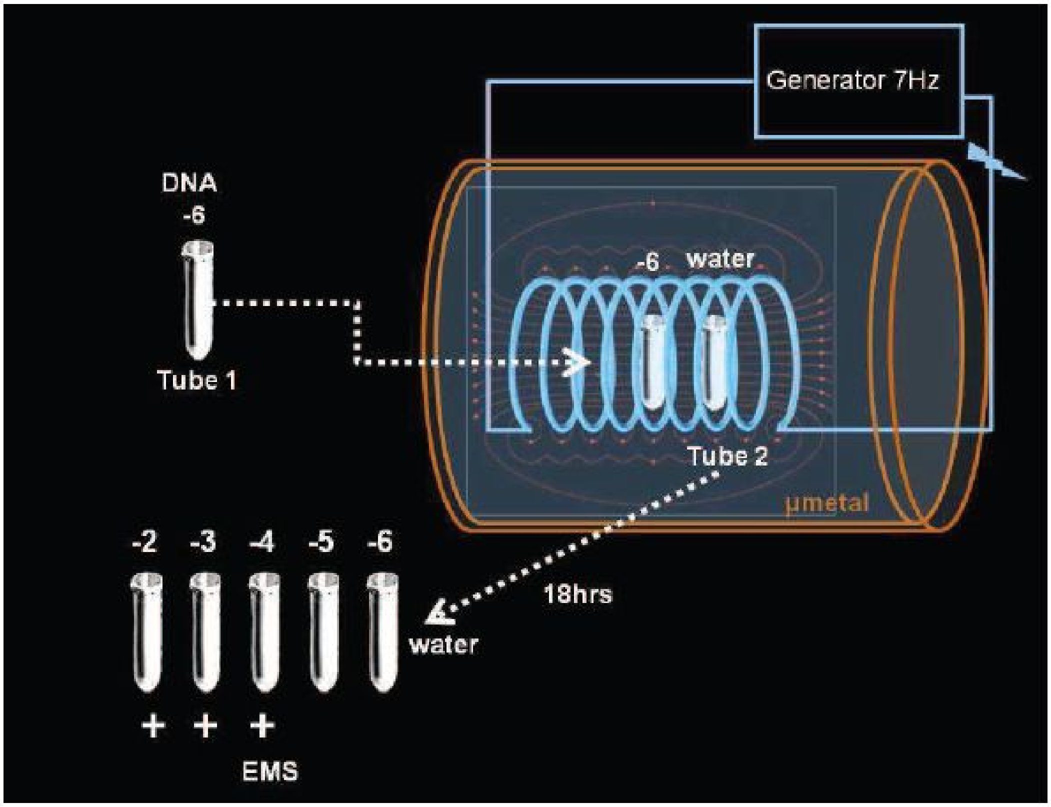

Figure 3. Transmission of DNA genetic information into water through electromagnetic waves. From Ref. 3

US8405379
System and method for the analysis of DNA sequences in biological fluids

---

WO2012142568
US2012024701
REMOTE TRANSMISSION OF ELECTROMAGNETIC SIGNALS INDUCING NANOSTRUCTURES AMPLIFIABLE INTO A SPECIFIC DNA SEQUENCE

Inventor: MONTAGNIER LUC // LAVALLEE CLAUDE

A general method for identifying both known and unknown DNA sequences at the origin of EMS, including DNA sequences in the plasma of patients suffering of chronic diseases such as Alzheimer, Parkinson, multiple sclerosis, rheumatoid arthritis, and other similar diseases, disorders and conditions. The invention is based on the discovery that: (1) The nanostructures induced by DNA sequences in water or other dipole solutions can faithfully reflect the information contained in these sequences at dilutions which do not contain anymore of that DNA, as evidenced by the fact that it can be retranscribed into the same DNA sequence by the polymerases and reagents used in classical polymerase chain reaction (PCR); (2) this information can be transmitted at a distance in water or other dipole solutions by EMS emitted by the nanostructures; and (3) EMS signatures of nanostructures containing this information can be detected, stored, transmitted, transduced and imprinted in water or other dipole solutions.

BACKGROUND OF THE INVENTION

[0004] 1. Field of the Invention

[0005] Induction, detection and transmission of electromagnetic signals (EMS) from self-replicating molecules like DNA.

Transduction of EMS from an EMS positive (EMS+) sample to a naïve, unsignalized sample. Methods for identifying a molecule like DNA in a sample by transducing its EMS signature to water, amplifying the signalized water to produce a DNA. Methods for detecting DNA associated with a condition, disorder or disease of incomplete or unknown etiology by inducing specific EMS emission from the sample at a particular frequency, signalizing a naïve sample with the emitted EMS, and detecting an EMS in the signalized water and/or amplifying the signalized water using a DNA amplification technique and analyzing the products of the amplification.

[0006] 2. Description of the Related Art

[0007] The inventors have previously described a method for selectively detecting DNA sequences of pathogenic microorganisms by their emission of low frequency electromagnetic waves (EMS) in water dilutions. U.S. application Ser. No. 12/560,772, filed Sep. 16, 2009, entitled "System and Method for the Analysis of DNA sequences in Biological Fluids" discloses a method for detecting electromagnetic waves derived from bacterial DNA, comprising extracting and purifying nucleic acids from a sample; diluting the extracted purified nucleic acids in an aqueous solvent; measuring a low frequency electromagnetic emission over time from the diluted extracted purified nucleic acids in an aqueous solvent; performing a signal analysis of the low frequency electromagnetic emission over time; and producing an output, based on the signal analysis, in dependence on the DNA in the sample. The products and procedures as well as other subject matter disclosed in this patent application are expressly incorporated by reference.

[0008] Methods for detecting some low electromagnetic frequency electromagnetic signals in diluted filtrates of the culture medium of certain bacteria and viruses, as well as in diluted plasma of patients infected by the same agents are disclosed by U.S. application Ser. No. 12/097,204, PCT/FR2007/001042, filed Jun. 22, 2007, and U.S. application Ser. No. 12/797,826, filed Jun. 10, 2010 each of which expressly incorporated by reference in their entirety. The electromagnetic signals (EMS) were believed to be produced by certain defined nanostructures induced by the microorganism, in high dilutions in water, after the microorganism had been removed by filtration.

[0009] Materials and methods for detecting replicating molecules such as DNA and methods for EMS detection as well as other subject matter pertinent to the present invention disclosed in these documents is incorporated by reference to the following documents:

[0010] U.S. Pat. No. 6,541,978, WO 00/17638 A (Digibio; Benveniste, Jacques; Guillonnet, Didier) 30 Mar. 2000 (2000-03-30).

[0011] U.S. Ser. No. 09/787,781, WO 00/17637 A (Digibio; Benveniste, Jacques; Guillonnet, Didier) 30 Mar. 2000 (2000-03-30);

[0012] U.S. Ser. No. 09/720,634, WO 00/01412 A (Digibio; Benveniste, Jacques; Guillonnet, Didier) 13 Jan. 2000 (2000-01-13);

[0013] FR 2,811,591 A (Digibio) 18 Jan. 2002 (2002-01-18);

[0014] FR 2,700,628 A (Benveniste Jacques) 22 Jul. 1994 (1994-07-22).

[0015] Benveniste J. et al: "Remote Detection Of Bacteria Using An Electromagnetic/Digital Procedure", Faseb Journal, Fed. Of American Soc. For Experimental Biology, Bethesda, Md., US, No. 5, Part 2, 15 Mar. 1999 (1999-03-15), page A852, XP008059562 ISSN: 0892-6638.

[0016] Thomas et al: "Activation Of Human Neutrophils By Electronically Transmitted Phorbol-Myristate Acetate" Medical Hypotheses, Eden Press, Penrith, US, vol. 54, no. 1, January 2000 (2000-01), pages 33-39, XP008002247, ISSN: 0306-9877;

[0017] Benveniste J. et al.: "Qed And Digital Biology" Rivista Di Biologia, Universita Degli Studi, Perugia, IT, vol. 97, no. 1, January 2004 (2004-01), pages 169-172, XP008059428 ISSN: 0035-6050;

[0018] Benveniste J. et al.: "A Simple And Fast Method For In Vivo Demonstration Of Electromagnetic Molecular Signaling (EMS) Via High Dilution Or Computer Recording" FASEB Journal, Fed. Of American Soc. For Experimental Biology, Bethesda, Md., US, vol. 13, no. 4, Part 1, 12 Mar. 1999 (1999-03-12), page A163, Abstr. No. 016209, XP008037356 ISSN: 0892-6638;

[0019] Benveniste J: "Biological effects of high dilutions and electromagnetic transmission of molecular signal" [Progress In Neonatology; 25th National Conference Of Neonatology] S. Karger Ag, P.O. Box, Allschwilerstrasse

10, CH-4009 Basel, Switzerland; S. Karger Ag, New York, N.Y., USA Series: Progres En Neonatologie (ISSN 0251-5601), 1995, pages 4-12, XP009070841; and 25ES Journees Nationales De Neonatologie; Paris, France; May 26-27, 1995 ISSN: 3-8055-6208-X;

[0020] Benveniste et al.: "Abstract 2392" FASEB Journal, Fed. Of American Soc. For Experimental Biology, Bethesda, Md., US, 22 Apr. 1998 (1998-04-22), page A412, XP009070843 ISSN: 0892-6638;

[0021] Benveniste et al.: "Abstract 2304" FASEB Journal, Fed. Of American Soc. For Experimental Biology, Bethesda, Md., US, 28 Apr. 1994 (1994-04-28), page A398, XP009070844 ISSN: 0892-6638; and

[0022] U.S. Pat. Nos. 7,412,340, 7,081,747, 6,995,558, and 6,952,652.

[0023] In some instances, it was demonstrated that the EMS could originate from specific genes or even from some fragmented DNA sequences. This was discovered to be the case for the adhesin gene of Mycoplasma pirum (U.S. Ser. No. 12/097,204, filed Dec. 14, 2006) and of the LTR (Long terminal repeat), nef and pol genes of Human Immunodeficiency Virus (HIV) (U.S. 61/186,610, filed Jun. 12, 2009 & U.S. Ser. No. 12/797,826, filed Jun. 10, 2010). However, for many microbial agents or diseases of unknown origin or etiology this identification was not possible. Consequently, the inventor developed new methods, disclosed herein for detecting and identifying biological molecules, specifically DNA or other nucleic acids, associated with these other disease or disorders.

BRIEF SUMMARY OF THE INVENTION

[0024] There are several nonlimiting aspects to the invention.

[0025] (1) A method for producing a solution, such an aqueous solution like water that contains nanostructures that characterize a molecule like DNA. This method involves dilution, usually serial dilution, of a sample containing DNA and agitation of the sample between dilutions to produce the water nanostructures.

[0026] (2) Measuring EMS characteristic of a molecule like DNA or of its nanostructure in an originating sample and transducing this signal into a second receiving sample, usually water that does not emit the EMS signal. This is performed without contacting the originating sample and the receiving sample.

[0027] (3) Electronic transmission of a detected or recorded EMS signal to a remove location and optionally imprinting it on a naïve sample and/or recovering DNA or other replicating molecule from the imprinted naïve sample.

[0028] (4) Detecting DNA or DNA like molecules in a sample suspected of containing a particular agent, like HIV or Borellia.

[0029] (5) Identifying DNA or similar molecules present in an unknown sample, such as from a sample from a subject having a disease of unknown etiology.

[0030] (6) Devices that detect, induce, transduce or transmit EMS signals.

BRIEF DESCRIPTION OF THE DRAWINGS

[0031] FIG. 1 illustration of apparatus and method for EMS signal transduction. Tube 1 contains a sample of DNA dilution positive for EMS. Tube 2 initially contains unsignalized or naïve water. After exposure inside coil to 7 Hz excitation signal, naïve sample converts and emits EMS when diluted up to D4 (10<-4>). D-4 LTR HIV DNA (104 bp) 7 Hz, 18 Hrs and then PCR (35 cycles) from D2 to D15 after filtration 450 and 20 nM; DW: Distilled Water; FD2: Dilution 10<-2 >after filtration at 450 nM and 20 nM.

[0032] FIG. 2 Detection by PCR of HIV1 LTR transduction in water.

[0033] FIG. 2A: HIV1 LTR DNA D6 (EMS positive) dilution was used as emitter using excitation frequency of 7 Hz during 18 hours in the apparatus described in FIG. 1 and placed close to the water receiver Tube 2. Like the latter, it was then diluted at 10<-2>, refiltered by 450 nM and 20 nM filters and diluted to 10<-15>. Each dilution was then amplified by PCR 35 cycles. Note the DNA bands detected at dilutions D2, (FD2), D3, D4, and D5.

[0034] FIG. 2B shows transmission in water of D6 dilution of LTR HIV DNA (104 bp). Method was performed using excitation frequency 7 Hz, an 18 hr exposure followed by 35 cycles of PCR from D-2 to D-15 after 450 nM and 20 nM filtration. DW denotes distilled water control. FD2-FD15, dilution to 10<-2>-10<-15>. Transmission in

water of D-4 LTR HIV DNA (104 bp) 7 Hz, 18 Hrs and then PCR (35 cycles) from D-2 to D-15 after filtration 450 and 20 nM. Note: DNA band formation is up to D-8.

[0035] FIG. 3. Illustration of method to generally identify an unknown DNA sample. DNA in plasma sample is induced to emit EMS and the EMS signal is transduced to a separate sample of water to produce signalized water. Water signalized by EMS is serially diluted and PCR is performed using random tag primers producing DNA. The sequence of the DNA is determined and can be compared to known DNA sequences to identify the DNA in the unknown sample. Example 3 describes such a method.

[0036] FIG. 4. Detection of unknown DNA sequences from a patient plasma DNA sample. DNA was extracted from the plasma of a patient suffering from chronic Lyme disease. A D9 (10<-9>) EMS positive dilution of the original DNA sample was transduced into water by excitation at 7 Hz for 18 hrs. PCR was performed on dilutions of the receiving water sample. FIG. 4 shows agarose gel electrophoresis of the transduced DNA obtained after PCR with Tag8N primers followed by a second PCR with the Tag primers only. Three DNA bands were observed. As shown at the left, results obtained when the tube of D9 DNA and the tube of water are placed side-by-side. At right, results obtained when the two tubes were placed at a distance of 4 cm from each other during the 7 Hz excitation. Dw denotes control, naïve, unsignalized water. Dw vor: denotes control naïve, unsignalized water agitated with a vortex. D0: water that was transduced but not diluted. D2 NF: same as D0 but diluted by 1:100 (D2). D2 same as D2 NF, but filtered. D3, D4, D5 represent further serial dilutions of D2 to factors of 1:1,000 (D3); 1:10,000 (D4) and 1:100,000 (D5). All serial dilutions were vortexed between each 1:10 dilution.

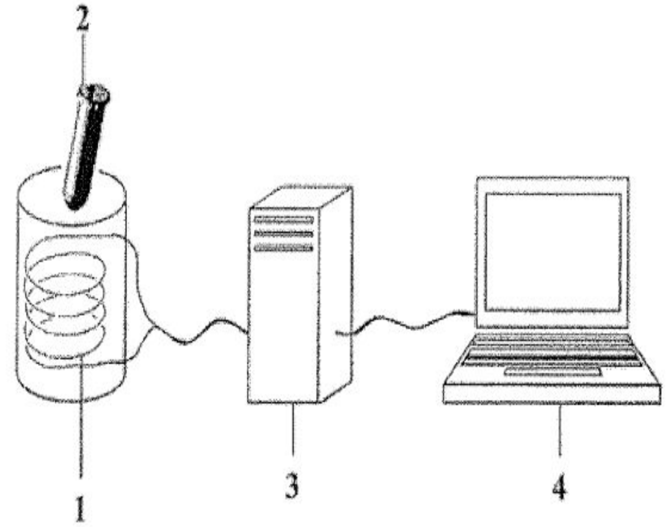

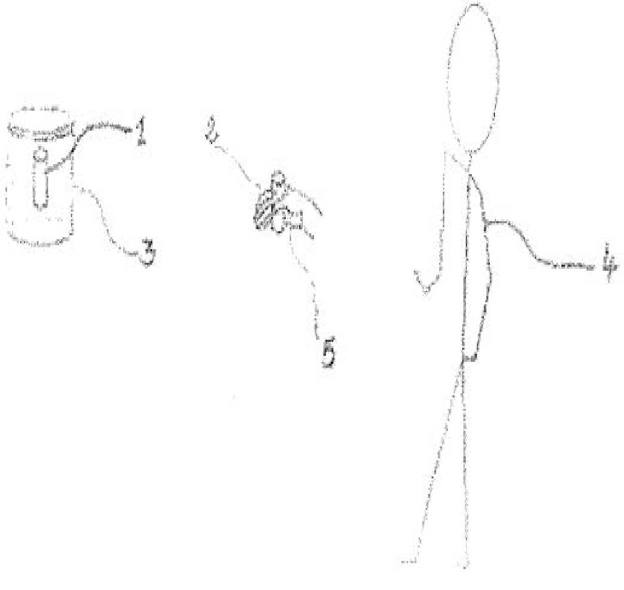

Figure 3

TUBE T3 (BACKGROUND NOISE)

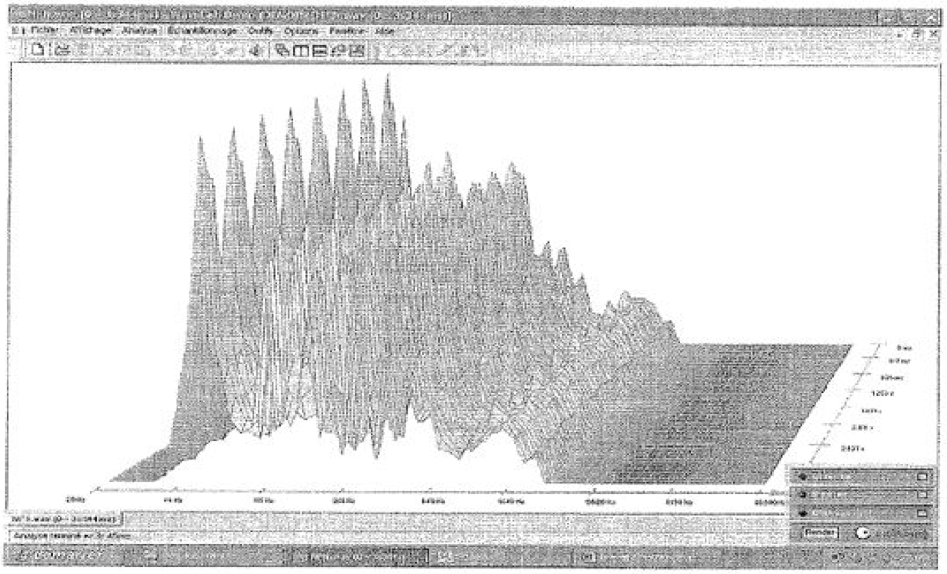
Fichier
Affichage
Analyse
Echantillonnage
Outils
Options
Fenêtre
Aide
Rendre

Figure 4

Tube 1

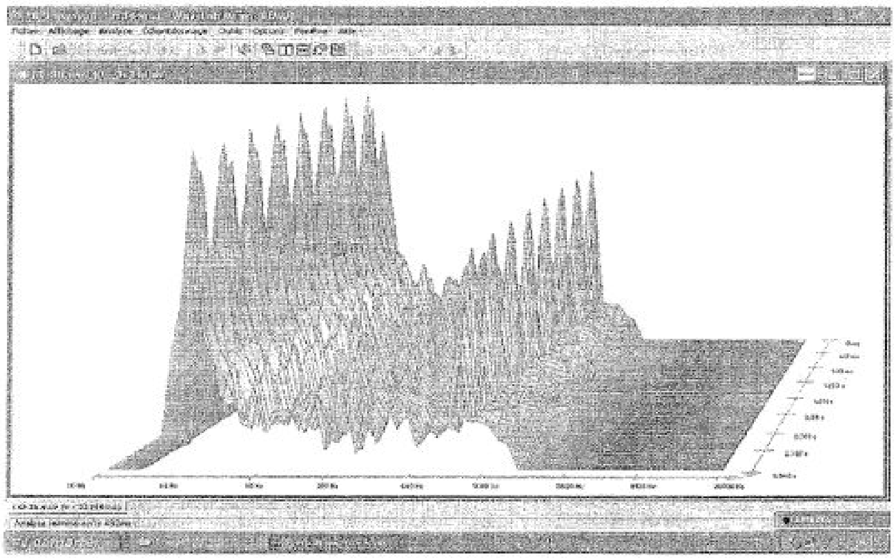
Fichier
Affichage
Analyse
Echantillonnage
Outils
Options
Fenêtre
Aide

Tube 2

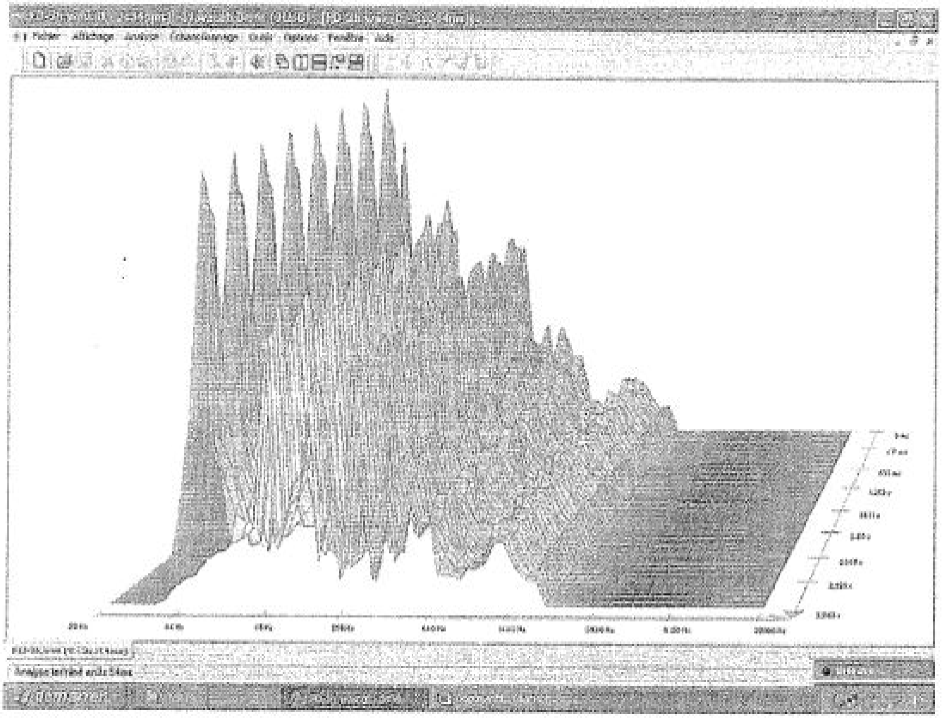
Fichier Affichage Analyse Echantillonnage Outils Options Fenêtre Aide
20 Hz
20000 Hz

DETAILED DESCRIPTION OF THE INVENTION

[0037] Definitions:

[0038] Nucleic acid: Includes single stranded, double stranded DNA, and RNA as well as modified polynucleotide sequences. Biological samples containing DNA associated with a disease or disorder are generally isolated or recovered in double stranded form.

[0039] Self-replicating molecule: A molecule, such as DNA, that under appropriate conditions, can reproduce the information content of its primary, second, tertiary or quaternary structure. For example, a DNA molecule can replicate itself in the presence of the appropriate enzymes, primers and nucleotides.

[0040] DNA Amplification: Methods for amplifying nucleic acids are known. Conventional methods including polymerase chain reaction (PCR) are known and are also incorporated by reference to Current Protocols in Molecular Biology (updated Apr. 5, 2010), Print ISSN: 1934-3639; Online ISSN: 1934-3647.

[0041] Nanostructures: These structures of water are induced by biological molecules like nucleic acids such as single stranded or double stranded DNA. While not being bound to any particular theory, according to the physical theory of diphasic water, filtration and mechanical agitation (succussion) are believed to induce in water a low energy potential favoring the formation of quantum coherent domains. These domains will become replicas of a DNA molecule and vibrate by resonance when properly diluted and excited; see Del Guidice, et al., Water as a Free Electric Dipole Laser, Phys. Rev. Lett. 61, 1085-1088 (1988). Hydrogen bonding networks in liquid water, such as those described by Cowan, et al., Nature 434 (7030): 199-202 (2005) have not been associated with nanostructures.

[0042] Serial Dilutions: Serial dilution is a well-known technique and involves the stepwise dilution of a substance, such as DNA, in a solvent, such as water, saline solution, aqueous buffer, or an aqueous alcohol solution. Generally, serial dilutions as performed herein are stepwise dilutions by a factor of 10, or dilution of 1 part of a more concentrated solution in 9 parts of a solvent.

[0043] EMS: Electromagnetic signal. EMS in the context of the methods herein generally involves those having frequencies ranging from 0 Hz to 20,000 Hz as well as all intermediate subranges and values. Components of the ambient electromagnetic field include Schumann resonances which represent a set of spectrum peaks in the extremely low frequency (ELF) portion of the Earth's electromagnetic field spectrum. Schumann resonances are global electromagnetic resonances excited by lightning discharges in the cavity formed by the Earth's surface and the ionosphere and are the principal background in the electromagnetic spectrum between 3 and 69 Hz. A representative Schumann resonance peak occurs in the Earth's electromagnetic spectrum and an ELF of about 7.83 Hz. By comparison, 60 Hz cycling of electricity is used in North America and 50 Hz elsewhere in the world.

[0044] EMS detection. Any suitable means for interrogating a sample and measuring its EMS may be employed. Exemplary systems, methods, and apparatuses for this purpose are disclosed by Butters, et al., WO 03/083439 A2, and are incorporated by reference to this document. Generally, these procedures will involve placing a sample into a container having electromagnetic and magnetic shielding, a source of Gaussian noise for injection in to the sample, a detector for detecting an electromagnetic time-domain signal composed of sample source radiation superimposed on the injected Gaussian noise, and a storage device for storing the time-domain signal and a time-domain signal separately detected from the same of a similar sample.

[0045] EMS Signature: The EMS characteristic of a particular biological molecule or a time domain signal associated with a material of interest. EMS signatures for various biological molecules are disclosed by U.S. Ser. No. 12/797,826, filed Jun. 10, 2010. Such EMS signatures as well as methods for producing samples suitable for EMS detection and methods for detecting an EMS signature are incorporated by reference to this patent application.

[0046] An EMS Signature of a particular molecule can be represented by a characteristic electromagnetic time domain signal. An EMS Signature may be recorded and replayed, undergo signal transformation or processing, or be transmitted.

[0047] Excitation Frequency: A frequency used to excite a sample in which an EMS signature has been detected and induce an EMS signature in a sample previously devoid of the EMS signature, e.g., pure water. These frequencies include those of 7 Hz or above, e.g., 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 30, 40, 45, 50, 55, 60, 65, 70 or more.

[0048] Originating Sample: A biological sample that contains an EMS signature, such as one characteristic of one or more biomolecules. An example would be a sample containing an EMS signature characteristic of DNA derived from human immunodeficiency virus.

[0049] Receiving or Signalized Sample: A sample, such as water or another aqueous buffer or dipole that has acquired or been imprinted with a nanostructure corresponding to a biological molecule, such as DNA. Methods for producing signalized water by serial dilution and agitation in water or in an aqueous solvent are disclosed herein.

[0050] Pathogenic Disease: Disease caused by or associated with a pathogen, such as a pathogenic parasite, yeast or fungus, bacterium, virus or infectious protein, such as a prion. Examples include parasitic diseases such as malaria or trypanosomiasis, fungal diseases, such as infections caused by or associated with Aspergillus, Candida, Histoplasma, Pneumocystis, Cryptococcus, Stachybotrys (black mold), bacterial infections such as Lyme Disease, sexually transmitted bacterial infections, tuberculosis, viral infections, including HIV infection, herpes virus infection, or hepatitis, and prion associated diseases such as Creutzfeldt-Jakob disease and so-called Mad Cow disease.

[0051] Autoimmune Disease, Degenerative Disease, Disorders or Conditions: These diseases, disorders or conditions may or may not have been previously associated with a particular biological molecule, such as a particular DNA molecule or its corresponding water nanostructure. Examples include allergic conditions, multiple sclerosis, rheumatoid arthritis, disorders associated with transplantation or replacement of body parts, Alzheimer's disease, Parkinson's disease and other diseases or disorders of unknown or incomplete etiology, such as Chronic Fatigue Syndrome, Gulf War Syndrome, or with exposure to particular biological, chemical or physical agents or with the sequela of such exposure.

[0052] Representative embodiments of the invention are described below.

[0053] (i) Originating and Signalized Samples.

[0054] Test samples used to produce an EMS will contain DNA or other replicating biological molecules that can form nanostructures or can be naïve samples signalized by EMS transduction to emit EMS or contain nanostructures representative of the DNA or other molecule. Representative test samples include blood, plasma, serum, CSF, joint fluid, saliva, mucous, semen, vaginal fluid, sweat, urine, and feces. Tissue samples and samples from other sources, including laboratory or hospital sources, foods, drinks and potable water may be used. These may be diagnostic samples, such as those obtained from a subject known to have or suspected of having a particular conditions, disorder or disease like AIDS or Lyme disease. Alternatively, they may be obtained from subjects having or suspected of having a condition, disorder or disease of unknown etiology, such as a parasitic or fungal disease or disorder, bacterial disease or disorder viral disease or disorder, an autoimmune disease, disorder or condition, diseases such as Alzheimer's Disease or Parkinson's Disease.

[0055] To produce a sample that emits detectable EMS, a test sample undergoes dilution, usually serial dilution, and agitation preferably between each serial dilution. A test sample is usually diluted by a factor of 10<3>, 10<4>, 10<5>, 10<6>, 10<7>, 10<8>, 10<9 >10<10>, 10<11>, 10<12>, 10<13 >or more. Though any intervening factor of dilution or other degrees of dilution that produce detectable EMS may also be used. The beginning concentration of a nucleic acid in a sample prior to dilution generally ranges from 1 ng/ml to 4 ng/ml.

[0056] Solutions for dilution and agitation as well as for containing an originating or receiving sample are preferably water, but other aqueous or dipolar solutions may be employed so long as they can provide nanostructures representative of DNA or other replicating molecules or induce detectable EMS when used. Examples of solutions include water, or other aqueous solutions, such as normal saline, phosphate buffered saline, physiologically acceptable aqueous solutions, buffered aqueous solutions, or alcohol and water mixtures, including 10, 20, 30, 40, 50, 60 and 70% or more of ethanol or other alcohol solutions or other solvents selected on a basis of their relevant properties depending on the molecule to be tested, may be employed in the methods described herein.

[0057] In some applications, control samples are required. The type of control sample may be selected by one of skill in the art depending on the particular application but in general will not emit the EMS signature of the molecule of interest or contain nanostructures corresponding to it. Often, such controls will constitute pure, unsignalized water, distilled water or pyrolyzed water or other solutions known to be nucleic acid free.

[0058] Signalized samples or solutions producing an EMS signature should not be boiled, heated or frozen for long periods of time so as to preserve the EMS signatures or nanostructures they contain. Preferably, these samples or solutions should be stored above freezing and less than 40[deg.] C.

[0059] Various forms and time periods for agitation are contemplated and are incorporated by reference to the documents mentioned above. Vortexing for a period of 15 seconds between serial dilutions is one representative method for producing a sample emitting detectable EMS.

[0060] (ii) EMS Transduction. The invention also relates to a method for producing an EMS signature in an aqueous buffer comprising placing an originating (EMS+) sample in an aqueous buffer and a receiving sample not having the EMS signature next to each other inside of an electromagnetically shielded container, applying an electromagnetic field for a time and under conditions sufficient to transfer the EMS signature from the originating sample to the receiving sample. The electromagnetic field is generally applied by a coil, such as a copper coil, located inside of an electromagnetically shielded container. Coils made of other electrically conducting metals or alloys may be employed or other devices that produce similar electromagnetic flux. The electromagnetic field can be applied to the sample for a time period ranging sufficient to produce an EMS signature, for example, from 12 to 24 hrs although other suitable time periods may be selected based on the nature of the sample, the sample dilution and the physical characteristics of the apparatus. Exposure time is chosen based on the amount of time required for transfer to occur. Some representative times include >0, 1, 2, 3, 4, 4-8, 8-12, 12-18, 18-24 and 24-48 hrs or longer. Signalized samples produced by this method as well as nucleic acids like DNA amplified from a signalized sample are also contemplated. Alternatively, an EMS signature may be imprinted in water or another aqueous buffer by contacting the one or more receiving samples with a recorded or transmitted and optionally amplified EMS signature previously obtained from an originating sample in an aqueous buffer having an EMS signature, for a time and under conditions sufficient to imprint the recorded or transmitted EMS signature of the originating sample onto the one or more receiving samples. Imprinting may be performed using means for applying an electromagnetic field, for example using a device, such as a copper coil or solenoid coil, optionally located inside of an electromagnetically shielded container. The electromagnetic field is applied to the sample for a time period sufficient to produce an EMS signature in the sample, for example for a period of 1 to 24 hrs. Other suitable time periods may be selected based on the nature of the sample, the sample dilution and the physical characteristics of the device or other means for applying the electromagnetic field. Signalized samples produced by this method as well as nucleic acids like DNA amplified from a signalized sample are also contemplated.

[0061] (iii) EMS Recording/Transmission. EMS signals once measured may be recorded on a tangible medium, such as a computer hard drive, a flash drive, DVD, or CD or other known media. They may be transmitted electronically, for example, over the internet, or by any other means that preserves the signal integrity. Recorded or received signals can be amplified and used to transduce EMS into a naïve solution as described above. This aspect of the invention can involve the recording, transducing, storing, and/or transmission of an EMS signature of a nucleic acid, such as that produced after serial dilution of a signalized sample. An EMS signature may be recorded by a suitable electronic device, such as a recorder, computer or computer network. The recorded EMS signature may undergo signal processing or signal transformation for example into a digital or analog signal, be transmitted by a communications device, such as via radio, telephone, modem, or Internet transmission to a receiver, such as a receiving computer, anywhere in the world.

[0062] A stored or transmitted EMS signature is then reconstituted and/or amplified and contacted with a receiving sample to imprint it with the EMS signature and produce nanostructures in the water or dipole solution of the receiving sample. Such a signal may be amplified prior to or after transmission, for example, using a commercial amplifier (e.g., Conrad). The electrical output from the amplifier containing the EMS signature is then applied to an electrically conducting coil (e.g., of copper wire) as described herein in which a plastic tube of pure non-signalized water or other dipole solution has been inserted for a time sufficient for imprinting of the EMS signature, generally for a period of at least one hour.

[0063] The production of EMS is then verified in water dilutions of the signalized water or dipole solution. The positive dilutions can be used for retrieving the DNA by PCR as described above. The DNA is then amplified by cloning and its sequence determined to be 98-100% identical to the initial DNA. This development will be useful for remote diagnosis or use in other telemedicine procedures or protocols.

[0064] The inventors previously discovered that an electromagnetic signal of low frequency (EMS) induced in a water dilution by the DNA of some kinds of bacteria and viruses can be transmitted at a distance into a naive or unsignalized water, aqueous medium or other dipole solution. It has also been discovered that such an EMS corresponding to a particular biomolecule like DNA (i.e., an EMS signature of a particular molecule), can be recorded. This involves recording EMS from DNA fragments obtained by PCR (polymerase chain reaction) with sequence specific primers in an electromagnetic coil. The resulting amplified current is connected to a computer and stored as a file, such as an analog or digital file (e.g., a digital sound file). The recorded EMS can then undergo signal processing, for example a digital sound file can be processed using computer software for storage, transmission, or use.

[0065] DNA may be reconstituted from its EMS signature. For example, the recorded or remotely transmitted EMS signature of a DNA molecule is input into a soundcard and the output from the soundcard is linked to an amplifier. Amplifier output is connected to a transducer solenoid into which an unsignalized water sample is placed. After a certain time, depending on the type of EMS signature, its intensity and the exposure time, the unsignalized water becomes signalized. In other words, the unsignalized water has memorized the EMS signature of the originating DNA molecule. By use of PCR the originating DNA molecule may be retrieved from the water signalized with its EMS signature. Verification of retrieval of the originating DNA sequence from the signalized water or verification of the fidelity of its reproduction can be verified by DNA sequencing.

[0066] Alternatively, prior to retrieval and synthesis of the DNA molecule by PCR, the signalization of the receiving sample with a DNA EMS signature may be determined by detecting the EMS emissions of the signalized sample using dilutions of the signalized water as previously described, e.g., by the device used to record the originating DNA sample's EMS signature in the first place. Only EMS positive dilutions will yield the DNA sequence. The procedure allows the transmission of DNA EMS signatures of medical interest as well as the remote retrieval of the corresponding originating DNA. Such transmission may be made by a medium of choice, for example, a digital signal may be transmitted over the internet or by sending USB keys (e.g., flashdrives) to remote laboratories or medical units.

[0067] (iv) Detection of a Known Nucleic Acid Sequence. Specific molecules known or suspected to be contained in a test sample may be screened using the methods described above. A test sample is diluted and agitated to produce an EMS+ sample and a nucleic acid amplification using specific known primers for the nucleic acid sequence of interest is performed. The test sample may be a sample produced by dilution and agitation or may be produced by tranduction of EMS into a naïve sample. An EMS+ test sample is incubated with primers for a specific nucleic acid sequence and the nucleic acid product by PCR amplification, usually DNA, is recovered. The recovered amplification products may be assayed indicate the presence of the particular nucleic acid in the test sample.

[0068] (v) Identification of an Unknown Nucleic Acid.

[0069] Another embodiment of the invention involves detecting a nucleic acid or nanostructures associated with an unknown nucleic acid in a test sample comprising amplifying a nucleic acid in a test sample using random nucleotide sequence or polynucleotides or primers; diluting and agitating during dilution the amplified nucleic acids in an aqueous solvent; measuring over time a low frequency electromagnetic emission from the diluted amplified nucleic acids; and optionally (i) identifying an EMS signature for amplified nucleic acid or its associated nanostructures by comparing the EMS of the test sample to the EMS of a control sample, and optionally (ii) comparing the results to relevant standard EMS signature(s). This method may further comprise performing a signal analysis of the low frequency electromagnetic emission over time, and/or producing an output, based on the signal analysis. This method may detect a biological molecule, such as a nucleic acid like DNA in a test sample and/or may detect a nanostructure derived from or associated with a nucleic acid such as DNA in the test sample. A suitable dilution of the test sample is selected for use within this method, for example, the test sample can be diluted by a factor of at least 10<4>, 10<5>, 10<6>, 10<7>, 10<8>, or 10<9>.

[0070] The test sample will usually be obtained from subject suffering from or at risk of developing a particular disease, disorder or condition. For example, the test sample can be obtained from a subject having or suspected of having a parasitic or fungal disease or disorder, a subject having or suspected of having a bacterial disease or disorder, a subject having or suspected of having viral disease or disorder, from a subject having or suspected of having had an autoimmune disorder, a subject having or suspected of having Alzheimer's Disease or Parkinson's Disease or any other neurological disease, a subject having or suspected to have a genetic disease or a gene alteration, or a subject having a disease, disorder or condition of unknown or incomplete etiology in comparison with a noninfected subject. For instance, an EMS signature of an HIV gene sequence, such as that of nef or pol, may be detected in a sample in comparison to a sample not containing the HIV gene sequence. Verification of the presence of a gene sequence in a sample may be made by PCR.

[0071] (vi) Devices. Various devices for use in conjunction with the different aspects of the invention are also disclosed. These include:

[0072] A device for producing an EMS signature in an aqueous buffer comprising at least two containers, at least one for an EMS originating sample and at least one for an EMS receiving sample, an electrically conducting coil that can emit a variable frequency ranging from 1 to 20,000 Hz, optionally connected to an external generator of alternating current having a variable frequency from 1 to 20,000 Hz, means for electromagnetic shielding the at least two containers and the electrically conducting coil.

[0073] A device or other means for transmitting at a distance EMS emitted by a biological sample or by nanostructures contained in a sample is also contemplated. Such a device will contain at least two containers, at least one to contain a sample determined to produce EMS characteristic of a DNA or a similar molecule in a first tube (originating sample), and another tube (receiving sample) to receive emitted EMS and contain signalized water produced. The device will contain an electrically conducting coil linked to an external generator of alternating current having a variable frequency from 1 to 20,000 Hz. The device will have shielding means, such as mu metal >=1 mm in thickness, capable of isolating external ambient electromagnetic signals or noise, enclosing a space into which will accommodate the coil and the containers. Any suitable material may be used to make the coil and the elements and design of the coil are selected based on the size of the samples, shielding, and other elements of the apparatus. One example of a coil is a copper coil with the following characteristics: bobbin with internal diameter 50 mm, length 80 mm, R=3.6 ohms, 3 layers of 112 turns of copper wire, field on the axis to the centre 44 Oe/A, and on the edge 25 Oe/A. An example of shielding is a cylinder of [mu] metal having a minimal thickness of 1 mm, closed at both ends in a manner that completely isolates the enclosed containers and coil from the external ambient electromagnetic noise.

[0074] The following Examples describe particular embodiments of the invention, but the invention is not limited to what is described in these Examples.

EXAMPLE 1

Production of Samples Containing an EMS Signature Characteristic of HIV DNA

[0075] Step A:

[0076] Synthesis of DNA by PCR

[0077] A particular DNA sequence is first synthesized by polymerase chain reaction (PCR) on a DNA template, for example, a region of the LTR sequence present in the viral DNA extracted from the plasma of a HIV infected patient or obtained from a purified infectious DNA clone of HIV1 Lai, is amplified by PCR and nested PCR with respectively LTR-derived outer and inner primers.

[0078] Those were chosen to pick up some conserved regions of the LTR, given to several subtypes of HIV1. This amplified DNA was sequenced and found 100% identical to the known sequence of the prototype strain of HIV1 subtype B, HIV1 LAI (3). The resulting amplicon was determined to be 488 bp long and the nested-PCR amplicon to be 104 bp long.

[0079] Filtration and Dilution: A sample of each amplicon is prepared at a concentration of 2 ng/ml in a final volume of 1 ml of pure water that had been previously filtered through a sterile 450 nM Millipore (Millex) filter and then to a 20 nM filter (Whatman, Anotop) to eliminate any contamination by viruses or bacteria. All manipulations are done under sterile atmosphere in a biological safety cabinet.

[0080] The DNA solution is diluted one in 100 (10<-2>) in 2 ml of water and filtered through a 450 nM Millex filter (Millipore) and filtered again through an Anotop filter of porosity size 20 nM (Whatman).

[0081] The resulting DNA filtrate (there is practically no DNA loss through filtration, as the DNA molecules do not bind to the filters), is then diluted serially 1 in 10 (0.1 ml in 0.9 ml of water in an Eppendorf sterile tube of 2 ml from 10<-2 >to 10<-15>.

[0082] A strong vortex agitation was performed at each dilution step for 15 seconds.

[0083] Each dilution in its stoppered plastic tube was placed on a coil under the ambient electromagnetic background at room temperature for 6 seconds; the resulting electric current is amplified 500 times and analyzed in a Sony laptop computer with specific software as previously described. The EMS positive vibrating dilutions (usually between 10<-4 >to 10<-8>) were detected not only by new peaks of frequency, but also quantitatively by the difference in amplitude and intensity of the signals measured in the software, as compared to the same parameters given by the background noise.

[0084] Table 1 shows the role of excitation frequency in inducing EMS from DNA into water. A fragment of LTR DNA (Tar region, 104 base pairs) was amplified by PCR with specific primers from the entire genomic HIV1 LAI DNA cloned in a plasmid (pLAI2). The fragment was purified by electrophoresis on an agarose gel; the DNA band was then cut and extracted with a Qiagen kit. Time of exposure DNA tube and water tube to the exciting frequency was 18 hrs.

[0000]

TABLE 1

Content	Frequency (Hz)	EMS	% over noise	Positive dilutions
LTR DNA 104 bp	2	+	33.3	D6-> D8
Water		-	1.2	
DNA	3	+	39.6	D4-> D7
Water		-	0.5	
DNA	4	+	43.9	D5-> D8
Water		-	1.5	
DNA	5	+	41.6	D5-> D8
Water		-	0	
DNA	6	+	33.5	D5-> D8
Water		-	1	
DNA	7	+	40	D6-> D8
Water		+	43.9	D5-> D8

[0085] Step B:

[0086] Producing a Signalized Sample from the Originating Sample

[0087] Tube 1 containing one of the dilutions found positive for EMS in step A (10<-6>) was placed in the vicinity of an identical tube 2 that had been previously filled with 1 ml of pure water under a separate safety cabinet different from the one utilized in step A for the DNA manipulation. Both tubes were placed inside a copper coil with the following characteristics: bobin with internal diameter 50 mm, length 80 mm, R=3.6 ohms, 3 layers of 112 turns of copper wire, field on the axis to the centre 44 Oe/A, and on the edge 25 Oe/A, linked to an external generator of alternate electric current of variable frequency from 1 to 20,000 Hz.

[0088] The tubes and the coil were enclosed in a cylinder of thick (1 mm) [mu]metal closed at both ends in order to isolate the system from the external ambient electromagnetic noise. A current intensity of 100 mA was applied to the coil, so that no significant heat was generated inside the cylinder.

[0089] The tubes were kept 18 Hrs at room temperature in an oscillating magnetic field strength of 25 Oe/A generated by the coil system. Afterwards, the signalized water of tube 2 is filtered on 450 nM and 20 nM filters and diluted from 10<-2 >to 10<-15>. As a control, the tube 1 was also filtered and diluted in the same way. EMS analysis revealed positive dilutions for EMS, starting at 10<-2 >which is explained if one takes into account that the emitter tube 1 was already at the 10<-6 >dilution (FIG. 1). As shown in Table 1 a minimal frequency of 7 Hz was found necessary and sufficient to induce the EMS in the naïve, unsignalized water filled tube 2. However, the intensity of the EMS signals was sometimes reduced by comparison to those found in tube 1. To determine conditions suitable for EMS transduction, the inventors also varied different parameters of the process. It was determined that the following conditions suppressed EMS emission from naïve tube 2 (receiving sample or sample to be signalized).

[0090] Time of exposure of the two tubes less than 16-18 hrs (Table 2).

[0091] No coil.

[0092] Generator of magnetic field turned off.

[0093] Frequency of excitation<6 Hz.

[0094] No use of DNA in tube 1.

[0095] Tube 2 frozen at -80[deg.] C. overnight and defrosted before recording the EMS.

[0096] Tube 2 heated at 95[deg.] C. for 60 minutes after the overnight exposure.

[0097] Based on the results in Table 1 and on testing of the process conditions and parameters it was concluded that excitation of tube 1 by a magnetic field of low frequency and of very low intensity has allowed the water nanostructures generated by the DNA fragment contained in this tube to be transmitted via waves to tube 2.

[0098] Step C:

[0099] Reconstitution by PCR of the LTR DNA from the Nanostructures in the Receiving or Signalized Sample.

[0100] A sample volume (5 [mu]l) of tube 2-signalized water was added to 45 [mu]l of an amplification mixture in a propylene 200 [mu]l PCR tube (Eppendorf).

[0101] The amplification mixture was composed of (buffer composition) 0.2 mM dNTP's, 10 [mu]M of each specific HIV-1 LTR primer containing the ingredients for synthesizing DNA, either from a positive dilution for EMS or in a lesser dilution, starting with 10<-2 >down to 10<-10>: and using 1 unit of Taq DNA polymerase.

[0102] Once the first cDNA strand is synthesized, cycling of denaturation, annealing and polymerization steps are performed as usually used for the PCR amplification.

[0103] The reaction (35 cycles, T[deg.] annealing 56[deg.] C.) yielded a DNA band of the size (in electrophoresis migration in agarose 1.5%) of the expected 104 bp sequence. This amplicon was then cloned in a bacterial plasmid (Topo Cloning, Invitrogen) which was used to transform bacterial competent cells. Plasmid clones were purified from isolated bacterial transformants and screened for the presence of the 104 bp insert by EcoRI digestion. Positive plasmid clones are then sequenced and the sequence of the insert shown to be 98% to 100% identical (difference of 2 nucleotides) to the original DNA of tube 1.

[0104] The first step of DNA synthesis using the nanostructures as templates can also be achieved by a reverse transcriptase (RT) and other more classical DNA polymerase, at lower temperature (42[deg.] C. for example for the reverse transcription step). HIV1 LTR DNA D6 (EMS positive) dilution was used as emitter using excitation frequency of 7 Hz during 18 hours in the apparatus described in FIG. 1, and placed close to the water receiver Tube 2. Like the latter, it was then diluted at 10<-2>, refiltered by 450 nM and 20 nM filters and diluted to 10<-15>.

[0105] Each dilution was then amplified by PCR for 35 cycles. Note the DNA bands detected at dilutions D2, (FD2), D3, D4, and D5. It has to be noted that the synthesis of the DNA LTR band is obtained in high water dilutions (up to 10<-9>) of the tube 2 containing the signalized water, indicating the transmission of the DNA information from tube to tube, in the presence of the ambient electromagnetic background. The same phenomenon was also observed in high dilutions of tube 1, indicating the synthesis of DNA at dilutions containing no DNA molecules.

[0106] This PCR technology can be applied to the detection of nanostructures in body fluid (plasma, urine) apparently devoid of the microorganisms from which they originate. In all cases, it is necessary to use mechanical agitation (vortex) at each water dilution in addition to the ambient or controlled electromagnetic background.

[0107] Table 2 shows the role of time of exposure to the 7 Hz frequency on EMS transmission from DNA to water. These results used the DNA LTR preparation as used for procedures reported in Table 1.

[0000]

TABLE 2

Content	Time of exposure (hr)	EMS	% over noise	Positive dilutions
Control DNA tube	2	+	57.3	D4-> D8
Water	2	-	0	
Water	4	-	0	
Water	6	-	0	
Water	8	+-	6.4	D4-> D8
Water	16	+	13.4	D5-> D8
Control DNA tube	16	+	63	D4-> D8

[0108] As shown above EMS were detected in the receiving sample after an exposure time of 8 or 16 hrs when the originating sample exhibited positive EMS at dilutions of D4 to D8 (10<-4 >to 10<-8>). No EMS was detected in water exposed for less than 8 hrs.

EXAMPLE 2

Identification of Unknown DNA Using Random Primers

[0109] Another aspect of the invention is directed to a general procedure for the identification of any unknown DNA sequence (or polynucleotide sequence) capable of producing EMS in biological fluids. The principle is shown by FIG. 3. The transmission of EMS in water allows the selective transmission of only the DNA sequences that were emitting the EMS under the induction conditions. The PCR method uses a combination of random and Tag primers. The random primer associated with the Tag has the following formula 5'-GGACTGACGAATTCCAGTGACTNNNNNNNN (SEQ ID NO: 1) in which are made all possible combinations of 8 nucleotides for the 4 possible bases (65,536). A detailed procedure is described below.

[0110] 1) DNA is purified from EDTA-collected human plasma extracted by the kit, QiaAMP, (Qiagen).

[0111] 2) The purified DNA samples are filtered through 0.45 and 0.1 [mu]m filters and then diluted to FD2-FD15 for analysis of EMS. FD2 refers to a filtered dilution of 1:100 or 10<-2>.

[0112] 3) The filtered and diluted samples are used to signalize water (molecular biology grade, 5Prime, 20 nm-filtered) with a dilution EMS+ of a patient DNA sample under an oscillating magnetic field of 7 Hz, 4V (coil in mu-metal) for 18 hours.

[0113] 4) Each EMS+ sample used is filtered, vortexed and diluted (FD2-FD5) the signalized water sample and proceed to EMS analysis.

[0114] 5) The samples of signalized water (EMS+), starting with FD2 are used as template for PCR amplification using random and Tag primers, following the protocol described below:

[0115] A 49 [mu]l PCR amplification mix containing 1* Advanced Taq buffer with Mg<2+> (available from 5Prime Co.), 200 [mu]M dNTPs, 20 nM of designed random primer Tag8N (SEQ ID NO: 1):

[0000]
(SEQ ID NO: 2)
(5'-GGACTGACGAATTCCAGTGACTNNNNNNNN)

[0116] 20 [mu]l of vortexed FD2 signalized water template, and 1 unit of Taq DNA polymerase (available from 5Prime Co.) is incubated stepwise at 8[deg.] C., 15[deg.] C., 20[deg.] C., 25[deg.] C., 30[deg.] C., 36[deg.] C., 42[deg.] C., and 46[deg.] C. for 2 min at each temperature to allow annealing of the random portion of the primer. An elongation step at 68[deg.] C. for 2-15 min was performed to allow synthesis of DNA, followed by a denaturation step at 95[deg.] C. for 3 min. One [mu]l of the designed primer Tag-ONLY (5'-GGACTGACGAATTCCAGTGACT) (SEQ ID NO: 3) is added to the mixture at a final concentration of 200 nM. The resulting sample is subjected to 40 cycles of amplification (95[deg.] C./30 s, 59[deg.] C./30 s, and 70[deg.] C./2 min), followed by an incubation at 70[deg.] C. for 10 min. PCR-amplified samples are subjected to electrophoresis in 1.3% agarose gel and stained with ethidium bromide to allow visualization of amplified DNA bands under UV light.

[0117] 6) If needed (if faint or no DNA bands a-re detected), sample can be reamplified by PCR using only the primer

Tag-ONLY, following the reamplification protocol described below:

[0118] A 50 [mu]l PCR amplification mix containing 1* Hot Start Taq buffer with Mg<2+> (available from 5Prime Co.), 200 [mu]M dNTPs, 200 nM of designed primer Tag-ONLY (5'-GGACTGACGAATTCCAGTGACT) (SEQ ID NO: 3), 1-10 [mu]l of PCR-amplified sample as template, and 1 unit of Hot Taq DNA polymerase (available from 5Prime Co.) is denatured at 95[deg.] C. for 3 min and subjected to 25-40 cycles of amplification (95[deg.] C./30 s, 59[deg.] C./30 s, and 70[deg.] C./2 min), followed by an incubation at 70[deg.] C. for 10 min.

[0119] 7) Isolation, purification and cloning of amplicons in pCR2.1-TOPO (InVitrogen) vector, followed by transformation of competent Escherichia coli cells, and screening for positive clones.

[0120] 8) DNA sequencing of amplicons using M13 universal primers (Eurofins MWG GmbH, Germany) and BLAST of the resulting sequences.

[0121] Application to a patient suffering from chronic Lyme disease:

[0122] A D9 (10<-9>) dilution of DNA extracted from the plasma of a patient suffering from chronic Lyme disease was transduced into water at an excitation frequency of 7 Hz for 18 hrs. PCR was performed on the water sample after transduction with Tag8N primers followed by a second PCR with Tag primers only. The PCR DNA products were resolved on agarose gels by electrophoresis and are shown in FIG. 4. As shown at the left, results obtained when the tube of D9 DNA and the naive tube of water are placed side-by-side. At right, results obtained when the two tubes were placed at a distance of 4 cm from each other.

[0123] Dw denotes control, naïve, unsignalized water.

[0124] Dw vor: denotes control naïve, unsignalized water agitated with a vortex.

[0125] D0: water that was transduced but not diluted.

[0126] D2 NF: same as D0 but diluted by 1:100 (D2).

[0127] D2 same as D2 NF, but filtered.

[0128] D3, D4, D5 represent further serial dilutions of D2 to factors of 1:1,000 (D3); 1:10,000 (D4) and 1:100,000 (D5). All serial dilutions were vortexed between each 1:10 dilution.

EXAMPLE 3

Recording and Transduction of EMS Signatures of HIV and Borrelia Burgdorferi

[0129] EMS signatures of HIV DNA and Borrelia DNA sequences are recorded and transduced as described below.

[0130] Step 1: Preparation of DNAs

[0131] 1. A fragment of HIV DNA taken from its long terminal repeat (LTR) sequence present in the viral DNA extracted from the plasma of a HIV-infected patient or obtained from a purified infectious DNA clone of HIV1 Lai, is amplified by PCR (487 base pairs) and nested PCR (104 base pairs) using specific primers: TR InS 5'-GCCTGTACTGGGTCTCT (SEQ ID NO: 4) and LTR InAS 5'-AAGCACTCAAGGCAAGCTTTA (SEQ ID NO: 5). A longer variant (300 bp) is obtained using the following primer: 5'-TGTTAGAGTGGAGGTTTGACA (SEQ ID NO: 6) in conjunction with the above primer InAS.

[0132] 2. A DNA sequence from Borrelia Burgdorferi, the agent of Lyme disease, is amplified by PCR (907 base pairs) and nested PCR (499 base pairs) with respectively Borrelia 16S outer and inner primers. Inner BORR16S inS 5'-CAATCYGGACTGAGACCTGC (SEQ ID NO: 7) and BORR16S inAS 5'-ACGCTGTAAACGATGCACAC (SEQ ID NO: 8). A shorter variant of 395 bp is obtained by using the following primer: 5'-GACGTCATCCTCACCTTCCT (SEQ ID NO: 9) in conjunction with the above primer inAS.

[0133] Step 2: Signal Recording

[0134] The resulting amplicons 104 bp and 300 bp for LTR and 499 bp and 395 bp for Borrelia were prepared at a concentration of 2 ng/ml in a final volume of 1 ml of DNAse/RNAse-free distilled water. The samples were read on an electromagnetic coil, connected to a Sound Blaster card (Creative Labs) itself connected to a microcomputer, (preferably Sony VGN-CS31) preferentially powered by its 12 volt battery. Each emission is recorded for 6 seconds, amplified 500 times and the digital file is saved, for example under the form of a sound file with the .wav format. This file can later undergo digital processing, by a specific software, Matlab (Mathworks), as for example digital amplification for calibrating the signal level, filtering for eliminating unwanted frequencies, or be analyzed by transformation into its spectrum by a discrete Fourier transform, preferably by the algorithm of FFT "Fast Fourier Transform".

[0135] Step 3: Signal Transduction in Water:

[0136] For transduction, the digital signal was converted by the digital/analog converter of the sound card into an analog signal. The output of the sound card of the microcomputer was linked to the input of a commercial amplifier (Kool Sound SX-250, www._conrad.com) having the following characteristics: passband from 10 Hz to 20 kHz, gain 1 to 20, input sensitivity 250 mV, output power RMS 140 W under 8 ohms.

[0137] The output of the amplifier was connected to a transducer solenoid which has the following characteristics: the bobbin has a length of 120 mm, an internal diameter of 25 mm, an external diameter of 28 mm, with 3 layers of 631 spirals of copper wire of 0.5 mm diameter and a resistance of 8 ohms, field on the axis to the centre 44 Oe/A, and on the edge 25 Oe/A. A measurement of 4.4 milliTesla (mT) was obtained when current, voltage and resistance were respectively, 100 mA, 4V and 8 ohms.

[0138] 50 ml of DNAse/RNAse-free distilled water (5-Prime Ref 2500010) are filtered first through a sterile 450 nM filter (Millex, Millipore, Cat N[deg.] SLHV033RS) and then to a 20 nM filter (Whatman, Anotop 25, Cat N[deg.] 6809-2002). For transduction, 1 ml of this filtered water in a Eppendorf sterile tube of 1.5 ml was placed at the center of the solenoid, itself installed at room temperature on an isolated (non metal) working bench. Alternatively, a sterile tube of 15 ml (Falcon-Becton Dickinson), filled with the filtered water can be used instead of the 1.5 ml Eppendorf tube.

[0139] The modulated electric current produced by the amplifier was applied to the transducer coil for 1 hr at the tension of 4 Volts. A current intensity of 100 mA was applied to the coil, so that no significant heat was generated inside the cylinder.

[0140] Step 4: Reconstitution by PCR of the DNA from the Signalized Water.

[0141] The water which has received the recorded specific signal is called "signalized water". The signalized water (kept in the same tube) was first agitated by strong vortex for 15 seconds at room temperature and then diluted 1/100 in non signalized DNAse/RNAse-free distilled water (30 [mu]l/3 ml). 1 ml was kept for control (NF, nonfiltered), the 2 mls remaining of signalized water were filtered through a sterile 450 nM filter and then through a 100 nM (Millex, Millipore, Cat N[deg.] SLVV033RS) for Borrelia DNA or 20 nM filter (Whatman, notop25) for HIV DNA. The filtrate was then diluted serially 1 in 10 (0.1 ml in 0.9 ml of DNAse/RNAse-free distilled water) in a Eppendorf sterile tube of 1.5 ml from 10<-2 >to 10<-15 >(D2 to D15). A strong vortex agitation was performed at each dilution step for 15 seconds. 5 [mu]l of each dilution is added to 45 [mu]l of the mix.

[0142] 1. Preparation of the mix for HIV LTR: The PCR mixture (50 [mu]l) contained 37.4 [mu]l of DNAse/RNAse-Free distilled water, 5 [mu]l of 10* Taq PCR buffer, 0.4 [mu]l of 25 mM dNTPs, 1 [mu]l of 50 [mu]M each appropriate primer Inner [LTR InS (5'-GCCTGTACTGGGTCTCT) (SEQ ID NO: 10) and LTR InAS (5'-AAGCACTCAAGGCAAGCTTTA) (SEQ ID NO: 11)], 0.2 [mu]l of 5 U/[mu]l Taq DNA Polymerase and 5 [mu]l of each dilution. The PCR was performed with the mastercycler ep (Eppendorf). The PCR mixtures were pre-heated at 68[deg.] C. for 3 min (elongation step), followed by 40 PCR cycles of amplification (95[deg.] C. for 30 s; 56[deg.] C. for 30 s; 70[deg.] C. for 30 sec). A final extension step was performed at 70[deg.] C. for 10 min.

[0143] 2. Preparation of the mix for Borrelia: The PCR mixture (50 [mu]l) contained 37.4 [mu]l of DNAse/RNAse-Free distilled water, 5 [mu]l of 10* Taq PCR buffer, 0.4 [mu]l of 25 mM dNTPs, 1 [mu]l of 50 [mu]M each appropriate primer Inner [BORR16S inS (5'-CAATCYGGACTGAGACCTGC) (SEQ ID NO: 7) and BORR16S inAS (5'-ACGCTGTAAACGATGCACAC) (SEQ ID NO: 8)], 0.2 [mu]l of 5 U/[mu]l Taq DNA polymerase and 5 [mu]l of each dilution. The PCR was performed with the mastercycler ep (Eppendorf). The PCR mixtures were pre-heated at 68[deg.] C. for 3 min (elongation step), followed by 40 PCR cycles of amplification (95[deg.] C. for 30 s; 61[deg.] C. for 30 s; 70[deg.] C. for 1 min). A final extension step was performed at 70[deg.] C. for 10 min.

[0144] Electrophoresis of the PCR products in 1.5% agarose gel: A band of 104 bp for HIV LTR and a band of 499 bp Borrelia DNA should be detected at several dilutions.

[0145] 3. Sequencing: The DNA bands are cut and DNA is extracted using a Qiagen kit which also describes classical conditions for cloning in E. coli. The amplified specific DNA is then sequenced to show its identity to the original DNA.

Method of Detecting Microorganisms with a Specimen
US2010323391

2010-12-23
Inventor(s): MONTAGNIER LUC [FR]; AISSA JAMAL
Classification: - international: C12M1/34; C12Q1/04 - European: C12Q1/04; G01N37/00
Also published as: FR2902883 // RU2009101670 // WO2007147982 // WO2007147982 // EP2044210

Abstract -- This invention concerns a process for preparing reagents intended for a microorganism detection test and notably an infection in humans or animals, wherein the following steps are included: a) Centrifuging a biological or artificial liquid medium containing a selected specific microorganism; b) Filtrating the supernatant obtained in step (a); c) Preparing a series of diluted samples corresponding to increasing dilutions of the filtrate obtained in step (b), down to a filtrate dilution of at least a factor of 10-15; d) Submitting said diluted samples obtained in step (c) to an electrical, magnetic, and/or electromagnetic exciting field; e) Analyzing the electrical signals detected using a solenoid, as well as digitally recording said electrical signal after analog/digital conversion of said signal; f) Selecting diluted samples with which the characteristic electrical signals were obtained in (e), i.e. signals whose amplitude is at least 1.5 times greater than background noise emitted by water and/or presenting a frequency displacement towards higher values; g) Placing the diluted samples selected in step (f) into protective enclosures, which are protecting said dilutions against external electromagnetic fields; h) Distributing one of the aforesaid diluted samples from step (g), volume by volume, into two tubes, T1 and T2, tube T1 remaining in a protective enclosure protecting diluted samples from external electromagnetic field interferences, and acting as a reference solution, tube T2 being also placed in a protective enclosure, and subjected subsequently to the presence or contact of a sample suspected to contain said selected specific microorganism.

Description

[0001] This invention has for object to reveal latent infections in humans and animals, by showing inhibition, through the examinee, of electromagnetic signals generated by a microorganism.

[0002] From the works by Dr. Jacques BENVENISTE and from patent application WO 00/17637, it has been known how to record and digitalize, after analog-to-digital conversion using a computer sound board, an electrical signal characteristic of a molecule possessing a biological activity.

[0003] Also known in prior art (WO 09417406) is a process and a device used to transmit biological activity from a first matter, so-called carrier, to a second matter, so-called target, the latter exempted of any traces from said carrier and physically separate from it, and the target not presenting initially the aforementioned biological activity. The method consists in (i) exposing the matter carrying the biological activity of interest to an electrical or electromagnetic signal sensor, (ii) amplifying said electromagnetic or electrical signals characteristic of the emitted biological activity feature, then (iii) exposing the target matter to an emitter of electrical or electromagnetic signals, said emitter being connected to aforesaid sensor through a transmission and amplification circuit, in order to transmit the signal characteristic of biological activity to said target.

[0004] In a previous French patent application 05/12686 filed on Dec. 14, 2005, not yet issued to this day, the inventor of this invention was describing a process for characterizing biochemical elements presenting a biological activity, microorganisms in this case, by analyzing low frequency electromagnetic signals, said process bringing improvements to prior art techniques. Said process also relates to biological analysis consisting in recording the electromagnetic or electrical "signatures" corresponding to known biochemical elements, and to compare such pre-recorded "signatures" to that of a biochemical element to be characterized. Said process implicates filtration and dilution steps in order to eliminate microorganisms and cells present within the original sample, the highest dilutions generating the most electrical or electromagnetic signals whereas the least diluted samples don't provide, most of the time, any electrical or electromagnetic signals. The inventor also showed that microorganisms of different nature, such as bacteria and viruses, produce "nanostructures" that persist in aqueous solutions, and that these very "nanostructures" are emitting electromagnetic signals. Said "nanostructures" behaves like polymers of a size less than 0.02 [mu]m for viruses, and less than 0.1/[mu]m for classic size bacteria, and present a density ranging from 1.12 and 1.30 g/ml.

[0005] The process described in this application is based on the astonishing observation that in absence of physical contact, the mere vicinity of a closed tube containing a sample of a bacterial or viral filtrate, little diluted and negative with regard to electrical or electromagnetic emitting signals, inhibits the signals produced by a more diluted sample of the same filtrate, initially positive with regard to electrical or electromagnetic signal emission. In this application, such inhibition will be indistinctly called "inhibitory effect" or "negativing effect". In the same way, in this application, to "inhibit" and "negativate" will be used indistinctly and have a similar meaning. This observation led the inventor to search for the same inhibitory phenomenon from an infected human being. It has been observed, in a patient suffering from an auto-immune microvascularitis of infectious origin, that the diluted samples of his plasma had an inhibitory effect on dilute filtrates of E. coli emitting electromagnetic signals (hereafter EMS), suggesting that the patient was suffering from a chronic infection by this or a related germ. It was also shown that the patient suffering from microvascularitis, as mentioned in the previous example, himself inhibits the EMS emitted by his filtered and diluted plasma, and also inhibits the EMS emitted by a filtered and diluted sample of E. coli culture present in a closed tube. In this case, a 5 minutes contact of a positive dilution in the patient's hand, or 10 minutes at a distance of up to 50 cm, are sufficient to observe said inhibitory effect.

[0006] Said inhibitory power thus involves both the emitting structures from one own plasma, and those of a specific bacterial germ, which could thus be used as a universal identification system.

[0007] The invention may therefore enable to determine a bacterial or viral origin in illnesses where such germs have not been identified.

[0008] A first object of the invention concerns a method for preparing reagents to be used in a test for detecting a microorganism and notably an infection in humans or animals. According to its most general acception, the method includes the following steps:

[0000] a) Centrifuging a biological or artificial liquid medium containing a selected specific microorganism;

b) Filtrating the supernatant obtained at step (a);

c) Preparing a series of diluted samples corresponding to increasing dilutions of the filtrate obtained in step (b), down to a filtrate dilution factor of at least 10<-15>;

d) Submitting the diluted samples obtained in step (c) to an electrical, magnetic and/or electromagnetic exciting field;

e) Analyzing the electrical signals detected using a solenoid and recording digitally aforesaid electrical signal, after analog/digital conversion of aforesaid signal;

f) Selecting diluted samples from which the characteristic electrical signals were obtained in (e), by characteristic signals one means signals whose amplitude is at least 1.5 times greater than background noise emitted by water, and/or presenting a frequency displacement towards higher values;

g) Introducing the diluted samples selected in step (f) in protective enclosures, which protect said dilutions from very low frequency external electromagnetic fields;

h) Distributing one of aforesaid diluted samples from step (g), volume by volume, in two tubes, T1 and T2, with T1 remaining in a protective enclosure protecting said diluted samples from external electromagnetic field interferences, said tube T1 acting as a reference solution, while tube T2, also placed in a protective enclosure, is subsequently being subjected to the presence or contact of a sample suspected of containing said selected specific microorganism.

[0009] By "a sample to be tested for presence or absence of aforesaid selected specific microorganism" one means: (i) a human or animal individual suspected to be infected by aforesaid selected specific microorganism, or (ii) a biological specimen or a biological or artificial fluid suspected of containing said selected specific microorganism, or (iii) a food component, a cosmetic, or a pharmaceutical composition susceptible to contain said selected specific microorganism.

[0010] By biological fluids, one means any human or animal fluid, e.g. blood, urine, various secretions. By artificial fluid, one means any reconstituted fluid for growing microorganisms, e.g. various culture media for bacteria, yeasts, and molds, and culture media for cells infected by a virus.

[0011] Another object of the invention concerns a system for detecting a microorganism within a sample. This system includes:

[0000] a) A tube T1 containing a reference sample emitting characteristic electrical signals, by characteristic signals one means signals whose amplitude is at least 1.5 times greater than background noise emitted by water, and/or presenting a frequency displacement towards higher values;

b) A tube T2 containing a sample emitting characteristic electromagnetic signal, said sample being identical to that contained in tube T1;

c) A protective enclosure for protecting tubes T1 and T2 against very low frequency external electromagnetic fields;

d) A tube T3 containing a control solution not presenting electromagnetic signal emission;

e) An equipment for receiving electromagnetic signals.

[0012] During detection, tube T2 will be subjected to the presence or contact of sample X to be tested for presence or absence of a selected specific microorganism.

[0013] Another object of the invention concerns a method for detecting a microorganism within a sample, characterized in that said method consists of the following steps:

[0000] a) A sample X, for which the presence of a suspected microorganism, e.g. E. coli, is to be established, is exposed to a sample as obtained after step (f) of the process according to one of claims 1 to 3, said sample obtained after step (f) being a dilution of a culture or biological medium filtrate containing said microorganism suspected to be present in sample X;

b) Comparing the electromagnetic signal emitted by sample X exposed to said sample obtained after step (f), obtained in step (a), with the electromagnetic signal emitted by an aliquot of the same sample obtained after step (f) and not submitted to sample X.

[0014] By "a sample X", one means (i) a human individual or animal suspected of being infected by aforesaid selected specific microorganism, or (ii) a biological specimen, or a biological or artificial fluid, suspected to contain said selected specific microorganism, or (iii) a food component, cosmetic, or pharmaceutical composition susceptible to contain said selected specific microorganism.

[0015] The methods according to the invention enable (i) to prepare reagents intended for a test to detect microorganisms implicated in chronic illnesses, and/or intended to detect systemic latent infections under circumstances where a quick and non invasive response is required, as it is in the case of e.g. avian flu virus detection, (ii) the identification of an infection in humans or animals.

[0016] Once the responsible microorganism identified, it is possible to confirm the presence of that germ using supersensitive PCR with specific oligonucleotidic promoters from such microorganism.

[0017] The invention shall be better understood by reading the following description, presenting in a non restrictive way examples of process embodiment according to the invention.

[0018] The figures in annex correspond to non restrictive examples of embodiment.

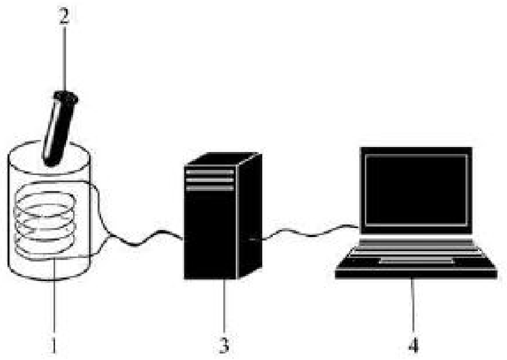

Figure 1. Device for the capture and analysis of em signals. (1) Coil made up of copper wire, impedance 300 Ohms. (2) Plastic stoppered tube containing 1 ml of the solution to be analyzed. (3) Amplifier. (4) Computer. From Ref. 1

EXAMPLE 1

A Lightly Dilute Bacterial Culture, not Emitting Electromagnetic Signals, "Negates" the Electromagnetic Signals Emitted by a Strong Dilution from the Same Culture

1) Sample Preparation

[0019] An Escherichia coli (E. coli) bacteria culture in LB (Luria broth) medium is centrifuged at 8000 rpm for 15 minutes in order to eliminate the cells. The bacterial supernatant is then filtered on a 0.45 [mu]m porosity PEVD Millipore filter, and the filtrate is then again filtered on a 0.1 [mu]m porosity Millipore filter.

[0020] From the resulting E. coli culture filtrate, which is completely sterile, one prepares a series of samples by diluting the filtrate from 10 to 10 into water down to 10<-15 >for injectable preparation. The successive dilutions are strongly agitated with a vortex for 15 seconds between each dilution.

[0021] The diluted samples are distributed in 1.5 ml Eppendorf conic plastic tubes. The fluid volume is in general of 1 milliliter.

[0000] 2) Selection of Diluted Samples Generating Electromagnetic signals.

[0022] Each dilute sample is tested for emission of low frequency electromagnetic signals.

[0023] The procedure for detecting EMS includes a step aimed at transforming the electromagnetic field from various diluted samples into one signal, namely an electrical signal, using a solenoid for capturing said electromagnetic field.

[0024] The transformation of the electromagnetic field coming from the diluted sample analyzed into an electrical signal is done as follows:

[0000] (i) Submitting the dilute sample being checked to an electrical, magnetic and/or electromagnetic exciting field;

(ii) Analyzing the electrical signals detected using a solenoid and digitally recording aforesaid electrical signal after analog/digital conversion of said signal;

(iii) Selecting the diluted samples generating characteristic electrical signals, by 'characteristic' one means signals whose amplitude is at least 1.5 times greater than background noise signals emitted by water and/or presenting a frequency displacement towards higher values, and placing them in Mumétal(R) protective enclosures for protecting said diluted samples against external electromagnetic field interferences.

[0025] Signal detection is carried out using the equipment schematically represented in FIG. 1. The equipment consists of a solenoid reading cell (1) sensitive from 0 to 20000 hertz, placed on a table made of insulating material. Said solenoid used in step (ii) includes a winding comprising a soft iron core. Said winding has an impedance of 300 ohms, an inside diameter of 6 mm, an outside diameter of 16 mm, and a length of 6 mm. The magnetic soft iron core is placed in contact with the external walls of the tube containing the dilution to be analyzed.

[0026] The diluted samples to be read are distributed in 1.5 ml Eppendorf (trade mark) conic plastic tubes (2). The fluid volume is in general of 1 milliliter.

[0027] Characteristic electrical signal acquisition is performed for a preset duration, i.e. ranging from 1 to 60s. In this example, each sample is read twice successively for 6 seconds.

[0028] The electrical signals delivered by the solenoid are amplified and converted into analog-digital signals using a signal acquisition board (sound card) (4) including a computer-built-in analog-to-digital converter (3). Said analog-to-digital converter has twice the sampling rate of the maximal frequency that one wants to digitalize, e.g. 44 kHz.

[0029] The digital file corresponding to said converted electrical signal is saved on a mass storage, e.g. as a WAV format audio file.

[0030] For processing the characteristic electrical signal, one uses e.g. Matlabs and SigViews (trademarks) software. The recorded digital file may possibly undergo digital processing, i.e. digital amplification for calibrating the signal level, filtering for eliminating undesired frequencies, calculating spectral power distribution (SPD), then such spectral power is truncated, e.g. only keeping frequency bands from 140 Hz to 20 kHz (Matlab), or is transformed in frequency components by Fourier transform (SigView).

3) Evaluating the Inhibitory Activity of a Non-Emitting Low Dilution on the Emission of Electromagnetic Signals Generated by an Active Dilution.

[0031] The diluted samples presenting characteristic electrical signals are samples diluted to 10<-8>, 10<-9>, 10<-10>. The 10<-2 >to 10<-6 >dilutions are negative (FIG. 2).

[0032] A closed tube containing a 10<-3 >dilution aliquot of E. coli is placed side by side with a closed tube containing a 10<-8 >diluted sample aliquot of E. coli, in an enclosure surrounded by a Mumétal(R) magnetic shield, and left 24 hours at room temperature. In parallel, a control series is realized. This control series consists of one tube containing a 10<-3 >diluted sample aliquot of E. coli, and of another containing a 10<-8 >diluted sample aliquot of E. coli that is processed in the same way, but in separate Mumétal(R) enclosures distant from one another. The placement in a Mumétal(R) enclosure eliminates very low active frequencies (5 to 100 Hertz) but not higher frequencies that could come from ambient electromagnetic noise.

[0033] After 24 hours, the tubes containing the diluted samples are again analyzed as describes above, revealing that the tube containing a 10<-8 >diluted sample aliquot and coupled to the tube containing a 10<-3 >diluted sample aliquot, no longer emits any electromagnetic signals, or much weaker ones. On the other hand, the control series tubes remained identical; the tube containing a 10<-8 >diluted sample aliquot protected from contact with the tube containing a 10<-3 >diluted sample aliquot remained positive for electromagnetic signal emission.

[0034] An important particularity of the invention is that the observed negating effect is specific, i.e. the lightly diluted, non-emitting sample and the greatly diluted electromagnetic signal-emitting sample must come from the same microorganism species.

[0035] Thus, the diluted E. coli-emitting samples are only "negated" by a weakly diluted non-emitting E. coli sample, but not by a lightly diluted non-emitting Streptococcus or Staphylococcus sample. Similarly, a diluted emitting Staphylococcus sample is only "negated" by a lightly diluted non-emitting sample of Staphylococcus and not by a lightly diluted non-emitting sample of Streptococcus or E. coli.

EXAMPLE 2

Quick and Non-Invasive Method for Detecting Infections in Humans and Animals

1) Preparations of Biological and Artificial Fluid Samples Containing Microorganisms.

[0036] A blood sample, collected with anticoagulant, preferably heparin, from a patient suffering from a neurological pathology consecutive to a bacterial infection, and an Escherichia coli (E. coli) bacteria K1 culture in suspension in LB (Luria broth) medium are centrifuged in order to eliminate the cells. The bacterial supernatant and/or the plasma collected are then diluted to 10<-2 >in RPMI medium. The solutions are filtered on 0.45[mu] Millipore PEVD filter, then the filtrate is again filtered on 0.02 [mu]m Whatman or 0.1 [mu]m Millipore filter.

[0037] From the plasma filtrates of infected individual and from the E. coli K1 culture, one prepares a series of diluted samples corresponding to increasing dilution levels, up to 10<-15>, in 10 to 10 dilutions in water for injectable preparation under laminar flow hood. The successive dilutions are strongly agitated with a vortex for 15 seconds between each dilution.

[0038] The diluted samples are then distributed in 1.5 ml conic Eppendorf plastic tubes. The fluid volume is in general of 1 milliliter.

2) Selection of Diluted Samples Generating Electromagnetic Signals.

[0039] The selection of the diluted samples emitting characteristic signals, signals whose amplitude is at least 1.5 times greater than the background noise signals and/or are of a frequency higher than background noise, is realized identically to what is described above in example 1, chapter 2. The method described as well as the material are identical to what is described above. Thus, the method includes a step for transforming the electromagnetic field from different dilutions into a signal, namely an electrical signal, by means of a solenoid capturing said electromagnetic field.

[0040] The transformation of the electromagnetic field from the analyzed dilution into an electrical signal is done by:

(i) Submitting the diluted sample being checked to an electrical, magnetic and/or electromagnetic exciting field;

(ii) Analyzing the electrical signals detected using a solenoid, and digitally recording said electrical signal after analog/digital conversion of aforesaid signal;

(iii) Selecting the diluted samples presenting characteristic electrical signals, by 'characteristic' one means signals whose amplitude is at least 1.5 times greater than background noise signals emitted by water, and/or presenting a frequency displacement towards higher values, and placing them in protective enclosures for protecting said diluted samples against external electromagnetic field interferences.

3) Evaluating an Infected Individual's Inhibitory Activity on the Electromagnetic Signal Emission Generated by a Microorganism.

[0044] The diluted samples selected at the previous step (item (iii)), from the plasma filtrate of the infected individual, from E. coli culture filtrate, i.e. the dilutions of filtered sample presenting a characteristic electrical signal, are distributed in Eppendorfs plastic tubes, at a rate of 1 ml per tube, and stored at +4[deg.] C. The diluted EMS emitting samples distributed in aliquots are protected from external influences by being placed in an enclosure protected from electromagnetic fields. Preferably, the enclosure is surrounded with a magnetic shield made of Mumétal(R), isolating the enclosure from very low frequency parasitic fields coming from the surroundings.

[0045] One of the diluted EMS emitting samples from the plasma filtrate of the infected individual, from E. coli culture filtrate, is distributed volume to volume in two tubes, T1 and T2, with T1 remaining in a protective enclosure protecting said diluted samples from external electromagnetic field interferences, that tube will act as reference solution; tube T2 will be subsequently subjected to the patient and is also placed in a protective enclosure.

[0046] Said protective enclosure being preferably surrounded with a Mumétal(R) shield.

[0047] **FIG. 2** represents schematically the steps to take when searching for the inhibitory effect. The search of the inhibitory effect is realized as follows:

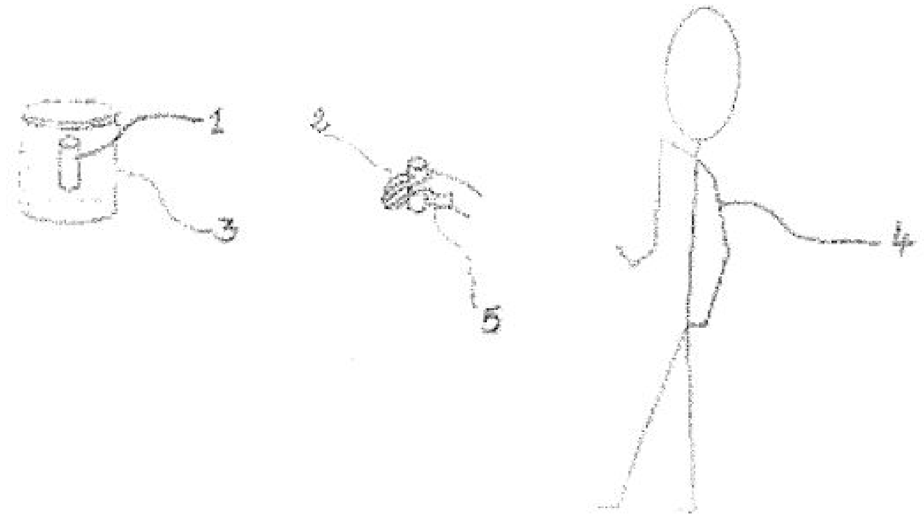

a) Tube T1, containing the reference solution, remains in an enclosure (3) surrounded by a Mumétal(R) magnetic shield, said tube T1 is thus protected from potential changes of the individual to be examined (4), whereas tube T2 is submitted to the influence of the infected individual to be examined (4) whose plasma present in tubes T1 and T2 comes from, said individual holds T2 in his/her hand (5) for a set period of time, e.g. 5 minutes;

b) Tube T2 is placed in an electromagnetic signal reception equipment, preferably a reading solenoid cell as described previously in chapter 2 of this example;

c) Electrical signals are then amplified, processed, converted into analog-digital signals as previously described in chapter 2;

d) Said analog-digital signals are possibly decomposed in harmonics by Fourrier transform.

[0052] The signals corresponding to tube T1 and those corresponding to tube T2, as well as the signals corresponding to water containing tube T3 (background noises) are compared.

[0053] The following figures represent the results obtained in the case where the active dilution comes from the examined infected individual plasma:

FIG. 3 represents a histogram in three dimensions (Matlab) of the electrical signals detected by the solenoid with tube T3 present (background noises);

Figure 3

TUBE T3 (BACKGROUND NOISE)

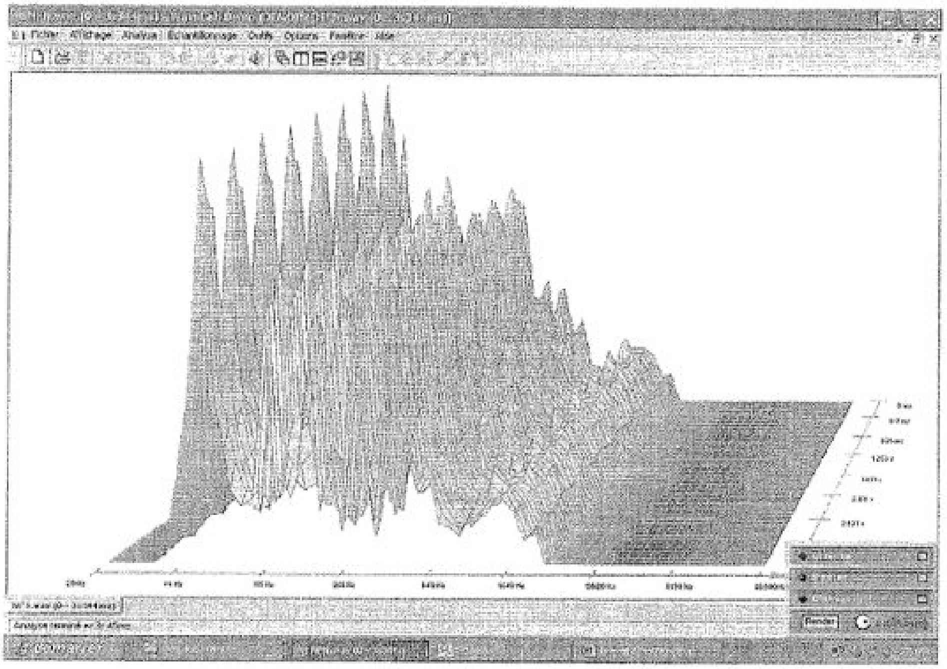

FIG. 4 represents a three dimension histogram of the frequency spectrum detected by the solenoid with tube 1 present;

Figure 4

Tube 1

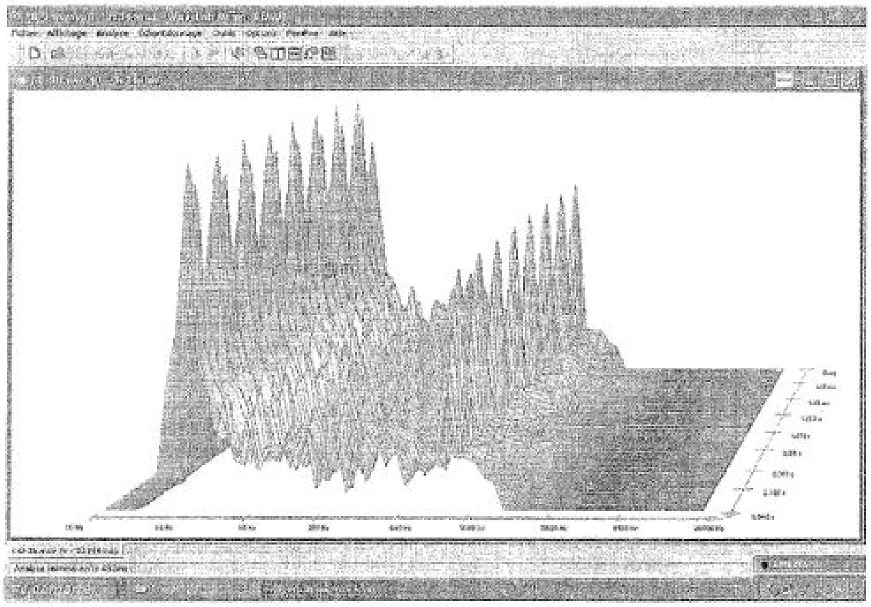

FIG. 5 represents a three dimension histogram of the frequency spectrum detected by the solenoid with tube 2 present;

Tube 2

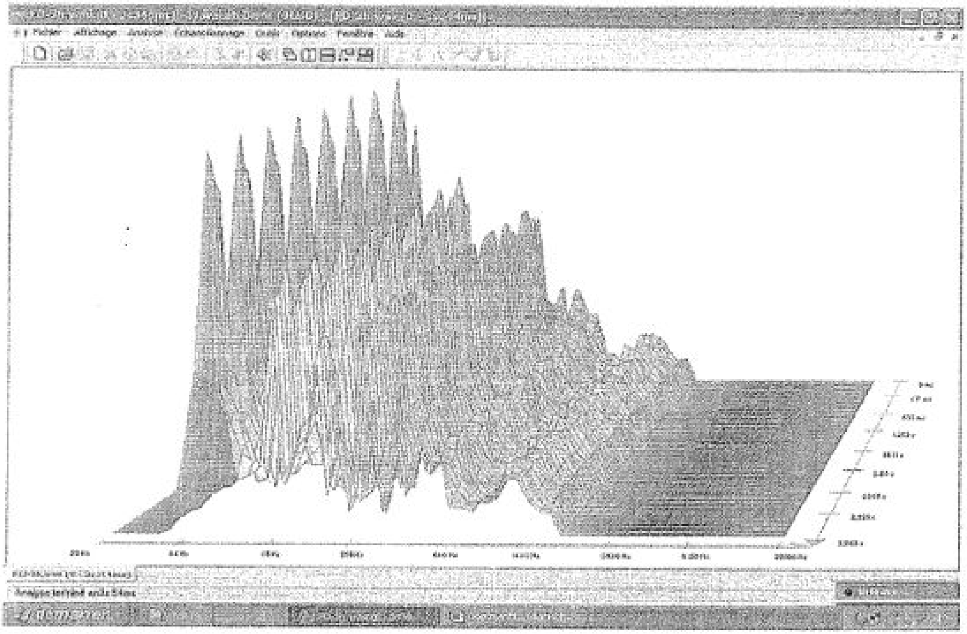
Fichier
Affichage
Analyse
Echantillonnage
Outils
Options
Fenêtre
Aide

FIG. 6 represents a Fourier analysis (SigView) of the same background noise (the harmonics of the non-filtered current of the power supply);

Figure 6

TUBE T3 (BACKGROUND NOISE)

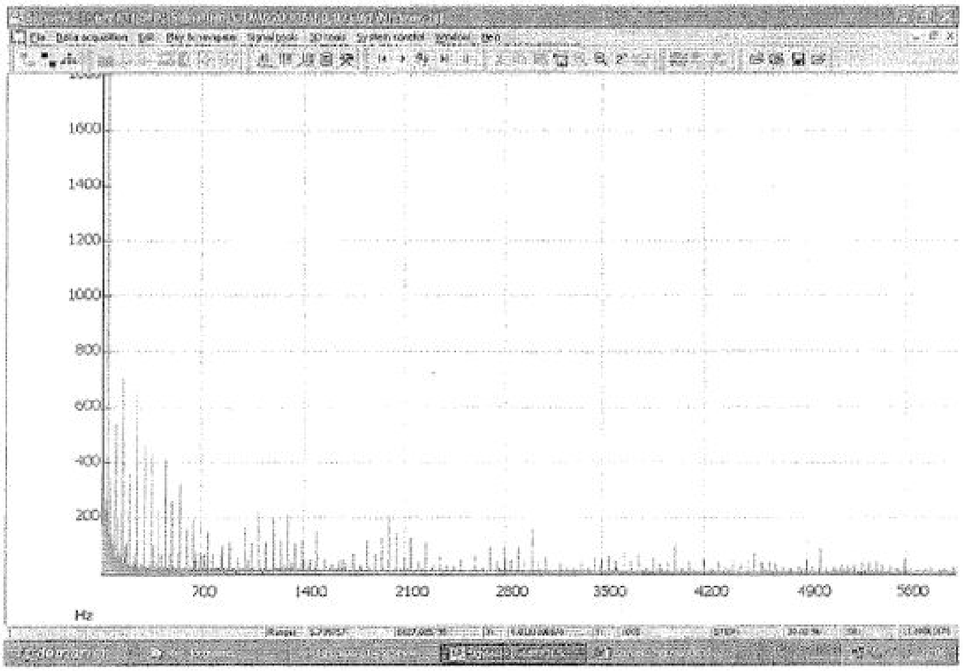
File
Data acquisition
Edit
Play & navigate
Signal tools
3D tools
System control
Window
Help
1600
1400
1200
1000
800
600
400
200
700
1400
2100
2800
3500
4200
4900
5600
Hz

FIG. 7 represents a Fourier analysis of the signal detected by the solenoid with tube 1 present;

Figure 7

Tube 1

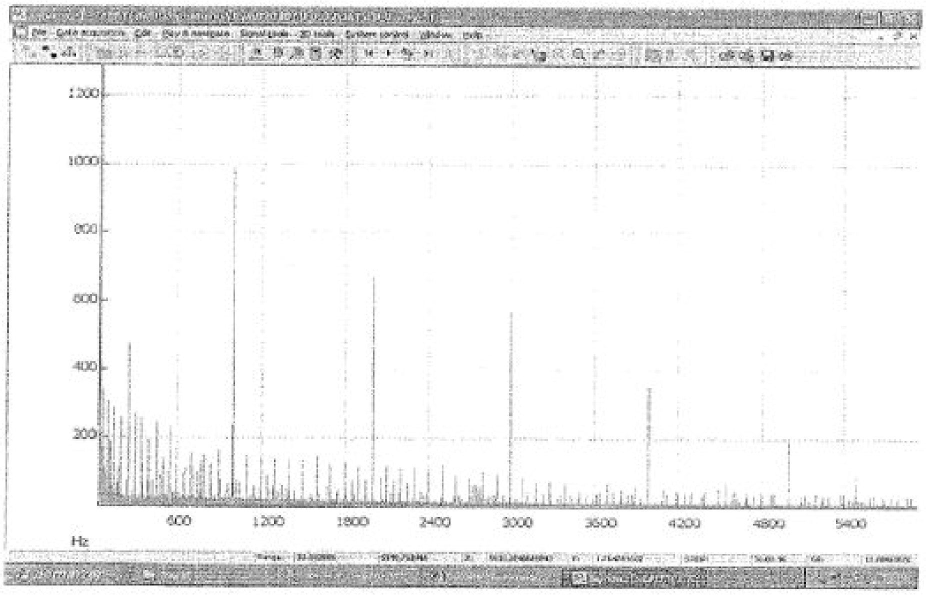
1200
1000
800
600
400
200
600
1200
1800
2400
3000
3600
4200
4800
5400
Hz

FIG. 8 represents a Fourier analysis of the frequency spectrum detected by the solenoid with tube 2 present, handled by the individual to be examined.

Figure 8

Tube 2

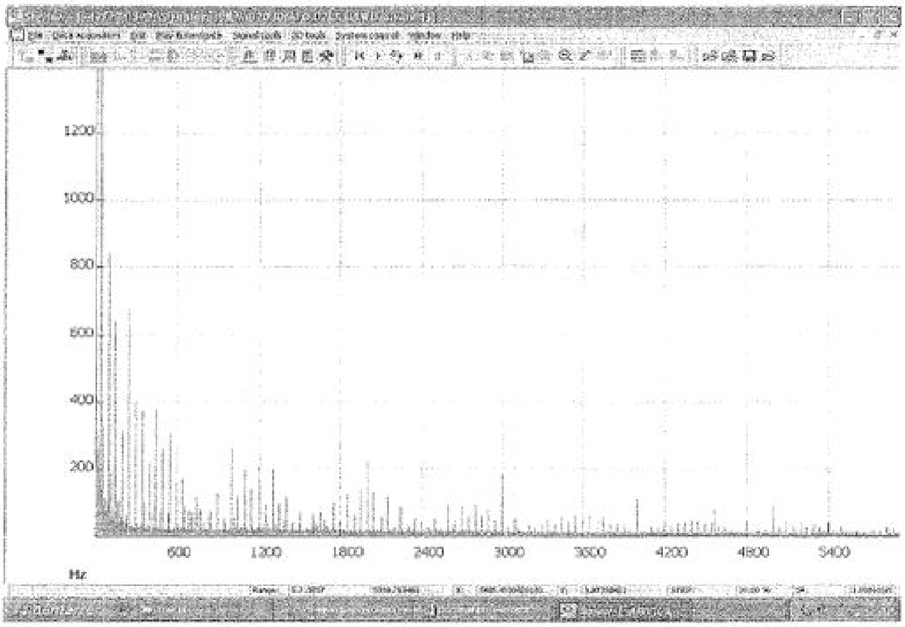
1200
1000
800
600
400
200
600
1200
1800
2400
3000
3600
4200
4800
5400
Hz

[0060] The analysis by 3 dimensions histogram, respectively for background noise (FIG. 3) and for the signal obtained with tube T1 present and containing the EMS emitting reference solution (FIG. 4), shows a displacement towards higher frequencies. On the other hand, when analyzing tube T2 containing the solution submitted to the influence of the individual to be examined (FIG. 5), no displacement toward higher frequencies is noted; the 3D histogram representing the signals of tube T2 is analogous to that obtained for background noise.

[0061] Fourier analysis of the positive frequencies generated by tube 1 (FIG. 7) revealed peaks at various frequencies. By decreasing order of signal intensity, the following frequencies presented signals: 1000, 2000, 3000, 4100, 5100 and 5500. On the other hand, Fourier analysis of tube T2 reveals results analogous to those obtained by background noise analysis: no significant peak was observed for background noise or for tube T2.

[0062] In conclusion, these analyses enable to deduct that the individual examined has a capacity for inhibiting electromagnetic signals emitted by a dilution of his/her own plasma.

[0063] Analogous results were obtained with the reference solution, derived from K1 E. coli.

[0064] Therefore, this inhibitory capacity concerns not only his/her own plasma but also E. coli emitting structures, suggesting that the individual is infected by an agent producing nanostructures close to those of E. coli.

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

rexresearch.com

Lucid Dreaming

MELIÃ REWARDS
2019 SUPER SALE
Up to 50% OFF
THE LOWEST RATES AT OUR HOTELS & RESORTS!
BOOK NOW

---

http://Wikibooks.org

Lucid Dreaming/Induction Techniques

This page describes a number of lucid dream induction techniques. It is recommended that you be able to recall *at least* one dream per night in order to maximize the effectiveness of these methods.

Certain elements are common to many of the lucidity-inducing techniques discussed later in this chapter. To better understand these techniques, these common components will be discussed first.

Sleep Interruption

An element shared by many of the techniques is sleep interruption. Sleep interruption is the process of purposefully awakening during your normal sleep period and falling asleep a short time later (10–60 minutes). This can be easily done by using a relatively quiet alarm clock to bring you to consciousness without fully waking you. If you find yourself resetting the clock in your sleep, it can be placed on the other side of the room, forcing you to get out of bed to turn it off. Other biorhythm-based options involve drinking lots of fluid (particularly water or tea, a known diuretic) prior to sleep, forcing one to get up to urinate.

Sleep interruption is a natural part of the MILD technique (described below) which trains you to arise immediately after your dreams end.

Sleep Continuity

If you have trouble initially falling asleep, avoid drinking water for about an hour before going to bed. Otherwise, you may find yourself running to the bathroom, disrupting any attempts at lucidity. Also, try to avoid caffeine and sugar before bed. However, depending on your sensitivity, caffeine may only stimulate your mind as opposed to your body. This extra grip on consciousness could be helpful in inducing lucid dreams. Exercising during the day is an excellent way of preparing your body for sleep. However, be sure to not exercise inside the three hours before bedtime, as your body will be stimulated for a short time afterwards. The morning or afternoon is the best time for this.

If you still have difficulty getting to sleep, try reading about lucid dreaming just before going to sleep. Your subconscious will likely absorb this information, increasing your chances of experiencing a lucid dream. If you *do* decide to read before going to sleep, keep a lamp next to your bed as physically getting up to turn off the lights may reawaken your body.

Reality checks

A reality check is a test you can perform to see if you're dreaming or awake. It might seem odd to test reality when you are sure that you're awake, but making a habit out of one or more of these reality checks will hugely increase your chances of having a lucid dream. If, say, you hold your nose and try to breathe in through it several times throughout the day then you're very likely to dream about doing it. And when you dream about performing a reality check, then of course the results should come out differently, in this case you'll find that you are somehow breathing in through your closed nostrils. You'll know that you're dreaming, and be able to take lucid control!

So here are some reality checks. You should be familiar with the entire list even if you only use a few.

		Reliability	**Speed**	**Discreetness**	**Overall**	**False positives**	**False negatives**
Breathing	Can you breathe with your fingers tightly sealing your nose and your mouth shut?	**5**	**5**	3	4.33	-	Very Rare
Jumping	When you jump, do you float back down?	**5**	**5**	1	3.67	-	Rare
Reading	Do sentences change when you read them? Read, turn away and repeat it to yourself, and then turn back and read it again. Do this twice.	**5**	4	4	4.33	-	Rare
Vision	Do you have perfect vision? This only works for people who have at least slightly blurry vision in the waking world. Alternatively, if you have *perfect* vision in the waking world, you may have blurred vision in the dream world.	4	**5**	**5**	4.67	-	Common
Hands	Do your hands have a strange color, too many fingers (sometimes they disappear and reappear when you try to count them!) or other abnormalities? Can you push your finger through your other hand?	4	**5**	**5**	4.67	-	Occasional
Time	Does your watch or clock tell a reasonable time? Are you even able to read the time off it? Sometimes	4	**5**	4	4.33	-	Occasional

	clocks have the wrong number of hands or have strange symbols. Try reading the time twice, like the Reading check above. Note: Digital clocks often work better for this reality check.						
Powers	Are you able to fly (just visualize it), unlock doors, or use other magical powers? Try to change the shape of your body, or walk through a wall, window, or mirror.	4	**5**	3	4	-	Common
Light switches	Does a light switch work?	4	3	1	2.67	Occasional	Occasional
Mirrors	Do you look normal in a mirror?	3	3	3	3	-	Occasional
Nose	Can you see your nose with one eye closed?	2	**5**	**5**	4	-	Rare
Memory	Are you able to remember how you got here, why you are here and what happened an hour ago?	2	3	**5**	3.33	Rare	Common
Identity	Do you have the same gender, age, job, relationships and responsibilities as usual?	2	3	**5**	3.33	-	Occasional

Choose a few reality checks which you will do regularly. Take them seriously, do not assume you are awake (even when you know you are). If you practice performing these checks very thoroughly while awake, then you're more likely to perform them thoroughly while dreaming. You should **always** carry out more than one reality check. If you find that it is not a dream, look around you and think of what would be different if it were a dream. If you do this it will make it more likely that you will do a reality check in a dream.

Apart from doing reality checks throughout the day, you also need to do a reality check immediately after you wake up. This helps you become lucid in false awakenings, when you dream that you have woken up but in fact are still dozing. Using a digital alarm clock or mobile phone display to do a Reading check, every single time you wake up, is a quick and reliable way to catch false awakenings.

If you have trouble bringing reality checks into your dreams then before going to bed imagine yourself in a dream, noticing odd details and doing a reality check. Then do a reality check in real life. If you do this a few times before bed you will find that you will do it more often in dreams.

If you are in a situation where you cannot do a reality check, such as at a public speaking, try to do one as soon as possible. You can do some reality checks very discreetly, such as looking at some text on a sign. If you start to say “well, I can't do a reality check now” you should not be surprised when you make this mistake in a dream!

Which reality checks are best?

When selecting reality checks, the most important properties of a reality check are *reliability*, *speed*, and *discreetness*.

- The **reliability** of each reality check is how likely you are to recognize the dream-sign's results as showing that you are dreaming **once you do them in a dream**. It changes for each person but some reality checks are overall more accurate than others. The figures in the table above are *rough only* and differ for each person.
- It is important for a reality check to be **fast**. It wastes dream time if you have to search around for a book or (perhaps worse) a mirror. Plus, it could also give your subconsciousness more time to produce real-life results, especially if you believe that the check will give real-life results.
- Third, a reality check should be **discreet**; that is, it should not draw too much attention to you when you do it in the waking world. Suddenly jumping in the air or trying to walk through a wall as a reality check could cause much embarrassment!
- Fourth, a reality check should not have "false positives". These occur when the reality check shows you are dreaming in real life. When you get a dream result, do another reality check to be sure!
- A final important note is whether or not a reality check should have "false negatives". These occur when a reality check shows you are in real life when you are actually dreaming. If you think you may be dreaming, keep doing tests until you prove yourself right!

On the table above, these are scored **out of 5**.

I have trouble remembering to do reality checks throughout the day. What reminders can I use?

You are lucky to have an interesting day and forget about lucid dreaming! You can tag your mind to remember dreaming when you think of certain things, like your friend or your homework. It isn't advisable to explicitly write “reality check” or “lucid” on your hand, as this could create an over dependence on this reminder, which may not exist in a dream. However, you might want to just draw a dot or small circle on your hand. This should be enough to remind you to do a reality check.

Try putting a little label on your clock, mobile phone, or watch, reminding yourself to do a reality check. (Some weird colors will make it more noticeable and it will take longer for you to get used to it and ignore it.) If you check these regularly during the course of your waking day, you will be doing lots of reality checks.

A simple coffee mug with a reminder such as "Are you dreaming?" printed on it or random alarms can also serve well, but try not to become too dependent on them. You can find examples of these at Byte Red.com and LD4All.com

Another technique is to write down three things you do regularly in a day. Examples include hearing your name, going through a doorway, turning on a TV, beginning to read a book, or seeing a stranger. In the morning, choose three such events and intend to do a reality check whenever they happen in the following day.

I did a reality check in a dream but it said that I was not dreaming. What went wrong?

Some reality checks work perfectly for some people and awfully for others. These are mostly the light switches one and the hands one. If you find that the light switch works or that your hands are perfectly normal, you need to change to a different technique.

I did a reality check in a dream but I did not quite realize I was dreaming. What went wrong?

An example of this is looking into a mirror and seeing some huge boils or a gray mist on your reflection and not realizing that you are dreaming. This is rare if you actually intended to look into the mirror as a reality check. You need to be more careful when doing your reality checks in real life or pick more reliable reality checks which show more obviously that you are dreaming. Also try to pick reality checks that are easy to do. For example, don't rely the Time Reality Check if you never wear a watch, and don't pick the Mirror Reality Check if you rarely look in the mirror during the day or you know that you won't find a mirror in your dream.

Another good remedy for this problem (and a good practice in general) is to perform two or three reality checks at a time. The Time Reality Check, for example, can be easily combined with attempting to push one hand through another. Or, for those with glasses, testing your ability to read text fits naturally with checking for "perfect eyesight w/o glasses".

Techniques

When you read through these techniques, remember that different techniques work for different people. There is no “best technique” and most techniques could be used to have 2–5 lucid dreams every night! You could have several lucid dreams in a night, but you will not know it unless you remember them!

However, you will probably want some advice as to which technique you should try first. Consider whether you want to use a method which starts from a dream or a method which starts from being awake.* If you master a technique which starts from being awake, you will eventually be able to have lucid dreams on demand when you sleep. For other techniques, you have to rely on your luck to give you lucid dreams after you have done your technique. Here are some advantages and disadvantages for specific techniques:

(Remember, it is *essential* to be able to recall at least one dream per night before attempting these techniques)

Technique	Summary	Advantages	Disadvantages	Best for...	Rating
WBTB (Wake-Back-To-Bed)	Wake after some sleep and then return to bed.	• Simple • Can be very reliable, especially when used with other techniques	• Disrupts sleep cycle	People who want to strengthen other techniques, or who wake up in the middle of the night anyway.	Green
Auto-suggestion	Let yourself genuinely believe that you'll become lucid—without *intending* to become lucid—so that you really will.	• Simple	• Less effective than some other techniques (such as MILD)	People who are susceptible to hypnosis or practice meditation.	Green
MILD (Mnemonic Induction of Lucid Dreams)	Fall asleep while focused on your intention to *remember* that you are dreaming.	• Simple	• Can be boring	People with a good prospective memory (remembrance of future intentions).	Green

WILD (Wake-Initiation of Lucid Dreams)	Keep your consciousness while falling asleep and go straight into a dream.	- Lets you truly induce lucid dreams at will	- Can cause frightening experiences - Can take long to master	People who want to reliably have lucid dreams.	Green
VILD (Visual Induction of Lucid Dreams)	By repetitive visualisation, incubate a dream in which you do a reality check.	- Also lets you induce lucid dreams at will - Works *extremely* well for some people...	- ...but not very well for others - Visualizing can keep you awake	People who have good visualization skills.	Green/ yellow
CAT (Cycle Adjustment Technique)	Adjust your sleep cycle to encourage awareness during the latter part of your sleep.	- Requires relatively little effort other than adjusting your sleep cycle - Is very effective	- Requires you to wake up early on some days - You are only likely to get a lucid dream on every other day (though this could easily be more frequently than with other techniques)	People who have a very regular sleep cycle.	Green/ yellow

[*] The usual acronyms in forums for this are DILD (Dream-Initiated Lucid Dream) and WILD (Wake-Initiated Lucid Dream). All the techniques that induce WILDs are described under WILD on this page.

WBTB

WBTB stands for “Wake-Back-To-Bed”.

Wake yourself up after 4 to 6 hours of sleep, get out of bed and stay up for anywhere between a few minutes to an hour before going back to bed. It is preferable that you do something related to lucid dreaming during this time (such as reading about lucid dreaming), but it is not required. This is best combined with other techniques; many people have amazing results with a MILD/WBTB combination. The WBTB technique significantly increases your chances of a lucid dream, and using MILD (see below) in conjunction increases your success rate if you are planning to sleep an hour or more after your WBTB session. However, you might need plenty of sleep time and therefore you may only be able to use it on weekends.

If you are sleeping too deeply to become lucid, then you can modify this technique. Try returning to sleep somewhere different than where you usually sleep, e.g. on a couch, a different bed, or even on the floor. If you are unable to do this, try changing the way you sleep, e.g. try sleeping with a lighter blanket or reversing the orientation of your body while in the bed (that is, swapping head and feet). Do this in order to teach your body that these different surroundings mean you want to have a more conscious sleep rather than a deeper sleep. In the beginning, different surroundings will also make you more alert, which can heighten your level of consciousness during sleep.

I am sometimes awake for very short times, but cannot pull myself together enough to get up and out of bed. What can I do?

Put a bright piece of paper on the wall or ceiling so that you will see it when you wake up. Other stimuli could be a hot water bottle, a light turned on under your bed, or an alarm clock. A good technique is to place an alarm clock out of arm's reach so that you are compelled to physically get up from bed and turn it off. If this is still insufficient to restore consciousness, try making a note of your intentions to remain awake and place the note on your alarm clock. After you get a lucid dream with this method, you'll find it easier and easier to get out of bed because you'll have more motivation.

Auto-suggestion

This technique describes how to use auto-suggestion to have lucid dreams. It can be especially effective for people who are highly susceptible to hypnosis or understand meditation, but for most people, MILD will probably be more effective.

As you are falling asleep, suggest to yourself that you will have a lucid dream either that night or in the near future. You can use a mantra (such as “I will recognize that I'm dreaming”) if you want, but make sure you don't try too hard to get a lucid dream. Instead of putting intentional effort into the suggestion, try to genuinely expect to have a lucid dream. Let yourself think expectantly about the lucid dream you are about to have, but be patient if you don't get one right away.

You may also use auto-suggestion to improve dream recall. Just use the technique as described above, but instead of suggesting that you'll have a lucid dream, suggest that you'll remember your dreams when you wake up. You could also use a mantra with this, such as “When I wake up, I will remember what I dreamt”. Just be careful not to put too much intentional effort into the mantra — try to genuinely expect to remember your dreams instead.

MILD

MILD stands for “Mnemonic Induction of Lucid Dreams", or sometimes, “Mnemonically Induced Lucid Dream". The MILD technique was developed by Stephen LaBerge, and is described fully in his book *Exploring the World of Lucid Dreaming*.

With the MILD technique, as you are falling asleep, you concentrate on your intention to *remember* to recognize that you are dreaming. Repeat a short mantra in your head, such as “Next time I'm dreaming, I will remember I'm dreaming”. Think about what this means (i.e., that you want to remember that you are dreaming—in the same way you might go to a grocery store and suddenly remember that you need bread), and imagine that you are back in a dream you've had recently, but this time you recognize that you are dreaming. For example, if you recently dreamed of flying, imagine realizing that it is a dream because you are flying. Keep repeating and visualizing the mantra until you are sure that your intention is set in your mind or you fall asleep. If you stop repeating and visualizing the mantra, then still try to make sure the last thing in your mind before falling asleep is your intention to remember to recognize that you are dreaming.

In general, the MILD technique can be practiced when you first go to bed at night, or after you have awakened from a dream during the night. If you practice the MILD technique after you have awakened from a dream, you should first run through the dream to ensure that you remember it. Some people find it helpful to jot down a few notes about their dream in their dream journal.

Once you have committed the dream to memory, go back to sleep and follow the steps above. But this time, visualize the dream you just had. Move through the dream in your mind until you encounter a dream-sign you originally missed. Now, instead of missing the dream-sign, imagine yourself recognizing it and becoming "lucid".

Repeat this until you have fallen asleep. You *will* re-enter the dream, and you *will* recognize the dream-sign and finally, become lucid.

WILD

WILD stands for "Wake-Initiated Lucid Dream", or "Wake-Initiation of Lucid Dreams" to refer to any technique that involves falling asleep consciously. These techniques are similar to self-hypnosis. Some people believe that WILDs are not actual dreams, but are instead astral projection. Various detailed resources are available under that moniker.

For most people, they are far easier to induce in the early morning after waking up or in afternoon naps, as the sleep cycle will continue with a REM period. Once you are experienced with inducing WILDs, you can try to induce them at other times.

For WILDs to occur, it is best for your body to be completely relaxed. When you go back to bed, lie down comfortably. Now tense and relax your body, starting from your shoulders and working downwards, then back up to the face. This (or similar relaxation, meditation, or trance techniques) should make your body feel slightly heavy and relaxed.

There are many different ways to induce WILDs, but they all involve simultaneously attempting to keep the mind aware while attempting to have the body fall asleep. A few techniques are detailed below.

If you pay attention to your physical body while using these techniques, then you will likely enter sleep paralysis (which usually happens after you are already asleep) without losing conscious awareness of your body. You will get a tingling and buzzing sensation (this might be unpleasant). These sensations might be so strong that you feel that you will die (e.g., you might feel a choking sensation), but don't worry, this is perfectly safe! In fact, this process happens to you every time you sleep, you are just not conscious during it. Sometimes you can simply wait until you fall asleep straight into a lucid dream. However, if you do not fall asleep, and you become completely paralyzed (with the exception of your eyes), do not try to move. Imagine your dream hand (or spirit hand if you prefer) going up and leaving your physical hand behind. Now you should have two separate bodies, a dream one and a real one. Control your dream body only — if you control your real one, you will wake up. Now you can try to roll out of bed into your dream world (alternatively, you can get up and walk through a mirror, or sink into your bed). There is a possibility that after waking up from a dream that you initiated using this technique, you may still be paralyzed. If this phenomenon occurs, it may be accompanied by hallucinations. For example, you may wake up from a lucid dream that you started using one of the WILD techniques, and you will still be paralyzed. A dark and scary figure may approach your bed, and you are unable to move, shout, or scream. Folklores have often given a name to this phenomenon: American people call it Old Hag , Japanese people call it kanashibari, Korean people call it kawi nullida (being pressed by a pair of scissors), and in old French it was called *Cauquemare*, which was later transformed into the term *cauchemar*.

Chakra Technique

Using Chakra ("third eye") meditation is one way to achieve WILDs. Basically, one has to focus solely on his third eye and breathing to achieve a lucid dream. This is a technique over 5000 years old, taught to Parvati by Shiva.[1] He is quoted as saying:

> With intangible breath in center of forehead, as this reaches heart at the moment of sleep, have direction over dreams and over death itself.[1]

Eyelid Patterns

This has been found to be very effective in many cases. However, it may lead to strange after effects, such as up to 15 dreams in one night, but otherwise, nothing harmful. This technique is as follows: When you go to sleep, ensure the room is completely dark. Then, with eyes closed, try to focus on the little dots that should appear to be moving on your eyelids. You will find that you can change their color at will. Continue focusing on these dots; make them dance around and form patterns and change colors, and eventually you should enter a Lucid Dream. This may take a little practice, but is usually very effective for summoning Lucid Dreams at will. Works better in conjunction with WBTB and other techniques mentioned above, but is still extraordinarily effective on its own.

Hypnagogic Imagery

Stimulate your thinking patterns by constantly switching your attention. After doing this long enough, the images and sounds should begin to take a momentum on their own (this is called hypnagogic imagery), and may they get very strange and illogical. You should enter a dream at this point and quickly become lucid. Otherwise, you will eventually realize you have entered sleep paralysis consciously (see above). Because hypnagogic sleep paralysis involves full consciousness, dreaming can sometimes be frighteningly real. There is often a feeling of being flipped upside down, spun, or being tugged upon by an outside force. Hypnagogic hallucinations may also include strange auditory hallucinations, dark beings and flying. It is possible to observe waking reality while in a hypnagogic state, but this is limited to the sensations of your physical body. Most hypnagogic sleep paralysis states occur when sleeping face up. There is evidence that the tendency toward experiencing Hypnagogic sleep paralysis may be hereditary.

Counting

Another technique is to count up to 100 in your head, optionally adding (for example) an “I'm dreaming” between each number. Alternatively, you can imagine stepping down stairs and reading each floor number, from 100 to 0. Try to make this image as vivid as possible — include not only what you see, but also what you hear, feel (touch the banister), and smell. At some point this image should continue into a dream or you will begin to get sleep paralysis as described above. It is easy to lose count, especially with counting up to 100 with an 'I'm dreaming' with each number. But stay focused: you are not going to sleep; your body is, and you must concentrate fully.

Sound Technique

This method is for people, who can hear the "tinnitus". The idea is pretty much the same as the other WILD methods, which is to remain conscious while entering the dream state. In order to use this method, you must sleep in a perfectly quiet place. You need to hear the inner buzzing sound inside your ears. Lay your body down and relax as much as possible while trying to hear the sound. If you are too tired, you will fall asleep too fast and it will be difficult to remain conscious - it this case you should combine it with WBTB. By time you realize that the buzzing sound will increase in intensity. This might frighten newcomers, but be assured - nothing bad is going to happen. No, you will not be deaf when you wake up, it’s perfectly safe! *It is just an effect caused by your brain is trying to change mode, from listening to the ambient sound, to listening to the sound of dreamland, which is not real sound but just electrical charge inputed to the part of the brain to create a sensation of hearing. By that time, you will enter the hypnagogic state. All you need to do is concentrate, do not be afraid or think of anything, just be still, and in time your dream body will float, separating from your physical body, and there you go, you arrive in the dreamworld.

Note: These sounds can be heard when you concentrate, even throughout the day - when you pay attention to them, but they shouldn't be heard much aloud if you are not in silence and concentrating on it - *as author said. The formula to hear it is quite simple: come into complete silence close your eyes and listen: is there only absolute silence or is there - something else? some...? do not create some sound just concentrate - :-) The best thing on this tech is - you can do it when you want.

Facts to know about Tinnitus

Tinnitus is a condition termed for a hissing and whistling sound that one hears in his/her ears. It's a medical term for ringing in the ears. Several reasons lead to this irritating condition and the most common cause is the damage to the nerves located inside the ear. However, there are some other factors as well such as earwax buildup, exposure to loud noises, ear infections and certain medications. Remember, the first step in finding a natural care is to determine the actual cause of tinnitus. Some Facts about Tinnitus.

The foremost thing to remember is that ringing in the ears is a symptom and not a disease. It's normal to a certain extent, however, the condition could be an indicator of a more serious condition. It's often related with hearing loss that is caused by aging. It can be a major nuisance when the ringing becomes very loud. The condition cannot be cured permanently but certainly can be eased.

Some Causes of Tinnitus

As mentioned above, ringing in the ear has some related causes and the most common out of them are damage of the cochlea due to aging (known as presbycusis), infection of the ear, exposure to loud noises and eardrum rupture. Amazingly, it has also been observed that excess ear wax buildup can also lead to this condition. Also, then there's the overuse of aspirin. Though, it's a very efficient pain reliever, its ototoxic properties (resulting into the damage to the ear) can result in tinnitus and worsen the condition in those that already have this ailment.

Tinnitus can also be developed by the usage of antibiotics, particularly amino glycoside antibiotics (a powerful form of infection fighting drug) and quinine. Also, many studies have proven the relationship between tinnitus and stress. When the brain is under stress, it has a tendency to read ordinary as well as daily sounds as fearful. After this, the tinnitus develops or becomes worse. Thus, it's very important and critical to manage stress. And if that's not enough, there are other causes for tinnitus that are not that common but are quite serious: Some of them are:

- A jaw misalignment disorder known as Temporo-Mandibular Joint (TMJ) Syndrome
- Meniere's disorder which is a disorder with recurring episodes of vertigo
- Head and neck injuries (resulting in tinnitus in one ear)
- A benign tumor known as acoustic neuroma which develops on the cranial nerve that runs from the brain to the inner ear

Treating Tinnitus

Putting in some hours with a therapist can really be beneficial and in several situations it's necessary to mental success even after the condition is alleviated. It's not easy to deal with anxiety and depression and knowing how to ignore the problem appropriately rather than indulging in it can be the best solution. It's also suggested to see a psychiatrist or psychologist who are experts with the patient suffering from the condition. This can be one of the cures for tinnitus relief and getting control over your mind can build up tolerance towards this condition which can also benefit you in dealing with other problems as well.

Masking the noise in your head can also provide you with a solution when you're trying to sleep or get rid of heavy waves of noise being produced. As tinnitus can appear loud in a silent room, leaving it on radio static can let you proper sleep. Remember, alleviating your tinnitus is only the cure for ringing in the ear as the condition can only be minimized and not completely cured.

Slight Physical Discomfort

For the purpose of helping to retain your conscious awareness, slight physical discomfort is useful while performing any WILD technique. This prevents you from just drifting off to sleep. If you are lying down on your bed to do WILD and you are totally comfortable then your chances of going to sleep instead of remaining conscious are very high. The WILD technique relies on a form of deep trance induction, and many people who induce trances for other reasons rely on slight physical discomfort — for example the lotus position, or sitting in a hard-backed chair. Depending on your own preferences and your requirements of discomfort for success, you could choose from the following methods, arranged in ascending order of discomfort:

1. Stacking pillows so that you can sit up in bed — the discomfort is caused by not being in a normal sleeping position.
2. Lying down on a hard floor.
3. Lifting your forearm vertically upwards, with the rest of your arm resting on the bed.
4. Sitting in a hard chair.

Incubating Dreams

To incubate a dream about a specific topic, you should first think of a phrase that summarizes that topic (e.g., “I want to go to Atlantis.”). It may help to write the phrase down. If there is something you want to do in the dream, think of a phrase to summarize that too (e.g., “I want to watch Atlantis sink into the ocean.”). If you want to become lucid in the dream, then you should probably write something like “When I dream of [the topic], I will remember that I'm dreaming.” beneath your topic phrase. Immediately go to sleep and focus on your topic phrase. Visualize yourself dreaming about the topic and (if you want to become lucid) realizing that you are dreaming. If there is something specific you want to do in the dream, visualize yourself doing it once you become lucid (not very likely to work if you don't become lucid in the dream). Think about your phrase and topic (and intention to become lucid) as you fall asleep. Make sure that the last thing in your mind before falling asleep is your intention to (lucidly) dream about the topic you want to dream about. You might want to wake yourself up when the dream starts to fade so that you remember more of the dream; you can do this by ignoring your perception of the dream environment — the opposite of dream stabilization techniques (just make sure you do a reality check when you wake up to make sure you are really awake).

Chaining Dreams

Dream-chaining or “chaining dreams” is a method to re-enter your dream *after you've woken up*. It can work for lucid and non-lucid dreams, but you will probably want to enter your dream lucidly.

Once you wake up from a dream (if you don't think you were dreaming before you woke up, it may not work well) you should stay still and keep your eyes closed. Small movements are okay, but the less movement, sensory stimulation, and less time awake, the better. Ideally, it should feel less like you've woken up, and more like you've taken a 30 second break from dreaming. Once you are prepared to go back to sleep, close your eyes and either visualize yourself back in your dream, or use the “spinning technique” given in the next chapter to imagine yourself spinning back “into” your dream. Spinning is a little faster than visualization. Be sure to maintain the fact that you are dreaming (unless you don't want to be lucid), or you may lose your lucidity while falling asleep. Once in the dream, stimulate your senses as early as possible (see the next chapter).

VILD

VILD stands for “Visual Induction of Lucid Dreams”, or sometimes, “Visually Incubated Lucid Dream”. This technique has been perfected by Peter Harrison, known as Pedro on the forums at ld4all.com. You may wish to read the main thread about the technique. The version described here has been adapted slightly.

First, make sure you are relaxed. You can use the relaxing technique mentioned in the description of the WILD technique. You can also imagine your brain emptying out and becoming sleepier. If you have a hard time falling asleep quickly, it should help to read a book (preferably about lucid dreaming) for a while before you go to sleep, until you feel very sleepy.

Now, you need to visualize a dream which you had prepared earlier. Here is an example of a prepared dream:

> I am in a red room with one door. A friend next to me asks me to show them what a reality check is. I do my reality checks (which show that I am dreaming), tell them that I am dreaming, and head towards the door.

Make sure you know exactly what the dream would be like, such as which friend, the exact words they say, and which reality checks you do. Reality checks that require no props, such as books or clocks, are recommended. Visualize this dream slowly three times, to make sure that you know every detail. Then, start going full-on and visualize the dream over and over. You should visualize the dream as though you are looking through your own eyes, not from a third-person perspective. If you find your thoughts drifting, ignore them and continue to visualize the dream continuously. You will need patience for this — don't just give up if you think it won't work.

When you actually dream this, you will not notice the difference — until you do your reality checks! Continue with the dream as you incubated it (e.g., remember to thank your friend!) before continuing through the door.

I tried to visualize the dream until I fell asleep, but I just stayed awake. What went wrong?

If visualizing keeps you awake, the VILD technique is not the technique for you! You should use a different technique.

LILD

LILD stands for “Lucid Induction of Lucid Dreams”, or sometimes, “Lucidly Induced Lucid Dream”.

To use this technique, you need to have a lucid dream in the first place, but it can help you to get more later. The idea is to do something in your dream that will help you to become lucid the next time you are dreaming. For example, you could ask a dream character for help — ask them to meet you the next night and tell you that you are dreaming. If it works out the way it should, then the next time you are dreaming, the dream character will walk up to you and tell you that you are dreaming, and so you'll (hopefully) become lucid. There are many variations on this technique; you could set up signs in your dreamworld that remind you to do a reality check or eat lucid pills instead! This technique is not likely to be very effective, but it *can* work; it relies on the chance that you'll subconsciously induce the reminder (i.e., the dream character or sign or whatever you used) during some later dream, and become lucid because of it.

Note that LILD is best used in conjunction with dream-signs and auto-suggested non-lucid dreams. The basic idea as explained above is to have something in your dream that triggers the transition from normal dream state to lucid dreaming. To simply tell a character to tell you that you are dreaming the next time you fall asleep is usually not enough. There is no guarantee that you will dream about that character and there is no guarantee that your subconscious will believe the character enough to make you snap into lucidity (make you realize that you are in fact dreaming).

Now as this technique suggests, you must have some previous alternate means of having a lucid dream. Whatever technique you employ to get into this initial lucid dream state is not really important, but you should try to remember to use this technique (LILD) once you do get into a lucid dream state. Thinking of this before falling asleep (MILD) sometimes helps and usually takes many lucid dreams before finally remembering. Once you are in a lucid dream, make up a dream-sign. It can be anything. It can be an object. It can be food or a drink (that doesn't taste like anything). It is usually best to pick something that isn't quite right. Something that on the surface would appear normal in the real world, but that upon closer inspection is not quite right. Food or drinks are good as they can have no taste or not be refreshing in a dream. But try and pick something that you dream about a lot so that there is a better chance of you dreaming about this dream-sign later on. Now pick something else that only appears or happens in your lucid dream. It can be anything. If there is nothing in your current lucid dream, create something really strange. Something that could never be confused with the real world. Now mentally associate the dream-sign (food) with this unusual item or event that could never happen in the real world. But at the same time, this unusual item or event should equate to "lucid dreaming". When you see the unusual item, it should only make you think of when you have a lucid dream as this should be the only time you encountered it. So we have a 3 item associative link. Do all of the above while in a lucid dream.

The next time you dream about your dream-sign, your subconscious will think of the unusual item or event. The unusual item or event will make you think of lucid dreaming. The two combined impossibilities (1. dream-sign that cannot exist in the real world 2. item or event that only appears in lucid dreams) will make your unconscious try to make a decision on all this. This will make your conscious mind come to the surface and hopefully you will come to the conclusion that you are dreaming. Many times, you will not want to deal with it because you are too tired (that is why you are sleeping, no?) and fall back into a normal dream state. This is why it can take a few tries. Eventually, your subconscious will start putting clear signs in your dreams like billboards that spell out "YOU ARE DREAMING". But once it triggers, it is quite the realization that an instant before, you had no real control over your actions and now you can do whatever you want. Another note... if it failed, you will usually know why. So next time, you can choose another dream-sign or slightly different technique or something more shocking. Once you get this working once, it is relatively easy to use over and over as the hard part just described is over with. Sometimes disassociative techniques are needed if used too much.

To sum up, this technique is a way to force a reality check while in a normal dream state where your subconscious has no choice but to come to the conclusion that you are in fact dreaming. Once your mind knows that you are dreaming, there will be no other conclusion than your conscious mind taking over. And this is what lucid dreaming is all about.

CAT

1. For one week, go to bed at the same time each night and get up 90 minutes earlier than you usually do. Spend those 90 minutes doing reality checks every 2–5 minutes.
2. Thereafter, on alternate days: follow the routine from step one, and set the intention to do your reality check routine at its regular time, while getting a full night sleep. This will cause the reality check conditioning to kick in during REM prime-time.

For detailed information on the Cycle Adjustment Technique, see the appendix on CAT : /wiki/Lucid_Dreaming/Appendices/CAT

Tibetan methods

Tibetan Buddhists practice what is known as Tibetan dream yoga. Probably the most time consuming way of inducing lucid dreams, it is also, according to the practitioners, the most rewarding. The basic practice is *awareness*. Awareness should be practiced while sleeping just as well as while being awake.

Meditating on the question “who is aware?” might catapult you into a higher degree of awareness according to Buddhist beliefs. Keeping this level of awareness is another matter. The Tibetans have developed many yogic exercises and disciplines to be practiced. Maybe the most interesting difference between Tibetan dream yoga and modern western methods of lucid dream induction is the Tibetan claim of the possibility to be aware during deep sleep, not only in the REM periods of sleep. For the reader who is interested in these methods a good start is to begin to regard all experience as a dream. After all, from the countless multitude of matter and radiation reaching our senses the nervous system tunes in only to a small fraction of this chaos. For members of the phalanx that believes we, more or less, create our own reality in the above sense this practice should feel natural. In general though, it is recommended to gain instruction from a teacher in the flesh rather than from books (like this one!).

The Tibetan Methods are not "time consuming" if the goal is to go into deeper than Lucid states. From a Raja Yoga stand point, from a Daoist standpoint, and from a Tibetan Yoga standpoint the goal is not to "play" in Lucid Dreams, but to dispel the delusional nature of what we call "reality." Also, to heal and to develop super-learning. Now "super-learning" is a relative term, it depends on your culture, modernity (in terms of entrepreneurialism - in 3rd world, 2nd world, 1st world sense; "are you developing a better rickshaw or are you developing an engine that runs off of water, what is relevant?"), and on the "energy" you find most attractive: innovative, creative, joyous, blissful, love, power, wisdom, mirth. Are you Ambitious-Worldly or Ambitious-Spiritual? All these factors play into it.

These techniques are very easy to learn and are rooted in a deeper science than the WILD, VILD, LILD -- one that does not use auto-suggestion or forced recall. In conclusion, the Tibetan methods are advanced yogic techniques that only an accomplished practitioner can fully teach. If one does not have access to a Tibetan dream yogi, one should probably concentrate on techniques that are more mundane from a Western point of view.

Other techniques

Many of these are combinations of other techniques with some addition or modification.

- **Inducing dream-signs** - You can become lucid by trying to induce specific dream-signs to watch for during your dream. You can use auto-suggestion (see above) to associate a specific dream-sign with doing a reality check, or you can just get used to doing a reality check whenever you encounter the dream-sign while awake. Some dream-signs you can use:
  - **Thirst** - Avoid drinking for very long. Wake up later in the night and put salt on your tongue or eat chili to make you even thirstier. Fill a glass of fresh cold water and take it with you back to bed. Hopefully, you'll dream of getting something to drink. Don't use this technique if you have trouble falling asleep while thirsty.
  - **False Awakening** - Set your intention as you fall asleep to wake up in the middle of the night. If you are a heavy sleeper, you'll hopefully dream of waking up in the middle of the night. If you are a light sleeper, you are probably more likely to really wake up.
  - **Bladder** - Drink large amounts of water before going to sleep. You should dream of having to go to the bathroom. You may wet your bed! This technique should be used carefully, to avoid the possibility of water intoxication.
- **Conditioning** - Strictly punish or reward yourself after a dream where you failed to realize you were dreaming or when you do have a lucid dream. This could increase motivation but not necessarily cause lucid dreams in itself. Cognitive psychology, however, states that this punishment/reward system is very counterproductive, because it ties our self-esteem to the outcome of the endeavor. The opposite of this system would be to see each attempt, including those that fail, as another step towards success.
- **Muscle Exhaustion** - Lift slight weights with one's arms and legs for about fifteen minutes before one begins to fall asleep. One should not, however, lift anything that strains the muscles too much, for that stimulates the sympathetic nervous system and excites the body. The method

should cause a release of tension in the muscles and tire them out. This combined with relaxation techniques and any of the -ILD's should produce a positive result.

Other methods

Food and drink

There are various foods and drinks that you can consume which seem to have some effect on sleeping and dreaming. Note that for most of these there is no explanation or scientific study of how they work, and some may simply be placebos.

Don't go overboard with the consumption of any of these, as overdosing could have nasty effects (well, milk should be safe unless you are allergic). **Don't experiment without accumulating enough knowledge first.** The authors in no way encourage the use of legal or illegal drugs.

- The amino acid tryptophan, which can be found in warm milk amongst other sources, is a precursor for the hormone serotonin, and has been proven to help you fall asleep.
- Vitamin B6 and others of the B group are important for neuronal functions.
- Melatonin is a hormone with neuronal effects. It is produced by the pineal gland in darkness and is suppressed by blue light. Wearing blue-blocking sunglasses or avoiding bright light in general the hours before bedtime maximizes night-time melatonin secretion.
- 5-HTP or L-5-HTP is a supplement that is related to serotonin, which some claim has induced lucid dreaming on approximately half the nights it is taken.
- Caffeine is useful in WILD techniques, as it helps the mind stay focused and think vividly. Please notice that caffeine is an addictive substance and may have negative effects on your health.
- An Amino Acid Blend made up of 2000mg L-aspartic acid, 4000mg L-glutamine, and 300mg L-theanine can substantially increase your odds of having a Lucid Dream.
- Galantamine - acetylcholinesterase inhibitors extract from plants such as *Lycoris radiata* (Red Spider Lily) or from chemical synthesis. Works best when taken together with choline and/or alpha-GPC.

Some people who practice Lucid dream (LD) or Out of body experience (OBE) use Galantamine to increase their odds to achieve LD or OBE. By taking small amount of Galantamine (around 4 to 8 mg) after 5 to 6 hours of deep sleep and practice the induction technique such as meditation, MILD or WILD many people report more success with Galantamine.

There also report that taking Galantamine without proper induction technique will not lead to LD or OBE but will result in only a vivid dream instead. It should also be noted that due to a long half life Galantamine will stay in the body for a period of up to and over 48 hours, as such it is advisable to space out the use of Galantamine over a period of three days so that the body does not build a resistance to the drug ruining its effectiveness.

Some people report mixing Galantamine with other Nootropic can enhance the degree of lucidity but this is still controversial since some mixtures may work for some people, but lead to failure for others.

Plants:

- Artemisia vulgaris
- Hypericum perforatum
- Korean Ginseng
- Valeriana officinalis
- Calea zacatechichi
- Ginkgo biloba
- Nutmeg - contains myristicin a deliriant.

For an in-depth guide to using supplements for lucid dreaming, see the book *Advanced Lucid Dreaming - The Power of Supplements*.

Drugs

Dissociatives and hallucinogens can be used to create a (more or less) lucid dream-like state, though whether or not these help with lucid dreaming is debatable. **The authors do not recommend use of these substances for induction of lucid dreams, nor do they urge the breaking of any applicable laws.**

Some dissociatives and hallucinogens are:

- Amanita muscaria
- Ayahuasca
- DMT - Common in the shamanic method of drinking Ayahuasca, to the result of entering the spirit realm. The herbal brew contains the psychedelic DMT, a drug is actually synthesized in minute quantities by the metabolism. It may be produced during dreams, adding to the enhanced experience.[2]
- DXM
- Ketamine
- LSD
- Mescaline
- Morning glory seeds
- PCP
- Psilocybe mushrooms
- Salvia divinorum

Gadgets

There are various gadgets you can use to become lucid easily. They generally detect when you are in the REM state and then provide a light and/or sound signal. This signal is supposed to be adjusted so that it doesn't wake you up but does enter your dream. The signal is then recognized as showing that you are dreaming, and you become lucid.

The most well-known device is the NovaDreamer from the Lucidity Institute. However, this product is no longer produced. Be sure to check for recommendations for devices from lucid dreaming forums.

A similar device is the DreamMaker. (http://www.wellnesstools.com/dreammaker_quickinfo.htm).The DreamMaker works very similarly to the NovaDreamer but without the Dream Alarm feature, which worked to wake the dreamer in the middle of the REM state. This device comes with a mask, a circuit board with adjustable controls, the batteries needed to operate it, a short owner's manual, a lucid dreaming workbook, and the Stephen LaBerge book *Exploring the World of Lucid Dreaming*. The circuit board is supplied completely ready to use, but you have to insert the batteries and put the circuit board into the mask yourself.

An alternative is the Kvasar (http://brindefalk.solarbotics.net/kvasar/kvasar.html). The Kvasar costs about $20 in raw materials, but needs to be constructed by somebody skilled in electronics as it is not sold commercially. It can also be hard to operate.

Another do-it-yourself alternative to commercial dreaming masks is Nate True's Lucid Dream Mask (http://cre.ations.net/creation/face-mounted-lucid-dreaming-mask),which does not bother with difficult-to-calibrate sensors and just uses a timer for flashing lights, and has (ostensibly) competitive results with all of the former gadgets.

Software

There are many programs for your computer that can assist with lucid dreaming. These can give out verbal cues while you sleep, or assist in doing your reality checks:

- iLucid Dream for the iPhone. Dream journal, reality checks, induction techniques, binaural beats for deep sleep and much more! http://cre.ations.net/creation/face-mounted-lucid-dreaming-mask. https://itunes.apple.com/us/app/ilucid-dream/id668103961?ls=1&mt=8
- Lightened Dream is available for Windows as freeware. It can be used as a dream journal and a dream-sign list. One of its more notable features is that you can have it play a voice or song each time you enter a REM sleep period. http://lucidcode.com/LightenedDream/
- DreamDoze A web application for Social Dream interpretation and dreams journal management. http://www.dreamdoze.com
- Gnaural is a multi-platform programmable binaural-beat generator. Gnaural is released as free software under the GNU General Public License. (For a browser based demo, see Gnaural Java) -- http://pantheon.yale.edu/%7Ebbl2/GnauralJavaApplet_signed.html .http://gnaural.sourceforge.net/
- Brainwave Generator is available for Windows. It works by playing binaural beats into your ears, changing your brainstate. http://www.bwgen.com
- Reality Check is available for Windows. It works by appearing on your computer screen at random times to remind you to do a reality check. http://bytered.com/software-realitycheck.html
- SBaGen is available for Windows, MacOS X, and Linux. It works by playing binaural beats into your ears, changing your brainstate. http://bytered.com/software-realitycheck.html
- Neuro-Programmer is available for Windows. This is powerful software that creates binaural beats and can work in conjunction with special goggles to induce certain mental states. http://www.transparentcorp.com
- LucidWeaver is available for mobile phones and PDA that support Java (J2ME). It includes a Dream alarm with sound cue which can be adjusted to a personal REM-sleep cycle for improving dream recall and lucidity training. Randomized reality tests can be set and it can be used as regular alarm clock.http://www.lucidweaver.com/
- Lucille 2.0 is an application to assist with doing reality checks. http://mythwell.com/lucille/index.html
- You Are Dreaming . http://www.sadglad.com/yad/yad.html
- Daldom a cross platform audio lucid dream induction software. http://www.sadglad.com/daldom/readme.html
- Lucid Dreaming Assistant DOS software (no home page) http://www.dreaminglife.org/downloads/lucid_dreaming_assistant.exe
- Unexplainable Store Lucid Dreaming Brainwave Entrainment MP3. http://www.world-of-lucid-dreaming.com/lucid-dreaming-mp3.html
- Sleep Check Reminder for Android. http://www.androlib.com/android.application.com-ropez-android-sleepcheckreminder-qDwmw.aspx
- iLucid for iOS (iPod, iPhone, iPad). https://itunes.apple.com/us/app/ilucid/id591703883?mt=8
- Sleep As Android is an Android sleep tracking app. It has a feature where it can play a sound file (by default, a soft voice that says "You're dreaming..." and echoes out) when it detects that the user is in a deep sleep. https://sites.google.com/site/sleepasandroid/

There is also a list of programs available at LD4all.com, under the “How” section.

Hypnosis

- Remotehypnosis.com has a hypnosis mp3 that can help you achieve lucid dreams.

Forums

- The World of Lucid Dreaming Forum is a good place to ask questions and receive feedback on how accurately you are performing these techniques. http://www.world-of-lucid-dreaming.com/forum/

References

1. ↑ *a b* 112 Meditations from Shiva to Parvati. *Internet Archive.* Accessed on 2009-09-18. http://web.archive.org/web/20080104094658/www.totallyok.com/secret/007m.htm
2. ↑ 5000 year old WILD technique. Very easy, and very effective. *DreamViews forums.* Accessed 2009-09-18. http://web.archive.org/web/20090406111627/http://www.dreamviews.com/community/showthread.php?t=71387

Patents

METHOD AND APPARATUS FOR ATTAINING LUCID DREAM STATE US2010130813

Inventor(s): DOZORTSEV DMITRI

The invention discloses an apparatus that facilitates attaining of lucid dreaming state.

0001] This application claims priority of U.S. Provisional Patent Application Ser. No. 61/117424 filed Nov. 24, 2008, which is incorporated herein by reference.

TECHNICAL FIELD OF INVENTION

[0002] The present invention is in the field of apparatuses for modulation of sleep pattern.

BACKGROUND OF THE INVENTION

[0003] It is also known that some people may experience so-called lucid dreams during REM where the person is aware that he or she is dreaming while the dream is in progress. During lucid dreams, it is possible to exert conscious control over the dream characters and environment and have them perform feats which would be physically impossible in the waking world. Lucid dreams can be extremely real and vivid depending on a person's level of self-awareness during the lucid dream.

[0004] The number of people who are naturally predisposed to lucid dreaming is small.

[0005] However, there are methods that allow a not predisposed individual to experience lucid dreaming.

[0006] For example, it is well known that during night sleep humans have several episodes of dreaming accompanied by rapid eye movements (REM).

[0007] The devices for assisting lucid dreaming usually detect REM phase by monitoring eye movement using detectors mounted within the mask and then use flashing light or sound to let person know that he is dreaming without waking him up completely.

[0008] It has also been known for a long time that most men and woman experience respectively penis or clitoris erections and changes in blood flow and heart rate during the REM phase.

SUMMARY OF THE INVENTION

[0009] The disclosed invention is designed to assist an individual to enter lucid dreaming filled with sexual experience, by providing sexual stimulation when REM is detected.

DETAILED DESCRIPTION OF THE INVENTION

[0010] The disclosed apparatus has two principal parts, one detecting sexual arousal during sleep and second that provides a sexually stimulating stimulus once such arousal is detected. Because sexual arousal during sleep coincides with dreaming phase, it becomes possible for a user to introduce a sexual connotation into the dream.

[0011] Furthermore, recognizing external sexual stimulus, exerted by the apparatus, user will become aware of being dreaming, thereby transforming normal dream into lucid with sexual content.

[0012] Alternatively, the monitoring unit of disclosed apparatus may detect dream state by other means, for example by monitoring REMs or heart rate or blood pressure, skin resistance etc and send the signal to the second unit that provides sexual stimulation.

[0013] In a more simplified form, the arousal is not being monitored or detected. Instead, sexual stimulation is applied at predetermined programmed intervals during sleep.

Method and apparatus for electrically generating signal for inducing lucid dreaming US8267851

Inventor:
KROLL JAMES M

BACKGROUND OF THE INVENTIONS

1. Technical Field

The present inventions relate to dream inducement and, more particularly, relate to a method and electrical apparatus for inducing lucid dreaming.

2. Description of the Related Art

A lucid dream is one in which the dreamer obtains conscious awareness while still remaining in the dream. While we are all familiar with ordinary non-lucid dreams, the notion of conscious awareness within ones dream-space is quite foreign to most. Lucid dreaming is better understood and leveraged in Eastern culture. For example, the use of lucid dreams for personal or spiritual development has been leveraged for thousands of years in Buddhist practice. In the West, most people are too busy to apply the mental training techniques required to achieve this state naturally.

Western researchers have tried to develop technological solutions in one sense or another to solve the challenge of having lucid dreams on demand. Secondary goals of such technology driven solutions have included maximizing dream vividness, recall, control or overall cognitive ability while in the dream.

The possible applications of lucid dreaming are limitless. Not surprisingly, entertainment seems to be the most common application of lucid dreaming. This might include flying like a bird, meeting up with long lost friends or even romantic fantasies. More esoteric and well known lucid dream tasks include the elimination of recurrent nightmares, practicing physical tasks and having that practice within your dream-space impact your actual wakeful performance, elimination of phobias, insights into your subconscious by engaging in conversation with dream characters and pursuing creative endeavors such as writing music or poetry under this slightly altered state of awareness.

Given the limitless number of applications for lucid dreaming, it is desired to identify a safe and predictable method for inducing lucid dreams. While lucid dream frequency is desired, it is also important to recognize that the recovery of full cognitive ability within the dream is crucial if some of the aforementioned goals are to be implemented by the dreamer.

It is known in the art that the quality of lucid dreams can vary greatly. Perhaps the single largest challenge once lucidity is obtained is the application of full cognitive ability. Lucid dreamers will often claim they were lucid, but that their cognitive ability was a fraction of what they expect when they are awake. Ramifications of this include poor ability to remember or implement goals, or an inability to remember the details of the dream once the dreamer awakens.

An early attempt at a technological solution to lucid dreaming was the concept of a lucid dream mask by LaBerge, et al in U.S. Pat. No. 5,507,716. This was essentially a facial mask worn during a lucid dream attempt. The mask has a compartment for a printed circuit board that contains a motion detector which aligns to the position of an eye. The circuit board was designed to detect rapid eye movement (REM). Upon this detection, the mask sent light or sound cues to the dreamer. These cues may manifest in the dream as light or sound anomalies. The dreamer had to train to use these cues in order to recognize that they are dreaming. The problems with these devices were several. First, most people find it difficult to sleep with a cumbersome mask strapped to their face. Second, the settings are very tricky. A little too much motion detection sensitivity and the cues can wake the dreamer before the dream is stable. Too little and the mask doesn't detect and deliver cues. Similar problems can be said of the brightness and sound settings. The worst part is that the brain becomes accustomed to this phenomenon and tends to either tune it out or wake upon light or sound stimulus. Other peculiarities can creep in such as the lucid dreamer feeling like the mask is on them during the dream leaving them unable to see. A review of various internet lucid dream sites will find few people claiming phenomenal success with such devices.

A more recent development in lucid dream induction was that provided by LaBerge in US Patent Application Publication No. 20040266659. In this application, LaBerge identified several classes of drugs that can positively impact dream lucidity and cognitive ability of the dreamer. The main focus of this development was the class of drugs known as acetylcholine esterase inhibitors. Fundamentally, these are drugs that slow the breakdown of acetylcholine in the brain. Commercial usage of drugs such as Galantamine or Huperzine-A for lucid dream induction resulted from this disclosure by LaBerge. While this class of drugs does seem to be safe and reasonably effective at lucid dream induction, they also exhibit a number of limitations. Insomnia is a common consequence of using these drugs. Not only can this leave the subject sleepy the next day, it eliminates any possibility of lucid dreaming. Since these drugs bias acetylcholine (AcH) levels, they tend to create a chemical balance that favors REM sleep over delta wave and other physically recuperative cycles of sleep. Again, this can often leave the subject run down and sleepy the next day due to a lack of recuperative sleep the prior evening. A noticeable tolerance to acetylcholine esterase inhibitors has also been exhibited and is a common problem known to those skilled in the art. A dose that used to cause a positive effect every 4th day (generally accepted as the minimal inter-use window for galantamine), can soon only realizes a positive effect once per week. Even once a week can become a challenge for users who have leveraged a drug like galantamine for a year or more. Discussion of this tolerance issue is well documented on internet based lucid dream bulletin boards. Even with the use of these AcH boosting drugs, lucid dreamers often report unsatisfactory cognitive ability and critical thinking recovery once lucid. The last and final challenge is the gastro-intestinal distress and intolerance some people realize.

Another approach to the generalized modification of brain state activity was cranial electrical stimulation (CES). Cranial electrical stimulation was researched and developed in the Soviet Union in the 1950's. At the time this application was termed "electro-sleep" because studies focused mostly on resolution of insomnia in human subjects. By the 1960's a significant amount of private and university level research was underway in the United States for both animal and human subjects. The typical applications explored during this time were reduction of pain, anxiety, depression and elimination of insomnia.

Cranial electrical stimulation differs from several other electrical or magnetic stimuli such as transcutaneous electrical nerve stimulation (TENS) or transcranial magnetic stimulation (TMS). Cranial electrical stimulation involves the delivery of a small micro current across the brain. Typically, this is delivered via electrodes clipped to the ear lobes. The currents involved are quite small, on the order of 50 [mu]A-5 mA. Research has demonstrated that the preferred waveforms for cranial electrical stimulation applications are square or rectilinear biphasic signals with zero mean voltage delivered over time. These types of biphasic signals are easily designed and implemented to have equal positive and negative portions to eliminate the possibility of elecrolysis of the blood. This positive/negative behavior of the signal better replicates the response of nerve impulses that are characterized by having one polarity upon application of pressure and a potential of equal but opposite polarity upon release of pressure. Pulse widths of 0.1 to 2.0 seconds are commonly used and the envelope of the signal typically used for traditional applications is designed to be periodic at a rate of 0.5-3.0 Hz. It is also recognized by those skilled in the art that dynamic modification of the signal is often preferred as the brain can otherwise develop a tolerance to a repetitive electrical stimulus and slowly become unresponsive to this external signal. Methods to dynamically modify the signal include but are not limited to using on/off pulses, changing the envelope frequency over time, using a low frequency envelope to modulate a higher frequency carrier wave, changing the time domain format of the periodic signal, making changes in signal magnitude, etc.

Other cited methods are different than cranial electrical stimulation. Transcutaneous electrical nerve stimulation (TENS) for example is typically applied to injured or uncomfortable muscles and/or soft tissue areas. TENS is typically delivered in the 50-200 Hz range at much higher currents. Transcranial magnetic stimulation (TMS) differs from cranial electrical stimulation in the stimulus itself. TMS is a magnetic alteration of brain response as opposed to an electrical stimulus.

While decades of research have been implemented in this area, very little has actually been quantified in terms of how cranial electrical stimulation actually works. Most data and conclusions are empirical in nature. Blind or double blind studies routinely prove that low frequency cranial electrical stimulation in the 0.5-3.0 Hz range has positive effects in the aforementioned diseases and discomforts. Certainly it is well understood via electroencephalogram (EEG) measurements of patients before and after cranial electrical stimulation treatment that cranial electrical stimulation tends to eliminate irregular or abnormal electro chemical activity especially in the alpha range of 8-12 Hz. However the precise mechanism that drives this response is poorly understood. Most research cites the generic theory that cranial electrical stimulation helps to drive neurotransmitter balance back to a normal homeostasis. Cranial electrical stimulation regiments typically follow the model of 30 minutes of treatment per day for 5-15 days. Depending on the ailment and individual, positive response is often subjectively perceived for anywhere from 1 week to 2 years.

DIETARY SUPPLEMENT AND A METHOD TO ENHANCE SLEEP AND LUCID DREAMING
US2008107754

Inventor(s): LUCIANO JEFF

A nutritional supplement for enhancing sleep and lucid dreaming in humans. It contains a combination of ingredients in proportions calculated to enhance lucid dreaming. The primary ingredients are Calea zacatechichi, L-5-Hydroxytryptophan (L-5-HTP), and Vinpocetine. In addition, the nutritional supplement may include the secondary ingredient Melatonin and the tertiary ingredients Wild Lettuce Extract, Mugwort Extract, Dimethylaminoethanol Powder (DMAE), Passionflower Extract and Green Tea Extract. Further, various Vitamins may be added such as certain B vitamins, D and C, as well as Zinc, Magnesium and Calcium. The selection and amounts of the ingredients of the nutritional supplement promotes sleep and lucid dreaming in people who have taken the nutritional supplement prior to going to sleep.

[0001] This application claims one or more inventions which were disclosed in Provisional Application No. 60/864,801, filed Nov. 8, 2006, entitled "DIETARY SUPPLEMENT TO ENHANCE LUCID DREAMING". The benefit under 35 USC [section]119(e) of the United States provisional application is hereby claimed, and the aforementioned application is hereby incorporated herein by reference.

FIELD OF THE INVENTION

[0002] The invention pertains to the field of dietary supplements. More particularly, the invention pertains to the use of a specific dietary supplement as a sleep aid and lucid dreaming enhancer in humans.

BACKGROUND OF THE INVENTION

[0003] Dreams have been recognized as sacred and spiritual experiences in many cultures around the world since the beginning of recorded history and perhaps even earlier. In the last century, the understanding of dreams has changed from one related exclusively to the spiritual to that pertaining to the physical/psychological realm. Toward the end of the twentieth century, the idea of lucid dreaming, or dreaming with deliberate intention and control, has moved from the world of shamans or spiritualists to realization that lucid dreaming can be achieved by anyone with the assistance of nutritional supplements and practice.

[0004] In order to facilitate the understanding of lucid dreaming, it is important to understand some basic elements about sleep and common, or non-lucid, dreaming. Humans have little understanding of why we sleep and dream. It is known, though, that the most vivid dreams occur during that phase of a sleep cycle called Rapid Eye Movement, or REM. It is known that healthy human adults go through about 3-6 cycles during a normal night's sleep, with each cycle being about 90-120 minutes long. REM sleep is usually the last phase of each cycle. Dreams can occur during any phase of any cycle, however, it is known that those dreams that occur during REM sleep tend to be the most vivid and visual and are usually not directly related to current events from the dreamer's experiences during the prior day.

[0005] As scientists began to learn more about the function of the human brain, it was discovered that the stimuli that produce images in dreams originate in a part of the brain not associated with thinking, desires or wishes, but a more primitive area. This lead to the neurophysiological or synthesis activation theory of dreams in which dreams result from random stimuli from the brain stem. In accordance with this theory, the cognitive parts of the brain play a minimal part in dreaming, probably limited to trying to make sense out of the random images originating from the primitive region of the brain stem.

[0006] It is surmised that dreaming is important in the development of memory. Human infants spend up to 16-18 hours a day sleeping and approximately 50% of this is in REM sleep. It has been determined that intense sessions of learning are followed by an increase in REM sleep. Dreaming also seems to play a role in emotional adjustment and in coping with traumatic experiences.

[0007] The term "Lucid Dreaming" was first used in 1913 by Fredrich van Eeden, a Dutch physician. Lucid dreaming is perhaps best defined as being aware that you are dreaming and is most likely to occur in REM sleep. Lucid dreaming can be either low level or high level. Low level lucid dreaming means that the dreamer is aware of being in a dream, but that not all the people, animals, things and scenes are constructs of his or her mind. In high level lucid dreaming, the dreamer is not only aware of being in a dream, but that all of the elements of the dream are being generated by their mind and are therefore potentially under their complete control.

[0008] According to studies done on lucid dreaming, these dreams may not only be potentially pleasurable or wish fulfilling, they can be tools for improving the quality of the dreamer's waking lives by facilitating healing, inspiring peak athletic performances, putting an end to recurrent nightmares, overcoming phobias, reducing or eliminating various anxieties and improving problem solving capabilities.

[0009] There are methods that have been developed by various researchers in the field, such as LaBerge and Rheingold. A common technique is used by the Lucidity Institute, founded by LaBerge, and is referred to as Mnemonic Induction of Lucid Dreams ("MILD"). The basics of the MILD technique involves various steps such as deliberately making yourself wake up during a dream cycle and immediately recalling the details of the dream. Over time, the dreamer is capable of recalling all of his or her dreams. The dreamer is then directed to use a "dream sign" to use as a cue to recognize that he or she is still dreaming. Becoming proficient at lucid dreaming requires a great deal of time, dedication and practice.

[0010] In order to enhance the training of a person to develop strong lucid dreaming skills, various devices have been devised. Such devices use light and/or sound to help trigger lucid dreams. Among the shamans of some primitive cultures, they would enhance their ability to have lucid dreams by using various drugs. However, as one can imagine, most of these substances are considered illegal in the U.S. Among the legal substances known to promote sleep (and possibly dreaming) are the B-vitamins, Calea zacatechichi, melatonin, mugwort and passionflower. Before using such a substance alone or in combinations, it is important to know how to use it and what side effects it might cause. There is a need, therefore, for a safe but effective nutritional supplement that helps a person go to sleep and which can enhance lucid dreaming even without extensive mental preparation or exercises.

SUMMARY OF THE INVENTION

[0011] The present invention is a nutritional supplement that acts as a sleep aid and has been found to enhance lucid dreaming in humans. It consists of a combination of ingredients in proportions calculated to provide these results. The primary ingredients are Calea zacatechichi, L-5-Hydroxytryptophan (L-5-HTP), and Vinpocetine. A secondary ingredient is Melatonin. In addition, the nutritional supplement of the invention may include Wild Lettuce Extract, Mugwort Extract, Dimethylaminoethanol Powder (DMAE), Passionflower Extract and Green Tea Extract. Further, various B-Vitamins, vitamin D, vitamin C, magnesium, calcium, as well as Zinc, may be added.

[0012] The relative amounts of these ingredients used in the nutritional supplement will be discussed in more detail in the Detailed Description of the Invention. The use of this nutritional supplement has been seen to assist a person in going to sleep as well as helping to promote lucid dreaming in people who take the supplement prior to going to sleep.

DETAILED DESCRIPTION OF THE INVENTION

[0013] The nutritional supplement of the invention is used as a sleep aid and a promoter of lucid dreaming in people who take it prior to going to sleep. To reiterate what was stated in the Background of the Invention, lucid dreaming is defined as being aware that the dreamer is dreaming. It is most likely to occur in REM sleep and can exist either at a low level or a high level. Low level lucid dreaming means that the dreamer is aware of being in a dream, but that not all the people, animals, things and scenes are constructs of his or her mind. In high level lucid dreaming, the dreamer is not only aware of being in a dream, but that all of the elements of the dream are being generated by their mind and are therefore potentially under their complete control.

[0014] The nutritional supplement of the invention is useful in promoting lucid dreaming in people who have not necessarily trained themselves to lucidly dream by performing extensive training techniques over a long period of time. It has been formulated to provide the best effect without the need to have undergone extensive mental conditioning, such as that taught by many organization and found in various publications, or subjecting oneself to some of the known mechanical devices invented by Laberge.

[0015] The primary ingredients are Calea zacatechichi powder, L-5 Hydroxytryptophan (L-5-HTP) and Vinpocetine. A secondary ingredient is Melatonin. Additional ingredients are Wild Lettuce Extract 4:1, Mugwort Extract 4:1, Dimethylaminoethanol Powder (DMAE), Passionflower Extract 10:1 and Green Tea Extract 10:1. Further, various vitamins may be added, such as B, C and D, as well as zinc, calcium and magnesium.

[0016] All of the ingredients of the nutritional supplement of the invention are known. For instance, Calea zacatechichi is a known dream herb, used by the Chontal natives of Mexico. Green Tea is a well known drink that is high in antioxidants. Mugwort has been used for medicinal purposes for centuries and is known to cause a dreamy state of consciousness. Passionflower (Passiflora incarnate) has mild sedative properties that may be used to treat insomnia and anxiety. Wild Lettuce is an ingredient in some sleep tonics and has been historically regarded as a mild sedative. L-5-Hydroxytryptophan is known to raise serotonin levels which, in turn, improves sleep quality. Dimethylaminoethanol is thought to increase the levels of acetylcholine (a neurotransmitter) in the brain. Vinpocetine (Vinca minor) is known to aid memory and improve mental functioning. Melatonin is a natural hormone that regulates sleep in mammals. The B Vitamins are known to regulate the body's energy processes and promote good health. Vitamin C is also a necessary ingredient in maintaining good health. Zinc is used by the body in hundreds of enzymes which regulate many of the body's functions. Vitamin D, Calcium and Magnesium are additional elements that are utilized by the body.

[0017] Relative weights for the amounts of the main ingredients of the nutritional supplement are as follows:

[0000]

Weight Range	Ingredient
300-600 mg	Calea zacatechichi powder
30-70 mg	Wild Lettuce Extract 4:1
30-70 mg	Mugwort Extract 4:1
10-40 mg	L-5-HTP
5-15 mg	DMAE
14-30 mg	Passionflower Extract 10:1
5-15 mg	Green Tea Extract 10:1
1.0-4.0 mg	Vinpocetine
300-600 mcg	Melatonin

[0018] Additional ingredients, by weight:

[0000]

Range	Ingredient
0.5-2.0 mg	Vitamin B-1
5-15 mg	Vitamin B-2
10-20 mg	Vitamin B-3
10-20 mg	Vitamin B-5
5-15 mg	Vitamin B-6
100-300 mcg	Vitamin B-9
50-200 mcg	Vitamin B-12
5-15 mg	Vitamin C
1-5 mg	Zinc
20-40 mg	Magnesium
50-70 mg	Calcium
1-4 mcg	Vitamin D

[0019] A preferred formula of the nutritional supplement by weight follows:

[0000]

	amount by weight
Primary Ingredients	
Calea zacatechichi powder	500 mg
L-5-HTP	20 mg
Vinpocetine	2.5 mg
Secondary Ingredient	
Melatonin	500 mcg
Tertiary Ingredients	
Wild Lettuce Extract 4:1	50 mg
Mugwort Extract 4:1	50 mg
DMAE	10 mg
Passionflower Extract 10:1	22.5 mg
Green Tea Extract 10:1	10 mg
Additional Ingredients	
Vitamin B-1	1.4 mg
Vitamin B-2	10 mg
Vitamin B-3	15 mg
Vitamin B-5	15 mg
Vitamin B-6	10 mg
Vitamin B-9	200 mcg
Vitamin B-12	100 mcg
Vitamin C	10 mg
Zinc	3 mg
Vitamin D	2 mcg
Magnesium	30 mg
Calcium	60 mg

[0020] When considering lucid dreaming, one of the key elements, especially at a high level, is that the dreamer not only realizes they are in a dream but that they have the potential to control the progress and outcome of the dream. The following examples are summaries of testimonials of people who took the nutritional supplement of the invention as a sleep aid and not only were able to lucidly dream but were able, through their heightened awareness and control, to change what had been recurrent patterns in their dreams and their lives.

EXAMPLES

Example 1

[0021] In this example, the individual kept having recurrent nightmares over a period of about 20 years. In the dream, the person would start across a bridge and be grabbed by a monster that would begin to crush him as he was being dragged under water. When he realized that he would soon drown, he would start screaming, waking not only himself but his bed partner, as well. This person started taking the nutritional supplement of the invention and, after several more dreams like this, gradually increased his awareness, or lucidity, to the point where he was able to confront the monster and eventually was able to exercise enough control over the outcome of the dream that he killed the monster. The nightmares stopped.

[0022] Apparently, the monster was a dream created manifestation of a real person that had abused him in the past. When the dreamer and the abuser faced each other in court to resolve a legal matter, he had gained the confidence to not be afraid of the abuser any more and won the case. In this instance, the nutritional supplement of the invention allowed him, through lucid dreaming, to conquer a long entrenched fear.

Example 2

[0023] The individual in this example had been similarly experiencing recurrent nightmares since he was about 10 years old. In the nightmare, this person was chased by a dark, faceless entity and would frequently awaken screaming. However, after beginning taking the nutritional supplement of the invention, he began dreaming more and more lucidly, to the point where he soon realized that he could control his reaction to the faceless entity. He soon was able to angrily confront the entity which allowed him to rid himself of this recurrent dream. He used his experience with the recurrent nightmare that he could exercise control, through lucidly dreaming, that he could adjust his normally negative reactions to anxiety producing events in his life into positive reactions, thus enabling him to function better in the world.

[0024] Accordingly, it is to be understood that the embodiments of the invention herein described are merely illustrative of the application of the principles of the invention. Reference herein to details of the illustrated embodiments is not intended to limit the scope of the claims, which themselves recite those features regarded as essential to the invention.

SUBSTANCES THAT ENHANCE RECALL AND LUCIDITY DURING DREAMING US2004266659

Inventor(s): LABERGE STEPHEN

BACKGROUND OF INVENTION

[0001] The term "lucid dreaming" was coined by Frederik van Eeden who used the word "lucid" in the sense of mental clarity. Lucidity usually begins in the midst of a dream when the dreamer realizes that the experience is not occurring in physical reality, but is a dream. Often this realization is triggered by the dreamer noticing some impossible or unlikely occurrence in the dream, such as flying or meeting the deceased. Sometimes people become lucid without noticing any particular clue in the dream; they just suddenly realize they are in a dream. A minority of lucid dreams (according to the research of LaBerge and colleagues, about 10 percent) are the result of returning to REM (dreaming) sleep directly from an awakening with unbroken reflective consciousness. The basic definition of lucid dreaming requires nothing more than becoming aware that you are dreaming. However, the quality of lucidity can vary greatly. When lucidity is at a high level, you are aware that everything experienced in the dream is occurring in your mind, that there is no real danger, and that you are asleep in bed and will awaken shortly. With low-level lucidity you may be aware to a certain extent that you are dreaming, perhaps enough to fly or alter what you are doing, but not enough to realize that the people are dream representations, or that you can suffer no physical damage, or that you are actually in bed.

[0002] In the circumstances described above, it is strongly desired in the art to develop a safe method of aiding dreamers attempts at more frequently achieving higher states of lucidity in their dreams. One class of preferred embodiments of this invention include neurotransmitter modulators such as Acetylcholine Esterase inhibitors.

[0003] A number of substances have been suggested to enhance the likelihood of lucid dreaming, from vitamins to prescription drugs. There are few good scientific studies to back such claims. Lucid dreaming is highly subject to the placebo effect; the belief that something will stimulate a lucid dream is very effective. Many prescription drugs as well as marijuana and alcohol alter the sleep cycle, usually by suppressing REM sleep. This leads to a phenomenon called "REM rebound," in which a person experiences intense, long REM periods after the drug has worn off. This can manifest as nightmares or, possibly, as lucid dreaming, since the brain is highly active. Drugs in the LSD family, including psilocybin and tryptamines actually stimulate REM sleep (in doses small enough to allow sleep), leading to longer REM periods. The potentially dangerous and illegal nature of these drugs is not conducive to their use as lucidity enhancers.

[0004] It is known in the scientific literature that Acetylcholine and its agonist as well as Acetylcholinesterase and its inhibitors/antagonist may be involved with REM and sleeping (Amatruda et al 1975; Baghdoyan et al 1984; Gillin et al 1985; Velazquez-Moctezuma et al 1991; Wauquier et al 1985). J. A. Hobson, on page 202 of his 1988 book, "The Dreaming Brain" states: "Cholinergic brainstem mechanisms cause REM sleep and dreaming". It has also been discovered that microinjection of the Acetylcholine agonist carbachol in the certain locations of the brain, such as the Pons, elicits extended periods of REM sleep, and that many of the neurons critical for REM sleep are responsive to Acetylcholine.

[0005] Neither Hobson, nor any other others, discuss the possibility of altering Cholinergic levels via Acetylcholine Esterase inhibitors as a means of enhancing dream recall and lucidity.

[0006] Historically, there are many cultures that have believed to have found naturally occurring substances that in some way alter dream consciousness. One such substance, Calea zacatechichi or "Dream Weed", is, according to Lilian Mayagoitia, et al, in a 1986 Journal of Ethnopharmacology article, "a plant used by the Chontal Indians of Mexico to obtain divinatory messages during dreaming." The neuro-active compounds of this herb are reported to cause thought to be sesquiterpene lactones, and as far as is currently known, unrelated to Acetylcholine Esterase inhibition or inhibitors.

[0007] Both ancient traditional medicines and modern popular speculation has assigned Oneirogenic, or "Dream-Inducing" properties to numerous herbs and other natural substances, including herbs like Valerian, Mugwort, Mullein, Kava Kava, Dittany of Crete, St. Johns Wort, Salvia Divinorum, Scutellaria Indica, Licorice Root, Vervain, Jasmine, Honeysuckle, Datura, Bee Pollen, Catnip, Hops, Sculicap, Mimosa, Lavender, Damiana, Withania Somnifera, Passionflower, Chamomile, Cardamom, Gotu Kola, Ginkgo Biloba, Ibogaine, Verbena, Rose, Cinnamon, Marigold, Nutmeg, Peppermint, Holly, Yarrow and Anise. Few if any of these have any scientific basis or support for these purported effects. Other supplements and/or drugs are also claimed to be linked to lucidity, including B-vitamins, Melatonin, DMAE, and the psychedelic DXM. Some of these drugs and herbs may be addictive, poisonous and/or illegal.

[0008] Lotsof, in U.S. Pat. No. 4,499,096, issued Feb. 12, 1985, teaches a "Rapid method for interrupting the narcotic addiction syndrome" via the drug Ibogaine. According to the patent, this drug induces "dream enhancement or hallucinatory effect". This claimed "oneirogenic" effect is due to the fact that it induces dream behavior with the ego perspective relatively intact. This is, as the patent mentions, more of a hallucinatory effect, not a lucid dreaming effect. Because of these hallucinatory and other side effects, however, this drug is illegal in the U.S., and is thus not a desirable means for enhancing sleep dream quality. This drug is also not considered to be related to Acetylcholine Esterase inhibitors or 'smart drugs'.

[0009] DMAE (2-dimethylaminoethanol) is a chemical that has been suggested by used to treat a number of conditions affecting the brain and central nervous system. It is thought to work by increasing production of the neurotransmitter Acetylcholine, although this has not been proven. Marketed as a memory and mood enhancer, DMAE is believed to enhance intellectual functioning. There are few good clinical studies supporting this belief. Such substances are known as "cholinergics" because of the belief that they increase Acetylcholine. They have been traditionally used to treat diseases such as Alzheimer's dementia, tardive dyskinesia, and Huntington's chorea. Because DMAE is believed to be a cholinergic, it has been suggested for these disorders, even though placebo-controlled studies have provided essentially negative results. There is continued controversy over whether DMAE actually increases Acetylcholine. Nonetheless, Sergio, W. Claims in the 1988 August edition of Medical Hypotheses, in the article: "Use of DMAE (2-dimethylaminoethanol) in the induction of lucid dreams", subjective experiences of himself and his wife experiencing enhanced lucidity through use of this supplement. It is unclear from the article to what extent his purported results derive from placebo response or any specific or cholinergic effect.

[0010] Also well known in the art is U.S. Pat. No. 5,507,716, awarded to LaBerge, et al. on Apr. 16, 1996, for Equipment and methods used to induce lucid dreams in sleeping persons. In this patent, a device was employed to assist people to achieve lucid dreams via the detecting and monitoring the eye and head movements of sleeping persons, where eye movement activity in the absence of head movement is used to indicate the presence of REM sleep. By then applying sensory stimuli to sleepers in REM sleep can cue them that they are dreaming without producing awakening. Other persons have developed equipment, such as Keith M. T. Hearne who illustrated and described his respiratory measuring device in his 1983 U.S. Pat. No. 4,420,001. His device sensed temperature changes of a person's respiration in his or her breathing passageway, or in airflows to and from his or her breathing passageway. Thermistors were used, in an electrical circuit, to sense the temperature changes of the person's respiration. When the rate of these temperature changes reached a high predetermined level, the signals created in the electrical circuit initiated an audible sound, either to help arouse a sleeping person from an unpleasant dream by awaking them or to help them enter into a lucid dream state.

[0011] None of the preceding references disclosed describe a method of enhancing lucid dreaming comprising administering to individuals the Acetylcholine Esterase inhibitor class of drugs. Use of therapeutic agents for Alzheimer"s Disease such as Donepizil (AriceptÂ(R)), Rivastigmin (ExelonÂ(R)), Galantamine (ReminylÂ(R), NivalinÂ(R)), Tacrine and Huperzine leads to low incidence of adverse events, such as insomnia and gastrointestinal symptoms, while significantly enhancing dreaming cognitive clarity, lucidity, self-reflection, recall, control, bizarreness, and visual vividness.

SUMMARY OF INVENTION

[0012] This invention relates to the field of Lucid Dreaming and the enhancement of dream recall and dream lucidity through memory enhancing drugs, including the class of substances that comprise

Acetylcholine Esterase inhibitors (AChEIs). Lucid dreaming involves dreaming while knowing that you are dreaming. AChE inhibitors (AchEI's) inhibit the normal metabolic inactivation of Acetylcholine (ACh) by inhibiting the enzyme, Acetylcholine Esterase (AChE), leading to accumulation of Ach. AchEI's are most commonly used to enhance memory, particularly in patients suffering from Alzheimer's disease. Ach is also well known to be important in REM and thus is suggested herein to enhance dreaming and lucidity.

DETAILED DESCRIPTION

[0013] While investigating the efficacy of donepizil (AriceptÂ(R)) as a means to enhance lucid dreaming and cognitive clarity during REM sleep in normal subjects, the following experiment was performed with the following results: A randomized, double blind, placebo-controlled crossover trial was performed with ten normal volunteers self-selected for high dream recall and interest in lucid dreaming (7 male, 3 female, age 22-55). Subjects collected dream content and other self-report measures on three nights, separated by a washout period of at least one week. At bedtime, they took capsules containing 0 (placebo), 5, or 10 mg of donepizil, with counter-balanced order of the three dosages. After each spontaneous awakening during the three nights, Ss evaluated dream content on a range of measures including bizarreness, complexity, and intensity, affect, cognitive clarity, lucidity, and control. They also estimated sleep quality, insomnia, and degree of any adverse effects.

[0014] RESULTS: Nine of the 10 Ss (90%) reported one or more lucid dreams on the experimental nights, with only one S reporting a lucid dream on a placebo night. Cognitive clarity, lucidity, recall, control, bizarreness, and visual vividness were all significantly elevated with donepizil compared to placebo. The effects were generally dose-related with 10 mg donepizil producing significantly higher levels of these variables than the 5 mg dose, which in turn produced significant elevations compared to placebo. Odds of lucid dreaming for the three conditions increased from 0.031 for placebo, to 0.429 for 5 mg donepizil, and 0.754 for 10 mg donepizil. The lucidity odds ratio for the 10 mg dose compared to placebo was 24.3 ($p<0.001$). Donepizil was also associated with a significantly elevated frequency of sleep paralysis and a 40% increase in estimated time awake during the night (176 vs. 126 min of a 9.0 hr night, $p<0.05$). The incidence of adverse events, notably mild insomnia and gastrointestinal symptoms, was low and primarily associated with the higher dose condition (with two subjects reporting nausea and one vomiting).

[0015] Thus, compared to placebo, both 5 and 10 mg dosages of donepizil significantly enhance self-reported cognitive clarity, self-reflection, and lucidity during dreaming. These effects may be further enhanced by combination of the acetylcholinesterase inhibitors with Acetylcholine precursors, agonists and/or lucidity inducing electronic devices.

[0016] In addition to donepizil data, we have data on ExelonÂ(R) (rivastigmin; 6-12 mg dose) and galantamine (ReminylÂ(R), NivalinÂ(R)8-16 mg) as well. Both work as well as AriceptÂ(R)), but with perhaps fewer side effects. We have also tested huperzine with promising results, but haven't yet worked out the dose correspondence with AriceptÂ(R) (but 5 mg Donepizil is more potent than 150 Apg huperzine).

[0017] In addition, we have reports of success with nicotine patches. There is also in the scientific literature studies showing triggering/intensification of REM sleep with arecoline. How much evidence do we need for these general claims: It should be noted that the REM altering, lucidity enhancing effects may be caused by a number of neurophysiological effects induced by these "smart drugs" besides Acetylcholinesterase inhibition. These include: 1. Using a cholinergic agonist (e.g., nicotine).

[0018] 2. Using muscarinic receptor agonist: Drugs that mimic the effect of ACh on muscarinic receptors: Inc: muscarinic (M1), M2 and nicotinic Agonists. (Such as arecoline or Recoline, a muscarinic receptor agonist)3. Using an antagonist of presynaptic receptors to activate the remaining

Acetylcholine neurons. (i.e. In rats, antisense oligonucleotide sequences that block Muscarinic M2 (but not M4) receptors increases extracellular Acetylcholine. This effect (Antisense nucleotide sequences are complementary to a sequence of messenger RNA. When antisense DNA or RNA is added to a cell, it binds to a specific messenger RNA molecule and inactivates it.) 4. Using allosteric modulators (such as allosterically potentiating ligands [APLs]) of Acetylcholine and nicotinic receptors (drugs that interact with the receptor through binding sites that are distinct from those for Acetylcholine and nicotinic agonists and antagonists).

[0019] Other strategies for potentiating Acetylcholine function for dream lucidity enhancement, including the use of other classes of compound working in similar ways should be apparent to those skilled in the art from a consideration of the enclosed descriptions.

Inducing and maintaining lucid dream state, uses stimulus response for testing state of consciousness, psychological conditioning and generating lucidity

DE19920376

Inventor(s): HORNUNG HEINZ

The invention relates to a method and apparatus for achieving and maintaining a lucid dream state. The process / device offers the possibility to achieve through the use of technical means the state of lucid dreaming (lucid dream). By this is meant that the dreaming of the state in which it is located, is aware that he would say that his present experience is a dream. This finding is a "normal" dreaming inaccessible. Furthermore, it offers / her the opportunity to retain or regain this status. First, the historical development of lucid dreaming is appreciated: While in the eastern culture very early discovered and applied the possibility of lucid dream experience - for Example, in the eighth century by Tibetan Buddhists, is this phenomenon in western culture until much later - the end of the 18th Century - at the moment few people attention. 1960-1970 was only about this topic in the modern scientific research input. In recent years, a relatively large number of related publications, published in academic journals and in book form. All methods for achieving lucid dreams described in this literature is based on the following principle. By frequently as possible (ideally permanent) Check the state of consciousness in the waking state (for Example, with the questions: Am I dreaming, I wake up? etc.) is a mental conditioning can be achieved on these issues, ie a special ability to buy this in the - that this finds its way into the dream experience hope - legitimate. This approach may thus lead to the desired result, namely the clear Will the dream, if the dream I have the same questions is how confident I am that day. However, experience shows that these practices are extremely laborious, requiring a lot of time, a very large degree of discipline and lead also unfortunately only rarely successful. It should be added that experience has shown that keeping the clarity of dream is often more difficult than achieving the same. The clarity is easily lost again and then you find yourself back in the normal dream state.

For the published prior art, the following is noted: From U.S. 5,507,716, U.S. 5,551,879 and FR 2,699,412 Devices and methods are known, with the help of lucid dreaming can be caused when a sleeping person. Here, the sleeping person are given through a face mask or a kind headband stimuli that should trigger a review of the state of consciousness. These devices offer the opportunity to register eye movements and thus REM sleep phases (here are more common dreams) to detect. If there is such, optical, or acoustic signals are triggered. However, these devices affect the one hand the sleeping person in their freedom of movement in no small measure, on the other hand, and this is very much a combination with the above described procedures appears (the conditioning during the waking state) and only to a limited extent, since they are not in the normal, everyday routine can be integrated. This possibility, DE 196 47 019 (same inventor / assignee as the present invention / application). It is the closest prior art. Here, a method and an apparatus is described which / to be applied during bedtime and waking the whole of a person. During the waking state the person (with the help of technical means, such, In the manner of a wristwatch with a vibration signal) to any selected time points (e.g. Example, using a random generator, so the user completely unexpected) External stimuli supplied. The using person checked at each stimulus their state of consciousness, ie there is a conditioning to that stimulus instead. If the person now mostly identical stimuli (which are below their awakening threshold, and are preferably added in the early morning hours because the longest and most frequent REM periods occur) during sleep is supplied, it will be on the learned on the day way (a conditioned reflex after learning theory) respond to this stimulus, namely by checking their state of consciousness, and they will then conclude with a good chance that they are dreaming must. At the times of the stimuli that are triggered in this process during sleep, the following should be noted: The charms can - take place in the (general) sleep state and may not necessarily fall into a REM stage - at any time.

According to recent findings of the suppression Fund dream research is the equation of "dream sleep REM sleep equal" wrong. It is dreamed in all the other stages of sleep. If the stimuli do not fall into a dream state, the user notices nothing flat, and only the stimuli that fall into a dream state, lead to the desired effect. The proportion of time spent dreaming periods of total sleep duration is so high that even with a random choice of the timing of the stimuli fall enough stimuli to dream phases. The applicability of the first-mentioned three devices is essentially limited to the bedtime of a person and is mainly based on the scientifically validated knowledge that sleeping outside stimuli register. Since, as mentioned, a conditioning to the stimuli (light flashes, sounds) that are released from the sleep mask / the headband in the guard time hardly seems possible, the question remains on which way processes the person these charms in her dream, that is, opportunity with which these stimuli lead to a lucid dream. Significantly better chance to really achieve a lucid dream state also offers the DE 196 47 019 by the applicability of this device is extended to the whole waking state. The desired response in a dream to a stimulus (ie, checking the state of consciousness) can now be learned easily in the waking state. Based on this, the present invention seeks to provide a method and apparatus that / the shortened compared to the above methods / devices, the state of consciousness verification time and relieved and / the total increases the chance lucid dream states to reach. Furthermore, the / the present method / device to hold, or recovering the clarity in lucid dreams is to enable. This requirement of maintaining the clarity and of recovering the same is taken into account by any of the devices mentioned above. The object is achieved by a method having the features of claim 1 and by an apparatus having the features of claim 23rd Advantageous embodiments are specified in the dependent claims.

In DE 196 47 019 A method and apparatus is described in which the user's response to a single stimulus outside the review of the state of consciousness, ie, is a purely mental activity (in the waking state and during sleep). It is a simple, but fundamental difference between waking and dream states is neglected here, namely, that action leads in the waking state, as opposed to actions in the dream, to changes in the real environment. This makes the / the present method / device advantage. It / she offers the user the opportunity to advance in the specified Reizgebungsverfahren (ie, for Example, when, how and how long stimuli are given) to intervene at any time and to change it temporarily. The basic function is not affected by the surgery and will resume automatically once the operation is completed, possibly with a lag. On the one hand, a stimulus (of which the date for Was as specified by the base function) during the period in which it acts on the person changed, eg As it can be stopped by means of a special key, on the other hand can be triggered randomly at any time an extra stimulus. Trying to consciously acting to intervene in the Reizgebungsverfahren will succeed in the waking state, of course, but inevitably fail in the dream state. Thus, a situation is created which allows an immediate and clear assessment of the state in which it is located (waking state or dream state). Succeed the test of the intervention (eg Example, by stopping a currently occurring stimulus), one knows one is awake. If you fail the test, you know that you must be dreaming. The sometimes lengthy Check the currently experienced situation with the doubts and uncertainties that occur often omitted, the desired state assignment is immediately possible and absolutely clear now. Become stimuli triggered while the user is already in an at least partially lucid dream state, in this case the failure of the attempts at transforming the stimuli for strengthening, recovering and maintaining the clarity will be extremely helpful. A common, custom steady release of the above mentioned extra stimulus in the waking state, the dream state also lead to more frequent attempts to initiate this stimulus. The inevitable failure of this option also allows an immediate and clear state assignment. This increases the chance again, using the / of the present method / device to achieve or maintain lucid dream states. Even with these extra stimuli is of course the possibility of change of stimuli, which further increases the chances of success / the method / apparatus. Very useful, the ability to trigger an extra stimulus, also for the following reason.

As already mentioned, the holding of clarity during a lucid dream is often more difficult than achieving the same. Is now in such a time of doubt and danger of losing the clarity, trying to trigger an extra stimulus, the clarity is either recovered immediately, or if, contrary to expectations should not occur alone the associated view, the feeling, the hands very helpful (when using the device in the form of a wristwatch), since according to the initially mentioned method for achieving lucid dream state is a very important method for maintaining the clarity without technical aids viewing and feeling of the hands.

This is equally well with the above-mentioned possibility of change (eg Example of stopping) of a single stimulus of importance. The / the inventive method / device offers a more substantial relief to preserve the clarity. It / she offers the possibility to follow up on every single stimulus automatically in certain, freely selectable intervals other stimuli. Thus, a single stimulus of such B. 5 to 10 sec Duration at a distance of, for As one or two minutes a further stimulus 5-10 sec Duration done, etc. Instead of a single stimulus is now seen as a consequence of stimuli (duration, for B. 5 to 10 min) cast. Here, (abandonment of a single stimulus) also such a sequence of stimuli to be completed (while awake) with the above-mentioned stop button.

As in the dream state, such a termination is not possible, such a sequence of stimuli, if the clarity threatens to be lost, be very helpful to keep or regain this.
Can reaching the clarity such a sequence of stimuli to be extremely helpful, if first stimuli should not have led to the desired effect, if such B. did not fall into a dream state, or the clarity was only partially achieved. Furthermore, such a sequence of stimuli - according to the way the individual stimulus - be triggered at an arbitrarily selectable time. This may, for As happened with a special key. Here, an arbitrary delay (for , 5 to 30 min) to the first stimulus of the sequence begins, can be specified. Pressing this hot key so leads with a time delay to a sequence of individual stimuli.

It is known that the falling-asleep, whether in the evening, with interruptions (especially in the early morning hours) or during an afternoon nap in a special degree to is to achieve a more or less direct transition from twilight in a lucid dream. Here, such a sequence of individual stimuli be with delayed onset a very big help to achieve the clarity and keep. According to one embodiment, each breaks again press this button, the currently running special sequence of stimuli, even during the lag phase, from, and will automatically reboot with the given initial conditions. This is very useful if you are not asleep during the selected time delay. The very significant advantage of this opportunity, a sequence of stimuli delayed the onset of a stimulus sequence, is that herewith - as opposed to random on the night distributed stimuli or stimulus sequences - a triggering stimuli to times when the chances of success very high are is selectively enabled. The device, as described in DE 196 47 019 can also be combined with other devices, and give due to physiological measurement data stimuli or stimulus consequences. These measurement data of the other devices may be, for Example, to rapid eye movement (REM), act REM sleep EEG, heart rate increase, etc.. Additional devices may be external or integrated in the device, for Example, as a heart rate monitor. Default settings for the device, such As stimulus duration, stimulus timings etc. can be preset, on the device itself, or for B. also be a connection to a PC with appropriately developed software to be entered or changed. For the purpose of / the method / apparatus following may be cited: First, it should be mentioned here that so far are not aware of any negative side effects of lucid dreaming. Experiencing a lucid dream is consistently (which incidentally mentioned very clearly is) very pleasant and dream about the experience, as well as in the subsequent memory in the waking state, as experienced "mind-expanding". This alone makes the commercial utility of this device, specifically for home users, more than clear.

Furthermore, there are many possible applications in the field of epistemological research, especially of consciousness research, as well as in use in the psychotherapeutic treatment (eg Example, nightmares). Further details of the procedure, in particular for education of the stimuli, and the device, in particular for the design of the sensor and the controller, the combination possibilities of the device with other devices 196 47 019 A1, seen from DE, which is fully incorporated by reference.
The encoder of the present devices has all the properties described in DE 196 47 019....

Method of obtaining lucid dreaming condition
DE19647019

Inventor(s): HORNUNG HEINZ DR

The method involves giving a number of stimuli to a person, which can be recognised through one of his senses. The person's response to the stimuli give proof of the conscious state of the person and conditions the person's response, as a conditioned reflex. When this person is in a state of sleep, the same stimuli are given to him to evoke the same responses as those which were obtained while he was awake. The response to the stimuli bring about a state of lucid dreaming in which the person knows that what he experiences is a dream.

DESCRIPTION DE19647019

The invention relates to a method and apparatus for achieving a lucid dream state.
The / be presented, the novel method / device for the first time offers the opportunity through the use of technical means the state of lucid dreaming (lucid dream) reach.
By this is meant that the dreaming of the state in which it is located, is aware that he would say that his present experience is a dream.
This finding is a "normal" dreaming inaccessible.

Historical development of lucid dreaming

While in the eastern culture very early discovered and applied the possibility of lucid dream experience - for Example, in the eighth century by Tibetan Buddhists (Gillespie, G. (1988) Lucid dreams in Tibetan Buddhism. In: JI Gackenbach and S. LaBerge (Eds.), Conscious Mind, Sleeping Brain: Perspectives on Lucid Dreaming. New York:. Plenary) - takes place this phenomenon in western culture until much later - the end of the 18th Century (LaBerge, S. (1985) Lucid Dreaming. New York: Ballantine Books. German edition: Wide awake in a dream. Munich: mvg-Verlag, 1991) - at the moment few people attention. Increased interest in scientists and philosophers since the end of the 19th Century determine (Hervey de Saint-Denys, MJL (1987) Les Rèves et les Moyens de les things. Paris: Amyot. Reprinting 1964: Cercle du Livre Precieux, Paris. Übers V McCreery. Engl Edition: Dreams and How to Guide Them, Translator v. N. Fry. London: Duckworth, 1982; Myers, FWH (1887) Automatic writing - III. Proceedings of the Society for Psychical Research, 4 Part II, 209-261; Delage, Y. (1919) Le Rève. Paris: Les Presses Universitaires de France, Arnold Foster, M. (1921) Studies in Dreams. London: Allen & Unwin; Brown, AE (1936) Dreams In Which the dreamer knows he is asleep. Journal of Abnormal Psychology, 31, 59-66; Moers Messmer, H. von (1938) Dreams with the simultaneous realization of the dream state (Dreams with concurrent knowledge of the dream state). Archives of Psychology, 102, 291-318). Since 1970 this topic is in the modern scientific research input (Green, CE (1968) Lucid Dreams. London: Hamish Hamilton). In recent years, a relatively large number of related publications - in scientific journals and in book form - appeared (range: Tholey, P. (1983) Techniques for inducing and manipulating lucid dreams. Perceptual and Motor Skills, 57, 79-90; Tholey, P. (1988) A model for lucidity Training as a Means of self-healing and psychological growth. In: Gackenbach and see LaBerge, LaBerge (1985); Brylowski A. (1986) H reflex in lucid dreams. Lucidity Letter, 5 (1), 116-118; Brylowski, A. (1990) Nightmares in crisis: clinical applications of lucid dreaming techniques. Psychiatric Journal of the University of Ottawa, 15, (2), 79-84; Gillespie (1988), and J. Gackenbach Bosveld J (1989) Control your Dreams. New York: Herper & Row. German edition: ruler of the realm of dreams. Brunswick. Aurum Verlag, 1991, Hearne, KMT (1990) The Dream Machine: Lucid Dreams and how to control Them. Wellingborough, Northamptonshire: The Aquarian Press, Green, CL and Charles McCreery (1994) Lucid Dreaming, London, Routledge. German edition: Navigate dreams consciously, Frankfurt, Wolfgang Krüger Verlag, 1996). Unfortunately, research has almost exclusively in the English language with lucid dreaming, but there are German translations of some books before that, for the German research and public Currently, however, still receive little attention.

State of "technology" to achieve the lucid dream state

The previously known and applied by few persons methods to achieve lucid dreams are all based on autosuggestion. The ability to dream lucid (clear, ie conscious) can be only one "work on themselves" are obtained in the waking state. The principle underlying all these methods is the following: By frequently as possible (ideally permanent) Check the current state of consciousness - the real experience - a day (for Example, with the following questions: Am I dreaming? I wake up? Where am I? How old am I? Is it the person with whom I speak at the moment in my current life, or was there it at an earlier time? Do I have to wonder about anything? Etc ...) to a mental conditioning to be made on these issues, ie a special ability to buy and this in the hope that this finds its way into the dream experience. This hope is justified, since it is known that recur both everyday and especially intensely felt experiences of the day as so-called day residue in dreams (Freud, S. (1925) The Interpretation of Dreams, after the issue of 1972. Frankfurt am Main: Fischer Taschenbuch Verlag, 1991; dressing and Riemann (1994) Diagnosis and treatment of sleep disorders. Stuttgart Jena: Gustav Fischer Verlag Lauer et al. (1987) Shortened REM latency: A Consequence of psychological strain. Psychophysiology 24, 262-271). This approach may thus lead to the desired result, namely the clear Will the dream, if the dream I have the same questions is how confident I am that day (detailed explanation and description of these procedures is given, for Example, Green (1994)).

These procedures are according to experience extremely laborious, requiring a lot of time, a very large degree of discipline and lead also, unfortunately, only very rarely successful. A good result means that one will reach a few times a month, the state of lucid dreaming . One author reported three lucid dreams in eight years constant effort (Myers (1887)). From U.S. 5,507,716 Devices and methods are known, with the help of lucid dreaming can be caused when a sleeping person.

Here, the sleeping person will be given a sleeping mask stimuli that should trigger a review of the state of consciousness. Such a mask affects the sleeping person in their freedom of movement in no small measure.

The role and application of / the inventive process / device Based on this, the object of the invention to provide a method and apparatus that facilitate the achievement of lucid dream states or allow much in the first place. The object is achieved by a method having the features of claim 1 and by an apparatus having the features of claim 14 Advantageous embodiments are specified in the dependent claims. The above-described "work itself" is very much facilitated by the / the inventive method / device.

Many people will this / allow s method / device to achieve lucid dreams in the first place. The / the new process / device based on principles of learning theory, it is a training for practicing a conditioned reflex (eg Textbook example of the psychology of Medinzinischen sink, H. et al (1977), Munich, Vienna, Baltimore: Urban and Schwarzenberg; psychology textbook by Zimbardo / Ruch (1978), Berlin, Springer Verlag).

In the waking state are the using person - given stimuli whose time points for - preferably from outside As determined by a random generator. The individual charm reminds the person from their current intense state of consciousness (I watch? Am I dreaming? Etc ...) to question, ie the using person conditioned to this stimulus, in other words, it learns in the waking state to a certain stimulus with a very specific answer to respond, namely by questioning their current state of consciousness intense. This requires only a period of a few seconds. These charms are for both day , 10 to 20 times repeated during a more or less constant repetition is necessary in the methods used so far. According to scientific learning theory can be no doubt about the success of the vorzustellenden method. It may be mentioned here because of its popularity, the experiment of Pavlov.
This has the concept of classical conditioning, ie the formation of a conditioned reflex, already in 1900 set up (animal experiments).

Many subsequent studies have confirmed the validity of this learning principle for humans (Textbook of Medical Psychology, Enke, 1977; psychology textbook, Zimbardo / Ruch, 1978). The basic idea of the method / apparatus based on the fact that the learned in the waking state conditioned reflex can also cause during a dream, as the ability to react to stimuli, obtained even in the absence waking consciousness is like the dream research proves (shrub / Meier (1992) The dreams on the track. Results of experimental dream research, Bern: Verlag Hans Huber). For these reasons, there can be no doubt about the success of the vorzustellenden method.

For the purposes of / the method / apparatus does this mean the following. During sleep preferably identical outer stimuli are as given in the waking state. The intensity of the external stimuli is determined by the using person in the waking state so that these stimuli remain under their awakening threshold, but are nevertheless perceived by her in the dream state. (Ie, checking the state of consciousness, which with very good chance now leads to the conclusion that the dreamer on the basis of diverse absurdities of dream events come to the conclusion that he must be in a dream) allows the learned in the waking state response to this stimulus so the transition to the lucid dream experience. As with any technology device, and A method (for As cycling, driving a car), the handling must therefore be learned here, with the effort to keep discipline, effort and time with the / procedures / device within narrow limits. According to the current state of knowledge of research on the learning of a conditioned reflex is the first occurrence of lucid dreams after days, or later than expected a few weeks, depending on the susceptibility of using person. In the further application of the device, the user will in many cases after a certain time can (with increasing learning effect) waive the application of the waking state.

The purpose of the apparatus

First to be on negative effects of lucid dreaming, which could possibly be feared, received here. A short-term risk factor could be seen in an impairment of the normal recovery function of sleep. Furthermore, would imagine that lucid dreaming could in the longer those for user people have a negative effect on the attitude towards daily life.

The current state of knowledge, the research provides no evidence for both fears (for detailed discussion on the subject in Green 1994). Experiencing a lucid dream is consistently experienced by lucid dreamers to be very pleasant and as for the waking experience and the dream experience as "mind-expanding". This already established the existence of a commercial utility of this / method / apparatus, since it opens up the possibility connected with very little effort, a new "dimension" of experiencing interested persons, and thus can also offer the opportunity for personal development. Moreover, great opportunities in the epistemological research and consciousness research results. Likewise, several possibilities are conceivable to incorporate lucid dreaming into the psychotherapeutic treatment of patients. So good results in the treatment of nightmares have already shown that in patients with nightmares of unknown cause in patients with post-traumatic appearance (such Example, after an accident, rape, torture, etc.), and do not forget to children and young people who often suffer from nightmares, lucid dreaming could be a very promising treatment approach (detailed in Green 1994). But here it is again emphasized that the vorzustellende invention can be applied by anyone without risk, so it is an invention for personal use and not a medical device. Overall, let the many possibilities that could provide lucid dreaming, according to experts at the time not yet foresee...

Equipment and methods used to induce lucid dreams in sleeping persons
US5507716

Inventor(s): LABERGE STEPHEN P [US]; RICH ROBERT S [US]; WRIGHT DAVID K [US]; KOTTKE DANIEL G [US] + (LABERGE, STEPHEN P, ; RICH, ROBERT S, ; WRIGHT, DAVID K, ; KOTTKE, DANIEL

Equipment and methods employed to assist people to achieve lucid dreams (dreams in which one is aware that one is dreaming) function by detecting and monitoring the eye and head movements of sleeping persons for the presence of sufficient eye movement activity in the absence of head movement to indicate the presence of REM sleep and then applying sensory stimuli to sleepers in REM sleep, which if incorporated into their dreams, can cue them that they are dreaming without producing awakening. With this equipment and methods, people are able to have more lucid dreams than they otherwise would. A preferred embodiment includes a face mask that contains two infrared emitter-detector pairs, one for sensing eye movements and one for sensing body movements, a state-test button, and components that produce low intensity sensory stimuli such as light and sound. A microprocessor monitors the fluctuating voltage from the infrared pairs for the occurrence of a predetermined sequence of voltages, adjustable in software, and when the correct sequence occurs, triggers the stimuli-producing components in the mask into activity, to cue the sleeper to become lucid. The methods include many procedures and adjustable parameters for optimizing the timing of the application of the sensory cues to the sleeper. Among these procedures and parameters are those used to locate the time in the sleeper's REM periods when the cue will be most effective, and those used to apply the sensory stimuli at these optimal times.

BACKGROUND

As set forth in the book entitled Exploring the World of Lucid Dreaming, written by Stephen LaBerge and Howard Rheingold, the idea of cultivating a state of mind while awake for the purpose of carrying it into a dream state as a means of inducing lucid dreams, has been used by Tibetan Buddhists for more than a thousand years. However as recently as the late 1970's, experts were convinced that dreaming with consciousness that you were dreaming was a contradiction in terms and therefore impossible. At this recent time Stephen LaBerge began his Ph.D. studies of lucid dreams at Stanford University to establish proof that lucid dreaming was real, by obtaining evidence from the dream world that a person knew he or she was dreaming during sleep.

During his studies, he, with the assistance of other persons, has developed experimental testing equipment for helping him and other persons to obtain such evidence from the dream world. In so doing, the developed experimental testing equipment, in many embodiments, has also been designed and built to help persons who are lucid dreamers, or wish to try to become lucid dreamers, to have lucid dreams and to have them more often and to have them more effectively. These lucid dreamers are referred to as being Oneironauts.

Also other persons have developed equipment, such as Keith M. T. Hearne who illustrated and described his respiratory measuring device in his U.S. Pat. No. 4,420,001 of 1983. His device sensed temperature changes of a person's respiration in his or her breathing passageway, or in airflows to and from his or her breathing passageway. Thermistors were used, in an electrical circuit, to sense the temperature changes of the person's respiration. When the rate of these temperature changes reached a high predetermined level, the signals created in the electrical circuit initiated an audible sound, either to help arouse a sleeping person from an unpleasant dream by awaking them or to help them enter into a lucid dream state.

At this time, a number of different techniques have been used to monitor sleep, in both the quiet phase and active phase, the latter also being called the rapid eye movement phase of sleep and referred to as REM sleep. These techniques include applications of: electrodes for EEG, ECG and EMG; infra-red sensing; and respiration measurement with strain-gauges or thermistors. Among these techniques, Stephen LaBerge and those helping him sought a method to reliably sense the presence of REM sleep, so that a device could automatically signal to a sleeper that he or she is dreaming. The requirements for such a device were portability, ease of use, low cost, and above all reliability and the successful induction of lucid dreams, which are those dreams wherein one is aware that one is dreaming.

SUMMARY

Stephen LaBerge and those helping him created better methods to help persons lucidly dream, and, in so doing, he and those helping him developed various embodiments of experimental test equipment that is used to make sleeping persons aware they are in the active phase of their sleep, and they are having a dream, while they still continue to dream or become lucid, and such equipment senses eye movements. They selected sensing movements of a person's eyes, via movement of his or her eyelids, because by using eye movement sensing, especially when combined with body movement sensing, and thereafter accounting for such body movements, to thereby ascertain accurately eye movements, per se, they developed very reliable test equipment for detecting the phases of sleep, and especially the active phase of sleep, called the rapid eye movement phase, which is further conveniently referred to as being REM sleep.

Their development of their methods and the experimental test equipment centering on eye movement sensing, undertaken by sensing eyelid movements, which so indicate eye movements, extended over a considerable period of time. Now equipment is soon to be made available for many persons to use in practicing their methods. In respect to a preferred embodiment of the equipment each person, during his or her sleep, is provided with a comfortably worn face mask covering face portions, and a headband supports the face mask about the person's head. The face mask, in turn, supports components of this equipment, which sense a person's eyelid movements during sleep, and which, then at the proper time, signals the person that he or she is dreaming, and thus should be able to be lucidly dreaming. The signal, preferably a light signal, occurs without awaking the person. Also a sound signal is selectively and optionally used.

The sensing components supported on the face mask utilize a low level infrared emitter positioned on the face mask to direct infrared light to the eyelid of a sleeping person, and a low level infrared detector to receive the reflections of this infrared light from the surface of this eyelid of this sleeping person. The signalling components supported on the face mask preferably utilize small flashable lights and/or small speakers to signal the sleeping person that he or she is dreaming, and thus should be able to be lucidly dreaming, without awaking the person.

A like arrangement of a low level infrared emitter, also positioned and supported on the face mask, but over a part of a person's face, which does not necessarily move with a person's eyelids, is also used with a low level infrared detector also positioned and supported on the mask, to detect face, head, or body movements, which are independent of a person's eyelid movements. These body movements are accounted for to avoid erroneous conclusions as to the eyelid movements, per se, and thereby to accurately determine whether or not a person is dreaming, or could be lucidly dreaming.

The other components of this preferred embodiment of the equipment are arranged together as a shelf or table supported subassembly, with circuitry extending to the sensing and signalling components supported and positioned on the face mask, These other components include: analog signal conditioning group of components, in turn including alternating current coupling, a preamplifier, a low pass filter, a computer gain setting unit, and an integrating unit to receive signals from the infrared detector detecting eyelid movements; an analog-digital converter for converting the analog signals of the eyelid movements into digital signals; analog signal conditioning group of components, in turn including an alternating current coupling, a preamplifier, a low pass filter, a computer gain setting unit, and an integrating unit to receive signals from another infrared detector, detecting face, head, and/or body movements; an analog-digital converter for converting the analog signals of the face, head, and/or body movements into digital signals; a computer to receive the digital signal information from both the analog-digital converters, to receive signals via software instructions, to direct data to data storage, to direct signals to activate sensing lights, and to activate sensing speaker sounds...

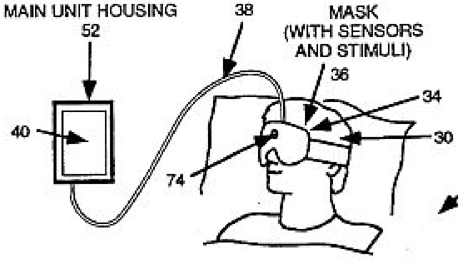

FIG. 1

LUCID DREAM STIMULATOR, SYSTEMS, AND RELATED METHODS
US2017304583

A dream stimulator may include a substrate and a plurality of electrodes positioned along a first side of the substrate. The stimulator also may include a stimulator coupled to the substrate. At least one lead may couple the plurality of electrodes to the stimulator. Additionally, a band may be positioned along a second side of the substrate and about each of the plurality of electrodes.

SYSTEMS AND METHODS FOR DELIVERING SENSORY INPUT DURING A DREAM STATE
US9620027

Embodiments of the invention provide apparatus, systems and methods for detecting neurological activity indicative of a dream state of a human. Many embodiments of the invention provide apparatus, systems and methods for detecting neurological activity of a human indicative of a dream state or the onset thereof and delivering an input to the user (such as an audio or other sensory input) during the dream state. Particular embodiments of the invention provide systems and methods for detecting neurological activity indicative of the onset or occurrence of a dream state of a human and delivering an audio or other sensory input during the user's dream state. The audio input may be used for learning, delivering messages to the user's subconscious, and/or promoting a state of relaxation.

EROTIC DREAM INDUCTION APPARATUS
US2016331924

An erotic dream induction apparatus includes an inflatable air bladder, such as a rubber balloon, to induce sexual arousal via genital stimulation. The air bladder is brought into contact with either the penis or vagina of the user, and/or is inserted into the anus and/or vagina of the user. In one embodiment, the air bladder is brought into contact with the penis, vagina, and/or anus of the user by being maintained in position near, against, or around the penis, vagina, and/or anus by means of straps or pockets in stimulation underwear, such that when the air bladder is inflated, the air bladder makes contact with at least a portion of the penis, vagina, and/or anus of the user. The air bladder is repeatedly inflated and deflated by means of pumps and/or valves.

Dream Enhancement Apparatus and Method
US2016310698

Dream stage enhancement uses a headband with EEG-EOG sensors, onboard processors, memory, coarse and fine time REM waveform detection modules, LEDs and an audio playback unit. After normalization to the user's EEG waveforms, the user's EEG-EOG signals are processed, REM and NREM stages detected and light, sound or AV stimuli are presented to the user based upon user-supplied light-sound-AV stimuli commands. To provide a reality check control ("RCC"), the head unit has a user actuatable RC interface whereby during sleep, RC stimuli are presented when the user depresses the RCC control which plays back the user supplied stimulus. In a "learning mode," the user selects "Recall" or "No Recall" ("NR") after the sleep period. If NR, then the system changes the color of light stimuli, light intensity, flash, audio sound type, audio intensity, and AV. If "Recall" the user supplied stimuli commands are carried out.

INDUCING LUCID DREAMING
WO2016209078

The present invention is in the field of induction of lucid dreams by Transcranial Alternating Current Stimulation (TACS). Lucid dreaming refers to a well-documented and scientifically accepted phenomenon that may occur during natural sleep. Lucid dreaming is characterized by awareness of the fact that one is dreaming, an increased chance of control over the events that are occurring within the dream and increased logical or rational thought during the dream.

Drug containing melatonin and phosphatidylserine
CN105708843

The invention discloses a drug containing melatonin and phosphatidylserine. The drug comprises raw materials in parts by weight as follows: 0.1-100 parts of the melatonin and 1-400 parts of phosphatidylserine. A drug preparation can be beverage, tea, tablets, granules or capsules. The drug maintains the effects of aging resistance, mental state maintenance, sleep improvement, resistance enhancement and human immune system support of the melatonin, and can treat symptoms of deficiency type dyspnea caused by long-time cough, wet dream and night emission, enuresis and frequent micturition, protracted diarrhea, spontaneous sweating, night sweat, body fluid deficiency, thirst, short breath and feeble pulse, diabetes caused by internal heat, palpitation and insomnia and the like, and due to addition of phosphatidylserine, the drug further can enhance the memory ability, prevent senile dementia, improve intelligence and the like.

Traditional Chinese medicine used for treating sexual intercourse in dream
CN104189132

The invention provides a traditional Chinese medicine used for treating sexual intercourse in dream. The traditional Chinese medicine is prepared from the following medicines: radix astragali, radix codonopsis, Chinese angelica, atractylodes macrocephala, longan, spina date seed, poria with hostwood, milkwort root, cassia twig, fossil fragments, oysters, radix glycyrrhizae preparata, Chinese dates and tuber fleeceflower stem. 700ml of water is added to first dose of the medicines to decoct the medicines until the decoction is 200ml, and 500ml of water is added to second dose of the medicines to decoct the medicines until the decoction is 200ml. The traditional Chinese medicine is taken twice a day on an empty stomach, once in the morning and once in the evening.

Non-Contact Electronic Tool for Dream Enhancement
US2013012762

This invention pertains to a non-contact electronic device for the enhancement of dreaming. By enhancement it is meant increasing the frequency, the length, the lucidity, and the color of dreaming. For example, one person who dreams in grey shades now dreams in color when using the device. The device will also allow better recall of the dreams. Dream enhancement is generally a pleasant process and promotes a pleasant feeling in the morning after use. The device is a passive electronic device, that is, it is not actively powered. The user places the device a few feet away during sleep. The device uses a piezo-crystal and an iron core. It is believed that ambient fluctuations in the piezo-crystal's voltage produce an influence field that is broadcasted through the iron core to the user a few feet away.

Method and apparatus for electrically generating signal for inducing lucid dreaming
US8267851

A lucid dream is induced in a subject. A brain state entrainment signal is generated in a circuit. A transducer applies a waveform of a kind sufficient to cause a lucid dream in the brain based on the brain state entrainment signal. The brain state entrainment signal modulates a carrier wave of a higher frequency than the entrainment frequency. The waveform is first applied to the subject and thereafter removed to attempt a lucid dream. The waveform can be electrical, light, sound or magnetic waveforms.

LUCID DREAM STIMULATOR, SYSTEMS, AND RELATED METHODS
US2017304583

A dream stimulator may include a substrate and a plurality of electrodes positioned along a first side of the substrate. The stimulator also may include a stimulator coupled to the substrate. At least one lead may couple the plurality of electrodes to the stimulator. Additionally, a band may be positioned along a second side of the substrate and about each of the plurality of electrodes.

SYSTEMS AND METHODS FOR DELIVERING SENSORY INPUT DURING A DREAM STATE
US9620027

Embodiments of the invention provide apparatus, systems and methods for detecting neurological activity indicative of a dream state of a human. Many embodiments of the invention provide apparatus, systems and methods for detecting neurological activity of a human indicative of a dream state or the onset thereof and delivering an input to the user (such as an audio or other sensory input) during the dream state. Particular embodiments of the invention provide systems and methods for detecting neurological activity indicative of the onset or occurrence of a dream state of a human and delivering an audio or other sensory input during the user's dream state. The audio input may be used for learning, delivering messages to the user's subconscious, and/or promoting a state of relaxation.

EROTIC DREAM INDUCTION APPARATUS
US2016331924
An erotic dream induction apparatus includes an inflatable air bladder, such as a rubber balloon, to induce sexual arousal via genital stimulation. The air bladder is brought into contact with either the penis or vagina of the user, and/or is inserted into the anus and/or vagina of the user. In one embodiment, the air bladder is brought into contact with the penis, vagina, and/or anus of the user by being maintained in position near, against, or around the penis, vagina, and/or anus by means of straps or pockets in stimulation underwear, such that when the air bladder is inflated, the air bladder makes contact with at least a portion of the penis, vagina, and/or anus of the user. The air bladder is repeatedly inflated and deflated by means of pumps and/or valves.

Dream Enhancement Apparatus and Method
US2016310698
Dream stage enhancement uses a headband with EEG-EOG sensors, onboard processors, memory, coarse and fine time REM waveform detection modules, LEDs and an audio playback unit. After normalization to the user's EEG waveforms, the user's EEG-EOG signals are processed, REM and NREM stages detected and light, sound or AV stimuli are presented to the user based upon user-supplied light-sound-AV stimuli commands. To provide a reality check control ("RCC"), the head unit has a user actuatable RC interface whereby during sleep, RC stimuli are presented when the user depresses the RCC control which plays back the user supplied stimulus. In a "learning mode," the user selects "Recall" or "No Recall" ("NR") after the sleep period. If NR, then the system changes the color of light stimuli, light intensity, flash, audio sound type, audio intensity, and AV. If "Recall" the user supplied stimuli commands are carried out.

INDUCING LUCID DREAMING
WO2016209078
The present invention is in the field of induction of lucid dreams by Transcranial Alternating Current Stimulation (TACS). Lucid dreaming refers to a well-documented and scientifically accepted phenomenon that may occur during natural sleep. Lucid dreaming is characterized by awareness of the fact that one is dreaming, an increased chance of control over the events that are occurring within the dream and increased logical or rational thought during the dream.

Drug containing melatonin and phosphatidylserine
CN105708843
The invention discloses a drug containing melatonin and phosphatidylserine. The drug comprises raw materials in parts by weight as follows: 0.1-100 parts of the melatonin and 1-400 parts of phosphatidylserine. A drug preparation can be beverage, tea, tablets, granules or capsules. The drug maintains the effects of aging resistance, mental state maintenance, sleep improvement, resistance enhancement and human immune system support of the melatonin, and can treat symptoms of deficiency type dyspnea caused by long-time cough, wet dream and night emission, enuresis and frequent micturition, protracted diarrhea, spontaneous sweating, night sweat, body fluid deficiency, thirst, short breath and feeble pulse, diabetes caused by internal heat, palpitation and insomnia and the like, and due to addition of phosphatidylserine, the drug further can enhance the memory ability, prevent senile dementia, improve intelligence and the like.

Traditional Chinese medicine used for treating sexual intercourse in dream
CN104189132
The invention provides a traditional Chinese medicine used for treating sexual intercourse in dream. The traditional Chinese medicine is prepared from the following medicines: radix astragali, radix codonopsis, Chinese angelica, atractylodes macrocephala, longan, spina date seed, poria with hostwood, milkwort root, cassia twig, fossil fragments, oysters, radix glycyrrhizae preparata, Chinese dates and tuber fleeceflower stem. 700ml of water is added to first dose of the medicines to decoct the medicines until the decoction is 200ml, and 500ml of water is added to second dose of the medicines to decoct the medicines until the decoction is 200ml. The traditional Chinese medicine is taken twice a day on an empty stomach, once in the morning and once in the evening.

Non-Contact Electronic Tool for Dream Enhancement
US2013012762
This invention pertains to a non-contact electronic device for the enhancement of dreaming. By enhancement it is meant increasing the frequency, the length, the lucidity, and the color of dreaming. For example, one person who dreams in grey shades now dreams in color when using the device. The device will also allow better recall of the dreams. Dream enhancement is generally a pleasant process and promotes a pleasant feeling in the morning after use. The device is a passive electronic device, that is, it is not actively powered. The user places the device a few feet away during sleep. The device uses a piezo-crystal and an iron core. It is believed that ambient fluctuations in the piezo-crystal's voltage produce an influence field that is broadcasted through the iron core to the user a few feet away.

Method and apparatus for electrically generating signal for inducing lucid dreaming
US8267851
A lucid dream is induced in a subject. A brain state entrainment signal is generated in a circuit. A transducer applies a waveform of a kind sufficient to cause a lucid dream in the brain based on the brain state entrainment signal. The brain state entrainment signal modulates a carrier wave of a higher frequency than the entrainment frequency. The waveform is first applied to the subject and thereafter removed to attempt a lucid dream. The waveform can be electrical, light, sound or magnetic waveforms.

Dream generator
CN202198951
The utility model relates to a dream generator, comprising a luminous source and a controller for controlling the luminous source to twinkle at intervals. The luminous source is connected on a clamp. The luminous source and the controller are supplied with electricity by a power source. The utility model provides a dream generator to generate dream for people entering into sleep.

METHOD AND APPARATUS FOR ATTAINING LUCID DREAM STATE
US2010130813
The invention discloses an apparatus that facilitates attaining of lucid dreaming state. The utility model relates to a dream generator, comprising a luminous source and a controller for controlling the luminous source to twinkle at intervals. The luminous source is connected on a clamp. The luminous source and the controller are supplied with electricity by a power source. The utility model provides a dream generator to generate dream for people entering into sleep.

COMPOSITIONS, METHODS, AND SYSTEMS FOR RAPID INDUCTION AND MAINTENANCE OF CONTINUOUS REM SLEEP
WO2009100071
Compositions, kits, methods, and systems to induce, maintain, monitor, and interpret a continuous, un-fragmented REM sleep state in humans, generate dreams, recover sleep, create brain neuronal plasticity and activate the brain for cognition enhancement and mood stabilization are described. If potentially combined with other medications, a platform is provided to develop new research and therapies in many fields treating the human brain. The procedure reliably and quickly generates a Continuous REM Sleep cycle of pre-determined time or indefinite duration depending on therapeutic goals, allowing the patient to experience a qualitatively superior dream sleep in a shortened period of time compared to a natural sleep cycle. Profuse positive (pleasant) dreams are produced as well. REM sleep, dreams and sleep recovery can be reliably generated, and various sleep, psychological and neurological illnesses and disorders can be treated or prevented either solely by this method or in combination with additional agents.

Your Support Maintains this Service --

BUY

The ***Rex Research Civilization Kit***

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

Don't buy a radio; Build one!

A Crystal Powered Steampunk Matchbox Radio

One of my friends recently had a birthday, and I got the idea to make him a simple steampunk themed matchbox radio. I'm not a steampunk expert, but I do know that devices should work without electricity. So, what better than a crystal radio?

Crystal radios in their most basic form require no electrical power, and are directly powered by the radio waves floating around us. Of course, you can add a little amplification to boost performance, like in my Altoids tin crystal radio, but it's not required.

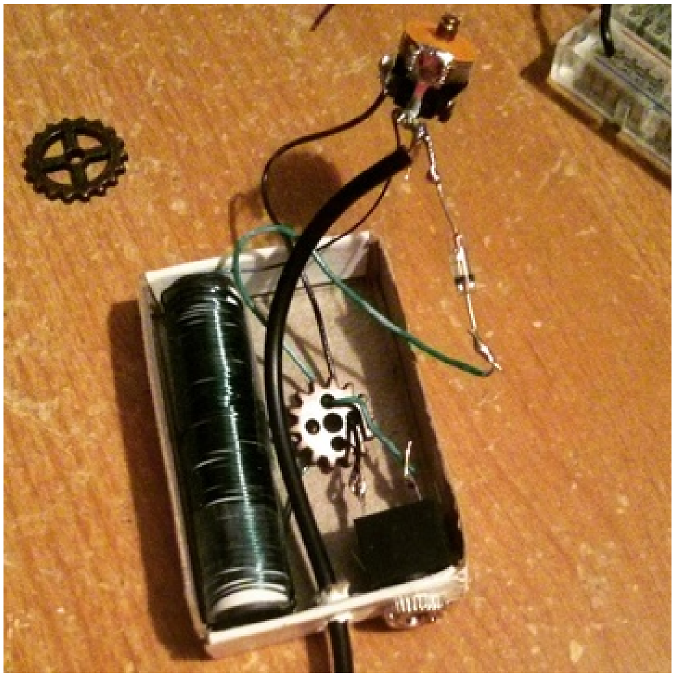

The heart of the crystal radio is a diode used as a detector, which has the ability to pull in amplitude modulated (AM) radio signals. The most popular diode for this purpose is the 1N34 germanium diode. You should be able to find 1N34 diodes in a number of online places, or at local hamfests. I picked up a small handful at Dayton last year.

Combine the detector with a LC (inductance / capacitance) circuit and an antenna, and you have a basic *tuned radio frequency* (TRF) radio. I used the same LC circuit used in in my rocket radio circuit (a 0-100 pF variable capacitor and about 70 turns of 26 gauge magnet wire on a ferrite core).

As for the antenna, the bigger the better. Often these radios will be small and provide an alligator clip that attach to an antenna or large piece of metal. I received pretty decent reception in my area, pulling in four local AM radio stations when clipping the radio to the cold water pipes or heating duct in my house. When clipping the antenna lead to my 44' doublet, wow! You can really hear some strong signals. It's hard to believe that there's nothing powering the circuit other than radio waves.

To hear the signal (without amplification), you will need a so-called crystal earpiece, which is a special type of high impedance earphone, which are much more sensitive than regular headphones.

The schematic for a basic crystal radio looks like this:

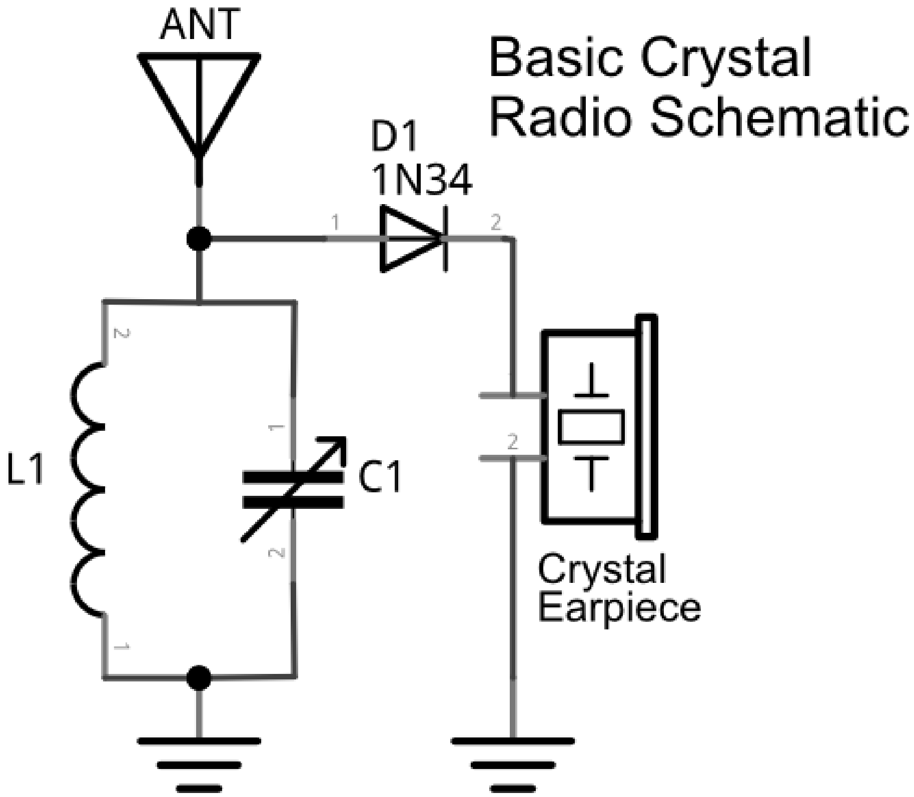
ANT
Basic Crystal
Radio Schematic
D1
1N34
1
2
2
1
L1
1
C1
2
2
Crystal
Earpiece

Here's a picture of the completed product. The steampunk gear tunes the frequency!

Posted: Feb 25, 2015

Keyword tags: crystal radiosteampunkschematicelectronicsbasicsmatchbox radio1N34 diode

Recent Posts

Ten Minute Transmitter

ARRL Field Day 2015

Dayton Hamvention 2015 Roundup

A Crystal Powered Steampunk Matchbox Radio

Make Your Own Rocket Radio with the TA7642

See more posts...

Sticky Posts

Fun With Crystal Oscillators: Part 1

Where to Locate Hard to Find RF Electronic Components

A Mighty Simple Shortwave Transmitter

So You Want to Be a Shortwave Pirate

shortwavetransmitterqrpschematicAM transmitterelectronics500mWarrlfield daycwradio operatingdayton hamventionham radiocrystal radiosteampunkbasicsmatchbox radio1N34 dioderocket radiota7642mk484zn414AM radiomake a radio

© 2018 makeRF

Making a Crystal Radio - The Little Whippersnapper's Parts List

The Little Whippersnapper’s Radio requires the following widely available low-cost parts. I hope that the cost of the circuit will be within the range of the little whippersnapper’s pocket money.

Hardware

- No. 6 Brass Screws ½-inch, quantity 16 from eBay
- No. 6 Brass Cups, quantity 16 from eBay
- Baseboard must be planed softwood ½-inch thick, 4-inch wide and 12-inch long.

Germanium Diode

- OA90, Maplin Code: QH71N

Main Parts

- 365 pF Tuning Capacitor
- Ferrite Rod 3/8 inch diameter, 3-inch or longer
- 36 s.w.g enamelled copper wire / Radio coil wire
- Connecting copper wire

Tools

- Pointed nose pliers
- Screwdriver
- Gimlet / Bradawl for the guide holes
- Wire Cutter or Stripper
- Sandpaper for cleaning terminals
- Cutter pliers
- Dad / Mum / Adult Supervision

US007053576B2

(12) **United States Patent**
Correa et al.

(10) **Patent No.: US 7,053,576 B2**
(45) **Date of Patent: May 30, 2006**

(54) ENERGY CONVERSION SYSTEMS

(76) Inventors: **Paulo N. Correa**, 42 Rockview Gardens, Concord, Ontario (CA) L4K 2J6; **Alexandra N. Correa**, 42 Rockview Gardens, Concord, Ontario (CA) L4K 2J6

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **10/270,154**

(22) Filed: **Oct. 15, 2002**

(65) **Prior Publication Data**

US 2006/0082334 A1 Apr. 20, 2006

Related U.S. Application Data

(63) Continuation of application No. 09/907,823, filed on Jul. 19, 2001, now abandoned.

(51) **Int. Cl.**
H05B 31/48 (2006.01)

(52) **U.S. Cl.** **318/558**; 315/111.21

(58) **Field of Classification Search** 310/308, 310/309; 318/558; 315/111.01, 111.21, 315/111.31

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

433,702	A	8/1890	Telsa	
645,576	A	3/1900	Telsa	
3,233,157	A *	2/1966	Stockman	318/116
4,095,118	A *	6/1978	Rathbun	290/2
4,213,797	A *	7/1980	Sher	136/254
4,473,736	A *	9/1984	Bloyet et al.	219/121.48
5,063,330	A *	11/1991	Leprince et al.	315/111.21
5,416,391	A	5/1995	Correa et al.	
5,449,989	A	9/1995	Correa et al.	
6,271,614	B1 *	8/2001	Arnold	310/233

OTHER PUBLICATIONS

Kuhn, T.S. (1978) "Black-body Theory and the Quantum Discontinuity, 1898-1912", The University of Chicago Press, pp. 246-249, 289-290.

Wang, LJ et al (2000) "Gain-assisted superluminal light propagation", Nature, 406:277.

Martin, Thomas Commerfold (1894) "The Inventions, Researches and Writlngs of Nikola Tesla", The Electrical Engineer, New York, p. 68.

Tesla, Nikola (1956) "Lectures, Patents, Articles", Nikola Tesla Museum, Beograd, Yugolsavia, L-70-71 & L-130-132 (Figure 16.II).

(Continued)

Primary Examiner—Bentsu Ro

(74) *Attorney, Agent, or Firm*—Ridout & Maybee LLP

(57) ABSTRACT

This invention relates to apparatus for the conversion of massfree energy into electrical or kinetic energy, which uses in its preferred form a transmitter and a receiver both incorporating Tesla coils, the distal ends of whose secondary windings are co-resonant and connected to plates of a chamber, preferably evacuated or filled with water, such that energy radiated by the transmitter may be picked up by the receiver, the receiver preferably further including a pulsed plasma reactor driven by the receiver coil and a split phase motor driven by the reactor. Preferably the reactor operates in pulsed abnormal gas discharge mode, and the motor is an inertially damped drag motor. The invention also extends to apparatus in which an otherwise driven plasma reactor operating in pulsed abnormal gas discharge mode in turn used to drive an inertially damped drag motor.

16 Claims, 11 Drawing Sheets

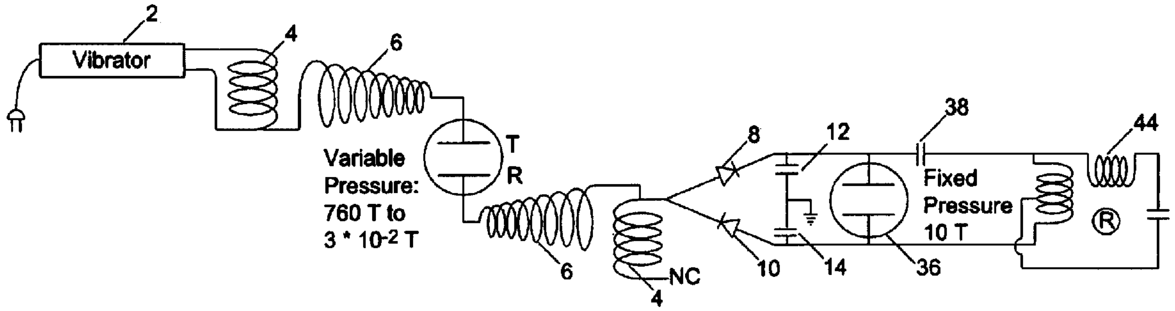

OTHER PUBLICATIONS

Reich, Wilhelm (1949) “A Motor Force in Orgone Energy. Preliminary Communications”, Orgone Energy Bulletin, vol. 1, #1, pp. 7-11.

Laithwaite, Eric (1957) “Induction Machines for Special Purposes”. p. 322-323.

Halliday, David & Resnick, Robert (1978)“Physics”, Parts I and II Combined Third Edition, John Wiley and Sons, p. 488-489.

* cited by examiner

FIG. 1

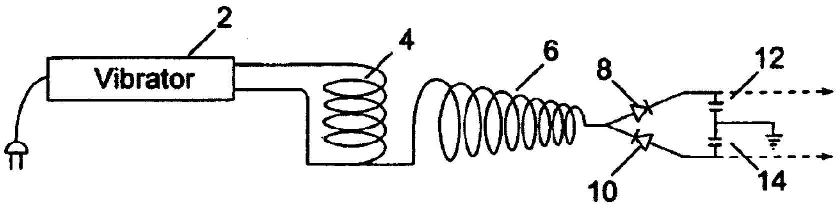
2
Vibrator
4
6
8
12
10
14

FIG. 2

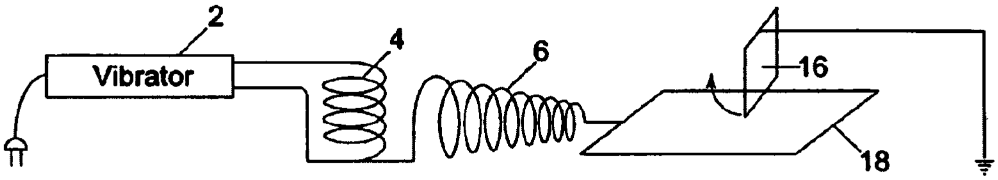
2
Vibrator
4
6
16
18

FIG. 3

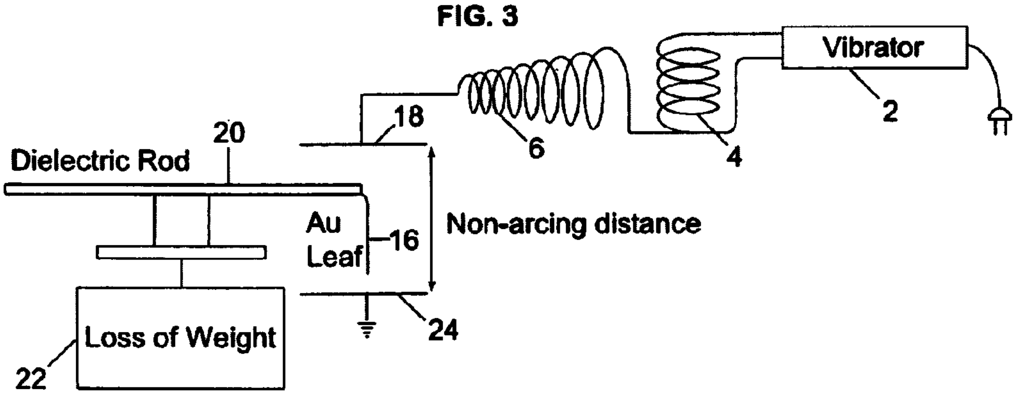
Vibrator
2
4
6
18
20
Dielectric Rod
Au Leaf
16
Non-arcing distance
24
Loss of Weight
22

FIG. 4

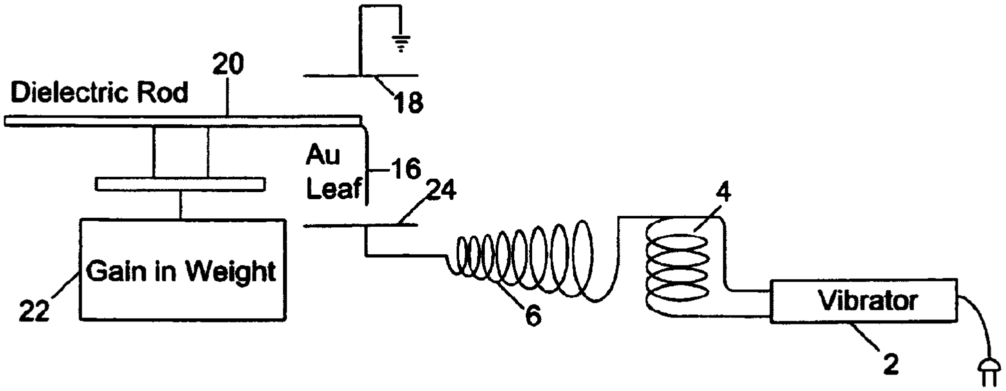
20
Dielectric Rod
18
Au Leaf
16
24
4
Gain in Weight
22
6
Vibrator
2

FIG. 5

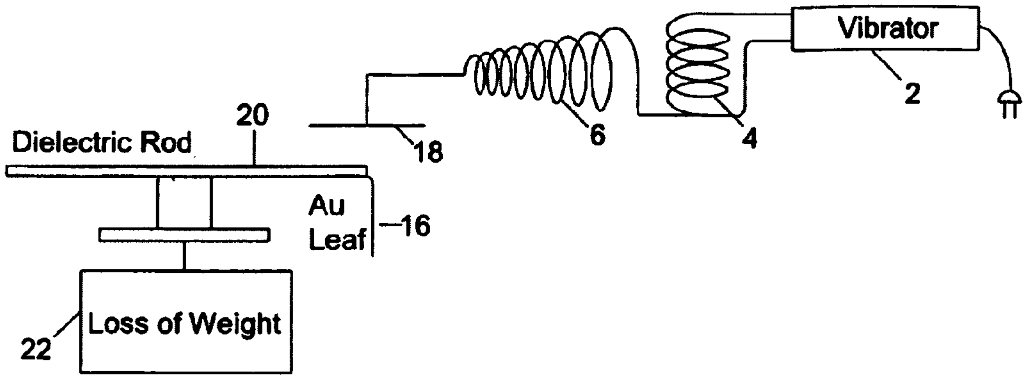
Vibrator
2
4
6
18
20
Dielectric Rod
Au
Leaf
16
Loss of Weight
22

FIG. 6

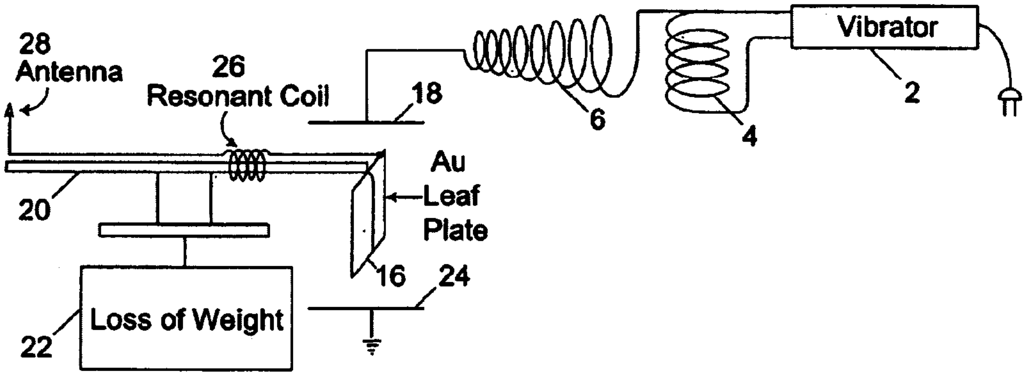
Vibrator
2
4
6
18
28
Antenna
26
Resonant Coil
20
Au
Leaf
Plate
16
24
Loss of Weight
22

FIG. 7

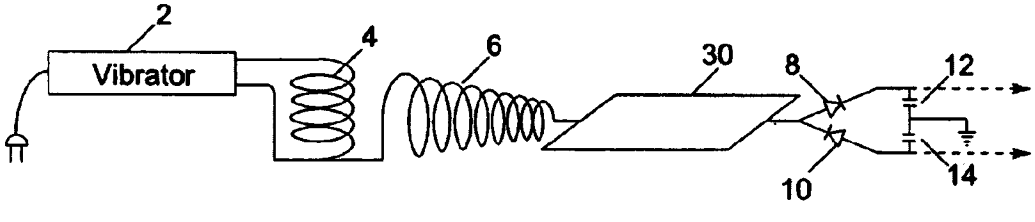
2
Vibrator
4
6
30
8
12
10
14

FIG. 8

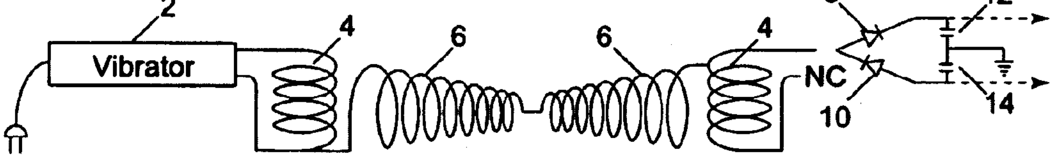
2
Vibrator
4
6
6
4
8
12
NC
10
14

FIG. 9

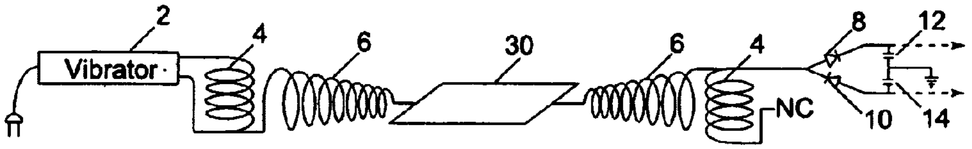
2
Vibrator
4
6
30
6
4
NC
8
12
10
14

FIG. 10

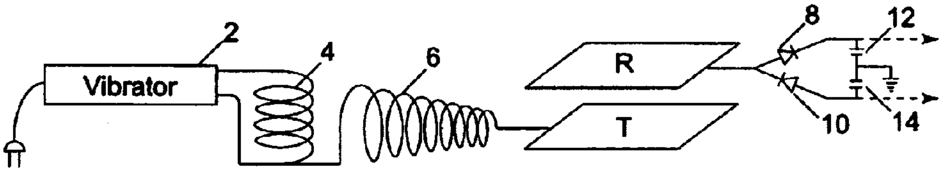
2
Vibrator
4
6
R
T
8
12
10
14

FIG. 11

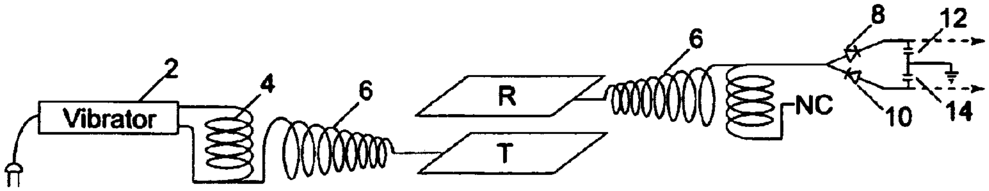
2
Vibrator
4
6
R
T
6
NC
8
12
10
14

FIG. 12

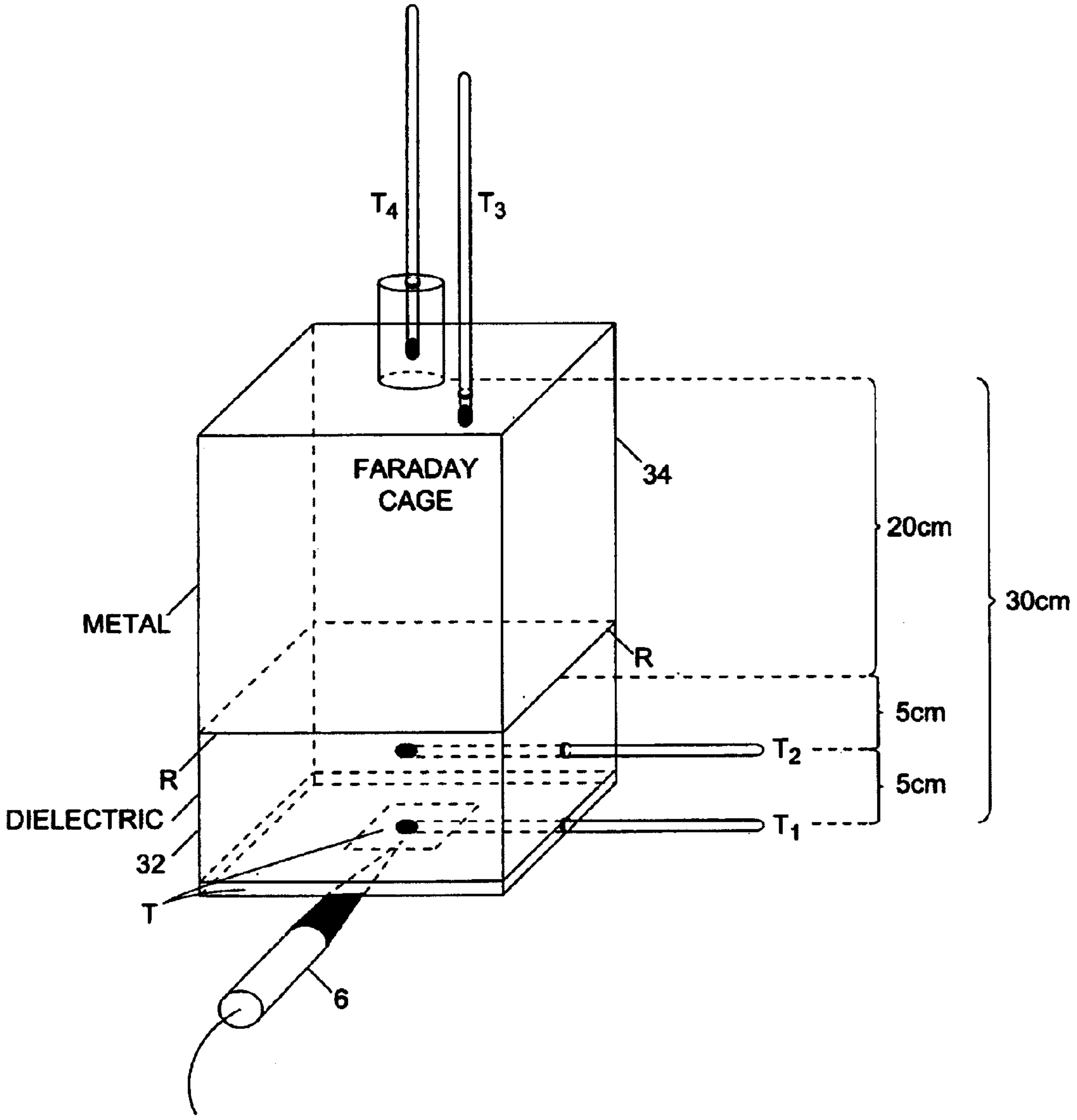
T4
T3
FARADAY
CAGE
34
20cm
30cm
METAL
R
R
5cm
T2
5cm
DIELECTRIC
T1
32
T
6

FIG. 13

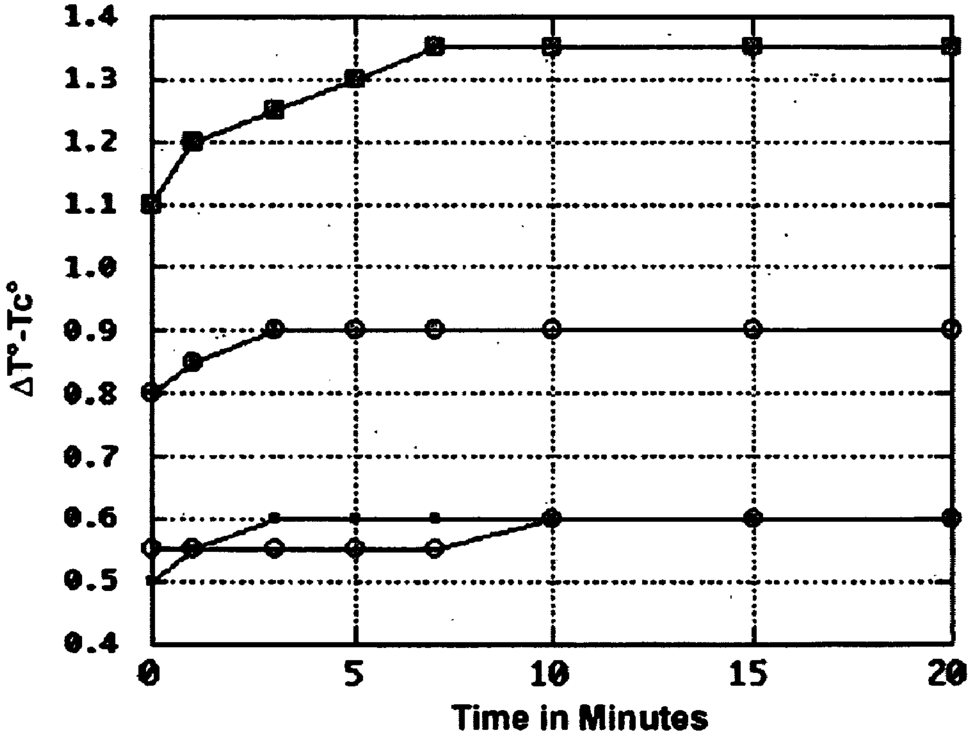

- ■ Above cage (T_4)
- ● R cage (T_3)
- ○ T/R cavity (T_2)
- ▪ T plate (T_1)

FIG. 14

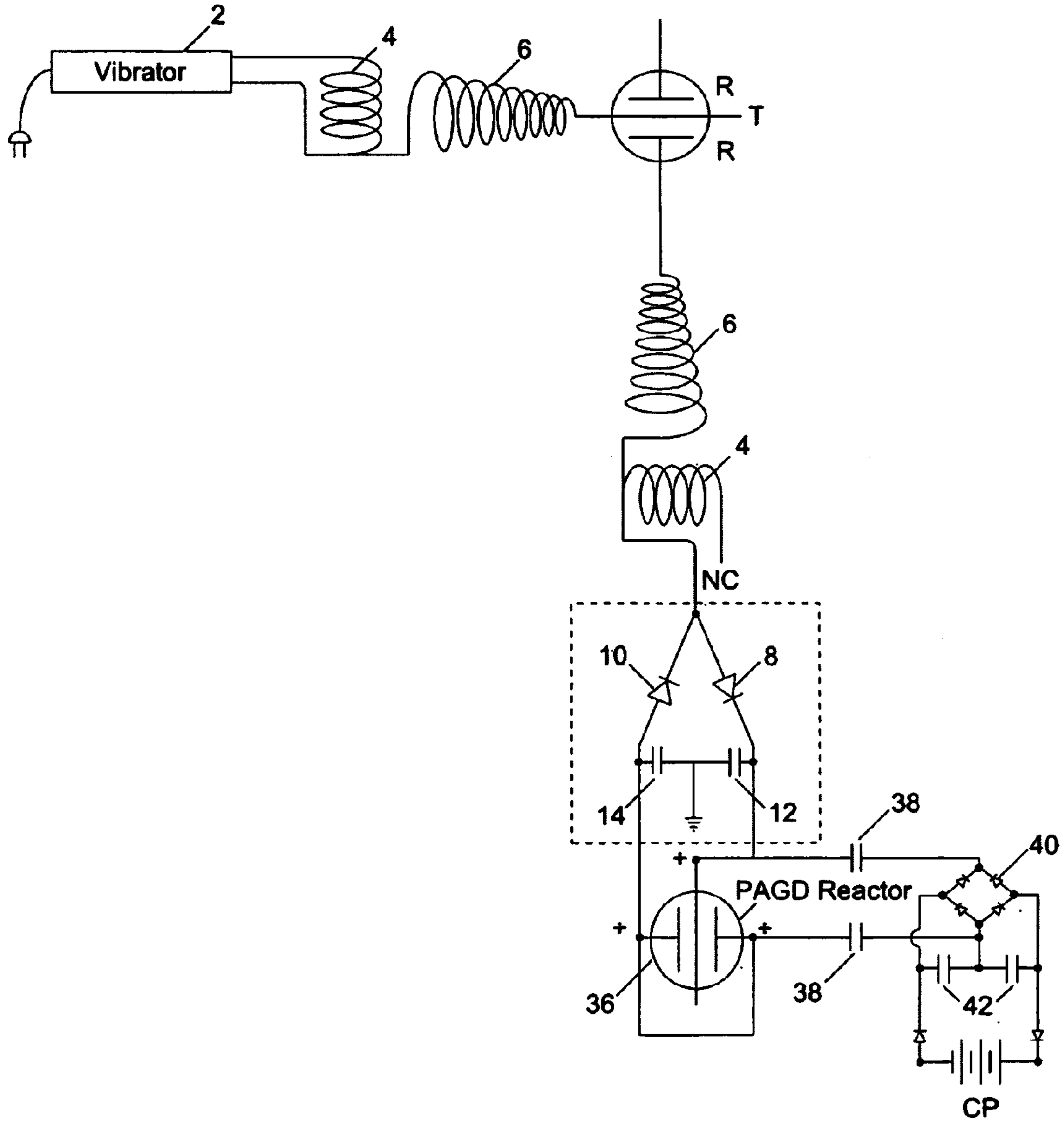
2
Vibrator
4
6
R
T
R
6
4
NC
10
8
14
12
38
40
PAGD Reactor
+
+
+
36
38
42
CP

FIG. 15

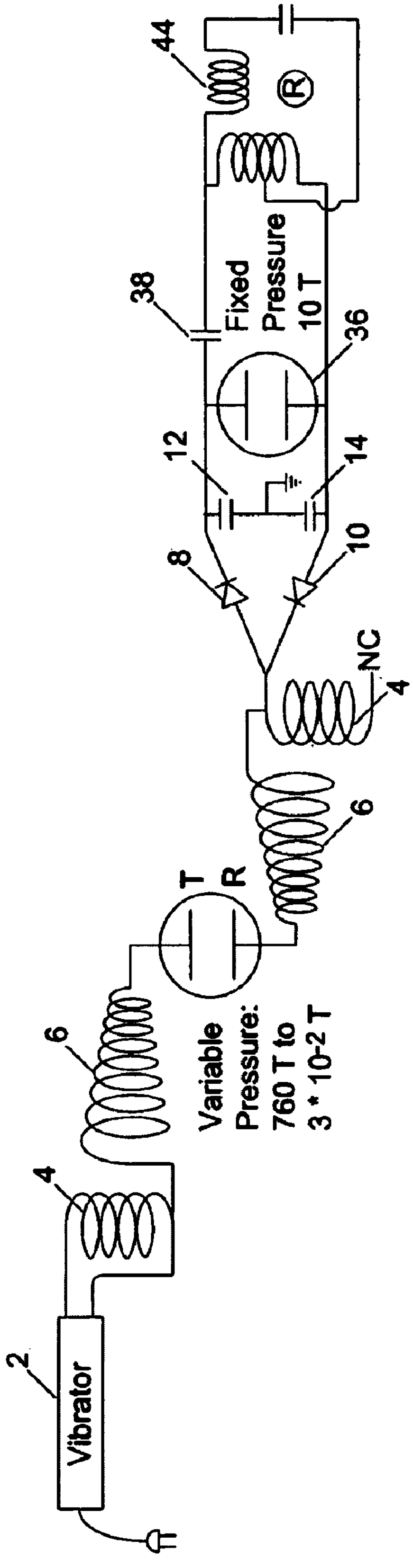
Vibrator
Variable Pressure: 760 T to 3 * 10-2 T
T
R
NC
Fixed Pressure 10 T
®

FIG. 16

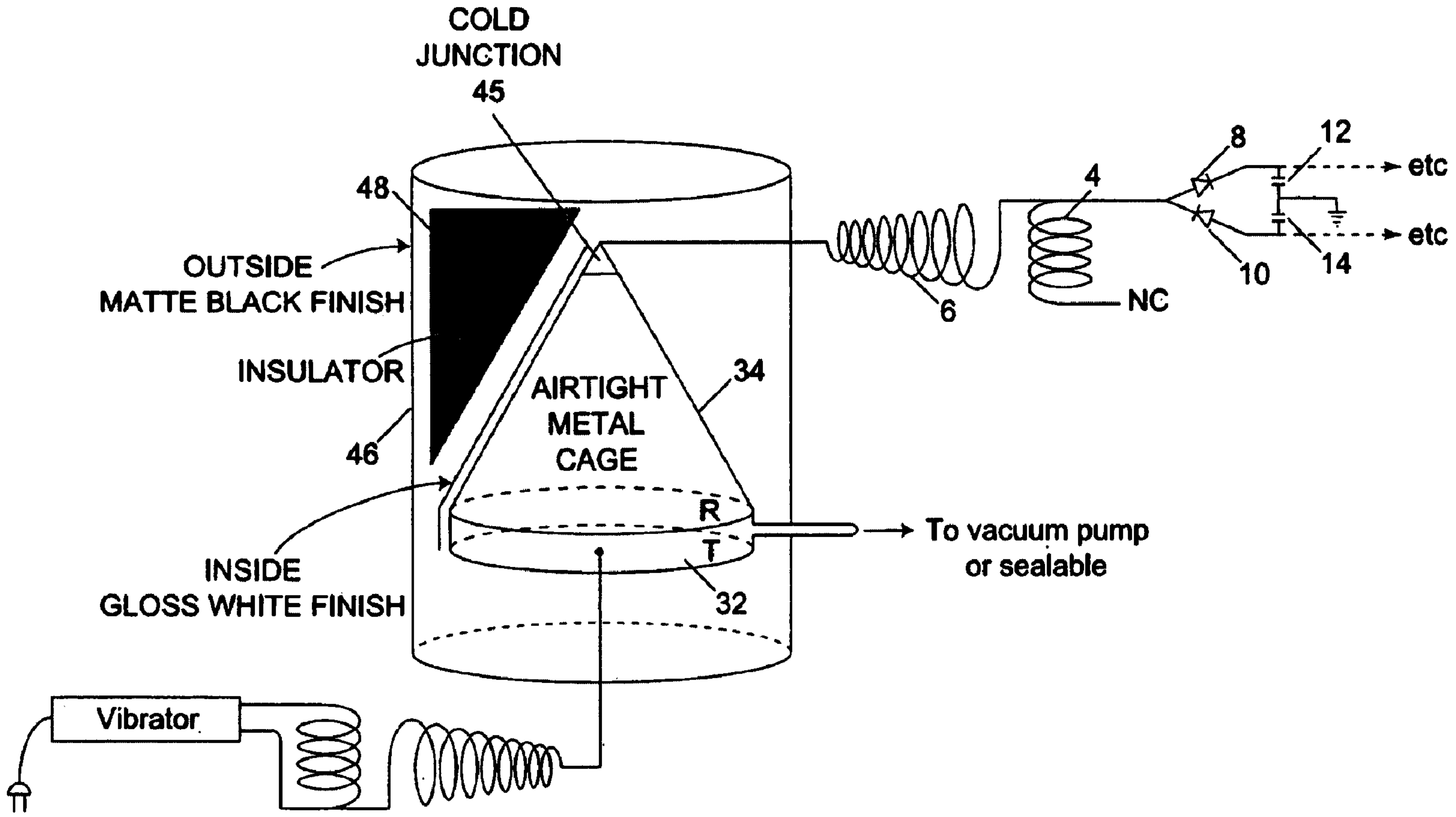
COLD JUNCTION 45
48
OUTSIDE MATTE BLACK FINISH
INSULATOR
46
INSIDE GLOSS WHITE FINISH
AIRTIGHT METAL CAGE
34
R
T
32
To vacuum pump or sealable
6
4
NC
8
10
12
14
etc
etc
Vibrator

FIG. 17

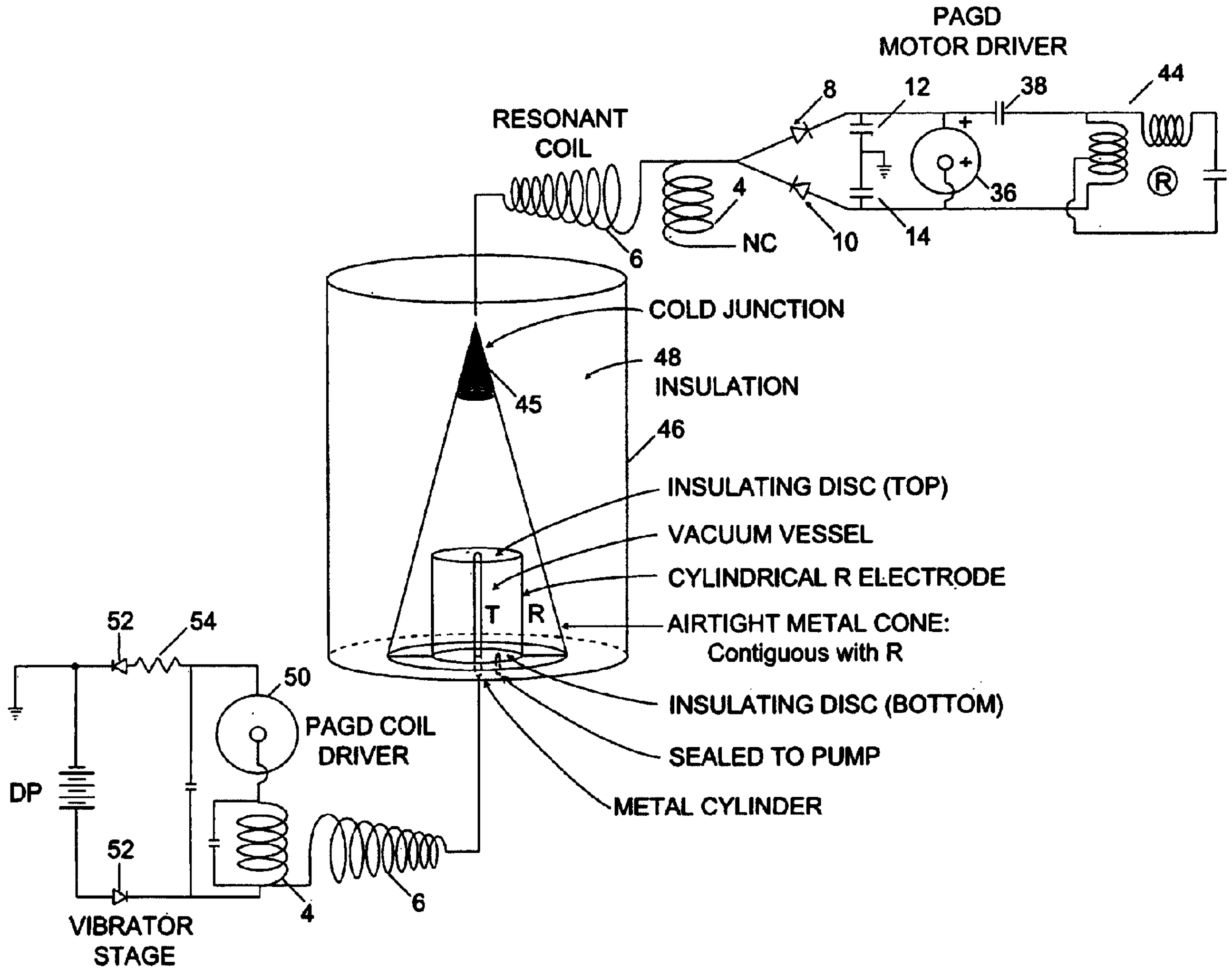
PAGD
MOTOR DRIVER
RESONANT
COIL
8
12
38
44
4
NC
10
14
36
6
R
COLD JUNCTION
48
INSULATION
45
46
INSULATING DISC (TOP)
VACUUM VESSEL
CYLINDRICAL R ELECTRODE
T
R
AIRTIGHT METAL CONE:
Contiguous with R
INSULATING DISC (BOTTOM)
SEALED TO PUMP
METAL CYLINDER
52
54
50
PAGD COIL
DRIVER
DP
52
4
6
VIBRATOR
STAGE

FIG. 18

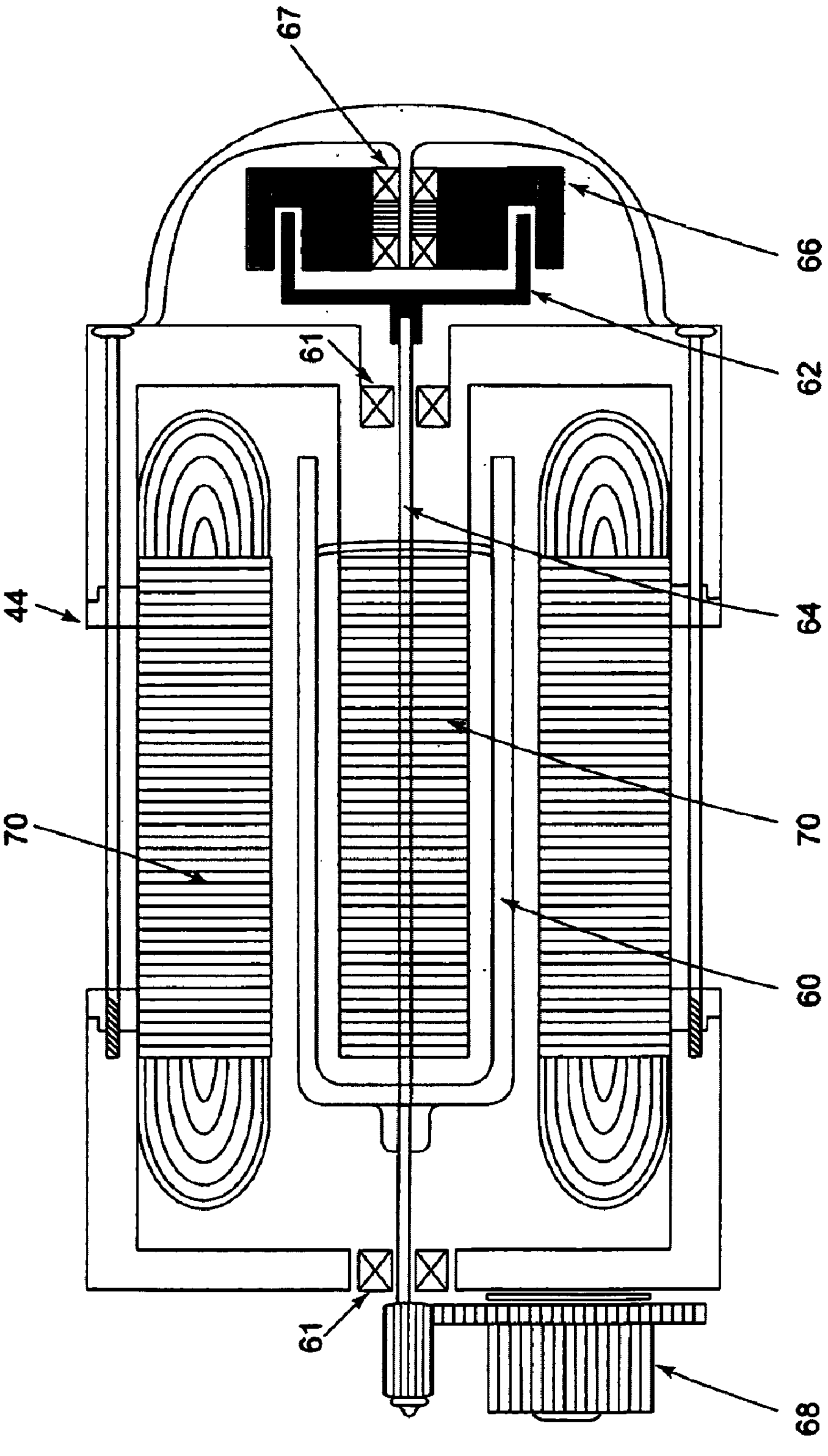
67
66
62
61
64
44
70
70
60
61
68

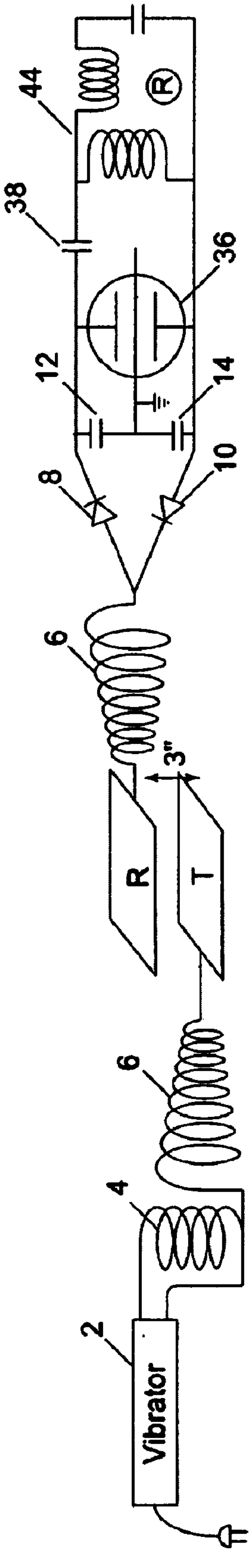
44
Ⓡ
38
36
12
14
8
10
6
3"
R
T
6
4
2
Vibrator

FIG. 19

ENERGY CONVERSION SYSTEMS

This is a continuation of application Ser. No. 09/907,823, filed Jul. 19, 2001 now abandoned.

FIELD OF THE INVENTION

This invention relates to systems for the conversion of energy, inter alia in the form of what we will refer to for convenience as Tesla waves (see below), to conventional electrical energy.

BACKGROUND OF THE INVENTION

Energy converters that are fed by local or environmental energy are usually explained by taking recourse to the notion that they convert zero point electromagnetic radiation (ZPE) to electric energy. The ZPE theories have gained a life of their own, as T. Kuhn has pointed out (in his “Black Body Theory and the Quantum”), after emerging from Planck’s second theory, specifically from the term ½ hυ in the new formula for oscillator energy. In 1913, Einstein and Stern suggested that motional frequencies contributing to specific heat fell into two categories—those that were independent of temperature and those that were not (e.g. rotational energy), leading them to conclude that zero-point energy on the order of ½ hυ was most likely. In the second part of their paper, however, they provided a derivation of Planck’s Law without taking recourse to discontinuity, by assuming that the value of the ZPE was simply hυ. It is worth noting that Einstein had already in 1905 (“Erzeugung und Verwandlung des Lichtes betreffenden heuristichen Gesichtspunkt”, Ann. d. Phys, 17, 132) framed the problem of discontinuity, even if only heuristically, as one of placing limits upon the infinite energy of the vacuum state raised by the Rayleigh-Jeans dispersion law. According to Einstein, the Rayleigh-Jeans law would result in an impossibility, the existence of infinite energy in the radiation field, and this was precisely incompatible with Planck’s discovery—which suggested instead that at high frequencies the entropy of waves was replaced by the entropy of particles. Einstein, therefore, could only hope for a stochastic validation of Maxwell’s equations at high frequencies “by supposing that electromagnetic theory yields correct time-average values of field quantities”, and went on to assert that the vibration-energy of high frequency resonators is exclusively discontinuous (integral multiples of hυ).

Since then, ZPE theories have gone on a course independent from Planck’s second theory. The more recent root of modern ZPE theories stems from the work of H. Casimir who, in 1948, apparently showed the existence of a force acting between two uncharged parallel plates. Fundamentally the Casimir effect is predicated upon the existence of a background field of energy permeating even the ‘vacuum’, which exerts a radiation pressure, homogeneously and from all directions in space, on every body bathed in it. Given two bodies or particles in proximity, they shield one another from this background radiation spectrum along the axis (i.e. the shortest distance) of their coupling, such that the radiation pressure on the facing surfaces of the two objects would be less than the radiation pressure experienced by all other surfaces and coming from all other directions in space. Under these conditions, the two objects are effectively pushed towards one another as if by an attractive force. As the distance separating the two objects diminishes, the force pushing them together increases until they collapse one onto the other. In this sense, the Casimir effect would be the macroscopic analogy of the microscopic van der Waals forces of attraction responsible for such dipole-dipole interactions as hydrogen bonding. However, it is worth noting that the van der Waals force is said to tend to establish its normal radius, or the optimal distance between dipoles, as the distance where the greatest attractive force is exerted, beyond which the van der Waals forces of nuclear and electronic repulsion overtake the attraction force.

Subsequently, another Dutch physicist, M. Sparnaay, demonstrated that the Casimir force did not arise from thermal radiation and, in 1958, went on to attribute this force to the differential of radiation pressure between the ZPE radiation from the vacuum state surrounding the plates and the ZPE radiation present in the space between them. Sparnaay’s proposal is that a classical, nonquantal, isotropic and ubiquitous electromagnetic zero-point energy exists in the vacuum, and even at a temperature of absolute zero. It is further assumed that since the ZPE radiation is invariant with respect to the Lorentz transformations, it obeys the rule that the intensity of its radiation is proportional to the cube of the frequency, resulting in an infinite energy density for its radiation spectrum.

What appeared to be the virtue of this reformulated theory was the notion that the vacuum no longer figured as pure space empty of energy, but rather as a space exposed to constantly fluctuating ‘fields of electromagnetic energy’.

Puthoff has utilized the isomorphism between van der Waals and Casimir forces to put forth the zero-point (ZP) energy theory of gravity, based on the interpretation that the virtual electromagnetic ZP field spectrum predicted by quantum electrodynamics (QED) is functionally equivalent to an actual vacuum state defined as a background of classical or Maxwellian electromagnetic radiation of random phases, and thus can be treated by stochastic electrodynamics (SED). Whereas in QED, the quanta are taken as virtual entities and the infinite energy of the vacuum has no physical reality, for SED, the ZPE spectrum results from the distortion of a real physical field and does not require particle creation. Gravity then, could be seen as but the macroscopic manifestation of the Casimir force.

We do not dispute the fact that even in space absent matter there is radiant energy present which is not of a thermal nature. But we claim that this energy is not electromagnetic, nor is its energy spectrum infinite. That this is so stems not just from our opinion that it is high time that Einstein’s heuristic hypothesis should be taken as literally factual—in the dual sense that all electromagnetic energy is photon energy and all photons are local productions, but above all from the fact that it is apparent, from the experiments of Wang and his colleagues (Wang, Li, Kuzmich, A & Dogariu, A. “Gain-assisted superluminal light propagation”, Nature 406; #6793; 277), that the photon stimulus can propagate at supraluminal speeds and lies therefore well outside of any scope of electromagnetic theory, be this Maxwell’s classical approach taken up by ZPE theories, or Einstein’s special relativistic phenomenology of Maxwell’s theory. The fact is that if the light stimulus can propagate at speeds greater than those of light, then what propagates is not light at all, and thus not energy configured electromagnetically. Light is solely a local production of photons in response to the propagation of a stimulus that itself is not electromagnetic.

It is critical to understand that the implication from this that—aside from local electromagnetic radiation and from thermal radiation associated with the motions of molecules (thermo-mechanical energy), there is at least another form of energy radiation which is everywhere present, even in space absent matter. Undoubtedly, it is that energy that prevents any attainment of absolute zero, for any possible local outpumping of heat is matched by an immediate local conversion of some of this energy into a minimum thermal radiation required by the manifolds of Space and Time. And undoubtedly also this radiation is ubiquitous and not subject to relativistic transformations (i.e. Lorentz invariant). What it is not, is electromagnetic radiation consisting of randomistic phases of transverse waves.

To understand this properly, one must summarize the differences from existing ZPE theories—and all these differences come down to the fact that this energy which is neither electromagnetic nor thermal per se (and is certainly not merely thermo-mechanical), has nevertheless identifiable characteristics both distributed across subtypes or variants and as well common to all of them.

Essentially the first subtype or variant consists of longitudinal massfree waves that deploy electric energy. They could well be called Tesla waves, since Tesla-type transformers can indeed be shown experimentally to radiate massfree electric energy, in the form of longitudinal magnetic and electric waves having properties not reduceable to photon energy or to 'electromagnetic waves', and having speeds of displacement that can be much greater than the limit c for all strictly electromagnetic interactions.

One may well denote the second subtype by the designation of massfree thermal radiation, since it contributes to temperature changes—and, as obviously indicated by the impossibility of reaching an absolute zero of temperature, this contribution occurs independently of the presence of matter, or mass-energy, in Space. In other words, not all thermal radiation can be reduced to vibration, rotation and translation (drift motion) of molecules, i.e. to thermomechanical energy, because the properties of pressure and volume that determine temperature and affect matter, appear indeed to a great extent to be independent from matter, a fact which itself is responsible for the observed catastrophic and unexpected phase changes of matter and has required to this day the insufficient explanation offered semi-empirically by the Van der Waals Force Law.

Finally the third subtype may be designated latent massfree energy radiation—since it deploys neither charge, nor thermal or baroscopic effects, and yet it is responsible for 'true latent heat' or for the 'intrinsic potential energy' of a molecule. It is also responsible for the kinetoregenerative phenomenon whereby an electroscope performs a variable charge-mediated work against the local gravitational field.

The common characteristic of all three subtypes of massfree energy radiation is that they share the same nonclassical fine structure, written as follows for any energy unit, where c is any speed of light wave function, and the wavelength λ and wave function W are interconnected as a function of the physical quality of the energy field under consideration:

$$E=\lambda cW$$

In the instance of longitudinal electric radiation, this takes on the directly quantizable form:

$$E=(\lambda_q c)W_\nu=p_e W_\nu=(h/\lambda_x)W_\nu=\int=q_\nu$$

where W_ν is the voltage-equivalent wave function corresponding to V, P_e constitutes the linear momentum corresponding to the conventional q or e, h is the Planck constant, λ_x the Duane-Hunt constant expressed as a wavelength, λ_q is a wavelength constant; and the sign=∫=signifies exact equality between an expression in the conventional dimensions of length, mass and time, and an expression in length and time dimensions alone.

In the instance of massfree thermal radiation (contributing to temperature changes), the transformation obeys Boltzmann's rule (k is now Boltzmann's constant and T is Kelvin-scale temperature):

$$E=\lambda_{n1}cW_{n1}=\lambda_{n1}(\lambda_V\zeta_p)(\lambda_\alpha\zeta_{n1})=\int=kT$$

and in the third instance—of latent massfree radiation, the transformation obeys the rule:

$$E=\lambda_{n1}cW_{n1}=\lambda_{n1}(\lambda_{n1}\zeta_{n1})(\lambda_{n1}f_{n1})=\lambda_{n1}{}^3\zeta_{n1}f_{n1}$$

where ζ and f are frequency functions, f being a specific gravitational frequency term, and f_{n1} being defined as equal to $(\lambda_{n1})^{-0.5}$ meter$^{0.5}$sec^{-1}. ζ_{n1} has the value of c/λ_{n1}.

If the electric variant of massfree radiation has a direct quantum equivalence, via the Duane-Hunt Law, none of the three primary aether energy variants possess either the classic form of electromagnetic energy which requires square superimposition of speed of light wave functions c, as c^2, or the quantum form of energy, requiring $E=h\upsilon$. The critical first step in the right direction may well be attributed to Dr. W. Reich, as it regards the fact that massfree energy couples two unequal wave functions, only one of which is electromagnetic and abides by the limit c. We then unravelled the threefold structure described above, and further showed that, in the case of longitudinal electric waves, the postulated equivalence ($q=\lambda_q c$) is merely phenomenological, as these waves are not restricted by the function c in their conveying of electric charge across space. It can further be demonstrated that all blackbody photons are bound by an upper frequency limit ($64*10^{14}$ Hz), above which only ionizing photons are produced, and that all blackbody photons arise precisely from the interaction of massfree electric radiation with molecules of matter (including light leptons), whereby the energy of that radiation is locally converted into photon or electromagnetic radiation. In other words, all nonionizing electromagnetic energy appears to be secondary energy that results locally from the interaction of matter with massfree electric energy. It cannot therefore consist of the primary energy that is present in the vacuum, an energy that is neither virtual nor electromagnetic, but actual and concrete in its electric, thermal and antigravitic manifestations. Lastly, gravitational energy, being either the potential or the kinetic energy responsible for the force of attraction between units of matter, is a manifestation that also requires, much as electromagnetic radiation does, coupling of massfree energy to matter or to mass-energy.

The Tesla coil is a generator of a massfree electric energy flux that it transmits both by conduction through the atmosphere and by conduction through the ground. Tesla thought it did just that, but it has been since regarded instead (because of Maxwell, Hertz and Marconi) as a transmitter of electromagnetic energy. The transmitter operates by a consumption of massbound electric power in the primary, and by induction it generates in the coupled secondary two electric fluxes, one massbound in the coil conductor, and the other massfree in the body of the solenoid. Tesla also proposed and demonstrated a receiver for the massfree energy flux in the form of a second Tesla coil resonant with the first. The receiver coil must be identical and tuned to the transmitter coil; the capacitance of the antenna plate must match that of the transmitter plate; both transmitter and receiver coils must be grounded; and the receiver coil input and output must be unipolar, as if the coil were wired in series.

The generators of massfree energy with which we are concerned provide current pulses associated with a damped wave (DW) oscillation of much higher frequency than the pulse repetition frequency. A particular problem in recovering the massfree energy content of such pulses is provided by the damped wave oscillations. Although in our U.S. Pat. No. 5,416,391 we describe arrangements incorporating split phase motors to recover such energy, their efficiency is a great deal less than what should theoretically be attainable. Other workers such as Tesla and Reich, have encountered the same problem to an even greater degree.

In XIXth century motor engineering terminology, dynamos capable of producing direct current by continuous homopolar induction were known as 'unipolar' generators. The term unipolar induction appears to have originated with W. Weber, to designate homopolar machines where the conductor moves continuously to cut the magnetic lines of one kind of magnetic pole only, and thus require sliding contacts to collect the generated current. Faraday's rotating copper disc apparatus was, in this sense, a homopolar generator when the disc was driven manually, or a homopolar motor when the current was provided to it. Where the rotating conductor continuously cuts the magnetic field of alternatingly opposite magnetic poles, the operation of a machine, whether a generator or a motor, is said to be heteropolar. Unipolar machines went on to have a life of their own in the form of low voltage and high current DC generators—from Faraday, through Plucker, Varley, Siemens, Ferraris, Hummel, to Lord Kelvin, Pancinoti, Tesla and others—almost exclusively in the form of disc dynamos, but some having Wound rotors. In Mordey's alternator, and in so-called 'inductor alternators', however, homopolar generators were employed to obtain alternating currents, with the use of rotors wound back and forth across the field. Use of smooth, unwound rotors in AC induction motors (as opposed to AC synchronous motors, such as hysteresis motors) was a later development than homopolar dynamos. By 1888, Tesla and Ferraris amongst still others, had independently produced rotating magnetic fields in a motor, by employing two separate alternate currents with the same frequency but different phase. Single phase alternate current motors were developed later, and split-phase motors were developed last. Ferraris (Ferraris, G (1888) "Rotazioni elettrodynamiche", Turin Acad, March issue.) proposed the elementary theory of the 2-phase motor, where the current induced in the rotor is proportional to the slip (the difference between the angular velocity of the magnetic field and that of the rotating cylinder), and the power of the motor is proportional to both the slip and the velocity of the rotor.

If an iron rotor is placed within the rotating magnetic field of a 2-phase stator, it will be set in rotation, but not synchronously, given that it is always attracted to the moving magnetic poles with a lag. But if an aluminum or copper rotor is used instead, it becomes 'dragged' around by the rotating stator field because of the eddy currents induced in it. If the aluminum or copper rotor were to rotate synchronously with the stator magnetic field, there would be no induced eddy currents and thus no motor action would result. The motor action depends, in this instance, upon the presence of asynchronous slip, since the function of the latter is to sustain the induction of those currents in the rotor that are responsible for the motor action of the dragged rotor. This then is the origin of the term 'AC drag motors'. Once the drag rotor evolved from a cylinder to a hollow cup, they earned the epithet of 'drag-cup motors'. Later, already in the XXth century, the cups were fitted over a central stator member, and the sleeve rotor 2 phase servomotor was born.

Tesla knew that impulse currents as well as CW sinusoidal currents could be used to drive AC motors. Regarding his invention of an hysteresis motor ('magnetic lag motor', as he called it), he stated: " . . . pulsatory as well as an alternating current might be used to drive the[se] motors . . . " (Martin, TC (1894) "The inventions, researches and writings of Nikola Tesla", Chapter XII, p. 68). In his search for efficient utilization of the high frequency DW impulse currents of his induction coils, Tesla began by employing an AC disc induction motor as shown in FIG. **17** of his famous 1892 address (Tesla, N (1892) "Experiments with alternate currents of high potential and high frequency", in "Nikola Tesla Lectures", 1956, Beograd, pp. L-70–71). This consisted of a copper or aluminum disc mounted vertically along the longitudinal axis of an iron core on which was wound a single motor coil which was series wired to the distal terminal of an induction coil at one end, and to a large suspended and insulated metal plate at the other. What was new about this was the implementation of an AC disc induction motor drive, where the exciting current traveled directly through the winding with just a unipolar connection to the coil secondary (under certain conditions, even the series connection to the plate could be removed, or replaced with a direct connection to the experimenter's body): "What I wish to show you is that this motor rotates with one single connection between it and the generator" (Tesla, N. (1892), op. cit., L-70, Tesla's emphasis). Indeed, he had just made a critical discovery that, unlike in the case of massbound charge where current flow requires depolarization of a bipolar tension, massfree charge engages current flow unipolarly as a mere matter of proper phase synchronization.

Tesla thought that his motor was particularly adequate to respond to windings that had 'high-self-induction', such as a single coil wound on an iron core. The basis of this self-induction is the magnetic reaction of a circuit, or an element of a circuit—an inductor—whereby it chokes, dims or dampens the amplitude of electric waves and retards their phase.

For the motor to respond to still higher frequencies, one needed to wind over the primary motor winding a partial overlap secondary, closed through a capacitor, since "it is not at all easy to obtain rotation with excessive frequencies, as the secondary cuts off almost completely the lines of the primary" (Idem, L-71.).

Tesla stated that "an additional feature of interest about this motor" was that one could run it with a single connection to the earth ground, although in fact one end of the motor primary coil had to remain connected to the large, suspended metal plate, placed so as to receive or be bathed by "an alternating electrostatic field", while the other end was taken to ground. Thus Tesla had an ordinary induction coil that transmitted this "alternating electrostatic field", an untuned Tesla antenna receiving this "field", and a receiver circuit comprising his iron-core wound motor primary, a closely coupled, capacitatively closed secondary, and the coupled nonferromagnetic disc rotor. Eventually, in his power transmission system, he would replace this transmitter with a Tesla coil, and place an identical receiving coil at the receiving end, to tune both systems and bring them into resonance. But his motor remained undeveloped, and so did the entire receiver system.

Tesla returned to this subject a year later: "on a former occasion I have described a simple form of motor comprising a single exciting coil, an iron core and disc" (Tesla, N (1893) "On light and other high frequency phenomena", in "Nikola Tesla Lectures", 1956, Beograd, pp. L-130, and L-131 with respect to FIG. **16**-II). He describes how he developed a variety of ways to operate such AC motors unipolarly from an induction transformer, and as well other arrangements for "operating a certain class of alternating motors founded on the action of currents of differing phase". Here, the connection to the induction transformer is altered so that the motor primary is driven from the coarse secondary of a transformer, whose finer primary is coupled, at one end, directly and with a single wire to the Tesla secondary, and at the other left unconnected. On this occasion, Tesla mentions that such a motor has been called a 'magnetic lag motor', but that this expression (which, incidentally, he had himself applied to his own invention of magnetic hysteresis motors) is objected to by "those who attribute the rotation of the disc to eddy currents when the core is finally subdivided" (Tesla,N (1893), op. cit., p. L-130).

In none of the other motor solutions, 2-phase or split-phase, that he suggests as unipolar couplings to the secondary of an induction coil, does the nonferromagnetic disc rotor motor again figure. But he returns to it a page later, and indirectly so, by first addressing the disadvantages of ferromagnetic rotors: "Very high frequencies are of course not practicable with motors on account of the necessity of employing iron cores. But one may use sudden discharges of low frequency and thus obtain certain advantages of high-frequency currents-without rendering the iron core entirely incapable of following the changes and without entailing a very great expenditure of energy in the core. I have found it quite practicable to operate, with such low frequency disruptive discharges of condensers, alternating-current motors."

In other words—whereas his experiments with constant wave (CW) alternating currents, and as well with high-voltage DW impulses from induction coils, indicated the existence of an upper frequency limit to iron core motor performance, one might employ instead high-current, DW impulses—of high DW frequencies but low impulse rates—to move these motors quite efficiently. Then he adds "A certain class of [AC] motors which I advanced a few years ago, that contain closed secondary circuits, will rotate quite vigorously when the discharges are directed through the exciting coils [emphasis added]. One reason that such a motor operates so well with these discharges is that the difference of phase between the primary and secondary currents is 90 degrees, which is generally not the case with harmonically rising and falling currents of low frequency. It might not be without interest to show an experiment with a simple motor of this kind, inasmuch as it is commonly thought that disruptive discharges are unsuitable for such purposes."

What he proposes next forms the basis of modern residential and industrial AC electric power meters, the AC copper disc motor whose rotor turns on the window of these meters, propelled forward by the supply frequency. But instead of employing any such CW input, Tesla uses the disruptive discharges of condensers, incipiently operating as current rectifiers. With the proper conditions, e.g. correct voltage from the generator, adequate current from the capacitor, optimum capacitance for the firing rate, and tuned spark-gap, to mention a few, Tesla found that the nonferromagnetic disc rotor turned but with considerable effort. But this hardly compared to the results obtained with a high-frequency CW alternator, which could drive the disc "with a much smaller effort". In summary then, Tesla went as far as being the first to devise a motor driven by Tesla waves, that employed a nonferromagnetic rotor, and whose arrangement encompassed both transmitter and receiver circuits. For this purpose, he employed a single phase method in which the signal is fed unipolarly to the winding, placed in series with a plate capacitance.

Tesla also later proposed driving a similar single-phase nonferromagnetic disc motor from bipolar capacitative discharges through an atmospheric spark-gap now placed in parallel with the main motor winding, and again simulating a split-phase by a closely wound secondary that was closed by a capacitance.

As Tesla admits, the results of all his AC eddy current motor solutions were meagre and limited by current and frequency problems. Likewise, the two phase arrangements proposed by Reich for his OR motor, involving a superimposition of the damped waves of a first phase on a fixed continuous wave second phase, require an external power source and a pulse amplifier circuit, and failed to meet Reich's own desiderata.

We have previously proposed the use of squirrel cage motors with capacitative splitting of phase to convert the damped wave (DW) output of plasma pulsers, but once a squirrel cage (SC) is introduced, the dampening effect which the nonferromagnetic copper cage exerts in being dragged by the revolving stator field is counteracted by the ferromagnetic cylinder of laminated iron, in which the copper cage is embedded, working to diminish the slip and bring the rotor to near synchronism. This is, in all likelihood, what limits SC motors to responding to the DC component of the DW impulse, and thus be limited to respond to fluxes of massbound charges. Historically, as we shall see, the obvious advantage of the SC servomotors lay in the fact that, in particular for 2-phase applications, they were far more efficient at performing work without evolution of heat. Indeed, if the eddy currents in the nonferromagnetic rotor are permitted to circulate in nonordered form, the rotor material and stator will heat up rapidly and consume much power in that heating. This is in fact considered to be a weakness of AC nonferromagnetic-rotor induction motors.

SUMMARY OF THE INVENTION

The present invention is concerned with conversion to conventional electrical energy of the variants of massfree energy radiation considered above, referred to for convenience as Tesla waves, massfree thermal radiation and latent massfree radiation. The first variant of such radiation was recognized, generated and at least partially disclosed by Tesla about a hundred years ago, although his work has been widely misinterpreted and also confused with his work on the transmission of radio or electromagnetic waves. The Tesla coil is a convenient generator of such radiation, and is used as such in many of the embodiments of our invention described below, but it should be clearly understood that our invention in its broadest sense is not restricted to the use of such a coil as a source of massfree radiation and any natural or artificial source may be utilized. For example, the sun is a natural source of such radiation, although interaction with the atmosphere means that it is largely unavailable at the earth's surface, limiting applications to locations outside of the earth's atmosphere.

According to the invention, a device for the conversion of massfree radiation (as herein defined) into electrical or mechanical energy comprises a transmitter of massfree electrical radiation having a damped wave component, a receiver of such radiation tuned to resonance with the damped wave frequency of the transmitter, a co-resonant output circuit coupled into and extracting electrical or kinetic energy from the receiver, and at least one of structure defining a transmission cavity between the transmitter and the receiver, a full-wave rectifier in the co-resonant output circuit, and an oscillatory pulsed plasma discharge device incorporated in the co-resonant output circuit. The output circuit preferably comprises a full wave rectifier presenting a capacitance to the receiver, or an electric motor, preferably a split phase motor, presenting inductance to the receiver. The transmitter and receiver each preferably comprise a Tesla coil and/or an autogenous pulsed abnormal glow discharge device. The transmission cavity is preferably at least partially evacuated, and comprises spaced plates connected respectively to the distal poles of the secondaries of Tesla coils incorporated in the transmitter and receiver respectively, the plates being parallel or concentric. The structure defining the cavity may be immersed in ion-containing water. The split-phase motor is preferably an inertially damped AC drag motor.

The invention, and experiments demonstrating its basis, are described further below with reference to the accompanying drawings.

SHORT DESCRIPTION OF THE DRAWINGS

FIG. **1** is a schematic view of a Tesla coil connected to a full-wave rectifier to form an energy conversion device;

FIG. **2** is a schematic view of a Tesla coil connected to a gold leaf electrometer;

FIGS. **3**–**6** show alternative electrometer configurations;

FIGS. **7**–**11** show modifications of the circuit of FIG. **1**;

FIG. **12** shows apparatus for investigating aspects of the experimental results obtained with the foregoing devices;

FIG. **13** is a graph illustrating results obtained from the apparatus of FIG. **12**;

FIGS. **14**–**17** show schematic diagrams of embodiments of energy conversion devices;

FIG. **18** is a diagrammatic cross section of an inertially damped drag cup motor; and

FIG. **19** is a schematic diagram of a further embodiment of energy conversion device incorporating such a motor.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Based upon observations of weight loss in metallic matter as induced by exposure to high frequency alternating electric fields, we developed an experimental method to optimize this weight loss, and from this a device that treats the forces causing weight loss as manifestations of intrinsic potential energy ΔU (or true "latent heat") of the molecules of matter, and converts both "true latent heat" energy present in the neighbourhood of a receiver, and "sensible" heat induced within that receiver, into electric energy that can be employed to drive a motor, flywheel or charge battery supplies.

It is commonly believed that the output of the Tesla coil is ionizing electromagnetic radiation. We have demonstrated that it is not; i.e. that it is neither electromagnetic radiation, nor ionizing electromagnetic radiation. The output of an air cored, sequentially-wound secondary, consists exclusively of electric energy: upon contact with the coil, a massbound AC current can be extracted at the resonant frequency, whilst across a non-sparking gap, massfree AC-like electric wave radiation having the characteristics of longitudinal waves, can be intercepted anywhere in adjacent space. Accordingly, the radiation output from such coils is distinct from electromagnetic radiation.

The basic demonstration that the output of a Tesla coil does not consist of ionizing radiation is that it does not accelerate the spontaneous discharge rate of electroscopes, whether positively or negatively charged. In fact, in its immediate periphery, the coil only accelerates the spontaneous discharge rate of the negatively charged electroscope (i.e. the charge leakage rate), whereas it arrests the discharge of the positively charged electroscope (i.e. the charge seepage rate falls to zero). But this dual effect is not due to any emission of positive ions from the secondary, even if it can positively charge a discharged electroscope brought to its proximity. This charging effect is in fact an artifact, in that metals but not dielectrics are ready to lose their conduction and outer valence band electrons when exposed to the massfree electric radiation of the coil. This is simply demonstrated by the apparatus of FIG. **1**, in which the distal terminal of the secondary winding **6** of a Tesla coil having a primary winding **4** driven by a vibrator **2** is connected to the input of a full-wave voltage wave divider formed by diodes **8** & **10** and reservoir capacitors **12** & **14** (the same reference numerals are used for similar parts in subsequent Figures). If the rectifiers employed are nondoped, the coil appears to only charge the divider at the positive capacitance **10**, but if doped rectifiers are employed, the coil will be observed to charge both capacitances equally. Whereas positive ionizers can charge either doped or undoped dividers positively, no positive ionizer can charge a doped divider negatively, clearly demonstrating that the Tesla coil does not emit positive ions.

The basic demonstration that the output of a Tesla coil is not nonionizing electromagnetic radiation of high frequency, such as optical radiation, or of lower frequency, such as thermal photons, is also a simple one. Placement of a sensitive wide spectrum photoelectric cell (capable of detecting radiation to the limits of vacuum UV), wired in the traditional closed circuit manner from a battery supply, at any distance short of sparking from the distal terminal of the coil will show in the dark that the light output from the coil is negligible. This rules out optical radiation at high frequency. The demonstration that the sensible heat output from the Tesla coil is also negligible will be addressed below.

Our theory proposed the existence of physical processes whereby massfree electric radiation can be converted into electromagnetic radiation. Such a process is at work whenever massfree electric wave radiation interacts with electrons, such as those that remain in the valence bands of atoms. This massfree electric energy interacts with charge carriers, such as electrons, to confer on them an electrokinetic energy which they shed in the form of light whenever that electrokinetic energy is dissociated from those carriers (e.g. by deceleration, collision or friction processes). Such a process is at work to a negligible extent in the coil itself and its usual terminal capacitance, hence the faint glow that can be seen to issue from it, but it can also be greatly amplified in the form of a corona discharge by connecting a large area plate to the output of the secondary, as Tesla himself did in his own experiments, and thus by increasing the capacitance of the coil system. Now, what is interesting in this process is that, in the absence of virtually any I^2R losses at the plate, and if the plate thus introduced is bent at the edges so that it has no pointed edges, or if it is in the form of a bowl, or in any other manner that precludes sparking at edges and specially corners, and thus enhances the corona discharge, any electroscope, whether negatively or positively charged, now brought close to the plate will show a tendency to arrest its spontaneous discharge rate. One might say that this is simply the result obtained in a Faraday cage that disperses charge on its outside and insulates electrically its interior, and indeed if an electroscope is placed inside a Faraday cage no amount of Tesla radiation on the outside of that cage, save direct sparking, adversely affects the leakage or seepage rate of the electroscope. In fact, since the effect of such a cage can be shown to be that of, by itself, inducing arrest of either spontaneous electroscopic discharge, this effect simply remains or is magnified when the cage is bathed by Tesla radiation. However, a cage constitutes an electrically isolated environment, whereas a plate with or without curved or bent edges does not. Furthermore, the change observed in the properties of the output radiation from a Tesla coil when certain metal plates or surfaces are directly connected to the distal terminal of the secondary, takes place whilst the capacitance of the coil is increased by the connected plate, and thus the plate is an electrically active element of the circuit—and hence the opposite of an electrically isolated element.

For a long time we believed that the anomalous cathode reaction forces observed in autoelectronic discharges (atmospheric sparks, autogenous PAGD (pulsed abnormal glow discharge) and vacuum arc discharges) were exclusive to an autoelectronic emission mechanism prompted by a direct potential between discharging electrodes. Sparking driven by AC potentials could sustain the same forces, but their mutual cancellation over time would not deploy a net force. In this sense, when a large gold leaf connected directly to the ground (via a water pipe or any other suitable connection) or to another large area plate suspended at some height above the ground, is vertically placed at a sparking distance above the surface of another plate connected to the secondary of a Tesla coil, one would not expect the AC spark to sustain any net force across the gap between the gold leaf and the plate. In terms of cathode reaction forces, one would expect their cancellation to be simply brought about by the high frequency of the current alternation in the coil, as both leaf and plate would alternate between being the emitting cathode or the receiving anode. However, this is not what is observed—instead, the gold leaf **16** lifts away from the plate **18** (FIG. **2**). Connecting instead the suspended gold leaf to the coil terminal, and the bottom plate to the ground in the same manner as described above, also yields the same result.

Even more curious is the finding that this anomalous reaction force deployed by an alternate current of mass-bound charges in the arc, remains present when the sparking is prevented and instead the corona effect is enhanced (by employing a large plate connected to the distal pole of the secondary, and by employing a distance at which sparking ceases), as if the lift itself were the property of the corona underlying the spark channels and not the property per se of the autoelectronic emission mechanism.

By mounting directly the suspended leaf **16** (41 mg of hammered 99.9996% pure gold) at the end of a long dielectric rod **20** balanced at the center and placed on a light stand over an electronic balance **22**, we sought to determine the observed lift of the leaf as weight lost. Surprisingly, and despite the most apparent lifting motion of the leaf, the balance registered a substantial weight gain, indicating the addition of 1 to 5 mg weight (with the same 14 W input to the vibrator stage), independently of whether the leaf was connected to the terminal of the coil or instead to the earth ground via a water pipe. This suggested to us that, whether formed as a DC or AC spark channel, or whether in the form of a corona discharge, the electric gap develops an expansion force (exactly opposite to a Casimir force) on both electrodes, independently of their polarity, which force is responsible for the observed repulsion. Yet, this expansion goes hand in hand with an increase in their weight such that some other process is at work in that electric gap.

To examine this problem further, we assembled a different experiment where the gold leaf **16** was suspended between two large metal plates **18**, **24** placed 20 cm apart, and the leaf was not electrically connected to them or to any other circuit, while attached to the dielectric rod employed to suspend it over the electronic balance. Given that the leaf is suitably and equally spaced from both plates, there is no arcing between it and either plate. The obvious expectation is that, since the electric field bathing the leaf alternates at high frequency (measured in hundreds of kilohertz), and the corona from both electrodes should equalize and balance any electric wind, no lift should be observed. In fact, no lift is apparent, but a most curious observation is made: depending upon which orientation is employed for the plates, the gold leaf either gains or loses 4–6% of its weight. This gain or loss is registered for as long as the coil is on. If the top plate is grounded and the bottom one connected to the different terminal of the secondary, a gain in weight is observed (FIG. **3**). If the connections are reversed, an equal weight loss is registered (FIG. **4**).

Furthermore, in this last instance, if the grounded plate **24** is entirely removed (FIG. **5**), and only the top plate remains connected to the distal terminal of the secondary, the observed loss of weight continues to occur such that effectively this reaction can be obtained with unipolar electric fields of high frequency, and provides a unidirectional force which, once exerted upon metallic objects bathed by its field, can be made to oppose or favour gravity.

Now, these effects can be greatly magnified, in the order of 10-fold, if the same gold leaf is made part of a simple series floating electric circuit where the leaf functions as a large area plate, and is wired in series with a coil **26** which, for best results, should be wound so as to be of a length resonant with the secondary of the Tesla-type coil employed; and this coil is connected in turn to a point antenna **28** upwardly oriented (FIG. **6**). The entire floating circuit is mounted on the rod **20** and this in turn is mounted over the sensitive balance. If both plates are kept as in FIGS. **3** & **4**, the observed weight loss and weight gain respectively vary between 30 to 95% of the total weight of the leaf. Again the gain or loss of weight is registered for as long as the coil is on.

These anomalous findings suggested that whatever is the nature of the energy responsible for the force observed in that high frequency alternating current gap, any metallic object placed in that gap will experience a force repelling it from the electric ground. This force will be maximal if the gap frequency is tuned to the elementary or molecular structure of the metallic object. If the electric ground is placed opposite the actual plane of the earth ground, that force will act in the direction of gravity. If, instead, the electric ground and the earth ground are made to coincide on the same plane, that force will act opposite the direction of gravity, i.e. will repel the metallic object from the ground.

No such weight alteration was observed with solid dielectrics, for instance with polyethylene and other thermoplastic sheets.

These facts rule out the possibility of a hidden electrostatic attraction force acting between the plate connected to the different terminal of the secondary and the gold leaf. First off, such an attraction would be able to lift entirely the gold leaf, as is easily observed with the unipole of any electrostatic generator operating with a few milliwatts output with either negative or positive polarity; secondly, the same attraction, if it existed and were the product of an electric force, would surely be manifested independently from whether the experimental leaf was metallic or a dielectric (as again is observed with electrostatic generators).

The results suggest therefore that whenever a large plate is connected to a Tesla-type coil, it induces in surrounding matter that is not part of its own circuit, a directional thrust oriented opposite the electric ground and, if the latter coincides with the gravitational ground, a thrust oriented opposite free fall.

When this thrust is made to oppose gravity, we believe that its effect upon the gold leaf can be compared to the lifting power imparted to the water molecule when it transits from the liquid to the vapor state and which is associated with the increase in internal (or intrinsic) potential 'thermal' energy ΔU (See Halliday D & ResnickR (1978) "Physics", Vol. 1, section 22–8, p. 489). The "specific latent heat" of water (m*L) contains indeed both an expression for the sensible radiant thermal work involving volume and pressure relations ($W=P(V_V-V_L)$, where P=1 Atm, and V_V and V_L, respectively, are the molar volumes in the vapor and liquid phases), and an expression for a quantity of "latent" energy (ΔU) which is associated with the molecule in the more rarefied state. Hence, the relation for the latter with respect to water vapor:

$$\Delta U=(mL)-[P(V_V-V_L)]$$

We propose that likewise, if a very small portion of the energy of the massfree electric waves is indirectly transformed by massbound charge carriers on that plate into blackbody photons (once those charge carriers shed their electrokinetic energy), the greater portion of those waves are directly transformed in the space adjacent to that plate into the latent energy equivalent to ΔU for the atoms of the surrounding air, and so on, until this process itself is also occurring for the atoms of that gold leaf, thus inducing their nonelectrical weight loss and suggesting the existence of a nonthermal "antigravitokinetic" energy term heretofore unknown to mankind other than as "latent heat" or "internal potential energy".

From this viewpoint, the energy released by any Tesla-type coil to its surroundings would be tantamount to a radiative injection of "internal potential energy" which would confer on local gas molecules a weight cancellation (a cancellation of gravitational mass occurring in the absence of any cancellation of inertial mass—a process which the inventors theorize is explained by the neutralisation of elementary gravitons), and the same process would be equally at work for metallic solids but not dielectric solids.

Gold vapour also deploys a substantial intrinsic potential energy. With an enthalpy of vaporization on the order of H_v=324 kJ mol^{-1}, the molar volumetric work performed by gold vapor at atmospheric pressure at the temperature of vaporization T_v (2,856° C., i.e. 3129 degrees Kelvin) is:

$$W=P\Delta V_{V\text{-}L}=23.58 \text{ kJ mol}^{-1}$$

where $\Delta V_{V\text{-}L}=0.2327$ m^{-3}. The intrinsic potential energy of gold vapor is then given by:

$$\Delta U=H_v-W=300.4 \text{ kJ mol}^{-1}$$

i.e. 12.74 times greater than the volumetric work performed during the phase transition.

It is our contention that this intrinsic potential energy, associated with molecules as their "latent heat", has fine structure that in turn is altered if this energy is released from these molecules and fails to gain a "sensible" thermal form. What is suggested is that the fine structure of "latent heat" is not electromagnetic and obeys instead the molecular function:

$$\Delta U/N_A=\lambda_{n2}^2 cf_{n2}$$

where N_A is Avogadro's number, the wavelength denoted as λ_{n2} is the wavelength-equivalent of the mass of the molecule to which the "latent heat" is associated, obtained by a conversion method proposed in these inventors' theory, and the frequency term f is a nonelectromagnetic frequency term, specifically in this case a gravitational frequency function. Employing the conversion of Joules into m^3 sec^{-2} proposed by these inventors as being exactly 1 J=10N_Am^3 sec^{-2}, and putting the wavelength λ_{n2} down as the wavelength-equivalent of the mass of the gold atom, λ_{Au}, at 1.9698 m, that frequency term f_{n2} can be obtained as being equal to 2.6*10^{-3} sec^{-1}. According to the present inventors' theory, the wavefunction c constitutive of the fine structure of "latent heat" associated with molecules of matter, carries the same wavelength λ_{Au} and its frequency is given in the usual manner by c/λ_{Au}=1.52*10^8 sec^{-1}. The resultant frequency for the non-Planckian unit quantum of "latent energy" associated with each gold atom at the vaporization temperature is then obtained by the geometric mean of the two synchronous frequency terms:

$$[(c/\lambda_{Au})f_{n2}]^{0.5}=624 \text{ Hz}$$

However, this is the signature of that intrinsic potential energy when associated with that gold atom at its vaporization temperature. It is not the signature of the energy quantum itself if it is released from that molecule, nor prior to being absorbed (i.e. in transit), at that same temperature. The fine structure of the same non-Planckian "latent" energy quantum varies to encompass different determinations of the constituent wavelength and frequency functions. The basic relation for the determination of the wavelength of a "latent thermal" energy quantum not associated with matter, but corresponding to one that is, is:

$$\lambda_{n1}=[\Delta U/N_A)/c]^{0.666} \text{ meters}^{-0.333} \text{ seconds}^{0.666}$$

which gives 0.046478 m for the unbound equivalent of the "latent heat" unit quantum of vaporization associated with the gold atom at one atmosphere. The fine structure of the free quantum is still parallel, as given by:

$$\Delta U/N_A=\lambda_{n1}^2 cf_{n1}$$

but now notice how the frequency terms have changed value, with the f_{n1} function having the value 4.65 sec^{-1} and c/λ_{n1} yielding 6.48*10^9 sec^{-1}. The geometric mean of the superimposition of the two frequencies is then—

$$[(c/\lambda_{n1})f_{n1}]^{0.5}=173.7 \text{ KHz}$$

We contend that it is at this frequency that the atoms of gold vapor absorb "latent heat".

However, this is just the overall scenario of what happens at the temperature of vaporization of gold. But at room temperature (e.g. 293 degrees Kelvin), and with respect to processes where there is no sublimation of the atoms of that gold leaf under way (and indeed, once the coil is turned off, the leaf returns to its normal weight), one must infer to a different phase of matter what portion of "latent heat" energy, if any, do the atoms of gold hold in the solid phase lattice. Assuming the same proportionality between the "sensible" and "latent" thermal energy terms for atoms of gold at room temperature, where the unit thermal energy is N_A kT=2.436 kJ mol^{-1}, the gold atom could absorb up to 12.74 times, we speculate, the value of this "sensible" thermal energy, and thus hold N_A kT=31.053 kJ more energy in its own micro-atmosphere.

If this speculation is correct, and employing the above novel methodology, then the mean geometric frequency of the maximal "latent heat" energy quantum of a gold atom at room temperature would be 538 KHz (versus 174 KHz at the vaporization temperature), and once absorbed its mean frequency mode would abate to 201.5 Hz (versus 630 Hz once the atom has vaporized).

To test this hypothesis we employed two different Tesla-type coils having output frequencies of 200 KHz and 394 KHz. The circuit tested was that shown in FIG. **6**, and both coils were operated at 50 KV outputs. Whereas the former coil, closer to the 174 KHz marker, could only systematically produce 10–11 mg of weight cancellation in the gold leaf of the floating circuit, the second coil, closer to the speculated 538 KHz marker, could produce 15–35 mg of weight cancellation in the same gold leaf. The empirical results appear therefore to suggest that our speculation may well be a valid one.

The above-mentioned full wave divider (see FIG. **1**) can be easily coupled to our autogenous pulsed abnormal glow discharge technology as described in our U.S. Pat. No. 5,416,391 to form an alternative source of direct current, ultimately powered by Tesla waves, and such a drive can equally be applied to any other vacuum device that can sustain endogenous oscillatory discharges, whether in the PAGD regime or any other pulsatory regime. For the purposes of experimental and visual determination of power outputs from the divider in question, we have utilized either 2 Torr vacuum tubes operating in the high-current PAGD regime, or 20–100 Torr spark tubes requiring high voltages (2 to 10 KV) for their spark breakdown. As taught in the above U.S. Patent, the output from the full wave voltage divider can be assessed by the energy spent in driving the tube and the motor, whose rotary speed is proportional, within the limits chosen, to the power input.

Two separate sets of experiments set forth in Table I below showed that direct connection of the wave divider to the distal terminal of the coil (set constantly at 6 clicks on the vibrator stage) (FIG. **1**) or to the same terminal but across a large (2 or 3 square feet) plate **30** that increased the capacitance of the secondary (FIG. **7**), presented the same power output in either case (the effect of the plate is to lower the voltage of the output proportional to the increase in current). A substantial increase in power output through the divider is observed only when an identically wound Tesla coil is connected in reverse (FIG. **8**) with the non-common end of its winding **4** not connected, in order to obtain a condition of resonance, and this observed increase is further augmented by now interposing either of the metal plates **18**, **24** between the two chirally connected and identical coils (FIG. **9**). The increase in plate area appears to have the effect of increasing the output for as long as the plate is isolated between the two chiral image coils. Throughout these experiments the input power to the vibrator was fixed at 14 W (60 Hz AC).

TABLE 1

Status	Pulse rate (PPS)	Motor rotation (RPM), M ± SEM
Expt A		
Tesla coil (TC) to divider	2.6	582.5 ± 3.9 (n = 4)
TC to inverted TC, to divider	4.4	621 7.6 (n = 4)
TC to 2 ft^2 plate, to inverted TC, to divider	5	775.25 ± 23.6 (n = 4)
Expt B		
Tesla coil (TC) to divider	2.2	613 ± 5.6 (n = 12)
TC to 3 ft^2 plate, to divider	2.3	605 ± 2.6 (n = 12)
TC to inverted TC, to divider	2.3	722 ± 5.7 (n = 12)
TC to 3 ft^2 plate, to inverted TC, to divider	4.2	877.6 ± 6.5 (n = 12)

In our loss of weight experiments described above, we noted that the phenomenon of weight loss by a metallic body placed in proximity of the coil output continued to be observed when only the plate connected to the distal pole of the secondary was retained. The leaf, although not part of the circuit of the secondary, could however be seen as part of a circuit for the capture of ambient radiant energy, specifically that generated by the coil and, as well, that also possibly picked up, in the process, from other ambient sources. To determine whether the last consideration is a possibility at all, or whether the energy picked up by an analogue of our metallic body or gold leaf in the experiments described above, is entirely a byproduct of the energy transmitted by the plate connected to the distal pole of the secondary, we next determined what would happen if the pick-up for the full wave-divider were placed not at the output from the secondary coil, but from an in all respects identical plate (the Receiver plate R, as opposed to the Transmitter plate T) placed a distance away from, and above, the first one. In other words, the gold leaf is replaced by a receiver plate, and this carries an attached test circuit identical to the test circuit employed to assess directly the coil output.

As shown in Table 2 below, the results of the experiment show that there is no loss of energy picked up at the R plate (FIG. **10**) when compared to the most favourable situation involving the plate **30** (FIG. **9**) interposed between the chirally connected coils. This observation is however not always the case. For best results one should employ iron, gold or silver plates placed parallel to the horizon, with the T plate underneath the R plate. In fact, if one employs instead aluminum plates and suspends these vertically, one can consistently register a loss of output at the divider when changing the divider input from the T to the R plates.

If however the plate R is connected in turn to a second identical coil, also wired in reverse, and this second coil in turn serves as input to the full wave divider (FIG. **11**), then a most curious occurrence takes place—the power output increases considerably (see Table 2), as if the divider circuit had undergone an energy injection not present at the source. Note that the circuits are in fact resonant, but the energy injection contributing nearly 60–66% (for both plate areas in the previous experiment) of the input that we refer to, is not caused by inductive resonance, since the effect of resonance can be ascribed to the set-up described in FIG. **9**. The distance between the plates, as well as their orientation with respect to the local horizon system of the observer also appear to matter, best results being achieved at optimal distances (e.g. for 2 square feet plates the best gap, at 43% RH and room temperature, was at least 6 inches).

TABLE 2

Status	T R distance	Pulse rate (PPS)	Motor rotation (RPM), M ± SEM
$2\ ft^2$ plates			
R plate to inverted TC, to divider	3"	6.7	882 ± 17.5 (n = 4)
	4"	8	906 ± 12.1 (n = 4)
	6"	10	936 ± 46.1 (n = 9)
$3\ ft^2$ plates			
TC to T plate, to divider	0	2.3	605 ± 2.6 (n = 12)
R plate to divider	6"	3.3	890.1 ± 3.8 (n = 12)
R plate to inverted TC, to divider	6"	5.1	1009.2 ± 4 (n = 12)
R plate to divider	8"	4.0	783.1 ± 11.3 (n = 12)
R plate to inverted TC, to divider	8"	5.1	1005.7 ± 6 (n = 12)

We tested the possibility that environmental heat produced by operation of the coil might be the source of the injected energy, the plate of the second system acting possibly as collector for the heat present in the gap. As it turned out, experiments showed repeatedly that in the gap between the T and R plates there was no significant thermal radiation propagating between one and the other. The more illustrative experiments are those in which we identified where the sensible thermal energy appears, and which involved coupling two cavities: the Transmitter-Receiver gap between plates T and R, and a Faraday cage enclosure **34** (see FIG. **12**). The first cavity appears to be much like that of a capacitor: the two identical parallel plates are surrounded by a thick dielectric insulator **32**, and a thermometer T**2** is inserted half-way through it. A thermometer T**1** is also affixed to the T plate, to read its temperature. The second cavity is a simple metal cage, also insulated, with a thermometer T**3** inserted 2 cm into its top. 2–4 cm above the top of the cage there is placed a fourth thermometer T**4**, within an insulated cylinder. If the TC is a source of thermal energy (e.g. IR radiation, microwaves, etc.) we would expect the T plate to be the hottest element from which, by radiation, thermal energy would reach the middle of the first cavity making the next thermometer T**2** second hottest, and that the third thermometer T**3** inside the second cavity, even if it might initially be slightly warmer than the other two, would, over time, become comparatively cooler than either one of the other two thermometers, despite the fact that the rising heat would still be seen to warm it up over time. One would expect a similar outcome for the fourth thermometer T**4**, above the cage. As shown by FIG. **13**, where only the temperature differences ($\Delta T^\circ - T_c^\circ$) between the experimental thermometers and the control thermometer reading the air temperature T_c° of the laboratory are shown, the surface of the T plate warms up by 0.1° C. at 3' after initiation of the run (closed squares), whereas in the space of the T/R gap a diminutive warming, by 0.05° C., is registered after 10' (open circles). Conversely, the temperature inside the cage, at the top (shaded circles) rises by 0.1° C. also by the third minute, and the temperature above the cage itself (shaded squares) rises by a much greater difference of 0.35° C., which remains stable after the eighth minute.

These results show that it is not sensible heat that radiates from the T plate. Rather, some other form of radiation traverses these cavities to generate sensible heat at their metallic boundaries, such that more heat is generated above the R plate (inside the cage) and again above the third plate, i.e. above the top of the cage, than is generated in the T/R gap, i.e. proximally to the T plate. This clearly shows that the Tesla coil is not a significant source of thermal radiation, and that sensible heat can be detected inside and atop the Faraday cage only as a further transformation of the radiant energy transmitted across the T/R cavity.

The same experiment also illustrates that, whatever is the nature of the additional environmental energy being injected at the surface of R plate (as shown by Table 2 results above), it is most likely not thermal radiation, at least not energy in the form of sensible heat. And whatever is the nature of this ambient radiant energy being mobilized by the electric radiant energy transmitted from the T plate, it can produce significant heat inside an enclosure contiguous with plate R.

Since we also know experimentally that this observation of an ambient energy injection at the R plate or R cage depends upon relative humidity, being most easily observable when the latter is low (<50% RH), and being virtually impossible to observe when air is saturated with water vapor, we can infer that water vapor is a good absorber of the electric massfree radiant energy emitted from the T plate. This strongly suggests that this absorption process is tantamount to increasing the potential intrinsic energy ΔU of the water vapor molecules adjacent to the T plate. In the absence of significant quantities of water vapor, when the atmosphere is dry, one may speculate that this absorption process is replaced by what one presumes is a parallel process involving the various gaseous molecules of air. However, either because the air molecules involve molecular species that readily give off this potential energy, as one might speculate is the case with molecular oxygen, hydrogen and nitrogen, or because the air molecules absorb far less "latent" energy (as appears to be the case with inert gases), and therefore there is more of it in the molecularly unbound state (as we explicitly propose as a possibility) and thus available for absorption by the appropriately tuned receiver, the increased ΔU of air molecules conferred by the absorption of the massfree electric radiation in the T/R gap is transferred to the R conductor together with the latent energy which those molecules already possessed before entering that gap. Hence the energy injection and its dependency upon the partial pressure of water vapor, which absconds instead with this "latent" energy and succeeds in withholding it from transmission to the R plate.

If the T/R gap can mobilize ambient energy which is neither electromagnetic nor thermal in nature, but which "latent" energy becomes injected into the divider circuit in electric form, the heat (i.e. sensible thermal energy) produced inside and atop the cage can also be mobilized electrically as input into the divider circuit. The obvious place to look for the emplacement of the cool junction that could convert sensible heat into electrokinetic energy of massbound charges is at the top of the cage, where it is warmest (See top curve of FIG. **13** in shaded squares). This is clearly observed from the results shown in Table 3 below, where the initial temperature difference between the top of the box and the T plate surface was 0.5° C., and the top of the box temperature rose by 0.2° C. after 2.5 minutes when the divider was connected at the junction, versus 0.35° C. when it was not (and the transmitter coil was on).

TABLE 3

Status 3 ft^2 plates	TR distance inches	Pulse rate PPS	Motor rotation RPM, M ± SEM(n = 12)
TC to T plate, to divider	NA	4.2	877.6 ± 6.5
R plate to inverted TC, to divider	6"	5.1	1009.2 ± 4
Top of naked R plate/cage to divider	6"	5.4	1047.1 ± 5.7
Top of insulated R plate/cage exposed to sun, to divider	6"	6.1	1072.4 ± 8.7

For the run performed with the naked R cage, the temperature directly above the top of the cage was 24.3° C., at the outset, versus the control room temperature of 23.9° C. For the run performed with the insulated R cage exposed directly to the sun at midday, on a cool and clear August day, the temperature directly above the top of the cage was 33° C., versus the control air temperature of 18.4° C. The temperature of the cool junction at the top of the cage was 31.9° C. while the run was performed.

It is apparent from the data of Table 3 how a second injection of energy has occurred in the apparatus. If, within the T/R gap, the energy injected appears to be on the order of absorption of "latent heat", at the top of the cage cavity, at the cool junction, the injection is one of radiant "sensible" heat. Moreover, this secondary energy addition could be further enhanced by placing strong insulation around the whole apparatus or the cage itself, and further so, by exposing the whole apparatus to solar radiation.

We next turned our attention to the T/R gap cavity with the intention of determining whether atmospheric conditions or vacua yield the same or different results. We could not, of course, test the same large area plates as have been employed for the studies undertaken at atmospheric pressures. For the present purpose we employed instead large area electrodes (ca 0.2 ft^2) made of high grade stainless steel or even aluminum. Preliminary results showed that these T/R gap tubes, when coupled to the divider circuit, yielded faster pulse rates in the secondary circuit when evacuated than at atmospheric pressure. The strength of the corona discharge also intensified, as it eventually became replaced by a normal glow discharge. For purposes of improved spatial capture of (1) the electric massfree energy radiated from the T electrode and (2) the non-radiant latent thermal energy mobilized by it to be collected electrically at the R plate, an axial cylindrical T electrode was inserted inside a larger concentric cylinder or between two common plates of large surface area (e.g. >100 cm^2) functioning as the R electrode(s), in a dielectric container suitable for evacuation (glass, polycarbonate), at a typical distance of at least 3 cm between electrodes, and the entire device was tested at different pressures. The secondary circuit connected downstream from the full wave divider was as shown in FIG. **14** (employing an autogenous pulsed abnormal glow discharge, or PAGD, converter circuit), with the PAGD reactor 36 set at 10 Torr (in light of the high-voltage input, which varied between 1500 and 3200V) and gave the results presented in Table 4 below. We should remark also that these pulses charged the charge pack CP through the coupling capacitors **38**, bridge rectifier **40** and reservoir capacitors **42**, and blocking diodes **44**, as expected from the prior art represented by our patents related to PAGD devices.

TABLE 4

T/R tube Pressure (Torr)	Pulse rate (PPS)
760	0.376
0.025	0.513

The effect of the vacuum in the T/R gap tube seems to be dual. By transforming the corona discharge into a normal glow discharge, it increases the local production of photons (probably associated to the formation and discharge of metastable states in the plasma), and at the same time, increases the pulse rate in the output circuit and thus, in all probability, the energy injected in the T/R gap cavity. But this did not yet permit us to confirm whether or not it is "latent heat" energy of the plasma molecules which is being tapped at the receiver plate, even if it be plausible in principle that plasmas may effect more efficient transfer of "latent heat" to tuned receivers than atmospheric gases.

The vacuum dependency of the pulse rate of the PAGD reactor employed as example in the secondary circuit downstream from the divider is also rather well marked, with the fastest pulse rates being registered at 1 Torr for the sample run shown in Table 5 below.

TABLE 5

T/R tube Pressure (Torr)	Pulse rate (PPS)	PAGD Reactor Pressure (Torr)	Voltage (across divider)
0.025	0.115	90	4.5 kV
0.025	0.1553	75	3.5 kV
0.025	0.183	60	3.3 kV
0.025	0.291	30	
0.025	0.513	15	1.6 kV
0.025	0.602	10	1.4 kV
0.025	2.9	2	0.53 kV
0.025	4.1	1	0.45 kV

It is worth noting here that the illustrated polarity of the wiring of the PAGD reactor tube, as shown in FIG. **14**, is best for purposes of sustaining regular autoelectronic emission at high voltage. The reverse configuration, with the center electrode negative and the plates positive favours instead heating of the cathode and a lapse into a normal glow discharge.

We tested a similar arrangement to that shown in FIG. **14** above, but with a PAGD motor circuit (see our U.S. Pat. No. 5,416,391). A split phase motor **44** replaces the rectifier and charge pack, and the PAGD reactor is operated at the same pressure of 15 Torr, as shown in FIG. **15**. The T/R gap tube tested had a longer plate distance (2"), with one plate now functioning as Transmitter and the other as Receiver. Note also the different wiring of the PAGD reactor. The results, as shown below in Table 6, present pulse per second (PPS) and motor revolutions per minute (RPM) curve trends that appear to be analogous and parallel to the well-known Paschen curves for breakdown voltage in vacuum—such that the T/R gap performs better either in the atmospheric corona discharge mode, or in the high vacuum normal glow discharge (NGD) mode, than in the low breakdown voltage range of the curve where the discharge forms a narrow channel and takes on the appearance of an "aurora" transitional region discharge (TRD).

TABLE 6

T/R tube Pressure (Torr)	Pulse rate (PPS)	Motor rotation (RPM), M ± SEM (n = 17)	Discharge Type
760	2.8	751.2 ± 7.1	Corona
100	2.1	611.5 ± 5.1	TRD
20	2.4	701.9 ± 4.6	TRD
0.006	2.8	748.4 ± 9.3	NGD
0.003	3.0	819.4 ± 6.3	NGD

These results suggest that plasmas with high lateral dispersion, i.e. formed over large electrode areas (e.g. corona and NGD plasmas) and thus devoid of pinch, are more likely to mobilize electrically the intrinsic potential energy of the molecular charges than pinch plasmas appear to be able to do (e.g. TRD plasmas). Apparently also, the greater the vacuum drawn from the T/R gap cavity, the more efficient does the transfer of this intrinsic potential energy, ie the massbound latent heat, to the electrokinetic energy of the charges circulating in the receiver circuit become. At about $6*10^{-2}$ Torr, this transfer in vacuo is comparable to that observed under atmospheric conditions and thus for a much greater density of molecules.

We investigated whether it Is possible to tap the latent heat energy of water molecules. It is possible that in the vapour phase they can effectively hold onto their latent energy—but once closely packed in liquid phase could they give off some of it? To test this hypothesis we immersed the T/R gap in a glass water tank. The motor employed for these tests was a high-speed 2 phase drag-cup motor (see FIG. **18** and associated description), wired in split-phase with two identical phase windings capacitatively balanced, and the galvanized iron plates had one foot square areas. The results are shown in Table 7 below, and clearly indicate that it is possible to tap—within the T/R cavity—the ‘latent heat’ of water in the liquid phase. As observed, immersion of the T/R cavity in water increased the motor output speed 22% (=12,117/9888)*100). This corresponds to a 50% increase in power output, from 18 W at 9888 rpm to 27 W at 12,117 rpm:

TABLE 7

	Pulse rate PPS	Motor rotation RPM M ± SEM	T/R distance cm
Direct from TC	0.3	8076 ± 89.3	NA
TC to T plate	0.5	9888 ± 78.7	NA
R plate	2.75	12117 ± 29.8	30
R plate	2.9	12203 ± 55.9	60

Thus the use of ion-containing water or other ion-containing aqueous liquid in the cavity promotes long distance propagation and a greater injection of latent and thermal energies in the receiver circuit. Such a result is not achieved if the cavity is filled with deionised water.

The preceding results lead therefore to the design of a presently preferred apparatus, based on these findings, for the conversion of massfree electric energy, “latent heat” energy and “sensible” heat energy into conventional electric energy, as shown in FIG. **16**, which integrates all the separate findings and improvements. The winding **6** of the Tesla coil at the bottom is driven in the usual manner employing a vibrator stage **2** to pulse the primary coil **4**. The distal pole of the secondary **6** is then connected to a circular metal plate T which is one end of an evacuated cylindrical cavity, connected to a vacuum pump or sealed at a desired pressure, or which forms a still containing water or other aqueous solution or liquid. This cavity constitutes the transmitter/receiver gap, and is therefore bounded by a dielectric envelope and wall structure **32**, with the circular receiver plate R as its top surface. In turn this plate R serves as the base of a conical Faraday cage **34**, preferably air-tight and at atmospheric pressure, but which could also be subject to evacuation, which conical structure carries at its apex provisions for a cold junction **45** and any possible enhancement of the same junction by surface application of different metallic conductors that may optimize the Peltier-Seebeck effect. The output from the cold junction where sensible thermal energy is added to the electrokinetic energy of charge carriers, is also the input to the distal end of the winding **6** of the chiral coil arrangement that sustains resonant capture of all three energy flows ((1) massfree electric waves of a longitudinal nature, (2) true “latent heat” or the intrinsic (thermal) potential energy, and (3) the thermokinetic energy of molecules, (i.e. “sensible” heat) and, placed in series with the input of the full wave divider **8**, **10**, feeds the circuit output from the series capacitors **12**, **14** grounded at their common tap. In the T/R gap, the transmitted electric longitudinal wave energy is captured along with any intrinsic potential energy shed by molecules caught in the field. Within the R element, expanded into an enclosure that guides “sensible” radiant heat, the latter is generated and then recaptured at the cold junction.

The apparatus consisting of the cylindrical T/R gap cavity and the contiguous conical cage is then preferably finished in gloss white and cylindrically enveloped within a matte black container **46** by effective thermal insulation **48**, the latter terminating at the height of the bottom disc T. Apparatus (not shown) may be provided to move the plate T vertically to adjust the T/R gap.

Another alternative embodiment of the apparatus is shown in FIG. **17**. Here the circuit driving the apparatus is as we have set forth in our prior patents, which employs an autogenous pulsed abnormal glow discharge tube **50** in the configuration shown, supplied by a battery pack DP through blocking diodes **52** and an RC circuit formed by resistor **54** and capacitor **56** to drive the primary 2 of a first Tesla coil to obtain at the distal pole of the secondary **6** the energy to be injected to plate T in the form of a central electrode of a coaxial vacuum chamber (sealed or not), of which the cylindrical metallic envelope forms the receiver plate R, the latter being placed centrally inside the conical cage **34** and contiguous with its walls and base. The top and bottom of the coaxial chamber carries suitable insulating discs, preferably with O-ring type fittings. Again, the apparatus is enclosed in insulation within a cylindrical container **46**, and the input into the capture circuit driven from the full wave divider is taken from the cold junction **45** at the apex of the air-tight cage. The output circuit is similar to that of FIG. **15**.

We have found however that even when the component values in the motor driver and motor circuits are carefully selected so that these circuits are co-resonant with the damped wave (DW) component of the motor driver pulses, the motor power output falls well short of that which should theoretically be attainable. In an endeavour to meet this problem, we replaced the squirrel cage type induction motor **44** by a drag cup motor of type KS 8624 from Western Electric in the expectation that the low inertia non-magnetic rotor would allow better response to the DW component. This motor is similar to one of the types used by Reich in his experiments. Although results were much improved they still fell short of expectations. Replacement of this motor by an inertially damped motor of type KS 9303, also from Western Electric, provided much better results as discussed below.

Fundamentally, the difficulties we encountered stemmed from the inability of motor couplings to respond efficiently and smoothly, and at the same time, to the pulse and wave components of DW impulses: that is, simultaneously to the high-intensity peak current pulses (the front end event), the DC-like component, and to the damped wave trains these cause, i.e. the pulse tails (or back end event) or AC-like component. This difficulty is present even when we just seek to run induction motors from the DW impulses of a Tesla coil, the very difficulty that led Tesla to abandon his project of driving a nonferromagnetic disc rotor mounted on an iron core bar stator with damped waves.

We believe that the key to the capture of the massfree energy flux output in electric form by Tesla transmitters, including any injected latent or thermal energy that have undergone conversion into electrical energy is to employ the tuned, unipolar, Y-fed, PAGD-plasma pulser driven split-phase motor drive we have invented (U.S. Pat. No. 5,416,391) in conjunction with an inertially damped AC servomotor-generator (see FIG. **18**): this has a motor shaft **64** which couples a drag-cup motor rotor **60**, preferably of aluminum, silver, gold or molybdenum, directly to a drag-cup generator rotor **62** that drives a permanent magnet (PM) flywheel **66**, freely rotatable in bearings **67**, that provides inertial damping. The shaft **64**, journalled by bearings **61** in the casing of the motor **44**, provides a power output through optional gearing **68**. The phase windings of the motor **44** are wound on a stator core **70** having concentric elements between which the rotor or cup **60** rotates. This structure makes it ideal for the capture of the DW impulses, whether sourced in the transmitter, amplified in the T/R cavity or sourced in the plasma pulser, all in synchrony. Effectively the motor couples the damping action of the drag-cup sleeve motor rotor, which action, as we have already found for the KS-8624 motors, is quite effective at absorbing the front-end DC-like event, with the inertial damping of the PM flywheel upon the drag-cup sleeve generator rotor, that in turn is quite efficient at absorbing the back-end AC-like wavetrain event.

The KS-9154 motor used by Reich was not an inertial damped AC drag-cup servomotor-generator. Had Reich succeeded in overcoming the limitations of his 2-phase OR Motor solution, as we have now shown it is possible to do—by applying the Function Y circuit to the PAGD split-phase motor drive which we invented, his motor would have suffered the same limitations which we encountered with the KS 8624 motor.

Any motor, by itself, has an internal or inherent damping whereby the acceleration only vanishes when the rotor is running at constant speed. For motors that operate on the basis of the drag principle, where the asynchronous slip is actually constitutive of the motor action, by inducing eddy currents in the rotor, the inherent damping is always more pronounced than for other induction motors. The damping or braking torque is produced when a constant current flows through a rotating drag disc or cup.

Aside from this inherent braking, dampers can also be applied to servomotors to further stabilize their rotation. They absorb energy, and the power output and torque of the motor is thereby reduced. Optimal operation of servomotors requires both rapid response on the part of the rotor to changes in the variable or control phase, and a stable response that is free from oscillation, cogging and overshooting. The rapid response is assured by employing low inertia rotors, such as drag-cups or cast alloy squirrel cages, and the overshooting and oscillation are reduced to a minimum by damping or a retarding torque that increases with increasing motor speed. Typically, in a viscous-damped servomotor, the damper is a drag-cup generator mounted rigidly on the shaft of the motor rotor, and the generator drag-cup rotates against the stator field of a static permanent magnet field. The generator develops a retarding torque directly proportional to speed, and the energy absorbed by the damper is proportional to speed squared. The damping can be adjusted and, as it increases, the same amount of input power yields lower torque and motor speeds. Inertial-damped servomotors differ from viscous damped motors in that the permanent magnet stator of the drag-cup generator is now mounted in its own bearings, either in the motor shaft or on a separate aligned shaft, forming a high-inertia flywheel. This means that, whereas the motor rotor always experiences a viscous damping in viscous-damped servomotors, in inertial-damped servomotors the drag cup motor rotor only experiences a viscous damping while accelerating the flywheel, with the damping torque always opposing any change in rotor speed. Once the flywheel rotates synchronously with the rotor, all damping ceases. Note that this viscous damping is carried out via the coupling of the drag-cup generator rotor, rigidly affixed to the motor rotor, to the PM flywheel, so that their relative motion generates the viscous torque proportional to the relative velocity. Use of drag-cup sleeve rotors in inertially damped servomotors was largely supplanted by squirrel cage rotors once the latter became produced as cast alloy rotors. Since inertially damped motors can be used in open and closed-loop servo applications, and present better stability—even in the presence of nonlinearities—and higher velocity characteristics than other induction motors do (Diamond, A (1965) "Inertially damped servo motors, performance analysis", Electro-Technology, 7:28–32.), they have been employed in antenna tracking systems, stable inertial-guidance platforms, analog to digital converters, tachometers and torque tables. The typical operation of an inertially damped servomotor is as follows: with the reference phase fully excited, the motor rotor—fixedly linked to the generator rotor, as well as the flywheel, remain immobile; once power is applied to the control phase, the motor rotor immediately responds but the flywheel remains at rest. However, as the drag-cup generator **62** is forced to move through the permanent magnetic field of the flywheel, it creates a drag torque that slows down the attached motor rotor proportionally to the acceleration that it imparts to the flywheel that it now sets into motion, thus creating the viscous damper. As the flywheel accelerates, the relative speed of the motor with respect to the flywheel, as well as the damping torque, decrease until both motor and flywheel rotate synchronously and no damping torque is exercised—at which point the drag on the motor cup exerted by the generator cup is negligible. The KS-9303 motor is an inertial damped servomotor but is differentiated with respect to other inertially damped motors, in that (1) it employs a drag-cup sleeve motor rotor made of aluminum, very much like that of the KS-8624, but with slightly altered dimensions and with a shaft extension for the drag-cup copper generator rotor, and (2) the moving flywheel structure was journalled on a separate, fixed shaft, as already described with reference to FIG. **18**. Now, in principle, even application of minimal damping decreases motor efficiency, resulting in diminished torque and speed. Whether the inertial-damped motor has a drag-cup rotor, a sleeve rotor or a squirrel cage rotor, the damping increases the rotor slip. Laithwaite considers drag-cup motors as being "dynamically inferior to their cage counterparts" (Laithwaite, ER

(1957) "Induction machines for special purposes", London, England, p. 323.). If we now add a viscous damping and retarding torque, we should not be able to get much more than a 55% efficiency in the best of conditions. On the other hand, the inertial damping arrangement described will only abstract or supply energy when the motor rotor is accelerating or decelerating relative to the flywheel.

These drag-cup motors, whether inertially damped or not, develop a constant torque at constant rpm for a given supply frequency and a suitable phase shift capacitance. For each frequency the motors respond to, there is an optimum resonant split-phase capacitance, but other values nearby are still suited for operation, and for each value of capacitance, there is an optimal frequency the motors respond to. For example the KS-8624 motor responds best at 450 Hz when a 1 μF capacitance is employed, responds best at 250 Hz when a capacitance of 10 μF is employed, and responds best at 60 Hz, when a capacitance of 100 μF is employed. As the capacitance increases, the resonant CW frequency of the motor is displaced to lower values. If we fix the capacitance at a value (e.g. 10 μF) suitable for testing the frequency response at a fixed voltage of 12 VAC, the observed result for both the KS-8624 and KS-9303 motors show a response distribution of the motor rotary velocity that has an identical peak at 250 Hz for both motors, with the response decreasing to zero smoothly on both sides of the peak. These results indicate that, when wired as a split-phase motor, the motor rotary velocity varies not as a function of voltage or current, but as a function of frequency when the phase-splitting capacitance is fixed within a suitable range, there being an optimal frequency mode for each value of suitable capacitance, with lower values of capacitance favouring higher frequency modes. For a given frequency and capacitance, the motor rotary velocity remains essentially constant and independent from voltage and current input, and thus at a plateau. Torque, in the same circuit arrangement, follows exactly the same pattern as rotary velocity, as a function of input frequency at a fixed potential. Torque is linearly proportional to rpm in these motors when they are split-phase wired, and rpm linearly proportional to CW frequency, which makes them ideal for experimentation and determination of power output computations. Moreover, since these are drag machines, the slip itself determines the rotor currents and these are susceptible of tuning such that their retardation and relative position in the field can find resonant modes for varying CW frequency and capacitance.

In the circuit of FIG. **17** when using the KS 9303 motor, the inertial damping of the flywheel coupling retards the motor rotor currents sufficiently to allow them to build up torque, with the entire motor assembly serving as the preferred sink for all of the energy, massfree and massbound, captured by the receiving coil circuit with a drawing action established by the motor on the circuit, and providing satisfactory absorption by an inertial damper of the combined, synchronized, damped wave impulses, those occurring at a low frequency as a result of the firing of the PAGD reactor, and those occurring at a higher superimposed frequency—sourced in the transmitter circuit and picked-up by the receiver plate and coil. The action of each DW impulse train itself comports two different events, the DC-like auto-electronic-like discontinuity that sets the motor into action and initiates the rotor currents, and the AC-like damped wavetrain that supports the consistency of those rotors. The concentration of current required to kickstart the motor is provided by the DW impulses of the PAGD reactor, whereas, once the motor is in motion, and particularly, once it is stabilized by the flywheel, the cumulative action of the higher frequency DW impulses makes itself felt by accelerating the rotor to an optimum rotary velocity.

For the next series of tests we employed the basic circuit diagram of the improved motor shown in FIG. **19**. The transmission station is the typical Tesla transmitter with a line-fed, 60 Hz vibrator stage. At the line input to the first stage, we place a calibrated AC wattmeter (Weston Model 432), and a Beckman 330B rms ammeter in series with the hot lead, we set the vibrator stage for 41 clicks, consuming between 28.5 and 35 W, as depending upon circumstances yet to be narrated. This consumption was confirmed by driving the coil from an inverter powered by a 12 volt battery. The inverter consumes 2.16 watts, and is 90% efficient. The total consumption from the battery was 42 watts (12V at 3.5 A); once the 2.16 watts is deducted and the efficiency taken into account, we obtain the same 36 W (vibrator stage at max, i.e. 47 clicks, in this experiment). The T/R gap is adjusted to 3", and 2 square ft plates are employed. Transmitter and receiver coils are tuned, and so are the plate capacitances, to 250 kHz, also the capacitances of the Function Y circuit connected at the output of the receiving coil. The rectified voltage and current generated by the transmitter secondary and by the transmitter plate was ascertained with a coil-tuned wave-divider (Function Y) circuit by loading it with different resistive values. The results constitute a measure of the massbound electrical power output directly from the transmitter apparatus. The same method was employed to ascertain the voltage, current and power of the massbound charges circulating in the receiving plate and coil circuit. The results are shown in Table 8 below:

TABLE 8

Massbound currents rectified by Function Y at the output of the Tesla transmitter, transmitter plate and receiver plate, as a function of the bleeding resistance employed in each of the function Y arms

	VDC (kilovolts)	ADC (amp)	WDC (watts)	R/arm (Mohm)
Direct from 20°	42–50	$3*10^{-5}$	1.26–1.5	500
From 2° (T) plate	26	$2*10^{-5}$	0.52	500
From 2° (R) plate	15.1	$1.25*10^{-5}$	0.189	500
Direct from 2°	20.4	$3.4*10^{-4}$	6.936	50
From 2° (T) plate	15.2	$2.4*10^{-4}$	3.648	50
From 2° (R) plate	9	$1.2*10^{-4}$	1.08	50
Direct from 2°	3.3	$1.75*10^{-3}$	5.775	1
From 2° (T) plate	3.5	$2*10^{-3}$	7.0	1
From 2° (R) plate	2.95	$1.6*10^{-3}$	4.72	1

The results indicate that the highest massbound power assembled by the secondary transmitter circuit does not exceed 7 watts—and this is directly output from the secondary **26** when the load is 50 Mohm, or from the transmitter plate when the load is 1 Mohm. The massbound electric power emulated by the receiving circuit (plate, coil and Function Y without the plasma pulser circuitry) never exceeds the massbound electric power output directly by the transmitter, and peaks when the resistive load value (1 Mohm) approaches the pre-breakdown resistance range of the vacuum tube, at 4.72 W. These findings then indicate that when the transmitter circuit is consuming a maximum of 35 W, a typical output from the secondary of the transmitter is 7 W, and at 3" of distance within the proximal field of the latter, the pick-up by a tuned receiver will be on the order of 5 W of massbound current duplicated within the receiving coil. The loss in the first stage is therefore on the order of sevenfold. Continuing with the description of the circuit of FIG. **19**, a 128 cm^2 plate area, 6 cm gap PAGD reactor is employed, connected as described in our prior art to a high-vacuum rotary pump (Correa, P & Correa, A (1995) "Energy conversion system", USPTO U.S. Pat. No. 5,449, 989.). Pressure readings were obtained with a thermocouple gauge during the operational runs. The KS-9303 motors to be tested are then connected to the PAGD reactor in the usual capacitatively-coupled, inverter fashion described in our prior art (Correa, P & Correa, A (1995) "Electromechanical transduction of plasma pulses", USPTO U.S. Pat. No. 5,416, 391.). Their rpm is detected by a stroboscopic tachometer and fed to a Mac Performa 6400 running a motor algorithm program calculating the power output. Motor measurements were made at five minutes into each run for the unloaded motors, and at ten minutes for the inertially damped motors. All experiments were carried out in the same work session. The experimental determination of the continuous rotary power output as a function of the reactor pulse rate confirmed that the improved circuit develops maximum rotary capture of the massfree energy in the receiver circuit at the lowest rates of pulsation, just as we have previously found for the conversion system of U.S. Pat. No. 5,449,989. Furthermore, the data showed that even motors of type KS-8624 are able to output power mechanically in excess of the massbound power output by the transmitter (7 W) or captured by the receiver (5 to a max of 7 W), once the PAGD rate decreases to 1.5 PPS. Such an anomaly can only be explained by the system having become able to begin capturing the massfree energy flux in the receiver circuit that we know already is output by the transmitter circuit. But this excess mechanical power is still, and clearly so, less than the power input into the transmitter. It represents a power gain with respect to the secondary, but a loss with respect to the primary. The full breadth of the capture of the massfree electric energy flux circulating in the receiver circuit is not seen until the motors are resonantly loaded because they are inertially damped.

The KS-9303 motors, once inertially damped, and thus loaded, are able to to recover enough power from the massfree energy field to develop a mechanical power, not just greatly in excess of the massbound power of the secondary, but also greatly in excess of the massbound power input to the vibrator stage and the primary, at 28 to 35 W. Once the pulse rate approaches the same 1.5 PPS marker, mechanical power in excess of the massbound electric power input to the primary becomes evident, peaking at nearly three times that input. In fact, the highest output recorded was also obtained with the lowest input to the transmitter circuit, the highest exact coefficient observed in this experiment being 100.8 W/28 W=3.6. Furthermore, with respect to the secondary massbound output, the same mechanical rotary output represents a much greater overunity coefficient of performance, on the order of 14.4 times greater. This is at least partly the result of the receiver and motor capture of the massfree electric energy output by the transmitter, and may be partly the result of massfree energy engrafted by the PAGD regime in the PAGD reactor. Reviewing the mechanical power output results as a function of increasing vacuum in the PAGD reactor and at different output power levels, any motor performance below the 5–7 W limit of the traditional massbound output power of the secondary represents an output mechanical power loss with respect to both the massbound secondary output and the massbound primary input. All the results for pressures down to 0.03 Torr fall in this category, and thus represent a very inefficient coupling to the PAGD regime. Any motor performance between 7 W and 28–35 W represent a loss with respect to the electrical power input to the transmitter system, but a net gain of power with respect to the massbound secondary power output. None of the non-inertially damped motors tested were able to perform outside of this area, under the test conditions. With more efficient primary to secondary couplings in the transmitter station, however, one could advantageously employ these motors alone to extract some of the massfree power of the secondary or to operate them in enclosed vessels without conventional external electrical connections. To reach satisfactory levels of recovery of massfree energy, one must dampen the superimposed DW impulses. Hence, all results showing outputs in excess of 35 W were obtained using the inertially damped KS-9303 motors, and represent a net overunity power gain over both the power input to the primary and the massbound power output by the secondary, or the massbound power emulated by the receiver circuitry. This happens when the PAGD pulse rate falls to 2 PPS, with the rotary power output steeply increasing as the rate falls to 1 PPS.

One of the interesting features of the motor circuitry we have proposed is that it can operate with pulsed plasmas in both the TRD and the AGD regions, the least efficient response occurring in the NGD region near the Paschen minimum. One might think that the voltage depression would allow increased current intensity supplied to the motors, but in fact that is not observed, with the flashing of the NGD yielding erratic oscillations and low values of current. In keeping with the notion that the TRD plasma is mainly composed of lagging positive ions, whereas the PAGD plasma is mostly an electron plasma, the observed direction of rotation of the motors is opposite in the TRD region to that of the AGD region. The NGD region therefore marks the depression where the velocity vectors change direction. In the second or PAGD region, motor operation is very quiet, unlike what is observed in the TRD region.

Part and parcel of the tuning of the circuit components is the selection of the optimum capacitances employed to couple the PAGD reactor to the motor circuit and split the phase to feed the auxiliary winding of the motor. We have experimented with capacitances ranging from 0.5 to 100 μF, and found that best results—for the specific circuit in question (including the characteristics of the transmission)—were such that the optimal value of the PAGD coupling capacitance lay near 4 pF, and the phase splitting capacitance, near 1 to 4 μF, depending upon weather conditions—in good weather days lower capacitance values can be implemented, and in bad weather days higher capacitances are needed. For ease of comparison in demonstrating the need to tune the circuit by employing optimal capacitances in those two couplings (reactor to motor, and motor phase coupling), we employed the same capacitances in both circuit locations. A comparison of tests using 1 and 4 μF values shows the difference caused by changing those capacitances from their optimal value: across all discharge regions of the pressure range that was examined, the four motors tested operating with greater motor speeds when the capacitances are set to 4 μF rather than to 1 μF. The less efficient performance obtained with 1 μF capacitances fits the inverse correlation of pulse power with increasing pulse frequency, such as we have found for the PAGD regime. This is made evident by a comparison of rpm versus pulse rate for the two capacitance values being considered. They demonstrate the higher pulse rates observed with the lower capacitance, that correlate with the slower motor speeds, and result in lower efficiency of the motor response. The results equally indicate that low capacitance values increase the pulse rate, but if this increase is out of tune with the rest of the circuit values, it results in power waste because it imposes a rate that is not optimal.

We have also experimentally determined that the efficiency of the system is affected by external weather conditions, higher efficiencies being noted on a fine bright day than under poor weather conditions even though the apparatus is not exposed to such conditions. This may reflect a diminution under poor weather conditions of latent massfree energy that can be taken up by the system.

The observed high efficiency of circuits including inertially damped motors indicates that the phenomenon does not reduce to a mere optimal capture of DC-like pulses produced by the reactor in what is essentially an AC motor circuit. Effectively, the pulsed plasma discharge deploys a front-end, DC-like pulse, or discontinuity, but this is followed by an AC-like damped wave of a characteristic frequency (having a half-cycle periodicity identical to that of the front-end pulse) to which the motor circuit also responds. Moreover, the massfree electric radiation from the transmitter circuit itself induces—in the receiver antenna, coil and circuit, and in the reactor discharge itself—the train of finer damped wave impulses responsible, after conversion through the wave-divider, for the massbound rectified current that is employed to charge the plasma reactor to begin with. Serving as trigger of the plasma discharges in the reactor are the DW impulses circulating in the receiver circuit, such that the two different lines of DW impulses, in the receiver circuit (for example 120 PPS for the pulses and 154 kHz for the waves) and from the reactor, are synchronized by interpolated coincidences, since their pulse and wave frequencies are different. Ideally, these two superimposed DW frequencies are harmonics or made identical. The receiver stage involves capture of the massfree electric energy received from the transmitter, duplication of the massbound current in the receiver coil, and injection of latent and sensible thermal energy in the T/R gap cavity that augments the emulated massbound current. The massbound current is employed to charge the wave-divider capacitance bridge and therefore the reactor. In turn, the plasma pulses from the reactor are superimposed with the DW impulses from the receiving coil, and together they are coupled to the split-phase motor drive. Hence the first receiver stage employs the totality of the energy captured in the T/R gap cavity-massfree electric energy transmitted by the T plate, latent and sensible thermal energy injected at the surface of the R plate—and produces in the receiving coil a massbound current comparable to that assembled in the transmitter coil by the action of the primary. The massbound current is stored in the wave-divider bridge and used to drive the plasma reactor in the PAGD region. Subsequently, the autogenous disruptive discharge that employs a substantial electron plasma generates both a concentrated, intense flux of massbound charges in the output circuit, and a massfree oscillation of its own. The damped motor is therefore fed directly with (1) the intense massbound current output from the reactor; (2) the pulse and wave components of the massfree electric energy captured by the receiver plate and coil (and matched by conduction through the earth), and which are gated through the wave-divider and the reactor for the duration of the PAGD channel; and (3) any massfree latent energy taken up from the vacuum by the PAGD event. Once the motor is set into motion, and is resonantly loaded with an inertial damper, we believe that it will also respond to the much weaker DW impulses captured by the receiver, since these impulses encompass both a DC-like front end—further enhanced by analytic separation through the wave-divider—and a damped wave at 154 kHz. Essentially the DW impulses that are ultimately sourced in the transmitter -and received unipolarly through the T/R gap—have sufficient DC-like potential (plus all the other requisite physical characteristics, such as frequency) to contribute directly to the motor response, once the motor has gained substantial speed (for they lack the current to set it into motion, one of the contributions from the plasma pulser). This is the case, provided that the motor itself is suited for absorption of both DC-like pulses and AC-like damped waves, which is precisely the case with motors of the type shown in FIG. **18** since the inertia of the flywheel is overcome through homopolar absorption of the damped oscillations simultaneously in the motor drag-cup rotor and in the generator drag-cup rotor.

We also tested these inertially damped motors in the traditional DC power supply-driven PAGD circuit we have taught in our previous patents, that is, circuits with an overt HV DC power source, and thus in the absence of any Function Y circuit or transmitter circuit. Here then, only the DW impulses generated by the PAGD reactor can account for the motor response. The tube employed (A31) had an area of 256 cm^2, and a gap distance of 4 cm. Coupling capacitances employed were 4 μF for the inverter coupling, and 1 μF for the split phase motor coupling. The DC power supply delivered up to 1 ampere of current between 150 and 1,000 VDC, and the ballast resistor was adjusted to 215 ohms. Having determined the basic physical characteristics of the reactor's behaviour in the circuit under consideration, we conducted our experiment in the PAGD region. We chose a pressure of 0.6 Torr, just off from the Paschen minimum, as we intended to benefit from the lower sustaining voltage that it affords. The experiment basically consisted of increasing the sustaining voltage at this fixed pressure in the PAGD regime, and measuring the diverse physical parameters of the circuit and motor response in order to ultimately ascertain the difference between the input electric DC power and the output mechanical rotary power. We first looked at how the motor rpm response varied as a function of the sustaining voltage (V_s): the results illustrate the importance of starting close to the Paschen minimum in the pressure scale, since the KS-9303 motors reach plateau response (at 17,000 rpm) when the reactor output voltage nears 450V. Any further increase in potential is simply wasted. Likewise, the same happened when we measured motor speed as a function of increasing peak DC current, plateau response being reached at 0.1 ADC. Again, any further increase in current is wasted. Essentially then, the optimal power input to the reactor when the latter's output is coupled to the motor, lies around 45 watts. This is a typical expenditure in driving a PAGD reactor. As for pulse rate we once again find a motor response that is frequency proportional in the low frequency range, between 10 and 40 PPS (all pulse rates now refer solely to PAGDs per sec), but once rates of >40 PPS are reached, the response of the motor also reaches a plateau. The observed increment in speed from 40 to 60 PPS translates only into an increase of 1,000 RPM, from 16,000 to 17,000 RPM. So, we can place the optimal PAGD rate at ca 40 PPS. The DC electric power input to drive the PAGD reactor was next compared to the rotary mechanical power output by the inertially loaded motor, driven in turn by the reactor. This comparison was first carried out with respect to the PAGD rates. The motor response far exceeds the conventional input power, indicating that the whole system can be tuned to resonance such that optimal power capture inside the reactor takes place, the critical limit rate lying at around 60 PPS, when the motor response is firmly within the pulse response plateau. At this juncture, the break-even efficiency for the measured rates of energy flux over time reach 700% (overunity coefficient of 7), in keeping with the observations and the values we have made in the PAGD conversion system. In the proportional part of the curve, before the plateau is reached, even greater rates of break-even efficiency—up to >1,000% were registered. These results constitute the first time we have been able to confirm the presence of output energy in excess of break-even over conventional mass-bound energy input in the PAGD inverter system, and the results are comparable to what we have observed and previously reported for the PAGD converter system. At pulse rates greater than 60 PPS a greater input power results in decreased efficiency, also translated into a noticeable heating of the reactor and motor. And this is all the more remarkable as experiments we have conducted with inductive tuning of PAGD reactors, or employing PAGD reactors as replacements for the primaries of Tesla coil assemblies, and still, more recently, with the PAGD inverter circuit driving motors, have all shown that it is possible to operate these reactors with minimal mirroring and heating, preserving essentially the cold-cathode conditions and yet focusing the plasma column so that deposition on the insulator is negligible. It appears that above a certain threshold of optimal efficiency, surplus input energy is just dissipated thermally by both the reactor and the motors.

It should be understood that the above described embodiments are merely exemplary of our invention, and are, with the exception of the embodiments of FIGS. **16**–**19** designed primarily to verify aspects of the basis of the invention. It should also be understood that in each of these embodiments, the transmitter portion may be omitted if an external or natural source of Tesla waves is available, provided that the receiver is tuned to the massfree radiation mode of the source. For example if solar radiation is available in which the massfree component has not interacted with the earth's atmosphere (as in space applications), the receiver is tuned to the voltage wave of the massfree radiation sourced in the sun, e.g. by using a Tesla coil in the receiver constructed to have an appropriate voltage wave close to the 51.1 kV characteristic of such radiation.

We claim:

1. A device for the conversion of massfree radiation into electrical or electrokinetic energy comprising a transmitter of massfree electrical radiation having a damped wave component, a receiver of such radiation tuned to resonance with the damped wave frequency of the transmitter, a co-resonant output circuit coupled into and extracting electrical or electrokinetic energy from the receiver, and at least one of a transmission cavity between the transmitter and the receiver, a full-wave rectifier in the co-resonant output circuit, and an oscillatory pulsed glow discharge device incorporated in the co-resonant output circuit.

2. A device according to claim **1**, wherein the output circuit comprises a full wave rectifier presenting a capacitance to the receiver.

3. A device according to claim **2**, wherein the output circuit comprises an electric motor presenting inductance to the receiver.

4. A device according to claim **3**, wherein the motor is a split phase motor.

5. A device according to claim **4**, wherein the motor is a drag motor having a non-magnetic conductive rotor.

6. A device according to claim **5**, wherein the motor has inertial damping.

7. A device according to claim **6**, wherein the motor has a shaft, a drag cup rotor on the shaft, and inertial damping is provided by a further drag cup on the shaft.

8. A device according to claim **6**, wherein the transmitter and receiver each comprise at least one of a Tesla coil and an autogenous pulsed abnormal glow discharge device.

9. A device according to claim **8**, wherein the transmitter and receiver both comprise Tesla coils, and further including a transmission cavity which comprises spaced plates connected respectively to the distal poles of the secondaries of Tesla coils incorporated in the transmitter and receiver respectively.

10. A device according to claim **9**, wherein the plates are parallel.

11. A device according to claim **9**, wherein the plates are concentric.

12. A device according to claim **9**, wherein at least the receiver comprises a Tesla coil driving a plasma reactor operating In PAGD (pulsed abnormal glow discharge) mode.

13. A device according to claim **12**, wherein the transmitter and receiver both comprise Tesla coils, and further Including a transmission cavity which comprises spaced plates connected respectively to the distal poles of the secondaries of Tesla coils incorporated in the transmitter and receiver respectively.

14. A device according to claim **1**, wherein the transmitter and receiver each comprise at least one of a Tesla coil and an autogenous pulsed abnormal glow discharge device.

15. A device according to claim **1** wherein a transmitter/receiver cavity is present and filled with an aqueous liquid.

16. A device for the conversion of massfree radiation into electrical or electrokinetic energy comprising a receiver of such radiation from a source of massfree electrical radiation having a damped wave component, the receiver being tuned to resonance with the damped wave frequency of the source, a co-resonant output circuit coupled into and extracting electrical or electrokinetic energy from the receiver, and at least one of a transmission cavity between the source and the receiver, a full-wave rectifier in the co-resonant output circuit, and an oscillatory pulsed glow discharge device incorporated in the co-resonant output circuit.

* * * * *

US008686575B2

(12) **United States Patent**
McCowen

(10) **Patent No.: US 8,686,575 B2**
(45) **Date of Patent: Apr. 1, 2014**

(54) **ENERGY COLLECTION**

(75) Inventor: **Clint McCowen**, Navarre, FL (US)

(73) Assignee: **Ion Power Group, LLC**, Navarre, FL (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **12/255,130**

(22) Filed: **Oct. 21, 2008**

(65) **Prior Publication Data**

US 2009/0040680 A1 Feb. 12, 2009

Related U.S. Application Data

(63) Continuation of application No. 11/358,264, filed on Feb. 21, 2006, now Pat. No. 7,439,712.

(51) **Int. Cl.**

F02B 63/04	(2006.01)
F03G 7/08	(2006.01)
H02K 7/18	(2006.01)
H02N 1/00	(2006.01)

(52) **U.S. Cl.**
USPC .... **290/1 R**; 322/2 A

(58) **Field of Classification Search**
USPC .... 290/1 R; 322/2 A; 126/660; 60/641.8, 60/641.11, 641.12, 659, 698; 244/158.2
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427	A		5/1901	Palencsar	307/149
911,260	A		2/1909	Pennock	307/149
1,014,719	A		1/1912	Pennock	307/149
2,473,819	A	*	6/1949	Pittman	315/36
3,532,959	A	*	10/1970	Erickson	320/166
3,780,722	A	*	12/1973	Swet	126/680
3,946,227	A		3/1976	Bingham	250/281
4,079,225	A	*	3/1978	Warner	219/110
4,092,250	A		5/1978	Sano et al.	210/500.29
4,104,696	A		8/1978	Cochran, Jr.	361/229
4,146,800	A	*	3/1979	Gregory et al.	290/44
4,180,056	A	*	12/1979	Schnabel et al.	126/660
4,201,197	A	*	5/1980	Dismer	126/600
4,206,396	A	*	6/1980	Marks	322/2 A
4,224,496	A	*	9/1980	Riordan et al.	219/110
4,307,936	A	*	12/1981	Ochiai	385/120
4,314,192	A	*	2/1982	Caro	322/2 R
4,346,478	A	*	8/1982	Sichling	398/107
4,425,905	A	*	1/1984	Mori	126/578
4,433,248	A	*	2/1984	Marks	290/44
RE31,678	E	*	9/1984	Ochiai	385/115

(Continued)

FOREIGN PATENT DOCUMENTS

WO	2007098341	8/2007

OTHER PUBLICATIONS

"Emissivity Values for Common Materials & Non-Metals", Raytek—A Fluke Company, http://www.raytek.com/Raytek/en-r0/IREducation/EmissivityNonMetals.htm, 1999.*

Primary Examiner — Pedro J Cuevas

(74) *Attorney, Agent, or Firm* — Benjamin A. Balser; Balser & Grell IP Law

(57) **ABSTRACT**

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support wire system supported by poles. The support wire system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but carbon and graphite are preferred. Diodes may be used to restrict the backflow or loss of energy.

25 Claims, 10 Drawing Sheets

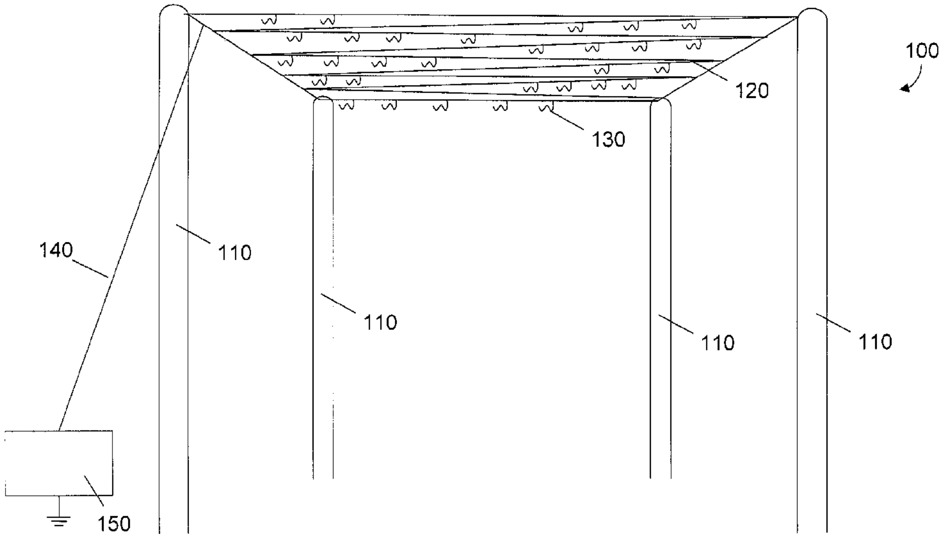

(56) References Cited

U.S. PATENT DOCUMENTS

4,483,311	A *	11/1984	Whitaker	126/602
4,489,269	A *	12/1984	Edling et al.	322/2 R
4,512,335	A *	4/1985	Mori	126/678
4,653,223	A *	3/1987	Mori	47/17
4,653,472	A *	3/1987	Mori	126/683
4,676,226	A *	6/1987	Mori	126/680
4,809,675	A *	3/1989	Mori	126/578
4,852,454	A *	8/1989	Batchelder	89/1.11
4,943,125	A *	7/1990	Laundre' et al.	385/35
4,988,159	A *	1/1991	Turner et al.	385/33
5,047,892	A	9/1991	Sakata et al.	361/231
5,114,101	A *	5/1992	Stern et al.	244/172.8
5,145,257	A *	9/1992	Bryant et al.	374/131
5,379,103	A *	1/1995	Zigler	356/73
5,851,309	A *	12/1998	Kousa	136/248
5,942,806	A *	8/1999	Veliadis	290/1 R
6,025,591	A	2/2000	Taylor et al.	250/292
6,038,363	A *	3/2000	Slater et al.	385/147
6,116,544	A *	9/2000	Forward et al.	244/158.2
6,173,922	B1 *	1/2001	Hoyt et al.	244/158.2
6,226,440	B1 *	5/2001	Lyons	385/119
6,419,191	B1 *	7/2002	Hoyt et al.	244/158.2
6,425,391	B1 *	7/2002	Davoren et al.	126/683
6,735,830	B1	5/2004	Merciel	250/324
6,765,212	B2 *	7/2004	Goetz et al.	250/339.11
6,846,447	B2 *	1/2005	Mbachu et al.	264/406
6,853,447	B2 *	2/2005	Goetz	356/237.1
6,894,772	B2 *	5/2005	Goetz et al.	356/237.1
6,925,852	B2 *	8/2005	Susko	73/23.2
6,974,110	B2	12/2005	Grandics	244/171.5
7,109,597	B1 *	9/2006	Bose	290/43
7,128,867	B2 *	10/2006	Mbachu et al.	264/406
7,231,809	B2 *	6/2007	Susko	73/31.05
7,259,475	B2 *	8/2007	Hong et al.	307/64
7,439,712	B2 *	10/2008	McCowen	322/2 A
7,449,668	B2 *	11/2008	Schutten et al.	250/214 SW
7,478,712	B2 *	1/2009	McCowen	191/45 R
7,532,819	B1 *	5/2009	Triebes et al.	398/121
2002/0026933	A1 *	3/2002	Gottlieb	126/628
2002/0108892	A1 *	8/2002	Goetz et al.	209/576
2002/0109094	A1 *	8/2002	Goetz et al.	250/339.11
2002/0109835	A1 *	8/2002	Goetz	356/237.1
2003/0107869	A1	6/2003	Pan et al.	361/306.2
2003/0125630	A1 *	7/2003	Furnish	600/461
2003/0193319	A1 *	10/2003	Wood et al.	322/2 A
2004/0011925	A1 *	1/2004	Grandics	244/62
2004/0083793	A1 *	5/2004	Susko	73/31.05
2004/0094853	A1 *	5/2004	Mbachu et al.	264/40.1
2004/0160711	A1	8/2004	Stumberger	361/20
2005/0101024	A1 *	5/2005	Mbachu et al.	436/85
2005/0136311	A1 *	6/2005	Ueda et al.	429/30
2005/0270525	A1 *	12/2005	Susko	356/326
2007/0107765	A1 *	5/2007	Schutten et al.	136/243
2007/0195481	A1 *	8/2007	McCowen	361/233
2007/0218323	A1 *	9/2007	Sudo et al.	429/13
2007/0223171	A1	9/2007	Guy et al.	361/126
2007/0273206	A1 *	11/2007	McCowen	307/9.1
2007/0297892	A1 *	12/2007	Kildegaard	415/1
2008/0145714	A1 *	6/2008	Kagami	429/13
2009/0107842	A1	4/2009	Park et al.	204/554
2009/0114495	A1 *	5/2009	McCowen	191/45 R
2009/0216292	A1 *	8/2009	Pless et al.	607/33
2010/0090562	A1 *	4/2010	McCowen	310/309
2010/0090563	A1 *	4/2010	McCowen	310/309
2012/0164555	A1 *	6/2012	Yamazaki et al.	429/482
2012/0299559	A1 *	11/2012	McCowen	322/2 A
2013/0276776	A1 *	10/2013	Te Kamp et al.	126/660

* cited by examiner

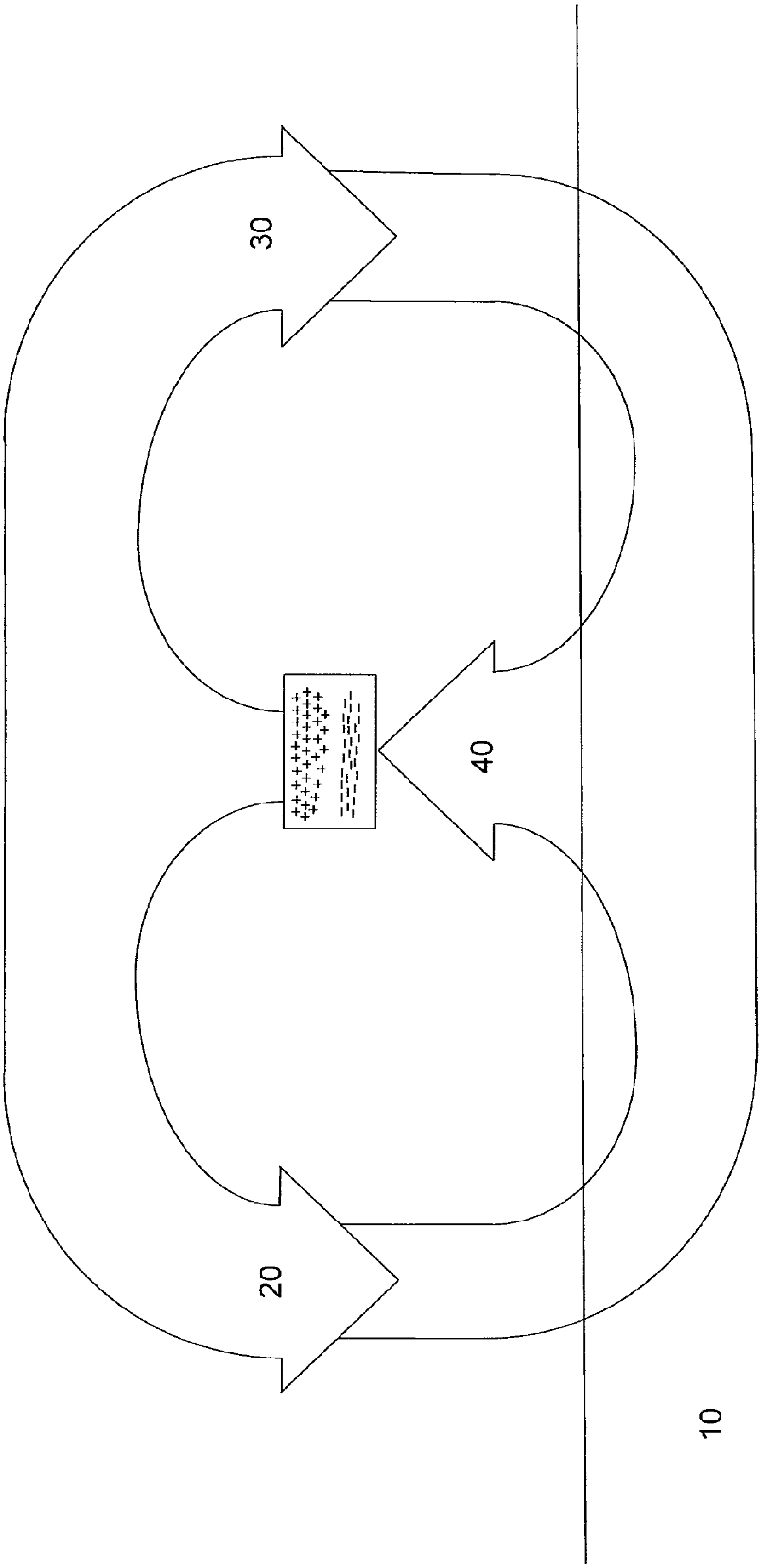

FIGURE 1

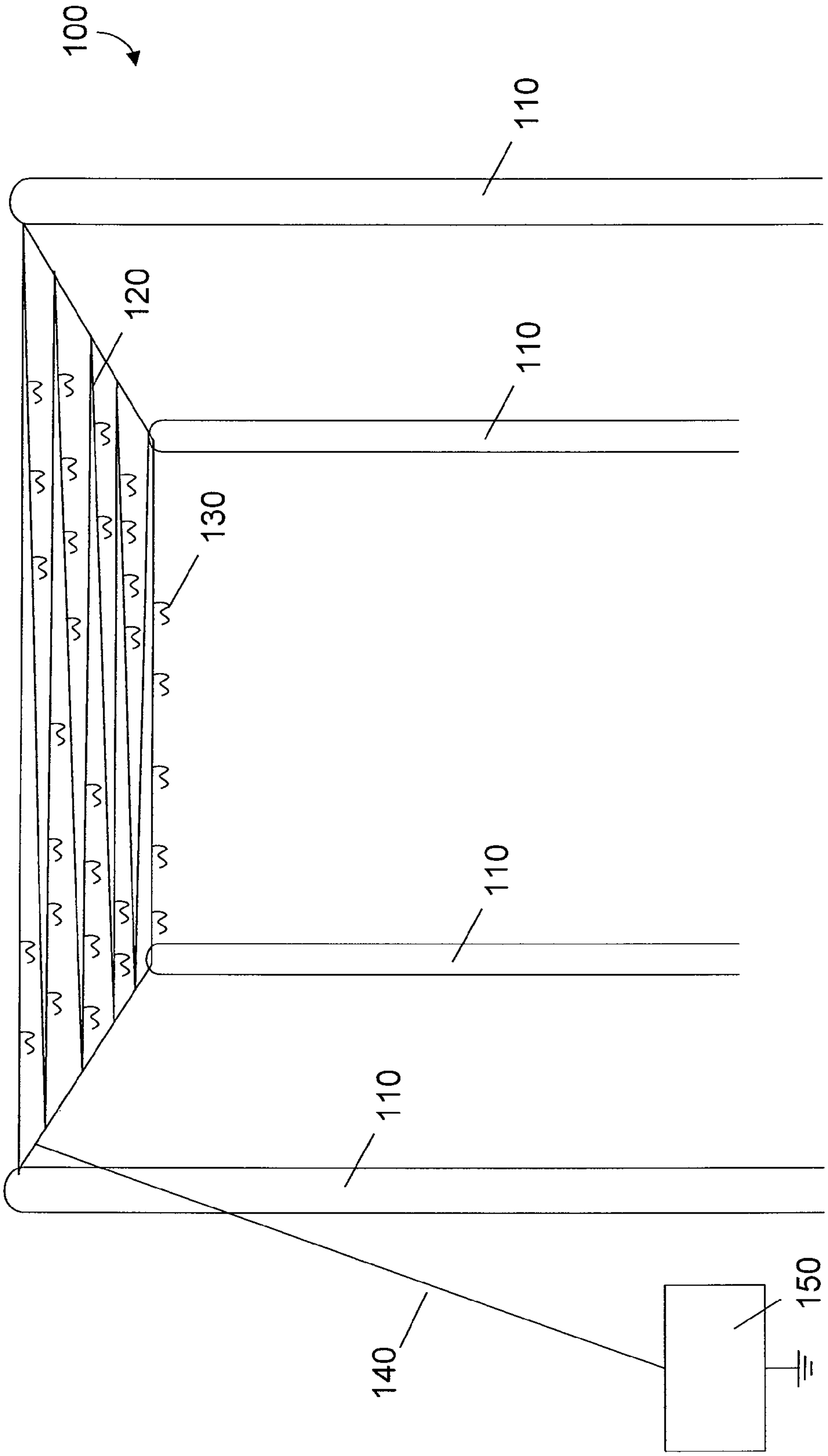
100
110
110
120
130
110
110
140
150

FIGURE 2

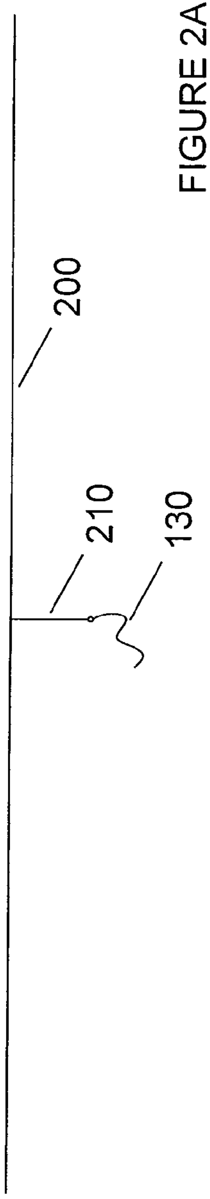

FIGURE 2A

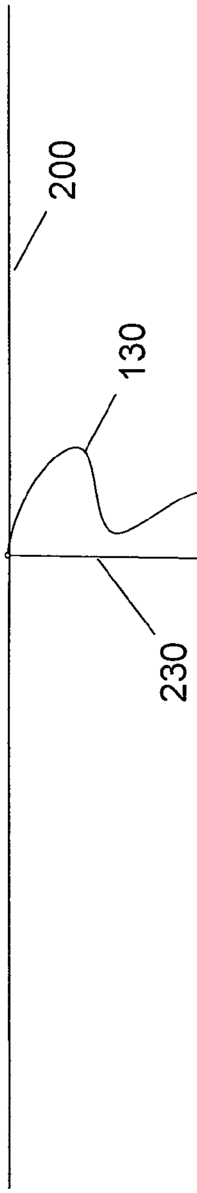

FIGURE 2B

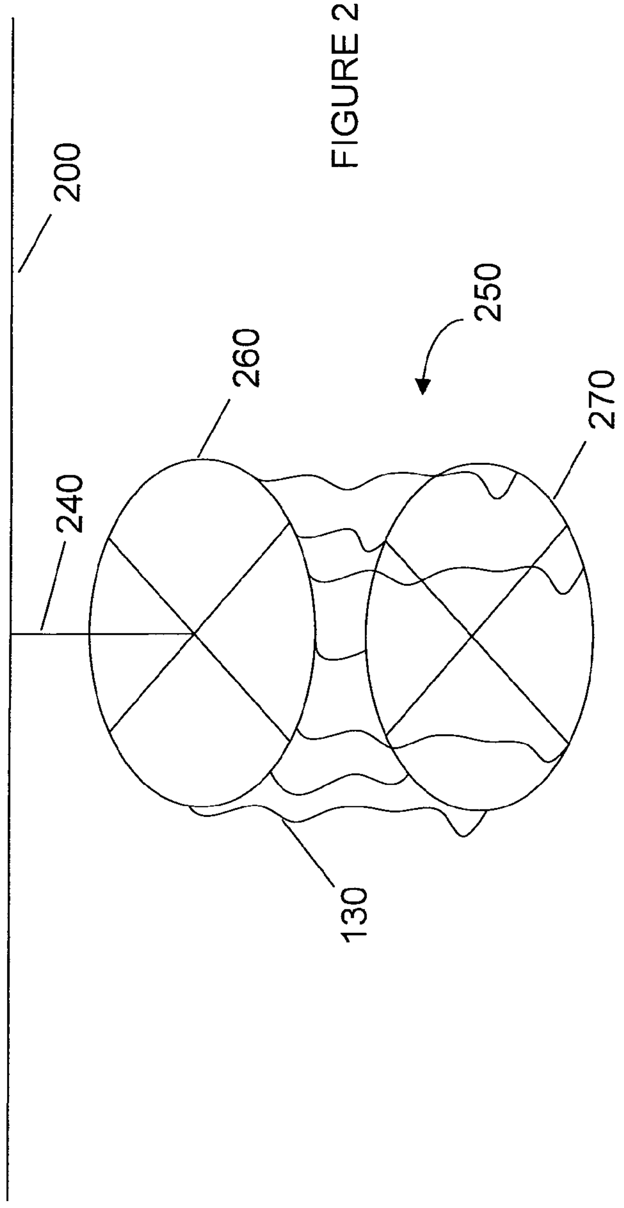
200
250
260
270
240
130

FIGURE 2C

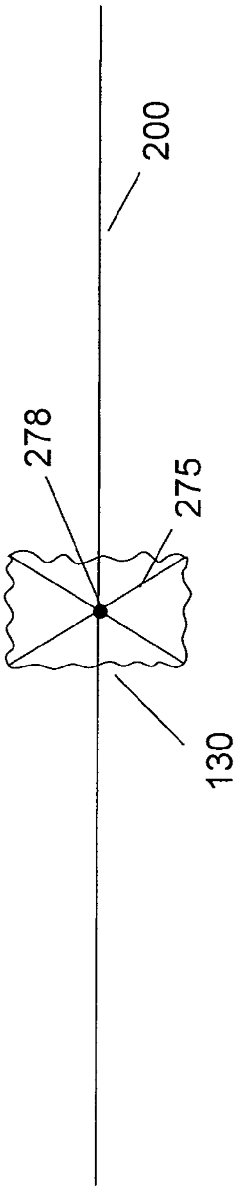
200
278
275
130

FIGURE 2D

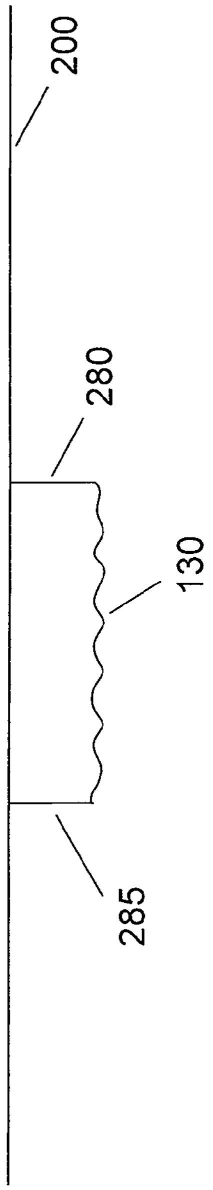
200
280
130
285

FIGURE 2E

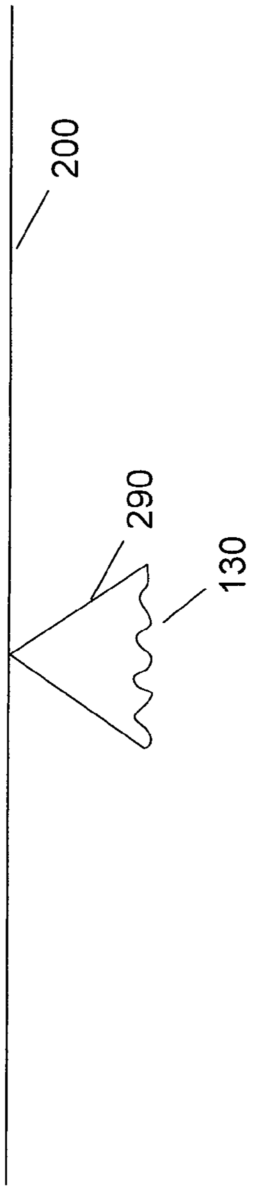

FIGURE 2F

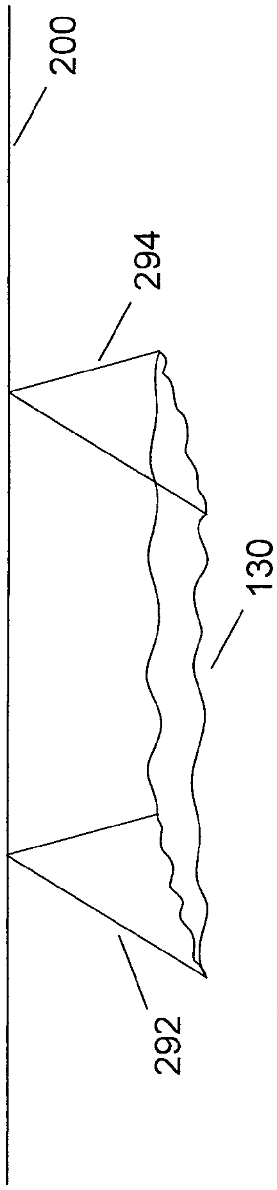
200
294
130
292

FIGURE 2G

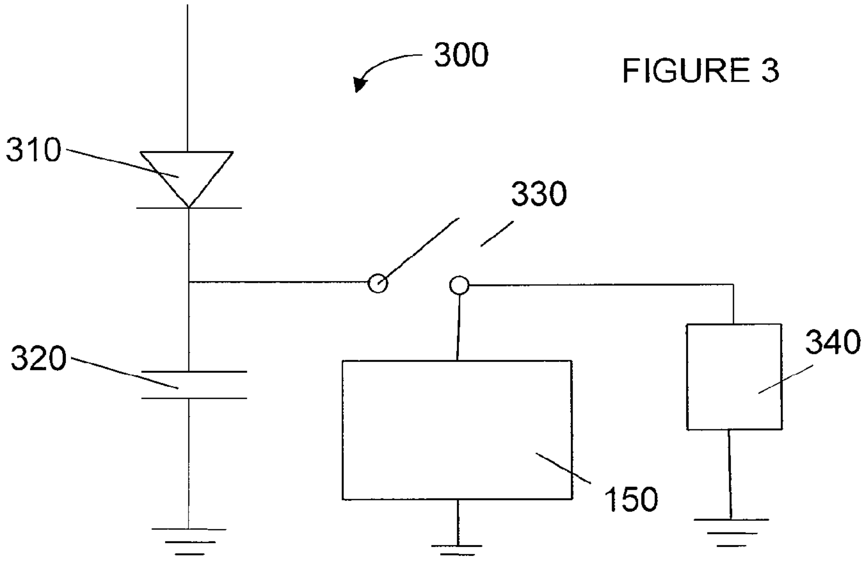
300
310
330
320
340
150

FIGURE 3

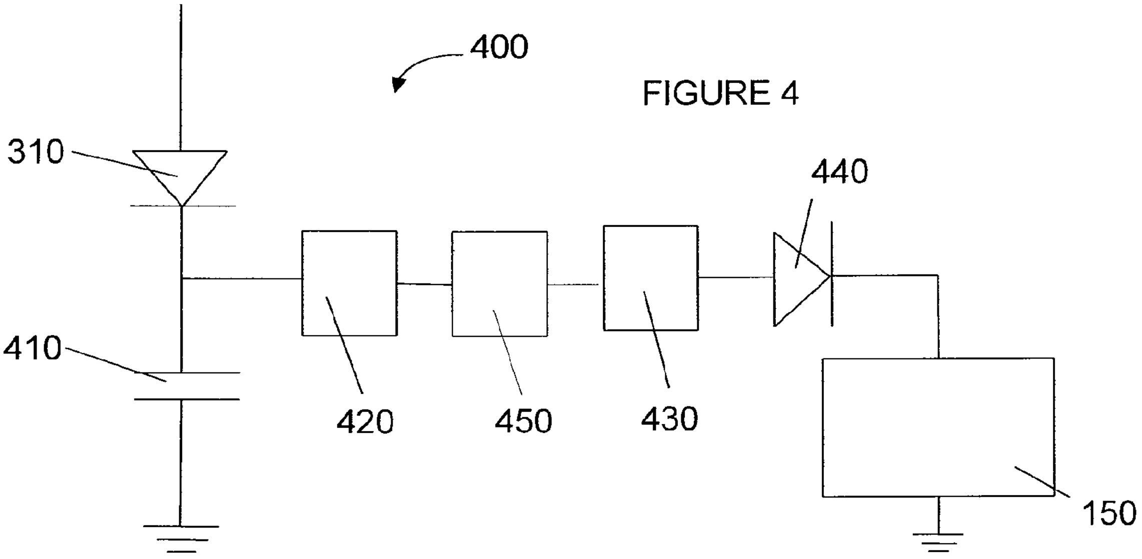
400
310
440
410
420
450
430
150

FIGURE 4

FIGURE 5

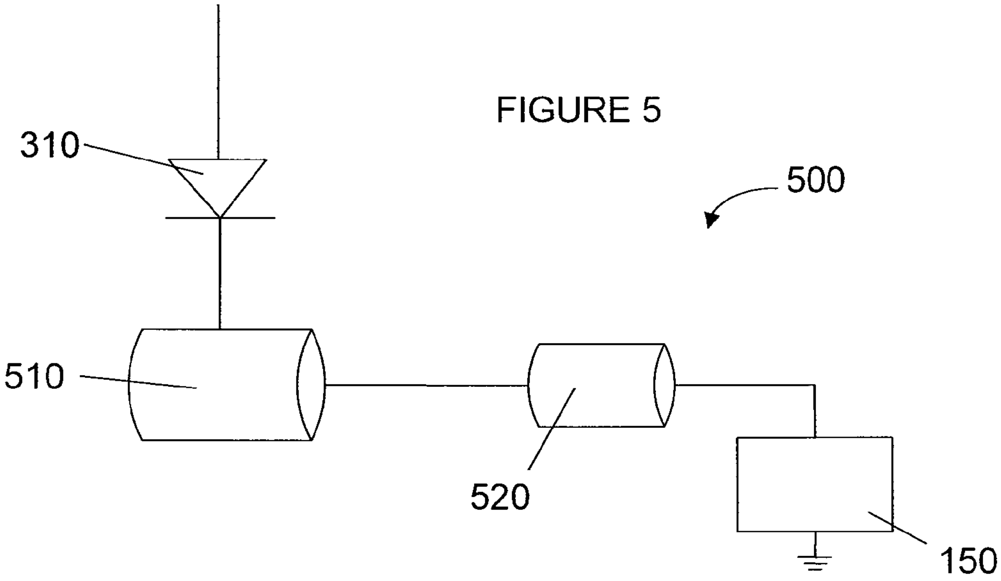
310
500
510
520
150

FIGURE 6

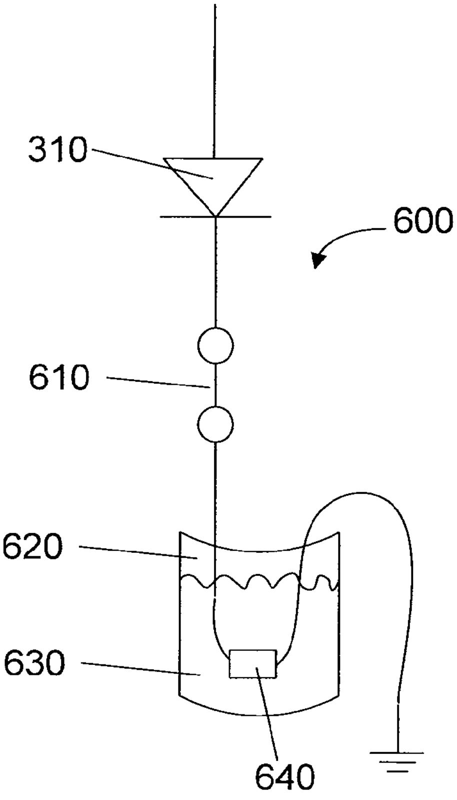
310
600
610
620
630
640

FIGURE 7

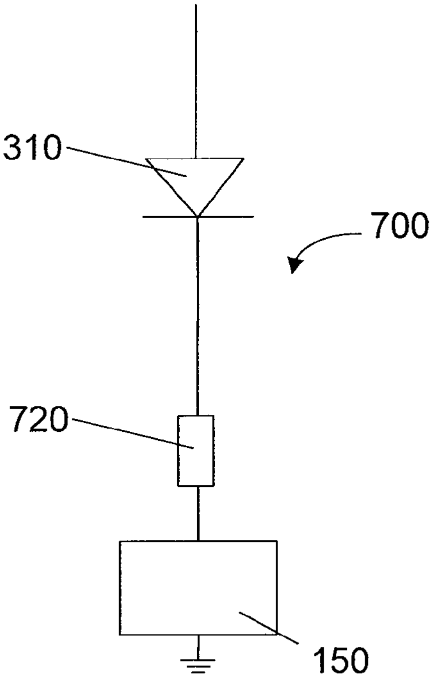
310
700
720
150

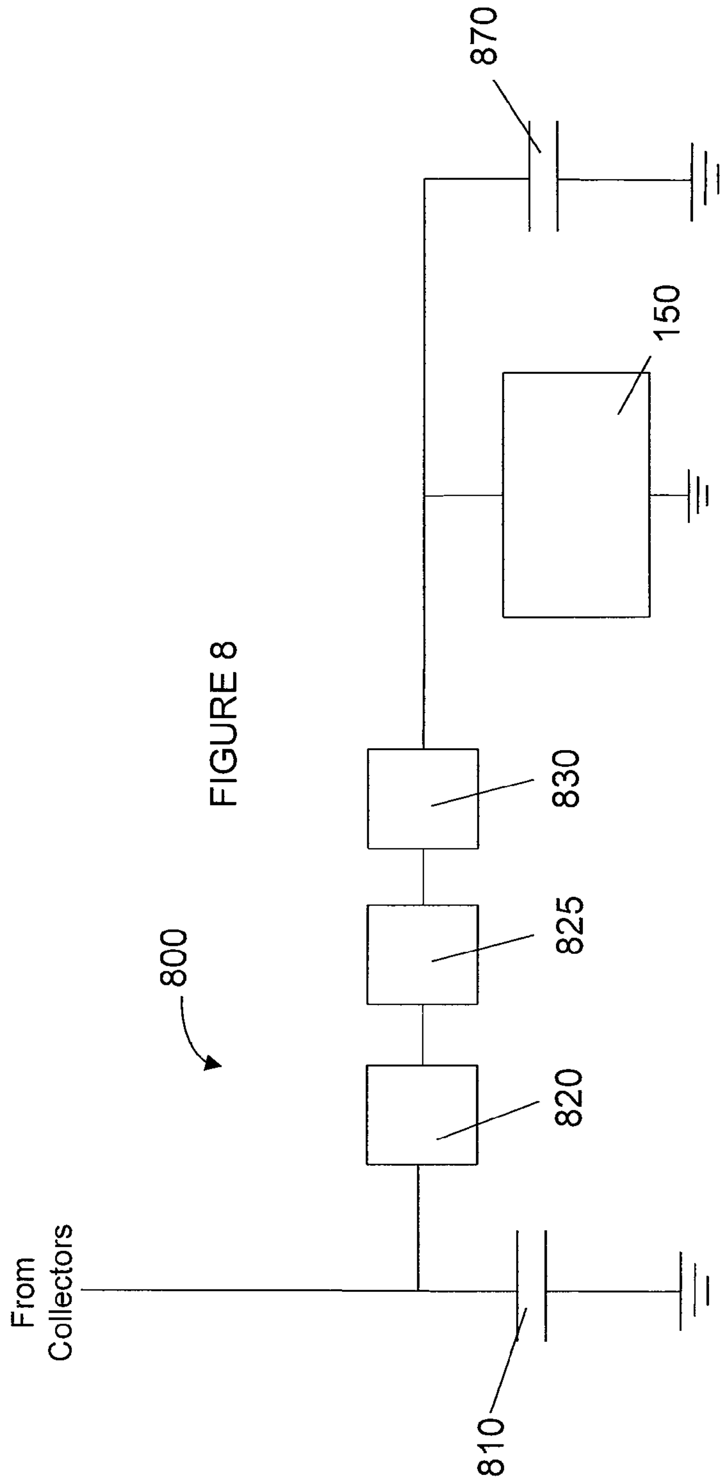
From Collectors
810
820
825
830
800
150
870

FIGURE 8

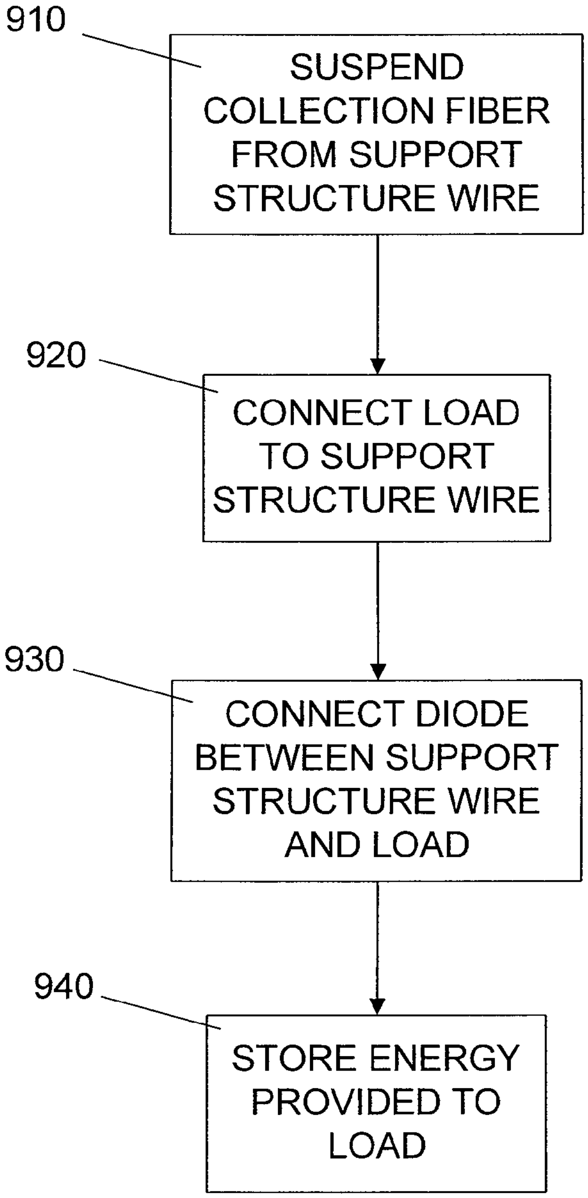
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

ENERGY COLLECTION

CROSS REFERENCE TO RELATED APPLICATION

This application is a continuation application of U.S. patent application Ser. No. 11/358,264, filed on Feb. 21, 2006, which is incorporated by reference herein.

TECHNICAL FIELD

The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure wire elevated above a ground level, at least one collection fiber electrically connected to the support structure wire; a load electrically connected to the support structure wire; and a diode electrically connected between the load and at least one collection fiber.

Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection fiber from a support structure wire elevated above ground level, the fiber electrically connected to the support structure wire; providing a load with an electrical connection to the support structure wire to draw current; and providing a diode electrically connected between the collection fiber and the load.

Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. **1** is a circuit diagram of a weather energy circuit.

FIG. **2** is a perspective view of an exemplary embodiment of many energy collectors elevated above ground by a structure.

FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

FIG. **2**B is a side view of an exemplary embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

FIG. **2**D is a side view of an exemplary embodiment of a support structure for multiple energy collection fibers.

FIG. **2**E is a side view of a support structure for an energy collection fiber.

FIG. **2**F is a side view of an exemplary embodiment of a support structure for an energy collection fiber.

FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

FIG. **3** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **4** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **5** is a circuit diagram of an exemplary embodiment of an energy collection circuit for powering a generator and motor.

FIG. **6** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

FIG. **7** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

FIG. **8** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy.

FIG. **9** is a flow diagram of an exemplary embodiment of collecting energy with a collection fiber.

DETAILED DESCRIPTION

Electric charges on conductors reside entirely on the external surface of the conductors, and tend to concentrate more

around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An exemplary embodiment of this disclosure takes advantage of this property, among others, to collect and use the energy generated by an electric field in the atmosphere. Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An exemplary embodiment utilizes carbon or graphite fibers for static electricity collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including aluminum or steel, but most notably, copper. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

An exemplary embodiment of the collection fibers as collection device **130** includes graphite or carbon fibers. Graphite and carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, or graphite fibers are examples that demonstrate exemplary collection ability.

In at least one exemplary embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than 100 volts per meter under some conditions. When support wire **120** is suspended in the air at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An exemplary embodiment of the collection device comprises the diode with no collection fiber. A preferred embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restricts energy that may be collected through one fiber from escaping through another fiber.

Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2A-2G** using a collection fiber embodiment. FIG. **2A** presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**.

Likewise, FIG. **2B** shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

FIG. **2C** presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

FIG. **2D** presents another exemplary embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2E** provides another exemplary embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

FIG. **2F** presents another exemplary embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2G** provides two supports as provided in FIG. **2F** such that at least two support members **292**, **294** may be connected

to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and between multiple support structures. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**.

FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load. Capacitor **320** may be connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

In an exemplary embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel. Non-limiting examples of spark gap **420** include mercury-reed switches and mercury-wetted reed switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground. The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes may be placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches and mercury-wetted reed switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel.

FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit. Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

FIG. **8** presents exemplary circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches and mercury-wetted reed switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

It should be emphasized that the above-described embodiments of the present disclosure, particularly, any "preferred" embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A method of collecting energy comprising:

suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to the support structure;

providing a load with an electrical connection to the at least one collection device to passively draw current; and

powering a fuel cell between the support structure and the load, the powering performed with energy collected from the at least one collection device.

2. The method of claim **1**, wherein the collection device comprises a diode.

3. The method of claim **1**, wherein the collection device comprises a collection fiber.

4. The method of claim **1**, wherein the collection device comprises a diode and a collection fiber and the diode is electrically connected between the collection fiber and the load.

5. The method of claim **1**, further comprising storing energy provided to the load.

6. The method of claim **5**, wherein storing energy provided to the load comprises storing energy in a capacitor or an inductor.

7. The method of claim **3**, wherein the collection fiber comprises carbon fiber or graphite fiber.

8. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the collection device electrically connected to and suspended from the support structure, the at least one collection device configured to passively draw current;

a load electrically connected to the at least one collection device; and

a diode electrically connected between the at least one collection device and the support structure.

9. The system of claim **8**, wherein the collection device comprises a diode.

10. The system of claim **8**, wherein the collection device comprises a collection fiber.

11. The system of claim **8**, wherein the collection device comprises a collection fiber and a diode electrically connected between the load and the collection fiber.

12. The system of claim **9**, wherein the diode is elevated relative to the ground level.

13. The system of claim **10**, wherein the collection fiber comprises a carbon fiber or a graphite fiber.

14. The system of claim **8**, further comprising:

a switch connected in series between the at least one collection device and the load; and

a capacitor connected in parallel with the switch and the load.

15. The system of claim **14**, wherein the switch comprises an interrupter connected between the at least one collection device and the load, and wherein the interrupter comprises at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

16. The system of claim **15**, further comprising a transformer connected between the interrupter and the load.

17. The system of claim **8**, further comprising:

a motor for providing power, the motor connected between the at least one collection device and the load; and

a generator powered by the motor.

18. The system of claim **8**, further comprising a fuel cell between the support structure and the load.

19. The system of claim **18**, wherein the fuel cell produces hydrogen and oxygen.

20. A system of collecting energy comprising:

means for suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to the means for suspending;

means for passively inducing current flow, the means for inducing current flow electrically connected to the means for suspending; and

means for restricting the backflow of charge carriers, the means for restricting the backflow of charge carriers electrically connected between the at least one collection device and the means for inducing current flow.

21. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the collection device electrically connected to and suspended from the support structure, the at least one collection device configured to passively draw current;

a load electrically connected to the at least one collection device; and

a switch connected in series between the at least one collection device and the load, the switch comprising an interrupter connected between the at least one collection device and the load, and wherein the interrupter comprises at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

22. The system of claim **21**, further comprising a transformer connected between the interrupter and the load.

23. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to and suspended from the support structure, the collection device configured to passively draw current;

a load electrically connected to the at least one collection device;

a motor for providing power, the motor connected between the at least one collection device and the load; and

a generator powered by the motor.

24. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to and suspended from the support structure the at least one collection device configured to passively draw current;

a load electrically connected to the at least one collection device; and

a fuel cell between the support structure and the load.

25. The system of claim **24**, wherein the fuel cell produces hydrogen and oxygen.

* * * * *

US007439712B2

(12) **United States Patent**
McCowen

(10) **Patent No.: US 7,439,712 B2**
(45) **Date of Patent: Oct. 21, 2008**

(54) ENERGY COLLECTION

(76) Inventor: **Clint McCowen**, 1902 Rue La Fontaine, Navarre, FL (US) 32566

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 509 days.

(21) Appl. No.: **11/358,264**

(22) Filed: **Feb. 21, 2006**

(65) **Prior Publication Data**

US 2007/0195481 A1 Aug. 23, 2007

(51) **Int. Cl.**
H02N 1/00 (2006.01)

(52) **U.S. Cl.** .... **322/2 A**; 307/149; 361/225; 361/117; 174/2; 320/101

(58) **Field of Classification Search** .... 361/220, 361/117, 212, 233, 221, 222, 225; 322/2 A, 322/2 R; 307/149; 174/2, 3; 320/101
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427	A *	5/1901	Palencsar	307/149
911,260	A *	2/1909	Pennock	307/149
1,014,719	A *	1/1912	Pennock	307/149
4,104,696	A	8/1978	Cochran, Jr.	
5,047,892	A	9/1991	Sakata et al.	
6,735,830	B1	5/2004	Merciel	
6,974,110	B2	12/2005	Grandics	
2003/0107869	A1	6/2003	Pan et al.	
2004/0160711	A1	8/2004	Stumberger	
2004/0212945	A1	10/2004	Sprenger et al.	

* cited by examiner

Primary Examiner—Mark T Le
(74) *Attorney, Agent, or Firm*—Thomas, Kayden, Horstemeyer & Risley, LLP

(57) ABSTRACT

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support wire system supported by poles. The support wire system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but carbon and graphite are preferred. Diodes may be used to restrict the backflow or loss of energy.

28 Claims, 10 Drawing Sheets

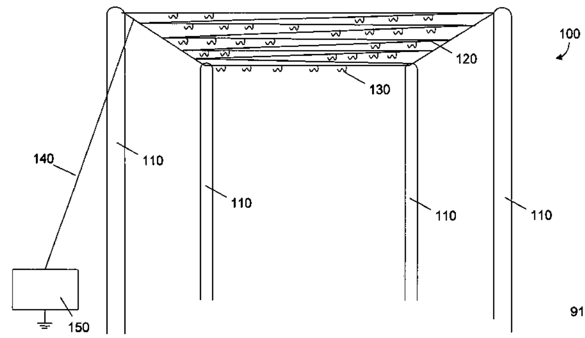

100
120
130
140
110
110
110
110
110
150

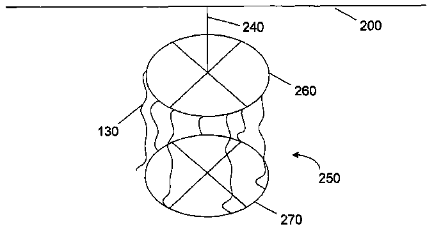

240
200
260
130
250
270

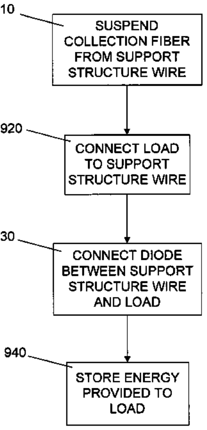

910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

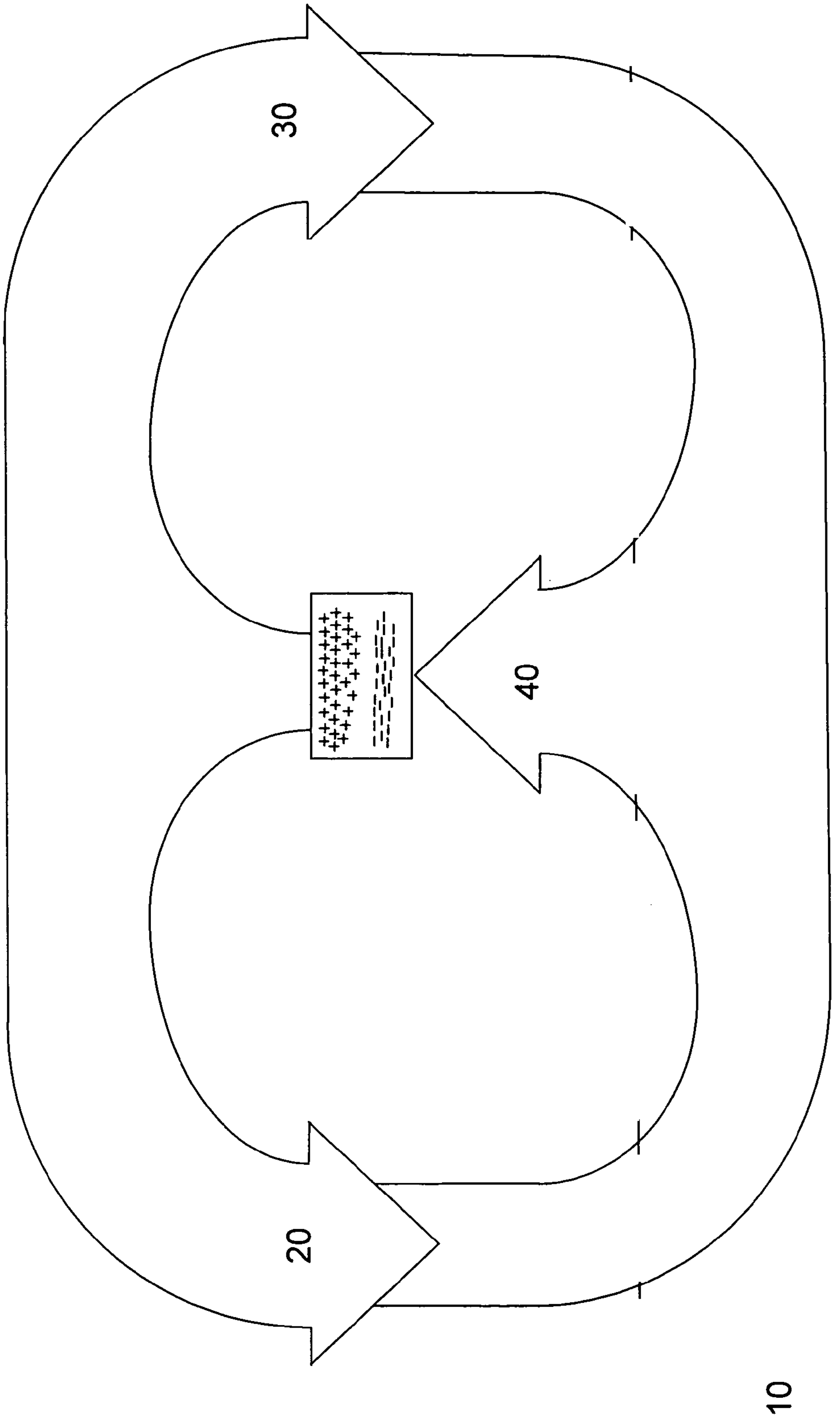

FIGURE 1

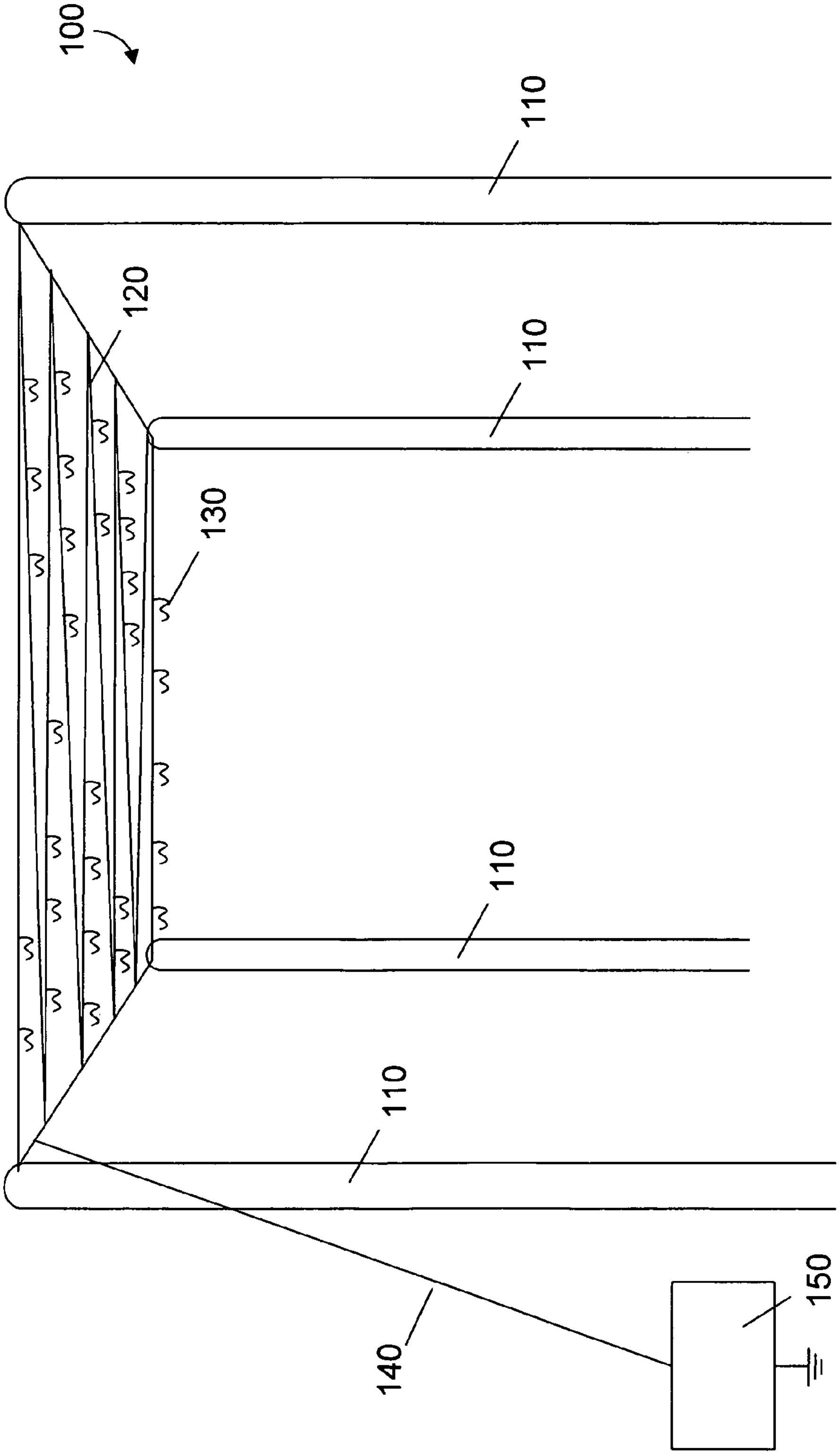
100
110
110
110
110
120
130
140
150

FIGURE 2

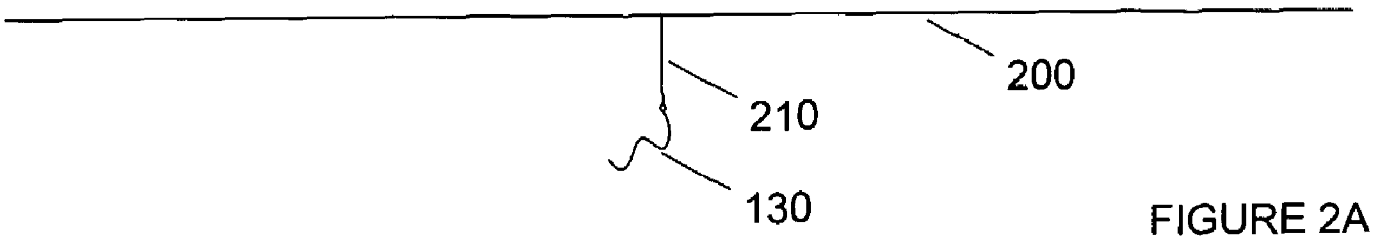

FIGURE 2A

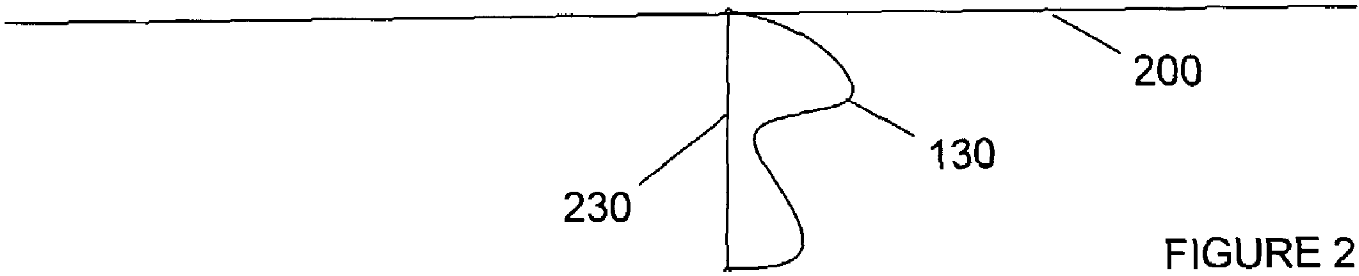

FIGURE 2B

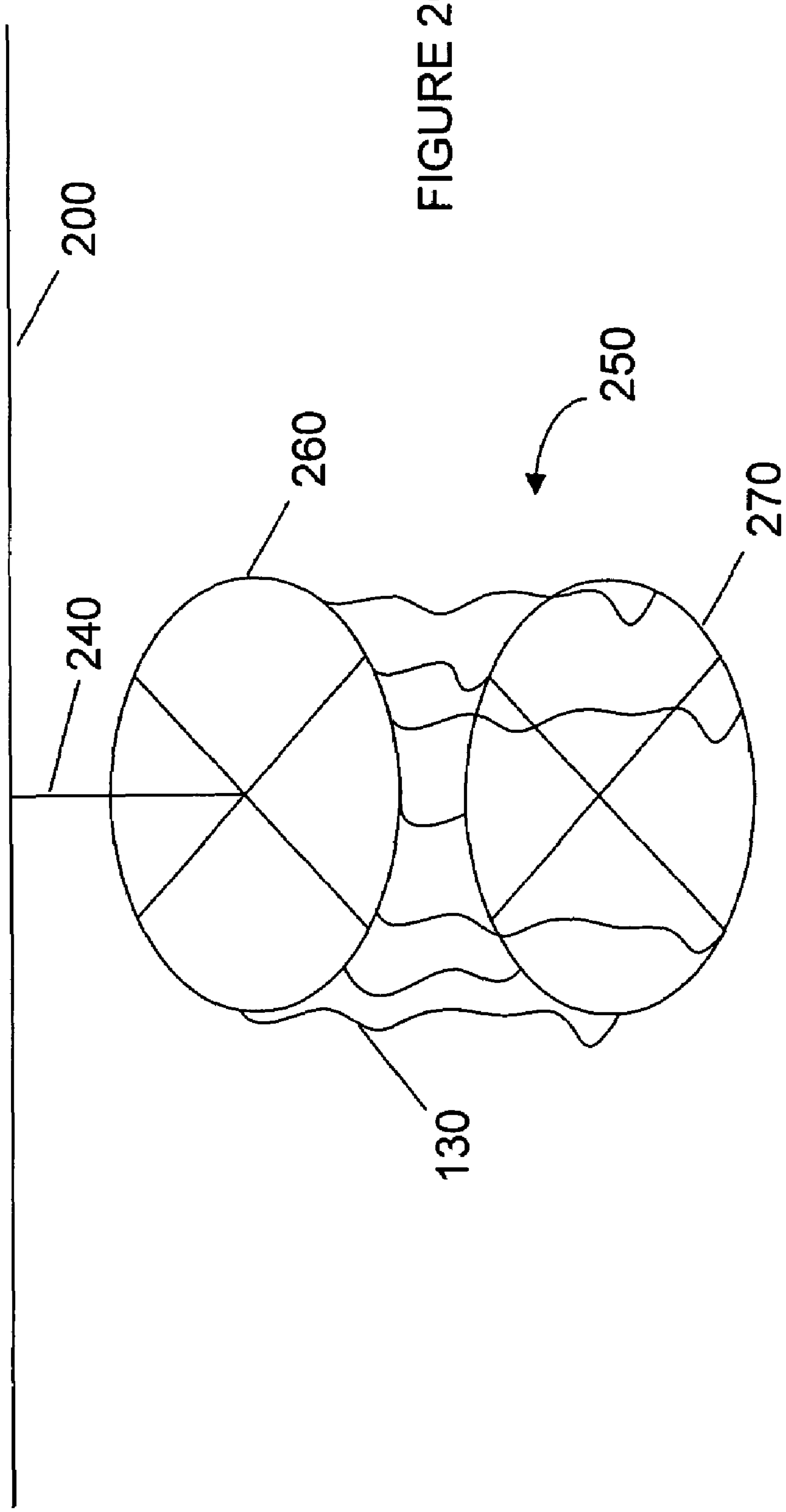
200
250
260
270
240
130

FIGURE 2C

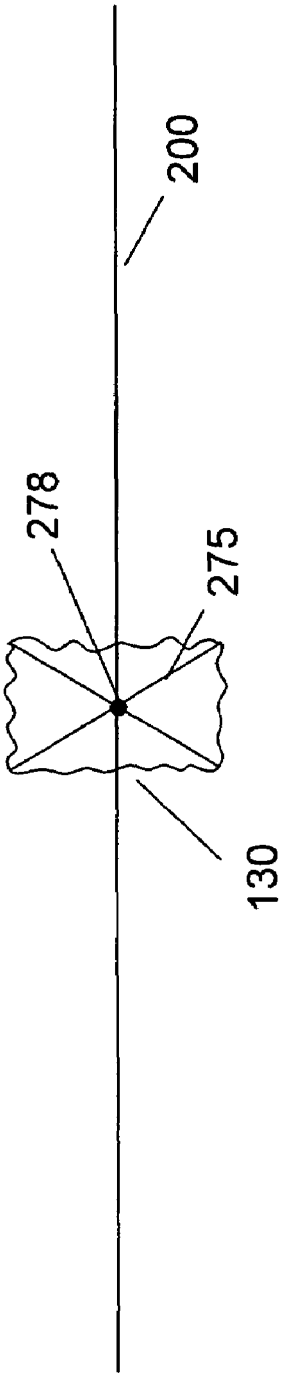
200
278
275
130

FIGURE 2D

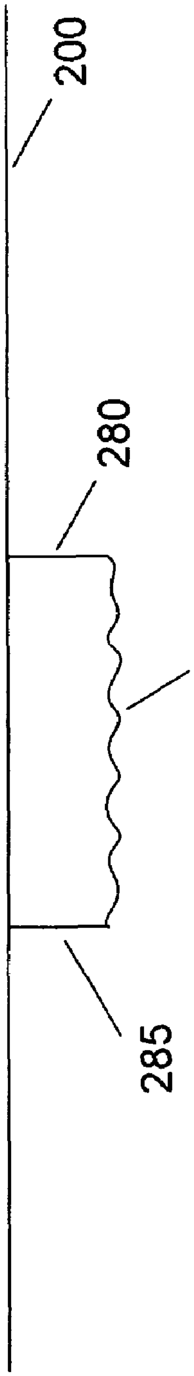
200
280
130
285

FIGURE 2E

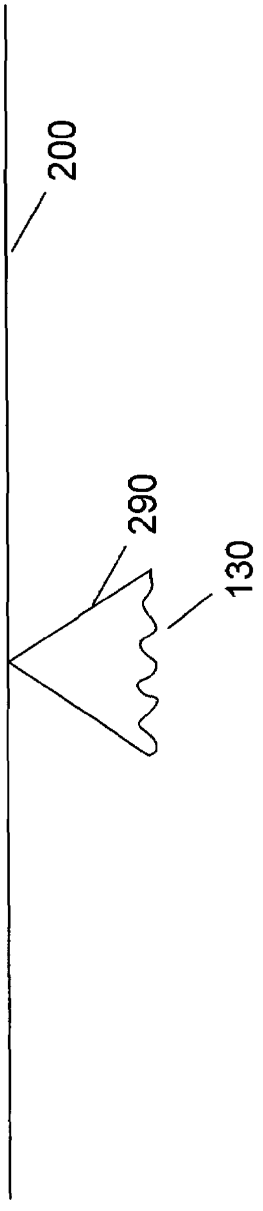

FIGURE 2F

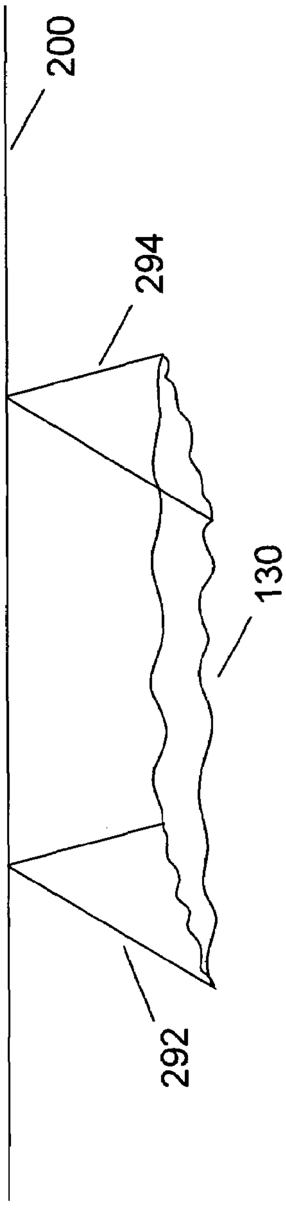
200
294
130
292

FIGURE 2G

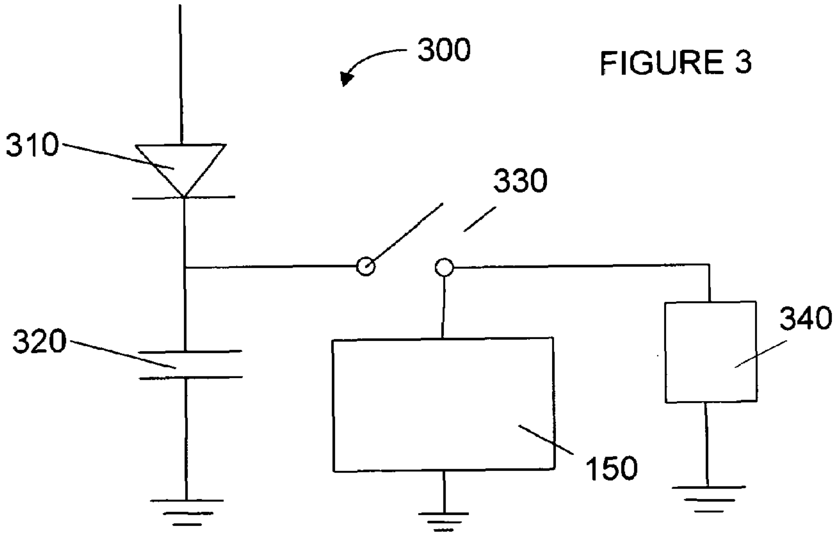
300
310
330
320
340
150

FIGURE 3

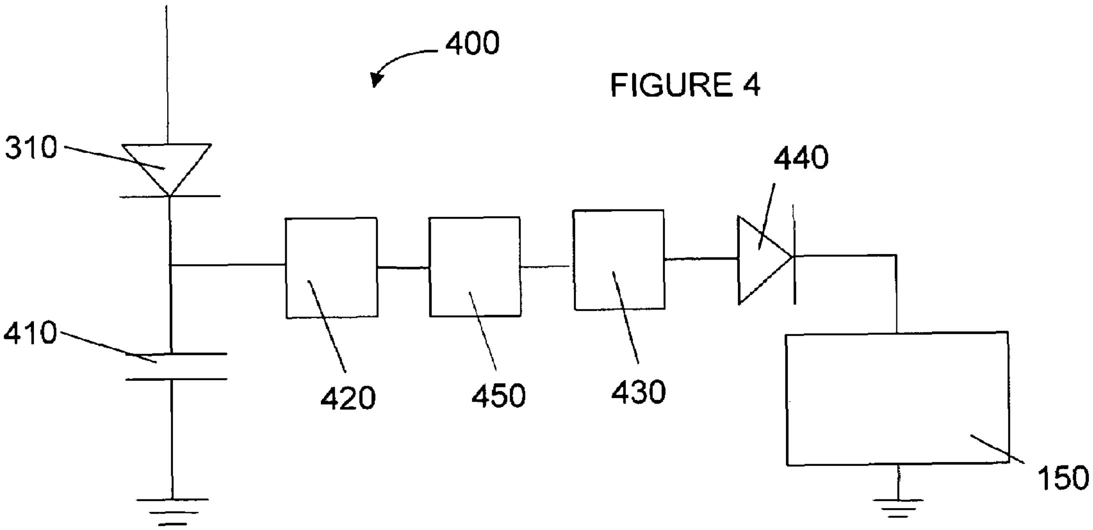
400
310
440
410
420
450
430
150

FIGURE 4

FIGURE 5

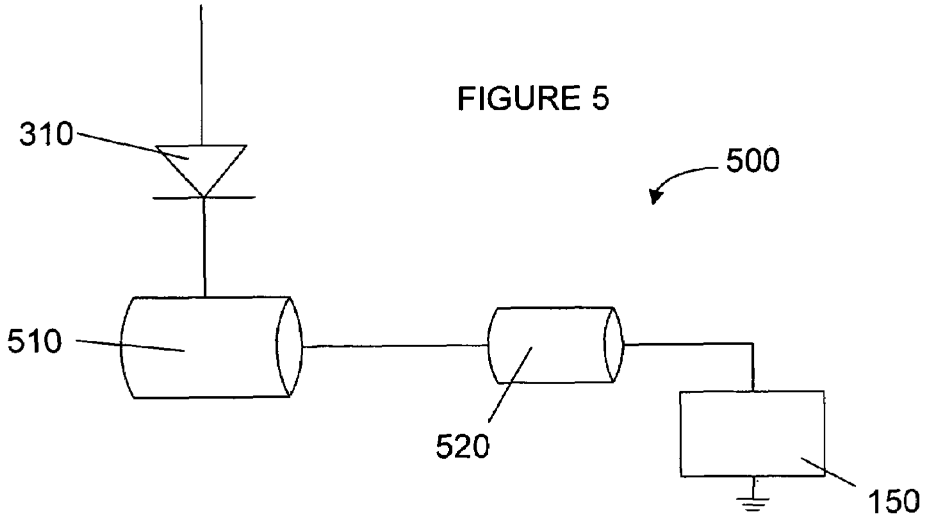
310
500
510
520
150

FIGURE 6

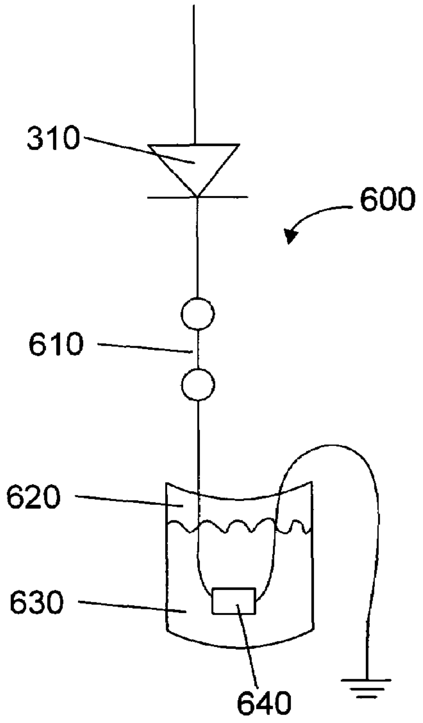
310
600
610
620
630
640

FIGURE 7

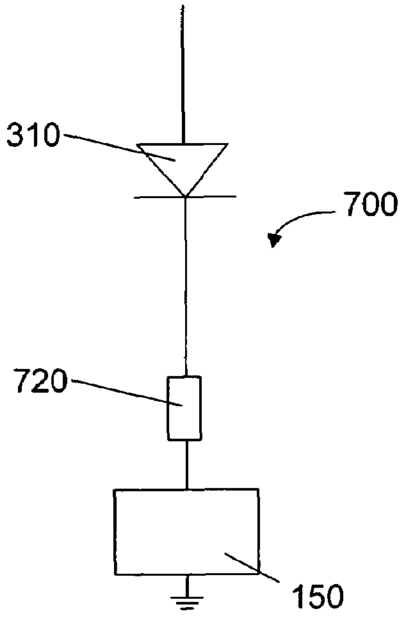
310
700
720
150

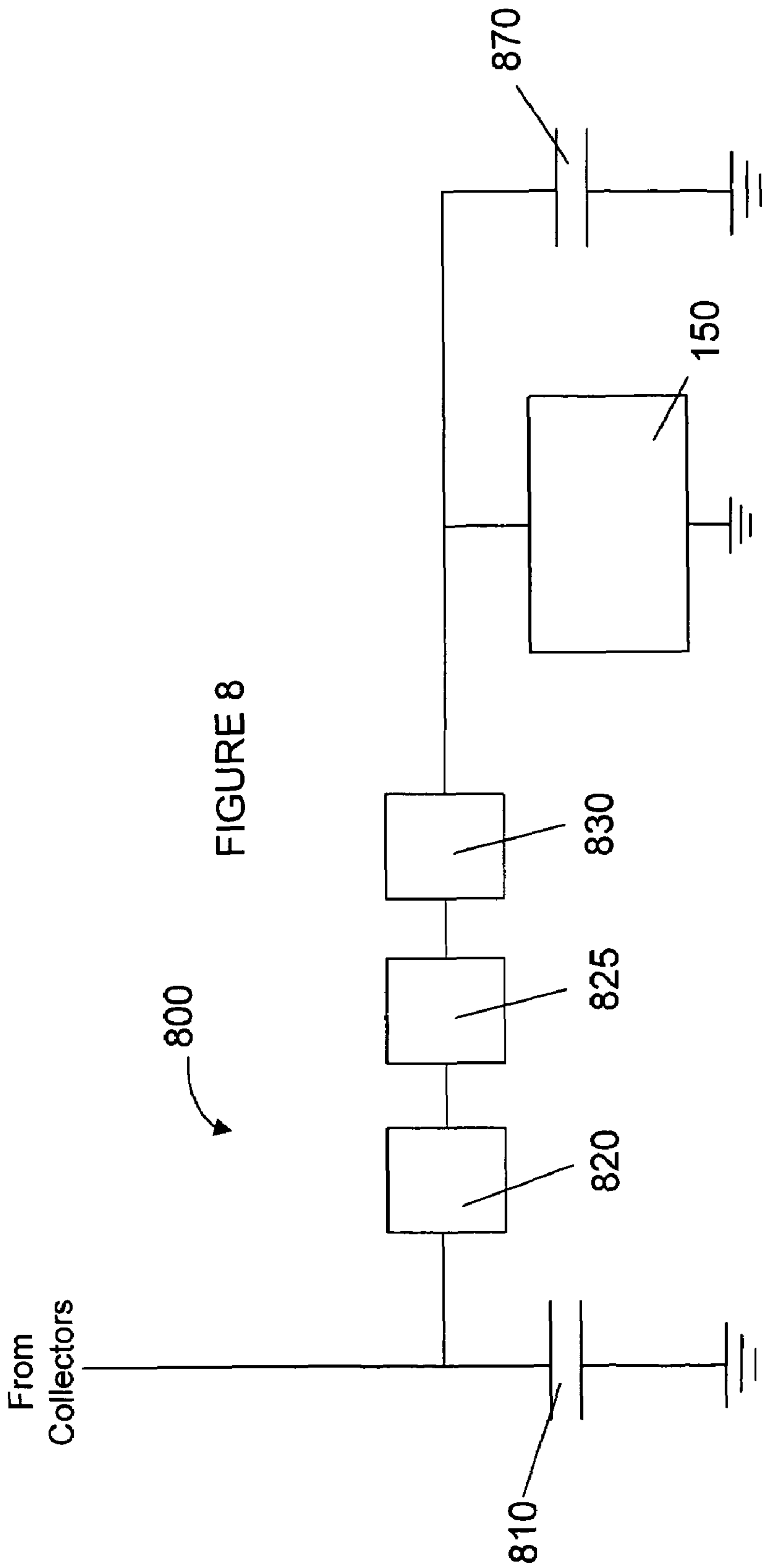
From Collectors
810
800
820
825
830
150
870

FIGURE 8

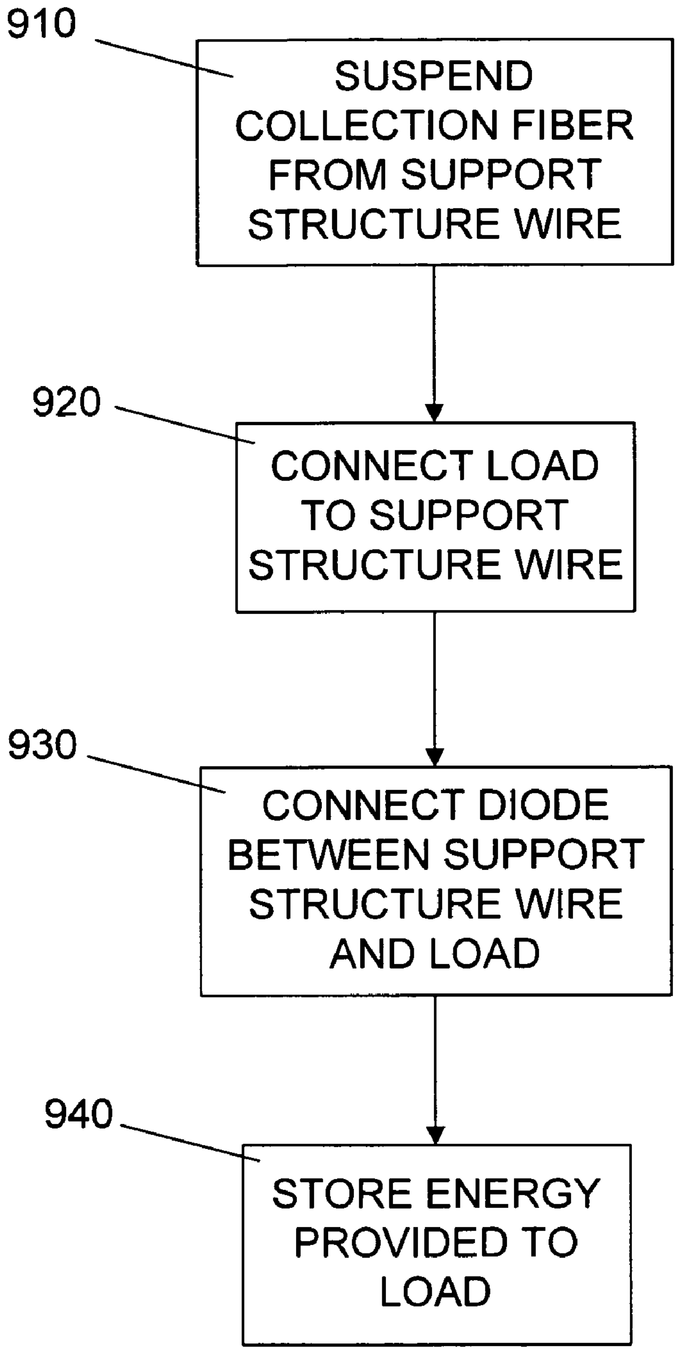
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

ENERGY COLLECTION

TECHNICAL FIELD

The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure wire elevated above a ground level, at least one collection fiber electrically connected to the support structure wire; a load electrically connected to the support structure wire; and a diode electrically connected between the load and at least one collection fiber.

Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection fiber from a support structure wire elevated above ground level, the fiber electrically connected to the support structure wire; providing a load with an electrical connection to the support structure wire to draw current; and providing a diode electrically connected between the collection fiber and the load.

Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. **1** is a circuit diagram of a weather energy circuit.

FIG. **2** is a perspective view of an exemplary embodiment of many energy collectors elevated above ground by a structure.

FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

FIG. **2**B is a side view of an exemplary embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

FIG. **2**D is a side view of an exemplary embodiment of a support structure for multiple energy collection fibers.

FIG. **2**E is a side view of a support structure for an energy collection fiber.

FIG. **2**F is a side view of an exemplary embodiment of a support structure for an energy collection fiber.

FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

FIG. **3** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **4** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **5** is a circuit diagram of an exemplary embodiment of an energy collection circuit for powering a generator and motor.

FIG. **6** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

FIG. **7** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

FIG. **8** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy.

FIG. **9** is a flow diagram of an exemplary embodiment of collecting energy with a collection fiber.

DETAILED DESCRIPTION

Electric charges on conductors reside entirely on the external surface of the conductors, and tend to concentrate more around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An exemplary embodiment of this disclosure takes advantage of this

property, among others, to collect and use the energy generated by an electric field in the atmosphere. Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An exemplary embodiment utilizes carbon or graphite fibers for static electricity collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including aluminum or steel, but most notably, copper. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

An exemplary embodiment of the collection fibers as collection device **130** includes graphite or carbon fibers. Graphite and carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, or graphite fibers are examples that demonstrate exemplary collection ability.

In at least one exemplary embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than **100** volts per meter under some conditions. When support wire **120** is suspended in the air at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An exemplary embodiment of the collection device comprises the diode with no collection fiber. A preferred embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restricts energy that may be collected through one fiber from escaping through another fiber.

Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2A-2G** using a collection fiber embodiment. FIG. **2A** presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**.

Likewise, FIG. **2B** shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

FIG. **2C** presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

FIG. **2D** presents another exemplary embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2E** provides another exemplary embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

FIG. **2F** presents another exemplary embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2G** provides two supports as provided in FIG. **2F** such that at least two support members **292**, **294** may be connected to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and between multiple support structures. A plurality of diodes may be

placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**.

FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load.

Capacitor **320** may be connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

In an exemplary embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel. Non-limiting examples of spark gap **420** include mercury-reed switches and mercury-wetted reed switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground.

The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes may be placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches and mercury-wetted reed switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel.

FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit. Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

FIG. **8** presents exemplary circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches and mercury-wetted reed switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

It should be emphasized that the above-described embodiments of the present disclosure, particularly, any "preferred" embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A method of collecting energy comprising:

suspending at least one collection device from a plurality of support structure wires elevated above a ground level, the collection device including a carbon or graphite fiber that collects electrical enemy from the air and is electrically connected to the support structure wire; and

providing a load with an electrical connection to the plurality of support structure wires to draw current.

2. The method of claim **1**, wherein the collection device comprises a diode.

3. The method of claim **1**, wherein the collection device comprises a diode electrically connected between the fiber and the load.

4. The method of claim **1**, further comprising storing energy provided to the load.

5. The method of claim **4**, wherein storing energy provided to the load comprises storing energy in a capacitor or an inductor.

6. The method of claim **2**, wherein the diode is elevated relative to the ground level.

7. A system of energy collection comprising:

a plurality of support structure wires elevated above ground level; at least one collection device including at least one carbon or graphite fiber that collects electrical enemy from the air and is electrically connected to the plurality of support structure wires; and

a load electrically connected to the plurality of support structure wires.

8. The system of claim **7**, wherein the collection device comprises a diode.

9. The system of claim **7**, wherein the collection device comprises a diode electrically connected between the load and the fiber.

10. The system of claim **8**, wherein the diode is elevated relative to the ground level.

11. The system of claim **7**, further comprising a diode electrically connected between the at least one collection device and the plurality of support structure wires.

12. The system of claim **7**, wherein the fiber comprises a first end and a second opposing end, and wherein the system further comprises a support structure connected to both ends of the fiber.

13. The system of claim **7**, comprising:

said at least one fiber including a plurality of collection fibers; a support frame having a top side and a bottom side; and a conducting connecting wire between the plurality of support structure wires and the support frame, wherein one end of each of a plurality of collection fibers is connected to the top side of the support frame, and the opposing end of each of a plurality of collection fibers is connected to the bottom side of the support frame.

14. The system of claim **7**, further comprising a rigid structure, the rigid structure comprising multiple supports extending outward from a single point of the support structure, wherein each end of the fiber is connected to an end of the multiple supports.

15. The system of claim **7**, further comprising a rigid structure, the rigid structure comprising multiple supports extending outward from multiple points of the support structure, wherein each end of the fiber is connected to an end of the multiple supports.

16. The system of claim **14**, further comprising a plurality of rigid structures wherein said at least one fiber includes collection fibers connected between a plurality of the rigid structures.

17. The system of claim **7**, further comprising:

a switch connected in series between the plurality of support structure wires and the load; and

a capacitor connected in parallel with the switch and the load.

18. The system of claim **17**, wherein the switch comprises an interrupter connected between the plurality of support structure wires and the load.

19. The system of claim **18**, wherein the interrupter comprises one of a fluorescent tube, a neon bulb, an AC light, or a spark gap.

20. The system of claim **18**, further comprising a transformer connected between the interrupter and the load.

21. The system of claim **7**, further comprising:

a motor for providing power the motor connected between the plurality of support structure wires and the load; and

a generator powered by the motor.

22. The system of claim **7**, wherein the load comprises a spark gap in a container of fluid, and the load is used to produce a chemical reaction.

23. The system of claim **22**, wherein the fluid comprises water and the chemical reaction comprises the production of hydrogen and oxygen.

24. The system of claim **7**, further comprising a fuel cell between the plurality of support structure wires and the load.

25. The system of claim **7**, wherein the load comprises a fuel cell.

26. The system of claim **24**, wherein the fuel cell produces hydrogen and oxygen.

27. The system of claim **24**, further comprising a diode connected between the plurality of support structure wires and the fuel cell.

28. A system of collecting energy comprising:

means for suspending carbon or graphite collection fibers, the means elevated above a ground level, the collection fibers collecting electrical enemy from the air and electrically connected to the means for suspending collection fibers;

means for inducing current flow, the means for inducing current flow electrically connected to the means for suspending collection fibers; and

means for restricting the backflow of charge carriers, the means for restricting the backflow of charge carriers electrically connected between the collection fibers and the means for inducing current flow.

* * * * *

US007478712B2

(12) **United States Patent**
McCowen

(10) **Patent No.: US 7,478,712 B2**
(45) **Date of Patent: Jan. 20, 2009**

(54) **ENERGY COLLECTION**

(76) Inventor: **Clint McCowen**, 1902 Rue La Fontaine, Navarre, FL (US) 32566

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **11/839,112**

(22) Filed: **Aug. 15, 2007**

(65) **Prior Publication Data**

US 2007/0273206 A1 Nov. 29, 2007

Related U.S. Application Data

(63) Continuation-in-part of application No. 11/358,264, filed on Feb. 21, 2006.

(51) **Int. Cl.**
H02N 1/00 (2006.01)

(52) **U.S. Cl.** .... **191/45 R**; 307/149; 361/225; 361/117; 174/2; 320/101; 322/2 A; 191/1 R

(58) **Field of Classification Search** .... 361/220, 361/117, 212, 233, 221, 222, 225; 322/2 A, 322/2 R; 307/149; 174/2, 3; 320/101; 191/2, 191/1 R, 45 R
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427	A *	5/1901	Palencsar	307/149
911,260	A *	2/1909	Pennock	307/149
1,014,719	A *	1/1912	Pennock	307/149
4,104,696	A	8/1978	Cochran, Jr.	
5,047,892	A	9/1991	Sakata et al.	
6,735,830	B1	5/2004	Merciel	
6,974,110	B2	12/2005	Grandics	
2003/0107869	A1	6/2003	Pan et al.	
2004/0160711	A1	8/2004	Stumberger	
2004/0212945	A1	10/2004	Sprenger et al.	

* cited by examiner

Primary Examiner—Mark T Le
(74) *Attorney, Agent, or Firm*—Thomas, Kayden, Horstemeyer & Risley, LLP

(57) **ABSTRACT**

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support wire system supported by poles. The support wire system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but carbon, graphite, carbon/carbon, or a mixture thereof are preferred. Diodes may be used to restrict the backflow or loss of energy.

18 Claims, 15 Drawing Sheets

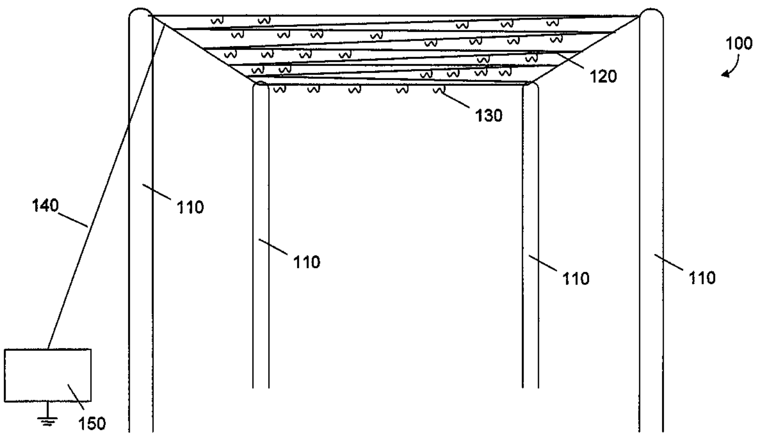
100
120
130
140
110
110
110
110
150

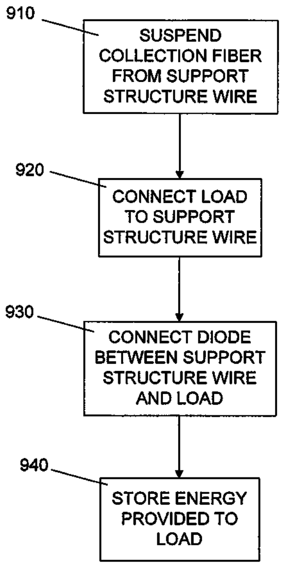
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

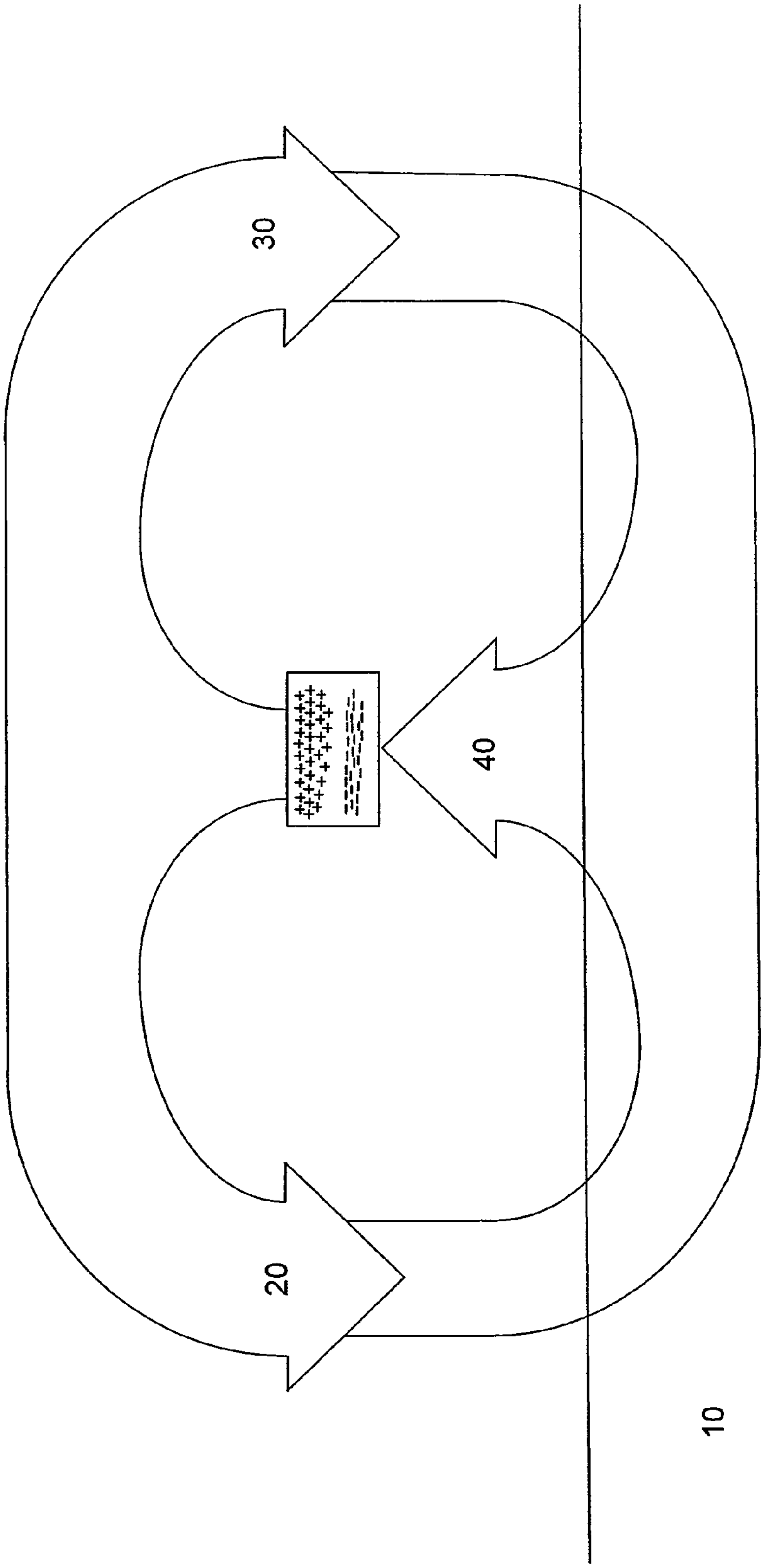

FIGURE 1

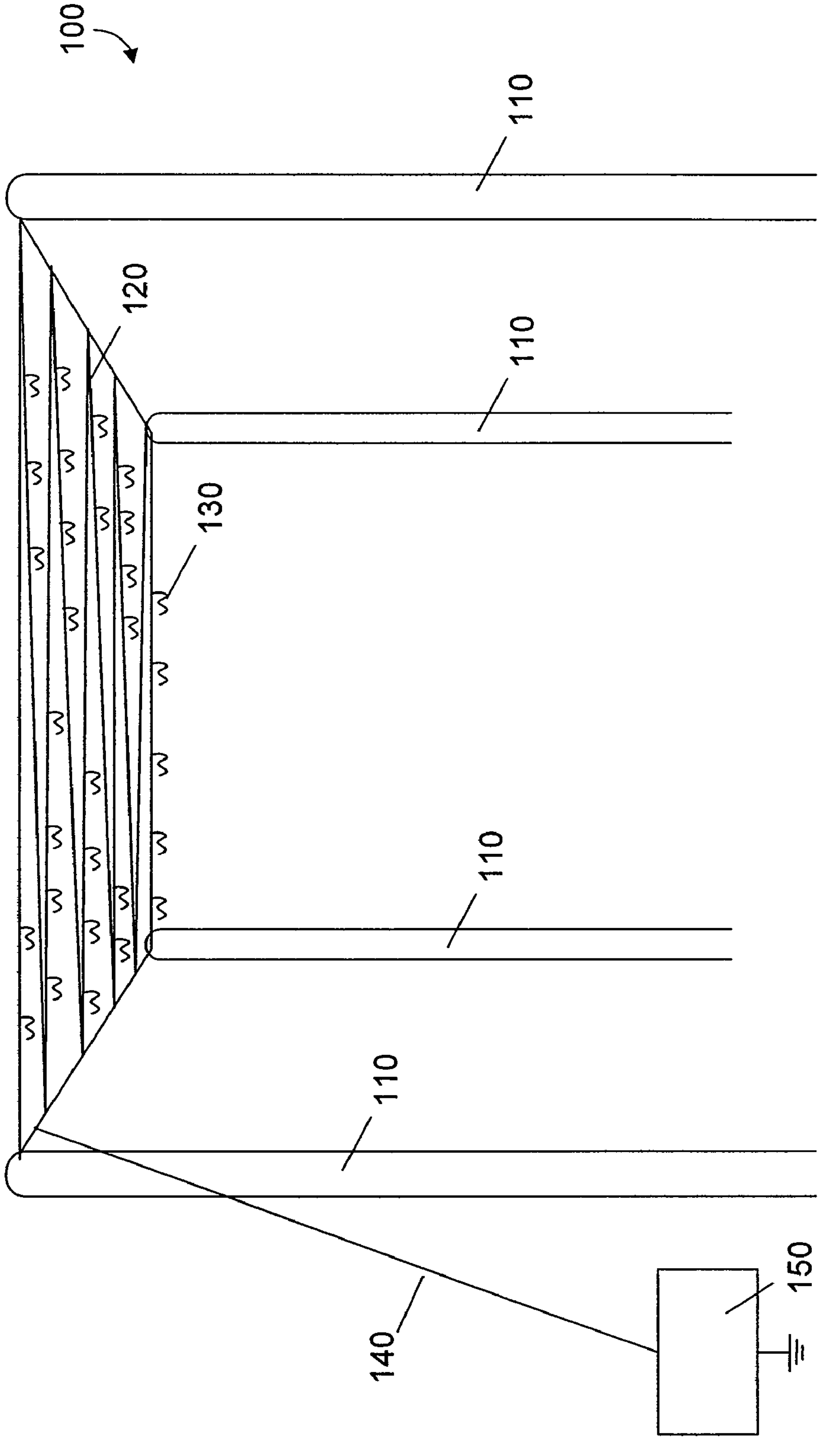
100
110
120
110
130
110
110
140
150

FIGURE 2

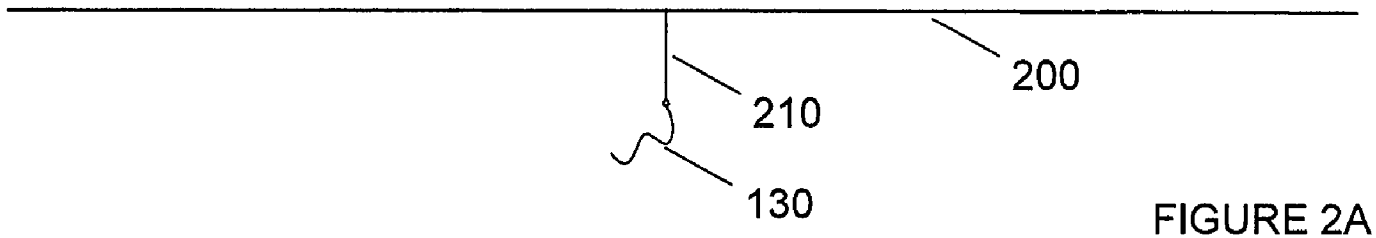

FIGURE 2A

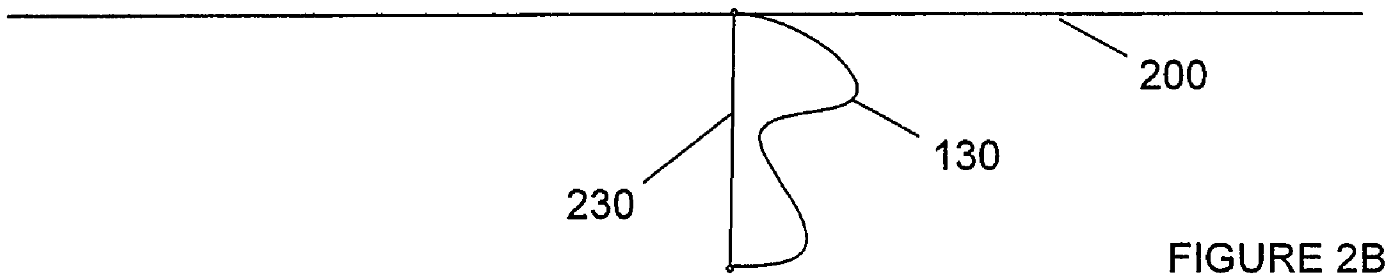

FIGURE 2B

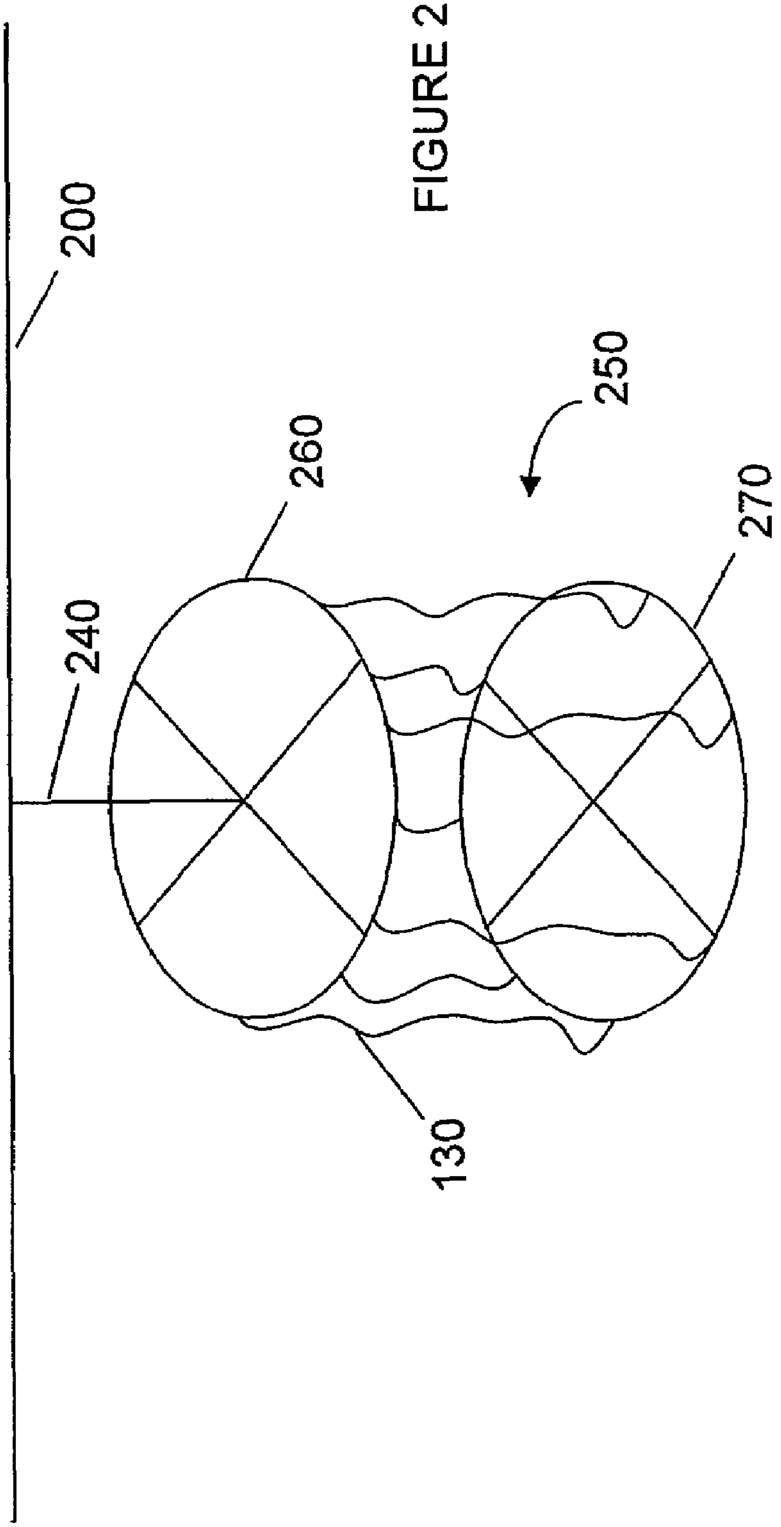
200
250
260
270
240
130

FIGURE 2C

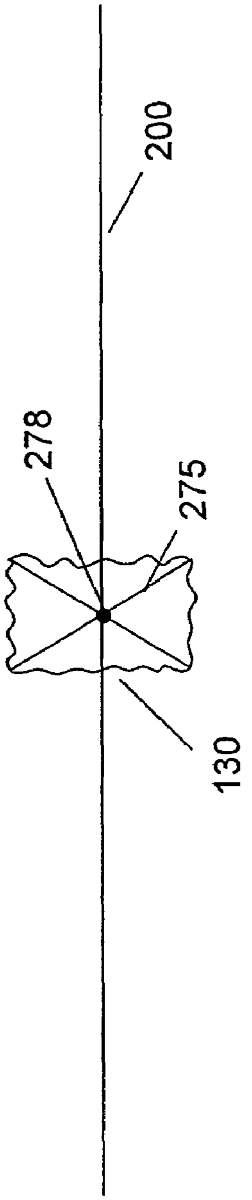
278
275
200
130

FIGURE 2D

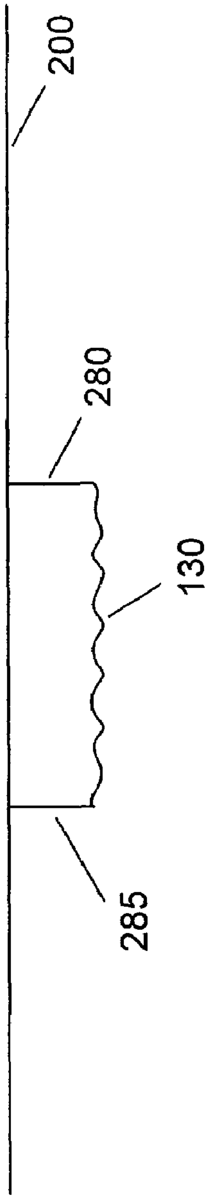
200
280
130
285

FIGURE 2E

FIGURE 2F

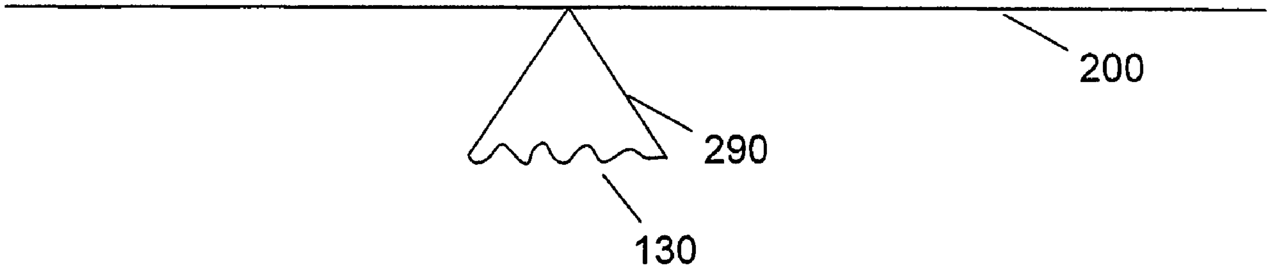

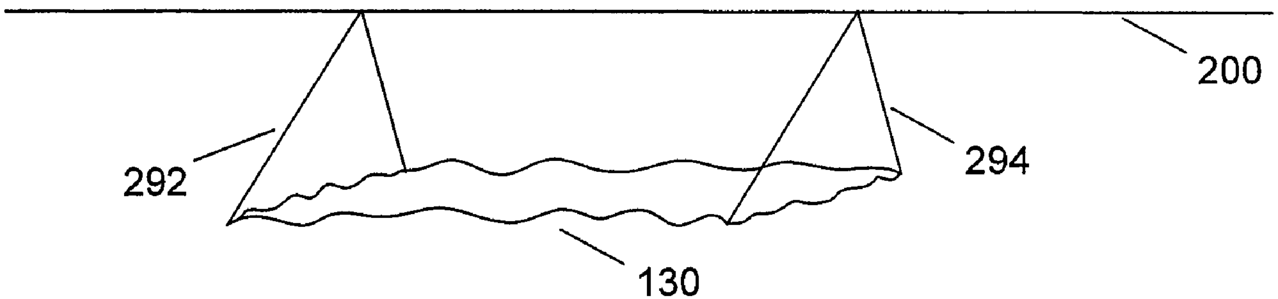
200
292
294
130

FIGURE 2G

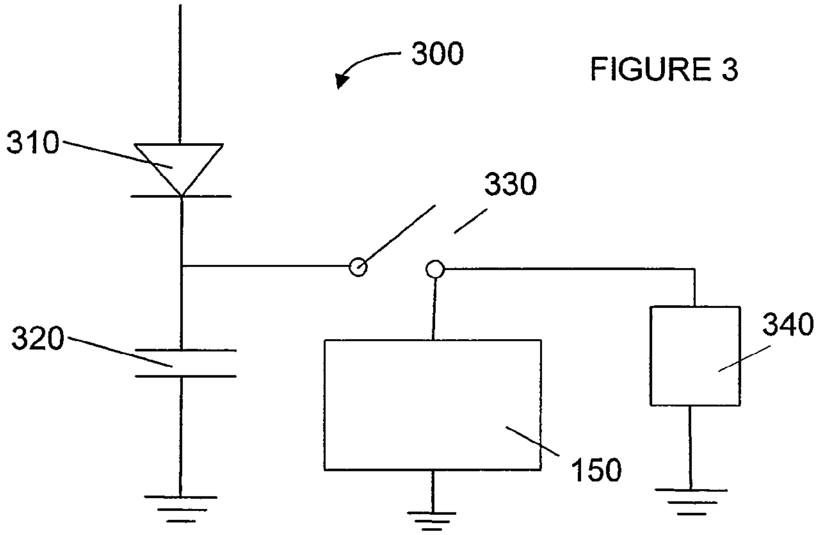
300
310
330
320
340
150

FIGURE 3

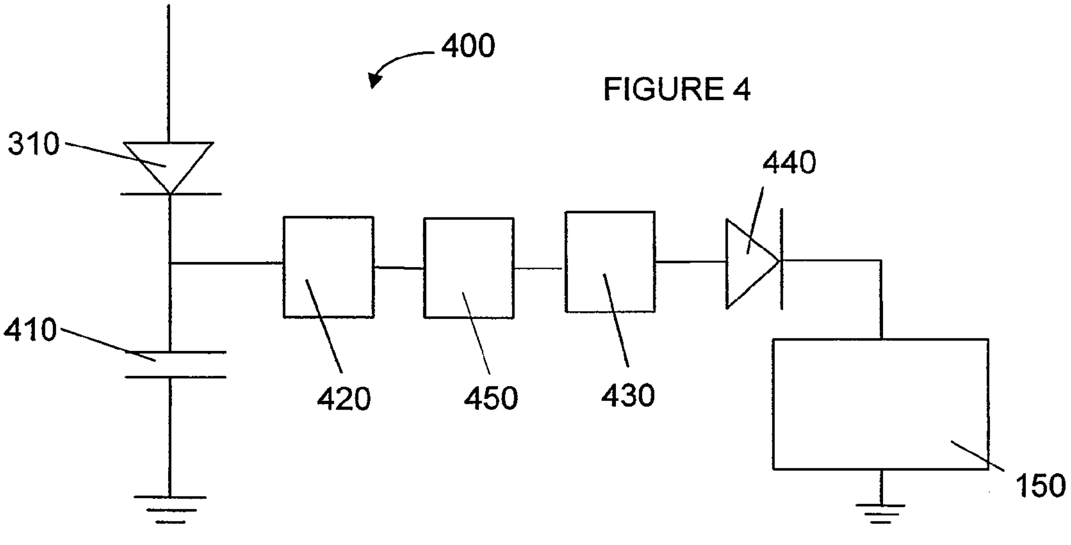
400
310
440
410
420
450
430
150

FIGURE 4

FIGURE 5

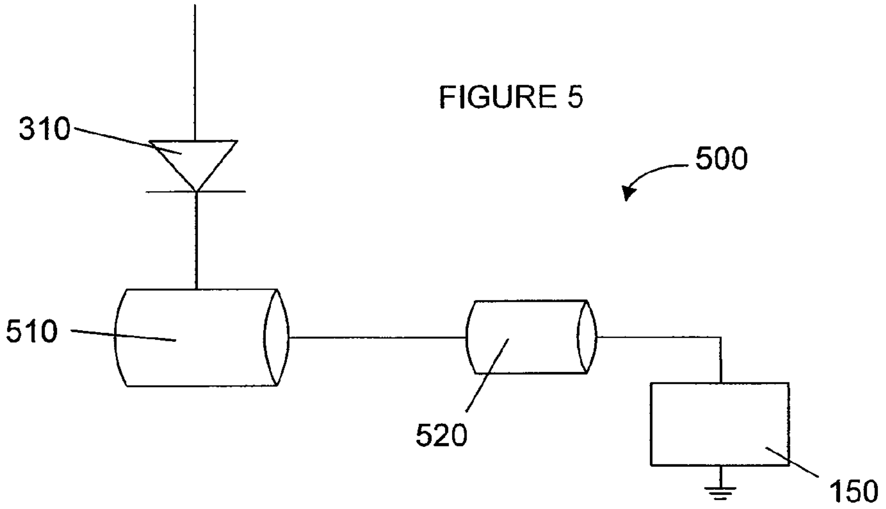
310
500
510
520
150

FIGURE 6

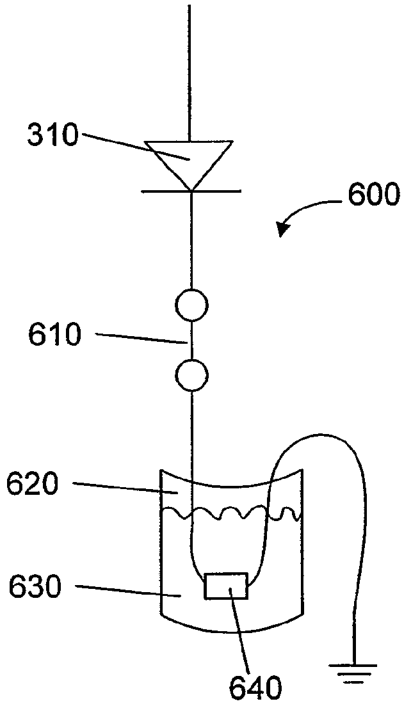
310
600
610
620
630
640

FIGURE 7

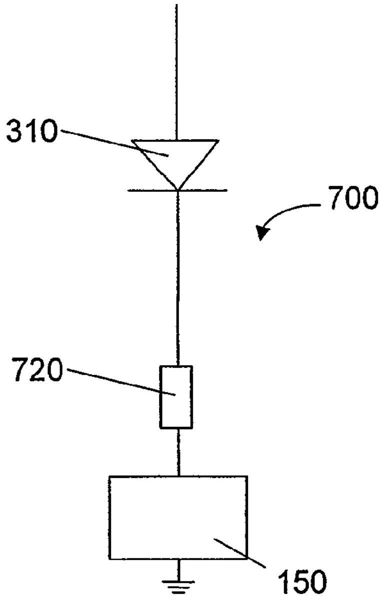
310
700
720
150

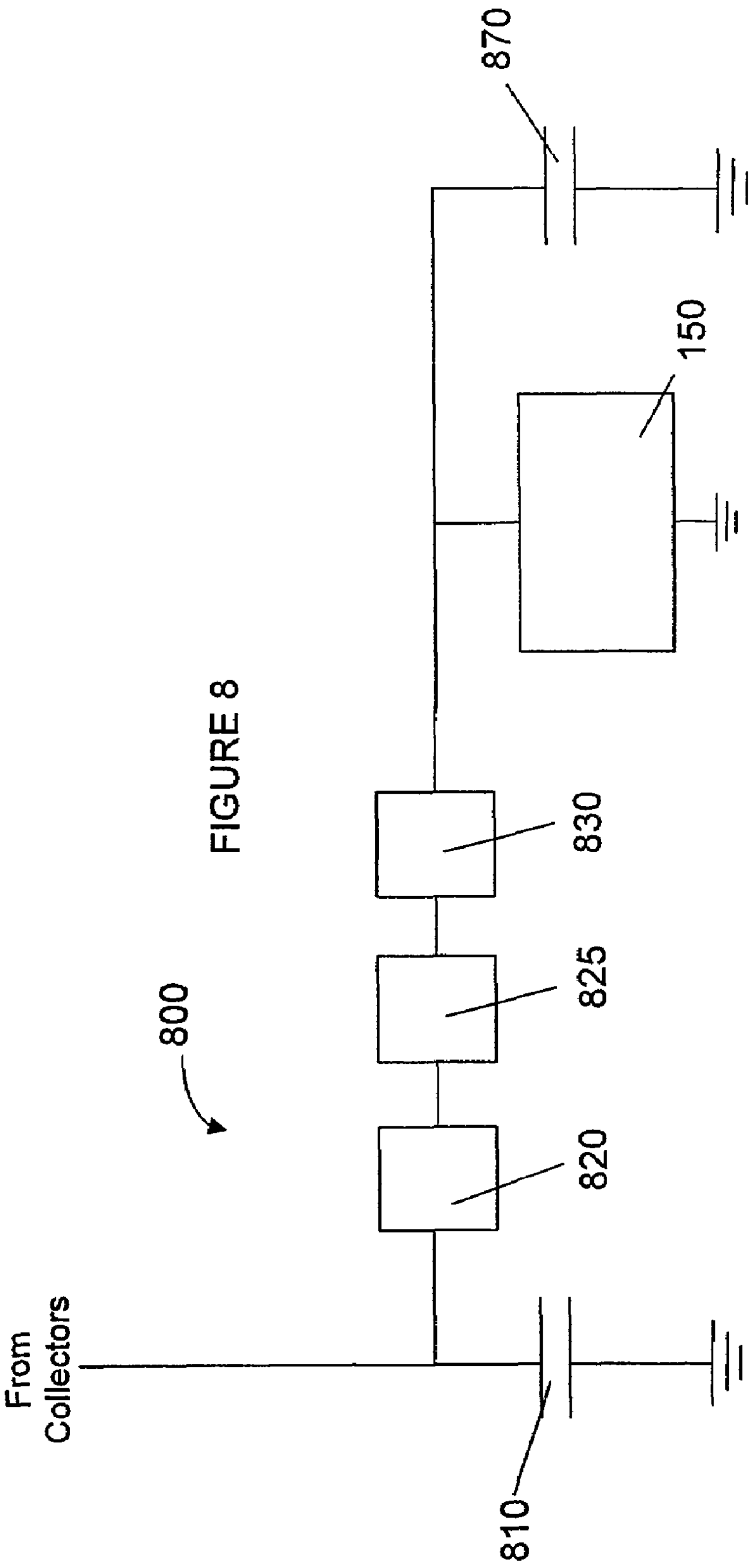
870
150
830
825
820
800
From Collectors
810

FIGURE 8

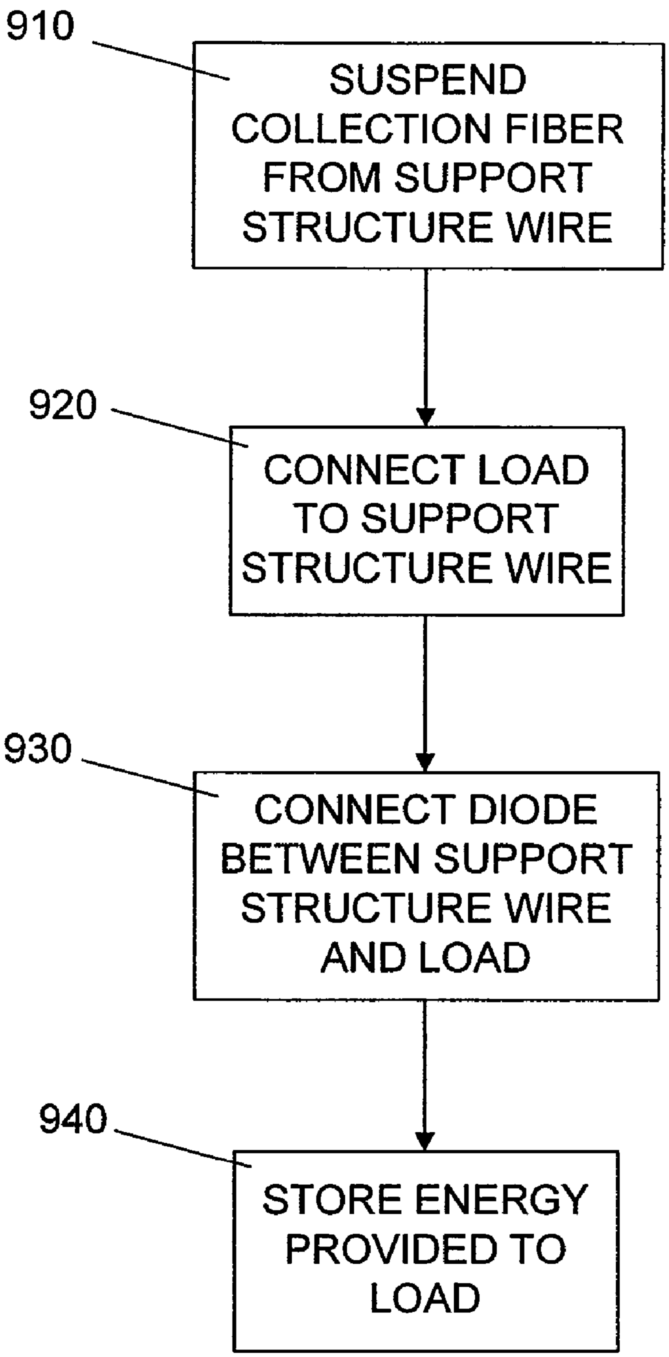
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

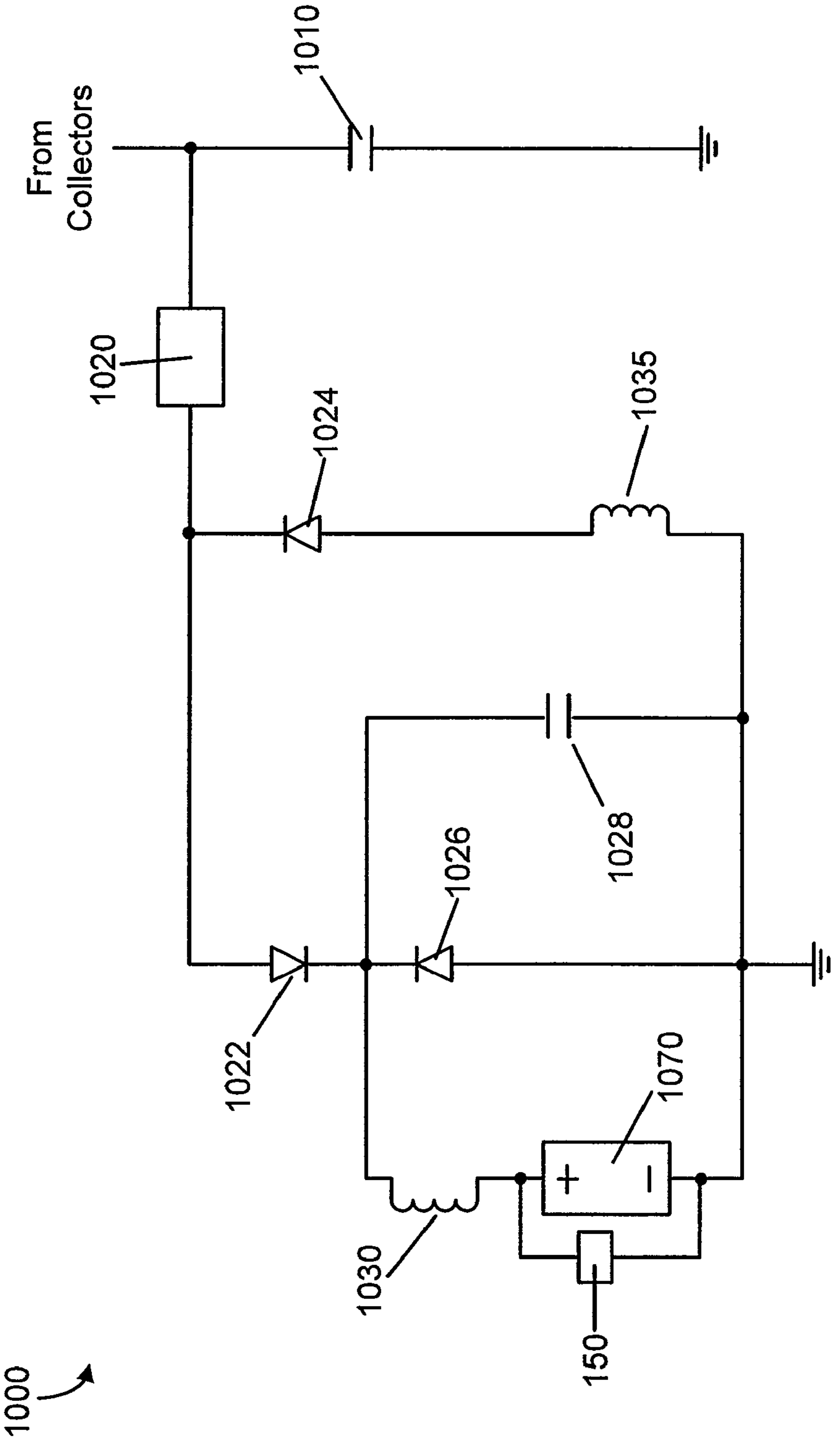
From Collectors
+
-

FIGURE 10

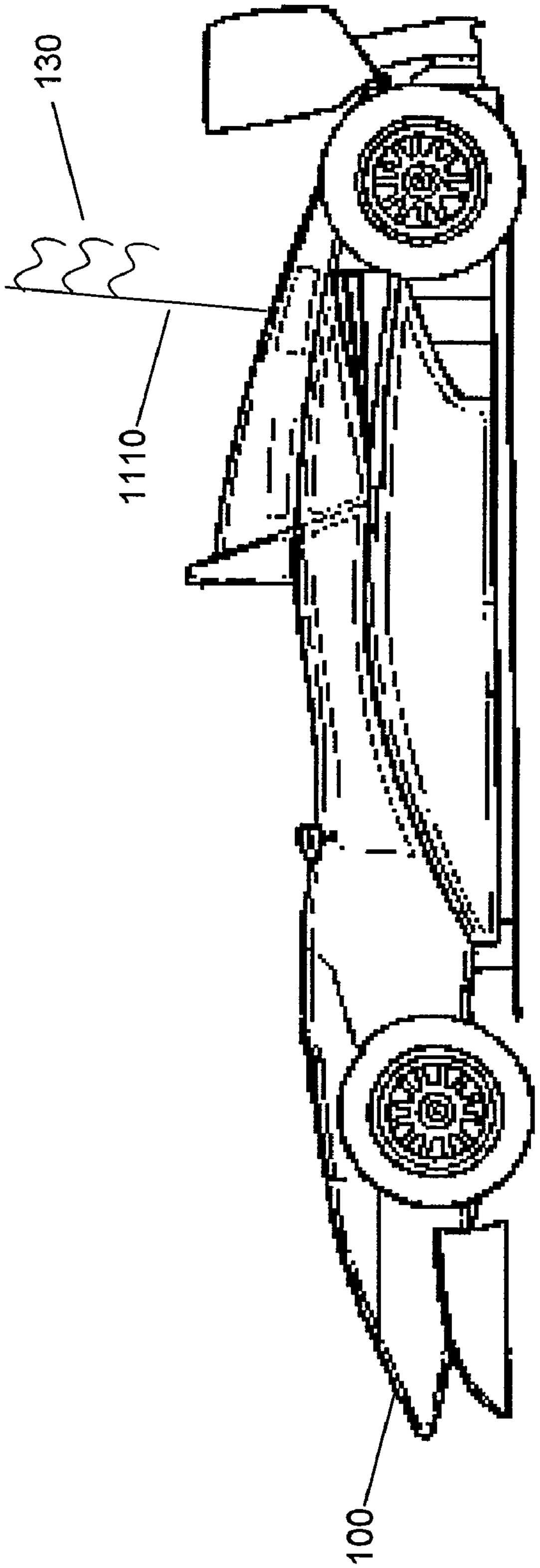
130
1110
1100

FIGURE 11

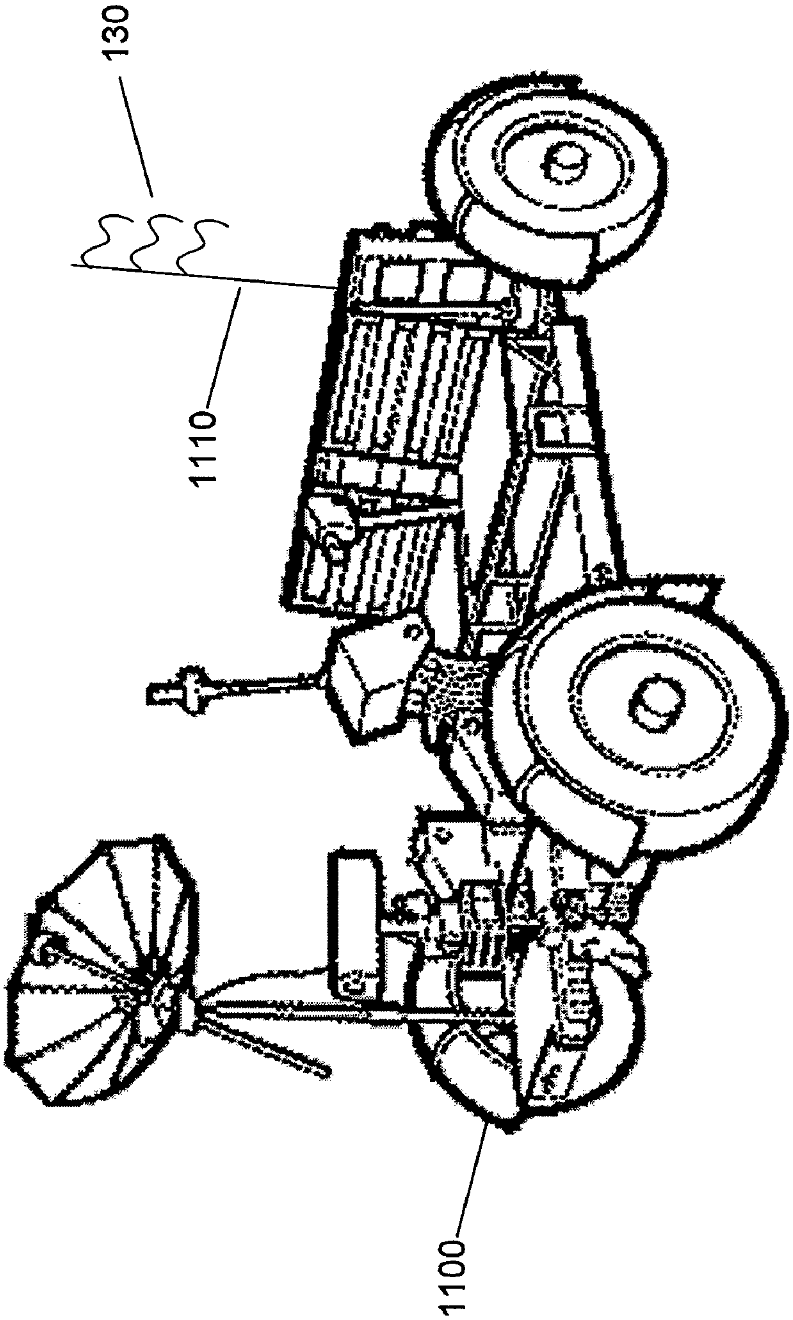
130
1110
1100

FIGURE 12

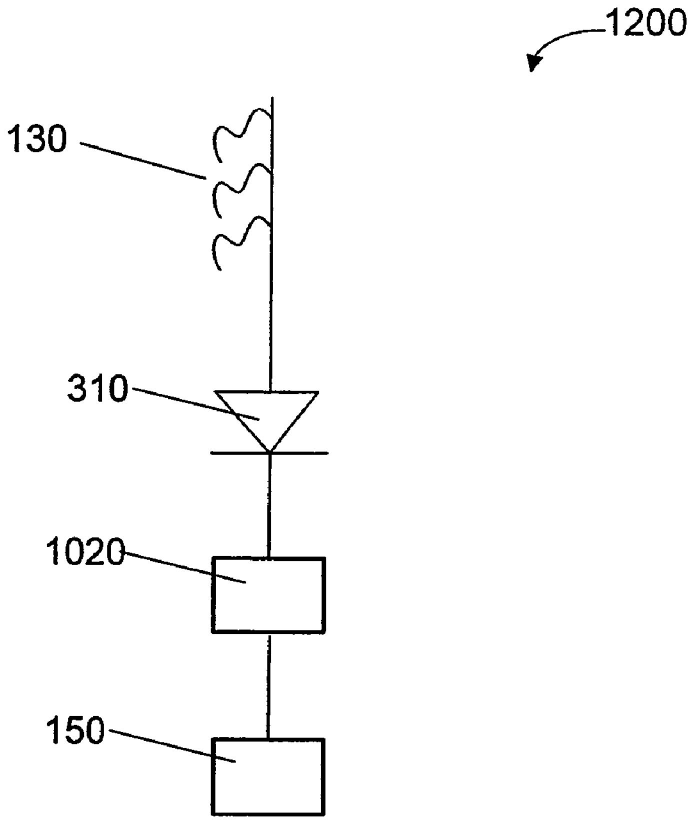
1200
130
310
1020
150

FIGURE 13

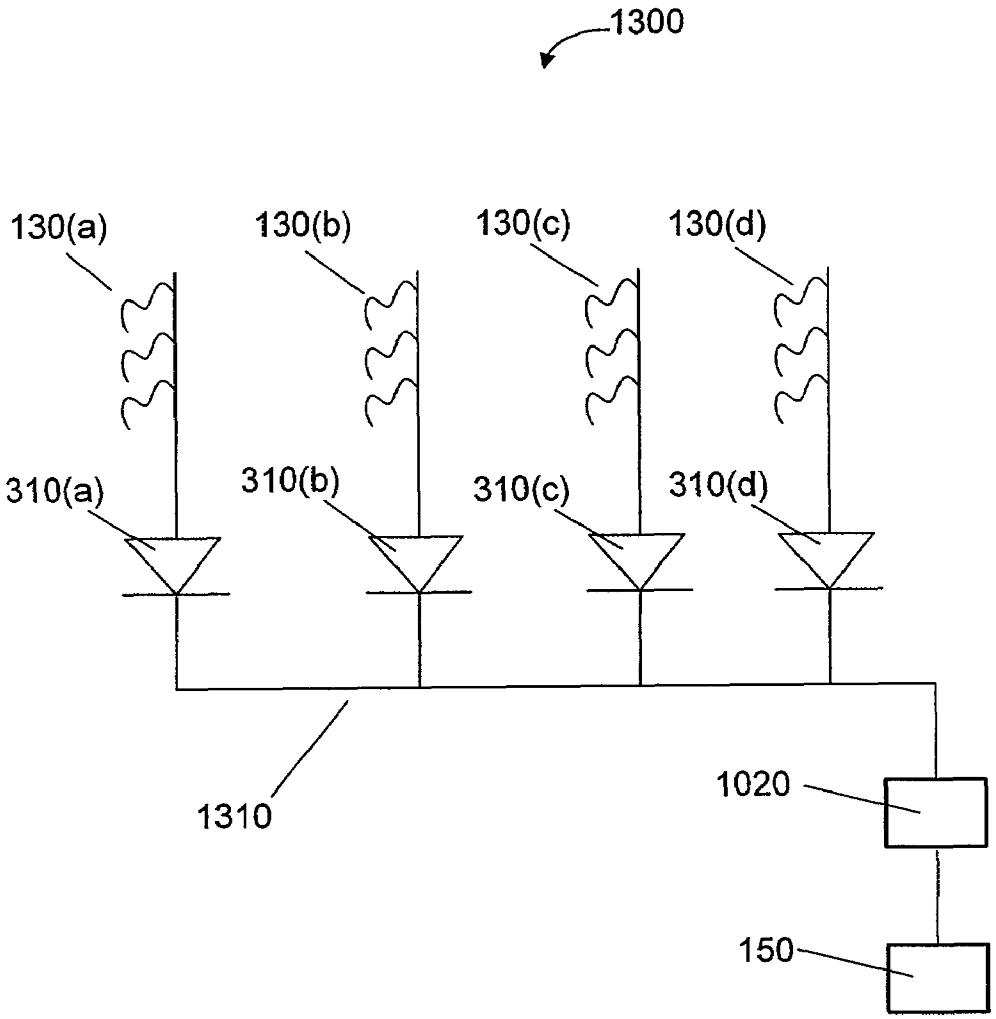
1300
130(a)
130(b)
130(c)
130(d)
310(a)
310(b)
310(c)
310(d)
1310
1020
150

FIGURE 14

ENERGY COLLECTION

CROSS REFERENCE TO RELATED APPLICATION

This application is a continuation-in-part application of U.S. patent application Ser. No. 11/358,264,filed on Feb. 21, 2006,which is incorporated by reference herein.

TECHNICAL FIELD

The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air or other gases. On Earth, clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the Earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the Earth's atmosphere or above the surface or in the atmosphere of other planets may include cosmic rays penetrating and interacting with the planet's surface and/or atmosphere, and also the creation and migration of electrically charged ions, as well as other effects yet to be fully studied.

Some of the ionization near the surface of a planet or in the lower atmosphere is caused by airborne radioactive substances. In most places on Earth, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

Alpha-active materials are also responsible for atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

While there is a large amount of usable energy available in the atmosphere of Earth, as well as above the surface or in the atmosphere of other bodies in space, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure wire elevated above a ground level, at least one collection fiber electrically connected to the support structure wire; a load electrically connected to the support structure wire; and a diode electrically connected between the load and at least one collection fiber.

Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection fiber from a support structure wire elevated above ground level, the fiber electrically connected to the support structure wire; providing a load with an electrical connection to the support structure wire to draw current; and providing a diode electrically connected between the collection fiber and the load.

Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. **1** is a circuit diagram of a weather energy circuit.

FIG. **2** is a perspective view of an example embodiment of many energy collectors elevated above ground by a structure.

FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

FIG. **2**B is a side view of an example embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

FIG. **2**D is a side view of an example embodiment of a support structure for multiple energy collection fibers.

FIG. **2**E is a side view of a support structure for an energy collection fiber.

FIG. **2**F is a side view of an example embodiment of a support structure for an energy collection fiber.

FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

FIG. **3** is a circuit diagram of an example embodiment of a circuit for the collection of energy.

FIG. **4** is a circuit diagram of an example embodiment of a circuit for the collection of energy.

FIG. **5** is a circuit diagram of an example embodiment of an energy collection circuit for powering a generator and motor.

FIG. **6** is a circuit diagram of an example embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

FIG. **7** is a circuit diagram of an example embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

FIG. **8** is a circuit diagram of an example embodiment of a circuit for collecting energy.

FIG. **9** is a flow diagram of an example embodiment of collecting energy with a collection fiber.

FIG. **10** is a circuit diagram of an example embodiment of a circuit for collecting energy from a dual polarity source.

FIG. **11** is a system diagram of an example embodiment of an energy collection system connected to an automobile vehicle.

FIG. **12** is a system diagram of an example embodiment of an energy collection system connected to a lunar rover vehicle.

FIG. **13** is a system diagram of an example embodiment of an energy collection system comprising collection devices with a diode.

FIG. **14** is a system diagram of an example embodiment of an energy collection system comprising multiple legs of the system of FIG. **13**.

DETAILED DESCRIPTION

Methods for collecting ambient energy have been known since the early 1900's. However, due to low efficiency, those methods did not become commercially feasible. One of the primary advantages of the present disclosure over past arts is the efficiency achievable by utilizing the collection fibers disclosed herein to collect or harvest said ambient energy. Another primary advantage over past arts is the dual-polarity circuit (FIG. **10**) which makes use of both positive and negative ambient energy collected by the collection fibers to convert said energy into a steady positive electrical output.

Electric charges on conductors may reside entirely on the external surface of the conductors, and tend to concentrate more around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An example embodiment of this disclosure takes advantage of this property, among others, to collect and use the energy generated by an electric field in the atmosphere or above the surface of a planet or heavenly body. “Energy” is not limited herein to energy generated by an electric field in the atmosphere. Energy as used herein may include any event, particle, or wave, including, but not limited to photons, protons, neutrons, electrons, neutrinos, leptons, muons, antimuons, mesons, baryons, positrons, bosons, gluons, gravitons, fermions, hadrons, ionic energy, gamma rays, cosmic rays, solar wind, quarks, energy generated by the triboelectric effect, convection lightning, and many other forms of energy known to one of ordinary skill in the art. These forms of energy, as well as others, may be present not only on Earth, but also on the moon of the Earth, on Mars, on Venus, and on other heavenly bodies and even in the relative vacuum of deep space.

Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An example embodiment utilizes carbon fibers, graphite fibers, carbon/carbon fibers, or a mixture thereof for energy collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including, but not limited to, aluminum, steel, copper, and super-conductive material. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

An example embodiment of the collection fibers as collection device **130** includes graphite, carbon fibers, carbon/carbon fibers, or a mixture thereof. Graphite, carbon fibers, and carbon/carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity, energetic particles, or waves may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, graphite fibers, carbon/carbon fibers, or a mixture thereof are examples that demonstrate example collection ability.

In at least one example embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than 100-300 volts per meter under some conditions on Earth. When support wire **120** is suspended at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An example embodiment of the collection device comprises the diode with no collection fiber. Another embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restrict energy that may be collected through one fiber from escaping through another fiber.

Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2A**-**2G** using a collection fiber embodiment. FIG. **2A** presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. One embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**. In an alternative embodiment, a diode is not placed between collection device **130** and connection member **210**.

Likewise, FIG. **2B** shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting.

A plurality of diodes may be placed at any position on the support structure wire. One embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**. In an alternative embodiment, a diode is not placed between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

FIG. **2**C presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. One embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**. In an alternative embodiment, a diode is not placed at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

FIG. **2**D presents another example embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. One embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**. In an alternative embodiment, a diode is not placed at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2**E provides another example embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. One embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**. In an alternative embodiment, one or more diodes are not placed at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

FIG. **2**F presents another example embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. One embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**. In an alternative embodiment, a diode is not placed at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2**G provides two supports as provided in FIG. **2**F such that at least two support members **292**, **294** may be connected to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and/or between multiple support structures. A plurality of diodes may be placed at any position on the support structure. One embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**. In an alternative embodiment, one or more diodes are not placed at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load. Capacitor **320** may be connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

In an example embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **420** include mercury-reed switches, mercury-wetted reed switches, open-gap spark gaps, and electronic switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground. The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes may be placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches, mercury-wetted reed switches, open-gap spark gaps, and electronic switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel or as components for breathable air, among other uses.

FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit.

Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

FIG. **8** presents example circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches, mercury-wetted reed switches, open-gap spark gaps, and electronic switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

FIG. **10** presents circuit **1000** as an example embodiment for the collection of energy from a dual polarity source. This may be used, for example, to collect atmospheric energy that reverses in polarity compared with the ground. Such polarity reversal has been discovered as occurring occasionally on Earth during, for example, thunderstorms and bad weather, but has also been observed during good weather. Such polarity reversal may occur on other planetary bodies, including Mars and Venus, as well. Energy polarity on other planets, in deep space, or on other heavenly bodies, may be predominantly negative or predominantly positive. Collector fibers (**130** from FIG. **2**) are capable of collecting positive energy and/or negative energy, and circuit **1000** is capable of processing positive and/or negative energy, providing an output which is always positive. Circuit **1000** may collect energy from one or more collection devices (**130** from FIG. **2**). Charging capacitor **1010** may be used to store a charge until the voltage at spark gap **1020** achieves the spark voltage. Capacitor **1010** is optional.

A plurality of diodes may be placed in a plurality of positions in circuit **1000**. The voltage from capacitor **1010** may be used to charge spark gap **1020** to a sufficient voltage. Spark gap **1020** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **1020** include mercury-reed switches, mercury-wetted reed switches, open-gap spark gaps, and electronic switches. When spark gap **1020** arcs, energy will arc from an emitting end of spark gap **1020** to a receiving end of spark gap **1020**. The output of spark gap **1020** is electrically connected to the anode of diode **1022** and the cathode of diode **1024**. The cathode of diode **1022** is electrically connected to the cathode of diode **1026** and inductor **1030**. Inductor **1030** may be a fixed value inductor or a variable inductor. The anode of diode **1026** is electrically connected to ground. Capacitor **1028** is electrically connected between ground and the junction of the cathodes of diode **1022** and diode **1026**. Inductor **1035** is electrically connected between ground and the anode of diode **1024**. Inductor **1035** may be a fixed value inductor or a variable inductor. Capacitor **1070**, the anode of diode **1026**, inductor **1035**, and load **1050** are electrically connected to ground. Capacitor **1070** may be placed in parallel with load **150**.

FIGS. **11** and **12** provide example embodiments of vehicle **1110**, which utilizes electricity, the vehicle employing systems of energy collection provided herein. Vehicle **1100** in FIG. **11** is shown as an automobile vehicle, but could be any means of locomotion that utilizes electricity, including a car, a train, a motorcycle, a boat, an airplane, robotic rovers, space craft, etc. Vehicle **1100** in FIG. **12** is shown as a lunar rover vehicle. In FIGS. **11** and **12**, support rod **1110** is electrically connected to an electrical system in vehicle **1100**. Energy collectors **130** are electrically connected to support rod **1110** and may be used to supply energy to electrical circuits within the vehicle. A non-limiting use includes a top-off charge for a battery system, on-board hydrogen production, and/or assisting in the same. Energy collectors **130** may be used to augment the efficiency of the locomotion that utilizes electrical energy as well.

FIG. **13** provides an example embodiment of energy collection system **1200** in which diode **310** is used to isolate collection devices **130** from spark gap **1020** and load **150**. Collection devices **130** may comprise graphite, carbon fibers, carbon/carbon fibers, or a mixture thereof.

FIG. **14** provides an example embodiment of energy collection system **1300** in which a plurality of energy collection systems, such as that provided in FIG. **13**, are combined. Each leg consisting of collection devices **130** and diode **310** are connected in parallel with other legs, each leg electrically connected to trunk wire **1310**. The legs could also be connected serially. Trunk wire **1310** is electrically connected to a collection circuit, which may comprise load **150** and spark gap **1020** in any configuration that has been previously discussed. Each leg may comprise one or more collection devices **130** and at least one diode electrically connected between the collection devices and the collection circuit. Although three collection devices **130** are shown on each leg, any number of collection devices may be used. Although four legs are shown, any number of legs may be used.

Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

It should be emphasized that the above-described embodiments of the present disclosure, particularly, any "preferred" embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A system of energy collection comprising: a plurality of support structure wires elevated above ground level; at least one collection device including at least one carbon or graphite fiber that collects electrical energy from the air and is electrically connected to the plurality of support structure wires; a load electrically connected to the plurality of support structure wires; and a dual polarity collection circuit electrically connected between the collection device and the load.

2. The system of claim **1**, wherein the collection device comprises a diode.

3. The system of claim **1**, wherein the collection device comprises:

a diode electrically connected between the load and the fiber.

4. The system of claim **2**, wherein the diode is elevated relative to the ground level.

5. The system of claim **1**, wherein the fiber comprises graphite, carbon fibers, carbon/carbon fibers, or a mixture thereof.

6. The system of claim **1**, further comprising a diode electrically connected between the at least one collection device and the plurality of support structure wires.

7. The system of claim **1**, wherein the dual polarity collection circuit comprises a plurality of diodes, at least one inductor, and at least one capacitor.

8. The system of claim **7**, wherein the capacitor is electrically connected in parallel with the load.

9. A vehicle employing a system of energy collection comprising: a means of locomotion; a support structure electrically connected to the vehicle; at least one collection device including at least one carbon or graphite fiber that collects electrical energy from the air and is electrically connected to the support structure; and a load electrically connected to the support structure.

10. The structure of claim **9**, wherein the collection device comprises a diode.

11. The structure of claim **9**, wherein the collection device comprises: a diode electrically connected between the load and the fiber.

12. The structure of claim **10**, wherein the diode is elevated relative to the vehicle.

13. The structure of claim **9**, wherein the fiber comprises graphite, carbon fibers, carbon/carbon fibers, or a mixture thereof.

14. The structure of claim **9**, further comprising a diode electrically connected between the at least one collection device and the support structure.

15. A system of collecting energy comprising:

a plurality of collection legs, each of the collection legs comprising:

at least one collection device including at least one carbon or graphite fiber that collects electrical energy from the air; and

a diode; and

a load electrically connected to ground.

16. The system of claim **15**, wherein the fiber comprises a surface having projections.

17. The system of claim **15**, further comprising a trunk line electrically connected between the plurality of collection legs and the load.

18. The system of claim **17**, further comprising a spark gap device electrically connected between the plurality of collection legs and the load.

* * * * *

US008686575B2

(12) **United States Patent**
McCowen

(10) **Patent No.: US 8,686,575 B2**
(45) **Date of Patent: Apr. 1, 2014**

(54) **ENERGY COLLECTION**

(75) Inventor: **Clint McCowen**, Navarre, FL (US)

(73) Assignee: **Ion Power Group, LLC**, Navarre, FL (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **12/255,130**

(22) Filed: **Oct. 21, 2008**

(65) **Prior Publication Data**

US 2009/0040680 A1 Feb. 12, 2009

Related U.S. Application Data

(63) Continuation of application No. 11/358,264, filed on Feb. 21, 2006, now Pat. No. 7,439,712.

(51) **Int. Cl.**
F02B 63/04 (2006.01)
F03G 7/08 (2006.01)
H02K 7/18 (2006.01)
H02N 1/00 (2006.01)

(52) **U.S. Cl.**
USPC .... **290/1 R**; 322/2 A

(58) **Field of Classification Search**
USPC .... 290/1 R; 322/2 A; 126/660; 60/641.8, 60/641.11, 641.12, 659, 698; 244/158.2
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427 A 5/1901 Palencsar 307/149
911,260 A 2/1909 Pennock 307/149
1,014,719 A 1/1912 Pennock 307/149
2,473,819 A * 6/1949 Pittman 315/36
3,532,959 A * 10/1970 Erickson 320/166
3,780,722 A * 12/1973 Swet 126/680
3,946,227 A 3/1976 Bingham 250/281
4,079,225 A * 3/1978 Warner 219/110
4,092,250 A 5/1978 Sano et al. 210/500.29
4,104,696 A 8/1978 Cochran, Jr. 361/229
4,146,800 A * 3/1979 Gregory et al. 290/44
4,180,056 A * 12/1979 Schnabel et al. 126/660
4,201,197 A * 5/1980 Dismer 126/600
4,206,396 A * 6/1980 Marks .... 322/2 A
4,224,496 A * 9/1980 Riordan et al. 219/110
4,307,936 A * 12/1981 Ochiai 385/120
4,314,192 A * 2/1982 Caro .... 322/2 R
4,346,478 A * 8/1982 Sichling 398/107
4,425,905 A * 1/1984 Mori 126/578
4,433,248 A * 2/1984 Marks 290/44
RE31,678 E * 9/1984 Ochiai 385/115

(Continued)

FOREIGN PATENT DOCUMENTS

WO 2007098341 8/2007

OTHER PUBLICATIONS

“Emissivity Values for Common Materials & Non-Metals”, Raytek—A Fluke Company, http://www.raytek.com/Raytek/en-r0/IREducation/EmissivityNonMetals.htm, 1999.*

Primary Examiner — Pedro J Cuevas
(74) *Attorney, Agent, or Firm* — Benjamin A. Balser; Balser & Grell IP Law

(57) **ABSTRACT**

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support wire system supported by poles. The support wire system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but carbon and graphite are preferred. Diodes may be used to restrict the backflow or loss of energy.

25 Claims, 10 Drawing Sheets

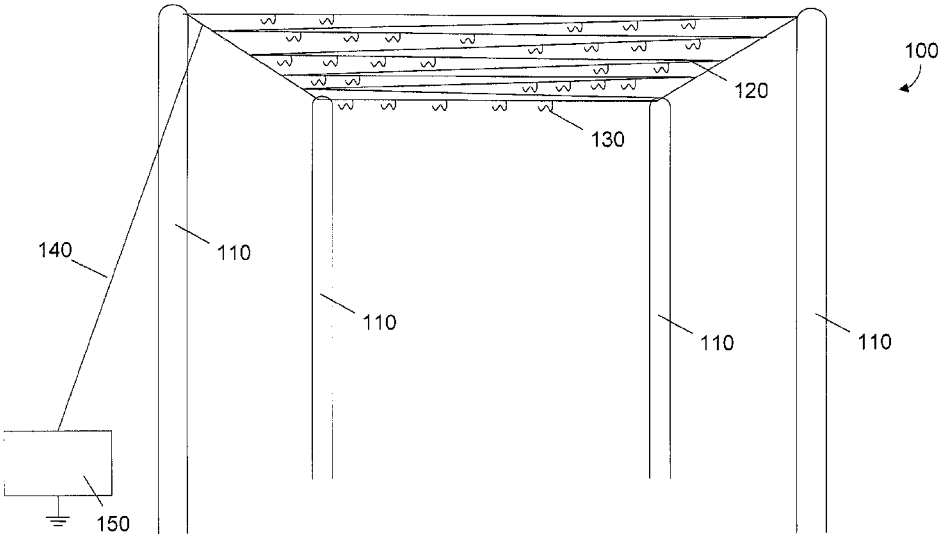

(56) References Cited

U.S. PATENT DOCUMENTS

4,483,311	A	*	11/1984	Whitaker	126/602
4,489,269	A	*	12/1984	Edling et al.	322/2 R
4,512,335	A	*	4/1985	Mori	126/678
4,653,223	A	*	3/1987	Mori	47/17
4,653,472	A	*	3/1987	Mori	126/683
4,676,226	A	*	6/1987	Mori	126/680
4,809,675	A	*	3/1989	Mori	126/578
4,852,454	A	*	8/1989	Batchelder	89/1.11
4,943,125	A	*	7/1990	Laundre' et al.	385/35
4,988,159	A	*	1/1991	Turner et al.	385/33
5,047,892	A		9/1991	Sakata et al.	361/231
5,114,101	A	*	5/1992	Stern et al.	244/172.8
5,145,257	A	*	9/1992	Bryant et al.	374/131
5,379,103	A	*	1/1995	Zigler	356/73
5,851,309	A	*	12/1998	Kousa	136/248
5,942,806	A	*	8/1999	Veliadis	290/1 R
6,025,591	A		2/2000	Taylor et al.	250/292
6,038,363	A	*	3/2000	Slater et al.	385/147
6,116,544	A	*	9/2000	Forward et al.	244/158.2
6,173,922	B1	*	1/2001	Hoyt et al.	244/158.2
6,226,440	B1	*	5/2001	Lyons	385/119
6,419,191	B1	*	7/2002	Hoyt et al.	244/158.2
6,425,391	B1	*	7/2002	Davoren et al.	126/683
6,735,830	B1		5/2004	Merciel	250/324
6,765,212	B2	*	7/2004	Goetz et al.	250/339.11
6,846,447	B2	*	1/2005	Mbachu et al.	264/406
6,853,447	B2	*	2/2005	Goetz	356/237.1
6,894,772	B2	*	5/2005	Goetz et al.	356/237.1
6,925,852	B2	*	8/2005	Susko	73/23.2
6,974,110	B2		12/2005	Grandics	244/171.5
7,109,597	B1	*	9/2006	Bose	290/43
7,128,867	B2	*	10/2006	Mbachu et al.	264/406
7,231,809	B2	*	6/2007	Susko	73/31.05
7,259,475	B2	*	8/2007	Hong et al.	307/64
7,439,712	B2	*	10/2008	McCowen	322/2 A
7,449,668	B2	*	11/2008	Schutten et al.	250/214 SW
7,478,712	B2	*	1/2009	McCowen	191/45 R
7,532,819	B1	*	5/2009	Triebes et al.	398/121
2002/0026933	A1	*	3/2002	Gottlieb	126/628
2002/0108892	A1	*	8/2002	Goetz et al.	209/576
2002/0109094	A1	*	8/2002	Goetz et al.	250/339.11
2002/0109835	A1	*	8/2002	Goetz	356/237.1
2003/0107869	A1		6/2003	Pan et al.	361/306.2
2003/0125630	A1	*	7/2003	Furnish	600/461
2003/0193319	A1	*	10/2003	Wood et al.	322/2 A
2004/0011925	A1	*	1/2004	Grandics	244/62
2004/0083793	A1	*	5/2004	Susko	73/31.05
2004/0094853	A1	*	5/2004	Mbachu et al.	264/40.1
2004/0160711	A1		8/2004	Stumberger	361/20
2005/0101024	A1	*	5/2005	Mbachu et al.	436/85
2005/0136311	A1	*	6/2005	Ueda et al.	429/30
2005/0270525	A1	*	12/2005	Susko	356/326
2007/0107765	A1	*	5/2007	Schutten et al.	136/243
2007/0195481	A1	*	8/2007	McCowen	361/233
2007/0218323	A1	*	9/2007	Sudo et al.	429/13
2007/0223171	A1		9/2007	Guy et al.	361/126
2007/0273206	A1	*	11/2007	McCowen	307/9.1
2007/0297892	A1	*	12/2007	Kildegaard	415/1
2008/0145714	A1	*	6/2008	Kagami	429/13
2009/0107842	A1		4/2009	Park et al.	204/554
2009/0114495	A1	*	5/2009	McCowen	191/45 R
2009/0216292	A1	*	8/2009	Pless et al.	607/33
2010/0090562	A1	*	4/2010	McCowen	310/309
2010/0090563	A1	*	4/2010	McCowen	310/309
2012/0164555	A1	*	6/2012	Yamazaki et al.	429/482
2012/0299559	A1	*	11/2012	McCowen	322/2 A
2013/0276776	A1	*	10/2013	Te Kamp et al.	126/660

* cited by examiner

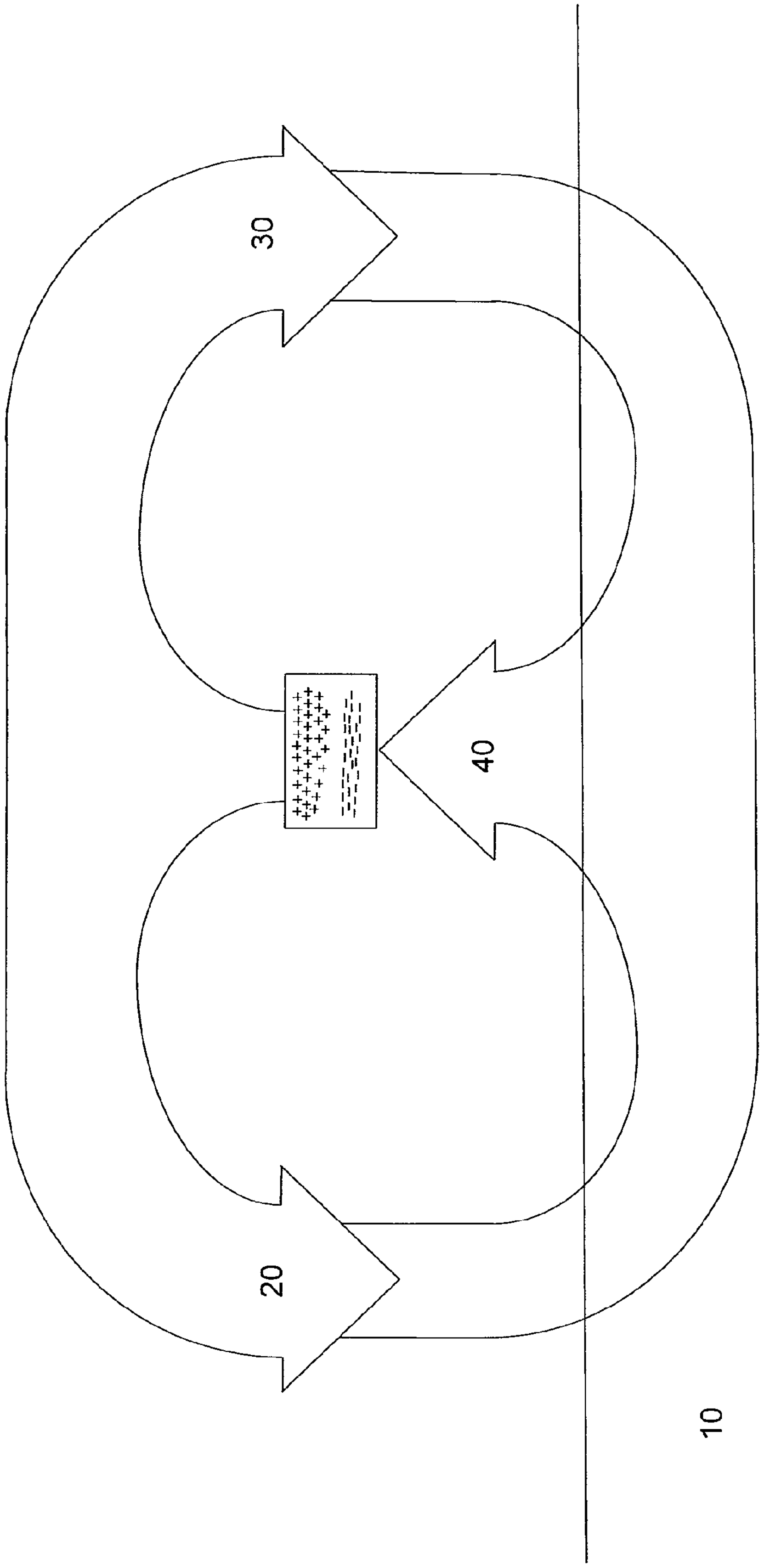

FIGURE 1

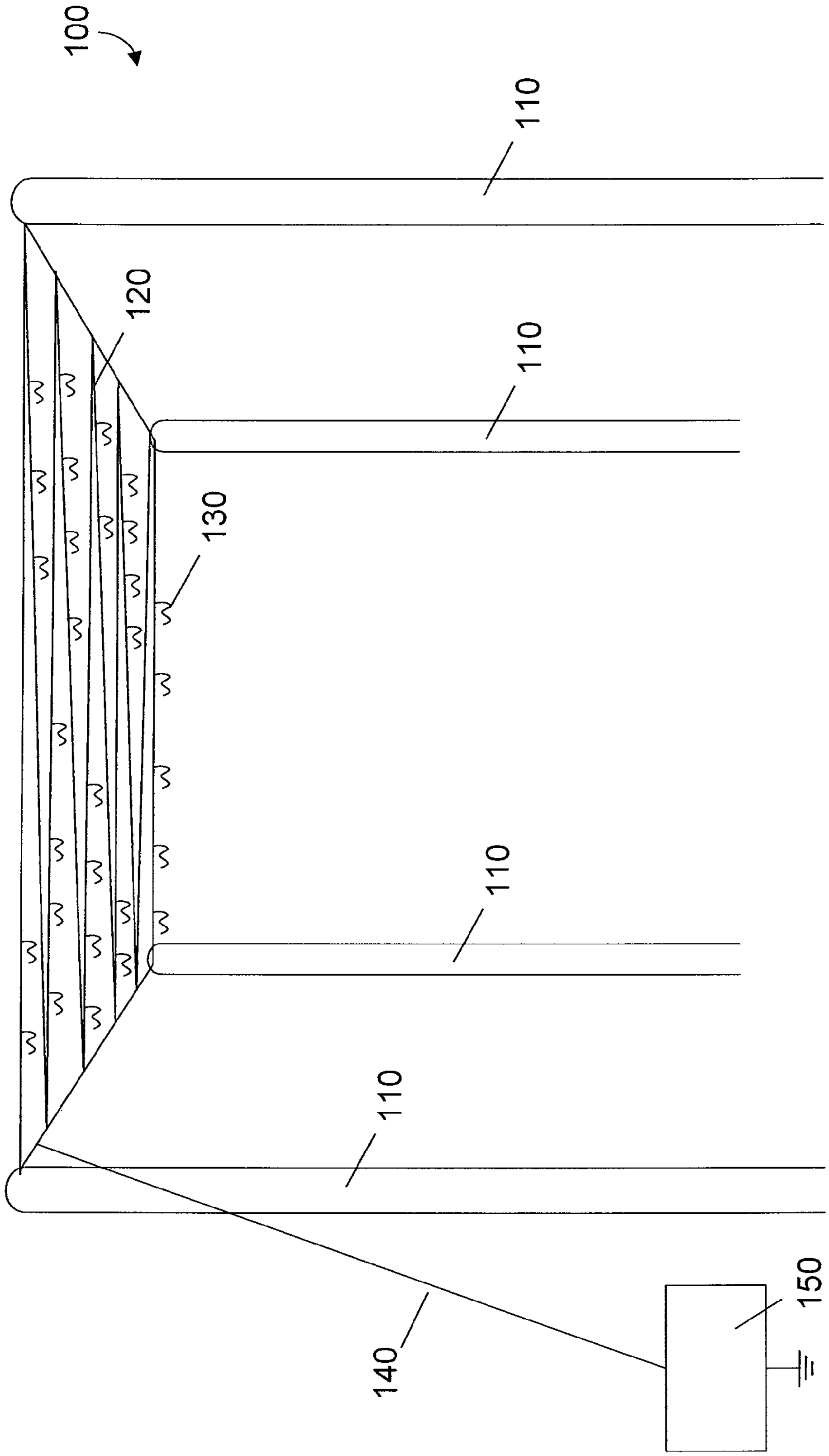
100
110
110
120
130
110
110
140
150

FIGURE 2

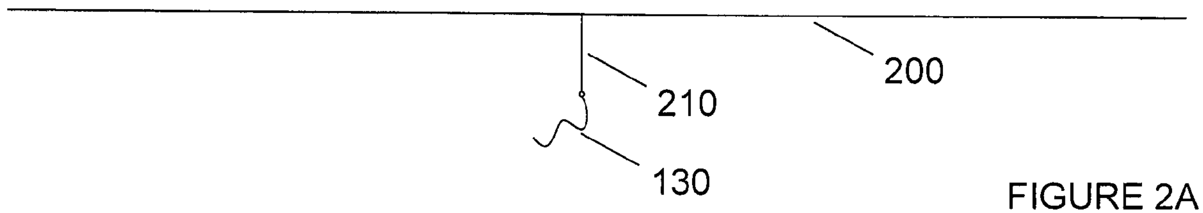

FIGURE 2A

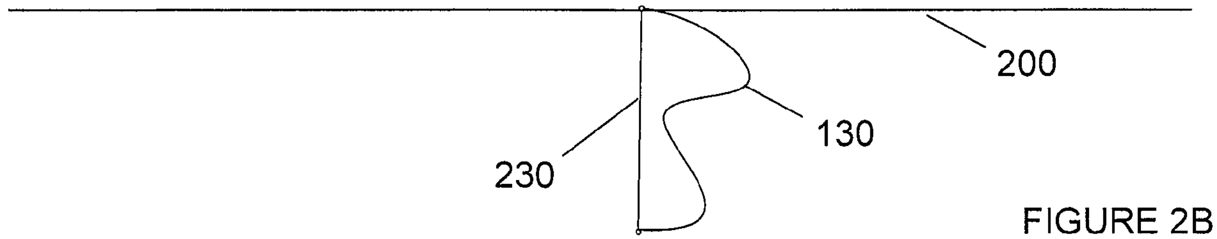

FIGURE 2B

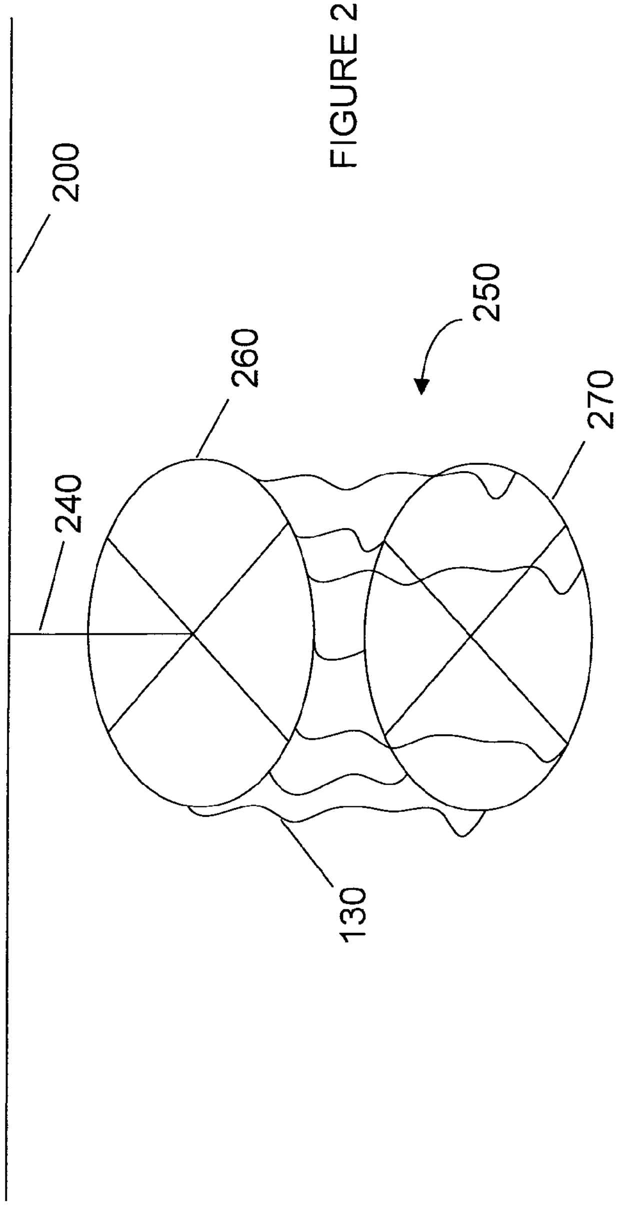
200
250
260
270
240
130

FIGURE 2C

FIGURE 2D

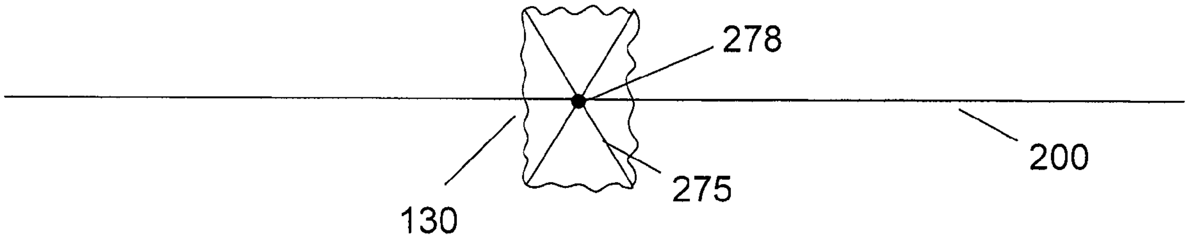
278
200
275
130

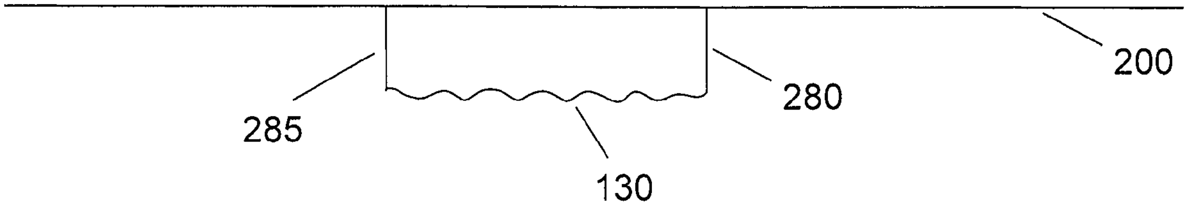
200
280
285
130

FIGURE 2E

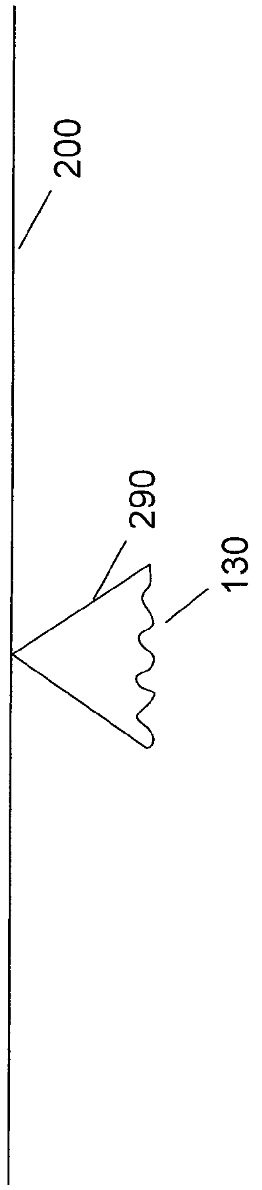

FIGURE 2F

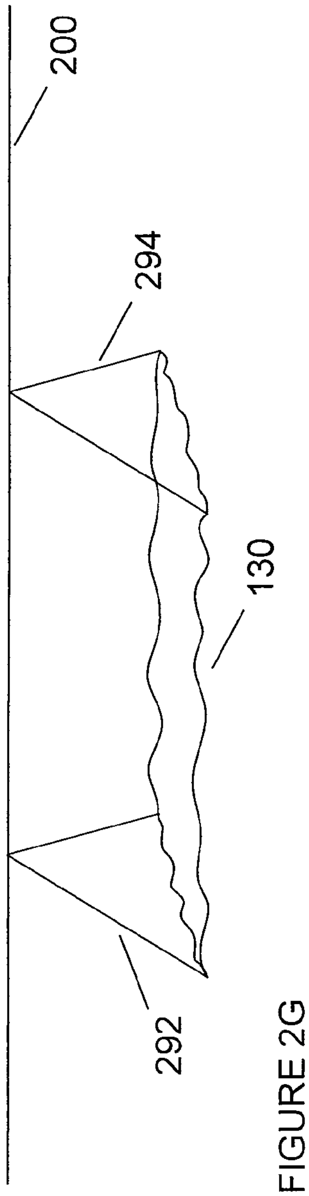
200
294
130
292

FIGURE 2G

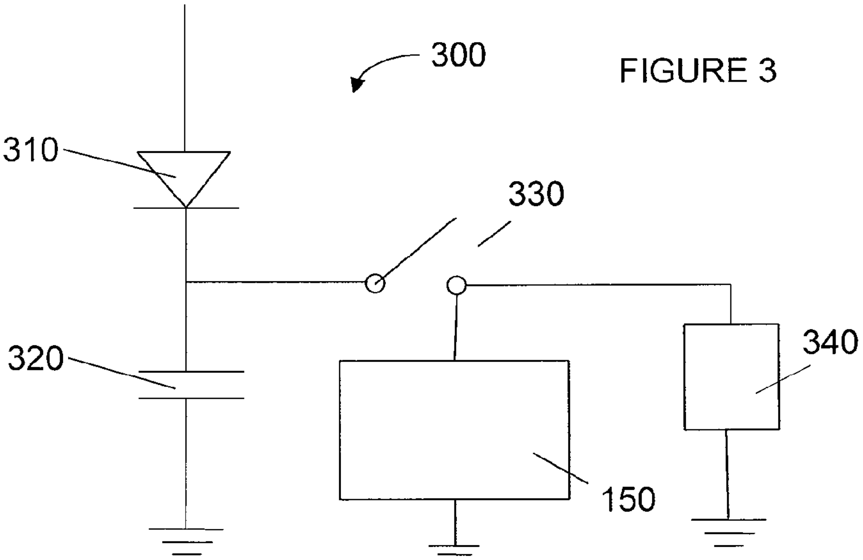
300
310
330
320
340
150

FIGURE 3

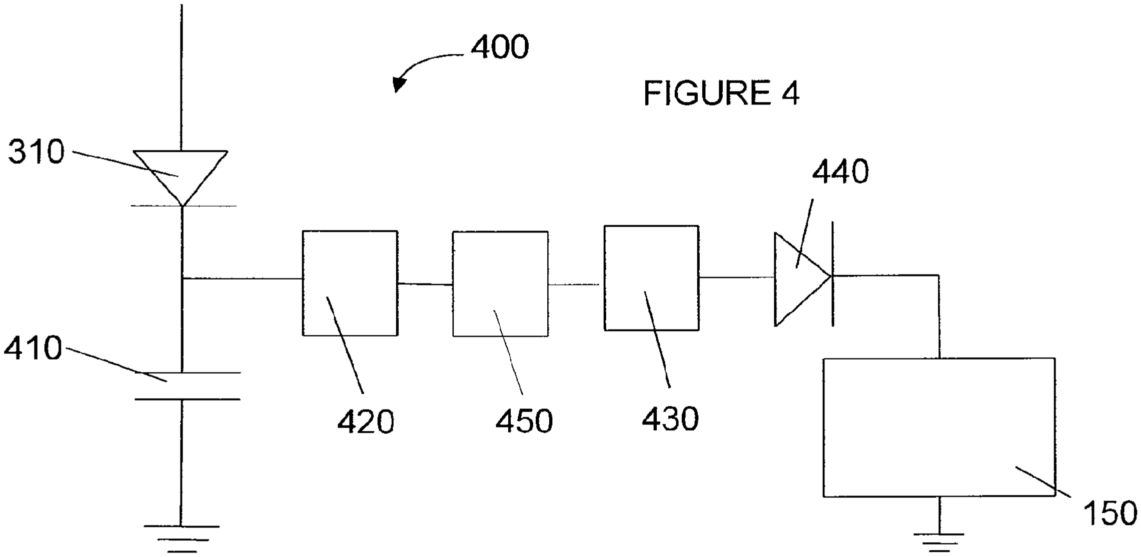
400
310
440
410
420
450
430
150

FIGURE 4

FIGURE 5

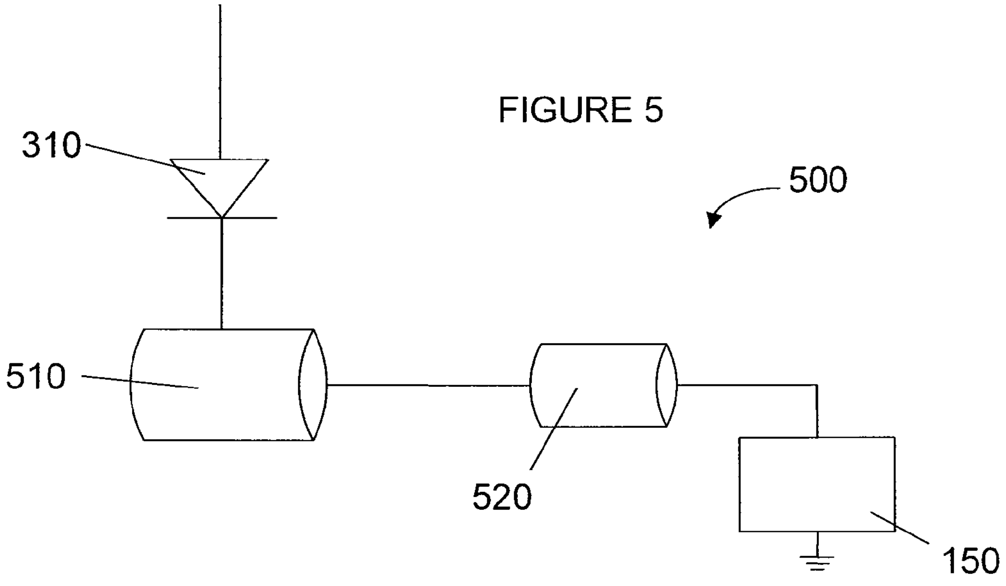
310
500
510
520
150

FIGURE 6

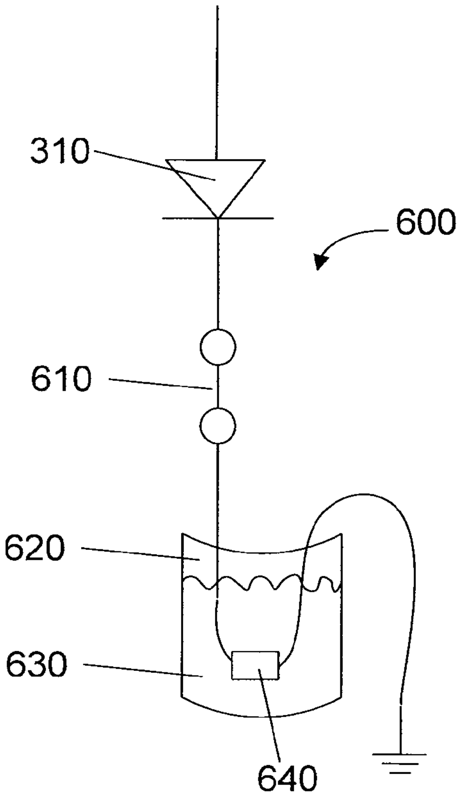
310
600
610
620
630
640

FIGURE 7

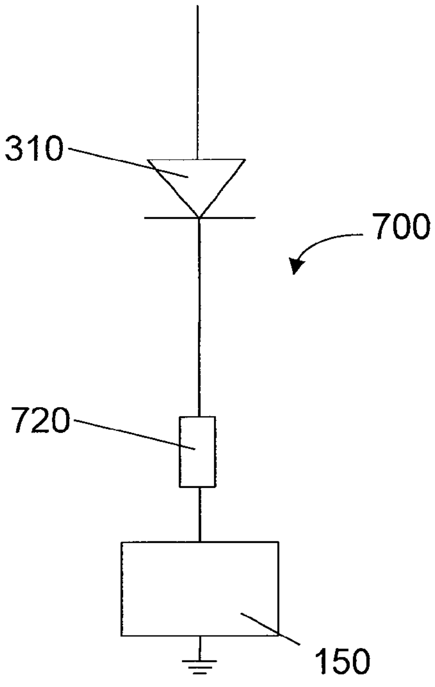
310
700
720
150

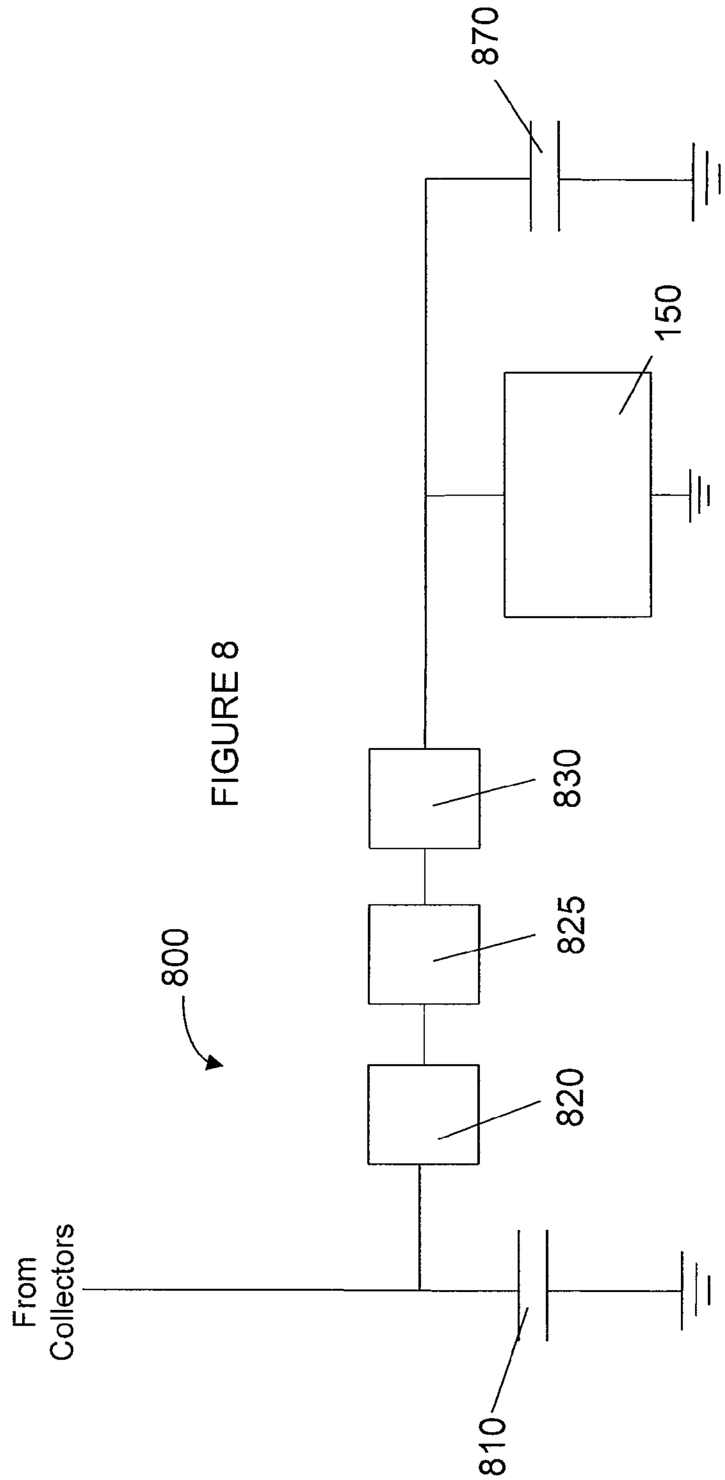
From Collectors
810
820
825
830
800
150
870

FIGURE 8

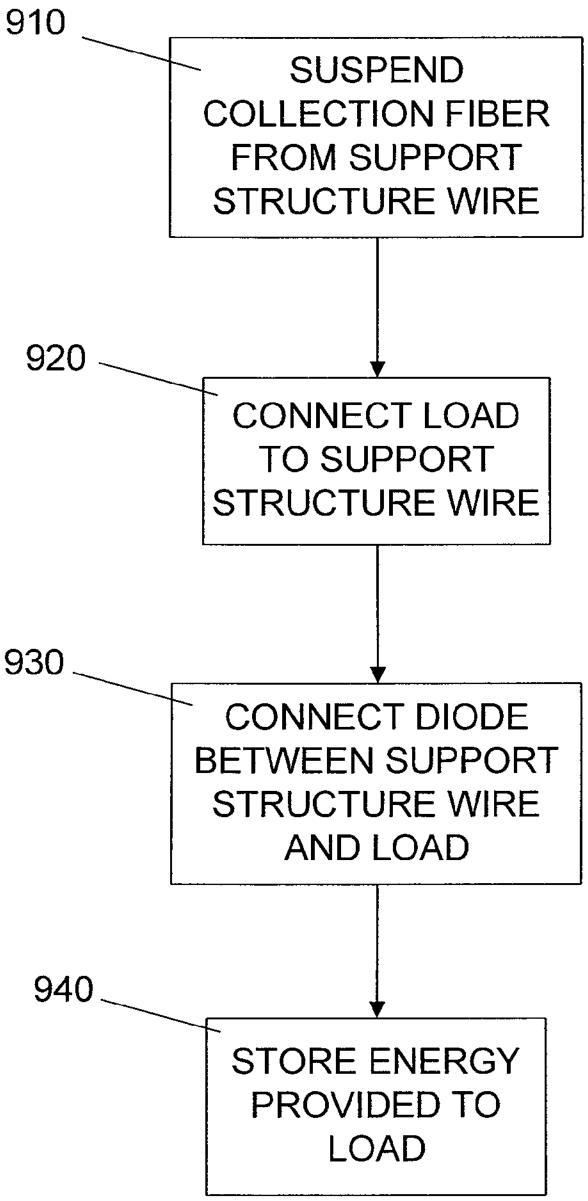
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

ENERGY COLLECTION

CROSS REFERENCE TO RELATED APPLICATION

This application is a continuation application of U.S. patent application Ser. No. 11/358,264, filed on Feb. 21, 2006, which is incorporated by reference herein.

TECHNICAL FIELD

The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure wire elevated above a ground level, at least one collection fiber electrically connected to the support structure wire; a load electrically connected to the support structure wire; and a diode electrically connected between the load and at least one collection fiber.

Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection fiber from a support structure wire elevated above ground level, the fiber electrically connected to the support structure wire; providing a load with an electrical connection to the support structure wire to draw current; and providing a diode electrically connected between the collection fiber and the load.

Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. **1** is a circuit diagram of a weather energy circuit.

FIG. **2** is a perspective view of an exemplary embodiment of many energy collectors elevated above ground by a structure.

FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

FIG. **2**B is a side view of an exemplary embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

FIG. **2**D is a side view of an exemplary embodiment of a support structure for multiple energy collection fibers.

FIG. **2**E is a side view of a support structure for an energy collection fiber.

FIG. **2**F is a side view of an exemplary embodiment of a support structure for an energy collection fiber.

FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

FIG. **3** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **4** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **5** is a circuit diagram of an exemplary embodiment of an energy collection circuit for powering a generator and motor.

FIG. **6** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

FIG. **7** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

FIG. **8** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy.

FIG. **9** is a flow diagram of an exemplary embodiment of collecting energy with a collection fiber.

DETAILED DESCRIPTION

Electric charges on conductors reside entirely on the external surface of the conductors, and tend to concentrate more

around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An exemplary embodiment of this disclosure takes advantage of this property, among others, to collect and use the energy generated by an electric field in the atmosphere. Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An exemplary embodiment utilizes carbon or graphite fibers for static electricity collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including aluminum or steel, but most notably, copper. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

An exemplary embodiment of the collection fibers as collection device **130** includes graphite or carbon fibers. Graphite and carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, or graphite fibers are examples that demonstrate exemplary collection ability.

In at least one exemplary embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than 100 volts per meter under some conditions. When support wire **120** is suspended in the air at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An exemplary embodiment of the collection device comprises the diode with no collection fiber. A preferred embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restricts energy that may be collected through one fiber from escaping through another fiber.

Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2A-2G** using a collection fiber embodiment. FIG. **2A** presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**.

Likewise, FIG. **2B** shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

FIG. **2C** presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

FIG. **2D** presents another exemplary embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2E** provides another exemplary embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

FIG. **2F** presents another exemplary embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2G** provides two supports as provided in FIG. **2F** such that at least two support members **292**, **294** may be connected

to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and between multiple support structures. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**.

FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load. Capacitor **320** may be connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

In an exemplary embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel. Non-limiting examples of spark gap **420** include mercury-reed switches and mercury-wetted reed switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground. The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes may be placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches and mercury-wetted reed switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel.

FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit. Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

FIG. **8** presents exemplary circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches and mercury-wetted reed switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

It should be emphasized that the above-described embodiments of the present disclosure, particularly, any "preferred" embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A method of collecting energy comprising:

suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to the support structure;

providing a load with an electrical connection to the at least one collection device to passively draw current; and

powering a fuel cell between the support structure and the load, the powering performed with energy collected from the at least one collection device.

2. The method of claim **1**, wherein the collection device comprises a diode.

3. The method of claim **1**, wherein the collection device comprises a collection fiber.

4. The method of claim **1**, wherein the collection device comprises a diode and a collection fiber and the diode is electrically connected between the collection fiber and the load.

5. The method of claim **1**, further comprising storing energy provided to the load.

6. The method of claim **5**, wherein storing energy provided to the load comprises storing energy in a capacitor or an inductor.

7. The method of claim **3**, wherein the collection fiber comprises carbon fiber or graphite fiber.

8. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the collection device electrically connected to and suspended from the support structure, the at least one collection device configured to passively draw current;

a load electrically connected to the at least one collection device; and

a diode electrically connected between the at least one collection device and the support structure.

9. The system of claim **8**, wherein the collection device comprises a diode.

10. The system of claim **8**, wherein the collection device comprises a collection fiber.

11. The system of claim **8**, wherein the collection device comprises a collection fiber and a diode electrically connected between the load and the collection fiber.

12. The system of claim **9**, wherein the diode is elevated relative to the ground level.

13. The system of claim **10**, wherein the collection fiber comprises a carbon fiber or a graphite fiber.

14. The system of claim **8**, further comprising:

a switch connected in series between the at least one collection device and the load; and

a capacitor connected in parallel with the switch and the load.

15. The system of claim **14**, wherein the switch comprises an interrupter connected between the at least one collection device and the load, and wherein the interrupter comprises at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

16. The system of claim **15**, further comprising a transformer connected between the interrupter and the load.

17. The system of claim **8**, further comprising:

a motor for providing power, the motor connected between the at least one collection device and the load; and

a generator powered by the motor.

18. The system of claim **8**, further comprising a fuel cell between the support structure and the load.

19. The system of claim **18**, wherein the fuel cell produces hydrogen and oxygen.

20. A system of collecting energy comprising:

means for suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to the means for suspending;

means for passively inducing current flow, the means for inducing current flow electrically connected to the means for suspending; and

means for restricting the backflow of charge carriers, the means for restricting the backflow of charge carriers electrically connected between the at least one collection device and the means for inducing current flow.

21. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the collection device electrically connected to and suspended from the support structure, the at least one collection device configured to passively draw current;

a load electrically connected to the at least one collection device; and

a switch connected in series between the at least one collection device and the load, the switch comprising an interrupter connected between the at least one collection device and the load, and wherein the interrupter comprises at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

22. The system of claim **21**, further comprising a transformer connected between the interrupter and the load.

23. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to and suspended from the support structure, the collection device configured to passively draw current;

a load electrically connected to the at least one collection device;

a motor for providing power, the motor connected between the at least one collection device and the load; and

a generator powered by the motor.

24. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to and suspended from the support structure the at least one collection device configured to passively draw current;

a load electrically connected to the at least one collection device; and

a fuel cell between the support structure and the load.

25. The system of claim **24**, wherein the fuel cell produces hydrogen and oxygen.

* * * * *

US008810049B2

(12) **United States Patent**
McCowen

(10) **Patent No.: US 8,810,049 B2**
(45) **Date of Patent: Aug. 19, 2014**

---

(54) **ENERGY COLLECTION**

(75) Inventor: **Clint McCowen**, Navarre, FL (US)

(73) Assignee: **Ion Power Group, LLC**, Navarre, FL (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 88 days.

(21) Appl. No.: **13/569,133**

(22) Filed: **Aug. 7, 2012**

(65) **Prior Publication Data**

US 2012/0299559 A1 Nov. 29, 2012

Related U.S. Application Data

(63) Continuation-in-part of application No. 12/255,130, filed on Oct. 21, 2008, now Pat. No. 8,686,575, which is a continuation of application No. 11/358,264, filed on Feb. 21, 2006, now Pat. No. 7,439,712.

(51) **Int. Cl.**

F02B 63/04	(2006.01)
F03G 7/08	(2006.01)
H02K 7/18	(2006.01)
F03D 9/00	(2006.01)
H02P 9/04	(2006.01)

(52) **U.S. Cl.**
USPC .... **290/1 R**; 290/44; 290/55; 310/300; 310/310; 416/95

(58) **Field of Classification Search**
CPC Y02E 10/721; Y02E 10/70; Y02E 10/722; B32B 2262/101; C01B 31/00
USPC .... 290/1 R, 44, 55
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427	A		5/1901	Palencsar	307/149
911,260	A		2/1909	Pennock	307/149
1,014,719	A		1/1912	Pennock	307/149
2,473,819	A		6/1949	Pittman	315/36
3,532,959	A		10/1970	Erickson	320/166
3,780,722	A		12/1973	Swet	126/680
3,946,227	A		3/1976	Bingham	250/281
4,079,225	A		3/1978	Warner	219/110
4,092,250	A		5/1978	Sano et al.	210/500.29
4,104,696	A		8/1978	Cochran, Jr.	361/229
4,146,800	A	*	3/1979	Gregory et al.	290/44
4,180,056	A		12/1979	Schnabel et al.	126/660
4,201,197	A	*	5/1980	Dismer	126/600
4,206,396	A	*	6/1980	Marks	322/2 A
4,224,496	A		9/1980	Riordan et al.	219/110
4,307,936	A		12/1981	Ochiai	385/120

(Continued)

FOREIGN PATENT DOCUMENTS

WO	2007098341	8/2007

OTHER PUBLICATIONS

U.S. Appl. No. 12/637,720, Dec. 14, 2009, McCowen.

Primary Examiner — Pedro J Cuevas
(74) *Attorney, Agent, or Firm* — Benjamin A. Balser; Balser & Grell IP Law

(57) **ABSTRACT**

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support wire system supported by poles. The support wire system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but carbon and graphite are preferred. Diodes may be used to restrict the backflow or loss of energy.

20 Claims, 11 Drawing Sheets

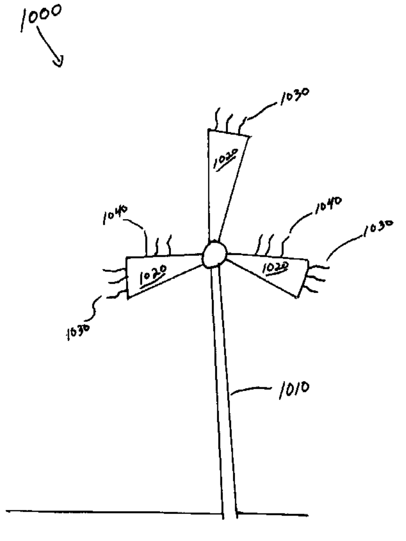

(56) References Cited

U.S. PATENT DOCUMENTS

4,314,192	A	2/1982	Caro	322/2 R
4,346,478	A	8/1982	Sichling	398/107
4,425,905	A	1/1984	Mori	126/578
4,433,248	A	2/1984	Marks	290/44
RE31,678	E	9/1984	Ochiai	385/115
4,483,311	A	11/1984	Whitaker	126/602
4,489,269	A	12/1984	Edling et al.	322/2 R
4,512,335	A	4/1985	Mori	126/678
4,653,223	A	3/1987	Mori	47/17
4,653,472	A	3/1987	Mori	126/683
4,676,226	A	6/1987	Mori	126/680
4,809,675	A	3/1989	Mori	126/578
4,852,454	A	8/1989	Batchelder	89/1.11
4,943,125	A	7/1990	Laundre' et al.	385/35
4,988,159	A	1/1991	Turner et al.	385/33
5,047,892	A	9/1991	Sakata et al.	361/231
5,114,101	A	5/1992	Stern et al.	244/172.8
5,145,257	A	9/1992	Bryant et al.	374/131
5,379,103	A	1/1995	Zigler	356/73
5,851,309	A	12/1998	Kousa	136/248
5,942,806	A	8/1999	Veliadis	290/1 R
6,025,591	A	2/2000	Taylor et al.	250/292
6,038,363	A	3/2000	Slater et al.	385/147
6,116,544	A	9/2000	Forward et al.	244/158.2
6,173,922	B1	1/2001	Hoyt et al.	244/158.2
6,226,440	B1	5/2001	Lyons	385/119
6,419,191	B1	7/2002	Hoyt et al.	244/158.2
6,425,391	B1	7/2002	Davoren et al.	126/683
6,457,943	B1 *	10/2002	Olsen et al.	416/230
6,735,830	B1	5/2004	Merciel	250/324
6,765,212	B2	7/2004	Faus et al.	250/339.11
6,846,447	B2	1/2005	Mbachu et al.	264/406
6,853,447	B2	2/2005	Goetz	356/237.1
6,894,772	B2	5/2005	Goetz et al.	356/237.1
6,925,852	B2	8/2005	Susko	73/23.2
6,974,110	B2	12/2005	Grandics	244/171.5
7,109,597	B1	9/2006	Bose	290/43
7,128,867	B2	10/2006	Mbachu et al.	264/406
7,231,809	B2	6/2007	Susko	73/31.05
7,259,475	B2	8/2007	Hong et al.	307/64
7,303,373	B2 *	12/2007	Viertl	416/1
7,439,712	B2 *	10/2008	McCowen	322/2 A
7,449,668	B2	11/2008	Schutten et al.	250/214 SW
7,478,712	B2 *	1/2009	McCowen	191/45 R
7,532,819	B1	5/2009	Triebes et al.	398/121
7,637,462	B2 *	12/2009	Pal	244/199.2
8,115,333	B2 *	2/2012	Parsche et al.	290/55
8,118,559	B2 *	2/2012	Llorente Gonzalez et al.	416/226
8,342,805	B2 *	1/2013	Mendez Hernandez et al.	416/146 R
8,519,596	B1 *	8/2013	Kim et al.	310/310
8,629,570	B1 *	1/2014	Kamenov	290/44
8,662,853	B2 *	3/2014	Vasudeva et al.	416/226
8,686,575	B2 *	4/2014	McCowen	290/1 R
2002/0026933	A1	3/2002	Gottlieb	126/628
2002/0108892	A1	8/2002	Goetz et al.	209/576
2002/0109094	A1	8/2002	Goetz et al.	250/339.11
2002/0109835	A1	8/2002	Goetz	356/237.1
2003/0125630	A1	7/2003	Furnish	600/461
2003/0193319	A1	10/2003	Wood et al.	322/2 A
2004/0011925	A1	1/2004	Grandics	244/62
2004/0083793	A1	5/2004	Susko	73/31.05
2004/0094853	A1	5/2004	Mbachu et al.	264/40.1
2004/0160711	A1	8/2004	Stumberger	361/20
2004/0212945	A1	10/2004	Sprenger et al.	361/212
2005/0101024	A1	5/2005	Mbachu et al.	436/85
2005/0136311	A1	6/2005	Ueda et al.	700/297
2005/0270525	A1	12/2005	Susko	356/326
2006/0051207	A1 *	3/2006	Becerra et al.	416/5
2007/0098551	A1 *	5/2007	Viertl	416/61
2007/0107765	A1	5/2007	Schutten et al.	136/243
2007/0195481	A1 *	8/2007	McCowen	361/233
2007/0217918	A1 *	9/2007	Baker et al.	416/227 R
2007/0218323	A1	9/2007	Sudo et al.	429/413
2007/0223171	A1	9/2007	Guy et al.	361/126
2007/0252047	A1 *	11/2007	Pal	244/35 R
2007/0273206	A1 *	11/2007	McCowen	307/9.1
2007/0297892	A1	12/2007	Kildegaard	415/1
2008/0145714	A1	6/2008	Kagami	429/413
2009/0040680	A1	2/2009	McCowen	361/233
2009/0107842	A1	4/2009	Park et al.	204/554
2009/0114495	A1	5/2009	McCowen	191/45 R
2009/0216292	A1	8/2009	Pless et al.	607/33
2010/0090562	A1	4/2010	McCowen	310/309
2010/0329881	A1 *	12/2010	Mendez Hernandez et al.	416/230
2012/0134826	A1 *	5/2012	Arocena De La Rua et al.	416/146 R
2012/0312918	A1 *	12/2012	Heppe	244/30
2013/0028739	A1 *	1/2013	Erichsen et al.	416/146 R
2013/0032671	A1 *	2/2013	Giles et al.	244/134 R
2013/0037650	A1 *	2/2013	Heppe	244/2
2013/0136598	A1 *	5/2013	Loewe et al.	416/95
2013/0170992	A1 *	7/2013	Loewe et al.	416/95
2013/0334824	A1 *	12/2013	Freda	290/55
2014/0084748	A1 *	3/2014	Wang et al.	310/300
2014/0086748	A1 *	3/2014	Peltola et al.	416/95
2014/0099208	A1 *	4/2014	Janiuk	416/210 R

* cited by examiner

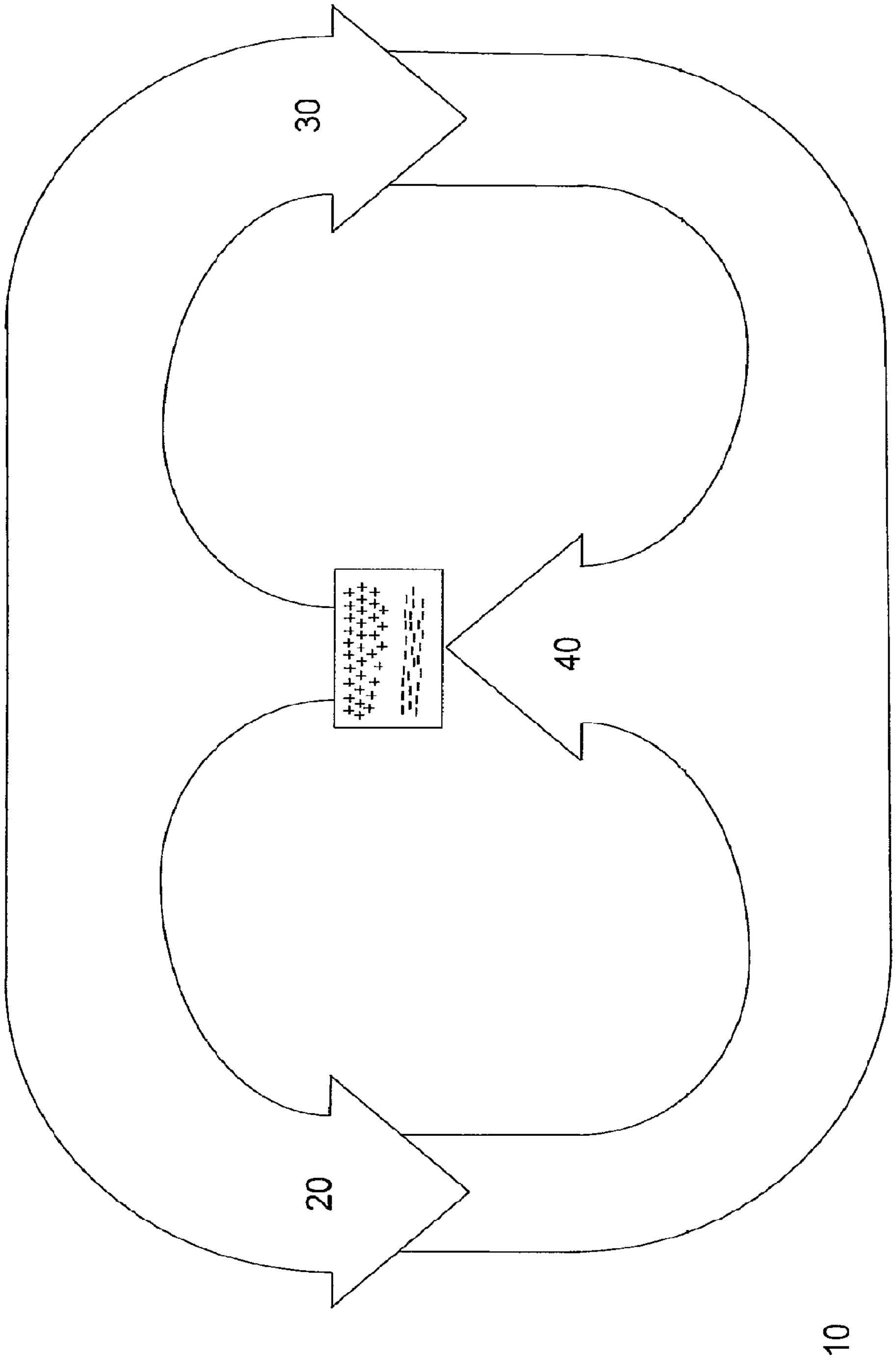

FIGURE 1

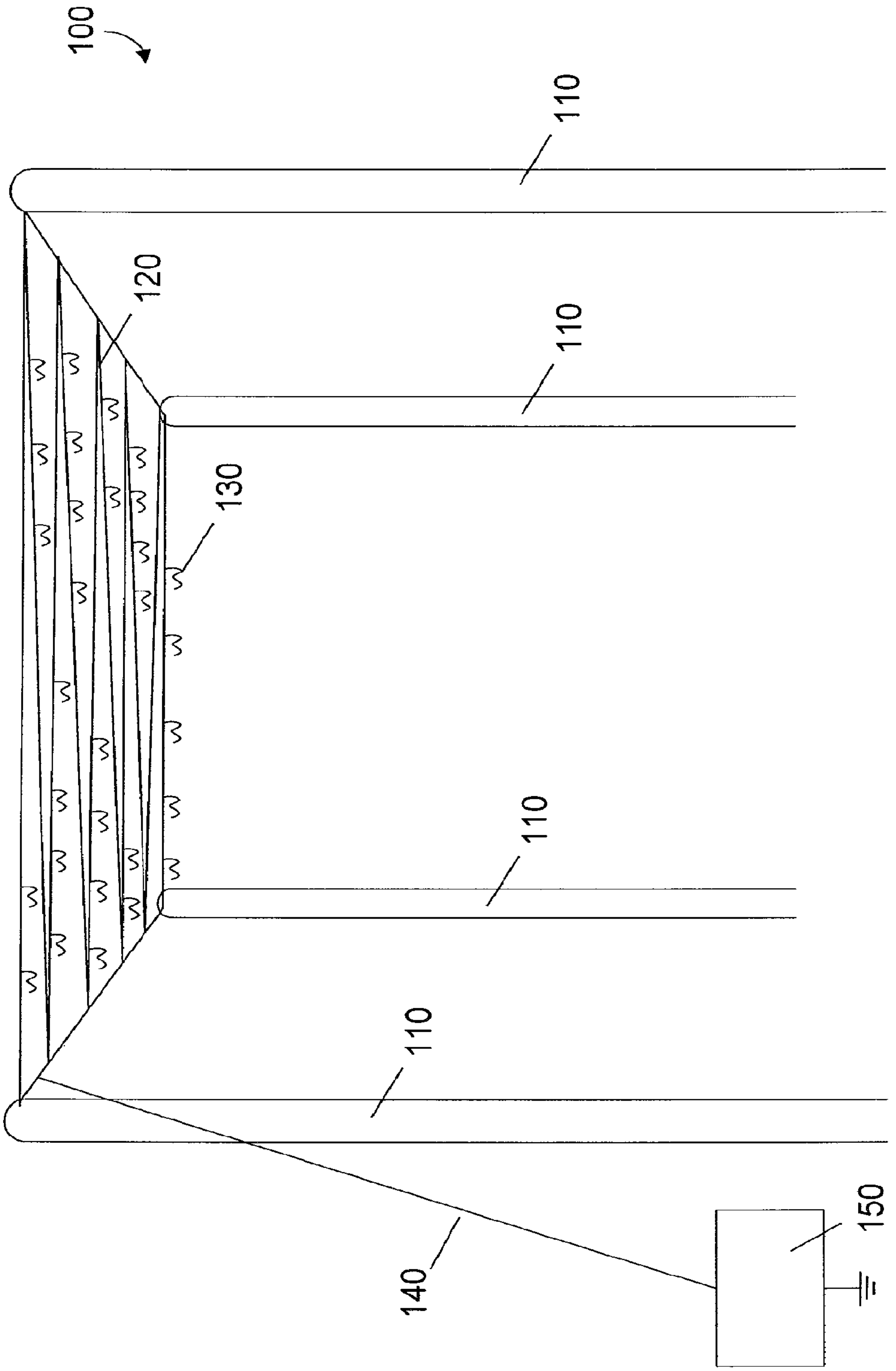
100
110
110
120
130
110
110
140
150

FIGURE 2

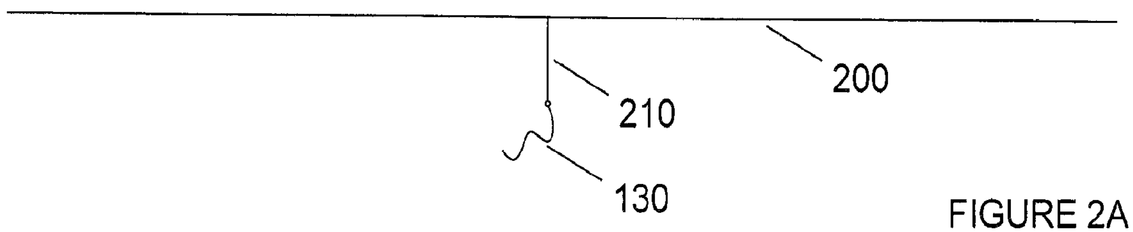

FIGURE 2A

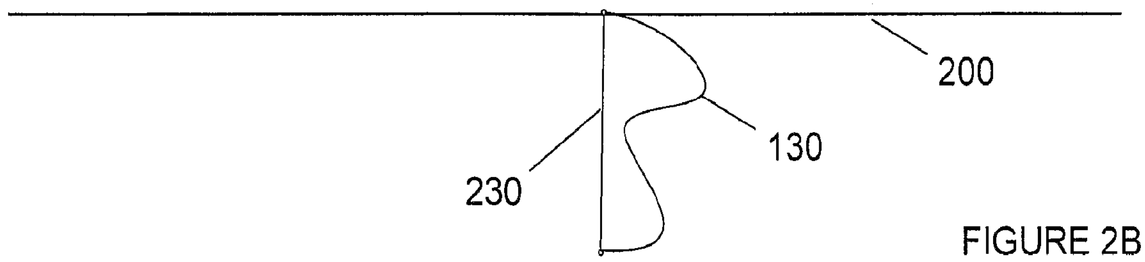

FIGURE 2B

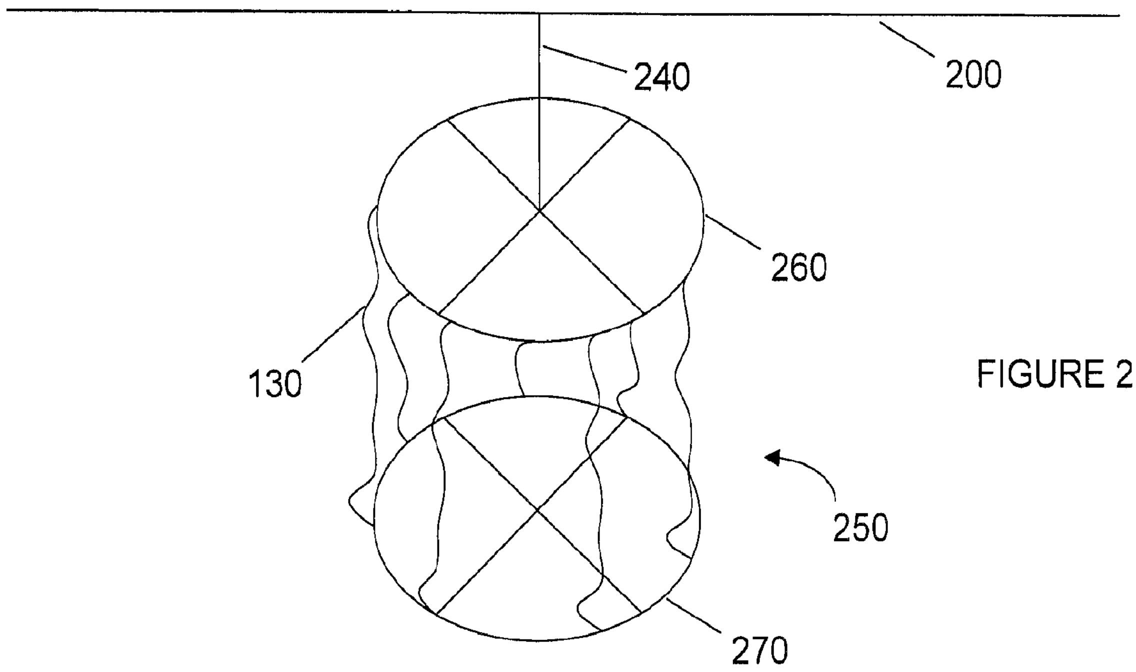
240
200
260
130
250
270

FIGURE 2C

FIGURE 2D

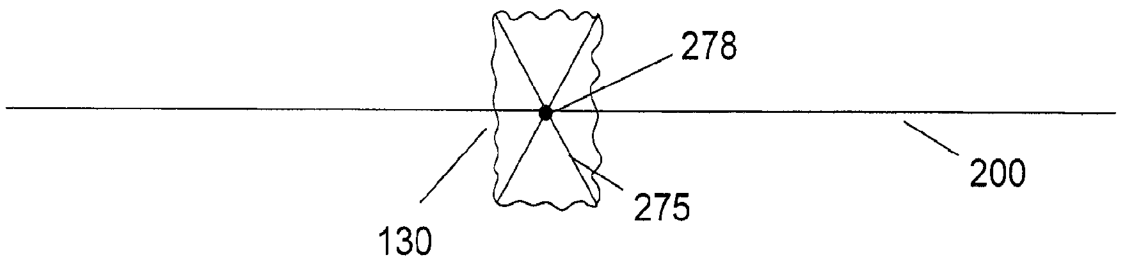
278
200
275
130

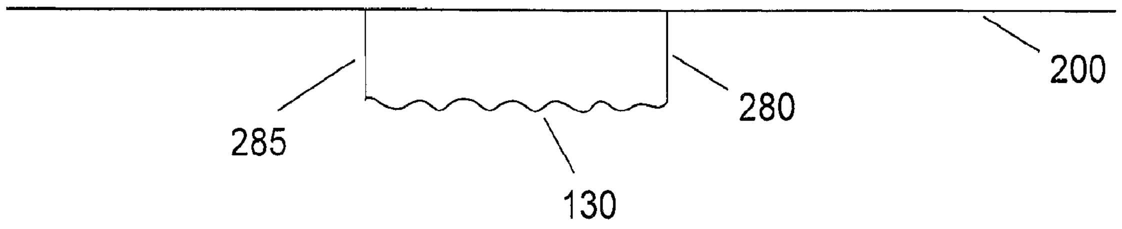
200
280
285
130

FIGURE 2E

FIGURE 2F

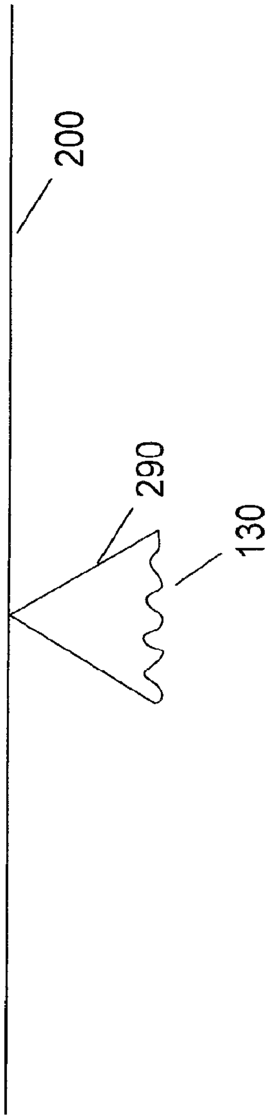

FIGURE 2G

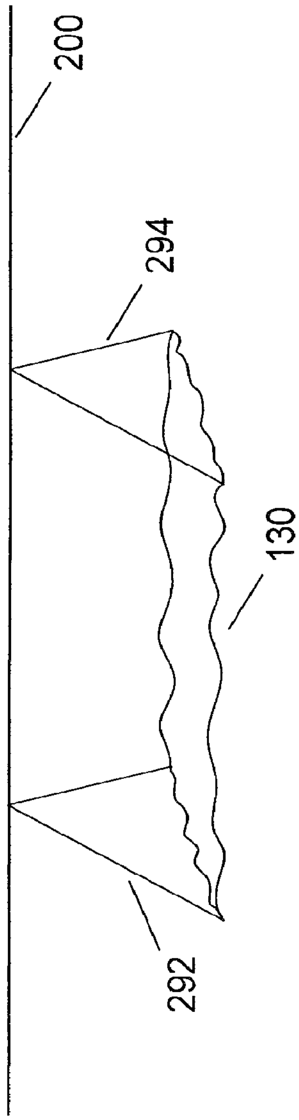
200
294
130
292

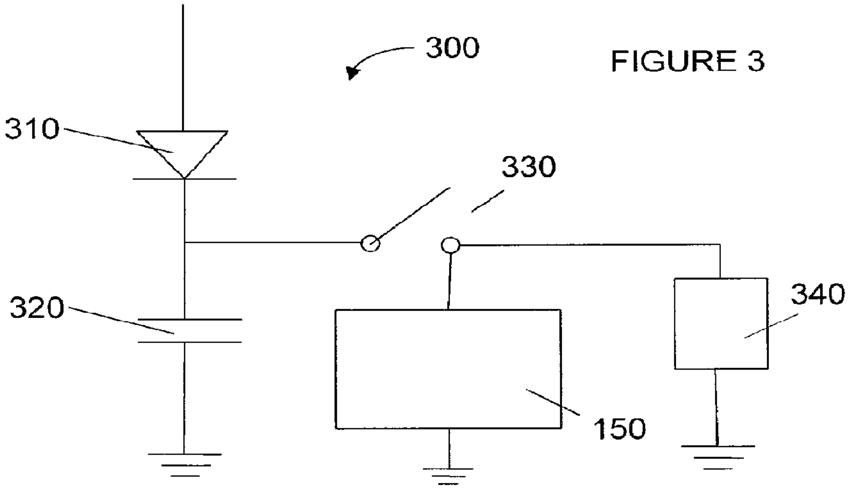
300
310
330
320
340
150

FIGURE 3

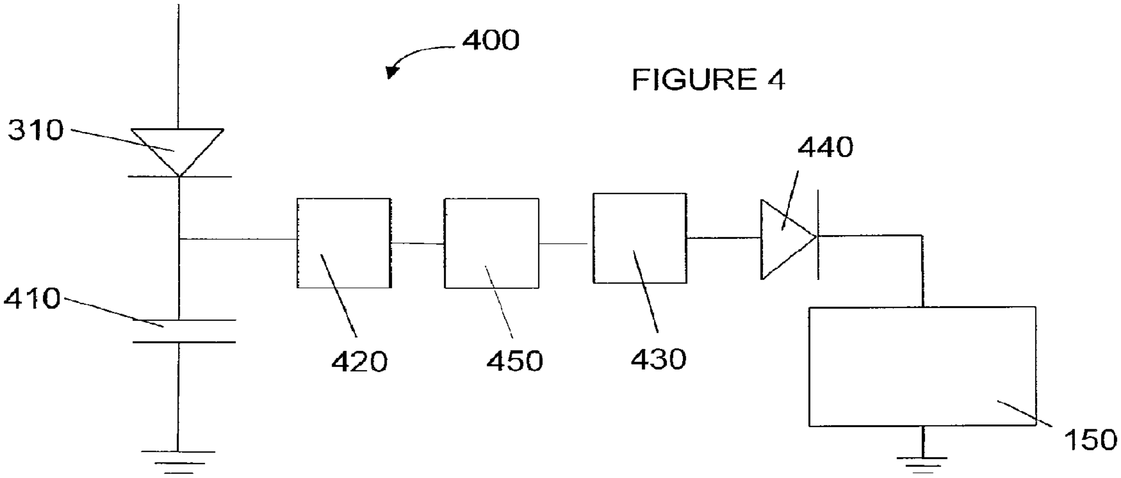
400
310
440
410
420
450
430
150

FIGURE 4

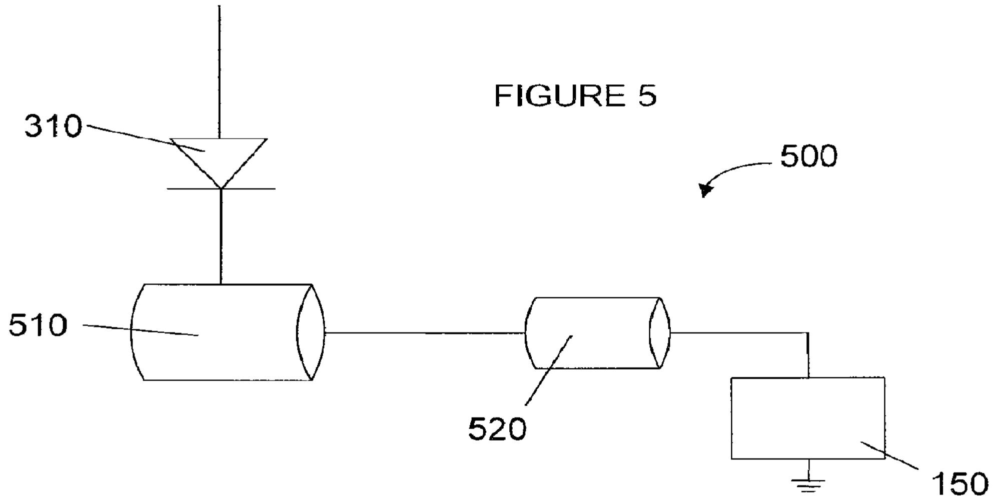
FIGURE 5
310
500
510
520
150

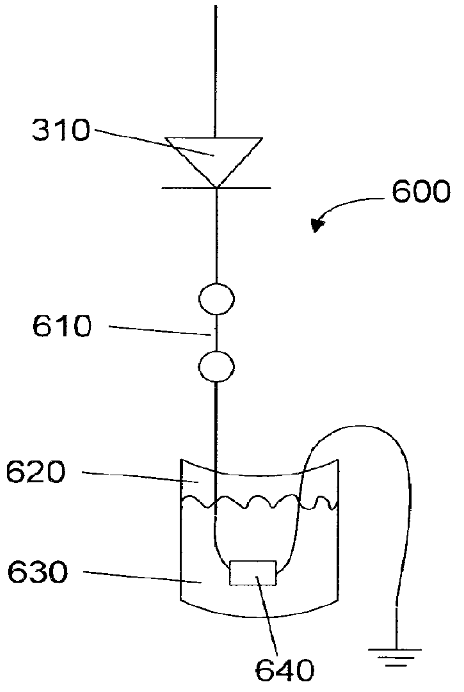
FIGURE 6
310
600
610
620
630
640

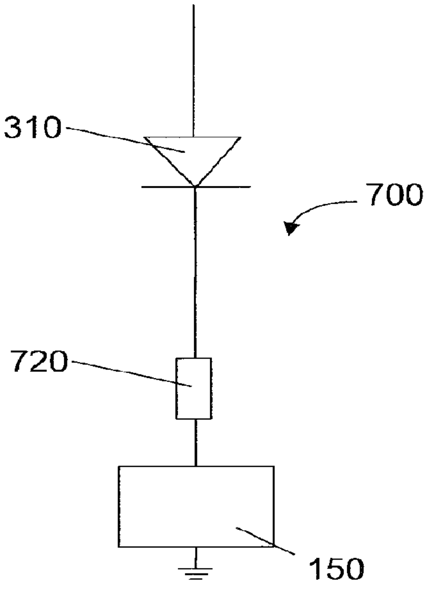
FIGURE 7
310
700
720
150

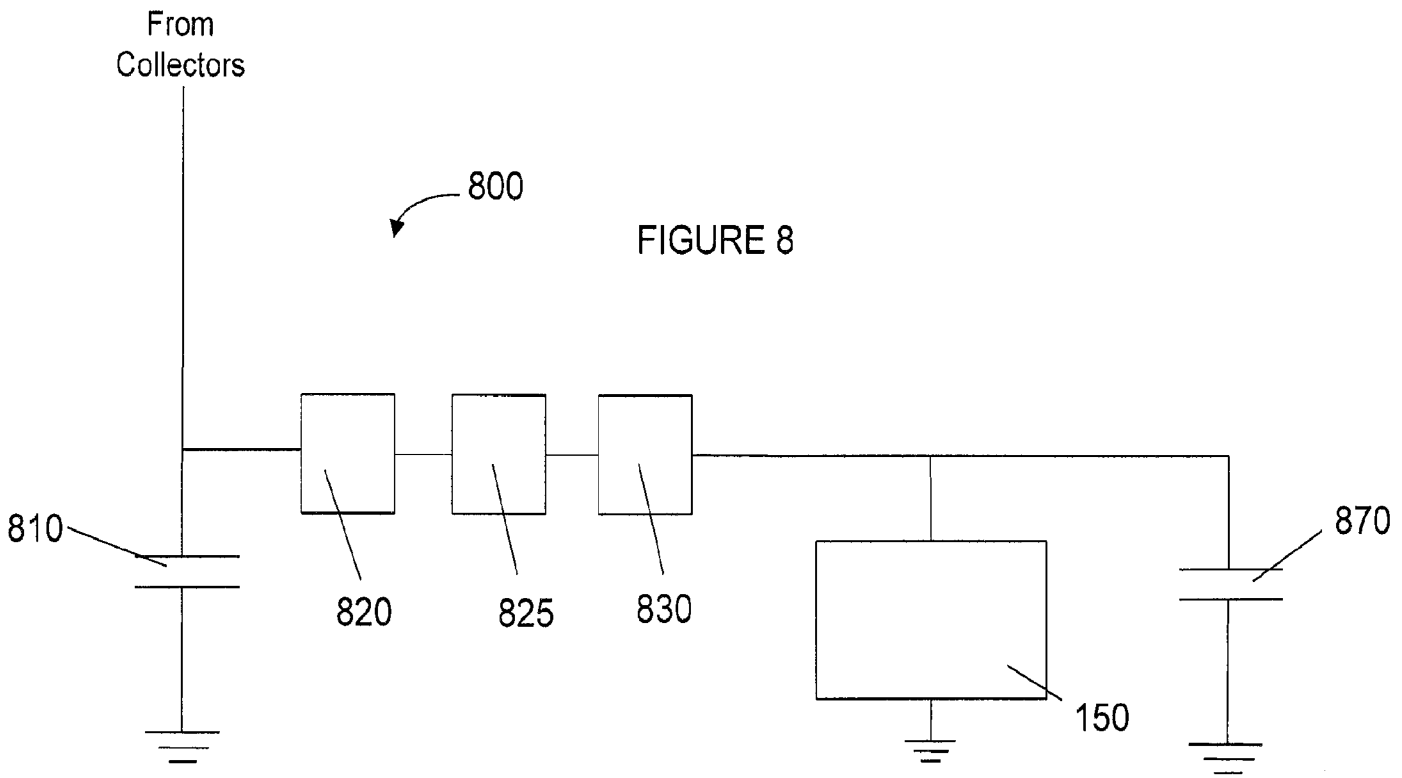
From
Collectors
800
810
820
825
830
150
870

FIGURE 8

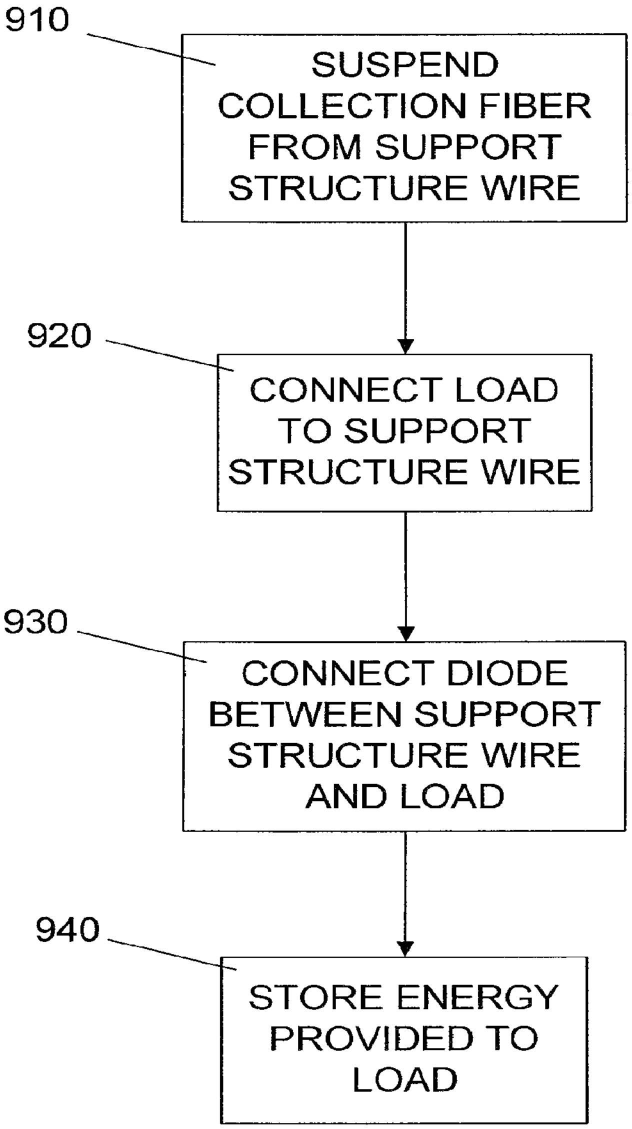
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

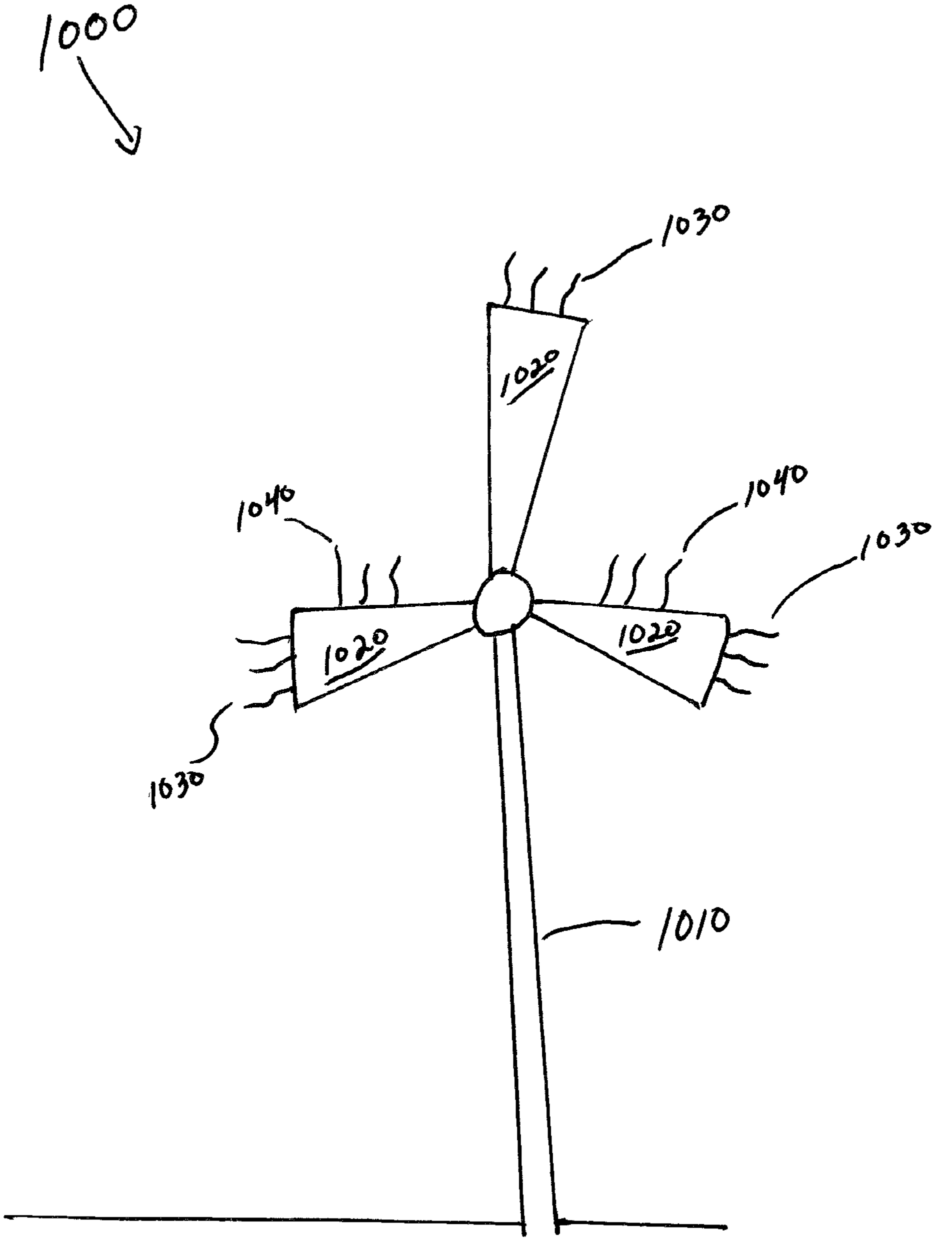
1000
1030
1020
1040
1040
1030
1020
1020
1030
1010

FIG. 10

ENERGY COLLECTION

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation in part application of U.S. patent application Ser. No. 12/255,130 filed on Oct. 21, 2008, which is a continuation application of U.S. patent application Ser. No. 11/358,264, filed on Feb. 21, 2006, which are both incorporated by reference herein.

TECHNICAL FIELD

The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure wire elevated above a ground level, at least one collection fiber electrically connected to the support structure wire; a load electrically connected to the support structure wire; and a diode electrically connected between the load and at least one collection fiber.

Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection fiber from a support structure wire elevated above ground level, the fiber electrically connected to the support structure wire; providing a load with an electrical connection to the support structure wire to draw current; and providing a diode electrically connected between the collection fiber and the load.

Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. **1** is a circuit diagram of a weather energy circuit.

FIG. **2** is a perspective view of an exemplary embodiment of many energy collectors elevated above ground by a structure.

FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

FIG. **2**B is a side view of an exemplary embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

FIG. **2**D is a side view of an exemplary embodiment of a support structure for multiple energy collection fibers.

FIG. **2**E is a side view of a support structure for an energy collection fiber.

FIG. **2**F is a side view of an exemplary embodiment of a support structure for an energy collection fiber.

FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

FIG. **3** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **4** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **5** is a circuit diagram of an exemplary embodiment of an energy collection circuit for powering a generator and motor.

FIG. **6** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

FIG. **7** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

FIG. **8** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy.

FIG. **9** is a flow diagram of an exemplary embodiment of collecting energy with a collection fiber.

FIG. **10** is a system diagram of an exemplary embodiment of a windmill with energy collectors.

DETAILED DESCRIPTION

Electric charges on conductors reside entirely on the external surface of the conductors, and tend to concentrate more around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An exemplary embodiment of this disclosure takes advantage of this property, among others, to collect and use the energy generated by an electric field in the atmosphere. Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An exemplary embodiment utilizes carbon or graphite fibers for static electricity collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including aluminum or steel, but most notably, copper. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

An exemplary embodiment of the collection fibers as collection device **130** includes graphite or carbon fibers. Graphite and carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, or graphite fibers are examples that demonstrate exemplary collection ability.

In at least one exemplary embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than 100 volts per meter under some conditions. When support wire **120** is suspended in the air at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An exemplary embodiment of the collection device comprises the diode with no collection fiber. A preferred embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restricts energy that may be collected through one fiber from escaping through another fiber.

Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2**A-**2**G using a collection fiber embodiment. FIG. **2**A presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**.

Likewise, FIG. **2**B shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

FIG. **2**C presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

FIG. **2**D presents another exemplary embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2**E provides another exemplary embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

FIG. **2**F presents another exemplary embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2**G provides two supports as provided in FIG. **2**F such that at least two support members **292**, **294** may be connected to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and between multiple support structures. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**.

FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load. Capacitor **320** maybe connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

In an exemplary embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel. Non-limiting examples of spark gap **420** include mercury-reed switches and mercury-wetted reed switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground. The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes maybe placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches and mercury-wetted reed switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel.

FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit. Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

FIG. **8** presents exemplary circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches and mercury-wetted reed switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

A windmill is an engine powered by the energy of wind to produce alternative forms of energy. They may, for example, be implemented as small tower mounted wind engines used to pump water on farms. The modern wind power machines used for generating electricity are more properly called wind turbines. Common applications of windmills are grain milling, water pumping, threshing, and saw mills. Over the ages, windmills have evolved into more sophisticated and efficient wind-powered water pumps and electric power generators. In an example embodiment, as provided in FIG. **10**, windmill tower **1000** of suitable height and/or propeller **1020** of windmill tower **1000** may be equipped with energy collecting fibers **1030**, **1040**. Collecting fibers **1030**, **1040** may turn windmill **1000** into a power producing asset even when there is not enough wind to turn propellers **1020**. During periods when there is enough wind to turn propellers **1020**, collecting fibers **1030**, **1040** may supplement/boost the amount of energy the windmill produces.

Windmill **1000**, properly equipped with ion collectors **1030**, **1040**, such as non-limiting example carbon fibers, can produce electricity: 1) by virtue of providing altitude to the carbon fiber to harvest ions, and 2) while the propeller is turning, by virtue of wind blowing over the carbon fiber producing electricity, among other reasons, via the triboelectric effect (however, it is also possible for the triboelectric effect to occur, producing electricity, in winds too weak to turn the propeller).

There are at least two ways that energy collectors may be employed on or in a windmill propeller to harvest energy. Propellers **1020** may be equipped with energy collectors **1030**, **1040** attached to, or supported by, propeller **1020** with wires (or metal embedded in, or on propeller **1020**) electrically connecting energy collectors **1030**, **1040** to a load or power conversion circuit. There may be a requirement to electrically isolate energy collectors **1030**, **1040**, which is added to propeller **1020**, from electrical ground, so that the energy collected does not short to ground through propeller **1020** itself or through support tower **1010**, but rather is conveyed to the load or power conversion circuit. Energy collectors may be connected to the end of propellers **1020** such as collectors **1030**. Alternatively, energy collectors may be connected to the sides of propellers **1020** such as collectors **1040**.

Alternatively, propeller **1020** may be constructed of carbon fiber or other suitable material, with wires (or the structural metal supporting propeller **1020** may be used) electrically connecting to a load or power conversion circuit. In the case of propeller **1020** itself being constructed of carbon fiber, for example, the fiber may be 'rough finished' in selected areas so that the fiber is "fuzzy." For example, small portions of it may protrude into the air as a means of enhancing collection efficiency. The fuzzy parts of collectors **1030**, **1040** may do much of the collecting. There may be a requirement to electrically isolate carbon fiber propeller **1020** from electrical ground, so that the energy it collects does not short to ground through metal support tower **1010**, but rather is conveyed to the load or power conversion circuit. Diodes may be implemented within the circuit to prevent the backflow of energy, although diodes may not be necessary in some applications.

In an alternative embodiment, windmill **1000** may be used as a base on which to secure an even higher extension tower to support the energy collectors and/or horizontal supports extending out from tower **1010** to support the energy collectors. Electrical energy may be generated via ion collection due to altitude and also when a breeze or wind blows over the collectors supported by tower **1010**.

Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

It should be emphasized that the above-described embodiments of the present disclosure, particularly, any "preferred" embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A method of collecting energy comprising:

suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment from a windmill, the at least one collection device electrically connected to the windmill; and

providing a load with an electrical connection to the at least one collection device to draw current.

2. The method of claim **1**, wherein the collection device collects energy by triboelectric effect.

3. The method of claim **1**, wherein the collection device comprises a collection fiber.

4. The method of claim **1**, wherein the collection device comprises a diode and a collection fiber and the diode is electrically connected between the collection fiber and the load.

5. The method of claim **1**, further comprising storing energy provided to the load.

6. The method of claim **5**, wherein storing energy provided to the load comprises storing energy in a capacitor or an inductor.

7. The method of claim **3**, wherein the collection fiber comprises carbon fiber or graphite fiber.

8. A system of energy collection comprising:

a windmill;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment electrically connected to the windmill; and

a load electrically connected to the at least one collection device.

9. The system of claim **8**, wherein the collection device collects energy by triboelectric effect.

10. The system of claim **8**, wherein the collection device comprises a collection fiber.

11. The system of claim **8**, wherein the collection device comprises a collection fiber and a diode electrically connected between the load and the collection fiber.

12. The system of claim **9**, wherein the diode is elevated relative to the ground level.

13. The system of claim **10**, wherein the collection fiber comprises a carbon fiber or a graphite fiber.

14. The system of claim **8**, further comprising a diode electrically connected between the at least one collection device and the support structure.

15. The system of claim **8**, further comprising: a switch connected in series between the at least one collection device and the load; and a capacitor connected in parallel with the switch and the load.

16. The system of claim **15**, wherein the switch comprises an interrupter connected between the plurality at least one collection device, and wherein the interrupter comprises at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

17. The system of claim **8**, further comprising: a motor for providing power, the motor connected between the at least one collection device and the load; and a generator powered by the motor.

18. The system of claim **8**, further comprising a fuel cell between the support structure and the load.

19. The system of claim **18**, wherein the fuel cell produces hydrogen and oxygen.

20. A system of collecting energy comprising:

means for suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment from a windmill structure, the at least one collection device electrically connected to the means for suspending;

means for inducing current flow, the means for inducing current flow electrically connected to the means for suspending; and

means for restricting the backflow of charge carriers, the means for restricting the backflow of charge carriers electrically connected between the at least one collection device and the means for inducing current flow.

* * * * *

US009479086B2

(12) United States Patent
McCowen

(10) **Patent No.: US 9,479,086 B2**
(45) **Date of Patent: *Oct. 25, 2016**

(54) **ENERGY COLLECTION**

(71) Applicant: **Ion Power Group LLC**, Navarre, FL (US)

(72) Inventor: **Clint McCowen**, Navarre, FL (US)

(73) Assignee: **Ion Power Group, LLC**, Navarre, FL (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **14/242,464**

(22) Filed: **Apr. 1, 2014**

(65) **Prior Publication Data**

US 2014/0210308 A1 Jul. 31, 2014

Related U.S. Application Data

(63) Continuation of application No. 12/255,130, filed on Oct. 21, 2008, now Pat. No. 8,686,575, which is a continuation of application No. 11/358,264, filed on Feb. 21, 2006, now Pat. No. 7,439,712.

(51) **Int. Cl.**

F02B 63/04	(2006.01)
F03G 7/08	(2006.01)
H02K 7/18	(2006.01)
H02N 1/10	(2006.01)
H02N 11/00	(2006.01)
H01M 16/00	(2006.01)

(52) **U.S. Cl.**

CPC ***H02N 1/10*** (2013.01); ***H02N 11/002*** (2013.01); *H01M 16/003* (2013.01)

(58) **Field of Classification Search**

USPC .... 290/1 R

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427 A		5/1901	Palencsar	
911,260 A		2/1909	Pennock	
1,014,719 A		1/1912	Pennock	
2,473,819 A		6/1949	Pittman	
3,532,959 A		10/1970	Erickson	
3,663,360 A	*	5/1972	Post	376/147
3,668,065 A	*	6/1972	Moir	376/147
3,780,722 A		12/1973	Swet	
3,946,227 A		3/1976	Bingham	

(Continued)

FOREIGN PATENT DOCUMENTS

CA	2647385	9/2008
EP	1999767 A2	12/2008

(Continued)

OTHER PUBLICATIONS

"Emissivity Values for Common Materials & Non-Metals", Raytek; http://raytek.com/Raytek/en-r0/IREducation/EmissivityNonMetals.htm; 1999.*

Primary Examiner — Pedro J Cuevas

(74) *Attorney, Agent, or Firm* — Benjamin A. Balser; Next ip Law Group

(57) **ABSTRACT**

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support wire system supported by poles. The support wire system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but carbon and graphite are preferred. Diodes may be used to restrict the backflow or loss of energy.

21 Claims, 10 Drawing Sheets

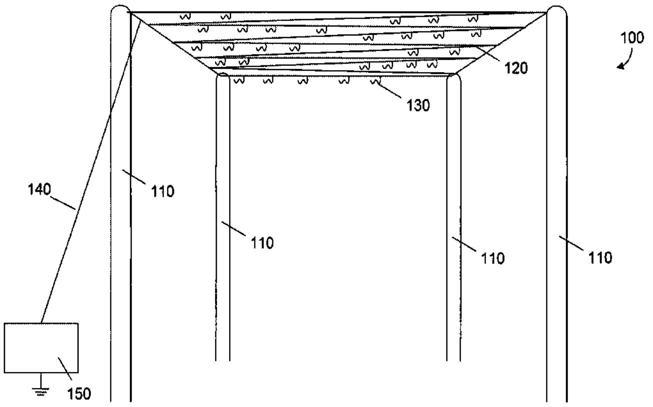

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,079,225	A	3/1978	Warner
4,092,250	A	5/1978	Sano et al.
4,104,696	A	8/1978	Cochran, Jr.
4,146,800	A	3/1979	Gregory et al.
4,180,056	A	12/1979	Schnabel et al.
4,201,197	A	5/1980	Dismer
4,206,396	A	6/1980	Marks
4,224,496	A	9/1980	Riordan et al.
4,307,936	A	12/1981	Ochiai
4,314,192	A	2/1982	Caro
4,346,478	A	8/1982	Sichling
4,425,905	A	1/1984	Mori
4,433,248	A	2/1984	Marks
RE31,678	E	9/1984	Ochiai
4,483,311	A	11/1984	Whitaker
4,489,269	A	12/1984	Edling et al.
4,512,335	A	4/1985	Mori
4,653,223	A	3/1987	Mori
4,653,472	A	3/1987	Mori
4,676,226	A	6/1987	Mori
4,809,675	A	3/1989	Mori
4,852,454	A	8/1989	Batchelder
4,943,125	A	7/1990	Laundre' et al.
4,988,159	A	1/1991	Turner et al.
5,047,892	A	9/1991	Sakata et al.
5,114,101	A	5/1992	Stern et al.
5,145,257	A	9/1992	Bryant et al.
5,379,103	A	1/1995	Zigler
5,851,309	A	12/1998	Kousa
5,942,806	A	8/1999	Veliadis
6,025,591	A	2/2000	Taylor et al.
6,038,363	A	3/2000	Slater et al.
6,116,544	A	9/2000	Forward et al.
6,173,922	B1	1/2001	Hoyt et al.
6,226,440	B1	5/2001	Lyons
6,419,191	B1	7/2002	Hoyt et al.
6,425,391	B1	7/2002	Davoren et al.
6,735,830	B1	5/2004	Merciel
6,765,212	B2	7/2004	Goetz et al.
6,846,447	B2	1/2005	Mbachu et al.
6,853,447	B2	2/2005	Goetz
6,894,772	B2	5/2005	Goetz et al.
6,925,852	B2	8/2005	Susko
6,974,110	B2	12/2005	Grandics
7,109,597	B1	9/2006	Bose
7,128,867	B2	10/2006	Mbachu et al.
7,231,809	B2	6/2007	Susko
7,259,475	B2	8/2007	Hong et al.
7,439,712	B2	10/2008	McCowen
7,449,668	B2	11/2008	Schutten et al.
7,478,712	B2	1/2009	McCowen
7,532,819	B1	5/2009	Triebes et al.
8,269,401	B1	9/2012	Kim et al.
8,686,575	B2	4/2014	McCowen
8,810,049	B2 *	8/2014	McCowen .... 290/1 R
2002/0026933	A1	3/2002	Gottlieb
2002/0108892	A1	8/2002	Goetz et al.
2002/0109094	A1	8/2002	Goetz et al.
2002/0109835	A1	8/2002	Goetz
2003/0125630	A1	7/2003	Furnish
2003/0193319	A1	10/2003	Wood et al.
2004/0011925	A1	1/2004	Grandics
2004/0083793	A1	5/2004	Susko
2004/0094853	A1	5/2004	Mbachu et al.
2004/0160711	A1	8/2004	Stumberger
2004/0212945	A1	10/2004	Sprenger et al.
2005/0101024	A1	5/2005	Mbachu et al.
2005/0136311	A1	6/2005	Ueda et al.
2005/0270525	A1	12/2005	Susko
2007/0107765	A1	5/2007	Schutten et al.
2007/0195481	A1	8/2007	McCowen
2007/0218323	A1	9/2007	Sudo et al.
2007/0223171	A1	9/2007	Guy et al.
2007/0273206	A1	11/2007	McCowen
2007/0297892	A1	12/2007	Kildegaard
2008/0145714	A1	6/2008	Kagami
2009/0040680	A1	2/2009	McCowen
2009/0107842	A1	4/2009	Park et al.
2009/0114495	A1	5/2009	McCowen
2009/0216292	A1	8/2009	Pless et al.
2010/0090562	A1	4/2010	McCowen
2011/0148248	A1	6/2011	Landa
2012/0043585	A1	2/2012	Guo et al.
2012/0299559	A1	11/2012	McCowen
2013/0049531	A1	2/2013	Wang et al.

FOREIGN PATENT DOCUMENTS

RU	2008137622	2/2007
WO	2007098341	8/2007

* cited by examiner

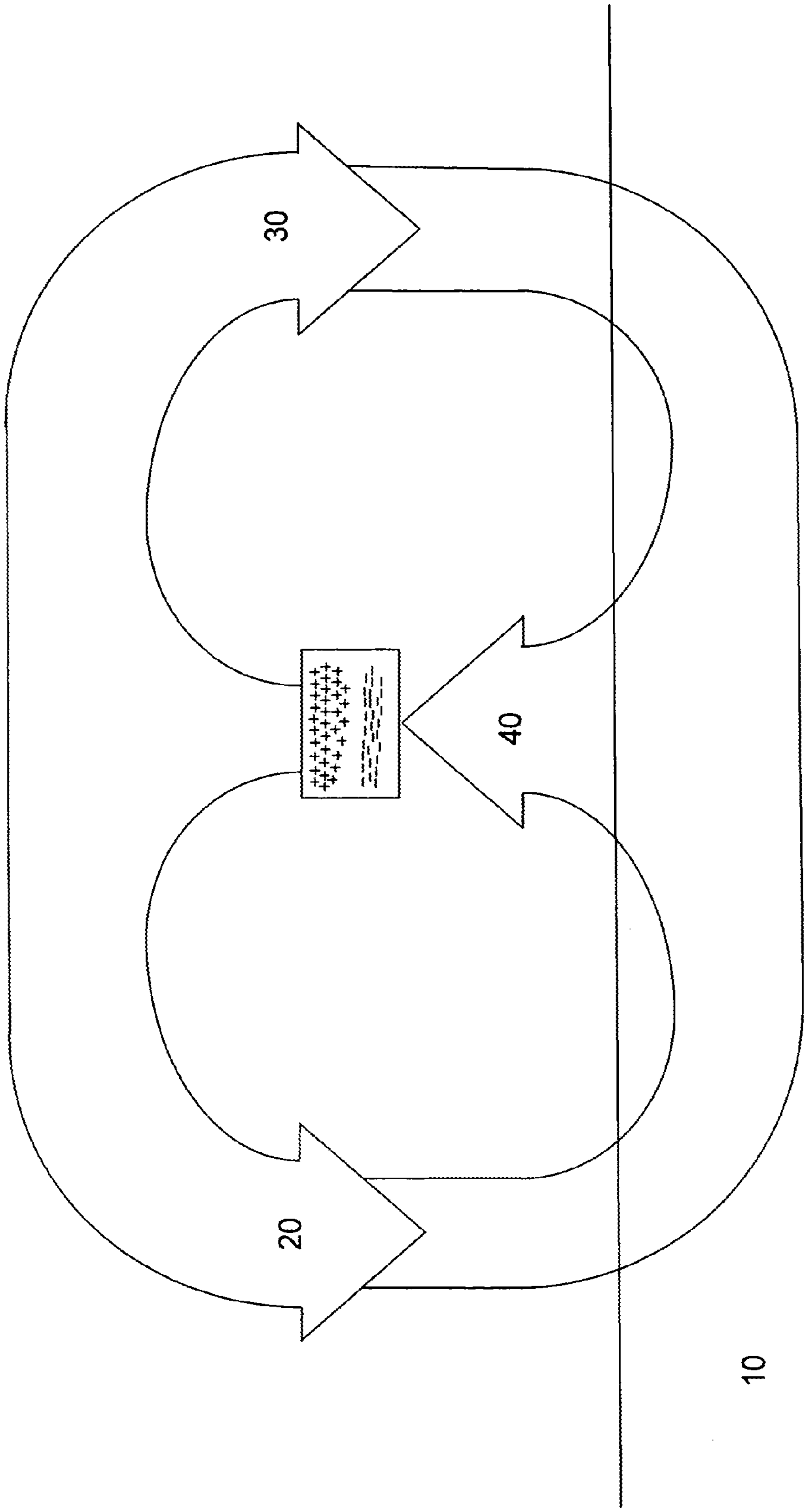

FIGURE 1

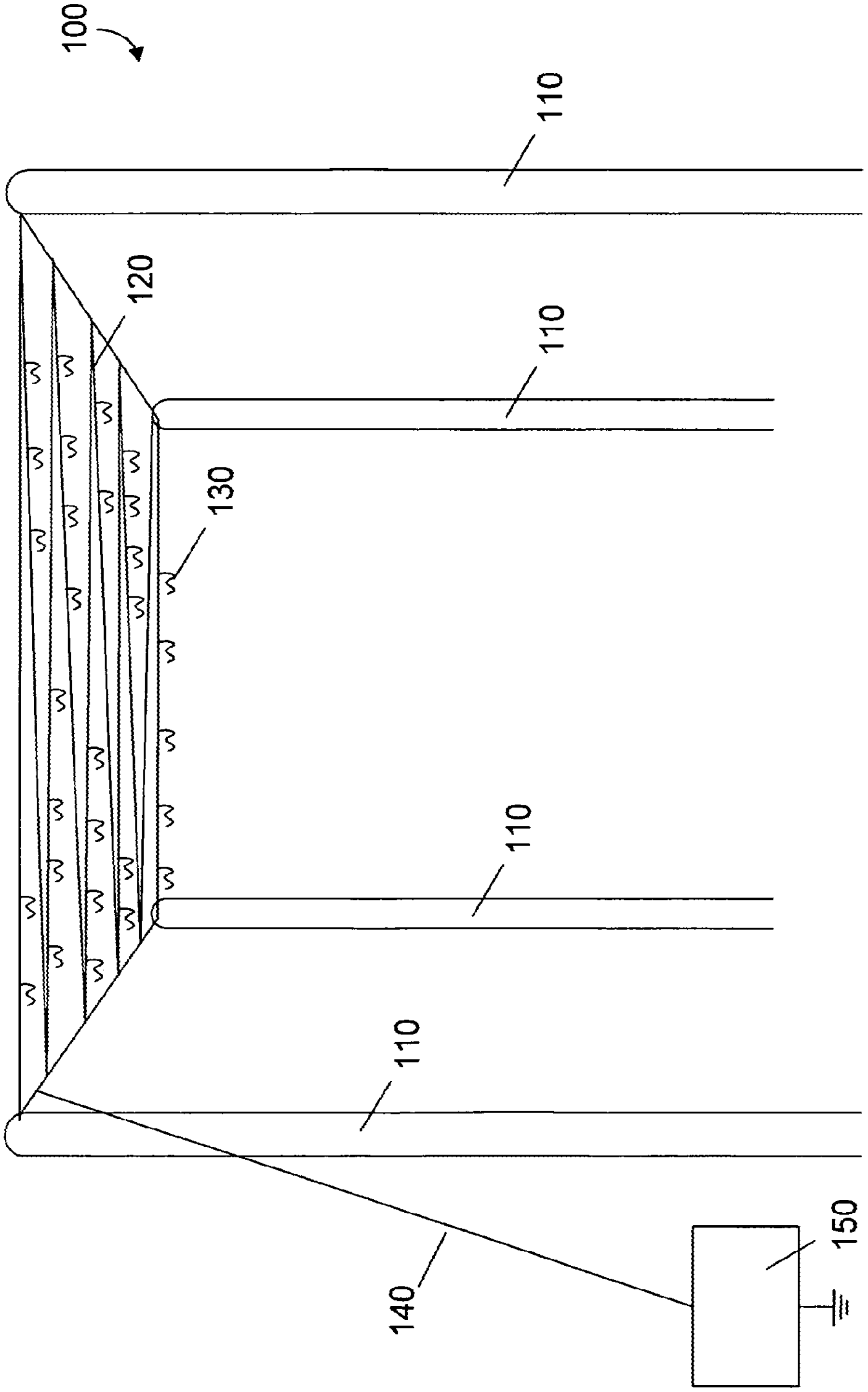
100
110
110
110
110
120
130
140
150

FIGURE 2

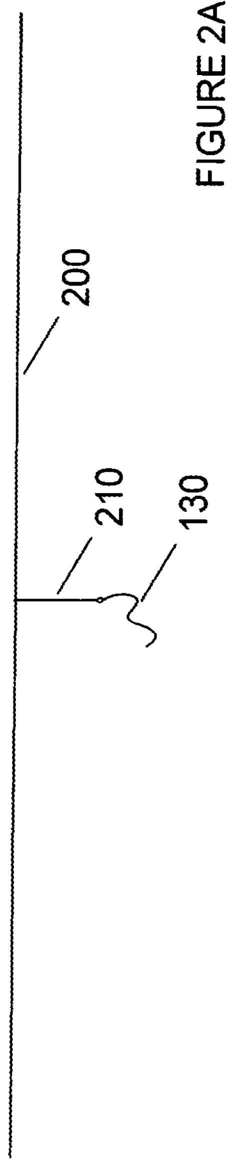

FIGURE 2A

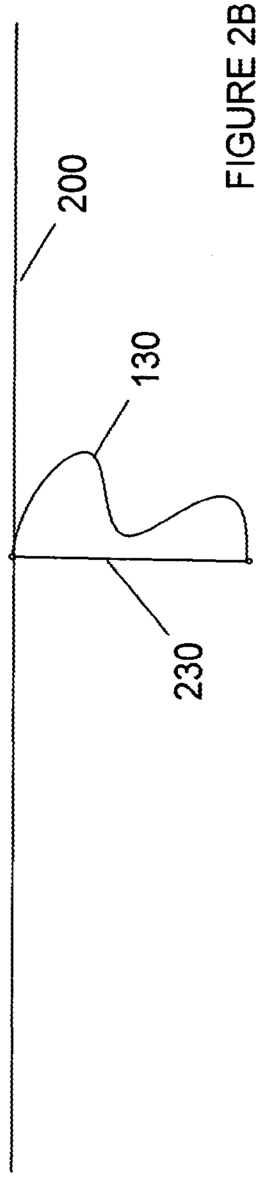

FIGURE 2B

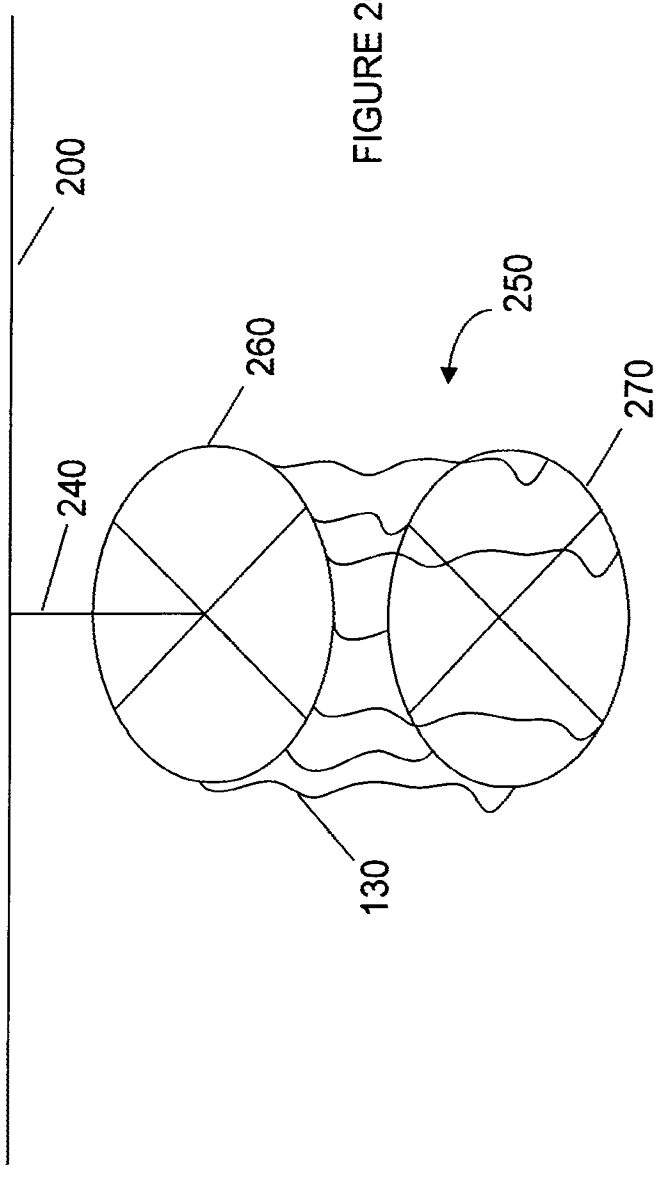
200
250
260
270
240
130

FIGURE 2C

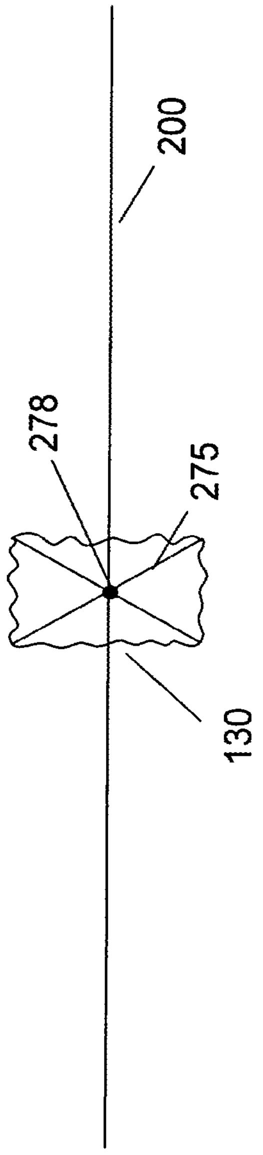
200
278
275
130

FIGURE 2D

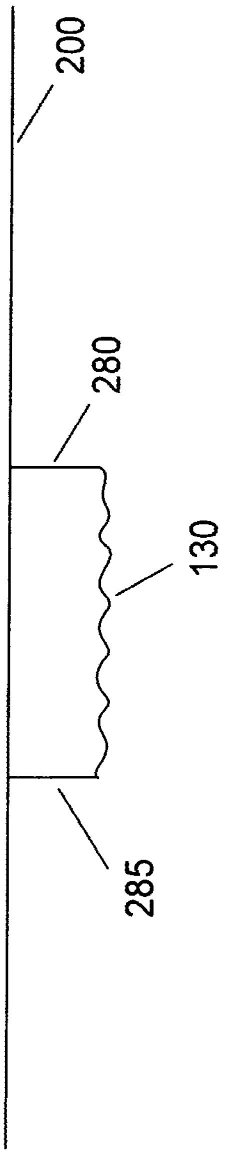
200
280
130
285

FIGURE 2E

FIGURE 2F

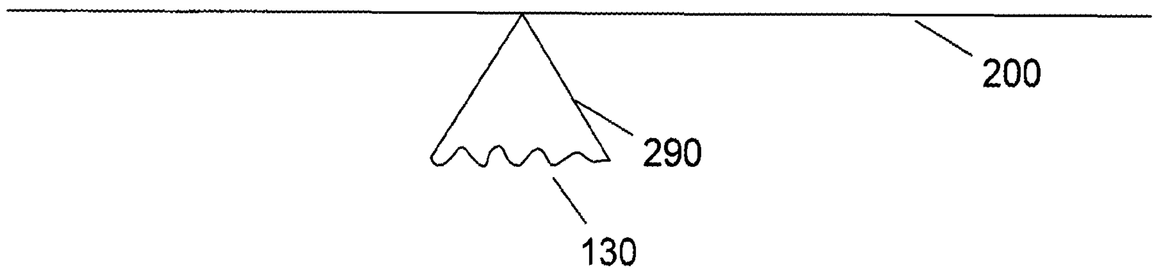

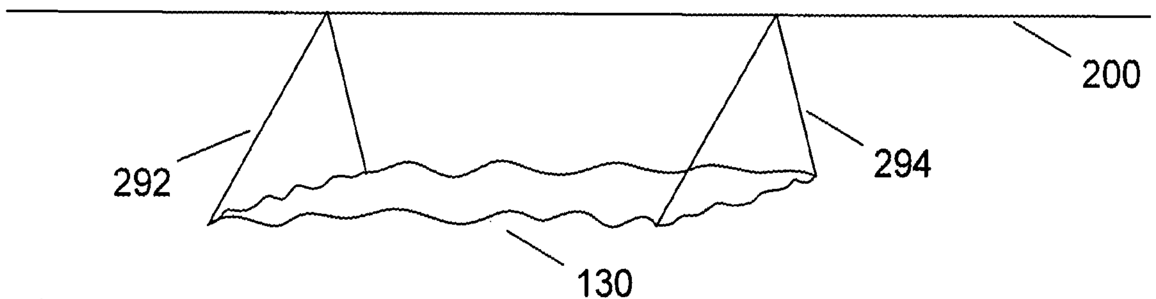
200
292
294
130

FIGURE 2G

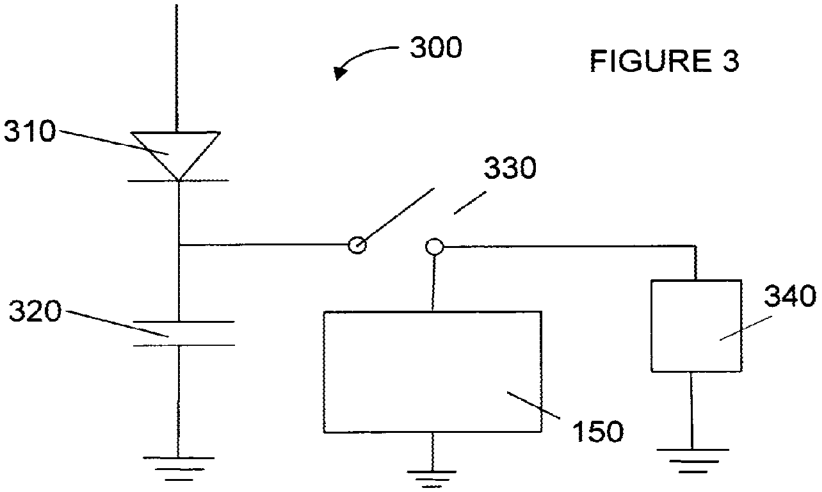
300
310
330
320
340
150

FIGURE 3

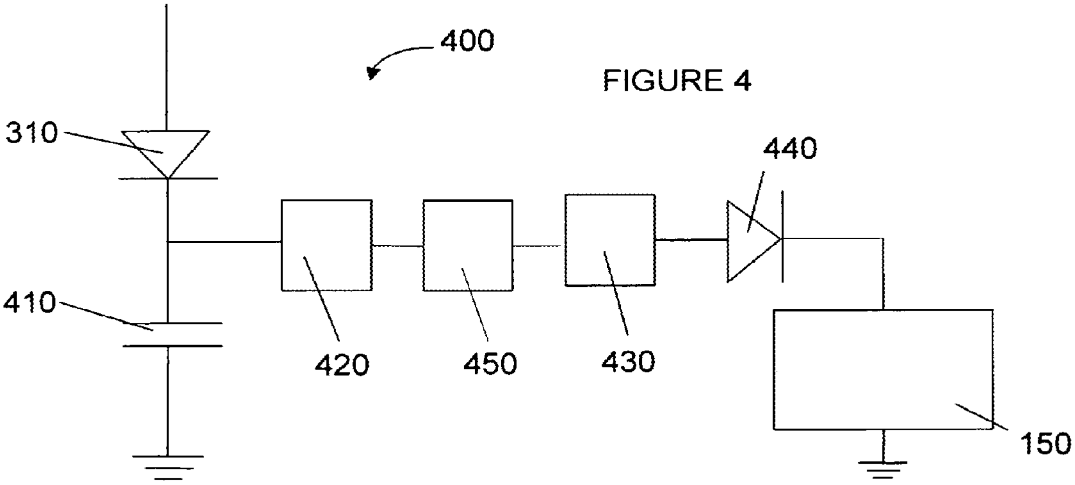
400
310
440
410
420
450
430
150

FIGURE 4

FIGURE 5

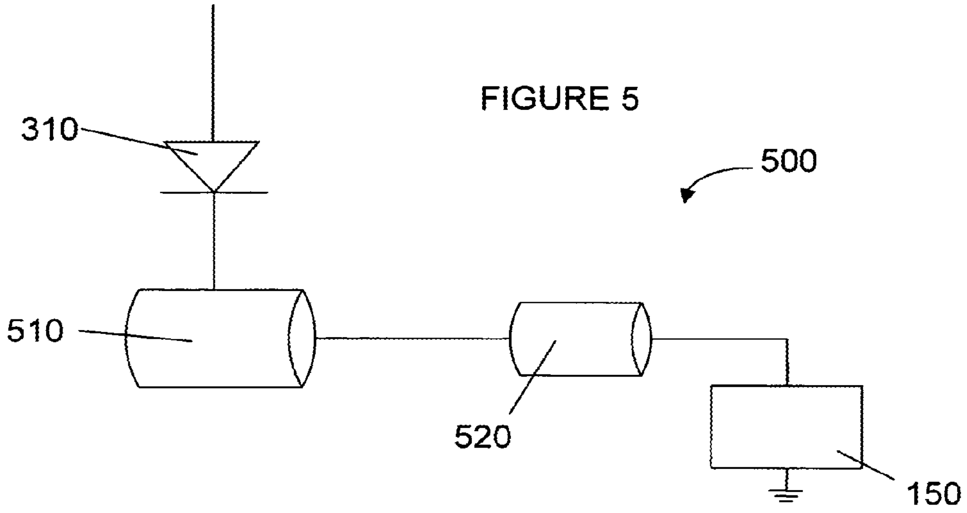
310
500
510
520
150

FIGURE 6

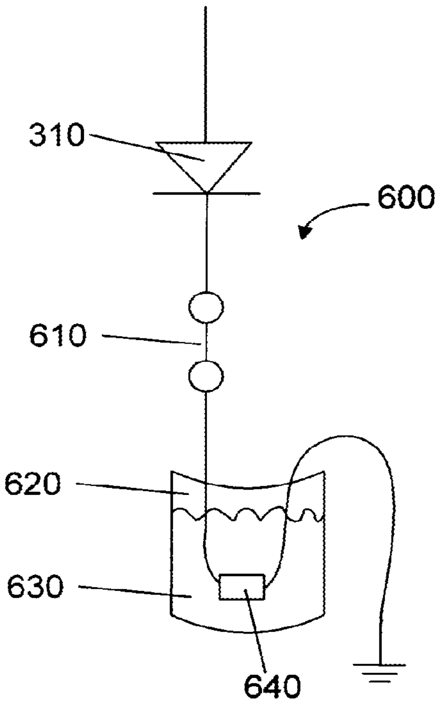
310
600
610
620
630
640

FIGURE 7

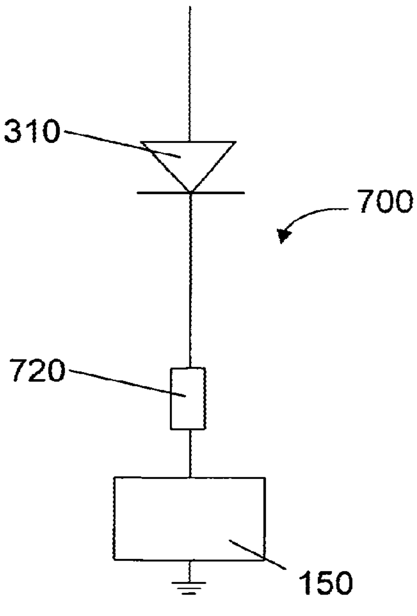
310
700
720
150

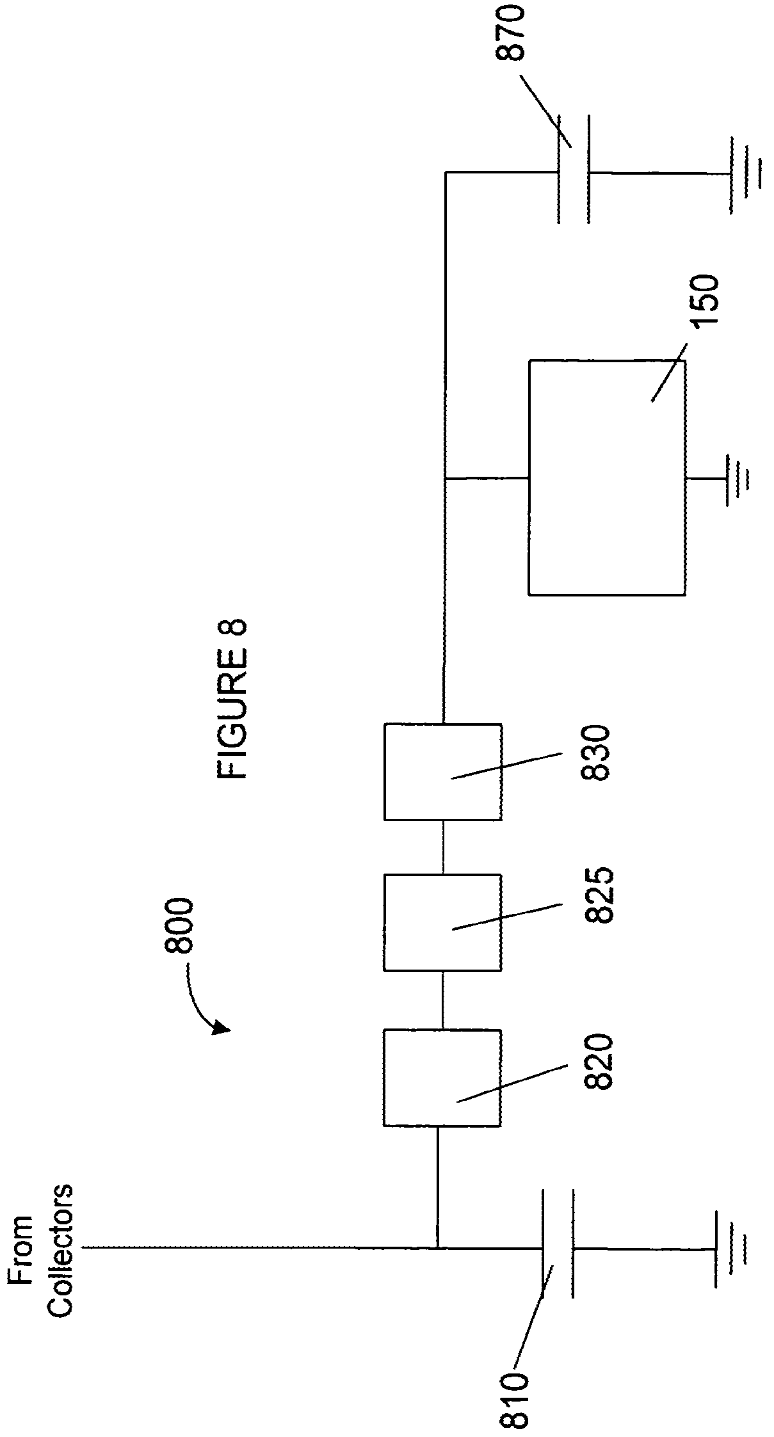
800
From Collectors
810
820
825
830
150
870

FIGURE 8

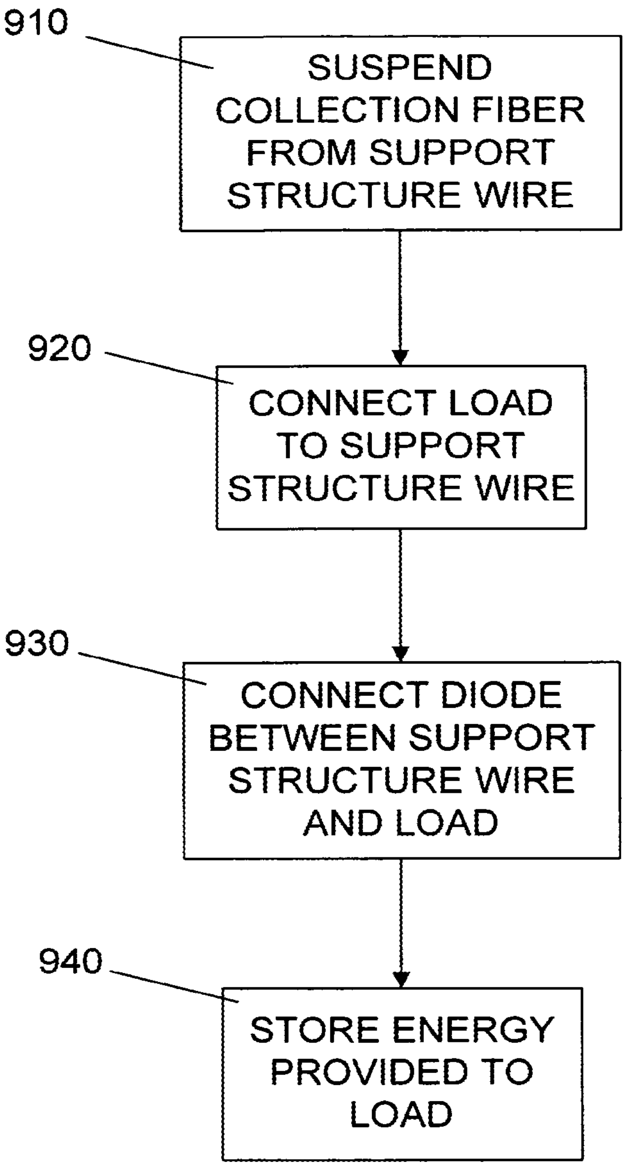
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

ENERGY COLLECTION

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation application of U.S. patent application Ser. No. 12/255,130, filed on Oct. 21, 2008 (now U.S. Pat. No. 8,686,575, issued on Apr. 1, 2014), which is a continuation application of U.S. patent application Ser. No. 11/358,264, filed on Feb. 21, 2006 (now U.S. Pat. No. 7,439,712, issued on Oct. 21, 2008), which are incorporated by reference herein.

TECHNICAL FIELD

The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure wire elevated above a ground level, at least one collection fiber electrically connected to the support structure wire; a load electrically connected to the support structure wire; and a diode electrically connected between the load and at least one collection fiber.

Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection fiber from a support structure wire elevated above ground level, the fiber electrically connected to the support structure wire; providing a load with an electrical connection to the support structure wire to draw current; and providing a diode electrically connected between the collection fiber and the load.

Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. **1** is a circuit diagram of a weather energy circuit.

FIG. **2** is a perspective view of an exemplary embodiment of many energy collectors elevated above ground by a structure.

FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

FIG. **2**B is a side view of an exemplary embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

FIG. **2**D is a side view of an exemplary embodiment of a support structure for multiple energy collection fibers.

FIG. **2**E is a side view of a support structure for an energy collection fiber.

FIG. **2**F is a side view of an exemplary embodiment of a support structure for an energy collection fiber.

FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

FIG. **3** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **4** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **5** is a circuit diagram of an exemplary embodiment of an energy collection circuit for powering a generator and motor.

FIG. **6** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

FIG. **7** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

FIG. **8** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy.

FIG. **9** is a flow diagram of an exemplary embodiment of collecting energy with a collection fiber.

DETAILED DESCRIPTION

Electric charges on conductors reside entirely on the external surface of the conductors, and tend to concentrate more around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An exemplary embodiment of this disclosure takes advantage of this property, among others, to collect and use the energy generated by an electric field in the atmosphere. Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An exemplary embodiment utilizes carbon or graphite fibers for static electricity collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including aluminum or steel, but most notably, copper. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

An exemplary embodiment of the collection fibers as collection device **130** includes graphite or carbon fibers. Graphite and carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, or graphite fibers are examples that demonstrate exemplary collection ability.

In at least one exemplary embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than **100** volts per meter under some conditions. When support wire **120** is suspended in the air at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An exemplary embodiment of the collection device comprises the diode with no collection fiber. A preferred embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restricts energy that may be collected through one fiber from escaping through another fiber.

Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2**A-**2**G using a collection fiber embodiment. FIG. **2**A presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**.

Likewise, FIG. **2**B shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

FIG. **2**C presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

FIG. **2**D presents another exemplary embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2**E provides another exemplary embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

FIG. **2**F presents another exemplary embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2**G provides two supports as provided in FIG. **2**F such that at least two support members **292**, **294** may be connected to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and between multiple support structures. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**.

FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load. Capacitor **320** maybe connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

In an exemplary embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel. Non-limiting examples of spark gap **420** include mercury-reed switches and mercury-wetted reed switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground. The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes maybe placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches and mercury-wetted reed switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel.

FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit. Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

FIG. **8** presents exemplary circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches and mercury-wetted reed switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

It should be emphasized that the above-described embodiments of the present disclosure, particularly, any "preferred" embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A method of collecting energy comprising:

suspending at least one ion collection device with, in operation, microscopic points of a cross section of the collection device exposed to the environment from a support structure; and

providing a load with an electrical connection to the at least one collection device to draw current.

2. The method of claim **1**, wherein the ion collection device comprises a diode.

3. The method of claim **1**, wherein the ion collection device comprises a collection fiber.

4. The method of claim **3**, wherein the collection fiber comprises carbon fiber and/or graphite fiber.

5. The method of claim **1**, wherein the ion collection device comprises a diode and a collection fiber and the diode is electrically connected between the collection fiber and the load.

6. The method of claim **1**, further comprising storing energy provided to the load in an energy storage device.

7. The method of claim **6**, wherein storing energy provided to the load comprises storing energy in a capacitor or an inductor.

8. A system of energy collection comprising:

a support structure;

at least one ion collection device with, in operation, microscopic points of a cross section of the collection device exposed to the environment, the ion collection device electrically connected to and suspended from the support structure; and

a load electrically connected to the at least one ion collection device.

9. The system of claim **8**, wherein the ion collection device comprises a diode.

10. The system of claim **9**, wherein the diode is elevated relative to the ground level.

11. The system of claim **10**, wherein the collection fiber comprises a carbon fiber or a graphite fiber.

12. The system of claim **8**, wherein the ion collection device comprises a collection fiber.

13. The system of claim **8**, wherein the ion collection device comprises a collection fiber and a diode electrically connected between the load and the collection fiber.

14. The system of claim **8**, further comprising a diode electrically connected between the at least one ion collection device and the support structure.

15. The system of claim **8**, further comprising:

a switch connected in series between the at least one ion collection device and the load; and

a capacitor connected in parallel with the switch and the load.

16. The system of claim **15**, wherein the switch comprises an interrupter connected between the at least one ion collection device and the load, the interrupter comprising at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

17. The system of claim **16**, further comprising a transformer connected between the interrupter and the load.

18. The system of claim **8**, further comprising: a motor for providing power, the motor connected between the at least one ion collection device and the load; and a generator powered by the motor.

19. The system of claim **8**, further comprising a fuel cell electrically connected between the ion collection device and the load.

20. The system of claim **19**, wherein the fuel cell produces hydrogen and oxygen.

21. A system of collecting energy comprising:

means for suspending at least one ion collection device with, in operation, microscopic points of a cross section of the collection device exposed to the environment;

means for inducing current flow, the means for inducing current flow electrically connected to at least one ion collection device; and

means for restricting the backflow of charge carriers, the means for restricting the backflow of charge carriers electrically connected between the at least one collection device and the means for inducing current flow.

* * * * *

US009331603B2

(12) **United States Patent**
McCowen

(10) **Patent No.: US 9,331,603 B2**
(45) **Date of Patent: May 3, 2016**

(54) **ENERGY COLLECTION**

(71) Applicant: **Ion Power Group LLC**, Navarre, FL (US)

(72) Inventor: **Clint McCowen**, Navarre, FL (US)

(73) Assignee: **Ion Power Group, LLC**, Navarre, FL (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 34 days.

(21) Appl. No.: **14/454,308**

(22) Filed: **Aug. 7, 2014**

(65) **Prior Publication Data**

US 2016/0043661 A1 Feb. 11, 2016

(51) **Int. Cl.**

F02B 63/04	(2006.01)
F03G 7/08	(2006.01)
H02K 7/18	(2006.01)
H02N 1/04	(2006.01)
F03D 9/00	(2006.01)
H02P 9/04	(2006.01)

(52) **U.S. Cl.**
CPC ***H02N 1/04*** (2013.01)

(58) **Field of Classification Search**
USPC .... 290/1 R, 44; 126/683; 307/149; 322/2 R
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427	A	5/1901	Palencsar
911,260	A	2/1909	Pennock
1,014,719	A	1/1912	Pennock
2,473,819	A	6/1949	Pittman
3,532,959	A	10/1970	Erickson
3,780,722	A	12/1973	Swet
3,946,227	A	3/1976	Bingham
4,079,225	A	3/1978	Warner
4,092,250	A	5/1978	Sano et al.
4,104,696	A	8/1978	Cochran, Jr.
4,146,800	A	3/1979	Gregory et al.
4,180,056	A	12/1979	Schnabel et al.
4,201,197	A	5/1980	Dismer
4,206,396	A	6/1980	Marks
4,224,496	A	9/1980	Riordan et al.
4,307,936	A	12/1981	Ochiai
4,314,192	A	2/1982	Caro
4,346,478	A	8/1982	Sichling
4,425,905	A	1/1984	Mori
4,433,248	A	2/1984	Marks
RE31,678	E	9/1984	Ochiai
4,483,311	A	11/1984	Whitaker

(Continued)

FOREIGN PATENT DOCUMENTS

CA	2647385	9/2008
CN	200780006209.7	8/2008

(Continued)

OTHER PUBLICATIONS

U.S. Appl. No. 12/637,720, filed Dec. 14, 2009, McCowen

Primary Examiner — Pedro J Cuevas
(74) *Attorney, Agent, or Firm* — Benjamin A. Balser; Next IP Law Group

(57) **ABSTRACT**

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support system. The support system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but graphene, carbon and graphite are preferred. Diodes may be used to restrict the backflow or loss of energy.

20 Claims, 16 Drawing Sheets

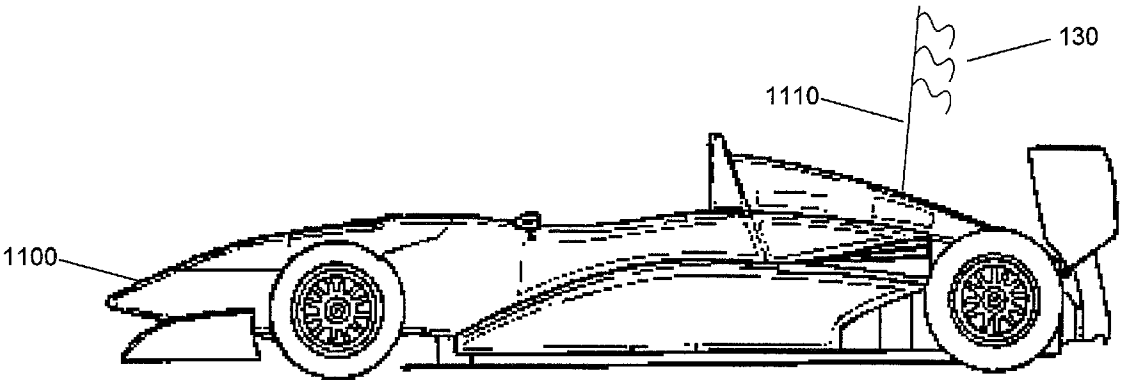

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,489,269	A	12/1984	Edling et al.	
4,512,335	A	4/1985	Mori	
4,653,223	A	3/1987	Mori	
4,653,472	A	3/1987	Mori	
4,676,226	A	6/1987	Mori	
4,809,675	A	3/1989	Mori	
4,852,454	A	8/1989	Batchelder	
4,943,125	A	7/1990	Laundre' et al.	
4,988,159	A	1/1991	Turner et al.	
5,047,892	A	9/1991	Sakata et al.	
5,114,101	A	5/1992	Stern et al.	
5,145,257	A	9/1992	Bryant et al.	
5,379,103	A	1/1995	Zigler	
5,851,309	A	12/1998	Kousa	
5,942,806	A	8/1999	Veliadis	
6,025,591	A	2/2000	Taylor et al.	
6,038,363	A	3/2000	Slater et al.	
6,116,544	A	9/2000	Forward et al.	
6,173,922	B1	1/2001	Hoyt et al.	
6,226,440	B1	5/2001	Lyons	
6,419,191	B1	7/2002	Hoyt et al.	
6,425,391	B1	7/2002	Davoren et al.	
6,550,639	B2	4/2003	Brown et al.	
6,735,830	B1	5/2004	Merciel	
6,765,212	B2	7/2004	Faus et al.	
6,846,447	B2	1/2005	Mbachu et al.	
6,853,447	B2	2/2005	Goetz	
6,894,772	B2	5/2005	Goetz et al.	
6,925,852	B2	8/2005	Susko	
6,974,110	B2	12/2005	Grandics	
7,109,597	B1	9/2006	Bose	
7,128,867	B2	10/2006	Mbachu et al.	
7,231,809	B2	6/2007	Susko	
7,259,475	B2	8/2007	Hong et al.	
7,439,712	B2 *	10/2008	McCowen	H02N 1/10 174/2
7,449,668	B2	11/2008	Schutten et al.	
7,478,712	B2 *	1/2009	McCowen	H02N 11/002 174/2
7,532,819	B1	5/2009	Triebes et al.	
7,592,783	B1	9/2009	Jarvinen	
8,269,401	B1	9/2012	Kim et al.	
8,519,596	B1	8/2013	Kim et al.	
8,686,575	B2 *	4/2014	McCowen	H02N 11/002 290/1 R
2002/0026933	A1	3/2002	Gottlieb	
2002/0108892	A1	8/2002	Goetz et al.	
2002/0109094	A1	8/2002	Goetz et al.	
2002/0109835	A1	8/2002	Goetz	
2003/0125630	A1	7/2003	Furnish	
2003/0193319	A1	10/2003	Wood et al.	
2004/0011925	A1	1/2004	Grandics	
2004/0083793	A1	5/2004	Susko	
2004/0094853	A1	5/2004	Mbachu et al.	
2004/0160711	A1	8/2004	Stumberger	
2004/0212945	A1	10/2004	Sprenger et al.	
2005/0101024	A1	5/2005	Mbachu et al.	
2005/0136311	A1	6/2005	Ueda et al.	
2005/0270525	A1	12/2005	Susko	
2007/0107765	A1	5/2007	Schutten et al.	
2007/0195481	A1	8/2007	McCowen	
2007/0218323	A1	9/2007	Sudo et al.	
2007/0223171	A1	9/2007	Guy et al.	
2007/0273206	A1	11/2007	McCowen	
2007/0297892	A1	12/2007	Kildegaard	
2008/0145714	A1	6/2008	Kagami	
2009/0040680	A1	2/2009	McCowen	
2009/0107842	A1	4/2009	Park et al.	
2009/0114495	A1	5/2009	McCowen	
2009/0216292	A1	8/2009	Pless et al.	
2010/0090562	A1	4/2010	Mccowen	
2011/0148248	A1	6/2011	Landa	
2012/0043858	A1	2/2012	Mahapatra	
2012/0299559	A1	11/2012	McCowen	
2013/0049531	A1	2/2013	Wang et al.	
2015/0001966	A1 *	1/2015	McCowen	H05F 7/00 307/149

FOREIGN PATENT DOCUMENTS

EP	1999767	A2	12/2008
JP	5552236		5/2014
RU	2008137622		2/2007
WO	2007098341		8/2007

* cited by examiner

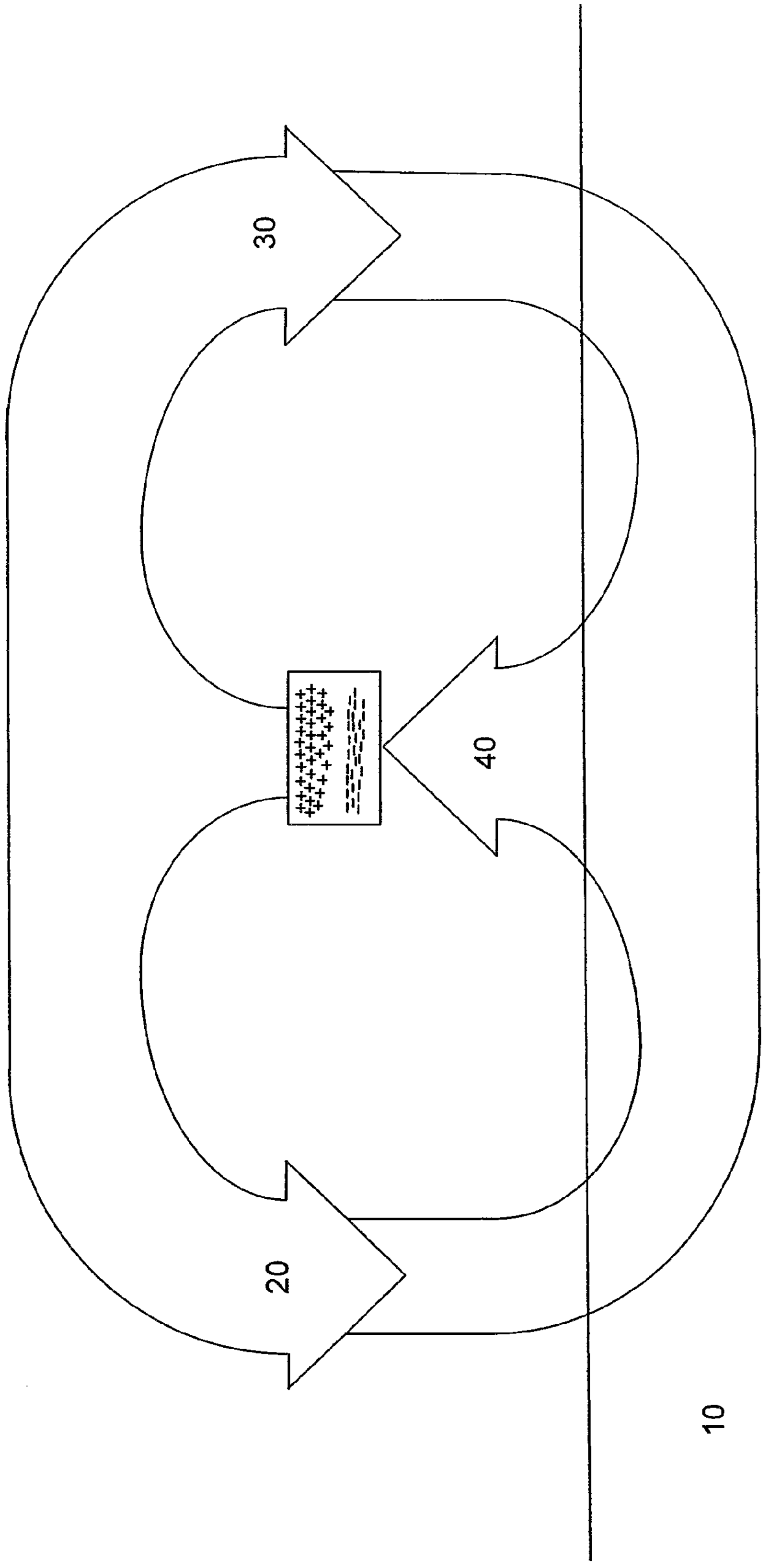

FIGURE 1

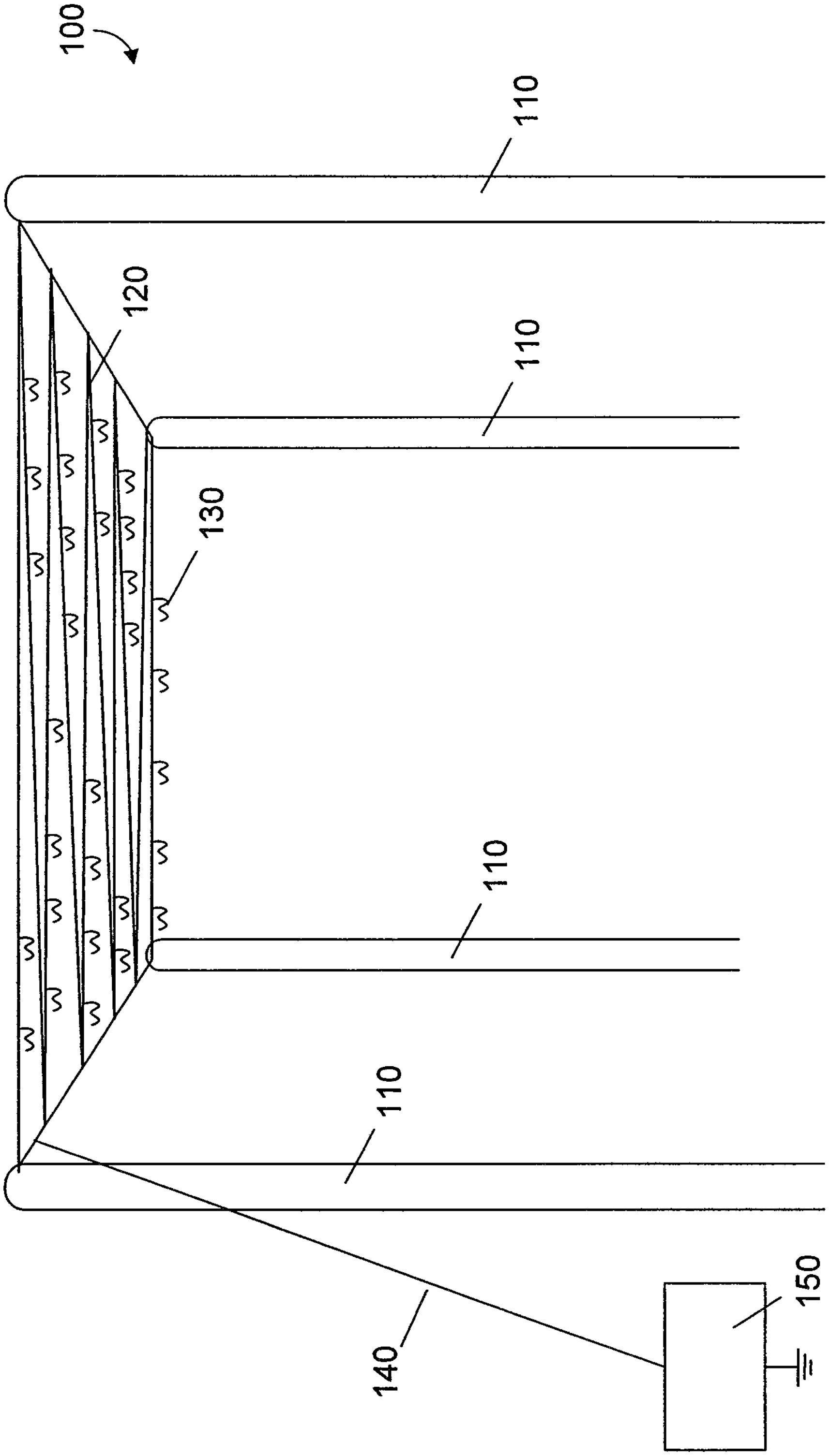
100
110
110
110
110
120
130
140
150

FIGURE 2

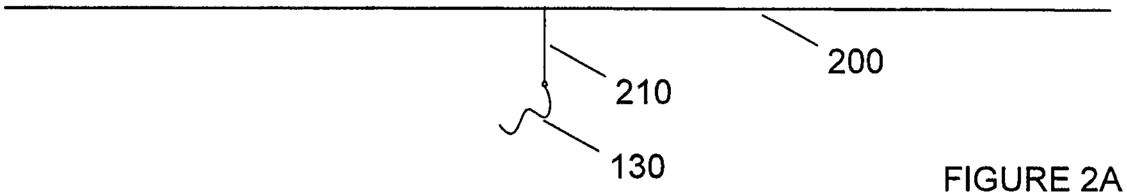

FIGURE 2A

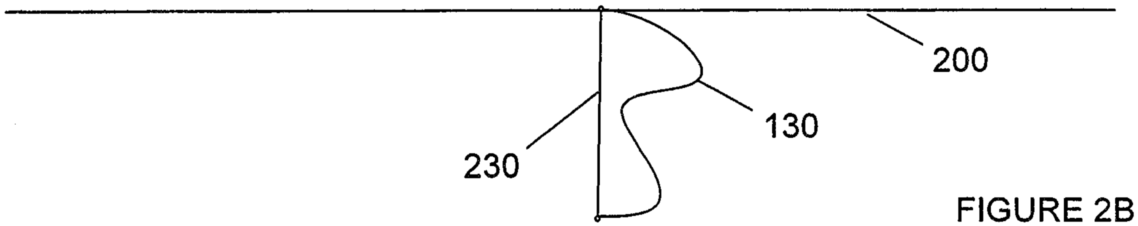

FIGURE 2B

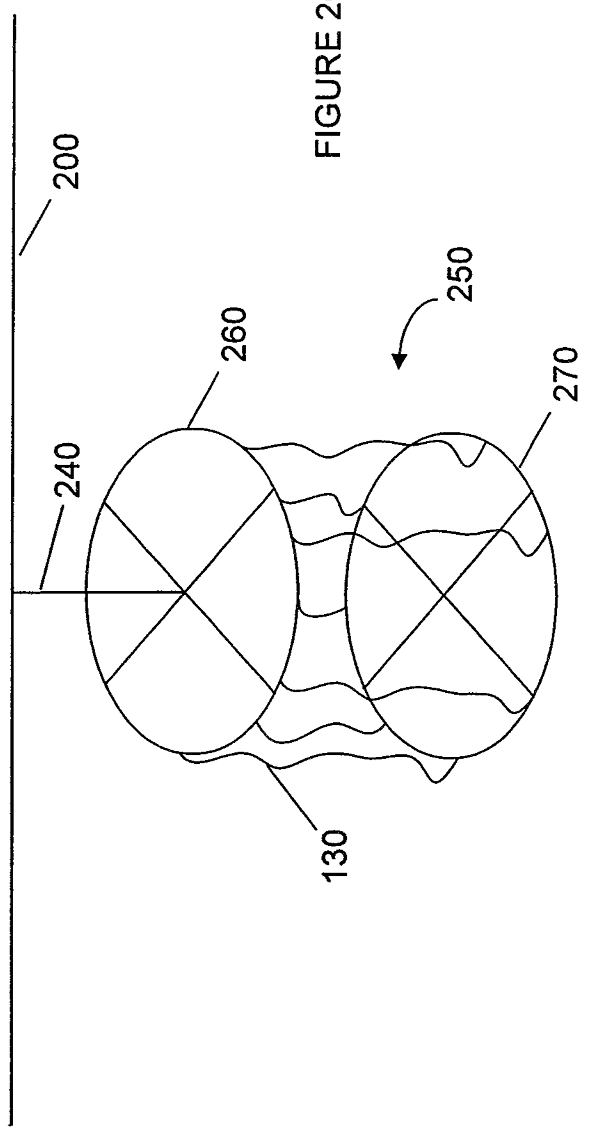
200
250
260
270
240
130

FIGURE 2C

FIGURE 2D

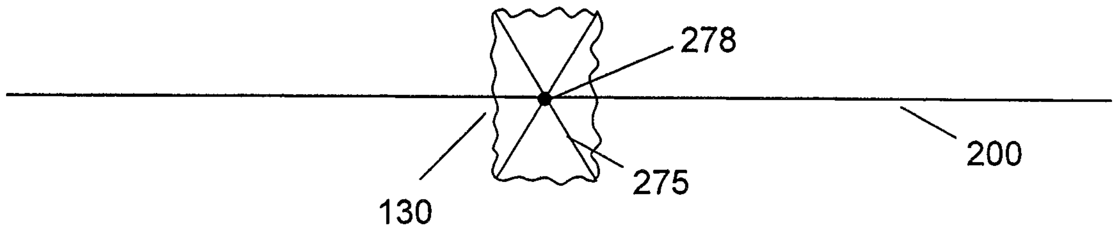
278
200
275
130

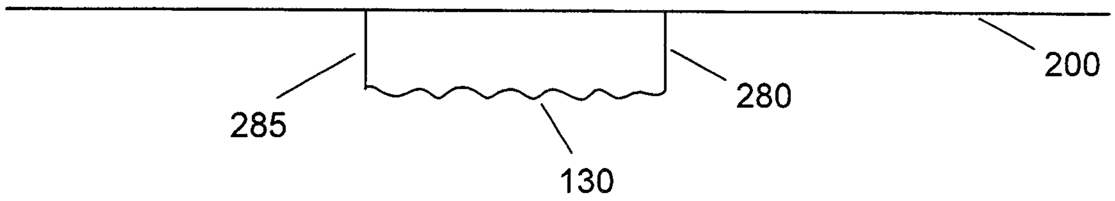
200
285
280
130

FIGURE 2E

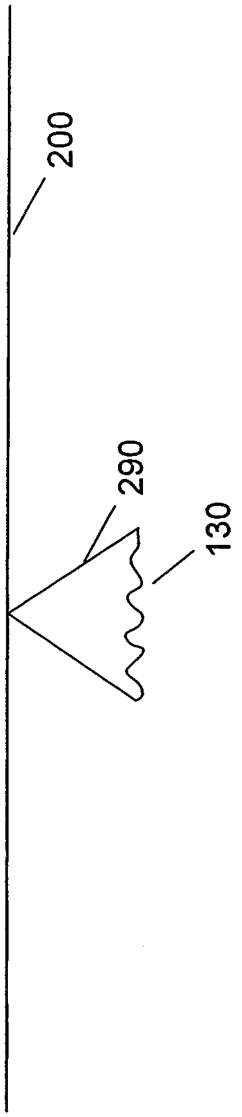

FIGURE 2F

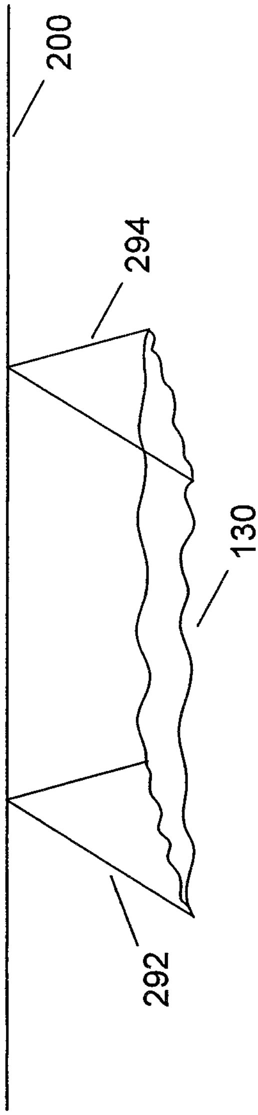
200
294
130
292

FIGURE 2G

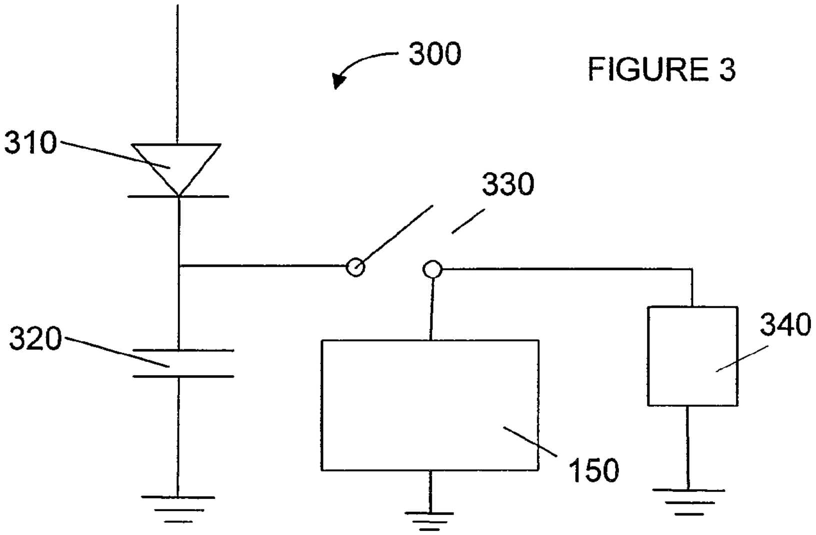
300
310
330
320
340
150

FIGURE 3

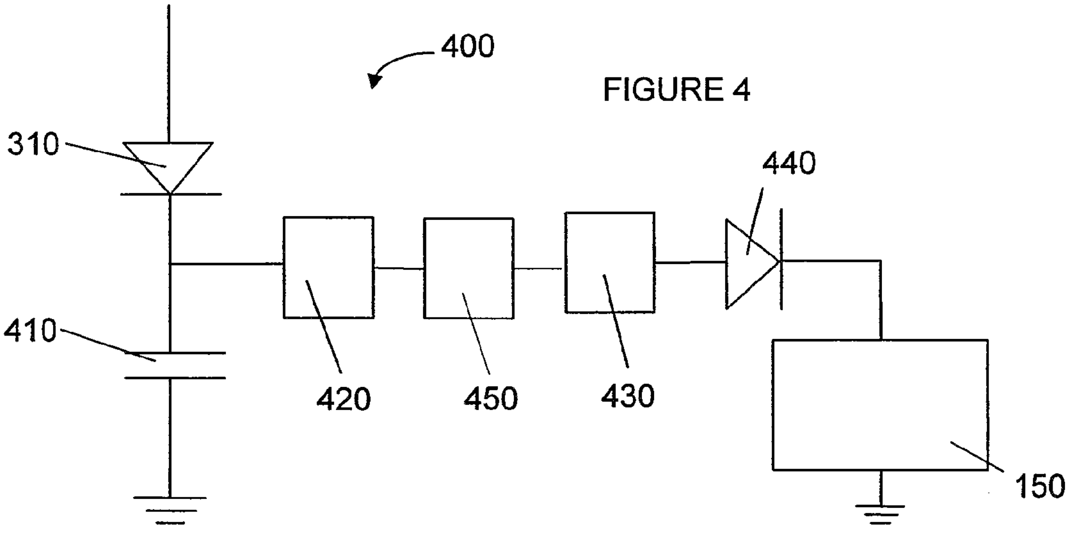
400
310
440
410
420
450
430
150

FIGURE 4

FIGURE 5

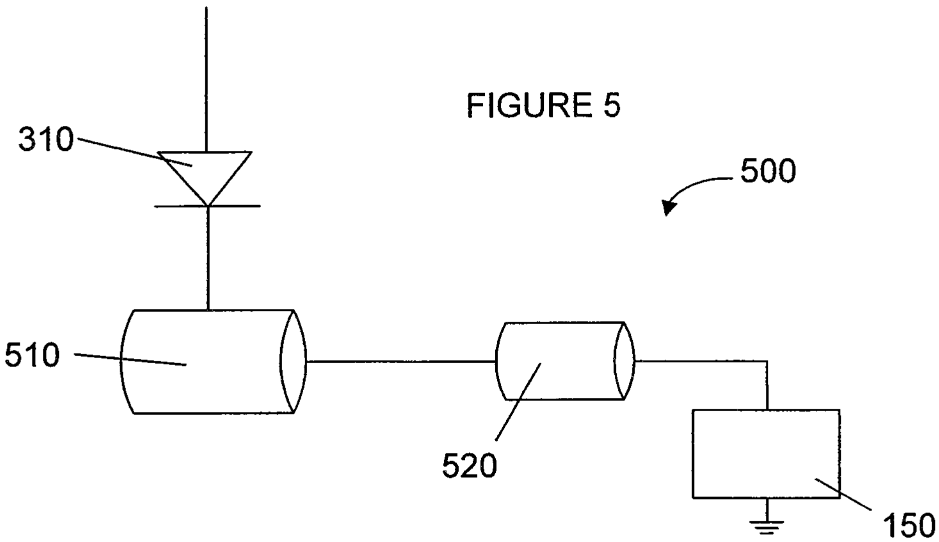
310
500
510
520
150

FIGURE 6

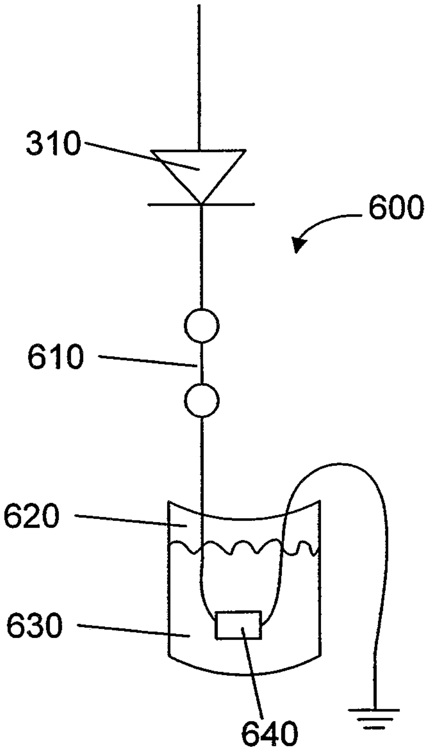
310
600
610
620
630
640

FIGURE 7

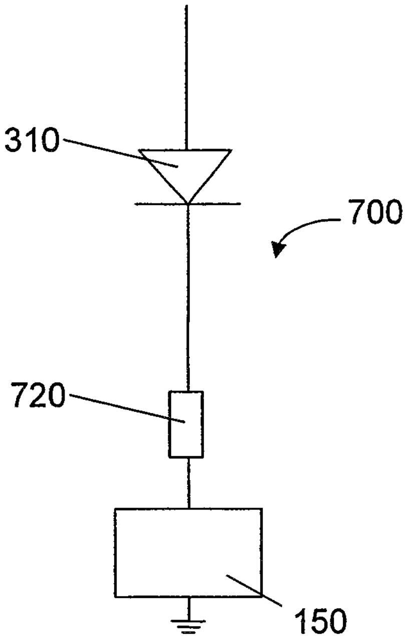
310
700
720
150

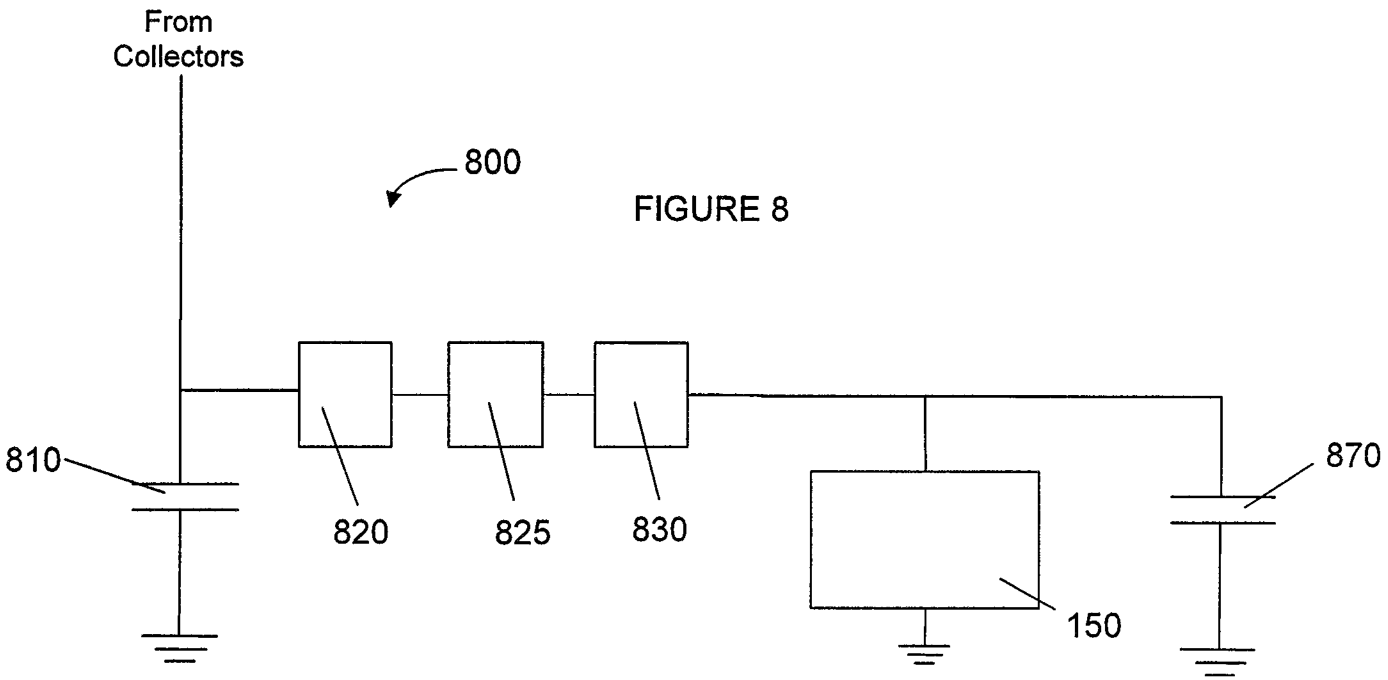
From
Collectors
800
810
820
825
830
150
870

FIGURE 8

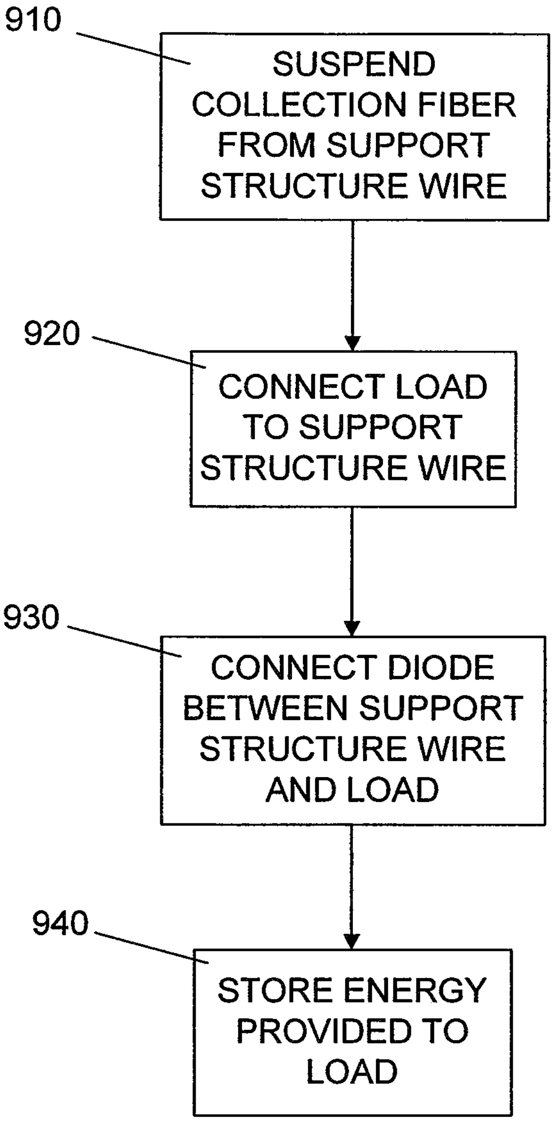
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

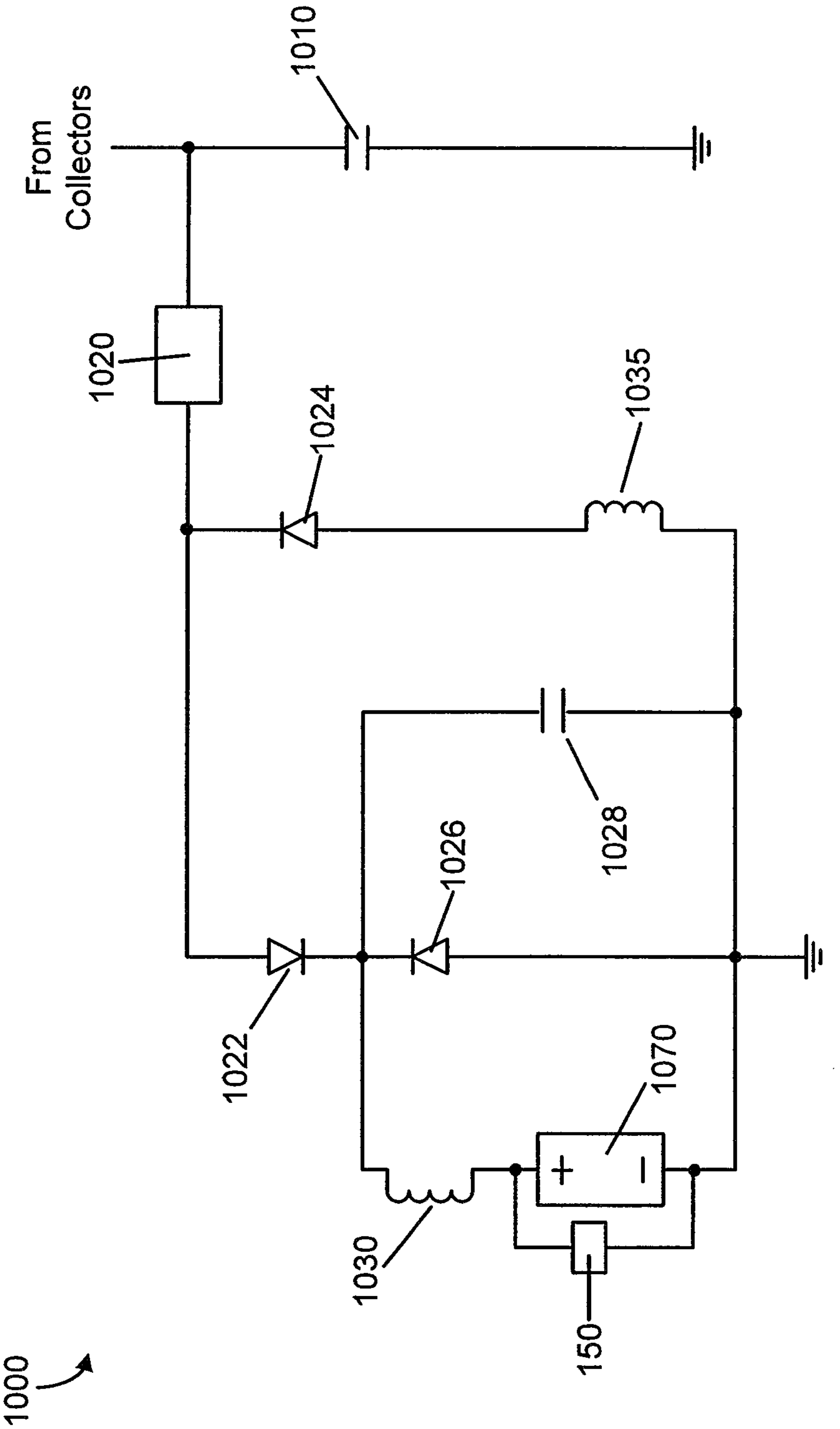
From Collectors
+
−

FIGURE 10

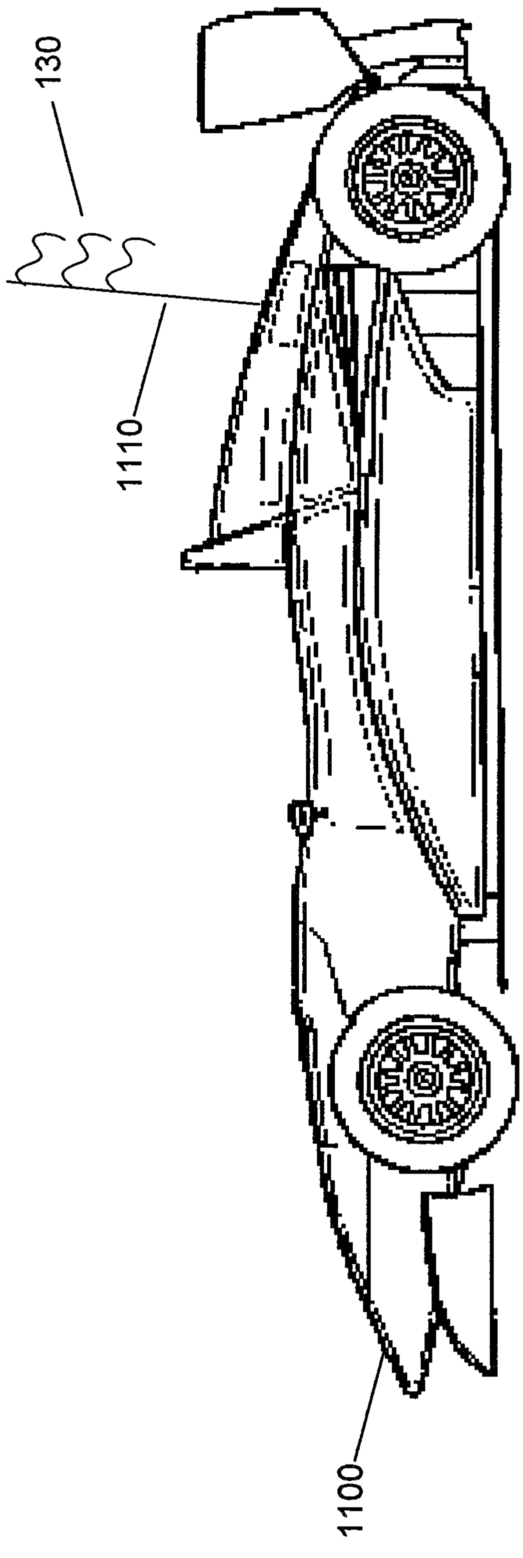
130
1110
1100

FIGURE 11

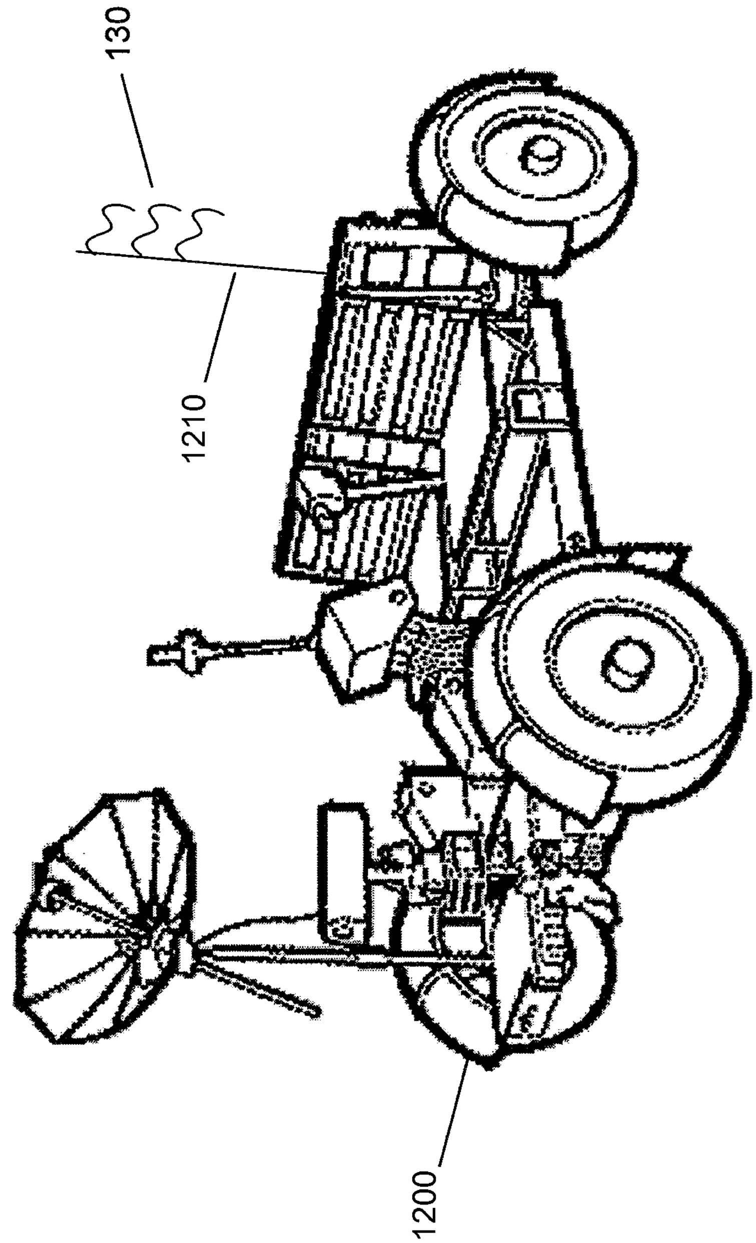
130
1210
1200

FIGURE 12

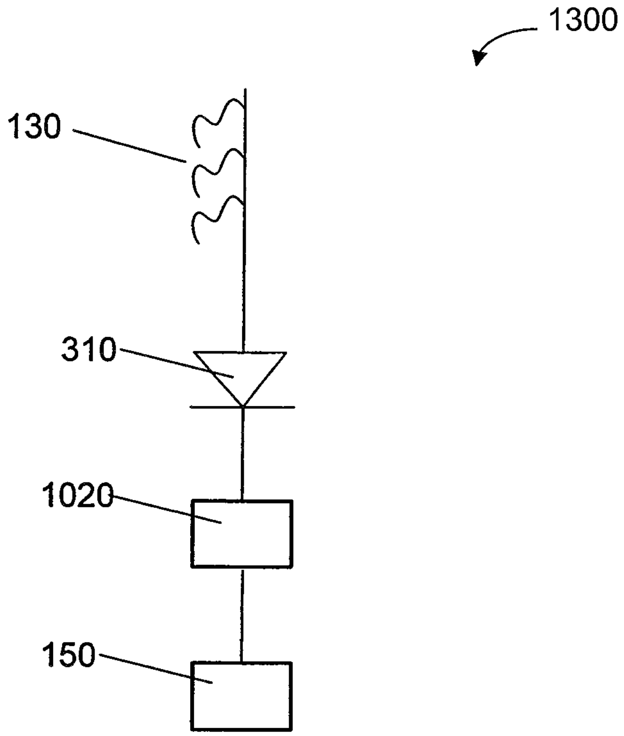
1300
130
310
1020
150

FIGURE 13

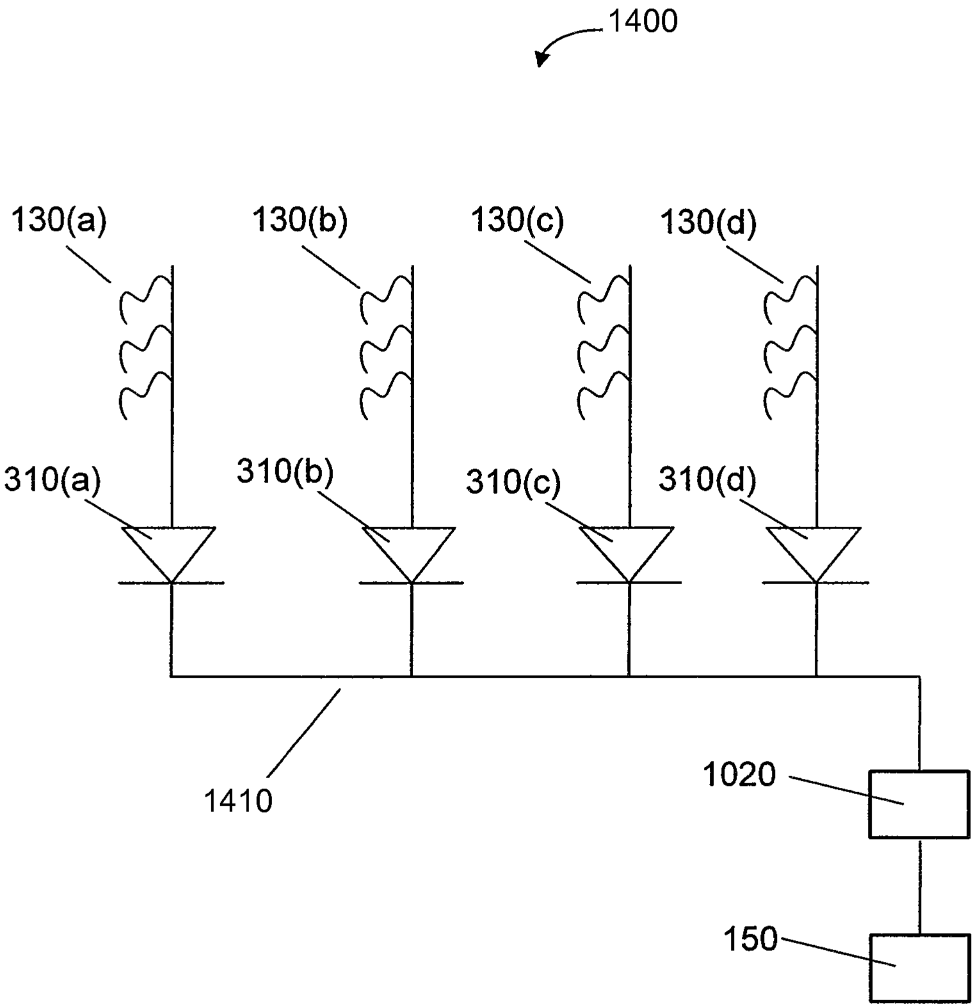
1400
130(a)
130(b)
130(c)
130(d)
310(a)
310(b)
310(c)
310(d)
1410
1020
150

FIGURE 14

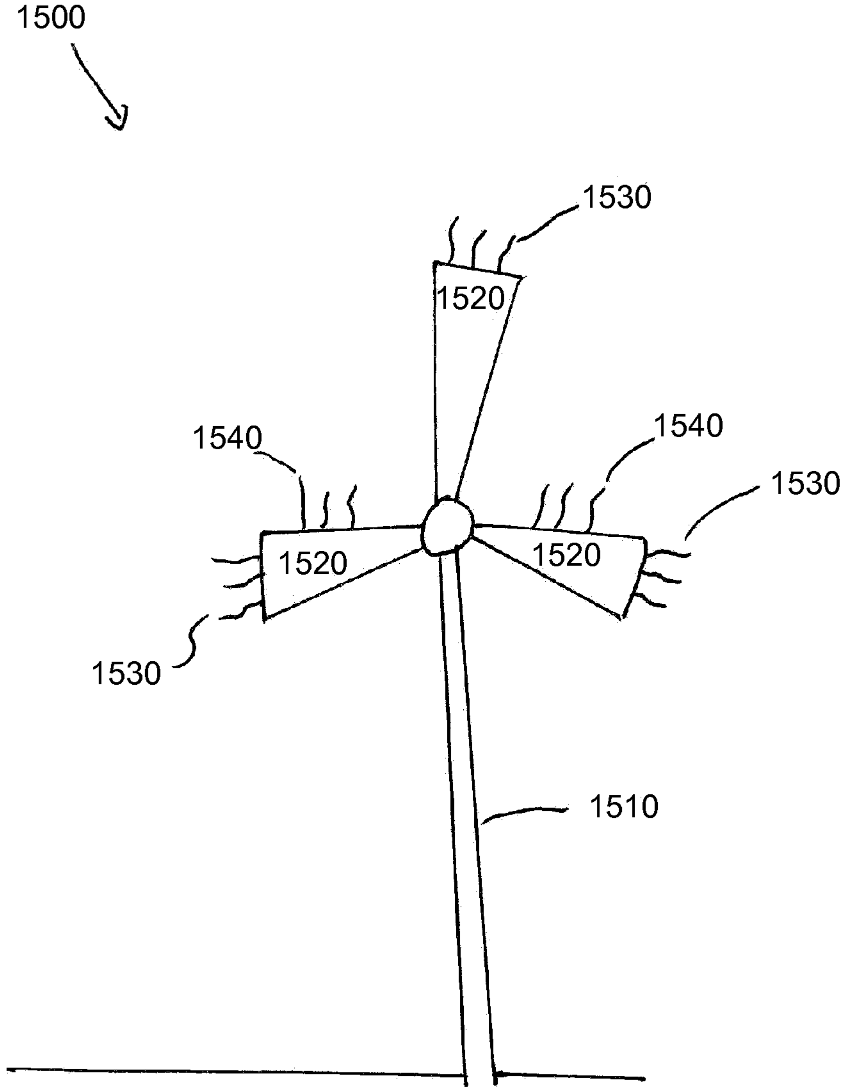
1500
1530
1520
1540
1540
1530
1520
1520
1530
1510

FIGURE 15

ENERGY COLLECTION

TECHNICAL FIELD

The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure, the support structure comprising at least one of an airplane, drone, blimp, balloon, kite, satellite, train, motorcycle, bike, skateboard, scooter, hovercraft, electronic device, electronic device case, billboard, cell tower, radio tower, camera tower, flag pole, telescopic pole, light pole, utility pole, water tower, building, sky scraper, coliseum, roof top, solar panel and a fixed or mobile structure exceeding 1 inch in height above ground or sea level; at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment electrically connected to the support structure; and a load electrically connected to the at least one collection device.

Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment from a support structure, the at least one collection device electrically connected to the support structure, the support structure comprising at least one of an airplane, drone, blimp, balloon, kite, satellite, train, motorcycle, bike, skateboard, scooter, hovercraft, electronic device, electronic device case, billboard, cell tower, radio tower, camera tower, flag pole, telescopic pole, light pole, utility pole, water tower, building, sky scraper, coliseum, roof top, solar panel and a fixed or mobile structure exceeding 1 inch in height above ground or sea level; and providing a load with an electrical connection to the at least one collection device to draw current.

Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. **1** is a circuit diagram of a weather energy circuit.

FIG. **2** is a perspective view of an example embodiment of many energy collectors elevated above ground by a structure.

FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

FIG. **2**B is a side view of an example embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

FIG. **2**D is a side view of an example embodiment of a support structure for multiple energy collection fibers.

FIG. **2**E is a side view of a support structure for an energy collection fiber.

FIG. **2**F is a side view of an example embodiment of a support structure for an energy collection fiber.

FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

FIG. **3** is a circuit diagram of an example embodiment of a circuit for the collection of energy.

FIG. **4** is a circuit diagram of an example embodiment of a circuit for the collection of energy.

FIG. **5** is a circuit diagram of an example embodiment of an energy collection circuit for powering a generator and motor.

FIG. **6** is a circuit diagram of an example embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

FIG. **7** is a circuit diagram of an example embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

FIG. **8** is a circuit diagram of an example embodiment of a circuit for collecting energy.

FIG. **9** is a flow diagram of an example embodiment of collecting energy with a collection fiber.

FIG. **10** is a circuit diagram of an example embodiment of a circuit for collecting energy from a dual polarity source.

FIG. **11** is a system diagram of an example embodiment of an energy collection system connected to an automobile vehicle.

FIG. **12** is a system diagram of an example embodiment of an energy collection system connected to a lunar rover vehicle.

FIG. **13** is a system diagram of an example embodiment of an energy collection system comprising collection devices with a diode.

FIG. **14** is a system diagram of an example embodiment of an energy collection system comprising multiple legs of the system of FIG. **13**.

FIG. **15** is a system diagram of an example embodiment of a windmill with energy collectors.

DETAILED DESCRIPTION

Electric charges on conductors reside entirely on the external surface of the conductors, and tend to concentrate more around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An example embodiment of this disclosure takes advantage of this property, among others, to collect and use the energy generated by an electric field in the atmosphere. Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An example embodiment utilizes carbon or graphite fibers for static electricity collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including aluminum or steel, but most notably, copper. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

An example embodiment of the collection fibers as collection device **130** includes graphite or carbon fibers. Graphite and carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, or graphite fibers are examples that demonstrate collection ability.

In at least one example embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than 100 volts per meter under some conditions. When support wire **120** is suspended in the air at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An example embodiment of the collection device comprises the diode with no collection fiber. A preferred embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restricts energy that may be collected through one fiber from escaping through another fiber.

Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2**A-**2**G using a collection fiber embodiment. FIG. **2**A presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**.

Likewise, FIG. **2**B shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

FIG. **2**C presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

FIG. **2D** presents another example embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2E** provides another example embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

FIG. **2F** presents another example embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2G** provides two supports as provided in FIG. **2F** such that at least two support members **292**, **294** may be connected to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and between multiple support structures. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**.

FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load. Capacitor **320** maybe connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

In an example embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel. Non-limiting examples of spark gap **420** include mercury-reed switches and mercury-wetted reed switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground. The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes maybe placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches and mercury-wetted reed switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel.

FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit. Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

FIG. **8** presents example circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches and mercury-wetted reed switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

FIG. **10** presents circuit **1000** as an example embodiment for the collection of energy from a dual polarity source. This may be used, for example, to collect atmospheric energy that reverses in polarity compared with the ground. Such polarity reversal has been discovered as occurring occasionally on Earth during, for example, thunderstorms and bad weather, but has also been observed during good weather. Such polarity reversal may occur on other planetary bodies, including Mars and Venus, as well. Energy polarity on other planets, in deep space, or on other heavenly bodies, may be predominantly negative or predominantly positive. Collector fibers (**130** from FIG. **2**), which may comprise graphene, silicene, and/or other like materials, are capable of collecting positive energy and/or negative energy, and circuit **1000** is capable of processing positive and/or negative energy, providing an output which is always positive. Circuit **1000** may collect energy from one or more collection devices (**130** from FIG. **2**). Charging capacitor **1010** may be used to store a charge until the voltage at spark gap **1020** achieves the spark voltage. Capacitor **1010** is optional.

A plurality of diodes may be placed in a plurality of positions in circuit **1000**. The voltage from capacitor **1010** may be used to charge spark gap **1020** to a sufficient voltage. Spark gap **1020** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **1020** include mercury-reed switches, mercury-wetted reed switches, open-gap spark gaps, and electronic switches. When spark gap **1020** arcs, energy will arc from an emitting end of spark gap **1020** to a receiving end of spark gap **1020**. The output of spark gap **1020** is electrically connected to the anode of diode **1022** and the cathode of diode **1024**. The cathode of diode **1022** is electrically connected to the cathode of diode **1026** and inductor **1030**. Inductor **1030** may be a fixed value inductor or a variable inductor. The anode of diode **1026** is electrically connected to ground. Capacitor **1028** is electrically connected between ground and the junction of the cathodes of diode **1022** and diode **1026**. Inductor **1035** is electrically connected between ground and the anode of diode **1024**. Inductor **1035** may be a fixed value inductor or a variable inductor. Capacitor **1070**, the anode of diode **1026**, inductor **1035**, and load **1050** are electrically connected to ground. Capacitor **1070** may be placed in parallel with load **150**.

FIGS. **11** and **12** provide example embodiments of vehicle **1110**, which utilizes electricity, the vehicle employing systems of energy collection provided herein. Vehicle **1100** in FIG. **11** is shown as an automobile vehicle, but could be any means of locomotion that utilizes electricity, including a car, a train, a motorcycle, a boat, an airplane, robotic rovers, space craft, etc. Vehicle **1200** in FIG. **12** is shown as a lunar rover vehicle. In FIGS. **11** and **12**, support rod **1110**, **1210** is electrically connected to an electrical system in vehicle **1100**, **1200**. Energy collectors **130**, which may comprise graphene, silicene, and/or other like materials, are electrically connected to support rod **1110**, **1210** and may be used to supply energy to electrical circuits within the vehicle. A non-limiting use includes a top-off charge for a battery system, on-board hydrogen production, and/or assisting in the same. Energy collectors **130** may be used to augment the efficiency of the locomotion that utilizes electrical energy as well.

FIG. **13** provides an example embodiment of energy collection system **1200** in which diode **310** is used to isolate collection devices **130** from spark gap **1020** and load **150**. Collection devices **130** may comprise graphite, carbon fibers, carbon/carbon fibers, graphene, silicene, and/or other like materials, or a mixture thereof.

FIG. **14** provides an example embodiment of energy collection system **1400** in which a plurality of energy collection systems, such as that provided in FIG. **13**, are combined. Each leg consisting of collection devices **130**, which may comprise graphene, silicene, and/or other like materials, and diode **310** are connected in parallel with other legs, each leg electrically connected to trunk wire **1410**. The legs could also be connected serially. Trunk wire **1410** is electrically connected to a collection circuit, which may comprise load **150** and spark gap **1020** in any configuration that has been previously discussed. Each leg may comprise one or more collection devices **130** and at least one diode electrically connected between the collection devices and the collection circuit. Although three collection devices **130** are shown on each leg, any number of collection devices may be used. Although four legs are shown, any number of legs may be used.

FIG. **15** presents a system diagram of an example embodiment of a windmill with energy collectors, which may comprise graphene, silicene, and/or other like materials in an example embodiment. A windmill is an engine powered by the energy of wind to produce alternative forms of energy. They may, for example, be implemented as small tower mounted wind engines used to pump water on farms. The modern wind power machines used for generating electricity are more properly called wind turbines. Common applications of windmills are grain milling, water pumping, threshing, and saw mills. Over the ages, windmills have evolved into more sophisticated and efficient wind-powered water pumps and electric power generators. In an example embodiment, as provided in FIG. **10**, windmill tower **1500** of suitable height and/or propeller **1520** of windmill tower **1500** may be equipped with energy collecting fibers **1530**, **1540**, which may comprise graphene, silicene, and/or other like materials in an example embodiment. Collecting fibers **1530**, **1540** may turn windmill **1500** into a power producing asset even when there is not enough wind to turn propellers **1520**. During periods when there is enough wind to turn propellers **1520**, collecting fibers **1530**, **1540** may supplement/boost the amount of energy the windmill produces.

A windmill is an engine powered by the energy of wind to produce alternative forms of energy. They may, for example, be implemented as small tower mounted wind engines used to pump water on farms. The modern wind power machines used for generating electricity are more properly called wind turbines. Common applications of windmills are grain milling, water pumping, threshing, and saw mills. Over the ages, windmills have evolved into more sophisticated and efficient wind-powered water pumps and electric power generators. In an example embodiment, as provided in FIG. **10**, windmill tower **1000** of suitable height and/or propeller **1020** of windmill tower **1000** may be equipped with energy collecting fibers **1030**, **1040**. Collecting fibers **1030**, **1040** may turn windmill **1000** into a power producing asset even when there is not enough wind to turn propellers **1020**. During periods when there is enough wind to turn propellers **1020**, collecting fibers **1030**, **1040** may supplement/boost the amount of energy the windmill produces.

Windmill **1500**, properly equipped with ion collectors **1530**, **1540**, such as a non-limiting example of fibers with graphene, silicene, and/or other like materials, can produce electricity: 1) by virtue of providing altitude to the fiber to harvest ions, and 2) while the propeller is turning, by virtue of wind blowing over the fiber producing electricity, among other reasons, via the triboelectric effect (however, it is also possible for the triboelectric effect to occur, producing electricity, in winds too weak to turn the propeller).

There are at least two ways that energy collectors may be employed on or in a windmill propeller to harvest energy. Propellers **1520** may be equipped with energy collectors

1530, **1540** attached to, or supported by, propeller **1520** with wires (or metal embedded in, or on propeller **1520**) electrically connecting energy collectors **1530**, **1540**, which may comprise graphene, silicene, and/or other like materials, to a load or power conversion circuit. There may be a requirement to electrically isolate energy collectors **1530**, **1540**, which are added to propeller **1520**, from electrical ground, so that the energy collected does not short to ground through propeller **1520** itself or through support tower **1510**, but rather is conveyed to the load or power conversion circuit. Energy collectors may be connected to the end of propellers **1520** such as collectors **1530**. Alternatively, energy collectors may be connected to the sides of propellers **1520** such as collectors **1540**.

Alternatively, propeller **1520** may be constructed of carbon fiber or other suitable material, with wires (or the structural metal supporting propeller **1520** may be used) electrically connecting to a load or power conversion circuit. In the case of propeller **1520** itself being constructed of carbon fiber, for example, the fiber may be 'rough finished' in selected areas so that the fiber is "fuzzy." For example, small portions of it may protrude into the air as a means of enhancing collection efficiency. The fuzzy parts of collectors **1530**, **1540** may do much of the collecting. There may be a requirement to electrically isolate carbon fiber propeller **1520** from electrical ground, so that the energy it collects does not short to ground through metal support tower **1510**, but rather is conveyed to the load or power conversion circuit. Diodes may be implemented within the circuit to prevent the backflow of energy, although diodes may not be necessary in some applications.

In an alternative embodiment, windmill **1500** may be used as a base on which to secure an even higher extension tower to support the energy collectors and/or horizontal supports extending out from tower **1510** to support the energy collectors. Electrical energy may be generated via ion collection due to altitude and also when a breeze or wind blows over the collectors supported by tower **1510**.

In alternative embodiments to windmill **1500**, other non-limiting example support structures include airplanes, drones, blimps, balloons, kites, satellites, cars, boats, trucks, (including automobile and other transportation conveyance tires), trains, motorcycles, bikes, skateboards, scooters, hovercraft (automobiles and conveyance of any kind), billboards, cell towers, radio towers, camera towers, flag poles, towers of any kind including telescopic, light poles, utility poles, water towers, buildings, sky scrapers, coliseums, roof tops, solar panel and all fixed or mobile structures exceeding 1 inch in height above ground or sea level.

An example embodiment of a support structure may also include cell phones and other electronic devices and their cases, including cases containing rechargeable batteries. For example, someone may set her cell phone or other electronic device or battery pack on the window ledge of a tall apartment building to help charge it. Other example support structures may include space stations, moon and Mars structures, rockets, planetary rovers and drones including robots and artificial intelligence entities.

Under some conditions, ambient voltage may be found to be 180-400 volts at around 6 ft, with low current. With the new generation of low current devices being developed, a hat containing ion harvesting material may provide enough charge, or supplemental charge, collected over time to help power low current devices such as future cell phones, tracking devices, GPS, audio devices, smart glasses, etc. Clothes may also be included as examples of support structures. Friction of the ion collection material (such as non-limiting examples of carbon, graphite, silicene and graphene) against unlike materials, such as wool, polyester, cotton, etc (used in clothes) may cause a voltage to be generated when rubbed together. Additionally, wind passing over the ion collection material has been demonstrated to generate voltage, even at low altitude. In an additional example embodiment, embedding collection devices into automobile tires (for example, in a particular pattern) could generate collectible voltage.

Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

It should be emphasized that the above-described embodiments of the present disclosure, particularly, any "preferred" embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A method of collecting energy comprising:

suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment from a support structure, the at least one collection device electrically connected to the support structure, the support structure comprising at least one of an airplane, drone, blimp, balloon, kite, satellite, train, motorcycle, bike, skateboard, scooter, hovercraft, electronic device, electronic device case, billboard, cell tower, radio tower, camera tower, flag pole, telescopic pole, light pole, utility pole, water tower, building, sky scraper, coliseum, roof top, solar panel and a fixed or mobile structure exceeding 1 inch in height above ground or sea level; and

providing a load with an electrical connection to the at least one collection device to draw current.

2. The method of claim **1**, wherein the collection device collects energy by triboelectric effect.

3. The method of claim **1**, wherein the collection device comprises a collection fiber.

4. The method of claim **1**, wherein the collection device comprises a diode and a collection fiber and the diode is electrically connected between the collection fiber and the load.

5. The method of claim **1**, further comprising storing energy provided to the load.

6. The method of claim **5**, wherein storing energy provided to the load comprises storing energy in a capacitor or an inductor.

7. The method of claim **3**, wherein the collection fiber comprises at least one of carbon, graphite, silicene and graphene.

8. A system of energy collection comprising:

a support structure, the support structure comprising at least one of an airplane, drone, blimp, balloon, kite, satellite, train, motorcycle, bike, skateboard, scooter, hovercraft, electronic device, electronic device case, billboard, cell tower, radio tower, camera tower, flag pole, telescopic pole, light pole, utility pole, water tower, building, sky scraper, coliseum, roof top, solar panel and a fixed or mobile structure exceeding 1 inch in height above ground or sea level;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment electrically connected to the support structure; and

a load electrically connected to the at least one collection device.

9. The system of claim **8**, wherein the collection device collects energy by triboelectric effect.

10. The system of claim **8**, wherein the collection device comprises a collection fiber.

11. The system of claim **8**, wherein the collection device comprises a collection fiber and a diode electrically connected between the load and the collection fiber.

12. The system of claim **11**, wherein the diode is elevated relative to the ground level.

13. The system of claim **10**, wherein the collection fiber comprises at least one of carbon, graphite, silicene, or graphene.

14. The system of claim **8**, further comprising a diode electrically connected between the at least one collection device and the support structure.

15. The system of claim **8**, further comprising: a switch connected in series between the at least one collection device and the load; and a capacitor connected in parallel with the switch and the load.

16. The system of claim **15**, wherein the switch comprises an interrupter connected between the load and at least one collection device, and wherein the interrupter comprises at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

17. The system of claim **8**, further comprising: a motor for providing power, the motor connected between the at least one collection device and the load; and a generator powered by the motor.

18. The system of claim **8**, further comprising a fuel cell between the support structure and the load.

19. The system of claim **18**, wherein the fuel cell produces hydrogen and oxygen.

20. A system of collecting energy comprising:

means for suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to the means for suspending, the means for suspending comprising at least one of an airplane, drone, blimp, balloon, kite, satellite, train, motorcycle, bike, skateboard, scooter, hovercraft, electronic device, electronic device case, billboard, cell tower, radio tower, camera tower, flag pole, telescopic pole, light pole, utility pole, water tower, building, sky scraper, coliseum, roof top, solar panel and a fixed or mobile structure exceeding 1 inch in height above ground or sea level;

means for inducing current flow, the means for inducing current flow electrically connected to the means for suspending; and

means for restricting the backflow of charge carriers, the means for restricting the backflow of charge carriers electrically connected between the at least one collection device and the means for inducing current flow.

* * * * *

US009479086B2

(12) **United States Patent**
McCowen

(10) **Patent No.: US 9,479,086 B2**
(45) **Date of Patent: *Oct. 25, 2016**

(54) **ENERGY COLLECTION**

(71) Applicant: **Ion Power Group LLC**, Navarre, FL (US)

(72) Inventor: **Clint McCowen**, Navarre, FL (US)

(73) Assignee: **Ion Power Group, LLC**, Navarre, FL (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **14/242,464**

(22) Filed: **Apr. 1, 2014**

(65) **Prior Publication Data**

US 2014/0210308 A1 Jul. 31, 2014

Related U.S. Application Data

(63) Continuation of application No. 12/255,130, filed on Oct. 21, 2008, now Pat. No. 8,686,575, which is a continuation of application No. 11/358,264, filed on Feb. 21, 2006, now Pat. No. 7,439,712.

(51) **Int. Cl.**

F02B 63/04	(2006.01)
F03G 7/08	(2006.01)
H02K 7/18	(2006.01)
H02N 1/10	(2006.01)
H02N 11/00	(2006.01)
H01M 16/00	(2006.01)

(52) **U.S. Cl.**
CPC ***H02N 1/10*** (2013.01); ***H02N 11/002*** (2013.01); *H01M 16/003* (2013.01)

(58) **Field of Classification Search**
USPC .... 290/1 R
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427	A		5/1901	Palencsar
911,260	A		2/1909	Pennock
1,014,719	A		1/1912	Pennock
2,473,819	A		6/1949	Pittman
3,532,959	A		10/1970	Erickson
3,663,360	A	*	5/1972	Post 376/147
3,668,065	A	*	6/1972	Moir 376/147
3,780,722	A		12/1973	Swet
3,946,227	A		3/1976	Bingham

(Continued)

FOREIGN PATENT DOCUMENTS

CA	2647385	9/2008
EP	1999767 A2	12/2008

(Continued)

OTHER PUBLICATIONS

"Emissivity Values for Common Materials & Non-Metals", Raytek; http://raytek.com/Raytek/en-r0/IREducation/EmissivityNonMetals.htm; 1999.*

Primary Examiner — Pedro J Cuevas
(74) *Attorney, Agent, or Firm* — Benjamin A. Balser; Next ip Law Group

(57) **ABSTRACT**

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support wire system supported by poles. The support wire system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but carbon and graphite are preferred. Diodes may be used to restrict the backflow or loss of energy.

21 Claims, 10 Drawing Sheets

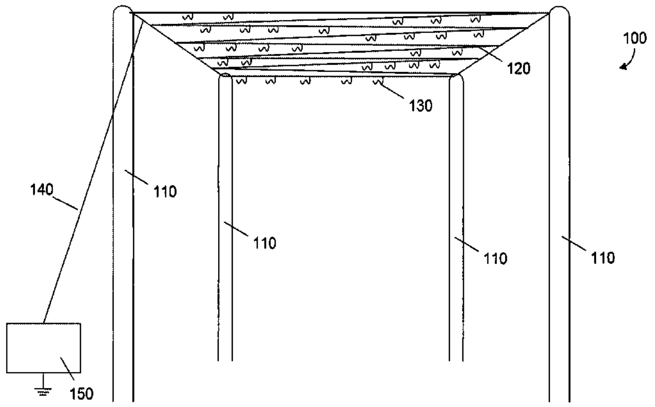

(56) References Cited

U.S. PATENT DOCUMENTS

4,079,225	A	3/1978	Warner	
4,092,250	A	5/1978	Sano et al.	
4,104,696	A	8/1978	Cochran, Jr.	
4,146,800	A	3/1979	Gregory et al.	
4,180,056	A	12/1979	Schnabel et al.	
4,201,197	A	5/1980	Dismer	
4,206,396	A	6/1980	Marks	
4,224,496	A	9/1980	Riordan et al.	
4,307,936	A	12/1981	Ochiai	
4,314,192	A	2/1982	Caro	
4,346,478	A	8/1982	Sichling	
4,425,905	A	1/1984	Mori	
4,433,248	A	2/1984	Marks	
RE31,678	E	9/1984	Ochiai	
4,483,311	A	11/1984	Whitaker	
4,489,269	A	12/1984	Edling et al.	
4,512,335	A	4/1985	Mori	
4,653,223	A	3/1987	Mori	
4,653,472	A	3/1987	Mori	
4,676,226	A	6/1987	Mori	
4,809,675	A	3/1989	Mori	
4,852,454	A	8/1989	Batchelder	
4,943,125	A	7/1990	Laundre' et al.	
4,988,159	A	1/1991	Turner et al.	
5,047,892	A	9/1991	Sakata et al.	
5,114,101	A	5/1992	Stern et al.	
5,145,257	A	9/1992	Bryant et al.	
5,379,103	A	1/1995	Zigler	
5,851,309	A	12/1998	Kousa	
5,942,806	A	8/1999	Veliadis	
6,025,591	A	2/2000	Taylor et al.	
6,038,363	A	3/2000	Slater et al.	
6,116,544	A	9/2000	Forward et al.	
6,173,922	B1	1/2001	Hoyt et al.	
6,226,440	B1	5/2001	Lyons	
6,419,191	B1	7/2002	Hoyt et al.	
6,425,391	B1	7/2002	Davoren et al.	
6,735,830	B1	5/2004	Merciel	
6,765,212	B2	7/2004	Goetz et al.	
6,846,447	B2	1/2005	Mbachu et al.	
6,853,447	B2	2/2005	Goetz	
6,894,772	B2	5/2005	Goetz et al.	
6,925,852	B2	8/2005	Susko	
6,974,110	B2	12/2005	Grandics	
7,109,597	B1	9/2006	Bose	
7,128,867	B2	10/2006	Mbachu et al.	
7,231,809	B2	6/2007	Susko	
7,259,475	B2	8/2007	Hong et al.	
7,439,712	B2	10/2008	McCowen	
7,449,668	B2	11/2008	Schutten et al.	
7,478,712	B2	1/2009	McCowen	
7,532,819	B1	5/2009	Triebes et al.	
8,269,401	B1	9/2012	Kim et al.	
8,686,575	B2	4/2014	McCowen	
8,810,049	B2 *	8/2014	McCowen	290/1 R
2002/0026933	A1	3/2002	Gottlieb	
2002/0108892	A1	8/2002	Goetz et al.	
2002/0109094	A1	8/2002	Goetz et al.	
2002/0109835	A1	8/2002	Goetz	
2003/0125630	A1	7/2003	Furnish	
2003/0193319	A1	10/2003	Wood et al.	
2004/0011925	A1	1/2004	Grandics	
2004/0083793	A1	5/2004	Susko	
2004/0094853	A1	5/2004	Mbachu et al.	
2004/0160711	A1	8/2004	Stumberger	
2004/0212945	A1	10/2004	Sprenger et al.	
2005/0101024	A1	5/2005	Mbachu et al.	
2005/0136311	A1	6/2005	Ueda et al.	
2005/0270525	A1	12/2005	Susko	
2007/0107765	A1	5/2007	Schutten et al.	
2007/0195481	A1	8/2007	McCowen	
2007/0218323	A1	9/2007	Sudo et al.	
2007/0223171	A1	9/2007	Guy et al.	
2007/0273206	A1	11/2007	McCowen	
2007/0297892	A1	12/2007	Kildegaard	
2008/0145714	A1	6/2008	Kagami	
2009/0040680	A1	2/2009	McCowen	
2009/0107842	A1	4/2009	Park et al.	
2009/0114495	A1	5/2009	McCowen	
2009/0216292	A1	8/2009	Pless et al.	
2010/0090562	A1	4/2010	McCowen	
2011/0148248	A1	6/2011	Landa	
2012/0043585	A1	2/2012	Guo et al.	
2012/0299559	A1	11/2012	McCowen	
2013/0049531	A1	2/2013	Wang et al.	

FOREIGN PATENT DOCUMENTS

RU	2008137622	2/2007
WO	2007098341	8/2007

* cited by examiner

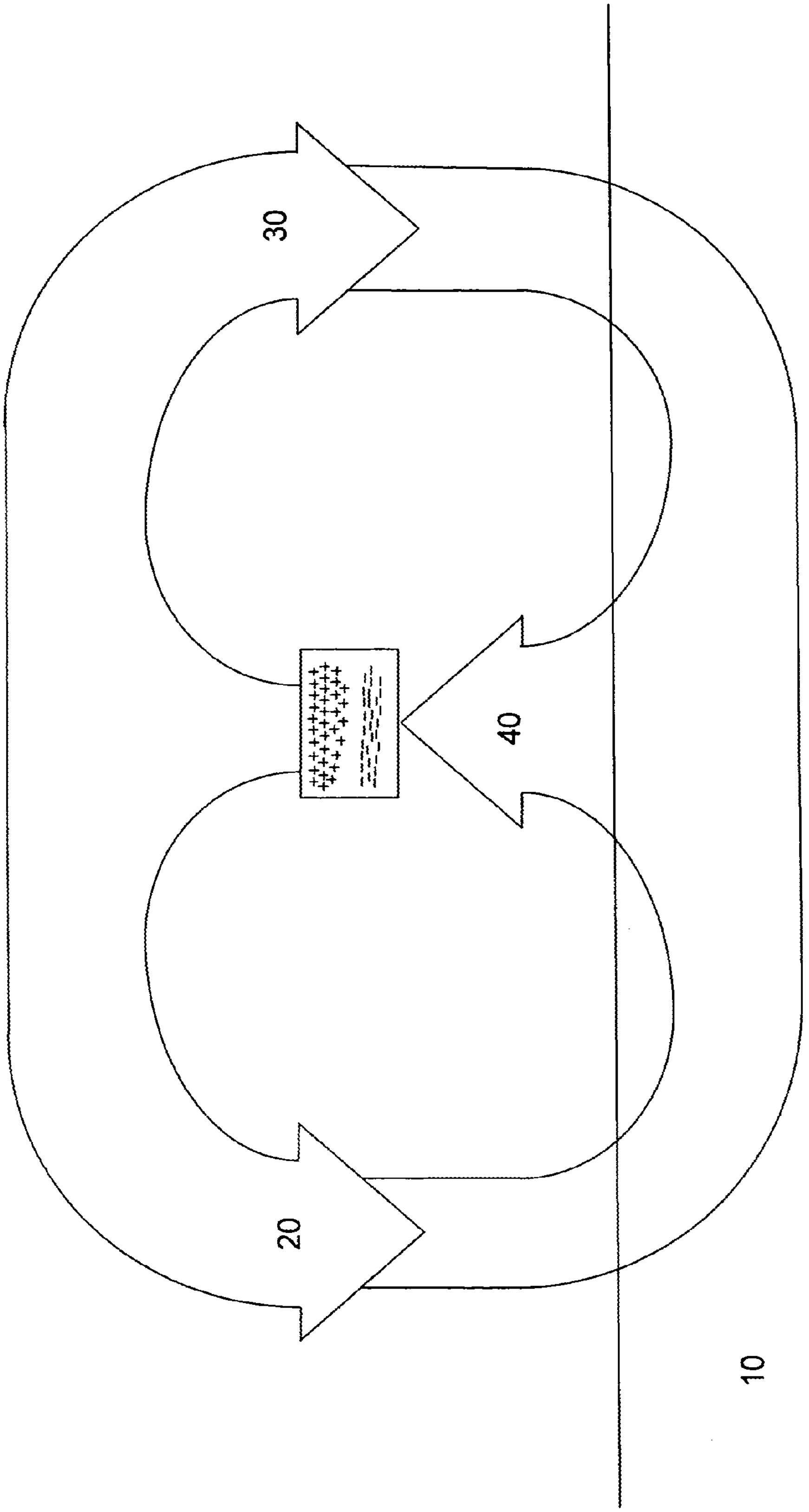

FIGURE 1

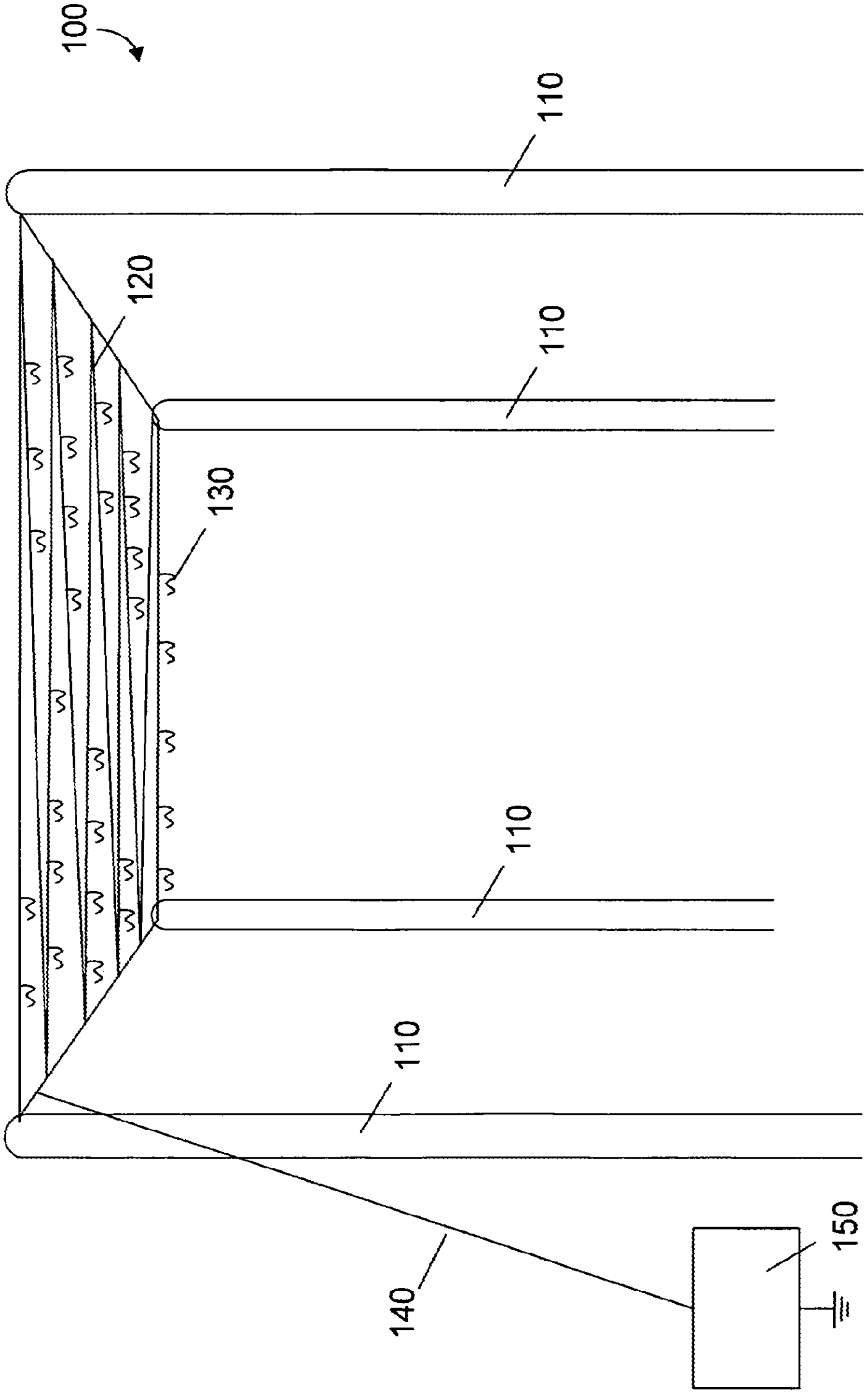
100
110
120
110
130
110
110
140
150

FIGURE 2

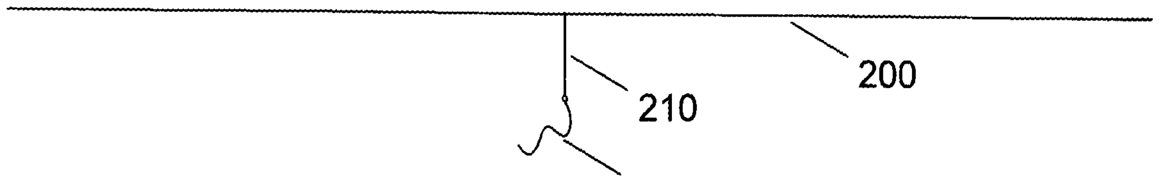

FIGURE 2A

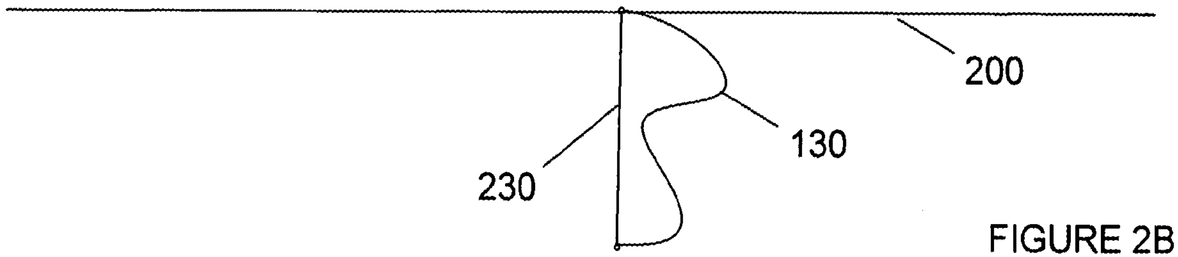

FIGURE 2B

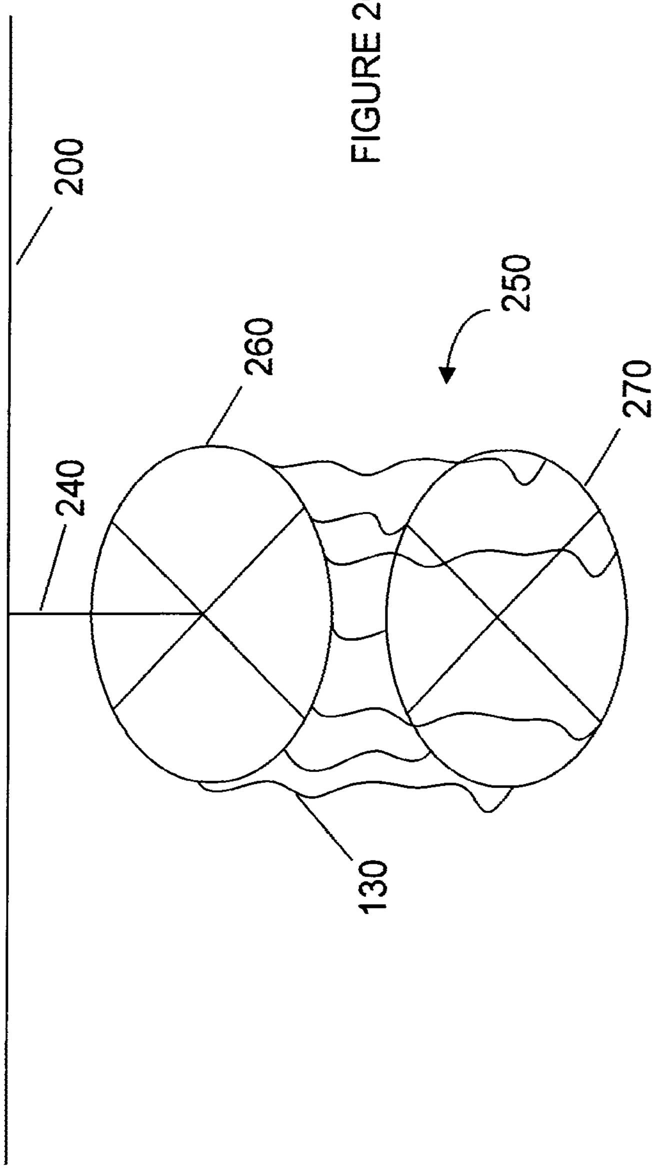
200
250
260
270
240
130

FIGURE 2C

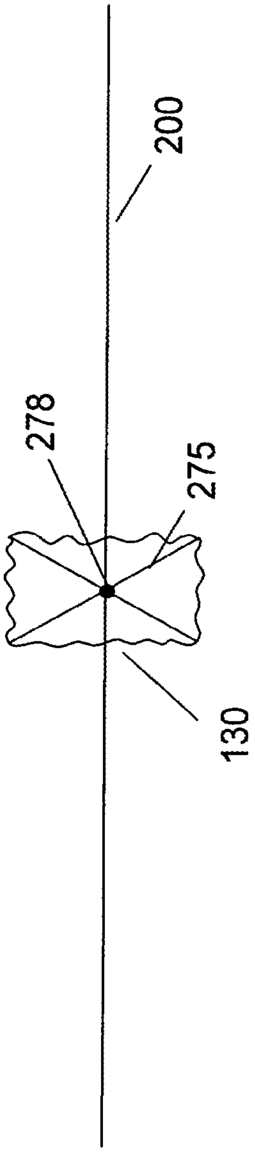
200
278
275
130

FIGURE 2D

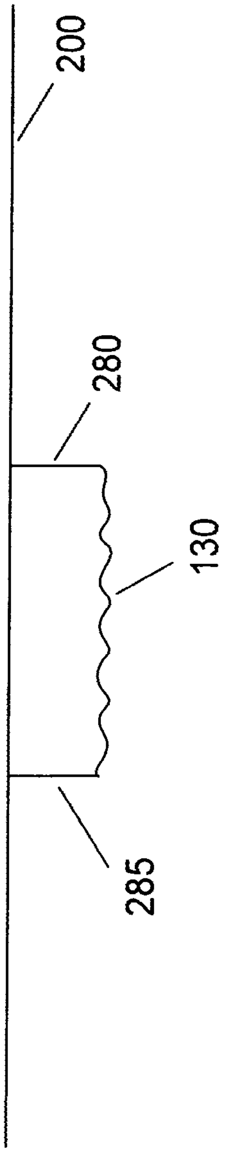
200
280
130
285

FIGURE 2E

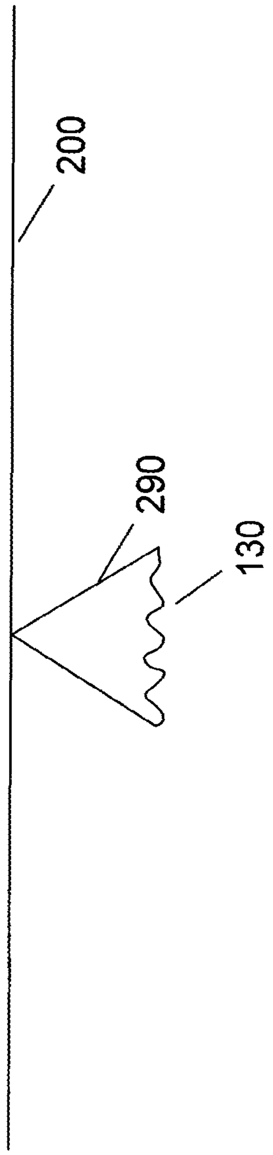

FIGURE 2F

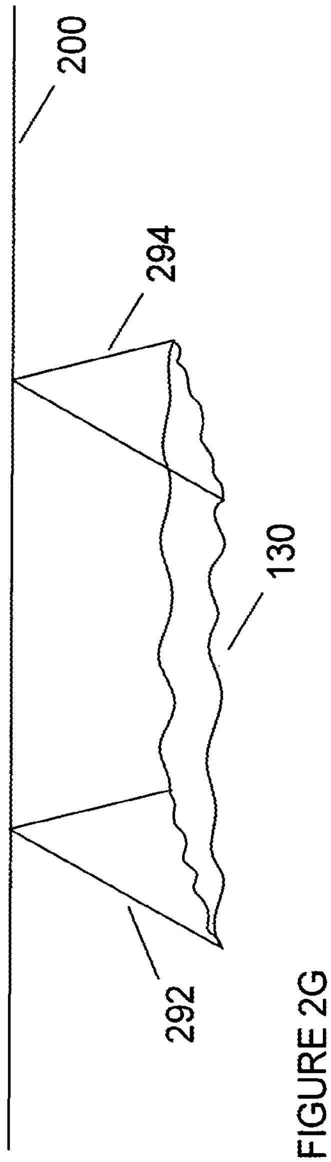
200
294
130
292

FIGURE 2G

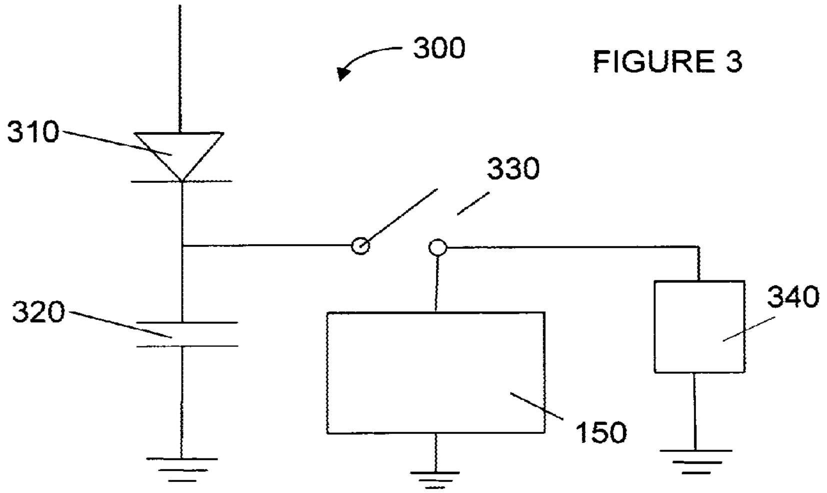
300
310
330
320
340
150

FIGURE 3

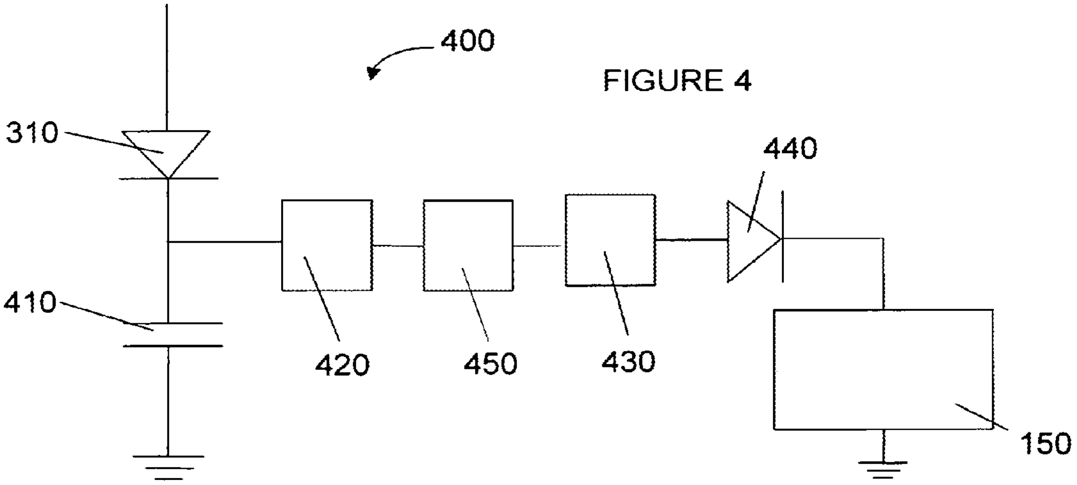
400
310
440
410
420
450
430
150

FIGURE 4

FIGURE 5

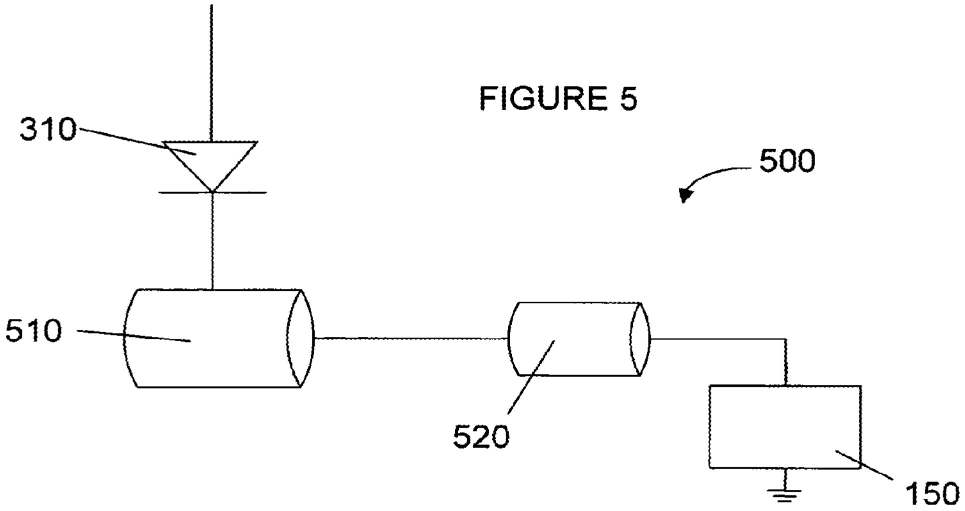
310
500
510
520
150

FIGURE 6

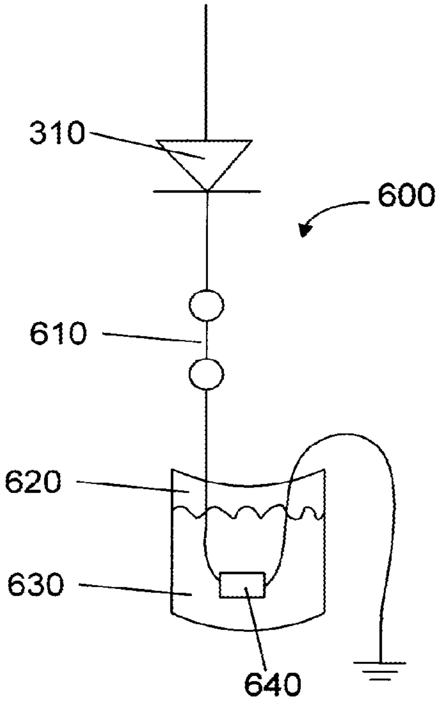
310
600
610
620
630
640

FIGURE 7

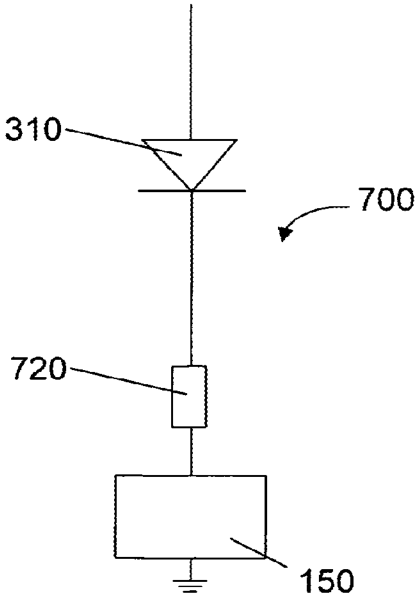
310
700
720
150

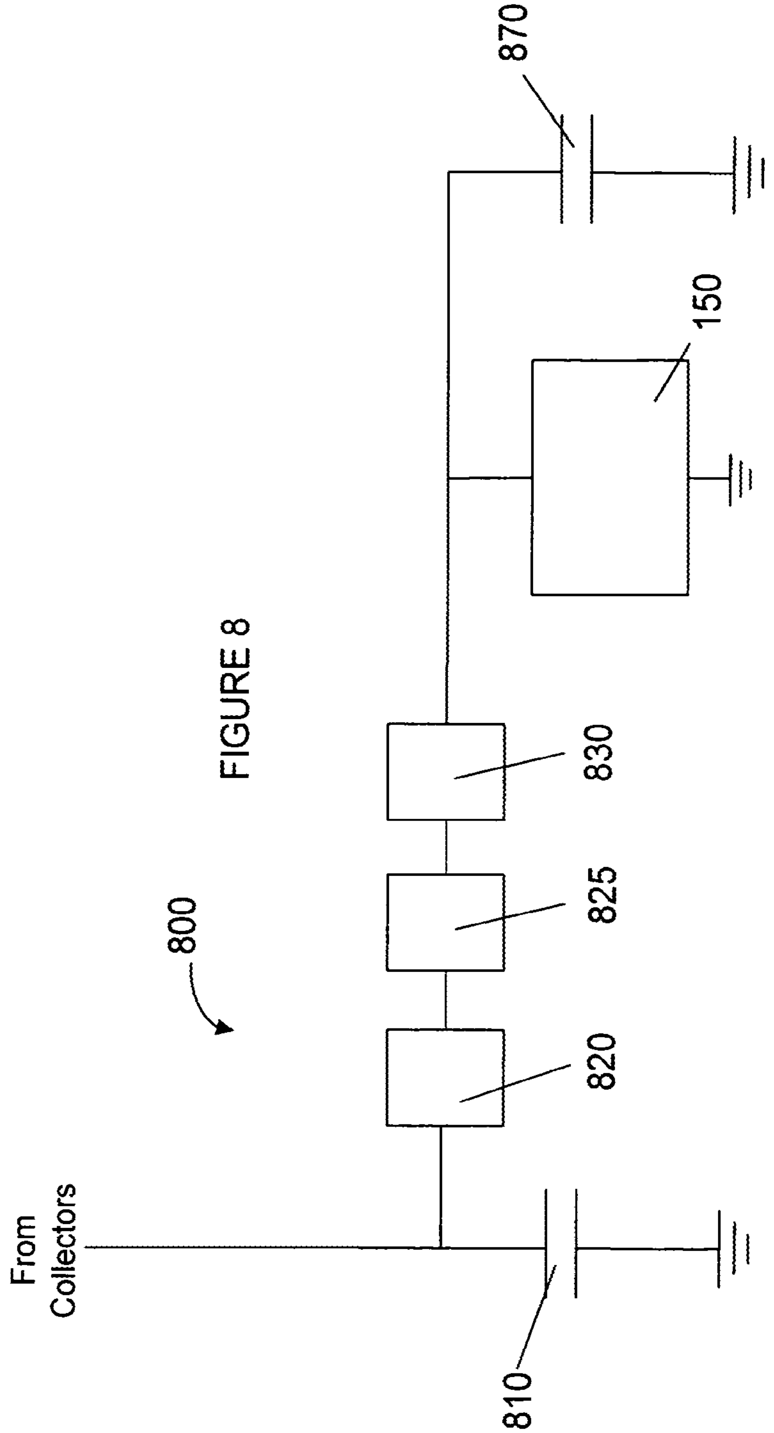
870
150
830
825
800
820
From Collectors
810

FIGURE 8

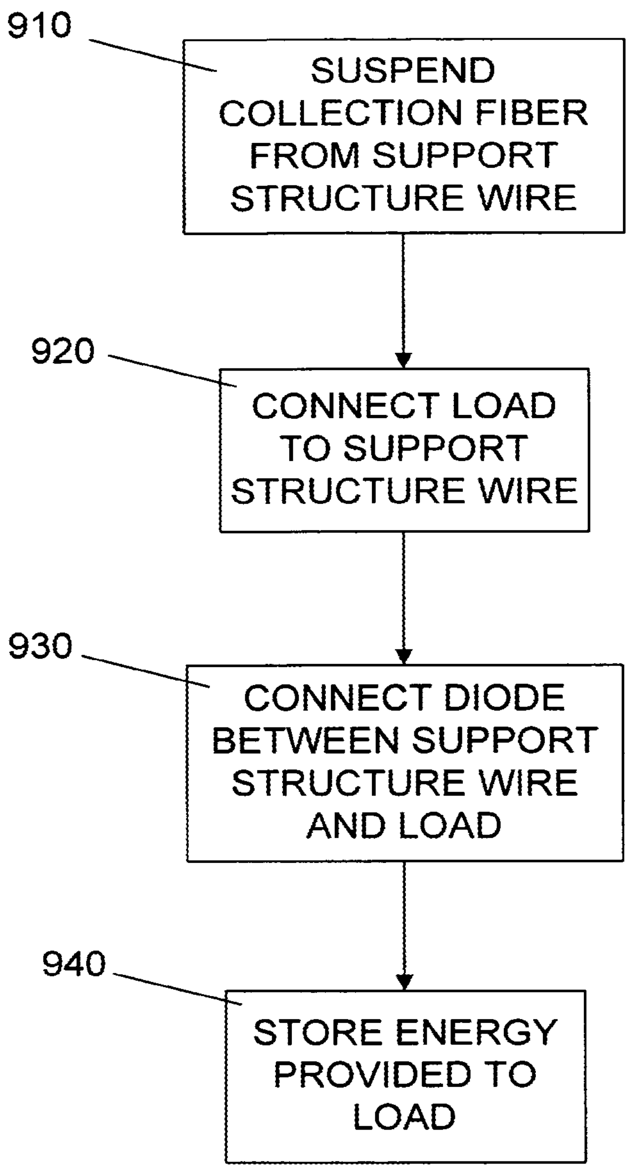
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

ENERGY COLLECTION

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation application of U.S. patent application Ser. No. 12/255,130, filed on Oct. 21, 2008 (now U.S. Pat. No. 8,686,575, issued on Apr. 1, 2014), which is a continuation application of U.S. patent application Ser. No. 11/358,264, filed on Feb. 21, 2006 (now U.S. Pat. No. 7,439,712, issued on Oct. 21, 2008), which are incorporated by reference herein.

TECHNICAL FIELD

The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure wire elevated above a ground level, at least one collection fiber electrically connected to the support structure wire; a load electrically connected to the support structure wire; and a diode electrically connected between the load and at least one collection fiber.

Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection fiber from a support structure wire elevated above ground level, the fiber electrically connected to the support structure wire; providing a load with an electrical connection to the support structure wire to draw current; and providing a diode electrically connected between the collection fiber and the load.

Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. **1** is a circuit diagram of a weather energy circuit.

FIG. **2** is a perspective view of an exemplary embodiment of many energy collectors elevated above ground by a structure.

FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

FIG. **2**B is a side view of an exemplary embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

FIG. **2**D is a side view of an exemplary embodiment of a support structure for multiple energy collection fibers.

FIG. **2**E is a side view of a support structure for an energy collection fiber.

FIG. **2**F is a side view of an exemplary embodiment of a support structure for an energy collection fiber.

FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

FIG. **3** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **4** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **5** is a circuit diagram of an exemplary embodiment of an energy collection circuit for powering a generator and motor.

FIG. **6** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

FIG. **7** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

FIG. **8** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy.

FIG. **9** is a flow diagram of an exemplary embodiment of collecting energy with a collection fiber.

DETAILED DESCRIPTION

Electric charges on conductors reside entirely on the external surface of the conductors, and tend to concentrate more around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An exemplary embodiment of this disclosure takes advantage of this property, among others, to collect and use the energy generated by an electric field in the atmosphere. Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An exemplary embodiment utilizes carbon or graphite fibers for static electricity collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including aluminum or steel, but most notably, copper. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

An exemplary embodiment of the collection fibers as collection device **130** includes graphite or carbon fibers. Graphite and carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, or graphite fibers are examples that demonstrate exemplary collection ability.

In at least one exemplary embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than **100** volts per meter under some conditions. When support wire **120** is suspended in the air at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An exemplary embodiment of the collection device comprises the diode with no collection fiber. A preferred embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restricts energy that may be collected through one fiber from escaping through another fiber.

Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2**A-**2**G using a collection fiber embodiment. FIG. **2**A presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**.

Likewise, FIG. **2**B shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

FIG. **2**C presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

FIG. **2**D presents another exemplary embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2**E provides another exemplary embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

FIG. **2**F presents another exemplary embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2**G provides two supports as provided in FIG. **2**F such that at least two support members **292**, **294** may be connected to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and between multiple support structures. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**.

FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load. Capacitor **320** maybe connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

In an exemplary embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel. Non-limiting examples of spark gap **420** include mercury-reed switches and mercury-wetted reed switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground. The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes maybe placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches and mercury-wetted reed switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel.

FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit. Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

FIG. **8** presents exemplary circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches and mercury-wetted reed switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

It should be emphasized that the above-described embodiments of the present disclosure, particularly, any "preferred" embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A method of collecting energy comprising:

suspending at least one ion collection device with, in operation, microscopic points of a cross section of the collection device exposed to the environment from a support structure; and

providing a load with an electrical connection to the at least one collection device to draw current.

2. The method of claim **1**, wherein the ion collection device comprises a diode.

3. The method of claim **1**, wherein the ion collection device comprises a collection fiber.

4. The method of claim **3**, wherein the collection fiber comprises carbon fiber and/or graphite fiber.

5. The method of claim **1**, wherein the ion collection device comprises a diode and a collection fiber and the diode is electrically connected between the collection fiber and the load.

6. The method of claim **1**, further comprising storing energy provided to the load in an energy storage device.

7. The method of claim **6**, wherein storing energy provided to the load comprises storing energy in a capacitor or an inductor.

8. A system of energy collection comprising:

a support structure;

at least one ion collection device with, in operation, microscopic points of a cross section of the collection device exposed to the environment, the ion collection device electrically connected to and suspended from the support structure; and

a load electrically connected to the at least one ion collection device.

9. The system of claim **8**, wherein the ion collection device comprises a diode.

10. The system of claim **9**, wherein the diode is elevated relative to the ground level.

11. The system of claim **10**, wherein the collection fiber comprises a carbon fiber or a graphite fiber.

12. The system of claim **8**, wherein the ion collection device comprises a collection fiber.

13. The system of claim **8**, wherein the ion collection device comprises a collection fiber and a diode electrically connected between the load and the collection fiber.

14. The system of claim **8**, further comprising a diode electrically connected between the at least one ion collection device and the support structure.

15. The system of claim **8**, further comprising:

a switch connected in series between the at least one ion collection device and the load; and

a capacitor connected in parallel with the switch and the load.

16. The system of claim **15**, wherein the switch comprises an interrupter connected between the at least one ion collection device and the load, the interrupter comprising at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

17. The system of claim **16**, further comprising a transformer connected between the interrupter and the load.

18. The system of claim **8**, further comprising: a motor for providing power, the motor connected between the at least one ion collection device and the load; and a generator powered by the motor.

19. The system of claim **8**, further comprising a fuel cell electrically connected between the ion collection device and the load.

20. The system of claim **19**, wherein the fuel cell produces hydrogen and oxygen.

21. A system of collecting energy comprising:

means for suspending at least one ion collection device with, in operation, microscopic points of a cross section of the collection device exposed to the environment;

means for inducing current flow, the means for inducing current flow electrically connected to at least one ion collection device; and

means for restricting the backflow of charge carriers, the means for restricting the backflow of charge carriers electrically connected between the at least one collection device and the means for inducing current flow.

* * * * *

US 20210036500A1

(19) **United States**

(12) **Patent Application Publication**

McCowen

(10) **Pub. No.: US 2021/0036500 A1**

(43) **Pub. Date: Feb. 4, 2021**

(54) **SYSTEMS AND METHODS OF LAUNCH PLATFORM ISOLATION**

(71) Applicant: **Clint McCowen**, Navarre, FL (US)

(72) Inventor: **Clint McCowen**, Navarre, FL (US)

(21) Appl. No.: **16/574,081**

(22) Filed: **Sep. 18, 2019**

Related U.S. Application Data

(60) Provisional application No. 62/732,100, filed on Sep. 17, 2018.

Publication Classification

(51) **Int. Cl.**

H02G 13/00	(2006.01)
B64B 1/50	(2006.01)

(52) **U.S. Cl.**

CPC ***H02G 13/40*** (2013.01); *H02G 13/20* (2013.01); ***H02G 13/80*** (2013.01); ***B64B 1/50*** (2013.01)

(57) **ABSTRACT**

Ion Harvesting Technology harvests high voltage electricity, including from atmospheric ions. A wire, or conductive tether, may be used to connect ion harvesting material (typically carbon, but including any materials such as metals, metamaterials, or others) located on or near an aerial platform to an anchor point. Because the harvested electricity is typically of high voltage, the electricity may arc between the conductive tether to nearby points of lesser or greater voltage. Such arcing represents a loss of power to the overall system, causing the overall system to be less efficient, or possibly non-operational, and in some cases may cause catastrophic system failure. Electrical isolators may be used to prevent the losses from the arcing.

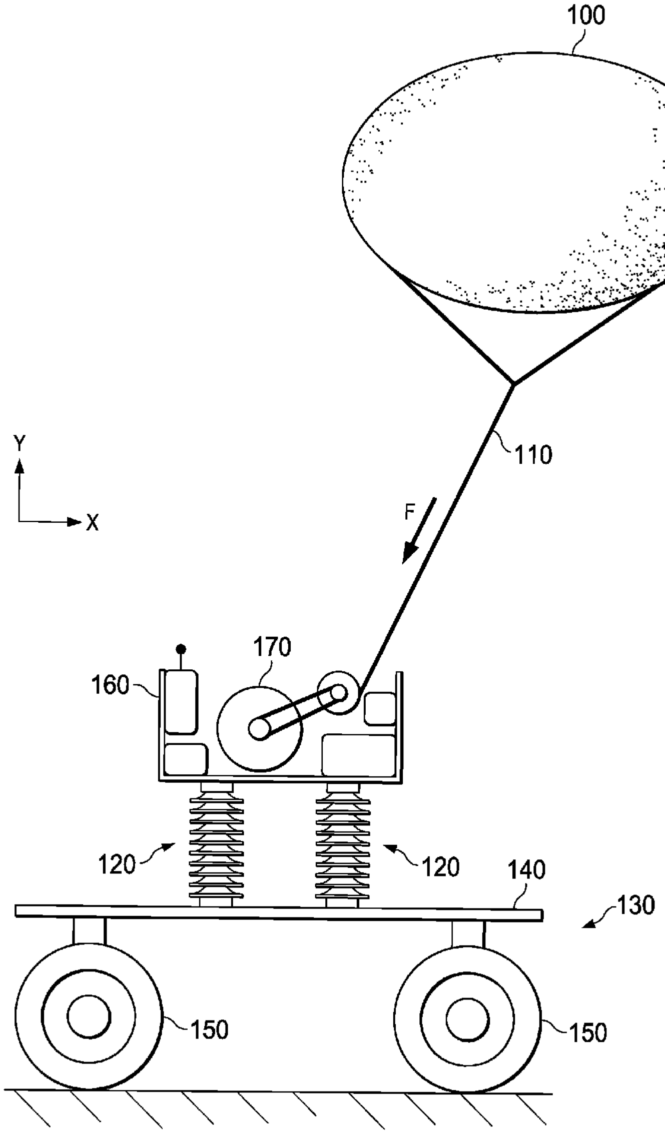

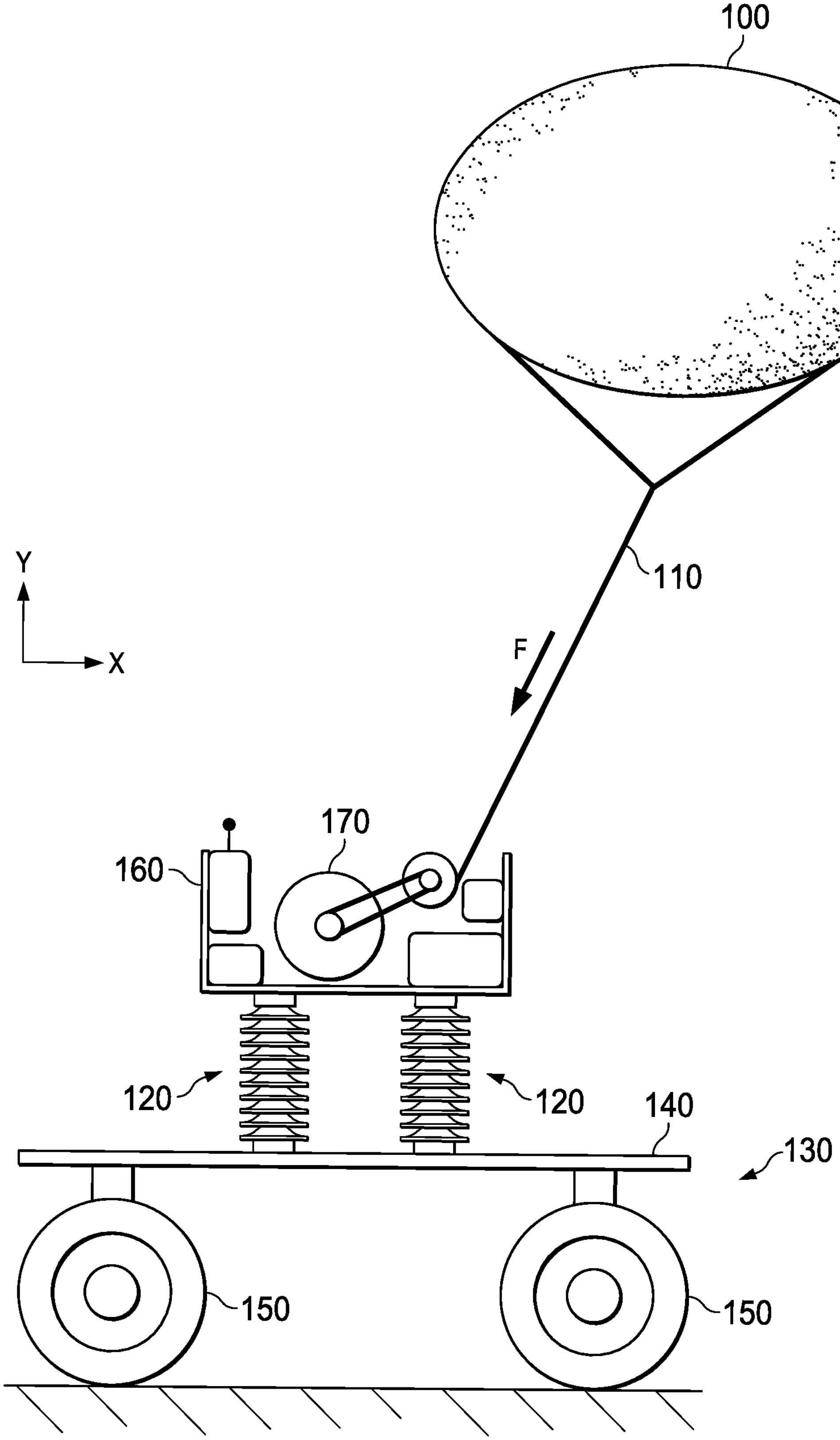
100
110
F
Y
X
170
160
120
120
140
130
150
150

FIG. 1

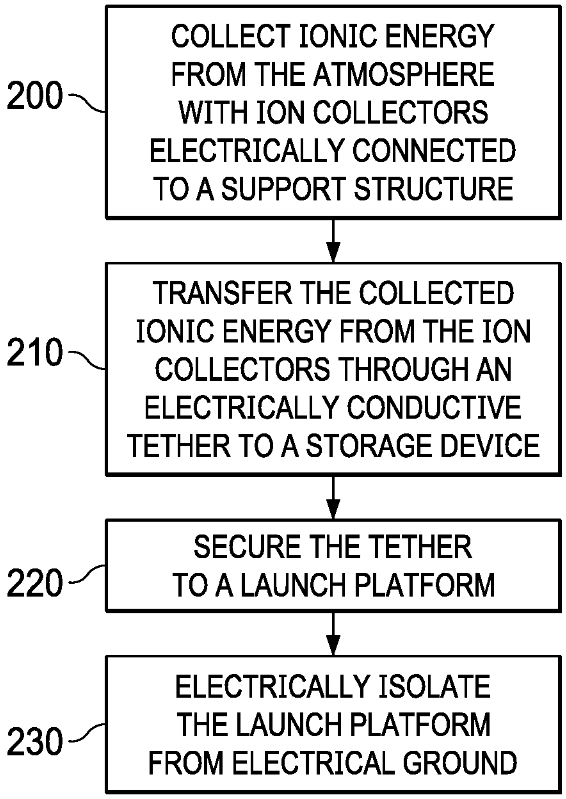
200
COLLECT IONIC ENERGY FROM THE ATMOSPHERE WITH ION COLLECTORS ELECTRICALLY CONNECTED TO A SUPPORT STRUCTURE
210
TRANSFER THE COLLECTED IONIC ENERGY FROM THE ION COLLECTORS THROUGH AN ELECTRICALLY CONDUCTIVE TETHER TO A STORAGE DEVICE
220
SECURE THE TETHER TO A LAUNCH PLATFORM
230
ELECTRICALLY ISOLATE THE LAUNCH PLATFORM FROM ELECTRICAL GROUND

FIG. 2

SYSTEMS AND METHODS OF LAUNCH PLATFORM ISOLATION

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims benefit to U.S. provisional patent application Ser. No. 62/732,100, filed on Sep. 17, 2018, which is incorporated by reference herein.

TECHNICAL FIELD

[0002] The present disclosure is generally related to energy and, more particularly, is related to systems and methods of isolating a launch platform.

BACKGROUND

[0003] The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightning storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy.

[0004] In a lightning storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

[0005] Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

[0006] Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

[0007] While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

[0008] Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by at least one ion collection device electrically connected to a support structure; a conductive tether electrically connected to the support structure; a launch platform configured to secure the support structure; and electrical isolators configured to isolate the launch platform from electrical ground.

[0009] Embodiments of the present disclosure can also be viewed as providing methods for isolating a launch platform. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: collecting, ionic energy from the atmosphere with ion collectors electrically connected to a support structure; transferring the collected ionic energy from the ion collectors through an electrically conductive tether to a storage device; securing the tether to a launch platform; and electrically isolating the launch platform from electrical ground.

[0010] Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] FIG. **1** is a system diagram of an example embodiment of a system of launch platform isolation.

[0012] FIG. **2** is a flow diagram of an example embodiment of a method of launch platform isolation.

DETAILED DESCRIPTION

[0013] Embodiments of the present disclosure will be described more fully hereinafter with reference to the accompanying drawings in which like numerals represent like elements throughout the several figures, and in which example embodiments are shown. Embodiments of the claims may, however, be embodied in many different forms and should not be construed as limited to the embodiments set forth herein. The examples set forth herein are non-limiting examples and are merely examples among other possible examples.

[0014] Ion Harvesting Technology harvests high voltage electricity, including from atmospheric ions. A wire, or conductive tether, may be used to connect ion harvesting material (typically carbon, but including any materials such as metals, metamaterials, or others) located on or near an aerial platform to an anchor point. In an example embodiment, the anchor point is a mechanical winch, but it may be any type of anchor point located above, near, on, or in the planet's ground. In an alternative embodiment, the anchor point includes attachment to aircraft, vehicles, drones, space stations, and space elevators, among others whether or not they are connected or related to planet ground. Because the harvested electricity is typically of high voltage, the electricity may arc between the conductive tether to nearby points of lesser or greater voltage. Such arcing represents a loss of power to the overall system, causing the overall system to be less efficient or possibly non-operational, and in some cases may cause catastrophic system failure.

[0015] FIG. **1** provides a system diagram of an example embodiment of a system of launch platform isolation. Aerial device **100** comprising ion harvesting material is connected to platform **130** with conductive tether **110**. Platform **130** may comprise one or more of wireless transmitter **160**, winch **170**, electrical isolators **120**, table **140**, and wheels **150**. Other possible elements include, but are not limited to, motor, gearbox, resistors, capacitors, batteries, energy storage modules, motor controller, wireless data transmitter, electrical measurement device, data recorders, and other support equipment.

[0016] To avoid the arcing problem, conductive tether **110** may be located away from objects of lesser or greater voltage potentials. For example, if winch **170** on which conductive tether **110** is rolled is not electrically isolated from the planet's ground (or artificial ground of greater or lesser potential), then arcing may occur. Utilizing a standard aerostat launching platform may not be suitable, nor compatible, with Ion Harvesting Technology as winch **170** and/or launch platform **130** may not be adequately electrically isolated from the planet's ground, potentially causing arcing.

[0017] In an example embodiment, winch **170** or other anchor device to which tether **110** is connected (including the 'drum' on which tether **170** is rolled) is electrically isolated from the planet's ground. Additionally, any mechanical or electrical connections to winch **170** may be likewise electrically isolated from the planet's ground. Failure to provide adequate electrical isolation to winch **170** or drum or other anchor devices connected to conductive tether **110** may result in catastrophic failure or undesirable operation of the entire Ion Harvesting System.

[0018] Example embodiments provided herein may utilize electrical isolators similar, but not necessarily identical, to electrical isolators used to provide electrical isolation to high voltage transmission lines on Earth grid systems. However, other designs are also included in this disclosure. Alternative embodiments include any rope, cable, pipe, straps, extensions, or the like or liquids, gases, materials, or metamaterials that possess sufficient electrical resistivity to prevent arcing or electrical leakage.

[0019] In an alternative embodiment, the winch support structure includes a point to which electrical isolation is provided.

[0020] In an alternative embodiment, launch platform **130**, including winch **170**, is electrically isolated from the planet's ground through driving or positioning the launch vehicle on another platform or area designed to provide electrical isolation to whatever is placed on it, including winch **170** or other anchor point with or without a launch platform or vehicle. Examples include an area, surface, deck, platform, or mat onto which launch platform **130** is positioned (or driven or towed) so as to provide a surface of high electrical resistivity to provide enhanced electrical isolation to launch platform **130**, including winch **170**. The insulating surface may be affixed or connected, permanently or temporarily, to launch platform **130**, or winch **170**, or any component thereof in order to provide suitable isolation.

[0021] Another example embodiment includes extensions made of high electrical resistivity used to separate or lift launch platform **130** away from the planet's surface. The definition of planet's ground herein, includes any form of ground, water, liquid, marsh, ice, mud, or any natural or man-made surface.

[0022] Another example embodiment, includes utilizing liquid, gas, or metamaterial of sufficient electrical resistivity to prevent arcing or electrical leakage on which winch **170** or launch platform **130** (anchor) may be placed upon or to float upon or in, thereby providing electrical isolation from the planet's ground, ocean, or surface.

[0023] An example embodiment of aerial platform **100** includes but is not limited to aerostats, balloons, kites, helikites, towers, poles, buildings, objects above the planet, in the atmosphere, or in space to which tether **110** is connected (including but not limited to a space elevator), and any device or means of providing altitude greater than 1 cm above the planet's surface or combinations thereof.

[0024] FIG. **2** provides a flow diagram of an example embodiment of a method of launch platform isolation. In block **200**, ionic energy is collected from the atmosphere with ion collectors electrically connected to a support structure. In block **210**, the collected ionic energy is transferred from the ion collectors through an electrically conductive tether to a storage device. In block **220**, a tether is secured to a launch platform. In block **230**, the launch platform is electrically isolated from electrical ground.

[0025] Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

[0026] It should be emphasized that the above-described embodiments of the present disclosure, particularly, any "preferred" embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A system comprising:

at least one ion collection device electrically connected to a support structure;

a conductive tether electrically connected to the support structure;

a launch platform configured to secure the support structure; and

electrical isolators configured to isolate the launch platform from electrical ground.

2. The system of claim **1**, wherein the ion collection device comprises at least one of carbon, graphite, graphene, and silicene.

3. The system of claim **1**, wherein launch platform comprises a winch configured to manipulate the altitude of the support structure.

4. The system of claim **1**, wherein the support structure comprises an aerial platform, the aerial platform comprising at least one of an aerostat, balloon kite, helikite, tower, pole, and building.

5. The system of claim **1**, wherein the electrical isolator at least one of an electrical, mechanical, and optical isolation barrier.

6. The system of claim **1**, wherein the electrical isolator comprises at least one of rope, cable, pipe, straps, extensions, liquids, gases, material, and metamaterial configure to prevent arcing or electrical leakage.

7. The system of claim **3**, wherein the electrical isolator is provided by positioning the launch platform in an area configured to provide electrical isolation to an item in the area.

8. A method comprising:

collecting ionic energy from the atmosphere with ion collectors electrically connected to a support structure;

transferring the collected ionic energy from the ion collectors through an electrically conductive tether to a storage device;

securing the tether to a launch platform; and

electrically isolating the launch platform from electrical ground.

9. The method of claim **8**, wherein the ion collectors comprise at least one of at least one of carbon, graphite, graphene, and silicene.

10. The method of claim **8**, further comprising manipulating the altitude of the support structure with a winch.

11. The method of claim **8**, wherein the support structure comprises an aerial platform, the aerial platform comprising at least one of an aerostat, balloon kite, helikite, tower, pole, and building.

12. The method of claim **9**, wherein electrically isolating the launch platform comprising isolating with at least one of an electrical, mechanical, and optical isolation barrier.

13. The method of claim **10**, wherein electrically isolating the platform comprises implementing at least one of rope, cable, pipe, straps, extensions, liquids, gases, material, and metamaterial configure to prevent arcing or electrical leakage.

14. The method of claim **8**, further comprising positioning the launch platform in an area configured to provide electrical isolation to an item in the area.

15. A system comprising:

a launch platform configured to secure a support structure supporting ionic energy collectors; and

electrical isolators configured to isolate the launch platform from electrical ground.

16. The system of claim **15**, wherein the ion collectors comprise at least one of at least one of carbon, graphite, graphene, and silicene.

17. The system of claim **15**, further comprising a winch configured to manipulate the altitude of the support structure.

18. The system of claim **15**, wherein the support structure comprises an aerial platform, the aerial platform comprising at least one of an aerostat, balloon kite, helikite, tower, pole, and building.

19. The system of claim **15**, wherein the electrical isolators comprise at least one of an electrical, mechanical, and optical isolation barrier.

20. The system of claim **15**, wherein the electrical isolators comprise at least one of rope, cable, pipe, straps, extensions, liquids, gases, material, and metamaterial configure to prevent arcing or electrical leakage.

* * * * *

The Spirit of Ma'at

Vol 2 Feb 2002

Methernitha

A Community That Runs
on Free Energy and Spiritual Values

by Carolina Hehenkamp

In the beautiful mountains of Switzerland lies Linden, a small, calm place surrounded by dozens of peaceful little villages.

The region is called Emmental, known for its Emmental cheese.

Linden itself is especially known for the Methernitha group that lives there, and for their Thestatika machine, built by Methernitha's founder, Paul Baumann.

This machine attracts many people from all over the world, for it produces electricity from what is called "free energy." Many think this Thestatika machine could be the answer to our world energy needs, but no-one outside of the Methernitha group knows how it works. And although the Methernitha people have demonstrated their machine to many scientists, the mystery of its operation remains a secret.

Methernitha — The Community

The story of Methernitha goes back a long time. Paul Baumann, its founder, had many visions that led him both to God and to the Bible. And in the fifties, he decided to form a community with others who, like himself, wanted to live their lives based on Christian principles. Thus, the Methernitha Alliance was formed, and the group started to live and work together as a social community, free from outside influences — and foreign money.

Paul Baumann grew up within the circles of the *Bruderverein* (brotherhood), a Lutheran society, and its influence can still be clearly felt. While it is emphasized that each individual must develop his own spirituality, it is still regarded as desirable that God's Law be adhered to in its entirety. For Baumann's wish was to found a cooperative clearly based upon Christian ideals. To this day many of the group (but not all) still follow strict guidelines, such as abstaining from alcohol.

The people are not organized like a club. They do not have a "guru," and there are no subscription fees. They have no desire to proselytize. They live on their own, and look after their own affairs.

Within Methernitha are departments specializing in precision mechanics, the making of appliances, gardening, and carpentry. All Methernitha members are aware of their responsibility within the social sphere and the balance of nature, and accordingly, an exemplary heating-system, based upon the incineration of wood-waste was installed.

Anyone can become a member of the cooperative who accepts its rules and is willing to do his best to supporting its aims.

The Mayor of Linden says of Methernitha:

> They work and live within their community. Their products are of high quality and are partly globally well known. There are no fences or walls between Methernitha inhabitants and the village. The inhabitants — roughly 120 people, including children — have their own free opinion and philosophy, and also have their personal freedom and freedom of action which they never try to impose on other people. On a positive note, one has to take into account that their supporters will be looked after in old age and, if necessary, be provided with medical care and social welfare until death.

So today, Methernitha is a social model, proving the quality of its ideals through having functioned so well for just over forty years.

Methernitha and Free Energy

Ever since the foundation of Methernitha, there has existed a department for research, development, and electronics, which has been concerned with the problem of alternative energy sources. Their focus is on technologies suited to exploit the inherent forces of nature, thus unlocking sources of energy without disturbing nature's ecological balance.

Many believe that humankind's technology must serve us in the long run and not just in immediate terms, but at Methernitha this idea is being fully put into practice.

Although interested in many energy producing fields, Methernithans started more than 30 years ago to concentrate their efforts on lesser known, and even generally unknown, sources of energy. The result of this scientific work was the Thestatika machine. This wonder-machine is fueled from nature — nothing else.

The Methernitha people are very specific about why they could "make" the machine, saying that it is their oneness with nature and their isolation from society that opened up the knowledge.

"We are fortunate in gaining the experience that paradoxically the most beautiful and useful results can be achieved by just using the most simple means," they say. "Never did we use any borrowed capital, because we want to stay free Swiss citizens, and not be hindered or bound in any way in the pursuit of our aims."

The Thestatika

It was not possible to speak to the Methernitha people in time to create this article, and not much is known about the Thestatika, so I will simply report what the Methernitha group writes about it on their own website:

> Two counterrotating discs generate an electrostatic charge. One disc represents the earth, the other the clouds. The charges are bound using grid electrodes. After that, they are collected by non-contacting so-called "antenna keys," and then sorted.
>
> After being initially turned on by hand, the discs rotate by themselves according to the electrostatic laws of attraction and repulsion. A rectifying diode keeps the cycles in a steady state. Otherwise, the impulses of attraction and repulsion would accumulate and cause the discs to run faster and faster — the correct speed is of great importance. For optimal power generation, the discs have to run quite steady and slow.
>
> By means of grid condensers, the energy is stored and then uniformly discharged, at the same time reducing the high voltage and building up power with additional devices.
>
> Finally, the machine supplies a uniform direct current, which varies according to the size of the model. The machine furnishes about 3-4 kW permanent output, depending on humidity, whereby the electric potential ranges from 270 to 320 volts.
>
> High humidity of the atmosphere prevents the build-up of electric potential, so the drier the air, the better. No doubt, through the so-far-achieved results, one main objective has been reached, namely, to prove that it is possible to use free energy.
>
> Nevertheless the research work is not yet completed.

The Methernithans realize that to the educated physicist, their machine may seem impossible or even crazy. But they point out that, "A trained specialist should remain free and independent in his thinking, and should avoid being limited by the temporal framework of publicly admitted knowledge in any science. It has to be remembered that the established science has already been forced many times to change or give up some of its most fundamental concepts."

The Methernithans compare today's science to the butterfly, which as a larva must shed its cocoon to reveal the beauty that lies within. This revelation, they feel, "is only possible when man becomes aware of his true role within the whole creation, and again learns to recognize his true tasks. Because the whole universe functions within a strict and precisely structured order according to the will and word of the creator. Therefore, also, man should recognize and realize the universal laws that are valid within the whole and within every part of this creation."

To evolve a free energy technology — no matter how ecologically sound this may be — is not enough.

Thestatika Demonstrations

Over the years the Thestatika has been demonstrated to various technicians and engineers who have been invited to view different versions of Thestatika machines. Most of these technicians and engineers are now retired.

Hans Holzherr from Switzerland was one of those who attended a demonstration. Here is part of his report, as translated by Stefan Harmann:

> I am referring in the following to the model with the 50-cm diameter disks. This machine was already running when the visitors stepped into the room, and was not halted during the whole time — we were there for about 1.5 hours.
>
> As a first load, a 1000-Watt lamp was connected for approximately 10 seconds, whose brightness did NOT diminish . . .
>
> The second load was a U-shaped heating element that Mr. Baumann handed to me. It became so hot within one second that I had to put it down immediately! What was particularly impressive was that while he pulled back one of the contact wires (that was with the lamp, I believe), a 1-cm long arc appeared between the output electrode and the connecting wire for approximately one second. The apparatus was under a Plexi-glass hood. Near the base it had two holes which Baumann used to insert the contact wires to touch the output electrodes. . . . .
>
> It was really impressive! One can hardly believe it, with this slow rotation. In any case, this cannot be explained in terms of bare electrostatics in the sense of the Wimshurst machine. The perforated sheets seem to have a key function. Beside the pick-up and the drive electrodes there are a number of small plexiglass blocks with glued-on perforated sheets, whose function is unknown.

There is a great deal available on various websites concerning the design and operation of the Thestatika machine. It's all very technical and some of it is quite detailed. But as those who live and work at Methernitha believe, there is probably not going to be a working prototype developed outside Methernitha except by attention to the same spiritual aspects with which the original Thestatika was developed.

As the Methernithans believe:

> To understand nature and to perceive its voice, man is obliged to experience silence and solitude. It was there that the knowledge about this technology was obtained. . . . in valleys and forests, in the mountains and on the shores of lakes, in places where, without being disturbed, one can study nature, one's own being, and the creator of all this universe in silence and concentration.

Top of Page Print Version

July 23, 1963 R. M. RICHARDSON 3,098,971

REMOTELY ACTUATED RADIO FREQUENCY POWERED DEVICES

Filed Sept. 26, 1961

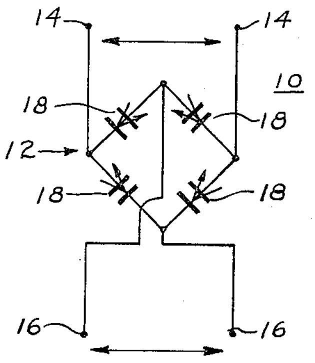
14
14
10
18
18
12
18
18
16
16

Fig. 1

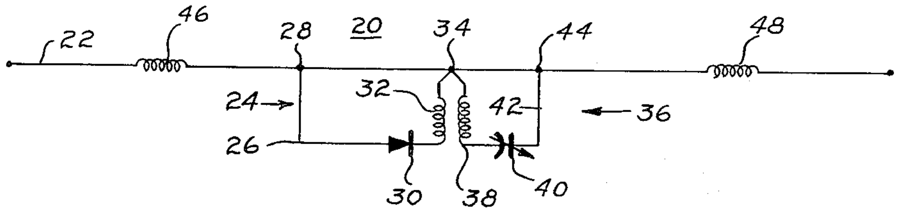
46
22
28
20
34
44
48
24
32
42
36
26
30
38
40

Fig. 2

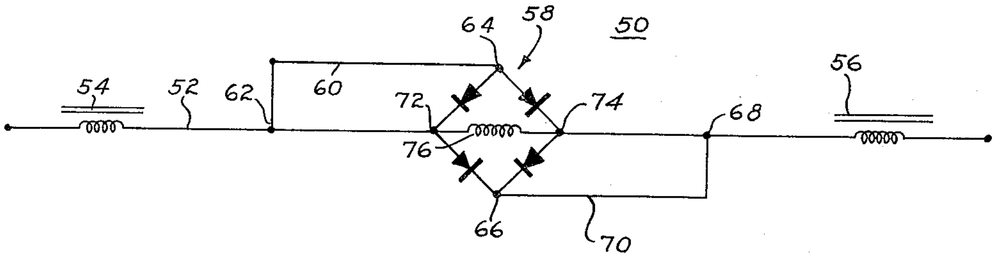
64
58
50
56
54
52
62
60
72
74
68
76
66
70

Fig. 3

INVENTOR
ROBERT M. RICHARDSON

BY Scrivener & Parker
ATTORNEYS

3,098,971

REMOTELY ACTUATED RADIO FREQUENCY POWERED DEVICES

Robert M. Richardson, Sterling, Va.

Filed Sept. 26, 1961, Ser. No. 140,892
2 Claims. (Cl. 325—9)

This invention relates to remotely actuated devices which derive their operating power from transmitted radio frequency energy and, more particularly, to radio receivers and transmitters or transponders operating in this manner.

This invention is particularly well adapted to the remote actuation and powering of hidden or "plant" transmitters and the like and will be described with respect to devices of this general type. Prior art devices accomplish remote actuation and powering of a plant unit by rectifying the radio frequency energy beamed on the plant unit from a central source and using the rectified energy to power suitable local circuit means. In the case of amplitude modulation the rectified energy powers a transistor oscillator and associated speech amplifier and modulator circuits. In the case of frequency modulation a reactance modulator circuit could be included in the plant unit which would require a portion of the rectified energy feeding the local oscillator or a capacitance microphone could be used to achieve the frequency modulation permitting all of the rectified energy to be utilized by the local oscillator.

Plant units such as those described in the foregoing paragraph are seen to be markedly inefficient when a comparison is made between the radio frequency power transmitted to the plant unit from a central source and the power output of the plant unit. In addition, plant units of this type require frequency stabilization of their local oscillators in order to maintain a narrow bandwidth, low noise system. The use of frequency stabilizing means adds to the complexity of the system and results in a further lowering of the operating efficiency.

It is a primary object of the present invention to provide a remotely actuated radio frequency powered device capable of a high conversion efficiency approaching the theoretical maximum. Another object of this invention is to provide a remotely actuated radio frequency powered transmitter comprising means to receive radio frequency energy at one frequency and transmit at a different carrier frequency, including a completely passive frequency multiplier as the local carrier frequency generator therein. Another object of this invention is to provide an improved remotely actuated radio frequency powered transmitter including a single dual band antenna for receiving radio frequency energy, and means to simultaneously multiply the frequency of the received radio frequency energy beamed thereon, for the purpose of retransmission, and rectify the received radio frequency energy to provide a source of power for the said transmitter.

In the drawings forming part of this specification:

FIGURE 1 is a schematic diagram of one embodiment of a frequency multiplier circuit in the present invention;

FIGURE 2 is a schematic diagram of one embodiment of the invention, and

FIGURE 3 is a schematic diagram of another embodiment of the invention.

Basically, the invention comprises a remotely located passive element receiver and/or transmitter adapted to be actuated and powered from a centrally located radio frequency energy source. The basic transmitter circuit comprises a tuned receiving circuit having frequency multiplying means therein to both rectify and multiply the frequency of the received radio frequency (hereinafter referred to as R.F.) energy to provide a source of carrier frequency and excite a transmitting circuit tuned to the same multiple of the frequency from the centrally located source. A local microphone is utilized to supply the modulation for the carrier frequency generated, the power for the modulating means being provided by the rectifying action of the frequency multiplying means. Information may also be received by incorporating a demodulating means such as an earphone type speaker powered by the rectified R.F. energy.

Referring in detail to the drawings and more particularly to FIGURE 1, a frequency multiplier circuit **10** is shown as comprising a full-wave rectifier bridge **12** having a pair of input terminals **14** and a pair of output terminals **16**. The rectifier bridge comprises four solid-state rectifier arms **18** comprising either germanium, gallium arsenide, gallium antimonide, silicon, varactor diodes, tunnel diodes, or any passive non-linear device, the latter being symbolically indicated in the drawing.

Since the rectifier bridge **12** cast on both half cycles of the input signal at the input terminals **14**, the frequency of the rectified signal appearing at the output terminals **16** of the rectifier bridge **12** contains both a rectified or direct current component and an R.F. harmonic component at a frequency equal to twice the input frequency. The relative magnitudes of the D.C. and the R.F. harmonic components depend on the power level at which the multiplier bridge is energized. Thus, both a rectifying action as well as a frequency doubling action is provided by the circuit **10**.

Because of its symmetry, the full wave rectifier bridge provides a 1:1 impedance ratio between input and output making possible an unlimited cascade of similar units to achieve even harmonic multiplication in powers of two.

An additional advantage is gained by the use of varactor diodes and/or back tunnel diodes in the rectifier bridge circuit **12** whereby conversion efficiencies in the range between 80 and 90 percent or greater are made possible.

Referring now to FIGURE 2, a transponder circuit **20** is shown as comprising a dual band antenna **22** tuned to a first or reception frequency and a harmonic multiple of the said first frequency hereinafter referred to as the transmission frequency.

A first passive resonant or input network **24** is provided comprising a first matching section **26** coupled at one end **28** to the antenna **22**, a diode **30** in series with the matching section **26** and a first inductance **32** connected in series with the said diode **30** and first matching section **26** between the diode and a terminal **34** on the antenna **22**.

A second passive resonant or output network **36** is inductively coupled to the first network **24** by means of a second inductance **38** connected to the same terminal **34** on the antenna **22** as the inductance **32** and properly positioned for coupling therewith. The remainder of the output network **36** comprises a tuning capacitor **40** connected in series with the second inductance **38** and a second matching section **42** also in series therewith between the tuning capacitor **40** and a terminal **44** on the antenna **22**.

A pair of selectively tuned trap coils **46** and **48** are provided so that the antenna presents a low impedance to both the reception frequency and the transmission frequency which is a harmonic multiple thereof.

In the embodiment shown, for example, the input network **24** is tuned to the second harmonic of the reception frequency. The received signal is fed through the first matching section **26** to the diode **30** where it is clipped, causing the input network **24** to resonate at the second harmonic of the input signal.

The resonant energy in the input network **24** is coupled to the output network **36** through the two inductors **32** and **38**. The output network **36** is tuned to the transmission frequency which is the same harmonic multiple of the reception frequency as the input network **24** but is coupled to the antenna **22** through the second matching section **42** which is selected to feed the antenna at the transmission frequency.

Thus, by the use of all passive elements and with no local power supply, radio frequency energy may be received by the transponder **20** at a given reception frequency and retransmitted at a preselected transmission frequency which is a preselected harmonic multiple of the said reception frequency.

A further embodiment of the transponder of FIGURE 2 is shown in FIGURE 3. The transponder circuit **50** in FIGURE 3 is shown to comprise a dual band antenna **52**, with appropriate trap coils **54** and **56**, feeding a first signal at a given reception frequency into a bridge type frequency doubler circuit **58**, similar to the multiplier circuit of FIGURE 1 and being fed by the said doubler circuit **58** at a transmission frequency which is the second harmonic of the reception frequency.

A first matching section **60** tuned to the reception frequency feeds the doubler circuit **58** from a terminal **62** on the antenna **52** through one terminal **64** of a diagonally arranged pair of terminals **64** and **66**. The other terminal **66** of the diagonal pair is an output terminal for the doubled reception or transmission frequency which is fed to a terminal **68** on the antenna **52** through a second matching section **70** selectively tuned therefore.

The antenna **52** is also coupled through the doubler bridge circuit **58** by the opposite pair of diagonally disposed terminals **72** and **74** which have an R.F. choke **76** connected therebetween.

The use of the two matching sections **60** and **70** and the manner in which they are connected with the doubler circuit **58** and the antenna **52** provides two resonant networks coupled through the bridge circuit **58** with the operation of the transponder circuit **50** substantially equivalent to that of the transponder **20** shown in FIGURE 2.

The efficiency of the circuit of FIGURE 3, however, is made much greater than that of FIGURE 2 by the use of the bridge type doubler circuit **58**.

As can be seen from the foregoing description this invention provides automatic transponder and transceiver communication systems wherein remotely located transmitters may be contructed of completely passive circuit elements capable of deriving all the necessary operating power from the R.F. energy received from the central station, simultaneously multiplying the received fundamental frequency transmit either a modulated or unmodulated harmonic of the fundamental frequency of the central station back to the said central station.

While I have described and illustrated one form which my invention may take, it will be apparent to those skilled in the art that other embodiments, as well as modifications of that disclosed, may be made and practiced without departing in any way from the spirit or scope of the invention, for the limits of which reference must be made to the appended claims.

What is claimed is:

1. A transponder circuit comprising a dual band antenna for receiving radio frequency energy at a fundamental frequency and re-transmitting said energy at a harmonic of said fundamental frequency, a first passive network resonant at said harmonic frequency selectively coupled with said antenna for receiving energy at said fundamental frequency and multiplying the fundamental frequency to the harmonic frequency, said first passive network comprising a first matching section tuned to feed radio frequency energy at the fundamental frequency from the antenna to the first network, a diode in series with the matching section for clipping said energy, a first inductance in series with said diode and connected through the antenna with the first matching section, the diode, first matching section and first inductance comprising a resonant circuit tuned to resonate at a desired harmonic of the fundamental frequency and being powered totally by the energy received by the antenna and clipped by the diode, and a second passive network resonant at the harmonic frequency coupled with the first network and energized thereby, the second network being selectively coupled to the antenna to feed energy to said antenna at the harmonic frequency whereby transmission of energy at said harmonic frequency is effected.

2. A transponder circuit comprising a dual band antenna for receiving radio frequency energy at a fundamental frequency and re-transmitting said energy at a harmonic of said fundamental frequency, a first passive network resonant at the harmonic frequency selectively coupled with the antenna for receiving energy at the fundamental frequency and multiplying the fundamental frequency to the harmonic frequency, and a second passive network resonant at the harmonic frequency coupled with the first network and energized thereby, the second network being selectively coupled to the antenna to feed energy to the antenna at the harmonic frequency whereby transmission of energy at the harmonic frequency is effected, said second passive network comprising a second inductance having a common connection at the antenna with the first inductance and inductively coupled therewith, a tuning capacitor in series with the second inductance and a second matching section in series with the second inductance and the tuning capacitor and connected through the antenna with the second inductance and tuned to feed radio frequency energy at the desired harmonic of the fundamental frequency to the antenna, the second resonant network being tuned to resonate at the harmonic frequency and energized by the first resonant network by means of the inductive coupling between the first and second inductances.

References Cited in the file of this patent

UNITED STATES PATENTS

2,927,321 Harris ---------------- Mar. 1, 1960

United States Patent [19]

Dudley et al.

[11] **3,933,323**

[45] **Jan. 20, 1976**

[54] **SOLID STATE SOLAR TO MICROWAVE ENERGY CONVERTER SYSTEM AND APPARATUS**

[75] Inventors: **Kenneth W. Dudley**, Sudbury; **George H. MacMaster**, Lexington, both of Mass.

[73] Assignee: **Raytheon Company**, Lexington, Mass.

[22] Filed: **Mar. 20, 1974**

[21] Appl. No.: **453,133**

[52] **U.S. Cl.** .... **244/1 R**; 250/199; 310/4 R; 322/2 R; 331/94.5 R

[51] **Int. Cl.** **B64c 39/02**

[58] **Field of Search** ...... 322/2 R; 310/4 R; 321/8 R; 331/DIG. 1, 94.5 R; 250/199; 136/206; 60/26, 203; 244/1 R, 158; 307/312, 311, 278

[56] **References Cited**

UNITED STATES PATENTS

3,225,208	12/1965	Wolfe	310/4 X
3,341,708	9/1967	Bilderback	331/DIG. 1
3,434,678	3/1969	Brown et al.	244/158
3,467,840	9/1969	Weiner	310/4
3,535,543	10/1970	Dailey	310/4 X
3,781,647	12/1973	Glaser	310/4 X

Primary Examiner—Trygve M. Blix
Assistant Examiner—Paul E. Sauberer
Attorney, Agent, or Firm—Edgar O. Rost; Joseph D. Pannone; Harold A. Murphy

[57] **ABSTRACT**

Solid state spaced-oriented means convert solar radiation into microwave energy which is amplified and beamed to earth as a source of electrical power. A large number of semiconductor devices operating in the current multiplication region are irradiated by single mode coherent light beam generation means which is pumped by solar energy. A neodymium yttrium aluminum iron garnet laser or CW gallium arsenide junction diode laser mode-locked at a predetermined microwave frequency, such as 3,000 MHz provide for generation of a light beam. The generated current is directly proportional to the radiant energy and is coupled from the semiconductor device array to antenna means for transmission to earth stations for conversion into low frequency electrical energy.

6 Claims, 3 Drawing Figures

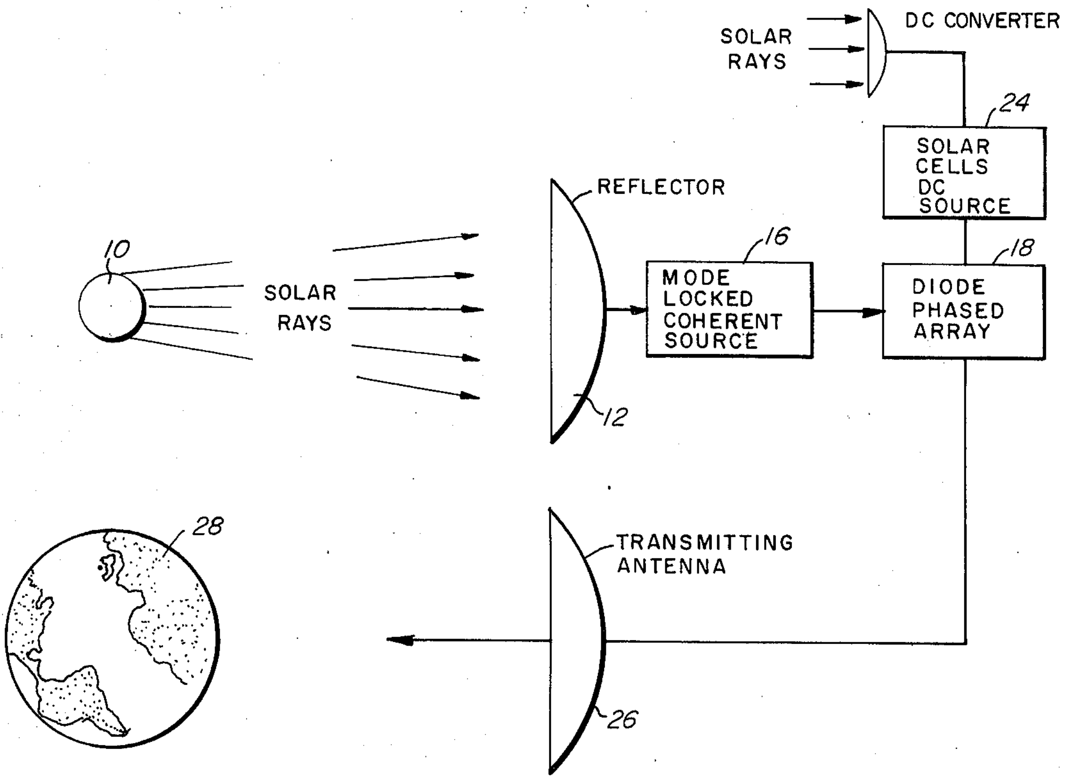

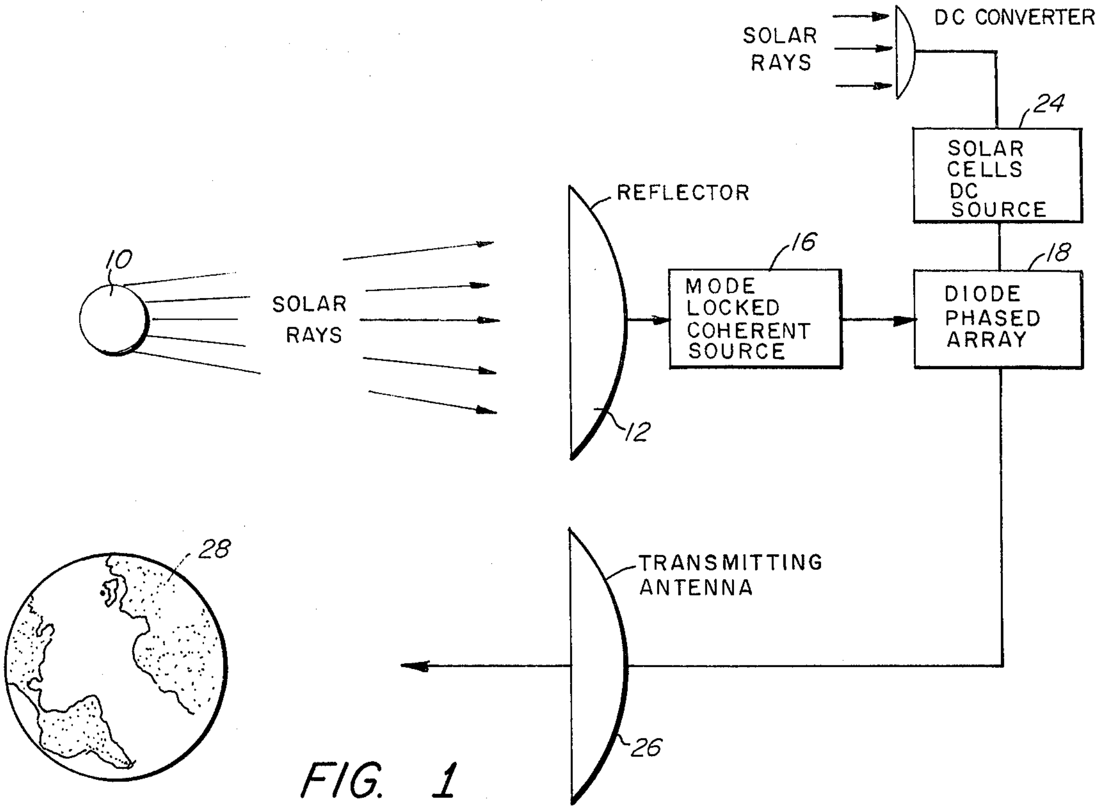
SOLAR
RAYS
DC CONVERTER
24
SOLAR
CELLS
DC
SOURCE
REFLECTOR
16
18
10
SOLAR
RAYS
MODE
LOCKED
COHERENT
SOURCE
DIODE
PHASED
ARRAY
12
28
TRANSMITTING
ANTENNA
26

FIG. 1

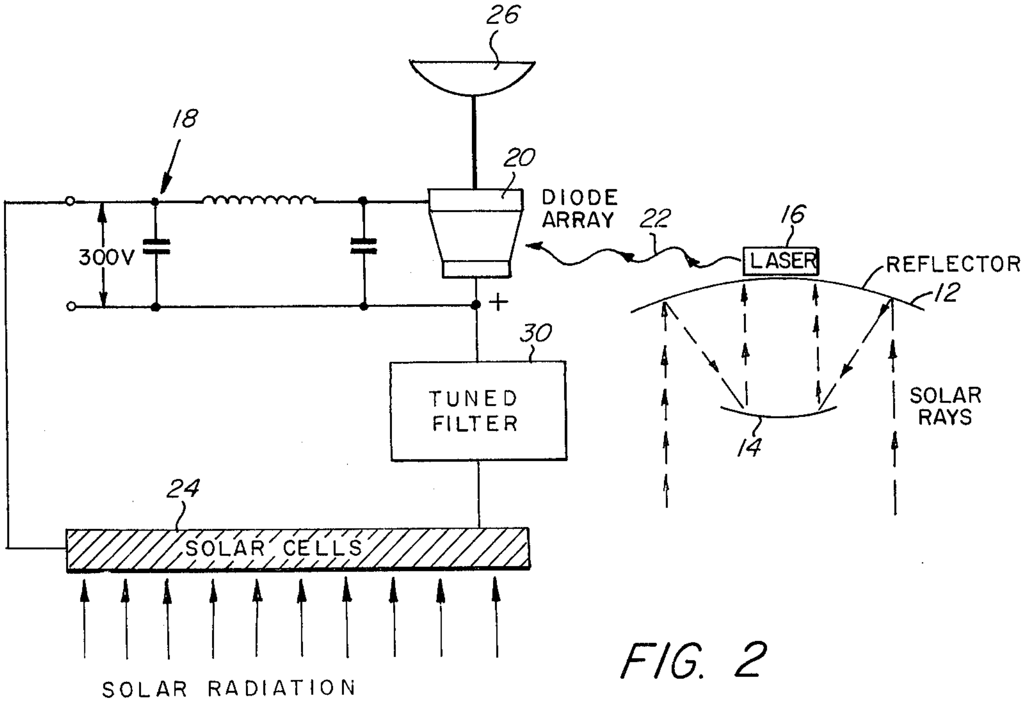
26
18
20
DIODE
ARRAY
22
16
300V
LASER
REFLECTOR
12
+
30
TUNED
FILTER
14
SOLAR
RAYS
24
SOLAR CELLS
SOLAR RADIATION

FIG. 2

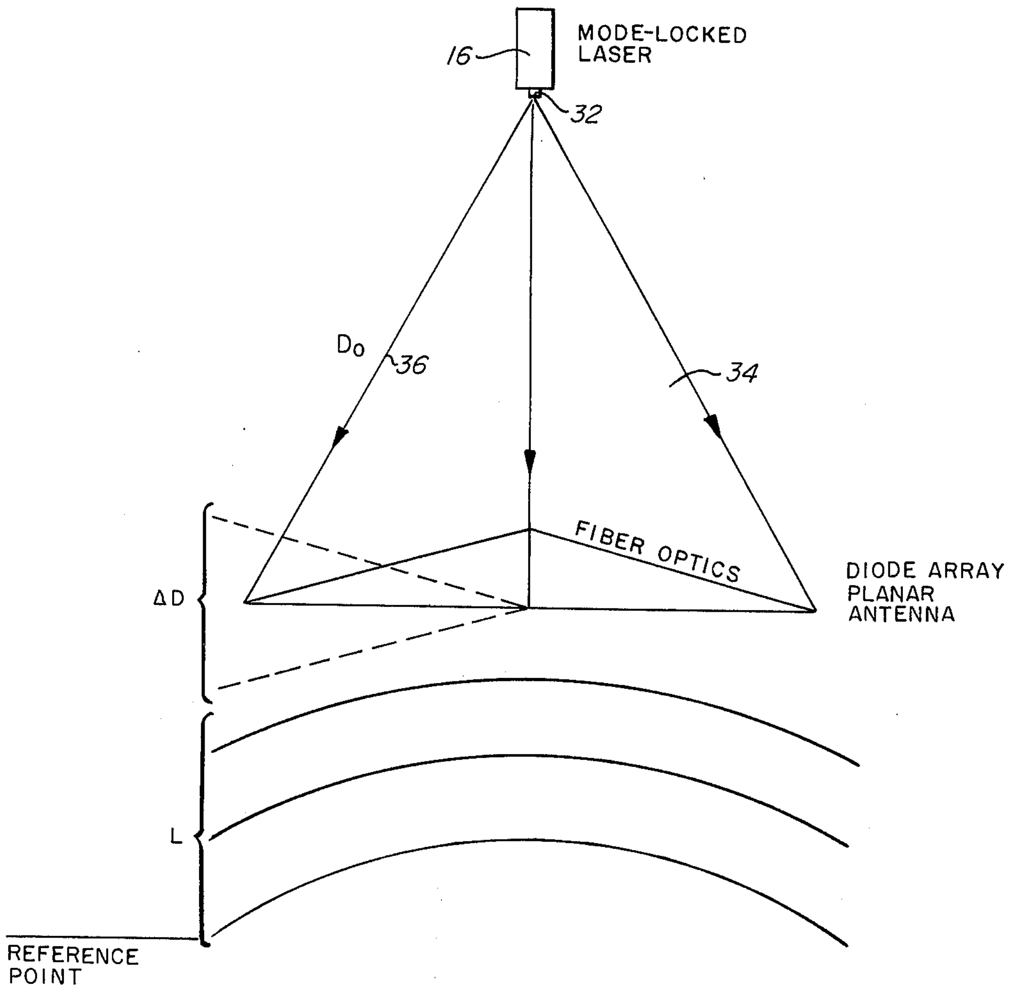
MODE-LOCKED LASER
16
32
Do
36
34
FIBER OPTICS
DIODE ARRAY PLANAR ANTENNA
ΔD
L
REFERENCE POINT

FIG. 3

SOLID STATE SOLAR TO MICROWAVE ENERGY CONVERTER SYSTEM AND APPARATUS

BACKGROUND OF THE INVENTION

1. Field of the Invention

The invention relates to a system and apparatus for conversion of solar radiation into microwave energy in space and conversion into low frequency electrical power utilizing solid state devices.

2. Description of the Prior Art

Conversion of microwave energy to low frequency electrical power has been disclosed in prior art embodiments utilizing antenna-rectifier arrays comprising a large number of devices, such as solid state devices, to propel space vehicles. U.S. Pat. No. 3,434,678 issued Mar. 25, 1969 to W. C. Brown et al and assigned to the assignee of the present invention provides an embodiment of a helicopter-type space vehicle which is self-supporting by means of microwave energy beamed into space and converted to electrical power for operation of an electric motor. Microwave energy is defined an electromagnetic energy in that portion of the spectrum having wavelengths in the order of 1 meter to 1 millimeter and frequencies in excess of 300 MHz. Such energy is preferred for wireless transmission since it is desirable to focus the transmitted energy with a relatively high power density upon a small area spaced from the source. In accordance with the laws of optics, the sharpness of a microwave beam varies as the ratio of antenna dimensions to the wavelength of the transmitted energy. Accordingly, for a given predetermined power density and beam sharpness, the very short wavelengths of microwave frequency oscillations facilitates a corresponding decrease in the dimensions of the antenna means for transmitting and receiving. In addition, in long wave or radio-type transmission, difficulties are encountered as a result of natural or man-made interference and noise and this problem is considerably less significant with the transmission of microwave frequencies. The recent United States space programs have shown that the employment of very short microwaves for beamed energy is preferable for penetration of certain atmospheric reflection layers, as well as, the advantages of accurate focusing and high power density of the transmitted energy.

Other examples of the utilization of microwave energy for space applications are found in U.S. Pat. No. 3,174,705, issued Mar. 23, 1965 to D. Shiff et al, as well as U.S. Pat. No. 3,083,528 issued Apr. 2, 1963 and U.S. Pat. No. 3,114,517, issued Dec. 17, 1963 to W. C. Brown. All of these are also assigned to the assignee of the present invention and disclose various conversion means utilizing heat exchangers for conversion of microwaves into electrical energy or heat which can be utilized for directly or indirectly propelling a space vehicle.

The current energy crisis throughout the world due to the growing shortage of fossil fuels, namely coal, oil and natural gas due to increases in the level of consumption and the decline in available resources has led to the search for new sources of energy. The use of solar energy has been proven in the space program to provide for electric power aboard the spacecraft. In addition, numerous attempts have been made on the earth to collect and convert solar energy into useful electrical power. An embodiment of a means for conversion of solar radiation in space into microwave energy utilizing a satellite system, including means for transmission to earth stations and conversion to electrical power is disclosed in U.S. Pat. No. 3,781,647, issued Dec. 25, 1973 to P. E. Glaser. The satellite system includes several geostationary satellites positioned in an energy receiving relationship to the sun. The solar energy conversion means include guidance and control means to correctly orient the collector means in relationship to the sun. The solar energy conversion means convert the radiation received into DC electrical energy which is subsequently utilized to operate high power generators of microwave energy for transmission by means of a focused beam to collection means on earth.

Solar energy offers an almost unlimited source for man's energy needs with the economics playing the most significant role in its utilization. In recent years solid state lasers and semiconductor devices for conversion of radiant energy into electrical energy have achieved very high efficiencies. Such semiconductors include PN silicon junction diodes which can be electrically biased to operate in the current multiplication region and the generated current then becomes directly proportional to the quanta of impinging light energy. The present invention describes a unique system and apparatus for generation of microwave energy from solar radiation in space applications utilizing controlled coherent light beams and semiconductor electrical conversion means.

SUMMARY OF THE INVENTION

In accordance with the present invention a system with apparatus utilizing solid state devices is provided for conversion of solar radiation into microwave energy and transmission to an earth station for conversion into electrical power. A large number of semiconductor devices are utilized in a phased array as electrical energy converters of a controlled frequency coherent laser beam irradiating the semiconductor array. Since microwave energy is desired for wireless transmission to earth a frequency of approximately 3,000 MHz is preferred. Pulsed coherent light energy from a solar energy pumped laser at this frequency when irradiating a large phased array of reverse-biased semiconductor diodes results in microwave power which is directly proportional to the quantra of light striking the semiconductor devices. The output current and power is modulated by the modulation on the laser beam. Solid state devices today operate at efficiencies as high as 80 to 90 percent. Mode-locking of the laser beam is achieved through the use of such devices as the neodymium-YAIG or continuous wave gallium arsenide diode junction lasers.

The output of the phased array of semiconductor devices is transmitted to an antenna for transmission to the earth station where the energy is converted into low frequency electrical power. The controlled phase irradiation of the large semiconductor device phased array planar antenna is accomplished by a means of a lens system and fiber optics to compensate for bending or flexing of the planar antenna. Phase coherency of the relatively stable outputs generated by the system in energy beamed to earth is a feature of the embodiment of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will now be described in the following detailed specification with reference directed to the accompanying drawings, wherein:

FIG. 1 is a diagrammatic view of the embodiment of the system of the invention;

FIG. 2 is a partial schematic view of the solar energy collection and microwave energy conversion portion of the system of the invention; and

FIG. 3 is a diagrammatic representation of the feature of the system relating to phase compensation of the coherent radiant energy beam generating apparatus.

DESCRIPTION OF THE PREFERRED EMBODIMENT

The system and space power station apparatus embodying the invention is illustrated in FIG. 1. A solar energy source **10** provides solar rays impinging on the reflector **12** of, illustratively, a Cassigranian-type antenna which is shown in detail in FIG. 2. Auxiliary concentrator **14** at approximately the focal point provides the primary source of solar energy for pumping a coherent radiant energy beam generation means **16**, preferably a mode-locked laser. Such beam generation means include a neodymium-yittrium aluminum iron garnet rod laser providing a mode-locked pulse output at, illustratively, a specific microwave frequency of 3,000 MHz. Approximately 25 watts of solar energy are required for the pumping of the coherent beam generation means **16** and the pulsed output beam **22** irradiates a phased array of a large number of semiconductor devices **18.** A variety of methods using intercavity modulation techniques are available to permit continuously pumped optical beam generators to be repetitively pulsed. Such techniques are generally referred to as mode-locking. The output of the coherent beam generation means **16** is a series of very narrow mode-locked pulses which, illustratively have a repetition frequency of $f = (c/2l)$, where l is the length of the modulation means. The pulse which is in the order of L/C where L is the length of the laser rod which corresponds to a mode-locked pulse width of 0.03 nanosecond for a 1 centimeter long rod. With a 2 percent efficiency for the neodymium-YAIG laser, a 0.5 watts input produces a 10 milliwatt output.

Other coherent beam means **16** include the recently developed junction lasers in which the laser action is produced by a high, nonequilibrium concentration of electrons and holes within a small region of a semiconductor crystal material. Mirrors form an integral part of the laser structure and may be formed by cleaving the crystal. Pulsed juncture lasers have now operated at as high as 40 percent external quantum efficiency. Continuous outputs of approximately 40 milliwatts have been obtained at room temperatures with even higher powers at low temperatures. Further details regarding this alternative apparatus involving solid state materials, such as Group III–V compounds and IV–VI compounds will be found in an article entitled "Junction Lasers" by L. A. D'Asaro and J. E. Ripper, Physics Today, March 1971, pps. 42–48 inclusive.

The output coherent optical beam **22** irradiates a large number of solid state semiconductor devices, such as reversed-bias PN silicon junction diodes **20**, disposed in a phased array **18** on the space power station. The details of the circuitry associated with an individual semiconductor device **20** is illustrated in FIG. 2. Preferably, the phased diode array operates as a Class "C" power amplifier with each of the semiconductor diodes being reversed-biased in a current multiplication region so that the current output is directly proportional to the quanta of the radiant energy in the optical beam impinging on the diodes. The output current and power of the phased diode array will follow the modulation on the light beam **22.** Illustratively, with a 10 milliwatt pulsed input from the mode-locked means **16**, the current flowing from each diode with anticipated gain of 40dB and efficiency of approximately 80 percent, would be 100 watts continuous wave energy at 3,000 MHz.

The phased array **18** is provided with DC reverse-biasing means, such as an array of solar cell converters **24** which may be of the photovoltaic, thermionic or thermoelectric type. Examples of the photovoltaic type include, but are not limited to, cadmium sulphide cells, NP silicon cells, gallium arsenide solar cells, as well as organic film solar cells. Examples of the thermionic type include high vacuum and plasma diodes. Examples of the third, or thermoelectric type, include bimetallic junctions displaying the Seebeck effect. Further details on such DC electrical energy conversion means is available in the hereinbefore referenced U.S. Pat. No. 3,781,647. The combined output of the semiconductor diode array of, illustratively, 100 watts CW per diode is transmitted by means of transmitting antenna **26** to a receiving station on earth **28** for reception of the microwave energy and conversion to low frequency electrical power. A tuned filter **30** between the DC solar energy converter means **24** and diodes **20** will assure the direction of substantially all of the microwave energy to antenna **26.**

All of the aforementioned apparatus is preferably carried on a space power station, which may be manned or unmanned, and conventional guidance means, such as sun sensors, star trackers, and horizon seekers or other known components for control of such space-oriented vehicles, would be included in the station. In addition, any protective devices, such as means for protecting the sensitive solar cells, and appropriate cooling systems for the operation of the station at the very high temperatures involved with the collection of the solar energy would also be required aboard the space power station.

Since a large number of solid state semiconductor microwave energy generators are utilized, the phase synchronization between each of the individual components is extremely critical. Referring now to FIG. 3, a system for maintaining phase compensation with respect to a phased diode array **18** planar antenna having a relatively large number of semiconductor microwave power generators, irradiated by mode-locked means **16**, includes a lens system **32** and optical fiber means **34.** Typically, such an antenna array is required to be rigid so that the phase of the combined output power of the semiconductor devices is in the proper phase for aiming to earth via the transmitting antenna **26.** To compensate for the bending or flexing of the planar antenna, it is possible to provide a method of compensation by using the principle that the velocity of the light waves and microwaves are essentially the same. Any deviation from a reference point is designated L for a beam indicated by the line **36** emitted from the mode-locked means **16.** For proper phase coherence at reference point L, the distance D from the antenna surface to the laser means **16**, together with ΔD which represents a change in the antenna surface, is considered. $D + L +$ (fiber optic delay $< 2\pi$) = proper phase coherence at point L. For light rays

$$\frac{(D \pm \Delta D)}{C_L}$$

For microwave energy the equation is

$$\frac{(L \mp \Delta D)}{C_{\mu\omega}}$$

Since the light and microwave wave velocities are substantially equal the sum of these two equations will equal

$$\frac{D+L}{C} = \text{proper phase coherency.}$$

Numerous other phase compensation and phase shifting means may be incorporated in the structure and programmed computerized control may also be utilized in conjunction with the laser lens and fiber optics system.

There is thus disclosed a unique system and apparatus for the conversion of solar radiation into microwave energy for transmission to earth and conversion into electrical power by utilizing solid state solar energy pumped coherent optical beam sources for irradiating solid state semiconductor generator means in a phased array system. The phase coherency feature of the beam generation means and substantially complete isolation of the semiconductor drive means will assure a high degree of stability of the generated microwave energy signals. With continued improvements resulting from space programs, numerous modifications, alterations and variations will be evident to those skilled in the art. The foregoing description of an illustrative embodiment, therefore, is to be considered in its broadest aspects and not in a limiting sense.

We claim:

1. A system for converting solar energy into microwave energy comprising in combination;

a source of solar energy radiation;

means for collecting said solar energy operatively associated with a pulsed coherent radiant energy mode-locked laser beam source operative at a fixed predetermined microwave energy frequency;

solid state semiconductor amplifier means adapted to be irradiated by said radiant energy laser beam to generate an amplified microwave energy output at substantially said predetermined frequency; and

means for transmitting said amplified microwave energy to a utilization load.

2. The system according to claim **1** wherein said coherent beam source comprises a neodymium yttrium aluminum iron garnet material laser.

3. The system according to claim **1** wherein said coherent beam source comprises a semiconductor material PN junction laser.

4. The system according to claim **1** wherein said solid state semiconductor means comprise a phased array of a plurality of interconnected electrically reversed-bias PN junction silicon diode operating in the current multiplication region.

5. The system according to claim **4** wherein said semiconductor diodes are electrically biased by a voltage source comprising an array of solar energy to DC electrical energy conversion means.

6. The system according to claim **1** wherein said solar energy collection means, coherent beam source, solid state semiconductor amplifier means and said microwave energy transmitting means are oriented in outer space and said utilization load comprises energy receiving and electrical power conversion means on earth.

* * * * *

United States Patent [19]

Martin et al.

[11] **Patent Number: 4,955,562**

[45] **Date of Patent: Sep. 11, 1990**

[54] **MICROWAVE POWERED AIRCRAFT**

[75] Inventors: **John E. Martin**, Ottawa; **James D. DeLaurier**, Concord; **George W. Jull**; **Arne Lillemark**, both of Nepean, all of Canada

[73] Assignee: **Her Majesty the Queen in right of Canada, as represented by the Minister of Communications**, Ottawa, Canada

[21] Appl. No.: **272,015**

[22] Filed: **Nov. 16, 1988**

[30] **Foreign Application Priority Data**

Nov. 24, 1987 [CA] Canada 552664

[51] **Int. Cl.**^{5} **B64C 39/02**

[52] **U.S. Cl.** .... **244/62**; 244/1 R; 343/705; 343/708

[58] **Field of Search** .... 244/62, 1 R, 45 R, 45 A, 244/87, 91; 343/705, 708

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,358,946	12/1967	Shye	244/45 R
3,434,678	3/1969	Brown et al.	244/158 R
3,656,164	4/1972	Rempt	343/705
4,593,288	6/1986	Fitzpatrick	244/118.1
4,635,067	1/1987	Fitzpatrick	343/705
4,662,588	5/1987	Henderson	244/118.2

Primary Examiner—Joseph F. Peters, Jr.
Assistant Examiner—Anne Sartelle
Attorney, Agent, or Firm—Nixon & Vanderhye

[57] **ABSTRACT**

A microwave-powered aircraft has lifting surfaces for exerting lifting forces on the aircraft in response to the propulsion of the aircraft and a rectenna array for receiving and rectifying microwave energy transmitted to the aircraft from a location remote from the aircraft. An electric motor for driving a propeller is energized by microwave energy received by the rectenna array, which is provided in a body at the underside of the aircraft. The body has its major dimensions extending horizontally and is relatively shallow with a periphery which is vertically curved so as to reduce turbulence in the airstream over the body during flight. The body is separate from the lifting surfaces and shaped to at least substantially avoid the generation of lifting forces by the body.

16 Claims, 2 Drawing Sheets

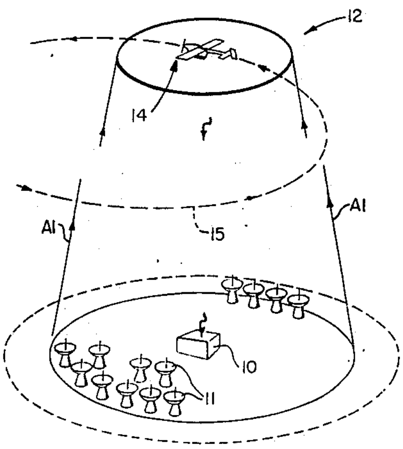

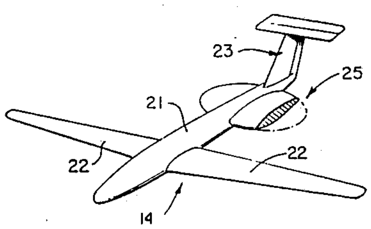

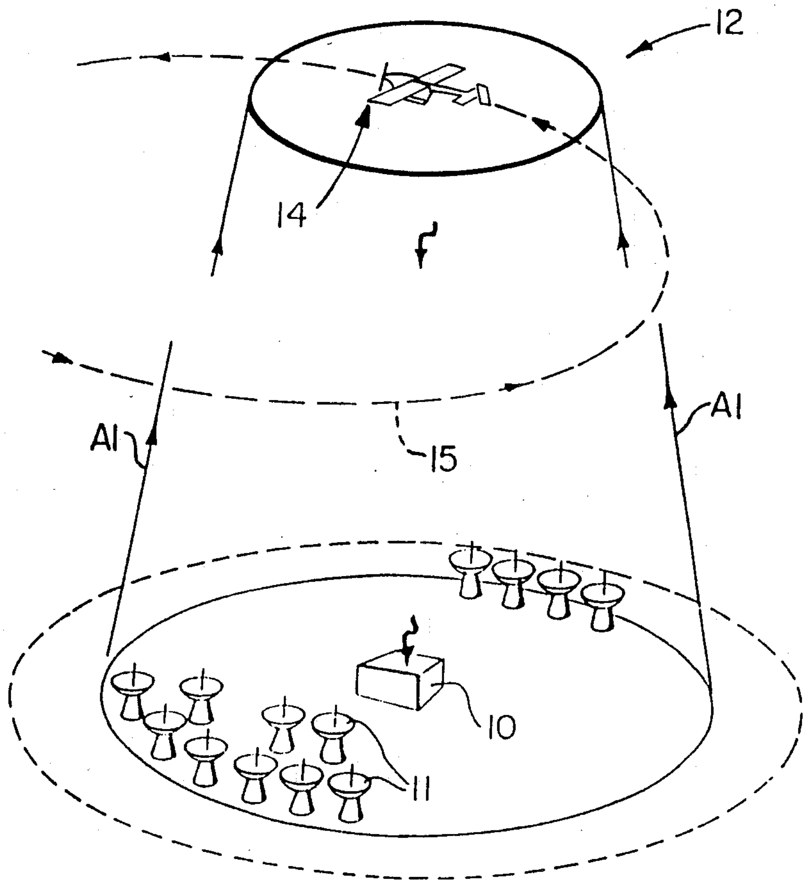
12
14
A1
A1
15
10
11

FIG. 1

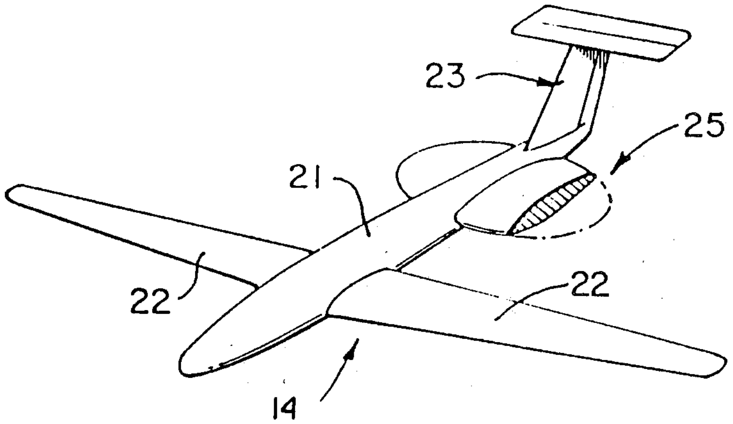
23
25
21
22
22
14

FIG. 2

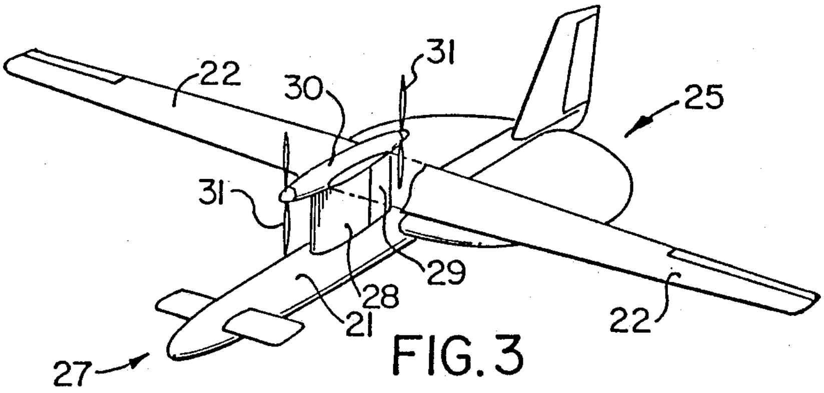
22
30
31
25
31
29
28
21
22
27

FIG. 3

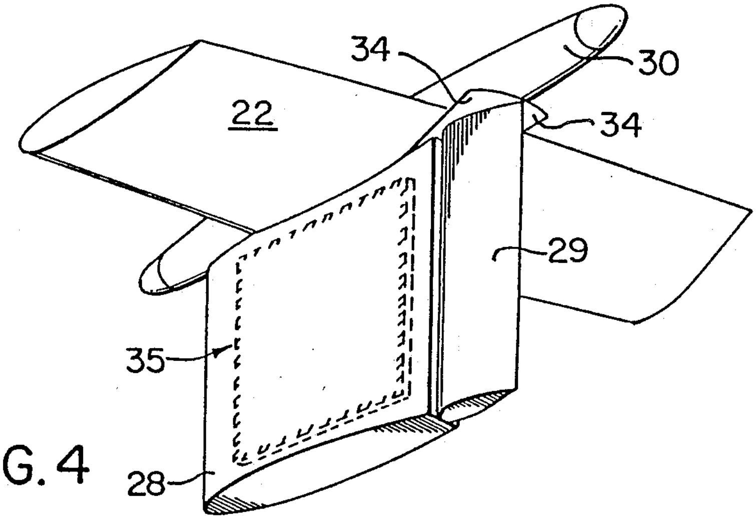
34
30
22
34
29
35
28

FIG. 4

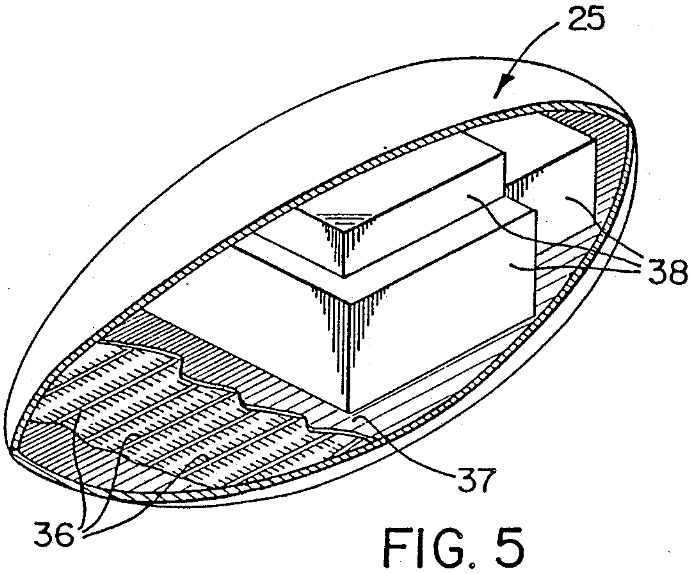

FIG. 5

MICROWAVE POWERED AIRCRAFT

FIELD OF THE INVENTION

The present invention relates to a microwave-powered aircraft and, more particularly, to a high altitude electrically-powered aircraft to which power is supplied in the form of microwaves, which are transmitted from a ground station, and which are rectified on the aircraft to provide a source of electrical power for the propulsion and on-board systems of the aircraft.

DESCRIPTION OF RELATED ART

It is already known, from the prior art, how to provide a microwave power collecting and conversion device, which is known as a "rectenna" (rectifying antenna) mounted on the lower surfaces of such an aircraft for receiving and rectifying microwave energy, the rectenna being provided in the form of an array of discrete components comprising half wavelength dipole antennas, each terminated in a rectifying diode, or in the form of one or more thin film printed circuits, which intercept the fraction of the power of a microwave beam which falls upon the surface of the rectenna. This microwave power is converted by the rectenna into direct current power, which is used to drive one or more electrical propulsion motors, for rotating one or more propellers, and to power a payload carried by the aircraft.

It has been suggested that such aircraft could be platforms for payloads employed, for example, for telecommunications signal relaying, TV broadcasting, remote sensing and conducting of surveillance and atmospheric environmental probing missions. For such missions, it is often desirable to confine the aircraft flight path to a constrained station-keeping volume above the ground station.

More particularly, in U.S. Pat. No. 3,144,517 there is disclosed a microwave-powered space vehicle which utilizes a heat exchanger method for electromagnetic energy conversion.

U.S. Pat. No. 3,434,678 discloses a combined antenna and conversion mechanism, i.e. a rectenna, for powering space vehicles, and discloses the use of such a mechanism for propulsion by incorporation of the rectenna in a small flying model helicopter.

In "High Altitude Powered Platform Cost and Feasibility Study", J. W. Sinko, SRI Project 5655-502, prepared for NASA under Contract NASW-2962, Oct. 1977, there is proposed a neutrally-buoyant dirigible lighter-than-air vehicle and a microwave-powered aircraft, both of which obtain propulsive power from a microwave beam through use of rectennae.

In "Design Definition of a Microwave Power Reception and Conversion System for use on a High Altitude Powered Platform" NASA/CR/156866, W. C. Brown, Raytheon, prepared for NASA under Contract NAS-6-3006, 1981, and "Design Study for a Ground Microwave Power Transmission System for use with a High Altitude Powered Platform", W. C. Brown, Raytheon, prepared for NASA under contract NAS-6-3200, May 1982, there are described microwave-power transmission and reception systems suitable for application in neutrally-buoyant air ships, employed as high altitude platforms, and "Design Study for Remotely-Piloted, High Altitude Airplanes Powered by Microwave Energy", C. E. K. Morris Jr., AIAA Applied Aeodynamics Conference, July 13–15, 1983, Danvers, Mass., and "Parametric Study of Microwave-Powered High-Altitude Airplane Platforms Designed for Linear Flight", C. E. K. Morris Jr., NASA Technical Paper 1918, Nov. 1981, disclose analytical research into the subject of microwave powered aircraft where the power is collected at the lower surfaces of the wings of the aircraft. "Research on the Technology of an Airplane Concept for a Stationary High-Altitude Relay Platform (SHARP)", J. Delaurier, B. Gagnon, J. Wong, R. Williams and C. Hayball, presented at the Thirty-Second Annual General Meeting of the Canadian Aeronautics and Space Institute, Montreal, May 27, 1985, describes microwave-powered aircraft research in which power is collected at the lower surfaces of monoplane and biplane airplane wings.

Economic analyses have shown that the ten year life cycle costs of a ground transmission system can exceed those associated with its high altitude airplane platform, constructed according to prior art, by factors of four or more. Therefore, increases in the cost and complexity of the aerial vehicle are fully justified provided that they result in a reduction in costs of the ground transmission system. Cost reductions would be achieved by reduction of transmitted power level and/or by a reduction of the beam steering range. This latter reduction would be achieved through reduction in the horizontal extent of the station keeping volume.

It is known that life cycle costs associated with a ground transmission system are proportional to the square root of the transmitted power level. They are also inversely proportional to the required beam cross-sectional diameter at the operating flight altitude, because, by well-known laws of microwave optics, a smaller focussed beam cross-section at altitude requires a larger aperture antenna on the ground, the costs of which are proportional to the aperture area.

The fraction of total transmitted power which can be intercepted on the lower surfaces of wings of microwave-powered aircraft can be shown to be very low indeed, for practical configurations of aircraft and power beam cross-sectional areas. For example, a focussed power beam cross-sectional shape is generally approximately circular, with a diameter of some tens of meters at an altitude of 20 km. On the other hand, the wing span of practical sizes of microwave-powered aircraft can be up to several tens of meters, with wing chords of only a few meters. As a consequence, the efficiency of transferring microwave power from ground to aircraft can be somewhat less than ten per cent in such practical cases. Hence, the transmitted power levels, and resulting power flux density levels required for flight at the operating altitude, must be increased to compensate for such low power transfer efficiency.

In addition, optimum tradeoffs between power beam steering angles and aircraft banking angles are important in increasing efficiency in microwave power transfer from ground to aircraft. These tradeoffs also impact effectiveness to carry out many types of missions. Analysis shows that aircraft configured according to the prior art either require a wide horizontal extent to the station-keeping volume and hence large beam steering angles or, alternately, must bank over large angles to execute required flight patterns within a station keeping volume of narrow horizontal extent. It is known that system costs increase for flight over a wide horizontal range because of the consequently additional wide angular steering demands placed on the power beam steering systems. For example, it has been calculated that life cycle costs of one type of ground transmission system would increase by a factor of two if the angular beam steering range were increased from ±3° to ±6°, corresponding to a horizontal diameter of the flight pattern increasing from two km to four km at an altitude of 20 km.

SUMMARY OF THE INVENTION

It is accordingly an object of the present invention to reduce the cost and to increase the mission payload capability of a microwave-powered aircraft system as compared with prior art systems.

The present invention provides a microwave-powered aircraft comprising means for propelling the aircraft, means defining lifting surfaces on the aircraft for exerting lifting forces on the aircraft in response to the propulsion of the aircraft, means for receiving and rectifying microwave energy transmitted to the aircraft from a location remote from the aircraft, the means for propelling the aircraft being operable in response to the microwave energy received by the microwave energy receiving and rectifying means, and the microwave energy receiving and rectifying means comprising a body at the underside of the aircraft, the body having its major dimensions extending horizontally and being relatively shallow with a periphery which is vertically curved so as to reduce turbulence in the airstream over the body during flight, and the body being separate from the lifting means and shaped to at least substantially avoid the generation of lifting forces by the body, and a plurality of downwardly-facing microwave antennas and associated rectifying elements at the underside of the body.

More particularly, the body may be of lenticular shape and substantially elliptical cross-sectional shape, in planes parallel to and transverse to the longitudinal axis of the aircraft, and of substantially circular plan shape.

The means defining the lifting surfaces preferably comprise wings located forwardly of the lenticular body, which may in that case serve as means for promoting longitudinal stability of the aircraft.

A microwave energy reflector may be provided within the lenticular body, above the microwave antennas, for shielding a space located above the microwave reflector within the lenticular body from microwave radiation.

The aircraft preferably includes a fuselage with a movable canard located forwardly of the lifting surfaces, or conventional tail surfaces of the fuselage, for promoting longitudinal stability of the aircraft, and a pylon of symmetrical aerofoil cross-section is preferably mounted on the fuselage to support the wings, the pylon including a flap at a trailing edge of the pylon for causing a side force to act on the aircraft independently of the angle of bank of the aircraft. The control of such a surface in conjunction with the other control surfaces of an aircraft, can produce turns with relative small banking angles.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be more readily understood from the following description of preferred embodiments thereof given, by way of example, with reference to the accompanying diagrammatic drawings, in which:

FIG. 1 shows a view in perspective of a microwave-powered aircraft system, including a ground station and an aircraft powered from the ground station;

FIGS. 2 and 3 show views in perspective of an aircraft embodying the present invention;

FIG. 4 shows a view in perspective of parts of the aircraft of FIG. 3; and

FIG. 5 shows a broken-away view in perspective of a lenticular body forming part of the aircraft of FIGS. 2 to 4.

DESCRIPTION OF PREFERRED EMBODIMENTS

Referring now to FIG. 1 of the accompanying drawings, there is shown in this figure a microwave-powered aircraft system of the type in which the present invention may be embodied. This system has, on the ground, a control center **10** associated with an array of microwave transmission antennas **11**, for transmitting upwardly a power beam, indicated by arrows **A1**, into a power spot, indicated generally by reference numeral **12**, at the height of an aircraft, indicated generally by reference numeral **14**, which is to be powered by the microwaves.

The aircraft **14** is shown flying around a circular flight pattern **15** and this circular flight pattern, or other appropriate closed flight pattern may, for example, be at a height of 20 km above ground level, the power spot **12** having a diameter of approximately 30 meters, a circular flight pattern **15** having e.g. a diameter of up to approximately 4 km and the microwave energy being transmitted from an array of antennas within a circular area, of approximately 70 meters diameter.

Thus, the antennas **11** direct and focus a high-powered microwave beam onto the aircraft **14**, which is powered by the microwave energy received by a rectenna (not shown) provided on the aircraft **14**. In this way, the aircraft is propelled around its flight path or circular course **15**, which diagrammatically represents the perimeter of the area of movement of the aircraft **14**, and which, as indicated above, has a diameter of 4 km, which corresponds to a beam steering range of approximately ±6° at the altitude of 20 km.

Referring now to FIG. 2 of the drawings, there are shown parts of an aircraft indicated generally by reference numeral **20**, which has a fuselage **21**, a pair of wings **22** and a conventional aircraft tail assembly indicated generally by reference numeral **23**, the means for propelling this aircraft being omitted to facilitate illustration.

To the rear of the aircraft wings **22**, the fuselage **21** is provided with a lenticular body indicated generally by reference numeral **25**, which has at its underside a downwardly facing rectenna (not shown).

In order to minimize the required microwave beam power flux density by providing a large area for accommodating the rectenna, but without necessarily contributing to the lift forces exerted on the aircraft **14** by the airflow past the aircraft **14**, the lenticular body **25** is shaped and oriented so as to avoid the generation of aerodynamic lift and, thus, to counteract the generation of lift-induced drag by the lenticular body **25**, which would otherwise increase the power required for propelling the aircraft.

This is in contrast with prior art aircraft, in which the provision of an adequate rectenna area on the lower wing surface conflicts with the necessity for high wing aspect ratio for reducing the required flight power.

Thus, to enable the rectenna area to be made large, the lenticular body **25** is designed to avoid the creation of lift and drag forces by the lenticular body and, for that purpose, is made elliptical in cross-section parallel to the longitudinal axis of the aircraft, i.e. in the streamwise direction, and has no sharp trailing edge, so as to avoid the generation of lifting forces.

As will be appreciated by those skilled in the art, in a real, viscous fluid such as air, there will in practice be some lifting forces generated by the lenticular body **25**. However, the effect of such forces is minimized by the choice of a low aspect ratio for the lenticular body, and by the adoption of a circular planform for the lenticular body **25**, which provides a lift-inefficient surface.

By thus separating the lifting surface area, i.e. the wings, from the lenticular body **25** of the rectenna, the wings **22** may be designed with an aerodynamically efficient high aspect ratio, thus producing a beneficial reduction in the power required for flight of the aircraft without necessitating an increase in the required power flux density in the microwave beam utilized to transmit that power to the aircraft.

The lenticular body **25** is formed by a skin of stretchable plastic material stretched over a framework (not shown) and coated with PTFE to protect the skin against the atmospheric environment, and particularly against atmospheric ozone.

In the embodiment of the lenticular body illustrated in FIG. **5**, the volume within the lenticular body **25** above the rectenna **36** is effectively shielded from the microwave power beam by a reflector surface **37**, associated with the rectenna itself. This volume is therefore available for the installation of power conditioning, controls and payload components **38**, many of which should desirably or even necessarily be isolated from the microwave power beam.

While the lift coefficient, and therefore the induced drag, of the lenticular body **25** is small, this body nevertheless presents a sufficiently large area, which is determined by the requirements for the rectenna, that the body can produce a sufficient force allowing it to act as a stabilizing horizontal tail, the lenticular body **25** being located, in the embodiment illustrated in FIG. **2**, aft of the main lifting surface provided by the wings **22**.

In this connection it will be appreciated that since the microwave-powered aircraft is intended for operation at high altitudes, it does not posses a high degree of natural longitudinal stability and, therefore, artificial augmentation of such stability is desirable. To this end, the control of the pitch of the aircraft may be effected by a separate control surface, either in the form of a conventional tail plane and elevator, as indicated by reference numeral **23** in FIG. **2**, by an all-moving tail plane or, alternatively, and in order to minimize aerodynamic resistance, by a forward control surface or canard, as indicated generally by reference numeral **27** in FIG. **3**. The size of the canard is selected so as to provide appropriate pitching moments to trim the aircraft throughout the operational speed range of the aircraft in response to steering control signals and to provide pitch damping moments in response to the signals from a pitch stability augmentation system (not shown).

In order to produce a lateral force acting upon the aircraft independently of the banking angle of the aircraft, so that the required banking angle can be minimized, a vertical aerofoil surface is provided in the neighbourhood of the location of a longitudinal center of gravity of the aircraft.

This aerofoil surface is provided by a pylon **28**, of horizontal symmetrical aerofoil cross-section, which is provided with a flap **29**. The pylon **28** is located between the fuselage **21** of the aircraft and a motor housing or nacelle **30**, containing electric motors (not shown) which are energized by direct current provided from the rectenna for driving propellers **31** at opposite ends of the nacelle **30**. This arrangement minimizes aerodynamic interference between the canard **27**, the wings **22** and the lenticular body **25**. Also, the location of the propellers **31** at opposite ends of the nacelle **30** on the pylon **28** serves to provide protection for the propellers **31** from damage during launching and recovery of the aircraft. As shown in FIG. **4**, end plate fairings **34**, which extend over the extent of travel of the flap **29**, are provided in order to maximize the efficiency of the flap **29** by counteracting the creation of turbulence at the ends of the flap **29**.

Since, as indicated above, the wings **22** are designed with a large wing aspect ratio in order to enhance the aerodynamic efficiency of the aircraft, and since a large aspect ratio wing can result in some degree of spiral instability, it may be necessary to ensure lateral-directional stability by artificial means. Consequently, the effect of the pylon **28**, acting as a forward fin, is not in itself a design criterion and, consequently, the dimensions of the pylon **28** may be selected sufficiently large to accommodate relatively large sideways-facing mission payload sensors, e.g. radar antennas, indicated generally by reference numeral **35** in FIG. **4**.

While, as indicated above, no propulsion means have been shown in FIG. **2**, it is to be understood that the aircraft shown in FIG. **2** is to be provided with a pylon **28**, a trailing edge flap **29**, a nacelle **30**, propellers **31** and one or more electrical propulsion motors as in the embodiment of FIG. **3**.

Various modifications may be made in the above-described aircraft for example, the rectenna may include microwave antennas and associated rectifying elements provided at the underside of the aircrafts wings, in addition to those provided at the lenticular body. Also, the invention is not restricted to fixed wing aircraft but may be utilized in rotary wing aircraft.

We claim:

1. A microwave-powered aircraft, comprising:

means for propelling the aircraft;

means defining lifting surfaces on the aircraft for exerting lifting forces on the aircraft in response to the propulsion of the aircraft;

means for receiving and rectifying microwave energy transmitted to the aircraft from a location remote from the aircraft;

the means for propelling the aircraft being operable in response to the microwave energy received by the microwave energy receiving and rectifying means; and

the microwave energy receiving and rectifying means comprising a rigid body of substantially circular plan form attached to the aircraft, the horizontal dimensions of the body being a substantial fraction of the aircraft wingspan and the body being relatively shallow with a periphery which is vertically curved so as to reduce turbulence in the airstream over the body during flight, and the body being separate from the lifting means and shaped to at least substantially avoid the generation of lifting forces by the body;

a plurality of downwardly-facing microwave antennas and associated rectifying elements at the underside of the body; and

microwave reflector means provided in the body above said microwave antenna for shielding a space located above said microwave reflector means and within the body from microwave radiation.

2. A microwave-powered aircraft as claimed in claim 1, wherein the body is of substantially elliptical vertical cross-sectional shape in planes parallel to the longitudinal axis of the aircraft.

3. A microwave-powered aircraft as claimed in claim 1, wherein the means defining lifting surfaces comprise wings located forwardly of the body, which serves as means for promoting longitudinal stability of the aircraft.

4. A microwave-powered aircraft as claimed in claim 1 or 2, wherein the aircraft includes a fuselage and a canard provided on the fuselage at a location disposed forwardly of the lifting surfaces for promoting longitudinal stability and control of the aircraft.

5. A microwave-powered aircraft as claimed in claim 1, wherein the aircraft includes a fuselage, the lifting surface defining means comprising a pair of wings, and a pylon of aerofoil cross-section is mounted on the fuselage and supports the wings.

6. A microwave-powered aircraft as claimed in claim 5, wherein the aerofoil cross-section of the pylon is symmetrical.

7. The microwave-powered aircraft as claimed in claim 5, wherein the pylon includes a flap at a trailing edge of the pylon for causing a side force to act on the aircraft independently of the angle of bank of the aircraft.

8. A microwave-powered aircraft as claimed in claim 7, further including end plate fairing means extending over the extent of travel of the flap for improving the efficiency of the flap.

9. A microwave-powered aircraft as claimed in claim 5, 6 or 7, wherein the means for propelling the aircraft include at least one propeller and electric motor means associated with the propeller for driving the propeller, the electric motor means being provided at the intersection of the pylon and the wings to protect the propeller.

10. A microwave-powered aircraft as claimed in claim 5, 6 or 7, wherein the wings comprise a pair of high aspect ratio wings, and side-facing sensors are provided on the pylon.

11. A microwave-powered aircraft as claimed in claim 1 or 2, wherein the lifting surface defining means comprise a pair of high aspect ratio wings.

12. The microwave-powered aircraft claimed in claim 1, further comprising:

means for laterally deflecting the course of travel of the aircraft without causing any substantial banking of the aircraft, wherein

the said laterally deflecting means comprises a vertically extending member having a horizontal cross-section of aerofoil shape and an adjustable flap extending along a vertically trailing edge of the deflecting member.

13. A microwave-powered aircraft as claimed in claim 12, further comprising a fuselage, the lifting surface defining means comprising a pair of fixed wings and the vertically extending member supporting the wings above the fuselage.

14. A microwave-powered aircraft as claimed in claim 12, wherein the laterally deflecting means are located at least approximately above the centre of gravity of the aircraft.

15. A microwave-powered aircraft as claimed in claim 13, wherein the means for propelling the aircraft comprise at least one propeller and electric motor means mounted above the vertically extending member for rotating the propeller.

16. A microwave-powered aircraft as claimed in claim 14, further comprising a canard provided forwardly of the lifting surfaces for promoting longitudinal stability and control of the aircraft.

* * * * *

US006967462B1

(12) **United States Patent**
Landis

(10) **Patent No.: US 6,967,462 B1**
(45) **Date of Patent: Nov. 22, 2005**

(54) **CHARGING OF DEVICES BY MICROWAVE POWER BEAMING**

(75) Inventor: **Geoffrey A. Landis**, Berea, OH (US)

(73) Assignee: **NASA Glenn Research Center**, Cleveland, OH (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 149 days.

(21) Appl. No.: **10/455,139**

(22) Filed: **Jun. 5, 2003**

(51) **Int. Cl.**^{7} **H01M 10/46**
(52) **U.S. Cl.** .... **320/101**; 320/109; 322/2 R
(58) **Field of Search** .... 320/101, 106, 320/110, 108, 109; 322/2 R

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,971,454	A	7/1976	Waterbury
3,989,994	A	11/1976	Brown
4,685,047	A	8/1987	Phillips, Sr.
5,210,804	A	5/1993	Schmid
5,260,639	A	11/1993	De Young et al.
5,396,538	A	3/1995	Hong
5,503,350	A	4/1996	Foote
5,733,313	A	3/1998	Barreras, Sr. et al.
5,982,139	A	11/1999	Parise
6,114,834	A	9/2000	Parise
6,127,799	A	10/2000	Krishnan
6,239,879	B1	5/2001	Hay
6,310,960	B1	10/2001	Saaski et al.
6,474,341	B1	11/2002	Hunter et al.
6,489,745	B1	12/2002	Koreis
6,498,455	B2	12/2002	Zink et al.
2002/0027390	A1	3/2002	Ichiki et al.

OTHER PUBLICATIONS

Prado, Mark "Environmental Effects—The PowerSat Beam and the Environment" 1983 P.E.R.M.A.N.E.N.T.

Primary Examiner—Edward H. Tso

(57) **ABSTRACT**

A system for providing wireless, charging power and/or primary power to electronic/electrical devices is described whereby microwave energy is employed. Microwave energy is focused by a power transmitter comprising one or more adaptively-phased microwave array emitters onto a device to be charged. Rectennas within the device to be charged receive and rectify the microwave energy and use it for battery charging and/or for primary power. A locator signal generated by the device to be charged is analyzed by the system to determine the location of the device to be charged relative to the microwave array emitters, permitting the microwave energy to be directly specifically towards the device to be charged. Backscatter detectors respond to backscatter energy reflected off of any obstacle between the device to be charged and the microwave array emitters. Power to any obstructed microwave array emitter is reduced until the obstruction is removed. Optionally, data can be modulated onto microwave energy beams produced by the array emitters and demodulated by the device, thereby providing means of data communication from the power transmitter to the device. Similarly, data can be modulated onto the locator signal and demodulated in the power transmitter, thereby providing means of data communication from the device to the power transmitter.

20 Claims, 4 Drawing Sheets

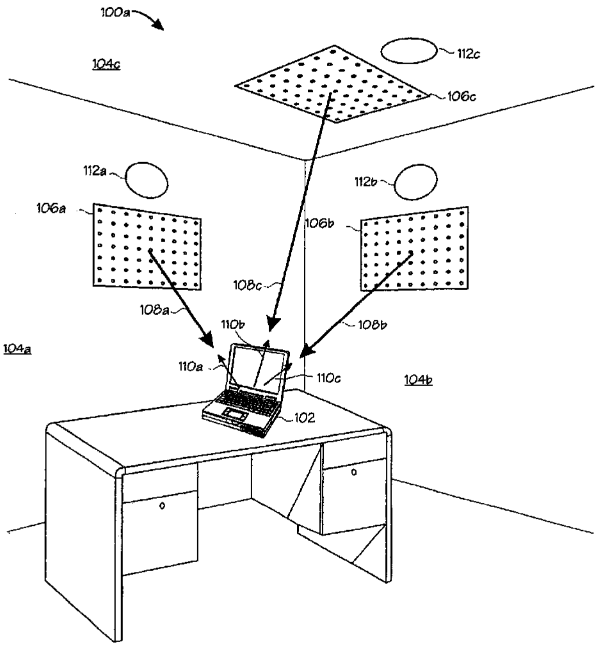

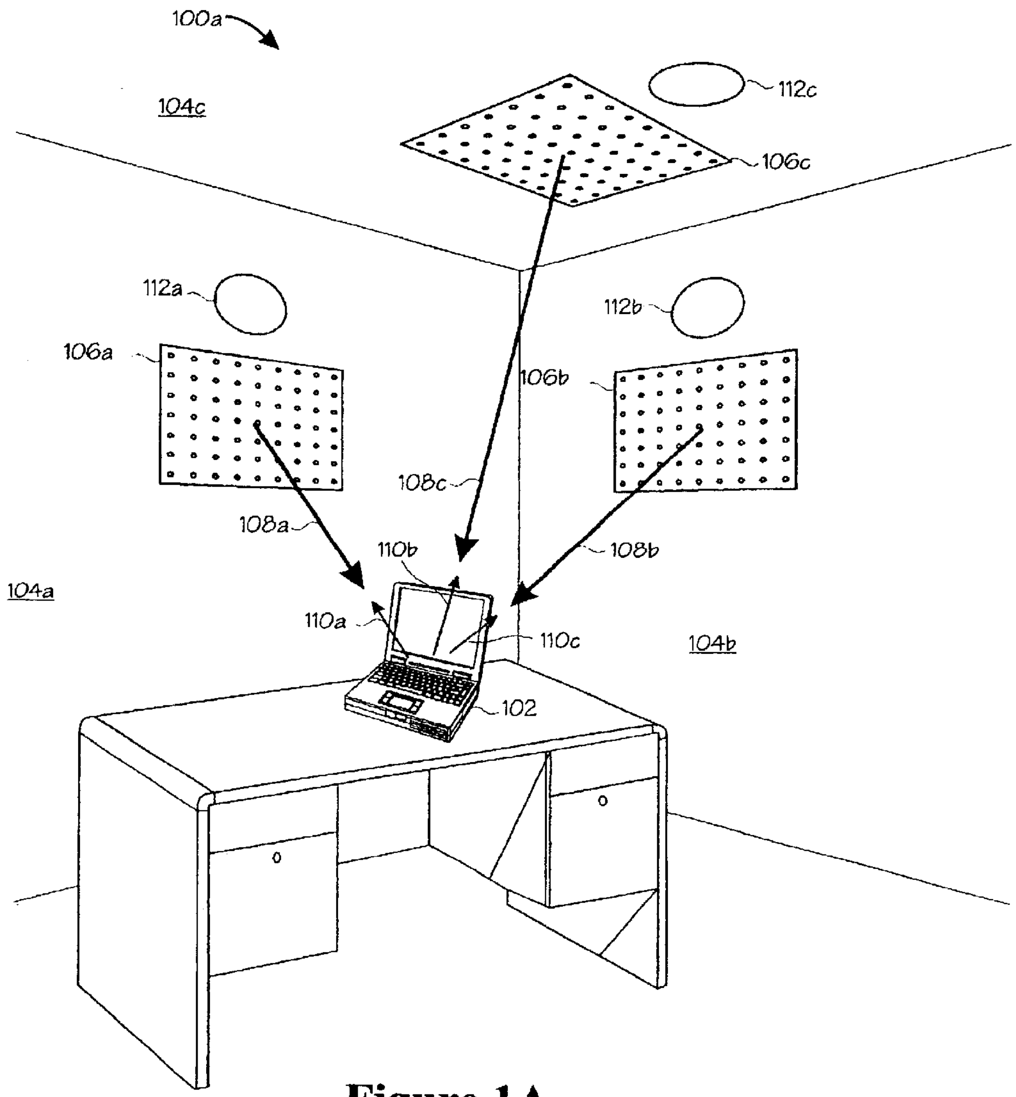

Figure 1A

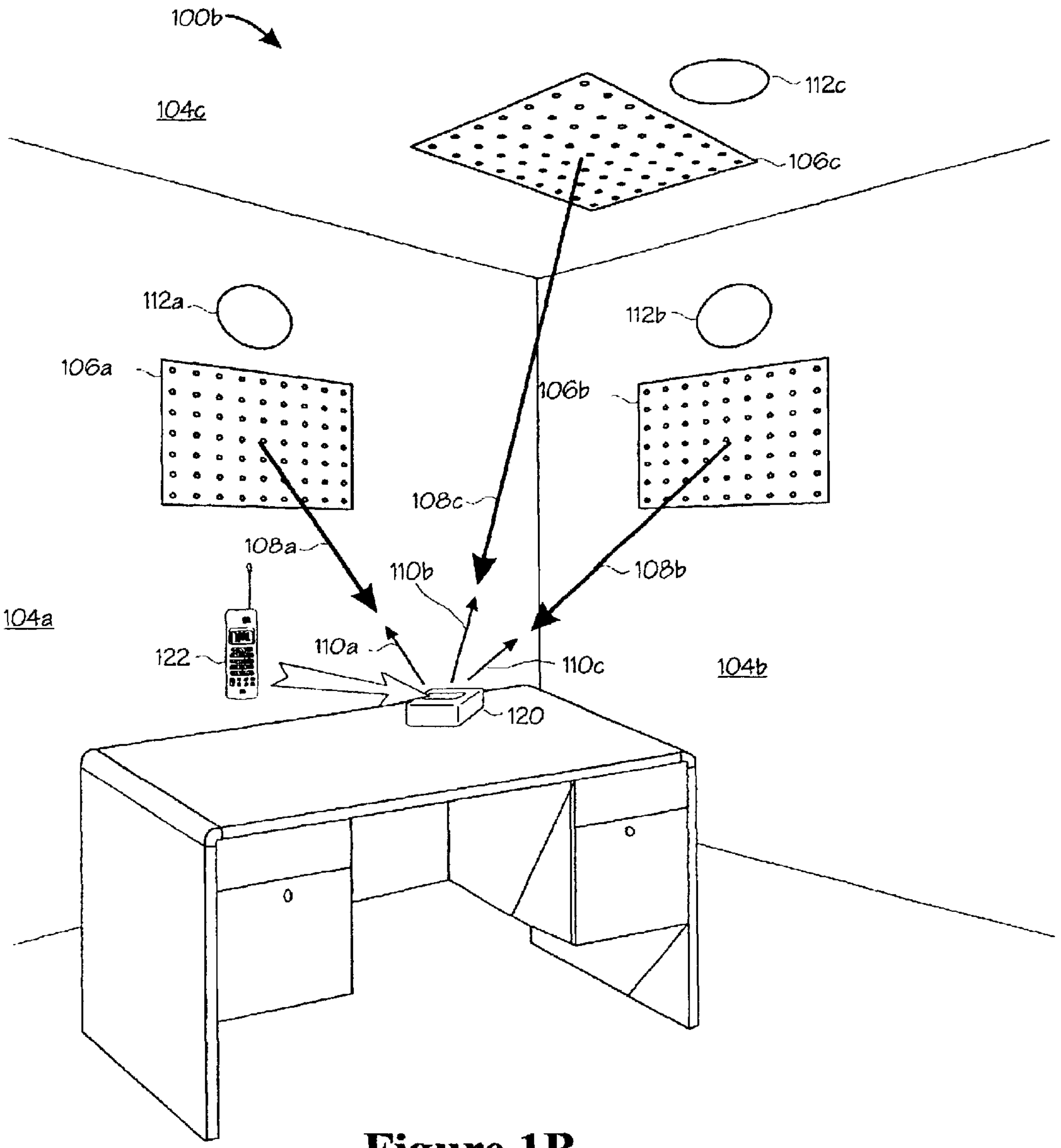

Figure 1B

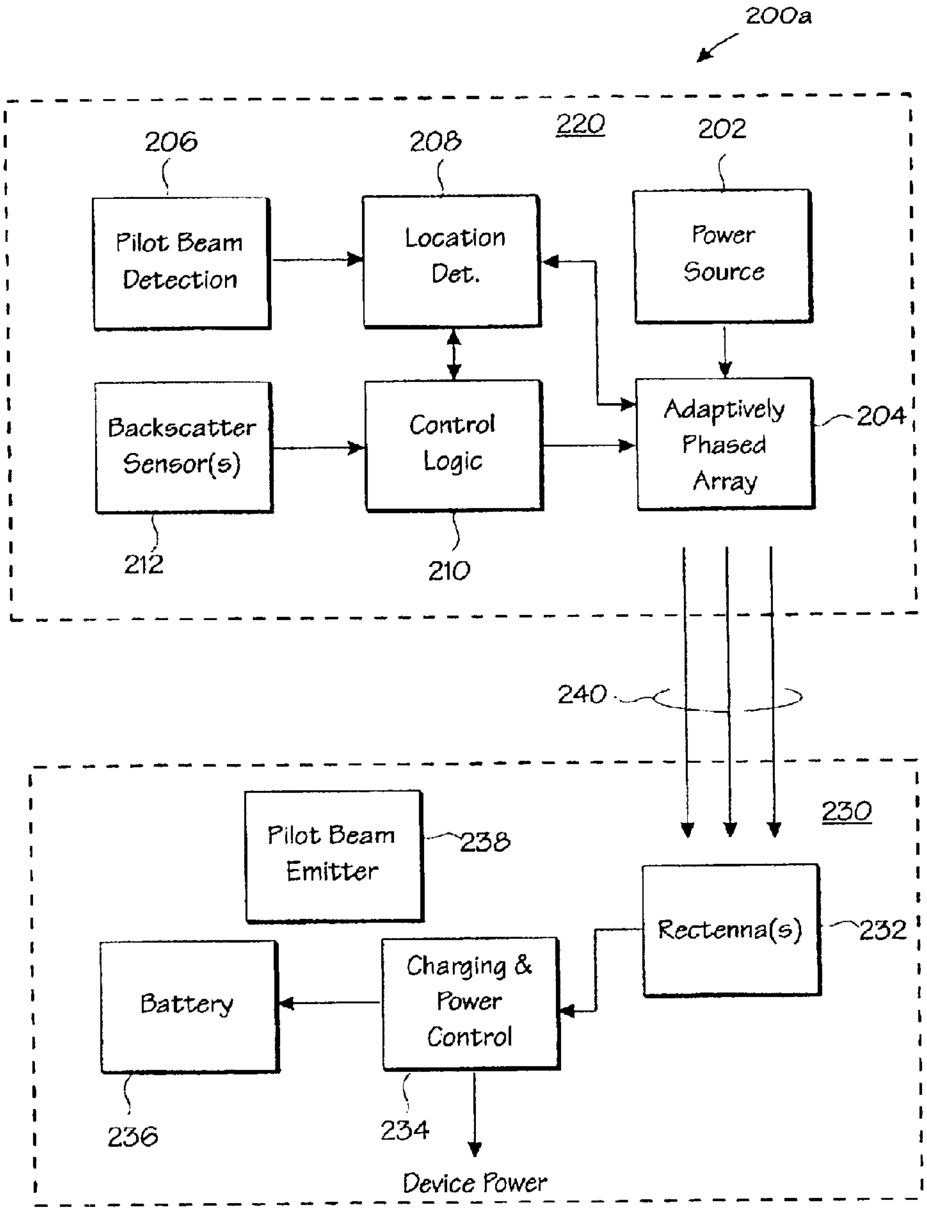
200a
220
206
Pilot Beam Detection
208
Location Det.
202
Power Source
Backscatter Sensor(s)
Control Logic
Adaptively Phased Array
204
212
210
240
230
Pilot Beam Emitter
238
Rectenna(s)
232
Battery
Charging & Power Control
236
234
Device Power

Figure 2A

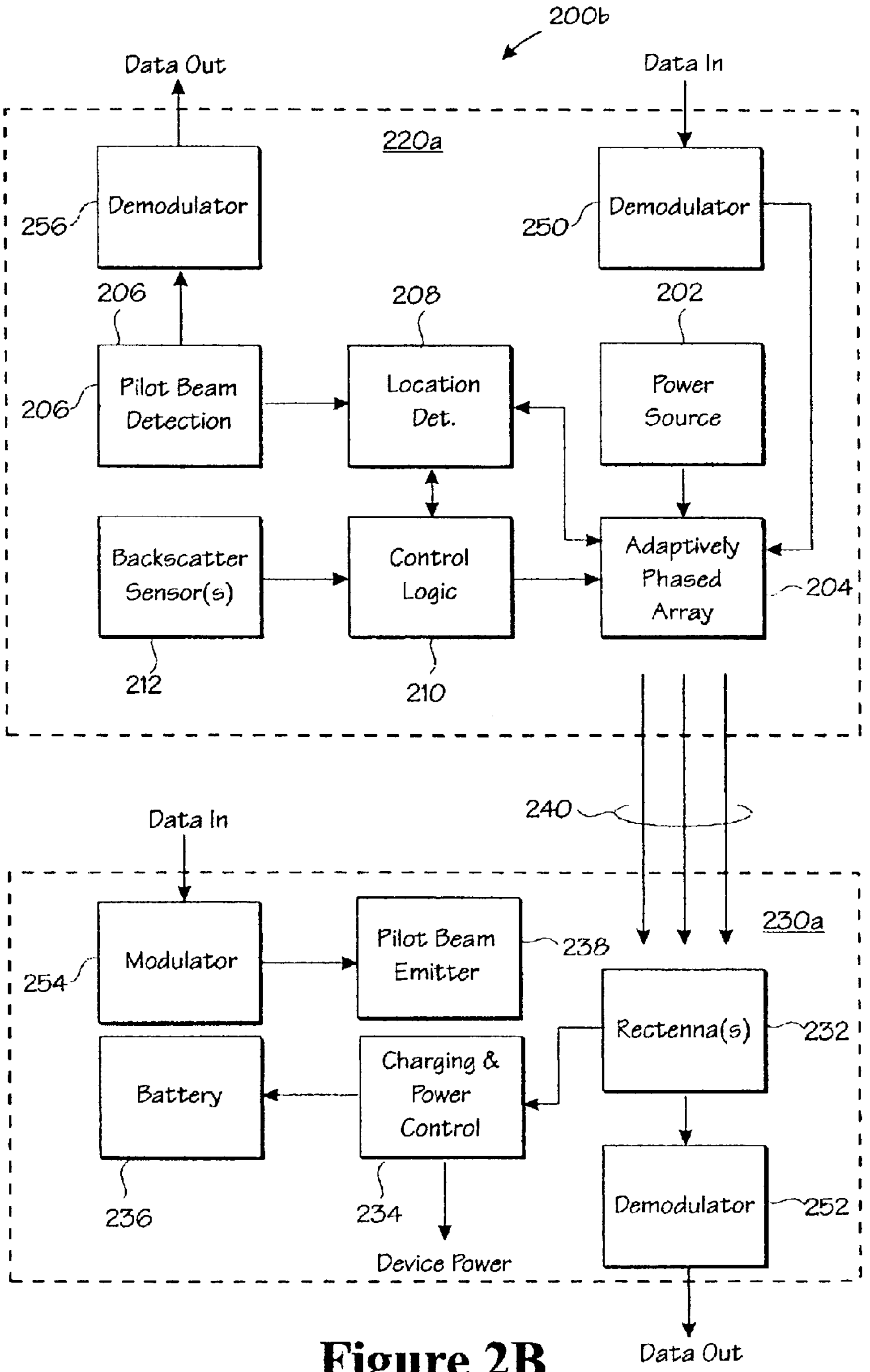
200b
Data Out
Data In
220a
Demodulator
256
Demodulator
250
206
208
202
Pilot Beam Detection
206
Location Det.
Power Source
Backscatter Sensor(s)
Control Logic
Adaptively Phased Array
204
212
210
240
Data In
230a
Modulator
254
Pilot Beam Emitter
238
Rectenna(s)
232
Battery
Charging & Power Control
236
234
Device Power
Demodulator
252
Data Out

Figure 2B

CHARGING OF DEVICES BY MICROWAVE POWER BEAMING

ORIGIN OF THE INVENTION

The invention described herein was made by an employee of the United States Government and may be manufactured and used by or for the Government for Government purposes without the payment of any royalties thereon or therefore.

TECHNICAL FIELD

The present invention relates generally to wireless charging and powering of battery-powered electrical/electronic devices, including consumer devices.

BACKGROUND ART

A recent trend in consumer electronic devices is small, battery operated devices with on-board rechargeable batteries that are recharged by using a small, external charging adapter. These consumer devices can be, e.g., cell phones, personal digital assistants (PDAs), personal stereo devices (e.g., "Walkman", or MP3 Player), laptop computers, calculators, pagers, etc. Typically, the charging adapters for these devices are wall-plug AC adapters that provide bulk, low-voltage AC or DC charging power to the consumer device for operating the device and/or charging the device's on-board battery. Often, cigarette lighter adapters and similar charging adapters are provided for charging the consumer devices from a car's electrical system.

A typical consumer device charging system comprises a source of charging power, a charging circuit and a rechargeable battery. The charging circuit typically controls the amount of charging current delivered to the consumer device's battery, sensing and adjusting the current according to the battery's state of charge. Often, such charging circuits provide additional power for operating the consumer device while it is connected to charging adapter, dividing available power between battery charging and device operation as required.

Sometimes the charging circuit resides within the consumer device; other times the charging circuit is built into the charging adapter. The practice of putting the charging circuit into the charging adapter can be particularly advantageous for very small consumer devices, as it eliminates circuitry from the consumer device itself. Although such charging circuits are typically very small themselves, they can be large in comparison to other circuitry in consumer devices and the space saved by eliminating them can have a significant effect on the size, complexity and cost of the consumer device.

Usually, the charging adapter plugs into the consumer device by means of a wired connection. Sometimes, a connector or a set of contacts on the exterior of the consumer device is arranged so that the act of placing it into a charging "cradle" establishes a physical/electrical connection for charging the device. In this case, the "cradle" is a part of the charging adapter. Most charging systems for consumer devices require a direct electrical connection of this type between the device to be charged and the charging adapter. Over time, these contacts can become worn, loose or dirty, compromising their physical and/or electrical integrity and making them unreliable.

One scheme for powering and/or charging electrical/electronic devices involves electromagnetically coupled coils. A first coil (or antenna) is disposed within a charging apparatus. A second coil (or antenna) is disposed within the device to be charged. A "transmitting" circuit drives the first coil with an AC waveform. The device to be charged is placed in close proximity to the charging apparatus such that an electro-magnetic field produced by the first coil induces a corresponding AC electrical signal in the second coil. By rectifying the electrical signal in the second coil, a source of charging energy for the device to be charged is realized. In effect, the two coils (or antennae) form a transformer by which electrical power and/or signals can be communicated between the charging apparatus and the device to be charged. Such non-contacting charging systems have also been adapted to provided data communication over the same electro-magnetic coupling that provides the charging energy by modulating the charging signal and/or load impedance.

SUMMARY OF THE INVENTION

The present inventive technique provides for wireless, charging power and/or primary power to electronic/electrical devices whereby microwave energy is employed. The microwave energy is focused by one or more adaptively-phased microwave array emitters in a power transmitter portion of the system onto a device to be charged. Rectennas within the device to be charged receive and rectify the microwave energy and use it for battery charging and/or for primary power. A locator signal generated by the device to be charged is analyzed by the system to determine the location of the device to be charged relative to the microwave array emitters, permitting the microwave energy to be directly specifically towards the device to be charged. Backscatter detectors respond to backscatter energy reflected off of any obstacle between the device to be charged and the microwave array emitters. Power to any obstructed microwave array emitter is reduced until the obstruction is removed. Optionally, data can be modulated onto microwave energy beams produced by the array emitters and demodulated by the device, thereby providing means of data communication from the power transmitter to the device. Similarly, data can be modulated onto the locator signal and demodulated in the power transmitter, thereby providing means of data communication from the device to the power transmitter.

The power transmitter portion of the system includes one or more (preferably planar) adaptively-phased microwave array emitters, and location detection means responsive to the location signal generated by the device to be charged. Typically, the array emitters would be mounted to walls and or floors of a room in which the device to be charged resides. The phase of microwave energy produced by the array emitters is continuously varied across the face(s) of the array emitter(s) to produce the effect of a focused beam of microwave energy, converging at the device. One or more rectennas within the device receive the transmitted microwave energy, converting and rectifying it into DC electrical energy useful for battery charging and/or device power.

According to an aspect of the invention, device location is determined by producing "pilot beams" (locator signal) at the location of the device to be charged to be received by the power transmitter. By analyzing timing characteristics of the received signal (e.g., by triangulation) the power transmitter determines the exact location of the device to be charged relative to its array emitters and adjusts phase of the microwave energy across the surface of the array emitters to focus the energy on the device.

According to another aspect of the invention, the array emitters produce continuous-wave microwave energy in the frequency range between 2 GHz and 10 GHz inclusive. Frequencies higher than 10 GHz can be employed (e.g., millimeter wave), but using current rectenna technology, there would be a loss of efficiency.

According to another aspect of the invention, each individual microwave array emitter produces energy at a frequency close to, but different from that produced by any other microwave array emitter in the system. This reduces loss of energy in sidelobes, and minimizes interference issues.

According to another aspect of the invention, the power transmitter includes backscatter detectors for detecting backscatter energy reflected off of any obstruction between the microwave array emitters and the device to be charged. When an obstruction is detected, the obstructed, the power output of the obstructed microwave array emitter is reduced to a low level until the obstruction is removed, thereby reducing lost power and preventing injury to humans or damage to objects that obstruct the power transmission path between the arrays and the device.

According to another aspect of the invention, conventional electrical/electronic devices (i.e., those not specifically adapted to microwave charging) can be accommodated by means of a microwave charging adapter/cradle that embodies the inventive technique for receiving and converting microwave energy. In this case, the charging cradle/adapter would generate the locator signal and receive the microwave energy.

According to another aspect of the invention, data can be modulated onto the microwave energy produced by the adaptively phased microwave array emitters by means of a modulator function, thereby producing microwave energy beams having a (DC) power component and a (AC) data component. A demodulator function in the device (or charging adapter/cradle) demodulates the data component of the microwave energy beam as received by the rectennas. This provides means of communicating data from the power transmitter to the device.

According to another aspect of the invention, data can be modulated onto the locator signal produced by the device to be charged (or by the charging cradle/adapter) by means of a modulator function. A corresponding demodulator function in the power transmitter demodulates the data modulated onto the locator signal, thereby providing means of communicating data from the device to the power transmitter.

BRIEF DESCRIPTION OF THE DRAWINGS

These and further features of the present invention will be apparent with reference to the following description and drawing, wherein:

FIG. **1**A is a view of a system for direct microwave charging of an electrical/electronic device, in accordance with the invention.

FIG. **1**B is a view of a system for microwave charging of an electrical/electronic device in a charging cradle, in accordance with the invention.

FIG. **2**A is a block diagram of a system for microwave charging, in accordance with the invention.

FIG. **2**B is a block diagram of a system for microwave charging, including means for data exchange, in accordance with the invention.

DETAILED DESCRIPTION OF THE INVENTION

The present inventive technique provides for charging and/or powering of an electrical/electronic device using microwave energy. One or more adaptive-phased arrays are used to focus one or more "rectennas" (rectifying antennas) disposed within the consumer device. Microwave energy reaching the rectenna(s) is converted into DC electrical energy that is used to charge a battery or other energy storage device within the consumer device. The DC electrical energy can also be used to provide primary power for the consumer device while the battery or storage device is being charged.

FIG. **1**A is a view of a system **100***a* for direct microwave charging of an electrical/electronic device **102**, wherein a plurality of adaptive-phased array microwave emitters **106***a*, **106***b* and **106***c* are disposed on walls **104***a*, **104***b* and ceiling **104***c*, respectively of a room in which the electrical/electronic device **102** resides. Pilot beams **100***a*, **100***b*, **110***c* from the electrical/electronic device **102** permit the system **100** to determine the exact location of the electrical/electronic device **102** within the room. Each of the adaptive-phased arrays **106***a*, **106***b* and **106***c* is then driven to emit continuous-wave microwave energy with varying phase across the array surface in order to effectively focus a respective beam **108***a*, **108***b*, **108***c* of microwave energy directly at the electrical/electronic device **102**. One or more rectennas (described in greater detail hereinbelow) within the electrical/electronic device **102** receive the microwave energy focused thereupon and convert it into a source of charging and/or operating power therefor. Backscatter detectors **112***a*, **112***b* and **112***c*, mounted in close proximity to adaptive-phased arrays **106***a*, **106***b* and **106***c*, respectively, detect "backscatter", i.e., microwave energy reflected off of any obstacle (e.g., a human) that might enter the beam path. When significant backscatter is detected by one of the backscatter detectors **112***a*, **112***b* or **112***c*, its respective adaptive-phased array **112***a* is either turned off completely or reduced to a low level of power emission until the beam is clear of the obstruction. This prevents transmission of wasted microwave power that will not reach the electrical/electronic device **102**, and also guards against injury or damage to a human (or other obstacle) that enters a beam path. If any beam is obstructed (**108***a*, **108***b* or **108***c*), the remaining beams can still power the electrical/electronic device **102**.

Preferably, the adaptive-phased arrays operate in the range of 2 to 10 GHz (2×10^9 Hz to 10×10^9 Hz). Higher frequencies, e.g., "millimeter wave" frequencies can be employed, but current millimeter wave rectenna technology is lower in efficiency than rectenna technology designed for the 2 to 10 GHz range. It is also preferable that the adaptive phase arrays (**106***a*, **106***b* and **106***c*) operate at slightly different frequencies, i.e., it is preferable that the adaptive-phased arrays **106***a*, **106***b*, **106***c* transmit mutually incoherent microwave signals. While having all of the adaptive-phased arrays operate on the same frequency would result in a smaller beam "spot" (focal point) on the electrical/electronic device to be charged/powered, it would also result in more of the beam power being scattered into sidelobes, yielding lower efficiency (wasted power in the sidelobes) and greater interference between arrays.

The "pilot beams" **110***a*, **110***b* and **110***c* are essentially a locator signal, and can be provided by any suitable means of identifying the location of the electrical/electronic device **102** relative to the adaptive-phased arrays **106***a*, **106***b* and **106***c*. This can be an RF signal, a microwave signal, or any other suitable locator signal from which the location of the electrical/electronic device **102** can be determined by the system **100***a*.

Data transfer between the charging system **100***a* and the electrical/electronic device **102** is readily accomplished by modulating data onto one or more of the charging beams **108***a*, **108***b*, **108***c*, and demodulating the modulated component in the electrical/electronic device. Similarly, the electrical/electronic device **102** can modulate either one or more of its pilot beams **110***a*, **110***b* or **110***c* (or whatever "locator" signal it employs) with data to be transferred to the system **100***a*. This data transfer mechanism is described in greater detail hereinbelow with respect to FIG. **2**B.

FIG. **1**B is a view of a similar system **100***b* for microwave charging of an electrical/electronic device **122** in a charging cradle/adpater **120**. This embodiment is essentially identical to that of FIG. **1**A, except that the electrical/electronic device **102** of FIG. **1**A had the beam-receiving rectennas and pilot beam generation built-in, while the implementation of FIG. **1**B separates the beam-receiving rectennas and pilot beam generation into a separate charging cradle/adapter **120**. In effect, the combination of electrical/electronic device **122** and charging cradle/adapter **120** in FIG. **1**B is comparable to the electrical/electronic device **102** of FIG. **1**A. Pilot beams **110***a*, **110***b*, and **110***c* can be "gated" such that they are only generated when the electrical/electronic device **122** is "docked" in its charging cradle/adapter **120**, thereby signaling to the system **100***b* that the adaptive-phased arrays can be powered down, since there is no need for powering/charging otherwise.

One advantage of the system **100***a* of FIG. **1**A over the system **100***b* of FIG. **1**B is that the fully-integrated electrical/electronic device **102** required no charging adapter, and can be charged or powered simply by being placed in the vicinity of a suitable charging system. The system **100***b* of FIG. **1**B has the advantage that it readily adapted to accept existing electrical/electronic devices such as cell phones and PDAs without modification.

FIG. **2**A is a block diagram of a system **200***a* for microwave charging an electrical/electronic device **230** by means of focused, microwave beams **240**. A power transmitter portion **220** of the system comprises a power source **202** that powers one or more adaptively-phased arrays **204** to produce one or more directed ("focused") beams **240** of microwave energy aimed at the electrical/electronic device **230**. Pilot beam detection **206**, detects a locator signal emitted by a pilot beam emitter **238** in the electrical/electronic device. A location detection function **208** analyzes the timing of the locator signal to determine the location of the electrical/electronic device **230** relative to the one or more adaptively-phased arrays **204**. The phase of transmitted microwave energy is varied continuously across the face of the adaptively-phased arrays **204** to produce beam(s) **240** that converge on one or more rectenna elements **232** in the electrical/electronic device. Microwave energy received by the rectennas **232** is rectified and converted thereby into DC electrical energy which is in turn presented to a charging and power control function block **234** for charging a battery **236** and/or for providing primary power to the electrical/electronic device **230**.

Backscatter sensors **212** in the power transmitter portion **220** detect reflected microwave energy indicative of an obstacle in the microwave beams. Control logic responds to signal from the backscatter detector and reduces (or cuts off) the power transmitted by any adaptively-phased array **204** whose beam path is obstructed. Preferably, the power of the obstructed beam is lowered to a "safe" level and the backscatter detectors are continually monitored to determine when the beam is clear, at which time full power to the affected beam can once again be restored. Additionally, the location detection function **208** can provide an indication of presence or absence of an electrical/electronic device **230** to be charged, enabling power transmission by the adaptively-phased arrays **204** only when an appropriately adapted electrical/electronic device **230** is present.

FIG. **2**B is a block diagram of a similar system **200***b* for microwave charging of an electrical/electronic device **230***a* (compare **230**) by a power transmitter portion **220***a* (compare **220**) of the system, but adding means for data exchange between the electrical/electronic device **230***a* and the power transmitter portion **220***a* As in the system **200** of FIG. **2**A, the power transmitter portion **220***a* of the system comprises a power source **202** that powers one or more adaptively-phased arrays **204** to produce one or more directed ("focused") beams **240** of microwave energy aimed at the electrical/electronic device **230***a*. In addition, however, a Modulator function **250** modulates the microwave energy generated by the adaptively phased arrays **204** such that the focused power beams **240** carry both power (a "DC" or continuous beam component) and data (an "AC" or varying beam component). A demodulator function **252** in the electrical/electronic device **230***a* demodulates or "decodes" the varying beam component as received by the rectennas **232**, and reproduces the data used to produce the modulation.

Pilot beam detection **206**, detects a locator signal emitted by a pilot beam emitter **238** in the electrical/electronic device **230***a* In this case, however, a modulator function **254** modulates data onto the locator signal produced by the pilot beam emitter. A demodulator function **256** in the power transmitter portion **220***a* decodes (demodulates) the data modulated onto the locator signal. A location detection function **208** analyzes the overall timing of the locator signal to determine the location of the electrical/electronic device **230***a* relative to the one or more adaptively-phased arrays **204**. The phase of transmitted microwave energy is varied continuously across the face of the adaptively-phased arrays **204** to produce beam(s) **240** that converge on one or more rectenna elements **232** in the electrical/electronic device. Microwave energy received by the rectennas **232** is rectified and converted thereby into DC electrical energy which is in turn presented to a charging and power control function block **234** for charging a battery **236** and/or for providing primary power to the electrical/electronic device **230***a*.

As in the system **200** of FIG. **2**A, backscatter sensors **212** in the power transmitter portion **220***a* detect reflected microwave energy indicative of an obstacle in the microwave beams. Control logic responds to signal from the backscatter detector and reduces (or cuts off) the power transmitted by any adaptively-phased array **204** whose beam path is obstructed. Preferably, the power of the obstructed beam is lowered to a "safe" level and the backscatter detectors are continually monitored to determine when the beam is clear, at which time full power to the affected beam can once again be restored. Additionally, the location detection function **208** can provide an indication of presence or absence of an electrical/electronic device **230***a* to be charged, enabling power transmission by the adaptively-phased arrays **204** only when an appropriately adapted electrical/electronic device **230***a* is present.

Those of ordinary skill in the art will immediately understand that the electronic/electronic device **230** of FIG. **2**A could be charged by the power transmitter portion **220***a* of FIG. **2**B, but that there would be no data exchange capability due to the lack of compatible data exchange circuitry in the electrical/electronic device **230**. Similarly, the electrical/electronic device **230***a* of FIG. **2**B could be charged by the power transmitter **220** of FIG. **2**A, but there would be no data exchange capability due to the lack of compatible data exchange circuitry in the power transmitter **220**.

Those of ordinary skill in the art will immediately understand that the systems **200** and **200***a* of FIGS. **2**A and **2**B, respectively, can be applied either to fully-integrated, microwave-chargeable electrical/electronics devices (e.g., **102**, FIG. **1**A) or to microwave charging stations for conventional electrical/electronic devices (e.g., **122**, **120**, FIG. **1**B) by making the appropriate functional divisions. Specifically, only the battery (e.g., **236**, FIG. **2**B) need reside within a conventional electrical/electronic device (e.g., **122**, FIG. **1**B). All other components of the microwave-chargeable electrical electronic device (see **230**, **230***a*, FIGS. **2**A, **2**B) can be integrated into a charging station (e.g., **120** FIG. **1**B).

Although the invention has been shown and described with respect to a certain preferred embodiment or embodiments, certain equivalent alterations and modifications will occur to others skilled in the art upon the reading and understanding of this specification and the annexed drawings. In particular regard to the various functions performed by the above described components (assemblies, devices, circuits, etc.) the terms (including a reference to a "means") used to describe such components are intended to correspond, unless otherwise indicated, to any component which performs the specified function of the described component (i.e., that is functionally equivalent), even though not structurally equivalent to the disclosed structure which performs the function in the herein illustrated exemplary embodiments of the invention. In addition, while a particular feature of the invention may have been disclosed with respect to only one of several embodiments, such feature may be combined with one or more features of the other embodiments as may be desired and advantageous for any given or particular application.

What is claimed is:

1. A system for charging of devices, comprising:

a power transmitter having a plurality of adaptively-phased microwave array emitters;

a device to be charged having one or more rectennas associated therewith; and

location determining means for determining a location of the device to be charged relative to the adaptively-phased array emitters;

wherein each individual microwave array emitter produces energy at a frequency close to, but different from that produced by any other microwave array emitter in the system.

2. A system according to claim **1**, further comprising:

means for exchanging data between the power transmitter and the device to be charged.

3. A system according to claim **2**, further comprising:

modulating means for modulating a data signal onto a locator signal produced by the device to be charged; and

demodulating means within the power transmitter for demodulating the data modulated onto the locator signal.

4. A system for charging of devices, comprising:

a power transmitter having one or more adaptively-phased microwave array emitters;

a device to be charged having one or more rectennas associated therewith;

location determining means for determining a location of the device to be charged relative to the adaptively-phased array emitters; and

one or more backscatter detectors associated with the adaptively-phased microwave array emitters for detecting microwave backscatter energy reflected off of any obstruction between the adaptively-phased microwave array emitters and the device to be charged.

5. A system according to claim **4**, wherein:

the location determining means are employed to focus energy produced by the adaptively-phased microwave array emitters onto the device to be charged.

6. A system according to claim **4**, wherein:

the location determining means further comprise:

a pilot beam emitter in the device to be charged for producing a locator signal;

a pilot beam detector in the power transmitter for detecting the locator signal; and

location detection means for analyzing the locator signal to determine the location of the device to be charged.

7. A system according to claim **4**, wherein:

the adaptively-phased microwave array emitters produce continuous-wave microwave energy in the frequency range between 2 GHz and 10 GHz inclusive.

8. A system according to claim **4**, wherein:

each separate adaptively-phased microwave array emitter produces microwave energy at a unique frequency close to but different from the frequency produced by any other adaptively-phased microwave array emitter in the system.

9. A system according to claim **4**, further comprising:

means for reducing power transmitted by any adaptively-phased microwave array emitter for which an obstruction exists between it and the device to be charged.

10. A system according to claim **9**, further comprising:

means for restoring full power to the obstructed adaptively-phased microwave array emitter upon removal of the obstruction.

11. A system according to claim **4**, wherein:

the device to be charged is connected to a charging cradle/adapter within which the rectennas are disposed that receives microwave energy transmitted by the one or more adaptively-phased microwave array emitters and transfers it to the device to be charged.

12. A system according to claim **4**, further comprising:

means for exchanging data between the power transmitter and the device to be charged.

13. A system according to claim **12**, wherein:

the location determining means are employed to focus energy produced by the adaptively-phased microwave array emitters onto the device to be charged.

14. A system according to claim **12**, wherein:

the location determining means further comprise:

a pilot beam emitter in the device to be charged for producing a locator signal;

a pilot beam detector in the power transmitter for detecting the locator signal; and

location detection means for analyzing the locator signal to determine the location of the device to be charged.

15. A system according to claim **12**, further comprising:

means for reducing power transmitted by any adaptively-phased microwave array emitter for which an obstruction exists between it and the device to be charged.

16. A system according to claim **15**, further comprising:

means for restoring full power to the obstructed adaptively-phased microwave array emitter upon removal of the obstruction.

17. A system according to claim **12**, further comprising:

modulating means for modulating data onto microwave energy beams produced by the one or more adaptively-phased microwave array emitters; and demodulating means for demodulating the data modulated onto the microwave energy beams.

18. A system according to claim **17**, wherein:

the rectennas and demodulating means reside within the device to be charged.

19. A system according to claim **17**, wherein:

the rectennas and demodulating means reside within a charger adapter/cradle to which the device to be charged is connected,

the device to be charged is situated in a charging cradle/adapter that receives microwave energy transmitted by the one or more adaptively-phased microwave array emitters and transfers it to the device to be charged.

20. A system for device charging, comprising:

a power transmitter portion further comprising:

a power source;

one or more adaptively-phased microwave array emitters;

means for receiving a location signal indicative of the location of a device to be charged;

location detection means for determining the location of the device to be charged from the locator signal;

means for focusing energy produced by the one or more adaptive-phase microwave array emitters onto the devices to be charged; and

control logic means for reducing the amount of energy produced by any obstructed adaptively phased microwave array emitter until the obstruction is removed; and

a device to be charged comprising:

a battery;

one or more rectennas for receiving and rectifying energy produced by the adaptively-phased microwave array emitters; and

charging and power control means for charging the battery from rectified energy received by the rectennas.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,967,462 B1
DATED : November 22, 2005
INVENTOR(S) : Landis

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 10,
Line 3, after "means for focusing energy produced by the one or more adaptively-phased microwave array emitters onto the device to be charged;" insert the following paragraph:
-- backscatter detection means for detecting microwave energy reflected off of any obstruction between the one or more adaptively-phased microwave array emitters and the device to be charged; and --.

Signed and Sealed this

Twenty-eighth Day of February, 2006

Jon W. Dudas

JON W. DUDAS
Director of the United States Patent and Trademark Office

US007521890B2

(12) **United States Patent**
Lee et al.

(10) **Patent No.: US 7,521,890 B2**
(45) **Date of Patent: Apr. 21, 2009**

(54) **SYSTEM AND METHOD FOR SELECTIVE TRANSFER OF RADIO FREQUENCY POWER**

(75) Inventors: **Yeechun Lee**, Mountain View, CA (US); **Victor Su**, Mountain View, CA (US)

(73) Assignee: **Power Science Inc.**, San Jose, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 324 days.

(21) Appl. No.: **11/319,818**

(22) Filed: **Dec. 27, 2005**

(65) **Prior Publication Data**

US 2007/0145830 A1 Jun. 28, 2007

(51) **Int. Cl.**
H01M 10/46 (2006.01)
(52) **U.S. Cl.** **320/108**
(58) **Field of Classification Search** 320/108, 320/109, 112, 114, 116
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,803,744 B1 * 10/2004 Sabo 320/108
6,954,180 B1 * 10/2005 Braun et al. 343/702
7,245,269 B2 * 7/2007 Sievenpiper et al. 343/909
2006/0043927 A1 * 3/2006 Beart et al. 320/108

* cited by examiner

Primary Examiner—Edward Tso
(74) *Attorney, Agent, or Firm*—Holland & Knight LLP; John P. Moran

(57) **ABSTRACT**

A system and method is provided for the inductive transfer of electric power between a substantially flat primary surface and a multitude of secondary devices in such a way that the power transfer is localized to the vicinities of individual device coils. The contact free power transfer does not require precise physical alignment between the primary surface and the secondary device and can allow the secondary device or devices to be placed anywhere and in arbitrary orientation with respect to the primary surface. Such power transfer is realized without the need of complex high frequency power switching network to turn the individual primary coils on or off and is completely scalable to almost arbitrary size. The local anti-resonance architecture of the primary device will block primary current from flowing when no secondary device or devices are in proximity to the local RF power network. The presence of a tuned secondary device detunes the local anti-resonance on the primary surface; thereby enable the RF power to be transferred from the local primary coils to the secondary device. The uniformity of the inductive coupling between the active primary surface and the secondary devices is improved with a novel multi-pole driving technique which produces an apparent traveling wave pattern across the primary surface.

25 Claims, 16 Drawing Sheets

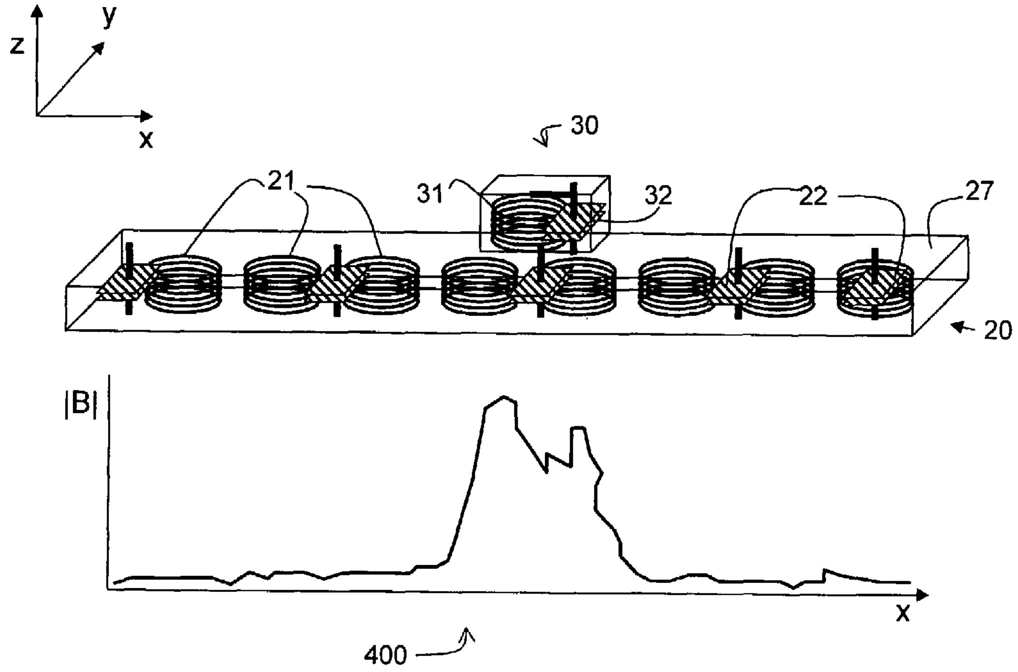

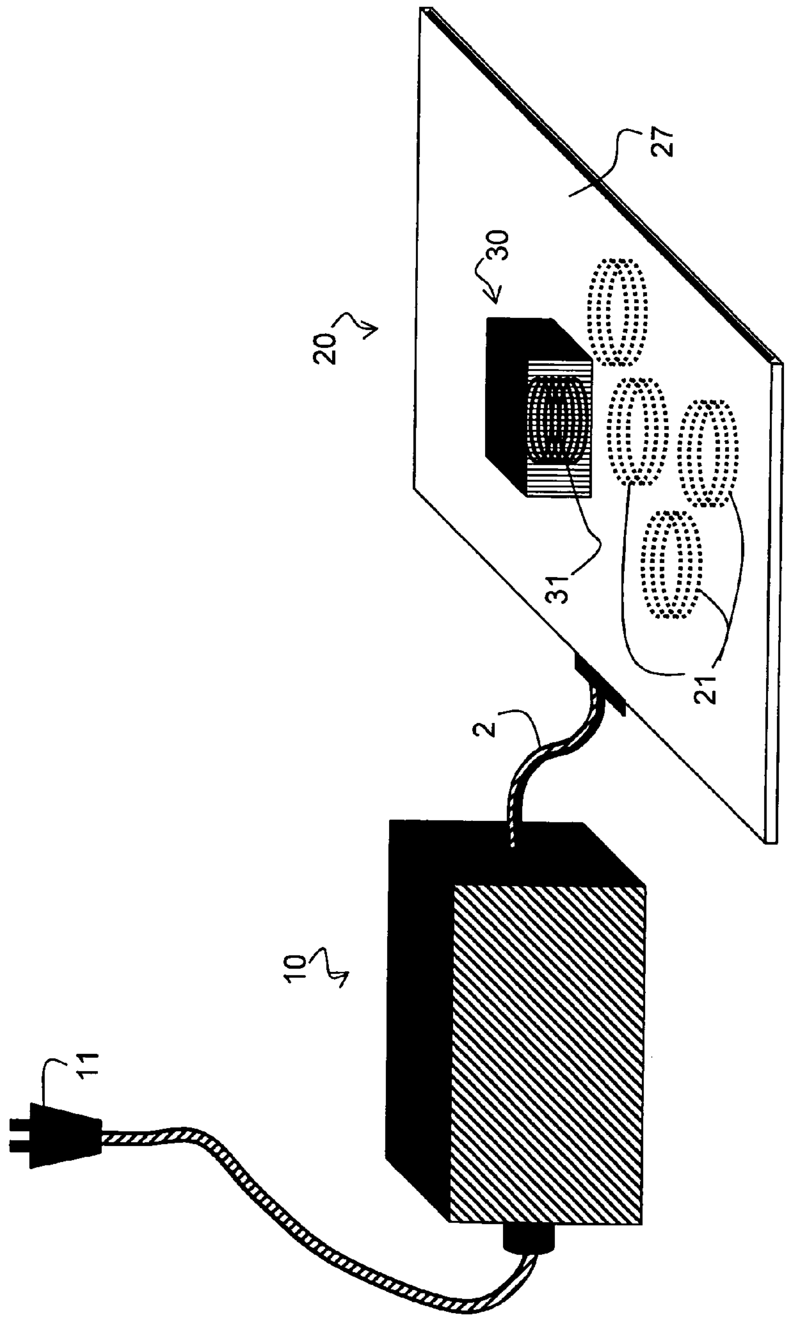
11
10
2
20
30
31
21
27

FIGURE 1

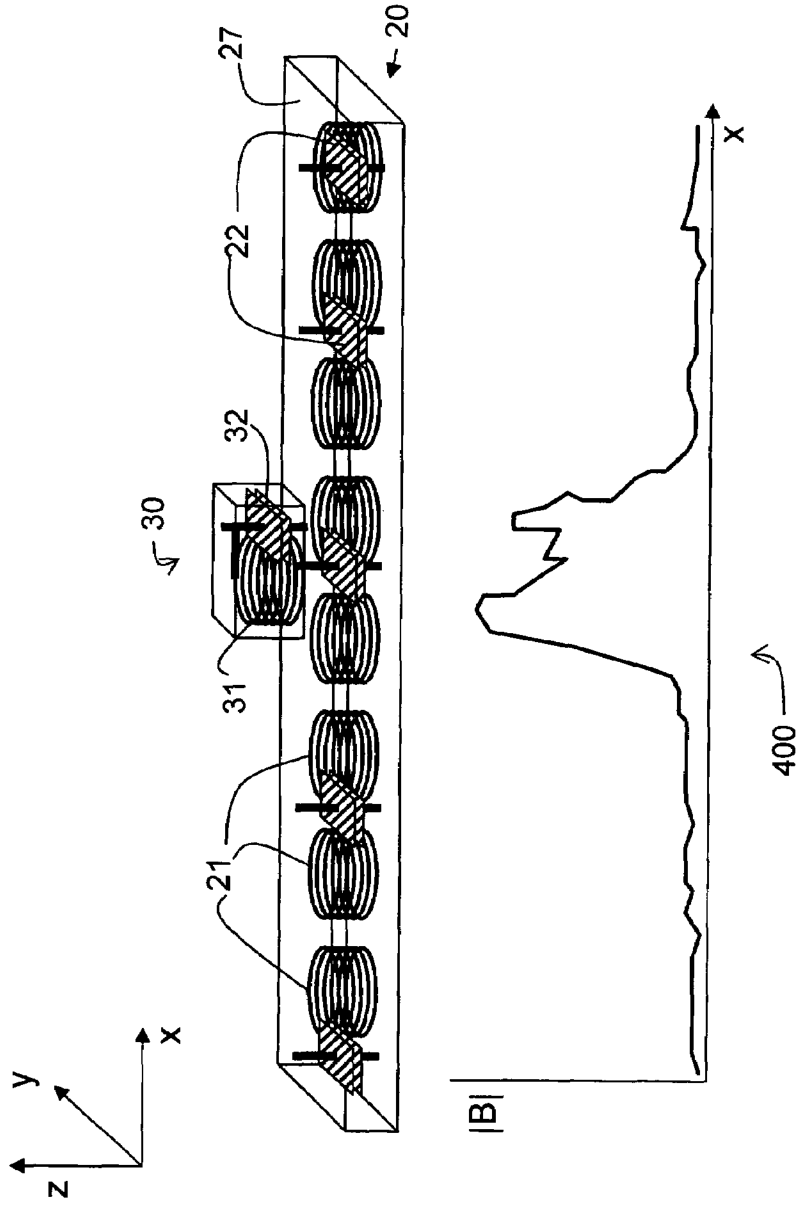
27
20
22
32
30
31
21
400
|B|
x
y
z
x

FIGURE 2

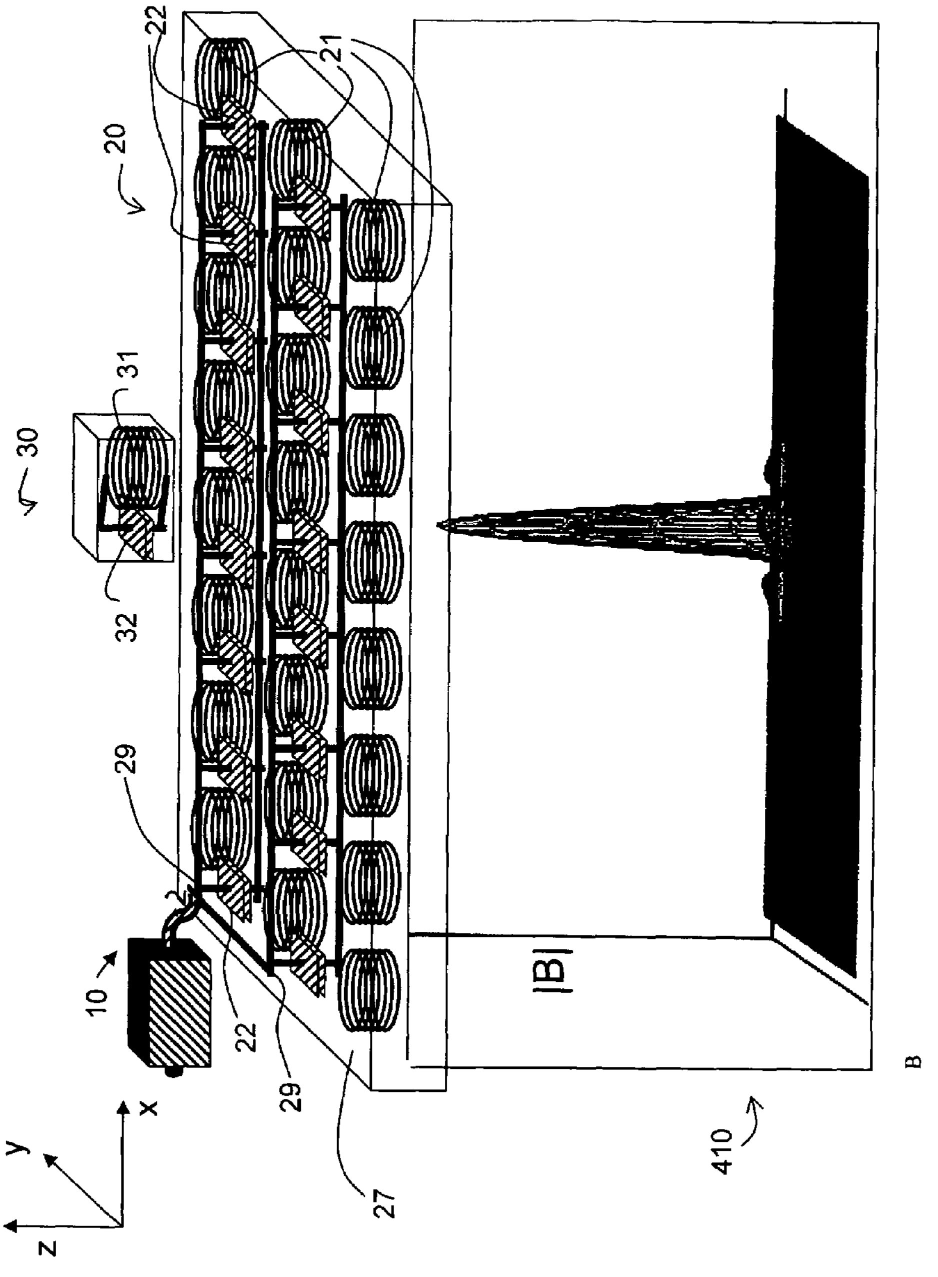
z
y
x
10
22
29
27
410
|B|
B
29
32
30
31
20
22
21

FIGURE 3

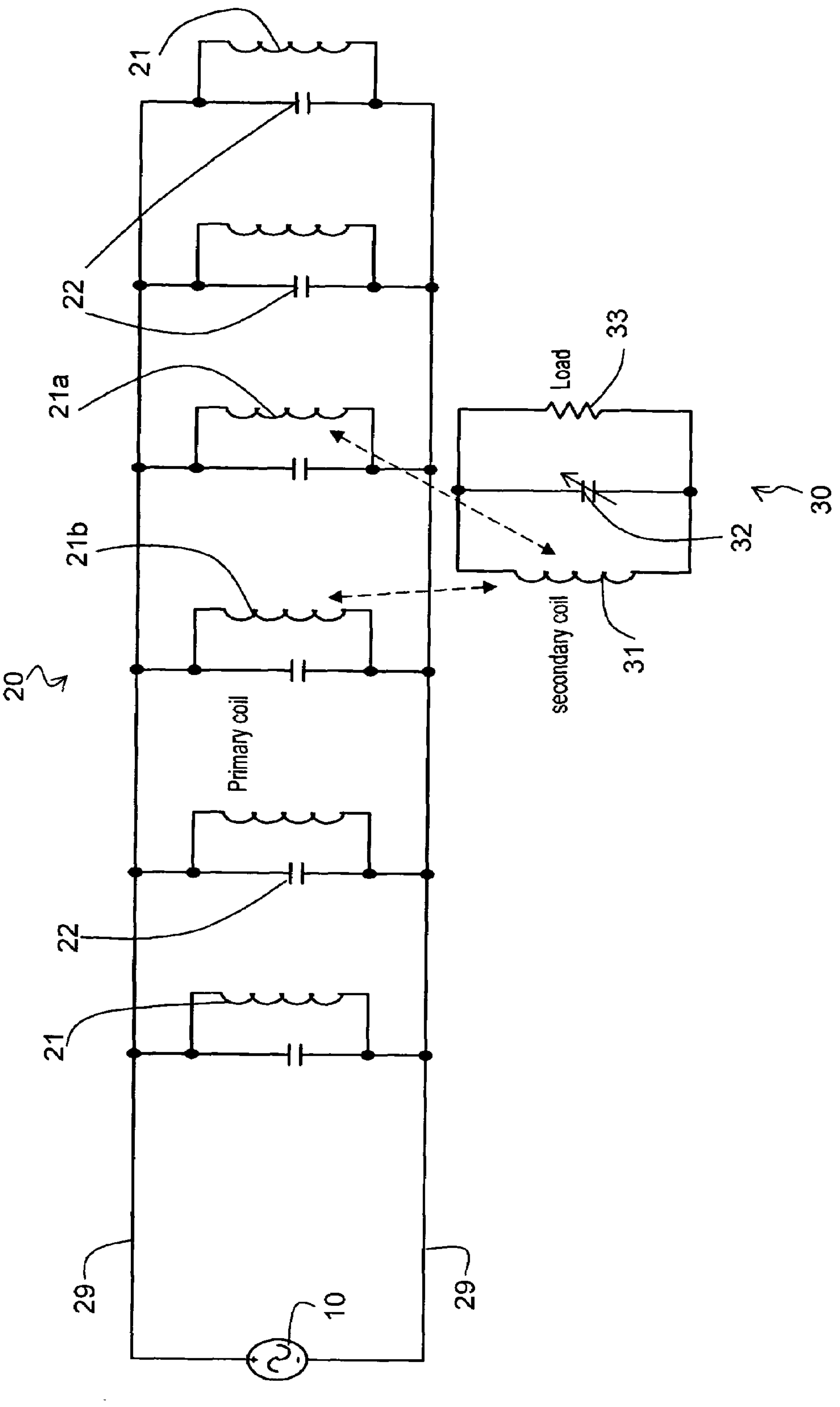
Primary coil
Load
secondary coil

FIGURE 4

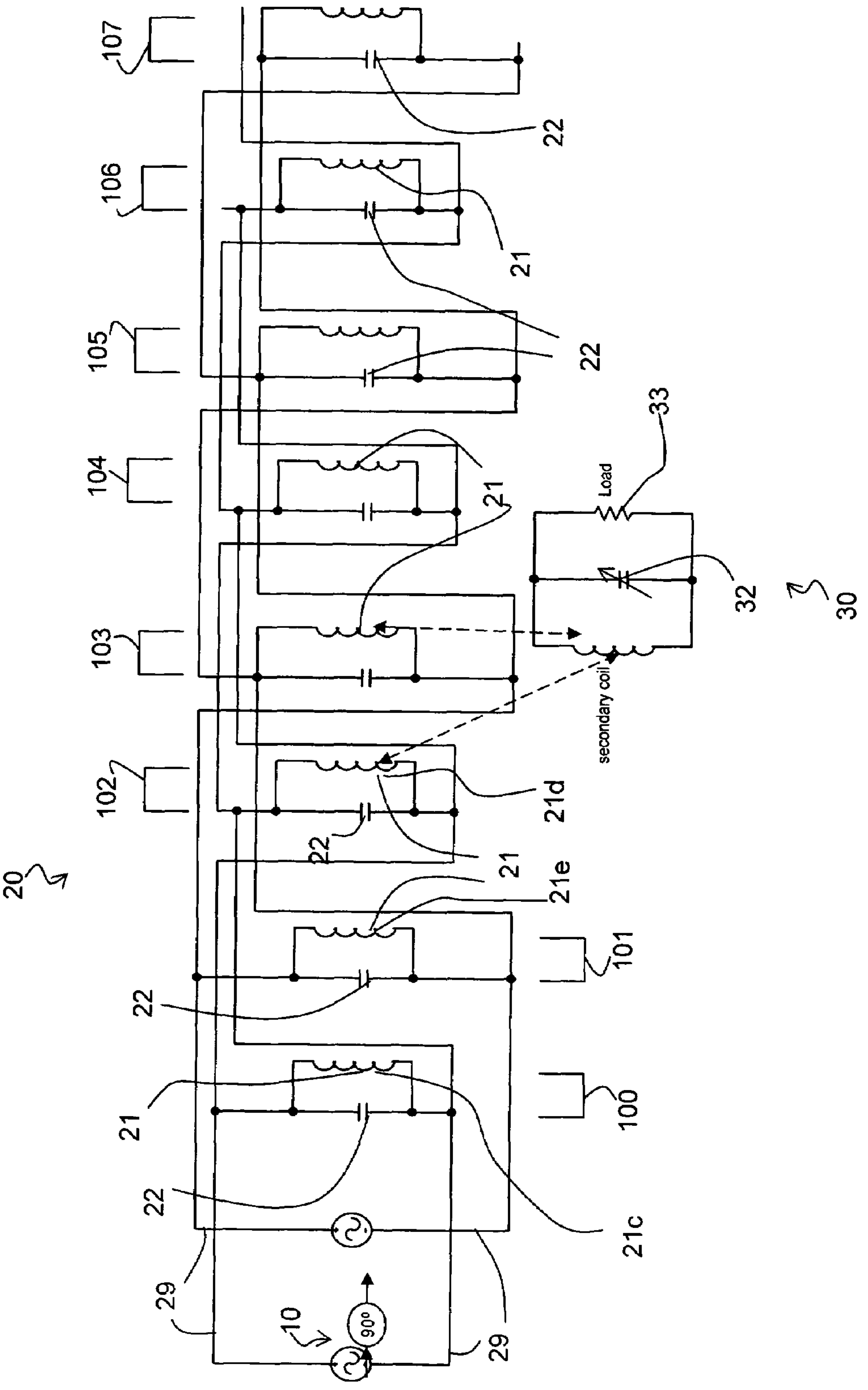
90°
secondary coil
Load

FIGURE 5

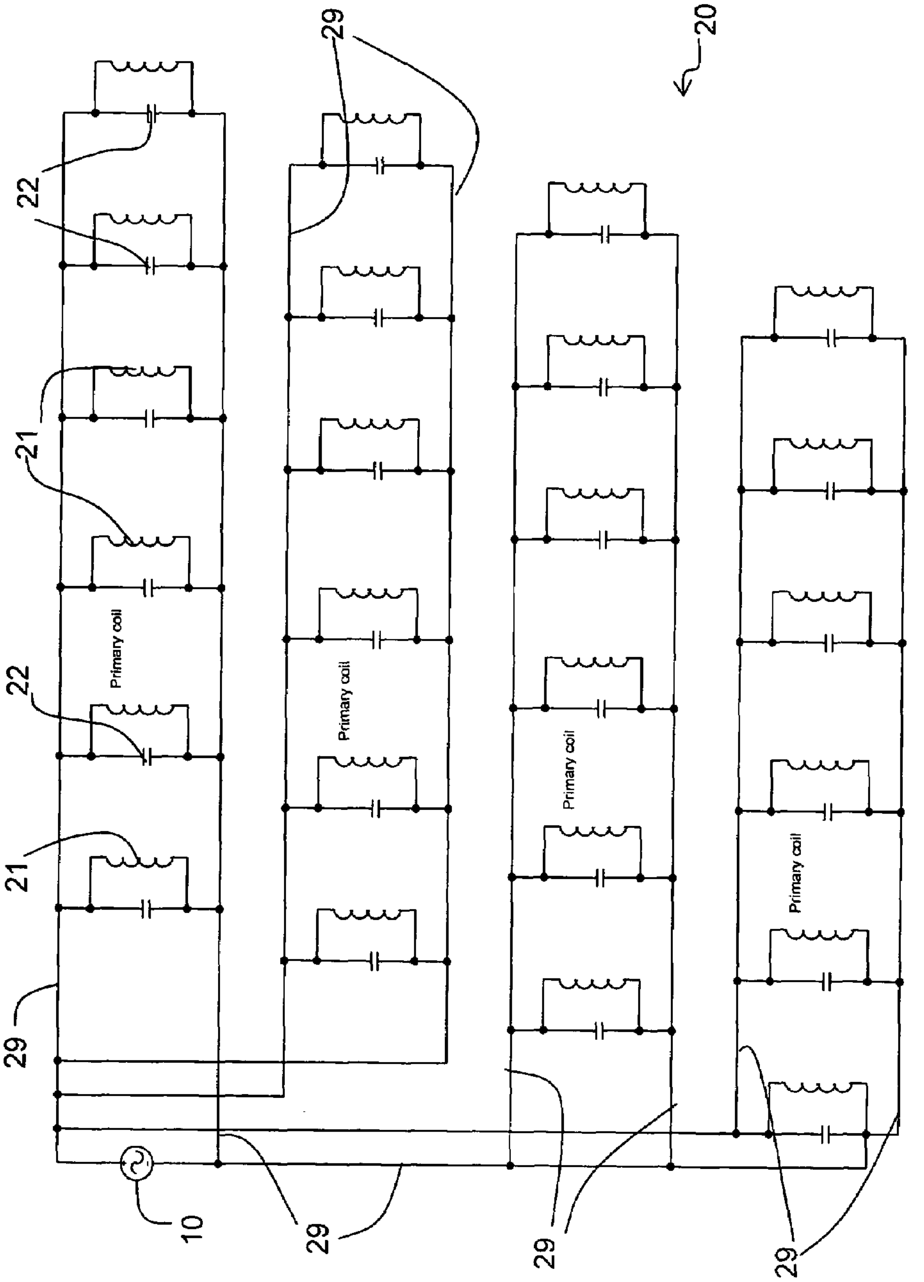
Primary coil
Primary coil
Primary coil
Primary coil

FIGURE 6

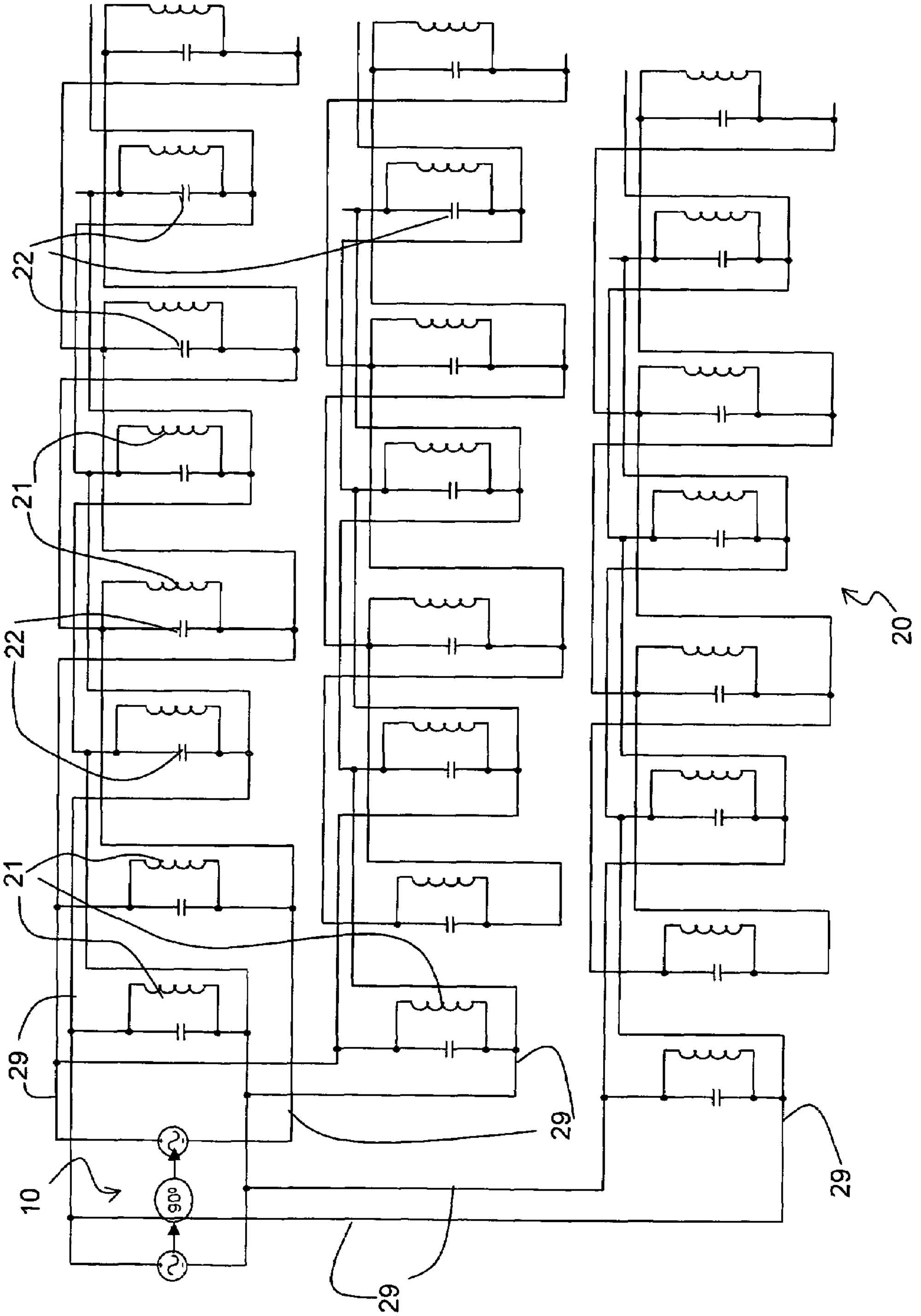
22
21
22
21
29
10
90°
29
29
29
29
20

FIGURE 7

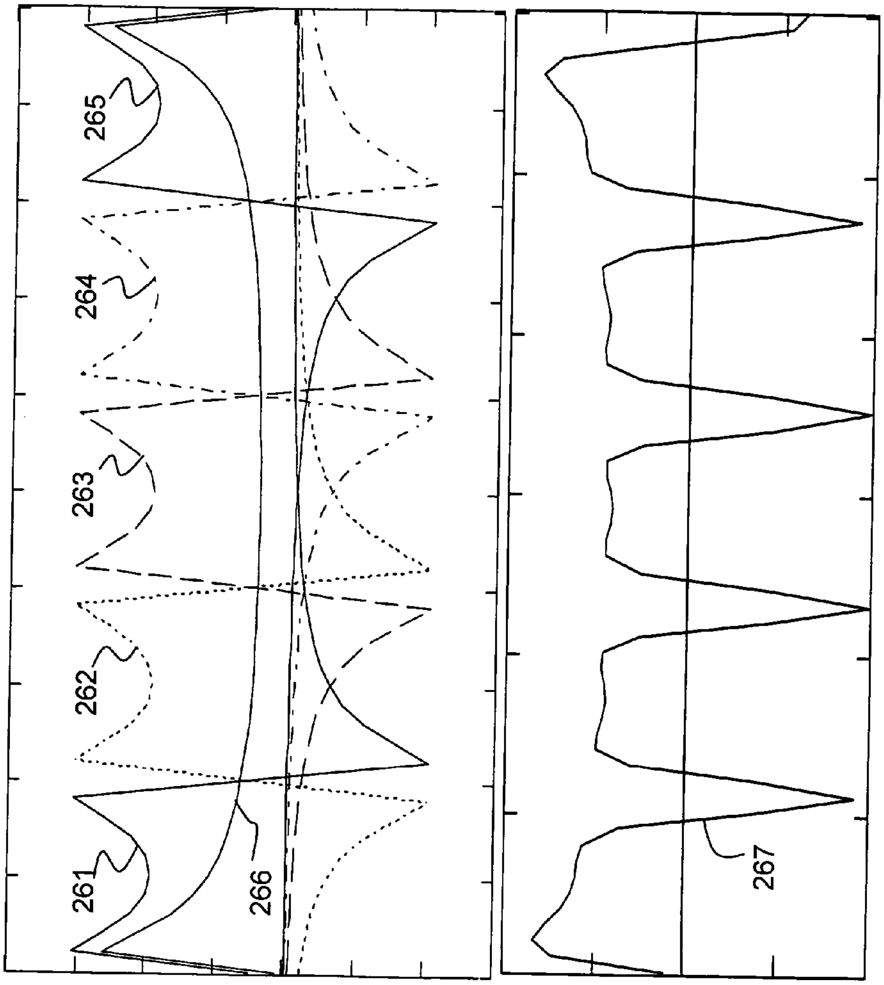

FIG. 8a

FIG. 8b

FIGURE 8

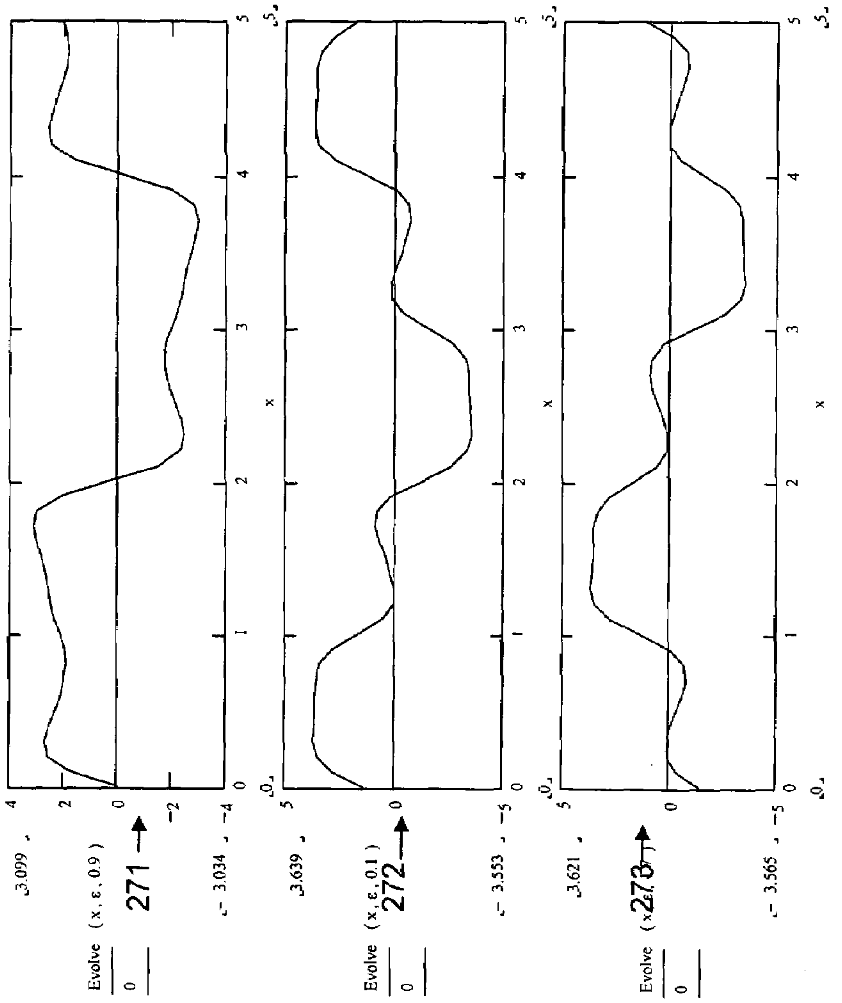
Evolve (x, ε, 0.9)
0
3.099
−3.034
271
4
2
0
−2
−4
Evolve (x, ε, 0.1)
0
3.639
−3.553
272
5
0
−5
Evolve (x, ε, ...)
0
3.621
−3.565
273
5
0
−5
0
1
2
3
4
5
x

FIGURE 9

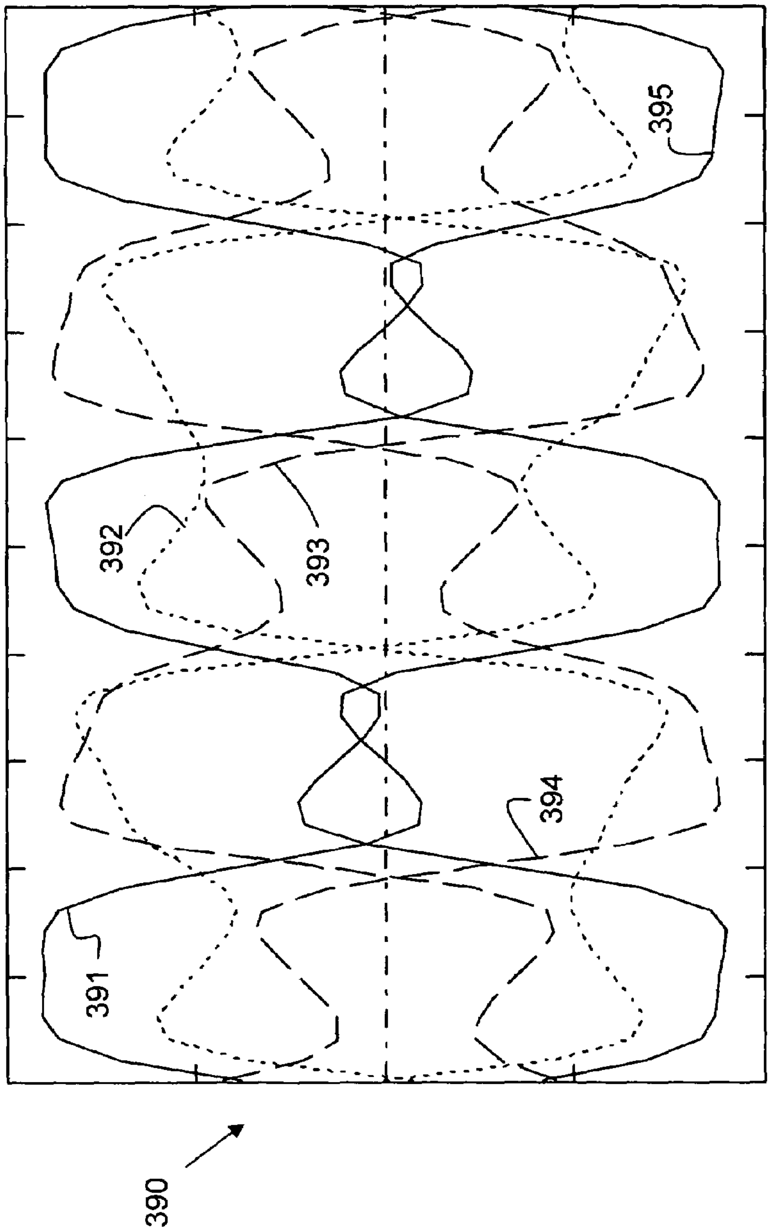

FIGURE 10

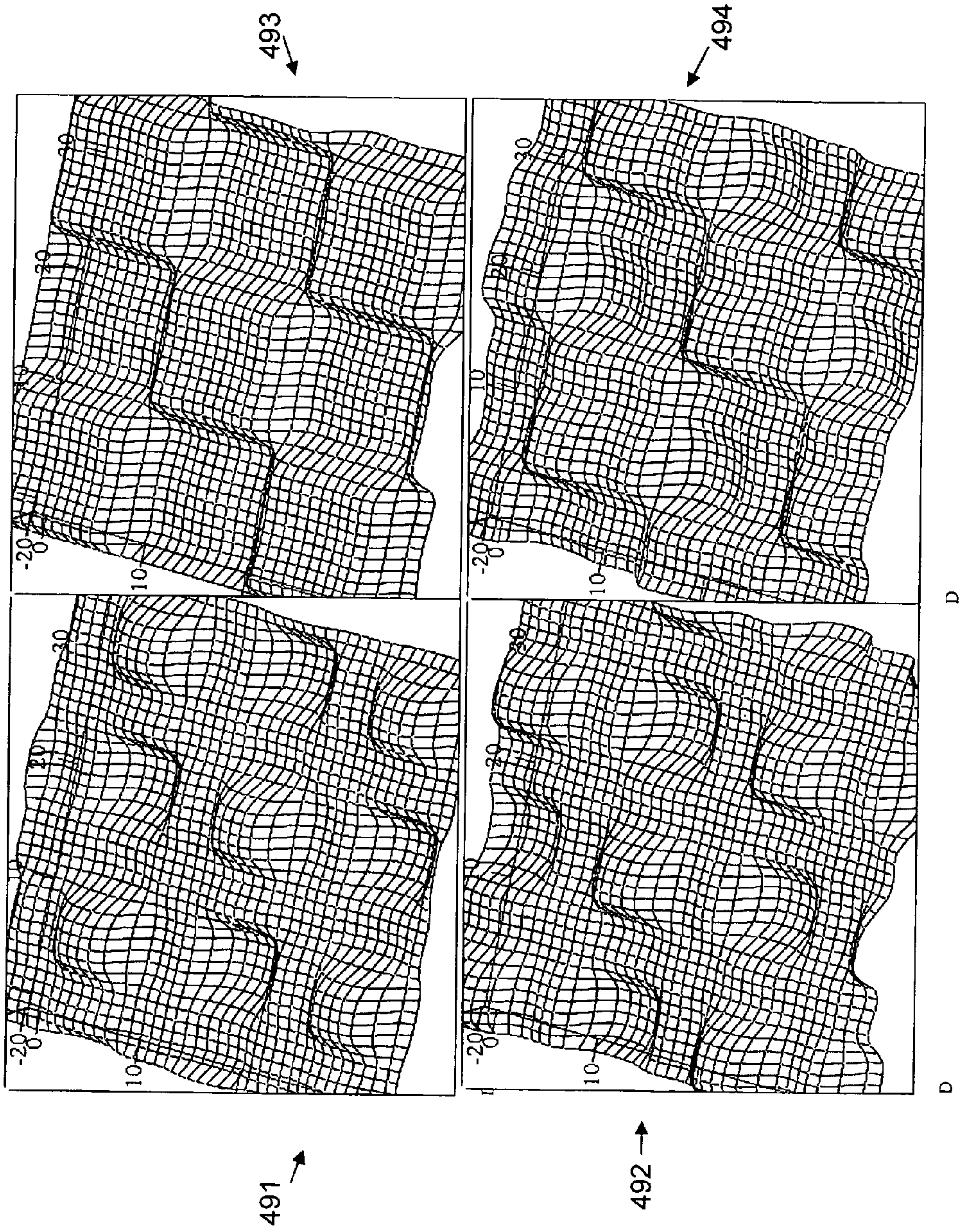

FIGURE 11

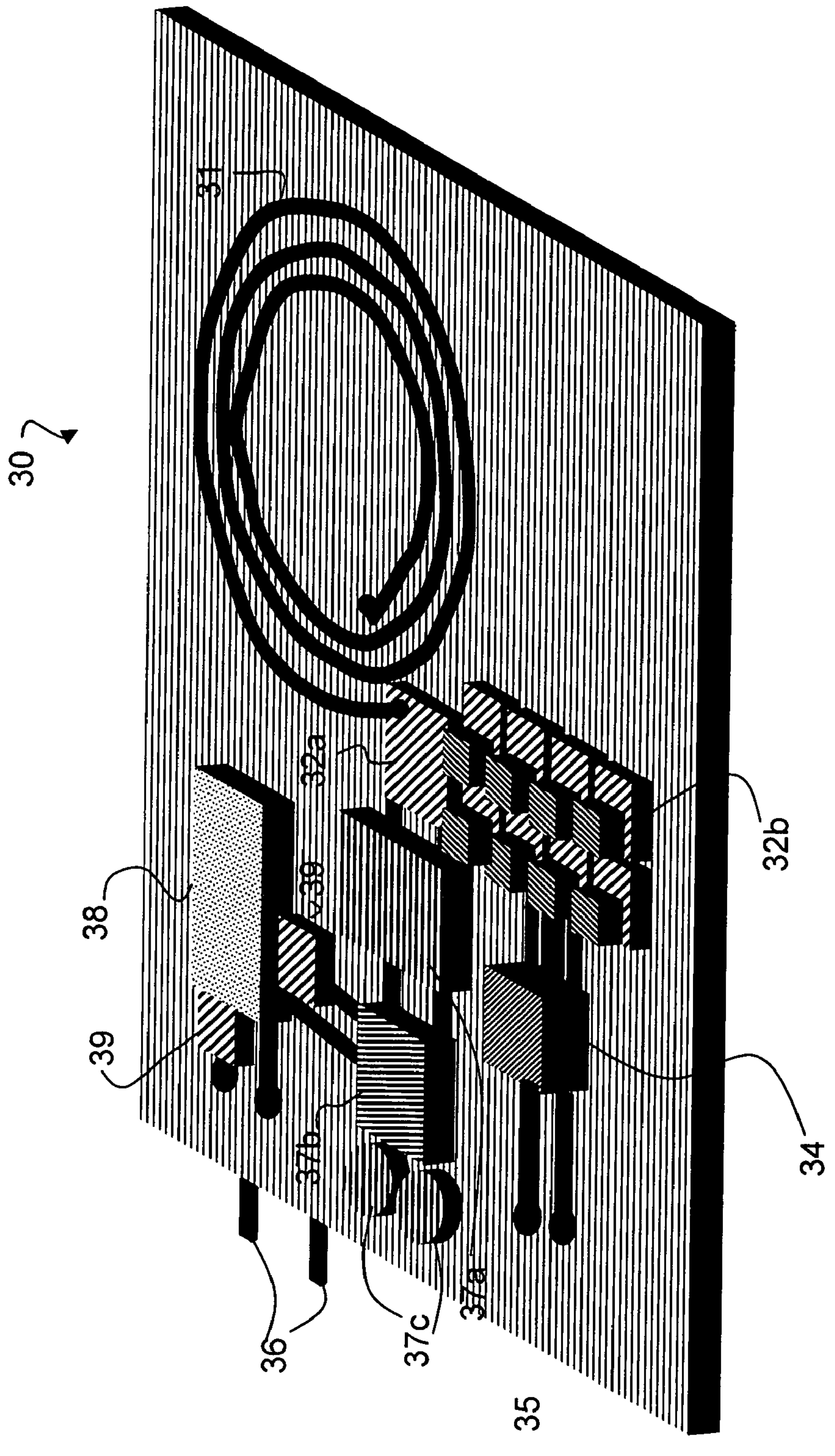
30
31
32a
32b
38
39
39
34
37b
37a
37c
36
35

FIGURE 12

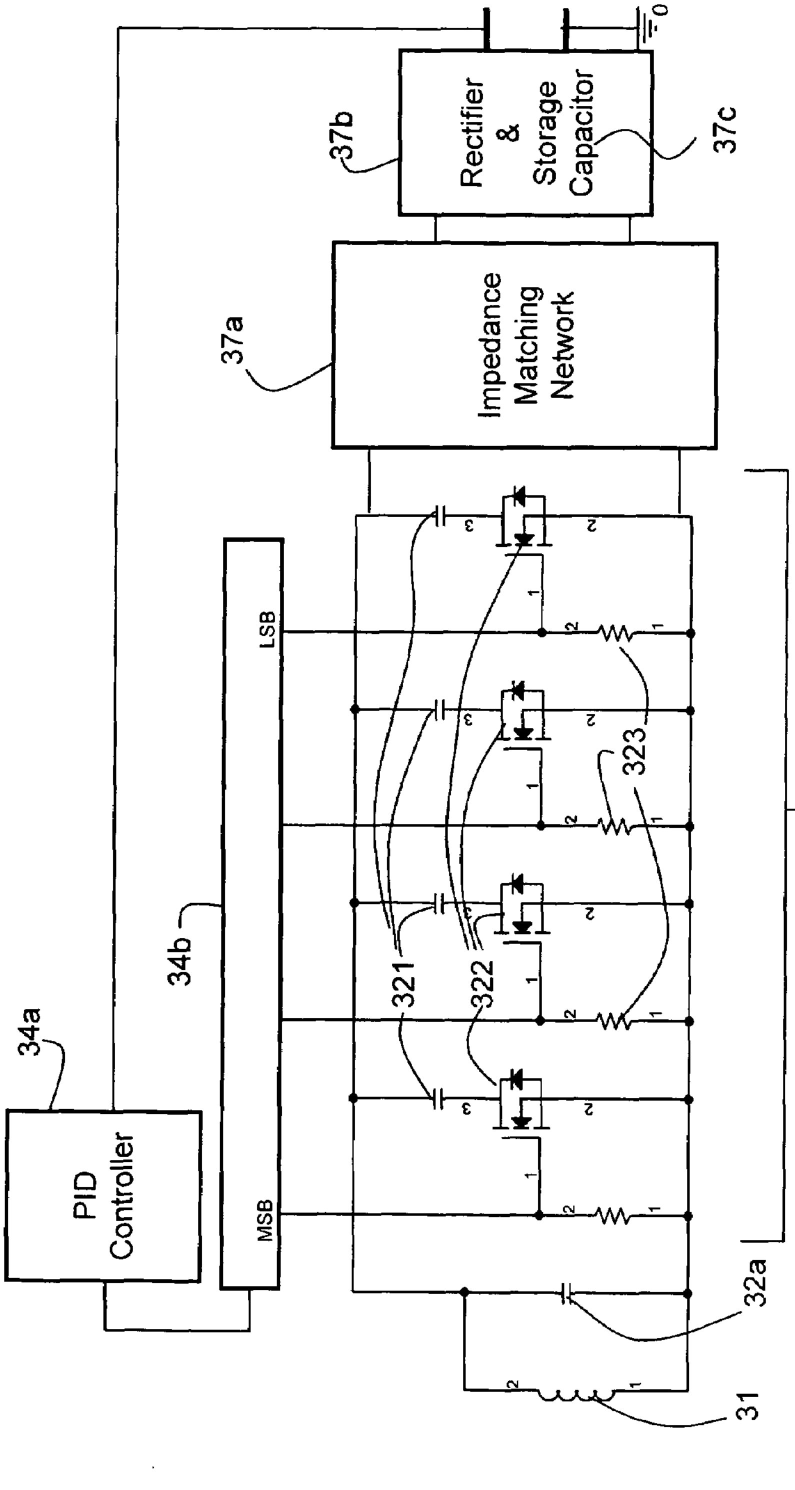
PID Controller
MSB
LSB
Impedance Matching Network
Rectifier & Storage Capacitor

FIGURE 13

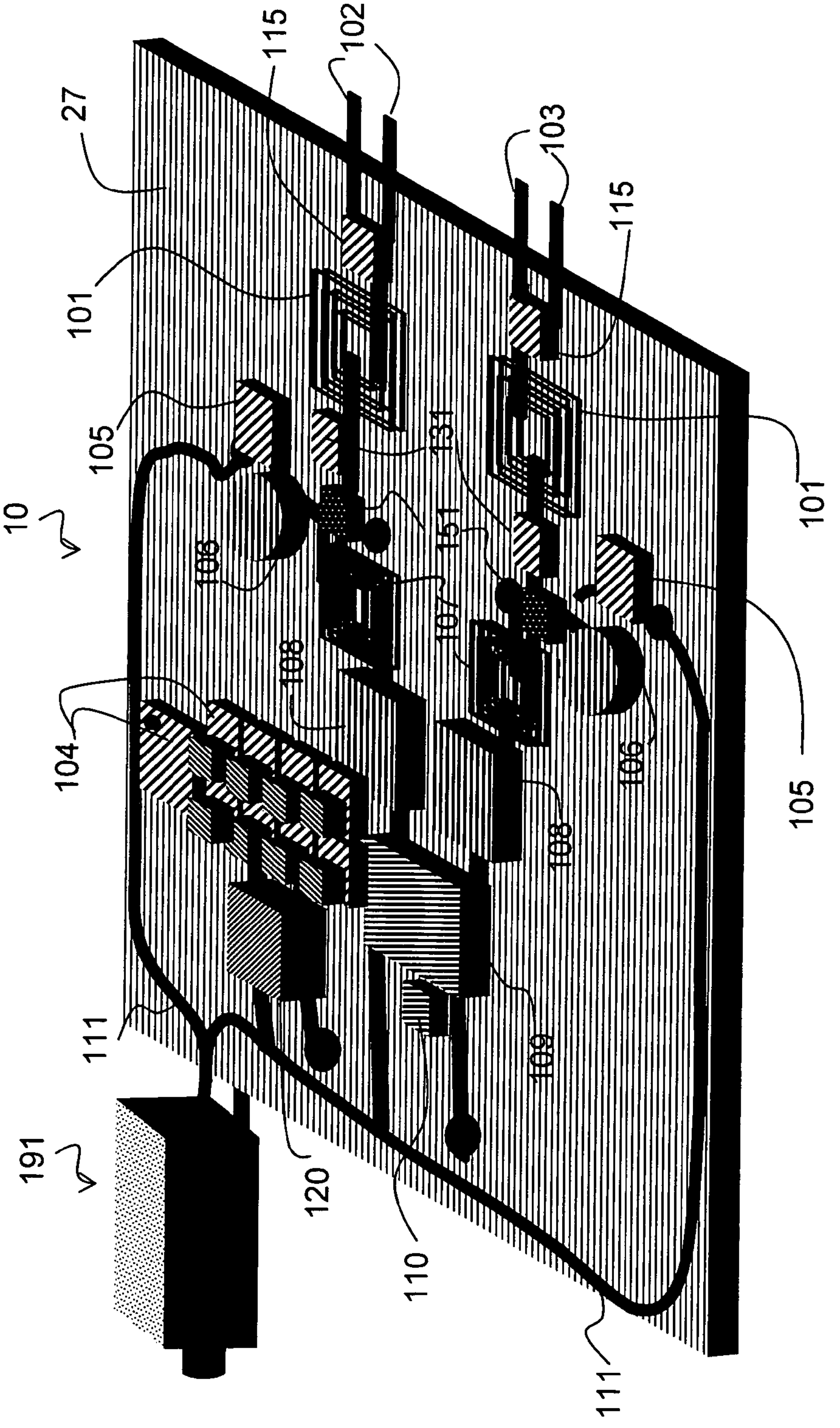

FIGURE 14

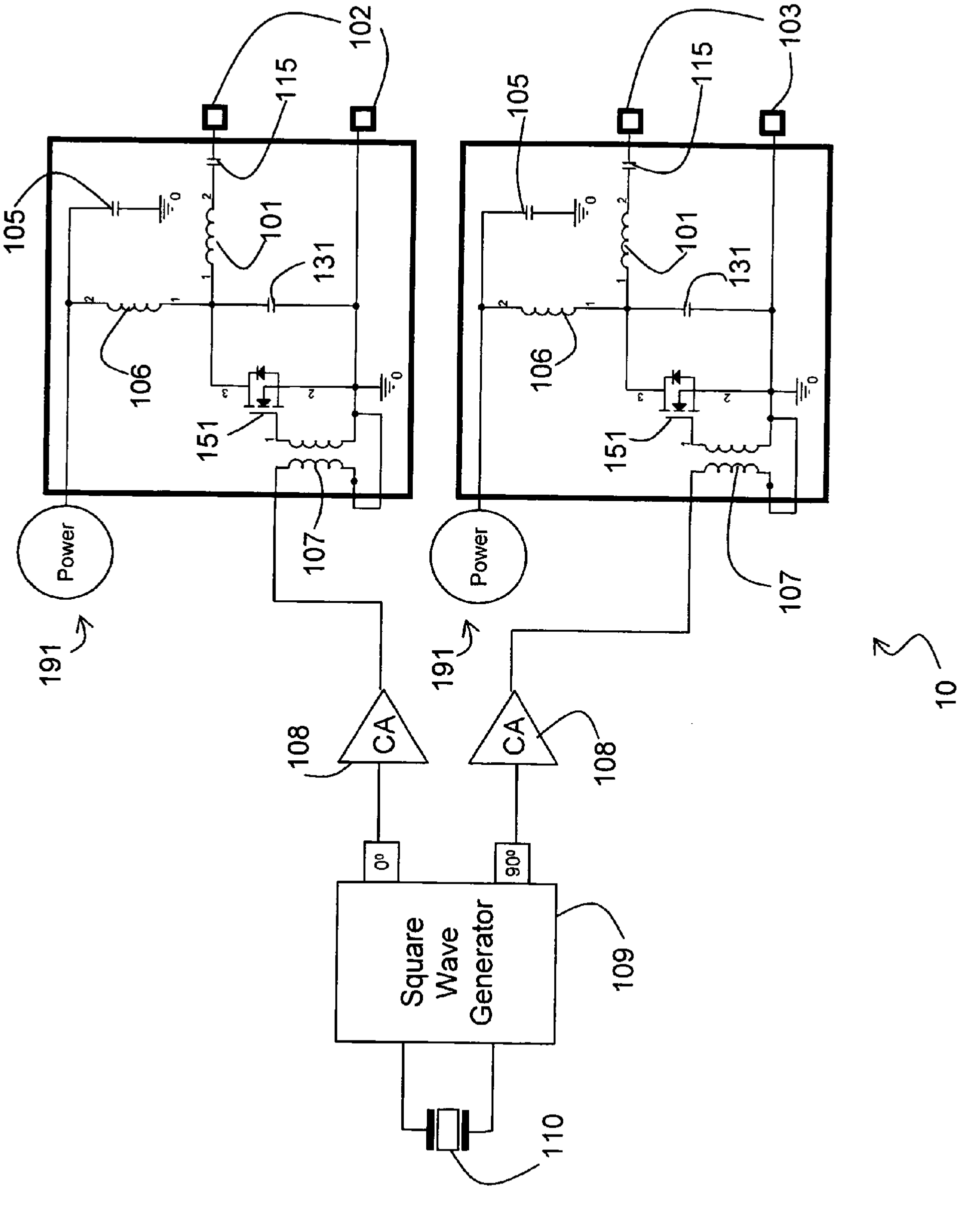
Power
CA
0°
90°
Square Wave Generator

FIGURE 15

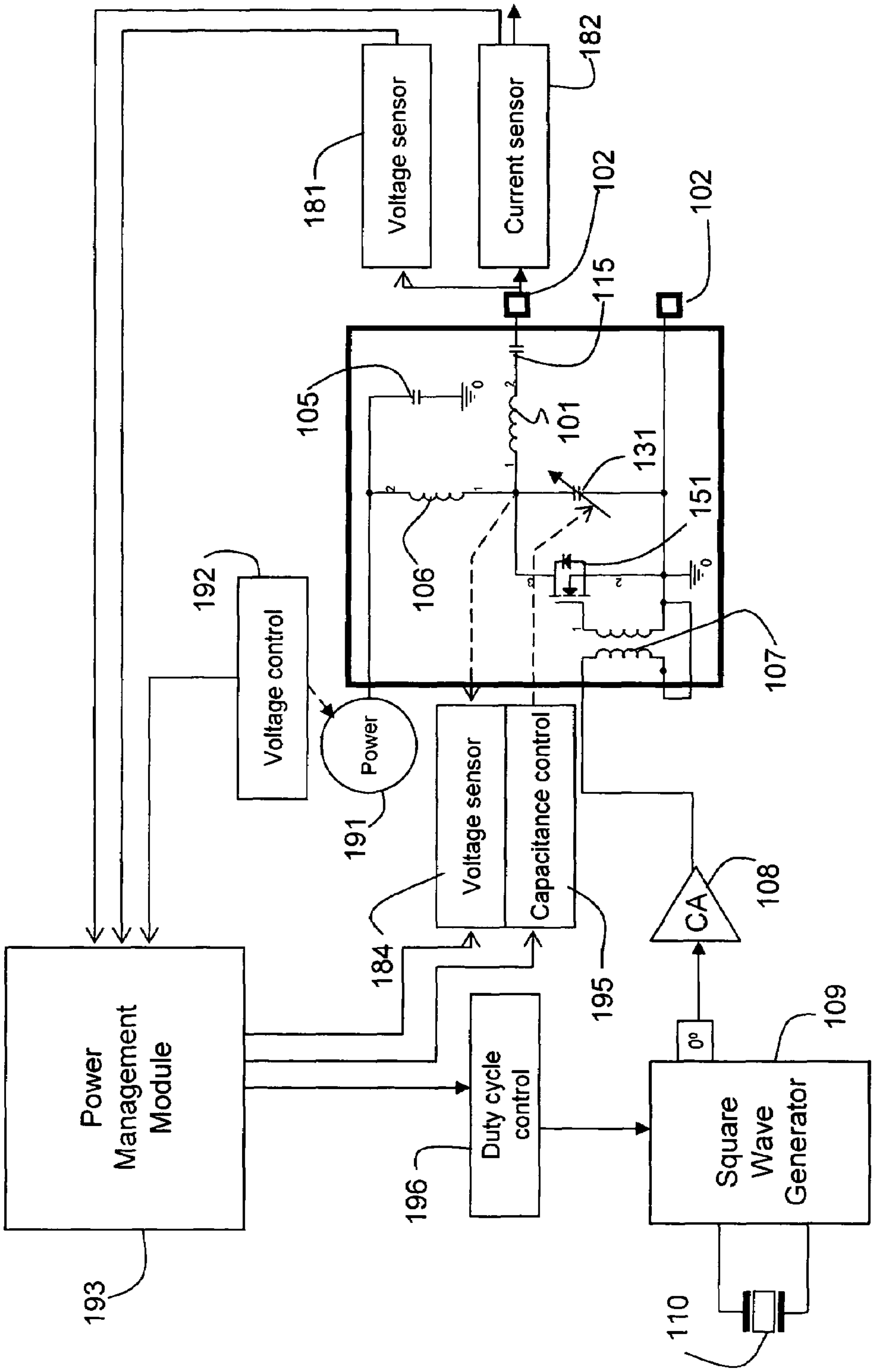
Voltage sensor
Current sensor
Voltage control
Power
Voltage sensor
Capacitance control
CA
0°
Square Wave Generator
Duty cycle control
Power Management Module

FIGURE 16

SYSTEM AND METHOD FOR SELECTIVE TRANSFER OF RADIO FREQUENCY POWER

FIELD OF INVENTION

One embodiment described herein relates to a system and method for efficient and safe transfer of electrical power without needing conductive contacts.

BACKGROUND OF THE INVENTION

Contact free power transfer has been around for a long time. A case in point is the transformer which converts high voltage AC power into low voltage AC power and vice versa through a shared inductive magnetic field. The inductive magnetic filed is generated by the primary coil which induces an electromotive force that propels electrons in the secondary conducting coil with the aid of a high magnetic permeability material core. There are no direct electrical conductive contacts between the primary and the secondary coils. A modern example of a transformer is the Braun Oral B toothbrush, which relies on the external casing of the toothbrush base to guide the toothbrush to mate with the internal shape of the charger module, thus allowing the primary coil, the secondary coil, and the high permeability cores to align perfectly so that charging can take place. Other induction charging device prior arts are similarly restrictive in the way the primary unit mates with the secondary unit. For those prior art inventions, the secondary device can only be charged by a dedicated primary device designed for it and charging is only possible if the primary unit is aligned accurately to receive the secondary unit it is designed for.

The contact free power transfer concept as embodied by the transformer inherently requires almost perfect mechanical alignment. The reason for this is that any “magnetic flux leakage” between the primary coil and the secondary coil, namely, the magnetic flux that is generated by the primary coil that is not enveloped by the secondary coil, will be seen by the primary unit as a series leakage inductance. The leakage inductance is directly proportional to both the self inductance of the primary coil and the fractional loss of magnetic coupling between the primary and the secondary. Since the power transfer rate is directly proportional to the primary self inductance as well as to the magnetic coupling coefficient and the frequency of the AC source current, the self inductance of the primary coil must necessarily be high. Hence even a 1% flux leakage will result in leakage impedance that is large enough to practically choke off the primary current. The leakage flux can only be minimized by properly aligning the secondary coil with the primary coil and by using the high permeability core to concentrate and guide the inductive magnetic flux through the volume enclosed by the coils with minimum stray magnetic fields. To further minimize leakage, the air gap between the primary magnetic core and the secondary magnetic core must be kept as small as possible. This explains why an improperly seated inductive toothbrush oftentimes will fail to get charged. The requirement for precise alignment presents a major obstacle to the popularity of such devices.

To surmount the difficulty of transformer-based inductive power transfer, one or more capacitive load may be introduced in series with the primary circuit in order to compensate for the large inductive leakage impedance. This is similar to “power factor correction” for power AC circuits. A switched capacitor bank can be used to adjust the capacitive load to accurately cancel out the inductive load caused by the leakage flux. A primary circuit in which the series capacitive load and series inductive load precisely cancels is called a series resonant circuit. Power transfer that relies on series resonance is appropriately called “resonant inductive power transfer.” Such transfer is no longer subjected to the stringent leakage free coupling requirement of the transformer power transfer method.

One prime example of resonant power transfer is the passive RFID tag. Passive RFID tag has no internal power supply. The minute RF current induced in the antenna of the tag by the incoming RF signal provides the power for the tag. The power is rectified and stored in a capacitor. When the voltage of the capacitor reached a threshold value, the capacitor activates the digital memory chip that is given a unique electronic product code. The RFID code is used to modulate the antenna to reflect the continuous RF signal transmitted by the RFID “interrogator” by varying the reflectivity of the antenna. The interrogator detects the changes in the reflected power and demodulates and decodes the received RFID signal. The passive RFID represents a diametrically opposite example to an AC transformer in that the leakage flux is so large that the coupling coefficient is for all practical purpose zero, i.e., very little flux from the primary interrogator coil is linked to the secondary passive RFID antenna coil. Hence the leakage coefficient is virtually 1. In such cases, there is no need of a switched capacitor bank. A single fixed capacitor tuned to the series resonance of the primary circuit is sufficient. Because the coupling coefficient is so small, very little power actually gets transferred between the primary unit (the interrogator) and the secondary unit (passive RFID tag), hence the energy conversion efficiency is extremely poor. However, the main purpose of RFID technology is not to provide power transfer, but to enable the interrogator to briefly power up the passive tag’s power unit with just enough energy to transmit back the product code information. The fact that there is power transfer between the primary unit and the secondary unit is just a means to achieve that goal without having to put an energy storage device inside the tag.

Another example of resonant power transfer is the Wacom’s wireless tablet technology. A Wacom pen or mouse can be used with their tablets as input devices without the need of any internal power supply such as batteries or supercapacitors. The tablet acts as a primary unit with a multitude of primary coil arranged in a matrix and emits a low power magnetic field whose intensity decreases rapidly away from the tablet. The pen or mouse comprises an antenna and a transponder unit that receives the incoming RF power, rectifies it, and retransmits a RF signal at a different frequency. The retransmitted signal is detected by two or more primary coils that are close to the antenna of the secondary unit. The received signals are used to triangulate the precise location of the secondary antenna. Wacom’s technology can be considered as a variation of the RFID technology in its use of a frequency converting transponder. Because of the closeness of the primary coils to the secondary antenna, very little power is required to transmit a beacon signal. Hence, as in RFID, the energy conversion efficiency does not play a role here, and the contactless power transfer is just a means to achieve the goal of precisely locating the battery-less pointing device.

US 20030210106A1 to Lily Ka Lai Cheng et al. discloses a system and method for transferring power without the need for direct electrical conductive contacts between a primary unit and one or more secondary units. A salient aspect of that patent is that the inductive magnetic field lines generated by the primary unit are substantially parallel to the plane of the surface within the laminar active surface of the primary unit, which minimizes the intensity and size of the magnetic field generated to reduce electromagnetic emission. Another salient aspect of that invention is the convenience of being able to allow secondary devices to be placed anywhere within the active vicinity thereby eliminating the need for plugging-in or placing secondary devices accurately relative to an adaptor or charger. A still another aspect of the prior art is the ability of the primary unit to supply power to a number of secondary different devices with different power requirements simultaneously. These are achieved by a unique design of two large base coils that generate magnetic fields which are substantially parallel to the plane of the laminar base surface and are substantially orthogonal to each other in space and time. The localization of the magnetic field is further aided by the use of high magnetic permeability material. This overcomes one of the main problems associated with inductive charging: the generation of large magnetic field profile. Since the intensity of electromagnetic emission is governed by regulatory limits, any device that is capable of generating a large inductive magnetic field profile runs the risk of violating such limits. In addition, numerous objects can be adversely affected by the presence of a large magnetic field, causing them to either be heated or lose precious stored data.

Despite these advances, this prior art approach suffers from two serious limitations: First, although the magnetic field generated by such approach has a low profile, it does not provide a method to localize the magnetic field around the secondary devices. Thus, if a metallic object is placed on or near the active area of the primary device, an eddy current will be induced, causing the metal to heat up, with a like effect on the primary unit, which sees the metallic object as an equivalent short, which may lead to irreparable damage to the primary unit. Second, the height of the magnetic field profile for the primary unit is directly proportional to the linear dimensions of the primary unit because of the similarity principle. Hence the concept can't easily be scaled up to large primary devices without exceeding regulatory limits. To date, the assignee of that patent, Splashpower limited, has demoed only small, mouse-mat-sized pad that can support up to two small power devices such as PDAs and cell phones simultaneously. Splashpower's devices have also suffered from low energy conversion efficiency.

Another approach to avoid the generation of large magnetic fields is to use an array of coils whereby only a few of them are activated as need arises. This was suggested by a paper published in the Journal of the Magnetics Society of Japan entitled "Coil Shape in a Desk-type Contactless Power Station System" (Nov. 29, 2001). In one embodiment of the concept, multiple position sensors sense the presence of a secondary device placed on the primary unit and relay that information to a central controlling unit. The control unit then sends power to the appropriate coils to energize the secondary coil in a localized fashion. The degree of localization of the magnetic field is determined by the number of primary coils. Since each coil requires a dedicated high frequency switching unit to activate or deactivate, and the central controller unit needs to have an input pin and an output pin for every coil, the complexity increases drastically as the number of primary coils increases.

The prior art solutions invariably use low induction frequencies in the audio frequency range (1 KHz to 50 KHz). The advantages of such solutions include the ready availability of power generation devices in this frequency range as well as the existence of very high magnetic permeability materials which can increase the induced magnetic flux drastically by offering a low reluctance path for the induction magnetic field, thereby increasing the power transfer. The drawbacks of the prior art approaches are: First, the power generation devices are typically bulky because of the need to use large inductors and capacitors at such low frequencies as well as the relatively large current requirement. Second, the high magnetic permeability materials typically have rather large loss tangent, and have nonlinear and hysteretic behaviors at large magnetic flux intensities. Further, the strong reliance of the prior art approaches on the use of high permeability materials in order to reduce the bulkiness of both the primary and the secondary device means that the only major high magnetic reluctance circuit resides within the air gap between the primary and the secondary devices. The consequence is that the power transfer decreases rapidly as a function of the air gap, rendering the prior art technologies no better than the prior art solution based on direct electrical contact in their ability to tolerate vertical separation. Fourth, in order to enable the power transfer through the air gap, the inductive magnetic flux through the air gap must be sufficiently large owing to the relatively low induction frequency. This is because the power transfer is proportional to the product of the induced voltage and the induced current, and the induced voltage is in turns proportional to the rate of change of magnetic flux linked. In other words, the induced voltage is proportional to the product of the driving frequency and the linked magnetic flux.

Within the air gap, the relative magnetic permeability is 1; hence the magnetic field strength within the air gap cannot be reduced by the presence of high permeability material elsewhere. Since the lower the driving frequency, the larger the magnetic field required for a given power transfer rate, this means that low frequency inductive power transfer necessarily entails a severe size/power tradeoff, namely, for a given size of the induction coil, there is a power limit. For audio frequency inductive methods, the power limit for an inductive coil that can fit into a typical cell phone is limited but sufficient to charge a cell phone. The situation gets worse for larger devices such as a notebook computer which often requires 20-40 times more power to charge its battery but lacks space for a coil large enough for that purpose.

The inability of the majority of prior arts to localize the inductive field presents thorny interference and safety issues, as well as energy conversion efficiency issues. The danger that is inherent in having stray magnetic flux that can cause metallic objects in its vicinity to heat up, no matter how slowly is a major safety hazard. An user wearing a ring on his or her finger, or having a metallic implant, is liable to have induced eddy current in and around his/her body. The electromagnetic interference caused by such stray field can also severely hamper the proper operation of a sensitive electronic device. Lastly, the presence of nonlocalized field which does not couple to client devices drastically reduce the energy conversion efficiency since the Q factor of the primary coil is not infinity in practice, hence the portion of the current that is responsible for the generation of such stray flux will always have a resistive component. The larger the unlinked flux, the higher the dissipation loss. Worse, the magnetic flux does not diminish even in the absence of any client devices in the active vicinity; hence the standby power consumption represents a 100% power loss.

Even though each of the aforementioned problems can be solved in theory, the solutions invariably increase system complexity and drive up cost as well as making the system less reliable and less user-friendly. The hazard associated with the presence of metallic objects in the vicinity of the primary device, for example, can be solved through the use of a proximity metal detector. The energy conversion efficiency can theoretically be improved by using conducting materials with very low resistivity such as room temperature superconductors, etc. The EMI interference issues can exemplarily be addressed by either limiting the size of the primary surface, or by switching micro-coils, as was suggested in the aforementioned Japanese paper (Nov. 23, 2001). The no-load power loss can be addressed by automatically turning off the power to the primary coils when no devices have been detected, and is only turned back on when at least one device has been detected. The detection of an eligible device can be performed by embedding a RFID in a device so that a RFID reader in the primary device can read the RFID tag to identify compatible client devices on the primary surface. The absence of any RFID signature will allow the primary unit to go to a standby mode with no electromagnetic emission.

Although a metal detector will make the prior art system safer, it significantly limits how the prior art system can be used. For example, a nearby metallic object will cause the primary device to go to a standby mode immediately, terminating the power transfer; hence the sensitivity of the metal detector must be tuned carefully to ensure that its envelope more or less coincides with that of the active region of the primary device. This way, a distant metallic object won't trigger an unnecessary shutdown. Even then, any small metallic object that accidentally falls on the primary surface will have to be removed, which is clearly a big annoyance. Worst, if an object such as steel wool whose metallic content is insufficient to trigger the metal detection happens to be on the primary surface, it may still be able to couple to the inductive magnetic field to generate a small Eddy current which causes it to slowly get hot enough to create a fire hazard. Similarly, if the sensitivity and range of the RFID reader is not properly tuned, it may either be unable to detect an eligible client device and therefore fails to turn on the primary coils, or there may be a client device that is too far to be charged by the primary device and yet close enough to be identified by the RFID reader to prevent the primary device from going to the standby mode.

SUMMARY OF THE INVENTION

In one embodiment, a system is provided for the inductive transfer of electric power between a primary surface and a multitude of secondary devices in such a way that the power transfer is localized to the vicinities of individual device coils. In one embodiment, such power transfer occurs without the need of complex high frequency power switching network to turn the individual primary coils on or off. In one embodiment, a modular way to provision power is provided.

In one embodiment, the inductive transfer of electric power between a primary surface and a multitude of secondary devices is unaffected by the presence of conductive or metallic objects either in direct contacts or in the vicinity of the primary inductive field to ensure that such power transfer can take place safely and uninterrupted.

In one embodiment, the inductive transfer of electric power between a primary surface and a multitude of secondary devices does not produce significant electromagnetic interference to sensitive electronic devices in the vicinity of the primary surface.

In one embodiment, the inductive transfer of electric power between a primary surface and a multitude of secondary devices does not require the use of expensive high magnetic permeability materials.

In one embodiment, a uniform inductive coupling between the primary device and the multitude of secondary devices is maintained in a uniform manner within the active region of the primary device.

In one embodiment, electric power is to be transferred in a contactless fashion comprising a primary unit and a multitude of secondary units, wherein the primary unit is a thin device with a laminar surface and a multitude of substantially flat induction coils embedded just underneath the laminar surface interconnected in parallel by a primary feed network, the multitude of conductive coils being accompanied by a multitude of counterbalance capacitors to reduce or substantially cancel out the primary current that flows in the primary feed network in areas where secondary units are absent.

In one embodiment, a multitude of counterbalance capacitors has a one-to-one relationship with the multitude of primary coils, with each primary coil being mated to the closest counterbalance capacitor. In this way, whenever a secondary unit is placed or held in close proximity to said primary coil, the counterbalance relationship between the primary coil and the matching capacitor is broken, leading to an increase in primary current to both the capacitor and the primary coil to effect a local transfer of electric power from the primary coil to the secondary unit.

In one embodiment, the plurality of primary coils is separated into four interweaving sets of coils which together form a 2-dimensional periodic pattern. The coils are driven by a 4-phase quadrature drive to generate a diagonally traveling induction field pattern. Alternatively, the plurality of coils can also be driven by other multi-phase driver such as a 6-phase driver to better simulate a traveling sine wave for the induction field. When a secondary unit is placed on or near the primary surface, the secondary unit will experience a sinusoidal inductive magnetic field irrespective of the location of the secondary unit relative to that of the primary unit. The closer the primary induction field resembles a traveling sine wave field, the more nearly constant the inductive electromotive forces which the secondary unit sees as a function of its position, and hence the more uniform is the power transfer. Note that a 4-phase driver can be constructed from just two mutually orthogonal phases, and a 6-phase driver can be built out of two phases that are 120° apart.

In one embodiment, there is provided a primary unit for transferring power wirelessly, wherein the primary unit includes at least one bank of switched capacitors that is adapted to adjusting the source impedance of the RF drive to match the impedance of the effective load of one or more secondary units that are coupled to the primary unit for the purpose of minimizing the reflected RF power, and thereby minimizing the heat generation on the RF power supply side when there is a load impedance-source impedance mismatch. Alternatively, the switched capacitor bank can be replaced by any other means of adjusting the capacitance value of a high voltage network electronically, such as a MEMS capacitor which can vary its capacitance by electrostatically varying the air gap between adjacent metallic membranes under analog or digital control.

In one embodiment, there is provided a primary unit for transferring power wirelessly, wherein the primary unit includes a load monitor which monitors the effective primary load impedance of secondary units to provide the needed feedback to the aforementioned source impedance matching means, as well as to act as a load sensor which provides the signal to switch the primary unit into the sleep mode when no secondary units are loaded.

In one embodiment, there is provided a primary unit for transferring power wirelessly, wherein the primary unit comprises a sleep mode of operation triggered by the absence of measured load and can return to normal operation by periodically waking up to operating mode to test the existence of secondary units. Alternatively the secondary units can generate keep-alive signals periodically by modulating the input impedance or by direct RF signal generation to either wake up the primary unit from sleep mode, or to prevent the primary unit from going into the sleep mode.

In one embodiment, there is provided a secondary unit for receiving power wirelessly, wherein the secondary unit includes a coil for coupling to the induction field of the primary unit. The coil is substantially flat and can comprise a single turn or multiple turns depending on the load resistance of the secondary unit, the voltage and/or current requirements.

In one embodiment, there is provided a secondary unit for receiving power wirelessly, wherein the secondary unit includes a variable capacitor means which provides an electronically adjustable capacitance for the purpose of matching the resonant frequency of the secondary receiving circuit with the drive frequency of the primary unit. The variable capacitor means can be a switched capacitor bank under the control of an analog or digital controller, or could be MEMS based variable capacitors.

In one embodiment, there is provided a secondary unit for receiving power wirelessly, wherein the secondary unit includes a fixed or programmable controller for regulating the received RF voltage through resonant frequency tuning by controlling the variable capacitor to tune in or out of the primary RF drive frequency.

In one embodiment, there is provided a secondary unit for receiving power wirelessly, wherein the secondary unit includes a rectifying circuit to convert the RF power into DC power to provide the power for the device served by the secondary unit. The rectifying circuit can be a simple full bridge passive rectifying circuit, or it could be an active synchronous rectifying circuit with reduced forward voltage drop to improve rectification efficiency.

In one embodiment, there is provided a secondary unit for receiving power wirelessly, wherein the secondary unit includes a signaling means for sending keep-alive signals periodically to the primary unit to prevent the primary unit from going to a sleep mode and to wake up the primary unit from a sleep mode.

In one embodiment, there is provided a secondary unit for receiving power wirelessly, wherein the secondary unit includes a proportional, integral, differential (PID) controller which takes the difference between the set voltage and the rectified DC voltage as the error signal and outputs a control signal or control vector to adjust the capacitance value to tune the resonant circuit so as to regulate the rectified voltage. This allows the secondary unit to provide the correct voltage or other power requirements to the device that is powered by the secondary unit.

In one embodiment, there is provided a secondary unit for receiving power wirelessly, wherein the secondary unit includes a transformer which transforms the RF voltage to a voltage more suitable for rectification to the desired output voltage for powering the device associated with the secondary unit.

In one embodiment, there is provided a secondary unit for receiving power wirelessly, wherein the secondary unit includes a RF shield surrounding the RF section of the secondary unit which is made of thin metallic film for the purpose of blocking the RF induction field to prevent it from interfering with the proper functions of the device associated with the secondary unit as well as minimizing parasitic coupling between the secondary coils and the metallic or electronic components of the associated device which would otherwise alter the resonant behaviors of the secondary RF circuit.

In one embodiment, there is provided a secondary unit for receiving power wirelessly, wherein the secondary unit includes a frequency translation and retransmission means which shifts the received RF frequency to a different frequency and retransmits it after it has been modulated with data. The primary unit can receive the retransmitted signal, use a Bandpass filter to filter out unwanted signals and demodulate and decode the RF signal to recover the data.

In one embodiment, there is provided a secondary unit for receiving power wirelessly, wherein the secondary unit includes a voltage multiplication means which can increase the received RF voltage to a value that is sufficient to power the PID controller to tune the resonant frequency to the primary drive frequency even though the default frequency of the secondary unit is too far from the pre-set value to allow strong enough induction voltage to power the PID controller. The voltage multiplication means jacks up the voltage so that, after rectification, the voltage is sufficient to power the PID controller. This increases the system's tolerance to detuning of the secondary unit when the secondary unit is not connected to an energy storage means such as a rechargeable battery.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is an isometric view, of a system of the subject invention shown during the process of wireless power transfer between a primary unit and a secondary unit in accordance with one embodiment;

FIG. **2** is an isometric view of a single array of inductive field coils of the primary unit together with a single coil of the secondary unit, as well as the localization of the inductive field when the secondary coil is brought close to the primary unit in accordance with one embodiment;

FIG. **3** is an isometric view of the primary unit which comprises a two-dimensional array of primary coils and attendant matching capacitors as well as the power rails for connecting the primary coils and capacitors to the primary RF power supply in accordance with one embodiment;

FIG. **4** is a schematic view showing the interaction between the primary induction coils of the primary unit and the receiving coil of the secondary unit in a simplified single primary coil array system for illustration purpose in accordance with one embodiment;

FIG. **5** is a schematic view of an alternative embodiment of the primary coil array driven by two coherent RF sources which have a 90° phase difference, showing the interaction between the primary induction coils of the primary unit and the receiving coil of the secondary unit in a simplified single primary coil array system for illustration purpose in accordance with one embodiment;

FIG. **6** is a schematic view of an embodiment of a two-dimensional primary coil array that is conceptually identical to that of FIG. **4** in accordance with one embodiment;

FIG. **7** is a schematic view of an alternative embodiment of a two-dimensional primary coil array that is conceptually identical to that of FIG. **5** in accordance with one embodiment;

FIGS. **8***a* and *b* shows the magnetic fields for a linear array of five coils and the total magnetic field when all five coils are energized in phase by currents of equal amount in accordance with one embodiment;

FIG. **9** displays a time sequence of magnetic field intensity to depict the traveling wave nature of the magnetic field excited by a long linear array of coils driven by the circuit shown in FIG. **5** in accordance with one embodiment;

FIG. **10** shows the magnetic fields excited by a long linear array of primary coils at various times in accordance with one embodiment;

FIG. **11** displays a time sequence of magnetic field intensity distribution to depict the traveling wave nature of the magnetic field excited by a large two dimensional array of coils driven by a circuit which is a two dimensional generalization of the circuit shown in FIG. **5** in accordance with one embodiment;

FIG. **12** is an isometric view of the secondary coil and circuit elements in the device adapter in accordance with one embodiment;

FIG. **13** is a more detailed schematic view of the secondary coil and circuit elements in the device adapter in accordance with one embodiment;

FIG. **14** is an isometric view of the RF generating module in the primary unit which generates two pairs of RF signals whose phases are 90 degrees apart in accordance with one embodiment;

FIG. **15** is a more detailed schematic view of the RF generating module in the primary unit which generates two sets of RF signals which are 90 degrees apart in their respective phases in accordance with one embodiment; and

FIG. **16** shows a power factor correction subsystem for the class E amplifier module depicted in FIG. **15** in accordance with one embodiment.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

In the following detailed description of the present invention, certain specific details are set forth to facilitate a thorough understanding of the principle concepts that is contained within the present invention. It will become obvious, however, to those skilled in the art that the present invention may be practiced without those specific details. In other instances, well-known methods, procedures, materials, components and circuitry have not been described in detail to avoid unnecessary obscuring aspects of the present invention. The detailed description focuses only on presenting a self-contained explanation with sufficient depth to cover certain simplified and idealized representations or embodiments of the present invention in order to elucidate the inventive concepts. These descriptions and representations are the means used by those experienced or skilled in the art to concisely and most effectively convey the substance of their work to others skilled in the art.

Specific terminology will be resorted to in describing the present invention for the sake of clarity and it is not intended for the present invention to be limited to the specific terms so chosen. Furthermore, it is to be understood that each specific term comprises all its technical equivalents that operate in a broadly similar manner to achieve a similar purpose. Reference herein to “one embodiment” or an “embodiment” means that a particular feature, structure, or characteristics described in connection with the embodiment can be included in at least one embodiment of the invention. The appearances of the phrase “in one embodiment” in various places in the specification are not necessarily all referring to the same embodiment, nor are separate or alternative embodiments mutually exclusive of other embodiments. Further, the order of process flow representing one or more embodiments of the invention do not inherently indicate any particular order nor imply any limitations of the invention.

Turning now to the drawings, FIG. **1** shows an isometric view, of a system of one embodiment invention during the process of wireless power transfer between a primary unit **20** and a secondary unit **30**. The primary unit converts the DC power supplied by a conventional DC power supply **10** connected to an AC outlet **11** to a multitude of radio frequency oscillations. The various RF oscillations can be of different RF frequencies, or they could all have the same frequency. The RF signals are used to drive a set of primary coil arrays which are in general two dimensional in nature to generate inductive magnetic field above a substantially laminar surface **27** in which the primary coils **21** are embedded. When a device with a device adapter equipped with one or more secondary coils **31** which can resonate with the inductive magnetic field excited by the primary coils **21**, the RF power is transferred from the primary unit **20** to the secondary unit **30** by a RF cable **2**. The transferred RF power is subsequently conditioned and rectified by the rectification circuit within the secondary unit **30** into a regulated DC power with a substantially constant voltage. The regulated DC power is then used to power the device to perform such tasks as charging a secondary battery, or to be used directly by the device to power its electronics, etc.

FIG. **2** depicts an isometric view of a single array of inductive field coils **21** of the primary unit **20** together with a single coil **31** of the secondary unit **30**. It also illustrates the localization of the inductive field when the secondary coil **31** is brought close to the primary unit **20** (see the |B| profile in **400**). The self-induced localization of the inductive magnetic field of the primary unit arises from both the local resonance nature of the primary RF circuit, and from the adaptive circuitry that resides in the secondary unit **30**. The primary unit has a local parallel LC (inductance **21**-capacitance **22**) balanced architecture to present a large primary load impedance to the RF power supply. Anytime the LC balance is strongly altered in any local region, the parallel LC impedance decreases drastically to allow primary RF load current to flow into the local region. When a secondary unit **30** is placed anywhere on the top surface **27** of the primary unit **20**, the resonance circuitry within the secondary unit **30** couples to the local LC circuit of the primary unit **20**, and together, they form a new local resonance substructure. In order for the new local resonance substructure to tune in to the applied RF frequency, the secondary unit **30** must adaptively adjust its own self impedance to pull the resonance in. The need of an adaptive component in the secondary unit **30** also means that it would be nearly impossible for any incompatible electronic device or a foreign object such as a metal bar, low reluctance magnetic material, or saline water to shift the local resonance enough to cause even a minute amount of power transfer. This greatly contributes to the safety of the present invention.

FIG. **3** is an isometric view of the primary unit **20** which comprises a two-dimensional array of primary coils **21** and attendant matching capacitors **22** as well as the power rails **29** for connecting the primary coils **21** and capacitors **22** to the primary RF power supply **10**. The primary coils **21** are arranged in groups such that coils **21** from one group do not have strong coupling to coils **21** from another group because of the way each group is driven by the RF source, and neighboring coils **21** from the same group likewise are only weakly coupled because of the physical and geometrical constraints. Each compensating capacitor **22** is placed in such a way that it can match the impedance of either a single primary coil **21** or a set of neighboring primary coils **21** to form a local parallel LC network. At its resonance frequency (or, more properly, the anti-resonance frequency, since the complex impedance of the parallel LC network has a pole in the complex frequency plane. Henceforth we will use the terms resonance and anti-resonance interchangeably to describe the impedance of a parallel LC network for simplicity, with the understanding that when a resonance is referred to a parallel LC network, it actually means anti-resonance), the LC network exhibits nearly infinite impedance to the RF source, hence the primary load current through such network is normally blocked provided that all the local LC networks are tuned to the primary driving frequency. Therefore, if no secondary unit **30** is placed in the vicinity of such a network, the amount of load current flowing through it is at its minimum.

When a compatible secondary unit **30** sits on top of the primary surface **27**, the parallel LC resonance frequency is shifted away from the driving frequency of the RF source **10**, which, in turns, causes the effective local LC impedance to decrease sufficiently to enable the load current to flow into the local LC network. The receiving coil in the secondary unit **30** likewise has a compensating capacitor **32** in parallel with the receiving induction coil **31**. The act of bringing the secondary unit **30** close to one or more of the local LC networks of the primary unit **20**, the inductive coupling is established between the primary unit **20** and the secondary unit **30**. This mutual inductive coupling creates a new local resonance substructure with its own resonance and anti-resonance frequencies. The increase in the local load current in turns results in the increase in the local magnetic field intensity which is shown in Graph **410**. The compensating capacitor **32** in the secondary unit **30** is electronically adjustable. By varying the compensating capacitance of the secondary unit **30**, the resonance frequency of the combined local resonant network can be changed.

FIG. **4** is a schematic view showing the interaction between the primary induction coils **21** of the primary unit **20** and the receiving coil **31** of the secondary unit **30** in a simplified single primary coil array system for illustration purpose. As shown, the primary unit **20** comprises a linear array of parallel LC networks all connected in parallel by the two RF feed rails **29**. Alternatively, the number of matching capacitors **22** can be reduced by using one capacitor **22** to balance the load impedance of several neighboring inductive coils **21**. In the extreme case, a single capacitor **22** can be used to balance the entire primary coil array. The last case corresponds to a global impedance compensating scheme.

In the more local impedance compensating geometry, there is very little load current that flows through the RF feeder rails **29** during standby phase. As the impedance compensation becomes progressively more global, there is a corresponding increase in the standby load current flowing in the feeder rails **29**. This is because the compensating capacitor **22** only blocks the load current that flows into the effective parallel LC network, but not the current that is internal to the effective parallel LC network. The RF load current in the feeder rails **29** introduces unwanted stray inductive magnetic field which increases ohmic heating, making the global compensation less desirable for large primary surfaces **27**. However, for smaller primary surface **27**, the need for local compensation is diminished.

The secondary unit **30** comprises a receiving coil **31** in parallel with a matching capacitor **32**. The device that the secondary unit **30** is attached to is simplistically represented by a resistive load **33** in series with the parallel LC network formed by the receiving coil **31** and the matching capacitor **32**. The variable nature of the capacitor **32** is indicated by the arrow across the capacitor symbol **32**. When the secondary unit **30** is brought to the proximity of the primary surface **20**, a mutual coupling is established between the secondary coil **31** and one or more of the proximal primary coils **21**. The matching capacitance of the secondary unit **30** is chosen so that its default value is at or near the value needed for a parallel LC resonance at the operating frequency of the primary unit **20**. The magnetic coupling of the primary coils **21** to the secondary coil **31** is greatly enhanced by such a resonance structure of the secondary unit **30**.

The resonance enhanced coupling drastically altered the local load impedance of the combined network, which in turns shifts the primary local parallel LC resonance away from the operating frequency of the primary RF source **10**. Perhaps more importantly, the resonant coupling introduces additional complex zeros and poles of the effective primary load impedance. Furthermore, the additional zeros and poles (or resonances and anti-resonances) are all clustered together around the original resonance frequency which is taken to be the operating frequency of the primary unit **20**. By varying the secondary capacitance, the dominant zero of the combined local network can be tuned to the driving frequency of the primary unit **20**, which in turns varies the magnitude of the RF voltage across the secondary load **33**. This makes it possible to regulate the load voltage of the secondary unit **30** using a feedback control scheme. The feedback control loop continuously monitors the output voltage (usually the rectified DC voltage after the rectification stage that is needed to convert the RF power into DC power to power an electronic device, instead of the un-rectified RF voltage) and compares it to a reference voltage. The deviation of the output voltage from the reference voltage is the error signal which is fed to a controller. The controller output a control signal which adjusts the capacitance value of the secondary RF matching capacitor **32** to tune the resonance frequency of the combined local resonance network either toward or away from the primary operating frequency in order to bring the output voltage close to the reference voltage. The details of the resonance tuning and output voltage regulation will be elaborated later when FIG. **12** and FIG. **13** are discussed.

FIG. **5** is a schematic view of an alternative embodiment of the primary coil array driven by two coherent RF sources **10** which have a 900 phase difference, showing the interaction between the primary induction coils **21** of the primary unit **20** and the receiving coil **31** of the secondary unit **30** in a simplified single primary coil array system for the purpose of elucidating the innovative concepts. This alternative embodiment overcomes a serious problem that arises from the way primary coils **21** are driven in the previous embodiment as described in FIG. **4**. In the previous RF driving arrangement, the coils are all driven in phase. Now the nature of the magnetic field generation by the individual primary coil **21** is such that the axial component of the magnetic field changes sign as it crosses from the interior of the coil **21** to the exterior through the coil boundary. It thus follows that there is a great deal of negative mutual inductive couplings between two adjacent primary coils **21***a* and **21***b*. When they are driven in phase, the inductive magnetic field generated by one coil **21***a* tends to oppose the field generated by the other coil **21***b*. For this reason the inductive magnetic field from one coil **21***a* is largely canceled by the fringe magnetic field generated by the adjacent coil **21***b*. When all the primary coils **21** are activated in phase, the combined magnetic field strength is drastically reduced relative to the magnetic field strength of a single coil **21***a* because of the mutual cancellation effect. The effect is most serious near the center of the primary coil array where cancellation effect is the largest. A graphic depiction of such cancellation effect is shown in FIG. **8**.

The highly non-uniform magnetic field distribution is clearly one drawback of the same-phase driving geometry since this would implies that the magnetic field strength at the maximum point will be excessively high if the magnetic field minimum has to be above a certain threshold value in order to provide effective inductive power transfer over the entire active area of the primary surface **27**. And since the field strength reduction at the center of the coil array depends strongly on the number and sizes of the primary coils **21**, the ability to scale up to larger primary surface area is severely hampered. Worst, owning to the fact that the inductive coil is made up of wire or trace of finite physical dimensions, and wires or traces between two adjacent coils **21***a* and **21***b* need to have a finite separation distance, the magnetic field between two adjacent coils **21***a* and **21***b* actually can become negative, i.e. point in a direction which is the reverse of the direction of the main magnetic field. As the magnetic field generated by the coil **21** varies in strength in inverse proportionality to the distance between it and the wire or trace that carries the RF current, the magnitude of such reverse magnetic field can be rather large. The upshot of this is that for certain small size secondary receiving coils **31**, the inductive coupling can actually go to zero. This occurs when there the total magnetic flux that is enclosed by the secondary coil **31** becomes zero as there are equal amount of positive and negative magnetic fields passing through the secondary coil **31** when the secondary coil **31** is positioned right along the border between the main field region and the reverse field region.

These drawbacks are overcome by dividing the primary coils **21** into two interleaving subgroups and drive the member of each subgroup with one of the four phases in a four phase drive scheme. For example, coil-capacitance network **100** is driven with an in-phase RF driver. Coil-capacitance network **101** is driven with a 900 quadrature phase driver. Coil-capacitance network **102** is driven by the same in-phase driver as coil-capacitance network **100** but in reverse, and coil-capacitance network **103** is driven by the quadrature phase driver in reverse. Coil-capacitance networks **104**, **105**, **106** and **107**, etc. are driven in turns by the same phase sequence over and over again until there is no more coil left. Within each subgroup, any two nearest coils are separate roughly by the diameter of one primary coil **21**, assuming all coils **21** are equal in size. Due to the fact that the stray return magnetic field diminishes in strength fairly quickly away from the coil, the two nearest coils **21***c* and **21***d* within the same subgroup will have almost negligible mutual inductive couplings because of the physical separation. Adjacent coil coupling is also drastically reduced by virtual of the fact that tow adjacent coils **21***c* and **21***e* belong to different subgroups. And except at the edges of the primary array, a coil belonging to one subgroup, such as **21***e* is always flanked by two coils belonging to the other subgroup, such as **21***c* and **21***d*. Consider Coil-capacitance networks **102**, **103** and **104**, for example; Coil-capacitance networks **102** and **104** are driven 1800 out of phase, hence the negative mutual inductance from **102** to **103** is substantially canceled by the positive mutual inductance from **104** to **103**. Thus it follows that **103** is essentially uncoupled from the subgroup that coil-capacitance networks **102** and **104** belong to. Thus the 900 driving geometry by and large decouples the coils **21** from one another. This drastically diminishes the non-uniformity of the magnetic field except near the edge of the primary active surface. It also makes scaling up to larger surface a trivial task because of the substantially uncoupled nature of the primary coils **21**. FIG. **9** and FIG. **10** describe the issue of coupling uniformity further.

FIG. **6** is a schematic view of an embodiment of a two-dimensional primary coil array that is conceptually identical to that of FIG. **4**. In this embodiment all primary coils **21** are driven exactly in phase. So it suffers from the same ills as in FIG. **4**, namely, it has strong coupling non-uniformity, is difficult to scale up, and has magnetic field null as well as field reversal regions which makes effective contact-less power transfer impossible in those areas. The existence of dead spots in the active area of the primary surface **27** is a serious drawback which will impede the usability of the system and limit its acceptance. One possible way of removing the dead spaces is to introduce another array of coils **21** underneath the first coil array. The second array is offset from the first array so that its own dead spots do not directly overlap with those of the first array. The second array can be driven by an independent RF power source **10** which has no fixed phase relationship with the RF source **10** for the first array. Alternatively, the second array can be driven 900 out of phase from the first one. In theory at least, when a secondary coil **31** is placed on a dead spot of the first coil array, it is still covered by the overlay coil array. In practice, the dead areas usually form long and narrow contours, hence some dead areas created by the original array and those created by the overlay array may still overlap. It is possible to design an overlay array network such that there are no overlapping dead spaces. However, even in those designs, there are still spots where the inductive magnetic field and hence the magnetic coupling is too weak to be useful. The introduction of the overlay array also increases the overall thickness as well as the system complexity of the primary unit **20**. There are other issues as well. For example, if the coils **21** of the stacked arrays are in the shape of a dense spiral formed on a layer of printed circuit board, then the amount of copper on the top array may well prevent the inductive field generated by the lower array from reaching the top surface, and vice versa. Thus, the introduction of an overlay array mitigates, but does not solve the coupling non-uniformity problem.

FIG. **7** is a schematic view of an embodiment of a two-dimensional primary coil array that is conceptually identical to that of FIG. **5**. The primary coils **21** are again divided into two groups in a checkerboard fashion. All physically adjacent coils are driven 900 out-of-phase and any two coils which are one coil apart and therefore belong to the same subgroup are driven 1800 out-of-phase. The same argument for the independence of primary coils **21** that applied to the simplified one dimensional array in FIG. **5** still applies here. The mutual inductances between two nearest coils that belong to the same subgroup still remains small although they are larger than those in the one dimensional case since the two coils are actually right next to each other along the diagonal. The coils **21** from the two subgroups have substantially no coupling between them because each interior coil from one subgroup is surrounded by four coils from the other subgroup whose net induction flux through the center coil is ideally zero.

Alternatively, the two-dimensional primary coil array can also be driven by other multi-phase scheme such as a 6-phase drive. Although for such drives, the independence of coils is no longer guaranteed. The larger mutual inductances among the coils **21** make it harder for the drivers to maintain their phase relationships because of the mutual couplings. The added complexity of the higher phased drives also leads to greater cost. The chief advantage of higher phased drive schemes is the increased uniformity of the inductive magnetic field.

FIG. **8***a* shows the magnetic fields **261**, **262**, **263**, **264**, **265** for an array of five coils and the total magnetic field when all five coils are energized in phase by currents of equal amount. It is noted that the magnetic field strength is the strongest immediately above the current carrying conductor that constitute the coil, and it dips to a minimum at the center of the coil. In addition, the perpendicular component of the magnetic field changes its sign as it crosses the outer boundary of the coil. The field reversal behavior is evident from the large negative dip just outside the coil in FIG. **8***a*. It can be seen from the graph that the field reversal behavior is significant only for up to a distance equal to about one diameter of the coil. Beyond that distance the strength of the reverse field diminishes rapidly. With all five coils driven in phase, the resultant magnetic field **266** is the sum of all the magnetic field generated by the five coils. Since the currents in adjacent conductors from different coils tend to cancel out, the five coils sum to an equivalent of one large coil. The magnetic field distribution **266** can be seen to be much lower in strength with a pronounced dip in the center of the equivalent large coil.

The large overlap in magnetic field distribution between adjacent coils means that they will have a very large mutual inductance, and the field reversal also implies that the mutual inductive coupling between the two coils will be negative. Because of the large mutual coupling between adjacent coils, it is no longer possible to treat the individual coils as independent. Hence, for example, the total inductance of two adjacent coils driven in parallel and in phase is not half of the inductance of a single coil, but is a lot less owning to the large negative mutual coupling between the two coils. This makes it difficult to use the design for a smaller version of the primary unit to design a scaled up version since all the inductances have to be recalculated. Furthermore, in the scaled up version, more turns are needed to compensate for the strong mutual cancellation of the magnetic field.

In calculating the sum of the magnetic fields **266** in FIG. **8***a*, it was assumed that the adjacent coils literally touch each other. It was also assumed that the coils are made of wires with no width. That explains why there are no reverse field regions for the summed field in FIG. **8***a*. In reality, coils are made of wires of finite size, and wires from adjacent coils need to be insulated from each other, usually by an air gap. FIG. **8***b* depicts the summation field **267** of the same five coils except that any two adjacent coils are separated by ⅕th the coil diameter. The rather pronounced negative peaks are the result of the overlap of the negative peaks from adjacent coils. In practice, the coils often take the shape of a spiral, comprising a series of turns having substantially linear sections. Those kinds of coils are often constructed using printed circuit board (PCB) technique. Under certain conditions, a plurality of such coils can also create large reverse field regions when they are driven in phase. The mere presence of field reversal regions is in itself not a significant issue since the direction, or the sign, of the magnetic field is an oscillating function of time; hence the sign of the magnetic field at any particular instant is unimportant. However, the very existence of the reverse field implies that there are zero field contours which are the borders between the normal field regions and the reverse field regions. A secondary coil **31** placed at or around one of the field nulls will see both positive magnetic flux and negative magnetic flux which approximately cancel each other. Since the magnetic coupling between the primary coils **21** and the secondary coil **31** is directly proportional to the amount of magnetic flux that is enclosed by the secondary coil **31**, the cancellation of the magnetic flux significantly reduces the coupling and therefore the power transfer.

Multi-phase driving of the primary coils addresses both the field reversal problem and the scalability problem. This can be seen in FIG. **9**, which displays a time sequence **271**, **272**, **273** of magnetic field intensity to depict the traveling wave nature of the magnetic field excited by a long linear array of coils driven by the circuit shown in FIG. **5**. FIG. **10** shows the magnetic fields **391**, **392**, **393**, **394**, **395** excited by a long linear array of primary coils **21** at various times, in which the complete absence of magnetic null is evident.

A multi-phase driving of the primary coil array is similar in nature to the technique used in driving a phased array antenna. Ideally, an N-phase drive, where N is a large number, will produce a magnetic field distribution which is close to a traveling sinusoidal waveform. In practice, N is usually limited to a small to moderate number, since a large N would entail the use of a large number of costly phase shifters. The minimum N is dictated by Nyquist Sampling Theorem. The 4-phased drive sequence depicted in FIG. **9** clearly has an appearance of a highly distorted traveling sine wave. Due to the distorted nature of the traveling wave profile, the RMS (root-mean-square) magnetic field strength is not strictly uniform, as it would if the profile were truly sinusoidal. The calculated RMS variation of the magnetic field strength for the sequence depicted in FIG. **10** is about ±22%. A more relevant measure of the field non-uniformity is to use the average field intensity of the computed magnetic field over a small aperture to calculate the RMS variation. By choosing an aperture which corresponds roughly to the physical dimension of a secondary coil, the computed RMS variation is reduced to ±13%. A similar computation is done for a 6-phased drive example, and the aperture averaged RMS variation is only ±6%. However, the superior RMS variation figure is tempered by the higher mutual coupling among coil elements which makes phased driving more difficult and costly.

Applying multi-phased drive to a two-dimensional primary coil array results in a traveling wave propagating along the diagonal direction. FIG. **11** displays a time sequence **491**, **492**, **493**, **494** of magnetic field intensity distribution that shows the traveling wave nature of the magnetic field excited by a two dimensional array of primary coils **21** driven by the circuit depicted in FIG. **7**. The two dimensional array has roughly one and a half times the RMS variation of its one-dimensional counterpart, or around ±20% for a 4-phased driving scheme while the 6-phased scheme has about ±10%.

Owing to the similarity of the multi-phase RF driving technique to the techniques employed in driving a phased array antennas, care must be taken to ensure that the RF radiation level be kept at an extremely low level. In a phased array antenna, bore sight direction is determined by the apparent phase velocity of the phased array elements. For example, when all the array elements are in phase, the apparent phase velocity becomes infinity, then it follows that the bore sight direction is the normal direction to the plane of the phased array (assuming planar phased array). When the apparent phase velocity is twice the speed of light, the bore sight direction is 300 from normal. In general, the deviation of the bore sight direction from normal is simply the arc-cosine of the ratio between the speed of light and the apparent phased speed of the array radiating elements. It should be apparent that the apparent phase velocity of the radiating elements should never be lower than the speed of light. Hence, in order to prevent the primary unit from radiating, the apparent phase speed of the diagonal traveling wave should be much smaller than the speed of light. In the 4-phased drive, the array pitch, which is roughly the diameter of a single coil, is ¼th the apparent phase wavelength, hence the apparent phase speed is the speed of light divided by 4 times the vacuum wavelength of the electromagnetic wave oscillating at the RF frequency and then by the array pitch. As an example, the RF frequency is taken to be 30 MHz, and the array pitch of the primary unit is assumed to be 10 cm, then the apparent phase speed is just 4% of the speed of light. This ensures that the primary array driven by a 4-phased RF driving technique does not radiate. By contrast, the single phase drive technique produces an apparent phase speed which is equal to infinity; hence RF radiation is not precluded. However, since the active area of the primary unit is typically much smaller than the RF wavelength, the level of RF radiation is still quite small since the primary unit behaves like a very poor antenna with negative antenna gain.

FIG. **12** is an isometric view of an exemplary embodiment of the secondary coil **31** and circuit elements in the device adapter. The secondary unit **30** comprises a spiral shaped PCB trace coil as the receiving coil **31**, a main capacitor **32***a* in parallel connection with the receiving coil **31**, and a bank of switched capacitors **32***b* which are also in parallel with the main capacitor **32***a* to form a variable and electronically adjustable capacitance, an impedance matching network **37***a*, a full bridge rectifier **37***b*, output capacitors **37***c* that smoothes out output voltage ripple, a feedback voltage controller **34** to regulate the DC output voltage even as the inductive coupling between the secondary unit **30** and the primary surface **27** is constantly changing. The output capacitor **37***c* also serves as the storage capacitor which accumulates energy to power the feedback controller during initial startup. During startup phase, the variable capacitor **32** of the secondary unit **30** is still at its default value and the rectified DC voltage is still unregulated, the unregulated DC current starts charging the storage capacitor **37***c* until **37***c* reaches a voltage level that is high enough to turn the controller **34** on. Once the feedback controller **34** is turned on, it starts to regulate the voltage by comparing the voltage of the storage capacitor **37***c* and a reference voltage and takes the difference which is used as an error signal to control the variable capacitor. The capacitance adjustment causes the resonant circuit of the secondary unit **30** to tune in to the RF driving frequency of the primary unit **20** until the voltage reaches the reference voltage value. The controller **34** can also compensate for any change in magnetic coupling or device load variation by tuning the secondary resonant circuit in or out of the primary RF driving frequency through continuous output voltage monitoring and capacitance adjustment.

The ability of the secondary unit **30** to regulate its output voltage is advantageous in many situations where the device to which the secondary unit **30** is attached requires a regulated voltage to power it. Examples of devices that require regulated power are mobile phones, personal digital assistants, laptop computers, MP3 players, ipods, digital cameras and camcorders, radios, portable game consoles or players, etc. Typically, such device is powered by a regulated switched mode power supply whose output voltage and current are specifically designed for its use. The need of a dedicated DC power supply is a major inconvenience for an end user who may carry a multitude of electronic devices with him or her and each device requires a different power supply. In view of the present invention, the obvious solution is to equip each device with a secondary unit **30** which can receive power from the primary unit **20** and convert it to whatever the voltage the device needs. However, if the output voltage from the secondary unit **30** is unregulated, it may need a linear voltage regulator if the power need of the device is small or a DC-DC converter if more efficient voltage regulation is required. Since the secondary unit **30** of the present invention can supply regulated voltage to the device, the need to have a separate voltage regulation device is obviated, making the conversion from the conventional power supply to the contactless one of the present invention painless. With the addition of a DC current monitoring element, such as a series current sense resistor, or a current mirrored sense resistor when MOSFETs are used in the full bridge rectifier section **37***b*, it would be possible to program the feedback controller **34** to control both the voltage and current to behave like a lithium ion/polymer charger with little or no additional cost.

This specific embodiment of the secondary unit **30** can also be profitably employed to construct a universal adapter for portable devices. Modern electronic devices are designed with a lot of built-in protection mechanisms, including over-voltage protection, under-voltage protection, over-current protection, etc. If such a device is connected to a variable voltage power supply, it would block the supply voltage until the supply voltage exceeds a lower voltage threshold. Once the supply voltage has the right value, the device's power interface module will connect the device's power unit to the power supply. During the operation, if the supply voltage spikes to a value that exceeds the threshold for over-voltage protection, the device's power interface module immediately disconnect the power from the power supply. Similarly, if the supply voltage dips to a value that is below a certain threshold, which may or may not be the same as the initial threshold, then the device also disconnect itself from the power supply, until the supply voltage returns to its normal range. Thus, it would be possible to program the feedback controller **34** of the secondary unit **30** to increase the regulated output voltage in steps, with a sufficient wait time to account for the initial delay before the power interface module of the device connects the device to the secondary unit's output. The controller **34** of the secondary unit **30** monitors the device current to determine whether the connection has been made. When the controller has detected the rise in output current beyond a certain current threshold, it determines that the connection has been made, and the voltage scanning stops. During the initial voltage scanning steps, the controller also imposes a strict current limit to prevent inadvertent over voltage situation which might damage the device if it is not equipped with over-voltage protection. This may occur when the device in question is in a standby mode and does not consume enough current for the feedback controller to determine whether the device power connection has been made, in which case the controller **34** will decide that the voltage is not right and proceeds to step up to another voltage. The controller **34** can detect fault conditions to determine whether over-voltage fault has occurred since under normal conditions the device's power interface module would insert a hysteresis time delay before connecting the power supply to the device, hence if that time delay is not there, then an over-voltage fault is indicated. Whenever a fault occurs, the controller **34** can reset the process and start the voltage scan sequence all over again until the power connection has been made without triggering a fault.

A more detailed schematic view of the circuit elements in the device adapter is presented in FIG. **13**. The proportional, differential, and integral (PID) controller **34***a* monitors the output voltage after the storage/smoothing capacitor **37***c* and compares it to a reference voltage. The value of the bias capacitor **32***a* is selected in such a way that, when it is coupled to the secondary receiving coil **31**, the combined LC network resonates with the primary RF driving frequency in the absence of any magnetic coupling with the primary unit **21** in conjunction with the receiving coil **31**. This choice of the bias capacitance **32***a* also ensures that even without any further adjustment of the parallel capacitance, the RF voltage induced by the primary coils **21** when the secondary unit **30** is in magnetic coupling with the primary unit **20** is sufficient to power the PID controller **34**. Furthermore, the nature of the coupling between the secondary **30** and the primary unit **20** is such that it tends to slightly raise the resonance frequency, which means that there needs to be additional parallel capacitance in order to restore the resonance to its original value so that it can stay matched with the RF driving frequency. This capacitance correction is provided by a bank of switched capacitors **32***b* shown in FIG. **13**. Only four switched capacitors **32***b* are shown, although it would be more advantageous to have eight such capacitors **32***b* to provide both the resolution and the range needed for a more robust power transfer. The four capacitors **32***b* are selected to have values which are powers of 2 of the smallest capacitance. So for example, if the smallest capacitance is 100 pF, then the other three capacitances are 200 pF, 400 pF and 800 pF. Each capacitor is in series with a MOSFET switch which can be turned on to enable the capacitor it is directly connected to, or to turn off to disable the capacitor. Since each MOSFET switch can be either ON, or OFF, the switched capacitor bank **32***b* can provide 24 or 16 different capacitance values which are more or less linearly distributed with values from 0 pF to 1,500 pF. The capacitance values are not precise because of the inherent variation of stock capacitors as well as the source drain capacitances of the MOSFETs. The source drain capacitance of the MOSFET is unimportant when the MOSFET is ON. However, when MOSFET is OFF, the capacitor it is connected to is not completely disabled because of the source drain capacitance which is in series with the external capacitor. The source drain capacitance of the MODFET is not a constant, either, but depends on the source drain voltage. Hence it would be difficult to modify the capacitance values of the external capacitors to compensate for the added OFF capacitances caused by MOSFETs.

Because of the binary nature of the switched capacitor bank **32***b*, the PID controller **34** can control the capacitance of the capacitor bank **32***b* simply by issuing a binary word. For example, a 4-bit word can be used in the capacitor bank **32***b* depicted in FIG. **13** to change the capacitance to a multiple of the smallest capacitance such that the integer multiple is simply the numerical value of the 4-bit word. In this sense, the switched capacitor bank **32***b* can be considered to be a DA (digital-to-analog) converter which converts a binary word into an analog capacitance value. For a standard 8-bit controller, it is more likely that the bank will have eight capacitors, because it can then be controlled by an 8-bit word directly.

To supply the RF power to the primary coil array, the primary unit **20** must first convert the main power from the wall jack to DC power **191**, which is typically a switched mode power supply. The DC power is subsequently converted to a multi-phased RF power supply **10**. FIG. **14** is an isometric view of one embodiment of the RF generating module in the primary unit **20** which generates two pairs of RF signals whose phases are 900, 1800, 2700, and 00, respectively. Shown are, square wave generator **109** with crystal oscillator **110**, a pair of current amplifiers **108**, a pair of transformers **107** to step down the voltage, a pair of power MOSFETs **151** for high-frequency power switching, a pair of chokes **106** and their attendant smoothing capacitors **105** to isolate the DC power supply from the switching noise and vice versa, a pair of source drain capacitors **131** to control switch timing, and a pair of series connected inductors **101** and capacitors **115** to provide series resonance circuits, and two pairs of output ports **102**, **103**, one pair **102** for the 00 and 1800 phases, and the other pair **103** for the 900 and 2700 phases. The inductors **131** and transformers **101** as well as the chokes **106** have either ferrite cores or iron powder cores, depending on the RF frequency, or other suitable high permeability magnetic materials to provide sufficient inductance values without needing a large number of turns for the coils. The use of magnetic materials for the transformers **107** and inductors **101** also reduces RF cross coupling due to the leakage inductive magnetic field. This is because the high permeability material can act as a guide for the magnetic flux since the magnetic field prefers to complete its loop through a low reluctance path which the magnetic material provides. Not shown are metal mesh Faraday cages which are placed over critical components such as the square wave generator **109**, the current amplifiers **108**, all the transformers **107** and inductors **101**, etc. in order to provide RF shielding of unwanted inductive magnetic field. This can prevent, for example, the unwanted coupling between a transformer **107** and the square wave generator **109**. Since the square wave generator **109** is capable of generating only a small amount of RF current, even a relatively small interfering stray RF magnetic field can significantly distort the waveform, thereby creating a vicious positive feedback cycle, causing the RF generator **10** to become unstable.

The power supplied by the RF module is transmitted to the primary coils **21** through RF waveguides **2**. The most common waveguides are coax cables, not shown, which are lightweight, flexible, and have relatively low transmission loss. Within the primary unit **20** the RF power is delivered to the individual coils simply via separate rails **29**. However, if the primary surface **27** is large, then the RF power would be more advantageously delivered via coplanar waveguides or their equivalent in order to minimize transmission loss, reflections, and radiation. Splitters are also needed in conjunction with the coplanar waveguides to allow the individual coils to tap the RF power delivered by the shared waveguides.

FIG. **15** is a more detailed schematic circuit diagram view of the RF generating module in the embodiment depicted in FIG. **14**. The square wave generator **109**, with a crystal oscillator **110** to provide a stable, precise frequency reference, generates two square wave signals that are 900 apart in phase. Each signal is passed through a current amplifier **108** and then a step down transformer **107** to boost its current. The current amplification is needed because the power MOSFET switches **151** require large gate currents to switch at radio frequencies. Further current amplification is provided by the transformer pair **151** which steps down the drive voltages. The step down transformers **151** also insulate the gate drive section from load variations and RF reflections further down stream, which reduces the destabilizing effect of the Miller capacitance.

The section of circuit immediately after the transformer **107** is a class E amplifier. The choke **106** is used basically as a large flywheel which both insulates the DC power supply **191** from large high-frequency fluctuation and provides a constant current source to the rest of the class E circuitry. The series LC resonator **101** and **115** provides a high "Q" environment to filter out higher harmonics and enhance the fundamental frequency to produce a substantially sinusoidal waveform at the output ports **102** and **103**. It is to be appreciated that a class D amplifier, or any other switched pulse amplifier design can be profitably employed to accomplish the RF generating function. Both class E and class D amplifiers have been known to have achieved power conversion efficiencies as high as 92% and up, far higher than the more traditional class AB or class C amplifiers which typically have below 50% power efficiencies.

Switched pulse amplifiers are quite sensitive to load variations and load power reflections, which can be somewhat mitigated by Faraday shielding of magnetic components and gate current amplification. Ideally, the power MOSFETs **151** are switched only when there is a zero crossing for the source-drain current. When a power MOSFET is switched while its source drain current is still nonzero, the flywheel actions of the choke and the main inductor will try to maintain the current during the switch transition, thereby creating large voltage spikes with high harmonic content which lowers the overall power efficiency as well as distorting the output waveform. Unfortunately, the zero crossing timing of the source drain current is a strong function of both the load and the reflected RF power when there is a load impedance mismatch. However, since the load as well as the load impedance changes as soon as a device is placed on or removed from the primary surface **27**. And even during wireless power transfer, there is no assurance that those values would not change. A notebook computer, for example, would vary its current consumption as work load varies, and can go into a sleep mode with a prolong period of inactivity. A cell phone consumes most of its power during voice and data transmission, or when it is performing its automated chores or when a movie, a video, or high fidelity music is being played. If the RF generation module is optimized for a particular load, then it may work poorly for another load such as placing two cell phones on the primary surface **27**, or when there is no device on the surface. To solve this problem, a power management module **193** which can perform power factor correction (PFC) is needed.

FIG. **16** shows a PFC subsystem for the class E amplifier module depicted in FIG. **15**. Current and voltage sensors **181**, **182** are used to measure the output current and voltage. The results of the measurement can provide information about equivalent load impedance, RMS voltage and current, and the harmonic content of the amplifier output. Another voltage sensor **184** is employed to monitor the drain voltage of the MOSFET transistor **151** whose harmonic content gives a very good indication of whether the class E amplifier is operating efficiently or not. Those data are fed to the power management module **193**. Based on those inputs, the power management module **193** sends control signals to voltage control module **192**, capacitance control module **195** and duty cycle control module **196** to adjust the voltage of the main power supply **191** to the amplifier, the capacitance of the source drain capacitor **131**, and the square wave generator **109**. The DC voltage control is achieved by changing the reference voltage of the fly back controller within the switched mode power supply. The capacitance adjustment is accomplished by an electronically switched capacitor bank **131**, and the duty cycle control is performed by a digital potentiometer. Because the control has multiple inputs and multiple outputs, or MIMO, standard feedback control schemes such as PID scheme would no longer work. More advanced control schemes such as fuzzy logic control, linear or recursive least mean square or artificial neural network controls are needed. These approaches generally start with an objective function defined as the mean square error of one or more linear or nonlinear functions of the inputs. By way of example, the objective function could be the ratio of the energy of the second harmonics to the fundamental of the source voltage. Using a hill climbing (actually minimum seeking) adaptive algorithm, the power management module can gradually learn to control its outputs to reduce the harmonic content at the source voltage.

The drawings FIG. **12** through FIG. **16** are for illustrative purpose, and are not meant to be comprehensive, nor are they intended to exclude other components or techniques which are similar in scope and can perform similar functions. The switched capacitor bank performs the function of an electronically controlled capacitor, and as such, it can be replaced by other electronically controlled capacitors such as varactors, or by a micro-electro-mechanical system (MEMS) based thin film capacitor. Different sensors and control elements can be substituted in FIG. **16** for the PFC without fundamentally altering the basic principle of adaptive power optimization of the RF power module. A switched mode DC power supply can also be replaced by a large primary battery, or a fuel cell battery, in which case the control of the DC supply voltage may be inhibited, or a DC-DC converter may becomes necessary. The PID controller in the secondary unit may be replaced by a hybrid scheme where the four most significant bits are exhaustively searched to find a coarse optimum, and a PID controller then takes over to perform an analog search. Another alternative is to exhaustively search all eight bits periodically to find and track the optimum point. The output voltage or current of the secondary unit may be unregulated, and the PID controller or its equivalent is only employed to pull in the resonance. The multi-phased RF driving techniques used to power the apparent traveling wave in the primary unit could also be substituted by a single phase driver that relies on the inherent slow wave structure of the primary coil array to excite a slow traveling wave across the primary coil array. Another alternative embodiment is to use a random multi-phased driver where the fixed phase relationships are replaced by random phase relationships.

It is to be understood that the scope of the present invention is not in any way limited to the specific embodiments heretofore disclosed, but only by the following claims.

What is claimed is:

1. A system comprising:

a first unit with a substantially flat top surface, operatively connectable to a RF power source having at least a first frequency, and including an array of local resonance circuits having a resonance frequency selected to present a large load impedance to the RF power source; and

a secondary unit, including a resonance circuit having an adaptive self impedance such that when the secondary unit couples to a local resonance circuit, the local resonance circuit presents a lower load impedance to the RF power source allowing radio frequency power to flow from the first unit to the secondary unit.

2. The system of claim **1** wherein;

said local resonance circuits include at least a capacitor in parallel connection with an inductive member.

3. The system of claim **1** wherein;

said secondary unit includes at least one capacitor in connection with an inductive member.

4. The system of claim **3** wherein;

said capacitor is operatively connected so as to be in LC resonance with said inductive member of the secondary unit.

5. The system of claim **1** wherein;

said array of local resonance circuits are arranged in a substantially periodic manner.

6. The system of claim **5** wherein;

said array of local resonance circuits are to be driven by corresponding radio frequency (RF) drivers such that each RF driver has a fixed phase relationship with respect to a separate RF driver.

7. The system of claim **6** wherein;

said phase relationships are integer multiples of a fixed phase to form a circular array of equally spaced phases that cover 3600.

8. The system of claim **7** wherein;

when said substantially periodic array of local resonance circuits is driven by said plurality of phased drivers to create a traveling wave pattern across the top surface of the first unit.

9. The system of claim **7** wherein;

said plurality of phases includes four phases that are 00, 900, 1800, and 2700.

10. The system of claim **7** wherein;

said plurality of phases includes six phases that are 00, 600, 1200, 1800, 2400, and 3000.

11. The system of claim **4** wherein;

said substantially periodic array of local resonance circuits are driven by a plurality of RF drivers such that each RF driver has a random phase relationship with respect to another RF driver.

12. The system of claim **11** wherein;

when said substantially periodic array of local resonance circuits is driven by said plurality of phased drivers, to create a randomly traveling wave pattern across the top surface of the first unit.

13. The system of claim **1** wherein;

said secondary unit comprises at least one electronically adjustable capacitor module.

14. The system of claim **13** wherein;

said electronically adjustable capacitor module comprises a bank of parallel capacitors switched by transistors.

15. The system of claim **14** wherein;

said transistors are MOSFETs.

16. The system of claim **13** wherein;

said secondary unit includes a controller to adjust a capacitance in response to variation of magnetic coupling between said secondary unit and the first unit.

17. The system of claim **16** wherein;

said controller of the secondary unit is operatively coupled to monitor an output voltage of the secondary unit and to adjust the capacitance to bring the output voltage to a predefined reference voltage regardless of a variation of the output current.

18. A method of transferring power from a first unit including local resonant circuits having a large load impedance at a first frequency, to a secondary unit that includes a resonant circuit having a self impedance, comprising:

supplying radio frequency voltages to the local resonant circuits in the first unit;

adjusting the self impedance of the secondary unit in response to the secondary unit being proximate to at least one of the local resonant circuits in the first unit; and

transferring power from the first unit to the secondary unit via the at least one of the local resonant circuits and the resonant circuit of the secondary unit.

19. The method of claim **18** wherein the adjusting the self impedance includes varying a capacitance of the resonant circuit of the secondary unit.

20. The method of claim **18** wherein the supplying radio frequency voltages further comprises supplying radio frequency voltages having a plurality of phases generating a traveling wave radio frequency magnetic field pattern across a top surface of the first unit.

21. The method of claim **20** further comprising: adjusting said variable capacitance to increase a radio frequency voltage across said resonant circuit of the secondary unit.

22. The method of claim **20** further comprising: adjusting said variable capacitance to stabilize a radio frequency voltage across said resonant circuit of the secondary unit relative to a predefined reference voltage.

23. The methods of claim **21**, wherein adjusting said variable capacitance further comprises monitoring the output voltage of said resonant circuit of the secondary unit.

24. The method of claim **23** wherein adjusting said variable capacitance further comprises using a feedback control.

25. The method of claim **24** wherein using the feedback control further comprises a PID control process.

* * * * *

US 20100039234A1

(19) **United States**

(12) **Patent Application Publication**

Soliven et al.

(10) **Pub. No.: US 2010/0039234 A1**

(43) **Pub. Date: Feb. 18, 2010**

(54) RF POWER CONVERSION CIRCUITS & METHODS, BOTH FOR USE IN MOBILE DEVICES

(75) Inventors: **Marcello Soliven**, Glendale, AZ (US); **Tamio Saito**, Cupertino, CA (US)

Correspondence Address:
TROUTMAN SANDERS LLP
BANK OF AMERICA PLAZA
600 PEACHTREE STREET, N.E., SUITE 5200
ATLANTA, GA 30308-2216 (US)

(73) Assignee: **IVI SMART TECHNOLOGIES, INC.**, New York, NY (US)

(21) Appl. No.: **12/542,522**

(22) Filed: **Aug. 17, 2009**

Related U.S. Application Data

(60) Provisional application No. 61/089,440, filed on Aug. 15, 2008.

Publication Classification

(51) **Int. Cl.**
H04Q 5/22 (2006.01)

(52) **U.S. Cl.** **340/10.1**

(57) ABSTRACT

This patent application teaches and describes radio frequency (RF) power conversion circuits and methods both for use in mobile devices (such as smart cards). Embodiments of the present invention include wireless personal ID cards or dongle including a fingerprint sensor. A fingerprint matching system can reside on cards. Power provided to the fingerprint sensor and on board processer(s) can be provided by a wireless signal provided to the card. The card can include an RF power conversion circuit configured to receive wireless RF energy and convert the wireless energy for powering electronics on the card. Other aspects, embodiments, and features of the present invention are also claimed and described.

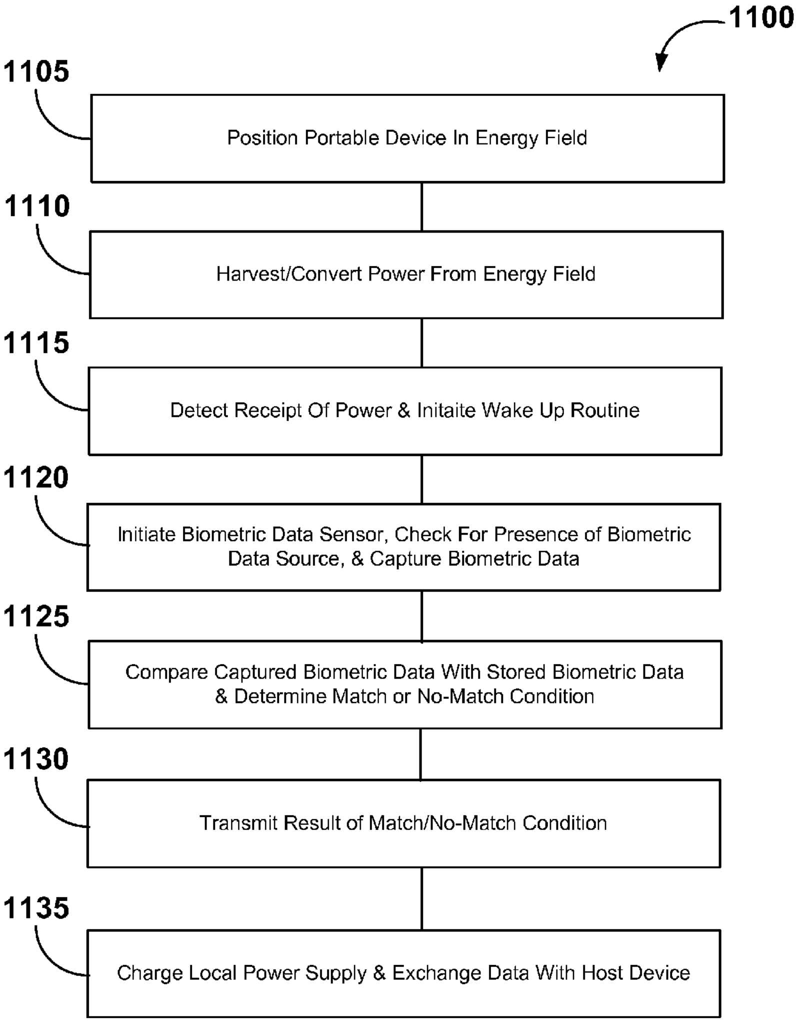
1100
1105
Position Portable Device In Energy Field
1110
Harvest/Convert Power From Energy Field
1115
Detect Receipt Of Power & Initaite Wake Up Routine
1120
Initiate Biometric Data Sensor, Check For Presence of Biometric Data Source, & Capture Biometric Data
1125
Compare Captured Biometric Data With Stored Biometric Data & Determine Match or No-Match Condition
1130
Transmit Result of Match/No-Match Condition
1135
Charge Local Power Supply & Exchange Data With Host Device

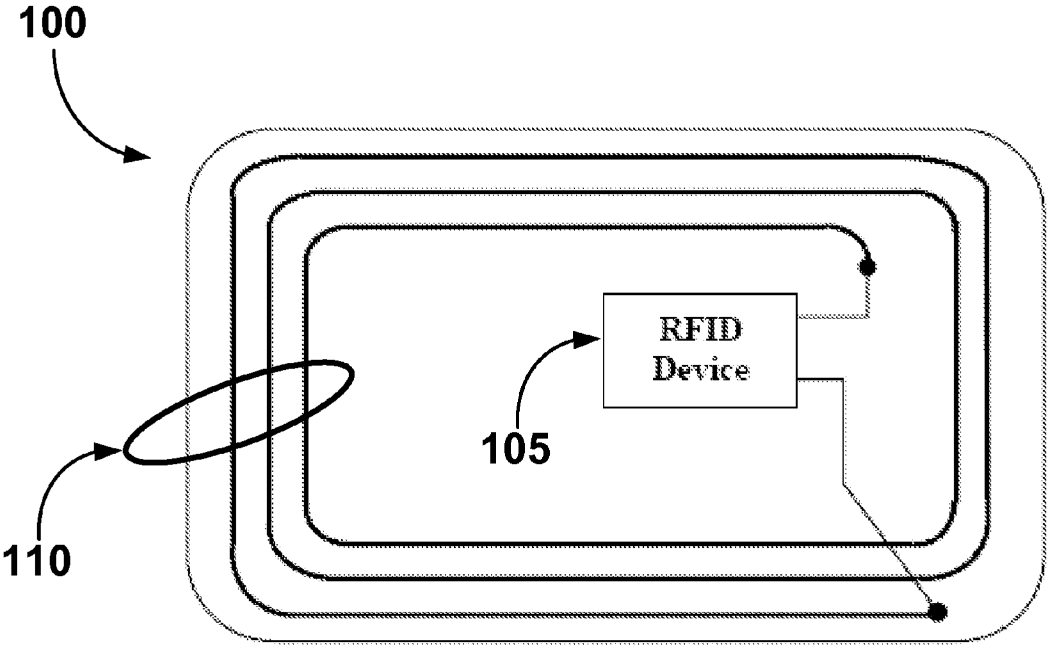
100
RFID Device
105
110

FIG. 1

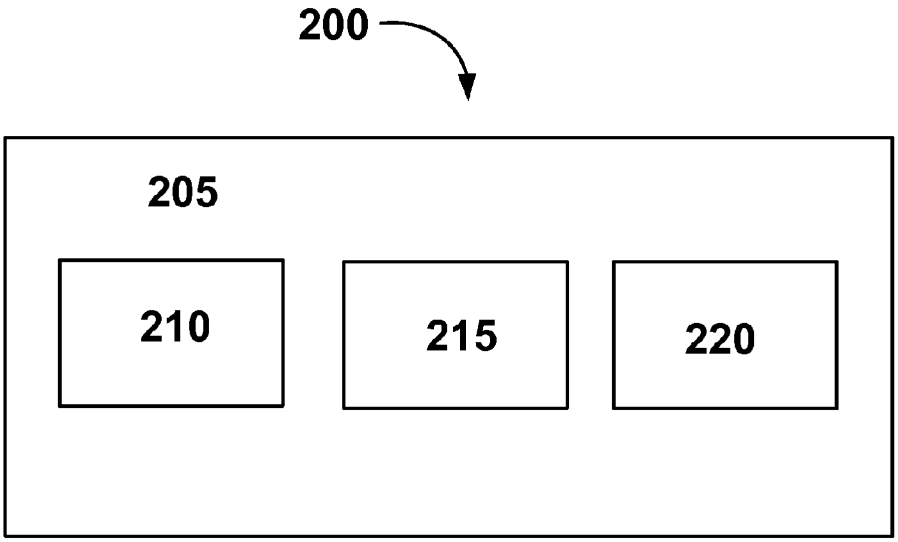
200
205
210
215
220

FIG. 2

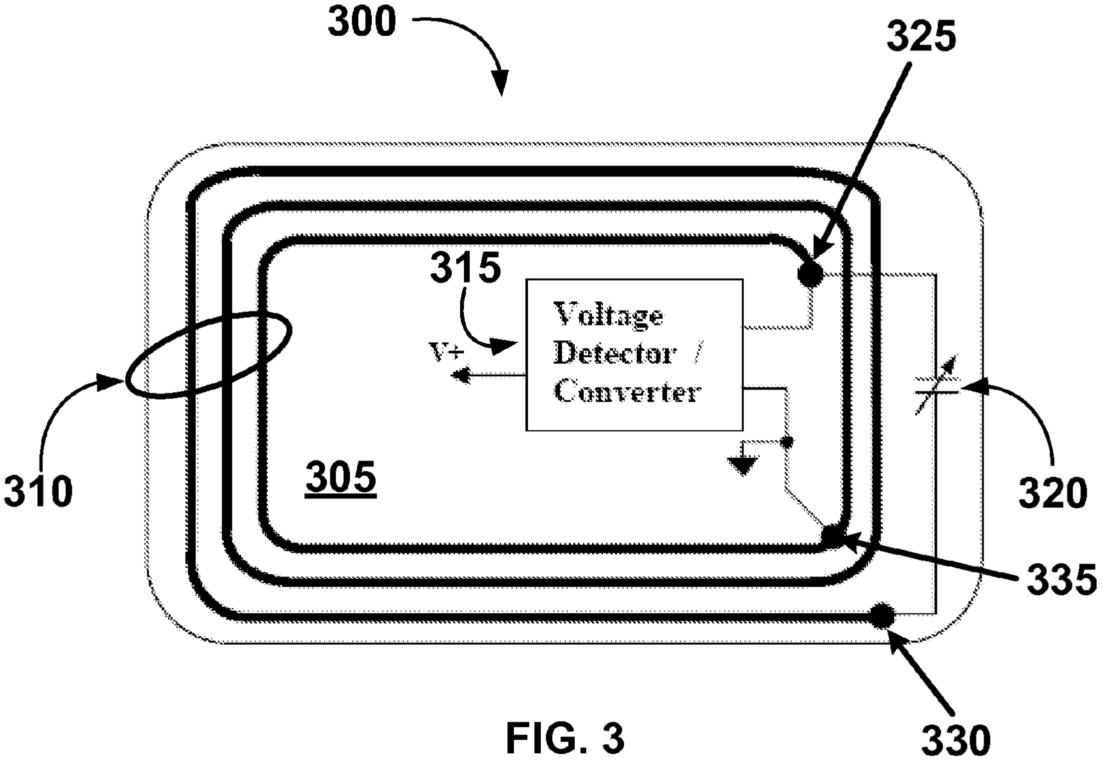
Voltage Detector / Converter
V+

FIG. 3

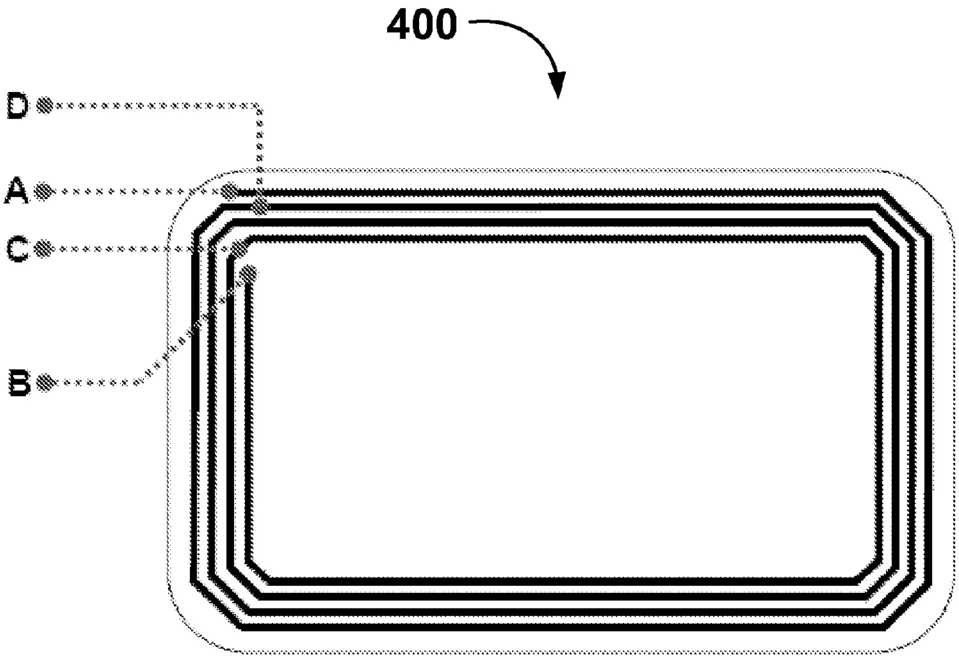
400
D
A
C
B

FIG. 4

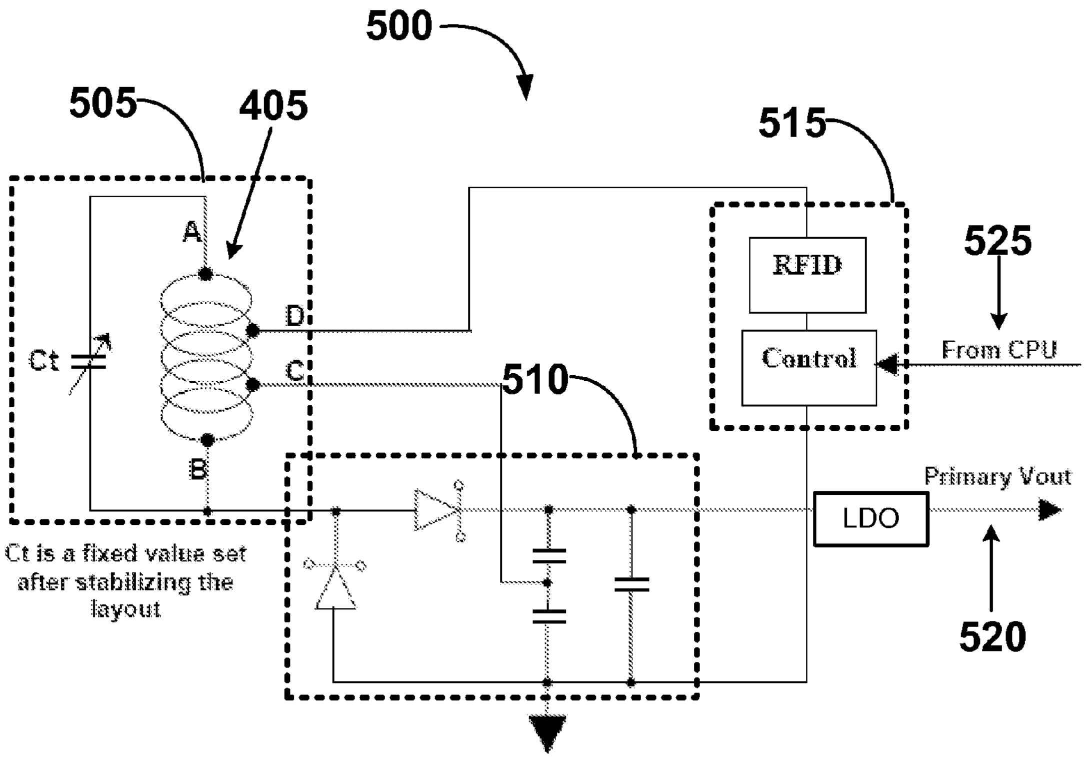
500
505
405
515
525
A
D
C
B
Ct
RFID
Control
From CPU
510
Primary Vout
LDO
520
Ct is a fixed value set after stabilizing the layout

FIG. 5

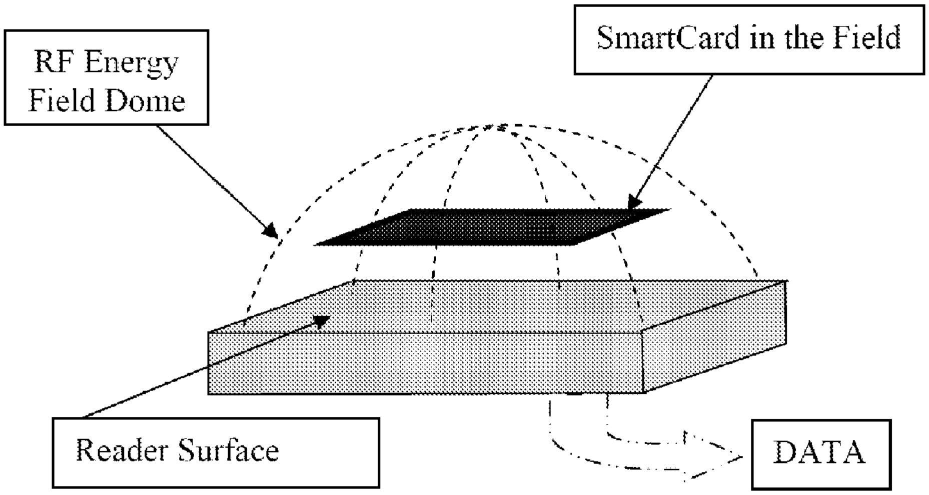
RF Energy Field Dome
SmartCard in the Field
Reader Surface
DATA

FIG. 6

FIG. 7

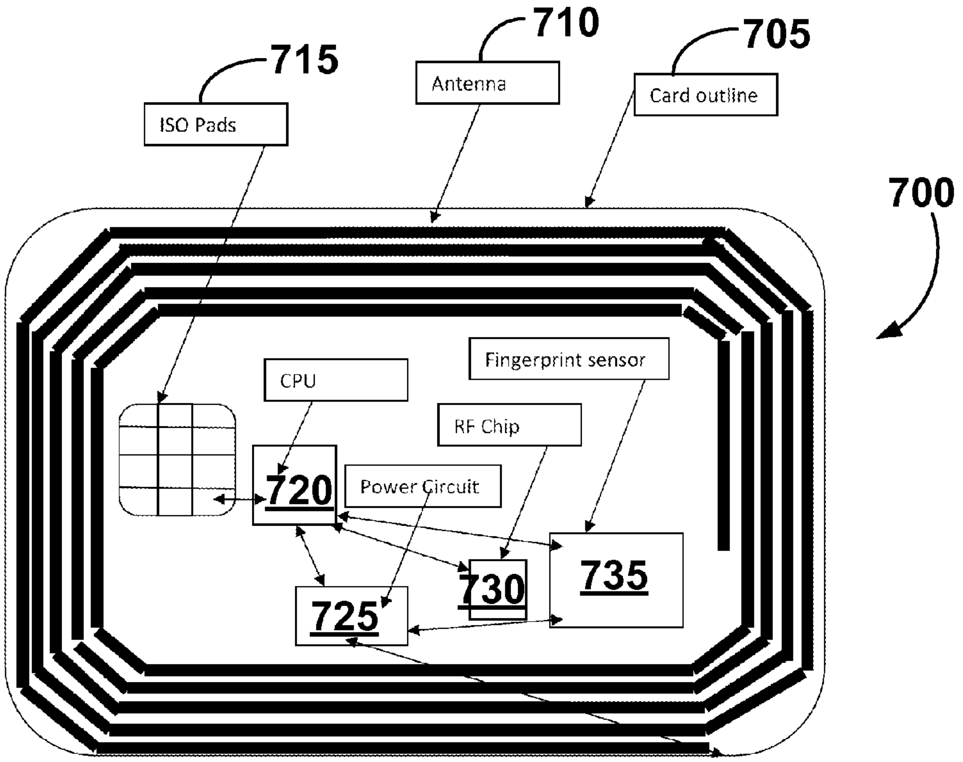
715
710
705
ISO Pads
Antenna
Card outline
700
CPU
Fingerprint sensor
RF Chip
Power Circuit
720
725
730
735

FIG. 8

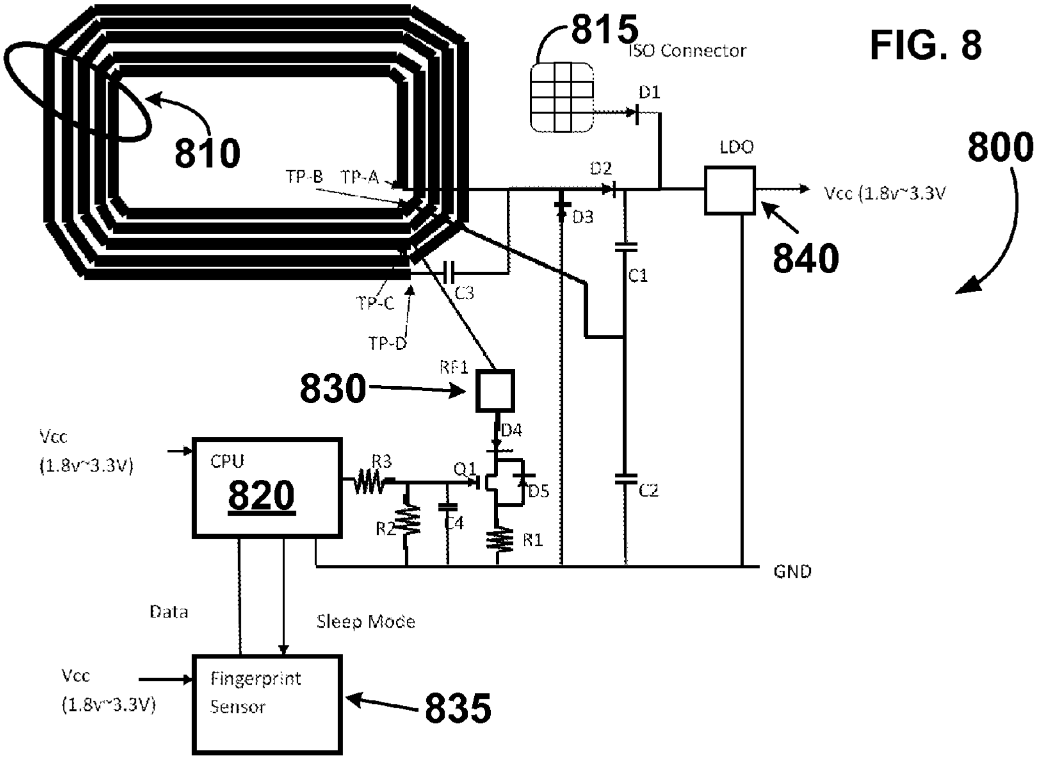
815
ISO Connector
810
D1
LDO
800
TP-B
TP-A
D2
D3
Vcc (1.8v~3.3V
840
C1
C3
TP-C
TP-D
RF1
830
D4
Vcc
(1.8v~3.3V)
CPU
820
R3
Q1
D5
C2
R2
C4
R1
GND
Data
Sleep Mode
Vcc
(1.8v~3.3V)
Fingerprint
Sensor
835

FIG. 9

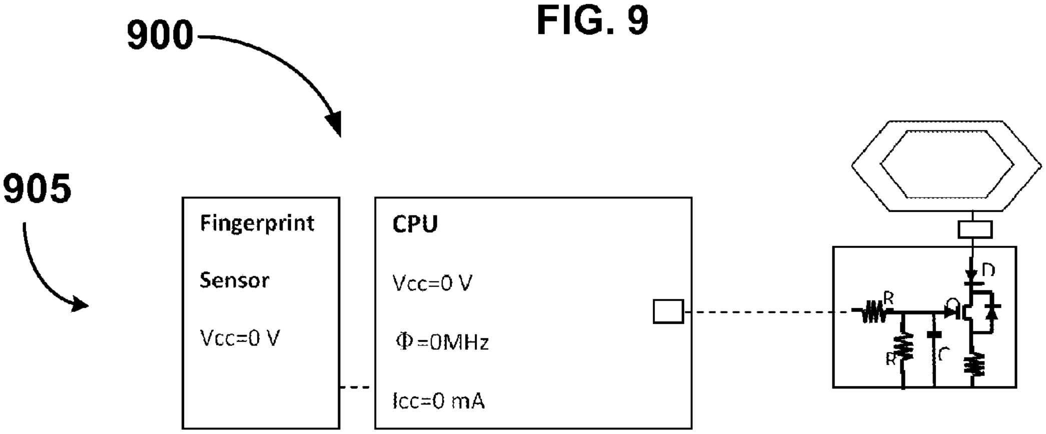
900
905
Fingerprint
Sensor
Vcc=0 V
CPU
Vcc=0 V
Φ=0MHz
Icc=0 mA
D
R
Q
R
C

(A) Initial state, No power exhists

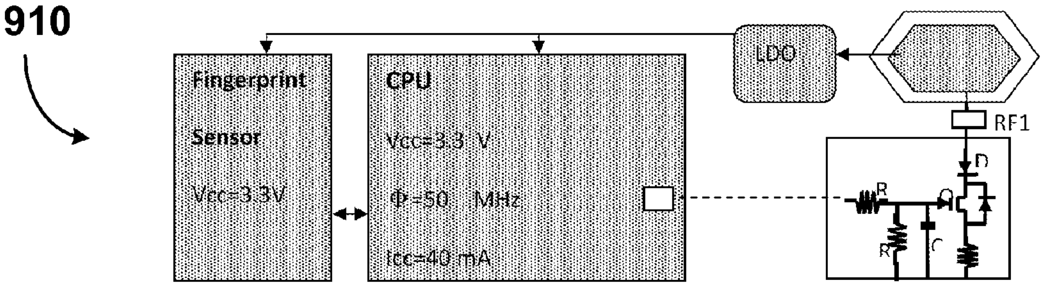
910
LDO
RF1
Fingerprint
Sensor
Vcc=3.3V
CPU
Vcc=3.3 V
Φ=50 MHz
Icc=40 mA
D
R
Q
R
C

(B) Full Power is applied to CPU and Fingerprint sensor for fingerprint matching calculations.

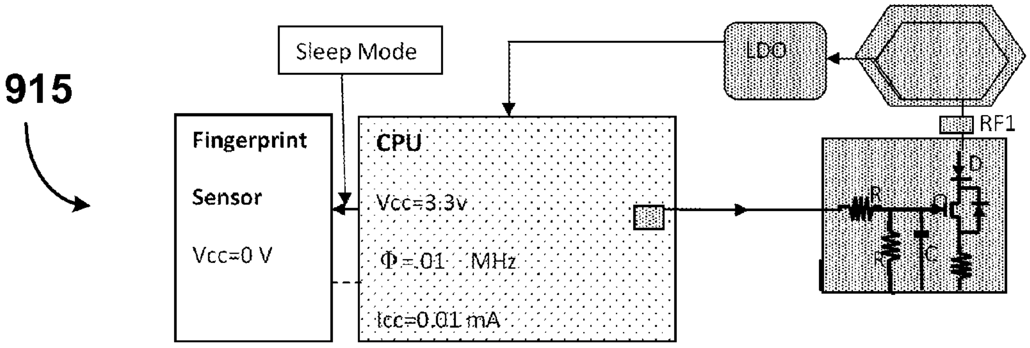
915
Sleep Mode
LDO
RF1
Fingerprint
Sensor
Vcc=0 V
CPU
Vcc=3.3v
Φ=.01 MHz
Icc=0.01 mA
D
R
Q
R
C

(C) CPU Power is minimized to maintain resister and I/O to drive transistor Q to drive RF chip RF1. CPU makes Fingerprint sensor as sleep mode.

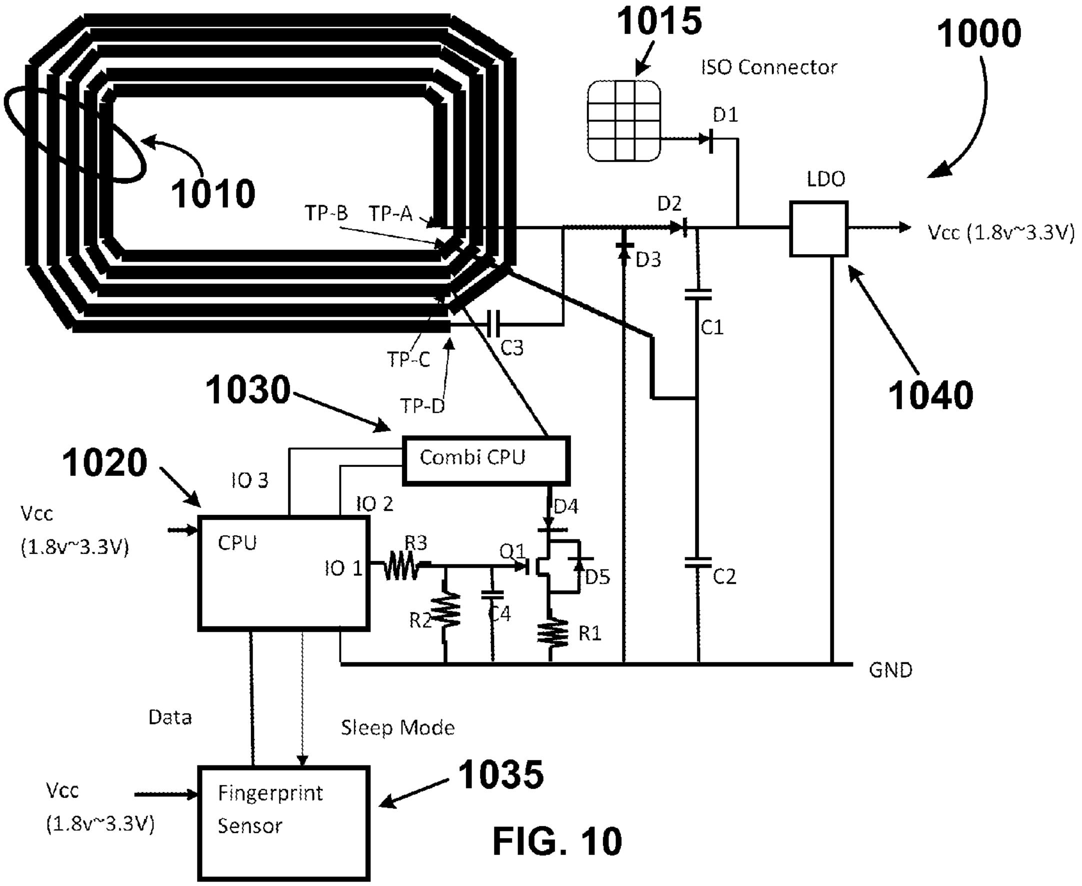
1015
ISO Connector
1000
D1
LDO
1010
TP-B
TP-A
D2
Vcc (1.8v~3.3V)
D3
C1
C3
TP-C
TP-D
1030
1040
1020
Combi CPU
IO 3
IO 2
Vcc
(1.8v~3.3V)
CPU
IO 1
R3
Q1
D4
D5
C2
R2
C4
R1
GND
Data
Sleep Mode
Vcc
(1.8v~3.3V)
Fingerprint Sensor
1035

FIG. 10

FIG. 12

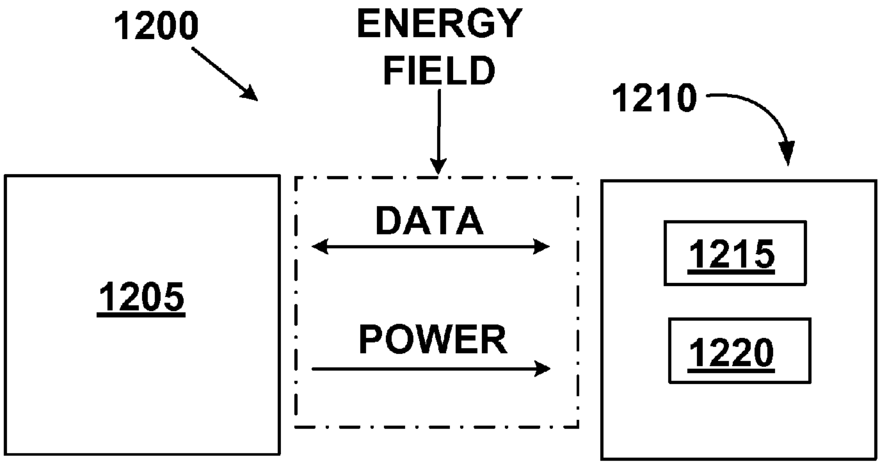
ENERGY FIELD
DATA
POWER

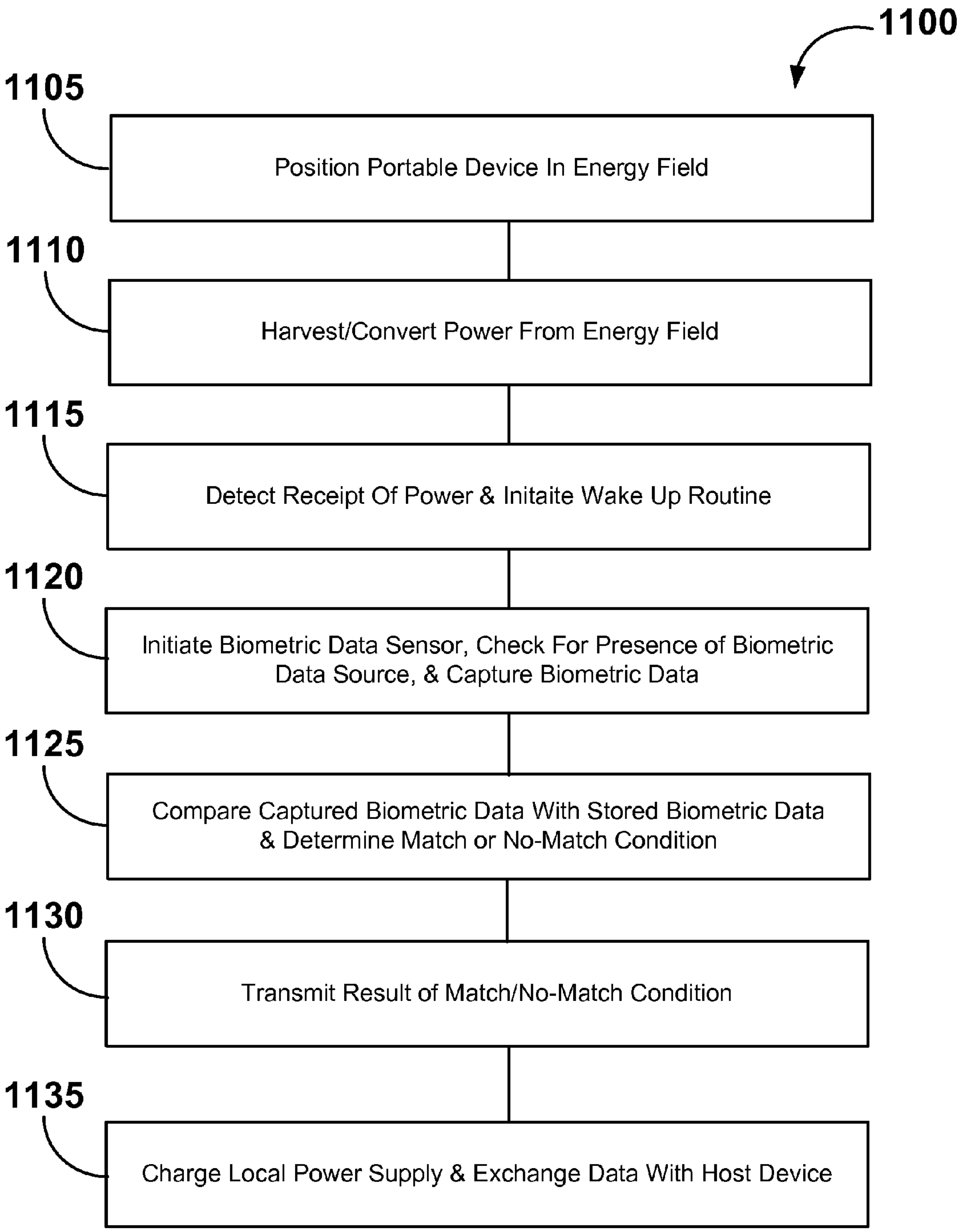
1100
1105
Position Portable Device In Energy Field
1110
Harvest/Convert Power From Energy Field
1115
Detect Receipt Of Power & Initaite Wake Up Routine
1120
Initiate Biometric Data Sensor, Check For Presence of Biometric Data Source, & Capture Biometric Data
1125
Compare Captured Biometric Data With Stored Biometric Data & Determine Match or No-Match Condition
1130
Transmit Result of Match/No-Match Condition
1135
Charge Local Power Supply & Exchange Data With Host Device

FIG. 11

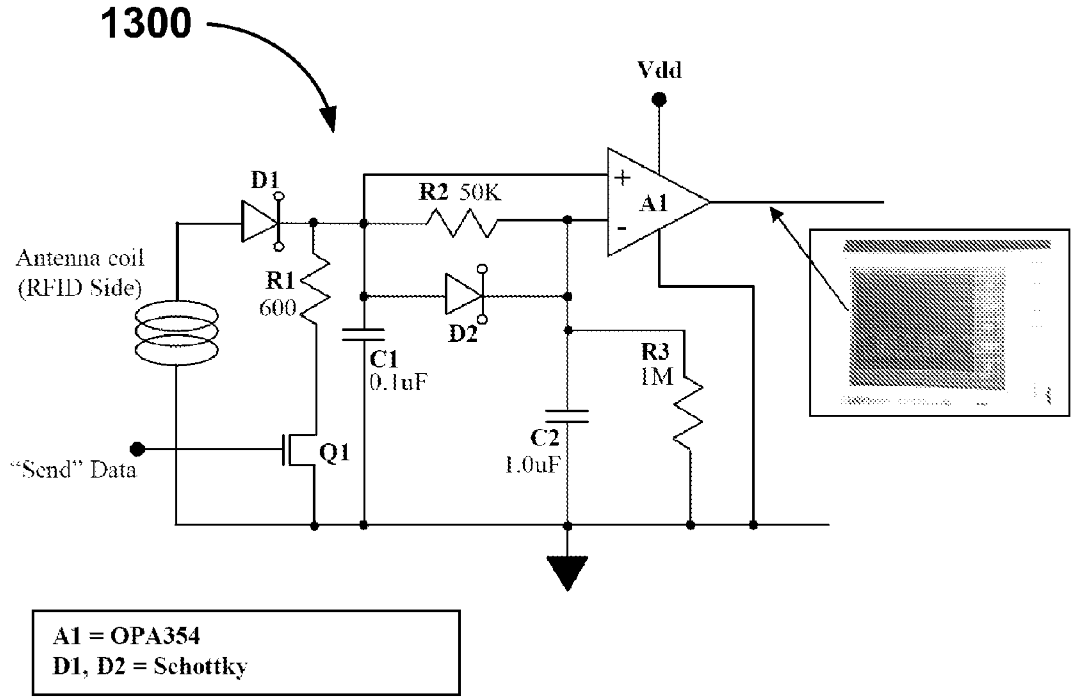
1300
Vdd
D1
R2 50K
+
A1
-
Antenna coil
(RFID Side)
R1
600
C1
0.1uF
D2
R3
1M
C2
1.0uF
"Send" Data
Q1
A1 = OPA354
D1, D2 = Schottky

FIG. 13

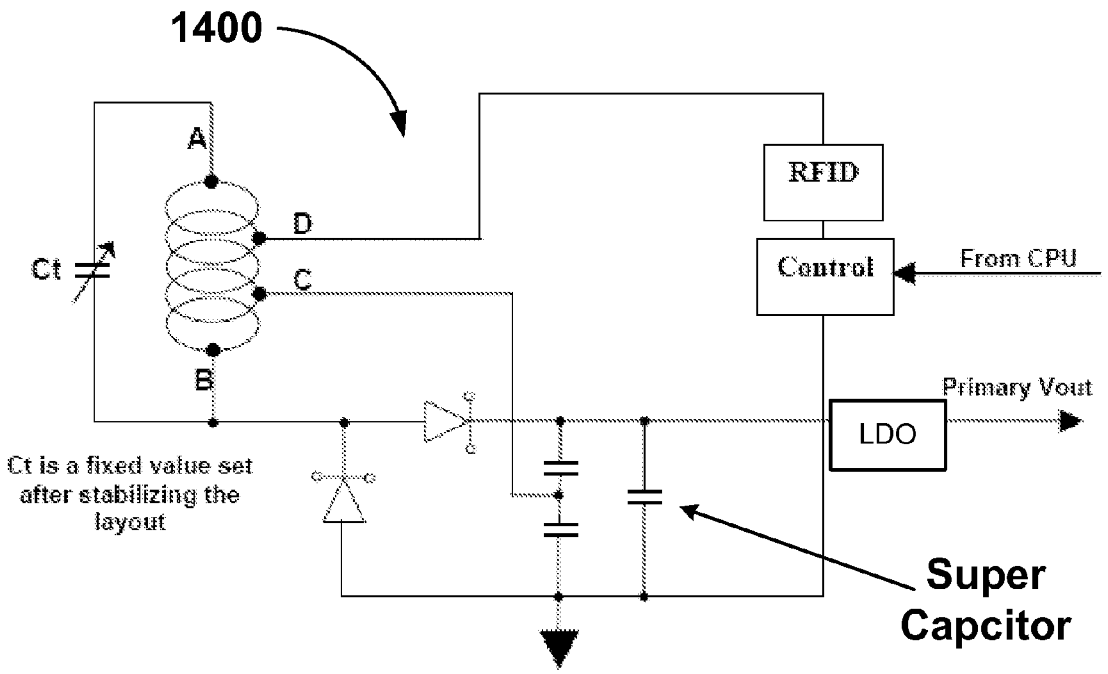
1400
A
D
C
B
Ct
RFID
Control
From CPU
Primary Vout
LDO
Ct is a fixed value set after stabilizing the layout
Super Capcitor

FIG. 14

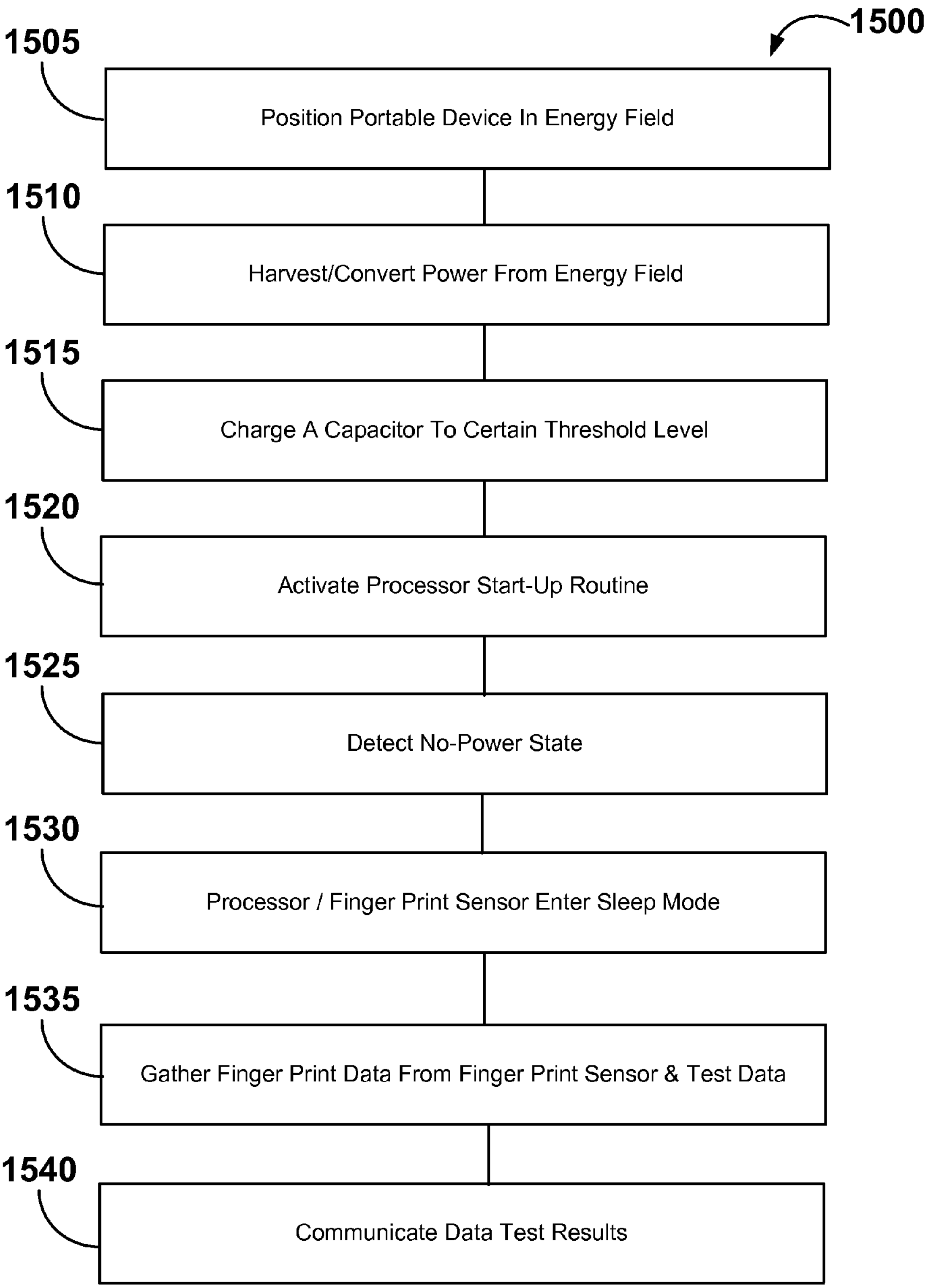
1500
1505
Position Portable Device In Energy Field
1510
Harvest/Convert Power From Energy Field
1515
Charge A Capacitor To Certain Threshold Level
1520
Activate Processor Start-Up Routine
1525
Detect No-Power State
1530
Processor / Finger Print Sensor Enter Sleep Mode
1535
Gather Finger Print Data From Finger Print Sensor & Test Data
1540
Communicate Data Test Results

FIG. 15

RF POWER CONVERSION CIRCUITS & METHODS, BOTH FOR USE IN MOBILE DEVICES

CROSS REFERENCE TO RELATED APPLICATIONS & PRIORITY CLAIM

[0001] This patent application claims the benefit of and priority to U.S. Provisional Patent Application No. 61/089,440, entitled RF POWER CONVERSION CIRCUIT, and filed 15 Aug. 2008, which is hereby incorporated herein by reference as if fully set forth below. Embodiments of the present invention may also utilize technology disclosed in U.S. Pat. No. 7,278,025 and PCT Application Publication Number WO 2005/104704; both of these publication disclosures are hereby incorporated herein by reference as if fully set forth below.

TECHNICAL FIELD

[0002] Embodiments of the present invention relate generally to portable verification devices, and more particularly, to a smart card having biometric data verification features and a dual purpose receiving antenna that can be used for wireless power transfer and data modulation.

BACKGROUND

[0003] RFID (Radio Frequency Identification) technology provides for near-field (short-range) wireless tracking of inventory as well as enabling users for near-secure access or transactions and other applications. The ISO-14443 specification defines near-field data communications between a reader device and one or more candidate devices. Candidate devices can include tags, badges, cards, or pocket devices commonly referred to as dongles, fobs, or smart cards.

[0004] When candidate devices are introduced into a reader devices' electromagnetic field, the candidate device detects the reader device's RF energy. A candidate device can then respond a data stream that modulates the energy field at the data rate. Candidate device(s) need not emit RF energy, but can provide a field load modulation that is detectable in the reader. A reader device can translate a candidate device's field load modulation into readable data (e.g., by using a microprocessor and supporting system).

[0005] Conventionally, a reader device emits RF energy at a frequency of 13.56 MHz. The energy field strength is specified at about 1.5 A/meter. Existing RFID systems are simple in that a candidate device's electronics may contain a secure code in the form of a sequence of up to 32 bytes. The candidate device responds to the entry of the RF field automatically with the data stream after introduction to the field in a process referred to as Answer to Reset (or ATR). The ATR data stream repeats while the candidate device is in the RF field. After the candidate device is removed from the field, the candidate device ceases to be active.

[0006] Traditional RFID candidate devices typically use a single track of coil windings to power the RFID-IC and respond with ATR data. This amounts to a current requirement that is typically on the order of about 5 to 10 mA during the ATR period which is less than two seconds. The design of conventional windings only delivers enough power to activate RFID transponder backscattering.

[0007] What is needed, therefore, are smart card systems enabling biometric data verification and using dual purpose antennas for wireless power transfer and data modulation. It is to the provision of such smart card devices, systems, and methods that the various embodiments of the present invention are directed.

SUMMARY OF EXEMPLARY EMBODIMENTS

[0008] Embodiments of the present invention provide a stand-alone, self-powered smart card without the use of a battery or a powered card-holding accessory. According to one embodiment, a receiving antenna is reformulated into a multi-purpose, multi-function component of the smart card. The receiving antenna is integrated with an efficient power detection and conversion circuit that produces a voltage and current suitable for powering the electronics of the smart card. As a result, the smart card operates without requiring power supplied by a battery or a powered card-holding accessory. In some embodiments, the present invention can be utilized to harness energy from a source of wireless energy for charging a local power supply. Antenna components used in embodiments of the present invention can be configured to have multiple segments. A segment can be used to increase rectified voltage, and another segment can be controlled to modulate antenna impedance. Taps or tap positions can be used to segment single antenna into multiple segments. Other embodiments of the present invention are summarized below.

[0009] In some embodiments, the present invention can be portable wireless devices used for event actuation. Generally, portable wireless device can generally comprise a wireless power harnessing module, a biometric data comparison module, and a communication module. The wireless power harnessing module can comprise an antenna tuned to a resonant frequency. The resonant frequency can be associated with a source of an energy field. The antenna can be tuned with a capacitor placed in parallel with the antenna. The antenna can comprise several windings. When the antenna is positioned proximate the energy field, the antenna can interact with the energy field to generate electrical energy. This enables the wireless power harnessing module to source power and provide power to other components.

[0010] The biometric data comparison module can be coupled to the wireless power harnessing module. This coupling enables the wireless power harnessing module to power the biometric data comparison module. The biometric data comparison can be configured to enter a powered state when receiving adequate power from the wireless power harnessing module. When in the powered-on state, the biometric data comparison module can be operatively configured to receive external biometric data. The external biometric data can be obtained from an external source. After obtaining the external biometric data, the comparison module can compare the external biometric date to stored biometric data. Stored biometric data can be stored in a flash memory. Results of the biometric data comparison can be communicated by the communication module. Communication can be done wirelessly with an RF chip in some embodiments. Based on the communicated results received at a host device, event actuation can take place.

[0011] Portable wireless devices of the present invention can also have additional features. For example, the wireless power harnessing module, the biometric data comparison module, and the communication module reside within an ISO-7816 defined card outline. The wireless power harnessing module can comprise a rectifier circuit. The rectifier can be coupled to a low impedance winding of the antenna and a common ground. The rectifier circuit can be configured to convert AC voltage provided by the antenna to DC voltage. Embodiments of the present invention can also include a capacitor located in parallel with the antenna. The relationship between the capacitor and the antenna defines the resonant frequency of the antenna. In some embodiments, portable wireless devices have no local power source.

[0012] Portable wireless devices of the present invention can still yet have additional features. For example, antennas can be divided up into segments. The segments can be segmented by taps disposed at various tap positions. Antenna segments enable a single antenna winding to have multiple segments configured to carry out multiple functions. For example, the wireless power harness module can harness power from the energy field simultaneously to the communication module transmitting and receiving data from the energy field. To do this, the two modules can be connected to the antenna at different tap positions to use different antenna segments. The antenna can be shaped in an antenna coil pattern wound in a concentric fashion that comprises inner and outer windings. And some embodiments, the antenna coil pattern can be a continuous planar copper trace having tap positions located at various places along the coil pattern so the antenna has multiple segments configured to have different functions. In some implementations, the wireless power harnessing module comprises a rectifier circuit as a voltage doubling circuit that comprises two Schottky barrier diodes arranged in a full wave rectifying arrangement.

[0013] Other embodiments of the present invention can be implemented as wireless access control devices. The device can generally comprise a power circuit and a processor. The power circuit can be configured to have a default non-energized state and an energized state. The power circuit can be configured to receive energy from an energy field to enter the energized state so that the power circuit can source electrical power. Preferably, the power circuit is finely tuned to a carrier frequency of the energy field. The processor is coupled to the power circuit to receive to receive electrical power when the power circuit enters the energized state. The processor can be further configured to receive data from a sensor. In response to the received data, the processor can generate a signal corresponding to an access level.

[0014] Wireless access control devices of the present invention can also have additional features. For example, the processor can receive power only from the power circuit when energized and the processor is not configured to receive power from any other power source. This allows embodiments of the present to not have a required battery for operations. In some embodiments, the power circuit can comprise a power detection stage, a power conversion stage, and a receiving antenna. The receiving antenna can be integrated with the power detection stage. The antenna can be shaped and sized to produce electrical power when placed into an energy field. The antenna can also be used to receive and transmit wireless data signals. Also the antenna can be finely tuned to the carrier frequency of the energy field. Tuning can be accomplished using a tuning capacitor in a tank circuit format. In addition, the processor can be configured to control data communication between the wireless access control device and the source of the energy field during the energized state.

[0015] Still yet other embodiments of the present invention can be implemented as portable wireless devices capable of harnessing wireless energy. The devices can include an antenna and a rectifier circuit. The antenna and a tuning capacitor can be connected in parallel to form a tank circuit. The tank circuit can be finely tuned to a resonant frequency associated with a carrier base frequency of source of an energy field. The antenna can have several windings which when proximate the energy field, result in the antenna sourcing electrical current and voltage. The antenna can be divided into a plurality of segments. The segments can set off by a plurality of taps. The taps can be disposed at various places along the length of the antenna. One of the segments can be configured to receive and transmit data with the energy field simultaneously as the antenna receiving energy from the energy field. The rectifier circuit can be connected to a first tap and a second tap of the antenna. The first tap can be located on an inner antenna winding. The second tap of the antenna can be in electrical communication with a common ground. The rectifier circuit configured to convert the sourced electrical current and voltage to a DC energy source.

[0016] Portable wireless device embodiments of the present invention can also have additional features. For example, devices can have an antenna driving circuit. Antenna driving circuits can be configured to drive the antenna for data communication. The antenna driving circuit can be connected to a third antenna tap. The third tap can be located on an outer antenna winding. Device embodiments of the present invention can also include a voltage divider network. The network can be a capacitor network coupled to the rectifier. The rectifier can comprise a pair of diodes (e.g., Schottky diodes). The cathode of a first diode can be connected to the anode of a second diode. The anode of the first diode can be connected to ground. The cathode of the second diode can be connected to the voltage divider capacitor network. The voltage divider capacitor network can comprise a first capacitor connected in parallel to two series connected capacitors. The cathode of the second diode can be connected to a positive terminal of the first and second capacitors. The anode of the first diode can be connected to a negative terminal of the first and third capacitors. Capacitors in the voltage divider capacitor range in value from about 1 pF to about 100 pF, tuning capacitors can range in value from about 10 pF to 500 pF, antennas can have between 1 to 10 coil windings, and coil windings can have a width ranging between about 1 mm to about 10 mm. In some embodiments, the voltage divider capacitor network can comprise an energy storage capacitor configured to store energy. The energy storage capacitor having a value ranging from about 0.5 micro-farads to about 1000 farads.

[0017] Still yet other embodiments of the present invention can be implemented as a method of harnessing electrical energy from an energy field while simultaneously communicating data with the energy field. The method can generally comprise configuring and/or providing a portable device and a processor. The portable device can have a tank circuit tuned to a center frequency of an energy field. The inductor (or antenna) of the tank circuit can interact with the energy field to convert wireless energy into electrical energy. This enables the inductor can to source electrical power. A processor can be located on the portable device to receive electrical power sourced by the inductor. The processor can be configured to receive and provide data for communication with a device emitting the energy field. Data can be received and transmitted using coils of the inductor while the inductor is sourcing energy.

[0018] Method embodiments of the present invention can also include other features. For example, methods can include configuring portable devices to receive external biometric data, to test the biometric data against a stored biometric set of data, and to communicate results of the test via the inductor. Methods can include configuring the processor to communicate data by modulating the field load of the energy field. Methods can also include providing a voltage conversion circuit on the portable device to convert the energy sourced by the inductor from AC to DC and to regulate the DC voltage relative to a predetermined threshold.

[0019] Still yet, embodiments of the present invention can be implemented as a computer program product embodied in a computer-readable medium for execution by a processor or engine. The computer program product can comprise one or more algorithms to manage actions carried out by a processor in managing power and testing biometric data. The method can generally comprise harvesting power, testing biometric data, and communications. More specifically, the method can detect an appropriate power level being sourced by an antenna that is finely tuned to resonate at a center carrier frequency of an energy field. The power level can be provided in electrical form after the antenna converts wireless energy to electrical energy. The method can also include communicating with a biometric sensor to determine if the sensor detects presence of biometric data and has captured external biometric data. The received biometric data can be tested against stored biometric data to determine if the captured external biometric data matches the stored biometric data. Methods can also include issuing communication signals for wireless transmission from the antenna to another component. The communication signals comprise data about results of the biometric data test.

[0020] Methods embodiments of the present invention can also include other features. For example, a method may include instructing one of the biometric sensor or a system processor to enter a sleep mode if a low power level state is detected or to preserve power. A method can also include testing received biometric data against stored biometric data includes by configuring a system processor to extract digital data from the captured external biometric data to place the external biometric data in the same format as the stored biometric data. Methods can also include testing received biometric data against stored biometric data by generating a score indicative of the data test and wherein the score can determine a positive or negative test result relative to a predetermined threshold. Still yet, methods can include testing received biometric data against stored biometric data by generating a false acceptance ratio and a false rejection ratio and wherein a match condition can be achieved with the false rejection rate is less than the false acceptance ratio.

[0021] Other aspects and features of embodiments of the present invention will become apparent to those of ordinary skill in the art, upon reviewing the following description of specific, exemplary embodiments of the present invention in conjunction with the accompanying figures. While features of the present invention may be discussed relative to certain embodiments and figures, all embodiments of the present invention can include one or more of the advantageous features discussed herein. In other words, while one or more embodiments may be discussed as having certain advantageous features, one or more of such features may also be used in accordance with the various embodiments of the invention discussed herein. In addition, while discussion contained herein may, at times, focus on insurance applications, embodiments of the present invention can also be used in other settings. In similar fashion, while exemplary embodiments may be discussed below as system or method embodiments it is to be understood that such exemplary embodiments can be implemented in various systems, and methods. It should be understood that use of the terms module, processor, or engine herein should be construed to mean singular or plural versions of these terms such that certain actions can be carried in separate fashion or integrated together in a single module, processor, or engine. Some embodiments of the present invention can be implemented with hardware and/or software.

BRIEF DESCRIPTION OF FIGURES

[0022] FIG. **1** illustrates a conventional RFID tag device with conventional tag circuitry.

[0023] FIG. **2** illustrates an RFID tag power circuit in accordance with some embodiments of the present invention.

[0024] FIG. **3** illustrates an RFID tag power circuit in accordance with some embodiments of the present invention.

[0025] FIG. **4** illustrates winding components of an RFID tag power circuit in accordance with some embodiments of the present invention.

[0026] FIG. **5** illustrates a schematic of an RFID tag power circuit in accordance with some embodiments of the present invention.

[0027] FIG. **6** graphically depicts an RF field in a proximate relationship with a smart card embodiment in accordance with some embodiments of the present invention.

[0028] FIG. **7** illustrates a block diagram of an RFID tag power circuit and biometric device in accordance with some embodiments of the present invention.

[0029] FIG. **8** illustrates a schematic of an RFID tag power circuit and biometric device in accordance with some embodiments of the present invention.

[0030] FIG. **9** illustrates a logical state diagram illustrating operational states of a biometric device in accordance with some embodiments of the present invention.

[0031] FIG. **10** illustrates a schematic of an alternative RFID tag power circuit and biometric device arrangement in accordance with some embodiments of the present invention.

[0032] FIG. **11** illustrates a functional block diagram of a power charging system in accordance with some embodiments of the present invention.

[0033] FIG. **12** illustrates a functional logic diagram showing a method of operating a power charging system in accordance with some embodiments of the present invention.

[0034] FIG. **13** illustrates a schematic diagram of a RFID transceiver module in accordance with some embodiments of the present invention.

[0035] FIG. **14** illustrates a schematic diagram of a RFID transceiver module circuit **1400** for use in charging applications in accordance with some embodiments of the present invention.

[0036] FIG. **15** illustrates a logical flow diagram **1500** of a method that can be used to implement embodiments of the present invention on a mobile device

DETAILED DESCRIPTION OF PREFERRED & ALTERNATIVE EMBODIMENTS

[0037] To facilitate an understanding of the principles and features of the various embodiments of the invention, various illustrative embodiments are explained below. Embodiments of the present invention may be described below with reference to RFID reader applications. The embodiments of the invention, however, are not so limited. Indeed, embodiments of the present invention can include any portable device having a default unenergized state that is capable of harnessing power from an energy field as discussed herein for use with biometric data verification. Other embodiments can be devices needing recharging of local power supply as such recharging can be accomplished by harnessing energy from a wireless energy field. Still yet, other embodiments can be used to harness energy from an energy field while enabling transceiving of data between a portable device and another device (which can be the source of the energy field).

[0038] Briefly described, in preferred form, an embodiment of present invention includes a portable device having a wireless power reception circuit capable of harnessing power from an energy field for supply to a biometric data verification stage. Upon receipt of power from the power reception circuit, the biometric data verification stage can receive biometric data and compare the received data to previously stored biometric data. The portable device can have a communication component to transmit a signal (or modulate an existing signal) that contains information about the results of the biometric data comparison. Advantageously, the portable device need not have an independent power supply since it can harness power from an energy field for use in conducting biometric data comparison.

[0039] Various words or phrases used herein at times have multiple meanings and should not be limited in certain instances unless expressly stated. For example, coupled can mean directly coupled or indirectly coupled. Also the phrase "in electrical communication" can mean that components are in the same electrical path or are electronically coupled together. In some instances, where specific advantages or features of an embodiment of the invention are discussed, it should be understood that such advantages or features can be applicable to the other various embodiments of the present invention.

[0040] Referring now to the figures, wherein like reference numerals in some instances represent like parts throughout the views, exemplary embodiments of the present invention are described in detail. FIG. **1** illustrates a conventional passive RFID tag device **100** with conventional tag circuitry **105** and an antenna **110**. The conventional RFID device **100** is tuned generally by design to receive energy from an RFID tag reader (not shown). The tuning is general in the sense that the antenna **110** is not tuned tightly to a specific frequency. For example, typically, RFID devices are tuned only by inductance, antenna features to about 17 MHz. The RFID tag device **100** is designed to only recover an RFID tag reader's magnetic field (H-Field) energy. Given the possibility that multiple RFID tag devices **100** can enter an RFID tag reader's energy field, the resonance of RFID tag devices is set to about 17 MHz (which is above the RFID tag reader's carrier frequency). This is purposefully done to enables the processing of multiple cards in close proximity within and RFID reader's RF field.

[0041] As pictured in FIG. **1**, the conventional passive RFID tag device **100** includes an antenna **110**. The antenna **110** is 3 turns of wire closely wound in a continuous, uninterrupted fashion. Electrically, this antenna **110** may be modeled as the secondary coil of an air core transformer. Energy is collected by the RFID tag device **100** and is used only for the short ART transmit period. General considerations for this antenna coil are for lower than optimum "Q" and loose tuning slightly above a 13.56 MHz frequency. As mentioned above, the loose tuning allows for multiple cards in an RFID's RF field and the detuning that occurs in that event. A typical RFID transponder (like the RFID device **105**) will use approximately 25 mW during the short transmit period. Because of the specific design of the conventional passive RFID tag device **100**, the device **100** is unable to harness sufficient amounts of RF energy for sourcing power (i.e., generating voltage and power) to adequately power electronics more complex than a simple RFID transponder (like the RFID device **105**).

[0042] FIG. **2** illustrates a functional block diagram of portable device **200** used for event actuation in accordance with some embodiments of the present invention. The device **200** can be formed in the shape of a card **205** in some embodiments. In other embodiments, the portable device **200** may be a fob, dongle, PDA, cell-phone, smart phone, computer, or many other portable devices. The device **200** may include a local power source (e.g., battery) in some embodiments, and in other embodiments, the device **200** may not include a local power source. In those embodiments without a local power source, the wireless power harnessing module **210** is configured to harness wireless energy sufficiently to power electronic circuitry more complex than a simple RFID transponder.

[0043] According to some embodiments, the device **200** can generally include a wireless power harnessing module **210**, a biometric data comparison module **215**, and a communication module **220**. In the embodiment pictured, the modules **210**, **215**, **220** can be coupled to each other to function and work together. In other embodiments, these modules **210**, **215**, **220** may be integrated together such that the functions of one or more modules can be combined in a single module. In smart card embodiments, it is currently preferred that the modules be sized and shape to fit within a card having sizes a defined in the ISO-7816 standard. Desired thicknesses range between about 0.7 mm to about 1 mm.

[0044] In embodiments with no local power source, the wireless power harnessing module **210** can be configured to recover energy from an energy field. The energy field can be, for example, an RF field emitted from a device (e.g., an RFID card reader). The wireless power harnessing module **210** can include an antenna having multiple coils windings of a conductor. Preferably the coil windings are planar in shape. As discussed further herein, the coil windings can be tapped at various places so that an antenna has multiple functions. Various tap points can be disposed on the antenna so that the antenna is a non-continuous, interrupted winding (as opposed to that shown in FIG. **1**). This configuration enables the antenna to dually function as for power recovery and data transmission. By virtue of being placed in an energy field, the antenna can generate a current thereby harnessing wireless energy for use by the biometric data comparison module **215** and communication module **220**.

[0045] The biometric data comparison module **215** can be configured to compare received external biometric data to stored biometric data. As such, the biometric data comparison module **215** can include a memory (e.g., flash memory) to store biometric data. In some embodiments, the stored biometric data can be a digital rendering of someone's fingerprint. The biometric data comparison module can also include a sensor (or other interface) to receive external biometric data. For example, the sensor can be a fingerprint sensor in some embodiments. When a finger is placed on the sensor, the sensor can capture external fingerprint data. The biometric data comparison module can also include a processor to receive captured external fingerprint data. The processor can be configured to compare the captured fingerprint data to stored fingerprint information. The results of the comparison can be provided as a score. If the score is above a certain threshold, then a match can be determined, and if the score is below a certain threshold, then a non-match can be determined.

[0046] Based on results of the comparison, the processor can instruct the communication module **220** to communicate information to a reader. Information can be communicated via a load modulation (or backscattering) protocol. If it is determined that a match occurred, the communication module can send this information to another device, and in response the device can actuate an event. For example, in the case of an access card, if a fingerprint match has been determined, then an RFID reader can send a signal to allow access.

[0047] It should be understood that embodiments of the present invention are not limited to access cards or access devices. For example, the device **200** can a fob, cell phone, smart phone, computer, dongle, or many other portable devices that may need power for functionality. In addition, the device **200** can be used for multiple applications. For example, the device **200** may used to authenticate a user prior to event actuating, including use of electronic devices and starting of vehicles. In other embodiments, the device **200** can be used as a source of power since it can harness power from wireless RF. The source of power may be used to charge an electronic device according to some embodiments.

[0048] FIG. **3** illustrates a bio-verification card **300** in accordance with some embodiments of the present invention. As shown, the bio-verification card **300** generally comprises an antenna **310**, a voltage detector/converter **315**, and a variable capacitor **320**. The bio-verification card **300** can be finely tuned to an energy field's center frequency. The tuning can be accomplished using the variable capacitor **320** to tune the antenna **310**. In some embodiments, the variable capacitor **320** can be a fixed capacitor assuming a used center frequency is used. For example, if an energy field has a center frequency of 13.56 MHz, the variable capacitor **320** can have a fixed value ranging from between about 5 pico-farads to about 30 pico-farads. When implemented, the capacitor can have a fixed value to finely tune an antenna to a specific frequency to that the sensitivity of the antenna matches with the energy field to create a resonance event.

[0049] By virtue of finely tuning the antenna **310** to a specific frequency that substantially matches an energy field's center frequency, maximum energy from the energy can be recovered. In some embodiments, only one the bio-verification card **300** is placed in an energy field at any given time. In such a case, the antenna **310** coils' outer turns are resonant in the energy field's electrical field (aka E-field or energy field). Resonance is achieved by a parallel inductor/capacitor (L-C) combination (e.g., the antenna **310** and capacitor **320**) which emulates an end-fed, monopole element. The resonance frequency and appropriate L/C values can be obtained using the resonance equation: resonance frequency (f) is equal to the inverse of 2 times Pi times the square root of L times C—$f=1/(2\pi\times\sqrt{(L\times C)})$. The antenna **310** configuration shown in FIG. **3** provides a transition in the coil's structure from electrical to magnetic when moving from the antenna's **310** outer turns toward the antenna's **310** inner turns. The innermost winding is a single-low-impedance winding that is the voltage source for the voltage detector/converter **315**.

[0050] FIG. **3** also illustrate various tap positions **325**, **330**, **335** being disposed in the antenna configuration. Placement of multiple taps in this illustration (and as described in other illustrations herein) enables a single antenna structure to be multi-functional. The tap positions break the single antenna into multiple antenna segments. This enables space savings within a confined area when multiple antennas can not be utilized (e.g., in a smart card application). Tap position **325** is located at the end of the innermost antenna winding, tap position **335** is tied to a common ground, and tap position **330** is located at the end of the outermost winding. The voltage detector/converter **315** can be disposed between tap positions **325**, **330** and the variable capacitor can be disposed between tap positions **330**, **335**.

[0051] FIGS. **4** illustrates an antenna arrangement **400** used for harnessing power of an energy field in accordance with some embodiments of the present invention. As shown in arrangement **400**, an antenna winding **405** is wound close to the outer periphery of a confined space **410** (e.g., internal area of a smart card). The antenna winding **405** comprises four windings. In other embodiments the antenna windings **405** can have between 2 and 10 windings. Other winding values are also possible in accordance with the present invention.

[0052] Also as shown, the antenna winding **405** has a plurality of tap positions. The tap positions can be placed at various locations along the antenna winding **405** to interrupt the continuous flow of the antenna winding **405**. Various tap positions also enable access to the varying impedance of the winding **405**. As shown, tap A is located at the end of the outermost winding, tap B is located at the end of the innermost winding, tab C is located at a position on the second innermost winding, and tab D is located at a position on the winding closest to the outer winding. In this arrangement, the outermost winding is the high impedance winding with the innermost winding being a low impedance winding. Although taps A, B, C, and D are located in these positions in this embodiment, various other embodiments could have various tap positions along the antenna winding.

[0053] The windings **405** can have various characteristics in accordance with the various embodiments of the present invention. For example, the windings **405** can have a planar shape having a thickness ranging between about 10 microns to about 100 microns. In currently preferred embodiments, the thickness of the windings **405** can range between about 13 to about 60 microns. In addition, the windings **405** can be arranged so that no sharp turns are provided in the windings **405**. As shown in FIG. **4**, the windings **405** are configured to have smooth transition between segments. In currently preferred embodiments, angular transitions have angular turns about 45 degrees or less. Also, the windings can be made of various conductive metals or metal alloys. In some embodiments, the windings can be made with substantially pure copper traces. In other embodiments, the windings can be made with copper foil, stamped copper, etched conductors, copper plating, milled copper, pressed copper wire, silver, and aluminum.

[0054] Now turning to FIG. **5**, there is shown a schematic diagram of a power recovery/conversion circuit **500** in accordance with some embodiments of the present invention. The circuit **500** generally includes three modules: a power harnessing module **505**, a power conversion module **510**, and a control RFID module **515**. The circuit **500** can also be configured to provide an output voltage (V_{OUT}) **520** and receive a control signal **525** from another component. The output voltage **520** can be provided from the interaction between the power harnessing module **505** and the power conversion module **510**.

[0055] As shown, the power harnessing module **505** includes a capacitor **530** and an antenna **535**. The capacitor **530** can be variable (as illustrated) or fixed at a certain value. The value of the capacitor **530** can be selected to tune the antenna **535** such that its winding can resonate at a certain frequency. The resonation frequency can be an energy field's center carrier frequency. Resonation enables maximum power transfer from an energy field to the antenna's **535** windings. As shown, the antenna's windings **535** can be tapped at various locations (similar those in FIG. **4**). Potential from the taps B, C, D can be provided as inputs to power conversion module **510** and the control/RFID module **515**.

[0056] When the antenna **535** encounters an energy field (e.g., an RF field) a current is generated thereby creating an AC voltage. This AC voltage can be accessed at tap B and provided to the power conversion module **510**. The power conversion module **510** includes a rectifier **540** to convert the AC voltage to a DC voltage. The rectifier **540** includes two diodes coupled in a full wave arrangement. In currently preferred embodiments, the diodes are Schottky diodes. This type of diode enables effective harnessing of power from high frequency energy fields.

[0057] The diodes can also be coupled to a capacitor network **545**. As shown, the capacitor network **545** can include two series capacitors in a parallel arrangement with a single capacitor. The capacitor network **545** can also be arranged in other configurations. The capacitor network **545** can filter the converted DC voltage. The capacitor network **545** can also include a super capacitor (ranging from 2 Farads to 10 Farads). In this arrangement, the circuit **500** can be used as fast charging device using only source energy provided from a wireless field.

[0058] The power recovery/conversion circuit **500** also includes a control/RFID module **515**. The module **515** can be used to communicate with a device that provides an energy field (such as an RFID reader). Communication can be done via load modulation (also known as backscattering). In this arrangement, embodiments of the present invention (e.g., the circuit **500**) can be used to advantageously simultaneously receive/convert power and transmit/receive data.

[0059] FIG. **6** graphically depicts an RF field in a proximate relationship with a portable device (e.g., smartcard) in accordance with some embodiments of the present invention. Projected at approximately 90 degrees from the reader surface, the RF energy field may be described as a "dome" of electromagnetic energy having a frequency of 13.56 MHz. Both electrical (E) and magnetic (H) fields are present; however the dominant field is magnetic. For ISO-14443, the typical field intensity is on the order of 1 Amp/meter. A typical RFID device uses 10 mA in the two second period (approximately) needed to detect and respond to reset conditions with the secured data stream. The dome of energy may extend for several inches from the reader surface. The reader's inductor or antenna is typically configured to permit a field height and radius of equal proportions. The typical dome provides useful energy within a dome of approximately 1 to 2 inches high. FIG. **4** depicts a model of the reader, the projected field, and a card in the field.

[0060] FIG. **7** illustrates a functional block diagram of a biometric device **700** in accordance with some embodiments of the present invention. As shown, the biometric device **700** can be sized and shaped as a card **705** (e.g., an access card or a smart card). In other embodiments, the biometric device **700** may be sized in shaped in other configurations. In some embodiments, the biometric device **700** may be integrated with other devices or a host device. These can include devices, such as fobs, dongles, cell phones, smart phones, computers, personal communication devices, and the like. When integrated with a host device, the biometric device **700** can be used to secure or enhance secure access to the host device.

[0061] As illustrated in FIG. **7**, the biometric device **700** can comprise various components. The components can include an antenna **710**, interface pads **715**, a processor or microcontroller **720** (CPU), a power circuit **725**, an RF chip **730**, and a biometric sensor **735** (e.g., a finger pint sensor). As illustrated, the biometric device **700** does not include a local power supply (in other embodiments, the biometric device may include a lower power supply, such as a battery or solar cell system). By not having a local power supply, the biometric device **700** can be arranged and confined to a small space.

[0062] Even though the biometric device **700** lacks a local power supply, the biometric device **700** is equipped with features capable of harnessing power from an RF energy field. The antenna **710** can be used to harness wireless energy for use as a power source. For example, the antenna can receive RF energy and convert the RF energy to AC power (i.e., and AC voltage and current). This AC voltage can be converted to DC using the power circuit **725**. The power circuit **725** can provide t his DC power (i.e., DC voltage and current) to the various other components. For example, the power circuit **725** can provide DC power to the CPU **720**, the RF chip **730**, and the fingerprint sensor **735**. As shown in FIG. **7**, the power circuit **725** can be electrically coupled to the CPU **720**, the RF chip **730**, and the fingerprint sensor **735**.

[0063] In operation, the biometric device **700** can be configured to authenticate a user's finger print to actuate an event (such as entry access). To implement the authentication ability, the CPU **720** can have a memory and biometric data (e.g., a finger print template) can be downloaded into this memory. This can be done via the interface pads **715** in some embodiments. The biometric data can be associated with one or more users. In currently preferred embodiments, the biometric data is a finger print for a unique user. The finger print data can be a digital representation of the finger print and can be stored as a fingerprint template.

[0064] Using the biometric device, the unique user can position the device close to a source of RF energy, such as an RFID reader. Typically, RFID readers are located near entryways to restrict access. When a user with the biometric device approaches the RFID reader, the antenna **710** harvests energy from the RF field, the power circuit **725** converts the harvested power to DC, and then the DC voltage is distributed for use.

[0065] When receiving power, the fingerprint sensor **735** activates and captures external finger print data. The external finger print data is provided to the CPU **720**. The CPU **720** compares the external finger print data to the stored fingerprint template. Based on the comparison, the CPU **720** can calculate a comparison score. The CPU **720** can then contrast the comparison score with a set threshold to determine if a match or no match condition has occurred. The threshold can be adjusted to ensure sensory integrity. If the CPU **720** determines that the external fingerprint data matches to stored template, the CPU **720** can proceed to take steps to communicate this information. For example, the CPU **720** can signal the RF chip **730** to generate a signal for wireless transmission by the antenna **710**. The RF chip **730** may not necessarily send a signal; rather it may modify an RF reader's energy field (via load modulation/backscattering) with a certain data modulation pattern. Upon detecting the modulation of its energy field, the RF reader can then actuate an event. This event actuation can include such things as unlocking the door to an entry way or sending a start signal to another device.

[0066] All necessary power for capturing external fingerprint, calculation of matching, RF chip power, or an optional LED/display on the card is supplied from RF power through antenna. Exemplary standards for the RF energy field can include, but are not limited to, ISO 14443 A/B/C, ISO 15693, Mifare, and Felica. Depending on the communication protocol, an energy field can have a certain carrier frequency. In some instances, the carrier frequency can be 13.56 MHz. To achieve maximum power transfer, the antenna **710** can be fine tuned to resonate at an energy field's carrier frequency. For example, the antenna **710** can be tuned with a capacitor to resonate at 13.56 MHz so that the antenna **710** can maximize energy harvesting from the energy field.

[0067] The biometric device **700** can enhance and improve upon legacy access card systems. In certain security applications, many use wireless door access cards. Legacy cards and card systems, however, have no functions to authenticate card holders. This deficiency results in a weakness of legacy card situations: cards can be stolen, faked, or replicated by fraudsters. This activity can lead to unauthorized access which can result in criminal activity. Embodiments of the present invention address the weakness of legacy card systems. In particular, embodiments of the present invention authenticate card holders.

[0068] Embodiments of the present invention also enable non-battery card systems. If batteries are used in cards, there is always a risk of running out of battery power in the battery and thus at an important event losing battery power can cause serious problems. Power supply environment has to be always guaranteed as long as power on the reader is guaranteed. Embodiments of the present invention are designed to have a low dissipating power system and utilize efficient energy acquisition through a novel tuned antenna design (as discussed herein).

[0069] FIG. **8** illustrates a schematic of a biometric device **800** in accordance with some embodiments of the present invention. This figure illustrates details of a RFID tag power circuit **805** (such as power circuit **725**). This figure also shows how the RFID power circuit **805** connects with antenna **810** at various tap positions.

[0070] As discussed herein, embodiments of the present invention can utilized a single antenna having various taps position along a single wound conductor. The various tap positions enable a single antenna winding to be multi-purpose: power harnessing and data transmission. As shown in FIG. **8**, the antenna **810** has four taps: A, B, C, and D. Tap A is positioned at the terminal end of the innermost winding, tab B is located at a corner position of the second innermost winding, tap C is located at a corner position of the third innermost winding, and tap D is located at the terminal winding of the outermost winding. By virtue of placing taps B and C between taps A and D, the single antenna **810** is divided into segments. The segments enable the single antenna to have multiple functions.

[0071] Shown connected to the various taps in the drawings are various logical devices and circuit components. For example, the antenna coil is terminated with a ceramic capacitor C**3** at taps A and D. C**3** can be used to tune the antenna to a certain frequency (e.g., 13.56 MHz±1 Mhz). The certain frequency that antenna is tuned to can be the center frequency of an energy field's carrier wave. In currently preferred embodiments, C**3** has a value ranging from 10 pF to 30 pF. The value can be more precisely 15 pF in some embodiments.

[0072] Tap C can be connected to the RF chip **830** so that the RF chip **830** can use the antenna for communication. In this arrangement, the RF chip **830** can generate signals for transmission using the antenna **810**. In addition, the antenna can be used to receive data (e.g., see FIGS. **13**-**14**). As a result, the single antenna **810** can be used to harvest power and communicate (receive data and transmit data). Data transmissions can be carried out by emitting wireless signals or modification of an energy's field load.

[0073] Certain of the antenna's **810** taps can be connected to devices to aid in harnessing power from an RF energy field. For example, and in currently preferred embodiments, tap A can be coupled to a rectifier. The rectifier can include two diodes: Schottky diodes D**2** and D**3**. Tap B can be coupled to the interconnection of C**1** and C**2**, with C**2**'s other terminal being tied to ground. This configuration enables the rectified voltage to be regulated by a voltage regulator **840** (e.g., a Low Drop Out (LDO)). The voltage regulator **840** can be more than 6 volts. This rating is high enough to supply an output voltage of 3.3 VDC. This output voltage can be utilized by logical/digital devices, such as CPU **820**. In currently preferred embodiments, C**1** and C**2** can be in the range of 1 micro-farad to 100 micro-farads. In some embodiments, C**1**=C**2**. Using the illustrated antenna and rectifying circuit, the LDO **840** can supply about 3.3 volt/50 mA to CPU **820** and a fingerprint sensor **835**. Preferably, the CPU is rated to consume between 30-40 mA at 60 MHz clocking operation and the fingerprint sensor **835** consumes 7-10 mA during the finger print capturing process.

[0074] When the biometric device **800** is positioned proximate an energy source (e.g., an RFID card reader), the device will begin to operate (FIG. **9** explains additional operational state details). When operations initiate, the CPU **820** can signal the finger print sensor **835** to capture the fingerprint of a card holder. In response, the finger print sensor **835** captures finger print data. The captured finger print data can be sent to the CPU **820**.

[0075] Upon receiving the captured finger print data, the CPU **820** can act on the data. The CPU **820** may take a digital rendering of the data or extract a simplified image from the raw, scanned finger print data. After acting on the captured finger print data, the CPU **820** can compare the extracted image to previously stored finger print data. To enable effective comparison, the stored finger print data and the captured finger should be obtained by the same method (e.g., same digital rendering algorithm or same data extraction method). Other finger print data simplification methods include but are not limited to thinning, noise removing, rotations, extracting Minutiae, and FFT (Fast Fourier Transfer).

[0076] To implement the comparison of the two data sets, the CPU **820** can implement a matching algorithm. The matching algorithm can retrieve the stored finger print data from memory and compare with the received external data. Results of the comparison may produce a comparison score. After obtaining the comparison score, the CPU **820** can determine if the score exceeds or falls below a predetermined threshold. In some embodiments, a comparison score that exceeds the threshold indicates a match condition and a comparison score that falls below the threshold indicates a no match condition.

[0077] After determining the existence of a match or no match condition, the CPU **820** can initiate control of the RF chip **830**. For example, upon confirming a match condition, the CPU **820** holds the register of an IO port to output a signal to enable Q1 to drive the antenna **810** at tap C (tap C can be located at roughly the center of the whole antenna **810**). By controlling the output signal to Q1, the CPU can toggle Q1's gate thereby turning Q1 off and on. This off and on modulates antenna transmission. The toggling activity, thus, enables the RF chip **830** to modulate data transmitted by the antenna **810**.

[0078] While the CPU **820** is controlling Q1, the CPU **820** can hold its IO port. When doing this, the CPU **820** can reduce its clock cycle to induce a sleep mode or a low frequency clock mode. When entering a sleep mode, the CPU **820** can also instruct the finger print sensor **835** to enter a sleep mode. By entering a sleep mode, the CPU **820** and the finger print sensor consume less power thereby preserving power for other components.

[0079] By operating in this fashion, the biometric device **800** can obtain full power from a wireless energy source. This full power can be initially used by the CPU **820** and the finger print sensor **835** to focus on calculations. Then the device **800** can focus on sending signal data via RF chip **830**. In testing of a prototype device, using a normal reader for one-fingerprint sensor ISO 14443A wireless card reader, a communication distance of **30**mm has been confirmed. This distance is the same distance of normal ISO14443A card communication distance. So, even though there are many power hungry components on the biometric device **800**, the biometric device **800** can communicate with the same reader at the same distance allowance.

[0080] The biometric device **800** can have various physical characteristics. For example, the antenna **810** is preferably made on a flexible PCB. The antenna's windings can be fabricated with planer copper traces. The antenna **810** preferably shares the same flexible PCB sheet with various other components and includes copper couplings to these other components. The other components can include the CPU **820**, the RF Chip **830**, a fingerprint sensor **835**, and a voltage regulator **840**. The flexible PCB sheet can be fabricated with, but no limited to, polyimide, mylar, PET, and kapton. The antenna's **810** windings can be made of laminated copper, plated copper, printed silver, combinations thereof, and many other conductive materials.

[0081] The antenna **810** can also have various other characteristics. For example, the antenna **810** coil can be made in a wound coil pattern. The wound coil pattern can be done so that a coil has a plurality of individual windings. The individual windings can have a thickness between about 10 microns to about 100 microns. The individual windings can also have a width ranging from about 50 microns to about 200 microns. Currently preferred embodiments have a width of about 100 microns with a thickness ranging between about 25 microns to about 35 microns. Thickness values should be selected to provide antennas having desired resistivity values. Such pattern can be patterned on one side of FPCB or both side of FPCB. Currently preferred coil winding embodiments include five windings with five turns. The windings can be positioned proximate the outer periphery edges of an access card. This configuration advantageously enables acceptance of a maximum magnetic flux from an RFID reader's energy field. The antenna **810** can be coiled so that the antenna **810** has angular turns less than about 45 degrees to limit eddy currents.

[0082] FIG. **9** illustrates a logical state diagram **900** illustrating operational states of a biometric device in accordance with some embodiments of the present invention. Generally, the several states show various operational stages of a portable device. A first state **905** shows a portable device in an initial state with no power, a second state **910** shows a portable device in range of an energy field at full power, and a third state **915** shows a portable device in range of an energy field with reduced power use to focus on data communication. Each of the states is discussed below in more detail with reference to an access card application. It should be understood, however, that the dual power harnessing and data communication states could be used in various other applications, including but not limited to, cell phone charging/data updating, smart phone charging/data updating, computer charging/data updating, personal music player charging/data updating.

[0083] Turning now to state **905**, an access card is in an initial no-power state. In this state, the card is likely outside the range of an energy field. As a result, the access card's antenna can not harvest any wireless energy. Access cards in this state will likely remain in the initial, no-power state until brought in the range of an energy field source. A no-power state could occur when multiple, fine-tuned cards are placed in close proximity of an energy source. Typically, in this situation, the card closest to the energy source pulls power from the energy source while those cards further away receive little to no power due to the existence of the closer power. The status information shown in state **905** indicated the existence zero volts and amps in the initial state.

[0084] A next state is shown as state **910**. State **910** can result when an access card is brought within range of an energy field (e.g., see FIG. **6**). When this occurs an access card's antenna and power conversion circuit can recover and harvest power from the energy field. This power can then be provided to electronics within the card. The electronics can include a processor and a fingerprint sensor. The processor and the finger print sensor can be used to scan a user's finger print and compare the scanned finger print against a known finger print. This procedure can authenticate someone holding an access card. Advantageously, this enables embodiments of the present invention to authenticate a card holder to ensure the card holder is properly associated with an access card. Access cards may not need to remain in a full energized state (such as state **910**) to carry out its functions. Indeed, to preserve energy and efficiently use harvested power, a process can be configured to switch on and off other components. An example of this is shown in state **915**.

[0085] State **915** illustrates a feature of some embodiments of the present invention, where certain components are switched off or instructed to enter a sleep mode. By instructing components to enter a sleep mode, power usage is minimized and or focused for use by other components. As shown in state **915**, the processor provides a sleep mode signal to the finger print sensor. When the finger print sensor is in sleep mode, the CPU can then direct adequate power to an RF chip.

When powered, the RF chip can communicate with a card reader. State **915** also represents the ability to continuously receive and harvest power from an energy field and at the same time, communicate with an access card reader. In currently preferred embodiments, communication can be accomplished via field load modulation.

[0086] FIG. **10** illustrates a schematic of an alternative RFID tag power circuit and biometric system **1000** in accordance with some embodiments of the present invention. In this system (which is similar to that shown in FIG. **8**), multiple processors are used and an RF chip is not used. In addition to CPU **1020**, a combination security CPU (Combi CPU) **1030** is used. The combination security CPU **1030** can be used for smart card embodiments and is capable of handling ISO7816 and ISO14443 wireless interface protocols. Further employment of the combination security CPU **1030** enables data transmission from ISO 7816 section to ISO 14443 section. Normally the ISO 14443 section is activated automatically when voltage (Vcc) to Combi CPU is off

[0087] In operation, system **1000** functions similar to the biometric device **800** (FIG. **8**). Upon a finger print match, however, the CPU **1020** gives power Vcc to the Combi CPU **1030** through IO**3**. The power can be 3.3 V 5 mA. At the same time, CPU **120** can output voltage from IO**3** to to enable Q1. Enabling Q1 allows data to be sent from from IO**2** (ISO7816 Protocol) to ISO7816 IO of Combi CPU **1030**. The data can be card holder name, matching result as the basic data and for security, send CPU ID, sensor ID and card UID or previous communication record, where all communication can be encrypted by such PKI.

[0088] Within Combi CPU, ISO7816 portion write the date in the shared memory of combi CPU, where shared memory can be read by ISO 14443 section and send such read data through antenna, when Vcc to Combi CPU is disconnected. Then upon all necessary data is transmitted from CPU to Combi CPU, then makes IO **3** to be floating. This enables Combi CPU to send data by reading the data in the shared memory written by ISO7816 portion trough antenna. At the same time, by reducing clock to CPU and making finger print sensor to be sleep mode, power consumption of the card is minimized as the ISO 14443 section of security CPU is only active. Regarding the power dissipations, this situation is almost same situation of normal ISO 14443 card and thus the invention can have the similar distance or normal ISO 14443 card, even though it contains intelligence and security.

[0089] FIG. **11** illustrates a functional logic diagram showing a method **1100** of a mobile (or portable) device being used in an energy field for energy harvesting and data exchange. At **1105**, a device is positioned in an energy field. The device can be a portable communication device or a portable access device. The energy field can be provided by any device capable of emitting or giving off an energy field. The energy field can be an RF energy field in some embodiments. In other embodiments, the energy field can higher or lower frequencies. In currently preferred embodiments, the device can be configured to interact with the energy field for multiple purposes.

[0090] As shown at **1110**, the method **1100** can also include harvesting and converting power from an energy field. This can be accomplished by configuring a portable device to convert wireless energy from the energy field into electrical power. For example, a portable device can include an antenna (as described herein) capable of interacting with an energy field to generate electrical current and voltage. The antenna can be sized and shaped to fit within a small area, like an access card. And in other embodiments, the antenna can be located within a portable communication device. When receiving power from an energy field, the power can be converted from AC to DC; DC power can be used to power both analog and digital devices. The AC can also be used to receive and transmit data.

[0091] When brought into an energy field and power transfer occurs, the method **1100** can also include at **1115** detecting receipt of power and an initialization procedure. In some embodiments, non-powered components can be in a sleep (or pause) mode until power detection occurs. Upon detection, for example, a processor can be configured to determine that adequate power is being provided and if so, enter an initialization procedure. The procedure can include ramping up of processor clocking speeds and signaling other components.

[0092] In some embodiments, a wake up routine can include a processor being configured to communicate with other components. For example, at **1120**, a processor can initiate operations of a biometric data sensor. The data sensor can check for presence of biometric data and capture biometric data. In currently preferred embodiments, biometric data sensors include finger print sensors. Other types of sensors can also be used.

[0093] After capturing biometric data with a sensor, the method can include testing of the captured data at **1125**. For example, the captured data can be compared with known data for authentication purposes. The comparison can result in a score which can be compared against a threshold. Results of the comparison against the threshold can result in a match or no-match condition.

[0094] The method **1100** can also include taking action on a determined match or no-match condition. For example, at **1130**, the method **1100** can include communicating the results of the data comparison. The communication can include an RF chip sending a wireless signal about the data comparison. The communication may also include modulating an existing energy field (e.g., field-load modulation). The communication can be in full duplex mode between a host/base device and a portable device.

[0095] Communication may occur simultaneously with other method actions. For example, at **1135**, while communication exchanges are occurring, power harvesting can be done in a manner to charge a local power supply. Harvesting wireless energy can result in doing away with physical cables/conductors normally required for power harvesting.

[0096] FIG. **12** illustrates a functional block diagram of a power harvesting/charging-data transmission system **1200** in accordance with some embodiments of the present invention. Generally, the system **1200** contains a host device **1205** and a portable wireless device **1210**. The host device **1205** can include many devices capable of sourcing an RF energy field and capable of receiving/detecting data modulations in an RF data field. The portable wireless device **1210** can include many portable devices capable of interacting with an RF energy field.

[0097] In some embodiments, the host device **1205** can be used to charge a power source (e.g., a battery) associated with the portable wireless device **1210**. For example, if the portable wireless device is a portable communication device, such as a cell phone or smart phone, having a battery, the portable communication device can include an RF power harnessing circuit as discussed herein. By harvesting the host's device RF energy field, the portable wireless device

1210 can charge its batteries. Since the charging can take place wirelessly, the need for charging devices or power cables does not exist. Charging times can range from fractions of seconds to multiple minutes.

[0098] In addition to enabling the charging of local power source, the portable wireless device **1210** can be equipped with circuitry to receive and transmit data from the host device's **1205** RF field. By being able to simultaneously transceive data and charge, the portable wireless data can share data with a network connected to the host device. Data exchanges can be accomplished at varying rates. For example, data exchange rates can include 106, 212, 424 and/or 848 kbit/s.

[0099] FIG. **13** illustrates a schematic diagram of a RFID transceiver module **1300** in accordance with some embodiments of the present invention. As shown, the module **1300** includes various analog and digital components. The antenna coil can interact with an energy field to generate AC power. The AC power can be fed to a rectifier (diode D**1**) for DC conversion. The converted DC can be provided to an op amp A**1** (e.g., Texas Instruments Op Amp OPA354 family) as an input source. The op amp can be configured as a comparator and use the converted DC as an input signal. The op amp also has as a reference input a floating reference. The floating reference is provided by a second diode (D**2**). The second diode D**2** allows current to pass so that it functions as a voltage disconnect. The output of the op amp A**1** can be data provided in a RF field, such as by an RFID card reader.

[0100] In addition to being able to receive data, the module **1300** can also simultaneously transmit data. Data transmission can be accomplished via transistor Q**1**. Toggling Q**1** on and off can result in voltage passing through diode D**1** to interact with the antenna coil. This interaction can result in load modulation. The load modulation can be detected/processed by a component to determine the toggling rate of Q**1**. The toggling rate of Q**1** can be used to encode data thereby transmitting data to another component. Using module **1300**, data handling can be simultaneous receipt and transmission all the while being powered by an RF energy field.

[0101] As discussed herein, the module **1300** can be used for wireless power configurations. For example, the module **1300** can be used in contact and wireless power mode applications (e.g., ISO 7816 and ISO 14443 A, B or Felica). An ISO pad can be used when a card is used as contact type IDO 7816. Voltage (5V or 3.3V) can be supplied through ISO pads. The voltage can be supplied to a voltage regulator, that regulates power to 3.3 V in this case. In case of contact mode, there is no wireless power in some embodiments. In wireless embodiments, there may be no power from ISO pads. The input to voltage regulator can be wired or from contact mode ISO 7816 via pads and Wireless Mode ISO 1443.

[0102] The voltage regulator can be used to supply power to a verification CPU and a finger print sensor. In either case, a finger is placed on to fingerprint sensor. Verification CPU supplies power to Dual mode CPU by logically control I/O Such as 10 mA supply able general IO of CPU. The verification CPU can enable or disables Dual mode CPU to send data through antenna or not. Verification CPU can send data from verification CPU to Dual SEC CPU through ISO 7816 IO, using URT IO of the verification CPU. The verification CPU can control the antenna directly so that at initial stage, while voltage is not strong enough. Antenna can not be activated by Dual SEC CPU at any moment, in another word, fail safe control.

[0103] FIG. **14** illustrates a schematic diagram of a RFID transceiver module circuit **1400** for use in charging applications in accordance with some embodiments of the present invention. The circuit **1400** is similar to that shown in FIG. **5** so for brevity, this discussion will not repeat identical details. The circuit **1400** can be used in power charging applications. For example, by using a super capacitor in parallel to a stacked pair of capacitors, the super capacitor can be used to charge a power source (e.g., a rechargeable battery). In some embodiments, the super capacitor can have a value ranging from several farads to many farads (e.g., 1 to 1000 farads). Preferably the super capacitor is sized and shaped to be small to fit within small, portable devices.

[0104] FIG. **15** illustrates a logical flow diagram **1500** of a method that can be used to implement embodiments of the present invention on a mobile device (e.g., an access card). Those skilled in the art will understand that method **1500** can be performed in various orders (including differently than illustrated in FIG. **15**), additional actions can be implemented as part of a method embodiment, and that some actions pictured in FIG. **15** or discussed below are not necessary. In addition, it should be understood that while certain actions illustrated in FIG. **15** may be discussed herein as including certain other actions, these certain other actions may be carried out in various orders and/or as parts of the other actions depicted in FIG. **15**. Method embodiments of the present invention, such as the one depicted in FIG. **15**, may be implemented with the devices and systems discussed herein. Method embodiments may also be coded in a programming language, stored in a memory, and implemented with a processor or microcontroller. Method embodiments can also include the use of component devices and a processor can be used to manage operation of component devices as desired.

[0105] The method **1500** can initiate in an initial setting. In an initial setting, there may be no power environment. And as a result, no action is made. When proximate an energy field at **1505**, energy can generated via a power circuit at **15 10**. The power circuit at **1515** can charge a capacitor. Charging a capacitor can increase voltage output from the capacitor.

[0106] If a harnessed voltage is over a CPU activation threshold, a CPU can initiate processing functions at **1520**. Or if a CPU does not have such functionality, a dedicated reset circuit can be used. This circuit can be made by RC charging voltage with Schmitt Trigger circuit. If necessary enabling time delay can be added to set up time. If a no-power state is detected at **1525**, then the CPU can go into sleep mode for saving energy at **1530**. If a no-power state is detected the CPU and biometric sensor can enter a sleep mode.

[0107] Sleep modes can also be implemented for power savings. For example, a CPU can enter sleep mode based on a finger print sensor's activity. A CPU can then monitor for a wake up trigger from a finger print sensor, if fingerprint sensor has finger detecting circuit. Such finger detecting circuit can be made by several lines of detection of finger out of whole cell activation. This can save more than 90% of energy of fingerprint sensor. Before detecting fingerprint, this is sleep mode of fingerprint sensor. By this sleep mode of fingerprint sensor and sleep mode of CPU, voltage across the voltage regulator ramps up at maximum speed.

[0108] Once a finger is placed in the field of energy where CPU and sensor can operate, the sensor can gather data at **1535**. For example, a fingerprint sensor can send wake up commands to CPU (verification CPU), and the CPU can start getting data from fingerprint sensor. As this occurs, only

CPU's interface, such as SPI interface and memory, is active to receive data from CPU. For example, a 128×256 bit cell has 8 bit gray scale (262 kb), 10 MHz reading speed takes 0.026 sec, and the power dissipation of sensor is between 0.1 mA to 7 mA depending on sensor type. The current of CPU can be 10 mA at 10 MHz. Total current at this phase is 17 mA.

[0109] Once data from fingerprint sensor is transferred to CPU memory, finger print sensor will be in sleep mode again even though finger is on the sensor. Then a taken image data can be processed as image processing to reduce data as convert gray scale data to skinny but continuous line data of fingerprint pattern. This process is done by filtering, such as two dimensional FFT. Then from skinned line data, crossing point, edge of line can be detected by Minutiae processing program. Through this process, minutiae vector can be obtained as personal ID vector data.

[0110] This data is compared to the stored reference data in CPU flash memory at **1535**. This comparison may not be 100% matching. Rather, the comparison can give a score of matching. Matching can be measured based on the numbers of minutiae to be compared. In the process of matching, angle is to be rotated by A'ffin transfer program. Also, a threshold like FAR (False Acceptance ratio) as 0.1%, as the card can be used only card holder. FRR (False Rejection Ratio as 0.01%) so that mismatch frustration. This threshold adjustment can be FRR<FAR.

[0111] After testing the finger print data, test results can be communicated at **1540** to another device. The data communication can include test result and other information. For example, when fingerprint is matched, the CPU can generate data comprising of the event of fingerprint match, Unique ID of CPU, Unique ID of sensor if available, assigned ID of card or along with card holder name, of if required picture of card holder, and or time stamp if useful.

[0112] Data transmission can be done in an encrypted fashion. The CPU can encrypt the data to increase security. Encryption protocols can include Triple DES, AREA, Camellio, AES, and RSA. Other encryption schemes can also be employed to meet whatever encryption required by users. PKI can be used as additional encryption and UID, time stamp, or part of Minutiae can be used as private key.

[0113] Those information can be generated in Verification CPU, but can be also generated by security CPU, such as dual mode CPU (SEC CPU hereafter), because SEC CPU has coprocessor of encryption. Wireless controller, which is for example, wireless portion of SEC CPU, start shaking hands with card reader. The communications can be wireless and sent through antenna by modulating load of antenna.

[0114] The embodiments of the present invention are not limited to the particular formulations, process steps, and materials disclosed herein as such formulations, process steps, and materials may vary somewhat. Moreover, the terminology employed herein is used for the purpose of describing exemplary embodiments only and the terminology is not intended to be limiting since the scope of the various embodiments of the present invention will be limited only by the appended claims and equivalents thereof.

[0115] Therefore, while embodiments of the invention are described with reference to exemplary embodiments, those skilled in the art will understand that variations and modifications can be effected within the scope of the invention as defined in the appended claims. Accordingly, the scope of the various embodiments of the present invention should not be limited to the above discussed embodiments, and should only be defined by the following claims and all equivalents.

We claim:

1. A portable wireless device used for event actuation, the portable wireless device comprising:

a wireless power harnessing module that comprises an antenna tuned to a resonant frequency associated with a source of an energy field, the antenna being tuned with a capacitor placed in parallel with the antenna;

the antenna comprising several windings which when proximate the energy field, result in the wireless power harnessing module sourcing power;

a biometric data comparison module coupled to the wireless power harnessing module, the biometric data comparison configured to enter a powered state when receiving adequate power from the wireless power harnessing module, wherein in the powered on state, the biometric data comparison module is operatively configured to receive external biometric data from an external source and compare the external biometric date to stored biometric data; and

a communication module configured to provide information responsive to the comparison of the external biometric date and stored biometric data.

2. The portable wireless device of claim **1**, wherein the wireless power harnessing module, the biometric data comparison module, and the communication module reside within an ISO-7816 defined card outline.

3. The portable wireless device of claim **1**, wherein the wireless power harnessing module comprises a rectifier circuit coupled to a low impedance winding of the antenna and a common ground, the rectifier circuit configured to convert AC voltage provided by the antenna to DC voltage.

4. The portable wireless device of claim **1**, further comprising a capacitor located in parallel with the antenna, wherein the relationship between the capacitor and the antenna defines the resonant frequency.

5. The portable wireless device of claim **1**, wherein the device has no local power source.

6. The portable wireless device of claim **1**, wherein the antenna is divided up into segments disposed at various tap positions such that antenna has multiple segments configured to carry out multiple functions.

7. The portable wireless device of claim **1**, wherein the wireless power harnessing module harness power from the energy field simultaneously to the communication module transmitting and receiving data from the energy field.

8. The portable wireless device of claim **1**, wherein the antenna comprises an antenna coil pattern wound in a concentric fashion that comprises inner and outer windings.

9. The portable wireless device of claim **8**, wherein the antenna coil pattern is a continuous planar copper trace having tap positions located at various places along the coil pattern so the antenna has multiple segments configured to have different functions.

10. The portable wireless device of claim **1**, wherein the wireless power harnessing module comprises a rectifier circuit is a voltage doubling circuits that comprises two Schottky barrier diodes arranged in a full wave rectifying arrangement.

11. A wireless access control device, the device comprising:

a power circuit configured to have a default non-energized state and an energized state, the power circuit configured to receive energy from an energy field to enter the energized state so that the power circuit can source electrical power, wherein the power circuit is finely tuned to a carrier frequency of the energy field; and

a processor coupled to the power circuit, the processor configured to receive electrical power when the power circuit enters the energized state, the processor further configured to receive data from a sensor, and in response to the received data, the processor further configured to generate a signal corresponding to an access level.

12. The wireless access control device of claim **11**, wherein the processor receives power only from the power circuit when energized and the processor is not configured to receive power from any other power source.

13. The wireless access control device of claim **11**, the power circuit comprising a power detection stage, a power conversion stage, and a receiving antenna, the receiving antenna being integrated with the power detection stage and being shaped and sized to produce electrical power when placed into an energy field.

14. The wireless access control device of claim **11**, the power circuit comprising an antenna finely tuned to the carrier frequency of the energy field.

15. The wireless access control device of claim **11**, wherein the processor is configured to control data communication between the wireless access control device and the source of the energy field during the energized state.

16. A portable wireless device capable of harnessing wireless energy comprising:

an antenna and a tuning capacitor connected in parallel to form a tank circuit, the tank circuit being finely tuned to a resonant frequency associated with a carrier base frequency of source of an energy field;

the antenna comprising several windings which when proximate the energy field, result in the antenna sourcing electrical current and voltage;

the antenna further comprising a plurality of segments set off by a plurality of taps disposed at various places along the length of the antenna, wherein one of the segments can be configured to receive and transmit data with the energy field simultaneously with receiving energy from the energy field; and

a rectifier circuit connected to a first tap and a second tap of the antenna, the first tap being located on an inner antenna winding and wherein the second tap of the antenna is in electrical communication with a common ground, the rectifier circuit configured to convert the sourced electrical current and voltage to a DC energy source.

17. The portable wireless device of claim **16**, further comprising an antenna driving circuit configured to drive the antenna for data communication, the antenna driving circuit being connected to a third tap, the third tap being located on an outer antenna winding.

18. The portable wireless device of claim **16**, further comprising a voltage divider capacitor network coupled to the rectifier, the rectifier comprising a pair of Schottky diodes with the cathode of a first diode connected to the anode of a second diode, anode of the first diode connected to ground, and the cathode of the second diode connected to the voltage divider capacitor network.

19. The portable wireless device of claim **18**, wherein the voltage divider capacitor network comprises first capacitor connected in parallel to two series connected capacitors, wherein the cathode of the second diode is connected to a positive terminal of the first and second capacitors, and the anode of the first diode is connected to a negative terminal of the first and third capacitors.

20. The portable wireless device of claim **18**, wherein capacitors in the voltage divider capacitor range in value from about 1 pF to about 100 pF, the tuning capacitor ranges in value from about 10 pF to 500 pF, the antenna has between 1 to 10 coil windings, and the coil windings have a width ranging between about 1 mm to about 10 mm.

21. The portable wireless device of claim **18**, wherein the voltage divider capacitor comprises an energy storage capacitor configured to store energy, the energy storage capacitor having a value ranging from about 0.5 micro-farads to about 1000 farads.

22. A method of harnessing electrical energy from an energy field while simultaneously communicating data with the energy field, the method comprising:

configuring a portable device with a tank circuit tuned to a center frequency of an energy field, wherein an inductor of the tank circuit can interact with the energy field to convert wireless energy into electrical energy so that the inductor can source electrical power; and

configuring a processor located on the portable device to receive electrical power sourced by the inductor and configuring the processor to receive and provide data for communication with a device emitting the energy field, wherein data can be received and transmitted using coils of the inductor while the inductor is sourcing energy.

23. The method of claim **22**, further comprising configuring the portable device to receive external biometric data, to test the biometric data against a stored biometric set of data, and to communicate results of the test via the inductor.

24. The method of claim **22**, further comprising configuring the processor to communicate data by modulating the field load of the energy field.

25. The method of claim **22**, further comprising providing a voltage conversion circuit on the portable device to convert the energy sourced by the inductor from AC to DC and to regulate the DC voltage relative to a predetermined threshold.

26. A computer program product embodied in a computer-readable medium for execution by a processor or engine, the computer program product comprising an algorithm to manage activated carried out by a processor in managing power and testing biometric data, the method comprising:

detecting an appropriate power level being sourced by an antenna that is finely tuned to resonate at a center carrier frequency of an energy field, wherein the power level is provided in electrical form after the antenna converts wireless energy to electrical energy;

communicating with a biometric sensor to determine if the sensor detects presence of biometric data and has captured external biometric data;

testing received biometric data against stored biometric data to determine if the captured external biometric data matches the stored biometric data; and

issuing communication signals for wireless transmission from the antenna to another component, the communication signals comprising data about results of the biometric data test.

27. The method of claim **26**, further comprising instructing one of the biometric sensor or a system processor to enter a sleep mode if a low power level state is detected or to preserve power.

28. The method of claim **26**, wherein testing received biometric data against stored biometric data includes configuring a system processor to extract digital data from the captured external biometric data to place the external biometric data in the same format as the stored biometric data.

29. The method of claim **26**, wherein testing received biometric data against stored biometric data includes generating a score indicative of the data test and wherein the score can determine a positive or negative test result relative to a predetermined threshold.

30. The method of claim **26**, wherein testing received biometric data against stored biometric data includes generating a false acceptance ratio and a false rejection ratio and wherein a match condition can be achieved with the false rejection rate is less than the false acceptance ratio.

* * * * *

US005159703A

United States Patent [19]

Lowery

[11] **Patent Number: 5,159,703**

[45] **Date of Patent: Oct. 27, 1992**

[54] **SILENT SUBLIMINAL PRESENTATION SYSTEM**

[76] Inventor: **Oliver M. Lowery**, 5188 Falconwood Ct., Norcross, Ga. 30071

[21] Appl. No.: **458,339**

[22] Filed: **Dec. 28, 1989**

[51] **Int. Cl.**^{5} **H04B 7/00;** H04R 25/00; H04R 3/02

[52] **U.S. Cl.** **455/42;** 455/46; 455/66; 381/73.1; 128/420.5

[58] **Field of Search** 455/46, 47, 66, 109, 455/110, 42-43; 381/73.1, 105, 124; 358/141-143; 600/28; 128/420.5; 380/38

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,060,795	10/1962	Corrigan et al.	352/131
3,278,676	10/1966	Becker	358/142
3,393,279	7/1968	Flanagan	128/420.5
3,712,292	1/1973	Zentmeyer, Jr.	600/28
4,141,344	2/1979	Barbara	600/28
4,395,600	7/1983	Lundy et al.	381/73.1
4,463,392	7/1984	Fischer et al.	360/30
4,777,529	10/1988	Schultz et al.	381/73.1
4,834,701	5/1989	Masaki	600/28
4,877,027	10/1989	Brunkan	128/420.5

Primary Examiner—Reinhard J. Eisenzopf
Assistant Examiner—Andrew Faile

[57] **ABSTRACT**

A silent communications system in which nonaural carriers, in the very low or very high audio frequency range or in the adjacent ultrasonic frequency spectrum, are amplitude or frequency modulated with the desired intelligence and propagated acoustically or vibrationally, for inducement into the brain, typically through the use of loudspeakers, earphones or piezoelectric transducers. The modulated carriers may be transmitted directly in real time or may be conveniently recorded and stored on mechanical, magnetic or optical media for delayed or repeated transmission to the listener.

3 Claims, 3 Drawing Sheets

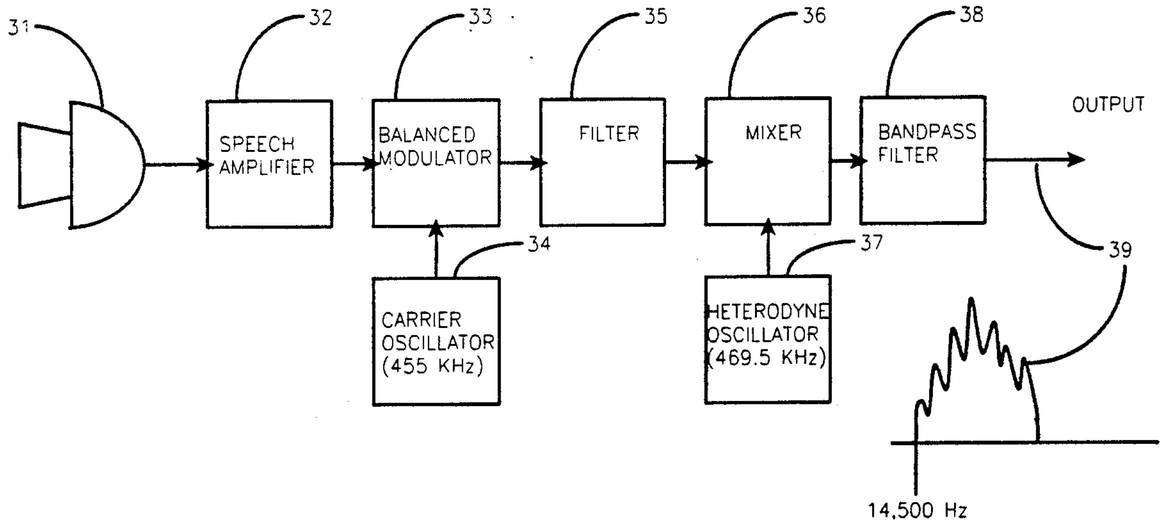
31
32
SPEECH AMPLIFIER
33
BALANCED MODULATOR
35
FILTER
36
MIXER
38
BANDPASS FILTER
OUTPUT
34
CARRIER OSCILLATOR (455 KHz)
37
HETERODYNE OSCILLATOR (469.5 KHz)
39
14,500 Hz

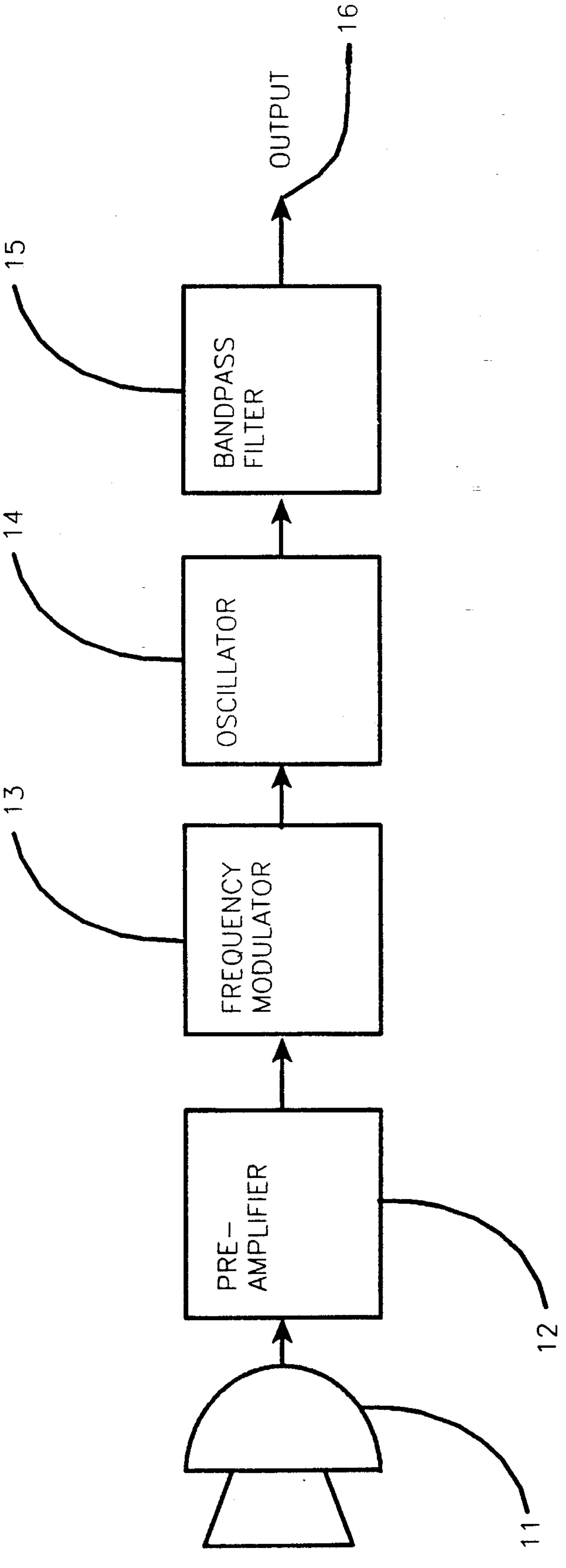
11
PRE-AMPLIFIER
12
FREQUENCY MODULATOR
13
OSCILLATOR
14
BANDPASS FILTER
15
OUTPUT
16

FIG. 1

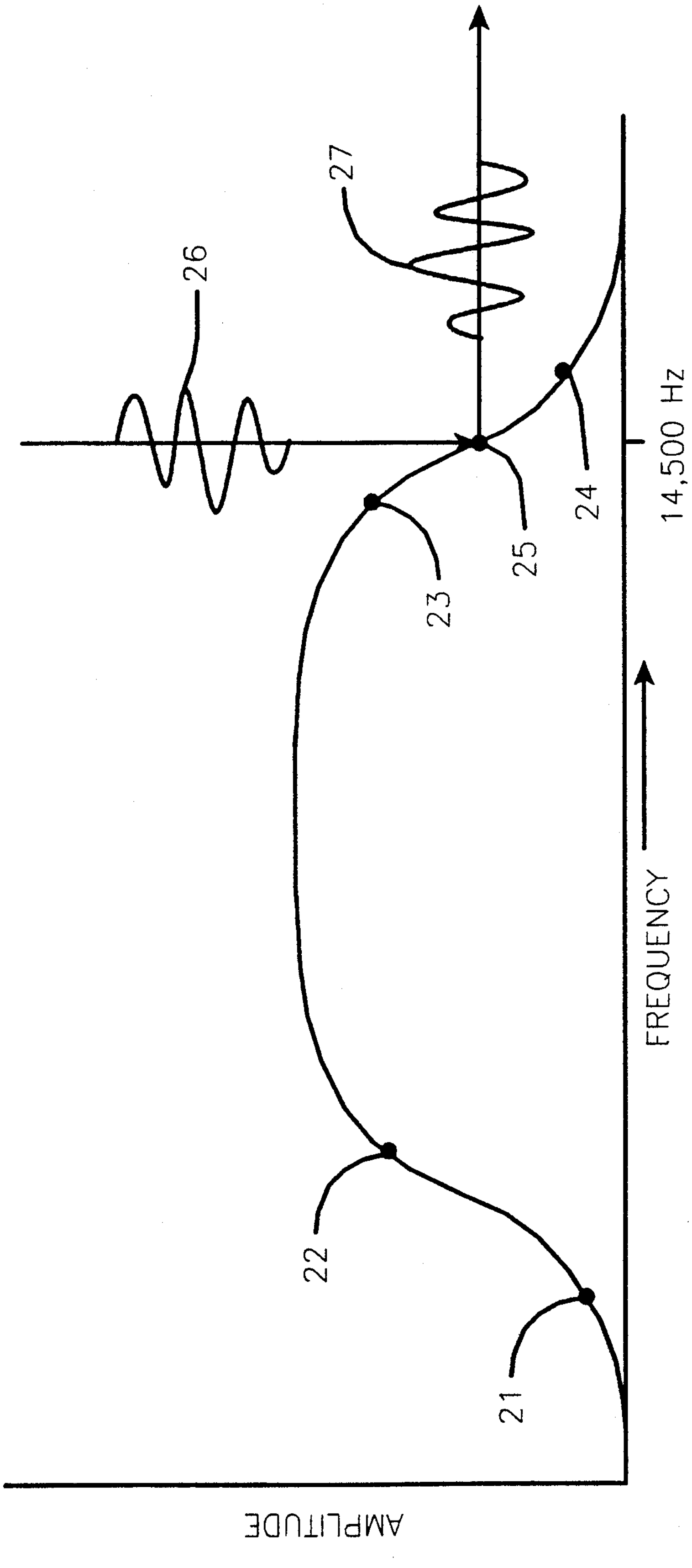

FIG. 2

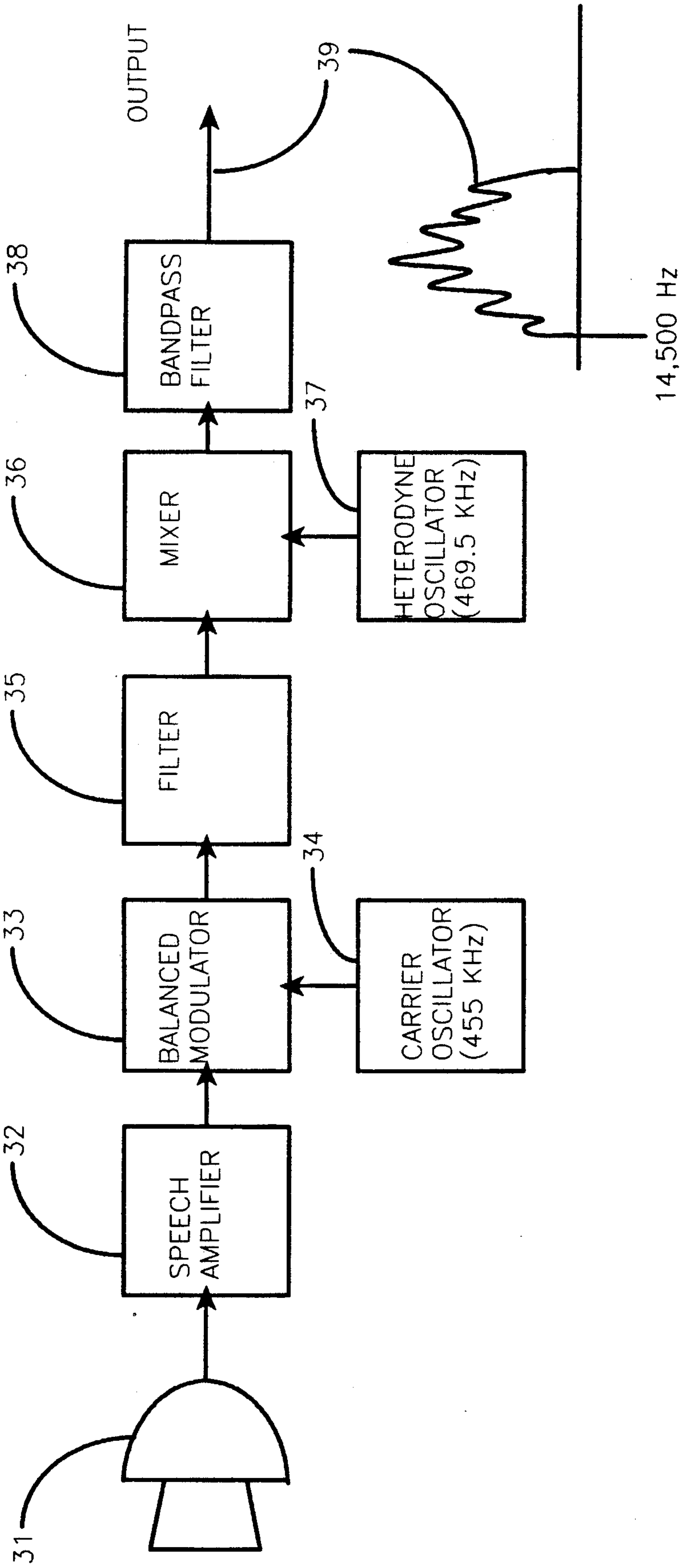
31
32
SPEECH AMPLIFIER
33
BALANCED MODULATOR
34
CARRIER OSCILLATOR (455 KHz)
35
FILTER
36
MIXER
37
HETERODYNE OSCILLATOR (469.5 KHz)
38
BANDPASS FILTER
OUTPUT
39
14,500 Hz

FIG. 3

SILENT SUBLIMINAL PRESENTATION SYSTEM

BACKGROUND—FIELD OF THE INVENTION

This invention relates in general to electronic audio signal processing and, in particular, to subliminal presentation techniques.

BACKGROUND—DESCRIPTION OF PRIOR ART

Subliminal learning enjoys wide use today and subliminal tapes are being manufactured by a number of companies in the United States alone. Several decades of scientific study indicate that subliminal messages can influence a human's attitudes and behavior. Subliminal, in these discussions, can be defined as "below the threshold of audibility to the conscious mind." To be effective however, the subliminally transmitted information (called affirmations by those in the profession) must be presented to the listener's ear in such a fashion that they can be perceived and "decoded" by the listener's subconscious mind. We are referring to audio information in this discussion, however, information could be inputted into the subject's subconscious mind through any of the body's sensors, such as touch, smell, sight or hearing. As an example, early development work in the subliminal field utilized motion pictures and slide projections as the medium. Early research into visual and auditory subliminal stimulation effects is exemplified by U.S. Pat. Nos. 3,060,795 of Corrigan, et al. and 3,278,676 of Becker. U.S. Pat. No. 4,395,600 of Lundy and Tyler is representative of later developments in today's subliminal message techniques.

The majority of the audio subliminal tapes available today are prepared using one basic technique. That is, the verbal affirmations are mixed with, and recorded at a lower level than, a "foreground" of music or sounds of ocean surf or a bubbling mountain brook or other similar "masking" sounds. The affirmations are generally recorded 5 decibels (db) or so below the "foreground" programming and regenerative automatic gain control is usually applied to permit the affirmations to change their recorded amplitude in direct proportion to the short term averaged amplitude of the continually varying "foreground" material. In other words, the volume of the affirmations will follow or track the volume changes of the "foreground" programming, but at a lower volume level. Circuit provisions are also usually included to "gate" the affirmations off when the music amplitude is low or zero. This insures that the affirmations cannot be heard during quiet program periods. Thus, today's subliminal affirmations can be characterized as being "masked" by music or other sounds, of constantly changing amplitude and of being reduced or cut off entirely during periods of low or quiet "foreground" programming.

One of the principal, and most widely objected to, deficiencies in available subliminal tape presentation techniques is that the presence of the "foreground" material is intrusive to both the listener and to anyone else in the immediate area. No matter what "foreground" material is chosen, the fact remains that this material can be heard by anyone within its range and presents a definite distraction to other activities such as conversation, thought, desire to listen to other programming such as radio or television, need to concentrate, etc. Additionally, and because the tapes are used repeatedly by the same listener, any "foreground" music or material eventually becomes monotonously tiring to that listener.

It is the purpose of the following described invention to eliminate or greatly reduce all of the above deficiencies. Although its application to the magnetic tape medium is described in the following discussion, the technique is equally applicable to most other desired transmission mediums, such as Compact Disc, videocassettes, digital tape recorders, Public Address (PA) systems, background music installations, computer software programs, random access memory (RAM), read only memory (ROM), "live", real time applications and other mediums now in existence or to be developed in the future.

Implemented on tape cassettes, for example, the subliminal presentation described here is inaudible i.e., high audio or ultrasonic frequencies, the affirmations are presented at a constant, high amplitude level, and they occupy their own "clear channel", non-masked frequency allocations. If desired, the previously described "foreground" music or other material can be added to the tape through use of an audio mixer. The "silent" recordings are inaudible to the user or by others present and are therefore very effective for use during periods of sleep or when in the presence of others. Additionally, the basic requirements of subliminal stimulation are met. That is, the affirmations are efficiently transmitted to the ear and, while undetected by the conscious mind, are perceived by and efficiently decoded by the subconscious mind.

OBJECTS OF THE INVENTION

Accordingly, several objects and advantages of my invention are:

(a) to provide a technique for producing a subliminal presentation which is inaudible to the listeners(s), yet is perceived and demodulated (decoded) by the ear for use by the subconscious mind.

(b) to provide a technique for transmitting inaudible subliminal information to the listener(s) at a constant, high level of signal strength and on a clear band of frequencies.

(c) to provide a technique for producing inaudible subliminal presentations to which music or other "foreground" programming may be added, if desired.

Still further objects and advantages will become apparent from a consideration of the ensuing description and drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

In the drawings, the first digit of each component number also refers to the figure number where that component can be located.

FIG. 1 represents the block diagram of a suitable system which will generate a frequency modulated (FM) signal at 14,500 Hz.

FIG. 2 represents an approximation of the frequency response curve of the human ear and the signal decoding process.

FIG. 3 represents the block diagram of a suitable system which will generate a single sideband, suppressed carrier, amplitude modulated (AM) signal at 14,500 Hz.

REFERENCE NUMERALS IN DRAWINGS	
11 microphone or other	14 low distortion

-continued

REFERENCE NUMERALS IN DRAWINGS	
audio input signal	audio oscillator
12 audio preamplifier if required	15 high pass or band pass audio filter
13 frequency modulation circuit	16 output to tape recorder or other device
21 point on low freq end response curve	25 midpoint on curve between points 23 and 24
22 point on low freq end of ear response curve	26 speaker output of FIG. 1 to ear
23 point on high freq end of ear response curve	27 demodulated subliminal audio inputted to ear
24 point on high freq end of ear response curve	31 microphone
32 speech amplifier	33 balanced modulator
34 carrier oscillator (455 KHz)	35 filter
36 mixer	37 heterodyne oscillator (469.5 KHz)
38 bandpass filter	39 output signal

DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT

Please refer now to FIG. **1** and FIG. **2**, which are drawings of a preferred implementation of the invention.

The principle of operation of the silent subliminal presentation system is as follows:

An audio signal in the upper frequency region of the audio spectrum (for example, 14,500 Hz) is modulated with the desired information. The type of modulation may be any type suitable for subliminal applications; frequency modulation (FM), phase modulation (PM), upper single sideband with suppressed carrier, amplitude modulation (AM), tone modulation, etc.

For broadest application, the high audio frequency selected as the carrier frequency must meet two basic criteria:

(1) be high enough in the audio spectrum that its presence to the human ear is essentially unnoticed or undetectable (without the listener being informed that the signal is actually present) and,

(2) be low enough in the audio spectrum that it (and its modulation content) can produce a useful output power from home entertainment type cassette or reel-to-reel magnetic recorders.

This would also include, of course, small portable and automobile tape decks.

Alternatively, the output of the system can be fed directly into an audio amplifier and its speaker/earphone system, Public Address system, etc.

FIG. **1** provides the block diagram of an example of a system capable of generating the desired silent frequency modulated carrier.

The modulation information is inputted into the microphone **11**. Other suitable input devices may be substituted for microphone **11**, such as a tape recorder or a radio. The microphone **11** is connected to the preamplifier **12** and should have provisions for adjusting its gain in order that the optimum modulation index can be set in the frequency modulator **13**. The frequency modulator **13** modulates the frequency of oscillator **14** which has been adjusted for an output of 14,500 Hz as described above. The output of oscillator **14** is fed through a suitable bandpass filter **15** into the tape recorder or directly into a suitable amplifier/speaker system. It is the purpose of the bandpass filter to remove or attenuate audible products of the modulation process in order to maintain as audibly silent an output as practical.

On the receiving end, FIG. **2** represents an approximate and idealized frequency response curve of the human ear. The frequency modulated carrier (centered at 14,500 Hz), as generated above and played through a tape recorder or amplifier/speaker system, is shown on FIG. **2** as speaker output **26**, impinging upon the upper slope of the ear's response curve at point **25**. The frequency modulated excursions of the speaker output **26** swing between points **23** and **24** on the ear's upper response curve. Because the response curve between points **23** and **25** is relatively linear, this action results in a relatively linear demodulation of the original modulation intelligence, which is passed on subliminally to the inner ear. The amplitude of the demodulated output is not high enough to be detected by the conscious mind but is sufficient in amplitude to be detected by the subconscious mind. In the field of communications engineering design, the above demodulation process in known as slope detection and was used in early FM receiver design. In those receivers, the response curve was formed by the action of a tuned (inductive/capacitance) circuit. In our case, the response curve is formed by the natural response curve of the human ear. The same slope detection technique can be performed at the low frequency end of the human ear response curve. This region is indicated on FIG. **2** as between points **21** and **22**. This region, however, has a much smaller available bandwidth and is therefore more restricted as to the amount of information that can be transmitted in an inaudible manner.

In practice, the listener adjusts the volume control of the tape recorder or amplifier to a level just below that at which the listener hears an audible sound or noise from the speaker of the tape recorder. If the recording process is properly done, a spectrum analyzer or a calibrated sound level meter will reveal a strong signal emanating from the tape recorder speaker. A calibrated sound level meter, at a distance of 1 meter (with C weighting and referenced to the standard of 0.0002 micro bar) will typically indicate a silent power output of from 60 to 70 decibels. This is equivalent to the audio power of a loud conversation, yet, in the described system, is inaudible or unnoticed by the listener.

FIG. **3** illustrates a system which generates a suitable amplitude modulated (AM) signal, instead of the frequency modulated (FM) system described above. The output is a modulated, single sideband (SSB), suppressed carrier (AM) signal at 14,500 Hz.

The block diagram represents a common scheme for generating an SSB signal and will be briefly described.

The desired subliminal information is spoken into microphone **31**. This signal is amplified by speech amplifier **32** and injected into one port of balanced modulator **33**. A continuous wave signal of **455** KHz is generated by carrier oscillator **34** and is injected into the second port of balanced modulator **33**. The output of balanced modulator **33** is a double sideband, suppressed carrier signal at 455 KHz. This signal is fed through filter **35**, causing one of the two sidebands to be removed. This signal is fed into one port of mixer **36**. A continuous wave signal at a frequency of 469.5 KHz from hetrodyne oscillator **37** is fed into the other port of mixer **36**, resulting in an output of the original subliminal audio information but translated 14,500 Hz higher in frequency. The bandpass filter **38** attenuates signals and noise outside of the frequencies of interest. The ampli-

tude modulated audio output signal is shown as output **39**.

Thus, as stated earlier, my invention provides a new system for subliminal presentations which is:

(a) silent,

(b) outputs a constant, high level modulated signal and,

(c) occupies a band of clear channel frequencies.

The foregoing description of the preferred embodiment of the invention has been presented for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed. Many modifications and variations are possible in light of the above discussions. It is intended that the scope of the invention be limited not only by this detailed description, but rather by the claims appended hereto.

What is claimed:

1. A silent communications system, comprising:

(a) amplitude modulated carrier means for generating signals located in non-aural portions of the audio and in the lower portion of the ultrasonic frequency spectrum said signals modulated with information to be perceived by a listener's brain and,

(b) acoustic and ultrasonic transducer means for propagating said signals, for inducement into the brain, of the listener, and,

(c) recording means for storing said modulated signals on mechanical, magnetic and optical media for delayed or repeated transmissions to the listener.

2. A silent communications system, comprising:

(a) frequency modulated carrier means for generating signals located in non-aural portions of the audio and in the lower portion of the ultrasonic frequency spectrum, said signals modulated with information to be perceived by a listener's brain, and;

(b) acoustic and ultrasonic transducer means for propagating said signals, for inducement into the brain of the listener, and;

(c) recording means for storing said modulated signals on mechanical, magnetic and optical media for delayed or repeated transmissions to the listener.

3. A silent communications system, comprising:

(a) a combination of amplitude and frequency modulated carrier means for generating signals located in non-aural portions of the audio and in the lower portion of the ultrasonic frequency spectrum, said signals modulated with information to be perceived by a listener's brain, and

(b) acoustic and ultrasonic transducer means for propagating said signals, for inducement into the brain of the listener;

(c) recording means for storing said modulated signals on mechanical, magnetic and optical media for delayed or repeated transmissions to the listener.

* * * * *

US005356386A

United States Patent [19]

Goldberg et al.

[11] **Patent Number: 5,356,386**

[45] **Date of Patent: * Oct. 18, 1994**

[54] **APPARATUS FOR LOCATING BODY CAVITIES**

[75] Inventors: **Edward M. Goldberg**, Glencoe; **Lev Melinyshyn**, Mt. Prospect; **Michael Jaron**, Des Plaines; **Jeffrey M. Stupar**, Chicago, all of Ill.

[73] Assignee: **Uresil Corporation**, Skokie, Ill.

[*] Notice: The portion of the term of this patent subsequent to May 31, 2005 has been disclaimed.

[21] Appl. No.: **924,407**

[22] Filed: **Aug. 3, 1992**

Related U.S. Application Data

[60] Continuation of Ser. No. 559,757, Jul. 30, 1990, abandoned, which is a division of Ser. No. 310,366, Feb. 13, 1989, Pat. No. 4,944,724, which is a continuation of Ser. No. 58,400, Jun. 5, 1987, abandoned.

[51] **Int. Cl.^{5}** **A61M 1/00**

[52] **U.S. Cl.** **604/118**; 604/323

[58] **Field of Search** 604/118, 318, 319, 320, 604/323

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,015,603	4/1977	Kurtz et al.	604/318
4,063,556	12/1977	Thomas et al.	604/318
4,543,095	9/1985	Jensen	604/318
4,820,284	4/1989	Hauri	604/318
4,944,724	7/1990	Goldberg et al.	604/118

Primary Examiner—Paul J. Hirsch

Attorney, Agent, or Firm—Laff, Whitesel, Conte & Saret

[57] **ABSTRACT**

An apparatus for locating a body cavity having a fluctuating fluid pressure by signaling the fluctuating pressure when the cavity is entered and then transporting fluid either into or out of the cavity.

15 Claims, 2 Drawing Sheets

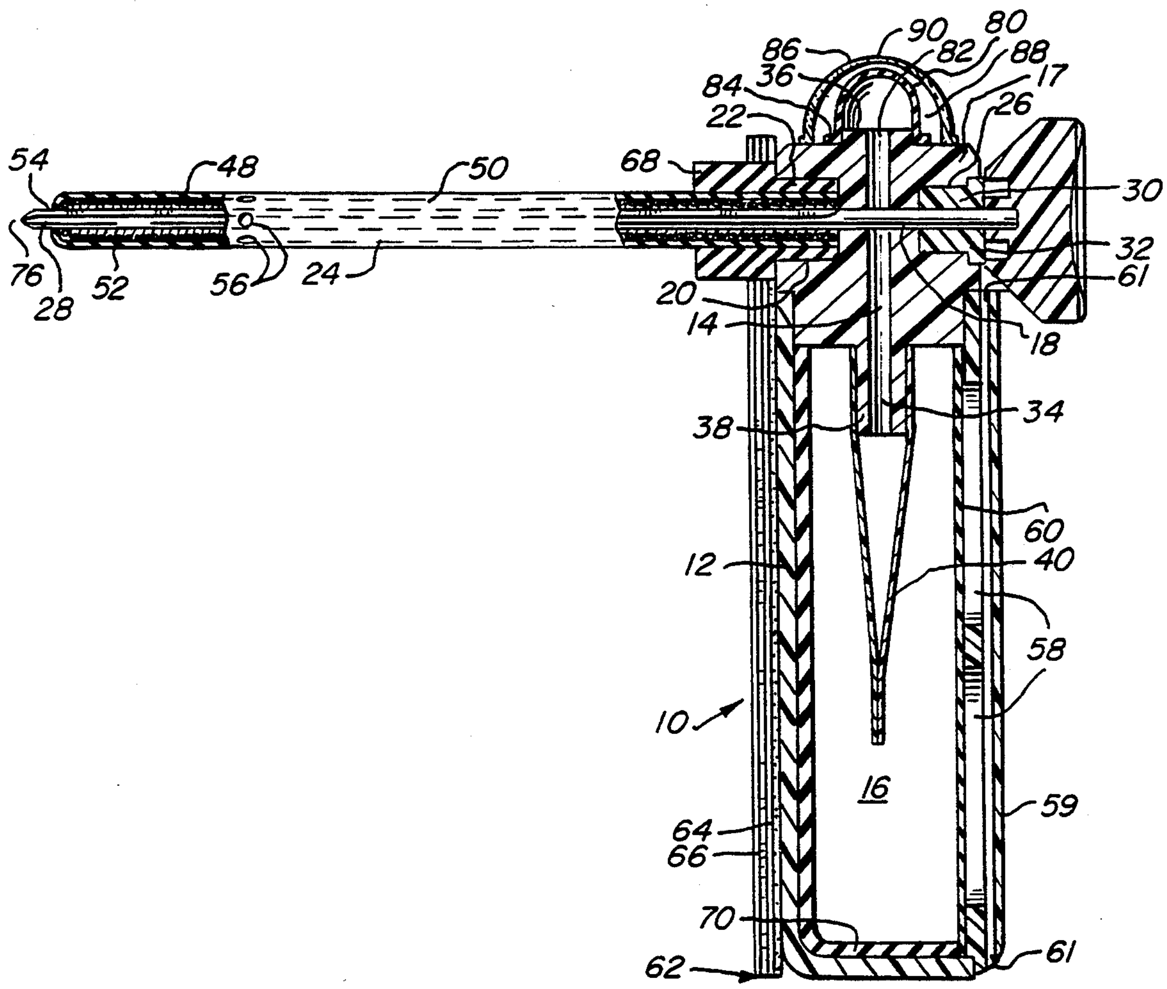

FIG. 1A

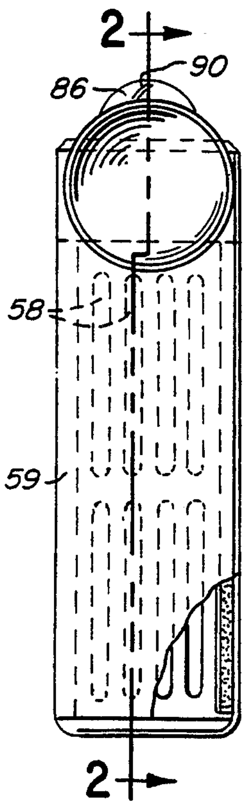
2
2
86
90
58
59

FIG. 1

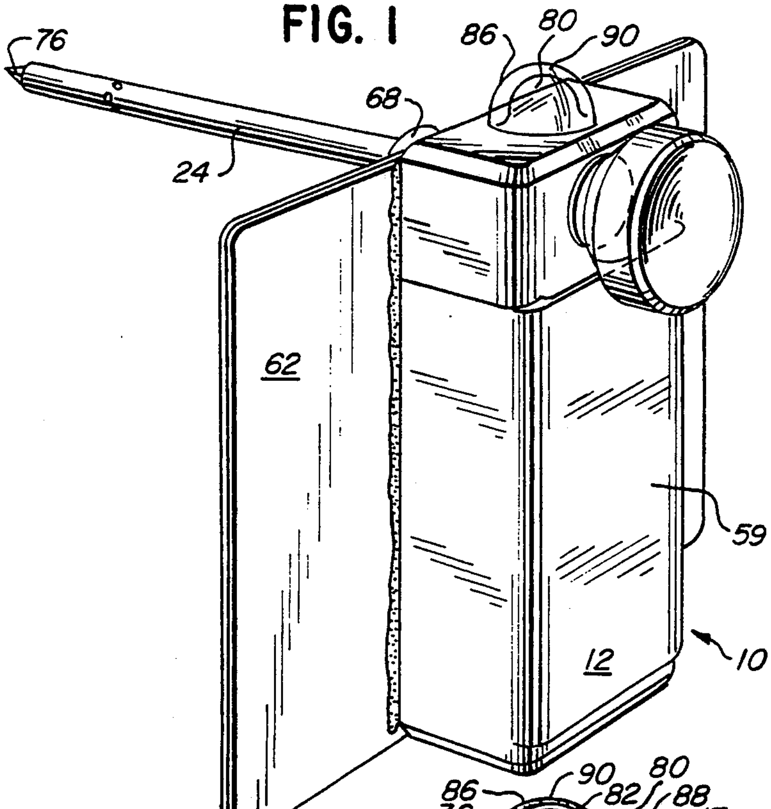
76
68
24
86
80
90
62
59
12
10

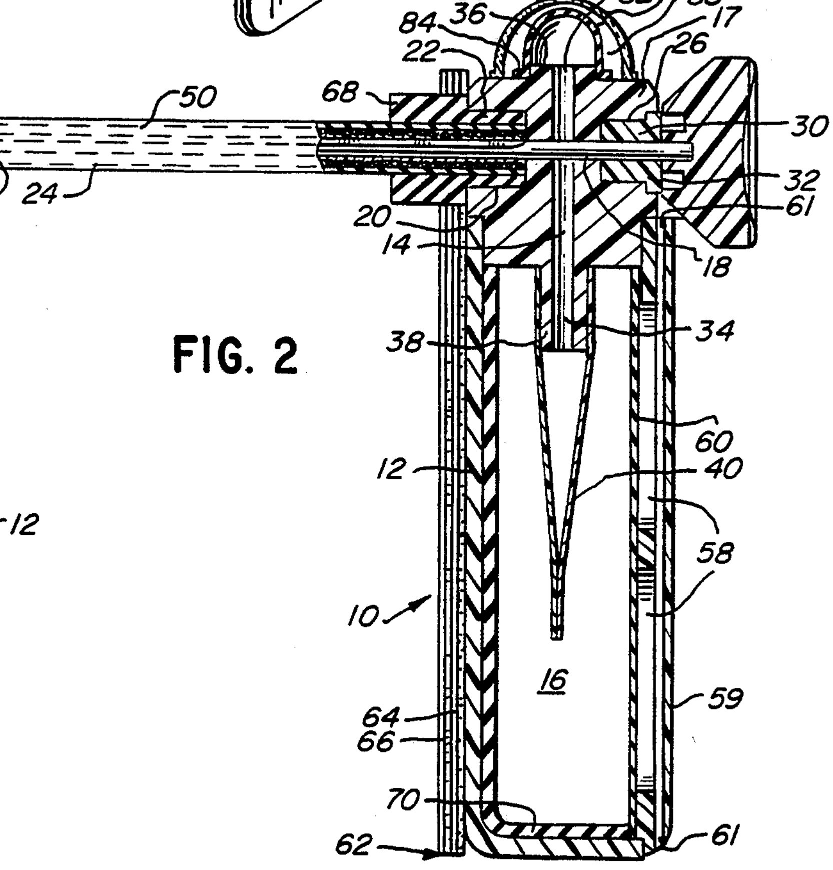
86
90
80
84
36
82
88
17
22
26
54
48
50
68
30
76
28
52
56
24
32
61
20
14
18
38
34
60
12
40
58
10
16
59
64
66
70
62
61

FIG. 2

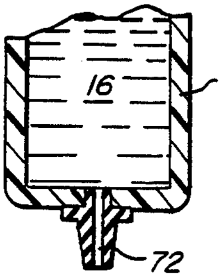

FIG. 2A

FIG. 2B

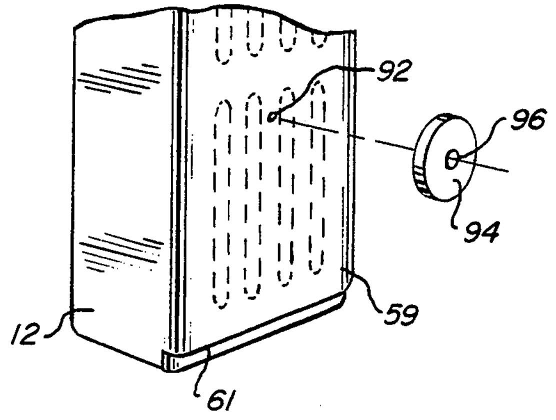
92
96
94
12
59
61

FIG. 3

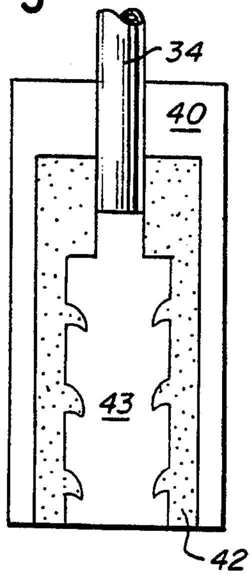

FIG. 4A

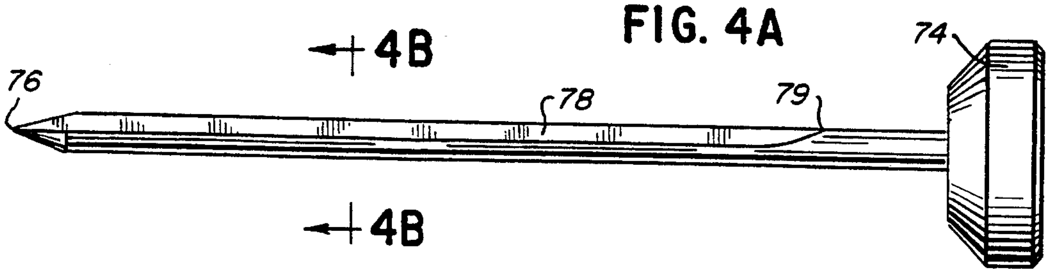
4B
74
76
78
79
4B

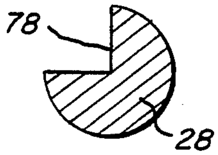

FIG. 4B

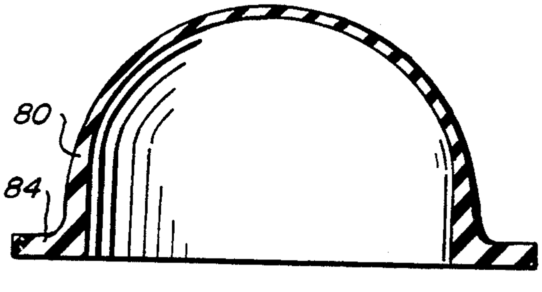

FIG. 5A

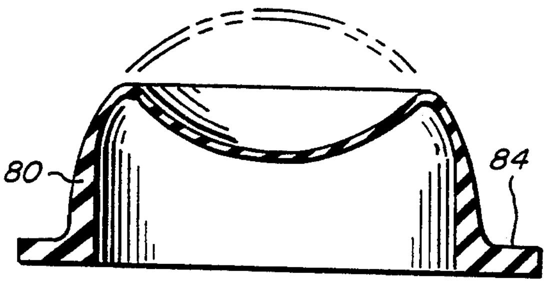

FIG. 5B

APPARATUS FOR LOCATING BODY CAVITIES

This is a continuation of application Ser. No. 559,757, now abandoned, filed Jul. 30, 1990 which is a divisional of copending application Ser. No. 310,366 now U.S. Pat. No. 4,944,724 filed on Feb. 13, 1989 which in turn is a continuation of application Ser. No. 07/058,400 filed on Jun. 5, 1987 and now abandoned.

BACKGROUND OF THE INVENTION

This invention relates generally to apparatus for draining fluids from body cavities and for introducing fluids into body cavities. More particularly, this invention relates to an improved apparatus for positively locating a body cavity having fluctuating fluid pressure and then either draining fluids from the cavity or introducing fluids into it.

An apparatus for draining fluids from body cavities, including particularly the pleural cavity, is described in U.S. Pat. No. 4,664,660. The apparatus described in that patent includes a vented housing having a fluid-receiving chamber, an anti-reflux valve mounted within the housing and a catheter, extending from the housing, in communication with the chamber through the anti-reflux valve.

The apparatus of U.S. Pat. No. 4,664,660 is used in the drainage of fluids from the pleural cavity by inserting a solid trocar into the end of the catheter. A skin incision is then prepared and the trocar/catheter assembly is introduced into the pleural space through the incision. When the trocar is removed from the catheter, fluid drains from the pleural cavity through the catheter and the vented housing.

While the above-described apparatus represents a very important contribution to the art, in using it one cannot ascertain whether the open distal end of the trocar/catheter assembly is properly located in the pleural cavity until the trocar is removed. Since drainage cannot begin until the trocar is removed, the proper positioning of the catheter in the pleural space cannot be determined with the trocar in place. In addition, damage to the lung, heart and surrounding tissue could occur if the trocar is not removed as soon as the tip of the trocar catheter assembly enters the pleural space.

Another apparatus for draining fluids from body cavities, particularly aspiration of liquids from the pleural cavity, is disclosed in U.S. Pat. No. 4,447,235. The apparatus described in this patent includes a catheter/hollow needle assembly which is inserted into the pleural cavity. During the insertion procedure, a vacuum is maintained in the needle with a syringe so that liquid will enter the syringe for observation by the surgeon when the needle enters the pleural space. This device cannot function unless a vacuum is maintained. Also, it cannot signal subsequent dislocation of the catheter from the pleural space or completion of drainage since the syringe is disconnected (and cannot be reintroduced) after the initial insertion of the catheter in the pleural cavity.

A medical suction device with an indicator flag to signal the pressure being developed by the device is described in U.S. Pat. No. 4,404,924. The indicator flag is designed to stand upright when the pressure in the device is relatively high and to collapse when suction is developed. The flag, unfortunately, indicates the state of fill of the suction device, not the satisfactory location of a catheter in the body cavity being aspirated or the completion of the aspiration procedure. In addition, it has no application in a procedure conducted without vacuum assistance.

Finally, U.S. Pat. No. 4,164,938 describes a device for diagnosing the presence of a tension pneumothorax. This device includes a sleeve with a needle at one end for puncturing the chest wall and extending into the pleural cavity and a diaphragm at the other end which expands when the pressure in the pleural cavity is greater than atmospheric. Since this device does not vent fluid from the pleural cavity, the diaphragm does not provide any indication that venting is proceeding or that the pneumothorax has been resolved.

SUMMARY OF THE INVENTION

It is therefore an object of the present invention to provide an apparatus for draining fluids from a body cavity or introducing fluids into the cavity which positively locates the body cavity.

It is a further object of the present invention to provide a catheter apparatus in which a trocar is used to facilitate insertion of the catheter into the cavity, and in which a positive indication that the cavity has been located is provided without removing the trocar.

It is yet another object of the present invention to provide a catheter apparatus for draining fluids from a body cavity which permits the drainage to commence as soon as the apparatus enters the cavity.

These and other objects of the present invention will be apparent from the discussion below.

The present invention is therefore directed to an apparatus for transporting fluid either into or out of a body cavity having a fluctuating fluid pressure. The apparatus includes a conduit for entering the cavity, means for signaling the fluctuation of the fluid pressure in the cavity and means for transporting the fluid either into or out of the cavity. The signaling means provides a positive indication that the cavity has been located.

In one important embodiment, the means for entering the cavity comprises a catheter and a housing, having an inlet chamber and a discharge chamber, supporting the catheter in sealed fluid communication with the inlet chamber. An inlet port is provided in the housing having a seal for reversibly receiving a trocar and enabling at least a portion of the trocar to pass through the inlet chamber and the catheter. The inlet port seal is adapted for sealing the chamber both in the absence of the trocar and upon its insertion and withdrawal. In addition, an anti-reflux valve is mounted in the housing, coupling the inlet chamber to the discharge chamber. The anti-reflux valve permits fluid entering the inlet chamber to pass into the discharge chamber, but prevents any fluid from flowing back into the inlet chamber. Finally, and most importantly, this embodiment of the invention also includes means for signaling the fluctuating fluid pressure in the body cavity, to thereby positively signal that the body cavity has been located.

In another important embodiment of the invention an improvement in apparatus having a conduit with a removable trocar to aid body penetration is provided. The improvement comprises a furrow in the trocar which permits fluid flow through the conduit while the trocar is still in place in the conduit.

BRIEF DESCRIPTION OF THE DRAWINGS

The features of this invention which are believed to be novel are set forth with particularly in the appended claims. The invention, together with its objects and the advantages thereof, may be best understood by reference to the following description, taken in conjunction with the accompanying drawings, in which like reference numerals identify like elements in the several figures and in which:

FIG. **1** is a perspective view of an apparatus in accordance with the present invention, showing a trocar inserted into the inlet chamber of the apparatus;

FIG. **1**A is an elevation view of the rear of the apparatus of FIG. **1**;

FIG. **2** is a sectional view of the apparatus of FIG. **1**, taken along lines 2—2 of FIG. **1**A;

FIG. **2**A is a partial sectional view of a discharge chamber of an alternate embodiment of the apparatus of FIG. **1** in which a discharge port is provided for removing liquids from the discharge chamber;

FIG. **2**B is a partial sectional view of an alternate embodiment of the apparatus of FIG. **1** in which an auditory signal indicative of the fluctuation of fluid pressure is produced;

FIG. **3** is a plan view of the film valve utilized in the apparatus of FIG. **1**;

FIG. **4**A is a front elevation view of a furrowed trocar used in the practice of the present invention;

FIG. **4**B is an enlarged sectional a view of the trocar of FIG. **4**, taken along lines **4**B—**4**B. of FIG. **4**; and

FIGS. **5**A and **5**B are enlarged sectional views of the domed signal utilized in the apparatus of FIG. **1**.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The apparatus of the present invention is intended to have application in locating and in moving fluids either into or out of body cavities such as the pleural cavity, the signal epidural space, blood vessels, the gallbladder, the urinary bladder, the kidney pelvis, the brain ventricles and the brain subdural and epidural spaces. However, for purposes of illustration, the discussion below is directed primarily to an embodiment of the invention particularly suited for locating and draining fluids from the pleural cavity.

An apparatus **10** in accordance with the present invention is illustrated in FIGS. **1**, **1**A and **2**. The apparatus includes a housing **12** having an inlet chamber **14** and a discharge chamber **16**. Housing **12** is made of any rigid, impervious material, such as an acrylic, ABS, polystyrene or polyvinyl chloride plastic.

Apparatus **10** includes a manifold **17** having a first bore **18** extending through the manifold. Bore **18** is enlarged at its distal end to form a recess **20** to which a cannula **22** is affixed and sealed. A polyethylene are one preferred material. When this laminate is used, the plastic film is used with the polyethylene sides facing each other.

The edges of the two film members of the valve are heat sealed to each other with a tortuous profile **42** at the closed edges of the valve in order to prevent fluid leakage. The cracking resistance or threshold pressure at which the valve permits fluid to flow through it can be adjusted by varying the thickness of the plastic films and the length and width of the valve passage **43**.

Other anti-reflux valves which are sufficiently sensitive to reflux pressures to produce a discernible movement in the signaling means (described below) can be used. The "Heimlich valve" utilized in U.S. Pat. No. 4,664,660, described above in the Background of the Invention, for example, can be used in the present application but is less preferred than the film valve, because the Heimlich valve is far less sensitive to reflux pressures. The Heimlich valve is described in U.S. Pat. No. 3,463,159.

The space defined by bore **18** and bore **34** together comprise the inlet chamber **14** of the apparatus. The inlet chamber is separated from outlet chamber **16** by anti-reflux valve **40**.

Catheter **24** is formed of a stainless steel spring core **48** covered with an elastomer **50**, which preferably is present in the form of a silicone coating on the spring. The tip **52** of the catheter is preferably made of a resilient polymer, such as polyurethane or silicone rubber, and is bonded to the end of the spring core.

The tip of the catheter is open at **54** to permit trocar **28** to pass through. Elastomer covering **50** of the catheter may have a plurality of ports **56**, in addition to the opening at the tip, to facilitate draining between the exposed coils of spring **48**, and to insure that drainage proceeds in the event that opening **54** is blocked by expanding lung tissue.

Discharge chamber **16** has a series of elongated vent openings **58** (FIG. **1**A) located behind a protective cover **59**. Cover **59** is affixed to housing **12** along its sides and undercut to create a passages **61** at its top and bottom through which gaseous fluids vent to the atmosphere. A hydrophobic filter **60** is fixed in place behind the vent openings to prevent aqueous liquids from escaping the discharge chamber.

A self-adhesive strip **62** is attached to the exterior of the distal wall **64** of housing **12**. The adhesive strip is covered by a protective release sheet **66** which is removed prior to installation of the catheter, so that the strip can be applied to the patient's skin upon installation of the apparatus.

In addition, a seal cuff **68** is provided about catheter **24**, adjacent wall **64** of the apparatus which will abut the patient's body. Cuff **68**, which is preferably a sponge silicone or other resilient material, is intended to prevent tissue emphysema by sealing the perimeter of the catheter at the wound site where the catheter enters the body.

An absorbent pad **70** is positioned at the bottom and side of the discharge chamber to absorb small quantities of blood or other liquids which may pass through the apparatus when, as in the present embodiment, the principal fluid being drained is air. In an alternative embodiment, where significant amounts of liquid are to be drained, a collection port **72** (FIG. **2**A) is provided so that the fluid may be continuously withdrawn either under suction or by gravity.

Trocar **28**, which is illustrated in FIGS. **4**A and **4**B, includes a handle **74** molded or bonded to its proximal end. The trocar is sized so that its distal, pointed tip **76** protrudes beyond the tip **52** of catheter **24** when the trocar is fully inserted in the apparatus. Trocar **28** includes a furrow **78** which extends from trocar tip **76** through a point **79** near handle **74**. The furrow permits fluid communication between the tip **52** of the catheter and inlet chamber **34** while the trocar is inserted in the catheter. Furrow **78** thus permits fluids to be vented from the plural cavity immediately upon introduction of the catheter, without removal of the trocar. Furrow **78** also permits the signaling means (described below) to being functioning immediately upon entry of the catheter tip into the pleural cavity.

Means for signaling fluctuation of fluid pressure in the pleural cavity is provided in the form of a resilient signal dome **80** which is in fluid communication with inlet chamber **14** by way of bore **82** in manifold **17**. Signal dome **80** has an annular lip **84** which is sealingly affixed to the top **36** of manifold **17** so that the dome caps and closes off bore **82**. A clear cover **86** is affixed to the top **36** of manifold **17**, to define a signal chamber **88**. A vent hole **90** is formed in the cap to permit the signal dome to freely expand and contract within it.

Signal dome **80** is made of a resilient material, preferably silicone, with a shape and resilience such that the dome will collapse (FIG. **5B**) and return to its normal configuration (FIG. **5A**) with the fluctuating pressures in the pleural cavity, producing a pulsating visual indication of the pleural cavity pressures. In the embodiment illustrated, the signal dome is molded of silicone and is wider than it is high. It has a wall thickness of about 0.012 to 0.015 inches at its base tapering to about 0.007 inches at its apex, as illustrated in FIGS. **5A–5B**. In other applications the signal dome sensitivity may be adjusted as necessary by varying, for example, the material from which the signal dome is fabricated, as well as the size and shape of the dome, and its wall thickness and taper.

The functioning of the signal means depends on: 1) changes in pressure (differential pressures) in the body cavity of interest; 2) cracking resistance of the anti-reflux valve, and 3) sensitivity of the signal dome. Since the differential pressures of the body cavity are not controllable, the anti-reflux valve cracking resistance and/or the signal dome sensitivity must be matched to the differential pressures of the particular body cavity in order to produce the desired pulsating movement in the signal dome. For example, in an application in the epidural space, where the pressures are generally lower than in the pleural cavity, a more sensitive dome and/or an anti-reflux valve with a higher cracking resistance will be needed.

Naturally, the signaling means include any resilient structures which pulsate with fluctuating pressure in the system. For example, a flat resilient membrane which expands with increasing pressure and collapses back to a generally flat position on decreasing pressure would provide a satisfactory alternative to the signal dome illustrated above.

In an alternate embodiment, an auditory signal is provided, either alone or in conjunction with the visual signal described above. Thus, there is illustrated in FIG. **2B** an embodiment in which a hole **92** is made in cover **59**, passages **61** are closed off and a low velocity whistle **94** is affixed to the cover so that the opening in the whistle **96** is aligned with the hole in cover **59**. Thus, the air passing though the discharge chamber will exit through the whistle, producing a rhythmic auditory signal indicative of the fluctuating pressure in the pleural cavity before the pneumothorax is resolved and either a continuous signal or silence when the pneumothorax is resolved. The loudness and sensitivity of the signal will be a function of the cracking resistance of the anti-reflux valve **40**, the resistance to air movement of filter **60**, and the sensitivity of the whistle.

In yet another alternative embodiment of the present invention, once the catheter is in place in the pleural cavity, a cannula sized to fit within the catheter **24** is passed through seal **30**, through bore **18** and into the catheter. A vacuum is then applied to the cannula to rapidly remove fluids from the pleural space.

In treating traumatic or idiopathic pneumothorax, or for evacuating air from the chest following thoracic surgery or percutaneous lung biopsy, the apparatus **10** may be used as follow:

1. If the apparatus is supplied in sterile kit form, the apparatus is removed from the kit under sterile conditions and the trocar is inserted through inlet port **26** so that its tip protrudes from the tip **52** of catheter **24**.
2. A site is selected, preferably in the second interspace at the midclavicular line.
3. The site is prepared with an antimicrobial agent and a drape is placed over the site.
4. Local anesthesia is applied at the site.
5. Using a scalpel, a small skin incision is made at the site.
6. The release sheet is removed from the adhesive strip.
7. With the thumb over the trocar handle, the device is positioned at the incision and the trocar/catheter assembly is passed into the pleural space immediately above the superior border of the rib. Catheter **24** is maintained in a straight condition by the trocar, so that the catheter will remain rigid and resist bending upon insertion into the body cavity. In addition, spring **48** prevents any significant compaction of the catheter on the trocar during the insertion process.
8. When the catheter tip enters the pleural space, signal dome **80** will begin to pulsate. When this occurs, the trocar is removed to increase the rate of fluid flow and prevent damage to the lung. The continuing pulsating movement of the signal dome is observed as the air is evacuated from the pleural cavity and the lung begins to reinflate.
9. When the pneumothorax is resolved and the lung is reinflated the signal dome stops pulsating and the apparatus can be removed.

While the present apparatus has been described in connection with the removal of fluids from the pleural cavity, the device may be used in a multitude of other applications, both for removing fluids and for introducing fluids. For example, it may be used in locating and in moving fluids either into or out of body cavities such as the pleural cavity, the signal epidural space, blood vessels, the gall bladder, the urinary bladder, the kidney pelvis, the brain ventricles, and the brain subdural and epidural spaces. In introducing fluids, the catheter is inserted in the cavity and then a cannula is introduced through seal **30** and into the catheter, whereby the fluids to be introduced are passed through the cannula by way of a syringe or other conventional injection device.

For example, in administering an epidural block, the anesthesiologist would position the trocar/catheter tip in the epidural space relying upon the pulsating signal dome to positively indicate that the epidural space has been located. He or she would then remove the trocar and insert a syringe containing an appropriate anesthetic and inject the anesthetic into the epidural space.

While particular embodiments of the invention have been shown and described, it will be obvious to those skilled in the art that various changes and modifications may be made therein without departing from the spirit and scope of the invention and, therefore, it is intended in the appended claims to cover all such changes and modifications which fall within the true spirit and scope of the invention.

What we claim is:

1. An apparatus for locating a body cavity having a fluctuating fluid pressure and transporting fluid either into or out of the cavity comprising:

a conduit comprising a flexible catheter having a spring covered by an elastomeric coating for entering the cavity;

means, in communication with said conduit, for producing a pulsating indication corresponding to the fluctuation of the fluid pressure in the cavity when said conduit enters the body cavity; and

means, in communication with said conduit, for transporting fluid either into or out of the cavity.

2. The apparatus of claim **1** wherein said conduit serves as said transporting means.

3. The apparatus of claim **1** wherein said signaling means comprises an elastomeric element which pulsates in response to the fluctuating fluid pressure in the body cavity.

4. An apparatus for locating a body cavity having a fluctuating fluid pressure and transporting fluid either in or out of the cavity comprising:

a conduit;

a housing having an inlet chamber and a discharge chamber, said housing supporting said conduit in sealed fluid communication with said inlet chamber;

an inlet port in said housing having inlet port sealing means for reversibly receiving a trocar and enabling at least a portion of a trocar to pass through said inlet chamber and said conduit, said inlet port sealing means being adapted for sealing said chamber both in the absence of a trocar and upon insertion and withdrawal of a trocar,

anti-reflux valve means mounted in said housing, coupling said inlet chamber with said discharge chamber, for permitting fluid entering said inlet chamber to pass into said discharge chamber while preventing fluid from flowing from said discharge chamber into said inlet chamber and

means for producing a pulsating visual indication corresponding to the fluctuating fluid pressure in the body cavity upon entry of the conduit into a body cavity.

5. The apparatus of claim **4** including a trocar positioned in said inlet port and passing through said inlet chamber and said conduit, said trocar having a furrow which permits fluid flow through said conduit and said inlet chamber.

6. The apparatus of claim **4** wherein said conduit is a flexible catheter comprising a spring covered by an elastomeric coating.

7. The apparatus of claim **4** wherein a resilient cuff is provided at the base of said conduit to seal against the body.

8. The apparatus of claim **4** wherein said anti-reflux valve is a film valve comprising two pieces of plastic film closed along their edges.

9. The apparatus of claim **8** wherein said closed edges of said film valve have a tortuous profile.

10. The apparatus of claim **8** wherein said plastic film pieces comprise a nylon/polyethylene laminate.

11. The apparatus of claim **4** wherein a collection port is provided in said discharge chamber for withdrawing liquids therefrom.

12. The apparatus of claim **4** wherein said signaling means comprises an elastomeric element which pulsates in response to the fluctuating fluid pressure in the body cavity.

13. The apparatus of claim **12** wherein said elastomeric element is dome-shaped and the thickness of the wall of said dome-shaped elastomeric element gradually tapers to its apex.

14. The apparatus of claim **4** wherein said signaling means comprises a whistle.

15. The apparatus of claim **4** including a cannula passing through said inlet port sealing means and said inlet chamber and into said conduit for transporting fluid either into or out of the cavity.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 5,356,386

DATED : Oct. 18, 1994

INVENTOR(S) : Edward M. Goldberg, Glencoe; Lev Melinyshyn, Mt. Prospect; Michael Jaron, Des Plaines; Jeffrey M. Stupar, Chicago; all of Ill.

It is certified that error appears in the above-indentified patent and that said Letters Patent is hereby corrected as shown below:

In Col. 1, line 29, after the word "the", (1st occurrence) insert --catheter so that the sharp tip of the trocar extends beyond the distal--.

In Col. 2, line 67, substitute --particularity-- for "particularly".

In Col. 3, line 35, substitute --spinal-- for "signal".

Signed and Sealed this

Fourteenth Day of February, 1995

Attest:

Bruce Lehman

BRUCE LEHMAN

Attesting Officer

Commissioner of Patents and Trademarks

US006506148B2

(12) **United States Patent**
Loos

(10) **Patent No.: US 6,506,148 B2**
(45) **Date of Patent: Jan. 14, 2003**

(54) **NERVOUS SYSTEM MANIPULATION BY ELECTROMAGNETIC FIELDS FROM MONITORS**

(76) Inventor: **Hendricus G. Loos**, 3019 Cresta Way, Laguna Beach, CA (US) 92651

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 8 days.

(21) Appl. No.: **09/872,528**

(22) Filed: **Jun. 1, 2001**

(65) **Prior Publication Data**

US 2002/0188164 A1 Dec. 12, 2002

(51) **Int. Cl.**^{7} **A61N 2/00**; A61B 5/04; A61M 21/00

(52) **U.S. Cl.** **600/27**; 600/545

(58) **Field of Search** 600/9–27, 545; 313/419; 324/318; 378/901; 434/236

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,592,965	A	*	7/1971	Diaz	313/419
4,800,893	A	*	1/1989	Ross et al.	600/545
5,169,380	A		12/1992	Brennan	600/26
5,304,112	A	*	4/1994	Mrklas et al.	434/236
5,400,383	A	*	3/1995	Yassa et al.	378/901
5,412,419	A	*	5/1995	Ziarati	324/318
5,450,859	A	*	9/1995	Litovitz	600/9
5,782,874	A		7/1998	Loos	607/2
5,800,481	A		9/1998	Loos	607/100
5,899,922	A		5/1999	Loos	607/2
5,935,054	A		8/1999	Loos	600/9
6,017,302	A		1/2000	Loos	600/28
6,081,744	A		6/2000	Loos	607/2
6,091,994	A		7/2000	Loos	607/100
6,167,304	A		12/2000	Loos	607/2
6,238,333	B1		5/2001	Loos	600/9

OTHER PUBLICATIONS

N.Wiener "Nonlinear problems in random theory" p.71–72 John Wiley New York 1958.

M.Hutchison "Megabrain" p.232–3 Ballantine Books New York 1991.

C.A.Terzuolo and T.H.Bullock "Measurement of imposed voltage gradient adequate to modulate neuronal firing" Proc. Nat. Acad. Sci, Physiology 42,687–94, 1956.

O.Kellogg"Foundations of Potential Theory"p. 191 Dover, 1953.

P.M.Morse and H.Feshbach"Methods of Theoretical Physics"p. 1267 McGraw–Hill New York, 1953.

* cited by examiner

Primary Examiner—Eric F. Winakur
Assistant Examiner—Nikita R Veniaminov

(57) **ABSTRACT**

Physiological effects have been observed in a human subject in response to stimulation of the skin with weak electromagnetic fields that are pulsed with certain frequencies near ½ Hz or 2.4 Hz, such as to excite a sensory resonance. Many computer monitors and TV tubes, when displaying pulsed images, emit pulsed electromagnetic fields of sufficient amplitudes to cause such excitation. It is therefore possible to manipulate the nervous system of a subject by pulsing images displayed on a nearby computer monitor or TV set. For the latter, the image pulsing may be imbedded in the program material, or it may be overlaid by modulating a video stream, either as an RF signal or as a video signal. The image displayed on a computer monitor may be pulsed effectively by a simple computer program. For certain monitors, pulsed electromagnetic fields capable of exciting sensory resonances in nearby subjects may be generated even as the displayed images are pulsed with subliminal intensity.

14 Claims, 9 Drawing Sheets

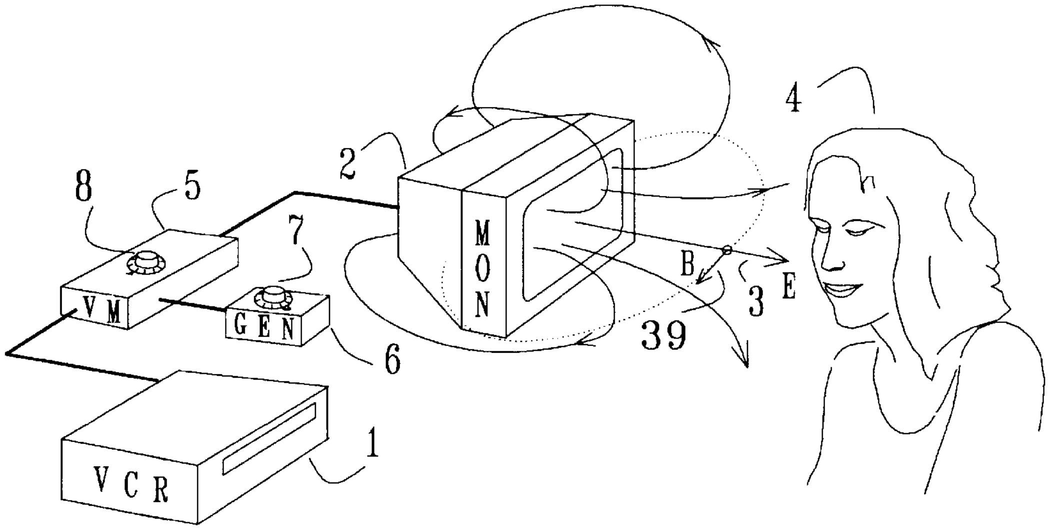

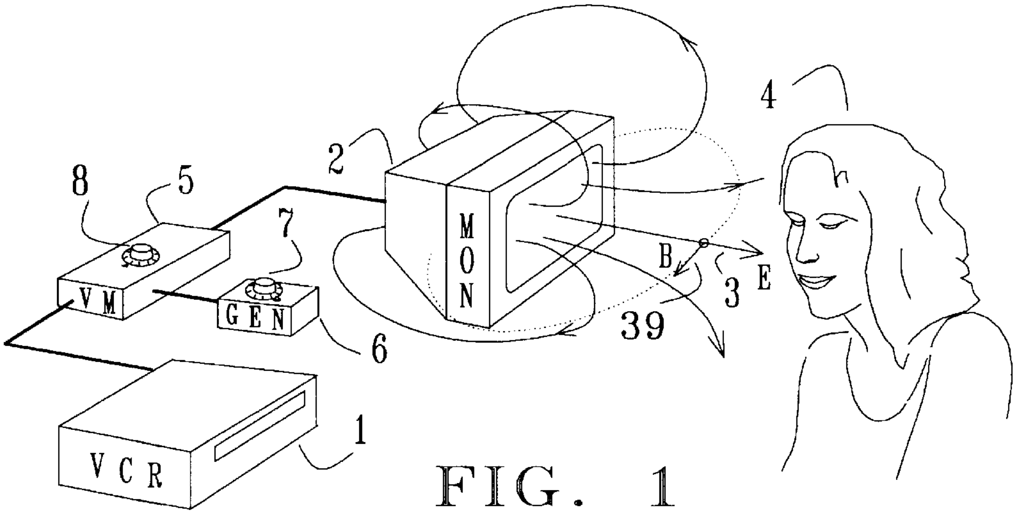
8
5
7
2
4
VM
GEN
MON
VCR
B
E
3
39
6
1

FIG. 1

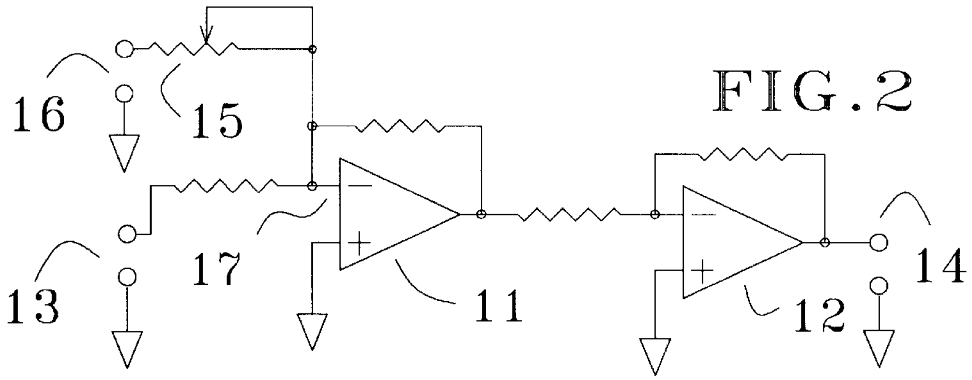
16
15
13
17
11
12
14

FIG. 2

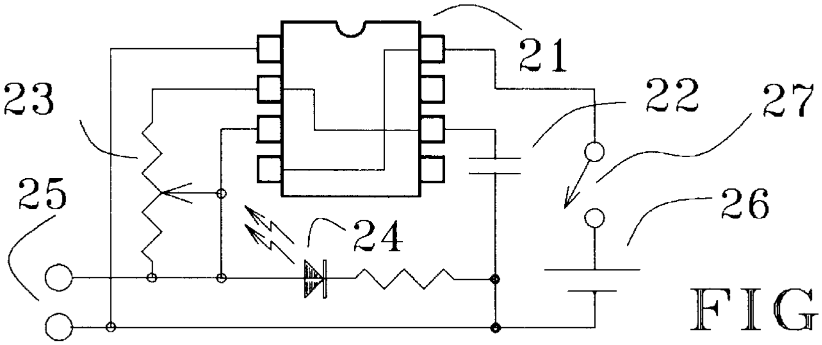
21
22
27
23
25
24
26

FIG. 3

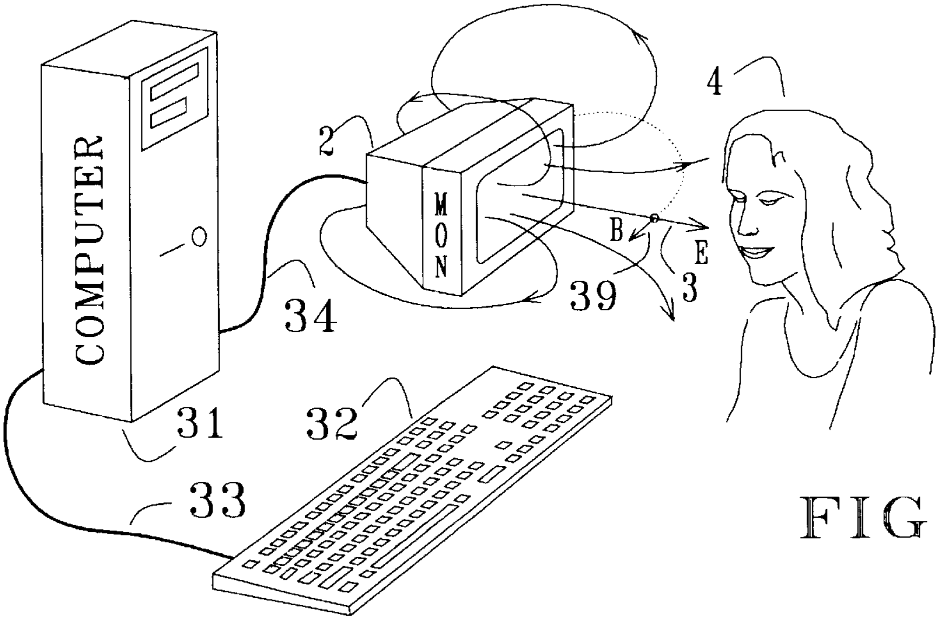
COMPUTER
MON
B
E

FIG. 4

FIG. 5

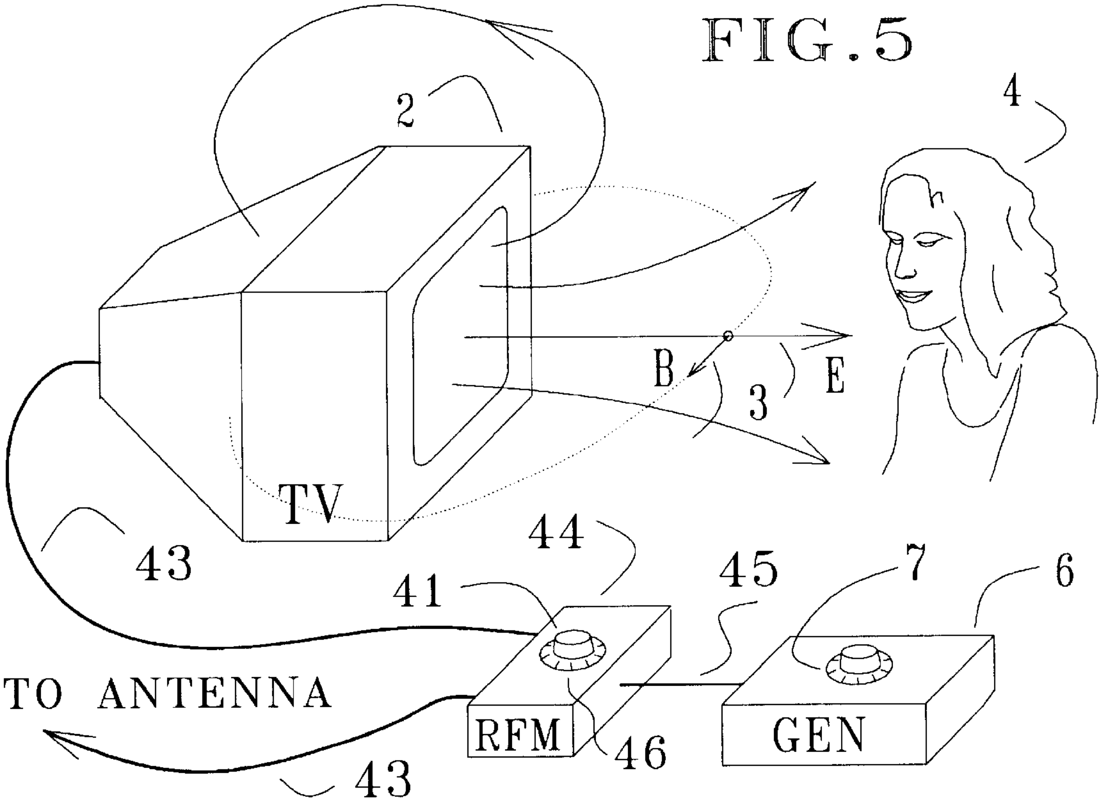
B
E
TV
TO ANTENNA
RFM
GEN

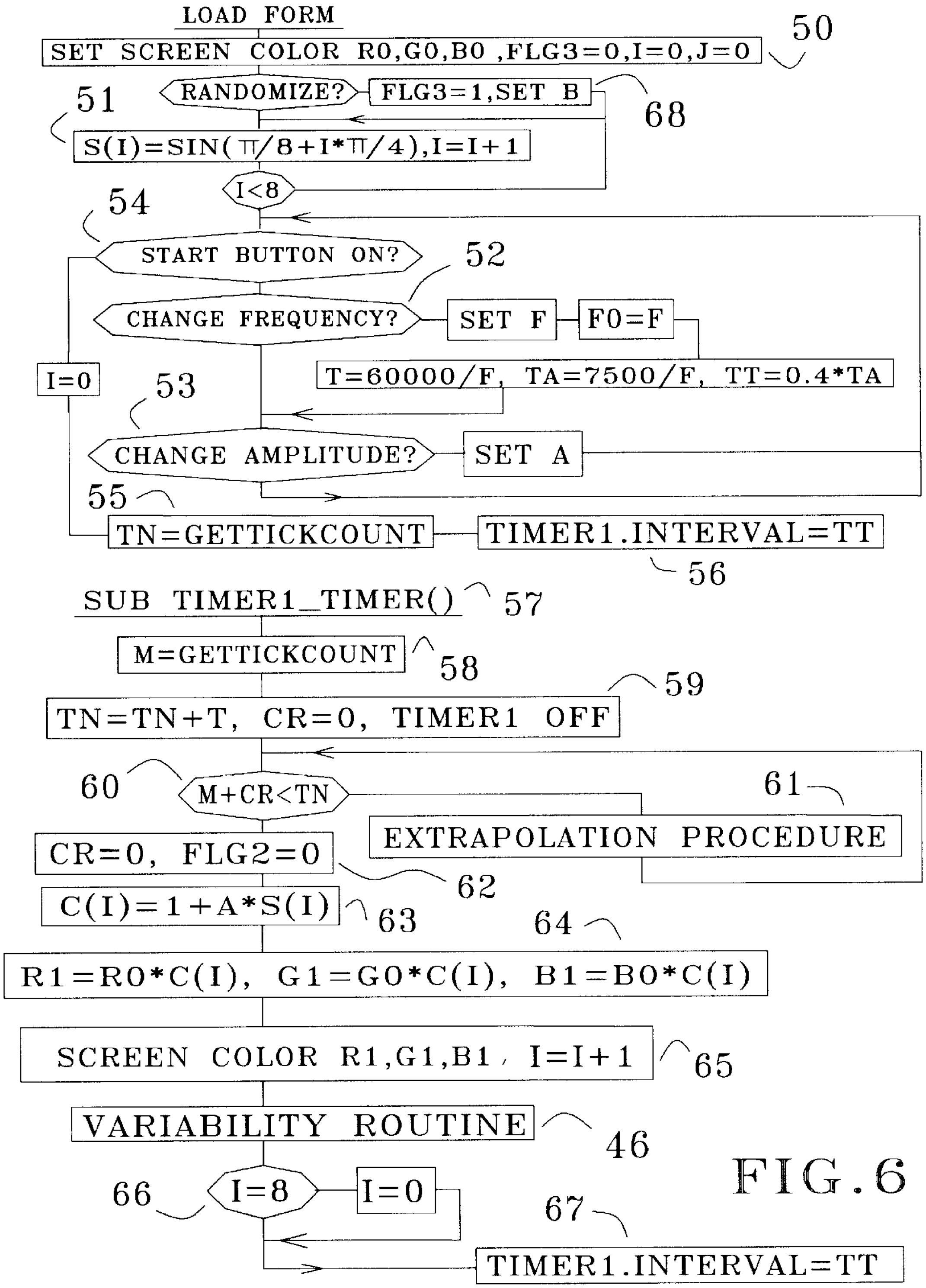
LOAD FORM
SET SCREEN COLOR R0,G0,B0 ,FLG3=0,I=0,J=0
50
RANDOMIZE?
FLG3=1,SET B
68
51
S(I)=SIN(π/8+I*π/4),I=I+1
I<8
54
START BUTTON ON?
52
CHANGE FREQUENCY?
SET F
F0=F
I=0
T=60000/F, TA=7500/F, TT=0.4*TA
53
CHANGE AMPLITUDE?
SET A
55
TN=GETTICKCOUNT
TIMER1.INTERVAL=TT
56
SUB TIMER1_TIMER()
57
M=GETTICKCOUNT
58
TN=TN+T, CR=0, TIMER1 OFF
59
60
M+CR<TN
61
EXTRAPOLATION PROCEDURE
CR=0, FLG2=0
62
C(I)=1+A*S(I)
63
64
R1=R0*C(I), G1=G0*C(I), B1=B0*C(I)
SCREEN COLOR R1,G1,B1, I=I+1
65
VARIABILITY ROUTINE
46
66
I=8
I=0
67
TIMER1.INTERVAL=TT

FIG.6

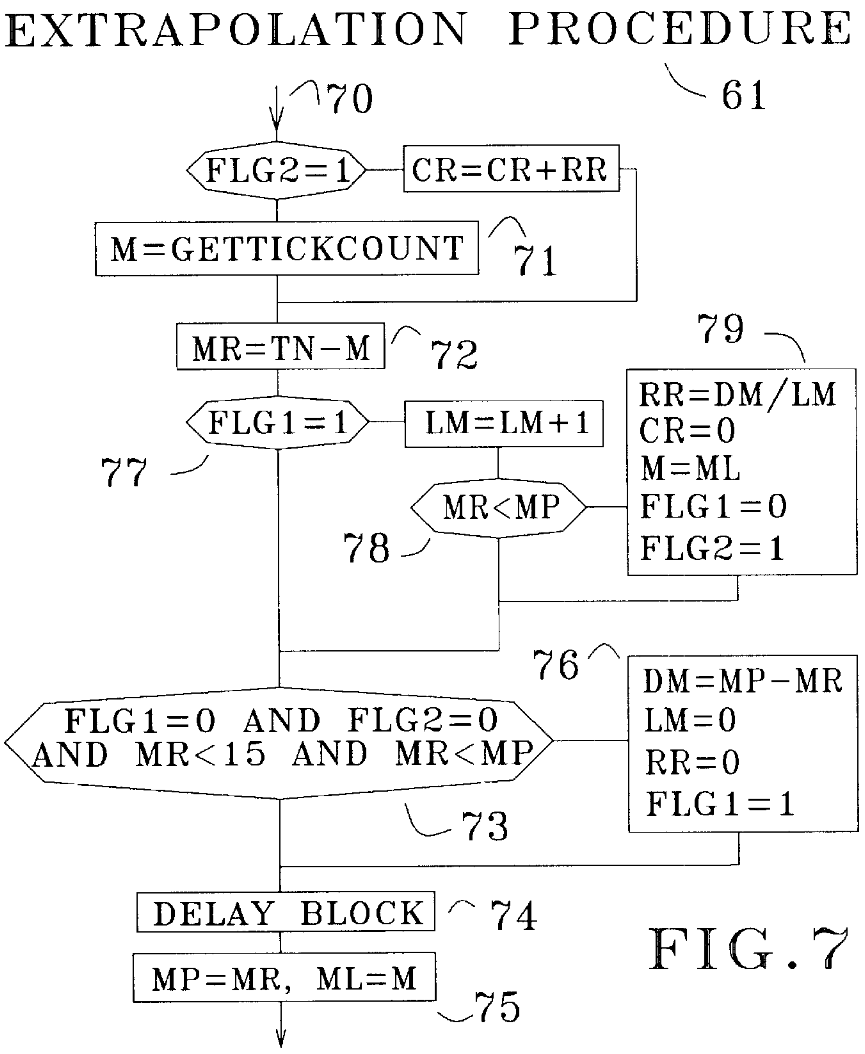
EXTRAPOLATION PROCEDURE
61
70
FLG2=1
CR=CR+RR
M=GETTICKCOUNT
71
MR=TN−M
72
79
FLG1=1
LM=LM+1
RR=DM/LM
CR=0
M=ML
FLG1=0
FLG2=1
77
MR<MP
78
76
FLG1=0 AND FLG2=0 AND MR<15 AND MR<MP
DM=MP−MR
LM=0
RR=0
FLG1=1
73
DELAY BLOCK
74
MP=MR, ML=M
75

FIG.7

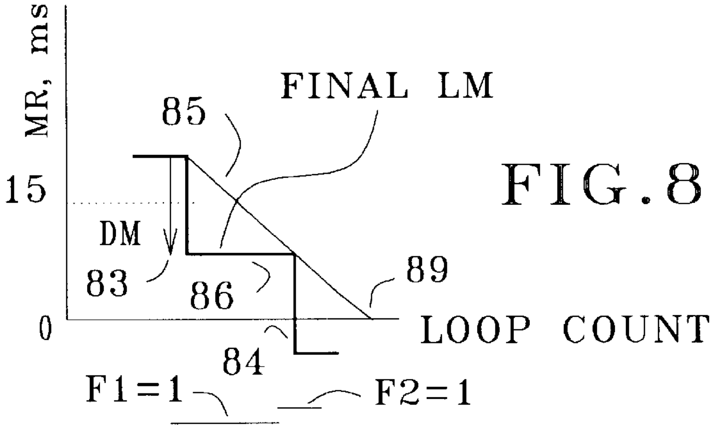
MR, ms
85
FINAL LM
15
DM
83
86
89
0
LOOP COUNT
84
F1=1
F2=1

FIG.8

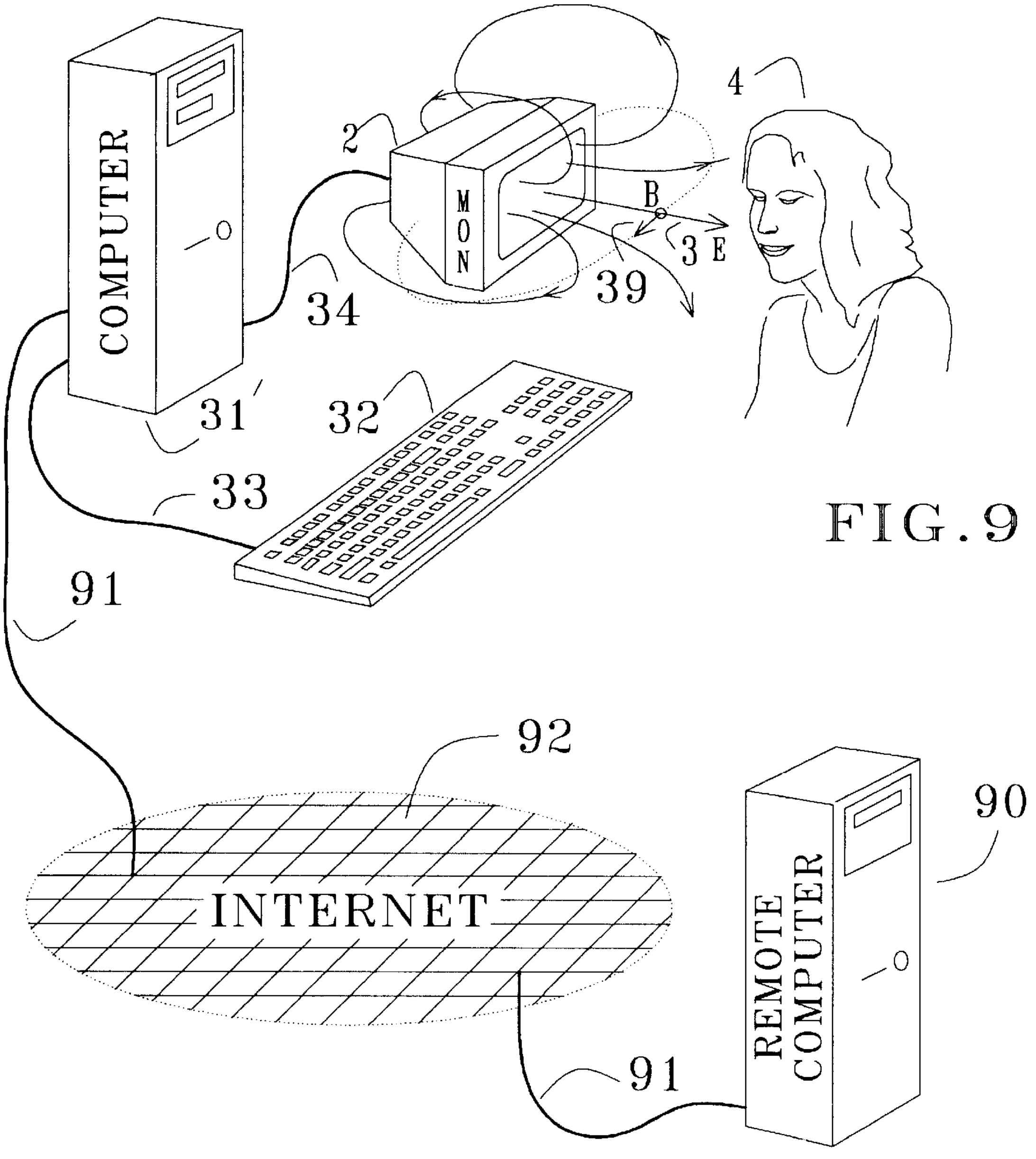
COMPUTER
MON
B
E
INTERNET
REMOTE COMPUTER

FIG. 9

FIG. 10

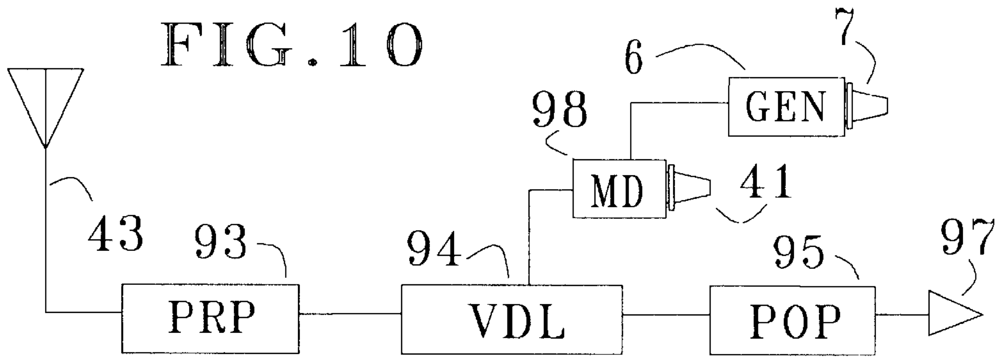
GEN
MD
PRP
VDL
POP

FIG. 11

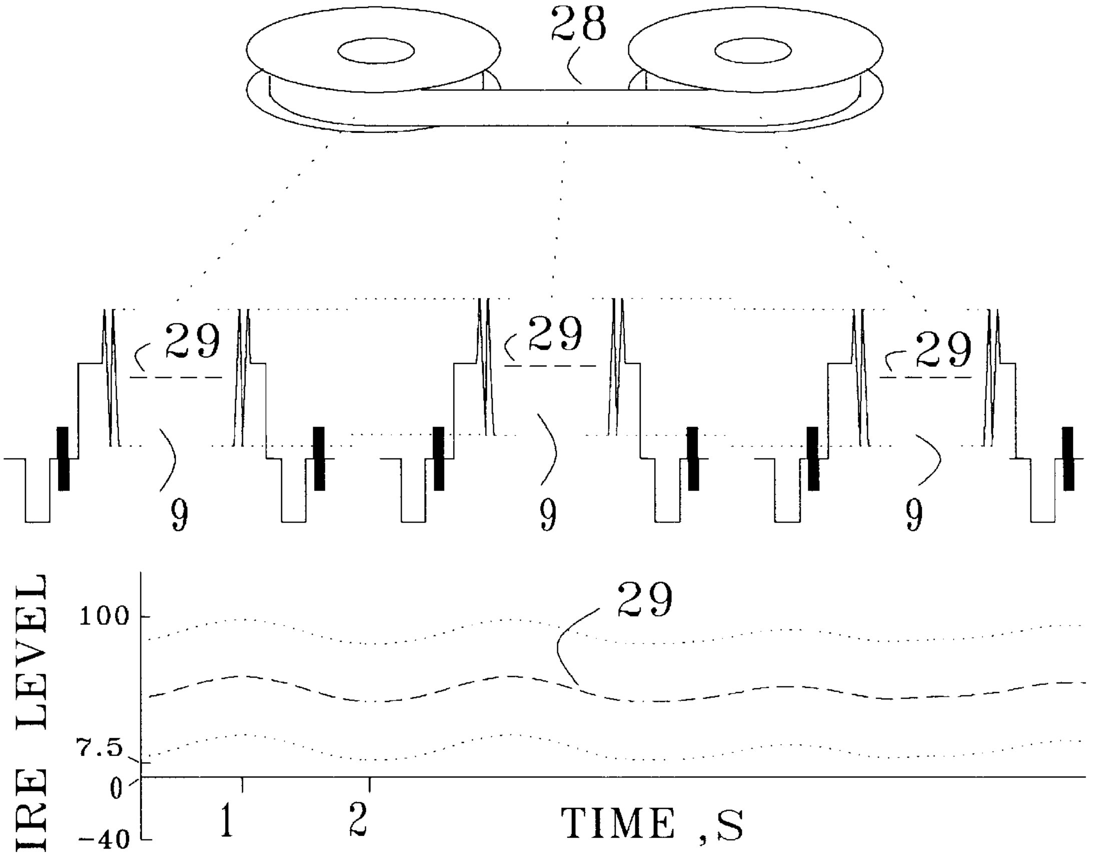
28
29
29
29
9
9
9
IRE LEVEL
100
29
7.5
0
-40
1
2
TIME ,S

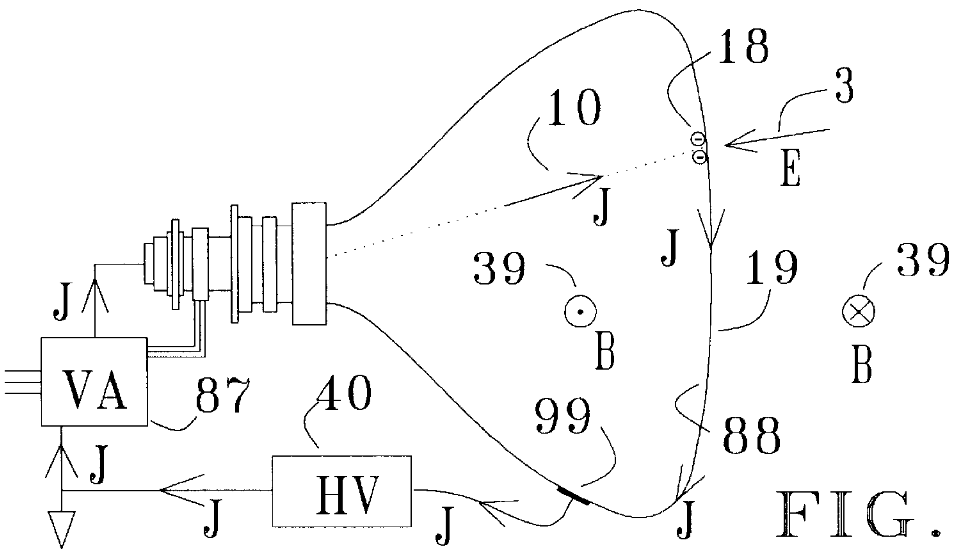
18
3
10
E
J
J
J
39
19
39
B
B
J
VA
87
40
99
88
J
HV
J
J
J

FIG. 14

FIG. 12

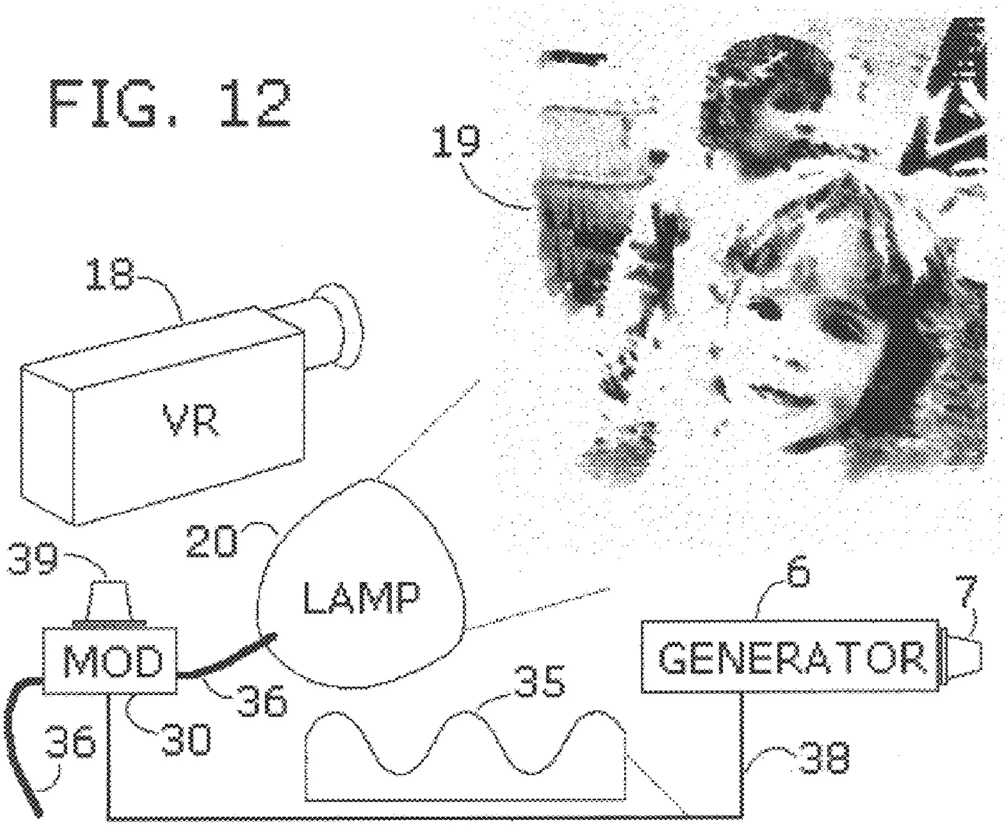
19
18
VR
20
LAMP
39
MOD
36
30
36
35
6
GENERATOR
7
38

VARIABILITY ROUTINE

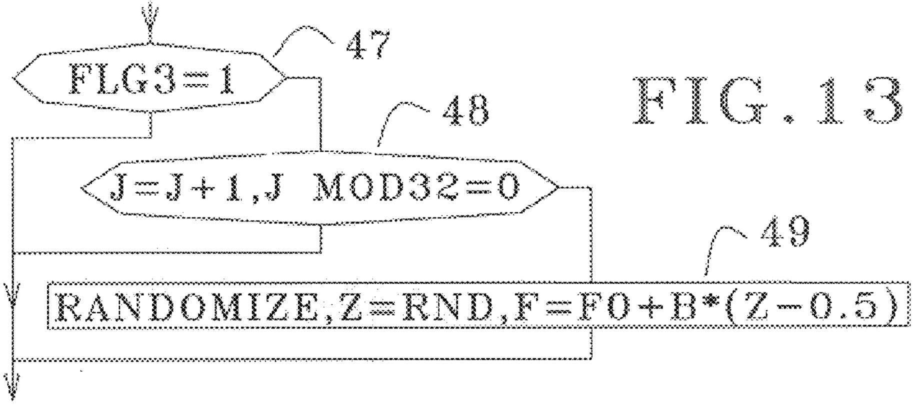
FLG3=1
47
48
J=J+1,J MOD32=0
49
RANDOMIZE,Z=RND,F=F0+B*(Z−0.5)

FIG. 13

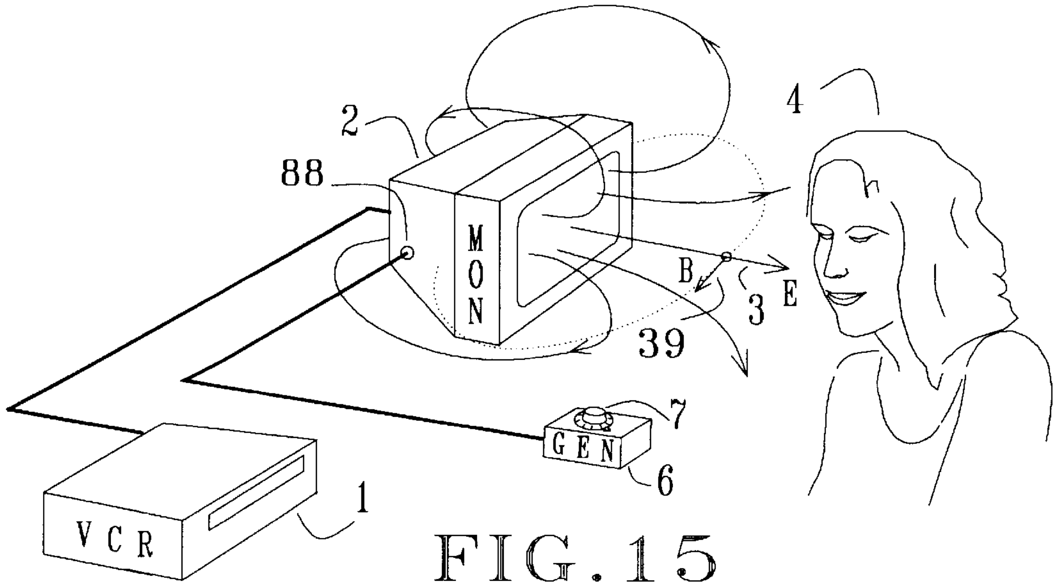
2
4
88
MON
B
E
3
39
7
GEN
6
VCR
1

FIG. 15

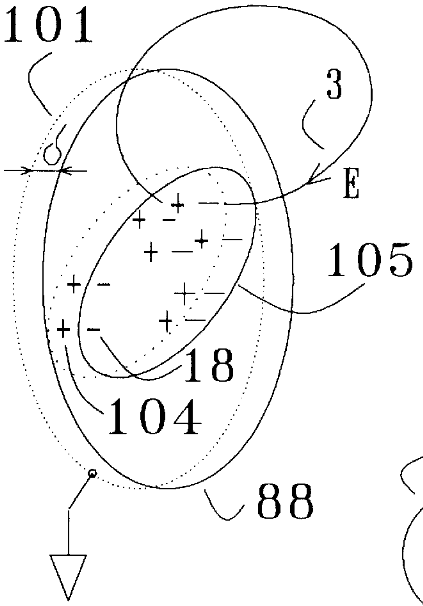
101
3
E
105
18
104
88

FIG. 16

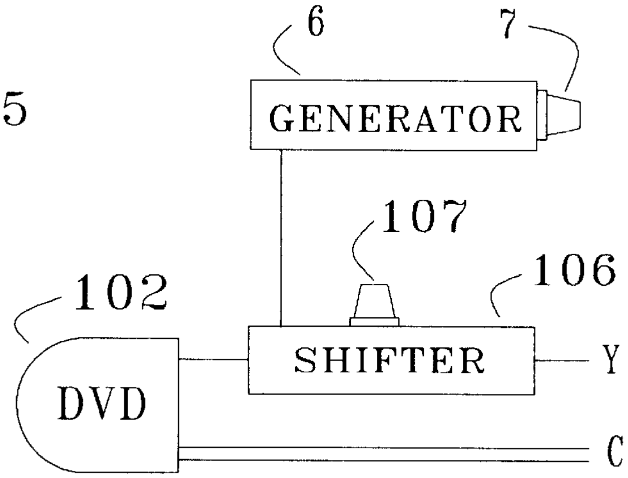
6
7
GENERATOR
107
106
102
SHIFTER
Y
DVD
C

FIG. 17

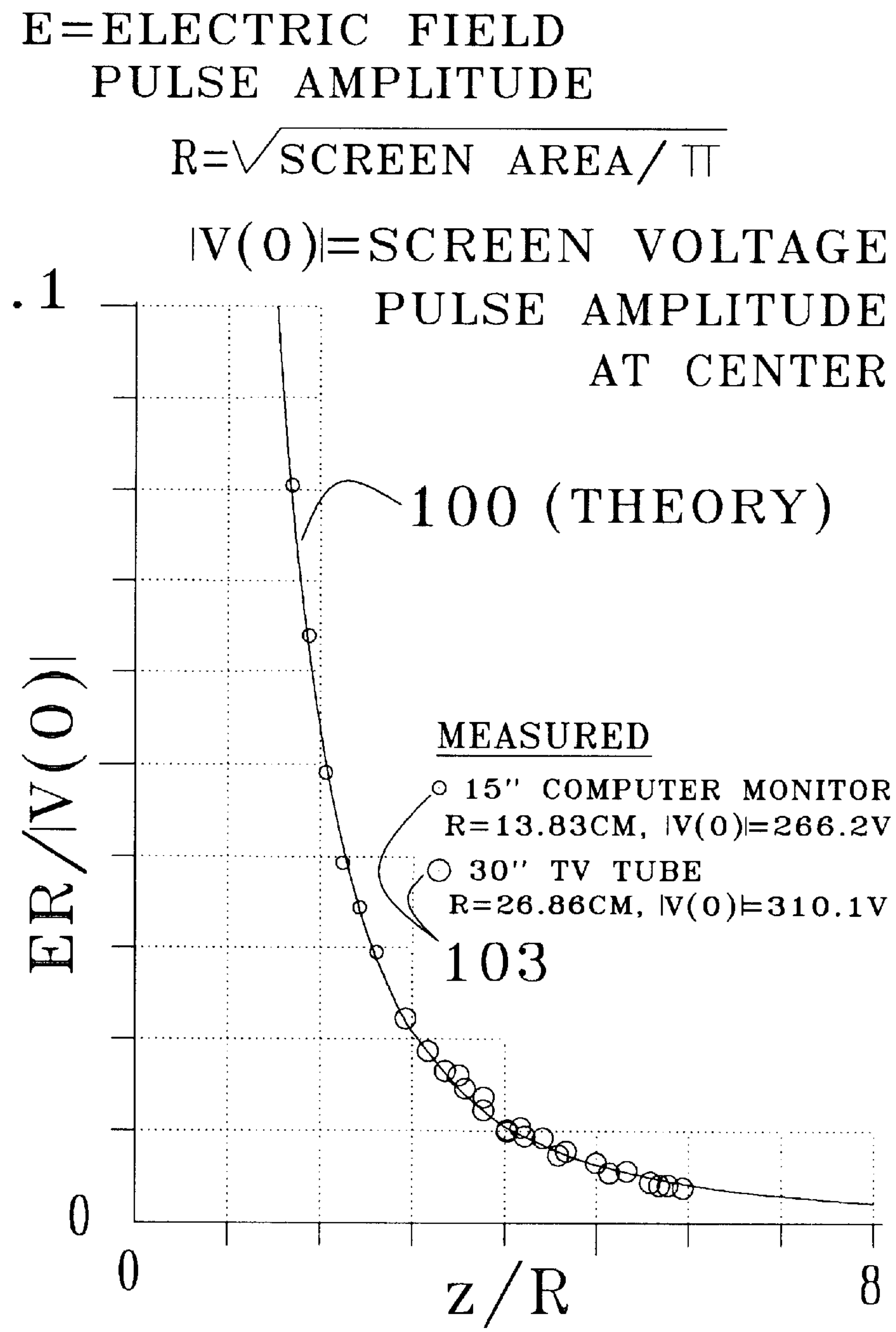

ELECTRIC FIELD AT A DISTANCE z FROM CRT SCREEN ON CENTER LINE

FIG. 18

NERVOUS SYSTEM MANIPULATION BY ELECTROMAGNETIC FIELDS FROM MONITORS

BACKGROUND OF THE INVENTION

The invention relates to the stimulation of the human nervous system by an electromagnetic field applied externally to the body. A neurological effect of external electric fields has been mentioned by Wiener (1958), in a discussion of the bunching of brain waves through nonlinear interactions. The electric field was arranged to provide “a direct electrical driving of the brain”. Wiener describes the field as set up by a 10 Hz alternating voltage of 400 V applied in a room between ceiling and ground. Brennan (1992) describes in U.S. Pat. No. 5,169,380 an apparatus for alleviating disruptions in circadian rythms of a mammal, in which an alternating electric field is applied across the head of the subject by two electrodes placed a short distance from the skin.

A device involving a field electrode as well as a contact electrode is the “Graham Potentializer” mentioned by Hutchison (1991). This relaxation device uses motion, light and sound as well as an alternating electric field applied mainly to the head. The contact electrode is a metal bar in Ohmic contact with the bare feet of the subject, and the field electrode is a hemispherical metal headpiece placed several inches from the subject’s head.

In these three electric stimulation methods the external electric field is applied predominantly to the head, so that electric currents are induced in the brain in the physical manner governed by electrodynamics. Such currents can be largely avoided by applying the field not to the head, but rather to skin areas away from the head. Certain cutaneous receptors may then be stimulated and they would provide a signal input into the brain along the natural pathways of afferent nerves. It has been found that, indeed, physiological effects can be induced in this manner by very weak electric fields, if they are pulsed with a frequency near ½ Hz. The observed effects include ptosis of the eyelids, relaxation, drowziness, the feeling of pressure at a centered spot on the lower edge of the brow, seeing moving patterns of dark purple and greenish yellow with the eyes closed, a tonic smile, a tense feeling in the stomach, sudden loose stool, and sexual excitement, depending on the precise frequency used, and the skin area to which the field is applied. The sharp frequency dependence suggests involvement of a resonance mechanism.

It has been found that the resonance can be excited not only by externally applied pulsed electric fields, as discussed in U.S. Pat. Nos. 5,782,874, 5,899,922, 6,081,744, and 6,167,304, but also by pulsed magnetic fields, as described in U.S. Pat. Nos. 5,935,054 and 6,238,333, by weak heat pulses applied to the skin, as discussed in U.S. Pat. Nos. 5,800,481 and 6,091,994, and by subliminal acoustic pulses, as described in U.S. Pat. No. 6,017,302. Since the resonance is excited through sensory pathways, it is called a sensory resonance. In addition to the resonance near ½ Hz, a sensory resonance has been found near 2.4 Hz. The latter is characterized by the slowing of certain cortical processes, as discussed in the ’481, ’922, ’302, ’744, ’944, and ’304 patents.

The excitation of sensory resonances through weak heat pulses applied to the skin provides a clue about what is going on neurologically. Cutaneous temperature-sensing receptors are known to fire spontaneously. These nerves spike somewhat randomly around an average rate that depends on skin temperature. Weak heat pulses delivered to the skin in periodic fashion will therefore cause a slight frequency modulation (fm) in the spike patterns generated by the nerves. Since stimulation through other sensory modalities results in similar physiological effects, it is believed that frequency modulation of spontaneous afferent neural spiking patterns occurs there as well.

It is instructive to apply this notion to the stimulation by weak electric field pulses administered to the skin. The externally generated fields induce electric current pulses in the underlying tissue, but the current density is much too small for firing an otherwise quiescent nerve. However, in experiments with adapting stretch receptors of the crayfish, Terzuolo and Bullock (1956) have observed that very small electric fields can suffice for modulating the firing of already active nerves. Such a modulation may occur in the electric field stimulation under discussion.

Further understanding may be gained by considering the electric charges that accumulate on the skin as a result of the induced tissue currents. Ignoring thermodynamics, one would expect the accumulated polarization charges to be confined strictly to the outer surface of the skin. But charge density is caused by a slight excess in positive or negative ions, and thermal motion distributes the ions through a thin layer. This implies that the externally applied electric field actually penetrates a short distance into the tissue, instead of stopping abruptly at the outer skin surface. In this manner a considerable fraction of the applied field may be brought to bear on some cutaneous nerve endings, so that a slight modulation of the type noted by Terzuolo and Bullock may indeed occur.

The mentioned physiological effects are observed only when the strength of the electric field on the skin lies in a certain range, called the effective intensity window. There also is a bulk effect, in that weaker fields suffice when the field is applied to a larger skin area. These effects are discussed in detail in the ’922 patent.

Since the spontaneous spiking of the nerves is rather random and the frequency modulation induced by the pulsed field is very shallow, the signal to noise ratio (S/N) for the fm signal contained in the spike trains along the afferent nerves is so small as to make recovery of the fm signal from a single nerve fiber impossibile. But application of the field over a large skin area causes simultaneous stimulation of many cutaneous nerves, and the fm modulation is then coherent from nerve to nerve. Therefore, if the afferent signals are somehow summed in the brain, the fm modulations add while the spikes from different nerves mix and interlace. In this manner the S/N can be increased by appropriate neural processing. The matter is discussed in detail in the ’874 patent. Another increase in sensitivity is due to involving a resonance mechanism, wherein considerable neural circuit oscillations can result from weak excitations.

An easily detectable physiological effect of an excited ½ Hz sensory resonance is ptosis of the eyelids. As discussed in the ’922 patent, the ptosis test involves first closing the eyes about half way. Holding this eyelid position, the eyes are rolled upward, while giving up voluntary control of the eyelids. The eyelid position is then determined by the state of the autonomic nervous system. Furthermore, the pressure excerted on the eyeballs by the partially closed eyelids increases parasympathetic activity. The eyelid position thereby becomes somewhat labile, as manifested by a slight flutter. The labile state is sensitive to very small shifts in autonomic state. The ptosis influences the extent to which the pupil is hooded by the eyelid, and thus how much light is admitted to the eye. Hence, the depth of the ptosis is seen by the subject, and can be graded on a scale from 0 to 10.

In the initial stages of the excitation of the ½ Hz sensory resonance, a downward drift is detected in the ptosis frequency, defined as the stimulation frequency for which maximum ptosis is obtained. This drift is believed to be caused by changes in the chemical milieu of the resonating neural circuits. It is thought that the resonance causes perturbations of chemical concentrations somewhere in the brain, and that these perturbations spread by diffusion to nearby resonating circuits. This effect, called "chemical detuning", can be so strong that ptosis is lost altogether when the stimulation frequency is kept constant in the initial stages of the excitation. Since the stimulation then falls somewhat out of tune, the resonance decreases in amplitude and chemical detuning eventually diminishes. This causes the ptosis frequency to shift back up, so that the stimulation is more in tune and the ptosis can develop again. As a result, for fixed stimulation frequencies in a certain range, the ptosis slowly cycles with a frequency of several minutes. The matter is discussed in the '302 patent.

The stimulation frequencies at which specific physiological effects occur depend somewhat on the autonomic nervous system state, and probably on the endocrine state as well.

Weak magnetic fields that are pulsed with a sensory resonance frequency can induce the same physiological effects as pulsed electric fields. Unlike the latter however, the magnetic fields penetrate biological tissue with nearly undiminished strength. Eddy currents in the tissue drive electric charges to the skin, where the charge distributions are subject to thermal smearing in much the same way as in electric field stimulation, so that the same physiological effects develop. Details are discussed in the '054 patent.

SUMMARY

Computer monotors and TV monitors can be made to emit weak low-frequency electromagnetic fields merely by pulsing the intensity of displayed images. Experiments have shown that the ½ Hz sensory resonance can be excited in this manner in a subject near the monitor. The 2.4 Hz sensory resonance can also be excited in this fashion. Hence, a TV monitor or computer monitor can be used to manipulate the nervous system of nearby people.

The implementations of the invention are adapted to the source of video stream that drives the monitor, be it a computer program, a TV broadcast, a video tape or a digital video disc (DVD).

For a computer monitor, the image pulses can be produced by a suitable computer program. The pulse frequency may be controlled through keyboard input, so that the subject can tune to an individual sensory resonance frequency. The pulse amplitude can be controlled as well in this manner. A program written in Visual Basic(R) is particularly suitable for use on computers that run the Windows 95(R) or Windows 98(R) operating system. The structure of such a program is described. Production of periodic pulses requires an accurate timing procedure. Such a procedure is constructed from the GetTimeCount function available in the Application Program Interface (API) of the Windows operating system, together with an extrapolation procedure that improves the timing accuracy.

Pulse variability can be introduced through software, for the purpose of thwarting habituation of the nervous system to the field stimulation, or when the precise resonance frequency is not known. The variability may be a pseudorandom variation within a narrow interval, or it can take the form of a frequency or amplitude sweep in time. The pulse variability may be under control of the subject.

The program that causes a monitor to display a pulsing image may be run on a remote computer that is connected to the user computer by a link; the latter may partly belong to a network, which may be the Internet.

For a TV monitor, the image pulsing may be inherent in the video stream as it flows from the video source, or else the stream may be modulated such as to overlay the pulsing. In the first case, a live TV broadcast can be arranged to have the feature imbedded simply by slightly pulsing the illumination of the scene that is being broadcast. This method can of course also be used in making movies and recording video tapes and DVDs.

Video tapes can be edited such as to overlay the pulsing by means of modulating hardware. A simple modulator is discussed wherein the luminance signal of composite video is pulsed without affecting the chroma signal. The same effect may be introduced at the consumer end, by modulating the video stream that is produced by the video source. A DVD can be edited through software, by introducing pulse-like variations in the digital RGB signals. Image intensity pulses can be overlaid onto the analog component video output of a DVD player by modulating the luminance signal component. Before entering the TV set, a television signal can be modulated such as to cause pulsing of the image intensity by means of a variable delay line that is connected to a pulse generator.

Certain monitors can emit electromagnetic field pulses that excite a sensory resonance in a nearby subject, through image pulses that are so weak as to be subliminal. This is unfortunate since it opens a way for mischievous application of the invention, whereby people are exposed unknowingly to manipulation of their nervous systems for someone else's purposes. Such application would be unethical and is of course not advocated. It is mentioned here in order to alert the public to the possibility of covert abuse that may occur while being online, or while watching TV, a video, or a DVD.

DESCRIPTION OF THE DRAWINGS

FIG. **1** illustrates the electromagnetic field that emanates from a monitor when the video signal is modulated such as to cause pulses in image intensity, and a nearby subject who is exposed to the field.

FIG. **2** shows a circuit for modulation of a composite video signal for the purpose of pulsing the image intensity.

FIG. **3** shows the circuit for a simple pulse generator.

FIG. **4** illustrates how a pulsed electromagnetic field can be generated with a computer monitor.

FIG. **5** shows a pulsed electromagnetic field that is generated by a television set through modulation of the RF signal input to the TV.

FIG. **6** outlines the structure of a computer program for producing a pulsed image.

FIG. **7** shows an extrapolation procedure introduced for improving timing accuracy of the program of FIG. **6**.

FIG. **8** illustrates the action of the extrapolation procedure of FIG. **7**.

FIG. **9** shows a subject exposed to a pulsed electromagnetic field emanating from a monitor which is responsive to a program running on a remote computer via a link that involves the Internet.

FIG. **10** shows the block diagram of a circuit for frequency wobbling of a TV signal for the purpose of pulsing the intensity of the image displayed on a TV monitor.

FIG. **11** depicts schematically a recording medium in the form of a video tape with recorded data, and the attribute of the signal that causes the intensity of the displayed image to be pulsed.

FIG. **12** illustrates how image pulsing can be embedded in a video signal by pulsing the illumination of the scene that is being recorded.

FIG. **13** shows a routine that introduces pulse variability into the computer program of FIG. **6**.

FIG. **14** shows schematically how a CRT emits an electromagnetic field when the displayed image is pulsed.

FIG. **15** shows how the intensity of the image displayed on a monitor can be pulsed through the brightness control terminal of the monitor.

FIG. **16** illustrates the action of the polarization disc that serves as a model for grounded conductors in the back of a CRT screen.

FIG. **17** shows the circuit for overlaying image intensity pulses on a DVD output.

FIG. **18** shows measured data for pulsed electric fields emitted by two different CRT type monitors, and a comparison with theory.

DETAILED DESCRIPTION

Computer monitors and TV monitors emit electromagnetic fields. Part of the emission occurs at the low frequencies at which displayed images are changing. For instance, a rythmic pulsing of the intensity of an image causes electromagnetic field emission at the pulse frequency, with a strength proportional to the pulse amplitude. The field is briefly referred to as "screen emission". In discussing this effect, any part or all what is displayed on the monitor screen is called an image. A monitor of the cathode ray tube (CRT) type has three electron beams, one for each of the basic colors red, green, and blue. The intensity of an image is here defined as

$$I=\int j\, dA, \qquad (1)$$

where the integral extends over the image, and

$$j=jr+jg+jb, \qquad (2)$$

jr, jg, and jb being the electric current densities in the red, green, and blue electron beams at the surface area dA of the image on the screen. The current densities are to be taken in the distributed electron beam model, where the discreteness of pixels and the raster motion of the beams are ignored, and the back of the monitor screen is thought to be irradiated by diffuse electron beams. The beam current densities are then functions of the coordinates x and y over the screen. The model is appropriate since we are interested in the electromagnetic field emision caused by image pulsing with the very low frequencies of sensory resonances, whereas the emissions with the much higher horizontal and vertical sweep frequencies are of no concern. For a CRT the intensity of an image is expressed in millamperes.

For a liquid crystal display (LCD), the current densities in the definition of image intensity are to be replaced by driving voltages, multiplied by the aperture ratio of the device. For an LCD, image intensities are thus expressed in volts.

It will be shown that for a CRT or LCD screen emissions are caused by fluctuations in image intensity. In composite video however, intensity as defined above is not a primary signal feature, but luminance Y is. For any pixel one has

$$Y=0.299R+0.587G+0.114B, \qquad (3)$$

where R, G, and B are the intensities of the pixel respectively in red, green and blue, normalized such as to range from 0 to 1. The definition (3) was provided by the Commission Internationale de l'Eclairage (CIE), in order to account for brightness differences at different colors, as perceived by the human visual system. In composite video the hue of the pixel is determined by the chroma signal or chrominance, which has the components R-Y and B-Y It follows that pulsing pixel luminance while keeping the hue fixed is equivalent to pulsing the pixel intensity, up to an amplitude factor. This fact will be relied upon when modulating a video stream such as to overlay image intensity pulses.

It turns out that the screen emission has a multipole expansion wherein both monopole and dipole contributions are proportional to the rate of change of the intensity I of (1). The higher order multipole contributions are proportional to the rate of change of moments of the current density j over the image, but since these contributions fall off rapidly with distance, they are not of practical importance in the present context. Pulsing the intensity of an image may involve different pulse amplitudes, frequencies, or phases for different parts of the image. Any or all of these features may be under subject control.

The question arises whether the screen emission can be strong enough to excite sensory resonances in people located at normal viewing distances from the monitor. This turns out to be the case, as shown by sensory resonance experiments and independently by measuring the strength of the emitted electric field pulses and comparing the results with the effective intensity window as explored in earlier work.

One-half Hertz sensory resonance experiments have been conducted with the subject positioned at least at normal viewing distance from a 15" computer monitor that was driven by a computer program written in Visual Basic(R), version 6.0 (VB**6**). The program produces a pulsed image with uniform luminance and hue over the full screen, except for a few small control buttons and text boxes. In VB**6**, screen pixel colors are determined by integers R, G, and B, that range from 0 to 255, and set the contributions to the pixel color made by the basic colors red, green, and blue. For a CRT-type monitor, the pixel intensities for the primary colors may depend on the RGB values in a nonlinear manner that will be discussed. In the VB**6** program the RGB values are modulated by small pulses ΔR, ΔG, ΔB, with a frequency that can be chosen by the subject or is swept in a predetermined manner. In the sensory resonance experiments mentioned above, the ratios ΔR/R, ΔG/G, and ΔB/B were always smaller than 0.02, so that the image pulses are quite weak. For certain frequencies near ½ Hz, the subject experienced physiological effects that are known to accompany the excitation of the ½ Hz sensory resonance as mentioned in the Background Section. Moreover, the measured field pulse amplitudes fall within the effective intensity window for the ½ Hz resonance, as explored in earlier experiments and discussed in the '874, '744, '922, and '304 patents. Other experiments have shown that the 2.4 Hz sensory resonance can be exited as well by screen emissions from monitors that display pulsed images.

These results confirm that, indeed, the nervous system of a subject can be manipulated through electromagnetic field pulses emitted by a nearby CRT or LCD monitor which displays images with pulsed intensity.

The various implementations of the invention are adapted to the different sources of video stream, such as video tape, DVD, a computer program, or a TV broadcast through free space or cable. In all of these implementations, the subject is exposed to the pulsed electromagnetic field that is generated by the monitor as the result of image intensity pulsing. Certain cutaneous nerves of the subject exhibit spontaneous spiking in patterns which, although rather random, contain sensory information at least in the form of average frequency. Some of these nerves have receptors that respond to the field stimulation by changing their average spiking frequency, so that the spiking patterns of these nerves acquire a frequency modulation, which is conveyed to the brain. The modulation can be particularly effective if it has a frequency at or near a sensory resonance frequency. Such frequencies are expected to lie in the range from 0.1 to 15 Hz.

An embodiment of the invention adapted to a VCR is shown in FIG. **1**, where a subject **4** is exposed to a pulsed electric field **3** and a pulsed magnetic field **39** that are emitted by a monitor **2**, labeled “MON”, as the result of pulsing the intensity of the displayed image. The image is here generated by a video casette recorder **1**, labeled “VCR”, and the pulsing of the image intensity is obtained by modulating the composite video signal from the VCR output. This is done by a video modulator **5**, labeled “VM”, which responds to the signal from the pulse generator **6**, labeled “GEN”. The frequency and amplitude of the image pulses can be adjusted with the frequency control **7** and amplitude control **8**. Frequency and amplitude adjustments can be made by the subject.

The circuit of the video modulator **5** of FIG. **1** is shown in FIG. **2**, where the video amplifiers **11** and **12** process the composite video signal that enters at the input terminal **13**. The level of the video signal is modulated slowly by injecting a small bias current at the inverting input **17** of the first amplifier **11**. This current is caused by voltage pulses supplied at the modulation input **16**, and can be adjusted through the potentiometer **15**. Since the noninverting input of the amplifier is grounded, the inverting input **17** is kept essentially at ground potential, so that the bias current is is not influenced by the video signal. The inversion of the signal by the first amplifier **11** is undone by the second amplifier **12**. The gains of the amplifiers are chosen such as to give a unity overall gain. A slowly varying current injected at the inverting input **17** causes a slow shift in the “pseudo-dc” level of the composite video signal, here defined as the short-term average of the signal. Since the pseudo-dc level of the chroma signal section determines the luminance, the latter is modulated by the injected current pulses. The chroma signal is not affected by the slow modulation of the pseudodc level, since that signal is determined by the amplitude and phase with respect to the color carrier which is locked to the color burst. The effect on the sync pulses and color bursts is of no consequence either if the injected current pulses are very small, as they are in practice. The modulated composite video signal, available at the output **14** in FIG. **2**, will thus exhibit a modulated luminance, whereas the chroma signal is unchanged. In the light of the foregoing discussion about luminance and intensity, it follows that the modulator of FIG. **2** causes a pulsing of the image intensity I. It remains to give an example how the pulse signal at the modulation input **16** may be obtained. FIG. **3** shows a pulse generator that is suitable for this purpose, wherein the RC timer **21** (Intersil ICM7555) is hooked up for astable operation and produces a square wave voltage with a frequency that is determined by capacitor **22** and potentiometer **23**. The timer **21** is powered by a battery **26**, controlled by the switch **27**. The square wave voltage at output **25** drives the LED **24**, which may be used for monitoring of the pulse frequency, and also serves as power indicator. The pulse output may be rounded in ways that are well known in the art. In the setup of FIG. **1**, the output of VCR **1** is connected to the video input **13** of FIG. **2**, and the video output **14** is connected to the monitor **2** of FIG. **1**.

In the preferred embodiment of the invention, the image intensity pulsing is caused by a computer program. As shown in FIG. **4**, monitor **2**, labeled “MON”, is connected to computer **31** labeled “COMPUTER”, which runs a program that produces an image on the monitor and causes the image intensity to be pulsed. The subject **4** can provide input to the computer through the keyboard **32** that is connected to the computer by the connection **33**. This input may involve adjustments of the frequency or the amplitude or the variability of the image intensity pulses. In particular, the pulse frequency can be set to a sensory resonance frequency of the subject for the purpose of exciting the resonance.

The structure of a computer program for pulsing image intensity is shown in FIG. **6**. The program may be written in Visual Basic(R) version 6.0 (VB**6**), which involves the graphics interface familiar from the Windows(R) operating system. The images appear as forms equipped with user controls such as command buttons and scroll bars, together with data displays such as text boxes. A compiled VB**6** program is an executable file. When activated, the program declares variables and functions to be called from a dynamic link library (DLL) that is attached to the operating system; an initial form load is performed as well. The latter comprises setting the screen color as specified by integers R, G, and B in the range 0 to 255, as mentioned above. In FIG. **6**, the initial setting of the screen color is labeled as **50**. Another action of the form load routine is the computation **51** of the sine function at eight equally spaced points, I=0 to 7, around the unit circle. These values are needed when modulating the RGB numbers. Unfortunately, the sine function is distorted by the rounding to integer RGB values that occurs in the VB**6** program. The image is chosen to fill as much of the screen area as possible, and it has spatially uniform luminance and hue.

The form appearing on the monitor displays a command button for starting and stopping the image pulsing, together with scroll bars **52** and **53** respectively for adjustment of the pulse frequency F and the pulse amplitude A. These pulses could be initiated by a system timer which is activated upon the elapse of a preset time interval. However, timers in VB**6** are too inaccurate for the purpose of providing the eight RGB adjustment points in each pulse cycle. An improvement can be obtained by using the GetTickCount function that is available in the Application Program Interface (API) of Windows 95(R) and Windows 98(R). The GetTickCount function returns the system time that has elapsed since starting Windows, expressed in milliseconds. User activation of the start button **54** provides a tick count TN through request **55** and sets the timer interval to TT miliseconds, in step **56**. TT was previously calculated in the frequency routine that is activated by changing the frequency, denoted as step **52**.

Since VB**6** is an event-driven program, the flow chart for the program falls into disjoint pieces. Upon setting the timer interval to TT in step **56**, the timer runs in the background while the program may execute subroutines such as adjustment of pulse frequency or amplitude. Upon elapse of the timer interval TT, the timer subroutine **57** starts execution with request **58** for a tick count, and in **59** an upgrade is computed of the time TN for the next point at which the RGB values are to be adjusted. In step **59** the timer is turned off, to be reactivated later in step **67**. Step **59** also resets the parameter CR which plays a role in the extrapolation procedure **61** and the condition **60**. For ease of understanding at this point, it is best to pretend that the action of **61** is simply to get a tick count, and to consider the loop controled by condition **60** while keeping CR equal to zero. The loop would terminate when the tick count M reaches or exceeds the time TN for the next phase point, at which time the program should adjust the image intensity through steps **63**–**65**. For now step **62** is to be ignored also, since it has to do with the actual extrapolation procedure **61**. The increments to the screen colors R**1**, G**1**, and B**1** at the new phase point are computed according to the sine function, applied with the amplitude A that was set by the user in step **53**. The number I that labels the phase point is incremented by unity in step **65**, but if this results in I=8 the value is reset to zero in **66**. Finally, the timer is reactivated in step **67**, initiating a new ⅛-cycle step in the periodic progression of RGB adjustments.

A program written in this way would exhibit a large jitter in the times at which the RGB values are changed. This is due to the lumpiness in the tick counts returned by the GetTickCount function. The lumpiness may be studied separately by running a simple loop with C=GetTickCount, followed by writing the result C to a file. Inspection shows that C has jumped every 14 or 15 milliseconds, between long stretches of constant values. Since for a ½ Hz image intensity modulation the ⅛-cycle phase points are 250 ms apart, the lumpiness of 14 or 15 ms in the tick count would cause considerable inaccuracy. The full extrapolation procedure **61** is introduced in order to diminish the jitter to acceptable levels. The procedure works by refining the heavy-line staircase function shown in FIG. **8**, using the slope RR of a recent staircase step to accurately determine the loop count **89** at which the loop controled by **60** needs to be exited. Details of the extrapolation procedure are shown in FIG. **7** and illustrated in FIG. **8**. The procedure starts at **70** with both flags off, and CR=0, because of the assignment in **59** or **62** in FIG. **6**. A tick count M is obtained at **71**, and the remaining time MR to the next phase point is computed in **72**. Conditions **77** and **73** are not satisfied and therefore passed vertically in the flow chart, so that only the delay block **74** and the assignments **75** are executed. Condition **60** of FIG. **6** is checked and found to be satisfied, so that the extrapolation procedure is reentered. The process is repeated until the condition **73** is met when the remaining time MR jumps down through the 15 ms level, shown in FIG. **8** as the transition **83**. The condition **73** then directs the logic flow to the assignments **76**, in which the number DM labeled by **83** is computed, and FLG**1** is set. The computation of DM is required for finding the slope RR of the straight-line element **85**. One also needs the "Final LM" **86**, which is the number of loops traversed from step **83** to the next downward step **84**, here shown to cross the MR=0 axis. The final LM is determined after repeatedly incrementing LM through the side loop entered from the FLG**1**=1 condition **77**, which is now satisfied since FLG**1** was set in step **76**. At the transition **84** the condition **78** is met, so that the assignments **79** are executed. This includes computation of the slope RR of the line element **85**, setting FLG**2**, and resetting FLG**1**. From here on, the extrapolation procedure increments CR in steps of RR while skipping tick counts until condition **60** of FIG. **6** is violated, the loop is exited, and the RGB values are adjusted.

A delay block **74** is used in order to stretch the time required for traversing the extrapolation procedure. The block can be any computation intensive subroutine such as repeated calculations of tangent and arc tangent functions.

As shown in step **56** of FIG. **6**, the timer interval TT is set to 4/10 of the time TA from one RGB adjustment point to the next. Since the timer runs in the background, this arrangement provides an opportunity for execution of other processes such as user adjustment of frequency or amplitude of the pulses.

The adjustment of the frequency and other pulse parameters of the image intensity modulation can be made internally, i.e., within the running program. Such internal control is to be distinguished from the external control provided, for instance, in screen savers. In the latter, the frequency of animation can be modified by the user, but only after having exited the screen saver program. Specifically, in Windows 95(R) or Windows 98(R), to change the animation frequency requires stopping the screen saver execution by moving the mouse, whereafter the frequency may be adjusted through the control panel. The requirement that the control be internal sets the present program apart from so-called banners as well.

The program may be run on a remote computer that is linked to the user computer, as illustrated in FIG. **9**. Although the monitor **2**, labeled "MON", is connected to the computer **31'**, labeled "COMPUTER", the program that pulses the images on the monitor **2** runs on the remoter computer **90**, labeled "REMOTE COMPUTER", which is connected to computer **31'** through a link **91** which may in part belong to a network. The network may comprise the Internet **92**.

The monitor of a television set emits an electromagnetic field in much the same way as a computer monitor. Hence, a TV may be used to produce screen emissions for the purpose of nervous system manipulation. FIG. **5** shows such an arrangement, where the pulsing of the image intensity is achieved by inducing a small slowly pulsing shift in the frequency of the RF signal that enters from the antenna. This process is here called "frequency wobbling" of the RF signal. In FM TV, a slight slow frequency wobble of the RF signal produces a pseudo-dc signal level fluctuation in the composite video signal, which in turn causes a slight intensity fluctuation of the image displayed on the monitor in the same manner as discussed above for the modulator of FIG. **2**. The frequency wobbling is induced by the wobbler **44** of FIG. **5** labeled "RFM", which is placed in the antenna line **43**. The wobbler is driven by the pulse generator **6**, labeled "GEN". The subject can adjust the frequency and the amplitude of the wobble through the tuning control **7** and the amplitude control **41**. FIG. **10** shows a block diagram of the frequency wobbler circuit that employs a variable delay line **94**, labelled "VDL". The delay is determined by the signal from pulse generator **6**, labelled "GEN". The frequency of the pulses can be adjusted with the tuning control **7**. The amplitude of the pulses is determined by the unit **98**, labelled "MD", and can be adjusted with the amplitude control **41**. Optionally, the input to the delay line may be routed through a preprocessor **93**, labelled "PRP", which may comprise a selective RF amplifier and down converter; a complimentary up conversion should then be performed on the delay line output by a postprocessor **95**, labelled "POP". The output **97** is to be connected to the antenna terminal of the TV set.

The action of the variable delay line **94** may be understood as follows. Let periodic pulses with period L be presented at the input. For a fixed delay the pulses would emerge at the output with the same period L. Actually, the time delay T is varied slowly, so that it increases approximately by LdT/dt between the emergence of consecutive pulses at the device output. The pulse period is thus increased approximately by

$$\Delta L = L\,dT/dt. \tag{4}$$

In terms of the frequency f, Eq. (4) implies approximately

$$\Delta f/f = -dT/dt. \tag{5}$$

For sinusoidal delay T(t) with amplitude b and frequency g, one has

$$\Delta f/f = -2\pi gb\cos(2\pi gt), \tag{6}$$

which shows the frequency wobbling. The approximation is good for gb<<1, which is satisfied in practice. The relative frequency shift amplitude 2πgb that is required for effective image intensity pulses is very small compared to unity. For a pulse frequency g of the order of 1 Hz, the delay may have to be of the order of a millisecond. To accomodate such long delay values, the delay line may have to be implemented as a digital device. To do so is well within the present art. In that case it is natural to also choose digital implementations for the pulse generator **6** and the pulse amplitude controller **98**, either as hardware or as software.

Pulse variability may be introduced for alleviating the need for precise tuning to a resonance frequency. This may be important when sensory resonance frequencies are not precisely known, because of the variation among individuals, or in order to cope with the frequency drift that results from chemical detuning that is discussed in the '874 patent. A field with suitably chosen pulse variability can then be more effective than a fixed frequency field that is out of tune. One may also control tremors and seizures, by interfering with the pathological oscillatory activity of neural circuits that occurs in these disorders. Electromagnetic fields with a pulse variability that results in a narrow spectrum of frequencies around the frequency of the pathological oscillatory activity may then evoke nerve signals that cause phase shifts which diminish or quench the oscillatory activity.

Pulse variability can be introduced as hardware in the manner described in the '304 patent. The variability may also be introduced in the computer program of FIG. **6**, by setting FLG**3** in step **68**, and choosing the amplitude B of the frequency fluctuation. In the variability routine **46**, shown in some detail in FIG. **13**, FLG**3** is detected in step **47**, whereupon in steps **48** and **49** the pulse frequency F is modified pseudo randomly by a term proportional to B, every 4th cycle. Optionally, the amplitude of the image intensity pulsing may be modified as well, in similar fashion. Alternatively, the frequency and amplitude may be swept through an adjustable ramp, or according to any suitable schedule, in a manner known to those skilled in the art. The pulse variability may be applied to subliminal image intensity pulses.

When an image is displayed by a TV monitor in response to a TV broadcast, intensity pulses of the image may simply be imbedded in the program material. If the source of video signal is a recording medium, the means for pulsing the image intensity may comprise an attribute of recorded data. The pulsing may be subliminal. For the case of a video signal from a VCR, the pertinent data attribute is illustrated in FIG. **11**, which shows a video signal record on part of a video tape **28**. Depicted schematically are segments of the video signal in intervals belonging to lines in three image frames at different places along the tape. In each segment, the chroma signal **9** is shown, with its short-term average level **29** represented as a dashed line. The short-term average signal level, also called the pseudo-dc level, represents the luminance of the image pixels. Over each segment, the level is here constant because the image is for simplicity chosen as having a uniform luminance over the screen. However, the level is seen to vary from frame to frame, illustrating a luminance that pulses slowly over time. This is shown in the lower portion of the drawing, wherein the IRE level of the short-term chroma signal average is plotted versus time. The graph further shows a gradual decrease of pulse amplitude in time, illustrating that luminance pulse amplitude variations may also be an attribute of the recorded data on the video tape. As discussed, pulsing the luminance for fixed chrominance results in pulsing of the image intensity.

Data stream attributes that represent image intensity pulses on video tape or in TV signals may be created when producing a video rendition or making a moving picture of a scene, simply by pulsing the illumination of the scene. This is illustrated in FIG. **12**, which shows a scene **19** that is recorded with a video camera **18**, labelled "VR". The scene is illuminated with a lamp **20**, labelled "LAMP", energized by an electric current through a cable **36**. The current is modulated in pulsing fashion by a modulator **30**, labeled "MOD", which is driven by a pulse generator **6**, labelled "GENERATOR", that produces voltage pulses **35**. Again, pulsing the luminance but not the chrominance amounts to pulsing the image intensity.

The brightness of monitors can usually be adjusted by a control, which may be addressable through a brightness adjustment terminal. If the control is of the analog type, the displayed image intensity may be pulsed as shown in FIG. **15**, simply by a pulse generator **6**, labeled "GEN", that is connected to the brigthness adjustment terminal **88** of the monitor **2**, labeled "MON". Equivalent action can be provided for digital brightness controls, in ways that are well known in the art.

The analog component video signal from a DVD player may be modulated such as to overlay image intensity pulses in the manner illustrated in FIG. **17**. Shown are a DVD player **102**, labeled "DVD", with analog component video output comprised of the luminance Y and chrominance C. The overlay is accomplished simply by shifting the luminance with a voltage pulse from generator **6**, labeled "GENERATOR". The generator output is applied to modulator **106**, labeled "SHIFTER". Since the luminance Y is pulsed without changing the chrominance C, the image intensity is pulsed. The frequency and amplitude of the image intensity pulses can be adjusted respectively with the tuner **7** and amplitude control **107**. The modulator **105** has the same structure as the modulator of FIG. **2**, and the pulse amplitude control **107** operates the potentiometer **15** of FIG. **2**. The same procedure can be followed for editing a DVD such as to overlay image intensity pulses, by processing the modulated luminance signal through an analog-to-digital converter, and recording the resulting digital stream onto a DVD, after appropriate compression. Alternatively, the digital luminance data can be edited by electronic reading of the signal, decompression, altering the digital data by software, and recording the resulting digital signal after proper compression, all in a manner that is well known in the art.

The mechanism whereby a CRT-type monitor emits a pulsed electromagnetic field when pulsing the intensity of an image is illustrated in FIG. **14**. The image is produced by an electron beam **10** which impinges upon the backside **88** of the screen, where the collisions excite phosphors that subsequently emit light. In the process, the electron beam deposits electrons **18** on the screen, and these electrons contribute to an electric field **3** labelled "E". The electrons flow along the conductive backside **88** of the screen to the terminal **99** which is hooked up to the high-voltage supply **40**, labelled "HV". The circuit is completed by the ground connection of the supply, the video amplifier **87**, labeled "VA", and its connection to the cathodes of the CRT. The electron beams of the three electron guns are collectively shown as **10**, and together the beams carry a current J. The electric current J flowing through the described circuit induces a magnetic field **39**, labeled "B". Actually, there are a multitude of circuits along which the electron beam current is returned to the CRT cathodes, since on a macroscopic scale the conductive back surface **88** of the screen provides a continuum of paths from the beam impact point to the high-voltage terminal **99**. The magnetic fields induced by the currents along these paths partially cancel each other, and the resulting field depends on the location of the pixel that is addressed. Since the beams sweep over the screen through a raster of horizontal lines, the spectrum of the induced magnetic field contains strong peaks at the horizontal and vertical frequencies. However, the interest here is not in fields at those frequencies, but rather in emissions that result from an image pulsing with the very low frequencies appropriate to sensory resonances. For this purpose a diffuse electron current model suffices, in which the pixel discreteness and the raster motion of the electron beams are ignored, so that the beam current becomes diffuse and fills the cone subtended by the displayed image. The resulting low-frequency magnetic field depends on the temporal changes in the intensity distribution over the displayed image. Order-of-magnitude estimates show that the low-frequency magnetic field, although quite small, may be sufficient for the excitation of sensory resonances in subjects located at a normal viewing distance from the monitor.

The monitor also emits a low-frequency electric field at the image pulsing frequency. This field is due in part to the electrons **18** that are deposited on the screen by the electron beams **10**. In the diffuse electron beam model, screen conditions are considered functions of the time t and of the Cartesian coordinates x and y over a flat CRT screen.

The screen electrons **18** that are dumped onto the back of the screen by the sum j(x,y,t) of the diffuse current distributions in the red, green, and blue electron beams cause a potential distribution V(x,y,t) which is influenced by the surface conductivity σ on the back of the screen and by capacitances. In the simple model where the screen has a capacitance distribution c(x,y) to ground and mutual capacitances between parts of the screen at different potentials are neglected, a potential distribution V(x,y,t) over the screen implies a surface charge density distribution

$$q=Vc(x,y), \tag{7}$$

and gives rise to a current density vector along the screen,

$$j_s=-\sigma \mathrm{grad}_s\ V, \tag{8}$$

where grad_s is the gradient along the screen surface. Conservation of electric charge implies

$$j=c\dot{V}-div_s\ (\sigma grad_s\ V), \tag{9}$$

where the dot over the voltage denotes the time derivative, and div_s is the divergence in the screen surface. The partial differential equation (9) requires a boundary condition for the solution V(x,y,t) to be unique. Such a condition is provided by setting the potential at the rim of the screen equal to the fixed anode voltage. This is a good approximation, since the resistance R_r between the screen rim and the anode terminal is chosen small in CRT design, in order to keep the voltage loss JR_r to a minimum, and also to limit low-frequency emissions.

Something useful can be learned from special cases with simple solutions. As such, consider a circular CRT screen of radius R with uniform conductivity, showered in the back by a diffuse electron beam with a spatially uniform beam current density that is a constant plus a sinusoidal part with frequency f. Since the problem is linear, the voltage V due to the sinusoidal part of the beam current can be considered separately, with the boundary condition that V vanish at the rim of the circular screen. Eq. (9) then simplifies to

$$V''+V'/r-i2\pi f c\eta\ V=-J\eta/A,\ r\leqq R, \tag{10}$$

where r is a radial coordinate along the screen with its derivative denoted by a prime, η=1/σ is the screen resistivity, A the screen area, J the sinusoidal part of the total beam current, and i=√(−1), the imaginary unit. Our interest is in very low pulse frequencies f that are suitable for excitation of sensory resonances. For those frequencies and for practical ranges for c and η, the dimensionless number $2\pi f cA\eta$ is very much smaller than unity, so that it can be neglected in Eq. (10). The boundary value problem then has the simple solution

$$V(r)=\frac{J\eta}{4\pi}(1-(r/R)^2). \tag{11}$$

In deriving **(11)** we neglected the mutual capacitance between parts of the screen that are at different potentials. The resulting error in **(10)** is negligible for the same reason that the $i2\pi f cA\eta$ term in **(10)** can be neglected.

The potential distribution V(r) of **(11)** along the screen is of course accompanied by electric charges. The field lines emanating from these charges run mainly to conductors behind the screen that belong to the CRT structure and that are either grounded or connected to circuitry with a low impedance path to ground. In either case the mentioned conductors must be considered grounded in the analysis of charges and fields that result from the pulsed component J of the total electron beam current. The described electric field lines end up in electric charges that may be called polarization charges since they are the result of the polarization of the conductors and circuitry by the screen emission. To estimate the pulsed electric field, a model is chosen where the mentioned conductors are represented together as a grounded perfectly conductive disc of radius R, positioned a short distance δ behind the screen, as depicted in FIG. **16**. Since the grounded conductive disc carries polarization charges, it is called the polarization disc. FIG. **16** shows the circular CRT screen **88** and the polarization disc **101**, briefly called "plates". For small distances δ, the capacitance density between the plates of opposite polarity is nearly equal to ε/δ, where ε is the permittivity of free space. The charge distributions on the screen and polarization disc are respectively $\epsilon V(r)/\delta+q_0$ and $-\epsilon V(r)/\delta+q_0$, where the $\epsilon V(r)/\delta$ terms denote opposing charge densities at the end of the dense field lines that run between the two plates. That the part q_0 is needed as well will become clear in the sequel.

The charge distributions $\epsilon V(r)/\delta+q_0$ and $-\epsilon V(r)/\delta+q_0$ on the two plates have a dipole moment with the density

$$D(r) = \epsilon V(r) = \frac{J\eta\epsilon}{4\pi}(1-(r/R)^2), \tag{12}$$

directed perpendicular to the screen. Note that the plate separation δ has dropped out. This means that the precise location of the polarization charges is not critical in the present model, and further that δ may be taken as small as desired. Taking δ to zero, one thus arrives at the mathematical model of pulsed dipoles distributed over the circular CRT screen. The field due to the charge distribution q_0 will be calculated later.

The electric field induced by the distributed dipoles (12) can be calculated easily for points on the centerline of the screen, with the result

$$E(z) = \frac{V(0)}{R}\{2\rho/R - R/\rho - 2|z|/R\}, \tag{13}$$

where V(**0**) is the pulse voltage (11) at the screen center, ρ the distance to the rim of the screen, and z the distance to the center of the screen. Note that V(**0**) pulses harmonically with frequency ∫, because in (11) the sinusoidal part J of the beam current varies in this manner.

The electric field (13) due to the dipole distribution causes a potential distribution V(r)/2 over the screen and a potential distribution of −V(r)/2 over the polarization disc, where V(r) is nonuniform as given by (11). But since the polarization disc is a perfect conductor it cannot support voltage gradients, and therefore cannot have the potential distribution −V(r)/2. Instead, the polarization disc is at ground potential. This is where the charge distribution $q_0(r)$ comes in; it must be such as to induce a potential distribution V(r)/2 over the polarization disc. Since the distance between polarization disc and screen vanishes in the mathematical model, the potential distribution V(r)/2 is induced over the screen as well. The total potential over the monitor screen thus becomes V(r) of (**11**), while the total potential distribution over the polarization disc becomes uniformly zero. Both these potential distributions are as physically required. The electric charges q_0 are moved into position by polarization and are partly drawn from the earth through the ground connection of the CRT.

In our model the charge distribution q_0 is located at the same place as the dipole distribution, viz., on the plane z=0 within the circle with radius R. At points on the center line of the screen, the electric field due to the monopole distribution q_0 is calculated in the following manner. As discussed, the monopoles must be such that they cause a potential ϕ_0 that is equal to V(r)/2 over the disc with radius R centered in the plane z=0. Although the charge distribution $q_0(r)$ is uniquely defined by this condition, it cannot be calculated easily in a straightforward manner. The difficulty is circumvented by using an intermediate result derived from Excercise 2 on page 191 of Kellogg (1953), where the charge distribution over a thin disc with uniform potential is given. By using this result one readily finds the potential $\phi^*(z)$ on the axis of this disc as

$$\phi^*(z) = \frac{2}{\pi}V^*\beta(R_1), \tag{14}$$

where $\beta(R_1)$ is the angle subtended by the disc radius R_1, as viewed from the point z on the disc axis, and V* is the disc potential. The result is used here in an attempt to construct the potential $\phi_0(z)$ for a disc with the nonuniform potential V(r)/2, by the ansatz of writing the field as due to a linear combination of abstract discs with various radii R_1 and potentials, all centered in the plane z=0. In the ansatz the potential on the symmetry axis is written

$$\phi_0(z) = \alpha\beta(R) + b\int_0^R \beta(R_1)dW, \tag{15}$$

where W is chosen as the function $1-R_1^2/R^2$, and the constants a and b are to be determined such that the potential over the plane z=0 is V(r)/2 for radii r ranging from 0 to R, with V(r) given by (11). Carrying out the integration in (15) gives

$$\phi_0(z)=\alpha\beta(R)-b\{(1+z^2/R^2)\beta(R)-|z|/R\}. \tag{16}$$

In order to find the potential over the disc r<R in the plane z=0, the function $\phi_0(z)$ is expanded in powers of z/R for 0<z<R, whereafter the powers z^n are replaced by $r^nP_n(\cos\theta)$, where the P_n are Legendre polynomials, and (r,θ) are symmetric spherical coordinates centered at the screen center. This procedure amounts to a continuation of the potential from the z-axis into the half ball r<R, z>0, in such a manner that the Laplace equation is satisfied. The method is discussed by Morse and Feshbach (1953). The "Laplace continuation" allows calculation of the potential ϕ_0 along the surface of the disc r<R centered in the plane z=0. The requirement that this potential be V(r)/2 with the function V(r) given by (11) allows solving for the constants a and b, with the result

$$a=-V(0)/\pi,\ b=-2V(0)/\pi. \tag{17}$$

Using (17) in (16) gives

$$\phi_0(z) = \frac{V(0)}{\pi}[(1+2z^2/R^2)\beta(R) - 2|z|/R], \tag{18}$$

and by differentiation with respect to z one finally finds

$$E_0(z) = \frac{V(0)}{\pi R}(z/|z|)[4-(R/\rho)^2 - 4\beta(R)|z|/R] \tag{19}$$

for the electric field on the center line of the screen brought about by the charge distribution $q_0(z)$.

The center-line electric field is the sum of the part (13) due to distributed pulsed dipoles and part (19) due to distributed pulsed monopoles. Although derived for circular screens, the results may serve as an approximation for other shapes, such as the familiar rounded rectangle, by taking R as the radius of a circle that has the same area as the screen.

For two CRT-type monitors the pulsed electric field due to image intensity pulsing has been measured at several points on the screen center line for pulse frequencies of ½ Hz. The monitors were the 15" computer monitor used in the sensory resonance experiments mentioned above, and a 30" TV tube. The experimental results need to be compared with the theory derived above. Since R is determined by the screen area, the electric fields given by (13) and (19) have as only free parameter the pulse voltage V(**0**) at the screen center. The amplitude of this voltage can therefore be determined for the tested monitors by fitting the experimental data to the theoretical results. Prior to fitting, the data were normalized to an image that occupies the entire screen and is pulsed uniformly with a 100% intensity amplitude. The results of the one-parameter fit are displayed in FIG. **18**, which shows the theoretical graph **100**, together with the normalized experimental data points **103** for the 15– computer monitor and for the 30" TV tube. FIG. **18** shows that the developed theory agrees fairly well with the experimental results. From the best fit one can find the center-screen voltage pulse amplitudes. The results, normalized as discussed above, are $|V(0)|=266.2$ volt for the 15" computer monitor and $|V(0)|=310.1$ volt for the 30" TV tube. With these amplitudes in hand, the emitted pulsed electric field along the center line of the monitors can be calculated from the sum of the fields (13) and (19). For instance, for the 15" computer monitor with 1.8% RGB pulse modulation used in the ½ Hz sensory resonance experiments mentioned above, the pulsed electric field at the center of the subject, located at z=70 cm on the screen center line, is calculated as having an amplitude of 0.21 V/m. That such a pulsed electric field, applied to a large portion of the skin, is sufficient for exciting the ½ Hz sensory resonance is consistent with experimental results discussed in the '874 patent.

In deriving **(11)**, the dimensionless number $2\pi\int cA\eta$ was said to be much smaller than unity. Now that the values for $|V(0)|$ are known, the validity of this statement can be checked. Eq. (11) implies that $|V(0)|$ is equal to $\eta|J|/4\pi$. The sum of the beam currents in the red, green, and blue electron guns for 100% intensity modulation is estimated to have pulse amplitudes $|J|$ of 0.5 mA and 2.0 mA respectively for the 15" computer monitor and the 30" TV tube. Using the derived values for $|V(0)|$, one arrives at estimates for the screen resistivity η as 6.7 MΩ/square and 1.9 MΩ/square respectively for the 15" computer monitor and the 30" TV tube. Estimating the screen capacity cA as 7 pf and 13 pf, $2\pi\int cA\eta$ is found to be 148×10^{-6} and 78×10^{-6}, respectively for the 15" computer monitor and the 30" TV tube. These numbers are very small compared to unity, so that the step from (10) to (11) is valid.

The following procedures were followed in preparing pulsed images for the field measurements. For the 15" computer monitor the images were produced by running the VB**6** program discussed above. The pulsed image comprised the full screen with basic RGB values chosen uniformly as R=G=B=127, with the exception of an on/off button and a few data boxes which together take up 17% of the screen area. The image intensity was pulsed by modifying the R, G, and B values by integer-rounded sine functions $\Delta R(t)$, $\Delta G(t)$, and $\Delta B(t)$, uniformly over the image, except at the button and the data boxes. The measured electric field pulse amplitudes were normalized to a pulsed image that occupies all of the screen area and has 100% intensity modulation for which the image pulses between black and the maximum intensity, for the fixed RGB ratios used. The image intensity depends on the RGB values in a nonlinear manner that will be be discussed. For the measurements of the pulsed electric field emitted by 30" TV tube, a similar image was used as for the 15" computer monitor. This was done by playing back a camcorder recording of the computer monitor display when running the VB**6** program, with 40% pulse modulation of R, G, and B.

In front of the monitor, i.e., for z>0, the parts **(13)** and **(19)** contribute about equally to the electric field over a practical range of distances z. When going behind the monitor where z is negative the monopole field flips sign so that the two parts nearly cancel each other, and the resulting field is very small. Therefore, in the back of the CRT, errors due to imperfections in the theory are relatively large. Moreover our model, which pretends that the polarization charges are all located on the polarization disc, fails to account for the electric field flux that escapes from the outer regions of the back of the screen to the earth or whatever conductors happen to be present in the vincinity of the CRT. This flaw has relatively more serious consequences in the back than in front of the monitor.

Screen emissions in front of a CRT can be cut dramatically by using a grounded conductive transparent shield that is placed over the screen or applied as a coating. Along the lines of our model, the shield amounts to a polarization disc in front of the screen, so that the latter is now sandwiched between to grounded discs. The screen has the pulsed potential distribution V(r) of (11), but no electric flux can escape. The model may be modified by choosing the polarization disc in the back somewhat smaller than the screen disc, by a fraction that serves as a free parameter. The fraction may then be determined from a fit to measured fields, by minimizing the relative standard deviation between experiment and theory.

In each of the electron beams of a CRT, the beam current is a nonlinear function of the driving voltage, i.e., the voltage between cathode and control grid. Since this function is needed in the normalization procedure, it was measured for the 15" computer monitor that has been used in the ½ Hz sensory resonance experiments and the electric field measurements. Although the beam current density j can be determined, it is easier to measure the luminance, by reading a light meter that is brought right up to the monitor screen. With the RGB values in the VB**6** program taken as the same integer K, the luminance of a uniform image is proportional to the image intensity I. The luminance of a uniform image was measured for various values of K. The results were fitted with

$$I=c_1K^{\gamma}, \tag{20}$$

where c_1 is a constant. The best fit, with 6.18% relative standard deviation, was obtained for $\gamma=2.32$.

Screen emissions also occur for liquid crystal displays (LCD). The pulsed electric fields may have considerable amplitude for LCDs that have their driving electrodes on opposite sides of the liquid crystal cell, for passive matrix as well as for active matrix design, such as thin film technology (TFT). For arrangements with in-plane switching (IPS) however, the driving electrodes are positioned in a single plane, so that the screen emission is very small. For arrangements other than IPS, the electric field is closely approximated by the fringe field of a two-plate condenser, for the simple case that the image is uniform and extends over the full screen. For a circular LCD screen with radius R, the field on the center line can be readily calculated as due to pulsed dipoles that are uniformly distributed over the screen, with the result

$$E_d(z)=(\tfrac{1}{2})VR^2/(z^2+R^2)^{3/2}, \tag{21}$$

where $E_d(z)$ is the amplitude of the pulsed electric field at a distance z from the screen and V is a voltage pulse amplitude, in which the aperture ratio of the LCD has been taken into account. Eq. (21) can be used as an approximation for screens of any shape, by taking R as the radius of a circle with the same area as the screen. The result applies to the case that the LCD does not have a ground connection, so that the top and bottom electrodes are at opposite potential, i.e., V/2 and −V/2.

If one set of LCD electrodes is grounded, monopoles are needed to keep these electrodes at zero potential, much as in the case of a CRT discussed above. The LCD situation is simpler however, as there is no charge injection by electron beams, so that the potentials on the top and bottom plates of the condenser in the model are spatially uniform. From **(14)** it is seen that monopoles, distributed over the disc of radius R in the plane z=0 such as to provide on the disc a potential V/2, induce on the symmetry axis a potential

$$\phi(z)=\frac{1}{\pi}V\beta(R). \tag{22}$$

Differentiating with respect to z gives the electric field on the symmetry axis

$$E_m(z)=\frac{zVR}{|z|\pi(z^2+R^2)}, \tag{23}$$

induced by the pulsed monopoles. For an LCD with one set of electrodes grounded, the pulsed electric field for screen voltage pulse amplitude V at a distance z from the screen on the center line has an amplitude that is the sum of the parts (21) and (23). The resultant electric field in the back is relatively small, due to the change in sign in the monopole field that is caused by the factor z/|z|. Therefore, screen emissions in front of an LCD can be kept small simply by having the grounded electrodes in front.

As a check on the theory, the pulsed electric field emitted by the 3" LCD-TFT color screen of the camcorder mentioned above has been measured at eleven points on the center line of the screen, ranging from 4.0 cm to 7.5 cm. The pulsed image was produced by playing back the video recording of the 15" computer monitor that was made while running the VB**6** program discussed above, for a image intensity pulse frequency of ½ Hz, R=G=B=K, modulated around K=127 with an amplitude ΔK=51. After normalization to a uniform full screen image with 100% intensity modulation by using the nonlinear relation (20), the experimental data were fitted to the theoretical curve that expresses the sum of the fields (21) and (23). The effective screen pulse voltage amplitude V was found to be 2.1 volt. The relative standard deviation in V for the fit is 5.1%, which shows that theory and experiment are in fairly good agreement.

Certain monitors can cause excitation of sensory resonances even when the pulsing of displayed images is subliminal, i.e., unnoticed by the average person. When checking this condition on a computer monitor, a problem arises because of the rounding of RGB values to integers, as occurs in the VB**6** program. For small pulse amplitude the sine wave is thereby distorted into a square wave, which is easier to spot. This problem is alleviated somewhat by choosing ΔR=0, ΔG=0, and ΔB=2, since then the 8 rounded sine functions around the unit circle, multiplied with the pulse amplitude ΔB=2 become the sequence 1, 2 11 2, 1, −1 −2, −2, −1, etc, which is smoother to the eye than a square wave. Using the VB**6** program and the 15" computer monitor mentioned above with R=71, G=71, and B=233, a ½ Hz pulse modulation with amplitudes ΔR=ΔG=0 and ΔB=2 could not be noticed by the subject, and is therefore considered subliminal. It is of interest to calculate the screen emission for this case, and conduct a sensory resonance experiment as well. A distance z=60 cm was chosen for the calculation and the experiment. Using Eq. (20), the image intensity pulse modulation for the case is found to be 1.0% of the maximum intensity modulation. Using R=13.83 cm together with |V(0)|=266.2 V for the 15" computer monitor, and the theoretical graph **100** of FIG. **18**, the pulsed electric field at z=60 cm was found to have an amplitude of 138 mV/m. In view of the experimental results discussed in the '874 and '922 patents, such a field, used at a pulse frequency chosen appropriately for the ½ Hz sensory resonance and applied predominantly to the face, is expected to be sufficient for exciting the ½ Hz sensory resonance. A confirmation experiment was done by running the VB**6** program with the discussed settings and the 15" monitor. The center of the subject's face was positioned on the screen center line, at a distance of 60 cm from the screen. A frequency sweep of −0.1% per ten cycles was chosen, with an initial pulse frequency of 34 ppm. Full ptosis was experienced by the subject at 20 minutes into the run, when the pulse frequency was f=31.76 ppm. At 27 minutes into the run, the frequency sweep was reversed to +0.1% per ten cycles. Full ptosis was experienced at f=31.66 ppm. At 40 minutes into the run, the frequency sweep was set to −0.1% per ten cycles. Full ptosis occurred at f=31.44 ppm. The small differences in ptosis frequency are attributed to chemical detuning, discussed in the Background Section. It is concluded that the ½ Hz sensory resonance was excited in this experiment by screen emissions from subliminal image pulsing on the 15" computer monitor at a distance of 60 cm. For each implementation and embodiment discussed, the image pulsing may be subliminal.

The human eye is less sensitive to changes in hue than to changes in brightness. In composite video this fact allows using a chrominance bandwidth that is smaller than the luminance bandwidth. But it also has the consequence that pulsing of the chrominance for fixed luminance allows larger pulse amplitudes while staying within the subliminal pulse regime. Eq. (3) shows how to pulse the chrominance components R-Y and B-Y while keeping Y fixed; for the change in pixel intensity one then has

$$\Delta I_h=0.491\Delta(R\text{-}Y)+0.806\Delta(B\text{-}Y). \tag{24}$$

Luminance pulses with fixed chrominance give a change in pixel intensity

$$\Delta I_1=3\Delta Y. \tag{25}$$

Of course, pure chrominance pulses may be combined with pure luminance pulses; an instance of such combination has been mentioned above.

The subliminal region in color space needs to be explored to determine how marginally subliminal pulses ΔR, ΔG, and ΔB depend on RGB values. Prior to this, the condition for image pulses to be subliminal should not be phrased solely in terms of the percentage of intensity pulse amplitude. The subliminal image pulsing case considered above, where the monitor is driven by a VB**6** computer program with R=G=71, B=233, and ΔR=ΔG=0, ΔB=2 for full-screen images will be referred to as "the standard subliminal image pulsing".

In the interest of the public we need to know the viewing distances at which a TV with subliminally pulsed images can cause excitation of sensory resonances. A rough exploration is reported here which may serve as starting point for further work. The exploration is limited to estimating the largest distance $z=z_{max}$ along the center line of the 30" TV at which screen emissions can excite the ½ Hz resonance, as determined by the ptosis test. The TV is to display an image which undergoes the standard subliminal pulsing as defined above. It would be best to perform this test with the 30" TV on which the subliminally pulsed images are produced by means of a video. Since such a video was not available, the ptosis test was conducted instead with a pulsed electric field source consisting of a small grounded doublet electrode of the type discussed in the '874 patent. The doublet was driven with a sinusoidal voltage of 10 V amplitude, and the center of mass of the subject was located on the center line of the doublet at a distance $z=z_d=323$ cm. The doublet electrodes are rectangles of 4.4 cm by 4.7 cm. At the large distance z_d there is whole-body exposure to the field, so that the bulk effect discussed in the '874 patent comes into play, as is expected to happen also at the distance z_{max} from the 30" TV monitor. The subject was facing the "hot" electrode of the doublet, so that at the subject center the electric field was the sum of the parts (21) and (23), for positive values of z. It was thought important to use a sine wave, since that would be the "commercially" preferred pulse shape which allows larger pulse amplitudes without being noticed. The only readily available sine wave generator with the required voltage was an oscillator with a rather coarse frequency control that cannot be set accurately, although the frequency is quite stable and can be measured accurately. For the experiment a pulse frequency of 0.506 Hz was accepted, although it differs considerably from the steady ptosis frequency for this case. The subject experienced several ptosis cycles of moderate intensity, starting 8 minutes into the experiment run. It is concluded that the ½ Hz sensory resonance was excited, and that the stimulating field was close to the weakest field capable of excitation. From Eqs. (21) and (23), the electric field pulse amplitude at the center of mass of the subject was found to be 7.9 mV/m. That an electric field with such a small pulse amplitude, applied to the whole body, is capable of exciting the ½ Hz sensory resonance is consistent with experimental results reported in the '874 patent, although these were obtained for the 2.4 Hz resonance. Next, the distance z_{max} was determined at which the 30" TV tube with 1% image intensity pulse amplitude produces an electric field with a pulse amplitude of 7.9 mV/m, along the center line of the screen. From Eqs. (13) and (19) one finds z_{max}=362.9 cm. At more than 11 feet, this is a rather large distance for viewing a 30" TV. Yet, the experiment and theory discussed show that the ½ Hz sensory resonance can be excited at this large distance, by pulsing the image intensity subliminally. Of course, the excitation occurs as well for a range of smaller viewing distances. It is thus apparent that the human nervous system can be manipulated by screen emissions from subliminal TV image pulses.

Windows 95, Windows 98, and Visual Basic are registered trademarks of Microsoft Corporation.

The invention is not limited by the embodiments shown in the drawings and described in the specification, which are given by way of example and not of limitation, but only in accordance with the scope of the appended claims.

I claim:

1. A method for manipulating the nervous system of a subject located near a monitor, the monitor emitting an electromagnetic field when displaying an image by virtue of the physical display process, the subject having a sensory resonance frequency, the method comprising:

creating a video signal for displaying an image on the monitor, the image having an intensity;

modulating the video signal for pulsing the image intensity with a frequency in the range 0.1 Hz to 15 Hz; and

setting the pulse frequency to the resonance frequency.

2. A computer program for manipulating the nervous system of a subject located near a monitor, the monitor emitting an electromagnetic field when displaying an image by virtue of the physical display process, the subject having cutaneous nerves that fire spontaneously and have spiking patterns, the computer program comprising:

a display routine for displaying an image on the monitor, the image having an intensity;

a pulse routine for pulsing the image intensity with a frequency in the range 0.1 Hz to 15 Hz; and

a frequency routine that can be internally controlled by the subject, for setting the frequency;

whereby the emitted electromagnetic field is pulsed, the cutaneous nerves are exposed to the pulsed electromagnetic field, and the spiking patterns of the nerves acquire a frequency modulation.

3. The computer program of claim **2**, wherein the pulsing has an amplitude and the program further comprises an amplitude routine for control of the amplitude by the subject.

4. The computer program of claim **2**, wherein the pulse routine comprises:

a timing procedure for timing the pulsing; and

an extrapolation procedure for improving the accuracy of the timing procedure.

5. The computer program of claim **2**, further comprising a variability routine for introducing variability in the pulsing.

6. Hardware means for manipulating the nervous system of a subject located near a monitor, the monitor being responsive to a video stream and emitting an electromagnetic field when displaying an image by virtue of the physical display process, the image having an intensity, the subject having cutaneous nerves that fire spontaneously and have spiking patterns, the hardware means comprising:

pulse generator for generating voltage pulses;

means, responsive to the voltage pulses, for modulating the video stream to pulse the image intensity;

whereby the emitted electromagnetic field is pulsed, the cutaneous nerves are exposed to the pulsed electromagnetic field, and the spiking patterns of the nerves acquire a frequency modulation.

7. The hardware means of claim **6**, wherein the video stream is a composite video signal that has a pseudo-dc level, and the means for modulating the video stream comprise means for pulsing the pseudo-dc level.

8. The hardware means of claim **6**, wherein the video stream is a television broadcast signal, and the means for modulating the video stream comprise means for frequency wobbling of the television broadcast signal.

9. The hardware means of claim **6**, wherein the monitor has a brightness adjustment terminal, and the means for modulating the video stream comprise a connection from the pulse generator to the brightness adjustment terminal.

10. A source of video stream for manipulating the nervous system of a subject located near a monitor, the monitor emitting an electromagnetic field when displaying an image by virtue of the physical display process, the subject having cutaneous nerves that fire spontaneously and have spiking patterns, the source of video stream comprising:

means for defining an image on the monitor, the image having an intensity; and

means for subliminally pulsing the image intensity with a frequency in the range 0.1 Hz to 15 Hz;

whereby the emitted electromagnetic field is pulsed, the cutaneous nerves are exposed to the pulsed electromagnetic field, and the spiking patterns of the nerves acquire a frequency modulation.

11. The source of video stream of claim **10** wherein the source is a recording medium that has recorded data, and the means for subliminally pulsing the image intensity comprise an attribute of the recorded data.

12. The source of video stream of claim **10** wherein the source is a computer program, and the means for subliminally pulsing the image intensity comprise a pulse routine.

13. The source of video stream of claim **10** wherein the source is a recording of a physical scene, and the means for subliminally pulsing the image intensity comprise:

pulse generator for generating voltage pulses;

light source for illuminating the scene, the light source having a power level; and

modulation means, responsive to the voltage pulses, for pulsing the power level.

14. The source of video stream of claim **10**, wherein the source is a DVD, the video stream comprises a luminance signal and a chrominance signal, and the means for subliminal pulsing of the image intensity comprise means for pulsing the luminance signal.

* * * * *

United States Patent [19]
Rabichev et al.

[11] **3,773,049**
[45] **Nov. 20, 1973**

[54] **APPARATUS FOR THE TREATMENT OF NEUROPSYCHIC AND SOMATIC DISEASES WITH HEAT, LIGHT, SOUND AND VHF ELECTROMAGNETIC RADIATION**

[76] Inventors: **Lev Yakovlevich Rabichev,** ulitsa 25 Oktyabrya, 80, kv. 7; **Vladislav Fedorovich Vasiliev,** ulitsa 25 Oktyabrya, 46, kv. 3; **Alexandr Sergeevich Putilin,** 1 pereulok Panfilova, 9-B, kv. 25; **Tatyana Grigorievna Ilina,** ulitsa Iskry, 25, kv. 1; **Petr Vasilievich Raku,** ulitsa Roz, 13, korpusl, kv. 73; **Leopold Pavlovich Kernitsky,** ulitsa Zelinskogo, 12, korpus4, kv. 55, all of Kishinev, U.S.S.R.

[22] Filed: **Nov. 13, 1970**

[21] Appl. No.: **89,374**

[52] **U.S. Cl.** .... **128/362,** 128/1 C
[51] **Int. Cl.** **A61n 1/40**
[58] **Field of Search** .... 128/1 C, 1 R, 1.3, 128/24.1, 362, 380, 399, 404, 410, 419 P, 422

[56] **References Cited**

UNITED STATES PATENTS

3,103,219	9/1963	Chadner	128/1 C
3,576,185	4/1971	Schulz et al.	128/1 C
3,213,851	10/1965	Currea	128/1 R
2,859,731	11/1958	Sutton	128/1 C
3,470,870	10/1969	Schoffer	128/1 C

FOREIGN PATENTS OR APPLICATIONS

619,129	3/1961	Italy	128/1 C
1,183,607	12/1964	Germany	128/1 C
826,766	1/1960	Great Britain	128/419 P

Primary Examiner—William E. Kamm
Attorney—Holman & Stern

[57] **ABSTRACT**

An apparatus for the treatment of neuropsychic and somatic disorders wherein light-, sound-, VHF electromagnetic field-pulses and radiation from light-, sound-, VHF electromcagnetic field- and heat-sources, respectively, are simultaneously applied by means of a control unit to the patient's central nervous system with a predermined repetition rate. The light radiation and sound radiation sources are made so as to exert an adequate and monotonous influence of the light-and sound-radiation on the patient's visual analyzers and auditory analyzers, respectively.

8 Claims, 2 Drawing Figures

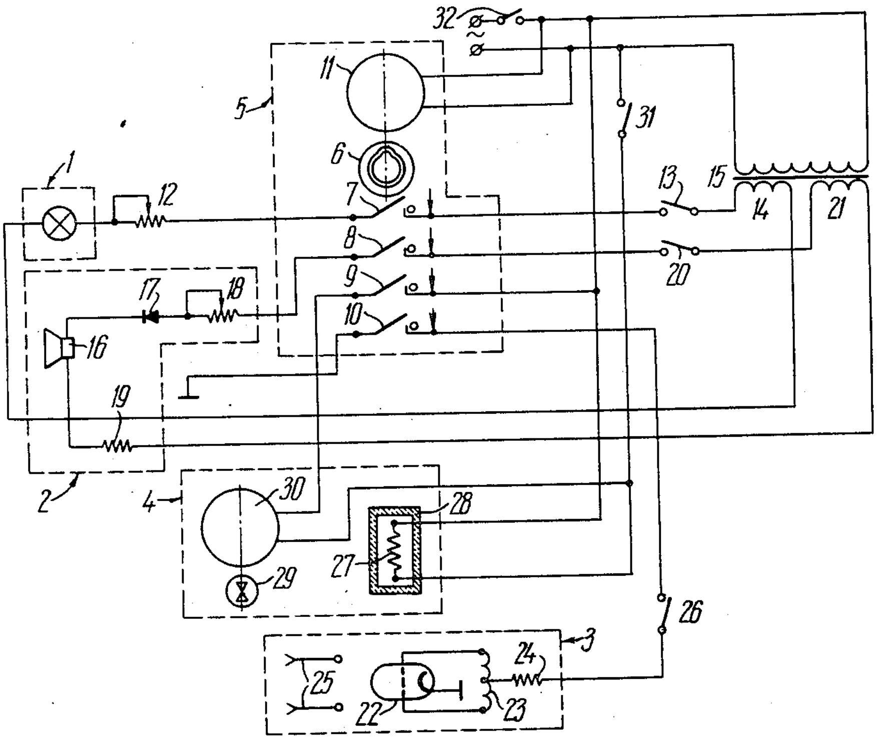

Patented Nov. 20, 1973 3,773,049

2 Sheets-Sheet 1

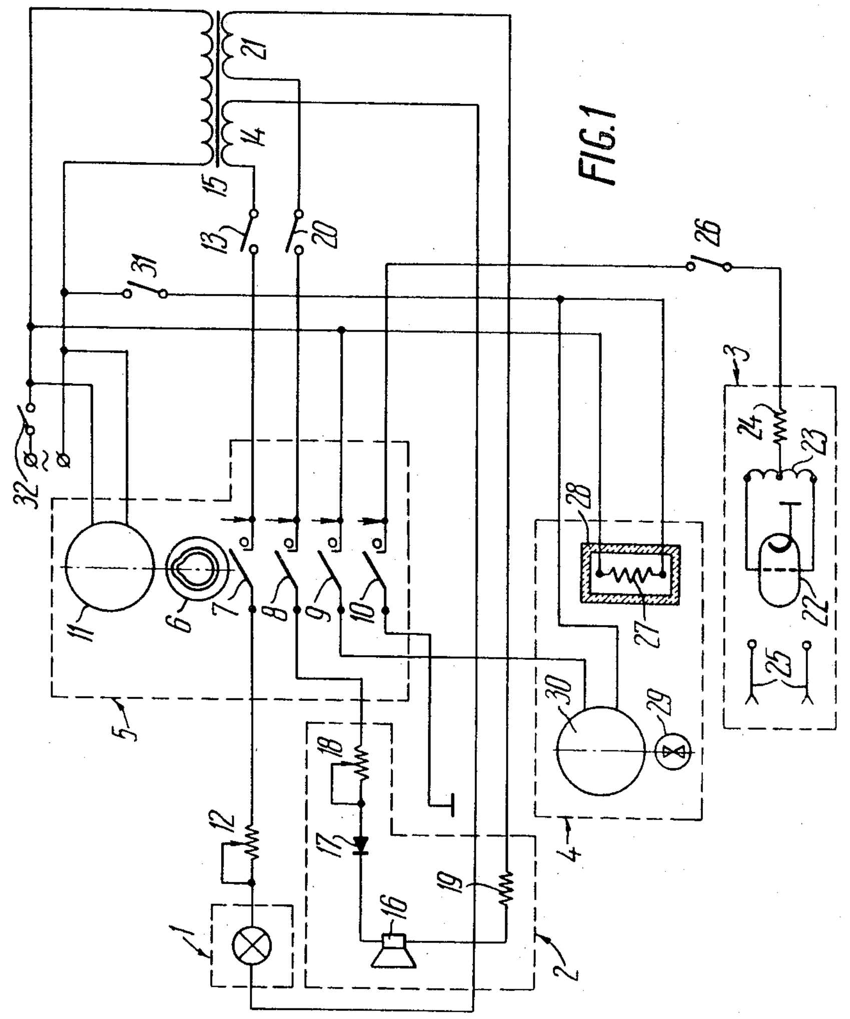
21
14
15
13
20
31
26
32
~
11
6
7
8
9
10
5
28
27
30
29
4
24
23
22
25
3
12
1
18
17
16
19
2

FIG.1

Patented Nov. 20, 1973 3,773,049

2 Sheets-Sheet 2

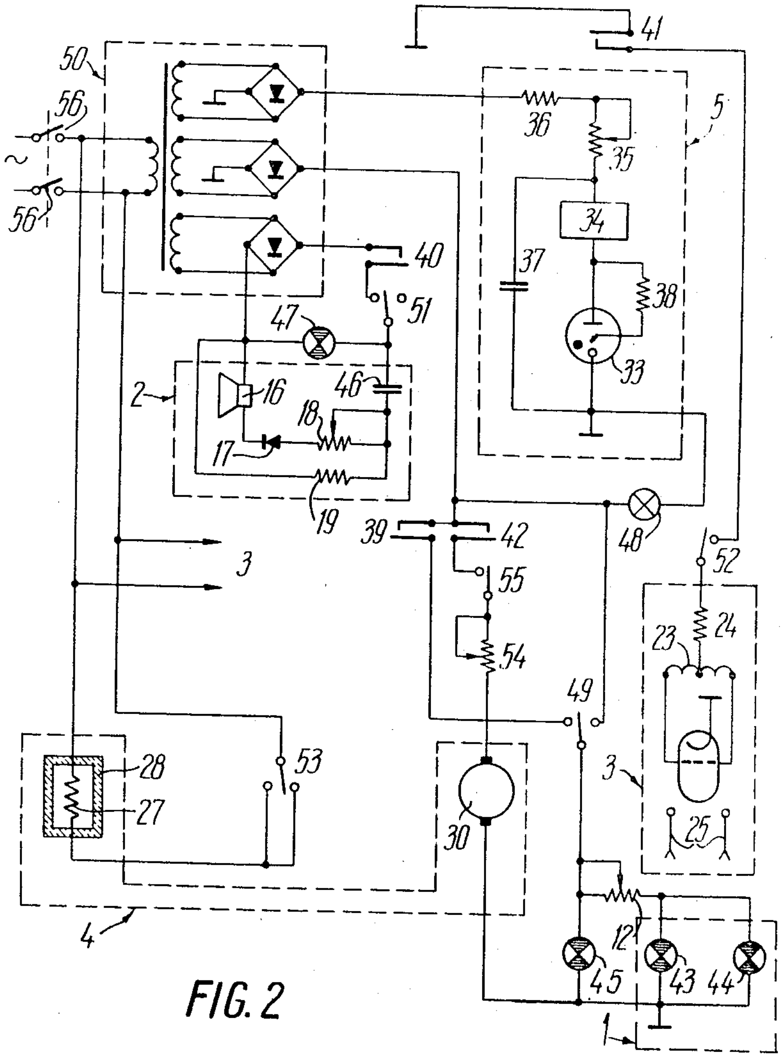
41
50
56
~
56
36
35
5
34
37
38
33
40
51
47
46
2
16
18
17
19
48
39
42
55
3
52
24
23
54
49
3
25
28
27
53
30
4
12
45
43
44
1

FIG. 2

APPARATUS FOR THE TREATMENT OF NEUROPSYCHIC AND SOMATIC DISEASES WITH HEAT, LIGHT, SOUND AND VHF ELECTROMAGNETIC RADIATION

The present invention relates to physiotherapeutic apparatus, and more particularly to apparatus for the treatment of inorganically caused neuropsychic and somatic disorders, such as neuroses, psychoses, insomnia, hypertension, tonic disease, ulcer, stammering, bronchial asthma, intestinal dyskinesia, cardioneurosis, vegeto-vascular neurosis, eczema, asthenic and reactive disturbances. The apparatus of the present invention may find successful application in sports- and military medicine, i.e., in such situations as:

the start of a crucial sports competition or a similar sports demonstration;

cases of extremely high neurotension, as observed in combatant unit soldiers and in patients of field military hospitals.

Known are apparatus for the treatment of neuropsychic and somatic disorders in which light and sound pulses are emitted by light radiation sources and sound radiation sources to act upon the patient's central nervous system.

In said apparatus use is made of light- and sound radiation sources. The sound emission includes speech sounds recorded on a magnetic tape, as well as ultrasound. Under the treatment conditions the patient's head and the upper portion of his body are enclosed in a chamber which is made in the form of a kind of sarcophagus.

Thus in the above-said apparatus, only one curative factor is employed, namely that of speech formulae of a hypnotic suggestion as recorded on a magnetic tape and designed to produce a suggestive effect via channels of the second signal system (intellect, mind, psyche). Whenever the patient is psychologically reluctant to admit suggestive formulae, such an apparatus is of no avail as far as that particular case is concerned. Light- and ultrasound emission, used in the apparatus of the prior art, do not, in fact, play an independent curative role, but are rather of an auxiliary nature, for their action on the patient's central nervous system is continuous and intended to inhibit the same so that an appropriate background could be prepared to facilitate the suggestive treatment itself. The background factors themselves are physiologically inadequate being beyond the ranges of aural and ocular perception and sometimes involving an adverse effect upon the patient's condition. It should be stressed that the chamber used in accomodating the patient's head and the upper portion of his body has an unfavourable effect upon the patient, because it is made, as said above in the form of a sarcophagus and may give rise to negative phychic responses.

The present invention is aimed at providing an apparatus for the treatment of neuropsychic and somatic diseases, which ensures a rhythmic and remotely controlled action of at least three physical factors upon the patient's nervous system.

With this and other objects in view, the invention resides in the apparatus for the treatment of neuropsychic and somatic diseases in which light pulses and sound pulses from light sources and sound sources, respectively, act upon the patient's central nervous system, wherein there is provided an VHF electromagnetic field source installed in the immediate proximity of the patient's ganglia, and a control unit which is electrically connected with the sources of light radiation, sound radiation and VHF electromagnetic field radiation and which effects control over these sources so as to enable said sources under control to send to the central nervous system simultaneous light pulses, sound pulses and VHF electromagnetic pulses at an assigned pulse repetition rate thereby causing a state of inhibition in the patient under treatment, the light radiation sources and sound radiation sources being embodied in such a way that an adequate and monotonous influence of the sound and light radiations is exerted upon the patient's visual analyser and auditory analyser.

In order to intensify the curative effect and to make for an accelerated transfer from vigilance to what is nearly normal physiological sleep it is reasonable that a source of heat radiation be added and installed in the immediate proximity to the patient's skin surface, said source being electrically connected with a control unit and controlled by the latter so that heat radiation is sent to the patient's skin surface simultaneously with the pulses from the VHF electromagnetic field source, light source and sound source, said pulses following at an assigned repetition rate causing the inhibition to be irradiated.

The light radiation source may be made in the form of at least one lamp, blue or green in colour, which is installed at the patient's eye level.

When using an even number of lamps, said lamps are installed symmetrically in respect of the patient's saggital sinus plane.

The control unit may be embodied as a chopper in conjunction with a cam mechanism providing for generation of pulses at a certain repetition rate.

In order to extend the pulse repetition range it is reasonable that an electronic timer be provided including at least one thyratron.

With due regard to the patient's individual features said electronic timer should be equipped with a potentiometer to control the pulse repetition rate of VHF electromagnetic field radiation, light radiation and sound radiation, as well as heat radiation, said electronic timer being electrically connected with a potentiometer.

Due to such construction the present apparatus for the treatment of neuropsychic and somatic disorders allows improvement of the mobility and intensity of nervous processes, normalization of the affected functional derangements as apparent from the removal of pathological symptoms, decrease in manifestations of autonomic angioneurosis and neurotic symptoms, a longer and sounder sleep.

Other objects and advantages of the present invention will be made apparent upon a consideration of the specific illustrative embodiments as described herein with reference to the drawings, wherein:

FIG.I shows a schematic electrical circuit of one embodiment of an apparatus for the treatment of neuropsychic and somatic disorders;

FIG.2 is another embodiment of same.

The apparatus for the treatment of neuropsychic disorders comprises, in accordance with the present invention, source I through 4 (FIGS I and 2) of light radiation, sound radiation and VHF electromagnetic field radiation, heat radiation respectively, wherein light pulses, sound pulses, VHF electromagnetic pulses and heat radiation act upon the central nervous system of a patient (not shown in the drawing). The apparatus also includes control unit 5 which is electrically connected with said sources 1 through 4 and controls the same so as to enable them to send to the central nervous system simultaneous light radiation pulses, sound radiation pulses, heat radiation pulses and VHF electromagnetic field radiation pulses at an assigned pulse repetition rate thereby causing a state of inhibition in the patient under treatment.

The sources 1 through 4 are oriented in respect of the patient as follows:

the light radiation source 1 is placed at the level of the patient's eyes at a distance of 40 to 50 cm therefrom;

the sound radiation source 2 is placed in the proximity of the patient under treatment at a distance of 40 to 50 cm him;

the VHF electromagnetic field radiation source 3 is placed in the immediate proximity of the patient's ganglia, preferably at the level of carotid sinuses, 3 to 4 cm away from the patient's skin surface, as is the case with the present apparatus;

the heat radiation source 4 is also placed close to the patient's skin surface, i.e., in the immediate proximity of the face skin surface.

The sources 1 and 2 of light radiation and sound radiation, respectively, provide for an adequate and monotonous action thereof upon the visual and auditory analyzers.

According to the first embodiment of the apparatus for the treatment of neuropsychic and somatic disorders (FIG.1) the control unit 5 is a chopper with a cam mechanism 6 and contacts 7 through 10. The cams of the cam mechanism 6 are rotated by an electric motor 11 fed from a power supply unit, which is mains supply in this specific case. The electric motor 11 makes 60 r.p.m. thereby providing for an assigned pulse repetion rate of sixty pulses per minute of light radiation pulses, sound radiation pulses, VHF electromagnetic field radiation pulses and heat radiation pulses.

Light radiation source 1 is an incandescent lamp, blue in colour, which is electrically connected with a secondary winding 14 of a transformer 15 through a resistor 12 which serves to control the intensity of light radiation, contacts 7 of the vibrator and a switch 13, the transformer 15 being in turn electrically connected with the a.c. mains.

Sound radiation source 2 is a small-size loudspeaker 16 which is electrically connected with another secondary winding 21 of the transformer 15 through a diode 17, variable resistor 18, constant resistor 19, contacts 8 of the vibrator and switch 20. The sound radiation source provides for an imitation of rain drop noise. Besides, it can be employed to imitate the noise of the surf.

The VHF electromagnetic field radiation sources 3 is a push-pull oscillator 22 with an inductance coil 23 and resistor 24. Electrodes 25 make the load of said oscillator 22. The VHF electromagnetic field source is fitted with a switch 26 and controlled by the contacts 10 of the chopper.

The heat radiation source 4 is a hot spiral wire 27 which is surrounded by a screen 28 made of thermal insulator. The heat flow from the hot wire is removed by means of a blower 29 rotated by an electric motor 30. The heat radiation 4 is controlled by a switch 31 via the contacts 9 of the vibrator.

The present device is equipped with a common mains switch 32. The sources 1, 2 and 4 of light radiation, sound radiation and heat radiation, respectively, are mounted in the same casing (not shown in the drawing).

According to the second embodiment of the present apparatus the control unit 5 is an electronic timer made of cold cathode thyratron 33 with an electromagnetic relay 34, potentiometer 35 and resistor 36 in the plate circuit thereof. The electronic timer also includes a capacitor 37 and a resistor 38. The electronic timer has contacts 39 through 42. The potentiometer 35 controls the assigned pulses repetition rate of VHF electromagnetic field radiation, light radiation, sound radiation and heat radiation within the 10 to 100 pulses per second range, depending upon the patient's individual features.

The light radiation source 1 includes two incandescent lamps 43 and 44, green in colour, which are installed symmetrically in respect of the patient's sagittal sinus plane, at the eye level. This position provides for a better therapeutic effect because the patient is no longer subject to light beam convergence action whenever use is made of the Paurquignet effect. The light radiation intensity is controlled by the resistor 12. The light source 1 is equipped with a signal lamp 45.

The sound radiation source 2 is not substantially different from that in the first embodyment as described above. The difference between them consists in the presence of a capacitor 46. The sound radiation source 2 is equipped with a signal lamp 47.

The VHF electromagnetic field source is the same as that of the first embodiment, with a signal lamp 48 added.

The heat radiation source 4 is the same as that of the first embodiment but the blower is not shown in FIG.2.

The light radiation source 1 is electrically connected with a bridge-type d.c. power supply unit 50 through a resistor 12, switch 49 and contacts 39 of the electronic timer.

The sound radiation source 2 is electrically connected with the d.c. power supply unit 50 through a switch 51 and contacts 40 of the electronic timer.

The VHF electromagnetic field source 3 is equipped with a switch 52.

The hot spiral wire 27 of the heat radiation source 4 is fed from the mains through a switch 53, while the electric motor 30 of the blower 29 is fed from the d.c. power supply unit 50 via a variable resistor 54, switch 55 and contacts 42 of the electronic timer.

In accordance with the second embodiment of the present apparatus it can be switched on and off by means of the mains switch 56.

The principle of operation of the apparatus according to the first embodiments as follows.

The patient is either put to bed or seated in a deep arm-chair. The apparatus is moved to the head of said bed or chair. The electrodes 25 of the VHF electromagnetic field source are mounted at the level of carotid sinuses 3 to 4 centimetres away from the surfice of the neck skin.

The casing, with the sources 1, 2 and 4 of light radiation, sound radiation and heat radiation mounted therein, is placed above the patient's head at a distance from 40 to 50 centimetres from his face. The apparatus is switched on by means of the mains switch 32. The switches 13, 20, 26 and 31 are used to turn on the sources 1 through 4 of light radiation, sound radiation, VHF electromagnetic field radiation and heat radiation.

Due to this procedure the patient's receptors are simultaneously acted upon by 60 p/min pulses of blue light radiation, sound radiation which is made similar to the rain drop noise, VHF electromagnetic field radiation as well as by heat radiation, all these having an effect on the retina, on the auditory sensory endings, carotide sensory ganglia and thermal receptors in the face skin, respectively.

A simultaneous transmission of the pulses and heat radiation is due to contacts **7** through **10** of the vibrator. The length of the pulse effect added to that of heat radiation as obtained by means of the cam mechanism **6** amounts to 15×10^{-2} sec.

The nerve pulses due to the action by stimuli, such as pulses of light radiation, sound radiation, VHF electromagnetic pulse radiation, as well as heat radiation, are transmitted to the thalamus opticus (sensory collector) via sensory nerve filament and sympathetic nerve trunk.

The simultaneous arrival of brain wave impulses at the thalamus opticus provides for an intensified rhythmic activity of its neural formations due to the interference effect. The rhythm of superimposed stimuli as then synchronized with the activities of the thalamus opticus neural formations (external synchronization). The rhythm as assigned by the apparatus is adopted by the subcortical neural formations being transmitted from the thalamus opticus to the subcortical fields of the visual, auditory and cutaneous analyzers. In the appropriate cortical centres of visual, auditory, and cutaneous analyzers. In the appropriate cortical centres of visual analysers the mono-tonous pulses cause focuses of inhibition to be formed which would intensify in strength and irradiate in space in the course of the same treatment session as well as due to the successive sessions. The primary focus inhibition, once and whenever it occured, would spread all over the cortical field bringing about natural sleep. During the initial treatment sessions the extent of irradiation may be quite insignificant because of the patient's inhibition inertia due to this or that neuropsychic disorder in the patient under treatment. In the course of successive treatment session the trace responses as retained in above-said cortical centres would cause the stimuli to be accumulated so that the inhibition tends to be intensified.

Besides, in the course of successive treatment sessions new conditioned reflex associations are put into effect either because of the procedure itself or due to its specific time, or because of any other isolated factor, all contributing to the irradiation of inhibition phenomena throughout the cortical centres in the cerebral hemispheres.

The whole system of stimuli which is addressed to the patient's organism makes use of the first signal system channels, i.e., the receptor zones of the appropriate analyzers, so that the second signal system channels (mind, intellect, psyche) are avoided thereby providing for a curative effect, no matter the patient's psychic condition or his attitude towards the treatment procedure. Besides, the second signal system channels, once unblocked, affords some opportunities for a simultaneous suggestive and rhythm therapy action thereby increasing the treatment effect.

The present invention makes use of a plurality of sources which are distant stimuli of different physical nature being such oscillatory phenomena as act upon the receptors to the adequate physiologic extents giving rise to no alterations which would exceed the physiologic constants limits involved. That is why the action exerted upon the patient's organism is mild and humane. Besides, provision is made both for a separate control of every factor and for a joint application of a required combination thereof so that the electrohypnotic treatment could be quite individual.

The second embodiment of this apparatus is based on the same principles of operation as the first embodiment thereof. The difference can be described as follows.

By means of a potentiometer **35** in conjunction with the electronic timer the pulse repetition rate of light pulses, sound pulses, VHF electromagnetic field pulses, as well as heat radiation, can be made to vary with due regard to the patient's individual characteristics.

The possibility of alterations both in pulse repetition rate and in the operation made of the apparatus under consideration provides for a controlled variation of biorhythms in the appropriate range as suggested by the principle of rhythm assimilation named after A.A.Ukhtomsky.

In the second embodiment of the apparatus, the light radiation pulses are sent by incandescent lamps **43** and **44**, green in color, which are mounted at eye level and symmetrical with respect to sinus plane, so that the convergence effect is eliminated (the eyeball convergence would give birth to a stimulation focus thereby preventing the progress of somnolescent inhibition).

The application of green light permits utilization of Paurquignet's effect to achieve an optimum influence on the visual analyzers.

The following procedure is recommended for the treatment of neuropsychic and somatic disorders using the apparatus of the present invention.

At a first treatment session the duration of the VHF exposure should be 10 minutes, at a second one - 15 minutes, at a third one - 20 minutes, at a fourth one - 25 minutes, and at all succeeding sessions it should be 30 minutes. The other radiation sources may be operated for 30 to 60 minutes. On the average, 30 minutes is required for one treatment-session. The patient may lie with his face upwards and eyes open, though it would be better for him to close his eyes to be ready to fall asleep. At the first treatment-session, the intensities of sound-, light-, and heat-radiation are selected to suit the particular patient.

Using the second embodiment of the apparatus, during the first minutes of the treatment session, the pulse repetition rate should be 20 to 40 pulses per minute (instead of 50 to 60 pulses per minute which is the normal case).

During the successive treatment-sessions the intensity of the exposure may remain the same as that used at the first session. The pulse repetition rate, however, may be gradually decreased later on from session to session.

The development of inhibition processes at a treatment-session is characterized by a very gradual progress. This manifests itself in a somewhat slower pulse, lower integument temperature and arterial pressure, relaxation of sceletal muscles, onset of somnolescence.

With each successive session the onset of all these phenomena takes less time while the phenomena themselves become more pronounced, bringing about a general improvement of the patient's state of health accompanied with a better sleep at night.

During the treatment session, the patient experiences gentle, tranquilizing sensations, which result in psychical relaxation and, gradual transference from vigilance to sleep. An active reproduction of inhibition of processes, if regularly repeated, each treatment-session makes a sort of training the patient for a better neurodynamic performance.

Clinical applications of the proposed apparatus are as follows. The apparatus of the present invention may be used for the treatment of patients in the 4 to 70 years bracket. One hundred patients, 4 to 18 years old, and 200 patients over **18** years old underwent the appropriate treatment. Among these patients were the following cases: neurasthenia, neurosis involving delusion, insomnia, asthenic, depressive and reactive states, postinfections and traumatic cerebrasthenia, diencephalic syndrome, vegeto-vascular dystonia, hypertension, bronchial asthma, stammering, rheumatic Sydenham's chorea.

The treatment was effected with respect to ambulatory clinic- and hospital patients.

No side effects, complications or harmful actions were observed.

Contraindications: infectious diseases, cancer, schizophrenia in its advanced stage (e.g., involving delirious phenomena), and hyperthyreosis.

The apparatus of the present invention for treating neuropsychic and somatic diseases is a physical therapy apparatus to be used for treating insomnia, hypertension, infantile stammering, and other diseases arising from nervous exhaustion and prostration, nervous break-down.

This apparatus is reliable, convenient, safe and simple to control and use, portable, and if necessary, may be placed at the bed-head when dealing with a bed case.

The use of the apparatus is not confined to any age-bracket. Even infants were effectively treated. The apparatus may be used in hospitals as well as in outpatient clinics.

The operation of the apparatus is quite economical: its power consumption is very small. The treatment sessions are effected by para medical personnel. Not to disturb the patient a remote control from an adjacent room may be exercised.

For the first time in medical practice, thanks to the apparatus of the present invention, use is made of a complex of four physical factors acting simultaneously on the respective receptors, said factors operating with predetermined rhythm pattern, and the sources of respective radiation being placed at a predetermined distance from the patient's receptors, so that it could bring about such a state of the nervous system which is characteristic of rest and sleep.

The synergetic action of the factors used in the apparatus result not in a simple sum of respective effects produced by each of said factors, but creates quite a new qualitative phenomenon, wherein interference - and resonance-processes, associated with the radiation from the pertinent sources, cause the inhibition process to be irradiated within the brain centres, which is a distinguishing feature of the present type rhythm therapy.

What is claimed is:

1. An apparatus for the treatment of neuropsychic and somatic diseases, said apparatus comprising a control unit means for simultaneously periodically activating apparatus connected thereto in a pulsed manner at a selected repetition rate; a light-radiation source essentially comprising at least one electric bulb electrically connected with said control unit means, such that light pulses are generated by said light-radiation source influencing the central nervous system of the patient having one of said diseases who faces said source; a sound radiation source essentially comprising an electric generator of pulse oscillations, a small dynamic loud-speaker means for converting said oscillations into acoustic signals imitating the sound of falling drops of water, said electric generator being electrically connected with said control unit means for actuation thereby whereby sound pulses of said sound-radiation source influence the central nervous system of the patient; a VHF electromagnetic-field source essentially comprising a generator of sinusoidal electric oscillations of the waves metric range, electrically connected with said control unit means and disposed in proximity to the patient's ganglia, pulses from said VHF electromagnetic field source acting upon the nervous system of the patient; said light-radiation source producing an adequate and monotonous action of the light-radiation of the patient's visual analyzer; said sound radiation source producing an adequate and monotonous influence of the sound radiation on the patient's auditory analyzer; said control unit means controlling said light-radiation, sound-radiation and VHF electromagnetic-field source so as to provide for a simultaneous generation of said light, sound and VHF electromagnetic pulses and directing said pulses to act upon the patient's central nervous system, all said pulses following at a predetermined repetition rate and ensuring suppression of the patient's nervous activity; and a power supply means to supply said lightradiation, sound-radiation and VHF electromagnetic-field sources and said control unit.

2. An apparatus according to claim **9**, further including a heat radiation source comprising an electrical coil surrounded by a heat-insulating shield and provided with an air-forcing motor, said heat source being disposed in the proximity of the patient's integument, said motor of said heat-radiation source being electrically connected with said control unit means and controlled thereby so as to ensure a heat irradiation flow onto the patient's integument at the same time as said VHF electromagnetic-field pulses and sound pulses are sent at an assigned repetition rate, so as to contribute to the inhibition of the patient's nervous activity.

3. An apparatus according to claim **1**, wherein said light-radiation source comprises at least one blue-light lamp adapted to be placed at the level of the patient's eyes.

4. An apparatus according to claim **1**, wherein said light-radiation source comprises at least one green-light lamp adapted to be placed at the level of the patient's eyes.

5. An apparatus according to claim **4**, wherein an even number of said green-light lamps are provided and are adapted to be arranged symmetrically with respect to the saggital plane of the patient and at a minimal angle relative to the axis of the respective eye.

6. An apparatus according to claim **1**, wherein said electromagnetic field source comprises an oscillator means having a control means, and wherein said control unit means is a chopper provided with a rotatable cam gear having electric contacts which periodically connect said power supply source with the circuit of the incandescent lamps, with the electric generator of pulse oscillations which are converted in a small dynamic loudspeaker into acoustic signals imitating the sound of falling drops of water, as well as grounding the control means of the oscillator means.

7. An apparatus according to claim **1**, wherein said electromagnetic field source comprises an oscillator means having a control means, and wherein said control unit means comprises an electronic timer means having at least one thyratron and a resistor-capacitor delay trigger circuit means therefor, said thyratron including an anode circuit comprising a series connected electro-mechanical relay having contacts connecting the power supply source with the circuit of the incandescent lamps, with the electric generator of pulse oscillations converted in a small dynamic loudspeaker into acoustic signals imitating the sound of falling drops of water, as well as grounding the control means of the oscillator means.

8. An apparatus according to claim **7**, further including a heat radiation source, and wherein said electronic timer having at least one thyratron with an anode circuit comprising a series connected electro-mechanical relay having contacts connecting the power supply source with the circuit of the incandescent lamps, with the electric generator of pulse oscillations converted in a small dynamic loudspeaker into acoustic signals imitating the sound of falling drops of water, as well as grounding the grid of the oscillator means, further connects the power supply with said heat radiation source, and has a potentiometer electrically connected with said timer to provide for a predetermined repetition rate of said VHF electromagnetic-field, light and sound pulses, and for the control of said heat radiation depending on the individual features of the patient.

* * * * *

United States Patent [19]

Malech

[11] **3,951,134**

[45] **Apr. 20, 1976**

[54] **APPARATUS AND METHOD FOR REMOTELY MONITORING AND ALTERING BRAIN WAVES**

[75] Inventor: **Robert G. Malech**, Plainview, N.Y.

[73] Assignee: **Dorne & Margolin Inc.**, Bohemia, N.Y.

[22] Filed: **Aug. 5, 1974**

[21] Appl. No.: **494,518**

[52] **U.S. Cl.** .... **128/2.1 B**
[51] **Int. Cl.^{2}** **A61B 5/04**
[58] **Field of Search** .... 128/1 C, 1 R, 2.1 B, 128/2.1 R, 419 R, 422 R, 420, 404, 2 R, 2 S, 2.05 R, 2.05 V, 2.05 F, 2.06 R; 340/248 A, 258 A, 258 B, 258 D, 229

[56] **References Cited**

UNITED STATES PATENTS

2,860,627	11/1958	Harden et al.	128/2.1 B
3,096,768	7/1963	Griffith, Jr.	128/420
3,233,450	2/1966	Fry	128/2.1 R
3,483,860	12/1969	Namerow	128/2.05 F
3,495,596	2/1970	Condict	128/1 C
3,555,529	1/1971	Brown et al.	128/2.1 R
3,773,049	11/1973	Rabichev et al.	128/1 C
3,796,208	3/1974	Bloice	128/2 S

Primary Examiner—William E. Kamm
Attorney, Agent, or Firm—Darby & Darby

[57] **ABSTRACT**

Apparatus for and method of sensing brain waves at a position remote from a subject whereby electromagnetic signals of different frequencies are simultaneously transmitted to the brain of the subject in which the signals interfere with one another to yield a waveform which is modulated by the subject's brain waves. The interference waveform which is representative of the brain wave activity is re-transmitted by the brain to a receiver where it is demodulated and amplified. The demodulated waveform is then displayed for visual viewing and routed to a computer for further processing and analysis. The demodulated waveform also can be used to produce a compensating signal which is transmitted back to the brain to effect a desired change in electrical activity therein.

11 Claims, 2 Drawing Figures

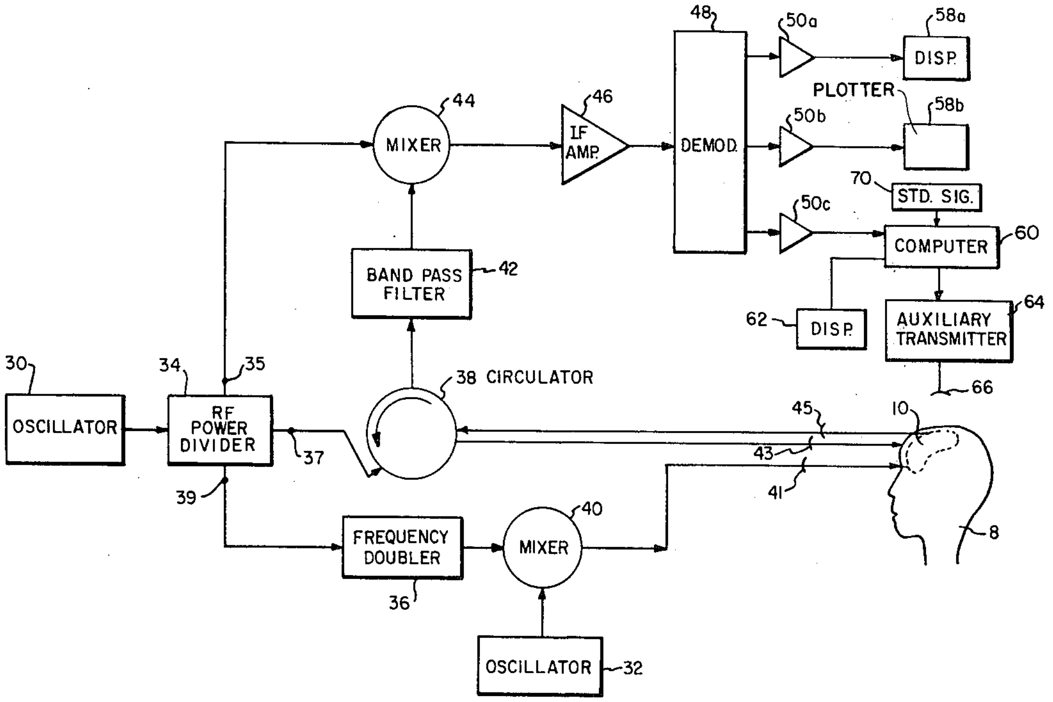
48
50a
58a
DISP.
PLOTTER
58b
50b
44
MIXER
46
IF AMP.
DEMOD.
70
STD. SIG.
50c
COMPUTER
60
BAND PASS FILTER
42
62
DISP.
AUXILIARY TRANSMITTER
64
66
30
OSCILLATOR
34
35
RF POWER DIVIDER
37
38 CIRCULATOR
45
10
43
41
8
39
FREQUENCY DOUBLER
36
MIXER
40
OSCILLATOR
32

FIG. 1

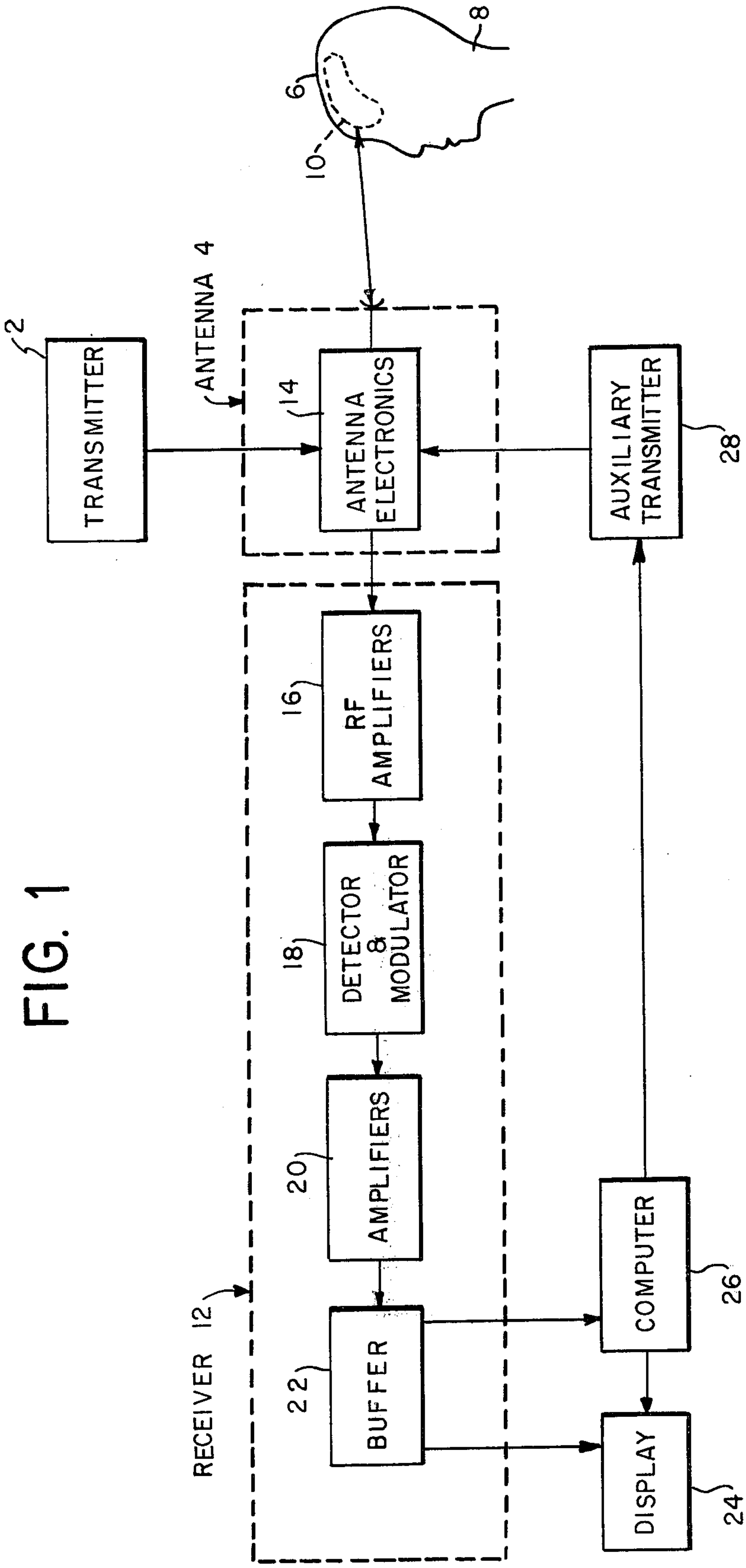
TRANSMITTER
2
ANTENNA 4
ANTENNA ELECTRONICS
14
AUXILIARY TRANSMITTER
28
6
8
10
RF AMPLIFIERS
16
DETECTOR & MODULATOR
18
AMPLIFIERS
20
BUFFER
22
RECEIVER 12
COMPUTER
26
DISPLAY
24

FIG. 2

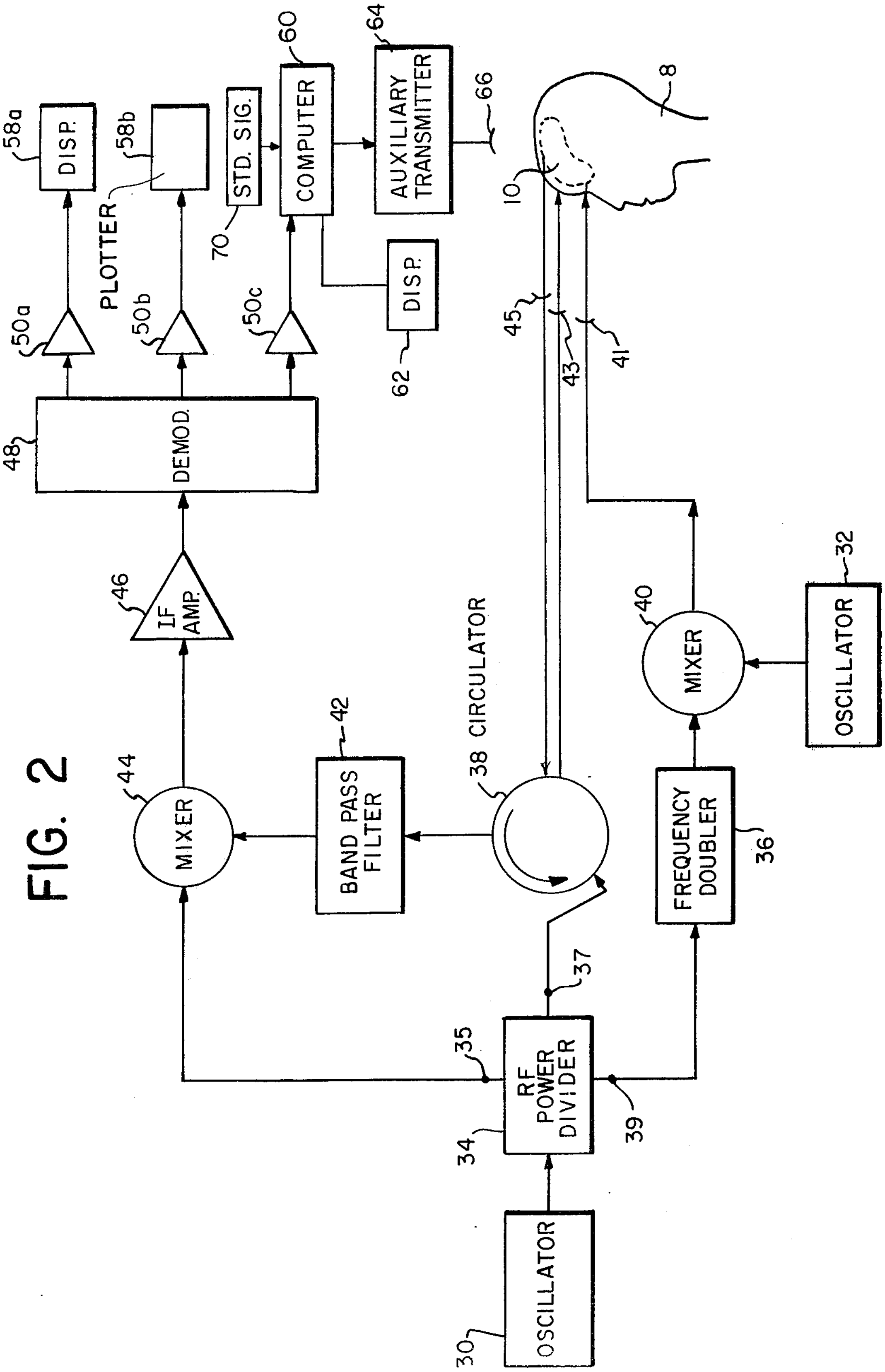
DISP.
PLOTTER
STD. SIG.
COMPUTER
AUXILIARY TRANSMITTER
DISP.
DEMOD.
IF AMP.
MIXER
BAND PASS FILTER
38 CIRCULATOR
MIXER
OSCILLATOR
FREQUENCY DOUBLER
RF POWER DIVIDER
OSCILLATOR

APPARATUS AND METHOD FOR REMOTELY MONITORING AND ALTERING BRAIN WAVES

BACKGROUND OF THE INVENTION

Medical science has found brain waves to be a useful barometer of organic functions. Measurements of electrical activity in the brain have been instrumental in detecting physical and psychic disorder, measuring stress, determining sleep patterns, and monitoring body metabolism.

The present art for measurement of brain waves employs electroencephalographs including probes with sensors which are attached to the skull of the subject under study at points proximate to the regions of the brain being monitored. Electrical contact between the sensors and apparatus employed to process the detected brain waves is maintained by a plurality of wires extending from the sensors to the apparatus. The necessity for physically attaching the measuring apparatus to the subject imposes several limitations on the measurement process. The subject may experience discomfort, particulary if the measurements are to be made over extended periods of time. His bodily movements are restricted and he is generally confined to the immediate vicinity of the measuring apparatus. Furthermore, measurements cannot be made while the subject is conscious without his awareness. The comprehensiveness of the measurements is also limited since the finite number of probes employed to monitor local regions of brain wave activity do not permit observation of the total brain wave profile in a single test.

SUMMARY OF THE INVENTION

The present invention relates to apparatus and a method for monitoring brain waves wherein all components of the apparatus employed are remote from the test subject. More specifically, high frequency transmitters are operated to radiate electromagnetic energy of different frequencies through antennas which are capable of scanning the entire brain of the test subject or any desired region thereof. The signals of different frequencies penetrate the skull of the subject and impinge upon the brain where they mix to yield an interference wave modulated by radiations from the brain's natural electrical activity. The modulated interference wave is re-transmitted by the brain and received by an antenna at a remote station where it is demodulated, and processed to provide a profile of the suject's brain waves. In addition to passively monitoring his brain waves, the subject's neurological processes may be affected by transmitting to his brain, through a transmitter, compensating signals. The latter signals can be derived from the received and processed brain waves.

OBJECTS OF THE INVENTION

It is therefore an object of the invention to remotely monitor electrical activity in the entire brain or selected local regions thereof with a single measurement.

Another object is the monitoring of a subject's brain wave activity through transmission and reception of electromagnetic waves.

Still another object is to monitor brain wave activity from a position remote from the subject.

A further object is to provide a method and apparatus for affecting brain wave activity by transmitting electromagnetic signals thereto.

DESCRIPTION OF THE DRAWINGS

Other and further objects of the invention will appear from the following description and the accompanying drawings, which form part of the instant specification and which are to be read in conjunction therewith, and in which like reference numerals are used to indicate like parts in the various views;

FIG. **1** is a block diagram showing the interconnection of the components of the apparatus of the invention;

FIG. **2** is a block diagram showing signal flow in one embodiment of the apparatus.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring to the drawings, specifically FIG. **1**, a high frequency transmitter **2** produces and supplies two electromagnetic wave signals through suitable coupling means **14** to an antenna **4**. The signals are directed by the antenna **4** to the skull **6** of the subject **8** being examined. The two signals from the antenna **4**, which travel independently, penetrate the skull **6** and impinge upon the tissue of the brain **10**.

Within the tissue of the brain **10**, the signals combine, much in the manner of a conventional mixing process technique, with each section of the brain having a different modulating action. The resulting waveform of the two signals has its greatest amplitude when the two signals are in phase and thus reinforcing one another. When the signals are exactly 180° out of phase the combination produces a resultant waveform of minimum amplitude. If the amplitudes of the two signals transmitted to the subject are maintained at identical levels, the resultant interference waveform, absent influences of external radiation, may be expected to assume zero intensity when maximum interference occurs, the number of such points being equal to the difference in frequencies of the incident signals. However, interference by radiation from electrical activity within the brain **10** causes the waveform resulting from interference of the two transmitted signals to vary from the expected result, i.e., the interference waveform is modulated by the brain waves. It is believed that this is due to the fact that brain waves produce electric charges each of which has a component of electromagnetic radiation associated with it. The electromagnetic radiation produced by the brain waves in turn reacts with the signals transmitted to the brain from the external source.

The modulated interference waveform is re-transmitted from the brain **10**, back through the skull **6**. A quantity of energy is re-transmitted sufficient to enable it to be picked up by the antenna **4**. This can be controlled, within limits, by adjusting the absolute and relative intensities of the signals, originally transmitted to the brain. Of course, the level of the transmitted energy should be kept below that which may be harmful to the subject.

The antenna passes the received signal to a receiver **12** through the antenna electronics **14**. Within the receiver the wave is amplified by conventional RF amplifiers **16** and demodulated by conventional detector and modulator electronics **18**. The demodulated wave, representing the intra-brain electrical activity, is amplified by amplifiers **20** and the resulting information in electronic form is stored in buffer circuitry **22**. From the buffers **22** the information is fed to a suitable visual display **24**, for example one employing a cathode ray tube, light emitting diodes, liquid crystals, or a mechanical plotter. The information may also be channeled to a computer **26** for further processing and analysis with the output of the computer displayed by heretofore mentioned suitable means.

In addition to channeling its information to display devices **24**, the computer **26** can also produce signals to control an auxiliary transmitter **28**. Transmitter **28** is used to produce a compensating signal which is transmitted to the brain **10** of the subject **8** by the antenna **4**. In a preferred embodiment of the invention, the compensating signal is derived as a function of the received brain wave signals, although it can be produced separately. The compensating signals affect electrical activity within the brain **10**.

Various configurations of suitable apparatus and electronic circuitry may be utilized to form the system generally shown in FIG. **1** and one of the many possible configurations is illustrated in FIG. **2**. In the example shown therein, two signals, one of 100 MHz and the other of 210 MHz are transmitted simultaneously and combine in the brain **10** to form a resultant wave of frequency equal to the difference in frequencies of the incident signals, i.e., 110 MHz. The sum of the two incident frequencies is also available, but is discarded in subsequent filtering. The 100 MHz signal is obtained at the output **37** of an RF power divider **34** into which a 100 MHz signal generated by an oscillator **30** is injected. The oscillator **30** is of a conventional type employing either crystals for fixed frequency circuits or a tunable circuit set to oscillate at 100 MHz. It can be a pulse generator, square wave generator or sinusoidal wave generator. The RF power divider can be any conventional VHF, UHF or SHF frequency range device constructed to provide, at each of three outputs, a signal identical in frequency to that applied to its input.

The 210 MHz signal is derived from the same 100 MHz oscillator **30** and RF power divider **34** as the 100 MHz signal, operating in concert with a frequency doubler **36** and 10 MHz oscillator **32**. The frequency doubler can be any conventional device which provides at its output a signal with frequency equal to twice the frequency of a signal applied at its input. The 10 MHz oscillator can also be of conventional type similar to the 100 MHz oscillator herebefore described. A 100 MHz signal from the output **39** of the RF power divider **34** is fed through the frequency doubler **36** and the resulting 200 MHz signal is applied to a mixer **40**. The mixer **40** can be any conventional VHF, UHF or SHF frequency range device capable of accepting two input signals of differing frequencies and providing two output signals with frequencies equal to the sum and difference in frequencies respectively of the input signals. A 10 MHz signal from the oscillator **32** is also applied to the mixer **40**. The 200 MHz signal from the doubler **36** and the 10 MHz signal from the oscillator **32** combine in the mixer **40** to form a signal with a frequency of 210 MHz equal to the sum of the frequencies of the 200 MHz and 10 MHz signals.

The 210 MHz signal is one of the signals transmitted to the brain **10** of the subject being monitored. In the arrangement shown in FIG. **2**, an antenna **41** is used to transmit the 210 MHz signal and another antenna **43** is used to transmit the 100 MHz signal. Of course, a single antenna capable of operating at 100 MHz and 210 MHz frequencies may be used to transmit both signals. The scan angle, direction and rate may be controlled mechanically, e.g., by a reversing motor, or electronically, e.g., by energizing elements in the antenna in proper synchronization. Thus, the antenna(s) can be of either fixed or rotary conventional types.

A second 100 MHz signal derived from output terminal **37** of the three-way power divider **34** is applied to a circulator **38** and emerges therefrom with a desired phase shift. The circulator **38** can be of any conventional type wherein a signal applied to an input port emerges from an output port with an appropriate phase shift. The 100 MHz signal is then transmitted to the brain **10** of the subject being monitored via the antenna **43** as the second component of the dual signal transmission. The antenna **43** can be of conventional type similar to antenna **41** herebefore described. As previously noted, these two antennas may be combined in a single unit.

The transmitted 100 and 210 MHz signal components mix within the tissue in the brain **10** and interfere with one another yielding a signal of a frequency of 110 MHz, the difference in frequencies of the two incident components, modulated by electromagnetic emissions from the brain, i.e., the brain wave activity being monitored. This modulated 110 MHz signal is radiated into space.

The 110 MHz signal, modulated by brain wave activity, is picked up by an antenna **45** and channeled back through the circulator **38** where it undergoes an appropriate phase shift. The circulator **38** isolates the transmitted signals from the received signal. Any suitable diplexer or duplexer can be used. The antenna **45** can be of conventional type similar to antennas **41** and **43**. It can be combined with them in a single unit or it can be separate. The received modulated 110 MHz signal is then applied to a band pass filter **42**, to eliminate undesirable harmonics and extraneous noise, and the filtered 110 MHz signal is inserted into a mixer **44** into which has also been introduced a component of the 100 MHz signal from the source **30** distributed by the RF power divider **34**. The filter **42** can be any conventional band pass filter. The mixer **44** may also be of conventional type similar to the mixer **40** herebefore described.

The 100 MHz and 110 MHz signals combine in the mixer **44** to yield a signal of frequency equal to the difference in frequencies of the two component signals, i.e., 10 MHz still modulated by the monitored brain wave activity. The 10 MHz signal is amplified in an IF amplifier **46** and channeled to a demodulator **48**. The IF amplifier and demodulator **48** can both be of conventional types. The type of demodulator selected will depend on the characteristics of the signals transmitted to and received from the brain, and the information desired to be obtained. The brain may modulate the amplitude, frequency and/or phase of the interference waveform. Certain of these parameters will be more sensitive to corresponding brain wave characteristics than others. Selection of amplitude, frequency or phase demodulation means is governed by the choice of brain wave characteristic to be monitored. If desired, several different types of demodulators can be provided and used alternately or at the same time.

The demodulated signal which is representative of the monitored brain wave activity is passed through audio amplifiers **50** *a*, *b*, *c* which may be of conventional type where it is amplified and routed to displays **58** *a*, *b*, *c* and a computer **60**. The displays **58** *a*, *b*, *c* present the raw brain wave signals from the amplifiers

50 *a*, *b*, *c*. The computer **60** processes the amplified brain wave signals to derive information suitable for viewing, e.g., by suppressing, compressing, or expanding elements thereof, or combining them with other information-bearing signals and presents that information on a display **62**. The displays can be conventional ones such as the types herebefore mentioned employing electronic visual displays or mechanical plotters **58***b*. The computer can also be of conventional type, either analog or digital, or a hybrid.

A profile of the entire brain wave emission pattern may be monitored or select areas of the brain may be observed in a single measurement simply by altering the scan angle and direction of the antennas. There is no physical contact between the subject and the monitoring apparatus. The computer **60** also can determine a compensating waveform for transmission to the brain **10** to alter the natural brain waves in a desired fashion. The closed loop compensating system permits instantaneous and continuous modification of the brain wave response pattern.

In performing the brain wave pattern modification function, the computer **60** can be furnished with an external standard signal from a source **70** representative of brain wave activity associated with a desired nuerological response. The region of the brain responsible for the response is monitored and the received signal, indicative of the brain wave activity therein, is compared with the standard signal. The computer **60** is programmed to determine a compensating signal, responsive to the difference between the standard signal and received signal. The compensating signal, when transmitted to the monitored region of the brain, modulates the natural brain wave activity therein toward a reproduction of the standard signal, thereby changing the neurological response of the subject.

The computer **60** controls an auxiliary transmitter **64** which transmits the compensating signal to the brain **10** of the subject via an antenna **66**. The transmitter **64** is of the high frequency type commonly used in radar applications. The antenna **66** can be similar to antennas **41**, **43** and **45** and can be combined with them. Through these means, brain wave activity may be altered and deviations from a desired norm may be compensated. Brain waves may be monitored and control signals transmitted to the brain from a remote station.

It is to be noted that the configuration described is one of many possibilities which may be formulated without departing from the spirit of my invention. The transmitters can be monostratic or bistatic. They also can be single, dual, or multiple frequency devices. The transmitted signal can be continuous wave, pulse, FM, or any combination of these as well as other transmission forms. Typical operating frequencies for the transmitters range from 1 MHz to 40 GHz but may be altered to suit the particular function being monitored and the characteristics of the specific subject.

The individual components of the system for monitoring and controlling brain wave activity may be of conventional type commonly employed in radar systems.

Various subassemblies of the brain wave monitoring and control apparatus may be added, substituted or combined. Thus, separate antennas or a single multimode antenna may be used for transmission and reception. Additional displays and computers may be added to present and analyze select components of the monitored brain waves.

Modulation of the interference signal retransmitted by the brain may be of amplitude, frequency and/or phase. Appropriate demodulators may be used to decipher the subject's brain activity and select components of his brain waves may be analyzed by computer to determine his mental state and monitor his thought processes.

As will be appreciated by those familiar with the art, apparatus and method of the subject invention has numerous uses. Persons in critical positions such as drivers and pilots can be continuously monitored with provision for activation of an emergency device in the event of human failure. Seizures, sleepiness and dreaming can be detected. Bodily functions such as pulse rate, heartbeat reqularity and others also can be monitored and occurrences of hallucinations can be detected. The system also permits medical diagnoses of patients, inaccessible to physicians, from remote stations.

What is claimed is:

1. Brain wave monitoring apparatus comprising

means for producing a base frequency signal,

means for producing a first signal having a frequency related to that of the base frequency and at a predetermined phase related thereto,

means for transmitting both said base frequency and said first signals to the brain of the subject being monitored,

means for receiving a second signal transmitted by the brain of the subject being monitored in response to both said base frequency and said first signals,

mixing means for producing from said base frequency signal and said received second signal a response signal having a frequency related to that of the base frequency, and

means for interpreting said response signal.

2. Apparatus as in claim **1** where said receiving means comprises

means for isolating the transmitted signals from the received second signals.

3. Apparatus as in claim **2** further comprising a band pass filter with an input connected to said isolating means and an output connected to said mixing means.

4. Apparatus as in claim **1** further comprising means for amplifying said response signal.

5. Apparatus as in claim **4** further comprising means for demodulating said amplified response signal.

6. Apparatus as in claim **5** further comprising interpreting means connected to the output of said demodulator means.

7. Apparatus according to claim **1** further comprising

means for producing an electromagnetic wave control signal dependent on said response signal, and

means for transmitting said control signal to the brain of said subject.

8. Apparatus as in claim **7** wherein said transmitting means comprises means for directing the electromagnetic wave control signal to a predetermined part of the brain.

9. A process for monitoring brain wave activity of a subject comprising the steps of

transmitting at least two electromagnetic energy signals of different frequencies to the brain of the subject being monitored,

receiving an electromagnetic energy signal resulting from the mixing of said two signals in the brain modulated by the brain wave activity and retrans-

mitted by the brain in response to said transmitted energy signals, and,

interpreting said received signal.

10. A process as in claim **9** further comprising the step of transmitting a further electromagnetic wave signal to the brain to vary the brain wave activity.

11. A process as in claim **10** wherein the step of transmitting the further signals comprises

obtaining a standard signal,

comparing said received electromagnetic energy signals with said standard signal,

producing a compensating signal corresponding to the comparison between said received electrogagnetic energy signals and the standard signal, and

transmitting the compensating signals to the brain of the subject being monitored.

* * * * *

A simple MK 484 IC tuned rf (trf) radio

(updated 4 Jun 01)

In 1970 Ferranti started the development of a single-chip AM tuner, and by 1973 it became widely available for use. The 414 is no longer in production, but a follow on IC, the MK484 recently became available, and incorporates essentially the same features. It is a 3 terminal device which packs ten transistors providing 3 stages of rf amplification with an input impedance of about 4 Megohms, and giving 72 dB of gain with about 30 millivolts of audio output (rms). Thanks to the editors of *Practical Wireless* a ham magazine in the UK, and Dr. Philip Miller Tate of Kingston University, who has probably built a dozen or so of these in one guise or another, and sent me one fully assembled, here is one adaptation of this versatile little chip which gives you a nice little portable radio which is stable and virtually idiot proof to build, needs only a 1.5 v battery, and requires a tiny thimble full of parts to give you near-superhet performance. Here is the schematic of the circuit that *PW* came up with:

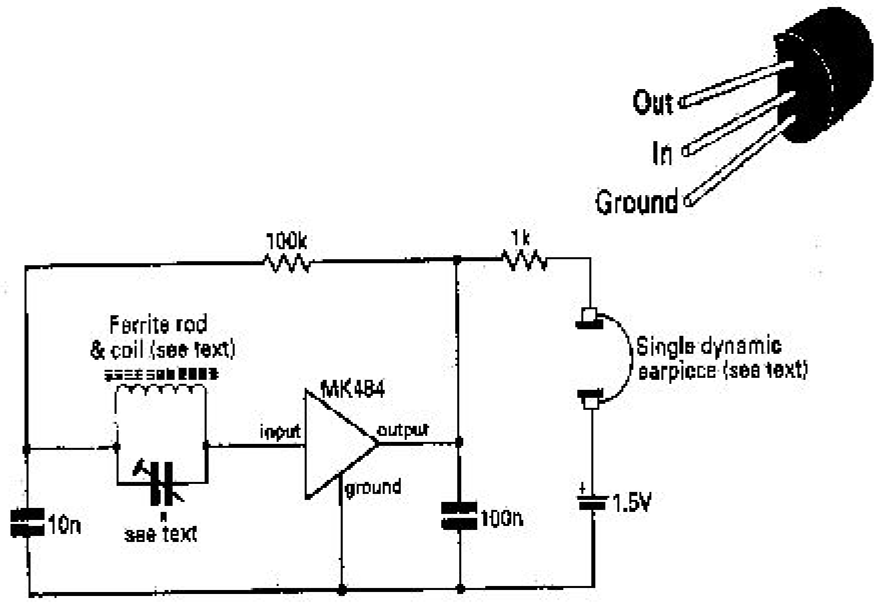

The frequency range of the chip is about 200 kHz to 3 MHz. The coil/capacitor selection for the tank circuit should fall into this range. For small size and shirt pocket portability, you can use a ferrite core coil as shown. I have tried it with a variety of tank coils, and they all work fine. Notice there is no antenna - don't need one. I have found that the set "likes" to have a ground connection. You can improve reception somewhat with an external antenna and ground, but do so lightly; I found that a small antenna coupling coil connected to antenna and ground in the vicinity of the tank circuit, loosely coupled, works nicely.

While the designers had in mind a pc board assembly, I have found that "dead bug squashed flat" works well too; so did the *PW* staff, as you could see by the rendition on their technical features page. Phil used a small masonite board and epoxied brass strips on it to make his solder connections (see it at Xtal Set Photos by Dr. Whiz). Mine was assembled on a 6 lug tie strip, which was then screwed into a board.

Now for some notes and fashion tips on putting this together and using it:

100 nF = 0.1 uF (that's a 104)   10 nF = 0.01 uF (103)

The dynamic earpiece mentioned can be one of the balanced armature types found in sound powered headsets (and most telephone earpieces - don't tell Ma Bell I told you). I used successfully magnetic earphones; 2000 ohm phones, if used, can replace the 1 kohm resistor. I also tried a 1 kohm dynamic earplug from Mouser, also replacing the 1 k resistor, and it worked, but it is not a very sensitive plug, and is not recommended except "in town". I also used a high impedance crystal earphone across the 1k resistor with good success - I think a 1.5 k resistor worked a little better, however.

You can connect a small audio amp to this set, connecting it across the 100 nF capacitor - the input capacitor to the amp should be about 100 nF.

Did I mention that this chip incorporates automatic gain control (AGC)? The resistance between the (+) side of the battery and the junction of the 1 k and 100 k resistors determines this, and you might want to play around with this a bit, depending on how you get the audio out. If you feed an audio amp with the set, the 1k AGC resistor might need to be adjusted, but don't go below about 500 ohms. The AGC and the Q of the tank circuit pretty much determine the ability of this set to perform as well as it does. Even with rather strong stations nearby, you will be slowly tuning and hearing a strong station in the background, all over the dial, so to speak (remember, this isn't a superhet), and then a station will pop up clearly out of the background noise. You can also null the strong locals a bit by turning the set so that its coil is pointed in a quieter direction. This set might be neat to use with a loop antenna for the coil.

You can get this set to oscillate (I did it when playing around with the AGC), but you don't want it to. You should separate the power and audio leads from the tuned circuit, and keep your leads short (separated is perhaps more important than short - I used long clip leads to connect different coil/cap combinations to the set).

Don't buy a new battery - this baby only draws a half a milliamp when its cooking. With a spanking new alkaline battery, the high gain front end is susceptible to oscillation, and a higher value R1 might be needed.

Austin Hellier from down under posted this on Rap 'n Tap for those of you looking for some mods to this simple rig which he claims will improve its performance - takes only a couple of 1N914 diodes and a bit of wire to accomplish:

"Thought you might be interested in a couple of mods for your MK484 receiver - these have proved very popular over here in Australia - Dick Smith Electronics puts out their Funway 1 series, and one of the projects is this kind of receiver, driven from a simple cell made of zinc coated nails, copper strips and - good old Aussie beer...

Now to the mods - take two 1N914 diodes in series from pin 1 to ground - this ensures no overvoltage to the chip - then take a 1K5 resistor to the 1.5 volt battery. You can use a 6K8 resistor instead, if you want to use a 9 volt battery.

Instead of using a single winding and parallel polycap, I use an inductive coupling method. Five turns of telephone jumper wire (brittle stuff, easy to work with though) and connect that between the junction of the 100K/0.01u - the other end goes to pin 2. Both ends of the main winding (50 turns around a 100 x 9mm rod in parallel with a 220pF cap)is then floating free from the rest of the circuit. This prevents loading of the tuning circuit, prevents acoustic feedback, and the volume, clarity and selectivity have to be heard to be believed - hope this helps you and your students, club, friends etc."

I suspect that either of the two modifications will improve the set. However, the diode trick will cost you some battery life, and the inductive coupling does cut down on the sensitivity with the unadorned set.

You Elmers can go HERE for a layout and construction instructions I use with my students.
Oh yeah, here is a picture of the "student model, without the battery installed or the capacitor knob. The base shown is 3 x 6, but I use 3 x 5 with the students; this is the prototype. Assembly time with prepared coil and capacitor was about 1.5 hours for very novice builders.

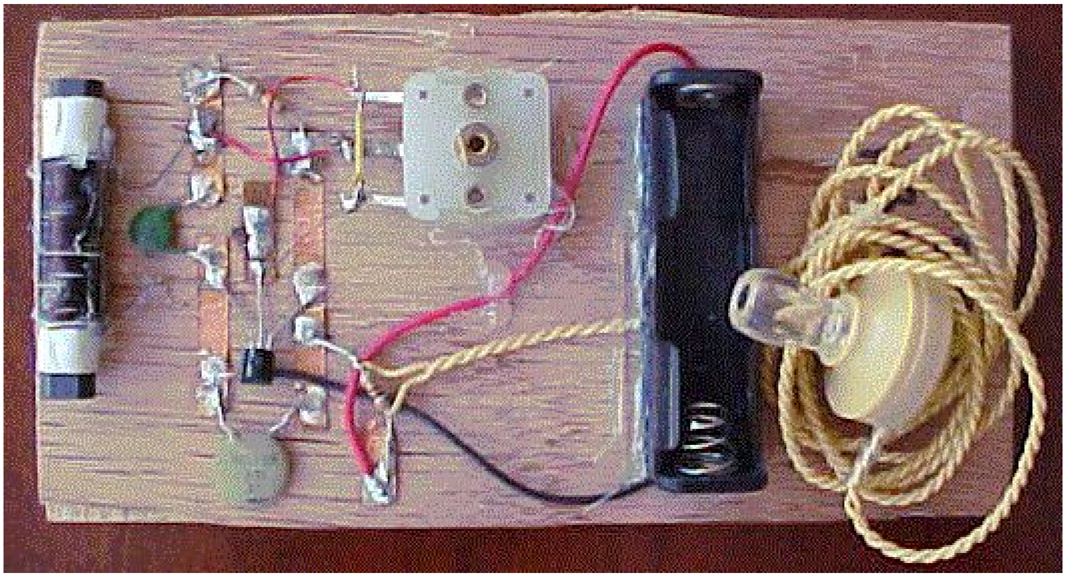

Suppliers of the MK 484:
Kanga , a UK supplier will sell you the chip for a pound note plus S&H.
Dick Smith Electronics has it listed at $2.29 AUS; get ten, they're cheap, and get your cub scout pack involved, or have a building contest at your next ham club meeting.
Ocean State Electronics has a pricey ZN414 listed.

If you run across any 414s, you can build the set here, just remembering that the pinout for the 414 in the TO92 case is reversed from the 484 above, center is still input. For the 414 in the TO39 case, the output is next to the metal tab, input is in the center, same as the base for a bipolar xistor, etc.

Once you have this one running, modify it slightly for more flexibility and dx by going HERE .

HOME LINKS
CONSTRUCTION TIPS ANTENNAS
RADIO SHACK PROJECT PROJECT RADIO

Mounting Your Own Crystals

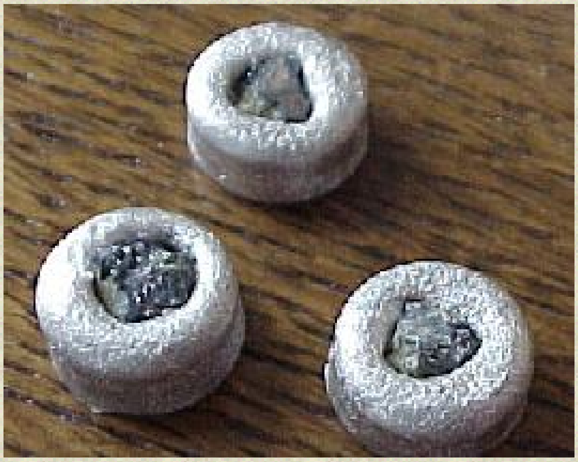

These instructions will show you how to make nicely mounted crystals just like the ones in the old crystal radios. Materials needed are:
-Minerals to mount, such as galena or iron pyrite
-Solder or Woods metal
-Mold for the mounts
-Soldering iron

You can make the mold by drilling a 7/16 inch hold in a piece of any material that will withstand the head of the molten solder. Some mounts are 1/2 inch. I use 1/4 inch thick acrylic and Woods metal that melts at 158 degrees F. I drill the 7/16 inch hole and then smooth and slightly bevel the inside edges of hole with sandpaper so the crystal will push out the bottom smoothly.
Although some users don't agree, I think the high temperatures necessary to melt solder diminish the sensitivity of the crystal.
I use another piece of acrylic underneath as a base plate clamped together so that the 2 pieces can be separated and the crystal pushed out the bottom of the mold when cool.

Break off a piece of mineral the size you want. I use wire cutters. It is best to test the crystal and find the best surface to leave facing upwards.

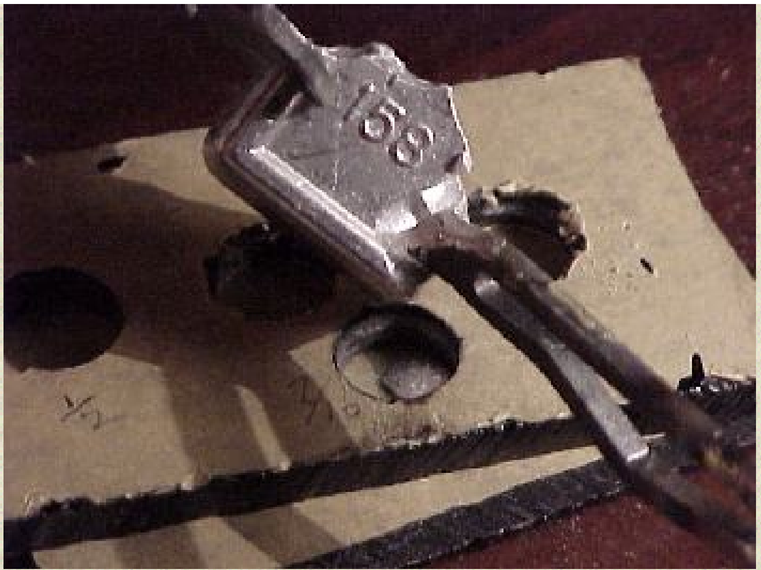

Next, melt some of your mounting metal in the hole about 2/3 full. Remember, the crystal will take up some room so don't fill it up all the way or it will uverflow when you put the rock in.

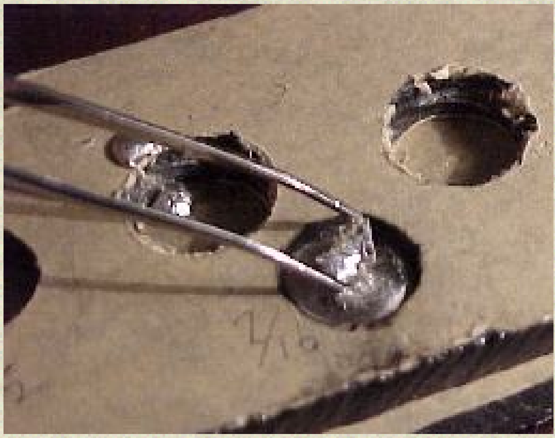

Then while the metal is still liquid, drop the rock on to the surface, keeping the best surface pointing upwards.

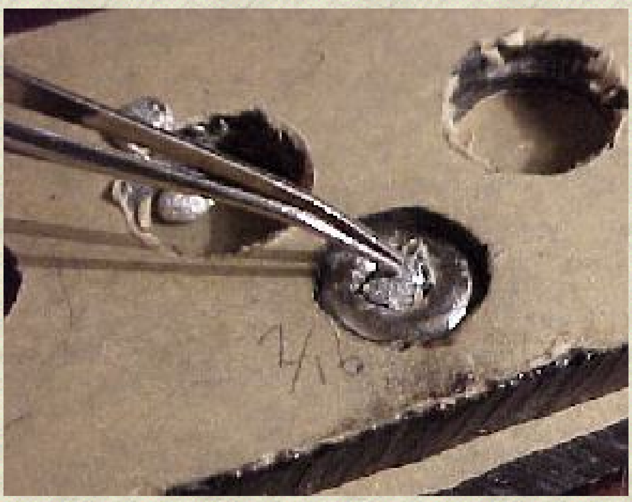

The rock will tend to float up so you have to hold it down while the metal cools and hardens.
When it is cool, separate the mold from the base plate and push the mounted crystal out the bottom.
Be careful, it will be HOT at first, epsecially if you use solder.

Email to Scottswim@aol.com

Back to Scott's Crystal Radios

US006882128B1

(12) **United States Patent**
Rahmel et al.

(10) **Patent No.: US 6,882,128 B1**
(45) **Date of Patent: Apr. 19, 2005**

(54) **METHOD AND SYSTEM FOR ENERGY RECLAMATION AND REUSE**

(75) Inventors: **Ronald S. Rahmel**, Encinitas, CA (US); **Gerald C. Gerace**, Chula Vista, CA (US)

(73) Assignee: **Science Applications International Corporation**, San Diego, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 776 days.

(21) Appl. No.: **09/670,613**

(22) Filed: **Sep. 27, 2000**

(51) **Int. Cl.**^{7} **H04B 1/38**; H04M 1/00

(52) **U.S. Cl.** **320/101**; 455/343.1; 455/572; 320/101; 320/124; 320/127

(58) **Field of Search** 455/343.1, 343.2, 455/343.3, 572, 573, 343.5; 320/101, 102, 124, 127, 137, 138; 343/702, 720

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,434,678	A		3/1969	Brown et al.	244/1
3,852,755	A		12/1974	Works et al.	343/701
3,927,375	A		12/1975	Lanoe et al.	325/363
4,031,449	A		6/1977	Trombly	320/2
4,079,268	A		3/1978	Fletcher et al.	307/151
4,360,741	A	*	11/1982	Fitzsimmons et al.	307/151
4,628,299	A		12/1986	Tate et al.	340/540
4,685,047	A	*	8/1987	Phillips, Sr.	363/126
5,043,739	A		8/1991	Logan et al.	343/701
5,416,486	A		5/1995	Koert et al.	342/42
5,440,300	A		8/1995	Spillman, Jr.	340/825.54
5,495,519	A		2/1996	Chen	379/58
5,710,988	A		1/1998	Suzuki et al.	455/90
5,842,838	A		12/1998	Berg	417/331
5,889,383	A		3/1999	Teich	320/107
5,898,932	A		4/1999	Zurlo et al.	455/573
5,970,393	A		10/1999	Khorrami et al.	455/66
6,046,676	A	*	4/2000	Ward et al.	340/572.1
6,154,664	A	*	11/2000	Chorey et al.	455/571
6,184,651	B1		2/2001	Fernandez et al.	320/108
6,198,250	B1	*	3/2001	Gartstein et al.	320/112
6,275,010	B1		8/2001	Neuteboom	320/137
6,289,237	B1	*	9/2001	Mickle et al.	600/509
6,369,759	B1	*	4/2002	Epp et al.	343/700 MS
6,426,917	B1	*	7/2002	Tabanou et al.	367/82
6,597,465	B1	*	7/2003	Jarchow et al.	358/1.12
6,609,013	B1	*	8/2003	Oh et al.	455/561

OTHER PUBLICATIONS

Noreen et al., "The system view," Abstract, *The Deep Space Network: A Radio Communication Instrumentation for Deep Space Exploration,* p. 25, 1983. (Deep Space Network).

Naumenko et al., "Trajectories of relativistic particles moving in free space," Abstract, *Radiotekhnika I Electronika,* vol. 29, No. 9, pp. 144–145, 1984. ("Relativistic Particles").

Yegerov, "The efficiency of energy transmission by an electromagnetic beam and its conversion in a rectenna," translated in *Soviet Journal of Communications Technology & Electronics,* vol. 30, No. 8, pp. 35–41, 1985. ("Energy Transmission").

Bianco, "Measurement Techniques for Evaluating Power System Interference to Instrument Landing Systems (ILS)," Abstract, *Canadian Electrical Association Contract/Grant Nos. CEA 100T219 and CEA 100T219A,* 1985. ("ILS").

(Continued)

Primary Examiner—Simon Nguyen
(74) *Attorney, Agent, or Firm*—Kilpatrick Stockton LLP

(57) **ABSTRACT**

The present invention relates to a system and method for harvesting ambient electromagnetic energy, and more particularly, to the integration of antennas and electronics for harvesting ubiquitous radio frequency (RF) energy, transforming such electromagnetic energy into electrical power, and storing such power for usage with a wide range of electrical/electronic circuits and modules.

3 Claims, 13 Drawing Sheets

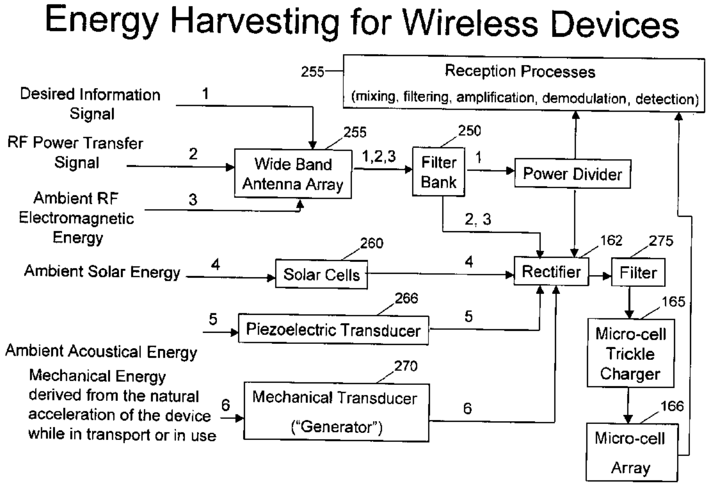
Energy Harvesting for Wireless Devices
255
Reception Processes
(mixing, filtering, amplification, demodulation, detection)
Desired Information Signal
1
RF Power Transfer Signal
2
Ambient RF Electromagnetic Energy
3
255
Wide Band Antenna Array
1,2,3
250
Filter Bank
1
Power Divider
2, 3
Ambient Solar Energy
4
260
Solar Cells
4
162
Rectifier
275
Filter
165
5
266
Piezoelectric Transducer
5
Micro-cell Trickle Charger
Ambient Acoustical Energy
Mechanical Energy derived from the natural acceleration of the device while in transport or in use
6
270
Mechanical Transducer ("Generator")
6
166
Micro-cell Array

OTHER PUBLICATIONS

Machina et al., "Power beaming—Energy transmission at 35 GHz and higher frequencies (for terrestrial and space applications with rectennas)," Abstract, *International Communication Satellite Systems Conference and Exhibit, 14th, Technical Papers, Part 3*, pp. 1671–1674, 1992. (Power Beaming).

Waldron, "Rectenna options for a lunar–solar power system," Abstract, 26th *International Conference on Environmental Systems,* 1996. ("Rectenna Options").

Keerthi, "Analysis of the multistage CM array for digital communication signals," Abstract, *Conference Record of the Thirtieth Asilomar Conference on Signals, Systems and Computers,* vol. 2, No. 23, p. 1416, 1997. ("Asilomar").

Takeshi et al., "Experimental study of microwave power transmission efficiency," Abstract, *IEIC Technical Report, Institute of Electronics, Information and Communication Engineers,* vol. 98, No. 474, pp. 21–24, 1998. ("IEIC").

Vasudev, "Telecom explosion drives RF ICs," *Electronic Engineering Times, No. 1016*, p. 81, 1998. ("RF ICs").

Takyuki et al., "Effective utilization of electromagnetic energy. Effective utilization of electromagnetic energy in ITS (4). Radiocommunication in turnpike automatic charge collection system," Abstract, *Denji Kankyo Kogaku Joho EMC,* vol. 11, No. 9, pp. 38–45, 1999. ("ITS").

Ohr, "Integrated, one–chip cell phone is still a stretch, panelists concur," *Electronic Engineering Times,* No. 1067, p. 64, 1999. ("One–chip cell phone").

Davies, "Operation of Information Satellites in an interference environment," Abstract, *Microwave Journal,* vol. 15, pp. 33–36, 1972. ("Microwave Journal").

Dickinson, "Satellite Power System (SPS) Microwave Subsystem Impacts and Benefits," Abstract, *Report No. NASA–CR–157951,* 1977. ("Dickinson SPS").

Tonelli et al., "Design and analysis of a 5000–MW GaA1As satellite power system," Abstract, *Proceedings of the 12th Intersociety Energy Conversion Engineering Conference Part II,* pp. 1412–1420, 1977. ("IECE 1").

Tonelli et al., "Design and evaluation of a 5 GW GaA1As solar power satellite (SPS)," Abstract, *Proceedings of the 13th Intersociety Energy Conversion Engineering Conference Part I,* pp. 156–161, 1978. ("IECE 2").

Hughes, "The interactions of a solar power satellite transmission with the ionosphere and troposphere," Abstract, *AGARD Conference Proceedings No. 284. Propagation Effects in Space/Earth Paths,* p. 38, 1980. ("AGARD").

Maynard, "Initial MPTS study results: Design considerations and issues," Abstract, *Johnson Space Center Solar Power Satellite Microwave Transmission and Reception,* pp. 28–32, 1980. ("MPTS").

* cited by examiner

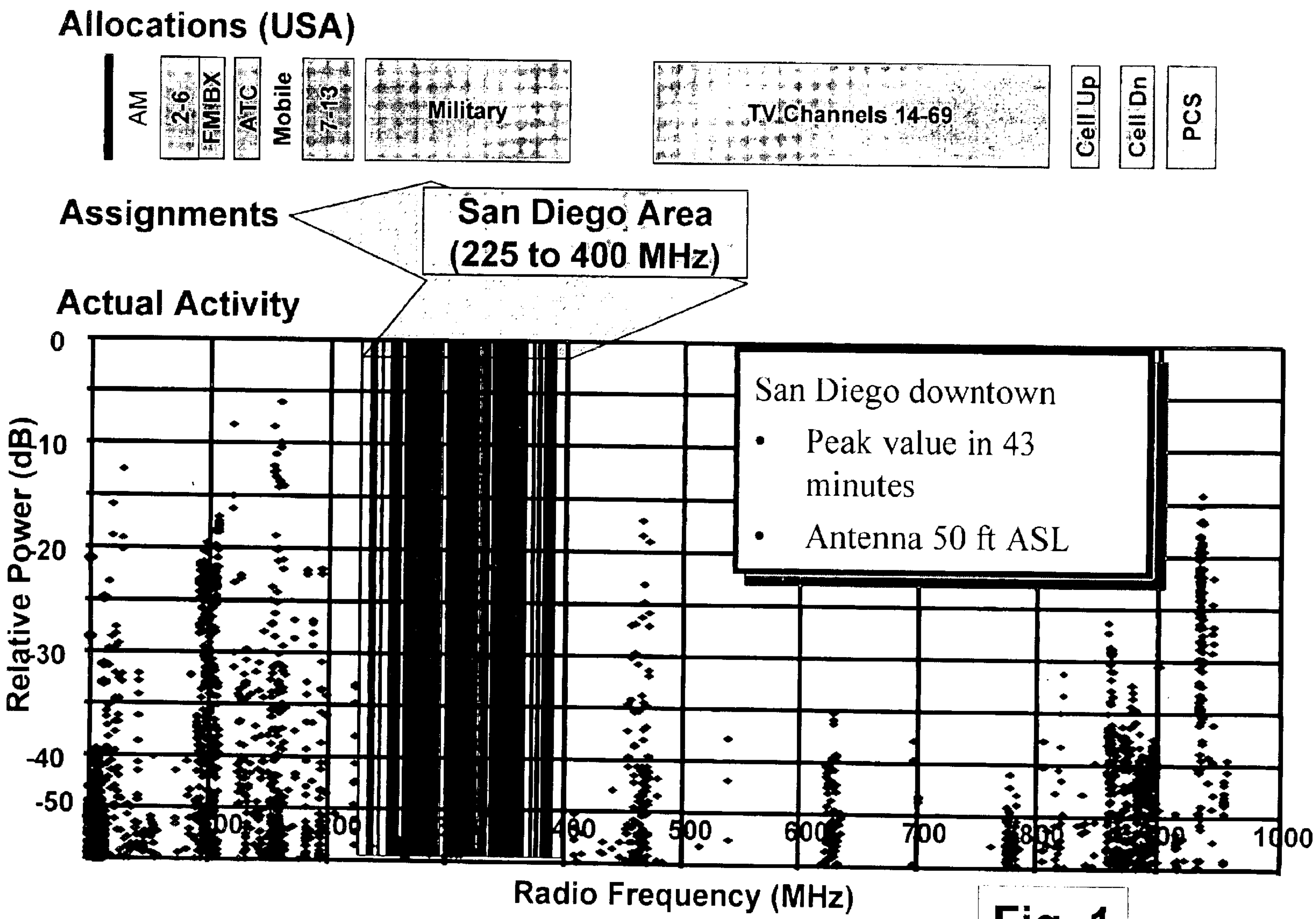
Allocations (USA)
AM
2-6
FM BX
ATC
Mobile
7-13
Military
TV Channels 14-69
Cell Up
Cell Dn
PCS
Assignments
San Diego Area
(225 to 400 MHz)
Actual Activity
0
-10
-20
-30
-40
-50
Relative Power (dB)
100
200
300
400
500
600
700
800
900
1000
Radio Frequency (MHz)
San Diego downtown
Peak value in 43 minutes
Antenna 50 ft ASL

Fig. 1

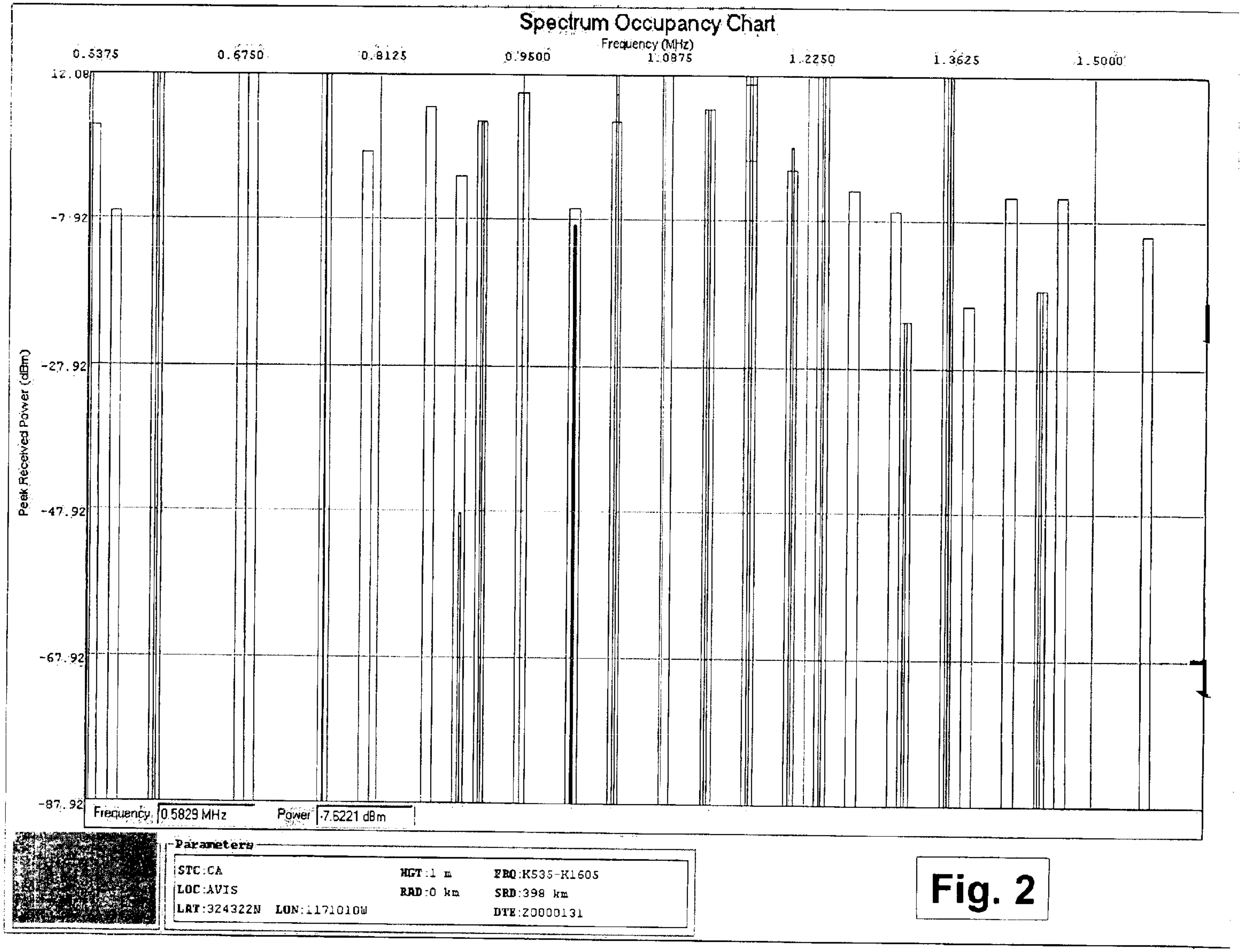
Spectrum Occupancy Chart
Frequency (MHz)
0.5375
0.6750
0.8125
0.9500
1.0875
1.2250
1.3625
1.5000
Peak Received Power (dBm)
12.08
-7.92
-27.92
-47.92
-67.92
-87.92

Frequency 0.5829 MHz Power -7.6221 dBm

Parameters

STC:CA HGT:1 m FRQ:K535-K1605

LOC:AVIS RAD:0 km SRD:398 km

LAT:324322N LON:1171010W DTE:20000131

Fig. 2

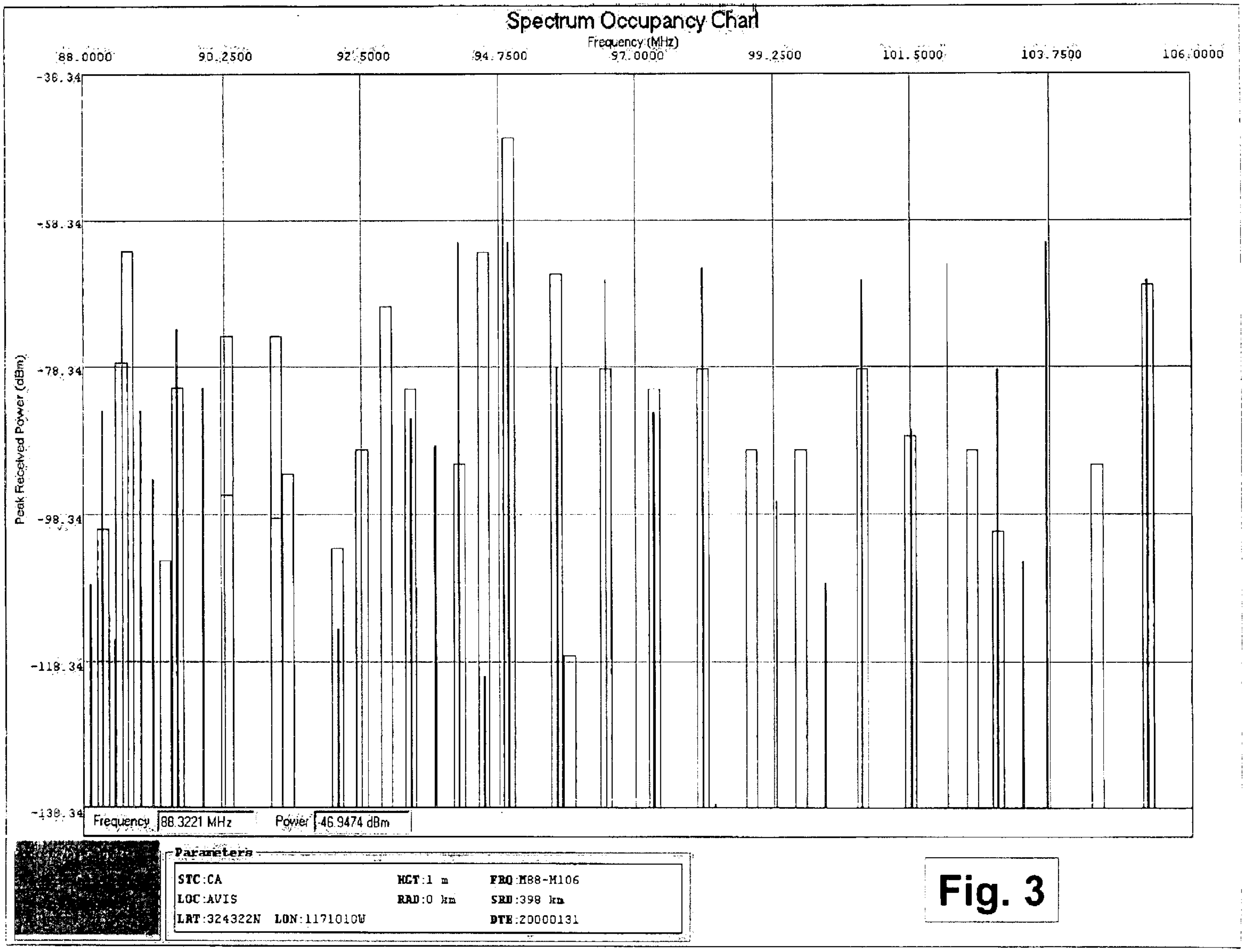

Frequency 88.3221 MHz Power -46.9474 dBm

Parameters

STC:CA HGT:1 m FRQ:M88-M106
LOC:AVIS RAD:0 km SRD:398 km
LAT:324322N LON:1171010W DTE:20000131

Fig. 3

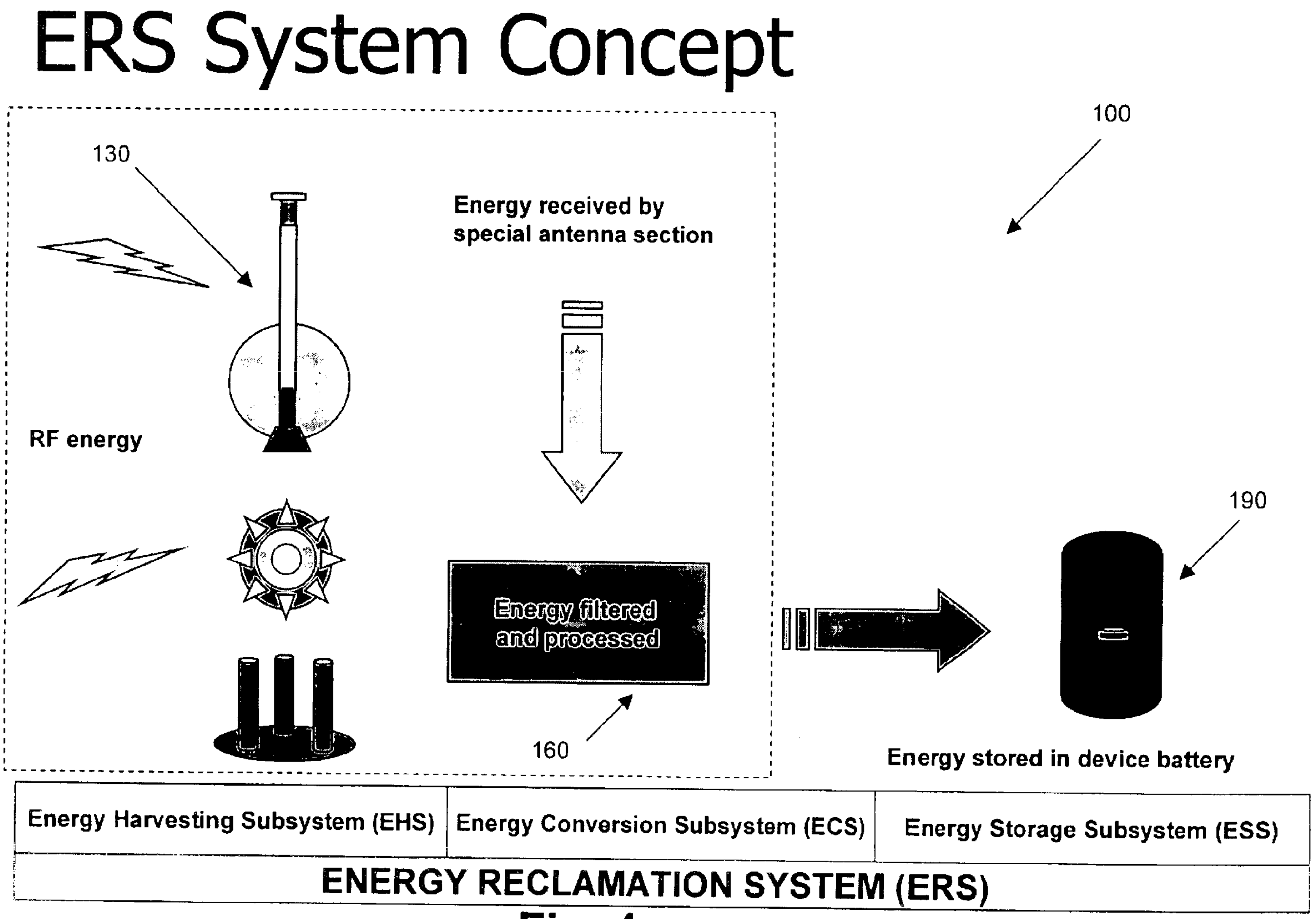
ERS System Concept
100
130
RF energy
Energy received by special antenna section
Energy filtered and processed
160
190
Energy stored in device battery
Energy Harvesting Subsystem (EHS)
Energy Conversion Subsystem (ECS)
Energy Storage Subsystem (ESS)
ENERGY RECLAMATION SYSTEM (ERS)

Fig. 4

Antenna Design

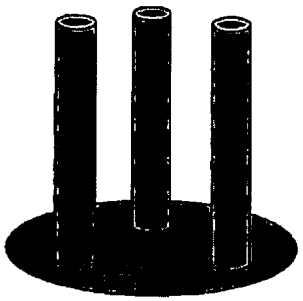

Fig. 5A

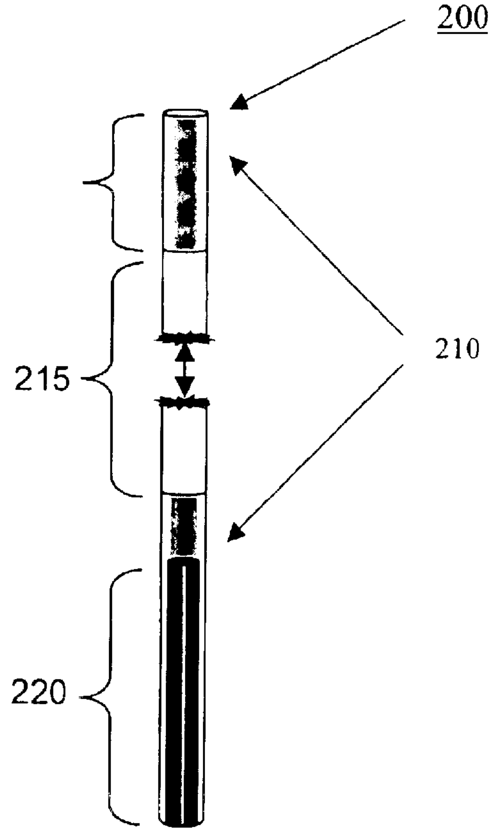
200
210
215
220

Fig. 5B

Alternate Antenna Design

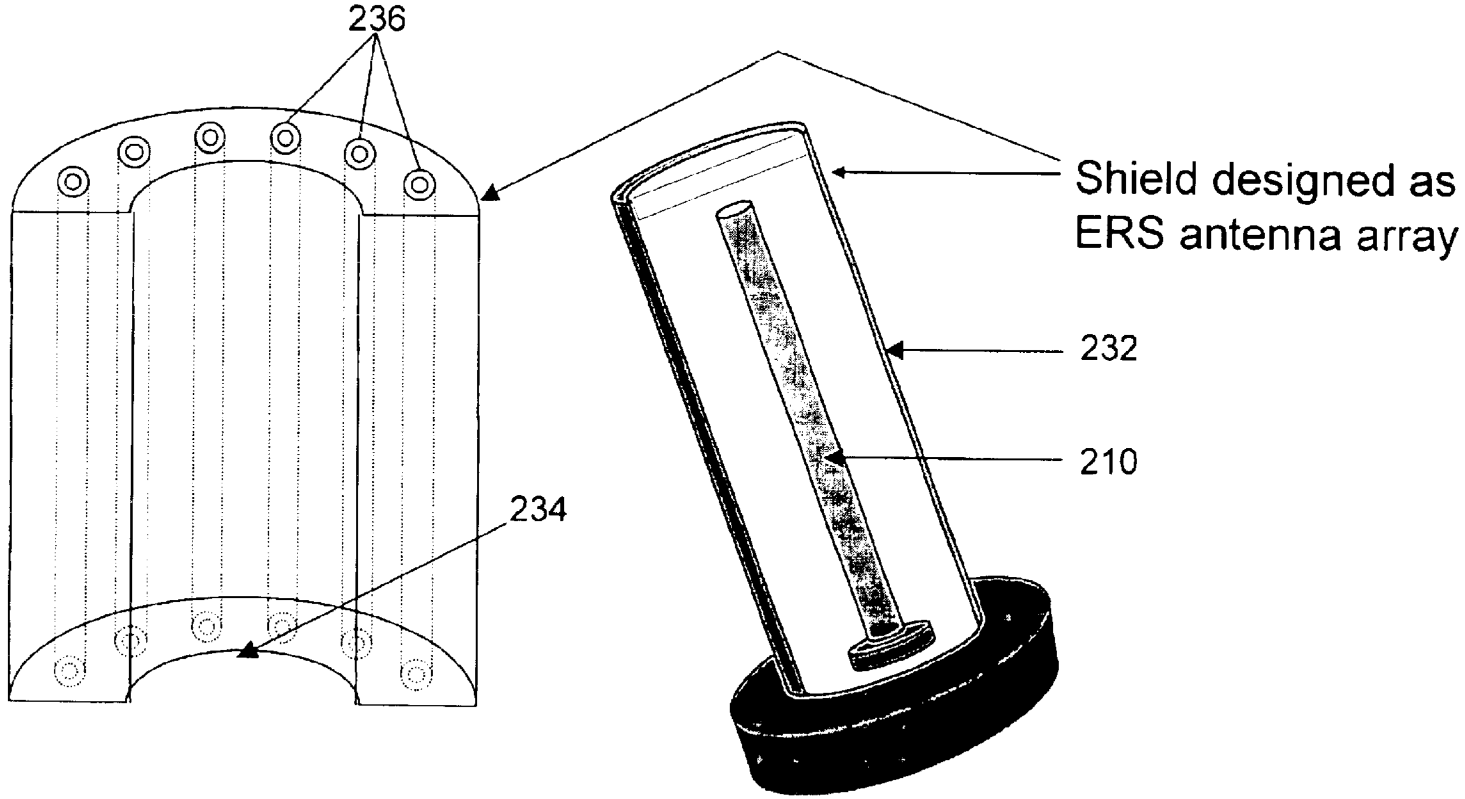
236
Shield designed as
ERS antenna array
232
210
234

Fig. 6

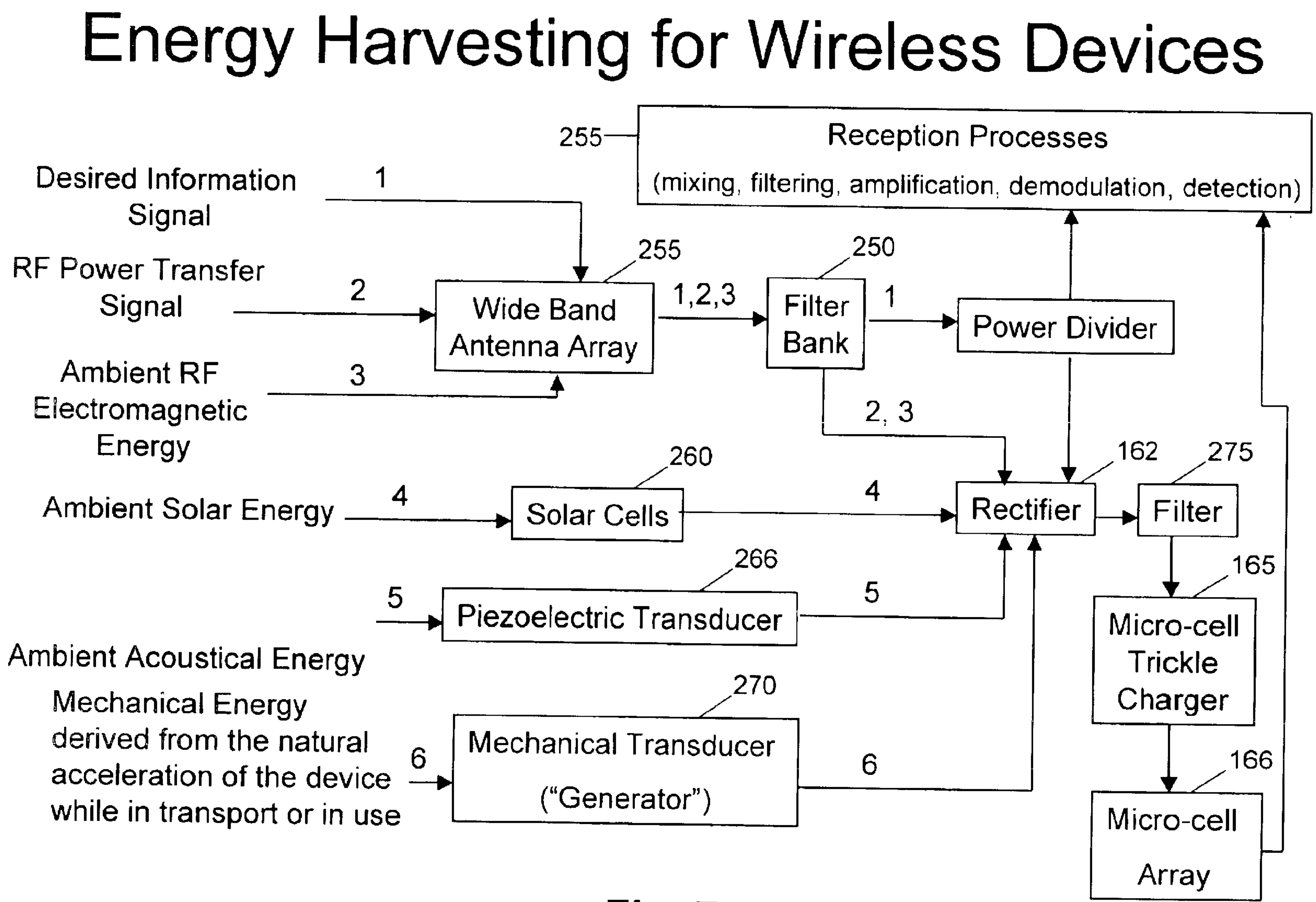
Energy Harvesting for Wireless Devices
255
Reception Processes
(mixing, filtering, amplification, demodulation, detection)
Desired Information Signal
1
RF Power Transfer Signal
2
Ambient RF Electromagnetic Energy
3
255
Wide Band Antenna Array
1,2,3
250
Filter Bank
1
Power Divider
2, 3
162
275
Rectifier
Filter
Ambient Solar Energy
4
260
Solar Cells
4
266
5
Piezoelectric Transducer
5
165
Micro-cell Trickle Charger
Ambient Acoustical Energy
Mechanical Energy derived from the natural acceleration of the device while in transport or in use
6
270
Mechanical Transducer
("Generator")
6
166
Micro-cell Array

Fig. 7

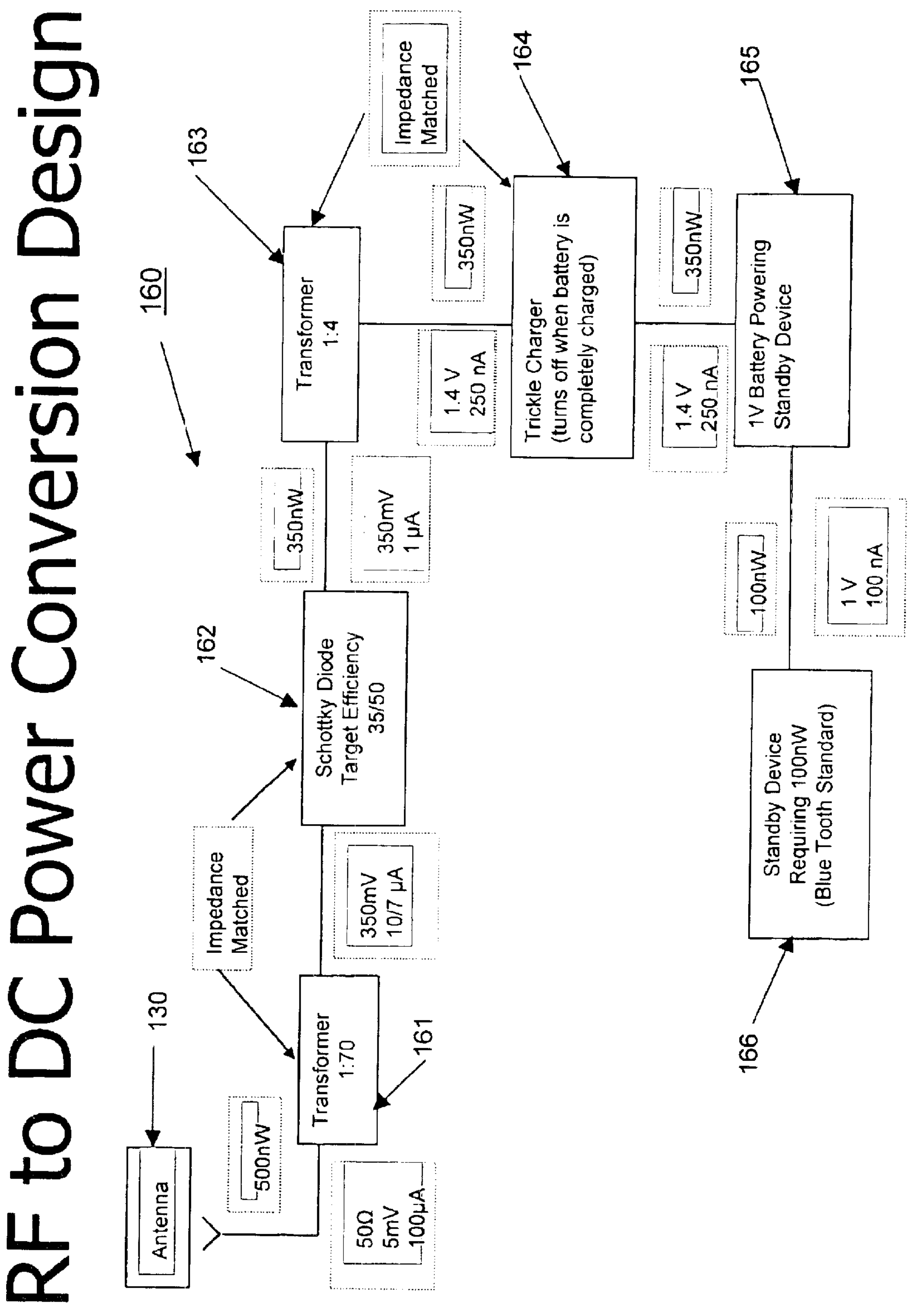
RF to DC Power Conversion Design
160
130
Antenna
500nW
50Ω
5mV
100μA
Impedance Matched
Transformer 1:70
161
350mV
10/7 μA
Schottky Diode
Target Efficiency
35/50
162
350nW
350mV
1 μA
Transformer 1:4
163
Impedance Matched
350nW
1.4 V
250 nA
Trickle Charger
(turns off when battery is
completely charged)
164
350nW
1.4 V
250 nA
1V Battery Powering
Standby Device
165
100nW
1 V
100 nA
Standby Device
Requiring 100nW
(Blue Tooth Standard)
166

Fig. 8

RF to DC Power Conversion/Trickle Charge/Power Storage Design Concept

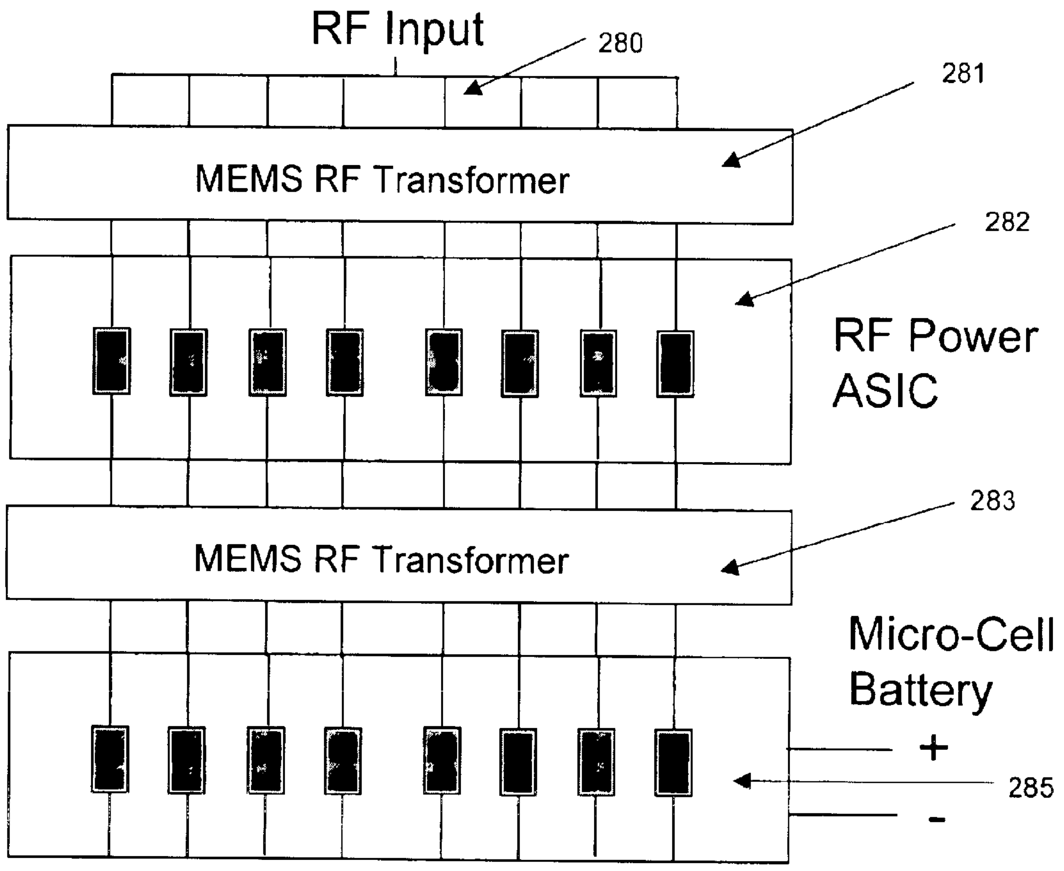
RF Input
280
281
MEMS RF Transformer
282
RF Power
ASIC
283
MEMS RF Transformer
Micro-Cell
Battery
+
285
-

Fig. 9

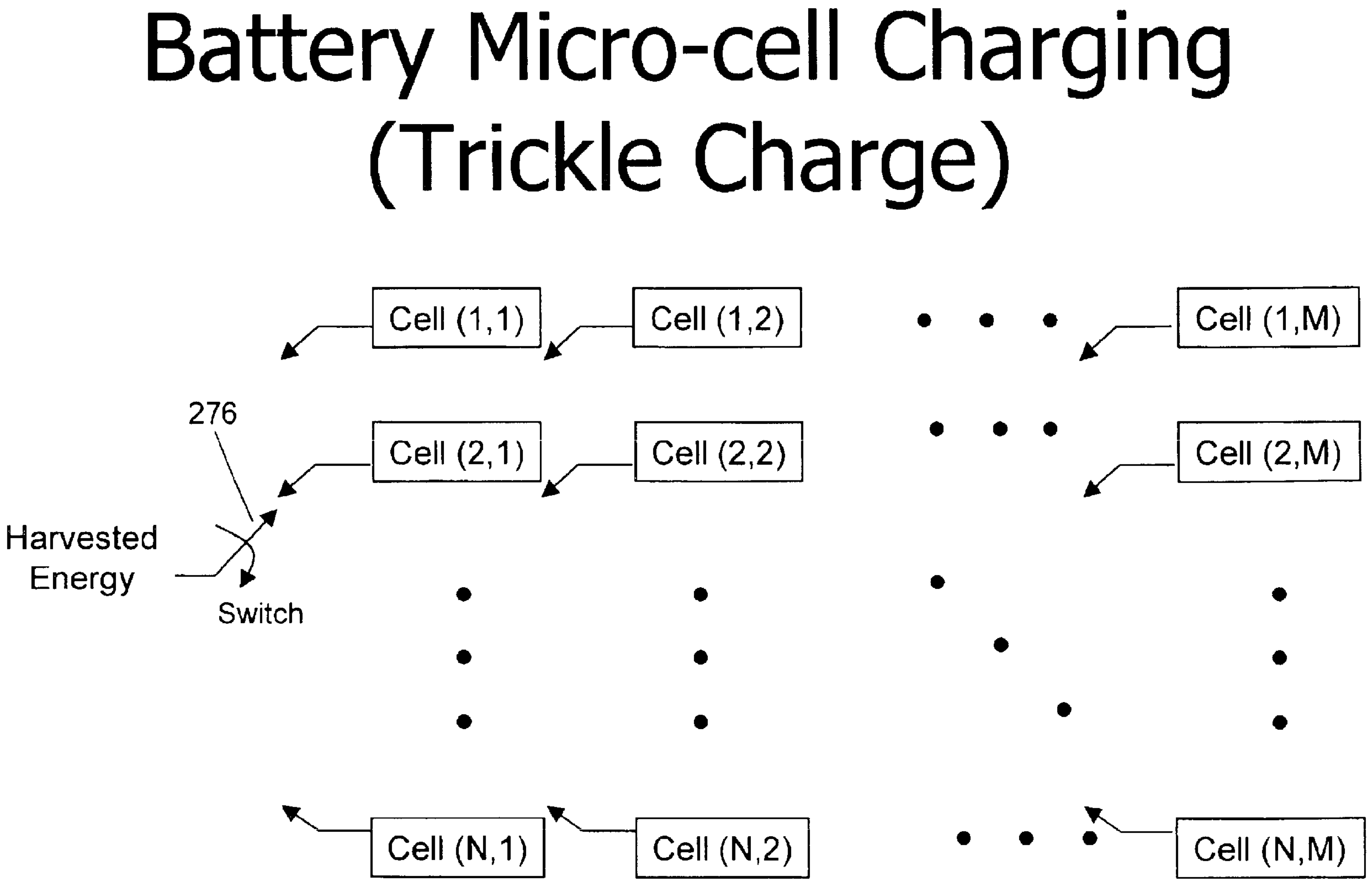
Battery Micro-cell Charging
(Trickle Charge)
Cell (1,1)
Cell (1,2)
Cell (1,M)
276
Cell (2,1)
Cell (2,2)
Cell (2,M)
Harvested Energy
Switch
Cell (N,1)
Cell (N,2)
Cell (N,M)

Fig. 10

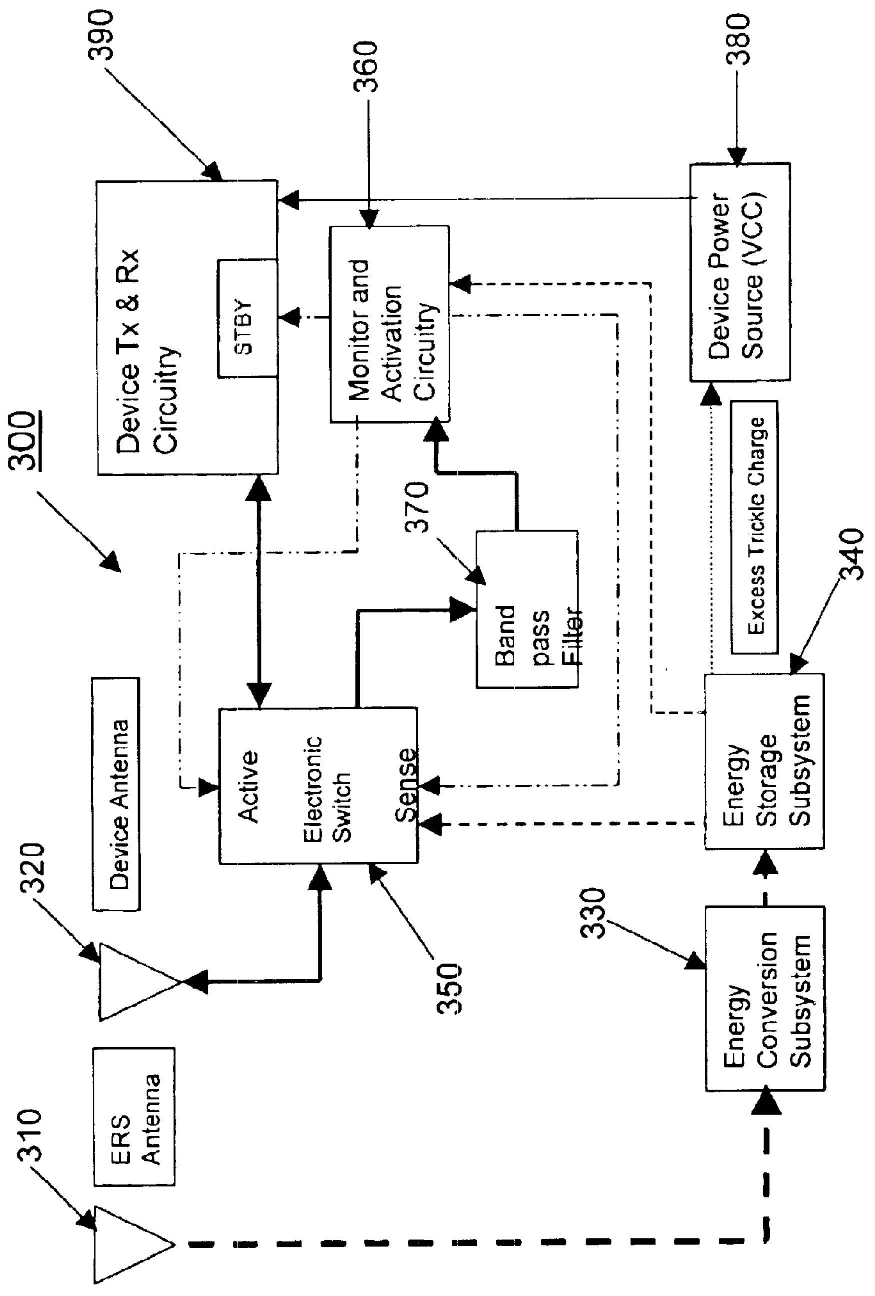
300
390
360
380
Device Tx & Rx Circuitry
STBY
Monitor and Activation Circuitry
Device Power Source (VCC)
370
Band pass Filter
Excess Trickle Charge
340
Device Antenna
Active
Electronic Switch
Sense
Energy Storage Subsystem
320
350
330
ERS Antenna
Energy Conversion Subsystem
310

Two or Multiple Array Antenna Sense and Active Mode

Fig. 11

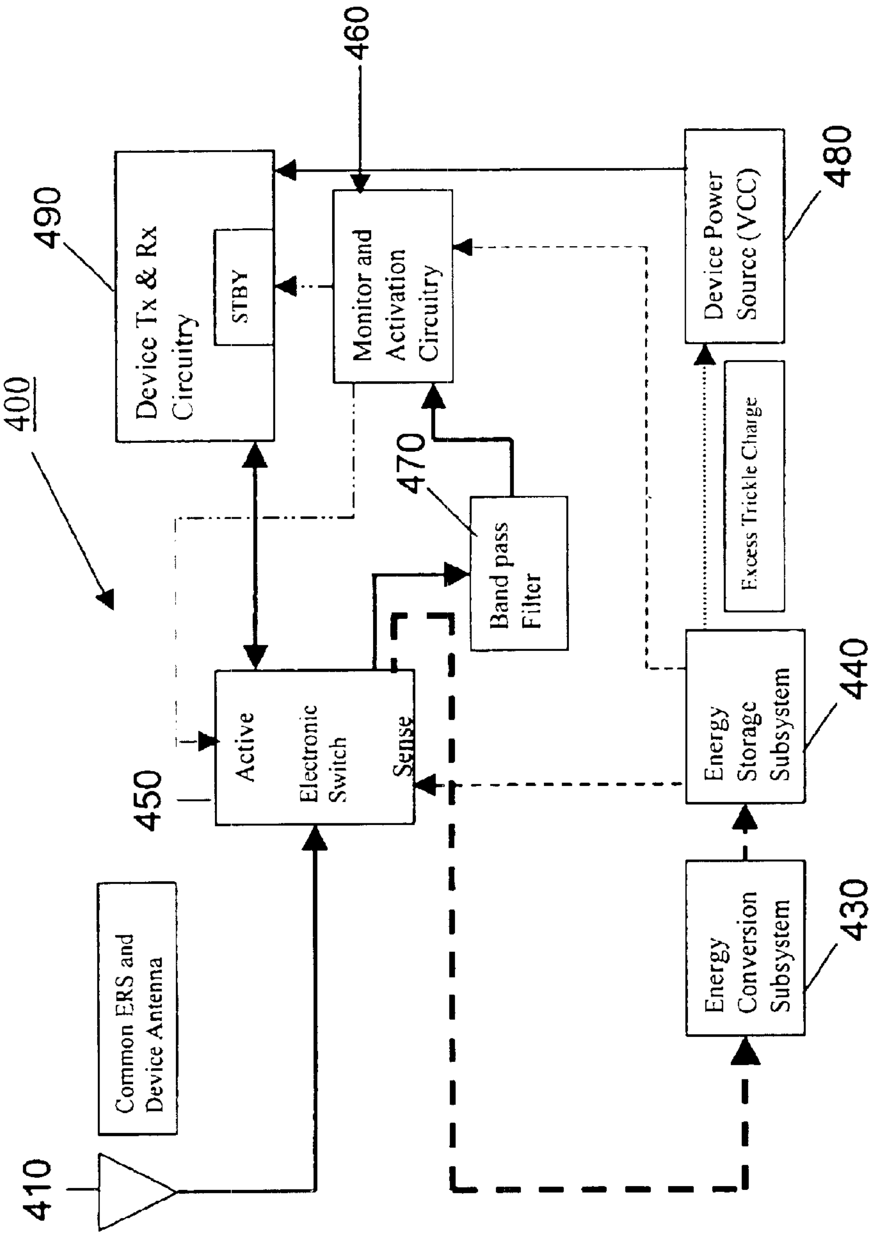
400
410
Common ERS and Device Antenna
450
Active
Electronic Switch
Sense
490
Device Tx & Rx Circuitry
STBY
460
Monitor and Activation Circuitry
470
Band pass Filter
480
Device Power Source (VCC)
Excess Trickle Charge
440
Energy Storage Subsystem
430
Energy Conversion Subsystem

Common Antenna used for Simultaneous Energy Harvesting, Sense and Active Mode

Fig. 12

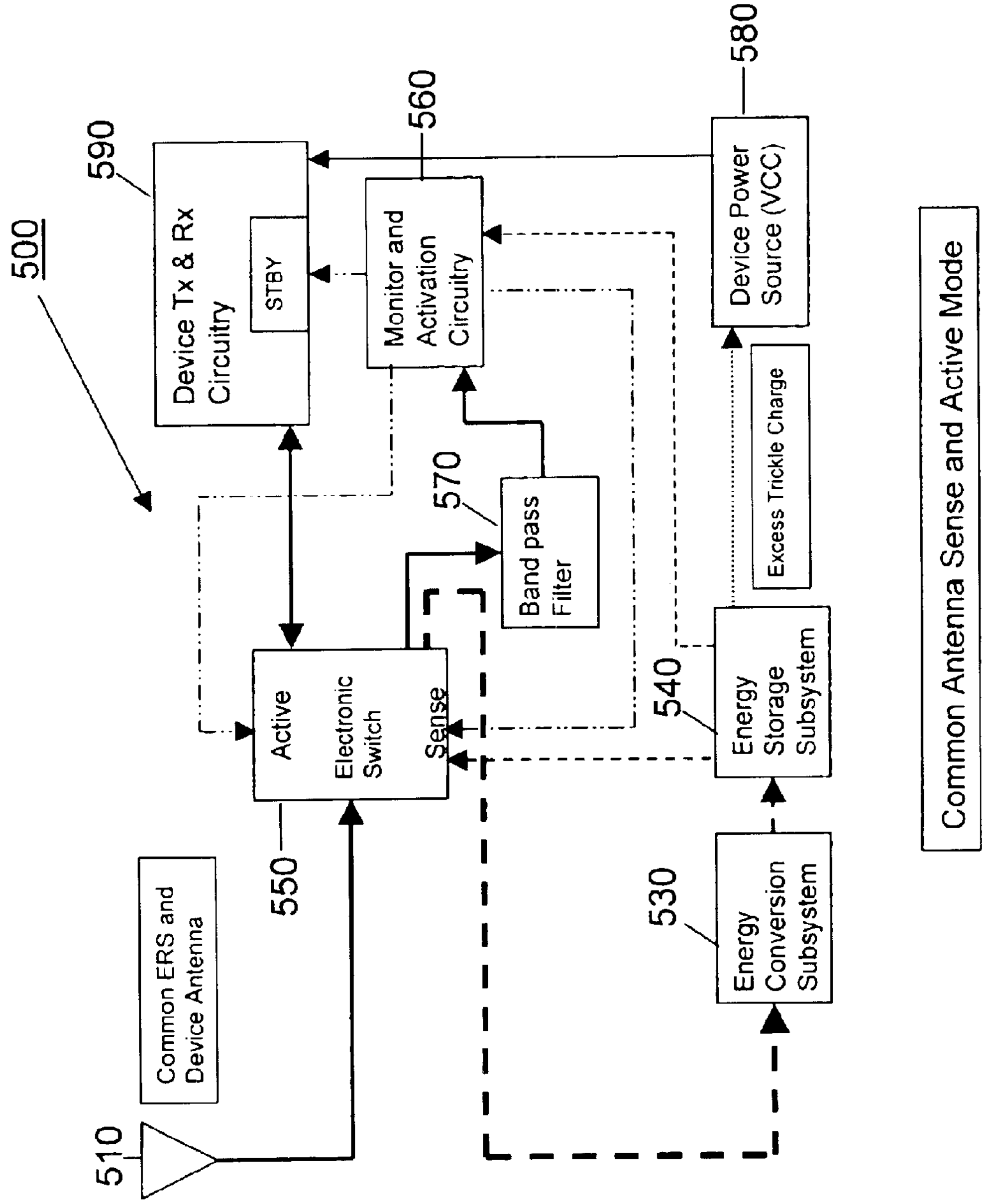
500
590
Device Tx & Rx Circuitry
STBY
560
Monitor and Activation Circuitry
580
Device Power Source (VCC)
570
Band pass Filter
Excess Trickle Charge
540
Energy Storage Subsystem
550
Active
Electronic Switch
Sense
530
Energy Conversion Subsystem
510
Common ERS and Device Antenna
Common Antenna Sense and Active Mode

Fig. 13

METHOD AND SYSTEM FOR ENERGY RECLAMATION AND REUSE

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates generally to wireless energy harvesting methods and systems. Specifically, the preferred embodiments of the present invention relate to the integration of antennas and electronics for harvesting ubiquitous electromagnetic energy, transforming that energy into useful power, and storing such power for usage.

2. Description of the Related Art

Typically, energy harvesting techniques and systems are focused on renewable energy such as solar energy, wind energy, and wave action energy from moving water. Solar energy is harvested by arrays of solar cells that convert radiant energy to DC power. Such energy is limited in low-light conditions. Cumbersome wind turbines that convert wind energy to electrical power depend on sufficient wind to spin the blades of the turbines to generate power. Wave action energy is harvested by water turbines or other apparatuses that are capable of converting the energy in the moving water into electrical power. This energy conversion is typically performed at complex hydroelectric power plants or by mechanical/electrical conversion apparatuses, such as those shown in U.S. Pat. Nos. 5,842,838 and 6,045,339 to Berg.

Another approach to energy harvesting is wireless power transmission via focused microwave signals. Unlike the above renewable energy technologies which harvest energy from natural sources, electrical energy of the wireless power transmission technology is remotely harvested and converted from focused microwave signals that are transmitted by a constructed source. For example, U.S. Pat. No. 5,043,739 to Logan et al. discloses a high frequency rectenna device for rectifying electromagnetic energy at microwave frequencies and higher and converting the energy into direct-current (DC) power to conduct wireless power transmission. Applications of this disclosure are focused on high-energy microwave radiation and high temperatures. One such application includes sending power to satellites or high altitude devices from the earth by electromagnetic waves. Another application involves collecting solar power from large space-based arrays of solar cells and transmitting the energy via electromagnetic waves to earth. A third application involves the direct conversion of microwaves generated by fusion reactions into a direct current.

BRIEF SUMMARY OF THE INVENTION

Currently, there is a large amount of renewable electromagnetic energy available in the radio frequency (RF) bands. For instance, AM radio, FM radio, TV, very high frequency (VHF), ultra high frequency (UHF), global system for mobile communications (GSM), digital cellular systems (DCS) and especially the personal communication system (PCS) bands are all lucrative energy sources. These untapped sources of energy grow more and more abundant as a result of the rapid growth in the wireless communication business. Hence, there is a desire to seek new methods and means of collecting the ubiquitous electromagnetic energy (ambient RF noise and signals) opportunistically present in the environment and transforming that energy into useful electrical power.

The preferred embodiments of the present invention can collect ubiquitous ambient radio frequency (RF) electro magnetic energy, convert it to a form useful for powering electronic devices, and store it for reuse. Thus, perpetual self contained subsystems or modules can be realized by using the reclaimed RF energy to power active antenna components, low power circuits, modules and other components such as transceivers, amplifiers, filters, etc. Hence, the preferred embodiments of the present invention can be applicable for any antenna or antenna array, for any antenna physical design (e.g., dipole, monopole, parabolic shape or dish, Yagi, etc.) and any frequency band. In some applications, this technology could be used advantageously and conveniently to transfer power remotely, without a physical connection, using a RF energy source intended for that purpose. In other applications, this technology could be used to maintain devices in a standby sleep mode indefinitely without ancillary power source. Then an RF signal could be used as a triggering mechanism to activate the device. Finally, for low power devices, this technology could provide primary power to the device and eliminate or reduce the traditional battery charge cycle.

Accordingly, the preferred embodiments of the present invention provide an integration of classical antenna physical materials and construction concepts with semiconductor and microprocessor technology to create an antenna structure capable of transforming electromagnetic waves into DC power. This antenna method and system collects and processes RF energy for reuse while simultaneously performing its primary function, such as signal reception, for a specific application. Furthermore, by using either a single antenna element or an array of elements configured to receive a broad range of frequencies in the present invention, energy collection and antenna transmit/receive performance can be maximized and optimized at the same time. The preferred embodiments of the present invention also provide an Energy Reclamation System (ERS) that collects, converts and stores energy for use from selected frequency bands over the entire electromagnetic spectrum. These selected frequency bands are not required to be contiguous and may based on the operational environment, the application, any physical size limitations, and the technology available to cost effectively integrate an ERS into a device or application. An ERS of the present invention is not restricted to a single energy source because technologies like micro electromechanical systems (MEMS) can be used to integrate the collection of other types of energy, such as electromagnetic energy, solar energy, acoustic energy and/or casual motion energy. Thus, energy diversity enhances an ERS' ability to provide power consistent with the demands of modern devices and applications.

The preferred embodiments of the present invention also provide a method and system for harvesting electromagnetic spectrum energy from unintentional power sources. Thus, the electromagnetic energy may be harvested opportunistically from the abundant ambient RF electromagnetic environment created by the emission of countless commercial RF sources typically present today.

The preferred embodiments of the present invention also provide a method and system for harvesting electromagnetic spectrum energy using efficient wideband omnidirectional antennas for maximizing the collection of RF energy. The antennas can be of any design such as monopole, dipole, Yagi, or contrawound.

The preferred embodiments of the present invention also provide an energy collection and conversion technology that may be used to transmit a control signal to select a device's operating mode (e.g., wake up call in the standby sleep mode). Incorporating this technology into some devices allows them to remain in the sleep mode indefinitely. Thus, the devices may use the continuously reclaimed energy to power the monitoring circuitry and initialize equipment activation as required.

The preferred embodiments of the present invention also provide a method and system for transferring low power levels from an intentional power source advantageously and conveniently. This may be done remotely without a physical connection by using a RF energy source with a focused or blanket radiation pattern and an ERS at the target device. This may be used for numerous applications, including autonomous sensor or communication type applications in any environment (e.g., ground, air, underwater) where the device may use the ERS as its power source or backup power source. For example, for underwater sensor locations, energy may come from sonar or other types of energy sources that can propagate with tolerable losses through the water and/or from RF energy delivered to a remote antenna extended above the surface of the water. If remaining operational for extended periods of time requires additional power, the device can be wirelessly and remotely supplied with power for storage and future use.

The preferred embodiments of the present invention further provide a conversion of the energy collected or captured by the antenna or transducer into an electrical current. The electrical current is then converted into a form suitable for input to a selected storage device. For instance, a rectifier and a low pass filter can be used to convert the electrical current into direct current (DC). This power can then be fed into a trickle charger for charging a power storage device, such as batteries, in wireless and other electrical devices. The type of electrical filter and power conditioning used depends on the type of battery or power storage device and the particular process of trickle charging used.

The preferred embodiments of the present invention additionally provide a method and system for harvesting electromagnetic energy that is designed to efficiently collect and convert RF energy using application specific integrated circuit (ASIC) technology. Hence, supporting electronics are integrated with the antenna itself and include, for example, amplifiers, mini-transceivers, trickle chargers for charging micro-cell batteries, filters efficient rectifier design and implementation (with different types of diodes), micro electromechanical systems (MEMS) RF transformers, etc. Thus, embedded circuitries are built into the antenna elements and optimized for a frequency range compatible with any given antenna design.

Additional aspects and novel features of the invention will be set forth in part in the description that follows, and in part will become more apparent to those skilled in the art upon examination of the disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

The preferred embodiments are illustrated by way of example and not limited in the following figures, in which:

FIG. **1** shows data taken from measurements in the RF spectrum between 3 MHz and 1 GHz in downtown San Diego, Calif., USA;

FIG. **2** shows the results of calculations predicting the signal levels in the AM radio band in downtown San Diego, Calif., USA;

FIG. **3** shows the results of calculations predicting the signal levels in the FM radio band in downtown San Diego, Calif., USA;

FIG. **4** shows the energy reclamation system (ERS) concept in accordance to one embodiment of the present invention;

FIGS. **5**A and **5**B show a physical structure of an energy harvesting antenna and an array of such antennas in accordance to an embodiment of the present invention;

FIG. **6** shows a communication antenna having an insulated shield acting as both a safety shield to the user and an energy harvesting antenna array in accordance to an embodiment of the present invention;

FIG. **7** shows a high level schematic diagram of an Energy Harvesting Subsystem (EHS) in accordance to one embodiment of the present invention;

FIG. **8** shows an energy conversion subsystem (ECS) of a perpetual active antenna (PAA) system in accordance to one embodiment of the present invention;

FIG. **9** shows an application specific integrated circuit (ASIC) chip implementation of the ECS circuitry shown in FIG. **8**;

FIG. **10** shows an array of battery micro-cells of an energy storage subsystem in accordance to an embodiment of the present invention;

FIG. **11** shows a high level schematic diagram of the Two or Multiple Array Sense and Active Mode in accordance to an embodiment of the present invention;

FIG. **12** shows a high level schematic diagram of the Common Antenna Sense and Active Mode in accordance to an embodiment of the present invention;

FIG. **13** shows a high level schematic diagram of the Common Antenna for Simultaneous Energy Harvesting, Sense and Active Mode in accordance to an embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The preferred embodiments of the present invention represent a new technology area for the reclamation and reuse of Radio Frequency (RF) power present due to either design or opportunity in certain frequency bands. It was found that RF signals and noise are available in useful power levels in many frequency bands-particularly in the AM, FM, and TV bands as well as the global system for mobile communications (GSM), digital cellular systems (DCS) and communication system (PCS) bands. The abundance of energy available for harvesting is clearly apparent from the measurements of the ambient RF spectrum. For example, FIG. **1** shows one such measurement in downtown San Diego Calif. The measurement indicates the relative power levels of RF signals between 3 MHz and 1 GHz that can be harvested at this location. As can be seen from the figure, there are ample RF signals from which to reclaim energy. FIG. **2** shows the results of calculations predicting the signal levels in just the AM radio band in downtown San Diego. It shows that many of the AM signals approach or are greater than 1 mW (0 dBm). This is because many commercial AM radio transmitters radiate thousands of watts of power. FIG. **3** shows the results of calculations predicting the signal levels in just the FM radio band in downtown San Diego. These calculations further indicate that there are ample opportunities for harvesting energy from the ambient electromagnetic environment.

Countless candidate devices can benefit from this new energy reclamation technology of the present invention. For instance, the technology can be used for wireless devices that are typically powered by exhaustible or rechargeable batteries. The present invention can be used to provide wireless devices located in certain environments a perpetual life in standby mode and extended normal operational life by using stored reclaimed energy as a power source. The Energy Reclamation System or ERS of the present invention can provide sufficient power (e.g., 0.1–1 mW) to perpetually operate low power wireless devices without external recharging. The ERS of the present invention can also able to provide higher power levels to meet higher demands placed on these wireless devices. According to one preferred embodiment of the present invention, the ERS of the present invention includes a common antenna designed for the energy abundant frequency bands with it's output connected to power conditioning or rectification circuitry appropriately designed for that frequency range. The conditioned or rectified power can then be supplied to a charger (e.g., a trickle charger) or distribution subsystem designed to meet the battery's or storage device's charging specification.

There are innumerable potential applications for the energy reclamation technology of the present invention. For instance, the power harvested from radio frequency (RF) signals can be used for recharging installed power sources, providing power directly to the devices, energizing a device to wake up, and providing emergency backup power for the devices. Countless potential users of this technology also exist. For instance, potential users of the ERS technology may be found in the cellular, personal communications services (PCS) and future generations of wireless communications device industry. Typical consumer wireless electronic products that can be enhanced by ERS technology include, but are not limited to, palm computers, personal digital assistants (PDAs), pocket phones, pagers, cameras, autonomous sensors, garage door openers, remote controls and auto security systems. The energy reclamation technology can also be used in emergency location systems such as GPS position finders, airplane and watercraft emergency beacons and fire alarm systems. The energy reclamation technology can also be applied to wireless personal devices such as video conferencing cell phones, Internet phones and Wireless Personal Area Networks (WPAN). The energy reclamation technology is also flexible enough to be adapted for future electrical and/or electronics applications and devices.

According to a preferred embodiment of the ERS of the present invention, the ERS includes three subsystems: an Energy Harvesting Subsystem (EHS), an Energy Conversion Subsystem (ECS), and an Energy Storage Subsystem (ESS). The Energy Harvesting Subsystem or EHS includes various components used to collect ambient electromagnetic energy via an antenna or antenna array and convert it to electrical current. If acoustic, solar, or motion is the energy source, the antenna or antenna array may be replaced by a transducer or transducer array for the respective energy source. The EHS may also include a combination of antennas and mechanical transducers to collect different types of energy at once or alternately. The Energy Conversion Subsystem or ECS includes circuits, components, and modules integrated together to convert and condition the noise like current out of the antenna or transducer into a form of energy suitable for storage. The Energy Storage Subsystem or ESS includes, for example, batteries, capacitors, and other such components that will store the reclaimed energy for future reuse. The ESS can also include appropriate circuitry (e.g. trickle charger) for charging those devices that the ERS supports.

FIG. **4** shows the ERS concept in accordance to the above preferred embodiment of the present invention. As shown, the ERS **100** includes three parts or subsystems: an Energy Harvesting Subsystem (EHS) **130**, a Power or Energy Conversion Subsystem (ECS) **160**, and a Power or Energy Storage Subsystem (ESS) **190**. The Energy Reclamation System provides for the collection and harvesting of ubiquitous ambient energy over a desired frequency range. In a preferred embodiment, the desired range is 500 KHz–2.4 GHz. For frequencies above 2.4 GHz and for intentionally-radiated high power harvesting applications, the ERS can incorporate the rectenna technology, such as that disclosed in U.S. Pat. No. 5,043,739, into the EHS and the ECS for connection to the ESS. The interconnection of the ESS to the rectenna technology is readily apparent to one skilled in the art.

According to another preferred embodiment of the present invention, the main objective of the ERS **100** is focused on the collection of ambient RF energy present in the ambient environment in the range of 5 watts and below and converting that energy into reusable energy to power numerous electronic devices. The converted energy can be full-wave or half-wave rectified and filtered current. The charging schemes depend on how the ECS **160** is used and its reuse requirements. The output of the ECS **160** can be distributed for reuse to several different types of circuitry. In one embodiment, the converted power may be fed to a trickle charger circuitry that charges a standby energy storage device in the ESS **190** that includes a battery, capacitor, etc. while any excess power is distributed to a charging circuitry for the primary power storage component. Another way to distribute the ECS output power is to feed it directly into circuitry appropriate for affecting indefinite operation of the electronic device in a standby or sleep mode.

The three subsystems of the ERS system are now described in further details.

Energy Harvesting Subsystem (EHS)

The EHS for RF energy collection includes a single antenna of any physical design and shape (e.g., dipole, Yagi, Contrawound Toroidal Helical Antenna, etc.) or an array of such antennas to collect RF energy from any frequency or bands of frequencies. Appropriate baluns (i.e., balanced to unbalanced converters) and impedance matching techniques may be integrated to maximize power transfer into the energy conversion circuitry of the ECS.

For the ERS to reclaim the maximum amount of energy, the energy-harvesting antenna is preferably designed to be a wideband, omnidirectional antenna or antenna array that has maximum efficiency at selected bands of frequencies containing the highest energy levels. If the EHS includes an array of antennas, each antenna in the array can be designed to have maximum efficiency at the same or different bands of frequency from one another. The collected RF energy is then converted into usable DC power using high speed switching semiconductor devices or a diode-type or other suitable rectifier. This power may be used to drive, for example, an amplifier/filter module connected to a second antenna system that is optimized for a particular frequency and application. Thus, one antenna acts as an energy harvester while the other antenna acts as a signal transmitter/receiver. Transforming and/or storing the power garnered by the ERS technology over time makes those devices employing such technology more efficient and extends battery life.

According to a preferred embodiment of the present invention, the antenna design and construction preferably begins with defining a wide band of frequencies over which collection will be required. Next, optimal length or element size is preferably calculated to achieve maximum efficiency over that frequency range. As is apparent to one skilled in the art, antenna efficiency depends on a number of factors related to, for example, the physical design and shape of the antenna. Other factors affecting the amount of energy col lected include signal strength at the receiving location from the transmitting source, which in turn depends on the paths from the transmitting source to the ERS of the present invention.

Once the optimal length or element size of the antenna is calculated, it is subsequently reduced to a fraction of a wavelength while integrating semiconductor amplification and impedance matching circuits in the construction of the antenna elements. The antenna circuit elements are preferably constructed using microprocessor wafer manufacturing techniques and shaped for integration or encapsulation in the antenna elements. This preferred design approach reduces the physical size of the antenna while retaining the energy collection efficiencies of a larger antenna. Antenna efficiency of 50% or more is preferred, for higher antenna efficiency resulting in higher energy collection.

FIG. **5**A shows the physical structure of an energy-harvesting antenna and FIG. **5**B shows an array of such antennas in accordance to an embodiment of the present invention. The cylindrical physical design shape was selected for easy explanation of the design concept; however, any physical shape (e.g., toroid, etc.) may evolve that still meets the ERS design concepts and goals. As shown, the antenna **200** includes a standard metal outer sleeve **210**. Under the sleeve **210** is a semiconductor material **215** used to create an apparent physical length optimized for reception to accommodate size limitations. Enclosed by the antenna outer sleeve **210** and semiconductor material **215** is electronic circuitry **220**, which is constructed using microprocessor and ASIC semiconductor technology to develop integrated support modules such as trickle charger, amplifier and filter for the ECS, which will be described later. The outer sleeve **210** may also be used as a heat sink for the electronic circuitry **220**.

According to another preferred embodiment of the present invention, the antenna can also be constructed to comply with recent radiation hazard concerns with wireless and cellular telephone technologies and at the same time increases power collection potential. There are a number of ways to direct the RF energy away from the user or shield the user using cupped shields in the antenna system design. Such design could be optimized to provide an insulated shield and enhance energy collection materials in the antenna system construction. FIG. **6** shows an antenna **210** covered by an insulated shield **232**. The antenna **210** performs its primary function for a specific application, such as receiving RF signals for normal wireless communication. The insulated shield **232** can be used as an electromagnetic shield and also as an ERS antenna or antenna array for energy harvesting. As shown in the figure, the RF shielding material is preferably located on the inside surface **234** of the shield **232**. As is known in the art, the shielding material may be of any material suitable for shielding RF radiation generated by the transmitting antenna **210**. Within the shield **232** is an antenna array **236**, with associated integrated modules of the ECS, for energy harvesting. Alternatively, within the shield **232** is a single antenna for energy harvesting. Each antenna in array **236** or the alternative single antenna is depicted in FIG. **5**A.

FIG. **7** shows a high level schematic diagram of an Energy Harvesting Subsystem (EHS) in accordance to one embodiment of the present invention. The EHS includes a wideband omnidirectional antenna or antenna array **255** for energy harvesting, as described earlier. The antenna **255** may be used to harvest energy from ambient RF signals **3** and/or RF power transfer signal **2** from a remotely-located, intentional power source. It may also be used to concurrently receive the desired communication signal **1** along with the energy-harvesting signals **2** and **3**. The combined signals are then separated by filter bank **250** which will be described in greater detail later.

FIG. **7** also shows that other energy harvesting technologies for collecting ambient solar, acoustical, and mechanical energy can be integrated as part of the EHS to supply electrical energy into the ECS. The electrical energy supplied by these other energy harvesting technologies may interface with the rectifier **162** and/or filter circuitries **250**, which is then fed to filter **275** in the ECS. As apparent to one skilled in the art, the ECS interface depends on the input frequency, quality, quantity and type (constant DC or some form of pulsating DC) of energy supplied.

For ambient solar energy, an array of solar cells or a solar energy conversion device **260** known in the art (such as a MEMS solar/heat electric generator) is preferably used to absorb solar energy from the environment and convert it to electrical energy. This energy is then delivered to a rectifier **162** of the ECS. An array of solar cells may be used to produce sufficient power for a variety of applications.

For ambient acoustical energy, a piezoelectric transducer **266** known in the art can be used to absorb ambient or intentional sound from the environment, which in turn causes a crystal to vibrate and through the piezoelectric effect produce an output voltage. The output energy is then delivered to the rectifier **162** of the ECS. An array of transducers may be used to produce sufficient power for a variety of applications.

For mechanical energy, the energy is preferably converted to electric energy through a MEMS Mechanical Transducer (Generator) **270** known in the art. The MEMS Mechanical Transducer **240** takes mechanical energy derived from the natural acceleration of a thing or person while in transport (e.g., person walking) or in use. These devices can wind a spring, force a piston to move or use some other method to store and convert acceleration energy collected into electrical energy. The output energy is then delivered to the rectifier **162** of the ECS. An array of such transducers may be used to produce sufficient power for a variety of applications.

Energy Conversion Subsystem (ECS)

According to a preferred embodiment of the present invention, the ECS for RF energy collection includes a power charger and other circuitry for performing RF to DC power conversion. Because the energy harvesting in the present invention is focused mainly on recovering small amounts of energy over long periods of time, charging energy storage devices (rechargeable batteries, etc.) may be done over a period of time by trickle charge. As is known in the art, trickle charging is a charging scheme in which a power storage component is charged at a fraction of its capacity rate. FIG. **8** shows an ECS **160** for use in conjunction with the EHS to perform RF to DC power conversion, with specifications in accordance to an embodiment of the present invention.

The ECS **160** preferably includes a transformer **161** connected to an energy harvesting antenna **130** for receiving the RF energy. The transformer **161** is also connected to a diode rectifier circuit **162**. The transformer **161** and the rectifier circuit **162** are designed to match impedance with one another to prevent undesired energy loss between these two elements. After rectification, the converted energy is sent through another transformer **163** and onward to a trickle charger **164**. Again, the second transformer **163** and the trickle charger **164** are designed to match impedance with one another to prevent undesired energy loss between these two elements. The trickle charger **164** is then used to deliver power to a battery **165** for storage and use with an electronic device, such as a standby device **166**. The battery is actually a part of the ESS, which will be described later.

If an array of energy harvesting antennas is used to receive the RF energy, the power collected from all channels (i.e., the power density at all antenna inputs) can be combined, rectified and forwarded to the trickle charger **164**. The power available in each channel is comparable; thus, for example, if 100 N nW is required for a load, the power collected in N channels at 100 nW each can be combined to generate the required power.

Although not shown in FIG. **8**, the converted and rectified power may be fed through a filter for signal purification prior to sending it to the transformer **163** and onward to the trickle charger **164**. FIG. **7** shows the use of such filter **275**, with like numbers for like elements shown in FIG. **8**. It should be noted that FIG. **7** depicts a broad overview of the ERS concept according to an embodiment of the present invention; therefore, the figure does not show all detailed components of the ECS as shown in FIG. **8**.

FIG. **9** shows an ASIC chip implementation of the ECS circuitry shown in FIG. **8**. The RF input **280** corresponds to the antenna **130** of FIG. **8** for receiving harvested RF energy; the MEMS RF transformer **281** corresponds to the transformer **161** of FIG. **8**; the RF power ASIC **282** corresponds to the rectifier circuit **162** and the trickle charger **164** of FIG. **8** and any other desired circuitry for the ECS, such as filters, impedance matching circuitry, etc.; and the micro-cell battery **285** corresponds to the battery **165** of FIG. **8**. The battery **165** or **285** is a part of the ESS, which is described next.

Energy Storage Subsystem (ESS)

According to a preferred embodiment of the present invention, the ESS includes a rechargeable battery **165**, which can be a complete battery or an array of micro-cells of battery. If battery **165** includes a complete battery, the trickle charger **164** preferably provides a larger potential difference between terminals and more power for charging during a period of time. If battery **165** includes individual battery cells, the trickle charger **164** preferably provides smaller amounts of power to each individual battery cell, with the charging proceeding on a cell by cell basis. FIG. **10** shows an array of battery micro-cells according to an embodiment of the present invention. The ESS charges one or more micro-cells as available energy allows. As the load requires power, the combined energy is drawn from a charged P×Q subset (sub-array) of the N×M set (array) of cells to power the device on standby or active mode. N, M, P and Q are all natural numbers (i.e., 1, 2, 3 . . .), wherein P is a number less than N, and Q is a number less than M. Charging of the remaining cells continues whenever ambient power is available. As the load depletes cells, switches such as switch **276** are used to replace the depleted cells with charged cells. The rotation of depleted cells and charged cells continues as required. As much energy as possible is banked to ensure adequate power is available on demand. Thus, energy is managed on a micro-cell basis.

According to the present invention, there are numerous ways in which the ERS technology can be configured to harvest RF energy. In one embodiment of the present invention, an electronic device, such as a cellular phone, is preferably designed to have two antennas, one for energy harvesting at a desired frequency band and one for performing RF communication at a different frequency band. An example of this dual-antenna design is previously shown in FIG. **6**. FIG. **11** shows a high level schematic diagram **300** of a dual-antenna design ERS in accordance to one embodiment of the present invention. As explained earlier, the frequency band at which an antenna receives and transmits signals depends on the physical dimensions of the antenna. Thus, the energy-harvesting antenna **310** and the communication antenna **320** are designed with different dimensions to receive signals at their respectively desired frequency bands. The energy-harvesting antenna **310** may comprise one antenna or an array of like antennas. In this embodiment, parallel operations are possible for energy harvesting and communication. In other words, a user can be using the phone for wireless communication in one frequency band, while at the same instance the energy harvesting antenna **310** and associated circuitry within the phone is collecting electromagnetic energy in another frequency band. The collected electromagnetic energy is passed to integrated circuitry that converts the RF energy into electrical energy. This electrical energy is then distributed for storage in batteries of the ESS **340** for use in powering various electrical components of the phone.

As shown in FIG. **11**, communication signals received by the device antenna **320** are sent to communication processing circuitry **390** for transmission and reception. The energy harvesting antenna **310**, on the other hand, is connected to an Energy Conversion Subsystem or ECS **330**, which is designed to match the impedance and frequency band or bands of the energy harvesting antenna **310** for maximizing input energy received. The ECS **330** is as described earlier with respect to FIG. **8**. Its output is conditioned power coming from a trickle charger designed to efficiently charge the power storage devices, such as NiCd, NiMH, or LiIon battery cells, in the energy storage subsystem or ESS **340**. Any excess conditioned power from the trickle charger of the ECS **330** is then distributed to the device's main power source **380**.

The ESS **340** is as described earlier with respect to FIGS. **8** and **10**. It stores the standby power and distributes it to the electronic switching circuitry **350** and the Monitor and Activation Circuitry **360** in both active and standby modes of the device. Alternatively, the ESS **340** may be used to power circuitries **350** and **360** in standby mode of the device, while the device power source **380** powers those circuitries in the device's active mode. The ESS **340** may also be used to complement the device power source **380**.

According to yet another embodiment of the present invention, the harvesting antenna **310** and associated circuitries **330** and **340** may be physically located outside of the cellular phone but yet capable of electrical coupling with the phone. For example, the harvesting antenna **310** and associated circuitries may be built into portable physical structures such as briefcases, suitcases, purses, and wearable apparel (cloths) or fixed physical structures by any known means in the art. Thus, when the phone is desired to be charged, it can be stored in any of the aforementioned physical structures and electrically coupled to the built-in harvesting antenna and circuitry for charging purposes. These separate structures facilitate the use of larger antennas (transducers) or arrays of smaller elements whichever is most efficient for collecting the ambient energy. In any case, the larger surface area intercepts more energy.

According to a further embodiment for harvesting RF energy in the present invention, an electronic device, such as a cellular phone, is designed to have two antennas that respectively harvest RF energy and perform wireless communications in the same desired frequency band. In this embodiment, the energy-harvesting antenna and the communication antenna are designed with the same physical dimensions. The communication antenna is connected to a signal level detection system for detecting the strength of RF signals in the desired frequency band. If the strength of the desired RF signals is sufficiently high to sustain both RF communication and energy harvesting, then the level detection system will cause the activation of the energy-harvesting antenna. Thus harvesting proceeds in parallel with communication and the RF energy is converted into electrical power useful for powering various components of the cellular phone without degrading phone service.

According to another embodiment for harvesting RF energy in the present invention, a single antenna or an array of like antennas in an electronic device, such as a cellular phone, is used for both energy harvesting and communication purposes. FIG. **12** shows a high level schematic diagram **400** of the Common Antenna for Simultaneous Energy Harvesting, Sense and Active Mode in accordance with this embodiment. This mode requires modifications to the device's antenna switching and filter circuitry to allow communication signals in along with collected RF energy for harvesting. The communication signals are filtered and directed to the receiver's communication circuitry while the noise or interference portions of the received RF signals are filtered out and sent to the energy harvesting circuitry within the electronic device. This noise filtering technique is illustrated back in FIG. **7**, where incoming communication signals **1** and the ambient RF energy **3** are filtered by a filter bank **250**, which may reside in the switching circuitry **450** shown in FIG. **12**. The communication signals **1** are then directed to the receiver's communication circuitry **255**, which is shown as communication processing circuitry **490** in FIG. **12**, while the noise or interference RF portions **3** are filtered out and sent to the energy harvesting circuitries, which include the ECS **430** and ESS **440**.

According to still another embodiment for harvesting RF energy in the present invention, a single antenna or an array of like antennas in an electronic device, such as a cellular phone, is again used for both energy harvesting and communication purposes. FIG. **13** shows a high level schematic diagram **500** of the Common Antenna Sense and Active Mode in accordance with this embodiment. Here, the single antenna or antenna array **510** is connected to the energy harvesting circuitries **530** and **540** and communication circuitry **590** by a switch **550**, which is slaved to the various powering modes of the electronic device. For example, when the phone is in the power-on mode, the switch **550** is programmed to connect the antenna **510** with the communication circuitry **590** to receive the RF communication signals and enable wireless communication with the phone. However, when the phone is in the standby mode, the switch is programmed to connect the antenna **510** with the harvesting circuitries; which then converts the electromagnetic energy in the same RF communication signals into electrical power for powering the electronic device in the standby mode.

The normal operation of an electronic device having the ERS technology as depicted in FIGS. **11** and **12** may also include a standby or sleep mode, as described above with regard to FIG. **13**. When the electronic device is set in a standby mode, the electronic switching circuitry (**350** or **450**) of the ERS (**300** or **400**) within the electronic device is placed in a Sense Mode. As shown in both FIGS. **11** and **12**, the monitor and activation circuitry (**360** or **460**) is used to set the electronic switching circuitry (**350** or **450**) in a sense mode. In this mode, signals received by the device's communication antenna (**320** or **410**) are sent to a band pass filter (**370** or **470**) to retrieve an activation signal, which is passed on to the monitor and activation circuitry (**360** or **460**) that is watching for the activation signal. The activation signal can be anything defined by the device's manufacturer, such as a trigger signal or an emergency beacon for activating a desired system. When a turn-on signal brings the electronic device out of standby mode, the monitor and activation circuitry (**360** or **460**) sends an active signal to the electronic switching circuitry (**350** or **450**) to connect the communication antenna (**320** or **410**) to the device's communication circuitry (**390** or **490**). The device's turn-on signal may be manually provided by a user turning on the electronic device, or it may be brought on by the reception of the activation signal by the monitor and activation circuitry (**360** or **460**).

Again, the harvesting circuitries in the above embodiments may be physically located apart from the communication circuitry. For example, the harvesting circuitries may be built into portable structures such as briefcases, suitcases, handbags, and purses by any known means in the art, while the communication circuitries and the switch are built into the phone itself. Thus, when the phone is desired to be charged, it can be stored in any of the aforementioned portable structures and electrically coupled to the harvesting circuitries built into the structure for charging purpose.

Although only a few exemplary embodiments of this invention have been described in detail above, those skilled in the art will readily appreciate that many modifications are possible in the exemplary embodiments without materially departing from the novel teachings and advantages of this invention. Accordingly, all such modifications are intended to be included within the scope of this invention as defined in the following claims. Furthermore, any means-plus-function clauses in the claims (invoked only if expressly recited) are intended to cover the structures described herein as performing the recited function and all equivalents thereto, including, but not limited to, structural equivalents, equivalent structures, and other equivalents.

What is claimed is:

1. A method for harvesting and utilizing electromagnetic energy, comprising:

receiving ambient electromagnetic energy;

converting the ambient electromagnetic energy into DC electrical power;

charging a power storage component with the DC electrical power, the power storage component comprises a N×M array of battery micro-cells, N and M are natural numbers;

providing a device power source for powering an electrical device; and

drawing power from the power storage component to power the electrical device, the drawing power from the power storage component to power the electrical device includes:

determining a charged P×Q sub-array of the N×M array of battery micro-cells, P and Q are natural numbers less than N and M, respectively; and

drawing power from the charged P×Q sub-array to power the electrical device.

2. The method of claim **1**, the charging the power storage component with the DC electrical power comprises includes:

charging at least one remaining micro-cell of battery in the N×M array that is not in the charged P×Q sub-array;

substituting the P×Q sub-array with the at least one remaining micro-cell of battery once the P×Q sub array is depleted of power; and

charging the depleted P×Q sub-array with the DC electrical power.

3. The method of claim **2**, the drawing power from the power storage component to power the electrical device further includes:

drawing power from the at least one remaining charged micro-cell of battery to power the electrical device.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,882,128 B1
DATED : April 19, 2005
INVENTOR(S) : Rahmel

Page 1 of 1

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Title page,
Item [56], **References Cited**, U.S. PATENT DOCUMENTS, please add the following:
-- 6,045,339 4/4/2000 Berg 417/332 --.

Column 12,
Line 51, please change “component with the DC electrical power comprises” to
-- component with the DC electrical power --.

Signed and Sealed this

Twenty-second Day of November, 2005

Jon W. Dudas

JON W. DUDAS
Director of the United States Patent and Trademark Office

The Mystery Crystal Set

This is my version of the famous Mystery Crystal Set from Australia. Click Here for a complete set of plans and a brief history of the set.

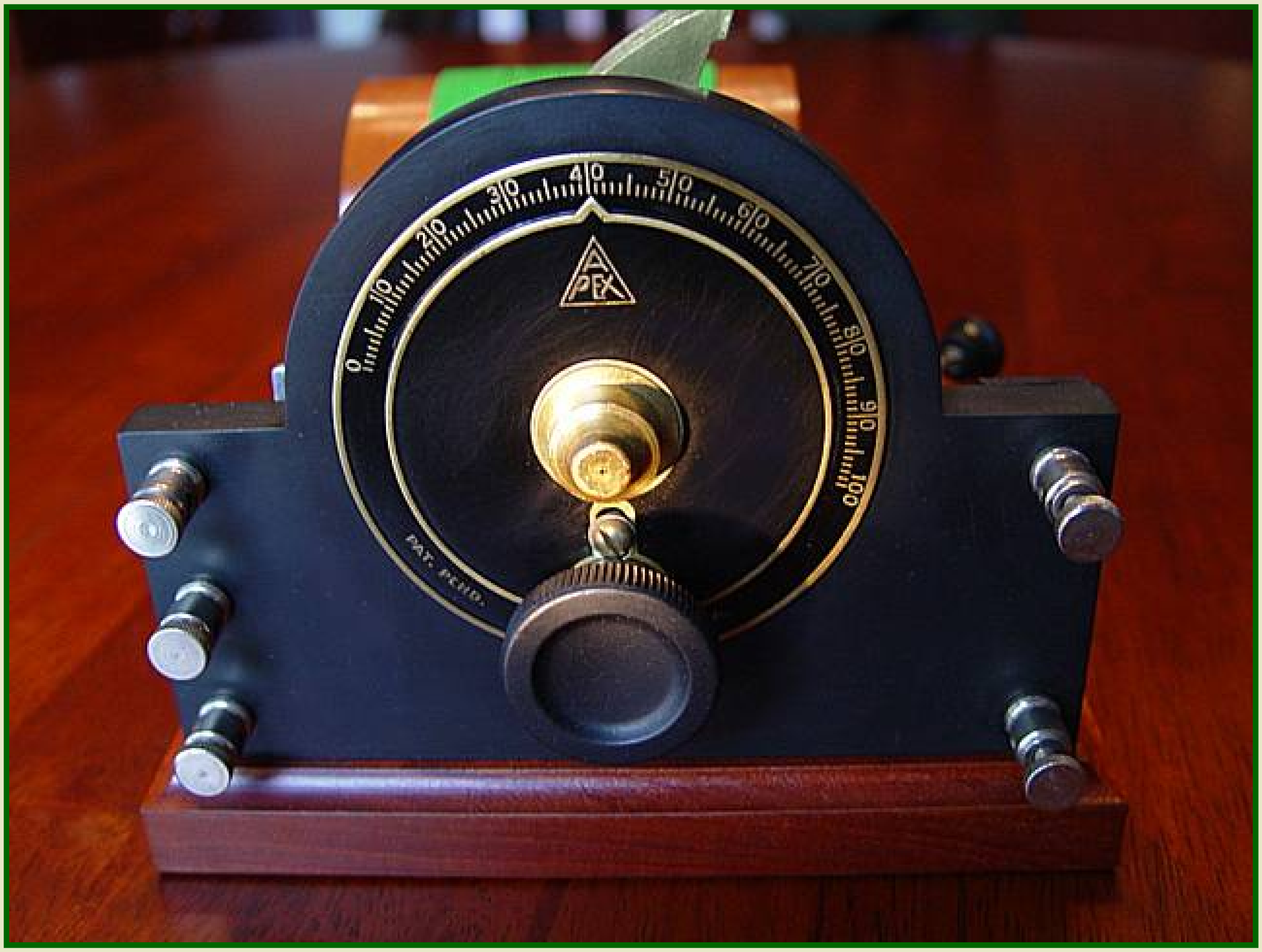

Front view showing the nice vintage Apex dial I used. The black plastic material is half-inch thick DuPont Delrin®. which looks like the old hard rubber that was used in the 1920's and 30's. It is very hard and machines easily as you can see by the fancy cut over the dial. This was the first time I used this, and I was very pleased with the results. The binding posts are authentic NOS 1920's style. The three on the left are for the various combinations of antenna and ground connections. The two on the right are for the headphones.

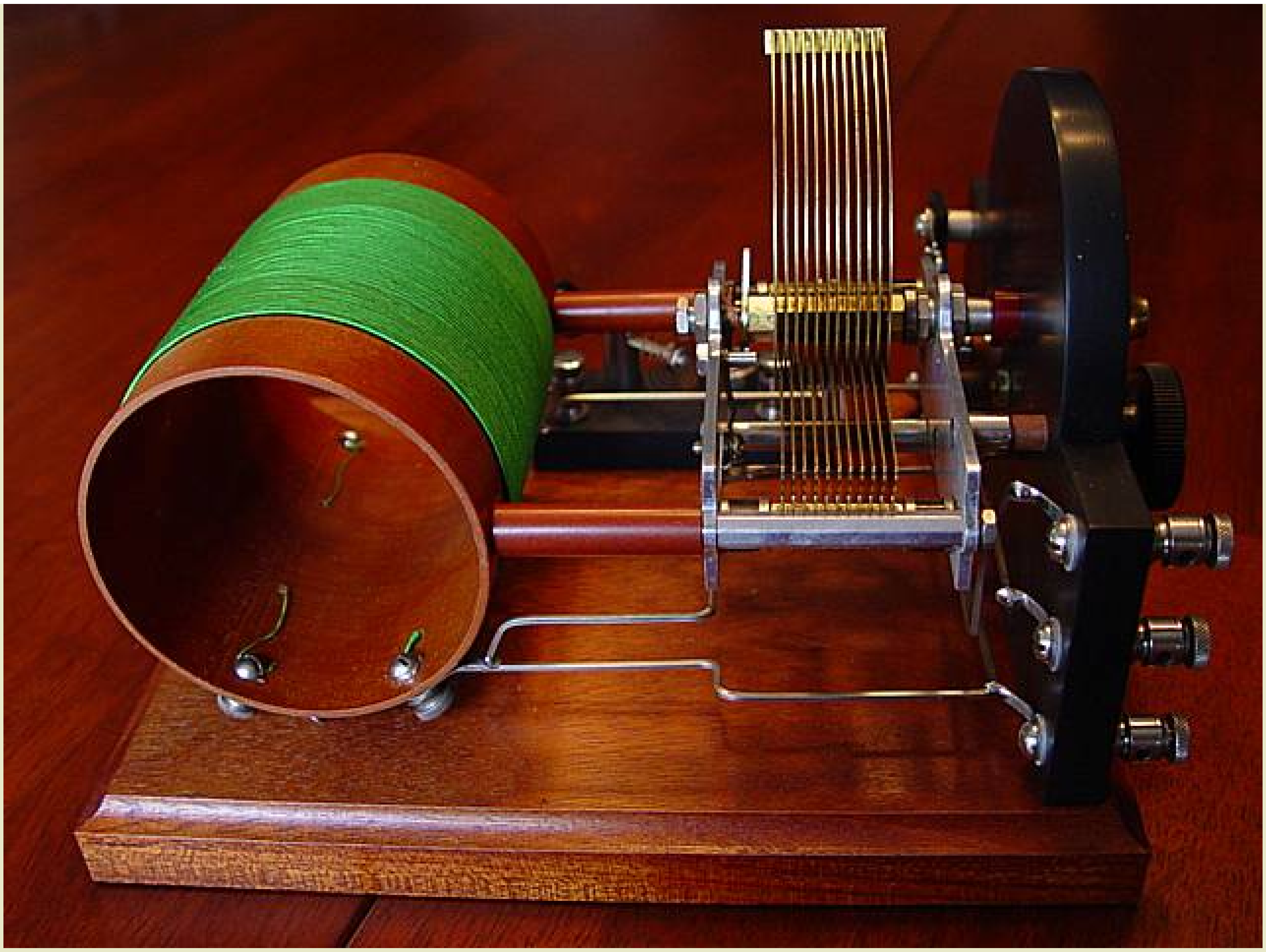

Left side view. I used solid mahogany for the base to which I applied 6 coats of clear gloss lacquer, sanding between each coat. There is no stain on the wood, what you see is the natural color. The wiring is all # 14 square buss wire I obtained from Antique Electronic Supply several years ago. Unfortunately, they were not suppling this size when I checked last. If you find a new source for this wire, please let me know. The variable capacitor is an AMSCO straight line frequency unit. Notice the shape of the plates. This makes the tuning spread more evenly across the dial. A very well made part. For the coil construction, I used a piece of three-inch diameter phenolic tubing. The wire is NOS authentic green silk-covered from my collection. I really hate to use any of this wire because it is irreplaceable. Again, if anyone knows a source for this kind of wire, please, please let me know. I usually apply a coat or two of polystyrene when I finish a winding, but I didn't coat this coil because I didn't want to risk changing the look of it. It seems to be very stable the way it is.

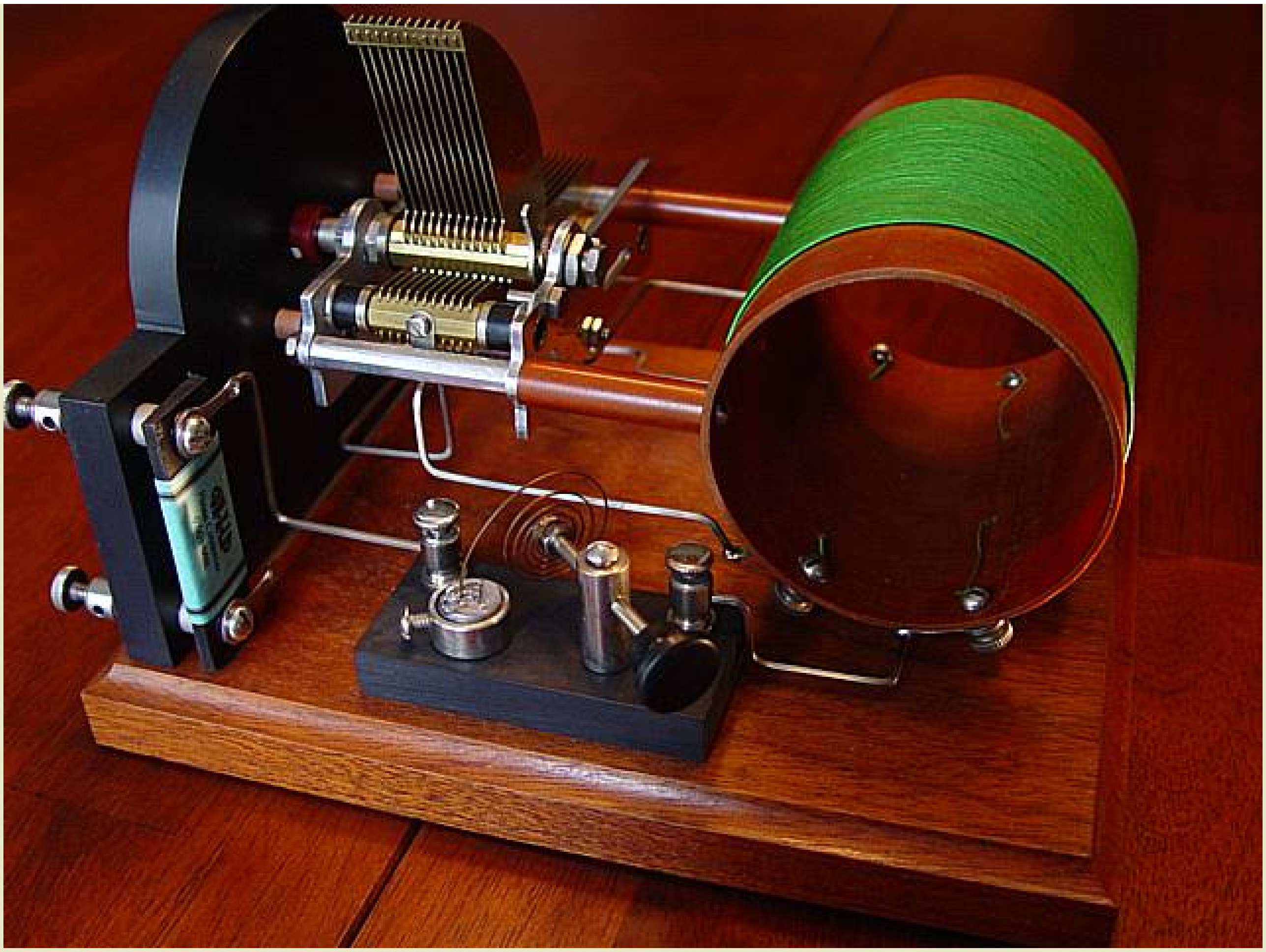

Right side view. Here, you can see one of my homebrew crystal detectors. The bypass capacitor is also homebrew with a scanned copy of an original capacitor label.

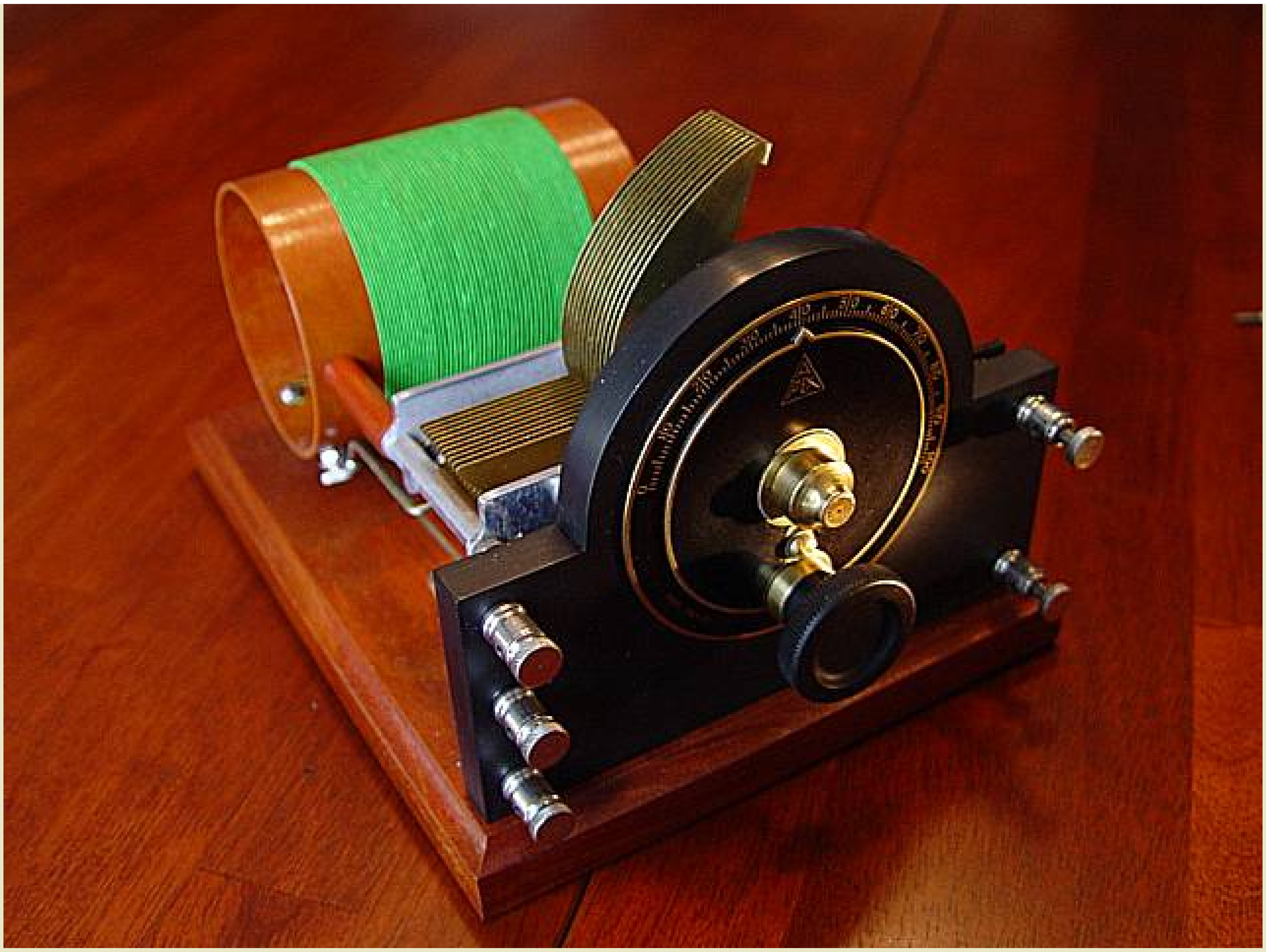

Another view of the AMSCO capacitor with the unusual plates. If you look closely at the coil, you can see the second winding in between the first.

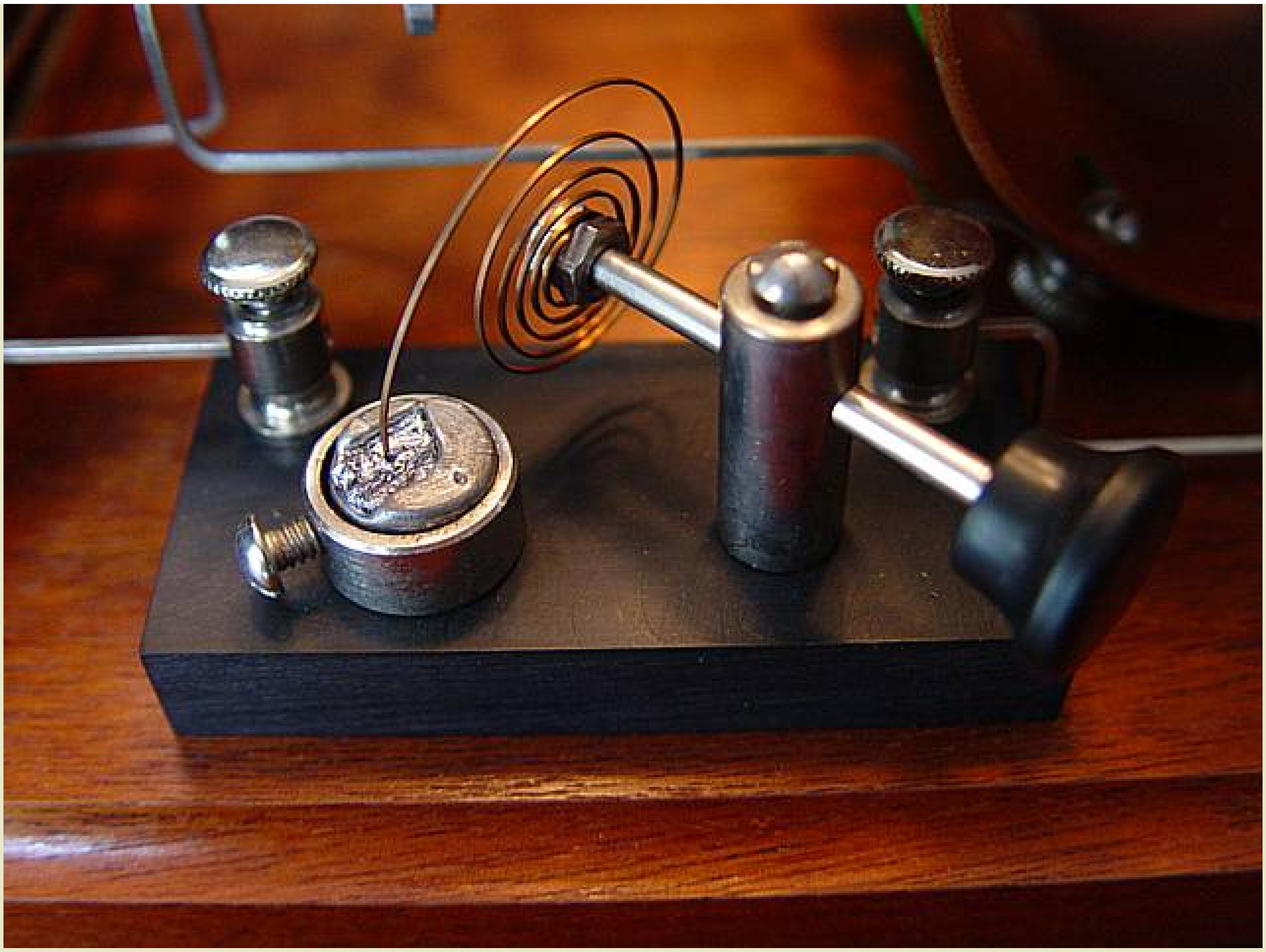

A closeup of the detector. This is a CARCO style from the 1920's. Here again, I used Delrin® for the base with vintage terminals and all other parts made of brass which I nickel plated. I really like the CARCO catswhisker construction, it is easy to adjust and stays where you leave it. The crystal is a standard Galena mounted in low temp alloy.

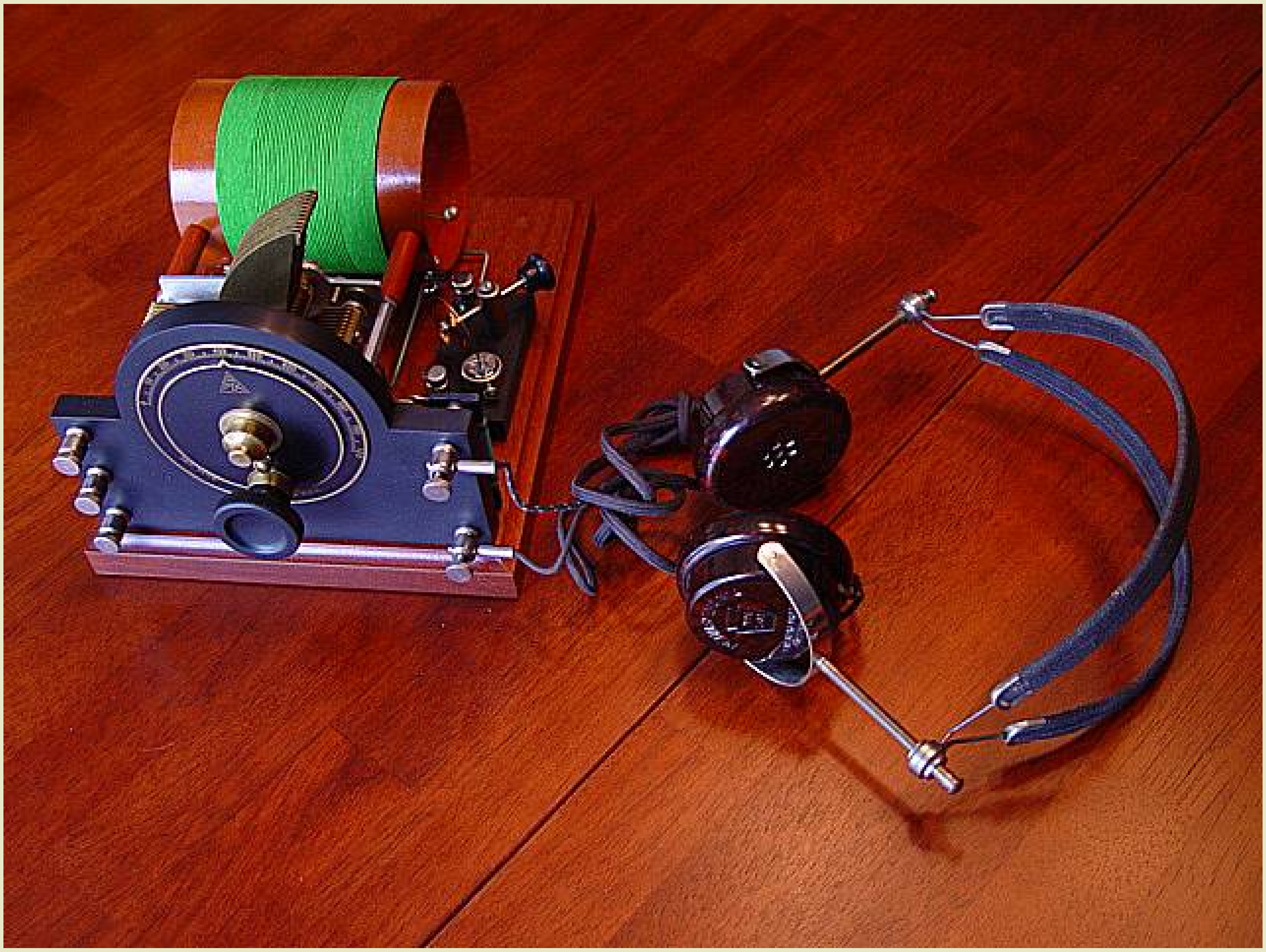

Here I am using a set of 3000 ohm Pennsylvania brand headphones. These work fairly well. I have had best results with a matching transformer and sound-powered phones. This set is very selective and you can get more volume on local stations by changing the input connections.

Click here to return to the VINTAGE HOMEBREW page

© 2006 ~ 2008 Michael R. Starcher ~ All Rights Reserved

Copyright 2001 ...[view copyright info]...

N5ESE's Crystal Checker

← PWR (Up = ON)
← TO FREQ COUNTER
N5FC Crystal Checker
built by N5FC Sep 2001

(click on any picture to see larger version)

NOTE: 'N5FC' is my former call.
This project was constructed while that call was valid, and you may observe references to it.

The N5ESE Crystal Checker is a weekend project based on a simple oscillator circuit designed to work readily with most any fundamental-mode crystal in the HF range. Used in conjunction with a frequency counter, it allows you to test the viability of unknown crystals, or match a bunch of crystals for use in filters.

The circuit is quite straightforward. It was adapted from a similar circuit by W1FB, presented in QST Jan 1990. The schematic, as built, appears below:

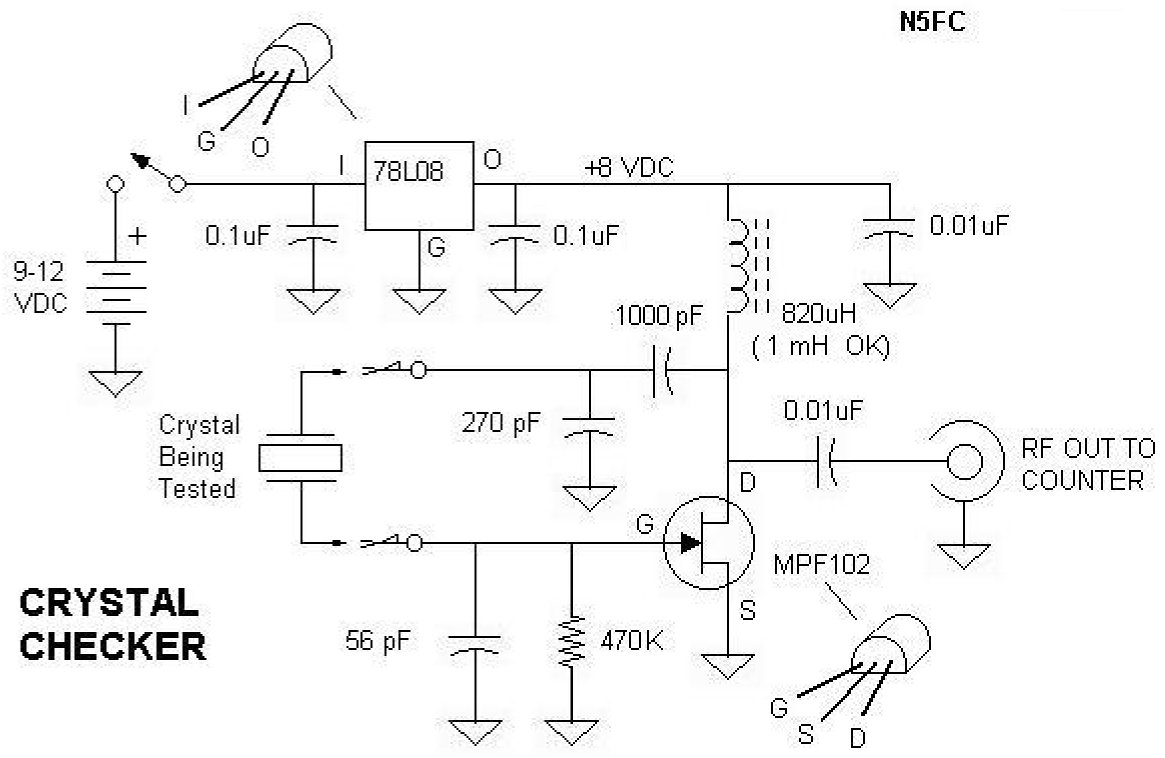
N5FC
I
G
O
78L08
I
O
G
+8 VDC
9-12
VDC
+
0.1uF
0.1uF
0.01uF
820uH
(1 mH OK)
1000 pF
Crystal
Being
Tested
270 pF
0.01uF
RF OUT TO
COUNTER
D
G
S
MPF102
CRYSTAL
CHECKER
56 pF
470K
G
S
D

I love putting things in Altoids boxes (could you tell?). This project was the perfect size for it. We built a small printed circuit board (PCB), cutting islands and traces using a hobby knife (CAUTION! Wear Safety Glasses!). The PCB is shown below:

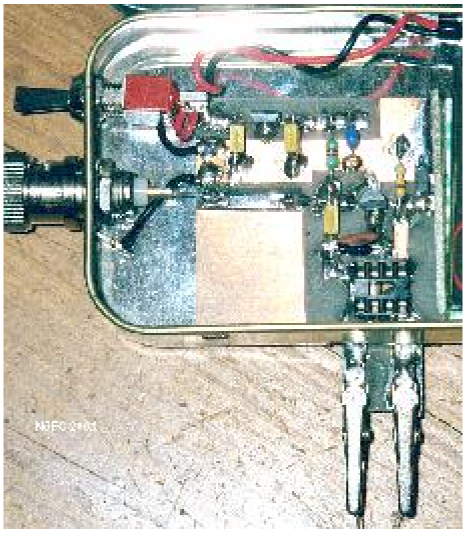

As you can see, the PCB extends through a hole in the Altoids box, where two alligator clips (RS 270-380) are solder-mounted. We can clip our crystal here, for testing. An 8-pin DIP socket is also included, surface-mounted to the top of the board, and we can mount a crystal here instead, if we like.

The Altoids box has plenty of room for a 9-Volt battery, making the unit self-contained and readily portable.

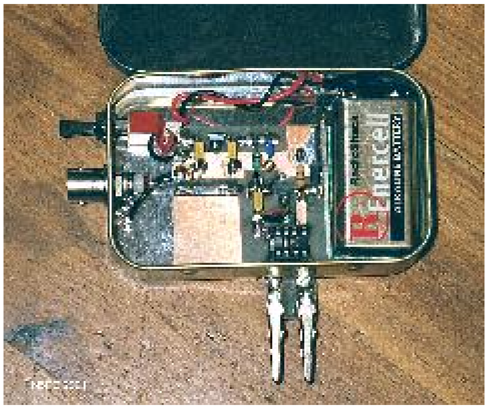

In use, we connect a frequency counter to the Crystal Checker's output, via a BNC cable, and connect a crystal for testing. See below, where we have connected it to the AzQRPions Stinger Singer Audible Frequency Counter (packaged N5ESE-style, in an Altoids box, of course...)

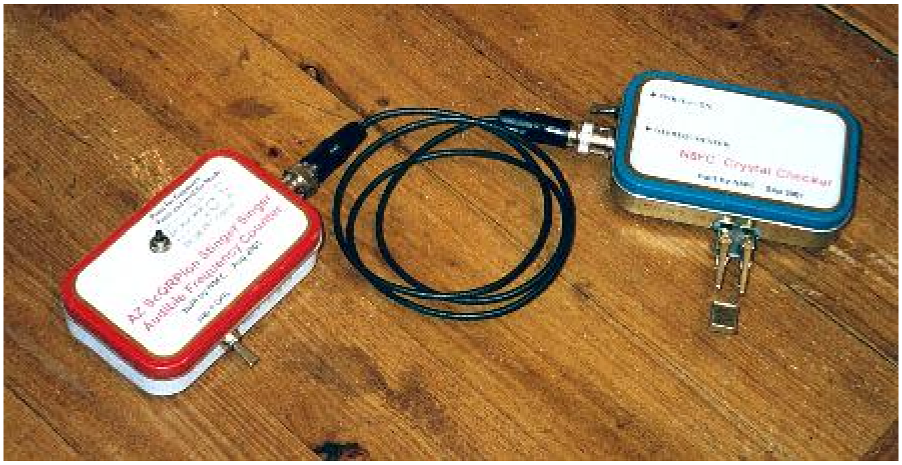
AZ BcQRPion Stinger Singer
Audible Frequency Counter
N5FC Crystal Checker

Return to N5ESE home page

Overseer: Monty Northrup ... e-mail ... leave e-mail ...

maddog homepage Visit our regular (non-ham, but very popular) homepage

(No Model.)

N. B. STUBBLEFIELD.

LIGHTING DEVICE.

No. 329,864. Patented Nov. 3, 1885.

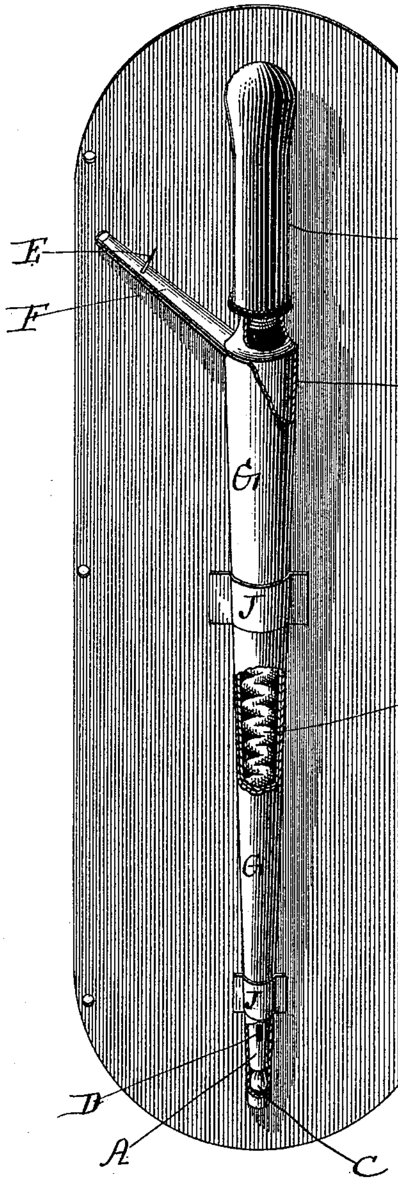

Witnesses:

W. F. Peterson
L C Linn

Inventor.

Nathan B. Stubblefield.

N. PETERS, Photo-Lithographer, Washington, D. C.

UNITED STATES PATENT OFFICE.

NATHAN B. STUBBLEFIELD, OF MURRAY, KENTUCKY.

LIGHTING DEVICE.

SPECIFICATION forming part of Letters Patent No. 329,864, dated November 3, 1885.

Application filed February 25, 1885. Serial No. 157,045. (No model.)

To all whom it may concern:

Be it known that I, NATHAN B. STUBBLEFIELD, a citizen of the United States, residing at Murray, in the county of Callaway and State of Kentucky, have invented a new and useful Improvement in Coal-Oil-Lamp Lighters, of which the following is a description.

The object of this invention is to light lamps which have glass chimneys without the removal of the latter.

To this end my invention consists in a tapering tube provided with a removable handle, a wick, a sheath fitting the tube, and a support for the sheath, constructed and combined as hereinafter described and claimed, reference being had to the accompanying drawing, which is a front view, in perspective, of my invention, about three-quarters full size and partly broken away to expose the interior.

A represents a tin tube tapering from its handle B to its point, which is open. The handle is fitted to screw air-tight into the tube.

C represents a common cotton wick, which should be long enough to nearly fill the tube when rammed or crowded in. This wick is to project from the point like a lamp-wick, and about a tea-spoonful of oil is to be placed in the tube with the wick. The air-tight fit of the handle prevents oil from escaping by leakage.

D represents a groove through which the wick may be picked. This being dirty work for scissors, &c., I have provided a pick in the form of a spur, E, projecting permanently from an arm or trough, F, which rises at one side of the tin sheath G.

The torch A B C is shown as inserted in the sheath, into which it fits closely, where it is kept when not in use, to prevent a disagreeable smell and for sake of neatness. This sheath is removably secured to a back, H, which may be tacked to the wall at any suitable place. That it may be removed for the purpose of being cleaned, it is held in bands J, which are permanently fixed to the back H.

When the wick is trimmed by the pick E, the charred ends, falling into the trough F, will slide down into the sheath, and there accumulate until it becomes necessary to clean the sheath. By removing the sheath from the bands it may be inverted and jarred over the fire, when its contents will be disposed of.

The torch may be lighted at a fire or with a match, and, being long enough to reach down into the chimney of a lamp, it will light the lamp without the necessity of removing the chimney. It may also be used for lighting gas and for other similar purposes.

What I claim as my invention, and desire to secure by Letters Patent, is—

1. The combination of the tapering tube A, the handle B, closely screwed therein, the back H, provided with bands J, and the sheath G, fitted to the said tapering tube and to the said bands, substantially as shown and described.

2. A sheath for lamp-lighters, having an open end and provided with a trough extended laterally from and opening into the sheath, and a spur or pick mounted on said trough, substantially as set forth.

3. A lighting device comprising a tube having an open end and a slot adjacent such end, and a sheath fitted to receive such tube, and provided with a lateral arm and a spur or pick mounted on said arm, substantially as set forth.

4. The lighting device herein described, consisting of the hollow tube having an open end and a slot adjacent such end, a sheath fitted to receive said tube, the latter being removable from the sheath, a trough extended from and opening into the sheath, and a spur or pick mounted on said trough, substantially as set forth.

NATHAN B. STUBBLEFIELD.

Attest:
S. L. HOLLAND,
L. C. LINN.

rexresearch.com

Nathan STUBBLEFIELD

Earth Battery

---

1 | Tesla Generator - Build Your Own Generator Eliminate Your Power Bill Easy. Do It Yourself. Great Discount Now! Just $49. tesla-energy.org/Generator
2 | Start Download - View PDF Convert From Doc to PDF, PDF to Doc Simply With The Free Online App! download.fromdoctopdf.com
3 | Start Download (Free) - PDF To Word Convert From Doc to PDF, PDF to Doc. 100% Free, Get PDF Baron! pdfbaron.com
1 | Tesla Generator - Build Your Own Generator
Eliminate Your Power Bill Easy. Do It Yourself. Great Discount Now! Just $49. tesla-energy.org/Generator
2 | Download PDF
To View PDF, Download Here free.propdfconverter.com

---

https://borderlandsciences.org/journal/vol/51/n03/Vassilatos_on_Stubblefield_Earth_Energy.html

Earth Energy & Vocal Radio: Nathan Stubblefield

by

Gerry Vassilatos

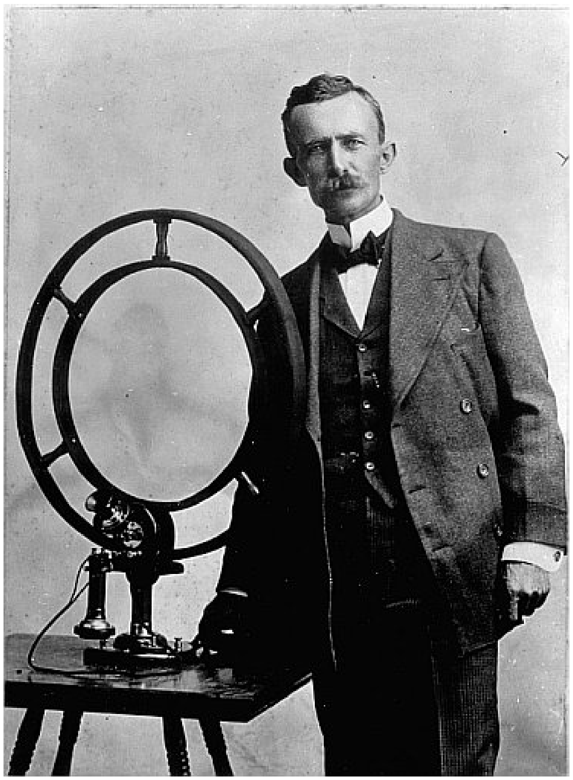

Inventor, Nathan Stubblefield, with wireless telephone.

The notion of drawing up electrical power from the ground sounds incredibly fanciful to conventional scientists, but numerous patents support the claim. A number of retrieved patents list compact batteries which can operate small appliances by drawing up ground electricity. Others describe methods whereby enough usable electrical power may be drawn out of the ground itself for industrial use.

Earth batteries have been detailed in a previous article. Their history can be traced back to experiments performed by Luigi Galvani on copper plates in deep stone water wells. Currents derived through these gave Galvani and his assistants "shivering thrills and joyous shocks".

The most notable earth battery patent is one which operated arc lamps by drawing "a constant electromotive force of commercial value" directly from the ground. In addition to this remarkable claim, a vocal radio broadcast system... through the ground.

It all began one hundred and fifty years ago with the advent of telegraphy. The meandering wire went through rich dark ever-green forests. Lush com flowing valleys sparkled and languidly waved as the linesmen drew their trail. Across meadows where wild flowers covered the earth in fragrant bouquets, there went the line in its curious path.

Over rolling hills which soared into the hazy sunlight, the telegraph linesmen sang as they went. And the lines followed a mysteriously winding trail which few discern. Sorciers and Templars alike called these trails "woivres". Anglo-Saxon geomancers called it VRIL, the black radiant organismic energy of earth.

Who is Nathan B. Stubblefield, and why do most citizens in the state of Kentucky justifiably revere his name? A native of Murray, Kentucky, Nathan B. Stubblefield had a love for the lonely wooded areas on the outskirts of town. Certain spots in these woods were mysterious and possessed of a strange magic all their own. Few would seek the magic of these places, and learn its true and deep power. There the song is sweet and deep, and still.

Vitalizing and sense-provocative, Stubblefield knew that specific locations could be unique natural energy sources. Rock outcrops, evergreens, and flowing springs each registered as strong sensual attractants. Could it be that they were sensual attractants because they conducted and projected special energetic ground currents? Can it he that we are enthralled and drawn into certain spots because of their projective energy? Furthermore, what is the exact nature of this energy? Does it contain or exceed the qualities of electricity?

A self-educated experimenter and avid reader of every kind of scientific literature, Nathan Stubblefield supplemented his living with farming. He remained a practical inventor of some of the most unusual electrical devices ever developed in America.. What he discovered and demonstrated before hundreds of qualified observers in his day seems to challenge many basic axioms of electrical dynamics.

He developed an extraordinary receiver of ground electricity (which produced great quantities of electric power) and numerous "vibrating telephones" which were used by local residents in 1887. The telephonic devices were patented in 1888 and represent the first commercial wireless telephones, using the ground as the transmission medium. The years when telephonic lines were suddenly made available to the world betrayed the fact that the new medium was one which only the very rich could afford. Common people could simply not be serviced with local telephones until prices were made cheaper.

While telegraphy employed thrifty iron wire, telephony demanded the expensive and better conducting copper lines. Telephones designed by A. G. Bell did not give powerful enough signals through iron wire at any distance because of the additional resistance represented. Other problems included the fact that the Bell telephone could not transmit or receive a strong and clear vocal signal without excessive battery power. The Bell System was not a truly "democratic" medium of communications.

A mysterious and unrecorded sequence of discoveries preceded Stubblefield's early developments, but he was able to dispense with wire connections entirely. His was not a "one-wire" system. Nathan Stubblefield performed the "impossible": he developed, tested, demonstrated, and established a small, democratic telephone service which did not require wire lines at all! His system utilized the ground itself as the conductive medium.

Mr. Stubblefield discovered that telephonic signals of exceptional clarity could be both transmitted and received through the ground medium. There was simply no precedent for this development. The first effect of this wonder was that common people could now have the much needed communications which both great distance and poverty prevented. Farms could be interlinked by the Stubblefield exchange by simply plugging both terminals into the ground. Wire would not take up the expense which the telephone exchange would later charge to the customers in addition to service. Signals were loud and clear. All those who experienced this kind of telephonic conversation declared that Stubblefield had discovered a true wonder.

We have photographs of his telephone sets. These reveal small, ruggedly built wooden cases which are surmounted by conventional transmitter-receiver sets. Heavy insulated cables rim to the outer ground from this apparatus. Stubblefield developed an "annunciator" (horn loudspeaker) which amplified the voice of callers. These sets of his appeared in numerous demonstrations on the east coast, from New York to Delaware. The signals were so loud and clear that they defied commercial levels of excellence provided by the now growing monopolies of American Telephone and Western Telegraph.

Thomas Edison broke the Bell telephone monopoly when he developed the carbon button microphone for Western Union. Sounds were louder when using the Edison carbon microphone. These carbon microphones needed excessive battery power, and batteries were not cheap. Some telephone companies began utilizing dynamo systems to power their lines. The fuel needs of dynamos drive customer costs much too far, prohibiting the ordinary people from having the system for their own moments of need.

Stubblefield developed remarkably unexpected systems by which available resources became the commodity. In the early Stubblefield system, twin terminals into the ground formed the initial bridge among telephones. System users were effectively joined together through the ground itself: wires were eliminated! The signals were exceptional, and did not fade or intensify with rain, a fact later to be considered in theoretical discussions. Those who experienced speech through the Stubblefield system each reported similar impressions.

Ordinary soil conduction telephonics require a certain degree of ground water for their operation. Stubblefield's system did not operate on this principle for reasons which will become more obvious as we continue recounting his tale.

Stubblefield developed a means by which calls could be individualized among customers. Later, his central telephone exchange included power-amplifying relays, set in the ground at specific distances. Calls were handled by an elaborate system of two-wire, ground connected automatic switches and relays. Signal purity was remarkable for the time, using a single carbon button for both transmission and reception.

Furthermore, Stubblefield's telephones could be left on for days without weakening the power system at all. Now hundreds of ordinary people in widely separated places could afford the [12] installation of telephone service.

The theoretical reasons why ground conduction telephony can occur had later been established by researchers in England, notably Sir William Preece. Preece successfully attempted only telegraphic signals across great distances of land and sea. Stubblefield was telephoning through greater distances with the legendary clarity and strength which became equal to his other mysterious developments.

The true difference between the Stubblefield system and these early "conduction telegraphy" systems becomes obvious as soon as we delve further into his biography. How were ground plugged relays acting as amplifiers in the Stubblefield system? This feature forms the core of an intensive investigation which several have reproduced in various forms today.

His discovery of an earth charging phenomenon permitted the development of an equally astounding invention, the Stubblefield earth battery. This device, an earthed electrode, drew up enough natural electric charge from the earth to operate motors, pumps, arc lamps, and his ground telephone system. The implementation of his earth energy technology would have changed the nature of American Society were it permitted free market expression in its day.

MEANDERING WAYS

As telephony gradually replaced the telegraph service, lines were also accommodated to telephony. Before becoming entirely reclusive, Mr. Stubblefield befriended a few employees of the telephone service. These friends obtained cast-off telephone equipment and parts for his experiments.

He became very familiar with the behavior of telephone exchange equipment in the natural environment. The telephonic systems of existing service companies were grounded systems. Each end of both telegraphic and telephonic lines were sunk into the ground, while the single expensive copper line formed the communications link. Ground sites terminated specific lengths of these service lines in special, thick metal plates. Plates were well-buried in selected ground. These plates were composed either of zinc or copper, and required specific ground placement for their continued operation.

Linesmen were taught to find "good ground" for these sites. Some later insensitivity among the growing numbers of hired crew members required the development of "ground location meters"; none of which were to give the special and anomalous characteristics observed in early linework.

Certain telephonic patents reveal extremely articulated" termination plates for these service lines: folded, stacked, coiled, and interleaved. Acting as accumulators of earth energies, these often became extremely charged. It was found that signals would both self-clarify and self-amplify to unexpected degrees when these special terminations were employed.

Telegraph linesman "felt" their way through woods, laying the paths for lines according to a peculiar intuition. Theirs was an intuitive path rather than a strictly mathematical one, carved through woods and vales in the artistic meandering way of the ancient "geomancers". Older linesmen recalled the days when line installations took their winding routes through woods, across meadows, and sinuously along ridges, lakes, and streams. Linesmen innately sensed the most favorable paths along which lines should pass.

Geomancy is the ancient qualitative science by which "holy spots" are discerned, and sacred edifices are properly founded. Intuitive discernment, rather than mathematical objectivity, governs the geomancer's aesthetic. Geomantic aesthetics governed the building of ancient villages and towns, and it is no wonder that most architects of any real artistic worth exercise these same aesthetics. Art-governed architects are natural geomancers.

Geomantic qualitative science precedes geologic quantitative science. A surveyor might simply draw a straight line across a section of land, and engineers would then employ powerful means to cut that straight path despite all natural barriers. Much of modern housing development is based on this "draw and cut" method. The sharp paths of engineers is effective and direct, but the old meandering rural roads dotted with their naturally placed homes are beautiful.

The old linesmen trekked across woods in a careful manner, turning aside from natural barricades. Maps of these first telegraphic lines may be consulted. It will be seen that these lines meandered with natural features. Telegraphic lines twisted and turned through the countryside and wilds; a twisting vine of iron on tar-covered wooden poles.

ELECTRICAL OCEAN

Properly ground-conformed telegraph lines were known to produce unexpected signal strengths, as well as unexpected signals. Night station operators were often "haunted" by spurious messages. These contained fragmentary words and sentences, and could not be traced to other station operators.

It is curious that the older placed lines demonstrated a remarkable and constant feature throughout their years, requiring few, or no batteries for their operation. This absolutely astounding fact is well documented in the telegraph and telephone literature of the day.

In these trade journals we find reports of lines in which current was everflowing! Company owners found this fascinating natural fact quite lucrative as well as surprising. The question was, where is the current coming from? The echo of the linesman resounded in the forest, the answer singing beneath his feet. Another equally remarkable fact involved the engineer's methodically driven lines across land and through mountains. These lines did not manifest electrical self-excitations. Clearly, the difference of methods had produced completely opposed energetic results; the one active and the other inert.

As companies expanded across greater regions of ground, engineers replaced the oldtime lineman's sense of "proper placement" with surveyor's charts. It is not unusual for corporate expansions to bring about such a dramatic loss of quality — in exchange for a growth of quantity.

MrStubblefield pondered the question of "taking current from the ground". He stated in very plain language that the earth was filled with "an electrical ocean". This electrical ocean was surging with huge "electrical waves" which could be felt in certain places. No doubt, he was one who felt the ground energy.

Telegraph lines were once two-wire lines: each completing the circuit among station receivers, batteries, and keys. It was quickly [13] discovered that single lines, terminated in the ground with heavy metal plates, could achieve similar results. The immense savings in wire, poles, insulators, and maintenance was an attractive feature of this method for company owners.

Telegraph linesmen required skill in finding proper ground terminals. Improperly placed ground plates could ruin a system by not conducting signals properly through time. Spurious conductivity in a line could ruin critical transmissions and receptions. Telegraphic lore is filled with discussions about both "good ground" and "bad ground".

The linesmen were workers in a yet primarily agrarian society, having experience with soil and earth in general. Many of them were farmboys who watched the oldtimers "divining" for water. Linesmen frequently discussed such natural means for discerning the "good ground" for terminating a telegraph line.

Thomas Edison adopted a method which could "valve" line signal by preventing unnecessary signal leakages into the ground. His method included the use of terminal rheostats in order to control the amount of current flowing to the ground during signal time. Several of these terminal rheostat patents have been found. One familiar model uses a thick cylinder of carbon with a slide spring contact.

The most amazing thing to the telegraph linesmen was the variation of rheostatic settings which each ground required before strong signalling could occur. Some terminal rheostats needed to be closed completely. Others could be opened full until signals were of sufficient strength to operate the system. Each ground had its own "character". Each ground was possessed of activities which defied description except but by poetry, song, and twisting green vine.

Telegraph line was not copper, neither was it insulated throughout its length. Telegraph line was bare iron wire, and was supported on porcelain insulators fixed to tarred wooden poles. Signal strength along such resistive wire would have theoretically been extremely poor, but was exceedingly strong at times. So great was the developed signal strength that operators could "remove battery cups" and work with almost no current at all. Where did this extra energy come from? From what mysterious depths did this strange power emerge? Examination of telegraph systems reveals them to be radionic tuners on a vast scale. I suggest that VRIL articulate energy, the dendritic living energy found in the ground, was at work in all these systems.

EARTH RESERVOIR

Nathan Stubblefield's experiments involved the development of earth batteries: buried metallic arrangements which produced electrical power. We find a good number of the earth battery designs in the Patent Registry. The earliest designs appeared in 1841 when Alexander Bain discovered the phenomenon. While working a telegraphic line, he chanced to discover that his leads had become immersed in water. This short-circuit through earthed water did not stop the actions which resounded through his system.

Mr. Bain took the next step to a greater distance, burying copper plates and zinc plates with a mile of ground between them. These, when connected to a telegraphic line performed remarkably well without any other battery assistance. Bain obtained a patent for his earth battery in 1841, using it to drive telegraph systems and clocks.

Stephen Vail (1837) observed the same effect, not knowing what caused it. The establishment of the first functional telegraph line seemed to require ever few batteries with time. Vail began with some 12 battery cups, reducing them gradually until 2 were needed. There came a point where he found it possible to remove even these, while operating the system.

This mystery persisted for years. I have heard such an account by a close friend and electrical engineer who reported that local telegraph stations remained in operation despite the fact that their batteries had not been recharged for a great number of years (W. Lehr).

J. W. Wilkins in England (1845) corroborated findings made by Bain, developing a similar earth battery for use in telegraphic service. An early English Patent appears in 1864 by John Haworth, the first true composite earth battery. This battery is drum shaped, having numerous solid discs mounted on an insulative axis, end braced, and buried. Their power was rated in terms of disc diameter and telegraph line distance: 1 foot diameter discs for 75 miles of line, 2 foot discs for up to 440 miles of line.

Patent Archives have revealed a great number of these devices including several remarkable operative descriptions. Earth batteries by Garratt (1868), Edard (1877), Mellon (1889), and Hicks (1890) yield therapeutic powers. Earth batteries by Bryan (1875), Cerpaux

(1876), Bear (1877), Dieckmann (1885), Drawbaugh (1879), Snow (1874), Spaulding (1885), and Stubblefield (1898) produce usable power. In addition to these marvelous patents, there are those which found their way into telephonic service later: designs by Strong (1880), Brown (1881), Tomkins (1881), Lockwood (1881) provided power assistance and primary power for telephonic systems through out the countryside.

The well reputed fame of "earth batteries" centers around their very anomalous electrical behavior. The central mystery about earth batteries is that they do not corrode to the degree in which their electrical production rate theoretically demands. Exhumed earth batteries reveal little surface corrosion.

Nathan Stubblefield knew that probes (placed into various soils) reveal an amazing degree of strong electrical activity. These currents show an amazing degree of variation across any chosen plot of ground. Wet soils often reverse the expected electrical strength: weakening, instead of strengthening their appearance. Proper placements of metallic probes can produce strong currents for use.

Touching a well-grounded iron rod is a good first experiment to try in these regards. Try and find a place where power leakage into the ground is minimal, such as a park or wooded area. Take a yard-long solid iron rod whose surface is free of shellac or insulator coatings. Carefully drive the rod into the ground with a hammer. Wetted hands on the iron should produce a mild electrical sensation. These voltages may be measured. They "pin" sensitive galvanometers. The current does not cease after several weeks of activity when properly placed.

Stubblefield's observations of natural electrical manifestations led him to consider the taking of "free" electrical energy from the earth. His initial revelation contended that such vast amounts of energy could be used to drive the engineworks of industry. Stubblefield sensed that the ground currents were arriving as [15] electrical waves.

In Stubblefield's view the electrical waves permeate the earth. Electrical waves were treated as ocean waves, constantly surging and cresting in specific locales. As ocean waves crash against fixed shores and rocks, so electrical waves also surge and crash against underground geological features. Stubblefield reasoned that this electroactivity should be extreme in certain locales: the "rocky shores" of the electrical ocean. Just as there are rocky shores, calm beaches, and surging ocean depths, Stubblefield clearly envisioned the mysterious dark waves of the vast and unsuspected electrical ocean.

Knowing these truths, Stubblefield arranged ground rods in specific locales in order to intercept the electrical waves for power. He knew that these electrical waves would only appear in very specific places, so he did not expect to find them everywhere in abundance. Stubblefield constantly spoke of "working the ground" before power could be taken from it.

Stubblefield observed the natural tides and boundaries of the electrical ocean in and around his lovely rural hometown. Concerning this earth energy, Tesla and other investigators later developed equivalent models. Tesla charted and used the earth waves in their surging impulses. Moray also intercepted these natural impulses in the "radiant energy" generator.

Some researchers believed that the vast electrical ground reservoir finds its source in the enormous solar efflux. Certainly daytime grounds yield a remarkable amount of static. Ground terminal shortwave reception is excessively "choked" during the daylight hours on certain bands. Despite the supposed insulative qualities of the atmosphere the solar efflux finds its way through space, eventually permeating the ground. Some researchers have referred to the ground-permeating solar energy as the "slow solar discharge".

The "slow discharge" represents the enormous drift of aether through the entire body of ground. The earth evidences a constantly self regenerating charge. Tesla opposed the notion that this potent field was the result of decaying radioactive materials deep in the crust.

Numerous other researchers would refer to this "electrical ocean" as the vast reservoir of untapped natural energy. Somehow this reservoir is regenerated in a constant swelling. Where did the energy come from? Earth static was presumed by Tesla to be a solar activity which manifested in and across the ground. The ever growing static of earth was problematic for physicists who could not see the source for such energetic growth potentials. Tesla believed that ultrafine corpuscles from the sun permeated the entire earth, manifesting as static charge, and further conjectured that these rays came primarily from the sun, since it was ejecting matter "at excessively high voltages". If this were so, reasoned Tesla, then sunlight contained something of this electroactive component, and it was certainly possible to derive electrical energy from sunlight.

Nikola Tesla announced these facts in 1894, finding only the silencing ridicule of academicians already hating his very name. When Tesla declared that "rays from space" were "bombarding the earth" he was absolutely rejected by the academic club who rejected these claims as "superstitious". Upbraiding his findings, they later claimed for themselves the very same discovery (Millikan, 1932).

Tesla stated that the electrical energy released by the sun is a far greater, more permeating supply than sunlight itself. He certainly believed it should be considered as a first rate natural electrical source of enormous potential for commercial applications. His assertion was based on experimentally verified facts when, measuring steadily growing charge states in vacuum tubes, it occurred to him that earth charge was sourced in solar activity.

Tesla also demonstrated the extraction of free electrical power from solar energy. A well grounded mica capacitor is surmounted by a highly polished zinc plate. This plate may be poised in a highly evacuated glass container to best advantage, the zinc not exposed to corrosive influences. The tube is elevated and exposed to sunlight. The mica capacitor is connected in series with the vacuum tube. After only several minutes of exposure time, the stored electrical energy is formidable: producing a powerful white arc discharge.

Tesla patented this device. Since earth absorbs the permeating solar efflux, then these energies can be extracted for aeons.

Others have viewed the generation of ground static as a natural "radiant process" from the ground itself. Static charge appears as the inert by-product of the mysterious VRIL, the self magnifying organismic ground energy. VRIL, according to medieval mystical philosophers, is the ground of being from which all material manifestations emerge. VRIL connectively fuses metaphysical realities (dream, vision, ideation, impulse) with physical realities. (mineralogy, botany, zoology). By the radionic method by which telegraph lines may be locally "tuned" we may well surmise that this ground based regenerative supply is the true source of static.

Samuel Morse originally planned the burial of telegraphic lines between cities. Having done so across several tens of miles at great expense and through great labor, Morse found his system utterly incapable of operation. Static had so flooded his receivers that no signalling was possible at all. This first bad experience with the static of ground presented such a discouragement, that he almost stopped the entire plan. The uneconomical task of elevating all his cables later became the normal procedure.

Early telegraphers observed a steady growth of static throughout night seasons. This growth continued despite the absence of winds or storm conditions anywhere along the line. Researchers have often referred to this kind of power as "free energy", meaning that the power source is extraterrestrial and natural in supply. Such an energy source would remain cost-free. The privatization of utility companies could conceivably be municipal and democratic; municipal groups could share the cost of installing the ground energy stations.

ENERGY RECEIVER

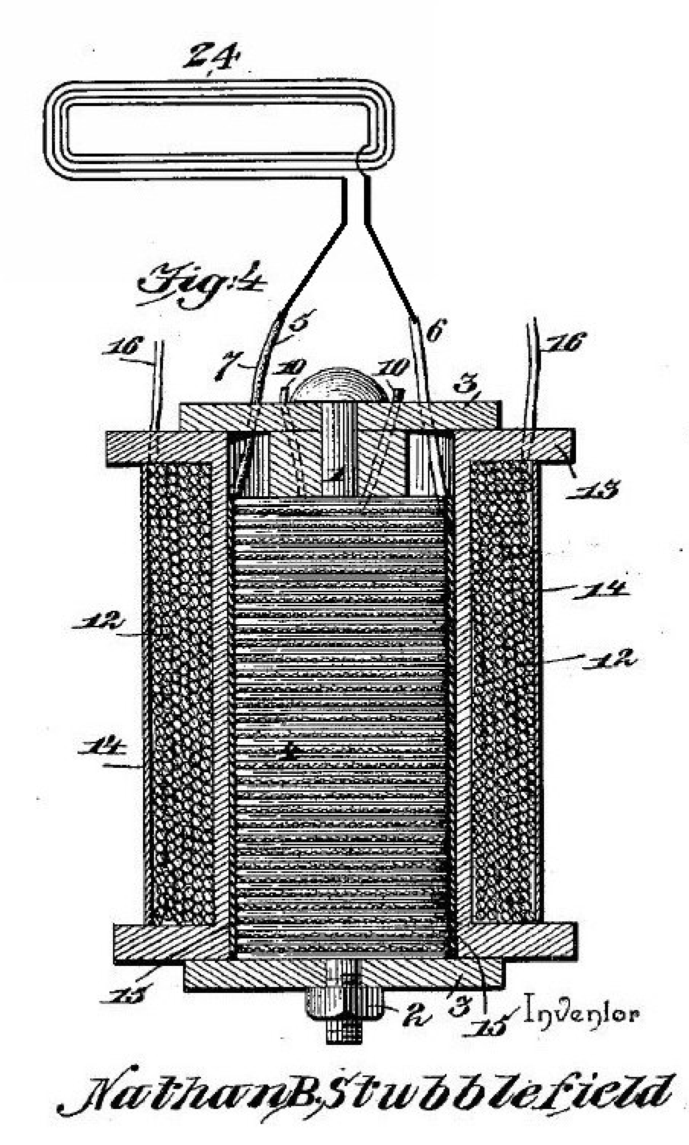

U.S. Patent No. 600,457 - N. B. Stubblefield, Electrical Battery - Patented Mar. 8, 1898.

Mr. Stubblefield developed a peculiar bi-metallic induction coil which could (when buried) draw up sufficient electrical power to operate lamps and other appliances which he designed and tested. The patent specification describes a terminal which draws electricity out of the ground. This device required very specific placement — it would not work with equal effectiveness in all locations. A very precise placement of the device required an equally precise knowledge, and Stubblefield shared this knowledge with only a few of his associates.

I spoke with an academician who had the extreme privilege of speaking with Mr. Stubblefield's son, Bernard Stubblefield. Bernard, by this time himself quite aged, told that his father's method in locating the "right spot" was deliberate. His father referred to the device as indeed a receptive terminal and not a battery.

Despite the insistence of Patent Officers in calling the device a "battery", Stubblefield declared it to be an energy receiver, a receptive cell for intercepting electrical ground waves. Its conductive ability somehow absorbs and directs enormous volumes of electrical energy. With this energy Nathan Stubblefield operated a score of arc lamps at full brightness for twenty four hours a day.

It becomes apparent that Mr. Stubblefield had witnessed (or experienced) some natural occurrence of discharging electrical energy in a telephonic system, and had determined the mode of its manifestation with simple means. His ground energy receiver (Pat. 600,457) remains a true electrician's mystery. There was a definite trigger by which this energy was stimulated and maintained. The induction coil which bears his name is equipped with three coils which are wrapped around upon a heavy iron core. Bare iron wire and cotton covered copper wire are wrapped side by side, comprising a primary coil body. Each layer of this primary coil body is covered by a band of cotton insulation, bringing four wire leads to the coil terminus. Two leads of iron and two of copper are external to the coil. Commercial-electrical power is obtained through these connective terminals.

In addition to this bimetallic winding, there is a third winding: the "secondary". This third coil is insulated from the primary bimetallic coil, serving as a trigger device. Presumably, a stimulating impulse shock was introduced into the tertiary coil, after which the upwelling electrical ground response brought forth powerful currents in both iron and copper coils.

Electrolytically (as a battery in acid or saltwater) the Stubblefield coil is disappointing, producing less than one volt according to those who have duplicated its construction. Stubblefield's bimetallic coil was a "plug": a receiver which intercepts the vast and free electrical reservoir of the ground itself. His patent and subsequent company brochures define the manner in which his earth battery was to be activated.

Technically, the Stubblefield device is a modified thermocouple (a bimetal in tight surface contact) but could not supply the degree of power which he reported. While this arrangement could develop a few milliwatts of power in appropriately hot ground spots, the thermoelectric explanation of the device cannot explain the phenomenal output reported in news reports of Stubblefield's demonstrations.

Furthermore, though the Stubblefield power receiver is wound like an induction coil, it produces a steady direct current output. This poses additional problems for the conventional engineers. Electrical induction only occurs with electrical alternations, oscillations, and impulses. Witnesses described ground-powered motors which ran unceasingly and unattended for months without need for replacing or replenishing the ground battery. Small machinery, clocks, and loud gongs were run by other ground-buried cells as reported by credible witnesses.

Mr. Stubblefield reported that the burial of his "earth energy cell" required time to build up charge. Once the cell was saturated, however, the cell became a conduit of earth charge and flowed over in "commercial electrical volumes". He did not claim complete knowledge of the phenomenon. He simply stated that (once the coil became saturated with the earth charge) it suddenly manifested an electromotive force "far greater than any known wet cell reaching into weeks and months of continuous work night and day" and poured this charge out for use.

Stubblefield used the cell as a "plug", drawing out the electrical charge of the ground. The cell coilings acted as a lumped conductor. Charge saturated this conductor and flowed up into it, powering any electrically connected appliance. After repeated exhumations, the copper element of these cells "is not acted on in a perceptible degree . . . even after repeated renewals".

Mr. Stubblefield described means by which such cells could be connected in series at short distances from one another. "With these, acting as electrodes . . . you draw from the electrical energy of the earth a constant E.M.F. of commercial value". The phrase "acting as electrodes" is the heart of the Stubblefield energy cell. It is not a battery. It absorbs and flows over with the stupendous energy of the earth's charge. Stubblefield may have discovered the auto-magnifying voltage effect of electrostatic induction in coils before Tesla, who later utilized the effect in his special electrostatic transformers.

Stubblefield's buried bifilar coils may have become saturated with earth electrostatic energy, travelling up the coil. In such a case, the mere battery power of the coil was replaced by the electrostatic flow, the coil acting as an electrode. This seemed obvious when considering the fact that its ordinary battery current (1 watt) was gradually replaced by a continually growing electrical current of far greater proportion.

TREE ROOTS

Experimenters have observed the "slow accumulation and creep" of current up through vertically buried coils and large solid rods. This current has growth characteristics which gains strength with lengthened burial time. Buried coils and rods do not give their full output until they have "developed" power over a few hours of time.

This behavior resembles nothing like a true electrical current. The best model to explain the phenomenon is vegetative growth — a biological expression. Only a full scale test of the reconstructed Stubblefield device in proper grounds will give conclusive and satisfying answers.

Witnesses convey that Mr. Stubblefield's batteries were usually buried at the roots of certain very old oak trees. From these sites it was possible for him to bring small arc lamps to their full candlepower. Tremendous amounts of energy are required for this expenditure of power. Not only was he remarkably able to draw such volumes of current from the ground reservoir for lamp lighting, but the power was available to him throughout the day. Arc lamps were hung in the trees themselves with their receiver coils buried in the roots. Such was the nature of this current that the lamps did not heat excessively, and seemed to burn forever.

Nathan was not replacing his lamps with the frequency demanded by such continuous operation. Obtained through his employment with the telephone company, he was able to recharge old wetcell [17] batteries with energy from these buried receivers for other experiment. Certain conventional thinkers claimed that the Stubblefield simply used wetcell power for his telephones. Later demonstrations indicate the fundamental error of this conventional view. Stubblefield ran most of his apparatus nonstop for days — without turning off the power.

It is more than likely that charged wetcells were used to "jump start" the ground electrode during certain seasons, since the patent reveals that an outer third coil could be added to the copper-iron bimetal. We do not know the secrets of the earth charge as Nathan determined. Others since this time have observed fluctuations at certain times of the year in ground energy. It may be that a sudden induction is required before the excess ground charge surges to the surface — like priming a pump.

The arc lamps could have been low pressure gas arc-lamps of the kind demonstrated by Daniel MacFarland-Moore; but these required high voltages. Nathan did not utilize such excessive voltages. Another paradox deals with the notion that Stubblefield simply connected hundreds of his small-wattage batteries together, producing a large and commercial output. Nowhere is this evidenced. Nathan showed that one or two such batteries were sufficient to draw off "the charge of the earth" . . . a very different kind of energy.

When properly placed, the weak power of the Stubblefield "battery" becomes an electrode for the powerful earth charge. But arc lighting and battery charging was not his only specialty; there were other marvels which he began developing in methodic succession. His bimetallic coil receiver intercepted electrical waves and produced enormous power outputs which could be modulated: superimposed with additional signals, sounds, and voices.

GROUND RADIO

Salva (1795) suggested several electrical schemes for long-distance, and even an aquatic telegraphy. He suggested that physiophonic telegraphy be the communications mode; where human recipients would receive the mild shocks of a distant signal station, and so convey messages.

Salva also believed that earthquakes had subterranean electric origins. Working on the hypothesis that subterranean electricity caused violent communications under vast earth strata, Salva suggested that ground and water be used to replace wires for electrical signalling.

Sommerring (1811) first attempted telegraphic transmissions through water-filled wooden tubs. The signals were effectively passed as if through wire conductors, the thought of wireless ground resulting.

James Lindsay (1830) first developed the notion of utilizing artificially generated electricity for special modes of lighting, motor-power, and communication. Mr. Lindsay suggested that submarine cables might be laid between land masses while using "earth batteries and bare wires" as the means for power transfer.

Steinheil (1838) demonstrated the remarkable passage of signals along one-wire to the ground. When trying to use earth as the "second line" he measured large currents. This complete success proved the great conductivity of ground; and so the "earth circuit" was born, liberating telegraph systems from the expense of using the two-wire system.

Morse (1842) sent telegraph signals across a river. Antonio Meucci (1852) had already demonstrated the transmission of vocal signals through seawater, but traversing the ground represents a different thing altogether. Mr. Stubblefield reasoned that, since electrical waves traverse the whole earth, it might be possible to send signals to distant places. These ground-permeating natural electrical waves might serve as carriers for the human voice. The ground would act as both power generator and signal conductor. Like a gale carrying messages downwind, these electrical waves could bring wireless communications instantly to any part of the world.

To this end, Mr. Stubblefield experimented with the buried power receiver and a system of telephone sets. He found it possible to send vocal signals through the ground to a distant receiver, referring to this system as a "ground telephone". Telephoning through the ground became routine for this remarkable man. Signals sent through the Stubblefield method were notable for their reported "great clarity". What is strange about this system is its elegant simplicity: Stubblefield's transmitting system evidences an almost crude minimalism which offends some researchers, while surprising others.

Numerous private and public demonstrations of this first system were made in Murray, Kentucky (1886-1892), where the mysterious "black boxes" were seen. Two metal rods were stuck into the ground a few feet apart from each distantly placed set. Speech between the two sets was clear, loud, and startling despite distances of 3500 to 6000 feet. These transmissions were made through the ground itself and used the Stubblefield cell for power.

As mentioned, in several photographs we see special loudspeaking telephones outfitted with long (1 foot) horns, designed to act as annunciators. Calls from these annunciators brought his son Bernard to the telephone transmitter. The system was never switched off. Power was limitless and did not diminish with time of day or length of use.

While Marconi and others were barely managing the transmission of telegraph signals for equivalent distances, Nathan Stubblefield was transmitting vocal dialogue. The clarity of these signals and their sheer volume was the most widely recognized feature of the Stubblefield system. He was developing the system to operate through far greater distances, using automatic relays to boost signals for very great distances.

He published an extraordinary brochure in 1898 to attract investors who had expressed interest in consolidating a small corporation. In this brochure, Stubblefield insisted that power for his device was not generated in the cell. He calmly stated that the cell received energy from the earth. In a less discussed portion of this brochure, Stubblefield stated that electrotherapeutic potentials were derived from the earth battery. Other researchers made similar claims for their earth batteries (Hicks, Mellon).

In 1902, Stubblefield set up one of his sets in a "Mainstreet" upper office — in a hardware shop. From that point to his farm (some 6000 feet distant) he conducted continuous conversations with his son Bernard. Tapping with a pencil on his one-piece transceiver, Bernard was quickly heard in aloud, very clear voice. This transceiver was a carbon button placed in a tin snuff box. Speech and response were transacted through the self-same device, [18] which acted as both microphone and loudspeaker. Cells were placed downstairs from the office in the ground. They were never removed and never wore out, though operating twenty four hours around the clock.

Nathan Stubblefield offered to construct a large scale power station for the town of Murray. His quoted initial installation cost were estimated at five thousand dollars. The town politicians declined the offer. Now, the technique of drawing up electricity from the earth remains a mystery.

Nathan Stubblefield demonstration with wireless telephone, 1908.

STATIONS

The Stubblefield ground radio system was demonstrated for approximately one thousand Murray residents (January 1902). Photographs of Stubblefield and his family, and a good crowd of witnesses from town show the cell lying on the ground among all his assembled inventions; a flower-pot sized coil of good volume. Other devices show motors and large capacitor stacks for aerial voice transmission experiments.

After the successful completion of these preliminary tests, Stubblefield travelled to Washington, D.C. for a public demonstration which was to be one of his crowning public achievements (March 1902). Stubblefield sent wireless messages from a steamship to stations on the shores of Georgetown. This successful test employed trail-wires in the river water. While successful, witnesses acknowledged that Stubblefield's ground telephony sounded louder and came through with greater clarity. Photographs of this event are all available.

He declared that news, weather, and other announcements could be broadcast through the ground across a great territory for private reception. Simultaneous messages and news of all kinds would soon be transmitted through the ground from a central distribution station.

He also stated that (while such broadcasts required wide transmissions) he was developing a means by which privacy of ground telephonic messages could be maintained among callers. This method of individuation would take place through the ground, insuring that no one could eavesdrop on anyone's conversations.

The Washington D.C. demonstrations were followed by a trip further north. Mr. Stubblefield took his apparatus to New York City for additional tests, preparing for a public demonstration in Manhattan's Central Park. The demonstration was to take place in less than twenty four hours after his arrival.

Stubblefield found (to his shock) that the ground was not permeable to earth charge in all places, and not conducive to easy ground telephony. He requested more time to discover the powerpoints before setting up the stations properly. Time to "work the stony earth" of the Park left a few investors foolishly wary of the system's worth. This demonstration was withdrawn.

His next public expositions were given in Philadelphia's Fairmont Park with greater success (May 1902). He now recognized, more than ever before, the role of geologic formations in determining and establishing stations.

Stubblefield published a prospectus for his WTCA (Wireless Telephone Company of America), stating that "I can telephone without wires a mile or more now, and when the more powerful apparatus I am working on is finished and combined with further developments, the distance will be unlimited". He sold only one telephonic system to another corporation: the Gordon Telephone Company of Charleston. This system was used to communicate with offshore islands. It would be interesting to retrieve this system and examine its contents.

He entered these commercial aspects with some trepidation. By June of the same year he withdrew from the project completely. A few persons managed to discover the reason for his quiet, sudden retreat. Because of his difficulty in instantly stationing his system in New York City, it was suggested that he adopt the method of burying lines to "fake" the operation — if just for the purpose of making a good show. Nathan declined.

After witnessing these demonstrations, another inventor (A. F. Collins) duplicated some of Nathan's inventions and filed a counter-patent for a ground telephonic system (patent 814,942 for "Wireless Telephony", 1906). One of the signing witnesses on the Collins patent was one Walton Harrison.

Harrison, a WTCA member later infringed on another Stubblefield experiment with his "Transmitter for Wireless Communication", a telegraphic-telephonic system (patent 1,119,952, 1914). It became apparent that certain WTCA members were trying to oust Stubblefield himself!

The WTCA now took on a life of its own. Stubblefield was thoroughly disgusted at the display of human greed and ambition, and left them to their own devising. Collins and others were later accused of petty crimes having to do with mail fraud, and the WTCA failed in time. Internal disputes over money, rather than technological progress and implementation, was their own death knell.

Marconi arrived with an inferior (though highly publicized) system. When Marconi began his work, the effective signal transmission distance was equal to that achieved by Stubblefield. Stubblefield was experimenting with ground radio since 1888, but did not patent his developments until much later. Credible witnesses saw his ground radio experiments in action during this time frame, establishing the historical priority of Stubblefield — a true and original American genius.

While Marconi could barely send telegraphic "dot and dash" signals with great difficulty (and static), Nathan had already transmitted the human voice with loud, velvet clarity. Others would adopt and implement the Collins system (Fessenden, DeForest, Bethenod, Braun).

Nikola Tesla performed double ground experiments with impulses as early as 1892, reporting these in lectures and patenting some embodiments in 1901. No one of these later systems ever achieved the same results of clarity, tone, and volume of Stubblefield ground telephony. Priority in this art belongs to Nathan Stubblefield. In addition, his was the only system in which natural energies were obtained, magnified, and entirely employed as the empowering source. All the other inventors used "artificial" sources (batteries, alternators, dynamos).

Following all these ground radio demonstrations, Stubblefield researched "magnetic waves" and developed several systems which did not use ground terminals for exchanging signals. Long distance wireless telephone communications were his aim. Many imagined this to be radio as we know it, but several features of the Stubblefield aerial system are distinctive and different. First, his transmitters and receivers were telephonic, [19] not telegraphic. In his preliminary experiments, the earth battery was used to energize an apparatus to which was connected a long horizontal aerial line. Marconi later adopted this "bent L" symmetry in conjunction with a grounded copper conduction screen. We do not have photographs of these arrays, but have handwritten manuscript copies of certain diary notes in which a progressively greater telephonic distance is reported. Nathan made steady progress in this form of telephonic transmission, but used neither alternators or spark discharge.

A second series of experiments reveal the development of stacked capacitors. Photographs reveal two large capacitor stacks, presumably for inductive transmission purposes. Some researchers induced ground oscillations of electrical current, while absorbing each "flyback" into large capacitors. This system evidenced the "hydraulic" model of electricity, popular during the latter Victorian Epoch.

Tesla would later show the essential differences between current species developed in various electrical machines. Each of these, when separated, evidenced entirely unsuspected new phenomena. It is most probable that Stubblefield was one of the first individuals to discover some of these strange effects, even before Tesla himself.

Direct current impulses have very different characteristics than alternating high frequency currents, used by Marconi. Vocal modulations of strong direct currents release polarized impulses. Impulses do not produce the inductive waves used in conventional radio transmission: they produce inductive rays which travel in straight lines. These inductive rays are penetrating and more like electrostatic energy than electromagnetic energy.

Photographs reveal a final form of Stubblefield's aerial telephone which utilizes a two foot in diameter single turn copper band. This outer copper band is spaced from a second inner copper band, and is mounted on wooden pedestals. A telephone is connected to this array. This compact apparatus transmitted inductive rays for great distances when earth energy was modulated by the human voice.

A truly honest and humble man, he justly considered the ambitious and aggressive (northern) investors as "scalawags and damned rascals". He became suspicious of others. Considering the time frame in which these events took place, we may understand his reaction. Rejecting their tempting swindle, he was compelled to leave for home in order to continue his beloved experiments in privacy. He became mysteriously compulsive about his privacy after this.

In the words of several persons with which I have had the good fortune to speak, "Nathan was honest to a fault". He, disappointed again in human behavior, packed away his equipment and went home. After this unfortunate time period, Mr. Stubblefield preferred to be alone. Some say he became increasingly intolerable to be with. These patterns mark the genius, the dreamer. Those who walk in the future, while being in the here and now. Finally, his wife left him.

HOMESTEAD

As visitors approached the Stubblefield farm, yet a good way off, Stubblefield would appear at the door to wave them away. This often occurred when they were simply too far away to be visually located. He refused to speak to anyone for long periods. This occurrence was reported during the night, when visibility from the cabin to the distant parts of his fields would be impossible. Nathan would always appear at the door, somehow knowingly, and wave them off.

Pranking schoolboys, intent on stealing vegetables or fruits, would ever so secretly crawl onto his farm quite out of possible sight. Nathan would always be right next to them laughing in no time, somehow mysteriously detecting their presence. In a later embodiment, bells would sound when anyone approached so much as a half-mile from his cabin. It has been suggested that he had developed a device which could actually indicate the positions of any intruder across a space of ground.

Some declared that Nathan, jealous of his privacy, rigged the whole farm with delicate trip wires in order to locate and surprise pranksters. Sometimes the intruders would be met by Stubblefield, waiting at the very spot where they were stealthily heading. No intruder ever managed to feel or find these supposed wires. Others would say that Nathan buried sound-sensors all over the farm. These, when pressed, could model a trace across a map of the farm inside the cabin. Each sensor, tied to an indicator, could show up on the map, and studying this map, he could see where intruders were on the fields. Nathan could then gleefully sneak up on them and chase them away. This tantalizing mystery has never been fully explored.

Methods of distant ranging and location were devised by Antonio Meucci, employing tone signals. These required receivers, however, at the distant end. But Nathan knew where the intruders were coming from and where they were going as well. Nathan may have developed ground-wireless relays which responded to ground-buried sensors. These may have transmitted a tonal signal to the cabin, where a receiver would he triggered. This receiver may have been the bell-sounding mechanism. How did he locate people with pin point. accuracy however? No complex array of detectors was ever found in his cabin when he died.

In light of all his experimentation with earth energy and wireless, we will assume that his last two mysterious inventions speak of utterly new and unknown (though related) ground energy phenomena. But, what natural phenomenon permitted him to achieve this feat?

Ocean waves often contour the shoreline, evidencing something of the shore outlines to distant places. Electrical waves might conceivably do this. But how would Nathan model this inside his cabin? No such map was ever really found. Also, if he were using some kind of ground impulse Doppler radar or sonar (electrical ground impulses outward) then what feature beneath the approaching intruders would signal an echo back to the receiver?

Some have suggested that Stubblefield was utilizing distant variable ground conductivity. Intruders would alter this by their weight and step. How would such a signal be transferred back to the measuring station? Such reciprocation in ground currents would require that they are irritable and sensitive. This would evidence an unsuspected permeating biological nature in geology; a song, a personality with which the old linesmen~dowsers were intimate.

MOTORS

A motor, designed by Stubblefield to operate entirely by [20] fluctuations in ground static, has been stored in a local museum. The device features several mobile pithballs around a compass-like perimeter, resembling the equally mysterious electrostatic hoop telegraphs of the 1700's. Students of Stubblefield's work have examined the pithball pendulum device and ignorantly concluded it to be useless.

Pithball (static) telegraphs of the early 1700's reveal this Stubblefield design to be a very special "find". Pithball telegraphs utilized a grounded metal hoop, an underlying dial, and a pendulum on which a pithball (cork) was hung. A single line (sometimes of silk) connected two such arrangements.

Signals were made and received in a very curious manner with pithball hoops, an equally historic mystery. Moving the pithball to a particular letter on the dial resulted in identical displacements in the receiver: an anomaly. These arcane devices managed the articulate transaction of messages by unknown phenomena approaching intelligent transfer.

Witnesses of these signalling devices were credible persons in the scientific community. No one questioned how it was possible to articulate such a transfer with static electricity. In any event, any researcher not familiar with the designs would pass over Stubblefield's "pithball table" without counting it as worthy of study.

Another device, found in Nathan's cabin after he passed away, is of singular mystery. One person actually thought that Nathan built it because it "looked really strange" — like some science artform. It sat upon a trunk of to the side of his cabin room. Bernard Stubblefield, his son, did not recognize the device. Nathan must have built it after Bernard was taken away with his mother. Bernard did not remember seeing the device before this moment. It was taken to a local museum, where it now resides.

The device is a square arrangement, having several insulator-mounted pithballs in each quadrant of the central square table. It is quite likely that this was the means by which Nathan detected movements and positions in his field. If this analysis proves true, then it represents a major leap in his earth power technology. I have surmised that this device is the Stubblefield long-range detector. Motions in a specific pithball pendulum gave the direction and position of the intruder. Such a device relies on phenomena which are unknown in conventional science.

Natural observations in systems lead to unexpected, theory-busting discoveries. Such an effect demonstrates that an articulate quasi-intelligent energy permeates the natural environment of which electricity is a minor part. The natural phenomenon which is responsible for this ability is truly remarkable, nothing short of the miraculous. In its realm, we see that nature is suffused with an almost biological organization which includes the supposed inert world of geology. This would be equivalent to acknowledging that geological structure is suffused with a neurological sensitivity; a thing which academic science is neither prepared nor equipped to endorse. Nevertheless, different aspects of this ground sensitivity were discovered and differently implemented throughout the following years.

T. H. Moray (1935) also discovered long-range articulate tuning through the ground from a fixed single site. His "radiant energy listening device" permitted him to scan a tract of land and actually eavesdrop on distant conversations and sounds with earphones. This device did not implement a microphone: the Moray Listening Device used a grounded rod and special large germanium detector. How does a stationary tuner sweep across land and pinpoint sound sources?

Stanley Rogers (1932) discovered the same long-range scanning effect when, using a radionic tuner for mineral detection, he found it possible to sweep a field or meadow with a variable capacitor. Adjustments on these grounded tuners could sweep across land, revealing and mapping every mineral contour. Dr. Ruth Drown (1951) independently, developed a compact device which could sweep, scan, and delve through subterranean grounds for the specific purpose of ore detection. This device permitted photographic detection of ores swept through the ground, isolating specifically sought mineral deposits.

The Stubblefield pithball pendulum represents a leap in ground power technology. It is an engine which operates without electrical transformations at all: a ground-powered "auric" engine.

SUNLIGHT

Two more mysteries have lingered from this latter period of invention in the Stubblefield biography. The nature of each reveals the extent to which he had developed and advanced his new technology. Nathan continued to pursue his experiments, but little was seen of him for long time periods. Alone and tired, Nathan stopped working his farm completely.

Later visitors felt sorry for Nathan, now aged and abandoned by his wife and children. Several of the town's many charitable ladies decided to take him some food. On one occasion, they arrived at his farm to find the ground "ablaze with light . . . like pure sunlight was coming right up out of the hillside". Later investigators entered his land area and found heavy wires leading from the roots of trees. To these wires were attached small arc lamps, hung in the trees. These were long extinguished. They imagined this to be the explanation of his hillside sunlight. Their hasty analysis proved problematic from stories which witnesses report.

The warm and diffuse sunlight which came from the ground itself around his house was not localized in specific lamps. The light came from the ground, not from the trees as before "a whole hillside that would blossom with light" [. . .] "lit up like daytime". These observations indicate that Stubblefield had managed indeed the direct conversion of earth energy to light and warmth.

This would be acceptable, were Mr. Stubblefield simply working on a newer means of drawing electricity from the ground to light small arc-lamps; a feat which he had accomplished earlier. But these kind persons could never find any evidence of arc-lighting or any other form of known lighting anywhere near the area. In their own words "the light seemed to come from the ground itself". In addition to the ground sunlight effect, many heard very loud and unfamiliar noises coming from the whole area surrounding his cabin. What could this be? Had he managed to directly transduce the natural impulses of the ground energy into audio?

His own last claim, made two weeks before he passed away was made to a kind neighbor: "The past is nothing. I have perfected now the greatest invention the world has ever known . . . I have taken light from the air and earth . . . as I did sound".

SUNSET

I was the quite fortunate recipient of an unexpected personal letter while writing my original treatise on Nathan Stubblefield. It was told by a gentleman who received the account through a man who witnessed the following story:

Neighbors had not seen Nathan for several days. As they were worried about his health, they attempted to call on him. The lock was secured from the inside. It was a lonely, cold, and rainy March day when old friends and neighbors broke the lock on Nathan's cabin and entered. He had passed away in his bed, the probable victim of malnutrition and fatigue.

They all noticed that the interior of the cabin was "toasty warm", as if heated by a strong fire. Moved to locate the source of this heat, town officials found "two highly polished metal mirrors which faced each other, radiating this great heat in rippling waves".

Now this, I must say, is a truly great discovery and last mystery. It fulfills what Nathan reported in his last testimony. Nathan's deepest confidence was in those kind and compassionate people who continued to seek him out with love and concern to his last days. Abandoned by all, he wished one of his dearest neighbors to write a biography. Perhaps he wished to explain his life, an apology for all his ways. He said "I have lived fifty years ahead of everybody else". While often sounding inspirational, these are words of deepest sorrow.

To live with a vision of the future is to experience the surprising, often disappointing rejection and resistance of all who surround. The conspiracy of human nature. Some said he was incapable of loving others. It was his love which coaxed the living sunshine out of hard, rocky ground. And his love brought up the resounding waves of an eternal sea of energy. Love, like the rose, often hides within its shelter of thorns; singing lonely, windy songs to the deep. In endless dreams of night, the stars listen.

Bibliography

Corliss, William R. Lightning, Auroras, Nocturnal Lights, and Related Luminous Phenomena: A Catalog of Geophysical Anomalies. Glen Arm, MD: Sourcebook Project, 1982. Print. <http://amzn.to/1zbg4Oz>

Devereux, Paul. Earth Lights Revelation: Ufos and Mystery Lightform Phenomena: the Earth's Secret Energy Force. London: Blandford, 1989. Print. <http://amzn.to/1FE8EHW>

Hoffer, Thomas W. "Nathan B. Stubblefield and His Wireless Telephone." Journal of Broadcasting. 15.3 (1971): 317-330. Print. <http://www.researchgate.net/publication/233131745>

Fawcett, Waldon. "The Latest Advance in Wireless Telephony." Scientific American. 86:21 (May 24, 1902): 363. Print. <https://archive.org/details/scientific-american-1902-05-24>

The Stubblefield Papers, Pogue Special Collections Library, Murray State University. <http://libguides.murraystate.edu/MS84-04>

Borderlands, Vril Compendium Volume 2, "Vril Telegraphy" (project page)

BSRF, Vril Compendium Volume 4, "Earth Batteries" (project page)

BSRF, Vril Compendium Volume 5, "Meucci, Grey" (project page)

BSRF, Vril Compendium Volume 6, "Telephonic Earth Batteries" (project page)

BSRF, Vril Compendium Volume 7, "Loomis, Stubblefield, Dolbear" (project page)

My many thanks to the warm and dear people of Murray, Kentucky. My very special thanks to Mrs. Dortha Bailey and Mr. E. R. Bailey, Mrs. Baker, Ms. Alexander. Thank you Wm. Lehr!

YouTube Videos by Lasersaber

https://www.youtube.com/watch?v=ZuQGuXJ02fo

Working Nathan Stubblefield Coil!

https://www.youtube.com/watch?v=xsuw12Qr8wk

How To Build A Nathan Stubblefield Coil Part 1

https://www.youtube.com/watch?v=28zBEvu5_g4

How To Build A Nathan Stubblefield Coil Part 2

https://www.youtube.com/watch?v=tD06J9IpbAA

How To Build A Nathan Stubblefield Coil Part 3

https://www.youtube.com/watch?v=O6a4goOGrgo

How To Build A Nathan Stubblefield Coil Part 4

https://www.youtube.com/watch?v=jLL7YKO7q8c

How To Build A Nathan Stubblefield Coil Part 5

https://www.youtube.com/watch?v=lcjxA7bYUp0

Stubblefield Electromagnet Effects?

http://www.youtube.com/watch?v=kbaub2kkkpA

Magnet motor NS coil 48 days and still going!

http://www.youtube.com/watch?v=PuMoaDXyFZU

Magnet motor baby NS coil!

https://www.youtube.com/watch?v=l5XoxYAmaJQ

Magnet motor baby NS coil - part 2

---

http://www.hereticalbuilders.com/showthread.php?t=387

Stubblefield Coil

I built a 6" acrylic frame for the coil. That way I can insert different cores, or no core. I also built this winding machine yesterday. I've had the motor for some time and it has a nice reduction gear box on it. I also put a foot switch on it so I can have my hands free.

Theres no speed control, but the price was right since I salvaged the motor. The speed is fine for most winds anyway.

I suppose I should also get a counter for it, but I don't generally use one. It would be nice for the new motor though. Balancing the coils would be better than guessing.

Anyway, I decided to start with 20 AWG wire. I got 150 feet of both bare copper and steel. I have a plan to weave the cotton into the wind so I won't need cloth covered wire.

We'll see how it goes...

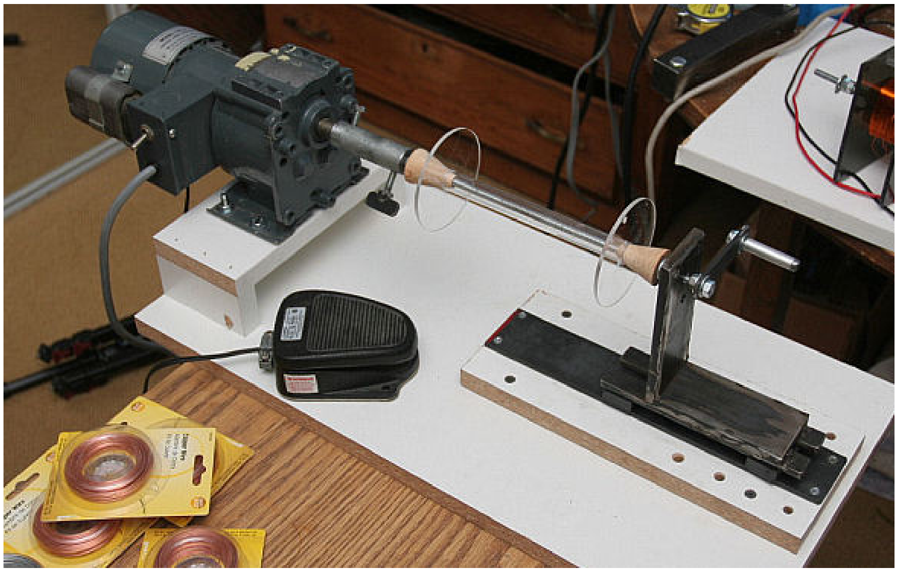

***

I got a few layers put on today. I wound one row of steel, then put a layer of cotton down. Then I wound a row of copper. I left enough space between the loops to keep them insulated from one another, and to have a space for the next winding to fit into.

Each row is a separate strand and I just connect them at the ends. This makes it much easier to wind.

I think in the end the result will be satisfactory since the copper will be right next to the steel, separated by a layer of cotton and all the layers will be in series.

I haven't taken any measurements yet except to check for shorts. I should have it done by tomorrow and then we'll see if it works.

Cheers,
Ted

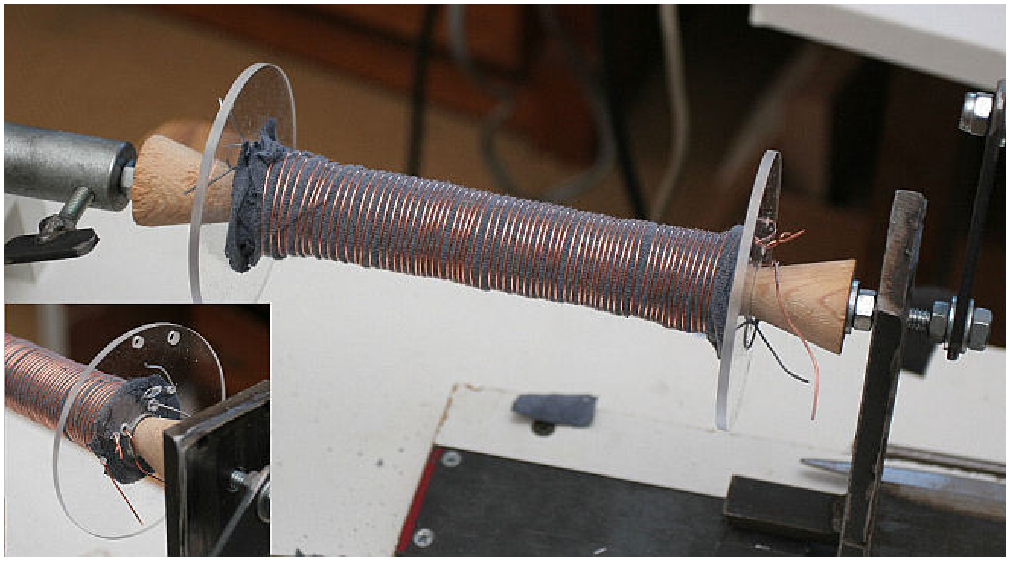

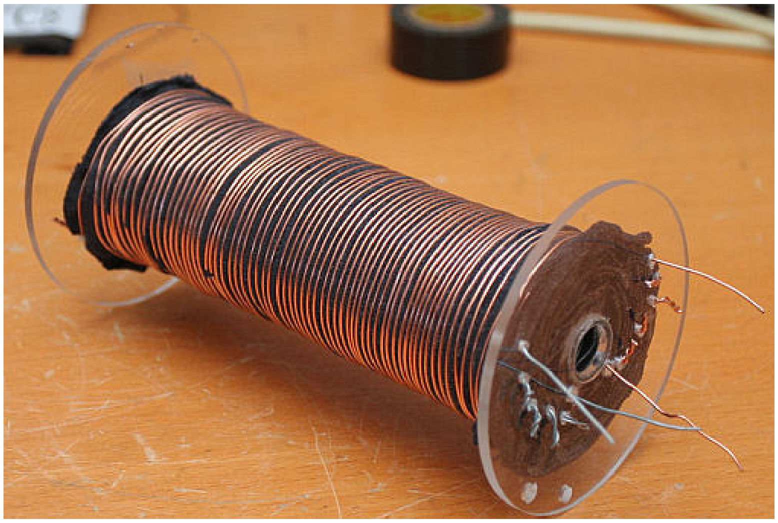

Finished it. It took forever to wind. I used about 200 feet of both copper and steel 20awg wire.

The dry voltage across the coil was .69 volts with the copper positive. Current was a minuscule 40 ua.

I sprayed it with some tap water, which got about a third of the layers wet, and the voltage went up to .75 volts. The current jumped up to 19 ma.

The current and voltage were unaffected by inserting an iron core into the coil. The iron did polarized and would deflect a compass at either end. However, I couldn't get any response from either the compass or a piece of steel wire when I shorted the coil.

I'm not used to these low power devices. If I fire one of my motor coils up on the bench, I get all kinds of metal crap flying around.

Anyway, I decided to soak it in salt water and see how it worked after that. It currently reads .82 volts at a respectable 85ma. I'll dry it out and see if those values hold up.

***

These things are weird. When I shorted the two ends of the copper together and short the two steel ends, the current between the two went up to 106ma. Instantaneous current is over 200ma, then it drops down to a steady 106.

I couldn't detect any solenoidal pull either. A steel rod partially out of the center exhibits no detectable movement when the leads are shorted, even when balanced on a very soft spring. This indicates little or no net current flow in one direction, which would be keeping with two generators (the two wires) pushing current in opposite directions. But then how do we get a polarity on the steel? A slight imbalance maybe?

I'm not really sure anymore how Laser's motor is running either. Perhaps there's some inductance from the spinning magnets that dictates polarity? I'll have to think about it.

***

Continuing along with the weird theme... I shorted all the separate windings together, both top and bottom... and the current increased. You can see when it's in the vertical position the current is around 119ma. When I tipped it over to show the bottom shorts, it gave me a few more ma's.

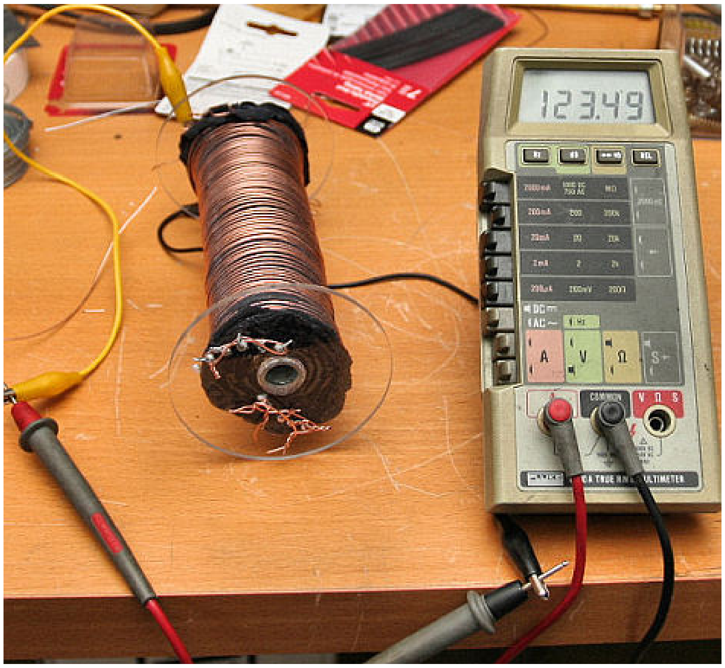

***

I don't know whether to consider this a battery or a self filling capacitor (is there a difference?). I doesn't act like a coil though. I'll bet you could get the same results with two sheets of metal coiled up like a capacitor with some cotton in between. That's essentially what I have here. I may try that next...

***

I went to check on it this morning and I think it died overnight. It read only .2 volts and the current was a paltry 10ma.
I stuck it outside in the sun to warm it up. Perhaps that will revive it. Maybe it needed a constant load to keep it going?
We'll see if the patient revives in a couple of hours.

***

I think I read somewhere that the copper has to be corroded. If so, you'd need to use bare copper wire and swab it with copper sulfate. You can get this 'antiquing' solution, which is called Copper Patina, at a stained glass store.

***

Thanks for the info. I'm using bare wire and I dipped it in salt water, so that's a start.
I did find a short and managed to isolate that section. The voltage is back up to .49 volts, but it's still nothing to write home about. I think you would have to make a really big coil for it to do much.
I really should get back to my sammich motor but I've lost the momentum. I need to find some inspiration somewhere. 10 hours more work will see it running, but I can't seem to kindle the spark.
I even cleaned the shop...

---

http://overunity.com/3500/nathan-stubblefield-earth-batteryself-generating-induction-coil-replications/4005/#.VaqPUPnRvIU

If you want the coil to be a self generating electromagnet then terminals 5&6 are tied together. If not then they are open for the voltaic couple.

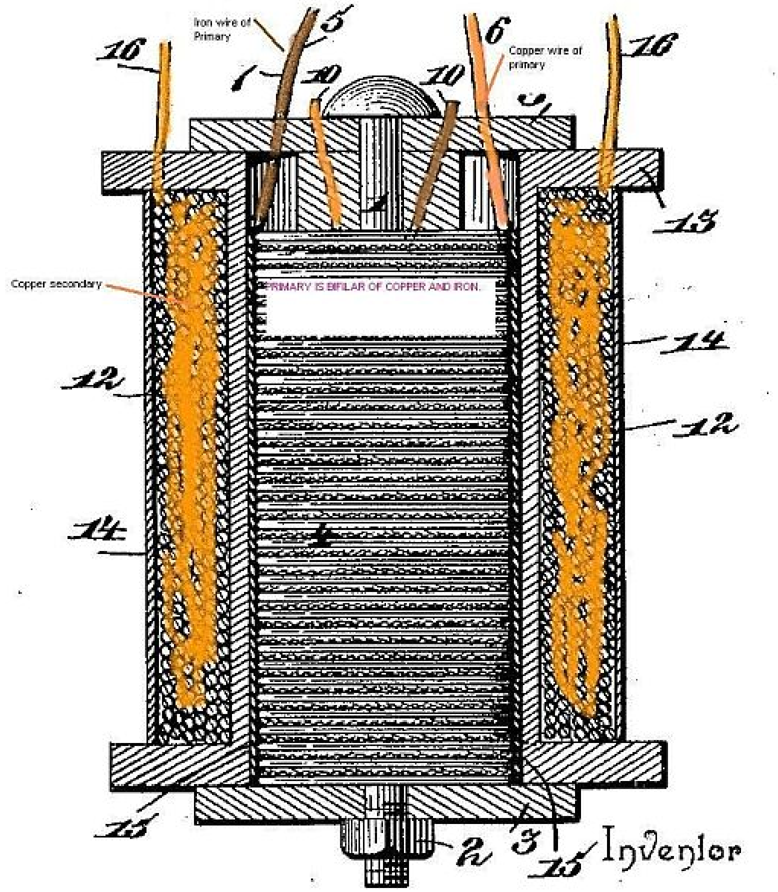
16
Iron wire of Primary
5
1
10
10
6
Copper wire of primary
9
16
13
Copper secondary
PRIMARY IS BIFILAR OF COPPER AND IRON.
14
12
12
14
4
15
2
3
15
Inventor

MURRAY, KY.
W. G. Love,

---

http://johnbedini.net/john34/stublefield1.html

We do not know the secret of the earth charge as Nathan Stubblefield determined it. Others since his time have observed fluctuations at certain times of the year in ground energy. It may be that a sudden induction is required before the excess ground charge surges to the surface..... like priming a pump.

So I did some experiments in building different types of cells. I used a 10 inch carriage bolt 3/4 inch diameter between two delron spacers. I used steel wire and copper wire that was coated. I would wind one layer and sand it off, check for shorts, and then wrap it with gauze and start the next winding. I then put the secondary induction coil windings on.

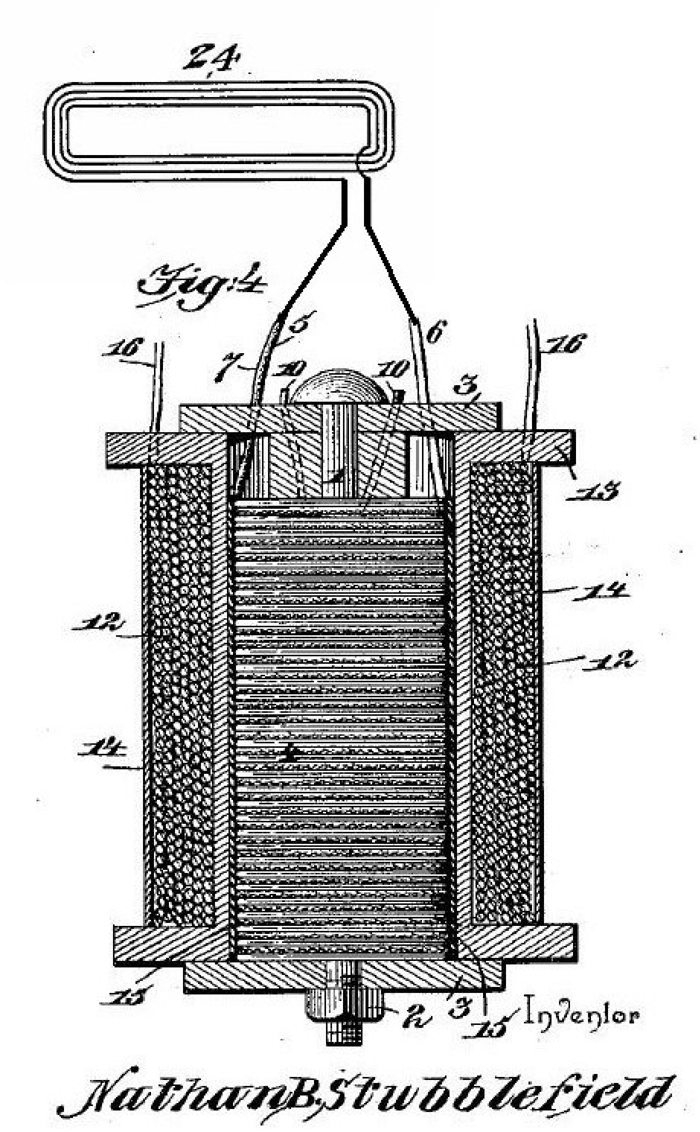

These are the results I got in a one gallon plastic jug filled with water. Volts .07, 250ma into dead short through meter shunt. I'm still working on a better cell to go into the ground at this time.

Stubblefield published an extraordinary brochure in 1898 to attract investors who had expressed interest in consolidating a small corporation around his work. In this brochure, Stubblefield insisted that power for his device was not generated in the cell. He calmly stated that the cell received its surplus energy from the earth. In a less discussed portion of the brochure, Stubblefield stated that "electrotherapeutic" devices had been developed from his earth battery. Other researchers made similar claims for their earth batteries (Hicks, Mellon). During this time, Stubblefield declared that news, weather, and other announcements could be broadcast through the ground across a great territory for private reception. He also added that simultaneous messages and news of all kinds would soon be transmitted through the ground from a central distribution station. (Shades of Tesla!)

In 1902 Stubblefield set up one of his sets in a "Mainstreet" upper office... in a hardware shop. From that point to his farm (some 6000 feet distant) he conducted continuous conversations with his son Bernard. Tapping with a pencil on his one-piece transceiver, Bernard was quickly heard in a loud, very clear voice. This transceiver was a carbon button placed in a tin snuff box. Speech and response were transacted through the same device, which acted as both microphone and loudspeaker. Cells (EARTH BATTERIES) were placed downstairs from the office in the ground. They were never removed and never wore out, though operating twenty-four hours a day around the clock.

Nathan Stubblefield offered to construct a large scale power station for the town of Murray. His quoted initial installation costs were estimated at five thousand dollars. The town politicians declined the offer. As a result, the technique of drawing up electricity from the earth remains a mystery.

The Stubblefield ground radio system was demonstrated for approximately one thousand Murray residents. Photographs of Stubblefield and his family, and a good crowd of witnesses from town, show the cell laying on the ground among all his assembled inventions; and a flower-pot sized coil of good volume. Other devices show motors and large capacitor stacks for aerial voice transmission experiments.

Stubblefield declares it to be an "energy receiver....a receptive cell for intercepting electrical ground waves". Its conductive ability somehow absorbs and directs the enormous volumes of earth energy.

Whether the current derived from this cell is electricity as we know it has been questioned. One indicator is not found when considering his use of the energy in lighting lamps. With this energy Nathan Stubblefield operated a score of arc lamps at full brightness for twenty-four hours a day. There was a definite trigger by which this energy was stimulated and maintained.

The induction coil which bears his name is equipped with three coils which are wrapped around and upon a heavy iron core. Bare iron wire and cotton-covered copper wire are wrapped side by side, comprising a primary coil body. Each layer of the primary coil body is covered by a band of cotton insulation, bringing four wire leads to the coil terminus. Two leads of iron and two of copper are external to the coil. Commercial electrical power is obtained through these connective terminals.

In addition to this bimetallic winding, there is a third winding: the "secondary". This third coil is insulated from the primary bimetallic coil, serving as a trigger device. Presumably, a stimulating impulse shock was introduced into the tertiary coil, after which the upwelling electrical ground response brought forth powerful currents in both iron and copper coils.

Electrolytically (as a battery in acid or saltwater) the Stubblefield coil is disappointing, producing less then one volt according to those who have duplicated its construction. Stubblefield's bimetallic coil was a "plug": a receiver which intercepts the vast and free electrical reservoir of the ground itself. His patent and subsequent company brochures define the manner in which his earth battery was to be activated.

Technically, the Stubblefield device is a modified thermocouple (a bimetal in tight surface contact) but a thermocouple could not supply the degree of power which he reported. While this arrangement could develop a few milliwatts of power in appropriately hot ground spots, the thermoelectric explanation of the device cannot explain the phenomenal output reported in the news reports of Stubblefield's demonstrations.

Furthermore, though the Stubblefield power receiver is wound like an induction coil, it produces a steady direct current output. This poses additional problems for the conventional engineer. Electrical induction only occurs with electrical alternations, oscillations, and impulses.

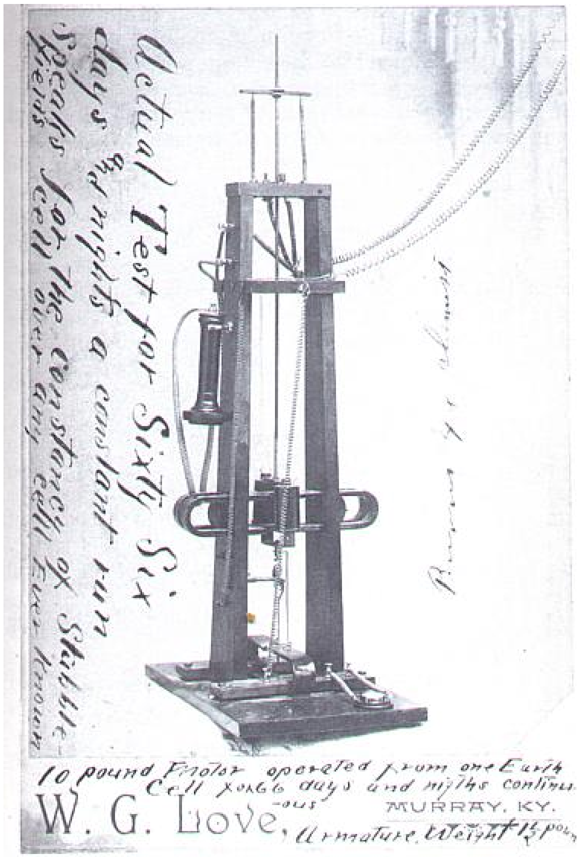
Actual Test for Sixty Six
days and nights a constant run
Speaks for the constancy of Stubble-
fields cell over any cells ever known
10 pound Motor operated from one Earth
Cell for 66 days and nights continu-
-ous
W. G. Love,
MURRAY, KY.
Armature Weight 1½ pou

Pogue Library

Electric Motor that Stubblefield operated non-stop for 66 days with one Earth Cell battery. There is no explanation for the telephone receiver.

January 1, 1902

Witnesses describe ground-powered motors which ran unceasingly and unattended for months without need for replacing or replenishing the ground battery. Small machinery, clocks, and loud gongs were run by other ground-buried cells as reported by credible witnesses. Stubblefield may have discovered the auto-magnifying voltage effect of electrostatic induction in coils before Tesla, who later utilized the effect in his special electrostatic Transformers.

These buried coils may have become saturated with earth electrostatic energy, which travelled from subterranean depths. In such a case, the mere battery power of the coil was replaced by the electrostatic flow, the coil acting as an electrode. This seems obvious when considering the fact that its ordinary battery current (1 watt) was gradually replaced by continually growing electrical current of far greater proportion.

Electrical Battery
US600457

[PDF]

(No Model.)

N. B. STUBBLEFIELD.
ELECTRICAL BATTERY.

No. 600,457. Patented Mar. 8, 1898.

Fig: 1.

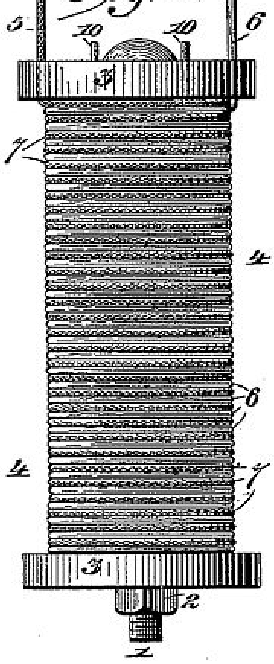
5
10
10
6
3
7
4
6
4
7
3
2
1

Fig: 2.

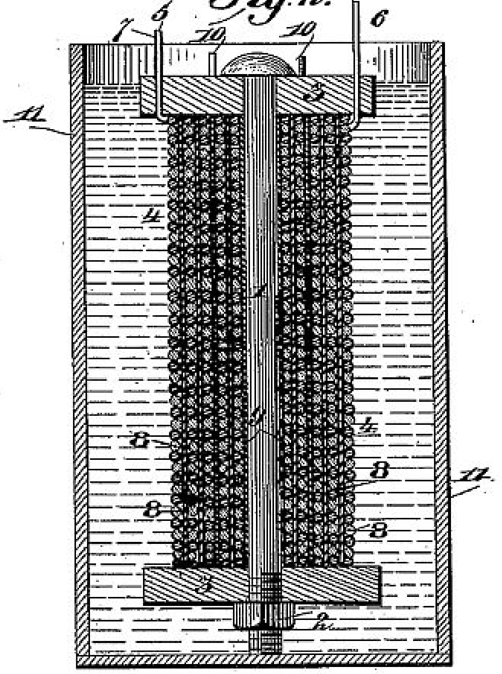
7
5
10
10
6
3
11
4
1
9
8
4
8
8
8
11
3
2

Fig: 3.

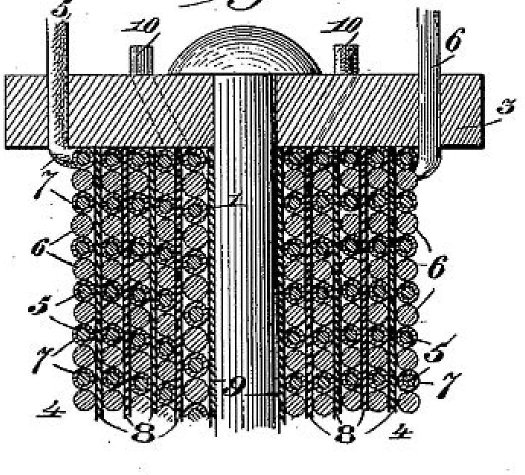
5
10
10
6
3
7
1
6
6
5
7
9
5
7
4
8
8
4

Fig: 4.

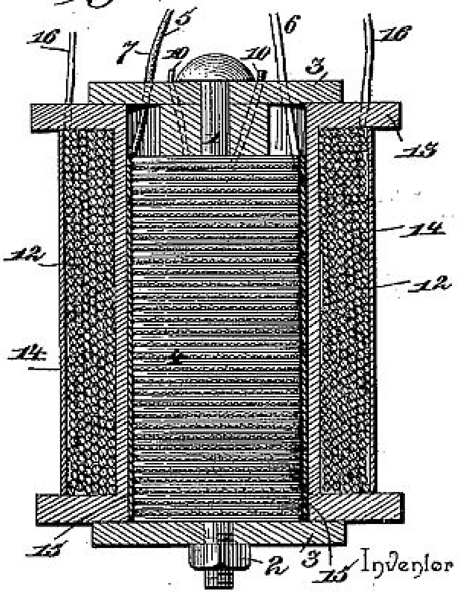
16
7
5
10
10
6
3
16
13
12
14
12
14
1
13
2
3
13
Inventor

Wireless Telephone
US887357

[PDF]

itnesses

Fig.1.

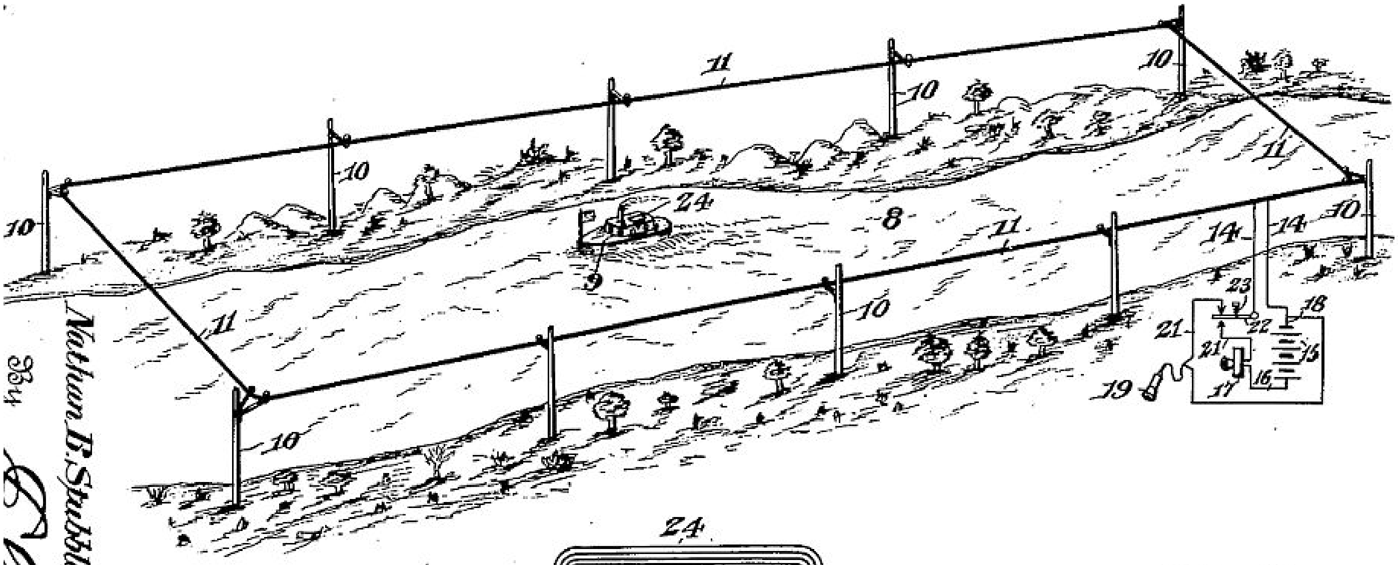
11
10
24
8
9
14
23
22
18
15
16
17
21
19

Fig.2.

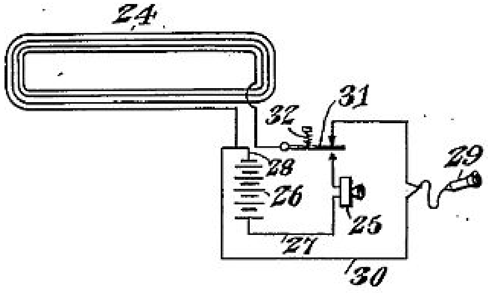
24
31
32
28
26
25
27
30
29

Fig.3.

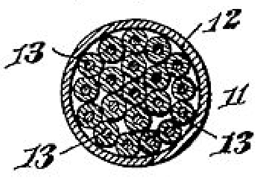

Nathan B. Stubblefield, Inventor

By

3 SHEETS—SHEET 1.

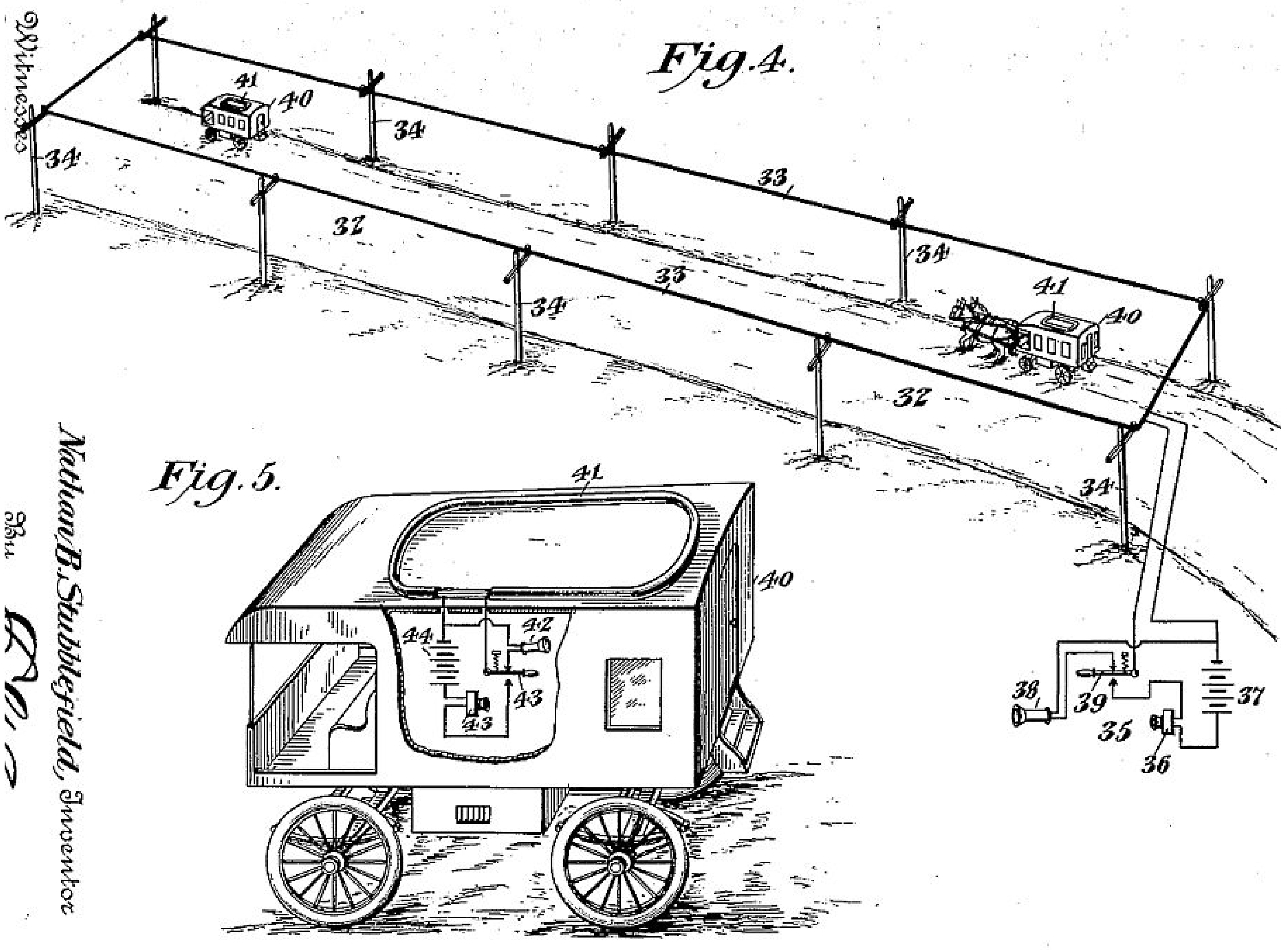
Fig. 4.

Fig. 5.

Witnesses
34
41
40
34
33
32
33
34
34
41
40
32
34
37
38
39
35
36
41
40
44
42
43
43
Nathan B. Stubblefield, Inventor
By
Atty.

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

(No Model.)

N. B. STUBBLEFIELD.

ELECTRICAL BATTERY.

No. 600,457. Patented Mar. 8, 1898.

Fig: 1.

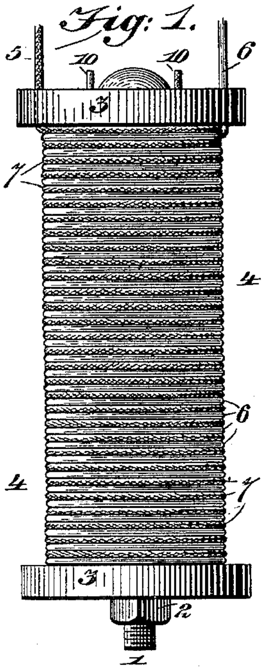
5
10
10
6
3
7
4
6
4
7
3
2
1

Fig: 2.

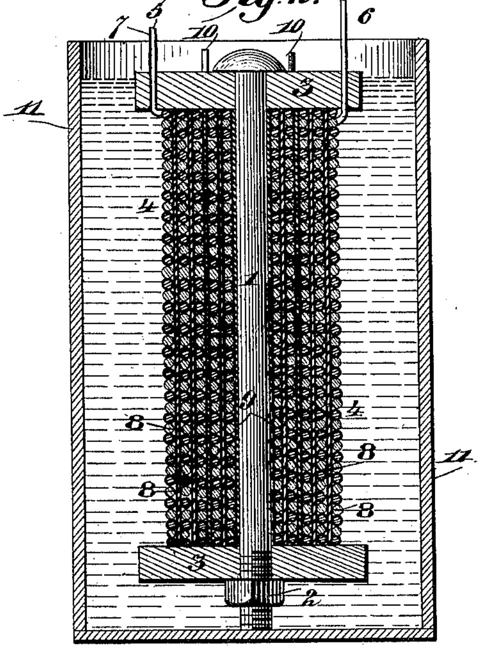
7
5
10
10
6
3
11
4
1
9
8
4
8
8
8
11
3
2

Fig: 3.

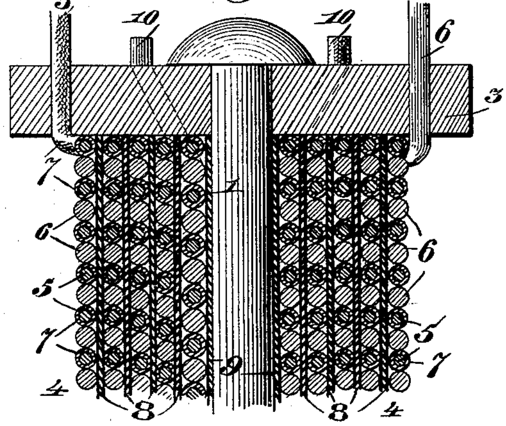
5
10
10
6
3
7
1
6
6
5
5
7
9
7
4
8
8
4

Fig: 4

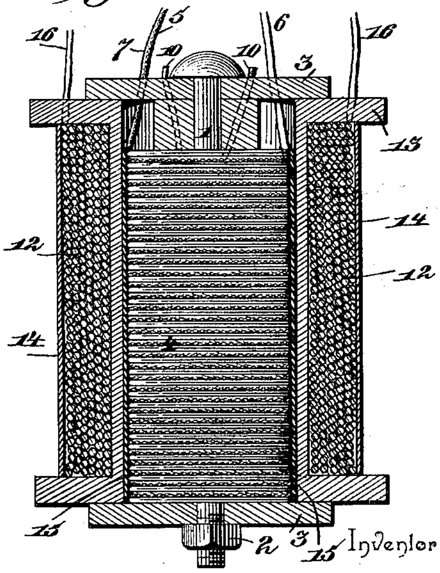
16
5
6
16
7
10
10
3
1
13
14
12
12
14
4
15
2
3
15

Inventor

Witnesses

H. G. Dieterich

S. P. Hollingsworth

Nathan B. Stubblefield

By his Attorneys,

C. A. Snow & Co.

THE NORRIS PETERS CO., PHOTO-LITHO., WASHINGTON, D. C.

UNITED STATES PATENT OFFICE.

NATHAN B. STUBBLEFIELD, OF MURRAY, KENTUCKY, ASSIGNOR OF ONE-HALF TO WILLIAM G. LOVE, OF SAME PLACE.

ELECTRICAL BATTERY.

SPECIFICATION forming part of Letters Patent No. 600,457, dated March 8, 1898.

Application filed October 24, 1896. Serial No. 609,969. (No model.)

To all whom it may concern:

Be it known that I, NATHAN B. STUBBLEFIELD, a citizen of the United States, residing at Murray, in the county of Calloway and State of Kentucky, have invented a new and useful Electrical Battery, of which the following is a specification.

This invention relates to electrical batteries; and it has for its object to provide a novel and practical battery for generating electrical currents of sufficient force for practical use, and also providing means for generating not only a constant primary current, but also an induced momentary secondary current.

It is well known that if any voltaic couple be immersed in water or placed in moist earth the positive element of the couple will undergo a galvanic action of sufficient intensity to produce a current when the terminals of the couple are brought in contact, and this form of battery is commonly known as the "water" battery, usually employed for charging electrometers, but not capable of giving any considerable current owing to their great internal resistance. Now the principle involved in this class of batteries is utilized to some extent in carrying out the present invention, but I contemplate, in connection with water or moisture as the electrolyte, the use of a novel voltaic couple constructed in such a manner as to greatly multiply or increase the electrical output of ordinary voltaic cells, while at the same time producing in operation a magnetic field having a sufficiently strong inductive effect to induce a current in a solenoid or secondary coil.

To this end the invention contemplates a form of voltaic battery having magnetic induction properties of sufficient intensity, so as to be capable of utilization for practical purposes, and in the accomplishment of the results sought for the invention further provides a construction of battery capable of producing a current of practically constant electromotive force and being practically free of the rapid polarization common in all galvanic or voltaic batteries.

With these and many other objects in view the invention consists in the novel construction, combination, and arrangement of parts hereinafter more fully described, illustrated, and claimed.

In the drawings, Figure 1 is a side elevation of an electrical battery constructed in accordance with this invention. Fig. 2 is a central longitudinal sectional view of the battery, showing the same immersed in water as the electrolyte. Fig. 3 is an enlarged sectional view of a portion of the battery, showing more clearly the manner of winding the voltaic couple or, in other words, the wires comprising the couple. Fig. 4 is a vertical sectional view of the battery, shown modified for use with an induction-coil.

Referring to the accompanying drawings, the numeral 1 designates a soft-iron core-piece extending longitudinally of the entire battery and preferably in the form of a bolt having at one end a nut 2, which permits of the parts of the battery being readily assembled together and also quite as readily taken apart for the purposes of repair, as will be readily understood. The central longitudinally-arranged core-piece 1 of the battery has removably fitted on the opposite ends thereof the oppositely-located end heads 3, confining therebetween the magnetic coil-body 4 of the battery, said heads 3 being of wood or equivalent material. The coil-body 4 of the battery is compactly formed by closely-wound coils of a copper and iron wire 5 and 6, respectively, which wires form the electrodes of the voltaic couple, and while necessarily insulated from each other, so as to have no metallic contact, are preferably wound in the manner clearly illustrated in Fig. 3 of the drawings.

In the preferred winding of the wires 5 and 6 the copper wire 5 is incased in an insulating-covering 7, while the iron wire 6 is a bare or naked wire, so as to be more exposed to the action of the electrolyte and at the same time to intensify the magnetic field that is created and maintained within and around the coil-body 4, when the battery is in operation and producing an electrical current. While the iron wire 6 is preferably bare or naked for the reasons stated, this wire may also be insulated without destroying the operativeness of the battery, and in order to secure the best results the wires 5 and 6 are wound side by side in each coil or layer of the windings, as clearly shown in Fig. 3 of the drawings, so that in each coil or layer of the windings there will be alternate convolutions of the copper and iron wires forming the voltaic couple, and it will of course be understood that there may be any number of separate coils or layers of the wires according to the required size and capacity of the battery. Each coil or layer of the windings is separated from the adjacent coils or layers by an interposed layer of cloth or equivalent insulating material 8, and a similar layer of insulating material 9 also surrounds the longitudinal core-piece 1 to insulate from this core-piece the innermost coil or layer of the windings.

The terminals 10 of the copper and iron wires 5 and 6 are disconnected so as to preserve the character of the wires as the electrodes of the voltaic couple; but the other or remaining terminals of the wires are brought in contact through the interposition of any electrical instrument or device with which they may be connected to cause the electric currents generated in the coil-body 4 to flow through such instrument or device.

In the use of the battery constructed as described the same may be immersed in a cell or jar 11, containing water as the electrolyte; but it is simply necessary to have the coil-body 4 moist to excite the necessary action for the production of a current in the couple, and it is also the contemplation of the invention to place the battery in moist earth, which alone is sufficient to provide the necessary electrolytic influence for producing an electric current.

It has been found that by reason of winding the couple of copper and iron wires into a coil-body the current traversing the windings of this body will produce a magnetic field within and around the body of sufficiently strong inductive effect for practical utilization by means of a solenoid or secondary coil 12, as illustrated in Fig. 4 of the drawings.

The solenoid or secondary coil 12 is of an ordinary construction, comprising a wire closely wound into a coil of any desired size on an ordinary spool 13 and incased within a protective covering 14 of mica, celluloid, or equivalent material. The spool 13 of the solenoid or secondary coil may be conveniently secured directly on the exterior of the coil-body 4 between the heads 3 with a suitable layer or wrapping of insulating material 15, interposed between the spool and the body 4, and the terminals 16 of the solenoid or secondary coil may be connected up with any instrument usually operated by secondary currents—such, for instance, as a microphone-transmitter or telegraphic relay. The magnetic field produced by the current traversing the coil-body 4 induces a secondary current in the solenoid or secondary coil 12, when the ordinary make and break of the primary current produced within the coil 4 is made between the terminals of said coil 4. It will therefore be seen that the construction of the battery illustrated in Fig. 4 is practically a self-generating induction-coil, and it can be used for every purpose that a coil of this character is used, for as long as the coil-body 4 is wet or damp with moisture electric currents will be produced in the manner described. It will also be obvious that by reason of the magnetic inductive properties of the coil-body 4 the core-piece 1 will necessarily be magnetized while a current is going through the body 4, so that the battery may be used as a self-generating electromagnet, if so desired, it being observed that to secure this result is simply required connecting the extended terminals of the wires 5 and 6 together after wetting or dampening the coil-body.

Many other uses of the herein-described battery will suggest themselves to those skilled in the art, and I will have it understood that any changes in the form, proportion, and the minor details of construction may be resorted to without departing from the principle or sacrificing any of the advantages of this invention.

Having thus described the invention, what is claimed, and desired to be secured by Letters Patent, is—

1. A combined electrical battery and electromagnet, for use with water as an electrolyte, comprising a soft-iron core-piece, and a voltaic couple of copper and iron wires insulated from each other and closely and compactly wound together in separate insulated layers to produce a solid coil-body surrounding the soft-iron core-piece, substantially as set forth.

2. An electrical battery for use with water as an electrolyte comprising a voltaic couple of insulated copper wire and bare iron wire closely wound into a coil-body, substantially as described.

3. An electrical battery for use with water as an electrolyte comprising a voltaic couple of insulated copper and bare iron wire wound side by side in separate insulated layers to produce a coil-body, substantially as described.

4. An electrical battery, for use with water as an electrolyte, comprising a voltaic couple having its separate electrodes insulated from each other and closely wound into a compact coil-body forming a self-generating primary coil when moistened and a solenoid or secondary coil fitted on the coil-body of the couple, substantially as set forth.

In testimony that I claim the foregoing as my own I have hereto affixed my signature in the presence of two witnesses.

NATHAN B. STUBBLEFIELD.

Witnesses:

JOHN H. SIGGERS,

W. B. HUDSON.

No. 887,357. PATENTED MAY 12, 1908.

N. B. STUBBLEFIELD.

WIRELESS TELEPHONE.

APPLICATION FILED APR. 5, 1907.

3 SHEETS—SHEET 1.

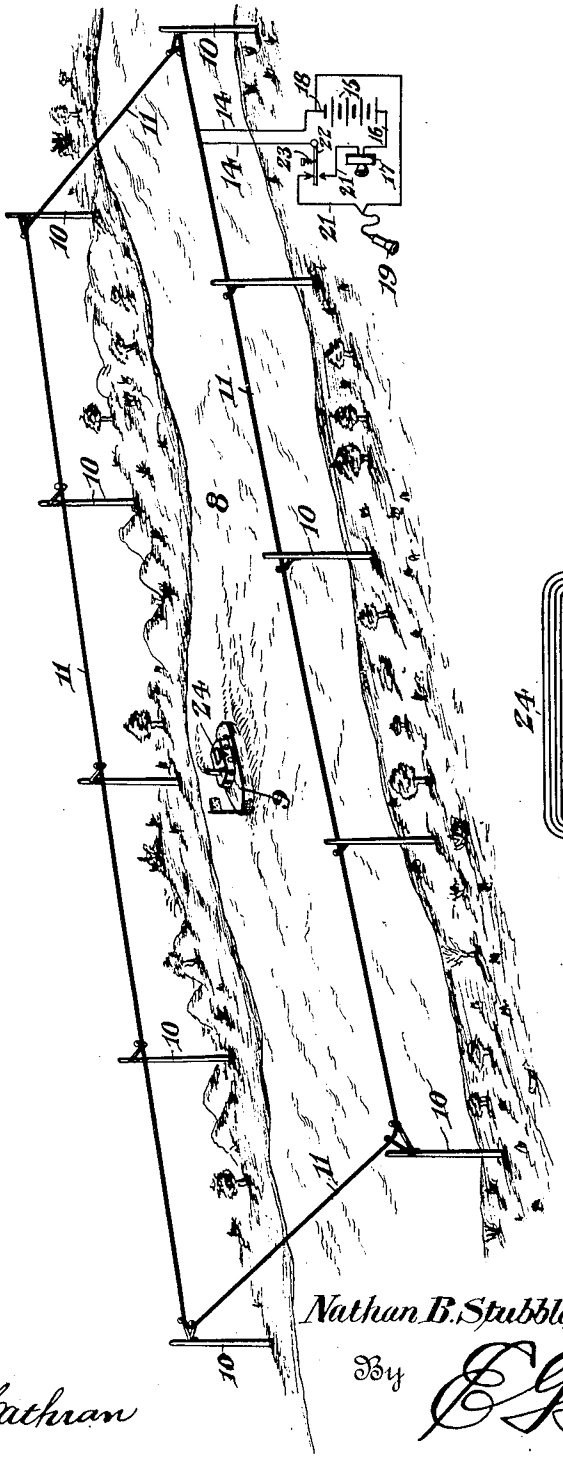
10
11
14
18
15
16
17
22
23
21
19
10
11
8
24
9
10
11

Fig. 1.

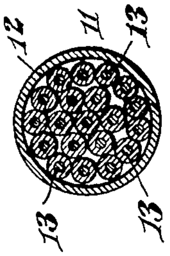
12
11
13
13
13

Fig. 3.

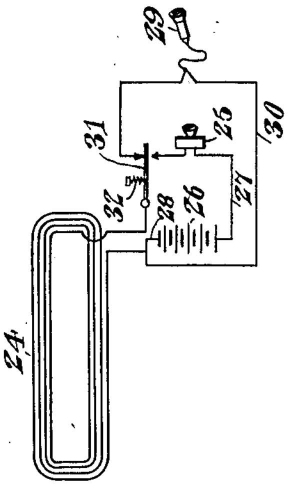
29
31
25
30
32
28
26
27
24

Fig. 2.

Nathan B. Stubblefield, Inventor

Witnesses

Jas. K. McCathran

B. G. Foster

By E. G. Siggers

Attorney

No. 887,357. PATENTED MAY 12, 1908.

N. B. STUBBLEFIELD.

WIRELESS TELEPHONE.

APPLICATION FILED APR. 5, 1907.

3 SHEETS—SHEET 2.

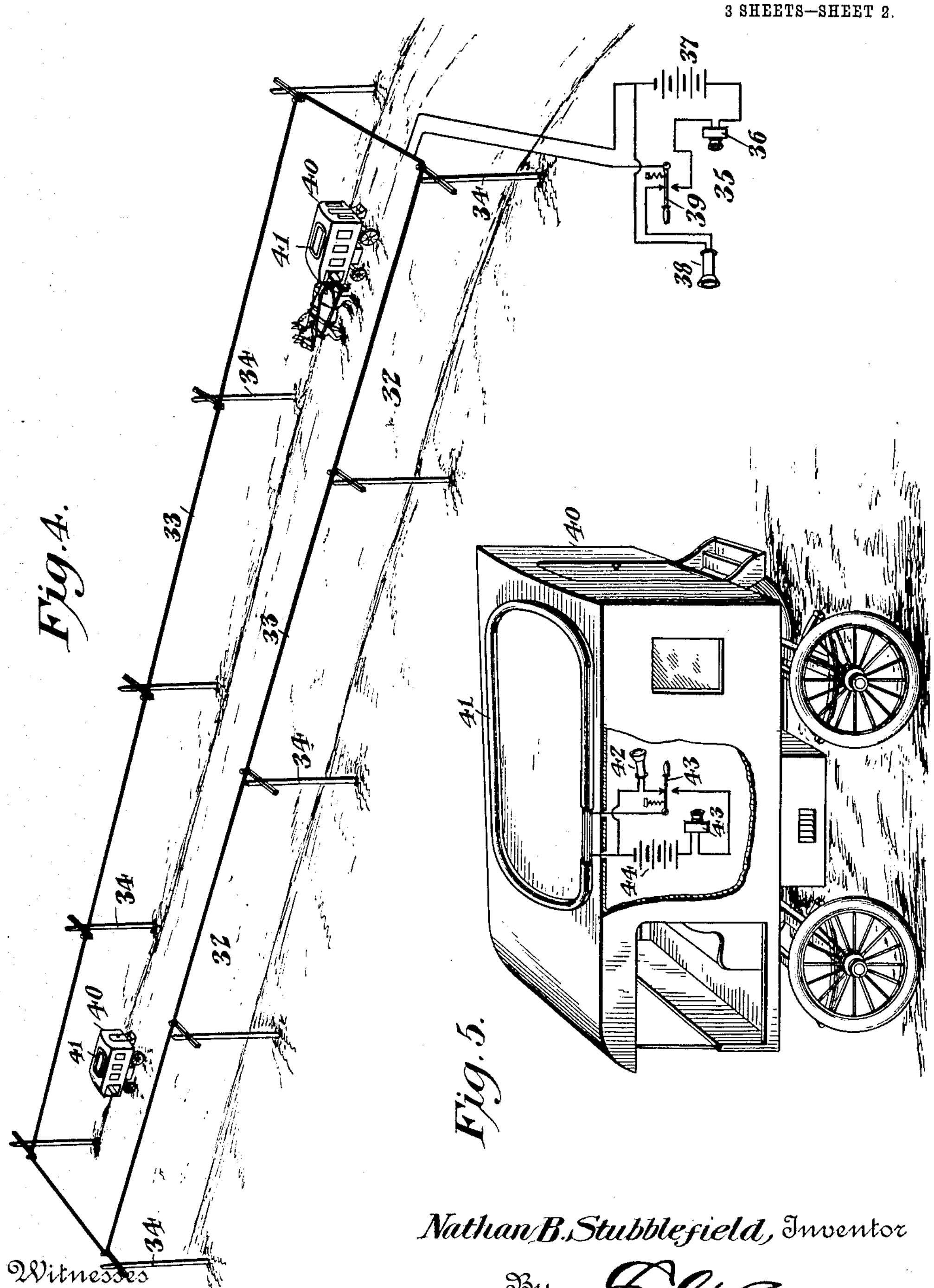

Fig. 4.

37
36
35
39
38
34
40
41
34
32
33
35
34
34
32
40
41
34

Fig. 5.

40
41
42
43
43
44

Nathan B. Stubblefield, Inventor

Witnesses

Jas. K. McCathran

By E. G. Siggers

Attorney

No. 887,357. PATENTED MAY 12, 1908.

N. B. STUBBLEFIELD.

WIRELESS TELEPHONE.

APPLICATION FILED APR. 5, 1907.

3 SHEETS—SHEET 3.

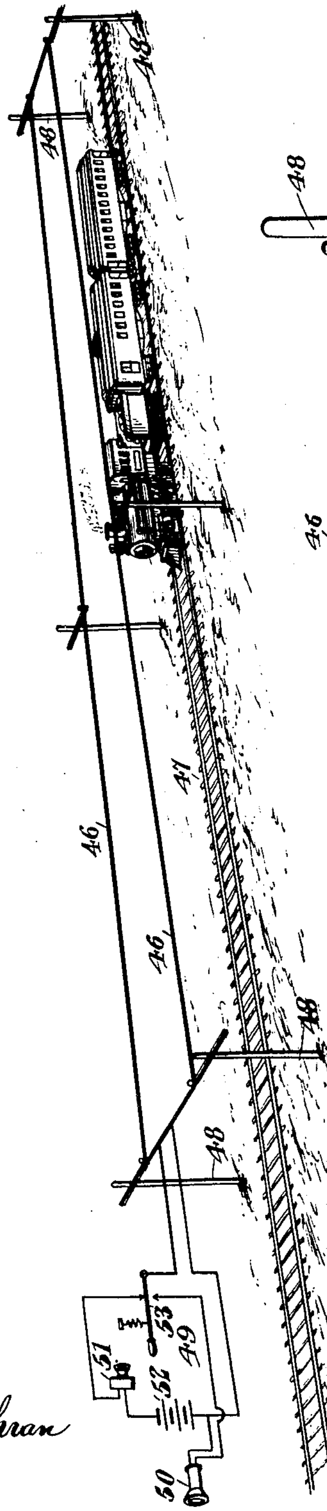
48
48
46
46
47
48
48
49
50
51
52
53

Fig. 6.

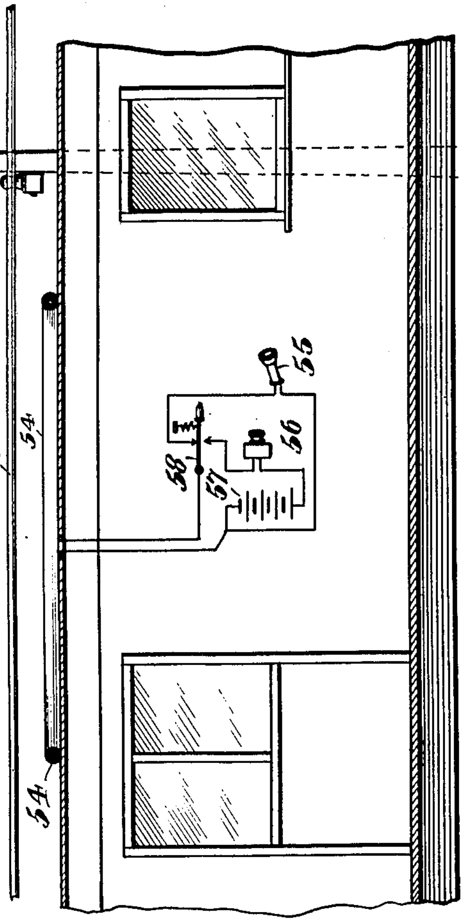
48
46
54
54
55
56
57
58

Fig. 7.

Witnesses

Jas. K. McCathran

B. G. Foster

Nathan B. Stubblefield,
Inventor

By E. G. Siggers
Attorney

UNITED STATES PATENT OFFICE.

NATHAN B. STUBBLEFIELD, OF MURRAY, KENTUCKY, ASSIGNOR OF TWELVE AND ONE-HALF ONE-HUNDREDTHS TO CONN LINN, FIVE ONE-HUNDREDTHS TO R. DOWNS, FIVE ONE-HUNDREDTHS TO B. F. SCHROADER, FIVE ONE-HUNDREDTHS TO GEORGE C. McLARIN, FIVE ONE-HUNDREDTHS TO JOHN P. McELRATH, TWO AND ONE-HALF ONE-HUNDREDTHS TO JEFF D. ROULETT, AND ONE-TWENTIETH TO SAMUEL E. BYNUM, ALL OF MURRAY, KENTUCKY.

WIRELESS TELEPHONE.

No. 887,357. Specification of Letters Patent. Patented May 12, 1908.

Application filed April 5, 1907. Serial No. 366,544.

To all whom it may concern:

Be it known that I, Nathan B. Stubblefield, a citizen of the United States, residing at Murray, in the county of Calloway and State of Kentucky, have invented a new and useful Wireless Telephone, of which the following is a specification.

The present invention relates to means for electrically transmitting signals from one point to another without the use of connecting wires, and more particularly comprehending means for securing telephonic communication between moving vehicles and way stations.

The principal object of the invention is to provide simple and practical means of a novel nature whereby clear and audible communication can be established, said means being simple and of a character that will permit certain of the station mechanisms to be small and compact.

In the accompanying drawings:—Figure 1 is a perspective view, showing means for establishing communication between a vessel and a shore station. Fig. 2 is a diagrammatic view of the mechanism mounted on the boat. Fig. 3 is a cross sectional view on an enlarged scale of the shore coil. Fig. 4 is a perspective view of a road-way, showing a system for establishing communication between road vehicles and a way-station, the latter being illustrated diagrammatically. Fig. 5 is a detail view of a vehicle equipped with one of the instruments, which is shown diagrammatically. Fig. 6 is a perspective view showing the system applied to a railway for establishing communication between a moving train and a way-station. Fig. 7 is a sectional view through a car showing in diagram the car mechanism illustrated in Fig. 6.

Similar reference numerals designate corresponding parts in all the figures of the drawings.

Referring to the embodiment illustrated in Figs. 1, 2 and 3, a water-way 8 is disclosed, upon which a vessel 9 operates. Surrounding the path of travel of the vessel, and preferably elevated on poles 10, is a coil 11 of considerable magnitude. This coil, as shown in Fig. 3, consists of an outer casing 12, within which is placed a conducting wire comprising a plurality of convolutions 13, each of which is insulated from the other. The terminals 14 of this coil extend to a suitable way-station, and at the station is located a powerful source of electrical energy 15, to which is connected by a suitable wire 16 an electrically operated transmitter 17. The battery or other source of electricity has a connection 18 with one of the leads 14. A receiver 19 of the ordinary type has a connection with the same lead 14, to which the battery is connected, and both the receiver and transmitter have connections 21 with the contacts of a switch 22. This switch has suitable means, as for instance, a spring 23, which normally maintains the receiver in circuit with the coil 11, as will be evident by reference to Fig. 1, but if the switch is thrown to break the circuit, it will then cut in the source of electrical energy 15 and the transmitter 17.

An outfit similar to the above, is located on the vehicle or boat 9, but the coil 24 thereof, shown in Fig. 2, is much smaller. As further illustrated in said figure, the mechanism mounted on the boat, consists of a transmitter 25, and a battery or other source of electrical energy 26 electrically connected, as shown at 27 and having a connection 28 with one of the leads of the coil. The receiver 29 also has a connection 30 with said lead. A switch 31 is connected to the other lead, and is normally held in a position by a spring 32 to maintain a closed circuit through the receiver 29 and the coil, though it may be moved to cut out said receiver and close the circuit through the coil, the source of electrical energy and the transmitter.

In this system, if it is desired to transmit from one station, as for instance, the shore-station, the switch 22 is moved downwardly to cut out the receiver and throw in the transmitter and source of electrical energy, while the operator upon the boat or vehicle leaving the mechanism in the condition shown in Fig. 2, holds the receiver 29 to his ear. If therefore the operator at the shore-station uses the transmitter in the ordinary manner, a varying current corresponding to that passing through the coil of great magnitude 11, will be induced in the coil 24, and the speech or other sounds will thus be transmitted to the operator on the boat. By reversing the arrangement, speech may be transmitted from the boat to the shore station.

The use of coils for both stations, each coil consisting of a plurality of convolutions has been found by experience to be of the utmost value, and furthermore experience has demonstrated that the employment of coils of different magnitudes is of great importance, for it has been found that while two small coils can be used to transmit but a short distance, if one large coil of the character set forth is employed, the other may be very small, and speech or sounds can be transmitted comparatively great distances from one to the other. These sounds are clearly audible.

The structure disclosed in Figs. 4 and 5 is of the same general character. A road-way 32 is disclosed surrounded by a coil 33 of great magnitude that is supported on suitable poles 34. The way-station 35 consists of a transmitter 36, a source of electrical energy 37 connected thereto, a receiver 38, and a switch 39, whereby the receiver or the transmitter and source of electrical energy can be thrown into circuit with the coil 33. The vehicles 40, which operate on the roadway, are provided with smaller coils 41 and instruments consisting of receivers 42, transmitters 43, sources of electrical energy 44 and switches 45 all arranged in the manner already described. In a system of this kind, it will be evident that the occupant of one vehicle may telephone to the home or way-station, and the message can be transmitted to another vehicle. Thus it will be evident that communication can be established between two moving vehicles or between a way-station and any vehicle desired which is within the range of the home- or way-station. The system is also capable of use in connection with railways, and in Figs. 6 and 7, such a system is disclosed in connection therewith. A comparatively great coil 46 is supported on opposite sides of the railway 47 by poles 48 and a station 49 has a receiver 50 and a transmitter 51, a source of electrical energy 52 and a switch 53, the last mentioned being employed for throwing either the receiver or the transmitter and source of electrical energy into closed circuit with the coil 46. One or more cars of a railway train is equipped with an outfit consisting of a coil 54, a receiver 55, a transmitter 56, a source of electrical energy 57, and a switch 58 for throwing either the receiver or the transmitter and source of electrical energy into circuit with the coil 54. It will be evident that the operation of these two last described systems are substantially the same as that first set forth, and no extended description thereof is believed to be necessary.

From the foregoing, it is thought that the construction, operation, and many advantages of the herein described invention will be apparent to those skilled in the art, without further description, and it will be understood that various changes in the size, shape, proportion, and minor details of construction, may be resorted to without departing from the spirit or sacrificing any of the advantages of the invention.

Having thus fully described my invention, what I claim as new, and desire to secure by Letters Patent, is:—

1. In a system of the character described, the combination with a vehicle, of a comparatively small coil of conducting material mounted thereon, electrical transmitting and receiving mechanism including a source of electrical energy connected to the small coil and carried by the vehicle, a stationary aerial coil of much greater magnitude than the small coil having its opposite stretches or sides extending along the opposite sides of the path of travel of the vehicle and elevated above the same and above the vehicle coil, and electrical transmitting and receiving mechanism connected to the greater coil and including a source of heavy electrical current.

2. In a system of the character described, the combination with a vehicle, of a coil of conducting material mounted thereon, electrical transmitting mechanism, a source of electrical energy connected thereto, receiving mechanism, means for connecting either the transmitting mechanism and source of electrical energy or the receiving mechanism to the coil, a stationary coil of greater magnitude surrounding the path of travel of the vehicle and comprising a plurality of convolutions of conducting material, the different convolutions being insulated one from the other, means for supporting the coil in an elevated position, electrical transmitting mechanism, a source of great electrical energy connected to said transmitting mechanism, electrical receiving mechanism, and means for electrically connecting either the transmitting mechanism and source of electrical energy or the receiving mechanism to the said coil of greater magnitude.

3. Means for communicating between a plurality of stations which consists of an aerial electrical coil of great magnitude, means for supporting the said coil, a station electrically connected to the great coil and comprising transmitting and receiving mechanism that includes a source of heavy electrical energy, and a plurality of other separate stations simultaneously in coacting relation with the aerial coil, each of said latter stations comprising a coil of conducting material spaced from but in coacting relation with said great coil and below the same, and transmitting and receiving mechanism connected to said other coil and including a source of electrical energy.

4. Means for communicating between a plurality of stations which consists of an aerial coil of conducting material of great magnitude, transmitting and receiving mechanism connected to said aerial coil and including a source of heavy electrical energy, a plurality of vehicles movable between the opposite sides or stretches of the great coil, coils carried by said vehicles and disposed within the field of action of the aerial coil, and transmitting and receiving mechanism mounted on each vehicle and including a source of electrical energy.

In testimony, that I claim the foregoing as my own, I have hereto affixed my signature in the presence of two witnesses.

NATHAN B. STUBBLEFIELD.

Witnesses:
J. P. McElrath,
J. H. Coleman.

US 20060027709A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2006/0027709 A1**

Pinto (43) **Pub. Date: Feb. 9, 2006**

(54) **APPARATUS COMPRISING OF PROPULSION SYSTEM**

(75) Inventor: **Fabrizio Pinto**, Monrovia, CA (US)

Correspondence Address:
DEMONT & BREYER, LLC
SUITE 250
100 COMMONS WAY
HOLMDEL, NJ 07733 (US)

(73) Assignee: **InterStellar Technologies Corporation**, Monrovia, CA

(21) Appl. No.: **11/197,787**

(22) Filed: **Aug. 4, 2005**

Related U.S. Application Data

(60) Provisional application No. 60/598,658, filed on Aug. 4, 2004.

Publication Classification

(51) **Int. Cl.**
B64G 1/40 (2006.01)

(52) **U.S. Cl.** **244/171.1**

(57) **ABSTRACT**

A propulsion system that does not consume fuel. The system operates to modify the dispersion force (i.e., van der Waals) that arises between particles, such as neutral atoms. A lifting force is generated as a result of this modification of the dispersion force. In the illustrative embodiment, the propulsion system includes particles, a particle trap, a source of electromagnetic energy, and a piston.

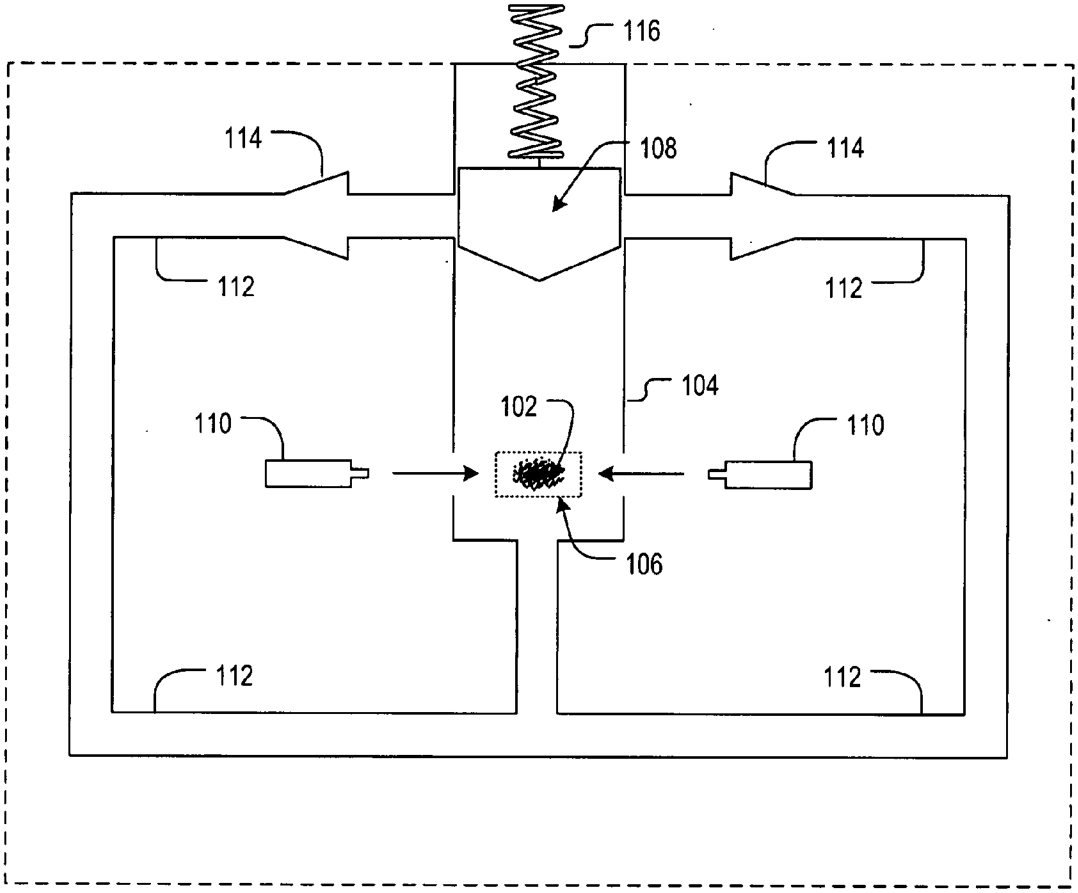

Figure 1

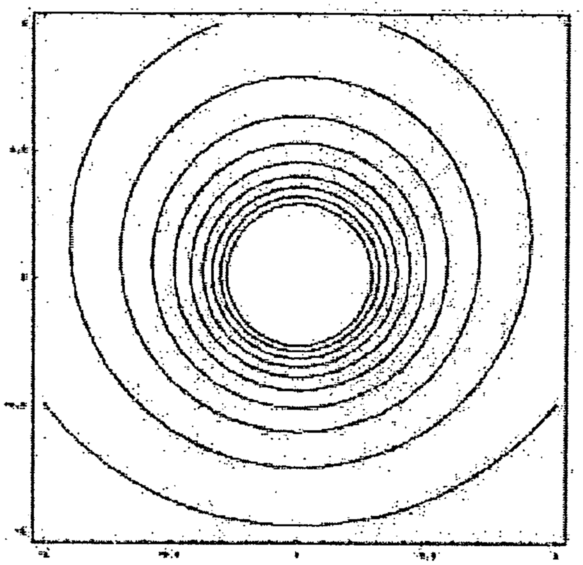

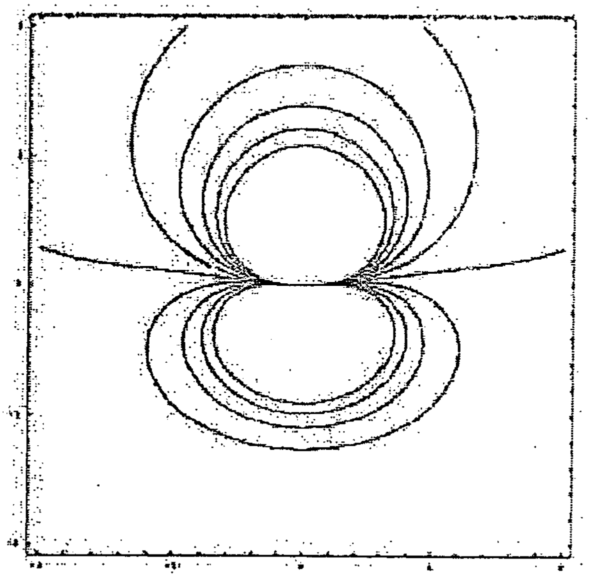

Figure 2

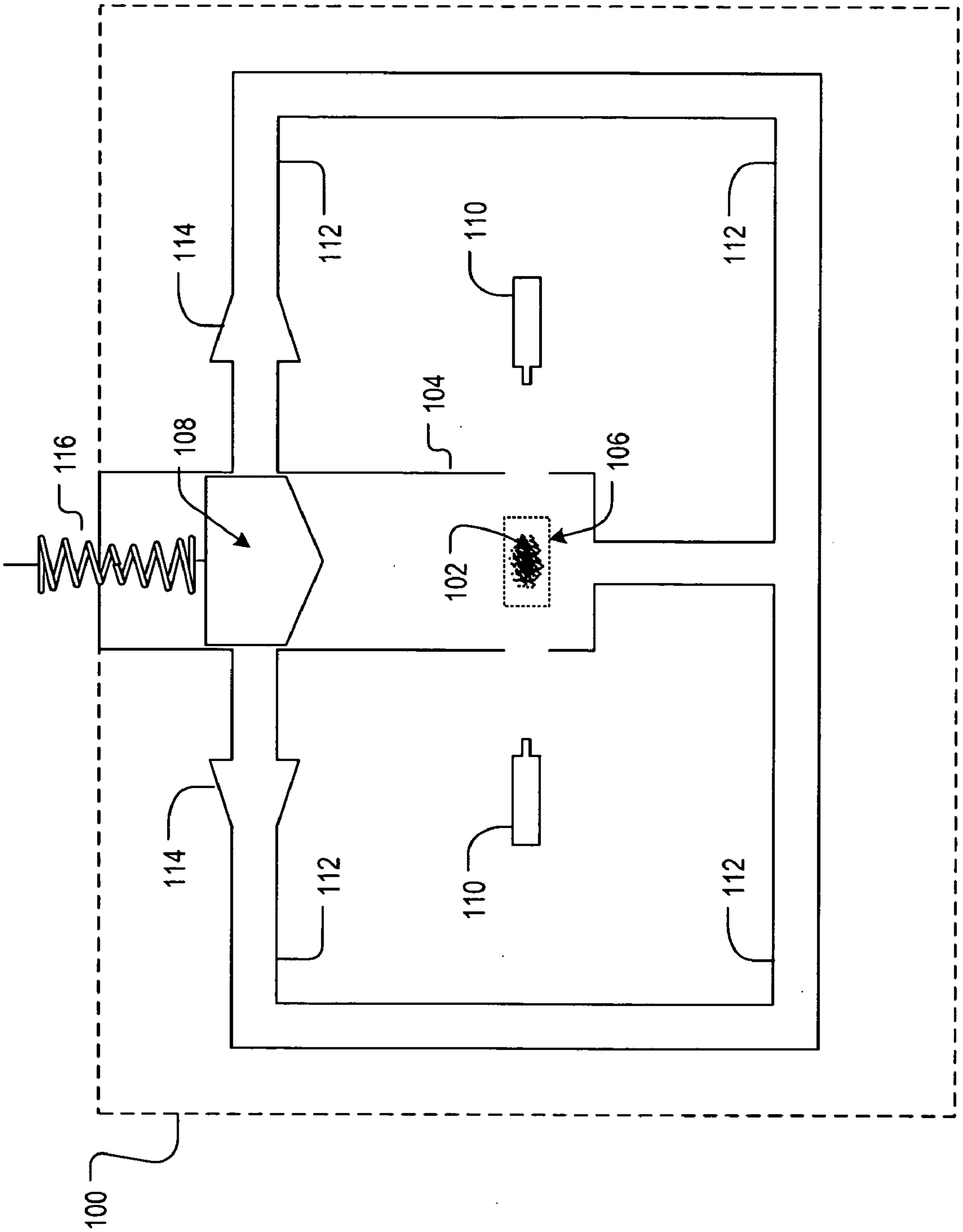
116
108
104
106
102
114
114
112
112
112
112
110
110
100

Figure 3

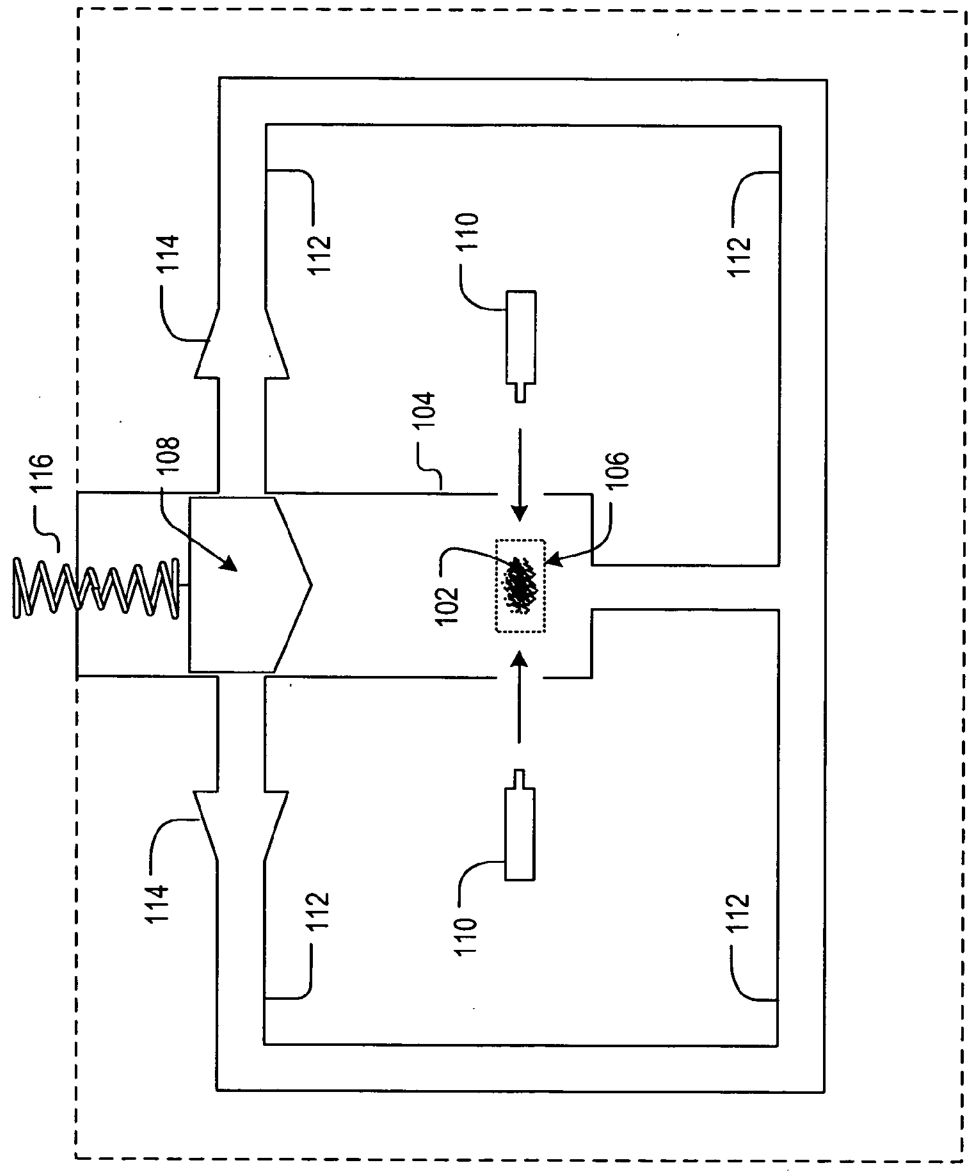
116
108
114
112
110
104
106
102
114
112
110
112
112
112

Figure 4

Figure 5

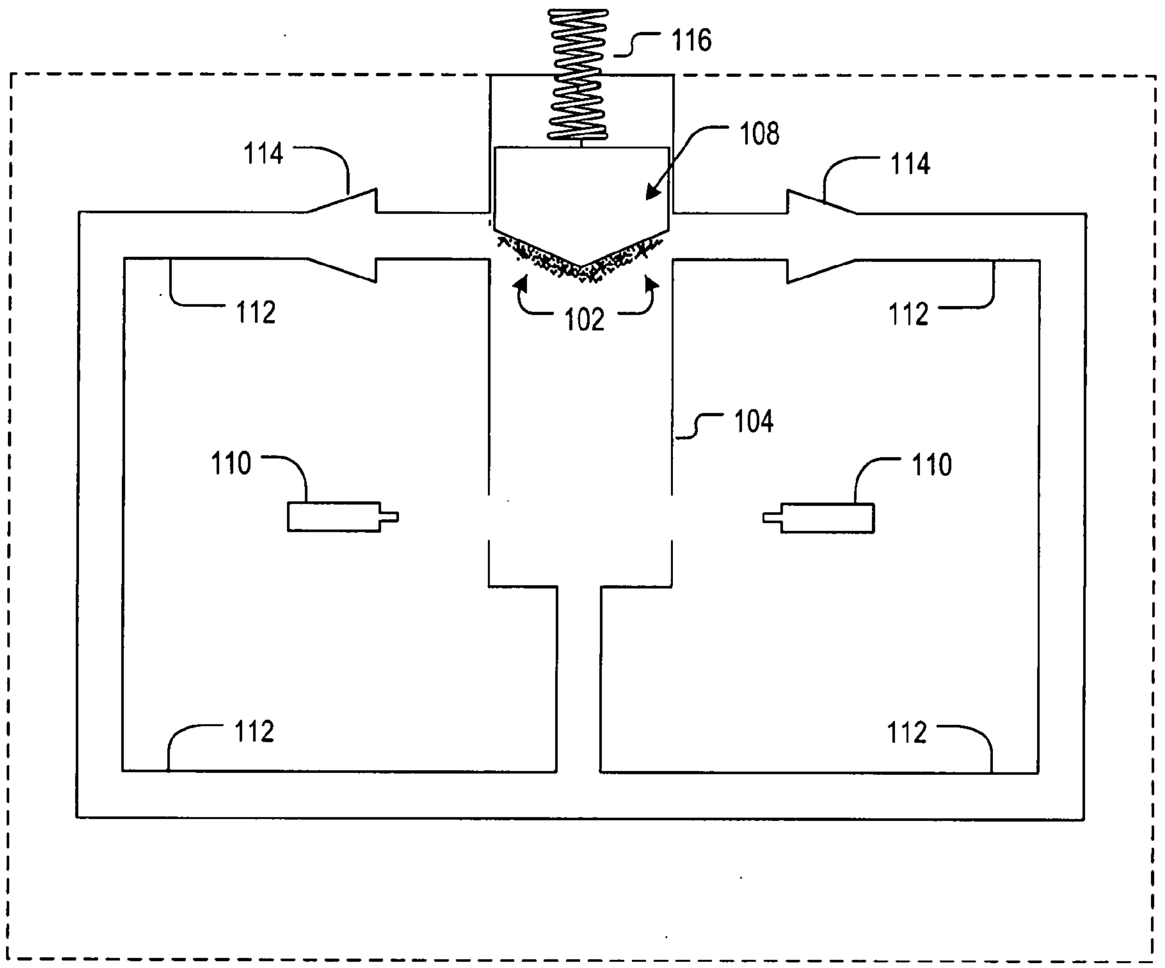
116
114
108
114
112
102
112
104
110
110
112
112

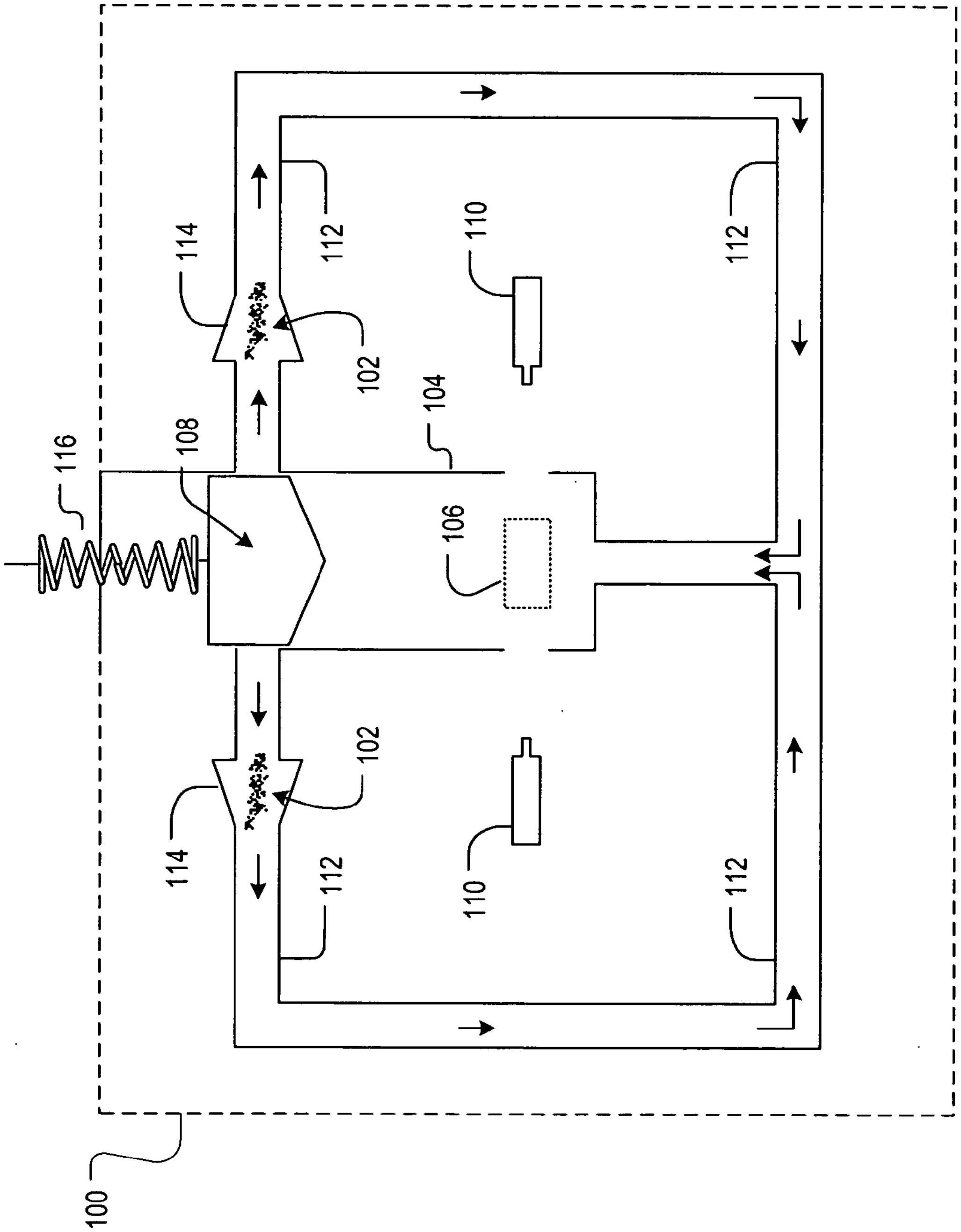
116
108
114
104
102
112
110
112
106
102
114
112
110
112
100

Figure 6

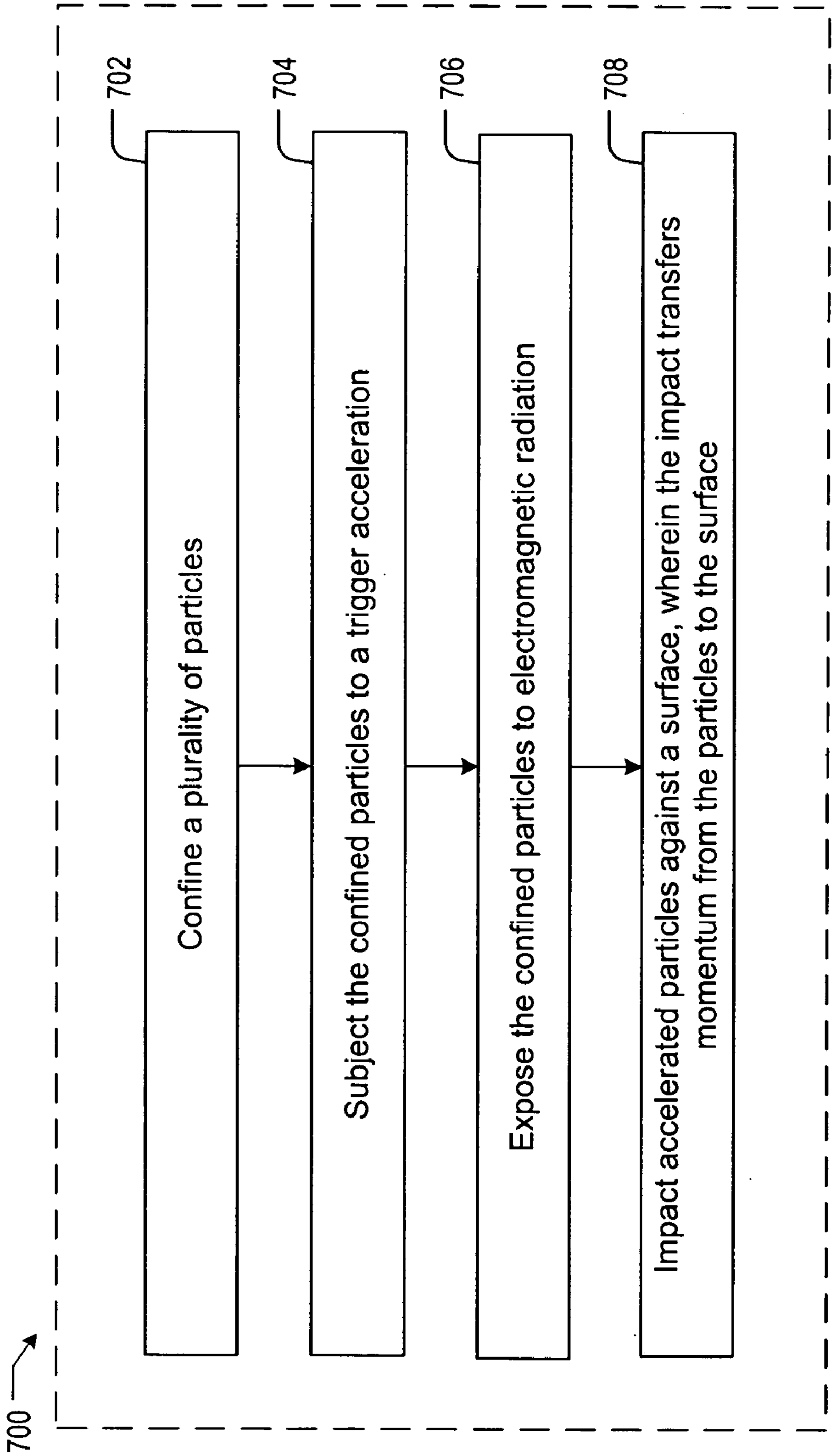
700
702
Confine a plurality of particles
704
Subject the confined particles to a trigger acceleration
706
Expose the confined particles to electromagnetic radiation
708
Impact accelerated particles against a surface, wherein the impact transfers momentum from the particles to the surface

Figure 7

Figure 8

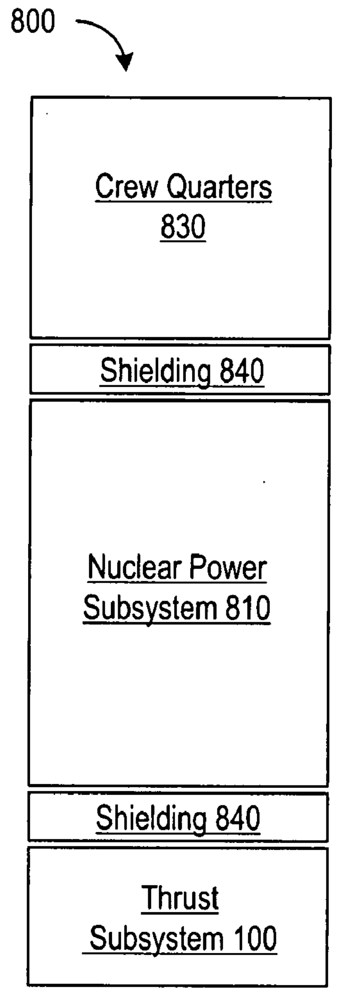
800
Crew Quarters 830
Shielding 840
Nuclear Power Subsystem 810
Shielding 840
Thrust Subsystem 100

Figure 9

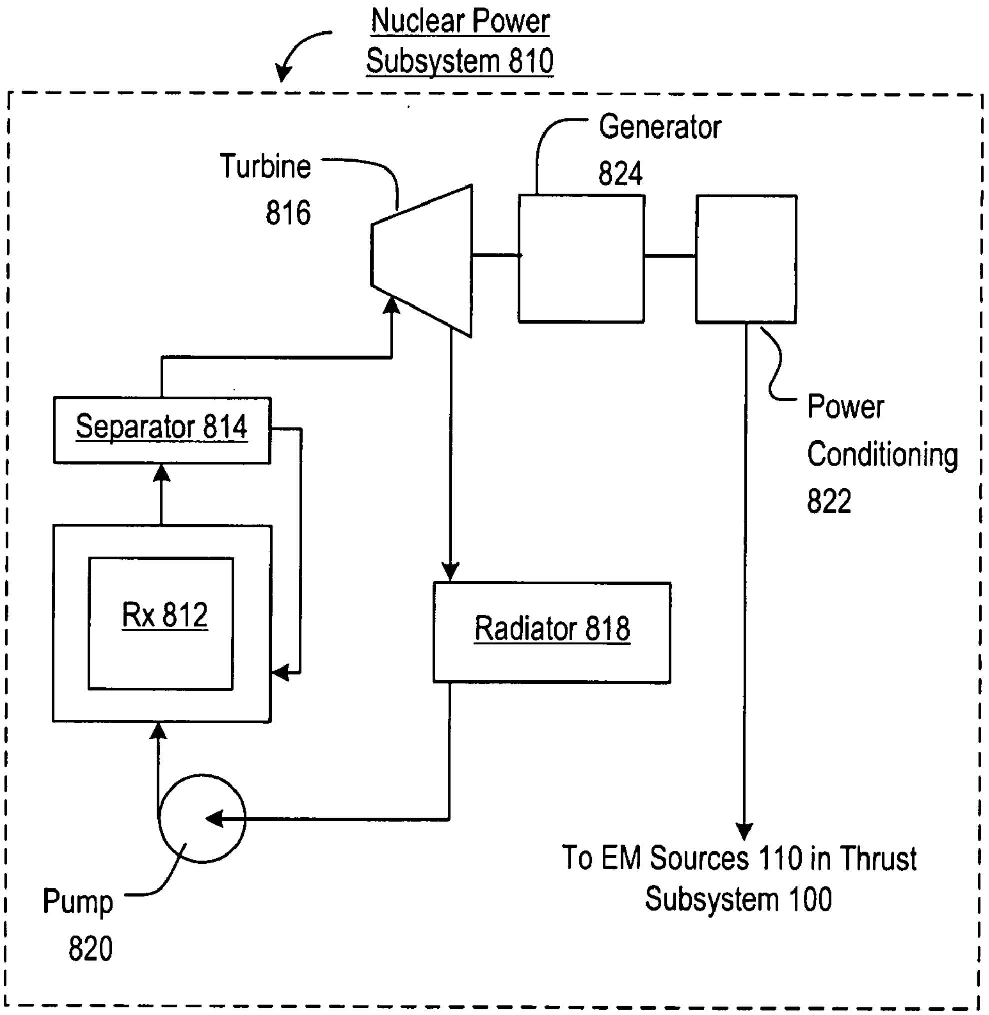
Nuclear Power Subsystem 810
Turbine 816
Generator 824
Power Conditioning 822
Separator 814
Rx 812
Radiator 818
Pump 820
To EM Sources 110 in Thrust Subsystem 100

APPARATUS COMPRISING OF PROPULSION SYSTEM

STATEMENT OF RELATED APPLICATION

[**0001**] This application claims priority of U.S. Provisional Patent Application 60/598,658, which was filed on Aug. 4, 2004 and is incorporated by reference herein.

FIELD OF THE INVENTION

[**0002**] The present invention relates generally to propulsion systems.

BACKGROUND OF THE INVENTION

[**0003**] Sixty-six years after the Wright brothers made their first, sustained powered flight, Neil Armstrong walked on the Moon. Incredible progress to be sure, but can this pace of innovation be sustained? Will we soon visit neighboring planets or the nearest stars? Can we reach these destinations with the technology that got us to the Moon? If we can't, what propulsion technologies might be able to take us to these unthinkably remote places?

[**0004**] Current propulsion technology is based on an action-reaction principle, whereby a gas is expelled at high-speed to propel a payload in the opposite direction. This technology is typically embodied as a chemical rocket engine. While a payload can be rapidly accelerated using a chemical rocket, fuel is quickly consumed to develop the required thrust. To illustrate the problem, consider that if a spacecraft could be powered to achieve a constant acceleration of only 1 g, the trip from Earth to Mars would require about 2-4 days. In fact, modern chemical rocket engines can achieve accelerations much greater than 1 g. But even at 1 g, the fuel would be exhausted within minutes. As a consequence, the trip to Mars from Earth via chemical rocket takes about six months.

[**0005**] With current chemical-rocket technology, most of the weight at launch is fuel. For example, a typical choice for a mission to Mars would involve the Boeing Delta II 7925 or 7925H rocket stages. In its common configuration, the RS-27A engine of the Delta II first stage, along with an additional nine strap-on sold rocket motors, will have a mass of about 285,000 kilograms at launch. But of this mass, only slightly more than 1000 kilograms will reach Mars.

[**0006**] As noted above, the delivery of several tons of a payload to Mars via chemical rockets is contemplated to take about six months, with total mission duration of about two to three years. For the majority of transit time, astronauts will be weightless, which is known to adversely affect the human body. Furthermore, the astronauts will be subject to exposure from harmful radiation. Additionally, the prospects of mounting a rescue or recovering from a serious malfunction are slim due to the transit times involved.

[**0007**] And the distances involved in interstellar travel are so large that with this technology, a trip to even the nearest star systems would take hundreds to thousands of years.

[**0008**] It seems clear that current technology does not provide a means to manned exploration of the Solar System or beyond. That being the case, can technological approaches be conceived that will send spacecraft from the Earth to destinations within the Solar System in a matter of days or weeks, as opposed to years or decades? Any such approach will face a daunting technological requirement. Namely, in order to drastically reduce travel time to "neighboring" planets and "nearby" stars, exceedingly large velocities must be achieved—velocities that are on the order of a significant fraction of the speed of light.

[**0009**] Proposals that meet this mission time requirement will, therefore, typically require what can only be described as "fantastic" technologies. From a feasibility perspective, perhaps the most "promising" of those technologies that have been proposed is the matter-antimatter drive. When combined, matter and antimatter will completely annihilate, releasing unfathomable quantities of energy. But even if we were able to develop a matter-antimatter drive, its use should be proscribed. The reason is that if antimatter were to leak from its containment chamber while in the vicinity of Earth, there is a distinct possibility that the resulting energy release would destroy Earth or at least cause the extinction of all life thereon.

[**0010**] Another exotic propulsion technology is the "solar sail." Although solar sails can produce momentum by reflecting a portion of the light that they receive from the sun, this approach, on its own, does not offer a solution to the problem of achieving interstellar or even interplanetary travel. More specifically, in order to deliver a space probe to a nearby star in less than a century, the sail must be driven by laser light aimed at it throughout the trip. The power requirement for the laser, which would be located on Earth, is on the order of hundreds of thousands of terawatts. For the sake of comparison, the current planetwide consumption of electricity is on the order of about 1 terawatt. And this approach has a further complication. Namely, the craft must be slowed from a non-trivial fraction of the speed of light to orbital velocity at its final destination using light that is coming from earth. This would require the coordination of very complex maneuvers that, if not carried out correctly, might result in the destruction of the ecosystem of the destination planet.

[**0011**] In the late 1990s, NASA established and funded a program, now defunct, called the "Breakthrough Propulsion Program." The program's charter was to evaluate entirely new propulsive principles that would enable interstellar or at least interplanetary travel. Technologies under consideration included the Schlicher thruster, Deep Dirac Energy, Podkletnov gravity shielding, Podkletnov force-beam, transient inertia, coupling between electromagnetism and spacetime, gravity modification schemes, anomalous heat effect, Biefeld-Brown effect, warp drives, wormholes, high-frequency gravitational waves, superluminal tunneling, the Slepian Drive, and the quantum vacuum (e.g., dynamical Casimir effect, etc.).

[**0012**] Unfortunately, none of these approaches were deemed to be promising. For example, one study pertaining to the quantum vacuum concluded that the acceleration of a spacecraft propelled by the dynamical Casimir effect would, after ten years under acceleration, be traveling at 0.1 meters per second!

[**0013**] In light of the foregoing, it seems likely that an as yet unidentified propulsion technology will be required to make routine, manned interplanetary and interstellar travel a reality.

SUMMARY

[0014] The illustrative embodiment of the present invention is a system and method for propulsion that avoids some of the drawbacks of the prior art. Unlike conventional propulsion technology, the propulsion system described in this specification does not consume fuel (although there is an energy requirement). In fact, the system does not even use fuel, as the term is commonly used.

[0015] The illustrative embodiment is grounded in accepted physics principles, albeit leading-edge theoretical, experimental and applied physics. The propulsion method does not violate basic physical "laws," such as the conversation of momentum. The equations on which the propulsion system is based are clearly established in the art, although extended to a domain of applicability and mode of use that has not been previously contemplated.

[0016] The propulsion system operates by modifying the dispersion force (i.e., van der Waals) that arises between particles, such as neutral atoms. The following two discoveries by the inventor enable the propulsion system:

[0017] (1) The dispersion force interaction between any two neutral atoms is affected by the presence of an external gravitational field in a way that results in a repulsive force upon the atomic pair.

[0018] (2) The distortion of the dispersion interaction described above, which is usually quite small and is proportional to the number of atoms present, can be magnified by many orders of magnitude. This requires that (a) the atom-atom interaction to be transformed from a relatively shorter-range interaction to a relatively longer-range interaction; and (b) a very large number of atoms are present for mutual interaction.

[0019] A method in accordance with the illustrative embodiment of the present invention comprises:

[0020] generating a lifting force by subjecting a plurality of confined particles to a trigger acceleration; and

[0021] exposing the particles to an amount of electromagnetic radiation that is sufficient to induce the lifting force to:

[0022] (i) exhibit relatively long-range interactions; and

[0023] (ii) increase the momentum of the particles; and

[0024] transferring at least a portion of the increase in momentum to a vehicle.

[0025] To begin the propulsion cycle, the particles must be subjected to acceleration, that is, the "trigger" acceleration. This can be accomplished, for example, by supporting a craft that contains the propulsion system in a gravitational field (i.e., the craft cannot be in free fall). Or, acceleration can be kinematic, such as by rotating the craft, or using a conventional propulsion system to accelerate the craft. The force is referred to as a "lifting" force because it's direction is opposite to the weight of the particles.

[0026] In the illustrative embodiment, the particles are neutral atoms, the atoms are confined in an atomic trap, and a laser provides the electromagnetic energy that is required to transform the atom-atom interaction from a short-range to a long-range interaction and to increase the momentum of the atoms.

[0027] Some of the increase in momentum of the atoms is transferred to a vehicle, such as a spacecraft. This transfer of momentum propels the vehicle. In the illustrative embodiment, this is accomplished by letting the atoms work against a piston, which in turn impacts against a part of the vehicle. Atoms that hit the piston are recycled to the atomic trap for the next propulsion cycle.

[0028] Several points of explanation or definition will be useful in understanding the illustrative embodiment of the present invention and its underlying principles.

[0029] The term "long range" interaction or force usually describes a force that decays with distance as $1/R^n$, where n is a positive number. The term "short range" interaction or force usually describes a force that decays with distance as $\exp[-R]$. Those conventions are not followed in this disclosure. Rather, for the purpose of this disclosure and the appended claims, terms that "short range" and "long range" are comparative or relative terms. For example, the language "inducing the lifting force to exhibit relatively long range interactions" means that the lifting force is induced to exhibit a relatively longer-range interaction than is normally the case.

[0030] Since the momentum that is donated by the particles propels the vehicle, there might be a tendency to characterize the particles as "fuel." But the particles are not "fuel" in any conventional sense of that term. They simply serve as a "momentum exchanging element."

[0031] It is important to recognize that the particles are not accelerated by the wall of the spacecraft but, rather, by their mutual dipole fields (van der Waals force) as distorted by the craft's acceleration (gravitational or otherwise). In other words, this method does not violate the action-reaction law, since the motion of the particles is not due to action of vehicle upon them.

[0032] An example of "action of the vehicle" on the particles is if the atoms were accelerated due to an explosion in the trap. In that case, the net of all internal forces on the system would be zero and, at the end of the process, the craft would not gain any momentum. In accordance with the illustrative method, however, the atoms accelerate towards the piston, etc., independently of the vehicle and do transfer a net amount of momentum to it during impact.

[0033] The lifting force that is "generated" by the method is not a new force. Rather, it is simply the vertical component of a known intermolecular force; in particular, the van der Waals force. This vertical component arises from an asymmetry in the van der Waals force that results from the introduction of a gravitational field (or acceleration). There would be no such asymmetry, nor vertical component of force, in flat space-time.

[0034] This result—that the interaction potential between two neutral atoms in their ground state depends upon the position of the atoms in a gravitational field—is new. Previous studies pertained to the distortion of the field of two point charges, not two dipoles (atoms). While this "new" force is measurable with presently existing technology, harnessing it, such as to lift an object, is not feasible, since this force amounts to an exceedingly small correction to the total weight of each atom.

[0035] It is useful to note that using many such atom pairs does not improve this situation, because that does not result in a larger force per atom. The reason for this is that the inter-atomic energy is a function, to a large power, of the reciprocal of distance. In other words, the atom-atom dispersion-force interaction is a realtively short-range force. As a consequence, the total "lifting" force on a large number of atoms is increased only minutely with respect to the lifting force acting on just one pair of atoms. That is, if the number of atoms is N, there will be about N pairs (for N>>1) to consider, but the mass of the system also goes up as N. So, there is no gain realized by adding atoms.

[0036] Critical to the present invention is the inventor's recognition that if this newly discovered "effective force" could be transformed from a relatively short-range interaction into a relatively long-range interaction, the lifting force that is available would be greatly increased. In particular, if particles could be made to interact over the long range, then the total energy of the system results from the interaction of every particle with all other particles present. For large groups of particles (N>>1), the interaction grows as N^2, while the total mass of the system is only growing proportionally to N. As a consequence, a large gain in energy can be realized by using large groups of particles.

[0037] A mechanism for transforming relatively short-range interactions into relatively long-range interactions was theoretically discovered several years ago and has been re-evaluated more recently as a way to introduce unusual behaviors in a cloud of trapped atoms. See, Kurizki et al., "New Regimes in Cold Gases Via Laser-Induced Long-Range Interactions," . . . . The method involves isotropic illumination of atoms by lasers. That technique, with several modifications, is utilized in conjunction with the illustrative embodiment.

[0038] As previously noted, the present propulsion system and method overcomes a key drawback of chemical engines; namely, the fact that at some point, the fuel is expended. In accordance with the illustrative embodiment of the present invention, it is possible to maintain acceleration without expelling high-speed gases. In other words, the propulsion system does not require fuel. Alternatively, if the "particles" are considered to be "fuel," then there is no consumption of fuel due to the process.

[0039] As previously mentioned, the illustrative propulsion systems and methods described herein are not energy free. In particular, to achieve the required transformations, an intense radiation field, such as can be generated by powerful lasers, must be developed throughout the region in which the particles are trapped. In the case of a craft destined for extremely long interplanetary or interstellar flights, the energy required to power the lasers is obtained, for example, from an on-board nuclear reactor, akin to the reactors powering some submarines.

[0040] The propulsion system described herein has many applications. In particular, in addition to its use as a propulsion system for spacecraft, it can be used to deliver a payload into low earth orbit without requiring orbital speeds. Furthermore, a small version of the propulsion system could be attached to literally any item (e.g., a pallet of goods, a railroad car, etc.) so that the item could be readily moved (e.g., in a warehouse, loaded onto a cargo ship, etc.) as needed. The propulsion system can, of course, also be used in conventional aircraft.

[0041] Additionally, the present propulsion system can be used to supplement a main, conventional propulsion system. In fact, this would facilitate phase-in to replace conventional technologies. For example, a propulsion system in accordance with the illustrative embodiment that is not sufficiently powered to bring a craft to a hover could be used to effectively reduce the mass of the craft, thereby improving the fuel consumption of the main propulsion system. Alternatively, it could be used as a supplemental system for emergencies.

BRIEF DESCRIPTION OF THE DRAWINGS

[0042] **FIG. 1** depicts distortion in the spherical symmetry of the field of a simple charge, as caused by the presence of a gravitational field.

[0043] **FIG. 2** depicts distortion in the cylindrical symmetry of the field of a classical dipole, as caused by the presence of a gravitational field.

[0044] **FIG. 3** depicts a propulsion system in accordance with the illustrative embodiment of the present invention.

[0045] **FIG. 4** depicts the propulsion system of **FIG. 3**, wherein lasers are illuminating confined particles.

[0046] **FIG. 5** depicts the propulsion system of **FIG. 3**, wherein particles impact against an elastically-bound piston.

[0047] **FIG. 6** depicts the propulsion system of **FIG. 3**, wherein particles are pumped back to a reservoir for use in a subsequent propulsion cycle.

[0048] **FIG. 7** depicts a method in accordance for propulsion in accordance with the illustrative embodiment of the present invention.

[0049] **FIG. 8** depicts a schematic of a vehicle that incorporates the propulsion system of **FIG. 3**.

[0050] **FIG. 9** depicts a schematic of the nuclear power subsystem of the vehicle of **FIG. 8**.

DETAILED DESCRIPTION

[0051] This Detailed Description proceeds with Section 1.1, which provides a description of propulsion system **100** and a method for propulsion in accordance with the illustrative embodiment of the present invention. Section 1.2 discloses a vehicle that incorporates propulsion system **100**. The remaining sections, which include Sections 2.1-2.4 and 4 provide a theoretical development for propulsion system **100** and performance estimates.

1.1 Propulsion System 100

[0052] **FIG. 3** depicts propulsion system **100** in accordance with the illustrative embodiment of the invention. Propulsion system **100** includes particles **102**, chamber **104**, piston **108**, source(s) of electromagnetic radiation **110**, return line(s) **112**, and pump(s) **114**, interrelated as shown.

[0053] **FIG. 7** depicts method **700** for propulsion, which can be used in conjunction with propulsion system **100**. In accordance with operation **702** of method **700**, a plurality of particles are confined.

[0054] **FIG. 3** depicts propulsion system **100** at the beginning phase of the propulsion cycle. Particles **102**, which in some embodiments are ground-state atoms, are confined in particle trap **106** of chamber **104** in known fashion and in accordance with operation **702**.

[0055] In operation **704**, particles **102** are subjected to a trigger acceleration. This can be accomplished, for example, by supporting propulsion system **100** in a gravitational field. Assuming propulsion system **100** is in a vehicle, such support is provided, for example, if the vehicle is at rest on the surface of the Earth or in flight, as long the vehicle is not in free fall. In some alternative embodiments, the trigger acceleration can be kinematic, such as by rotating the craft, or by using a conventional propulsion system to accelerate the craft.

[0056] At operation **706**, the particles are exposed to an amount of electromagnetic radiation that is sufficient to:

[0057] 1. induce an effective inter-particle force that arises between said particles to exhibit long-range interactions; and

[0058] 2. increase the momentum of the particles.

This operation is depicted in **FIG. 4**, wherein sources of electromagnetic radiation **110** ("EM sources **110**") are activated and directed toward particles **102**. In some embodiments, the EM sources are high-power lasers. In the **FIG. 4**, two EM sources **110** are shown. Depending upon the power required for a given embodiment, far more EM sources might be required. Power requirements for driving the propulsion system are described in

[0059] The EM radiation causes an upward acceleration of particles **102** with respect to a vehicle, etc., that houses propulsion system **100**. In the case of an ideal propulsion system, particles **102** remain trapped in place (in particle trap **106**) as they mutually interact and the craft is accelerated upward by the reaction of the atoms themselves against whatever forces are used to keep them in trap **106**.

[0060] It is possible, if not likely, that once particles **102** are accelerated by conducting operations **704** and **706**, they will escape from particle trap **106**. This is a non-ideal situation, which yields less than the ideal momentum. But, if particles are allowed to escape, this relaxes the constraints on particle traps **106**. That is, suitable traps can be readily constructed with existing technologies. See, e.g., H. J. Metcalf and P. van der Straten, *Laser Cooling and Trapping* (Springer, N.Y., 1999); http://ww.rle.mit.edu/cua/research/project 02/project 02.vandp.htm. These references describe techniques for trapping atoms at extremely low temperatures.

[0061] Although it is necessary for the generation of the lifting force itself, it is desirable to trap particles at very low temperatures because the thermal speed of, for example, atoms, even at room temperature, is comparable to the maximum speeds that can be obtained by this method. It is, therefore, "easier" to illuminate the atoms for an appropriate length of time if they are not moving at very high thermal speeds.

[0062] Continuing with the description of method **700**, operation **708** recites impacting the accelerated particles against a surface, thereby transferring some of the momentum of the particles to the surface. This operation is depicted in **FIG. 5**, wherein particles **102** impact piston **108** at high speed.

[0063] Piston **108** functions as a shock absorber. That is, the piston provides an area against which particles **102** can impact and which can transfer momentum to the vehicle non-destructively at every forward stroke. In propulsion system **300**, non-destructive momentum transfer is indicated by spring **116**, which elastically couples piston **108** to a vehicle. Shock absorber technology for aerospace applications is well developed within the context of pyrotechnic release technology. See, e.g., N. Butterfield, *Pyrotechnic Release Devices, in Space Vehicle Mechanisms*, P. Conley, Ed. (Wiley, N.Y., 1998).

[0064] As illustrated in **FIG. 5**, in propulsion system **100**, as piston **108** moves upward, return lines **112** are accessed. The return lines provide a route back to particle trap **106**. As particles move away from piston **108**, the piston drops back to a seated position against chamber **104**.

[0065] **FIG. 6** depicts particles **102** in return lines **112**, being pumped via pumps **114** toward a gas reservoir (not depicted) for reuse in a subsequent propulsion cycle. The propulsion cycles occur at a rapid and substantially continuous pace.

[0066] It is very important to recognize that in propulsion system **100**, and in accordance with method **700**, particles **102** are NOT being accelerated by the walls of the vehicle or by the chamber in which they reside. Rather, they are accelerated by their mutual dipole fields, as distorted by vehicle acceleration (i.e., gravitational or otherwise). There is, therefore, no concern that this scheme violates the action-reaction law, since the motion of particles **102** is NOT due to an action upon them by the vehicle.

[0067] An example of a situation in which the walls of the vehicle are acting on particles **102** is if the particles were accelerated by an explosion in chamber **104**. In such a case, the net sum of all internal forces on the system would be zero and, at the end of the process, the vehicle would not gain any net momentum. But using the methods and apparatus described herein, particles **102** are accelerated toward piston **108** INDEPENDENTLY of the vehicle and, on impact, transfer a net amount of momentum to it.

1.2 Vehicle Incorporating Propulsion System 100

[0068] **FIG. 8** depicts vehicle **800**, which incorporates propulsion system **100** in accordance with the illustrative embodiment of the present invention. As depicted in **FIG. 8**, vehicle **800** includes propulsion subsystem **100**, nuclear power subsystem **810**, crew quarters **830**, and shielding **840**, arranged as shown.

[0069] The presence of nuclear power subsystem **810** requires the use of shielding **840** to protect crew quarters **830** and propulsion subsystem **100**. Those skilled in the art will be capable of designing and building shielding suitable for this purpose.

[0070] The purpose for nuclear power subsystem **810** is to generate electricity to power EM source(s) **110**. Nuclear power is used as a power source due to the ability of a nuclear reactor to provide continuous power for extended periods of time (e.g., several years, etc.). Operation of nuclear power subsystem **810** is described in more detail below in conjunction with **FIG. 9**.

[0071] Propulsion subsystem **100** couples to shielding **840**, which receives momentum transferred from piston **108** (see **FIGS. 3-7** and the accompanying description). This substantially continuous transfer of momentum from particles **102** to piston **108** to vehicle **800** (e.g., shielding **840**) drives the vehicle.

[0072] In the embodiment that is depicted in **FIG. 8**, propulsion subsystem **100** is disposed an end of vehicle **800**. This location draws maximum advantage from a rotational trigger acceleration while providing the crew, in crew quarters **830**, with appropriate gravity-like conditions.

[0073] **FIG. 9** depicts an embodiment of nuclear subsystem **810** suitable for use to provide electricity to drive EM sources **110** (e.g., lasers, etc.) in propulsion subsystem **100**. The embodiment that is depicted in **FIG. 9** is a direct Rankine cycle continuous power system. (See, e.g., M. W. Edenburn, "Models for Multimegawatt Space Power Systems," Sandia Report SAND86-2742 (June 1990). This type of system is suitable for use with vehicle **800** due to its ability to provide continuous power for several years.

[0074] Nuclear subsystem **810** includes nuclear reactor **812**, separator **814**, turbine **816**, radiator **818**, pump **820**, generator **822**, and power conditioning unit **824**.

[0075] Reactor **812**, which is liquid metal cooled, boils potassium and sends the saturate vapor to turbine **816** for power generation. Since the fluid leaving the "hot" end of reactor **812** is unlikely to be pure vapor, separator **814** is used to separate the saturated vapor from its accompanying liquid. The liquid is recirculated to the "cold" end of reactor **812**.

[0076] Waste heat is rejected by space radiator **818**. Since the system rejects heat from a condensing working fluid, the radiator operates nearly isothermally and radiates a relatively large amount of heat per unit area. Condensed liquid is returned to the "cold" end of reactor **812** via pump **820**.

[0077] Electricity that is produced by generator **822** is appropriately conditioned in power conditioning unit **822** to provide EM sources **110** with a suitable supply of electrical power.

2.1 Distorted Dipole-Dipole Potential

[0078] It is already a well-known fact that a gravitational field can introduce novel forces acting on a single charge or on a dipole. An example is the self-interaction of a point charge in a Schwarzschild geometry [8], ultimately due to the term Linet [9] discovered has to be added to the Copson potential [10] in order to satisfy the appropriate asymptotic boundary conditions for this problem. Commenting about the very extreme conditions nearby a miniblack-hole, Smith and Will wrote that "[I]t is amusing to note that . . . the test particle's electrostatic self-force would suffice to support it against the hole's gravity, without the help of any external force."

[0079] Unfortunately, however, such fascinating conclusion is undermined by the fact that, as pointed out by these authors, "it is meaningless to talk of an electron being held fixed at, say, 10^{-13} cm from a miniblack hole, when the Compton wavelength of an electron is two orders of magnitude larger than this." Following this approach, the self-interaction of a static dipole has also been calculated [11], but Parker has shown that this force has no effect on the Hamiltonian of a neutral atom in free-fall [12].

[0080] Another example is the "electrostatic levitation of a dipole," predicted on the basis of the distortion caused by a uniform gravitational field [13]. This author found that "one is unlikely to witness such levitation," which could only be observed in a fixed classical dipole whose electron charge separation is 1.4×10^{-15} m. The outlook for detection of these field distortion phenomena was effectively summarized by Boyer, who stated that "Clearly our example may be instructive from a theoretical point of view, but it does not lend itself to easy experimental measurement."[14]

[0081] In this section, we consider the effect of a weak gravitational field upon intermolecular forces. In particular, the effect of gravitation on the van der Waals hydrogenic interatomic potential in the unretarded regime is discussed within non-relativistic first-order perturbation theory. The quantitative conclusion of these computations is that the system proposed herein shows extreme promise for direct experimental verification although the effect is certainly too small to be of any practical engineering use in the field of propulsion.

[0082] The first step to obtain the distorted dipole-dipole potential is the calculation of the electrostatic potential, and thus of the electric field, of a point dipole in the presence of gravitation. For this purpose, let us start by considering the potential field of a single point charge q located at a position $r_0=x_0^i$ in the quasi-homogeneous gravitational field caused by a relatively distant spherically symmetrical mass distribution M located at a radial distance R from the dipole. This has been the subject of several investigations, starting with the pioneering work of Whittaker [15].

[0083] Since we are considering a charge in a gravitational field g antiparallel to the z-axis and located at a position other than the origin, we transform the unprimed Rindler coordinates defined by the usual metric

$$ds^2 = \left(1 + \frac{gz}{c^2}\right)^2 c^2 dt^2 - (dx^2 + dy^2 + dz^2), \tag{1}$$

by introducing new primed coordinates given by $t=(1+gz_0/c^2)^{-1}t'$, $x^i=x_0^i+x'^i$. By substituting these definitions into Eq. (1), it is simple to show that the metric in the new coordinates is:

$$ds^2 = \left(1 + \frac{G_R z'}{c^2}\right)^2 c^2 dt'^2 - (dx'^2 + dy'^2 + dz'^2), \tag{2}$$

where $G_R \equiv g/(1+gz_0/c^2)$. With this result, we can transform Whittaker's expression for the electrostatic potential in Kottler-Whittaker coordinates into our transformed Rindler frame (for simplicity of notation, we neglect to write the primes in what follows):

$$V(r; r_0) = \frac{q}{|r - r_0|} \frac{\left[1 + \frac{g(z + z_0)}{c^2} + \frac{g^2}{2c^4}(|\rho - \rho_0|^2 + (z^2 + z_0^2)\right]}{\left(1 + \frac{g z_0}{c^2}\right)\sqrt{1 + \frac{g}{c^2}(z + z_0) + \frac{g^2}{4c^4}|r - r_0|^4}} \tag{3}$$

where $|r-r_0|^2=(x-x_0)^2+(y-y_0)^2+(z-z_0)^2$, $|\rho-\rho_0|^2=(x-x_0)^2+(y-y_0)^2$. This expression of course approaches the Coulomb potential as gx^i/c^2, $gx_0^i/c^2 \to 0$. In this paper, we shall neglect the Linet term [9] responsible for the self-interaction discussed above since this contribution will be shown to be negligible with respect to the effects treated herein.

[0084] In order to write down the electrostatic dipole field in the presence of a gravitational field, we use the formulation of Léauté and Linet, originally designed to calculate the self-interaction of an electric dipole [11]. For a point dipole A of moment $d_A=d_A^i$, with k=1 . . . 3, located at r_A, the result found by these authors, neglecting the self-interaction term [12], can be written as:

$$U_{dip,W}(x^k, x_A^k) = d_A^i \left.\frac{\partial V_C(x^k, x_0^k)}{\partial x_0^i}\right|_{r_A}, \tag{4}$$

where V_C is the Copson potential, which, in the quasi-homogeneous field limit, coincides with our solution above.

[0085] Since, to the best of this author's knowledge, this potential has never been graphically represented, it is shown, along with the corresponding electric field lines, in **FIG. 3** for a point dipole $d_A^i=d_Ak$.

[0086] By computing the electric field as usual, after some very lengthy algebra [13] one obtains the general expression for the interaction potential energy $W_{dd}(r, r_0; d_A^i, d_B^i)$ of two point dipoles of moments d_A^i and d_B^i, placed at positions r_0 and r, respectively. In order to illustrate the physical meaning of this result, let us write it to second order in gR/c^2 for the case of dipole A placed at the origin and dipole B placed at r=(R, 0, 0):

$$W_{dd} \approx \frac{q^2}{R^3}\left[(x_Ax_B + y_Ay_B - 2z_Az_B) + \frac{1}{2}\left(\frac{gR}{c^2}\right)\left(1 - \frac{gR}{c^2}\right)(x_Ax_B + y_Ay_B)\right], \tag{5}$$

which again yields the usual Minkowski space result [17] if the gravitational field is absent. Interestingly, the asymmetry due to the presence of the gravitational field causes a net vertical force upon each dipole in this geometry where there would of course be none in flat space-time. For instance, in the case of two antiparallel dipoles $d_A^i=-d_B^i=d_Ak$, we find that this force $F_{dd}=-\partial W_{dd}/\partial z$ is, again to second order:

$$F_{dd,z} = \frac{3}{2}\frac{q^2}{R^3}\frac{g}{c^2}z_A^2, \tag{6}$$

which is the dipole-dipole analogy of the levitation force upon a single dipole mentioned at the beginning of this section [15].

[0087] The problem of a single hydrogenic atom either held fixed or in free-fall within various assigned metrics has been discussed extensively by starting from generally covariant expressions of the Dirac equation in curved space-time [18]. The present author has discussed realistic astrophysical settings for the observation of the perturbative effects of gravitational fields on freely-falling atoms both in the static case and within the framework of possible remote gravitational wave detection [19].

[0088] Tourrenc and Grossiord, in their treatment of a hydrogen atom held fixed in a Schwarzschild geometry, have shown that the most significant contribution to the perturbative Hamiltonian by far derives from what can be interpreted as the classical weight of the electron in the gravitational field. However, since the corresponding energy shift for a ground-state atom is found to be $\Delta E \sim 2(GMm^e/R^2)a_0 \approx 6\times10^{-21}$ eV, we shall neglect it here and use the unperturbed hydrogenic wavefunctions to evaluate the interatomic gravitational self-force.

[0089] As in the undistorted van der Waals case, the general expression for the potential $W_{dd}(r, r_0; d_A^i, d_B^i)$ contains only bilinear forms of the type $x_A^ix_B^j$ and thus yields no first-order contribution in the case of symmetrical hydrogenic nS. The second order correction on the other hand is [20]:

$$\Delta E_{vdW}^{(2)} = \sum{}' \frac{|\langle\phi_{n,l,m}^A; \phi_{n',l',m'}^B|W_{dd}(r, r_0; d_A^i, d_B^i)|\phi_{1,0,0}^A; \phi_{1,0,0}^B\rangle|^2}{-2E_I - E_n - E_{n'}}, \tag{7}$$

where $|\phi_{n,l,m}^A\rangle$ are the unperturbed eigenfunctions of energy E_n, the total energy of the unperturbed atomic pair is $-2E_I$, and the primed summation indicates that the $|\phi_{1,0,0}^A; \phi_{1,0,0}^B\rangle$ term is excluded.

[0090] By inspecting Eq. (5), it is evident that, in the gravitational case, the second order intermolecular potential, to second order in gx^i/c^2, takes on the general form:

$$\Delta E_{vdW}^{(2)} = \frac{C_0}{R^6} + \left(\frac{g}{c^2}\right)\frac{C_1}{R^5} + \left(\frac{g}{c^2}\right)^2\frac{C_2}{R^4}, \tag{8}$$

where C_0, C_1, and C_2 are appropriate dimensional constants, which in principle depend on the variables r, r_0, d_A^i, and d_B^i. For instance, again in the geometry used in the examples above, one can quickly recover the well-known result that $C^0 \approx -6e^2a_0^5$ as well as show that $C_1 \approx -2e^2a_0^5$, and

$$C_2 \approx +\frac{3}{2}e^2a_0^5,$$

where a_0 is the Bohr radius.

[0091] In our case, however, we are not here interested in the modification of the intermolecular forces due to the presence of gravitation, but rather we want to pursue isolating the vertical component of the gravitational self-force due to the dipole-dipole interaction of two hydrogenic atoms in the $|1,0,0>$eigenstate located at $r_0=0$ and $r=(R, 0, 0)$, respectively:

$$F_{sf,z}^{(2)} = -\frac{\partial}{\partial z}\Delta E_{vdW}^{(2)} \approx -\frac{\left|\left\langle \phi_{1,0,0}^{A}; \phi_{1,0,0}^{B}\right| \frac{\partial}{\partial z} W_{dd}(r, r_0; d_A^i, d_B^i)\left|\phi_{1,0,0}^{A}; \phi_{1,0,0}^{B}\right\rangle\right|^2}{-2E_I}, \tag{9}$$

Explicit evaluation [13] yields the following result to first order in gx^i/c^2:

$$F_{sf,z}^{(2)} \approx \frac{e^4}{2E_I R^6}\frac{g}{c^2}\langle\phi_{1,0,0}^{A}; \phi_{1,0,0}^{B}|(4x_A^2x_B^2 - 4x_Ax_By_Ay_B + y_A^2y_B^2 - 3z_A^2x_B^2 - 14x_Ax_Bz_Az_B + 4y_Ay_Bz_Az_B - 3x_A^2z_B^2 + 3z_A^2z_B^2)|\phi_{1,0,0}^{A}; \phi_{1,0,0}^{B}\rangle. \tag{10}$$

By again making use of the fact that the cross terms vanish in the 1s state and that $<\phi_{1,0,0}{}^{A}|x_A{}^{2}|\phi_{1,0,0}{}^{A}>=<\phi_{1,0,0}{}^{A}|\tfrac{1}{3}r_A{}^{2}|\phi_{1,0,0}{}^{A}>$ and equally for all the other squared terms, we finally find:

$$\begin{aligned} F_{sf,z}^{(2)} &\approx +\frac{e^4}{2E_I R^6}\frac{g}{c^2}\times 2\left|\left\langle\phi_{1,0,0}^{A}\right|\frac{1}{3}r_A^2\left|\phi_{1,0,0}^{A}\right\rangle\right|^2 \\ &= +2\frac{e^2a_0^5}{R^6}\frac{g}{c^2} \\ &= +\frac{1}{16}\hbar\omega_0\frac{\alpha_0^2}{R^6}\frac{g}{c^2}, \end{aligned} \tag{11}$$

where the last equality was obtained by writing the ionization energy and the polarizability as $E_I=e^2/2a_0$ and $\alpha_0=(2a_0)^3$, respectively. Notice that this result does not coincide with what one might expect from näively calculating one half of the negative contribution to the gravitational mass of the London binding energy of the pair, since in that case the coefficient would be ⅜ and not 1⁄16.

[0092] An explicit estimate of this result in the case of two hydrogen atoms in their ground state at, for instance, 20 a_0, yields a relative acceleration $a_{lift,H}$~4×10^{-13} cm/s^2. The situation improves dramatically if one considers two positronium (Ps) atoms, in which case we find $a_{lift,Ps}$~8×10^{-10} cm/s^2, which is in principle detectable via atomic interferometry, at the price of dealing with the added complications of interferometry of atoms with a finite lifetime.

2.2 Relevance of the Above Results

[0093] The above results show that the distortion of the classical dipole-dipole field caused by gravity results in a net force upon each dipole, which is anti-parallel to the weight of each atom. This phenomenon has been known for some time, although its implications at the quantum level have not been fully explored and engineering implications do not appear to have attracted their due attention. Interestingly, the warping of the Coulomb field due to gravitation renders their mutual interaction non-central, which in turn implies that Newton's action-reaction law will not be satisfied by this system. Although there is no study of this exotic problem in the literature, it is clear that, if the problem were to be considered from the standpoint of full quantum electrodynamics (QED), it would result that the virtual photon field responsible for the charge-charge interaction is perturbed by the gravitational acceleration so as to carry a net momentum flux, part of which is simply transferred to the dipole, or dipoles, thus resulting in their upward motion.

[0094] Importantly, this phenomenon makes the interaction energy between two hydrogen atoms in their ground state dependent upon their position within the gravitational field, which results in a force upon the pair even in the quantum case. The existence of this additional force acting on an atomic pair is by itself a new result, since previous studies had concentrated on the distortion of the field of two point charges, and not of two dipoles. Despite the fact that this new and additional force is measurable with presently existing technology, its use in an actual lifting device is unlikely, since it amounts to a very small correction of the total weight of each atom.

[0095] The important consideration, critical to the present invention, is that adding many atomic pairs does not result in a larger force per atom. The reason for this is that the interatomic energy is a function of a large power of the reciprocal of the distance—which results in the atom-atom dispersion interaction being a short-range force. Therefore, the total lifting force on a large number of atoms is increased only very slightly with respect to that acting on just one pair. Therefore, if the number of atoms is N, there will be ~N pairs to consider but the mass of the of the system also went up as N. Thus no gain is made.

[0096] The critical, and non-obvious, point of the present invention is to transform the new force outlined above from a short-range force into a long-range force, as we consider in the next section. In principle, if point-like particles interact via a long-range force, the total energy of the system results form the interaction of every atom with all the others, something which is not possible in the short-range. For instance, if a system is made up of three particles, the total energy will result from the interaction of particle 1 with 2, particle 1 with 3, and particle 2 with 3. However, if the number of particle is now ten, we will have to consider the interaction of particle 1 with 2, 3 . . . , 10, and so fort, which results, in the case of a large number of particles. For large N (N>>1), in this case the number of interaction grows as N^2, while the total mass of the system is still only growing proportionally to N.

[0097] This point would only be of philosophical importance if we did not have available a mechanism to indeed transform atom-atom interactions from short-range into long-range. Such mechanism was discovered theoretically several years ago and has been reproposed recently as a way to introduce unusual behaviors in a cloud of trapped atoms. Atomic traps have become one of the hottest subjects of scientific research in recent years. The method proposed by the researchers to cause the atom-atom interaction to become a long-range force can of course also be used in our case to leverage the presence of a large number of atoms so as to make the lifting force due to the gravitational distortion we have seen above much larger by many orders of magnitude. The technological price to pay is that, in order for this transformation to occur, powerful lasers must be pumping an intense radiation field throughout the region where the atoms are trapped.

2.3 Trapped Gases in Curved Space-Time: the Effect of Radiation Fields

[0098] It is well-known that an intense directional radiation field, such as that produced by a laser, alters the nature of intermolecular forces [21]. For instance, in the near zone region, where $k_L R << 1$ and k_L is the laser light wavenumber, the power dependence of the force becomes $\equiv -1/R^3$. Importantly, if a molecular pair is allowed to "tumble" with equal probability in all directions with respect to the radiation field, the unretarded force averages out and the only term left is that due to the retarded part of the Hamiltonian. The resulting contribution, which can be produced by appropriate laser beams [22] and is still attractive, is [21]:

$$\Delta E \simeq -\frac{44}{15}\left(\frac{I\pi k_L^2}{c}\right)\frac{\alpha^A(k_L)\alpha^B(k_L)}{R}, \quad k_L R \ll 1, \tag{12}$$

where a $\alpha^{A,\,B}(k_L)$ are the dynamic polarizabilities of the two atoms, I is the intensity of the beams, and the domain of validity of this result is everywhere in space except where the atom-atom exchange interactions become important. It is simple to see numerically that this energy is much smaller than the usual van der Waals force at near range but, as originally pointed out by Thirunamachandran [21], the long-range nature of the force offers the potential to actually achieve remarkable effects.

[0099] Let us now consider the distortion of Thirunamachandran's long-range interaction due to an external gravitational field, such as that present in a ground-based laboratory. One is fully justified by both our results in the unretarded case and by dimensional considerations to assume that a molecular pair interacting through the gravity-like attractive long-range force at Eq. (12) will also undergo a lifting force of the type:

$$F_{lift,z}^{AB} \sim \left(\frac{Ik_L^2}{c}\right)\frac{\alpha^A(k_L)\alpha^B(k_L)}{R}\frac{g}{c^2}. \tag{13}$$

For the purpose of order of magnitude estimation, let us write the total potential energy of N atoms contained in a spherical volume of diameter D interacting through a mean field determined by Eq. (12) simply as $U_{gas} \sim -N^2(Ik_L^2/c)\alpha^A(k_L)\alpha^B(k_L)/D$. Therefore, the total lifting force acting upon the center-of-mass of the trapped gas is $F_{lift,z}^{gas} \sim N^2 F_{lift,z}^{AB}$, assuming all atoms to be of species A, and the corresponding acceleration is:

$$\begin{aligned}\frac{a_{lift,z}^{gas}}{g} &= \frac{F_{lift,z}^{gas}}{Nm_A g} \\ &= N\left(\frac{Ik_L^2}{c}\right)\frac{[\alpha^A(k_L)]^2}{D}\frac{1}{m_A c^2}.\end{aligned} \tag{14}$$

The relevant figures of merit to judge the feasibility to bring such an atomic cluster to a hover are the number of atoms in the trap and its size, the intensity and wavelength of the laser light, and the average intermolecular distance. Consider $N=10^2$ atoms in a trap with $D \approx 2\times10^{-7}$ cm, which yields an intermolecular distance $R \sim D/N^{1/3} \approx 10a_0$. Now let us have 18 (six triads [22]) high-power lasers each outputting 2.5 kW into a 0.5 cm diameter beam at a wavelength $\lambda_L \approx 1000$ Å. Appropriately focused onto the trap size, this would yield an intensity $I \approx 1.3\times10^{17}$ W/cm^2. By moderate off-resonance detuning, it is possible to obtain dramatic increases in atomic polarizability over its static value [23] (another approach may consist of using a cold gas of highly excited Rydberg atoms, since, in this case, the polarizability is proportional to n^7, where n is the principal quantum number). By adopting $\alpha^A(k_L) \approx 3\times10^5\alpha_0$, and by substituting the above numerical values into Eq. (14) in c.g.s. units, we find $a_{lift,z}^{gas}/g \approx 1.5$, that is, the system will hover unsupported in the gravitational field of the earth or accelerate upward.

2.4 Relevance of the Above Results

[0100] The mechanism outlined above represents the first novel and realistic proposal to achieve lift in the history of flight since the great inventions of the airplane and of the rocket in the 20th century. As we shall see in the detailed numerical estimates below, more atoms than just 10^2 must be present in the trap for this mechanism to be technologically convenient, although it is possible to trade off a higher laser power for a lower number of atoms in the trap. What is important to stress at this time is that this mechanism represents a non-obvious use of well understood quantum laws in the presence of gravitation for the purpose of creating fuel-free propulsion. Of course fuel-free propulsion does not imply energy-free propulsion. In other words, a source of energy is still needed to achieve the needed thrust. In the case of extremely long interplanetary and interstellar flights it is expected that on-board nuclear energy production will continue to grow in engineering importance, given the fact that no other source has been able to achieve similarly convenient power outputs.

[0101] The above invention, however, removes the greatest problem in the way of achieving fuel-free propulsion, that is, the fact that, independently of the energy source used, at some point the fuel available on board is exhausted. For instance, if one could power a spacecraft so as to achieve a constant acceleration of 1 g, the trip from Earth to Mars would require approximately 2-4 days. Although the acceleration of 1 g is smaller than the larger accelerations achievable by present-day rocket technologies, it is absolutely impossible to maintain those accelerations for times longer than minutes at the most, simply because of quick fuel exhaustion. With the scheme outlined in this invention, on the other hand, it is possible to maintain the needed acceleration without any need to expel high speed gases, provided that the required laser illumination is constantly at work transforming the atom-atom interactions from short-range into long-range ones.

[0102] It is important to stress another characteristic of the propulsive method of this invention. All the calculations carried out so far imply that the atoms are "at rest," that is, not in "free-fall." This is extremely important, since it is only if the atoms are somehow supported by an external force against the gravitational field that a relative acceleration due to that field can affect their mutual interactions. Such is not the case if the atoms are freely falling. In that case, in fact, the acceleration felt by a freely falling object is rigorously zero, because of the Principle of Equivalence at the foundation of the General Relativity theory. To use a well-known popular example, if two atoms are at rest with respect to the walls of a freely falling elevator, they will not feel the presence of any gravitational field—the acceleration of the elevator exactly cancels exactly the gravitational acceleration. In other words, locally, there is no distortion of the dipole-dipole field and thus no change to the van der Waals force (see below for further subtle clarifications on this point).

[0103] The Principle of Equivalence can be stated by saying that, locally, there exists no experiment that can indicate the difference between the acceleration due to the presence of a gravitational field and that due to the kinematics of the system. For instance, once could simulate the presence of a gravitational field in the same elevator travelling through outer space by simply accelerating it "upwards" at the same rate as the free-fall acceleration it would have in the gravitational field to simulate. In fact, our Eq. (1) above was obtained exactly by making this assumption. Therefore, the atoms will undergo the lifting force at the basis of this invention whether or not there is a gravitational field against which to lift—their behavior is due to their being within an accelerated reference frame no matter what the reason for the existence of such frame.

[0104] In principle, this represents an operational limitation of the present invention, in the sense that, if a spacecraft were to be left to freely-fall in the gravitational field of a massive body, the lifting mechanism would not be operating. For this purpose, the vehicle must be provided with an initial acceleration through other means in order for the thrust cycle described below to commence. For instance, this happens if the craft is at ground level initially, or somehow hovering under the action of an external force. Once the cycle starts, it is only necessary to coordinate the laser illumination of the (n+1)-th cycle to occur during the transfer of momentum due to the atoms that were accelerated during the n-th cycle. The dipole-dipole field during that time will behave as though under the effect of a gravitational field of that acceleration, because of the Principle of Equivalence.

[0105] From the practical standpoint, it is appropriate to stress that every propulsive or lifting system has an appropriate envelope of performance which, is exceeded or not met, will result in insufficient or abnormal behavior. For instance, the lifting force due to the wings of an airplane will cease to be effective if the airflow detaches from the wings because of a stall condition. Therefore, pilots are trained to operate so as to remain well clear of the conditions that might lead to a stall of the airfoils, such as, for instance, excessively low speed for a fixed wing aircraft. Similarly, it is expected that the propulsive system of this invention, in its simplest embodiment, must be operated under appropriate conditions of initial acceleration, in both magnitude and direction. This is much less complicated than it may appear. For instance, in outer space, an initial acceleration can be imparted by causing the entire spacecraft to rotate around an axis by means of a reaction wheel. Once the entire vehicle is rotating like a rigid disk, from the standpoint of the atoms in the propulsive subsystem the rotational acceleration will be indistinguishable from that caused by gravity. Once the thrust cycle is successfully started, the spacecraft can then be despun while the engine provides its own acceleration. In a fixed wing application, it might be possible to make use of the lift provided by the wings themselves in order to start the process.

[0106] Finally, it is also important to stress that the Principle of Equivalence only rigorously applies to a volume of space that is infinitely small. Therefore, there exists some distortion of the dipole-dipole field even in the case of freely-falling atoms, although this distortion is far smaller than that of supported atoms, in the sense described above. The effects of this distortion were studied by this author in several papers (see for instance [19]) and it is therefore possible that lift might be obtained in some embodiments even if the craft is initially in free-fall.

4. Performance

[0107] Estimating the performance of an aerospace propulsion system based upon entirely novel physical principles naturally presents some difficulties. For instance, the typical concept of specific impulse [29] is undefined in the case in which thrust is obtained without the ejection of high speed gases. At the same time, since the approach calls for the use of high power lasers to engineer the atom-atom interactions into a long-range force, it is of interest to determine whether the thrust thus obtained is in fact larger than that which would be obtained if the laser power utilized were, for instance, projected from the spacecraft into a particular direction in space or whether a laser beam of the same power were to be aimed at a hypothetical laser sail on the spacecraft [3]. In the following subsections we obtain some important order of magnitude estimates both in equation and in graphic form of a few important quantities in order to gain a more realistic understanding of the potential capabilities of a vehicle propelled by means of the physical principle of the present invention.

[0108] The conclusions below will clearly establish that the propulsion concept of this invention offers great potential from the standpoint of realistic engineering applications, although such parameters as the exact laser wavelength and power, trap size, atomic mass, and number of atoms of course will have to be optimized according to both accurate theoretical modeling and prototype testing.

[0109] In what follows, in order to make a firm connection between the theoretical treatment, which was developed here in the c.g.s. system (centimeter-gram-second), and the more typical engineering M.K.S. units (meter-kilogram-second), the cgs or MKS subscripts will be appended as appropriate. If no subscript is used, the quantity should be assumed as expressed in the cgs system. No use is made if English units throughout (such as, for instance, lbf for thrust). Also, for improved legibility, all order of magnitude signs will be replaced by equal signs.

4.1 Fundamental Equations

4.1.1 Atomic Physics of Trapped Atoms in the Accelerated Propulsive System

[0110] Let us consider a gas of N_A identical atoms of mass m_A, polarizability $\alpha_A^2(k_L)$, confined within an appropriate trap of such dimensions as to correspond to an average interatomic distance R. In what follows, we shall assume that the number of atoms, N_A, the size of the trap, D, and the average interatomic distance, R, are related simply as D~

$\bar{R}N_A^{1/3}$. In addition, Thirunamachandran's theory of dispersion forces under the effect of illumination also requires the constraint that $\lambda_L >> \bar{R}$[21].

[0111] The polarizability $\alpha_A^2(k_L)$ can be made several orders of magnitude larger than its static value, $\alpha_0=(2a_0)^3$, where Bohr's radius is $a_0=\hbar^2/\mu_e e^2$ and μ_e is the reduced electron mass ($\mu \approx m_e$), by choosing an appropriate near-resonance wavelength. Without getting into the details of atomic physics calculations, in this section we shall rely upon the well-established theoretical and experimental fact that such near-resonance condition can be satisfied, that is, the polarizability can be made larger than the static value by a factor, α_{nr}, which can be as large as $\alpha_{nr} \sim \mathbf{10}^5$ [23].

[0112] Another strategy to produce values of the polarizability that are vastly larger than the static value is to use Rydberg atoms. A qualitative argument in favor of this choice is that, as we have seen above, the static polarizability is $\alpha_0=(2a_0)^3$, where a_0 is Bohr's radius. If the atom is in a Rydberg state, that is, in an excited state with relatively large principal quantum number n>>1, the atomic radius can be replaced by $a_n=a_0 n^2$. On the strength of this argument alone, the atomic polarizability of a Rydberg state appears proportional to n^6. In fact, if one accounts for all states with different orbital quantum number l, the static polarizability of a Rydberg atom results proportional to n^7. For instance, the atomic radius of atoms in one-electron Rydberg states with $n \sim 10^2$, which are routinely created in the laboratory and are present in interstellar space, is $\sim 10^4 a_0$—similar to the size of an Ebola virus! Under these circumstances, the static polarizability is a stunning $\sim 10^{14}$ times larger than its static value [19, 30, and Refs. therein].

[0113] From the practical standpoint, it is important to notice that Rydberg atoms gases have already been "frozen" and trapped in order to study, among others, the very dipole-dipole interactions we discussed at the very beginning of this disclosure [31]. The atoms themselves are prepared by causing them to absorb laser light of wavelength appropriate to induce a radiative transition to the desired excited state. Noticeably, the radiative lifetime of Rydberg atoms can be quite long, even compared to the relatively short crossing time within the propulsive system. The choice of atoms in states other than the ground state, such as Rydberg atoms, imposes an additional constraint upon the interatomic distance, since the radius of these atoms can be macroscopic. It is therefore important to require that the interatomic distance be much larger than the Rydberg atom radius, which in turn affects the size of the entire propulsive system.

[0114] It is appropriate at this time to again stress the difference between the atom-atom force, which is related to the dependence of the dispersion energy on the interatomic distance and is a central force, and the vertical lifting force acting upon all atoms as a consequence of the modification of their interaction potential in an accelerated reference frame. For instance, let us again consider the potential energy of a pair of atoms (Eq. (12)) and let us rewrite it to highlight the close similarity to the gravitational case:

$$\Delta E = -G_{QED}\frac{\alpha_A(k_L)\alpha_B(k_L)}{R},\quad k_L R \ll 1, \tag{12a}$$

where the equivalent quantum-electro-dynamical "gravitational constant" is, by definition:

$$G_{QED} \equiv \frac{176\pi^3}{15c}\frac{I}{\lambda_L^2}, \tag{15}$$

and the dynamic polarizabilities play the role of the "gravitational mass." The total potential energy can then be written as usual:

$$U_{gas} = -\frac{1}{2}G_{QED}\sum_{i \neq j}\frac{\alpha_i(k_L)\alpha_j(k_L)}{R_{ij}}, \tag{16}$$

where the summation is meant over all pairs and R_{ij} is the magnitude of the interparticle distance vector, R_{ij}.

[0115] Every atom in the gas is acted upon by a gravitational-like self-force and thus undergoes an acceleration g_{QED} towards the center of the cloud (assumed approximately spherical) similar to a typical gravitational acceleration. This is in complete analogy to the collapse of a "cold" gas sphere taking place whenever the gravitational pressure is vastly larger than any opposing gas pressure gradient. The order of magnitude of the time required for the entire cloud to collapse to its center is an important characteristic time in stellar evolution and in stellar dynamics, and it is referred to as the free-fall time [32-33]. In order to generalize this quantity to our case, let us consider the equation of motion of a particle (atom) in the above potential at a distance r from the center of the cloud with $\alpha_A(k_L)=\alpha_B(k_L)$ and $N_A>>1$. In this case it is well-known from elementary mechanics that Gauss' Theorem allows us to only consider the force exerted by the atoms inside a sphere of radius equal to r and we find:

$$\ddot{r} = -G_{QED}\frac{\alpha_A^2(k_L)}{r^2}\frac{1}{m_A}. \tag{17}$$

Since we are only estimating an order of magnitude, let us assume, as done typically, that the acceleration is approximately constant during the free-fall process. Thus:

$$\frac{1}{2}\ddot{r}t_{ff}^2 \sim D,$$

which yields, for r~D

$$t_{ff} \sim \sqrt{\frac{2D}{\ddot{r}}} \sim \sqrt{\frac{2N_A\bar{R}^3}{G_{QED}\alpha_A^2(k_L)m_A}}. \tag{18}$$

Finally, by substituting Eq. (15) into this result, we obtain:

$$t_{ff} \sim \sqrt{\frac{2N_A\overline{R}^3}{\alpha_A^2(k_L)}\frac{15c}{176\pi^3}\frac{\lambda_L^2}{I}m_A}\,. \tag{19}$$

The importance of this result lies with the fact that the physics of the system after such free-fall time must be expected to be substantially different than in its initial state. For instance, in the case of atoms in their Rydberg states, a drastic evolution of the system towards a higher density configuration can be expected to result in the transformation of the gas into a neutral plasma, with consequent complete loss of thrust.

[0116] The condition to be required so that the atoms do not have the time to evolve into an extremely different, and technologically useless, state is that the free-fall time above be much longer than the time the atoms spend in the trap before the lifting force causes them to be ejected, Δt_A. Evidently, if the gas evolves into, for instance, a plasma upon ejection, that is of no consequence to the momentum it will transfer to the vehicle. By using the result below at Eq. (40), we can write this requirement as:

$$t_{ff} \gg \Delta t_A, \tag{20}$$

or

$$\sqrt{\frac{2N_A\overline{R}^3}{\alpha_A^2(k_L)}\frac{15c}{176\pi^3}\frac{\lambda_L^2}{I}m_A} \gg \sqrt{\frac{\overline{R}N_A^{1/3}}{\left[\frac{4\pi^2}{c^3}N_A\left(\frac{I}{\lambda_L^2}\right)\frac{\alpha_A^2(k_L)}{\overline{R}N_A^{1/3}m_A}-1\right]g}}, \tag{21}$$

If the lift acceleration is much larger than g, the term in square brackets in the denominator of the right-hand-side will become much larger than unity so that term can be neglected and we easily find the limiting condition:

$$N_A\frac{Dg}{c^2} \gg \frac{22}{15}\pi. \tag{22}$$

This condition, which of course results also by requiring that $a_A>>\ddot{r}$, can be satisfied by realistic values of the geometry of the propulsive system, although it clearly points to the usefulness of employing larger traps, for which $N_A>>1$.

[0117] That the physical state of the gas can in fact be extreme after a few free-fall times if this is not accomplished, can be seen by writing the condition that the gas be in equilibrium under the action of this gravity-like interaction. As well-known, gravitationally bound systems do not display what can be properly referred to as equilibrium configurations in the thermodynamical sense. This is well illustrated by an appropriate similitude between the cold gas in the propulsive system of this invention and a globular star cluster—a spherical system in which thousands to hundreds of thousands of stars are bound by their mutual gravitational interaction—or, alternatively, a star [22].

[0118] In a globular cluster (or in a star), the distribution of velocities of the constituent particles “relaxes” to a quasi-Maxwellian velocity distribution after a time properly referred to as the “relaxation” time of the object. However, a Maxwellian velocity distribution contemplates a finite number of particles whose speeds at any given time are higher than the escape velocity from the system. Therefore there occurs a process of constant evaporation, which clearly forbids the existence of any equilibrium configuration [32-35].

[0119] However, it is possible to define a condition of quasi-equilibrium, in which a gravitationally bound system does not change drastically over many relaxation times. Under these conditions, the object obeys a general theorem referred to as the virial theorem, which links its total average kinetic and potential energies. This connection is very powerful, as it allows one to obtain estimates of the average speeds of its particles, whether they be stars or atoms [35-36]. For the purposes of our estimates here, this very general theorem can be written as

$$2<K_{gas}>+<U_{gas}>=0, \tag{23}$$

where K_{gas} is the total kinetic energy and the triangular brackets indicate the time-average. By writing the total kinetic energy as

$$\begin{aligned} K_{gas} &= \frac{3}{2}N_Ak_BT_{gas} \\ &= \frac{1}{2}N_Am_A\langle v_A^2\rangle, \end{aligned}$$

and by approximating <1/D>=1/<D>, the virial theorem yields:

$$T_{gas} = \frac{8\pi^2}{3k_Bc}N_A\frac{I}{\lambda_L^2}\frac{\alpha_A^2(k_L)}{D}, \tag{24}$$

and

$$v_A = \sqrt{\frac{8\pi^2}{3c}\frac{N_A}{m_A}\frac{I}{\lambda_L^2}\frac{\alpha_A^2(k_L)}{D}}, \tag{25}$$

The physical significance of these equations can be found by imposing the condition that the average kinetic energy be equal to the ionization potential of the atoms E_I in the gas so that the atoms may become ionized as they collide, in analogy to the Saha and Boltzmann equations of stellar astrophysics [33]:

$$\frac{3}{2}k_BT_{gas} \sim \frac{E_0}{n^2}, \tag{26}$$

where the right-hand-side of the above equation contains the ionization potential of a hydrogen atom in its n-th state and E_0 is the ionization energy of the ground state. By using Eq. (24), this yields:

$$\frac{4\pi^2}{c}N_A\frac{I}{\lambda_L^2}\frac{\alpha_A^2(k_L)}{D}\sim\frac{E_0}{n^2}, \tag{27}$$

which can be expressed in even more fundamental terms in the case of Rydberg atoms as:

$$\frac{4\pi^2}{c}N_A\frac{I}{\lambda_L^2}\frac{(2a_0)^6n^{14}}{D}\sim\frac{e^2}{2a_0n^2}, \tag{27}$$

or

$$\frac{4\pi^2}{e^2c}N_A\frac{I}{\lambda_L^2}\frac{(2a_0)^7n^{16}}{D}\sim 1. \tag{28}$$

In the near-resonance case we can instead write:

$$\frac{4\pi^2}{E_0c}N_A\frac{I}{\lambda_L^2}\frac{(2a_0)^6\alpha_{nr}^2}{D}\sim 1, \tag{29}$$

Replacing the numerical values we shall produce below in a few realistic examples, it is immediate to conclude that the almost immediate ionization of the entire atomic population is highly likely in all technologically meaningful cases, and that, as already pointed out in this section, it is imperative to choose parameters that ensure the expulsion of the gas from the trap well before this process is completed.

4.1.2 Validity of the Present Approach

[0120] In this subsection we concern ourselves with the possible limitations of our treatment. Let us first of all notice that we have so far dealt with atoms in the trap as "classical" objects, that is, as material particles whose positions and velocities can be likened to those of stars in a cluster, for instance. This is approximately correct only if the de Broglie wavelength λ_A of the atoms is much smaller than the interatomic distance, that is, if:

$$\lambda_A\equiv\frac{h}{p_A}\approx\frac{h}{m_Av_A}\ll\overline{R}, \tag{30}$$

where p_A is the momentum of the atoms and the middle step is warranted for non-relativistic speeds. This condition may be violated for extremely low temperatures and for high densities, such as those of artificial "white dwarfs," and in these cases the gas must be treated as a quantum gas of the appropriate statistics as is the case, for instance, in white dwarf stars [22, 33]. Although such extreme conditions are only marginally important to the present work, a well-established theoretical framework exists in the literature to describe them.

[0121] Secondly, the basic result at Eq. (12) is rigorously valid only within fourth-order perturbation theory [21]. An order of magnitude of the regime within which our results are certainly warranted can thus be found by requiring that the energy shift per atom due to the total intermolecular force of the gas cloud be much smaller than the energy of the n-th unperturbed state occupied by the atoms, that is:

$$|E_n|>>N_A|\Delta E|, \tag{31}$$

or, for hydrogen atoms:

$$\begin{aligned}\frac{e^2}{2a_n}\gg\frac{176\pi^3}{15c}N_A\frac{I}{\lambda_L^2}\frac{\alpha_A^2(\lambda_L)}{D}&=\frac{176\pi^3}{15c}N_A\frac{W}{18N_A^{2/3}\overline{R}^2}\frac{1}{\lambda_L^2}\frac{\alpha_A^2(\lambda_L)}{N_A^{1/3}\overline{R}}\\&=\frac{176\pi^3}{270c}\frac{W}{\overline{R}^3}\frac{1}{\lambda_L^2}\alpha_A^2(\lambda_L),\end{aligned} \tag{32}$$

where the relationship between total power and intensity at Eq. (48) was used. Let us consider only the s-state polarizability as $\alpha_n(\lambda_L)=\alpha_{nr}(2a_n)^3$ and let us write the average intermolecular distance in terms of the atomic radius as $R=\bar{r}a_n$. Finally, by recalling that $a_n=a_0n^2$, we find:

$$\frac{e^2}{2a_0n^2}\gg\frac{1.2\times10^3}{c}\frac{W}{\bar{r}^3a_0^3n^6}\frac{1}{\lambda_L^2}\alpha_{nr}^2(\lambda_L)a_0^6n^{12}, \tag{33}$$

or

$$\frac{W}{\bar{r}^3}\frac{\alpha_{nr}^2(\lambda_L)}{\lambda_L^2}n^8\ll\frac{e^2}{2a_0^4}\frac{c}{1.2\times10^3}\sim4\times10^{21}. \tag{34}$$

The dynamical consequences of this constraint can be seen by recalling our basic Eq. (14) written by making use of the first two terms of Eq. (32):

$$\frac{a_{lift,z}}{g}=\frac{4\pi}{c}N_A\left(\frac{I}{\lambda_L^2}\right)\frac{\alpha_A^2(\lambda_L)}{D}\frac{1}{m_Ac^2}\ll\frac{60}{176\pi^2}\frac{e^2}{2a_n}\frac{1}{m_Ac^2}. \tag{35}$$

In order to extract a physical meaning from this requirement, let us consider the problem of bringing the gas cloud to a hover from a different standpoint. The condition of balance between weight and lift of one atom is given by Eq. (14) above:

$$\frac{a_{lift,z}}{g}\sim\frac{F_{lift,z}^{gas}}{N_Am_Ag}=N_A\left(\frac{Ik_L^2}{c}\right)\frac{[\alpha^A(k_L)]^2}{D}\frac{1}{m_Ac^2}. \tag{14}$$

By using the expression for the potential of one pair at Eq. (12), we can express this condition by introducing the total intermolecular energy of one atom due to all the others, $N_A\Delta E$:

$$\frac{a_{lift,z}}{g}\sim\frac{\Delta E}{m_Ac^2}. \tag{36}$$

With the usual strong caveats of using our non-quantum "intuition" to interpret results in the realm of quantum-electro-dynamics (QED) in curved space-time, a possible qualitative view of this hovering condition is that this requirement corresponds to having the total (negative) intermolecular potential of the gas cloud cancel its total gravitational mass, thus resulting in atoms which are, for all practical purposes, "weightless." By going further, it is possible to consider intermolecular potentials which are negative and larger in magnitude than the gravitational mass of the atoms, thus causing the total energy of the cloud to become effectively negative. According to Newton's Law of Gravitation, this would require the atoms to be "repelled by gravity," instead of being attracted. Although this is an interesting and useful image—which has in fact been pursued by Boyer in the past [16]—one has to be extremely careful to adopt it as an "explanation."

[0122] In the scheme of the present invention, the upward lifting force does not result from such view, but from the well-understood distortion of the Coulomb potential due to the presence of the gravitational field. Although Boyer's work proved these two points of view to be equivalent [16], his study dealt with completely classical dipoles within the realm of special relativity and very weak gravity, and not with fully quantized atoms with the framework of QED in curved space-time. Since one possible interpretation of dispersion forces between molecules involves a modification of the zero-point-energy of the quantum vacuum [37], such simple "semi-classical" explanations should be looked at only as useful mental pictures, since the "system" under consideration is actually an open system.

[0123] In this context, two possible objections must be addressed. The first is whether it is at all correct to use perturbation theory in this case. After all, in order to bring a cloud of hydrogen atoms to a hover, the intermolecular potential must contribute an energy approximately equal to the rest-mass of a hydrogen atoms, or ≈1 GeV, whereas the ionization potential of a hydrogen atom is 13.59 eV if the atom is in its ground state and much less if it is in a Rydberg state. The intermolecular potential will be vastly larger if we want to achieve an upward acceleration. The answer rests with the key fact that, in our case, the processes we are studying take place on relatively "short" time-scales. At some time after the cloud of atoms has been cooled and trapped, a very intense beam of radiation is turned on within a total time that will be assumed to be much shorter than that of any possible atomic transition. For instance, the radiative lifetimes of very high n Rydberg atoms can be even fractions of a second, whereas lasers can be turned on in fractions of a nanosecond (10^{-9} s) and that radiation will require a similar time to cross the entire trap at the speed of light.

[0124] A well-known feature of the perturbation theory of "sudden" interactions is that the perturbation does not have to be small for theory to be used, unlike most other applications of perturbation theory, as stated, for instance in [38] (see also [39-41]): "The transition probabilities in instantaneous perturbations can also be found in cases where the perturbation is not small . . . If the change in the Hamiltonian occurs instantaneously (i.e. in a time short compared with the periods $1\omega_{if}$ of transitions from the given state i to other states), then the wavefunction of the system is "unable" to vary and remains the same as before the perturbation. It will no longer, however, be an eigenfuction of the new Hamiltonian $\hat{H}$ of the system . . . "

[0125] By making further use of this same mental picture, we can say that in the description of the fundamental physical process at the basis of the present invention we assume that the phase of upward acceleration of the atoms takes place in the very early stages immediately following the application of the laser radiation field and before the eventual modification of the wavefunctions intervenes. In other words, there exists a relatively brief time span immediately following the laser turn-on time, during which the atomic dipoles still behave as such, although, eventually, the gas evolves towards a very different state, such as a plasma for instance. Part of the refining work of the present invention will be to consider in detail the dynamics of the evolution of the atoms in the intense radiation field. However, the references quoted show that there is a solid logical foundation in the use of perturbation theory in the present case due to the instantaneous application of the radiation field.

[0126] The second objection to consider is whether it is against any fundamental law of physics to even consider the existence of a volume of space where the total energy is negative, as appears to be necessary in order to obtain a gravitational "repulsion" of the atoms. We have already provided a first hint above that one answer to this objection is that it is not at all necessary to look at this process as being due to a "negative" energy. In fact, one alternative and much more satisfactory way to explain the upward motion of the atoms is to appeal to the distortion of the dipole-dipole field due to the acceleration in the presence of a gravitational field. The distortion of the field lines due to the presence of gravitational appears to be a much firmer concept than that of a "negative" energy.

[0127] However, even if one wants to engage in the widespread debate concerning the existence of negative energy, it is interesting to point out that dispersion forces in general and Casimir forces in particular are in fact commonly used as examples to the affirmative. In other words, whereas there exist strict quantum limits, similar to the uncertainty principle, as to the length of time a very negative energy density can be imposed upon a volume of space, dispersion forces clearly afford an example where the energy density can remain negative indefinitely (although the same limitations apply if one attempts to decrease the energy even further). The important point to the present invention is that there is no fundamental physical reason that dispersion forces cannot be made negative for the brief time needed to accelerate the atoms upward as required to provide thrust. On the other hand, much more exotic and fantastic applications of negative energy to "warp-drives,""time-machines," and "faster-than-light" travel, which now appear forbidden by fundamental laws of quantum mechanics in curved space-time [42-45] are not involved in the physical processes of this invention.

4.1.3 Thrust

[0128] By using Eq. (13) above, we can write the total lifting force on the trapped atoms in the case that $\alpha_A(k_L)=\alpha_B(k_L)$, by assuming that the average of the interatomic distance between any given atom and all others is ~D, as:

$$F_{gas} = N_A^2 F_z^{AB} = \frac{4\pi^2}{c^3} N_A^2 \left(\frac{I}{\lambda_L^2}\right) \frac{\alpha_A^2(k_L)}{D} g. \tag{37}$$

Let us now consider the total momentum acquired by the gas as it flows out of the atomic trap. Without getting into the details of the dynamics of such process, which certainly deserve further consideration, let us consider the acceleration of the atoms under the action of the lifting force, F_{gas}. In an inertial reference frame, the total acceleration undergone by the atoms would be $F_{gas}/N_A m_A$, but, in this non-inertial frame, the acceleration will be $(F_{gas}-N_A m_A g)/N_A m_A=(F_{gas}/N_A m_A)-g$. For instance, if the lifting force is exactly equal to $N_A m_A g$, the atoms will not accelerate with respect to the vehicle, but only hover, as we have seen in the above theoretical treatment. Therefore:

$$a_A=\frac{F_{gas}}{N_A m_A}-g=\left[\frac{4\pi^2}{c^3}N_A\left(\frac{I}{\lambda_L^2}\right)\frac{\alpha_A^2(k_L)}{(\bar{R}N_A^{1/3})m_A}-1\right]g. \tag{38}$$

At this rate, the average final velocity of the atoms will be $v_{A,fin}=\sqrt{2a_A(D/2)}$:

$$\begin{aligned} v_{A,fin} &= \sqrt{\left(\frac{F_{gas}}{N_A m_A}-g\right)\bar{R}N_A^{1/3}} \\ &= \sqrt{\left[\frac{4\pi^2}{c^3}N_A\left(\frac{I}{\lambda_L^2}\right)\frac{\alpha_A^2(k_L)}{(\bar{R}N_A^{1/3})m_A}-1\right]g\bar{R}N_A^{1/3}}, \end{aligned} \tag{39}$$

where, of course, the application considered in this invention considers only circumstances in which the radical is real.

[0129] It is also of interest to find an order of magnitude for the time required to leave the atomic trap defined as

$$\frac{1}{2}D=\frac{1}{2}a_A(\Delta t_A)^2,$$

which sets a lower minimum to the "reload" time of the thrust cycle:

$$\begin{aligned} \Delta t_A &= \sqrt{\frac{D}{a_A}} \\ &= \sqrt{\bar{R}N_A^{1/3}\frac{1}{\dfrac{F_{gas}}{N_A m_A}-g}} \\ &= \sqrt{\frac{\bar{R}N_A^{1/3}}{\left[\dfrac{4\pi^2}{c^3}N_A\left(\dfrac{I}{\lambda_L^2}\right)\dfrac{\alpha_A^2(k_L)}{\bar{R}N_A^{1/3}m_A}-1\right]g}}, \end{aligned} \tag{40}$$

From Eq. (39) we can immediately write the total momentum of the gas, ΔP_{gas}, for each cycle at it approaches the shock absorber:

$$\Delta P_{gas}=m_{gas}v_{A,fin}=N_A m_A\sqrt{\left(\frac{F_{gas}}{N_A m_A}-g\right)\bar{R}N_A^{1/3}}, \tag{41}$$

or

$$\Delta P_{gas}=N_A m_A\sqrt{\left[\frac{4\pi^2}{c^3}N_A\left(\frac{I}{\lambda_L^2}\right)\frac{\alpha_A^2(k_L)}{\bar{R}N_A^{1/3}m_A}-1\right]g\bar{R}N_A^{1/3}}, \tag{42}$$

Finally, let us estimate the total momentum transferred to the vehicle in the assumption that the impact of the gas against it is dissipative, that is, by assuming the complete conservation of momentum. In this case, the final speed change of the (gas+craft) system at the end of the n-th cycle will be given by $m_{gas}v_{A,fin}=(m_{gas}+M_{craft})\Delta v_n$. That is, by solving for Δv_n,

$$\Delta v_n=\frac{N_A m_A}{N_A m_A+M_{craft}}\sqrt{\left[\frac{4\pi^2}{c^3}N_A\left(\frac{I}{\lambda_L^2}\right)\frac{\alpha_A^2(k_L)}{(\bar{R}N_A^{1/3})m_A}-1\right]g\bar{R}N_A^{1/3}}. \tag{43}$$

Although the acceleration of the vehicle varies over time during the cycle, it is useful to introduce an average acceleration over the cycle itself, we the understanding that the acceleration experienced by the atoms during their phase of acceleration may be different, and usually higher, than this value because of the cycle synchronization we discussed earlier. We write:

$$a_{craft}=\frac{\Delta v_n}{\Delta t_A+\Delta t_{Reload}}, \tag{44}$$

where Δt_{Reload} is the time required to prepare the new atomic cloud in the trap for the following cycle. This definition allows us to also define a nominal thrust for the engine as:

$$F_{engine}=(N_A m_A+M_{craft})a_{craft}. \tag{45}$$

[0130] In the limit in which the reload time vanishes ($\Delta t_{Reload}\to 0$), of course we find, as expected:

$$a_{craft}\to\frac{\Delta v_n}{\Delta t_A}=\frac{N_A m_A}{N_A m_A+M_{craft}}a_A, \tag{46}$$

and

$$F_{engine}\to F_{gas}. \tag{47}$$

In this ideal limit, the hovering condition becomes that $F_{gas}=(N_A m_A+M_{craft})g$, that is, the total force on the atomic cloud must be equal to the weight of the entire vehicle. This is also the thrust that could be obtained if the atomic cloud could be trapped permanently within the chamber even as it acts on the vehicle via the action-reaction law.

4.1.4 Energy Considerations and Thruster Efficiency

[0131] The total radiation power utilized will be estimated as $W=18ID^2$, which corresponds to six laser triads (this is an upper estimate of the power needed as it is possible to induce gravitation-like behavior in a particular direction by using less power than this maximum). This yields:

$$W=18IN_A^{2/3}R^2. \tag{48}$$

By solving this expression for the radiation flux I, we find:

$$I=\frac{W}{18N_A^{2/3}\bar{R}^2}. \tag{49}$$

This result allows us to obtain a realistic estimate of the relationship between the dynamics of the process and the total amount of energy needed. For instance, by substituting it into Eq. (38), we obtain the following useful expression:

$$a_A=\left[\frac{2\pi^2}{9c^3}\left(\frac{W\alpha_A^2(k_L)}{\bar{R}^3\lambda_L^2m_A}\right)-1\right]g, \tag{50}$$

which exposes the interesting fact that, in the present approximation, the atomic acceleration as a function of the total laser power irradiated is independent of the number of atoms in the trap. By means of this equation, it is possible to write the total power needed to be focused onto the trap to achieve a hover, that is, a vanishing acceleration of the atoms with respect to the vehicle. This corresponds to:

$$W_{hover}=\frac{9c^3}{2\pi^2}\frac{\bar{R}^3\lambda_L^2m_A}{\alpha_A^2(k_L)}, \tag{51}$$

independently of g. For practical reasons, let us rewrite this result in units of Megawatts (MW) in terms of the wavelength in micrometers, $\lambda(\mu m)$, of the average interatomic distance in units of Bohr radii, R/a_0, and by involving the dimensionless polarizability factor as $\alpha_A(k_L)=\alpha_{nr}(k_L)\alpha_0$:

$$W_{hover}=2.18\times10^9\frac{(\bar{R}/a_0)^3\lambda_L^2\,(\mu\text{m})}{\alpha_{nr}^2(k_L)}\text{MW}, \tag{52}$$

Of course this result should be interpreted as allowing the atoms to be brought to a hover for a time no longer than the free-fall time given at Eq. (19) above unless an independent trapping approach is employed to hold the atoms at constant intermolecular distances, such as in an optical crystal.

[0132] Similarly for Eqs. (17), (18), (20), and (21):

$$v_{A,fin}=\sqrt{\left[\frac{2\pi^2}{9c^3}\left(\frac{W\alpha_A^2(k_L)}{\bar{R}^3\lambda_L^2m_A}\right)-1\right]g\bar{R}N_A^{1/3}}\,. \tag{39a}$$

$$\Delta t_A=\sqrt{\frac{\bar{R}N_A^{1/3}}{\frac{2\pi^2}{9c^3}\left(\frac{W\alpha_A^2(k_L)}{\bar{R}^3\lambda_L^2m_A}\right)-1}\frac{1}{g}}\,, \tag{40a}$$

-continued

$$\Delta P_{gas}=N_Am_A\sqrt{\left[\frac{2\pi^2}{9c^3}\left(\frac{W\alpha_A^2(k_L)}{\bar{R}^3\lambda_L^2m_A}\right)-1\right]g\bar{R}N_A^{1/3}}\,, \tag{42a}$$

$$\Delta v_n=\frac{N_Am_A}{N_Am_A+M_{craft}}\sqrt{\left[\frac{2\pi^2}{9c^3}\left(\frac{W\alpha_A^2(k_L)}{\bar{R}^3\lambda_L^2m_A}\right)-1\right]g\bar{R}N_A^{1/3}}\,, \tag{43a}$$

with the same restrictions on the duration of the impulse.

[0133] The ratio of thrust to radiation pressure: This dimensionless quantity compares the thrust obtained from the present thrust mechanism to the radiation pressure one would obtain by simply radiating away the power used instead to alter the intermolecular potential. Clearly one expects this quantity to be much larger than unity in order for the approach of this invention to be of practical use. Should that not be so, the case could be made that focusing the radiation onto an appropriately large sail would be a more efficient use of that energy. We have, in the ideal case in which the reload time vanishes,

$$\frac{F_{gas}}{P_{rad}}=\frac{1}{W/c}N_Am_A\left[\frac{2\pi^2}{9c^3}\left(\frac{W\alpha_A^2(k_L)}{\bar{R}^3\lambda_L^2m_A}\right)-1\right]g. \tag{53}$$

In the limit in which $a_A>>g$, we have, simply:

$$\frac{F_{gas}}{P_{rad}}\to N_A\frac{2\pi^2}{9c^2}\left(\frac{\alpha_A^2(k_L)}{\bar{R}^3\lambda_L^2}\right)g. \tag{54}$$

Typical estimates of this quantity in interesting cases indeed indicate values much larger than unity for this quantity.

[0134] As in any other system that does not make use of the traditional expulsion of high speed gases to generate momentum (e.g. light sails), one must introduce new figures of merit. For instance, the typical concept of the specific impulse (the ratio of engine thrust to the weight of the material ejected in the unit time) [29], requires special attention in this case. Let us consider the case of a thrust system in which all radiation emitted by the high power lasers is permanently lost and radiated away from the spacecraft. Because of the mass-energy equivalency, this will correspond to a net mass-loss of the vehicle at a rate $M_{craft}=-W/c^2$. In the literature, the "photon rocket," which annihilates matter and antimatter to eject the corresponding radiation in a particular direction, is defined by theoreticians as the "perfect" rocket engine, because it yields the highest terminal speed at burnout for the same final to initial mass ratio [46]. However, whereas in that case the thrust is directly due to the reaction to the emitted photons, in our case the presence of the radiation in the chamber where the atoms are trapped is only a catalyst to create the thrust upon the vehicle. Thus, since the origin of the thrust in our case is not the lost radiation one could just as convincingly argue that the specific impulse becomes undefined in this case.

[0135] From the quantitative standpoint, therein lies the great interest of the approach of the present invention. That is, for the same amount of energy expended, the thrust derived is much higher than that obtainable in the photon rocket [46]. At the same time, this high thrust does not come at the price of an impulsive propulsive system as in chemical engines, but one that can actually provide high accelerations over almost indefinite periods of time [47].

[0136] Thruster efficiency: This is the ratio of the kinetic energy of the atomic cloud as it emerges from the trap to the total energy used during the acceleration of the gas. By using the above expression for the final velocity of the atoms, this quantity can be written as:

$$\frac{\frac{1}{2}N_A m_A v_{A,fin}^2}{\Delta t_A}\frac{1}{W} = \frac{N_A m_A}{\sqrt{2}}a_A^{3/2}D^{1/2}\frac{1}{W} = \frac{N_A m_A}{\sqrt{2}}\left[\frac{2\pi^2}{9c^3}\left(\frac{W\alpha_A^2(k_L)}{\overline{R}^3\lambda_L^2 m_A}\right)-1\right]^{3/2} \frac{g^{3/2}\overline{R}^{1/2}N_A^{1/6}}{W}. \tag{55}$$

In the same limit $a_A >> g$ we find:

$$\frac{\frac{1}{2}N_A m_A v_{A,fin}^2}{\Delta t_A}\frac{1}{W} \to \left(\frac{W}{2m_A}\right)^{1/2}\frac{1}{\overline{R}^4}\left[\frac{2\pi^2}{9c^3}\left(\frac{\alpha_A^2(k_L)}{\lambda_L^2}\right)\right]^{3/2} g^{3/2}N_A^{7/6}. \tag{56}$$

Special care must be exercised in interpreting the results obtained by extrapolating this equation to values of the efficiency that exceed unity, since this may be an indication that other phenomena are becoming important. Furthermore, we have here neglected other forms of energy that are also required, such as, for instance, the radiation required to excite the atoms if Rydberg states are used and the energy used for the initial trapping. As shall become clear in the examples below, the present invention represents a useful and revolutionary technological innovation even in regimes where the thruster efficiency is below unity.

4.2 Start-Up, Maneuvers, Cruise, and Turn-Off

[0137] The fundamental, basic physics principle of this invention is the distortion of the field of a dipole in an accelerated reference frame and the resulting "lifting" dipole-dipole force. In order for this principle to operate, an acceleration must be induced upon the dipole system before the thrust on it can appear. In other words, it is not possible to turn on the thrusting system of this invention from a cruise phase (approximately at constant speed). Because of the Principle of Equivalence, such initial acceleration can be provided by means of any combination of two different, but equivalent mechanisms. In the former, the acceleration is due to gravitation, which provided the initial motivation for the present invention.

[0138] Because of the Principle of Equivalence, the effect of a gravitational field is indistinguishable from that a uniformly accelerated reference frame (if we neglect the Linet term mentioned in our comment of Eq. (3) and in Ref. [9] and quantitatively unimportant to this invention). Therefore, the trigger acceleration can be provided by the gravitational field as a supported (not freely falling system) rests, for instance on the ground. Because of the Principle of Equivalence, if the spacecraft is freely falling its acceleration exactly cancels that of the gravitational field and it becomes indistinguishable from a laboratory at a large distance from any other mass, to the extent that its size is relatively small.

[0139] Because of the mechanism of this invention, if the spacecraft is resting on the ground, upon laser turn-on, the intermolecular forces will be distorted in such a way as to cause an upward acceleration of the atoms, which can be made, for instance, as large as needed for the spacecraft to hover. On the other hand, if the spacecraft is in outer space at a large distance from any other celestial objects, such as in interplanetary flight, the initial acceleration must be provided by independent means. This can be achieved in a variety of ways. For instance: (1) an initial forward acceleration can be produced by a traditional thrusting system until the first few cycles of the present engine are produced—in this case the traditional thruster becomes an "ignition system" for the present invention; (2) the entire craft can be put into rotation around an appropriate axis, thus reproducing Einstein's rotating disk [46]; (3) on re-entry, a spacecraft undergoes an aerodynamic deceleration that can be used as the trigger.

[0140] The same above sample triggers can be also used for maneuvering, that is, to create an initial acceleration in a direction different than that of the present motion. This can be very useful to transform the present invention in a vehicle for motion near the surface of the earth. A slight acceleration parallel to the ground could be maintained with relatively little radiative energy while a larger amount of radiation would keep the vehicle hovering safely. A slow decrease in the power of the lasers, or any other change in the parameters determining the thrust, such as intermolecular distance or total atom number, would result in a decreased acceleration of the atoms and, thus, of the vehicle. In order to bring the vehicle to a cruise (constant speed) the only necessary action is to turn off the high power lasers, or, alternatively, to empty the atomic trap system of atoms. The acceleration upon the vehicle would then immediately stop along with the dipole-dipole field distortion, thus bringing the craft to constant velocity. Finally, a trigger acceleration can be exercised in a direction opposite to the instantaneous velocity to start an opposite thrust that brings the vehicle to a deceleration and to a new hover.

[0141] Although the notion of flying by exploiting an initial acceleration may appear counterintuitive, it must be said that a similar transition also challenges some pilots at the beginning of their training. In fact, flying through air requires speed to create lift from the airfoils and a slower and slower aircraft is unable to keep altitude. This concept requires constant training in beginning pilots today to fight the instinct to "pitch-up" to regain lost altitude on landing, as that can result in a stall. Similarly, the lifting mechanism of this invention requires the user to become sensitive to acceleration—as opposed to speed—and to the notion that an accelerating vehicle can maintain thrust, whereas one at constant speed will loose thrust.

4.3 Numerical Examples

[0142] The following numerical examples were generated by making use of a Mathematica notebook [13] which encoded the equations discussed in this document. The design of any of the missions below could (and does) take many years of study, but the goal of this section is to provide the information necessary to appreciate the fact that the physics of this invention leads to realistic engineering demands. In other words, whereas relativistic travel is usually assumed to require harnessing absolutely fantastic amounts of energy, the method below yields numbers that can be realistically contemplated to be put into feasible design with presently existing technologies.

[0143] Any space mission must start with some requirements, which are to some extent arbitrary or descend from other constraints. In the examples below, we assumed that all travel takes place at a given spacecraft acceleration $a/g<1$. This is to contrast the approach to space travel of this invention with typical designs, which either impose high (several g) accelerations for a short periods of time (chemical rockets) or provide very low thrust over very long periods of time, such in the case of ion propulsion, in which the author of this invention has been directly involved [47]. Let us recall that the present method requires that the intermolecular potential be distorted by the acceleration of the vehicle (or by gravitation if the vehicle is held at rest, as required by the Principle of Equivalence). Therefore, the acceleration of the vehicle due to the impact of a gas cloud into the plate determines the efficiency of the atomic acceleration in the following cycle.

[0144] According to the present method, space travel is most efficient when it is significantly accelerated. This is not a drawback, since, for instance, it is now understood that the dangers to human health of very long periods of weightlessness are significant. The data presented herein were thus produced by assuming a spacecraft acceleration $a=g$ and then by demanding self-consistency, that is, by requiring that the spacecraft acceleration produced by any given gas cloud also be equal to g. Even with this requirement, the choices shows below represent only one of many possibilities chosen for their being realistic from the engineering standpoint or for their being representative of a contrast between the present invention and present-day technology. For simplicity, the reload time was everywhere assumed to be negligible.

4.2.1 A Robotic Low-Thrust Delivery System to the planet Mars

Propulsive System Specifications

[0145] $W=35.0$ MW$=3.5\times10^{14}$ erg/s

[0146] $\alpha_{nr}=1\times10^{6}$ (quasi-resonant response)

[0147] $n=1$ (ground state)

[0148] $\bar{R}=5a_0$

[0149] $\lambda_L=1000$ Å (quasi-Lyman-α transition)

[0150] $M_{craft}=10^{7}$ g=10 metric tons

[0151] $N_A=4.66\times10^{26}$

[0152] $m_{gas}=7.86\times10^{3}$ g=0.786 kg

[0153] Chamber Size $D=20.3$ cm

[0154] $W_{hover}(a_{craft}=1\text{ g})=2.7$ kW

[0155] $a_A=1.28\times10^{5}$ cm/s^2

[0156] $v^{A,fin}=2.29\times10^{3}$ cm/s

[0157] $\Delta t_A=1.79\times10^{-2}$ s

[0158] $t_{ff}=60.6\times10$ s

[0159] Efficiency$=3.3\times10^{-4}$

[0160] Δv_{craft}/cycle=2.62 cm/s

[0161] $a_{craft}=10$ cm/s^2 (self-consistent)

[0162] Thrust=10.0 kN

[0163] (Thrust/Total Radiation Pressure)$=8.57\times10^{4}$

[0164] Earth-Mars distance at opposition (Aug. 27, 2003) 55.758×10^{11} cm [48]

[0165] Total Travel time$=1.52\times10^{6}$ s=17 d 13 h 17 min 46.7 s

[0166] Maximum Speed (reached after$=6\times10^{5}$ s)$=6.0\times10^{6}$ cm/s=60 km/s

[0167] Maximum Kinetic Energy (reached after 6×10^{5} s)$=1.8\times10^{20}$ erg

[0168] Total Energy Radiated (up to maximum speed)$=2.1\times10^{20}$ erg

[0169] Time to Mach 1 (Speed of Sound $v_{sound}\cong331$ m/s)$=3.31\times10^{3}$ s

[0170] Time to Clear Low-Earth-Orbit (400 km)$=2.8\times10^{3}$ s=47.1 min

[0171] Time to Clear the Earth-Moon System (4×10^{5} km)$=8.94\times10^{4}$ s=1 d 50.7 min

Comments

[0172] In order to obtain the above estimates for the transfer to Mars, we greatly simplified the problem by neglecting the gravitational force of all objects involved, including the Sun, the Earth, and Mars. Of course, the gravitational force of the Earth is actually an integral part of the engine start-up mechanism of the present invention and in that sense its effect is accounted for here. However, since the gravitational field of all near-by objects is cleared in a matter of minutes, its influence on the dynamics of flight was neglected in these order-of-magnitude calculations.

[0173] It is useful to make come comparisons between the specifications of the system above and present-day delivery systems, since that helps elucidate the great technological relevance of the present invention. A typical choice for such recent space missions to Mars as the ones of Spirit and Opportunity is the Boeing Delta II 7925 or 7925H (the letter H indicates the more powerful high performance version) [49]. In its common configuration, the RS-27A engine of the Delta II first stage, along with the additional nine strap-on solid rocket motors, generates approximately 8.9×10^{5} Newtons of thrust, which are necessary to lift the total "wet" (fueled up) vehicle mass of 285, 228 kg off the launch pad. This thrust is almost one order of magnitude larger than that of the engine described in this document until the Main Engine Cut Off (MECO), approximately 265 s after lift-off. The thrust of the following stages is smaller, with the thrust of the third stage at approximately 6.6×10^{4} N.

[0174] Since the initial phase of ascent takes place under the thrust of a chemical rocket, not surprisingly we see that both in the case of Spirit and Opportunity the Delta II vehicle passes through Mach 1 in a much shorter time than the one propelled by the engine of this invention (32.4 s and 29.6 s, respectively) [50]. However, very importantly, by a very good approximation, the entire launch mass of 10 metric tons (10^4 kg) is propelled towards Mars in the present case whereas, in the traditional approach, only 1.070×10^3 kg out of the initial 285, 228 kg represent the useful remaining payload. Of this surviving mass, only 533 kg actually lands on Mars in the traditional case, since approximately 250 kg are allocated to the cruise stage alone.

[0175] Finally, the entire vehicle propelled by the engine of this invention arrives at Mars in a matter of less than twenty days, whereas both Spirit and Opportunity, extremely reduced in mass, arrive approximately six months later. It is very significant that the gravitational field of Mars, added to the deceleration of the spacecraft, makes the process of landing much slower and completely safe, in contrast with what NASA/JPL itself defines as the "six minutes of terror," during which the vehicle must be slowed down from 12,000 miles/hr to 0 miles/hr corresponding to average accelerations $a_{craft}\sim1.5$ g, although during the last drop the acceleration can reach ~40 g [51]! In contrast, since the present vehicle is approaching Mars decelerating at $a_{craft}=10$ cm/s^2, the same process can be executed within over a week-long time and distributed over a path at uniform acceleration that concludes with the entire vehicle safely hovering at a desired height above the ground of the planet.

[0176] An obvious question when carrying out a comparison between traditional propulsion technology and the mechanism described in this document concerns the feasibility of generating $\sim10^{-3}$-10^{-2} s laser pulses requiring over 30 MWe (that is, MW of electric power) on board of a space vehicle. Is this possible? Has this possibility ever been carefully considered in the past? The answer to such legitimate question is that the study of high power, high efficiency, low-mass nuclear reactors for use in space applications is actually extremely advanced, although political and public opinion considerations tend to hide the enormous amount of available information away from the non-technical readership. A small selection of such literature, which cannot be cited here even partially because of its sheer size, can be found at [52] and References therein. General historic motivations behind this technical effort are to be found in the large energy needs of any hypothetical "Star Wars" defense system, in research into the possibility of interstellar travel, and, finally, in the recently renewed commitment to the human colonization of space made by the President of the United States and NASA. What is important to the evaluation of the engineering feasibility of the present invention is that, typically, reactor power densities in the order of ~1-5 kW/kg are surely possible, with very wide variations in either direction of that estimate depending on the specifics of design and shielding requirements. However, the figure above is sufficient to make the case that, by means of presently existing technology, it is absolutely appropriate to consider a 10 ton-vehicle carrying a reactor able to produce pulses in the 30 MWe range in space.

[0177] The implications for human flight to Mars by means of the present invention are very significant as well. At present, the delivery of several tens of tons of payload to Mars is contemplated to take approximately 180 days, with a typical mission duration for the crew of 2-3 years. The overwhelming majority of the mission duration would be spent in complete weightless conditions during transit and possibly under exposure from harmful radiation. In addition, the prospects of recovery in case of an even minor malfunction are dire not to speak of a major accident of the type that occurred on Apollo 13. Being able to reduce the travel time to a matter of days, while transporting the crew under at least partial gravitational conditions completely changes the prospects for successful colonization of the Red Planet as well as the potential for a rescue mission should that be necessary [53].

4.2.4 A Thrust System for Safe, Low-Speed, Near-Earth Human Transportation

Propulsive System Specifications

[0178] $a_A=6.29\times10^6$ cm/s^2

[0179] $v_{A,fin}=1.61\times10^4$ cm/s

[0180] $\Delta t_A=2.55\times10^3$ s

[0181] $t_{ff}=60.6\times10$ s

[0182] Efficiency=11.2%

[0183] Δv_{craft}/cycle=18.35 cm/s

[0184] $a_{craft}=g/2$ (self-consistent)

[0185] Thrust=0.491 MN

[0186] (Thrust/Total Radiation Pressure)=4.20×10^6

[0187] Earth-to-LEO distance=4×10^7 cm

[0188] Total Travel time=4.97×10^2 s=8 min 17.5 s

[0189] Maximum Speed (reached at midpoint)=1.22×10^5 cm/s=1.22 km/s

[0190] Maximum Kinetic Energy (reached at midpoint)=7.44×10^{16} erg

[0191] Total Energy Radiated (at midpoint)=1.74×10^{17} erg

[0192] Time to Mach 1 (Speed of Sound $v_{sound}\cong331$ m/s)=67.5 s

[0193] Note: The quantities not repeated are unchanged with respect to the previous example.

Comments

[0194] The interest of this particular case lies not in the acceleration, which is less favorable than by means of already available technology, but in its ability to deliver the entire payload to high altitude at rest. This allows us to consider an entirely new philosophy or air transportation and space travel around the Earth (or other planets). Whereas the key objective to reach extreme altitudes with ordinary technologies must be the achievement of high speeds, as that is the only strategy which allows the vehicle to be injected into a permanent orbit, in the case of the present invention it is possible to deliver a payload to a high hovering altitude without requiring orbital speeds.

[0195] Similarly, the descent maneuver of our vehicle does not require the fiery but unavoidable re-entry of typical deorbiting, thus avoiding the accompanying extreme heating and grave dangers to the crew, as in the recent Columbia tragedy. In fact, the vehicle of this example would not reach speeds higher than Mach 4 before decelerating to its hovering point, as opposed to the orbital re-entry speed of the Shuttle of approximately Mach 24. This achievement would represent nothing less than a revolution in aerospace technology. Interestingly, the present approach also lends itself to being phased-in as it replaces traditional propulsion technologies. In other words, it is conceivable that a system of lower thrust, unable by itself to attain a complete hover, could be placed into service for the only purpose to provide additional breaking in an emergency at those speeds that make parachute deployment am impossible option.

I claim:

1. An apparatus comprising:

at least one trap for confining particles;

a device for delivering electromagnetic radiation to the confined particles, wherein said device delivers an amount of electromagnetic radiation that is sufficient to:

(i) induce long-range interactions between said particles; and

(ii) cause said particles to either accelerate or hover.

2. The apparatus of claim 1 wherein said trap is an atomic trap.

3. The apparatus of claim 1 wherein said device is a laser.

4. The apparatus of claim 1 wherein said particles are characterized by a polarizability, wherein said particles have a polarizability that is greater than a static polarizability.

5. The apparatus of claim 1 wherein said electromagnetic radiation has a wavelength, and wherein said wavelength is a near-resonance wavelength.

6. The apparatus of claim 1 wherein said particles are Rydberg atoms.

7. The apparatus of claim 1 wherein said particles are neutral atoms.

8. The apparatus of claim 1 wherein said apparatus includes at least 1×10^6 traps.

9. The apparatus of claim 1 further comprising a current controller, wherein said current controller causes said device for delivering electromagnetic radiation to deliver, at a minimum, an amount, W, of electromagnetic radiation, given by the expression:

$$W = 2.18\times10^{9}\frac{(\bar{R}/a_0)^3\lambda_L^2}{\alpha_{nr}^2(k_L)}$$

wherein:

W: is power, in Megawatts;

R/a_0: is the average interatomic distance, in Bohr radii;

λ_L: is the wavelength of the electromagnetic radiation, in micrometers;

α_{nr}: is a factor (dimensionless) by which the static polarizability of a particle is increased at near resonance; and

k_L: is the laser light wave number (dimensionless).

10. The apparatus of claim 1 wherein said apparatus comprises a propulsion system, and wherein said propulsion system is coupled to a vehicle.

11. The apparatus of claim 10 further comprising an arrangement whereby said vehicle receives momentum from said particles.

12. The apparatus of claim 10 further comprising a surface upon which said particles are impacted, wherein said surface is coupled to said vehicle.

13. The apparatus of claim 12 wherein said surface is a part of a piston.

14. The apparatus of claim 10 wherein said vehicle is selected from the group consisting of helicopter, prop-driven aircraft, jet-aircraft, and space vehicle.

15. The apparatus of claim 10 further comprising a conventional propulsion system, wherein said conventional propulsion system is selected from the group consisting of a turboprop engine, a turbojet engine, a turbo-fan engine, ramjet engine, and chemical (rocket) engine.

16. The apparatus of claim 1 wherein said apparatus is coupled to freight.

17. An apparatus comprising:

a hull;

a first propulsion system, wherein said first propulsion system is disposed within said hull and comprises:

(a) at least one trap for confining particles; and

(b) a device for delivering electromagnetic radiation to the confined particles, wherein said device delivers an amount of electromagnetic radiation that is sufficient to:

(i) induce long-range interactions between said particles; and

(ii) cause said particles to accelerate;

(c) a surface against which said particles are impacted, wherein said surface couples to said hull; and

a source of energy for powering said device.

18. The apparatus of claim 17 comprising a second propulsion system, wherein said second propulsion system is selected from the group consisting of a turboprop engine, a turbojet engine, a turbo-fan engine, a ramjet engine, and chemical (rocket) engine.

19. An apparatus comprising:

at least one atomic trap for confining particles;

at least one laser for delivering electromagnetic radiation to the confined particles;

a power source for powering said laser; and

a controller for controlling said laser, wherein said controller, in conjunction with said power source, provide an amount of current to said laser that is sufficient to cause said laser to deliver an amount of electromagnetic radiation that is sufficient to:

(i) induce long-range interactions between said particles; and

(ii) cause said particles to accelerate or hover.

20. The apparatus of claim 19 further comprising a conventional propulsion system.

* * * * *

advertisement

See why millions trust us for Home & Auto.

Contact Me ▶

Darren Reuter, Agent
Insurance Lic. #: 0D97352
San Rafael, CA 94901

StateFarm®

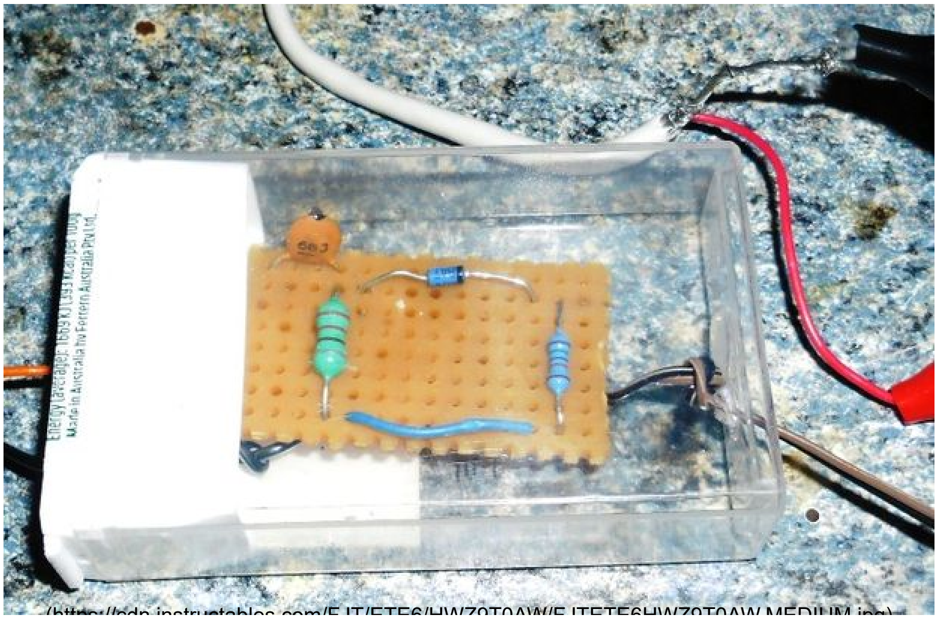

(https://cdn.instructables.com/FJT/FTF6/HWZ0T0AW/FJTFTF6HWZ0T0AW.MEDIUM.jpg)

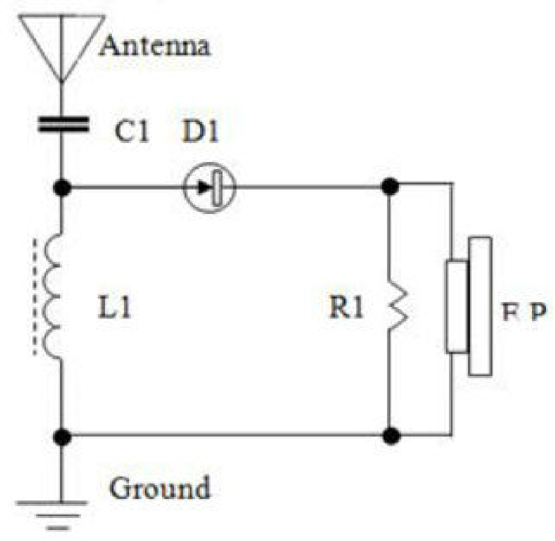

(https://cdn.instructables.com/F8Q/OFKC/HX8HK0L5/F8Q(

(https://cdn.instructables.com/F7L/A

advertisement

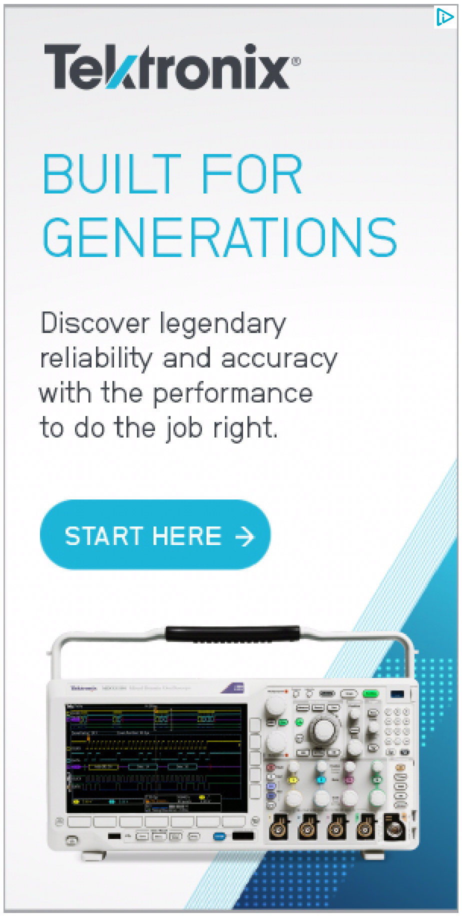

Tektronix®

BUILT FOR GENERATIONS

Discover legendary reliability and accuracy with the performance to do the job right.

START HERE →

About This Instructable

10,581 views

84 favorites

License: Public Domain

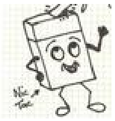

mk484 (/member/mk484/)

(http://)

(/member/mk484/)

Follow

82

Hi there! **This Instructable is all about building a basic crystal radio set that is so simple to build and understand, that a child could do it - with help from mum or dad** - or even at school, as a class project. Parts can be bought "off the shelf" at Jaycar Electronics and other suppliers, or purchased online via Ebay and Paypal.

I built my first ever crystal set when I was 9 years old, but that was a different world. We're going to build a 21st century style radio with all modern components which are relatively easy to obtain at your local electronics store or on the Internet.

The radio is contained inside a Tic Tac box, and is called "The Nic Nac Tic Tac Radio". It is built on a square of matrix board, with the component leads poked through the holes in the board, and joined underneath. It is a very simple crystal set, designed to initially receive only one station, but you can expand the tuning range with the addition of only one extra component.

The Tic Tac radio can be constructed by simply twisting the component leads together, underneath the matrix board and attaching antenna/ground wires as well as the earphone wires to the board, using that same technique. You don't have to solder any wires, but there is another way to join them all together - more on that later.

The Tic Tac Radio is a very easy set to put together - you don't need to wind bulky coils or use heavy tools and screw strange looking parts onto a large breadboard. See step 1 for the parts list you will need.

Step 1: Parts and Tools You Will Need

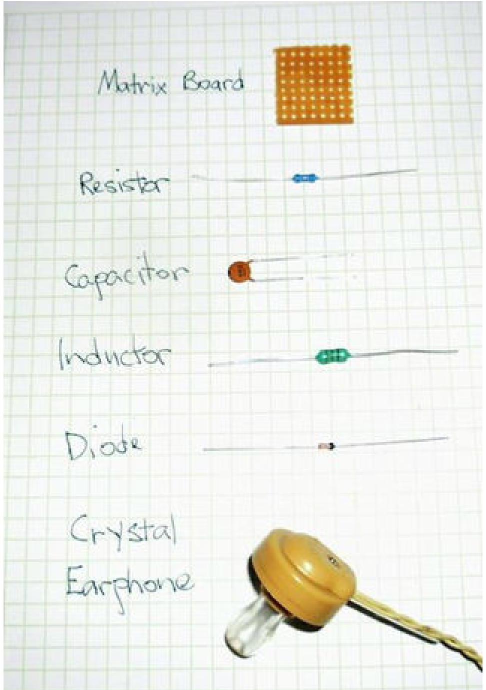

(https://cdn.instructables.com/F9B/WLFJ/HX5H54T

(https://cdn.instructables.com/F99/H

(https://cdn.instructables.com/F

(https://cdn.instructables.com/

You will need to acquire the following components from a local electronics shop or an online electronics business.

Bio: Started as a hobbyist at 9 - built my first crystal radio on one of mum's prized cutting boards (eeek) - Worked in 2 electrical/electronics ... More » (/member/mk484/)

I am adding the catalogue numbers for Jaycar Electronics, so if you live in Australia or North America, you may be able to visit a Jaycar store and buy the parts over the counter, if not then you may be able to do an online mail order via Paypal:

Parts List

Resistor - 47k - yellow-purple-black-red and brown - RR 0612 (pkt of 8)

Capacitor - 68pF - ceramic x 2 - RC 5322 (pkt of 2) and a 100 or 120 pF value as well for experiments.

Inductor - 220 uH - red-red-brown silver - LF 1538 (resistive type)

Polyvaricon tuning capacitor - 220 pF - RV 5728 - with knob and mounting screws

Diode - BAT46 - ZR 1141 (You can also use a 1N34A Germanium Diode too if you have one at home)

Ceramic Earphone - AS 3305 *

A 25 meter roll of yellow hook up wire for the Antenna wire and a 3 meter length of wire for the Ground wire.

Please note that some Jaycar parts come in multiples of 2 or more per packet. **And please note the following:**

PLEASE NOTE: THIS PROJECT/KIT CONTAINS SMALL PARTS THAT MAY FORM A CHOKING HAZARD FOR SMALL CHILDREN OR PETS. NOT SUITABLE FOR CHILDREN UNDER FIVE (5) YEARS OLD.

*A normal crystal radio earphone is OK, but if you can't get one of these, or if the one you bought goes dead (as they sometimes do,) you can use a substitute, such as the Murata PKM44EW passive transducer (see picture above) which is available from an old Telstra TF200 touchphone, (the one on the left in the diagram above,) or an equivalent, such as the ARIO transducer, from an old Telstra T1000 pushbutton phone.

The ARIO unit is soldered to the phone's pc board so you'll need to be able to unsolder the three mounting pins underneath the board, or find someone in the neighbourhood who is able.

Take the back off the TF200 (if you've obtained one of these phones,) and you'll see a black disc shaped object 2" round by 1/2' thick - with a red and black wire. Unplug the wires from the circuit board, and unscrew any retaining screws and remove the transducer. Cut the mini plug off, **carefully** strip the insulation from the ends of the wires and extend them by about 18" with 2 thin lengths of hookup wire. These pietzo devices make good earphones for crystal sets and can be housed in an old pair of ear muffs.

Miscellaneous Materials:

A Tic Tac box (smaller size)

A piece of matrix board at least 7 holes long by 8 holes across. Cut the board to fit neatly inside the Tic Tac box.

A short length of 2 differently coloured wires 60 mm in length and 2 crocodile clips with red and black plastic covers

A length of antenna wire at least 25 metres long and a 3 metre length of a different colour for the Ground wire.

A metal rod or cold water pipe for the ground stake. **Be careful which pipes you connect your Ground wire to.**

You will need a small sharp object for punching holes in the Tic Tac box.

A small pair of wire cutters for cutting and stripping wires.

Step 2: How a Crystal Radio Works

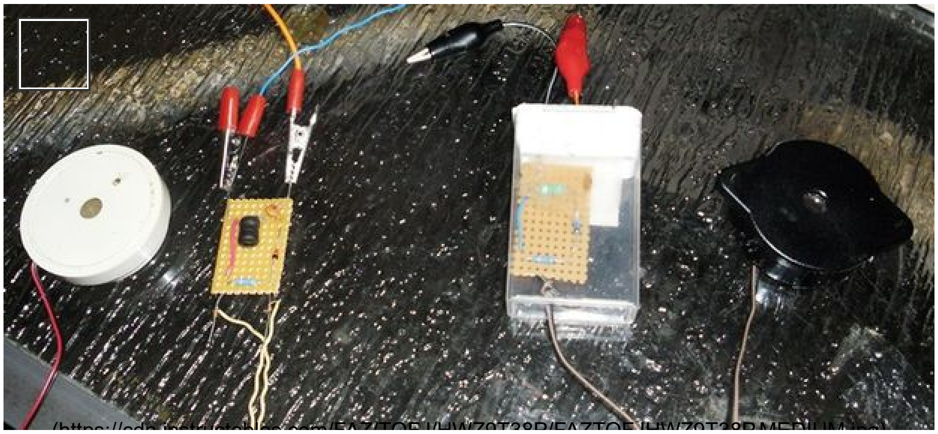

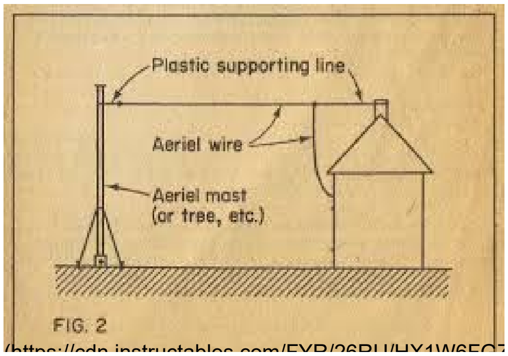

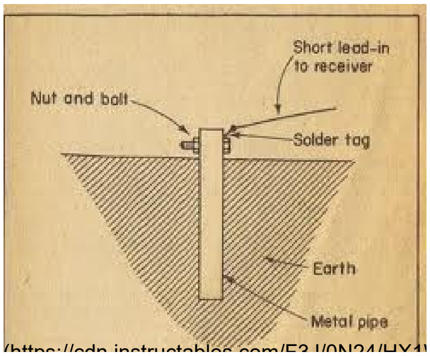

Radio signals consists of two parts - the 'carrier wave' which is the AM radio station's frequency of operation, and the 'program signal' which is mixed with the carrier wave for transmission.

Radio waves travel out from the AM transmitter tower through the atmosphere. We want to capture one specific frequency so we can listen to it, so we need the antenna/ground wire system to capture that signal. We also need a 'tuned circuit' that will filter out the desired AM signal, and discard the rest, so that all the other unwanted radio signals pass out through the ground wire to earth.

Two components in our circuit will perform that task for us. The **capacitor** C1, together with L1 **inductor**, form a basic 'series tuned' circuit. Their respective values will determine just which local AM radio station we will capture.

We also need a **diode** to 'detect' the voice and music, so we can hear them in our **earphone**, which transduces electrical signals into sound waves that we can hear.

In the photo above you can see a completed Tic Tac crystal radio. It is already inside the box. The other set is connected to the antenna/ground wire circuit, undergoing a 'soak test'. It is necessary to do this to ensure that the radio will work once inside the box!

The first diagram shows a typical antenna wire installation. Coming out of a window, the wire is anchored to the building and then over some distance (10 metres +) to a nearby tree or other building.

You must take great care not to erect antenna wires near to power or telephone cables, near your home!

The ground wire comes out the same window and is anchored to a metal pipe/water pipe or metal ground stake, embedded in soft, moist soil.

You must not connect ground wires to electrical mains wiring installations, including wall power outlets!

Another electrical hazard to consider is **lightning strikes!** Although it is very rare for anyone to be seriously injured or die from a lightning strike it is **not impossible**. So, if you hear a thunderstorm coming your way (you may hear the lightning 'crashes' in your earphone first,) then **disconnect your antenna wire immediately**, connect it to the ground wire and put it well up out of the way. Stay well clear of this temporary antenna/ground connection until the storm has completely passed away from your general area - miles away!

Step 3: How to Build the Tic Tac Crystal Radio Set

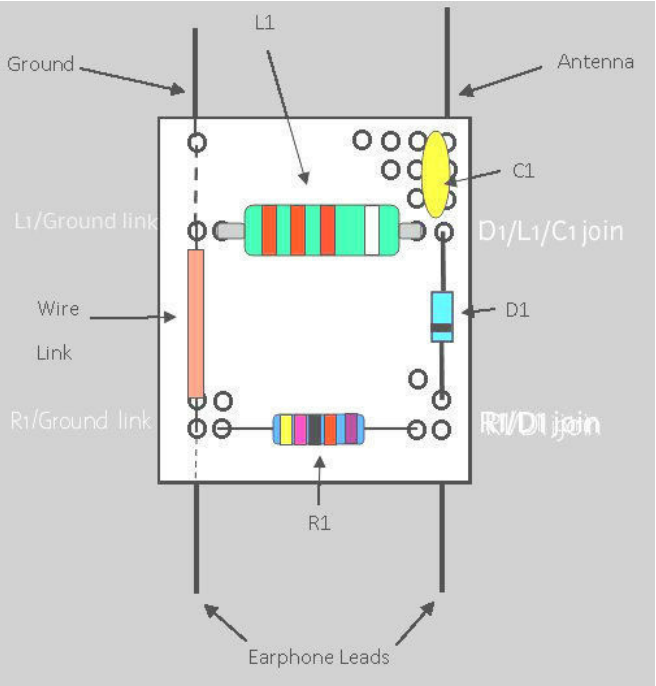

(https://cdn.instructables.com/FA5/FW4F/HXA9BXF0/FA5FW4FHXA9

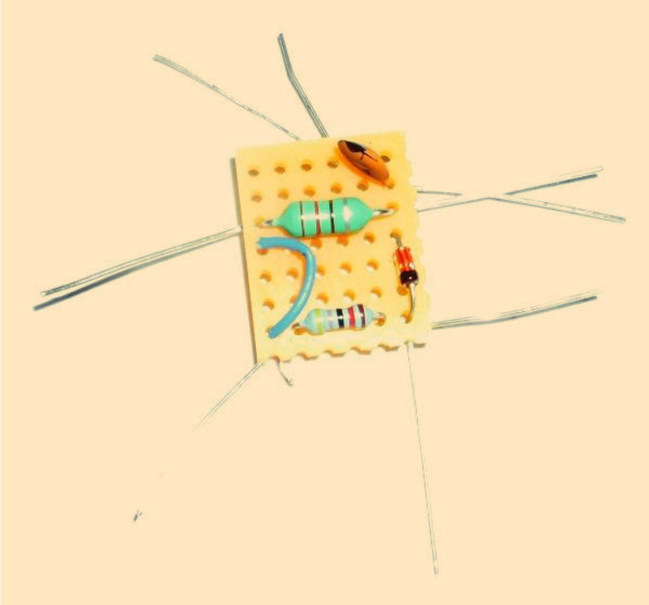

(https://cdn.instructables.com/FS8/N068/HXA9BY3O/FS8N068HXA9

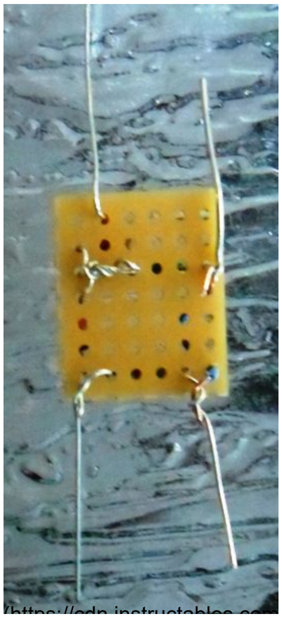

(https://cdn.instructables.com

1. Take the components and lay them out on a clean surface.

2. Take the square of matrix board (this is also called 'perf board' because of all the holes or perforations in the material) and beginning with the 220uH (uH = micro-Henry) and place it as shown in the diagram.

3. Then take the diode, capacitor and resistor and place them in the places shown for them, making sure that the coloured band at one end of the diode joins with the resistor wire, as shown

4. Then take each junction where wires come together through their respective holes and gently twist them together, until they form a neat, tight bundle. Take your side cutters and cut off any excess length, taking care not to cut any one wire too short, so that it comes undone from the join.

.5. Take the link wire, strip 2 centimetres of insulation from each end of the wire and install that wire between the free ends of the inductor and resistor, and the matrix board construction is complete.

6. Then take the ceramic earphone, cut the plug off the end, and strip the 2 wire ends about 1.5 centimetres in length. Wrap each earphone wire around each end of the resistor component, underneath the matrix board.

7. Finally, strip the insulation off both ends of the 60mm long differently coloured wires, and attach them to the matrix board - one goes to the inductor/link wire junction and this will be the Ground connection wire. Attach the other one to the free end of capacitor C1 and this will be the antenna wire connection. Both connections are made underneath the matrix board. The Nic Nac Tic Tac Crystal Radio is now ready for testing.

Step 4: Final Checks and Installation of Wires

24g NET
FRESHNESS IN 2 CALORIES
tic tac
Strawberry Fields

(https://cdn.instructables.com/FX9/8BCY/HX8HJEU3/FX98BCYHX

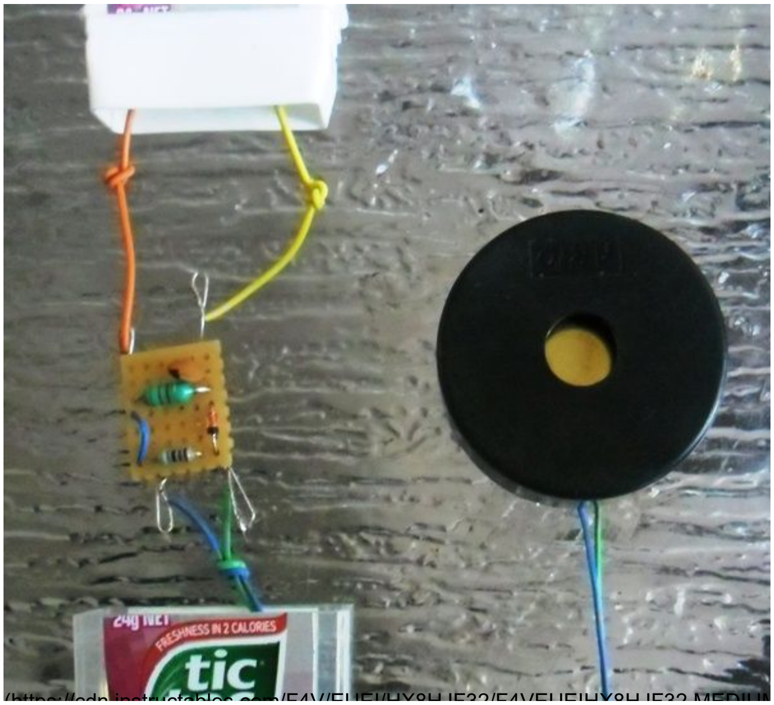

(https://cdn.instructables.com/F4V/EHEI/HY8HJF32/F4VEHEIHY8HJF32.MEDIUM

When you've finished constructing the matrix board circuit and have clipped all the excess component leads off, puncture three (3) holes in the Tic Tac box - 2 small holes on the top lid of the box, about an inch (24mm) apart, so that your antenna and ground wire leads can come through the box lid, and be connected to the matrix board at those 2 points.

Feed the wires through the holes and then tie small knots in each one, near the underside of the box lid, so that they won't pull back out if strained, and disconnect themselves from the matrix board.

Then make one larger hole in the centre of the bottom of the box (clear part) so that your earphone wires can be fed through to the connecting points on that part of the matrix board. Tie a larger knot in the earphone lead so as to prevent it from pulling out of the box if it is strained. Strip the insulation off the earphone leads, and wrap them around the matrix board leads at those 2 points. You are now ready to test out the Nic Nac Tic Tac Crystal Radio.

Step 5: Testing Out the Tic Tac Radio

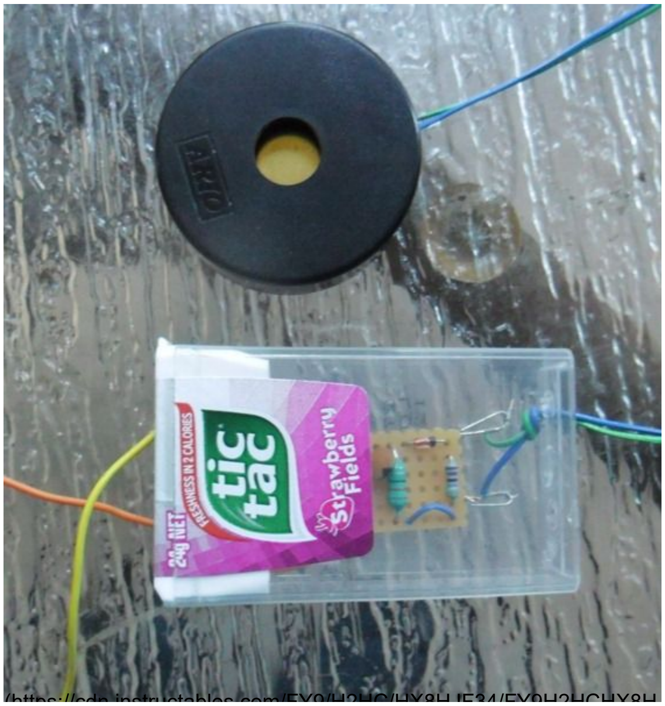
tic tac
Strawberry Fields
24g NET
FRESHNESS IN 2 CALORIES

(https://cdn.instructables.com/FX9/H2HC/HX8HJF34/FX9H2HCHX8H

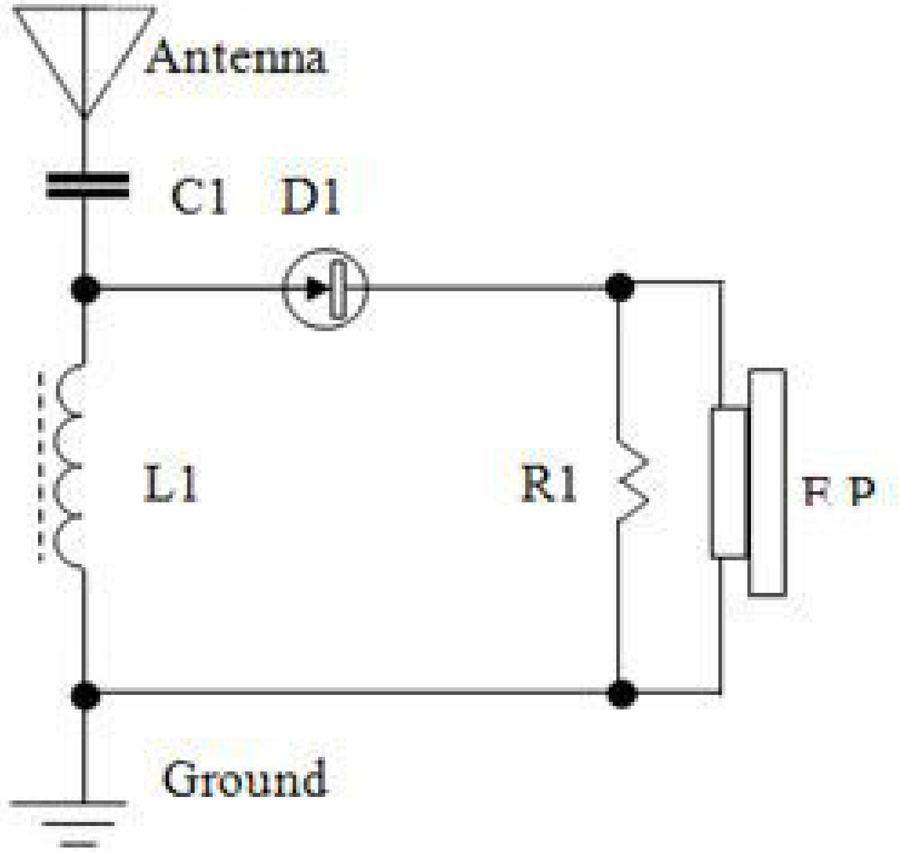
Antenna
C1
D1
L1
R1
F P
Ground

(https://cdn.instructables.com/FL6/41CU/HX5H55A2/FL641CUHX5H55A2.MEDIUM.jpg)

Take your finished Tic Tac Radio (with the lid part gently pushed back inside the top of the box) and connect your Antenna and Ground wires to the lead outs from the box lid. Place the earphone in/over your ear,and listen carefully for a local AM radio station. This crystal radio is a simple one, and you may have to make one or two adjustments to the components, before you succeed in receiving one or more local AM radio stations, in your area.

If you can't hear anything in the earphone, don't panic. It might just be a simple wiring mistake, which is easily fixed. Go back over all of the steps, making sure that you have the right value components from the electronics store. Make sure that each component is in the right place on the matrix board, (don't confuse the L1 inductor with the R1 resistor - they look a lot like each other!) and that no component wires have come undone from the twisted joins you have just made. I'll be writing up a troubleshooting step soon, so if you run into any problems, post your questions in the 'comments' section down at the bottom of the page, and I'll try and answer them as soon as possible.

Make sure that your antenna and ground wires aren't snagged on anything metal or anything **dangerous!**

If you have ANY doubts about the electrical safety of your antenna or ground wires, then consult a licensed electrical trades person, who will be able to advise you on electrical safety principles and procedures!

Always remember that electrical safety is your responsibility! If you don't think it is safe to proceed, then don't!

You will need at least 10 to 15 metres of antenna wire, strung between 2 insulating points (not connected to anything metallic, or that gets wet,) at least 2 to 3 metres in height - anything less than this minimum arrangement may mean that you cannot receive any signals at all.

Some places are regarded as 'radio dead spots', so you may need to try an open space, such as a park or a remote corner of a beach. If you do erect antenna and ground wires in public places, hang some streamers or balloons off of them, to alert people to their presence, otherwise people going past may become entangled with them - and get cranky with you!!!

As a final word for now, you'll be happy to know that the completed "Nic Nac Tic Tac Radio" shown in the picture above, picked up local AM radio station **1116 khz 4BC** here in Brisbane, with a very clear signal and quite good volume. It works! So be safe kids, have some fun and look forward to more... mk484

Step 6: Nic Nac Extras

(https://cdn.instructables.com/FGL/BHOH/HX8HJG9V/FGLBHOHHX8HJG9V.MEDIUM.jpg)

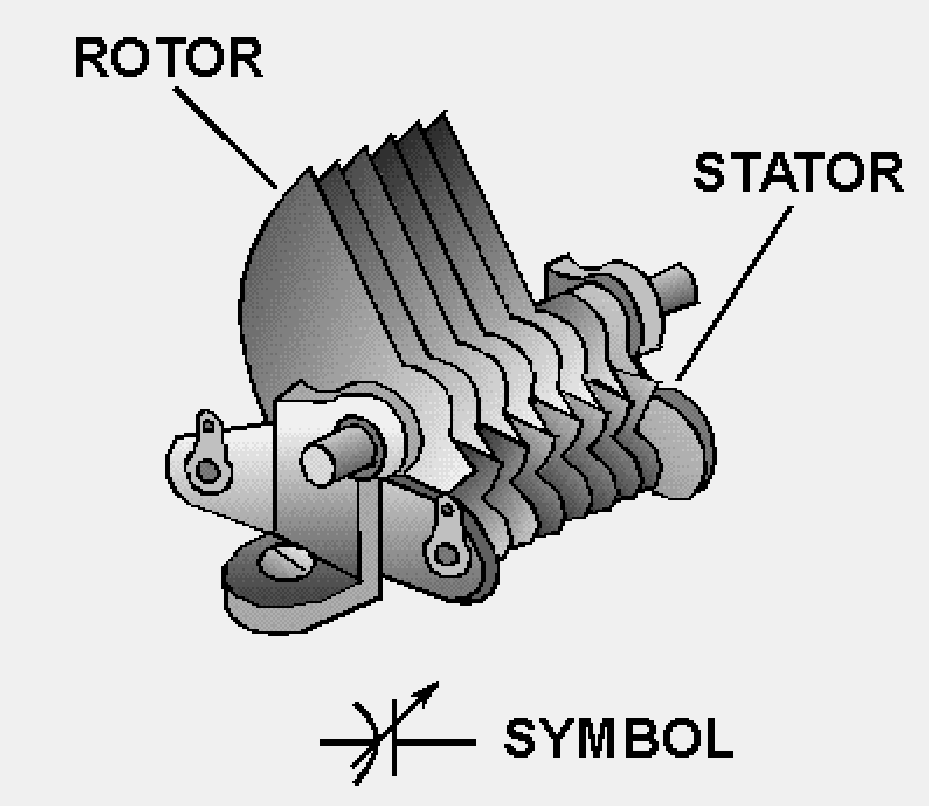

(https://cdn.instructables.com/FRL/AV48/HX8HK020/FRLAV48HX8HK020.MEDIUM.gif)

If you've built your Tic Tac crystal set and found that you can't receive a local AM radio station yet - don't panic - help is in on the way. You have just built the simplest version of the Tic Tac crystal radio and you may need to add one more component part for it to work properly. This is called a 'polyvaricon' - a miniature tuning capacitor, which can vary the frequency that your radio will receive at. You can see a picture of one up above - the small white box with the black knob next to it. It has 3 connecting tags - the one in the middle ('G') goes to the moving plates and the shaft, while the 2 outer tags ('O' and 'A') go to 2 sets of fixed plates. the smaller set of plates has a value of 60pF - pF is short for 'picoFarad' - a unit of measurement for capacitance. The larger set of plates is valued at 160 pF so that the combined value of the polyvaricon is 220 pF - or 220 picoFarads.

The other picture shows you what happens inside a basic tuning capacitor. there are 2 sets of metal plates - one set is fixed and the other set moves on a rotating shaft, connected to the tuning knob of your radio set. Both sets of plates are mounted on an insulating frame so that they won't 'short out' by touch each other.

The fixed capacitor C1, was chosen to tune somewhere close to the middle of the AM radio band. This band of frequencies starts at 531 kilohertz (Khz) and goes as far as 1701 Khz here in Australia. So we need a combination of coil and capacitor which will tune across all of those frequencies. Our simple Tic Tac radio is known as a 'series tuned' set. If you look at the circuit diagram, you can see electronic symbols for all components in the radio set.If you start at the top with the antenna symbol, you can see the capacitor C1 underneath that, the inductor L1 underneath C1 and then finally, the ground symbol - all wired in series with each other.

The diode D1 detector, the resistor and the earphone plus the link wire, can be considered as one block - the "detector unit". The Antenna wire, ground wire, inductor/coil and capacitor/polyvaricon, can be also be considered as another block - the "tuned circuit". So joining both blocks together, we have the tuned circuit that tunes in only one frequency, passing all other unwanted signals out through the ground wire to earth. This one 'tuned frequency' passes through the diode detector, which strips away the 'carrier wave' and leaves only the 'program signal' (voice, music etc,) behind, which is then fed via the resistor into the earphone. The earphone changes electrical impulses from the diode detector, into sound waves that we can hear. You need the R1 resistor to provide a pathway for the signals going through the diode, out to the ground wire connection. Without this resistor, the signals would sound very distorted and you couldn't hear the program signal very clearly.

Step 7: Adding the Polyvaricon Tuning Capacitor

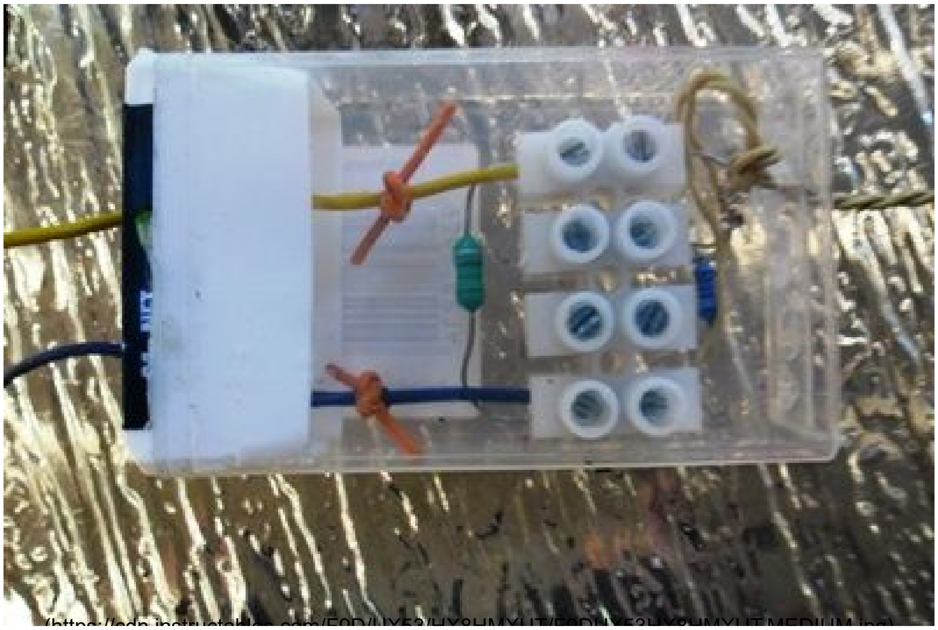

(https://cdn.instructables.com/F9D/HX53/HY8HMYUT/F9DHX53HY8HMYUT.MEDIUM.jpg)

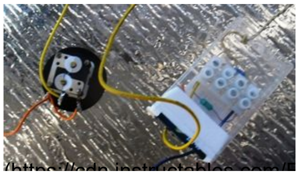

(https://cdn.instructables.com/F

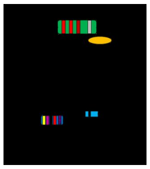

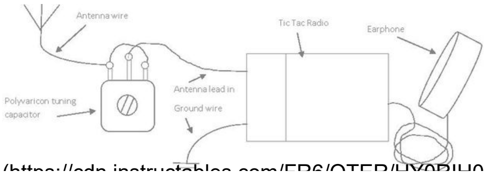

(https://cdn.instructables.com/FB6/OTEB/HY0BIH0

If you want to expand the tuning range of your Tic Tac Radio, than all you have to do is a simple modification (change) to the circuit of your radio set. You can see from the 2 pictures above, that there's an alternative way of building the crystal radio - you can use the matrix board method or you can use a 4 way screw terminal strip.

To use the screw terminal version, cut yourself a 4 way strip of terminals as shown in the diagram and pictures. Undo the screws right out as far as they will go without falling out of the strip. Connect the component leads and wires from the antenna/ground system as well as the earphone wires. Cut off any excess from component leads that you don't need. Wrap thin wires from the antenna/ground system and the earphones, around the thicker component leads before screwing the screws in tight.

Regardless of which method you have used, all we will now wire up the tuning capacitor between the actual Antenna wire and the junction (join) where the diode D1, the fixed capacitor C1 and one end of the inductor meet. Remove the capacitor from the matrix board, and connect the Antenna lead out wire straight to the join of the diode and inductor. Then take the polyvaricon and another piece of wire. Strip both ends of that wire and join the two outer tags (tagged 'O' and 'A' - the centre on is tagged 'G') and then connect you actual Antenna wire to one of the outer tags. Connect the antenna lead out wire, coming out of the box, to the 'G' (middle) tag of the tuning capacitor,so that the antenna wiring now looks like the picture up above.

If you're having problems following the pictures, then refer to the diagrams, which clearly shows all of the connections Make sure your ground wire is connected to the set, and then, listening with your earphone in/over your ear, slowly turn the tuning cap's flat knob, until you hear one or more stations. Congratulations - you now have a "tuneable" Tic Tac crystal set! Happy listening! And don't forget to post in your results, questions, problems etc... mk484 :)

Step 8: Tic Tac One Transistor Amplifier - Use Your Earbuds and Hear Great Sound...

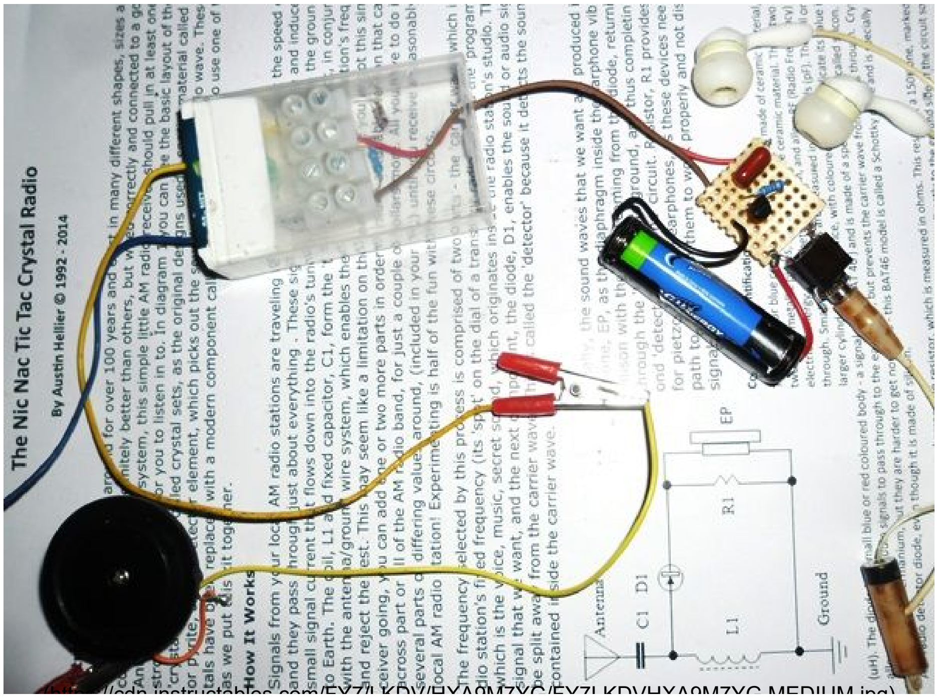
The Nic Nac Tic Tac Crystal Radio
By Austin Hellier © 1992 - 2014
How It Works
Antenna
C1
D1
L1
R1
EP
Ground

(http://cdn.instructables.com/FYZ/LKDV/HYA9MZYC/FYZLKDVHYA9MZYC.MEDIUM.jpg)

Hi there folks - this isn't really an extra 'Step' as such - it's a sneak preview at a picture of my upcoming Instructable for a one transistor amplifier, which will connect to the Tic Tac Radio - and give you some really good volume - in your Iphone earbuds!

Yes, it's entirely possible nowadays, with modern circuit design, to fit a one transistor amp in such a small space (yes kiddies - it WILL fit inside the smaller Tic Tac box...) and at the same time, get that great sound that comes from those "inside your ear" type earbuds.

This circuit took me about a half hour to complete, uses only 3 electronic components, costing about $1.00, a stereo earphone jack and a 1.5 volt AAA battery and plastic holder costing about another $3 to $4, so you can build the amplifier for about $5 all up - don't forget to shop around and - mums and dads - cheap batteries will be OK for this project, and the battery can simply be replaced without soldering or undoing half a dozen screws....

I can hear my Tic Tac Radio with great volume and clarity - and so will you, so "stay tuned" (ha ha ha) to this series of Instructables, and you will hear great sound too...

mk484

advertisement

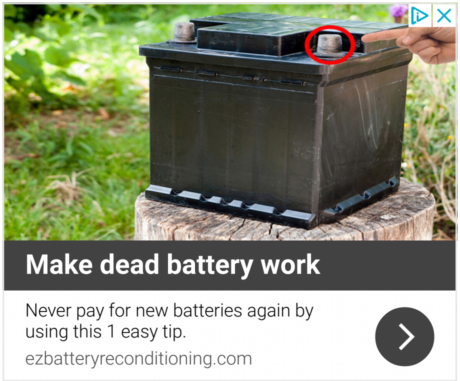

Make dead battery work

Never pay for new batteries again by using this 1 easy tip.

ezbatteryreconditioning.com

AIS Auto Insurance Specialists

COMPARE QUOTES

We Search.
We Find.
You Save.

Call toll free today (877) 215-2607

Comments

We have a be nice comment policy.
Please be positive and constructive.

I Made it! Add Images Post Comment

Colaris (/member/Colaris/) 2014-10-29 Reply

Can't you use regular headphones for both ears?

---

mk484 (/member/mk484/) ▸ Colaris (/member/Colaris/) 2014-10-29 Reply

Hi Colaris - earphones for crystal radios need to exhibit a fairly high impedance to the circuit (diode end and ground) - in the order of 2 to 4,000 ohms, as a bare minimum.

Low impedance phones (8 to 64 ohms like your average earbuds/large headphones) won't have anywhere near the impedance required. Trying to drive a low impedance 'load' such as a pair of earbuds, is like trying to drive a mains powered with a 9 volt battery - it's not going to happen...

While crystal earphones or the Murata substitute may not be perfect, they are more than adequate for kids and adults who are just experimenting with basic radio receivers.

---

mk484 (/member/mk484/) ▸ mk484 (/member/mk484/) 2016-01-16 Reply

Hi Colaris - what I meant to say was that

"Trying to drive a low impedance 'load' such as a pair of earbuds, is like trying to drive a mains powered (toaster) with a 9 volt battery - it's not going to happen..."

TravisM18 (/member/TravisM18/) 2015-11-03 Reply

I think you may need a tuning capacitor in there some where?

nupogodi (/member/nupogodi/) 2015-09-10 Reply

It is so pitty that in the near future all AM radiostations at least in USA and Europian Union will change to digital modulation so those amazing constructs will become useless :-(

mk484 (/member/mk484/) ▸ nupogodi (/member/nupogodi/) 2015-09-12 Reply

Thereare no plans to dump off AM or FM radio here, or in many other countries. Singapore has gone all digital but many other Asian countries are keeping their AM and FM transmitters for now

mk484 (/member/mk484/) 2015-07-15 Reply

Hello there - can any body tell me if they actually built this Tic Tac crystal radio, and if so, how did you go with it?

Thanks

MK484 :)

ViktorV1 (/member/ViktorV1/) 2014-09-28 Reply

So is there any way to have a smaller antenna or less grounding for this setup? I was thinking of making a "walkie-talkie" type thing with this and a simple AM transmitter. Are there things I can add to the circuit that would reduce the need for a big stationary antenna?

mk484 (/member/mk484/) ▸ ViktorV1 (/member/ViktorV1/) 2014-09-28 Reply

Hi Viktor - if you live really close to an AM transimitter or two, you could try a 2 foot length of very thick pvc covered wire as a portable antenna. You may still need a ground wire though. Walkie talkies on the AM band would tend to be very noisy though.

mk484 (/member/mk484/) 2014-07-10 Reply

Thanks to everyone who has marked this project as a favourite

jessyratfink (/member/jessyratfink/) 2014-07-08 Reply

What a neat build - I could see this getting assembled in schools really easily. :D

mk484 (/member/mk484/) ▸ jessyratfink (/member/jessyratfink/) 2014-07-08 Reply

Hi jessyratfink - I "invented" the Tic Tac xtal radio back in 1992 and used to sell it as a kit at flea markets - I also ran classes as a one of at a local community centre for adults. One little old lady who came and built her set, used a curtain rod in her kitchen as the antenna and the cold water tap in the sink as her ground. She told me that she used to sit there each night and listen to the ABC and smile whenever she reminded herself that SHE built it!

mk484 (/member/mk484/) 2014-07-07 Reply

Hi folks - the Tic Tac Radio article is finished - no more mods or additions for now, so feel free to post in with questions, problems in building your radio etc - mk484

---

mk484 (/member/mk484/) 2015-11-03 Reply

Hi Travis - this set was designed as a basic 'bare bones' one for the beginner. In the later construction pages, I do add a small polyvaricon tuner cap in place of the fixed ceramic cap so you can vary the frequency you can receive at.

The set is wired to be 'series tuned' so reception may not be perfect at first, but you can experiment with the height and length of the antenna wire to try and get better reception - eg: more stationsor better volume.

ALWAYS REMEMBER FOLKS TO CHECK OUT THE ELECTRICAL SAFETY OF YOUR ANTENNA AND GROUND WIRE SYSTEMS BEFORE CONNECTNG YOUR CRYSTAL SET TO THEM...

---

charris8 (/member/charris8/) 2015-10-15 Reply

Do you really need it grounding? What are the chances of an electrical surge from a non powered device?

---

mk484 (/member/mk484/) ▸ charris8 (/member/charris8/) 2015-10-15 Reply

Hi - a ground wire provides a return path so that the AM radio signal can be fully received by the set. If you leave the ground wire off, you may not get very good volume in the earphone, or worst still, you may not get a signal at all.

You need to install a safe ground wire - one that goes straight to a metal pole, embedded in soft, wet soil. Make sure there are no pipes or cables underneath the ground where you are going to do this. If you have any doubts, consult a qualified electrician who can advise you on how to do that safely.

As for electrical shocks, they are possible during thunderstorms, due to potential lightening strikes, so it's best to disconnect your crystal set from both the antenna and ground wires for the duration of the storm. Quite often you can hear the 'crackle' of distant thunder/lightening in the distance, long before you can see/hear them.

⬇ More Comments

Nic Nac Tic Tac Crystal Radio Set - an Ideal Primary School Project...

by mk484 (/member/mk484/) in science (/technology/science/)

by mk484 (/member/mk484/) in science (/technology/science/)

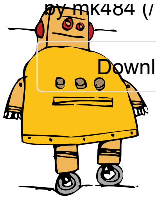

Download

(/id/Nicu-Nax-Tiel-Taco-Crystal-Radio-Set-C-Austin-Hellier/)

8 Steps

+ Collection

I Made it!

Favorite

Share

Newsletter

Let your inbox help you discover our best projects, classes, and contests. Instructables will help you learn how to make anything!

enter email | I'm in!

About Us

Who We Are (/about/)

Advertise (/advertise/)

Contact (/about/contact.jsp)

Jobs (/community/Positions-available-at-Instructables/)

Help (/id/how-to-write-a-great-instructable/)

advertisement

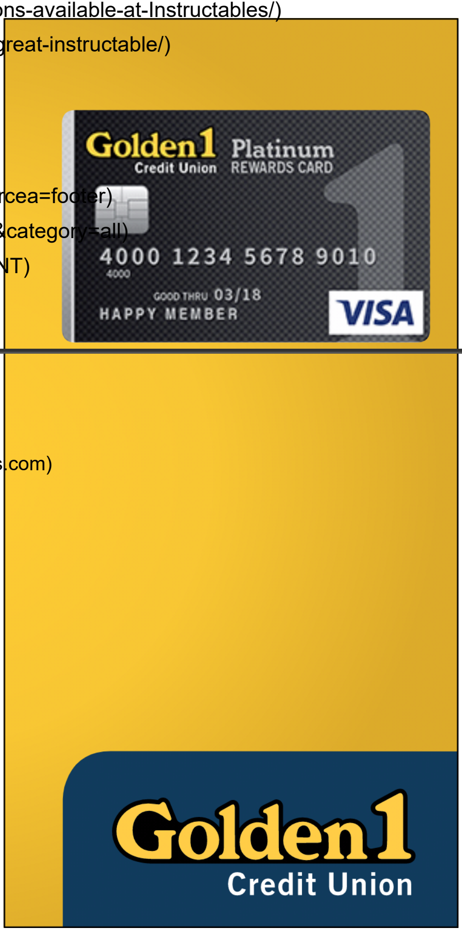

Find Us

Facebook (http://www.facebook.com/instructables)

Youtube (http://www.youtube.com/user/instructablestv)

Twitter (http://www.twitter.com/instructables)

Pinterest (http://www.pinterest.com/instructables)

Google+ (https://plus.google.com/+instructables)

Resources

For Teachers (/teachers/)

Residency Program (/pier9residency)

Gift Premium Account (/account/give?sourcea=footer)

Forums (/community/?categoryGroup=all&category=all)

Answers (/tag/type-question/?sort=RECENT)

Sitemap (/sitemap/)

Terms of Service (http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=21959721) |
Privacy Statement (http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=21292079) |
Legal Notices & Trademarks (http://usa.autodesk.com/legal-notices-trademarks/) | Mobile Site (https://www.instructables.com)
(http://usa.autodesk.com/adsk/servlet/pc/index?id=20781545&siteID=123112)
© 2017 Autodesk, Inc.

advertisement

See why millions trust us for Home & Auto.
Contact Me ▶
Darren Reuter, Agent
Insurance Lic. #: 0D97352
San Rafael, CA 94901

Queer news and culture.
For everyone.
INTO MORE.COM

START HERE →

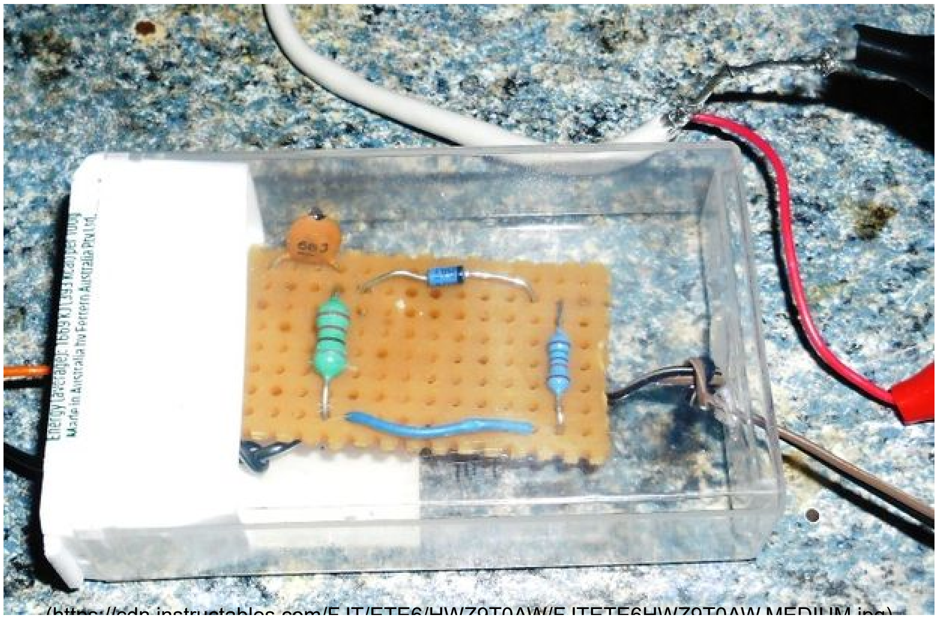
(https://cdn.instructables.com/FJT/FTE6/HWZ9T0AW/FJTFTE6HWZ9T0AW.MEDIUM.jpg)

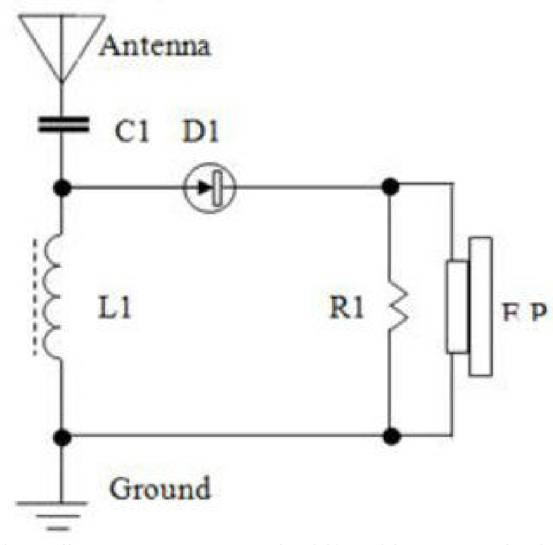

(https://cdn.instructables.com/F8O/OFKG/HX8HK0L5/F8O(

(https://cdn.instructables.com/FZL/AJM

About This Instructable

10,581 views
84 favorites

License: Public Domain

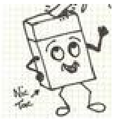
mk484 (/member/mk484/)
(http://)
(/member/mk484/)
Follow
82

Hi there! **This Instructable is all about building a basic crystal radio set that is so simple to build and understand, that a child could do it - with help from mum or dad** - or even at school, as a class project. Parts can be bought "off the shelf" at Jaycar Electronics and other suppliers, or purchased online via Ebay and Paypal.

I built my first ever crystal set when I was 9 years old, but that was a different world. We're going to build a 21st century style radio with all modern components which are relatively easy to obtain at your local electronics store or on the Internet.

The radio is contained inside a Tic Tac box, and is called "The Nic Nac Tic Tac Radio". It is built on a square of matrix board, with the component leads poked through the holes in the board, and joined underneath. It is a very simple crystal set, designed to initially receive only one station, but you can expand the tuning range with the addition of only one extra component.

The Tic Tac radio can be constructed by simply twisting the component leads together, underneath the matrix board and attaching antenna/ground wires as well as the earphone wires to the board, using that same technique. You don't have to solder any wires, but there is another way to join them all together - more on that later.

The Tic Tac Radio is a very easy set to put together - you don't need to wind bulky coils or use heavy tools and screw strange looking parts onto a large breadboard. See step 1 for the parts list you will need.

Step 1: Parts and Tools You Will Need

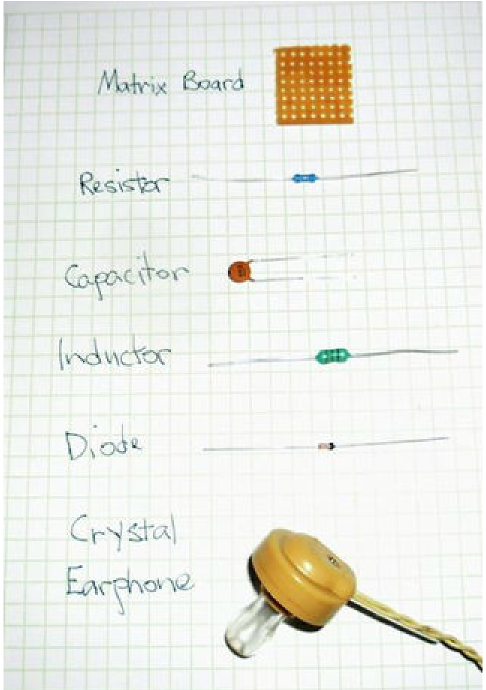

(https://cdn.instructables.com/F9B/WLFJ/HX5H54T

(https://cdn.instructables.com/F99/U

(https://cdn.instructables.com/F

(https://cdn.instructables.com/

You will need to acquire the following components from a local electronics shop or an online electronics business.

Bio: Started as a hobbyist at 9 - built my first crystal radio on one of mum's prized cutting boards (eeek) - Worked in 2 electrical/electronics ... More » (/member/mk484/)

Queer news and culture.

For everyone.

INTO MORE.COM

I am adding the catalogue numbers for Jaycar Electronics, so if you live in Australia or North America, you may be able to visit a Jaycar store and buy the parts over the counter, if not then you may be able to do an online mail order via Paypal:

Parts List

Resistor - 47k - yellow-purple-black-red and brown - RR 0612 (pkt of 8)

Capacitor - 68pF - ceramic x 2 - RC 5322 (pkt of 2) and a 100 or 120 pF value as well for experiments.

Inductor - 220 uH - red-red-brown silver - LF 1538 (resistive type)

Polyvaricon tuning capacitor - 220 pF - RV 5728 - with knob and mounting screws

Diode - BAT46 - ZR 1141 (You can also use a 1N34A Germanium Diode too if you have one at home)

Ceramic Earphone - AS 3305 *

A 25 meter roll of yellow hook up wire for the Antenna wire and a 3 meter length of wire for the Ground wire.

Please note that some Jaycar parts come in multiples of 2 or more per packet. And please note the following:

PLEASE NOTE: THIS PROJECT/KIT CONTAINS SMALL PARTS THAT MAY FORM A CHOKING HAZARD FOR SMALL CHILDREN OR PETS. NOT SUITABLE FOR CHILDREN UNDER FIVE (5) YEARS OLD.

*A normal crystal radio earphone is OK, but if you can't get one of these, or if the one you bought goes dead (as they sometimes do,) you can use a substitute, such as the Murata PKM44EW passive transducer (see picture above) which is available from an old Telstra TF200 touchphone, (the one on the left in the diagram above,) or an equivalent, such as the ARIO transducer, from an old Telstra T1000 pushbutton phone.

The ARIO unit is soldered to the phone's pc board so you'll need to be able to unsolder the three mounting pins underneath the board, or find someone in the neighbourhood who is able.

Take the back off the TF200 (if you've obtained one of these phones,) and you'll see a black disc shaped object 2" round by 1/2' thick - with a red and black wire. Unplug the wires from the circuit board, and unscrew any retaining screws and remove the transducer. Cut the mini plug off, **carefully** strip the insulation from the ends of the wires and extend them by about 18" with 2 thin lengths of hookup wire. These pietzo devices make good earphones for crystal sets and can be housed in an old pair of ear muffs.

Miscellaneous Materials:

A Tic Tac box (smaller size)

A piece of matrix board at least 7 holes long by 8 holes across. Cut the board to fit neatly inside the Tic Tac box.

A short length of 2 differently coloured wires 60 mm in length and 2 crocodile clips with red and black plastic covers

A length of antenna wire at least 25 metres long and a 3 metre length of a different colour for the Ground wire.

A metal rod or cold water pipe for the ground stake. **Be careful which pipes you connect your Ground wire to.**

Queer news and culture.
For everyone.
INTO MORE.COM

You will need a small sharp object for punching holes in the Tic Tac box.

A small pair of wire cutters for cutting and stripping wires.

Step 2: How a Crystal Radio Works

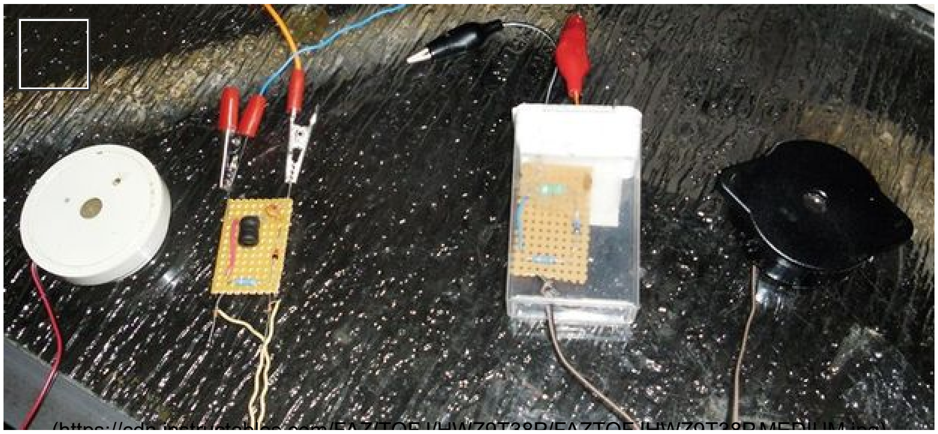

(https://cdn.instructables.com/FAZ7TOEJHW79T38P/FAZ7TOEJHW79T38P.MEDIUM.jpg)

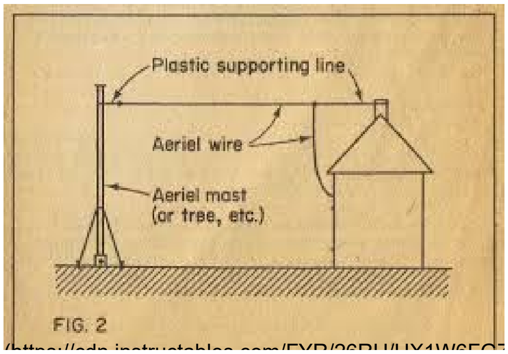

FIG. 2

(https://cdn.instructables.com/FXB/26BU/HX1W6FG7

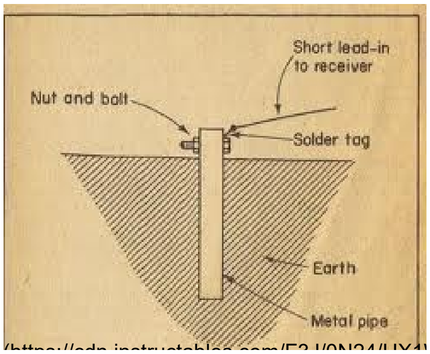

(https://cdn.instructables.com/F3J/0N24/HX1V

Queer news and culture.

For everyone.

INTO MORE.COM

Radio signals consists of two parts - the 'carrier wave' which is the AM radio station's frequency of operation, and the 'program signal' which is mixed with the carrier wave for transmission.

Radio waves travel out from the AM transmitter tower through the atmosphere. We want to capture one specific frequency so we can listen to it, so we need the antenna/ground wire system to capture that signal. We also need a 'tuned circuit' that will filter out the desired AM signal, and discard the rest, so that all the other unwanted radio signals pass out through the ground wire to earth.

Two components in our circuit will perform that task for us. The **capacitor** C1, together with L1 **inductor**, form a basic 'series tuned' circuit. Their respective values will determine just which local AM radio station we will capture.

We also need a **diode** to 'detect' the voice and music, so we can hear them in our **earphone**, which transduces electrical signals into sound waves that we can hear.

In the photo above you can see a completed Tic Tac crystal radio. It is already inside the box. The other set is connected to the antenna/ground wire circuit, undergoing a 'soak test'. It is necessary to do this to ensure that the radio will work once inside the box!

The first diagram shows a typical antenna wire installation. Coming out of a window, the wire is anchored to the building and then over some distance (10 metres +) to a nearby tree or other building.

You must take great care not to erect antenna wires near to power or telephone cables, near your home!

The ground wire comes out the same window and is anchored to a metal pipe/water pipe or metal ground stake, embedded in soft, moist soil.

You must not connect ground wires to electrical mains wiring installations, including wall power outlets!

Another electrical hazard to consider is **lightning strikes!** Although it is very rare for anyone to be seriously injured or die from a lightning strike it is **not impossible**. So, if you hear a thunderstorm coming your way (you may hear the lightning 'crashes' in your earphone first,) then **disconnect your antenna wire immediately**, connect it to the ground wire and put it well up out of the way. Stay well clear of this temporary antenna/ground connection until the storm has completely passed away from your general area - miles away!

Queer news and culture.
For everyone.
INTO MORE.COM

Step 3: How to Build the Tic Tac Crystal Radio Set

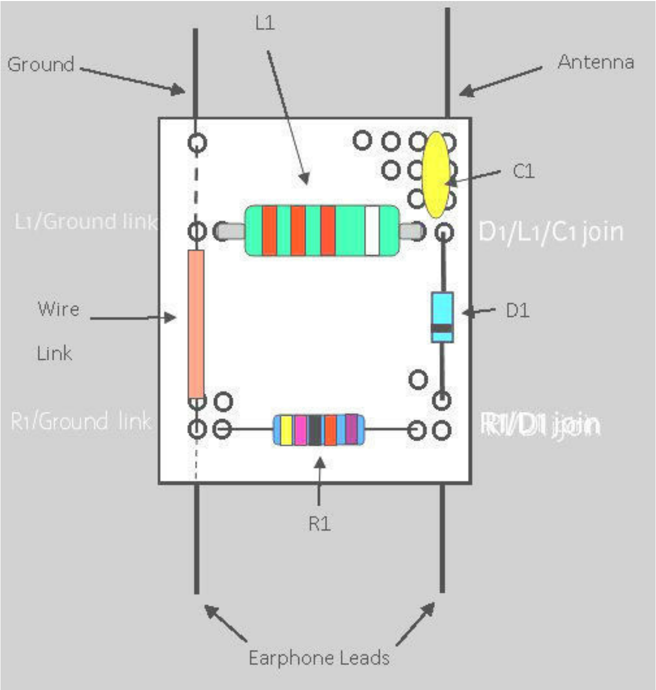

(https://cdn.instructables.com/FA5/EW4E/HXA9BXE0/FA5EW4EHXA

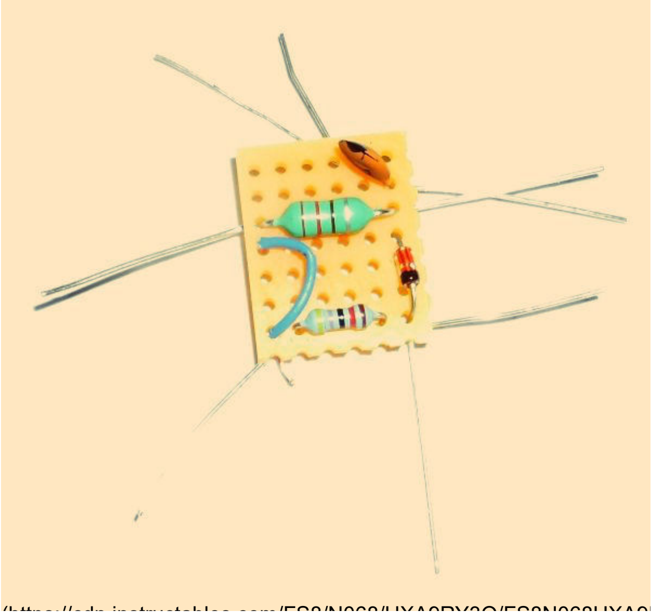

(https://cdn.instructables.com/FS8/N063/HXA9BX3O/FS8N063HXA

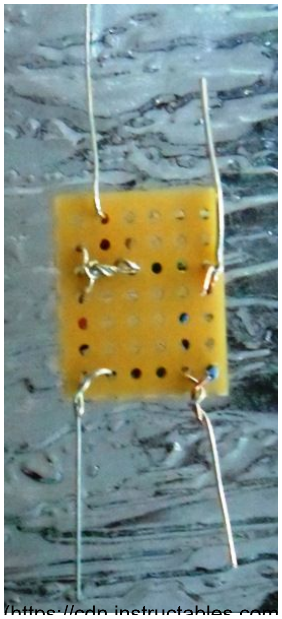

(https://cdn.instructables.com

1. Take the components and lay them out on a clean surface.

2. Take the square of matrix board (this is also called 'perf board' because of all the holes or perforations in the material) and beginning with the 220uH (uH = micro-Henry) and place it as shown in the diagram.

3. Then take the diode, capacitor and resistor and place them in the places shown for them, making sure that the coloured band at one end of the diode joins with the resistor wire, as shown

4. Then take each junction where wires come together through their respective holes and gently twist them together, until they form a neat, tight bundle. Take your side cutters and cut off any excess length, taking care not to cut any one wire too short, so that it comes undone from the join.

.5. Take the link wire, strip 2 centimetres of insulation from each end of the wire and install that wire between the free ends of the inductor and resistor, and the matrix board construction is complete.

6. Then take the ceramic earphone, cut the plug off the end, and strip the 2 wire ends about 1.5 centimetres in length. Wrap each earphone wire around each end of the resistor component, underneath the matrix board.

7. Finally, strip the insulation off both ends of the 60mm long differently coloured wires, and attach them to the matrix board - one goes to the inductor/link wire junction and this will be the Ground connection wire. Attach the other one to the free end of capacitor C1 and this will be the antenna wire connection. Both connections are made underneath the matrix board. The Nic Nac Tic Tac Crystal Radio is now ready for testing.

Queer news and culture.

For everyone.

INTO MORE.COM

Step 4: Final Checks and Installation of Wires

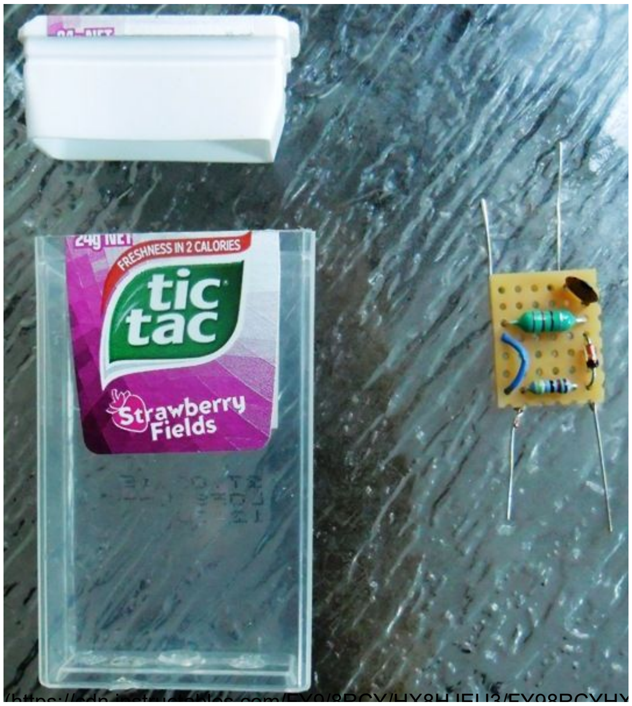

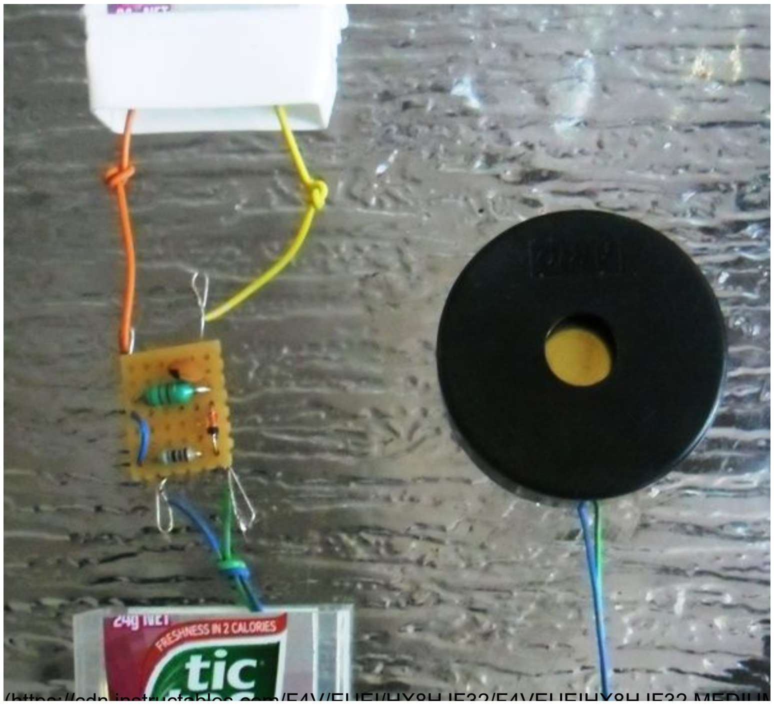

(https://cdn.instructables.com/F4V/FUFI/HY8HJF32/F4VFUFIHY8HJF32.MEDIU

When you've finished constructing the matrix board circuit and have clipped all the excess component leads off, puncture three (3) holes in the Tic Tac box - 2 small holes on the top lid of the box, about an inch (24mm) apart, so that your antenna and ground wire leads can come through the box lid, and be connected to the matrix board at those 2 points.

Feed the wires through the holes and then tie small knots in each one, near the underside of the box lid, so that they won't pull back out if strained, and disconnect themselves from the matrix board.

Then make one larger hole in the centre of the bottom of the box (clear part) so that your earphone wires can be fed through to the connecting points on that part of the matrix board. Tie a larger knot in the earphone lead so as to prevent it from pulling out of the box if it is strained. Strip the insulation off the earphone leads, and wrap them around the matrix board leads at those 2 points. You are now ready to test out the Nic Nac Tic Tac Crystal Radio.

Step 5: Testing Out the Tic Tac Radio

Queer news and culture.
For everyone.
INTO MORE.COM

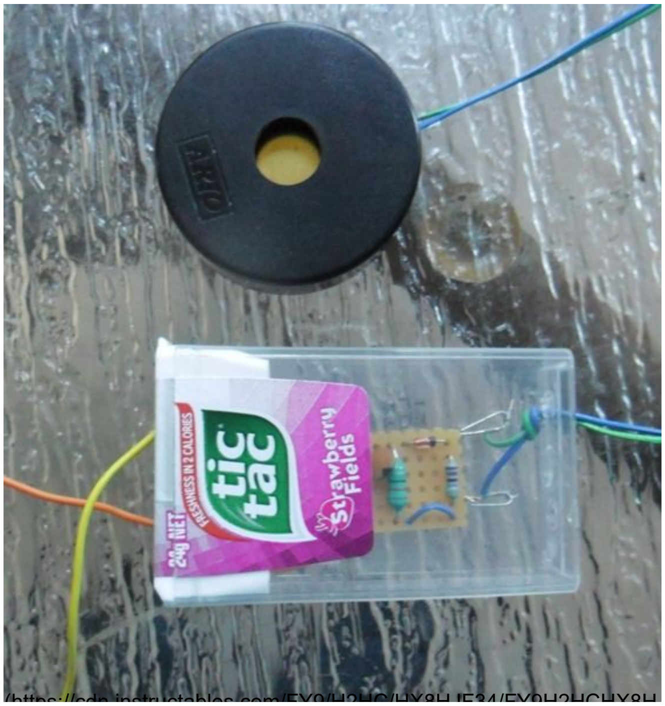
tic
tac
Strawberry Fields
24g NET
FRESHNESS IN 2 CALORIES

(https://cdn.instructables.com/FX9/H2HC/HX8HJE34/FX9H2HCHX8H

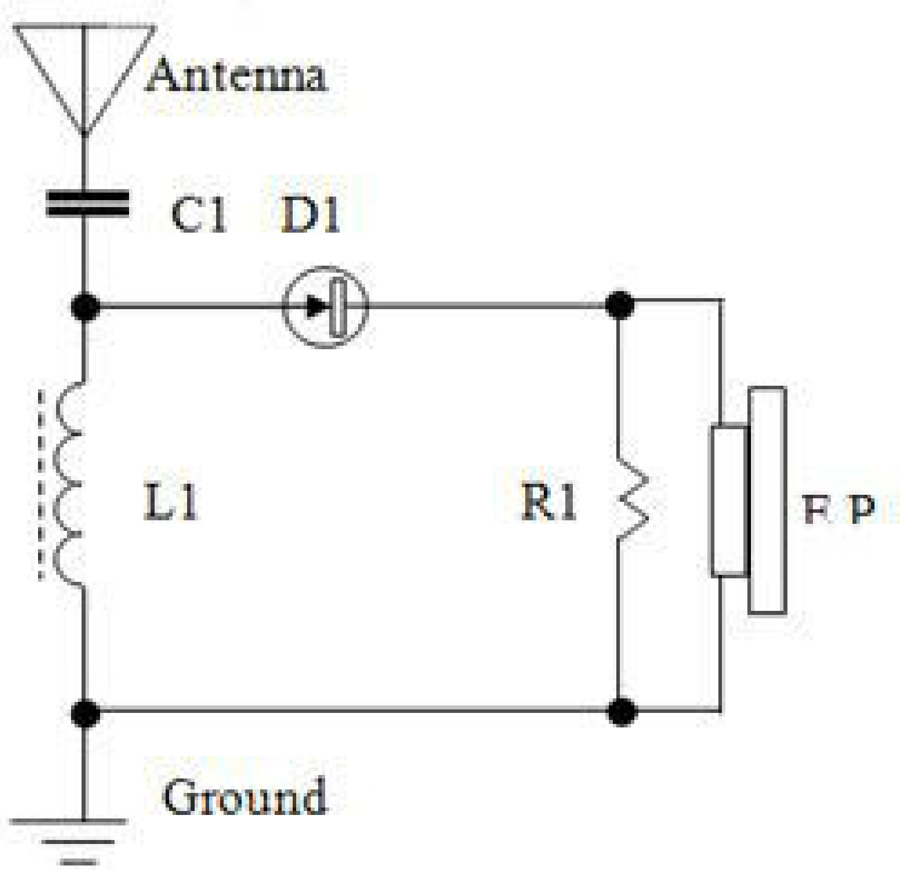
Antenna
C1
D1
L1
R1
E P
Ground

(https://cdn.instructables.com/FL6/41CU/HX5H55A2/FL641CUHX5H55A2.MEDIUM.jpg)

Queer news and culture.
For everyone.
INTO MORE.COM

Take your finished Tic Tac Radio (with the lid part gently pushed back inside the top of the box) and connect your Antenna and Ground wires to the lead outs from the box lid. Place the earphone in/over your ear,and listen carefully for a local AM radio station. This crystal radio is a simple one, and you may have to make one or two adjustments to the components, before you succeed in receiving one or more local AM radio stations, in your area.

If you can't hear anything in the earphone, don't panic. It might just be a simple wiring mistake, which is easily fixed. Go back over all of the steps, making sure that you have the right value components from the electronics store. Make sure that each component is in the right place on the matrix board, (don't confuse the L1 inductor with the R1 resistor - they look a lot like each other!) and that no component wires have come undone from the twisted joins you have just made. I'll be writing up a troubleshooting step soon, so if you run into any problems, post your questions in the 'comments' section down at the bottom of the page, and I'll try and answer them as soon as possible.

Make sure that your antenna and ground wires aren't snagged on anything metal or anything **dangerous!**

If you have ANY doubts about the electrical safety of your antenna or ground wires, then consult a licensed electrical trades person, who will be able to advise you on electrical safety principles and procedures!

Always remember that electrical safety is your responsibility! If you don't think it is safe to proceed, then don't!

You will need at least 10 to 15 metres of antenna wire, strung between 2 insulating points (not connected to anything metallic, or that gets wet,) at least 2 to 3 metres in height - anything less than this minimum arrangement may mean that you cannot receive any signals at all.

Some places are regarded as 'radio dead spots', so you may need to try an open space, such as a park or a remote corner of a beach. If you do erect antenna and ground wires in public places, hang some streamers or balloons off of them, to alert people to their presence, otherwise people going past may become entangled with them - and get cranky with you!!!

As a final word for now, you'll be happy to know that the completed "Nic Nac Tic Tac Radio" shown in the picture above, picked up local AM radio station **1116 khz 4BC** here in Brisbane, with a very clear signal and quite good volume. It works! So be safe kids, have some fun and look forward to more... mk484

Queer news and culture.

For everyone.

INTO MORE.COM

Step 6: Nic Nac Extras

(https://cdn.instructables.com/FGL/BHOH/HX8HJG9V/FGLBHOHHX8HJG9V.MEDIUM.jpg)

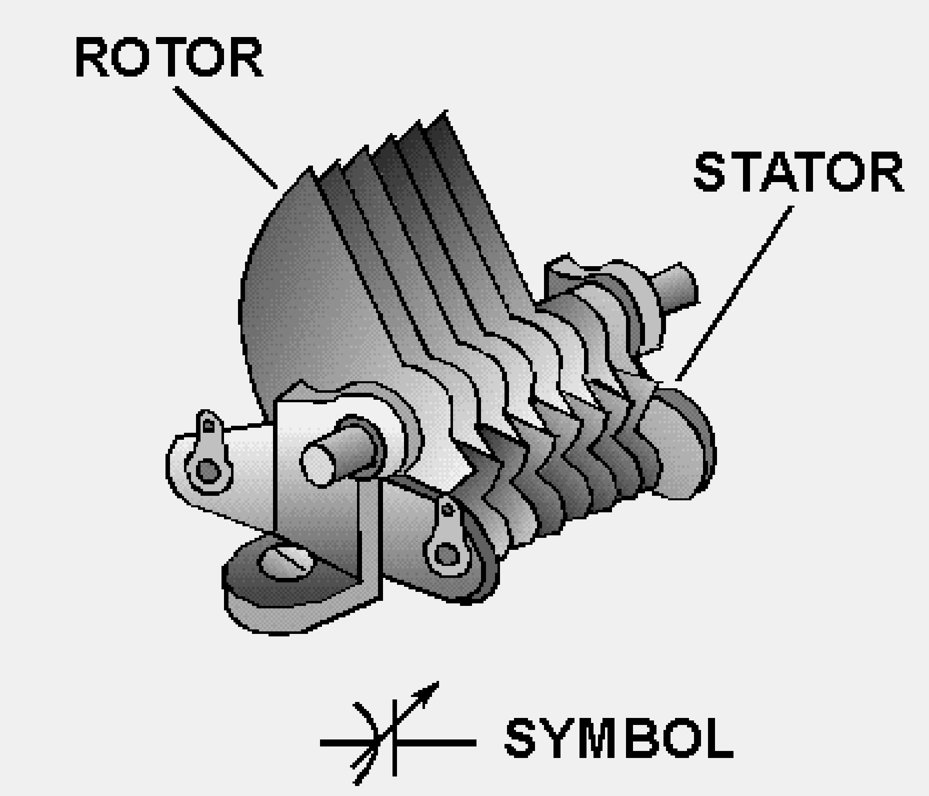

(https://cdn.instructables.com/FBL/AV48/HX8HK020/FBLAV48HX8HK020.MEDIUM.gif)

Queer news and culture.
For everyone.
INTO MORE.COM

If you've built your Tic Tac crystal set and found that you can't receive a local AM radio station yet - don't panic - help is in on the way. You have just built the simplest version of the Tic Tac crystal radio and you may need to add one more component part for it to work properly. This is called a 'polyvaricon' - a miniature tuning capacitor, which can vary the frequency that your radio will receive at. You can see a picture of one up above - the small white box with the black knob next to it. It has 3 connecting tags - the one in the middle ('G') goes to the moving plates and the shaft, while the 2 outer tags ('O' and 'A') go to 2 sets of fixed plates. the smaller set of plates has a value of 60pF - pF is short for 'picoFarad' - a unit of measurement for capacitance. The larger set of plates is valued at 160 pF so that the combined value of the polyvaricon is 220 pF - or 220 picoFarads.

The other picture shows you what happens inside a basic tuning capacitor. there are 2 sets of metal plates - one set is fixed and the other set moves on a rotating shaft, connected to the tuning knob of your radio set. Both sets of plates are mounted on an insulating frame so that they won't 'short out' by touch each other.

The fixed capacitor C1, was chosen to tune somewhere close to the middle of the AM radio band. This band of frequencies starts at 531 kilohertz (Khz) and goes as far as 1701 Khz here in Australia. So we need a combination of coil and capacitor which will tune across all of those frequencies. Our simple Tic Tac radio is known as a 'series tuned' set. If you look at the circuit diagram, you can see electronic symbols for all components in the radio set.If you start at the top with the antenna symbol, you can see the capacitor C1 underneath that, the inductor L1 underneath C1 and then finally, the ground symbol - all wired in series with each other.

The diode D1 detector, the resistor and the earphone plus the link wire, can be considered as one block - the "detector unit". The Antenna wire, ground wire, inductor/coil and capacitor/polyvaricon, can be also be considered as another block - the "tuned circuit". So joining both blocks together, we have the tuned circuit that tunes in only one frequency, passing all other unwanted signals out through the ground wire to earth. This one 'tuned frequency' passes through the diode detector, which strips away the 'carrier wave' and leaves only the 'program signal' (voice, music etc,) behind, which is then fed via the resistor into the earphone. The earphone changes electrical impulses from the diode detector, into sound waves that we can hear. You need the R1 resistor to provide a pathway for the signals going through the diode, out to the ground wire connection. Without this resistor, the signals would sound very distorted and you couldn't hear the program signal very clearly.

Queer news and culture.

For everyone.

INTO MORE.COM

Step 7: Adding the Polyvaricon Tuning Capacitor

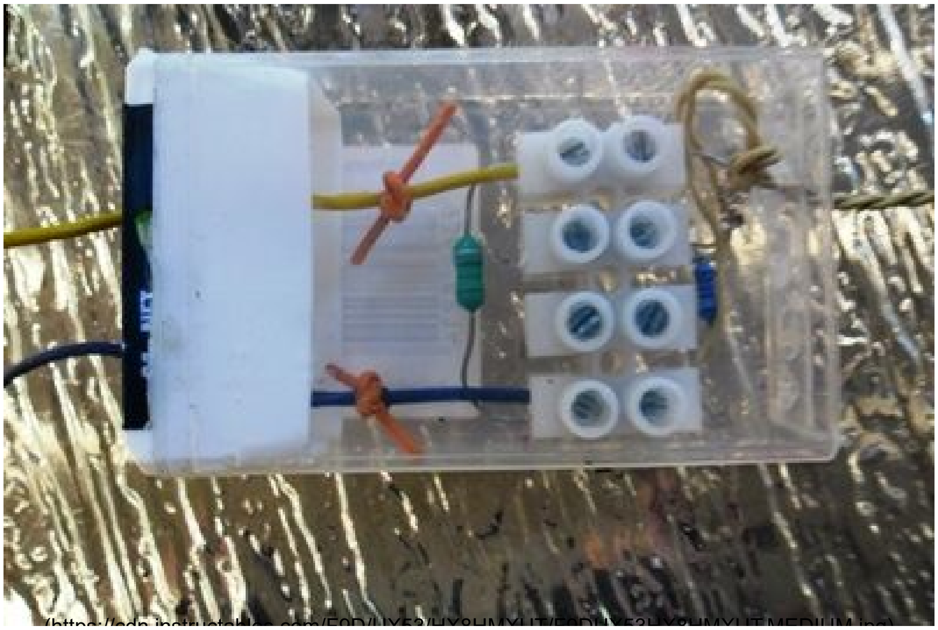

(https://cdn.instructables.com/F9D/UX53/HY8HMYUT/F9DUX53HY8HMYUT.MEDIUM.jpg)

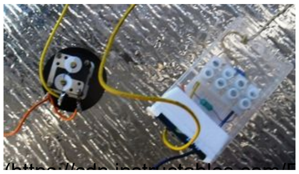

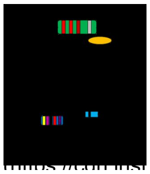

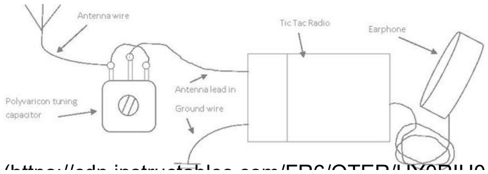

(https://cdn.instructables.com/F (https://cdn.instructables.com/FB6/OTEP/HY0BIU0

If you want to expand the tuning range of your Tic Tac Radio, than all you have to do is a simple modification (change) to the circuit of your radio set. You can see from the 2 pictures above, that there's an alternative way of building the crystal radio - you can use the matrix board method or you can use a 4 way screw terminal strip.

To use the screw terminal version, cut yourself a 4 way strip of terminals as shown in the diagram and pictures. Undo the screws right out as far as they will go without falling out of the strip. Connect the component leads and wires from the antenna/ground system as well as the earphone wires. Cut off any excess from component leads that you don't need. Wrap thin wires from the antenna/ground system and the earphones, around the thicker component leads before screwing the screws in tight.

Regardless of which method you have used, all we will now wire up the tuning capacitor between the actual Antenna wire and the junction (join) where the diode D1, the fixed capacitor C1 and one end of the inductor meet. Remove the capacitor from the matrix board, and connect the Antenna lead out wire straight to the join of the diode and inductor. Then take the polyvaricon and another piece of wire. Strip both ends of that wire and join the two outer tags (tagged 'O' and 'A' - the centre on is tagged 'G') and then connect you actual Antenna wire to one of the outer tags. Connect the antenna lead out wire, coming out of the box, to the 'G' (middle) tag of the tuning capacitor,so that the antenna wiring now looks like the picture up above.

If you're having problems following the pictures, then refer to the diagrams, which clearly shows all of the connections Make sure your ground wire is connected to the set, and then, listening with your earphone in/over your ear, slowly turn the tuning cap's flat knob, until you hear one or more stations. Congratulations - you now have a "tuneable" Tic Tac crystal set! Happy listening! And don't forget to post in your results, questions, problems etc... mk484 :)

Queer news and culture.

For everyone.

INTO MORE.COM

Step 8: Tic Tac One Transistor Amplifier - Use Your Earbuds and Hear Great Sound...

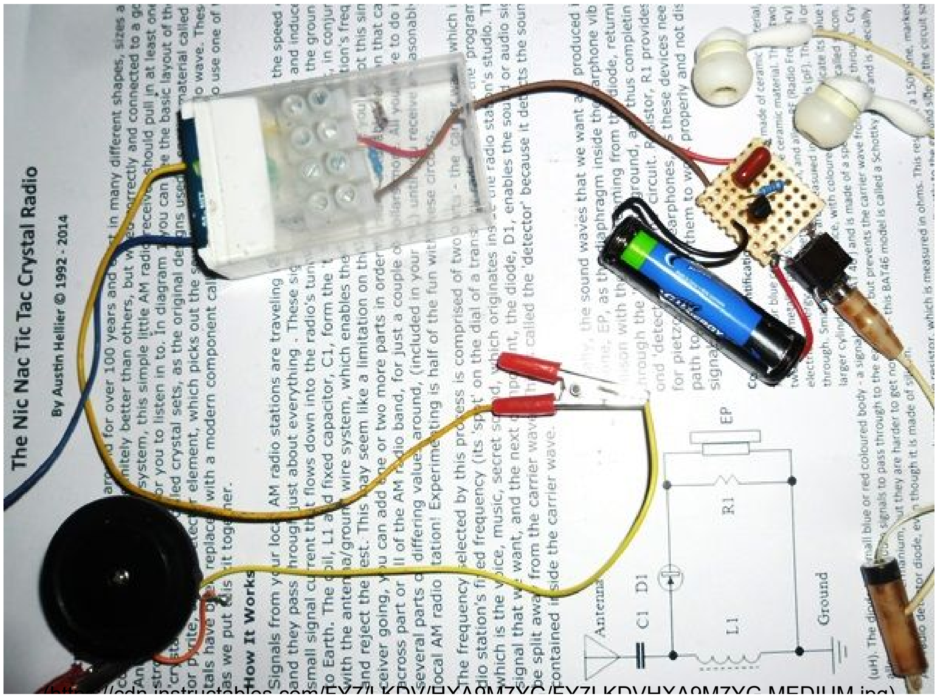
The Nic Nac Tic Tac Crystal Radio
By Austin Hellier © 1992 - 2014
How It Works
Antenna
C1
D1
L1
R1
EP
Ground

//cdn.instructables.com/FYZ/LKDV/HYA9MZYC/FYZLKDVHYA9MZYC.MEDIUM.jpg)

Hi there folks - this isn't really an extra 'Step' as such - it's a sneak preview at a picture of my upcoming Instructable for a one transistor amplifier, which will connect to the Tic Tac Radio - and give you some really good volume - in your Iphone earbuds!

Yes, it's entirely possible nowadays, with modern circuit design, to fit a one transistor amp in such a small space (yes kiddies - it WILL fit inside the smaller Tic Tac box...) and at the same time, get that great sound that comes from those "inside your ear" type earbuds.

This circuit took me about a half hour to complete, uses only 3 electronic components, costing about $1.00, a stereo earphone jack and a 1.5 volt AAA battery and plastic holder costing about another $3 to $4, so you can build the amplifier for about $5 all up - don't forget to shop around and - mums and dads - cheap batteries will be OK for this project, and the battery can simply be replaced without soldering or undoing half a dozen screws....

I can hear my Tic Tac Radio with great volume and clarity - and so will you, so "stay tuned" (ha ha ha) to this series of Instructables, and you will hear great sound too...

mk484

advertisement

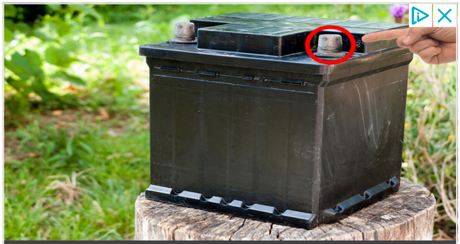

Make dead battery work

Never pay for new batteries again by using this 1 easy tip.

ezbatteryreconditioning.com

AIS Auto Insurance Specialists

COMPARE QUOTES

We Search.
We Find.
You Save.

Call toll free today (877) 215-2607

Queer news and culture.

For everyone.

INTO MORE.COM

Comments

be nice

We have a be nice comment policy.
Please be positive and constructive.

I Made it! Add Images Post Comment

Colaris (/member/Colaris/) 2014-10-29 Reply

Can't you use regular headphones for both ears?

mk484 (/member/mk484/) ▸ Colaris (/member/Colaris/) 2014-10-29 Reply

Hi Colaris - earphones for crystal radios need to exhibit a fairly high impedance to the circuit (diode end and ground) - in the order of 2 to 4,000 ohms, as a bare minimum.

Low impedance phones (8 to 64 ohms like your average earbuds/large headphones) won't have anywhere near the impedance required. Trying to drive a low impedance 'load' such as a pair of earbuds, is like trying to drive a mains powered with a 9 volt battery - it's not going to happen...

While crystal earphones or the Murata substitute may not be perfect, they are more than adequate for kids and adults who are just experimenting with basic radio receivers.

mk484 (/member/mk484/) ▸ mk484 (/member/mk484/) 2016-01-16 Reply

Hi Colaris - what I meant to say was that

"Trying to drive a low impedance 'load' such as a pair of earbuds, is like trying to drive a mains powered (toaster) with a 9 volt battery - it's not going to happen..."

TravisM18 (/member/TravisM18/) 2015-11-03 Reply

I think you may need a tuning capacitor in there some where?

nupogodi (/member/nupogodi/) 2015-09-10 Reply

It is so pitty that in the near future all AM radiostations at least in USA and Europian Union will change to digital modulation so those amazing constructs will become useless :-(

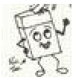

mk484 (/member/mk484/) ▸ nupogodi (/member/nupogodi/) 2015-09-12 Reply

Thereare no plans to dump off AM or FM radio here, or in many other countries. Singapore has gone all digital but many other Asian countries are keeping their AM and FM transmitters for now

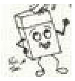

mk484 (/member/mk484/) 2015-07-15 Reply

Hello there - can any body tell me if they actually built this Tic Tac crystal radio, and if so, how did you go with it?

Thanks

MK484 :)

ViktorV1 (/member/ViktorV1/) 2014-09-28 Reply

So is there any way to have a smaller antenna or less grounding for this setup? I was thinking of making a "walkie-talkie" type thing with this and a simple AM transmitter. Are there things I can add to the circuit that would reduce the need for a big stationary antenna?

mk484 (/member/mk484/) ▸ ViktorV1 (/member/ViktorV1/) 2014-09-28 Reply

Hi Viktor - if you live really close to an AM transimitter or two, you could try a 2 foot length of very thick pvc covered wire as a portable antenna. You may still need a ground wire though. Walkie talkies on the AM band would tend to be very noisy though.

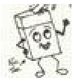

mk484 (/member/mk484/) 2014-07-10 Reply

Thanks to everyone who has marked this project as a favourite

jessyratfink (/member/jessyratfink/) 2014-07-08 Reply

What a neat build - I could see this getting assembled in schools really easily. :D

mk484 (/member/mk484/) ▸ jessyratfink (/member/jessyratfink/) 2014-07-08 Reply

Hi jessyratfink - I "invented" the Tic Tac xtal radio back in 1992 and used to sell it as a kit at flea markets - I also ran classes as a one of at a local community centre for adults. One little old lady who came and built her set, used a curtain rod in her kitchen as the antenna and the cold water tap in the sink as her ground. She told me that she used to sit there each night and listen to the ABC and smile whenever she reminded herself that SHE built it!

Queer news and culture.

For everyone.

INTO MORE.COM

mk484 (/member/mk484/) 2014-07-07 Reply

Hi folks - the Tic Tac Radio article is finished - no more mods or additions for now, so feel free to post in with questions, problems in building your radio etc - mk484

mk484 (/member/mk484/) 2015-11-03 Reply

Hi Travis - this set was designed as a basic 'bare bones' one for the beginner. In the later construction pages, I do add a small polyvaricon tuner cap in place of the fixed ceramic cap so you can vary the frequency you can receive at.

The set is wired to be 'series tuned' so reception may not be perfect at first, but you can experiment with the height and length of the antenna wire to try and get better reception - eg: more stationsor better volume.

ALWAYS REMEMBER FOLKS TO CHECK OUT THE ELECTRICAL SAFETY OF YOUR ANTENNA AND GROUND WIRE SYSTEMS BEFORE CONNECTNG YOUR CRYSTAL SET TO THEM...

Queer news and culture.

For everyone.

INTO MORE.COM

charris8 (/member/charris8/) 2015-10-15 Reply

Do you really need it grounding? What are the chances of an electrical surge from a non powered device?

mk484 (/member/mk484/) ▸ charris8 (/member/charris8/) 2015-10-15 Reply

Hi - a ground wire provides a return path so that the AM radio signal can be fully received by the set. If you leave the ground wire off, you may not get very good volume in the earphone, or worst still, you may not get a signal at all.

You need to install a safe ground wire - one that goes straight to a metal pole, embedded in soft, wet soil. Make sure there are no pipes or cables underneath the ground where you are going to do this. If you have any doubts, consult a qualified electrician who can advise you on how to do that safely.

As for electrical shocks, they are possible during thunderstorms, due to potential lightening strikes, so it's best to disconnect your crystal set from both the antenna and ground wires for the duration of the storm. Quite often you can hear the 'crackle' of distant thunder/lightening in the distance, long before you can see/hear them.

More Comments

by mk484 (/member/mk484/) in science (/technology/science/)

Download

(/id/Nir-Nax-Tie-Taco-Crystal-Radio-Set-C-Austin-Hellier/)

8 Steps

Collection

I Made it!

Favorite

Share

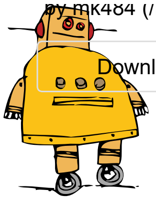

Newsletter

Let your inbox help you discover our best projects, classes, and contests. Instructables will help you learn how to make anything!

enter email I'm in!

About Us

Who We Are (/about/)

Advertise (/advertise/)

Contact (/about/contact.jsp)

Jobs (/community/Positions-available-at-Instructables/)

Help (/id/how-to-write-a-great-instructable/)

Find Us

- Facebook (http://www.facebook.com/instructables)
- Youtube (http://www.youtube.com/user/instructablestv)
- Twitter (http://www.twitter.com/instructables)
- Pinterest (http://www.pinterest.com/instructables)
- Google+ (https://plus.google.com/+instructables)

Resources

For Teachers (/teachers/)

Residency Program (/pier9residency)

Gift Premium Account (/account/give?sourcea=footer)

Forums (/community/?categoryGroup=all&category=all)

Answers (/tag/type-question/?sort=RECENT)

Sitemap (/sitemap/)

advertisement

Queer news and culture.

For everyone.

INTO MORE.COM

Terms of Service (http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=21959721) |
Privacy Statement (http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=21292079) |
Legal Notices & Trademarks (http://usa.autodesk.com/legal-notices-trademarks/) | Mobile Site (https://www.instructables.com)
(http://usa.autodesk.com/adsk/servlet/pc/index?id=20781545&siteID=123112)
© 2017 Autodesk, Inc.

(No Model.)

2 Sheets—Sheet 1.

C. LEAVITT.

MECHANISM FOR UTILIZING WAVE POWER.

No. 321,229. Patented June 30, 1885.

FIG. 2.

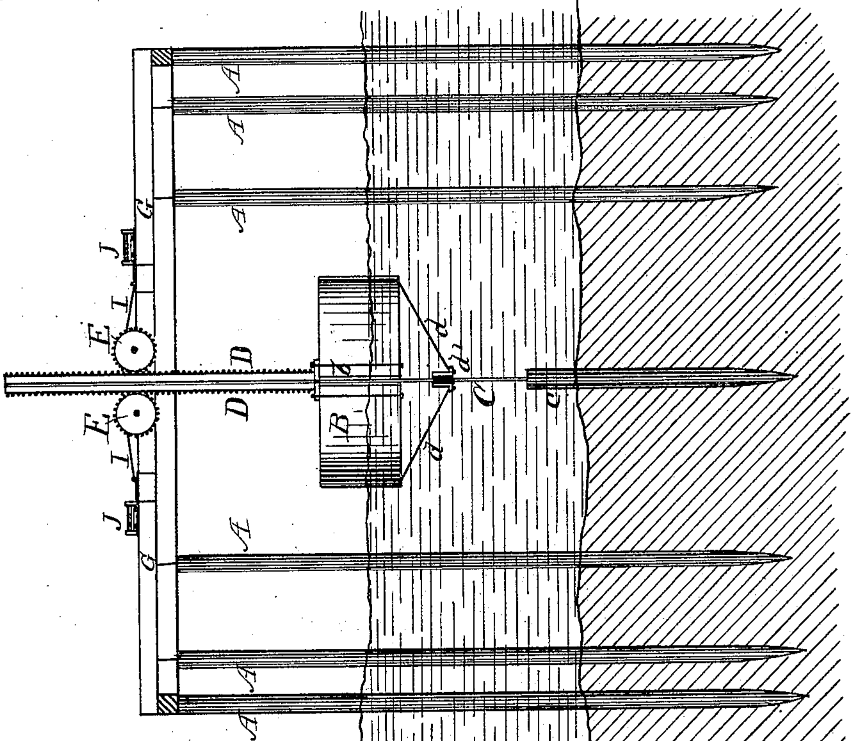
A
A
A
G
J
I
E
D
D
B
d
d
d^1
C
c
d
E
I
J
G
A
A
A

FIG. 1.

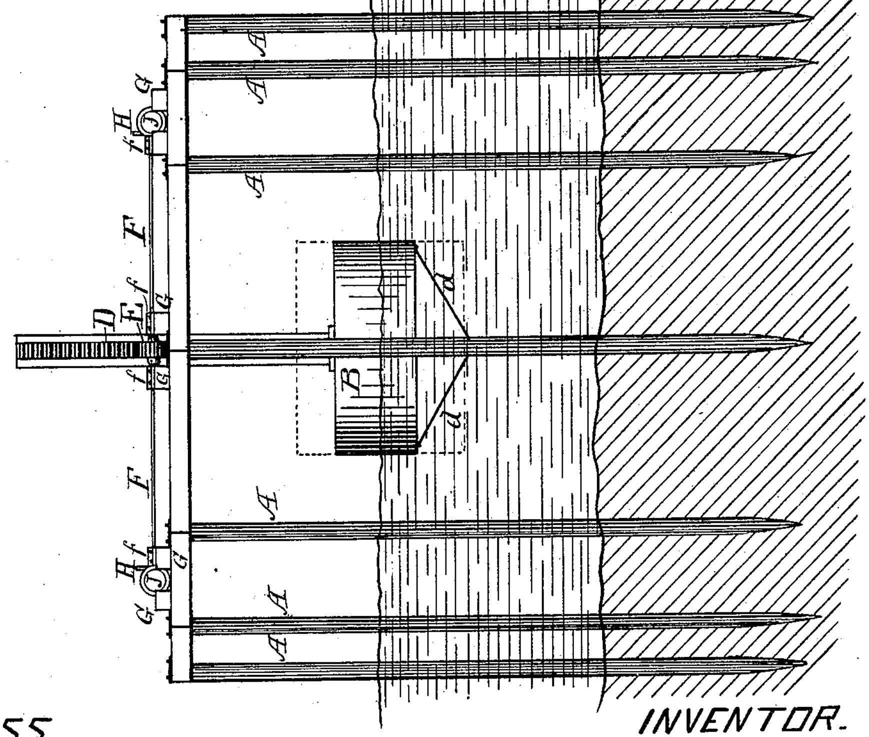
A
A
G
H
f
J
G
F
E
f
D
f
G
B
d
d
F
A
A
H
f
J
G
A
A

WITNESS
E. W. Laird
F. R. Tibbitts

INVENTOR.
Charles Leavitt,
per Geo. W. Tibbitts atty

N. PETERS, Photo-Lithographer, Washington, D. C.

(No Model.) 2 Sheets—Sheet 2.

C. LEAVITT.

MECHANISM FOR UTILIZING WAVE POWER.

No. 321,229. Patented June 30, 1885.

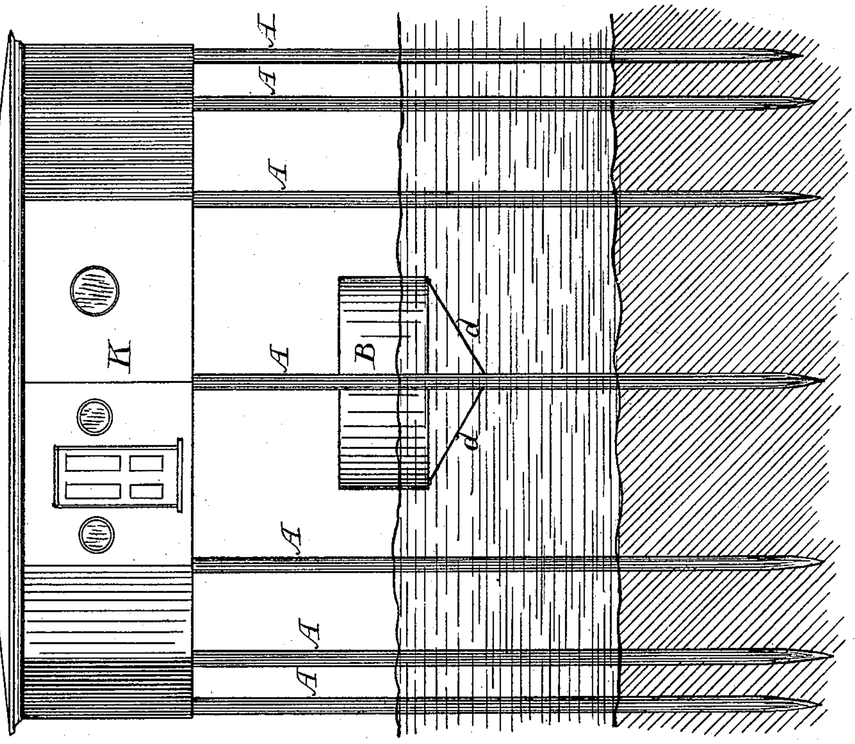
A
A
A
A
A
A
A
K
B
d
d

FIG. 4.

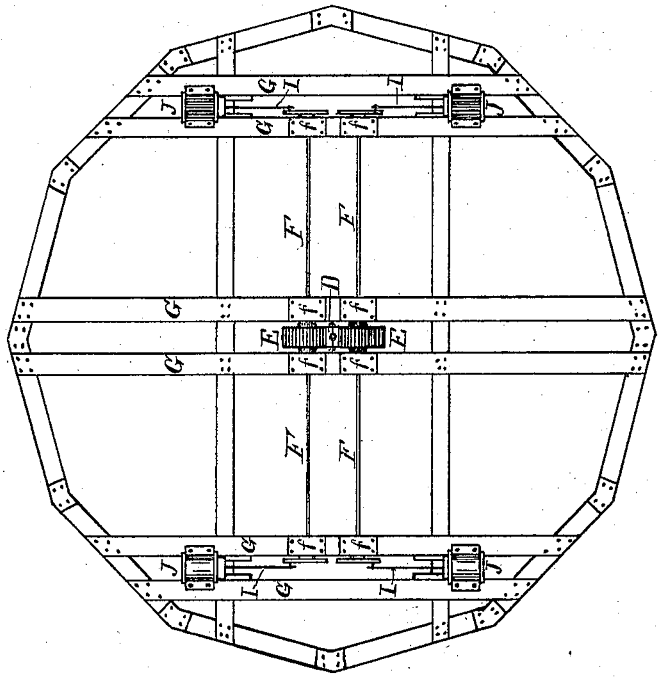
J
G
I
G
f
f
I
J
F
F
D
G
E
E
f
f
f
f
G
F
F
J
G
I
G
f
f
I
J

FIG. 3.

WITNESS.

Ew. Laird
F. R. Tibbitts.

INVENTOR.

Charles Leavitts.
per Geo. W. Tibbitts Atty.

N. PETERS, Photo-Lithographer, Washington, D. C.

UNITED STATES PATENT OFFICE.

CHARLES LEAVITT, OF CLEVELAND, OHIO.

MECHANISM FOR UTILIZING WAVE-POWER.

SPECIFICATION forming part of Letters Patent No. 321,229, dated June 30, 1885.

Application filed February 27, 1885. (No model.)

To all whom it may concern:

Be it known that I, CHARLES LEAVITT, of Cleveland, in the county of Cuyahoga and State of Ohio, have invented certain new and useful Improvements in Mechanism for Utilizing Wave-Power, of which the following is a specification.

This invention has for its object to employ the force of the waves on large bodies of water; and it consists of a float placed on the surface of the water, and inclosed within a suitable piling for protection to said float, and for supporting a superstructure for the employment of the power generated by the rise and fall of said float. The float is made to ride on a perpendicular rod, and is provided with a rack or racks, which mesh with and operate pinions or gears, from which power may be transmitted for pumping water or for air-compressing purposes.

In the accompanying drawings, Figure 1 is a side elevation of my mechanism, showing the float surrounded by piling, which is represented as standing in the water and the earth under the water. Fig. 2 is a transverse section of the same. Fig. 3 is a plan view of the frame-work and floor supporting the machinery operated by the aforesaid float. Fig. 4 is a side elevation of the mechanism, like Fig. 1, showing a house supported on the said piling and covering the aforesaid machinery.

A A represent piling driven into the ground, under water of suitable depth, in circular form, and in sufficient numbers to form a convenient inclosure capable of withstanding the vehemence of great storms, and to provide a protection to the operating parts of the mechanism.

B is a float or buoy, made preferably of iron and in circular form. Said float is maintained in a central position within the said inclosure by means of a strong perpendicular rod or bar, C, secured near the bottom of the water by being anchored in a pile, *c*, firmly driven into the bed of the lake. Said rod extends upward through the floor and forms a guide for the movements of the float, which is provided with a central tube, *b*, through which the said rod passes. To the under side of said float are attached braces *d d*, secured to a sleeve, *d'*, on the rod C, which rides thereon, together with the float, and which serve to prevent any tilting of the float and binding it, so as to impede its vertical movements.

D D are rack-bars secured to the upper side of the float B, which extend upward through the floor and mesh with and operate gears E E, attached to shafts F F, having their bearings in boxes *f f*, fixed on timbers G G, lying across the floor. Said shafts carry cranks H H, to which the pitmen I I of pumping-cylinder J J are attached. The rise and fall of the float operates through the medium of the racks to rotate the pinions or gears in both directions. The pitmen operate the same upon the pumps, however, whether the shafts rotate to the right or left.

K represents a house built upon the said floor, for covering the machinery and providing a suitable protection for same and the workmen or attendant, and is preferably made round, so as to resist the force of great wind-storms.

From the foregoing it will be seen that the rise and fall of the float gives motion through the medium of the racks to the gears, thereby operating the pumps; and it will also be seen that said pumps work at short strokes or partial strokes as well as at full stroke, so that if the water is at all agitated to give rise and fall to the float, the pumps will be operated.

From the pump-cylinders are attached pipes, which may be united to one common main pipe, and leading to storage-tanks located in suitable position and at convenient distance from said pumping mechanism. In the instance of pumping water, force-pumps with air-chambers are to be used, and when the purpose is for air-compressing suitable air-pumps may be employed.

The great objects of my invention are to utilize the forces of the waves for operating pumps or other machinery, whereby great economy is the result, enabling the accumulation of water in reservoirs or the storage of compressed air to be accomplished at greatly-reduced expenditure of time, labor, and money.

I am aware that a float having racks attached for operating the mechanisms for pumping and the like have been used, the float, rising and falling on the waves, actuating such mechanism. I do not therefore claim, broadly, such mechanism, but the combination, with a float provided with racks, of the mechanism substantially as described in my specification and shown in my drawings.

Having described my invention, I claim—

In a mechanism for utilizing wave-power, the combination of the following elements, viz: the piling A, float or buoy B, riding on the upright rod C, supported centrally within the inclosure of piling, and provided with the racks D D, operating pinions or gears E E, attached to shafts F F, supported in boxes *f f*, resting on the timbers and frame-work G, and having cranks H, connected by pitmen I to pumping-cylinders J, all constructed and arranged to operate substantially as and for the purposes described.

CHARLES LEAVITT.

Witnesses:
E. W. LAIRD,
GEO. W. TIBBITTS.

US005167786A

United States Patent [19]

Eberle

[11] **Patent Number: 5,167,786**

[45] **Date of Patent: Dec. 1, 1992**

[54] **WAVE-POWER COLLECTION APPARATUS**

[76] Inventor: **William J. Eberle**, 102 Oak Forrest
Trail, Euless, Tex. 76039

[21] Appl. No.: **646,353**

[22] Filed: **Jan. 25, 1991**

[51] **Int. Cl.**^{5} **C25B 1/10**; C25B 1/12; C25B 9/00; C25B 15/02

[52] **U.S. Cl.** **204/229**; 204/242; 204/264; 204/266

[58] **Field of Search** 204/129, 275-278, 204/263-266, 229, 242; 210/242.1, 170

[56] **References Cited**

U.S. PATENT DOCUMENTS

242,233	5/1881	Swales .	
321,229	6/1885	Leavitt .	
366,768	6/1887	Elias .	
581,067	4/1897	Fletcher .	
706,620	8/1902	Williams .	
739,538	9/1903	Fredson .	
787,182	4/1905	Huchings .	
791,366	5/1905	Rapp .	
852,232	4/1907	Kohler .	
879,992	2/1908	Wilson .	
884,080	4/1908	Fallis .	
975,157	11/1910	Quedens .	
1,377,163	5/1921	Pool	60/497
1,389,445	8/1921	Hare .	
1,403,702	1/1922	Melvin .	
1,471,222	10/1923	Taylor .	
1,623,341	4/1927	Hare .	
1,647,025	10/1927	Stich .	
2,668,918	2/1954	Howell .	
2,706,077	4/1955	Searcy	230/67
2,749,085	6/1956	Searcy	253/10
3,060,119	10/1962	Carpenter	210/242.1 X
3,487,228	12/1969	Kriegel et al.	290/52
3,652,431	3/1972	Reynolds	204/129
4,125,463	11/1978	Chenoweth	210/170 X
4,206,601	6/1980	Eberle	60/398
4,235,693	11/1980	Rowe et al.	204/129 X
4,335,576	6/1982	Hopfe	210/242.1 X
4,452,696	6/1984	Lopez	210/242.1 X
4,490,232	12/1984	Lapeyre	204/278

Primary Examiner—Donald R. Valentine
Attorney, Agent, or Firm—Woodcock Washburn Kurtz Mackiewicz & Norris

[57] **ABSTRACT**

A novel tidal and wave-power collection apparatus is disclosed which is adapted to produce hydrogen and/or oxygen gas. The apparatus is particularly well suited for oceanic installations. In the preferred embodiment, a matrix of gas producing units is easily assembled at sea, each being bridged and reinforced by catwalks extending therebetween which form trilateral gas producing modules having such units at each of the corners thereof. Each gas producing unit comprises a central tubular pilaster embedded in the ocean floor around which is disposed a toroidal float which moves up and down with the waves and tide and which acts to drive a DC generator while rising and falling. The output of the generator is utilized to electrolytically produce hydrogen and/or oxygen gas, which is compressed for storage in the float's chamber. The apparatus is efficient, economical and capable of capturing and storing large quantities of gas. In an alternative embodiment, the DC generator drives a high pressure pump which forces sea water through the microporous membrane of a reverse osmosis unit, thereby converting the sea water into potable water.

18 Claims, 8 Drawing Sheets

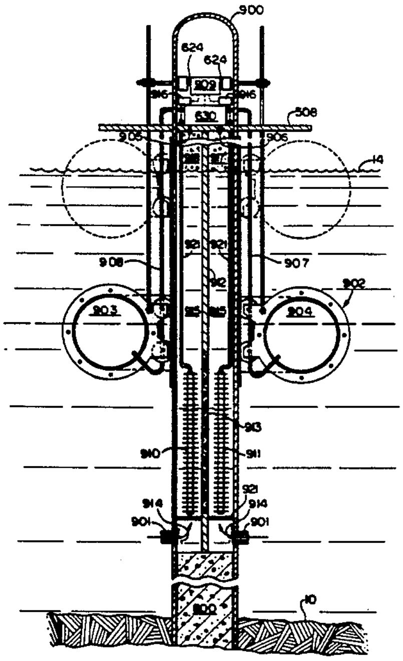

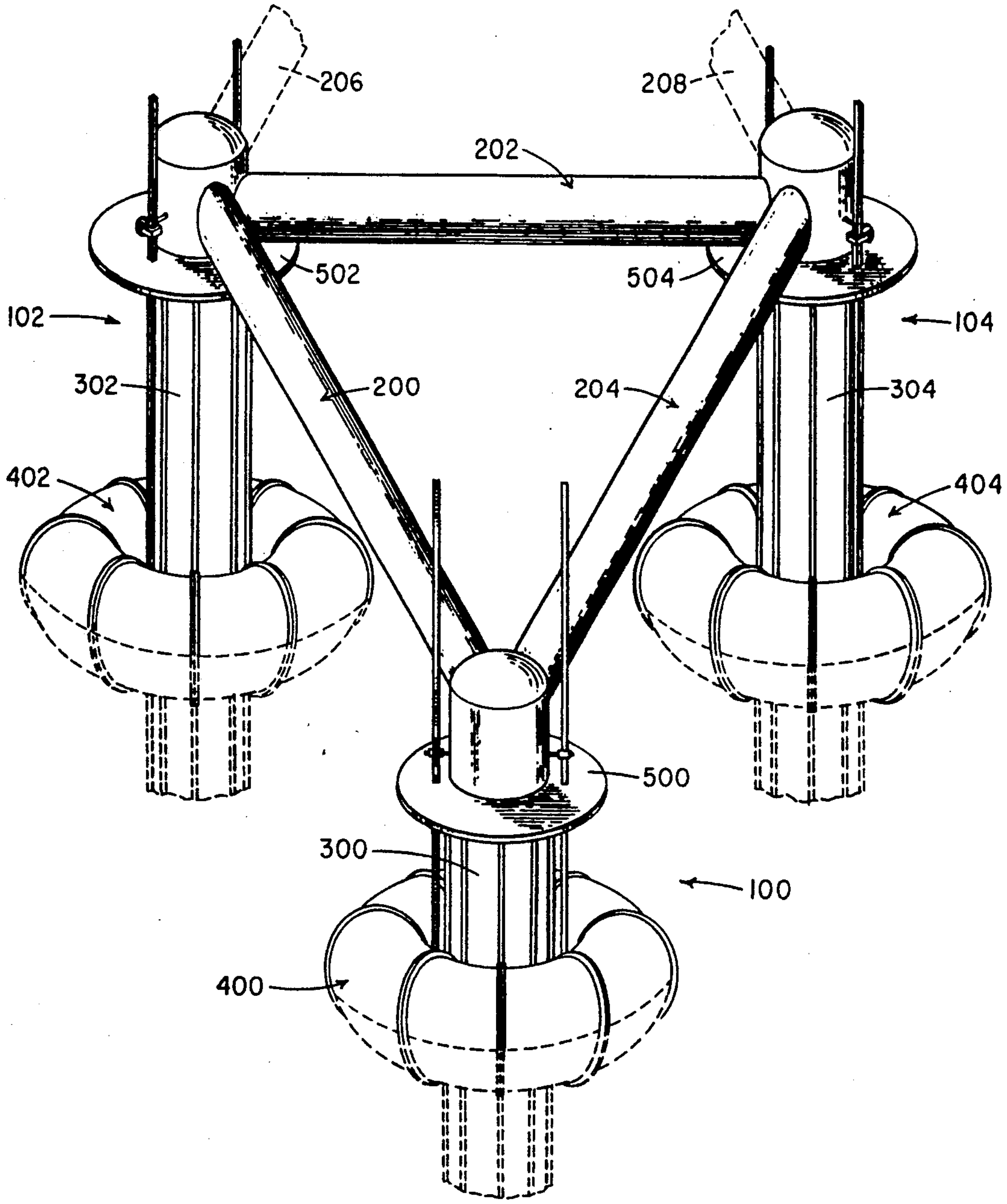

FIG. 1

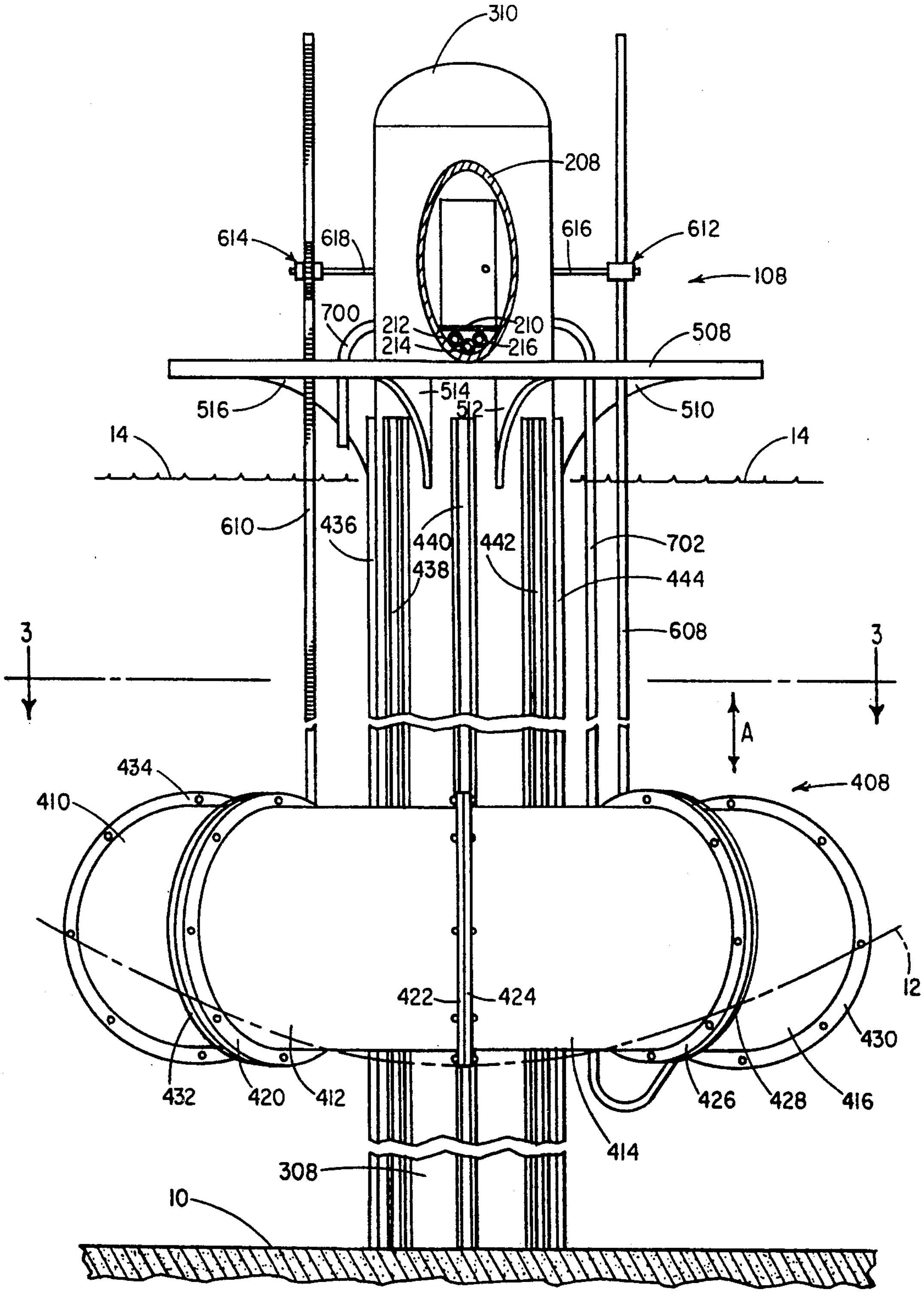
A

FIG. 2

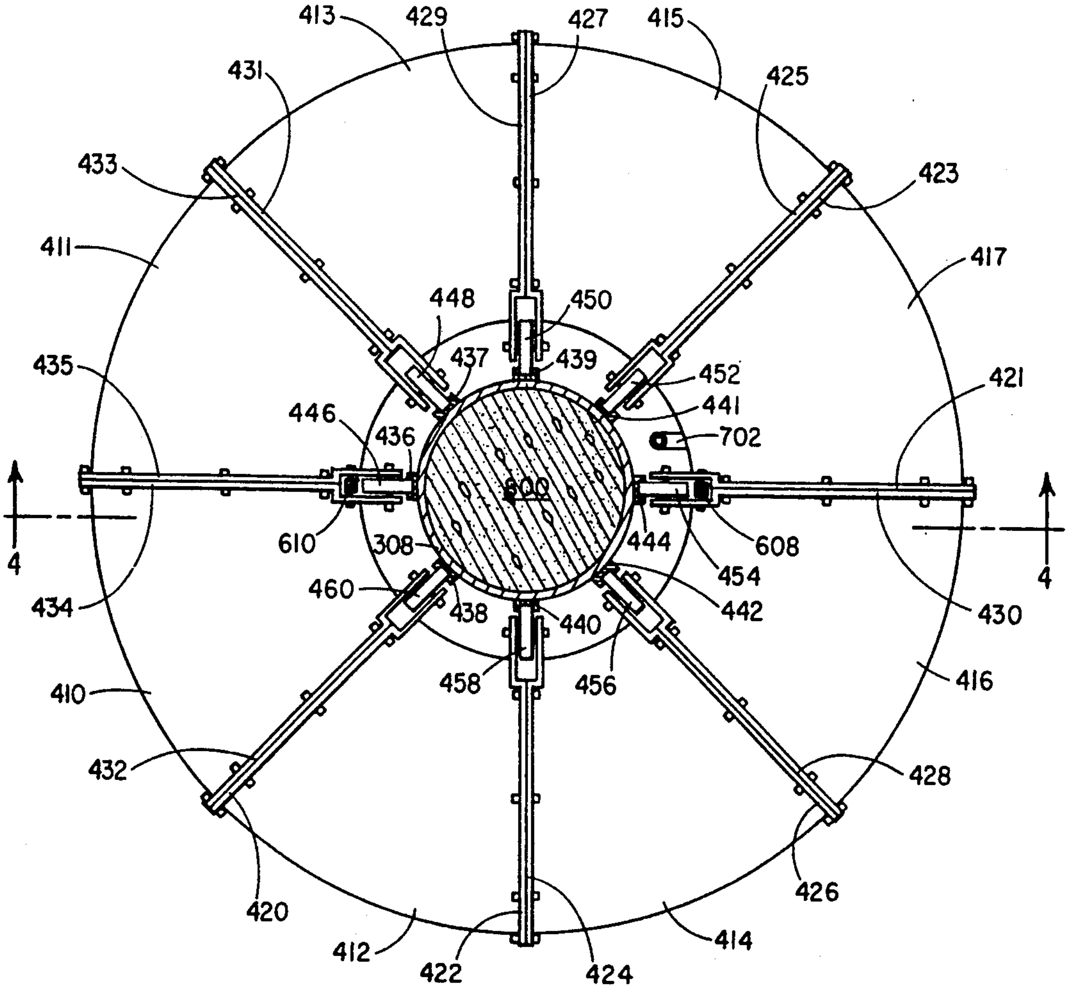

FIG.3

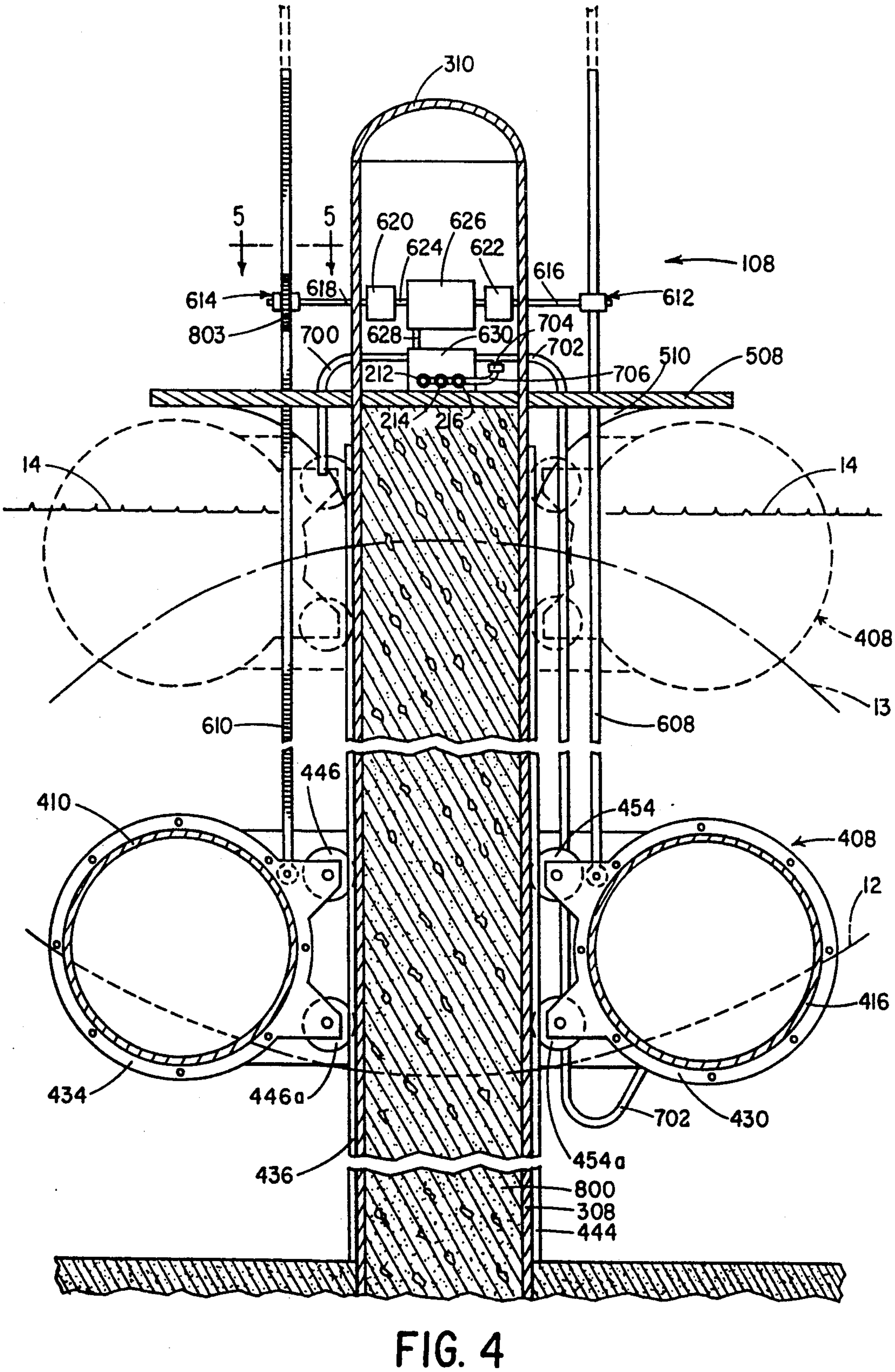

FIG. 4

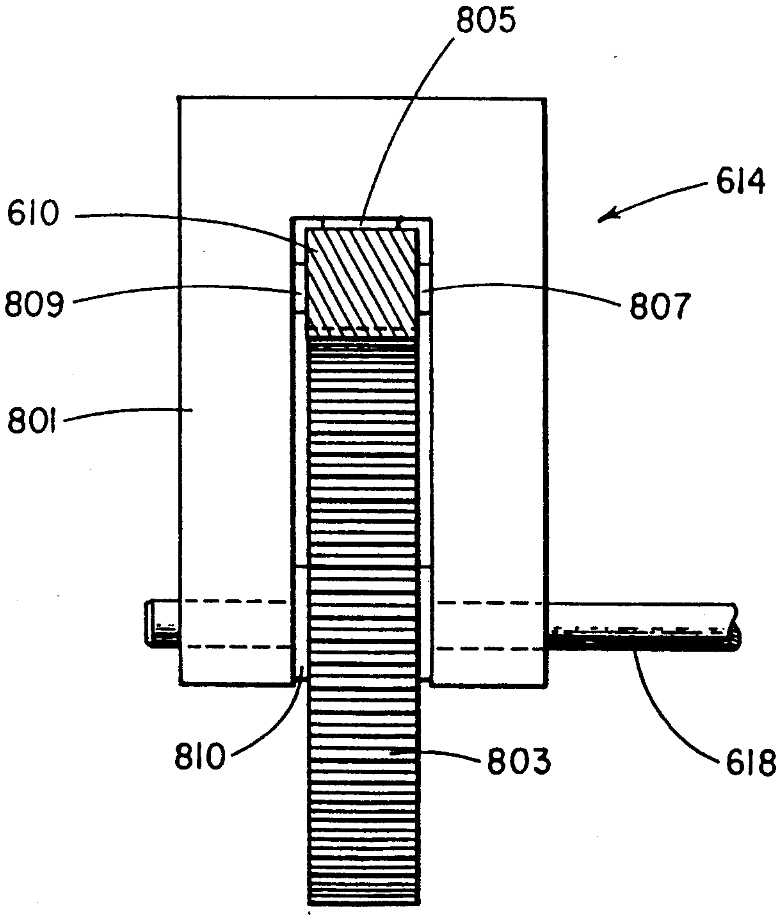
805
614
610
809
807
801
810
803
618

FIG. 5

FIG. 6

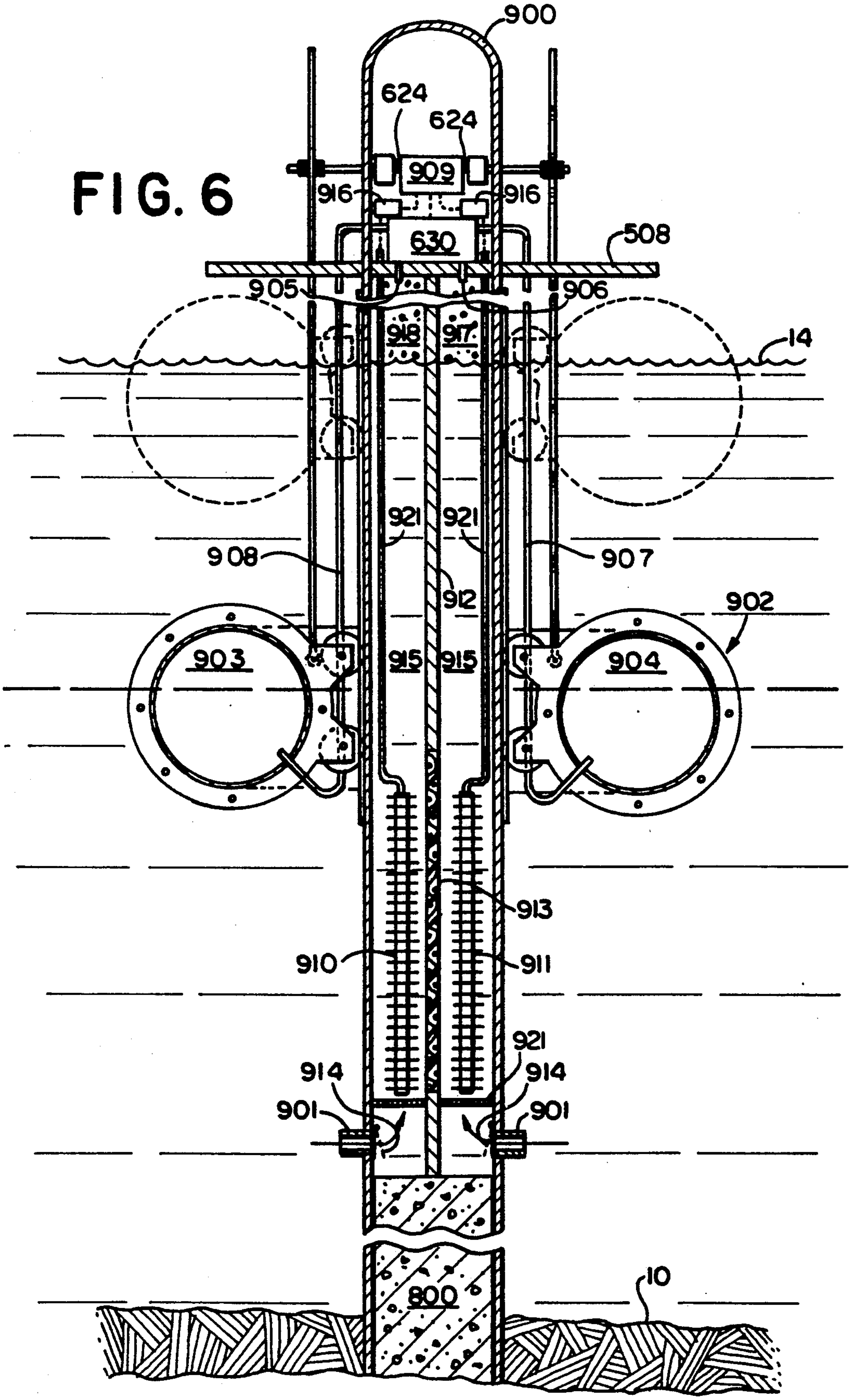

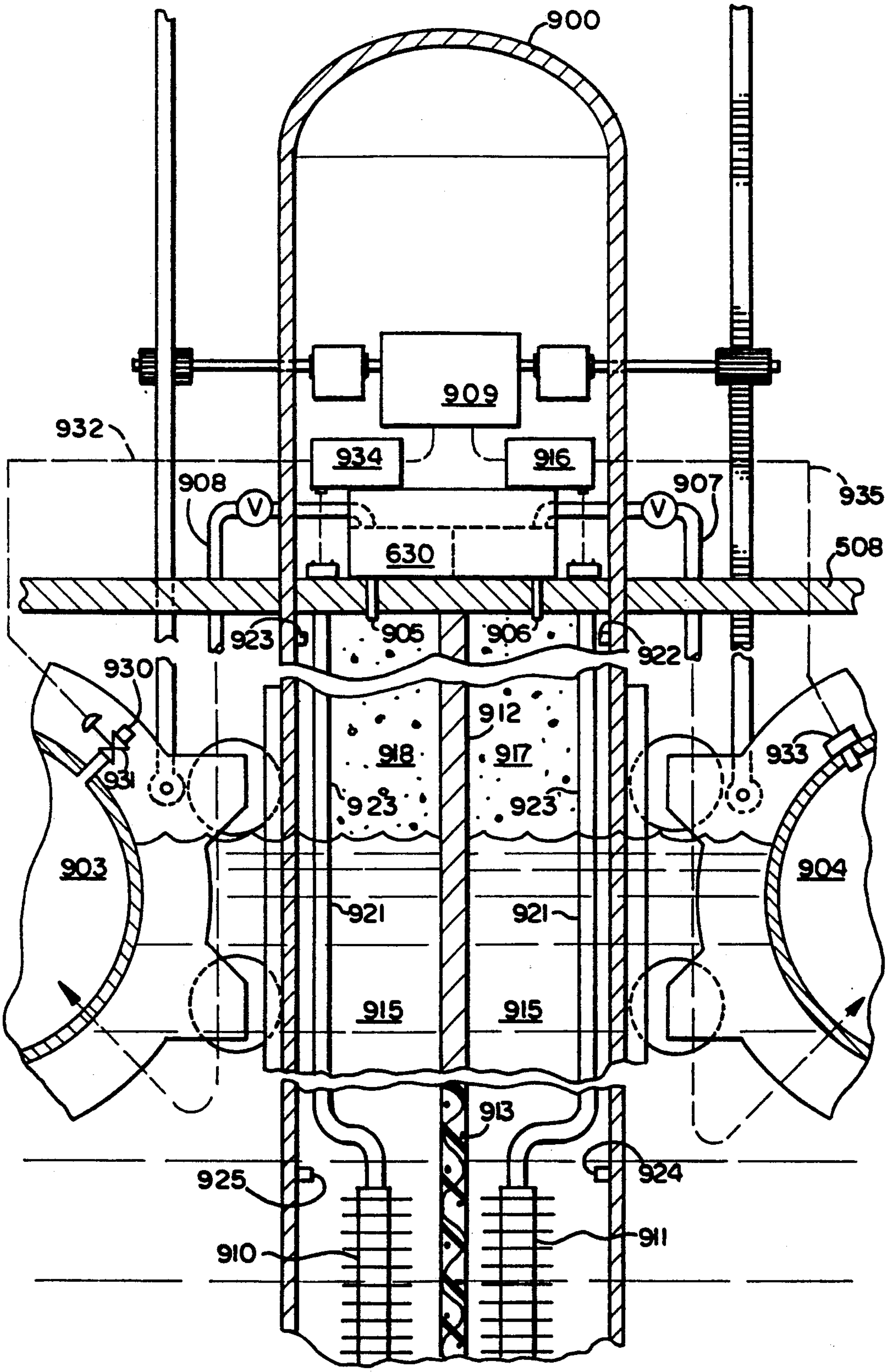
V

FIG. 7

FIG. 8

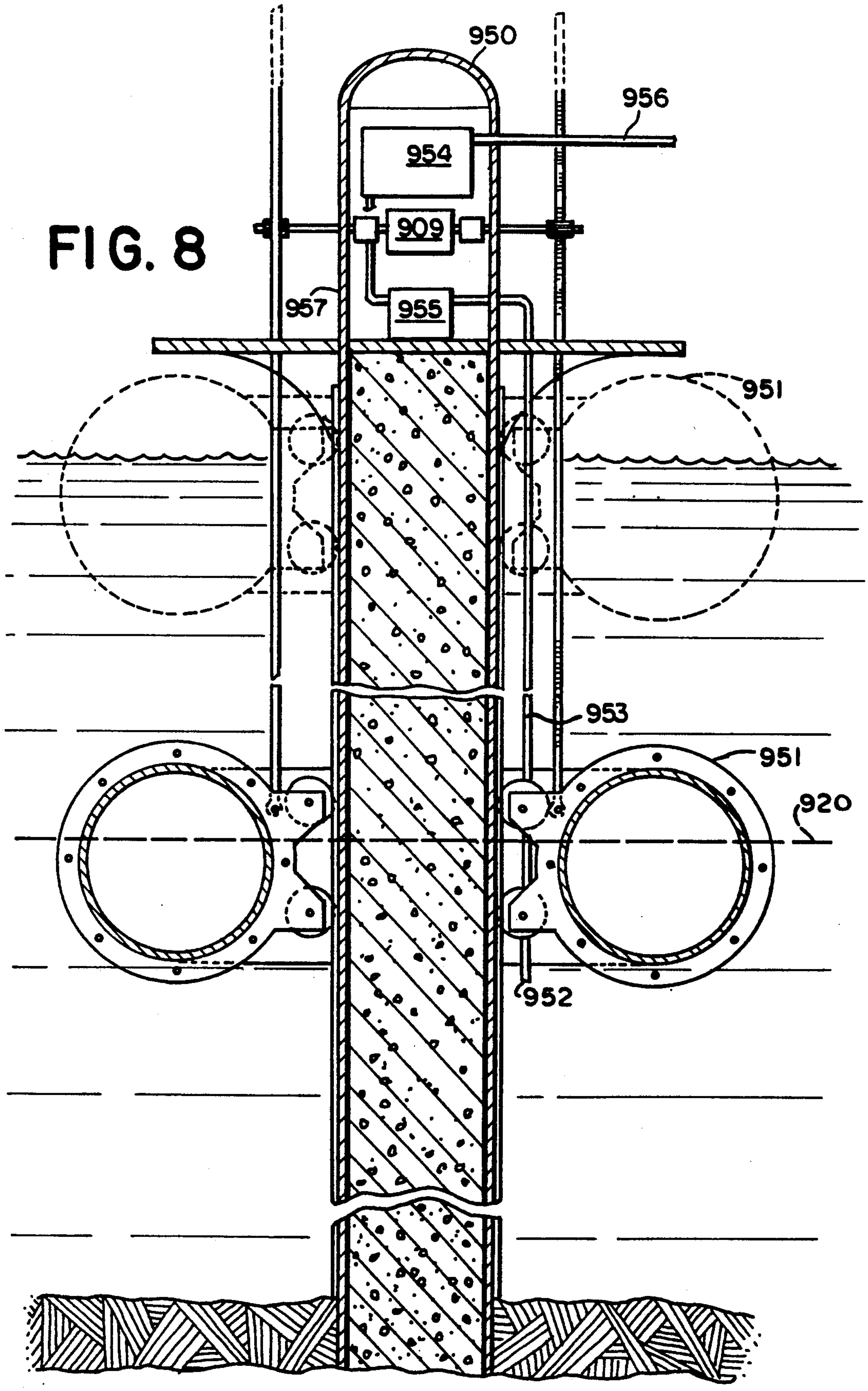
950
956
954
909
957
955
951
953
951
920
952

WAVE-POWER COLLECTION APPARATUS

CROSS-REFERENCE TO RELATED PATENT

The present application is related to my prior U.S. Pat. No. 4,206,601, dated Jun. 10, 1980, which patent is hereby incorporated by reference as if fully set forth herein.

BACKGROUND OF THE INVENTION

It has long been recognized that waves and/or tidal forces contain tremendous amounts of potential energy which, if harnessed, could not only supply substantial power but might also, if appropriately installed, protect large areas of shore line from the ravages of oceanic erosion.

In the prior art, extensive patent literature exists dealing with the subject of "wave motors" and relating to various ways of utilizing floats which rise and fall with passing waves, the movement of the floats being translated by various means into mechanical energy, and then into electrical energy. For example, U.S. Pat. Nos. 321,229 and 1,647,025 disclose compressed air storage in combination with a wave motor. U.S. Pat. No. 2,706,077 is titled "Ocean Wave Air Compressor". Other forms of energy storing means are known to this art, including means utilizing balance wheels or flywheels in combination with various mechanisms for supplying power from waves. Additionally, the art teaches the use of multiple pontoons, or floats, the floats being arranged in various ways for improving the continuity of power derived from the passing waves, while occasionally also showing means for deriving energy from the lateral motion of waves as well as from the up and down motion of waves. In these regards, please refer to U.S. Pat. Nos. 242,233; 321,229; 366,768; 581,067; 706,620; 739,538; 787,182; 791,366; 852,232; 884,080; 879,992; 975,157; 1,377,163; 1,389,445; 1,403,702; 1,471,222; 1,623,341; 1,647,025; 2,668,918; 2,706,077; 2,749,085 and 3,487,228.

While many of the devices disclosed in the prior art are no doubt operative, the failure of such devices to gain widespread acceptance suggests that these devices may, in practice, be too expensive, complex, and difficult to maintain to serve as practical energy sources.

SUMMARY OF THE INVENTION

The present invention overcomes many of the disadvantages of the prior art by providing an extremely simple, economical device for capturing substantial amounts of energy and converting that energy into hydrogen and/or compressed oxygen which is stored within the floats of each power unit. These gas(es) may then be collected, or supplied through conduits, as needed, to satisfy energy and or raw material demands.

Applicant has recognized that a drawback of prior art wave power collection devices is the inability to collect and store large amounts of energy. The present invention takes advantage of the ocean's plentiful supply of highly conductive sea water to serve as a reservoir for the electrolytic production of hydrogen, and if desired, oxygen, gas. Production of hydrogen gas in large quantities is particularly timely, since that gas is currently in high demand as an alternate, non-polluting fuel source for automobiles.

The preferred embodiment device of the present invention comprises a plurality of tubular members which have been designed to facilitate on site "at sea" assembly Each of the power collection units comprises a central, substantially cylindrical pilaster around which a toroidal (donut-shaped) float is trackingly mounted. One end of the cylindrical pilaster is embedded in the ocean floor and the pilaster extends vertically through the water to protrude some distance above the highest expected high water line. Each pilaster is interconnected to at least two adjacent pilasters by tubular bridging members which serve to reinforce the pilasters, to house high pressure conduits, and to act as pedestrian catwalks to facilitate servicing of each energy collection unit. In the preferred embodiment, the floats are manufactured in accurate tubular sections which have flanges formed thereon which, upon assembly, create high pressure seals between sections to define a gas chamber within the float. Portions of these flanges additionally act in combination with axles mounted therebetween to position wheels which track up and down the outside surface of the pilaster. Portions of these flanges also act as attachment sites for several rack and pinion drives which power a drive shaft in response to the ascension and descension of the float with respect to the pilaster. The drive shaft drives a DC generator which is connected to electrodes located in a hollow well formed in the core of the tubular support member. This well, which is in liquid communication with the surrounding ocean, is partially filled with sea water which totally submerse the electrodes. Baffles and appropriate separators separate the electrodes so that the gases generated in the vicinity of each may be separately collected, or separately vented to the atmosphere, as desired. A compressor, driven either manually by the drive shaft or electrically by the output of the generator, compresses at least hydrogen gas drawn from the appropriate, hydrogen containing chamber of the well headspace, for transmission and storage in the toroidal floats.

In an alternative embodiment of the invention, the DC generator drives a high pressure pump which forces sea water through the microporous membrane of a reverse osmosis unit, thereby converting the sea water into potable water.

As seen from the above, a primary object of the present invention is the provision of a superior tidal and wave-power collection apparatus which is easily assembled and installed at sea.

Another object of the present invention is the provision of such an apparatus which may produce, store, and supply substantial amounts of hydrogen and/or oxygen gas(es) which has been captured by electrolytic conversion of tidal and wave energy.

Still another object of the present invention is the provision of such an apparatus which may produce potable water from sea water using the tidal and wave energy collected.

These and other objects of the present invention will become apparent from the following more detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is an isometric view of a single power collection module having three preferred embodiment power collection units disposed at the corners thereof, connections to other unit(s) and portions of each unit disposed below the water line being represented in phantom;

FIG. 2 is a front view on an enlarged scale of a power collection unit similar to that illustrated in FIG. 1, the height of which has been foreshortened for purposes of illustration, and portions of an elliptical tubular bridging member and high pressure transmission conduits being shown in cross-section;

FIG. 3 is a cross-sectional top view of the embodiment illustrated in FIG. 2 taken as indicated by the lines and arrows 3—3 in FIG. 2;

FIG. 4 is a cross-section of the embodiment shown in FIG. 2 taken as indicated by the lines and arrows 4—4 in FIG. 3;

FIG. 5 is a cross-sectional top view on a greatly enlarged scale of a portion of the embodiment shown in FIG. 4 taken as indicated by the lines and arrows 5—5, in FIG. 4;

FIG. 6 is a cross-section showing a first alternate embodiment of the invention shown in FIG. 4;

FIG. 7 is an enlarged view of the upper portion of the cross-section shown in FIG. 6;

FIG. 8 is a cross-section showing a second alternate embodiment of the invention shown in FIG. 4.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Although specific forms of the invention have been selected for illustration in the drawings, and the following description is drawn in specific terms for the purpose of describing these forms of the invention, this description is not intended to limit the scope of the invention which is defined in the appended claims.

Referring now to the figures, and particularly FIG. **1**, the preferred embodiment wave power collection apparatus should be installed in a relatively large body of water for converting wave and tidal energy into high pressure compressed air. The preferred embodiment apparatus further is capable of collecting and storing air in the apparatus, and for transmitting the compressed air to a preselected consumption site. The preferred embodiment apparatus generally comprises a plurality of power collection units which are structurally interconnected by bridging means for rigidly connecting the power collection means to create a rigid connection module. In the preferred embodiment, at least three power collection units are interconnected to form a trilateral power collection module which is capable of withstanding any lateral forces which might be applied there against by ocean currents or winds In FIG. **1**, power collection units designated generally **100**, **102**, and **104** are illustrated which are interconnected by bridging members designated generally **200**, **202**, and **204**. Each power collection unit comprises a central pilaster, such as pilasters **300**, **302**, and **304** which are substantially cylindrical structures which are driven into the sea bed or floor and which extend as shown in FIG. **1** for a preselected distance above the expected high water mark for a given location. Floats designated generally **400**, **402**, and **404** are trackingly mounted for substantially vertical movement up and down their respective pilasters, such that the movement of these floats will drive rack and pinion mechanisms associated therewith to drive a fly wheel and ultimately a compressor which are disposed in the hollow upper portion of the pilaster. Compressed air produced thereby is fed into the floats **400**, **402**, and **404** for storage, and/or is fed into conduits which are housed in the bridging members **200**, **202**, and **204** for eventual transmission from unit to unit and then to a remote power consumption site.

FIG. **1** illustrates the minimum number of power collection units which may be used in a single installation. In phantom in FIG. **1**, however, connections to an additional power collection unit as illustrated by the phantom bridging members **206** and **208** which may be used to interconnect this additional unit to power collection units **100** and **102** to expand the apparatus and to create therewith a second trilateral collection module. Accordingly, as power requirements increase, it is a relatively simple matter to expand the system to accommodate the needs of any particular power consumption site.

As shown particularly in FIG. **1**, the bridging means, such as bridging members **200**, **202**, and **204**, are substantially tubular and have elliptical cross-sections. In the preferred embodiment, these members are sized to act as catwalks which facilitate pedestrian movement between one power collection unit and the next. By orienting the axis of the elliptical cross-section substantially vertically, the bridging members are not only configured for easy pedestrian movement therethrough, but are also configured to permit the power collection units to be disposed at substantial distances apart, yet be spanned by relatively simple bridging structures. Finally, due to the enclosed nature of the catwalks maintenance workers servicing the power modules are well protected from the environment.

In order to limit the maximum vertical travel of each of the floats in an upward direction and also to provide a deck for use in servicing various portions of the apparatus, a substantially planar, annular deck is provided such as decks **500**, **502**, and **504**. During assembly these decks may act to temporarily support the bridging members prior to their permanent attachment to the pilasters.

The float **408** is seen in FIG. **2** to encircle the pilaster and to be trackingly mounted with respect thereto for movement in response to variations in water level, that is, so that the float may travel in directions which are substantially parallel to the axis of the pilaster. The float itself comprises a plurality of float sections, such as float sections **410**–**417**. Each of these sections is preferably identical to the other. At either end of each section are outwardly depending flanges, such as outwardly depending flanges **420** and **422** for section **412**; **424** and **426** for section **414**; **428** and **430** for section **416**; and **432** and **434** for section **410**. As shown particularly in FIGS. **2** and **3**, these outwardly depending flanges are complementally configured to bolt to the outwardly extending flanges of the adjacent float sections to thereby form a high pressure seal therebetween, and to create a compression chamber within the float which is not only water tight so that the float will not become swamped with water which might otherwise seep between the flanges, but also gas-tight to contain compressed air within the float so that the float may act as a compressed air reservoir. In order to stabilize and guide the movement of the float **408** a plurality of tracks, such as tracks **436**–**442** and **444** are applied to the exterior surface of pilaster **308**, one for each intersection between adjacent float sections, which tracks receive and guide wheels (two for each track) generally disposed between the float **408** and the pilaster **308**, as illustrated particularly in FIG. **4**. In FIG. **2** the water level, indicated by line **12**, is shown buoying the float **408** in a relatively low position with respect to the pilaster and other equipment mounted thereon. Attached to various places on the float are a plurality of racks, such as racks **608** and **610** which are pivotally connected at one end to portions of the float **408**. These racks extend through openings in the deck **508** and engage pinion assemblies designated generally **612** and **614** which are oscillated in opposite rotational directions in response to the up and down movement of the racks **608** and **610**. This rack movement is produced by movement of float **408** in the directions illustrated by double ended arrow A. The pinion assemblies **612** and **614** are mounted on remote ends of clutched drive shafts **616** and **618**.

A fluid intake conduit for supplying fluids to the air compressor located within the upper portion of the pilaster extends through an opening in deck **508** downwardly to a point between the highest expected high water line **14** and the undersurface of the deck **508**. By orienting the fluid intake means in this position, the fluid intake will normally supply air to the compressor but will be under water in the event that the expected high water mark is exceeded, as for example under hurricane conditions. In this event, weather conditions may be so severe that protection of the apparatus from destruction by such forces may become a matter of primary importance. Accordingly, the fluid intake conduit **700** will, under these conditions, supply sea water to the compressor. Since the output of the compressor communicates through output conduit **702** with the interior of the float **408**, water introduced through the fluid intake conduit will be channeled through output conduit **702** into the interior of the float, causing the compressor to act as a liquid pump, thereby lowering the float in the water and lowering the center of gravity of the apparatus. Depending upon the ferocity of the storm, the float will eventually settle down the pilaster until caused to stop at an appropriate position below the water's surface, yet at a position which is readily retrievable once weather conditions have improved. Assuming the float is not completely swamped (in which case auxiliary bilge pumps may become necessary), by reversing the compressor (and by providing negative pressure valves in the interior of the float chamber) the compressor may automatically pump out the interior of a float, discharging the sea water through fluid intake means **700** to reestablish the float as a gas filled chamber.

FIG. **2** has been taken so that a portion of the bridging means **208** connecting it with an adjacent power collection unit is shown in cross-section, and will be seen to have a floor **210** under which are disposed a plurality of high pressure compressed air transmission conduits **212**, **214**, and **216**.

Referring now in particular to FIG. **3**, which is a cross-sectional top view of the apparatus shown in FIG. **2**, the interrelationship between the toroidal float **408** and pilaster **308** is clearly illustrated. In FIG. **3** toroid sections **410**, **411**, **412**, **413**, **414**, **415**, **416**, and **417** are illustrated as are outwardly depending flanges **420**–**435**. As seen particularly in FIG. **3**, these flanges mate with complemental flanges of adjacent sections for almost their entire length, however, the portions thereof which are disposed toward the toroid's center are differentiated into spaced apart flange portions which act as an axle mounting assemblies for wheels **446**, **448**, **450**, **452**, **454**, **456**, **458**, and **460** which move along tracks **436**, **437**, **439**, **441**, **444**, **442**, **440**, and **448** respectively. Ballast **800** is provided in the center of the pilaster in order to weight the entire unit and increase both its strength and ability to withstand the tendency of ocean currents etc. to displace the unit.

Referring now in particular to FIG. **4**, many of the features discussed above with reference to the other figures are also illustrated in FIG. **4**. Additionally, FIG. **4**, which is a cross-section of the power collection unit shown in FIGS. **2** and **3**, depicts the machinery disposed within the upper hollow portion of pilaster **308**, as well as additional wheel mounting and float details. In particular, clutched drive shafts **616** and **618** extend inwardly from pinion assemblies **612** and **614** to pass through openings in the pilasters to engage unidirectional clutches **620** and **622**. Since the tracks **608** and **610** are disposed on opposite sides of their respective pinion assemblies, both of these clutches may be either clockwise or counter-clockwise clutches so that one of the racks, such as rack **610** will act through pinion assembly **614**, drive shaft **618**, and clutch **620** to drive the main drive shaft **624** in a preferred rotational direction in response to the ascension of the float, while the other rack, rack **608**, will act through pinion assembly **612**, clutch drive shaft **616**, and clutch **622** to continue driving the main drive shaft in that direction during the descension of the float. The clutches **620** and **622** allow their respective clutch shafts to idle with respect to the drive shaft upon the return strokes of the racks. A flywheel **626** is diagrammatically illustrated disposed on main drive shaft **624**, and a belt or other drive means **628** is shown spanning between the flywheel and compressor **630** so that the compressor is continuously driven by the flywheel and main drive shaft even though the power supplied thereto by either of the racks is discontinuous.

The output of compressor **630** is connected to output conduit **702** and also through a ball check valve **704** to transmission supply conduit **706**. The ball check valve prevents any water which might enter the system from inadvertently being introduced into transmission conduits **212**, **214** or **216**, and additionally seals off those transmission conduits when the gas pressure within conduit **702** is less than the pressure within conduit **706**.

FIG. **4** additionally shows the relative vertical orientations of each wheel set, such as wheel set **446** and **446***a* and wheel set **454** and **454***a*, which wheel sets are spaced apart and vertically aligned to move within a single pilaster track. By mounting two vertically spaced apart wheels between opposing flange portions from each of the sections, any tendency of the float to twist with respect to the pilaster is effectively resisted.

FIG. **4** also shows the pivotal attachment of racks **608** and **610** between opposing portions of flanges **430** and **421** for rack **608** and **434** and **435** for rack **610**. An aperture formed at the float end of each rack is sized to receive a rod extending between these opposing flange portions. Since the rack aperture is somewhat oversized, some lateral play is accordingly created between the racks and the float.

In FIG. **4**, the float is shown in its solid lined position buoyed by water level **12** which represents a wave or tidal trough, whereas the float **408** is shown in dotted outline buoyed to a greater nearly maximum height with respect to the pilaster on a wave or tidal crest, as indicated by phantom water level height **13**. It should be noted in FIG. **4** that the height of the apparatus illustrated in FIG. **4** has been foreshortened in several places. While the deck **508** is designed to resist any forces which might be applied thereto by interference with the float **408**, as described hereinabove, the height of the pilaster should not be made so short that the travel of the float is unduly limited by the deck position, but rather under most weather conditions the float should be free to move between its crest and trough positions. At the same time, by designing the apparatus with reference to an expected high water mark and by providing the preferred embodiment apparatus with various means for withstanding occasional water heights which may be substantially higher than the maximum expected high water line, expenditures for unduly high pilasters and other mechanisms which might otherwise be used only very infrequently are avoided.

In FIG. **5**, which is a top cross-sectional view on a greatly enlarged scale, the details of the pinion assembly designated generally **614** is illustrated. The pinion assembly is seen to comprise a substantially U-shaped rack follower assembly sol, the ends of which are journalled on the end of clutched drive shaft **618**. A main drive gear **803** is attached to shaft **618** and is positioned between the two arms of the rack follower assembly. The rack **610**, which is shown in cross-section in FIG. **5**, is journalled between main drive gear **803** and bearings **805**, **807**, and **809** which are mounted on the inner surfaces of the rack follower assembly near the interior base of the "U". These bearings preferably are biased towards the rack yet provide enough forgiveness to prevent undue stress on the rack member. Similar bearings **810** are provided between the main drive gear **803** and the interior surfaces of the arms of the rack follower assembly **801** so that the movement of the rack **610** will ride smoothly so that axial movement of the rack **610** with respect to the main drive gear will produce a positive rotational driving of clutched shaft **618**. Although not illustrated in the drawings, according to one embodiment of the present invention, a unidirectional clutch may be incorporated between the main drive gear **803** and the shaft on which it is mounted. In this instance, the main drive shaft **624** would extend substantially through the pinion assemblies, with the unidirectional clutching action taking place within those assemblies. It is also within the scope of the present invention, if desired, to provide additional bearing means at various places along the clutched and main drive shafts to further stabilize the pinion, flywheel, and clutched assemblies.

In FIG. **6**, a first alternate embodiment of the invention is shown. In this embodiment a pilaster **900**, driven into the sea floor **10**, extends a preselected distance above the expected high water line **14**, as before. However, in this embodiment, ballast **800** is provided in only the lower portion of the pilaster **900** so that a hollow well **915** is formed within the pilaster above the ballast. A plurality of apertures **901** are disposed in the pilaster **900** below the low water line **920** so as to allow water to enter the well **915** through flapper type check valves **914**. A filter **921**, which may be a stainless steel strainer, removes debris from the water entering into the well **915** through apertures **901**. Two finned electrodes **910** and **911**, Which may be of the type commonly used in lead batteries, are disposed in the well **915** below the low water line **920**. The electrodes **910** and **911** are oppositely charged so that an electrical potential is created between them, causing current to flow through the water between the electrodes. As a result of this current, a portion of the water dissociates into hydrogen and oxygen molecules by an electrolysis process, the hydrogen molecules forming on the cathode electrode **910** and the oxygen molecules forming on the anode electrode **911**.

A microporous barrier **913**, which may be of the type commonly used as battery separators, is mounted in the pilaster **900** between the electrodes **910** and **911**. The microporous barrier **913** allows the water to pass freely across it but is impervious to gases. Thus, the microporous barrier **913** ensures that the hydrogen and oxygen molecules which form on the electrodes **910** and **911** remain separated.

After dissociation, the hydrogen and oxygen molecules form bubbles which rise through the water and collect in a headspace formed between the water level and the deck **508**. A baffle **912** is disposed above, and is contiguous with, the microporous barrier **913** and divides the headspace into two portions **917** and **918** so that the hydrogen and oxygen remain separate as they collect in the headspace **917**, **918**. From the headspace, the hydrogen and oxygen are drawn into a compressor **630** for compression therein. The compressor **630** is disposed on the deck **508** above the headspace and the hydrogen and oxygen enter the compressor **630** through intake conduits **906** and **905**, respectively, which extend through the deck. The compressed hydrogen and oxygen from the compressor **630** are directed, via conduits **907** and **908**, respectively, to the float **902** for storage. In the float **902**, the oxygen and hydrogen are stored in separate portions **903** and **904**, respectively, formed in the float by two vertically extending baffles (not shown) disposed within two of the torus sections.

Alternatively, if it were deemed undesirable or uneconomical to store both the hydrogen and the oxygen produced, the oxygen could be vented to atmosphere, rather than drawn into the compressor **630** as shown in FIG. **6**. In this case, the aforementioned baffles would be unnecessary and the entire volume within the float could be utilized for the storage of hydrogen.

In any event, after being stored in the floats, the hydrogen/oxygen gas is fed into conduits housed in the bridging members **200**, **202** and **204**, shown in FIG. **1**, for eventual transmission from unit to unit and then to a remote consumption site, as was done with the compressed air produced by the embodiment shown in FIG. **4**.

In the embodiment shown in FIG. **6**, the float **902** drives the drive shaft **624** through the rack and pinion arrangement, as previously discussed. Although in this embodiment the compressor **630** could be mechanically driven via the fly wheel, as before, in the preferred arrangement, the drive shaft **624** turns a D.C. generator **909** which provides electrical power to drive the compressor **630**. The generator **909** is also electrically connected to the electrodes **910** and **911** via bus bars **923** so as to create the aforementioned electrical potential.

With respect to the energy which can be produced by the generator **909**, as driven by the oscillations of the float **902** due to wave action, it should be noted that the energy flux in an incident wave can be approximated by the equation:

$$E_f=\rho g H^2 L\,[(1+2kh)/\sinh\,(2kh)]/16T \qquad (1)$$

where

E_f=energy flux per unit wave crest width
ρ=density of water
g=acceleration due to gravity
H=wave height
L=wave length
k=wavenumber, 2ρ/L

h=depth of water

T=wave period.

Based on dispersion, the wave length L can be determined from the equation:

$$\omega^2 = gk \tanh (kh) \qquad (2)$$

where ω=wave frequency, $2\pi/T$. By combining equations (1) and (2), the total power contained in an incident wave can be estimated for a given float outside diameter. For a water depth of 30 feet and a float outside diameter of 30 feet, the total power estimated for waves of various heights and periods, using the aforementioned analysis, is shown in Table 1.

TABLE 1

Wave Height, Ft.	Wave Period, Sec.	Wave Power, kW
3	6	54.8
3	10	74.4
3	20	85.8
6	6	219.2
6	10	297.6
6	20	343.2

It has been further estimated that the apparatus disclosed herein, taking into account the movement of the float induced by the incident wave field and the canceling effect of the waves radiated by the float itself, is capable of extracting approximately 50% of the incident wave power. Thus, these theoretical estimates suggest that ample energy will be produced in the generator **909** to compress the gas to a relatively high pressure.

It will be appreciated that to avoid damage to the electrodes **910** and **911**, they must remain submerged at all times. To this end, flapper type check valves **914** are disposed at the discharges of the apertures **901**. The check valves are oriented so as to open only in the inward direction. Thus, water will enter the well **915** only when the pressure in the surrounding water is greater than that in the well. Initially, the gravity head in the surrounding water will cause the water in the well **915** to rise to the level of surrounding water. Any further rises in the level of the surrounding water will cause a similar rise in the water level in the well **915**. Should the level of the surrounding water drop, the check valves will prevent the loss of water from the well **915**. Thus, regardless of the action of the surrounding sea, the electrodes **910** and **911** will remain submerged except in the unlikely event that the surrounding water drops below its normal low level **920** for a sufficiently long period of time to enable the quantity of water stored in the well **915** above the electrodes to be consumed by the electrolysis process. Note that as the electrolysis progresses, the concentration of brine in the well water will increase. Therefore, it is expected that it will be necessary to periodically pump the fluid out of the well **915** to allow fresh sea water to re-enter the well.

Notwithstanding the safety features designed into the apparatus as described above, in the preferred embodiment, logic devices **916** and **934**, electrically coupled to the generator **909**, are utilized to ensure the reliable operation of the apparatus. As shown in FIG. 7, level sensors **924** and **925**, which may be level switches of the type commonly used in boilers, are mounted in the well **915**. In the event either of the level sensors **924** and **925** sense that the level of the water in the well **915** has dropped below a predetermined level—that is, the level at which they are mounted—the logic devices **916** and **934** will cut off power to the electrodes **910** and **911** to protect them from damage.

In the event of a severe storm or other unusual sea conditions, the level of the surrounding sea water may rise excessively high. Such rise in the surrounding water will cause the water level in the well **915** to rise. Accordingly, level switches **922** and **923** are also mounted in the well **915**. As shown in FIGS. 7, the level switches **922** and **923** are disposed in the head space portions **917** and **918**, respectively, at the level of the compressor intakes **915** and **916**. The signals from the level switches **922** and **923** are directed to logic devices **916** and **934**, respectively, so that the logic devices will shut down the compressor **630** should the water reach the level of the switches. Moreover, as shown in FIG. 7, a sea cock **930** is mounted in portion **903** of the float. In the event level sensors **922** and **923** indicate the aforementioned high water level, logic device **934** will actuate valve **931** installed on the sea cock **930** to flood the portion **903** of the float, thereby causing it to sink deeper into the water where it may safely ride out the storm. A second sea cock and valve (not shown) is mounted in portion **904** of the float and is actuated in a similar fashion by logic device **916**.

As shown in FIG. 7, a pressure sensor **933**, which may be a transducer, is installed in portion **904** of the float **902** to measure the pressure therein. Electrical line **935** transmits the signal from the pressure sensor **933** to the logic device **916**. Should the pressure of the hydrogen in portion **904** exceed a predetermined level, logic device **916** will shut down the compressor. A similar pressure sensor (not shown) is installed in portion **903** of the float and transmits its signal to logic device **934** for similar protection against oxygen over pressure.

FIG. 8 shows a second alternate embodiment of the invention shown in FIG. 4. In this embodiment, the float **951** powers a high pressure pump **955** via the generator **909** discussed with respect to the embodiment shown in FIG. 6. Alternatively, the pump **955** could be mechanically driven via the flywheel **626** discussed with respect to the embodiment shown in FIG. 4. In either case, the pump **955** draws in sea water through intake conduit **953** having an inlet **952** disposed well below the low water line **920**. Pump **955** directs the sea water, via conduit **957**, to a reverse osmosis unit **954**, thereby producing potable water which is then transported to a consumption or storage site through discharge conduit **956** The reverse osmosis unit **954** may be of the microporous membrane type, relying on the pump **955** to force the water through the membrane.

It will be understood that various changes in the details, materials and arrangement of parts which have been herein described and illustrated in order to explain the nature of this invention may be made by those skilled in the art within the principle and scope of the invention as expressed in the following claims.

It will further be understood that the "Abstract of the Disclosure" set forth above is intended to provide a non-legal technical statement of the contents of the disclosure in compliance with the Rules of Practice of the U.S. Patent and Trademark Office, and is not intended to limit the scope of the invention described and claimed herein.

What is claimed is:

1. A wave-power collection apparatus for installation in a body of water for converting wave energy into at least hydrogen gas and for at least collecting said gas in said apparatus, comprising:

(a) pilaster means located extending to protrude for a preselected distance above the high water line of said body of water;

(b) float means trackingly mounted for movement along said pilaster means in response to tidal and wave motion;

(c) generating means mounted on said pilaster, for generating current from the movement of said float means;

(d) electrolysis means electrically connected to said generating means, and in liquid communication with said body of water, said electrolysis means generating hydrogen and oxygen utilizing said current generated by said generating means, said electrolysis means comprising at least two opposingly charged electrodes;

(e) gas separation means mounted on said pilaster means generally disposed between said electrodes, for separating oxygen and hydrogen gases produced by said electrolysis means;

(f) compressor means mounted on said pilaster means, for compressing at least hydrogen gas separated by said gas separation means; and,

(g) gas storage means for at least storing hydrogen gas, said storage means comprising at least a portion of said float means, said gas storage means being in fluid communication with the output of said compressor means.

2. The apparatus of claim **1**, wherein said pilaster means comprises a hollow tubular member defining a hollow well portion, and wherein said electrodes are disposed below a predetermined low water line within said hollow well portion.

3. The apparatus of claim **2**, wherein said gas separation means comprises a microporous barrier between said electrodes.

4. The apparatus of claim **3**, wherein said hollow tubular member defines a head space and wherein said gas separation means comprises a baffling means for separating portions of said headspace of said well portion.

5. The apparatus of claim **4**, wherein the compressor means compresses said hydrogen gas generated by said electrolysis means and transmits said compressed hydrogen gas to said float means.

6. The apparatus of claim **5**, further comprising means for placing said well portion in liquid communication with said body of water at a location below said low water line.

7. The apparatus of claim **6**, wherein said compressor means is powered by current produced by said generator means.

8. The apparatus of claim **6**, wherein said compressor means is mechanically powered by the movement of said float means.

9. The apparatus of claim **6**, wherein said liquid communication means comprises a plurality of apertures formed in said pilaster means.

10. The apparatus of claim **9**, further comprising filtering means disposed in said pilaster means for filtering debris from water entering said well portion.

11. A wave-power collection apparatus for installation in a body of water, comprising:

(a) a tubular member forming a hollow well therein, said tubular member having a first end attached to the floor of said body of water and a second end extending above the surface of said body of water, said hollow well in flow communication with said body of water, whereby said water fills at least a portion of said hollow well;

(b) a toroidal float adapted for vertical movement along said tubular member in response to tidal and wave motion in said body of water;

(c) a shaft mounted in said tubular member and means for converting said vertical movement of said float into rotation of said shaft;

(d) a generator adapted to be driven by said rotation of said shaft and produce electrical current thereby; and

(e) first and second electrodes mounted in said portion of said hollow well filled with said water and electrically coupled to said generator, said electrodes adapted to cause current to flow therebetween through said water.

12. The apparatus of claim **11** wherein said electrodes are adapted to dissociate a portion of said water through which said current flows, thereby producing hydrogen and oxygen, and further comprising a compressor for compressing said hydrogen.

13. The apparatus of claim **12** wherein said compressor is driven by said rotation of said shaft.

14. The apparatus of claim **12** wherein said compressor is driven by said electrical current produced by said generator.

15. The apparatus of claim **11** further comprising means for flooding said float with water from said body of water, said flooding means responsive to the level of said water filling said well.

16. The apparatus of claim **11** further comprising means for severing said electrical couple between said electrodes and said generator if the level of said water filling said well drops below a predetermined level.

17. The apparatus of claim **11** wherein said electrical couple severing means comprises logic means.

18. The apparatus of claim **17** wherein said electrical couple severing means further comprises level sensing means.

* * * * *

US 20120286623A1

(19) **United States**
(12) **Patent Application Publication**
Ogram

(10) **Pub. No.: US 2012/0286623 A1**
(43) **Pub. Date: Nov. 15, 2012**

---

(54) **ATMOSPHERIC ENERGY COLLECTION**

(75) Inventor: **Mark E. Ogram**, Tucson, AZ (US)

(73) Assignee: **Sefe, Inc.**, Tempe, AZ (US)

(21) Appl. No.: **13/103,963**

(22) Filed: **May 9, 2011**

Publication Classification

(51) **Int. Cl.**
H02N 3/00 (2006.01)

(52) **U.S. Cl.** **310/309**

(57) **ABSTRACT**

The subject matter described herein is an atmospheric energy collector. The atmospheric energy collector includes of a windsock arrangement that has a large up-wind opening on one side and that tapers from the larger up-wind opening on the one side to a small down-wind opening on the other side. The up-wind side is secured to a tether such that an electrically conducting material (e.g. metal) included in construction of the atmospheric energy collector is connected to the tether. The windsock arrangement is extended outwards by wind and the like atmospheric conditions such that the electrically conducting material collects the atmospheric energy and transfers the collected energy to the tether.

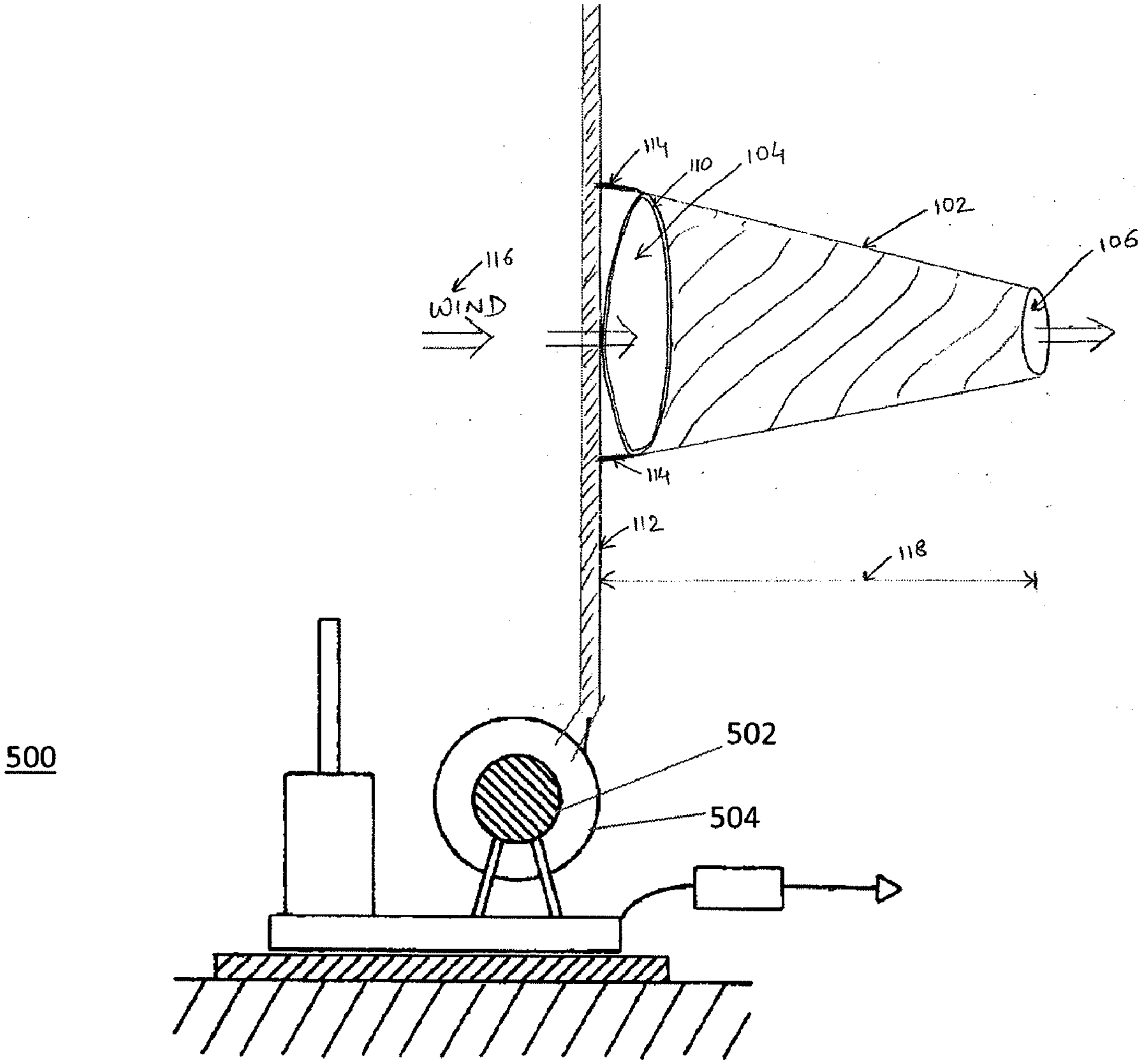

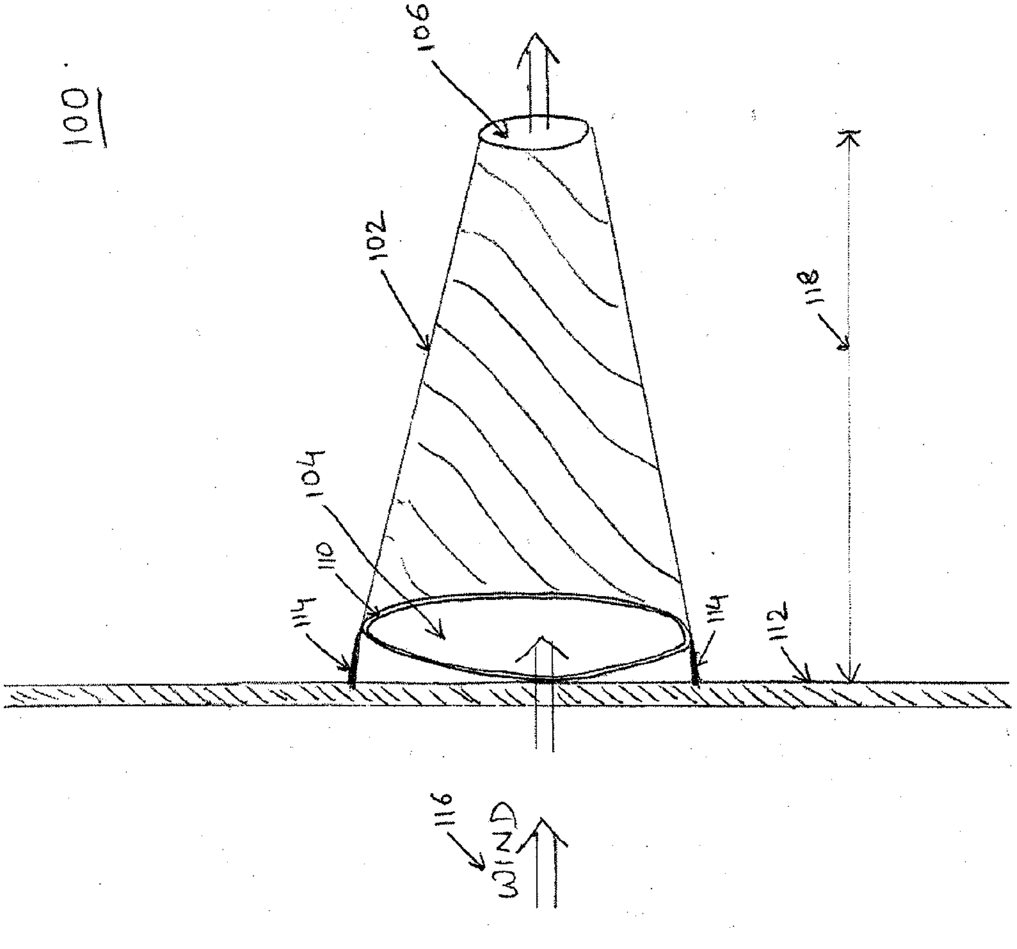
100
106
102
118
104
110
114
114
112
116
WIND

FIG. 1A

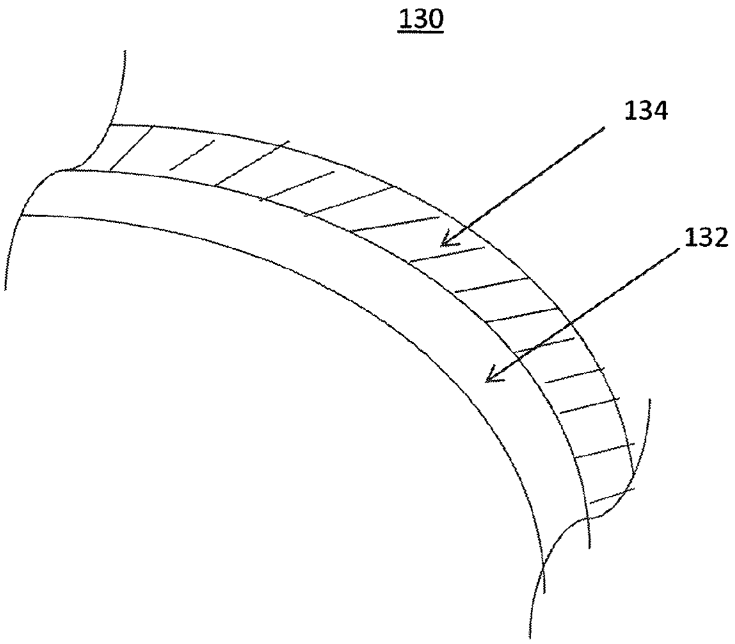

FIG. 1B

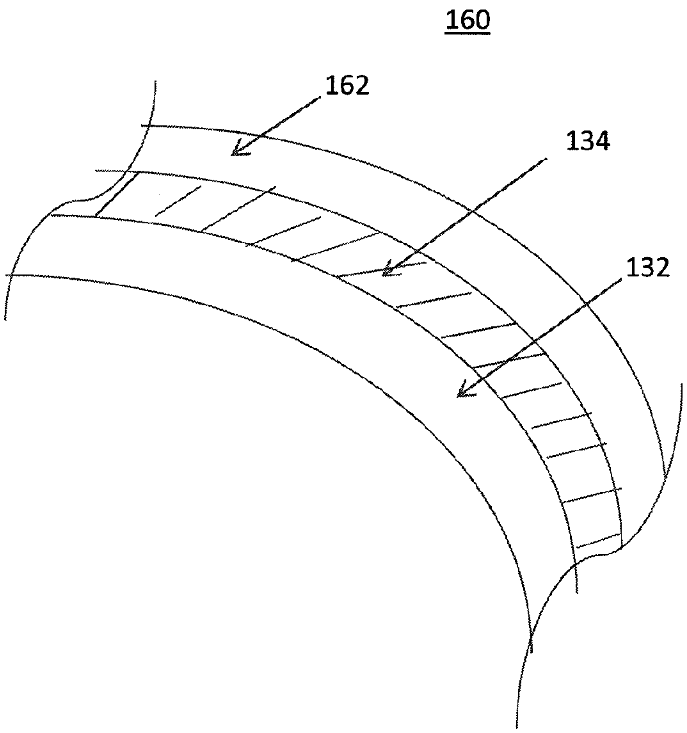
160
162
134
132

FIG. 1C

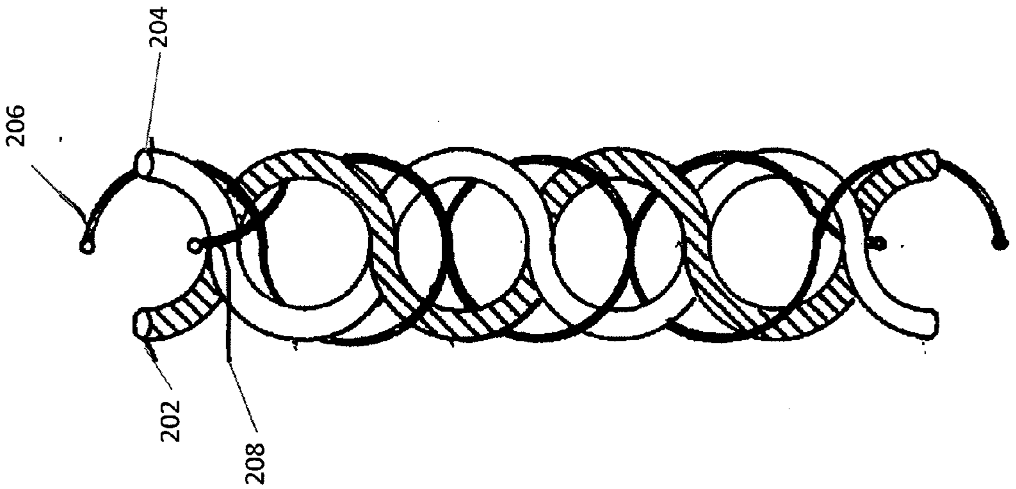
206
204
202
208

FIG. 2

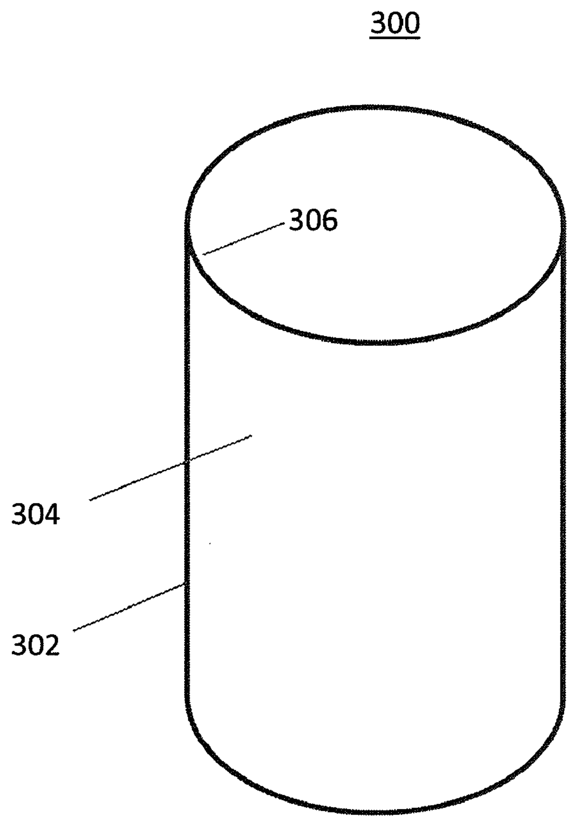
300
306
304
302

FIG. 3A

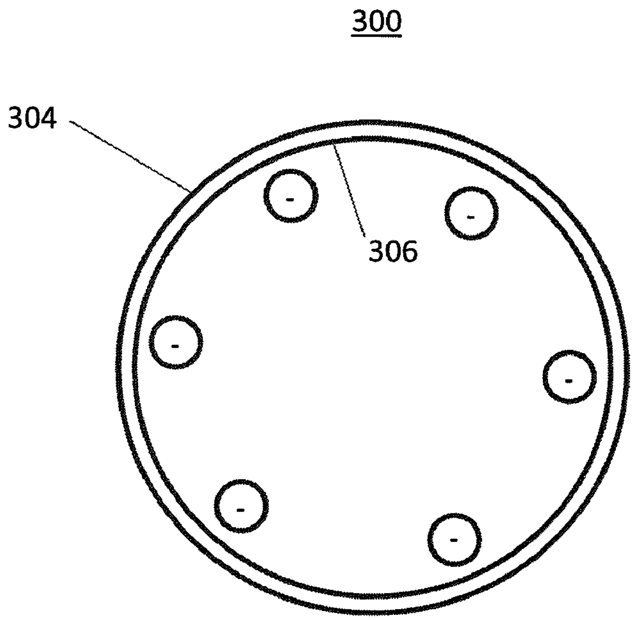

FIG. 3B

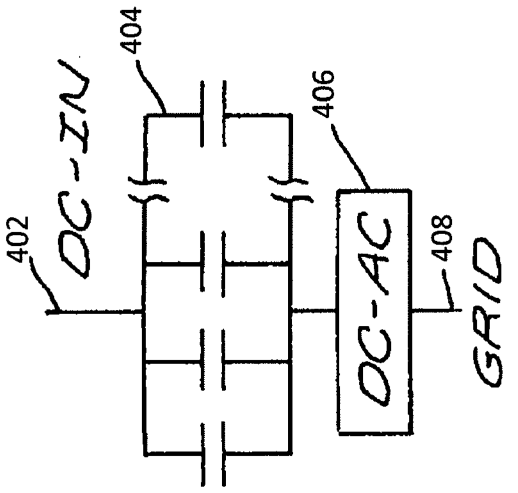
402
DC-IN
404
406
DC-AC
408
GRID

FIG. 4A

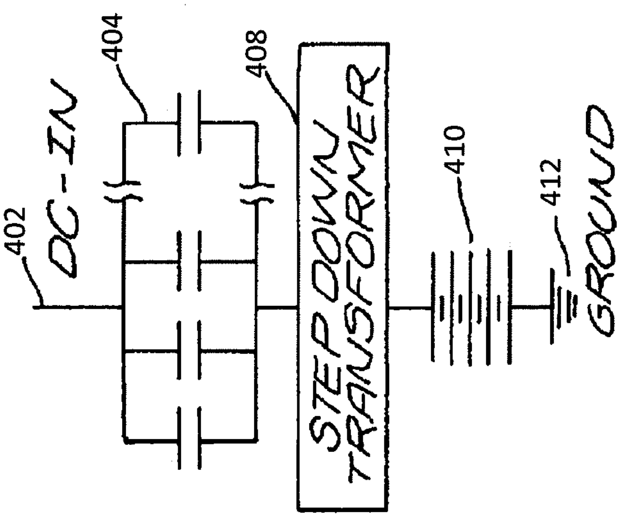
402
DC-IN
404
408
STEP DOWN TRANSFORMER
410
412
GROUND

FIG. 4B

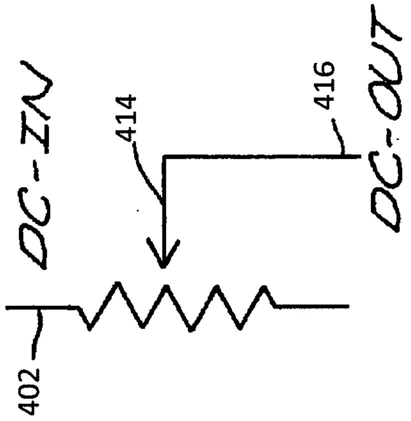
402
DC-IN
414
416
DC-OUT

FIG. 4C

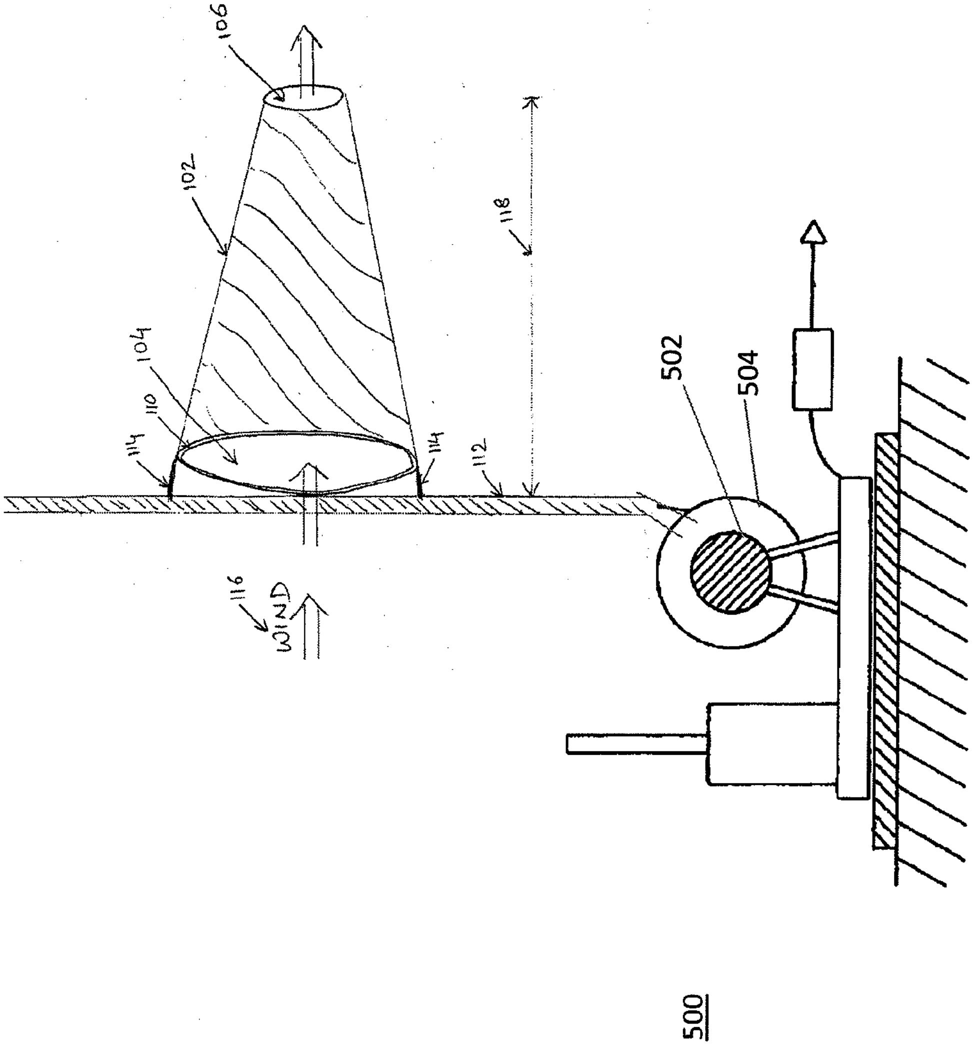
106
102
118
104
110
114
114
112
116
WIND
502
504
500

FIG. 5

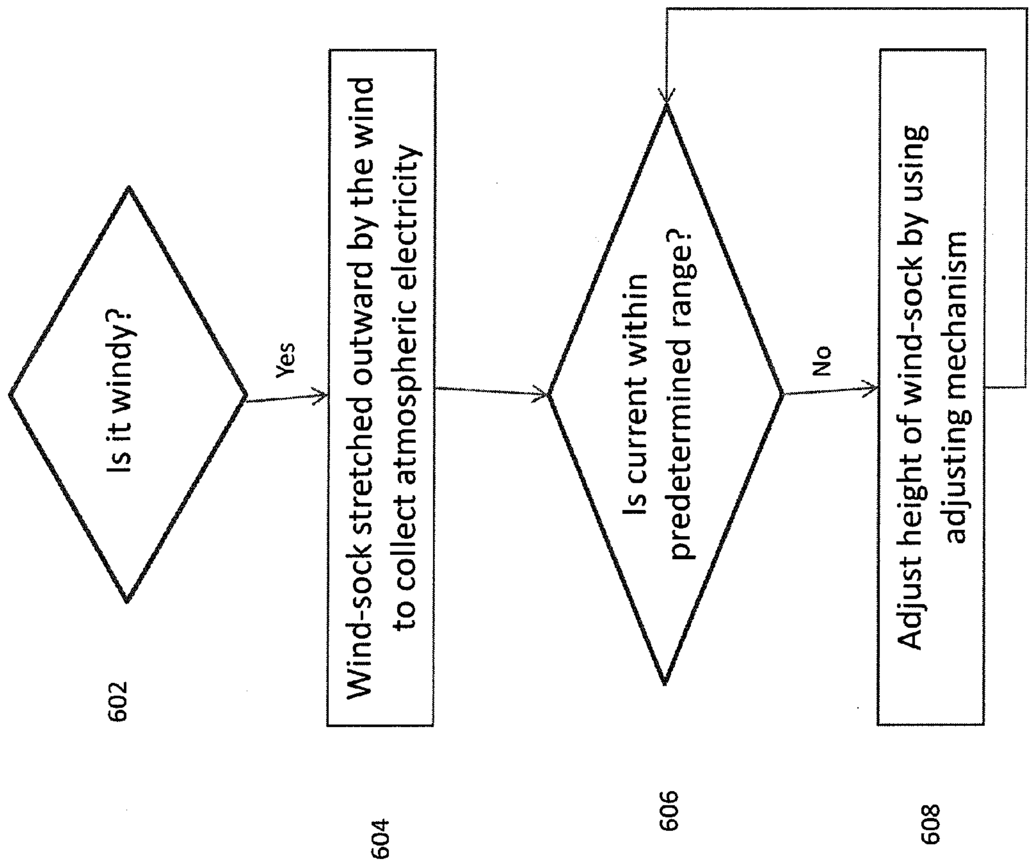
602
Is it windy?
Yes
604
Wind-sock stretched outward by the wind to collect atmospheric electricity
606
Is current within predetermined range?
No
608
Adjust height of wind-sock by using adjusting mechanism

FIG. 6

ATMOSPHERIC ENERGY COLLECTION

TECHNICAL FIELD

[0001] The subject matter described herein relates to a light-weight, extendible electricity collecting windsock arrangement comprising an enhanced collection surface for atmospheric electrical energy collection.

BACKGROUND

[0002] The atmosphere above the earth is known to include electrical charge. The earth's surface is negatively charged, while the air above it is positively charged. All atmospheric effects are a result of an interplay between these two huge areas of opposite electrical energy. The potential difference between the positively charged atmosphere and the negatively charged earth surface causes atmospheric electrical energy to be developed. Some of this atmospheric electrical energy may be experienced through thunderstorms.

[0003] Even though there is a huge amount of atmospheric electrical energy in the atmosphere, collection of the atmospheric electrical energy still remains a problem. Due to high voltage and low current conditions, there is a need for large collection surfaces for maximizing the collection of atmospheric electrical energy. Note that the terms atmospheric electricity, atmospheric energy, and atmospheric electrical energy have been used interchangeably in this specification.

[0004] The problem of low current is compounded by a heavy weight of the collectors of atmospheric electrical energy. The heavy weight of the collectors increases payload requirements of a lift mechanism, which may be used to lift the collectors to an appropriate height or altitude so that the collection of atmospheric electrical energy is maximized. Thus, a need exists for large yet light-weight collectors such that payload requirements for the lift mechanism remain reasonable.

SUMMARY

[0005] The subject matter described herein relates to an atmospheric energy collector. The atmospheric energy collector includes an electricity collecting windsock arrangement that has a large up-wind opening on one side and that tapers from the larger up-wind opening on the one side to a small down-wind opening on the other side. The up-wind side is secured to a tether such that an electrically conducting material (e.g. metal) included in construction of the atmospheric energy collector is connected to the tether. The electricity collecting windsock arrangement is extended outwards by wind and the like atmospheric conditions such that the electrically conducting material collects the atmospheric energy and transfers the collected energy to the tether. Further, the electricity collecting windsock includes light-weight collectors such that payload requirements for the lift mechanism remain reasonable (i.e. remain within a predetermined value).

[0006] In one aspect, an electricity collection apparatus includes a tether and a windsock. The windsock is formed of an insulation material with a metal deposited on at least a portion of the insulation material. The metal of the windsock is electrically connected to the tether. The windsock extends in a direction of ambient wind to provide a surface area based on the ambient wind, and the surface area provides the metal to collect electrical energy from the ambient wind, the collected electrical energy being transferred to the tether.

[0007] In another aspect, an electricity collection apparatus includes a windsock formed of an insulation material the windsock to extend in a direction of ambient wind to provide a surface area based on the ambient wind. The apparatus further includes one or more electrical conductors provided on at least a portion of the insulation material to collect electrical energy from the ambient wind. The apparatus further includes an electrically conductive tether connected with windsock to anchor the windsock in the direction of the ambient wind, the electrically conductive tether being electrically connected to the one or more electrical conductors provided on at least a portion of the insulation material to transfer the collected electrical energy from the one or more electrical conductors to an electrical storage.

[0008] The subject matter described herein provides many advantages. For example, large collection surfaces allow a maximized collection of atmospheric energy. Moreover, the light-weight collectors allow payload requirements for the lift mechanism to remain reasonable (i.e. remain within a predetermined value).

[0009] The details of one or more variations of the subject matter described herein are set forth in the accompanying drawings and the description below. Other features and advantages of the subject matter described herein will be apparent from the description and drawings, and from the claims.

DESCRIPTION OF DRAWINGS

[0010] The accompanying drawings, which are incorporated in and constitute a part of this specification, show certain aspects of the subject matter disclosed herein and, together with the description, help explain some of the principles associated with the disclosed implementations. In the drawings,

[0011] FIG. **1**A illustrates extendible electricity collecting windsock arrangement that is used for collection of atmospheric electrical energy in accordance with implementations of the current subject matter;

[0012] FIG. **1**B illustrates a cross-section of one implementation of the electrically conducting windsock in accordance with implementations of the current subject matter;

[0013] FIG. **1**C illustrates a cross-section of another implementation of the electricity collecting windsock in accordance with implementations of the current subject matter;

[0014] FIG. **2** illustrates an exemplary conductive line in accordance with implementations of the current subject matter;

[0015] FIG. **3**A illustrates a front view of a conductive line in accordance with implementations of the current subject matter;

[0016] FIG. **3**B illustrates a top view of the conductive line shown in FIG. **3** in accordance with implementations of the current subject matter;

[0017] FIGS. **4**A, **4**B, and **4**C illustrate electrical schematics for handling the static charge from the atmosphere in accordance with implementations of the current subject matter;

[0018] FIG. **5** illustrates an adjusting apparatus (winch motor and spool) that enables adjustment of location/height of the electricity collecting windsock arrangement in accordance with implementations of the current subject matter; and

[0019] FIG. **6** is a process flow diagram illustrating aspects of a method in accordance with implementations of the current subject matter.

[0020] When practical, similar reference numbers denote similar structures, features, or elements.

DETAILED DESCRIPTION

[0021] To address these and potentially other issues with currently available solutions, one or more implementations of the current subject matter provide methods, systems, articles or manufacture, and the like that can, among other possible advantages, provide an energy collector formed as a windsock and having an enhanced collection surface for atmospheric electrical energy collection. In preferred implementations, an electricity collecting windsock is formed as a flexible cylinder, a truncated cone, or a cone with a metal surface, and is mounted to a mast or other tether so as to capture ambient wind which extends the windsock outward from the mast, in turn forming the largest possible surface area for the wind conditions to enable the metal surface to collect atmospheric amperage. The collected amperage is transferred from the metal surface of the electricity collecting windsock to the mast or tether, and eventually to an electricity storage device. Although implementations in which windsock arrangements having one windsock each are discussed below, those or ordinary skill in the art understand that a windsock arrangement may include two or more windsocks for enhanced atmospheric electricity collection.

[0022] FIG. **1**A illustrates an implementation of electricity collecting windsock arrangement **100** that is used to collect atmospheric electrical energy. The electricity collecting windsock arrangement **100** includes an extendible electricity collecting windsock **102**. The extendible electricity collecting windsock **102** is made of a material that has a high tensile strength and provides electrical insulation. This electrically insulating material **134** (discussed below with respect to FIG. **1**B) may be a plastic or a flexible glass, such as polyester like polyethylene terephthalate (PET) or biaxially-oriented polyethylene terephthalate (BoPET). BoPET is known in the industry by different trade names, some of which are Mylar, Melinex and Hostaphan.

[0023] As is described later with respect to FIGS. **1**B and **1**C, which illustrate cross-sections of implementations of electricity collecting windsock **102**, a thin coat of an electrically conducting material **132**, **162** is deposited on the electrically insulating material **134**. The deposited electrically conducting material **132**, **162** is a good conductor of electricity, and helps electricity collecting windsock **102** to tap atmospheric electrical energy. The electrically conducting material may be any one of a metal or a suitable alloy, such as one of or a combination of one or more of gold, silver, copper, aluminum, and the like.

[0024] The electricity collecting windsock **102** has a large up-wind opening up-wind opening **104** on one side. The electricity collecting windsock **102** tapers from the large up-wind opening **104** on the one side to a small down-wind opening **106** on the other side of the electricity collecting windsock **102**. The up-wind opening **104** has a metal loop **110** on the circumference of the up-wind opening **104**. The metal loop **110** is attached to a tether **112** using an attachment mechanism **114**. The attachment mechanism **114** may include a soldering mechanism. In some implementations, other attachment mechanisms **114** are also known to be used, such as nut and bolt mechanism, threading mechanism, gluing mechanism, and the like. The electricity collecting windsock **102** is extended outwards by wind **116** and/or like atmospheric conditions such that the electrically conducting material (e.g. metal) collects the atmospheric electrical energy and transfers the collected energy to the tether **112**.

[0025] The tether **112** includes a conductive line that is isolated from the ground. The conductive line is used to transfer the collected atmospheric electrical energy from the electricity collecting windsock **102** to an electricity storage device (not shown) where the collected atmospheric charge may be gathered for later or simultaneous use.

[0026] When wind **116** strikes against the electricity collecting windsock arrangement **100** in the direction illustrated, the electricity collecting windsock **102** extends outward, thus allowing the electrically conducting material **132** (discussed below with respect to FIG. **1**B) to collect maximum possible atmospheric electricity. In one implementation, the outward extension **118** is fifty feet or more. When there is no wind, the electricity collecting windsock **102** may be in a relatively compressed state such that the length of the electricity collecting windsock **102** is less than the outward extension **118** caused by windy conditions or like atmospheric conditions. The collected atmospheric electrical energy is communicated to the tether **112**.

[0027] FIG. **1**B illustrates a cross-section **130** of one implementation of the electrically conducting windsock **102**. The cross-section **130** illustrates the electrically conducting material **132** deposited on the inner surface of the electrically insulating material **134**. The electricity collecting windsock **102** includes the electrically conducting material **132** and the electrically insulating material **134** such that the electrically conducting material **132** forms the inner surface of the electricity collecting windsock **102**. When the wind **116** strikes the electricity collecting windsock arrangement, the electricity collecting windsock **102** extends outward, thus allowing the electrically conducting material **132** to collect maximum possible atmospheric electricity. The electrically conducting material **132** may include multiple collectors that are spaced apart such that the collection surface is greatly enhanced. The multiple collectors may be spaced apart in patterns, such as spaced apart lines, squares, rectangles, circles, and the like, or any other pattern that may maximize the collection of the atmospheric electrical energy. As noted above, the electrically insulating material **134** may be a plastic or a flexible glass, such as polyester like polyethylene terephthalate (PET) or biaxially-oriented polyethylene terephthalate (BoPET). The electrically conducting material **132** may be a metal or alloy, such as one of or a combination of one or more of gold, silver, copper, aluminum, and the like.

[0028] FIG. **1**C illustrates a cross-section **160** of another implementation of the electricity collecting windsock **102**. The electrically conducting material (e.g. metal deposit) **132**, **162** may be placed on both sides of the electrically insulating material **134** such that the collection ability is increased even further. Multiple collectors are implemented and are spaced apart on electrically conducting material **132**, **162** such that the collection surface is greatly enhanced. The multiple collectors may form one or more patterns, as noted above with respect to FIG. **1**B. Furthermore, the electrically insulating material **134** (e.g. mylar) used in the electricity collecting windsock **102** is light, thereby minimizing the overall weight of the electricity collecting windsock **102** and reducing the weight for payload calculations. The electrically conducting materials **132**, **162** may be an electrically conducting metal or alloy, such as one of or a combination of at least one of gold, copper, aluminum, silver, and the like. The electrically conducting materials **132**, **162** may be made of the same metal/alloy or a different metal/alloy.

[0029] The electrically conducting materials **132**, **162** may be attached to the electrically insulating material **134** by a gluing mechanism. In some implementations, other attachment mechanisms are known to be implemented, such as paint coating mechanism, a threading mechanism, a nut and bolt mechanism, soldering mechanism, and the like. The attachment mechanism between the electrically conducting material **132** and the electrically insulating material **134** may or may not be the same as the attachment mechanism between the electrically conducting material **162** and the electrically insulating material **134**.

[0030] FIG. **2** illustrates an exemplary conductive line in accordance with one implementation. This type of conductive line is commonly referred to as poly-wire. The conductive line consists of multiple interwoven strands of plastic **202** and **204** woven into a cord or rope arrangement having intertwined therein exposed metal wires **206** and **208**. Although FIG. **2** illustrates two plastic strands and two metal wires, any number of possible combinations of plastic strands and metal wires is possible. The exposed metal wires **206** and **208** attract the atmospheric static charge and transmit the charge down to the electricity storage device (not shown).

[0031] FIG. **3**A illustrates a front view of a conductive line **300** in accordance with another implementation. FIG. **3**B illustrates a top view of the conductive line **300** illustrated in FIG. **3**. The conductive line **300** creates an ionized pathway for the flow of the static charges from the atmosphere to the electricity storage device via the electricity collecting windsock arrangement **100**. This conductive line utilizes a tube **302** having an outer layer **304** of PET Film (Biaxially-oriented polyethylene terephtalate polyester film). The tube **302** provides exceptionally high tensile strength and is chemically and dimensionally stable. In one implementation, the tube **302** may have an ideal diameter of between two and three inches. An interior metal coating **306** provides an initial conduit for the flow of static charge. The static charge through the metal may force the tube **302** to expand due to the repulsion experienced by like charges. Further, the flow of electricity causes the interior of the tube **302** to become ionized to provide an additional pathway for the atmospheric static charges to the electricity storage device (not shown).

[0032] FIGS. **4**A, **4**B, and **4**C illustrate electrical schematics for handling the static charge from the atmosphere. By maintaining the voltage being collected in a prescribed range, an electrical conversion system is easily designed. While FIGS. **4**A, **4**B, and **4**C illustrate some electrical configurations, those of ordinary skill in the art readily recognize a variety of other configurations that may serve the same function.

[0033] Referencing FIG. **4**A, Direct Current In (DC IN) **402** is buffered by a gang of capacitors **404** before being communicated to a DC/AC converter **406**. The DC/AC converter converts the direct current into alternating current suitable for placement over an existing electrical grid **408** such as normally found from a power-plant. Those of ordinary skill in the art readily recognize a variety of DC/AC converters that may work in this capacity.

[0034] FIG. **4**B illustrates an electrical arrangement suitable for use in charging a battery. DC IN **402** is buffered by capacitor bank **404** before entering into a step down transformer **408**. Step down transformer **408** reduces the voltage so that the voltage can safely be introduced into battery **410**, which is connected to ground **412** at the battery's other pole. Those of ordinary skill in the art readily recognize a variety of batteries that may work in this capacity.

[0035] In FIG. **4**C, DC IN **402** is fed into an adjustable rheostat **414**, which is controlled by the controller so that the DC OUT **416** falls within a specified range.

[0036] FIG. **5** illustrates an adjusting apparatus **500** that enables adjustment of height of the electricity collecting windsock arrangement **100** in accordance with one implementation. As per one implementation, the height of the electricity collecting windsock **102** can be adjusted using a winch motor **502** and a spool **504**. The winch motor **502** and the spool **504** can release or withdraw the tether **112** to adjust the height of the electricity collecting windsock **102** that is connected to the tether **112**. In one implementation, this release and the withdrawal may be performed manually. In another implementation, this release and the withdrawal may be performed automatically based on information obtained from a sensor system that measures atmospheric electrical energy being collected per unit time. If the sensor indicates that a current flow is diminishing, then the tether **112** is released/extended from the spool **504** to increase the altitude of the electricity collecting windsock **102** such that more static charge from the atmosphere is gathered. If the sensor indicates that collected atmospheric electrical energy exceeds a preset level per unit time, the tether **112** is withdrawn onto the spool **504** to decrease the static charge being collected from the atmosphere.

[0037] FIG. **6** is a process flow diagram illustrating aspects of a method consistent with implementations of the current subject matter. At step **602**, the functioning of the electricity collecting windsock arrangement **100** depends on whether atmospheric conditions are windy. If the atmospheric conditions are windy, the wind **116** stretches/extends outward the electricity collecting windsock **102**—Step **604**. Next, it is determined whether the atmospheric electrical energy (atmospheric current) that is collected is in a predetermined range—Step **606**. If the collected atmospheric electrical energy is in the predetermined range, the flow goes back to step **604**. If the current is not in the predetermined range, the height of the electricity collecting windsock **102** may be adjusted using adjusting apparatus **500**, such as the winch motor **502** and spool **504**. One skilled in the art understands that other implementations may include any other adjusting apparatus **500**, such as a pulley, a wheel mechanism, or the like.

[0038] The implementations set forth in the foregoing description do not represent all implementations consistent with the subject matter described herein. Instead, they are merely some examples consistent with aspects related to the described subject matter. Although a few variations have been described in detail herein, other modifications or additions are possible. In particular, further features and/or variations can be provided in addition to those set forth herein. For example, the implementations described above can be directed to various combinations and sub-combinations of the disclosed features and/or combinations and sub-combinations of one or more features further to those disclosed herein. In addition, the logic flows depicted in the accompanying figures and/or described herein do not necessarily require the particular order shown, or sequential order, to achieve desirable results. The scope of the following claims may include other implementations or embodiments.

What is claimed is:

1. An electricity collection apparatus comprising:

a tether; and

a windsock formed of an insulation material with a metal deposited on at least a portion of the insulation material, the metal of the windsock being electrically connected to the tether, the windsock to extend in a direction of ambient wind to provide a surface area based on the ambient wind, the surface area providing the metal to collect electrical energy from the ambient wind, the collected electrical energy being transferred to the tether.

2. The apparatus in accordance with claim **1**, wherein the metal is deposited on at least one of opposite sides of the insulation material.

3. The apparatus in accordance with claim **2**, wherein the metal is a metal selected from a group of metals comprising gold, silver, copper and aluminum.

4. The apparatus in accordance with claim **1**, wherein the insulation material is a polyester film.

5. The apparatus in accordance with claim **1**, wherein the transferred electrical energy at the tether is further transferred from the tether to an electricity storage device.

6. The apparatus in accordance with claim **1**, wherein the metal comprises conductors that are light-weight thereby allowing payload requirements for adjusting the windsock to an optimum location to be within a predetermined value, the optimum location comprising a height of the windsock from ground level.

7. The apparatus in accordance with claim **6**, wherein the deposited metal conductors form a pattern that maximizes the collection of the electrical energy at the optimum location.

8. An electricity collection apparatus comprising:

a windsock formed of an insulation material having a metal deposited on at least a portion of the insulation material, the windsock to extend in a direction of ambient wind to provide a surface area based on the ambient wind, the surface area providing the metal to collect electrical energy from the ambient wind; and

an electrically conductive tether connected with windsock to anchor the windsock in the direction of the ambient wind, the electrically conductive tether being electrically connected to the metal of the insulation material, the collected electrical energy being transferred to the tether.

9. The apparatus in accordance with claim **8**, wherein the metal is deposited on at least one of opposite sides of the insulation material.

10. The apparatus in accordance with claim **9**, wherein the metal is a metal selected from a group of metals comprising gold, silver, copper and aluminum.

11. The apparatus in accordance with claim **8**, wherein the insulation material is a polyester film.

12. The apparatus in accordance with claim **8**, wherein the transferred electrical energy at the tether is further transferred from the tether to an electricity storage device.

13. The apparatus in accordance with claim **8**, wherein the metal comprises conductors that are light-weight thereby allowing payload requirements for adjusting the windsock to an optimum location to be within a predetermined value, the optimum location comprising a height of the windsock from ground level.

14. The apparatus in accordance with claim **13**, wherein the metal conductors form a pattern that maximizes the collection of the electrical energy at the optimum location.

15. An electricity collection apparatus comprising:

a windsock formed of an insulation material the windsock to extend in a direction of ambient wind to provide a surface area based on the ambient wind;

one or more electrical conductors provided on at least a portion of the insulation material to collect electrical energy from the ambient wind; and

an electrically conductive tether connected with windsock to anchor the windsock in the direction of the ambient wind, the electrically conductive tether being electrically connected to the one or more electrical conductors provided on at least a portion of the insulation material to transfer the collected electrical energy from the one or more electrical conductors to an electrical storage.

* * * * *

US007855476B2

(12) **United States Patent**
Ogram

(10) **Patent No.: US 7,855,476 B2**
(45) **Date of Patent: Dec. 21, 2010**

(54) ATMOSPHERIC ELECTRICAL GENERATOR

(76) Inventor: **Mark Ellery Ogram**, 780 S. Freeman Rd., Tucson, AZ (US) 85748

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 254 days.

(21) Appl. No.: **12/218,297**

(22) Filed: **Jul. 14, 2008**

(65) **Prior Publication Data**

US 2010/0007218 A1 Jan. 14, 2010

(51) **Int. Cl.**
H02G 11/00 (2006.01)

(52) **U.S. Cl.** **307/145**; 307/149; 361/218; 361/230

(58) **Field of Classification Search** 307/149, 307/145; 244/30, 31; 174/2; 361/212, 215–218, 361/230, 231
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427	A *	5/1901	Palencsar	307/149
911,260	A *	2/1909	Pennock	307/149
1,014,719	A *	1/1912	Pennock	307/149
1,540,998	A *	6/1925	Plauson	310/309
3,121,196	A *	2/1964	Kasemir	324/464
4,234,167	A *	11/1980	Lane	254/291
4,486,669	A *	12/1984	Pugh	290/44
4,842,219	A *	6/1989	Jakubowski et al.	244/31
5,203,542	A	4/1993	Coley et al.	
6,254,034	B1 *	7/2001	Carpenter	244/153 R
6,829,911	B2	12/2004	Davis et al.	
6,961,662	B2	11/2005	Murphy	
7,016,785	B2	3/2006	Makela et al.	
7,200,418	B2	4/2007	Kaikuranta et al.	
7,330,366	B2	2/2008	Lee et al.	
7,375,492	B2	5/2008	Calhoon et al.	
7,378,181	B2	5/2008	Skinlo	
7,388,350	B1	6/2008	Wright	
7,394,671	B2	7/2008	Fukumoto et al.	
7,397,220	B2	7/2008	Uchida et al.	

* cited by examiner

Primary Examiner—Fritz M Fleming
(74) *Attorney, Agent, or Firm*—Mark Ogram

(57) ABSTRACT

A mechanism to tap an electrical source which uses an aircraft (preferably a lighter than air balloon) tethered by a conductive line. The conductive line is extended/withdrawn by a winch motor to adjust the altitude of the aircraft. The conductive line is isolated from the ground and an electrical conductor is connected to the conductive line and to an electrical load. In this manner, static electricity generated in the atmosphere is gathered for use.

15 Claims, 4 Drawing Sheets

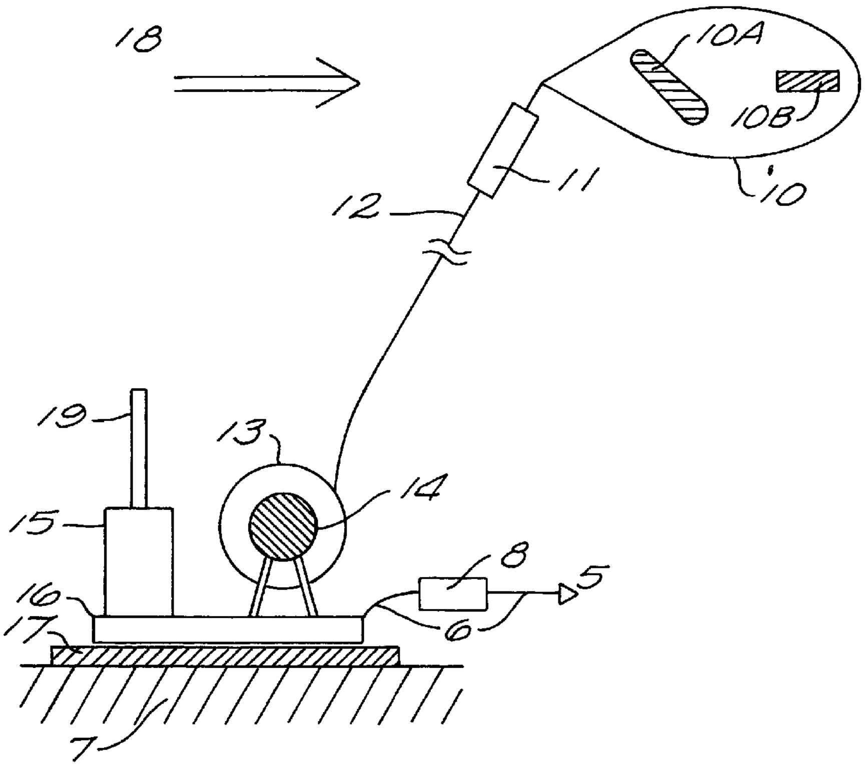

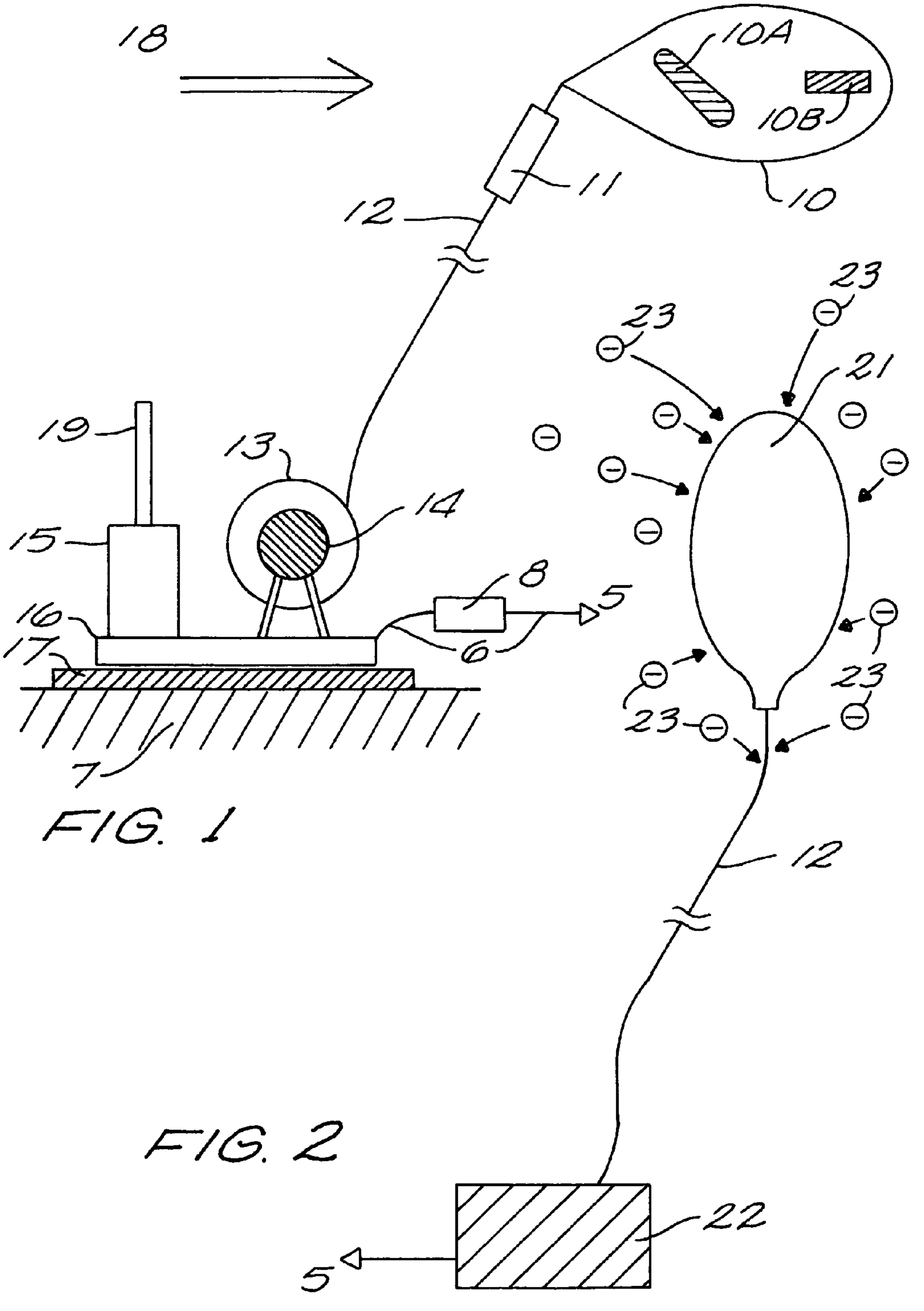
18
10A
10B
10
11
12
19
13
14
15
8
5
16
6
17
7
23
21
23
23
23
12
22
5

FIG. 1

FIG. 2

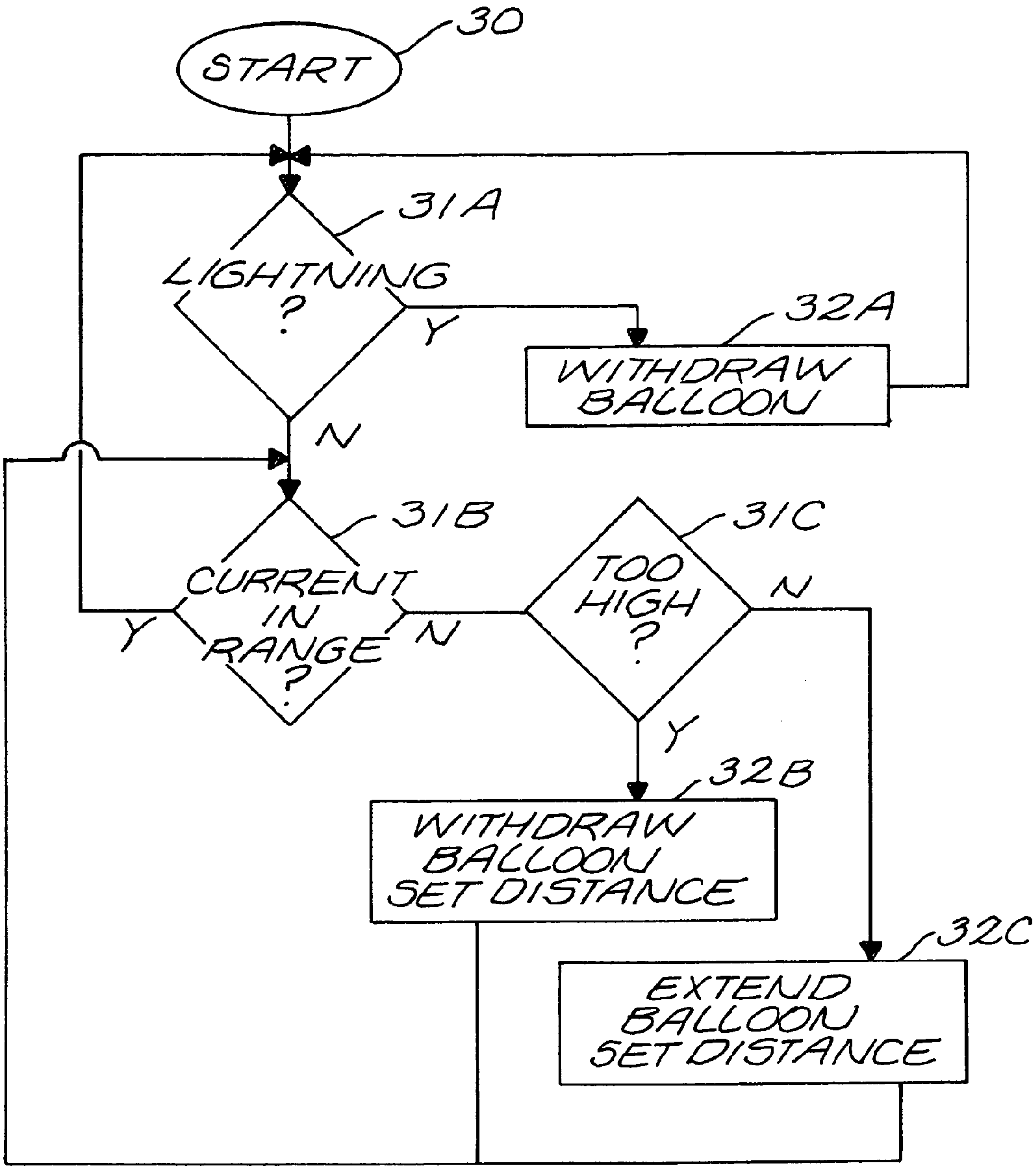
30
START
31A
LIGHTNING ?
Y
32A
WITHDRAW BALLOON
N
31B
CURRENT IN RANGE ?
Y
N
31C
TOO HIGH ?
N
Y
32B
WITHDRAW BALLOON SET DISTANCE
32C
EXTEND BALLOON SET DISTANCE

FIG. 3

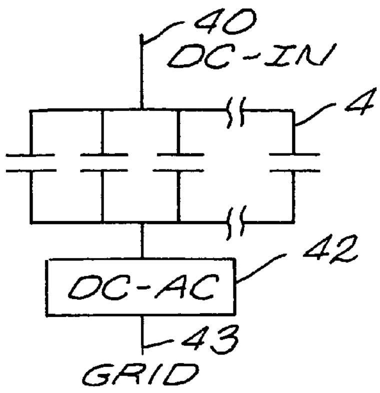
40
DC-IN
41
42
DC-AC
43
GRID

FIG. 4A

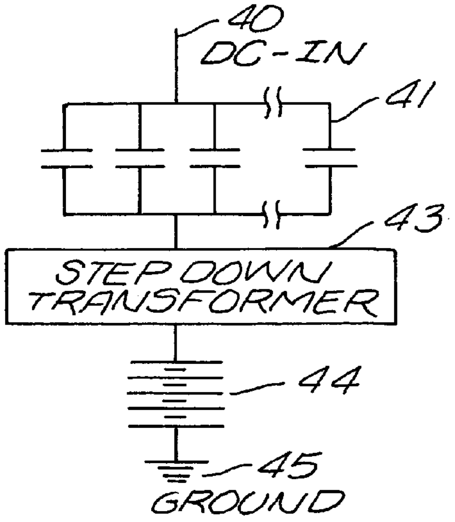
40
DC-IN
41
43
STEP DOWN TRANSFORMER
44
45
GROUND

FIG. 4B

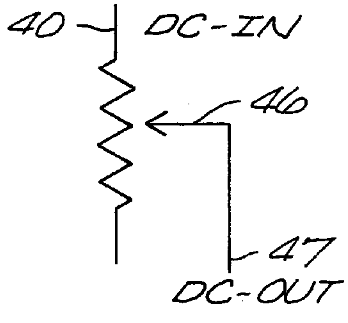
40
DC-IN
46
47
DC-OUT

FIG. 4C

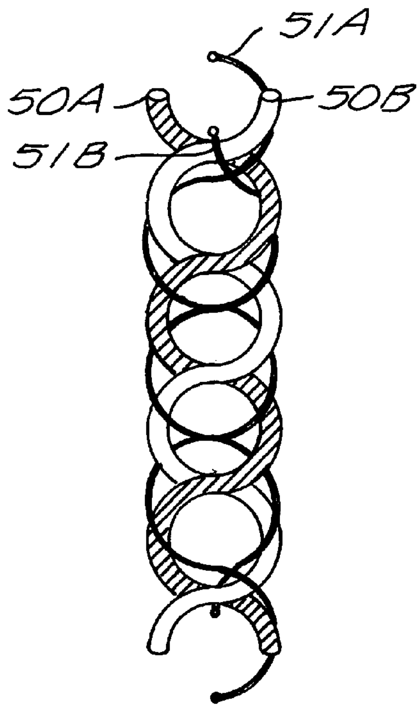
51A
50A
50B
51B

FIG. 5

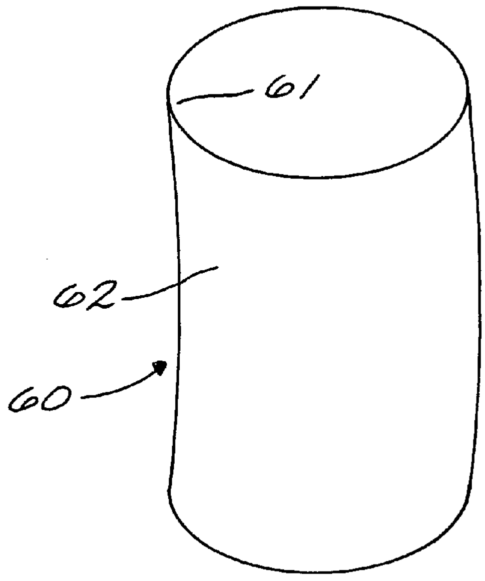

FIG. 6A

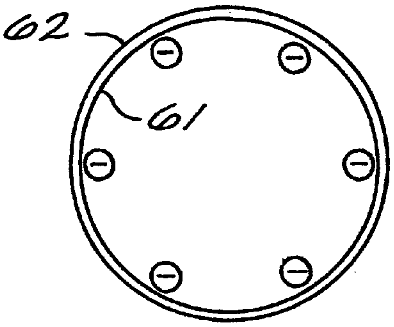

FIG. 6B

REPLACEMENT

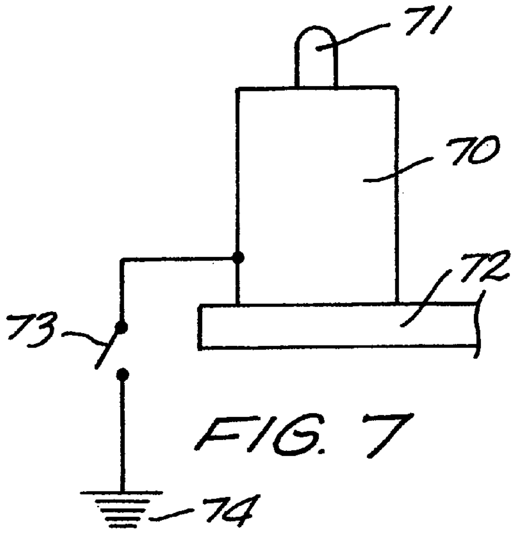
71
70
72
73
74

FIG. 7

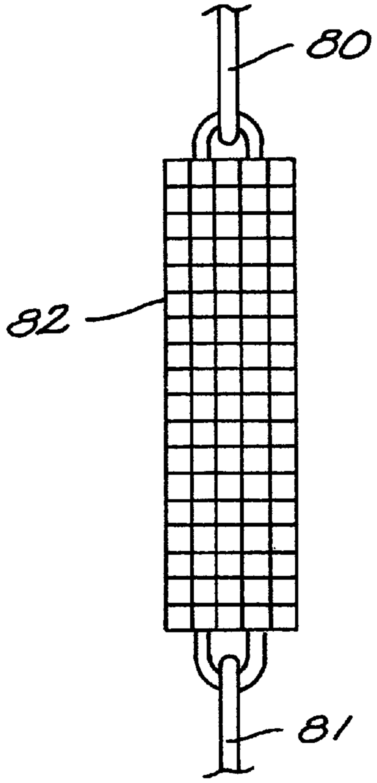

FIG. 8A

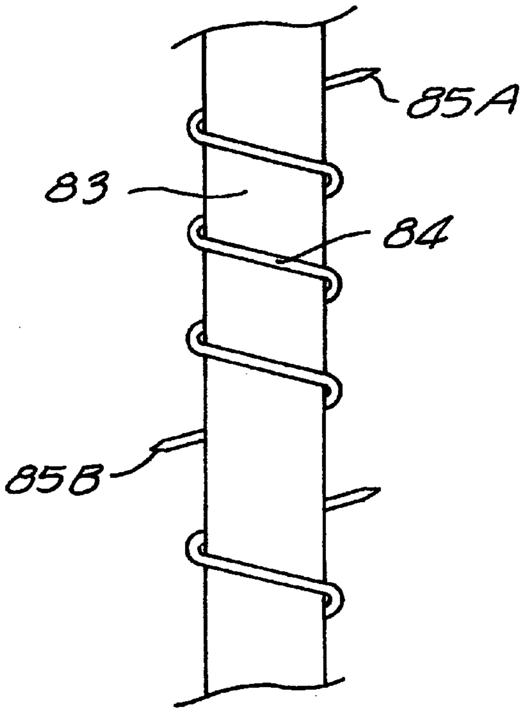
85A
83
84
85B

FIG. 8B

ATMOSPHERIC ELECTRICAL GENERATOR

BACKGROUND OF THE INVENTION

This invention relates to the production of electrical energy and more particularly to the production of electrical energy from the atmosphere.

Everyone is familiar with Benjamin Franklin's kite experiment of 1752. Using a kite whose string had become wet, negative charges from the passing clouds flowed into the string, down to the suspended key, and then into a Leyden jar via a thin metal wire. Franklin was protected by a dry silk string; but, when Franklin's knuckle came too close to the key, he received a strong shock. Fortunately, Benjamin Franklin was not killed, others who tried this same experiment were not so lucky.

Since then, the formation of lightning has remained something of a mystery. Lightning bolts are triggered when a negatively charged cloud base induces a positive charge from the ground, thereby forming a "pathway" for the discharge of the collected electrical energy.

Lightning travels up to 60,000 miles per hour with a flash that is brighter than ten million 100-watt lightbulbs. This wattage is as much power as is produced by all of the electricity plants in the United States and with a voltage of up to 300 million volts.

It is this very fact, the power within lightning is immense, that has prevented any successful collection of the electrical energy from lightning. The electricity in lightning is far too extreme for current technology to harness.

While lightning has attracted a energy starved industrial world, no one has developed any technique to harness this naturally occurring electrical source.

It is clear there is a continuing need for an electrical source other than carbon-based fuels and that the naturally occurring electricity in the atmosphere is being ignored.

SUMMARY OF THE INVENTION

The invention is a mechanism which taps into the naturally occurring static electricity in the atmosphere. Whereas heretofore, the attempt to garner electricity from the atmosphere has focused exclusively on capturing lightning, the present invention syphons off the static electricity which is generated from any agitated air and avoids lightning.

Lightning is only the final discharge of the static electricity, whether that lightning is intra-cloud lightning, cloud-to-ground lightning, or inter-cloud lightning. Other types of final discharges are known as heat lightning, summer lightning, sheet lightning, ribbon lightning, silent lightning, ball lightning, bead lightning, elves, jets, and sprites. Well before these discharges are observed, as the atmosphere becomes agitated by wind or thermal activity, static electricity is being generated.

The present invention recognizes that this static electricity is being formed and creates a mechanism to capture it.

The mechanism of this invention utilizes an aircraft such as a lighter than air balloon. While the preferred embodiment uses a foil balloon, a variety of other aircraft are obvious to those of ordinary skill in the art, including, but not limited to: gliders, rubber balloons (such as weather balloons), biaxially-oriented polyethylene terephthalate polyester film balloons, and latex balloons.

Within this discussion, the balloon is referenced, but, the invention is not intended to be limited solely to balloons.

The balloon is sent aloft and is tethered by a conductive line. In this context, the conductive line may be any obvious to those of ordinary skill in the art. For the preferred embodiment, the conductive line is a generically referred to as a "poly-rope" and is commercially available through a variety of sources. A suitable conductive line is described in U.S. Pat. No. 5,203,542, entitled "Apparatus for Improved Electric Fence Wire Construction for use with Intensive Grazing" issued Apr. 20, 1993, to Coley, et al. and incorporated hereinto by reference.

The conductive line is played out of a winch to control the altitude of the balloon. The motor controlling the winch is able to reverse direction to both extend and withdraw the conductive line which is wrapped around a spool on the winch. The winch/spool combination are part of a base unit.

In some embodiments of the invention, the spool is constructed of rubber so as to insulate the conductive line from the winch assembly. In this embodiment, only the conductive line is charged by the atmospheric static electricity while the winch remains neutral.

In yet another embodiment, the winch/spool are part of a base unit which is itself isolated from the ground by an insulator. In this embodiment, the entire base unit is charged by the atmospheric static electricity.

A conductor, such as an insulated wire, is electrically connected to the conductive line. In one embodiment, where the conductive line is electrically isolated from the spool and winch motor, the conductor is connected to the conductive line. In the embodiment where the conductive line is electrically connected to the base unit, then the conductor is connected anywhere on a metallic base unit.

The other end of the conductor is connected to a load. The load in this case can be any of a variety of electrical loads well known to those of ordinary skill in art, including, but not limited to a motor, a battery system, or the electrical grid for the system.

In the preferred embodiment, a sensor array is used to monitor the activities both at the base unit (such as electrical flow within the conductor) and in the surrounding locale.

A sensor monitoring the electrical flow (i.e. voltage and/or current) within the conductor is used to monitor the electrical activity within the conductor.

In the preferred embodiment, a lightning sensor monitors for lightning activity within the locale. As noted earlier, the electrical characteristic of lightning is so extreme that ideally this discharge is avoided as it might damage the mechanism of this invention.

The sensor array is utilized by a controller, such as microprocessor, programmed to operate the mechanism as outlined herein.

The controller operates the winch motor to extend or withdraw the conductive line and by extension the altitude of the balloon. The controller is programmed to operate the winch by monitoring the electrical characteristics of the conductor and adjusting the balloon's altitude to maintain these characteristics within the conductor within a preset range.

This preset range is established either in the base programming of the controller or is established by an operator of the system.

As example, by controlling the amount of current being withdrawn from the atmosphere, the mechanism operates within a safe range and also provides a relatively stable current flow from which a variety of activities can take place (such as DC-AC conversion).

The controller also utilizes the lightning sensor to protect the mechanism from a lightning strike. Should lightning be detected within a pre-determined range (as established by the software or defined by an operator), then the balloon is pulled down to minimize the risk of damage from a lightning strike.

The invention, together with various embodiments thereof will be more fully explained by the following description of the accompanying drawings.

DRAWINGS IN BRIEF

FIG. **1** diagrams the preferred embodiment of the invention.

FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

DRAWINGS IN DETAIL

FIG. **1** diagrams the preferred embodiment of the invention.

Balloon **10** is an aircraft which, in this illustration, is a lighter than air balloon. Wings **10**A, extending from the body of balloon **10**, provide additional lift in air flow **18**. Tail **10**B helps to stabilize balloon **10**.

Balloon **10** is tethered to the ground via conductive line **12**. As noted earlier, a variety of configurations and materials are available to serve as conductive line **12**. In this illustration, a poly-wire is used Poly-wire is commercially available through a variety of vendors, including, but not limited to: Jeffers Livestock and Zareba Systems, Inc, of Ellendale, Minn.

In this embodiment, located proximate to balloon **10**, is an electrical collection enhancement lead **11** which assists in the collection of the static electrical charge in the atmosphere. Electrical collection enhancement lead **11** is configured to attract the static charge and conduct the charge into the conductive line **12**.

The electricity flows down the conductive line into spool **13**, where the conductive line **12** is collected and either withdrawn or dispensed through operation of winch motor **14**.

Winch motor **14** and spool **13** are mounted onto base unit **16** which is electrically isolated from ground **7** using insulator **17**. Note, in this embodiment of the invention, when electricity is being collected from the atmosphere, the entire base unit **16** becomes charged. In another embodiment of the invention, spool **13** is constructed of rubber, thereby preventing base unit **16** from becoming charged, thereby restricting the charging from the atmosphere to only conductive line **12**.

In this embodiment, conductor **6** is connected to base unit **16** (since the entire base unit **16** is charged and the base unit is metallic) to communicate the electrical current to load **5**. Conductor **6** is ideally an insulated wire.

The electrical current through conductor **6** is measured using sensor **8**.

In the alternative embodiment discussed above, where only the conductive line **12** is charged, then conductor **6** is connected to conductive line **12**.

Controller **15**, located in this embodiment on base unit **16**, operates winch motor **14** in response to signals from sensor **8** (measuring the current being discharged to load **5**) to maintain the current flow within a pre-defined range. As the current flow diminishes, then the conductive line **12** extended from spool **13** to increase the altitude of balloon **10** to that more static charge from the atmosphere is gathered; as the current flow falls exceeds a preset level, conductive line **12** is withdrawn onto spool **13** to decrease the static charge being collected from the atmosphere.

The range of current flow through conductor **6** is ideally set by the program, although some embodiments of the invention permit an operator to establish this range of operation.

In an alternative embodiment, the sensor monitoring conductor **6** monitors the voltage therein.

In the preferred embodiment of the invention, controller **15** is also equipped with a lightning sensor **19**. In this embodiment, when lightning is sensed within a preset range, then substantially all of conductive line **12** is wound onto spool **13** to pull balloon **10** near the ground and protect the entire mechanism from being damaged from a lightning discharge.

In the preferred embodiment, the "safe" distance form lightning is set in the programming of controller **15** and is ideally two miles; other embodiments permit the operator to "safe" distance.

There are a variety of lightning sensors well known to those of ordinary skill in the art, including, but not limited to those described in: U.S. Pat. No. 7,016,785, entitled "Lightning Detection" issued to Makela, et al. on Mar. 21, 2006; U.S. Pat. No. 6,829,911, entitled "Lightning Detection and Prediction Alarm Device" issued to Jones, et al. on Dec. 7, 2004; U.S. Pat. No. 7,200,418, entitled "Detection of Lightning" issued to Karikuranta, et al. on Apr. 3, 2007; and U.S. Pat. No. 6,961,662, entitled "Systems and Methods for Spectral Corrected Lightning Detection" issued to Murphy on Nov. 1, 2005; all of which are incorporated hereinto by reference.

In another embodiment of the invention, controller **15** is not located on base unit **16**, rather it is remote and communicates its control signals to winch motor **14** using radio waves.

FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

Static charges **23** are generated in the atmosphere by agitated air. These static charges are often collected at the bottom of clouds, but exist in other environments as well.

Balloon **21** is extended into this strata of static charges **23** which are then attracted to conductive line **12** to flow to base unit **22** and then onto load **5**.

By increasing or decreasing the altitude of balloon **21** (defined by the length of the extended conductive line **12**), conductive line **12** is selectively exposed to varying densities and levels of the static charge strata, and by extension, the current flow or voltage is increased or decreased.

FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

Once the program starts **30**, the lightning sensor is checked to determine if lightning has occurred within the unsafe range **31**A, if it has, then the balloon is lowered **32**A, and the program continues monitoring the status of lightning until no lightning is detected.

When the lightning status is acceptable, then the current within the conductor is checked to see if the current is within the prescribed range **31**B. If the current is acceptable (within range) the program returns to check the lightning status **31**A; otherwise a determination is made to see if the current is above the prescribed range **31**C.

If the current is above the prescribed range, then the altitude of the balloon is withdrawn a set amount **32**B (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

If the current is below the prescribed range, then the altitude of the balloon is extended a set amount **32**C (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

In this manner of feed-back and minor adjustments in the altitude of the balloon, the current is maintained within a prescribed range which can be handled by the downstream electrical system.

As noted earlier, some embodiments of the invention monitor the voltage instead of the current.

FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

By maintaining the voltage being collected in a prescribed range, an electrical conversion system is easily designed. While FIGS. **4**A, **4**B, and **4**C illustrate some electrical configurations, those of ordinary skill in the art readily recognize a variety of other configurations which will serve the same function.

Referencing FIG. **4**A, Direct Current In (DC IN) **40** is buffered by a gang of capacitors **41** before being communicated to a DC/AC converter **42**. The DC/AC converter converts the direct current into a an alternating current suitable for placement over an existing electrical grid **43** such as normally found from a power-plant.

Those of ordinary skill in the art readily recognize a variety of DC/AC converters, including, but not limited to: U.S. Pat. No. 7,394,671, entitled “Controller IC, DC-AC Conversion Apparatus, and parallel running system of DC-AC Conversion Apparatuses” issued to Fukumoto, et al. on Jul. 1, 2008; and, U.S. Pat. No. 7,330,366, entitled “DC-AC Converter” issued to Lee, et al. on Feb. 12, 2008; all of which are incorporated hereinto by reference.

FIG. **4**B illustrates an electrical arrangement suitable for use in charging a battery. DC IN **40** is buffered by capacitor bank **41** before entering into a step down transformer **43**. Step down transformer **43** reduces the voltage so that the voltage can safely be introduced into battery **44** which is connected to ground **45** at the battery’s other pole.

Those of ordinary skill in the art readily recognize a variety of batteries which will work in this capacity, including, but not limited to those described in: U.S. Pat. No. 7,378,181, entitled “Electric Storage Battery Construction and Manufacture” issued to Skinlo on May 27, 2008; U.S. Pat. No. 7,388,350, entitled “Battery with Electronic Compartment” issued to Wright on Jun. 17, 2008; U.S. Pat. No. 7,397,220, entitled “Connection Member and Battery Pack” issued to Uchida, et al. on Jul. 8, 2008; and, U.S. Pat. No. 7,375,492, entitled “Inductively Charged Battery Pack” issued to Calhoon, et al. on May 20, 2008; all of which are incorporated hereinto by reference.

In FIG. **4**C, DC IN **40** is fed into an adjustable rheostat **46** which is controlled by the controller so that the DC OUT **47** falls within a specified range.

FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

This type of conductive line is commonly called poly-wire and consists of multiple interwoven strands of plastic **50**A and **50**B woven into a cord or rope arrangement having intertwined therein exposed metal wires **51**A and **51**B. While this illustration shows two plastic strands and two metal wires, any number of possible combinations is possible.

The exposed metal wires **51**A and **51**B attract the atmospheric static charge and transmit the charge down to the base unit (not shown).

FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

This conductive line utilizes a tube **60** having an outer layer **62** of PET Film (Biaxially-oriented polyethylene terephtalate polyester film) which provides exceptionally high tensile strength and is chemically and dimensionally stable. The tube has an ideal diameter of between two and three inches.

An interior metal coating **61** provides an initial conduit for the flow of static charge. The static charge through the metal forces the tube to expand due to the repulsion experienced by like charges. Further, the flow of electricity causes the interior of the tube **60** to become ionized to provide an additional pathway for the atmospheric static charges to the base unit (not shown).

Because outer layer **62** provides a gas barrier, the resulting ionization is not dissipated by air currents, thereby providing a highly stable pathway.

FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

In this embodiment, controller box **70**, resting on insulating pad **72**, is in communication with the sensors as described above. Using the input from these sensors, when there is flow of electricity through the base unit, warning flashing light **71** is illuminated. To electrically neutralize the mechanism, switch **73** is activated to pass any existing current into the ground **74**.

FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

Referencing FIG. **8**A, enhanced electrical collection lead **82** is a wire mesh which is in electrical communication with conductive line **81** and balloon **80**. Because of the significant amount of metal exposed by enhanced electrical collection lead **82**, more static electricity from the atmosphere is drawn to the collection lead **82**, and then down conductive line **81** to the base unit (not shown).

Conductive lead **82** is positioned proximate to balloon **80**.

In FIG. **8**B, poly-wire **83** has enhanced electrical collection leads **84** wrapped therearound. Collection leads **84** have pointed ends **85**A and **85**B which have a propensity to attract more electricity than rounded ends do.

It is clear from the foregoing that the present invention captures an entirely new source of electrical energy.

What is claimed is:

1. A mechanism to tap an electrical source comprising:

a) a lighter than air balloon suspended in the atmosphere;

b) a base unit having a spool of conductive line on a winch motor, one end of said conductive line secured to said lighter than air balloon, a portion of said conductive line collecting electricity in the atmosphere, said winch motor capable of selectively extending or withdrawing said conductive line from said spool;

c) an insulator electrically isolating said conductive line from ground;

d) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load being powered by collected electricity from said conductive line;

e) an electrical flow sensor monitoring electrical flow through said conductor and generating an electrical flow indicia indicative of said electrical flow in said conductor; and,

f) a controller receiving said electrical flow indicia and selectively operating said winch motor such that said electrical flow indicia remains within a selected operating range.

2. The mechanism to tap an electrical source according to claim **1**, wherein said selected operating range is established by an operator.

3. The mechanism to tap an electrical source according to claim **2**, further including a warning light activated when said electrical flow indicia is non-zero.

4. The mechanism to tap an electrical source according to claim **1**,

a) further including a lightning sensor generating a lightning presence indicia indicative of lightning within a prescribed range, said presence indicia being communicated to said controller; and,

b) wherein said control mechanism, in response to said lightning presence indicia, operates said winch motor to withdraw substantially all of said conductive line onto said spool.

5. The mechanism to tap an electrical source according to claim **4**, wherein said prescribed range is established by an operator.

6. The mechanism to tap an electrical source according to claim **1**, further including an electrical collection enhancement lead in electrical contact with a first end of said conductive material, said electrical collection enhancement lead configured to attract static electricity.

7. The mechanism to tap an electrical source according to claim **6**, wherein said electrical collection enhancement lead is positioned proximate to said lighter than air balloon.

8. The mechanism to tap an electrical source according to claim **7**, wherein said electrical collection enhancement lead includes at least two pointed electrical conductors.

9. The mechanism to tap an electrical source according to claim **1**, wherein said insulator electrically isolates said conductive line from said winch motor.

10. The mechanism to tap an electrical source according to claim **1**, wherein said insulator electrically isolates said base unit from ground.

11. A mechanism comprising:

a) an airborne aircraft having a conductive line secured to a winch capable of extending or withdrawing said conductive line from a spool, when said airborne aircraft is aloft, said conductive line collecting electricity from the atmosphere;

b) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load such that collected electricity from said conductive line powers said load;

c) an electrical flow sensor monitoring electrical flow through said conductor and generating an electrical flow indicia indicative of said electrical flow in said conductor; and,

d) a controller receiving said electrical flow indicia adjusting an altitude of said airborne aircraft via said winch such that said electrical flow indicia remains within a selected operating range.

12. The mechanism to tap an electrical source according to claim **11**,

a) further including a lightning sensor communicating with said controller; and,

b) wherein said controller, in response to selected signals from said lightning sensor withdraws substantially all of said conductive line.

13. A mechanism to tap an electrical source comprising:

a) a lighter than air balloon suspended in the atmosphere;

b) a base unit having a spool of conductive line on a winch motor, one end of said conductive line suspended by said balloon, a portion of said conductive line collecting electricity from the atmosphere, said winch motor capable of adjusting an altitude of said lighter than air balloon by selectively extending or withdrawing said conductive line from said spool;

c) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load being powered by said collected electricity from said conductive line;

d) a sensor array having,

 1) an electrical flow sensor monitoring electrical flow through said conductor and generating an electrical flow indicia indicative of said electrical flow in said conductor, and,

 2) a lightning sensor monitoring existence of proximate lightning; and,

e) a controller responsive to said electrical flow indicia from said sensor array to selectively operate said winch motor.

14. The mechanism to tap an electrical source according to claim **13**,

a) further including a warning light; and,

b) wherein said controller activates said warning light when said electrical flow in said conductor is non-zero.

15. The mechanism to tap an electrical source according to claim **14**, further including an electrical collection enhancement lead in electrical contact with said conductive material, said electrical collection enhancement lead configured to attract static electricity and positioned proximate to said lighter than air balloon.

* * * * *

US 20100007321A1

(19) **United States**

(12) **Patent Application Publication**

Ogram

(10) **Pub. No.: US 2010/0007321 A1**

(43) **Pub. Date: Jan. 14, 2010**

(54) **DYNAMIC ELECTRICAL CONVERTER SYSTEM**

(76) Inventor: **Mark Ellery Ogram**, Tucson, AZ (US)

Correspondence Address:
Mark E. Ogram
Ste. 110-186, 6990 E. 22nd Street
Tucson, AZ 85710-5191 (US)

(21) Appl. No.: **12/231,663**

(22) Filed: **Sep. 4, 2008**

Related U.S. Application Data

(63) Continuation-in-part of application No. 12/218,297, filed on Jul. 14, 2008.

Publication Classification

(51) **Int. Cl.**
G05F 5/00 (2006.01)
H02H 3/22 (2006.01)

(52) **U.S. Cl.** **323/299**; 361/111

(57) **ABSTRACT**

A dynamic electrical converter system configured to receive a variable voltage input which is then converted to a proper electrical configuration for a load. The converter assembly utilized multiple converters, each of which is configured to accept a unique range of voltages and from these voltages creating the desired electrical output configuration. A monitor checks the incoming voltage and by controlling switches going to each of the converters, selectively closes the circuit to the appropriate converter; thereby, routing the incoming voltage to the converter capable of addressing the incoming voltage.

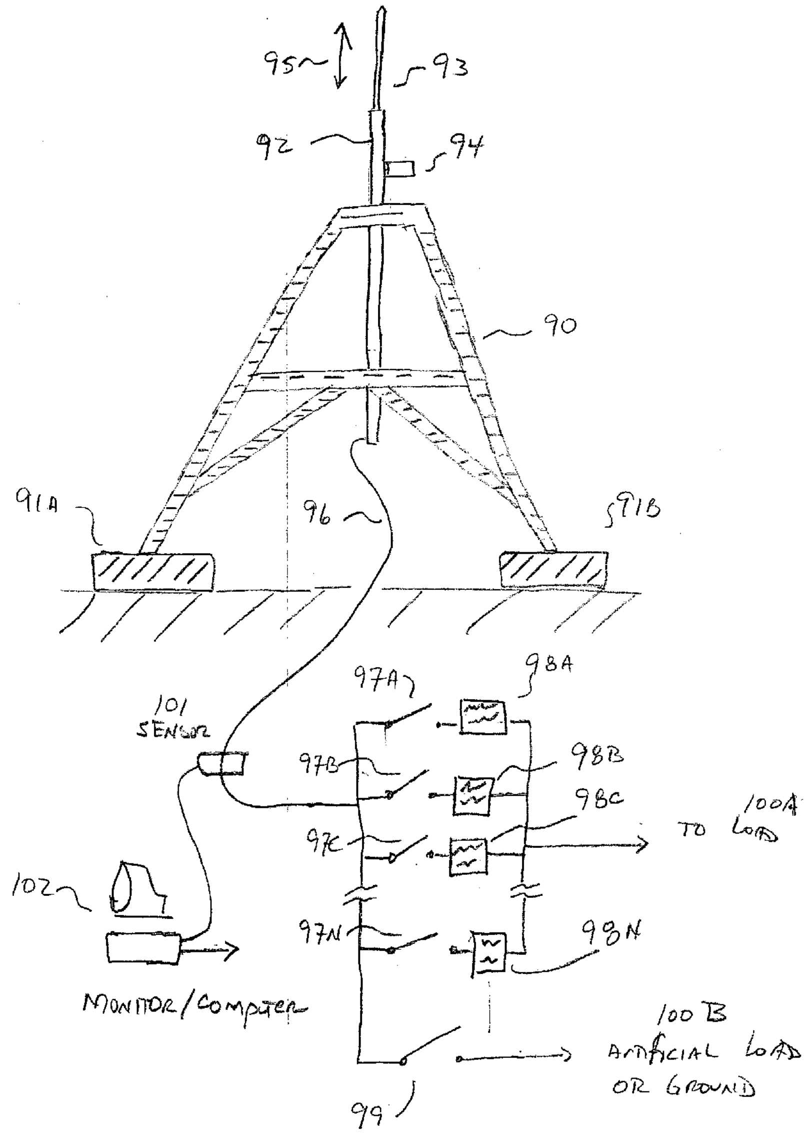
95
93
92
94
90
91A
91B
96
97A
98A
101
SENSOR
97B
98B
98C
100A
TO LOAD
97C
102
97N
98N
MONITOR/COMPUTER
100B
ARTIFICIAL LOAD
OR GROUND
99

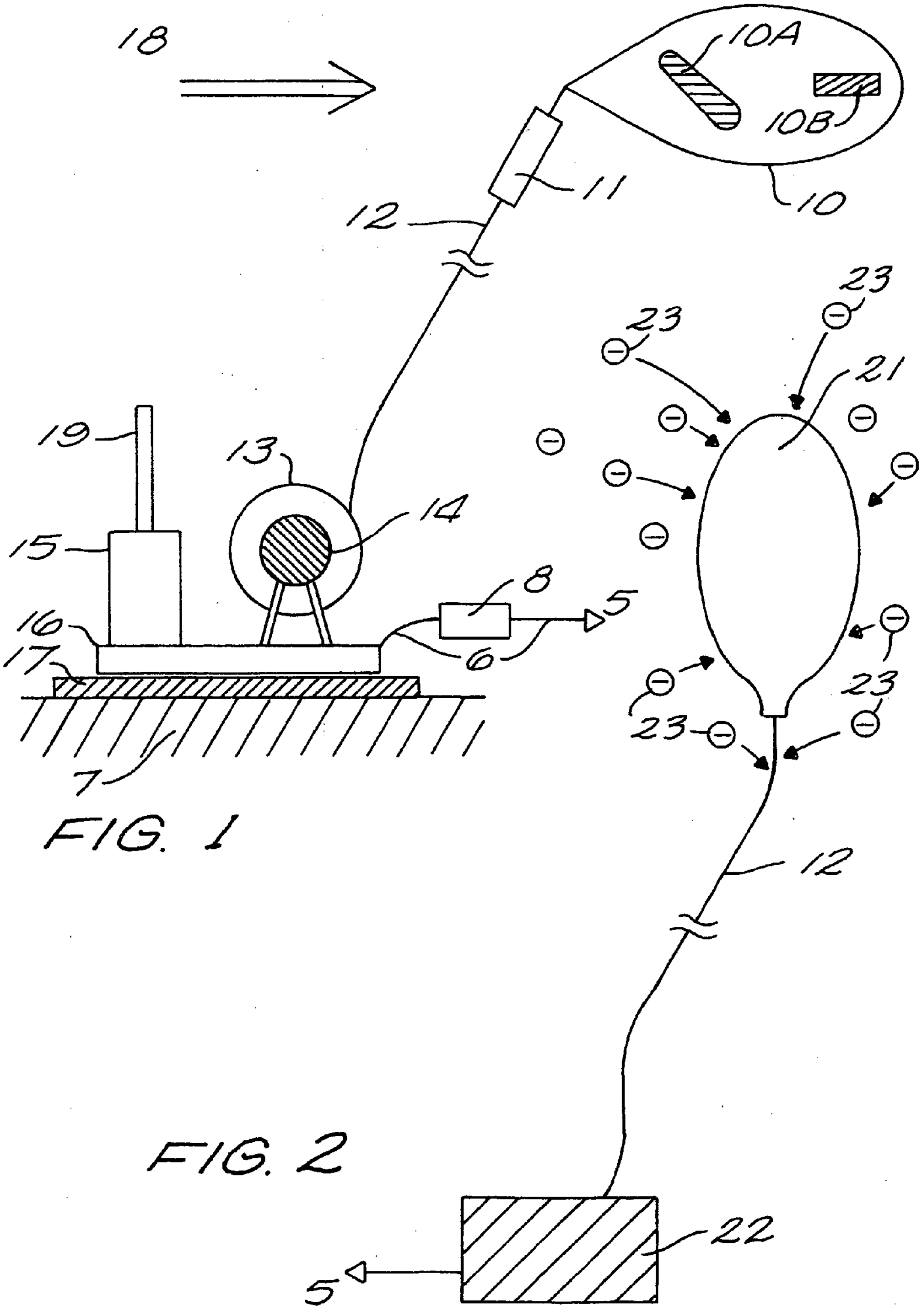
18
10A
10B
10
11
12
19
13
14
15
8
5
16
6
17
7
23
23
21
23
23
23-
12
22
5

FIG. 1

FIG. 2

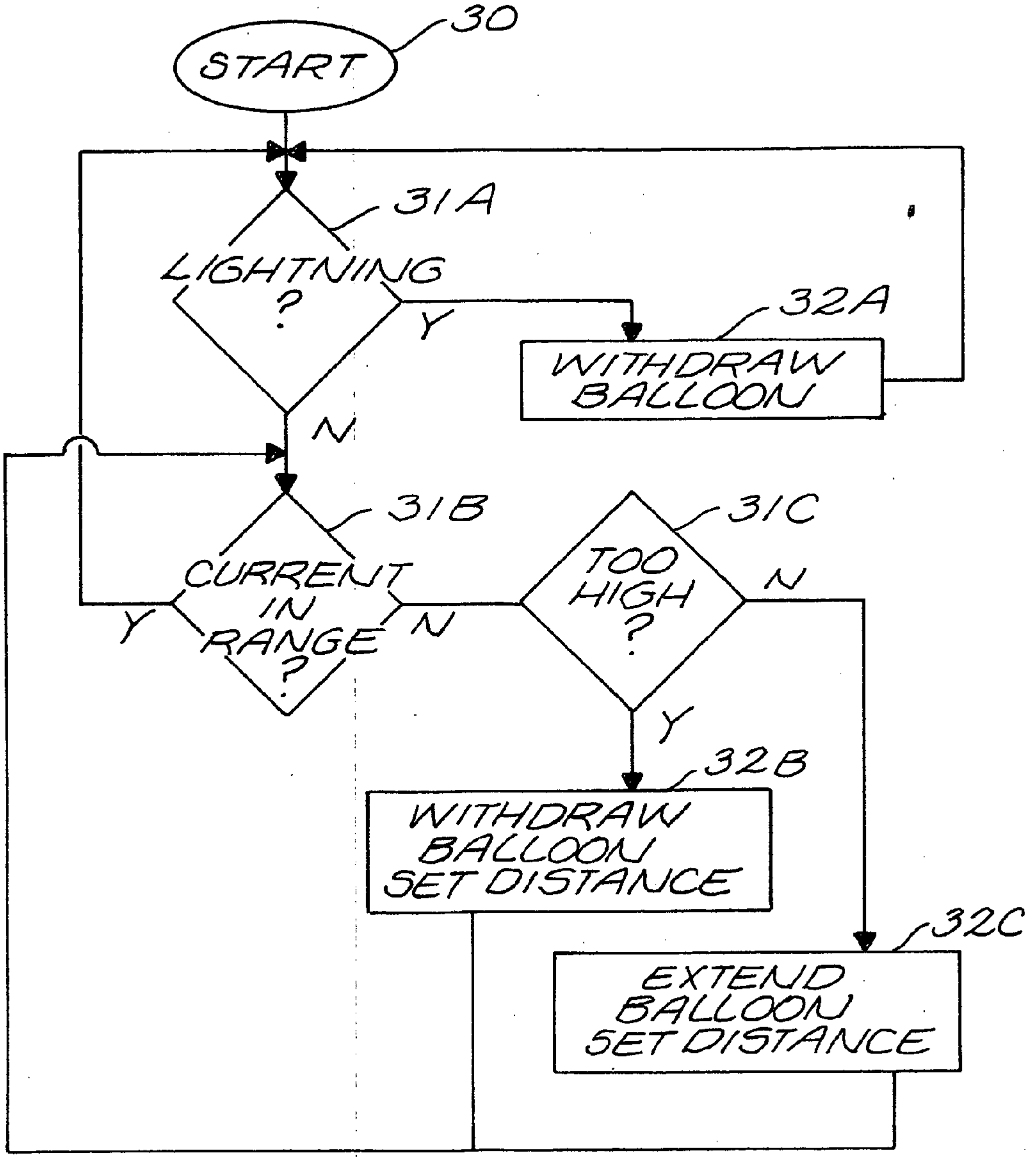
30
START
31A
LIGHTNING ?
Y
N
32A
WITHDRAW BALLOON
31B
CURRENT IN RANGE ?
Y
N
31C
TOO HIGH ?
N
Y
32B
WITHDRAW BALLOON SET DISTANCE
32C
EXTEND BALLOON SET DISTANCE

FIG. 3

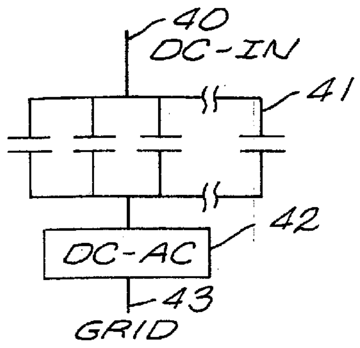
40
DC-IN
41
42
DC-AC
43
GRID

FIG. 4A

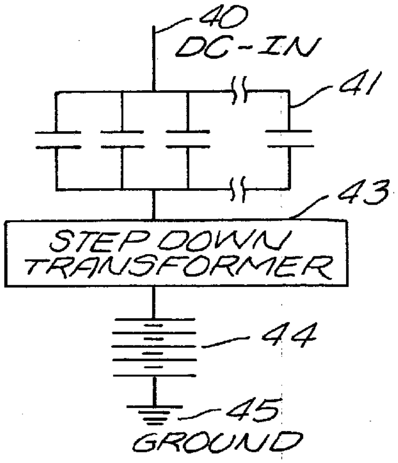
40
DC-IN
41
43
STEP DOWN TRANSFORMER
44
45
GROUND

FIG. 4B

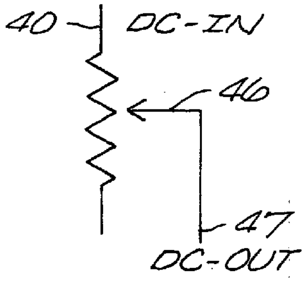
40
DC-IN
46
47
DC-OUT

FIG. 4C

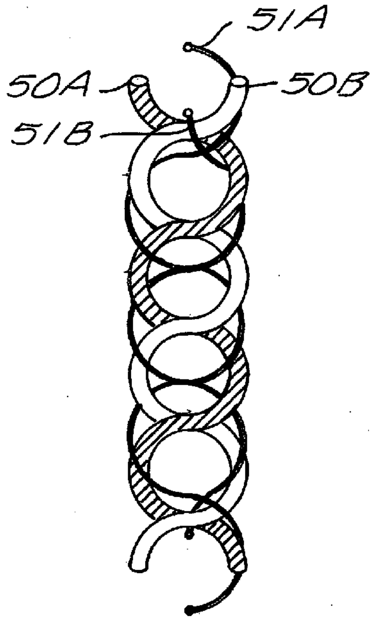
51A
50A
50B
51B

FIG. 5

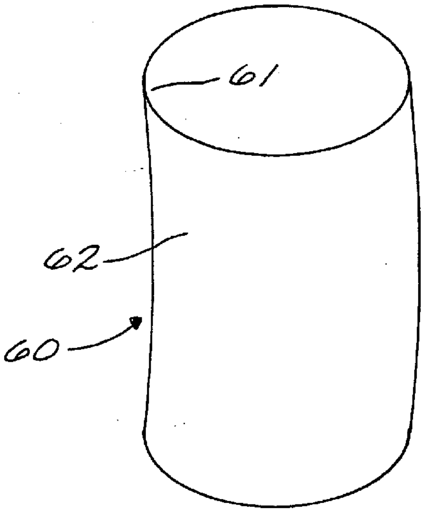

FIG. 6A

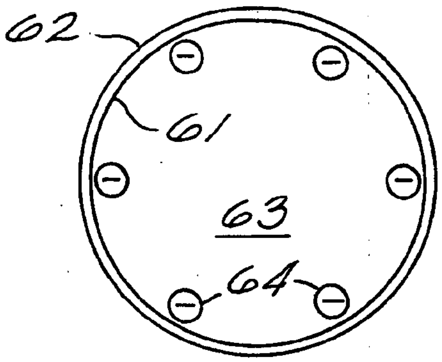

FIG. 6B

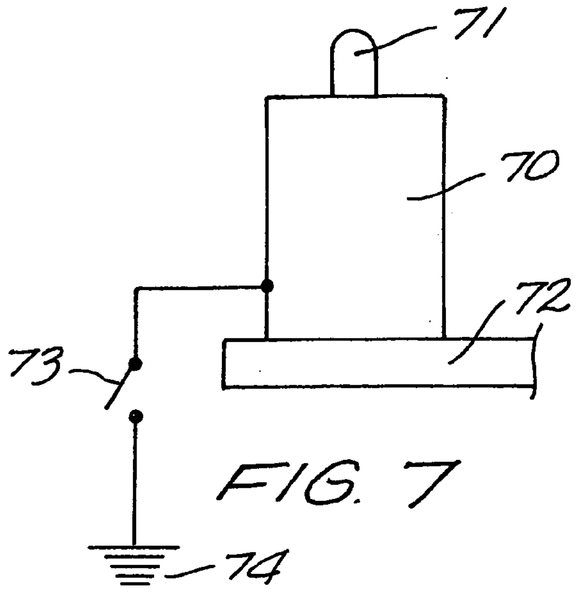
71
70
72
73
74

FIG. 7

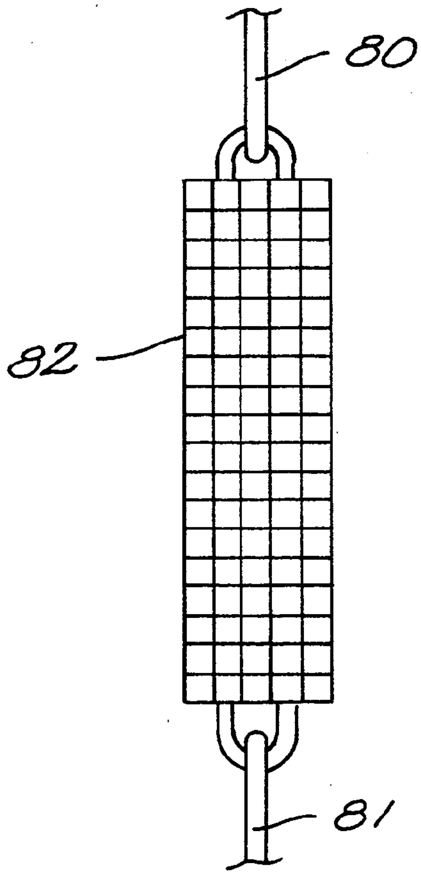

FIG. 8A

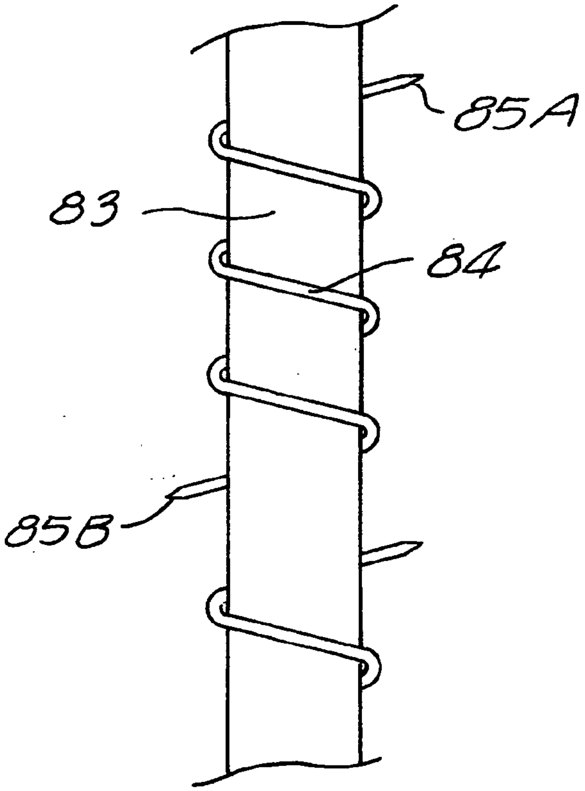
85A
83
84
85B

FIG. 8B

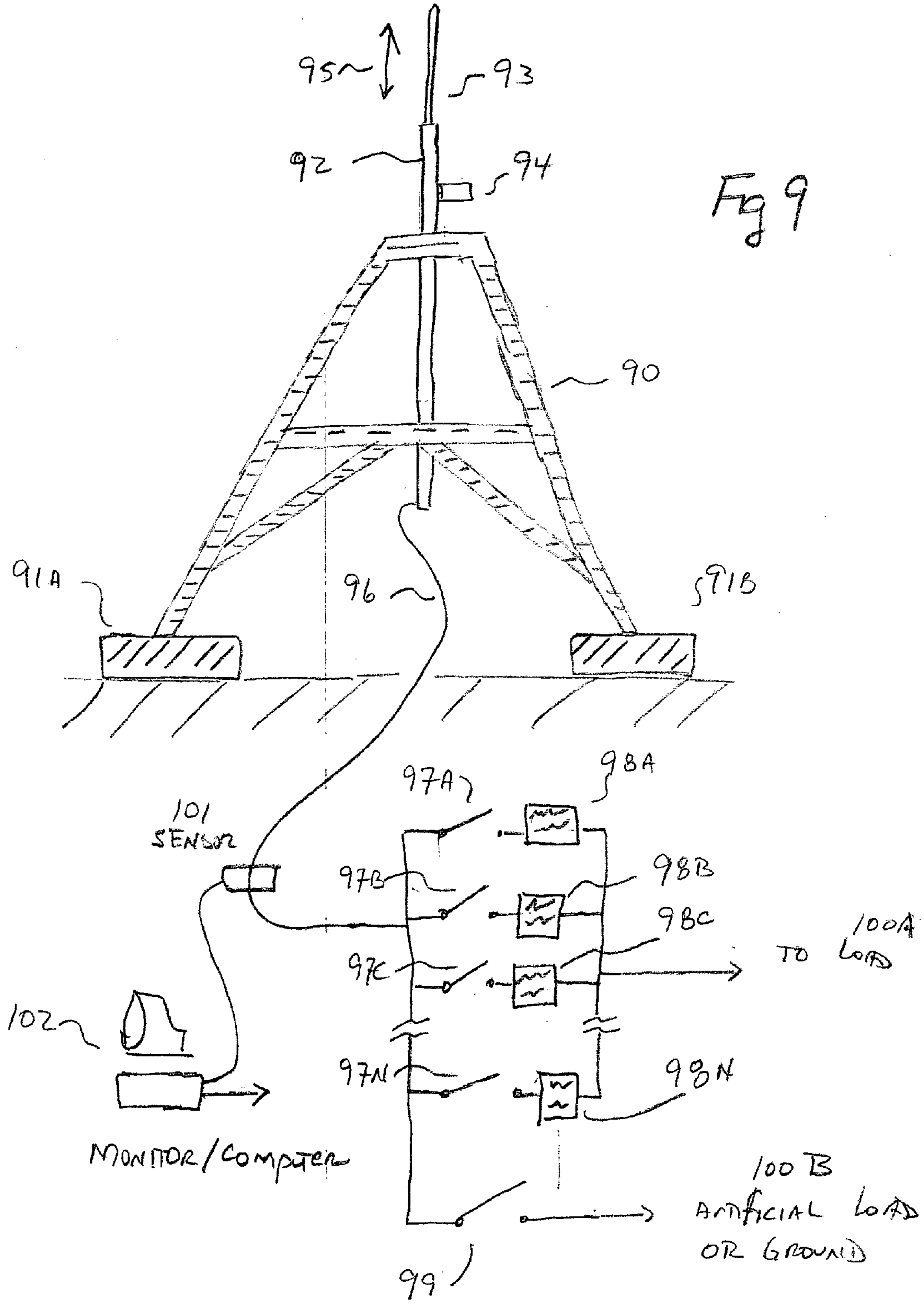
Fig 9
SENSOR
TO LOAD
97N
98N
MONITOR/COMPUTER
ARTIFICIAL LOAD
OR GROUND

DYNAMIC ELECTRICAL CONVERTER SYSTEM

BACKGROUND OF THE INVENTION

[0001] This is a continuation in part of U.S. patent application Ser. No. 12/218,297, filed on Jul. 14, 2008, and entitled, "Atmospheric Electrical Generator".

[0002] This invention relates to the production of electrical energy and more particularly to the a conversion system for variable electrical voltages.

[0003] Everyone is familiar with Benjamin Franklin's kite experiment of 1752. Using a kite whose string had become wet, negative charges from the passing clouds flowed into the string, down to the suspended key, and then into a Leyden jar via a thin metal wire. Franklin was protected by a dry silk string; but, when Franklin's knuckle came too close to the key, he received a strong shock. Fortunately, Benjamin Franklin was not killed, others who tried this same experiment were not so lucky.

[0004] Since then, the formation of lightning has remained something of a mystery. Lightning bolts are triggered when a negatively charged cloud base induces a positive charge from the ground, thereby forming a "pathway" for the discharge of the collected electrical energy.

[0005] Lightning travels up to 60,000 miles per hour with a flash that is brighter than ten million 100-watt lightbulbs. This wattage is as much power as is produced by all of the electricity plants in the United States and with a voltage of up to 300 million volts.

[0006] It is this very fact, the power within lightning is immense, that has prevented any successful collection of the electrical energy from lightning. The electricity in lightning is far too extreme for current technology to harness.

[0007] While lightning has attracted a energy starved industrial world, no one has developed any technique to harness this naturally occurring electrical source.

[0008] It is clear there is a continuing need for an electrical source other than carbon-based fuels and that the naturally occurring electricity in the atmosphere is being ignored.

SUMMARY OF THE INVENTION

[0009] The invention is a mechanism which taps into the naturally occurring static electricity in the atmosphere. Whereas heretofore, the attempt to garner electricity from the atmosphere has focused exclusively on capturing lightning, the present invention syphons off the static electricity which is generated from any agitated air and voids lightning.

[0010] Lightning is only the final discharge of the static electricity, whether that lightning is intra-cloud lightning, cloud-to-ground lightning, or inter-cloud lightning. Other types of final discharges are known as heat lightning, summer lightning, sheet lightning, ribbon lightning, silent lightning, ball lightning, bead lightning, elves, jets, and sprites. Well before these discharges are observed, as the atmosphere becomes agitated by wind or thermal, static electricity is being generated.

[0011] The present invention recognizes that this static electricity is being formed and creates a mechanism to capture it.

[0012] The mechanism of this invention utilizes an aircraft such as a lighter than air balloon. While the preferred embodiment uses a foil balloon, a variety of other aircraft are obvious to those of ordinary skill in the art, including, but not limited to: gliders, rubber balloons (such as weather balloons), biaxially-oriented polyethylene terephthalate polyester film balloons, and latex balloons.

[0013] Within this discussion, the balloon is referenced, but, the invention is not intended to be limited solely to balloons.

[0014] The balloon is sent aloft and is tethered by a conductive line. In this context, the conductive line may be any obvious to those of ordinary skill in the art. For the preferred embodiment, the conductive line is a generically referred to as a "poly-rope" and is commercially available through a variety of sources. A suitable conductive line is described in U.S. Pat. No. 5,203,542, entitled "Apparatus for Improved Electric Fence Wire Construction for use with Intensive Grazing" issued Apr. 20, 1993, to Coley, et al. and incorporated hereinto by reference.

[0015] The conductive line is played out of a winch to control the altitude of the balloon. The motor controlling the winch is able to reverse direction to both extend and withdraw the conductive line which is wrapped around a spool on the winch. The winch/spool combination are part of a base unit.

[0016] In some embodiments of the invention, the spool is constructed of rubber so as to insulate the conductive line from the winch assembly. In this embodiment, only the conductive line is charged by the atmospheric static electricity while the winch remains neutral.

[0017] In yet another embodiment, the winch/spool are part of a base unit which is itself isolated from the ground by an insulator. In this embodiment, the entire base unit is charged by the atmospheric static electricity.

[0018] A conductor, such as an insulated wire, is electrically connected to the conductive line. In one embodiment, where the conductive line is electrically isolated from the spool and winch motor, the conductor is connected to the conductive line. In the embodiment where the conductive line is electrically connected to the base unit, then the conductor is connected anywhere on a metalic base unit.

[0019] The other end of the conductor is connected to a load. The load in this case can be any of a variety of electrical loads well known to those of ordinary skill in art, including, but not limited to a motor, a battery system, or the electrical grid for the system.

[0020] In the preferred embodiment, a sensor array is used to monitor the activities both at the base unit (such as electrical flow within the conductor) and in the surrounding locale.

[0021] A sensor monitoring the electrical flow (i.e. voltage and/or current) within the conductor is used to monitor the electrical activity within the conductor.

[0022] In the preferred embodiment, a lightning sensor monitors for lightning activity within the locale. As noted earlier, the electrical characteristic of lightning is so extreme that ideally this discharge is avoided as it might damage the mechanism of this invention.

[0023] The sensor array is utilized by a controller, such as microprocessor, programmed to operate the mechanism as outlined herein.

[0024] The controller operates the winch motor to extend or withdraw the conductive line and by extension the altitude of the balloon. The controller is programmed to operate the winch by monitoring the electrical characteristics of the conductor and adjusting the balloon's altitude to maintain these characteristics within the conductor within a preset range.

[0025] This preset range is established either in the base programming of the controller or is established by an operator of the system.

[0026] As example, by controlling the amount of current being withdrawn from the atmosphere, the mechanism operates within a safe range and also provides a relatively stable current flow from which a variety of activities can take place (such as DC-AC conversion).

[0027] The controller also utilizes the lightning sensor to protect the mechanism from a lightning strike. Should lightning be detected within a pre-determined range (as established by the software or defined by an operator), then the balloon is pulled down to minimize the risk of damage from a lightning strike.

[0028] Another aspect of the invention relates to the electrical system which accepts the fluctuating atmospheric charge and changes it into an acceptable configuration for either the desired load or for the existing power grid.

[0029] Power grids in the United States operate with a frequency of 60 hertz in an alternating current arrangement. While this basic configuration seems to be universally accepted, the voltage within the grid varies dramatically, such as 15 kv, 34 kv, 69 kv, and even 112 kv.

[0030] Each atmospheric generator is placed proximate to or within easy access to a specific grid; this establishes the required electrical output configuration (i.e. that which is accepted by the power grid). As example, one of the atmospheric electrical collector units as described above collects the atmospheric electrical power as direct current and then supplies the appropriate power grid a specific flow (as example, AC, 60 hertz, at 69 kv).

[0031] The difficulty lies the fact that the DC current being garnered from the atmosphere varies depending on the actual agitation being generated in the atmosphere. This means that the source of DC current is fluctuating.

[0032] The present invention uses a monitoring system which checks the input DC voltage. Depending on the actual voltage being received, the appropriate converter is connected to the input DC voltage so that the desired output is obtained.

[0033] As example, suppose the DC input voltage is 1500 volts, the monitoring system, sensing this input, closes the switch connecting the DC voltage to a converter which accepts DC voltage in the range of 1000-2000 volts which then delivers an AC, 60 hertz 69 kv signal to the power grid. If the DC input voltage increases to 2100 volts, then the monitoring system opens the switch to the first converter (1000-2000 volts) and closes the switch to a second converter (such as 2000-4000 volts) to deliver the desired output of AC (60 hertz, 69 kv) for the power grid.

[0034] In this manner, regardless of the fluctuating input DC voltage, the electrical grid is supplied with a fully configured electrical input conforming to the needs of that specific electrical grid.

[0035] Another aspect of the present invention is the use of a tower or permanent structure instead of an aircraft. In this embodiment, the building or tower is electrically isolated from the ground and a rod (similar to a lightning rod) is extended into the atmosphere. The rod collects the atmospheric charge which is conveyed via an electrical conduit (ideally insulated) where the collected DC charge is reconfigured to meets the need of the locale.

[0036] In this context, for one embodiment of the invention, a tower is placed onto the top of a building. The tower is electrically isolated from the building using such mechanisms well known to those of ordinary skill in the art such as rubber mats. A rod ideally extends from the top of the tower to facilitate the collection of the DC electrical energy.

[0037] A converter mechanism such as described above, is connected to the tower to flow the DC electricity to a converter which modifies the DC current for the specific application. In one application, the DC current is converted to the electrical needs of the building, thereby providing at least some of the electrical requirements of the building itself.

[0038] As noted earlier, the dynamic converter system of the present invention allows a power generator to address a variable voltage in an efficient manner. This makes the dynamic converter system ideal for a variety of alternative energy sources such as the above described atmospheric electrical generator and other alternative energy sources such as wind and wave powered systems. In these systems, the energy being generated must be converted to a proper electrical configuration for a identified load. This may be a particular motor or connection to the power grid which act as a load to the power generating mechanism.

[0039] For these energy generating systems, the converter assembly of this invention utilizes multiple converters. Each converter is configured to accept a unique range of voltages and from these voltages, create the desired electrical configuration. By using multiple converters, a full range is available, from a minimum voltage input to a maximum voltage input.

[0040] As example, suppose the output voltage needs to be 55 kv AC, but the input voltage usually ranges from 20 kv DC to 100 kv DC. By using four different converts, this 80 kv DC range is broken into:

[0041] 1) converter A accepting 20 kv DC to 40 kv DC and converting it to 55 kv AC;

[0042] 2) converter B accepting 40 kv DC to 60 kv DC and converting it to 55 kv AC;

[0043] 3) converter C accepting 60 kv DC to 80 kv DC and converting it to 55 kv AC; and,

[0044] 4) converter A accepting 80 kv DC to 100 kv DC and converting it to 55 kv AC.

[0045] Those of ordinary skill in the art readily recognize a variety of other arrangement.

[0046] A monitor checks the incoming voltage recognizes this voltage as being addressed by a particular converter, and, by controlling switches going to each of the converters directs the input voltage appropriately. Carrying on with the above example, assume the monitor determines that the incoming voltage is 56 kv DC, then converter B is activated by selectively closing the switch associated with converter B and keeping the switches for the other converters open.

[0047] In this manner, only the appropriate converter is used for whatever the incoming voltage is.

[0048] For the situation of wind power, this permits the wind power generator to operate even in “light” winds and “heavy” winds when the amount of electricity being generated varies significantly from the desired levels. With the increased operating capability, the wind power generator is able to provide a much larger overall energy flow.

[0049] The invention, together with various embodiments thereof will be more fully explained by the following description of the accompanying drawings.

DRAWINGS IN BRIEF

[0050] FIG. **1** diagrams the preferred embodiment of the invention.

[0051] FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

[0052] FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

[0053] FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

[0054] FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

[0055] FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

[0056] FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

[0057] FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

[0058] FIG. **9** diagrams the electrical converter system being employed on a stationary structure and the dynamic converter system.

DRAWINGS IN DETAIL

[0059] FIG. **1** diagrams the preferred embodiment of the invention.

[0060] Balloon **10** is an aircraft which, in this illustration, is a lighter than air balloon. Wings **10**A, extending from the body of balloon **10**, provide additional lift in air flow **18**. Tail **10**B helps to stabilize balloon **10**.

[0061] Balloon **10** is tethered to the ground via conductive line **12**. As noted earlier, a variety of configurations and materials are available to serve as conductive line **12**. In this illustration, a poly-wire is used. Poly-wire is commercially available through a variety of vendors, including, but not limited to: Jeffers Livestock and Sareba Systems, Inc. of Ellendale, Minn.

[0062] In this embodiment, located proximate to balloon **10**, is an electrical collection enhancement lead **11** which assists in the collection of the static electrical charge in the atmosphere. Electrical collection enhancement lead **11** is configured to attract the static charge and conduct the charge into the conductive line **12**.

[0063] The electricity flows down the conductive line into spool **13**, where the conductive line **12** is collected and either withdrawn or dispensed through operation of winch motor **14**.

[0064] Winch motor **14** and spool **13** are mounted onto base unit **16** which is electrically isolated from ground **7** using insulator **17**. Note, in this embodiment of the invention, when electricity is being collected from the atmosphere, the entire base unit **16** becomes charged. In another embodiment of the invention, spool **13** is constructed of rubber, thereby preventing base unit **16** from becoming charged, thereby restricting the charging from the atmosphere to only conductive line **12**.

[0065] In this embodiment, conductor **6** is connected to base unit **16** (since the entire base unit **16** is charged and the base unit is metalic) to communicate the electrical current to load **5**. Conductor **6** is ideally an insulated wire.

[0066] The electrical current through conductor **6** is measured using sensor **8**.

[0067] In the alternative embodiment discussed above, where only the conductive line **12** is charged, then conductor **6** is connected to conductive line **12**.

[0068] Controller **15**, located in this embodiment on base unit **16**, operates winch motor **14** in response to signals from sensor **8** (measuring the current being discharged to load **5**) to maintain the current flow within a pre-defined range. As the current flow diminishes, then the conductive line **12** extended from spool **13** to increase the altitude of balloon **10** to that more static charge from the atmosphere is gathered; as the current flow falls exceeds a preset level, conductive line **12** is withdrawn onto spool **13** to decrease the static charge being collected from the atmosphere.

[0069] The range of current flow through conductor **6** is ideally set by the program, although some embodiments of the invention permit an operator to establish this range of operation.

[0070] In an alternative embodiment, the sensor monitoring conductor **6** monitors the voltage therein.

[0071] In the preferred embodiment of the invention, controller **15** is also equipped with a lightning sensor **19**. In this embodiment, when lightning is sensed within a preset range, then substantially all of conductive line **12** is wound onto spool **13** to pull balloon **10** near the ground and protect the entire mechanism from being damaged from a lightning discharge.

[0072] In the preferred embodiment, the "safe" distance form lightning is set in the programming of controller **15** and is ideally two miles; other embodiments permit the operator to "safe" distance.

[0073] There are a variety of lightning sensors well known to those of ordinary skill in the art, including, but not limited to those described in: U.S. Pat. No. 7,016,785, entitled "Lightning Detection" issued to Makela, et al. on Mar. 21, 2006; U.S. Pat. No. 6,829,911, entitled "Lightning Detection and Prediction Alarm Device" issued to Jones, et al. on Dec. 7, 2004; U.S. Pat. No. 7,200,418, entitled "Detection of Lightning" issued to Karikuranta, et al. on Apr. 3, 2007; and U.S. Pat. No. 6,961,662, entitled "Systems and Methods for Spectral Corrected Lightning Detection" issued to Murphy on Nov. 1, 2005; all of which are incorporated hereinto by reference.

[0074] In another embodiment of the invention, controller **15** is not located on base unit **16**, rather it is remote and communicates its control signals to winch motor **14** using radio waves.

[0075] FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

[0076] Static charges **23** are generated in the atmosphere by agitated air. These static charges are often collected at the bottom of clouds, but exist in other environments as well.

[0077] Balloon **21** is extended into this strata of static charges **23** which are then attracted to conductive line **12** to flow to base unit **22** and then onto load **5**.

[0078] By increasing or decreasing the altitude of balloon **21** (defined by the length of the extended conductive line **12**), conductive line **12** is selectively exposed to varying densities and levels of the static charge strata, and by extension, the current flow or voltage is increased or decreased.

[0079] FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

[0080] Once the program starts **30**, the lightning sensor is checked to determine if lightning has occurred within the unsafe range **31**A, if it has, then the balloon is lowered **32**A, and the program continues monitoring the status of lightning until no lightning is detected.

[0081] When the lightning status is acceptable, then the current within the conductor is checked to see if the current is within the prescribed range **31**B. If the current is acceptable (within range) the program returns to check the lightning status **31**A; otherwise a determination is made to see if the current is above the prescribed range **31**C.

[0082] If the current is above the prescribed range, then the altitude of the balloon is withdrawn a set amount **32**B (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

[0083] If the current is below the prescribed range, then the altitude of the balloon is extended a set amount **32**C (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

[0084] In this manner of feed-back and minor adjustments in the altitude of the balloon, the current is maintained within a prescribed range which can be handled by the downstream electrical system.

[0085] As noted earlier, some embodiments of the invention monitor the voltage instead of the current.

[0086] FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

[0087] By maintaining the voltage being collected in a prescribed range, an electrical conversion system is easily designed. While FIGS. **4**A, **4**B, and **4**C illustrate some electrical configurations, those of ordinary skill in the art readily recognize a variety of other configurations which will serve the same function.

[0088] Referencing FIG. **4**A, Direct Current In (DC IN) **40** is buffered by a gang of capacitors **41** before being communicated to a DC/AC converter **42**. The DC/AC converter converts the direct current into a an alternating current suitable for placement over an existing electrical grid **43** such as normally found from a power-plant.

[0089] Those of ordinary skill in the art readily recognize a variety of DC/AC converters, including, but not limited to: U.S. Pat. No. 7,394,671, entitled "Controller IC, DC-AC Conversion Apparatus, and parallel running system of DC-AC Conversion Apparatuses" issued to Fukumoto, et al. on Jul. 1, 2008; and, U.S. Pat. No. 7,330,366, entitled "DC-AC Converter" issued to Lee, et al. on Feb. 12, 2008; all of which are incorporated hereinto by reference.

[0090] FIG. **4**B illustrates an electrical arrangement suitable for use in charging a battery. DC IN **40** is buffered by capacitor bank **41** before entering into a step down transformer **43**. Step down transformer **43** reduces the voltage so that the voltage can safely be introduced into battery **44** which is connected to ground **45** at the battery's other pole.

[0091] Those of ordinary skill in the art readily recognize a variety of batteries which will work in this capacity, including, but not limited to those described in: U.S. Pat. No. 7,378,181, entitled "Electric Storage Battery Construction and Manufacture" issued to Skinlo on May 27, 2008; U.S. Pat. No. 7,388,350, entitled "Battery with Electronic Compartment" issued to Wright on Jun. 17, 2008; U.S. Pat. No. 7,397,220, entitled "Connection Member and Battery Pack" issued to Uchida, et al. on Jul. 8, 2008; and, U.S. Pat. No. 7,375,492, entitled "Inductively Charged Battery Pack" issued to Calhoon, et al. on May 20, 2008; all of which are incorporated hereinto by reference.

[0092] In FIG. **4**C, DC IN **40** is fed into an adjustable rheostat **46** which is controlled by the controller so that the DC OUT **47** falls within a specified range.

[0093] FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

[0094] This type of conductive line is commonly called poly-wire and consists of multiple interwoven strands of plastic **50**A and **50**B woven into a cord or rope arrangement having intertwined therein exposed metal wires **51**A and **51**B. While this illustration shows two plastic strands and two metal wires, any number of possible combinations is possible.

[0095] The exposed metal wires **51**A and **51**B attract the atmospheric static charge and transmit the charge down to the base unit (not shown).

[0096] FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

[0097] This conductive line utilizes a tube **60** having an outer layer **62** of PET Film (Biaxially-oriented polyethylene terephtalate polyester film) which provides exceptionally high tensile strength and is chemically and dimensionally stable. The tube has an ideal diameter of between two and three inches.

[0098] An interior metal coating **61** provides an initial conduit for the flow of static charge. The static charge through the metal forces the tube to expand due to the repulsion experienced by like charges. Further, the flow of electricity causes the interior of the tube **60** to become ionized to provide an additional pathway for the atmospheric static charges to the base unit (not shown).

[0099] Because outer layer **62** provides a gas barrier, the resulting ionization is not dissipated by air currents, thereby providing a highly stable pathway.

[0100] FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

[0101] In this embodiment, controller box **70**, resting on insulating pad **72**, is in communication with the sensors as described above. Using the input from these sensors, when there is flow of electricity through the base unit, warning flashing light **71** is illuminated. To electrically neutralize the mechanism, switch **73** is activated to pass any existing current into the ground **74**.

[0102] FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

[0103] Referencing FIG. **8**A, enhanced electrical collection lead **82** is a wire mesh which is in electrical communication with conductive line **81** and balloon **80**. Because of the significant amount of metal exposed by enhanced electrical collection lead **82**, more static electricity from the atmosphere is drawn to the collection lead **82**, and then down conductive line **81** to the base unit (not shown).

[0104] Conductive lead **82** is positioned proximate to balloon **80**.

[0105] In FIG. **4**B, poly-wire **83** has enhanced electrical collection leads **84** wrapped therearound. Collection leads **84** have pointed ends **85**A and **85**B which have a propensity to attract more electricity than rounded ends do.

[0106] FIG. **9** diagrams the electrical converter system being employed on a stationary structure and the dynamic converter system.

[0107] This illustration uses a structure for the collection of the atmospheric energy. The circuit described in this illustration is not to be construed only to this energy collection mechanism, but, is adapted work with any number of energy collection/generation mechanisms, including, but not limited to: solar collectors; wind power generators; wave motion power generators; and, the atmospheric energy collection mechanisms described herein.

[0108] The structure of this embodiment is tower **90** which is electrically isolated from the ground via insulators **91**A and **91**B. In this manner, any atmospheric energy collected is maintained within tower **90**.

[0109] Conduit **90** contains antenna **93** which is moved up/down as illustrated by arrow **95** by motor **94**. Motor **94** is controlled by monitor/computer **102** based upon the electrical energy being collected by antenna **93** and communicated via insulated conductor **96**, monitored by sensor **101**.

[0110] The collected energy is communicated to a bank of switches (**97**A, **97**B, **97**C, through **97**N, N being any number of switches). In this manner, each switch receives the electrical energy being gathered by antenna **93**. As noted earlier, this electrical energy varies depending on the ambient conditions.

[0111] Each switch addresses a single converter. As illustrated, switch **97**A, when closed, communicates the electrical energy to converter **98**A; switch **97**B to converter **98**B; switch **97**C to converter **98**C; and so on until switch **97**N to converter **98**N.

[0112] Each converter is designed to accept a range of electrical input and generate a uniform electrical configuration to load **100**A.

[0113] Monitor/computer **102** monitors the varying electrical input from antenna **93** via sensor **101**. Based upon the electrical input, monitor/computer **102** closes a selected one of the switches (and opens all of the rest) so that the voltage being received is directed to the converter designed to handle that voltage.

[0114] In some situations, the voltage is so large that (or during emergency) that emergency switch **99** must be closed to direct the voltage to an artificial load or ground either during the switching of switches **97**A-**97**N or to "shut down" the system for repairs/maintenance or if lightning is sensed within a "dangerous" range of tower **90**.

[0115] In this context, an artificial load is anything that expends the incoming energy, such as lifting a weight to let it fall once switch **99** is opened, or pumping water out of and then into the same reservoir.

[0116] It is clear from the foregoing that the present invention captures an entirely new source of electrical energy.

What is claimed is:

1. An electrical converter system comprising:
- a) N converters, each converter configured to accept a unique range of voltages and producing a uniform electrical output;
- b) N switches, each switch receiving an input voltage and connected to one of said N converters; and,
- c) a monitor having,
  - 1) a sensor identifying a level of the input voltage, and,
  - 2) in response to said level of the input voltage, selectively closing one of said N switches.

2. The electrical converter system according to claim **1**,
- a) further including a load;
- b) a load switch connecting said input voltage to said load; and,
- c) wherein said monitor closes said load switch prior to closing one of said N switches.

3. The electrical converter system according to claim **2**, wherein said load switch routes said input voltage to ground.

4. The electrical converter system according to claim **2**, wherein said monitor closes said load switch if said input voltage exceeds the unique ranges of voltages for all of said N converters.

5. The electrical converter system according to claim **2**,
- a) further including a lighting sensor generating a signal indicative of a distance to a lightning activity; and,
- b) wherein said monitor closes said load switch when said signal indicative of a distance to a lightning activity is within a prescribed range.

6. The electrical converter system according to claim **2**, wherein said monitor opens said load switch after closing one of said N switches.

7. The electrical converter system according to claim **1**, wherein said input voltage is direct current.

8. The electrical converter system according to claim **1**, wherein said input voltage is alternating current.

9. An electrical system comprising:
- a) a power generating mechanism generating a variable voltage; and,
- b) a power grid electrically connected to consumers of electricity, said power grid operating at a uniform electrical configuration; and,
- c) a conversion system having,
  - 1) N converters, each converter configured to accept a unique range of voltages and producing said uniform electrical configuration to said power grid,
  - 2) N switches, each switch receiving the variable voltage and connected to one of said N converters, and,
  - 3) a monitor having a sensor identifying a level of the variable voltage, and, in response to said level of the variable voltage, selectively closing one of said N switches.

10. The electrical system according to claim **9**, wherein said variable voltage is direct current.

11. The electrical system according to claim **10**, wherein said power generating mechanism is reactive to atmospheric voltage.

12. The electrical system according to claim **11**,
- a) further including a lighting sensor generating a signal indicative of a distance to a lightning activity; and,
- b) wherein said monitor closes said artificial load switch when said signal indicative of a distance to a lightning activity is within a prescribed range.

13. The electrical converter system according to claim **9**, wherein said variable voltage is alternating current.

14. The electrical system according to claim **13**, wherein said power generating mechanism is reactive to wind power.

15. An electrical converter system receiving a variable voltage input and connected to a load requiring a set electrical configuration, said system comprising:
- a) converter assembly having,
  - 1) N converters, each converter configured to accept a unique range of voltages and producing the set electrical configuration, and,
  - 2) N switches, each switch receiving the variable voltage input and connected to one of said N converters;
- b) a sensor identifying a level of the variable voltage input, and,
- c) a monitor, in response to said level of the input voltage, selectively closing one of said N switches.

16. The system according to claim **15**, further including,
- a) a load; and,
- b) a load switch connecting said variable voltage input to said load; and,
- c) wherein said monitor closes said load switch prior to closing one of said N switches.

17. The system according to claim **16**, wherein said load switch routes said variable voltage input to ground.

18. The system according to claim **15**, wherein said monitor closes said load switch if said variable voltage input exceeds the unique range of voltages for all of said N converters.

19. The system according to claim **16**,

a) further including a lighting sensor generating a signal indicative of a distance to a lightning activity; and,

b) wherein said monitor closes said load switch when said signal indicative of a distance to a lightning activity is within a prescribed range.

* * * * *

US008102082B2

(12) **United States Patent**
Ogram

(10) **Patent No.: US 8,102,082 B2**
(45) **Date of Patent: *Jan. 24, 2012**

(54) **ATMOSPHERIC STATIC ELECTRICITY COLLECTOR**

(75) Inventor: **Mark Ellery Ogram**, Tucson, AZ (US)

(73) Assignee: **Sefe, Inc.**, Scottsdale, AZ (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 295 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **12/321,306**

(22) Filed: **Jan. 16, 2009**

(65) **Prior Publication Data**

US 2010/0008011 A1 Jan. 14, 2010

Related U.S. Application Data

(63) Continuation-in-part of application No. 12/218,297, filed on Jul. 14, 2008, now Pat. No. 7,855,476.

(51) **Int. Cl.**
H02G 11/00 (2006.01)

(52) **U.S. Cl.** **307/145**; 307/149; 307/151; 361/218; 361/230

(58) **Field of Classification Search** 361/212, 361/215, 216–218, 230, 231; 307/145, 149, 307/151; 174/2; 244/30, 31
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427 A * 5/1901 Palencsar 307/149
1,014,719 A * 1/1912 Pennock 307/149
1,540,998 A * 6/1925 Plauson 310/309
4,486,669 A * 12/1984 Pugh 290/44

* cited by examiner

Primary Examiner — Fritz M Fleming
(74) *Attorney, Agent, or Firm* — Mintz Levin Cohn Ferris Glovsky and Popeo, P.C.

(57) **ABSTRACT**

An antenna for the collection of atmospheric static electricity in which an electrically conductive hub is suspended from a balloon or blimp via a tether. The hub is either solid or uses a spoke/arm arrangement. A number of rods extend from the hub enhance the collection of atmospheric static electricity. The collected atmospheric electricity is conducted from the rods to an electrical connection where the electricity is conducted to earth via a conductive line.

16 Claims, 7 Drawing Sheets

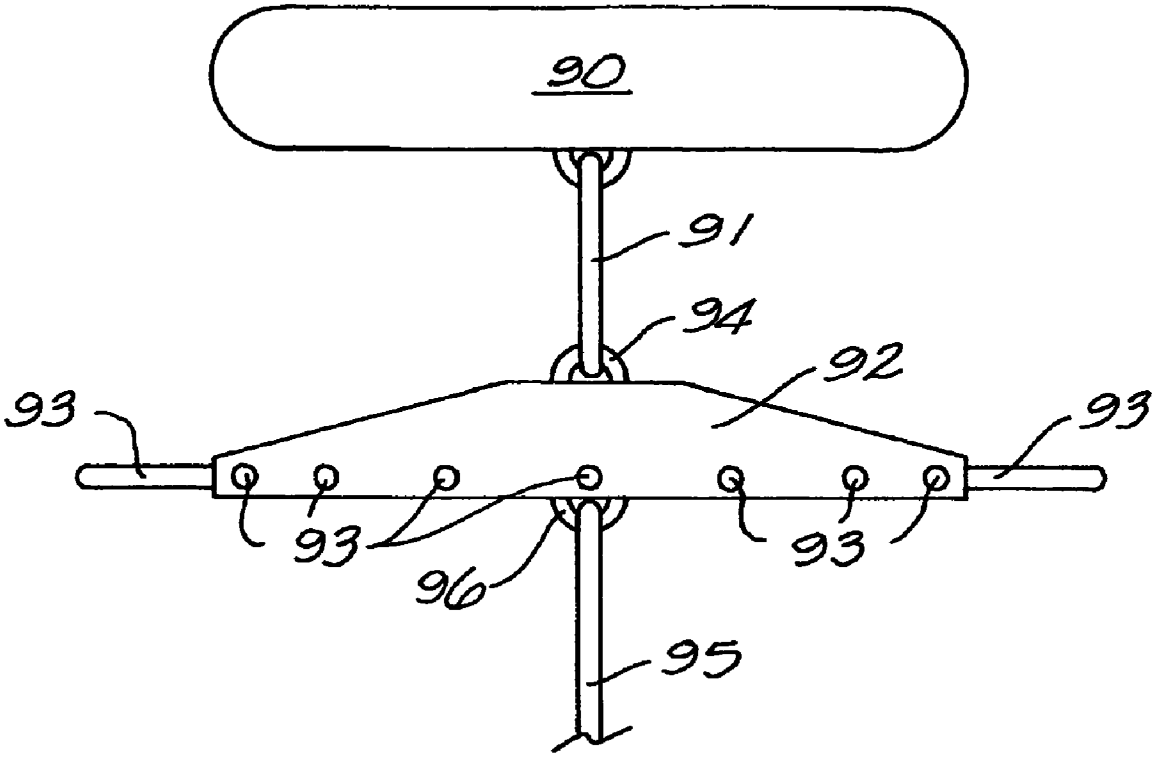

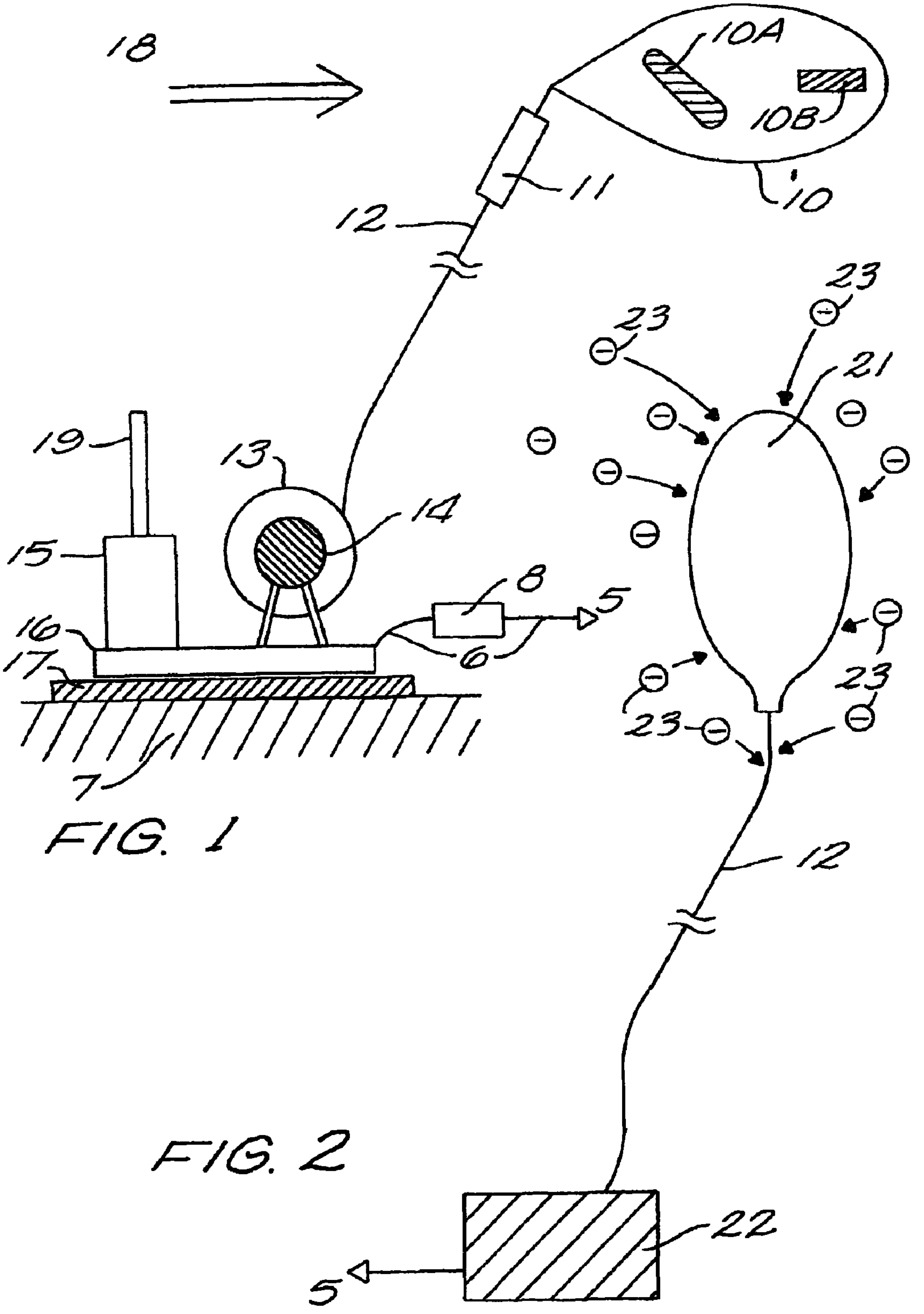
18
10A
10B
10
11
12
19
13
14
15
8
5
16
6
17
7
23
21
23
23
23
12
22
5

FIG. 1

FIG. 2

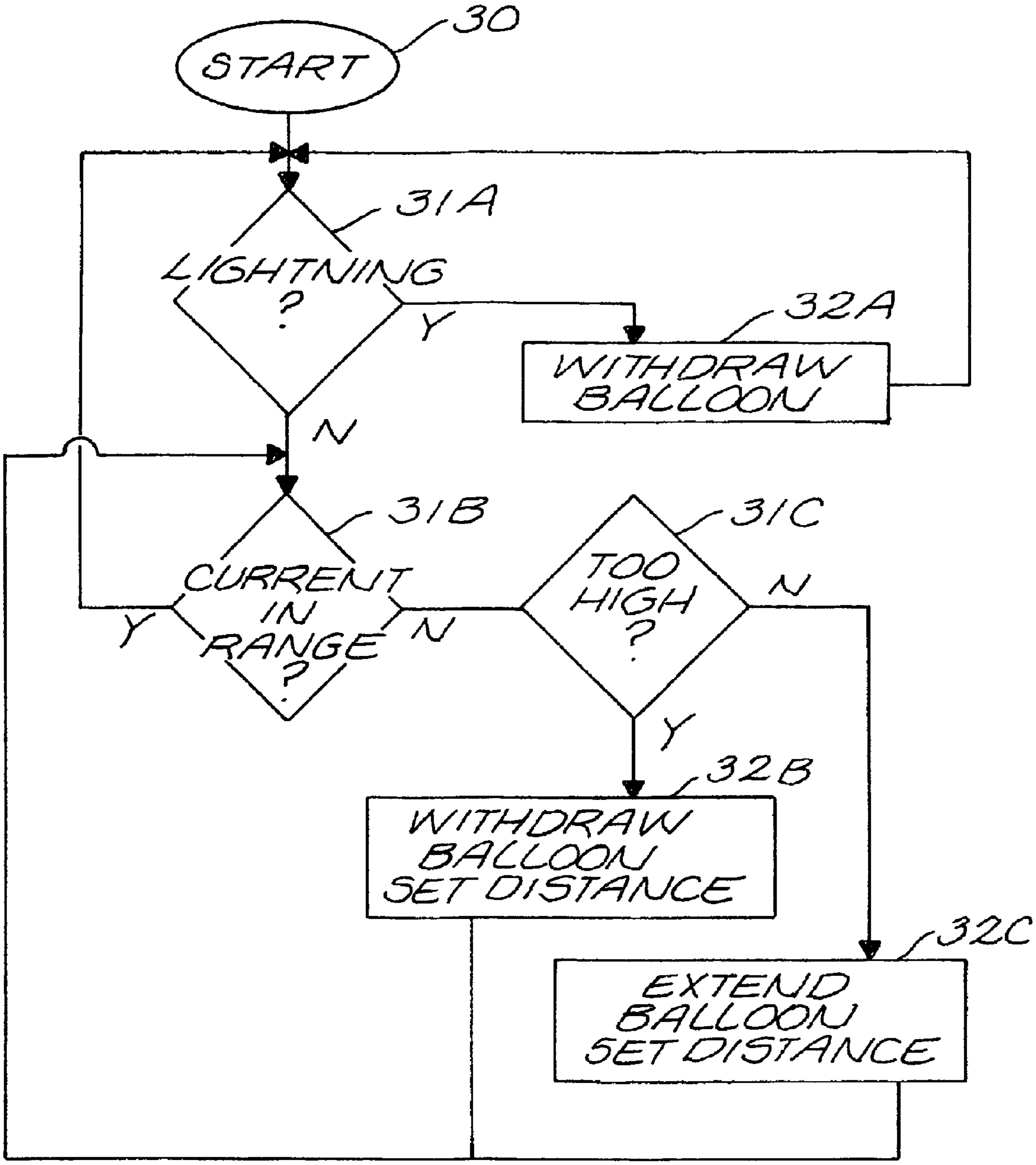
30
START
31A
LIGHTNING ?
Y
32A
WITHDRAW BALLOON
N
31B
CURRENT IN RANGE ?
Y
N
31C
TOO HIGH ?
N
Y
32B
WITHDRAW BALLOON SET DISTANCE
32C
EXTEND BALLOON SET DISTANCE

FIG. 3

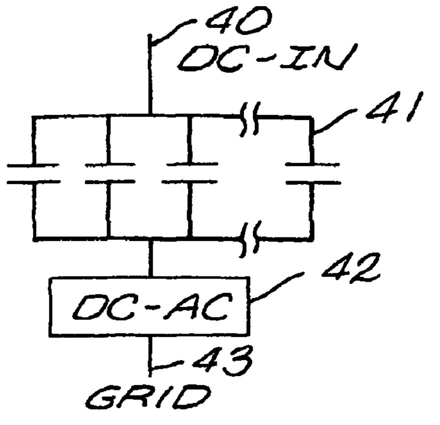
40
DC-IN
41
42
DC-AC
43
GRID

FIG. 4A

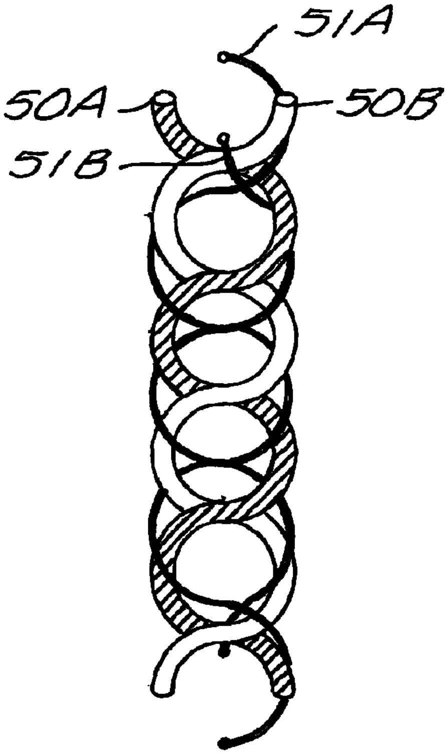
51A
50A
50B
51B

FIG. 5

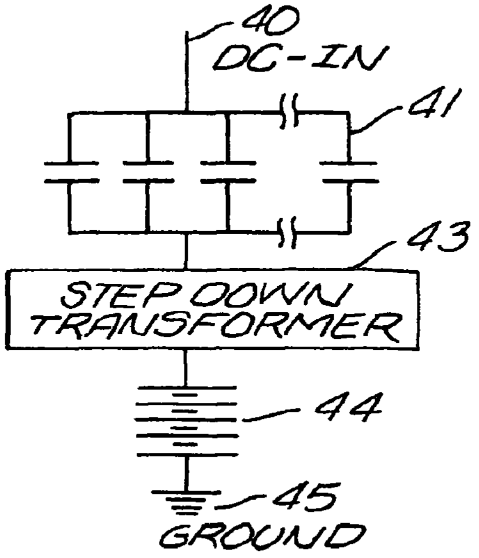
40
DC-IN
41
43
STEP DOWN TRANSFORMER
44
45
GROUND

FIG. 4B

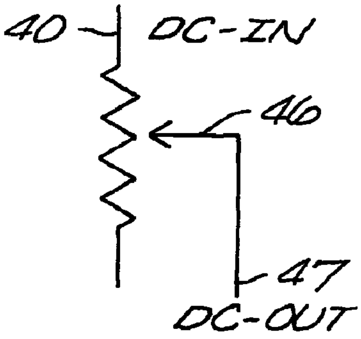
40
DC-IN
46
47
DC-OUT

FIG. 4C

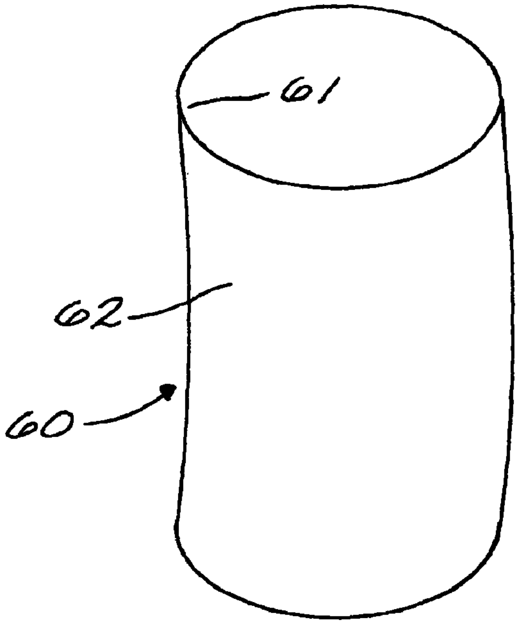

FIG. 6A

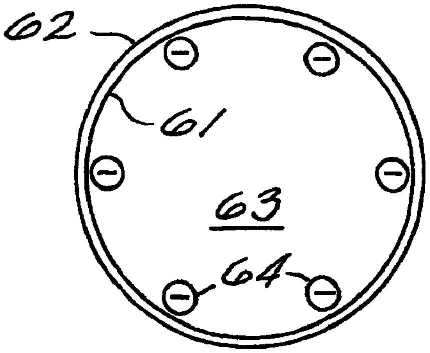

FIG. 6B

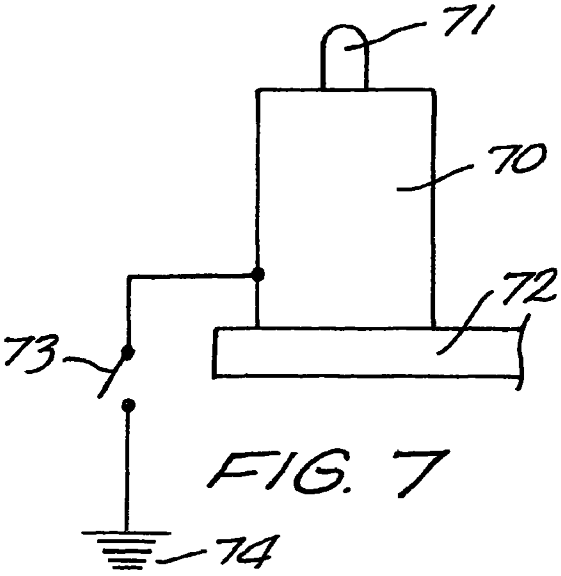
71
70
72
73
74

FIG. 7

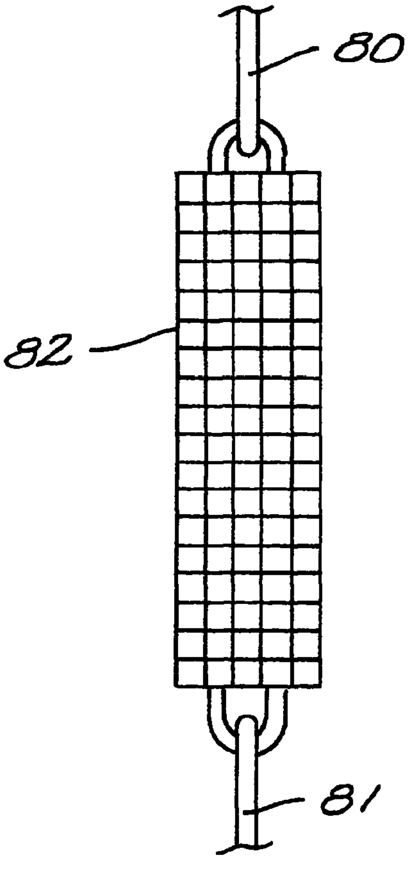

FIG. 8A

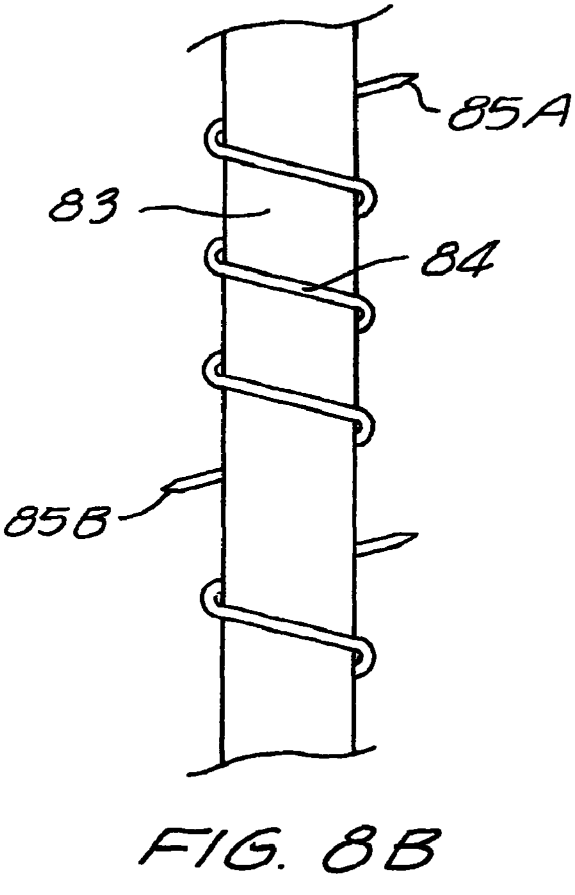
85A
83
84
85B

FIG. 8B

FIG. 9A

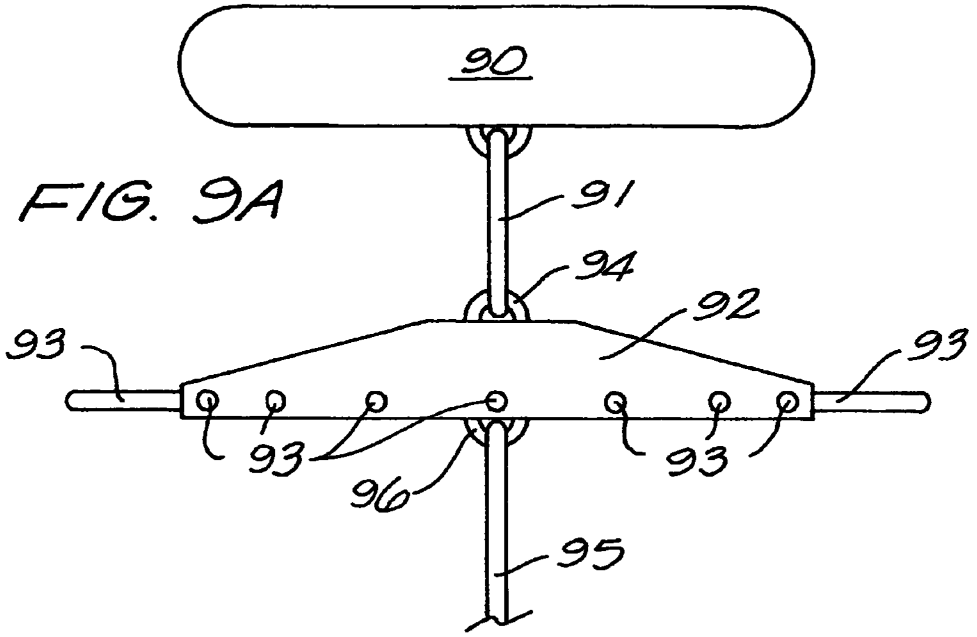
90
91
94
92
93
93
93
93
93
96
95

FIG. 9B

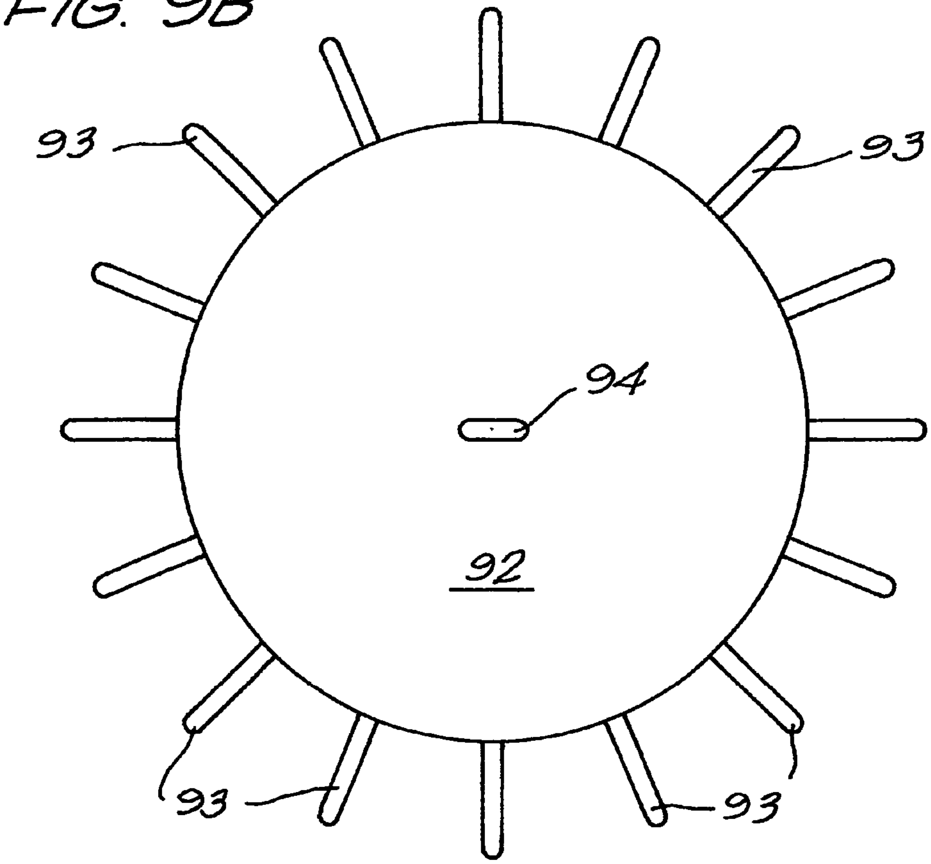
93
93
94
92
93
93

FIG. 10

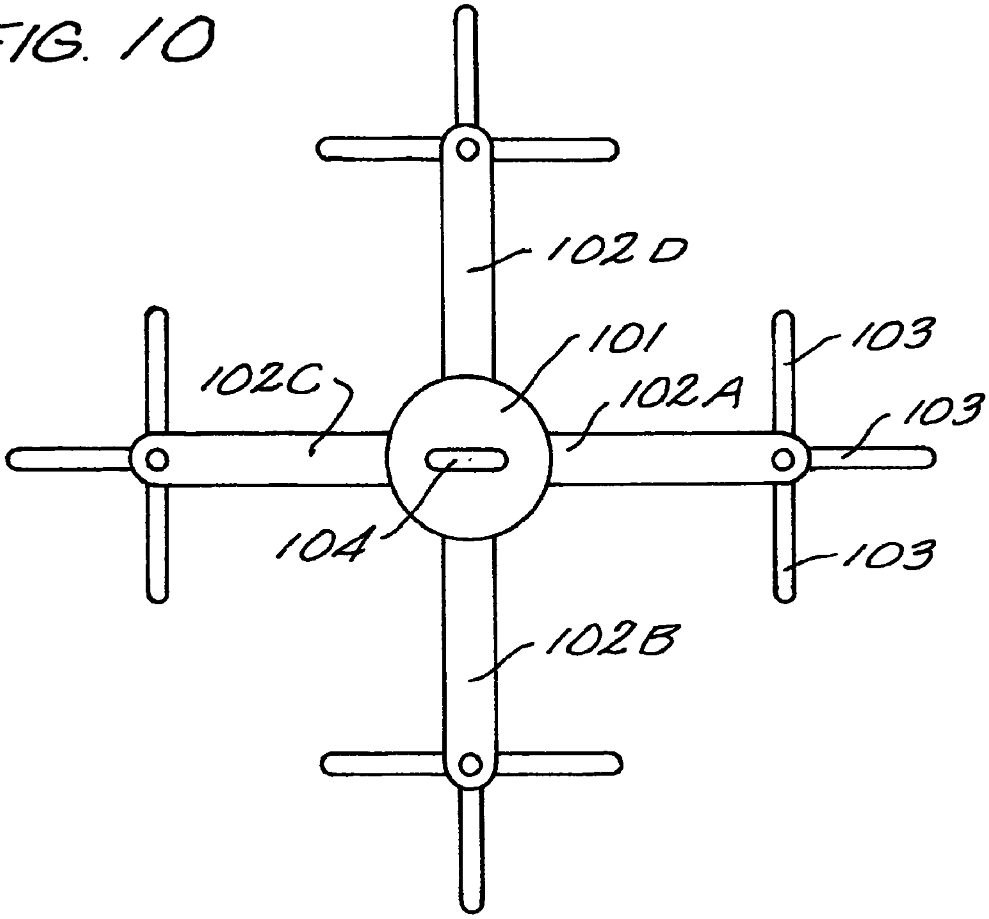
102D
101
103
102C
102A
103
104
103
102B

FIG. 11

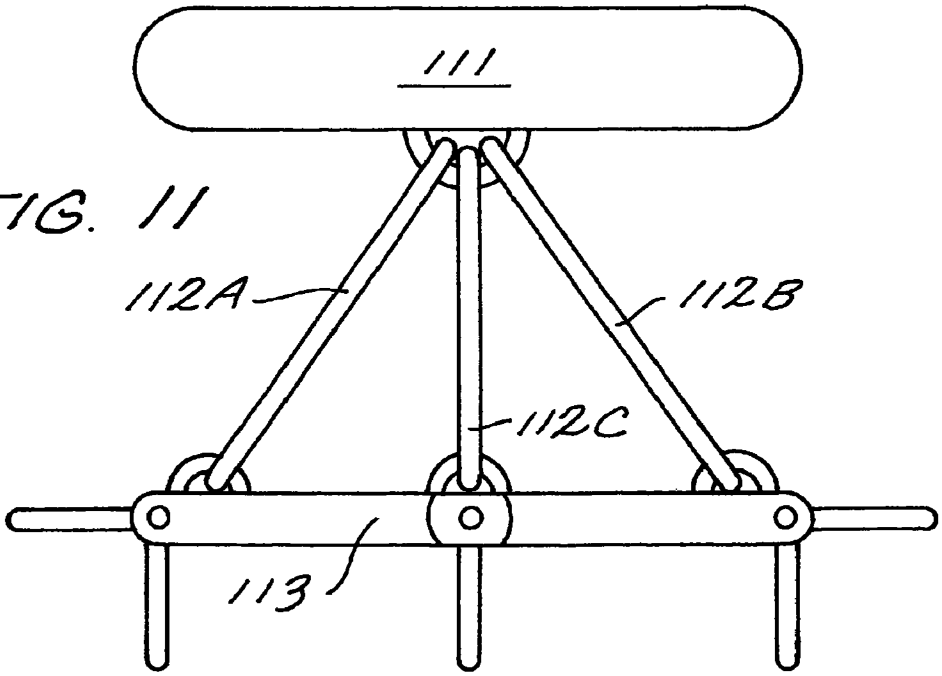
111
112A
112B
112C
113

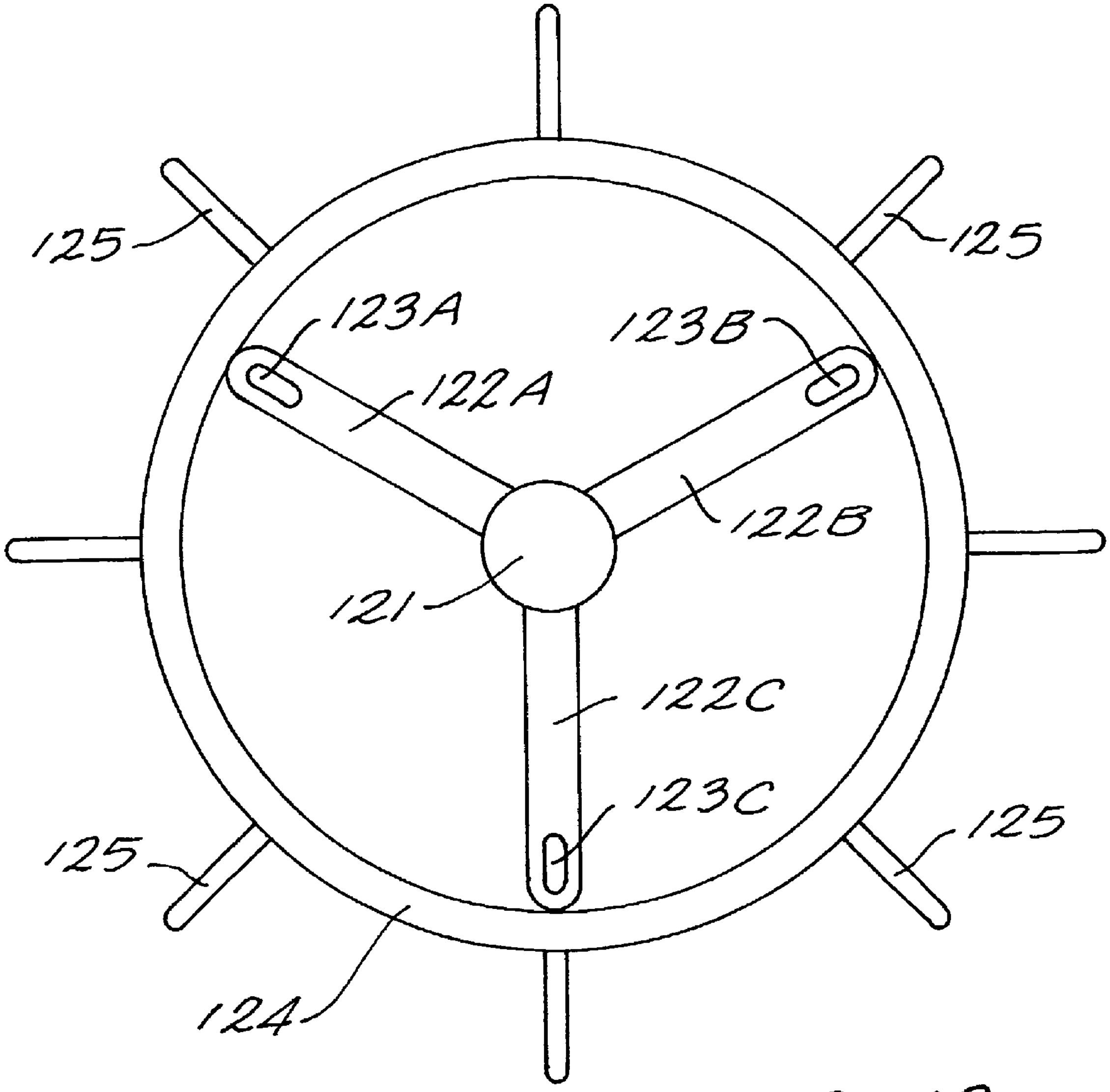
125
125
123A
123B
122A
122B
121
122C
123C
125
125
124

FIG. 12

ATMOSPHERIC STATIC ELECTRICITY COLLECTOR

This is a continuation in part of U.S. patent application Ser. No. 12/218,297, filed on Jul. 14, 2008, now U.S. Pat. No. 7,855,476 and entitled, "Atmospheric Electrical Generator".

BACKGROUND OF THE INVENTION

This invention relates to the production of electrical energy and more particularly to the collection of static electricity from the atmosphere.

Everyone is familiar with Benjamin Franklin's kite experiment of 1752. Using a kite whose string had become wet, negative charges from the passing clouds flowed into the string, down to the suspended key, and then into a Leyden jar via a thin metal wire. Franklin was protected by a dry silk string; but, when Franklin's knuckle came too close to the key, he received a strong shock. Fortunately, Benjamin Franklin was not killed, others who tried this same experiment were not so lucky.

Since then, the formation of lightning has remained something of a mystery. Lightning bolts are triggered when a negatively charged cloud base induces a positive charge from the ground, thereby forming a "pathway" for the discharge of the collected electrical energy.

Lightning travels up to 60,000 miles per hour with a flash that is brighter than ten million 100-watt lightbulbs. This wattage is as much power as is produced by all of the electricity plants in the United States and with a voltage of up to 300 million volts.

It is this very fact, the power within lightning is immense, that has prevented any successful collection of the electrical energy from lightning. The electricity in lightning is far too extreme for current technology to harness.

While lightning has attracted a energy starved industrial world, no one has developed any technique to harness this naturally occurring electrical source.

It is clear there is a continuing need for an electrical source other than carbon-based fuels and that the naturally occurring electricity in the atmosphere is being ignored.

SUMMARY OF THE INVENTION

The invention is a mechanism which taps into the naturally occurring static electricity in the atmosphere. Whereas heretofore, the attempt to garner electricity from the atmosphere has focused exclusively on capturing lightning, the present invention syphons off the static electricity which is generated from any agitated air and voids lightning.

Lightning is only the final discharge of the static electricity, whether that lightning is intra-cloud lightning, cloud-to-ground lightning, or inter-cloud lightning. Other types of final discharges are known as heat lightning, summer lightning, sheet lightning, ribbon lightning, silent lightning, ball lightning, bead lightning, elves, jets, and sprites. Well before these discharges are observed, as the atmosphere becomes agitated by wind or thermal, static electricity is being generated.

The present invention recognizes that this static electricity is being formed and creates a mechanism to capture it.

The mechanism of this invention utilizes an aircraft such as a lighter than air balloon. While the preferred embodiment uses a foil balloon, a variety of other aircraft are obvious to those of ordinary skill in the art, including, but not limited to: gliders, rubber balloons (such as weather balloons), biaxially-oriented polyethylene terephthalate polyester film balloons, and latex balloons.

Within this discussion, the balloon is referenced, but, the invention is not intended to be limited solely to balloons.

The balloon is sent aloft and is tethered by a conductive line. In this context, the conductive line may be any obvious to those of ordinary skill in the art. For the preferred embodiment, the conductive line is a generically referred to as a "poly-rope" and is commercially available through a variety of sources. A suitable conductive line is described in U.S. Pat. No. 5,203,542, entitled "Apparatus for Improved Electric Fence Wire Construction for use with Intensive Grazing" issued Apr. 20, 1993, to Coley, et al. and incorporated hereinto by reference.

The conductive line is played out of a winch to control the altitude of the balloon. The motor controlling the winch is able to reverse direction to both extend and withdraw the conductive line which is wrapped around a spool on the winch. The winch/spool combination are part of a base unit.

In some embodiments of the invention, the spool is constructed of rubber so as to insulate the conductive line from the winch assembly. In this embodiment, only the conductive line is charged by the atmospheric static electricity while the winch remains neutral.

In yet another embodiment, the winch/spool are part of a base unit which is itself isolated from the ground by an insulator. In this embodiment, the entire base unit is charged by the atmospheric static electricity.

A conductor, such as an insulated wire, is electrically connected to the conductive line. In one embodiment, where the conductive line is electrically isolated from the spool and winch motor, the conductor is connected to the conductive line. In the embodiment where the conductive line is electrically connected to the base unit, then the conductor is connected anywhere on a metalic base unit.

The other end of the conductor is connected to a load. The load in this case can be any of a variety of electrical loads well known to those of ordinary skill in art, including, but not limited to a motor, a battery system, or the electrical grid for the system.

In the preferred embodiment, a sensor array is used to monitor the activities both at the base unit (such as electrical flow within the conductor) and in the surrounding locale.

A sensor monitoring the electrical flow (i.e. voltage and/or current) within the conductor is used to monitor the electrical activity within the conductor.

In the preferred embodiment, a lightning sensor monitors for lightning activity within the locale. As noted earlier, the electrical characteristic of lightning is so extreme that ideally this discharge is avoided as it might damage the mechanism of this invention.

The sensor array is utilized by a controller, such as microprocessor, programmed to operate the mechanism as outlined herein.

The controller operates the winch motor to extend or withdraw the conductive line and by extension the altitude of the balloon. The controller is programmed to operate the winch by monitoring the electrical characteristics of the conductor and adjusting the balloon's altitude to maintain these characteristics within the conductor within a preset range.

This preset range is established either in the base programming of the controller or is established by an operator of the system.

As example, by controlling the amount of current being withdrawn from the atmosphere, the mechanism operates within a safe range and also provides a relatively stable current flow from which a variety of activities can take place (such as DC-AC conversion).

The controller also utilizes the lightning sensor to protect the mechanism from a lightning strike. Should lightning be detected within a pre-determined range (as established by the software or defined by an operator), then the balloon is pulled down to minimize the risk of damage from a lightning strike.

An aspect of the present invention is the use of an antenna which are used to collect the atmospheric static electricity. The antenna is shaped as a hub which is suspended from the blimp/balloon. The hub is ideally spoked shaped although an alternative embodiment uses a solid hub.

A number of rods extend from the hub so as collect atmospheric static electricity These rods are ideally rounded at the ends to enhance the attraction of the atmospheric static electricity.

The collected atmospheric electricity is conducted from the rods to an electrical connection on the hub where the electricity is conducted to a power plant on earth such as described above.

The invention, together with various embodiments thereof will be more fully explained by the following description of the accompanying drawings.

DRAWINGS IN BRIEF

FIG. **1** diagrams the preferred embodiment of the invention.

FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

FIGS. **9**A and **9**B are side views and top views of an embodiment of the antenna used to collected atmospheric electricity.

FIG. **10** is a top view of an alternative embodiment of the antenna of this invention.

FIG. **11** is a side view of yet another alternative embodiment of the antenna used to collect atmospheric electricity.

FIG. **12** is the preferred embodiment of the antenna of this invention.

DRAWINGS IN DETAIL

FIG. **1** diagrams the preferred embodiment of the invention.

Balloon **10** is an aircraft which, in this illustration, is a lighter than air balloon. Wings **10**A, extending from the body of balloon **10**, provide additional lift in air flow **18**. Tail **10**B helps to stabilize balloon **10**.

Balloon **10** is tethered to the ground via conductive line **12**. As noted earlier, a variety of configurations and materials are available to serve as conductive line **12**. In this illustration, a poly-wire is used. Poly-wire is commercially available through a variety of vendors, including, but not limited to: Jeffers Livestock and Sareba Systems, Inc. of Ellendale, Minn.

In this embodiment, located proximate to balloon **10**, is an electrical collection enhancement lead **11** which assists in the collection of the static electrical charge in the atmosphere. Electrical collection enhancement lead **11** is configured to attract the static charge and conduct the charge into the conductive line **12**.

The electricity flows down the conductive line into spool **13**, where the conductive line **12** is collected and either withdrawn or dispensed through operation of winch motor **14**.

Winch motor **14** and spool **13** are mounted onto base unit **16** which is electrically isolated from ground **7** using insulator **17**. Note, in this embodiment of the invention, when electricity is being collected from the atmosphere, the entire base unit **16** becomes charged. In another embodiment of the invention, spool **13** is constructed of rubber, thereby preventing base unit **16** from becoming charged, thereby restricting the charging from the atmosphere to only conductive line **12**.

In this embodiment, conductor **6** is connected to base unit **16** (since the entire base unit **16** is charged and the base unit is metalic) to communicate the electrical current to load **5**. Conductor **6** is ideally an insulated wire.

The electrical current through conductor **6** is measured using sensor **8**.

In the alternative embodiment discussed above, where only the conductive line **12** is charged, then conductor **6** is connected to conductive line **12**.

Controller **15**, located in this embodiment on base unit **16**, operates winch motor **14** in response to signals from sensor **8** (measuring the current being discharged to load **5**) to maintain the current flow within a pre-defined range. As the current flow diminishes, then the conductive line **12** extended from spool **13** to increase the altitude of balloon **10** to that more static charge from the atmosphere is gathered; as the current flow falls exceeds a preset level, conductive line **12** is withdrawn onto spool **13** to decrease the static charge being collected from the atmosphere.

The range of current flow through conductor **6** is ideally set by the program, although some embodiments of the invention permit an operator to establish this range of operation.

In an alternative embodiment, the sensor monitoring conductor **6** monitors the voltage therein.

In the preferred embodiment of the invention, controller **15** is also equipped with a lightning sensor **19**. In this embodiment, when lightning is sensed within a preset range, then substantially all of conductive line **12** is wound onto spool **13** to pull balloon **10** near the ground and protect the entire mechanism from being damaged from a lightning discharge.

In the preferred embodiment, the "safe" distance form lightning is set in the programming of controller **15** and is ideally two miles; other embodiments permit the operator to "safe" distance.

There are a variety of lightning sensors well known to those of ordinary skill in the art, including, but not limited to those described in: U.S. Pat. No. 7,016,785, entitled "Lightning Detection" issued to Makela, et al. on Mac. 21, 2006; U.S. Pat. No. 6,829,911, entitled "Lightning Detection and Prediction Alarm Device" issued to Jones, et al. on Dec. 7, 2004; U.S. Pat. No. 7,200,418, entitled "Detection of Lightning" issued to Karikuranta, et al. on Apr. 3, 2007; and U.S. Pat. No. 6,961,662, entitled "Systems and Methods for Spectral Corrected Lightning Detection" issued to Murphy on Nov. 1, 2005; all of which are incorporated hereinto by reference.

In another embodiment of the invention, controller **15** is not located on base unit **16**, rather it is remote and communicates its control signals to winch motor **14** using radio waves.

FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

Static charges **23** are generated in the atmosphere by agitated air. These static charges are often collected at the bottom of clouds, but exist in other environments as well.

Balloon **21** is extended into this strata of static charges **23** which are then attracted to conductive line **12** to flow to base unit **22** and then onto load **5**.

By increasing or decreasing the altitude of balloon **21** (defined by the length of the extended conductive line **12**), conductive line **12** is selectively exposed to varying densities and levels of the static charge strata, and by extension, the current flow or voltage is increased or decreased.

FIG. **3** is a flowchart of the operation of the controller for the preferred embodiment of the invention.

Once the program starts **30**, the lightning sensor is checked to determine if lightning has occurred within the unsafe range **31**A, if it has, then the balloon is lowered **32**A, and the program continues monitoring the status of lightning until no lightning is detected.

When the lightning status is acceptable, then the current within the conductor is checked to see if the current is within the prescribed range **31**B. If the current is acceptable (within range) the program returns to check the lightning status **31**A; otherwise a determination is made to see if the current is above the prescribed range **31**C.

If the current is above the prescribed range, then the altitude of the balloon is withdrawn a set amount **32**B (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

If the current is below the prescribed range, then the altitude of the balloon is extended a set amount **32**C (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

In this manner of feed-back and minor adjustments in the altitude of the balloon, the current is maintained within a prescribed range which can be handled by the downstream electrical system.

As noted earlier, some embodiments of the invention monitor the voltage instead of the current.

FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

By maintaining the voltage being collected in a prescribed range, an electrical conversion system is easily designed. While FIGS. **4**A, **4**B, and **4**C illustrate some electrical configurations, those of ordinary skill in the art readily recognize a variety of other configurations which will serve the same function.

Referencing FIG. **4**A, Direct Current In (DC IN) **40** is buffered by a gang of capacitors **41** before being communicated to a DC/AC converter **42**. The DC/AC converter converts the direct current into a an alternating current suitable for placement over an existing electrical grid **43** such as normally found from a power-plant.

Those of ordinary skill in the art readily recognize a variety of DC/AC converters, including, but not limited to: U.S. Pat. No. 7,394,671, entitled "Controller IC, DC-AC Conversion Apparatus, and parallel running system of DC-AC Conversion Apparatuses" issued to Fukumoto, et al. on Jul. 1, 2008; and, U.S. Pat. No. 7,330,366, entitled "DC-AC Converter" issued to Lee, et al. on Feb. 12, 2008; all of which are incorporated hereinto by reference.

FIG. **4**B illustrates an electrical arrangement suitable for use in charging a battery. DC IN **40** is buffered by capacitor bank **41** before entering into a step down transformer **43**. Step down transformer **43** reduces the voltage so that the voltage can safely be introduced into battery **44** which is connected to ground **45** at the battery's other pole.

Those of ordinary skill in the art readily recognize a variety of batteries which will work in this capacity, including, but not limited to those described in: U.S. Pat. No. 7,378,181, entitled "Electric Storage Battery Construction and Manufacture" issued to Skinlo on May 27, 2008; U.S. Pat. No. 7,388,350, entitled "Battery with Electronic Compartment" issued to Wright on Jun. 17, 2008; U.S. Pat. No. 7,397,220, entitled "Connection Member and Battery Pack" issued to Uchida, et al. on Jul. 8, 2008; and, U.S. Pat. No. 7,375,492, entitled "Inductively Charged Battery Pack" issued to Calhoon, et al. on May 20, 2008; all of which are incorporated hereinto by reference.

In FIG. **4**C, DC IN **40** is fed into an adjustable rheostat **46** which is controlled by the controller so that the DC OUT **47** falls within a specified range.

FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

This type of conductive line is commonly called poly-wire and consists of multiple interwoven strands of plastic **50**A and **50**B woven into a cord or rope arrangement having intertwined therein exposed metal wires **51**A and **51**B. While this illustration shows two plastic strands and two metal wires, any number of possible combinations is possible.

The exposed metal wires **51**A and **51**B attract the atmospheric static charge and transmit the charge down to the base unit (not shown).

FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

This conductive line utilizes a tube **60** having an outer layer **62** of PET Film (Biaxially-oriented polyethylene terephtalate polyester film) which provides exceptionally high tensile strength and is chemically and dimensionally stable. The tube has an ideal diameter of between two and three inches.

An interior metal coating **61** provides an initial conduit for the flow of static charge. The static charge through the metal forces the tube to expand due to the repulsion experienced by like charges. Further, the flow of electricity causes the interior of the tube **60** to become ionized to provide an additional pathway for the atmospheric static charges to the base unit (not shown).

Because outer layer **62** provides a gas barrier, the resulting ionization is not dissipated by air currents, thereby providing a highly stable pathway.

FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

In this embodiment, controller box **70**, resting on insulating pad **72**, is in communication with the sensors as described above. Using the input from these sensors, when there is flow of electricity through the base unit, warning flashing light **71** is illuminated. To electrically neutralize the mechanism, switch **73** is activated to pass any existing current into the ground **74**.

FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

Referencing FIG. **8**A, enhanced electrical collection lead **82** is a wire mesh which is in electrical communication with conductive line **81** and balloon **80**. Because of the significant amount of metal exposed by enhanced electrical collection lead **82**, more static electricity from the atmosphere is drawn to the collection lead **82**, and then down conductive line **81** to the base unit (not shown).

Conductive lead **82** is positioned proximate to balloon **80**.

In FIG. **8**B, poly-wire **83** has enhanced electrical collection leads **84** wrapped therearound. Collection leads **84** have pointed ends **85**A and **85**B which have a propensity to attract more electricity than rounded ends do.

FIGS. **9**A and **9**B are side views and top views of an embodiment of the antenna used to collected atmospheric electricity.

Blimp **90** supports antenna **92** via tether **91** which is attached to antenna **92** by connection **94**. In this embodiment, antenna **92** is made of electrically conductive material and is ideally light in weight to lessen the payload requirements on blimp **90**.

Encircling antenna **92** and extending therefrom are rods **93** which are rounded at their distal ends so as to enhance attraction of the atmospheric static electricity. The static electricity is communicated to connection **96** and then to electrical line **95** which communicates the electricity to the ground based station (not shown) as described above.

FIG. **10** is a top view of an alternative embodiment of the antenna of this invention.

In this embodiment of the antenna a central hub **101** has the blimp connection **104** (capable of receiving the tether to the blimp) secured thereto. In this illustration, four arms **102**A, **102**B, **102**C, and **102**D extend from hub **101**. Hub **101**, and arms **102**A, **102**B, **102**C, and **102**D are all electrically conductive in this embodiment.

At the end of each arm are rods **103** which are used to enhance the collection of the atmospheric static electricity.

Although this embodiment illustrates four arms, the invention is not intended to be limited to four arms, rather, any number of arms may be used and the number of rods extending from the distal ends of the arms also varies.

FIG. **11** is a side view of yet another alternative embodiment of the antenna used to collect atmospheric electricity.

In this embodiment, several tethers **112**A, **112**B, and **112**C, are used to secure the antenna **113** to the blimp **111**. This arrangement of several tethers provides heightened stability of the antenna by reducing the affects wind will have on the antenna.

FIG. **12** is the preferred embodiment of the antenna of this invention.

In this embodiment of the antenna, arms **122**A, **122**B, and **122**C extend from a central hum and are electrically connected to rim **124**. Tether connectors **123**A, **123**B, and **123**C, are used to secure the antenna to the blimp or balloon.

Rods **125** extend from rim **124** to increase the collection of the static charges in the atmosphere.

It is clear from the foregoing that the present invention captures an entirely new source of electrical energy.

What is claimed is:

1. A system for collection of atmospheric static electricity comprising:

a balloon;

a hub that is electrically conductive and connected to the balloon by a tether, the hub having:

at least three electrically conductive arms, a first end of each of said conductive arms electrically connected to said hub; and

a plurality of rods, a first end of each of said rods being electrically connected to a second end of at least one of said at least three arms;

a base unit having a spool of conductive line on a winch motor, one end of said conductive line secured to said hub, a portion of said conductive line and said hub collecting electricity in the atmosphere, said winch motor selectively extending or withdrawing said conductive line from said spool;

a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load being powered by collected electricity from said conductive line;

an electrical flow sensor monitoring electrical flow through said conductor and generating an electrical flow indicia indicative of said electrical flow in said conductor; and

a controller receiving said electrical flow indicia and selectively operating said winch motor such that said electrical flow indicia remains within a selected operating range.

2. The system according to claim **1**, wherein each of said plurality of rods has a rounded second end.

3. The system according to claim **2**, wherein at least two rods are attached to the second end of each of said at least three arms.

4. The system according to claim **1**, further comprising an electrically conductive rim connected to a second end of said electrically conductive arms.

5. The system according to claim **4**, wherein at least a portion of said plurality of electrically conductive arms are connected substantially at right angles to an exterior of said electrically conductive rim.

6. A system for collection of atmospheric static electricity comprising:

a radial arrangement of at least three electrically conductive arms;

a plurality of rods, each of said rods being electrically connected to at least one of said at least three arms; and

a balloon tether connected between a balloon and said radial arrangement of at least three electrically conductive arms;

a base unit having a spool of conductive line on a winch motor, one end of said conductive line secured to said radial arrangement of at least three electrically conductive arms, a portion of said conductive line and said radial arrangement of at least three electrically conductive arms collecting electricity in the atmosphere, said winch motor selectively extending or withdrawing said conductive line from said spool;

a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load being powered by collected electricity from said conductive line;

an electrical flow sensor monitoring electrical flow through said conductor and generating an electrical flow indicia indicative of said electrical flow in said conductor; and

a controller receiving said electrical flow indicia and selectively operating said winch motor such that said electrical flow indicia remains within a selected operating range.

7. The system according to claim **6**, wherein each of said plurality of rods has a rounded end distal from said electrically conductive arm.

8. The system according to claim **7**, further including a balloon tether connection secured to the first end of said at least three arms.

9. The system according to claim **8**, further including an electrical connection secured to the first end of said at least three arms.

10. The system according to claim **9**, wherein said at least two rods attached to said at least three arms is five rods.

11. The system according to claim **9**, further including an electrically conductive rim connected to a second end of said electrically conductive arms.

12. The system according to claim **11**, wherein said plurality of arms are electrically connected to said electrically conductive rim.

13. The system according to claim **12**, wherein at least a portion of said plurality of rods are connected to an exterior of said electrically conductive rim substantially at right angles thereto.

14. A system for the collection of atmospheric static electricity comprising:

a generally circular electrically conductive body;

a balloon tether connected to a first side of said generally circular electrically conductive body and to a balloon; and

an electrical connection connected to a second side of said generally circular electrically conductive body;

a base unit having a spool of conductive line on a winch motor, one end of said conductive line secured to said electrical connection, a portion of said conductive line and said generally circular electrically conductive body collecting electricity in the atmosphere, said winch motor selectively extending or withdrawing said conductive line from said spool;

a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load being powered by collected electricity from said conductive line;

an electrical flow sensor monitoring electrical flow through said conductor and generating an electrical flow indicia indicative of said electrical flow in said conductor; and

a controller receiving said electrical flow indicia and selectively operating said winch motor such that said electrical flow indicia remains within a selected operating range.

15. The system according to claim **14**, wherein said generally circular electrically conductive body comprises:

an electrically conductive hub;

at least three electrically conductive arms, each of said arms connected at a first end to said hub; and

an electrically conductive rim connected to the second end of each of said at least three arms.

16. The system according to claim **15**, further comprising a plurality of electrically conductive rods, a first end of each of said rods connected to said rim and extending therefrom substantially at rights angles thereto.

* * * * *

US 20100007218A1

(19) **United States**
(12) **Patent Application Publication** (10) **Pub. No.: US 2010/0007218 A1**
Ogram (43) **Pub. Date: Jan. 14, 2010**

(54) ATMOSPHERIC ELECTRICAL GENERATOR

(76) Inventor: **Mark Ellery Ogram**, Tucson, AZ (US)

Correspondence Address:
Mark E. Ogram (30343)
Ste. 110-186, 6990 E. 22nd Street
Tucson, AZ 85710-5192 (US)

(21) Appl. No.: **12/218,297**

(22) Filed: **Jul. 14, 2008**

Publication Classification

(51) **Int. Cl.**
H02G 11/00 (2006.01)
(52) **U.S. Cl.** **307/145**

(57) ABSTRACT

A mechanism to tap an electrical source which uses an aircraft (preferably a lighter than air balloon) tethered by a conductive line. The conductive line is extended/withdrawn by a winch motor to adjust the altitude of the aircraft. The conductive line is isolated from the ground and an electrical conductor is connected to the conductive line and to an electrical load. In this manner, static electricity generated in the atmosphere is gathered for use.

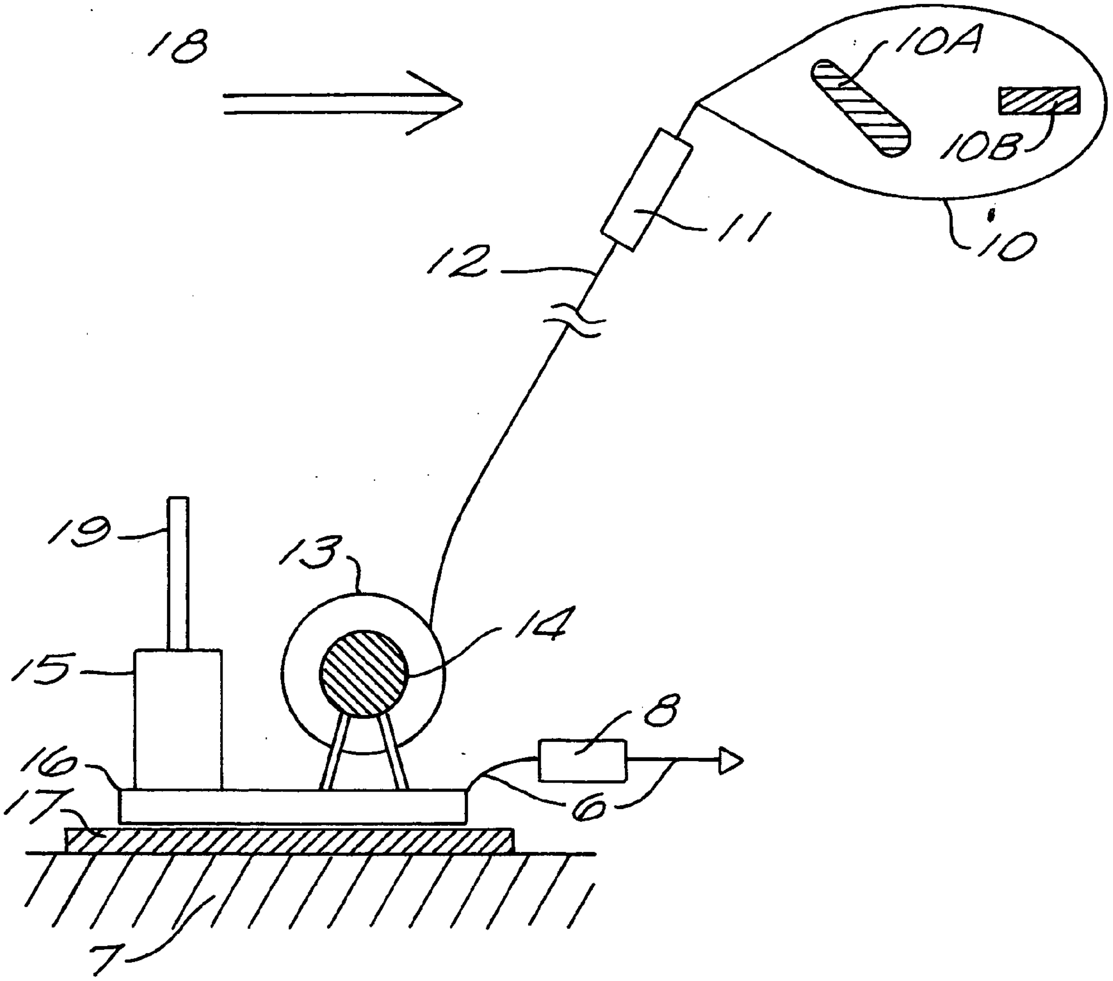

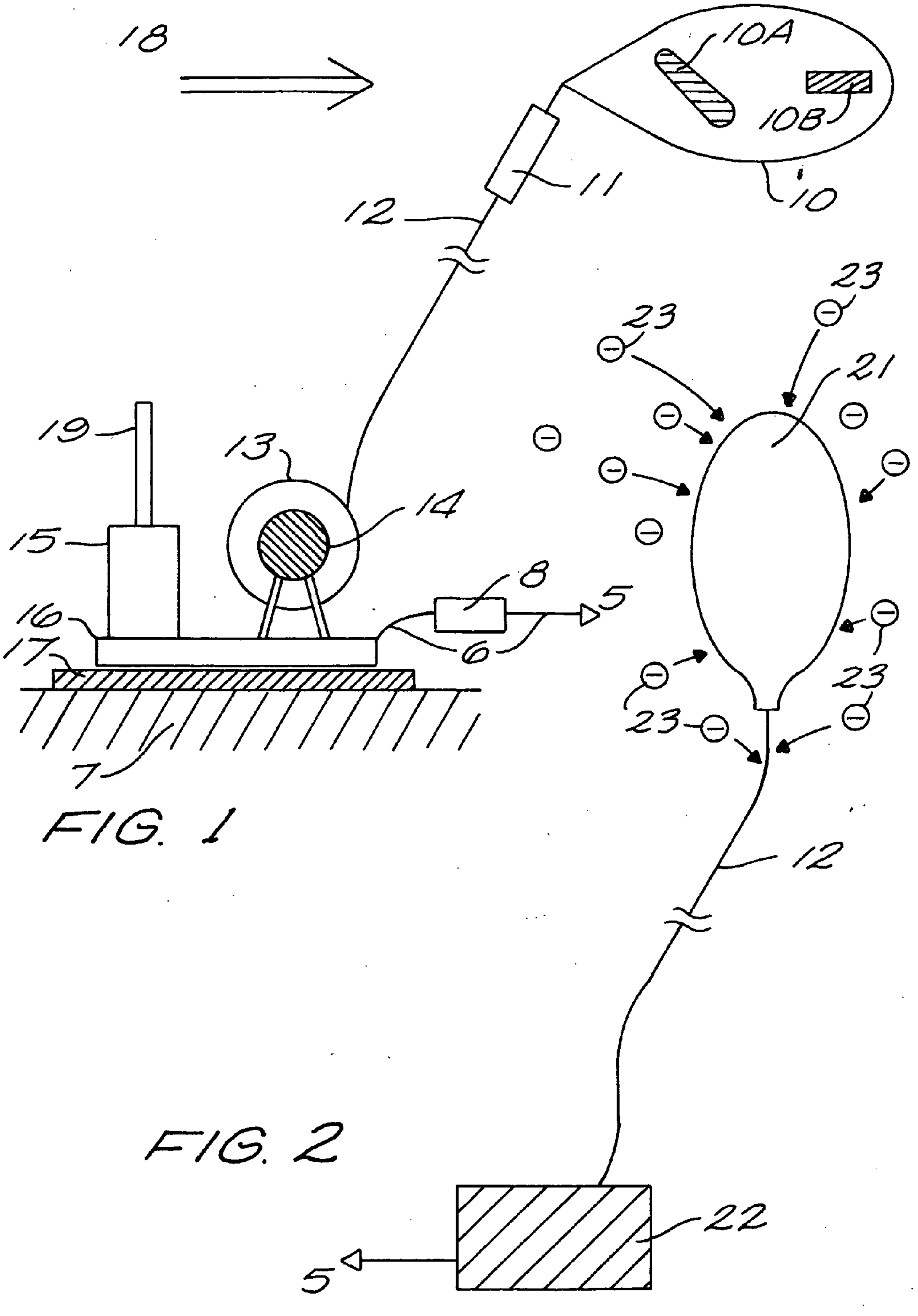
18
10A
10B
10
11
12
19
13
14
15
8
5
16
6
17
7
23
23
21
23
23
12
22
5

FIG. 1

FIG. 2

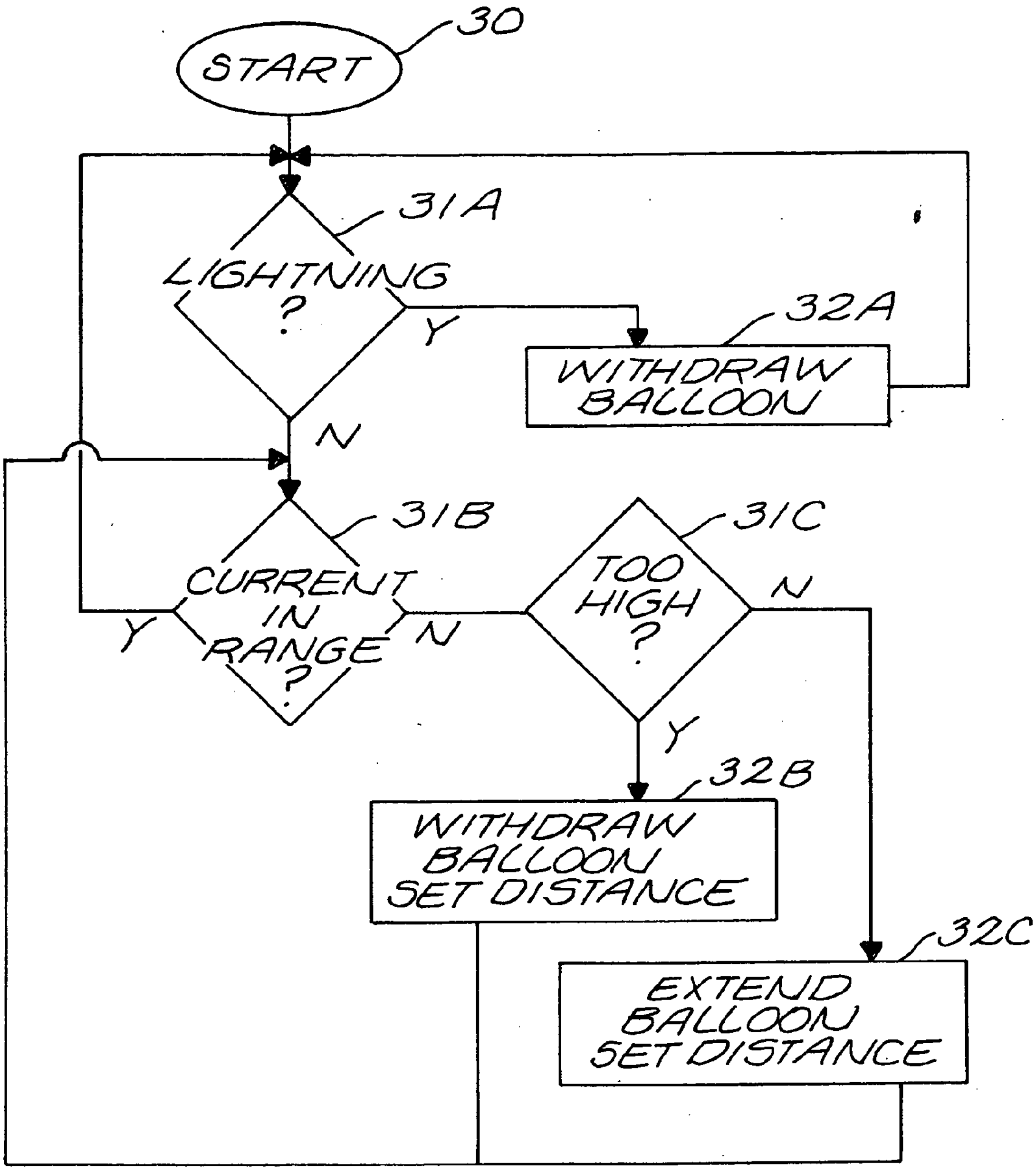
30
START
31A
LIGHTNING ?
Y
N
32A
WITHDRAW BALLOON
31B
CURRENT IN RANGE ?
Y
N
31C
TOO HIGH ?
N
Y
32B
WITHDRAW BALLOON SET DISTANCE
32C
EXTEND BALLOON SET DISTANCE

FIG. 3

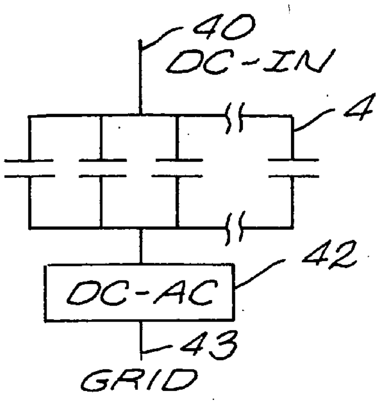
40
DC-IN
41
DC-AC
42
43
GRID

FIG. 4A

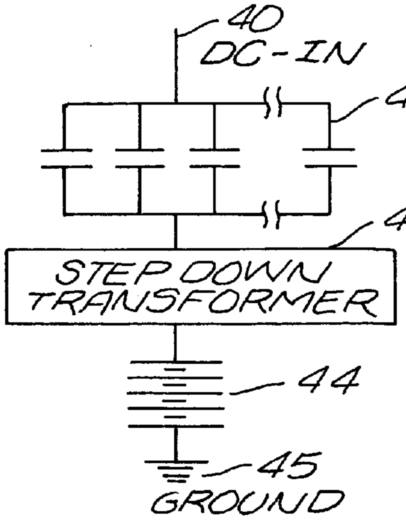
40
DC-IN
41
43
STEP DOWN TRANSFORMER
44
45
GROUND

FIG. 4B

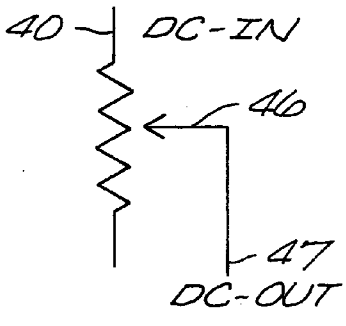
40
DC-IN
46
47
DC-OUT

FIG. 4C

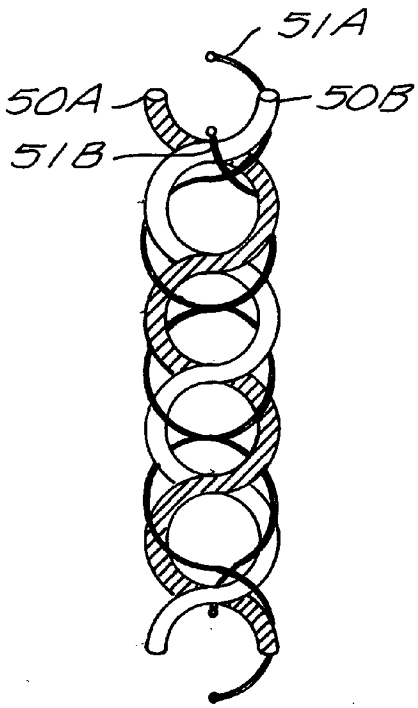
51A
50A
50B
51B

FIG. 5

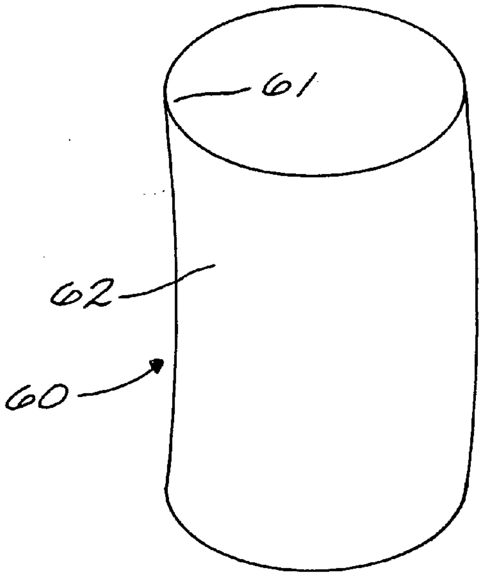

FIG. 6A

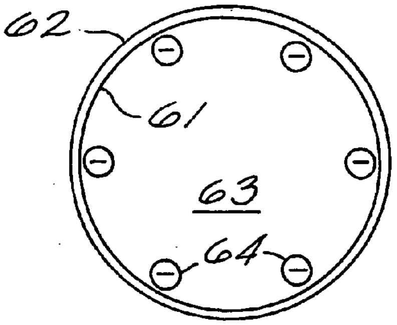

FIG. 6B

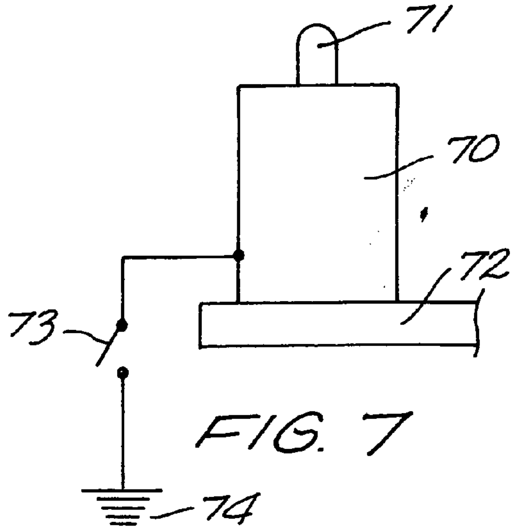
71
70
72
73
74

FIG. 7

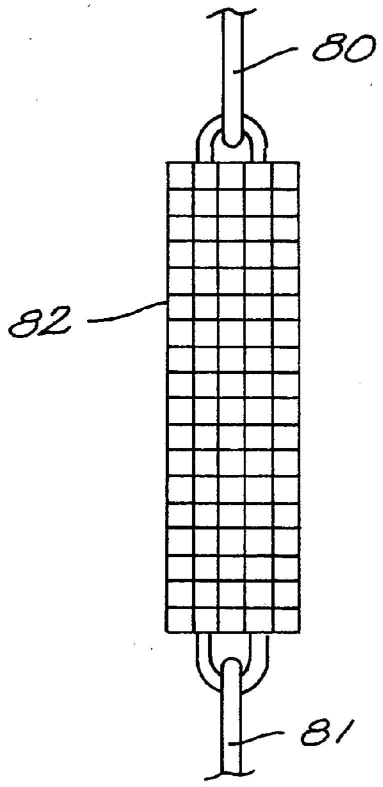

FIG. 8A

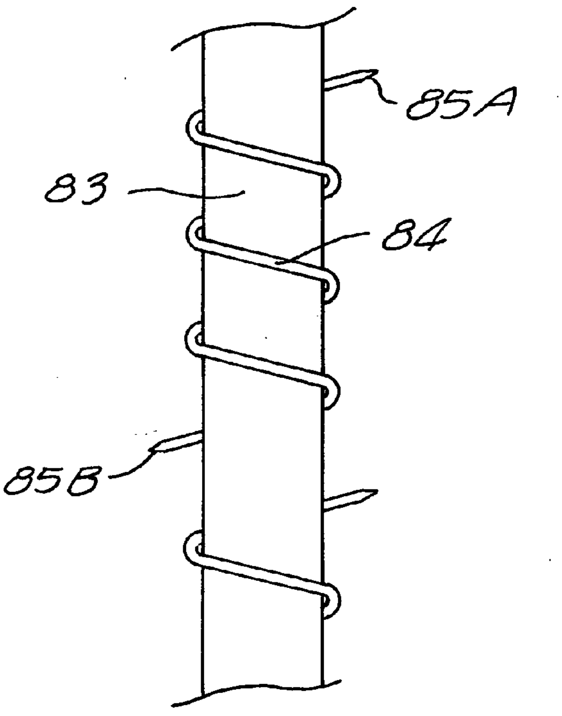
85A
83
84
85B

FIG. 8B

ATMOSPHERIC ELECTRICAL GENERATOR

BACKGROUND OF THE INVENTION

[0001] This invention relates to the production of electrical energy and more particularly to the production of electrical energy from the atmosphere.

[0002] Everyone is familiar with Benjamin Franklin's kite experiment of 1752. Using a kite whose string had become wet, negative charges from the passing clouds flowed into the string, down to the suspended key, and then into a Leyden jar via a thin metal wire. Franklin was protected by a dry silk string; but, when Franklin's knuckle came too close to the key, he received a strong shock. Fortunately, Benjamin Franklin was not killed, others who tried this same experiment were not so lucky.

[0003] Since then, the formation of lightning has remained something of a mystery. Lightning bolts are triggered when a negatively charged cloud base induces a positive charge from the ground, thereby forming a "pathway" for the discharge of the collected electrical energy.

[0004] Lightning travels up to 60,000 miles per hour with a flash that is brighter than ten million 100-watt lightbulbs. This wattage is as much power as is produced by all of the electricity plants in the United States and with a voltage of up to 300 million volts.

[0005] It is this very fact, the power within lightning is immense, that has prevented any successful collection of the electrical energy from lightning. The electricity in lightning is far too extreme for current technology to harness.

[0006] While lightning has attracted a energy starved industrial world, no one has developed any technique to harness this naturally occurring electrical source.

[0007] It is clear there is a continuing need for an electrical source other than carbon-based fuels and that the naturally occurring electricity in the atmosphere is being ignored.

SUMMARY OF THE INVENTION

[0008] The invention is a mechanism which taps into the naturally occurring static electricity in the atmosphere. Whereas heretofore, the attempt to garner electricity from the atmosphere has focused exclusively on capturing lightning, the present invention syphons off the static electricity which is generated from any agitated air and voids lightning.

[0009] Lightning is only the final discharge of the static electricity, whether that lightning is intra-cloud lightning, cloud-to-ground lightning, or inter-cloud lightning. Other types of final discharges are known as heat lightning, summer lightning, sheet lightning, ribbon lightning, silent lightning, ball lightning, bead lightning, elves, jets, and sprites. Well before these discharges are observed, as the atmosphere becomes agitated by wind or thermal, static electricity is being generated.

[0010] The present invention recognizes that this static electricity is being formed and creates a mechanism to capture it.

[0011] The mechanism of this invention utilizes an aircraft such as a lighter than air balloon. While the preferred embodiment uses a foil balloon, a variety of other aircraft are obvious to those of ordinary skill in the art, including, but not limited to: gliders, rubber balloons (such as weather balloons), biaxially-oriented polyethylene terephthalate polyester film balloons, and latex balloons.

[0012] Within this discussion, the balloon is referenced, but, the invention is not intended to be limited solely to balloons.

[0013] The balloon is sent aloft and is tethered by a conductive line. In this context, the conductive line may be any obvious to those of ordinary skill in the art. For the preferred embodiment, the conductive line is a generically referred to as a "poly-rope" and is commercially available through a variety of sources. A suitable conductive line is described in U.S. Pat. No. 5,203,542, entitled "Apparatus for Improved Electric Fence Wire Construction for use with Intensive Grazing" issued Apr. 20, 1993, to Coley, et al. and incorporated hereinto by reference.

[0014] The conductive line is played out of a winch to control the altitude of the balloon. The motor controlling the winch is able to reverse direction to both extend and withdraw the conductive line which is wrapped around a spool on the winch. The winch/spool combination are part of a base unit.

[0015] In some embodiments of the invention, the spool is constructed of rubber so as to insulate the conductive line from the winch assembly. In this embodiment, only the conductive line is charged by the atmospheric static electricity while the winch remains neutral.

[0016] In yet another embodiment, the winch/spool are part of a base unit which is itself isolated from the ground by an insulator. In this embodiment, the entire base unit is charged by the atmospheric static electricity.

[0017] A conductor, such as an insulated wire, is electrically connected to the conductive line. In one embodiment, where the conductive line is electrically isolated from the spool and winch motor, the conductor is connected to the conductive line. In the embodiment where the conductive line is electrically connected to the base unit, then the conductor is connected anywhere on a metallic base unit.

[0018] The other end of the conductor is connected to a load. The load in this case can be any of a variety of electrical loads well known to those of ordinary skill in art, including, but not limited to a motor, a battery system, or the electrical grid for the system.

[0019] In the preferred embodiment, a sensor array is used to monitor the activities both at the base unit (such as electrical flow within the conductor) and in the surrounding locale.

[0020] A sensor monitoring the electrical flow (i.e. voltage and/or current) within the conductor is used to monitor the electrical activity within the conductor.

[0021] In the preferred embodiment, a lightning sensor monitors for lightning activity within the locale. As noted earlier, the electrical characteristic of lightning is so extreme that ideally this discharge is avoided as it might damage the mechanism of this invention.

[0022] The sensor array is utilized by a controller, such as microprocessor, programmed to operate the mechanism as outlined herein.

[0023] The controller operates the winch motor to extend or withdraw the conductive line and by extension the altitude of the balloon. The controller is programmed to operate the winch by monitoring the electrical characteristics of the conductor and adjusting the balloon's altitude to maintain these characteristics within the conductor within a preset range.

[0024] This preset range is established either in the base programming of the controller or is established by an operator of the system.

[0025] As example, by controlling the amount of current being withdrawn from the atmosphere, the mechanism oper-

ates within a safe range and also provides a relatively stable current flow from which a variety of activities can take place (such as DC-AC conversion).

[0026] The controller also utilizes the lightning sensor to protect the mechanism from a lightning strike. Should lightning be detected within a pre-determined range (as established by the software or defined by an operator), then the balloon is pulled down to minimize the risk of damage from a lightning strike.

[0027] The invention, together with various embodiments thereof will be more fully explained by the following description of the accompanying drawings.

DRAWINGS IN BRIEF

[0028] FIG. **1** diagrams the preferred embodiment of the invention.

[0029] FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

[0030] FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

[0031] FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

[0032] FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

[0033] FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

[0034] FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

[0035] FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

DRAWINGS IN DETAIL

[0036] FIG. **1** diagrams the preferred embodiment of the invention.

[0037] Balloon **10** is an aircraft which, in this illustration, is a lighter than air balloon. Wings **10**A, extending from the body of balloon **10**, provide additional lift in air flow **18**. Tail **10**B helps to stabilize balloon **10**.

[0038] Balloon **10** is tethered to the ground via conductive line **12**. As noted earlier, a variety of configurations and materials are available to serve as conductive line **12**. In this illustration, a poly-wire is used. Poly-wire is commercially available through a variety of vendors, including, but not limited to: Jeffers Livestock and Sareba Systems, Inc. of Ellendale, Minn.

[0039] In this embodiment, located proximate to balloon **10**, is an electrical collection enhancement lead **11** which assists in the collection of the static electrical charge in the atmosphere. Electrical collection enhancement lead **11** is configured to attract the static charge and conduct the charge into the conductive line **12**.

[0040] The electricity flows down the conductive line into spool **13**, where the conductive line **12** is collected and either withdrawn or dispensed through operation of winch motor **14**.

[0041] Winch motor **14** and spool **13** are mounted onto base unit **16** which is electrically isolated from ground **7** using insulator **17**. Note, in this embodiment of the invention, when electricity is being collected from the atmosphere, the entire base unit **16** becomes charged. In another embodiment of the invention, spool **13** is constructed of rubber, thereby preventing base unit **16** from becoming charged, thereby restricting the charging from the atmosphere to only conductive line **12**.

[0042] In this embodiment, conductor **6** is connected to base unit **16** (since the entire base unit **16** is charged and the base unit is metallic) to communicate the electrical current to load **5**. Conductor **6** is ideally an insulated wire.

[0043] The electrical current through conductor **6** is measured using sensor **8**.

[0044] In the alternative embodiment discussed above, where only the conductive line **12** is charged, then conductor **6** is connected to conductive line **12**.

[0045] Controller **15**, located in this embodiment on base unit **16**, operates winch motor **14** in response to signals from sensor **8** (measuring the current being discharged to load **5**) to maintain the current flow within a pre-defined range. As the current flow diminishes, then the conductive line **12** extended from spool **13** to increase the altitude of balloon **10** to that more static charge from the atmosphere is gathered; as the current flow falls exceeds a preset level, conductive line **12** is withdrawn onto spool **13** to decrease the static charge being collected from the atmosphere.

[0046] The range of current flow through conductor **6** is ideally set by the program, although some embodiments of the invention permit an operator to establish this range of operation.

[0047] In an alternative embodiment, the sensor monitoring conductor **6** monitors the voltage therein.

[0048] In the preferred embodiment of the invention, controller **15** is also equipped with a lightning sensor **19**. In this embodiment, when lightning is sensed within a preset range, then substantially all of conductive line **12** is wound onto spool **13** to pull balloon **10** near the ground and protect the entire mechanism from being damaged from a lightning discharge.

[0049] In the preferred embodiment, the "safe" distance form lightning is set in the programming of controller **15** and is ideally two miles; other embodiments permit the operator to "safe" distance.

[0050] There are a variety of lightning sensors well known to those of ordinary skill in the art, including, but not limited to those described in: U.S. Pat. No. 7,016,785, entitled "Lightning Detection" issued to Makela, et al. on Mach 21, 2006; U.S. Pat. No. 6,829,911, entitled "Lightning Detection and Prediction Alarm Device" issued to Jones, et al. on Dec. 7, 2004; U.S. Pat. No. 7,200,418, entitled "Detection of Lightning" issued to Karikuranta, et al. on Apr. 3, 2007; and U.S. Pat. No. 6,961,662, entitled "Systems and Methods for Spectral Corrected Lightning Detection" issued to Murphy on Nov. 1, 2005; all of which are incorporated hereinto by reference.

[0051] In another embodiment of the invention, controller **15** is not located on base unit **16**, rather it is remote and communicates its control signals to winch motor **14** using radio waves.

[0052] FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

[0053] Static charges **23** are generated in the atmosphere by agitated air. These static charges are often collected at the bottom of clouds, but exist in other environments as well.

[0054] Balloon **21** is extended into this strata of static charges **23** which are then attracted to conductive line **12** to flow to base unit **22** and then onto load **5**.

[0055] By increasing or decreasing the altitude of balloon **21** (defined by the length of the extended conductive line **12**),

conductive line **12** is selectively exposed to varying densities and levels of the static charge strata, and by extension, the current flow or voltage is increased or decreased.

[0056] FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

[0057] Once the program starts **30**, the lightning sensor is checked to determine if lightning has occurred within the unsafe range **31**A, if it has, then the balloon is lowered **32**A, and the program continues monitoring the status of lightning until no lightning is detected.

[0058] When the lightning status is acceptable, then the current within the conductor is checked to see if the current is within the prescribed range **31**B. If the current is acceptable (within range) the program returns to check the lightning status **31**A; otherwise a determination is made to see if the current is above the prescribed range **31**C.

[0059] If the current is above the prescribed range, then the altitude of the balloon is withdrawn a set amount **32**B (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

[0060] If the current is below the prescribed range, then the altitude of the balloon is extended a set amount **32**C (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

[0061] In this manner of feed-back and minor adjustments in the altitude of the balloon, the current is maintained within a prescribed range which can be handled by the downstream electrical system.

[0062] As noted earlier, some embodiments of the invention monitor the voltage instead of the current.

[0063] FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

[0064] By maintaining the voltage being collected in a prescribed range, an electrical conversion system is easily designed. While FIGS. **4**A, **4**B, and **4**C illustrate some electrical configurations, those of ordinary skill in the art readily recognize a variety of other configurations which will serve the same function.

[0065] Referencing FIG. **4**A, Direct Current In (DC IN) **40** is buffered by a gang of capacitors **41** before being communicated to a DC/AC converter **42**. The DC/AC converter converts the direct current into a an alternating current suitable for placement over an existing electrical grid **43** such as normally found from a power-plant.

[0066] Those of ordinary skill in the art readily recognize a variety of DC/AC converters, including, but not limited to: U.S. Pat. No. 7,394,671, entitled "Controller IC, DC-AC Conversion Apparatus, and parallel running system of DC-AC Conversion Apparatuses" issued to Fukumoto, et al. on Jul. 1, 2008; and, U.S. Pat. No. 7,330,366, entitled "DC-AC Converter" issued to Lee, et al. on Feb. 12, 2008; all of which are incorporated hereinto by reference.

[0067] FIG. **4**B illustrates an electrical arrangement suitable for use in charging a battery. DC IN **40** is buffered by capacitor bank **41** before entering into a step down transformer **43**. Step down transformer **43** reduces the voltage so that the voltage can safely be introduced into battery **44** which is connected to ground **45** at the battery's other pole.

[0068] Those of ordinary skill in the art readily recognize a variety of batteries which will work in this capacity, including, but not limited to those described in: U.S. Pat. No. 7,378,181, entitled "Electric Storage Battery Construction and Manufacture" issued to Skinlo on May 27, 2008; U.S. Pat. No. 7,388,350, entitled "Battery with Electronic Compartment" issued to Wright on Jun. 17, 2008; U.S. Pat. No. 7,397,220, entitled "Connection Member and Battery Pack" issued to Uchida, et al. on Jul. 8, 2008; and, U.S. Pat. No. 7,375,492, entitled "Inductively Charged Battery Pack" issued to Calhoon, et al. on May 20, 2008; all of which are incorporated hereinto by reference.

[0069] In FIG. **4**C, DC IN **40** is fed into an adjustable rheostat **46** which is controlled by the controller so that the DC OUT **47** falls within a specified range.

[0070] FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

[0071] This type of conductive line is commonly called poly-wire and consists of multiple interwoven strands of plastic **50**A and **50**B woven into a cord or rope arrangement having intertwined therein exposed metal wires **51**A and **51**B. While this illustration shows two plastic strands and two metal wires, any number of possible combinations is possible.

[0072] The exposed metal wires **51**A and **51**B attract the atmospheric static charge and transmit the charge down to the base unit (not shown).

[0073] FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

[0074] This conductive line utilizes a tube **60** having an outer layer **62** of PET Film (Biaxially-oriented polyethylene terephtalate polyester film) which provides exceptionally high tensile strength and is chemically and dimensionally stable. The tube has an ideal diameter of between two and three inches.

[0075] An interior metal coating **61** provides an initial conduit for the flow of static charge. The static charge through the metal forces the tube to expand due to the repulsion experienced by like charges. Further, the flow of electricity causes the interior of the tube **60** to become ionized to provide an additional pathway for the atmospheric static charges to the base unit (not shown).

[0076] Because outer layer **62** provides a gas barrier, the resulting ionization is not dissipated by air currents, thereby providing a highly stable pathway.

[0077] FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

[0078] In this embodiment, controller box **70**, resting on insulating pad **72**, is in communication with the sensors as described above. Using the input from these sensors, when there is flow of electricity through the base unit, warning flashing light **71** is illuminated. To electrically neutralize the mechanism, switch **73** is activated to pass any existing current into the ground **74**.

[0079] FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

[0080] Referencing FIG. **8**A, enhanced electrical collection lead **82** is a wire mesh which is in electrical communication with conductive line **81** and balloon **80**. Because of the significant amount of metal exposed by enhanced electrical collection lead **82**, more static electricity from the atmosphere is drawn to the collection lead **82**, and then down conductive line **81** to the base unit (not shown).

[0081] Conductive lead **82** is positioned proximate to balloon **80**.

[0082] In FIG. **4**B, poly-wire **83** has enhanced electrical collection leads **84** wrapped therearound. Collection leads **84** have pointed ends **85**A and **85**B which have a propensity to attract more electricity than rounded ends do.

[0083] It is clear from the foregoing that the present invention captures an entirely new source of electrical energy.

What is claimed is:

1. A mechanism to tap an electrical source comprising:

a) a lighter than air balloon;

b) a base unit having a spool of conductive line on a winch motor, one end of said conductive material secured to said lighter than air balloon, said winch motor capable of selectively extending or withdrawing said conductive line from said spool, one end of said conductive line connected to said lighter than air balloon;

c) an insulator electrically isolating said conductive line from ground; and,

d) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load.

2. The mechanism to tap an electrical source according to claim **1**, further including a controller, said controller selectively operating said winch motor based upon a voltage level in said conductor.

3. The mechanism to tap an electrical source according to claim **2**,

a) further including an electrical flow sensor monitoring electrical flow through said conductor and generating an electrical flow indicia indicative of said electrical flow in said conductor, said electrical flow indicia communicated to said controller; and,

b) wherein said controller operates said winch motor such that said electrical flow indicia remains within a selected operating range.

4. The mechanism to tap an electrical source according to claim **3**, wherein said selected operating range is established by an operator.

5. The mechanism to tap an electrical source according to claim **4**, further including a warning light activated when said electrical flow indicia is non-zero.

6. The mechanism to tap an electrical source according to claim **3**,

a) further including a lightning sensor generating a lightning presence indicia indicative of lightning within a prescribed range, said presence indicia being communicated to said controller; and,

b) wherein said controller, in response to said lightning presence indicia, operates said winch motor to withdraw substantially all of said conductive line onto said spool.

7. The mechanism to tap an electrical source according to claim **6**, wherein said prescribed range is established by an operator.

8. The mechanism to tap an electrical source according to claim **3**, further including an electrical collection enhancement lead in electrical contact with a first end said conductive material, said electrical collection enhancement lead configured to attract static electricity.

9. The mechanism to tap an electrical source according to claim **8**, wherein said electrical collection enhancement lead is positioned proximate to said lighter than air balloon.

10. The mechanism to tap an electrical source according to claim **9**, wherein said electrical collection enhancement lead includes at least two pointed electrical conductors.

11. The mechanism to tap an electrical source according to claim **1**, wherein said insulator electrically isolates said conductive line from said winch motor.

12. The mechanism to tap an electrical source according to claim **1**, wherein said insulator electrically isolates said base unit from ground.

13. A mechanism comprising:

a) an airborne aircraft tethered by a conductive line secured to a winch capable of extending or withdrawing said conductive line from a spool, thereby adjusting an altitude of the airborne aircraft;

b) an insulator electrically isolating said conductive line from ground; and,

c) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load.

14. The mechanism according to claim **13**, further including a controller, said controller selectively operating said winch motor based upon a voltage level in said conductor.

15. The mechanism to tap an electrical source according to claim **14**,

a) further including a lightning sensor communicating with said controller; and,

b) wherein said controller, in response to selected signals from said lightning sensor withdraws substantially all of said conductive line.

16. A mechanism to tap an electrical source comprising:

a) a lighter than air balloon;

b) a base unit having a spool of conductive line on a winch motor, one end of said conductive material secured to said balloon, said winch motor capable of adjusting an altitude of said lighter than air balloon by selectively extending or withdrawing said conductive line from said spool;

c) an insulator electrically isolating said conductive line from ground; and,

d) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load;

e) a sensor array having,

1) an electrical flow sensor monitoring electrical flow through said conductor, and,

2) a lightning sensor monitoring existence of proximate lightning; and,

f) a controller responsive to signals from said sensor array to selectively operate said winch motor.

17. The mechanism to tap an electrical source according to claim **16**,

a) further including a warning light; and,

b) wherein said controller activates said warning light when said electrical flow in said conductor is non-zero.

18. The mechanism to tap an electrical source according to claim **17**, further including an electrical collection enhancement lead in electrical contact with said conductive material, said electrical collection enhancement lead configured to attract static electricity and positioned proximate to said lighter than air balloon.

* * * * *

Nov. 10, 1959 O. T. CARR 2,912,244

AMUSEMENT DEVICE

Filed Jan. 22, 1959 5 Sheets-Sheet 1

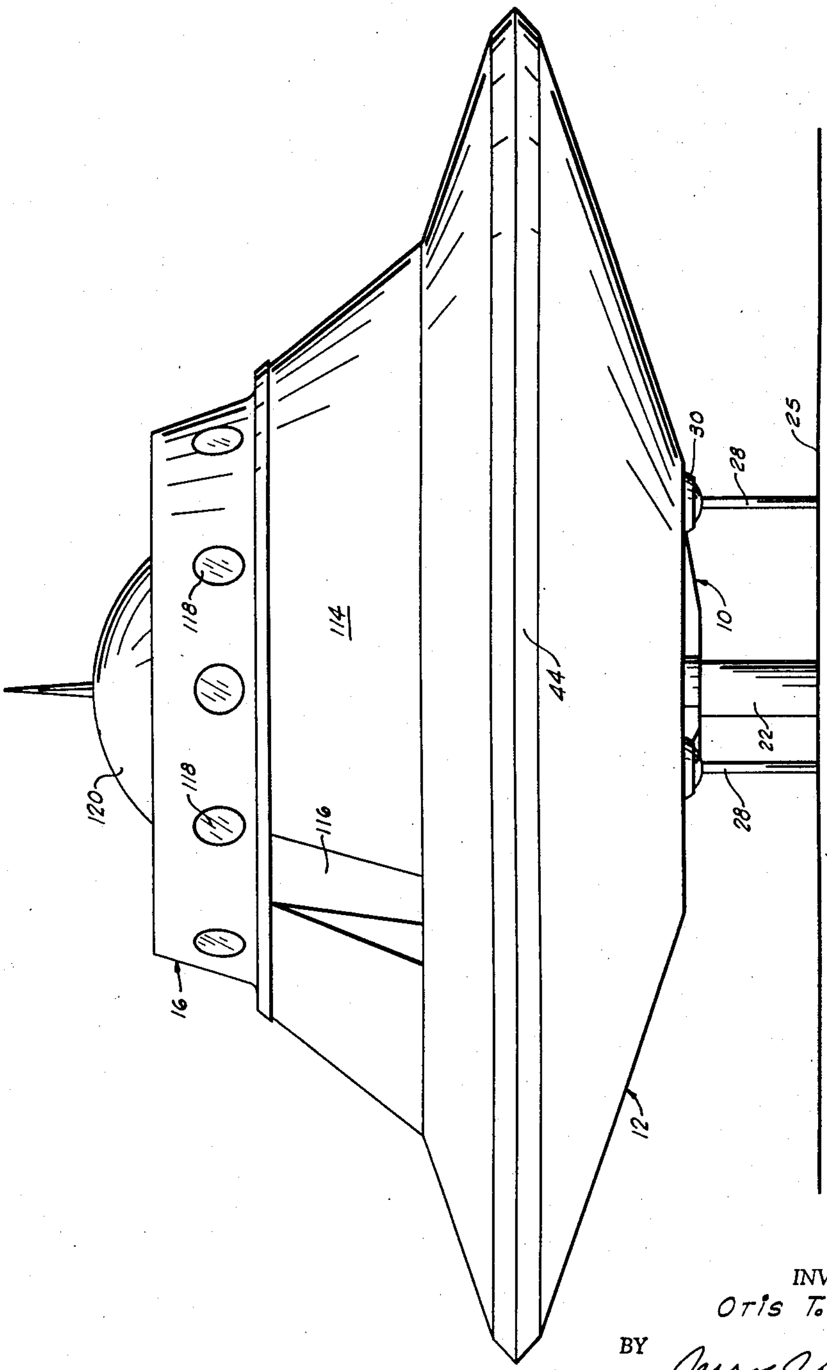
30
28
25
10
22
28
44
114
118
118
120
116
16
12

Fig. 1

INVENTOR.
Otis T. Carr
BY Jerry J. Dunlap
ATTORNEY

Nov. 10, 1959 **O. T. CARR** **2,912,244**

AMUSEMENT DEVICE

Filed Jan. 22, 1959 5 Sheets-Sheet 2

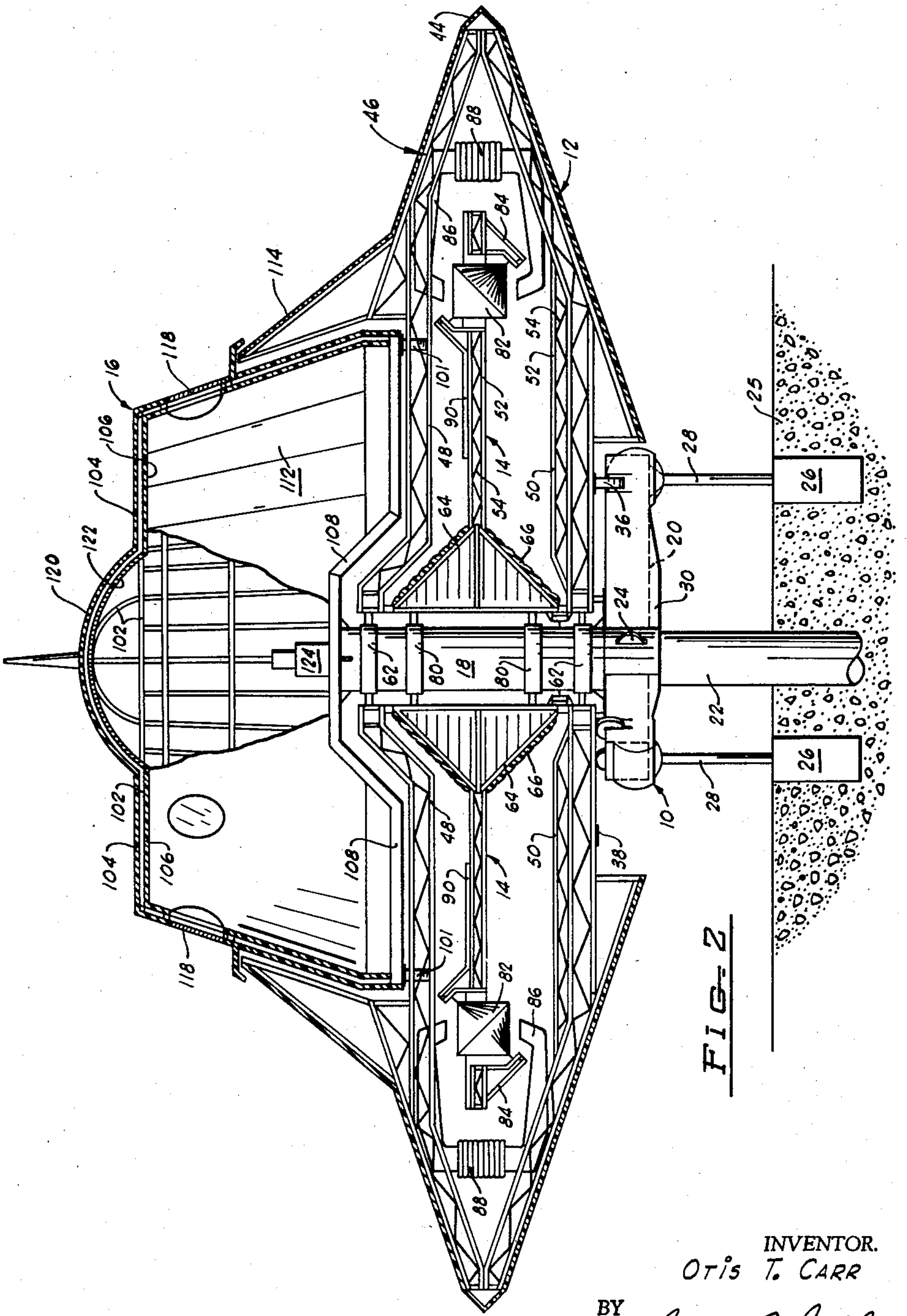
44
46
88
12
84
86
114
82
54
118
16
101
52
90
106
48
104
14
112
28
25
108
54
50
64
66
26
36
122
20
120
30
24
102
62
80
18
80
62
22
124
26
10
28
102
64
66
48
50
38
104
106
108
14
90
118
101
82
86
84
88

FIG-2

INVENTOR.
Otis T. Carr
BY Jerry J. Dunlap
ATTORNEY

Nov. 10, 1959 **O. T. CARR** **2,912,244**

AMUSEMENT DEVICE

Filed Jan. 22, 1959 5 Sheets-Sheet 3

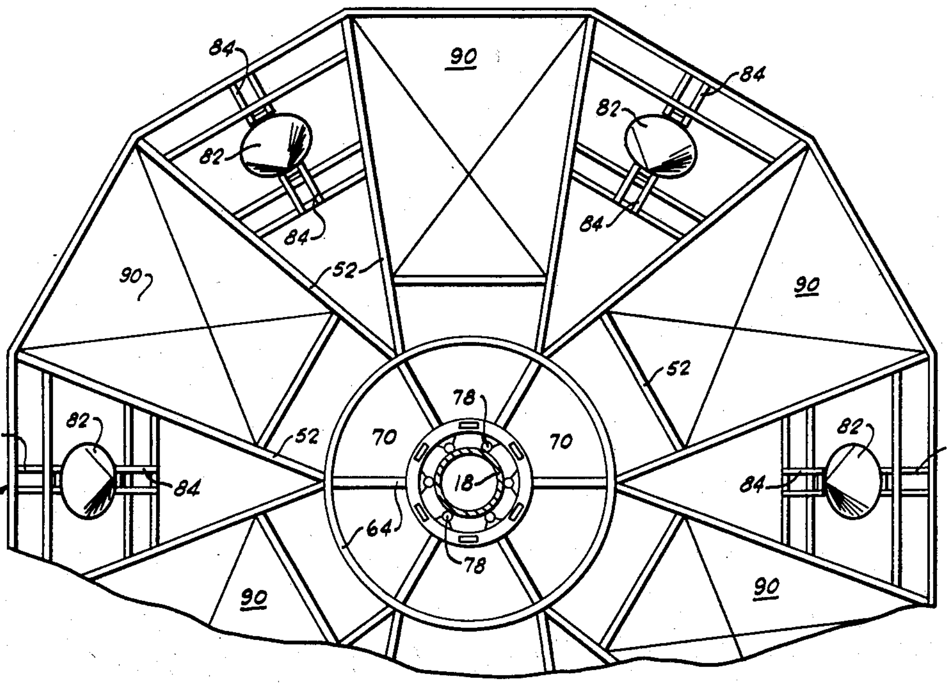
84
90
84
82
82
82
84
84
52
90
90
52
78
52
70
70
82
82
84
84
18
84
84
14
64
78
90
90

Fig. 5

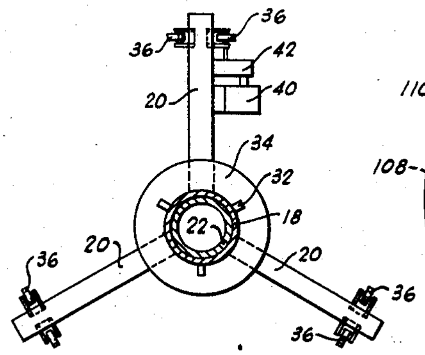
36
36
42
20
40
34
32
22
18
20
20
36
36
36

Fig. 3

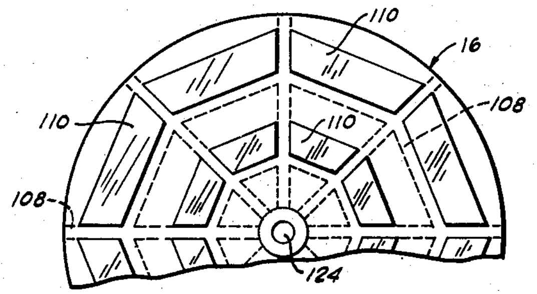
110
16
108
110
110
108
124

Fig. 7

INVENTOR.
Otis T. Carr

BY Jerry J. Dunlap
ATTORNEY

Nov. 10, 1959 O. T. CARR 2,912,244

AMUSEMENT DEVICE

Filed Jan. 22, 1959 5 Sheets-Sheet 4

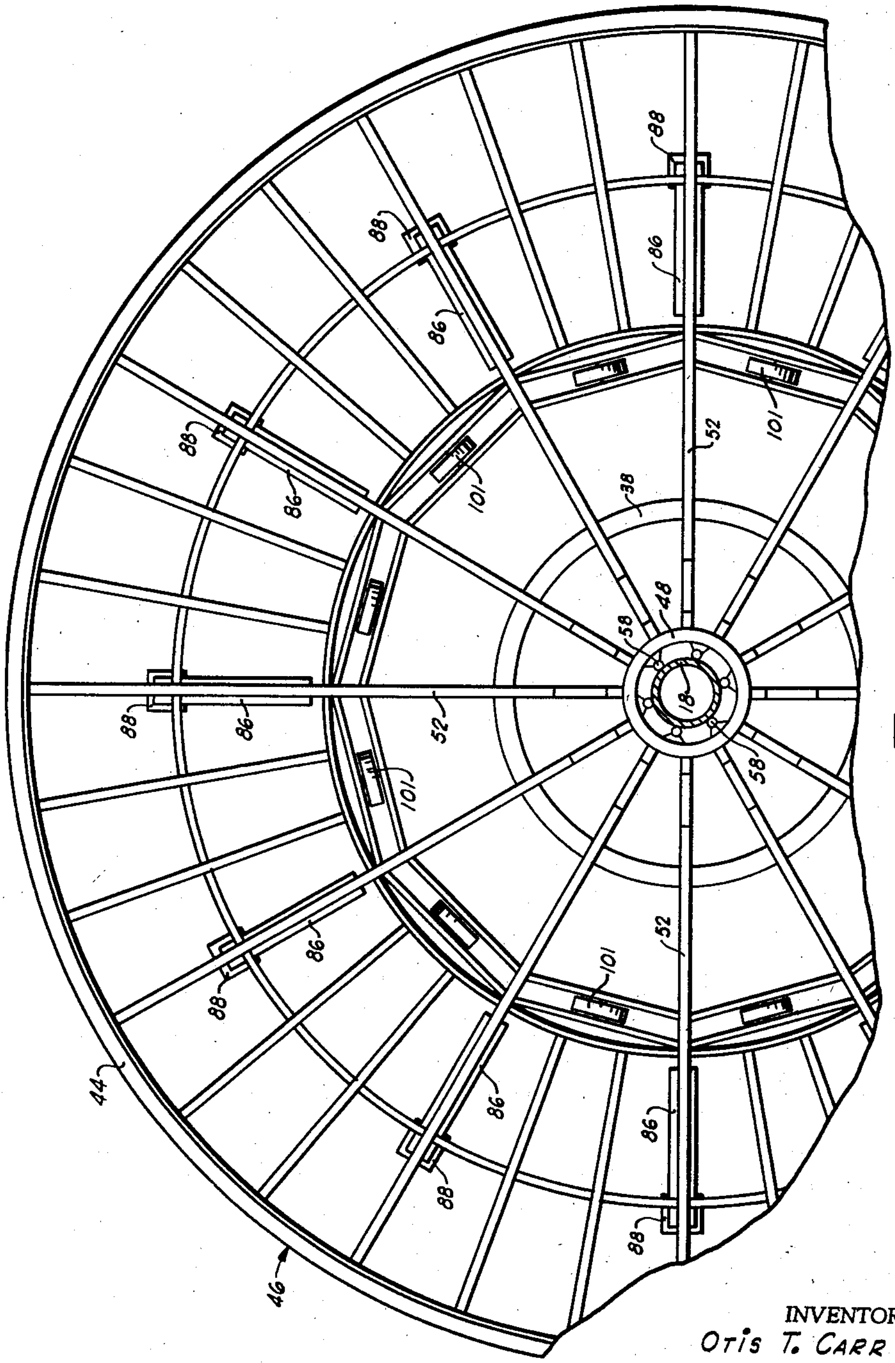
88
86
101
52
38
48
58
18
44
46

FIG. 4

INVENTOR.
OTIS T. CARR
BY Jerry J Dunlap
ATTORNEY

Nov. 10, 1959 **O. T. CARR** **2,912,244**

AMUSEMENT DEVICE

Filed Jan. 22, 1959 5 Sheets-Sheet 5

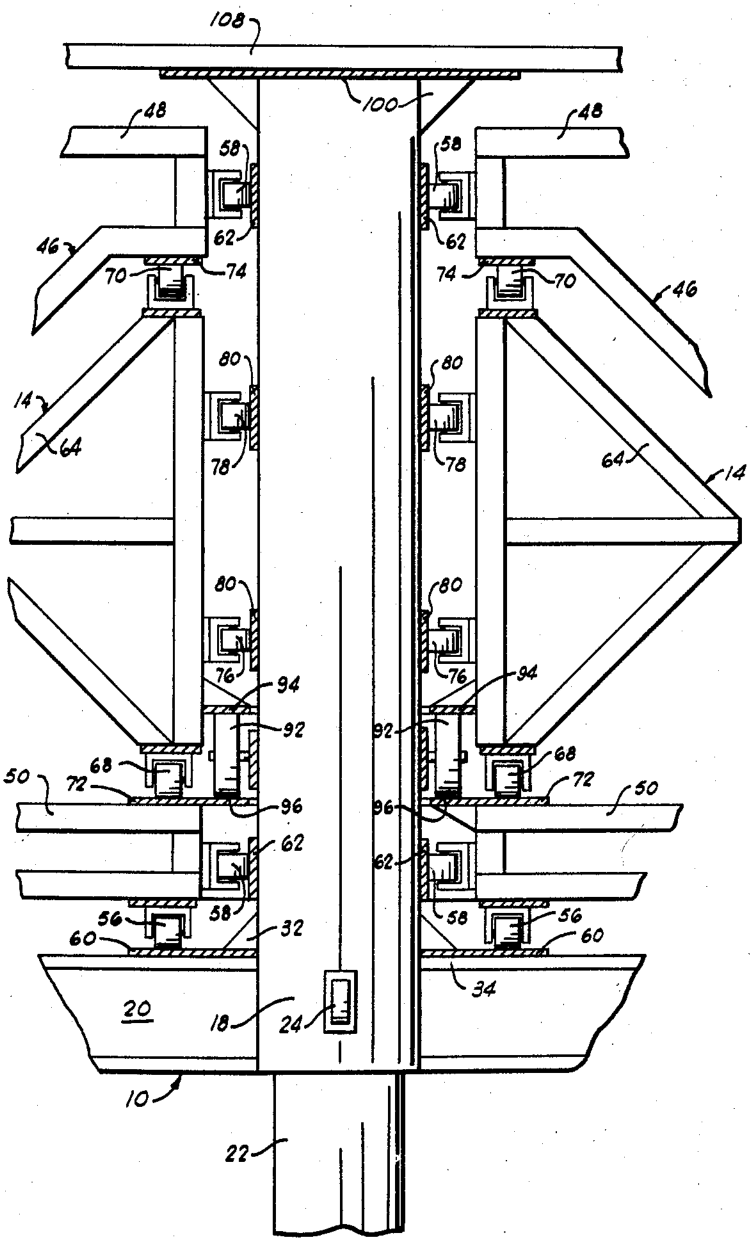
108
100
48
48
58
58
62
62
46
46
70
74
74
70
80
80
14
64
78
78
64
14
80
80
76
76
94
94
92
92
68
68
50
72
72
50
96
96
62
62
58
58
56
56
60
60
32
34
20
18
24
10
22

FIG. 6

INVENTOR.
Otis T. Carr

BY Jerry J Dunlap

ATTORNEY

2,912,244
Patented Nov. 10, 1959

2,912,244

AMUSEMENT DEVICE

Otis T. Carr, Baltimore, Md.

Application January 22, 1959, Serial No. 788,392
9 Claims. (Cl. 272—18)

This invention relates generally to improvements in amusement devices, and more particularly, to an improved amusement device of the type wherein the passengers will receive the impression of riding in an interplanetary space craft.

The present invention contemplates a novel amusement device having the overall configuration of a space craft and being formed in various sections, with portions of the sections being rotated in opposite directions to give the impression of movement to the passengers. It is also contemplated to move the craft up and down and display an animated movie of heavenly bodies above the passengers to give the passengers an impression of leaving the earth and approaching a distant planet or the like. This invention further contemplates the provision of electro-magnets, cones and the like carried by the oppositely rotating sections of the device, and windows through which the passengers may view these simulated objects to give the impression to the passengers that they are viewing the inner workings of an interplanetary space craft during flight of the craft. Finally, this invention contemplates the construction of the various portions of an amusement device simulating an interplanetary space craft in such a manner that the device may be easily transported in sections from one location to another and reassembled in a minimum of time.

An important object of this invention is to provide an amusement device wherein a passenger in the device receives the impression of flying in an interplanetary space craft.

Another object of this invention is to provide an amusement device having a general configuration of an interplanetary space craft and having oppositely rotating sections simulating the movement of various portions of such a space craft during flight.

A further object of this invention is to provide an amusement device simulating an interplanetary space craft and having a passenger cabin wherein the passengers may view various moving portions of the device to impress the passengers with flight in a space craft.

Another object of this invention is to provide an amusement device simulating an interplanetary space craft wherein the passengers will be physically moved a very minor distance and yet will receive the impression of flying in a space craft.

A further object of this invention is to provide a portable amusement device simulating an interplanetary space craft which may be easily disassembled and transported from one location to another.

A still further object of this invention is to provide an amusement device which is simple in construction, may be economically manufactured, and which will have a long service life.

Other objects and advantages of the invention will be evident from the following detailed description, when read in conjunction with the accompanying drawings which illustrate my invention.

In the drawings:

Figure 1 is an elevational view of an amusement device constructed in accordance with this invention.

Figure 2 is a vertical sectional view through the device illustrated in Fig. 1.

Figure 3 is a plan view of the landing gear for the amusement device, with the housing around the legs of the landing gear removed.

Figure 4 is a plan view of the supporting structure or frame work for the outer rotating shell of the amusement device shown in Figs. 1 and 2.

Figure 5 is a plan view of the central rotating assembly which rotates within the outer shell frame illustrated in Figs. 2 and 4.

Figure 6 is an enlarged sectional view illustrating the support and the drive between the inner rotating assembly and the outer rotating assembly.

Figure 7 is a plan view of the floor portion of the passenger cabin.

Referring to the drawings in detail, and particularly Figs. 1 and 2, a preferred embodiment of the present invention comprises, generally speaking, a supporting structure and landing gear **10**, an outer rotating shell **12**, an inner rotating assembly **14** (shown only in Fig. 2) and a passenger cabin **16** extending downwardly into the upper end portion of the outer shell **12**. The landing gear and supporting mechanism **10** comprises a vertically extending tubular member **18** around which the outer shell **12** and inner rotating assembly **14** are journaled, as will be hereinafter set forth, and a plurality of legs **20** extending radially outward in circumferentially spaced relation from the lower end of the tubular member **18**. The tubular member **18** is telescoped over a supporting stanchion **22** which may be in the form of a pipe, and the tubular member **18** is slidingly supported on the stanchion **22** by suitable wheels or rollers **24** to facilitate vertical movement of the member **18** on the stanchion **22**, as will be hereinafter set forth. The stanchion **22** is preferably supported in a suitable foundation **25** to provide adequate strength for the amusement device.

A suitable jack **26**, preferably a hydraulic jack, is anchored in the foundation **25** underneath the outer end portion of each of the landing gear legs **20** and is provided with an extension **28** extending upwardly into connection with the respective leg **20**. The jacks **26** are operated simultaneously to raise and lower the landing gear and supporting structure **10**, as will be described. A suitable shaped housing **30** is preferably secured around each of the landing gear legs **20** to provide a streamlined appearance. It may be noted that each of the legs **20** is preferably constructed out of a structurally strong member, such as an I-beam, and is secured to the lower end portion of the tubular member **18** by suitable braces **32** and plates **34** to assure a rigid and adequate connection of the legs **20** to the tubular member **18**.

A pair of rollers **36** (see also Fig. 3) are rotatably supported adjacent the outer end portion of each leg **20** of the landing gear and are arranged to engage a circular track **38** (Figs. 2 and 4) for partially supporting the outer shell **12** and inner rotating assembly **14** as will be described. Also, a suitable drive motor **40** (Fig. 3) is mounted on one of the legs **20** of the landing gear and is connected through a suitable gear box **42** to one of the rollers **36** to rotate the outer shell **12** with respect to the landing gear and supporting structure **10**. The roller **36** connected to the drive motor **40** may be suitably chained (not shown) to its companion roller **36** to assure an adequate drive connection between the drive motor **40** and the circular track **38**.

The shell **12** (Fig. 2) is tubular in configuration, with the diameter of the central portion thereof substantially larger than the diameter at the opposite ends thereof, and is supported in an upright position to give the general appearance of two saucers placed face-to-face. The outer edge portion 44 of the shell 12 is preferably formed of partially transparent material, such as Plexiglas, and is illuminated from within (not shown) to provide a band of light around the central portion of the shell 12. The remainder of the shell 12 is formed of any desired opaque and relative strong material, such as aluminum-laminated Masonite sheets.

The shell 12 is supported on a frame generally designated by reference character 46 and illustrated in Figs. 2 and 4. The frame 46 is circular in plan, as viewed in Fig. 4, and is substantially U-shaped in cross-section (as shown in Fig. 2) to provide an upper horizontally extending portion 48 and a lower horizontally extending portion 50 which extend from adjacent the support-member 18 outwardly to the central portion of the shell 12. The frame 46 may be easily formed with industrial square steel tubing 52 formed into a truss with bracing of steel rods 54. With this construction, the frame 46 may be easily fabricated in sections for convenient assembly, or the entire frame 46 may be lifted off of the supporting member 18 and transported to another location.

Support of the frame 46 and the outer shell 12 is provided by the track 38 (previously mentioned) which is secured on the lower face of the lower horizontally extending portion 50 of the frame 46 in a position to engage the rollers 36 carried by the landing gear legs 20. Additional support for the frame 46 is provided by rollers 56 (Fig. 6) secured around the inner edge of the lower portion 50 of the frame, and rollers 58 secured around the inner periphery of both portions 48 and 50 of the frame. The rollers 56 are arranged in circumferentially spaced relation around the inner periphery of the frame 46 and engage a circular track 60 supported by the legs 20 of the landing gear. The track 60 may be an extension of the upper plate 34 shown in Fig. 3. The rollers 58 engage tracks 62 extending circumferentially around the supporting member 18 to accommodate any radial forces imposed on the frame 46. It will be understood that any desired number of the rollers 58 may be provided around the inner periphery of the frame 46 to provide the necessary support. It will thus be apparent that when the drive motor 40 is placed in operation, the rollers or wheels 36 operated by the motor 40 will engage the track 38 to rotate the shell 12 in one direction around the longitudinal axis of the supporting member 18. The rollers 56 will engage the track 60, and the rollers 58 will engage the tracks 62 during such rotation of the outer shell 12 and frame 46.

The inner rotating assembly 14 is generally annular-shaped, as illustrated in Fig. 5, for movement around the supporting member 18 between the upper and lower horizontally extending portions 48 and 50 of the frame 46. The assembly 14 is also preferably constructed as a frame out of square steel tubing 52 formed into a truss with steel rods 54 for bracing, and is provided with a double conically-shaped framework 64 adjacent the supporting member 18. The framework 64 is provided to support a covering material 66 of any suitable type, such as laminated metal or plastic, to simulate a solid member in the central portion of the amusement device, which in turn simulates a portion of a space craft, such as the power unit for the craft. The covering material 66 may be corrugated if desired. The framework 64, as illustrated in Fig. 6, is also provided to support the rotating assembly 14 on the supporting member 18 and the frame 46. A plurality of rollers 68 are provided around the lower face of the framework 64, and a plurality of rollers 70 are provided around the upper face of the framework 64 to engage circular tracks 72 and 74, respectively, which are mounted on the frame 46. The rollers in each set of rollers 68 and 70 are provided in circumferentially spaced relation around the inner periphery of the rotating assembly 14 to accommodate vertical thrusts imposed on the rotating assembly 14. Additional rollers 76 and 78 are provided around the inner periphery of the framework 64 to engage tracks 80 extending circumferentially around the supporting member 18 to accommodate radial thrusts imposed on the central rotating assembly 14. It will be apparent that any desired number of rollers 76 and 78 may be provided around the inner periphery of the framework 64 to adequately support the rotating assembly 14.

I also prefer to support a plurality of conically-shaped members 82 in the outer edge portion of the central rotating assembly 14. Each conically-shaped member 82 may be easily formed out of a suitable sheet material, such as laminated aluminum or plastic, and is supported at its opposite ends by suitable trunnions 84 to extend through the truss comprising the frame of the central rotating assembly 14. The conically-shaped members 82 are preferably provided in circumferentially spaced relation around the assembly 14, as illustrated in Fig. 5, and are positioned to move through simulated electro-magnets 86 supported around the frame 46, as illustrated in Figs. 2 and 4. Each simulated electro-magnet 86 may be formed in horse shoe shape out of any suitable sheet material, and is provided with tubing or the like 88 around a portion thereof to simulate the winding of an electro-magnet. I further prefer to secure a plurality of plates 90 in spaced relation around the upper face of the rotating assembly 14 to simulate capacitor plates in a space craft. The plates 90 may be formed out of any suitable material, such as aluminum laminated Masonite, to provide a striking appearance.

The inner rotating assembly 14 is driven by a friction or dead wheel 92 pivotally supported on the tubular member 18, as illustrated in Fig. 6, engaging circular plate 94 and 96 carried by the inner rotating assembly 14 and the frame 46, respectively. The plates 94 and 96 are arranged in vertically spaced relation with sufficient distance therebetween to frictionally engage opposite portions of the wheel 92. It will therefore be apparent that when the frame 46 is turned in one direction, the plate 96 will engage the periphery of the wheel 92 to drive the wheel 92 and in turn drive the central rotating assembly 14 in an opposite direction by frictional engagement of the wheel 92 with the plate 94. Therefore, the various conically-shaped members 82 will be moved successively through the simulated electro-magnets 86 and give the appearance of the generation of electrical energy.

The cabin 16 is supported on the upper end of the tubular supporting member 18 and is rigidly secured to the member 18 by bracing 100 to move with the supporting member 18 and to be prevented from rotating with the shell 12 or rotating assembly 14. Additional support is provided by a plurality of rollers 101 carried by the frame 46 engaging a complementary track on the bottom of the cabin. The cabin 16 is constructed out of a suitable framework 102 covered by suitable sheet material 104, such as aluminum laminated with Masonite, on the outer surface of the framework and any other suitable material 106, such as Masonite, around the inner surface of the framework. It will be understood that substantially strong members 108 must be provided in the floor of the cabin 16 to adequately support passengers in the cabin. As illustrated in Fig. 7, a plurality of transparent windows 110 are provided adjacent the outer and inner peripheries of the floor of the cabin 16, through which passengers in the cabin may adequately view the inner rotating assembly 14 and the frame 46.

Passengers may either stand or sit in the cabin 16 and may gain entrance and exit from the cabin through suitable doors 112 as illustrated in Fig. 2. It will also be noted that the upper edge portion **114** of the outer shell **12** is extended upwardly to overlap the lower end portion of the cabin **16**. A portion of the upper edge **114** of the shell **12** is removed, as at **116** in Fig. 1, such that passengers may enter the cabin **16** through the doors **112**. Simulated port holes **118** are provided around the upper edge portion of the cabin **16** to further simulate an interplanetary space craft. It will be noted, however, that the port holes **118** are preferably covered with an opaque material on the inside of the cabin **16** to prevent the passengers from seeing familiar objects outside of the cabin **16**.

A dome **120** is preferably provided in the top central portion of the cabin **16** and the under surface thereof is covered with a suitable material **122**, such as cloth or the like, which will display a movie scene. A projector **124** is suitably mounted in the central portion of the floor of the cabin **16** to project an animated movie onto the screen **122** during operation of the amusement device. The movie displayed by the projector **124** is preferably an animated movie of heavenly bodies made in such a manner that the heavenly bodies displayed on the screen **122** will appear to come closer to and move away from the passengers during operation of the amusement device to give the impression of interplanetary flight.

In summarizing the operation of the preferred amusement device, the passengers are directed into the cabin **16** through the opening **116** in the upper edge portion of the shell **12** and through the doors **112**. The doors **112** are then securely fastened in closed positions and the passengers are thereafter prevented from viewing any object outside of the amusement device during the simulated flight. The drive motor **40** is then placed in operation to rotate the shell **12** and frame **46** in one direction, while the friction wheel **92** drives the inner rotating assembly **14** in an opposite direction. Passengers in the cabin **16** will then view the movement of the inner rotating assembly **14** and the frame **46** through the windows **110** to receive the impression of watching the inner workings of a space craft. Also, suitable lighting effects (not shown) are preferably provided in the shell **12** to illuminate the moving parts. Simultaneously with rotation of the shell **12** and the inner rotating assembly **14**, the jacks **26** are actuated to raise the supporting member **18** on the stanchion **22**. This upward movement of the supporting member **18** will be rather minor, but will be sufficient to give the sensation of rising to the passengers in the closed cabin **16**. It will be observed that when the supporting member **18** is raised, the shell **12** and inner rotating assembly **14** are simultaneously raised to retain the vertical relationship between the cabin **16** and the remainder of the apparatus viewed by passengers in the cabin. Also, simultaneously with rotation of the shell **12** and rotating assembly **14**, the movie is displayed by the projector **124** on the screen **122** on the dome of the cabin to give the passengers the impression of approaching one or more heavenly bodies. The jacks **26** are then operated in the opposite direction to lower the cabin **16** and the movie projected by the projector **124** may then provide an impression on the screen **122** that the passengers are returning to earth.

From the foregoing it will be apparent that the present invention provides an amusement device which will give the impression and sensation to patrons or passengers of the device that they are taking an interplanetary flight. The device includes counter-rotating sections simulating the inner workings of a space craft and these moving sections may be easily viewed by passengers in the device. Also, the passengers are slightly raised during operation of the device to simulate flight from the earth and a movie is simultaneously displayed to give the impression of approaching a heavenly body. It will also be apparent that the construction of the amusement device is such that the device may be easily disassembled and moved from one location to another. It will further be apparent that the present amusement device is simple in construction, may be economically manufactured and will have a long service life.

Changes may be made in the combination and arrangement of parts or elements as heretofore set forth in the specification and shown in the drawings, it being understood that changes may be made in the embodiment disclosed without departing from the spirit and scope of the invention as defined in the following claims.

I claim:

1. In an amusement device simulating a space ride, the combination of: an outer tubular shell, the diameter of the central portion of said outer shell being substantially larger than the diameter at the opposite ends thereof, means for supporting the outer shell in an upright position with the largest diameter thereof extending horizontally, means for rotating the outer shell about its longitudinal axis, a plurality of simulated electro-magnets carried in circumferentially spaced relation in the outer shell, an annular-shaped frame rotatably supported concentrically in the outer shell, means for simultaneously turning said frame in a direction opposite to the rotation of the outer shell, conically-shaped members carried by said frame in positions for movement through the simulated electro-magnets in the outer shell, a passenger cabin having a floor therein, means for non-rotatably supporting the cabin above the outer shell with the floor of the cabin positioned within the upper end portion of the outer shell, and transparent windows in the floor of the cabin for viewing movement of the outer shell and said frame from the interior of the cabin.

2. An amusement device as defined in claim 1 wherein said means for supporting said outer shell comprises a tubular support member extending downwardly from the cabin through the outer shell and said frame, a base portion on the lower end of said tubular support member simulating a landing gear, cooperating rollers and tracks supporting the outer shell on the base portion, and means for raising and lowering said base portion during rotation of the outer shell and said frame.

3. In an amusement device simulating a space ride, the combination of: a vertically extending supporting member, a passenger cabin mounted on the upper end of the supporting member and having transparent windows in the floor thereof, an outer shell extending downwardly and outwardly from the lower end portion of the cabin and then inwardly and downwardly toward the supporting member, a first frame extending radially inward from the outer shell and having a generally U-shaped cross section with upper and lower horizontally extending portions, means rotatably supporting said first frame on the supporting member, a second frame rotatably supported on the supporting member for rotary movement horizontally between the upper and lower horizontally extending portions of the first frame, and means for rotating said first and second frames in opposite directions to simulate movement of the inner workings of a space craft as viewed from the cabin.

4. An amusement device as defined in claim 3 characterized further to include at least one jack connected to said supporting member for raising and lowering the cabin and first and second frames during rotary movement of said frames.

5. An amusement device as defined in claim 3 wherein said supporting member is a tube, and characterized further to include a fixed stanchion telescoped into the supporting member for holding the supporting member vertical, legs extending outwardly in circumferentially spaced relation from the supporting member below the lower horizontal portion of the first frame simulating a landing gear for the device, rollers carried by each leg, a circular track on the lower horizontal portion of the first frame arranged to ride on said rollers during rotation of the first frame, and a jack connected to each leg of the

landing gear for raising and lowering the supporting member on the stanchion during rotation of said frames.

6. An amusement device as defined in claim 5 wherein said means for rotatably supporting the first frame on the supporting member includes a track around the supporting member opposite each of the upper and lower horizontal portions of the first frame, and a plurality of rollers carried by each of the upper and lower horizontal portions of the first frame in circumferentially spaced relation in positions to engage the respective track.

7. An amusement device as defined in claim 5 wherein said means for rotatably supporting the second frame on the supporting member comprises a pair of vertically spaced tracks around the supporting member, two sets of circumferentially spaced rollers carried by the second frame in positions to engage said tracks on the supporting member, circular tracks on the upper and lower horizontal portions of the first frame, and circumferentially spaced rollers on the upper and lower faces of the second frame for engaging the last-mentioned circular tracks.

8. An amusement device as defined in claim 3 wherein said means for rotating said frames in opposite directions comprises concentric annular plates secured around the first and second frames in vertically spaced relation adjacent the supporting member, a friction wheel carried by the supporting member in a position to engage said plates and drive the second frame in one direction upon rotation of the first frame in an opposite direction, and means for rotating the first frame.

9. An amusement device as defined in claim 3 characterized further to include a movie projector in the cabin arranged to project a movie of heavenly bodies on the interior of the cabin and provide passengers in the cabin with a sensation of interplanetary space flight.

References Cited in the file of this patent

UNITED STATES PATENTS

754,532	Austen et al. ----------	Mar. 15, 1904
868,369	Stewart et al. ----------	Oct. 15, 1907
981,672	Napp -----------------	Jan. 17, 1911
1,818,647	Pronnecke --------------	Aug. 11, 1931

rexresearch.com

Otis T. CARR

UTRON

readmystock
shutterstock

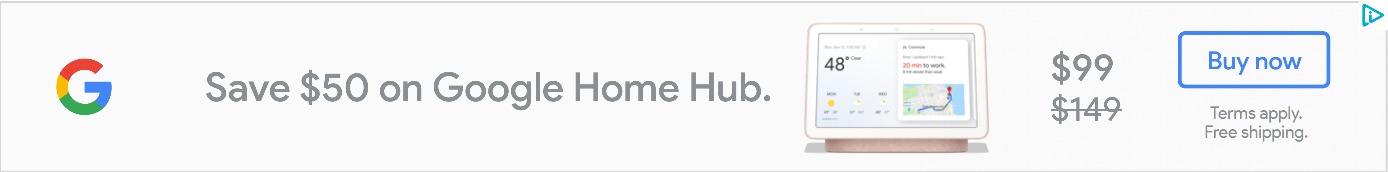
Save $50 on Google Home Hub.
$99
~~$149~~
Buy now
Terms apply.
Free shipping.

Introduction
Interview: "Long John" Nebel & Otis Carr, *et al.* (WOR Radio, NY, 1959)
Information Bulletin #3 (December 23, 1957)
US Patent # 2,912,244: Amusement Device
OTC-X1 Model Construction Plans
Fate Magazine (May 1958): "Gravity Machine?"
Fate Magazine (August 1959): "The Saucer That Didn't Fly"

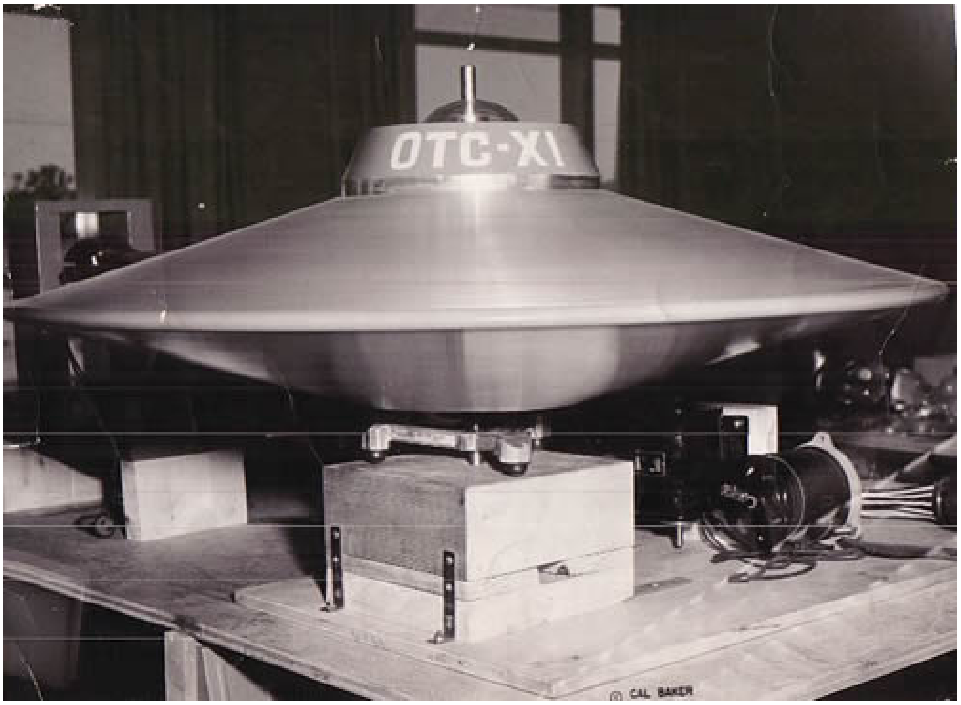

[Click Photo to View Large Image]

Introduction

In the late 1950s, Otis Carr claimed to have discovered what he called the "Utron", a "molecular shell" that houses the atoms. The cohesion of molecular shells is brought about by inertial systems in the gravity field, manifesting in the two poles of natural magnetism and electromagnetism. He claimed that by "hatching" the atomic egg, particularly of paramagnetic and diamagnetic elements (especially bismuth), atomic energy could be released gently. The Carr affair is generally considered to be a fraud. This file is a collection of excerpts of technical factoids from articles, interviews, etc., on the subject, without consideration of Carr's botched attempts to make a business of it. The technology actually makes some sense (to me, anyhow, with qualifications and reservations). I was told by Ralph Ring (one of Carr's engineers) that he himself had flown in a short test flight of a small model of the craft. Carr's business methods (fraudulent stock offerings and amateurish publicity tainted with hubris and non-performance) doomed his efforts. His engineer Norman Colton was an alcoholic who stupidly threatened a local newspaper publisher, thus generating more negative feedback. It was a sad pass from Carr's earlier statement that "We shall not stop until we have penetrated our great discoveries into all the ends of the earth, or even as far into outer space as man desires to go!"

The historic details as published in various newspaper articles and particularly in *True, The Man's Magazine* ("King of the Non-Flying Saucers", January 1961), are too long to be transcribed here for any useful purpose. This file is composed only of technical factoids gleaned from the collection of papers offered by Rex Research (albeit as poor-quality photocopies, caveat emptor), plus a couple of short articles cribbed from Keelynet/Fate magazine.

According to Carr, "Any vehicle accelerated to an axis rotation relative to its attractive inertial mass, immediately becomes activated by free-space-energy and acts as an independent force... We have shown that a charged body, accelerated to an axis rotation relative to this attractive inertial mass, indicates polarity in a given direction. The dip-needle points, say, up toward the top of the body. But mount this while rotating body, with its spindle, on another platform and rotate this platform on a spindle, then if the counter-rotation is greater than the inertial forward rotation of the body, a dip-needle on the second platform will point down while the first dip-needle points up, indicating complete relativity of polarity. When the exact counter-rotation matches the forward rotation the body loses its polarity entirely and immediately becomes activated by free-energy (tensor stresses in space) and acts as an independent force... The above-described assembly of counter-rotating charged masses becomes weightless and will escape the immediate attraction of gravitational forces..."

The Utron

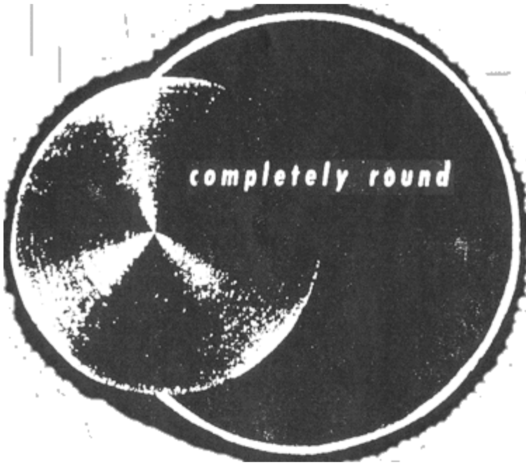

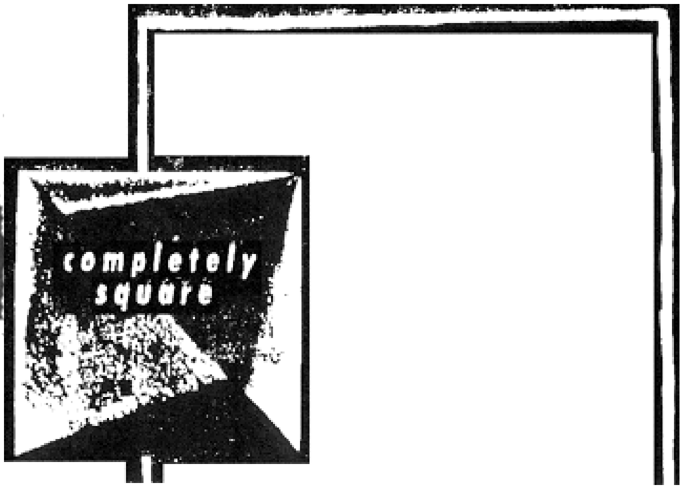

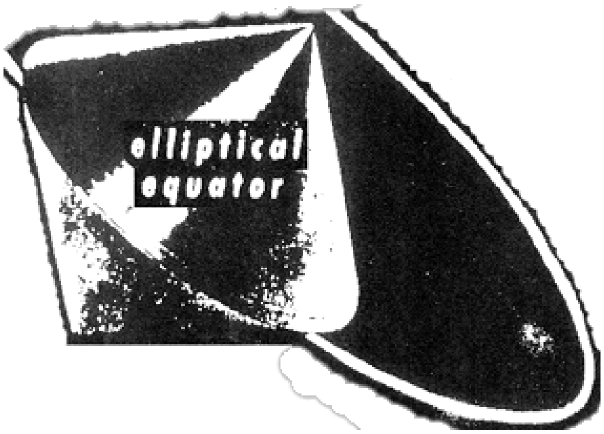

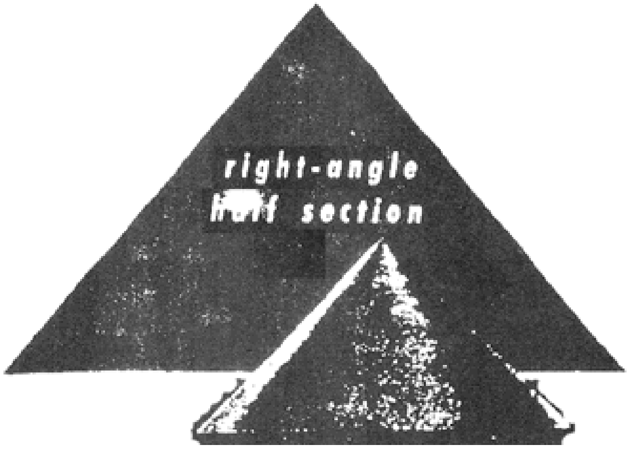

Interview: "Long John" Nebel (WOR Radio, NY) with Otis Carr, *et al.* (1959) ---

OTC (Otis T. Carr): "Our system utilizes gravity, electromagnetism, and electromotive force and a relative field to get its functional operation. We use an electrified sender. It's a sensor power core. Now this is what we call an accumulator... It is a storage cell, an accumulation of storage cells which provide an electromotive force in the same manner that any known battery produces and electromotive force...

This [the Utron] is a dimensional product. It was designed with the dimensions of space itself. We say it is truly the geometric form of space, because it is completely round and completely square. It has been proven in scientific laboratories that the very smallest unit of mass matter ever photographed in the electron microscope are square in shape... We have applied this principle into an electrified system, which is the power core of our space vehicle. Now what makes this unique and novel from a battery is the fact that this is a piece of moving machinery that rotates. Our average storage battery is an inanimate object set in an inertial spot and then the electromotive force is conducted by wires from this battery to animate some object.

LJ ["Long John" Nebel]: Let me interrupt for a moment; I'll try to describe it further. Well, it looks like two ice cream cones put together at the wide ends, but he angle is a lot wider [90 degrees] than that of the ice cream cones. There are a series of ridges; they look like gears would fit in. Is that correct?

OTC: No, those are in a sense turbine; they are reactive channels and where there is atmosphere a flow of air there aids in rotation.

LJ: Well, then, this is one of the components of the drive; this is the battery.

OTC: This is the center power core.

LJ: This generates electricity?

OTC: This is right. This is a storage cell for electrical energy. In operation it generates electricity at the same time it puts out electromotive force. This is the central power system for our spacecraft.

LJ: I can tell this: that it opened up and it appears to be hollowed out on the inside... it's circular, the inside; when the two parts are put one on top of the other and they fit into place, the cavity inside is circular...

OTC: It is a sphere, yes. And each unit is a hemisphere. We call the center of this large dimension the equator and of course it contracts and expands to a point on each side It's the union of two presto-conical sections; that is what it is Two right angle sections, and we say it is the dimensions of space and we have shown how this comes about...

LJ: In the literature you've provided us with, there's a mathematical formula that crops up here that intrigues me: minus zero divided by plus zero equals zero. Can you tell me more about it and how you stumbled on this idea?

OTC: The equation is brought about by the shape of our Utron electrical accumulator; this is the name given to our central power system. In our operation of working models and in checking out experiments, we had to find the formula that fir the reason for the action and reaction we were getting. So in exploring nature and studying the great inspirational work of the Dr. Einstein on relativity, we came upon this formula of linear correlation. And when we study linear correlation in geometric form we have to have a starting point and this is the point. And from there it explains through the cross and through the circle. And the only mathematical way we can express it is in the symbolism of zero X or 0X, and this formula brings us to that. We claim that this is the true unified field theory in physical practice.

In a further study of Dr Einstein's great inspirational work --- and we corresponded with him and we had the great good fortune of being advised by him at one time --- we learned that all measurements of time and space had to be considered in relationship to the observer and therefore there never was a fixed quantity of any number. Now in physical form, this is something else again...

On this craft, insofar as the individuals are concerned, we can travel in the same manner as a pressurized airliner. We don't have the problem of a heat shield... because the electromagnetic system sets up a protective shield in our craft which enables us to overcome this [thermal] barrier without any discomfort to the occupants of the craft. And we can very slowly rise, and once we are outside the atmosphere we can accelerate to tremendous velocities up to the speed of light itself.

LJ: How would it land, sir?

OTC: Very simply, we can fall at a very slow velocity of 100 feet/minute or less and we can set it down as gently as a feather because we have joined part of the operation of our craft to universal systems. This is a relative velocity craft and the minute it reaches the relative velocity of the attractive inertial mass, it becomes weightless as regards this inertial attraction. Individually, it is not weightless; t has the same weight as before, but when it reaches the relative location it becomes an independent system just as a planet is an independent system.

LJ: Is there any gravitational pull at this point, sir?

OTC: None whatsoever. They will have the same feeling of pressure or weight that they have right now because we will maintain as far as possible the atmospheric pressure of the earth at sea level inside the craft.

LJ: I am under the impression that the only reason I'm able to sit in this chair is because of gravitational pull.

OTC: We have around 14 lb/square inch pressure within our atmosphere. We have been able to be sealed off away from such a condition and then artificially with atmospheric pressure the pressure in the cabin is maintained. We have it very well in submarines. The same may be used in our craft.

LJ: In other words, under sea where a submarine may be there is no gravitational pull; is that what you're saying?

OTC: There is gravitational pull at all times, but we are speaking about the atmosphere of the occupants inside a sealed unit.

LJ: Is that necessary to keep the occupants in the position they desire?

OTC: Absolutely, because in a vacuum they are at the mercy of any velocity.

LJ: What would happen, sir, if there were some kind of instrument that you could turn on and eliminate the gravitational pull that was in this room?

OTC: You would in a sense become very buoyant and this is not in itself a novelty but it certainly does not have any disastrous effect on humanity.

LJ: Would I remain in this position?

OTC: You could, but any movement would move you out of it.

LJ: Would objects remain in position?

OTC: Until they were brought into any movement. Any movement would make them buoyant themselves.

LJ: I have a lead pencil; if I hold it in the air and release my fingers it would fall because of gravitational pull. If we had this other condition which you so aptly described a moment ago, if I released my fingers would the pencil remain in midair?

OTC: This is true; it would stay there.

LJ: I believe what you're saying is that you'd be creating an artificial gravitational field within the body of the spacecraft and yet there would not be any gravity on the outside?

OTC: Exactly correct.

LJ: And this is done by the battery which I attempted to describe, spinning around and producing its own gravitational influence?

OTC: Yes, this is the beginning of an answer to your question: we have capacitor plates and electromagnets as a part of this system. Now, this is counter-rotating; the electromagnets rotate in one direction, and the batteries rotate in another. The capacitor plates rotate in conjunction with the battery so that we have a clockwise and counterclockwise rotation. Now the third system is the cabin that maintains the crew. This does not rotate; it is fixed due to the fact that the two bodies are rotating clockwise and counterclockwise. Therefore the system causes the craft to escape from the gravity pull. The craft itself due to this system still has internal gravity because it still has the same weight that it had in the beginning.

LJ: What charges this battery?

OTC: This starts out electrochemically the same as other batteries, but we do have a regenerating system that is very unique. We are able here, for the first time to our knowledge, to use atmospheric electricity as a recharging system. This is done as part of the operational principle of the craft.

LJ: You say you use atmospheric electricity. What happens when you leave the atmosphere?

OTC: We have electrochemical systems to provide us with all the energy that we need and have a regenerating system in the manner of a regenerative coil that recharges this battery in the same manner that the storage battery in the automobile is recharged now, by a generator.

LJ: What you have doe is made the first perpetual motion machine.

OTC: There is nothing perpetual about our machine. The energies which cause it to operate are perpetual. You cannot destroy matter, and you cannot destroy energy; molecular flow is perpetual and that has been proven in the laboratory. It has been proven that electricity itself is immortal. When we take away resistance we can set up a spark of electricity and it will continue to operate; therefore we have perpetual energy. No machine that we can conceive of made by man would be perpetual, but it is free energy. It's self-energizing and as long as all parts function and do not wear out this is truly a self-energizing machine.

Ben: About this formula: were using conventional algebraic methods?

OTC: No, we weren't. Partially conventional, but we were joining actual space forms. We arrived at satisfactory equations for ourselves which can be demonstrated.

Ben: Are the physical laws upon which your invention works expressed in mathematical terms?

OTC: Possibly, but I wouldn't say that I'm qualified. We're satisfied with this formula.

Ben: Well, it's like saying plus 4 divided by minus 4 equals 4.

OTC: Sometimes these solutions are not always what appear. As we know in synergy, we know 1 plus 1 equals 3.

Ben: 1 plus 1 equals 3, when?

OTC: Because two conditions always produce a third.

Ben: the third condition is 2, isn't it?

OTC: Not necessarily.

Ben: Could you go over these conditions?

OTC: If one condition operates one way and another operates another way, when you join them you have another condition and their sum is 3.

Ben: Well, that is a little bit over my head. I've been looking at this prototype you have here and I notice a wooden frame or scaffolding you have a larger model of that you have on a turbine and around it is a wooden ring and it seems to be filled with electromagnets.

OTC: That's right; this is a wood model of an operational model. What we have here is the cones, our Utron electric accumulator; that is the power system. This system activates the electromagnets and in turn activates.

Ben: Does the system --- the thing inside --- activate the electromagnets on the outside?

OTC: This is true. We do this by contacting this lead wire from the positive and negative poles of these batteries to the electromagnets and then we have circuit breakers from these electromagnets and we have counter-rotation. These electromagnets will rotate counter-clockwise while the internal area is rotating clockwise.

Ben: Are these spools of wire on the model itself, are they magnetized also?

OTC: The coils of wire inside the ring are regenerative coils; they are electromotive force coils and they assist in regenerating the battery, because they are loops of wire brought through a magnetic field which sets up an electromotive force. These others are capacitor plates and these are also activated by the central power core, but these plates, which can accept a very high charge in neutral conductance also through the process of ionization utilize atmospheric electricity.

Ben: The fact that you have two cones meeting at their bases, so that you have something that is also square. I don't see how you get a square.

OTC: Dimensionally it is; it is square in these dimensions and when this rotation starts and builds up to a certain velocity, this form is very important because we have the total equation of action and reaction. Now this is done by a system of coil winding wherein we start at a point, expand to an equator, and continue our winding down to a point. With this physical expansion and contraction, is an electromagnetic field. Gravity enters the picture in the form of this relative rotation. When the relative rotation ---- inertial effective mass ---, it's a matter of dimension. So that is the earth, say, is 8,000 miles in diameter, we know its fixed rotation is 1 in 24 [blurred text, possible additional digit]. If we were 1 mile in diameter its rotation would be 8000 in 24 [blurred text, possible additional digit]. And by the same system, our 45-foot craft has a rotation of 580 rpm and when it reaches this rotation it is totally independent of its inertial attractive mass, in an electromagnetic field.

Ben: If your models get up to 580 rpm, will they take off?

OTC: This model was spun at 40,000 rpm and when it did it set a pressure pattern of 1,000 tons; the horsepower reading was a little over 700. Six engineers checked this out. Now the relative rotation of this model would be about 68,000 rpm, and when it reaches this rotation, it would immediately take off.

LJ: Would the time factor be involved with this craft?

OTC: In our solar system, the time factor would be involved, yes. We evaluate time on the velocity of light and in certain systems, if we exceed the velocity of light, unquestionably time would slow up.

Ben: Your craft can exceed the velocity of light?

OTC: We don't say this; I say I other systems.

Ben: I thought nothing could do this. I thought it was a constant factor, one of Einstein's factors.

OTC: Possibly in our system, but not necessarily true in other systems.

Ben: Anything approaching the sped of light becomes pure energy.

OTC: Pure energy, but in other systems it could change.

Ben: What other systems?

OTC: Other solar systems. We are completely controlled by our system and the velocity of light is our yardstick and our pattern and our craft is designed around this.

Ben: You don't just upset one of the basic principles of the universe.

Mel Salomey: Doesn't Einstein say any measurement is relative?

Ben: Except this one; it's the first axiom.

Mel: What is an axiom?

Ben: Self-evident truth.

Mel. Thank you. Wasn't Einstein theorizing; wasn't he assuming?

[Discussion of the speed of light, not included in transcript of the interview]

Roy: What are these coils? To describe this the best I can, imagine if you will a circle about 18 inches in diameter. Two circles that form a sort of a -- In other words, one circle fits over another circle and from the top of the circle and coming down from the other an inverted round pyramid. There are a number of what appear to be copper wound coils around the edge of it. And then there's a sort of a framework on top and underneath there's sort of braces. That's the general idea. I noticed these cones placed mouth to mouth with some coils, the coils on the edge of this thing; it revolves within this structure. A tough thing to describe.

Q: What is this over here, Mr Colton?

Colton: This is a paper mockup to show the counter-rotation principle and its outside section that Roy is describing looking at the other mockup. It contains the electrified horseshoe magnets. This would rotate in one direction counter-clockwise while the center section with the electron accumulator [Utron] would rotate clockwise in the other direction.

Q: Would these coils in the outer rim?

Colton: Rotate counter-clockwise -- It has a rotation speed relatively equal the earth's rotation speed, at about 580 rpm.

Q: When you say 580 rpm do you mean the outer rim is rotating on one direction 580 rpm and the inner rotating in the other direction at the same speed, giving a total rotation, one relative to the other, of 1160?

Colton: Exactly, although I don't know if 1160 has any bearing on it or not.

Q: Well, it would be twice the rotation in reference to the earth.

Colton: We're not giving it for a certain rotation for the sake of rotation but for the sake of relativity to the attractive mass. The earth at 8,000 miles diameter rotating once in 24 hours is relatively equal to a 45-foot craft rotating at 580 rpm, and 580 rpm would calculate to be the approximate rotational speed of an automobile tire on a car moving at about 25-30 mph.

LJ: Could a small craft be made to take off?

OTC: We plan to build a prototype model as a demonstration device. Now I would like to state that certain models have been built by me and tested. Each one has been airborne. One was lost entirely in space. We had a control system and this one didn't function. This has already been done.

Q: As was calculated, the speed of the circumference was 1263 mph. It gets kind of warm at that, doesn't it?

OTC: No, it won't because it has its own protection field which is its electromagnetic actuation. We described it as a self-contained unit. This ionization of the capacitor plates sets up a glow brilliantly with a very soft luminescent light.

LJ: What color?

OTC: It would be in the nature of blue-green or very similar to the electric arc you see in welding. This is the field we are testing. It does not have a heat barrier in forward velocity at all. This electromagnetic field is being tested out now in conventional aircraft and proved very efficient.

Q: Would you suppose you could put rockets toward the edges of the cone and have the thing spin by rocket propulsion?

OTC: We don't need it. We have a tremendous spin here. An electric motor operates the same way. You set up an electromotive force inside a magnetic field and you get rotation. So what we actually have here is an improved electrical motor which in itself is a circular device, and we say we make energy out of the air, from another dimension.

To clear your analogy up also, we would like to demonstrate the fact that this earth itself is literally a space craft demonstrating what we're talking about; it's rotating and orbiting at a certain constant speed with a magnetic field and it is in itself a spacecraft.

Q: Mr Colton, we assume the moon has a gravitational field. How does it make a gravitational field and yet it does not rotate on its axis.

Colton: It does not rotate on its axis?

Q: No, the moon does not rotate on its axis.

Ben: Sure it does, 1 rotation for 1 revolution.

Q: How long does it take?

Ben: 28 days.

Q: And the earth takes one day.

OTC: And that's how engineers and scientists have evaluated the velocity of the craft we call the earth by the orbit of it. Upon the pattern already set up, by the amount of time it takes the moon to rotate once around the earth from the center of the earth's core: 28 days. The distance being 245,000 miles. It is easy to calculate.

Q: These cones seem to revolve over intricately wound copper coils. Do you supply any motor power to this?

OTC: All energy comes from these two cones [Utron]. This in vernacular is a battery. The big novelty is that we have put a battery in motion. We have designed it within the accepted knowledge of total dimensions of space-matter and we have activated it electrochemically [electrolyte in the hollow center] and used the force through chemical activation to activate the entire craft, after which we have motion as the feature of this accumulator.

Ben: Mr Osgood's telegram brought up a very important and perhaps crucial point, mainly, James Clerk Maxwell demonstrated that light is an electromagnetic radiation, also verified by hertz who laid the foundation for modern radio. Now the speed of electromagnetic radiation such as radio waves also travels 186,000 miles per second. In other words, light and all forms of electromagnetic radiation travel at the speed of 186,000 miles per second. Now if it were possible for your craft to travel faster than the speed of light, it could, therefore, travel faster than the speed of electromagnetic radiation. So once it exceeded 186,00 miles per second, you wouldn't be gaining all this energy from this electromagnetic radiation you're generating and wouldn't you fall rapidly down?

OTC: There is a continuous falling in space which in itself can bring velocity and can bring you to another system. We mentioned conjecturally that in other systems, there could be different velocities. We're not applying them to our craft. We don't identify them wit our craft. Relatively, we could not go faster than the speed of light unless we were in a system that permitted it. In our solar system, which we have mentioned now 3 times, it is designed on known principles. We conjectured about other systems. If we go beyond the speed of light in other systems the conditions within that system would make it possible for us to have power.

Roy: Let's assume, Mr Carr, this vehicle is at rest. What is it that originally overcomes the inertia of this rotor and starts the rotor moving?

OTC: The electromagnetic force stored in the energy of the Utron electric-accumulator, which in vernacular is a battery.

Roy: Now the Utron electric accumulator is these cones that are inverted to each other with bases together. I know of zinc batteries, nickel-cadmium, lead-acid cells. Could you use those?

OTC: We could use any of the kinds you mentioned; what we have here is tremendous power size in comparison to other batteries; therefore, it's very easy to put 1000 2-volt cells inside this one unit as you see it. It has functioned very well. In our 45-foot craft we plant to have 12,000-volt batteries which will extend an electromotive force which will energize the electromagnets and the capacitor plates. The generative coils will put back into the batteries in this system the same amount of volts going out until there is a breakdown of electro-chemicals or wear-out of equipment. But it could last as long as average storage batteries in automobiles.

Roy: In order to start the motor off originally, is it necessary to cause a flow of electricity through one set of the coils?

OTC: That's true.

Roy: Is there any magnetic force in the other set of coils at this point?

OTC: They individually operate by circuit breakers and the first motion begins to start a repetition. The same we have in a motor that has the opposite of a commutator, which is an accumulation of contact points where each coil is energized as the current flows through this coil. Then this starts the motion, the repetition of this motion brings the whole motor into phase in the same sense our accumulator and magnets become speeded up and the circuits are made and broken as they rotate.

Roy: Where the inner rotor rotates in one direction and the outer in the other direction. and if the cabin is located on top of the rotating mechanism, what keeps it from rotating in one direction or the other?

OTC: We have this cabin as the center of the craft and the battery below the cabin and the electromagnets are the total outside of the circular foil. The shaft of the accumulator goes through the cabin and there is a bearing. Now, just as this stays stationary when this is rotated, so will the cabin because there are two rotating forces. You have the clockwise rotation of the accumulator, the capacitor plates, the generative coils, you have a counter-rotation of the entire circular area of the craft, the larger diameter which houses the electromagnets; therefore, when you have rotation in both directions, the cabin itself is like a bearing and extension of the shaft. We've built models and proved this is correct.

Roy: OTC XI was accomplished?

OTC: Six crafts were airborne, one escaped; we used circuit breakers of various types and fuses burned through the switch and we lost one craft.

Roy: What was the size?

OTC: The largest and the one lost was 6 feet in diameter.

Roy: You also talk of the Caroto Gravity Motor and you mention it requires no fixed location in which to function, and you also say a lot of other wonderful things about the possibility for this motor. Is it something else than the spacecraft?

OTC: They are two separate packages. The spacecraft utilizes the electrical accumulator and the gravity motor uses the energy of the inertial attractive forces. We have learned how to take this energy and key it to a working shaft and get work power which we call free energy because it is. We don't make any part of it. Now in the rather vulgar vernacular sense this would have been classified as perpetual motion. It is nothing of the sort. It's free energy. Now we have learned that all masses that are smaller than the masses to which they are attracted exert energy. Even this ashtray; if it takes a pound of energy to lift it, then it is exerting a pound of energy. We have a true gravity motor. Its functional operation is to produce power continuously without any dissipation of the energy which causes it to operate, and we have built models of this and they operate and function and we are in the patent procedure with this.

Mel: Back to a question a while ago regarding an analogy of the earth as a spacecraft. Taking another look at the mock up prototype power package there in front of you. It resembles the solar system itself, and as a matter of fact one of the statements in the brochure published in 1957, stated it illustrated the geometry of the universe. It seems that this device miniaturizes and essentially duplicates the motion of bodies in the solar system. It must have the ability to miniaturize their energy. So in a sense, the bodies in the solar system, in all time, have maintained their constant motion perpetually in those motions and this is why people characterize this power package as a perpetual motion machine.

OTC: But we do not make any such claim.

Colton: The Utron has many applications, has many forms, many variation. In one sense you might describe it as an energized armature or in other words a motor with a self-contained moving battery, also capable of continuously re-energizing itself. The offer in regards to spacecraft applies to government and industry alike. We will only make total disclosure demonstration after we have procured a firm order. In other words, we're not looking for anyone to evaluate our development. Mr Carr has come a long way in his research. He doesn't need any risk money from taxpayers or industries for further exploration and development. We will disclose to anyone who is a purchaser, but not to anyone who comes along for curiosity.

Mel: What does the word "Utron" mean?

Colton: "Utron" is a coined word, a word Mr Carr put together: the letter "U" and "tron", U meaning the direction or shape of motion as applied and used and equated in this accumulator or battery we described -- U is the plane, the geometrical figure that is the portrait of the wave, you might say. The letter U as described on paper, the two-dimensional, is a portrait on paper of the wave or the wave motion with the cut field, with the straight line, the pressure energy in the Utron accelerator.

OTC: To me there's no such thing as a completed curve; you only go half way, just like you only go half way into the woods, then you're coming out. This is the same. A bisection of a total sphere is its exact curve and one half of it is primarily U-shaped. Because in magnets there are always two poles and one normal way to show them is in a U-shape, but if it's in a bar magnet, there are still two poles and the shape is still the same. We can only put a rope one half way around the tree and it's coming back the other way, and this is true in all wave motion. Now, if you extend this into velocity, this is the pattern in the sine wave and definitely the electromagnetic wave.

Ben: I see... the vortexian of the wave

OTC: All motion is relative to all other motion, so this serpentine, spiral state.

Ben: So it's not the form of that motion, the graph and the equation. I'll accept the fact that it's the form of the utron motion though.

OTC: Also the form as well in two dimensions is intensified in the geometry of our accumulator.

Q: Why do you refer to this vehicle as a 4th dimensional vehicle?

OTC: Because the geometry of the accumulator is such as the 4th dimension. To me the application of space and time, a vibratory field and electricity as we know it is a vibratory force in motion. This is symbolic of it and when it is activated it becomes such. Now this may be a little bit difficult to understand, but nevertheless the very smallest electronic particles of matter have been shown under the highest type of electron microscope to be one-dimensional squares. To me this is verification that this is truly a space dimension because it is the shape of matter. Without matter you couldn't have space.

Q: I'm holding a thing that looks like two loudspeakers in a simplified form, placed mouth to mouth, two cones mouth to mouth, like two tops together so they have a point at each end and actually like a thickened flying saucer... Now what is it that comes from that: is it a high voltage?

OTC: The voltage is whatever we wish to make it by design.

Q: In other words, then, this is a battery?

OTC: That's right.

Q: The battery then goes through these magnets?

OTC: The battery rotates in this magnetic field. The average armature today in any electrical system is usually the permeability, iron wound with copper, then through a magnetic field acts as a motor, or it becomes a generator, depending upon the lead. The great novelty here in the area in which an armature is normally used, we have a [power unit, and this a battery, and this is a moving power unit.

Q: In other words, you might call this a self-contained power supply, right? How would this generator --- maybe I'm using the wrong word --- gather additional energy from outside?

OTC: This is due to its circular motion. Electrical forces are motions where they manifest. Now we have cycles in alternating current; AC gives you 60 cycles per second; we have discovered in our experiments that there is a space cycle related to electricity, and if we join the cycle we get energy from it.

Q: Mr Colton, will you try to describe to our listeners how they can draw at home a facsimile of an Utron?

Colton: You can take a pencil and draw four lines to form an open square. When you have a square, draw a straight line from one point of the square to an opposite point and you'll have two right angle triangles. Now if you convert the line you've just drawn into a small lip you'll begin to see two inverted cones, the base of which form a circular equator. While you started with a square, you now have two cones. Obviously, the base of the cone is a circle or completely round as we describe it, and you have the device which is described as completely round and completely square, the Utron electric accumulator. The cavity in the center of these which is a hemisphere when the two cones are put together have a hollow sphere. This is the cavity which contains the electrolyte which would be used in some of the applications of the accumulator.

Q: Of what value is the term "completely round and completely square" apart from its obvious redundancy.

Colton: It is Carr's definition of the geometry or the basic space form or the basic form of all matter large or small relatively as we describe it. It is the definition of the terminal motions of universal energies in what we call space.

Q: In one of your brochures you make mention of a "photon gun" and you say, "This is primarily a development that works outside of the earth's atmosphere. We are entering an age of space flight and the use of solar energy is practically unlimited"... What is a photon gun?

OTC: I am using the word gun as a reaction principal instead of as a weapon, nevertheless it is a gun and in fact fires billions of rays of solar energy at right angles to the reception. By placing them through a certain chamber, we have been able to get a reaction and whenever there is a reaction, we can get power, we can get force from it. So we feel outside the earth's atmosphere new systems of propulsion even beyond our own of electromagnetism will make themselves apparent.

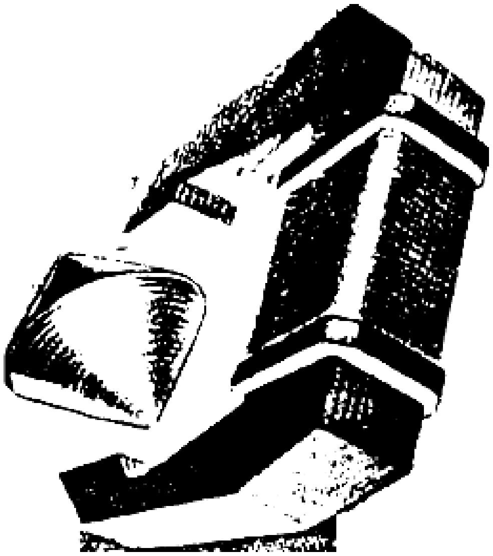

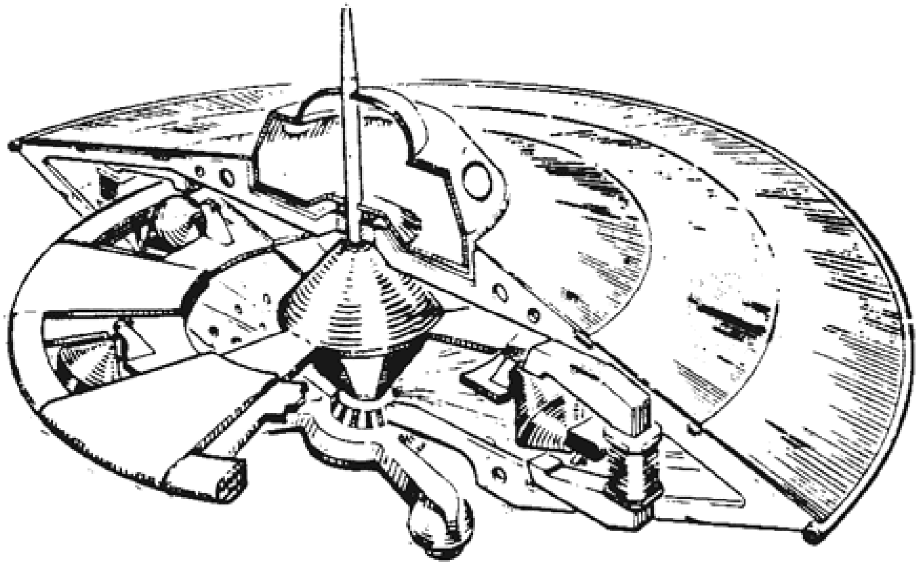

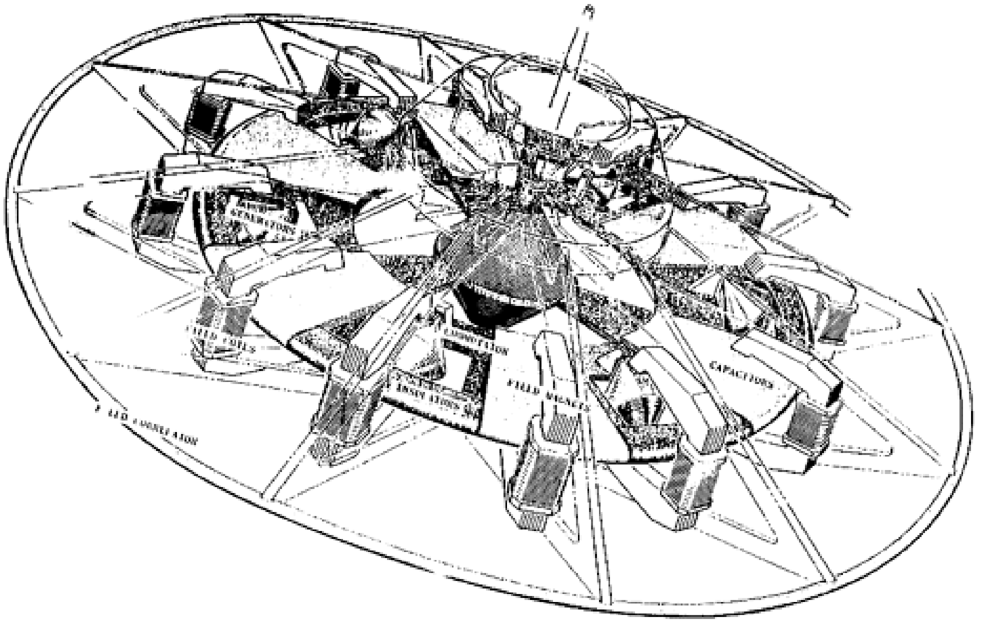
GENERATORS
FIELD COILS
CAPACITORS
FIELD MAGNETS

***Information Bulletin* #3 (December 23, 1957) [Excerpts]**

"Carr's Principles of Free Energy Space Flight Accorded Highest Scientific Verification"

Otis T. Carr, President of OTC Enterprises in Baltimore and inventor of a manmade spacecraft, which he has offered to build and deliver to the government without one penny of risk to the American taxpayer, publicly revealed the propulsion principle of his spacecraft in the following statement:

"Any vehicle accelerated to an axis rotation relative to its attractive inertial mass, immediately becomes activated by free-space-energy and acts as an independent force..."

An excellent example of this principle applied in practice is the fact that the two satellites now orbiting the earth, have achieved space-energy activation by an equivalent inertial velocity, and were it not for their susceptibility to internal or inherent physical deterioration, they would theoretically and factually be kept orbiting by natural space forces until the end of time in our solar system. These statements in evaluation of natural physical phenomena as first discerned and made public by Mr Carr, have been physically tested in laboratory experiment by an impartial scientific laboratory and confirmed as follows:

"We have shown that a charged body, accelerated to an axis rotation relative to this attractive inertial mass, indicates polarity in a given direction. The dip-needle points, say, up toward the top of the body. But mount this while rotating body, with its spindle, on another platform and rotate this platform on a spindle, then if the counter-rotation is greater than the inertial forward rotation of the body, a dip-needle on the second platform will point down while the first dip-needle points up, indicating complete relativity of polarity. When the exact counter-rotation matches the forward rotation the body loses its polarity entirely and immediately becomes activated by free-energy (tensor stresses in space) and acts as an independent force... The above-described assembly of counter-rotating charged masses becomes weightless and will escape the immediate attraction of gravitational forces.

"Your discovery therefore is distinctly a possibility, and its creative genius lies in its inventor's ability to provide it with a process for generating its own power as a result of its self-rotation."

Translated into familiar terms, this scientific explanation means that the above-described assembly of counter-rotating charged masses becomes weightless and will escape the immediately attractive gravitational forces...

Our experiments and investigations have made known to us that the mighty power from the Sun can be utilized when the geometry of design is properly evaluated. We are aware of the resistance factors involved, due to the earth's atmosphere, and our developments have been predicated on overcoming these, which in some instances has been done.

We have invented a Fourth Dimensional Vehicle. This Fourth Dimensional Vehicle is powered by the revolutionary Utron electric accumulator, which is completely round and completely square, and generates its own power, which is electromagnetism and electromotive force. It can be launched form the earth, and does not have to be blasted into space. The basic design was brought to the attention of the United States Government, and a patent application was filed in 1949.

Some tests were independently made by qualified engineers at one of America's leading plants and it was estimated that a small model, only 4 inches in diameter, produced energy equivalent to 1,000 tons. This was in 1952, and Mr Carr was unable to interest anyone in his efforts at that time. Things have changed, and the Government has now put its blessing on space flight development.

At the present time the rocket enthusiasts are commanding topmost attention. But there will be changes in thinking, and new directions of endeavor very soon. Our experiments with circular-foil design, as relates to gravity, electromagnetism, and electromotive force, have led us to the certain conclusion that it is possible to use the inverse-proportion laws of Newton as successfully as the third law of motion applies to accelerated rockets. We will produce the OTC-X1 circular-foil spacecraft as the first operational prototype as symbolized in this conception drawing.

Claims Christ Disclosed True Nature of the Universe in Symbolism of the Cross ~

Visionary Mr Carr claims that Christ disclosed the true nature of the Universe in the Symbolism of the Cross. He pointed out that "Christ is truly expressed in the sign of the Cross -- the true Cross -- the absolute, the finite expression of the cosmos that we call the Universe". The inventor further stated:

"A universe has only two basic properties: the motions and the directions. The correlation of these properties manifest in several forms, their two categories being called energy and matter.

"These two forms are equal to each other, and are the product of each other, matter being energy under the pressure of motion and having relative finity because of the harmonious correlation of the direction of such pressure, and energy being matter released from pressure by any change in the direction of motion.

"In any universal system, these motions and directions manifest their finite totality in what we call space, and describe their total form in what we call straight lines and curves, the geometry of which is expressed in a finite form whose dimensions are a constant correlation of the directions and motions. We have equated the shape of this universal space and translated it into physical form which we have called the Utron Electrical Accumulator, which expresses the linear correlation of space motions and directions in a form that is therefore completely round and completely square.

"In its linear proportions the Utron Electrical Accumulator is the union of straight lines brought together in and from all directions, and in its correlation of and with curved lines it expands from a point to the circumference of its equator and contracts from that equator to a point, always being a correlation of complete curves. The aim of these linear correlative dimensions is three points, the exact center point of which is +0 and -0, the exact peripheral points of which are +X and -X.

The space ship designer then mathematically points out his theory:

-0 / +0 = 0

-X / +X = X

0 + X = 0X

Continued Mr Carr:

"We were inspired to complete this work in the service of the Christ Jesus by Albert Einstein's search for a unified field. This relativity is a unified field, expressed in physical form is the square in the circle bisected by the Cross. The Utron Electrical Accumulator is the full expression of the dot that was the beginning of 0X, extended in all directions to form the cross, which in motion continues on to form the square and the circle.

"In the beginning was the Word. The Word was 0X [actually: AeTZI], and in the revelation of Jesus, was given for all who wanted to see, in the symbolism of the true cross.

"Set the cross in motion and you will see the halo that is Light and which we also call Energy.

"If there is not time
And there is no space
There can be no other place.
And perhaps there is no far.
Perhaps the Universe
Is right where we are."

Peace and plenty through the Application of Free Energy to Supply All Things to All People.

US Patent # 2,912,244

Amusement Device

(November 10, 1959)

Otis T. Carr

This invention relates generally to implements in amusement devices, and more particularly to an improved amusement device of the type where the passengers will receive the impression of riding in an interplanetary spacecraft.

The present invention contemplates a novel amusement device having the overall configuration of a spacecraft and being formed in various sections, with portions of the sections being rotated in opposite directions to give the impression of movement to the passengers. It is also contemplated to move the craft up and down and display an animated movie of heavenly bodies above the passengers to give the passengers an impression of leaving the earth and approaching a distant planet or the like. This invention further contemplates the provision of electromagnets, cones, and the like carried by the oppositely rotating sections of the device, and windows through which the passengers may view these simulated objects to give the impression that they are viewing the inner workings of an interplanetary spacecraft during flight of the craft. Finally, this invention contemplates the construction of the various portions of an amusement device simulating an interplanetary spacecraft in such a manner that the device may be easily transported in sections from one location to another and reassembled in a minimum of time.

An important object of this invention is to provide an amusement device wherein a passenger in the device receives an impression of flying in an interplanetary spacecraft.

Another object of this invention is to provide an amusement device having a general configuration of an interplanetary spacecraft and having oppositely rotating sections simulating the movement of various portions of the device to impress the passengers with flight in a spacecraft.

Another object of this invention is to provide an amusement device simulating an interplanetary spacecraft wherein the passengers will be physically moved a very minor distance and yet will receive the impression of flying in a spacecraft.

A further object of this invention is to provide a portable amusement device simulating an interplanetary spacecraft which may be easily disassembled and transported from one location to another.

A still further object of this invention is to provide an amusement device which is simple in construction, may be economically manufactured, and which will have a long service life.

Other objects and advantages of the invention will be evident from the following detailed description, when read in conjunction with the accompanying drawings which illustrate my invention.

In the drawings:

Figure 1 is an elevational view of an amusement device constructed in accordance with this invention.

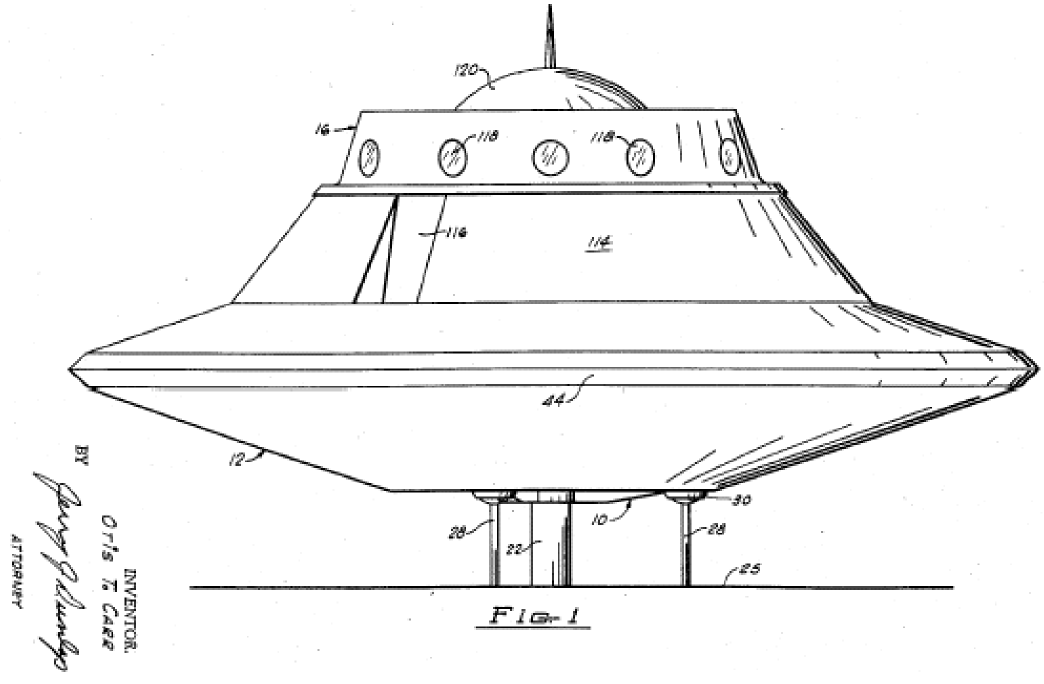
120
16
118
118
116
114
44
12
30
10
28
22
28
25
FIG. 1
INVENTOR.
OTIS T. CARR
BY
ATTORNEY

Figure 2 is a vertical sectional view through the device illustrated in Figure 1.

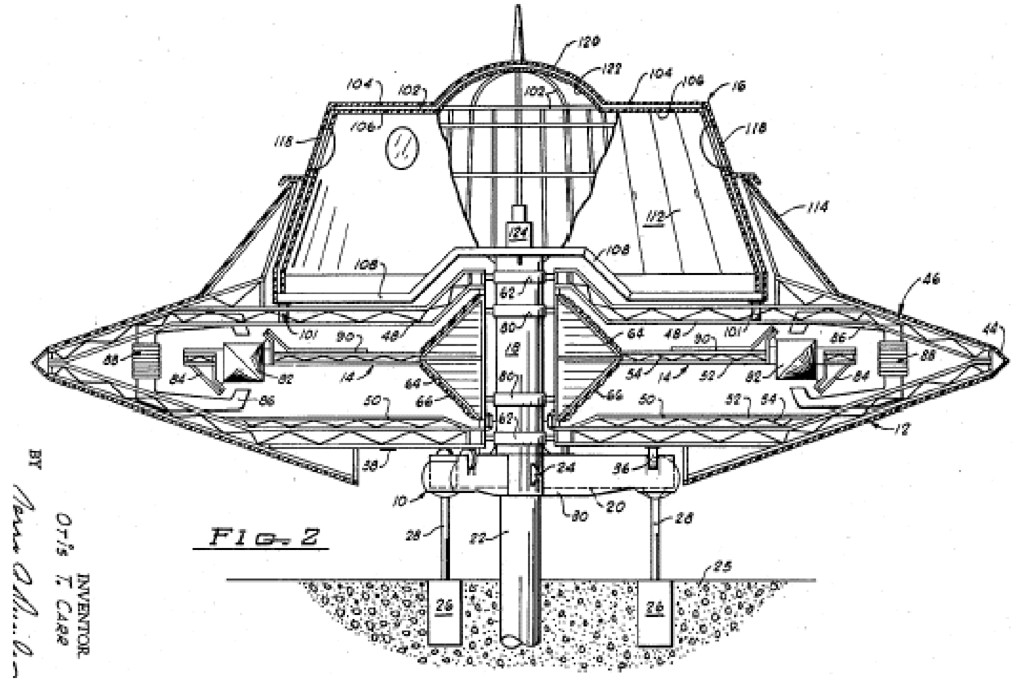
120
122
104
102
102
104
106
16
118
106
118
114
112
124
108
108
46
44
62
101
90
48
60
64
48
90
101
86
88
84
82
14
18
54
14
52
82
84
88
86
50
64
66
60
66
50
52
54
62
12
38
24
36
10
30
20
28
22
28
25
26
26
FIG. 2
INVENTOR.
OTIS T. CARR
BY

Figure 3 is a plan view of the landing gear for the amusement device, with the housing around the legs of the landing gear removed.

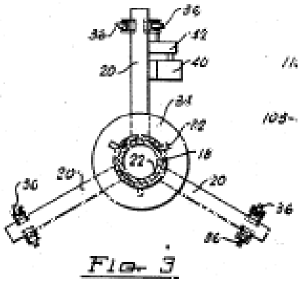
36
36
42
40
20
34
32
22
18
36
20
20
36
36
FIG. 3

Figure 4 is a plan view of the supporting structure or framework for the outer rotating shell of the amusement device shown in Figures 1 and 2.

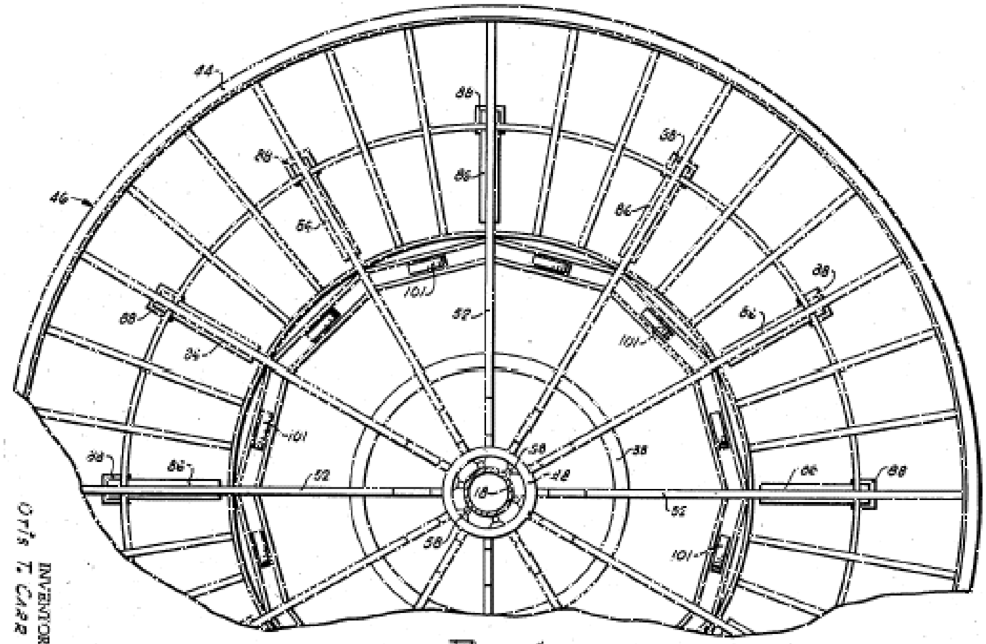
44
46
88
86
88
58
86
56
101
52
101
88
86
88
84
88
86
52
101
58
48
18
38
56
86
88
58
101
INVENTOR.
Otis T. Carr

Fig. 4

Figure 5 is a plan view of the central rotating assembly which rotates within the outer shell frame illustrated in Figures 2 and 4.

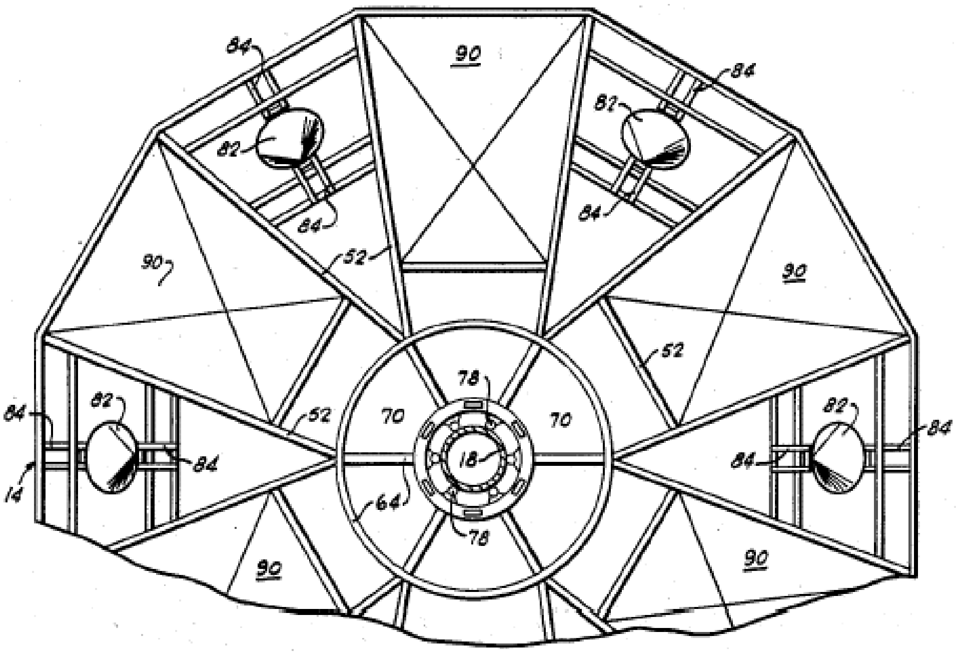
84
90
84
82
82
82
84
84
52
90
90
52
78
82
82
84
52
70
70
84
84
84
18
84
14
64
78
90
90

Fig. 5

Figure 6 is an enlarged sectional view illustrating the support and the drive between the inner rotating assembly and the outer rotating assembly.

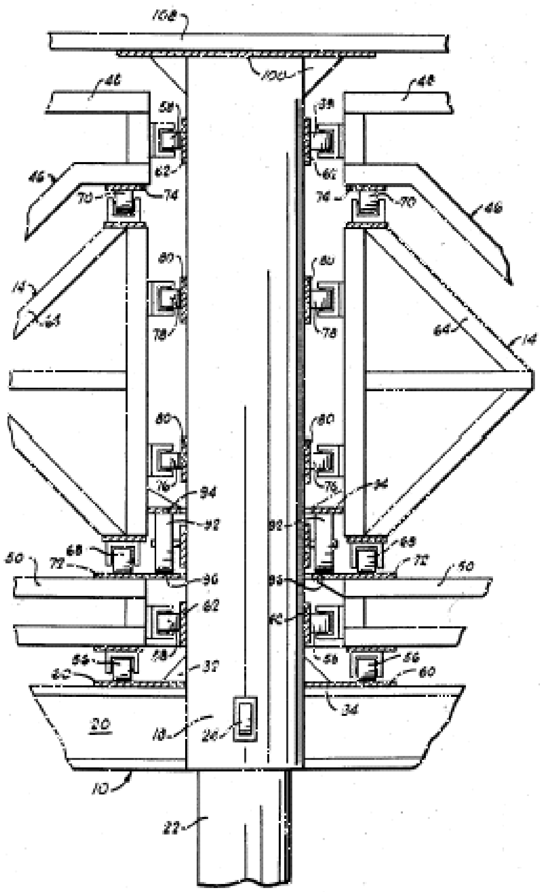

FIG. 6

INVENTOR.
OTIS T. CARR

Figure 7 is a plan view of the floor portion of the passenger cabin.

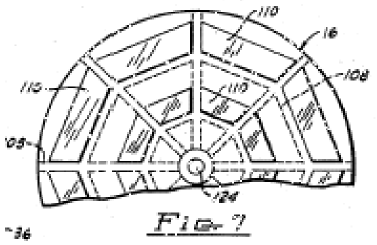

FIG. 7

Referring to the drawings in detail, and particularly Figures 1 and 2, a preferred embodiment of the present invention comprises, generally speaking, a supporting structure and landing gear 10, an outer rotating shell 12, an inner rotating assembly 14 (shown only in Figure 2) and a passenger cabin 16 extending downwardly into the upper end portion of the outer shell 12. The landing gear and supporting mechanism 20 comprises a vertically extending tubular member 18 around which the outer shell 12 and inner rotating assembly 14 are journaled as will be hereinafter set forth, and a plurality of legs 20 extending radially outward in circumferentially spaced relation from the lower end of the tubular member 18. The tubular member 18 is telescoped over a supporting stanchion 22 which may be in the form of a pipe, and the tubular member 18 is slidingly supported on the stanchion 22 by suitable wheels or rollers 24 to facilitate vertical movement of the member 18 on the stanchion 22, as will be hereinafter set forth. The stanchion 22 is preferably supported in a suitable foundation 25 to provide adequate strength for the amusement device.

A suitable jack 26, preferably a hydraulic jack, is anchored in the foundation 23 underneath the outer end portion of each of the landing gear legs 20 and is provided with an extension 28 extending upwardly into connection with the respective leg 20. The jacks 26 are operated simultaneously to raise and lower the landing gear and supporting structure 10, as will be described. A suitable shaped housing 20 is preferably secured around each of the landing gear legs 20 to provide a streamlined appearance. It may be noted that each of the legs 20 is preferably constructed out of a structurally strong member, such as an I-beam, and is secured to the lower end potions of the tubular member 18 by suitable braces 32 and plates 34 to assure a rigid and adequate connection of the legs to the tubular member 18.

A pair of rollers 36 (see also Figure 3) are rotatably supported adjacent the outer end portion of each leg 20 of the landing gear and are arranged to engage a circular track 38 (Figures 2 and 4) for partially supporting the outer shell 12 and the inner rotating assembly 14 as will be described. Also, a suitable drive motor 40 (Figure 3) is mounted on one of the legs 20 of the landing gear and is connected through a suitable gear box 42 to one of the rollers 36 to rotate the outer shell 12 with respect to the landing gear and the supporting structure 10. The roller 36 connected to the drive motor 40 may be suitably chained (not shown) to its companion roller 36 to assure an adequate drive connection between the drive motor and the circular track 38.

The shell (Figure 2) is tubular in configuration, with the diameter of the central portion thereof substantially larger than the diameter at the opposite ends thereof, and is supported in an upright position to give the general appearance of two saucers placed face-to-face. The outer edge portion 44 of the shell 12 is preferably formed of partially transparent material, such as Plexiglass, and is illuminated from within (not shown) to provide a band of light around the central portion of the shell 12. the remainder of the shell 12 is formed of any desired opaque and relatively strong material, such as aluminum-laminated masonite sheets.

The shell 12 is supported on a frame generally designated by reference character 46 and illustrated in Figures 2 and 4. the frame 46 is circular in plan, as viewed in Figure 4, and is substantially U-shaped in cross-section (as shown in Figure 2) to provide an upper horizontally extending portion 48 and a lower horizontally extending portion 50 which extend from adjacent the support member 18 outwardly to the central portion of the shell 12. The frame 46 may be easily formed with industrial steel tubing 52 formed into a truss with bracing of steel rods 54. With this construction, the frame 46 may be easily fabricated in sections for convenient assembly, or the entire frame 46 may be lifted off of the supporting member 18 and transported to another location.

Support of the frame 46 and the outer shell 12 is provided by the track 38 (previously mentioned) which is secured on the lower face of the lower horizontally extending portion 50 of th3 frame 46 in a position to engage the roller 36 carried by the landing gear legs 20. Additional support for the frame 46 is provided by rollers 56 (Figure 6) secured around the inner edge of the lower portion 50 of the frame, and rollers 58 secured around the inner periphery of both portions 48 and 50 of the frame. The rollers 56 are arranged in circumferentially spaced relation around the inner periphery of the frame 46 and engage a circular track 60 supported by the legs 20 of the landing gear. The track 60 may be an extension of the upper plate 34 shown in Figure 3. The rollers 58 engage tracks 62 extending circumferentially around the supporting member 18 to accommodate any radial forces imposed on the frame 46. It will be understood that any desired number of the rollers or wheels 36 operated by the motor 40 will engage the track 38 to rotate the shell 12 in one direction around the longitudinal axis of the supporting member 18. The rollers 56 will engage the track 60, and the rollers 58 will engage the tracks 62 during such rotation of the outer shell 12 and frame 46.

The inner rotating assembly 14 is generally annular-shaped, as illustrated in Figure 5, for movement around the supporting member 18 between the upper and lower horizontally extending portions 48 and 50 of the frame 46. The assembly 14 is also preferably constructed as a frame out of square steel tubing 52 formed into a truss with steel rods 54 for bracing, and is provided with a double conically-shaped framework 64 adjacent the supporting member 18. The framework 64 is provided to support a covering material 66 of any suitable type such as laminated metal or plastic, to simulate a solid member in the central portion of the amusement device, which in turn simulates a portion of a spacecraft, such as the power unit for the craft. The covering material 66 may be corrugated if desired. The framework 64, as illustrated in Figure 6, is also provided to support the rotating assembly 14 on the supporting member 18 and the frame 46. A plurality of rollers 68 are provided around the lower face of the framework 64, and a plurality of rollers 70 are provided around the upper face of the framework 64 to engage circular tracks 72 and 74, respectively, which are mounted on the frame 46. The rollers in each set of rollers 68 and 70 are provided in circumferentially spaced relation around the inner periphery of the rotating assembly 14 to accommodate vertical thrusts imposed on the rotating assembly. Additional rollers 76 and 78 are provided around the inner periphery of the framework 64 to engage tracks 80 extending circumferentially around the supporting member 18 to accommodate radial thrusts imposed on the central rotating assembly 14. It will be apparent that any desired number of rollers 76 and 78 may be provided around the inner periphery of the framework 64 to adequately support the rotating assembly 14.

I also prefer to support a plurality of conically-shaped member 82 in the outer edge portion of the central rotating assembly 14. Each conically-shaped member 82 may be easily formed out of a suitable sheet material, such as aluminum or plastic, and is supported at its opposite ends by suitable trunions 84 to extend through the truss comprising the frame of the central rotating assembly 14. The conically-shaped members 82 are preferably provided in circumferentially spaced relation around the assembly 14, as illustrated in Figure 5, and are positioned to move through simulated electromagnets 86 supported around the frame 46, as illustrated in Figures 2 and 4. Each simulated electromagnet 86 may be formed in horseshoe shape out of any suitable sheet material, and is provided with tubing or the like 88 around a portion thereof to simulate the winding of an electromagnet. I further prefer to secure a plurality of plates 90 in spaced relation 14 to simulate capacitor plates in a spacecraft. The plates 90 may be formed out of any suitable material such as aluminum-laminated masonite to provide a striking appearance.

The inner rotating assembly 14 is driven by a friction or dead wheel 92 pivotally supported on the tubular member 18, as illustrated in Figure 6, engaging circular plate 94 and 96 carried by the inner rotating assembly 14 and the frame 46, respectively. The plates 94 and 96 are arranged in vertically spaced relation with sufficient distance therebetween to frictionally engage opposite portions of the wheel 92. It will therefore be apparent that when the frame 46 is turned in one direction, the plate 96 will engage the periphery of the wheel 92 to drive the wheel 92 and in turn drive the central rotating assembly in the opposite direction by frictional engagement of the wheel 92 with the plate 94. Therefore, the various conically-shaped members 82 will be moved successively through the simulated electromagnets 86 and give the appearance of the generation of electrical energy.

The cabin 16 is supported on the upper end of the tubular supporting member 18 and is rigidly secured to the member 18 by bracing 100 to move with the supporting member 18 and to be prevented from rotating with the shell 12 or rotating assembly 14. Additional support is provided by a plurality of rollers 101 carried by the frame 46 engaging a complementary track on the bottom of the cabin. The cabin 6 is constructed out of a suitable framework 102 such as aluminum laminated with masonite, on the outer surface of the framework and any other suitable material 106, such as masonite, around the inner surface of the framework. It will be understood that substantially strong members 108 must be provided in the floor of the cabin 16 to adequately support passengers in the cabin. As illustrated in Figure 7, a plurality of transparent windows 110 are provided adjacent the outer and inner peripheries of the floor of the cabin 16, through which the passengers in the cabin may adequately view the inner rotating assembly 14 and the frame 46.

Passengers may either stand or sit in the cabin 16 and may gain entrance and exit from the cabin through suitable doors 112 as illustrated in Figure 2. It will also be noted that the upper edge portion 114 of the outer shell 12 is extended upwardly to overlap the lower end portion of the cabin 16. A portion of the upper edge 114 of the shell 12 is removed, as at 116 in Figure 1, such that passengers may enter the cabin 16 through the doors 112. Simulated portholes 118 are provided around the upper edge portion of the cabin 16 to further simulate an interplanetary spacecraft. It will be noted, however, that the portholes 118 are preferably covered with an opaque material on the inside of the cabin 16 to prevent the passengers from seeing familiar objects outside the cabin 16.

A dome 120 is preferably provided in the top central portion of the cabin 16 and the under surface thereof is covered with a suitable material 122, such as cloth or the like, which will display a movie scene. A projector 124 is suitably mounted in the central portion of the cabin 16 to project an animated movie onto the screen 122 during the operation of the amusement device to give passengers the impression of interplanetary flight.

In summarizing the operation of the referred amusement device, the passengers are directed into the cabin 16 through the opening 116 in the upper edge portion of the shell 12 and through the doors 112. The doors

112 are then securely fastened in close depositions and the passengers are thereafter prevented from viewing any object outside the amusement device during the simulated flight. The drive motor 40 is then placed in operation to rotate the shell 12 and frame 46 in one direction, while the friction wheel 92 drives the inner rotating assembly 14 in an opposite direction. Passenger sin the cabin 16 will then view the movement of the inner assembly 124 and the frame 46 through the windows 110 to receive the impression of watching the inner workings of a spacecraft. Also, suitable lighting effects (not shown) are preferably provided in the sell 12 to illuminate the moving parts. Simultaneously with rotation of the shell 12 and the inner rotating assembly 14, the jacks 26 are actuated to raise the supporting member 18 on the stanchion 22. This upward movement of the supporting member 18 will be rather minor, but will be sufficient to give the sensation of rising to the passengers in the closed cabin 16. It will be observed that when the supporting member 18 is raised, the shell 12 and the inner rotating assembly 14 are simultaneously raised to retain the vertical relationship between the cabin and the remainder of the apparatus viewed by passengers in the cabin. Also, simultaneously with rotation of the shell 12 and rotating assembly 14, the movie is displayed by the projector 124 on screen 122 on the dome of the cabin to give the passengers the impression of approaching one or more heavenly bodies. The jacks 26 are then operated in the opposite direction to lower the cabin 16 and the movie projected by the projector 124 may then provide an impression on the screen 122 that the passengers are returning to earth.

From the foregoing it will be apparent that the present invention provides an amusement device which will give the impression and sensation to patrons or passengers of the device that they are taking an interplanetary flight. The device includes counter-rotating sections simulating the inner workings of a spacecraft and these moving sections may be easily viewed by passengers in the device. Also, the passengers are slightly raised during operation of the device to simulate flight from the earth and a movie is simultaneously displayed to give the impression of approaching a heavenly body. It will also be apparent that the construction of the amusement device is such that the device may be easily disassembled and moved from one location to another. It will further be apparent that the present amusement device is simple in construction, may be economically manufactured and will have a long service life.

Changes may be made in the combination and arrangement of parts or elements as heretofore set forth in the specification and shown in the drawings, it being understood that changes may be made in the embodiment disclosed without departing from the spirit and scope of the invention as defined in the following claims.

I claim: --- [Claims not included here]

Model OTC-X1 Construction Plans

As your eyes take in the shapes and relationships in these illustrations of the miniature model you are about to build, your reasoning will let you understand much about the invisible forces that contribute to the noise-free, pressure-free, and vibrationless sensation of floating flight to be experienced in the passenger living rooms aboard the OTC-X1. It will be very much like reclining in an overstuffed armchair on our comfortable old planet Earth.

It mechanical behavior being like that of a giant gyroscope, and electrically like a planet with its own atmosphere and force field, the OTC-X1 craft will know nothing of sound barriers, shockwaves, skin friction, turbulence, or G-strains.

Neither will its occupants need fear any harm from cosmic forces, gamma rays, x-rays, or other radiation, because the OTC-X1 craft will either be using or absorbing all such forces, or by the pressure and polarity of its own force field, be shielded against their penetration.

While the commutation system for the actual energized flight model craft is not shown together with the other components on this on this sheet, its design for the circuitry of electro-gravitational forces permits an environmental surroundment of equal and opposite pressures in and from all direction, whose circumference is nowhere and whose center is everywhere relative to the universal field.

Thus, by imitating nature, we become one with nature --- by translating her shapes and duplicating her motions we join forces with natural energies and become magnetically independent of manmade limitations, man-constructed barriers, man-invented feats, and man-imposed controls! [&c, &c., yada yada yada]

Construction Plan Thumbnails: Click on the Images to open the Full-Size Drawings ~

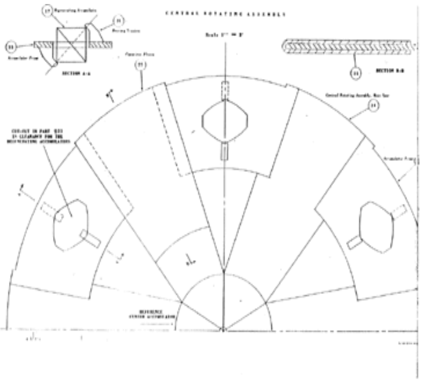

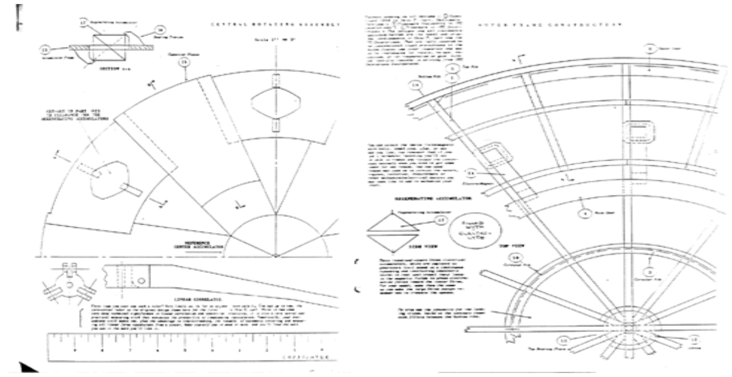

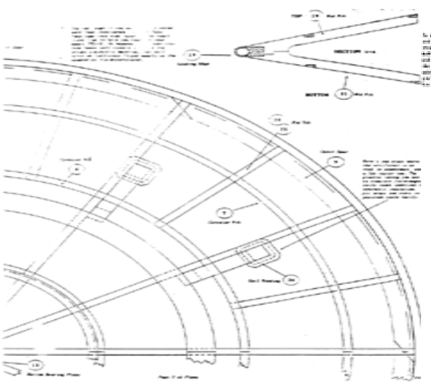

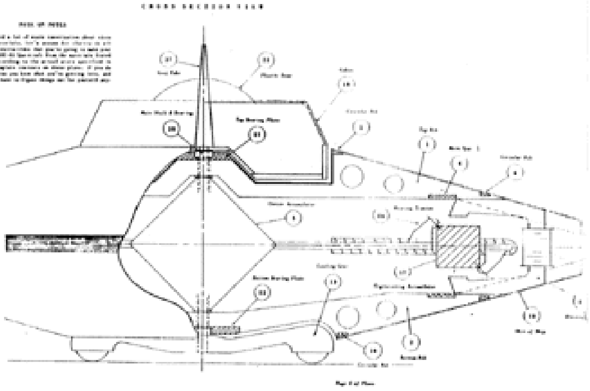

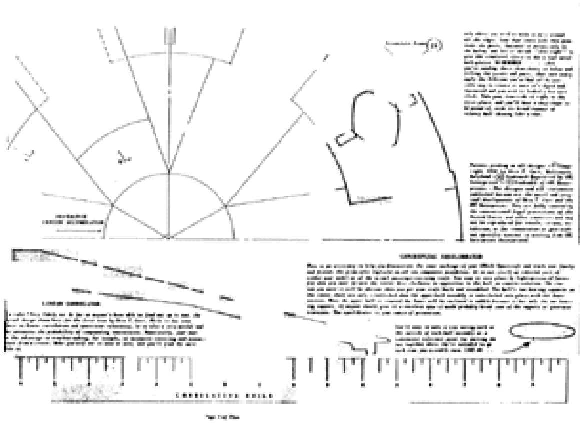

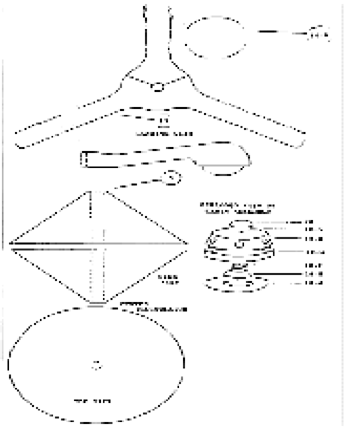

Additional Large Specifications Images (Links open in new windows): [1] ~ [2] ~ [3] ~ [4] ~ [5]

---

http://www.projectcamelot.org/ralph_ring.html

Excerpts from an Interview with Ralph Ring

Ralph Ring is a technician who worked with Otis T. Carr in the late 1950s and early 1960s. Ring stresses that resonance is the key to working with nature: "Resonance. You have to work with nature, not against her." According to Ring, when the model Utrons were powered up to operational speed, "...the metal turned to Jell-o. You could push your finger right into it. It ceased to be solid. It turned into another form of matter, which was as if it was not entirely here in this reality. That's the only way I can attempt to describe it. It was uncanny, one of the weirdest sensations I've ever felt."

Ralph Ring describes Otis Carr thus: "He was an unquestioned genius. Tesla had recognized his quality immediately and had taught him everything he knew. He was inspired, and – like Tesla – seemed to know exactly what to do to get something to work. He was a private man and was also very metaphysical in his thinking. I think the fact that he was not formally trained in physics helped him. He was not constrained by any preconceived ideas. As crazy as it sounds now, he was determined to fly to the moon and really believed it could be done. I believed it. We all did."

Ring described his "flight" in the Utron craft, saying that "Fly is not the right word. It traversed distance. It seemed to take no time. I was with two other engineers when we piloted the 45' craft about ten miles. I thought it hadn't moved --- I thought it had failed. I was completely astonished when we realized that we had returned with samples of rocks and plants from our destination. It was a dramatic success. It was more like a kind of teleportation.

"What's more, time was distorted somehow. We felt we were in the craft about fifteen or twenty minutes. We were told afterwards that we'd been carefully timed as having been in the craft no longer than three or four minutes. I still have no complete idea how it worked. We just built it exactly according to Carr's instructions. Everything had to be perfect... it all had to be just so, or it he said it would not work: a kind of symbiotic state between man and machine.

"The Utron was the key to it all. Carr said it accumulated energy because of its shape, and focused it, and also responded to our conscious intentions. When we operated the machine, we didn't work any controls. We went into a kind of meditative state and all three of us focused our intentions on the effect we wanted to achieve. It sounds ridiculous, I know. But that's what we did, and that's what worked. Carr had tapped into some principle which is not understood, in which consciousness melds with engineering to create an effect. You can't write that into equations. I have no idea how he knew it would work. But it did.

"...There are two secrets to making the alien saucers work. One is their advanced engineering, and the other is their mental ability."

"I've lost count of the number of people who have refused to believe what happened. I no longer talk about it. It's no fun being laughed at and ridiculed. But I've described it exactly as it occurred. One day someone will build the disk just as we did and they'll have the same experience. All his blueprints still exist. Nowadays, it would all be done with digital and solid state circuitry --- no moving parts would be necessary.

"I've heard that the aliens use the same principle to operate their craft. Their physics seems to work in harness with their consciousness. The craft amplifies the power of their minds. Their craft won't operate without the pilots. I've heard that's why we can't operate their craft --- or, at any rate, we can't operate them the way they do. We're just not adept enough mentally and spiritually. So there are two secrets to making the saucers work. One is the advanced engineering, and the other is the mental and spiritual ability. We may have duplicated some of the first, but we may be a long way from the second yet."

***FATE* (May 1958), page 17**

"Gravity Machine?"

The following summary was sent from a Ship's Paper of October 30,1957, after being copied from a CW News broadcast while at sea. It is unusual because no other report of this announcement reached us. It certainly is sensational --- if true.

Baltimore, MD (October 29):--- A group of inventors claimed Monday they have been able to utilize gravity in circular motion machines capable of powering everything from hearing aids to space cruisers.

Otis T. Carr, president of OTC Enterprises, Inc., detailed his claims in an interview and demonstration of a crude model of a circular motion machine which he said is the principle of a "free energy circular foil" spacecraft he can build, if someone puts up the money.

He said the machine can be adapted to devices of any size to produce continuous power absolutely free of dissipation.

Its immediate application, Carr said, would be in a spacecraft --- which would be able to fly among the planets in controlled flight.

It could land or take off as desired on the earth, the moon or any planet in the earth's solar system, he said.

Carr and his associates said their claims are based on the most simple practical applications of natural laws and discoveries in science and mathematics. They have no formal education in science or engineering.

He said the same "free energy" which causes the earth to rotate on its axis and orbit around the sun will turn a machine he described as two cones joined at their circular bases.

When the rotation of such a machine reaches a certain velocity relative to the earth's orbital velocity, Carr said, it will take off.

Carr said the core of his space ship would be a huge battery which would spin at the velocity of the external craft and which would be recharged, he said, by its motion. Carr declared such a battery, built to any size, could be designed to power the largest electric generating plant, operate an automobile, heat a house or power any conceivable machine or device.

The principle on which Carr said such circular motion machines would operate is that "any vehicle accelerated to an axis rotation relative to its attractive inertial mass (the earth) immediately becomes activated by free space energy and acts as an independent force."

***FATE* (August 1959), page 32.**

"The Saucer That Didn't Fly"

by

W. E. Du Soir

The serious field of UFO's and flying saucer research received a setback at Oklahoma City in late April when a highly publicized launching attempt by O.T.C. Enterprises of Baltimore, Md., resulted in failure. Hundreds of persons had been invited to Oklahoma City by Otis T. Carr to watch him "launch a six-foot prototype model of the O.T.C. X-1, a space craft which works on 'utron' energy." Those who were there came away disappointed. The flying saucer did not fly. The six-foot model, well-hidden in a warehouse, never even got to the launching site, or to the amusement park called Frontier City, where it was to have been displayed beside a 45-foot version of the same X-1 which will be an amusement "ride" in the near future.

The launching site was to have been a gravel pit eight miles north of Oklahoma City where Carr claimed the saucer would rise "400 to 600 feet." Launching time was set for 3:00 PM Sunday, April 19. Two and a half hours past this time Major Wayne Aho finally announced that the launching had to be postponed because of "technical difficulties." Later it was stated by O.T.C. officials that one bearing housing was "off one-sixteenth of an inch." Most of the major flying saucer clubs had ignored the "launching". Except for a story in early April there was no advance news on Carr's experiment in Oklahoma City newspapers.

One Oklahoma City television reporter expressed the general feeling of the townspeople, "This thing will never leave the ground. And I feel that a great deal of the ballyhoo they're giving out is tied in with the ride at Frontier City. I have tried constantly to get in to see the saucer model, but they've kept it hidden." Equally well hidden was Carr himself. He was finally located in Mercy Hospital, Room 302, by John Nebel, the famed "Long John" of WOR, New York City. Carr did not show up at a meeting at a local church the day before the scheduled launching but a taped speech was delivered to the 70 people who attended. "Barring any flat tires," Carr said, "I feel that history will be made Sunday afternoon when the model of the O.T.C. X-1 is launched here."

Carr did not leave his hospital room until Monday, when he was taken to the home of Major Aho, the man in charge of publicity for the company. A scheduled "victory" dinner was pretty dismal Sunday night, where again a taped message from Carr was presented. Monday morning at 5 AM a rumor spread that the model launching would again be attempted. At that early hour a group of a half-dozen people, including Norman Colton, Carr's chief aide, went to the warehouse. But it was three hours later before anyone else showed up and most of those who had come to see the model had given up in disgust and left. Many believers in UFO's, although skeptical, had traveled to Oklahoma City to see what developed. All of them had left town by Monday afternoon.

One member of a Pennsylvania saucer club commented, "I don't know what's going on but I feel they never had any intention of trying to launch the model. I could not see any plans in sight for the model and, in fact, I understand, that a Mr. Maywood Jones presented only what he called "three-dimensional illustrations" of Carr's ideas." Monday afternoon there was more activity in the warehouse and finally, around 4 PM, the model was "turned on" and revolved, although it showed no signs of taking off from the table on which it sat. "We hoped to try to launch it again but unfortunately one of the seams burst during the test," an O.T.C. official said. "We still hope to have a launching shortly," another member of Carr's group added. Left unanswered were such questions as:

1) Why was there no publicity about the launching in Oklahoma City for two weeks before the model was supposed to fly?

2) Why did Carr ready a taped message for delivery at the "victory dinner" a day before the scheduled launching?

3) How could any engineer make a mistake of one-16th of an inch in a precision part? This would be similar to a navigator landing an airplane 300 miles off course. A mistake of one-10,000th of an inch can be a major engineering calamity.

Carr was apparently undaunted by the weekend's developments. He again commented, "This indeed is an historic occasion." Questioned about the failures of his model, he added, "The theory itself has been proven. The model will go up, if not this week, then next week or next month." Mr. Carr also has stated that he and Major Wayne Aho will "fly to the moon in a flying saucer on December 7, 1959."

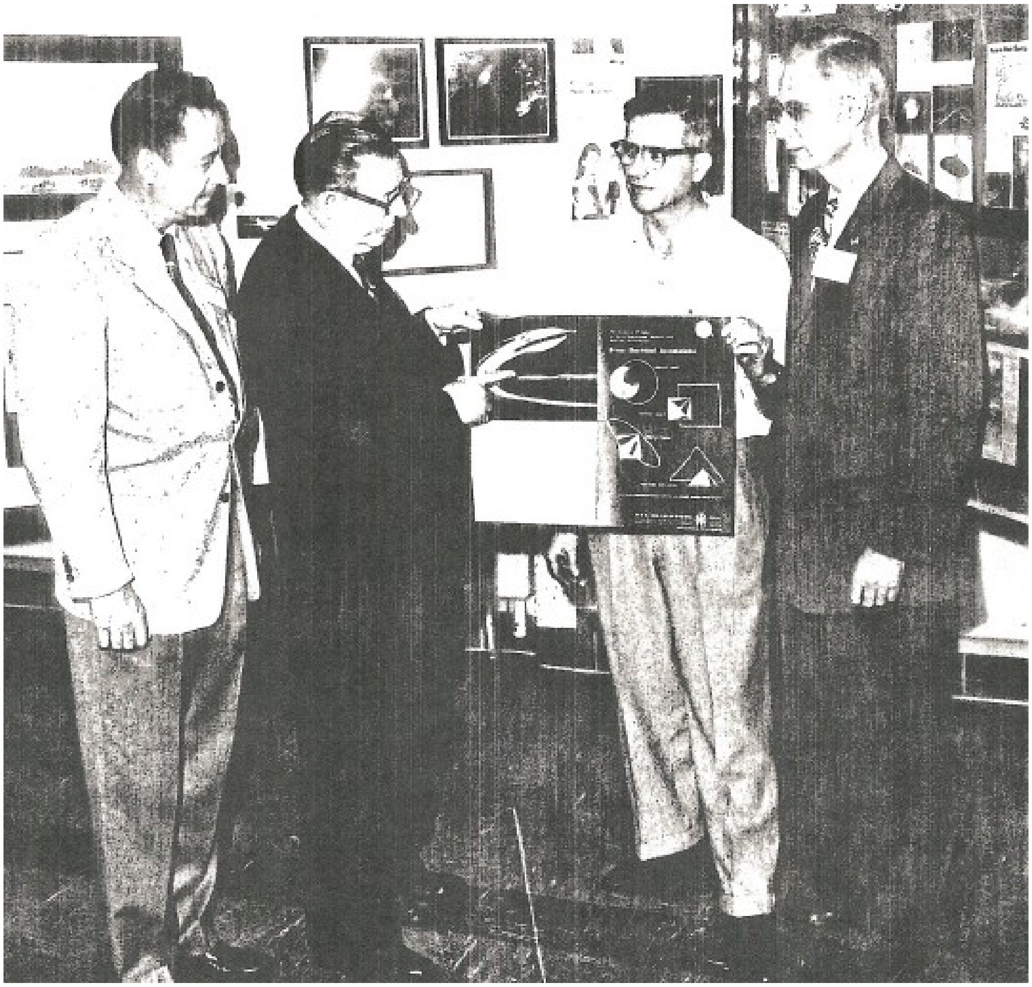

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

rexresearch.com

Otis T. CARR

UTRON

ALLIED ELECTRONICS & AUTOMATION Customizable Han-Modular® Connectors HARTING Pushing Performance LEARN MORE

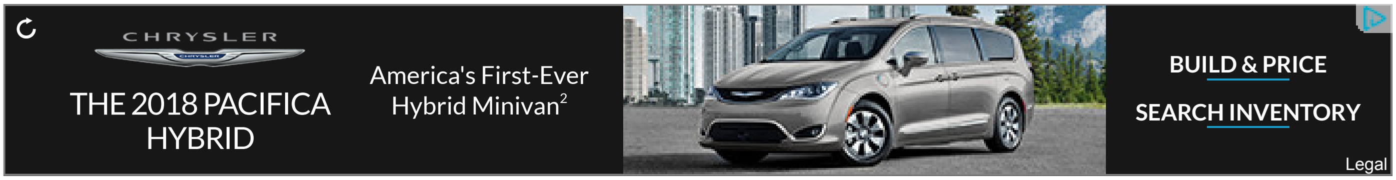
CHRYSLER THE 2018 PACIFICA HYBRID America's First-Ever Hybrid Minivan² BUILD & PRICE SEARCH INVENTORY Legal

Introduction
Interview: "Long John" Nebel & Otis Carr, *et al.* (WOR Radio, NY, 1959)
Information Bulletin #3 (December 23, 1957)
US Patent # 2,912,244: Amusement Device
OTC-X1 Model Construction Plans
Fate Magazine (May 1958): "Gravity Machine?"
Fate Magazine (August 1959): "The Saucer That Didn't Fly"

OTC-XI
© CAL BAKER

[Click Photo to View Large Image]

Introduction

In the late 1950s, Otis Carr claimed to have discovered what he called the "Utron", a "molecular shell" that houses the atoms. The cohesion of molecular shells is brought about by inertial systems in the gravity field, manifesting in the two poles of natural magnetism and electromagnetism. He claimed that by "hatching" the atomic egg, particularly of paramagnetic and diamagnetic elements (especially bismuth), atomic energy could be released gently. The Carr affair is generally considered to be a fraud. This file is a collection of excerpts of technical factoids from articles, interviews, etc., on the subject, without consideration of Carr's botched attempts to make a business of it. The technology actually makes some sense (to me, anyhow, with qualifications and reservations). I was told by Ralph Ring (one of Carr's engineers) that he himself had flown in a short test flight of a small model of the craft. Carr's business methods (fraudulent stock offerings and amateurish publicity tainted with hubris and non-performance) doomed his efforts. His engineer Norman Colton was an alcoholic who stupidly threatened a local newspaper publisher, thus generating more negative feedback. It was a sad pass from Carr's earlier statement that "We shall not stop until we have penetrated our great discoveries into all the ends of the earth, or even as far into outer space as man desires to go!"

The historic details as published in various newspaper articles and particularly in *True, The Man's Magazine* ("King of the Non-Flying Saucers", January 1961), are too long to be transcribed here for any useful purpose. This file is composed only of technical factoids gleaned from the collection of papers offered by Rex Research (albeit as poor-quality photocopies, caveat emptor), plus a couple of short articles cribbed from Keelynet/Fate magazine.

According to Carr, "Any vehicle accelerated to an axis rotation relative to its attractive inertial mass, immediately becomes activated by free-space-energy and acts as an independent force... We have shown that a charged body, accelerated to an axis rotation relative to this attractive inertial mass, indicates polarity in a given direction. The dip-needle points, say, up toward the top of the body. But mount this while rotating body, with its spindle, on another platform and rotate this platform on a spindle, then if the counter-rotation is greater than the inertial forward rotation of the body, a dip-needle on the second platform will point down while the first dip-needle points up, indicating complete relativity of polarity. When the exact counter-rotation matches the forward rotation the body loses its polarity entirely and immediately becomes activated by free-energy (tensor stresses in space) and acts as an independent force... The above-described assembly of counter-rotating charged masses becomes weightless and will escape the immediate attraction of gravitational forces..."

The Utron

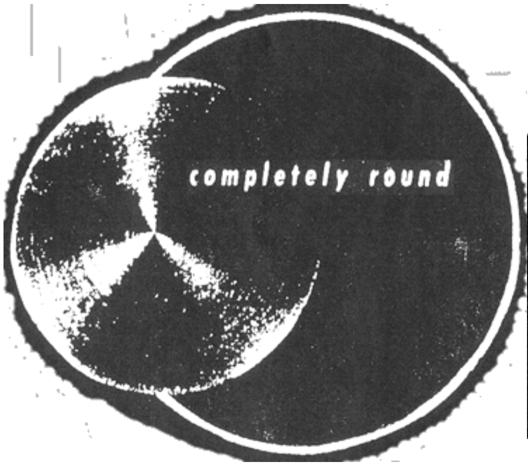

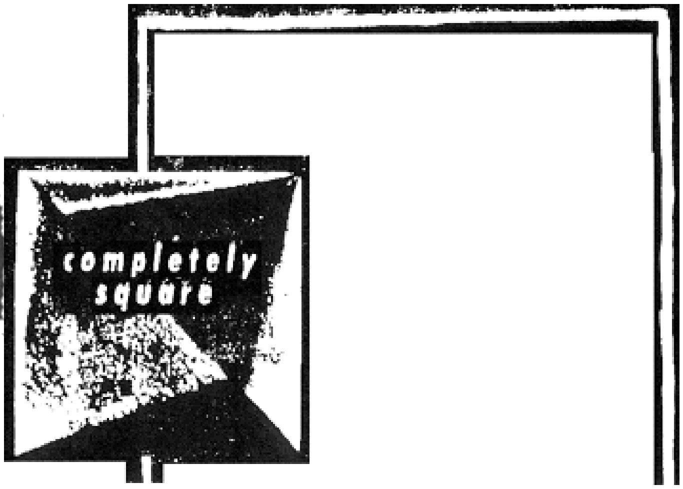

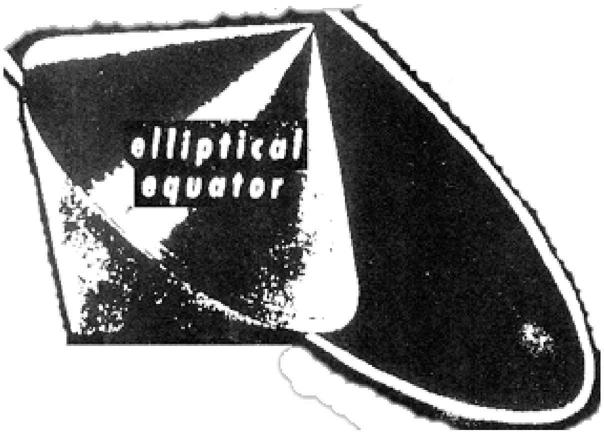

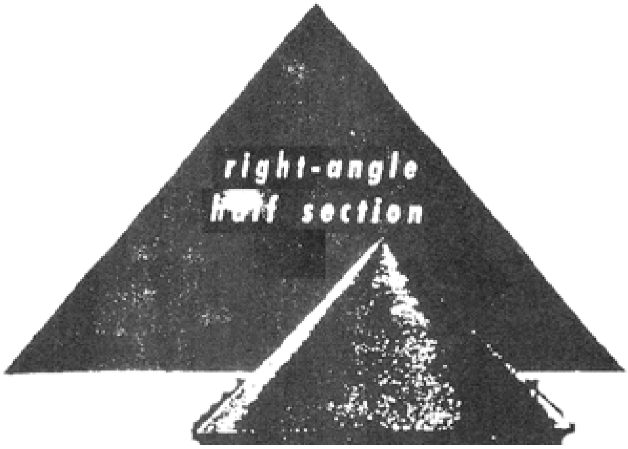

Interview: "Long John" Nebel (WOR Radio, NY) with Otis Carr, *et al.* (1959) ---

OTC (Otis T. Carr): "Our system utilizes gravity, electromagnetism, and electromotive force and a relative field to get its functional operation. We use an electrified sender. It's a sensor power core. Now this is what we call an accumulator... It is a storage cell, an accumulation of storage cells which provide an electromotive force in the same manner that any known battery produces and electromotive force...

This [the Utron] is a dimensional product. It was designed with the dimensions of space itself. We say it is truly the geometric form of space, because it is completely round and completely square. It has been proven in scientific laboratories that the very smallest unit of mass matter ever photographed in the electron microscope are square in shape... We have applied this principle into an electrified system, which is the power core of our space vehicle. Now what makes this unique and novel from a battery is the fact that this is a piece of moving machinery that rotates. Our average storage battery is an inanimate object set in an inertial spot and then the electromotive force is conducted by wires from this battery to animate some object.

LJ ["Long John" Nebel]: Let me interrupt for a moment; I'll try to describe it further. Well, it looks like two ice cream cones put together at the wide ends, but he angle is a lot wider [90 degrees] than that of the ice cream cones. There are a series of ridges; they look like gears would fit in. Is that correct?

OTC: No, those are in a sense turbine; they are reactive channels and where there is atmosphere a flow of air there aids in rotation.

LJ: Well, then, this is one of the components of the drive; this is the battery.

OTC: This is the center power core.

LJ: This generates electricity?

OTC: This is right. This is a storage cell for electrical energy. In operation it generates electricity at the same time it puts out electromotive force. This is the central power system for our spacecraft.

LJ: I can tell this: that it opened up and it appears to be hollowed out on the inside... it's circular, the inside; when the two parts are put one on top of the other and they fit into place, the cavity inside is circular...

OTC: It is a sphere, yes. And each unit is a hemisphere. We call the center of this large dimension the equator and of course it contracts and expands to a point on each side It's the union of two presto-conical sections; that is what it is Two right angle sections, and we say it is the dimensions of space and we have shown how this comes about...

LJ: In the literature you've provided us with, there's a mathematical formula that crops up here that intrigues me: minus zero divided by plus zero equals zero. Can you tell me more about it and how you stumbled on this idea?

OTC: The equation is brought about by the shape of our Utron electrical accumulator; this is the name given to our central power system. In our operation of working models and in checking out experiments, we had to find the formula that fir the reason for the action and reaction we were getting. So in exploring nature and studying the great inspirational work of the Dr. Einstein on relativity, we came upon this formula of linear correlation. And when we study linear correlation in geometric form we have to have a starting point and this is the point. And from there it explains through the cross and through the circle. And the only mathematical way we can express it is in the symbolism of zero X or 0X, and this formula brings us to that. We claim that this is the true unified field theory in physical practice.

In a further study of Dr Einstein's great inspirational work --- and we corresponded with him and we had the great good fortune of being advised by him at one time --- we learned that all measurements of time and space had to be considered in relationship to the observer and therefore there never was a fixed quantity of any number. Now in physical form, this is something else again...

On this craft, insofar as the individuals are concerned, we can travel in the same manner as a pressurized airliner. We don't have the problem of a heat shield... because the electromagnetic system sets up a protective shield in our craft which enables us to overcome this [thermal] barrier without any discomfort to the occupants of the craft. And we can very slowly rise, and once we are outside the atmosphere we can accelerate to tremendous velocities up to the speed of light itself.

LJ: How would it land, sir?

OTC: Very simply, we can fall at a very slow velocity of 100 feet/minute or less and we can set it down as gently as a feather because we have joined part of the operation of our craft to universal systems. This is a relative velocity craft and the minute it reaches the relative velocity of the attractive inertial mass, it becomes weightless as regards this inertial attraction. Individually, it is not weightless; t has the same weight as before, but when it reaches the relative location it becomes an independent system just as a planet is an independent system.

LJ: Is there any gravitational pull at this point, sir?

OTC: None whatsoever. They will have the same feeling of pressure or weight that they have right now because we will maintain as far as possible the atmospheric pressure of the earth at sea level inside the craft.

LJ: I am under the impression that the only reason I'm able to sit in this chair is because of gravitational pull.

OTC: We have around 14 lb/square inch pressure within our atmosphere. We have been able to be sealed off away from such a condition and then artificially with atmospheric pressure the pressure in the cabin is maintained. We have it very well in submarines. The same may be used in our craft.

LJ: In other words, under sea where a submarine may be there is no gravitational pull; is that what you're saying?

OTC: There is gravitational pull at all times, but we are speaking about the atmosphere of the occupants inside a sealed unit.

LJ: Is that necessary to keep the occupants in the position they desire?

OTC: Absolutely, because in a vacuum they are at the mercy of any velocity.

LJ: What would happen, sir, if there were some kind of instrument that you could turn on and eliminate the gravitational pull that was in this room?

OTC: You would in a sense become very buoyant and this is not in itself a novelty but it certainly does not have any disastrous effect on humanity.

LJ: Would I remain in this position?

OTC: You could, but any movement would move you out of it.

LJ: Would objects remain in position?

OTC: Until they were brought into any movement. Any movement would make them buoyant themselves.

LJ: I have a lead pencil; if I hold it in the air and release my fingers it would fall because of gravitational pull. If we had this other condition which you so aptly described a moment ago, if I released my fingers would the pencil remain in midair?

OTC: This is true; it would stay there.

LJ: I believe what you're saying is that you'd be creating an artificial gravitational field within the body of the spacecraft and yet there would not be any gravity on the outside?

OTC: Exactly correct.

LJ: And this is done by the battery which I attempted to describe, spinning around and producing its own gravitational influence?

OTC: Yes, this is the beginning of an answer to your question: we have capacitor plates and electromagnets as a part of this system. Now, this is counter-rotating; the electromagnets rotate in one direction, and the batteries rotate in another. The capacitor plates rotate in conjunction with the battery so that we have a clockwise and counterclockwise rotation. Now the third system is the cabin that maintains the crew. This does not rotate; it is fixed due to the fact that the two bodies are rotating clockwise and counterclockwise. Therefore the system causes the craft to escape from the gravity pull. The craft itself due to this system still has internal gravity because it still has the same weight that it had in the beginning.

LJ: What charges this battery?

OTC: This starts out electrochemically the same as other batteries, but we do have a regenerating system that is very unique. We are able here, for the first time to our knowledge, to use atmospheric electricity as a recharging system. This is done as part of the operational principle of the craft.

LJ: You say you use atmospheric electricity. What happens when you leave the atmosphere?

OTC: We have electrochemical systems to provide us with all the energy that we need and have a regenerating system in the manner of a regenerative coil that recharges this battery in the same manner that the storage battery in the automobile is recharged now, by a generator.

LJ: What you have doe is made the first perpetual motion machine.

OTC: There is nothing perpetual about our machine. The energies which cause it to operate are perpetual. You cannot destroy matter, and you cannot destroy energy; molecular flow is perpetual and that has been proven in the laboratory. It has been proven that electricity itself is immortal. When we take away resistance we can set up a spark of electricity and it will continue to operate; therefore we have perpetual energy. No machine that we can conceive of made by man would be perpetual, but it is free energy. It's self-energizing and as long as all parts function and do not wear out this is truly a self-energizing machine.

Ben: About this formula: were using conventional algebraic methods?

OTC: No, we weren't. Partially conventional, but we were joining actual space forms. We arrived at satisfactory equations for ourselves which can be demonstrated.

Ben: Are the physical laws upon which your invention works expressed in mathematical terms?

OTC: Possibly, but I wouldn't say that I'm qualified. We're satisfied with this formula.

Ben: Well, it's like saying plus 4 divided by minus 4 equals 4.

OTC: Sometimes these solutions are not always what appear. As we know in synergy, we know 1 plus 1 equals 3.

Ben: 1 plus 1 equals 3, when?

OTC: Because two conditions always produce a third.

Ben: the third condition is 2, isn't it?

OTC: Not necessarily.

Ben: Could you go over these conditions?

OTC: If one condition operates one way and another operates another way, when you join them you have another condition and their sum is 3.

Ben: Well, that is a little bit over my head. I've been looking at this prototype you have here and I notice a wooden frame or scaffolding you have a larger model of that you have on a turbine and around it is a wooden ring and it seems to be filled with electromagnets.

OTC: That's right; this is a wood model of an operational model. What we have here is the cones, our Utron electric accumulator; that is the power system. This system activates the electromagnets and in turn activates.

Ben: Does the system --- the thing inside --- activate the electromagnets on the outside?

OTC: This is true. We do this by contacting this lead wire from the positive and negative poles of these batteries to the electromagnets and then we have circuit breakers from these electromagnets and we have counter-rotation. These electromagnets will rotate counter-clockwise while the internal area is rotating clockwise.

Ben: Are these spools of wire on the model itself, are they magnetized also?

OTC: The coils of wire inside the ring are regenerative coils; they are electromotive force coils and they assist in regenerating the battery, because they are loops of wire brought through a magnetic field which sets up an electromotive force. These others are capacitor plates and these are also activated by the central power core, but these plates, which can accept a very high charge in neutral conductance also through the process of ionization utilize atmospheric electricity.

Ben: The fact that you have two cones meeting at their bases, so that you have something that is also square. I don't see how you get a square.

OTC: Dimensionally it is; it is square in these dimensions and when this rotation starts and builds up to a certain velocity, this form is very important because we have the total equation of action and reaction. Now this is done by a system of coil winding wherein we start at a point, expand to an equator, and continue our winding down to a point. With this physical expansion and contraction, is an electromagnetic field. Gravity enters the picture in the form of this relative rotation. When the relative rotation ---- inertial effective mass ---, it's a matter of dimension. So that is the earth, say, is 8,000 miles in diameter, we know its fixed rotation is 1 in 24 [blurred text, possible additional digit]. If we were 1 mile in diameter its rotation would be 8000 in 24 [blurred text, possible additional digit]. And by the same system, our 45-foot craft has a rotation of 580 rpm and when it reaches this rotation it is totally independent of its inertial attractive mass, in an electromagnetic field.

Ben: If your models get up to 580 rpm, will they take off?

OTC: This model was spun at 40,000 rpm and when it did it set a pressure pattern of 1,000 tons; the horsepower reading was a little over 700. Six engineers checked this out. Now the relative rotation of this model would be about 68,000 rpm, and when it reaches this rotation, it would immediately take off.

LJ: Would the time factor be involved with this craft?

OTC: In our solar system, the time factor would be involved, yes. We evaluate time on the velocity of light and in certain systems, if we exceed the velocity of light, unquestionably time would slow up.

Ben: Your craft can exceed the velocity of light?

OTC: We don't say this; I say I other systems.

Ben: I thought nothing could do this. I thought it was a constant factor, one of Einstein's factors.

OTC: Possibly in our system, but not necessarily true in other systems.

Ben: Anything approaching the sped of light becomes pure energy.

OTC: Pure energy, but in other systems it could change.

Ben: What other systems?

OTC: Other solar systems. We are completely controlled by our system and the velocity of light is our yardstick and our pattern and our craft is designed around this.

Ben: You don't just upset one of the basic principles of the universe.

Mel Salomey: Doesn't Einstein say any measurement is relative?

Ben: Except this one; it's the first axiom.

Mel: What is an axiom?

Ben: Self-evident truth.

Mel. Thank you. Wasn't Einstein theorizing; wasn't he assuming?

[Discussion of the speed of light, not included in transcript of the interview]

Roy: What are these coils? To describe this the best I can, imagine if you will a circle about 18 inches in diameter. Two circles that form a sort of a -- In other words, one circle fits over another circle and from the top of the circle and coming down from the other an inverted round pyramid. There are a number of what appear to be copper wound coils around the edge of it. And then there's a sort of a framework on top and underneath there's sort of braces. That's the general idea. I noticed these cones placed mouth to mouth with some coils, the coils on the edge of this thing; it revolves within this structure. A tough thing to describe.

Q: What is this over here, Mr Colton?

Colton: This is a paper mockup to show the counter-rotation principle and its outside section that Roy is describing looking at the other mockup. It contains the electrified horseshoe magnets. This would rotate in one direction counter-clockwise while the center section with the electron accumulator [Utron] would rotate clockwise in the other direction.

Q: Would these coils in the outer rim?

Colton: Rotate counter-clockwise -- It has a rotation speed relatively equal the earth's rotation speed, at about 580 rpm.

Q: When you say 580 rpm do you mean the outer rim is rotating on one direction 580 rpm and the inner rotating in the other direction at the same speed, giving a total rotation, one relative to the other, of 1160?

Colton: Exactly, although I don't know if 1160 has any bearing on it or not.

Q: Well, it would be twice the rotation in reference to the earth.

Colton: We're not giving it for a certain rotation for the sake of rotation but for the sake of relativity to the attractive mass. The earth at 8,000 miles diameter rotating once in 24 hours is relatively equal to a 45-foot craft rotating at 580 rpm, and 580 rpm would calculate to be the approximate rotational speed of an automobile tire on a car moving at about 25-30 mph.

LJ: Could a small craft be made to take off?

OTC: We plan to build a prototype model as a demonstration device. Now I would like to state that certain models have been built by me and tested. Each one has been airborne. One was lost entirely in space. We had a control system and this one didn't function. This has already been done.

Q: As was calculated, the speed of the circumference was 1263 mph. It gets kind of warm at that, doesn't it?

OTC: No, it won't because it has its own protection field which is its electromagnetic actuation. We described it as a self-contained unit. This ionization of the capacitor plates sets up a glow brilliantly with a very soft luminescent light.

LJ: What color?

OTC: It would be in the nature of blue-green or very similar to the electric arc you see in welding. This is the field we are testing. It does not have a heat barrier in forward velocity at all. This electromagnetic field is being tested out now in conventional aircraft and proved very efficient.

Q: Would you suppose you could put rockets toward the edges of the cone and have the thing spin by rocket propulsion?

OTC: We don't need it. We have a tremendous spin here. An electric motor operates the same way. You set up an electromotive force inside a magnetic field and you get rotation. So what we actually have here is an improved electrical motor which in itself is a circular device, and we say we make energy out of the air, from another dimension.

To clear your analogy up also, we would like to demonstrate the fact that this earth itself is literally a space craft demonstrating what we're talking about; it's rotating and orbiting at a certain constant speed with a magnetic field and it is in itself a spacecraft.

Q: Mr Colton, we assume the moon has a gravitational field. How does it make a gravitational field and yet it does not rotate on its axis.

Colton: It does not rotate on its axis?

Q: No, the moon does not rotate on its axis.

Ben: Sure it does, 1 rotation for 1 revolution.

Q: How long does it take?

Ben: 28 days.

Q: And the earth takes one day.

OTC: And that's how engineers and scientists have evaluated the velocity of the craft we call the earth by the orbit of it. Upon the pattern already set up, by the amount of time it takes the moon to rotate once around the earth from the center of the earth's core: 28 days. The distance being 245,000 miles. It is easy to calculate.

Q: These cones seem to revolve over intricately wound copper coils. Do you supply any motor power to this?

OTC: All energy comes from these two cones [Utron]. This in vernacular is a battery. The big novelty is that we have put a battery in motion. We have designed it within the accepted knowledge of total dimensions of space-matter and we have activated it electrochemically [electrolyte in the hollow center] and used the force through chemical activation to activate the entire craft, after which we have motion as the feature of this accumulator.

Ben: Mr Osgood's telegram brought up a very important and perhaps crucial point, mainly, James Clerk Maxwell demonstrated that light is an electromagnetic radiation, also verified by hertz who laid the foundation for modern radio. Now the speed of electromagnetic radiation such as radio waves also travels 186,000 miles per second. In other words, light and all forms of electromagnetic radiation travel at the speed of 186,000 miles per second. Now if it were possible for your craft to travel faster than the speed of light, it could, therefore, travel faster than the speed of electromagnetic radiation. So once it exceeded 186,00 miles per second, you wouldn't be gaining all this energy from this electromagnetic radiation you're generating and wouldn't you fall rapidly down?

OTC: There is a continuous falling in space which in itself can bring velocity and can bring you to another system. We mentioned conjecturally that in other systems, there could be different velocities. We're not applying them to our craft. We don't identify them wit our craft. Relatively, we could not go faster than the speed of light unless we were in a system that permitted it. In our solar system, which we have mentioned now 3 times, it is designed on known principles. We conjectured about other systems. If we go beyond the speed of light in other systems the conditions within that system would make it possible for us to have power.

Roy: Let's assume, Mr Carr, this vehicle is at rest. What is it that originally overcomes the inertia of this rotor and starts the rotor moving?

OTC: The electromagnetic force stored in the energy of the Utron electric-accumulator, which in vernacular is a battery.

Roy: Now the Utron electric accumulator is these cones that are inverted to each other with bases together. I know of zinc batteries, nickel-cadmium, lead-acid cells. Could you use those?

OTC: We could use any of the kinds you mentioned; what we have here is tremendous power size in comparison to other batteries; therefore, it's very easy to put 1000 2-volt cells inside this one unit as you see it. It has functioned very well. In our 45-foot craft we plant to have 12,000-volt batteries which will extend an electromotive force which will energize the electromagnets and the capacitor plates. The generative coils will put back into the batteries in this system the same amount of volts going out until there is a breakdown of electro-chemicals or wear-out of equipment. But it could last as long as average storage batteries in automobiles.

Roy: In order to start the motor off originally, is it necessary to cause a flow of electricity through one set of the coils?

OTC: That's true.

Roy: Is there any magnetic force in the other set of coils at this point?

OTC: They individually operate by circuit breakers and the first motion begins to start a repetition. The same we have in a motor that has the opposite of a commutator, which is an accumulation of contact points where each coil is energized as the current flows through this coil. Then this starts the motion, the repetition of this motion brings the whole motor into phase in the same sense our accumulator and magnets become speeded up and the circuits are made and broken as they rotate.

Roy: Where the inner rotor rotates in one direction and the outer in the other direction. and if the cabin is located on top of the rotating mechanism, what keeps it from rotating in one direction or the other?

OTC: We have this cabin as the center of the craft and the battery below the cabin and the electromagnets are the total outside of the circular foil. The shaft of the accumulator goes through the cabin and there is a bearing. Now, just as this stays stationary when this is rotated, so will the cabin because there are two rotating forces. You have the clockwise rotation of the accumulator, the capacitor plates, the generative coils, you have a counter-rotation of the entire circular area of the craft, the larger diameter which houses the electromagnets; therefore, when you have rotation in both directions, the cabin itself is like a bearing and extension of the shaft. We've built models and proved this is correct.

Roy: OTC XI was accomplished?

OTC: Six crafts were airborne, one escaped; we used circuit breakers of various types and fuses burned through the switch and we lost one craft.

Roy: What was the size?

OTC: The largest and the one lost was 6 feet in diameter.

Roy: You also talk of the Caroto Gravity Motor and you mention it requires no fixed location in which to function, and you also say a lot of other wonderful things about the possibility for this motor. Is it something else than the spacecraft?

OTC: They are two separate packages. The spacecraft utilizes the electrical accumulator and the gravity motor uses the energy of the inertial attractive forces. We have learned how to take this energy and key it to a working shaft and get work power which we call free energy because it is. We don't make any part of it. Now in the rather vulgar vernacular sense this would have been classified as perpetual motion. It is nothing of the sort. It's free energy. Now we have learned that all masses that are smaller than the masses to which they are attracted exert energy. Even this ashtray; if it takes a pound of energy to lift it, then it is exerting a pound of energy. We have a true gravity motor. Its functional operation is to produce power continuously without any dissipation of the energy which causes it to operate, and we have built models of this and they operate and function and we are in the patent procedure with this.

Mel: Back to a question a while ago regarding an analogy of the earth as a spacecraft. Taking another look at the mock up prototype power package there in front of you. It resembles the solar system itself, and as a matter of fact one of the statements in the brochure published in 1957, stated it illustrated the geometry of the universe. It seems that this device miniaturizes and essentially duplicates the motion of bodies in the solar system. It must have the ability to miniaturize their energy. So in a sense, the bodies in the solar system, in all time, have maintained their constant motion perpetually in those motions and this is why people characterize this power package as a perpetual motion machine.

OTC: But we do not make any such claim.

Colton: The Utron has many applications, has many forms, many variation. In one sense you might describe it as an energized armature or in other words a motor with a self-contained moving battery, also capable of continuously re-energizing itself. The offer in regards to spacecraft applies to government and industry alike. We will only make total disclosure demonstration after we have procured a firm order. In other words, we're not looking for anyone to evaluate our development. Mr Carr has come a long way in his research. He doesn't need any risk money from taxpayers or industries for further exploration and development. We will disclose to anyone who is a purchaser, but not to anyone who comes along for curiosity.

Mel: What does the word "Utron" mean?

Colton: "Utron" is a coined word, a word Mr Carr put together: the letter "U" and "tron", U meaning the direction or shape of motion as applied and used and equated in this accumulator or battery we described -- U is the plane, the geometrical figure that is the portrait of the wave, you might say. The letter U as described on paper, the two-dimensional, is a portrait on paper of the wave or the wave motion with the cut field, with the straight line, the pressure energy in the Utron accelerator.

OTC: To me there's no such thing as a completed curve; you only go half way, just like you only go half way into the woods, then you're coming out. This is the same. A bisection of a total sphere is its exact curve and one half of it is primarily U-shaped. Because in magnets there are always two poles and one normal way to show them is in a U-shape, but if it's in a bar magnet, there are still two poles and the shape is still the same. We can only put a rope one half way around the tree and it's coming back the other way, and this is true in all wave motion. Now, if you extend this into velocity, this is the pattern in the sine wave and definitely the electromagnetic wave.

Ben: I see... the vortexian of the wave

OTC: All motion is relative to all other motion, so this serpentine, spiral state.

Ben: So it's not the form of that motion, the graph and the equation. I'll accept the fact that it's the form of the utron motion though.

OTC: Also the form as well in two dimensions is intensified in the geometry of our accumulator.

Q: Why do you refer to this vehicle as a 4th dimensional vehicle?

OTC: Because the geometry of the accumulator is such as the 4th dimension. To me the application of space and time, a vibratory field and electricity as we know it is a vibratory force in motion. This is symbolic of it and when it is activated it becomes such. Now this may be a little bit difficult to understand, but nevertheless the very smallest electronic particles of matter have been shown under the highest type of electron microscope to be one-dimensional squares. To me this is verification that this is truly a space dimension because it is the shape of matter. Without matter you couldn't have space.

Q: I'm holding a thing that looks like two loudspeakers in a simplified form, placed mouth to mouth, two cones mouth to mouth, like two tops together so they have a point at each end and actually like a thickened flying saucer... Now what is it that comes from that: is it a high voltage?

OTC: The voltage is whatever we wish to make it by design.

Q: In other words, then, this is a battery?

OTC: That's right.

Q: The battery then goes through these magnets?

OTC: The battery rotates in this magnetic field. The average armature today in any electrical system is usually the permeability, iron wound with copper, then through a magnetic field acts as a motor, or it becomes a generator, depending upon the lead. The great novelty here in the area in which an armature is normally used, we have a [power unit, and this a battery, and this is a moving power unit.

Q: In other words, you might call this a self-contained power supply, right? How would this generator --- maybe I'm using the wrong word --- gather additional energy from outside?

OTC: This is due to its circular motion. Electrical forces are motions where they manifest. Now we have cycles in alternating current; AC gives you 60 cycles per second; we have discovered in our experiments that there is a space cycle related to electricity, and if we join the cycle we get energy from it.

Q: Mr Colton, will you try to describe to our listeners how they can draw at home a facsimile of an Utron?

Colton: You can take a pencil and draw four lines to form an open square. When you have a square, draw a straight line from one point of the square to an opposite point and you'll have two right angle triangles. Now if you convert the line you've just drawn into a small lip you'll begin to see two inverted cones, the base of which form a circular equator. While you started with a square, you now have two cones. Obviously, the base of the cone is a circle or completely round as we describe it, and you have the device which is described as completely round and completely square, the Utron electric accumulator. The cavity in the center of these which is a hemisphere when the two cones are put together have a hollow sphere. This is the cavity which contains the electrolyte which would be used in some of the applications of the accumulator.

Q: Of what value is the term "completely round and completely square" apart from its obvious redundancy.

Colton: It is Carr's definition of the geometry or the basic space form or the basic form of all matter large or small relatively as we describe it. It is the definition of the terminal motions of universal energies in what we call space.

Q: In one of your brochures you make mention of a "photon gun" and you say, "This is primarily a development that works outside of the earth's atmosphere. We are entering an age of space flight and the use of solar energy is practically unlimited"... What is a photon gun?

OTC: I am using the word gun as a reaction principal instead of as a weapon, nevertheless it is a gun and in fact fires billions of rays of solar energy at right angles to the reception. By placing them through a certain chamber, we have been able to get a reaction and whenever there is a reaction, we can get power, we can get force from it. So we feel outside the earth's atmosphere new systems of propulsion even beyond our own of electromagnetism will make themselves apparent.

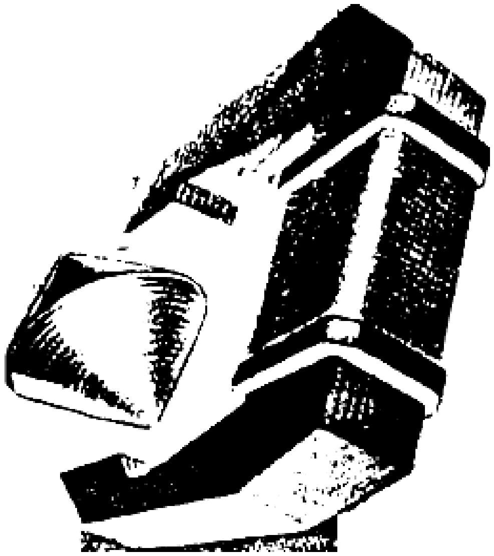

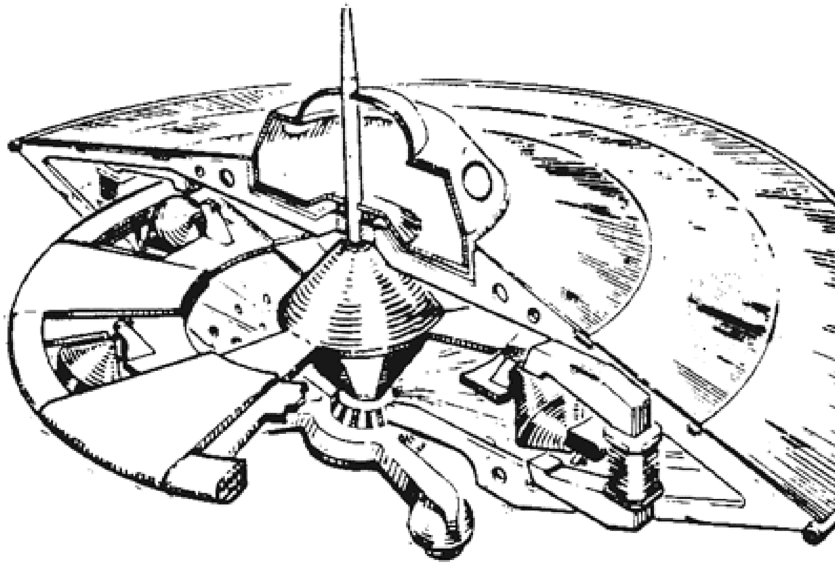

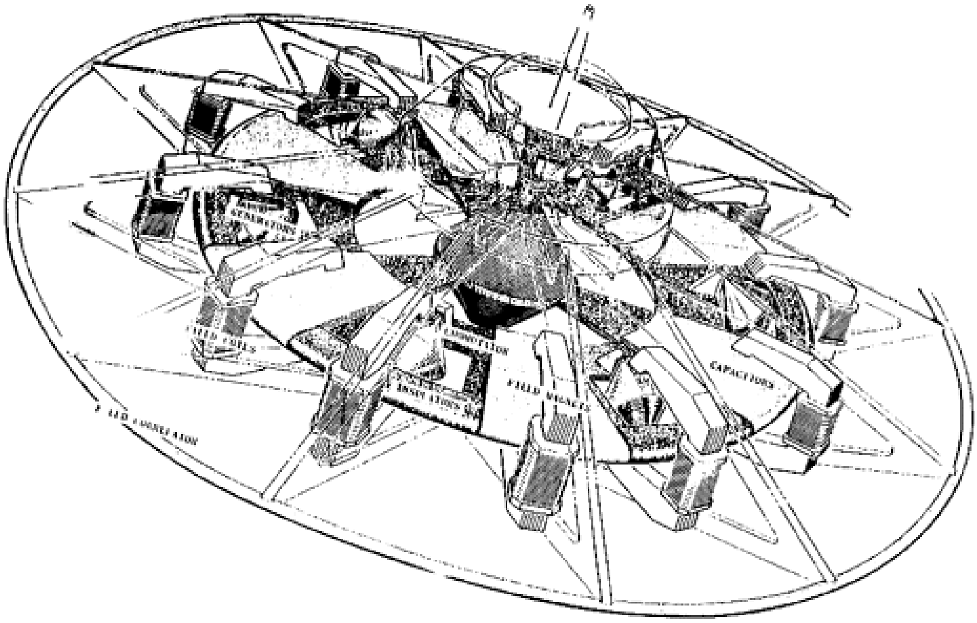
GENERATORS
FIELD COILS
FIELD MAGNETS
CAPACITORS

***Information Bulletin* #3 (December 23, 1957) [Excerpts]**

"Carr's Principles of Free Energy Space Flight Accorded Highest Scientific Verification"

Otis T. Carr, President of OTC Enterprises in Baltimore and inventor of a manmade spacecraft, which he has offered to build and deliver to the government without one penny of risk to the American taxpayer, publicly revealed the propulsion principle of his spacecraft in the following statement:

"Any vehicle accelerated to an axis rotation relative to its attractive inertial mass, immediately becomes activated by free-space-energy and acts as an independent force..."

An excellent example of this principle applied in practice is the fact that the two satellites now orbiting the earth, have achieved space-energy activation by an equivalent inertial velocity, and were it not for their susceptibility to internal or inherent physical deterioration, they would theoretically and factually be kept orbiting by natural space forces until the end of time in our solar system. These statements in evaluation of natural physical phenomena as first discerned and made public by Mr Carr, have been physically tested in laboratory experiment by an impartial scientific laboratory and confirmed as follows:

"We have shown that a charged body, accelerated to an axis rotation relative to this attractive inertial mass, indicates polarity in a given direction. The dip-needle points, say, up toward the top of the body. But mount this while rotating body, with its spindle, on another platform and rotate this platform on a spindle, then if the counter-rotation is greater than the inertial forward rotation of the body, a dip-needle on the second platform will point down while the first dip-needle points up, indicating complete relativity of polarity. When the exact counter-rotation matches the forward rotation the body loses its polarity entirely and immediately becomes activated by free-energy (tensor stresses in space) and acts as an independent force... The above-described assembly of counter-rotating charged masses becomes weightless and will escape the immediate attraction of gravitational forces.

"Your discovery therefore is distinctly a possibility, and its creative genius lies in its inventor's ability to provide it with a process for generating its own power as a result of its self-rotation."

Translated into familiar terms, this scientific explanation means that the above-described assembly of counter-rotating charged masses becomes weightless and will escape the immediately attractive gravitational forces...

Our experiments and investigations have made known to us that the mighty power from the Sun can be utilized when the geometry of design is properly evaluated. We are aware of the resistance factors involved, due to the earth's atmosphere, and our developments have been predicated on overcoming these, which in some instances has been done.

We have invented a Fourth Dimensional Vehicle. This Fourth Dimensional Vehicle is powered by the revolutionary Utron electric accumulator, which is completely round and completely square, and generates its own power, which is electromagnetism and electromotive force. It can be launched form the earth, and does not have to be blasted into space. The basic design was brought to the attention of the United States Government, and a patent application was filed in 1949.

Some tests were independently made by qualified engineers at one of America's leading plants and it was estimated that a small model, only 4 inches in diameter, produced energy equivalent to 1,000 tons. This was in 1952, and Mr Carr was unable to interest anyone in his efforts at that time. Things have changed, and the Government has now put its blessing on space flight development.

At the present time the rocket enthusiasts are commanding topmost attention. But there will be changes in thinking, and new directions of endeavor very soon. Our experiments with circular-foil design, as relates to gravity, electromagnetism, and electromotive force, have led us to the certain conclusion that it is possible to use the inverse-proportion laws of Newton as successfully as the third law of motion applies to accelerated rockets. We will produce the OTC-X1 circular-foil spacecraft as the first operational prototype as symbolized in this conception drawing.

Claims Christ Disclosed True Nature of the Universe in Symbolism of the Cross ~

Visionary Mr Carr claims that Christ disclosed the true nature of the Universe in the Symbolism of the Cross. He pointed out that "Christ is truly expressed in the sign of the Cross -- the true Cross -- the absolute, the finite expression of the cosmos that we call the Universe". The inventor further stated:

"A universe has only two basic properties: the motions and the directions. The correlation of these properties manifest in several forms, their two categories being called energy and matter.

"These two forms are equal to each other, and are the product of each other, matter being energy under the pressure of motion and having relative finity because of the harmonious correlation of the direction of such pressure, and energy being matter released from pressure by any change in the direction of motion.

"In any universal system, these motions and directions manifest their finite totality in what we call space, and describe their total form in what we call straight lines and curves, the geometry of which is expressed in a finite form whose dimensions are a constant correlation of the directions and motions. We have equated the shape of this universal space and translated it into physical form which we have called the Utron Electrical Accumulator, which expresses the linear correlation of space motions and directions in a form that is therefore completely round and completely square.

"In its linear proportions the Utron Electrical Accumulator is the union of straight lines brought together in and from all directions, and in its correlation of and with curved lines it expands from a point to the circumference of its equator and contracts from that equator to a point, always being a correlation of complete curves. The aim of these linear correlative dimensions is three points, the exact center point of which is +0 and -0, the exact peripheral points of which are +X and -X.

The space ship designer then mathematically points out his theory:

-0 / +0 = 0

-X / +X = X

0 + X = 0X

Continued Mr Carr:

"We were inspired to complete this work in the service of the Christ Jesus by Albert Einstein's search for a unified field. This relativity is a unified field, expressed in physical form is the square in the circle bisected by the Cross. The Utron Electrical Accumulator is the full expression of the dot that was the beginning of 0X, extended in all directions to form the cross, which in motion continues on to form the square and the circle.

"In the beginning was the Word. The Word was 0X [actually: AeTZI], and in the revelation of Jesus, was given for all who wanted to see, in the symbolism of the true cross.

"Set the cross in motion and you will see the halo that is Light and which we also call Energy.

"If there is not time
And there is no space
There can be no other place.
And perhaps there is no far.
Perhaps the Universe
Is right where we are."

Peace and plenty through the Application of Free Energy to Supply All Things to All People.

US Patent # 2,912,244

Amusement Device

(November 10, 1959)

Otis T. Carr

This invention relates generally to implements in amusement devices, and more particularly to an improved amusement device of the type where the passengers will receive the impression of riding in an interplanetary spacecraft.

The present invention contemplates a novel amusement device having the overall configuration of a spacecraft and being formed in various sections, with portions of the sections being rotated in opposite directions to give the impression of movement to the passengers. It is also contemplated to move the craft up and down and display an animated movie of heavenly bodies above the passengers to give the passengers an impression of leaving the earth and approaching a distant planet or the like. This invention further contemplates the provision of electromagnets, cones, and the like carried by the oppositely rotating sections of the device, and windows through which the passengers may view these simulated objects to give the impression that they are viewing the inner workings of an interplanetary spacecraft during flight of the craft. Finally, this invention contemplates the construction of the various portions of an amusement device simulating an interplanetary spacecraft in such a manner that the device may be easily transported in sections from one location to another and reassembled in a minimum of time.

An important object of this invention is to provide an amusement device wherein a passenger in the device receives an impression of flying in an interplanetary spacecraft.

Another object of this invention is to provide an amusement device having a general configuration of an interplanetary spacecraft and having oppositely rotating sections simulating the movement of various portions of the device to impress the passengers with flight in a spacecraft.

Another object of this invention is to provide an amusement device simulating an interplanetary spacecraft wherein the passengers will be physically moved a very minor distance and yet will receive the impression of flying in a spacecraft.

A further object of this invention is to provide a portable amusement device simulating an interplanetary spacecraft which may be easily disassembled and transported from one location to another.

A still further object of this invention is to provide an amusement device which is simple in construction, may be economically manufactured, and which will have a long service life.

Other objects and advantages of the invention will be evident from the following detailed description, when read in conjunction with the accompanying drawings which illustrate my invention.

In the drawings:

Figure 1 is an elevational view of an amusement device constructed in accordance with this invention.

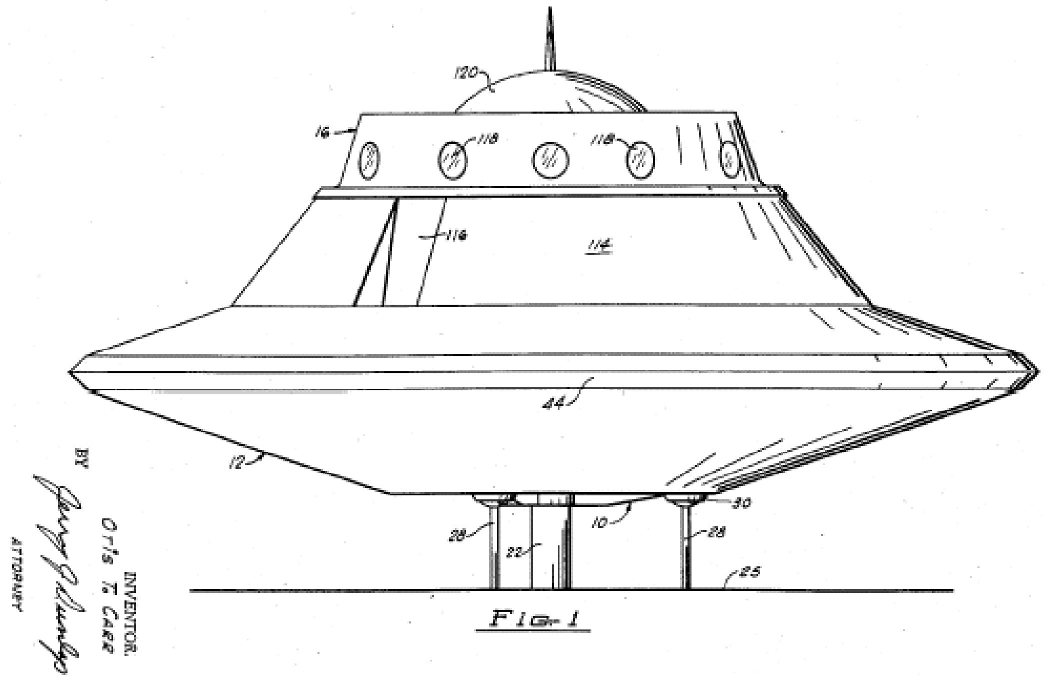
120
16
118
118
116
114
44
12
30
10
28
22
28
25
INVENTOR.
Otis T. Carr
BY
ATTORNEY

FIG-1

Figure 2 is a vertical sectional view through the device illustrated in Figure 1.

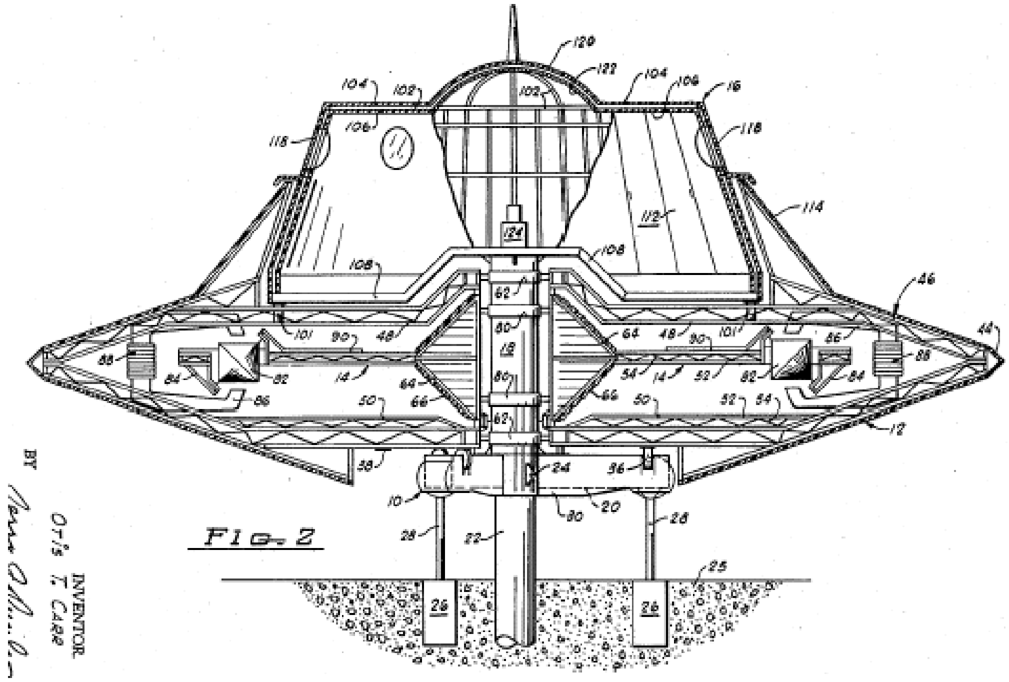
120
122
104
102
104
106
16
102
106
118
118
124
112
114
108
108
46
44
62
101
90
48
60
64
48
90
101
86
88
18
88
84
82
14
54
14
52
82
84
60
64
86
50
66
66
50
52
54
62
12
38
24
36
10
30
20
28
22
28
25
26
26
INVENTOR.
Otis T. Carr
BY

FIG-2

Figure 3 is a plan view of the landing gear for the amusement device, with the housing around the legs of the landing gear removed.

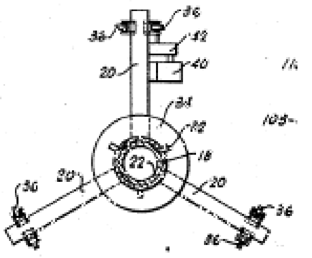
36
36
42
40
20
34
32
22
18
20
20
36
36
86

FIG-3

Figure 4 is a plan view of the supporting structure or framework for the outer rotating shell of the amusement device shown in Figures 1 and 2.

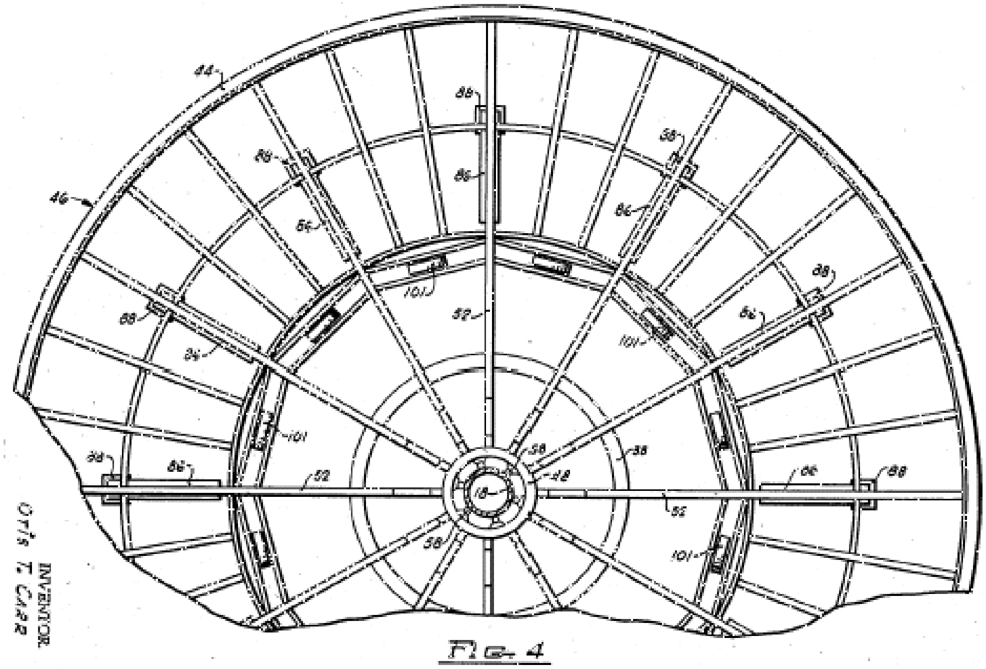
44
46
88
86
56
88
84
101
52
58
18
48
38
55
101
88
86
84
88
INVENTOR
Otis T. Carr

FIG. 4

Figure 5 is a plan view of the central rotating assembly which rotates within the outer shell frame illustrated in Figures 2 and 4.

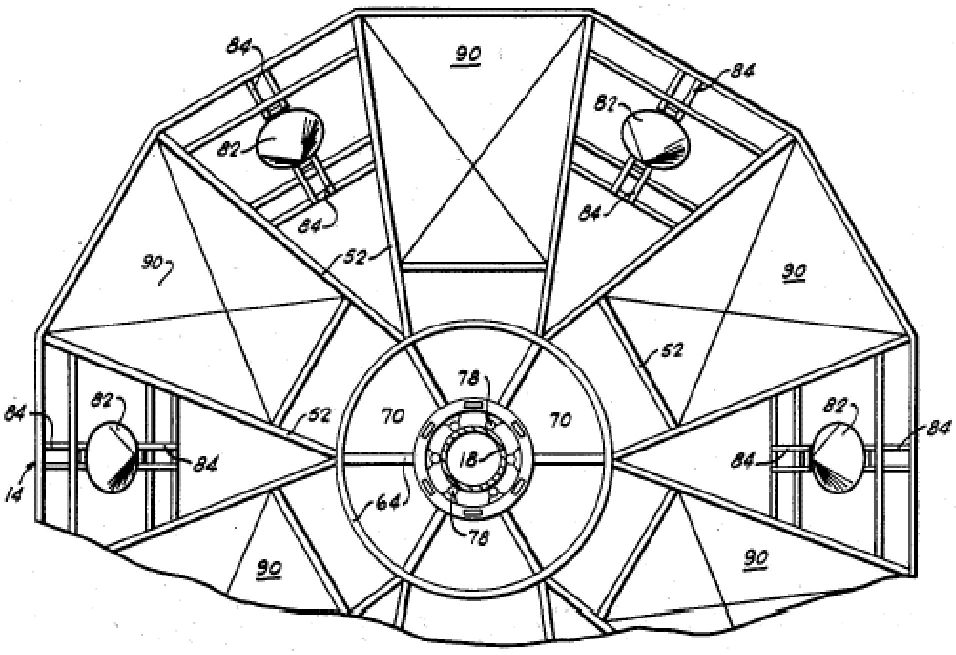
84
90
84
82
82
82
84
84
90
52
90
52
78
82
84
84
52
70
70
82
84
84
18
14
64
78
90
90

FIG. 5

Figure 6 is an enlarged sectional view illustrating the support and the drive between the inner rotating assembly and the outer rotating assembly.

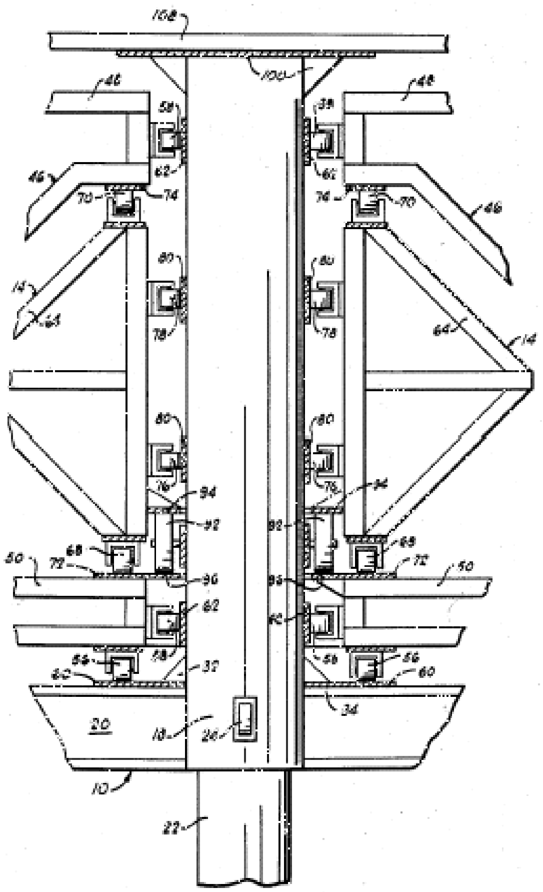

FIG. 6

INVENTOR.
OTIS T. CARR

Figure 7 is a plan view of the floor portion of the passenger cabin.

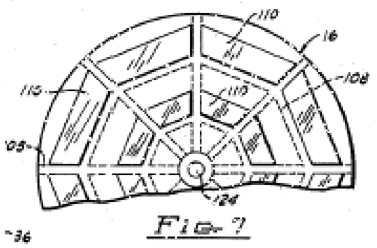

FIG. 7

Referring to the drawings in detail, and particularly Figures 1 and 2, a preferred embodiment of the present invention comprises, generally speaking, a supporting structure and landing gear 10, an outer rotating shell 12, an inner rotating assembly 14 (shown only in Figure 2) and a passenger cabin 16 extending downwardly into the upper end portion of the outer shell 12. The landing gear and supporting mechanism 20 comprises a vertically extending tubular member 18 around which the outer shell 12 and inner rotating assembly 14 are journaled as will be hereinafter set forth, and a plurality of legs 20 extending radially outward in circumferentially spaced relation from the lower end of the tubular member 18. The tubular member 18 is telescoped over a supporting stanchion 22 which may be in the form of a pipe, and the tubular member 18 is slidingly supported on the stanchion 22 by suitable wheels or rollers 24 to facilitate vertical movement of the member 18 on the stanchion 22, as will be hereinafter set forth. The stanchion 22 is preferably supported in a suitable foundation 25 to provide adequate strength for the amusement device.

A suitable jack 26, preferably a hydraulic jack, is anchored in the foundation 23 underneath the outer end portion of each of the landing gear legs 20 and is provided with an extension 28 extending upwardly into connection with the respective leg 20. The jacks 26 are operated simultaneously to raise and lower the landing gear and supporting structure 10, as will be described. A suitable shaped housing 20 is preferably secured around each of the landing gear legs 20 to provide a streamlined appearance. It may be noted that each of the legs 20 is preferably constructed out of a structurally strong member, such as an I-beam, and is secured to the lower end potions of the tubular member 18 by suitable braces 32 and plates 34 to assure a rigid and adequate connection of the legs to the tubular member 18.

A pair of rollers 36 (see also Figure 3) are rotatably supported adjacent the outer end portion of each leg 20 of the landing gear and are arranged to engage a circular track 38 (Figures 2 and 4) for partially supporting the outer shell 12 and the inner rotating assembly 14 as will be described. Also, a suitable drive motor 40 (Figure 3) is mounted on one of the legs 20 of the landing gear and is connected through a suitable gear box 42 to one of the rollers 36 to rotate the outer shell 12 with respect to the landing gear and the supporting structure 10. The roller 36 connected to the drive motor 40 may be suitably chained (not shown) to its companion roller 36 to assure an adequate drive connection between the drive motor and the circular track 38.

The shell (Figure 2) is tubular in configuration, with the diameter of the central portion thereof substantially larger than the diameter at the opposite ends thereof, and is supported in an upright position to give the general appearance of two saucers placed face-to-face. The outer edge portion 44 of the shell 12 is preferably formed of partially transparent material, such as Plexiglass, and is illuminated from within (not shown) to provide a band of light around the central portion of the shell 12. the remainder of the shell 12 is formed of any desired opaque and relatively strong material, such as aluminum-laminated masonite sheets.

The shell 12 is supported on a frame generally designated by reference character 46 and illustrated in Figures 2 and 4. the frame 46 is circular in plan, as viewed in Figure 4, and is substantially U-shaped in cross-section (as shown in Figure 2) to provide an upper horizontally extending portion 48 and a lower horizontally extending portion 50 which extend from adjacent the support member 18 outwardly to the central portion of the shell 12. The frame 46 may be easily formed with industrial steel tubing 52 formed into a truss with bracing of steel rods 54. With this construction, the frame 46 may be easily fabricated in sections for convenient assembly, or the entire frame 46 may be lifted off of the supporting member 18 and transported to another location.

Support of the frame 46 and the outer shell 12 is provided by the track 38 (previously mentioned) which is secured on the lower face of the lower horizontally extending portion 50 of th3 frame 46 in a position to engage the roller 36 carried by the landing gear legs 20. Additional support for the frame 46 is provided by rollers 56 (Figure 6) secured around the inner edge of the lower portion 50 of the frame, and rollers 58 secured around the inner periphery of both portions 48 and 50 of the frame. The rollers 56 are arranged in circumferentially spaced relation around the inner periphery of the frame 46 and engage a circular track 60 supported by the legs 20 of the landing gear. The track 60 may be an extension of the upper plate 34 shown in Figure 3. The rollers 58 engage tracks 62 extending circumferentially around the supporting member 18 to accommodate any radial forces imposed on the frame 46. It will be understood that any desired number of the rollers or wheels 36 operated by the motor 40 will engage the track 38 to rotate the shell 12 in one direction around the longitudinal axis of the supporting member 18. The rollers 56 will engage the track 60, and the rollers 58 will engage the tracks 62 during such rotation of the outer shell 12 and frame 46.

The inner rotating assembly 14 is generally annular-shaped, as illustrated in Figure 5, for movement around the supporting member 18 between the upper and lower horizontally extending portions 48 and 50 of the frame 46. The assembly 14 is also preferably constructed as a frame out of square steel tubing 52 formed into a truss with steel rods 54 for bracing, and is provided with a double conically-shaped framework 64 adjacent the supporting member 18. The framework 64 is provided to support a covering material 66 of any suitable type such as laminated metal or plastic, to simulate a solid member in the central portion of the amusement device, which in turn simulates a portion of a spacecraft, such as the power unit for the craft. The covering material 66 may be corrugated if desired. The framework 64, as illustrated in Figure 6, is also provided to support the rotating assembly 14 on the supporting member 18 and the frame 46. A plurality of rollers 68 are provided around the lower face of the framework 64, and a plurality of rollers 70 are provided around the upper face of the framework 64 to engage circular tracks 72 and 74, respectively, which are mounted on the frame 46. The rollers in each set of rollers 68 and 70 are provided in circumferentially spaced relation around the inner periphery of the rotating assembly 14 to accommodate vertical thrusts imposed on the rotating assembly. Additional rollers 76 and 78 are provided around the inner periphery of the framework 64 to engage tracks 80 extending circumferentially around the supporting member 18 to accommodate radial thrusts imposed on the central rotating assembly 14. It will be apparent that any desired number of rollers 76 and 78 may be provided around the inner periphery of the framework 64 to adequately support the rotating assembly 14.

I also prefer to support a plurality of conically-shaped member 82 in the outer edge portion of the central rotating assembly 14. Each conically-shaped member 82 may be easily formed out of a suitable sheet material, such as aluminum or plastic, and is supported at its opposite ends by suitable trunions 84 to extend through the truss comprising the frame of the central rotating assembly 14. The conically-shaped members 82 are preferably provided in circumferentially spaced relation around the assembly 14, as illustrated in Figure 5, and are positioned to move through simulated electromagnets 86 supported around the frame 46, as illustrated in Figures 2 and 4. Each simulated electromagnet 86 may be formed in horseshoe shape out of any suitable sheet material, and is provided with tubing or the like 88 around a portion thereof to simulate the winding of an electromagnet. I further prefer to secure a plurality of plates 90 in spaced relation 14 to simulate capacitor plates in a spacecraft. The plates 90 may be formed out of any suitable material such as aluminum-laminated masonite to provide a striking appearance.

The inner rotating assembly 14 is driven by a friction or dead wheel 92 pivotally supported on the tubular member 18, as illustrated in Figure 6, engaging circular plate 94 and 96 carried by the inner rotating assembly 14 and the frame 46, respectively. The plates 94 and 96 are arranged in vertically spaced relation with sufficient distance therebetween to frictionally engage opposite portions of the wheel 92. It will therefore be apparent that when the frame 46 is turned in one direction, the plate 96 will engage the periphery of the wheel 92 to drive the wheel 92 and in turn drive the central rotating assembly in the opposite direction by frictional engagement of the wheel 92 with the plate 94. Therefore, the various conically-shaped members 82 will be moved successively through the simulated electromagnets 86 and give the appearance of the generation of electrical energy.

The cabin 16 is supported on the upper end of the tubular supporting member 18 and is rigidly secured to the member 18 by bracing 100 to move with the supporting member 18 and to be prevented from rotating with the shell 12 or rotating assembly 14. Additional support is provided by a plurality of rollers 101 carried by the frame 46 engaging a complementary track on the bottom of the cabin. The cabin 6 is constructed out of a suitable framework 102 such as aluminum laminated with masonite, on the outer surface of the framework and any other suitable material 106, such as masonite, around the inner surface of the framework. It will be understood that substantially strong members 108 must be provided in the floor of the cabin 16 to adequately support passengers in the cabin. As illustrated in Figure 7, a plurality of transparent windows 110 are provided adjacent the outer and inner peripheries of the floor of the cabin 16, through which the passengers in the cabin may adequately view the inner rotating assembly 14 and the frame 46.

Passengers may either stand or sit in the cabin 16 and may gain entrance and exit from the cabin through suitable doors 112 as illustrated in Figure 2. It will also be noted that the upper edge portion 114 of the outer shell 12 is extended upwardly to overlap the lower end portion of the cabin 16. A portion of the upper edge 114 of the shell 12 is removed, as at 116 in Figure 1, such that passengers may enter the cabin 16 through the doors 112. Simulated portholes 118 are provided around the upper edge portion of the cabin 16 to further simulate an interplanetary spacecraft. It will be noted, however, that the portholes 118 are preferably covered with an opaque material on the inside of the cabin 16 to prevent the passengers from seeing familiar objects outside the cabin 16.

A dome 120 is preferably provided in the top central portion of the cabin 16 and the under surface thereof is covered with a suitable material 122, such as cloth or the like, which will display a movie scene. A projector 124 is suitably mounted in the central portion of the cabin 16 to project an animated movie onto the screen 122 during the operation of the amusement device to give passengers the impression of interplanetary flight.

In summarizing the operation of the referred amusement device, the passengers are directed into the cabin 16 through the opening 116 in the upper edge portion of the shell 12 and through the doors 112. The doors

112 are then securely fastened in close depositions and the passengers are thereafter prevented from viewing any object outside the amusement device during the simulated flight. The drive motor 40 is then placed in operation to rotate the shell 12 and frame 46 in one direction, while the friction wheel 92 drives the inner rotating assembly 14 in an opposite direction. Passenger sin the cabin 16 will then view the movement of the inner assembly 124 and the frame 46 through the windows 110 to receive the impression of watching the inner workings of a spacecraft. Also, suitable lighting effects (not shown) are preferably provided in the sell 12 to illuminate the moving parts. Simultaneously with rotation of the shell 12 and the inner rotating assembly 14, the jacks 26 are actuated to raise the supporting member 18 on the stanchion 22. This upward movement of the supporting member 18 will be rather minor, but will be sufficient to give the sensation of rising to the passengers in the closed cabin 16. It will be observed that when the supporting member 18 is raised, the shell 12 and the inner rotating assembly 14 are simultaneously raised to retain the vertical relationship between the cabin and the remainder of the apparatus viewed by passengers in the cabin. Also, simultaneously with rotation of the shell 12 and rotating assembly 14, the movie is displayed by the projector 124 on screen 122 on the dome of the cabin to give the passengers the impression of approaching one or more heavenly bodies. The jacks 26 are then operated in the opposite direction to lower the cabin 16 and the movie projected by the projector 124 may then provide an impression on the screen 122 that the passengers are returning to earth.

From the foregoing it will be apparent that the present invention provides an amusement device which will give the impression and sensation to patrons or passengers of the device that they are taking an interplanetary flight. The device includes counter-rotating sections simulating the inner workings of a spacecraft and these moving sections may be easily viewed by passengers in the device. Also, the passengers are slightly raised during operation of the device to simulate flight from the earth and a movie is simultaneously displayed to give the impression of approaching a heavenly body. It will also be apparent that the construction of the amusement device is such that the device may be easily disassembled and moved from one location to another. It will further be apparent that the present amusement device is simple in construction, may be economically manufactured and will have a long service life.

Changes may be made in the combination and arrangement of parts or elements as heretofore set forth in the specification and shown in the drawings, it being understood that changes may be made in the embodiment disclosed without departing from the spirit and scope of the invention as defined in the following claims.

I claim: --- [Claims not included here]

Model OTC-X1 Construction Plans

As your eyes take in the shapes and relationships in these illustrations of the miniature model you are about to build, your reasoning will let you understand much about the invisible forces that contribute to the noise-free, pressure-free, and vibrationless sensation of floating flight to be experienced in the passenger living rooms aboard the OTC-X1. It will be very much like reclining in an overstuffed armchair on our comfortable old planet Earth.

It mechanical behavior being like that of a giant gyroscope, and electrically like a planet with its own atmosphere and force field, the OTC-X1 craft will know nothing of sound barriers, shockwaves, skin friction, turbulence, or G-strains.

Neither will its occupants need fear any harm from cosmic forces, gamma rays, x-rays, or other radiation, because the OTC-X1 craft will either be using or absorbing all such forces, or by the pressure and polarity of its own force field, be shielded against their penetration.

While the commutation system for the actual energized flight model craft is not shown together with the other components on this on this sheet, its design for the circuitry of electro-gravitational forces permits an environmental surroundment of equal and opposite pressures in and from all direction, whose circumference is nowhere and whose center is everywhere relative to the universal field.

Thus, by imitating nature, we become one with nature --- by translating her shapes and duplicating her motions we join forces with natural energies and become magnetically independent of manmade limitations, man-constructed barriers, man-invented feats, and man-imposed controls! [&c, &c., yada yada yada]

Construction Plan Thumbnails: Click on the Images to open the Full-Size Drawings ~

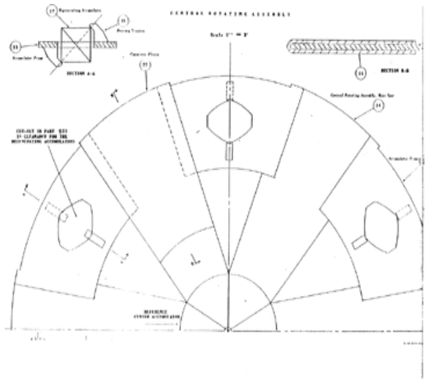
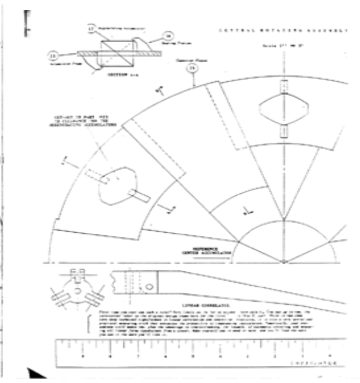
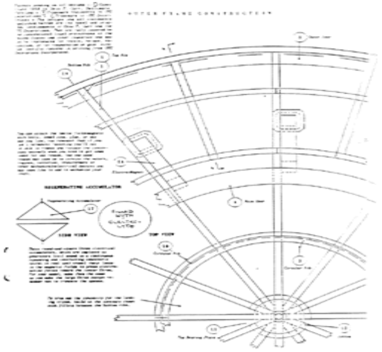
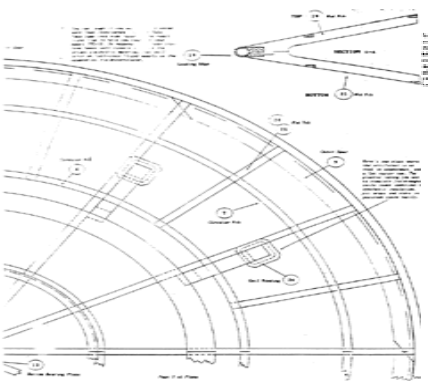
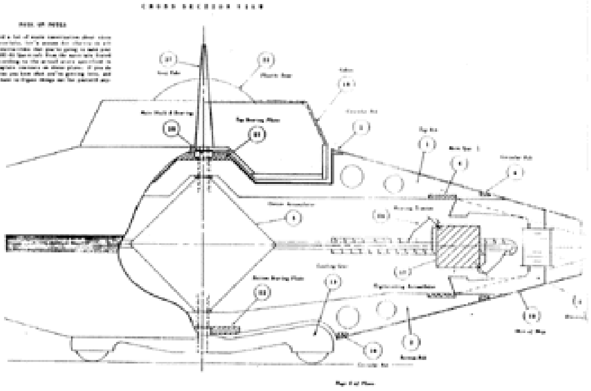
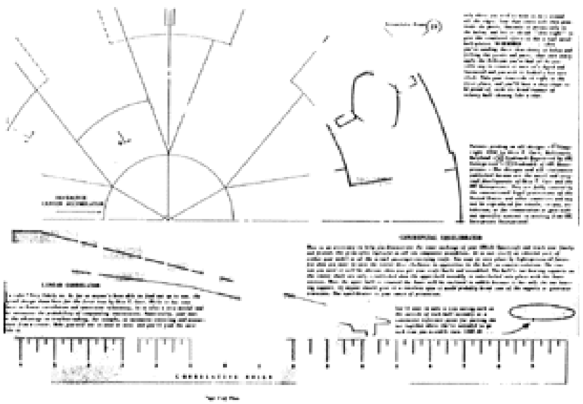
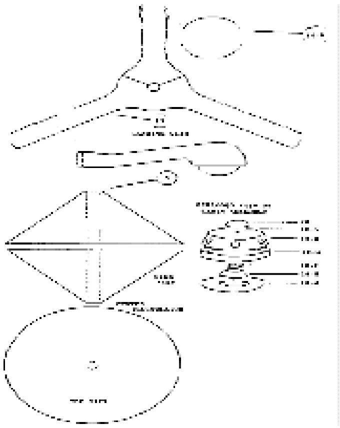

Additional Large Specifications Images (Links open in new windows): [1] ~ [2] ~ [3] ~ [4] ~ [5]

---

http://www.projectcamelot.org/ralph_ring.html

Excerpts from an Interview with Ralph Ring

Ralph Ring is a technician who worked with Otis T. Carr in the late 1950s and early 1960s. Ring stresses that resonance is the key to working with nature: "Resonance. You have to work with nature, not against her." According to Ring, when the model Utrons were powered up to operational speed, "...the metal turned to Jell-o. You could push your finger right into it. It ceased to be solid. It turned into another form of matter, which was as if it was not entirely here in this reality. That's the only way I can attempt to describe it. It was uncanny, one of the weirdest sensations I've ever felt."

Ralph Ring describes Otis Carr thus: "He was an unquestioned genius. Tesla had recognized his quality immediately and had taught him everything he knew. He was inspired, and – like Tesla – seemed to know exactly what to do to get something to work. He was a private man and was also very metaphysical in his thinking. I think the fact that he was not formally trained in physics helped him. He was not constrained by any preconceived ideas. As crazy as it sounds now, he was determined to fly to the moon and really believed it could be done. I believed it. We all did."

Ring described his "flight" in the Utron craft, saying that "Fly is not the right word. It traversed distance. It seemed to take no time. I was with two other engineers when we piloted the 45' craft about ten miles. I thought it hadn't moved --- I thought it had failed. I was completely astonished when we realized that we had returned with samples of rocks and plants from our destination. It was a dramatic success. It was more like a kind of teleportation.

"What's more, time was distorted somehow. We felt we were in the craft about fifteen or twenty minutes. We were told afterwards that we'd been carefully timed as having been in the craft no longer than three or four minutes. I still have no complete idea how it worked. We just built it exactly according to Carr's instructions. Everything had to be perfect... it all had to be just so, or it he said it would not work: a kind of symbiotic state between man and machine.

"The Utron was the key to it all. Carr said it accumulated energy because of its shape, and focused it, and also responded to our conscious intentions. When we operated the machine, we didn't work any controls. We went into a kind of meditative state and all three of us focused our intentions on the effect we wanted to achieve. It sounds ridiculous, I know. But that's what we did, and that's what worked. Carr had tapped into some principle which is not understood, in which consciousness melds with engineering to create an effect. You can't write that into equations. I have no idea how he knew it would work. But it did.

"...There are two secrets to making the alien saucers work. One is their advanced engineering, and the other is their mental ability."

"I've lost count of the number of people who have refused to believe what happened. I no longer talk about it. It's no fun being laughed at and ridiculed. But I've described it exactly as it occurred. One day someone will build the disk just as we did and they'll have the same experience. All his blueprints still exist. Nowadays, it would all be done with digital and solid state circuitry --- no moving parts would be necessary.

"I've heard that the aliens use the same principle to operate their craft. Their physics seems to work in harness with their consciousness. The craft amplifies the power of their minds. Their craft won't operate without the pilots. I've heard that's why we can't operate their craft --- or, at any rate, we can't operate them the way they do. We're just not adept enough mentally and spiritually. So there are two secrets to making the saucers work. One is the advanced engineering, and the other is the mental and spiritual ability. We may have duplicated some of the first, but we may be a long way from the second yet."

***FATE* (May 1958), page 17**

"Gravity Machine?"

The following summary was sent from a Ship's Paper of October 30,1957, after being copied from a CW News broadcast while at sea. It is unusual because no other report of this announcement reached us. It certainly is sensational --- if true.

Baltimore, MD (October 29):--- A group of inventors claimed Monday they have been able to utilize gravity in circular motion machines capable of powering everything from hearing aids to space cruisers.

Otis T. Carr, president of OTC Enterprises, Inc., detailed his claims in an interview and demonstration of a crude model of a circular motion machine which he said is the principle of a "free energy circular foil" spacecraft he can build, if someone puts up the money.

He said the machine can be adapted to devices of any size to produce continuous power absolutely free of dissipation.

Its immediate application, Carr said, would be in a spacecraft --- which would be able to fly among the planets in controlled flight.

It could land or take off as desired on the earth, the moon or any planet in the earth's solar system, he said.

Carr and his associates said their claims are based on the most simple practical applications of natural laws and discoveries in science and mathematics. They have no formal education in science or engineering.

He said the same "free energy" which causes the earth to rotate on its axis and orbit around the sun will turn a machine he described as two cones joined at their circular bases.

When the rotation of such a machine reaches a certain velocity relative to the earth's orbital velocity, Carr said, it will take off.

Carr said the core of his space ship would be a huge battery which would spin at the velocity of the external craft and which would be recharged, he said, by its motion. Carr declared such a battery, built to any size, could be designed to power the largest electric generating plant, operate an automobile, heat a house or power any conceivable machine or device.

The principle on which Carr said such circular motion machines would operate is that "any vehicle accelerated to an axis rotation relative to its attractive inertial mass (the earth) immediately becomes activated by free space energy and acts as an independent force."

***FATE* (August 1959), page 32.**

"The Saucer That Didn't Fly"

by

W. E. Du Soir

The serious field of UFO's and flying saucer research received a setback at Oklahoma City in late April when a highly publicized launching attempt by O.T.C. Enterprises of Baltimore, Md., resulted in failure. Hundreds of persons had been invited to Oklahoma City by Otis T. Carr to watch him "launch a six-foot prototype model of the O.T.C. X-1, a space craft which works on 'utron' energy." Those who were there came away disappointed. The flying saucer did not fly. The six-foot model, well-hidden in a warehouse, never even got to the launching site, or to the amusement park called Frontier City, where it was to have been displayed beside a 45-foot version of the same X-1 which will be an amusement "ride" in the near future.

The launching site was to have been a gravel pit eight miles north of Oklahoma City where Carr claimed the saucer would rise "400 to 600 feet." Launching time was set for 3:00 PM Sunday, April 19. Two and a half hours past this time Major Wayne Aho finally announced that the launching had to be postponed because of "technical difficulties." Later it was stated by O.T.C. officials that one bearing housing was "off one-sixteenth of an inch." Most of the major flying saucer clubs had ignored the "launching". Except for a story in early April there was no advance news on Carr's experiment in Oklahoma City newspapers.

One Oklahoma City television reporter expressed the general feeling of the townspeople, "This thing will never leave the ground. And I feel that a great deal of the ballyhoo they're giving out is tied in with the ride at Frontier City. I have tried constantly to get in to see the saucer model, but they've kept it hidden." Equally well hidden was Carr himself. He was finally located in Mercy Hospital, Room 302, by John Nebel, the famed "Long John" of WOR, New York City. Carr did not show up at a meeting at a local church the day before the scheduled launching but a taped speech was delivered to the 70 people who attended. "Barring any flat tires," Carr said, "I feel that history will be made Sunday afternoon when the model of the O.T.C. X-1 is launched here."

Carr did not leave his hospital room until Monday, when he was taken to the home of Major Aho, the man in charge of publicity for the company. A scheduled "victory" dinner was pretty dismal Sunday night, where again a taped message from Carr was presented. Monday morning at 5 AM a rumor spread that the model launching would again be attempted. At that early hour a group of a half-dozen people, including Norman Colton, Carr's chief aide, went to the warehouse. But it was three hours later before anyone else showed up and most of those who had come to see the model had given up in disgust and left. Many believers in UFO's, although skeptical, had traveled to Oklahoma City to see what developed. All of them had left town by Monday afternoon.

One member of a Pennsylvania saucer club commented, "I don't know what's going on but I feel they never had any intention of trying to launch the model. I could not see any plans in sight for the model and, in fact, I understand, that a Mr. Maywood Jones presented only what he called "three-dimensional illustrations" of Carr's ideas." Monday afternoon there was more activity in the warehouse and finally, around 4 PM, the model was "turned on" and revolved, although it showed no signs of taking off from the table on which it sat. "We hoped to try to launch it again but unfortunately one of the seams burst during the test," an O.T.C. official said. "We still hope to have a launching shortly," another member of Carr's group added. Left unanswered were such questions as:

1) Why was there no publicity about the launching in Oklahoma City for two weeks before the model was supposed to fly?

2) Why did Carr ready a taped message for delivery at the "victory dinner" a day before the scheduled launching?

3) How could any engineer make a mistake of one-16th of an inch in a precision part? This would be similar to a navigator landing an airplane 300 miles off course. A mistake of one-10,000th of an inch can be a major engineering calamity.

Carr was apparently undaunted by the weekend's developments. He again commented, "This indeed is an historic occasion." Questioned about the failures of his model, he added, "The theory itself has been proven. The model will go up, if not this week, then next week or next month." Mr. Carr also has stated that he and Major Wayne Aho will "fly to the moon in a flying saucer on December 7, 1959."

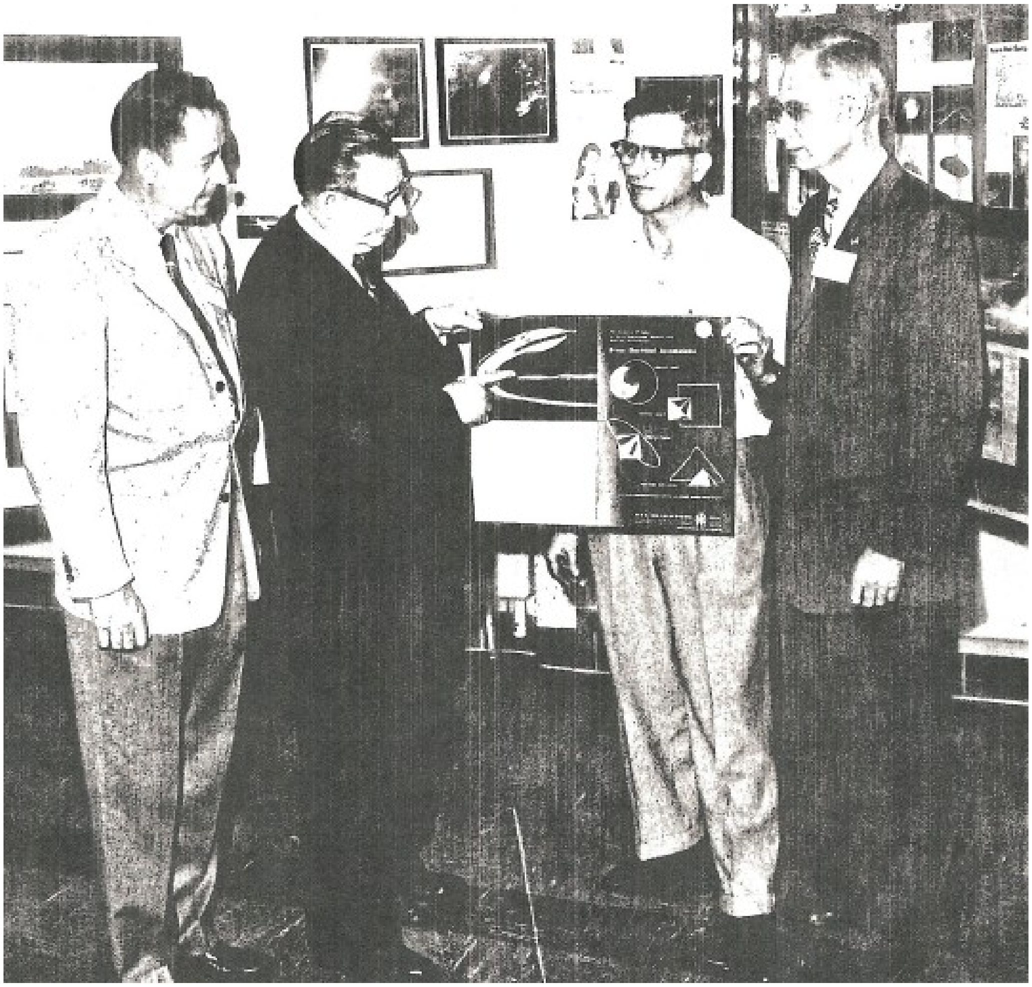

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

No. 674,427. Patented May 21, 1901.

A. PALENCSÁR.

APPARATUS FOR COLLECTING ATMOSPHERIC ELECTRICITY.

(Application filed July 10, 1900.)

(No Model.) 3 Sheets—Sheet 1.

Fig: 1.

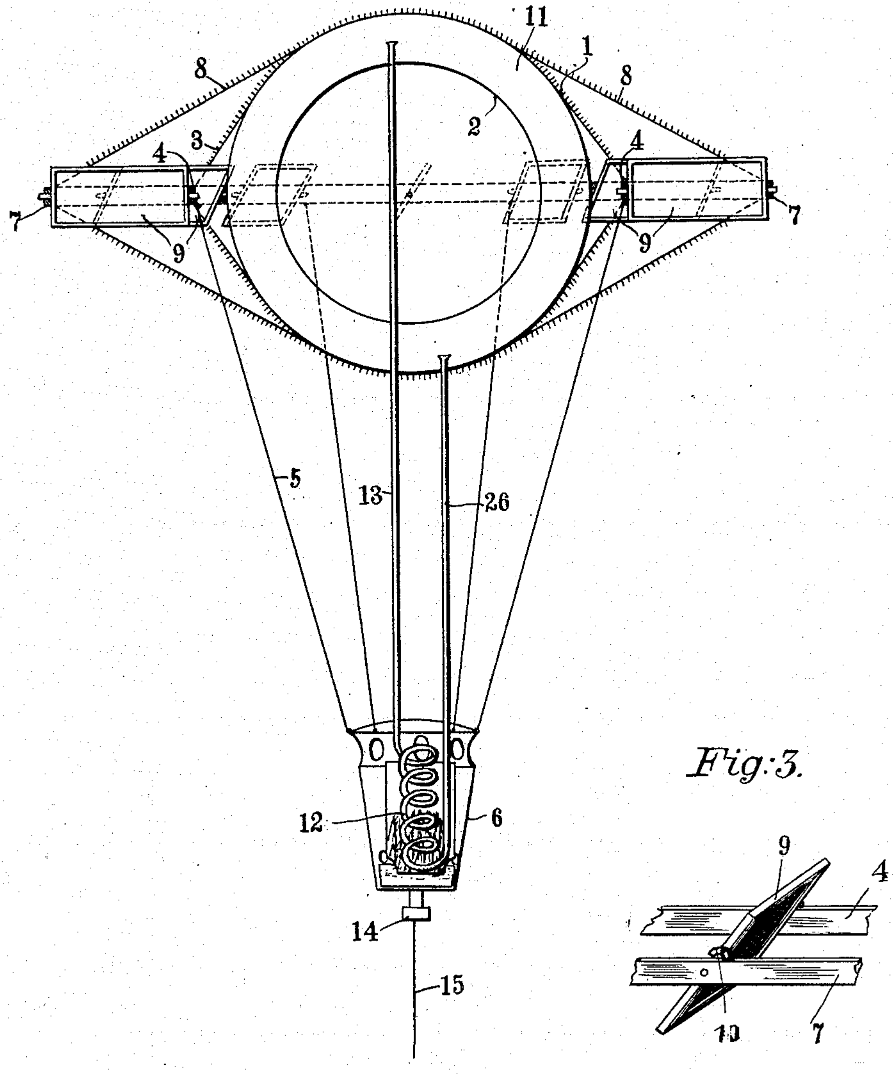
11
1
8
8
2
3
4
4
7
7
9
9
9
9
5
13
26
12
6
14
15

Fig: 3.

9
4
10
7

WITNESSES:
Ella L. Giles
Otto Munk

INVENTOR
Andor Palencsár
BY Richards
ATTORNEYS

THE NORRIS PETERS CO., PHOTO-LITHO., WASHINGTON, D. C.

No. 674,427. Patented May 21, 1901.

A. PALENCSÁR.

APPARATUS FOR COLLECTING ATMOSPHERIC ELECTRICITY.

(Application filed July 10, 1900.)

(No Model.) 3 Sheets—Sheet 2.

Fig: 2.

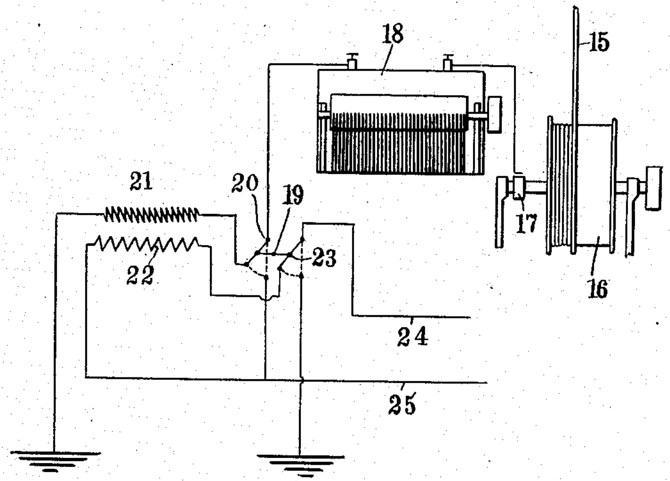

WITNESSES:
Ella L. Giles
Otto Munk

INVENTOR
Andor Palencsár
BY
Richards & Co
ATTORNEYS

THE NORRIS PETERS CO., PHOTO-LITHO., WASHINGTON, D. C.

No. 674,427. Patented May 21, 1901.

A. PALENCSÁR.

APPARATUS FOR COLLECTING ATMOSPHERIC ELECTRICITY.

(Application filed July 10, 1900.)

(No Model.) 3 Sheets—Sheet 3.

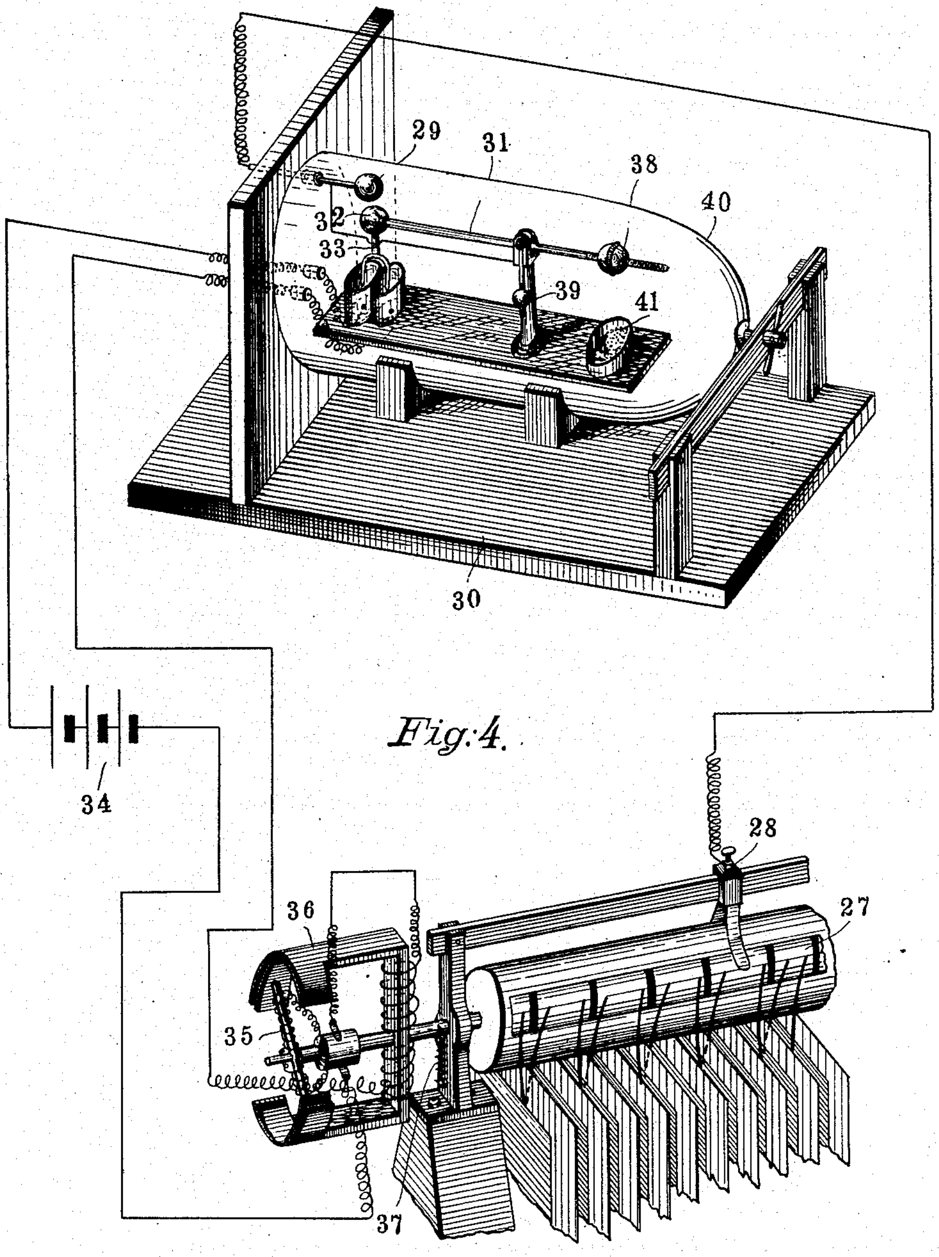
29
31
38
40
32
33
39
41
30
34
28
27
36
35
37

Fig. 4.

WITNESSES:

Ella L. Giles

Otto Munk

INVENTOR

Andor Palencsár

BY

Richards & Co

ATTORNEYS

THE NORRIS PETERS CO., PHOTO-LITHO., WASHINGTON, D. C.

UNITED STATES PATENT OFFICE.

ANDOR PALENCSÁR, OF BUDA-PESTH, AUSTRIA-HUNGARY.

APPARATUS FOR COLLECTING ATMOSPHERIC ELECTRICITY.

SPECIFICATION forming part of Letters Patent No. 674,427, dated May 21, 1901.

Application filed July 10, 1900. Serial No. 23,102. (No model.)

To all whom it may concern:

Be it known that I, ANDOR PALENCSÁR, a subject of the King of Hungary, residing at Buda-Pesth, Austria-Hungary, have invented a certain new and useful Apparatus for Collecting Atmospherical Electricity, of which the following is a full, clear, and exact specification.

The present invention concerns an apparatus for collecting and driving or conducting atmospherical electricity which renders it practicable to obtain material quantities of the same in serviceable form.

The experiments made hitherto have been limited to deviating or conducting the electricity by simple lightning-rods without making allowance for the progress of the theory of atmospherical electricity. It is obvious that only extremely small quantities of electricity can be collected in this manner, and, moreover, it was obtained in a form which entirely precluded its utilization. It is evident that a source of electricity of irregular yield can only be utilized by means of accumulators and for charging accumulators only a current of constant potential can be employed. The potential of electricity derived or deviated by means of a lightning-rod varies within wide limits, and, moreover, it is so high that it cannot be used at all for charging accumulators. All these drawbacks are remedied by the present invention, by which the atmospheric electricity is obtained in larger quantities with as low and constant a potential as may be desired.

The idea underlying the invention is based on the modern theory of atmospheric electricity, according to which it is produced by the condensation of steam or aqueous vapors, and that the increase of potential is effected by the concentration of the small drops of water into larger ones, as the proportion of the surface of the drops to the volume of same is materially reduced thereby. According to this theory the water-droplets floating in the layers of air are considered as vehicles or carriers of the electricity, and a rational system of the deviation of the atmospheric electricity must derive it from the water-drops. This is attained by the present invention in the following manner: A collecting-body of as large a surface as possible is provided with sharp needles. This is moved in the higher layers of air, while being continually heated by a suitable heating device. Owing to the heat the water-drops immediately ambient to the collecting-body will be evaporated, their capacity is gradually reduced, while the potential of the charge grows until it reaches an infinite height with infinite smallness of the drops. It is readily apparent that the whole charge of the drops as soon as they have been evaporated will have passed to the collector or collecting-body, from which it can be conducted. For the purpose of replacing the evaporated drops, which have been deprived of their charge by new-charged drops, the collector is moved in relation to the ambient air.

One way of carrying out the invention is represented in diagram in the accompanying drawings, in which—

Figure 1 is a longitudinal section of the collecting-balloon. Fig. 2 shows the deviating and conversion device, and Fig. 3 a detail. Fig. 4 is a diagrammatic perspective view of a form of automatic regulator for the rheostatic machine.

The apparatus consists of a balloon having two walls and covered with a light wire net, preferably of aluminium wire, said net being studded with needles. Besides, the balloon carries the net 3, on which the ring 4, made of a solid but light material, (wood, cane, &c.,) is fixed. This ring carries the basket 6 by means of the cords or ropes 5. On a level with the ring 4 is the ring 7, which is kept spread by the blades or wings 9, which are journaled in a manner to rotate easily. The ring 7 is kept fixedly in position by the cords or ropes 8. The blades or wings consist of a frame covered with light material, and their rotation in either direction is limited by the stops or ledges 10. The blades or wings form advantageously an angle of sixty to seventy degrees with the vertical line.

All wire nets, ropes, rings, and blades or wings may be covered with small metal needles which are electrically connected with each other.

From the space 11, formed between the double walls of the balloon, the pipe 26 leads from the lowest point of the same to the serpentine 12, arranged in the basket 6, which in turn is

connected with the pipe 13, ending in the upper part of the space 11 between the two walls 12. The serpentine is heated by means of a suitable source of heat, whereby a warm current of gas or air circulates continually between the double walls of the balloon.

Under the basket the ball-bearing 14 is arranged in electrical connection with the wire nets, and its stud is electrically connected with the carefully-insulated light though sufficiently strong cable 15.

On the earth's surface is a winch 16, Fig. 2, by means of which the balloon can be made to ascend or descend as soon as the interior space of the balloon is filled with illuminating gas or hydrogen.

The end of the cable-core is soldered to a collector arranged upon and insulated from the axle of the winch, and the electricity is conducted from this collector by means of a sliding contact. The collecting of the electricity takes place by moving the balloon continuously up and down by means of the winch. In this movement the balloon is turned by means of the wings or blades 9, which are adapted to turn the balloon always in the same direction whether ascending or descending, as in the change from ascending into descending of the balloon, or vice versa, the blades are turned over by the aerial resistance, and thus impart the rotating motion to the balloon in the same direction. In order to avoid torsion of the cable, the ball-bearing 14 is provided. This up-and-down motion and rotating of the balloon accomplishes the purpose of bringing it into contact with as many water particles floating in the air as possible.

As the electricity conducted from the collector-wing 17 possesses a much too high and varying potential for making its direct application practicable, and as it is usual with an irregular source of electricity to first charge accumulators and to further utilize the easily-regulated current of the same only, it becomes necessary to seek to maintain the electricity conducted from the collector 17 for the charging of the accumulators at a constant potential and convert the potential to a much lower one; but as we deal in this case with a direct current ordinary converters cannot be used for this purpose. Moreover, the electricity possesses in this case a much too high potential, so that with the employment of ordinary converters the greatest part of the collected electricity would be lost again. The only practical method for this purpose is the converting by means of the Planté rheostatic machine, by which this high-voltage electricity can be transformed almost without any loss whatever. Thus the question of conversion would be solved, and only the question of maintenance of constant potential remains. This is obtained in the following manner: The rheostatic machine 18 or only a part of the plates is connected with an electrometer of any construction, the movable part of which closes a contact which eventually actuates an electromagnet which effects the switching or reversion of the rheostatic machine. After the switching of the rheostatic machine it is discharged, the potential falls to zero, (*nil*,) and the electrometer resumes its initial position, whereby the current of the electromagnet which effects the reversion is interrupted and the plates of the rheostatic machine are reconnected to potential. The machine is then ready for renewed charge and is again discharged when the determined potential is reached. This action is continually repeated as long as the apparatus is in operation.

A form of the automatic regulator for the rheostatic machine is represented in Fig. 4. 27 is the contact-cylinder of the rheostatic machine, on which, for clearness sake, only the contacts for the charging position of the condenser-plates are shown, while the contacts for the discharging position, which come into action after the contact-cylinder has been turned, are omitted. The coating of the condenser-plates is connected in electrical circuit with the stationary ball 29 and the movable ball 32 of the charge-meter 30. When the charge of the rheostatic machine rises, the ball 32, arranged on the one extremity of the lever 31, is repelled and at a certain stroke actuated by means of the fork 33, which is fastened on the ball 32, and by dipping into a mercury-cup it closes the circuit of the source of current 34. This current passes through the windings of the anchor or armature 35, fixed on the shaft of the contact-cylinder, and it also passes through the electromagnet 36. Thereby the contact-cylinder is turned by a certain angle and the rheostatic machine is thus reversed. Then if the tension decreases by discharge and the repulsion of the balls 29 32 declines so far that the contact is interrupted at 33 the spring 37 turns the contact-cylinder into its normal position and the rheostatic machine is again switched to tension. The regulation of the electrometer is effected by the adjustable weights 38 39. When larger quantities of electricity shall be derived or deviated, two rheostatic machines may operate alternately so, that while the one is being discharged the other can be charged. It is readily apparent that if the capacity of the rheostatic machine is not changed the switching or reversion by the electrometer will always take place at the same potential of the rheostatic machine, and as the number of plates, and hence the proportion of conversion, remains the same the current impulses derived or deviated from the rheostatic machine will also have the same potential.

The irregularities of the source of electricity change the interval of time in which the charges follow each other; but as long as the potential remains constant this has no injurious effect on the charge of the accumulators. The current derived. or deviated from the

rheostatic machine can be further transformed by an ordinary converter 21 22, and whenever it is sufficiently constant it can be utilized without the intervention of the accumulators. The converters can be connected or disconnected by means of the double switch 19 20 23.

24 and 25 are the conducting-wires, which run either directly to the place of consumption or to an accumulator-battery.

Having now described my invention, what I claim as new, and desire to secure by Letters Patent, is—

1. An apparatus for collecting atmospheric electricity comprising a collecting-body adapted to be kept in motion, heating means for said body, and a rheostatic machine and a converter connected with the said body electrically, substantially as described.

2. An apparatus for collecting atmospheric electricity for storage comprising a collecting-body adapted to be kept in motion in the ambient air, heating means for said body, a conductor leading from said body, a rheostatic machine connected with the said conductor, an electrometer connected with said rheostatic machine electrically, a contact controlled by the electrometer, and an electromagnet controlled by said contact, said electromagnet controlling the reversing of the rheostatic machine, substantially as described.

3. In combination, the balloon-like collecting-body, having the collecting-points, means for heating the interior space of said body, a conductor leading from the balloon-like body, and electrical devices for receiving the current therefrom, substantially as described.

4. In combination, the balloon-like collecting-body, means for turning the same constantly in one direction in both ascending and descending, conducting means leading from the balloon and electrical devices for receiving the current therefrom, substantially as described.

In witness whereof I have hereunto set my hand in presence of two witnesses.

ANDOR PALENCSÁR.

Witnesses:
EUGÈNE HARRÄNJO,
PAUL BÖLESKEY.

United States Patent [19]

Malech

[11] **3,951,134**

[45] **Apr. 20, 1976**

---

[54] **APPARATUS AND METHOD FOR REMOTELY MONITORING AND ALTERING BRAIN WAVES**

[75] Inventor: **Robert G. Malech**, Plainview, N.Y.

[73] Assignee: **Dorne & Margolin Inc.**, Bohemia, N.Y.

[22] Filed: **Aug. 5, 1974**

[21] Appl. No.: **494,518**

[52] **U.S. Cl.** .... **128/2.1 B**
[51] **Int. Cl.**^{2} **A61B 5/04**
[58] **Field of Search** .... 128/1 C, 1 R, 2.1 B, 128/2.1 R, 419 R, 422 R, 420, 404, 2 R, 2 S, 2.05 R, 2.05 V, 2.05 F, 2.06 R; 340/248 A, 258 A, 258 B, 258 D, 229

[56] **References Cited**

UNITED STATES PATENTS

2,860,627	11/1958	Harden et al.	128/2.1 B
3,096,768	7/1963	Griffith, Jr.	128/420
3,233,450	2/1966	Fry	128/2.1 R
3,483,860	12/1969	Namerow	128/2.05 F
3,495,596	2/1970	Condict	128/1 C
3,555,529	1/1971	Brown et al.	128/2.1 R
3,773,049	11/1973	Rabichev et al.	128/1 C
3,796,208	3/1974	Bloice	128/2 S

Primary Examiner—William E. Kamm
Attorney, Agent, or Firm—Darby & Darby

[57] **ABSTRACT**

Apparatus for and method of sensing brain waves at a position remote from a subject whereby electromagnetic signals of different frequencies are simultaneously transmitted to the brain of the subject in which the signals interfere with one another to yield a waveform which is modulated by the subject's brain waves. The interference waveform which is representative of the brain wave activity is re-transmitted by the brain to a receiver where it is demodulated and amplified. The demodulated waveform is then displayed for visual viewing and routed to a computer for further processing and analysis. The demodulated waveform also can be used to produce a compensating signal which is transmitted back to the brain to effect a desired change in electrical activity therein.

11 Claims, 2 Drawing Figures

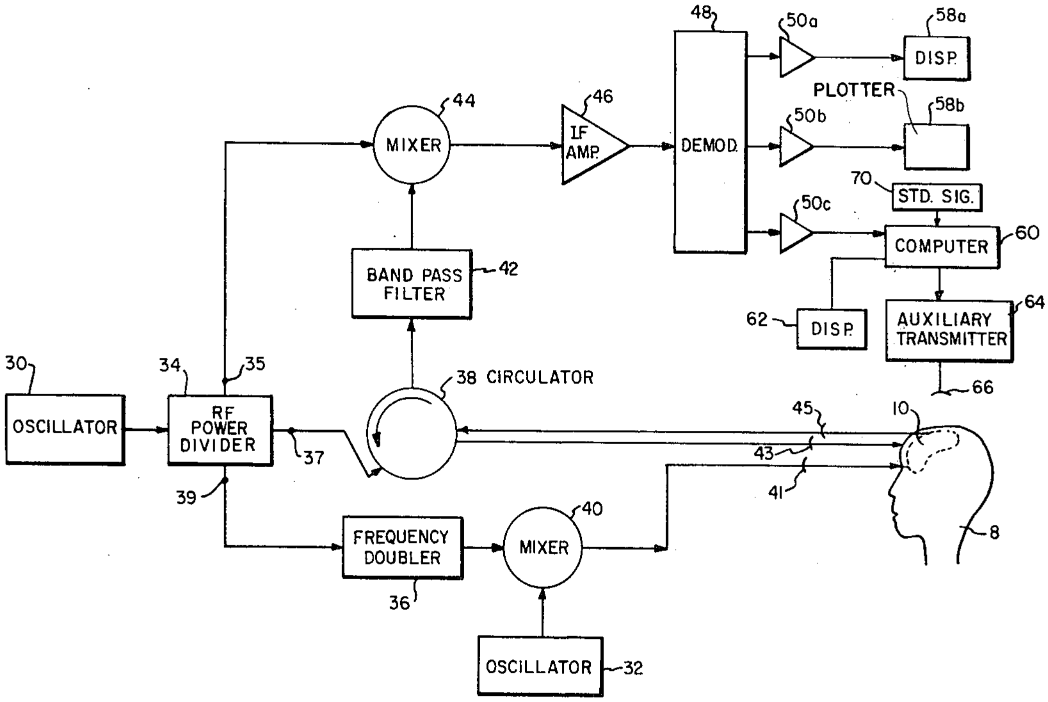
48
50a
58a
DISP.
PLOTTER
58b
44
MIXER
46
IF AMP.
DEMOD.
50b
70
STD. SIG.
50c
60
COMPUTER
BAND PASS FILTER
42
62
DISP.
AUXILIARY TRANSMITTER
64
66
30
OSCILLATOR
34
35
RF POWER DIVIDER
37
38 CIRCULATOR
45
10
43
41
8
39
FREQUENCY DOUBLER
36
MIXER
40
OSCILLATOR
32

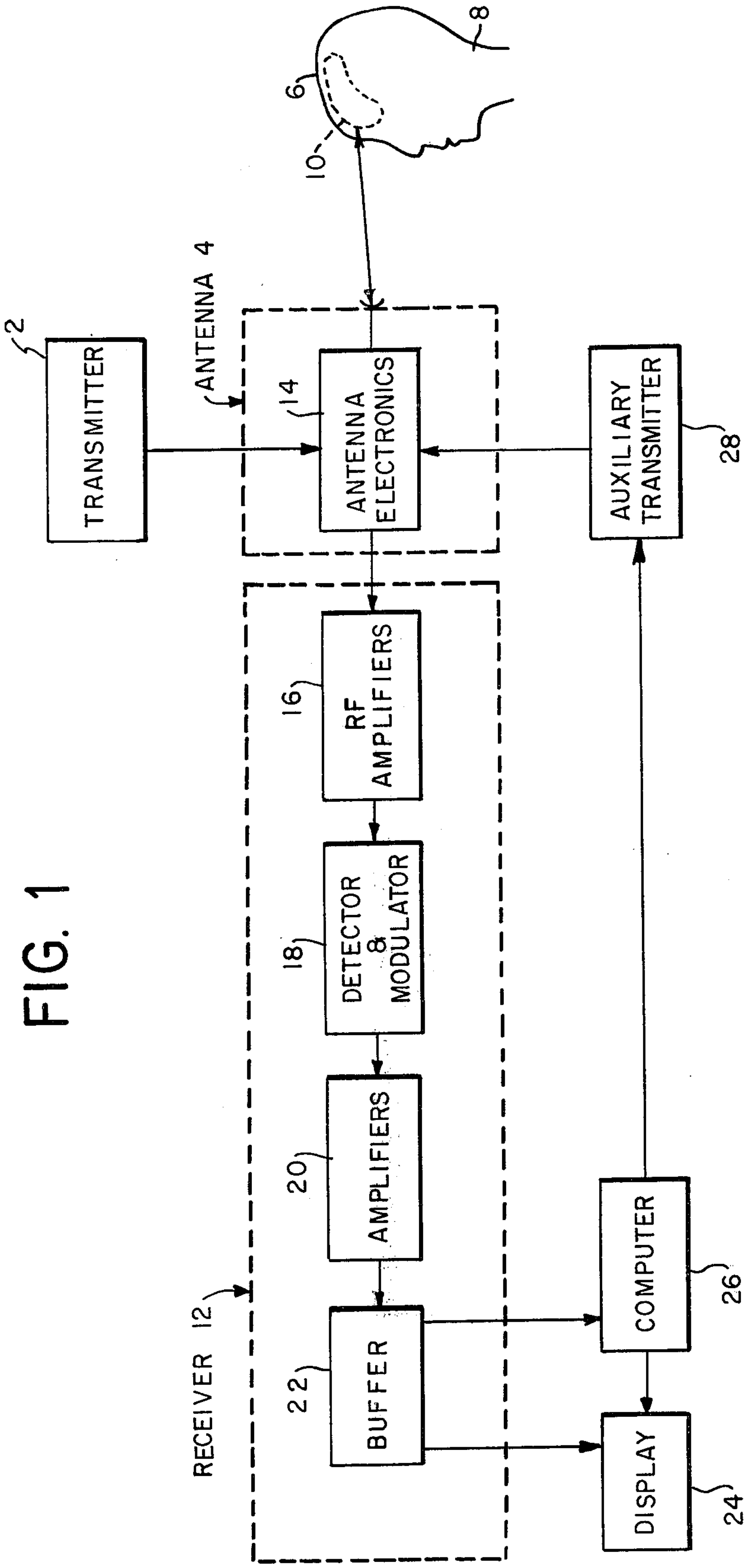
TRANSMITTER
2
ANTENNA 4
ANTENNA ELECTRONICS
14
AUXILIARY TRANSMITTER
28
6
8
10
RF AMPLIFIERS
16
DETECTOR & MODULATOR
18
AMPLIFIERS
20
BUFFER
22
RECEIVER 12
COMPUTER
26
DISPLAY
24

FIG. 1

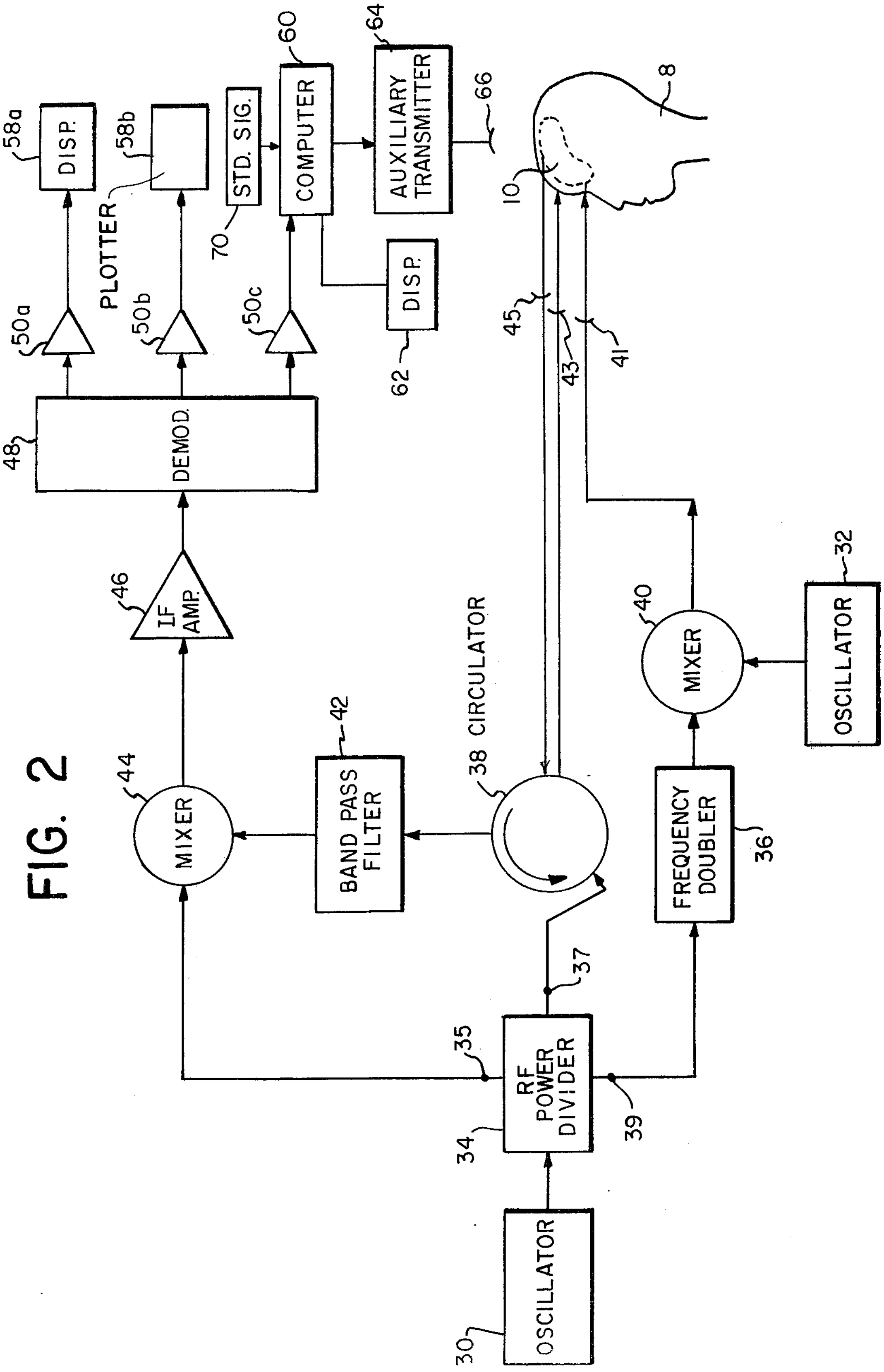
FIG. 2
OSCILLATOR
RF POWER DIVIDER
FREQUENCY DOUBLER
MIXER
OSCILLATOR
38 CIRCULATOR
BAND PASS FILTER
MIXER
IF AMP.
DEMOD.
DISP.
PLOTTER
STD. SIG.
COMPUTER
AUXILIARY TRANSMITTER
DISP.

APPARATUS AND METHOD FOR REMOTELY MONITORING AND ALTERING BRAIN WAVES

BACKGROUND OF THE INVENTION

Medical science has found brain waves to be a useful barometer of organic functions. Measurements of electrical activity in the brain have been instrumental in detecting physical and psychic disorder, measuring stress, determining sleep patterns, and monitoring body metabolism.

The present art for measurement of brain waves employs electroencephalographs including probes with sensors which are attached to the skull of the subject under study at points proximate to the regions of the brain being monitored. Electrical contact between the sensors and apparatus employed to process the detected brain waves is maintained by a plurality of wires extending from the sensors to the apparatus. The necessity for physically attaching the measuring apparatus to the subject imposes several limitations on the measurement process. The subject may experience discomfort, particulary if the measurements are to be made over extended periods of time. His bodily movements are restricted and he is generally confined to the immediate vicinity of the measuring apparatus. Furthermore, measurements cannot be made while the subject is conscious without his awareness. The comprehensiveness of the measurements is also limited since the finite number of probes employed to monitor local regions of brain wave activity do not permit observation of the total brain wave profile in a single test.

SUMMARY OF THE INVENTION

The present invention relates to apparatus and a method for monitoring brain waves wherein all components of the apparatus employed are remote from the test subject. More specifically, high frequency transmitters are operated to radiate electromagnetic energy of different frequencies through antennas which are capable of scanning the entire brain of the test subject or any desired region thereof. The signals of different frequencies penetrate the skull of the subject and impinge upon the brain where they mix to yield an interference wave modulated by radiations from the brain's natural electrical activity. The modulated interference wave is re-transmitted by the brain and received by an antenna at a remote station where it is demodulated, and processed to provide a profile of the suject's brain waves. In addition to passively monitoring his brain waves, the subject's neurological processes may be affected by transmitting to his brain, through a transmitter, compensating signals. The latter signals can be derived from the received and processed brain waves.

OBJECTS OF THE INVENTION

It is therefore an object of the invention to remotely monitor electrical activity in the entire brain or selected local regions thereof with a single measurement.

Another object is the monitoring of a subject's brain wave activity through transmission and reception of electromagnetic waves.

Still another object is to monitor brain wave activity from a position remote from the subject.

A further object is to provide a method and apparatus for affecting brain wave activity by transmitting electromagnetic signals thereto.

DESCRIPTION OF THE DRAWINGS

Other and further objects of the invention will appear from the following description and the accompanying drawings, which form part of the instant specification and which are to be read in conjunction therewith, and in which like reference numerals are used to indicate like parts in the various views;

FIG. 1 is a block diagram showing the interconnection of the components of the apparatus of the invention;

FIG. 2 is a block diagram showing signal flow in one embodiment of the apparatus.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring to the drawings, specifically FIG. **1**, a high frequency transmitter **2** produces and supplies two electromagnetic wave signals through suitable coupling means **14** to an antenna **4**. The signals are directed by the antenna **4** to the skull **6** of the subject **8** being examined. The two signals from the antenna **4**, which travel independently, penetrate the skull **6** and impinge upon the tissue of the brain **10**.

Within the tissue of the brain **10**, the signals combine, much in the manner of a conventional mixing process technique, with each section of the brain having a different modulating action. The resulting waveform of the two signals has its greatest amplitude when the two signals are in phase and thus reinforcing one another. When the signals are exactly 180° out of phase the combination produces a resultant waveform of minimum amplitude. If the amplitudes of the two signals transmitted to the subject are maintained at identical levels, the resultant interference waveform, absent influences of external radiation, may be expected to assume zero intensity when maximum interference occurs, the number of such points being equal to the difference in frequencies of the incident signals. However, interference by radiation from electrical activity within the brain **10** causes the waveform resulting from interference of the two transmitted signals to vary from the expected result, i.e., the interference waveform is modulated by the brain waves. It is believed that this is due to the fact that brain waves produce electric charges each of which has a component of electromagnetic radiation associated with it. The electromagnetic radiation produced by the brain waves in turn reacts with the signals transmitted to the brain from the external source.

The modulated interference waveform is re-transmitted from the brain **10**, back through the skull **6**. A quantity of energy is re-transmitted sufficient to enable it to be picked up by the antenna **4**. This can be controlled, within limits, by adjusting the absolute and relative intensities of the signals, originally transmitted to the brain. Of course, the level of the transmitted energy should be kept below that which may be harmful to the subject.

The antenna passes the received signal to a receiver **12** through the antenna electronics **14**. Within the receiver the wave is amplified by conventional RF amplifiers **16** and demodulated by conventional detector and modulator electronics **18**. The demodulated wave, representing the intra-brain electrical activity, is amplified by amplifiers **20** and the resulting information in electronic form is stored in buffer circuitry **22**. From the buffers **22** the information is fed to a suitable visual display **24**, for example one employing a cathode ray tube, light emitting diodes, liquid crystals, or a mechanical plotter. The information may also be channeled to a computer **26** for further processing and analysis with the output of the computer displayed by heretofore mentioned suitable means.

In addition to channeling its information to display devices **24**, the computer **26** can also produce signals to control an auxiliary transmitter **28**. Transmitter **28** is used to produce a compensating signal which is transmitted to the brain **10** of the subject **8** by the antenna **4**. In a preferred embodiment of the invention, the compensating signal is derived as a function of the received brain wave signals, although it can be produced separately. The compensating signals affect electrical activity within the brain **10**.

Various configurations of suitable apparatus and electronic circuitry may be utilized to form the system generally shown in FIG. **1** and one of the many possible configurations is illustrated in FIG. **2**. In the example shown therein, two signals, one of 100 MHz and the other of 210 MHz are transmitted simultaneously and combine in the brain **10** to form a resultant wave of frequency equal to the difference in frequencies of the incident signals, i.e., 110 MHz. The sum of the two incident frequencies is also available, but is discarded in subsequent filtering. The 100 MHz signal is obtained at the output **37** of an RF power divider **34** into which a 100 MHz signal generated by an oscillator **30** is injected. The oscillator **30** is of a conventional type employing either crystals for fixed frequency circuits or a tunable circuit set to oscillate at 100 MHz. It can be a pulse generator, square wave generator or sinusoidal wave generator. The RF power divider can be any conventional VHF, UHF or SHF frequency range device constructed to provide, at each of three outputs, a signal identical in frequency to that applied to its input.

The 210 MHz signal is derived from the same 100 MHz oscillator **30** and RF power divider **34** as the 100 MHz signal, operating in concert with a frequency doubler **36** and 10 MHz oscillator **32**. The frequency doubler can be any conventional device which provides at its output a signal with frequency equal to twice the frequency of a signal applied at its input. The 10 MHz oscillator can also be of conventional type similar to the 100 MHz oscillator herebefore described. A 100 MHz signal from the output **39** of the RF power divider **34** is fed through the frequency doubler **36** and the resulting 200 MHz signal is applied to a mixer **40**. The mixer **40** can be any conventional VHF, UHF or SHF frequency range device capable of accepting two input signals of differing frequencies and providing two output signals with frequencies equal to the sum and difference in frequencies respectively of the input signals. A 10 MHz signal from the oscillator **32** is also applied to the mixer **40**. The 200 MHz signal from the doubler **36** and the 10 MHz signal from the oscillator **32** combine in the mixer **40** to form a signal with a frequency of 210 MHz equal to the sum of the frequencies of the 200 MHz and 10 MHz signals.

The 210 MHz signal is one of the signals transmitted to the brain **10** of the subject being monitored. In the arrangement shown in FIG. **2**, an antenna **41** is used to transmit the 210 MHz signal and another antenna **43** is used to transmit the 100 MHz signal. Of course, a single antenna capable of operating at 100 MHz and 210 MHz frequencies may be used to transmit both signals. The scan angle, direction and rate may be controlled mechanically, e.g., by a reversing motor, or electronically, e.g., by energizing elements in the antenna in proper synchronization. Thus, the antenna(s) can be of either fixed or rotary conventional types.

A second 100 MHz signal derived from output terminal **37** of the three-way power divider **34** is applied to a circulator **38** and emerges therefrom with a desired phase shift. The circulator **38** can be of any conventional type wherein a signal applied to an input port emerges from an output port with an appropriate phase shift. The 100 MHz signal is then transmitted to the brain **10** of the subject being monitored via the antenna **43** as the second component of the dual signal transmission. The antenna **43** can be of conventional type similar to antenna **41** herebefore described. As previously noted, these two antennas may be combined in a single unit.

The transmitted 100 and 210 MHz signal components mix within the tissue in the brain **10** and interfere with one another yielding a signal of a frequency of 110 MHz, the difference in frequencies of the two incident components, modulated by electromagnetic emissions from the brain, i.e., the brain wave activity being monitored. This modulated 110 MHz signal is radiated into space.

The 110 MHz signal, modulated by brain wave activity, is picked up by an antenna **45** and channeled back through the circulator **38** where it undergoes an appropriate phase shift. The circulator **38** isolates the transmitted signals from the received signal. Any suitable diplexer or duplexer can be used. The antenna **45** can be of conventional type similar to antennas **41** and **43**. It can be combined with them in a single unit or it can be separate. The received modulated 110 MHz signal is then applied to a band pass filter **42**, to eliminate undesirable harmonics and extraneous noise, and the filtered 110 MHz signal is inserted into a mixer **44** into which has also been introduced a component of the 100 MHz signal from the source **30** distributed by the RF power divider **34**. The filter **42** can be any conventional band pass filter. The mixer **44** may also be of conventional type similar to the mixer **40** herebefore described.

The 100 MHz and 110 MHz signals combine in the mixer **44** to yield a signal of frequency equal to the difference in frequencies of the two component signals, i.e., 10 MHz still modulated by the monitored brain wave activity. The 10 MHz signal is amplified in an IF amplifier **46** and channeled to a demodulator **48**. The IF amplifier and demodulator **48** can both be of conventional types. The type of demodulator selected will depend on the characteristics of the signals transmitted to and received from the brain, and the information desired to be obtained. The brain may modulate the amplitude, frequency and/or phase of the interference waveform. Certain of these parameters will be more sensitive to corresponding brain wave characteristics than others. Selection of amplitude, frequency or phase demodulation means is governed by the choice of brain wave characteristic to be monitored. If desired, several different types of demodulators can be provided and used alternately or at the same time.

The demodulated signal which is representative of the monitored brain wave activity is passed through audio amplifiers **50** *a, b, c* which may be of conventional type where it is amplified and routed to displays **58** *a, b, c* and a computer **60**. The displays **58** *a, b, c* present the raw brain wave signals from the amplifiers

50 *a*, *b*, *c*. The computer **60** processes the amplified brain wave signals to derive information suitable for viewing, e.g., by suppressing, compressing, or expanding elements thereof, or combining them with other information-bearing signals and presents that information on a display **62**. The displays can be conventional ones such as the types herebefore mentioned employing electronic visual displays or mechanical plotters **58***b*. The computer can also be of conventional type, either analog or digital, or a hybrid.

A profile of the entire brain wave emission pattern may be monitored or select areas of the brain may be observed in a single measurement simply by altering the scan angle and direction of the antennas. There is no physical contact between the subject and the monitoring apparatus. The computer **60** also can determine a compensating waveform for transmission to the brain **10** to alter the natural brain waves in a desired fashion. The closed loop compensating system permits instantaneous and continuous modification of the brain wave response pattern.

In performing the brain wave pattern modification function, the computer **60** can be furnished with an external standard signal from a source **70** representative of brain wave activity associated with a desired nuerological response. The region of the brain responsible for the response is monitored and the received signal, indicative of the brain wave activity therein, is compared with the standard signal. The computer **60** is programmed to determine a compensating signal, responsive to the difference between the standard signal and received signal. The compensating signal, when transmitted to the monitored region of the brain, modulates the natural brain wave activity therein toward a reproduction of the standard signal, thereby changing the neurological response of the subject.

The computer **60** controls an auxiliary transmitter **64** which transmits the compensating signal to the brain **10** of the subject via an antenna **66.** The transmitter **64** is of the high frequency type commonly used in radar applications. The antenna **66** can be similar to antennas **41, 43** and **45** and can be combined with them. Through these means, brain wave activity may be altered and deviations from a desired norm may be compensated. Brain waves may be monitored and control signals transmitted to the brain from a remote station.

It is to be noted that the configuration described is one of many possibilities which may be formulated without departing from the spirit of my invention. The transmitters can be monostratic or bistatic. They also can be single, dual, or multiple frequency devices. The transmitted signal can be continuous wave, pulse, FM, or any combination of these as well as other transmission forms. Typical operating frequencies for the transmitters range from 1 MHz to 40 GHz but may be altered to suit the particular function being monitored and the characteristics of the specific subject.

The individual components of the system for monitoring and controlling brain wave activity may be of conventional type commonly employed in radar systems.

Various subassemblies of the brain wave monitoring and control apparatus may be added, substituted or combined. Thus, separate antennas or a single multimode antenna may be used for transmission and reception. Additional displays and computers may be added to present and analyze select components of the monitored brain waves.

Modulation of the interference signal retransmitted by the brain may be of amplitude, frequency and/or phase. Appropriate demodulators may be used to decipher the subject's brain activity and select components of his brain waves may be analyzed by computer to determine his mental state and monitor his thought processes.

As will be appreciated by those familiar with the art, apparatus and method of the subject invention has numerous uses. Persons in critical positions such as drivers and pilots can be continuously monitored with provision for activation of an emergency device in the event of human failure. Seizures, sleepiness and dreaming can be detected. Bodily functions such as pulse rate, heartbeat reqularity and others also can be monitored and occurrences of hallucinations can be detected. The system also permits medical diagnoses of patients, inaccessible to physicians, from remote stations.

What is claimed is:

1. Brain wave monitoring apparatus comprising
means for producing a base frequency signal,
means for producing a first signal having a frequency related to that of the base frequency and at a predetermined phase related thereto,
means for transmitting both said base frequency and said first signals to the brain of the subject being monitored,
means for receiving a second signal transmitted by the brain of the subject being monitored in response to both said base frequency and said first signals,
mixing means for producing from said base frequency signal and said received second signal a response signal having a frequency related to that of the base frequency, and
means for interpreting said response signal.

2. Apparatus as in claim **1** where said receiving means comprises
means for isolating the transmitted signals from the received second signals.

3. Apparatus as in claim **2** further comprising a band pass filter with an input connected to said isolating means and an output connected to said mixing means.

4. Apparatus as in claim **1** further comprising means for amplifying said response signal.

5. Apparatus as in claim **4** further comprising means for demodulating said amplified response signal.

6. Apparatus as in claim **5** further comprising interpreting means connected to the output of said demodulator means.

7. Apparatus according to claim **1** further comprising
means for producing an electromagnetic wave control signal dependent on said response signal, and
means for transmitting said control signal to the brain of said subject.

8. Apparatus as in claim **7** wherein said transmitting means comprises means for directing the electromagnetic wave control signal to a predetermined part of the brain.

9. A process for monitoring brain wave activity of a subject comprising the steps of
transmitting at least two electromagnetic energy signals of different frequencies to the brain of the subject being monitored,
receiving an electromagnetic energy signal resulting from the mixing of said two signals in the brain modulated by the brain wave activity and retrans-

mitted by the brain in response to said transmitted energy signals, and,

interpreting said received signal.

10. A process as in claim **9** further comprising the step of transmitting a further electromagnetic wave signal to the brain to vary the brain wave activity.

11. A process as in claim **10** wherein the step of transmitting the further signals comprises

obtaining a standard signal,

comparing said received electromagnetic energy signals with said standard signal,

producing a compensating signal corresponding to the comparison between said received electrogagnetic energy signals and the standard signal, and

transmitting the compensating signals to the brain of the subject being monitored.

* * * * *

N. TESLA.

APPARATUS FOR TRANSMITTING ELECTRICAL ENERGY.

APPLICATION FILED JAN. 18, 1902. RENEWED MAY 4, 1907.

1,119,732. Patented Dec. 1, 1914.

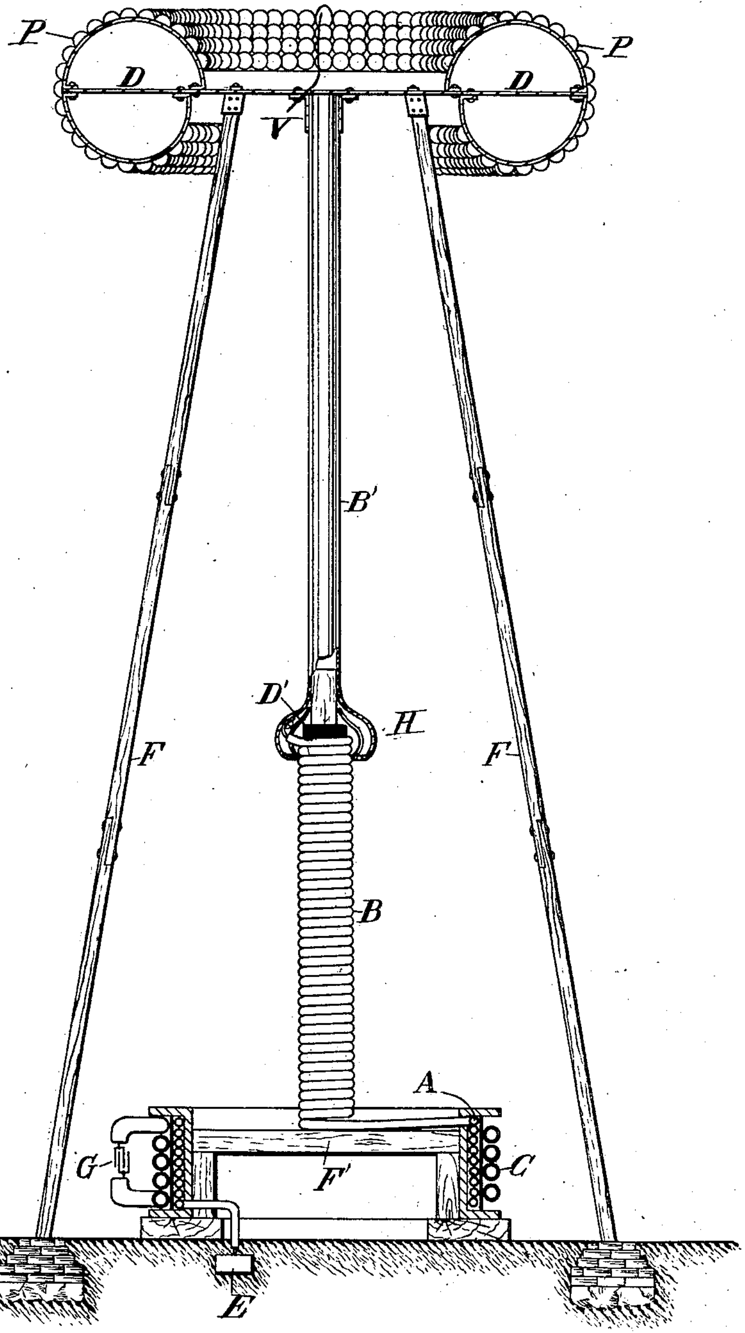

WITNESSES:

M. Lawson Dyer

Benjamin Miller.

Nikola Tesla, INVENTOR,

BY Kerr, Page & Cooper,

his ATTORNEYS.

UNITED STATES PATENT OFFICE.

NIKOLA TESLA, OF NEW YORK, N. Y.

APPARATUS FOR TRANSMITTING ELECTRICAL ENERGY.

1,119,732. Specification of Letters Patent. **Patented Dec. 1, 1914.**

Application filed January 18, 1902, Serial No. 90,245. Renewed May 4, 1907. Serial No. 371,817.

To all whom it may concern:

Be it known that I, Nikola Tesla, a citizen of the United States, residing in the borough of Manhattan, in the city, county, and State of New York, have invented certain new and useful Improvements in Apparatus for Transmitting Electrical Energy, of which the following is a specification, reference being had to the drawing accompanying and forming a part of the same.

In endeavoring to adapt currents or discharges of very high tension to various valuable uses, as the distribution of energy through wires from central plants to distant places of consumption, or the transmission of powerful disturbances to great distances, through the natural or non-artificial media, I have encountered difficulties in confining considerable amounts of electricity to the conductors and preventing its leakage over their supports, or its escape into the ambient air, which always takes place when the electric surface density reaches a certain value.

The intensity of the effect of a transmitting circuit with a free or elevated terminal is proportionate to the quantity of electricity displaced, which is determined by the product of the capacity of the circuit, the pressure, and the frequency of the currents employed. To produce an electrical movement of the required magnitude it is desirable to charge the terminal as highly as possible, for while a great quantity of electricity may also be displaced by a large capacity charged to low pressure, there are disadvantages met with in many cases when the former is made too large. The chief of these are due to the fact that an increase of the capacity entails a lowering of the frequency of the impulses or discharges and a diminution of the energy of vibration. This will be understood when it is borne in mind, that a circuit with a large capacity behaves as a slackspring, whereas one with a small capacity acts like a stiff spring, vibrating more vigorously. Therefore, in order to attain the highest possible frequency, which for certain purposes is advantageous and, apart from that, to develop the greatest energy in such a transmitting circuit, I employ a terminal of relatively small capacity, which I charge to as high a pressure as practicable. To accomplish this result I have found it imperative to so construct the elevated conductor, that its outer surface, on which the electrical charge chiefly accumulates, has itself a large radius of curvature, or is composed of separate elements which, irrespective of their own radius of curvature, are arranged in close proximity to each other and so, that the outside ideal surface enveloping them is of a large radius. Evidently, the smaller the radius of curvature the greater, for a given electric displacement, will be the surface-density and, consequently, the lower the limiting pressure to which the terminal may be charged without electricity escaping into the air. Such a terminal I secure to an insulating support entering more or less into its interior, and I likewise connect the circuit to it inside or, generally, at points where the electric density is small. This plan of constructing and supporting a highly charged conductor I have found to be of great practical importance, and it may be usefully applied in many ways.

Referring to the accompanying drawing, the figure is a view in elevation and part section of an improved free terminal and circuit of large surface with supporting structure and generating apparatus.

The terminal D consists of a suitably shaped metallic frame, in this case a ring of nearly circular cross section, which is covered with half spherical metal plates P P, thus constituting a very large conducting surface, smooth on all places where the electric charge principally accumulates. The frame is carried by a strong platform expressly provided for safety appliances, instruments of observation, etc., which in turn rests on insulating supports F F. These should penetrate far into the hollow space formed by the terminal, and if the electric density at the points where they are bolted to the frame is still considerable, they may be specially protected by conducting hoods as H.

A part of the improvements which form the subject of this specification, the transmitting circuit, in its general features, is identical with that described and claimed in my original Patents Nos. 645,576 and 649,621. The circuit comprises a coil A which is in close inductive relation with a primary C, and one end of which is connected to a ground-plate E, while its other end is led through a separate self-induction coil B and a metallic cylinder B′ to the terminal D.

The connection to the latter should always be made at, or near the center, in order to secure a symmetrical distribution of the current, as otherwise, when the frequency is very high and the flow of large volume, the performance of the apparatus might be impaired. The primary C may be excited in any desired manner, from a suitable source of currents G, which may be an alternator or condenser, the important requirement being that the resonant condition is established, that is to say, that the terminal D is charged to the maximum pressure developed in the circuit, as I have specified in my original patents before referred to. The adjustments should be made with particular care when the transmitter is one of great power, not only on account of economy, but also in order to avoid danger. I have shown that it is practicable to produce in a resonating circuit as E A B B′ D immense electrical activities, measured by tens and even hundreds of thousands of horse-power, and in such a case, if the points of maximum pressure should be shifted below the terminal D, along coil B, a ball of fire might break out and destroy the support F or anything else in the way. For the better appreciation of the nature of this danger it should be stated, that the destructive action may take place with inconceivable violence. This will cease to be surprising when it is borne in mind, that the entire energy accumulated in the excited circuit, instead of requiring, as under normal working conditions, one quarter of the period or more for its transformation from static to kinetic form, may spend itself in an incomparably smaller interval of time, at a rate of many millions of horse power. The accident is apt to occur when, the transmitting circuit being strongly excited, the impressed oscillations upon it are caused, in any manner more or less sudden, to be more rapid than the free oscillations. It is therefore advisable to begin the adjustments with feeble and somewhat slower impressed oscillations, strengthening and quickening them gradually, until the apparatus has been brought under perfect control. To increase the safety, I provide on a convenient place, preferably on terminal D, one or more elements or plates either of somewhat smaller radius of curvature or protruding more or less beyond the others (in which case they may be of larger radius of curvature) so that, should the pressure rise to a value, beyond which it is not desired to go, the powerful discharge may dart out there and lose itself harmlessly in the air. Such a plate, performing a function similar to that of a safety valve on a high pressure reservoir, is indicated at V.

Still further extending the principles underlying my invention, special reference is made to coil B and conductor B′. The latter is in the form of a cylinder with smooth or polished surface of a radius much larger than that of the half spherical elements P P, and widens out at the bottom into a hood H, which should be slotted to avoid loss by eddy currents and the purpose of which will be clear from the foregoing. The coil B is wound on a frame or drum D^1 of insulating material, with its turns close together. I have discovered that when so wound the effect of the small radius of curvature of the wire itself is overcome and the coil behaves as a conductor of large radius of curvature, corresponding to that of the drum. This feature is of considerable practical importance and is applicable not only in this special instance, but generally. For example, such plates at P P of terminal D, though preferably of large radius of curvature, need not be necessarily so, for provided only that the individual plates or elements of a high potential conductor or terminal are arranged in proximity to each other and with their outer boundaries along an ideal symmetrical enveloping surface of a large radius of curvature, the advantages of the invention will be more or less fully realized. The lower end of the coil B—which, if desired, may be extended up to the terminal D—should be somewhat below the uppermost turn of coil A. This, I find, lessens the tendency of the charge to break out from the wire connecting both and to pass along the support F′.

Having described my invention, I claim:

1. As a means for producing great electrical activities a resonant circuit having its outer conducting boundaries, which are charged to a high potential, arranged in surfaces of large radii of curvature so as to prevent leakage of the oscillating charge, substantially as set forth.

2. In apparatus for the transmission of electrical energy a circuit connected to ground and to an elevated terminal and having its outer conducting boundaries, which are subject to high tension, arranged in surfaces of large radii of curvature substantially as, and for the purpose described.

3. In a plant for the transmission of electrical energy without wires, in combination with a primary or exciting circuit a secondary connected to ground and to an elevated terminal and having its outer conducting boundaries, which are charged to a high potential, arranged in surfaces of large radii of curvature for the purpose of preventing leakage and loss of energy, substantially as set forth.

4. As a means for transmitting electrical energy to a distance through the natural media a grounded resonant circuit, comprising a part upon which oscillations are impressed and another for raising the ten-

sion, having its outer conducting boundaries on which a high tension charge accumulates arranged in surfaces of large radii of curvature, substantially as described.

5. The means for producing excessive electric potentials consisting of a primary exciting circuit and a resonant secondary having its outer conducting elements which are subject to high tension arranged in proximity to each other and in surfaces of large radii of curvature so as to prevent leakage of the charge and attendant lowering of potential, substantially as described.

6. A circuit comprising a part upon which oscillations are impressed and another part for raising the tension by resonance, the latter part being supported on places of low electric density and having its outermost conducting boundaries arranged in surfaces of large radii of curvature, as set forth.

7. In apparatus for the transmission of electrical energy without wires a grounded circuit the outer conducting elements of which have a great aggregate area and are arranged in surfaces of large radii of curvature so as to permit the storing of a high charge at a small electric density and prevent loss through leakage, substantially as described.

8. A wireless transmitter comprising in combination a source of oscillations as a condenser, a primary exciting circuit and a secondary grounded and elevated conductor the outer conducting boundaries of which are in proximity to each other and arranged in surfaces of large radii of curvature, substantially as described.

9. In apparatus for the transmission of electrical energy without wires an elevated conductor or antenna having its outer high potential conducting or capacity elements arranged in proximity to each other and in surfaces of large radii of curvature so as to overcome the effect of the small radius of curvature of the individual elements and leakage of the charge, as set forth.

10. A grounded resonant transmitting circuit having its outer conducting boundaries arranged in surfaces of large radii of curvature in combination with an elevated terminal of great surface supported at points of low electric density, substantially as described.

NIKOLA TESLA.

Witnesses:
M. Lamson Dyer,
Richard Donovan.

Nov. 12, 1957

L. R. CRUMP

POWERING ELECTRICAL DEVICES WITH ENERGY
ABSTRACTED FROM THE ATMOSPHERE

Filed March 12, 1954

2,813,242

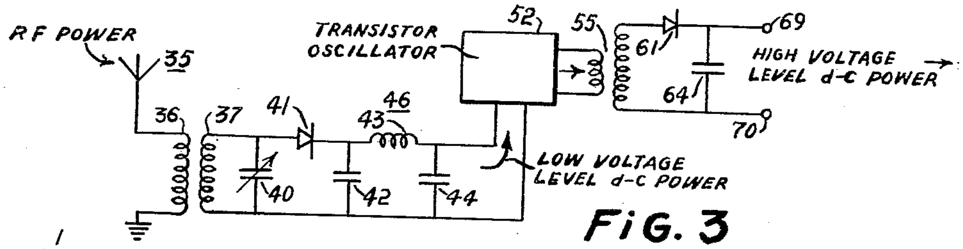
RF POWER
35
36
37
41
46
43
40
42
44
TRANSISTOR OSCILLATOR
52
55
61
64
69
70
HIGH VOLTAGE LEVEL d-c POWER
LOW VOLTAGE LEVEL d-c POWER

FiG. 3

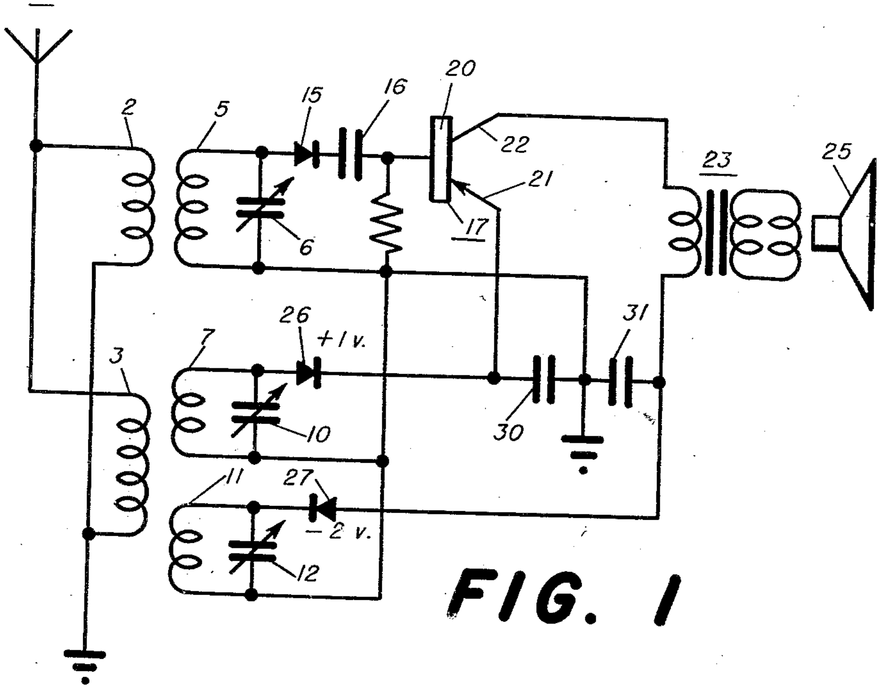
1
2
5
15
16
20
22
21
17
6
23
25
26
+1 v.
7
3
10
31
30
11
27
−2 v.
12

FIG. 1

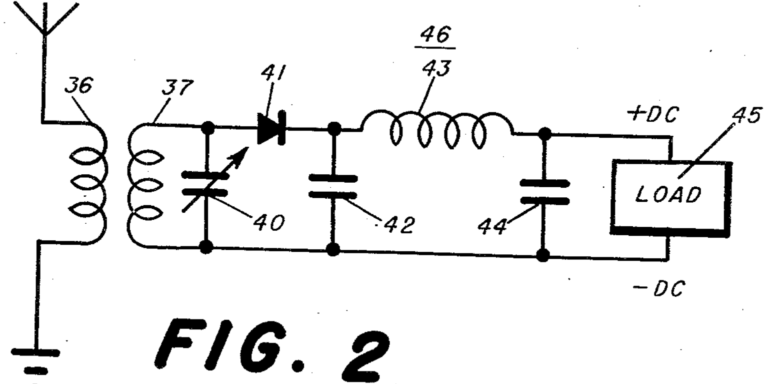
35
36
37
41
46
43
+DC
45
LOAD
40
42
44
−DC

FIG. 2

INVENTOR

Lloyd R. Crump

BY W.E. Thibodeau + A.W. Dew

ATTORNEYS

United States Patent Office

2,813,242

Patented Nov. 12, 1957

2,813,242

POWERING ELECTRICAL DEVICES WITH ENERGY ABSTRACTED FROM THE ATMOSPHERE

Lloyd R. Crump, Silver Spring, Md.

Application March 12, 1954, Serial No. 415,986

1 Claim. (Cl. 321—2)

(Granted under Title 35, U. S. Code (1952), sec. 266)

The invention described herein may be manufactured and used by or for the Government for governmental purposes without payment to me of any royalty thereon.

This invention relates to the convenient and economical provision of power for the operation of electronic circuits and devices using transistors, and of other electrical devices having modest power requirements.

A great advantage of transistors, and a major reason for their enthusiastic reception since their introduction a few years ago, is the fact that they will operate satisfactorily with very low supply voltages and currents. One milliwatt or even less is sufficient to power a transistor in many applications. Various batteries have been developed to provide, in a minimum of space, the relatively minute amounts of power needed by transistors.

My invention provides methods and means that permit transistor circuits, and also other low-powered electrical devices, to be economically and conveniently operated without any batteries whatever, and indeed without any power supply whatever as power supplies are ordinarily conceived.

The invention centers around my discovery that it is practicable to construct operative transistor circuits that are able to abstract from the atmosphere sufficient electromagnetic energy to provide all necessary supply voltages and currents for their own operation. Circuits and devices powered according to my invention will operate indefinitely without any local power source whatever.

I have successfully constructed and demonstrated such circuits. For example, I have constructed a batteryless transistor radio receiver on which I have listened to either nearby or distant broadcast stations as desired, using either headphones or a loudspeaker; this receiver has been powered entirely by electromagnetic energy abstracted from the atmosphere.

From the successful operation of this receiver, and from other experimental work, it becomes clear that, by the methods and means of the invention, a great variety of practical and useful transistor circuits can be powered entirely by energy abstracted from the atmosphere.

Furthermore, as will become apparent below, my invention is applicable to the powering of other electrical devices requiring relatively small amounts of power.

An object of the present invention is to provide methods and means for powering transistor circuits entirely from radiofrequency energy abstracted from the atmosphere.

Another object is to provide methods and means for powering remote radio receivers, low-powered radio transmitters, and other low-powered electrical devices, with energy received by radio from a master station, so that no local power supplies are needed by the devices and so that the powering or non-powering of the remote device is under the control of the master station.

A further object is to provide methods and means for powering transistor circuits and other low-powered electrical devices with radiofrequency energy received from one or more remote radio transmitters.

Other objects, aspects, uses, and advantages of the invention will become apparent from the following description and from the drawing.

Figure 1 is a schematic diagram of a transistor radio receiver in which all necessary power is supplied by energy abstracted from the atmosphere in accordance with the invention.

Figure 2 is a schematic diagram showing a general application of the invention to provide direct-current power to a load.

Figure 3 is a schematic diagram of a system for obtaining a high energy D.-C. source at a high voltage level using energy abstracted from the atmosphere.

Referring to Figure 1, a receiving antenna **1** is connected to antenna coupling coils **2** and **3**, the other ends of which are connected to ground. A parallel resonant circuit consisting of coil **5** and capacitor **6** is coupled to coil **2**. A second parallel resonant circuit consisting of coil **7** and capacitor **10** is coupled to coil **3**. A third parallel resonant circuit consisting of coil **11** and capacitor **12** is also coupled to coil **3**.

Coil **5** and capacitor **6** are tuned to the frequency of a radio transmitter from which it is desired to receive information—for instance, an amplitude-modulated standard broadcast station. The signal received from this transmitter need not be strong. The signal is detected by diode **15** to obtain an audio-frequency information signal. This audio signal is coupled through a capacitor **16** and is amplified by a circuit that includes a transistor **17** having a base **20**, an emitter **21**, and a collector **22**. The amplified audio output of the transistor is coupled through an audio transformer **23** to an electroacoustical transducer, preferably a permanent-magnet dynamic loudspeaker **25** as shown.

The novelty of the invention lies largely in the method and means by which the necessary direct-current power is supplied to the emitter and collector circuits of transistor **17**. This method and means will now be described.

Coil **7** and capacitor **10**, and also coil **11** and capacitor **12**, are tuned to receive radio signals of relatively high strength. It does not matter whether these signals contain information. These power signals are rectified by diodes **26** and **27** to provide direct-current power that is filtered by capacitors **30** and **31**. The D.-C. power thus obtained is utilized to power the transistor **17**.

In the circuit shown, two tuned circuits (coil **7** and capacitor **10**, and coil **11** and capacitor **12**) are tuned to power signals and the D.-C. voltages obtained from each are connected in series. The tuned power circuits may be tuned to the same or different power signals. Under certain circumstances it may be desirable to use more than two tuned power circuits and to tune them to more than two power signals; in this way power can be obtained from several signals and combined. On the other hand, if a strong power signal is available, a single tuned power circuit may suffice to give the needed D.-C. power.

Even weak information signals can be received successfully. A plurality of transistor amplifier stages can be used if desired, or other circuits such as superheterodyne circuits can be used. It is merely necessary that a sufficiently strong power signal or signals be available to provide the small amount of D.-C. needed to power the transistors.

If the information signal happens to be strong, it can be used as the power signal; all of the tuned circuits (coil **5** and capacitor **6**, coil **7** and capacitor **10**, coil **11** and capacitor **12**) are tuned to the information signal.

Engineers who have observed my invention in operation have been surprised at the unexpectedly good results obtained, even with readily available power signals of quite moderate strength. For instance, sufficient power for satisfactory operation of a loudspeaker at low volumes is

readily obtained from a 5-kilowatt standard broadcast station 5 miles away, using only an indoor antenna to pick up the power signal as well as information signals. In typical operation under these conditions a D.-C. voltage of about 2.5 to 3 volts is obtained between the emitter and the collector, at a current of about 250 microamperes; D.-C. power input to the transistor is thus of the order of 0.5 to 1 milliwatt. So far as I am aware, no one has ever before discovered and demonstrated the practicability of this method of powering a radio receiver.

Because existing broadcast stations within a radius of a number of miles provide adequate power signals, the invention is readily practicable with existing power signals in almost any location in or near any city in the United States.

Although I have described a transistor radio receiver powered by my invention, it will be readily apparent that the invention is applicable to the powering of any transistor circuit using one or a number of transistors, and to the powering of other devices requiring relatively small amounts of power. For instance, sensitive electromechanical, electrochemical, or electrothermal devices can be operated by the method of the invention.

Referring to Figure 2, which shows a more general embodiment of my invention, an antenna **35** picks up radiofrequency energy from the atmosphere. This energy flows through coil **36**, which is coupled to a tuned circuit consisting of coil **37** and capacitor **40**. The radiofrequency voltage across capacitor **40** is rectified by diode **41** and filtered by a low-pass filter **46** consisting of capacitors **42** and **44** and choke coil **43**. The resulting D.-C. voltage is applied to a load **45**.

In the practice of my invention, larger amounts of power can be obtained for short periods of time by storing received energy in a suitable energy storage device. Stored energy may then be withdrawn at intervals at a more rapid rate than that at which it was received and put into the storage device. In this way the invention can be used to provide short pulses of relatively very high electrical energy. This result can be readily obtained by charging a relatively large capacitor with direct current and then discharging the capacitor rapidly into a load when desired. This rapid discharge can be initiated automatically when the voltage across the capacitor reaches a certain level, or it can be initiated when a transistor radio receiver receives a certain information signal.

Higher voltages can be obtained with the invention by means of well known devices for raising D.-C. voltages as shown in Figure 3. The D.-C. voltage output from the capacitor **44** can be used to power a low frequency transistor oscillator **52** whose A.-C. output is raised to a higher voltage level by the transformer **55**. This relatively high A.-C. voltage can then be rectified by a diode **61** and fed to a capacitor **64** to provide a high energy D.-C. source at a relatively high voltage level at the terminals **69** and **70**. If desired, energy can now be withdrawn from the capacitor **64** at intervals in short pulses of high energy at a high voltage level. Pulsed radio transmission is one of the possible uses for this form of the invention. Other uses would be to provide a single relatively powerful pulse needed to actuate an electrothermal or electromechanical device.

As has been indicated above, in many locations and particularly anywhere in or near most American cities, power signals normally present in the atmosphere are readily available for the easy and convenient practice of the invention. However, the invention also has important applications in systems in which the necessary power signal is generated and transmitted specifically for the operation of the particular system. Such systems can, for example, comprise a master station transmitting all the power that is needed for hundreds or thousands of fixed or mobile transistor receivers or other remote devices over

a range of many miles. This eliminates the need for hundreds or thousands, as the case may be, of local power supplies. At the same time, such a system has the advantage that all of the remote devices can be simultaneously activated or deactivated at the will of the master station, simply by starting or stopping the transmission of the power signal. In such systems it will often be advantageous to use power signals of frequencies sufficiently high to permit the use of resonant receiving antennas of small physical dimensions for signal pickup at the remote devices. In addition to the power signal, the master station may transmit an information signal on the same or a different carrier.

Certain types of devices powered entirely by received radio waves are of course well known. The well-known "crystal set" of the early days of radio, which used a diode rectifier to demodulate an amplitude-modulated radiofrequency signal, is an outstanding example of such a device. My invention is readily distinguishable from such prior devices, however. In typical prior devices a modulated radiofrequency signal is applied to a diode to obtain unidirectional half-wave pulses whose amplitudes vary with modulation. These pulses are integrated by means of a capacitor to obtain a unidirectional signal the amplitude of which follows the audiofrequency modulation envelope. If the radiofrequency signal is received with sufficient strength the audio signal may have sufficient power to operate headphones or similar utilization device without power amplification; but the signal is utilized for its information content, rather than to supply non-information-containing power.

My invention, on the other hand, entails the utilization of received radiofrequency energy to supply power to at least one pair of circuit points (across capacitor **31** in Fig. 1, for example), such circuit points requiring power solely for its power content and not for any information or modulation it may contain. In other words, my invention entails the utilization of radiofrequency energy to supply power that would otherwise have to be supplied by batteries, generator, or other local power source.

It will be apparent that the embodiments shown are only exemplary and that various modifications can be made in construction and arrangement within the scope of the invention as defined in the appended claim.

I claim:

An electrical device for obtaining a high energy D.-C. source at a high voltage level using energy abstracted from the atmosphere, said device comprising in combination: resonant means for receiving radio waves, first rectifier means for converting said radio waves into first direct current energy, first capacitor means for storing said first direct current energy, an oscillator powered by said direct current energy, said oscillator producing an A.-C. output, transformer means for raising said A.-C. output to an increased voltage level, second rectifier means for converting the A.-C. output of increased voltage level from said transformer into second direct current energy, and second capacitor means for storing said second direct current energy.

References Cited in the file of this patent

UNITED STATES PATENTS

1,540,998	Plauson	June 9, 1925
2,056,824	Cawley	Oct. 6, 1936
2,553,734	Adler	May 22, 1951
2,569,687	Pan	Oct. 2, 1951
2,647,957	Mallinckrodt	Aug. 4, 1953

OTHER REFERENCES

Turner: "A crystal receiver with transistor amplifier," Radio and Television News, January 1950.

Dec. 14, 1971 H. W. WALLACE 3,626,605

METHOD AND APPARATUS FOR GENERATING A SECONDARY
GRAVITATIONAL FORCE FIELD

Filed Nov. 4, 1968 6 Sheets-Sheet 1

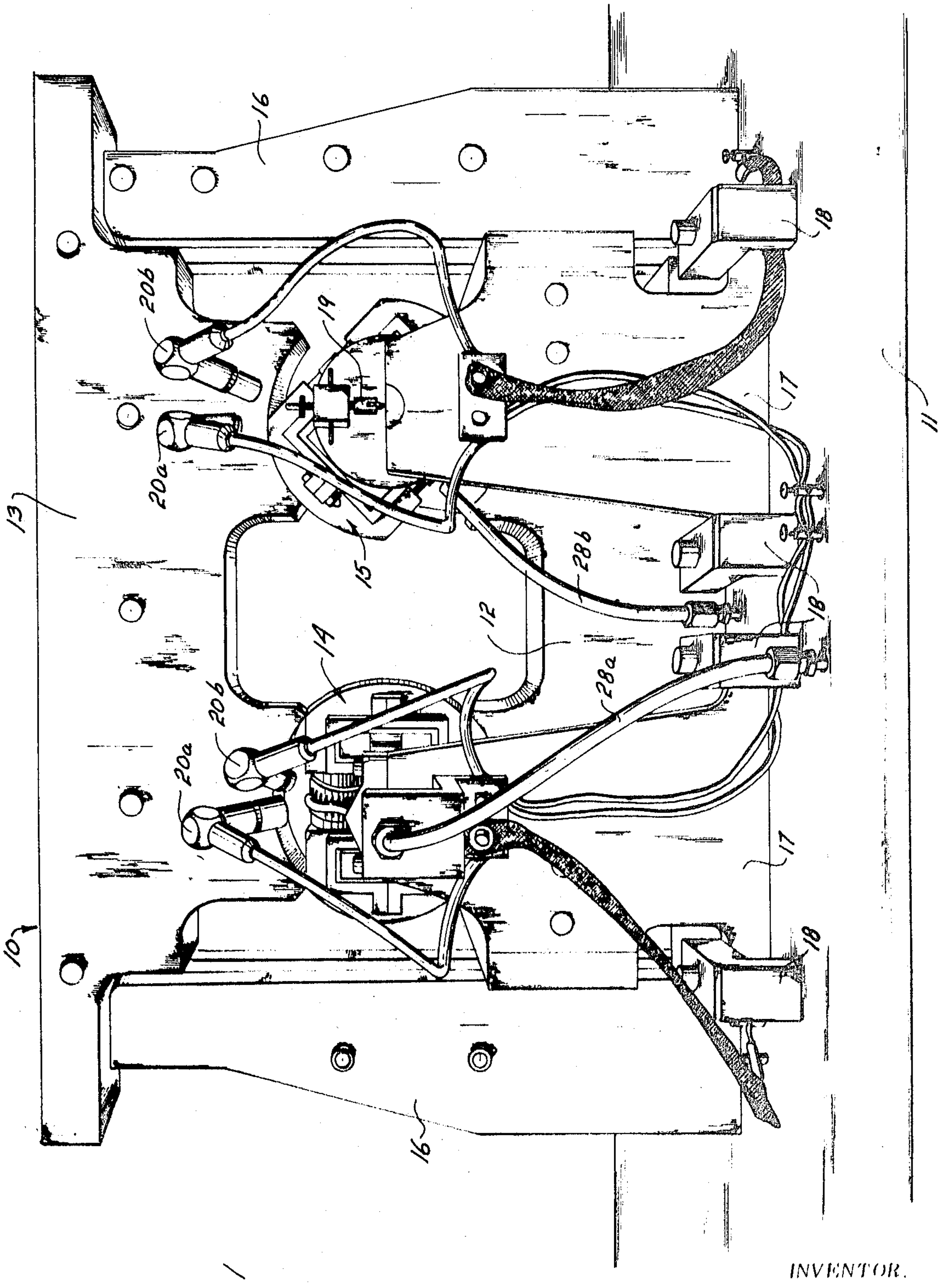
16
18
11
17
19
20b
20a
13
15
28b
12
14
18
28a
20b
20a
17
10
18
16

FIG-1

INVENTOR.
HENRY W. WALLACE
BY
Synnestvedt & Lechner
ATTORNEY

Dec. 14, 1971 H. W. WALLACE **3,626,605**

METHOD AND APPARATUS FOR GENERATING A SECONDARY GRAVITATIONAL FORCE FIELD

Filed Nov. 4, 1968 6 Sheets-Sheet 3

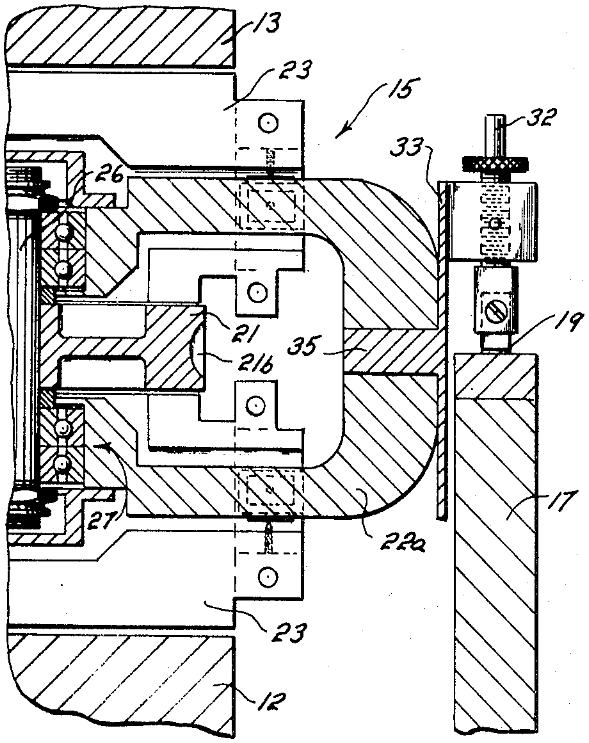
13
23
15
32
33
26
21
21b
35
19
17
27
22a
23
12

FIG-3A

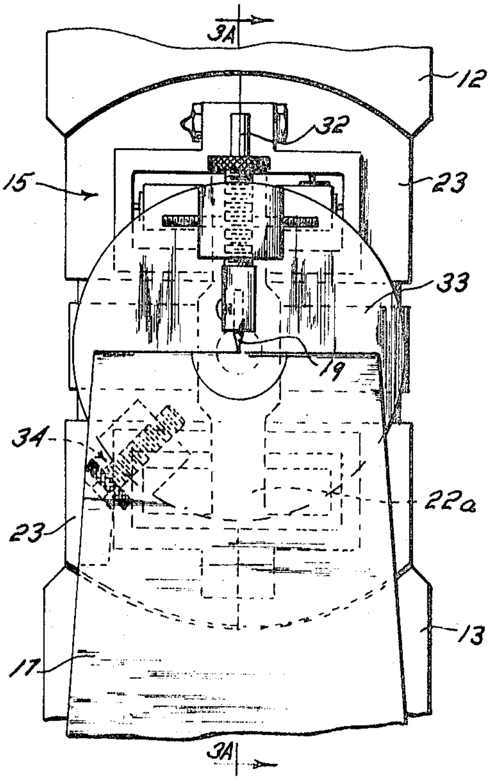
3A
12
32
15
23
33
19
34
22a
23
13
17
3A

FIG-3B

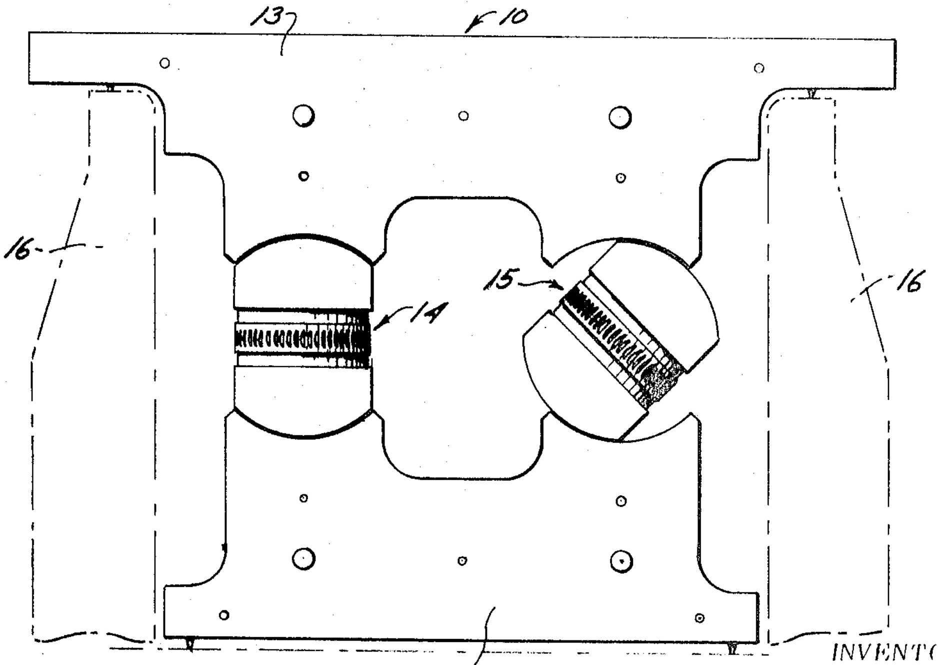
13
10
16
15
14
16
12

FIG-2

INVENTOR.
HENERY W. WALLACE
BY
Synnestvedt & Lechner
ATTORNEYS

Dec. 14, 1971 H. W. WALLACE 3,626,605

METHOD AND APPARATUS FOR GENERATING A SECONDARY GRAVITATIONAL FORCE FIELD

Filed Nov. 4, 1968 6 Sheets-Sheet 3

FIG-3

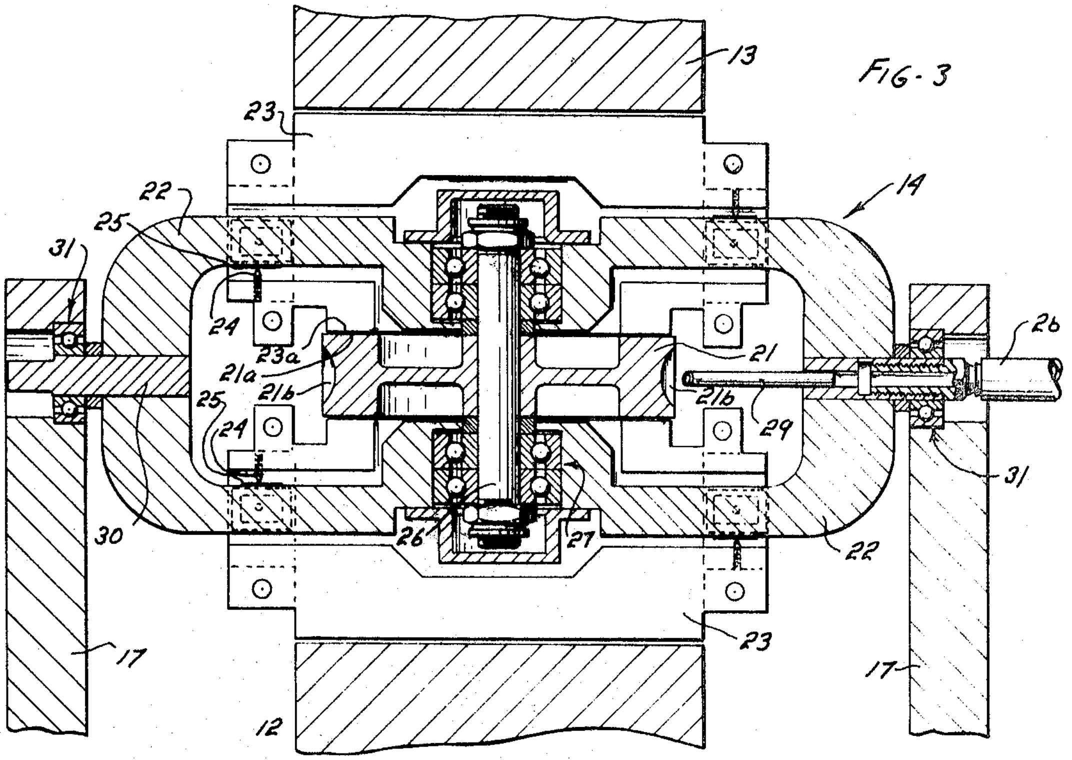
13
23
14
22
25
31
24
23a
21a
21b
21
26
21b
29
25
24
31
30
26
27
22
23
17
17
12

FIG-4

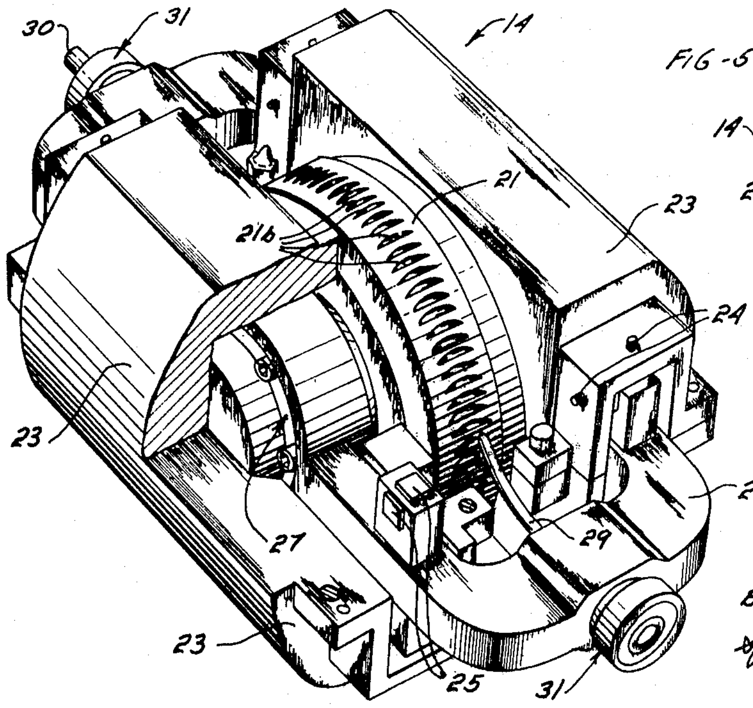
30
31
14
21
23
21b
24
23
22
27
29
23
25
31

FIG-5

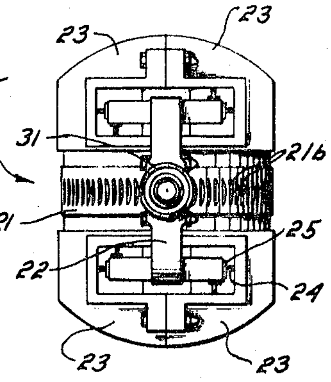
23
23
14
31
21b
21
25
22
24
23
23

INVENTOR.
HENRY W. WALLACE
By
Symes Duede & Lechner
ATTORNEYS

METHOD AND APPARATUS FOR GENERATING A SECONDARY GRAVITATIONAL FORCE FIELD

Filed Nov. 4, 1968

FIG-6

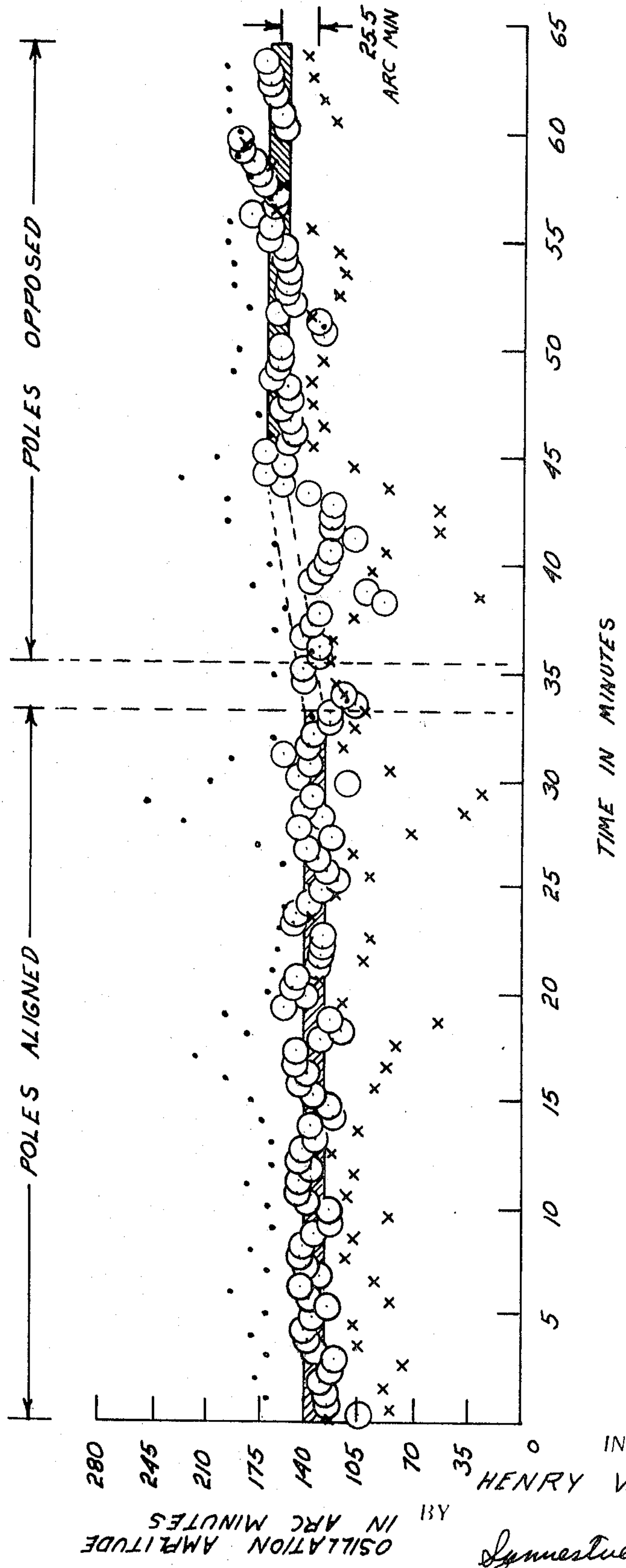

INVENTOR.
HENRY W. WALLACE

BY Synnestvedt & Lechner

ATTORNEYS

FIG-7

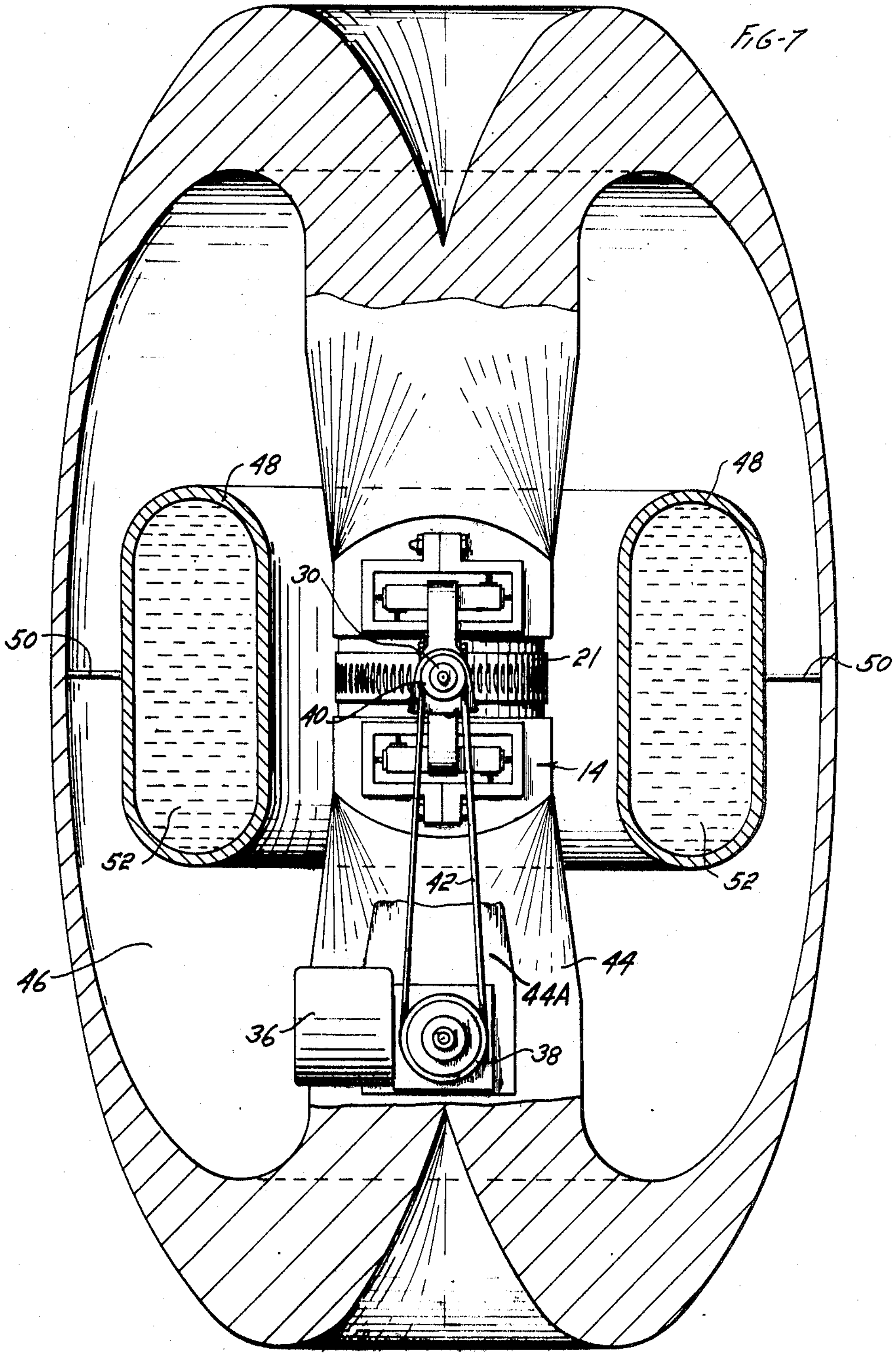
48
48
30
50
21
50
40
14
52
52
42
44
46
44A
36
38

INVENTOR.
HENRY W. WALLACE
BY
Synnestvedt & Lechner
ATTORNEYS

Dec. 14, 1971 H. W. WALLACE 3,626,605

METHOD AND APPARATUS FOR GENERATING A SECONDARY GRAVITATIONAL FORCE FIELD

Filed Nov. 4, 1968 6 Sheets-Sheet 6

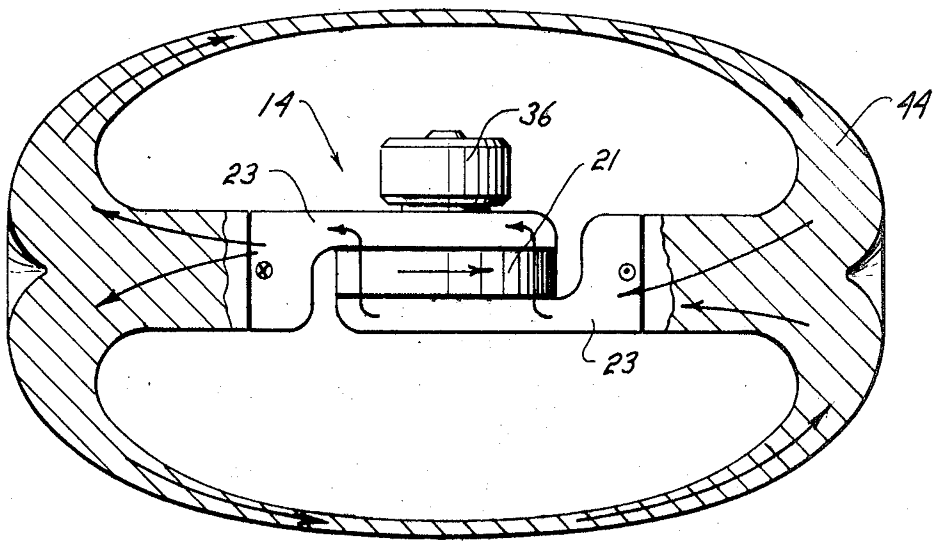

14
36
44
23
21
23

FIG-7A

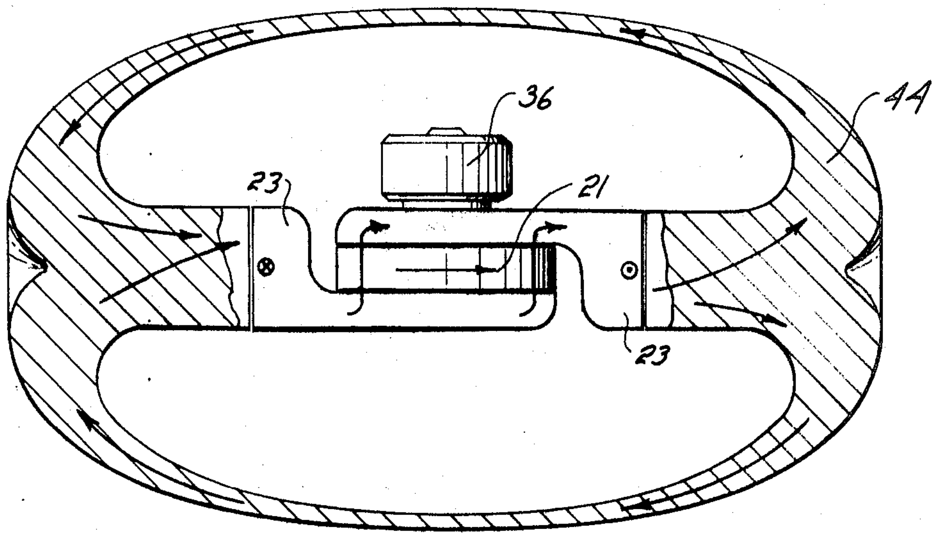

36
44
23
21
23

FIG-7B

INVENTOR.
HENRY W. WALLACE
BY
Synnestvedt & Lechner
ATTORNEY

United States Patent Office

3,626,605
Patented Dec. 14, 1971

3,626,605

METHOD AND APPARATUS FOR GENERATING A SECONDARY GRAVITATIONAL FORCE FIELD

Henry Wm. Wallace, Ardmore, Pa.
(803 Cherry Lane, Laurel, Miss. 39440)

Filed Nov. 4, 1968, Ser. No. 773,051
Int. Cl. G09b 23/06
U.S. Cl. 35—19 10 Claims

ABSTRACT OF THE DISCLOSURE

Apparatus and method for generating a time variant non-electromagnetic force field due to the dynamic interaction of relatively moving bodies and for transforming such force fields into energy for doing useful work.

The method of generating such time variant force fields including the steps of juxtaposing in field series relationship a stationary member of spin nuclei material, and a member capable of assuming relative motion with respect to said stationary member also characterized by spin nuclei material; initiating relative motion by rotation of said one member with respect to the other, which rotational motion may occur both about an axis within the plane of said other member and about an axis perpendicular thereto; whereby the rotational motion of said one member about the axis perpendicular to the plane of said other member results in the dual interaction of the angular momentum property of said one member with inertial space and also with respect to the angular momentum property of said other member thereby resulting in a dynamic interaction field arising through gravitational coupling which dynamic interaction field is further characterized by its nonelectromagnetic nature and its mass-proximity and relative motion dependency; the rotation of said one member about the axis within the plane of said other member further resulting in an undulation of the dynamic interaction field within said other member which in turn gives rise to a secondary time-variant gravitational field in the surrounding space.

The present invention relates to an apparatus and method for generating a time-variant force field due to the relative motion of moving bodies; which force field exhibits itself in the form of an induced secondary gravitational force. As such, this invention constitutes a continuation-in-part of an application filed Nov. 4, 1968 by the same inventor, entitled Method and Apparatus for Generating a Dynamic Force Field and bearing Ser. No. 773,116.

In the practice of the present invention it has been found that when bodies composed of certain material are placed in relative motion with respect to one another there is generated an energy field therein not heretofore observed. This field is not electromagnetic in nature; being by theoretical prediction related to the gravitational coupling of relatively moving bodies.

The initial evidence indicates that this nonelectromagnetic field is generated as a result of the relative motion of bodies constituted of elements whose nuclei are characterized by half integral "spin" values, the spin of the nuclei being associated with the net angular momentum of the nucleons thereof. The nucleons in turn comprise the elemental particles of the nucleus; i.e., the neutrons and protons. For purposes of the present invention the field, generated by the relative motion of materials characterized by a half integral spin value, is referred to as a "kinemassic" force field.

It will be appreciated that relative motion occurs on various levels, i.e., there may be relative motion of discrete bodies as well as of the constituents thereof including, on a subatomic level, the nucleons of the nucleus. The kinemassic force field under consideration is a result of such relative motion, being a function of the dynamic interaction of two relatively moving bodies including the elemental particles thereof. The value of the kinemassic force field created, by reason of the dynamic interaction of the bodies experiencing relative motion, is the algebraic sum of the fields created by reason of the dynamic interaction of both elementary particles and of the discrete bodies.

For a closed system comprising only a stationary body, the kinemassic force, due to the dynamic interaction of the subatomic particles therein, is zero because of the random distribution of spin orientations of the respective particles. Polarization of the spin components so as to align a majority thereof in a preferred direction establishes a flux field aligned with the spin axes of the elementary particles. The present invention is in part comprised of an apparatus for polarizing its spin nuclei material, while additional means are provided to induce an alternating or undulating effect in the kinemassic field so generated.

Accordingly, a primary object of the present invention concerns the provision of means for generating a time-variant kinemassic field within a permeable field body due to the dynamic interaction of relatively moving bodies and the relative rotation of said generating means with respect to the permeable field body.

The kinemassic force field finds theoretical support in the laws of physics, being substantiated by the generalized theory of relativity. According to the general theory of relativity three exists not only a static gravitational field but also a dynamic component thereof due to the gravitational coupling of relatively moving bodies.

This theory proposes that two spinning bodies will exert force on each other. Heretofore the theoretical predictions have never been experimentally substantiated; however, as early as 1896, experiments were conducted in an effort to detect predicted centrifugal forces on stationary bodies placed near large, rapidly rotating masses. The results of these early experiments were inconclusive, and little else in the nature of this type of work is known to have been conducted.

It is therefore another object of the present invention to set forth an operative technique for generating a measurable time-variant force field due to the gravitational coupling of relatively moving bodies.

In carrying out the present invention, means are provided to enable the relative rotational motion of a first member with respect to a stationary member positioned in close proximity thereto; the construction of one embodiment of the first member being such as to enable portions thereof to assume rotational motion about an axis perpendicular to the plane of said stationary member. The effect of the rotation of said member about the axis perpendicular to the plane of the stationary member is to establish a kinemassic force field in the nature of that referred to in the aforementioned copending application of the same inventor. The rotation of said member about the axis lying within the plane of said stationary member results in an undulation of the dynamic interaction field within said field members which in turn induces a secondary time-variant gravitational field in the surrounding space.

Accordingly, another more specific object of the present invention concerns a method of generating a nonelectromagnetic force field due to the dynamic interaction of relatively moving bodies and for utilizing such force field to further generate a secondary gravitational field.

The foregoing objects and features of novelty which characterize the present invention, as well as other objects of the invention, are pointed out with particularity in the claims annexed to and forming a part of the present specification. For a better understanding of the invention, its advantages and specific objects allied with its use, reference should be made to the accompanying drawings and descriptive matter in which there is illustrated and described a preferred embodiment of the invention.

In the drawings:

FIG. 1 shows an overall view of equipment constructed in accordance with the principles of the present invention, including means to demonstrate the effect of a time-variant kinemassic force field;

FIG. 2 is an isolation schematic of apparatus components comprising the time-variant kinemassic field circuit of FIG. 1;

FIGS. 3, 3A, 3B, 4 and 5 show the details of construction of the generator and detector assemblies of FIGS. 1 and 2;

FIG. 6 represents measured changes in operating characteristics of the apparatus of FIGS. 1 and 2 demonstrating the time-variant nature of the kinemassic force field so produced; and

FIGS. 7, 7A and 7B are sectioned views of various embodiments of the present invention for demonstrating the time-variant nature of the kinemassic force field as used in establishing a secondary gravitational component.

Before getting into a detailed discussion of the apparatus and the steps involved in the practice of the present invention, it should be helpful to an understanding if the present invention of consideration is first given to certain defining characteristics thereof, many of which bear an analogous relationship to electromagnetic field theory.

A first feature is that the kinemassic field is vectorial in nature. The direction of the field vector is a function of the geometry in which the relative motion between mass particles takes place.

The second significant property of the kinemassic field relates the field strength to the nature of the material in the field. This property may be thought of as the kinemassic permeability by analogy to the concept of permeability in magnetic field theory. The field strength is apparently a function of the density of the spin nuclei material comprising the field circuit members. Whereas the permeability in magnetic field theory is a function of the density of unpaired electrons, the kinemassic permeability is a function of the density of spin nuclei and the measure of magnitude of their half integral spin values. As a consequence of this latter property, the field may be directed and confined by interposing into it denser portions of desired configuration. For example, the field may be in large measure confined to a closed loop of dense material starting and terminating adjacent a system wherein relative motion between masses is occurring.

A further property of the kinemassic force field relates field strength to the relative spacing between two masses in relative motion with respect to one another. Thus, the strength of the resultant field is a function of the proximity of the relatively moving bodies such that relative motion occurring between two masses which are closely adjacent will result in the generation of a field stronger than that created when the same two relatively moving bodies are spaced farther apart.

As mentioned above, a material consideration in generating the kinemassic force field concerns the use of spin nuclei material. By spin nuclei material is meant materials in nature which exhibit a nuclear external angular momentum. This includes both the intrinsic spin of the unpaired nucleon as well as that due to the orbital motion of these nucleons.

Since the dynamic interaction field arising through gravitational coupling is a function of both the mass and proximity of two relatively moving bodies, then the resultant force field is predictably maximized within the nucleus of an atom due to the relatively high densities of the nucleons plus the fact that the nucleons possess both intrinsic and orbital components of angular momentum. Such force fields may in fact account for a significant portion of the nuclear binding force found in all of nature.

It has been found that for certain materials, namely those characterized by a half integral spin value, the external component of angular momentum thereof will be accompanied by a force due to the dynamic interaction of the nucleons.

This is the so-called kinemassic force which on a submacroscopic basis exhibits itself as a field dipole moment aligned with the external angular momentum vector. These moments are of sufficient magnitude that they interact with adjacent, or near adjacent, spin nuclei field dipole moments of neighboring atoms.

This latter feature gives rise to a further analogy to electromagnetic field theory in that the interaction of adjacent spin nuclei field dipole moments gives rise to nuclear domain-like structures within matter containing a sufficient portion of spin nuclei material.

Although certain analogies exist between the kinemassic force field and electromagnetic field theory, it should be remembered that the kinemassic force is essentially non-responsive to or affected by electromagnetic force phenomena. This latter condition further substantiates the ability of the kinemassic field to penetrate through and extend outward beyond the ambient electromagnetic field established by the moving electrons in the atomic structure surrounding the respective spin nuclei.

As in electromagnetic field theory, in an unpolarized sample the external components of angular momentum of the nuclei to be subjected to a kinemassic force field are originally randomly oriented such that the material exhibits no residual kinemassic field of its own. However, establishing the necessary criteria for such a force field effects a polarization of the spin components of adjacent nuclei in a preferred direction thereby resulting in a force field which may be represented in terms of kinemassic field flux lines normal to the direction of spin.

The fact that spin nuclei material exhibits external kinemassic forces suggests that these forces should exhibit themselves on a macroscopic basis and thus be detectable, when arranged in a manner similar to that for demonstrating the Barnett effect when dealing with electromagnetic phenomena.

In the Barnett effect a long iron cylinder, when rotated at high speed about its longitudinal axis, was found to develop a measurable component of magnetization, the value of which was found to be proportional to the angular speed. The effect was attributed to the influence of the impressed rotation upon the revolving electronic systems due to the mass property of the unpaired electrons within the atoms.

In an apparatus constructed in accordance with the foregoing principles it was found that a rotating member such as a wheel composed of spin nuclei material exhibits a kinemassic force field. The interaction of the spin nuclei angular momentum with inertial space causes the spin nuclei axes of the respective nuclei of the material being spun to tend to reorient parallel with the axis of the rotating member. This results in the nuclear polarization of the spin nuclei material. With sufficient polarization, an appreciable field of summed dipole moments emanates from the wheel rim flange surfaces to form a secondary dynamic interaction with the dipole moments of spin nuclei contained within the facing surface of a stationary body positioned immediately adjacent the rotating member.

When the stationary body, composed of suitable spin nuclei material, is connected in spatial series with the rotating member, a circuitous form of kinemassic field is created; the flux of which is primarily restricted to the field circuit.

If now means are provided to periodically reverse the direction of rotation of the wheel with respect to the facing surfaces of the stationary body positioned in immediate proximity thereto, then the resultant time-varying kinemassic field generates or induces an accompanying time-varying secondary gravitational field in the space immediately surrounding. That is to say, if the time-varying kinemassic field is made to undulate typically sinusoidally, there will be induced an accompanying undulating secondary gravitational field which is phase-related to the kinemassic field. In this repsect the induced secondary gravitational field is generated in a manner analogous to electromagnetic induction theory.

By properly configuring the undulating kinemassic field, the resultant secondary gravitational field may be essentially restrained to or confined within an enclosed space. Although numerous specific geometric configurations may be proposed, the necessary conditions are established in the preferred embodiment of the present invention by enclosing the kinemassic field generating apparatus, including the rotational members and at least a portion of the stationary member thereof, within an enclosure, the material portions of which obey the rules concerning kniemassic permeability.

The kinemassic field when so configured, will shield the enclosed space both with respect to the secondary gravitational field induced therein and with respect to the ambient gravitational field caused by the earth and other cosmic bodies, existing externally of the shielded space. The undulating kinemassic field, which gives rise to the enclosed secondary undulating gravitational field, is effective in reducing the quantity of flux lines within the space surrounded by the undulating kinemassic field contributed by the ambient gravitational field, thus reducing the mutual force of gravitational attraction acting between this structure and the earth or other cosmic bodies dependent upon their relative contribution to the local gravitational flux density.

It is well known that nature opposes heterogeneous field flux densities. If the normal local flux density contributed by the earth and the other cosmic bodies within the space occupied by and surrounded by the undulating kinemassic field permeable structure were added to by the forcibly enclosed flux of the induced secondary undulating gravitational field, this increased flux density would be in opposition to nature. Although the induced secondary undulating gravitational field would cause an undulating variation of the cosmic or primary gravitational field flux lines of force to penetrate within the kinemassic field permeable structure, if this undulation were sinusoidal, for example, the RMS or 0.707 value of peak reduction in mutual gravitational attraction would apply.

Means for increasing the relative magnitude of the undulating kinemassic field is effected by positioning a mass circuit within the induced secondary field space. The mass circuit in linking with the undulating kinemassic field circuit results in an increase in the kinemassic field and in the same sense effectively intensifies the primary gravitational field shielding. A partial parallel can be found in electromagnetic field theory, where it is known that a shorted secondary turn enhances iron saturation.

The mass circuit located in the induced secondary field space need not be comprised of material having a spin nuclei characteristic; rather, it is more important that this mass circuit have high mass density. A further desirable characteristic of this mass circuit is that it have a capability for mass flow with respect to the undulating kinemassic field structure. Mercury has the desired combination of properties and while other materials may be used, mercury is the most effective thus far known.

As indicated above, the effective result of generating a secondary undulating gravitational field within the space enclosed by the undulating kinemassic field is a relative reduction in apparent weight of the kinemassic field permeable structure, with respect to its apparent weight without such an undulating kinemassic field. The explanation of this phenomenon may be readily conceived as caused by the generation of a field force vector antiparallel in direction to the local gravitational field force vector. If the shielding is sufficiently effective to reduce the density of gravitational field flux lines within the shielded space to the equivalent of the ambient flux line density, there will be no net local distortion of the gravitational field flux line pattern in the space occupied by the kinemassic field permeable material or the space enclosed by the kinemassic field configuration. Without distortion of the local flux line pattern the two bodies cease to mutually attract and, in effect, become weightless, one with respect to the other.

Although similar in result, the technique for effecting the state of weightlessness in the present invention differs from conventional apparatus for achieving such a state of weightlessness. The latter, in general, utilize the force of radial acceleration to effectively "balance" the gravitational forces acting on a body.

The relative magnitude of the undulating kinemassic force field and the kinemassic permeability characterizing the associated structure are both influential in determining the effective shielding of the kinemassic field permeable structure. If the shielding is sufficiently effective so as to reduce the primary gravitational field flux line density within the enclosed space to less than that of the ambient, the distortion of the local gravitational field flux line pattern of reduced flux density would result in the antiparallel field force vector magnitudes exceeding that force of the kinemassic field body's initial weight, i.e., the effective weight of the kinemassic field permeable structure absent the generated undulating kinemassic field. This condition would effectively endow the kinemassic field structure with a negative weight characteristic. As a consequence, the two bodies, that is the kinemassic field body and the earth or other cosmic body, would experience relative motion separating one another along the local primary gravitational field flux lines unless acted upon by other forces.

The hardware required to generate and sustain such an undulating kinemassic field is in part comprised of components which continue to retain a "non-field-energized" weight during the period of field generation. Therefore, the ambient gravitational field flux line pattern within the structure will simultaneously experience zones of both reduced and increased densities. It is the average density of all the zones that determines the magnitude of the antiparallel field force vecor in its opposition to the ambient gravitational field force vector. Bodies located within the shielded space enclosed by the undulating kinemassic field will lose their weight with respect to the body earth in direct proportion to the reduction of ambient gravitaional field flux lines which are common to it and the body earth.

As a consequence of the above, the shielding which results in a reduction of mutual attraction between bodies screened by the shielding effected by the undulating kinemassic force field does not violate the principle of equivalence. Thus, two free bodies of different masses, located within the shielded space, will fall within this space toward or away from earth with equal accelerations. Also, the force of mutual gravitational attraction of two or more bodies located within the shielded space will be unaffected by the various degrees of shielding although their free-fall acceleration toward one another will be effected.

Having now further defined the substantiating theory giving rise to the kinemassic forces operative in the present invention, reference is now made to the aforementioned drawings depicting in general an apparatus embodying the defining characteristics outlined above.

From the foregoing discussion, it will be appreciated that both for the purpose of detecting the kinemassic field forces operative in the present invention, and for transforming such forces into energy for doing useful work, several basic apparatus elements are necessary. First, apparatus is needed to enable masses to be placed in relative motion to one another; which relative motion may occur in two mutually orthogonal directions. In order to maximize field strength the apparatus should be capable of generating high velocities between the particles in relative motion. Furthermore, the apparatus should be configured so that the proximity of the particles which are in relative motion is maximized. The necessity of using relatively dense material comprising half integral spin nuclei for the field circuit portion of the apparatus has already been stressed. These and other features are discussed in greater detail below in explanation of the drawings depicting an implementation of the invention.

In considering the drawings, reference will first be made to the general arrangement of components shown in FIGS. 1 and 2. As viewed in FIG. 1, the equipment is mounted upon a stationary base comprising a horizontal structural element **11** which rests upon poured concrete, precast concrete pilings not shown, or other suitable structurally rigid material. It should be made clear at the outset, that the stationary base although not a critical element in its present form nevertheless serves an important function in the subject invention. Thus, the stationary base acts as a stabilized support member for mounting the equipment and, perhaps more significantly, the horizontal portion thereof is of such material that it tends to localize the kinemassic force field to the kinemassic force field generating apparatus proper. This latter feature is discussed in more detail below. The surface uniformity of the horizontal structural element **11** also facilitates the alignment of equipment components. In the reduction to practice embodiment of the present invention a layer of shock absorbing material, not shown, was interposed between the stationary base and the floor.

Shown mounted on the horizontal structural element **11** is the kinemassic force field generating apparatus indicated generally as **10**, the lower portion of which is referred to as the lower mass member **12**. The lower mass member **12** is not to be confusingly associated with the mass circuit mentioned above as being positioned in the space experiencing the secondary gravitational field. The nature and specific identity of the latter mentioned mass circuit will be more fully explained in connection with the explanation of FIG. 7.

An upper mass member **13** is positioned in mirrored relationship with respect to member **12** and separated somewhat to provide two air gaps therebetween. The lower and upper mass members **12** and **13** function as field circuit members in relationship to a generator **14** and a detector **15** positioned with respective one's of said two gaps. The spatial relationship of the generator **14**, the detector **15** and the mass members **12** and **13** is such as to form a kinemassic force field series circuit.

All of the material members of the field circuit are comprised of half integral spin material. For example, the major portion of the generator **14**, the detector **15**, as well as the upper and lower mass members **13** and **12** are formed of a particular brass alloy containing 89% copper of which both isotopes provide a three-halves proton spin, 10% zinc, and 1% lead as well as traces of tin and nickel. The zinc possesses one spin nuclei isotope which is 4.11% in abundance and likewise the lead also contains one spin nuclei isotope which is 22.6% in abundance. In order to gain an estimate of apparatus size, the upper mass member **13** has an overall length of 56 centimeters and a mass of 43 kilograms.

It will be seen that, by far, the constituents of the mass members are such as satify the criteria of half integral spin nuclei material for those apparatus parts associated with the field and the use of non-spin nuclei material for those parts where it is desired to inhibit the field. Accordingly, all support or structural members, such as the horizontal structural element **11**, consist of steel. The iron and carbon nuclei of these structural members are classed as no-spin nuclei and thus represent high relative reluctance to the kinemassic field. Supports **16** are provided to accommodate the suspension of the upper mass member **13**. The supports **16** are of steel the same as the horizontal support element **11**. The high relative reluctance of steel to the kinemassic field minimizes the field flux loss created in the field series circuit of mass members **12** and **13**, the generator **14** and the detector **15**. The loss of field strength is further minimized by employing high-reluctance isolation bridges at the point of contact between the lower and upper mass members **12** and **13**, and the structural support members **11** and **16**.

Shunt losses within the apparatus are in general minimized by employing the technique of minimum mass contact; the use of low field permeability material at the isolation bridges or structural connections; and avoiding bulk mass proximity of non-field generating components.

A number of techniques were devolped for optimizing the isolation bridge units including carboloy cones and spherical spacers. As is depicted more clearly in FIGS. 3, 4 and 5, the structural connection unit ultimately utilized consisted of a hardened 60° steel cone mounted within a set screw and bearing against a hardened steel platen. The contact diameter of the cone against the platen measured approximately 0.007 inch and was loaded within elastic limits. Adjustment is made by means of turning the set screw within a mated, threaded hole.

FIG. 2 is presented in rather diagrammatic form; however, the diagrammatic configuration emphasizes that it consists of a rotatable member corresponding to the generator **14** of FIG. 1 which is "sandwiched" between a pair of generally U-shaped members corresponding to the lower and upper mass members **12** and **13** of FIG. 1. The wheel of generator **14** is mounted for rotation about an axis lying in the plane of the drawing. The generator assembly is also shown as being mounted for rotation about an axis perpendicular to the plane of the drawing; however, the generator assembly may alternatively be oriented to rotate in the plane of the drawing. When member **14** is rotated rapidly with respect to the U-shaped members **12** and **13**, a kinemassic field is generated which is normal to the plane defined by the rotating member and within the plane of the drawing. As such, it may be represented in the drawing of FIG. 2, as taking a generally counter-clockwise direction with respect to the field series circuit members.

Referring once more to FIG. 1, it is seen that support for the generator unit **14** is provided by way of a support assembly **17**, also fabricated of steel components. The support assembly **17** is in turn clamped to the horizontal structural element **11** by way of bracket assemblies **18**.

The detector **15** is of similar configuration to the generator assembly **14**, the exception being that the detector assembly is mounted for limited rotation about the axis normal to the plane of the paper. The limited rotational capability is effected by a knife-edge mounting **19** more clearly discernible in FIGS. 3A and 3B. As will become more readily apparent from the discussion of the operation of the embodiment of FIG. 1 which follows, the knife-edge mounting enables a slow sinusoidal oscillation of the detector assembly about its knife-edge axis.

A pair of light-emitting and light-sensing elements **20***a* and **20***b* respectively are shown in FIG. 1 in operative relationship to the generator and detector assemblies **14** and **15**. The function of the light-generating and-sensing members **20***a* and **20***b* is to measure the rate of spin of the generator and detector wheels respectively. For this purpose every other quadrant on the rim surface is painted black. Accordingly, light directed at the rim of the wheel will be reflected by the unpainted quadrants into light-sensing cells associated with a rate-sensing circuit of conventional design. Since the rate-measuring circuit forms no part of the present invention it has not been depicted in the actual drawing nor is it the subject of further explanation.

Compressed air or nitrogen is used to drive the generator and detector wheels. In this respect the compressed gas is directed against turbine buckets **21***b* cut in the rim of the wheel **21** of both the generator and detector assemblies and such buckets are more readily discernible in FIGS. 3, 4 and 5. The compressed gas is supplied to the generator and detector assemblies by way of air supply lines **28***a* and **28***b*.

FIGS. 3, 4 and 5 present details of the generator and detector assemblies of FIGS. 1 and 2. In particular, these figures disclose the relationship between a freely rotatable wheel **21**, a bearing frame **22**, and a pair of pole pieces **23**. The bearing frame **22** is of structural steel, and functions to spatially orient the three generator parts without shunting the generated field potential as well as to maintain this orientation against the force moment stresses of precession.

Positioning of the generator wheel **21** with respect to the cooperative faces of the pole pieces **23** is effected by way of the bearing frame upon which the generator wheel is mounted. In this respect the high-reluctance isolation bridges mentioned with respect to FIGS. 1 and 2 are herein shown as set screws **24** which are adjustably positioned to cooperate with hardened steel platens **25**. The set screws **24** are mounted on the pole pieces **23** and are adjustably positioned with respect to steel platens **25** cemented to the bearing frame **22** so as to facilitate the centering of the generator wheel **21** with respect to the interface surfaces **23***a* of the pole pieces **23**.

In the implementation of the present invention the air gap formed between the generator wheel rim flanges and the stationary pole pieces **23** was adjusted to a light-rub relationship when the wheel was slowly rotated; as such this separation was calculated to be 0.001 centimeter for a wheel spin rate of 28,000 revolutions per minute due to the resulting hoop tension. In the drawing of FIG. 3 the spacing between the pole pieces **23** and the generator wheel rim flange has been greatly exaggerated to indicate that in fact such a spacing does exist.

The generator wheel **21** utilized in the implementation of the present invention has a 8.60 centimeter diameter and an axial rim dimension of 1.88 centimeters. The rim flange surfaces **21***a* which are those field emanating areas closely adjacent the surfaces **23***a* of the pole pieces **23**, are each 29.6 square centimeters. The rim portion of the wheel has a volume of 55.7 cubic centimeters neglecting the rim turbine slots **21***b*.

The generator wheel **21** and an associated mounting shaft **26** are mounted on the bearing frame **22** by means of enclosed double sets of matched high speed bearings **27**.

Shaft members **30** carry suitable bearing members **31** for rotatably mounting the generator assembly will respect to a second axis. The support assembly **17** of FIG. 1 is partially represented in FIG. 3, and as noted above provides the mounting means for positioning the generator assembly **14** with respect to the lower and upper mass members **12** and **13**.

Reference is now made to FIGS. 3A and 3B which disclose a portion of the detector **15** of FIG. **1** including the knife-edge mounting **19** of FIG. 1. Adjusting means **32** are shown connected to the bearing frame **22***a* of the detector assembly **15** by means of a disc-like member **33**. Attached to the lower portion of the disc **33**, and depicted in the end view of the detector assembly of FIG. 3B, is shown a second adjusting member **34**, which in combination with equivalent members **32** and **34** mounted on the other end of the detector assembly, provide means for symmetrically aligning the detector assembly within the gap provided by the lower and upper mass members **12** and **13**. This further means that the knife-edge assembly is mounted so that the knife-edge axis is coincident with the geometric axis of the detector assembly. At the same time, the center of mass of the detector assembly is located below the geometric center of the detector assembly thereby providing a righting moment to the assembly due to the asymmetry of the mass center with respect to the knife-edge axis. The adjusting means **32** is shown as bearing against the support assembly **17**, thereby, in combination with the knife-edge mounting at either end of the detector assembly, providing an effective four point suspension for symmetrically positioning the detector assembly **15** within the end poles of the upper and lower mass members.

In FIGS. 1, 2 and 3 the detector assembly **15** is shown in three different positions. As will become apparent from the discussion of the operation of the subject system which follows, the facility to so reposition the detector assembly is necessary to demonstrate its operative capabilities. Accordingly, the bearing frame **22***a* is rotatably mounted with respect to the disc **33** by means of a bearing surface interfacing the frame **22***a* with the shaft **35**, the latter being affixed to the face of the disc **33**.

Proceeding now to an explanation of the operation of the embodiment of the invention thus far disclosed, it will be appreciated that in accordance with the theory of operation of the present apparatus when the generator wheel is made to spin at rates upwards of 10 to 20 thousand revolutions per minute, effective polarization of spin nuclei within the wheel structure gradually occurs. This polarization gradually gives rise to domain-like structures which continue to grow so as to extend their field dipole moment across the interface separating the rim **21** from the pole pieces **23**. Secondary dynamic interaction of gravitational coupling increases the field flux lines around the kinemassic force generating assembly, thus resulting in ever increasing total nuclear polarization of half integral spin nuclei.

The non-electromagnetic forces so generated within the subject apparatus are primarily channeled through the high-kinemassic permeability material defining the series field circuit of the apparatus. The fact that the high speed rotatable wheels of both the generator and detector assemblies are capable of being positioned in a series aiding or series opposing relationship, facilitates the determination of the effective influence of the energies generated in one on the other.

The detector, when carefully balanced on its knife-edges as shown in FIGS. 3A and 3B, exhibits an oscillation period of 11 seconds. When the wheels are energized a stiffening action is induced due to the reaction of the compressed gas impingement against the wheel bucket **21***b*, since the jet nozzle is fixed with respect to the apparatus base. This results in a reduction of the oscillation period to approximately 6 seconds. A light image not shown is directed against the mirrored face of the knife-edge **19** and reflected onto a calibrated wall screen. Measurements were taken with the apparatus so operative, which measurements established the oscillatory extremes of the reflected light beam for a pole-aligned relationship of the spinning generator and detector wheels. The results of one such set of measurements are recorded in FIG. 6. Therein, the *x*'s and dots represent extremes in deviation while the larger circles represent the mean thereof. The mean was in turn used to establish a null line to be compared with a similar null line derived from poles-opposed orientation of the generator and detector wheels. As a result, a displacement from equilibrium of approximately 13 arc minutes is shown.

In order to maximize the shift of the null line, the field circuit polar relationship of the generator and detector poles was reversed every 30 or 40 minutes from a relation of poles aligned, to poles opposed, to poles aligned. An average null shift of 26 arc minutes is indicated in FIG. 6. That the interaction between generator and detector was in fact accountable for the recorded results was demonstratively supported when the upper mass member was raised so as to create two air gaps one centimeter in length respectively. Predictably, the disruption to the field circuit continuity resulted in the failure of the apparatus to register a shift in the null lines upon reversal of the poles.

Reference is now made to FIG. 7 which discloses an apparatus constructed in accordance with the principles of the present invention for generating a time-variant secondary gravitational field. This apparatus is a mere modification of the apparatus of FIGS. 1 and 2 wherein one detector assembly **15** has been removed and supplementary means are provided to mechanically implement the rotation of the generator assembly **14** about the axis perpendicular to the plane of the paper. These supplementary means are in the nature of an auxiliary motor **36** having a drive pulley **38** adapted to spin the generator assembly **14** about an axis normal to the plane of the drawing and coincident with that of the shaft **30**. The shaft **30** carries a pulley **40** which is driven by the motor and pulley assembly **36–38** by way of a conventional drive belt **42**. The wheel **21** of the generator assembly **14** is driven in the manner outlined above, namely by means of a source of compressed air not shown.

The supporting assembly depicted in FIG. 7 in partially sectioned form as member **44**, is in reality the equivalent of the series mass circuit of FIGS. 1 and 2, inverted or turned inside out so as to form a shield for the kinemassic field generating apparatus. Included as part of the supporting assembly is member **44A** which is provided to position the generator assembly **14** in the discontinuity formed in the mass circuit. The kinemassic field generated within the apparatus of FIG. 7 upon energization of the wheel **21** is directed in an enveloping fashion about its generator, being confined in general to the shell. The cross sectional thickness of the shell along equipotential lines must be equal in order to ensure a homogeneous field within the structure. If now the spin rate of the wheel **21** is made to vary, or if the generator assembly **14** is made to rotate about the axis defined by the shaft **30**, a time-variant secondary gravitational field is induced in the toroidal space **46**.

The secondary gravitational field undulates in a sinusoidal manner with the undulating kinemassic field confined to the series mass circuit. Since the kinemassic field in the dense mass circuit **44** has been restricted through permeability, into an enveloping shell about the generator **14**, it follows that the induced undulating secondary gravitational field is likewise restricted primarily to the enclosure **46** as the flux lines of both fields must interlink. In accordance with analogous electromagnetic field theory, the kinemassic field flux lines and the secondary gravitational field flux lines interlink in such manner that, as the kinemassic field alternates, these interlinking loops decay and build up in alternate vector directions in proper phase relation.

A hollow ring member **48** is positioned within the toroidal space **46** and supported thereby a series of fine steel wire spokes **50** secured to the ring and the outer portion of the inverted core housing **44** preferably along points of equipotential of the kinemassic field. Within the hollow ring **48** is contained a dense fluid such as mercury depicted in FIG. 7 generally as member **52**. Alternatively, the ring-fluid combination may take the form of a single solid mass. In the latter event the mass circuit would be supported on bearings facilitating its rotation about an axis common to the axis of the generator wheel **21** in order to permit mass flow or rotation of the mass circuit under the influence of the alternating secondary gravitational field. The shielding effected by the design considerations of the toroidal shell **44** with respect to the primary gravitational field reduces the inertial parameter of mass acceleration within the toroidal space **46** in proportion to the ambient gravitational shielding effect. With reduced inertia there will be an appreciable rotational flow displacement of the mass circuit **48–52** for each half cycle of the induced secondary gravitational field, thereby further strengthening the coupling effected between the effective field forces i.e., the primary gravitational field, the kinemassic field and the secondary gravitational field.

Consider now that the apparatus of FIG. 7 is energized such that the wheel **21** spins about its axis creating a uniformly distributed kinemassic field throughout the entire field circuit referred to generally as that encompassed within the inverted core housing **44**. As the generator assembly **14** is energized to rotate about the axis passing through shaft **30**, a uniformly distributed alternating kinemassic field is established throughout the field circuit.

The presence of the undulating kinemassic field produces a shielding effect within the inverted housing effectively restricting the induced secondary gravitational field while at the same time tending to shield or force out the flux due to the ambient gravitational field. As the spin rates of the wheel and the generator assembly about their respective axes are increased, there results a stronger undulating kinemassic force field of higher frequency. The spin rates may be so varied that a mean gravitatinal flux line density within the apparatus of FIG. 7 exists which is equivalent to the primary gravitational flux line density i.e., that due to the earth and other cosmic bodies. This condition establishes a state of weightlessness or zero gravitational force of attraction with respect to other masses such as earth, for that particular value of gravitational field gradient.

If the spin rates of the wheel and the generator assembly are further increased there results a "bowing-out" or spreading of the gravitational flux lines within the immediate proximity of the apparatus of FIG. 7 so as to result in a lesser local flux line density, thus resulting in the propulsion of the apparatus along the local gravitational field lines of force in a direction diametrically opposed to the local gravitational field force vector.

Because of the nature of the interaction of the primary gravitational field, the secondary gravitational field and the kinemassic field, the secondary gravitational field forces will continue to act upon the apparatus as it passes into lesser gravitational field gradients; however, it will do so with diminishing magnitudes until the local gravitational flux line density about the apparatus of FIG. 7 is no longer effectively diminished thereby. The energy required to propel a vehicle powered by an engine, such as is described above, is accounted for by way of the gravitational field potential energy gained by such a vehicle as it passes to areas of lesser gravitational field intensities. Energy input into this engine would appear as the product of torque and rotational values about the spin axes of both the wheel and the generator assembly, and especially about the latter axis which is responsible for alternating the kinemassic field and thereby generating the secondary gravitational component.

As was mentioned above in explanation of the embodiment of FIGS. 1 and 2, the wheel **21** and the generator assembly **14** are mounted so as to be rotatable in mutually orthogonal directions. It was further mentioned that such orthogonal rotation is not an absolute necessity, it being only necessary that relative motion be established between the wheel **21** and the stationary pole pieces **23**. The generator assembly is made to rotate thereby effecting an undulation in the kinemassic field flux in the associated mass circuit. FIG. 7A and 7B disclose a variation of the apparatus of FIG. 7 which satisfies the basic requirements outlined above while at the same time providing certain advantages not available in the aforementioned structure.

In this respect FIGS. 7A and 7B disclose an embodiment wherein the spin axis of the equivalent wheel structure **21** and the generator assembly **14** are concentric thereby eliminating precessional forces present in the embodiment of FIG. 7 due to the rotation of the respective members about the two mutually orthogonal axes. The absence of precessional forces permits a close tolerance to be established between the cooperating faces of the wheel structure **21**, the pole pieces **23** and the mass circuit **44**.

The embodiment of FIGS. 7A and 7B is also to be preferred to that of FIG. 7 in that the design of the generator assembly of the former permits the energization of the independently rotatable members **21** and **23** by means of a single motor **36** differentially geared so as to effect the rotation of the wheel **21** at a speed far in excess of that of the generator assembly, and as indicated, in a reverse direction thereto.

Also indicated in the embodiment of FIGS. 7A and 7B is the orientation of the flux within the mass circuit, the latter being constructed preferredly of bismuth. It should be understood that the direction of flux within the mass circuit reverses with each reverseal in orientation of the equivalent pole pieces **23** due to the rotation of the generator assembly **14**.

It will be apparent from the foregoing description that there has been provided an apparatus for generating time-variant kinemassic forces due to the dynamic interaction of relatively moving bodies. Although in its disclosed application, the time-variant kinemassic force has been described in relation to its function of generating a secondary gravitational force, it should be readily apparent that other equally basic applications of these forces are contemplated.

Thus, in addition to providing an effective propulsion technique, the principles of the present invention may be utilized for the purpose of generating localized areas of gravitational shielding for housing medical patients for which such weight reductions would be beneficial. In addition, the principles may be adapted to laboratory use, as for example the analysis of the effects of a sustained reduction of "*g*" value upon astronauts and for specialized manufacturing techniques.

While in accordance with the provisions of the statutes there has been illustrated and described the best form of the invention known, it will be apparent to those skilled in the art that changes may be made in the apparatus described without departing from the spirit of the invention as set forth in the appended claims, and that in some cases, certain features of the invention may be used to advantage without a corresponding use of other features.

Having now described the invention, what is claimed as new and for which it is desired to secure Letters Patent is:

1. An apparatus for establishing a time-variant kinemassic force field resulting from the relative motion of moving bodies, comprising a generator assembly independent portions of which are mounted to assume relative rotational motion about at least a single axis located within said generator assembly, a mass circuit of dense material of discontinuous configuration, means for positioning said generator assembly within said mass circuit discontinuity, and means for initiating independent relative rotational motion of said generator assembly portions whereby an undulating kinemassic force field is established within said mass circuit.

2. Apparatus according to claim **1** further characterized in that said mass circuit and said relatively moving portions are comprised of spin nuclei material.

3. An apparatus according to claim **1** wherein said mass circuit is further characterized by first and second U-shaped members positioned in mirrored relationship with respect to each other and displaced somewhat so as to form two gaps therebetween, one of said gaps corresponding to said mass circuit discontinuity and being adapted to receive said generator assembly and the other said gap being adapted to receive a detector assembly.

4. Apparatus constructed in accordance with claim **1** wherein said mass circuit is further characterized by a shell of generally toroidal configuration having a cylindrical central portion within which is located said mass circuit discontinuity.

5. An apparatus constructed in accordance with claim **2** wherein said generator assembly mounted within said mass circuit discontinuity further comprises a rotatable member, a frame, means for mounting said rotatable member on said frame, pole pieces mounted on said frame on either side of said rotatable member, each pole piece presenting a generally circular face in close proximity to but spaced from a face of said rotatable member, means for effecting the rotation of said rotatable member about a first axis, and means for rotating said frame about a second axis oriented perpendicular to said first axis.

6. Apparatus constructed in accordance with claim **4** and further characterized by a dense mass ring mounted within the walls of said shell structure by mounting means establishing small area contact between said mass ring and said shell structure.

7. An apparatus constructed in accordance with claim **6** wherein said dense mass ring is further comprised of a hollow shell housing a liquid metal of suitable density.

8. Apparatus according to claim **6** wherein said dense mass ring has as its axis the axis of revolution defining the surface of said shell.

9. Apparatus constructed according to claim **3** wherein said shell is further characterized as being of equal cross sectional area normal to the kinemassic field lines of force.

10. A method of generating a time-variant kinemassic force field including the steps of:

juxtaposing in field series relationship a first member comprised of spin nuclei material of half integral spin value and a second member similarly constituted, portions of said first member being adapted to assume relative rotational motion about at least a single axis;

initiating the independent rotation of said first member about at least a single axis whereby an undulating kinemassic force field is established therein;

and so configuring said second member as to confine said undulating kinemassic force field thereto whereby a time-variant secondary gravitational force field is induced in the surrounding space.

No references cited

HARLAND S. SKOGQUIST, Primary Examiner

UNITED STATES PATENT OFFICE
CERTIFICATE OF CORRECTION

Patent No. 3,626,605 Dated December 14, 1971

Inventor(s) Henry W. Wallace

It is certified that error appears in the above-identified patent and that said Letters Patent are hereby corrected as shown below:

IN THE SPECIFICATION

Column 1, lines 19 to 42, should appear as a part of the "Abstract of the Disclosure".

Column 2, line 31, "three" should read --there--.

Column 3, line 31, "if" should read --of--.

Column 3, line 32, "of" should read --if--.

Column 7, line 51, "with" should read --within--.

Column 8, line 73, "rate-sensing" should read --rate-measuring--.

Column 11, line 10, "one" should read --the--.

Column 13, line 15, "reverseal" should read --reversal--.

Signed and sealed this 5th day of September 1972.

(SEAL)
Attest:

EDWARD M.FLETCHER,JR.
Attesting Officer

ROBERT GOTTSCHALK
Commissioner of Patents

Dec. 14, 1971 **H. W. WALLACE** **3,626,606**

METHOD AND APPARATUS FOR GENERATING A DYNAMIC FORCE FIELD

Filed Nov. 4, 1968 4 Sheets-Sheet 1

FIG-1

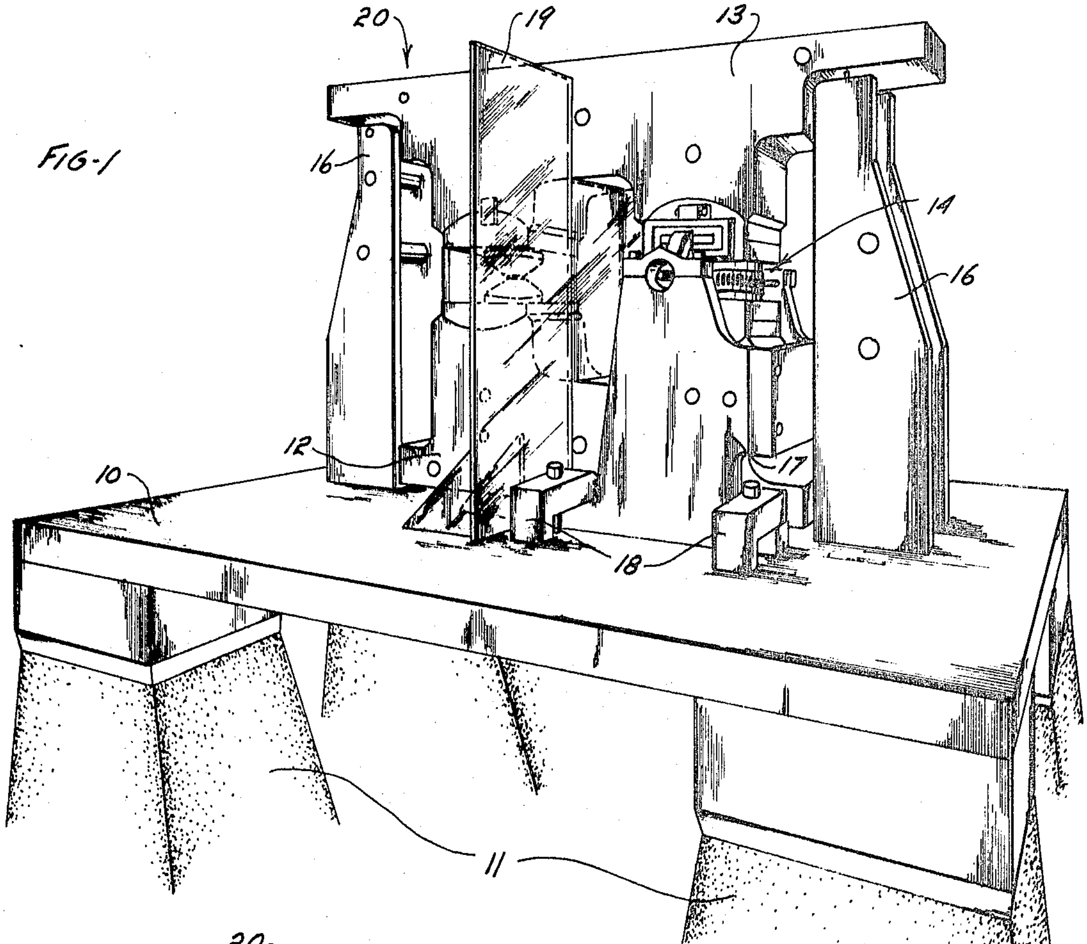

20
19
13
16
14
16
12
10
17
18
11

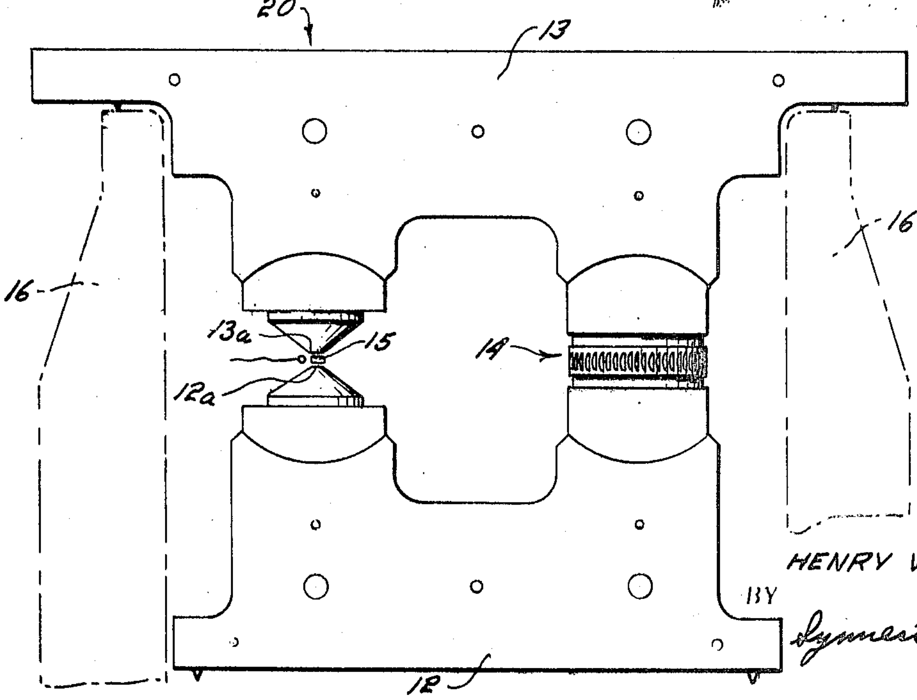

20
13
16
16
13a
15
14
12a
12

FIG-2

INVENTOR.

HENRY W. WALLACE

BY Synnestvedt & Lechner

ATTORNEYS

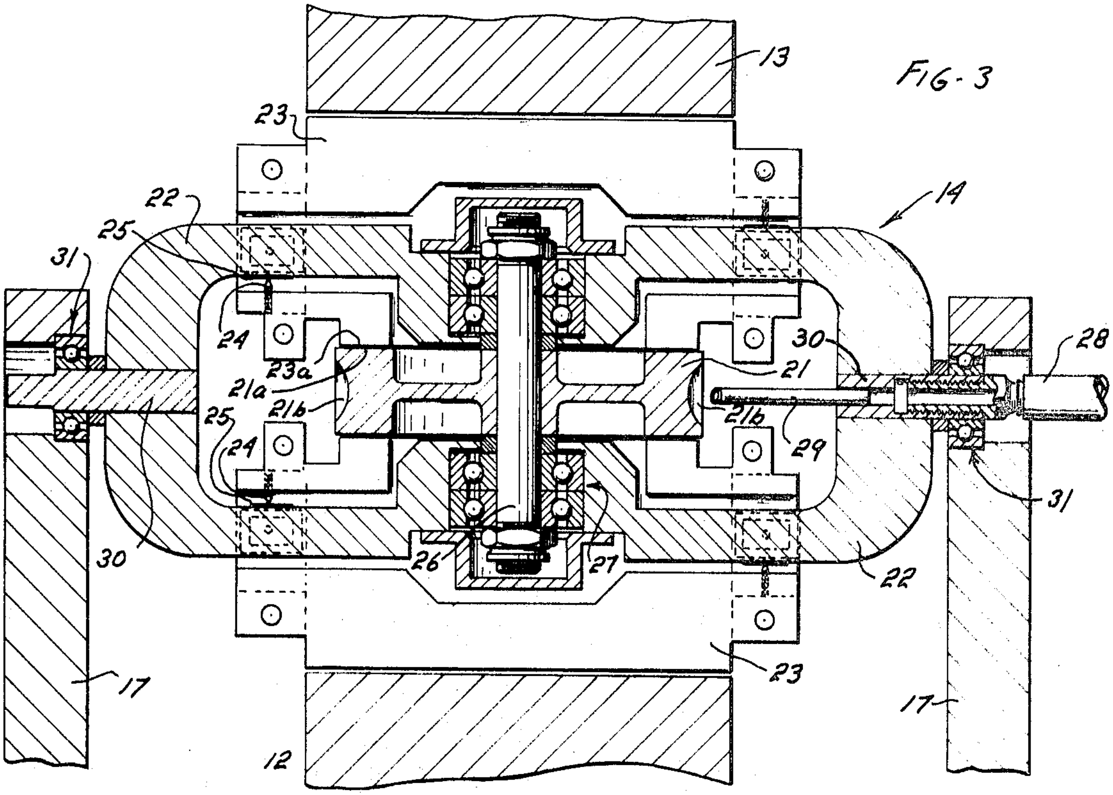
13
23
14
22
25
31
24
23a
21a
21b
30
21
28
25
24
21b
29
31
30
26
27
22
23
17
17
12

FIG-3

FIG-4

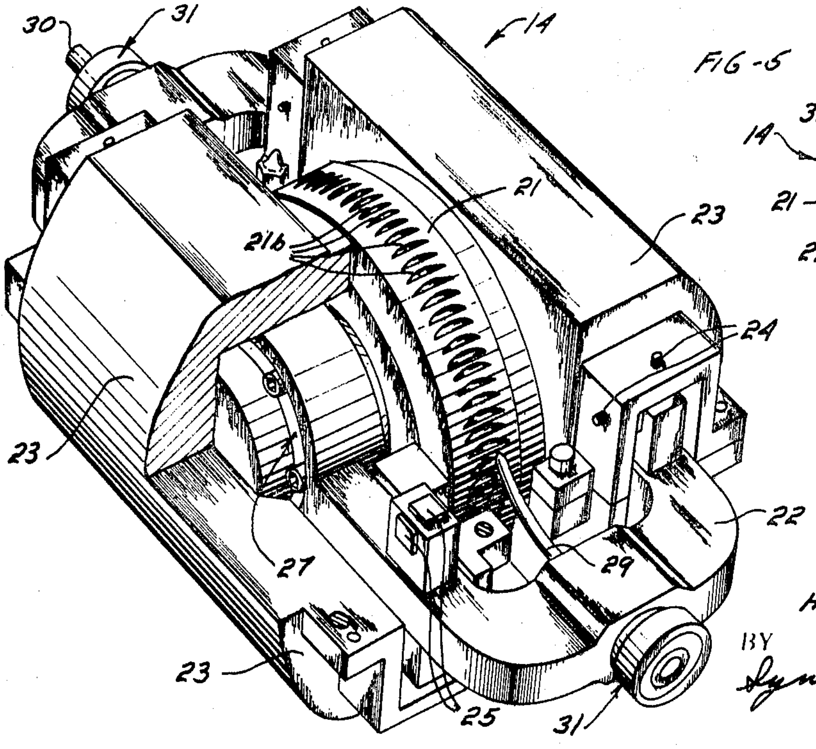
30
31
14
21
21b
23
24
23
22
27
29
23
25
31

FIG-5

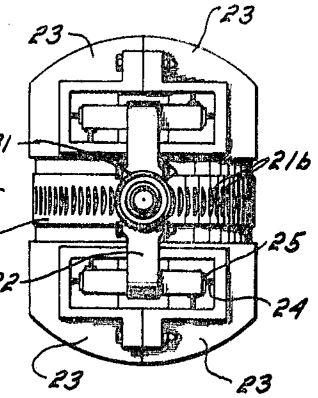
23
23
31
21b
14
21
25
22
24
23
23

INVENTOR.
HENRY W. WALLACE
BY
Synnestvedt & Lechner
ATTORNEYS

FIG-6

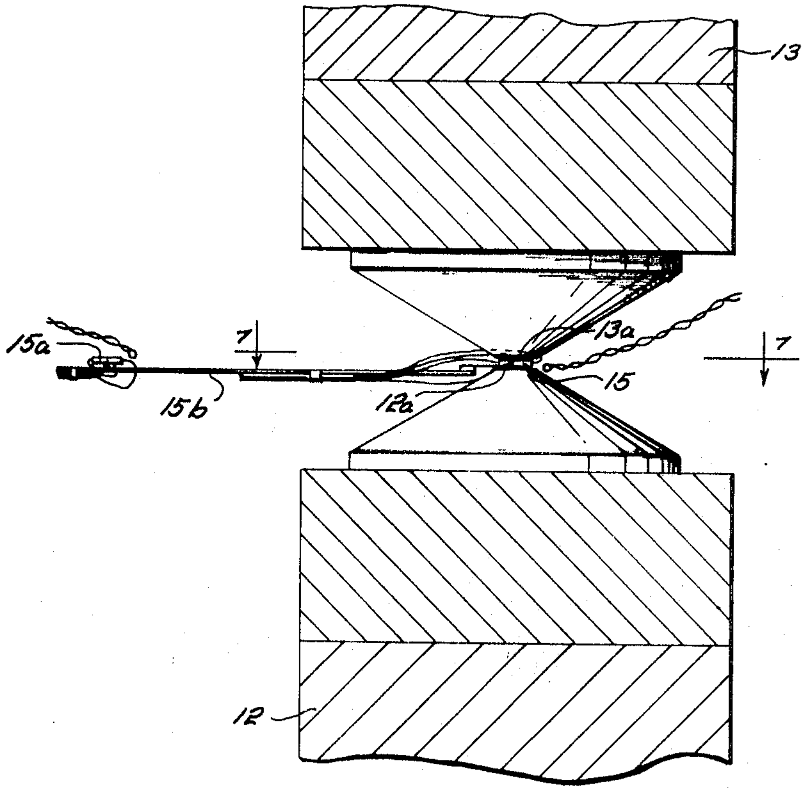
13
13a
7
7
15a
15b
12a
15
12

FIG-7

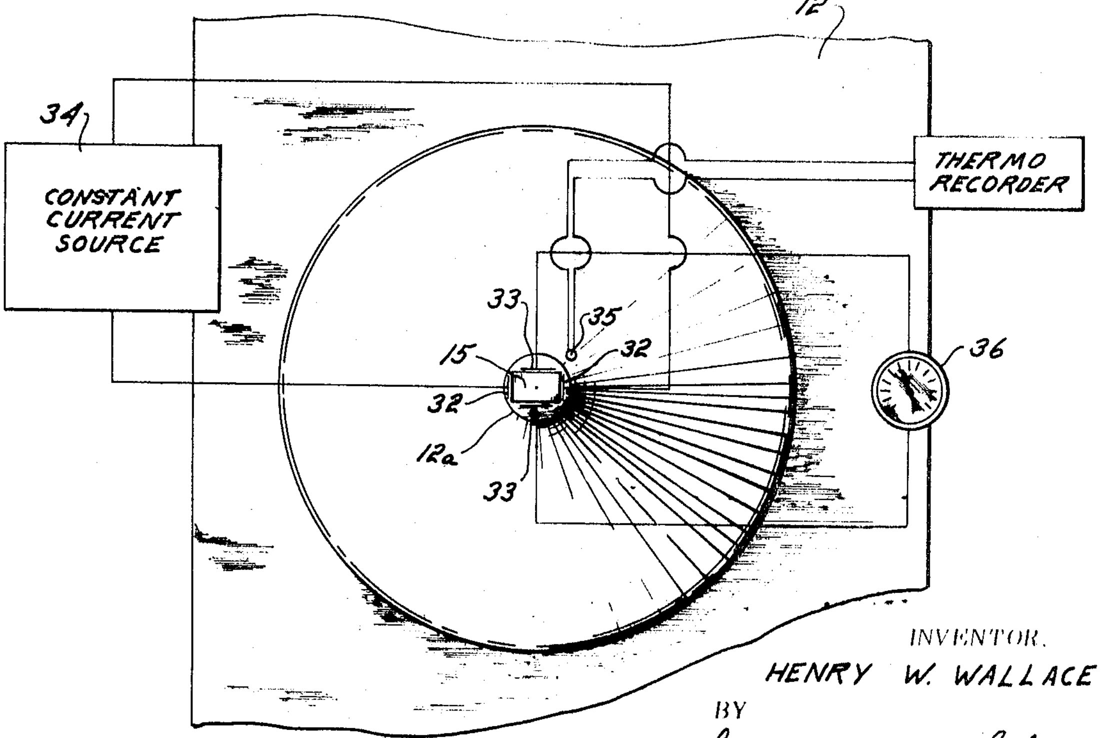
12
34
CONSTANT CURRENT SOURCE
THERMO RECORDER
33
35
32
15
32
36
12a
33

INVENTOR.
HENRY W. WALLACE
BY
Synnestvedt & Lechner
ATTORNEYS

Dec. 14, 1971 H. W. WALLACE **3,626,606**

METHOD AND APPARATUS FOR GENERATING A DYNAMIC FORCE FIELD

Filed Nov. 4, 1968 4 Sheets-Sheet 4

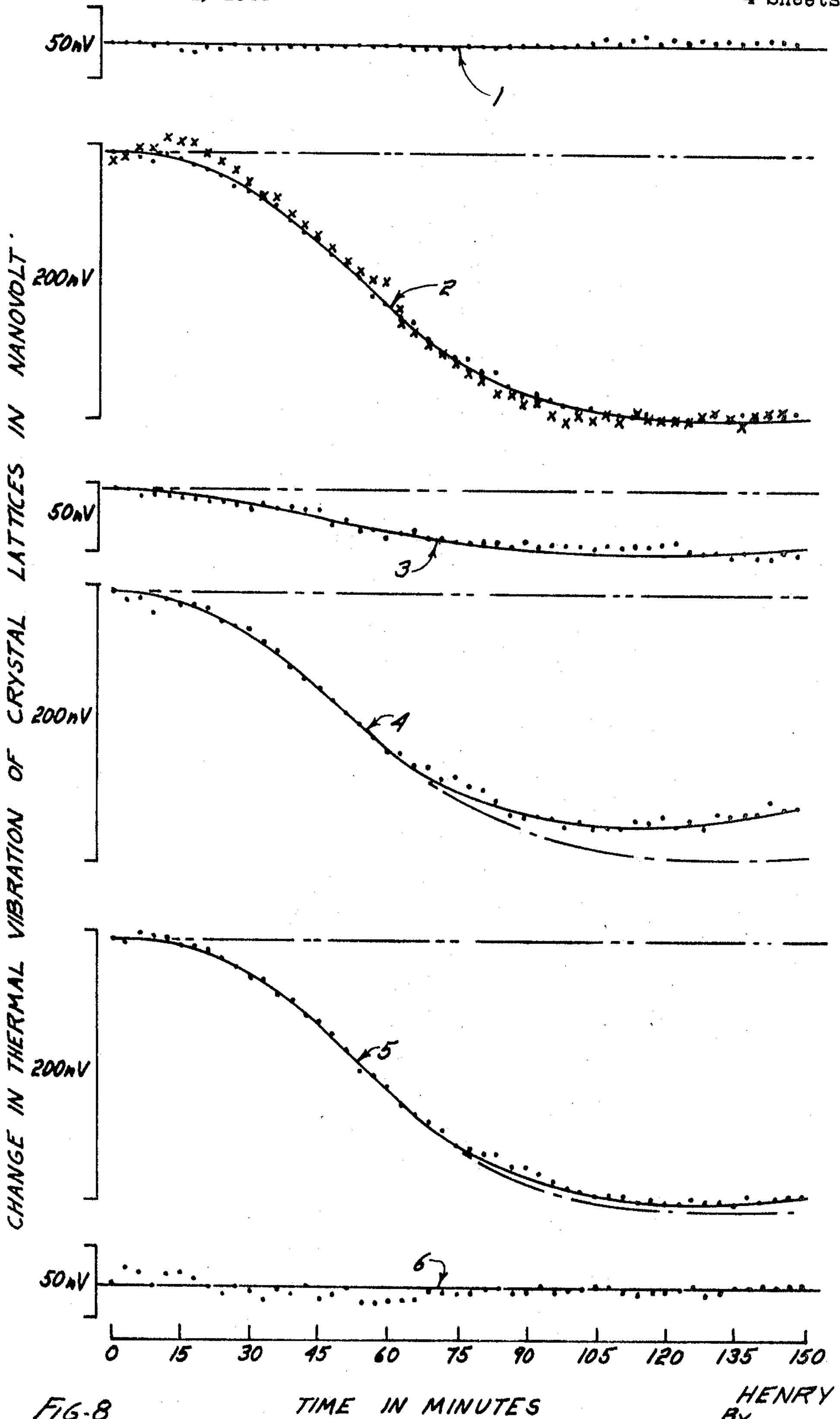

FIG-8

INVENTOR
HENRY W. WALLACE
By.
ATTORNEYS

United States Patent Office 3,626,606
Patented Dec. 14, 1971

3,626,606

METHOD AND APPARATUS FOR GENERATING A DYNAMIC FORCE FIELD

Henry W. Wallace, Ardmore, Pa.
(803 Cherry Lane, Laurel, Miss. 39440)

Filed Nov. 4, 1968, Ser. No. 773,116
Int. Cl. G09b 23/06
U.S. Cl. 35—19 10 Claims

ABSTRACT OF THE DISCLOSURE

Apparatus and method for generating a non-electromagnetic force field due to the dynamic interaction of relatively moving bodies through gravitational coupling, and for transforming such force fields into energy for doing useful work.

The method of generating such non-electromagnetic forces includes the steps of juxtaposing in field series relationship a stationary member, comprising spin nuclei material further characterized by a half integral spin value, and a member capable of assuming relative motion with respect to said stationary member and also characterized by spin nuclei material of one-half integral spin value; and initiating the relative motion of said one member with respect to the other whereby the interaction of the angular momentum property of spin nuclei with inertial space effects the polarization of the spin nuclei thereof, resulting in turn in a net component of angular momentum which exhibits itself in the form of a dipole moment capable of dynamically interacting with the spin nuclei material of the stationary member, thereby further polarizing the spin nuclei material in said stationary member and resulting in a usable non-electromagnetic force.

This invention relates to an apparatus and method for use in generating energy arising through the relative motion of moving bodies and for transforming such generated energy into useful work. In the practice of the present invention it has been found that when bodies composed of certain material are placed in relative motion with respect to one another there is generated an energy field therein not heretofore observed. This field is not electromagnetic in nature; being by theoretical prediction related to the gravitational coupling of moving bodies.

The initial evidence indicates that this nonelectromagnetic field is generated as a result of the relative motion of bodies constituted of elements whose nuclei are characterized by half integral "spin" values; the spin of the nuclei being synonymous with the angular momentum of the nucleons thereof. The nucleons in turn comprise the elemental particles of the nucleus; i.e., the neutrons and protons. For purposes of the present invention, the field generated by the relative motion of materials characterized by a half integral spin value is referred to as a "kinemassic" force field.

It will be appreciated that relative motion occurs on various levels, i.e., there may be relative motion of discrete bodies as well as of the constituents thereof including, on a subatomic level, the nucleons of the nucleus. The kinemassic force field under consideration is a result of such relative motion, being a function of the dynamic interaction of two relatively moving bodies including the elemental particles thereof. The value of the kinemassic force field, created by reason of the dynamic interaction of the bodies experiencing relative motion, is the algebraic sum of the fields created by reason of the dynamic interaction of both elementary particles and of the discrete bodies.

For a closed system comprising only a stationary body, the kinemassic force, due to the dynamic interaction of

the particles therein, is zero because of the random distribution of spin orientations of the respective particles. Polarization of the spin components so as to align a majority thereof in a preferred direction establishes a field gradient normal to the spin axis of the elementary particles. The present invention is concerned with an apparatus for establishing such a preferred orientation and as a result generating a net force component capable of being represented in various useful forms.

According, the primary object of the present invention concerns the provision of means for generating a kinemassic field due to the dynamic interaction of relatively moving bodies.

A further object of the present invention concerns a force field generating apparatus wherein means are provided for polarizing material portions of the apparatus so as to reorient the spin of the elementary nuclear components thereof in a preferred direction thereby generating a detectable force field.

The kinemassic force field finds theoretical support in the laws of physics, being substantiated by the generalized theory of relativity. According to the general theory of relativity there exists not only a static gravitational field but also a dynamic component thereof due to the gravitational coupling of relatively moving bodies. This theory purposes that two spinning bodies will exert force on each other. Heretofore the theoretical predictions have never been experimentally substantiated however, as early as 1896, experiments were conducted in an effort to detect predicted centrifugal forces on stationary bodies placed near large, rapidly rotating masses. The results of these early experiments were inconclusive, and little else in the nature of this type of work is known to have been conducted.

It is therefore another object of the present invention to set forth an operative technique for generating a measurable force field due to gravitational coupling of relatively moving bodies.

Another more specific object of the present invention concerns a method of generating a non-electromagnetic force field due to the dynamic interaction of relatively moving bodies and for utilizing such forces for temperature control purposes including the specific application of such forces to the control of lattice vibrations within a crystalline structure thereby establishing an appreciable temperature reduction, these principles being useful for example in the design of a heat pump.

The foregoing objects and features of novelty which characterize the present invention as well as other objects of the invention are pointed out with particularity in the claims annexed to and forming a part of the present spection. For a better understanding of the invention, its advantages and specific objects allied with its use, reference should be made to the accompanying drawings and descriptive matter in which there is illustrated and described a preferred embodiment of the invention.

In the drawings:

FIG. 1 is an overall perspective view of equipment constructed according to the present invention, this equipment being designed especially for demonstrating the useful applications of kinemassic force fields;

FIG. 2 is an isolation schematic of apparatus components comprising the kinemassic field circuit of the apparatus of FIG. 1, showing the field series relationship of generator and detector units;

FIGS. 3, 4 and 5 show the generator of FIGS. 1 and 2 in greater detail;

FIG. 6 is an enlarged view of the detector working air gap area of the apparatus of FIGS. 1 and 2;

FIG. 7 is a sectional view of FIG. 6 showing associated control and monitoring equipment; and

FIG. 8 represents measured changes in the operating characteristics of a crystalline target subject to a kinemassic force field generated in the apparatus of FIGS. 1 and 2.

Before getting into a detailed discussion of the apparatus and steps involved in the practice of the present invention it should be helpful to an understanding of the present invention if consideration is first given to certain defining characteristics many of which bear an analogous relationship to electromagnetic field theory. A first feature is that the kinemassic field is vectorial in nature. The direction of the field vector is a function of the geometry in which the relative motion between mass particles takes place.

The second significant property of the kinemassic field relates the field strength to the nature of the material in the field. This property may be thought of as the kinemassic permeability by analogy to the concept of permeability in magnetic field theory. The field strength is apparently a function of the density of the spin nuclei material comprising the field circuit members. Whereas the permeability in magnetic field theory is a function of the density of unpaired electrons, the kinemassic permeability is a function of the density of spin nuclei and the measure of magnitude of their half integral spin values. As a consequence of this latter property, the field may be directed and confined by interposing into it denser portions of desired configuration. For example, the field may be in large measure confined to a closed loop of dense material starting and terminating adjacent a system wherein relative motion between masses is occuring.

A further property of the kinemassic force field relates field strength to the relative spacing between two masses in relative motion with respect to one another. Thus, the strength of the resultant field is a function of the proximity of the relatively moving bodies such that relative motion occurring between two masses which are closely adjacent will result in the generation of a field stronger than that created when the same two relatively moving bodies are spaced farther apart.

As mentioned above, a material consideration in generating the kinemassic force field concerns the use of spin nuclei material. By spin nuclei material is meant materials in nature which exhibit a nuclear external angular momentum component. This includes both the intrinsic spin of the unpaired nucleon as well as that due to the orbital motion of these nucleons.

Since the dynamic interaction field arising through gravitational coupling is a function of both the mass and proximity of two relatively moving bodies, then the resultant force field is predictably maximized within the nucleus of an atom due to the relatively high densities of the nucleons, both in terms of mass and relative spacing, plus the fact that the nucleons possess both intrinsic and orbital components of angular momentum. Such force fields may in fact account for a significant portion of the nuclear binding force found in all of nature.

It has been found that for certain materials, namely those characterized in a half integral spin value, the external component of angular momentum thereof will be accompanied by a force due to the dynamic interaction of the nucleons. This is the so-called kinemassic force which on a submacroscopic basis exhibits itself as a field dipole moment aligned with the external angular momentum vector. These moments are of sufficient magnitude that they interact with adjacent, or near adjacent spin nuclei field dipole moments of neighboring atoms.

This latter feature gives rise to a further analogy to electromagnetic field theory in that the interaction of adjacent spin nuclei field dipole moments gives rise to nuclear domain-like structures within matter containing sufficient spin nuclei material.

Although certain analogies exist between the kinemassic force field and electromagnetic field theory, it should be remembered that the kinemassic force is essentially nonresponsive to or affected by electromagnetic force phenomena. This latter condition further substantiates the ability of the kinemassic field to penetrate through and extend outward beyond the ambient electromagnetic field established by the moving electrons in the atomic structure surrounding the respective spin nuclei.

As in electromagnetic field theory, in an unpolarized sample, the external components of angular momentum of the nuclei to be subjected to a kinemassic force field, are originally randomly oriented such that the material exhibits no residual kinemassic field of its own. However, establishing the necessary criteria for such a force field effects a polarization of the spin components of adjacent nuclei in a preferred direction thereby resulting in a force field which may be represented in terms of kinemassic field flux lines normal to the direction of spin.

The fact that spin nuclei material exhibits external kinemassic forces suggests that these forces should exhibit themselves on a macroscopic basis and thus be detectable, when arranged in a manner similar to that for demonstrating the Barnett effect when dealing with electromagnetic phenomena.

In the Barnett effect a long iron cylinder, when rotated at high speed about its longitudinal axis, was found to develop a measurable component of magnetization, the value of which was found to be proportional to the angular speed. The effect was attributed to the influence of the impressed rotation upon the revolving electronic systems due to the mass property of the unpaired electrons within the atoms.

In the apparatus constructed in accordance with the foregoing principles it was found that a rotating member composed of spin nuclei material exhibits a kinemassic force geld. The interaction of the spin nuclei angular momentum with inertial space causes the spin nuclei axes of the respective nuclei of the material being spun to tend to reorient parallel with the axis of the rotating member. This results in the nuclear polarization of the spin nuclei material. With sufficient polarization, an appreciable field of summed dipole moments emanates from the wheel rim flange surfaces to form a secondary dynamic interaction with the dipole moments of spin nuclei contained within the facing surface of a stationary body positioned immediately adjacent the rotating member.

When the stationary body, composed of suitable spin nuclei material, is connected in spatial series with the rotating member, a circuitous form of kinemassic field is created; the flux of which is primarily restricted to the field circuit.

Having now further defined the substantiating theory giving rise to the kinemassic forces operative in the present invention, reference is now made to the aforementioned drawings depicting in general an apparatus embodying the defining characteristics outlined above.

From the foregoing discussion, it will be appreciated that for both the purpose of detecting and exploiting the kinemassic field, several basic apparatus elements are necessary. First, apparatus is needed to enable masses to be placed in relative motion to one another. In order to maximize field strength the apparatus should be capable of generating high velocities between the particles in relative motion. Furthermore, the apparatus should be configured so that the proximity of the particles which are in relative motion is maximized. The necessity of using relatively dense material comprising half integral spin nuclei for the field circuit has already been stressed. These and other features are discussed in greater detail below in explanation of the drawings depicting an implementation of the invention, primarily for detection of the kinemassic field.

In considering the drawings, reference will first be made to the general arrangement of components, as particularly shown in FIGS. 1 and 2. As viewed in FIG. 1, the equipment is mounted upon a stationary base comprising a horizontal structure element **10** which rests upon permanent pilings of poured concrete **11** or other suitable structurally rigid material. It should be made clear at the outset that the stationary base although not a critical element in its present form nevertheless serves an important function in the subject invention. Thus, the stationary base acts as a stabilized support member for mounting the equipment and, perhaps more significantly, the horizontal portion thereof is of such material that it tends to localize the kinemassic force field to the kinemassic force field generating apparatus proper. This latter feature is discussed in more detail below. The surface uniformity of the horizontal structural element **10** also facilitates the alignment of equipment components. In the reduction to practice embodiment of the present invention a layer of shock absorbing material (not shown) was interposed beneath the stationary base and the floor.

Shown mounted on the horizontal structural element **10** is the kinemassic force field generating apparatus indicated generally as **20**, the lower portion of which is referred to as the lower mass member **12**. An upper mass member **13** is positioned in mirrored relationship with respect to member **12** and separated somewhat to provide two air gaps therebetween. The lower and upper mass members **12** and **13** function as field circuit members in relationship to a generator **14** and a detector **15** positioned within respective ones of said two gaps. The spatial relationship of the generator, the detector and the mass members is such as to form a kinemassic force field series circuit.

All of the material members of the field circuit are comprised of half integral spin material. For example the major portion of the generator **14**, and the upper and lower mass members **13** and **12**, respectively, are formed of a particular brass alloy containing 89% copper, of which both isotopes provide a three-halves proton spin, 10% zinc, and 1% lead, as well as traces of tin and nickel. The zinc atom possesses one spin nuclei isotope which is 4.11% in abundance and likewise the lead also contains one spin nuclei isotope which is 22.6% in abundance. In order to gain an estimate of apparatus size, the upper circuit member has an overall length of 56 centimeters and a mass of 43 kilograms.

It will be seen that the constituents of the mass members are such as satisfy the criteria of half integral spin nuclei material for those apparatus parts associated with the field and the use of non-spin nuclei material for those parts where it is desired to inhibit the field. Accordingly, all support or structural members, such as the horizontal structural element **10**, consist of steel. The iron and carbon nuclei of these structural members are classed as no-spin nuclei and thus represent high relative reluctance to the kinemassic field. Supports **16** are provided to accommodate the suspension of the upper mass member **13**. The supports **16** are of steel the same as the horizontal support element **10**. The high relative reluctance of steel to the kinemassic field minimizes the field flux loss created in the field series circuit of mass members **12** and **13**, the generator **14**, and the detector **15**. The loss of field strength is further minimized by employing high-reluctance isolation bridges at the points of contact between the lower and upper mass members **12** and **13**, and the structural support members **10** and **16**.

Shunt losses within the apparatus were, in general, minimized by employing the technique of minimum mass contact; the use of low field permeability material at the isolation bridges or structural connections; and avoiding bulk mass proximity.

A number of techniques were developed for optimizing the isolation bridge units including Carboloy cones and spherical spacers. As is depicted more clearly in FIGS. 3, 4 and 5, the structural connection unit ultimately utilized consisted of a hardened 60° steel cone mounted within a setscrew and bearing against a hardened steel platen. The contact diameter of the cone against the platen measured approximately 0.007 inch and was loaded within elastic limits. Adjustment is made by means of turning the setscrew within a mated, threaded hole.

FIG. 2 is presented in rather diagrammatic form; however, the diagrammatic configuration emphasizes that it consists of a rotatable member corresponding to the generator **14** of FIG. 1 which is "sandwiched" between a pair of generally U-shaped members corresponding to the lower and upper mass members **12** and **13** of FIG. 1. The wheel of generator **14** is mounted for rotation about an axis lying in the plane of the drawing. When member **14** is rotated rapidly with respect to the U-shaped members **12** and **13**, a kinemassic field is generated which is normal to the plane defined by the rotating member and within the plane of the drawing.

As such, it may be represented in the drawing of FIG. 2 as taking a generally counterclockwise direction with respect to the field series circuit members.

Referring once more to FIG.1, it is seen that support for the generator unit **14** is provided by way of a support assembly **17**, also fabricated of steel components. The support assembly **17** is in turn clamped to the horizontal structural element **10** by way of bracket assemblies **18**.

In the embodiment of the present invention depicted in FIGS. 1 and 2, the lower and upper mass members **12** and **13** are fashioned into conical sections terminating in conical pole faces **12***a* and **13***a* in the area of the detector **15**. This configuration tends to maximize the flux density in this area.

For isolation purposes, a curtain of transparent plastic material **19** is positioned so as to geometrically bisect the detector portion of the field circuit from the generator portion thereof. The function of the transparent curtain is to provide a degree of thermal isolation between the generator and detector units. Although not actually shown in FIG. 2 the transparent curtain is of H configuration and forms a vertical plane normal to the plane of the drawing and symmetrically positioned with respect thereto.

Not shown in the drawings are a tunnel of transparent material and a film of flexible plastic material which surround the detector **15** and associated equipment and thus serve to further stabilize the temperature conditions, thereby diminishing the adverse effects due to thermal gradients.

Before proceeding with the explanation of the operation of the apparatus disclosed in FIGS. 1 and 2, a more detailed description of certain portions of the structure will be given.

FIGS. 3, 4 and 5 present the generator assembly **14** of FIGS. 1 and 2 in greater detail. In particular, these figures disclose the relationship between a freely rotatable wheel **21**, a bearing frame **22**, and a pair of pole pieces **23**. The bearing frame **22** is of structural steel, and functions to spatially orient the three generator parts without shunting the generated field potential.

Positioning of the generator wheel **21** with respect to the cooperative faces of the pole pieces **23** is effected by way of the bearing frame upon which the generator wheel is mounted. In this respect the high-reluctance isolation bridges mentioned with respect to FIGS. 1 and 2 are herein shown as setscrews **24** which are adjustably positioned to cooperate with hardened steel platens **25**. The setscrews **24** are mounted on the pole pieces **23** and are adjustably positioned with respect to steel platens **25** cemented to the bearing frame **22** so as to facilitate the centering of the generator wheel **21** with respect to the interface surfaces **23***a* of the pole pieces **23**.

In the implementation of the present invention the air gap formed between the generator wheel rim flanges and the stationary pole pieces **23** was adjusted to a light-rub relationship when the wheel was slowly rotated; as such this separation was calculated to be 0.001 centimeter for a wheel spin rate of 28,000 revolutions per minute due to the resulting hoop tension. In the drawing of FIG. 3 the spacing between the pole pieces **23** and the generator wheel rim flange has been greatly exaggerated to indicate that in fact such a spacing does exist.

The generator wheel **21** utilized in the implementation of the present invention has an 8.60 centimeter diameter and an axial rim dimension of 1.88 centimeters. The rim flange surfaces **21***a* which are those field emanating areas closely adjacent the surfaces **23***a* of the pole pieces **23**, are each 29.6 square centimeters. The rim portion of the wheel has a volume of 55.7 cubic centimeters neglecting the rim turbine slots **21***b*.

The generator wheel **21** and an associated mounting shaft **26** are mounted on the bearing frame **22** by means of enclosed double sets of matched high speed bearings **27**. Compressed air or nitrogen is used to drive the generator wheel by means of gas impingement against turbine buckets **21***b* cut in the wheel rim. The compressed gas is supplied through the supply line **28** and emanates from the air jet tube **29**. Rate of spin is sensed by light rays reflected from the rim. For this purpose every other quadrant on the rim surface was painted black. Accordingly, light directed at the rim of the wheel will be reflected by the unpainted quadrants into light-sensing cells associated with a rate-measuring circuit of conventional design. Since the rate-detecting means form no part of the present invention they have not been depicted in the actual drawing.

Shaft members **30** carry suitable bearing members **31** for rotatably mounting the generator asembly with respect to a second axis. The support assembly **17** of FIG. 1 is partially represented in FIG. 4, and, as noted above it provides the mounting means for positioning the generator assembly **14** with respect to the lower and upper mass members **12** and **13**.

Before proceeding with an explanation of the operation of the generator assembly with respect to the apparatus of FIG. 1, reference is made to FIGS. 6 and 7 which disclose an enlarged view of the detector **15**. The lower and upper mass members **12** and **13** are given a conical configuration so as to maximize kinemassic field densities in the area of the working air gap, within which the detector is positioned. FIG. 7 represents a sectional view taken across the working air gap, showing the projection of the conical section of the upper mass member upon the conical section of the lower mass member. Although symmetrical in shape, the projection of the conical surface of the upper mass member onto the corresponding surface of the lower mass member has been slightly reduced for purposes of illustration. In the subject apparatus the two conical brass pole faces **12***a* and **13***a* form a working air gap measuring 0.114 centimeter across. Each disc shape pole face measures 0.71 square centimeter in area.

The detector or probe **15** is of indium arsenide and is inserted in the detector air gap with a spacing from either pole face of 0.02 centimeter, the target thickness measuring 0.07 centimeter. Both indium and arsenic process 100% isotope abundance of half integral spin nuclei; arsenic nuclei consists of one isotope of three halves proton spin, while indium nuclei are of two isotopes, both being of nine-halves proton spin.

A second probe of similar semi-conductor material **15***a* is shown in FIG. 6 as being positioned in close proximity to the first detector. Both probes **15** and **15***a* are shown mounted on a boom **15***b* which is shock mounted by means not shown. Shock mounting of the components is important due to the relatively close spacing between the probe and conical pole faces. Lateral displacement of the second probe from the vicinity of the working air gap measured as 25 centimeters.

Although not critical to the overall theory of the present invention, the selection of a semi-conductor probe of the nature heretofore described and the effective results realized through the positioning of the probe **15** and the associated probe **15***a* with respect to the working air gap between the conical pole faces as well as the manner in which signals measured by the two probes

is correlated, are important to an understanding of the forces involved. In this respect it is important to realize that the first and second semi-conductor probes were differentially connected in terms of electrical output and are polarity-sensitive to magnetic field measurements. Together the two probes constitute a differential magnetic probe for an FW Bell Gaussmeter. As conventionally used, such probes provide a measure of the magnetic field intensity from both AC and DC sources, via the Hall effect. The Hall effect is a well known phenomenon whereby a potential gradient is developed in a direction transverse to the direction of current flow within a conductor when the conductor is positioned in a magnetic field. It should be clearly understood, however, that no magnetic field phenomenon is associated with the present invention. Thus the lateral voltages which are measured in the present arrangement are not Hall voltages. This statement is substantiated by the explanation which follows, clearly establishing the absence of any Hall voltage indicative of magnetic fields. In this respect, the two probes are differentially connected for magnetic field measurements to eliminate errors due to ambient magnetic field changes whereas they are additively connected for sensing changes in thermal vibration of crystal lattices. Although polarity-sensitive to the magnetic field, the differential magnetic probe is not polarity-sensitive to changes in thermal vibration of crystal lattices.

The fact that the probes are polarity-sensitive with respect to magnetic field but not with respect to the direction of crystal lattice vibrations means that when the probes are reversed with respect to polarity any discernible difference in the output readings might be attributed to a magnetic field induced into the system by the rotating wheel. Inasmuch as the field conductive portions of the apparatus are comprised predominately of brass which is a paramagnetic material, no appreciable magnetic field should be detected. This in fact corresponds to the actual results in that no measurable difference in magnetic flux was recorded when the polarity of the probes was changed. It is thus possible to realistically discount magnetic fields as influencing operating results.

As seen in FIG. 7, the detector **15** has associated therewith two pairs of contacts **32** and **33**, the first of which represents current contacts connected in turn to a source of constant current **34** of conventional design. The second set of contacts **33** are voltage contacts connected to detect any potential gradient transverse to the direction of current flow within the detector. The meter **36** represents means for detecting such potential differences and may be in the form of a very sensitive galvanometer.

A thermocouple **35** is positioned in close proximity to the detector **15** to monitor the temperature thereof. Temperature differences, as recorded by the thermocouple **35**, are used for purposes of providing correction figures to the test results. A similar thermocouple is used in conjunction with the second detector **15***a*, as well as with the upper mass member particularly in the area of the generator wheel. Thermocouples are used for temperature monitoring since the energy change of their conducting electrons, by which they sense temperature change, are not measurably affected by the kinemassic field.

Proceeding now to an explanation of the operation of the subject invention, it will be appreciated that in accordance with the theory of operation of the present apparatus when the generator wheel is made to spin at rates upwards of 10 or 20 thousand revolutions per minute, effective polarization of spin nuclei within the wheel structure gradually occurs. This polarization gradually gives rise to domain-like structures which continue to grow so as to extend their field dipole moment across the interface separating the rim **21** from the pole pieces **23**. Secondary dynamic interactions of gravitational coupling between respective dipoles increase the field flux lines around the apparatus field circuit, thus resulting in ever increasing total nuclear polarization of half integral spin nuclei.

The non-electromagnetic forces so generated within the subject apparatus are directed to the working air gap within which is positioned the semi-conductor probe **15**. Therein the kinemassic forces are constructively used to reduce the vibrational degrees of freedom of the crystal lattice structure of the semi-conductor probe resulting in a change of its electrical conductivity property. More specifically, the kinemassic field, due to the dynamic interaction of the gravitational coupling of the mass components of the wheel in relation to the stationary portions of the pole pieces in immediate proximity therewith, is restricted to the relatively high permeability material comprising the lower and upper mass members, and is concentrated at the working air gap by means of the conical pole pieces. Inserted in the air gap is the probe of semi-conductor spin nuclei material.

Control circuitry connected to two of the four contacts on the semi-conductor probe **15** is designed to maintain a constant current flow across these contacts. At the same time the ambient temperature of the area surrounding the equipment is permitted to increase. In fact the increase in ambient temperature is initiated well in advance of the initiation of rotation of the generator wheel giving rise to the non-electromagnetic kinemassic force field. The constant increase in temperature is meant to mask out otherwise positive and negative temperature variations resulting in a reduced signal-to-noise ratio of measurement.

In light of the gradual and constant increase in temperature of both the equipment and ambient conditions surrounding the equipment, it might be expected that the thermal vibrations of crystal lattice of the semi-conductor probe would likewise increase. In actuality, a measurable decrease in crystal lattice vibrations is detected within the semi-conductor probe. The actual measurements recorded are in terms of nanovolts of meter movement, and correspond to a decrease in lateral voltages measured across the semi-conductor probe. These values can only be accounted for by an effective polarization of the spin nuclei of the lattice structure due to the polarizing effects of the applied kinemassic force field. The polarization results in a change in the specific heat property of the crystal material, which in turn reflects itself as an increase in electrical conductivity measurable by the galvanometer.

Reference is now made to FIG. 8 which discloses in graphical relationship the results achieved by various test arrangements of the semi-conductor probe with respect to the subject apparatus.

In the interpretation of the graphical relationships of FIG. 8 it should be understood that corrections for temperature variations have already been applied. These temperature corrections account for the heat applied to the system, that generated within the apparatus due to frictional heating, as well as that due to the change in specific heat property of the apparatus principally the brass members due to their relative bulk. The latter component represents a positive contribution to the ambient temperature due to the decrease in degrees of freedom of the crystal lattice structure of the spin nuclei material when subjected to the kinemassic force field. The above mentioned heat factors result in the increased temperature of the brass members of the apparatus; these increases being monitored by way of the thermocouples positioned in proximity to the kinemassic field generating apparatus, member **35** of FIG. 7 being an example thereof.

Curve **1** of FIG. 8 represents a static test conducted over a period of 150 minutes, values being recorded at 3 minute intervals which was standard procedure for the entire test series. Information gathered in respect to curve **1** was useful in determining compensating factors for ambient temperature changes. In curve **1** as well as each of the other curves of FIG. 8, the ordinate values measure a level of thermal vibration, in nanovolts of meter movement, of InAs lattice structure against time in which ambient temperature change of the two probes has been quantitatively compensated.

Curve **2** represents a portion of a standard test run, the portion shown being the active portion of the curve, i.e., that portion of the curve for which measurable results were recorded due to the spinning of the generator wheel. Not included in curve **2** are measurements taken during a 78 minute preenergizing thermal calibration period typical of the initial portion of each test run conducted. The pre-energizing thermal calibration period is effected in order to illustrate the ambient temperature compensation of the probes and as such is similar to that of the static test of curve **1**.

The first 45 minutes of the indicated 150 minute test period of curve **2** represents the time during which the wheel was made to spin at a rate of 28,000 r.p.m. The continuity of the negatively sloping curve prior to, during and following the time interval of the wheel returning to its no spin state, and somewhat subsequently (an indication of a return toward thermal equilibrium percentage distribution of spin angular moment) is consistent with the explanation advanced above concerning the force field generated due to the dynamic interaction of relatively moving bodies. It should be noted with respect to curve **2** that separate test runs conducted some six weeks apart tend to corroborate the independent test results. The results of the two separate tests are superimposed in curve **2**. These two tests, in addition to being spaced in time, were spaced many test runs apart. The two test results further establish the repeatibility of the operation.

The change in thermal vibration of the InAs crystal lattice for the test run of curve **2** is approximately equivalent to an 11° centigrade reduction in probe temperature. This figure has been substantiated by computer studies. The computer has also been used to statistically analyze the test data and establish the probability of error in terms of the information recorded. In this respect the results of the computerized study indicate a probability of error of 1 in 1 billion. Since any ratio in excess of 1 in 20 eliminates the probability of chance occurrence, the results obtained in the present instance should be above reproach.

In order to substantiate the distance dependency of the gravitational coupling force due to the dynamic interaction of relatively moving bodies it was predicted that increasing the separation between the generator rim flange **21*a*** and the cooperating surface of the pole pieces **23*a*** should measurably reduce the results obtained. The results obtained when this separation was increased to 0.006 centimeters appears in curve **3**. A comparison of these results with those of curve **2** seemingly substantiates the conclusion that upon widening the gap a lessening of the dynamic interaction due to gravitational coupling between the spinning wheel and the stationary pole piece actually occurs.

The data of curve **4** was taken with the air gap separation of the wheel to pole piece established at 0.001 centimeter as in the arrangement of curve **2**; however, the duration of wheel spin was decreased from 45 minutes to 30 minutes. Curve **2** results are shown superimposed on the solid line of curve **4**. The relative magnitudes of curves **2** and **4**, when so contrasted with their respective wheel spin periods, would appear to indicate a degree of half integral spin nuclei polarization saturation.

Curve **5** depicts the results achieved by way of a shunt test wherein two lead bars were secured to the stationary brass bodies of the generator assembly so as to measure the effect of shunting the field at zones of maximum field potential. As contrasted with the results of curve **2**, superimposed thereon, a statistically as well as visually significant difference is associated with the experimental results which, realistically, may be attributed to the shunting effect. The statistical study mentioned above, substantiates the distinguishable nature of the data groups resulting in curves **2** and **5**.

Curve **6** depicts the results of a test run in which the field permeability has been eliminated by removal from the test apparatus of the upper mass member and the two detector conical pole faces. The lower mass member has also been adjusted downward so as to rest on the horizontal structural element **10**. At the same time the spatial relationship between the generator assembly and the two differentially connected probes was not altered. As may be observed from the curve **6**, there occurred no change in thermal vibration of the InAs crystal lattice. The plot scatter observable during the 45 minutes wheel spin period is attributable to increased temperature gradients which developed between the probes and the respective thermocouples in the absence of the various field circuit member thermal masses.

Further experimental results are available to substantiate the heretofore stated conclusions concerning the operating characteristics of the subject apparatus. In this respect reference is made to the copending application of the present inventor entitled Method and Apparatus For Generating a Secondary Gravitational Force Field, filed Nov. 4, 1968 and bearing Ser. No. 773,051, the subject of which concerns an apparatus for establishing a time variant kinemassic force field.

It will be apparent from the foregoing description that there has been provided an apparatus for generating and transforming kinemassic forces due to a dynamic interaction field arising through gravitational coupling of relatively moving bodies. Although in its original application the kinemassic force has been applid to the reduction of thermal vibrations in the lattice structure of a crystal, it should be readily apparent that other more significant uses of these forces are contemplated. In this respect the principles of the present invention may well be applied to any system in which bodies are nonresponsive or only partially responsive to conventional forces such as electromagnetic force fields. Thus, the present invention should have particular applicability to the stabilization of plasma particles, pursuant to controlled thermal nuclear fusion, or in the governing of temperatures and thermal energies within matter.

While in accordance with the provisions of the statutes, there has been illustrated and described the best forms of the invention known, it will be apparent to those skilled in the art that changes may be made in the apparatus described without departing from the spirit of the invention as set forth in the appended claims and that in some cases, certain features of the invention may be used to advantage without a corresponding use of other features.

Having now described the invention, what is claimed as new and for which it is desired to secure Letters Patent is:

1. An energy generating and transforming apparatus comprising a first member, said first member further comprised of spin nuclei material and mounted so as to be freely rotatable about an axis located within said first member, at least one stationary member, said stationary member comprised of spin nuclei material and positioned immediately adjacent said first member, and means for effecting the rotation of said first member whereby it is effective in impressing a non-electromagnetic force onto said stationary member.

2. A method for generating a non-electromagnetic force field and for converting such force field into useful work comprising the steps of mounting a first member comprised of preferred material in a manner which enables said first member to assume a degree of relative motion with respect to a second member also comprised of preferred material, establishing a degree of relative motion between said first and said second members, and sensing the resultant energy due to the dynamic interaction of the relatively moving members.

3. The method of claim **2** wherein the sensing further comprises the steps of positioning a member of preferred material within said non-electromagnetic force field and measuring the change in the physical characteristics thereof.

4. An apparatus comprising two U-shaped members of spin nuclei material, non-spin nuclei material means for positioning said U-shaped members in mirrored relationship with one another and separated by two gaps, means including a freely rotatable member of spin nuclei material mounted in one of said two gaps, means including a detector mounted in the other one of said two gaps, and means for effecting the rotation of said freely rotatable member whereby a non-electromagnetic force is impressed upon said detector.

5. The apparatus of claim **4** wherein the detector positioned within the second of said two gaps comprises a crystalline structure of spin nuclei material such that the non-electromagnetic force impressed upon said crystalline structure is effective in polarizing said spin nuclei material sufficiently to reduce the specific heat properties of the crystalline structure so as to effect a substantial increase in the temperature thereof.

6. An energy generating apparatus comprising a first member, a second member, and means for establishing relative motion between said first and second members whereby a non-electromagnetic force is generated within said first and second members due to the dynamic interaction of said relatively moving members.

7. An energy generating and transforming apparatus comprising a mass circuit constructed of spin nuclei material of half integral spin value, said mass circuit having two gaps therein, field generating means rotatably mounted in one of said mass circuit gaps, said field generator means further comprising a frame for rotatably mounting thereon a member comprising spin nuclei material of half integral spin value, the axis of rotation of said rotatable member lying in the plane of said mass circuit, a pair of pole pieces mounted on said frame, said pole pieces being disposed on said frame on opposite sides of said rotatable member, each pole piece presenting a generally circular face in close proximity to but spaced from a face of said rotatable member, said pole pieces being further configured to substantially fill the gap in said mass circuit, means for rotating the rotatable member of said field generator means at high velocity, and means mounted in the other gap of said mass circuit for detecting a field in said circuit.

8. An energy generating and transforming apparatus comprising: a mass circuit of dense material, and having two gaps therein, mounting means for said mass circuit, said mounting means having restricted contact area with said mass circuit, field generator means rotatably mounted in one of said mass circuit gaps; said generator means further comprising a frame, a rotatable member mounted on said frame for rotation, the axis of rotation of said rotatable member lying in the plane of said mass circuit throughout all relative positions of said frame, a pair of pole pieces mounted on said frame by mounting means establishing restricted contact area between each pole piece and said frame, said pole pieces being disposed on said frame on opposite sides of said rotatable member, each pole piece presenting a generally circular face in close proximity to but spaced from a face of said rotatable member, said pole pieces being further configured to substantially fill the gap in said mass circuit, means for rotating the rotatable member of said generator means at high velocity, and means mounted in the other gap of said mass circuit for demonstrating a change in physical characteristics within said gap region due to the field generated within said mass circuit.

9. The apparatus of claim **8**, wherein said means mounted in the other gap of said mass circuit comprises a member whose atomic structure is such that it is affected by said field generated within said mass circuit.

10. A method for controlling the temperature in a crystalline structure by subjecting the crystalline structure to non-electromagnetic forces capable of altering the specific heat properties thereof, including the steps of: connecting in field series relation a mass circuit constructed of dense spin nuclei material of half integral spin value, a field generator constructed essentially of spin nuclei material having a half integral spin value and rotatably mounted in one of said mass circuit gaps, and a crystalline structure also of spin nuclei material having a half integral spin value positioned in the other of said mass circuit gaps; initiating the rotation of said field generator whereby the external angular momentum of spin nuclei material within said rotating field generator interacts with inertial space to effect the polarization of the spin nuclei thereof, resulting in turn in a net component of angular momentum which dynamically interacts with the spin nuclei material of the mass circuit thereby further polarizing the nuclei of the material therein; and concentrating the resultant field within said field series circuit onto said crystalline structure within the second of said mass circuit gaps whereby the spin nuclei material of said crystalline structure is sufficiently polarized to reduce the specific heat properties of the crystalline structure due to a reduction in degrees of freedom of the lattice vibrations of said crystalline structure thereby effecting a substantial temperature increase in the body thereof.

No references cited.

HARLAND S. SKOGQUIST, Primary Examiner

UNITED STATES PATENT OFFICE
CERTIFICATE OF CORRECTION

Patent No. 3,626,606 Dated December 14, 1971

Inventor(s) Henry W. Wallace

It is certified that error appears in the above-identified patent and that said Letters Patent are hereby corrected as shown below:

IN THE SPECIFICATION

Column 2, line 10, "According should read --Accordingly--.

Column 2, line 25, "purposes" should read --proposes--.

Column 2, line 51, "spection" should read --specification--.

Column 4, line 32, "geld" should read --field--.

Column 7, line 56, "process" should read --possess--.

Column 11, line 34, "applid" should read --applied--.

Signed and sealed this 4th day of July 1972.

(SEAL)
Attest:

EDWARD M. FLETCHER, JR.
Attesting Officer

ROBERT GOTTSCHALK
Commissioner of Patents

United States Patent [19]

Fletcher et al.

[11] **3,760,257**

[45] **Sept. 18, 1973**

---

[54] **ELECTROMAGNETIC WAVE ENERGY CONVERTER**

[76] Inventors: **James C. Fletcher,** Administrator of the National Aeronautics and Space Administration with respect to an invention of; **Robert L. Bailey,** Gainesville, Fla.

[22] Filed: **Sept. 27, 1972**

[21] Appl. No.: **292,698**

[52] **U.S. Cl.** **321/1.5,** 136/89, 250/212
[51] **Int. Cl.** **H02m**
[58] **Field of Search** 136/89; 250/212; 333/21; 321/1.5

[56] **References Cited**

UNITED STATES PATENTS

3,026,439	3/1962	Geer	250/212 X
3,121,648	2/1964	Jensen	136/89
3,232,795	2/1966	Gillette et al.	250/212 X
3,350,234	10/1967	Ule	250/212 X
3,427,200	2/1969	Lapin et al	136/89
3,448,273	6/1969	Webb	250/212 X
3,615,853	10/1971	Paine	136/89

Primary Examiner—William M. Shoop, Jr.
Attorney—R. F. Kempf et al.

[57] **ABSTRACT**

Electromagnetic wave energy is converted into electric power with an array of mutually insulated electromagnetic wave absorber elements each responsive to an electric field component of the wave as it impinges thereon. Each element includes a portion tapered in the direction of wave propagation to provide a relatively wideband response spectrum. Each element includes an output for deriving a voltage replica of the electric field variations intercepted by it. Adjacent elements are positioned relative to each other so that an electric field subsists between adjacent elements in response to the impinging wave. The electric field results in a voltage difference between adjacent elements that is fed to a rectifier to derive d.c. output power. The element pairs may be arranged in a two-dimensional array to provide power conversion of randomly polarized electromagnetic waves, such as sunlight.

19 Claims, 3 Drawing Figures

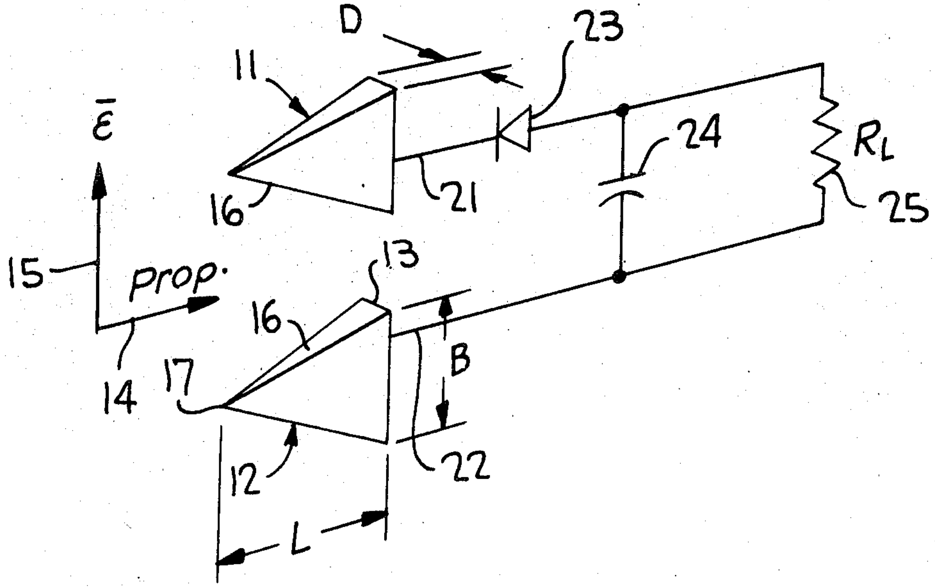

FIG. 1

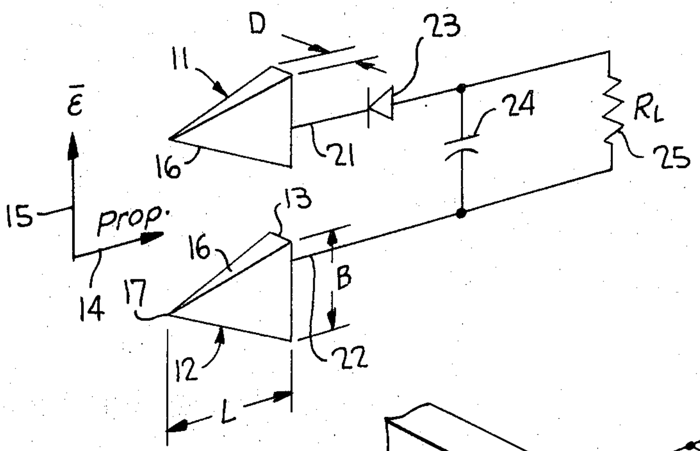
D
23
11
$\bar{\varepsilon}$
24
R_L
16
21
25
15
prop.
13
16
B
17
14
12
22
L

FIG. 2

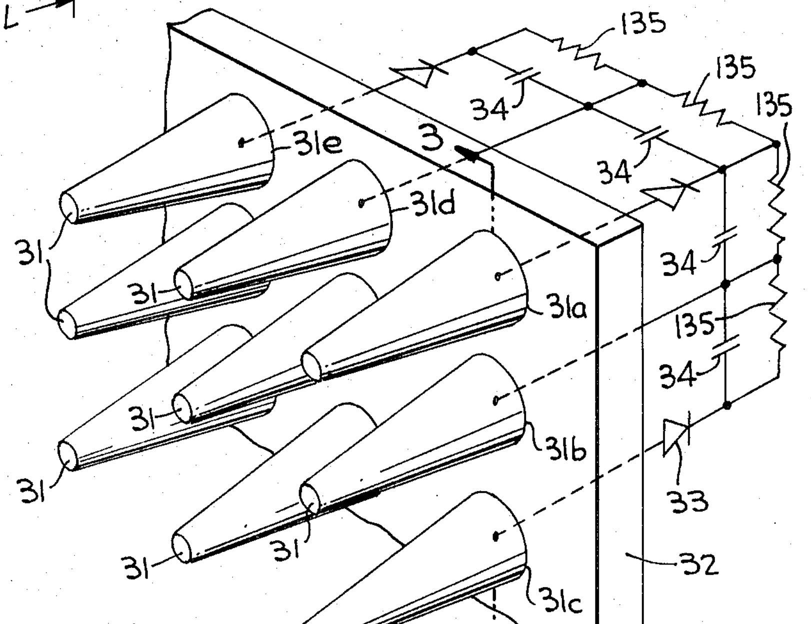
135
135
135
34
3
34
31e
31d
31
31
31
31a
34
135
34
31
31
31b
33
31
31
31
32
31c
31
3

FIG. 3

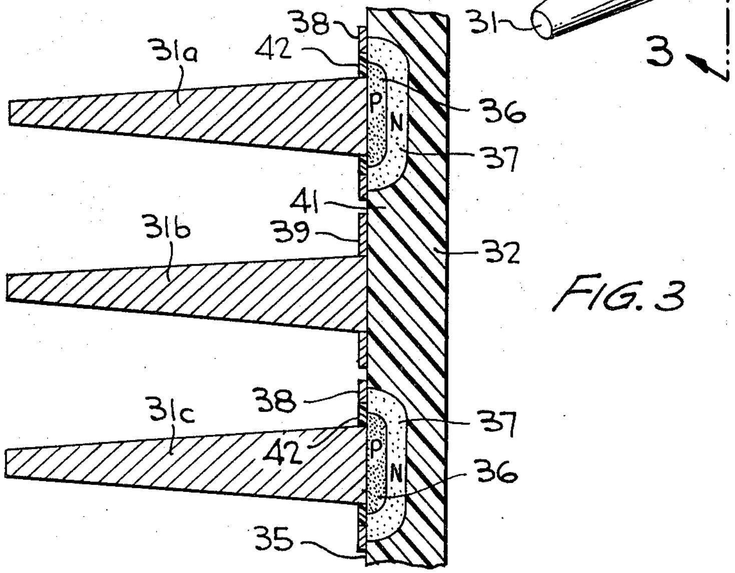
38
42
31a
P
N
36
37
41
31b
39
32
38
37
31c
42
P
N
36
35

ELECTROMAGNETIC WAVE ENERGY CONVERTER

ORIGIN OF THE INVENTION

The invention described herein was made in the performance of work under a NASA contract and is subject to the provisions of Section 305 of the National Aeronautics and Space Act of 1958, Public Law 85-568 (72 Stat. 435; 42 U.S.C. 2457).

FIELD OF INVENTION

The present invention relates generally to devices for converting electromagnetic wave energy into electric power and, more particularly, to a device including a number of relatively closely spaced electromagnetic wave absorber elements having tapered portions responsive to wide band electromagnetic wave radiation.

BACKGROUND OF THE INVENTION

Devices for converting radiant energy into usable power have been considered extensively in the past. Presently, the most commonly utilized devices for converting radiant energy into usable power are solar cells which are adapted to convert solar energy directly into electricity for power generating purposes. Solar cells generally include oppositely doped semiconductor junctions which generate current in response to solar energy impinging thereon. Solar cells depend for their operation on quantum properties of light energy to provide charge separation and hence current flow in the junction region. To date, the maximum efficiency of solar cells in converting solar energy into electric energy is approximately 13 percent. In addition to the relatively low conversion efficiency of solar energy into electric power, solar cells are very expensive, as well as fragile, and usually must be mounted on a rigid, preferably flat substrate. The requirement for a rigid, flat substrate frequently substantially increases the combined cost of the solar cells and their supporting structure; if the supporting structure is a spacecraft panel, problems of folding the panel and maintaining it in a rigid condition in use are encountered.

BRIEF DESCRIPTION OF THE INVENTION

In accordance with the present invention, the wave properties of electromagnetic energy are utilized to convert wide band electromagnetic wave energy into usable electric power. The wide band energy impinges on an array of tapered absorber elements mutually insulated from each other. The length, base area, and separation of the elements are such that relatively wide band energy can be absorbed. To achieve maximum conversion of wide band energy, each element has a length, in the direction of wave propagation, equal to several wavelengths of the longest wavelength energy to be converted. Adjacent elements have spaced, sloping converging sides to enable maximum conversion of energy wavelengths in the spectrum to be converted. A voltage replica of the electric field variations absorbed by each element is derived at an output of each element. The difference between the voltage variations derived from adjacent elements is derived and supplied to a rectifier to provide the power conversion. To maximize the voltage difference, bases of adjacent absorber elements are spaced from each other by less than one wave length and preferably less than ½ wavelength, of the received energy. The close spacing, relatively long length and tapered sides also enable the absorber elements to function effectively as absorbers and prevent substantial re-radiation and reflection of the energy impinging thereon because the wave radiation is trapped in the converging interstices between the abosrbing elements.

In a preferred embodiment, the absorbers are formed of metallic elements, such as copper or other suitable material; however, absorbers may consist of dielectric elements which function to direct the electromagnetic wave energy onto an electromagnetic wave energy-to-electric voltage converter that derives a voltage replica of the wave energy.

The invention may be utilized in conjunction with converting wave energy from the microwave region through the visible light spectrum. If the invention is utilized for wave lengths in the solar spectrum, it is ideally suited for use as a solar energy-to-electric power converter. In such a configuration, the absorber elements and electric components utilized for converting the voltage wave replicas into electric power might be formed utilizing integrated circuit type manufacturing processes.

If the device is utilized for absorbing plane polarized electromagnetic waves, adjacent elements in the direction of the wave electric field are aligned with each other in pairs such that maximum electric field variations are derived between them but adjacent elements at right angles to the electric field are packed as close as possible to each other without electric contact, to maximize the active absorber area. If the electromagnetic wave radiation is circularly or randomly polarized, as in the case of solar energy, a two-dimensional array of elements is provided. The elements of the two-dimensional array are positioned in mutually orthogonal directions having aligned columns and rows, and the elements preferably have symmetrical bases, such as a square or circle. Voltage differences are derived between adjacent pairs in both orthogonal directions to provide maximum conversion efficiency of the circularly or randomly polarized wave energy.

Because the conversion process is in response to the wave properties of the impinging electromagnetic wave energy, rather than the quantum properties of such waves, and because of the small number of loss mechanisms individually and collectively optimal, it is believed possible to achieve conversion efficiencies considerably greater than existing solar cell type devices. A major advantage of the present invention is the separation of the wave absorption means and the conversion means permitting each to be individually optimized for the incident wave electromagnetic power spectrum. Another major advantage is that by suitable choice of geometry for the absorbing elements the device can be made to match the incident radiation spectrum. There is no known means of achieving this desirable result with present art solar cells. Another advantage of the present invention is that it does not utilize temperature-sensitive active semiconductor elements, for the basic absorption process; the only semiconductor components are passive diodes. A further advantage of the present invention is that the substrate on which the elements are mounted can be mechanically flexible, to substantially eliminate many of the problems inherent with prior art fragile solar cells mounted on rigid substrates.

It is, accordingly, an object of the present invention to provide a new and improved device for converting electromagnetic wave energy into electric power.

Another object of the invention is to provide a new and improved device for converting wide band electromagnetic wave energy into electric power with relatively high efficiency, utilizing the wave properties of the energy.

An additional object of the invention is to provide a new and improved device for converting circularly or randomly polarized electromagnetic wave energy into electric power in response to electric field gradients.

A further objet of the invention is to provide a relatively inexpensive device for converting electromagnetic wave energy into electric power, which device is relatively temperature-insensitive and can be mounted on a flexible substrate.

An additional object is to provide a new and improved integrated circuit type device for converting solar energy into electric power.

An additional object is to provide a new and improved means of matching the response of the device to the incident electromagnetic wave spectrum.

The above and still further objects, features and advantages of the present invention will become apparent upon consideration of the following detailed description of several specific embodiments thereof, especially when taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWING

FIG. 1 is a perspective view schematically illustrating the principles of the present invention;

FIG. 2 is a perspective diagram illustrating an embodiment of the invention particularly adapted for converting solar energy into electric power; and

FIG. 3 is a side sectional view of the embodiment illustrated in FIG. 2, along the lines 3—3.

DETAILED DESCRIPTION OF THE DRAWING

Reference is now made to FIG. 1 of the drawing wherein is illustrated a pair of mutually insulated, substantially identical, pyramidal or conical, aligned and relatively closely packed, preferably metallic, as copper or other suitable material, electromagnetic wave absorber elements **11** and **12** which are responsive to plane polarized wide band electromagnetic wave energy that is propagating in a direction indicated by arrow **14** and has an electric field with a gradient in a direction indicated by arrow **15.** Elements **11** and **12** are constructed so that the shapes of every cross section parallel to their bases **13** are similar; the cross-sectional areas decrease for increasing distances from the bases **13** of the elements. Each of elements **11** and **12** includes a co-planar base **13**, which may have any desired configuration, such as a rectangle, square or circle, that lies in a plane at right angles to the direction of wave propagation and is parallel to the electric field gradient **15.** Elements **11** and **12** have a substantial height, L, or length in the direction of wave energy propagation, between their base **13** and apex **17.** The height is generally several (up to approximately seventy five) wave lengths of the longest wavelength of the incident electromagnetic energy. Elements **11** and **12** include adjacent, tapered side surfaces **16** that converge toward base **13** and are positioned relative to each other to intercept the electric field **15** and absorb energy over a wide band of the electromagnetic wave. The apex of absorber elements **11** and **12** may be truncated in which case L is the length of the truncated absorber elements.

Voltage variations derived in elements **11** and **12** in response to the electric field variations of the electromagnetic wave energy impinging on surfaces **16** are transduced or converted into an electric voltage at terminals **21** and **22** of the elements. The voltages at terminals **21** and **22** are replicas of the electric field variations impinging on absorber elements **11** and **12,** whereby the voltage difference between terminals **21** and **22** is proportional to the electric field gradient established by the plane polarized wave between facing surfaces **16** of elements **11** and **12.**

The centerline spacing between adjacent elements **11** and **12** is such that for the incident electromagnetic wave spectrum, maximum voltage between **21** and **22** occurs. It is of the order of a wavelength or less of the incident electromagnetic energy. The present invention responds to this voltage difference as derived over the desired spectrum of the incident wave energy, to convert the wave energy into electric power.

The voltage difference between terminals **21** and **22** is converted into useful d.c. electric power by a rectifier **23** and a filter capacitor **24,** the voltage across which is supplied to a suitable external load, such as resistor **25.** The cathode is connected between the anode of the rectifier diode and terminal **22.** It is to be understood that a full wave rectifier can be employed and that the d.c. voltage developed across external load **25** can be supplied to any suitable device, such as a d.c.-to-a.c. power frequency device.

The breadth,B, and depth,D, of base **13** as well as the length, L, and spacing between the elements **11** and **12** enable the band width to which the elements **11** and **12** are responsive to be adjusted to provide optimum matching to the spectrum of the electromagnetic wave energy impinging on absorber elements **11** and **12.** Close spacing between adjacent elements **11** and **12,** in addition to enabling the electric field gradient to be coupled with the greatest voltage difference to terminals **21** and **22,** enables the electric field to be effectively trapped in the converging interstices between facing, adjacent elements **11** and **12.** Trapping of the electric field variations occurs because elements **11** and **12** are absorptive to the electric electromagnetic wave energy, whereby the energy is not reflected and re-radiated from the elements.

Because the device of FIG. **1** is responsive to a plane polarized electromagnetic wave, it is preferable for the depth, D, of each of the elements **11** and **12** to be relatively narrow, less than ¼ wave length, so that adjacent elements (not shown) can be packed as closely as possible to either side of elements **11** and **12** in the horizontal direction. Close packing of the elements enables maximum conversion of the electromagnetic wave energy into electric power over the entire area of an array which may be fabricated out of a multiplicity of element pairs as illustrated in FIG. **1.** It is to be noted that the device of FIG. **1** is responsive most efficiently to electromagnetic waves polarized so that the E field is in a plane parallel or coplanar with the long dimension of bases **13.** If the electromagnetic wave polarization direction were rotated 90°, there would be no substantial electric field variation between the adjacent elements **11** and **12** and there would be substantially zero

voltage developed between terminals **21** and **22** with a resulting zero conversion of electromagnetic wave energy into electric power.

The power of the electromagnetic wave incident on absorber elements **11** and **12** is the incident power density times the effective area of the two elements and can thereby be stated approximately as: 2 $(\bar{E}\times\bar{H})$ BD, where $\bar{E}$ and $\bar{H}$ are respectively the electric and magnetic fields of the wave energy. The power indicated by the equation is available so that it can be converted into electric power supplied to a load, except for losses in elements **11** and **12**, rectification, and stray losses. The losses of elements **11** and **12** are principally due to skin effect or di-electric losses, while the rectifier losses are the series resistance of the diode and stray losses are of the capacitor, and supporting substrate for elements **11** and **12**. Stray losses can be minimized by locating elements **11** and **12** on a low loss dielectric substrate to minimize loss currents between the abosrber terminals.

The maximum wavelength restriction (about 75λ of the longest wavelength of the received spectrum) is provided to enable the shortest wavelength of the desired spectrum to be trapped between adjacent elements and to provide a relatively large surface area for extracting power from it. If the maximum wavelength restriction, which corresponds approximately with the length of a cone in an eyeball of a mammal (about 75λ of the longest wavelength of the received spectrum), is exceeded by the incident electromagnetic energy spectrum, then the voltage output between terminals **21–22** of the device decreases.

If the device of FIG. **1** is utilized to convert plane polarized microwave electromagnetic wave energy into electric power, the rectifier circuit can be connected to the terminals by a coaxial cable having a center conductor connected directly to terminal **21** and a shielded outer conductor, possibly with a conventional balun, connected to terminal **22**. In the alternative, the rectifier circuit can be connected to elements **11** and **12** by a wave guide excited by the voltage difference between terminals **21** and **22** so that propagation in the wave guide is in one of the transverse electric modes, such as TE_{01}. In such a configuration, upper and lower conducting surfaces of a rectangular wave guide, between which the electric field is developed, are respectively connected to terminals **21** and **22**.

One microwave absorber actually constructed and tested in accordance with the present invention includes a pair of sheet copper pyramidal elements **11** and **12** having a length (L) of 13.3 centimeters, a breadth (B) of 6.3 centimeters, and a depth (D) of 2.0 centimeters. This absorber pair had a center frequency of approximately 475 Mhz and a pass band between approximately 200 and 700 Mhz. Increasing the length of the abosrber elements, without any changes in the base dimensions of the pyramid, lowered the pass band center frequency, as well as the upper and lower cut-off frequencies.

The basic converter structure of FIG. **1** can be modified, as illustrated in FIG. **2**, to respond to circularly polarized electromagnetic waves or randomly polarized electromagnetic waves, as subsist in solar energy. In FIG. **2**, a multiplicity of absorber elements **31** are arranged in a two-dimensional array of aligned orthogonally directed columns and rows. Elements **31** are mutually insulated from each other and mounted on a dielectric substrate **32**, which is preferably flexible but may be rigid. Each of elements **31** preferably has a symmetrical base and cross section parallel to the base, which may be either circular or square, to enable equal voltages to be derived between adjacent ones of elements **31** in the two orthogonal directions in response to orthogonal electric field components at the same wavelength. Thereby, the total array power output of FIG. **2** is insensitive to the polarization direction of the energy impinging thereon and the array can respond to circularly polarized, as well as randomly polarized electromagnetic wave energy.

To convert the electromagnetic wave energy impinging on elements **31** into electric power, adjacent elements are interconnected with each other by a half wave rectifying network similar to that illustrated in FIG. **1**. In particular, aligned vertical elements **31*a***, **31*b***, and **31*c*** are interconnected so that the anodes of rectifier diodes **33**, connected to elements **31*a*** and **31*c*** (which are separated from each other by element **31*b***) are connected directly to the elements, while the cathodes of the rectifier diodes are connected to one electrode of different capacitors **34**, the other electrodes which have a common connection to element **31*b***. Similarly, horizontally aligned elements **31*a***, **31*d***, and **31*e*** are connected to rectifier circuits such that the center element of the triad has a common connection to a pair of capacitors **34** which are connected to the cathodes of diodes **33**, having anodes connected to be responsive to the voltage replicas respectively derived by absorber element pairs **31*a*–31*d*** and **31*e*–31*d***. D. C. voltages developed across capacitors **34** are supplied to a matrix of load resistors **135**, one of which is connected across each of the capacitors.

Elements **31** of FIG. **2** are illustrated as frustoconical structures having upper bases with considerably smaller areas than the bases of the elements that are secured to substrate **32**. The frusto-conical configuration is preferred in certain instances because sharp points, as illustrated in the pyramidal elements **11** and **12**, FIG. **1**, may have a tendency to fracture.

To enable the structure illustrated in FIG. **2** to function as a device for converting solar electromagnetic wave energy, which has the majority of its power in the spectrum from 0.3 to 1.1 microns, into d.c. power, the element dimensions and inter-element spacing must be on the order of a micron, whereby integrated circuit techniques are preferably employed in fabrication. To this end, a solar energy converter in accordance with the present invention may take the form illustrated by the cross-sectional view illustrated in FIG. **3** but not limited to this embodiment. In FIG. **3**, electrically conductive cones **31*a***, **31*b***, and **31*c*** project from surface **35** of substrate **32**. Alternate ones of cones **31**, such as cones **31*a*** and **31*c***, are ohmically connected to a P-doped regions **36** of substrate **32**, which are formed on surface **35** of substrate **32** in super-position with the bases of cones **31*a*** and **31*c***. A junction is formed between P-doped regions **36** and N-doped regions **37** which are formed in substrate **32** beneath P-doped regions **36**. Regions **36** and **37** respectively correspond with the anode and cathode of diodes **33** connected to cones **31*a*** and **31*c***, FIG. **2**. Annular, dielectric oxide films **42** are formed on face **35** over the intersections between the otherwise exposed portions of the junctions between regions **36** and **37**. Oxide films **42** cover the peripheries of P-doped regions **36** and the inner circumference of N-doped regions **37**. On the portion of

P-doped regions **36** not covered by oxide films **42** and the inner radial portion of the films, metal elements **31***a* and **31***c* are placed. The bases of elements **31***a* and **31***c* are ohmically connected to P-doped regions **36.** Annular, metal films **38** are formed on face **35** in superposition with the otherwise exposed regions **37** to form ohmic contacts with the N-doped regions. Metal films **38** are electrically insulated from elements **31***a* and **31***c* by dielectric films **42** which are disposed between metal films and elements. Metal film **39** is formed on face **35** at the base of element **31***b* and is electrically connected to the base. The areas **41** of substrate **32** between adjacent metal films **38** and **39** form dielectrics for capacitors **34,** having electrodes comprised of the metal film portions **38** and **39.** A d.c. load is connected between metal film portions **38** and **39** so that the film portions form load terminals.

The structure of FIG. 3 is formed by diffusing P and N regions **36** and **37** onto surface **35** of substrate **32,** utilizing conventional integrated circuit techniques. Regions **36** and **37** are diffused only onto the areas beneath and slightly to the sides of where cones **31***a* and **31***d* are to be located. Thereafter, annular oxide regions **42** are formed over the exterior portion of P-doped region **36** and over a slight segment of the interior of N-doped region **37** on surface **35.** After oxide regions **42** have been formed, thin metal film contacts **38** are vacuum vapor deposited on the exposed portions of N regions **37.** Simultaneously with the formation of contacts **38** thin, metal film, annular contact **39** is vacuum vapor deposited on surface **35** in the region outside of cone **31***b*. After metal films **38** and **39** have been deposited, cones **31***a*, **31***b*, and **31***d* are suitably attached.

While metal is believed preferable in most instances for the absorber elements, it is to be understood that dielectric elements can also be utilized as elements for directing electromagnetic wave radiation to an optical to d.c. converter diode. The dielectric elements would preferably have the same configuration and dimensions as indicated supra while the optical to d.c. converter diode may be of a type described by Javin in the IEEE Spectrum, October, 1971, page 91.

While there have been described and illustrated several specific embodiments of the invention, it will be clear that variations in the details of the embodiments specifically illustrated and described may be made without departing from the true spirit and scope of the invention as defined in the appended claims.

I claim:

1. A device for converting electromagnetic wave energy having an electric field into electric power comprising a plurality of mutually insulated electromagnetic wave absorber elements each responsive to the electric field, adjacent ones of said elements converging toward each other in the direction of the wave propagation, each element including means for deriving a voltage replica of variations of the electric field intercepted by the element, adjacent ones of said elements being positioned relative to each other so that an electric field subsists between said adjacent elements, and means responsive to the voltage difference between the derived voltage replicas between two adjacent elements.

2. The device of claim **1** wherein each of the absorber elements is metallic.

3. The device of claim **1** where each of the absorber elements comprises a metallic surface and has a length in the direction of wave propagation several wavelengths of the impinging electromagnetic wave.

4. The device of claim **3** wherein the adjacent elements have substantially coplanar bases, and the spacing between the adjacent elements at the bases thereof is of the order of no more than a wavelength of the electromagnetic energy.

5. The device of claim **1** wherein the adjacent elements have substantially coplanar bases, and the spacing between the adjacent elements at the bases thereof is of the order of no more than a wavelength of the electromagnetic energy.

6. The device of claim **1** wherein the conversion means responsive to the voltage difference includes means for rectifying the voltage difference to derive a d.c. voltage.

7. A device for converting circularly polarized or randomly polarized electromagnetic wave energy having an electric field into electric power comprising a two-dimensional array of mutually insulated electromagnetic wave absorber elements each responsive to a component of the electric field, each elements converging toward each other in the direction of propagation of the wave and including means for deriving a voltage replica of variations of the electric field intercepted by the element, adjacent ones of said elements being positioned relative to each other so that an electric field subsists between said adjacent elements, and means responsive to the voltage difference between the voltage replicas derived between two adjacent elements, said pairs being spaced at right angles to each other in the array.

8. The device of claim **7** wherein each of the absorber elements is metallic.

9. The device of claim **7** where each of the absorber elements comprises a metallic surface and has a length in the direction of wave propagation several wavelengths of the electromagnetic wave.

10. The device of claim **9** wherein the adjacent elements have substantially co-planar bases and the spacing between the adjacent elements at the bases thereof of the order of one wavelength or less of the electromagnetic energy.

11. The device of claim **7** wherein the adjacent elements have substantially coplanar bases and the spacing between the adjacent elements at the bases thereof is of the order of one wavelength or less of the electromagnetic energy.

12. A device for converting solar, randomly polarized electromagnetic wave energy in the wave length band from approximately 0.3 to 1.1 microns into electric power comprising a dielectric substrate having a face substantially at right angles to the direction of propagation of the solar electromagnetic wave energy, a plurality of mutually insulated electromagnetic wave, metallic absorber elements mounted on said face, each of said elements including tapered portion extending away from the face in the direction of propagation of the wave energy, each of said elements having a decreasing cross-sectional area as the distance from the face increases and a length in the direction of propagation of the wave several wavelengths of the shortest wavelength of the spectrum, an integrated circuit diode in said substrate connected to be responsive to wave energy intercepted by alternately spaced ones of said

elements, a metallic film on said substrate ohmically connected to the remaining metallic elements, whereby a capacitor is formed in the substrate dielectric between adjacent ones of said metallic films.

13. The device of claim **12** wherein each of the elements includes a base with a symmetrical geometry in contact with said face, each element having a cross section parallel to the base similar to the base.

14. The device of claim **13** wherein said face has a square cross section.

15. The device of claim **13** wherein said face has a circular cross section.

16. The device of claim **13** wherein each of said elements is a frustum.

17. The device of claim **12** wherein the spacing between adjacent ones of said elements is of the order of one wavelength or less of the shortest wavelength of the spectrum.

18. The device of claim **12** wherein the several wavelengths are between 1.25 and 75.

19. A device for converting solar, randomly polarized electromagnetic wave energy in the wave length band from approximately 0.3 to 1.1 microns into electric power comprising a dielectric substrate having a face substantially at right angles to the direction of propagation of the solar electromagnetic wave energy, a plurality of mutually insulated electromagnetic wave, metallic absorber elements mounted on said face, each of said elements including a tapered portion extending away from the face in the direction of propagation of the wave energy, each of said elements having a decreasing cross-sectional area as the distance from the face increases, adjacent ones of said elements being spaced from each other by a distance on the order of a wavelength or less of said energy, an integrated circuit diode in said substrate connected to be responsive to wave energy intercepted by alternately spaced ones of said elements, a first metallic film on said substrate ohmically connected to one electrode of each of said diodes, a second metallic film on said substrate ohmically connected to the remaining metallic elements, said first and second films including load terminals whereby a capacitor is formed in the substrate dielectric between adjacent ones of said metallic films, and said load terminals.

* * * * *

United States Patent [19]
Johnson

[11] **Patent Number: 4,596,119**
[45] **Date of Patent: Jun. 24, 1986**

[54] **COMPRESSED AIR PROPULSION SYSTEM FOR A VEHICLE**

[75] Inventor: **John M. Johnson,** Amboy, Ind.

[73] Assignee: **Earl L. Alderfer,** Fort Wayne, Ind.

[21] Appl. No.: **556,085**

[22] Filed: **Nov. 29, 1983**

[51] **Int. Cl.**⁴ **F15B 11/06**
[52] **U.S. Cl.** **60/407;** 60/412; 91/275; 180/165; 180/302
[58] **Field of Search** 60/370, 407, 412; 180/165, 302; 91/275

[56] **References Cited**

U.S. PATENT DOCUMENTS

992,661	5/1911	Hautier .	
1,380,795	6/1921	Graziano .	
1,688,591	10/1928	Mitton .	
1,831,983	11/1931	Wheeler .	
1,932,698	10/1933	Jose .	
3,036,253	5/1962	Bramley	91/275
3,379,008	4/1968	Manganard .	
3,666,038	5/1972	Hudspeth et al. .	
3,688,859	9/1972	Hudspeth et al. .	
3,765,180	10/1973	Brown .	
3,847,058	11/1974	Manor	60/370
3,913,699	10/1975	Dyer .	
3,925,984	12/1975	Holleyman .	
3,980,152	9/1976	Manor .	
3,986,575	10/1976	Eggmann .	
4,018,050	4/1977	Murphy .	
4,292,804	10/1981	Rogers, Sr.	60/407
4,337,842	7/1982	Spangler et al. .	
4,355,508	10/1982	Blenke et al.	60/416
4,361,204	11/1982	Earle .	
4,383,589	5/1983	Fox	60/407
4,404,800	9/1983	Penney	60/370

FOREIGN PATENT DOCUMENTS

2326403	12/1974	Fed. Rep. of Germany .
WO82/00615	3/1982	PCT Int'l Appl. .

Primary Examiner—Robert E. Garrett
Assistant Examiner—John M. Husar
Attorney, Agent, or Firm—Jeffers, Irish & Hoffman

[57] **ABSTRACT**

A compressed air propulsion system is provided to be adapted for use with a vehicle, and comprises a transaxle mechanism adapted to be mounted to and driven by the vehicle wheels, and an air compressor coupled to the output shaft of the transaxle mechanism to be driven thereby. The air compressor supplies a flow of compressed air to an air distributor system connected to the vehicle engine. The air distributor system comprises a plurality of solenoids connected to the air lines leading to the engine cylinders, and a respective plurality of microswitches are electrically connected to the solenoids to sequentially operate the solenoids for sequentially supplying compressed air to the engine cylinders. The microswitches are sequentially actuated mechanically by a rotor secured to the distributor shaft, which is coupled to the vehicle crankshaft. An accelerator is also provided for controllably varying the flow rate of compressed air to the air distributor system.

7 Claims, 13 Drawing Figures

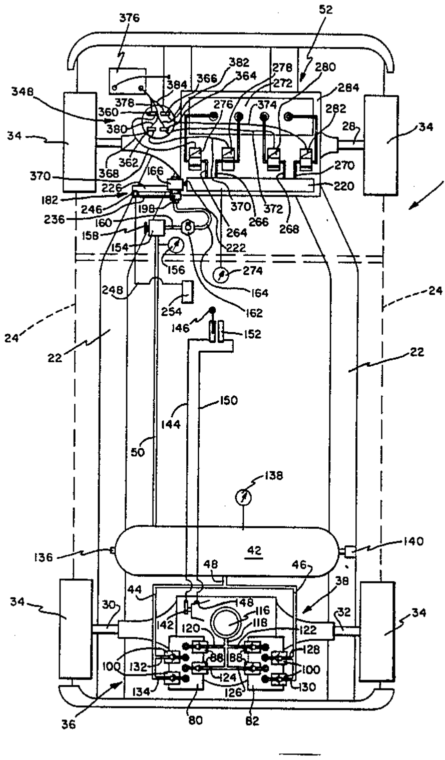
376
52
348
382
366
364
278
280
384
272
378
276
374
284
360
282
34
380
28
34
362
370
368
166
270
20
226
220
182
246
198
370
372
236
160
266
268
158
264
154
222
274
156
248
164
24
254
162
24
146
152
22
22
150
144
50
138
136
42
140
48
46
44
38
148
34
30
116
34
142
120
118
122
32
128
88
88
100
132
124
100
134
126
130
36
80
82

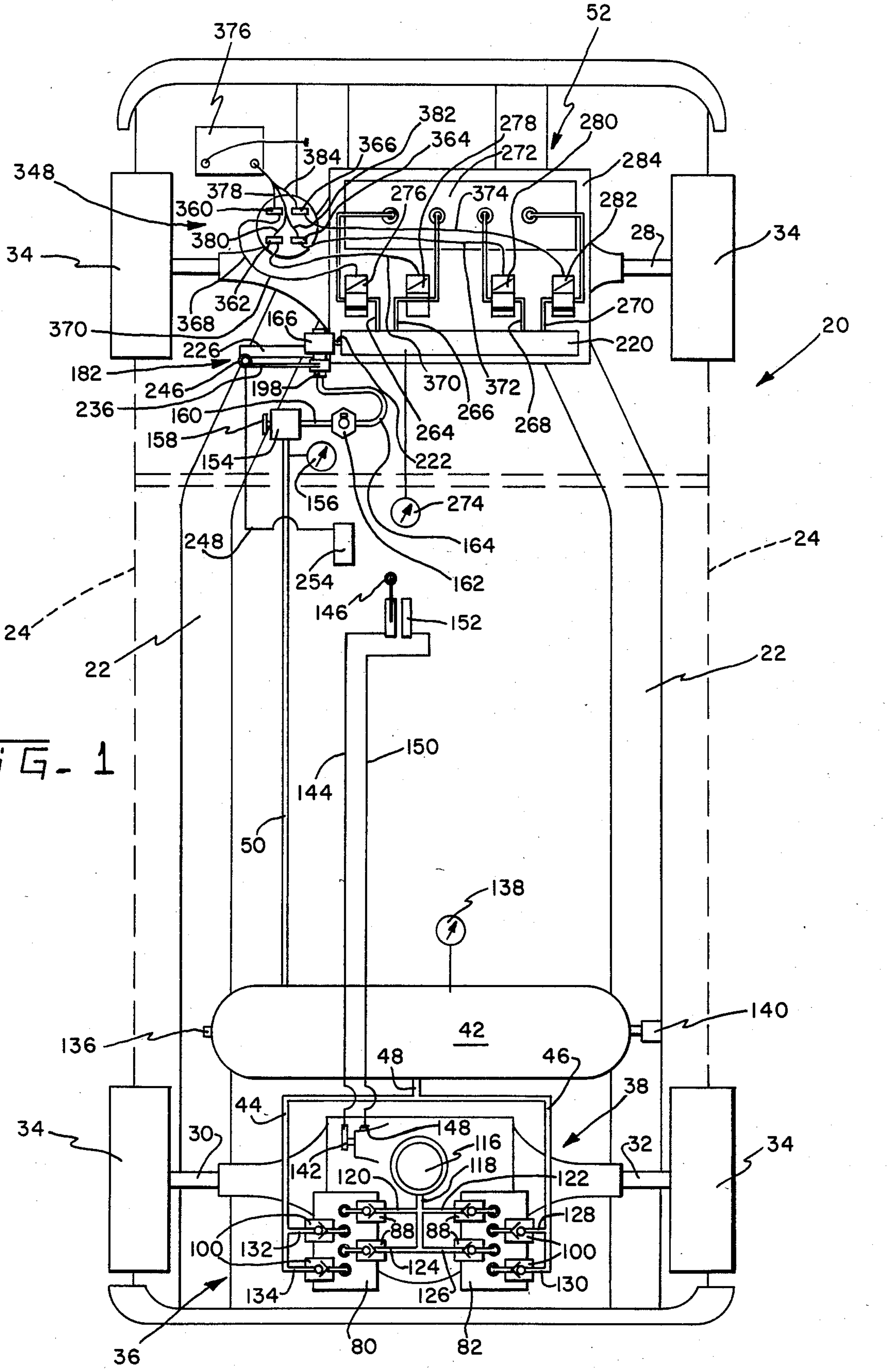

FIG. 1

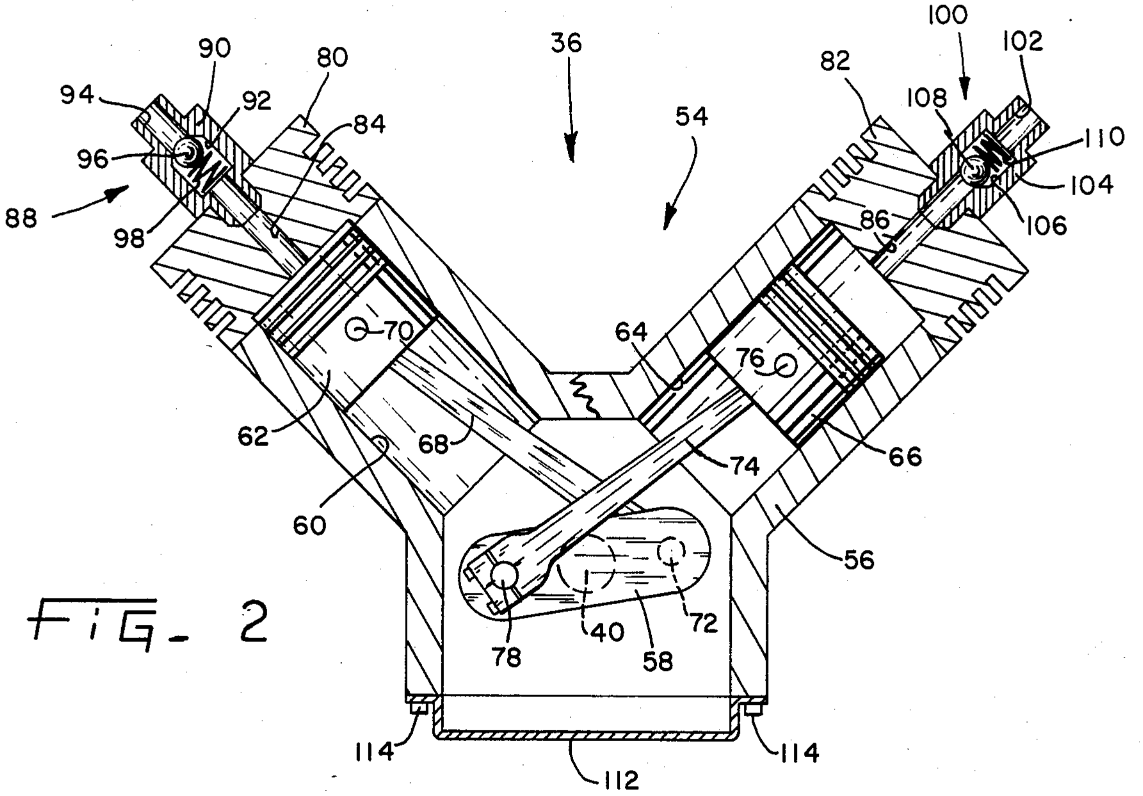

FIG. 2

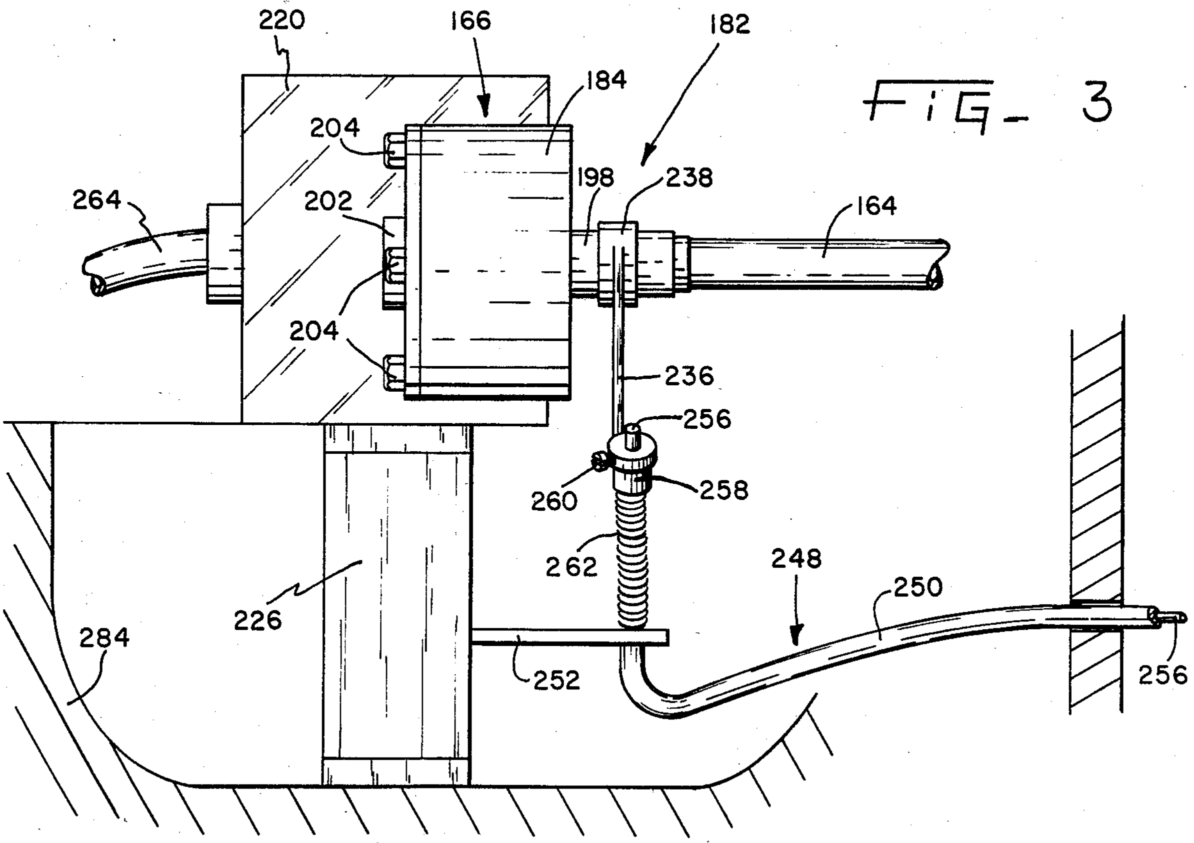

FIG. 3

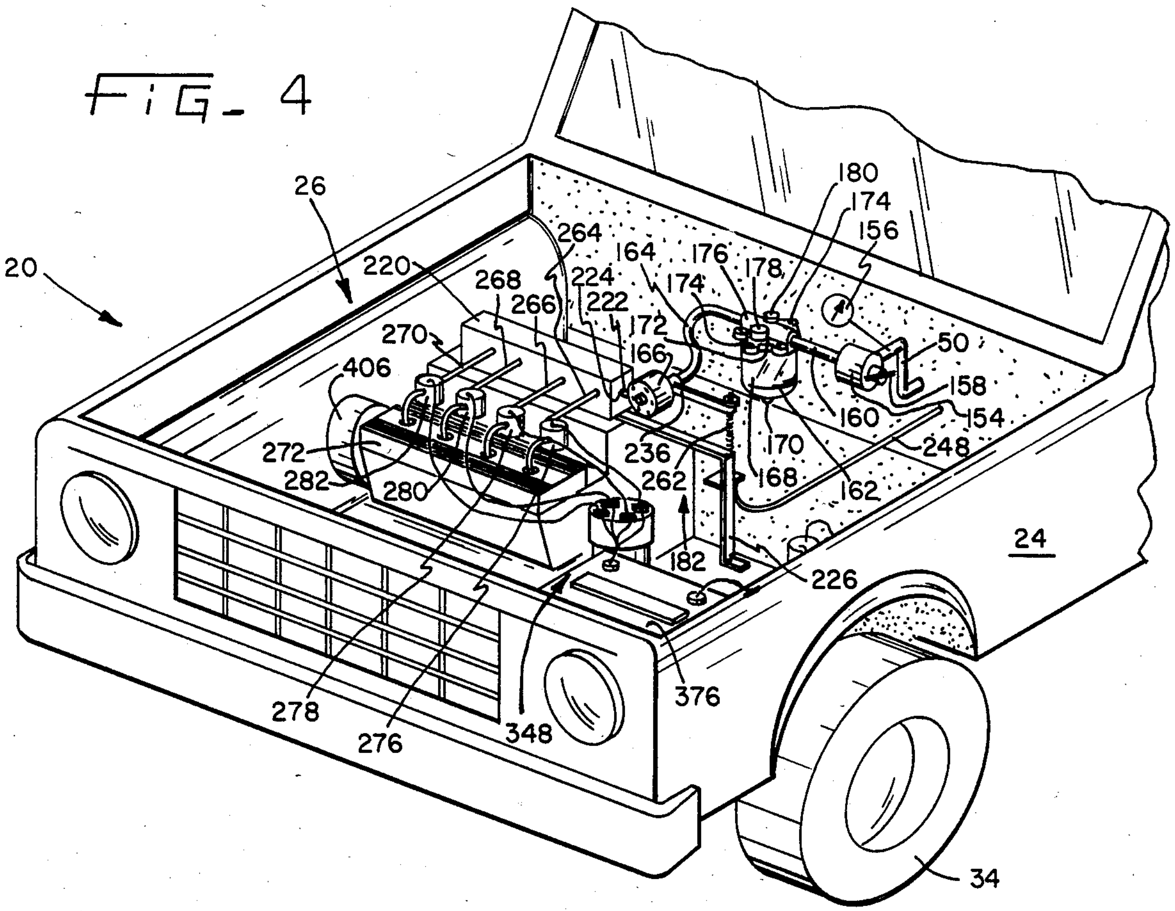
FIG. 4

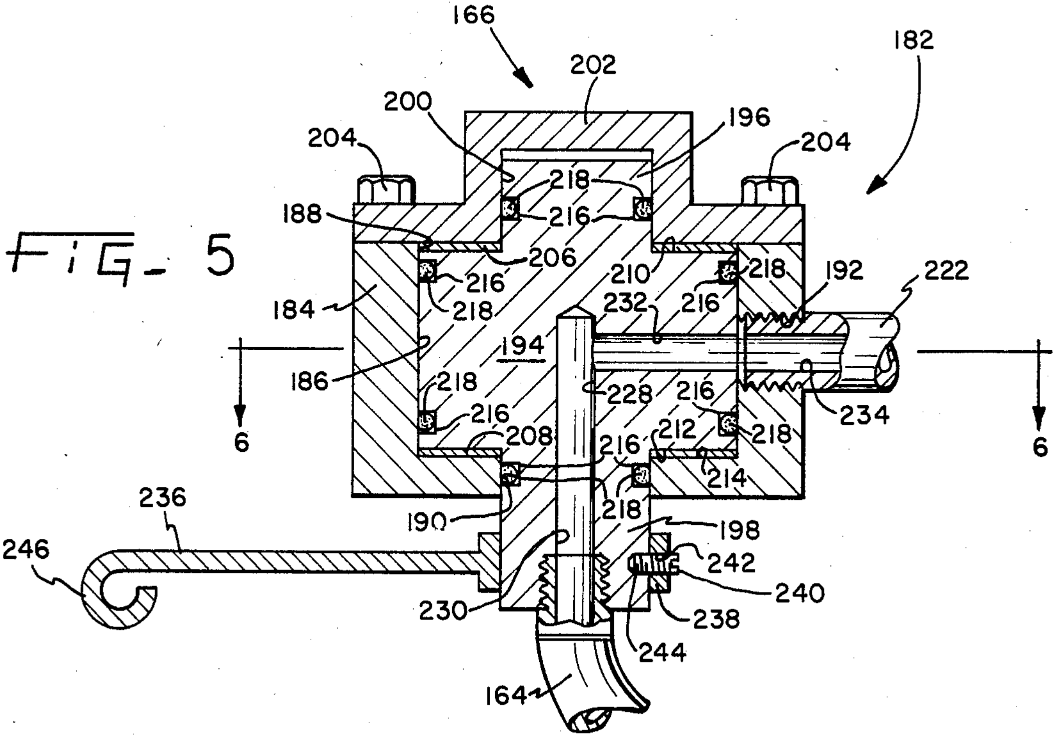
FIG. 5

FIG. 6

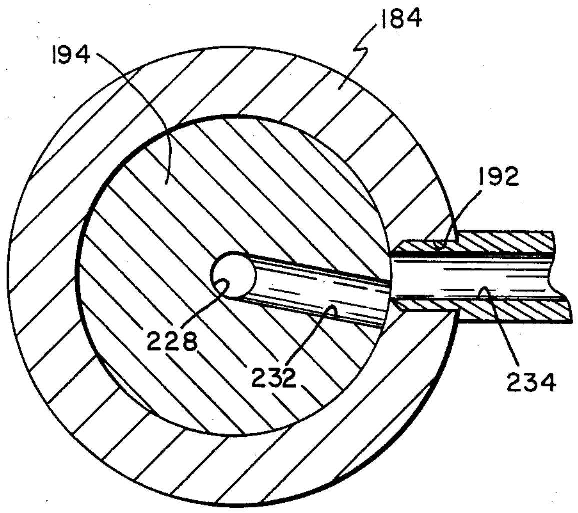

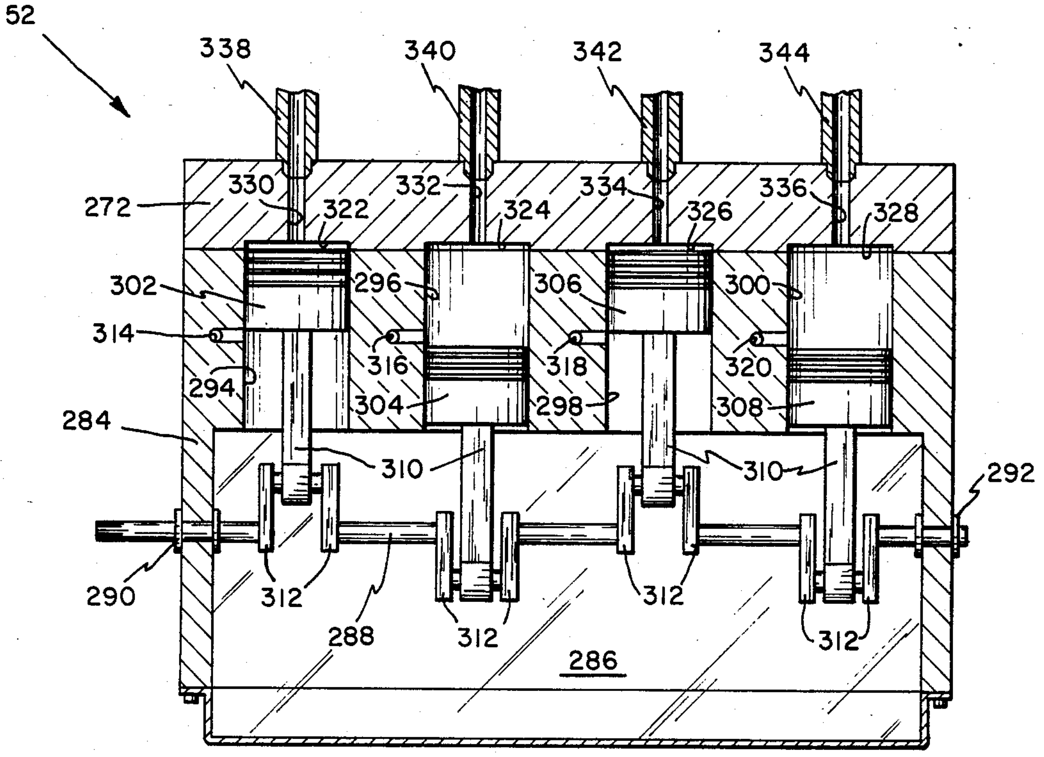

FIG. 7

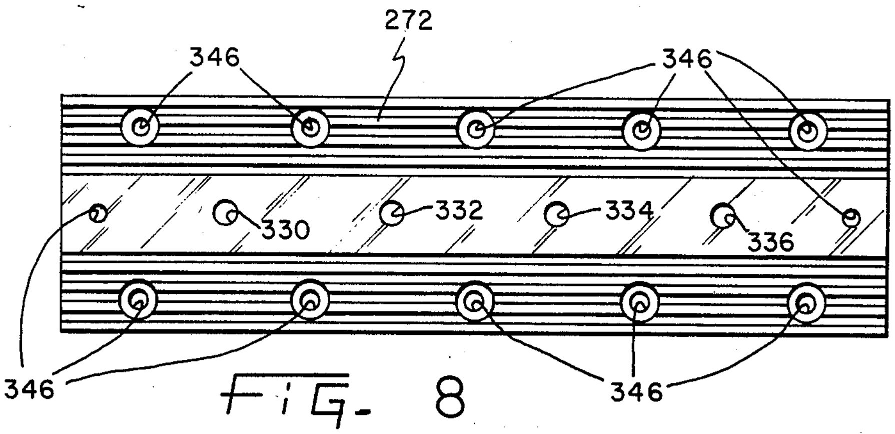
272
346
346
330
332
334
336
346
346

FIG. 8

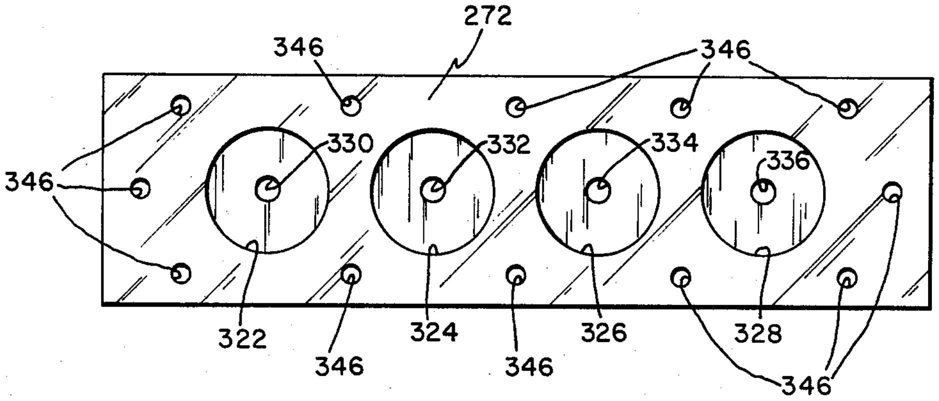
272
346
346
346
330
332
334
336
322
324
326
328
346
346
346

FIG. 9

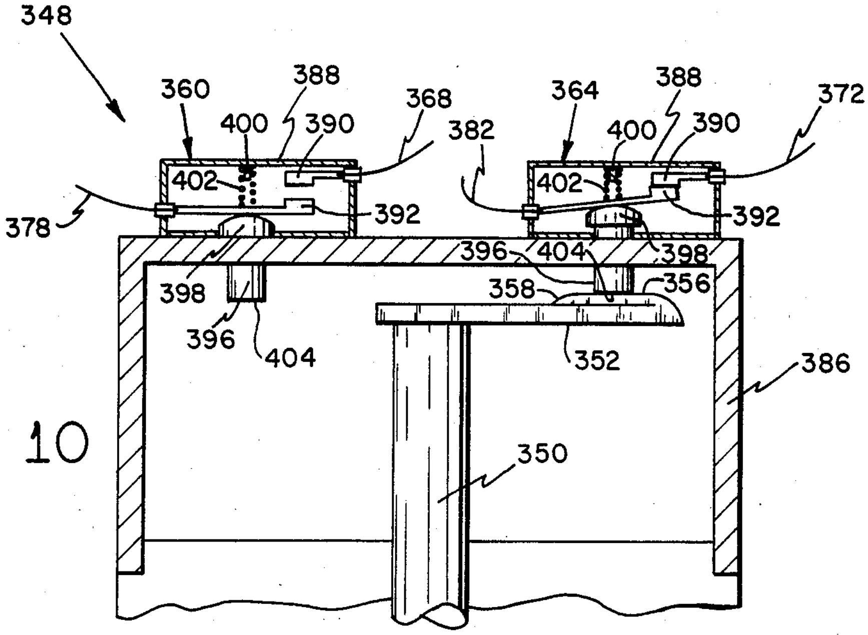
348
360
388
400
390
368
382
364
388
400
390
372
402
402
378
392
392
398
396
404
398
396
404
358
356
352
386
350

FIG. 10

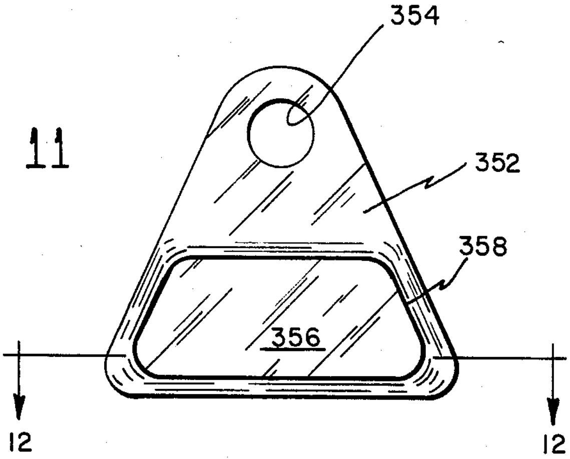
354
352
358
356
12
12

FIG. 11

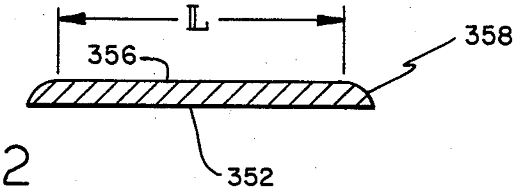
L
356
358
352

FIG. 12

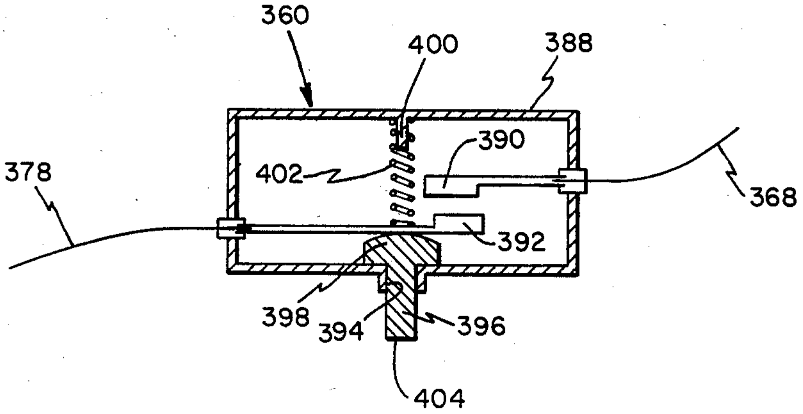
360
400
388
390
378
402
368
392
398
394
396
404

FIG. 13

COMPRESSED AIR PROPULSION SYSTEM FOR A VEHICLE

BACKGROUND OF THE INVENTION

This invention pertains to an air propulsion system, and more particularly to an improved compressed air propulsion system for a vehicle.

Many compressed air driven vehicles have been devised in attempting to provide a means of transportation that is inexpensive and pollution free. Most of these vehicles include a conventional engine having a piston-cylinder arrangement connected to a crankshaft, and a conventional, yet complex, valve assembly including camlifters, rocker arms, springs, and the like connected to the crankshaft or camshaft for supplying compressed air to the cylinders to reciprocate the pistons therein. Compressed air is delivered to the cylinders through air lines leading from the supply of compressed air and connected to the intake valves, and is then exhausted through the exhaust valves in the conventional manner.

One of the primary disadvantages with this type of compressed air driven engine is the continued use of the conventional valve assembly operated by the complex arrangement of camlifters, rocker arms, springs, and the like. In any mechanical apparatus, the more moving parts or elements required for the operation of the apparatus, the more complex it becomes and the more susceptible it is to mechanical or material failure.

In still other different types of compressed air driven vehicles, solenoid valves are connected to the air lines leading to the cylinders and are electrically operated to supply and terminate a flow of compressed air. The electrical circuitry operating the solenoids is in someway mechanically related to the rotation of the crankshaft or camshaft. For example, the reciprocative motion of the rocker arms driven by the crankshaft or camshaft can be adapted to open and close the electrical circuitry between the battery and solenoid valves, thereby supplying and terminating the flow of compressed air to the cylinders.

Although the use of solenoids increases the efficiency in delivering compressed air to the cylinders, operation of the solenoids by the conventional arrangement of valves, camlifters, rocker arms, springs, and the like still poses the problem of a greater probability of mechanical or material failure.

Compressed air can be supplied to the engine in a number of ways, one of which is to provide an air compressor driven by a battery and mounted on the vehicle. Naturally, the major drawback associated with this type of battery-driven air compressor is the fact that the battery will eventually require recharging or replacing, either of which may occur at a most undesirable time and place.

Another prior art method of supplying compressed air is to mount both an air compressor and a combustion engine on the vehicle frame. The fuel-driven combustion engine is connected by its crankshaft to the air compressor for the operation thereof. Again, the primary disadvantage with this type of air compressor is the undesirable use of the polluting fuel-driven combustion engine. Any use of a combustion engine defeats the purpose of a pollution-free compressed air driven vehicle.

Still another method of supplying a source of compressed air is to provide a plurality of rechargeable compressed air tanks mounted on the vehicle. The major problem associated with this type of air compressor is similar to that of the battery-operated air compressor in that continual monitoring of the system is required, i.e., the operator must continually monitor the pressure available to insure proper operation of the air driven engine. Furthermore, using rechargeable air tanks presents the possibility of the tanks requiring recharging when a source of compressed air is not available.

SUMMARY OF THE INVENTION

The present invention overcomes the problems and disadvantages of the prior art by providing an improved compressed air propulsion system for use with a vehicle.

The compressed air propulsion system of the present invention is used in combination with an engine having a piston-cylinder arrangement, and eliminates the complex arrangement of camlifters, rocker arms, springs, and the like interconnecting the valve assembly of the cylinders with the crankshaft or camshaft. The head mounted on the cylinders in the engine block has only one opening therein for each cylinder, and a single air line is connected to each opening for supplying compressed air to the respective cylinder. The respective air lines leading from the cylinders are joined to a common manifold to which the air line leading from the source of compressed air is connected. The flow of compressed air to the cylinders is controlled by a uniquely constructed air distributor system which utilizes the rotation of the distributor shaft, which is operatively coupled to the engine crankshaft or camshaft, for sequentially actuating solenoids connected to the respective air lines leading to the cylinders.

A rotor mounted on the end portion of the rotating distributor shaft sequentially and mechanically actuates a plurality of electrical microswitches, which are electrically connected to the respective solenoids. The microswitches are arranged in a generally circular array about the rotor, and upon rotation of the rotor, each microswitch is individually mechanically actuated to close a pair of spaced-apart electrical conductors therein, thereby providing current from the battery through the microswitch to its respective solenoid. The surface or contact length of the rotor, with which each microswitch is engageable, determines the length of time a microswitch will remain closed to supply current to the solenoid, thereby controlling the amount of compressed air delivered through the solenoid to its respective cylinder.

By eliminating the complex arrangement of camlifters, rocker arms, springs, and the like interconnecting the valves and camshaft or crankshaft, the air propulsion system of the present invention completely eliminates problems due to their material or mechanical failure. Moreover, in place thereof, the air propulsion system of the present invention provides a unique air distributor system for distributing the compressed air to the cylinders, and which is much less complex in construction and operation.

The air propulsion system of the present invention further includes a source of compressed air that is run neither by batteries, a combustion engine, or solely by rechargeable tanks. The source of compressed air provided by the present invention includes a transaxle mounted on the vehicle and an air compressor coupled to the output shaft of the transaxle. During movement of the vehicle, the transaxle is operated by the wheels to which it is connected and in turn operates the air compressor. Compressed air for engine start-up and engine operation when the vehicle is stationary is provided by a rechargeable air tank connected between the air compressor and engine. The rechargeable air tank is continually maintained in a charged condition by the air compressor.

In one form of the present invention there is provided a compressed air propulsion system in a vehicle comprising an engine having a plurality of cylinders slidably receiving therein a respective plurality of pistons. The engine is operatively connected to an axle of the vehicle for driving the same. A transaxle mechanism is also provided and is mounted on and driven by the vehicle, and an air compressor is coupled to the transaxle mechanism for supplying a flow of compressed air. An air line assembly leads from the air compressor to the engine for delivering compressed air thereto, and an air distributor in the air line assembly sequentially distributes compressed air to individual ones of the cylinders.

Another aspect of the present invention provides in combination with a compressed air driven vehicle including an engine having a plurality of pistons reciprocatively received in a respective plurality of cylinders, the pistons being operatively coupled to a crankshaft; a supply of compressed air; and an air line assembly connected between the supply of compressed air and the engine; an air distributor system comprising a plurality of electrically-operated valves connected between respective ones of the cylinders and the air line assembly, wherein the valves are sequentially electrically actuable to distribute compressed air to the cylinders. A plurality of electrical switches are arranged in a generally circular array and are electrically connected to respective ones of the valves for supplying electric current for the actuation thereof, and a rotor is disposed generally within the array of switches and is operatively coupled to the crankshaft to be rotated thereby. Upon rotation, the rotor is sequentially engageable with respective ones of the electrical switches to close respective ones thereof.

It is an object of the present invention to provide an improved compressed air propulsion system in a vehicle.

Another object of the present invention is to provide a unique air distributor system for sequentially distributing air to the engine cylinders.

Yet another object of the present invention is to provide an improved air distributor system comprising a plurality of electrical switches mechanically actuated by a rotatable rotor for supplying electric current to solenoids connected to respective air lines leading to the engine cylinders.

A further object of the present invention is to provide an air compressor operatively coupled to and driven by a transaxle mechanism mounted on the vehicle.

Further objects of the present invention will appear as the description proceeds.

BRIEF DESCRIPTION OF THE DRAWINGS

The above mentioned and other features and objects of this invention, and the manner of attaining them, will become more apparent and the invention itself will be better understood by reference to the following description of an embodiment of the invention taken in conjunction with the accompanying drawings, wherein:

FIG. 1 is a schematic view of a preferred embodiment of the present invention;

FIG. 2 is a sectional view of the compressor in the embodiment in **FIG. 1**;

FIG. 3 is a broken-away side elevational view of the accelerator assembly in the embodiment in **FIG. 1**;

FIG. 4 is a broken-away perspective view of the vehicle engine compartment containing the engine and portions of the embodiment in **FIG. 1**;

FIG. 5 is a broken-away sectional view of the accelerator in the embodiment in **FIG. 1**;

FIG. 6 is a sectional view of **FIG. 5** taken along line **6—6** and viewed in the direction the arrows;

FIG. 7 is a broken-away sectional view of the engine in the embodiment in **FIG. 1**;

FIG. 8 is top plan view of the head for the engine in **FIG. 7**;

FIG. 9 is bottom plan view of the head in **FIG. 8**;

FIG. 10 is a broken-away sectional view of the distributor in the embodiment in **FIG. 1**;

FIG. 11 is a top plan view of the rotor used in the accelerator in **FIG. 10**;

FIG. 12 is a sectional view of **FIG. 11** taken along line **12—12** and viewed in the direction of the arrows; and

FIG. 13 is a broken-away sectional view of a microswitch in the accelerator in **FIG. 10**.

DESCRIPTION OF A PREFERRED EMBODIMENT

Referring primarily to FIGS. **1** and **4**, vehicle **20** is illustrated and comprises frame **22**, body **24**, engine compartment **26**, front axle **28**, rear axles **30**, **32**, and wheels **34** connected to axles **28**, **30**, **32**.

Transaxle **38** is mounted in the rear of vehicle **20** and includes an output shaft (not shown) connected to crankshaft **40** (FIG. **2**) of air compressor **36**, and is coupled to rear axles **30**,**32**. Air reservoir **42** is connected to air compressor **36** by air lines **44**, **46**, **48** and air line **50** is connected to air reservoir **42** and leads forward to engine compartment **26** for delivering compressed air to engine **52**, as will be explained in greater detail hereinafter. Front axle **28** is conventionally connected to a transmission (not shown), which in turn is conventionally connected to engine **52** for driving front wheels **34**.

Referring now to FIGS. **1** and **2**, compressor **36** includes piston cylinder arrangement **54** coupled to transaxle **38**. Piston-cylinder arrangement **54** comprises block **56** rotatably housing therein crankshaft **40**, which has secured thereto cranklever **58**. Block **56** further includes cylinder **60** slidably containing piston **62** and cylinder **64** slidably containing piston **66**. Although FIG. **2** only illustrates two cylinders, piston-cylinder arrangement **54** may include more or fewer cylinders. Piston rod **68** is pivotally connected to piston **62** by pin **70** and is pivotally connected at its opposite end to one end of cranklever **58** by pin **72**. Piston rod **74** in a like manner is pivotally connected to piston **66** by pin **76** and is pivotally connected at its opposite end to the other end of cranklever **58** by pin **78**.

Each cylinder **60**, **64**, or a set of cylinders in a V-arrangement, has a respective head **80**, **82** attached thereto. FIG. **2** illustrates head **80** sectioned to illustrate intake port **84**, and head **82** sectioned to illustrate exhaust port **86** to assist in describing the structure and operation of compressor **36**. However, it should be understood that each cylinder **60**, **64** in compressor **36** is provided with both an intake port and exhaust port.

Intake port **84** of head **80** is provided with a one-way check valve **88** comprising connector **90** securely seated in intake port **84**, and having an enlarged chamber **92** and passage **94** which communicate with intake port **84**. Ball **96** is biased upwardly by spring **98** to block passage **94**, the diameter of ball **96** being greater than the diameter of passage **94**. As described, valve **88** will only permit the flow the compressed air in one direction, that being through passage **94** and into intake port **84** when the pressure of the air flow forces ball **96** against the spring tension of spring **98**, thereby unseating ball **96** away from passage **94**. When the pressure of the flow of compressed air is less than the spring tension of spring **98**, ball **96** is upwardly biased against passage **94**, thereby preventing reverse air flow through intake port **84** and passage **94**.

In a similar manner, one-way check valve **100** permits air flow in only one direction, that being from cylinder **64** through exhaust port **86** and passage **102** of connector **104**. Chamber **106** in passage **102** contains ball **108** biased downwardly against passage **102** by spring **110**.

Block **56** further includes bottom plate **112** secured thereto by suitable fasteners, such as screws **114**. The operation of piston-cylinder arrangement **54** is accomplished by the rotation of rear wheels **34** turning rear axles **30**, **32**, which are operatively coupled to crankshaft **40** by transaxle **38**.

Referring to FIG. **1**, air filter **116** is mounted on compressor **36** and includes air line **118** leading therefrom and between heads **80**, **82**. Each intake valve **88** is connected by a respective air line to air line **118**; the upper left intake valve **88** being connected to air line **118** by air line **120**, the upper right intake valve **88** being connected to air line **118** by air line **122**, the lower left intake valve **88** being connected to air line **118** by air line **124**, and the lower right intake valve being connected to air line **118** by air line **126**. Although only one air filter **116** is illustrated, individual air filters may be provided for each intake valve **88**.

Continuing to refer to FIG. **1**, the two exhaust check valves **100** on the right side of compressor **36** are connected to air line **46** by air lines **128**, **130**, and the two exhaust check valves **100** on the left side of compressor **36** are connected to air line **44** by air lines **132**, **134**.

Compressor **36**, as described, need not be limited to a piston-cylinder arrangement for providing compressed air. For example, compressor **36** may be a rotary compressor manufactured with a self-contained internal valve system and having a single compressed air discharge line connected to air line **48**. Further, compressor **36** is manufactured as, or modified to be, a continuous non-oil pump compressor, which will operate to provide compressed air whether vehicle **20** is moving forward or in reverse. The non-oil pump compressor refers to a compressor which is lubricated by an oil slinger device connected to the crankshaft. Upon rotation of the crankshaft, the oil slinger device is rotated through the oil sump and throws oil toward cavities machined in the compressor block or crankcase. The cavities include oil passages leading to bearings and other moving parts for providing lubrication thereof. Compressor **36** is capable of providing compressed air at a pressure of about 1000–1100 pounds per square inch.

Air reservoir **42** is mounted just forward of compressor **36** on frame **22** and includes external connection **136**, which can be used, if necessary, to charge reservoir **42** with compressed air, air pressure gauge **138**, and relief valve **140**. Air pressure gauge **138** registers the pressure within reservoir **42** and may be located in either the trunk or on the dashboard in view of the vehicle operator. Air reservoir **42** is designed and constructed to contain compressed air at a pressure of at least **1500** psi, and relief valve **140** is designed to open when the air pressure exceeds **1500** psi.

Transaxle **38** includes gear shift link **142** connected by control cable **144** to gear shift lever **146** near the vehicle operator, and clutch link **148** connected by cable **150** to clutch **152**, also near the vehicle operator. Clutch **152** can be designed to be either hand or foot operated, and allows gear shift link **142** to shift transaxle **38** between high and low gears. Gear shift lever **146** shifts transaxle **38** between high gear and low gear either to decrease or increase the output of compressor **36**. For example, during acceleration or travel up steep inclines, low gear would be selected to increase the output of compressor **36**, while high gear would be selected to maintain a steady-state velocity along a level or less hilly road. Preferably, the low gear supplies twice as much compressed air as does high gear. Naturally, should compressor **36** be replaced by some other suitable compressor, it should also be capable of supplying a variable amount of compressed air. In this embodiment, transaxle **38** was removed from a Volkswagon Beetle and mounted on vehicle **20**, and fourth gear is used as the high gear and third gear as the low gear.

Referring now to FIGS. **1** and **4**, air line **50** leads from compressor **36** into engine compartment **26** and is connected to regulator **154**, which is a manually operated valve for regulating the air pressure delivered to engine **52**. Air pressure gauge **156** is connected to air line **50** to indicate the air pressure at the inlet side of regulator **154**. Regulator **154** is manually adjusted by handle **158** and compressed air delivered therefrom passes through airline **160** to oiler **162**, and thence through flexible hose **164** to accelerator **166**.

Referring to FIGS. **1** and **4**, oiler **162** comprises container **168** having an opening (not shown) in the bottom thereof with drain valve **170** disposed therein. Container **168** is filled with a suitable lubricant, and drain valve **170** is utilized to drain old lubricant from container **168**. Cover **172** is attached to container **168** by screws **174** and includes a hollow tubular portion **176** to which is connected air line **160** and flexible hose **164**. Cover **172** includes an opening (not shown) with a filler cap **178** disposed therein, and is used to refill container **168** with lubricant. An adjusting valve **180** is disposed in hollow tubular portion **176** and is utilized to regulate the flow of lubricant delivered from oiler **162** into flexible hose **164** and thence through accelerator **166** to the cylinders of engine **52**. In this embodiment, oiler **162** is adjusted to provide **38** drops of lubricant per hour for lubricating moving parts and other assemblies downstream of oiler **162**. Oil added to the compressed air system does not enter the oil sump of engine **52** and is exhausted through the exhaust system of vehicle **20**. Lubricant provided by oiler **162** is in addition to the conventional lubrication system of vehicle **20**. Oiler **162** can be a type 041-2 oiler manufactured by the Norgren Company of Denver, Colo., and is generally used for pneumatic tools and generally referred to as an "automatic machine oiler."

Referring now to FIGS. **1**, **3**, **4**, **5**, and **6**, accelerator assembly **182** will be described. Accelerator **166** comprises an outer housing **184** having a generally cylindrically shaped interior **186**. One end of outer housing **184** has opening **188**, which has a diameter the same as the diameter of interior **186**, and a smaller opening **190** in the other end thereof; and further includes hole **192** radially disposed therein.

Rotatably received in interior **186** is accelerator valve **194** having a pair of smaller ends **196**, **198** extending axially therefrom. End **198** is rotatably received through opening **190**, and end **196** is rotatably received in closed opening **200** of cover **202**. Cover **202** is secured to outer housing **184** by fasteners, such as screws **204**. A pair of washers **206**, **208** are disposed between radially extending shoulder **210** and cover **202** and between radially extending shoulder **212** and annular surface **214** of outer housing **184**.

Valve **194** also includes a plurality of grooves **216** containing a respective plurality of O-rings **218**.

Accelerator **166** is connected to manifold **220** of engine **52** by air line **222**, which is threadedly secured in hole **192** in outer housing **184** and opening **224** in manifold **220**. Outer housing **184** is also secured to an L-shaped bracket **226**, which is secured to engine **52**.

Accelerator valve **194** further includes air passage **228** having an axially extending portion **230** that extends through end **198** and into the central interior of valve **194**, and a radially extending portion **232**, which has an inner diameter of about $\frac{3}{8}$ (three-eighths) of an inch and which can be rotated to be axially aligned with opening **234** of air line **222**. Flexible. hose **164** is connected to portion **230** of air passage **228** for delivering compressed air thereto.

Lever **236** has a generally circularly shaped end **238** and is secured to accelerator valve end **198** by screw **240** received through hole **242** in end **238** and threadedly secured in threaded opening **244** in end **198**. Lever **236** has its opposite end formed into an eye **246** for connecting accelerator cable **248** thereto. Accelerator cable **248** comprises a hollow sheath **250**, which is connected at one end thereof to flange **252** of bracket **226** and at its opposite end near accelerator pedal **254**. A cable **256** is slidably received in sheath **250** and has one end thereof received through eye **246** with a fastener **258** secured thereto by set screw **260**. A spring **262** is slidably received over cable **256** and disposed between lever **236** and flange **252** to bias lever **236** upwardly as viewed in FIG. 3. The opposite end of cable **256** is connected to accelerator pedal **254**. By depressing pedal **254**, cable **256** pulls downwardly on lever **236** to rotate accelerator valve **194** in outer housing **184**. Rotation of valve **194** selectively moves portion **232** of air passage **228** between a position where it is unaligned with opening **234** of air line **222** and a second position wherein portion **232** and opening **234** are coaxially aligned. FIG. 6 illustrates a partial alignment between portion **232** and opening **234** to illustrate the manner in which the flow of compressed air through accelerator **166** can be selectively varied by the operator. Upon release of accelerator pedal **254**, spring **262** biases lever **236** upwardly to its original position. The alignment of portion **232** and opening **234** is adjusted so that there is partial alignment between the two when accelerator pedal **254** is not depressed, thereby providing a flow of compressed air to permit engine **52** to idle.

Referring to FIGS. **1**, **4**, **7**, **8**, and **9**, manifold **220**, which is connected to air line **222** leading from accelerator **166**, is mounted on engine **52** and has four air lines **264**, **266**, **268**, **270** connected thereto and leading to engine head **272**. An air pressure gauge **274** is connected to manifold **220** to indicate to the vehicle operator the air pressure therein. The preferred air pressure in manifold **220** for engine **52** of the present embodiment is between about 125 psi to 170 psi. Naturally, this range of air pressure will vary depending on the particular type of engine used.

In this particular embodiment, engine **52** is a 1975 Volkswagon Rabbit 4-cylinder engine having a cubic inch displacement of 90 CID and of the engine family 32NC. Other engines can naturally be used, however, the air pressures and other parameters and specifications disclosed herein pertain to this particular type of engine **52**.

The inner diameters of air lines **264**, **266**, **268**, **270** is about 9/32 (nine-thirty secondths) inch, and they have coupled therein respective 12 volt solenoids **276**, **278**, **280**, **282**.

Engine **52** comprises engine block **284**, crankcase **286**, crankshaft **288** rotatably mounted in crankcase **286** by journals **290**, **292**, and cylinders **294**, **296**, **298**, **300** having reciprocatively received therein pistons **302**, **304**, **306**, **308**, respectively. Pistons **302**–**308** are conventionally connected pivotally to rods **310**, which are conventionally connected pivotally to respective crankarms **312** of crankshaft **288**. Each cylinder **294**-**300** includes a respective exhaust port **314**, **316**, **318**, **320**, which have a diameter of about 9/16 (nine-sixteenths) of an inch. Crankshaft **288** is connected to the transmission (not shown) for driving front axle **28** and front wheels **34**.

It should be noted that exhaust ports **314**–**320** are radially disposed in respective cylinders **294**–**300** at a point where exhaust ports **314**–**320** are just slightly above the top surfaces of respective pistons **302**-**308** when the pistons are at the bottom of their power strokes.

Referring primarily to FIGS. **7**, **8** and **9**, engine head **272** includes annular recesses **322**, **324**, **326**, **328**, which are respectively aligned with cylinders **294**, **296**, **298**, **300**. Recesses **322**–**328** include respective openings **330**, **332**, **334**, **336**, and in which are tightly secured respective connectors **338**, **340**, **342**, **344** for connecting air lines **264**–**270** to head **272**. Head **272** also includes holes **346** for receiving therethrough suitable fasteners for rigidly securing head **272** to engine block **284**.

Referring to FIGS. **1**, **4**, **10**–**13**, air distributor assembly **348** comprises distributor shaft **350** (FIG. **10**) conventionally coupled to crankshaft **288** of engine **52** to be rotated thereby, and rotor **352** secured to distributor shaft **350** to rotate therewith. Unique to the invention is the absence of the usual condenser and points in a conventional distributor for sequentially supplying electric current to the spark plugs of a conventional engine.

Referring to FIGS. **11** and **12**, rotor **352** is generally triangular in shape and includes hole **354** for connecting rotor **352** to distributor shaft **350** and an elevated or raised camming surface **356**, which includes a sloping peripheral portion **358**. Upon rotation of rotor **352**, camming surface **356** mechanically actuates four microswitches **360**, **362**, **364**, **366**, which are respectively electrically connected to solenoids **276**, **278**, **280**, **282** by electrical lines **368**, **370**, **372**, **374**, respectively. Microswitches **360**–**366** are also electrically connected to battery **376** by electrical lines **378**, **380**, **382**, **384**, respectively.

Microswitches **360**–**366** are mounted on distributor cover **386** and are generally circularly arrayed about distributor shaft **350** and rotor **352**. Referring to FIG.

13, since microswitches **360-366** are identical, only microswitch **360** will be described. Microswitch **360** comprises an outer shell **388** having a pair of spaced-apart electrical conductors **390**, **392** mounted on opposite sides thereof by any suitable fastening means. Electrical conductor **392** is made of a resilient material and has electrical line **378** electrically connected thereto, and electrical conductor **390** has electrical line **368** connected thereto. Shell **388** includes opening **394** and a pin member **396** slidably received therein. Pin member **396** has an enlarged head portion **398**, which supports pin member **396** in shell **388** and which is also in close proximity to or slightly abutting against electrical conductor **392**. Directly above pin member **396** is spring guide **400** attached to or integral with shell **388**, and spring **402** is received over spring guide **400** and abuts against shell **388** and the top of electrical conductor **392** so as to bias conductor **392** away from conductor **390**, thereby maintaining conductor **390**, **392** in a spaced-apart relation with no external force applied to pin member **396**. Pin member **396** includes cam follower surface **404** engageable by camming surface **356** and sloping peripheral portion **358**.

In operation, rotation of rotor **352** will cause sloping peripheral portion **358** and camming surface **356** to sequentially engage and move respective pin members **396** of microswitches **360-366** upwardly to force electrical conductors **392** to electrically contact electrical conductors **390**, thereby sequentially closing the circuit between the battery **376**, and respective solenoids **276-282**. In this particular embodiment, solenoids **276-282** are required to be open for a time interval of approximately five milliseconds to provide a sufficient amount of compressed air to enter respective cylinders **294-300**. The time interval of five milliseconds during which solenoids **276-282** are open is a function of the period of time during which electrical conductors **390** and **392** are in contact. The length L of camming surface **356** determines the time interval during which a respective pin member **396** is forcibly engaged by camming surface **356**. Of course, the duration during which solenoids **276-282** remain open can be increased or decreased by lengthing or shortening the length L of camming surface **356**. Perferably, rotor **352** and pin member **396** are made of Bakelite or a similar material.

As earlier mentioned, lubrication of the air propulsion system is accomplished by oiler **162** which delivers oil to lubricate accelerator **166**, solenoids **276-282**, and cylinders **294-300** and pistons **302-308**. The oil is then exhausted through exhaust ports **314-320**. Although not necessary to the operation of vehicle **20**, an exhaust manifold (not shown) may be connected to exhaust ports **314-320**, and a tail pipe and muffler system (not shown) may also be utilized with vehicle **20**.

It can also be seen now that no pollution control devices are required with vehicle **20** since engine **52** is operated by compressed air rather than by fuel combustion. It should also be pointed out that vehicle **20** utilizes other standard features or accessories of a typical vehicle, such as a transmission, brake system, lighting system, and the like, except that vehicle **20** does not require a radiator for cooling engine **52**.

In operation, vehicle **20** is initially started the same as a conventional vehicle in that a starter (not shown) is utilized to crank engine **52**, thereby rotating distributor shaft **350** and rotor **352** in order to sequentially supply electrical current to solenoids **276-282**. Since air reservoir **42** is maintained in a charged condition, engine **52** can be started by depressing accelerator pedal **254**, thereby permitting a flow of compressed air from reservoir **42** through air line **50**, regulator **154**, oiler **162**, accelerator **166**, and into manifold **220**. From manifold **220**, the flow of compressed air is.then sequentially delivered to cylinders **294-300** by solenoids **276-282**. When engine **52** is running, alternator **406**, which is conventionally coupled thereto, is operated to continually charge battery **376**. Vehicle **20** is then driven in the normal manner using accelerator pedal **254** and the conventional automatic or manual transmission. If required for greater acceleration, or in traveling in hilly areas, clutch **152** can be used to allow gear shift lever **146** to operate gear shift link **142** to shift transaxle **38** from high gear to low gear. When vehicle **20** is in motion, the rotation of rear wheels **34** will rotate rear axles **30**, **32**, which rotation is transferred by transaxle **38** to compressor **36**, which will then supply a flow of compressed air to engine **52** and maintain air reservoir **42** in a charged condition.

Compressed air delivered to cylinders **294**, **296**, **298**, **300** is exhausted through respective ports **314**, **316**, **318**, **320** when the top surfaces of respective pistons **302**, **304**, **306**, **308** are below the level of their respective ports **314-320**. An exhaust manifold and exhaust pipe (not shown) then deliver the exhausted air to the atomsphere.

Engine shut down is performed in the usual manner by turning the ignition system (not shown) off, thereby removing electrical current from air distributor system **348**.

While this invention has been described as having a preferred embodiment, it will be understood that it is capable of further modifications. This application is therefore intended to cover any variations, uses, or adaptations of the invention following the general principles thereof, and including such departures from the present disclosure as come within known or customary practice within the art to which this invention pertains and falls within the limits of the appended claims.

What is claimed:

1. A compressed air propulsion system for a vehicle, comprising:

an engine including a plurality of cylinders reciprocatingly receiving therein a respective plurality of pistons, and an exhaust means connected to said cylinders, said engine drivingly connected to an axle of said vehicle to propel said vehicle;

a transaxle operatively mounted on and rotatably driven by said vehicle;

air compressing means rotatingly driven by said transaxle for supplying a flow of compressed air to an air reservoir;

an air supply line connecting said air reservoir to said engine for supplying compressed air thereto;

an air distributor connected to said air supply line and said cylinders and including a rotor operatively coupled to said pistons for sequentially actuating a plurality of solenoid valve means to sequentially supply compressed air to individual ones of said cylinders;

accelerator means including a housing having an opening therein and connected to said air reservoir by said air line, a valve member received in said housing and having an air passage therein connected to said engine by said air line, said valve member selectively adjustable in said housing between a first position wherein said opening and said air passage are unaligned and a second position wherein said opening and said air passage are aligned, the alignment of said opening and air passage being adjustable to vary the flow rate of compressed air therethrough;

said air distributor including a plurality of electrical switch means, each said electrical switch means comprising; a pair of spaced-apart electrical conductors, one of said electrical conductors being electrically connected to a respective one of said plurality of solenoid valve means; and cam follower means engageable by a camming surface of said rotor for mechanically moving a first one of said electrical conductors between a first position wherein said first one of said electrical conductors is spaced apart from the second one of said electrical conductors and a second position wherein said first one of said electrical conductors is in electrical contact with said second one of said electrical conductors, whereby said respective one of said solenoid valve means is electrically actuated when said cam follower means mechanically moves said first one of said electrical conductors into electrical contact with said second one of said electrical conductors.

2. The system of claim **1** wherein said air compressing means is a continuous non-oil pump compressor.

3. The system of claim **1** wherein said exhaust means comprises a plurality of exhaust ports respectively disposed in said cylinders, each said exhaust port being radially disposed in a lower portion of a respective said cylinder, said lower portion being just slightly above the top surface of the respective said piston when said piston is at the bottom of its stroke.

4. The vehicle of claim **1** further comprising a manifold connected to said air supply line means between said accelerator means and said plurality of valve means, and having a plurality of air conduits leading therefrom, said plurality of valve means being connected between respective ones of said cylinders and respective ones of said air conduits.

5. The vehicle of claim **1** wherein said accelerator housing is stationary and generally cylindrically shaped with said opening being radially disposed therein, said valve member being rotatable and said air passage being radially disposed in said valve member.

6. A compressed air propulsion system for a vehicle, comprising:

an engine including a plurality of cylinders reciprocatingly receiving therein a respective plurality of pistons, and an exhaust means connected to said cylinders, said engine drivingly connected to an axle of said vehicle to propel said vehicle;

a transaxle operatively mounted on and rotatably driven by said vehicle;

air compressing means rotatingly driven by said transaxle for supplying a flow of compressed air to an air reservoir;

an air supply line connecting said air reservoir to said engine for supplying compressed air thereto;

an air distributor connected to said air supply line and said cylinders and including a rotor operatively coupled to said pistons for sequentially actuating a plurality of solenoid valve means to sequentially supply compressed air to individual ones of said cylinders;

accelerator means including a housing having an opening therein and connected to said air reservoir by said air line, a valve member received in said housing and having an air passage therein connected to said engine by said air line, said valve member selectively adjustable in said housing between a first position wherein said opening and said air passage are unaligned and a second position wherein said opening and said air passage are aligned, the alignment of said opening and air passage being adjustable to vary the flow rate of compressed air therethrough;

said air distributor including a plurality of electric switches, each of said switches including a pair of electrical conductors, one of said conductors being electrically connected to a respective one of said solenoid valve means; and a cam follower element engageable by said rotor for moving a first one of said electrical conductors between an open position wherein said first one of said electrical conductors is spaced apart from the second one of said electrical conductors and a closed position wherein said first one of said electrical conductors is in electrical contact with said second one of said electrical conductors to supply electrical current to a respective one of said valve means, said rotor including a contoured camming surface engageable with said cam follower element when rotated by said pistons; said camming surface moving said cam follower element between a first position wherein cam follower element is at least adjacent to said first one of said electrical conductors at said open position and a second position wherein said cam follower element forcibly abuts against said first one of said electrical conductors to move said first one of said electrical conductors from said open position to said closed position.

7. The vehicle of claim **6** wherein the length of said camming surface engageable by said cam follower element is predetermined by the length of time during with each said valve means is open to distribute compressed air to its respective said cylinder.

* * * * *

US005159703A

United States Patent [19]

Lowery

[11] **Patent Number: 5,159,703**

[45] **Date of Patent: Oct. 27, 1992**

[54] **SILENT SUBLIMINAL PRESENTATION SYSTEM**

[76] Inventor: **Oliver M. Lowery**, 5188 Falconwood Ct., Norcross, Ga. 30071

[21] Appl. No.: **458,339**

[22] Filed: **Dec. 28, 1989**

[51] **Int. Cl.⁵** **H04B 7/00;** H04R 25/00; H04R 3/02

[52] **U.S. Cl.** **455/42;** 455/46; 455/66; 381/73.1; 128/420.5

[58] **Field of Search** 455/46, 47, 66, 109, 455/110, 42–43; 381/73.1, 105, 124; 358/141-143; 600/28; 128/420.5; 380/38

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,060,795	10/1962	Corrigan et al.	352/131
3,278,676	10/1966	Becker	358/142
3,393,279	7/1968	Flanagan	128/420.5
3,712,292	1/1973	Zentmeyer, Jr.	600/28
4,141,344	2/1979	Barbara	600/28
4,395,600	7/1983	Lundy et al.	381/73.1
4,463,392	7/1984	Fischer et al.	360/30
4,777,529	10/1988	Schultz et al.	381/73.1
4,834,701	5/1989	Masaki	600/28
4,877,027	10/1989	Brunkan	128/420.5

Primary Examiner—Reinhard J. Eisenzopf
Assistant Examiner—Andrew Faile

[57] **ABSTRACT**

A silent communications system in which nonaural carriers, in the very low or very high audio frequency range or in the adjacent ultrasonic frequency spectrum, are amplitude or frequency modulated with the desired intelligence and propagated acoustically or vibrationally, for inducement into the brain, typically through the use of loudspeakers, earphones or piezoelectric transducers. The modulated carriers may be transmitted directly in real time or may be conveniently recorded and stored on mechanical, magnetic or optical media for delayed or repeated transmission to the listener.

3 Claims, 3 Drawing Sheets

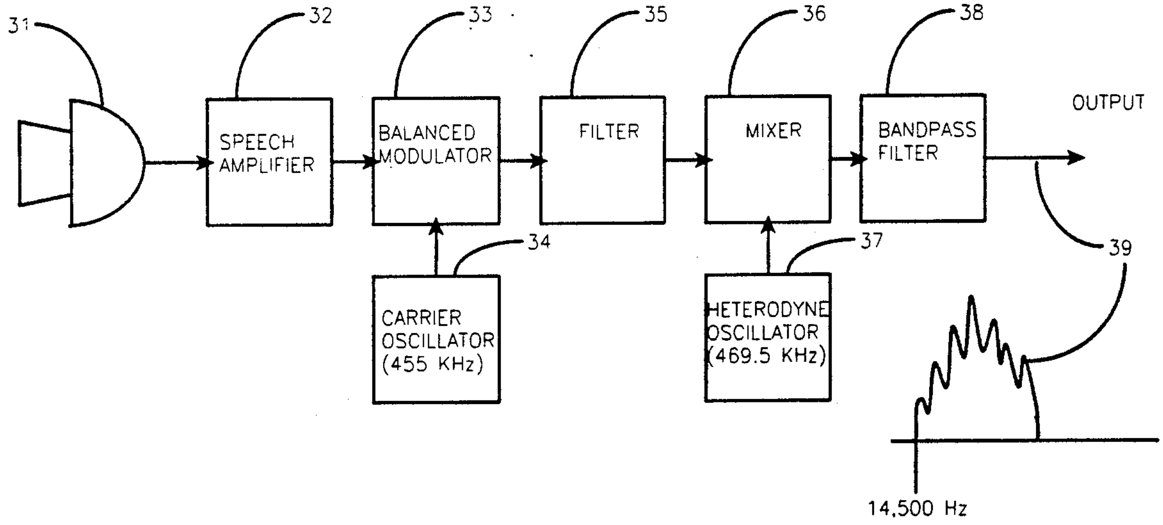
31
32
33
35
36
38
SPEECH AMPLIFIER
BALANCED MODULATOR
FILTER
MIXER
BANDPASS FILTER
OUTPUT
34
37
39
CARRIER OSCILLATOR (455 KHz)
HETERODYNE OSCILLATOR (469.5 KHz)
14,500 Hz

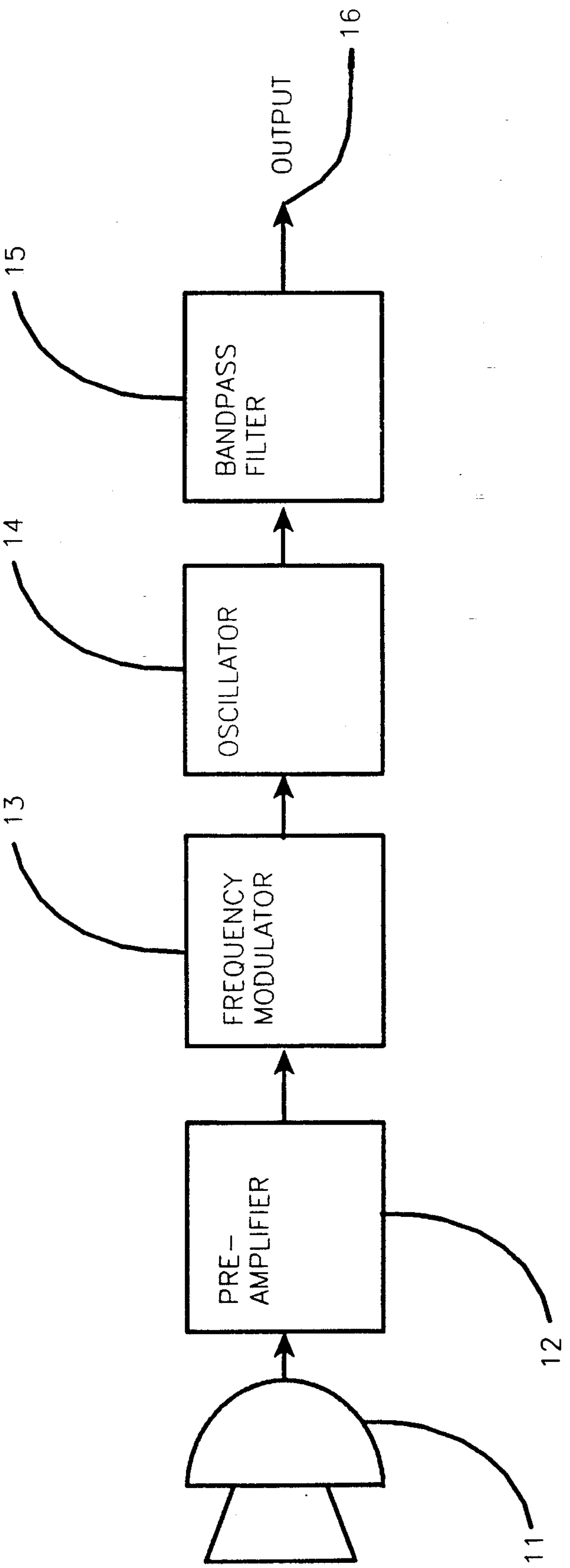
11
12
PRE-AMPLIFIER
13
FREQUENCY MODULATOR
14
OSCILLATOR
15
BANDPASS FILTER
OUTPUT
16

FIG. 1

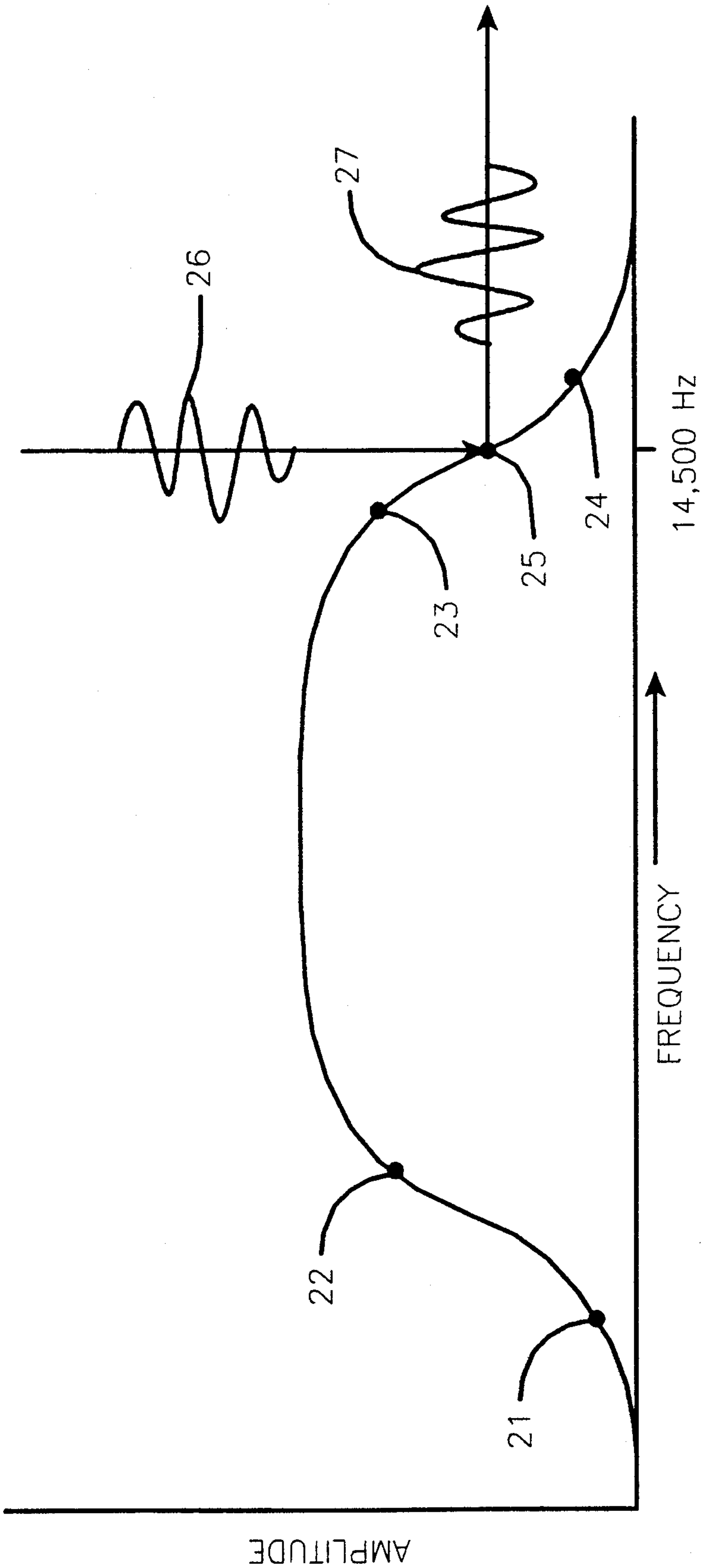

FIG. 2

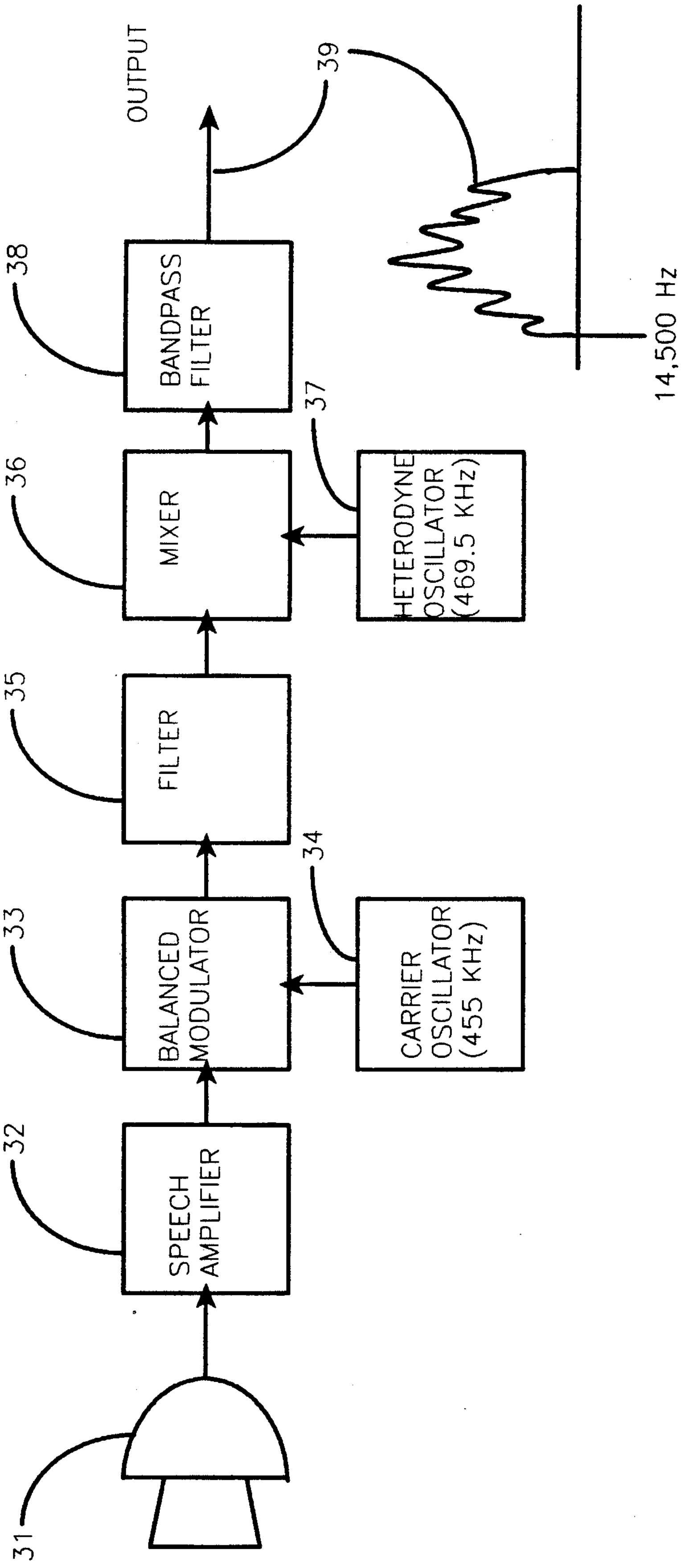
31
32
SPEECH AMPLIFIER
33
BALANCED MODULATOR
34
CARRIER OSCILLATOR (455 KHz)
35
FILTER
36
MIXER
37
HETERODYNE OSCILLATOR (469.5 KHz)
38
BANDPASS FILTER
OUTPUT
39
14,500 Hz

FIG. 3

SILENT SUBLIMINAL PRESENTATION SYSTEM

BACKGROUND—FIELD OF THE INVENTION

This invention relates in general to electronic audio signal processing and, in particular, to subliminal presentation techniques.

BACKGROUND—DESCRIPTION OF PRIOR ART

Subliminal learning enjoys wide use today and subliminal tapes are being manufactured by a number of companies in the United States alone. Several decades of scientific study indicate that subliminal messages can influence a human's attitudes and behavior. Subliminal, in these discussions, can be defined as "below the threshold of audibility to the conscious mind." To be effective however, the subliminally transmitted information (called affirmations by those in the profession) must be presented to the listener's ear in such a fashion that they can be perceived and "decoded" by the listener's subconscious mind. We are referring to audio information in this discussion, however, information could be inputted into the subject's subconscious mind through any of the body's sensors, such as touch, smell, sight or hearing. As an example, early development work in the subliminal field utilized motion pictures and slide projections as the medium. Early research into visual and auditory subliminal stimulation effects is exemplified by U.S. Pat. Nos. 3,060,795 of Corrigan, et al. and 3,278,676 of Becker. U.S. Pat. No. 4,395,600 of Lundy and Tyler is representative of later developments in today's subliminal message techniques.

The majority of the audio subliminal tapes available today are prepared using one basic technique. That is, the verbal affirmations are mixed with, and recorded at a lower level than, a "foreground" of music or sounds of ocean surf or a bubbling mountain brook or other similar "masking" sounds. The affirmations are generally recorded 5 decibels (db) or so below the "foreground" programming and regenerative automatic gain control is usually applied to permit the affirmations to change their recorded amplitude in direct proportion to the short term averaged amplitude of the continually varying "foreground" material. In other words, the volume of the affirmations will follow or track the volume changes of the "foreground" programming, but at a lower volume level. Circuit provisions are also usually included to "gate" the affirmations off when the music amplitude is low or zero. This insures that the affirmations cannot be heard during quiet program periods. Thus, today's subliminal affirmations can be characterized as being "masked" by music or other sounds, of constantly changing amplitude and of being reduced or cut off entirely during periods of low or quiet "foreground" programming.

One of the principal, and most widely objected to, deficiencies in available subliminal tape presentation techniques is that the presence of the "foreground" material is intrusive to both the listener and to anyone else in the immediate area. No matter what "foreground" material is chosen, the fact remains that this material can be heard by anyone within its range and presents a definite distraction to other activities such as conversation, thought, desire to listen to other programming such as radio or television, need to concentrate, etc. Additionally, and because the tapes are used repeatedly by the same listener, any "foreground" music or material eventually becomes monotonously tiring to that listener.

It is the purpose of the following described invention to eliminate or greatly reduce all of the above deficiencies. Although its application to the magnetic tape medium is described in the following discussion, the technique is equally applicable to most other desired transmission mediums, such as Compact Disc, videocassettes, digital tape recorders, Public Address (PA) systems, background music installations, computer software programs, random access memory (RAM), read only memory (ROM), "live", real time applications and other mediums now in existence or to be developed in the future.

Implemented on tape cassettes, for example, the subliminal presentation described here is inaudible i.e., high audio or ultrasonic frequencies, the affirmations are presented at a constant, high amplitude level, and they occupy their own "clear channel", non-masked frequency allocations. If desired, the previously described "foreground" music or other material can be added to the tape through use of an audio mixer. The "silent" recordings are inaudible to the user or by others present and are therefore very effective for use during periods of sleep or when in the presence of others. Additionally, the basic requirements of subliminal stimulation are met. That is, the affirmations are efficiently transmitted to the ear and, while undetected by the conscious mind, are perceived by and efficiently decoded by the subconscious mind.

OBJECTS OF THE INVENTION

Accordingly, several objects and advantages of my invention are:

(a) to provide a technique for producing a subliminal presentation which is inaudible to the listeners(s), yet is perceived and demodulated (decoded) by the ear for use by the subconscious mind.

(b) to provide a technique for transmitting inaudible subliminal information to the listener(s) at a constant, high level of signal strength and on a clear band of frequencies.

(c) to provide a technique for producing inaudible subliminal presentations to which music or other "foreground" programming may be added, if desired.

Still further objects and advantages will become apparent from a consideration of the ensuing description and drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

In the drawings, the first digit of each component number also refers to the figure number where that component can be located.

FIG. 1 represents the block diagram of a suitable system which will generate a frequency modulated (FM) signal at 14,500 Hz.

FIG. 2 represents an approximation of the frequency response curve of the human ear and the signal decoding process.

FIG. 3 represents the block diagram of a suitable system which will generate a single sideband, suppressed carrier, amplitude modulated (AM) signal at 14,500 Hz.

REFERENCE NUMERALS IN DRAWINGS	
11 microphone or other	14 low distortion

-continued

REFERENCE NUMERALS IN DRAWINGS	
audio input signal	audio oscillator
12 audio preamplifier if required	15 high pass or band pass audio filter
13 frequency modulation circuit	16 output to tape recorder or other device
21 point on low freq end response curve	25 midpoint on curve between points 23 and 24
22 point on low freq end of ear response curve	26 speaker output of FIG. 1 to ear
23 point on high freq end of ear response curve	27 demodulated subliminal audio inputted to ear
24 point on high freq end of ear response curve	31 microphone
32 speech amplifier	33 balanced modulator
34 carrier oscillator (455 KHz)	35 filter
36 mixer	37 heterodyne oscillator (469.5 KHz)
38 bandpass filter	39 output signal

DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT

Please refer now to FIG. **1** and FIG. **2**, which are drawings of a preferred implementation of the invention.

The principle of operation of the silent subliminal presentation system is as follows:

An audio signal in the upper frequency region of the audio spectrum (for example, 14,500 Hz) is modulated with the desired information. The type of modulation may be any type suitable for subliminal applications; frequency modulation (FM), phase modulation (PM), upper single sideband with suppressed carrier, amplitude modulation (AM), tone modulation, etc.

For broadest application, the high audio frequency selected as the carrier frequency must meet two basic criteria:

(1) be high enough in the audio spectrum that its presence to the human ear is essentially unnoticed or undetectable (without the listener being informed that the signal is actually present) and,

(2) be low enough in the audio spectrum that it (and its modulation content) can produce a useful output power from home entertainment type cassette or reel-to-reel magnetic recorders.

This would also include, of course, small portable and automobile tape decks.

Alternatively, the output of the system can be fed directly into an audio amplifier and its speaker/earphone system, Public Address system, etc.

FIG. **1** provides the block diagram of an example of a system capable of generating the desired silent frequency modulated carrier.

The modulation information is inputted into the microphone **11**. Other suitable input devices may be substituted for microphone **11**, such as a tape recorder or a radio. The microphone **11** is connected to the preamplifier **12** and should have provisions for adjusting its gain in order that the optimum modulation index can be set in the frequency modulator **13**. The frequency modulator **13** modulates the frequency of oscillator **14** which has been adjusted for an output of 14,500 Hz as described above. The output of oscillator **14** is fed through a suitable bandpass filter **15** into the tape recorder or directly into a suitable amplifier/speaker system. It is the purpose of the bandpass filter to remove or attenuate audible products of the modulation process in order to maintain as audibly silent an output as practical.

On the receiving end, FIG. **2** represents an approximate and idealized frequency response curve of the human ear. The frequency modulated carrier (centered at 14,500 Hz), as generated above and played through a tape recorder or amplifier/speaker system, is shown on FIG. **2** as speaker output **26**, impinging upon the upper slope of the ear's response curve at point **25**. The frequency modulated excursions of the speaker output **26** swing between points **23** and **24** on the ear's upper response curve. Because the response curve between points **23** and **25** is relatively linear, this action results in a relatively linear demodulation of the original modulation intelligence, which is passed on subliminally to the inner ear. The amplitude of the demodulated output is not high enough to be detected by the conscious mind but is sufficient in amplitude to be detected by the subconscious mind. In the field of communications engineering design, the above demodulation process in known as slope detection and was used in early FM receiver design. In those receivers, the response curve was formed by the action of a tuned (inductive/capacitance) circuit. In our case, the response curve is formed by the natural response curve of the human ear. The same slope detection technique can be performed at the low frequency end of the human ear response curve. This region is indicated on FIG. **2** as between points **21** and **22**. This region, however, has a much smaller available bandwidth and is therefore more restricted as to the amount of information that can be transmitted in an inaudible manner.

In practice, the listener adjusts the volume control of the tape recorder or amplifier to a level just below that at which the listener hears an audible sound or noise from the speaker of the tape recorder. If the recording process is properly done, a spectrum analyzer or a calibrated sound level meter will reveal a strong signal emanating from the tape recorder speaker. A calibrated sound level meter, at a distance of 1 meter (with C weighting and referenced to the standard of 0.0002 micro bar) will typically indicate a silent power output of from 60 to 70 decibels. This is equivalent to the audio power of a loud conversation, yet, in the described system, is inaudible or unnoticed by the listener.

FIG. **3** illustrates a system which generates a suitable amplitude modulated (AM) signal, instead of the frequency modulated (FM) system described above. The output is a modulated, single sideband (SSB), suppressed carrier (AM) signal at 14,500 Hz.

The block diagram represents a common scheme for generating an SSB signal and will be briefly described.

The desired subliminal information is spoken into microphone **31**. This signal is amplified by speech amplifier **32** and injected into one port of balanced modulator **33**. A continuous wave signal of **455** KHz is generated by carrier oscillator **34** and is injected into the second port of balanced modulator **33**. The output of balanced modulator **33** is a double sideband, suppressed carrier signal at 455 KHz. This signal is fed through filter **35**, causing one of the two sidebands to be removed. This signal is fed into one port of mixer **36**. A continuous wave signal at a frequency of 469.5 KHz from hetrodyne oscillator **37** is fed into the other port of mixer **36**, resulting in an output of the original subliminal audio information but translated 14,500 Hz higher in frequency. The bandpass filter **38** attenuates signals and noise outside of the frequencies of interest. The ampli-

tude modulated audio output signal is shown as output **39**.

Thus, as stated earlier, my invention provides a new system for subliminal presentations which is:

(a) silent,

(b) outputs a constant, high level modulated signal and,

(c) occupies a band of clear channel frequencies.

The foregoing description of the preferred embodiment of the invention has been presented for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed. Many modifications and variations are possible in light of the above discussions. It is intended that the scope of the invention be limited not only by this detailed description, but rather by the claims appended hereto.

What is claimed:

1. A silent communications system, comprising:

(a) amplitude modulated carrier means for generating signals located in non-aural portions of the audio and in the lower portion of the ultrasonic frequency spectrum said signals modulated with information to be perceived by a listener's brain and,

(b) acoustic and ultrasonic transducer means for propagating said signals, for inducement into the brain, of the listener, and,

(c) recording means for storing said modulated signals on mechanical, magnetic and optical media for delayed or repeated transmissions to the listener.

2. A silent communications system, comprising:

(a) frequency modulated carrier means for generating signals located in non-aural portions of the audio and in the lower portion of the ultrasonic frequency spectrum, said signals modulated with information to be perceived by a listener's brain, and;

(b) acoustic and ultrasonic transducer means for propagating said signals, for inducement into the brain of the listener, and;

(c) recording means for storing said modulated signals on mechanical, magnetic and optical media for delayed or repeated transmissions to the listener.

3. A silent communications system, comprising:

(a) a combination of amplitude and frequency modulated carrier means for generating signals located in non-aural portions of the audio and in the lower portion of the ultrasonic frequency spectrum, said signals modulated with information to be perceived by a listener's brain, and

(b) acoustic and ultrasonic transducer means for propagating said signals, for inducement into the brain of the listener;

(c) recording means for storing said modulated signals on mechanical, magnetic and optical media for delayed or repeated transmissions to the listener.

* * * * *

US005590031A

United States Patent [19]

Mead, Jr. et al.

[11] **Patent Number: 5,590,031**

[45] **Date of Patent: Dec. 31, 1996**

[54] SYSTEM FOR CONVERTING ELECTROMAGNETIC RADIATION ENERGY TO ELECTRICAL ENERGY

[76] Inventors: **Franklin B. Mead, Jr.**, 44536 Avenida Del Sol, Lancaster, Calif. 93535; **Jack Nachamkin**, 12314 Teri Dr., Poway, Calif. 92064

[21] Appl. No.: **281,271**

[22] Filed: **Jul. 27, 1994**

[51] **Int. Cl.**^{6} **H02M 1/00**

[52] **U.S. Cl.** **363/8**; 363/178; 342/6

[58] **Field of Search** 363/8, 13, 178; 342/6, 61, 73, 173, 175

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,882,503	5/1975	Gamara	343/100 R
4,725,847	2/1988	Poirier	343/840
5,008,677	4/1991	Trigon et al.	342/17

Primary Examiner—Peter S. Wong
Assistant Examiner—Adolf Berhane
Attorney, Agent, or Firm—Chris Papageorge

[57] **ABSTRACT**

A system is disclosed for converting high frequency zero point electromagnetic radiation energy to electrical energy. The system includes a pair of dielectric structures which are positioned proximal to each other and which receive incident zero point electromagnetic radiation. The volumetric sizes of the structures are selected so that they resonate at a frequency of the incident radiation. The volumetric sizes of the structures are also slightly different so that the secondary radiation emitted therefrom at resonance interfere with each other producing a beat frequency radiation which is at a much lower frequency than that of the incident radiation and which is amenable to conversion to electrical energy. An antenna receives the beat frequency radiation. The beat frequency radiation from the antenna is transmitted to a converter via a conductor or waveguide and converted to electrical energy having a desired voltage and waveform.

14 Claims, 8 Drawing Sheets

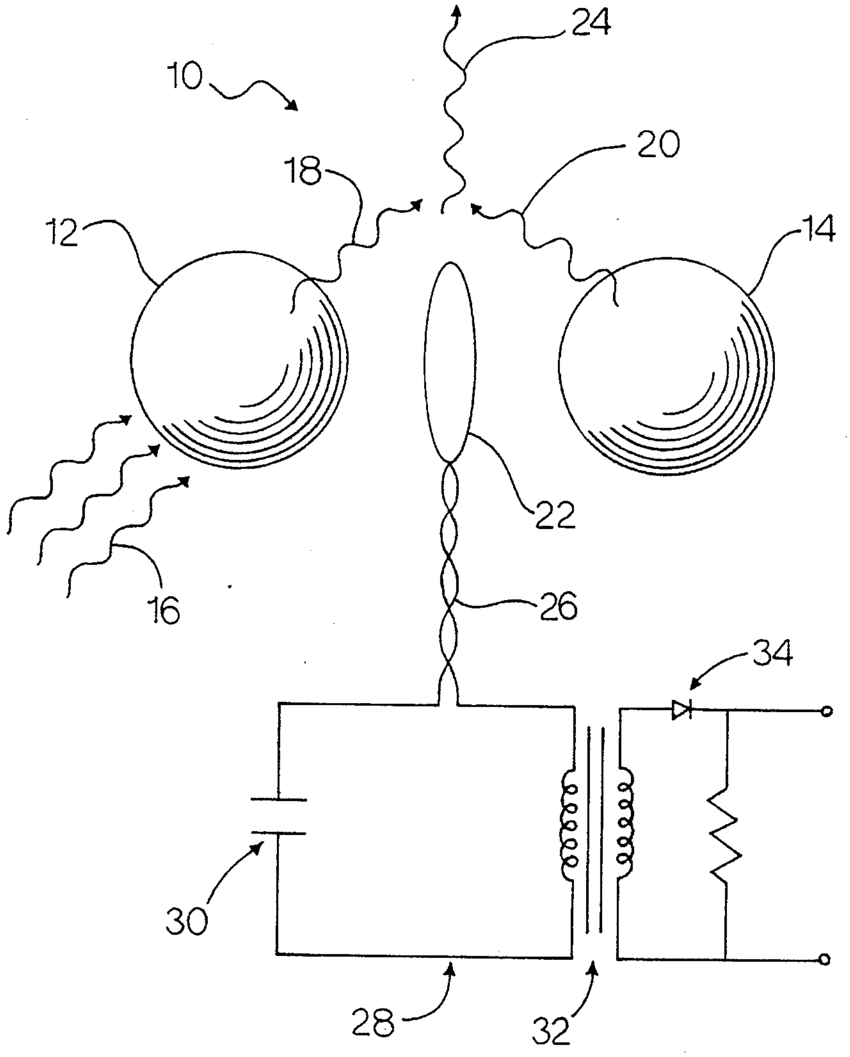

FIG. 1

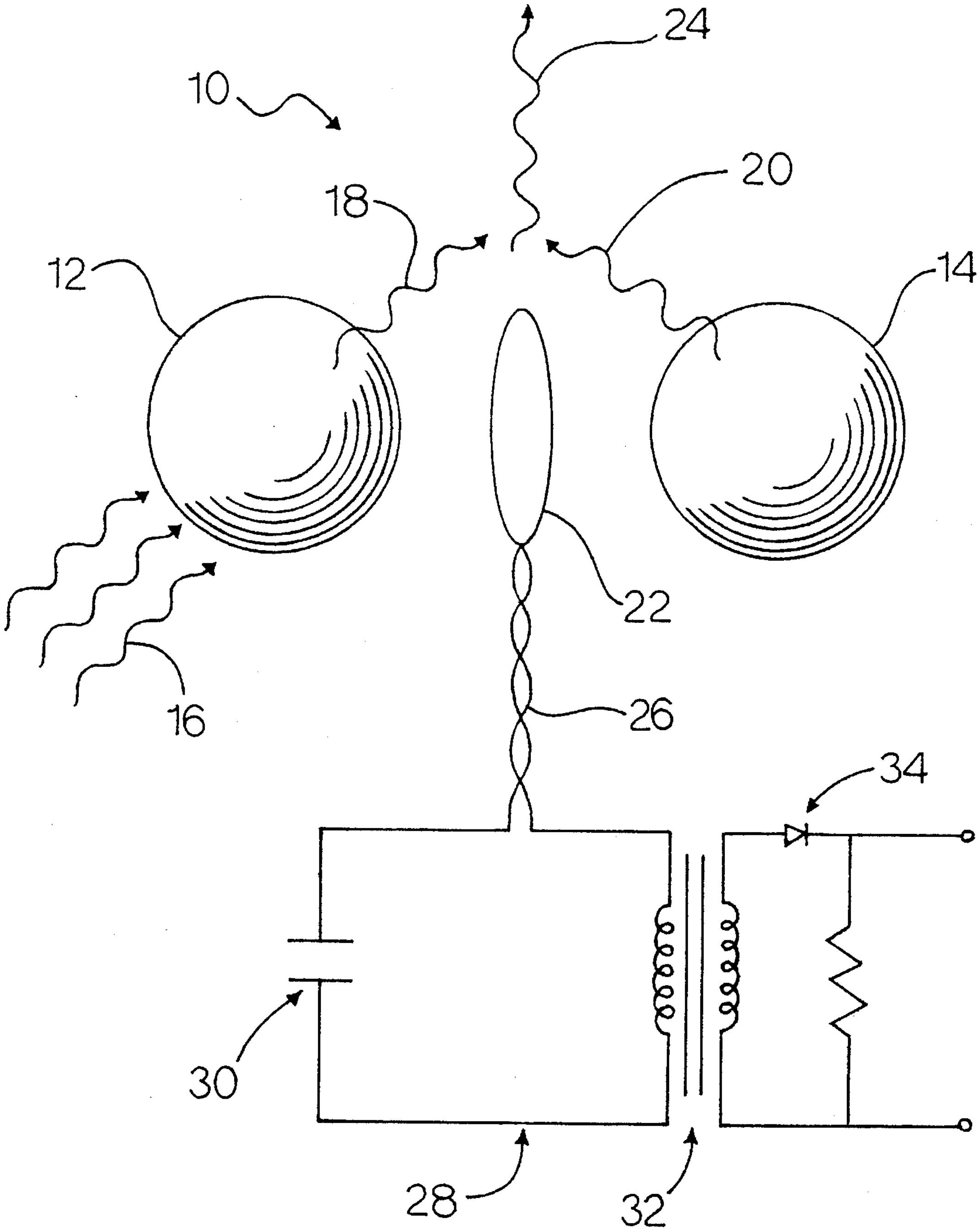
10
24
18
20
12
14
22
16
26
34
30
28
32

FIG. 2

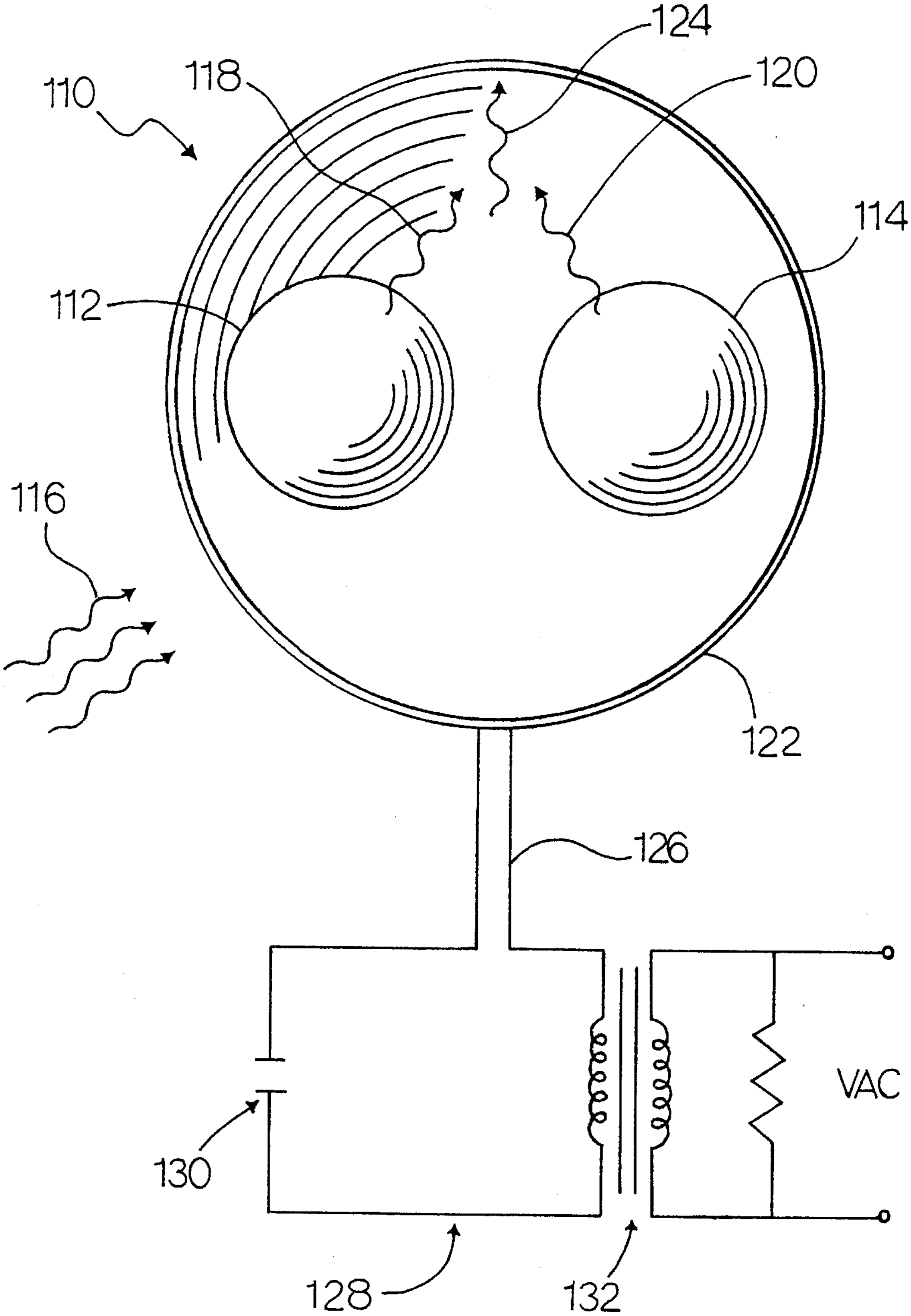
124
110
118
120
114
112
116
122
126
VAC
130
128
132

FIG. 3

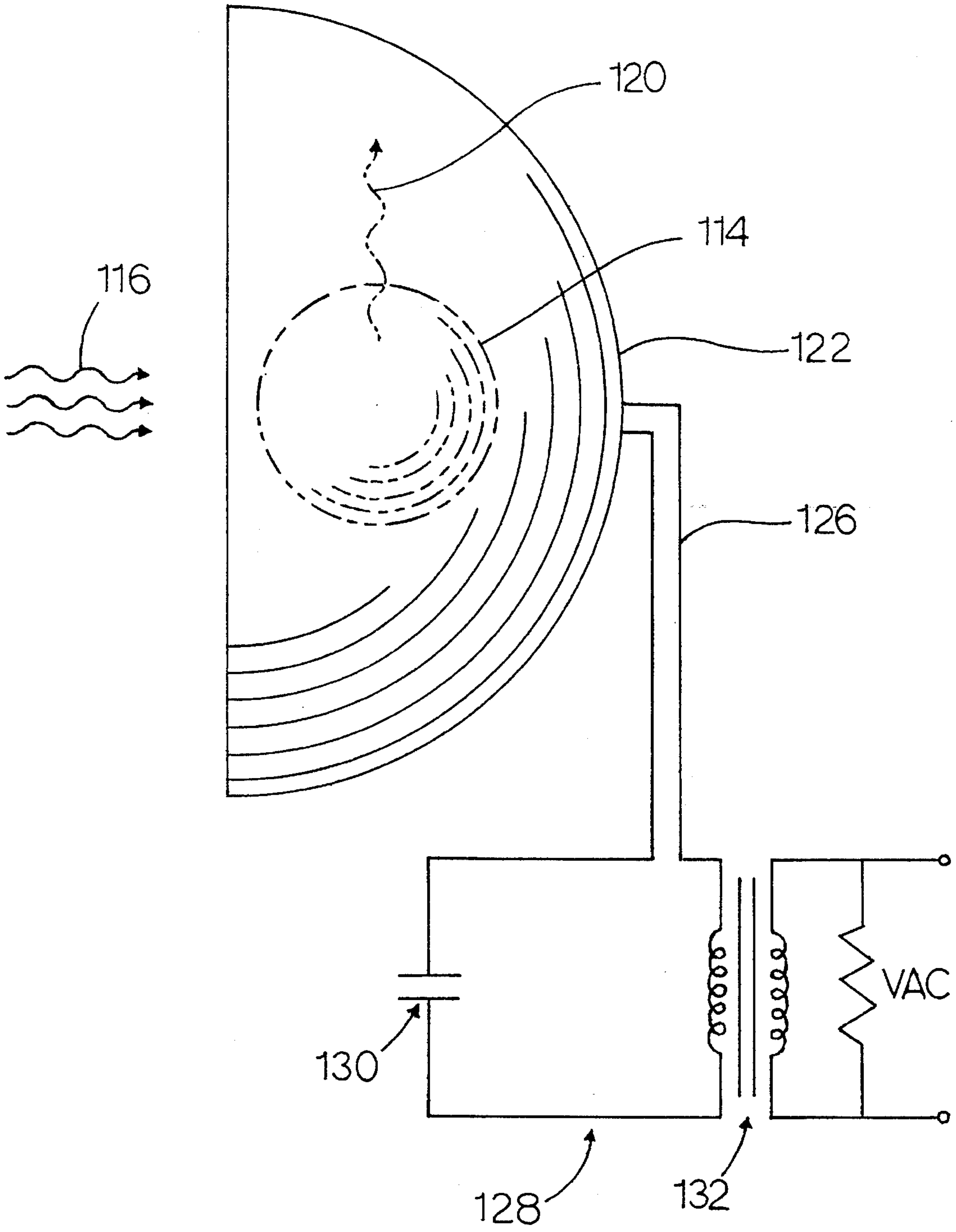
120
114
116
122
126
VAC
130
128
132

FIG. 4

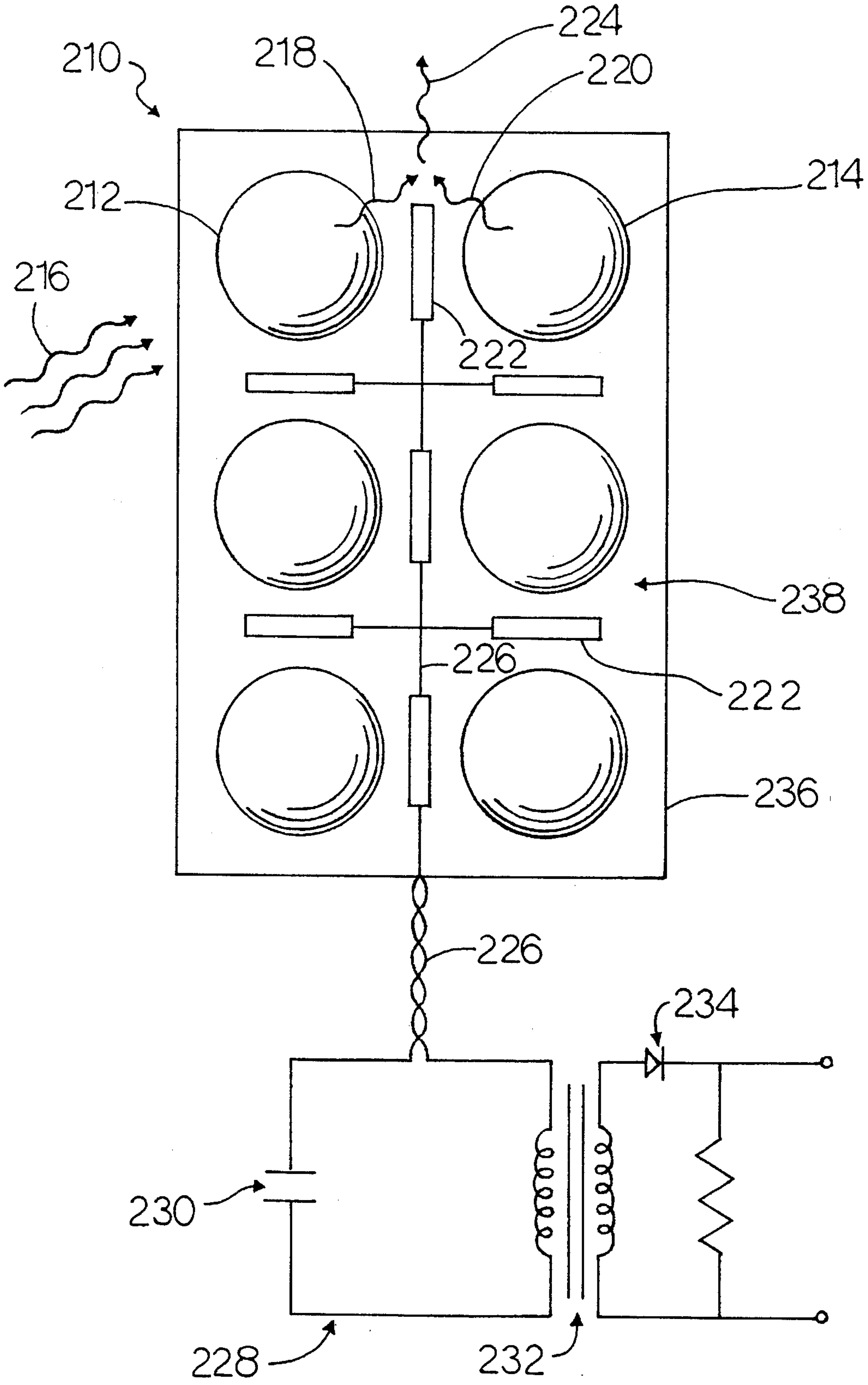

FIG. 5

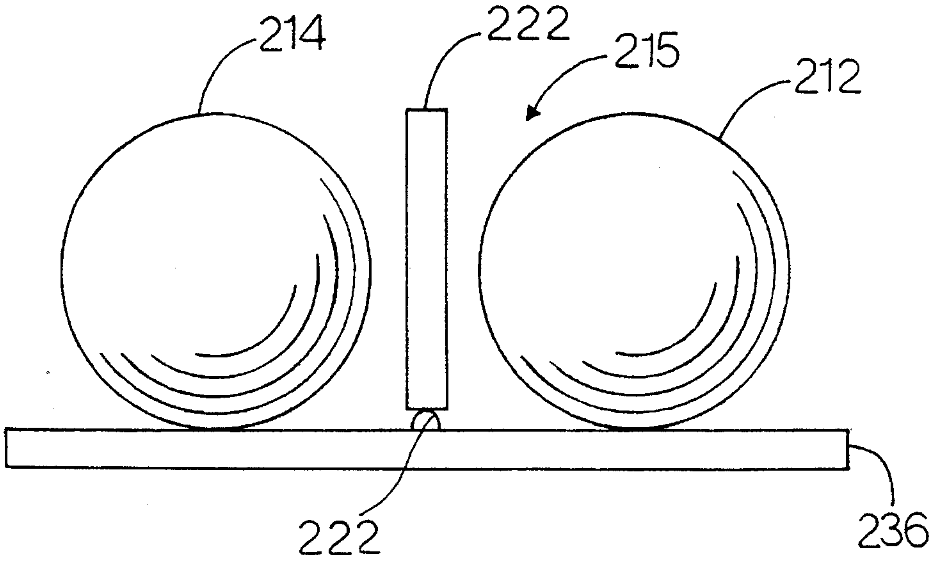
214
222
215
212
222
236

FIG. 6

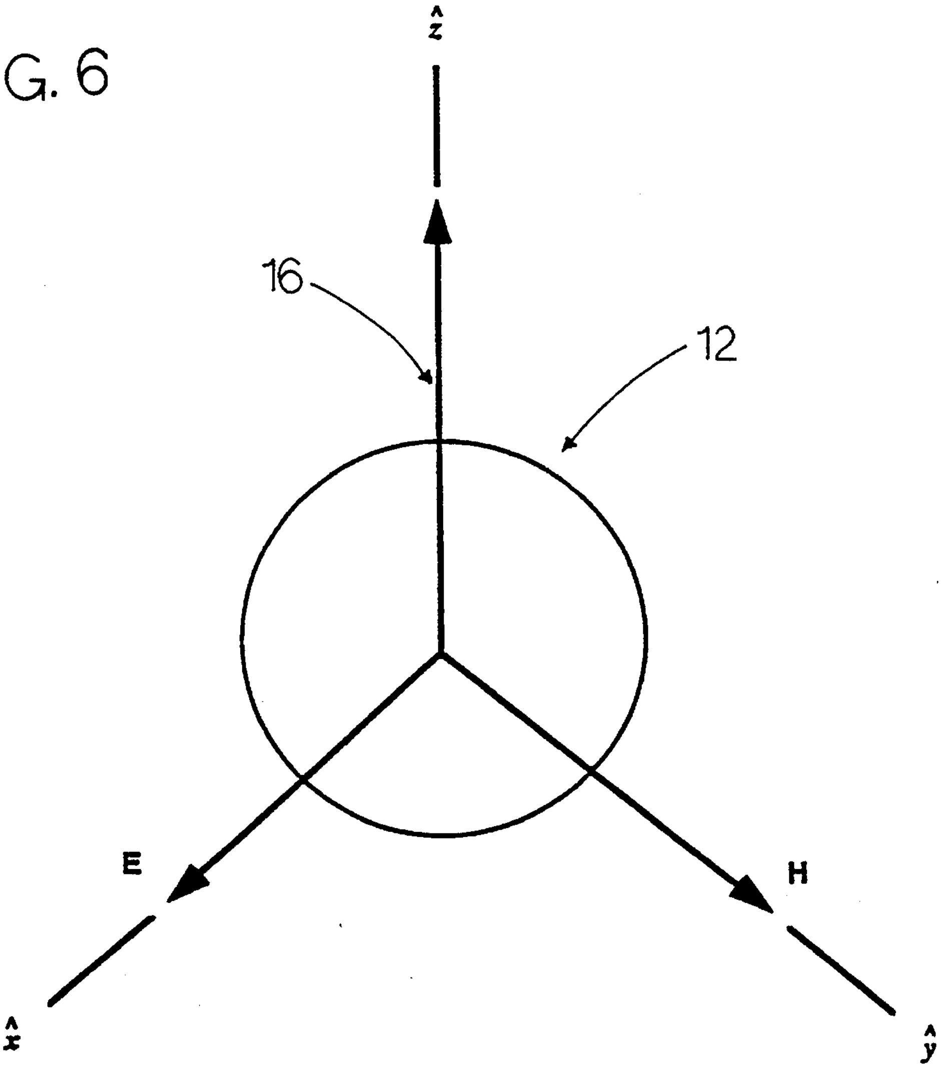
$\hat{z}$
16
12
E
H
$\hat{x}$
$\hat{y}$

FIG. 7

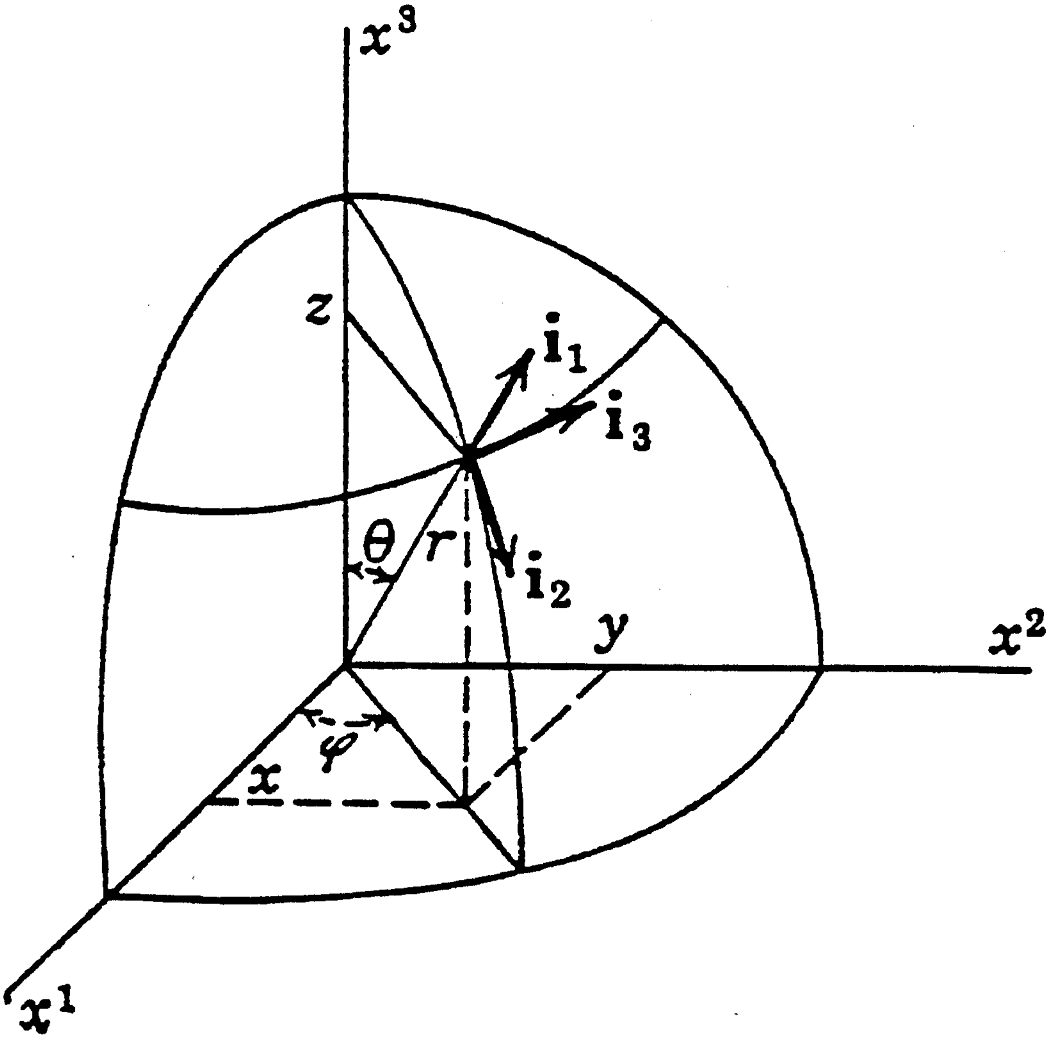
x^3
z
$\mathbf{i}_1$
$\mathbf{i}_3$
θ
r
$\mathbf{i}_2$
y
x^2
φ
x
x^1

FIG. 8

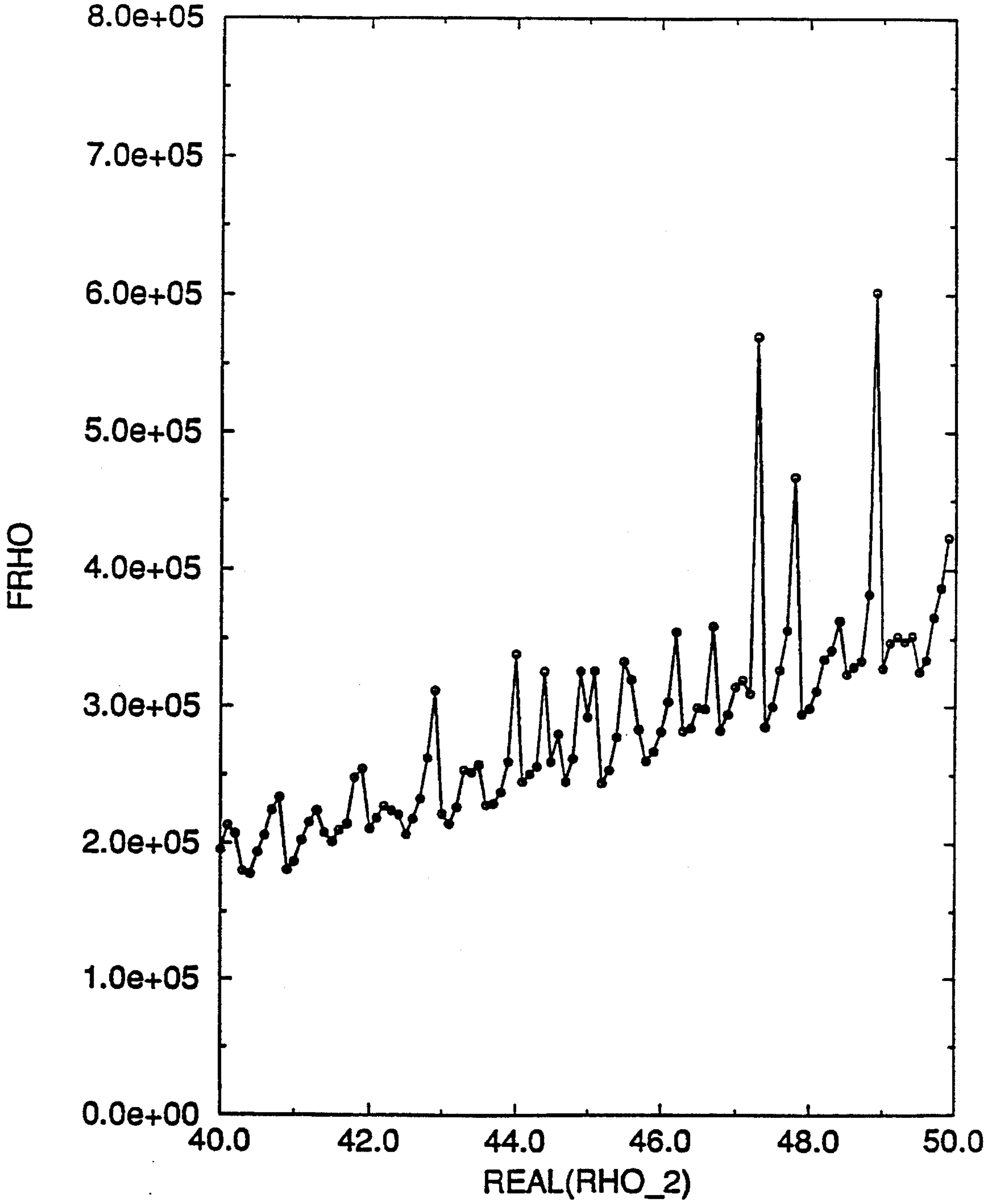
8.0e+05
7.0e+05
6.0e+05
5.0e+05
4.0e+05
3.0e+05
2.0e+05
1.0e+05
0.0e+00
FRHO
40.0
42.0
44.0
46.0
48.0
50.0
REAL(RHO_2)

FIG. 9

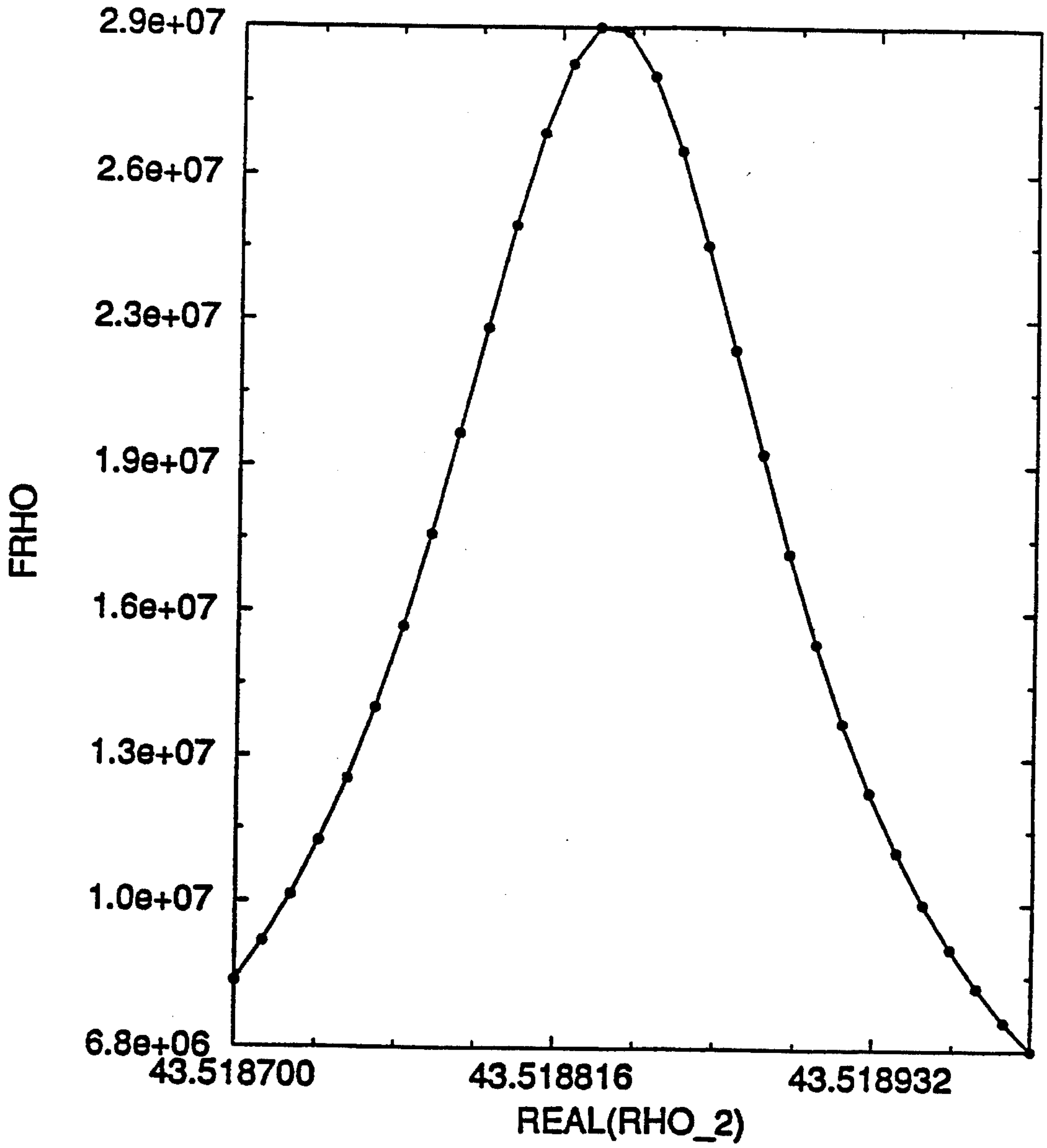
2.9e+07
2.6e+07
2.3e+07
1.9e+07
1.6e+07
1.3e+07
1.0e+07
6.8e+06
FRHO
43.518700
43.518816
43.518932
REAL(RHO_2)

SYSTEM FOR CONVERTING ELECTROMAGNETIC RADIATION ENERGY TO ELECTRICAL ENERGY

BACKGROUND OF THE INVENTION

The invention relates generally to conversion of electromagnetic radiation energy to electrical energy, and, more particularly, to conversion of high frequency bandwidths of the spectrum of a type of radiation known as zero point electromagnetic radiation to electrical energy.

The existence of zero point electromagnetic radiation was discovered in 1958 by the Dutch physicist M. J. Sparnaay. Mr. Sparnaay continued the experiments carried out by Hendrik B. G. Casimir in 1948 which showed the existence of a force between two uncharged parallel plates which arose from electromagnetic radiation surrounding the plates in a vacuum. Mr. Sparnaay discovered that the forces acting on the plates arose from not only thermal radiation but also from another type of radiation now known as classical electromagnetic zero point radiation. Mr. Sparnaay determined that not only did the zero point electromagnetic radiation exist in a vacuum but also that it persisted even at a temperature of absolute zero. Because it exists in a vacuum, zero point radiation is homogeneous and isotropic as well as ubiquitous. In addition, since zero point radiation is also invariant with respect to Lorentz transformation, the zero point radiation spectrum has the characteristic that the intensity of the radiation at any frequency is proportional to the cube of that frequency. Consequently, the intensity of the radiation increases without limit as the frequency increases resulting in an infinite energy density for the radiation spectrum. With the introduction of the zero point radiation into the classical electron theory, a vacuum at a temperature of absolute zero is no longer considered empty of all electromagnetic fields. Instead, the vacuum is now considered as filled with randomly fluctuating fields having the zero point radiation spectrum. The special characteristics of the zero point radiation which are that it has a virtually infinite energy density and that it is ubiquitous (even present in outer space) make it very desirable as an energy source. However, because high energy densities exist at very high radiation frequencies and because conventional methods are only able to convert or extract energy effectively or efficiently only at lower frequencies at which zero point radiation has relatively low energy densities, effectively tapping this energy source has been believed to be unavailable using conventional techniques for converting electromagnetic energy to electrical or other forms of easily useable energy. Consequently, zero point electromagnetic radiation energy which may potentially be used to power interplanetary craft as well as provide for society's other needs has remained unharnessed.

There are many types of prior art systems which use a plurality of antennas to receive electromagnetic radiation and provide an electrical output therefrom. An example of such a prior art system is disclosed in U.S. Pat. No. 3,882,503 to Gamara. The Gamara system has two antenna structures which work in tandem and which oscillate by means of a motor operatively attached thereto in order to modulate the radiation reflected from the antenna surfaces. The reflecting surfaces of the antennas are also separated by a distance equal to a quarter wavelength of the incident radiation. However, the Gamara system does not convert the incident radiation to electrical current for the purpose of converting the incident electromagnetic radiation to another form of readily useable energy. In addition, the relatively large size of the Gamara system components make it unable to resonate at and modulate very high frequency radiation.

What is therefore needed is a system which is capable of converting high frequency electromagnetic radiation energy into another form of energy which can be more readily used to provide power for transportation, heating, cooling as well as various other needs of society. What is also needed is such a system which may be used to provide energy from any location on earth or in space.

SUMMARY OF THE INVENTION

It is a principle object of the present invention to provide a system for converting electromagnetic radiation energy to electrical energy.

It is another object of the present invention to provide a system for converting electromagnetic radiation energy having a high frequency to electrical energy.

It is another object of the present invention to provide a system for converting zero point electromagnetic radiation energy to electrical energy.

It is another object of the present invention to provide a system for converting electromagnetic radiation energy to electrical energy which may used to provide such energy from any desired location on earth or in space.

It is another object of the present invention to provide a system for converting electromagnetic radiation energy to electrical energy having a desired waveform and voltage.

It is an object of the present invention to provide a miniaturized system for converting electromagnetic radiation energy to electrical energy in order to enhance effective utilization of high energy densities of the electromagnetic radiation.

It is an object of the present invention to provide a system for converting electromagnetic radiation energy to electrical energy which is simple in construction for cost effectiveness and reliability of operation.

Essentially, the system of the present invention utilizes a pair of structures for receiving incident electromagnetic radiation which may be propagating through a vacuum or any other medium in which the receiving structures may be suitably located. The system of the present invention is specifically designed to convert the energy of zero point electromagnetic radiation; however, it may also be used to convert the energy of other types of electromagnetic radiation. The receiving structures are preferably composed of dielectric material in order to diffract and scatter the incident electromagnetic radiation. In addition, the receiving structures are of a volumetric size selected to enable the structures to resonate at a high frequency of the incident electromagnetic radiation based on the parameters of frequency of the incident radiation and propagation characteristics of the medium and of the receiving structures. Since zero point radiation has the characteristic that its energy density increases as its frequency increases, greater amounts of electromagnetic energy are available at higher frequencies. Consequently, the size of the structures are preferably miniaturized in order to produce greater amounts of energy from a system located within a space or area of a given size. In this regard, the smaller the size of the receiving structures, the greater the amount of energy that can be produced by the system of the present invention.

At resonance, electromagnetically induced material deformations of the receiving structures produce secondary fields of electromagnetic energy therefrom which may have eva-

nescent energy densities several times that of the incident radiation. The structures are of different sizes so that the secondary fields arising therefrom are of different frequencies. The difference in volumetric size is very small so that interference between the two emitted radiation fields, and the receiving structures at the two different frequencies produces a beat frequency radiation which has a much lower frequency than the incident radiation. The beat frequency radiation preferably is at a frequency which is sufficiently low that it may be relatively easily converted to useable electrical energy. In contrast, the incident zero point radiation has its desirable high energy densities at frequencies which are so high that conventional systems for converting the radiation to electrical energy either cannot effectively or efficiently so convert the radiation energy or simply cannot be used to convert the radiation energy for other reasons.

The system of the present invention also includes an antenna which receives the beat frequency radiation. The antenna may be a conventional metallic antenna such as a loop or dipole type of antenna or a rf cavity structure which partially encloses the receiving structures. The antenna feeds the radiation energy to an electrical conductor (in the case of a conventional dipole or comparable type of antenna) or to a waveguide (in the case of a rf cavity structure). The conductor or waveguide feeds the electrical current (in the case of the electrical conductor) or the electromagnetic radiation (in the case of the waveguide) to a converter which converts the received energy to useful electrical energy. The converter preferably includes a tuning circuit or comparable device so that it can effectively receive the beat frequency radiation. The converter may include a transformer to convert the energy to electrical current having a desired voltage. In addition, the converter may also include a rectifier to convert the energy to electrical current having a desired waveform.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a plan view of the receiving structures and antenna of a first embodiment of the system of the present invention with a schematic view of the conductor and converter thereof and also showing the incident primary and emitted secondary electromagnetic radiation.

FIG. **2** is a front view of the receiving structures, antenna and waveguide of a second embodiment of the system of the present invention with a schematic view of the converter thereof and also showing the incident primary and emitted secondary electromagnetic radiation.

FIG. **3** is a perspective view of the receiving structures, antenna and waveguide of the second embodiment shown in FIG. **2** with a schematic view of the converter thereof and also showing the incident primary and emitted secondary electromagnetic radiation.

FIG. **4** is a front view of the substrate and a plurality of pairs of the receiving structures and a plurality of antennas of a third embodiment of the system of the present invention with a schematic view of the conductor and converter thereof and also showing the incident primary and emitted secondary electromagnetic radiation.

FIG. **5** is a top view of some of the components of the third embodiment of the system of the present invention showing two of the plurality of pairs of receiving structures and two of the plurality of antennas mounted on the substrate.

FIG. **6** is a diagram of a receiving structure of the system of the present invention showing an incident electromagnetic plane wave impinging on the receiving structure and illustrating the directions of the electric and magnetic field vectors thereof.

FIG. **7** is a diagram of a spherical coordinate system as used in the formulas utilized in the system of the present invention.

FIG. **8** is a graph showing an imaginary ρ parameter plotted against a real ρ parameter illustrating the values thereof at resonance as well as values thereof at other than resonance.

FIG. **9** is a graph showing a portion of the graphical representation shown in FIG. **8** illustrating the real and imaginary ρ values at or near a single resonance.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring to the drawings, a first embodiment of the present invention is generally designated by the numeral **10**. The system **10** includes a first and second means for receiving **12** and **14** incident electromagnetic radiation **16**. The means for receiving **12** and **14** are preferably a pair of spherical structures **12** and **14** which are preferably composed of a dielectric material. Alternatively, the spheres **12** and **14** may be cubical structures or any other suitable shape. The spheres **12** and **14** may be mounted on a suitable foundation by any suitable mounting means (not shown), or spheres **12** and **14** may be suspended from a suitable foundation by any suitable suspension means (not shown). The spheres **12** and **14** are preferably composed of a dielectric material. The dielectric spheres **12** and **14** scatter and concentrate electromagnetic waves. At very sharpely defined frequencies, the spheres **12** and **14** will have resonances wherein the internal energy densities can be five orders of mangitude larger than the energy density of the incident electromagnetic field driving the spheres **12** and **14**. At resonance, the electromagnetic stresses, equivalent to pressures proportional to the energy density, can cause material deformation of the spheres **12** and **14** which produce a secondary electromagnetic field. The spheres **12** and **14** are preferably positioned proximal to each other, as shown in FIG. **1**. Although the proximity of the spheres to each other will adversely affect the resonances, the very high "Q"s of the isolated-sphere resonances results in such adverse affect being relatively small. However, the proximity of the spheres **12** and **14** allows the spheres to interact electromechanically which increases the magnitude of the secondary radiation emitted therefrom.

The electromagnetic radiation incident upon the spheres **12** and **14** which drives the spheres to resonance is preferably zero point radiation **16**. However, other types of electromagnetic radiation may also be used to drive the spheres **12** and **14**, if desired.

The effect of a dielectric sphere such as **12** or **14** on an incident electromagnetic radiation such as a plane wave thereof is shown in FIG. **6**. The plane wave propagates in the z axis direction and is diffracted by the sphere **12** resulting in scattering thereof. This scattering is commonly known as Mie scattering. The incident radiation wave has an electric vector component which is linearly polarized in the x axis direction and a magnetic vector component which is linearly polarized in the y axis direction.

An electromagnetic wave incident upon a structure produces a forced oscillation of free and bound charges in synch with the primary electromagnetic field of the incident electromagnetic wave. The movements of the charges produce a

secondary electromagnetic field both inside and outside the structure. The secondary electromagnetic radiation comprising this secondary electromagnetic field is shown in FIG. **1** and designated by the numerals **18** and **20**. An antenna which is shown simply as a loop antenna but may also be a dipole or any other suitable type of antenna is also shown in FIG. **1** and designated by the numeral **22**. The nonlinear mutual interactions of the spheres produces interference between the secondary electromagnetic radiation **18** and **20** produces a beat frequency radiation **24** which is preferably at a much lower frequency than the primary radiation **16**. It is this beat frequency radiation **24** which is desired for conversion into electrical energy because it preferably is within the frequency range of rf radiation which may be converted into electrical energy by generally conventional systems. Thus, the radiation **24** received by the antenna **22** is fed via an electrical conductor **26** to a means for converting the beat frequency radiation **24** to electrical energy. This means for converting is designated by the numeral **28** and preferably includes a tuning capacitor **30** and a transformer **32** and a rectifier (preferably a diode) **34**. Instead of including the capacitor **30**, transformer **32** and rectifier **34**, the converter **28** may alternatively include an rf receiver of any suitable type.

The resultant field at any point is the vector sum of the primary and secondary fields. For the equations that follow, the structure receiving the incident plane wave is a sphere of radius a having a propagation constant k_1 positioned in an infinite, homogeneous medium having a propagation constant k_2. The incident plane wave propagates in the z axis direction and is as shown in FIG. **6**. The spherical coordinate system used for the vector spherical wave functions is shown in FIG. **7**. Expansion of the incident field provides:

$$E_i=E_o e^{-i\omega t}\sum_{n=1}^{\infty} i^n\frac{2n+1}{n(n+1)}\left(m^{(1)}_{01n}-in^{(1)}_{eln}\right)$$

$$H_i=-\frac{k_2}{\omega\mu_2}E_o e^{-i\omega t}\sum_{n=1}^{\infty} i^n\frac{2n+1}{n(n+1)}\left(m^{(1)}_{eln}+in^{(1)}_{01n}\right)$$

where E is the electric field and H is the magnetic field; and

$$\overset{(1)}{m}_{\substack{01n\\ e1n}}=\pm\frac{1}{\sin\theta}j_n(k_2R)P_n{}^1(\cos\theta)\begin{matrix}\cos\\ \sin\end{matrix}\phi i_2-j_n(k_2R)\frac{\partial p_n{}^1}{\partial\theta}\begin{matrix}\sin\\ \cos\end{matrix}\phi i_3$$

$$\overset{(1)}{n}_{\substack{0ln\\ e1n}}=\frac{n(n+1)}{k_2R}j_n(k_2R)P_n{}^1(\cos\theta)\begin{matrix}\sin\\ \cos\end{matrix}\phi i_1+\frac{1}{k_2R}[k_2Rj_n(k_2R)]'\times$$

$$\frac{\partial P_n{}^1}{\partial\theta}\begin{matrix}\sin\\ \cos\end{matrix}\phi i_2\pm\frac{1}{k_2R\sin\theta}[k_2Rj_n(k_2R)]'P_n{}^1(\cos\theta)\begin{matrix}\cos\\ \sin\end{matrix}\phi i_3.$$

The electric and magnetic fields of the incident wave transmitted into the sphere i.e., R<a, can be similarly expanded:

$$E_t=E_0e^{-i\omega t}\sum_{n=1}^{\infty} i^n\frac{2n+1}{n(n+1)}\left(a_n^{tm}\,{}^{(1)}_{01n}-ib_n^{tn}\,{}^{(1)}_{e1n}\right)$$

$$H_t=\frac{k_2}{\phi\mu_1}E_0e^{-i\omega t}\sum_{n=1}^{\infty} i^m\frac{2n+1}{n(n+1)}\left(b_n^{tm}\,{}^{(1)}_{e1n}-ia_n^{tn}\,{}^{(1)}_{01n}\right)$$

If $j_n(k_2R)$ is replaced by $h_n^{(1)}(k_2R)$ in the previous equations, the functions $m^{(1)}$ and $n^{(1)}$ become $m^{(3)}$ and $n^{(3)}$. The outgoing fields i.e., R>a, are represented by:

$$E_R=E_0e^{-i\omega t}\sum_{n=1}^{\infty} i^n\frac{2n+1}{n(n+1)}\left(a_n^{rm}\,{}^{(3)}_{01n}-ib_n^{rn}\,{}^{(3)}_{e1n}\right)$$

$$H_r=\frac{k_2}{\phi\mu_1}E_0e^{-i\omega t}\sum_{n=1}^{\infty} i^m\frac{2n+1}{n(n+1)}\left(b_n^{rm}\,{}^{(3)}_{e1n}-ia_n^{rn}\,{}^{(3)}_{01n}\right)$$

where H_r represents the resultant wave in the medium surrounding the sphere. At resonance, the values of ρ at resonance require that the $a_n{}^t$ and $b_n{}^t$ coefficients be infinite. In order to determine these values of $a_n{}^t$ and $b_n{}^t$, the boundary conditions at the sphere radius are needed. Since there must be continuity of the E and H values at the surface, the following equations are used:

$i_1\times(E_i+E_r)=i_1\times E_t$ and

$i_1\times(H_i+H_r)=i_1\times H_t$

which lead to two pairs of inhomogeneous equations:

$a_n{}^tj_n(N\rho)-a_n{}^rh_n{}^{(1)}(\rho)=j_n(\rho)$

$\mu_2a_n{}^t[N\rho j_n(N\rho)]'-\mu_1a_n{}^r[\rho h_n{}^{(1)}(\rho)]'=\mu_1[\rho j_n(\rho)]'$ and

$\mu_2Nb_n{}^tj_n(N\rho)-\mu_1b_n{}^rh_n{}^{(1)}(\rho)=\mu_1j_n(\rho)$

$b_n{}^t[N\rho j_n(N\rho)]'-Nb_n{}^r[\rho h_n{}^{(1)}(\rho)]'=N[\rho j_n(\rho)]'$

where $k_1=Nk_2$, $\rho=k_2a$, $k_1a=N\rho$. Spherical Bessel functions of the first kind are denoted by j_n, while those of the third kind are denoted by $h_n{}^{(1)}$. The resulting equations are:

$$a_n{}^t=\frac{\mu_1j_n(\rho)[\rho h_n^{(1)}(\rho)]'-\mu_1h_n^{(1)}\rho)[\rho j_n(\rho)]'}{\mu_1j_n(N\rho)[\rho h_n^{(1)}(\rho)]'-\mu_2h_n^{(1)}(\rho)[N\rho j_n(N\rho)]'}$$

and

$$b_n{}^t=\frac{\mu_1N\mathrm{j}_n(\rho)[\rho h_n^{(1)}(\rho)]'-\mu_1\mathrm{N}h_n^{(1)}\rho)[\rho j_n(\rho)]'}{\mu_2N^2j_n(N\rho)[\rho h_n^{(1)}(\rho)]'-\mu_1h_n^{(1)}(\rho)[N\rho j_n(N\rho)]'}$$

At a resonance, the denominator of either $a_n{}^t$ or $b_n{}^t$ will be zero. Thus, ρ values are found using the above equations that correspond to a resonant combination of angular frequency (ω) and radius (a) for a given sphere material and given surrounding medium. In determining such values of ρ, the following equations are also specifically used:

$$\rho=ak_2=a\omega\sqrt{\epsilon_2\mu_2}\ \text{and}$$

$$\rho_1=(k_I/k_2)\rho$$

where ρ_1 corresponds to the sphere material. An iterative method is preferably used to find the desired values of ρ at resonance. In calculating ρ utilizing the above equations for purposes of example, it was assumed that $\mu_1=\mu_2=\mu_0=4\pi\times10^{-7}$ and $\epsilon_2=\epsilon_0=8.85419\times10^{-12}$.

One major root of ρ which was found has a value of:

Real (ρ)=+66.39752607619131

Imaginary (ρ)=−0.6347867071968998.

These particular values are not shown in FIG. **8**. However, other values of ρ found using the equations set forth herein are shown in FIG. **8**. The peaks in FIG. **8** are the resonances. One of these resonances shown in FIG. **8** is shown in detail in FIG. **9**. These resonance values are shown for purposes of example. Other resonances also exist which have not been determined; thus, not all possible resonance values are shown in FIGS. **8** and **9**.

Calculation of these values also allows the determination of a possible am combination which would have these root values. For ρ, ϵ (epsilon)=ϵ_0 and $\mu=\mu_0$, and

$$\rho = a\omega\sqrt{\epsilon_0\mu_0} = a\omega/c.$$

Expressed in SI units, the speed of light $c=2.99792458\times10^{14}$ m/s. If an a value of 10^{-6} m is assumed for the examples shown herein, then:

$$\omega=\rho c/a\approx1.9919\times10^{16}-i1.9044\times10^{14} \text{ radians/s.}$$

This is an example of the angular frequency required within the impingent EM radiation in order to create a resonant situation. Examples of other resonances were indicated, and these are shown in FIG. **8**. No complex-frequency plane waves exist. Therefore, the calculations were made by considering only the real portion of the above root and setting the imaginary portion equal to zero. However, upon doing this, the iterative calculation procedure becomes insensitive to any root in the vicinity of the root's real portion. In the iterative calculation procedure, initially a range of ρ values is input into the equations. These ρ values are in the neighborhood of the prospective root. A range of ρ values is subsequently studied to find any imaginary ρ i.e., fρ (a function of ρ), peaks in that range. Next, once a peak has been chosen, the function order n giving the dominant fρ is determined. This also gives a clue as to whether the peak is due to a magnetic resonance (a_n approaches infinity) or an electrical resonance (b_n approaches infinity). A large number of Newton-Raphson iterations is preferably performed in order to converge upon a root ρ value.

FIGS. **2** and **3** show a second embodiment of the present invention generally designated by the numeral **110**. Embodiment **110** is essentially the same as embodiment **10** except that the antenna is a rf cavity structure **122** which feeds the received beat frequency radiation **124** to a waveguide **126**. Embodiment **110** also preferably includes two spheres **112** and **114** which receive the primary incident electromagnetic radiation **116** and emit the secondary electromagnetic radiation **118** and **120**. As with the spheres **18** and **20** of embodiment **10**, spheres **118** and **120** are preferably composed of a dielectric material. Embodiment **110** also includes converter **128**, capacitor **130**, transformer **132** and rectifier **134** which are essentially identical to the correspondingly numbered elements of embodiment **10**. Therefore, a description of these components of embodiment **110** will not be repeated in order to promote brevity. In addition, the same equations and method of calculation set forth above with regard to embodiment **10** also apply to embodiment. Therefore, their description will not be repeated in order to promote brevity.

FIGS. **4** and **5** show a third embodiment of the present invention generally designated by numeral **210**. Embodiment **210** is essentially identical to the first embodiment **10** except that the embodiment **210** includes a plurality of pairs **215** of receiving means (spheres) **212** and **214** mounted on a substrate **236**. The spheres **212** and **214** are thus in the form of an array **238**. The pairs **215** of the array **238** are preferably positioned proximal to each other in order to maximize the amount of energy extracted from a particular area or space of a given size. Since, as set forth hereinabove, the energy density of the zero point radiation increases as the frequency of the radiation increases, it is desirable that the spheres resonate at as high a bandwidth of frequencies as possible. Because the spheres **212** and **214** must be small in direct proportion to the wavelength of the high frequencies of the incident electromagnetic radiation **216** at which resonance is desirably obtained, the spheres **212** and **214** are preferably microscopic in size. Current lithographic techniques are capable of manufacturing such microscopically small spheres mounted on a suitable substrate thereby providing a suitably miniaturized system **210**. A miniaturized system enhances the energy output capability of the system by enabling it to resonate at higher frequencies at which there are correspondingly higher energy densities. Consequently, utilization of array **238** in the system **210** enhances the maximum amount of electrical energy provided by the system **210**.

Lithographic techniques may be more amenable to manufacturing microscopically small receiving structures **212** and **214** which may be disc shaped, semispherical or have another shape other than as shown in FIGS. **4** and **5**. Consequently, the receiving means **212** and **214** may accordingly have such alternative shapes rather than the spherical shape shown in FIGS. **4** and **5**. In addition, a large number of small spheres may be manufactured by bulk chemical reactions. Packing a volume with such spheres in close proximity could enhance the output of energy.

Embodiment **210** also includes a plurality of antennas **222** positioned preferably between the spheres **212** and **214** which receive the beat frequency radiation **224** produced by the interference between the secondary radiation **218** and **220**. The antennas **222** are shown as loop antennas **222** but may be any other suitable type of antennas as well.

Embodiment **210** has a plurality of electrical conductors **226** which preferably include traces mounted on the substrate **236** which occupies a finite volume. The electrical conductors **226** feed the electrical output from the antennas **222** to a suitable converter **228** which preferably includes tuning capacitor **230**, transformer **232** and rectifier **234**, as with embodiments **10** and **110**. Except as set forth above, the components of embodiment **210** are identical to embodiment **10** so the detailed description of these components will not be repeated in order to promote brevity. In addition, the same equations and method of calculation set forth above for embodiment **10** also apply to embodiment **210**. Therefore, the description of these equations and method of calculation will not be repeated in order to promote brevity.

Accordingly, there has been provided, in accordance with the invention, a system which converts high frequency zero point electromagnetic radiation into electrical energy effectively and efficiently and thus fully satisfies the objectives set forth above. It is to be understood that all terms used herein are descriptive rather than limiting. Although the invention has been specifically described with regard to the specific embodiments set forth herein, many alternative embodiments, modifications and variations will be apparent to those skilled in the art in light of the disclosure set forth herein. Accordingly, it is intended to include all such alternatives, embodiments, modifications and variations that fall within the spirit and scope of the invention as set forth in the claims hereinbelow.

What is claimed is:

1. A system for converting incident electromagnetic radiation energy to electrical energy, comprising:

a first means for receiving incident primary electromagnetic radiation, said means for receiving producing emitted secondary electromagnetic radiation at a first frequency, said first means for receiving having a first volumetric size selected to resonate at a frequency within the frequency spectrum of the incident primary electromagnetic radiation in order to produce the secondary electromagnetic radiation at the first frequency at an enhanced energy density;

a second means for receiving the incident primary electromagnetic radiation, said means for receiving producing emitted secondary electromagnetic radiation at a second frequency, the secondary radiation at the first frequency and the secondary radiation at the second frequency interfering to produce secondary radiation at a lower frequency than that of the incident primary radiation, said second means for receiving having a second volumetric size selected to resonate at a frequency within the frequency spectrum of the incident primary electromagnetic radiation in order to produce the emitted secondary electromagnetic radiation at the second frequency at an enhanced energy density;

an antenna for receiving the emitted secondary electromagnetic radiation at the lower frequency, said antenna providing an electrical output responsive to the secondary electromagnetic radiation received;

a converter electrically connected to said antenna for receiving electrical current output from said antenna and converting the electrical current output to electrical current having a desired voltage and waveform.

2. The system of claim **1** wherein:

said first means for receiving is composed of a dielectric material; and

said second means for receiving is composed of a dielectric material.

3. The system of claim **1** wherein:

said first means for receiving is spherical; and

said second means for receiving is spherical.

4. A system for for converting incident zero point electromagnetic radiation energy to electrical energy, comprising:

a first means for receiving incident primary zero point electromagnetic radiation, said means for receiving producing emitted secondary electromagnetic radiation at a first frequency;

a second means for receiving the incident primary zero point electromagnetic radiation, said means for receiving producing emitted secondary electromagnetic radiation at a second frequency, the secondary radiation at the first frequency and the secondary radiation at the second frequency interfering to produce secondary radiation at a beat frequency which is lower than that of the incident primary radiation;

an antenna for receiving the emitted secondary electromagnetic radiation at the lower frequency, said antenna providing an electrical output responsive to the secondary electromagnetic radiation received;

means for transmitting the emitted secondary electromagnetic radiation at the beat frequency from said antenna, said means for transmitting connected to said antenna;

a converter connected to said means for transmitting for receiving the emitted secondary electromagnetic radiation at the beat frequency from said antenna and converting the same to electrical current having a desired voltage and waveform.

5. The system of claim **4** wherein:

said first means for receiving has a first volumetric spherical size selected to resonate in response to the incident primary electromagnetic radiation in order to produce the secondary electromagnetic radiation at the first frequency at an enhanced energy density; and

said second means for receiving has a second volumetric spherical size selected to resonate in response to the incident primary electromagnetic radiation in order to produce the emitted secondary electromagnetic radiation at the second frequency at an enhanced energy density, said first and second volumetric sizes selected based on parameters of propagation constant of said first and second means for receiving, propagation constant of medium in which said first and second means for receiving are located and frequency of the incident primary electromagnetic radiation.

6. The system of claim **5** wherein the first and second volumetric sizes are selected by utilizing the formulas:

$$a_n{}^t=\frac{\mu_1 j_n(\rho)[\rho h_n^{(1)}(\rho)]'-\mu_1 h_n^{(1)}\rho)[\rho j_n(\rho)]'}{\mu_1 j_n(N\rho)[\rho h_n^{(1)}(\rho)]'-\mu_2 h_n^{(1)}(\rho)[N\rho j_n(N\rho)]'}$$

$$b_n{}^t=\frac{\mu_1 N\mathrm{j}_n(\rho)[\rho h_n^{(1)}(\rho)]'-\mu_1 \mathrm{N} h_n^{(1)}\rho)[\rho j_n(\rho)]'}{\mu_2 N^2 j_n(N\rho)[\rho h_n^{(1)}(\rho)]'-\mu_1 h_n^{(1)}(\rho)[N\rho j_n(N\rho)]'}$$

$$\rho=a\omega\sqrt{\epsilon_2\mu_2}$$

wherein at a resonance, the denominator of either equation for $a_n{}^t$ or $b_n{}^t$ will be approximately zero and wherein k_1=propagation constant of the means for receiving, k_2=propagation constant of medium through which the incident electromagnetic radiation propagates, a is the radius of either means for receiving, $N=k_1/k_2$, $\rho=k_2a$, $k_1a=N\rho$, $a_n{}^t$=magnitude of oscillations of the electric field of the nth order, $b_n{}^t$=magnitude of oscillations of the magnetic field of the nth order, ω=angular frequency of the incident electromagnetic radiation, ϵ is the permittivity of the material or medium and μ is the permeability of the material or medium.

7. The system of claim **6** wherein the radius of the first means for receiving is different from the radius of the second means for receiving, difference between the radius of said first means for receiving and the radius of said second means for receiving selected so that the beat frequency resulting from the difference is a frequency which facilitates conversion of the beat frequency electromagnetic radiation to electrical energy.

8. The system of claim **4** wherein:

said first means for receiving is composed of a dielectric material; and

said second means for receiving is composed of a dielectric material.

9. The system of claim **4** wherein:

said first means for receiving is spherical; and

said second means for receiving is spherical.

10. The system of claim **4** wherein said antenna is positioned generally between said first and second means for receiving.

11. The system of claim **4** wherein said antenna is a loop antenna.

12. The system of claim **4** wherein said antenna is a generally concave shell partially enclosing said first and second means for receiving.

13. The system of claim **4** wherein said means for transmitting is a waveguide.

14. A system for for converting incident zero point electromagnetic radiation energy to electrical energy, comprising:

a substrate;

a plurality of pairs of first means for receiving incident primary zero point electromagnetic radiation and second means for receiving incident primary zero point electromagnetic radiation, said plurality of pairs of means for receiving mounted on said substrate, said first means for receiving producing emitted secondary

electromagnetic radiation at a first frequency, said second means for receiving the incident primary zero point electromagnetic radiation producing emitted secondary electromagnetic radiation at a second frequency, the secondary radiation at the first frequency and the secondary radiation at the second frequency interfering to produce secondary radiation at a beat frequency which is lower than that of the incident primary radiation, said first means for receiving having a first volumetric size selected to resonate in response to the incident primary electromagnetic radiation in order to produce the secondary electromagnetic radiation at the first frequency at an enhanced energy density, and said second means for receiving having a second volumetric size selected to resonate in response to the incident primary electromagnetic radiation in order to produce the emitted secondary electromagnetic radiation at the second frequency at an enhanced energy density, said first and second volumetric sizes selected based on parameters of propagation constant of said first and second means for receiving, propagation constant of medium in which said first and second means for receiving are located and frequency of the incident primary electromagnetic radiation, said first and second volumetric sizes being different from each other;

a plurality of antennas for receiving the emitted secondary electromagnetic radiation at the lower frequency, said antenna providing an output responsive to the secondary electromagnetic radiation received, said plurality of antennas mounted on said substrate, each of said plurality of antennas receiving the emitted secondary electromagnetic radiation of one of said pairs of first and second means for receiving;

means for transmitting the emitted secondary electromagnetic radiation at the beat frequency from said antenna, said means for transmitting connected to said plurality of antennas;

a converter connected to said means for transmitting for receiving the emitted secondary electromagnetic radiation at the beat frequency from said antenna and converting the same to electrical current having a desired voltage and waveform.

* * * * *

US005676977A

United States Patent [19]
Antelman

[11] **Patent Number: 5,676,977**
[45] **Date of Patent: Oct. 14, 1997**

[54] METHOD OF CURING AIDS WITH TETRASILVER TETROXIDE MOLECULAR CRYSTAL DEVICES

[75] Inventor: **Marvin S. Antelman**, Rehovot, Israel

[73] Assignee: **Antelman Technologies Ltd.**, Providence, R.I.

[21] Appl. No.: **658,955**

[22] Filed: **May 31, 1996**

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 310,859, Sep. 22, 1994, abandoned.

[51] **Int. Cl.**^{6} **A61K 33/38**
[52] **U.S. Cl.** **424/618**; 514/495
[58] **Field of Search** 424/618; 514/495

[56] References Cited

U.S. PATENT DOCUMENTS

4,415,565	11/1983	Wysor	424/618
4,915,955	4/1990	Gömöri	424/616
4,952,411	8/1990	Fox, Jr. et al.	424/618
5,073,382	12/1991	Antelman	424/604
5,078,902	1/1992	Antelman	424/618
5,089,275	2/1992	Antelman	424/602
5,211,855	5/1993	Antelman	424/618
5,223,149	6/1993	Antelman	424/618
5,336,499	8/1994	Antelman	424/405
5,571,520	11/1996	Antelman	424/618

OTHER PUBLICATIONS

"Is The AIDS Virus A Science Fiction?" by Peter H. Duesberg and Bryan J. Ellison, *Policy Review*, Summer 1990, pp. 40–51.

Primary Examiner—Amy Hulina
Attorney, Agent, or Firm—Salter & Michaelson

[57] ABSTRACT

The diamagnetic semiconducting molecular crystal tetrasilver tetroxide (Ag_4O_4) is utilized for destroying the AIDS virus, destroying AIDS synergistic pathogens and immunity suppressing moieties (ISM) in humans.

A single intravenous injection of the devices is all that is required for efficacy at levels of about 40 PPM of human blood. The device molecular crystal contains two mono and two trivalent silver ions capable of "firing" electrons capable of electrocuting the AIDS virus, pathogens and ISM. When administered into the bloodstream, the device electrons will be triggered by pathogens, a proliferating virus and ISM, and when fired will simultaneously trigger a redox chelation mechanism resulting in divalent silver moieties which chelate and bind active sites of the entities destroying them. The devices are completely non-toxic. However, they put stress on the liver causing hepatomegaly, but there is no loss of liver function.

3 Claims, No Drawings

METHOD OF CURING AIDS WITH TETRASILVER TETROXIDE MOLECULAR CRYSTAL DEVICES

This application is a continuation-in-part of patent application Ser. No. 08/310,859 filed Sep. 22, 1994, now abandoned.

BACKGROUND OF THE INVENTION

The present invention relates to the employment of molecular crystals as anti-AIDS devices, but more particularly to the molecular crystal semiconductor tetrasilver tetroxide Ag_4O_4 which has two monovalent and two trivalent silver ions per molecule, and which through this structural configuration enables intermolecular electron transfer capable of killing viruses and binding them to the resulting silver entity so that a single intravenous injection will completely obliterate acquired immune deficiency syndrome (AIDS) in humans. Furthermore, said devices are capable of killing pathogens and purging the bloodstream of immune suppressing moieties (ISM) whether or not created by the AIDS virus (HIV); so as to restore the immune system.

The present invention is based on concepts previously elucidated in applicant's U.S. Pat. No. 5,336,499 which discloses the destruction and inhibition of bacteria, algae and the AIDS virus in nutrient life supporting systems by using said silver oxide devices. Example 3 of said patent discloses that 18 PPM of said crystal devices could totally suppress the AIDS virus (page 6, line 5). Subsequent to the filing of the aforementioned patent, further testing revealed complete 100% destruction of the AIDS virus in vitro at 20 PPM, and the fact that said devices were harmless when ingested and inhaled, being non-toxic.

Encouraged by these evaluations and successes, applicant obtained permission to evaluate the crystals in vitro against murine acquired immune deficiency syndrome (MAIDS). Only one facility in the State of Israel is licensed for these evaluations, namely, the Kaplan Hospital in Rehovot, Israel, which is affiliated with the Hebrew University-Hadassah Medical School where said evaluations were done.

The initial evaluations entailed experimenting with various silver moieties cited in applicant's aforementioned patent, concentrations, non-reactive buffers and modes of administration. After about 18 months of judicious efforts and initial failures, success was finally achieved in destroying the MAIDS virus in C57BL mice with a single intravenous injection. The results of this test program comprise Example 5 of U.S. Pat. No. 5,336,499. After success with mice, the inventor was able to test the efficacy of said devices on two select etiological groups of terminal AIDS patients in a clinic in Tegucigalpa, Honduras, Central America.

The AIDS patients comprised the etiological subgroups, Candidiasis and Wasting Syndrome. Current indicator diseases for diagnosing AIDS which have been expanded by the CDC, fall into the following five major categories with the approximate percent distribution among AIDS patients:

1. P. carinii pneumonia	51%
2. Wasting syndrome	19%
3. Candidiasis	13%
4. Kaposi's sarcoma	11%
5. Dementia	6%

This invention concerns itself with the treatment and cure of candidiasis and wasting syndrome AIDS patients with Tetrasil*. These two groups account for approximately one third of AIDS cases.

*Trademark of Holipharm Corporation (of Israel) for Ag_4O_4

Stedman's Medical Dictionary (Williams & Wilken's 26th Ed., 1995) defines wasting syndrome "as a condition of 10% weight loss in conjunction with diarrhea or fever . . . Associated with AIDS (p. 1744)."

OBJECTS OF THE INVENTION

The main object of the invention is to provide for a molecular scale device of a single tetrasilver tetroxide crystalline molecule capable of restoring the immunity of AIDS afflicted humans of the two AIDS etiological subgroups, candidiasis and wasting syndrome.

Another object of the invention is to provide for immunity restoration in said AIDS afflicted humans through a single injection.

Another object of this invention is to destroy ISM in humans manifesting AIDS diseases of said AIDS etiological subgroups irrespective as to whether said ISM was HIV induced, since it is known that humans may manifest AIDS and still be HIV negative, and thus restore the immune system in said humans.

Another object of this invention is to destroy the AIDS virus when present in the systems of said AIDS afflicted humans.

SUMMARY OF THE INVENTION

This invention relates to a molecular scale device not only capable of destroying the AIDS virus, but of purging the human bloodstream of pathogens and restoring immunity to AIDS patients of the candidiasis and wasting syndrome categories. Said molecular device consists of a single crystal of tetrasilver tetroxide (Ag_4O_4). The crystal lattice of this molecule has a unique structure since it is a diamagnetic semiconducting crystal containing two mono and two trivalent silver ions, which in effect are capable of "firing" electrons under certain conditions which will destroy AIDS viruses, other pathogens and immune suppressing moieties (ISM), not only through the electrocution mode, but also by a binding process which occurs simultaneously with electron firing, namely, binding and chelation of divalent silver, i.e., the resulting product of the electron transfer redox that occur when the monovalent silver ions are oxidized and the trivalent ions are reduced in the crystal. The binding/chelation effect occurs at active sites of the AIDS virus, pathogens and ISM. Because of the extremely minute size of a single molecule of this crystal, several million of these devices may be employed in concert to destroy a virus colony to purge a life support system of ISM and pathogens with the consumption of only parts per trillion of the crystal devices. Thus an optimum of 40 PPM of the devices by weight of human blood was found to be sufficient to completely obliterate AIDS. This concentration is slightly over double of the optimum concentration recommended in applicant's aforementioned U.S. patent for the destruction of the human AIDS virus in vitro. Other details concerning the structure of the crystal and its mechanism against pathogens, the AIDS virus and ISM would analogously hold here, and have already been further elucidated in said patent.

The actual destruction of pathogens, ISM and the AIDS virus is effectuated by injection of a suspension of these devices in distilled or deionized water with a non-reacting electrolyte directly, i.e. intravenously, into the bloodstream. A single injection is all that is required under these conditions. Accordingly, humans injected in this manner, upon being inspected after three weeks or more had elapsed and compared with similar humans that had been given placebos, were completely cured of AIDS. The control group still manifested AIDS. Accordingly, the tetrasilver tetroxide device performed in concert with and in full conformity with the ultimate objects of this invention. Furthermore, three out of four wasting syndrome terminal patients and four out of the five candidiasis terminal patients were still alive in 1995 after a year and a half had elapsed from their initial injection. By that time all the AIDS patients had been released from the clinic and allowed to return home.

Other objects and features of the present invention shall become apparent to those skilled in the art when the present invention is considered in view of the accompanying examples. It should, of course, be recognized that the accompanying examples illustrate preferred embodiments of the present invention and are not intended as a means of defining the limits and scope of the present invention.

EXAMPLE 1

Five patients afflicted with AIDS of the candidiasis etiological category were segregated for Tetrasil treatment. The rationale for selecting them was based on facts presented in an article by Peter H. Duesberg and Brian J. Ellison entitled "Is The AIDS Virus A Science Fiction?" (*Policy Review*, Summer 1990 pp. 40–51). Only the factual presentations of the article were utilized and the hypothesis of the authors was ignored. The facts presented in the article related to the method of selecting AIDS patients based on the five aforementioned etiological subgroups targeted by the CDC, and the evidence presented, that there is AIDS without HIV as well as with it so that an anti-viral agent in most instances will not necessarily restore the immunity system.

Evaluations with Tetrasil were conducted on AIDS patients at Lucha Contra el Sida, Comayaguela, Honduras. The patients two weeks prior to inoculation were removed from their AZT, AIDS therapy. Tetrasil was administered at approximately 40 PPM of blood volume per patient as a suspension in a proprietary buffer solution (pH=6.5), supplied by Holipharm Corporation.

The results of evaluations with candidiasis are tabulated in Table I under its disease category. All patients evaluated were terminal. Some, however, were in moderate (m) condition and others in poor (p) as designated in the Table. The I and F designations refer to initial and final values as shown. WBC indicates white cell blood count. The H column, following CD 8, indicates whether hepatomegaly occurred. This was an unfortunate consequence of the treatment which resulted in enlarged livers in all patients except the second one. Despite hepatomegaly, there was no interference with liver function.

The onset of hepatomegaly was not spontaneous and varied from patient to patient, being in the range of 4–16 days.

It should also be noted that shortly after injection of Tetrasil there were indications of fever (symbolized by T in the Ag_4O_4 column), sometimes accompanied by fatigue (F). The body temperature was invariably 38.5° C. (101.3° F.). This was indicative of restoration of the immune response of the body, since normally the body will destroy pathogens when the immune system is functional by raising the temperature. The patient who died; first responded favorably to Diflucan, which previously gave no response. He was cured of his candidiasis, but unfortunately succumbed to his previous body damage. All the other candidiasis syndrome people who previously did not respond to the indicated medications subsequently responded after the Tetrasil treatment. Further evidence of the recovery of the AIDS patients manifested itself 30 days after the initial injection when white blood cell counts were taken. They are shown in Table I under the WBC column, which gives the initial and final WBC. All candidiasis patients showed a dramatic increase in their white blood cell counts, indicative of the restoration of their immunity systems.

EXAMPLE 2

The above protocol of Example 1 was repeated with AIDS patients exhibiting wasting syndrome. The results of their treatment are tabulated in Table I under the disease category of said syndrome. It should be noted that two of the four wasting syndrome patients showed improved white blood counts. The female patient, whose condition improved from poor and terminal to be among the living, showed a decrease in the WBC. However, she showed an increase in body temperature which was indicative of immune response. The test results indicate that one cannot rely on a single factor to indicate the demise of AIDS. The usual HIV marker CD 4 initial and final are irrelevant. ISM suppression appears to be more critical than the destruction of HIV. AIDS was suppressed, any permanent damage that had been done to the patients in the course of their succumbing to AIDS was not obviously cured or corrected by said crystal device treatment, rather said injury persisted and the patient was improved with respect to AIDS but still suffered from said permanent injury or impairment previously inflicted.

TABLE I

Response of AIDS Patients to Single 40 PPM Ag_4O_4 Inoculation

DISEASE	PATIENT			Date Inoc.	WBC		CD 4				DEATH	Weight Lbs.		
Group	Sex	Age	Medictn	1994	I	F	I	F	CD 8	H	1944	I	F	Ag_4O_4
Candidiasis	M p	28	Diflucan	5/5	1,200	4,200	41	—	221	+	6/11	82	76	T
	F m	33	"	5/5	6,000	6,700	554	872	394	–		98	98	T
	F m	33	Ketaconzl	5/27	2,600	3,850	248	181	951	+		123	123	T
	M p	62	"	6/2	3,300	3,700	89	237	59	+		105	92	F
	F m	31	Pentamidn	6/2	2,400	3,050	9	181	65	+		121	118	Pain
Wasting Syndrome	M m	27		5/27	3,600	4,600	39	14	709	+		119	120	T
	M m	28		5/27	2,750	—	10	—	60	+	7/19	121	119	T, F
	F p	43		5/27	3,600	2,700	68	246	248	+		101	98	T, F
	M m	19		5/10	3,850	5,400	137	36	48	+		103	106	T, F

As this invention may be embodied in several forms without departing from the spirit or essential characteristics thereof, the present embodiments are therefore illustrative and not restrictive, since the scope of the invention is defined by the appended claims rather than by the description preceding them, and all changes that fall within the metes and bounds of the claims or that form their functional as well as conjointly cooperative equivalents, are therefore intended to be embraced by these claims.

What is claimed is:

1. A method of treating AIDS-afflicted humans comprising injecting a multitude of tetrasilver tetroxide molecular crystals into the bloodstream of the human subject.

2. A method for increasing white blood cell counts in AIDS-afflicted humans comprising injecting a multitude of tetrasilver tetroxide molecular crystals into the bloodstream of the human subject.

3. Methods of treating AIDS-affilicted humans according to claims **1–2** where the concentration of said molecular crystals is approximately 40 PPM of the total blood weight of the human subject.

* * * * *

US005852993A

United States Patent [19]

Anderson

[11] **Patent Number: 5,852,993**

[45] **Date of Patent: Dec. 29, 1998**

[54] **FUEL SYSTEM FOR INTERNAL COMBUSTION SYSTEM AND ADAPTER FOR USE IN SAME**

[75] Inventor: **Herman P. Anderson**, Brentwood, Tenn.

[73] Assignee: **Herman P. Anderson Technologies, LLC**, Brentwood, Tenn.

[21] Appl. No.: **905,862**

[22] Filed: **Aug. 4, 1997**

[51] **Int. Cl.**^{6} **F02M 57/06**

[52] **U.S. Cl.** **123/297**; 123/DIG. 12; 123/527

[58] **Field of Search** 123/297, 527, 123/DIG. 12; 313/120

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,926,169	12/1975	Leshner et al.	123/297
3,980,061	9/1976	Mcalister	123/297
4,319,552	3/1982	Sauer et al.	123/297
4,343,272	8/1982	Buck	123/297
4,383,198	5/1983	Hosking	313/120
4,448,160	5/1984	Vosper	123/297
4,546,740	10/1985	Clements et al.	123/297
5,531,199	7/1996	Bryant et al.	123/297

Primary Examiner—Erick R. Solis
Attorney, Agent, or Firm—I. C. Waddey, Jr.; Waddy & Patterson

[57] **ABSTRACT**

The internal combustion engine fuel system described includes a structure for mixing the alternative fuel, preferably hydrogen, with oxygen in ambient air to stratify the fuel. The system includes an adapter, and the adapter includes a housing mounted between spark plug and cylinder of the internal combustion engine. A plug is placed within the housing. The plug has ridges or grooves on its outer surface that act as mixing structures. Thus, when hydrogen is introduced into the adapter housing it is mixed with ambient oxygen within the chamber as it flows over the plug. The mixing structures in the housing create a vortexing action as the hydrogen flows over the plug and towards the cylinder of the engine. An electrode protrudes from the plug towards the cylinder. The electrode is preferably platinum and generates the necessary spark to create combustion of the hydrogen/air mixture adjacent to the cylinder to thereby power the cylinder in the engine. A platinum electrode is preferably used because it enhances a catalytic conversion of combustion by-products to more environmentally compatible products.

38 Claims, 6 Drawing Sheets

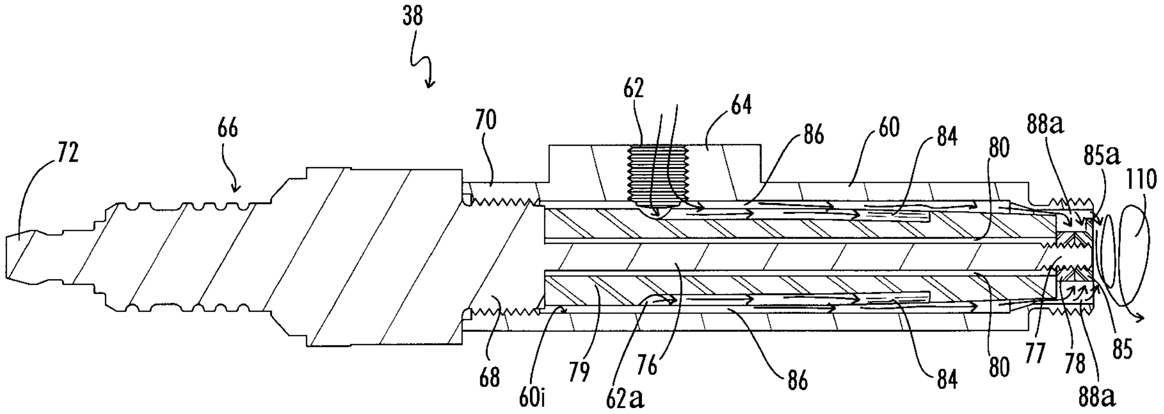

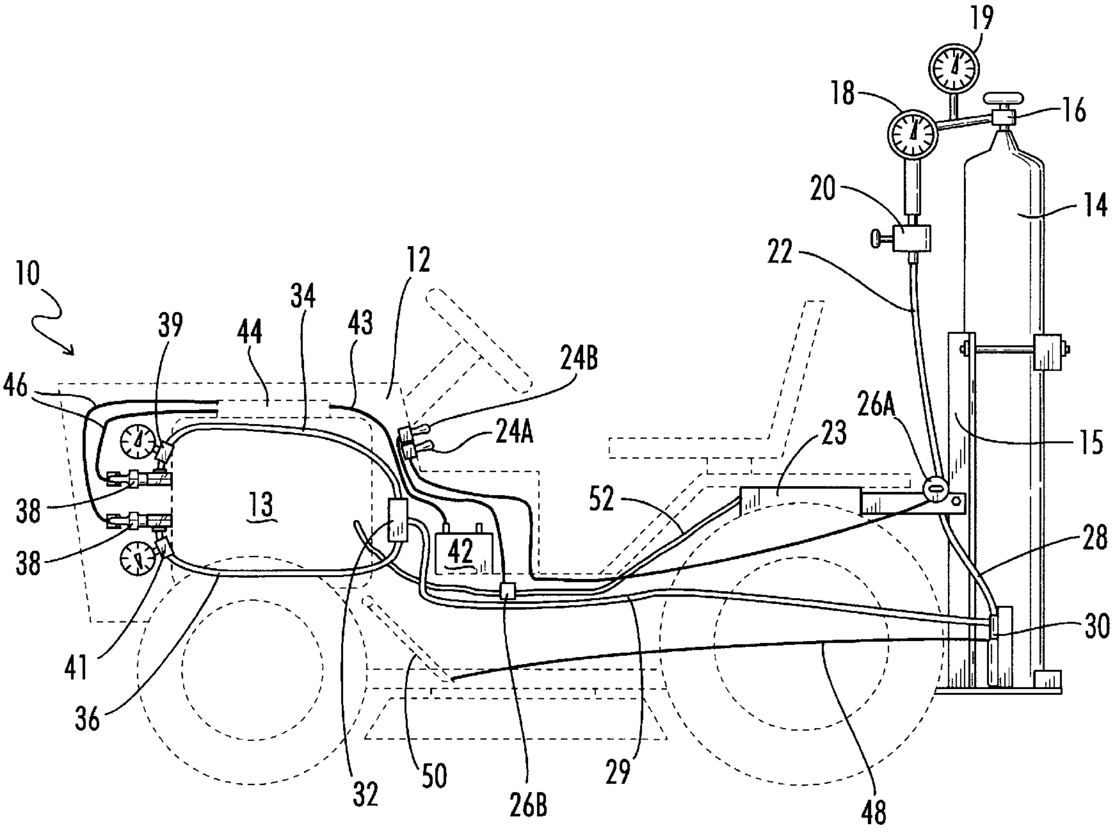

FIG. 1

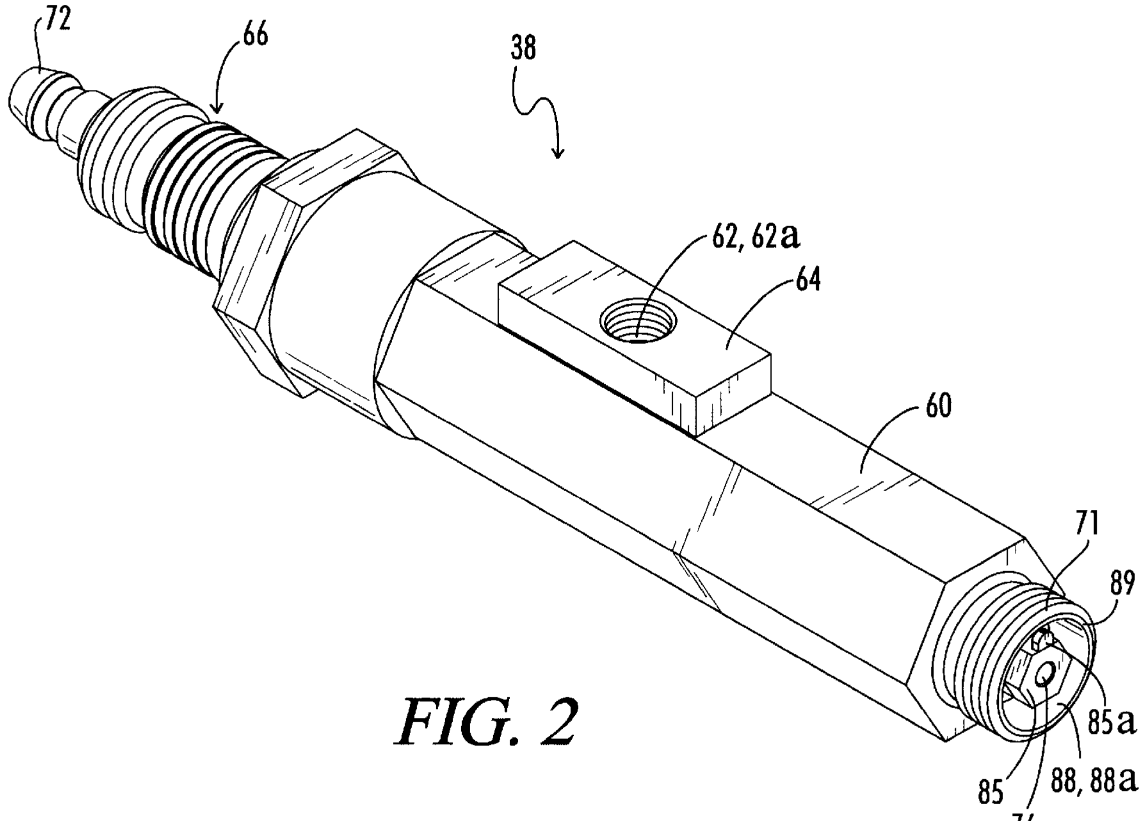
72
66
38
62, 62a
64
60
71
89
85a
88, 88a
85
76

FIG. 2

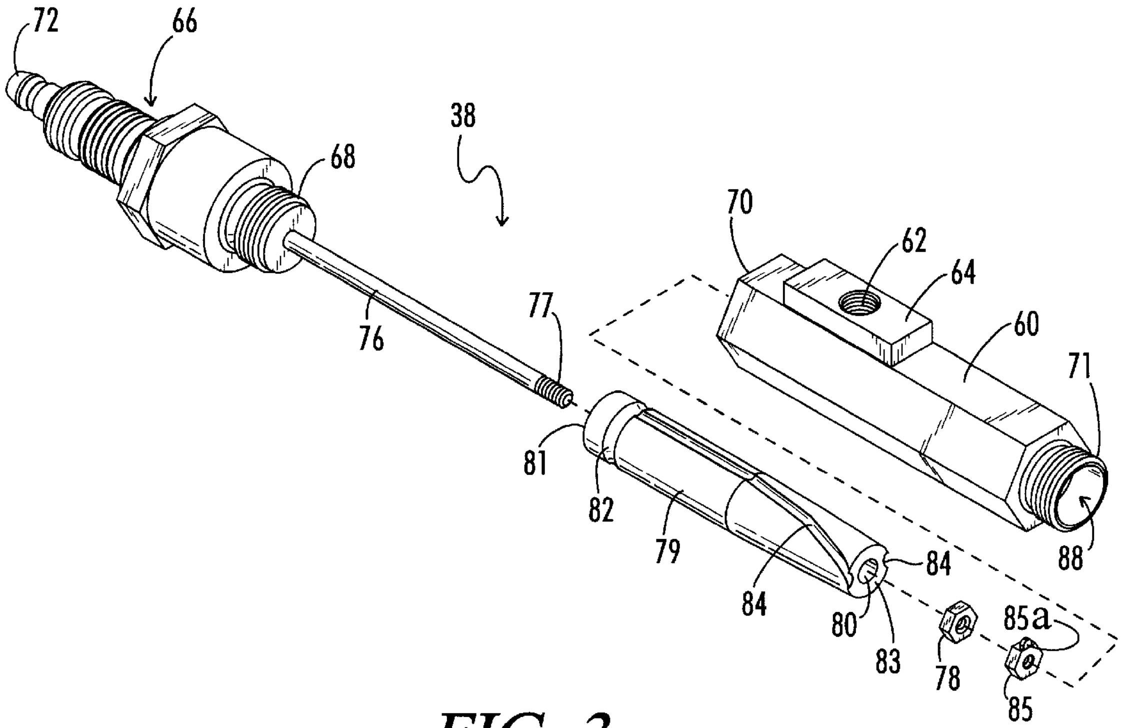
72
66
38
68
70
62
64
60
77
76
71
81
82
79
84
84
88
80
83
78
85a
85

FIG. 3

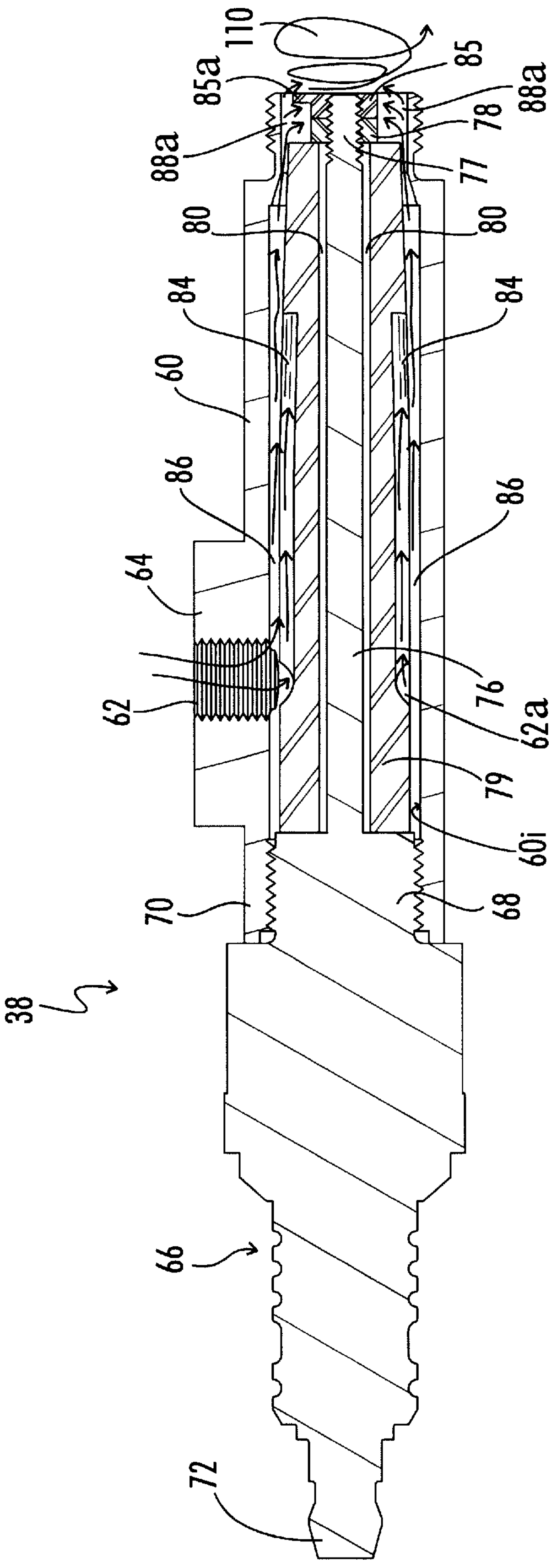

FIG. 4

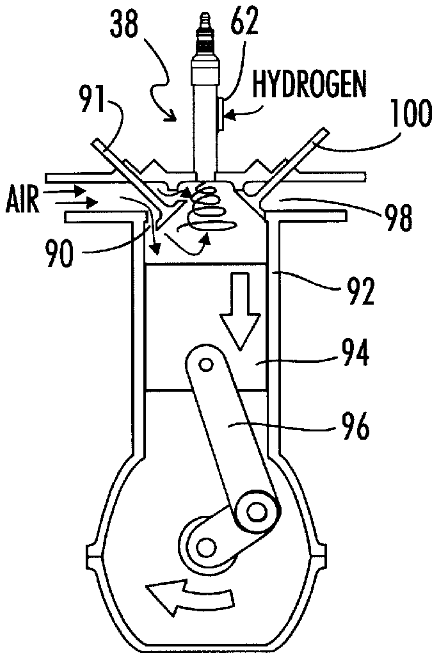
38
62
HYDROGEN
91
100
AIR
98
90
92
94
96

INTAKE

FIG. 5A

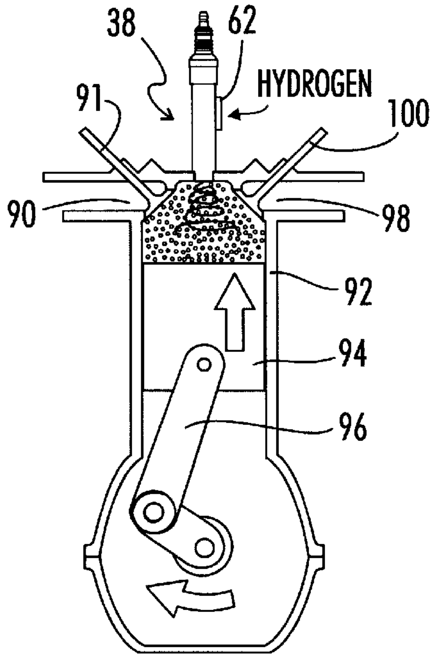
38
62
HYDROGEN
91
100
90
98
92
94
96

COMPRESSION

FIG. 5B

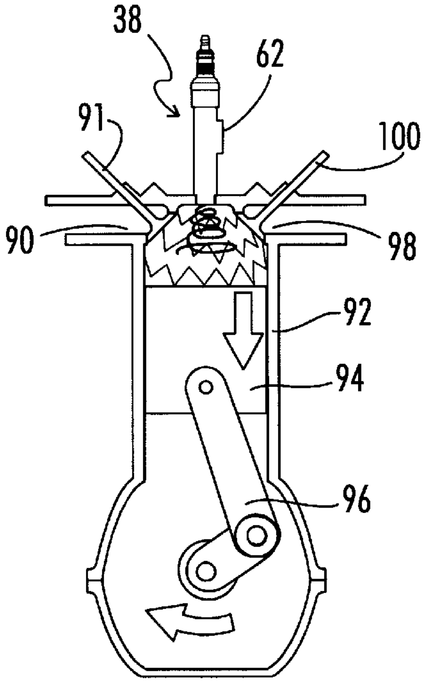
38
62
91
100
90
98
92
94
96

POWER

FIG. 5C

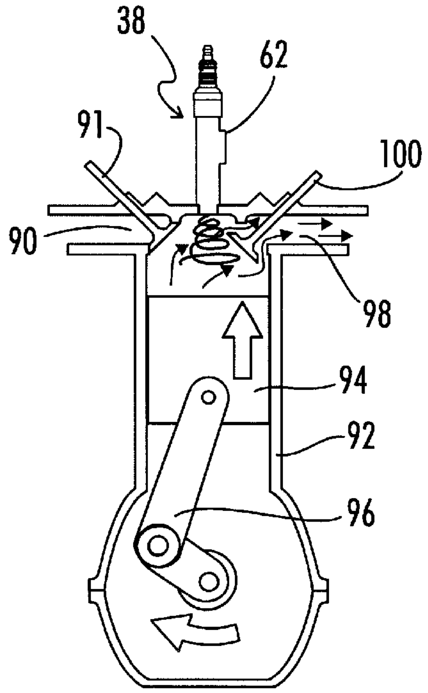
38
62
91
100
90
98
94
92
96

EXHAUST

FIG. 5D

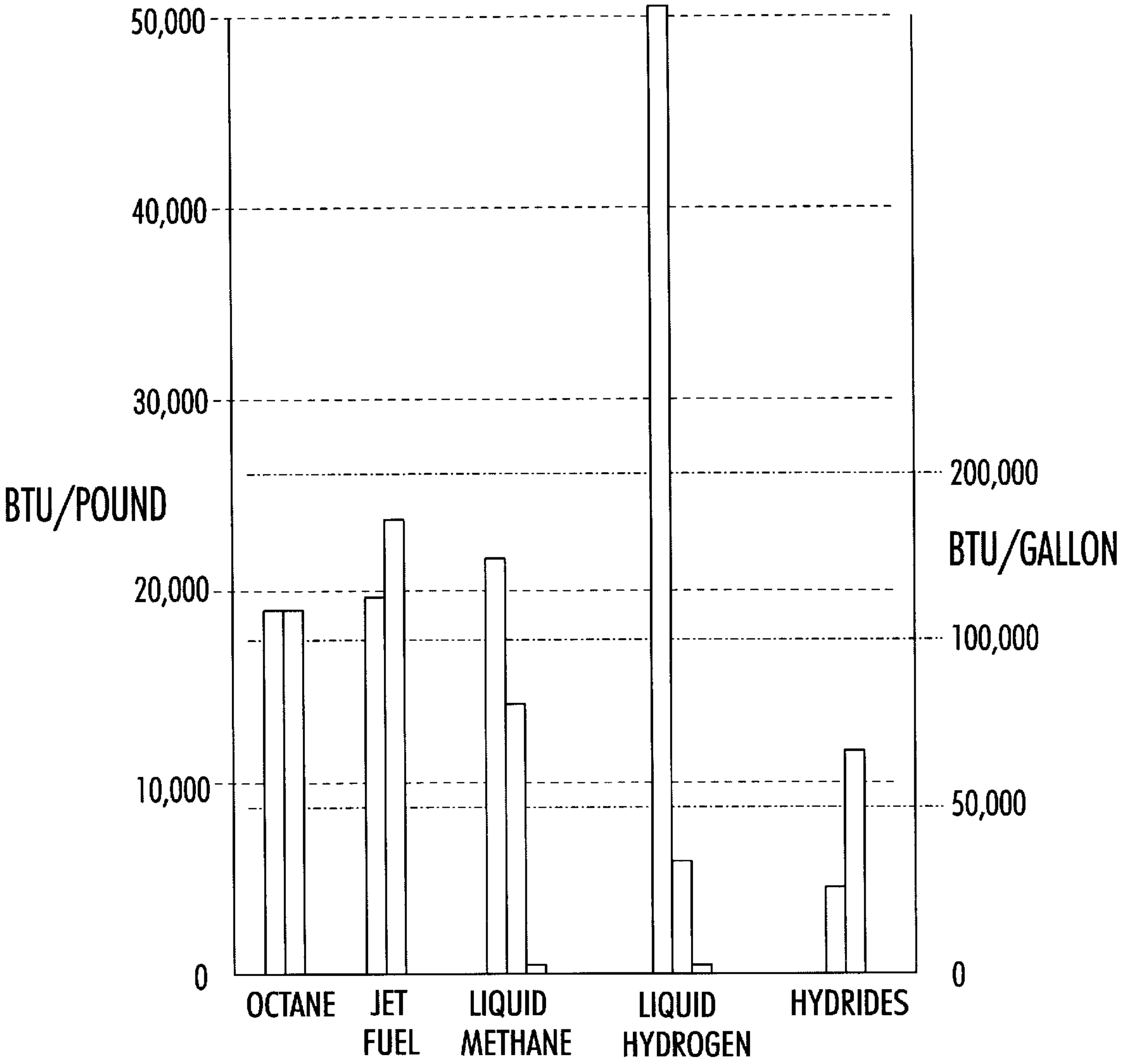

METHANE GAS - 140 BTU/GAL
HYDROGEN GAS - 40 BTU/GAL

FIG. 6

FUEL SYSTEM FOR INTERNAL COMBUSTION SYSTEM AND ADAPTER FOR USE IN SAME

BACKGROUND OF THE INVENTION

The present invention relates generally to a fuel system for internal combustion engines that enables the invention to utilize alternative fuels; and, more particularly to an improved fuel system for an internal combustion engine and an adapter for use in same.

The problems of air pollution caused by automobiles, lawn mowers and other vehicles that have internal combustion engines powered by gasoline is well documented. This problem is particularly described in an article in the San Diego Union Tribune, Wednesday, Sep. 25, 1996 and in Popular Science, August, 1996. Indeed, under the Clean Air Act, in the next few years cars that do not run on standard fossil fuels like gasoline will be required in several states. Currently, there are no vehicles available to meet this need.

Thus, it will be appreciated by those skilled in the art that there has been a push to seek other fuels for use in internal combustion engines. These fuels include methane, propane and gasohol.

Hydrogen has also been characterized as a promising fuel source. See Department of Energy publication entitled "Hydrogen Fuel", printed in 1978, U.S. Government Printing Office, U.S. Department of Energy, Office of Public Affairs, Washington, D.C. 20585. Particularly, the graph presented in FIG. **6**, which is copied from this publication, shows the potential for hydrogen in BTUs/Lb. However, the mass energy density for hydrogen is low; and thus, hydrogen needs to be compressed before its energy potential can be realized. No prior art fuel system has been able to capture the energy of hydrogen.

Several fuel systems for use with hydrogen have been proposed. Examples are described in U.S. Pat. Nos. 4,167,919; 4,253,428; 4,016,836; 4,178,882; 5,222,993; 5,085,176; and 5,085,176. The problem with these prior art fuel systems is they cannot can take advantage of the alternative fuels to generate sufficient power from the fuels to drive an internal combustion engine for use in a lawn mower, automobile or other vehicle. As noted above, this is a particular problem when hydrogen is used as a fuel. Additionally, the prior art systems do not address the problem of back fire, i.e. undesired explosion of the fuel in the cylinder during the exhaust stroke of the internal combustion engine.

What is needed, then, is a fuel system for an internal combustion engine that provides a way to generate sufficient power from the alternative fuels and that reduces backfire problems. Such a system is lacking in the prior art.

SUMMARY OF THE INVENTION

It is an object of this invention to provide an improved fuel system for use with alternative fuels.

It is a further object of this invention to provide an improved fuel system for use particularly with hydrogen, and particularly to enhance the mass energy density of hydrogen.

It is still a further object of this invention to provide an adapter for use in a fuel system for an internal combustion engine that enhances the power obtained from the alternative fuels.

It is still a further object of this invention to provide an improved fuel system for an internal combustion engine that does not present the environmental hazards of a standard gasoline engine by providing reduced or zero emission of environmentally harmful gases.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that can use an alternative fuel and can use gasoline to facilitate the transition from gasoline to alternative fuels.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that reduces backfire problems.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that reduces backfire problems by cooling the electrode in the spark plug with the hydrogen gas fuel.

It is still a further object of this invention to provide an adapter for use in a fuel system for an internal combustion engine that enhances the power obtained from the alternative fuels.

It is yet a further object of this invention to provide a direct injection fuel system wherein fuel is injected directly into the cylinder.

It is yet a further object of this invention to provide a way to convert the standard induction fuel system of a lawnmower to a direct injection fuel system wherein fuel is injected directly into the cylinder.

It is yet a further object of this invention to provide a way to convert the indirect injection fuel system of an automobile to a direct injection fuel system wherein fuel is injected directly into the cylinder.

The internal combustion engine fuel system of this invention includes a structure for mixing the alternative fuel, preferably hydrogen, with oxygen in ambient air to stratify the fuel. The term "alternative fuels" is known in the art and is meant to refer to fuels other than conventional gasoline.

The invention comprises an adapter, the adapter including a housing mounted between spark plug and cylinder of the internal combustion engine. A plug is placed within the housing. The plug has ridges or grooves on its outer surface that act as mixing structures. Thus, when hydrogen is introduced into the adapter housing it is mixed with ambient oxygen within the chamber as it flows over the plug. The mixing structures on the plug create a vortexing action as the hydrogen flows over the plug and towards the cylinder of the engine.

An electrode protrudes from the plug towards the cylinder. The electrode is preferably platinum and generates the necessary spark to create combustion of the hydrogen/air mixture adjacent to the cylinder to thereby power the cylinder in the engine. A platinum electrode is preferably used because it enhances a catalytic conversion of combustion by-products to more environmentally compatible products.

Thus, the combination of the compression action generated by the conventional piston and push rod within the cylinder of the engine in conjunction with the hydrogen/air mixing structure that vortexes the hydrogen/air create compressed hydrogen that is mixed with a combustion facilitator, i.e. oxygen in the ambient air to enable the system of this invention to capture the abundant energy found in hydrogen fuel.

The enhanced mixing capabilities of this invention is also adaptable for use with other alternative fuels such as natural gas, alcohol, propane and the like.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a plan view of the fuel system of this invention with a lawn mower shown in phantom.

FIG. **2** is a perspective view of the adapter of the fuel system of this invention.

FIG. **3** is an exploded perspective view of the adapter of the fuel system of this invention.

FIG. **4** is a cross-sectional view of the adapter of the fuel system of this invention.

FIGS. **5A**–**5D** are a schematic of the 4 cycle combustion in an internal combustion engine including the adapter of this invention.

FIG. **6** is a graph which shows the energy potential for hydrogen in BTUs/Lb.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring now to the drawings, wherein like reference numerals refer to like parts throughout, the fuel system is referred to generally at **10**. The fuel system of this invention is contemplated for use with an internal combustion power system, such as an internal combustion engine or such as a home heating system. Fuel system **10** is portrayed in this case as connected to a standard lawn mower. The lawn mower is shown in phantom at **12** in FIG. **1**. The lawn mower is a conventional lawn mower as can be purchased from a company such as Murray of Ohio. The internal combustion engine in the lawn mower is a standard 4-cycle engine with two cylinders and develops 18 horsepower. A suitable example can be purchased from Briggs & Stratton, Milwaukee, Wis., 53201.

Referring particularly to FIG. **1**, system **10** includes fuel tank **14** that is mounted to lawn mower **12** via mounting assembly **15**. In the preferred embodiment, fuel tank **14** includes hydrogen gas. Accordingly, for the purposes of the foregoing detailed description, hydrogen gas will be described.

A line **22** begins at fuel tank **14** and proceeds to back-flow valve **16**. Back-flow valve **16** is a standard valve and is available from AirCo Gases Company, 575 Mountain Avenue, Murray Hill, N.J. 07974, part no. WCS CV-4M. PSI tank gauge **19** is inserted in line **22** between back-flow valve **16** and flow pressure gauge **18**. PSI tank gauge **19** is available from AirCo Gases Company also, part no. WCS CV-5M, and serves as an indicator of fuel remaining in the tank **14**. Flow pressure gauge **18** is used to monitor the flow of the hydrogen in pressure units. The flow is started using valve **16**. The optimal flow pressure is 25 pounds of flow pressure from the tank **14** to the cylinder of the engine.

The flow of hydrogen is regulated via valve **20** that is part of gauge **18**. Thus, when valve **20** is opened, hydrogen proceeds through line section **22** to hydrogen flow solenoid **26**A. The user can access the hydrogen fuel by flipping hydrogen flow switch **24**A into the "on" mode so that hydrogen gas in line **22** can proceed through hydrogen flow solenoid **26**A and into line section **28**.

Throttle **30** regulates the flow of hydrogen from line section **28** into line section **29** and ultimately, to engine **13**. Throttle **30** is connected to throttle pedal **50** via includes throttle cable **48**. Throttle pedal **50** also operates the butterfly valves in the carburetor in engine **13** in a conventional manner.

As an option, the throttle can be controlled electronically by a digital fuel controller product. Such a product is commercially available from Autotronic Controls Corp., 1490 Henry Brennan Drive, El Paso, Tex. 79936.

Line portion **29** proceeds through splitter **32** and is split into upper line **34** and lower line **36**. Lines **34** and **36** proceed to adapters **38**. Additional back-flow prevention valves **39** and **41** are mounted between lines **34** and **36** and adapters **38**. Braided Teflon® is the preferred material for the fuel lines of the system.

Battery **42** is connected via battery cable **43** to magneto **44**. Magneto **44** then controls the current through spark plug lines **46** to adapters **38** in a conventional manner. Thus, the regulation of the spark that ignites the fuel in the cylinder of engine **13** is conventional. Additionally, the intake of air into the engine **13** is accomplished through the standard carburetor system of engine **13**. In a car, the sparking of the spark plugs can be controlled by the standard distributor/battery ignition system.

A second solenoid **26**B controls the flow of gasoline from tank **23** through gasoline line **52**. Solenoid **26**B is controlled by switch **24**B. Throttle **30** is connected to gasoline line **52** and can also be used to control the flow of gasoline to engine **13**. Thus, this invention can be used with both an alternative fuel and gasoline.

Referring now to FIGS. **2**, **3** and **4**, the adapter **38** of this invention is shown in more detail. Adapter **38** includes housing **60**. A threaded port **62** is formed into housing **60** through port block **64**. Hydrogen lines are threadably connected to the adapter **38** via port **62**. Port **62** is also referred to as an entrance port.

Adapter **38** also includes spark plug **66** which is effectively a standard spark plug except for the modification to the electrode as described below. Spark plug **66** includes a threaded end **68** which is threadably received in the threaded upper end **70** of housing **60**. Spark plug wires **46** (as seen in FIG. **1**) are connected to connector end **72** of spark plug base **66**.

Electrode **76** protrudes from spark plug **66** and is of a length sufficient so that the threaded tip **77** (best seen in FIGS. **3** and **4**) of the electrode **76** is proximate to the cylinder of the internal combustion engine when the adapter **38** is mounted. In one embodiment the electrode **76** is co-terminus with an exit port **88** of the adapter **38**. This is most clearly shown in FIG. **4**. Electrode **76** is preferably platinum to enhance catalysis of combustion by-products, as will be more fully discussed below.

Adapter **38** also includes plug **79**. Plug **79** includes an axial internal void space **80**. Thus, plug **79** fits over electrode **76** via internal void space **80**. Plug **79** preferably comprises a ceramic material that acts as a good insulator. A suitable source for ceramic plug **79** is AM/PRO Machinery, Inc. 134 Church Street, Plain City, Ohio 43064-1321.

An important aspect of this invention is found on the external surface of plug **79**. An annular groove **82** is formed on the surface of plug **79** near its distal end **81** so as to be proximal to the port **62** where hydrogen enters housing **60**. Two stratifying grooves **84** proceed along the surface of plug **79** until they reach its distal end **83**. It is these mixing structures, i.e. annual groove **82** and stratifying grooves **84**, that facilitate the improved power generation from alternative fuels as provided by this invention. It will be apparent to those skilled in the analogous arts that the mixing structure may be formed on an inner surface **60***i* of the adapter housing **60**. Alternatively, a combination of vortex producing structures may be attached to, and preferably integral with, both the plug **79** and the adapter housing inner surface **60***i*. Also, any conventional vortex producing means would be sufficient, i.e. grooves, per se, are not required. An entrance area **62***a* proximate the entrance port **62** is typically defined between the housing inner surface **60***i* and the external surface of plug **79**. The external surface of plug **79** typically includes the grooves **84**. So, the entrance area **62***a* is the volume of space within the housing, proximate the entrance port **62**, which the gas will be able to occupy. This is most clearly shown in FIG. **4**.

On threaded tip **77** of electrode **76**, there are mounted a spacer nut **78** and a spark gap nut **85**. As best seen in FIG. **2**, the spark gap nut **85** includes a protrusion **85***a* that provides the spark gap between electrode **76** and periphery **89** of the exit port **88** of housing **60**. The exit port **88** includes an exit area **88***a*. This is shown most clearly in FIGS. **2** and **4**. The exit area **88***a* is no larger than the entrance area **62***a*. If the exit area **88***a* is smaller than the entrance area **62***a*, the pressure on the gas will increase. This increases the rate at which the gas is expelled from the plug **79**, typically out through grooves **84**. The electrode **76** is substantially co-terminus with the plug distal end **83** (seen most clearly in FIG. **2**). Thus, the electrode **76** is substantially co-terminus with the exit port **88** or exit area **88***a*. Note that tip of the electrode **77** is preferably proximate with the exit area **88***a* because this is the location of greatest mixing. This is most clearly seen in FIG. **4**. Preferably, the spark gap nut **85** is brass and the preferred spark gap distance is 25/1000 of an inch.

As best seen in FIG. **4**, where arrows represent gas flow, hydrogen gas flows into chamber **60** via port **62**. The hydrogen gas cools electrode **76**, which facilitates combustion. Stratifying grooves **84**, preferably there are two, act to mix the hydrogen that is flowing into housing **60** with ambient air therein. Though, more or less vortex creating grooves (a.k.a. stratifying grooves) may be desired. Typically, this will depend on the relative dimensions of the grooves and diameter of the plug as well as the pressure and velocity of the mixture in the housing. The mixing occurs within void space **86** of housing **60** and creates a vortexing or tornado action. Plug **79** tapers as it proceeds from its distal end **81** to its proximal end **83**. Thus, proximal end **83** is smaller in diameter than distal end **81**. This provides additional space within void space **86** for the mixing of the hydrogen and air. The stratified hydrogen/air are twisting in a vortex fashion when leaving housing **60** via exit port **88**. The vortexing action produced by grooves **84** on the plug **79** creates a "tornado" **110** of hydrogen/air with a hollow center that generates a vacuum. Similar effect, or vortex flow path, would result were the grooves to be placed on the housing. The vacuum propels the hydrogen/air mixture into the cylinder of the engine for combustion. Such mixing cannot be accomplished in a conventional induction system.

Referring now to FIG. **5**, a schematic of the system of this invention is portrayed, with arrows representing gas flow. Hydrogen flows in through port **62** through adapter **38** and into cylinder **92**. Air enters cylinder **92** through air intake **90** and air intake valve **91**. The stratified hydrogen/air mixture is compressed in the compression stroke of the engine via piston **94** and push rod **96**, which are standard structures in an internal combustion engine. In the power stroke shown in FIG. **5**-C, a spark is generated in the compressed stratified hydrogen/air mixture ignites and drives piston **94** and push rod **96** to power the engine.

The mass energy density of hydrogen is enhanced through the increased pressure on the hydrogen exerted by piston **94** on the hydrogen/air mixture within the cylinder **92** in the uptake stroke in an internal combustion engine. The pressurized and well-mixed hydrogen/air mixture is rapidly ignited through the spark because hydrogen has a fast flame front. This reduces problems with timing. Thus, the energy store found in hydrogen is captured through this system via the compression of the hydrogen gas within cylinder **92** and via the thorough mixing of hydrogen and air provided by adapter **38**.

In addition, in the exhaust phase in a four cycle engine a second spark is generated from the plug to ensure complete combustion of by-products. This is enhanced by using an electrode made of platinum. Exhaust leaves cylinder **92** through exhaust port **98**. Exhaust port **98** is closed during other phases by exhaust outlet valve **100**.

In the exhaust stroke, because air valve **91** is closed, the substantial component of the gases in cylinder **92** is hydrogen. This is also accomplished due to the fact that in the system of this invention, the hydrogen is continuously flowing. When this spark interacts with what is predominately hydrogen, a plasma is formed. This plasma blocks back-flow of air from exhaust port **98**. Prevention of back-flow of air prevents the hydrogen fuel system from backfiring, which is another common problem of hydrogen fuel systems. In effect, then, the plasma formed by the hydrogen acts as a virtual valve in that the plasma blocks the back-flow of air.

When the plasma of the hydrogen is generated as described above, UV radiation is generated which also facilitates consumption of combustion by-products. This increases the environmental friendliness of the system.

In addition to hydrogen the following fuels can be used in this system: natural gas; alcohol; gasohol; propane and buthane. A mixture of hydrogen and natural gas can also be used. As noted above, gasoline can be used in this system when solenoid **26** is used to stop the flow of hydrogen.

Further, it is contemplated that an adapter that is described herein for use with an internal combustion engine can also be used with a home heating system such as a home heating system that may currently use natural gas. The vortexing action created by the plug would enable appropriate power to be captured from the hydrogen fuel so as to effectively heat a home.

Thus, the apparatus and system of this invention provides an effective means to use alternative fuels. The use of alternative fuels can enhance the environment by reducing the use of conventional gasoline and the environmentally detrimental by-products caused by the consumption thereof.

Thus, although there have been described particular embodiments of the present invention of a new and useful "Fuel System For Internal Combustion System And Adapter For Use In Same", it is not intended that such references be construed as limitations upon the scope of this invention except as set forth in the following claims. Further, although there have been described certain dimensions used in the preferred embodiment, it is not intended that such dimensions be construed as limitations upon the scope of this invention except as set forth in the following claims.

What I claim is:

1. An adapter for use in an internal combustion power system, the adapter comprising a housing having an internal void space; a plug mounted within the internal void space of the housing, said housing and said plug combining to define an entrance area though which fuel can be introduced, and said housing and said plug combining to define an exit area through which fuel introduced to said entrance area exits, said exit area being no larger than the entrance area; a vortex means for causing fuel exiting through said exit area to have a vortex flow path; and an entrance port into the housing for accepting a fuel supply.

2. The adapter according to claim **1** further comprising an electrode mounted within said plug.

3. The adapter according to claim **1** wherein an end of the plug is smaller in diameter that an other end of the plug.

4. The adapter according to claim **1** wherein the vortex means comprises at least one groove formed along a surface of the plug.

5. The adapter according to claim **4** wherein the vortex means further comprises an annular groove formed on the plug adjacent to said entrance port.

6. The adapter according to claim **1** wherein the mixing structure comprises a ridge.

7. The adapter according to claim **2** wherein the electrode comprises platinum.

8. The adapter of claim **2**, wherein the electrode is co-terminus with the exit area.

9. The adapter of claim **1**, wherein the housing further comprises an inner surface, the housing inner surface including the vortex means.

10. The adapter of claim **9**, wherein the plug comprises complementary vortex means compatible with the vortex means included in the housing inner surface.

11. A fuel system for an internal combustion power system comprising a fuel source, the fuel source in fluid communication with the internal combustion power system; and an adapter connected between the fuel source and the internal combustion power system, the adapter comprising a housing having an internal void space; a plug mounted within the internal void space of the housing, said housing and said plug combining to define an entrance area though which fuel can be introduced, and said housing and said plug combining to define an exit area through which fuel introduced to said entrance area exits, said exit area being no larger than the entrance area; a mixing structure within the housing for causing fuel exiting through said exit area to have a vortex flow path; and an entrance port into the housing for accepting a fuel supply from the fuel source.

12. The fuel system according to claim **11** further comprising an electrode mounted within said plug.

13. The fuel system according to claim **11** wherein an end of the plug is smaller in diameter that an other end of the plug.

14. The fuel system according to claim **11** wherein the mixing structure comprises at least one groove formed along a surface of the plug.

15. The fuel system according to claim **14** wherein the mixing structure further comprises an annular groove formed on the plug adjacent to said port.

16. The fuel system according to claim **11** wherein the mixing structure comprises a ridge.

17. The fuel system according to claim **12** wherein the electrode comprises platinum.

18. The fuel system of claim **12**, wherein the electrode is co-terminus with the exit area.

19. The fuel system of claim **11**, wherein the housing further comprises an inner surface, the housing inner surface including the mixing structure.

20. The adapter of claim **19**, wherein the plug comprises complementary mixing structure compatible with the mixing structure included in the housing inner surface.

21. An adapter for use in an internal combustion engine, the adapter comprising a housing having an internal void space; a plug mounted within the internal void space of the housing, said housing and said plug combining to define an entrance area though which fuel can be introduced, and said housing and said plug combining to define an exit area through which fuel introduced to said entrance area exit; a mixing structure in the housing for causing fuel exiting through said exit area to have a vortex flow path; an electrode axially mounted with the plug and co-terminus with the exit area of the housing; and an entrance port into the housing for accepting a fuel supply.

22. The adapter according to claim **21** wherein an end of the plug is smaller in diameter that an other end of the plug.

23. The adapter according to claim **21** wherein the mixing structure comprises at least one groove formed along a surface of the plug.

24. The adapter according to claim **23** wherein the mixing structure further comprises an annular groove formed on the plug adjacent to said port.

25. The adapter according to claim **21** wherein the mixing structure comprises a ridge.

26. The adapter according to claim **21** wherein the electrode comprises platinum.

27. The adapter according to claim **22** wherein the end of the plug having a smaller diameter is adjacent to a cylinder of the internal combustion engine when the adapter is mounted to an internal combustion engine.

28. The fuel system of claim **21**, the housing further comprises an entrance area no smaller than the exit area.

29. The fuel system of claim **21**, wherein the housing further comprises an inner surface, the housing inner surface including the mixing structure.

30. A fuel system for an internal combustion engine comprising a fuel source, the fuel source in fluid communication with the internal combustion power system; and an adapter connected between the fuel source and the internal combustion engine, the adapter comprising the adapter comprising a housing having an internal void space; a plug mounted within the internal void space of the housing said housing and said plug combining to define an entrance area though which fuel can be introduced, and said housing and said plug combining to define an exit area through which fuel introduced to said entrance area exit; a mixing structure in the housing for causing fuel exiting through said exit area to have a vortex flow path; an electrode axially mounted with the plug and co-terminus with the exit area of the housing; and an entrance port into the housing for accepting the fuel supply.

31. The adapter according to claim **30** wherein an end of the plug is smaller in diameter that an other end of the plug.

32. The adapter according to claim **30** wherein the mixing structure comprises at least one groove formed along a surface of the plug.

33. The adapter according to claim **32** wherein the mixing structure further comprises an annular groove formed on the plug adjacent to said port.

34. The adapter according to claim **30** wherein the mixing structure comprises a ridge.

35. The adapter according to claim **30** wherein the electrode comprises platinum.

36. The adapter according to claim **31** wherein the end of the plug having a smaller diameter is adjacent to a cylinder of the internal combustion engine when the adapter is mounted to an internal combustion engine.

37. The adapter of claim **30**, wherein the housing further comprises an entrance area no smaller than the exit area.

38. The adapter of claim **30**, wherein the housing further comprises an inner surface, the housing inner surface including the mixing structure.

* * * * *

US006119651A

United States Patent [19]

Anderson

[11] **Patent Number: 6,119,651**

[45] **Date of Patent: Sep. 19, 2000**

[54] **HYDROGEN POWERED VEHICLE, INTERNAL COMBUSTION ENGINE, AND SPARK PLUG FOR USE IN SAME**

[75] Inventor: **Herman P. Anderson**, Brentwood, Tenn.

[73] Assignee: **Herman P. Anderson Technologies, LLP**, Brentwood, Tenn.

[21] Appl. No.: **09/220,286**

[22] Filed: **Dec. 23, 1998**

Related U.S. Application Data

[63] Continuation-in-part of application No. 08/905,862, Aug. 4, 1997, Pat. No. 5,852,993.

[51] **Int. Cl.**^{7} **F02M 57/06**

[52] **U.S. Cl.** **123/297**; 123/635; 123/DIG. 12

[58] **Field of Search** .... 123/3, DIG. 12, 123/527, 297, 636, 637, 635, 169 EL; 313/120

[56] **References Cited**

U.S. PATENT DOCUMENTS

2,208,030	7/1940	Holmes	123/169 EL
3,165,099	1/1965	Vanderpoel	123/637
3,926,169	12/1975	Leshner et al.	123/297
3,980,061	9/1976	McAlister	123/297
4,164,912	8/1979	Beyler	123/26
4,186,712	2/1980	Fitzner et al.	123/633
4,319,552	3/1982	Sauer et al.	123/297
4,343,272	8/1982	Buck	123/297
4,383,198	5/1983	Hosking	313/120
4,448,160	5/1984	Vosper	123/297
4,820,957	4/1989	Zivkovich	315/209 T
5,852,993	12/1998	Anderson	123/297

FOREIGN PATENT DOCUMENTS

686943	8/1930	France	123/297

Primary Examiner—Noah P. Kamen
Attorney, Agent, or Firm—Lucian Wayne Beavers Waddey & Patterson

[57] **ABSTRACT**

The internal combustion engine fuel system described includes a structure for mixing the alternative fuel, preferably hydrogen, with oxygen in ambient air to stratify the fuel. The system includes an adapter, and the adapter includes a housing mounted between spark plug and cylinder of the internal combustion engine. A plug is placed within the housing. The plug has ridges or grooves on its outer surface that act as mixing structures. Thus, when hydrogen is introduced into the adapter housing it is mixed with ambient oxygen within the chamber as it flows over the plug. The mixing structures in the housing creates a vortexing action as the hydrogen flows over the plug and towards the cylinder of the engine. An electrode protrudes from the plug towards the cylinder. The electrode is preferably platinum and generates the necessary spark to create combustion of the hydrogen/air mixture adjacent to the cylinder to thereby power the cylinder in the engine. A platinum electrode is preferably used because it enhances a catalytic conversion of combustion by-products to more environmentally compatible products. The present invention also teaches a spark plug producing a hotter spark for a hydrogen fuel system. Also taught is a hydrogen powered vehicle with reduced emissions by producing a spark during the power stroke and the exhaust stroke. Methods for reducing exhaust pollution are also taught. Methods of reducing exhaust pollution by generating a plasma are taught as well.

13 Claims, 12 Drawing Sheets

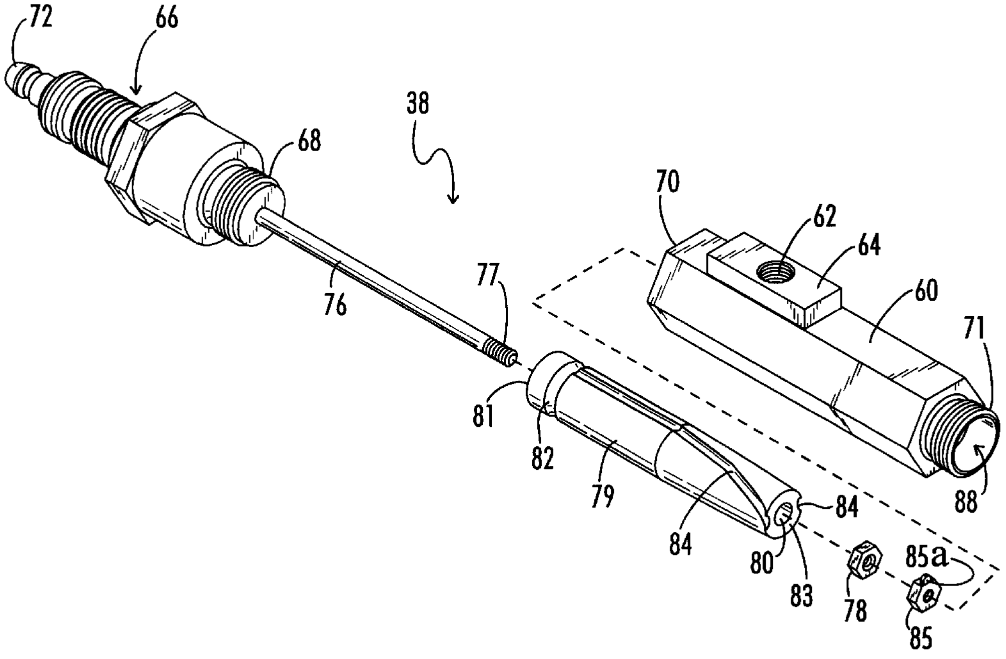

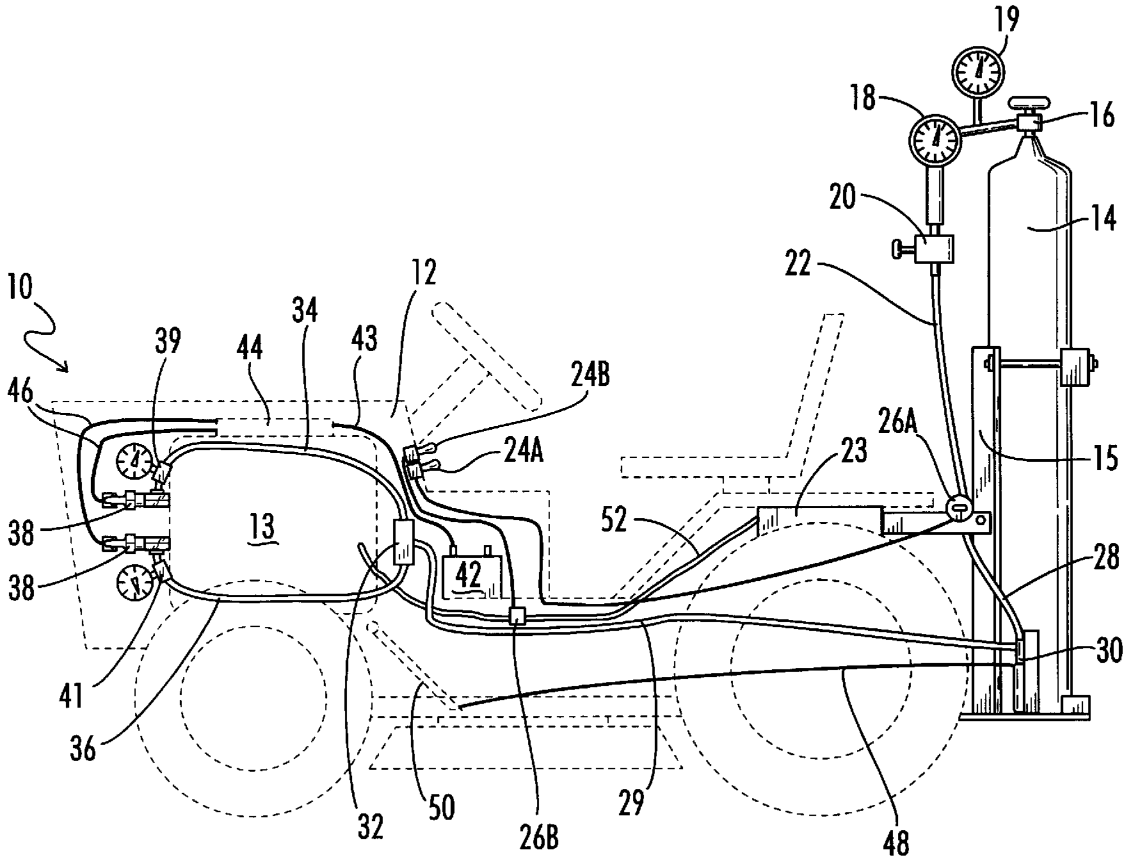

FIG. 1

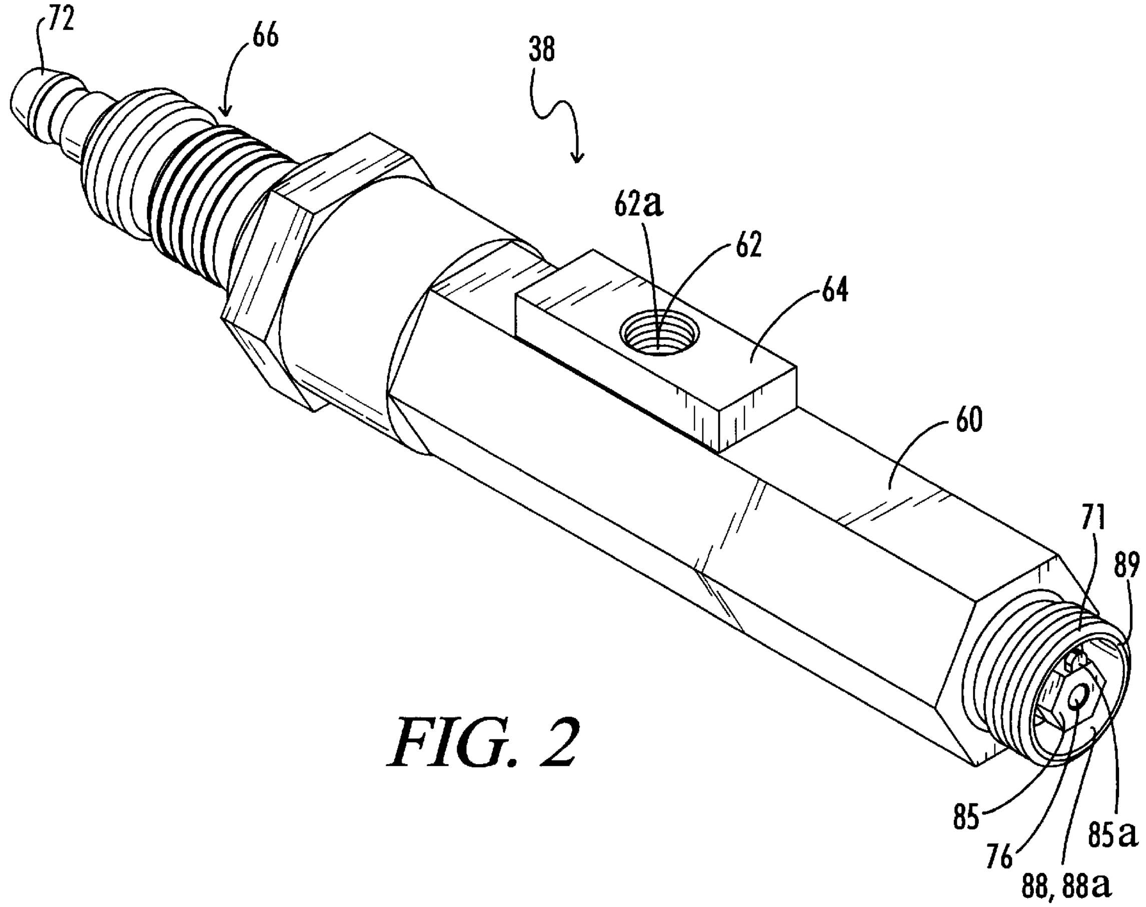
72
66
38
62a
62
64
60
71
89
85
85a
76
88, 88a

FIG. 2

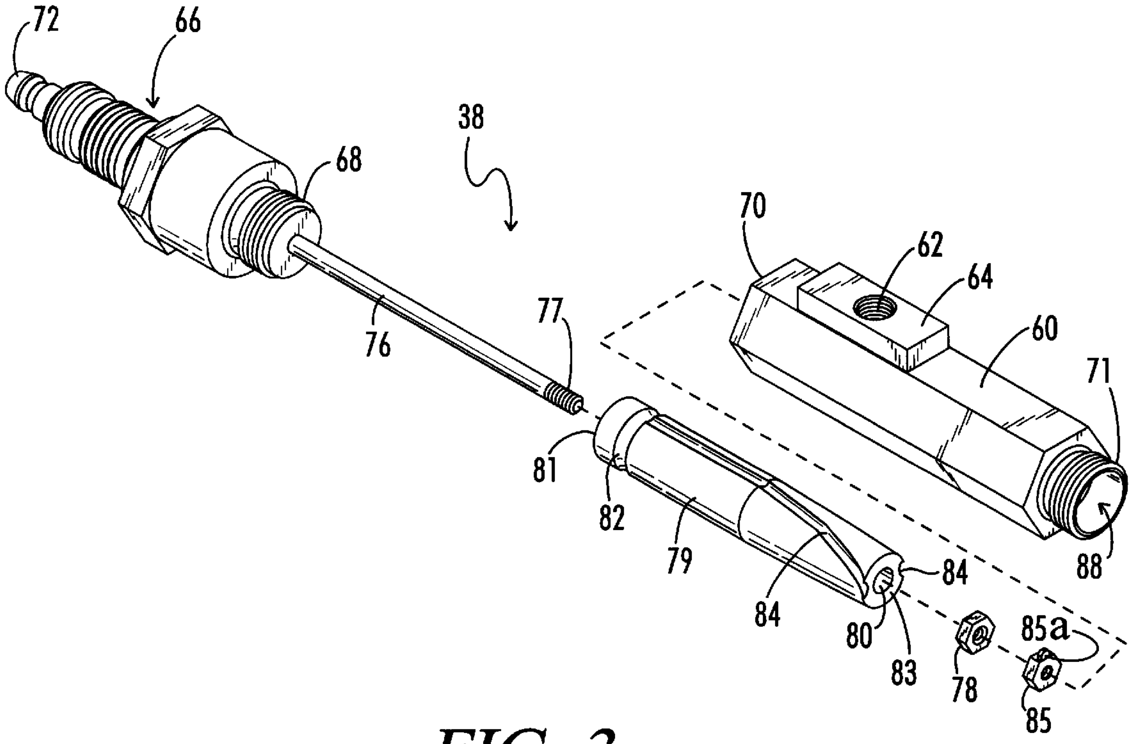
72
66
68
38
76
77
70
62
64
60
71
81
82
79
84
84
80
83
78
85a
85
88

FIG. 3

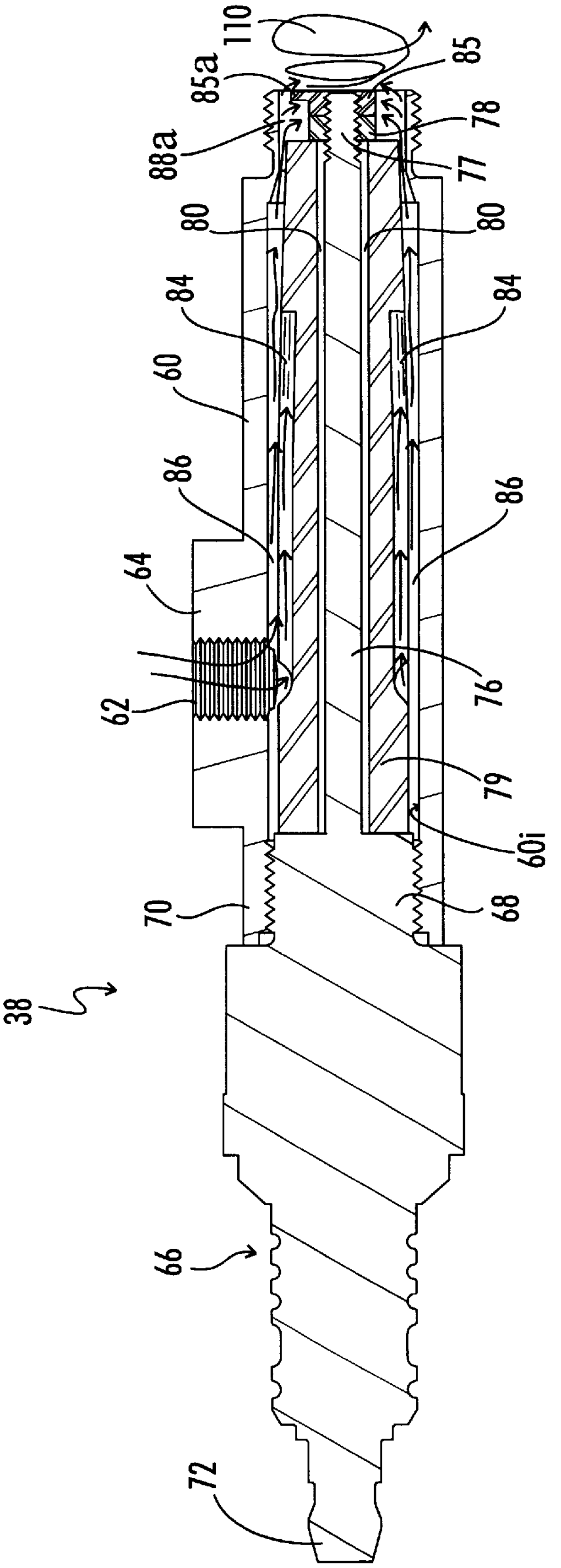

FIG. 4

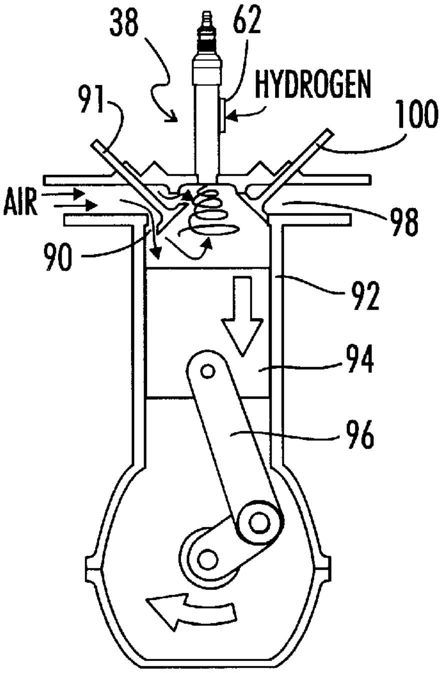
38
62
HYDROGEN
91
100
AIR
98
90
92
94
96

INTAKE

FIG. 5A

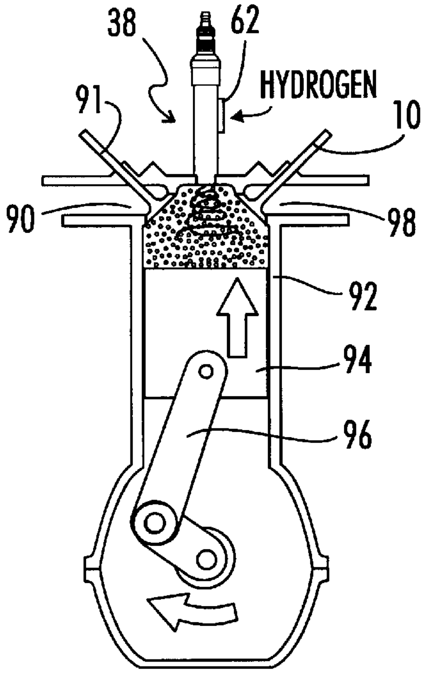
38
62
HYDROGEN
91
100
90
98
92
94
96

COMPRESSION

FIG. 5B

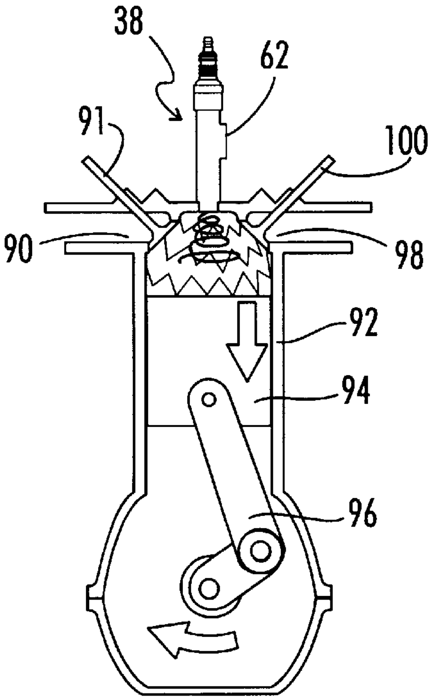
38
62
91
100
90
98
92
94
96

POWER

FIG. 5C

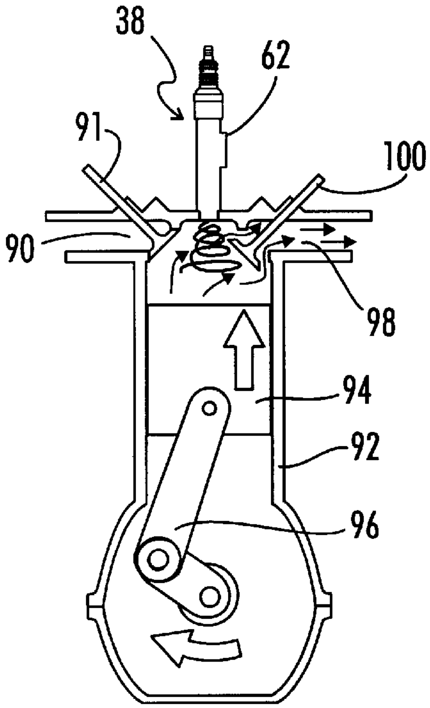
38
62
91
100
90
98
94
92
96

EXHAUST

FIG. 5D

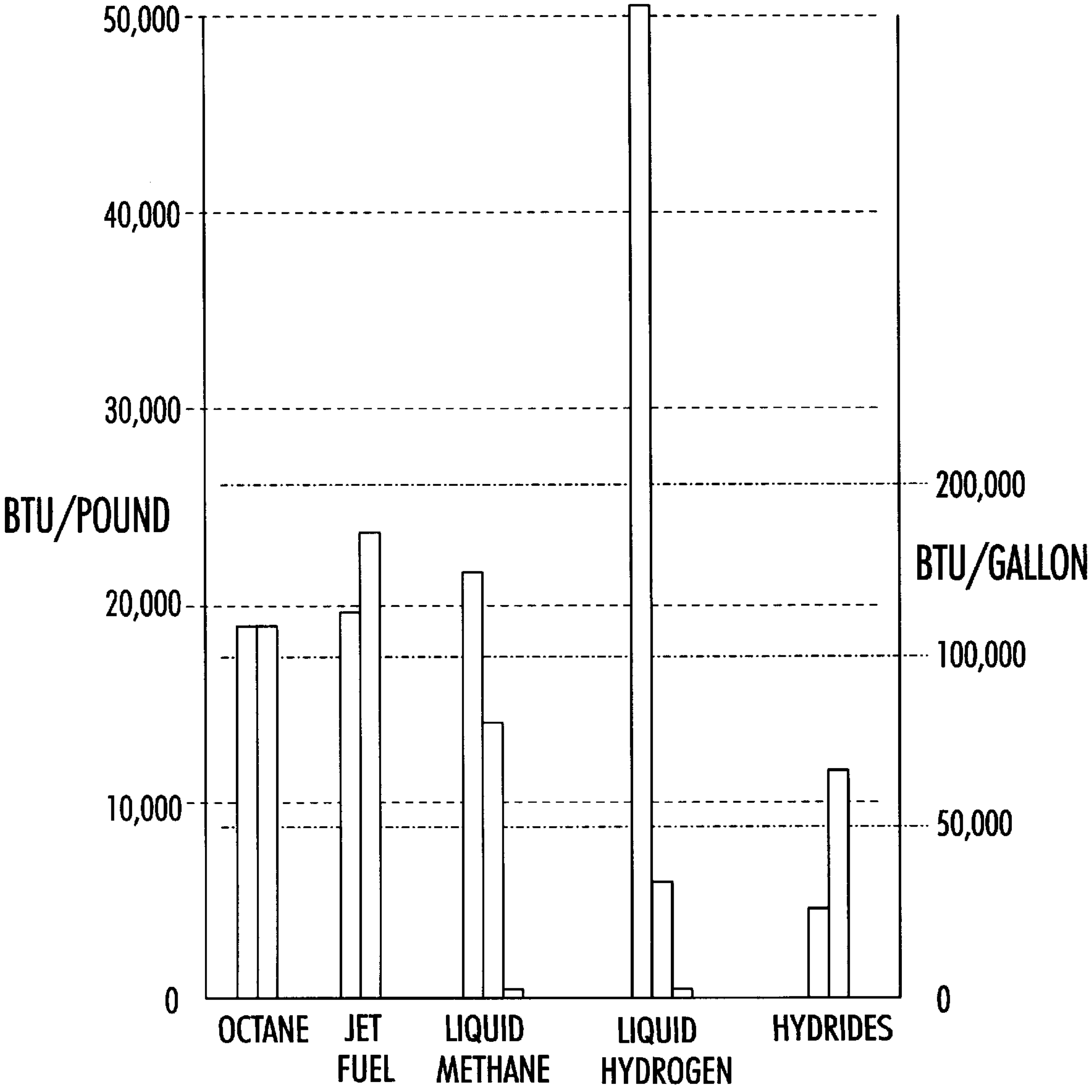

METHANE GAS - 140 BTU/GAL
HYDROGEN GAS - 40 BTU/GAL

FIG. 6

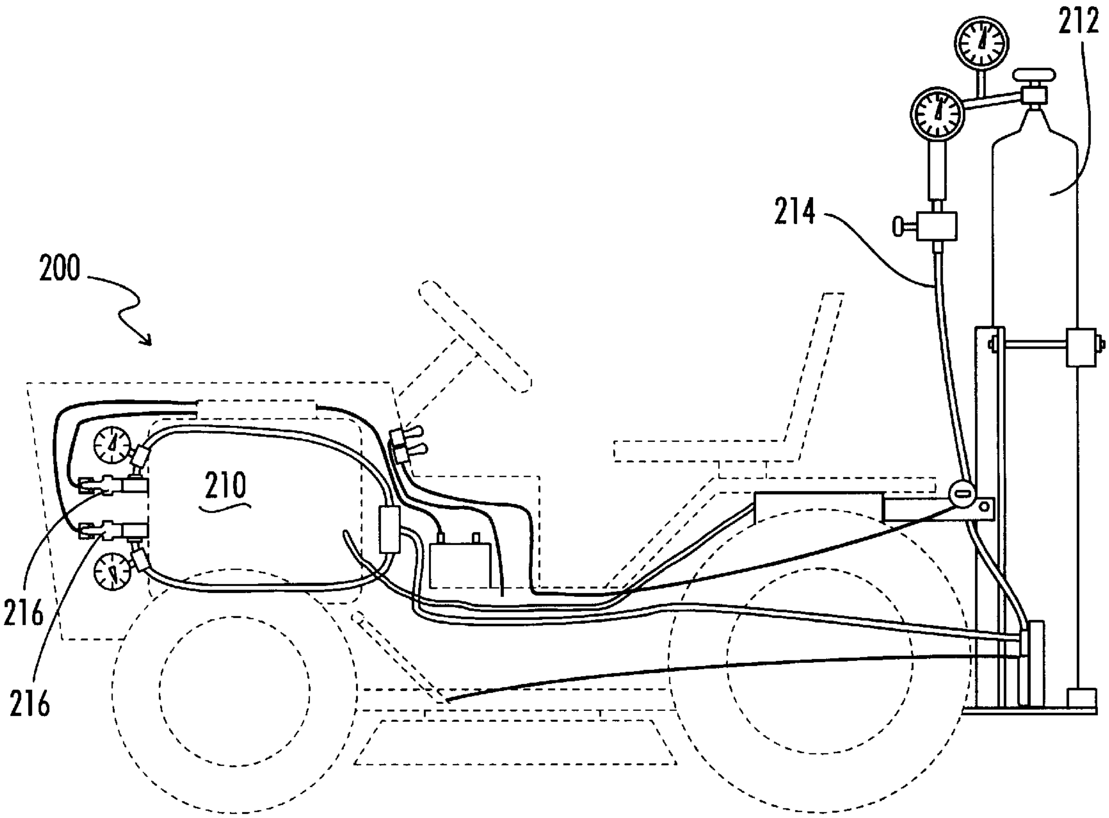
200
212
214
210
216
216

FIG. 7

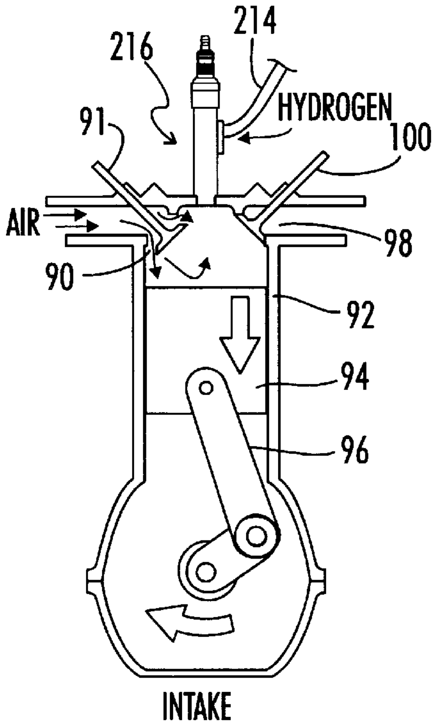

216
214
HYDROGEN
91
100
AIR
98
90
92
94
96
INTAKE

FIG. 8

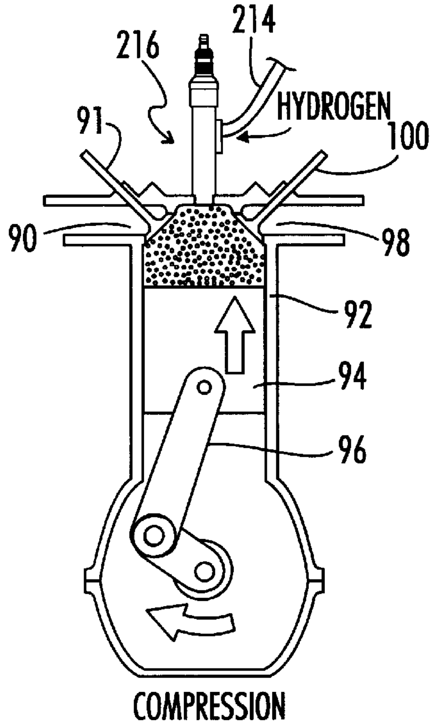

216
214
HYDROGEN
91
100
90
98
92
94
96
COMPRESSION

FIG. 9

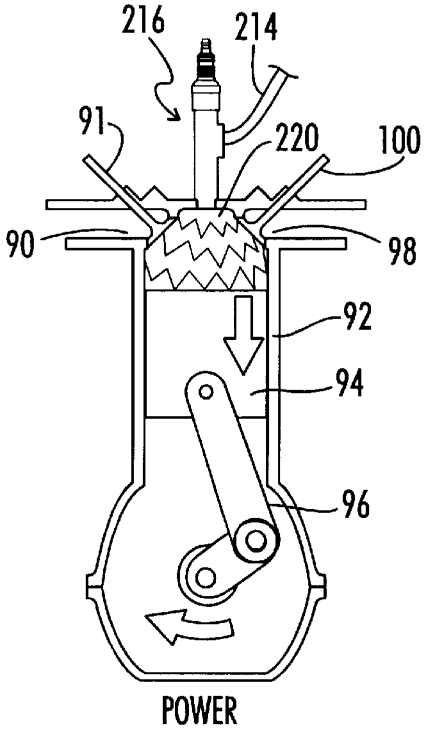

216
214
220
91
100
90
98
92
94
96
POWER

FIG. 10

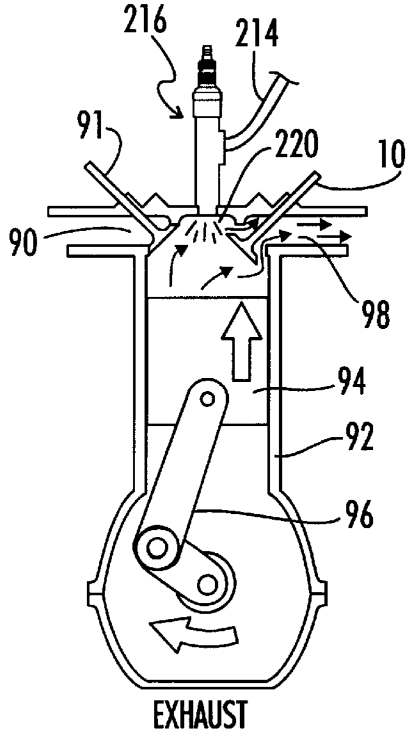

216
214
91
220
100
90
98
94
92
96
EXHAUST

FIG. 11

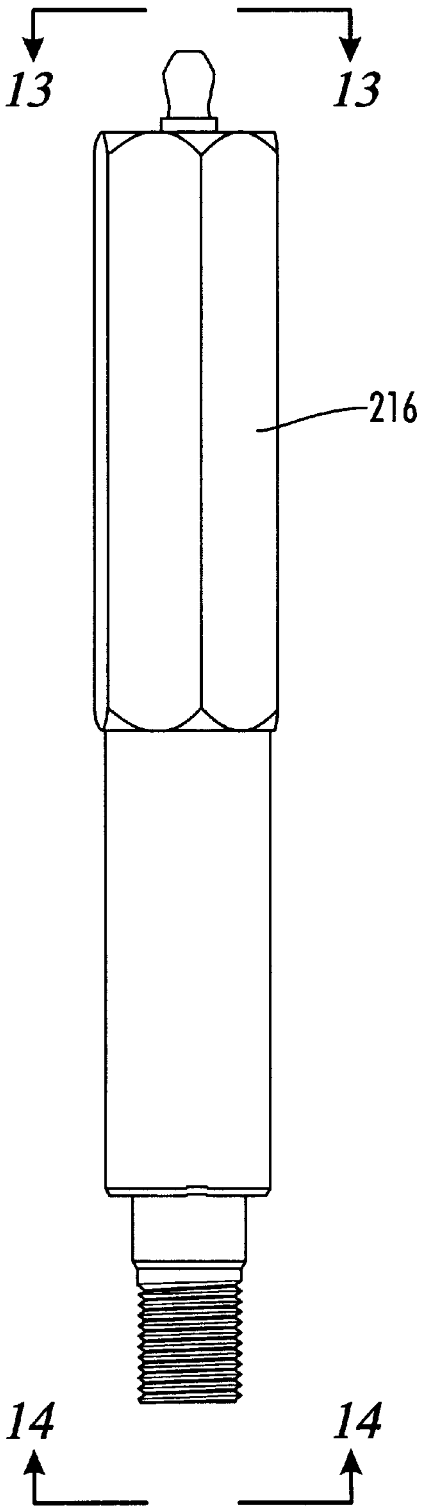
13
13
216
14
14

FIG. 12

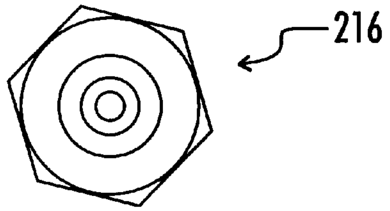

FIG. 13

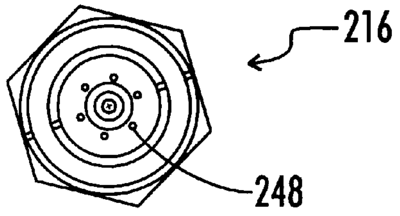

FIG. 14

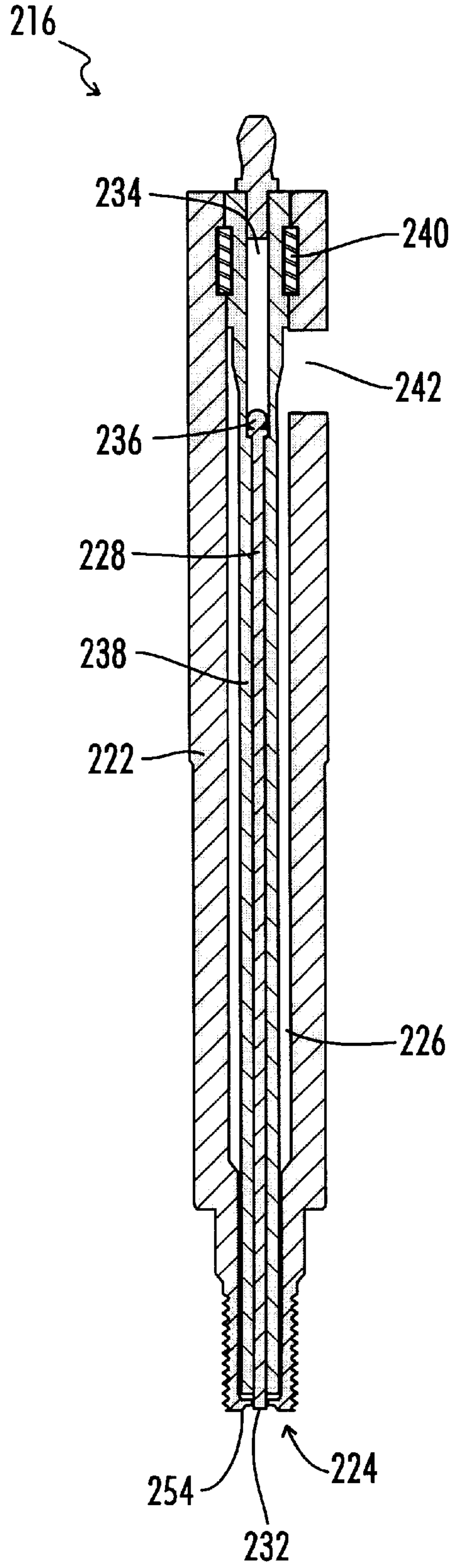

FIG. 15

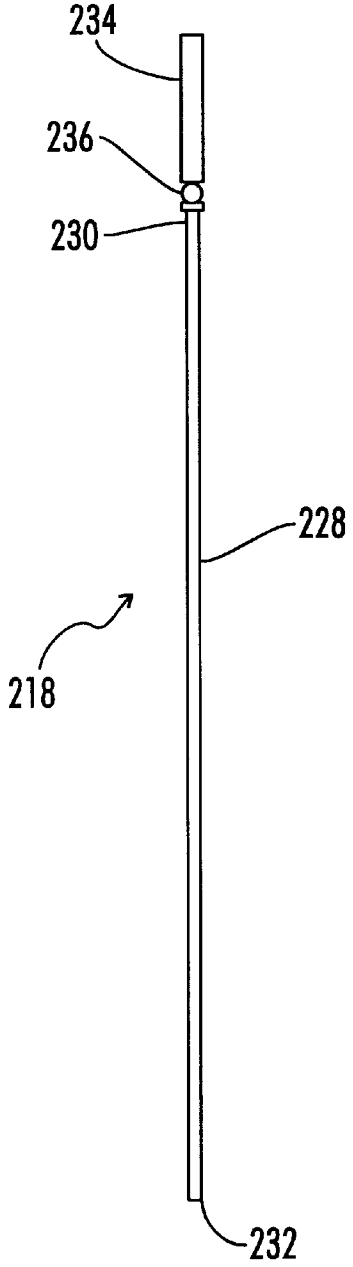

FIG. 16

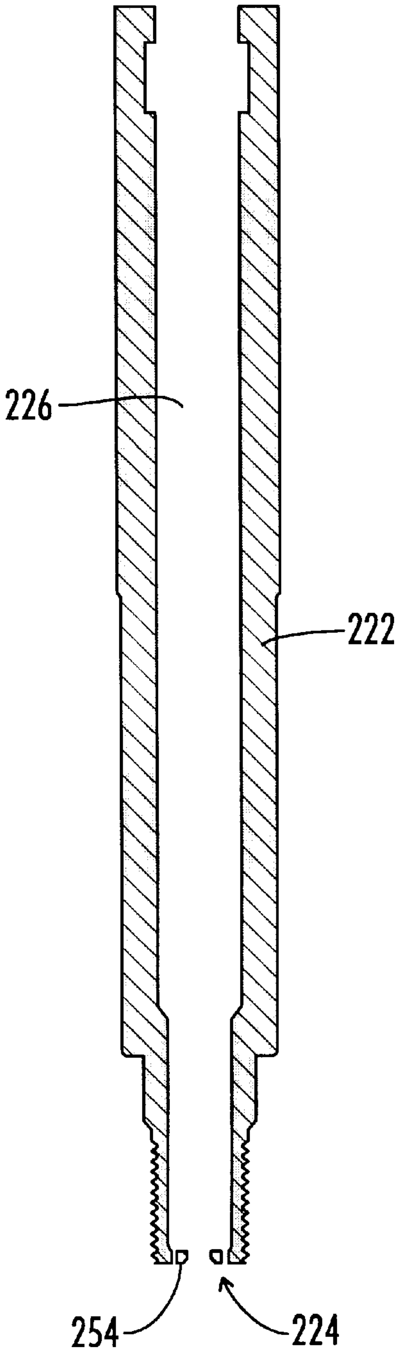
226
222
254
224

FIG. 17

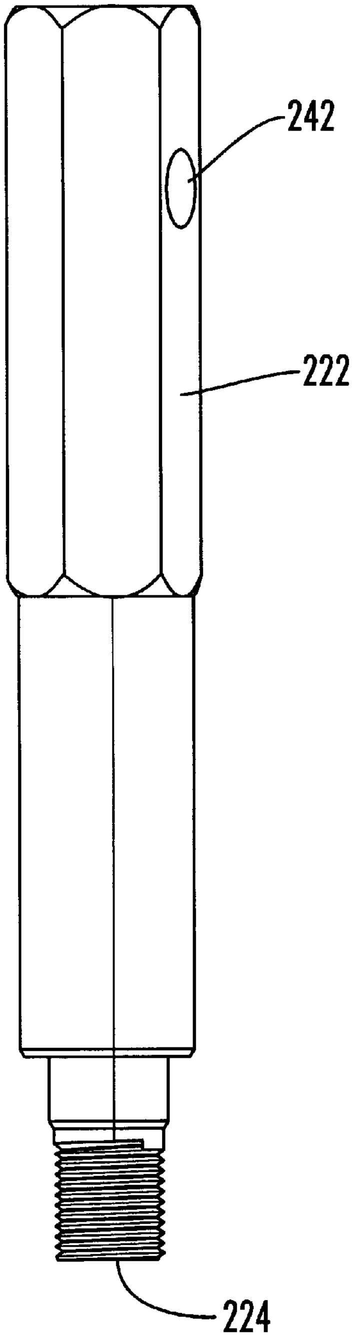

FIG. 18

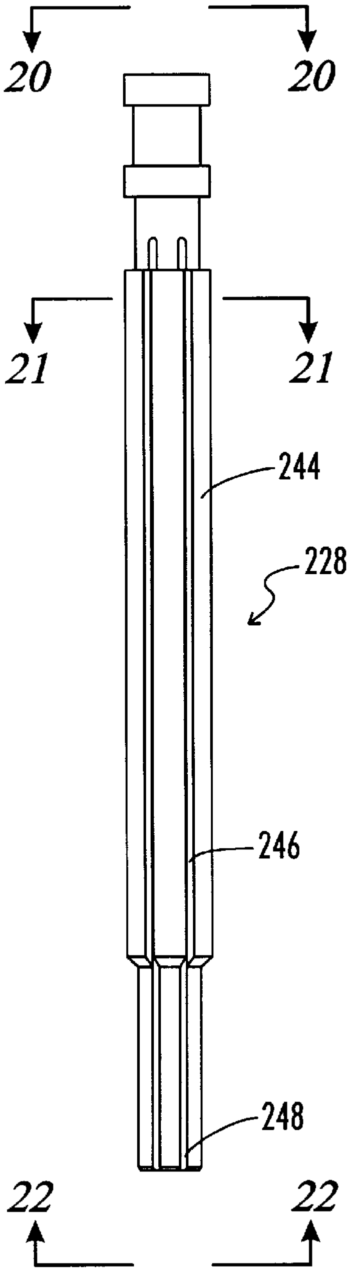
20
20
21
21
244
228
246
248
22
22

FIG. 19

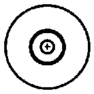

FIG. 20

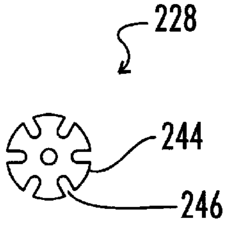

FIG. 21

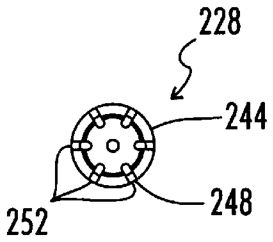
228
244
252
248

FIG. 22

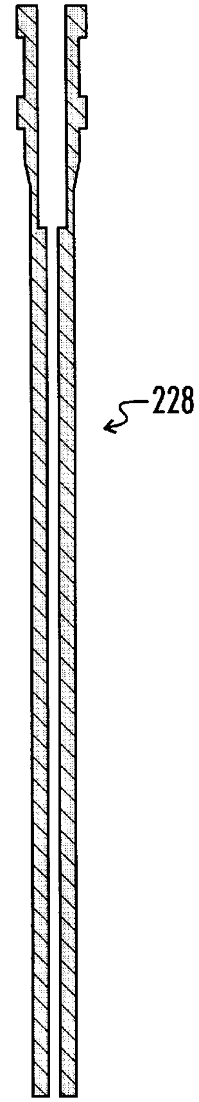

FIG. 23

HYDROGEN POWERED VEHICLE, INTERNAL COMBUSTION ENGINE, AND SPARK PLUG FOR USE IN SAME

This application is a continuation-in-part of U.S. patent application Ser. No. 08/905,862 filed Aug. 4, 1997 by Herman P. Anderson, entitled "Fuel System For Internal Combustion System And Adapter For Use In Same" , now U.S. Pat. No. 5,852,993. U.S. patent application Ser. No. 08/905,862 is hereby incorporated herein by reference.

BACKGROUND OF THE INVENTION

The present invention relates generally to a fuel system for internal combustion engines that enables the invention to utilize alternative fuels; and, more particularly to an improved fuel system for an internal combustion engine and an adapter and spark plug for use in same.

The problems of air pollution caused by automobiles, lawn mowers and other vehicles that have internal combustion engines powered by gasoline is well documented. This problem is particularly described in an article in the San Diego Union Tribune, Wednesday, Sep. 25, 1996 and in Popular Science, August 1996. Indeed, under the Clean Air Act, in the next few years cars that do not run on standard fossil fuels like gasoline will be required in several states. Currently, there are no vehicles available to meet this need.

Thus, it will be appreciated by those skilled in the art that there has been a push to seek other fuels for use in internal combustion engines. These fuels include methane, propane and gasohol.

Hydrogen has also been characterized as a promising fuel source. See Department of Energy publication entitled "Hydrogen Fuel", printed in 1978, U.S. Government Printing Office, U.S. Department of Energy, Office of Public Affairs, Washington, DC 20585. Particularly, the graph presented in FIG. **6**, which is copied from this publication, shows the potential for hydrogen in BTUs/Lb. However, the mass energy density for hydrogen is low; and thus, hydrogen needs to be compressed before its energy potential can be realized. Hydrogen is a "dry fuel" that is more difficult to ignite than "wet fuels." It becomes more difficult to ignite when it is cooled. However, cooling is preferred in most systems to avoid premature ignition of the fuel. Other systems deliver the hydrogen under pressure to prevent ignition in the supply line, which is a form of pre-ignition. No prior art fuel system has been able to capture the energy of hydrogen.

Several fuel systems for use with hydrogen have been proposed. Examples are described in U.S. Pat. Nos. 4,167,919; 4,253,428; 4,016,836; 4,178,882; 5,222,993; 5,085,176; and 5,085,176. The problem with these prior art fuel systems is they cannot take advantage of the alternative fuels to generate sufficient power from the fuels to drive an internal combustion engine for use in a lawn mower, automobile or other vehicle. As noted above, this is a particular problem when hydrogen is used as a fuel. Additionally, the prior art systems do not address the problem of back fire, i.e. undesired explosion of the fuel in the cylinder during the exhaust stroke of the internal combustion engine.

Premature ignition or backfire is often a problem with hydrogen fueled systems. U.S. Pat. No. 4,383,198 (Hosking) teaches a fuel injection spark plug with a pre-cooling chamber to avoid premature ignition indentination of a gaseous fuel supplied to the spark plug. U.S. Pat. No. 4,383,198, Hosking, entitled "Fuel Injection Spark Plug", issued May 10, 1983, is hereby incorporated herein by reference. However, cooling hydrogen makes it more difficult to ignite. This is because hydrogen is a dry fuel whereas gas is generally considered a wet fuel. Thus to reduce fuel pollution one prefers to utilize a fuel such as hydrogen and ignite it with a hotter spark.

One patent discussing very high energy (VHE) ignition systems is U.S. Pat. No. 5,207,208 issued to Ward, entitled "Integrated Convertor High Power CD Ignition", issued May 4, 1993. U.S. Pat. No. 5,207,208 is hereby incorporated herein by reference. Pat. '208 teaches very high energy ignition systems, for instance on the order of 36K volts. It teaches voltage doubling utilizing pulsating sparks of peak current on the order of two amps. While 36K volts may at times be adequate, it is preferable to achieve a higher voltage in the ignition chamber when using hydrogen.

It is known in the art to use a direct ignition system, also referred to as a distributorless ignition system, to increase the spark plug voltage. Such direct ignition systems use a "waste spark" method of spark distribution. Each cylinder is paired with its opposing cylinder in the firing order (1-4, 2-3, on a four cylinder, 1-4, 2-5, 3-6 on a V6) so that one cylinder on compression fires simultaneously with its opposing cylinder on exhaust stroke. Since the cylinder on exhaust stroke requires very little of the available voltage to fire its plug, most of the voltage is used to fire the cylinder on the compression stroke.

U.S. Pat. No. 4,462,380 discusses a distributorless ignition system. U.S. Pat. No. 4,462,380 entitled "Enhanced Spark Energy Distributorless Ignition System" issued Jul. 31, 1984 by Asik is hereby incorporated herein by reference. Pat. '380 teaches a distributorless ignition system of an internal combustion engine which has a supplementary spark energy module to increase spark energy. Each high voltage terminal is connected to a single spark plug and each ignition coil primary is alternately energized and quickly de-energized, producing opposite polarity ignition voltages at each coil terminal. As a result, pairs of spark plugs are alternately fired, with each firing pair occurring in a compression or exhaust stroke and thereby providing the proper ignition to the engine. Asik asserts that firing a spark plug does not affect engine performance or emissions. As taught by the instant invention this is incorrect for hydrogen fuels when the exhaust is sparked at a sufficient voltage.

U.S. Pat. No. 5,146,882 (Brinkman) teaches use of specific wet fuel mixtures in conjunction with direct ignition systems (D.I.S.). It teaches that technical breakdown voltages for typical spark plugs are on the order of 15–20 kV and DIS typically yield between 30 kV–40 kV. U.S. Pat. No. 5,146,882, entitled "Method And Apparatus For Cold Starting A Spark Ignited Internal Combustion Engine Fueled With An Alcohol-Based Fuel Mixture", by Brinkman, et al., issued Sep. 15, 1992, is hereby incorporated herein by reference. The previously referenced prior art discusses alternate fuels and fuel systems but does not teach a method of achieving a sufficient voltage with a hydrogen system.

What is needed, then, is a fuel system for an internal combustion engine that provides a way to generate sufficient power from the alternative fuels and that reduces backfire problems. Such a system is lacking in the prior art. Also lacking in prior art is a system utilizing alternate fuels with reduced pollution capabilities. Also lacking is a way of generating sufficient voltage spark to ignite the alternate fuel.

The following U.S. Patents discuss spark plugs designed for gas internal combustion engines to reduce radio frequency noise and interference. U.S. Pat. No. 4,224,554, entitled "Spark Plug Having A Low Noise Lever" issued to Nishio et al., Sep. 23, 1980 is hereby incorporated herein by reference. U.S. Pat. No. 4,029,990 entitled "Spark Plug Construction" issued to Nagy et al., Jun. 14, 1977 is hereby incorporated herein by reference.

SUMMARY OF THE INVENTION

It is an object of this invention to provide an improved fuel system for use with alternative fuels.

It is a further object of this invention to provide an improved fuel system for use particularly with hydrogen, and particularly to enhance the mass energy density of hydrogen.

It is an object to provide a hotter sparking plug to ignite the fuel.

It is still a further object of this invention to provide an adapter for use in a fuel system for an internal combustion engine that enhances the power obtained from the alternative fuels.

It is still a further object of this invention to provide an improved fuel system for an internal combustion engine that does not present the environmental hazards of a standard gasoline engine by providing reduced or zero emission of environmentally harmful gases.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that can use an alternative fuel and can use gasoline to facilitate the transition from gasoline to alternative fuels.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that reduces backfire problems.

It is yet a further object of this invention to provide an improved fuel system for an internal combustion engine that reduces backfire problems by cooling the electrode in the spark plug with the hydrogen gas fuel.

It is an object to reduce backfire by providing hydrogen at a sufficient pressure to reduce backfire and igniting it with a hotter spark (i.e. a higher voltage).

It is still a further object of this invention to provide an adapter for use in a fuel system for an internal combustion engine that enhances the power obtained from the alternative fuels.

It is a further object to generate a plasma field in a cylinder to burn residual exhaust and reduce emissions.

It is yet a further object of this invention to provide a direct injection fuel system wherein fuel is injected directly into the cylinder.

It is an object to provide a "2-fire" direct injection system, wherein the fuel is ignited during the power stroke and the exhaust stroke.

It is yet a further object of this invention to provide a way to convert the standard induction fuel system of a lawnmower to a direct injection fuel system wherein fuel is injected directly into the cylinder.

It is yet a further object of this invention to provide a way to convert the indirect injection fuel system of an automobile to a direct injection fuel system wherein fuel is injected directly into the cylinder.

The internal combustion engine fuel system of this invention includes a structure for mixing the alternative fuel, preferably hydrogen, with oxygen in ambient air to stratify the fuel. The term "alternative fuels" is known in the art and is meant to refer to fuels other than conventional gasoline.

The invention in one embodiment comprises an adapter, the adapter including a housing mounted between spark plug and cylinder of the internal combustion engine. A plug is placed within the housing. The plug has ridges or grooves on its outer surface that act as mixing structures. Thus, when hydrogen is introduced into the adapter housing it is mixed with ambient oxygen within the chamber as it flows over the plug. The mixing structures on the plug create a vortexing action as the hydrogen flows over the plug and towards the cylinder of the engine.

An electrode protrudes from the plug towards the cylinder. The electrode is preferably platinum and generates the necessary spark to create combustion of the hydrogen/air mixture adjacent to the cylinder to thereby power the cylinder in the engine. A platinum electrode is preferably used because it enhances a catalytic conversion of combustion by-products to more environmentally compatible products.

Thus, the combination of the compression action generated by the conventional piston and push rod within the cylinder of the engine in conjunction with the hydrogen/air mixing structure that vortexes the hydrogen/air create compressed hydrogen that is mixed with a combustion facilitator, i.e. oxygen in the ambient air to enable the system of this invention to capture the abundant energy found in hydrogen fuel.

The enhanced mixing capabilities of this invention is also adaptable for use with other alternative fuels such as natural gas, alcohol, propane and the like.

One embodiment of the present invention is for a hydrogen powered vehicle including an internal combustion engine operably connected to power the vehicle and adapted to connect to a hydrogen supply line. The internal combustion engine comprises a cylinder operably connected to the supply line; a piston operably disposed in the cylinder; and a spark plug having an electrode assembly in operable sparking communication with the cylinder, wherein the spark plug is capable of producing a spark of at least 45 kV.

Preferably the spark plug comprises a housing having a sparking end and a hollow chamber containing the electrode assembly. The electrode assembly comprises an electrode having a connector end and a sparking end, wherein the sparking end is positioned adjacent to the housing sparking end. The electrode assembly preferably comprises a coil electrically connected to the electrode connector.

Preferably the housing comprises a fuel inlet port in fluid communication with the supply line, and an insulator surrounding the electrode assembly, wherein the insulator is positioned in the hollow chamber. Preferably the insulator comprises an outer surface defining a first fuel channel having a first respective fuel exit port, wherein the first fuel channel is in fluid communication with the fuel inlet port and the fuel exit port exits into the cylinder.

The present invention also includes a method of operating an internal combustion engine comprising supplying hydrogen fuel to the engine and creating a spark in the cylinder during the power stroke and the exhaust stroke, wherein the spark created during the power stroke is at least 45 kV. In one embodiment the spark is at least 70 kV. In certain embodiments the method comprises generating plasma during the exhaust stroke. The exhaust fuel converts to an exhaust emission and coalesces with the plasma, thereby reducing fuel pollution.

Accordingly, it is an object in the invention to provide a spark plug producing a sufficiently hot spark to ignite hydrogen fuel, including cooled hydrogen fuel.

Another object of the present invention is to provide a device for reducing fuel exhaust emissions.

Another object of the present invention is to provide a cleaner burning engine.

Another object is to provide a spark plug for reducing emissions.

Another object of the present invention is to provide a device functional with dry fuels. Use of the term hydrogen, as discussed herein, is intended to encompass such dry fuels. However the term as used excludes such wet fuels as gas and alcohol.

Another object of the present invention is to provide an integrated fuel system adapted to provide these and objectives.

Other objects and advantages of the present invention will be apparent to those of skill in the art by reference to the teachings disclosed herein, including the attached drawings, detailed description of exemplary embodiments, and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a plan view of the fuel system of this invention with a lawn mower shown in phantom.

FIG. **2** is a perspective view of the adapter of the fuel system of this invention.

FIG. **3** is an exploded perspective view of the adapter of the fuel system of this invention.

FIG. **4** is a cross-sectional view of the adapter of the fuel system of this invention.

FIGS. **5A–5D** are a schematic of the 4 cycle combustion in an internal combustion engine including the adapter of this invention.

FIG. **6** is a graph which shows the energy potential for hydrogen in BTUs/Lb.

FIG. **7** shows a vehicle of the present invention utilizing hydrogen fuel.

FIG. **8** depicts an intake stroke of a cylinder connected to a hydrogen source.

FIG. **9** depicts a compression stroke of a cylinder connected to a hydrogen source.

FIG. **10** depicts a power stroke of a cylinder connected to a hydrogen source. The spark plug shown in FIG. **10** is depicted sparking.

FIG. **11** depicts a cylinder in an exhaust stroke connected to a hydrogen source, wherein the spark plug is sparking during the exhaust stroke.

FIG. **12** shows an elevated side view of a spark plug of the present invention.

FIG. **13** shows a plan view in the direction of view line **13—13** of the spark plug shown in FIG. **12**.

FIG. **14** depicts a plan view in the direction of view line **14—14** of the spark plug shown in FIG. **12**.

FIG. **15** depicts an elevated side section view of the spark plug shown in FIG. **12**.

FIG. **16** shows a electrode assembly which is disposed in the hollow chamber depicted in FIG. **15**.

FIG. **17** shows a section view of the spark plug housing clearly showing the hollow chamber and a torridal ground electrode.

FIG. **18** depicts an elevated side view of the spark plug housing and a fuel inlet port.

FIG. **19** depicts an insulator comprising fuel channels and fuel exit ports.

FIG. **20** is a top plan view of the insulator shown in FIG. **19** viewed in the direction of view line **20—20**.

FIG. **21** is a cross section view of the insulator shown in FIG. **19** along section line **21—21**.

FIG. **22** is a cross section view of the insulator shown in FIG. **19** viewed in the direction of view line **22—22**.

FIG. **23** is an elevated section view of the insulator shown in FIG. **19**.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring now to the drawings, wherein like reference numerals refer to like parts throughout, the fuel system is referred to generally at **10**. The fuel system of this invention is contemplated for use with an internal combustion power system, such as an internal combustion engine or such as a home heating system. Fuel system **10** is portrayed in this case as connected to a standard lawn mower. The lawn mower is shown in phantom at **12** in FIG. **1**. The lawn mower is a conventional lawn mower as can be purchased from a company such as Murray of Ohio. The internal combustion engine in the lawn mower is a standard 4-cycle engine with two cylinders and develops 18 horsepower. A suitable example can be purchased from Briggs & Stratton, Milwaukee, Wis. 53201.

Referring particularly to FIG. **1**, system **10** includes fuel tank **14** that is mounted to lawn mower **12** via mounting assembly **15**. In the preferred embodiment, fuel tank **14** includes hydrogen gas. Accordingly, for the purposes of the foregoing detailed description, hydrogen gas will be described.

A line **22** begins at fuel tank **14** and proceeds to back-flow valve **16**. Back-flow valve **16** is a standard valve and is available from AirCo Gases Company, 575 Mountain Avenue, Murray Hill, N.J. 07974, part no. WCS CV-4M. PSI tank gauge **19** is inserted in line **22** between back-flow valve **16** and flow pressure gauge **18**. PSI tank gauge **19** is available from AirCo Gases Company also, part no. WCS CV5M, and serves as an indicator of fuel remaining in the tank **14**. Flow pressure gauge **18** is used to monitor the flow of the hydrogen in pressure units. The flow is started using valve **16**. The optimal flow pressure is 25 pounds of flow pressure from the tank **14** to the cylinder of the engine.

The flow of hydrogen is regulated via valve **20** that is part of gauge **18**. Thus, when valve **20** is opened, hydrogen proceeds through line section **22** to hydrogen flow solenoid **26A**. The user can access the hydrogen fuel by flipping hydrogen flow switch **24A** into the "on" mode so that hydrogen gas in line **22** can proceed through hydrogen flow solenoid **26A** and into line section **28**.

Throttle **30** regulates the flow of hydrogen from line section **28** into line section **29** and ultimately, to engine **13**. Throttle **30** is connected to throttle pedal **50** via includes throttle cable **48**. Throttle pedal **50** also operates the butterfly valves in the carburetor in engine **13** in a conventional manner.

As an option, the throttle can be controlled electronically by a digital fuel controller product. Such a product is commercially available from Autotronic Controls Corp., 1490 Henry Brennan Drive, El Paso, Tex. 79936.

Line portion **29** proceeds through splitter **32** and is split into upper line **34** and lower line **36**. Lines **34** and **36** proceed to adapters **38**. Additional back-flow prevention valves **39** and **41** are mounted between lines **34** and **36** and adapters **38**. Braided Teflon® is the preferred material for the fuel lines of the system.

Battery **42** is connected via battery cable **43** to magneto **44**. Magneto **44** then controls the current through spark plug lines **46** to adapters **38** in a conventional manner. Thus, the regulation of the spark that ignites the fuel in the cylinder of engine **13** is conventional. Additionally, the intake of air into the engine **13** is accomplished through the standard carburetor system of engine **13**. In a car, the sparking of the spark plugs can be controlled by the standard distributor/battery ignition system.

A second solenoid **26**B controls the flow of gasoline from tank **23** through gasoline line **52**. Solenoid **26**B is controlled by switch **24**B. Throttle **30** is connected to gasoline line **52** and can also be used to control the flow of gasoline to engine **13**. Thus, this invention can be used with both an alternative fuel and gasoline.

Referring now to FIGS. **2**, **3** and **4**, the adapter **38** of this invention is shown in more detail. Adapter **38** includes housing **60**. A threaded port **62** is formed into housing **60** through port block **64**. Hydrogen lines are threadably connected to the adapter **38** via port **62**. Port **62** is also referred to as an entrance port. The entrance port has an entrance area **62***a*.

Adapter **38** also includes spark plug **66** which is effectively a standard spark plug except for the modification to the electrode as described below. Spark plug **66** includes a threaded end **68** which is threadably received in the threaded upper end **70** of housing **60**. Spark plug wires **46** (as seen in FIG. **1**) are connected to connector end **72** of spark plug base **66**.

Electrode **76** protrudes from spark plug **66** and is of a length sufficient so that the threaded tip **77** (best seen in FIGS. **3** and **4**) of the electrode **76** is proximate to the cylinder of the internal combustion engine when the adapter **38** is mounted. In one embodiment the electrode **76** is coterminus with an exit port **88** of the adapter **38**. This most clearly is shown in FIG. **4**. Electrode **76** is preferably platinum to enhance catalysis of combustion by-products, as will be more fully discussed below.

Adapter **38** also includes plug **79**. Plug **79** includes an axial internal void space **80**. Thus, plug **79** fits over electrode **76** via internal void space **80**. Plug **79** preferably comprises a ceramic material that acts as a good insulator. A suitable source for ceramic plug **79** is AM/PRO Machinery, Inc. 134 Church Street, Plain City, Ohio 43064-1321.

An important aspect of this invention is found on the external surface of plug **79**. An annular groove **82** is formed on the surface of plug **79** near its distal end **81** so as to be proximal to the port **62** where hydrogen enters housing **60**. Two stratifying grooves **84** proceed along the surface of plug **79** until they reach its distal end **83**. It is these mixing structures, i.e. annual groove **82** and stratifying grooves **84**, that facilitate the improved power generation from alternative fuels as provided by this invention. It will be apparent to those skilled in the analogous arts that mixing structure may be formed in the inner surface of the adapter housing **60**.

Alternatively a combination of vortex producing structures is integral with both the plug **79** and adapter housing **60**. Also any conventional vortex producing means should be sufficient groves, per se, are not required.

On threaded tip **77** of electrode **76**, there are mounted a spacer nut **78** and a spark gap nut **85**. As best seen in FIG. **2**, the spark gap nut **85** includes a protrusion **85***a* that provides the spark gap between electrode **76** and periphery **89** of the exit port **88** of housing **60**. The exit port **88** includes an exit area **88***a*. Preferably, the exit area **88***a* is no larger than the entrance area **62***a*, though this is not critical. Preferably, the spark gap nut **85** is brass and the preferred spark gap distance is $^{25}/_{1000}$ of an inch. Of course, the electrode may be coterminus with the plug distal end **83**.

As best seen in FIG. **4**, where arrows represent gas flow, hydrogen gas flows into chamber **60** via port **62**. The hydrogen gas cools electrode **76**, which facilitates combustion. Stratifying grooves **84**, preferably there are two, act to mix the hydrogen that is flowing into housing **60** with ambient air therein. Though more or less vortex creating grooves (stratifying grooves) may be desired. Typically this will depend on the relative dimensions of the grooves and diameter of the plug as well as the pressure or velocity of the mixture in the housing. The mixing occurs within void space **86** of housing **60** and creates a vortexing or tornado action. Plug **79** tapers as it proceeds from its distal end **81** to its proximal end **83**. Thus, proximal end **83** is smaller in diameter than distal end **81**. This provides additional space within void space **86** for the mixing of the hydrogen and air. The stratified hydrogen/air are twisting in a vortex fashion when leaving housing **60** via exit port **88**. The vortexing action produced by grooves **84** on the plug **79** creates a "tornado" **110** of hydrogen/air with a hollow center that generates a vacuum. Similar effect would result were t he grooves to be placed on the housing. The vacuum propels the hydrogen/air mixture into the cylinder of the engine for combustion. Such mixing cannot be accomplished in a conventional induction system.

Referring now to FIG. **5**, a schematic of the system of this invention is portrayed, with arrows representing gas flow. Hydrogen flows in through port **62** through adapter **38** and into cylinder **92**. Air enters cylinder **92** through air intake **90** and air intake valve **91**. The stratified hydrogen/air mixture is compressed in the compression stroke of the engine via piston **94** and push rod **96**, which are standard structures in an internal combustion engine. In the power stroke shown in FIG. **5**-C, a spark is generated in the compressed stratified hydrogen/air mixture ignites and drives piston **94** and push rod **96** to power the engine.

The mass energy density of hydrogen is enhanced through the increased pressure on the hydrogen exerted by piston **94** on the hydrogen/air mixture within the cylinder **92** in the uptake stroke in an internal combustion engine. The pressurized and well-mixed hydrogen/air mixture is rapidly ignited through the spark because hydrogen has a fast flame front. This reduces problems with timing. Thus, the energy store found in hydrogen is captured through this system via the compression of the hydrogen gas within cylinder **92** and via the thorough mixing of hydrogen and air provided by adapter **38**.

In addition, in the exhaust phase in a four cycle engine a second spark is generated from the plug to ensure complete combustion of by-products. This is enhanced by using an electrode made of platinum. Exhaust leaves cylinder **92** through exhaust port **98**. Exhaust port **98** is closed during other phases by exhaust outlet valve **100**.

In the exhaust stroke, because air valve **91** is closed, the substantial component of the gases in cylinder **92** is hydrogen. This is also accomplished due to the fact that in the system of this invention, the hydrogen is continuously flowing. When this spark interacts with what is predominately hydrogen, a plasma is formed. This plasma blocks back-flow of air from exhaust port **98**. Prevention of back-flow of air prevents the hydrogen fuel system from backfiring, which is another common problem of hydrogen fuel systems. In effect, then, the plasma formed by the hydrogen acts as a virtual valve in that the plasma blocks the back-flow of air.

When the plasma of the hydrogen is generated as described above, UV radiation is generated which also facilitates consumption of combustion by-products. This increases the environmental friendliness of the system.

In addition to hydrogen the following fuels can be used in this system: natural gas; alcohol; gasohol; propane and buthane. A mixture of hydrogen and natural gas can also be used. As noted above, gasoline can be used in this system when solenoid **26** is used to stop the flow of hydrogen.

Further, it is contemplated that an adapter that is described herein for use with an internal combustion engine can also be used with a home heating system such as a home heating system that may currently use natural gas. The vortexing action created by the plug would enable appropriate power to be captured from the hydrogen fuel so as to effectively heat a home.

Another embodiment of the invention is a hydrogen powered vehicle **200** shown in FIG. **7**. The hydrogen powered vehicle **200** includes an internal combustion engine **210** operably connected to power the vehicle **200** and adapted to connect to a hydrogen supply **212** through a supply line **214**.

Referring to FIG. **8**, the internal combustion engine **210** comprises a cylinder **92** operably connected to the supply line **214**. A piston **94** is operably disposed in the cylinder **92**. A spark plug **216** having an electrode assembly **218** (not shown in FIG. **8**) is in operable sparking communication with the cylinder **92**, wherein the spark plug **216** is capable of producing a spark **220** (shown in FIG. **10**) of at least 45 kilovolts. FIGS. **8**–**11** are substantially similar to FIGS. **5**A–**5**D depicting schematic representations of a 4-cycle piston-cylinder engine. Other piston and cylinder embodiments will be apparent to those with skill in the art. FIG. **8** represents an intake stroke. FIG. **9** depicts a compression stroke. FIG. **10** depicts a power stroke, in which the spark plug **216** sparks. FIG. **11** depicts an exhaust stroke, where the spark plug **216** is sparking. Refer to FIGS. **5**A–**5**D discussed earlier for further description of similar components of FIGS. **8**–**11**.

FIG. **12** shows one embodiment of the spark plug **216** in an elevated side view. FIG. **13** shows a plan view of the spark plug **216** along line **13**—**13** shown in FIG. **12**. FIG. **14** shows a plan view of the spark plug **216** in the direction of line **14**—**14** shown in FIG. **12**.

FIG. **15** depicts a section view of the spark plug **216**. FIG. **15**, section view, depicts a representative spark plug **216**. The spark plug shown in FIG. **15** comprises a housing **222** having a sparking end **224** and a hollow chamber **226** containing the electrode assembly **218**.

FIG. **16** depicts the electrode assembly **218** removed from the hollow chamber **226**. The electrode assembly **218** depicted in FIG. **16** comprises an electrode **228**. The electrode **228** has a connector end **230** and a sparking end **232**. As shown in FIG. **15** the sparking end **232** of the electrode **228** is positioned adjacent to the housing sparking end **224**. The electrode assembly further comprises a coil **234** electrically connected to the electrode connector end **230**.

Referring to FIGS. **10** and **11**, wherein the piston **94** moves through a power stroke in FIG. **10** and an exhaust stroke in FIG. **11**, the spark plug **216** is shown to fire during the power stroke and the exhaust stroke.

The electrode assembly **218** shown in FIG. **16** comprises a steel ball **236** positioned between the coil **234** and the electrode connector end **230**. Other operable ball-coil-electrode arrangements will be apparent to those of skill in the art.

The spark plug **216** shown in FIG. **15** also comprises an insulator **238** surrounding the electrode assembly **218** and positioned in the hollow chamber **226**. A ceramic pore **240** is shown stabilizing the insulator **238** in the hollow chamber **226**. Other methods of locating the insulator in the hollow chamber will be apparent to those of skill in the art.

An elevated section view of the housing **222** is shown in FIG. **17**. An elevated side view of the housing **222** is shown in FIG. **18**. In the exemplary embodiment of the housing **222** shown in FIG. **18**, the housing **222** comprises a fuel inlet port **242** in fluid communication with the supply line **214** (not shown). FIG. **19** depicts an exemplary embodiment of the insulator **238**. The insulator **238**, shown in FIG. **19**, comprises an outer surface **244** defining a first fuel channel **246** having a first respective fuel exit port **248**. The first fuel channel **246** is in fluid communication with the fuel inlet port **242**. The fuel exit port **248** exits into the cylinder **292**.

FIG. **20** shows a plan view of the insulator **228** shown in FIG. **19** along line **20**—**20**. FIG. **21** shows a section view of the insulator **228** shown in FIG. **19** viewed along section line **21**—**21**. FIG. **22** depicts a bottom plan view of the insulator **228** shown in FIG. **19** along line **22**—**22**. FIG. **23** shows a section view of the insulator **228** cut through section line **23**—**23** shown in FIG. **19**.

In the embodiment of the insulator **228** shown in FIGS. **19**–**22**, the insulator outer surface **244** defines a plurality of fuel channels **250**, including the first fuel channel **246**. The fuel channels **250** include a plurality of respective fuel exit ports **252**, including the first fuel exit port **248**. This is shown clearly in FIG. **22**. Preferably, the plurality of fuel channel exit ports **252** are spaced equally, or equi-distant, about the electrode **228**. This is shown well in FIGS. **21** and **22**.

Referring to FIGS. **15** and **17**, one embodiment of the housing **222** sparking and **224** comprises a torridal ground electrode **254** positioned about the electrode sparking end **232** (shown in FIG. **15**). Preferably the housing sparking end **224** and the electrode sparking end **232** are co-terminous. In one embodiment the first respective fuel exit port **248** is co-terminous with the housing sparking end **224** and the electrode sparking end **232**.

More generally the present invention includes an internal combustion engine **210** adapted to operate on hydrogen fuel supplied through a supply line **214**. Typically the engine comprises a first cylinder **92** in fluid communication with the supply line **214**; a first piston **94** disposed in the first cylinder **92** to move through a power stroke and an exhaust stroke (See FIGS. **10** and **11**) in the first cylinder **92**. The engine **210** also comprises a spark plug **216** having a housing **222** including a sparking end **224** and an electrode assembly **218** located in the housing **222**, wherein the sparking end **224** is in operable communication with the cylinder **92** and capable of producing a spark **220** of at least 45 kV. Preferably, the spark plug **216** fires during the power stroke and the exhaust stroke. In one embodiment the housing sparking end **224** comprises a torridal ground electrode **254**. The electrode assembly **218** comprises an electrode **228** having a sparking end **230** capable of sparking toward the torridal ground electrode **254**. Preferably the electrode assembly **218** comprises a coil **234** in electrical communication with the electrode **228**.

In one embodiment the electrode assembly **218** comprises an electrode **228** having a sparking end **230** and a coil **234** electrically connected to the electrode **228** opposite the electrode sparking end **230**. This is the embodiment shown in FIG. **15**. Typically the spark plug **216** comprises an insulator **228** surrounding the electrode assembly **218**. Preferably the insulator **228** includes a fuel channel **246** having a fuel exit port **248** near the electrode sparking end **232**.

It will be apparent to those with skill in the art that the present invention also comprises a method of operating an internal combustion engine **210** comprising the steps of supplying hydrogen fuel to the engine **210**; moving a piston **94** in a cylinder **92** through a power stroke and an exhaust stroke; creating a spark **220** in the cylinder **92** during the power stroke and the exhaust stroke, wherein the spark created during the power stroke is at least 45 kV; and supplying the fuel to a location sufficiently near the spark to ignite the fuel. This is shown well in FIGS. **10**, **15** and **19**.

The present invention also includes a method wherein the step of creating a spark **220** comprises passing an electric current (not shown) through an electrode assembly **218** including a coil **234** electrically connected to an electrode **218** having a sparking end **232**; insulating the electrode assembly **218**; and arcing the electric current from the electrode sparking end **232** to a ground electrode **254**. The ground electrode **254** as shown in FIGS. **15** and **17** a toroidal ground electrode **254**.

In certain embodiments the step of creating a spark **220** includes creating a spark **220** having at least 70 kV. Creating a spark of sufficient voltage will generally reduce the exhaust pollution. Typically this happens by generating a plasma or plasma field during the exhaust stroke. Thus, one method of the present invention further comprises the steps of generating plasma during the exhaust stroke; and allowing exhaust fuel converted to an exhaust state to coalesce with the plasma. This reduces exhaust pollution.

Thus, the apparatus and system of this invention provides an effective means to use alternative fuels. The use of alternative fuels can enhance the environment by reducing the use of conventional gasoline and the environmentally detrimental by-products caused by the consumption thereof.

Thus, although there have been described particular embodiments of the present invention of a new and useful "Hydrogen Powered Vehicle, Internal Combustion Engine, And Spark Plug For Use In Same", it is not intended that such references be construed as limitations upon the scope of this invention except as set forth in the following claims. Further, although there have been described certain dimensions used in the preferred embodiment, it is not intended that such dimensions be construed as limitations upon the scope of this invention except as set forth in the following claims.

What is claimed is:

1. A hydrogen powered vehicle including an internal combustion engine operably connected to power the vehicle and adapted to connect to a hydrogen supply through a supply line, the internal combustion engine comprising:

- a cylinder operably connected to the supply line;
- a piston operably disposed in the cylinder;
- a spark plug having an electrode assembly in operable sparking communication with the cylinder, wherein the spark plug is capable of producing a spark of at least 45 kV, the spark plug including:
  - a housing having a sparking end and hollow chamber, the housing including a fuel inlet port in fluid communication with the supply line;
  - an electrode assembly, received in the hollow chamber and including an electrode having a connector end and a sparking end, the sparking end positioned near the housing sparking end;
  - an insulator surrounding the electrode assembly and positioned in the hollow chamber,
  - the insulator including an outer surface defining a first fuel channel having a first respective fuel exit port, wherein the first fuel channel is in fluid communication with the fuel inlet port and the fuel exit port exits in the cylinder.

2. The apparatus of claim **1**, wherein the piston moves through a power stroke and an exhaust stroke in the cylinder, and the spark plug fires during the power stroke and the exhaust stroke.

3. The apparatus of claim **2**, wherein:

- the spark plug comprises a housing having a sparking end and hollow chamber containing the electrode assembly; and
- the electrode assembly comprises an electrode having a connector end and a sparking end, the sparking end positioned adjacent to the housing sparking end and the electrode assembly further comprises a coil electrically connected to the electrode connector end.

4. The apparatus of claim **1**, wherein the electrode assembly comprises a steel ball positioned between a coil and the electrode connector end.

5. The apparatus of claim **3**, wherein the insulator outer surface defines a plurality of fuel channels, including the first fuel channel, wherein the fuel channels include a plurality of respective fuel exit ports, including the first respective fuel exit port.

6. The apparatus of claim **5**, wherein the plurality of fuel channel exit ports are spaced equally about the electrode.

7. The apparatus of claim **6**, wherein the housing sparking end comprises a toroidal ground electrode positioned about the electrode sparking end, and wherein the housing sparking end and the electrode sparking end are co-terminous.

8. The apparatus of claim **7**, wherein the housing sparking end comprises a toroidal ground electrode positioned about the electrode sparking end, and wherein the housing sparking end, the electrode sparking end, and the first respective fuel exit port are co-terminous.

9. An internal combustion engine adapted to operate on hydrogen fuel supplied through a supply line, the engine comprising:

- a first cylinder in fluid communication with the supply line;
- a first piston operably disposed in the first cylinder to move through a power stroke and an exhaust stroke in the first cylinder;
- a spark plug having housing including a sparking end, and an electrode assembly located in the housing, wherein the sparking end is in operable communication with the cylinder and capable of producing a spark of at least 45 kV; and wherein
  - the spark plug fires during the power stroke and the exhaust stroke;
  - the spark plus comprises an insulator located in the housing around the electrode assembly;
  - the housing comprises a fuel inlet port in fluid communication with the supply line; and
  - the insulator includes a first fuel channel in fluid communication with the fuel inlet port, the first fuel channel having a respective first fuel exit port exiting into the cylinder.

10. The device of claim **9**, wherein the housing sparking end comprises a toroidal ground electrode, and wherein the electrode assembly comprises an electrode having a sparking end capable of sparking toward the toroidal ground electrode.

11. The device of claim **10**, wherein the electrode assembly comprises a coil in electrical communication with electrode.

12. The device of claim **11**, wherein the electrode sparking end, the toroidal ground electrode, and the first respective fuel exit port are co-terminus.

13. The device of claim **12**, wherein:

the electrode assembly comprises an electrode having a sparking end, and a coil electrically connected to the electrode opposite the electrode sparking end.

* * * * *

US006882128B1

(12) **United States Patent**
Rahmel et al.

(10) **Patent No.: US 6,882,128 B1**
(45) **Date of Patent: Apr. 19, 2005**

(54) **METHOD AND SYSTEM FOR ENERGY RECLAMATION AND REUSE**

(75) Inventors: **Ronald S. Rahmel**, Encinitas, CA (US); **Gerald C. Gerace**, Chula Vista, CA (US)

(73) Assignee: **Science Applications International Corporation**, San Diego, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 776 days.

(21) Appl. No.: **09/670,613**

(22) Filed: **Sep. 27, 2000**

(51) **Int. Cl.**^{7} **H04B 1/38**; H04M 1/00

(52) **U.S. Cl.** **320/101**; 455/343.1; 455/572; 320/101; 320/124; 320/127

(58) **Field of Search** 455/343.1, 343.2, 455/343.3, 572, 573, 343.5; 320/101, 102, 124, 127, 137, 138; 343/702, 720

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,434,678	A		3/1969	Brown et al.	244/1
3,852,755	A		12/1974	Works et al.	343/701
3,927,375	A		12/1975	Lanoe et al.	325/363
4,031,449	A		6/1977	Trombly	320/2
4,079,268	A		3/1978	Fletcher et al.	307/151
4,360,741	A	*	11/1982	Fitzsimmons et al.	307/151
4,628,299	A		12/1986	Tate et al.	340/540
4,685,047	A	*	8/1987	Phillips, Sr.	363/126
5,043,739	A		8/1991	Logan et al.	343/701
5,416,486	A		5/1995	Koert et al.	342/42
5,440,300	A		8/1995	Spillman, Jr.	340/825.54
5,495,519	A		2/1996	Chen	379/58
5,710,988	A		1/1998	Suzuki et al.	455/90
5,842,838	A		12/1998	Berg	417/331
5,889,383	A		3/1999	Teich	320/107
5,898,932	A		4/1999	Zurlo et al.	455/573
5,970,393	A		10/1999	Khorrami et al.	455/66
6,046,676	A	*	4/2000	Ward et al.	340/572.1
6,154,664	A	*	11/2000	Chorey et al.	455/571
6,184,651	B1		2/2001	Fernandez et al.	320/108
6,198,250	B1	*	3/2001	Gartstein et al.	320/112
6,275,010	B1		8/2001	Neuteboom	320/137
6,289,237	B1	*	9/2001	Mickle et al.	600/509
6,369,759	B1	*	4/2002	Epp et al.	343/700 MS
6,426,917	B1	*	7/2002	Tabanou et al.	367/82
6,597,465	B1	*	7/2003	Jarchow et al.	358/1.12
6,609,013	B1	*	8/2003	Oh et al.	455/561

OTHER PUBLICATIONS

Noreen et al., "The system view," Abstract, *The Deep Space Network: A Radio Communication Instrumentation for Deep Space Exploration,* p. 25, 1983. (Deep Space Network).

Naumenko et al., "Trajectories of relativistic particles moving in free space," Abstract, *Radiotekhnika I Electronika,* vol. 29, No. 9, pp. 144–145, 1984. ("Relativistic Particles").

Yegerov, "The efficiency of energy transmission by an electromagnetic beam and its conversion in a rectenna," translated in *Soviet Journal of Communications Technology & Electronics,* vol. 30, No. 8, pp. 35–41, 1985. ("Energy Transmission").

Bianco, "Measurement Techniques for Evaluating Power System Interference to Instrument Landing Systems (ILS)," Abstract, *Canadian Electrical Association Contract/Grant Nos. CEA 100T219 and CEA 100T219A,* 1985. ("ILS").

(Continued)

Primary Examiner—Simon Nguyen
(74) *Attorney, Agent, or Firm*—Kilpatrick Stockton LLP

(57) **ABSTRACT**

The present invention relates to a system and method for harvesting ambient electromagnetic energy, and more particularly, to the integration of antennas and electronics for harvesting ubiquitous radio frequency (RF) energy, transforming such electromagnetic energy into electrical power, and storing such power for usage with a wide range of electrical/electronic circuits and modules.

3 Claims, 13 Drawing Sheets

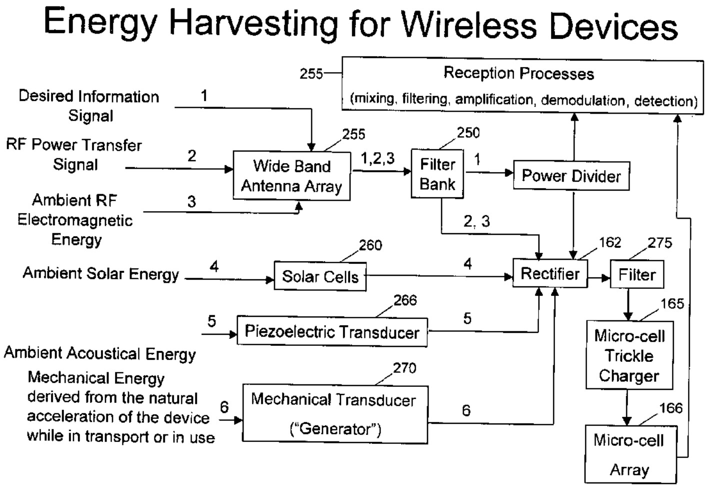
Energy Harvesting for Wireless Devices
255
Reception Processes
(mixing, filtering, amplification, demodulation, detection)
Desired Information Signal
1
RF Power Transfer Signal
2
255
Wide Band Antenna Array
1,2,3
250
Filter Bank
1
Power Divider
Ambient RF Electromagnetic Energy
3
2, 3
Ambient Solar Energy
4
260
Solar Cells
4
162
Rectifier
275
Filter
266
5
Piezoelectric Transducer
5
165
Micro-cell Trickle Charger
Ambient Acoustical Energy
Mechanical Energy derived from the natural acceleration of the device while in transport or in use
6
270
Mechanical Transducer ("Generator")
6
166
Micro-cell Array

OTHER PUBLICATIONS

Machina et al., "Power beaming—Energy transmission at 35 GHz and higher frequencies (for terrestrial and space applications with rectennas)," Abstract, *International Communication Satellite Systems Conference and Exhibit, 14th, Technical Papers, Part 3*, pp. 1671–1674, 1992. (Power Beaming).

Waldron, "Rectenna options for a lunar–solar power system," Abstract, 26th *International Conference on Environmental Systems,* 1996. ("Rectenna Options").

Keerthi, "Analysis of the multistage CM array for digital communication signals," Abstract, *Conference Record of the Thirtieth Asilomar Conference on Signals, Systems and Computers,* vol. 2, No. 23, p. 1416, 1997. ("Asilomar").

Takeshi et al., "Experimental study of microwave power transmission efficiency," Abstract, *IEIC Technical Report, Institute of Electronics, Information and Communication Engineers,* vol. 98, No. 474, pp. 21–24, 1998. ("IEIC").

Vasudev, "Telecom explosion drives RF ICs," *Electronic Engineering Times, No. 1016*, p. 81, 1998. ("RF ICs").

Takyuki et al., "Effective utilization of electromagnetic energy. Effective utilization of electromagnetic energy in ITS (4). Radiocommunication in turnpike automatic charge collection system," Abstract, *Denji Kankyo Kogaku Joho EMC,* vol. 11, No. 9, pp. 38–45, 1999. ("ITS").

Ohr, "Integrated, one–chip cell phone is still a stretch, panelists concur," *Electronic Engineering Times,* No. 1067, p. 64, 1999. ("One–chip cell phone").

Davies, "Operation of Information Satellites in an interference environment," Abstract, *Microwave Journal,* vol. 15, pp. 33–36, 1972. ("Microwave Journal").

Dickinson, "Satellite Power System (SPS) Microwave Subsystem Impacts and Benefits," Abstract, *Report No. NASA–CR–157951,* 1977. ("Dickinson SPS").

Tonelli et al., "Design and analysis of a 5000–MW GaA1As satellite power system," Abstract, *Proceedings of the 12th Intersociety Energy Conversion Engineering Conference Part II,* pp. 1412–1420, 1977. ("IECE 1").

Tonelli et al., "Design and evaluation of a 5 GW GaA1As solar power satellite (SPS)," Abstract, *Proceedings of the 13th Intersociety Energy Conversion Engineering Conference Part I,* pp. 156–161, 1978. ("IECE 2").

Hughes, "The interactions of a solar power satellite transmission with the ionosphere and troposphere," Abstract, *AGARD Conference Proceedings No. 284. Propagation Effects in Space/Earth Paths,* p. 38, 1980. ("AGARD").

Maynard, "Initial MPTS study results: Design considerations and issues," Abstract, *Johnson Space Center Solar Power Satellite Microwave Transmission and Reception,* pp. 28–32, 1980. ("MPTS").

* cited by examiner

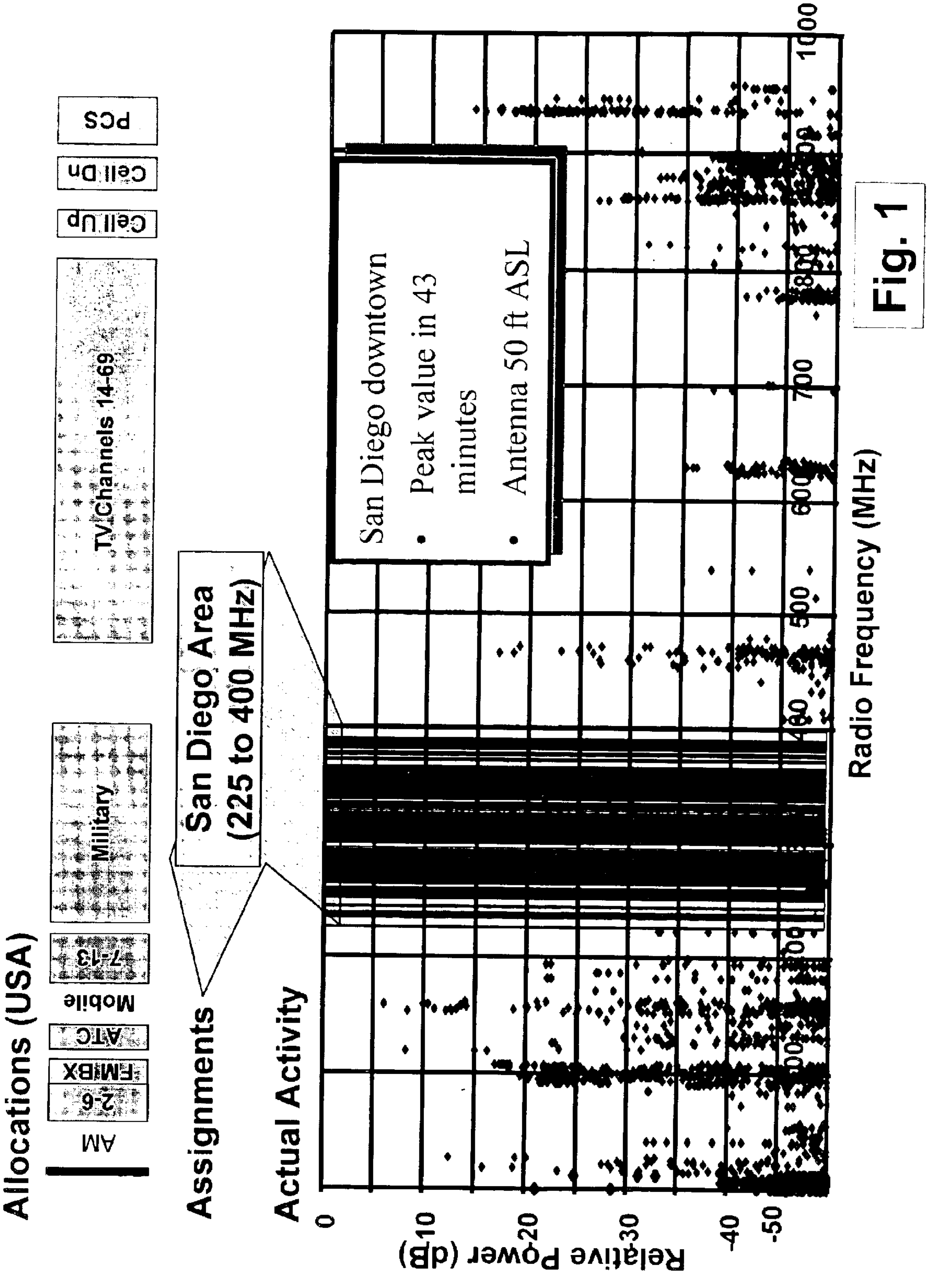
Allocations (USA)
AM
2-6
FM/BX
ATC
Mobile
7-13
Military
TV Channels 14-69
Cell Up
Cell Dn
PCS
Assignments
San Diego Area
(225 to 400 MHz)
Actual Activity
San Diego downtown
• Peak value in 43 minutes
• Antenna 50 ft ASL
0
-10
-20
-30
-40
-50
Relative Power (dB)
100
200
300
400
500
600
700
800
900
1000
Radio Frequency (MHz)

Fig. 1

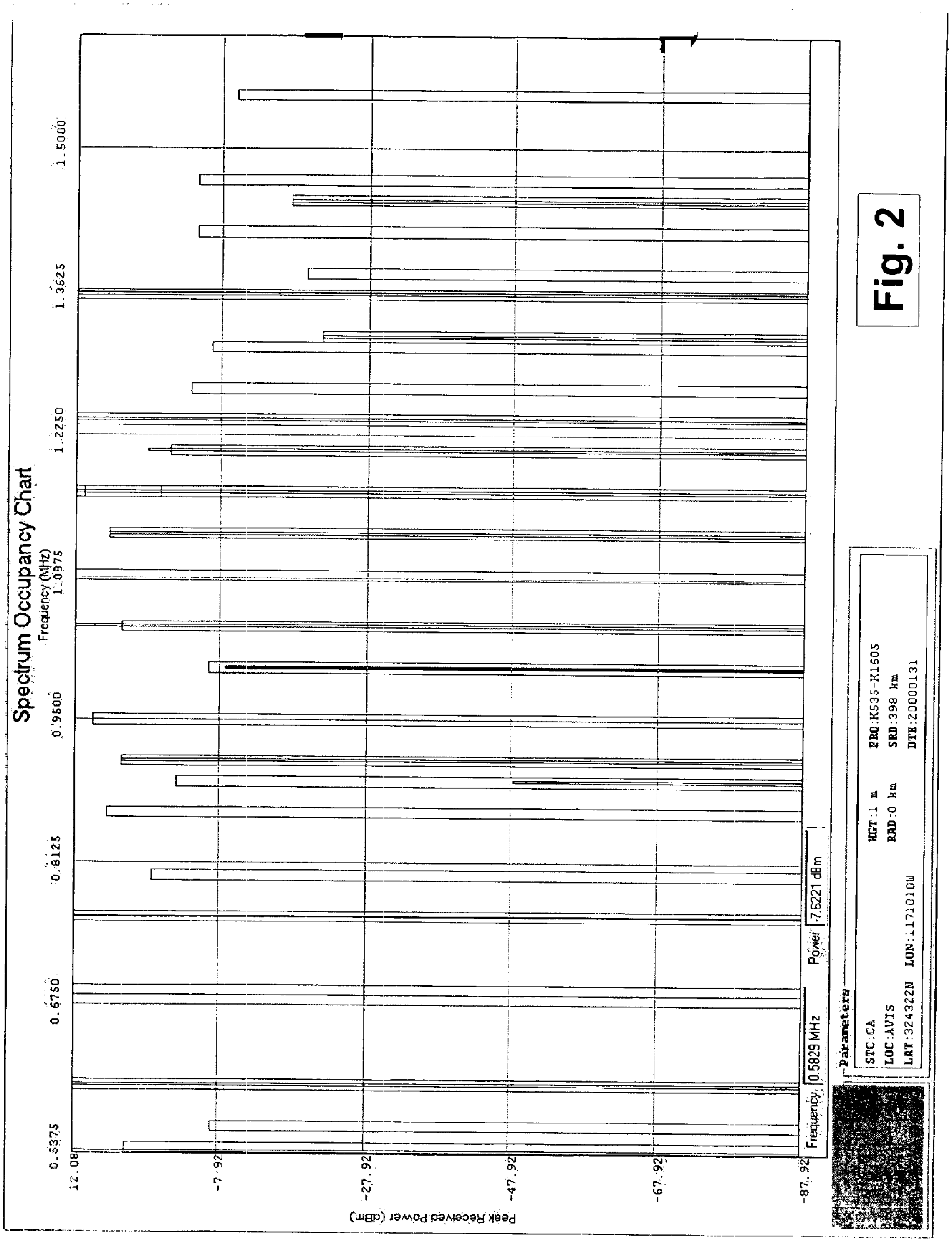
Spectrum Occupancy Chart
Frequency (MHz)
0.5375
0.6750
0.8125
0.9500
1.0875
1.2250
1.3625
1.5000
Peak Received Power (dBm)
12.08
-7.92
-27.92
-47.92
-67.92
-87.92
Frequency
0.5829 MHz
Power
-7.6221 dBm
Parameters
STC:CA
LOC:AVIS
LAT:324322N
HGT:1 m
RAD:0 km
LON:1171010W
FRQ:K535-K1605
SRD:398 km
DTE:Z0000131

Fig. 2

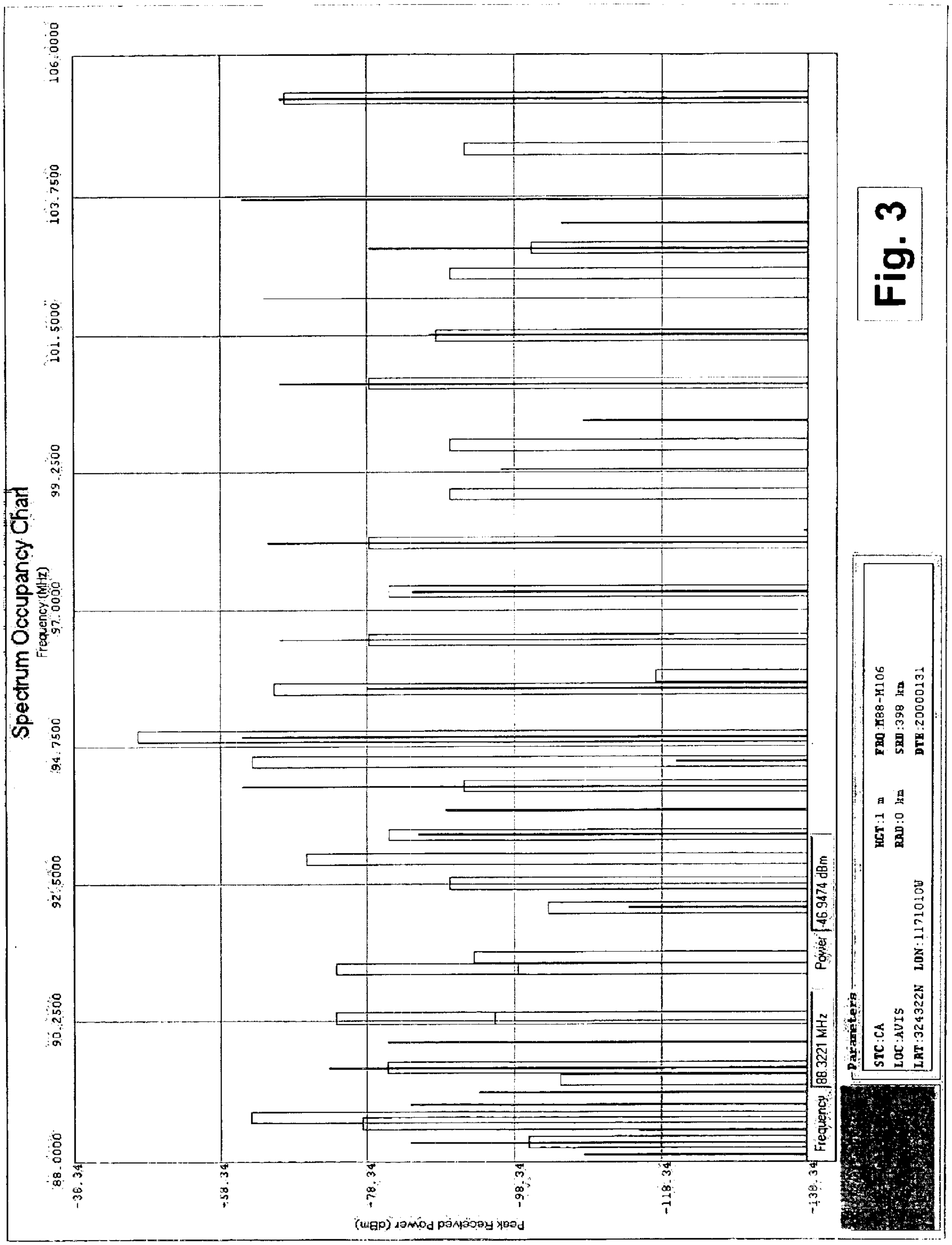
Spectrum Occupancy Chart
Frequency (MHz)
88.0000
90.2500
92.5000
94.7500
97.0000
99.2500
101.5000
103.7500
106.0000
Peak Received Power (dBm)
-38.34
-58.34
-78.34
-98.34
-118.34
-138.34
Frequency
88.3221 MHz
Power
46.9474 dBm
Parameters
STC:CA
LOC:AVIS
LAT:324322N
HGT:1 m
RAD:0 km
LON:1171010W
FRQ:M88-M106
SRD:398 km
DTE:20000131

Fig. 3

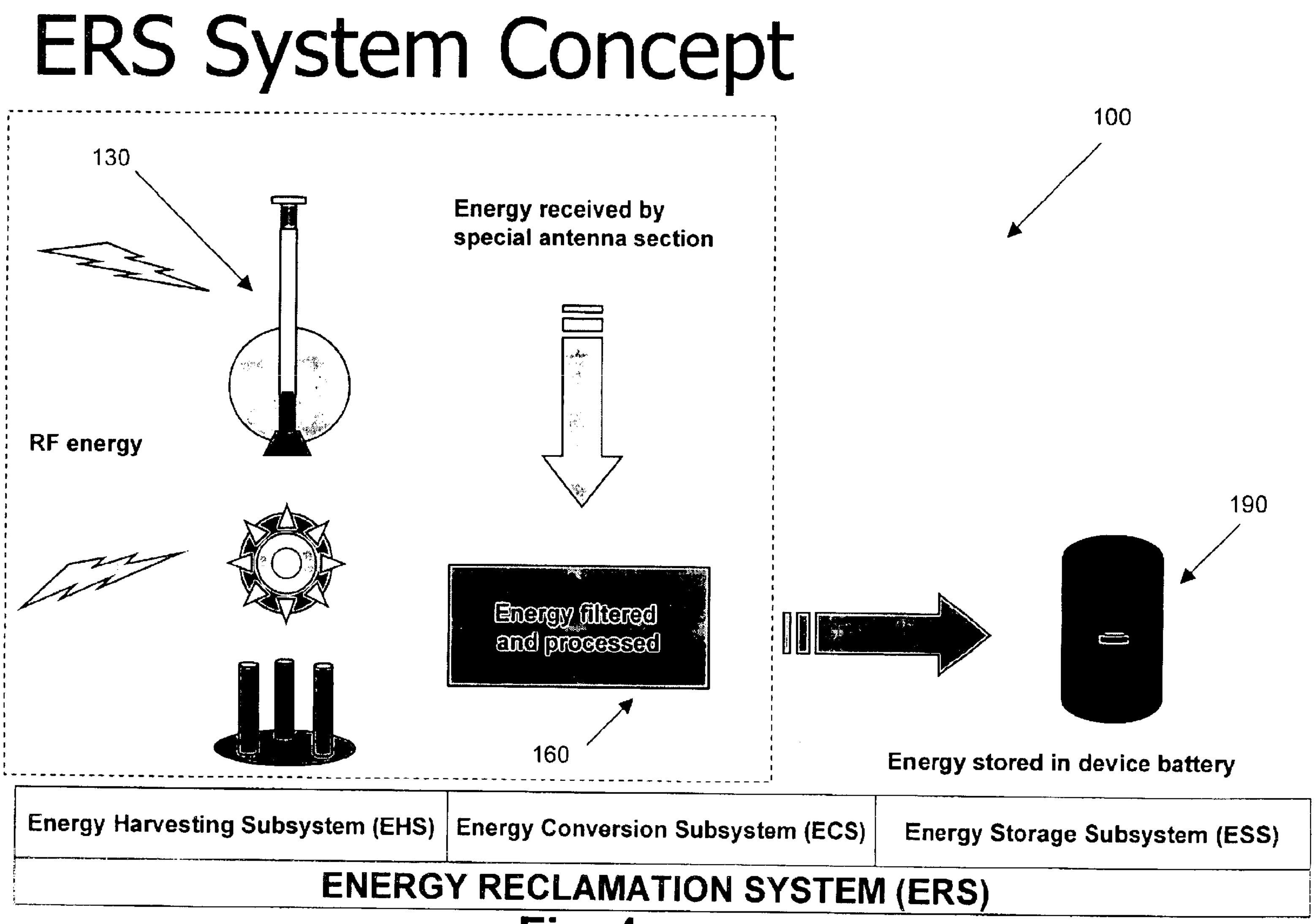
ERS System Concept
100
130
RF energy
Energy received by special antenna section
Energy filtered and processed
160
190
Energy stored in device battery
Energy Harvesting Subsystem (EHS)
Energy Conversion Subsystem (ECS)
Energy Storage Subsystem (ESS)
ENERGY RECLAMATION SYSTEM (ERS)

Fig. 4

Antenna Design

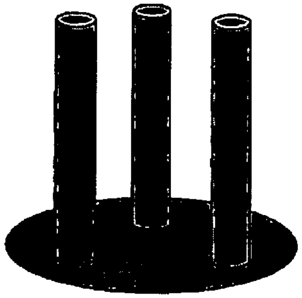

Fig. 5A

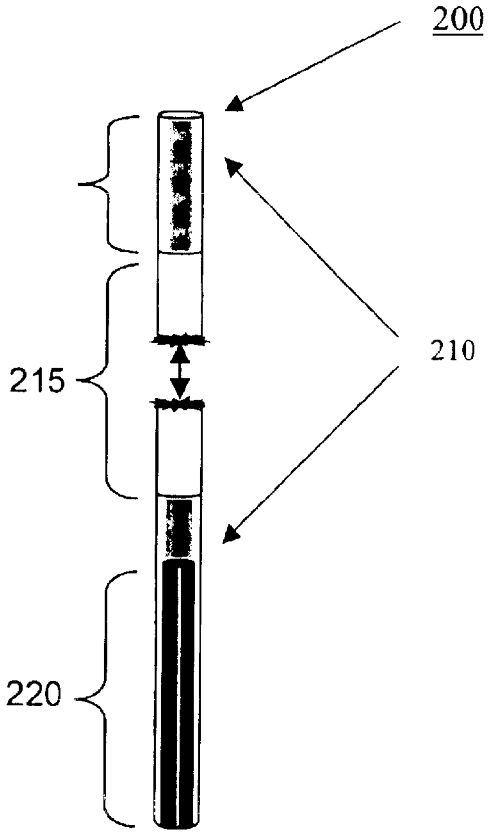
200
210
215
220

Fig. 5B

Alternate Antenna Design

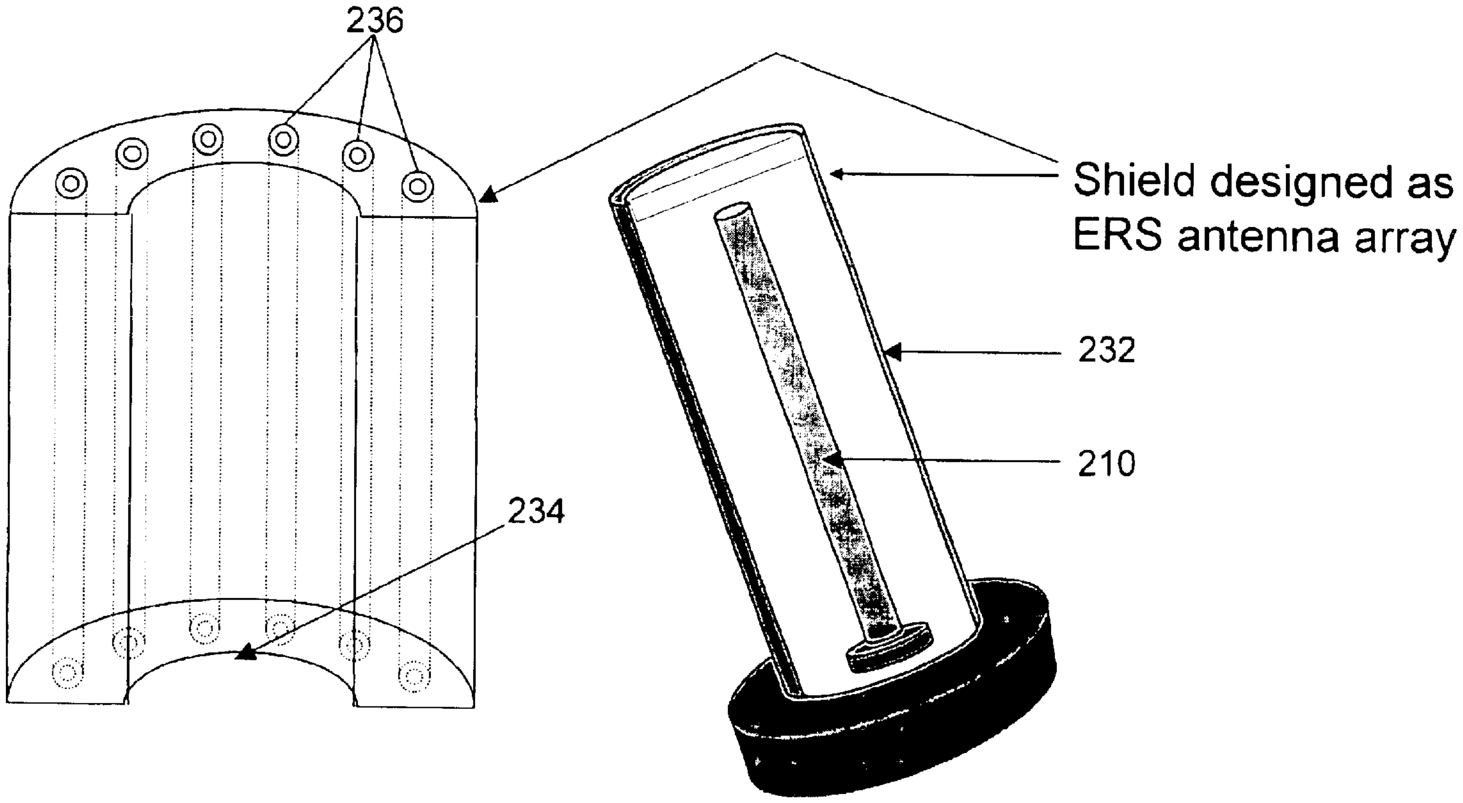
236
Shield designed as ERS antenna array
232
210
234

Fig. 6

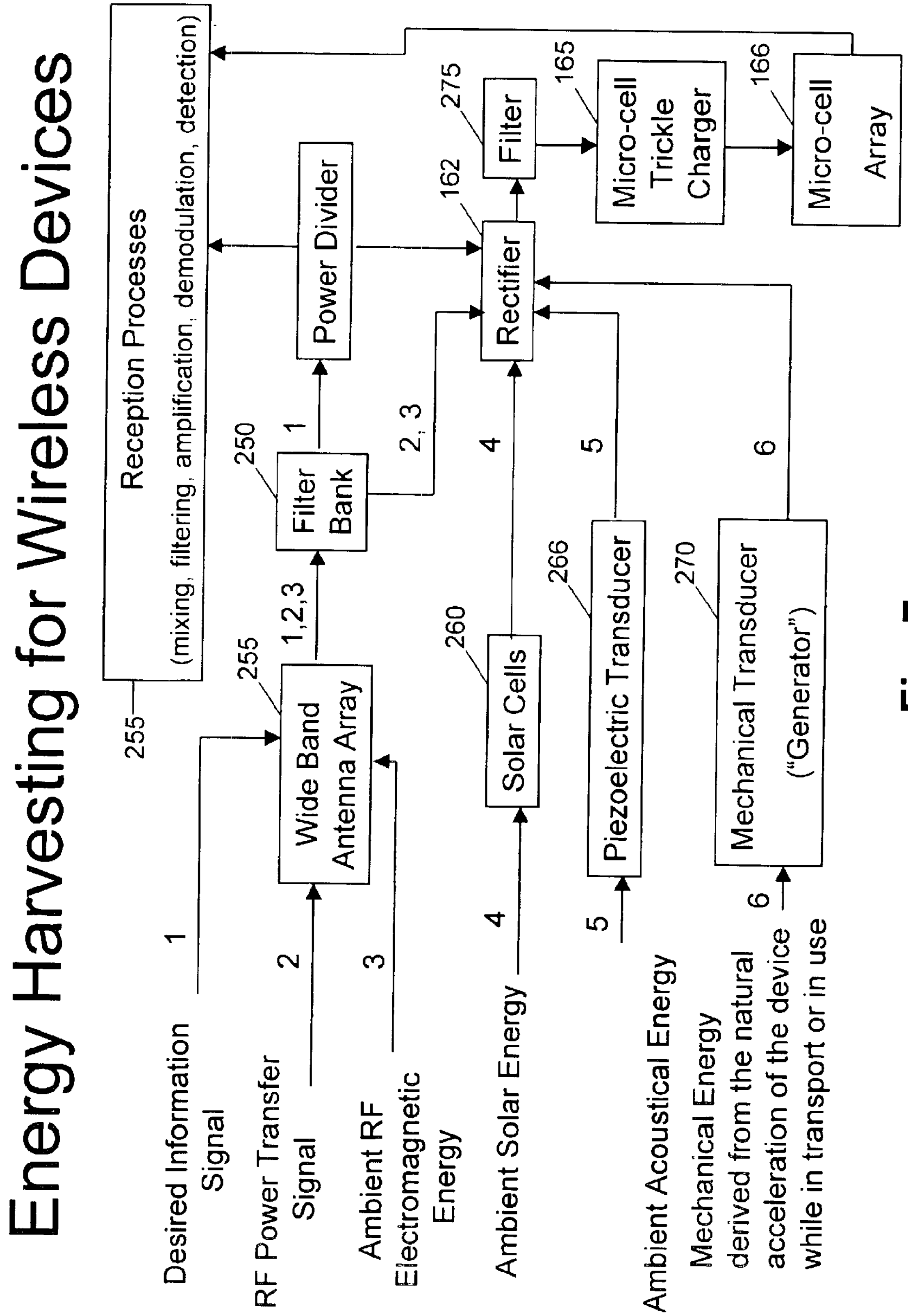
Energy Harvesting for Wireless Devices
Desired Information Signal
RF Power Transfer Signal
Ambient RF Electromagnetic Energy
Ambient Solar Energy
Ambient Acoustical Energy
Mechanical Energy derived from the natural acceleration of the device while in transport or in use
1
2
3
4
5
6
255
Reception Processes (mixing, filtering, amplification, demodulation, detection)
255
Wide Band Antenna Array
1,2,3
250
Filter Bank
1
2, 3
Power Divider
260
Solar Cells
4
266
Piezoelectric Transducer
5
270
Mechanical Transducer ("Generator")
6
162
Rectifier
275
Filter
165
Micro-cell Trickle Charger
166
Micro-cell Array

Fig. 7

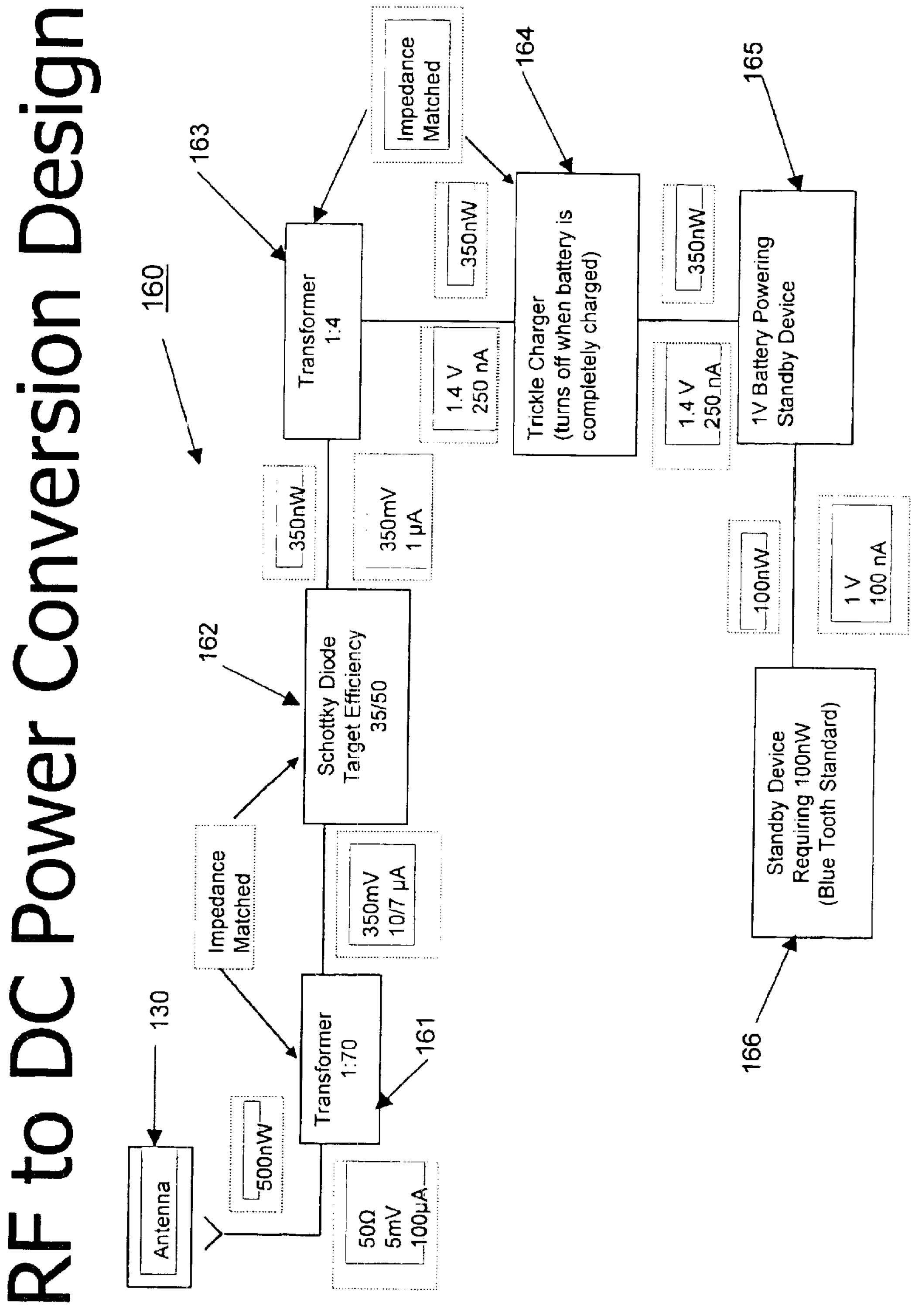
RF to DC Power Conversion Design
160
Antenna
130
500nW
50Ω 5mV 100µA
Transformer 1:70
161
Impedance Matched
350mV 10/7 µA
Schottky Diode Target Efficiency 35/50
162
350nW
350mV 1 µA
Transformer 1:4
163
Impedance Matched
350nW
1.4 V 250 nA
Trickle Charger (turns off when battery is completely charged)
164
350nW
1.4 V 250 nA
1V Battery Powering Standby Device
165
100nW
1 V 100 nA
Standby Device Requiring 100nW (Blue Tooth Standard)
166

Fig. 8

RF to DC Power Conversion/Trickle Charge/Power Storage Design Concept

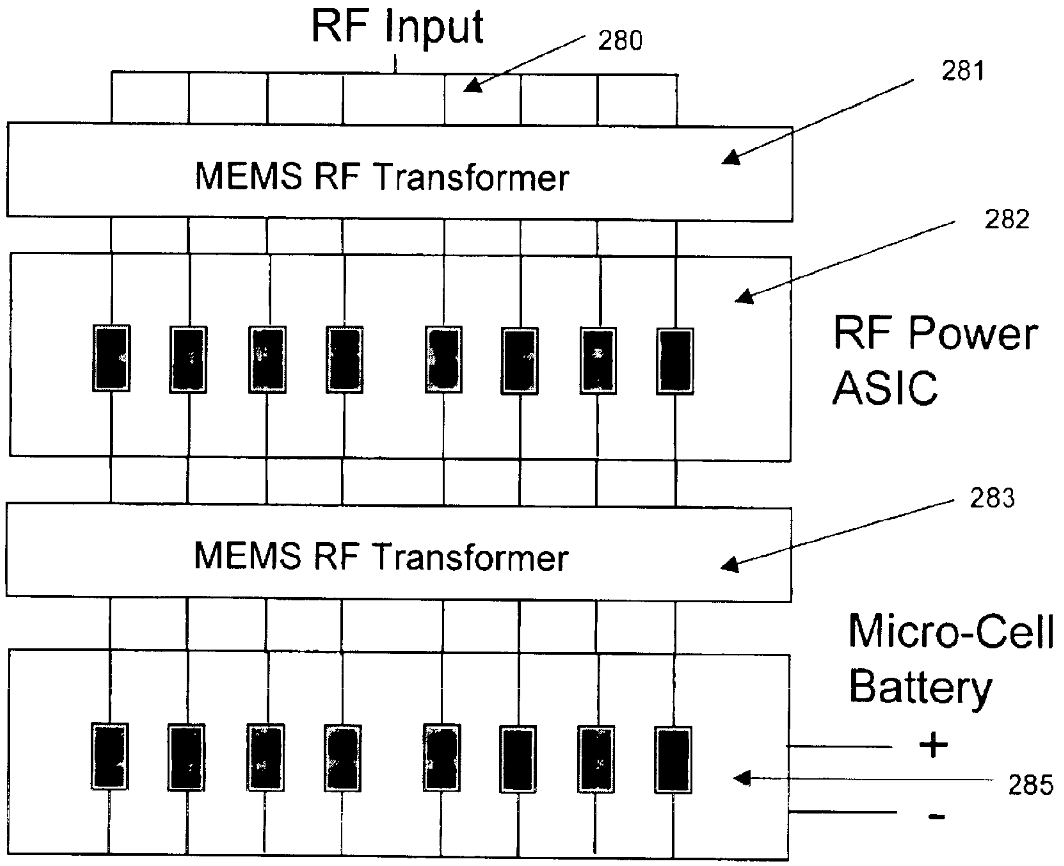
RF Input
280
281
MEMS RF Transformer
282
RF Power
ASIC
283
MEMS RF Transformer
Micro-Cell
Battery
+
285
-

Fig. 9

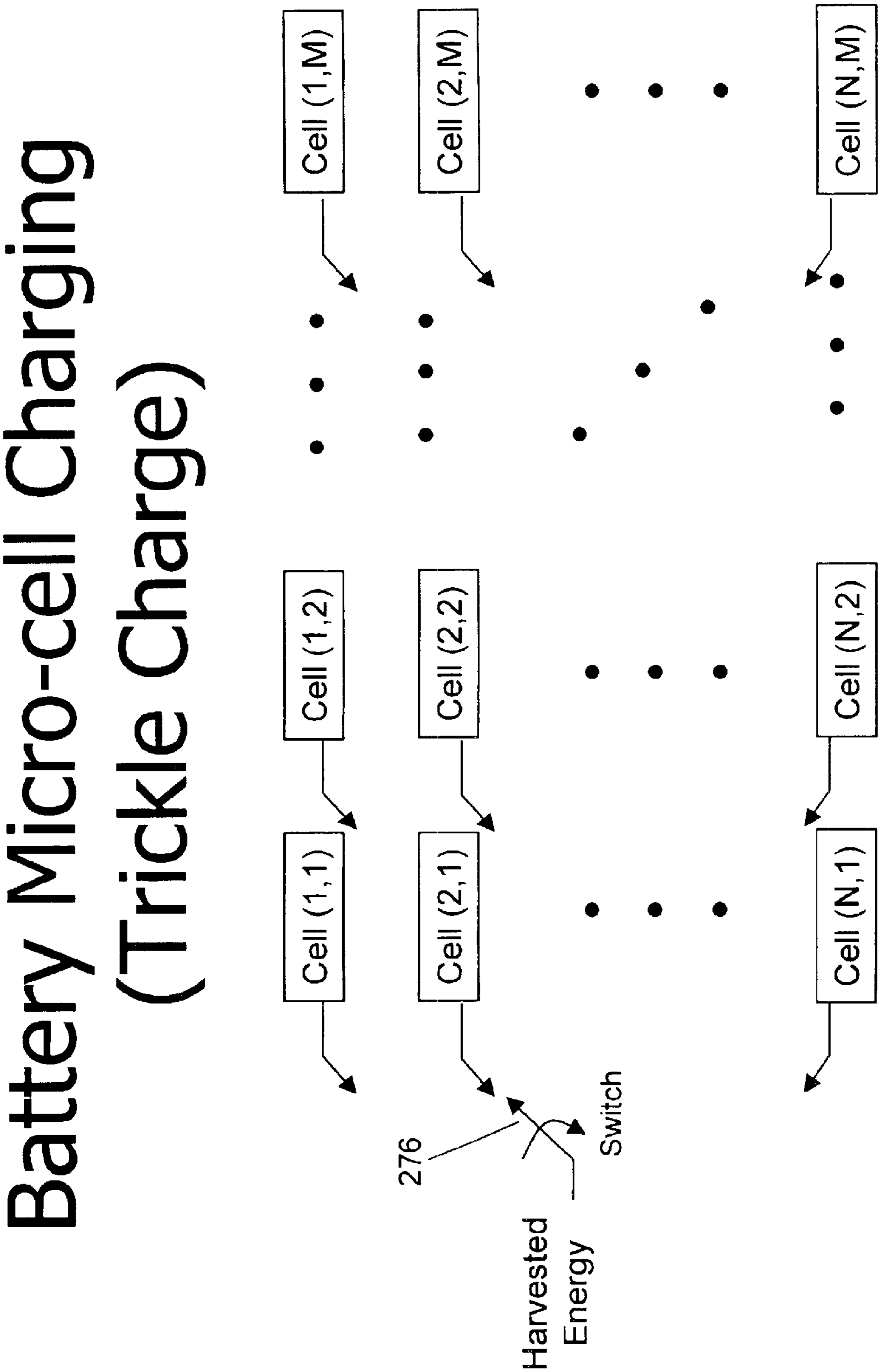
Battery Micro-cell Charging
(Trickle Charge)
Cell (1,1)
Cell (1,2)
Cell (1,M)
Cell (2,1)
Cell (2,2)
Cell (2,M)
Cell (N,1)
Cell (N,2)
Cell (N,M)
276
Switch
Harvested Energy

Fig. 10

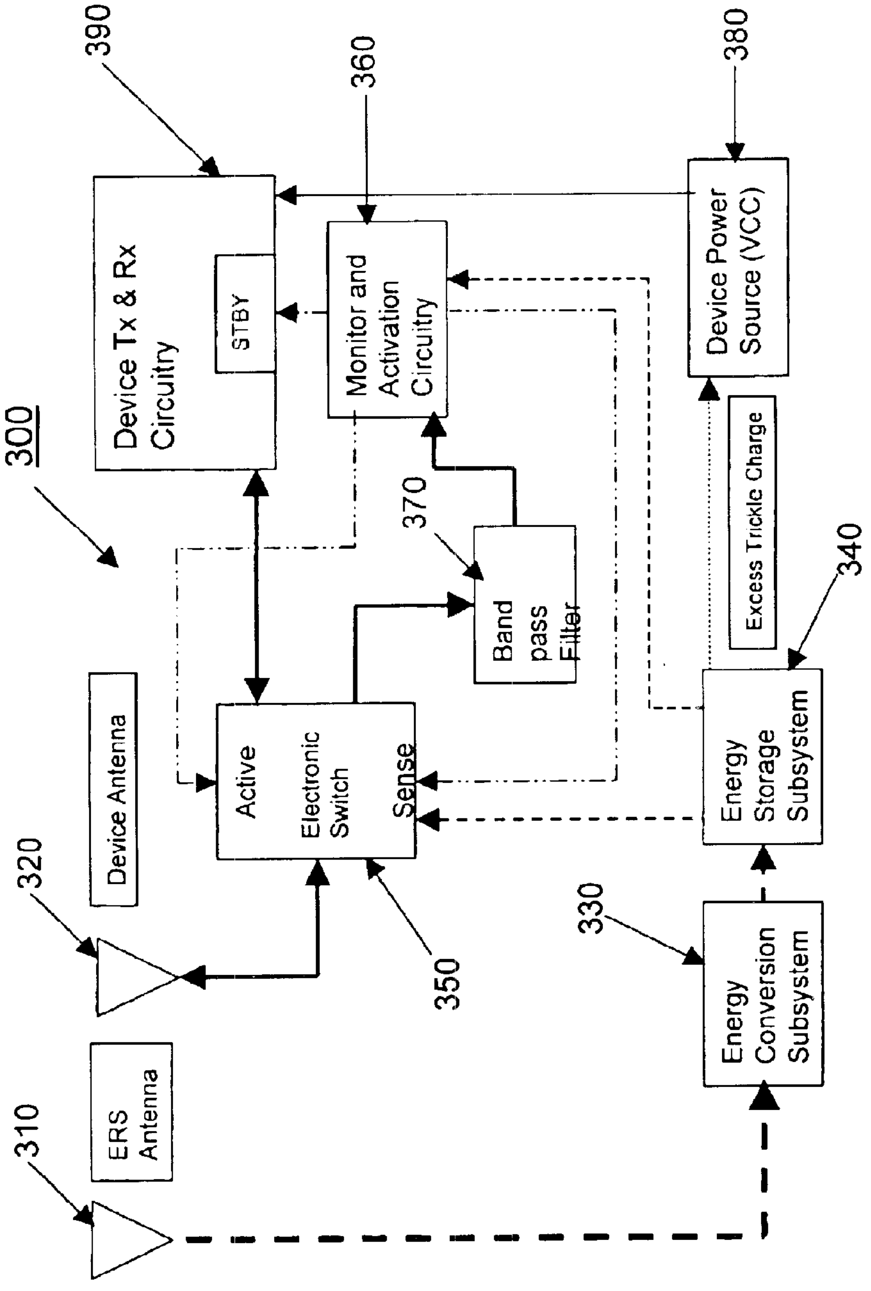
300
390
Device Tx & Rx Circuitry
STBY
360
Monitor and Activation Circuitry
380
Device Power Source (VCC)
Excess Trickle Charge
370
Band pass Filter
Active
Electronic Switch
Sense
350
340
Energy Storage Subsystem
330
Energy Conversion Subsystem
320
Device Antenna
310
ERS Antenna

Two or Multiple Array Antenna Sense and Active Mode

Fig. 11

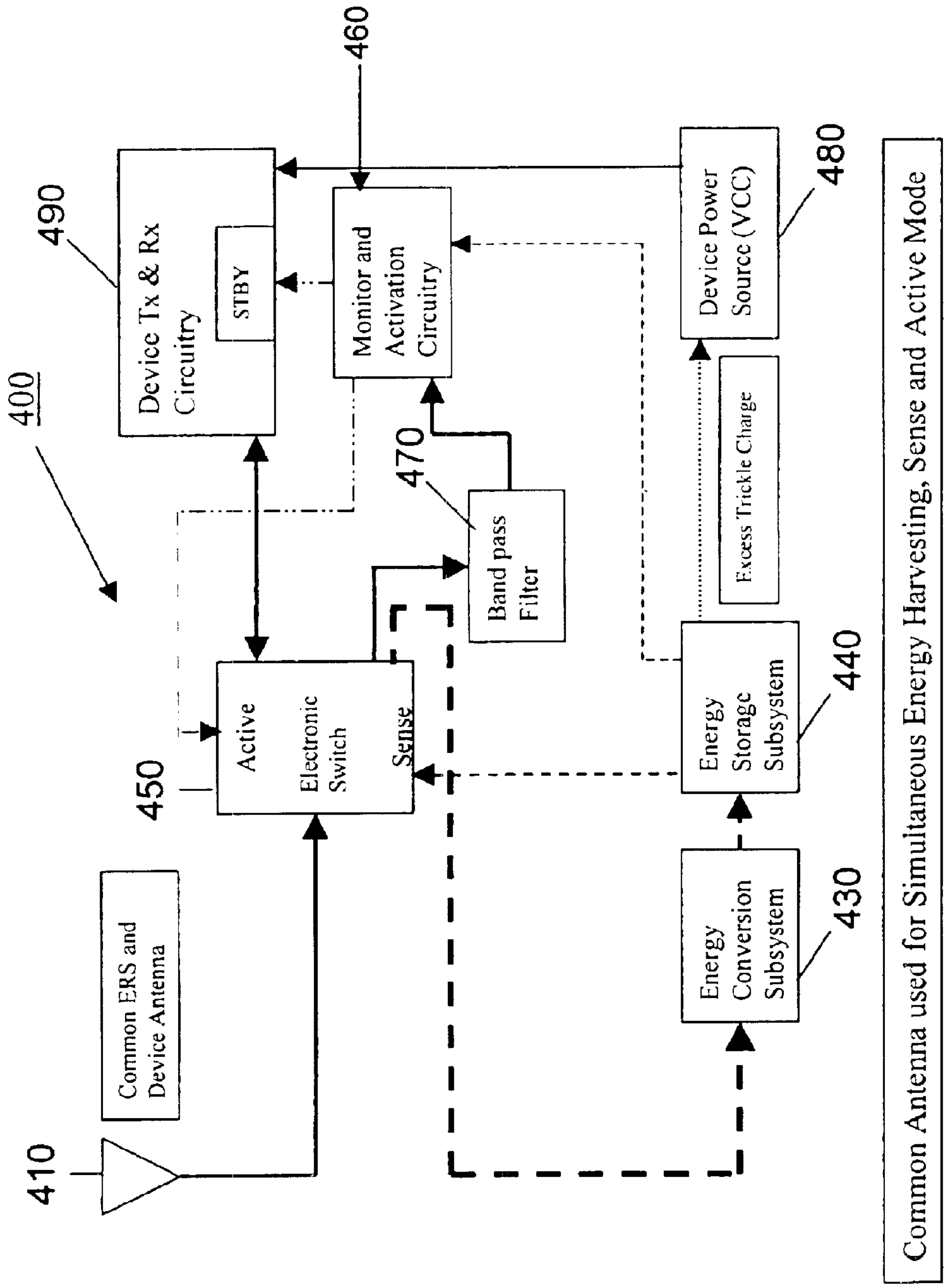
400
490
Device Tx & Rx Circuitry
STBY
460
Monitor and Activation Circuitry
470
Band pass Filter
480
Device Power Source (VCC)
Excess Trickle Charge
450
Active Electronic Switch
Sense
440
Energy Storage Subsystem
430
Energy Conversion Subsystem
410
Common ERS and Device Antenna
Common Antenna used for Simultaneous Energy Harvesting, Sense and Active Mode

Fig. 12

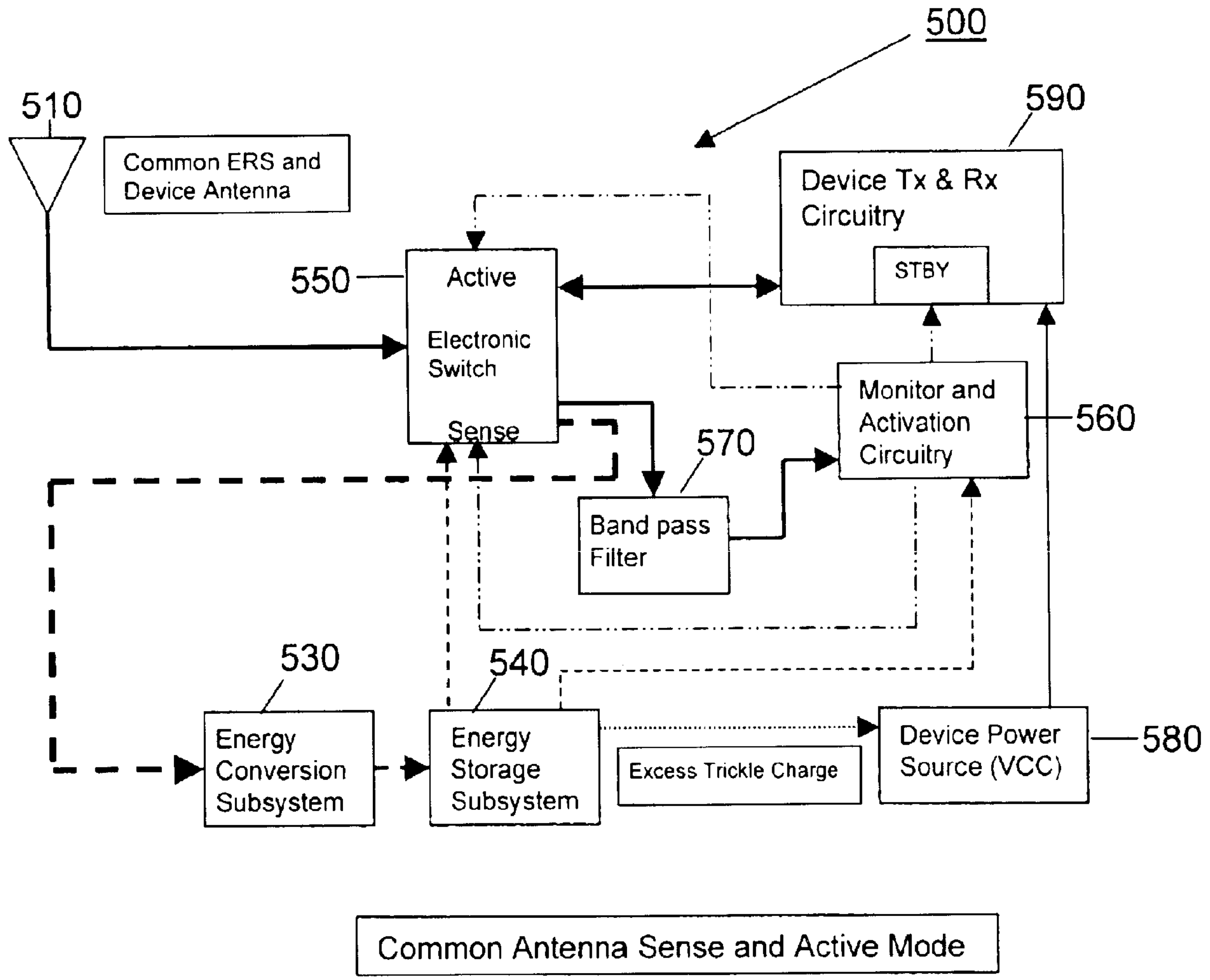
500
510
Common ERS and Device Antenna
590
Device Tx & Rx Circuitry
STBY
550
Active
Electronic Switch
Sense
570
Band pass Filter
Monitor and Activation Circuitry
560
530
Energy Conversion Subsystem
540
Energy Storage Subsystem
Excess Trickle Charge
Device Power Source (VCC)
580
Common Antenna Sense and Active Mode

Fig. 13

METHOD AND SYSTEM FOR ENERGY RECLAMATION AND REUSE

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates generally to wireless energy harvesting methods and systems. Specifically, the preferred embodiments of the present invention relate to the integration of antennas and electronics for harvesting ubiquitous electromagnetic energy, transforming that energy into useful power, and storing such power for usage.

2. Description of the Related Art

Typically, energy harvesting techniques and systems are focused on renewable energy such as solar energy, wind energy, and wave action energy from moving water. Solar energy is harvested by arrays of solar cells that convert radiant energy to DC power. Such energy is limited in low-light conditions. Cumbersome wind turbines that convert wind energy to electrical power depend on sufficient wind to spin the blades of the turbines to generate power. Wave action energy is harvested by water turbines or other apparatuses that are capable of converting the energy in the moving water into electrical power. This energy conversion is typically performed at complex hydroelectric power plants or by mechanical/electrical conversion apparatuses, such as those shown in U.S. Pat. Nos. 5,842,838 and 6,045,339 to Berg.

Another approach to energy harvesting is wireless power transmission via focused microwave signals. Unlike the above renewable energy technologies which harvest energy from natural sources, electrical energy of the wireless power transmission technology is remotely harvested and converted from focused microwave signals that are transmitted by a constructed source. For example, U.S. Pat. No. 5,043,739 to Logan et al. discloses a high frequency rectenna device for rectifying electromagnetic energy at microwave frequencies and higher and converting the energy into direct-current (DC) power to conduct wireless power transmission. Applications of this disclosure are focused on high-energy microwave radiation and high temperatures. One such application includes sending power to satellites or high altitude devices from the earth by electromagnetic waves. Another application involves collecting solar power from large space-based arrays of solar cells and transmitting the energy via electromagnetic waves to earth. A third application involves the direct conversion of microwaves generated by fusion reactions into a direct current.

BRIEF SUMMARY OF THE INVENTION

Currently, there is a large amount of renewable electromagnetic energy available in the radio frequency (RF) bands. For instance, AM radio, FM radio, TV, very high frequency (VHF), ultra high frequency (UHF), global system for mobile communications (GSM), digital cellular systems (DCS) and especially the personal communication system (PCS) bands are all lucrative energy sources. These untapped sources of energy grow more and more abundant as a result of the rapid growth in the wireless communication business. Hence, there is a desire to seek new methods and means of collecting the ubiquitous electromagnetic energy (ambient RF noise and signals) opportunistically present in the environment and transforming that energy into useful electrical power.

The preferred embodiments of the present invention can collect ubiquitous ambient radio frequency (RF) electro magnetic energy, convert it to a form useful for powering electronic devices, and store it for reuse. Thus, perpetual self contained subsystems or modules can be realized by using the reclaimed RF energy to power active antenna components, low power circuits, modules and other components such as transceivers, amplifiers, filters, etc. Hence, the preferred embodiments of the present invention can be applicable for any antenna or antenna array, for any antenna physical design (e.g., dipole, monopole, parabolic shape or dish, Yagi, etc.) and any frequency band. In some applications, this technology could be used advantageously and conveniently to transfer power remotely, without a physical connection, using a RF energy source intended for that purpose. In other applications, this technology could be used to maintain devices in a standby sleep mode indefinitely without ancillary power source. Then an RF signal could be used as a triggering mechanism to activate the device. Finally, for low power devices, this technology could provide primary power to the device and eliminate or reduce the traditional battery charge cycle.

Accordingly, the preferred embodiments of the present invention provide an integration of classical antenna physical materials and construction concepts with semiconductor and microprocessor technology to create an antenna structure capable of transforming electromagnetic waves into DC power. This antenna method and system collects and processes RF energy for reuse while simultaneously performing its primary function, such as signal reception, for a specific application. Furthermore, by using either a single antenna element or an array of elements configured to receive a broad range of frequencies in the present invention, energy collection and antenna transmit/receive performance can be maximized and optimized at the same time. The preferred embodiments of the present invention also provide an Energy Reclamation System (ERS) that collects, converts and stores energy for use from selected frequency bands over the entire electromagnetic spectrum. These selected frequency bands are not required to be contiguous and may based on the operational environment, the application, any physical size limitations, and the technology available to cost effectively integrate an ERS into a device or application. An ERS of the present invention is not restricted to a single energy source because technologies like micro electromechanical systems (MEMS) can be used to integrate the collection of other types of energy, such as electromagnetic energy, solar energy, acoustic energy and/or casual motion energy. Thus, energy diversity enhances an ERS' ability to provide power consistent with the demands of modern devices and applications.

The preferred embodiments of the present invention also provide a method and system for harvesting electromagnetic spectrum energy from unintentional power sources. Thus, the electromagnetic energy may be harvested opportunistically from the abundant ambient RF electromagnetic environment created by the emission of countless commercial RF sources typically present today.

The preferred embodiments of the present invention also provide a method and system for harvesting electromagnetic spectrum energy using efficient wideband omnidirectional antennas for maximizing the collection of RF energy. The antennas can be of any design such as monopole, dipole, Yagi, or contrawound.

The preferred embodiments of the present invention also provide an energy collection and conversion technology that may be used to transmit a control signal to select a device's operating mode (e.g., wake up call in the standby sleep mode). Incorporating this technology into some devices allows them to remain in the sleep mode indefinitely. Thus, the devices may use the continuously reclaimed energy to power the monitoring circuitry and initialize equipment activation as required.

The preferred embodiments of the present invention also provide a method and system for transferring low power levels from an intentional power source advantageously and conveniently. This may be done remotely without a physical connection by using a RF energy source with a focused or blanket radiation pattern and an ERS at the target device. This may be used for numerous applications, including autonomous sensor or communication type applications in any environment (e.g., ground, air, underwater) where the device may use the ERS as its power source or backup power source. For example, for underwater sensor locations, energy may come from sonar or other types of energy sources that can propagate with tolerable losses through the water and/or from RF energy delivered to a remote antenna extended above the surface of the water. If remaining operational for extended periods of time requires additional power, the device can be wirelessly and remotely supplied with power for storage and future use.

The preferred embodiments of the present invention further provide a conversion of the energy collected or captured by the antenna or transducer into an electrical current. The electrical current is then converted into a form suitable for input to a selected storage device. For instance, a rectifier and a low pass filter can be used to convert the electrical current into direct current (DC). This power can then be fed into a trickle charger for charging a power storage device, such as batteries, in wireless and other electrical devices. The type of electrical filter and power conditioning used depends on the type of battery or power storage device and the particular process of trickle charging used.

The preferred embodiments of the present invention additionally provide a method and system for harvesting electromagnetic energy that is designed to efficiently collect and convert RF energy using application specific integrated circuit (ASIC) technology. Hence, supporting electronics are integrated with the antenna itself and include, for example, amplifiers, mini-transceivers, trickle chargers for charging micro-cell batteries, filters efficient rectifier design and implementation (with different types of diodes), micro electromechanical systems (MEMS) RF transformers, etc. Thus, embedded circuitries are built into the antenna elements and optimized for a frequency range compatible with any given antenna design.

Additional aspects and novel features of the invention will be set forth in part in the description that follows, and in part will become more apparent to those skilled in the art upon examination of the disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

The preferred embodiments are illustrated by way of example and not limited in the following figures, in which:

FIG. **1** shows data taken from measurements in the RF spectrum between 3 MHz and 1 GHz in downtown San Diego, Calif., USA;

FIG. **2** shows the results of calculations predicting the signal levels in the AM radio band in downtown San Diego, Calif., USA;

FIG. **3** shows the results of calculations predicting the signal levels in the FM radio band in downtown San Diego, Calif., USA;

FIG. **4** shows the energy reclamation system (ERS) concept in accordance to one embodiment of the present invention;

FIGS. **5**A and **5**B show a physical structure of an energy harvesting antenna and an array of such antennas in accordance to an embodiment of the present invention;

FIG. **6** shows a communication antenna having an insulated shield acting as both a safety shield to the user and an energy harvesting antenna array in accordance to an embodiment of the present invention;

FIG. **7** shows a high level schematic diagram of an Energy Harvesting Subsystem (EHS) in accordance to one embodiment of the present invention;

FIG. **8** shows an energy conversion subsystem (ECS) of a perpetual active antenna (PAA) system in accordance to one embodiment of the present invention;

FIG. **9** shows an application specific integrated circuit (ASIC) chip implementation of the ECS circuitry shown in FIG. **8**;

FIG. **10** shows an array of battery micro-cells of an energy storage subsystem in accordance to an embodiment of the present invention;

FIG. **11** shows a high level schematic diagram of the Two or Multiple Array Sense and Active Mode in accordance to an embodiment of the present invention;

FIG. **12** shows a high level schematic diagram of the Common Antenna Sense and Active Mode in accordance to an embodiment of the present invention;

FIG. **13** shows a high level schematic diagram of the Common Antenna for Simultaneous Energy Harvesting, Sense and Active Mode in accordance to an embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The preferred embodiments of the present invention represent a new technology area for the reclamation and reuse of Radio Frequency (RF) power present due to either design or opportunity in certain frequency bands. It was found that RF signals and noise are available in useful power levels in many frequency bands-particularly in the AM, FM, and TV bands as well as the global system for mobile communications (GSM), digital cellular systems (DCS) and communication system (PCS) bands. The abundance of energy available for harvesting is clearly apparent from the measurements of the ambient RF spectrum. For example, FIG. **1** shows one such measurement in downtown San Diego Calif. The measurement indicates the relative power levels of RF signals between 3 MHz and 1 GHz that can be harvested at this location. As can be seen from the figure, there are ample RF signals from which to reclaim energy. FIG. **2** shows the results of calculations predicting the signal levels in just the AM radio band in downtown San Diego. It shows that many of the AM signals approach or are greater than 1 mW (0 dBm). This is because many commercial AM radio transmitters radiate thousands of watts of power. FIG. **3** shows the results of calculations predicting the signal levels in just the FM radio band in downtown San Diego. These calculations further indicate that there are ample opportunities for harvesting energy from the ambient electromagnetic environment.

Countless candidate devices can benefit from this new energy reclamation technology of the present invention. For instance, the technology can be used for wireless devices that are typically powered by exhaustible or rechargeable batteries. The present invention can be used to provide wireless devices located in certain environments a perpetual life in standby mode and extended normal operational life by using stored reclaimed energy as a power source. The Energy Reclamation System or ERS of the present invention can provide sufficient power (e.g., 0.1–1 mW) to perpetually operate low power wireless devices without external recharging. The ERS of the present invention can also able to provide higher power levels to meet higher demands placed on these wireless devices. According to one preferred embodiment of the present invention, the ERS of the present invention includes a common antenna designed for the energy abundant frequency bands with it's output connected to power conditioning or rectification circuitry appropriately designed for that frequency range. The conditioned or rectified power can then be supplied to a charger (e.g., a trickle charger) or distribution subsystem designed to meet the battery's or storage device's charging specification.

There are innumerable potential applications for the energy reclamation technology of the present invention. For instance, the power harvested from radio frequency (RF) signals can be used for recharging installed power sources, providing power directly to the devices, energizing a device to wake up, and providing emergency backup power for the devices. Countless potential users of this technology also exist. For instance, potential users of the ERS technology may be found in the cellular, personal communications services (PCS) and future generations of wireless communications device industry. Typical consumer wireless electronic products that can be enhanced by ERS technology include, but are not limited to, palm computers, personal digital assistants (PDAs), pocket phones, pagers, cameras, autonomous sensors, garage door openers, remote controls and auto security systems. The energy reclamation technology can also be used in emergency location systems such as GPS position finders, airplane and watercraft emergency beacons and fire alarm systems. The energy reclamation technology can also be applied to wireless personal devices such as video conferencing cell phones, Internet phones and Wireless Personal Area Networks (WPAN). The energy reclamation technology is also flexible enough to be adapted for future electrical and/or electronics applications and devices.

According to a preferred embodiment of the ERS of the present invention, the ERS includes three subsystems: an Energy Harvesting Subsystem (EHS), an Energy Conversion Subsystem (ECS), and an Energy Storage Subsystem (ESS). The Energy Harvesting Subsystem or EHS includes various components used to collect ambient electromagnetic energy via an antenna or antenna array and convert it to electrical current. If acoustic, solar, or motion is the energy source, the antenna or antenna array may be replaced by a transducer or transducer array for the respective energy source. The EHS may also include a combination of antennas and mechanical transducers to collect different types of energy at once or alternately. The Energy Conversion Subsystem or ECS includes circuits, components, and modules integrated together to convert and condition the noise like current out of the antenna or transducer into a form of energy suitable for storage. The Energy Storage Subsystem or ESS includes, for example, batteries, capacitors, and other such components that will store the reclaimed energy for future reuse. The ESS can also include appropriate circuitry (e.g. trickle charger) for charging those devices that the ERS supports.

FIG. **4** shows the ERS concept in accordance to the above preferred embodiment of the present invention. As shown, the ERS **100** includes three parts or subsystems: an Energy Harvesting Subsystem (EHS) **130**, a Power or Energy Conversion Subsystem (ECS) **160**, and a Power or Energy Storage Subsystem (ESS) **190**. The Energy Reclamation System provides for the collection and harvesting of ubiquitous ambient energy over a desired frequency range. In a preferred embodiment, the desired range is 500 KHz–2.4 GHz. For frequencies above 2.4 GHz and for intentionally-radiated high power harvesting applications, the ERS can incorporate the rectenna technology, such as that disclosed in U.S. Pat. No. 5,043,739, into the EHS and the ECS for connection to the ESS. The interconnection of the ESS to the rectenna technology is readily apparent to one skilled in the art.

According to another preferred embodiment of the present invention, the main objective of the ERS **100** is focused on the collection of ambient RF energy present in the ambient environment in the range of 5 watts and below and converting that energy into reusable energy to power numerous electronic devices. The converted energy can be full-wave or half-wave rectified and filtered current. The charging schemes depend on how the ECS **160** is used and its reuse requirements. The output of the ECS **160** can be distributed for reuse to several different types of circuitry. In one embodiment, the converted power may be fed to a trickle charger circuitry that charges a standby energy storage device in the ESS **190** that includes a battery, capacitor, etc. while any excess power is distributed to a charging circuitry for the primary power storage component. Another way to distribute the ECS output power is to feed it directly into circuitry appropriate for affecting indefinite operation of the electronic device in a standby or sleep mode.

The three subsystems of the ERS system are now described in further details.

Energy Harvesting Subsystem (EHS)

The EHS for RF energy collection includes a single antenna of any physical design and shape (e.g., dipole, Yagi, Contrawound Toroidal Helical Antenna, etc.) or an array of such antennas to collect RF energy from any frequency or bands of frequencies. Appropriate baluns (i.e., balanced to unbalanced converters) and impedance matching techniques may be integrated to maximize power transfer into the energy conversion circuitry of the ECS.

For the ERS to reclaim the maximum amount of energy, the energy-harvesting antenna is preferably designed to be a wideband, omnidirectional antenna or antenna array that has maximum efficiency at selected bands of frequencies containing the highest energy levels. If the EHS includes an array of antennas, each antenna in the array can be designed to have maximum efficiency at the same or different bands of frequency from one another. The collected RF energy is then converted into usable DC power using high speed switching semiconductor devices or a diode-type or other suitable rectifier. This power may be used to drive, for example, an amplifier/filter module connected to a second antenna system that is optimized for a particular frequency and application. Thus, one antenna acts as an energy harvester while the other antenna acts as a signal transmitter/receiver. Transforming and/or storing the power garnered by the ERS technology over time makes those devices employing such technology more efficient and extends battery life.

According to a preferred embodiment of the present invention, the antenna design and construction preferably begins with defining a wide band of frequencies over which collection will be required. Next, optimal length or element size is preferably calculated to achieve maximum efficiency over that frequency range. As is apparent to one skilled in the art, antenna efficiency depends on a number of factors related to, for example, the physical design and shape of the antenna. Other factors affecting the amount of energy col lected include signal strength at the receiving location from the transmitting source, which in turn depends on the paths from the transmitting source to the ERS of the present invention.

Once the optimal length or element size of the antenna is calculated, it is subsequently reduced to a fraction of a wavelength while integrating semiconductor amplification and impedance matching circuits in the construction of the antenna elements. The antenna circuit elements are preferably constructed using microprocessor wafer manufacturing techniques and shaped for integration or encapsulation in the antenna elements. This preferred design approach reduces the physical size of the antenna while retaining the energy collection efficiencies of a larger antenna. Antenna efficiency of 50% or more is preferred, for higher antenna efficiency resulting in higher energy collection.

FIG. **5**A shows the physical structure of an energy-harvesting antenna and FIG. **5**B shows an array of such antennas in accordance to an embodiment of the present invention. The cylindrical physical design shape was selected for easy explanation of the design concept; however, any physical shape (e.g., toroid, etc.) may evolve that still meets the ERS design concepts and goals. As shown, the antenna **200** includes a standard metal outer sleeve **210**. Under the sleeve **210** is a semiconductor material **215** used to create an apparent physical length optimized for reception to accommodate size limitations. Enclosed by the antenna outer sleeve **210** and semiconductor material **215** is electronic circuitry **220**, which is constructed using microprocessor and ASIC semiconductor technology to develop integrated support modules such as trickle charger, amplifier and filter for the ECS, which will be described later. The outer sleeve **210** may also be used as a heat sink for the electronic circuitry **220**.

According to another preferred embodiment of the present invention, the antenna can also be constructed to comply with recent radiation hazard concerns with wireless and cellular telephone technologies and at the same time increases power collection potential. There are a number of ways to direct the RF energy away from the user or shield the user using cupped shields in the antenna system design. Such design could be optimized to provide an insulated shield and enhance energy collection materials in the antenna system construction. FIG. **6** shows an antenna **210** covered by an insulated shield **232**. The antenna **210** performs its primary function for a specific application, such as receiving RF signals for normal wireless communication. The insulated shield **232** can be used as an electromagnetic shield and also as an ERS antenna or antenna array for energy harvesting. As shown in the figure, the RF shielding material is preferably located on the inside surface **234** of the shield **232**. As is known in the art, the shielding material may be of any material suitable for shielding RF radiation generated by the transmitting antenna **210**. Within the shield **232** is an antenna array **236**, with associated integrated modules of the ECS, for energy harvesting. Alternatively, within the shield **232** is a single antenna for energy harvesting. Each antenna in array **236** or the alternative single antenna is depicted in FIG. **5**A.

FIG. **7** shows a high level schematic diagram of an Energy Harvesting Subsystem (EHS) in accordance to one embodiment of the present invention. The EHS includes a wideband omnidirectional antenna or antenna array **255** for energy harvesting, as described earlier. The antenna **255** may be used to harvest energy from ambient RF signals **3** and/or RF power transfer signal **2** from a remotely-located, intentional power source. It may also be used to concurrently receive the desired communication signal **1** along with the energy-harvesting signals **2** and **3**. The combined signals are then separated by filter bank **250** which will be described in greater detail later.

FIG. **7** also shows that other energy harvesting technologies for collecting ambient solar, acoustical, and mechanical energy can be integrated as part of the EHS to supply electrical energy into the ECS. The electrical energy supplied by these other energy harvesting technologies may interface with the rectifier **162** and/or filter circuitries **250**, which is then fed to filter **275** in the ECS. As apparent to one skilled in the art, the ECS interface depends on the input frequency, quality, quantity and type (constant DC or some form of pulsating DC) of energy supplied.

For ambient solar energy, an array of solar cells or a solar energy conversion device **260** known in the art (such as a MEMS solar/heat electric generator) is preferably used to absorb solar energy from the environment and convert it to electrical energy. This energy is then delivered to a rectifier **162** of the ECS. An array of solar cells may be used to produce sufficient power for a variety of applications.

For ambient acoustical energy, a piezoelectric transducer **266** known in the art can be used to absorb ambient or intentional sound from the environment, which in turn causes a crystal to vibrate and through the piezoelectric effect produce an output voltage. The output energy is then delivered to the rectifier **162** of the ECS. An array of transducers may be used to produce sufficient power for a variety of applications.

For mechanical energy, the energy is preferably converted to electric energy through a MEMS Mechanical Transducer (Generator) **270** known in the art. The MEMS Mechanical Transducer **240** takes mechanical energy derived from the natural acceleration of a thing or person while in transport (e.g., person walking) or in use. These devices can wind a spring, force a piston to move or use some other method to store and convert acceleration energy collected into electrical energy. The output energy is then delivered to the rectifier **162** of the ECS. An array of such transducers may be used to produce sufficient power for a variety of applications.

Energy Conversion Subsystem (ECS)

According to a preferred embodiment of the present invention, the ECS for RF energy collection includes a power charger and other circuitry for performing RF to DC power conversion. Because the energy harvesting in the present invention is focused mainly on recovering small amounts of energy over long periods of time, charging energy storage devices (rechargeable batteries, etc.) may be done over a period of time by trickle charge. As is known in the art, trickle charging is a charging scheme in which a power storage component is charged at a fraction of its capacity rate. FIG. **8** shows an ECS **160** for use in conjunction with the EHS to perform RF to DC power conversion, with specifications in accordance to an embodiment of the present invention.

The ECS **160** preferably includes a transformer **161** connected to an energy harvesting antenna **130** for receiving the RF energy. The transformer **161** is also connected to a diode rectifier circuit **162**. The transformer **161** and the rectifier circuit **162** are designed to match impedance with one another to prevent undesired energy loss between these two elements. After rectification, the converted energy is sent through another transformer **163** and onward to a trickle charger **164**. Again, the second transformer **163** and the trickle charger **164** are designed to match impedance with one another to prevent undesired energy loss between these two elements. The trickle charger **164** is then used to deliver power to a battery **165** for storage and use with an electronic device, such as a standby device **166**. The battery is actually a part of the ESS, which will be described later.

If an array of energy harvesting antennas is used to receive the RF energy, the power collected from all channels (i.e., the power density at all antenna inputs) can be combined, rectified and forwarded to the trickle charger **164**. The power available in each channel is comparable; thus, for example, if 100 N nW is required for a load, the power collected in N channels at 100 nW each can be combined to generate the required power.

Although not shown in FIG. **8**, the converted and rectified power may be fed through a filter for signal purification prior to sending it to the transformer **163** and onward to the trickle charger **164**. FIG. **7** shows the use of such filter **275**, with like numbers for like elements shown in FIG. **8**. It should be noted that FIG. **7** depicts a broad overview of the ERS concept according to an embodiment of the present invention; therefore, the figure does not show all detailed components of the ECS as shown in FIG. **8**.

FIG. **9** shows an ASIC chip implementation of the ECS circuitry shown in FIG. **8**. The RF input **280** corresponds to the antenna **130** of FIG. **8** for receiving harvested RF energy; the MEMS RF transformer **281** corresponds to the transformer **161** of FIG. **8**; the RF power ASIC **282** corresponds to the rectifier circuit **162** and the trickle charger **164** of FIG. **8** and any other desired circuitry for the ECS, such as filters, impedance matching circuitry, etc.; and the micro-cell battery **285** corresponds to the battery **165** of FIG. **8**. The battery **165** or **285** is a part of the ESS, which is described next.

Energy Storage Subsystem (ESS)

According to a preferred embodiment of the present invention, the ESS includes a rechargeable battery **165**, which can be a complete battery or an array of micro-cells of battery. If battery **165** includes a complete battery, the trickle charger **164** preferably provides a larger potential difference between terminals and more power for charging during a period of time. If battery **165** includes individual battery cells, the trickle charger **164** preferably provides smaller amounts of power to each individual battery cell, with the charging proceeding on a cell by cell basis. FIG. **10** shows an array of battery micro-cells according to an embodiment of the present invention. The ESS charges one or more micro-cells as available energy allows. As the load requires power, the combined energy is drawn from a charged P×Q subset (sub-array) of the N×M set (array) of cells to power the device on standby or active mode. N, M, P and Q are all natural numbers (i.e., 1, 2, 3 . . .), wherein P is a number less than N, and Q is a number less than M. Charging of the remaining cells continues whenever ambient power is available. As the load depletes cells, switches such as switch **276** are used to replace the depleted cells with charged cells. The rotation of depleted cells and charged cells continues as required. As much energy as possible is banked to ensure adequate power is available on demand. Thus, energy is managed on a micro-cell basis.

According to the present invention, there are numerous ways in which the ERS technology can be configured to harvest RF energy. In one embodiment of the present invention, an electronic device, such as a cellular phone, is preferably designed to have two antennas, one for energy harvesting at a desired frequency band and one for performing RF communication at a different frequency band. An example of this dual-antenna design is previously shown in FIG. **6**. FIG. **11** shows a high level schematic diagram **300** of a dual-antenna design ERS in accordance to one embodiment of the present invention. As explained earlier, the frequency band at which an antenna receives and transmits signals depends on the physical dimensions of the antenna. Thus, the energy-harvesting antenna **310** and the communication antenna **320** are designed with different dimensions to receive signals at their respectively desired frequency bands. The energy-harvesting antenna **310** may comprise one antenna or an array of like antennas. In this embodiment, parallel operations are possible for energy harvesting and communication. In other words, a user can be using the phone for wireless communication in one frequency band, while at the same instance the energy harvesting antenna **310** and associated circuitry within the phone is collecting electromagnetic energy in another frequency band. The collected electromagnetic energy is passed to integrated circuitry that converts the RF energy into electrical energy. This electrical energy is then distributed for storage in batteries of the ESS **340** for use in powering various electrical components of the phone.

As shown in FIG. **11**, communication signals received by the device antenna **320** are sent to communication processing circuitry **390** for transmission and reception. The energy harvesting antenna **310**, on the other hand, is connected to an Energy Conversion Subsystem or ECS **330**, which is designed to match the impedance and frequency band or bands of the energy harvesting antenna **310** for maximizing input energy received. The ECS **330** is as described earlier with respect to FIG. **8**. Its output is conditioned power coming from a trickle charger designed to efficiently charge the power storage devices, such as NiCd, NiMH, or LiIon battery cells, in the energy storage subsystem or ESS **340**. Any excess conditioned power from the trickle charger of the ECS **330** is then distributed to the device's main power source **380**.

The ESS **340** is as described earlier with respect to FIGS. **8** and **10**. It stores the standby power and distributes it to the electronic switching circuitry **350** and the Monitor and Activation Circuitry **360** in both active and standby modes of the device. Alternatively, the ESS **340** may be used to power circuitries **350** and **360** in standby mode of the device, while the device power source **380** powers those circuitries in the device's active mode. The ESS **340** may also be used to complement the device power source **380**.

According to yet another embodiment of the present invention, the harvesting antenna **310** and associated circuitries **330** and **340** may be physically located outside of the cellular phone but yet capable of electrical coupling with the phone. For example, the harvesting antenna **310** and associated circuitries may be built into portable physical structures such as briefcases, suitcases, purses, and wearable apparel (cloths) or fixed physical structures by any known means in the art. Thus, when the phone is desired to be charged, it can be stored in any of the aforementioned physical structures and electrically coupled to the built-in harvesting antenna and circuitry for charging purposes. These separate structures facilitate the use of larger antennas (transducers) or arrays of smaller elements whichever is most efficient for collecting the ambient energy. In any case, the larger surface area intercepts more energy.

According to a further embodiment for harvesting RF energy in the present invention, an electronic device, such as a cellular phone, is designed to have two antennas that respectively harvest RF energy and perform wireless communications in the same desired frequency band. In this embodiment, the energy-harvesting antenna and the communication antenna are designed with the same physical dimensions. The communication antenna is connected to a signal level detection system for detecting the strength of RF signals in the desired frequency band. If the strength of the desired RF signals is sufficiently high to sustain both RF communication and energy harvesting, then the level detection system will cause the activation of the energy-harvesting antenna. Thus harvesting proceeds in parallel with communication and the RF energy is converted into electrical power useful for powering various components of the cellular phone without degrading phone service.

According to another embodiment for harvesting RF energy in the present invention, a single antenna or an array of like antennas in an electronic device, such as a cellular phone, is used for both energy harvesting and communication purposes. FIG. **12** shows a high level schematic diagram **400** of the Common Antenna for Simultaneous Energy Harvesting, Sense and Active Mode in accordance with this embodiment. This mode requires modifications to the device's antenna switching and filter circuitry to allow communication signals in along with collected RF energy for harvesting. The communication signals are filtered and directed to the receiver's communication circuitry while the noise or interference portions of the received RF signals are filtered out and sent to the energy harvesting circuitry within the electronic device. This noise filtering technique is illustrated back in FIG. **7**, where incoming communication signals **1** and the ambient RF energy **3** are filtered by a filter bank **250**, which may reside in the switching circuitry **450** shown in FIG. **12**. The communication signals **1** are then directed to the receiver's communication circuitry **255**, which is shown as communication processing circuitry **490** in FIG. **12**, while the noise or interference RF portions **3** are filtered out and sent to the energy harvesting circuitries, which include the ECS **430** and ESS **440**.

According to still another embodiment for harvesting RF energy in the present invention, a single antenna or an array of like antennas in an electronic device, such as a cellular phone, is again used for both energy harvesting and communication purposes. FIG. **13** shows a high level schematic diagram **500** of the Common Antenna Sense and Active Mode in accordance with this embodiment. Here, the single antenna or antenna array **510** is connected to the energy harvesting circuitries **530** and **540** and communication circuitry **590** by a switch **550**, which is slaved to the various powering modes of the electronic device. For example, when the phone is in the power-on mode, the switch **550** is programmed to connect the antenna **510** with the communication circuitry **590** to receive the RF communication signals and enable wireless communication with the phone. However, when the phone is in the standby mode, the switch is programmed to connect the antenna **510** with the harvesting circuitries; which then converts the electromagnetic energy in the same RF communication signals into electrical power for powering the electronic device in the standby mode.

The normal operation of an electronic device having the ERS technology as depicted in FIGS. **11** and **12** may also include a standby or sleep mode, as described above with regard to FIG. **13**. When the electronic device is set in a standby mode, the electronic switching circuitry (**350** or **450**) of the ERS (**300** or **400**) within the electronic device is placed in a Sense Mode. As shown in both FIGS. **11** and **12**, the monitor and activation circuitry (**360** or **460**) is used to set the electronic switching circuitry (**350** or **450**) in a sense mode. In this mode, signals received by the device's communication antenna (**320** or **410**) are sent to a band pass filter (**370** or **470**) to retrieve an activation signal, which is passed on to the monitor and activation circuitry (**360** or **460**) that is watching for the activation signal. The activation signal can be anything defined by the device's manufacturer, such as a trigger signal or an emergency beacon for activating a desired system. When a turn-on signal brings the electronic device out of standby mode, the monitor and activation circuitry (**360** or **460**) sends an active signal to the electronic switching circuitry (**350** or **450**) to connect the communication antenna (**320** or **410**) to the device's communication circuitry (**390** or **490**). The device's turn-on signal may be manually provided by a user turning on the electronic device, or it may be brought on by the reception of the activation signal by the monitor and activation circuitry (**360** or **460**).

Again, the harvesting circuitries in the above embodiments may be physically located apart from the communication circuitry. For example, the harvesting circuitries may be built into portable structures such as briefcases, suitcases, handbags, and purses by any known means in the art, while the communication circuitries and the switch are built into the phone itself. Thus, when the phone is desired to be charged, it can be stored in any of the aforementioned portable structures and electrically coupled to the harvesting circuitries built into the structure for charging purpose.

Although only a few exemplary embodiments of this invention have been described in detail above, those skilled in the art will readily appreciate that many modifications are possible in the exemplary embodiments without materially departing from the novel teachings and advantages of this invention. Accordingly, all such modifications are intended to be included within the scope of this invention as defined in the following claims. Furthermore, any means-plus-function clauses in the claims (invoked only if expressly recited) are intended to cover the structures described herein as performing the recited function and all equivalents thereto, including, but not limited to, structural equivalents, equivalent structures, and other equivalents.

What is claimed is:

1. A method for harvesting and utilizing electromagnetic energy, comprising:

receiving ambient electromagnetic energy;

converting the ambient electromagnetic energy into DC electrical power;

charging a power storage component with the DC electrical power, the power storage component comprises a N×M array of battery micro-cells, N and M are natural numbers;

providing a device power source for powering an electrical device; and

drawing power from the power storage component to power the electrical device, the drawing power from the power storage component to power the electrical device includes:

determining a charged P×Q sub-array of the N×M array of battery micro-cells, P and Q are natural numbers less than N and M, respectively; and

drawing power from the charged P×Q sub-array to power the electrical device.

2. The method of claim **1**, the charging the power storage component with the DC electrical power comprises includes:

charging at least one remaining micro-cell of battery in the N×M array that is not in the charged P×Q sub-array;

substituting the P×Q sub-array with the at least one remaining micro-cell of battery once the P×Q sub array is depleted of power; and

charging the depleted P×Q sub-array with the DC electrical power.

3. The method of claim **2**, the drawing power from the power storage component to power the electrical device further includes:

drawing power from the at least one remaining charged micro-cell of battery to power the electrical device.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 6,882,128 B1
DATED : April 19, 2005
INVENTOR(S) : Rahmel

Page 1 of 1

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Title page,
Item [56], **References Cited**, U.S. PATENT DOCUMENTS, please add the following:
-- 6,045,339 4/4/2000 Berg 417/332 --.

Column 12,
Line 51, please change “component with the DC electrical power comprises” to
-- component with the DC electrical power --.

Signed and Sealed this

Twenty-second Day of November, 2005

Jon W. Dudas

JON W. DUDAS
Director of the United States Patent and Trademark Office

US007383792B1

(12) **United States Patent**
Sharpe

(10) **Patent No.: US 7,383,792 B1**
(45) **Date of Patent: Jun. 10, 2008**

(54) **HYDROGEN GAS INJECTOR PLUG**

(76) Inventor: **Thomas H. Sharpe**, 55 Dorr Dr., North Augusta, SC (US) 29841

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **11/902,588**

(22) Filed: **Sep. 24, 2007**

Related U.S. Application Data

(60) Provisional application No. 60/877,138, filed on Dec. 27, 2006.

(51) **Int. Cl.**
F02B 43/00 (2006.01)
F02B 43/04 (2006.01)

(52) **U.S. Cl.** .... **123/1 A**; 123/25 A

(58) **Field of Classification Search** .... 123/3, 123/DIG. 12, 1 A, 25 B, 25 A, 25 R, 25 D, 123/468, 470
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,799,124	A *	3/1974	Swain	123/1 A
4,037,568	A	7/1977	Schreiber	
4,140,090	A	2/1979	Lindberg	
4,448,160	A *	5/1984	Vosper	123/297
4,702,894	A	10/1987	Cornish	
5,085,176	A	2/1992	Brinkley	
5,676,100	A *	10/1997	Dam et al.	123/145 A
5,690,902	A	11/1997	Werth	
5,852,993	A *	12/1998	Anderson	123/297
6,062,185	A *	5/2000	Chiu et al.	123/145 A
6,119,651	A *	9/2000	Anderson	123/297
6,941,901	B2 *	9/2005	Taxon	123/25 E
2004/0149256	A1 *	8/2004	Dye et al.	123/297

* cited by examiner

Primary Examiner—Mahmoud Gimie
(74) *Attorney, Agent, or Firm*—Richard C. Litman

(57) **ABSTRACT**

The hydrogen gas injector plug for generating hydrogen gas from ambient air drawn into the piston chamber of an engine includes an array of nozzles fabricated from a metal or a mixture of metals. When heated to a pre-determined temperature, the metallic nozzles react with atmospheric water to disassociate hydrogen gas. The hydrogen gas is injected from the nozzles into the piston chamber and mixed with air. An ignition device ignites the mixture of hydrogen and air so that it burns to power the piston in a conventional manner. The metallic nozzle is heated via a copper conductor connected to a source of electric current.

14 Claims, 3 Drawing Sheets

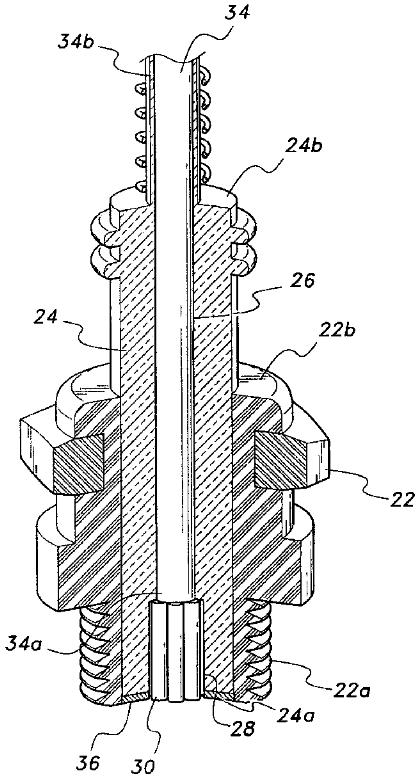

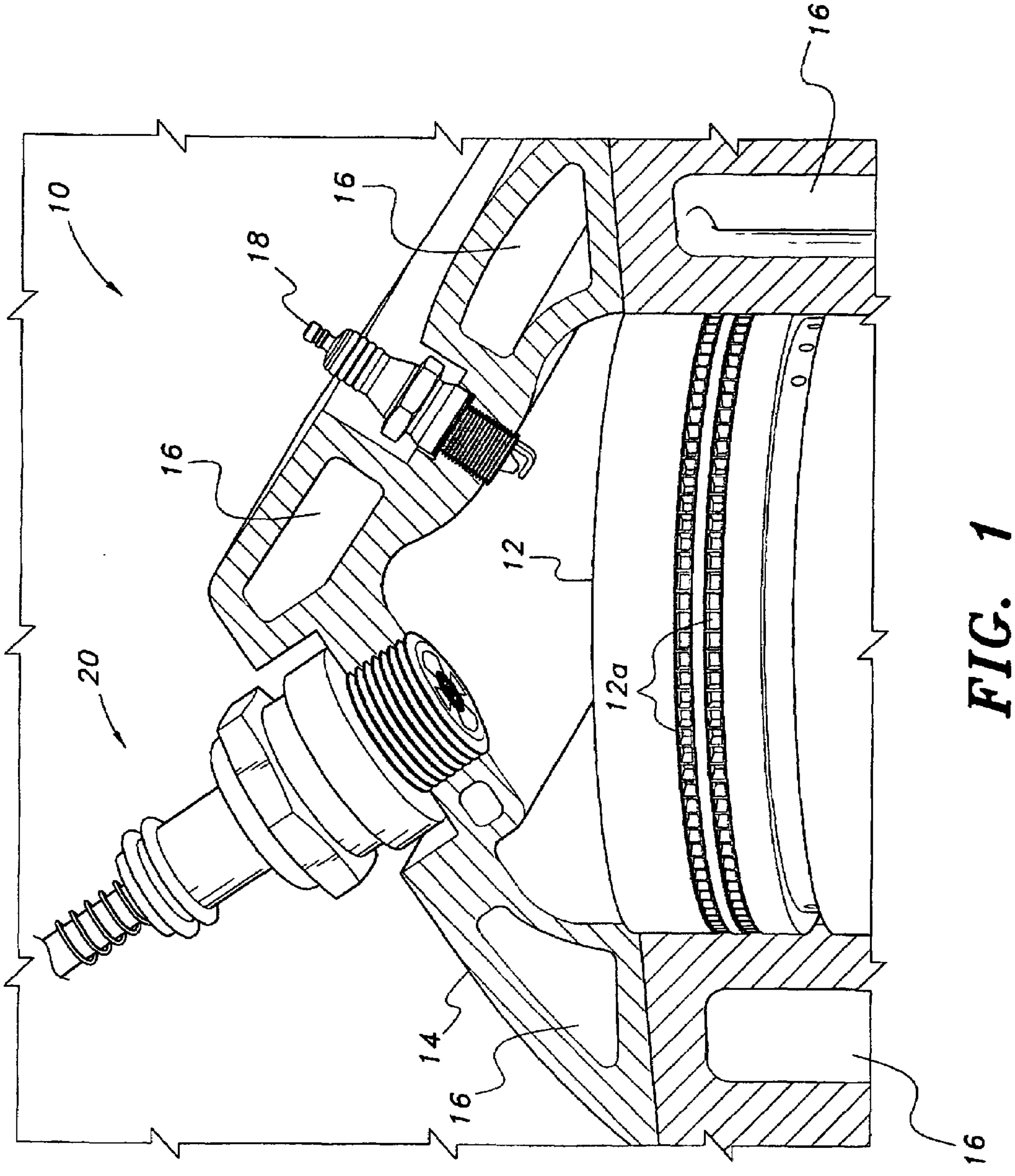
10
18
16
16
16
16
16
12
12a
20
14
16

FIG. 1

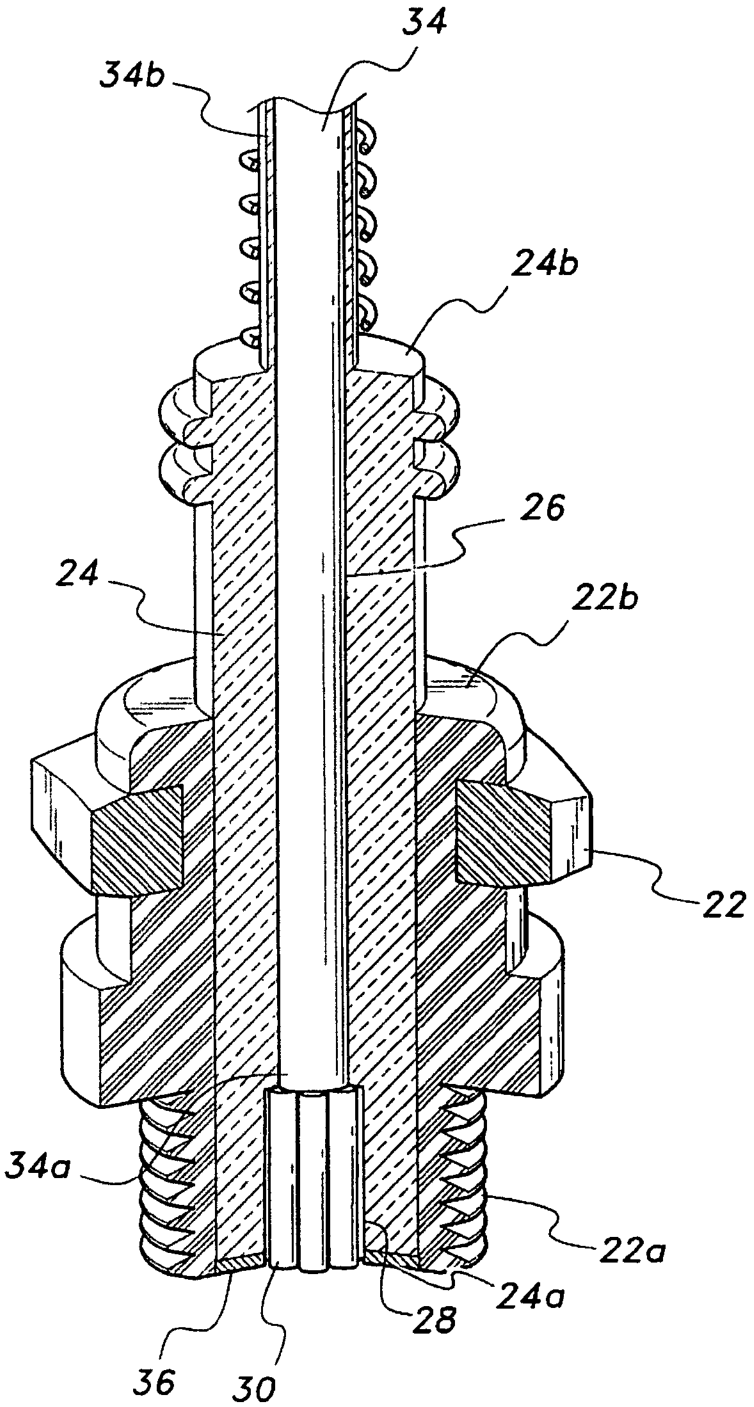
34
34b
24b
26
24
22b
22
34a
22a
24a
28
36
30

FIG. 2

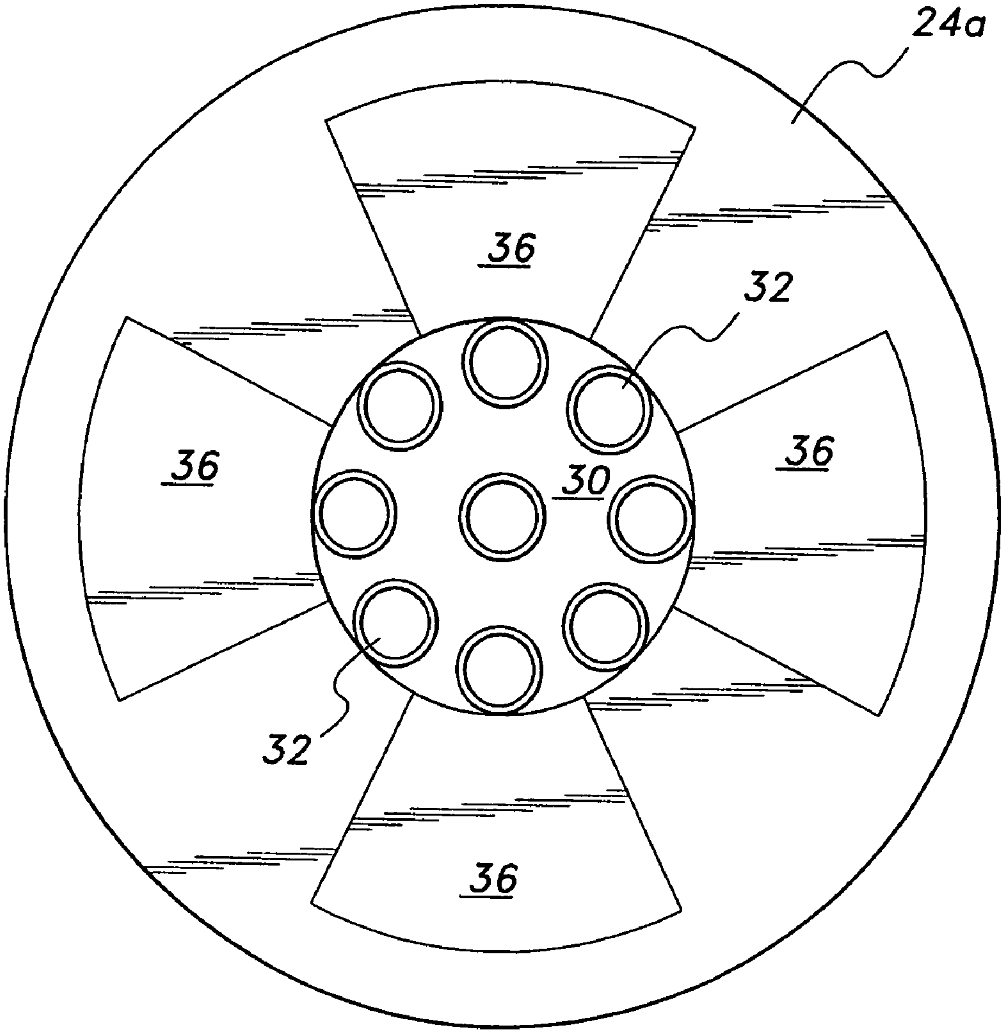
24a
36
32
36
30
36
32
36

FIG. 3

HYDROGEN GAS INJECTOR PLUG

CROSS-REFERENCE TO RELATED APPLICATION

This application claims the benefit of U.S. Provisional Patent Application Ser. No. 60/877,138, filed Dec. 27, 2006.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention generally relates to internal combustion engines. More specifically, the present invention is drawn to a device for injecting hydrogen gas into the combustion chamber of an engine for use as a fuel for powering the engine.

2. Description of the Related Art

The rising cost and diminishing supply of hydrocarbon fuels, i.e. gasoline, have increased the criticality of developing or finding alternative fuels. Furthermore, pollution caused by burning hydrocarbon fuels is suspected of creating a "greenhouse" effect in the atmosphere, thereby creating problems that may have a bearing on the future course of human civilization. The art would certainly welcome a device that could utilize a virtually inexhaustible supply of a common element to power internal combustion engines, which device would also cause production of pollution-free byproducts.

The use of hydrogen as a fuel to power engines has been contemplated for many years. Hydrogen is one of the most abundant elements on earth and combustion of this abundant element produces pollution-free water. Unfortunately, hydrogen poses many risks when stored in large quantities, thus creating many problems in making the gas available to the general public. Thus, a hydrogen gas injector plug solving the aforementioned problems is desired.

SUMMARY OF THE INVENTION

The hydrogen gas injector plug is used for generating hydrogen gas from ambient air drawn into the piston chamber of an engine. The plug comprises an array of nozzles fabricated from one or a mixture of metals. When heated to a pre-determined temperature, the metallic nozzles become "white-hot" and generate photons that react with atmospheric water to disassociate hydrogen gas therefrom. The nozzle is designed so that no reaction will occur if the temperature is below 350° F. The hydrogen gas is injected from the nozzles into the piston chamber and mixed with air. An ignition device ignites the mixture of hydrogen and air so that it burns to power the piston in a conventional manner. The metallic nozzle is heated via a copper conductor connected to a source of electric current.

Accordingly, the invention presents a hydrogen gas generator capable of generating small amounts of hydrogen gas from water available in the atmosphere. The invention provides for improved elements thereof in an arrangement for the purposes described that are inexpensive, dependable and fully effective in accomplishing their intended purposes.

These and other features of the present invention will become readily apparent upon further review of the following specification and drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is an environmental, perspective view of a hydrogen gas generator injector plug according to the present invention.

FIG. **2** is a side, sectional view of a hydrogen gas generator injector plug according to the present invention.

FIG. **3** is an end view of a hydrogen gas generator injector plug according to the present invention.

Similar reference characters denote corresponding features consistently throughout the attached drawings.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Attention is first directed to FIG. **1** wherein an engine cylinder chamber is generally indicated at **10**. A piston **12** having rings **12***a* is housed within chamber **10**. The head **14** and the cylinder wall are surrounded by coolant chamber **16**. Conventional intake and exhaust valves (not shown) are also disposed on head **14**. The structure and arrangement of the aforementioned items are conventional and are not part of the inventive concept, per se. A hydrogen gas generator injector plug **20** and a conventional spark plug **18** are mounted on head **14**. Injector plug **20** and conventional spark plug **18** each have proximate ends disposed within cylinder chamber **10**.

As best seen in FIGS. **2** and **3**, hydrogen gas generator injection plug **20** comprises a circular member **22** fabricated from steel and having a threaded proximate end **22***a* and a distal end **22***b*. A passageway extends through member **22** from the distal end **22***b* to proximate end **22***a*. An insulation member **24** is disposed in the passageway. Insulation member **24** is fabricated from a material having high temperature insulation characteristics. Member **24** has a first end **24***a* that is disposed adjacent to, but spaced slightly inwardly of proximate end **22***a*. The second end **24***b* of insulation member **24** extends above the distal end **22***b* of member **22**. A passageway **26** extends through insulation member **24** from second end **24***b* and terminates at a cavity **28**, which cavity is formed in the first end **24***a* of insulation member **24**. A nozzle member **30** is positioned in cavity **28**. Nozzle member **30** is fabricated from a metal or a metal mixture that has the ability to react with water at relatively high temperatures to oxidize and generate hydrogen gas. Nozzle member **30** must also have the ability to become "white hot" very quickly when an adequate electrical current is applied thereto. As currently contemplated, iron, certain alloys of iron and tungsten exhibit the requisite abilities needed to function as the nozzle member. It should be noted however, that any metal or alloy could be utilized if suitable. An array of hydrogen gas injector passages **32** is formed in nozzle member **30**. It has been determined that nine injector passages will provide an adequate amount of hydrogen gas to power piston **12**. An electric current conductor **34**, preferably fabricated from copper, extends through passageway **26**. The lower end **34***a* of conductor **34** abuts nozzle member **30**. The upper end defines a terminal for an electrical connection **34***b*. Four electrical conductors **36** abut the second end **24***a* of member **24** and are evenly spaced therearound. Conductors **36** are fabricated from a nickel-copper alloy.

In use, ambient air containing water vapor is drawn into the cylinder chamber via conventional intake valves (not shown) on the intake stroke. A portion of the air water vapor mixture is forced into injector passages **32** on the compression stroke. An electric current is applied to nozzle member

30 via conductor **34** during the latter part of the compression stroke, heating the nozzle member to a "white-hot" temperature. The water vapor reacts with the metal in nozzle member **30** to oxidize the metal and produce hydrogen gas. The hydrogen gas expands through the passages **32** back into the cylinder chamber where it mixes with the oxygen in the air. Spark plug **18** is fired to ignite the mixture, which mixture burns thereby creating a high gas pressure to drive the piston. Occasionally (and in some climates), the humidity may be very low and the ambient air will not contain enough water vapor to produce an adequate amount of hydrogen gas to properly drive the piston. In such instances additional water vapor is injected into the intake valve.

It is to be understood that the present invention is not limited to the embodiment described above, but encompasses any and all embodiments within the scope of the following claims.

I claim:

1. A hydrogen gas injector plug comprising:

a circular member having a proximate end and a distal end;

a first passageway extending through said circular member from said proximate end to said distal end;

a heat insulating member disposed in said first passageway, said heat insulating member having a first end and a second end, said first end terminating adjacent said proximate end of said circular member;

a second passageway extending through said heat insulating member, said second passageway extending from said first end to said second end of said heat insulating member;

a cavity formed in said second passageway at said first end;

a nozzle array disposed in said cavity; and

a first electrical conductor disposed in said second passageway, said electrical conductor in abutment with said nozzle array.

2. The hydrogen gas injector plug according to claim **1**, wherein said circular member has an outer surface having threads disposed on said outer surface at said proximate end.

3. The hydrogen gas generator injector plug according to claim **1**, further including plural electrical conductors disposed on said first end of said heat insulating member.

4. The hydrogen gas injector plug according to claim **1**, wherein said circular member is fabricated from steel.

5. A hydrogen gas injector plug, comprising:

a circular member fabricated from steel and having a proximate end, a distal end and an outer surface;

screw threads disposed on said outer surface at said proximate end;

a first passageway extending through said circular member from said proximate end to said distal end;

a heat insulating member disposed in said first passageway, said heat insulating member having a first end and a second end, said first end terminating adjacent said proximate end of said circular member;

a second passageway extending through said heat insulating member, said second passageway extending from said first end to said second end of said heat insulating member;

a cavity formed in said second passageway at said first end;

a nozzle array disposed in said cavity; and

a copper electrical conductor disposed in said second passageway, said electrical conductor in abutment with said nozzle array.

6. The hydrogen gas injector plug according to claim **5**, wherein the number of nozzles in said array is nine.

7. The hydrogen gas injector plug according to claim **5**, wherein the number of nozzles in said array is nine, the hydrogen gas injector plug further including plural electrical conductors disposed on said first end of said heat insulating member.

8. The hydrogen gas injector plug according to claim **5**, wherein each nozzle in said array is fabricated of material from the group consisting of iron, iron alloys and tungsten.

9. The hydrogen gas injector plug according to claim **5**, wherein the number of nozzles in said array is nine, the hydrogen gas injector plug further including plural electrical conductors disposed on said first end of said heat insulating member, each of said plural electrical conductors being fabricated from a nickel-copper alloy.

10. A hydrogen gas injector plug in combination with a piston chamber of an internal combustion engine, the combination comprising:

a circular member fabricated from steel and having a proximate end, a distal end and an outer surface;

screw threads disposed on said outer surface at said proximate end for removably mounting said circular member on the piston chamber;

a first passageway extending through said circular member from said proximate end to said distal end;

a heat insulating member disposed in said first passageway, said heat insulating member having a first end and a second end, said first end terminating adjacent said proximate end of said circular member;

a second passageway extending through said heat insulating member, said second passageway extending from said first end to said second end of said heat insulating member;

a cavity formed in said second passageway at said first end;

a nozzle array disposed in said cavity;

a copper electrical conductor disposed in said second passageway, said electrical conductor in abutment with said nozzle array; and

a spark plug mounted on the piston chamber.

11. The hydrogen gas injector plug combination according to claim **10**, wherein the number of nozzles in said array is nine.

12. The hydrogen gas injector plug combination according to claim **10**, wherein the number of nozzles in said array is nine, the combination further including plural electrical conductors disposed on said first end of said heat insulating member.

13. The hydrogen gas injector plug combination according to claim **10**, wherein each nozzle in said array is fabricated of material from the group consisting of iron, iron alloys and tungsten.

14. The hydrogen gas injector plug combination according to claim **10**, wherein the number of nozzles in said array is nine, the combination further including plural electrical conductors disposed on said first end of said heat insulating member, each of said plural electrical conductors being fabricated from a nickel-copper alloy.

* * * * *

US008981261B1

(12) **United States Patent**
Tillotson

(10) **Patent No.: US 8,981,261 B1**
(45) **Date of Patent: Mar. 17, 2015**

(54) **METHOD AND SYSTEM FOR SHOCKWAVE ATTENUATION VIA ELECTROMAGNETIC ARC**

(75) Inventor: **Brian J. Tillotson**, Kent, WA (US)

(73) Assignee: **The Boeing Company**, Chicago, IL (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 336 days.

(21) Appl. No.: **13/483,995**

(22) Filed: **May 30, 2012**

(51) **Int. Cl.**
F42D 5/045 (2006.01)
F41H 5/007 (2006.01)

(52) **U.S. Cl.**
USPC **219/383**; 219/201; 219/202; 89/36.01; 89/36.07; 89/36.08; 89/36.09

(58) **Field of Classification Search**
None
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

1,195,042	A	8/1916	Leon	
2,405,694	A	8/1946	Nicolas	
2,513,279	A	7/1950	Albert	
3,050,707	A	8/1962	Baker et al.	
3,660,951	A	5/1972	Cadwell	
3,773,168	A	11/1973	Meinass	
3,875,844	A	4/1975	Hicks	
3,943,870	A	3/1976	Paslay	
4,215,630	A	8/1980	Hagelberg et al.	
4,313,181	A	1/1982	Holm	
4,543,872	A	10/1985	Graham et al.	
5,020,411	A *	6/1991	Rowan	89/1.11
H1231	H	9/1993	Richards	
5,341,718	A	8/1994	Woodall et al.	
5,394,786	A	3/1995	Gettle et al.	
5,400,688	A	3/1995	Eninger et al.	
5,739,458	A	4/1998	Girard	
6,029,558	A	2/2000	Stevens et al.	
6,256,263	B1	7/2001	Stevens	
6,266,926	B1	7/2001	Figge et al.	
6,279,449	B1	8/2001	Ladika et al.	
6,412,391	B1	7/2002	Stevens et al.	
6,595,102	B2	7/2003	Stevens et al.	
6,653,972	B1	11/2003	Krikorian et al.	
6,782,790	B2 *	8/2004	Barrett	89/1.11

(Continued)

FOREIGN PATENT DOCUMENTS

WO	97/16697	5/1997
WO	2011/148165	12/2011

OTHER PUBLICATIONS

"Review of methods to attenuate Shock/Blast waves," Igra et al, Dec. 2012.*

Primary Examiner — Joseph M Pelham
(74) *Attorney, Agent, or Firm* — Thompson Hine LLP

(57) **ABSTRACT**

A method and system for attenuating a shockwave propagating through a first medium by heating a selected region of the first fluid medium rapidly to create a second, transient medium that intercepts the shockwave and attenuates its energy density before it reaches a protected asset. The second medium may attenuate the shockwave by one or more of reflection, refraction, dispersion, absorption and momentum transfer. The method and system may include a sensor for detecting a shockwave-producing event, determining a direction and distance of the shockwave relative to a defended target and calculating a firing plan, and an arc generator for creating the second medium. The arc generator may create the second medium by creating an electric arc that travels along an electrically conductive path utilizing at least one of high intensity laser pulses, pellets forming a conductive ion trail, sacrificial conductors, projectiles trailing electrical wires, and magnetic induction.

23 Claims, 7 Drawing Sheets

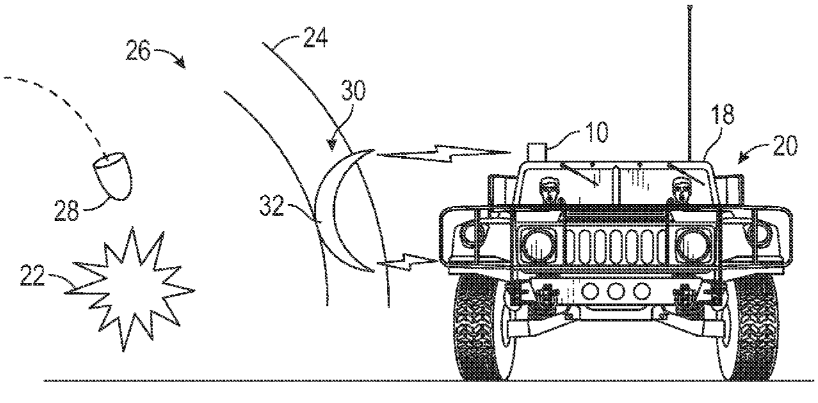

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,077,049	B2	7/2006	Shumov et al.	
7,213,494	B2	5/2007	James	
7,350,447	B1 *	4/2008	Smith	89/1.13
7,421,936	B2	9/2008	Barger et al.	
7,437,987	B1	10/2008	Ohnstad et al.	
7,827,900	B2	11/2010	Beach et al.	
7,866,250	B2	1/2011	Farinella et al.	
7,878,103	B2	2/2011	Imholt et al.	
8,042,449	B2	10/2011	Farinella et al.	
8,051,762	B2	11/2011	Beach et al.	
8,141,470	B1	3/2012	Farinella et al.	
8,151,710	B2	4/2012	Fu et al.	
8,436,730	B2	5/2013	Fischbach et al.	
8,437,223	B2	5/2013	Barger et al.	
8,448,559	B2	5/2013	Hunn et al.	
8,555,768	B1	10/2013	Barker et al.	
8,863,666	B2 *	10/2014	Tillotson et al.	102/215
2003/0233931	A1 *	12/2003	Nemtsev	89/1.11
2004/0107827	A1	6/2004	Edberg et al.	
2004/0118270	A1 *	6/2004	Barrett	89/1.11
2007/0006723	A1	1/2007	Waddell, Jr. et al.	
2007/0180983	A1	8/2007	Farinella et al.	
2008/0190276	A1	8/2008	Barger et al.	
2009/0114084	A1	5/2009	Thinn et al.	
2009/0266226	A1	10/2009	Beach et al.	
2010/0319524	A1	12/2010	Farinella et al.	
2010/0319526	A1	12/2010	Imholt et al.	
2011/0120294	A1	5/2011	Beach et al.	
2011/0168004	A1	7/2011	Henegar	
2011/0297031	A1	12/2011	Fu et al.	
2012/0060677	A1	3/2012	Farinella	

* cited by examiner

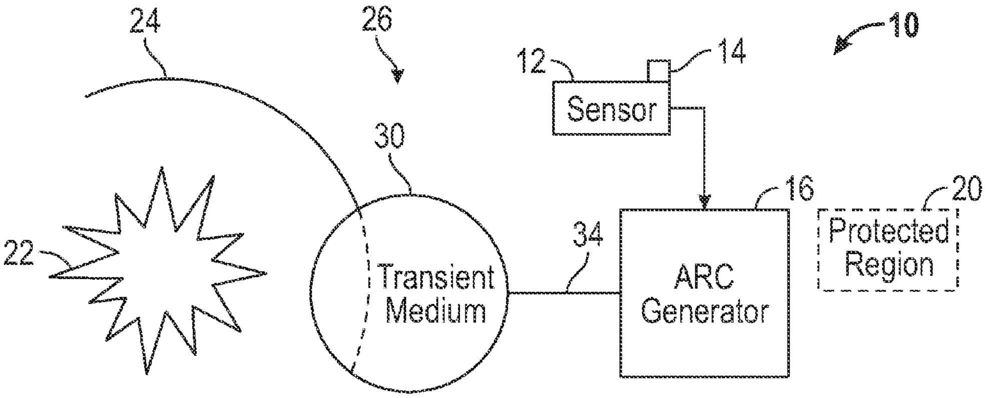
Transient Medium
Sensor
ARC Generator
Protected Region

FIG. 1

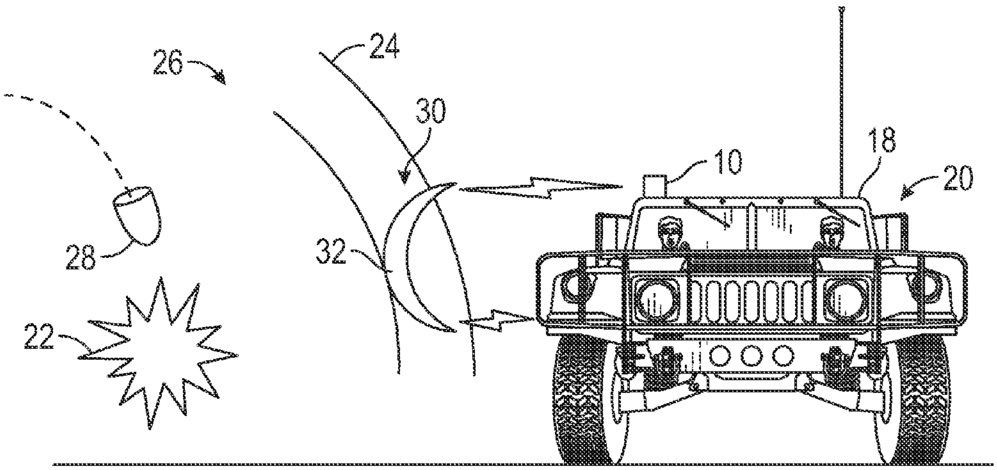

FIG. 2

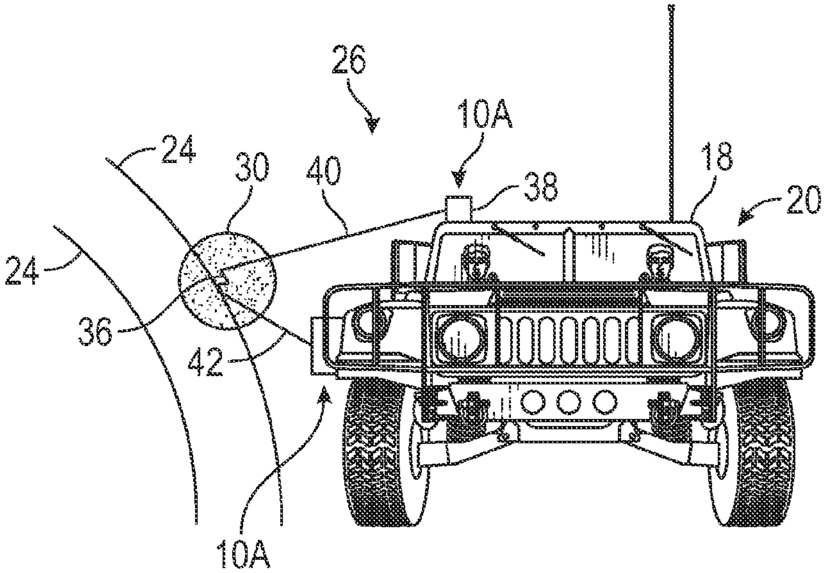

FIG. 3

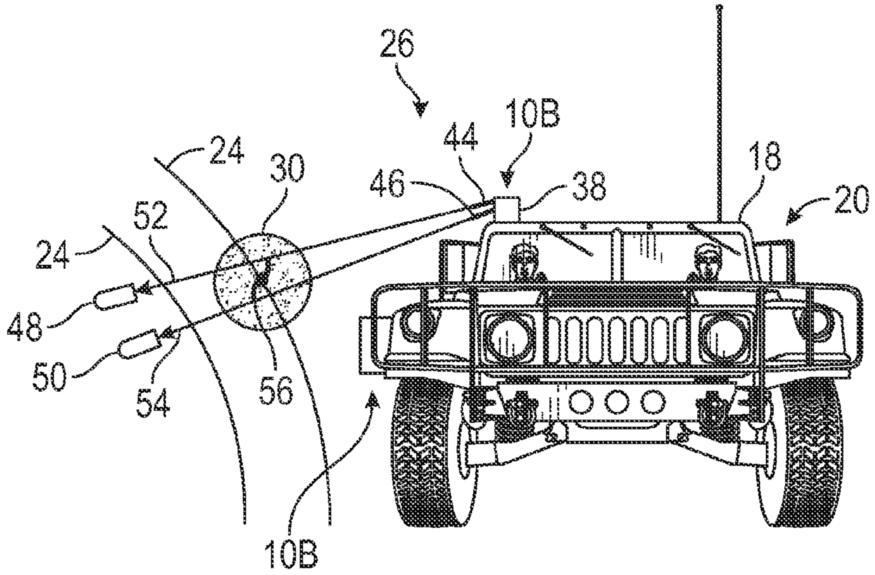

FIG. 4

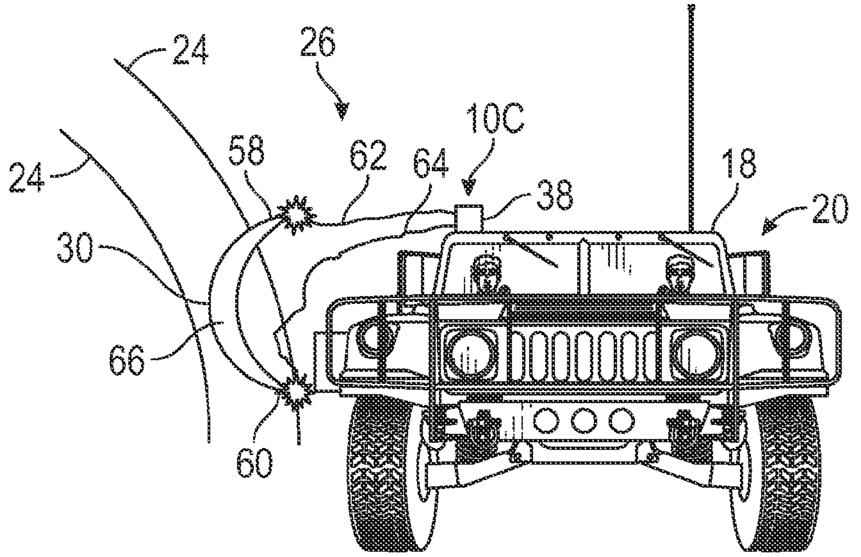
24
26
24
58
62
64
10C
38
18
20
30
66
60

FIG. 5

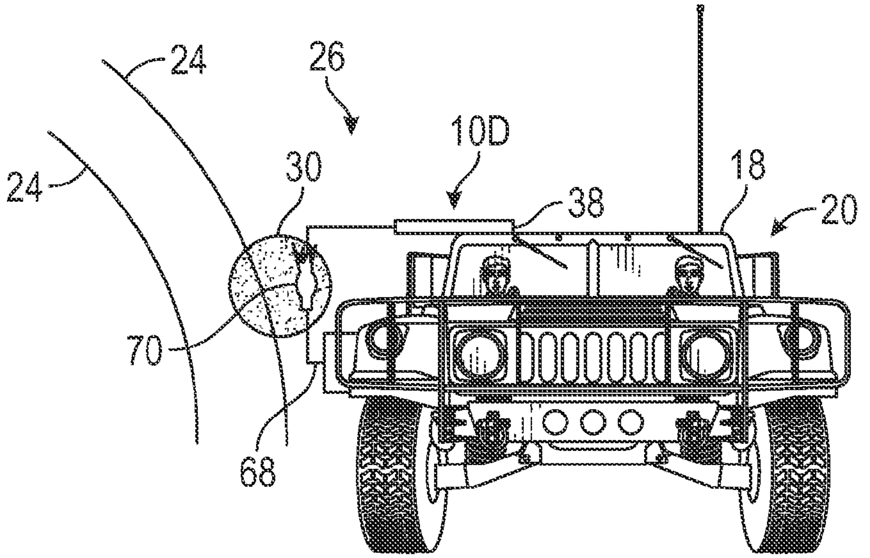
24
26
24
10D
30
38
18
20
70
68

FIG. 6

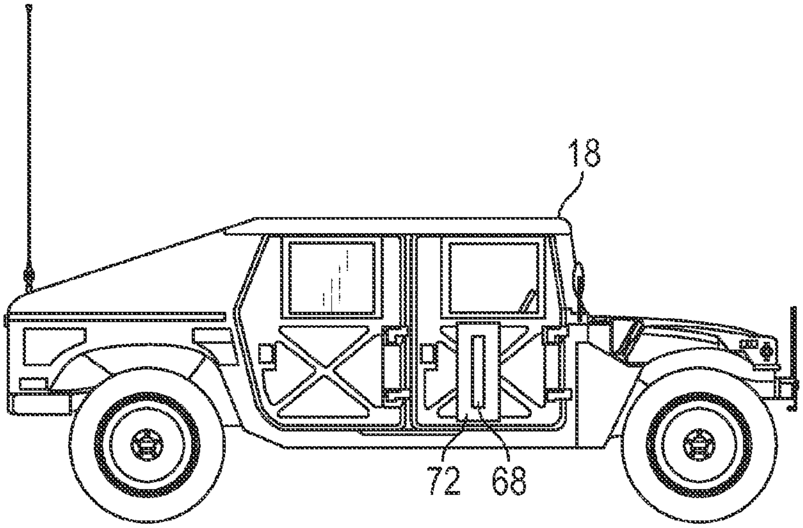

FIG. 7

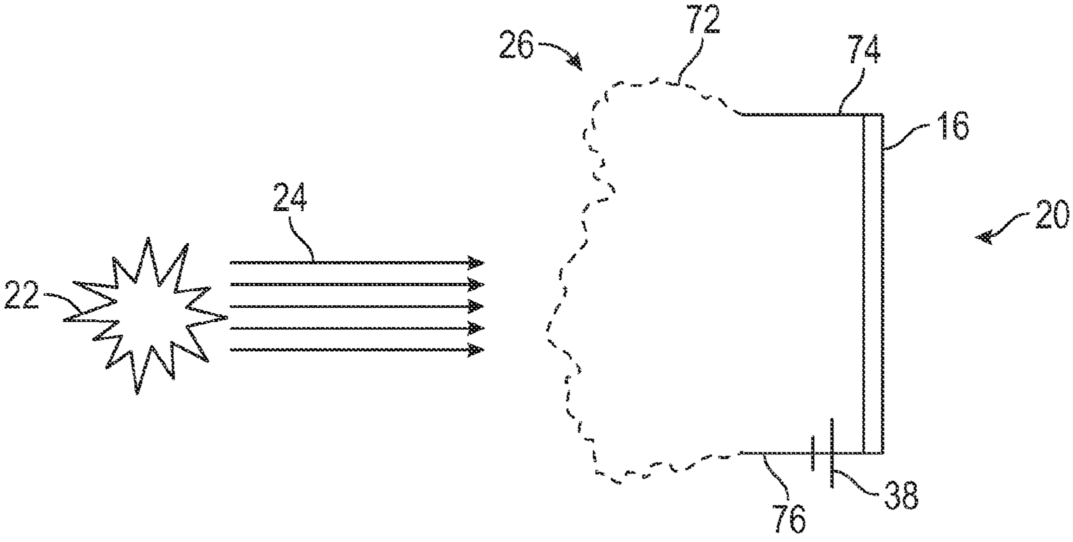
26
72
74
16
20
24
22
76
38

Prior to Shock Interaction

FIG. 8A

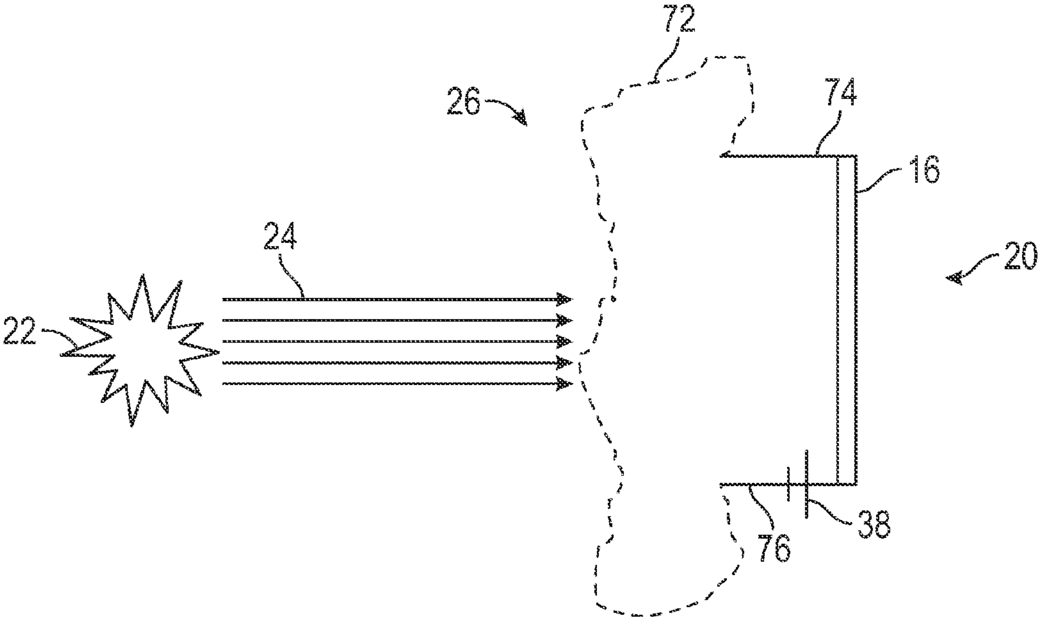
72
26
74
16
24
20
22
38
76

After Shock Interaction

FIG. 8B

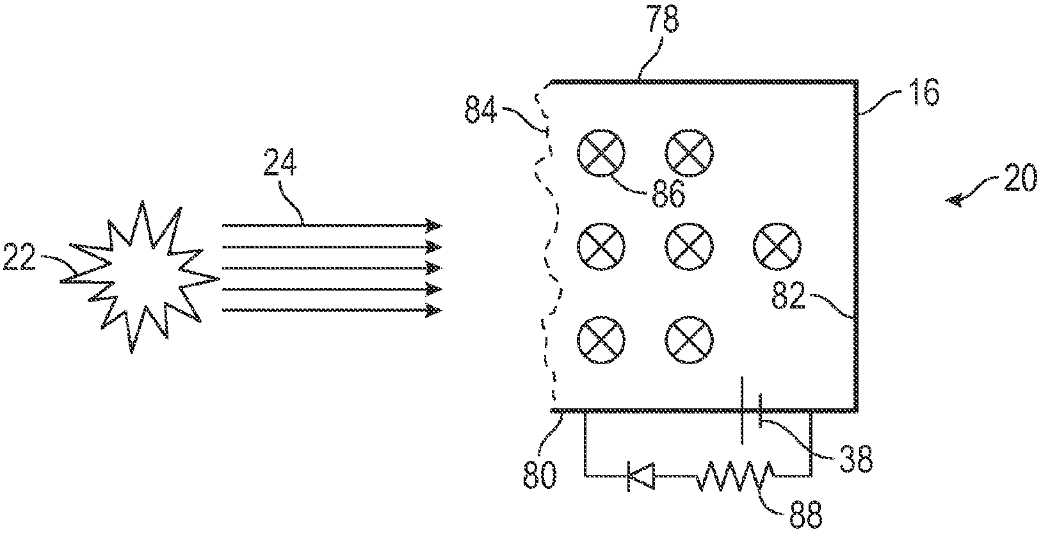
78
84
16
24
86
20
22
82
38
80
88

Prior to Shock Interaction

FIG. 9A

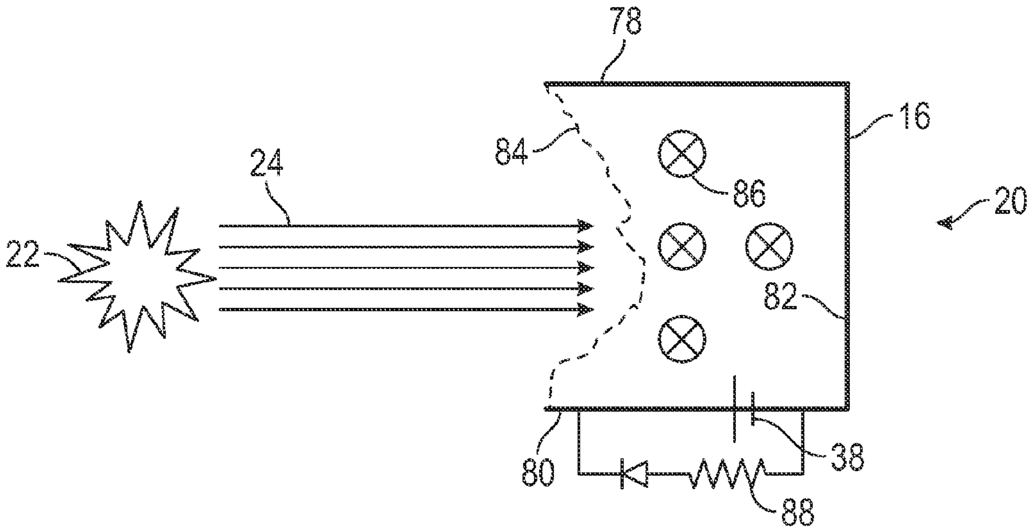
78
16
84
24
86
20
22
82
80
38
88

After Shock Interaction

FIG. 9B

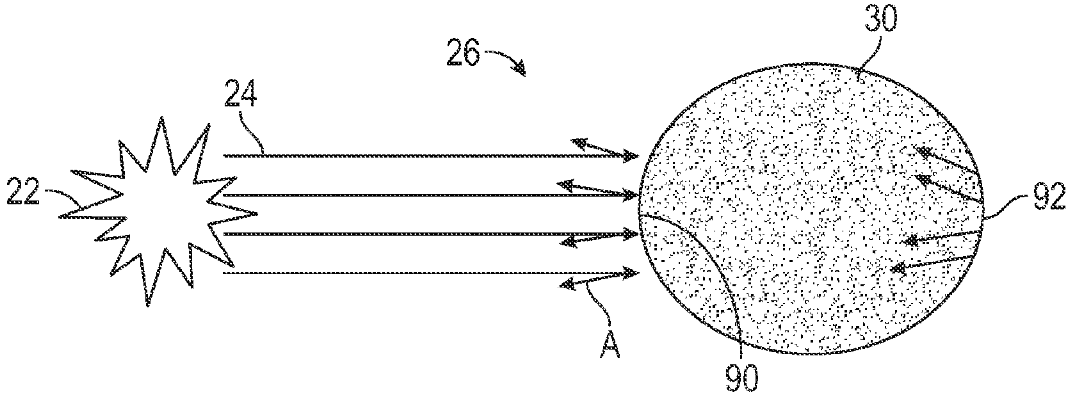
30
26
24
22
92
A
90

FIG. 10

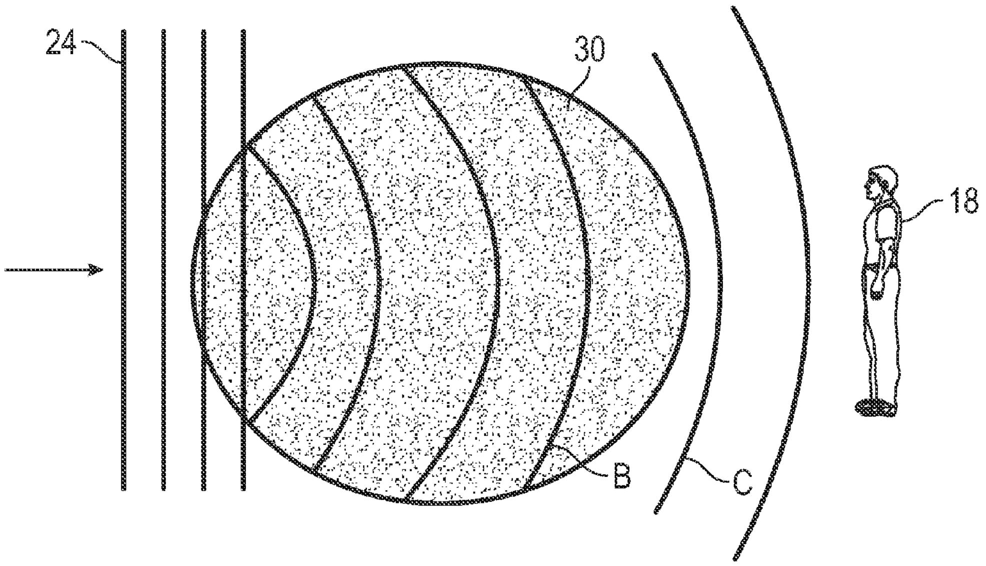

FIG. 11

METHOD AND SYSTEM FOR SHOCKWAVE ATTENUATION VIA ELECTROMAGNETIC ARC

FIELD

The disclosure relates to methods and systems for shockwave attenuation, and more particularly to methods and systems for attenuating shockwaves by rapidly heating air to interpose an intermediate medium between the shockwave and a protected region.

BACKGROUND

Explosive devices are being used increasingly in asymmetric warfare to cause damage and destruction to equipment and loss of life. The majority of the damage caused by explosive devices results from shrapnel and shockwaves. Shrapnel is material, such as metal fragments, that is propelled rapidly away from the blast zone and may damage stationary structures, vehicles, or other targets. Damage from shrapnel may be prevented by, for example, physical barriers. Shockwaves are traveling discontinuities in pressure, temperature, density, and other physical qualities through a medium, such as the ambient atmosphere. Shockwave damage is more difficult to prevent because shockwaves can traverse an intermediate medium, including physical barriers.

Damage from shockwaves may be lessened or prevented by interposing an attenuating material between the shockwave source and the object to be protected. This attenuating material typically may be designed or selected to absorb the energy from the shockwave by utilizing a porous material that distorts as the energy of the shockwave is absorbed.

U.S. Pat. No. 5,394,786 to Gettle et al. describes a shockwave attenuation device that utilizes an absorbing medium. That assembly includes porous screens that form an enclosure filled with a pressure wave attenuating medium. This attenuating medium may be an aqueous foam, gas emulsion, gel, or granular or other solid particles. However, as shown and described in the drawings of that patent, the shockwave attenuating assembly must be positioned before the explosion occurs and surround the area to be protected. For example, the assembly may be positioned on the side of a vehicle to prevent damage to the vehicle or passengers within.

A similar shockwave attenuation device is described in U.S. Patent Publication No. 2007-0006723 to Waddell, Jr. et al. That device includes a number of cells filled with an attenuating material, such as aqueous foams. However, like the device described in Gettle et al., the pressure-attenuating material and device must be positioned on a structure, surface, or person desired to be protected by the system before the explosion occurs.

One feature common among prior art shockwave attenuation systems is that they require an intermediate medium or structure that acts to attenuate the force of the shockwave by absorbing the energy of the shockwave. Although only a portion of the shockwave may pass through the medium, the energy of the shockwave is nevertheless significantly reduced by the intermediate medium. However, because these systems are structural, they must be fixed in place before a shockwave is created. Further, these shockwave attenuation systems may not protect an entire vehicle or person. For example, attenuating panels are not transparent and therefore cannot be placed over windows or used as facemasks in helmets. They also may be bulky and heavy, and therefore negatively impact the performance of a vehicle on which they are mounted.

Such prior art shockwave attenuation systems may not be effective to protect highly mobile land assets for which an incoming threat may be in the form of a ballistic shell, rocket, IED, or landmine, or waterborne assets for which an incoming threat may be in the form of a torpedo, ballistic shell, bomb or a naval mine. Therefore, a need exists for a shockwave attenuation device that is capable of dynamically interposing a medium between an explosion source and a protected asset. There is also a need for an intermediate medium that effectively attenuates the energy from a shockwave and that allows for protection of a protected asset in a marine environment.

SUMMARY

Presented is a system and method for attenuating a shockwave propagating in a first medium by detecting a shockwave-producing event, determining a direction of the shockwave relative to a protected asset, and interposing a second, transient medium, different from the first medium, between the shockwave and the protected asset such that a shockwave produced by the event contacts the second medium and is attenuated in energy thereby prior to reaching the protected asset. The second medium may be formed by rapidly heating a region of the first medium so that the second medium differs from the first medium in at least one of temperature, density and composition.

In one embodiment, a system for attenuating a shockwave propagating in a first medium may include a sensor for detecting a source of the shockwave and generating a detection signal, an arc generator in communication with the sensor and configured to receive the detection signal therefrom, and in response thereto create an electromagnetic arc to heat a selected region of the first fluid medium rapidly to create the second, transient medium, different from the first medium, interposed between the shockwave and the protected region such that the shockwave contacts the second, transient medium. The arc generator may be configured to heat the selected region by generating an electromagnetic arc, such as an electric arc or a laser or microwave arc, between the protected region and the incoming shockwave. In one embodiment, the arc generator may include a power supply for generating the arc and may provide a conducting path.

In embodiments, the arc generator may be configured to generate a focused microwave beam or a focused laser beam. In each case, the beam may rapidly heat the fluid medium in the selected region to create the second medium. In one embodiment, in which the fluid medium is atmospheric air, the focused beam rapidly heats the air in the selected region and changes its temperature, density and composition, the latter the result of the creation of free electrons.

In other embodiments, the arc generator may be adapted to develop and drive a large electric current through the fluid medium rapidly. In the embodiment in which the medium is atmospheric air, the second medium may differ from ambient air in temperature, density and composition. With these embodiments, the arc generator may be adapted to create a conducting path for the electric current. Accordingly, the arc generator may be configured to generate one or more of a laser-induced plasma channel (LIPC) from converging laser beams, ionizing tracer pellets fired along converging paths, and projectiles trailing fine electrical wires fired along converging paths. In each of these embodiments, an electric arc may be generated to travel along a conducting path created by dielectric breakdown of ionized ambient air at the selected region.

In still other embodiments, the arc generator may include a sacrificial conductor that may not rely on current travel along a path of ionized air. Rather, the arc generator may include a power supply that applies current to a conductor in the form of one or more strips or wires of conductive material. The high current flowing through the conductor from the power supply may vaporize the conductor, forming a conductive channel of vapor that may rapidly heat and ionize the air in the selected region to create a rapidly expanding second medium. An advantage of this embodiment is that the sacrificial conductor may be attached directly to the protected asset, such as a vehicle.

Such embodiments as described above may reduce the energy density of the shockwave by creating a second medium in the path of the advancing shockwave that reflects, refracts, absorbs and deflects at least a portion of the shockwave. This may result from creating a second medium that differs from the ambient medium (e.g., atmospheric air) in density, temperature and/or composition. Such differences may change the index of refraction of the wave front as it contacts the second medium, causing at least some of the shockwave to be reflected from the surface of the second medium, to diverge as the shockwave travels through the medium, and to be reflected and diverge further as the shockwave contacts the rearward boundary of the second medium. The second medium, some embodiments, acts to absorb the energy of the shockwave as the medium may be increased in temperature.

In yet other embodiments, the arc generator create the second medium by magnetic induction. The arc generator may be adapted to create channels or ionized air. When contacted by an advancing shockwave, the conducting channel may be deformed as ionized air is pushed inward. This movement does mechanical work, which removes energy from the shockwave, making it weaker. In one aspect, the channels may not be rigid, and may be pushed by the shockwave against the ambient air that the channels are displacing, which may transfer energy from the shockwave to kinetic energy of displaced air. In another aspect, parts of the conductive channels are not rigid, but the channels of ionized air pushed ahead of the magnetic flux are disrupted and broken until they form new channels through the air that heat the air. At least a portion of the shockwave energy may be transformed to ionization energy of air ions and into pressure-volume-temperature energy of expanded hot air.

In still another aspect, the conductive channels of magnetic flux are rigid may not be deformed by contact with the advancing shockwave. Instead, the flux may be compressed by contact with the shockwave, which may increase the electric current in the conductive channels. This increased current results from energy lost by the shockwave. The arc generator may be configured such that the excess current may bypass the power supply and be shunted to heat a resistive load, or charge a capacitor where it may be used later to power the arc generator.

With such embodiments, the advancing shockwave is diminished in force as the energy of the shockwave is converted either into mechanical energy, as when it deforms magnetic induction channels, or into electrical energy, as when it interacts with rigid magnetic flux channels. It is within the scope of this disclosure to provide a system and method in which combinations of the foregoing embodiments are deployed to defend a protected asset, or in which an embodiment is deployed multiple times against an incoming threat.

The features, functions, and advantages that have been discussed can be achieved independently in various embodiments of the present invention or may be combined in yet other embodiments further details of which can be seen with reference to the following description, drawings and claims.

BRIEF DESCRIPTION OF DRAWINGS

FIG. **1** is a schematic diagram of the elements of a disclosed system, according to an embodiment;

FIG. **2** is an elevational, schematic view of the disclosed system mounted on a protected asset generating an arc to encounter a shockwave from an explosion;

FIG. **3** is an elevational, schematic view of the disclosed system configured to utilize converging laser or microwave beams;

FIG. **4** is an elevational, schematic view of the disclosed system configured to utilize pellets leaving conductive trails;

FIG. **5** is an elevational, schematic view of the disclosed system configured to utilize projectiles trailing conductive wires;

FIG. **6** is an elevational, schematic view of the disclosed system configured to utilize a sacrificial conductor;

FIG. **7** is an elevational, schematic view of the system of FIG. **6** surface mounted on a protected asset;

FIGS. **8**A and **8**B are schematic views of an embodiment of the heating element of the disclosed system configured to generate an arc using flexible electrodes;

FIGS. **9**A and **9**B are schematic views of an embodiment of the arc generator of the disclosed system configured to generate an arc using relatively rigid electrodes; and

FIGS. **10** and **11** are diagrams showing some of the effects of a shockwave impacting the second, fluid medium.

DETAILED DESCRIPTION

The disclosed shockwave attenuation method and system may utilize a second, transient fluid medium that may be dynamically deployed in a first fluid medium between an explosion and a protected asset within a protected region. When contacted by an advancing shockwave produced by the explosion traveling through the first fluid medium, the second fluid medium may attenuate the energy from the shockwave through several vectors, rather than simply by absorbing the energy of the shockwave. While the following discussion describes various embodiments of the disclosed system and method as operating in a first fluid medium of atmospheric air, it is to be understood that the first fluid medium may be water, such that the disclosed system and method may be deployed in a marine environment.

As shown in FIG. **1**, in one embodiment, the system for attenuating a shockwave via electromagnetic arc, generally designated **10**, may include a sensor **12** including or incorporating a computer **14**, and an arc generator, generally designated **16**. The sensor **12** and computer **14** may be mounted on or incorporated in a protected asset, such as a vehicle **18**, which may be a military vehicle as shown in FIG. **2**, or they may be physically separate from the vehicle. Although FIG. **2** shows vehicle **18** as a military vehicle, it is to be understood that the depicted vehicle **18** also may represent other types of land vehicles, such as automobiles, armored vehicles such as tanks and armored personnel carriers; fixed structures such as encampments, bunkers, revetments, buildings or portions of buildings such as balconies; marine vessels such as surface ships, submarines, a hovercraft or air-cushion vehicles, offshore platforms, and other structures that operate in, under or adjacent a body of water; aircraft including fixed wing aircraft, vertical takeoff and landing (VTOL) craft and helicopters; and people and animals. Each of the foregoing would be placed in a protected region **20**.

Referring to FIG. **1**, the sensor **12** may be selected to provide measurements that enable the computer **14** to estimate the location and time of an explosion **22**, either before or after it occurs, and direct the arc generator **16** to respond. In one embodiment, the sensor **12** is configured to detect an explosion **22** that has occurred, but before the shockwave **24** caused by the explosion, traveling through a first fluid medium **26**, which in this embodiment is atmospheric air at ambient temperature and pressure, reaches the protected region **20**. This type of sensor **12** may be configured to detect any one or more of the electromagnetic signals emitted during chemical detonations, including microwave bursts, flashes of infrared radiation, flashes of visible light, flashes of ultraviolet light, and x-ray bursts. In an embodiment a sensor **12** may be configured to detect two or more of the foregoing types of electromagnetic radiation, which may result in the sensor detecting fewer false positives. In a variation of such an embodiment, the sensor **12** may be in the form of two discrete sensors, each configured to detect a different type of electromagnetic radiation. In an embodiment in which the sensor **12** is selected for use in a marine environment (e.g., to detect underwater explosions from naval mines or torpedoes), the sensor **12** may be configured to detect electromagnetic radiation at wavelengths for which water is substantially transparent, such as visible light near-ultraviolet light and near-infrared light.

In some embodiments, once the sensor **12** has detected an explosion **22**, the sensor may be configured to estimate one or more of the magnitude, distance, and azimuthal position of the explosion and provide the estimates to the computer **14** or to the arc generator **16**. In some embodiments, the computer **14** may measure the signature of the explosion **22** and compare it with stored known explosion signatures of a plurality of different known explosive devices. The computer **14** may then determine a probability that the explosion is from one of the known explosive devices. The computer **14** may then calculate an estimated probability distribution function (p.d.f.) of explosion magnitudes and locations relative to the protected asset **18** or protected region **20** based on at least one of stored data about the type of explosive device, location of the explosion, and the shape, relative orientation and relative motion of the protected asset, and make a determination to counter the explosion **22**, including determining when and where to activate the arc generator **16**, or not counter the explosion, based on one of stored data, models of vulnerability of the protected asset to shockwaves, and data from at least one of data and models of performance of the arc generator **16** with respect to attenuating shockwaves from at least one of an estimated explosion magnitude and location. Such calculations and estimates may prevent the deployment of the arc generator **16** in the event that the explosion **22** is too far away or too weak to generate a shockwave **24** that damages the protected region **20** significantly.

In another embodiment, the sensor **12** may be configured to detect an incoming threat **28** containing an explosive device or devices, such as a ballistic shell, bomb, torpedo, depth charge, naval mine or bomb-laden surface vessel. In such an embodiment, the sensor **12** may be configured to use radar, visible or infrared light, passive or active acoustic sensors, or other threat-detection method known to those skilled in the art, as well as trajectory tracking and prediction methods known to those skilled in the art.

In yet another embodiment, the sensor **12** may be configured to detect both the incoming threat **28** and the explosion **22** from the threat. In one embodiment, two systems **10** may be deployed on a protected asset **18** in which one system is configured to detect an incoming threat **28** and the other system is configured to detect an explosion **22**. In yet another variation of such an embodiment, the sensor **12** may be in the form of two discrete sensors: one configured to detect an incoming threat **28**, and the other configured to detect an explosion **22** from that incoming threat.

In some embodiments, the computer **14** may receive measurements from the sensor **12**, estimate where and when an incoming threat **28** will detonate, or has already detonated, and directs the system **10** to deploy. When used with a sensor **12** that may detect an explosion that already has occurred, the computer **14** may be configured to receive information from the sensor pertaining to one or more of the direction, location, time, distance and magnitude of the explosion **22**, the computer determines when and where to activate the arc generator **16**.

In an embodiment wherein the sensor **12** is configured to detect an incoming threat **28** before explosion **22** has occurred, the computer **14** may be configured to compare the signature of the incoming threat **28** with stored known signatures of various threats (e.g., particular missiles). The computer **14** then estimates how probable each type of threat is, and, based on the stored data about the type of warhead for each threat, the measured motion of the incoming object (with associated uncertainty), and the shape, trajectory, orientation, speed and motion of the protected asset **18** in the protected region **20**, the computer estimates a probability distribution function (p.d.f.) of explosion magnitudes and locations relative to the protected asset.

Based on data or models of vulnerability of the protected asset **18** to shockwaves **24** of various magnitudes from various directions (including crew injuries likely to result from shockwaves) stored in a database (either locally or available over a network), together with data or models of what the arc generator **16** can do to attenuate shockwaves in what positions and in what time interval, the computer **14** may then form a firing plan to counter the threat at minimum cost. Cost may include not only the cost to operate the arc generator **16**, but also the probable cost of damage from the attenuated shockwave. In cases where the probable explosion yield is small and the probable distance of the explosion **22** from the protected region is large, the lowest cost plan may be not to deploy the system **10**.

As shown in the embodiments of FIGS. **1** and **2**, the arc generator **16** may be configured to create a transient second medium **30** between the advancing shockwave **24** resulting from the explosion **22** and the protected region **20**, which may contain a protected asset **18**. As shown in FIG. **2**, generally speaking, the sensor **12** may generate a detection signal in response to detecting at least one of the explosion **22** or explosive device such as incoming threat **28**. The detection signal is received by arc generator **16**, and in response, may create an electromagnetic arc through the first medium **26** which rapidly heats the medium to change one or more of its composition and temperature to create the second, transient medium **30**, different from the first medium **26**. The location of the second medium **30** may be in a selected region calculated by the computer **14** of the sensor **12** to be interposed between the shockwave **24**, or predicted shockwave path, and the protected region **20**. Thus, the shockwave **24** contacts the second, transient medium **30** and is attenuated in energy density before it reaches the protected region **20**. The arc generator **16** shown schematically in FIG. **1** may take the form of multiple, discrete arc generators, each connected to and controlled by the sensor **12**.

In one embodiment, shown schematically in FIG. **1**, the arc generator **16** is configured to direct a beam **34** of electromagnetic energy to focus in a selected region where the second medium **30** may be created. The beam **34** may be in the form of a relatively powerful microwave beam or laser beam. The electric field of the intense beams **34** may cause dielectric breakdown of the ambient air at the focal point of the focused beams, in which the electrons are separated from the molecules in the air. The free electrons are accelerated by the electric field and strike other molecules to knock other electrons loose, creating a cascade of electrons and ions.

As the density of free charge carriers in the air increases, the air becomes opaque to the incoming electromagnetic beams and rapidly absorbs energy from the beams as heat, which raises the temperature of the air as well as its density and composition, thus creating the second medium **30** to intercept the advancing shockwave **24**. The composition changes of the second fluid medium **30** may include adding free electrons, which have a relatively low molar mass, ionization of molecules so that they interact more strongly and therefore propagate shockwaves at higher speeds, and breaking diatomic molecules such as molecular oxygen into single atoms, which reduces the average molar mass.

In some embodiments, the beam **34** may be a microwave beam. The arc generator **16** may include a vacuum tube amplifier (e.g., a magnetron) and focused by a static focusing device (e.g., a dish antenna or a Fresnel plate), or produced and focused by other means of combinations apparent to those skilled in the art. In other embodiments, the beam **34** is a laser beam. The laser beam may be a single beam focused to a point by optics, or multiple beams converging to a common point. In either case, the beam **34** may create single or multiple arcs **32** (FIG. **2**). In some embodiments, the arc generator **16** may use many converging beams **34** to create one arc **32**. In other embodiments one beam **34** may be used multiple times in quick succession to create many arcs **32**. The beams may be fixed or steerable.

As shown in FIG. **3**, in another embodiment of the system **10**A the arc generator **16** (FIG. **1**) may be configured to generate an electric arc **36** through a volume of ambient air **26** to create the second medium **30**. Electrical resistance to the current heats the air, changing its temperature, density and composition to create the second medium **30** to attenuate the shockwave **24**. The change in composition may include adding free electrons, which have very low molar mass, ionization of molecules so they interact at longer distance and therefore propagate shockwaves at higher speed, and breaking diatomic molecules like oxygen into single atoms, which reduces the average molar mass of the fluid in the second medium **30**.

The arc generator **16** (FIG. **1**) of the system **10**A of this embodiment may include a power supply **38** configured to initiate a large electric current quickly—typically on the order of a few milliseconds or less. Such a power supply **38** may include a capacitor, a superconducting storage coil, and an explosive flux compression generator. The first two examples may require fast-acting, high-current switchgear, such as a gas-insulated switch, to turn on a large current quickly (for the capacitor) or to divert it quickly from a shunt to the electric arc (for the storage coil).

In one aspect of this embodiment, the arc generator **16** creates an electrically conductive path from the power supply **38** to the selected region between the shockwave **24** and the protected region **20** to establish the arc, and back to the power supply. In one embodiment, shown in FIG. **3**, the arc generator **16** creates a laser-induced plasma channel (LIPC). The arc generator **16** fires two high-intensity laser beams **40**, **42** along converging paths. The laser wavelengths and intensities are selected to ionize the air along their paths, thereby forming the plasma channels. The power supply **38** applies voltage across the terminals at the bases of the two channels **40**, **42**, and the voltage is sufficient to form the arc **36** through the air **26** where the beams are close to each other. Current flows through the channels of the LIPC and induces a magnetic field. The current interacts with the magnetic field and produces an outward force (J×B). This force acts to widen the area enclosed by the current loop.

As shown in FIG. **4**, in another embodiment of the system **10**B the arc generator **16** (FIG. **1**) is a gun, which may have at least two or multiple barrels **44**, **46** that fire two pellets **48**, **50** along converging paths. The pellets **48**, **50** are configured to create trails **52**, **54** of ions, for example, by burning silver iodide, in the manner of tracer bullets. The two trails **52**, **54** form conductive channels through the air (the first fluid medium), and the power supply **38** applies voltage across terminals at the bases of the two conductive channels, preferably the barrels **44**, **46**. The voltage is sufficient to form an arc **56** through the air **26** between the two trails **52**, **54** (i.e., the conductive channels) where they are at their closest. The arc **56** heats the air **26** to create the second transient fluid medium **30**. In other embodiments, the pellets **48**, **50** may be selected to melt or burn up completely, thereby avoiding creating a hazard for by standers.

As shown in FIG. **5**, in another embodiment of the system **10**C the arc generator **16** (FIG. **1**) is a projectile launcher configured to launch projectiles **58**, **60** trailing conductive wires **62**, **64** much like a Taser®. The power supply **38** applies voltage across the two wires sufficient to form an arc **66** through the air **26** where the paths of the projectiles **58**, **60** are closest to each other. The arc **66** heats the air **26**, forming the second fluid medium **30** in a selected region between the advancing shockwave **24** and the protected asset **18** in the protected region **20**. As electrical current flows through the wires **62**, **64**, which form conductive channels, the current further heats and ionizes the air **26**. In some embodiments, the current is sufficient to vaporize the wires **62**, **64**, thereby heating additional nearby air **26**.

As shown in FIGS. **6** and **7**, in another embodiment the system **10**D includes a arc generator **16** (FIG. **1**) in the form of one or more sacrificial conductors **68** connected to the power supply **38**. The heating mechanism of this embodiment does not rely on dielectric breakdown of the air **26** to create a current channel. The sacrificial conductor **68** may be in the form of strips or wires made of electrically conductive material. In some embodiments, the conductor **68** may be mounted on the protected asset **18**.

To deploy the system **10**D, the arc generator **16** actuates the power supply **38** to supply current to one of more of the conductors **68**, first vaporizing it, then ionizing it to form a conductive channel of vapor **70**. Due to both the resultant magnetic field (J×B) force and the expansion of the hot vapor created by vaporizing the conductor, the ionized vapor moves outward from its initial position and heats and ionizes the nearby air to create the second fluid medium **30**.

In one embodiment, shown in FIG. **7**, the sacrificial conductor **68** may be attached to a protected asset **18**, which may be the land vehicle shown. The conductor **68** may be attached to a wider insulating strip **72** that in turn is mounted on the asset **18**. With such a placement of the conductor **68**, the conductor **68** may be protected from routine vehicle operations, and may be further protected by overlaying it with a coat of paint. When energized by the power supply **38** (FIG. **6**), the conductor **68**, which may be a metallic strip, vaporizes, and the heated and ionized air may expand from the vehicle surface and move away from the vehicle to create the second medium **30** in the selected region.

In yet another embodiment, the sacrificial conductor **68** may be made partly or entirely of lithium. Lithium has a very low molecular weight, and consequently a higher shock speed and lower effective index of refraction than other metallic vapors. Further, it disperses into nearby air **26** more quickly, which helps to heat the air more rapidly.

In other embodiments, the arc generator **16** may be configured to attenuate the advancing shockwave by magnetic induction. As shown in FIG. **8**A, the arc generator **16** may be powered by the power supply **38** to create an electric arc **72** that forms a conducting channel between two electrodes **74**, **76** that are not rigid, such as thin rod antennae. The conducting channel **72** interacts with its own magnetic field to produce an outward force F=J×B (where F is the outward force, J is the current density and B is the magnetic field flux).

When the shockwave **24** created by an explosion **22** reaches the arc **72**, as shown in FIG. **8**B, the shockwave deforms the arc in directions perpendicular to the conductive channels. Specifically, the shockwave **24** pushes the ionized air created by or making up the arc **72** inward, and the current tends to flow where the air **26** is ionized. This movement by the shockwave **24** may do mechanical work against the magnetic field force F and remove energy from the shockwave, making it weaker. Although not shown, the movement by the shockwave **24** also may deform the electrodes **74**, **76**, which also represents physical work that drains energy from the shockwave **24**.

In the embodiment shown in FIGS. **8**A and **8**B, because the electrodes **74**, **76** may be relatively non-rigid, these parts and the conductive channel **72** push against the air **26** they are displacing. This may transform energy from the shockwave **25** into kinetic energy of displaced air. In some cases it may be possible for the conductive channels **72** and electrodes **74**, **76** to push the air so fast that they may create new shockwaves moving away from the protected region **20**.

In a variation, the parts of the conductive channels **72** and electrodes **74**, **76** may not be rigid, but the channels of ionized air pushed ahead of the magnetic flux are disrupted (broken) when they flow rapidly through the air **26**. Each time a channel **72** breaks, the electric current briefly stops. When it stops, the trapped magnetic flux creates an electromotive force strong enough to ionize a new channel through the air **26** and then heat that air. This may transform energy from the shockwave **24** into ionization energy of air ions and into pressure-volume-temperature energy of expanded hot air.

In the embodiment shown in FIGS. **9**A and **9**B, the arc generator **16** may be configured to have rigid electrodes **78**, **80** joined by a rear wall **82** to form three sides of a box. An arc **84** may be formed by the power supply **38** of the arc generator **16** between electrodes **78**, **80**. When the arc **84** is contacted by the shockwave **24**, the magnetic flux (represented by the crossed circles **86**) is deflected inward but movement is constrained by the electrodes **78**, **80** and wall **82**. Instead, the flux **86** is compressed and increases the electric current in the electrodes **78**, **80**. In one embodiment, the power supply **16** is configured so that this additional electric current bypasses the power supply and instead is directed to a resistive load **88**. Resistive load **88** also may take the form of a capacitor, which may receive a charge from the additional electric current. Thus, with the embodiments of FIGS. **8** and **9**, magnetic induction is used to remove energy from the advancing shockwave by the additional mechanism of converting the shockwave energy to mechanical energy or converting the shockwave energy to current.

A system **10** may be comprised of multiple copies of each embodiment. In an embodiment, a system **10** may include a single sensor **12** connected to and controlling multiple, discrete arc generators **16**, each mounted on the protected asset **18**. In some applications, an embodiment may produce a relatively narrow, substantially linear arc of hot, ionized air. Multiple copies of each embodiment may be used to increase the frontal area as desired. Alternatively, a single copy of an embodiment may be used multiple times in rapid succession, producing multiple arcs that collectively cover the desired protected region **20**. Further, these embodiments may be combined. For example, a system **10** may utilize ionizing tracer pellets to protect an asset **18** from explosions **22** relatively far away, and employ sacrificial conductors to protect the same asset from explosions at short range.

With each of the embodiments discussed, the system **10** is deployed to attenuate the energy of an advancing shockwave **24** form an explosion **22** by creating a second fluid medium **30** that differs from the first fluid medium **26**, which may be ambient air, positioned so that it interacts with the shockwave. As shown in FIG. **10**, as the shockwave contacts the interface **90** between the first fluid medium **26** and the second fluid medium **30**, the difference in refractive index reflects a fraction of the incoming energy toward the explosion **22**, as indicated by arrows A. This partial reflection occurs a second time as the shockwave passes through the second fluid medium **30** and contacts the interface **92** between the second medium and the ambient **26** as it exits the second medium. All gradients or discontinuities in the medium provide a reflection point for the incoming shockwave **24**. For example, if the second medium **30** is non-uniform, reflection will occur at each of many places within the medium.

As shown in FIG. **11**, shockwaves **24** obey Fermat's theory of least time and therefore an effective refractive index for the shockwave can be defined that is inversely proportional to the shock speed. The properties or composition of the second medium **30** are chosen such that the effective refractive index of the second medium **30** differs from the first medium **26** in at least one of temperature, molecular weight and composition. As the shockwave passes into or out of the second medium **30**, the difference in effective refractive index refracts the wave, as shown by lines B, diverting it and defocusing it away from the protected asset **18**. In the disclosed embodiments, the second medium **30** is created such that the shockwave travels faster in the second medium **30** than in the first medium **26**, so the refractive index of the second medium is less than that of the first medium. Further, the second medium is created to have a convex shape and therefore acts as a divergent lens, so that the energy of the shockwave **24** spreads out, as shown by lines C, so its intensity drops as it approaches the protected asset **18**.

In addition, the second medium **30** may absorb some shock energy as the shock travels through it. Factors contributing to the absorption of energy include energy retained in the molecules of the second medium itself (e.g., enhanced rotational energy, excited molecular bonds, excited electrons, molecular decomposition, and ionization) and shock energy converted to electromagnetic energy through blackbody emission from hot particles or photon emission from de-exciting various excited states.

A further mechanism for attenuating the energy density of the shockwave **24** is momentum exchange. If the second medium **30** is moving relative to the first medium **26**, then it will exchange momentum with the shockwave **24**. The result is a combination of reflection, slowing, and redirection of the shockwave. Any or all of the foregoing mechanisms may operate in a given embodiment. The composition, temperature, speed and location of the second medium **30** may be chosen or created to create any one or all of the aforementioned mechanisms.

While the method and forms of apparatus disclosed herein constitute preferred aspects of the disclosed shockwave attenuation apparatus and method, other methods and forms of apparatus may be employed without departing from the scope of the invention.

What is claimed is:

1. A shockwave attenuation system, comprising:

a sensor for generating a detection signal based on at least one of detecting an explosion capable of producing a shockwave traveling through a first fluid medium to a protected region, and estimating a location and time of the explosion, and detecting an explosive device and estimating a location and time of an explosion from the explosive device that is capable of producing the shockwave traveling through the first fluid medium; and

an arc generator in communication with the sensor for receiving the detection signal therefrom, and in response thereto heat a selected region of the first fluid medium rapidly to create a second, transient medium, different from the first medium, interposed between the shockwave and the protected region such that the shockwave contacts the second, transient medium and is attenuated in energy density before it reaches a protected asset in the protected region.

2. The system of claim **1**, wherein the second medium differs from the first medium in at least one of temperature, density and composition.

3. The system of claim **1**, wherein the arc generator heats the region of the first fluid medium to create the second medium by using one of an electric arc, a laser-induced arc and a microwave-induced arc.

4. The system of claim **1**, wherein the arc generator includes a power supply, and the arc generator rapidly creates the second medium by creating an electrically conductive path from the power supply to the selected region and back to the power supply.

5. The system of claim **4**, wherein the arc generator creates the electrically conductive path utilizing at least one of high intensity laser pulses to form a laser-induced plasma channel (LIPC), pellets that leave a conductive trail of ions, sacrificial conductors, projectiles trailing electrical wires fired along converging paths, magnetic induction utilizing flexible channels of ionized air, and magnetic induction utilizing substantially rigid conductors.

6. The system of claim **1**, wherein the protected region includes at least one protected asset, and at least one of the sensor and the arc generator is mounted on the protected asset.

7. The system of claim **1**, wherein the sensor detects at least two bands of electromagnetic radiation generated by the explosion.

8. The system of claim **1**, wherein the sensor detects at least one of a shape, trajectory and speed of an incoming threat containing the explosive device, and to calculate a signature of the incoming threat, the sensor also using the signature to determine likely explosion characteristics of the explosive device.

9. The system of claim **8**, wherein the explosion characteristics include at least one of a location of the explosion, a time of the explosion, and a magnitude of the explosion.

10. The system of claim **9**, wherein the sensor uses the explosion characteristics to calculate a location of the selected region.

11. The system of claim **6**, wherein the protected asset is one of a surface vessel, a submarine vessel, an offshore platform, a land vehicle, a land structure, and a human, and wherein the sensor determines the selected region based on one or more predetermined vulnerabilities of the protected asset.

12. The system of claim **1**, further comprising multiple arc generators, each connected to the sensor.

13. The system of claim **12**, wherein the multiple arc generators are adapted to be mounted on a protected asset.

14. The system of claim **1**, wherein the first fluid medium is ambient air, and the arc generator utilizes at least one of electric, microwave and laser energy to produce at least one of relatively hot and ionized air to form the second, transient medium in the selected region, such that the shockwave contacts the second, transient medium and is attenuated in energy density by at least one of reflection, refraction, absorption, momentum transfer and magnetic induction.

15. A method of attenuating a shockwave, the method comprising:

detecting with a sensor at least one of an explosion capable of producing a shockwave traveling through a first fluid medium to a protected region, and an explosive device:

estimating a location and time of the explosion of the at least one of an explosion from the explosive device and the explosive device that is capable of producing the shockwave;

generating a detection signal by the sensor in response to detecting at least one of the explosion and the explosive device;

heating a selected region of the first fluid medium rapidly to create a second, transient medium, different from the first medium, by an arc generator in communication with the sensor, in response to the detection signal, the second medium being interposed between the shockwave and the protected region such that the shockwave contacts the second, transient medium and is attenuated in energy density before it reaches the protected region.

16. The method of claim **15**, wherein the second, transient medium attenuates the energy density of the shockwave by at least one of reflection, refraction, absorption, momentum transfer and magnetic induction.

17. The method of claim **15**, wherein detecting includes detecting, with one or more sensors, at least two bands of electromagnetic radiation from the explosion.

18. The method of claim **15**, wherein heating includes rapidly creating the second medium by creating an electrically conductive path from a power supply to the selected region and back to the power supply.

19. The method of claim **18**, wherein creating an electrically conductive path utilizes at least one of high intensity laser pulses to form a laser-induced plasma channel (LIPC), pellets that leave a conductive trail of ions, sacrificial conductors, projectiles trailing electrical wires fired along converging paths, magnetic induction utilizing flexible channels of ionized air, and magnetic induction utilizing substantially rigid channels of ionized air.

20. A method of attenuating a shockwave, the method comprising:

detecting with a sensor at least one of an explosive device and an explosion from the explosive device;

estimating a location and time of the explosion of the at least one of an explosion from the explosive device and the explosive device and predicting an explosion therefrom that is capable of producing the shockwave;

calculating with a computer a firing plan based upon at least one of data and models of vulnerability of a protected asset, and at least one of data and models of effectiveness of actuating an arc generator to heat a selected region adjacent the protected region; and

if the firing plan determines that it is cost effective to execute the firing plan in view of a cost to operate an arc generator and probable cost of damage from a shockwave from the explosion, actuating the arc generator to heat a selected region of a first fluid medium rapidly to create a second, transient medium, different from the first medium, the second medium being interposed between the shockwave and the protected region such that the shockwave contacts the second, transient medium and is attenuated in energy density before it reaches the protected asset.

21. The method of claim **20**, wherein detecting includes

measuring a signature of an incoming threat carrying the explosive device;

comparing the signature with known signatures of a plurality of different threats;

determining a probability the incoming threat is one of the plurality of different threats; and

calculating includes estimating a probability distribution function of explosion magnitudes and locations relative to the protected asset based on at least one of stored data about the type of explosive device, measured motion of the incoming threat, and a shape, relative orientation and relative motion of the protected asset; and making a determination to counter the incoming threat or not counter the incoming threat, based on one of stored data and models of vulnerability of the protected asset to shockwaves, and data from at least one of data and models of performance of the arc generator with respect to attenuating shockwaves from the estimated explosion magnitudes and locations.

22. The method of claim **20**, wherein detecting includes

measuring a signature of the explosion from the explosive device;

comparing the signature with stored signatures of a plurality of known different explosive devices;

determining a probability the explosion is from one of the plurality of known different explosive devices; and

calculating includes estimating a probability distribution function of explosion magnitudes and locations relative to the protected asset based on at least one of stored data about the type of explosion, the location of the explosion and the shape, relative orientation and relative motion of the protected asset; and making a determination to counter the explosion or not counter the explosion, based on one of stored data and models of vulnerability of the protected asset to shockwaves from the explosion, and data from at least one of data and models of performance of the arc generator with respect to attenuating shockwaves from at least one of an estimated explosion magnitude and location.

23. The method of claim **20**, wherein calculating includes calculating with a computer a firing plan based upon at least one of data and models of vulnerability of the protected asset within a protected region.

* * * * *

US010006446B2

(12) United States Patent
Purvis

(10) **Patent No.: US 10,006,446 B2**
(45) **Date of Patent: Jun. 26, 2018**

(54) **ELECTROMAGNETIC SEGMENTED-CAPACITOR PROPULSION SYSTEM**

(71) Applicant: **James Wayne Purvis**, Albuquerque, NM (US)

(72) Inventor: **James Wayne Purvis**, Albuquerque, NM (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 766 days.

(21) Appl. No.: **14/591,585**

(22) Filed: **Jan. 7, 2015**

(65) **Prior Publication Data**

US 2016/0197543 A1 Jul. 7, 2016

(51) **Int. Cl.**
F03H 99/00 (2009.01)
H02K 53/00 (2006.01)

(52) **U.S. Cl.**
CPC ***F03H 99/00*** (2013.01); ***H02K 53/00*** (2013.01)

(58) **Field of Classification Search**
CPC F03H 1/0037; F03H 3/00; F03H 99/00; H01F 7/066
USPC 335/223–226
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2,949,550	A	8/1960	Brown
2,958,790	A	11/1960	Bahnson
3,018,394	A	1/1962	Brown
3,187,206	A	6/1965	Brown
3,223,038	A	12/1965	Bahnson
3,227,901	A	1/1966	Bahnson
3,626,605	A	12/1971	Wallace
4,891,600	A	1/1990	Cox
5,142,861	A	9/1992	Schlicher et al.
5,182,958	A	2/1993	Black
5,197,279	A	3/1993	Taylor
5,207,760	A	5/1993	Dailey et al.
5,280,864	A	1/1994	Woodward
5,377,936	A	1/1995	Mitchell
2,886,976	A	5/1995	Dean
6,098,924	A	8/2000	Woodward
6,317,310	B1	11/2001	Campbell
6,347,766	B1	2/2002	Woodward
6,411,493	B2	6/2002	Campbell

(Continued)

FOREIGN PATENT DOCUMENTS

GB	2229865	10/1990
GB	2334761	9/1999

(Continued)

OTHER PUBLICATIONS

Powell "Did NASA Validate an Impossible Space Drive In a Word, No" 2015.*

(Continued)

Primary Examiner — Gerald L Sung
Assistant Examiner — William Breazeal

(57) **ABSTRACT**

The present invention is a spacecraft electromagnetic propulsion system which provides thrust without expelling propellant. The system includes a multi-element capacitor with rotating segmented cathodes, and electromagnetic coils for generating magnetic fields. Interactions between the electromagnetic fields and the rotating charged capacitor elements produce unidirectional forces, while reaction momentum is carried away by the Poynting Vector fields in conformity with the currently understood principles of modern physics.

11 Claims, 5 Drawing Sheets

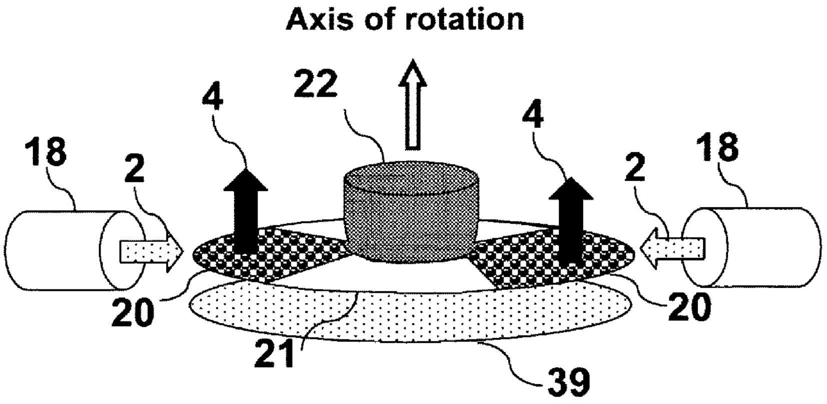

(56) References Cited

U.S. PATENT DOCUMENTS

6,492,784	B1	12/2002	Serrano	
6,775,123	B1	8/2004	Campbell	
7,913,954	B2	3/2011	Levin	
8,066,226	B2	11/2011	Fiala et al.	
8,575,790	B1	11/2013	Ogilvie	
2014/0013724	A1	1/2014	Fetta	
2014/0265688	A1*	9/2014	Post	H02K 7/09 310/90.5

FOREIGN PATENT DOCUMENTS

GB	2399601	9/2004
GB	2537119	10/2016

OTHER PUBLICATIONS

Stafford, Thomas P., "America at the Threshold: Report of the Synthesis Group on America's Space Ezploration Initiative", 1991, NASA.

NASA, "Breakthrough Propellantless Propulsion, Workshop Proceedings", Glenn Research Center, 1997.

Shawyer, Roger, "A Theory of Microwave Propulsion for Spacecraft (Theory Paper v9.3)", New Scientist, Sep. 2006.

Hambling, David, "EmDrive: China's radical new space drive", Technology, Feb. 6, 2013.

Hambling, David, "NASA validates 'impossible' space drive", http://www.wired.co.uk, Jul. 31, 2014.

Feynman Richard P., et.al., "The Feynman Lectures on Physics", vol. II, 1997.

Tipler, Paul A., "Physics for Scientists and Engineers", Fourth Edition, 1999.

Cullwick, E..G, Electromagnetic Momentum and Newton's Third Law, Nature, 170-425, 1952.

Maimon, Ron, http://physics.stackexchange.com/questions/43269/violation-of-newtons-3rd-law-and-momentum-conservation, Nov. 2, 2012.

Stafford, Thomas P., "America at the Threshold: Report of the Synthesis Group on America's Space Exploration Initiative",1991, pp. 64-68,NASA.

NASA, "Breakthrough Propulsion Project, Workshop Proceedings", NASA TM/1998-208400, pp. 4-9,1998.

Resnick, D., and Halliday, R., "Physics, Parts I and II", pp. 986, 995-998., John Wiley & Sons, 1967.

Sebens, C.T., "Forces on Fields", arXiv:1707.04198 [physics.hist-ph], pp. 1-21, 2017.

Cullwick, E.G., Nature, vol. 470, No. 125, 1952.

White, Harold et.al.,AIAA Journal of Propulsion and Power, vol. 33, No. 4, pp. 830-841.

Feynman, R.P., et.al.,The Feynman Lectures on Physics, vol. II, Ch. 26-2, 1997.

* cited by examiner

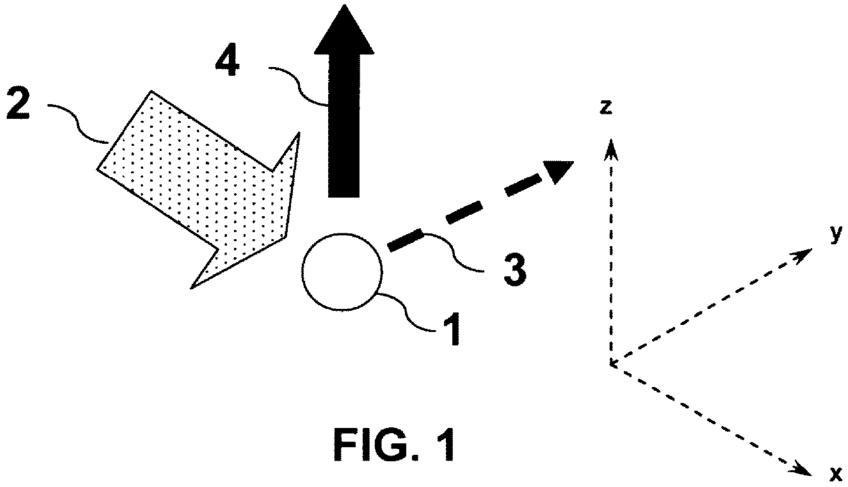

FIG. 1

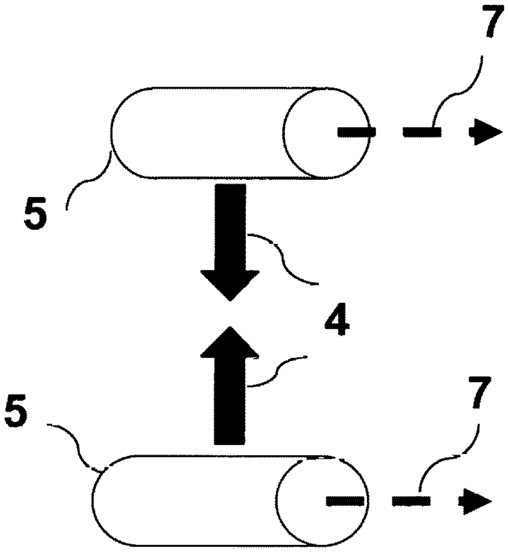

FIG. 2

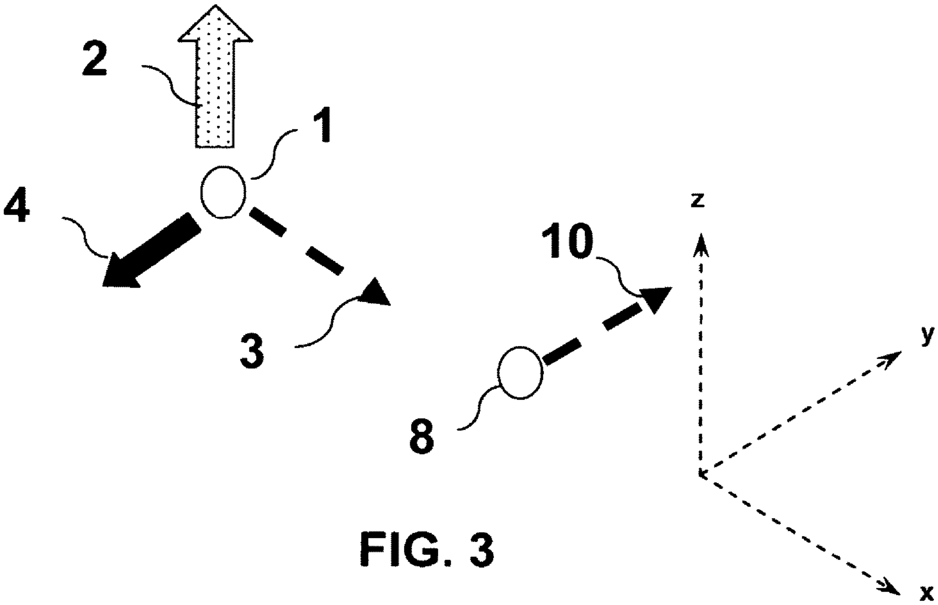
2
1
4
3
10
8
z
y
x

FIG. 3

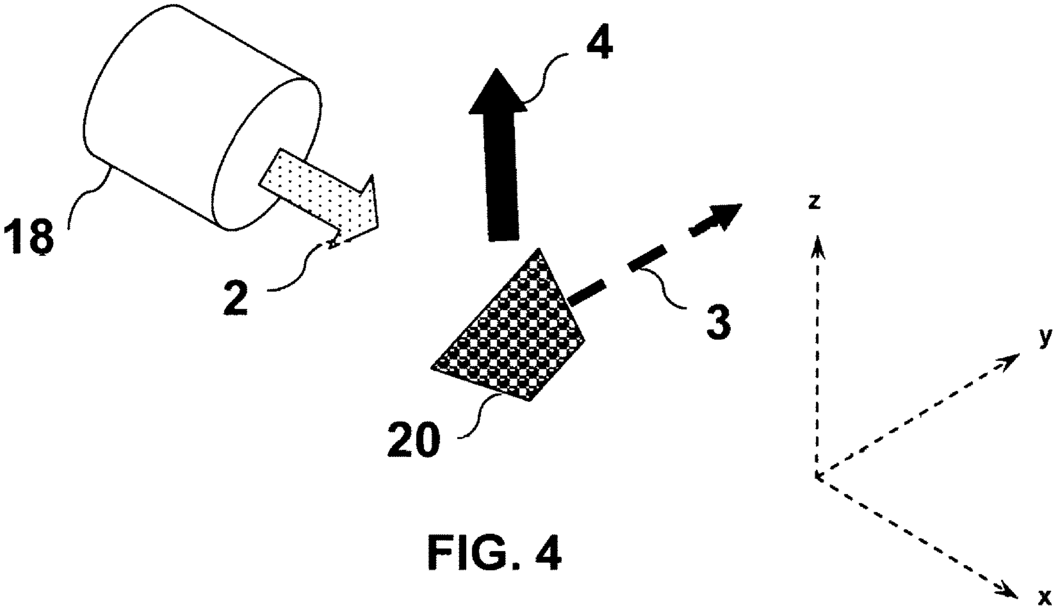
18
2
4
3
20
z
y
x

FIG. 4

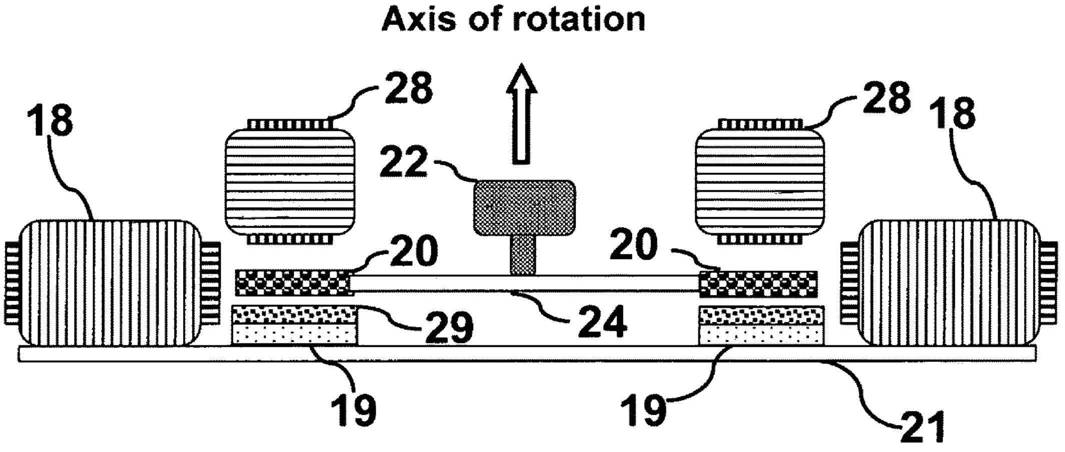
Axis of rotation
28
28
18
18
22
20
20
29
24
19
19
21

FIG. 5

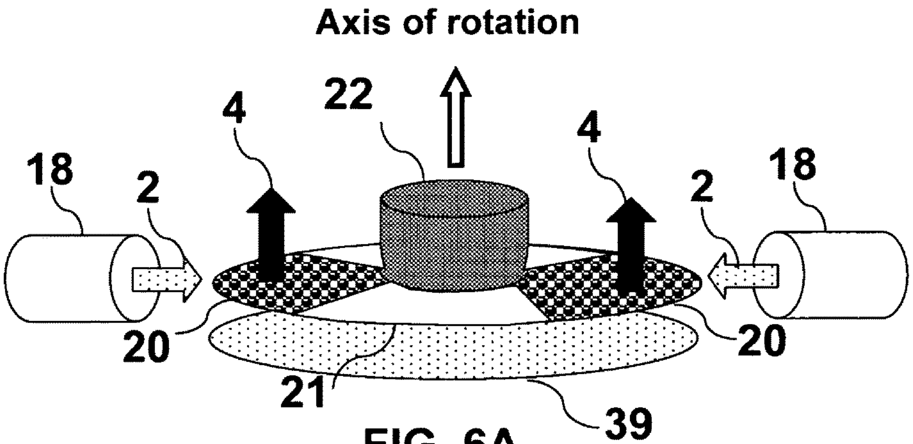
Axis of rotation

FIG. 6A

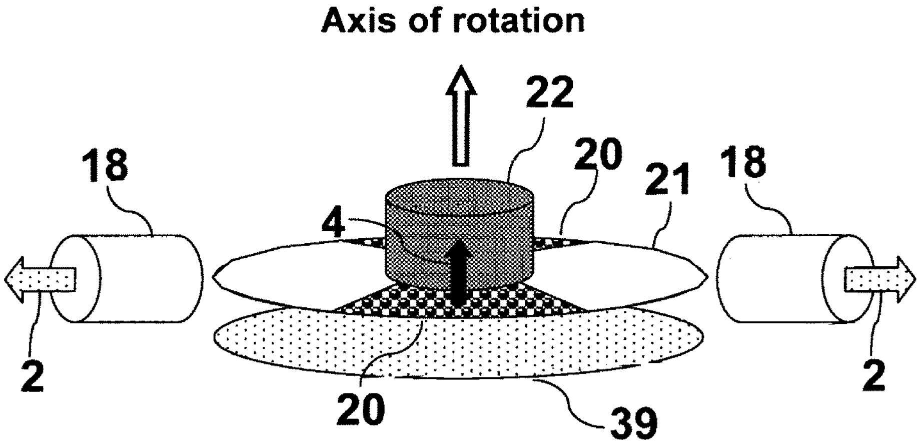
Axis of rotation

FIG. 6B

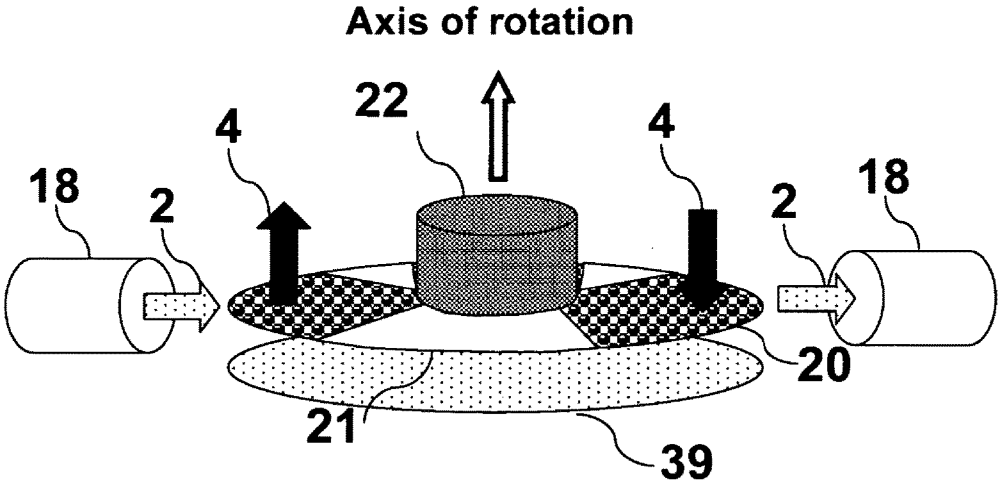
Axis of rotation

FIG. 7

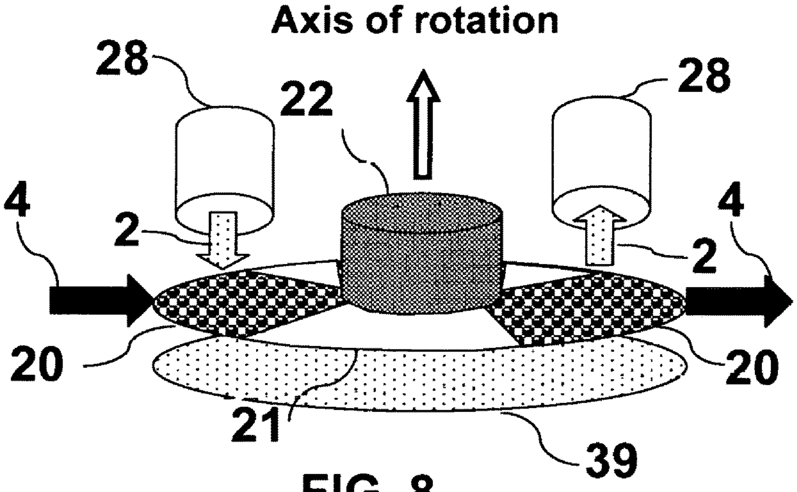
Axis of rotation

FIG. 8

ELECTROMAGNETIC SEGMENTED-CAPACITOR PROPULSION SYSTEM

FIELD OF THE INVENTION

This invention relates generally to the field of spacecraft propulsion. More specifically, the invention relates to electromagnetic spacecraft propulsion systems which provide thrust without expelling propellant.

BACKGROUND OF THE INVENTION

One of the most significant metrics for space propulsion systems is specific impulse, which is the ratio of thrust produced to the rate of propellant consumed. Specific impulse has units of seconds, and is essentially the number of seconds that a pound of propellant will produce a pound of thrust. The higher the specific impulse the less propellant mass and associated tankage required for a given space mission. The Stafford Synthesis Group (Stafford, *America at the Threshold*) concluded that future space exploration will require advanced propulsion technologies. Subsequently, NASA conducted the *Breakthrough Propulsion Physics Program* (NASA/TM-1998-208400), one of the three main goals of which was to "discover new propulsion methods that eliminate or dramatically reduce the need for propellant. This implies discovering fundamentally new ways to create motion, presumably by manipulating inertia, gravity, or by any other interactions between matter, fields, and space-time." Therefore, an electromagnetic spacecraft propulsion system such as the present invention, that does not require expendable propellant and thus has an effectively infinite specific impulse, would address these needs and greatly reduce the cost of doing business in space.

Sir Isaac Newton's well-known Third Law of Motion, which states "To every action there is always an opposed equal reaction; or, the mutual actions of two bodies upon each other are always equal and directed to contrary parts", has been interpreted as all-encompassing for over three hundred years. The Newtonian interpretation excluded the possibility of "reactionless propulsion" of a solid body. However, the discovery of electricity and magnetism, the new branch of electrodynamics resulting from Maxwell's Equations, and the subsequent rise of relativistic electrodynamics and quantum electrodynamics, led to a new interpretation of the famous "Third Law of Motion".

One of the important characteristics of an electromagnetic (EM) wave is that it can transport energy from point to point. The rate of energy flow per unit area in a plane electromagnetic wave is described by the Poynting Vector S, defined as the cross-product of the instantaneous E-field and B-field comprising the EM wave. The direction of S gives the direction in which the energy moves. (Resnick and Halliday, *Physics Parts I and II*). Less familiar is the fact that EM waves also transport linear momentum. Just as Poincare noted in 1900, Stebens explains that the electromagnetic field closely resembles a relativistic fluid, composed of quantum, particles (photons), which transmits properties and responds to forces in much the same way as an Eulerian fluid (Stebens, *Forces on Fields*). In electromagnetism, as in Newton's solid body mechanics, the force on matter from the electromagnetic field is balanced by an equal and opposite force from matter on the field. Stebens presents an excellent proof which can be derived from Maxwell's equations and the Lorentz force law. The proof relates the reaction momentum change of the electromagnetic field to the time rate of change of the Poynting Vector, the divergence of the Maxwell Stress Tensor, and Einstein's relativistic mass-energy relation (*Forces and Fields*, op. cit.). By recognizing that electromagnetic fields can transport both energy and momentum, apparent "violations" of the Third Law are mathematically resolved by including modern field transport physics in the analysis of the reaction process.

Field propulsion, which employs electromagnetic field effects for generating propulsion forces, expels no reaction mass, and therefore effectively has an infinite specific impulse. As is well known to anyone skilled in the art, a moving charged particle generates a magnetic field. It is also well known that a magnetic field generates a force on a moving charged particle, namely the magnetic component of the full Lorentz Force, which component is proportional to the vector cross-product of the particle velocity vector and the magnetic field vector at the particle location. A common example is the well-known mutual equal-and-opposite forces on parallel conductors, which may be calculated by anyone skilled in the art through the use of the Biot-Savart Law. Prior to modern electrodynamics, the required compliance with Newton's Third Law (NTL), it has previously been accepted that the aforementioned magnetic interactions could not be used to produce a propellantless propulsion system. As discussed above, modern electrodynamics has shown that this is no longer a valid assumption.

Physicists have known since at least 1952 that apparent violations of NTL, in cases where force interactions involve charged particles and electromagnetic fields, simply do not account for the momentum carried in the fields themselves. As noted by Cullwick over sixty years ago: "The following simple relations between electromagnetic momentum, the Poynting vector for energy transport, and mass-energy equivalence do not appear to have received general recognition in connection with the validity of Newton's Third Law when applied to electromagnetic forces (Cullwick, *Nature*)." Since then, over thirty US patents and at least two dozen foreign patents have been issued for propellantless propulsion devices. Most importantly, recent experimental investigations conducted by NASA with an EM drive device (White, *AIAA Journal of Propulsion and Power*) have definitively confirmed that EM field propulsion is a reality.

Nobel Laureate Richard Feynman (*The Feynman Lectures on Physics*), among others, noted that the magnetic interaction between two charged particles moving orthogonally to each other apparently does not satisfy NTL in classical Newtonian dynamics. But, as discussed above, if the changing momentum of the electromagnetic fields of the two particles is included, then overall momentum is conserved. Engineering exploitation of this situation, to date, has not been effected due to the "circuit completion" problem. While isolated moving charges may apparently violate NTL, when they are considered as part of a complete closed circuit by application of Gauss' Law, the net forces on each circuit are found to be equal and opposite in accordance with NTL.

Prior Art for Electromagnetic Propulsion

The electrokinetic phenomenon known in the scientific literature as the Biefeld-Brown Effect, wherein electrical energy input into asymmetrical capacitors can be converted to mechanical energy, provides a reactionless force for propelling an object. T. T. Brown, a co-discoverer of the Biefeld-Brown Effect, has been issued U.S. Pat. No. 2,949,550, U.S. Pat. No. 3,018,394, and U.S. Pat. No. 3,187,206 for devices based on utilization of the effect. Townsend's colleague A. H. Bahnson has similarly been issued U.S. Pat. No. 2,958,790 U.S. Pat. No. 3,223,038, and U.S. Pat. No. 3,227,901 utilizing the effect. NASA is still experimentally investigating the use of Brown's discovery, as disclosed in U.S. Pat. No. 6,317,310, U.S. Pat. No. 6,411,493, and U.S. Pat. No. 6,775,123 to Campbell (assigned to NASA). Another such device is disclosed in U.S. Pat. No. 6,492,784 to Serrano, which generates the Biefeld-Brown Effect using stacked-disc asymmetrical capacitors, and which has just been approved for spaceflight testing. Debate is ongoing in the literature as to whether the Biefeld-Brown Effect will work in the vacuum of space. Another limiting consideration for using the effect may be the scalability potential, since asymmetrical capacitor devices to date have only generated tens of milli-newtons of thrust from tens of watts of input power.

Electrodynamic structures, as disclosed in U.S. Pat. No. 7,913,954 to Levin, include a power system, a plurality of collectors, a plurality of emitters, and conductive paths for moving payloads through the Earth's magnetic field. The inertial propulsion device disclosed in U.S. Pat. No. 8,066,226 to Fiala utilizes several interconnected gyroscopic elements within Earth's gravity field to move without propellant. U.S. Pat. No. 5,377,936 to Mitchell, discloses a method for using the electromagnetic spectrum particles which continually irradiate the earth from space for gravity guidance and propulsion of geosynchronous satellites, and other satellites and space vehicles within the gravitational influence of the earth. The superconducting electromagnetic turbine, as disclosed in U.S. Pat. No. 8,575,790 to Ogilvie, uses a pair of counter-rotating electrodynamic superconductor rotors to displace the surrounding geomagnetic field. These devices do not have general space-based utility because they are restricted to operations within either the gravity field or the magnetic field of Earth.

Other approaches to propellantless propulsion include mechanical devices such as U.S. Pat. No. 2,886,976 to Dean, which discloses a device for converting rotary motion into unidirectional motion. Similarly, U.S. Pat. No. 5,182,958 to Black discloses a mechanical device with wheels inside comprising both a nonlinear propulsion as well as an energy conversion system. A device for propellantless propulsion using radiated electromagnetic fields is disclosed in U.S. Pat. No. 5,142,861 to Schlicher et al., comprising an electromagnetic propulsion system based on an extremely low frequency radiating antenna structure driven by a matched high current pulsed power supply. U.S. Pat. No. 5,197,279 to Taylor also discloses an engine for imparting momentum to a vehicle by generating and transmitting in predetermined directions massive quantities of electromagnetic field energy in successive pulses.

Field modification approaches to propellantless propulsion include apparent reductions in gravitational mass or inertial mass. U.S. Pat. No. 3,626,605 to Wallace discloses a method and apparatus for generating a time-variant non-electromagnetic force field due to the relative motion of moving bodies constituted of elements whose nuclei have half integral "spin" values, with said force field exhibiting itself in the form of an induced secondary gravitational force. U.S. Pat. No. 5,280,864 to Woodward discloses a method for producing transient fluctuations in the inertial masses of material objects by employing an effect that is a consequence of relativistic theories of gravitation. Applying pulsed thrust synchronously with the mass fluctuations thus produced facilitate the transport of massive objects. Also to Woodward, both U.S. Pat. No. 6,098,924 and U.S. Pat. No. 6,347,766 disclose inventions for producing propellantless thrust by using piezoelectric force transducers attached to resonant mechanical structures, in accordance with Mach's principle and local Lorentz-invariance predictions of transient rest mass fluctuations in accelerated objects.

The most current example of a propellantless field propulsion system is an electromagnetic drive system as disclosed in British Patents GB2229865, GB2334761, GB2399601 and UK Patent Application GB2537119 to Shawyer, as well as U.S. Pat. Appl. No. 20140013724 to Fetta. This system includes an axially-asymmetric resonant cavity with a conductive inner surface adapted to support a standing electromagnetic (EM) wave. The internal standing EM wave induces a unidirectional force on the resonant cavity, thus generating thrust without reaction mass. Experimental versions of these EM devices have produced thrust levels of micro-newtons up to milli-newtons from several kilowatts of input power (*AMA Journal of Propulsion and Power*, op. cit.)

SUMMARY OF THE INVENTION

The current disclosure presents a novel electromagnetic spacecraft propulsion system which does not expell propellant and requires only electrical power. The basic elements of the invention are a segmented capacitor cathode, an anode with a dielectric, and electromagnetic coils. The charged cathode segments are rotated through electric fields from the electromagnetic coils in such a manner as to produce a unidirectional Lorentz Force on the coil-cathode system. Thus electric power is used to generate momentum on the system through the electromagnetic production of a unidirectional Lorentz Force, while reaction momentum is carried away by the Poynting Vector fields in conformity with the currently understood principles of modern physics.

The current invention is capable of producing continuously variable thrust, including reverse thrust, torques, and lateral forces. Embodiments of the present invention are novel in that thrust is generated in a manner distinct from the devices and methods of Brown, Campbell, Serrano, Shawyer, Fetta, and others as mentioned above. Engineering calculations indicate that the present invention is a significant improvement in thrust level with several orders of magnitude less power consumption as compared to other such propulsion devices. No evidence has been found in the literature of any device designed to take advantage of the electromagnetic field effect as herein described for the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and form part of the specification, illustrate various principles of operation and examples of the best mode of the present invention.

FIG. **1** is a schematic diagram illustrating the Lorentz Force acting on a charged particle moving through a magnetic field.

FIG. **2** is a schematic diagram illustrating the mutual Lorentz Forces acting on segments of parallel conductors.

FIG. **3** is a schematic diagram illustrating the unidirectional Lorentz Force acting on two charged particles moving orthogonally in the same plane.

FIG. **4** is a schematic diagram illustrating the unidirectional Lorentz Force induced by an electromagnetic field on a group of charges attached to a moving conducting plate.

FIG. **5** is a cross-section side view schematic diagram illustrating the elements of the best mode of the present invention.

FIGS. **6**A and **6**B present schematic diagram perspective views illustrating the production of axial forces.

FIG. **7** is a schematic diagram illustrating the production of a torque.

FIG. **8** is a schematic diagram illustrating the production of a lateral translation force.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

The present invention will now be described more fully hereinafter with reference to the accompanying drawings, in which preferred embodiments of the invention are shown. This invention may, however, be embodied in many different forms and should not be construed as limited to the embodiments set forth herein. Rather, these embodiments are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the invention to those skilled in the art. Like numbers refer to like elements throughout.

By way of background, and with reference initially to FIG. **1**, it is well known to those skilled in the art that a charged particle **1**, with velocity vector **3** parallel to the x-axis as shown, moving through magnetic field **2** which is parallel to the y-axis, will experience a Lorentz Force component **4**, parallel to the z-axis, as determined by the sign of the particle charge and the vector cross-product of the velocity vector and the magnetic field vector. The necessary equations to calculate the force magnitude and direction are well known in the current physics literature.

By way of further background, and with reference to FIG. **2**, it is also well known to those skilled in the art that segments of two parallel conductors **5** each carrying a current **7** will experience equal and opposite Lorentz Forces **4** as shown. The forces will be attractive if the currents are in the same direction, otherwise the forces will be repulsive. The necessary equations to calculate the force magnitude per unit length of conductor are well known in the current physics literature.

By way of further background, the situation presented in FIG. **3** is also well known to those skilled in the art and familiar with the works of Feynman and Tipler. The figure illustrates two positively charged coplanar particles **1** and **8** moving with respective orthogonal velocities **3** and **10**, with velocity vector **10** being parallel to the y-axis and velocity vector **3** being parallel to the x-axis of the indicated coordinate system. At the instant shown, particle **8** induces a magnetic field **2** on particle **1**, while particle **1** induces no magnetic field effect on particle **8**. As a result, a Lorentz Force component **4** acts on particle **1**, while no reaction force acts on particle **8**. This situation has been well-studied in the literature, both as an apparent violation of Newton's Third Law (e.g. Cullick, Maimon) and as a potential for devising a propellantless propulsion device. However, experimental attempts to date have the charged particles confined in conductors as part of continuous current loops, and consequently when all forces are properly accounted for on the continuous current loops, no net unidirectional system force is produced on the system. Thus to date no propellantless propulsion device has been developed exploiting this situation.

It is the genius innovation of the present invention that exploits the situation of FIG. **3** by physically engineering an actual current segment that is not part of a continuous current loop. With reference to FIG. **4**, an electromagnetic coil **18** is positioned such that the axis of symmetry of the coil is in the direction of the y-axis of the coordinate system shown. When energized with a current, the electromagnetic coil **18** produces a magnetic field **2** with strongest vector field component along the x-axis, which is the axis of symmetry of the electromagnetic coil **18**. A negatively charged plate **20** is moved with velocity **3** in the y-direction such that it passes through the axis of symmetry of the electromagnetic coil **18** as shown. At the appropriate time during which the charged plate **20** is aligned with the axis-of-symmetry of the electromagnetic coil **18**, the electromagnetic coil **18** is energized to produce magnetic field **2**, which results in a net Lorentz Force component **4** on the system comprised of the electromagnetic coil **18** and the charged plate **20**. This net unidirectional Lorentz Force **4** arises from the fact that most individual electrons comprising the current moving in the conducting coils of the electromagnetic coil **18** have z-axis velocity components which are orthogonal to the y-axis velocities **3** of each charge bound to the charged plate **20**. For each electromagnetic coil-charged plate electron pair wherein a z-axis velocity component exists, the electrodynamic situation of FIG. **3** is realized. Because the charges on the charged plate **20** are a discrete packet and are not part of a continuous current loop, there is thus a net unidirectional Lorentz Force realized on the system comprised of the electromagnetic coil **18** and the charged plate **20**. Comprehensive engineering calculations have verified that a net force is produced on the said coil-plate system, while reaction momentum is carried away by the Poynting Vector fields in conformity with the currently understood principles of modern physics.

FIG. **5** presents a side section view of the elements of the best mode of the present invention in which two parallel non-conductive surfaces **21** and **24** form the structural basis of a capacitor assembly. For the purpose of nomenclature hereinafter, the naming convention for the charged parts of a capacitor assembly will follow that used for polarized capacitors, i.e., any negatively charged part of a capacitor assembly will be referred to as a cathode, and any positively charged part of a capacitor assembly will be referred to as an anode. The upper surface **24**, which forms the basis of the rotating cathode of the capacitor assembly, is rigidly attached to the means of rotation **22**. It is well-known to those skilled in the art that the usual means of rotation for components on spacecraft are electric motors designed for operation in the space environment. The lower surface **21** forms the basis of the non-rotating anode of the capacitor assembly and also serves as both the static mount for multiple elements of the device as well as a structural base for attaching the propulsion system to a spacecraft.

A plurality of negatively charged conductive capacitor plates **20** are attached to the circumference of the disc **24**, so as to form collectively the rotating cathode of the capacitor assembly. A plurality of positively charged conductive capacitor plates **19**, each having a surface coating of dielectric material **29**, are attached to the surface **21**, so as to form collectively the stationary anode of the capacitor assembly. The function of the anode and cathode is to maintain a separation of positive and negative charge, as in the usual function of capacitors, but in a configuration such that the charged cathode can be rotated relative to the anode.

A plurality of electromagnetic coils **18** and **28** are rigidly attached to the surface **21**. The function of the electromagnetic coils **18** and **28** is to generate magnetic fields which interact with the charges on conductive capacitor plates **20**. The electromagnetic coils **18** are positioned radially out from and circumferentially aligned with the conductive capacitor plates **19**. The electromagnetic coils **18** are further positioned so as to radially direct a generated magnetic field onto the charged conductive capacitor plates **20** when said plates **20** rotate to a position just above the oppositely charged conductive capacitor plates **19** during operation of the device. The electromagnetic coils **28** are positioned radially out from and circumferentially aligned above the conductive capacitor plates **19**. The function of the electromagnetic coils **28** is to direct a generated magnetic field parallel to the axis of rotation onto the conductive capacitor plates **20** when said plates **20** rotate to a position just above conductive capacitor plates **19** during operation of the device.

It is well-known to those skilled in the art that electrical power for spacecraft components is provided from a standardized spacecraft power bus, which is in turn supplied by any electricity source commonly used on spacecraft, such as space-qualified batteries, advanced storage capacitors, solar panels, radioisotopic thermal generators, or any combination thereof. Spacecraft integration specialists design the electrical power connections for operation of specific components such as electromagnetic coils, motors, and capacitor charging systems. Propulsion system thrust profiles for spacecraft, which are both payload and mission dependent, include thrust level variations, thrust durations and thrust directions during a specific mission. It is therefore assumed hereinafter that a spacecraft propulsion system engineer will design the appropriate electric power system, to include on/off power switching, power levels, and power polarity, for desired operation of the present invention.

FIGS. **6**A and **6**B illustrate the use of the best mode embodiment of the present invention to generate an axially directed propulsive force. For clarity, only the minimum necessary components from FIG. **5** are shown. With reference to FIG. **6**A, charged cathode plates **20**, which are attached to non-conductive disc **24**, are rotated by means **22**. Oppositely charged anode plates **19**, dielectrics **29**, and non-conductive surface **21** are shown represented as the single composite disc element **39** solely for clarity of illustration. The composite disc element **39** and magnetic coils **18** are stationary. As cathode plates **20** rotate to a position aligned with the axes of symmetry of coils **18**, the coils are energized so as to produce magnetic fields **2** with the polarities as shown. The interaction of the magnetic fields **2** with the rotating charges on cathode plates **20** produces the unidirectional Lorentz Force components **4**. With reference to FIG. **6**B, as the charged cathode plates **20** are further rotated out of radial alignment with the axes of coils **18**, the current in coils **18** is reversed. This reverses the polarity of the generated magnetic fields **2**. At the non-aligned location of cathode plates **20**, the components of fields **2** are of lower strength but are still directed radially inward. The interaction of the these components of the magnetic fields **2** with the rotating charges on cathode plates **20** produce the reduced unidirectional Lorentz Force components **4**. The net effect during one complete cycle of rotation is a continuous unidirectional Lorentz Force on the system in the direction of the axis of rotation.

With reference to FIG. **7**, charged cathode plates **20**, which are attached to non-conductive disc **21**, are rotated by means **22**. Oppositely charged anode plates **19**, dielectrics **29**, and non-conductive surface **21** are shown represented as the single composite disc element **39** solely for clarity of illustration. The composite disc element **39** and magnetic coils **18** are stationary. As cathode plates **20** rotate to a position aligned with the axes of symmetry of coils **18**, the coils are energized so as to produce magnetic fields **2** with the polarities as shown, in which opposing coils **18** have polarities reversed with respect to each other. The interaction of the magnetic fields **2** with the rotating charges on cathode plates **20** produces the unidirectional Lorentz Force components **4**. This reversal of the polarity of the generated magnetic fields **2** causes the generated Lorentz Forces **4** to act in opposite directions, thus causing no net unidirectional force on the system but producing a torque on the system perpendicular to the axis of rotation.

With reference to FIG. **8**, charged cathode plates **20**, which are attached to non-conductive disc **21**, are rotated by means **22**. Oppositely charged anode plates **19**, dielectrics **29**, and non-conductive surface **21** are shown represented as the single composite disc element **39** solely for clarity of illustration. The composite disc element **39** and magnetic coils **28** are stationary. As cathode plates **20** rotate to a position aligned with the axes of symmetry of coils **28**, the coils are energized so as to produce magnetic fields **2** with the polarities as shown, in which the coils **28** have polarities reversed with respect to each other. The interaction of the magnetic fields **2** with the rotating charges on cathode plates **20** produce the unidirectional Lorentz Force components **4** acting in the plane of rotation of the cathode plates **20**. The sum of the generated Lorentz Forces **4** results in a net unidirectional translation force on the system which is perpendicular to the axis of rotation.

It is to be understood that even though numerous characteristics and advantages of the present invention have been set forth in the foregoing description, together with details of the structure and function of the invention, the disclosure is illustrative only, and changes may be made in detail, especially in matters of shape, size and arrangement of parts within the principles of the invention to the full extent indicated by the broad general meaning of the terms in which the appended claims are expressed.

What is claimed is:

1. An electromagnetic spacecraft propulsion system comprising:

a capacitor of two or more parts having charged conductive elements;

a means for rotating parts of the capacitor, wherein said means for rotating sets said charged conductive elements in motion;

at least two electromagnetic coils for generating magnetic fields;

wherein the charged conductive elements and the at least two electromagnetic coils are arranged in certain geometric configurations whereby a net unidirectional Lorentz force is created by an action of the magnetic fields generated by the at least two electromagnetic coils on the charged conductive elements in motion.

2. The electromagnetic spacecraft propulsion system according to claim **1**, wherein each part of the capacitor consists of an insulator carrying one or more separate charged conductive elements.

3. The electromagnetic spacecraft propulsion system according to claim **2**, wherein one or more of the separate charged conductive elements have a dielectric coating.

4. The electromagnetic spacecraft propulsion system according to claim **1**, wherein the separate charged conductive elements of the capacitor are each charged to a desired voltage and polarity.

5. The electromagnetic spacecraft propulsion system according to claim **1**, wherein one or more parts of the capacitor are rotated about an axis of symmetry with respect to other static parts of the capacitor.

6. The electromagnetic spacecraft propulsion system according to claim **1**, wherein one or more parts of the capacitor are counter-rotated about an axis of symmetry with respect to other rotating parts of the capacitor.

7. The electromagnetic spacecraft propulsion system according to claim **1**, wherein two or more of the at least two electromagnetic coils are positioned circumferentially around a rotational axis of symmetry capacitor.

8. The electromagnetic spacecraft propulsion system according to claim **1**, wherein magnetic fields of desired polarity, intensity and duration are generated from the at least two electromagnetic coils.

9. The electromagnetic spacecraft propulsion system according to claim **1**, wherein the net unidirectional Lorentz force is generated parallel to an axis of rotation of the capacitor.

10. The electromagnetic spacecraft propulsion system according to claim **1**, wherein the net unidirectional Lorentz force is generated perpendicular to an axis of rotation of the capacitor.

11. The electromagnetic spacecraft propulsion system according to claim **1**, wherein the net unidirectional Lorentz force is a torque on the electromagnetic spacecraft propulsion system, and wherein the torque is perpendicular to an axis of rotation of the capacitor.

* * * * *

Dublin Institute of Technology
ARROW@DIT

Doctoral

Engineering

2015

Study of UHF and VHF Compact Antennas

Abraham Loutridis
Dublin Institute of Technology

Follow this and additional works at: http://arrow.dit.ie/engdoc

Part of the Electrical and Computer Engineering Commons

Recommended Citation

Loutridis, A. (2015). *Study of UHF and VHF compact antennas*. Doctoral thesis. Dublin Institute of Technology. doi:10.21427/D7NW2T

This Theses, Ph.D is brought to you for free and open access by the Engineering at ARROW@DIT. It has been accepted for inclusion in Doctoral by an authorized administrator of ARROW@DIT. For more information, please contact yvonne.desmond@dit.ie, arrow.admin@dit.ie, brian.widdis@dit.ie.

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License

STUDY OF UHF AND VHF COMPACT ANTENNAS

Abraham Loutridis

Doctor of Philosophy

Supervisors:
Prof. Max J. Ammann
Dr. Matthias John

Dublin Institute of Technology
School of Electrical & Electronic Engineering

July 2015

Abstract

This thesis presents and describes designs of small antennas that operate in UHF and VHF frequency bands. The proposed antennas are designed for integration into small volumes, therefore low profile, compact size and good radiation properties are the key parameters in this work. A further investigation on miniaturization techniques, as well as the ground plane effects on the general performance, is also made.

The main objective is the design of novel compact sized geometries, lightweight and cost efficient, operating in the lower UHF and VHF frequency bands. The groundplane size and the antenna position with respect to it, are two parameters which are investigated and contribute to optimum design performance. Compact solutions are realised in this work based on folded, meander-line and inverted-F geometries providing broadband operation and omnidirectional radiation properties.

The investigation of broadband properties of a dual band folded monopole led to a controllable frequency-ratio with wide range, operating in the WLAN frequency spectrum. The proposed solution offers high efficiency and gain and stable omnidirectionality across the operating frequency band.

The study also deals with planar inverted-F antennas (PIFA) operating in the LTE frequency bands. The two highly efficient broadband antennas provide compactness, gain

stability and are fabricated using low-cost materials. By configuring an optimised position of the PIFA on the groundplane, the impedance bandwidth, the gain and the total efficiency can be significantly improved. A more compact solution of a dual band PIFA structure is provided with omnidirectional radiation characteristics and large frequency ratio for machine-to-machine applications.

A novel tuneable meander line structure operating over the frequency range of 412 − 475 MHz is designed for integration into smart meter devices. The resonant frequency of this antenna can be tuned using a sliding via connector. A matching stub is introduced into the proposed geometry to improve the impedance matching and to shift the resonant frequency to lower values. This innovative solution overcomes material loading problems when installed on a concrete wall, as well as the S_{11} characteristic are not impaired with the small sized ground plane.

Finally, a dual band meander line folded monopole antenna in the lower UHF and VHF frequency bands is proposed for smart metering and Wireless M-Bus applications. The miniaturization of the proposed solution is based on a double-sided meandering structure which also offers good isolation between the two sections and an easily controlled large frequency-ratio. The introduction of a shunt lumped inductor improves the impedance matching at both frequencies. The antenna despite its compact size offers high total efficiency and gain across the operating frequency bands.

Declaration

I certify that this thesis which I now submit for examination for the award of PhD, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work.

This thesis was prepared according to the regulations for postgraduate study by research of the Dublin Institute of Technology and has not been submitted in whole or in part for another award in any other third level institution.

The work reported on in this thesis conforms to the principles and requirements of the DIT's guidelines for ethics in research.

The Institution has permission to keep, to lend or to copy this thesis in whole or in part, on condition that any such use of the material of the thesis be duly acknowledged.

Signature: ________________________ Date ______________________
Abraham Loutridis

Acknowledgements

I am sincerely thankful to my supervisors, Professor Max J. Ammann and Doctor Matthias John, for the guidance they provided me during the elaboration of my work. I would also like to thank them for the patience, the support and the encouragement shown to me all this time.

I thank all my colleagues at the Antenna and High Frequency Research Centre for their support and interesting discussions. In particular, I am very grateful to Giuseppe and to Kansheng for their assistance and our discussions regarding to issues related to antenna miniaturization techniques. I also wish to express my gratitude to my colleagues Padraig, Antoine, Xiulong, Afshin, Oisin, Adam and Domenico with whom I had the great pleasure to work and socialize for the last four years.

I thank my colleagues at the Centre for Telecommunications Value-Chain Research group for the extraordinary learning experience we shared those years. I am also very grateful to Science Foundation Ireland for supporting my research work.

My heartfelt thanks to all my friends here in Dublin and back in Greece for their invaluable company has been always greatly supportive and valuable to me.

Very special thanks go to my housemates Panos for our countless hours of philosophising and Rasa for making me feel at home all the time and for the great fun we had altogether during the last year of my studies.

I would like to express all my gratitude to my close family back in Greece for their lasting love. Finally, I dedicate this thesis to my parents, to my sister to my grandmother and to my brother in law for their support and love that they showed me all these years.

Abbreviations

AMR	Automatic Meter Reading
AUT	Antenna Under Test
CDMA	Code Division Multiple Access
CST	Computer Simulation Technology GmbH
ECC	Electronic Communication Committee
EIRP	Equivalent Isotropically Radiated Power
ESA	Electrically Small Antenna
FIT	Finite Integration Technique
GPRS	General Packet Radio Service
GSM	Global System for Mobiles
IEEE	Institute of Electrical and Electronic Engineers
LTE	Long Term Evolution
MWS	*CST* Microwave Studio
M2M	Machine-to-Machine
PCB	Printed Circuit Board
PCS	Personal Communication System
PIFA	Planar Inverted-F Antenna
RF	Radio Frequency
SGH	Standard Gain Horn
VNA	Vector Network Analyzer
VSWR	Voltage Standing Wave Ratio
WLAN	Wireless Local Area Network

Contents

ABSTRACT .... **I**

DECLARATION .... **III**

ACKNOWLEDGEMENTS .... **IV**

ABBREVIATIONS .... **VI**

1. INTRODUCTION **1**

1.2 ANTENNA REQUIREMENTS FOR WIRELESS COMMUNICATION SYSTEMS 6
1.3 MOTIVATION FOR ELECTRICALLY SMALL ANTENNAS 7
1.4 OUTLINE OF THE THESIS 9

2. BACKGROUND **11**

2.1 ANTENNA THEORY 11
2.1.1 DIPOLE AND MONOPOLE ANTENNAS 11
2.1.2 PLANAR INVERTED-F ANTENNAS 13
2.1.3 MEANDER LINE ANTENNAS 14
2.1.4 INPUT IMPEDANCE 15
2.1.5 RETURN LOSS AND BANDWIDTH 16
2.1.6 RADIATION PATTERN 17
2.1.7 DIRECTIVITY AND GAIN 18
2.1.8 ANTENNA EFFICIENCY 19
2.1.9 FUNDAMENTAL LIMITATIONS OF ELECTRICALLY SMALL ANTENNAS 19
2.1.10 WHEELER CAP METHOD 21
2.1.11 COMPARISON WITH OTHER METHODS 25
2.2 SOFTWARE SIMULATION AND MODELLING TOOLS 25
2.2.1 TIME & FREQUENCY DOMAIN SOLVER 27
2.3 PROTOTYPING 28
2.4 MEASUREMENT SETUP 28
2.4.1 VECTOR NETWORK ANALYZER (VNA) 30
2.4.2 ANECHOIC CHAMBER 31

2.4.3 STANDARD GAIN ANTENNA (SGA) 31
2.4.4 COORDINATE SYSTEM 32

3. PRINTED FOLDED MONOPOLE ANTENNA 33

3.1 ANTENNA DESIGN 33
3.2 PARAMETRIC STUDY 36
3.3 GROUND PLANE INVESTIGATION 40
3.4 SIMULATED AND MEASURED RESULTS 46
3.5 CONCLUSIONS 49

4. PLANAR INVERTED-F ANTENNAS 51

4.1 DUAL BAND LTE PIFA FOR M2M APPLICATIONS 51
4.1.1 ANTENNA DESIGN 52
4.1.2 PARAMETRIC STUDY 54
4.1.3 GROUND PLANE INVESTIGATION 58
4.1.4 SIMULATED AND MEASURED RESULTS 65
4.2 DUAL LTE PIFA FOR M2M APPLICATIONS 69
4.2.1 ANTENNA DESIGN 69
4.2.2 PARAMETRIC STUDY 72
4.2.3 SIMULATED AND MEASURED RESULTS 76
4.3 CONCLUSIONS 82

5. MEANDER LINE MONOPOLE ANTENNA 83

5.1 MEANDERED ANTENNA FOR M2M APPLICATIONS 84
5.2 ANTENNA CONFIGURATION 84
5.3 MATCHING STUB 87
5.4 GROUND PLANE SIZE 88
5.5 MEASURED AND SIMULATED RESULTS 89
5.6 INSTALLED PERFORMANCE 95
5.7 CONCLUSIONS 98

6. VHF & UHF MEANDERED MONOPOLE ANTENNA 99

6.1 VHF & UHF MEANDERED MONOPOLE ANTENNA 100
6.2 ANTENNA CONFIGURATION 100
6.3 PARAMETRIC INVESTIGATION 105
6.4 SIMULATED AND MEASURED RESULTS 109

7. CONCLUSIONS AND FUTURE WORK 114

BIBLIOGRAPHY 118

APPENDIX A. 130

List of Figures

Figure 1.1. Wireless Communication Technology. 3

Figure 1.2. The radiansphere. 6

Figure 1.3. (a) λ/2 dipole, (b) λ/4 monopole on a ground plane..................................8

Figure 2.1. Voltage and Current distribution of (a) half wave and (b) full wave dipole. 12

Figure 2.2. Voltage and Current distribution of quarter wave monopole. 13

Figure 2.3. The evolution of the PIFA design. 14

Figure 2.4. Planar Inverted-F Antenna (PIFA) structure. 14

Figure 2.5. Meander Line Monopole Antenna structure. 15

Figure 2.6. The three radiation pattern types (a) isotropic, (b) omnidirectional and (c) directional [1]. 17

Figure 2.7. Measurement Wheeler Cap method [57]. 22

Figure 2.8. Different shapes of shields is used in the Wheeler Cap method [57]. ...23

Figure 2.9. The (a) Tetrahedral and the (b) Hexahedral mesh type.........................26

Figure 2.10. Measurement setup. 29

Figure 2.11. Near R1 and Far R2 field boundaries. 30

Figure 2.12. Standard Spherical Coordinate System. 32

Figure 3.1. Evolution of the folded monopole antenna. 34

Figure 3.2. Antenna geometry. 35

Figure 3.3. Manufactured folded monopole antenna prototype. 35

Figure 3.4. The simulated S_{11} dependence on the height (h) of the monopole.........36

Figure 3.5. The simulated S_{11} dependence on the width (*a*) of the horizontal strip of the monopole. 37

Figure 3.6. The simulated S_{11} dependence on the length (*l*) of the monopole. 38

Figure 3.7. The simulated S_{11} dependence on the width (*f*) of the feeding strip of the monopole 39

Figure 3.8. The current distribution at (a) 2.45 GHz and (b) 5.8 GHz. 39

Figure 3.9. The simulated S_{11} dependence on the size of the ground plane. 41

Figure 3.10. Simulated Total Efficiency for various ground plane sizes. 42

Figure 3.11. Simulated radiation patterns for various ground plane sizes at 2.45 GHz. 43

Figure 3.12. Simulated radiation patterns for various ground plane sizes at 5.8 GHz. 44

Figure 3.13. The simulation model with the antenna located in the (a) centre of the plastic enclosure and the antenna located in the (b) corner of the plastic enclosure. 45

Figure 3.14. Simulated S_{11} for two different locations of the antenna into the plastic enclosure. 45

Figure 3.15. Simulated and measured S_{11} for the printed folded monopole antenna. 46

Figure 3.16. The printed folded monopole antenna in the anechoic chamber. 47

Figure 3.17. Measured and simulated radiation patterns at 2.45 GHz. 48

Figure 3.18. Measured and simulated radiation patterns at 5.8 GHz. 49

Figure 4.1. Rear view (a), Front view (b), Left view (c), General view (d), CST simulation model and the coordinate system (e). 53

Figure 4.2. Manufactured dual band LTE PIFA prototype. 54

Figure 4.3. The simulated S_{11} dependence on the feeding strip position (*a*) of the PIFA. 55

Figure 4.4. The simulated S_{11} dependence on the height (*h*) of the PIFA. 55

Figure 4.5. The simulated S_{11} dependence on the length of the rear folded section (*b*) of the PIFA. 56

Figure 4.6. The simulated S_{11} dependence on the length of the front folded section (*d*). 57

Figure4.7. The simulated S_{11} dependence on the shorting strip position (*s*) of the PIFA. 58

Figure 4.8. The parameters X and Y which define the position of the PIFA antenna. 59

Figure 4.9. Ground plane current distribution at 825 MHz for three different position of the PIFA. 60

Figure 4.10. Simulated S_{11} when moving the antenna along y-axis. 60

Figure 4.11. Simulated S_{11} when moving the antenna along x-axis. 61

Figure 4.12. Radiation patterns for different positions of the PIFA on the ground plane. 62

Figure 4.13. The simulated S_{11} dependence on the size of the ground plane. 63

Figure 4.14. Simulated Total Efficiency for various ground plane sizes. 63

Figure 4.15. Current Distribution at 867, 1917 and 2652 MHz. 64

Figure 4.15. Simulated and measured S_{11} for the PIFA antenna. 65

Figure 4.16. The simulated total and radiation efficiency. 66

Figure 4.17. Measured and simulated (a) azimuth (x-y), (b) elevations (z-y) and (c) (z-x) plane patterns at 825 MHz. 67

Figure 4.18. Measured and simulated (a) azimuth (x-y), (b) elevations (z-y) and (c) (z-x) plane patterns at 2200 MHz. 68

Figure 4.19. Schematic illustration of dual LTE PIFA design evolution. 70

Figure 4.20. Antenna Model and the coordinate system. 71

Figure 4.21. Antenna geometry and parameters. 71

Figure 4.22. Dual PIFA LTE Antenna prototype 72

Figure 4.23. Simulated S_{11} variation for different values of (*a*). 73

Figure 4.24. Simulated S_{11} variation for different values of (*b*). 73

Figure 4.25. Simulated S_{11} variations for different values of (*c*) 74

Figure 4.26. Simulated S_{11} variations for different values of (*w*) 75

Figure 4.27. The current distribution at (a) 760 MHz and (b) 1860 MHz. 76

Figure 4.28. Simulated and measured S_{11} for the PIFA antenna. 77

Figure 4.29. StarLab spherical coordinate system [74]. 78

Figure 4.30. Measured and simulated total efficiency and realized gain. 78

Figure 4.31. Measured and Simulated Peak Directivity. 79

Figure 4.32. Measured and simulated (a) azimuth (x-y), (b) elevations (z-y) and (c) (z-x) plane patterns at 760 MHz. 80

Figure 4.33. Measured and simulated (a) azimuth (x-y), (b) elevations (z-y) and (c) (z-x) plane patterns at 1860 MHz. 81

Figure 5.1. Illustration of the antenna designed steps. 84

Figure 5.2. Simulated S_{11} for the two different antenna models. 85

Figure 5.3. Antenna geometry and the coordinate system. 86

Figure 5.4. (a) Improved impedance matching and control of frequency, (b) Smith chart. 87

Figure 5.5. Simulated S_{11} dependence on the ground plane size. 88

Figure 5.6. Simulated and measured S_{11}. 91

Figure 5.7. Current distribution at 450 MHz. 92

Figure 5.8. Total measured and simulated efficiency of the antenna at 420, 450 and 470 MHz. 92

Figure 5.9. The antenna inside the anechoic chamber. 93

Figure 5.10. Azimuth radiation patterns in the xz-plane at 450 MHz. 93

Figure 5.11. Elevation radiation patterns in the xy-plane at 450 MHz. 94

Figure 5.12. Elevation radiation patterns in the yz-plane at 450 MHz. 94

Figure 5.13. The simulation model with ground plane (a) parallel to the wall and (b) perpendicular to the wall. 95

Figure 5.14. Simulated and measured S_{11} for the parallel case. 97

Figure 5.15. Simulated and measured S_{11} for the perpendicular case. 97

Figure 6.1. Schematic illustration of the three basic design models. 101

Figure 6.2. Simulated S_{11} for the three basic design models. 101

Figure 6.3. Folded meander line monopole antenna and coordinate system. 103

Figure 6.4. Improved impedance matching and control of frequency. 104

Figure 6.5. Manufactured meandered monopole prototype (a) Front View, (b) Rear View. 104

Figure 6.6. The simulated S_{11} dependence on the height *r* of the back meander line section. 105

Figure 6.7. The simulated S_{11} dependence on the height f of the front meander line section. 106

Figure 6.8. The simulated S_{11} dependence on the height f_1 of the last front four meander line strips. 107

Figure 6.9. The current distribution of the antenna in (a) 169 MHz and (b) 433 MHz. 108

Figure 6.10. Simulated and measured S_{11}. 110

Figure 6.11. Wheeler Cap measurement setup (a) without cap, (b) with cap. 111

Figure 6.12. Azimuth radiation patterns in the xy-plane at 433 MHz. 112

Figure 6.13. Elevation radiation patterns in the xz-plane at 433 MHz. 112

List of Tables

Table 1.1. The frequency spectrum. 4

Table 1.2. Wireless Communication Frequency Spectrum. 4

Table 2.1. Properties of FR4 and Brass used for antenna prototyping. 28

Table 4.1. Measured and Simulated Total Efficiency and Realized Gain. 79

Table 5.1. Comparison of theoretical and measured total efficiency. 91

1. Introduction

The rapid evolution of wireless cellular communications started with the deployment of the 1st generation of mobile networks in the early 1980s and led to the requirement for compact terminal antennas. The new communication systems and newly introduced applications operating at frequencies in the ultra-high frequency (UHF) and super-high frequency (SFH) bands (Table 1.1) set strict requirements for the antenna in terms of size and efficiency. The portability and mobility of wireless technology demands compact and handheld devices with small enough size to be easily carried, of light weight and with low cost.

Dipole and monopole antennas were the two most practically used antennas types for wireless communication systems in the early stage of their development. The broadband radiation characteristics and manufacturing simplicity made monopoles the most widely used antennas for portable devices such as mobile and cordless phones, laptops, wireless metering devices, etc. The introduction of Global System for Mobiles (GSM), General packet radio service (GPRS) and other mobile services such as Wireless Local Area Network (WLAN), Bluetooth, ad-hoc networks, etc. (Table 1.2) as well the availability of the frequency spectrum generated further specifications and requirements for antenna designers.

In 1889, Heinrich Hertz was the first developer of a wireless communication system in which he forced an electrical spark to occur in the gap of a dipole antenna. For the receiver he used a loop antenna. By 1901, Marconi was transmitting data across the Atlantic Ocean by using several vertical wires attached to the ground. Across the Atlantic Ocean, the receive antenna was a 200 meter wire held up by a kite [1]. Since then, an extensive evolution of electrically small antennas (ESA) and compact integrated antennas followed making it one of the important elements that make modern wireless devices possible.

Modern mobile communication applications require large available channel capacity, large quantities of transmitting data, high efficiency, flexibility and capability to handle multiband frequencies. Modern mobile terminals define additional requirements for the antennas used, like compact size, low profile and low-cost, omnidirectional radiation characteristics, adopted low and safe levels of Specific Absorption Rate (SAR) values [2, 3] for devices operating within close proximity to the human body as specified in [4]. The trend in universal wireless access connectivity and the compatibility of different devices drive the motivation for miniaturization of antennas able to work well across a broadband bandwidth with good efficiency.

Nowadays, wireless communication technology (Fig. 1.1) has replaced the wired networks in an efficient way, increasing the distance beyond the capabilities of typical cabling. Wireless communication technology is used to meet many needs such as:

- To provide a backup connection in case of network failure,
- To link portable or temporary workstations,
- To overcome problems where cabling is difficult or financially impractical,
- To remotely connect mobile users or networks.

A large range of wireless applications involves home security systems, Wi-Fi, cellular phones, wireless power transfer, wireless sensors and smart meters. The implementation and optimization of existing wireless technology can be addressed also in automotive, airborne and medical applications offering reliability and low operating cost.

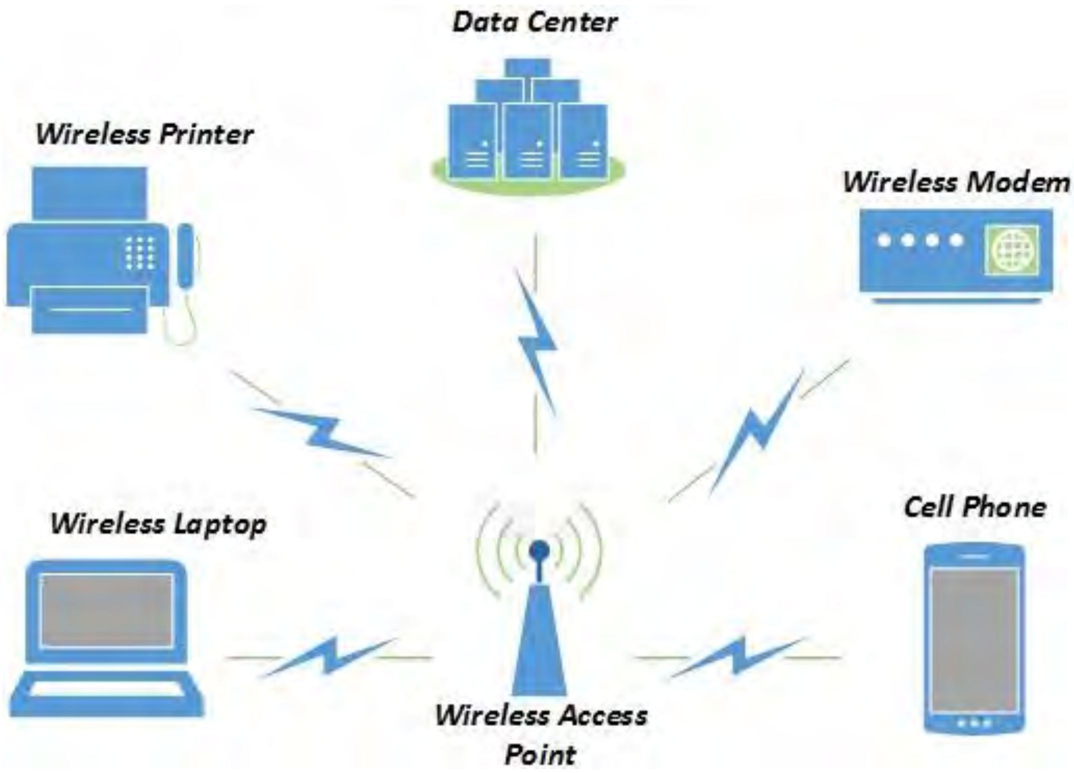
Data Center
Wireless Printer
Wireless Modem
Wireless Laptop
Cell Phone
Wireless Access Point

Figure 1.1. Wireless Communication Technology.

Table 1.1. The frequency spectrum.

Band	Frequencies		Wavelengths*	
VLF	3 kHz	30 kHz	100 km	10 km
LF	30 kHz	300 kHz	10 km	1 km
ML	300 kHz	3 MHz	1 km	100 m
HF	3 MHz	30 MHz	100 m	10 m
VHF	30 MHz	300 MHz	10 m	1 m
UHF	300 MHz	3 GHz	1 m	10 cm
SHF	3 GHz	30 GHz	1 cm	1 mm

*Wavelength: $\lambda(m) = 300/f(MHz) \Rightarrow \lambda(cm) = 30/f(GHz)$

Table 1.2. Wireless Communication Frequency Spectrum.

Wireless System	Frequency Range
AM/FM Radio	535-1605 KHz/88-108 MHz
Broadcast TV (Ch. 2-6)	54-88 MHz
Broadcast TV (Ch. 7-13)	174-216 MHz
ISM Band (LPD433)	433.05-434.79 MHz
DVB-T/TVWS (UHF)	470-862 MHz
1G/2G Cell Phones	806-902 MHz/1.85-1.99 GHz
3G UMTS	746-798 MHz, 814-894 MHz
3G UMTS	1.7-1.85 GHz, 2.5-2.69 GHz
Satellite Digital Radio	2.32-2.325 GHz
4G LTE	690-960 MHz, 1.7-2.69 GHz
ISM Band (Bluetooth, 802.11b WLAN)	2.4-2.4835 GHz
5.8 GHz ISM Band	5.725-5.875 GHz
Digital Broadcast Satellite	12.2-12.7 GHz
Local Multipoint Distribution Service (LMDS)	27.5-29.5 GHz, 31-31.3 GHz
Fixed Wireless Services	38.6-40 GHz

1.1 ELECTRICALLY SMALL ANTENNAS

In according to Harold A. Wheeler, a small antenna is one in which the maximum dimension is less than the “radianlength”. The radianlength is equal to 1/2π wavelength [5]. Another expression to define a compact antenna is that it fulfils the condition $ka<1$, which $k=2\pi/\lambda$ where, a is the radius of the minimum sphere (radiansphere) [6] that encloses the antenna (Fig. 1.1). For over 60 years, extensive research took place on the theoretical and practical field of fundamental limitations and small antennas by Chu [7], Wheeler [5, 6], Harrington [8], Fano [9] and other pioneers providing facts and results which are invaluable for the antenna engineers.

Over the years new miniaturization techniques have been documented to reduce the size of an antenna for a given frequency. Using high dielectric constant substrates enables more miniaturization as more of the antenna fields are coupled in the substrate [1, 10]. A variety of different folded [11, 12, 13, 14], spiral [15, 16], fractal [17, 18, 19, 20] and meander line [21, 22, 23, 24] geometries are reported in order to reduce the antenna size.

Antenna designers also employ lumped elements (capacitors and inductors) and varactor diodes [25, 26, 27, 28, 29] to improve impedance matching and antenna bandwidth as well to control the resonant frequency. Lumped elements can be used to achieve a good impedance matching to the source but they introduce additional losses into the antenna system. On the other hand, incomplete matching is more critical and has stronger impact on the antenna total efficiency than inserted losses from lumped elements. Despite of this disadvantage, these kind of compact antennas with external matching networks exist in many real applications because of their simple design methodology.

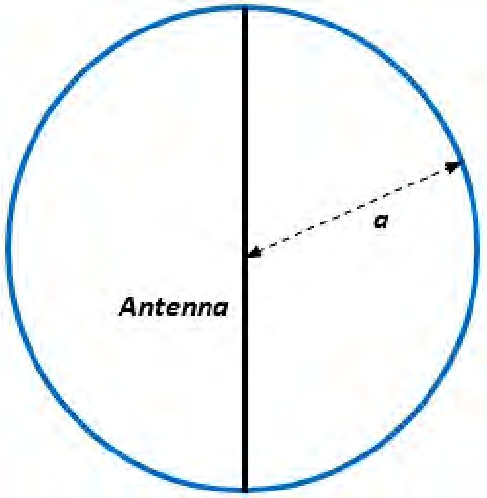

Figure 1.2. The radiansphere.

1.2 ANTENNA REQUIREMENTS FOR WIRELESS COMMUNICATION SYSTEMS

Monopole antennas have been used as the best choice for various portable and mobile equipments. The simplicity of the antenna design, the fulfilment of the performance specifications, particularly with respect to bandwidth and efficiency, are the basic criteria for the antenna designers to create electrically small antennas that are compatible with modern wireless technology, which will also operate on a small ground plane.

The ongoing evolution of wireless communication and the demand for high speed and high quality data transfer using portable and personal communication (PCS) services [30, 31] has led to the need for a reorientation of antenna design specifications as a basic part of any wireless system. The modern handheld devices require antennas which are embedded on a small area, a relatively short distance from a user body tissue operating in accordance with Electromagnetic Compatibility (EMC) Regulations.

Modern wireless units tend to have small and compact size, low-cost manufacturability, flexibility to integrate with the wireless communication system and an omnidirectional radiation performance in at least one plane. The consideration of frequency of operation, impedance matching, broadband or wideband operation, high gain and efficiency, omnidirectional radiation pattern, ground plane size and Specific Absorption Rate (SAR) are the fundamental electrical characteristics that drove the designers to introduce a wide variety of antenna structures and topologies to balance the particular demands for modern wireless communication systems.

1.3 MOTIVATION FOR ELECTRICALLY SMALL ANTENNAS

At its resonant frequency f_0, the length L of the half-wavelength (λ/2) dipole is equal to a half of the wavelength λ_0. The quarter-wavelength (λ/4) vertical monopole consists of one arm of a half-wavelength (λ/2) dipole, placed on a horizontal ground plane (Fig. 1.3).

Monopoles and dipoles have good radiation properties but for a 4G mobile phone which operates at 800 MHz, a quarter-wavelength (λ/4) is almost 9.4 cm, hence a typical dipole would not be suitable to embed in a modern portable device.

Moreover, some constrains such as the narrow bandwidth (up to 10% FBW) performance, as well as the balanced feeding, (need for baluns in some cases) and the protrusive and extended structure make them less attractive solutions for compact designs especially for low frequencies.

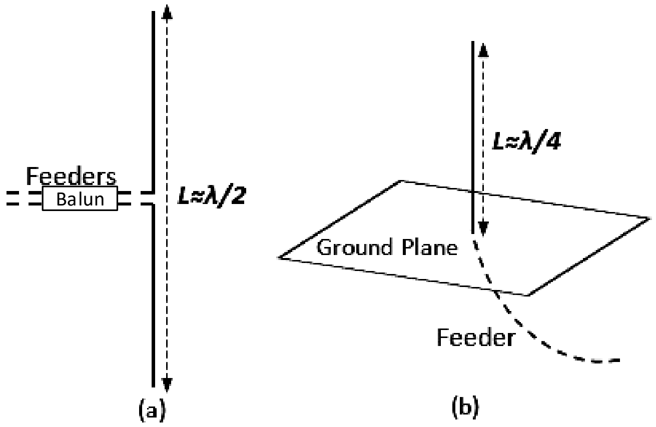

Figure 1.3. (a) λ/2 dipole, (b) λ/4 monopole on a ground plane.

Miniaturization techniques and methods enable antenna designers to enclose monopole antennas with large electrical length into a compact physical size. Electrically small and compact antennas including monopoles, dipoles, normal-mode helix antennas [32, 33], planar inverted-F antennas (PIFA's) [34, 35, 36, 37, 38], microstrip patches [39, 40] and meander line antennas [21, 22, 23, 24] used mostly for small terminals and handheld units. Additionally, miniaturization techniques give more freedom to the available space and allow the integration of other components on the same Printed Circuit Board (PCB). The low manufacturing cost is favoured from this integration.

This thesis presents an investigation of electrically small and compact antennas and the use of miniaturization techniques. A variety of topologies that include folded and meander-line monopoles and Planar Inverted-F Antennas (PIFA's) are designed and studied, as well an extensive investigation of the ground plane size which is an integral part of the electrically small antennas. Miniaturisation techniques are analysed and developed for application to antennas which are ready to be integrated into different types of mobile and compact units.

1.4 OUTLINE OF THE THESIS

Initially, in Chapter 2 an overview of dipole, monopole, PIFA and meander line antennas is provided. A theoretical background on the operation principle of antennas is also presented. The simulation software, prototyping method, measurement setup and techniques used in the thesis are introduced and described.

A printed folded monopole antenna for WLAN applications is investigated in Chapter 3. The effects of key geometrical parameters of the monopole and the role of the ground plane size in the general performance of the antenna are studied.

In Chapter 4, two LTE Planar Inverted-F Antennas (PIFAs) for M2M application are introduced. Both configurations are optimised and developed for better impedance matching and bandwidth. The PIFA position on the ground plane and the ground plane size are investigated.

Chapter 5 deals with the design of a novel tuneable meander line monopole operating in the lower UHF band. A parametric study of the ground plane size is made and the antenna is optimised for impedance matching. The installed performance of the monopole embedded in a plastic housing and placed on a concrete wall is investigated.

A dual-band meander-line monopole operating in VHF and lower UHF bands for smart meters is presented in Chapter 6. The proposed design is optimised in both frequency bands for impedance matching. A geometrical parametric investigation is made and the frequency-ratio performance is also reported. The radiation and total efficiency of proposed antenna is measured based on the Wheeler Cap Method providing good agreement with the simulated results.

Chapter 7 gives conclusions and outlines as well as possibilities for future work. A List of Journal and Conference Publications are listed in chronological order given in Appendix A.

The novelty which is provided in this work is summarized in the following paragraph. A compact low profile dual band folded monopole for WLAN applications enable to control large frequency ratio with more than 90% efficiency and 2 dBi gain across the two bands is presented. In an effort to move to lower frequencies and to maintain the compact size of the design, two LTE PIFAs with large and easily controlled frequency ratio ranges are provided with more than 80% efficiency and 2 dBi gain across the operating frequency range. A novel forked feed is introduced providing compactness. A tuneable novel UHF meander line monopole for M2M applications with very compact size and with good and stable efficiency (21%-24%) across the tuneable band is proposed. The tuneability of the antenna is based on a novel tuneable sliding via connector which controls the antenna electrical length offering adjustability to its radiation performance. Finally, the first dual band printed meander line monopole antenna is presented operating at 169 MHz and 433 MHz offering more than 20% and 50% efficiency at the first and second resonant frequencies, respectively.

2. Background

An antenna can be defined as an electrical device which converts electric power into radio waves and vice versa [41]. The IEEE defines the antenna as a means for radiating or receiving radio waves [42]. In this chapter the definitions of fundamental parameters and performance metrics of an antenna are given.

The design and software simulation tools used in this work are presented. Additionally, facilities used for the analysis, fabrication and measurement of the proposed antennas are described.

2.1 ANTENNA THEORY

2.1.1 DIPOLE AND MONOPOLE ANTENNAS

A dipole antenna is the simplest and most widely used antenna type which consists of two identical conductive wires. Between the two halves of the dipole, a current source is applied, connected to the two adjacent ends. Dipole antennas are resonant, with the current flowing back and forth between the ends of the two elements. A half-wave dipole has two wire elements of approximately a quarter wavelength long (λ/4). In Fig. 2.1 a half wave and a full wave dipole with their voltage and current distribution along the two structures is shown.

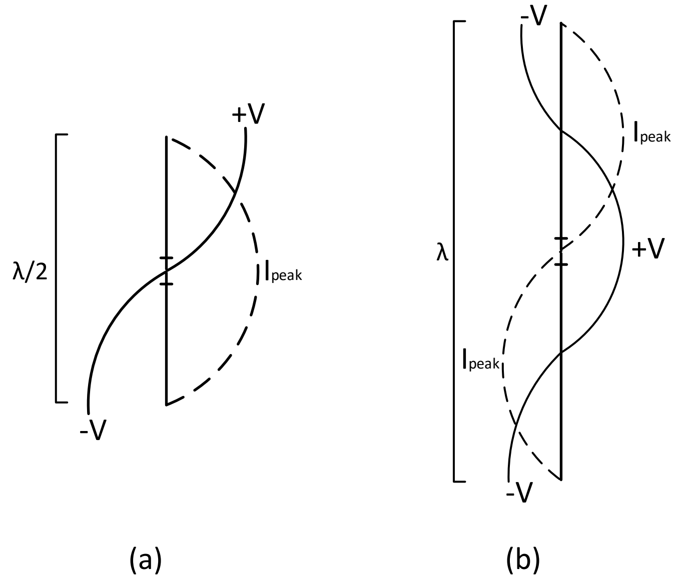

Figure 2.1. Voltage and Current distribution of (a) half wave and (b) full wave dipole.

The basic principle of a conventional quarter wavelength (*λ/4*) monopole mounted over a ground plane is that the length L, is equal to a quarter of the wavelength *λ/4* at its fundamental resonant frequency f_0 ($L_0 = \lambda_0/4\,, f_0 = 1/\lambda_0 \Rightarrow \lambda_0 = 1/f_0$) [1]. Fig 2.2 below shows the corresponding voltage and current distribution along a quarter wavelength monopole structure. The reflections from the ground plane produce a virtual identical monopole below it (mirror effect) and the monopole antenna can be evaluated in much the same way as dipole antenna.

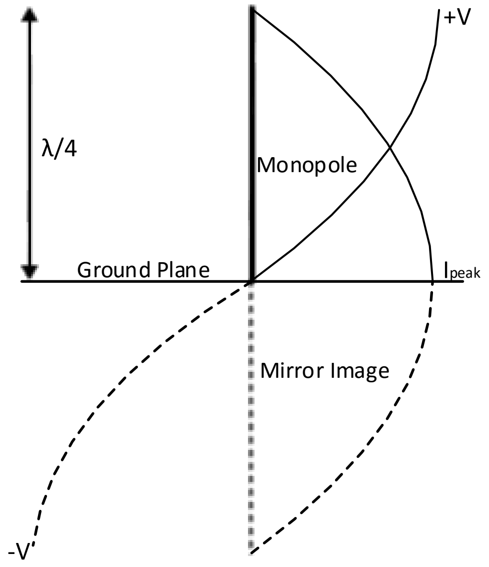

Figure 2.2. Voltage and Current distribution of quarter wave monopole.

The radiation pattern for the monopole antenna is strongly affected by a finite sized ground plane while its impedance is minimally affected [43]. Whereas a ground plane with diameter of 10λ or larger (infinite size) has a fairly small effect on the feed-point impedance of a monopole antenna [44].

2.1.2 PLANAR INVERTED-F ANTENNAS

Planar Inverted-F Antennas (PIFAs) are a widely used terminal antenna type with many advantages which make them suitable for integration in portable and compact devices [45]. The evolution of the PIFA design [46, 47, 48] can be started from a quarter wavelength (λ/4) wire monopole (Fig. 2.3 (a)) which is placed above a ground plane and it is folded by 90° as it shown in Fig. 2.3 (b). By connecting the monopole to the ground plane through a shorting pin, a significant impedance matching improvement can be achieved. The overall height is reduced while the antenna maintains the good performance.

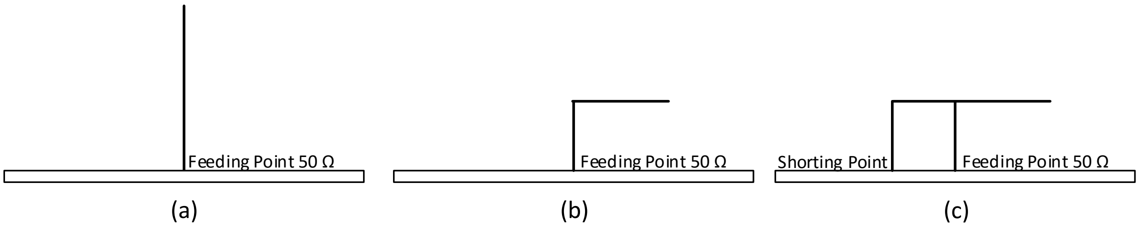
Feeding Point 50 Ω
(a)
Feeding Point 50 Ω
(b)
Shorting Point
Feeding Point 50 Ω
(c)

Figure 2.3. The evolution of the PIFA design.

Finally, the bandwidth of the antenna can be improved if the wire is replaced by a planar radiating patch, as it shown in Fig. 2.4.

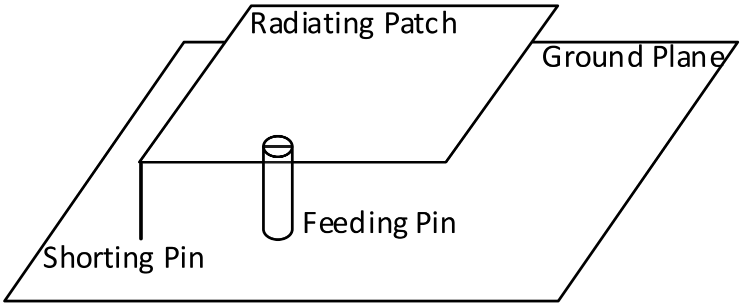
Radiating Patch
Ground Plane
Feeding Pin
Shorting Pin

Figure 2.4. Planar Inverted-F Antenna (PIFA) structure.

2.1.3 MEANDER LINE ANTENNAS

Meander Line Antennas [49, 50, 51] are a type of monopole antenna which can achieve miniaturization in size. The starting evolution point of the basic meander monopole design is achieved by folding the conductor back and forth, decreasing the overall height of the antenna and introducing more wire into the structure (Fig. 2.5).

A meander line monopole is a set of horizontal and vertical wires which form turns. For a fixed wired length, as the number of turns increases, the antenna volume decreases and as the separation between the wires increases, the resonant frequency decreases [23].

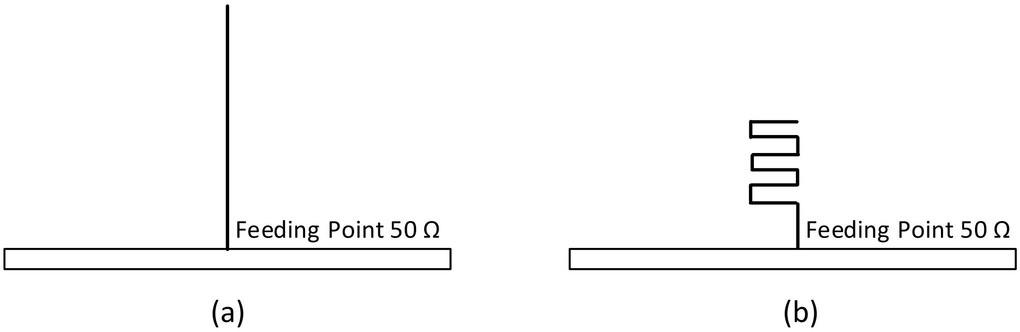

Figure 2.5. Meander Line Monopole Antenna structure.

2.1.4 INPUT IMPEDANCE

The antenna input impedance $Z_i(f)$ relates the voltage and current at the antenna input port. The real part $R_i(f)$ of the input impedance represents the power that the antenna radiates or absorbs. The imaginary part $jX_i(f)$ of the input impedance represents the power that is stored in the near field of the antenna. The input impedance is frequency dependent [52]:

$$Z_i(f) = R_i(f) + jX_i(f) \tag{2.1}$$

2.1.5 RETURN LOSS AND BANDWIDTH

The return loss is the ratio that compares the power reflected P_r at the antenna port to the power P_i that is fed into the antenna from the transmission line. It is usually expressed as a ratio in decibels (dB):

$$RL(dB) = 10\log P_i/P_r \tag{2.2}$$

The return loss is also expressed as the negative of the magnitude of the reflection coefficient in dB,

$$RL(dB) = -20\log|\Gamma|, \tag{2.3}$$

where the reflection coefficient is given by

$$\Gamma = \frac{Z_i - Z_0}{Z_i + Z_0} \tag{2.4}$$

The return loss is related to the voltage standing wave ratio (VSWR) and expressed by the following equation:

$$VSWR = \frac{1 + |\Gamma|}{1 - |\Gamma|} \tag{2.5}$$

The impedance bandwidth (BW) is another fundamental antenna parameter. It describes the frequency range over which the antenna has a return loss remains below −10 dB (VSWR 2:1). For electrically small antennas which operate at relatively low frequencies, a return loss below −6 dB (VSWR 3:1) is acceptable. The fractional bandwidth (FBW) is the ratio of the bandwidth divided by the centre frequency and is given by:

$$FBW = \frac{f_u - f_l}{f_0} \times 100\%, \qquad (2.6)$$

where f_u is the upper frequency and f_l is the lower frequency.

2.1.6 RADIATION PATTERN

The radiation pattern of an antenna is the three dimensional spatial distribution of radiated energy as a function of the antenna position along a constant radius [1]. Radiation properties include radiation intensity, field strength, phase and polarization.

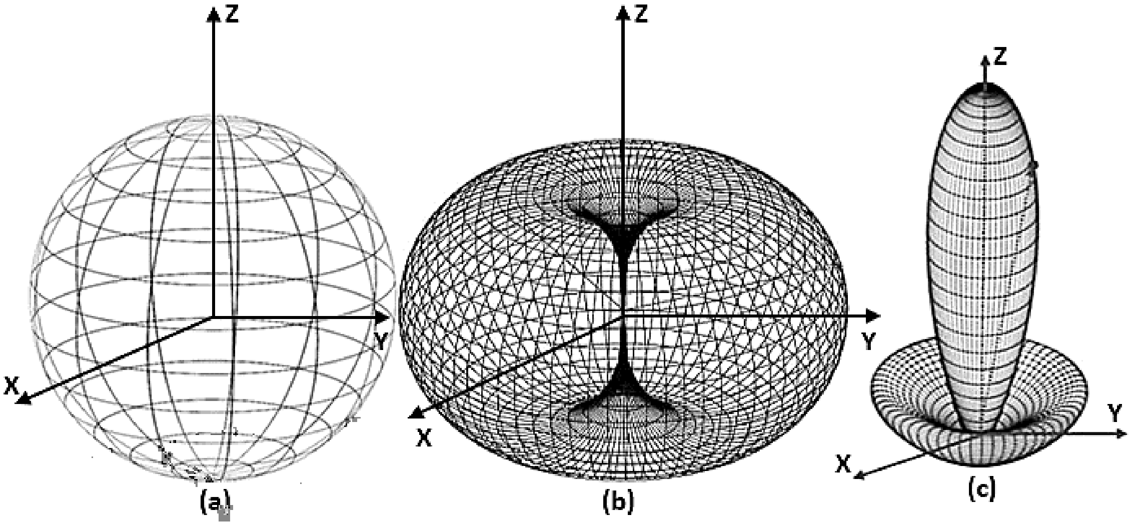

Figure 2.6. The three radiation pattern types (a) isotropic, (b) omnidirectional and (c) directional [1].

An antenna with isotropic radiation pattern is an ideal antenna having equal radiation in all directions. A directional antenna radiates more effectively in some directions than in others. Some antennas may also be described as omnidirectional which have an isotropic radiation pattern in a single plane. An example of omnidirectional antenna is the dipole antenna (Figure 2.6 (b)).

2.1.7 DIRECTIVITY AND GAIN

The antenna directivity *D* is defined the ratio of the radiation intensity in a certain direction to the radiation intensity of an isotropic radiator [1]:

$$D = \frac{U}{U_0} = \frac{4\pi U}{P_{rad}}, \tag{2.7}$$

where U is the radiation intensity, U_0 is the radiation intensity of the isotropic radiator and P_{rad} is the total radiated power.

The antenna gain *G* is related to directivity *D* and to the radiation efficiency η of an antenna [1]

$$G(\theta, \varphi) = \eta \times D(\theta, \varphi) \tag{2.8}$$

The antenna gain is usually expressed in dBi and signifies the ratio of radiated power in a given direction relative to that of an isotropic radiator which is radiating the electrical power uniformly in all directions.

According to the IEEE standards, gain does not include losses arising from impedance and polarization mismatches [53]. The term realized gain is used when mismatch effects are included.

2.1.8 ANTENNA EFFICIENCY

The antenna radiation efficiency η is defined as the ratio of radiated power P_{rad} to the power delivered (input power) P_{input} to the antenna:

$$\eta = \frac{P_{radiated}}{P_{input}} \tag{2.9}$$

An efficient antenna radiates most of the input power. An inefficient antenna has more of the power absorbed as losses within the antenna. The losses can be due to limited conductivity of the antenna or due to dielectric losses.

The Total Efficiency η_T of an antenna is the Radiation Efficiency η multiplied by the impedance mismatch loss $(1 - |\Gamma|^2)$ of the antenna:

$$\eta_T = (1 - |\Gamma|^2) \times \eta \tag{2.10}$$

For electrically small antennas, 40 to 50% (−3.9 to −3 dB) of Total Efficiency is usually the acceptable level in the design phase and it is challenging to obtain a fully embedded antenna.

2.1.9 FUNDAMENTAL LIMITATIONS OF ELECTRICALLY SMALL ANTENNAS

In recent years there has been some emphasis on establishing a link between the volume occupied by an electrically small antenna and the Quality Factor (Q) and its bandwidth (BW). Early studies on the fundamental limitations and performance of small antennas were published by Harold Wheeler [5, 6]. Chu determined the lowest possible Quality Factor Q (2.11) of a linearly-polarized omnidirectional antenna and introduced the concept of maximal and practical gain [7]:

$$Q > \frac{1}{k^3a^3} + \frac{1}{ka}, \qquad (2.11)$$

where k is the wave number in free space, $k=2\pi/\lambda$ and a is the radius of the minimum sphere that encloses the antenna.

Following Chu's work and same principle for a circularly-polarized antenna, McLean obtains the smallest possible Q [53]:

$$Q = \frac{1}{2}\left[\left(\frac{1}{ka}\right)^3 + \frac{2}{ka}\right], \qquad (2.12)$$

In 1958, Harrington [8] extended Chu's previous obtained results and defined the natural gain limit for a practical antenna:

$$G_{max} = (ka)^2 + (ka), \qquad (2.13)$$

Based on the theory of the evaluation of the energy stored around the antenna, Fano [9], Collin [55] and Fante [56] derived expressions for the radiation Q which verify Chu's earlier results.

The theoretical investigation on the fundamental limits of an electrically small antenna expands to the relationship between the bandwidth (BW) which is an important parameter for antennas and the Quality Factor (Q). Considering the antenna as a resonant RLC circuit, the Fractional Bandwidth (FBW) can be defined by:

$$FBW = \frac{1}{Q} \qquad (2.14)$$

In practice, an antenna's Fractional Bandwidth is also defined in terms of VSWR. The relationship between Q and VSWR is given by:

$$Q = \frac{VSWR - 1}{FBW\sqrt{VSWR}} \tag{2.15}$$

Fano introduced an expression for the maximum Fractional Bandwidth that can be achieved with a lossless matching network that matches to a resonant circuit (antenna) [9].

$$FBW = \frac{27}{Q}\frac{1}{RL}, \tag{2.16}$$

where RL is the desired return loss.

2.1.10 WHEELER CAP METHOD

One of the simplest, cheapest and most accurate methods to determine the antenna radiation efficiency is the Wheeler Cap method which was presented by H. A. Wheeler in 1959 [6]. Harold Wheeler presented a method for measuring the efficiency of electrically small antennas (ESAs) and it is more accurate compared with other methods, such as, gain comparison, radiometric and pattern integration.

The wheeler cap method is based on the "radiansphere" concept which defines the boundary between the near field and the far field of any small antenna (Fig. 2.7). The distance between the Antenna Under Test (AUT) and the walls of wheeler cap is called "radianlength" and it is λ/2π. The measurement procedure separates into two parts which are without cap (measurement in free space) and with cap and they will respectively determine the radiation resistance and loss resistance (Fig. 2.7). The cap consists of a conducting radiation metal shield.

Wheeler suggested that the placement of the antenna within a conductive cavity would reduce the radiation in the far-field without affecting the losses of the antenna. The method is based on the assumption that the placement of the antenna into the cavity shorts the radiation resistance R_{rad}, while the losses R_{loss} remains constant. Wheeler indicates that the size of the cavity must be very large to avoid the excitations of the near-field and small enough to prevent the emergence of resonances at the frequency of measurement.

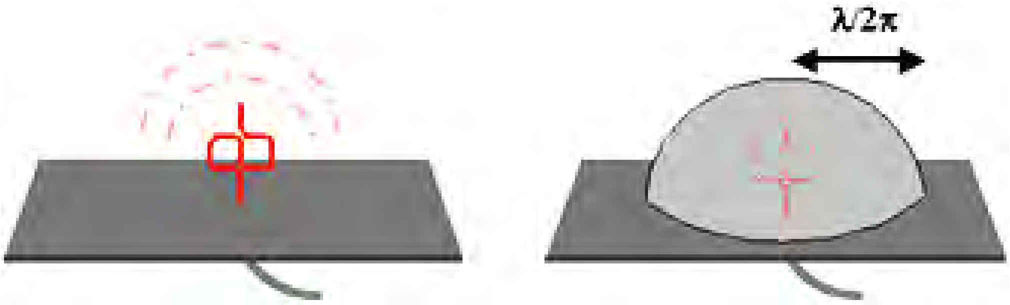

Figure 2.7. Measurement Wheeler Cap method [57].

In the practice, there are many factors that influence the measurement accuracy, such as the thickness of the shield and the reflections by walls, floor and ceiling of the room if the measurements is made indoors. Otherwise, the choice of material and shape of the caps will affect results during measurement.

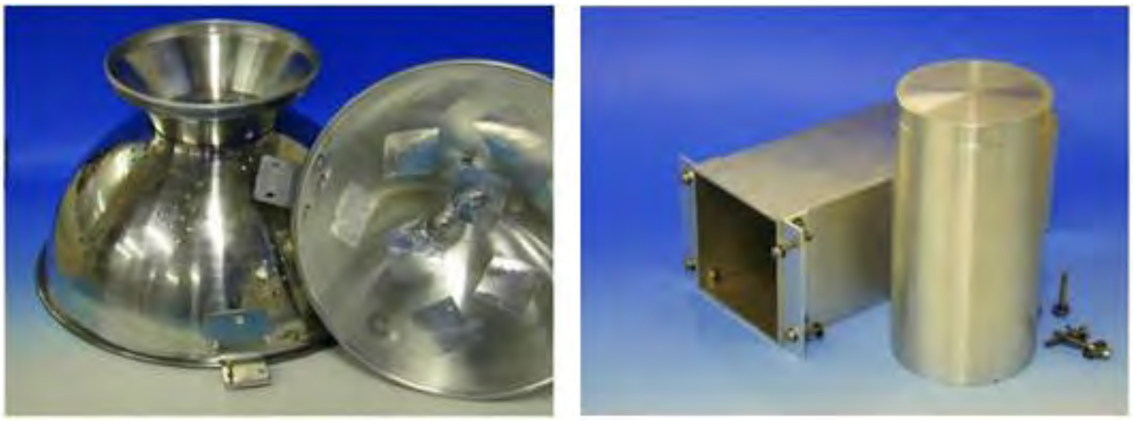

Figure 2.8. Different shapes of shields is used in the Wheeler Cap method [57].

- ***The Constant Power Loss Method***

The radiation efficiency can be defined as radiated power is directly proportional to input power, where input power is the sum of radiated power and loss power. Thus the radiation efficiency, η, is given by:

$$\eta=\frac{P_{rad}}{P_{rad}+P_{loss}}=\frac{I^2R_{rad}}{I^2R_{rad}+I^2R_{loss}}=\frac{R_{rad}}{R_{rad}+R_{loss}}, \qquad (2.17)$$

where

P_{rad} is radiation power of antenna

P_{in} is input power

P_{loss} is loss power (reflection power)

I is the input current

R_{rad} is radiation resistance of antenna

R_{loss} is loss resistance.

With the Wheeler Cap on, the radiation resistance is zero and the antenna reflection coefficient is measured and referred to as $\boldsymbol{S_{11WC}}$.

$$\frac{P_{loss}}{P_{in}} = 1 - S^2_{11WC} \tag{2.18}$$

Without the Wheeler Cap the radiation resistance is that of the antenna radiating into free space and the antenna reflection coefficient is measured and referred to as $\boldsymbol{S_{11FS}}$.

$$\frac{P_{rad}+P_{loss}}{P_{in}} = 1 - S^2_{11FS} \tag{2.19}$$

The antenna efficiency becomes

$$\eta = \frac{P_{rad}}{P_{rad}+P_{loss}} = \frac{(1-|S_{11FS}|^2)-(1-|S_{11WC}|^2)}{(1-|S_{11FS}|^2)} = \frac{|S_{11WC}|^2-|S_{11FS}|^2}{1-|S_{11FS}|^2} \tag{2.20}$$

- ***The Constant Loss Resistor Method***

Also, the antenna reflection coefficient $\boldsymbol{S_{11WC}}$ can be represented as

$$S_{11WC} = (R_{loss}-R_S)/(R_{loss}+R_S) \tag{2.21}$$

The antenna reflection coefficient $\boldsymbol{S_{11FS}}$ can be represented as

$$S_{11FS} = ((R_{rad}+R_{loss})-R_S)/((R_{rad}+R_{loss})+R_S), \tag{2.22}$$

where R_s is the source resistance.

Equations (2.17), (2.21) and (2.22) are transposed to become

$$R_{loss}/R_S = (1+S_{11WC})/(1-S_{11WC}) \tag{2.23}$$

$$((R_{rad}+R_{loss}))/R_S = (1+S_{11FS})/(1-S_{11FS}) \tag{2.24}$$

The radiation efficiency now is

$$\eta = \frac{R_{rad}}{R_{rad}+R_{loss}} = \frac{\left(\frac{1+S_{11FS}}{1-S_{11FS}}\right) - \left(\frac{1+S_{11WC}}{1-S_{11WC}}\right)}{\left(\frac{1+S_{11FS}}{1-S_{11FS}}\right)} = 1 - \frac{(1-S_{11FS})(1+S_{11WC})}{(1+S_{11FS})(1-S_{11WC})} \tag{2.25}$$

2.1.11 COMPARISON WITH OTHER METHODS

The most common method used to measure the radiation efficiency of an antenna is the Gain /Directivity method which is based on anechoic chamber measurements [58]. The advantage of G/D method is the ability for accurate measures. However, this method has some limitations, such as requiring large space for installation [59]. In order to reduce the costs of design and development of an antenna, a simpler and quicker method, such as the Wheeler Cap method, was used [60].

Another method to determine small antennas efficiency in multipath environment is the reverberation chamber (RC) method [61]. A reverberation chamber is a screened room with a maximum of reflections of electromagnetic energy. It operates as a multi-mode cavity with one or two mechanical stirrers to perturb the boundary conditions of the electromagnetic environment or stir the inevitable standing waves. Inside the chamber, two antennas can be allocated; a transmit antenna which excites the chamber and an antenna under test (AUT) with a radiation efficiency which is to be determined [62].

The Wheeler Cap method was used due to the ease to implement, the fast accurate results which are in a very good agreement with simulations.

2.2 SOFTWARE SIMULATION AND MODELLING TOOLS

CST Microwave Studio (CST MWS) [63] is a 3D electromagnetic software simulation tool which is used to design and to simulate the antennas presented in this work. In a comprehensive design environment (CAD interface), CST MWS allows modelling of the antenna geometry, the properties of the materials and the surrounding space. The electromagnetic behaviour of the design can be analysed. The simulation delivers a variety of parameters including far field radiation pattern, gain, the total and radiation efficiency, S-parameter information and VSWR. Electric and magnetic field as well the surface current distribution are also available. The time domain and frequency domain solver are mainly used for the simulations. A unique meshing and sub-gridding technique (Tetrahedral or Hexahedral mesh type) is used for every antenna structure (Fig. 2.9).

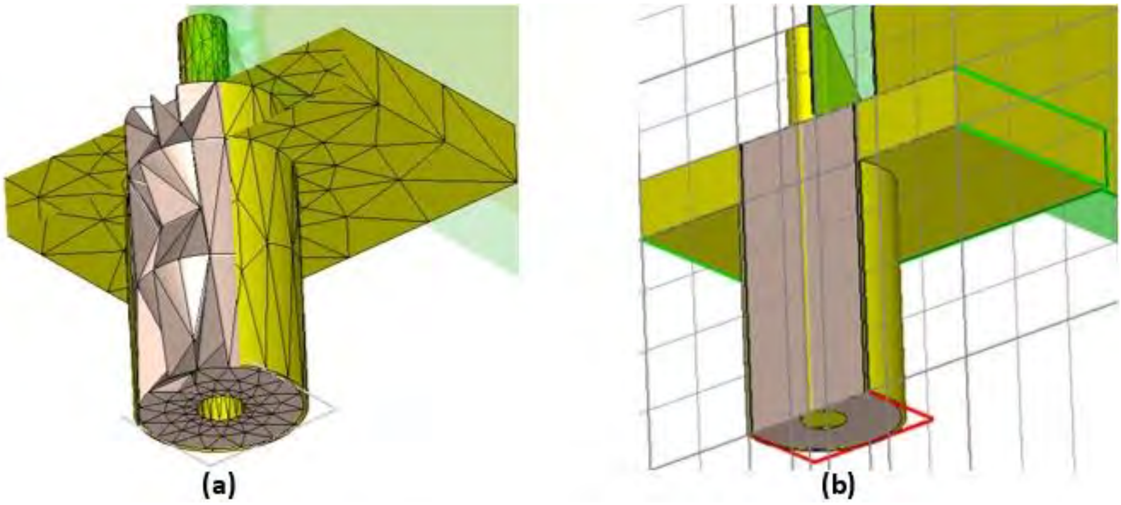

Figure 2.9. The (a) Tetrahedral and the (b) Hexahedral mesh type.

CST MWS also allows exporting of parameter data for further analysis and processing and the geometry data in a format readable by the circuit board plotter software used to prototype the antennas.

2.2.1 TIME & FREQUENCY DOMAIN SOLVER

The Time Domain Solver (TDS) calculates the electromagnetic fields through time at discrete locations and at discrete time samples. It calculates the transmission of energy between various ports or other excitation sources of the designed structure to the boundaries. A Time Domain Solver (TDS) is efficient for High Frequency structures such as connectors, transmission lines and antennas obtaining the broadband frequency behaviour of the simulated device.

The Frequency Domain Solver (FDS) calculates by using a broadband frequency sweep (frequency by frequency) in order to derive the full broadband spectrum from a relatively small number of frequency samples. Frequency Domain Solver (FDS) is suitable for compact structures or devices with a high Quality Factor (Q) value.

CST MWS is based on the Finite Integration Technique (FIT) which was proposed in 1977 by Thomas Weiland [64] and it is a spatial discretization numerical method to solve electromagnetic field problems in time and frequency domains. It delivers the spatial properties of the continuous equations such as conservation of charge and energy.

The basis of this approach is to apply Maxwell equations in integral form suitable for complex geometries. The Finite Integration Technique (FIT) is a method with high flexibility in geometric modelling and boundary handling as well as incorporation of arbitrary material distributions and material properties such as anisotropy, non-linearity and dispersion. Additionally, the use of a three-dimensional (3D) Cartesian mesh combined with a time integration scheme leads to efficient calculation algorithms, which allow simulating real world electromagnetic field problems [65].

2.3 PROTOTYPING

For the manufacturing of prototypes, a LPKF ProMat C60 Circuit Board (PCB) plotter was used to mill prototype designs from a PCB material layer. The substrates used are single or double-sided FR4 laminates with specific properties [66] (Table 2.1). Additionally, thin layers of Brass were used as main components of some prototypes. To facilitate measurements 50 Ω SMA (Sub Miniature version A) connectors were used to excite the prototypes providing good performance from DC to 18 GHz and also good mechanical durability.

Table 2.1. Properties of FR4 and Brass used for antenna prototyping.

Properties	FR4	Brass
ε_r	4.2-4.5±0.2 (1 GHz)	-
μ	1	1
$\tan\delta$	0.02 (1 MHz) – 0.014 (1 GHz)	-
Electric Conductivity	-	2.74×10^7 S m^{-1}
Thermal Conductivity	0.3 K m W^{-1}	109 K m W^{-1}
Thickness	1.52 mm	0.15 mm
Copper Layer Thickness	0.035 mm	-

2.4 MEASUREMENT SETUP

The accuracy of the measurement results are due to the use of the measurement setup which consists of a partially anechoic chamber, the Antenna Under Test (AUT) and a Standard Gain Horn Antenna (SGH) which are both mounted in the chamber (Fig. 2.10).

The Antenna Under Test is placed on the top of a fiberglass mast which is centred on the turntable base which rotates by 360° in Azimuth. Both antennas are connected to a vector network analyzer. The turntable and measurement data from the vector network analyzer are gathered, processed and plotted in a PC using software which controls both devices. The measured radiation patterns are obtained in far field conditions (Fig. 2.11).

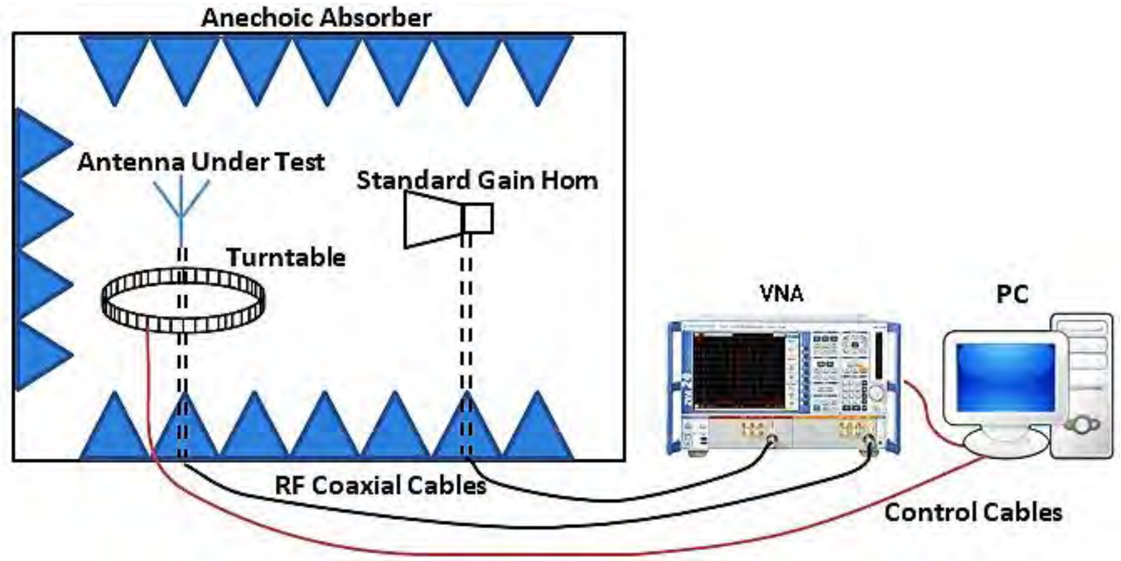
Anechoic Absorber
Antenna Under Test
Standard Gain Horn
Turntable
VNA
PC
RF Coaxial Cables
Control Cables

Figure 2.10. Measurement setup.

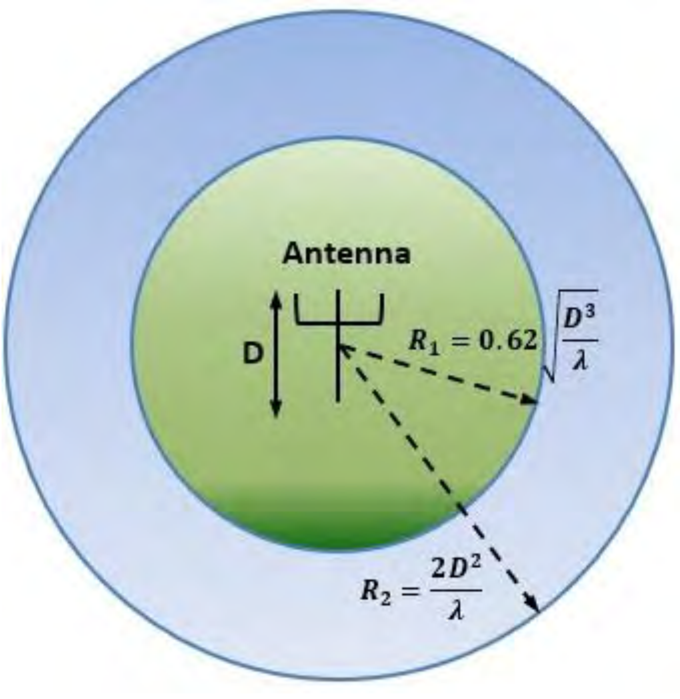
Antenna
D
$R_1 = 0.62\sqrt{\frac{D^3}{\lambda}}$
$R_2 = \frac{2D^2}{\lambda}$

Figure 2.11. Near R1 and Far R2 field boundaries.

2.4.1 VECTOR NETWORK ANALYZER (VNA)

For measuring the performance of the prototype antennas a Vector Network Analyzer (VNA) *Rohde&Schwarz ZVA 24* is used [67, 68]. A Vector Network Analyzer measures the amplitude and phase of the wave quantities and uses these values to calculate a complex S-parameter, VSWR as well as Smith Chart. The real and the imaginary part of the impedance are also available.

The VNA is connected to a PC and can be controlled since all its functions are available to the PC via a network. Proprietary software controls the measurement of radiation pattern.

2.4.2 ANECHOIC CHAMBER

A radio frequency anechoic chamber is an electromagnetic shielded space whose walls are covered with absorbing materials that can absorb almost all of the incident electromagnetic radiation, providing an approximate free space environment. The partial anechoic chamber used in the measurement setup has absorbing materials which were placed in strategic positions on the walls to absorb the incident power and to degrade the signal reflections from the antenna that could impair and affect the accuracy of the measurement results.

The absorbing material used in the partial chamber is *EMERSON & CUMING ECCOSORB* VHP-18 pyramidal carbon loaded urethane foam absorber of 45.7 cm total height and a maximum reflectivity of −30 dB at 500 MHz, −40 dB at 1 GHz, −45 dB at 3 GHz and −50 dB above 5 GHz [69].

2.4.3 STANDARD GAIN ANTENNA (SGA)

Horn antennas of known gain and S-parameters value have been used as reference antennas to measure and to calculate the gain of the antennas under test. The measurement setup uses two standard gain horn antennas, one covering from 1 GHz (3.8 dBi) to 10 GHz (17.7 dBi) and the other from 500 MHz (5.3 dBi) to 2.9 GHz (11.1 dBi). The manufacturer is *SCHWARZBECK* [70].

The calculation of the AUT's gain is made by the Friis transmission equation [1], which relates the power P_r which is delivered to the receiver to the input power of the transmitter P_t:

$$(G_{AUT})_{dB} = 20\log\left(\frac{4\pi R}{\lambda}\right) + 10\log\left(\frac{P_r}{P_0}\right) - (G_{SGH})_{dB}, \qquad (2.26)$$

where $(G_{SGH})_{dB}$ is the known gain of the horn antenna and R is the distance between the two antennas.

2.4.4 COORDINATE SYSTEM

Based on the IEEE Standard Test Procedures for Antennas [71], the antenna's coordinate system presented in the following Figure 2.5. The radiation pattern of an antenna under test which located at the centre of the spherical coordinate system is dependent on the two angular coordinates Θ and Φ. The distance R from the antenna under test (AUT) to the measuring point (SGH) is fixed and only the two angular coordinates (Θ, Φ) are variables in a given radiation pattern.

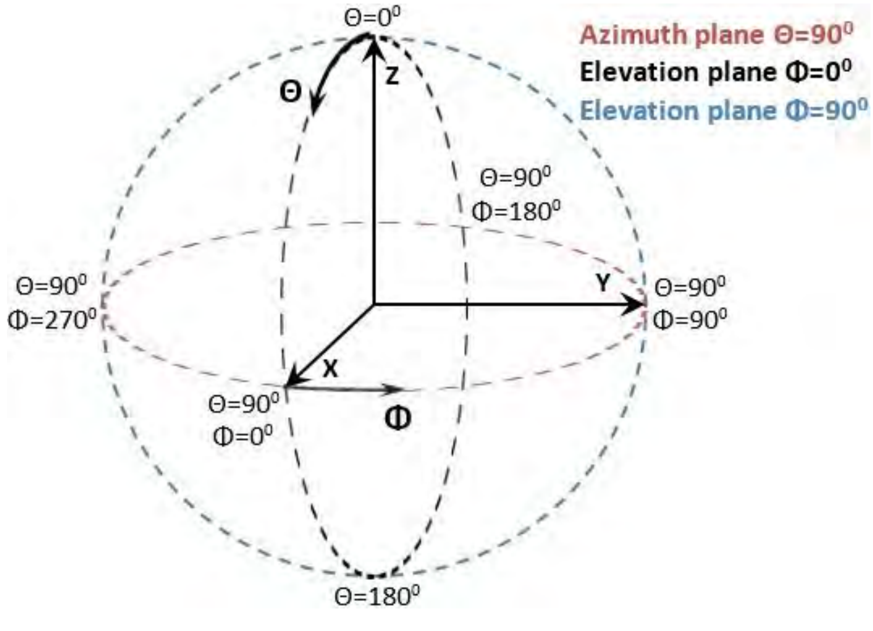
Θ=0⁰
Azimuth plane Θ=90⁰
Elevation plane Φ=0⁰
Elevation plane Φ=90⁰
Θ
Z
Θ=90⁰
Φ=180⁰
Θ=90⁰
Φ=270⁰
Y
Θ=90⁰
Φ=90⁰
X
Θ=90⁰
Φ=0⁰
Φ
Θ=180⁰

Figure 2.12. Standard Spherical Coordinate System.

3. Printed Folded Monopole Antenna

Printed Folded Monopole Antennas [11, 12, 13, 14] are designed with the main advantage of the low profile size and ease of integration to compact size devices, offering a very easy and a low cost prototyping and fabrication.

In this chapter a probe-fed printed folded monopole antenna placed on a conventional rectangular ground plane for WLAN applications is presented and discussed. The dual band antenna is operating at 2.45 GHz and 5.8 GHz. It is a low profile size antenna with omni-directional radiation characteristics in both bands. A parametric study of key geometrical parameters is reported as well as the effect of the ground plane size on the radiation performance.

3.1 ANTENNA DESIGN

The evolution of the folded monopole design can be started from a conventional quarter wavelength (λ/4) PIFA antenna placed on a ground plane. By reducing the length of the extended horizontal arm of the PIFA as it shown in Fig. 3.1 a folded monopole antenna shorted to the ground plane, with dual band performance can be achieved.

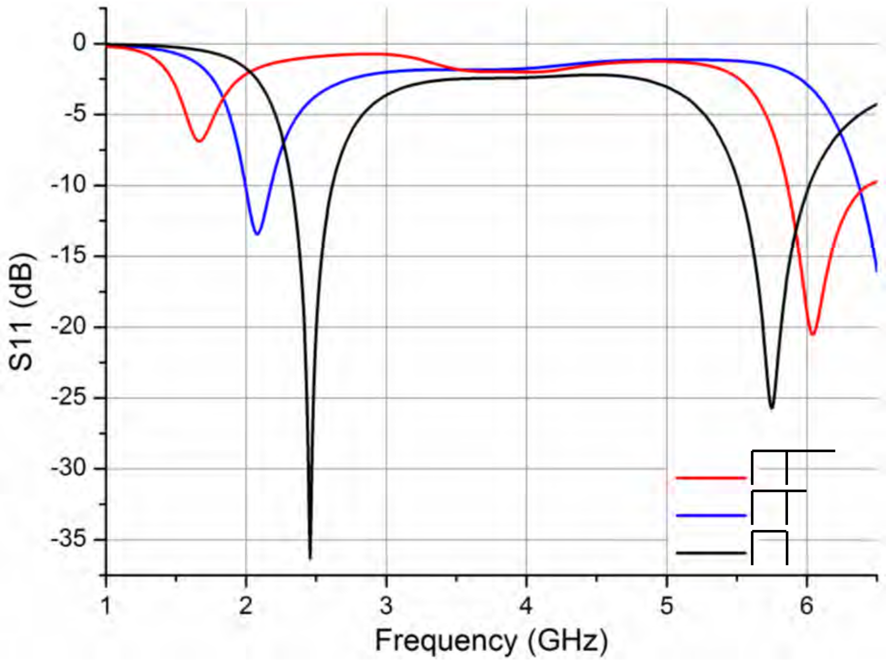

Figure 3.1. Evolution of the folded monopole antenna.

A 70 mm × 70 mm brass ground plane (Table 2.1) was used. The folded monopole with perimetric length of 53.38 mm (≈ $0.44\lambda_0$ with $\lambda_0 = 122.4$ mm at 2.45 GHz) and height of $h = 20.2$ mm (≈ $0.165\lambda_0$) is printed on a single-side FR-4 substrate (Table 2.1) with dimensions 22.2 mm × 17 mm. The width of the shorting and the feeding strip is s =0.5 mm and f = 1.98 mm respectively. The length (l) of the horizontal arm of the antenna is l = 10.5 mm and the width a = 3.57 mm. The monopole is fed via a 50 Ω SMA connector through the ground plane (Fig. 3.2).

From the fabrication point of view, the FR-4 substrate is touched on the metallic ground plane and does not affect the measured and simulated results. The connection between the two parts (FR-4 substrate and ground plane) is via shorting strip which is soldered to it.

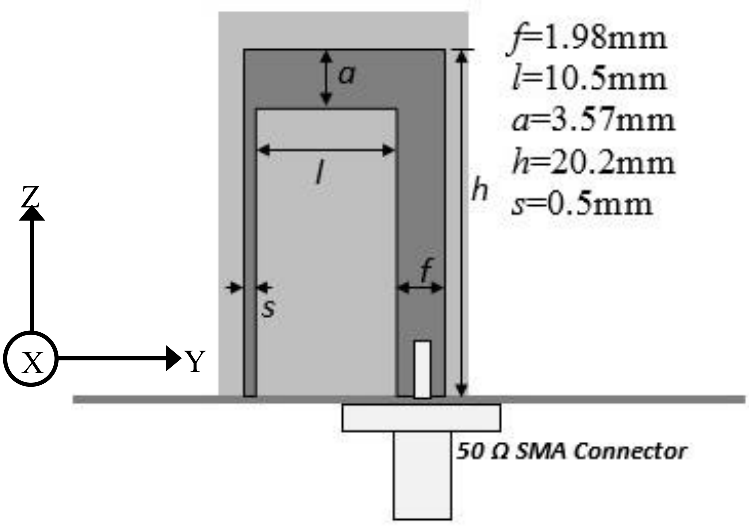
a
l
h
s
f
Z
X
Y
50 Ω SMA Connector

f=1.98mm
l=10.5mm
a=3.57mm
h=20.2mm
s=0.5mm

Figure 3.2. Antenna geometry.

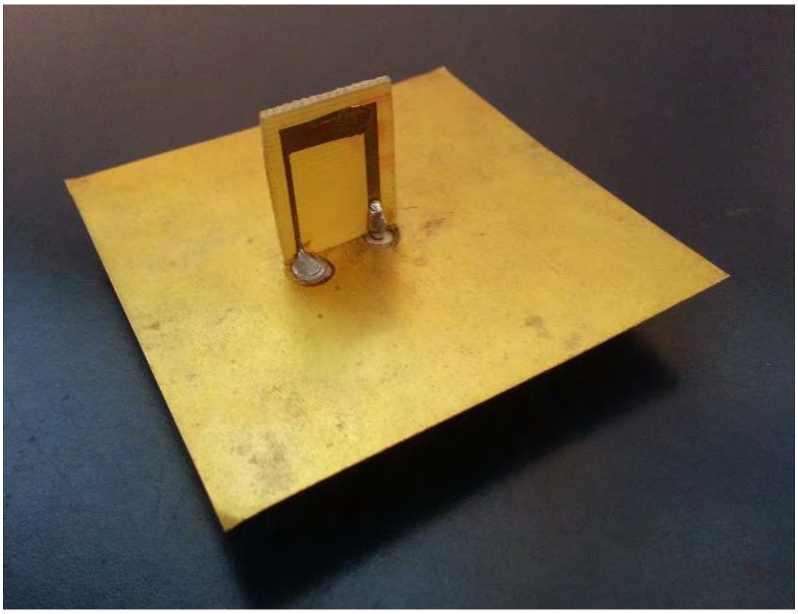

Figure 3.3. Manufactured folded monopole antenna prototype.

3.2 PARAMETRIC STUDY

A parametric investigation of the variation of S_{11} with respect to four geometric parameters of the printed folded monopole shown in Fig. 3.2 is made.

The frequency of operation of the monopole is strongly dependant on its height (*h*). The height (*h*) of the monopole was varied from 16.2 mm to 22.2 mm in 2 mm increments. Fig. 3.4 shows the simulated S_{11} for different values of height (*h*) of the antenna. The results indicate that as the height (*h*) is increased both resonant frequencies shifted upwards with a significant impedance matching improvement for the second resonance. By selecting the proper value of *h* = 20.2 mm can be tuned both resonances to the desired values.

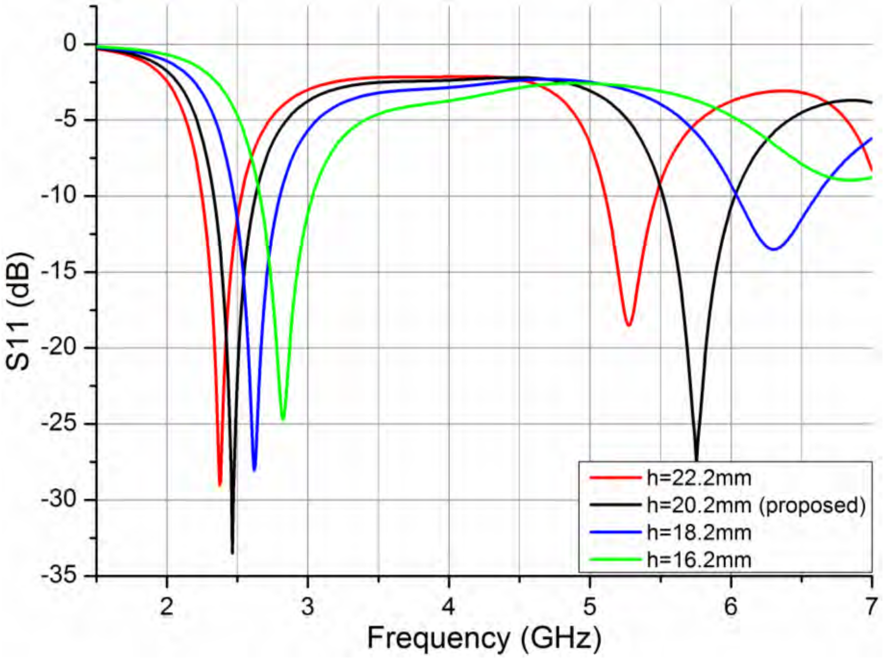

Figure 3.4. The simulated S_{11} dependence on the height (*h*) of the monopole.

The width (*a*) of the horizontal strip of the monopole was found to have a significant influence on the frequency-ratio range of the antenna. The width (*a*) of the horizontal strip was varied from 0.57 mm to 5.07 mm in 1.5 mm increments. From the graph (Fig. 3.5) it is clearly visible that increasing the value (*a*) shifts the second resonance upwards with little effect on the lower resonance. The frequency-ratio between the upper and the lower resonance $F_r = f_u/f_l$ continuously increases from 1.88 to 3.82 as the value of (*a*) changes from 0.57 mm to 5.07 mm. The desirable frequency-ratio for the proposed antenna is 2.36.

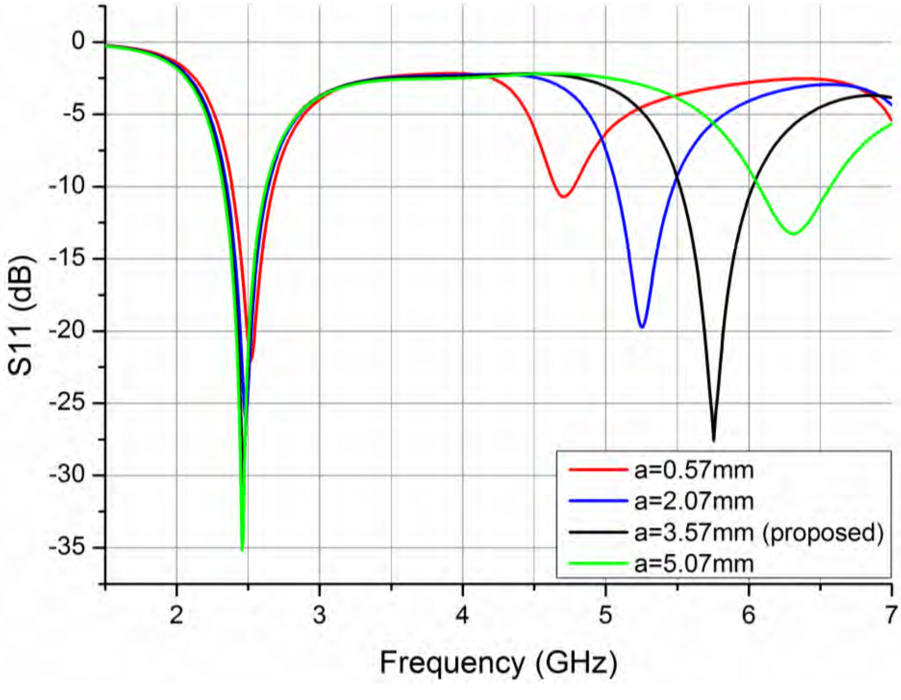

Figure 3.5. The simulated S_{11} dependence on the width (*a*) of the horizontal strip of the monopole.

The horizontal arm length (*l*) is one of the geometrical parameters which is also investigated and varied from 2.5 to 12.5 mm with a step size of 4 mm. The results (Fig. 3.6) clearly indicate that as the length (*l*) increases the lower resonance shifts upwards while in the upper resonance a significant impedance matching improvement occurs as well as a frequency shifting. In that case the proposed value is 10.5 mm.

The dependence of the S_{11} on the width of the feeding (*f*) strip was investigated for the two bands. The results are shown in the following Fig. 3.7. The width of the feeding strip as increased from 0.98 mm to 3.98 mm an upward shift of the first resonance with little effect of the impedance matching on both resonances occurred.

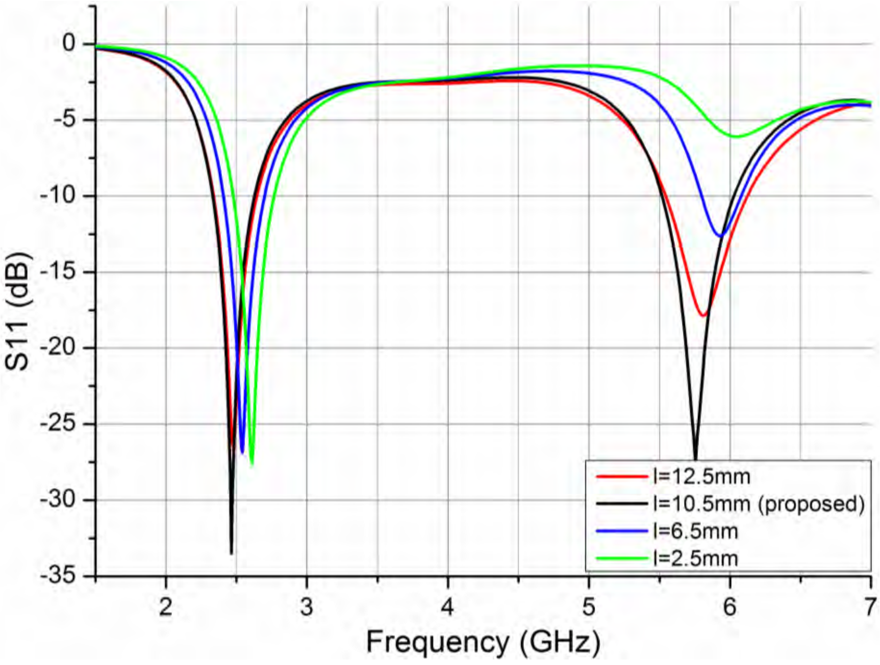

Figure 3.6. The simulated S_{11} dependence on the length (*l*) of the monopole.

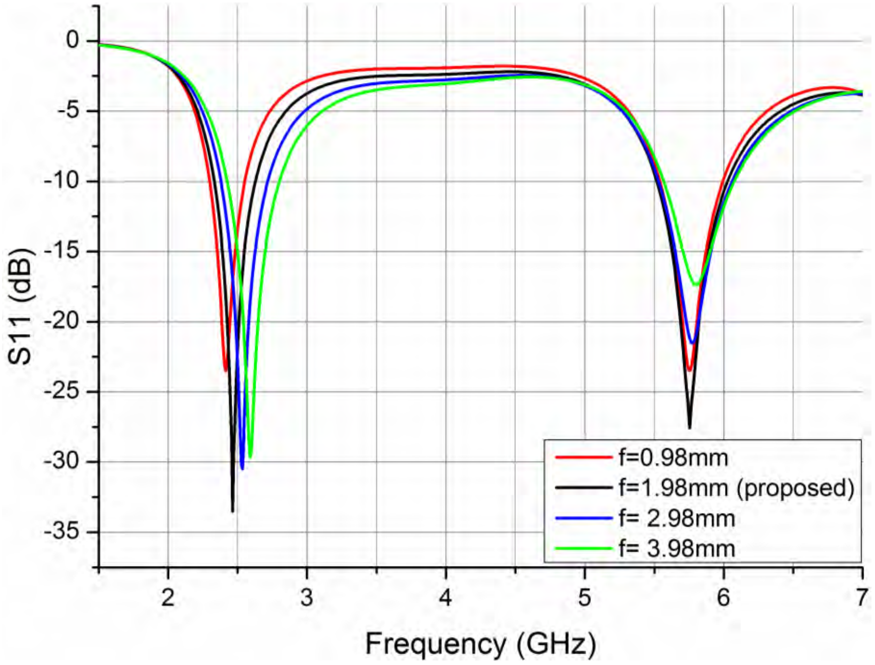

Figure 3.7. The simulated S_{11} dependence on the width (*f*) of the feeding strip of the monopole

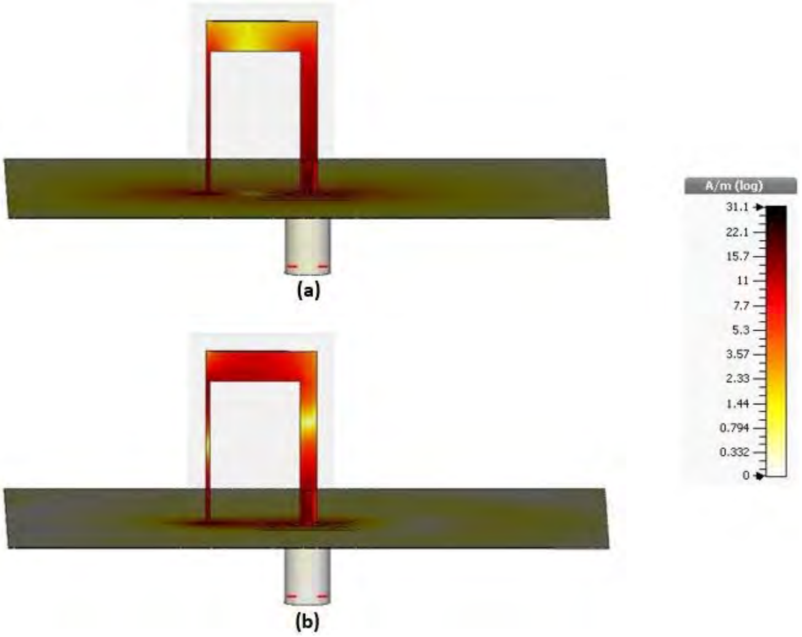
(a)
(b)
A/m (log)
31.1
22.1
15.7
11
7.7
5.3
3.57
2.33
1.44
0.794
0.332
0

Figure 3.8. The current distribution at (a) 2.45 GHz and (b) 5.8 GHz.

The current distribution at both frequencies is shown in Fig. 3.8. For the first resonant frequency (Fig. 3.8(a)) the main part of the monopole which resonates is the feeding and the shorting strips. The electrical length of the two strips is almost 33 mm and it is slightly over $\lambda_0/4$ at 2.45 GHz ($\lambda_0/4$ = 30.6 mm). The distance between the horizontal strip and the ground plane is less than $\lambda_0/4$ and due to the mirror effect, there is a current cancelation between the two parts. For the second resonance (Fig. 3.8(b)) the main current distribution comes from the horizontal part of the monopole and partially from the two vertical strips where the current is strong but with opposite phase. The length of the horizontal strip is almost 13 mm which is slightly over $\lambda_0/4$ at 5.8 GHz (≈ 12.9 mm) and the distance from the ground plane (≈ 16.63 mm) allows the contribution given by these currents to radiation to become dominant.

3.3 GROUND PLANE INVESTIGATION

The size of the ground plane was found to have a significant influence on the radiation performance of the folded monopole antenna. A parametric study of the ground plane size was made and simulations were carried out for ground plane sizes of 30 × 30 mm^2 to 90 × 90 mm^2. Fig. 3.9 shows a plot of the simulated S_{11} results as a function of ground plane size. From the results it becomes obvious that the lower resonant frequency can be effectively controlled (impedance matching and frequency shifting) by the size of the ground plane. The upper resonance is unaffected by the ground plane size which is electrically large and almost twice wavelength (99 mm ≈ $2\lambda_0$) at 5.8 GHz (λ_0 ≈ 51.7 mm).

Fig. 3.10 shows a plot of the simulated total efficiency (%) for various ground plane sizes. The efficiency for the lower resonance is very high (more than 92%) for its centre frequency *f₀* for each ground plane size. For the upper resonance the total efficiency is stable without notable changes. The simulated maximum value of the total efficiency at 2.45 and 5.8 GHz is 98% and 90% respectively for ground plane size of 70 × 70 mm.

Radiation patterns cuts (x-y, z-x and z-y) were simulated using CST MWS and are presented in Figs. 3.11 and 3.12 for both frequencies in 10 dB/division scaled plots. The patterns are shown for ground plane sizes of 30 × 30 mm^2, 50 × 50 mm^2, 70 × 70 mm^2 and 90 × 90 mm^2.

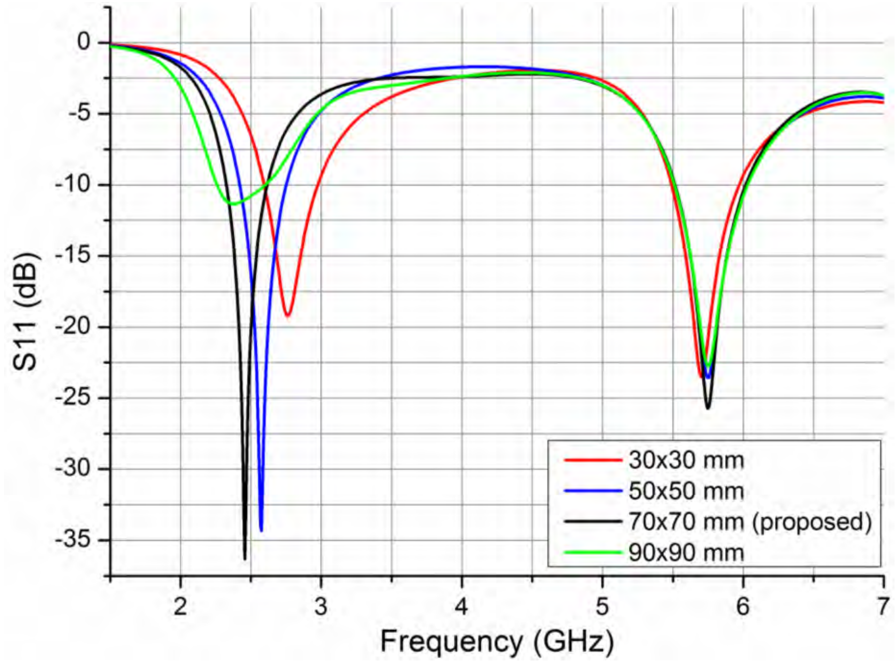

Figure 3.9. The simulated S_{11} dependence on the size of the ground plane.

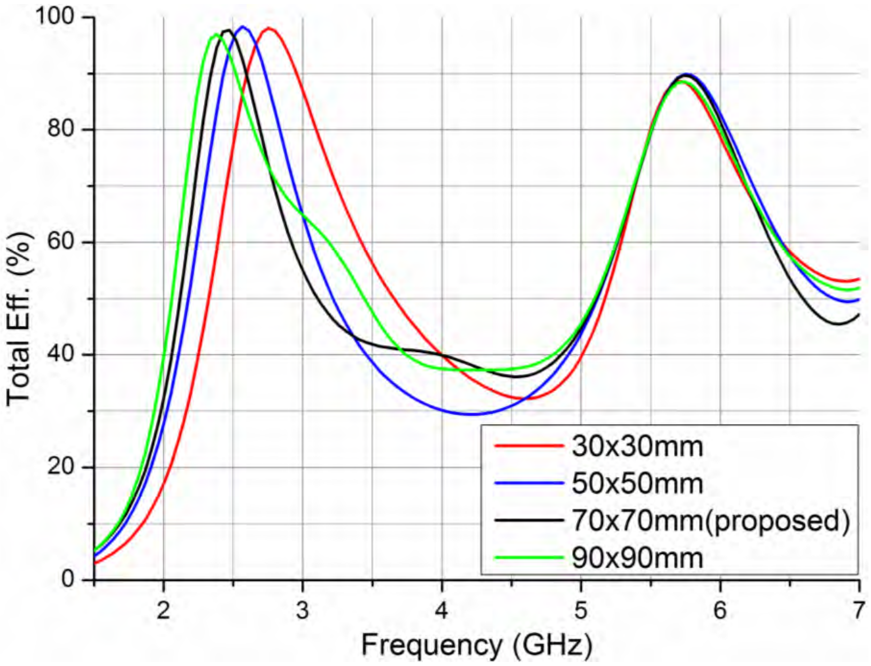

Figure 3.10. Simulated Total Efficiency for various ground plane sizes.

For the azimuth plane (x–y) the radiation patterns are omnidirectional for both frequencies with a small variation of gain as the ground plane size varies. The two elevation radiation patterns (z–x) and (z–y) at the lower resonant frequency are found to be more dipole-like with two visible nulls at θ = 0° and θ = 180° as the ground plane size is increased from 30 mm × 30 mm to 90 mm × 90 mm.

As the size of the ground plane is increased the gain at θ = 30° and θ = 330° is increased. For the upper resonant frequency, in both elevation planes (z–x) and (z–y) (Fig. 3.12) the gain below the ground plane (θ = 180°) is decreased and respectively increased above the ground plane (θ = 0°) as the size increases.

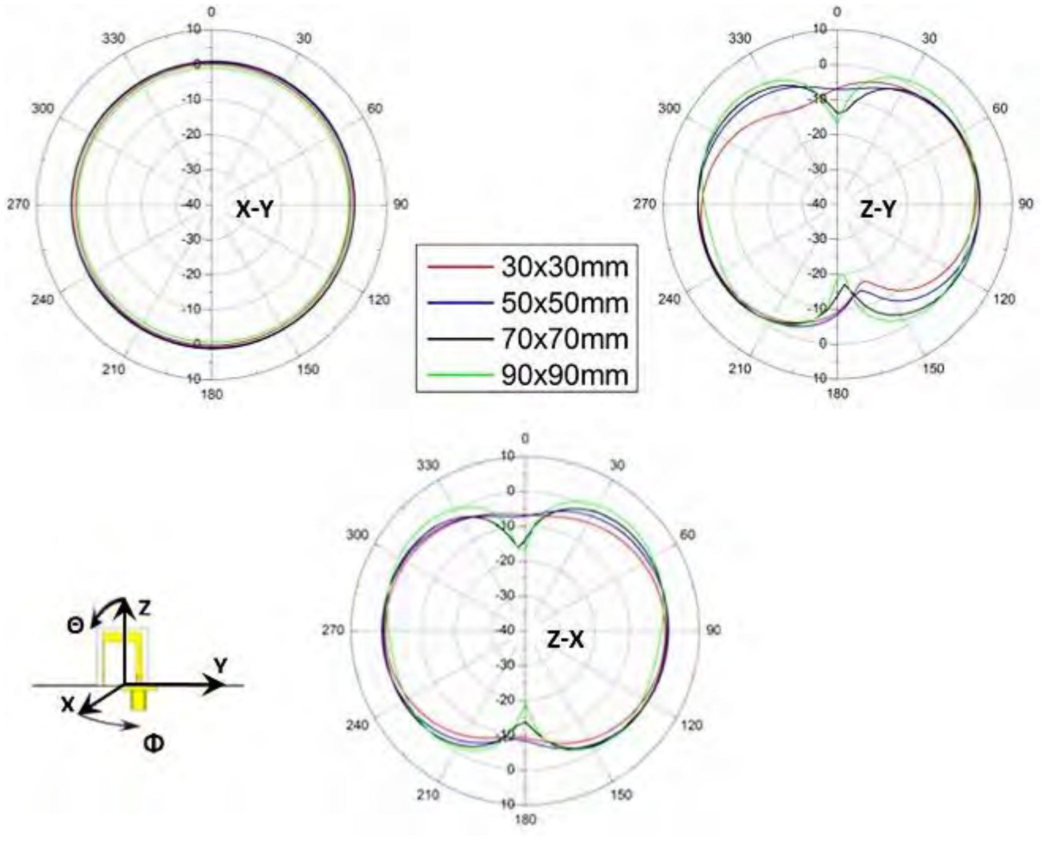

Figure 3.11. Simulated radiation patterns for various ground plane sizes at 2.45 GHz.

To investigate the installed performance of the proposed antenna, a simulation model is created (Fig. 3.13) with the antenna in a plastic enclosure made from Plexiglas (εr = 3.6, tanδ = 0.001 and thickness = 3 mm) with outer dimensions 190 × 190 × 40 mm^3 (L × W × H). Two scenarios were evaluated, with the antenna located in the centre and in the corner of the plastic enclosure as shown in the Fig. 3.13.

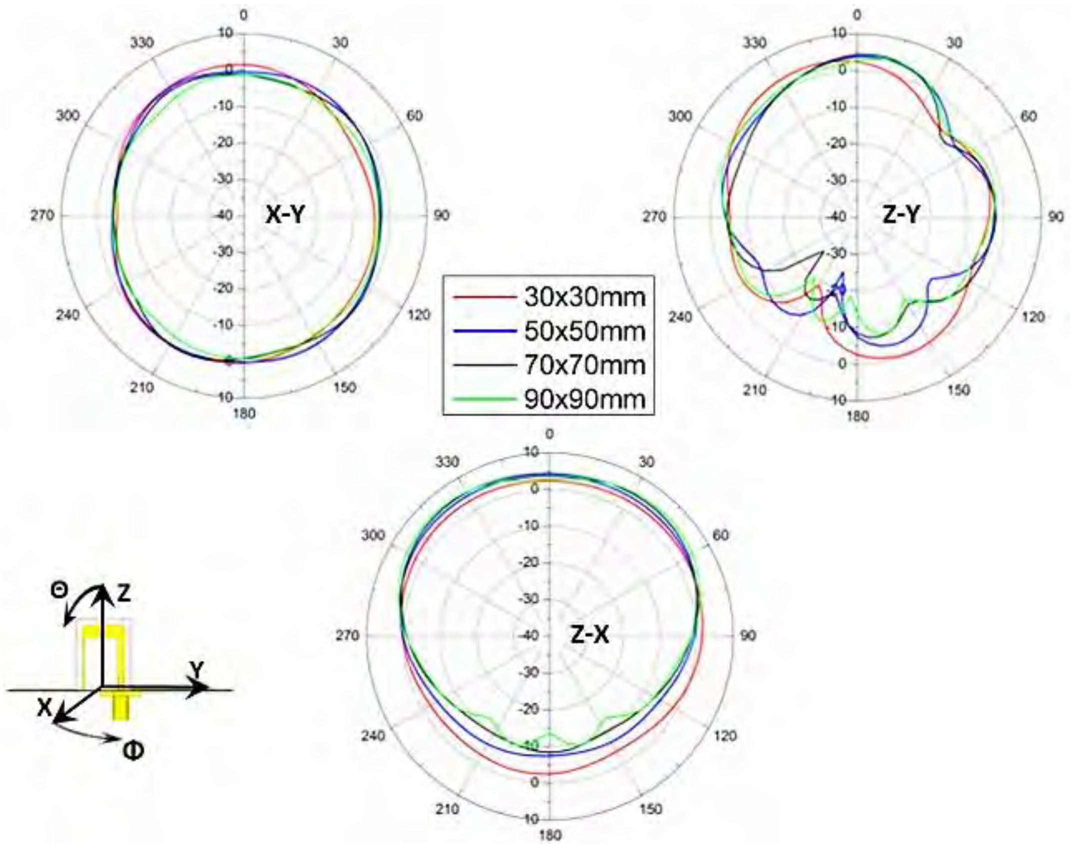

Figure 3.12. Simulated radiation patterns for various ground plane sizes at 5.8 GHz.

Fig. 3.14 show the simulated S_{11} for the two different positions. From inspection of the results, the first resonant frequency shifts slightly downwards due to the wall permittivity loading and this becomes more acute when the monopole is closer to the wall surface (corner). For the case that the antenna is located in the centre of the enclosure the resonant frequency shifted downwards by 0.01 GHz and for the corner located case the frequency shifted downwards by 0.05 GHz.

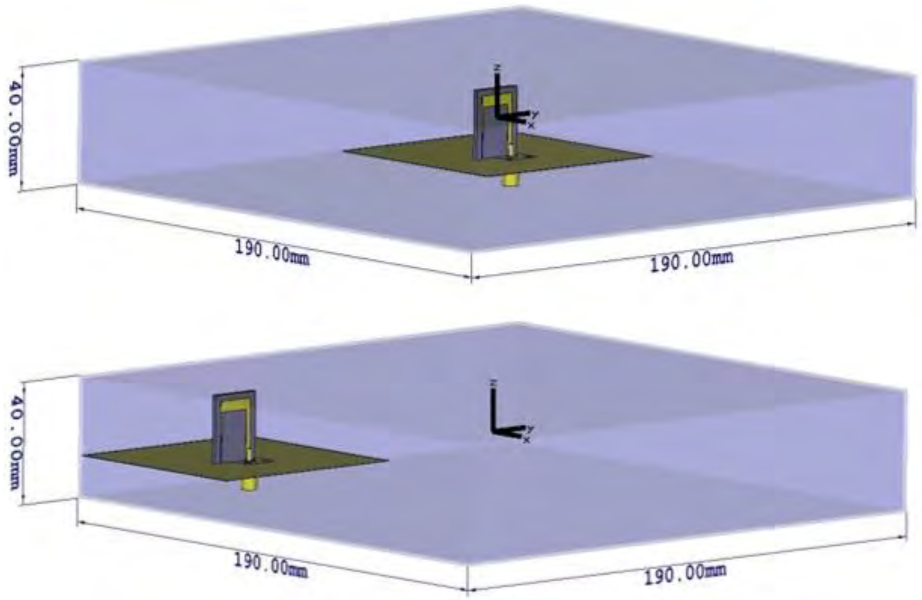
40.00mm
190.00mm
190.00mm
40.00mm
190.00mm
190.00mm

Figure 3.13. The simulation model with the antenna located in the (a) centre of the plastic enclosure and the antenna located in the (b) corner of the plastic enclosure.

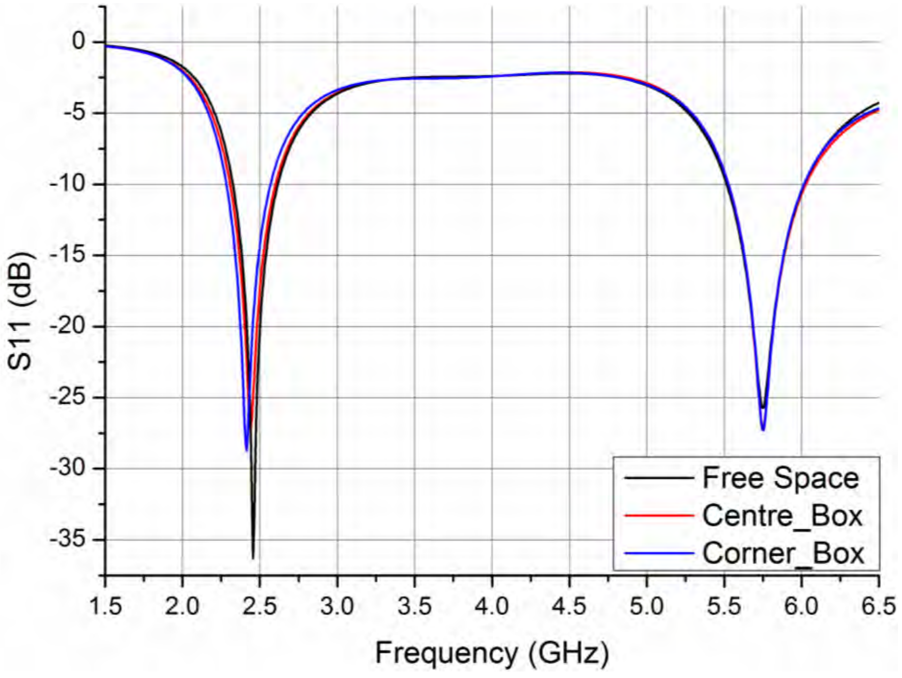

Figure 3.14. Simulated S_{11} for two different locations of the antenna into the plastic enclosure.

3.4 SIMULATED AND MEASURED RESULTS

The simulated and measured S_{11} for the folded monopole with the 70 × 70 mm^2 ground plane is shown in Fig. 3.15. The simulated and measured results are in good agreement. The lower band has a −10 dB impedance bandwidth of 300 MHz (2.34 − 2.64 GHz) for the simulated results and 280 MHz (2.32 − 2.6 GHz) for the measured results. For the upper band the simulated results provide 530 MHz (5.56 − 6.09 GHz) bandwidth, while the measured results provide 500 MHz (5.61 − 6.11 GHz) bandwidth. The fractional bandwidth for the first band is 11.4% and for the second 8.6%.

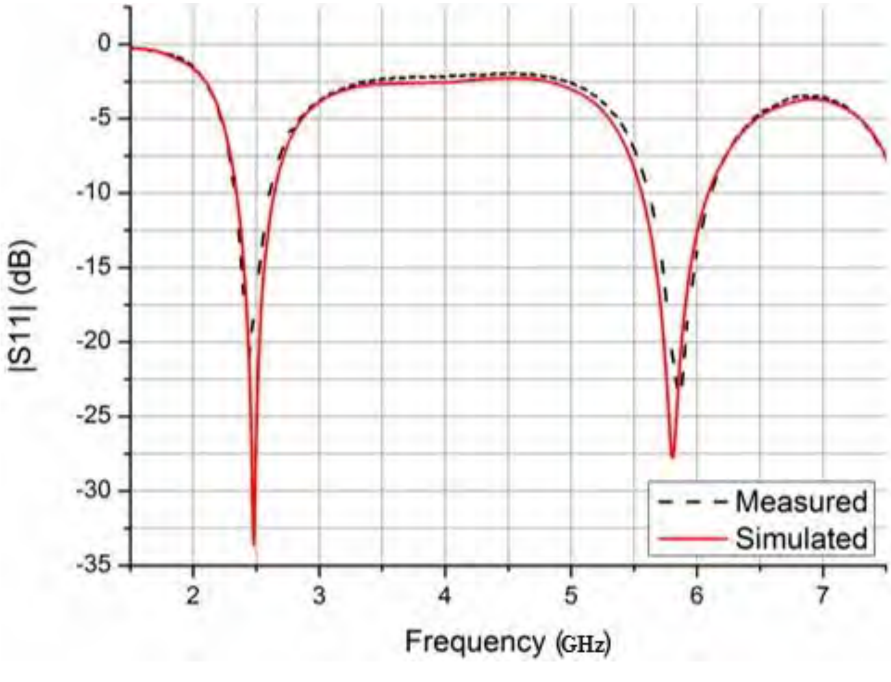

Figure 3.15. Simulated and measured S_{11} for the printed folded monopole antenna.

The simulated and measured azimuth (y–x) and elevation (z–x) plane radiation pattern at both frequencies are illustrated in Figs. 3.17 and 3.18. All the measured gain patterns are illustrated against the simulated patterns in 10 dB/division scaled plots. For each radiation pattern the cross-polar and the co-polar components are presented together.

From the obtained results it is seen that at the lower frequency (Fig. 3.17) the omnidirectional pattern provides good polarization discrimination in the azimuth (y–x) plane. Cross-polar components in the higher frequency (Fig. 3.18) in both planes are due to the greater flow of current in the horizontal strip of the monopole. The measured peak realized gain is 2.0 dBi for the first resonance and 5.7 dBi for the second resonance.

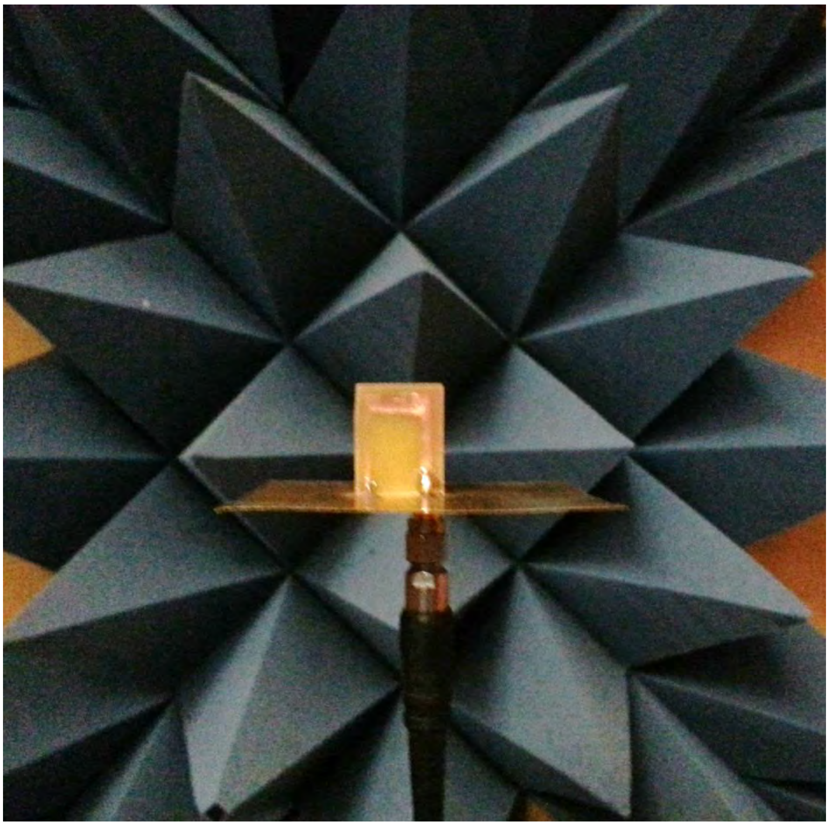

Figure 3.16. The printed folded monopole antenna in the anechoic chamber.

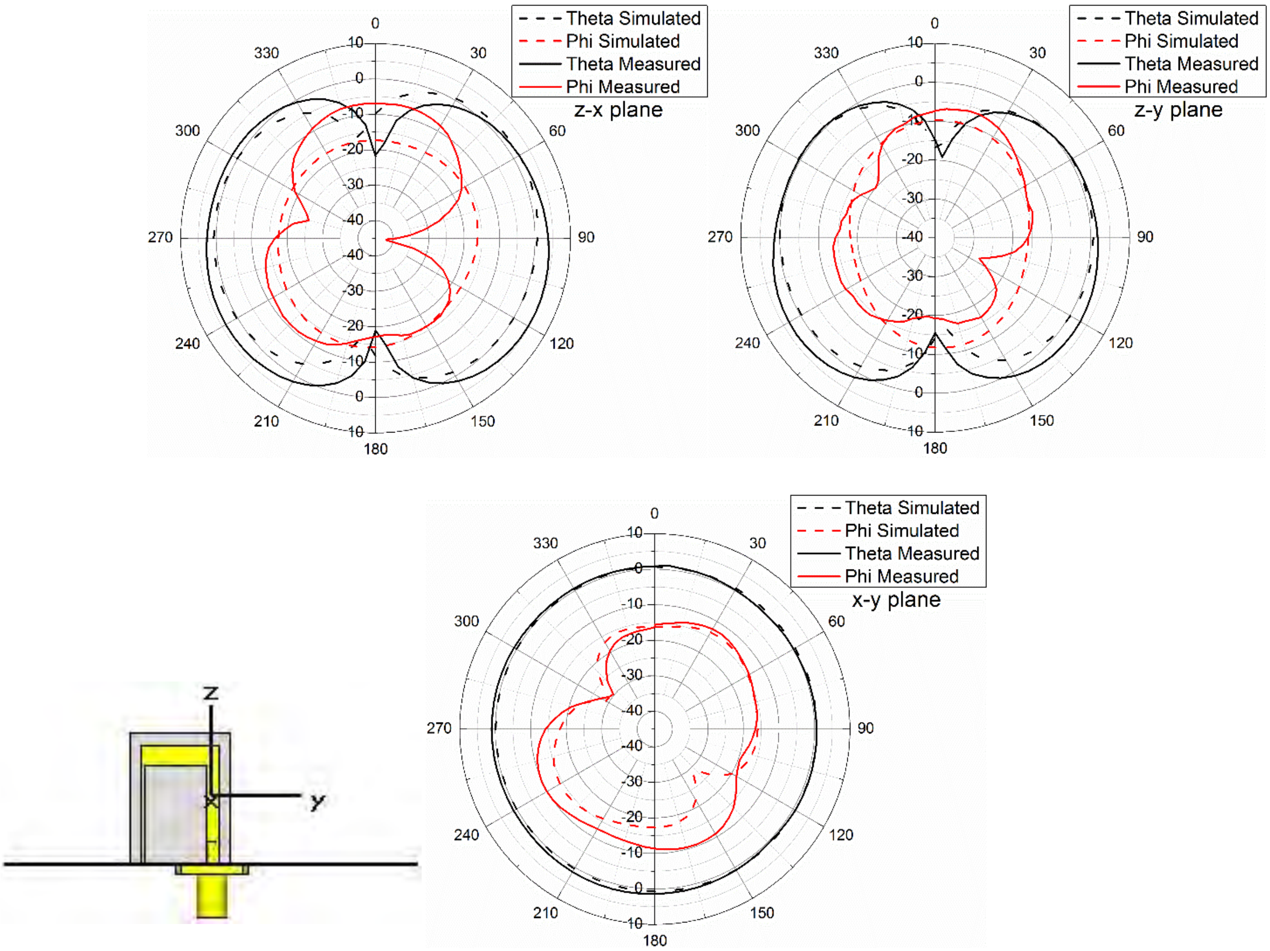

Figure 3.17. Measured and simulated radiation patterns at 2.45 GHz.

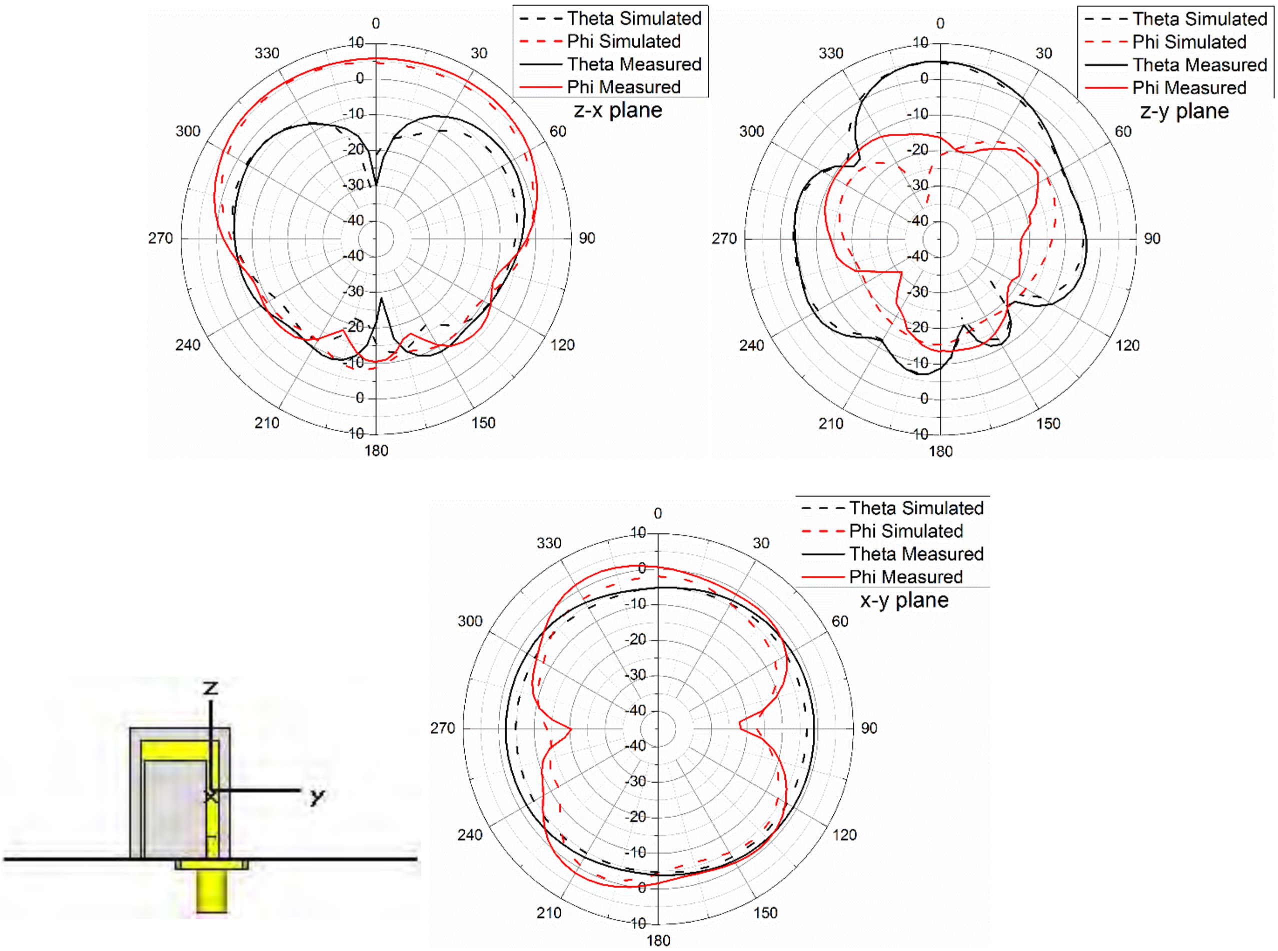

Figure 3.18. Measured and simulated radiation patterns at 5.8 GHz.

3.5 CONCLUSIONS

This chapter presented the study of the performance of a printed folded monopole antenna in the WLAN frequency range which is placed on a conventional rectangular ground plane. A parametric investigation of key geometrical parameters of the monopole as well as the effect of each parameter to the general radiation performance of the monopole is reported. The folded monopole is studied also in terms of the ground plane size. The obtained results showed that the ground plane as an integral part of the antenna effectively controls the first resonant frequency (impedance matching and frequency shifting), improving also the total efficiency over the operating frequency range. The effect of the ground plane on radiation pattern as its size increased is turned to be a more dipole-like at the E–plane for the first resonance. For the second resonance when the ground plane size is increased the antenna becomes more directive at $\theta = 0°$ direction and less directive at $\theta = 180°$. The proposed antenna provides good omnidirectional radiation characteristic at both bands, is low profile, with additional benefit of a large frequency-ratio range which is easy to control.

4. Planar Inverted-F Antennas

Planar Inverted-F Antennas [32, 33, 34, 35, 36] are generally used in the portable devices market due to their low-cost and low profile architecture. They provide broad band and omnidirectional radiation properties.

This chapter describes the implementation of two Planar Inverted-F antennas (PIFAs) which can be used for machine-to-machine (M2M) applications in a broad frequency range. Performance metrics include impedance bandwidth and radiation patterns.

In the first section a dual-band Planar Inverted-F antenna covering the lower and the upper LTE band is described. The second section introduces a dual PIFA antenna which is more compact and covers the whole lower LTE band and partially the upper LTE band. A parametric investigation of key geometrical parameters of both antennas is given as well as the effect of the ground plane size on the general radiation performance.

4.1 DUAL BAND LTE PIFA FOR M2M APPLICATIONS

A dual band PIFA antenna was designed for machine-to-machine (M2M) applications to enable broadband functionality at both LTE frequency bands 690 − 960 MHz and 1710 − 2690 MHz. It is required to meet standards specifications of −6 dB S_{11} and 50%

total efficiency or greater across the proposed frequency range. The antenna size is suitable to integrate in terminal devices for mesh networking applications, automotive and vehicles and generally for indoor and outdoor assets.

4.1.1 ANTENNA DESIGN

The geometry and coordinate system of the antenna is shown in Fig. 4.1. A 178 mm × 247 mm ground plane consists of a single sided FR-4 substrate (Table 2.1). The front and the rear part of the PIFA antenna (Fig. 4.1(a), (b)) are printed on FR-4 layers (Table 2.1) with dimensions l = 67 mm (≈ $0.18\lambda_0$ with λ_0 = 481.15 mm at 623.5 MHz) mm × h = 30 mm (≈ $0.08\lambda_0$). The top part of the PIFA antenna consists of a thin brass layer (Table 2.1) with dimensions l = 67 mm × w=33 mm (≈ $0.09\lambda_0$). The PIFA is fed via a 50 Ω SMA connector through the ground plane to a feeding strip of width f = 2.7 mm which is located a = 20.5 mm from the left edge of the antenna. The fold-over metallization lengths for the front and back sections are d = 23 mm and b = 12 mm respectively. The shorting strip has width (c) of 7.4 mm and is printed on a thin FR-4 layer (thickness = 0.2 mm) it touches the inner part of the FR-4 rear side layer and connects to the top face of the antenna. The shorting strip is located 20 mm (s) away from the left edge of the antenna.

In Fig. 4.2 a general view of the manufactured dual band LTE PIFA prototype is demonstrated. Several fabrication details can be considered as typical for the considered antenna prototype. The soldering of the top plate with the two vertical FR-4 layers was very crucial as well as the connection of the very thin shorting strip with the brass top plate and the ground plane was challenging and precise execution was needed.

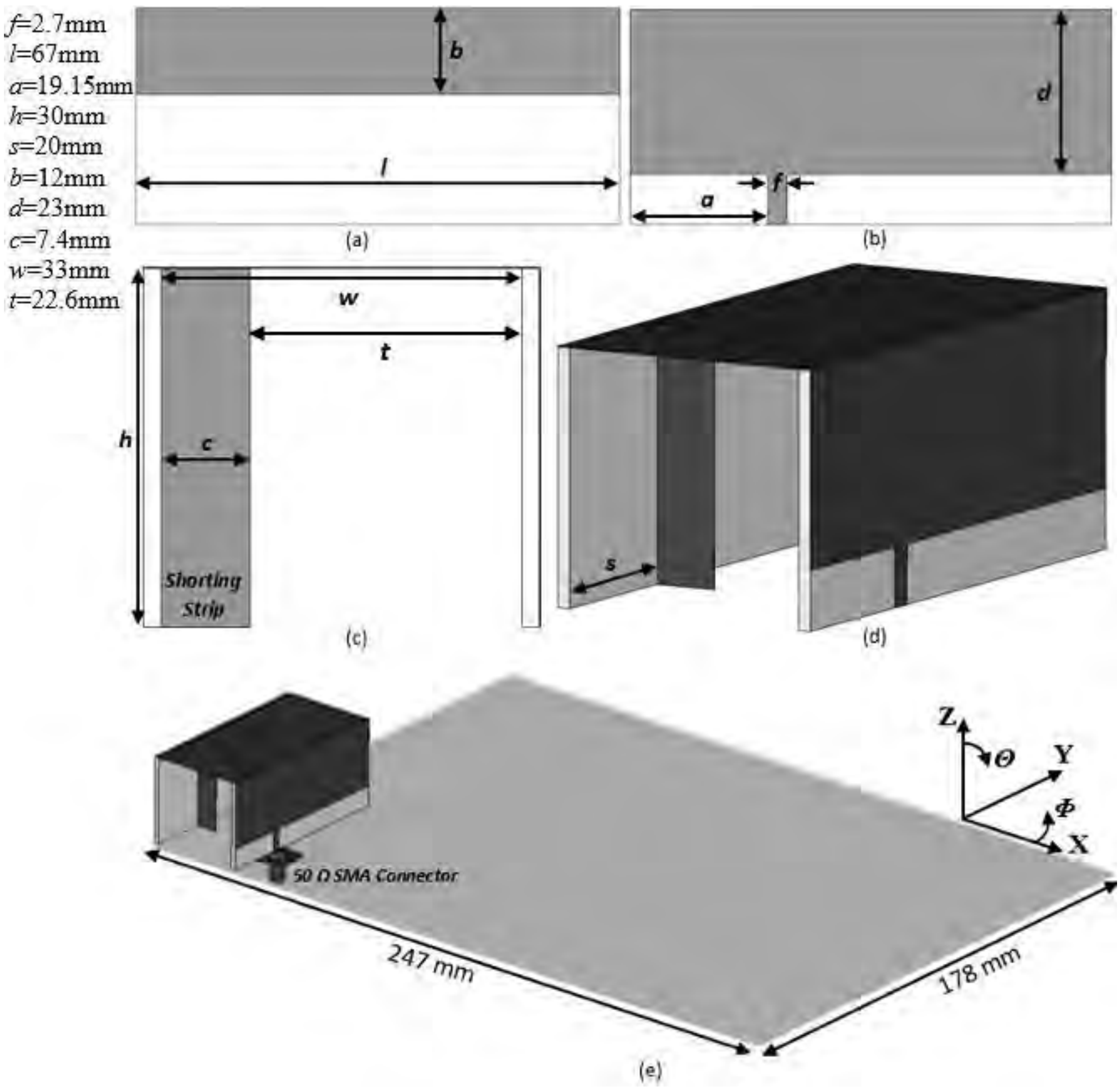
f=2.7mm
l=67mm
a=19.15mm
h=30mm
s=20mm
b=12mm
d=23mm
c=7.4mm
w=33mm
t=22.6mm
b
l
(a)
d
f
a
(b)
w
t
h
c
Shorting Strip
(c)
s
(d)
Z
Θ
Y
Φ
X
50 Ω SMA Connector
247 mm
178 mm
(e)

Figure 4.1. Rear view (a), Front view (b), Left view (c), General view (d), CST simulation model and the coordinate system (e).

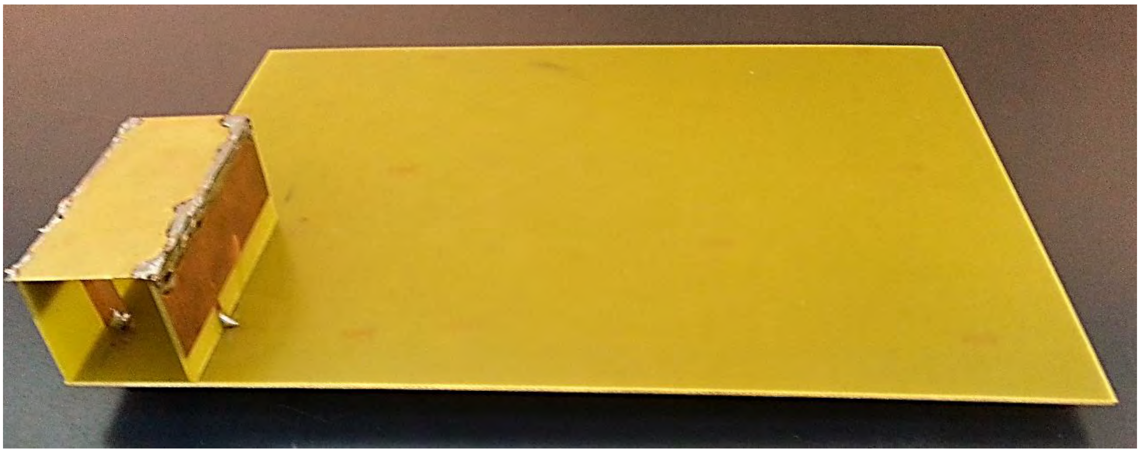

Figure 4.2. Manufactured dual band LTE PIFA prototype.

4.1.2 PARAMETRIC STUDY

A parametric study of five key geometrical parameters of the antenna is made. The position of feeding strip (*a*) is optimised to obtain the desired resonance bands. From the S_{11} (Fig. 4.2) it is seen that moving the feeding strip of the PIFA antenna from the left to the right side shifts the first resonance upwards and effectively controls the impedance matching of both bands. The optimised value of the parameter (*a*) for the proposed antenna is 19.15 mm.

The operating frequency of the antenna is strongly dependent on its height (*h*). The height (*h*) is varied from 25 mm to 40 mm in 5 mm increments. As the height increases (Fig. 4.3), the impedance matching at the first resonance improves. For the second and third resonance as the height increases, the impedance matching deteriorates. The optimized value of the height (*h*) for the proposed antenna is 30 mm, providing good matching for both LTE bands.

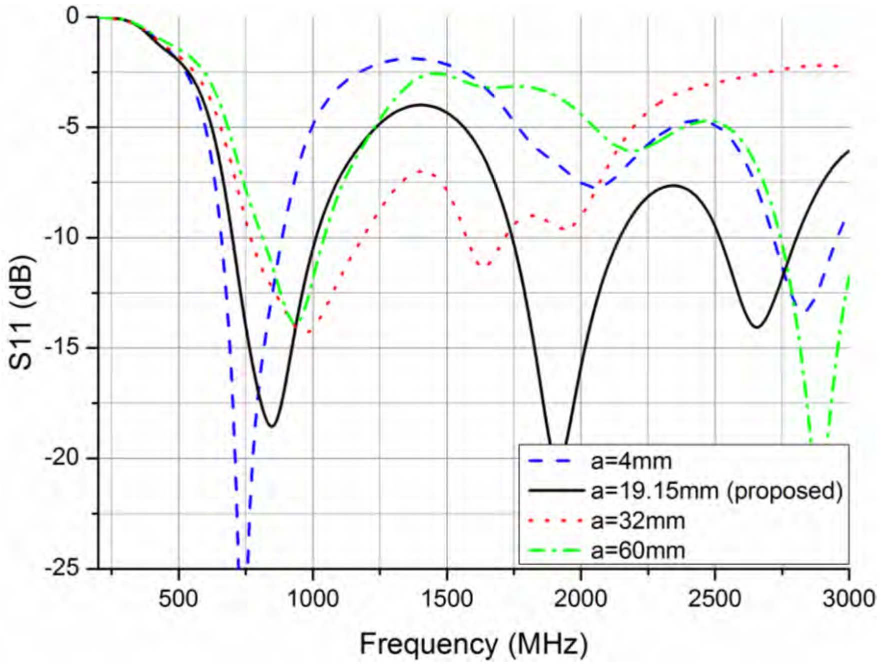

Figure 4.3. The simulated S_{11} dependence on the feeding strip position (*a*) of the PIFA.

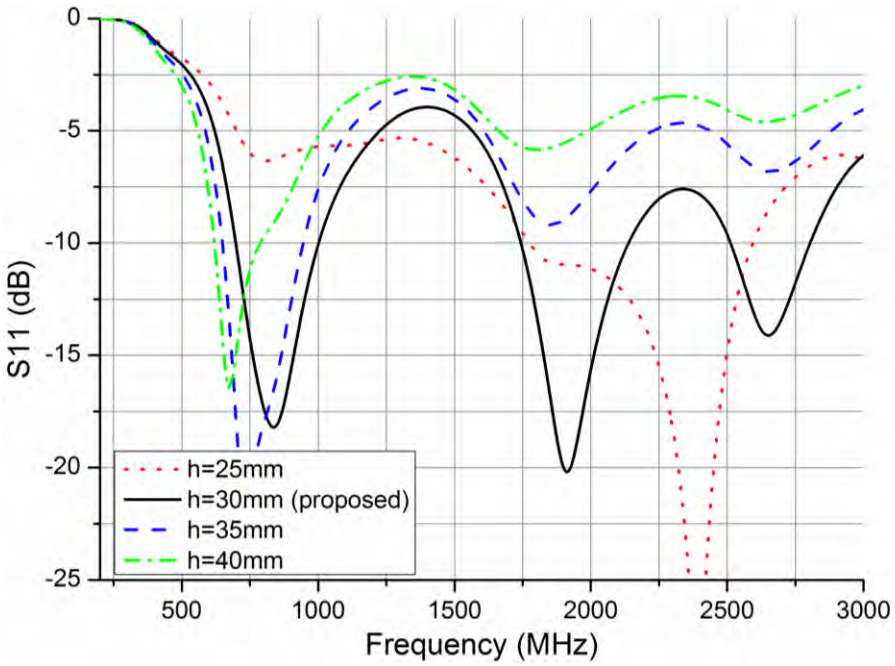

Figure 4.4. The simulated S_{11} dependence on the height (*h*) of the PIFA.

The height (*b*) of the back folded section is investigated. In Fig. 4.4 simulated results for the parameter (*b*) are shown. As the value of the back folded section (*b*) increases from 5 mm to 27 mm, the second band shifts downwards with a little effect on the lower resonance. A proposed value for the PIFA length (*b*) was chosen to be 12 mm.

The height of the front folded section (*d*) is also investigated. From the obtained results (Fig. 4.5) it is clear that the impedance matching of the second resonance is improved as the parameter (*d*) is increased. However, the impedance matching of the first resonance is also strongly dependent on this parameter. The proposed value of the parameter (*d*) is 23 mm.

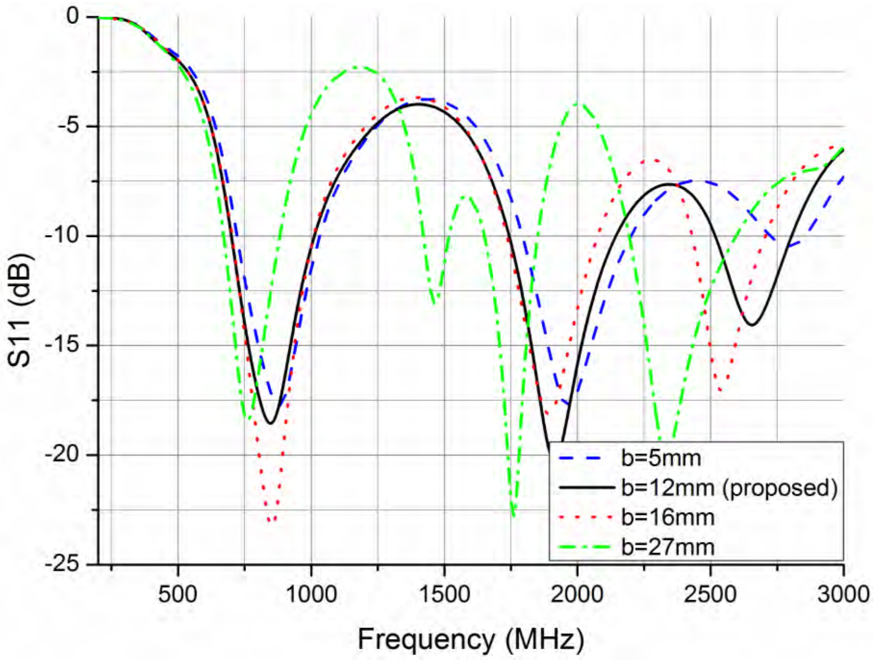

Figure 4.5. The simulated S_{11} dependence on the length of the rear folded section (*b*) of the PIFA.

Another way to obtain the desirable frequency bands is to define the position (*s*) of the shorting strip. The positon of the shorting strip is varied from the left side (0 mm) of the antenna to the right side (67 mm). The obtained results clearly show that all resonant frequencies can be controlled (impedance matching and frequency shifting) by varying the parameter (*s*). By considering the results of Fig. 4.6 the best value of the length (*s*) is 20 mm.

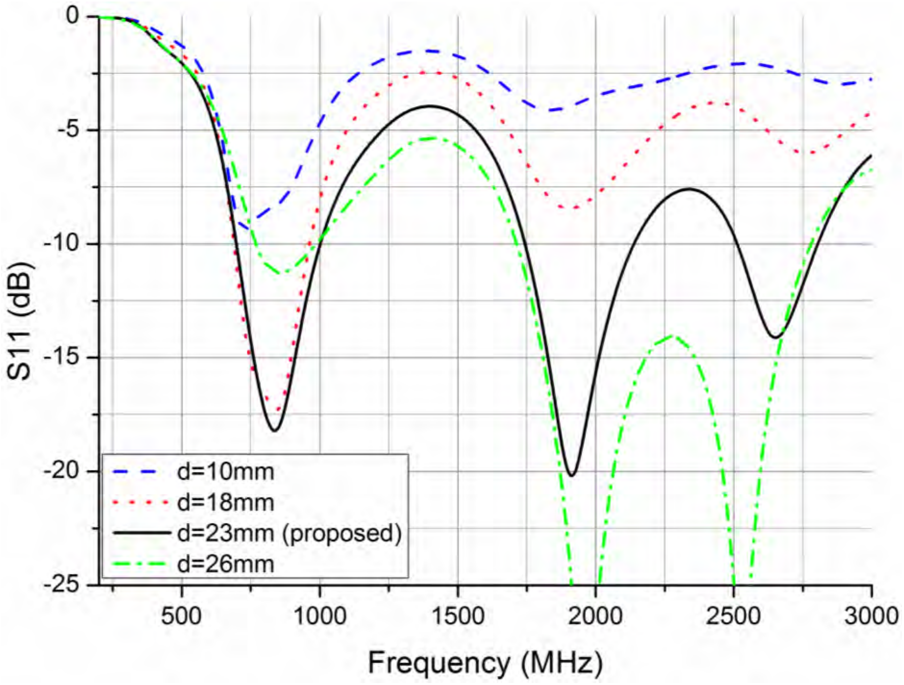

Figure 4.6. The simulated S_{11} dependence on the length of the front folded section (*d*).

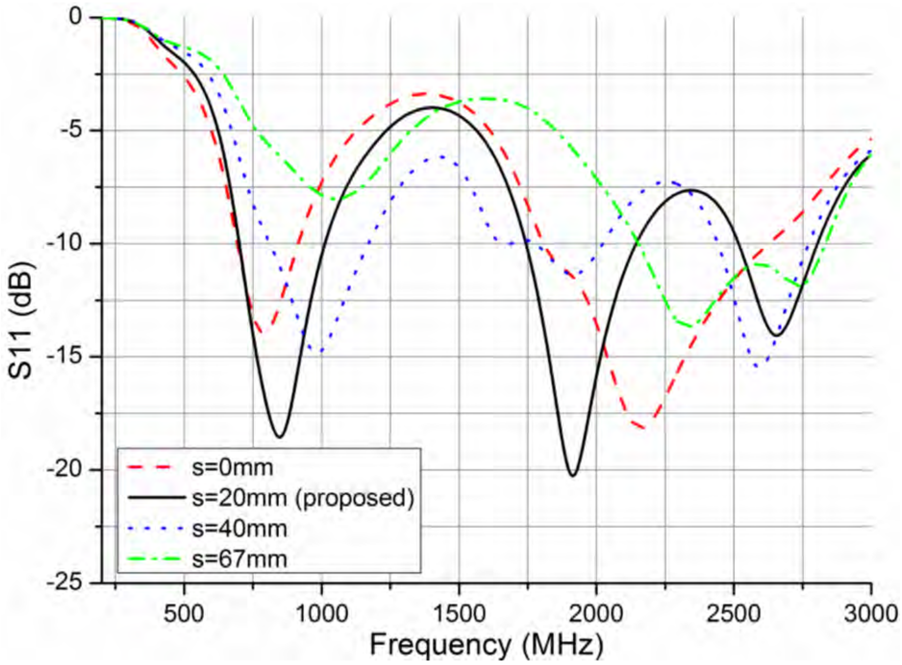

Figure4.7. The simulated S_{11} dependence on the shorting strip position (*s*) of the PIFA.

4.1.3 GROUND PLANE INVESTIGATION

The position of the PIFA antenna on the ground plane (Fig. 4.7) as well as the ground plane size, has a significant influence on the antenna performance [72]. A parametric study of the PIFA position on the ground plane demonstrates that the optimum position is the bottom left corner of the ground plane. For the y-axis the antenna is moved from the bottom left corner (0 mm) of the ground plane to the top right corner (111 mm) with a step size of 37 mm. For the x-axis the antenna is moved from the bottom left corner (0 mm) to the bottom left corner (210 mm) with a step of 70 mm.

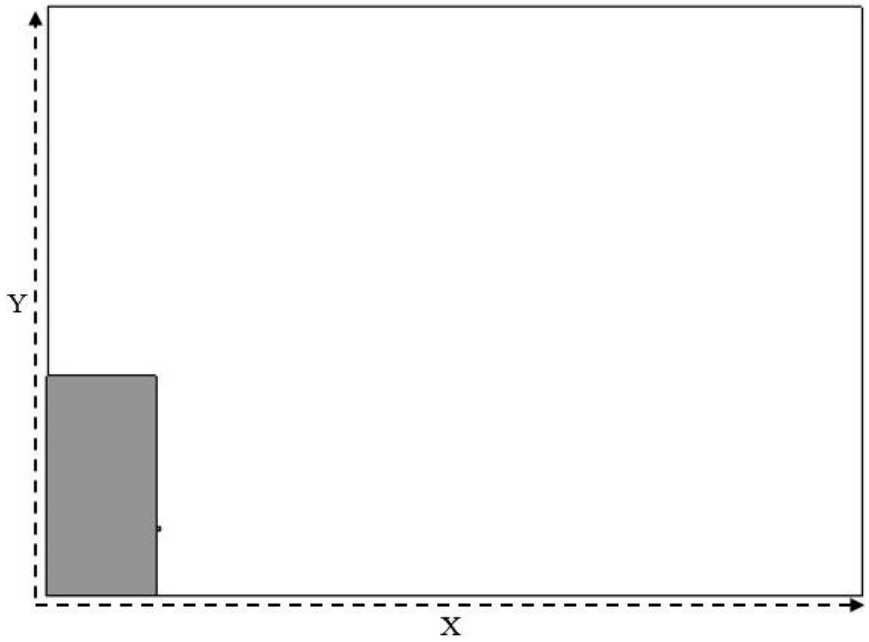

Figure 4.8. The parameters X and Y which define the position of the PIFA antenna.

Figs. 4.9 and 4.10 show that the position of the PIFA along the edge greatly influences the impedance matching and the bandwidth of the lower resonances of the antenna. The position of the PIFA also effects the radiation pattern and the gain (Fig. 4.11). This can be explained by the finite ground plane having distinct characteristic modes depending on its size and the frequency [73]. The surface current of the ground plane for three different positions of the PIFA at 825 MHz is simulated using CST MWS and is presented in Fig. 4.8. When the PIFA is positioned in the bottom right corner of the ground plane a strong current distribution across the right side of the ground plane and especially at the two corners is excited. The ground plane excitation modes is turned to be a more dipole-like. On the other hand, when the PIFA is positioned in the middle of the bottom side and the right side of the ground plane the current distribution is weak without significant contribution to the antenna radiation performance. In that case, the ground plane excitation modes is more monopole-like.

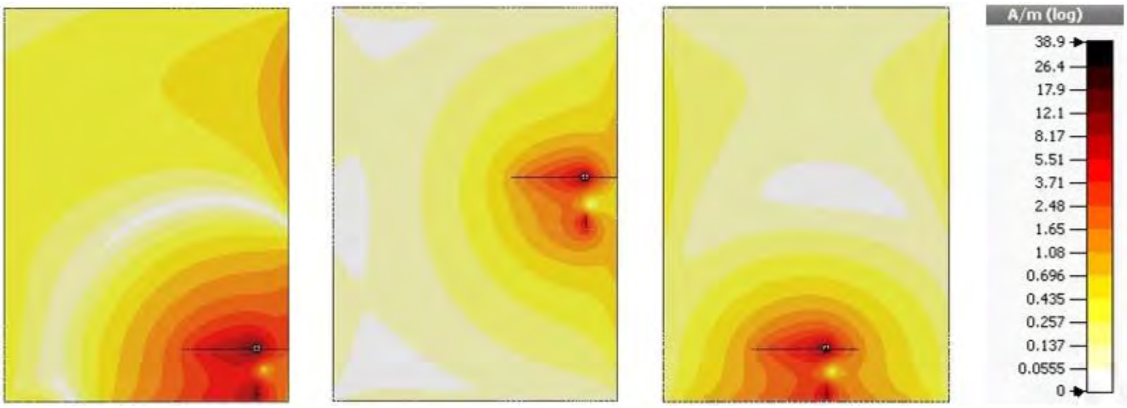
A/m (log)
38.9
26.4
17.9
12.1
8.17
5.51
3.71
2.48
1.65
1.08
0.696
0.435
0.257
0.137
0.0555
0

Figure 4.9. Ground plane current distribution at 825 MHz for three different position of the PIFA.

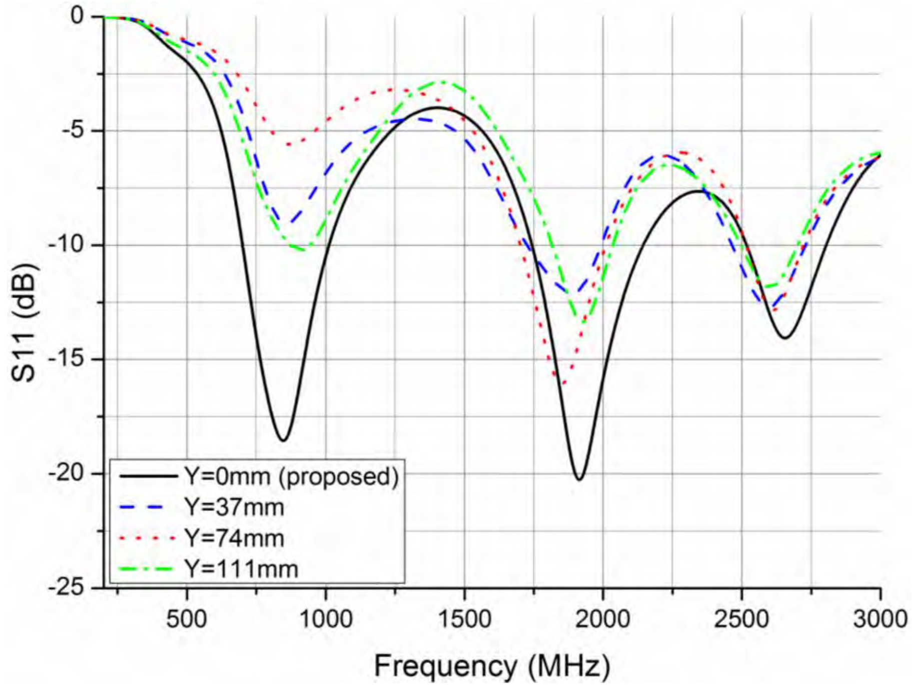

Figure 4.10. Simulated S_{11} when moving the antenna along y-axis.

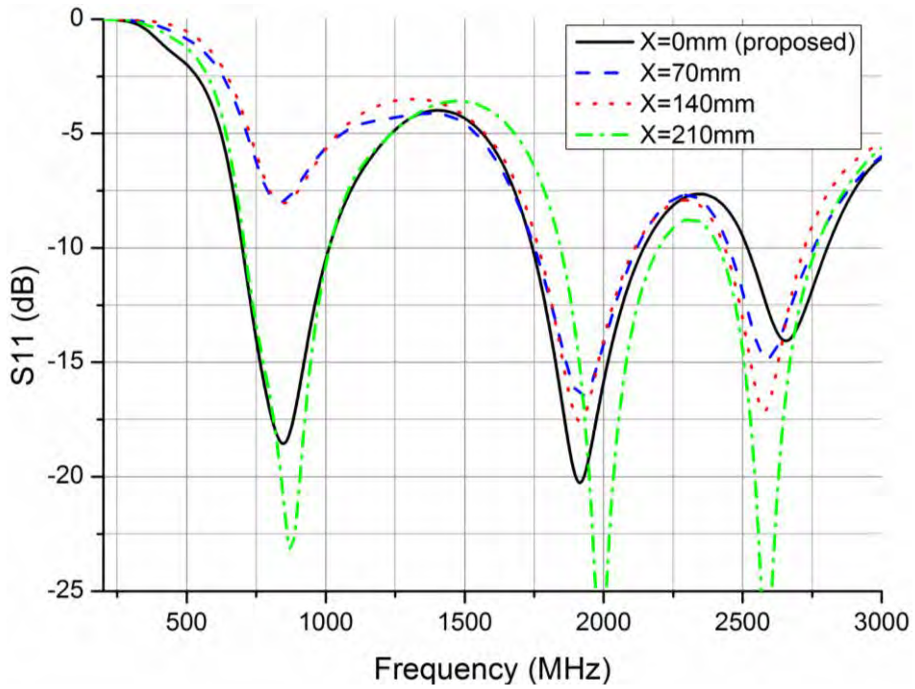

Figure 4.11. Simulated S_{11} when moving the antenna along x-axis.

A parametric study of the ground plane size was made and simulations were carried out for ground plane sizes of 133 × 178 mm^2 to 237 × 346 mm^2. From the graph (Fig. 4.13) it is clearly visible that the antenna is not heavily dependent on the ground plane size due its electrical large size (< 0.84λ_0 at 825 MHz) and does not significantly affect the three resonances.

Fig. 4.14 shows a plot of the simulated total efficiency (%) for various ground plane sizes. From the obtained results it is understood that the total efficiency of the antenna does not change significantly as the ground plane size varies.

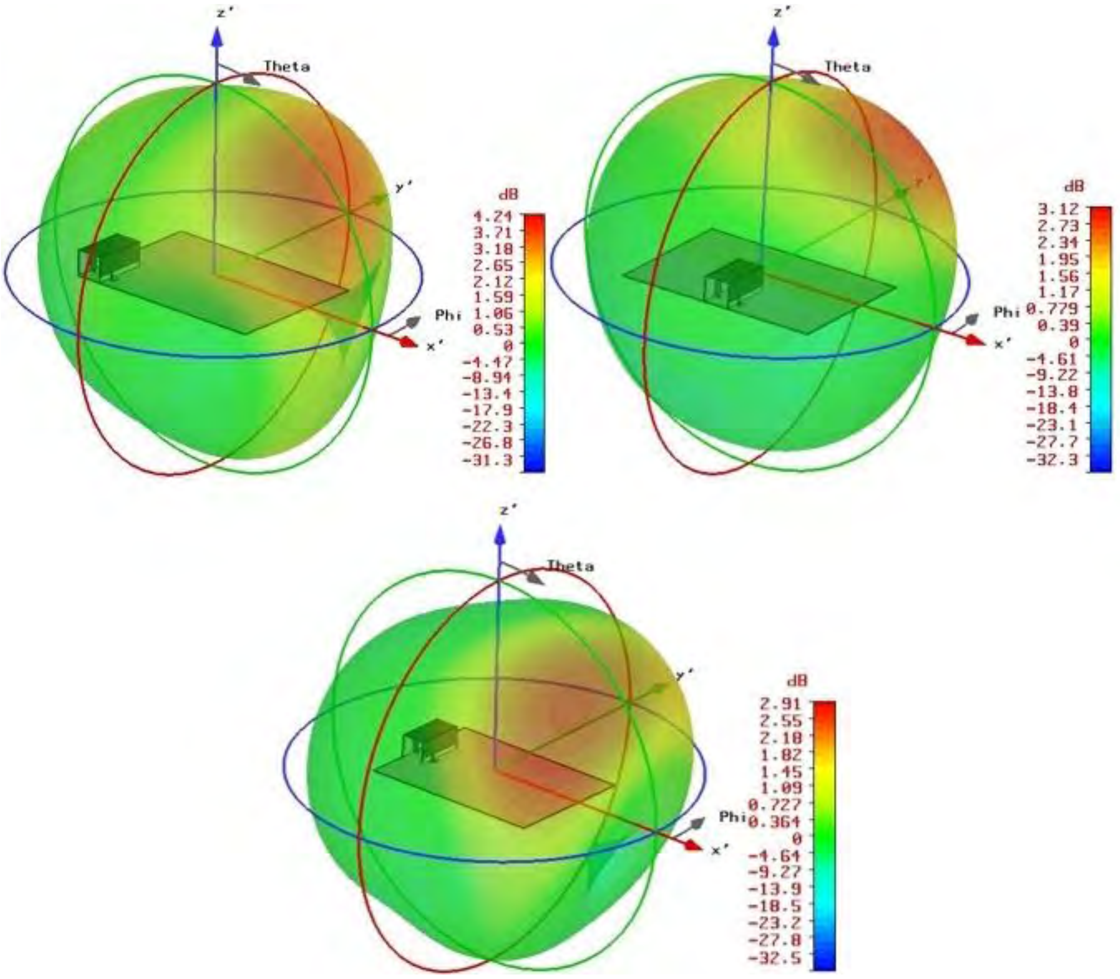
z'
Theta
y'
Phi
x'
dB
4.24
3.71
3.18
2.65
2.12
1.59
1.06
0.53
0
-4.47
-8.94
-13.4
-17.9
-22.3
-26.8
-31.3
z'
Theta
y'
Phi
x'
dB
3.12
2.73
2.34
1.95
1.56
1.17
0.779
0.39
0
-4.61
-9.22
-13.8
-18.4
-23.1
-27.7
-32.3
z'
Theta
y'
Phi
x'
dB
2.91
2.55
2.18
1.82
1.45
1.09
0.727
0.364
0
-4.64
-9.27
-13.9
-18.5
-23.2
-27.8
-32.5

Figure 4.12. Radiation patterns for different positions of the PIFA on the ground plane.

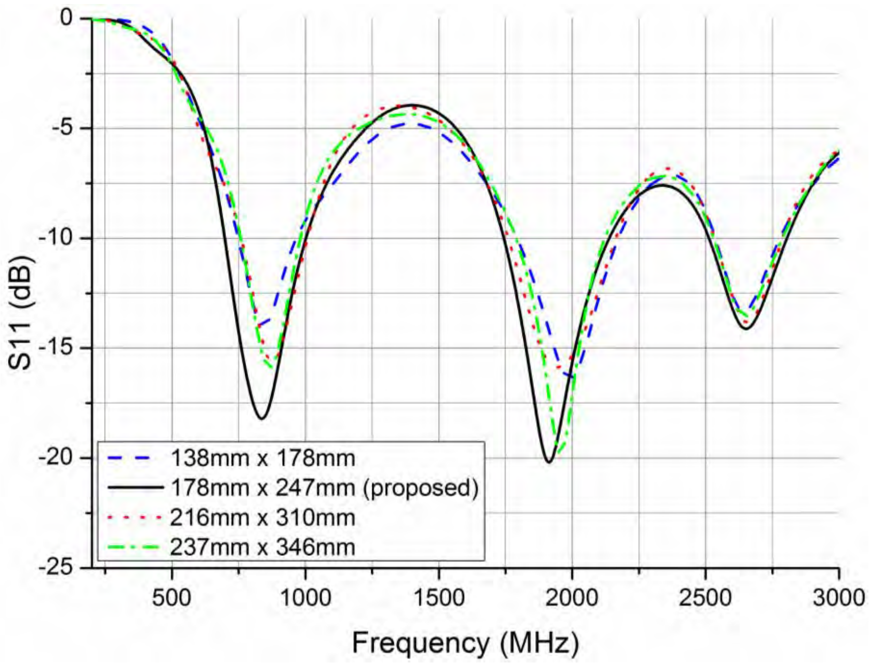

Figure 4.13. The simulated S_{11} dependence on the size of the ground plane.

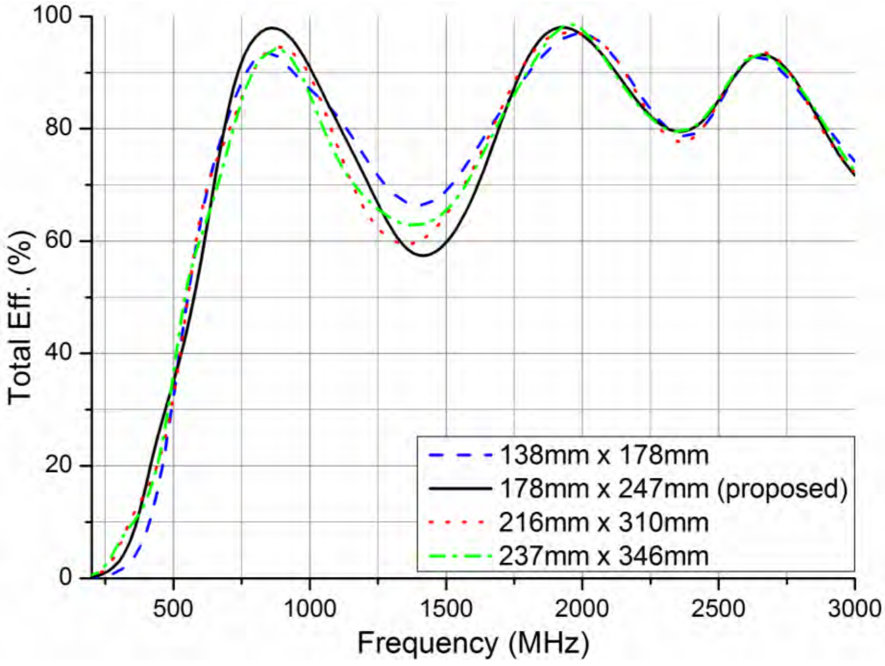

Figure 4.14. Simulated Total Efficiency for various ground plane sizes.

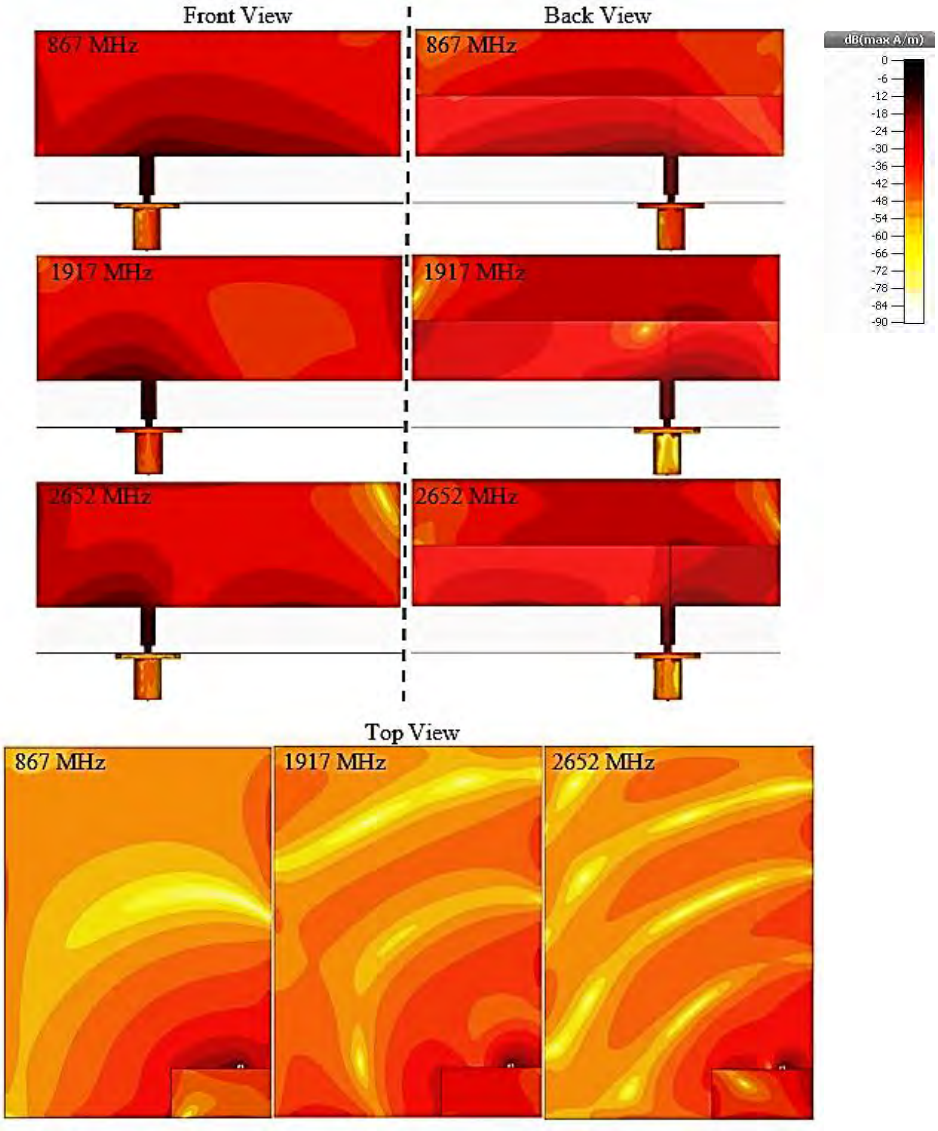
Front View
Back View
867 MHz
867 MHz
1917 MHz
1917 MHz
2652 MHz
2652 MHz
Top View
867 MHz
1917 MHz
2652 MHz
dB(max A/m)
0
-6
-12
-18
-24
-30
-36
-42
-48
-54
-60
-66
-72
-78
-84
-90

Figure 4.15. Current Distribution at 867, 1917 and 2652 MHz.

The surface current distribution at three frequencies is shown in Fig. 4.15. For the first resonant frequency it is evident that the main part of the antenna which resonates is the front folded part of PIFA and partially the back folded one. For the second and the third resonance the back folded part of PIFA contributes more to the radiation of the antenna.

From the top view it is particularly interesting be stated that the ground plane is an integral part of the antenna and contributes as a main part of the radiator at the three frequencies.

4.1.4 SIMULATED AND MEASURED RESULTS

The measured and simulated S_{11} for the PIFA antenna placed on the 178 × 247 mm^2 ground plane are in good agreement as seen in Fig. 4.14. The S_{11} was less than −10 dB over the range 688.5 − 1009.8 MHz, 1736.3 − 2121.5 MHz and 2459 − 2698 MHz giving fractional bandwidths of 40.6%, 20.87% and 9.31%, respectively. The frequency range for S_{11} less than −6 dB is 623.5 − 1111.8 MHz and 1621.74 − 3007.04 MHz, (62% and 75%, respectively).

The simulated total efficiency at 825 MHz and 2200 MHz is 85% and 97% respectively. The maximum value is 98% at 1915.4 MHz (Fig. 4.15).

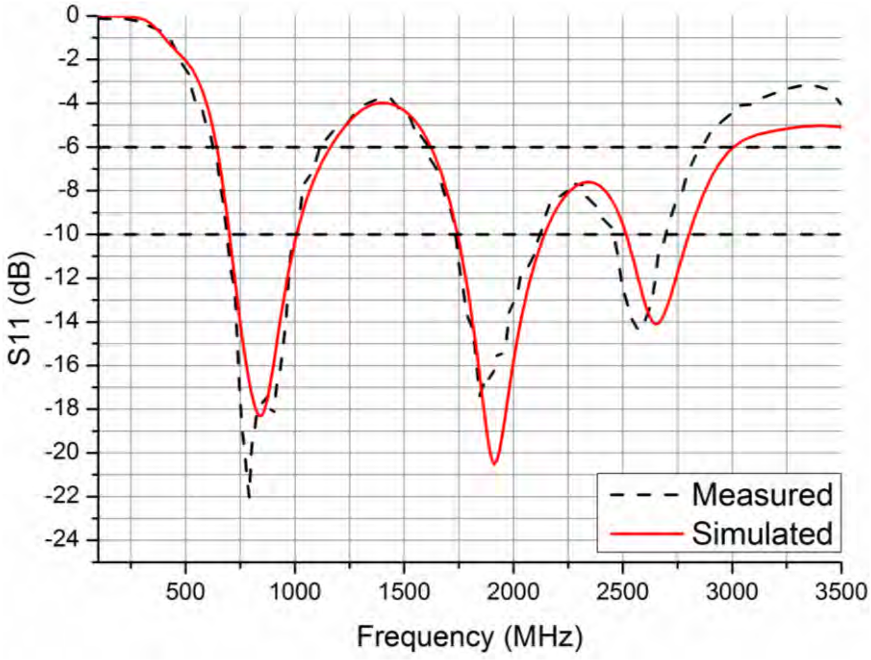

Figure 4.15. Simulated and measured S_{11} for the PIFA antenna.

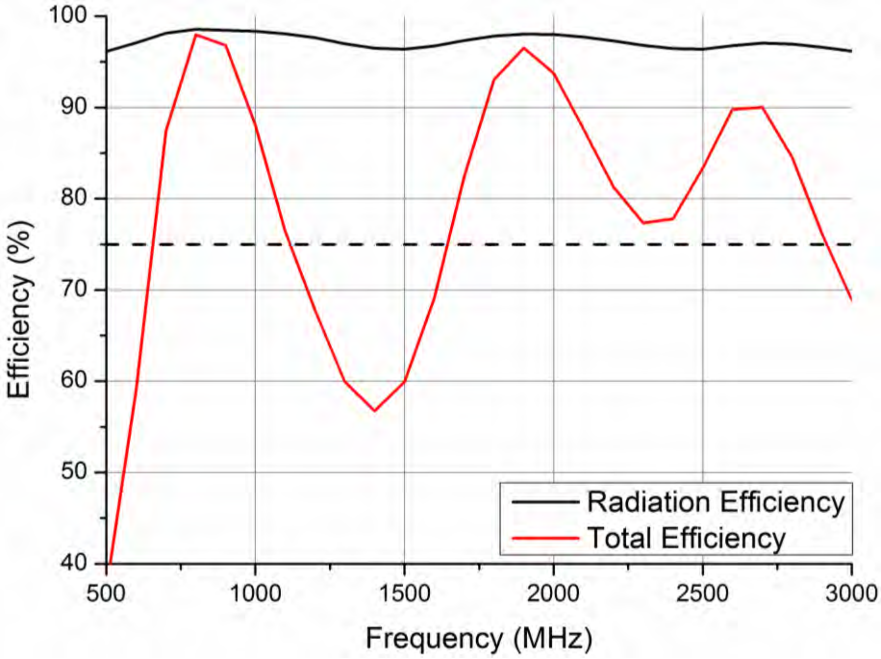

Figure 4.16. The simulated total and radiation efficiency.

In Figs. 4.16 and 4.17 the simulated and measured radiation patterns for (x–y) azimuth and (z–y), (z–x) elevation planes at 825 MHz and 2200 MHz are illustrated.

In the three planes at 825 MHz it is observed that the phi (φ) component provides good omnidirectional characteristics, which is also observed at the azimuth (x–y) plane of 2200 MHz for the theta (θ) component. Moreover, at 825 MHz in the elevation (z–y) plane good discrimination between the two components is provided.

The measured maximum gain for the lower and upper band is 4.5 dBi and 6.4 dBi, respectively. The measured gain, averaged over the principle plane cuts (x–y, z–y and z–x) was −1.1, −1.2 and −2.1 dBi for 825 MHz and −1.4, −1.7 and −1.6 dBi for 2200 MHz.

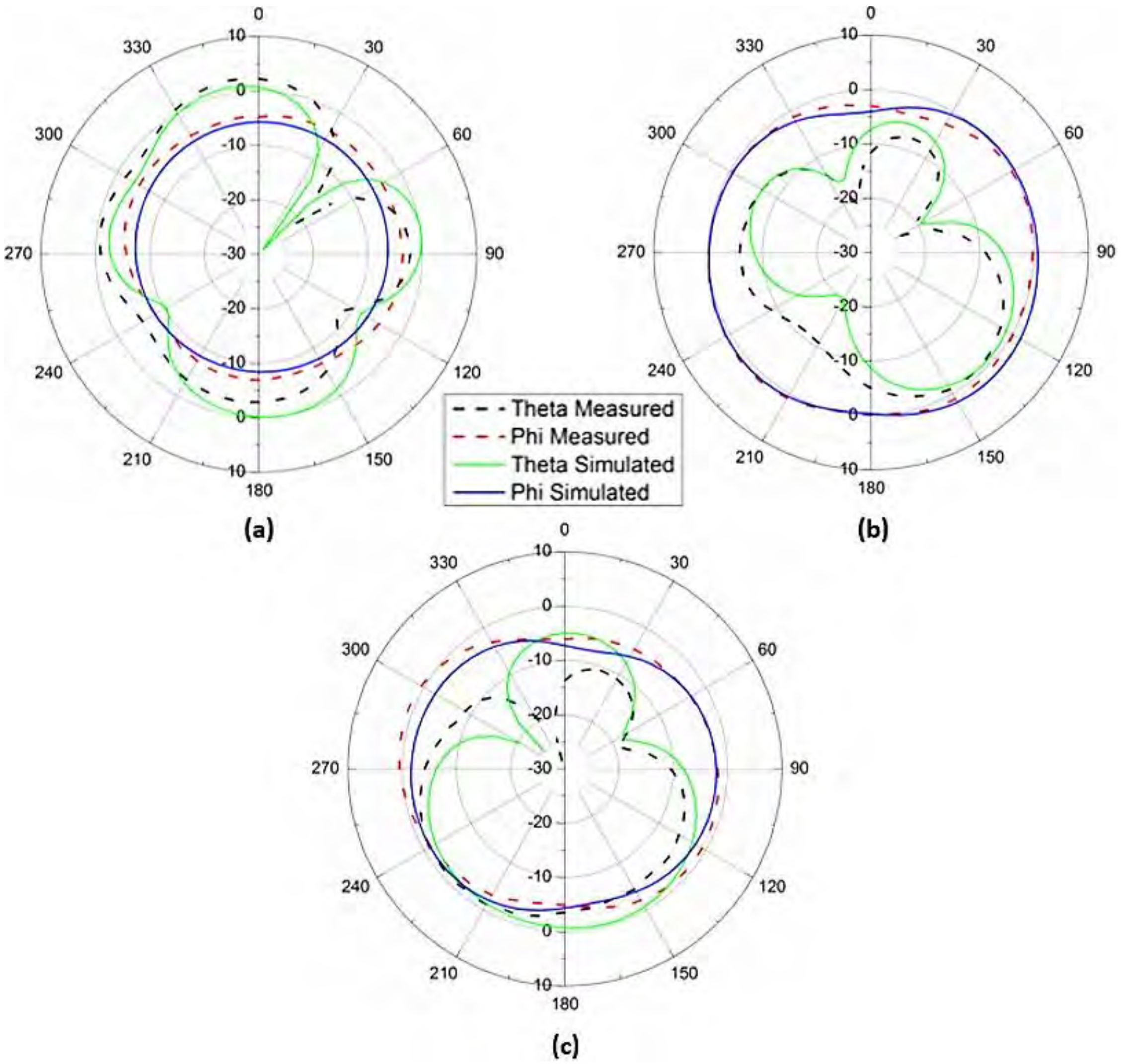

Figure 4.17. Measured and simulated (a) azimuth (x-y), (b) elevations (z-y) and (c) (z-x) plane patterns at 825 MHz.

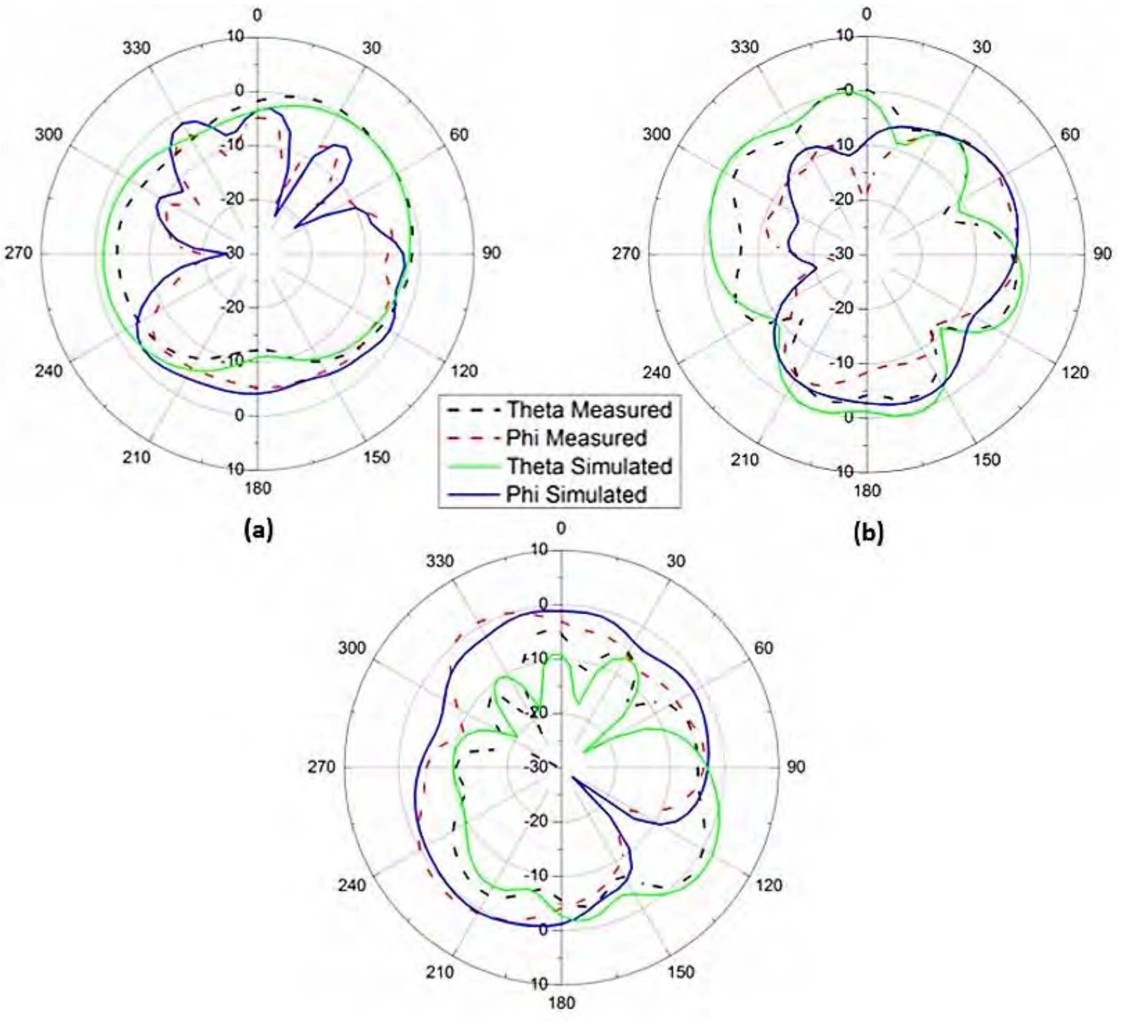

(a)

(b)

(c)

Figure 4.18. Measured and simulated (a) azimuth (x-y), (b) elevations (z-y) and (c) (z-x) plane patterns at 2200 MHz.

4.2 DUAL LTE PIFA FOR M2M APPLICATIONS

A Dual Planar Inverted-F-Antenna [32, 33, 34, 35, 36] placed on a 130 mm × 70 mm ground plane which is a typical M2M device PCB size. The antenna was designed with smaller size compared to the antenna of the first section and covers the lower LTE band (690 – 960 MHz) and partially the upper LTE band (1670 – 2067 MHz). The antenna is suitable for machine-to-machine (M2M) applications.

4.2.1 ANTENNA DESIGN

As a first step of the dual LTE PIFA antenna implementation is a single PIFA which resonates and provide one resonance. A long arm PIFA antenna resonates at lower frequency f_1 as well as a short arm PIFA resonates at higher frequency f_2 as it shown in Figure 4.19. A novel forked feed is introduced to combine the two PIFAs in the same geometry sharing the same feed, providing the dual band operation.

The antenna structure and coordinate system is depicted in Fig. 4.20. The radiating element is located on a FR-4 (Table 2.1) rectangular ground plane with dimensions 13 × 70 × 1.5 mm^3. The sides of the prototyped antenna are printed on thin FR-4 layers of thickness 0.2 mm with dimensions 70 mm ($\approx 0.15\lambda_0$ with $\lambda_0 = 455.9$ mm at 658 MHz) ×19 mm ($\approx 0.041\lambda_0$). The top part of the antenna consists of two thin copper layers (thickness = 0.15 mm). The top left (59.61 mm × 11 mm) and the top right (9.92 mm × 11 mm) part have a fold-over segments of length $a = 9.7$ mm and $b = 18.3$ mm respectively. The antenna comprises two PIFAs sharing the same feed. A small gap (0.2 mm) separating the two PIFAs (Fig. 4.21). The shorting strip of each PIFA has a width of 0.2 mm. The antenna is fed via a 50 Ω SMA connector through the ground plane to a 5.1 mm wide feeding strip, forked at the top with a small gap (0.2 mm) to separate the two feeds.

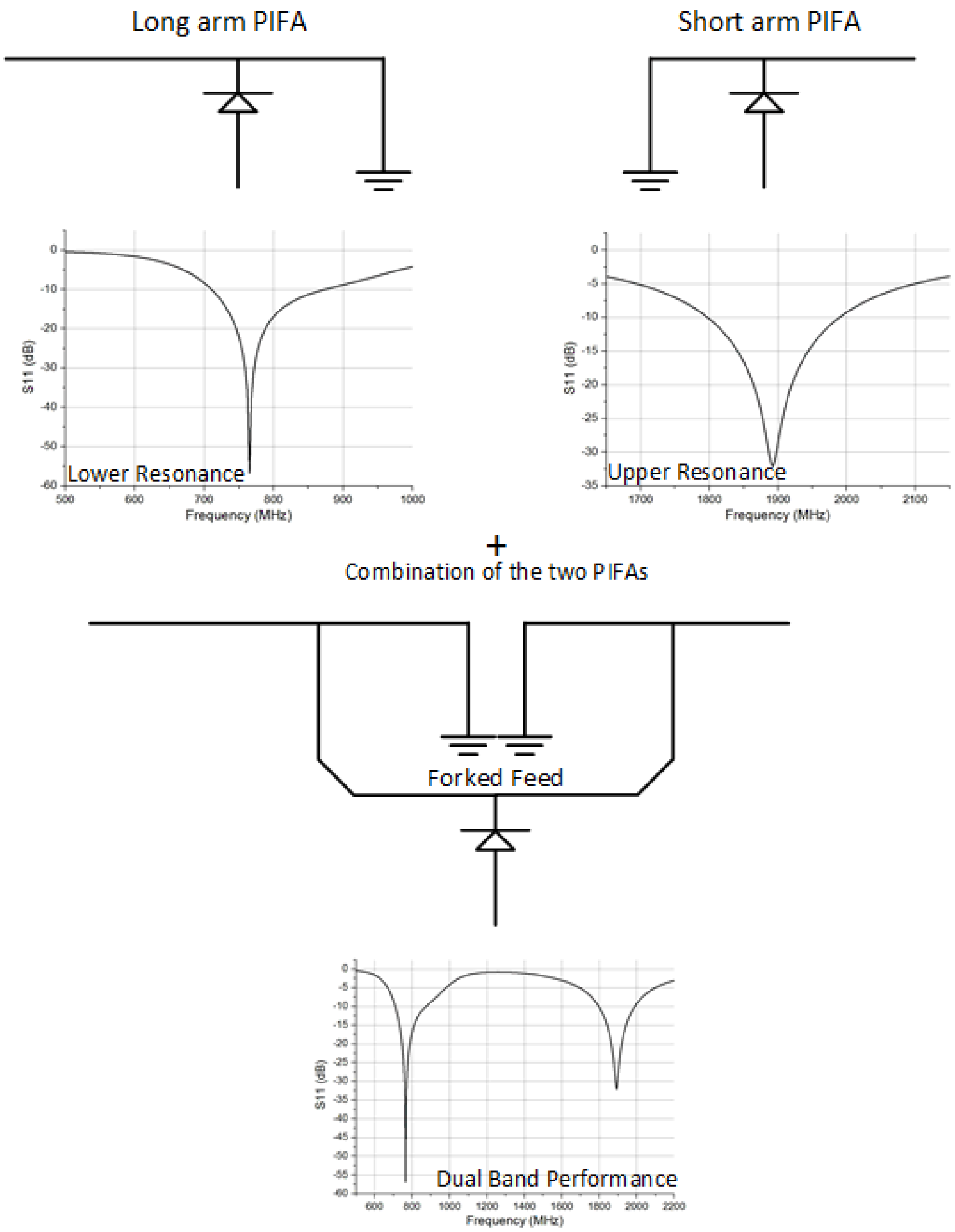
Long arm PIFA
Short arm PIFA
S11 (dB)
0
-10
-20
-30
-40
-50
-60
Lower Resonance
500
600
700
800
900
1000
Frequency (MHz)
S11 (dB)
0
-5
-10
-15
-20
-25
-30
-35
Upper Resonance
1700
1800
1900
2000
2100
Frequency (MHz)
+
Combination of the two PIFAs
Forked Feed
S11 (dB)
0
-5
-10
-15
-20
-25
-30
-35
-40
-45
-50
-55
-60
Dual Band Performance
600
800
1000
1200
1400
1600
1800
2000
2200
Frequency (MHz)

Figure 4.19. Schematic illustration of dual LTE PIFA design evolution.

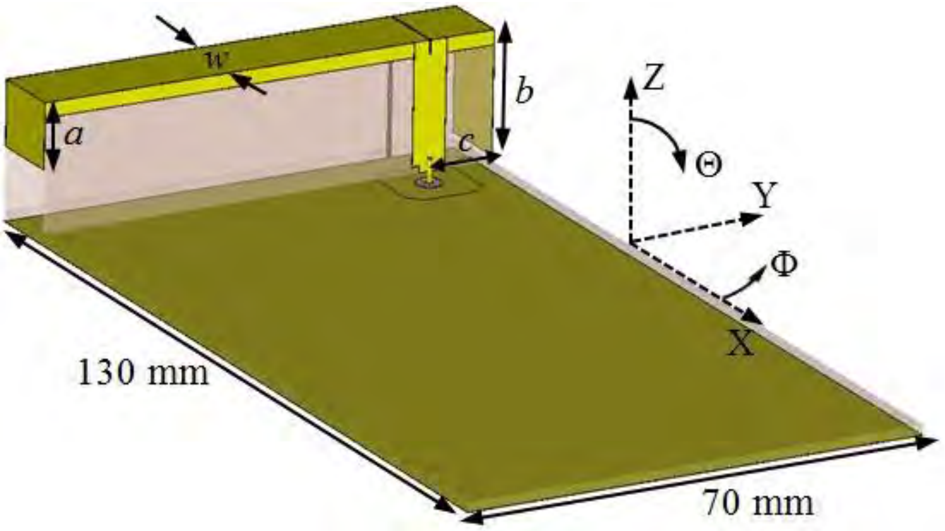
w
a
b
c
Z
Θ
Y
Φ
X
130 mm
70 mm

Figure 4.20. Antenna Model and the coordinate system.

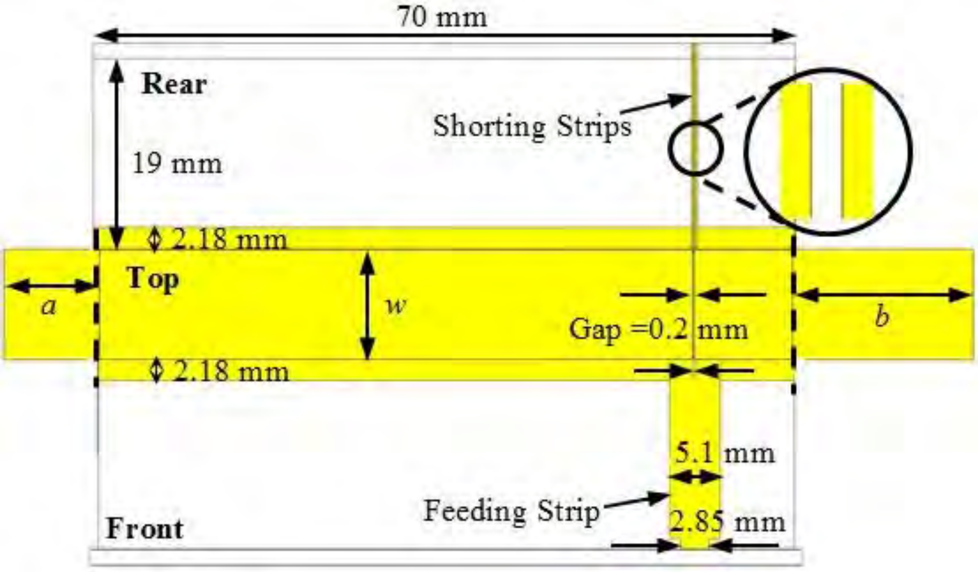
70 mm
Rear
19 mm
Shorting Strips
2.18 mm
Top
a
w
Gap =0.2 mm
b
2.18 mm
5.1 mm
Feeding Strip
2.85 mm
Front

Figure 4.21. Antenna geometry and parameters.

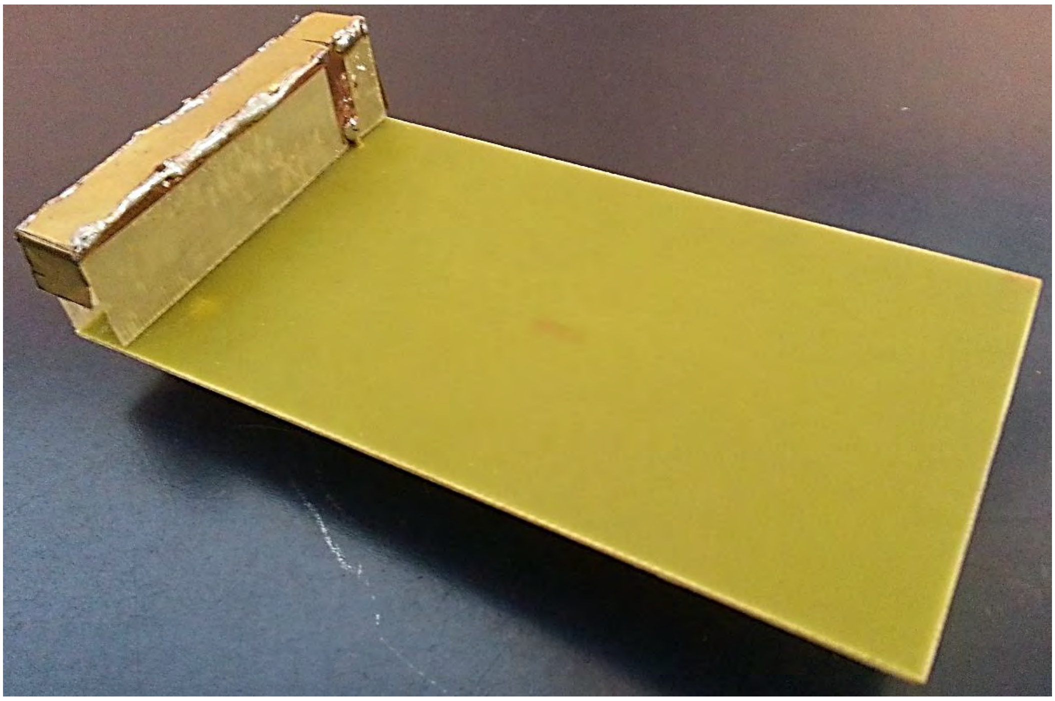

Figure 4.22. Dual PIFA LTE Antenna prototype.

4.2.2 PARAMETRIC STUDY

The dependence of the lower resonance on the height (*a*) of the left folded part was investigated for four different values. It is apparent that the matching and first resonant frequency is heavily dependent on this parameter as well as some effects on the second resonance. The optimum value of this parameter (*a*) is 9.7 mm.

The upper resonance is strongly dependent on the height (*b*) of the right folded section. The height (*b*) is varied from 16 mm to 19 mm. From the obtained result (Fig. 4.24) it is evident that the second resonant frequency can be controlled (impedance matching and frequency shifting) by varying the parameter *b*. The proposed value of the folded length (*b*) is 18.3 mm.

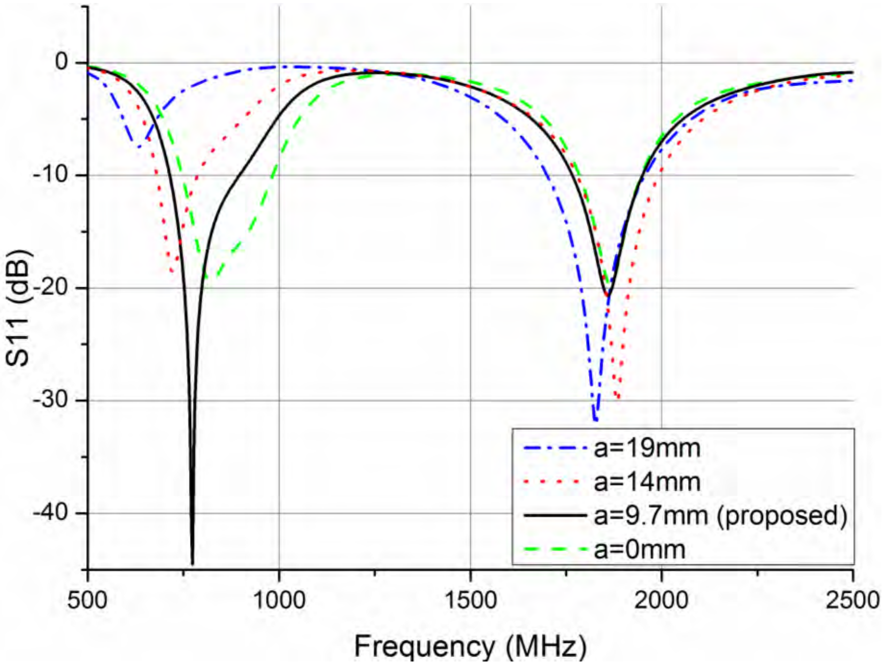

Figure 4.23. Simulated S_{11} variation for different values of (*a*).

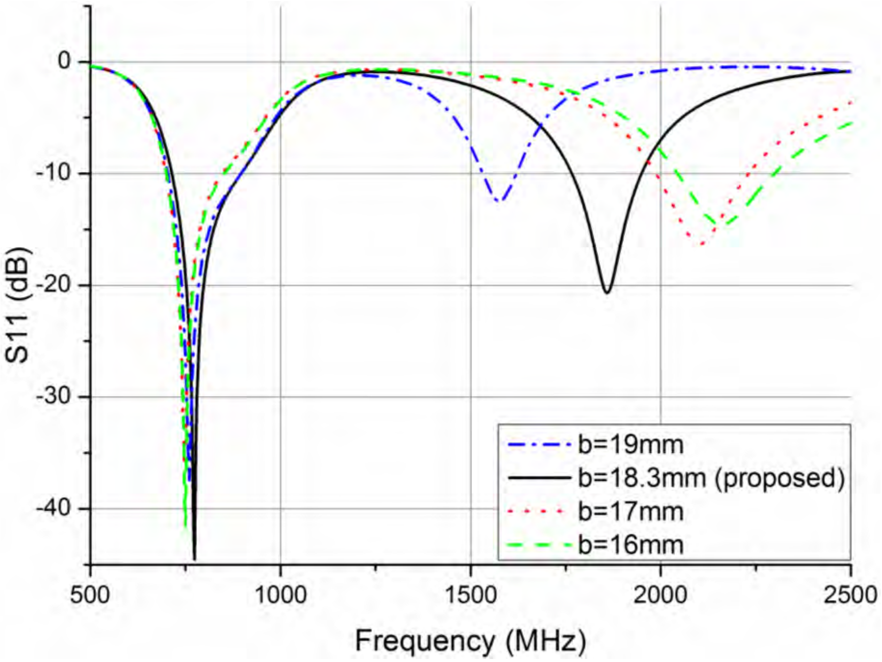

Figure 4.24. Simulated S_{11} variation for different values of (*b*).

The position (*c*) of the feed and the shorting point can determine the electrical length of both PIFAs. As the feed and shorting point moved from the right side (5.97 mm) of the antenna to the centre (35 mm) both resonances are shifted. The frequency ratio $F_r = f_u/f_l$ between the upper and the lower resonance continuously increases from 1.43 to 2.99 as the value of (*c*) changes from 35 mm to 5.97 mm. The proposed value of the parameter (*c*) is 9.97 mm.

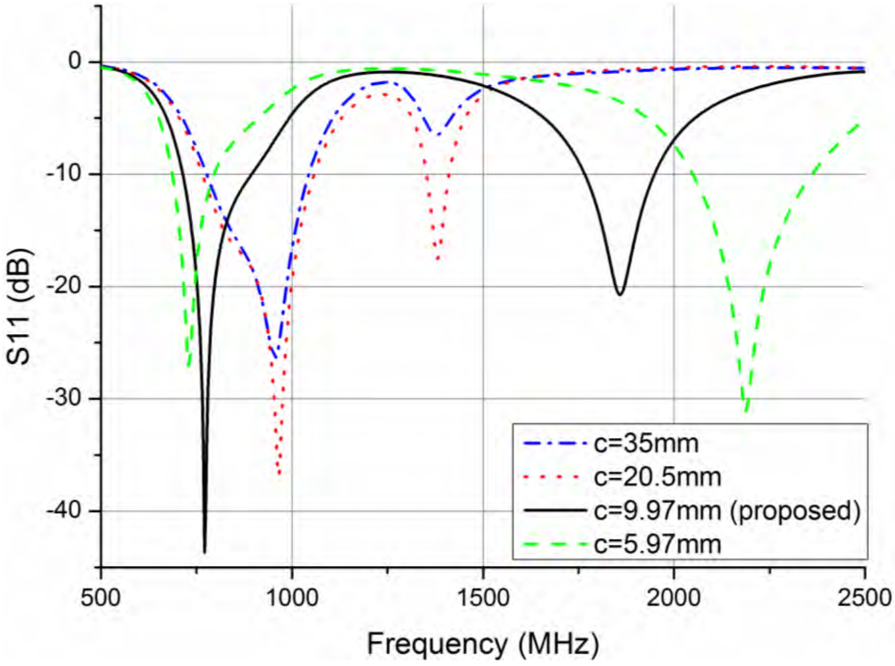

Figure 4.25. Simulated S_{11} variations for different values of (*c*).

Finally, the last key parameter which is studied for this antenna is the width (*w*) of the top part of the two PIFAs. From the obtained results it is clear that the second resonance is effectively controlled and shifted downwards as the parameter (*w*) increased from 5 mm to 14 mm. The width (*w*) has a strong impact and improves the impedance matching of the first resonant frequency. The proposed value of (*w*) is 11 mm.

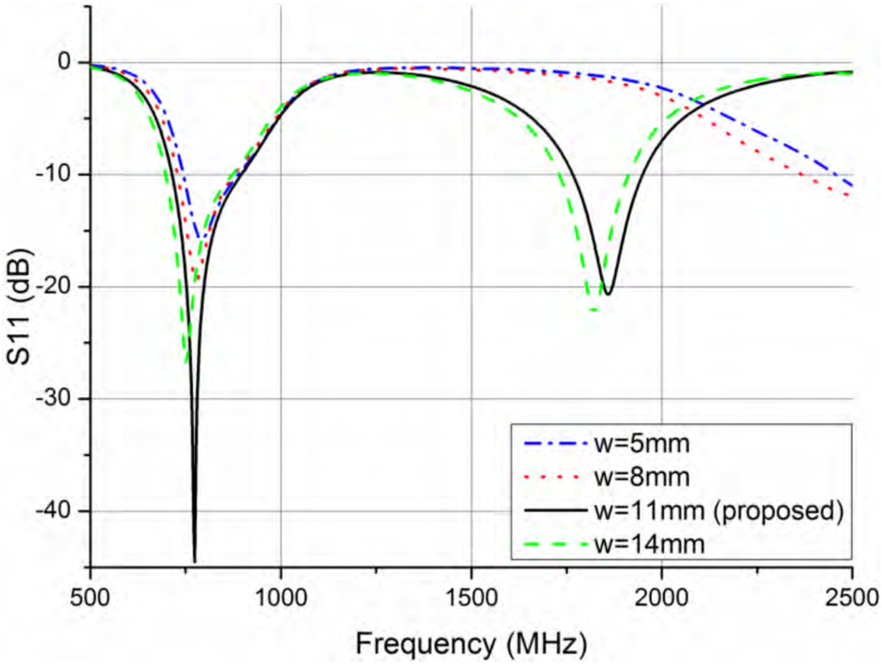

Figure 4.26. Simulated S_{11} variations for different values of (*w*).

The current distribution at both frequencies is shown in Fig. 4.27. For the first resonant frequency (Fig. 4.27(a)) the main part of the antenna which resonates is the longest arm PIFA. The electrical length of the longest PIFA is almost 90 mm which is almost a quarter wavelength at 760 MHz ($\lambda_0/4 = 98.6$ mm). For the second resonance (Fig. 4.27(b)) the main current distribution comes from the shorter arm PIFA of the antenna. The length of the shortest PIFA is almost 46 mm which is slightly over a quarter wavelength at 1860 MHz ($\lambda_0/4 \approx 40.3$ mm).

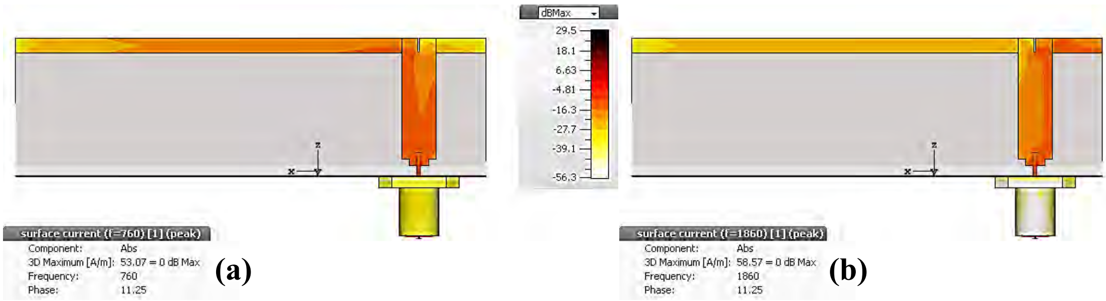

surface current (f=760) [1] (peak)
Component: Abs
3D Maximum [A/m]: 53.07 = 0 dB Max
Frequency: 760
Phase: 11.25

surface current (f=1860) [1] (peak)
Component: Abs
3D Maximum [A/m]: 58.57 = 0 dB Max
Frequency: 1860
Phase: 11.25

Figure 4.27. The current distribution at (a) 760 MHz and (b) 1860 MHz.

4.2.3 SIMULATED AND MEASURED RESULTS

In Fig. 4.28 the measured and simulated S_{11} show good agreement. The antenna operates at 760 MHz with a −10 dB and −6 dB impedance bandwidth of 208 MHz (682 – 890 MHz) and 306 MHz (658 – 964 MHz) respectively for the lower band and at 1860 MHz with a −10 dB and −6 dB impedance bandwidth of 220 MHz (1750 – 1970 MHz) and 396 MHz (1671 – 2067 MHz) respectively for the upper band. The simulated results provide a −10 dB and −6 dB impedance bandwidth of 162 MHz (711 – 873 MHz) and 281 MHz (680 – 961 MHz) at the centre frequency of 765 MHz for the first band and a −10 dB and −6 dB impedance bandwidth of 192 MHz (1797 – 1989 MHz) and 341 MHz (1725 – 2066 MHz) at the centre frequency of 1893 MHz for the second band.

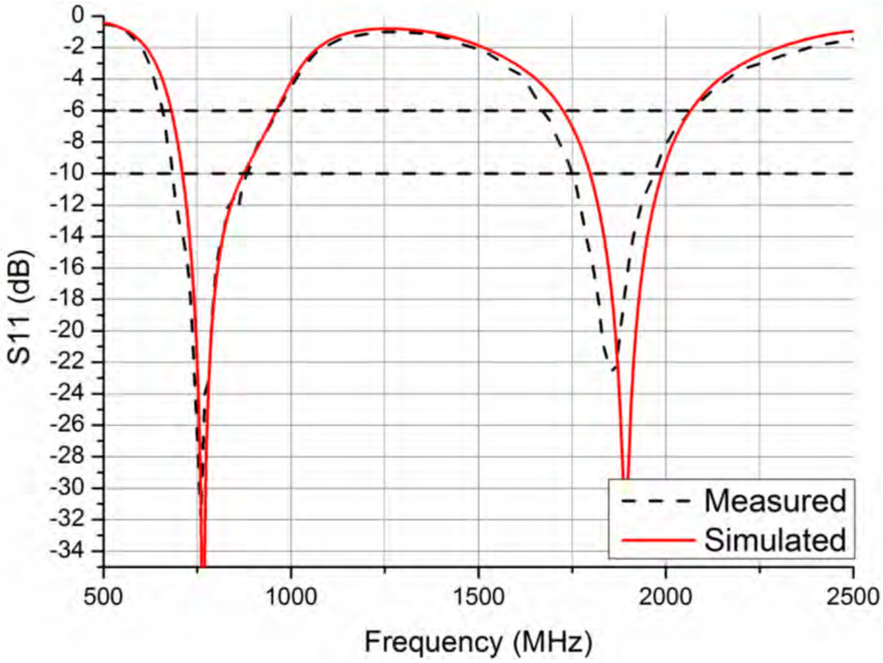

Figure 4.28. Simulated and measured S_{11} for the PIFA antenna.

In Fig. 4.30 the measured and the simulated total efficiency and realized gain are given. There is a reasonable agreement between the simulated and measured results. The total efficiency and the realized gain of the antenna were measured by Microwave Vision Group (MVG) in StarLab 0.6 – 18 GHz and are listed in Table 4.1. This chamber uses the nearfield measurement technique to calculate the equivalent far-field data of the AUT. The antenna under test is positioned in the centre of a circular ‘arch’ which contains 15 separate measurement probes spaced equally apart as can be seen in Fig. 4.29. The antenna is rotated horizontally through 360° and the combination of this rotation and the array of probes allows a full 3D scan of the antenna to be carried out, allowing full 3D radiation patterns to be measured, plotted and analysed. Efficiency, gain and directivity information can be transformed and calculated from the far-field radiation pattern data [74].

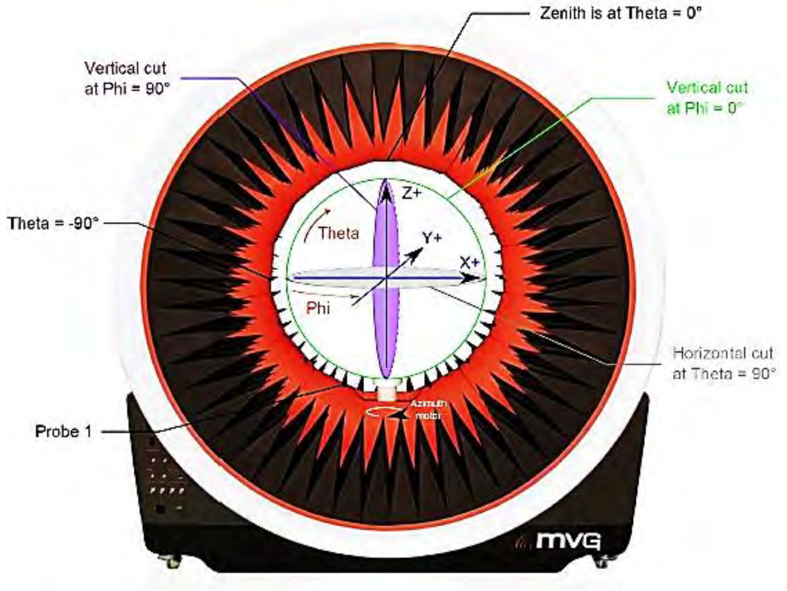
Zenith is at Theta = 0°
Vertical cut at Phi = 90°
Vertical cut at Phi = 0°
Theta = -90°
Z+
Theta
Y+
X+
Phi
Horizontal cut at Theta = 90°
Probe 1
Azimuth motor
mvg

Figure 4.29. StarLab spherical coordinate system [74].

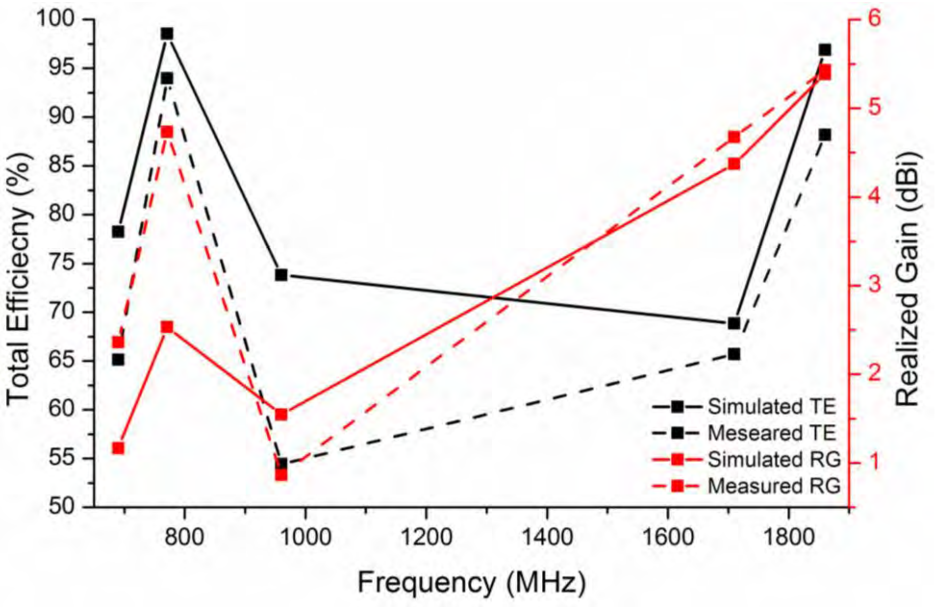

Figure 4.30. Measured and simulated total efficiency and realized gain.

Table 4.1. Measured and Simulated Total Efficiency and Realized Gain.

Frequency (MHz)	Total Efficiency (%)		Realized Gain (dBi)	
	Simulated	Measured	Simulated	Measured
690	78.2	64.8	1.2	1.9
760	98.5	83.6	2.5	3.6
960	73.8	58	1.5	0.8
1710	68.8	65.2	4.4	3.5
1860	96.9	79.8	5.4	4.1

The radiation patterns of the three plane cuts (x–y, z–x and z–y) for 760 MHz and 1860 MHz are shown in Figs. 4.32 and 4.33 respectively. All the measured patterns are illustrated against the simulated patterns in 10 dB/division scaled plots.

It is observed that the theta (θ) component provides good omnidirectional characteristics in the (x–y) plane for both frequencies, which is also observed in the (z–y) plane for 760 MHz for the phi (φ) component.

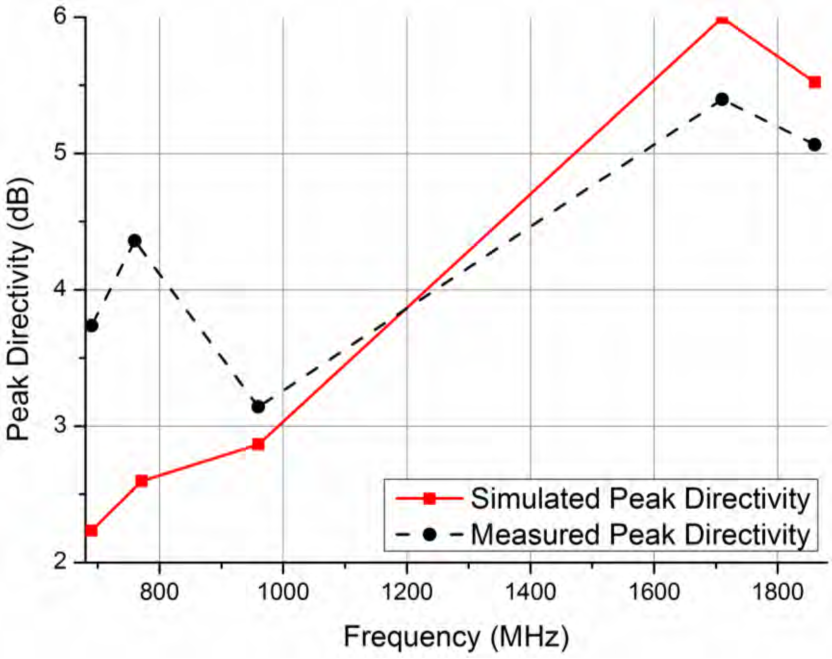

Figure 4.31. Measured and Simulated Peak Directivity.

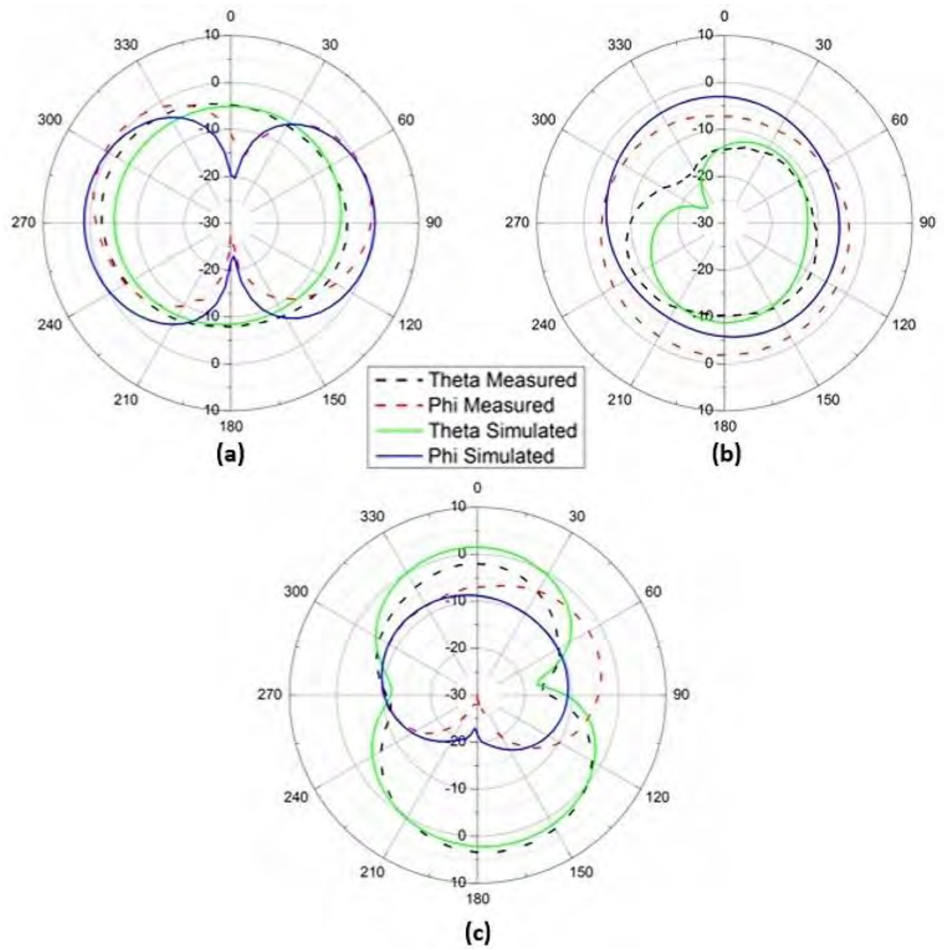

Figure 4.32. Measured and simulated (a) azimuth (x-y), (b) elevations (z-y) and (c) (z-x) plane patterns at 760 MHz.

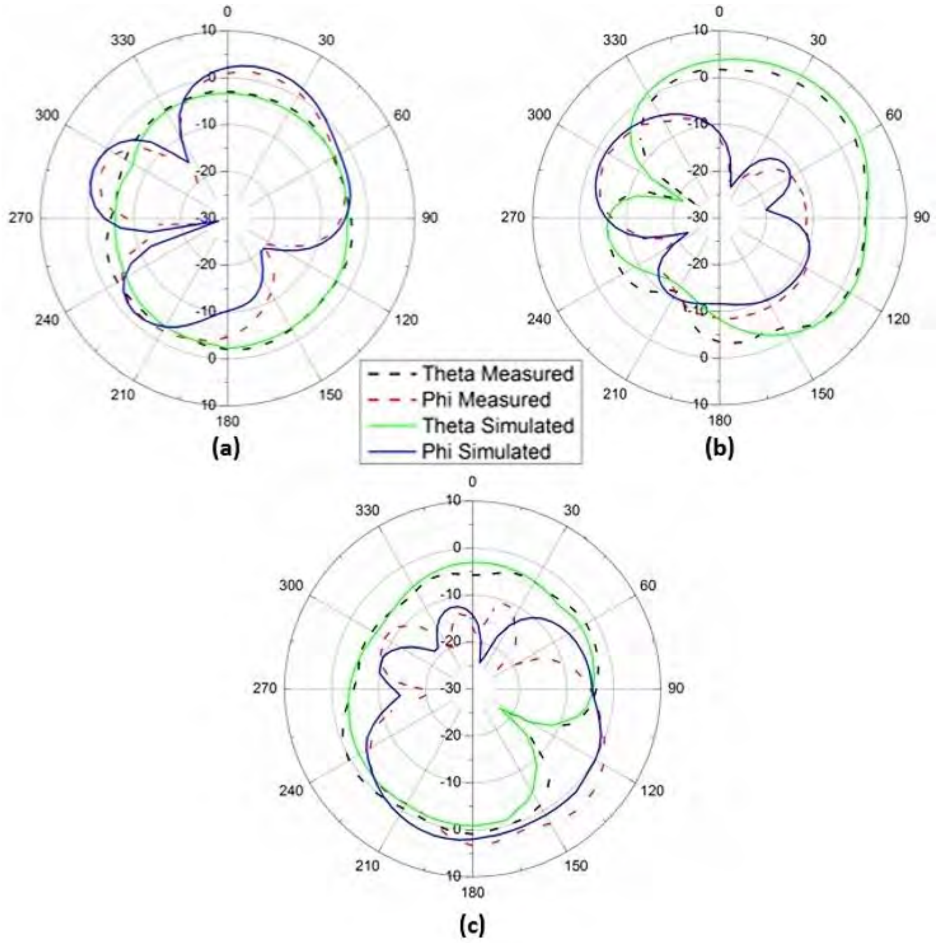

Figure 4.33. Measured and simulated (a) azimuth (x-y), (b) elevations (z-y) and (c) (z-x) plane patterns at 1860 MHz.

4.3 CONCLUSIONS

This chapter presented the study of the performance of two Planar Inverted-F antennas on rectangular ground planes in the LTE frequency range respectively. A parametric investigation of key geometrical parameters of the two antennas are reported.

The first PIFA antenna is studied in terms of the ground plane size and positioning. The obtained results show that the position of the PIFA on the ground plane effectively controls both resonant frequencies (impedance matching and frequency shifting). The S_{11} shows stronger dependence on the ground plane size than the radiation properties and the total efficiency does not change significantly.

The second PIFA antenna which is smaller and mounted on a typical device size ground plane covers the lower LTE band and partially the upper. The proposed antenna has been optimised with good omnidirectional radiation characteristic at both bands, high efficiency and gain, is low profile, with additional benefit of a large frequency-ratio range $f_u/f_l = 2.44$ which is easy to control.

5. Meander Line Monopole Antenna

Meander line monopole antennas [21 22, 23, 24] are mainly used in compact and portable smart devices. With this state of the art technology, antenna designers are able to integrate compact meander line structures in small envelopes achieving broadband performance, providing also high gain and efficiency for its compact size.

This chapter proposes a geometry based on a folded meander line structure which is suitable for automated meter reading (AMR) systems [75]. The Electronic Communication Committee (ECC) allocates narrowband channels in the 410 – 470 MHz range [76]. CDMA and more recently eLTE operate in this band with channel bandwidths in the region of ~1 − 5 MHz [77, 78]. Performance metrics include impedance bandwidth, radiation patterns and total efficiency.

In this section a tuneable folded meander-line monopole antenna covering the lower UHF band (412 – 475 MHz) is studied. The antenna configuration is optimized for impedance matching. The effect of the ground plane size on the general radiation properties is analysed. Finally, the installed performance of the proposed antenna is investigated with the antenna embedded in a plastic housing and placed on a concrete wall.

5.1 MEANDERED ANTENNA FOR M2M APPLICATIONS

The antenna is designed for integration into a compact smart meter device for cellular communications technologies [79] which provides a tuneable functionality over the frequency range of 412 – 475 MHz. According to specifications for 450 MHz meter reading system applications [80], there is a requirement for an omnidirectional antenna and a S_{11} of better than −6 dB over the proposed frequency range with narrow channel bandwidths and maximum EIRP of +27 dBm. Designing a small-size, high-efficiency, broadband antenna, is a considerable challenge for the antenna designer.

5.2 ANTENNA CONFIGURATION

The initial implementation idea which leads to the final meander line design is a folded monopole antenna as it shown in Fig. 5.1. The position of the feeding and the shorting strips is changed to the back side of the FR4 substrate allows more available space to confine the electrical length (meander line) into a compact physical size and to move to lower frequencies (Fig. 5.2).

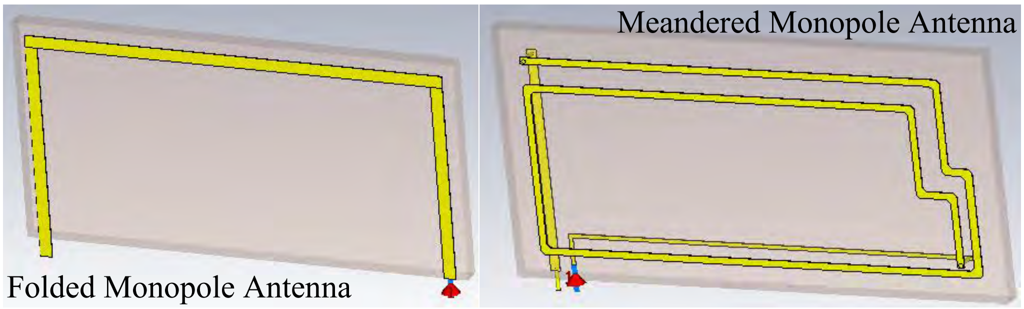

Figure 5.1. Illustration of the antenna designed steps.

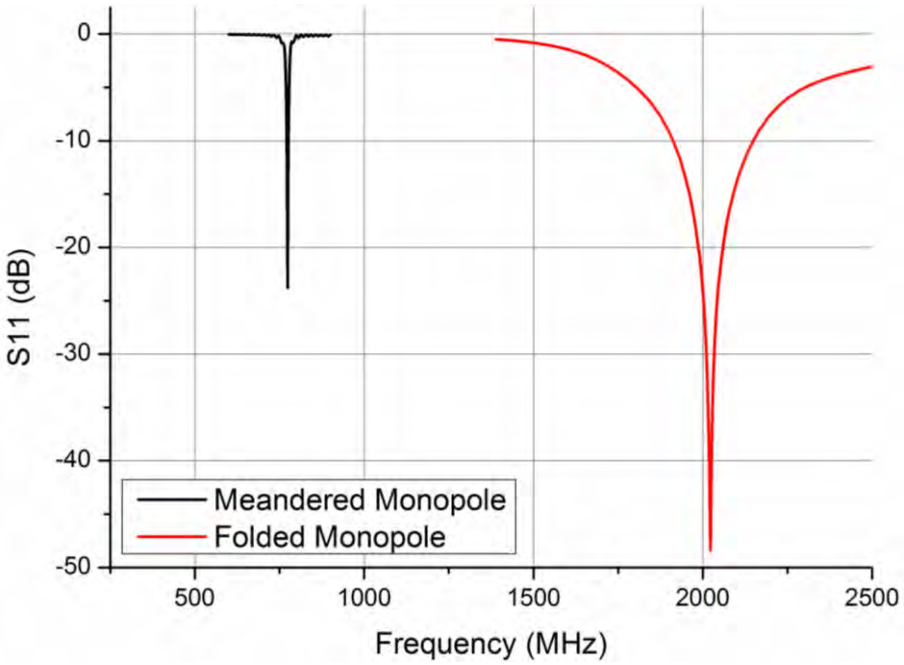

Figure 5.2. Simulated S_{11} for the two different antenna models.

The antenna geometry, shown in Fig. 5.3, is printed on double-sided FR-4 substrate (Table 2.1). The antenna of height h = 15 mm ≈ 0.02λ_0 (λ_0 = 724.5 mm at 414 MHz) and length l = 42.3 mm ≈ 0.058λ_0 is located on the corner of a 130 mm × 70 mm ground plane which represents a typical device PCB. The feeding and shorting strips (width = 0.5 mm and 0.7 mm respectively) are located on the rear side of the substrate and connected to the meander line (width = 0.5 mm) on the front side using a via and a slider connector.

The separation between the horizontal strips a_1, a_2 and a_3 is 0.85 mm. The gap between the vertical strips b_1, b_2 and b_3 is 0.83 mm. The lower horizontal strips c_1 and c_2 have a separation of 0.5 mm and between c_2 and c_3 the separation is 0.68 mm. The gap between the two vertical strips d_1 and d_2 is 0.8 mm. The separation between the strips is optimised to obtain the best impedance matching for the desired frequency.

The moveable U-shaped aluminium via-slider is mounted through a slot in the substrate, connecting the two sides of the PCB. The start and end positions of the via-slider are at 0 mm ≤ P ≤ 33.5 mm (Fig. 5.3). The monopole antenna is fed using a 50 Ω microstrip line (width = 1.3 mm) with a SMA connector below the ground plane.

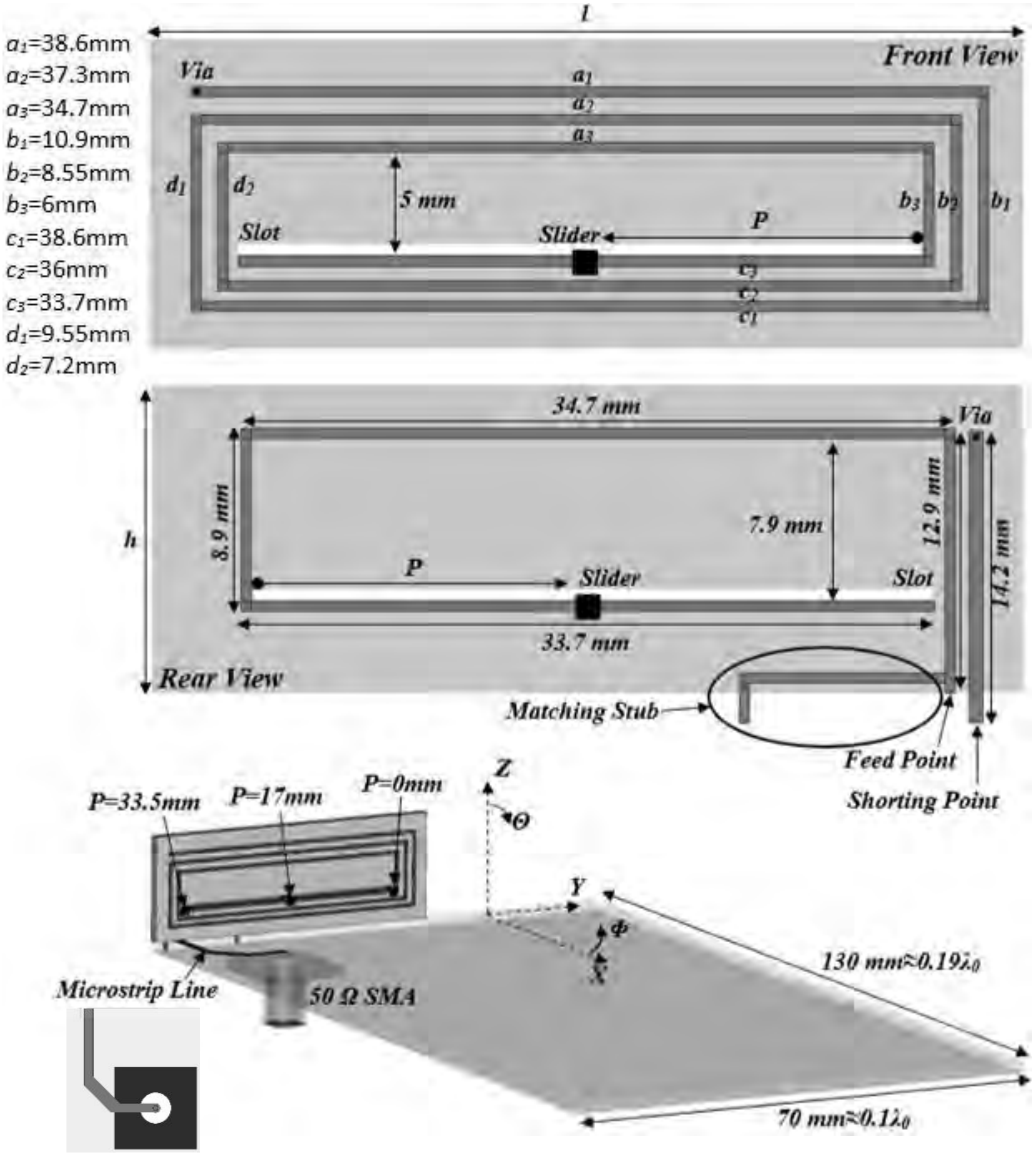
a_1=38.6mm
a_2=37.3mm
a_3=34.7mm
b_1=10.9mm
b_2=8.55mm
b_3=6mm
c_1=38.6mm
c_2=36mm
c_3=33.7mm
d_1=9.55mm
d_2=7.2mm
l
Front View
Via
a_1
a_2
a_3
d_1
d_2
5 mm
Slot
Slider
P
b_3
b_2
b_1
c_3
c_2
c_1
34.7 mm
Via
h
8.9 mm
7.9 mm
12.9 mm
14.2 mm
P
Slider
Slot
33.7 mm
Rear View
Matching Stub
Feed Point
Shorting Point
P=33.5mm
P=17mm
P=0mm
Z
Θ
Y
Φ
X
130 mm≈0.19λ_0
70 mm≈0.1λ_0
Microstrip Line
50 Ω SMA

Figure 5.3. Antenna geometry and the coordinate system.

5.3 MATCHING STUB

To improve the impedance matching of the antenna, a matching stub is added and shorted to the ground plane as shown in Fig 5.4 (a) and also shifts downwards the resonant frequency by introducing additional capacitance and reducing the real impedance towards 50 Ω (Fig. 5.4 (b)). The matching stub consists of two strips (width = 0.5 mm) with horizontal and vertical length 10 mm and 2.5 mm, respectively.

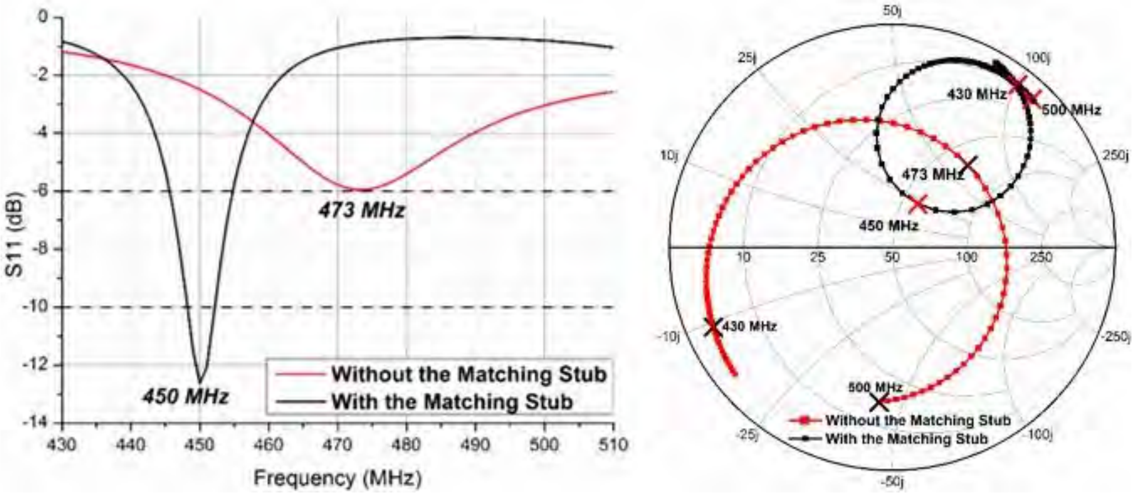

Figure 5.4. (a) Improved impedance matching and control of frequency, (b) Smith chart.

5.4 GROUND PLANE SIZE

The ground plane size was found not to have significant influence on the antenna performance. A parametric study of the ground plane size was made and simulations were carried out for ground plane sizes of 40 × 100 mm^2 to 170 × 230 mm^2. Fig. 5.5 shows a plot of the S_{11} for different ground plane sizes. From the graph it is clear that the ground plane size does not significantly affect the resonant frequency and the impedance matching of the antenna. For the given ground plane sizes the resonant frequency does not change more than 4 MHz.

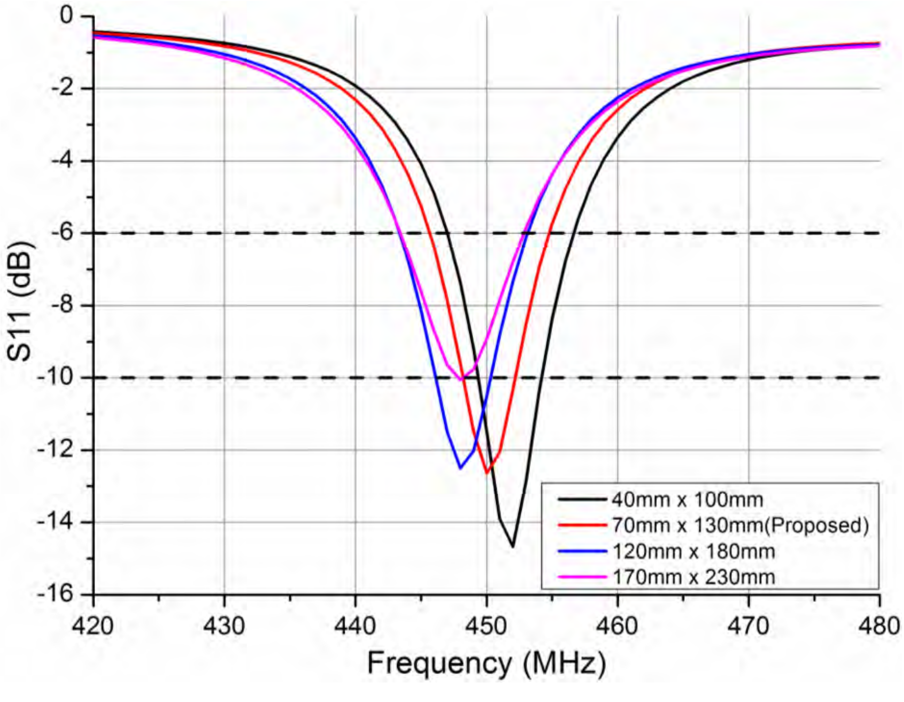

Figure 5.5. Simulated S_{11} dependence on the ground plane size.

5.5 MEASURED AND SIMULATED RESULTS

In Fig. 5.6 the measured and simulated S_{11} for 5 different tuning positions of the via-slider connector *P*, which controls the resonant frequency are shown. The measured −10 dB and −6 dB tuneable impedance bandwidth, is 412.4 – 474.8 MHz (BW = 62.4 MHz) and 71.6 MHz (409.5 – 481.1 MHz) respectively. The measured −6 dB and −10 dB bandwidths at 414 MHz (λ_0 ~ 724.5 mm) are ~8 MHz and ~3 MHz respectively.

The position of the via-slider for resonance at 450 MHz is *P* = 17 mm with a −10 dB impedance bandwidth of 7.2 MHz (446.5 – 453.7 MHz) and a −6 dB impedance bandwidth of 12.6 MHz (444.1 – 456.7 MHz).

The tuneability of the proposed antenna could be also achieved with diodes or varactors. However, the installation of diodes and varactors in a limited available space with the requirement of bias and DC suppliers is a constraint for implementation.

The meander-line monopole is the dominant part of the antenna which contributes and resonates at 450 MHz. The current distribution is shown in Fig. 5.7. The two open circuited lines after the slider connector are an integral part of the antenna reconfigurability and have been included in all simulations and measurements including total efficiency. If the open circuited strips are cut after tuning the antenna, this would cause a slight upward shift in frequency.

The total efficiency was simulated and measured and is illustrated in Fig. 5.8 for three different frequencies and indicates 21% total measured efficiency at 420 MHz, 24% at

450 MHz and 21.7% at 470 MHz. It is observed that there is a reasonable agreement between the simulated and measured results.

In an attempt to calculate the maximum theoretical total efficiency of the proposed antenna and to define how close is the measured values to the fundamental limitations, a mathematical explanation is described.

Harrington [7] defined the natural gain limit for a practical antenna: $G_{max}=(ka)^2+(ka)$. In [81] the maximum directivity of an antenna appointed with the following reasonable formula: $D_{max}=(ka)^2+3$. The relation between the antenna gain and directivity is expressed through the following formula:

$$G_{max} = \eta \times D_{max} = G_{max}/D_{max} = (ka)^2 + ka/(ka)^2 + 3 \qquad (5.1)$$

Based on the previous expression, the maximum total efficiency is equal to:

$$\eta_T = (1-|\Gamma|^2) \times \eta = (1-|\Gamma|^2) \times [(ka)^2 + ka/(ka)^2 + 3], \qquad (5.2)$$

where k is the wave number in free space, $k = 2\pi/\lambda$ and a is the radius of the minimum sphere that encloses the antenna. The theoretical and measured total efficiency values of the antenna are listed in Table 5.1. The radius a of the proposed antenna's radiansphere was calculated at 74.2 mm which is $0.11\lambda_0$ of the respective wavelength $\lambda_0 \approx 667$ mm at 450 MHz. The wavenumber k at 450 MHz is 0.0094 and $ka = 0.0094 \times 74.2 \approx 0.69$. The theoretical maximum total efficiency is calculated through the formula (5.2):

$$\eta_T=(1-|\Gamma|^2)\times [(ka)^2+ka/(ka)^2+3]=(1-|\Gamma|^2)\times 0.335=0.98\times 33.5=0.328\approx 32.8\%.$$

As can been seen from the acquired results in Table 5.1 the measured total efficiency is very close to the theoretical maximum values, which shows that the proposed antenna is quite efficient compared to its size and to the operating frequency range.

Table 5.1. Comparison of theoretical and measured total efficiency.

Total Efficiency (%)	420 MHz	450 MHz	470 MHz
Theoretical	28	32.8	30
Measured	21	24	21.7

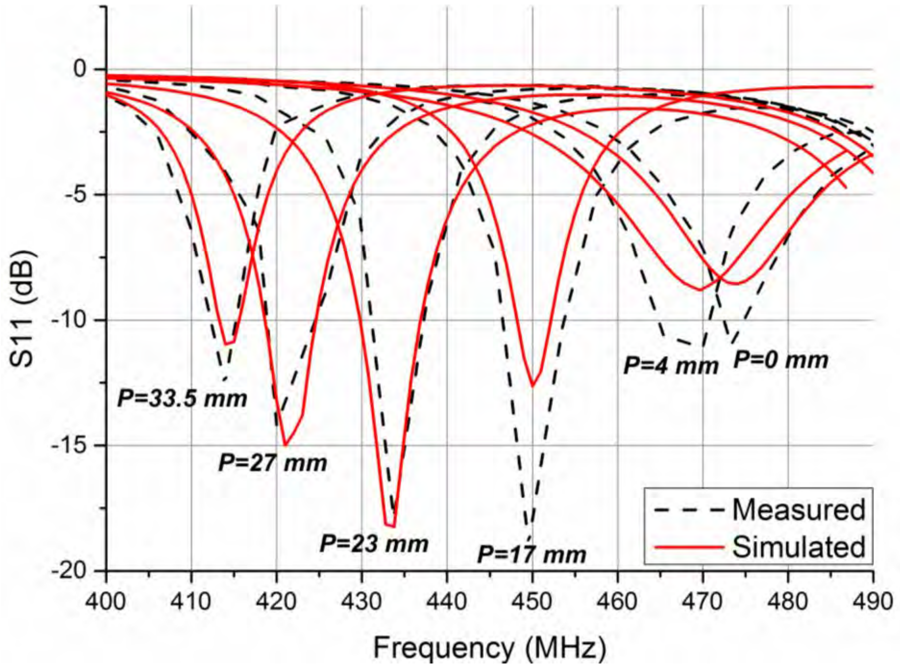

Figure 5.6. Simulated and measured S_{11}.

In Figs. 5.10 to 5.12 the measured and simulated azimuth (x–z) and elevation (x–y) and (y–z) plane radiation patterns at 450 MHz are illustrated. In Fig. 5.11 it is seen that the omnidirectional pattern provides good polarization discrimination in the elevation (x–y) plane. The maximum measured gain is −6.1 dBi. There is good agreement between the measured and the simulated results. The radiation patterns are stable across the tuneable band.

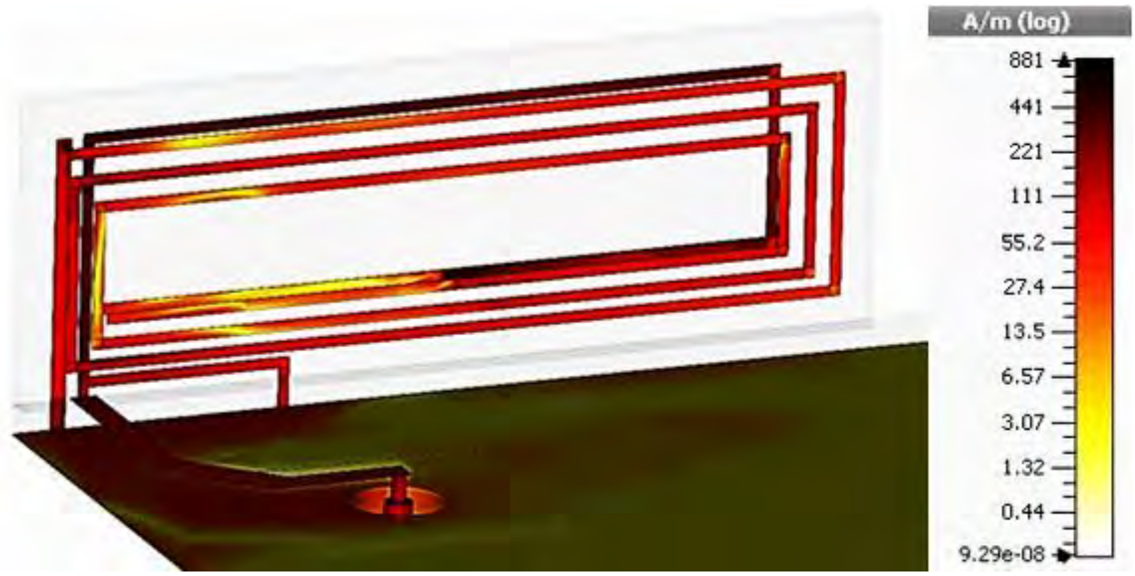
A/m (log)
881
441
221
111
55.2
27.4
13.5
6.57
3.07
1.32
0.44
9.29e-08

Figure 5.7. Current distribution at 450 MHz.

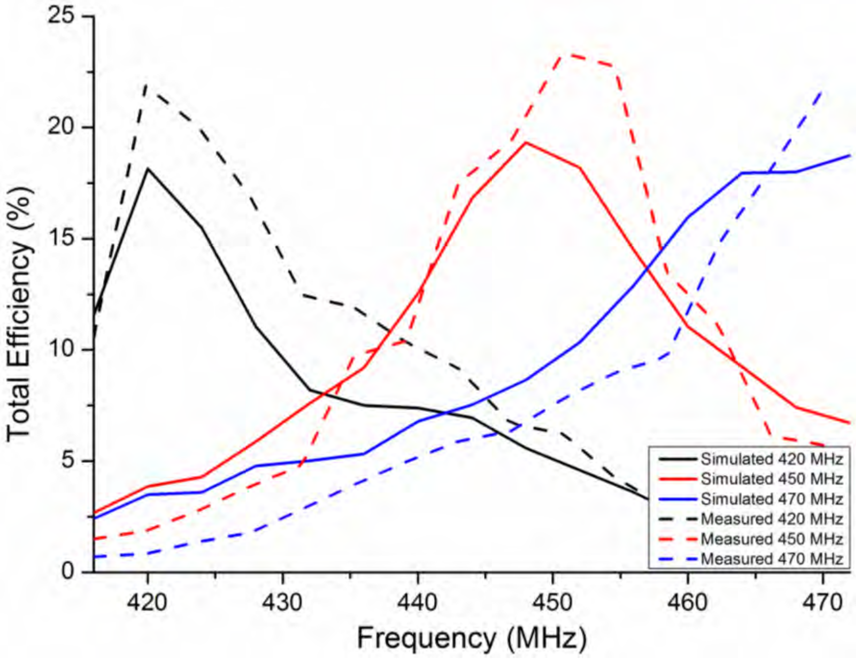

Figure 5.8. Total measured and simulated efficiency of the antenna at 420, 450 and 470 MHz.

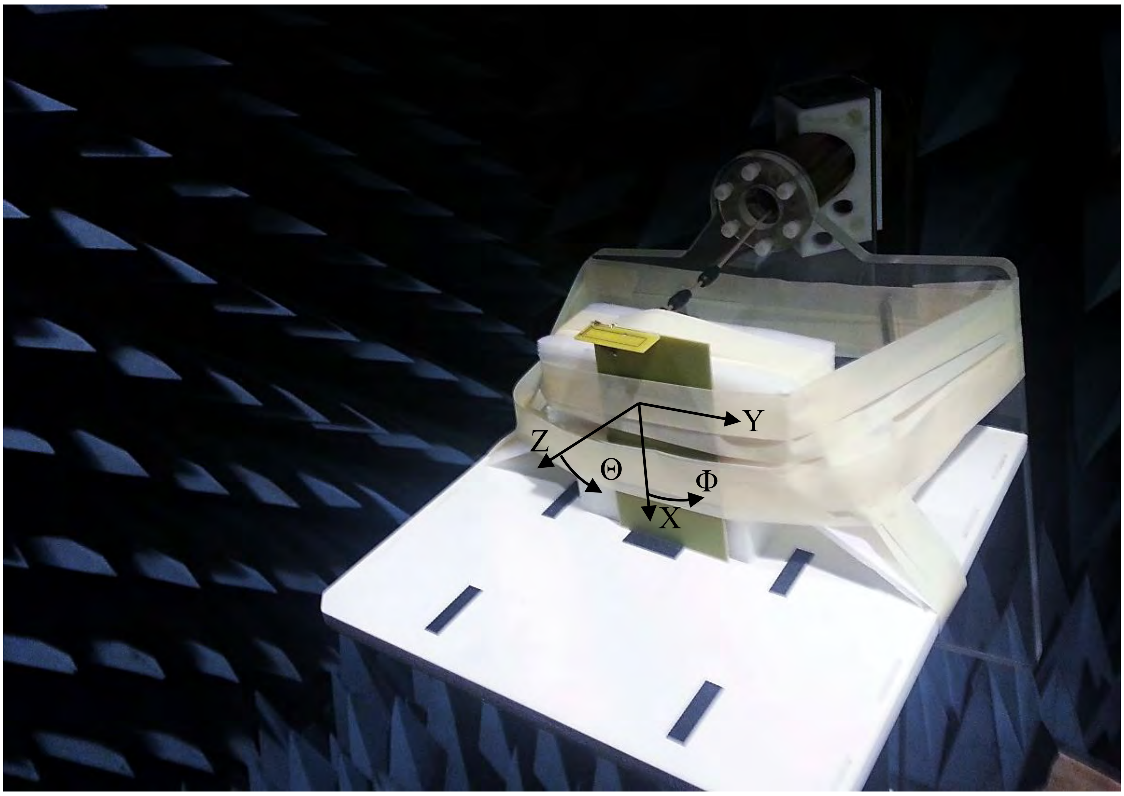

Figure 5.9. The antenna inside the anechoic chamber.

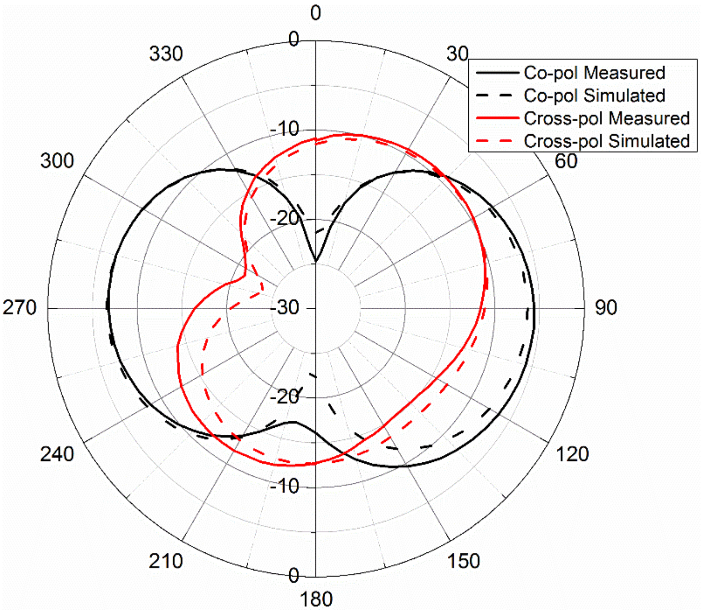

Figure 5.10. Azimuth radiation patterns in the xz-plane at 450 MHz.

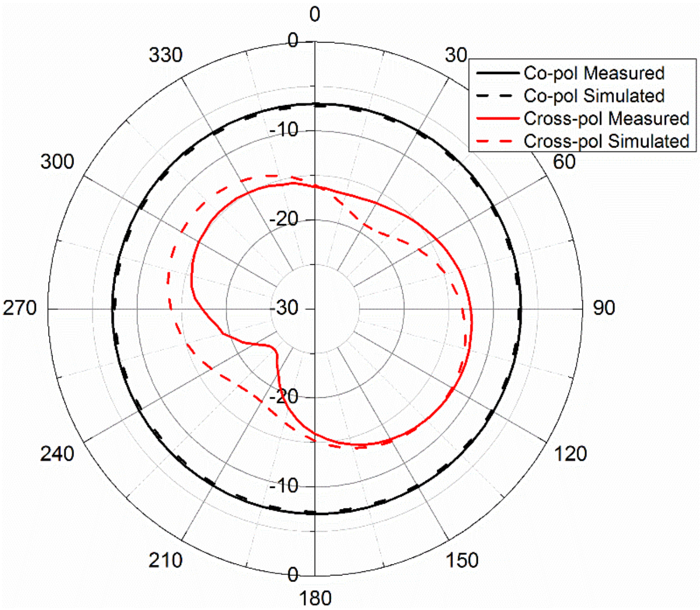

Figure 5.11. Elevation radiation patterns in the xy-plane at 450 MHz.

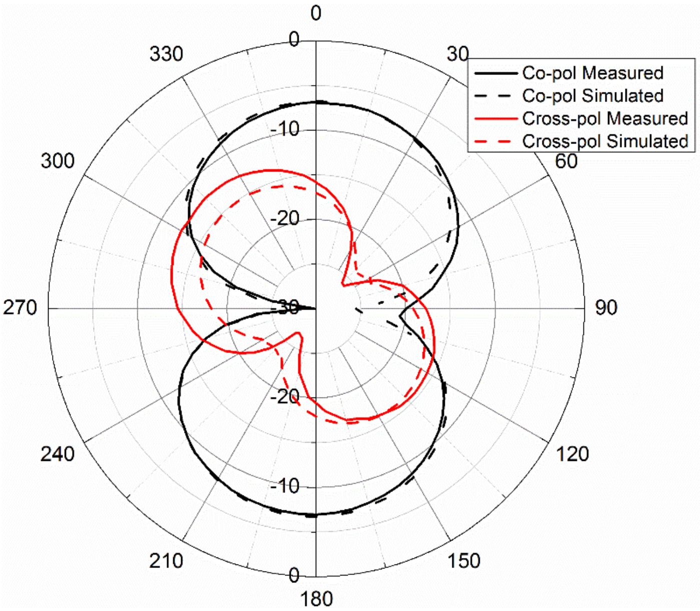

Figure 5.12. Elevation radiation patterns in the yz-plane at 450 MHz.

5.6 INSTALLED PERFORMANCE

To investigate the installed performance of the proposed antenna, a simulation model is created (Fig. 5.13) with the antenna in a Plexiglas housing ($\varepsilon_r = 3.6$, tanδ = 0.001 and thickness = 3 mm) with outer dimensions 81 × 266 × 172 mm^3 (H × L × W) and placed on a concrete wall ($\varepsilon_r = 5.8$ at 450 MHz, thickness = 200 mm). Two scenarios were evaluated, with the antenna ground plane oriented parallel and perpendicular to the wall surface as shown in the Fig. 5.13.

Figs. 5.14 and 5.15 show the measured and simulated S_{11} for the two different orientations. From inspection of the results, the resonant frequency shifts downwards due to the wall permittivity loading and this becomes more acute when the monopole is closer to the wall surface (parallel case).

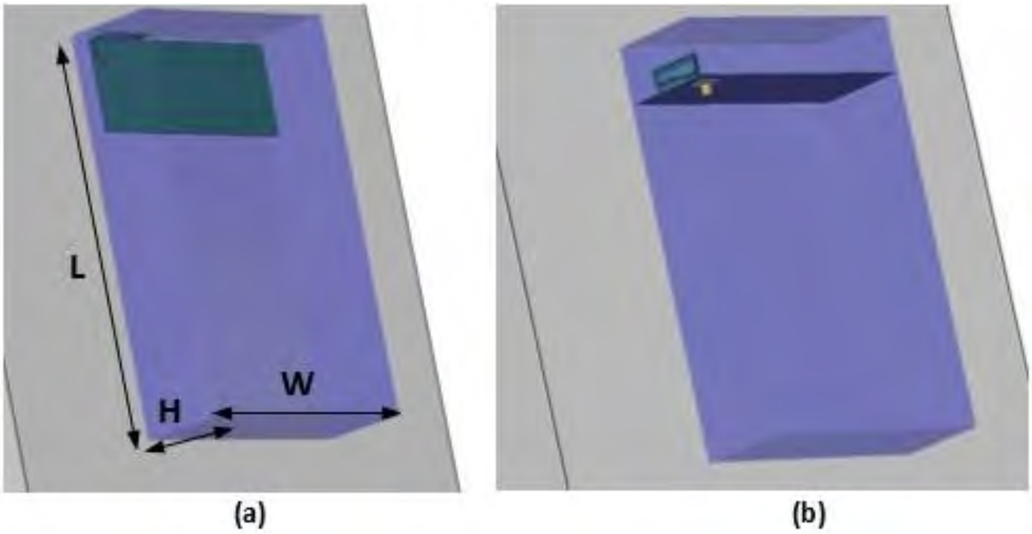

(a) (b)

Figure 5.13. The simulation model with ground plane (a) parallel to the wall and (b) perpendicular to the wall.

The distance between the edge of the ground plane and the wall for the parallel and perpendicular case was respectively 13.1 mm and 3.5 mm. For the parallel case the resonant frequency shifted downwards by 8 MHz for simulation and by 9 MHz measured. For the perpendicular case the frequency shifted downwards by 7 MHz for simulation and by 7.5 MHz for the measured case.

To mitigate the wall proximity loading detuning effects, the slider was adjusted to tune the antenna back to 450 MHz. This also improved the total efficiency for both cases. The simulated total efficiency for the parallel case before retuning was 17.6% and after retuning was 22.9%. For the perpendicular case the simulated total efficiency before retuning was 14.5% and after was 19.7%. In both cases some energy is absorbed by the concrete wall. This absorption is greater for the perpendicular case where the total efficiency is 3.2 percentage points less than for the parallel case. For both cases the via-slider was re-adjusted to tune the resonant frequency to 450 MHz. The via-slider positions were $P = 21.3$ mm and $P = 20.3$ mm for the parallel and perpendicular cases respectively.

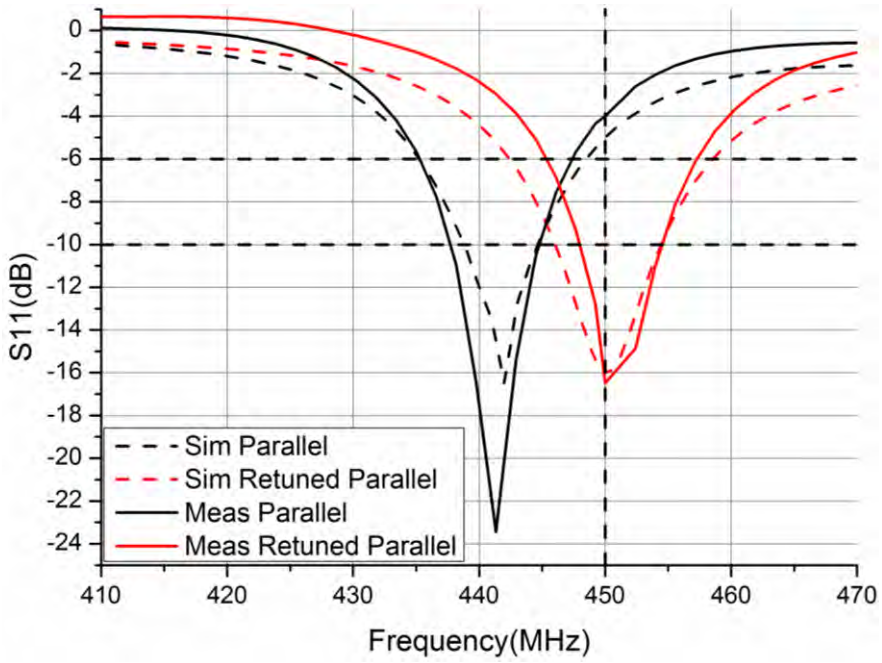

Figure 5.14. Simulated and measured S_{11} for the parallel case.

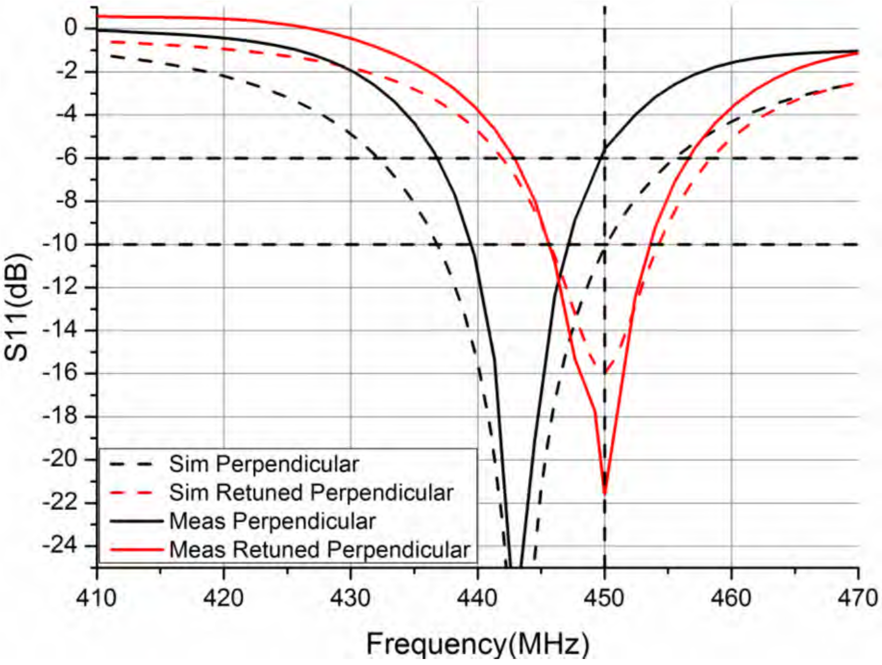

Figure 5.15. Simulated and measured S_{11} for the perpendicular case.

5.7 CONCLUSIONS

In this chapter a study of a meander-line folded monopole antenna for emerging applications in smart metering at 450 MHz was presented. The proposed compact antenna was studied in terms of the ground plane size, the matching stub role and the via-slider connector with respect to the monopole performance. The results show that the S_{11} characteristic is not heavily dependent on the ground plane size. The frequency is shifted downwards by 23 MHz with a significant impedance matching improvement by adding a matching stub into the antenna geometry. The electrical length of the monopole is increased and controlled by introducing a via-slider which connects the sides of the PCB. The rear feeding and grounding position allows more freedom of available space for the meander-line and confines the electrical length into a compact physical size. The antenna provides 21% to 24% measured total-efficiency across the tuneable bandwidth.

The packaged antenna performed well despite material loading and when installed on a concrete wall. The via-slider tuneable position can re-adjusted to tune the resonant frequency of the monopole to 450 MHz.

6. VHF & UHF Meandered Monopole Antenna

Intelligent metering systems for the measurement of water, gas, heat and electricity consumption enable modern smart cities. These smart metering applications require a long range link between the sensor grid nodes and the data aggregators, overcoming building penetration losses. The demand for harmonised frequency bands for several existing and new applications led the Electronic Communication Committee (ECC) [82] to reconsider the use of the 169 MHz and 433 MHz bands. The Wireless Metering Bus (WM-Bus) specification, introduced a narrow band channel of 75 kHz with a maximum EIRP of +27 dBm (+500 mW) [83] providing long range solutions.

The size of a smart metering device is typically around 150 × 150 mm^2 with a volume which must house the electronic components, battery and the integrated antenna. Meander line structures [21, 22, 23, 24] due to their low compact size are suitable for integration in small envelopes achieving broadband performance with omnidirectional characteristics.

In this chapter a dual-band folded meander line monopole antenna for Advanced Metering Infrastructure (AMI) [84] and Wireless-M-Bus applications operating in the

VHF and lower UHF bands is presented. Performance measures include impedance bandwidth, radiation patterns, total and radiation efficiency.

6.1 VHF & UHF MEANDERED MONOPOLE ANTENNA

A dual-band printed antenna is designed for integration in smart meter devices for Wireless M-Bus and M2M applications operating at 169 MHz and 433 MHz. The antenna configuration is optimized for impedance matching at both frequencies. A parametric investigation of key geometrical parameters is reported. The proposed antenna provides omnidirectional radiation characteristics and excellent measured total efficiency.

6.2 ANTENNA CONFIGURATION

The basic procedure steps for designing a novel dual-band miniaturized antenna based on a meander line topology is presented in Fig. 6.1. As the first step, a basic folded monopole (Fig. 6.1 (a)) is employed to design the antenna. The antenna resonates at around 1100 MHz with poor impedance matching (Fig. 6.2 blue line). The two horizontal strips (feeding and shorting) are meander (Fig. 6.1 (b)) in order to introduce more electrical length to the antenna geometry shifting the resonant frequency downward by 500 MHz (Fig. 6.2 red line). The final design step is defined with the shorting strip not be shorted anymore, but with an open end and positioning to the back side of the substrate (Fig. 6.1 (c)) providing better isolation from the feeding strip and increasing the electrical length of the antenna. The last antenna is dual band (Fig. 6.2 black line) with the lowest resonant be shifted downwards more than 300 MHz from the previous one.

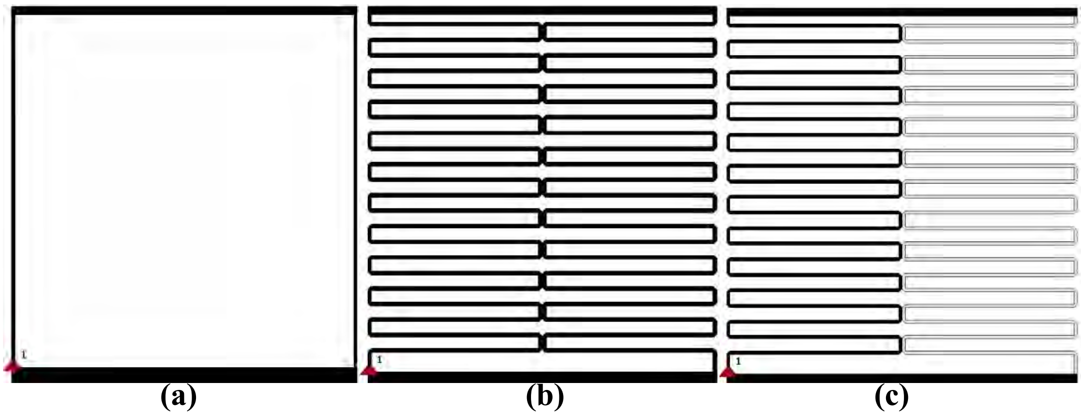

Figure 6.1. Schematic illustration of the three basic design models.

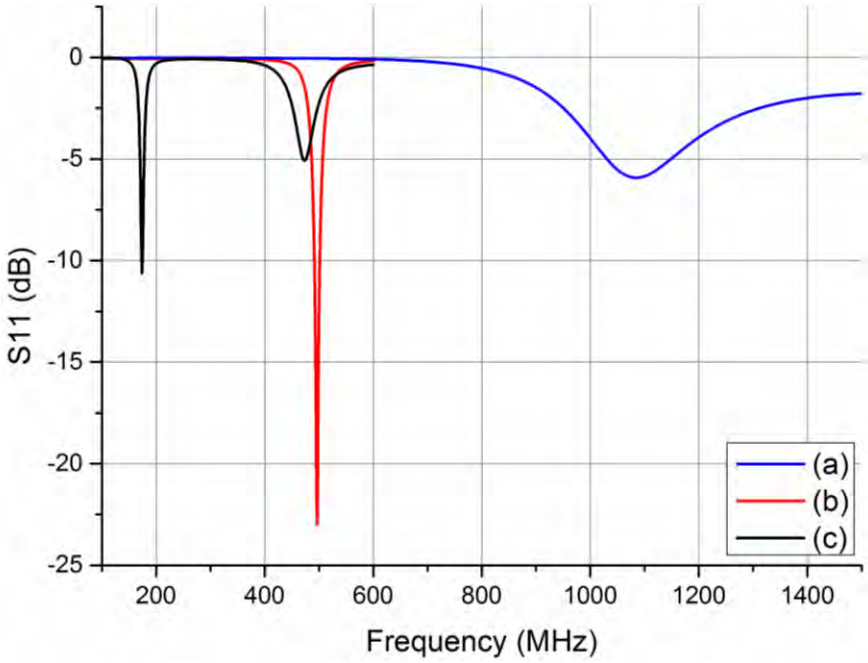

Figure 6.2. Simulated S_{11} for the three basic design models.

The antenna (Fig. 6.3) is printed on double-sided FR4 substrate, (Tabled 2.1) with dimension of 150 mm × 150 mm ≈ ($0.085\lambda_0$, $\lambda_0 = 1775.15$ mm at 169 MHz) representing the meter motherboard. The meandered monopole of height $h = 40.23$ mm ($\approx 0.02\lambda_0$) and length $l = 45.02$ mm ($\approx 0.025\lambda_0$) is located on the corner of the ground plane. The antenna is fed with an SMA connector through a 50 Ω microstrip line (width = 2.94 mm) which is stepped to a short section (length = 0.85 mm and width = 0.35 mm) which connects to the antenna through via 1. The rear meander line (width = 0.35 mm) with uniform spacing runs up the substrate to via 2 which connects to an extended meander line section located at the front with an 4.25 mm open end which terminates 0.31 mm away from the ground plane level. The introduction of a shunt lumped inductor (L = 56 nH, 0.2 Ω) 20.15 mm away from via 1 improves the matching at both frequencies as shown in Fig. 6.4.

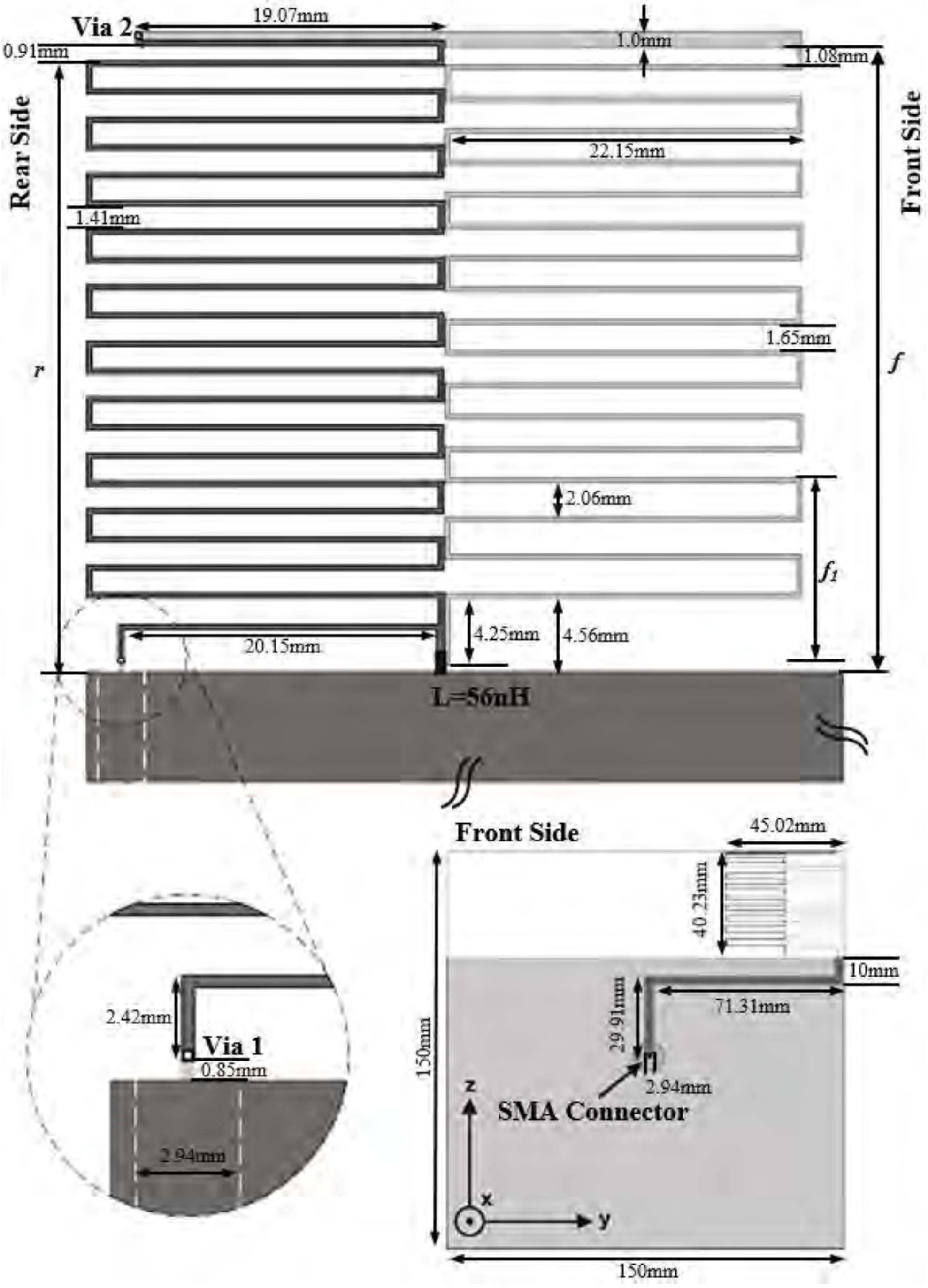
Via 2
19.07mm
1.0mm
0.91mm
1.08mm
Rear Side
Front Side
22.15mm
1.41mm
1.65mm
r
f
2.06mm
f1
4.25mm
4.56mm
20.15mm
L=56nH
Front Side
45.02mm
40.23mm
10mm
71.31mm
29.91mm
2.42mm
Via 1
0.85mm
150mm
z
x
y
2.94mm
SMA Connector
2.94mm
150mm

Figure 6.3. Folded meander line monopole antenna and coordinate system.

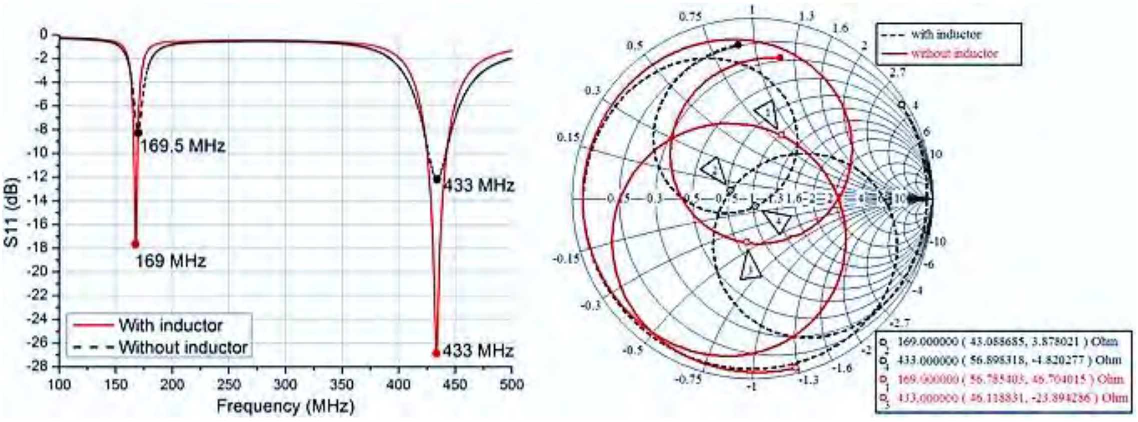

Figure 6.4. Improved impedance matching and control of frequency.

Practically, the presence of lumped elements (inductors/capacitors) introduce additional losses to antenna, but the main advantage of such designs is to improve impedance matching (Fig. 6.4), reducing the missmatch losses and thus, to increase the total efficiency of the antenna system. Looking at the S_{11} results in Fig. 6.4 it becomes evident that the introduction of an inductor reduce the missmatch losses from 0.75 dB to 0.07 dB for the first resonance and from 0.28 dB to 0.01 dB for the second one.

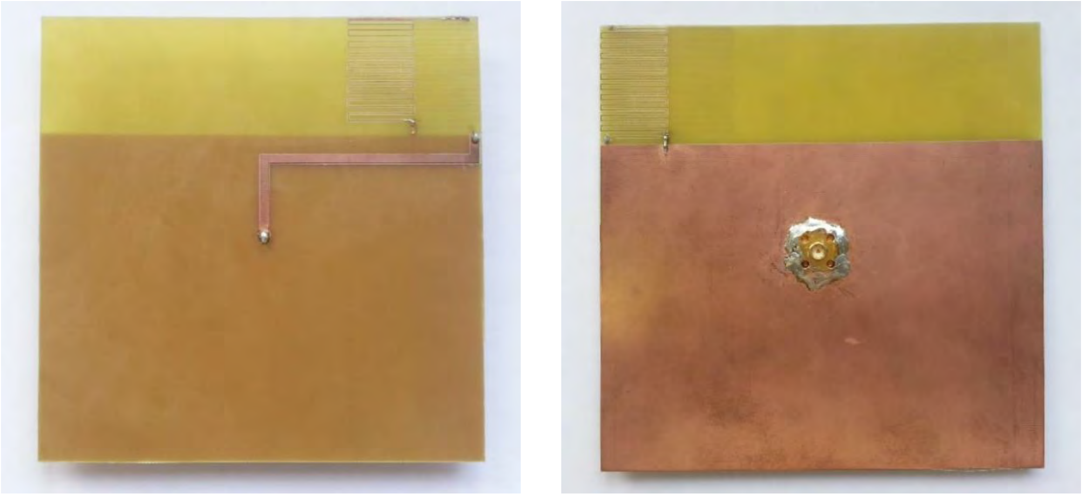

(a) **(b)**

Figure 6.5. Manufactured meandered monopole prototype (a) Front View, (b) Rear View.

6.3 PARAMETRIC INVESTIGATION

A parametric investigation of the variation of S_{11} with respect to three geometric parameters of the meander line monopole shown in Fig. 6.1 is made.

The dual band operation of the monopole is strongly dependence on the height (*r*) of the rear meandered section, which was varied from 34 mm to 58 mm. The results indicate (Fig. 6.6) that as the height (*r*) is increased the first resonant frequency is shifted upwards while the second resonance is shifted downwards. From the current distribution Fig. 6.9(b) it can be seen that the rear meandered section mainly determinates the upper resonance. By selecting the value of *r* = 38.3 mm both resonances can be tune to the desired values.

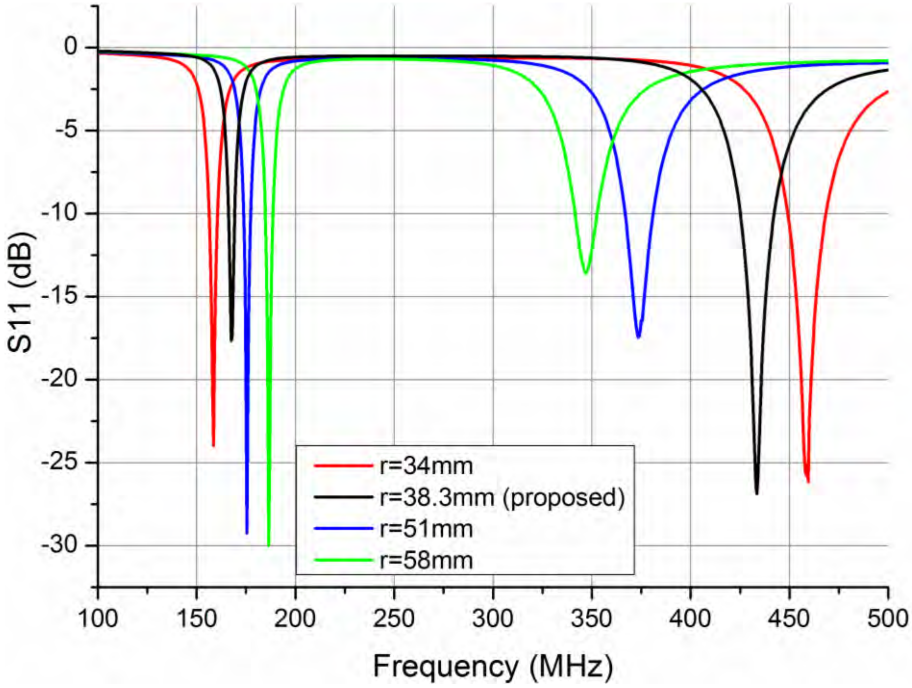

Figure 6.6. The simulated S_{11} dependence on the height *r* of the back meander line section.

The height (*f*) of the front meander line section of the monopole was found to have a significant influence on the frequency-ratio range of the antenna. This height was varied from 30 mm to 48 mm. The simulations were made with uniform separation between the meander line strips. From the S_{11} (Fig. 6.7) it is clearly visible that increasing the value (*f*) shifts the first resonance downwards with little effect on the upper resonance. The frequency-ratio between the upper and the lower resonance $F_r = f_u/f_l$ continuously increases from 1.61 to 3.11 as the value of (*f*) changes from 30 mm to 48 mm.

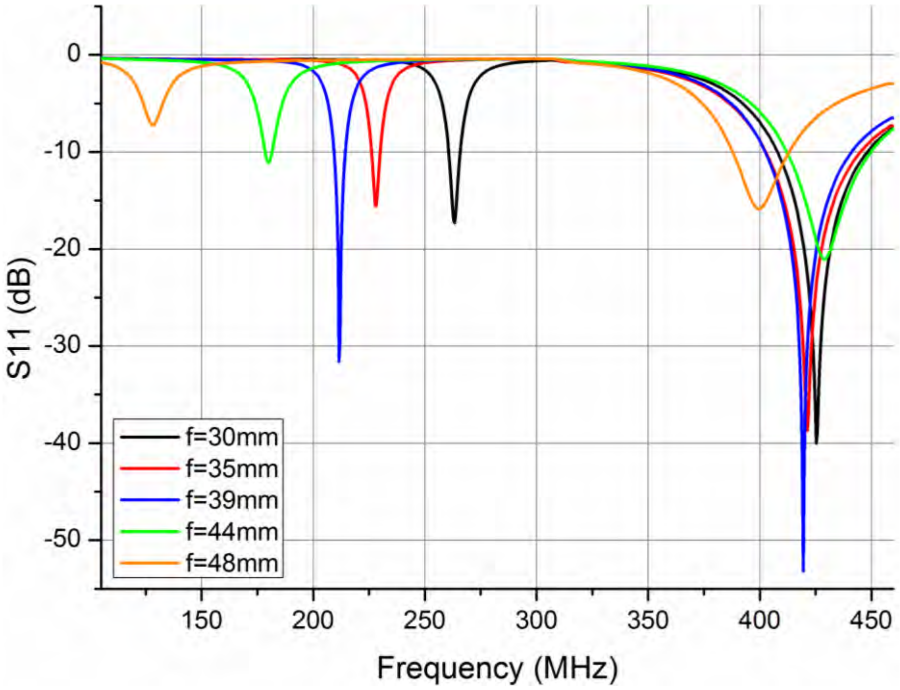

Figure 6.7. The simulated S_{11} dependence on the height *f* of the front meander line section.

The dependence of the first resonance on the height (f_1) of the last two meander lines on the front of the monopole was investigated. The results are shown in Fig. 6.8. When the height (f_1) of the two last meander line strips is increased from 10 mm to 17.7 mm the first resonance is shifted from 203 MHz to 144.5 MHz with little effect on the impedance matching. The second resonant frequency is not heavily dependent on from the height (f_1). This is confirmed by the current distribution (Fig. 6.9(b)) where the four last meander line strips do not carry a strong current. The proposed value is f_1 = 13.34 mm.

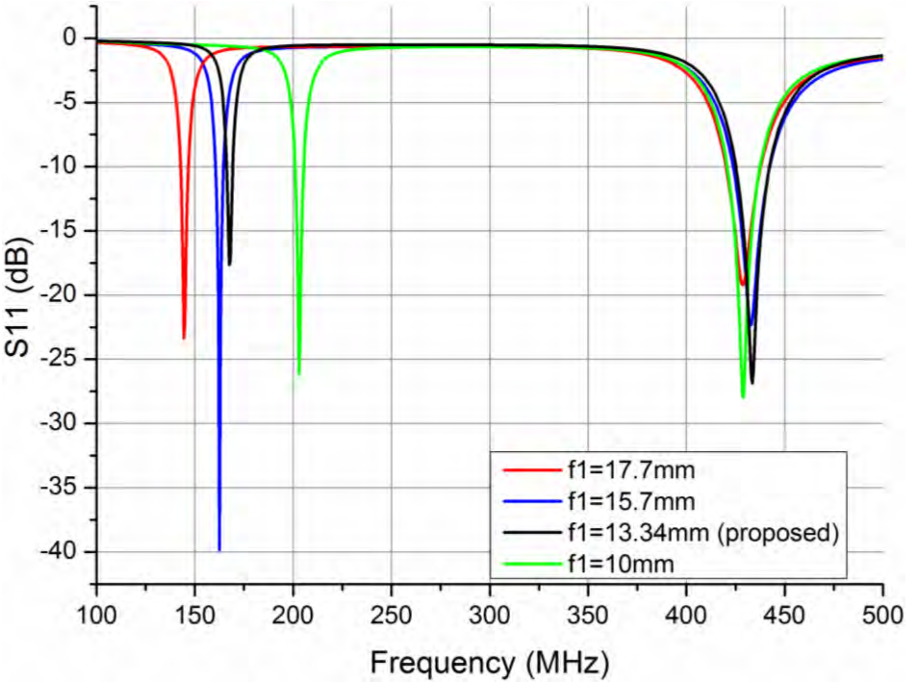

Figure 6.8. The simulated S_{11} dependence on the height f_1 of the last front four meander line strips.

The performance of the antenna is based the fundamental frequency at which the structure resonates and radiates. That fundamental frequency is related to the antenna's resonant length which is caused by current paths (Fig. 6.9) within the structure. For the proposed antenna the lower rear side meandered section and the upper front side meandered sections determine the upper resonance (Fig. 6.9(b)) and the combined line sections provide the lower resonance (Fig. 6.9(a)). In both resonant frequencies the ground plane excites strong currents at the horizontal and the vertical side which have same phase with the currents of the monopole. In the proposed antenna the ground plane is an integral part and reinforces and contributes to the radiation performance. It should be noted that in the upper resonance (Fig. 6.9 (b)) the current reversed the direction after a full wavelength along the rear side meandered section.

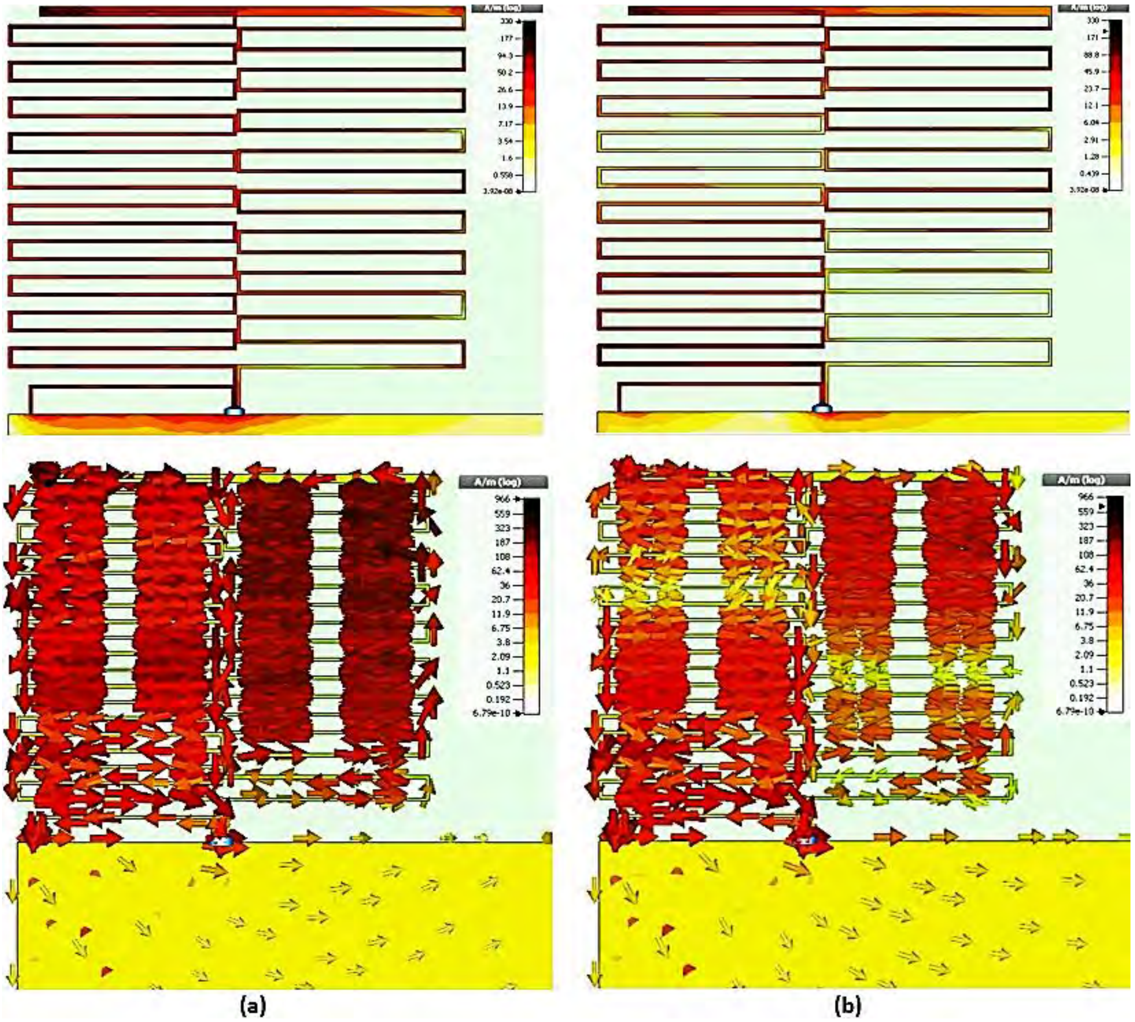

Figure 6.9. The current distribution of the antenna in (a) 169 MHz and (b) 433 MHz.

Placing meander line sections on opposite sides of the PCB provides isolation, miniaturization and greater freedom in controlling the frequency ratio. The frequency ratio f_u/f_l of 2.56 is tuned and optimized by non-uniformity in the front meander line section as well as some coupling of the front and back meander lines.

6.4 SIMULATED AND MEASURED RESULTS

In Fig. 6.10 the measured and simulated S_{11} is presented providing good agreement. The measured −6 dB and −10 dB impedance bandwidth was 5.8 MHz (165.7 − 171.5 MHz) and 3.3 MHz (167 − 170.3 MHz) respectively at 169 MHz and 53.9 MHz (417.5 − 471.4 MHz) and 26.4 MHz (424.7 − 451.1 MHz) at 433 MHz. The fractional bandwidths in −6 dB is 3.4% for the first resonance and 12.4% for the second, while the fractional bandwidths in −10 dB is 1.95% for the first resonant frequency and 6.1% for the second.

The simulated −6 dB and −10 dB impedance bandwidth was 5.7 MHz (164.9 − 170.6 MHz) and 3.1 MHz (166.2 − 169.3 MHz) at 169 MHz and 30.4 MHz (419.1 − 449.5 MHz) and 17.2 MHz (425.1 − 442.3 MHz) at 433 MHz.

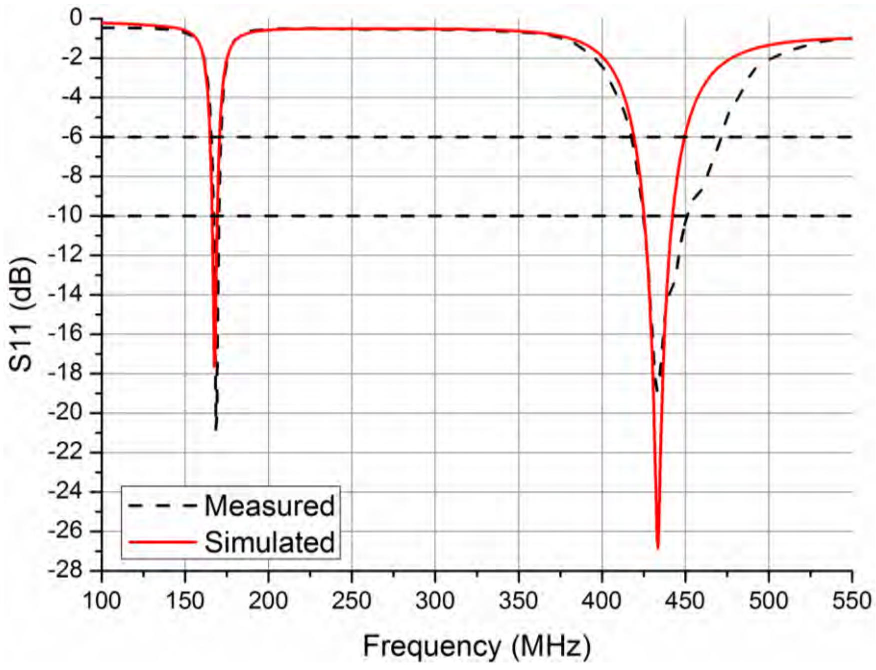

Figure 6.10. Simulated and measured S_{11}.

Efficiency measurements are performed using the Wheeler cap method [85] (Fig 6.11) using a metallic box of 610 × 610 × 360 mm^3. The measured radiation and total efficiency for the first resonance is 20.4%. For the second resonance the measured radiation and total efficiency is 50% and 49.4% respectively. These results are in good agreement with simulation which indicate a radiation and total efficiency of 18.3% and 17.8% at 169 MHz and 48.2% and 47.9% at 433 MHz respectively.

In Figs. 6.12 and 6.13 the measured and simulated azimuth (x–y) and elevation (x–z) plane radiation patterns at 433 MHz are illustrated. The maximum measured gain is 4.6 dBi. There is a good agreement between measurement and simulation. The maximum simulated realized gain at 169 MHz is −1.8 dBi. In AHFR Lab there is no suitable SGH available to test and measure the radiation patterns at 169 MHz.

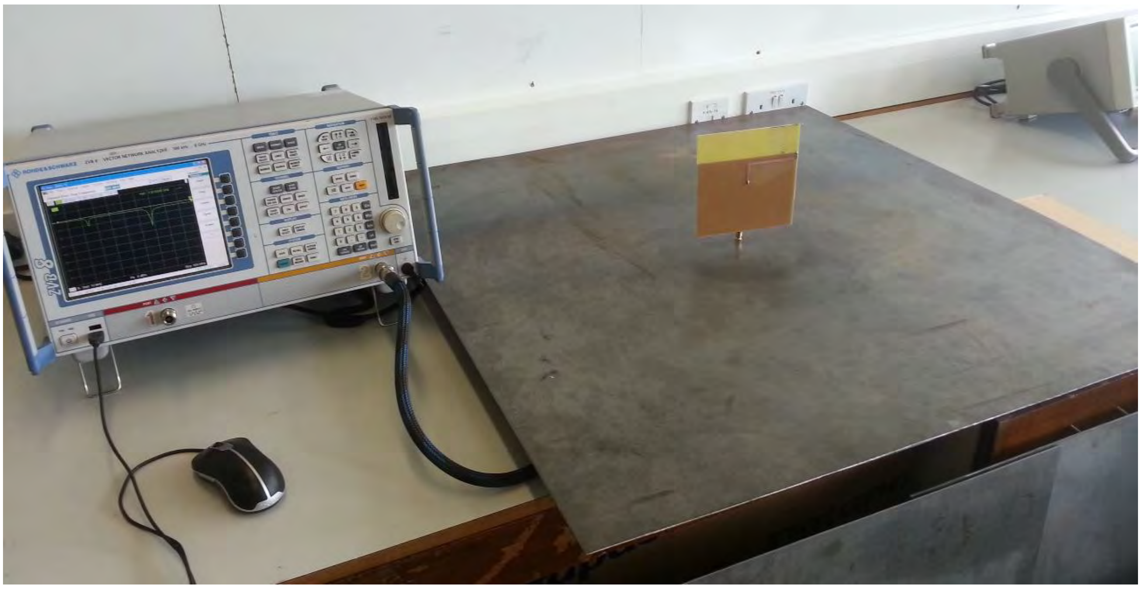

(a)

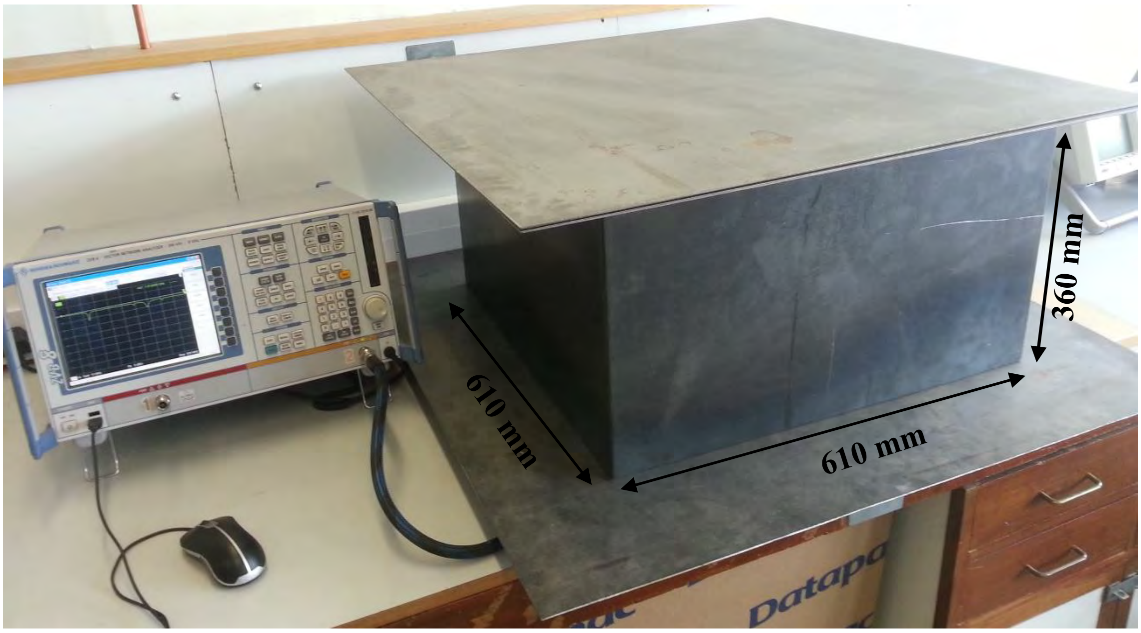
610 mm
610 mm
360 mm

(b)

Figure 6.11. Wheeler Cap measurement setup (a) without cap, (b) with cap.

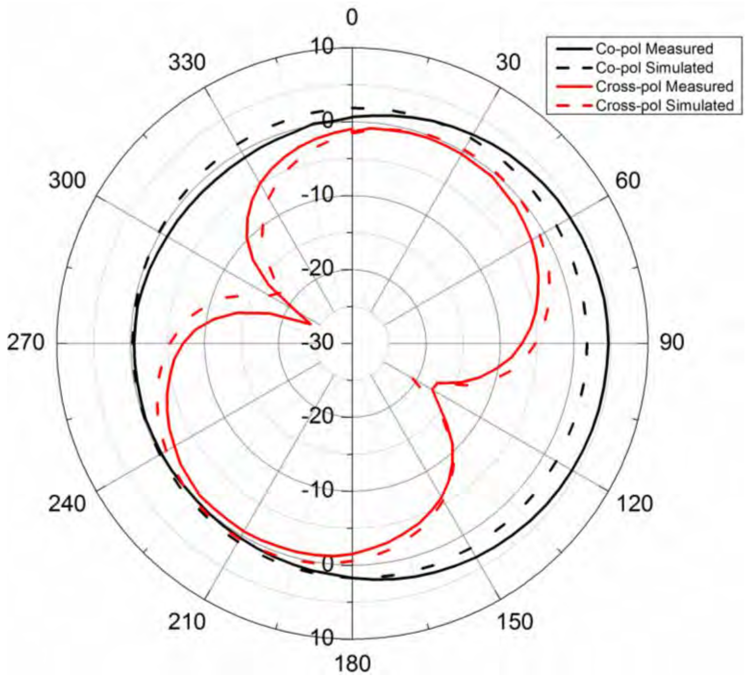

Figure 6.12. Azimuth radiation patterns in the xy-plane at 433 MHz.

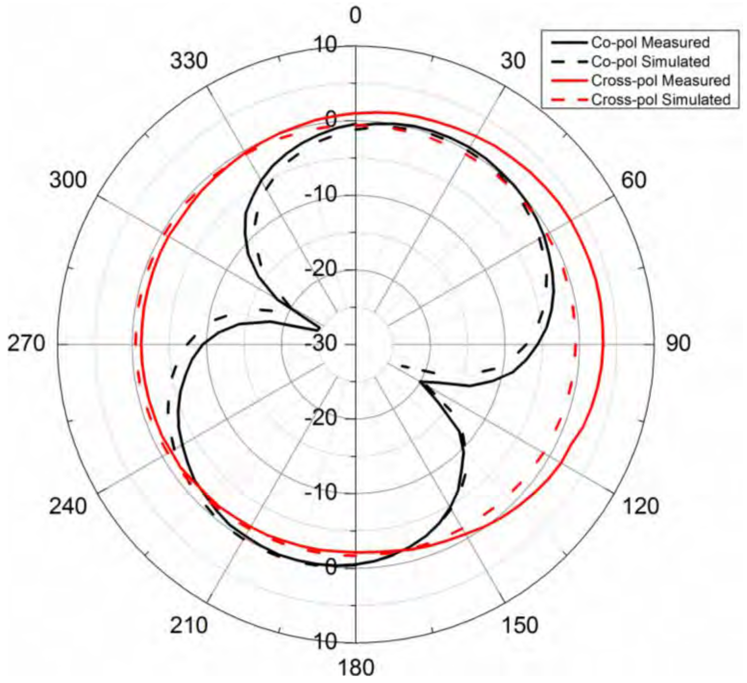

Figure 6.13. Elevation radiation patterns in the xz-plane at 433 MHz.

6.5 CONCLUSIONS

In this chapter a dual band folded meander line monopole antenna for emerging smart metering and M2M applications at 169 and 433 MHz is presented. Three key geometrical parameters of the proposed electrically small antenna were studied as well as the effect to the frequency ratio $F_r = f_u/f_l$ is reported. The introducing of a shunt lumped inductor was found to significantly improve the impedance matching of both resonant frequencies. The antenna provides good omnidirectional radiation characteristic at 433 MHz and more than 20% and 49% measured total efficiency at 169 MHz and 433 MHz respectively. The measured gain at 433 MHz is 4.6 dBi and the simulated gain at 169 MHz is −1.8 dBi. The antenna is low cost and easy fabricated with additional benefit of a large easy controlled frequency-ratio range.

7. Conclusions and Future Work

7.1 CONCLUSIONS

A number of small novel antenna designs have been researched, developed and are described in this thesis. Parametric studies were carried out on folded and meander line monopoles and planar inverted-F antennas. The proposed geometries are optimised and designed to provide sufficient bandwidth, good impedance matching, omnidirectional radiation characteristics with high gain and efficiency, maintaining compact size.

The main contributions of this dissertation are the following:

A folded and shorted printed monopole antenna has been proposed. A compact low profile structure has been designed based on these principles offering an easy controlled large frequency-ratio range and stable broad band omnidirectional radiation characteristics. Moreover, this approach provides a good polarization purity between horizontal and vertical polarized signals which is essential for maintaining the good link between the receiver and the transmitters.

Two Planar Inverted-F structures have been designed operating broad band, with high gain and efficiency. The effects of their key dimensional parameters have been parametrically studied. Moreover, the antenna performance for different locations on the ground plane and for various dimensions of ground plane layout have been investigated accordingly. A size reduction of almost 50% was achieved when two PIFAs combined in one antenna share the same feed. The study has demonstrated a controllable large frequency-ratio range when the feeding and shorting point change position.

By seeking more compact solutions for low frequencies, two novel meander line folded monopole antennas were designed. Miniaturization has been provided by double sided PCB printed technique allowing integrating more electrical length into a compact physical size and isolating the two printed sides. Matching stub structures have been applied in order to improve the impedance matching. A via slider connector has been used to connect the two sides of the PCB, enabling tuneability of the impedance matching across a large frequency range and offering stable omnidiarectionality over the tuneable range. The strong dependence of the frequency-ratio range on the meander line separation was investigated in depth.

7.2 FUTURE WORK

During the development and test phase of this dissertation some research topics which are worth further investigation have been identified and are outlined here.

Further investigation could be made in topics related with installed performance of various antennas when they are integrated in actual devices. When an antenna is integrated in housing, the impedance bandwidth was slightly increased as well as the resonant frequency was shifted downwards [86]. The walls of the enclosure are effectively decreasing the gain and the efficiency of the installed antenna. Future work could focus on investigating the positioning of an antenna inside an enclosure in order to find the optimum position and orientation.

Another interesting topic for extensive investigation is the packaging impact to an integrated antenna in an actual device which contains numerous metal structures that could potentially interfere its performance. Metal structures near antennas can change input impedances and phase of received signals. Design guidelines to eliminate the electromagnetic interference (EMI) emitted from the RF components and circuits could be developed [87, 88]. To eliminate the electromagnetic emissions and keep EMI to a minimum it would be very important to arrange the layout of ground planes to minimize EMI [89, 90].

Further development could be done on ground planes which have been found to affect the general performance of antennas [91, 92, 93]. The position and the orientation of the antenna on the ground plane can dramatically affect the general performance [94]. Further research could be devoted to the extensive investigation of different ground plane shapes.

Two compact solutions suitable for integration into packaging were proposed in this thesis for smart metering applications. The double sided PCB printed meandered shape and the matching stub technique were the two techniques for miniaturization providing isolation between the meander line sections and improving the impedance matching across the operating frequency range. A further study on broad-banding meander line antennas should be made as well as an investigation on providing more than 2 bands [95, 96].

A compact VHF antenna is proposed suitable for integration into packaging. Test and measurements of the radiation patterns at 169 MHz were not carried out due to the non-availability of a large sized anechoic chamber and a suitable Standard Gain Horn antenna. Further research efforts should focus on the outdoor measurements [97, 98] where many factors, such as the directivity of the antenna under test, the frequency range and mechanical features of the antenna (size, weight and volume) should be considered to ensure the performance metrics are measured with sufficient accuracy.

Bibliography

[1] C. A. Balanis, *Antenna Theory*, New Jersey: John Wiley & Sons, 2005.

[2] J.-H. Kim and H.-M. Lee, "Efficient Electrically Small Wide Band Planar Antenna for SAR Reduction," *in 20th International Zurich Symposium on Electromagnetic Compatibility*, Zurich, 2009, pp. 269-272.

[3] F. C. C. (FCC), "Wireless Devices and Health Concerns," 2014. [Online]. Available: http://transition.fcc.gov/cgb/consumerfacts/mobilephone.pdf.

[4] *Standard for Safety Levels with Respect to Human Exposure to Radio Frequency Electromagnetic Fields, 3 kHz to 300 GHz*, IEEE International Committee on Electromagnetic Safety IEEE Std C95.1-1991, April 19 2006.

[5] H. Wheeler, "Fundamental limitations of small Antennas," *IRE Proc*, vol. 35, no. 12, pp. 1479-1484, December 1947.

[6] H. Wheeler, "The Radiansphere Around a Small Antenna," *IEEE Proceedings of The I.R.E.*, vol. 47, no. 8, pp. 1325-1331, August 1959.

[7] L. Chu, "Physical limitations of omni-directional antennas," *J. Appl. Phys.*, vol. 19, no. 12, pp. 1163-1175, May 1948.

[8] R. Harrington, "On the Gain and Beamwidth of Directional Antennas," *IEEE Transactions on Antennas and Propagation*, vol. 6, no. 3, pp. 219-225, July 1958.

[9] R. Fano, "Theoretical limitations on the broadband matching of arbitrary impedances," *J. Franklin Inst.*, vol. 249, no. 1, pp. 57-83, January 1950.

[10] W. Stutzman and G. Thiele, *Antenna Theory and Design*, New York: John Wiley & Sons, 1998.

[11] C. Lin, "Dual-Band Folded Monopole Antenna with Slotted Ground Plane for WLAN Applications," vol. 15, pp. 53-60, 2010.

[12] A. Loutridis, M. John, G. Ruvio and M. J. Ammann, "A 2.45/5.8 GHz folded monopole antenna for WLAN applications," *in Antennas and Propagation Conference (LAPC)*, 2012 Loughborough, UK, 2012.

[13] T. Ito, T. Oki, H. Morishita, "Current reduction on the ground plane using additional element for WiMAX/WLAN folded monopole antenna," *in IEEE International Workshop on Electromagnetics (iWEM)*, Sapporo, 2014.

[14] G. Augustin, B.P. Chacko, T.A. Denidni, "Uniplanar folded monopole antenna for mobile phone applications in LTE/GSM/UMTS/WiFi band," *in IEEE Antennas and Propagation Society International Symposium (APSURSI)*, Memphis, 2014.

[15] J. Volakis, C. Chen, J. Halloran and S. Koulouridis, "Miniature VHF/UHF conformal spirals with inductive and ferrite loading," *in IEEE Antennas and Propagation Society International Symposium*, Honolulu, 2007, pp. 5-8.

[16] A. Loutridis, K. Yang, M. John and M. J. Ammann, "Compact Printed Spiral FM Antenna," *in Progress in Electromagnetics Research Symposium, Prague (PIERS)*, Czech Republic, 2015, pp. 37.

[17] C. Borja and J. Romeu, "On the behaviour of Koch island fractal boundary microstrip patch antenna," *IEEE Antennas and Propagation Society*, vol. 51, no. 6, pp. 1281 - 1291, June 2003.

[18] J. Gonzalez-Arbesu, S. Blanch and J. Romeu, "Are space-filling curves efficient small antennas?" *IEEE Antennas and Propagation Society*, vol. 2, no. 1, pp. 147 - 150, 2003.

[19] S. R. Best, "On the Resonant Behaviour of the Small Koch Fractal Monopole Antenna," *Microwave and Optical Technology Letters*, vol. 35, Issue: 4, pp. 311-315. Nov. 2002.

[20] K. J. Vinoy, K. A. Jose, V. K. Varadan and V. V. Varadan, "Hilbert Curve Fractal Antenna: A Small Resonant Antenna for VHF / UHF Applications," Microwave and Optical Technology Letters, vol. 29, Issue: 4, pp. 215-219. May 2001.

[21] A. Loutridis, M. John and M. J. Ammann, "Printed Folded Meander Line Dual-Band Monopole for TV White Space and GSM," *in 8th European Conference on Antennas and Propagation (EuCAP)*, The Hague, Netherlands, 2014, pp. 2848-2852.

[22] C.-N. Hu, W. Chen, J.-W. Huang and J. Chiu, "Design of a multiband coupled-meander-line monopole antenna for M2M applications," *in International Conference on Applications of Electromagnetism and Student Innovation Competition Awards (AEM2C)*, Taipei, 2010, pp. 264-268.

[23] S. R. Best and J. D. Morrow, "Limitations of inductive circuit model representations of meander line antennas," *in Antennas and Propagation Society International Symposium*, Columbus, OH, USA, 2003, pp. 852-855.

[24] T. J. Warnagiris and T. J. Minardo, "Performance of a meandered line as an electrically small transmitting antenna," *Transactions on Antennas and Propagation*, vol. 46, no. 12, pp. 1797 - 1801, 1998.

[25] V. Gonzalez-Posadas, C. Martin-Pascual, J. Jiménez-Martín and D. Segovia-Vargas, "Lumped-Element Balun for UHF UWB Printed Balanced Antennas," *IEEE Antennas and Propagation Society*, vol. 56, no. 7, pp. 2102 - 2107, July 2008.

[26] Niamien, M. A. C., Sharaiha, A., Collardey, S. Mahdjoubi, K., "An electrically small frequency reconfigurable antenna for DVB-H," *in IEEE International Workshop on Antenna Technology (IWAT)*, Tucson, March 2012, pp. 245-248.

[27] M. -I. Lai, T. -Y. Wu, J. -C. Hsieh, C. -H. Wang, S. -K. Jeng, " Design of reconfigurable antennas based on an L-shaped slot and PIN diodes for compact wireless devices," *IET Microwaves Antennas 7 Propagation*, vol. 3, Issue: 1, pp. 47-54, Feb. 2009.

[28] A. Alhawari, A. Ismail and M. Mahdi, " Reconfigurable Loaded Planar Inverted-F Antenna Using Varactor Diodes," *IEEE Antennas and Wireless Propagation Letters,* vol. 10, pp. 466-468, May 2011.

[29] S. -H. Chang, W. -J. Liao, "A Broadband LTE/WWAN Antenna Design for Tablet PC," *in IEEE Transactions on Antennas and Propagation*, vol. 60, Issue: 9, pp. 4354-4359, Aug. 2012.

[30] R. Mittra and S. Dey, "Challenges in PCS antenna design," *in IEEE Antennas and Propagation Society International Symposium*, vol. 1, Orlando USA, 1999, pp. 544-547.

[31] A.K. Skrivervik, J. F. Zurcher, O. Staub, J.R. Mosig, "PCS antenna design: the challenge of miniaturization," *IEEE Antennas and Propagation Magazine*, vol. 43, Issue: 4, pp. 12-27, 2001.

[32] K. Noguchi, S. Betsudan, T. Katagi and M. Mizusawa, "A compact broad-band helical antenna with two-wire helix," *IEEE Transactions on Antennas and Propagation*, vol. 51, no. 9, pp. 2176 - 2181, Sep. 2003.

[33] Sooliam Ooi, Boon Ping Koh, "Single-fed Dual Band UHF-GPS Helical Antenna," *in IEEE International Workshop on Antenna Technology (IWAT)*, March 2006, pp 184-187.

[34] D. Kearney, M. John and M. J. Ammann, "Miniature Ceramic PIFA for UWB Band Group 3 & 6," *IEEE Antennas and Wireless Propagation Letters*, vol. 9, pp. 28-31, Jan. 2010.

[35] A. Loutridis, M. John, M. J. Ammann, "A Dual Band LTE PIFA Antenna for M2M Applications," *Microwave and Optical Technology Letters*, vol. 57, Issue: 7, pp. 1655-1658. July 2015.

[36] H.-K. Ryu, E. Kim and J.-M. Woo, "Miniaturisation of printed inverted-F antenna using chip coupler for Bluetooth applications," *Electronics Letters*, vol. 46, no. 13, pp. 883 - 885, June 2010.

[37] D. Nashaat, H. Elsadek and H. Ghali, "Single Feed Compact Quad-Band PIFA Antenna for Wireless Communication Applications," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 8, pp. 2631 - 2635, Aug. 2005.

[38] L. Wei-Yu, W. Chun-Yih, W. Kin-Lu and T. Ming-Fang, "Internal Small-Size PIFA for LTE/GSM/UMTS Operation in the Mobile Phone," *in IEEE Antennas and Propagation Society International Symposium (APSURSI)*, Toronto, 2010.

[39] S. Bokhari, J.-F. Zurcher, J. Mosig and F. Gardiol, "A small microstrip patch antenna with a convenient tuning option," *IEEE Transactions on Antennas and Propagation*, vol. 44, no. 11, pp. 1521 - 1528, Nov. 1996.

[40] K. Wong, J. Kuo and T. Chiou, "Compact microstrip antennas with slots loaded in the ground plane," *in 11th International Conference on Antennas and Propagation*, vol. 2, Manchester, 2001, pp. 623-626.

[41] R. F. Graf, *Modern Dictionary of Electronics*, Newnes, 1999, p. 29.

[42] *Standard Definitions of Terms for Antennas*, IEEE Antennas and Propagation Society Std 145-1993, Mar. 18 1993.

[43] L. Kibona, "Analysis of the Effects of Rectangular Ground Plane on Radiation Pattern of the Monopole Antenna," *International Journal of Scientific and Research Publications (IJSRP)*, vol. 3, 2013.

[44] R. C. Johnson and H. Jasik, *Antenna Engineering Handbook*, 2nd ed., New York: McGraw – Hill, 1984.

[45] Z.N. Chen, Ed., *Antennas for Portable Devices*, 1st ed. Wiley-VCH, 2007

[46] Z.N. Chen, M.Y.W. Chia, *Broadband Planar Antennas*, Wiley, 2006

[47] K.-L. Wong, *Planar Antennas For Wireless Communications*, New Jersey: Wiley, 2003.

[48] G. Ruvio and M. J. Ammann, "A Novel Wideband Semi-Planar Miniaturized Antenna," *IEEE Transactions on Antennas & Propagation*, vol. 55, no. 6, pp. 1760-1767, 2007.

[49] T. Endo, Y. Sunahara, S. Satoh and T. Katagi. "Resonant Frequency and Radiation Efficiency of Meander Line Antennas," *Electronics and Communications in Japan. Part 2*, vol. 83, no. 1, pp. 52-58, 2000.

[50] T. J. Warnagirism and T. J. Minardo. "Performance of a Meandered Line as an Electrically Small Transmitting Antenna," *IEEE Trans. Antennas Propagation* vol. 46, issue: 12, pp. 1797-1801, Dec. 1998.

[51] K. Fujimoto, A. Henderson, K. Hirasawa and J.R. James, *Small Antennas*, Research Studies Press, England; John Wiley and Sons, New York 1987.

[52] W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, John Wiley & Sons, 2012.

[53] *Standard Definitions of Terms for Antennas*, IEEE Antennas and Propagation Society Std 145-1983, June. 22 1983.

[54] J. S. McLean, "A Re-Examination of the Fundamental Limits on the Radiation Q of Electrically Small Antennas," *IEEE Transactions on Antennas and Propagation*, vol. 44, no. 5, pp. 672-676, May 1996.

[55] R. E. Colin and S. Rothschild, "Evaluation of Antenna Q," *IEEE Transactions on Antennas and Propagation*, vol. 12, issue: 1, pp. 23-27, 1964.

[56] R. L. Fante, "Quality factor of general ideal antenna," *IEEE Transactions on Antennas and Propagation*, vol. 17, issue: 2, pp. 151-155, 1696.

[57] J. M. Gonzalez, J. Romeu, E. Cabot and J. R, "Exploring the limits of Fractal Electrodynamics for the future telecommunication technologies," IST-2001-33055, 2003.

[58] D. Pozar, B. Kaufman, "Comparison of Three Methods for the Measurement of Printed Antenna Efficiency," *IEEE Trans. Antennas Propagation*, vol. 36, No. 1, pp. 136-139, January 1988.

[59] R.H. Johnston, J.G. McRory, "An improved small antenna radiation-efficiency measurement method," *IEEE Antennas and Propagation Magazine*, vol. 40, issue: 5, pp. 40-48, Oct. 1998

[60] W. E. McKinzie III, "A Modified Wheeler Cap Method for Measuring Antenna Efficiency," *in IEEE Antennas and Propagation Society International Symposium*, vol. 1, Montreal, Quebec, Canada, 1997, pp. 542-545.

[61] C.L. Holloway, H.A. Shah, R.J. Pirkl, W.F. Young, D.A. Hill, J. Ladbury, "Reverberation Chamber Techniques for Determining the Radiation and Total Efficiency of Antennas," *IEEE Trans. Antennas Propagation*, vol. 60, issue: 4, pp. 1758-1770, January 2012.

[62] G. Le Fur, C. Lemoine, P. Besnier, A. Sharaiha, "Performances of UWB Wheeler Cap and Reverberation Chamber to Carry Out Efficiency Measurements of Narrowband Antennas," *IEEE Antennas and Wireless Propagation Letters*, vol. 8, pp. 332-335, September 2008.

[63] CST GmbH – Computer Simulation Technology. 2014. [Online]. Available: https://www.cst.com/.

[64] T. Weiland, "A discretization method for the solution of Maxwell's equations for six-component fields," *Electronics and Communication (AEÜ)*, vol. 31, p. 116–120, 1997.

[65] M. Clemens and T. Weiland, "Discrete electromagnetism with the finite," *Progress In Electromagnetics Research (PIER)*, vol. 32, p. 65–87, 2001.

[66] M. J. Ammann, "A comparison of some low cost laminates for antennas operating in the 2.45 GHz ISM band," *in IEE Colloquium on Low Cost Antenna Technology*, London, 1998.

[67] M. Hiebel, *Fundamentals of Vector Network Analysis*, Munchen: Rohde&Schwarz GmbH & Co., 2014.

[68] Rohde&Schwarz, "R&S Vector Network Analyzers ZVA 8 / ZVA 24 / ZVA 40 Operating Manual," Rohde&Schwarz, 2005.

[69] E.A.C. NV, "ECCOSORB VHP-NRL very high Performance Broadband Pyramidal Absorber," EMERSON & CUMING Anechoic Chambers, 2010.

[Online]. Available: http://www.ecanechoicchambers.com/TB/EB-100%20-%20VHP-NRL.pdf.

[70] "Schwarzbeck Mess Elektronik," Schwarzbeck Mess Elektronik, 2014. [Online]. Available: http://www.schwarzbeck.com/.

[71] *Standard Test Procedures for Antennas*, IEEE Std. No. 149-1979, 1979.

[72] A. T. Arkko, "Effect of ground plane size on the free-space performance of a mobile handset PIFA antenna," *in Proc. 12th IEE International Conference on Antennas and Propagation*, vol. 1, Exeter, UK, Mar. 2003, pp. 316–319.

[73] R. F. Harrington, J. R. Mautz, "Theory of characteristic modes for conducting bodies," *IEEE Transactions on. Antennas and Propagation*, vol. 19, no. 5, pp. 622–628, Sept. 1971.

[74] MVG StarLab 650 MHz – 18 GHz, 2014. [Online]. Available: http://www.mvg-world.com/system/files/starlab_2014.pdf

[75] C. Brasek, "Urban utilities warm up to the idea of wireless automatic meter reading," *IEE Computing & Control Engineering Journal*, vol. 15, issue: 6, pp. 10-14, Jan. 2004.

[76] E. C. C. Report 39, "The technical impact of introducing CDMA-PAMR on 12.5 / 25 kHz PMR/PAMR technologies in the 410-430 and 450-470 MHz bands."

[77] R. E. Badra, "Compared RF Performance of 1.4 MHz-LTE and EVDO rev. A in Rural Environments at 450 MHz," *in 78th Vehicular Technology Conference (VTC Fall)*, Las Vegas, NV, 2013.

[78] B. Holfeld, S. Jaeckel, L. Thiele, T. Wirth, K. Scheppelmann, "Smart Grid Communications: LTE Outdoor Field Trials at 450 MHz," *in 81st Vechicalr Technology Conference (VCT Spring)*, Glasgow, UK, 2015.

[79] P.-S. Kildal, S.R. Best, "Further investigations of fundamental directivity limitations of small antennas with and without ground planes," *in IEEE Antennas and Propagation Society International Symposium (APSURSI)*, San Diego, 2008, pp. 1-4.

[80] C. Hagerling, C. Ide, C. Wietfeld, "Coverage and capacity analysis of wireless M2M technologies for smart distribution grid services," *in IEEE International Conference on Smart Grid Communications*, Venice, 2014, pp. 368-373.

[81] Z. M. Fadlulah, M. M. Fouda, N. Kato, A. Takeuchi, N. Iwasaki and Y. Nozaki, "Toward intelligent machine-to-machine communications in smart grid," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 60-65, 2011.

[82] Electronic Communication Committee (ECC) Decision (05)02, "A harmonised frequency plan for the use of the band 169.4-169.8125 MHz," Amended 8 November 2013.

[83] European Committee for Standardization, "EN 13757-4," 2013.

[84] D. Li, B. Hu, "Advanced metering standard infrastructure for smart grid," *in China Intrenational Conference on Electricity Distribution (CICED)*, China, 2012.

[85] E. Higashi, Y. Iida, Y. Omura, "A Study of Antenna Efficiency Measurements by Wheeler Cap Method Applied to Planer Inverted-F Antenna (PIFA)," *in International Symposium on Signals, Systems and Electronics (ISSSE)*, Montreal, 2007, pp. 599-601.

[86] M. John, P. McEvoy and M. J. Ammann, "Evaluation of ABS enclosed printed monopole antenna supplied by Benetel Limited," Dublin Institute of Technology, Tech. Rep., Sep. 2007.

[87] Clayton R. Paul, *Introduction to Electromagnetic Compatibility*, Wiley-Interscience Publication, 1992.

[88] Mohit Arora, *The Art of Hardware Architecture: Design Methods and Techniques Digital Circuits*, Springer Science & Business Media, LLC 2012.

[89] M. Steer, *Microwave and RF Design: A Systems Approach*, 2nd edition. Edison, NJ: SciTech Publishing, 2013.

[90] M. Golio and J. Golio, Eds., *RF and Microwave Circuits, Measurements and Modeling*, 2nd edition. Boca Raton, FL: CRC Press, 2007.

[91] N.L. Bohannon, J.T. Bernhard, "Ground plane effects on planar inverted-F antennas," *in IEEE Antennas and Propagation Society International Symposium (APSURSI)*, Chicago, 2012.

[92] L. Cheng-Tse, W. Kin-Lu, "Internal WWAN Clamshell Mobile Phone Antenna Using a Current Trap for Reduced Ground Plane Effects," *IEEE Transactions on Antennas and Propagation*, vol. 57, Issue 10, pp. 3303-3308, 2009.

[93] Y. Lu, Y. Huang, H.T. Chattha, Y. Shen, "Technique for minimising the effects of ground plane on planar ultra-wideband monopole antennas," *IET Microwaves, Antennas & Propagation*, vol. 6, Issue 5, pp. 510-518, 2012.

[94] C.C. Chiau, C. Xiaodong, C. G. Parini, "Effect of antenna's positioning on a handset for a miniature dielectric-loaded folded half-loop antenna," *in 1st European Conference on Antennas and Propagation*, Nice, 2006.

[95] K. S. Sultan, H. H. Abdullah, E. A. Abdallah, E. A. Hashish, "Low SAR, compact and multiband antenna for mobile and wireless communication," *in Middle East Conference on Antennas and Propagation (MECAP)*, Cairo, Dec. 2012.

[96] M. K. Taher, "Small size multiband meander line antenna for wireless applications," *in Loughborough Antennas and Propagation Conference (LAPC)*, Loughborough, UK, March 2008, pp. 401-404.

[97] R. Hartman and Jack Berlekamp, "Fundamentals of Antenna Test and Evaluation," *in Microwave Systems New and Communications Tracking*, June 1988.

[98] J.S. Hollis, T.J. Lyon and L. Clayton, eds., *Microwave Antenna Measurements*, Scientific-Atlanta, Inc., 1985.

Appendix A.

LIST OF PUBLICATIONS

In this dissertation many of the ideas and designs have been published in various journals and international conferences. These publications are listed here in reverse chronological order.

Journal Publications

[JP1] K. Yang, A. Loutridis, X. Bao, G. Ruvio and M. J. Ammann, "Printed Inverted-F Antenna with Reconfigurable Pattern and Polarization", *IEEE Transactions Antennas and Propagation Letters*; (In review).

[JP2] A. Loutridis, M. John and M. J. Ammann, "Folded Meandered Monopole for Emerging Smart Metering & M2M Applications in the Lower UHF Band," *IEEE Antenna & Propagation Magazine*; (In review).

[JP3] A. Loutridis, M. John and M. J. Ammann, "Folded Meander Line Antenna for Wireless M-Bus in the VHF and UHF Bands," *Electronics Letters IET*, vol. 51, no. 15, pp. 1138-1140, July 2015.

[JP4] A. Loutridis, M. John, M. J. Ammann, "A Dual Band LTE PIFA Antenna for M2M Applications," *Microwave and Optical Technology Letters*, vol. 57, no. 7, pp. 1655-1658, July 2015.

[JP5] A. Loutridis, M. John, M. J. Ammann, "Dual Band LTE Planar Inverted-F Antenna for M2M Applications," *Microwave and Optical Technology Letters*, vol. 22, no. 12, pp. 2925–2929, Dec 2013.

Conference Proceedings

[CP1] A. Loutridis, K. Yang, M. John and M. J. Ammann, "Dual Band Printed Antenna for M2M Applications in ISM and GSM bands," *Loughborough Antennas & Propagation Conference (LAPC)*, Loughborough, UK. 2015; (Accepted).

[CP2] A. Loutridis, K. Yang, M. John and M. J. Ammann, "Compact Printed Spiral FM Antenna," *Progress in Electromagnetics Research Symposium, (PIERS)*, Prague, Czech Republic, July 2015.

[CP3] A. Loutridis, M. John and M. J. Ammann, "Printed Folded Meander Line Dual-Band Monopole for TV White Space and GSM," *European Conference of Antennas and Propagation (EuCAP)*, Hague, Netherlands, Feb. 2014, pp. 2848-2852.

[CP4] A. Loutridis, M. John, M. J. Ammann, G. Ruvio, "A Wideband Printed Inverted-F Antenna for M2M Applications in the Lower UHF Band," *Loughborough Antennas & Propagation Conference (LAPC)*, Loughborough, UK, Nov. 2013, pp. 242-245.

[CP5] A. Loutridis, M. John, G. Ruvio, M. J. Ammann, "A 2.45/5.8 GHz Folded Monopole Antenna for WLAN Applications," *Loughborough Antennas & Propagation Conference (LAPC)*, Loughborough, UK, Nov. 2012.

Wireless Engineering and Technology, 2014, 5, 107-116
Published Online October 2014 in SciRes. http://www.scirp.org/journal/wet
http://dx.doi.org/10.4236/wet.2014.54012

An Investigation of Wideband Rectennas for Wireless Energy Harvesting

Jing-Wei Zhang, Yi Huang, Ping Cao

Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, UK
Email: zjw131@hotmail.com

Received 7 July 2014; revised 9 August 2014; accepted 27 August 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

This paper is focused on a wireless energy harvesting system using a rectifying antenna (rectenna). The proposed device consists of a wideband cross-dipole antenna, a microwave low-pass filter and a doubling rectifying circuit using Shottcky diodes as rectifying elements. Previously, a few of wideband rectennas have been investigated at 1.7 to 2.5 GHz. The originality of this paper is on the new wideband rectenna design which can harvest the ambient radio frequency (RF) power at 1.7 to 2.5 GHz. In this system, a new wideband cross dipole is designed and used to achieve the required bandwidth and duel-polarization. In addition, the voltage doubling rectifying circuit is optimized to achieve the best performance at power density levels < 200 μW/cm² which are typical in urban environments. The characteristics of the proposed rectenna over the desired frequency range are investigated, and the integrated rectenna is simulated, made and tested for low input power densities from 5 to 200 μW/cm². The simulation and measurement results of the rectenna are compared and a good agreement is achieved. The results demonstrate that the maximum rectenna conversion efficiency is nearly 57% around 1.7 GHz and over 20% over the wideband of interest for the incident power density of 120 μW/cm². It is noted that the impedance matching is one of the main factors affecting the rectenna energy conversion efficiency. This new wideband rectenna has great potential to harvest wireless energy in GSM/3G/4G and ISM 2.4 GHz bands.

Keywords

Rectennas, Wireless Energy Harvesting, Wideband Dipoles, Rectenna Conversion Efficiency

1. Introduction

With the rapid progress of wireless communications, there have been many wireless systems developed and installed in our cities. The Radio Frequency (RF) power density has become quite considerable in many areas.

How to cite this paper: Zhang, J.-W., Huang, Y. and Cao, P. (2014) An Investigation of Wideband Rectennas for Wireless Energy Harvesting. *Wireless Engineering and Technology*, **5**, 107-116. http://dx.doi.org/10.4236/wet.2014.54012

Currently, the RF energy harvesting has attracted much attention to supply power for some small devices such as wearable and battery-free sensors [1]-[3]. The core device of such a wireless energy harvesting system is the rectifying antenna which is also called as the rectenna. The rectenna is a combination of a rectifier and an antenna, and can receive electromagnetic waves and convert them to direct current (DC). The block diagram of a rectenna device is plotted in **Figure 1** which consists of an antenna, a microwave low-pass filter (MLPF)/microwave band-pass filter (MBPF), a rectifier, a DC filter and a load. The receiving antenna connects the MLPF/MBPF which matches the impedance of antenna to the impedance of rectifier and prevents higher order harmonics generated by the rectifier (diodes) re-radiating back to the free space by the antenna. The cut-off frequency of filters is slightly higher than the fundamental frequency. In addition, the DC filter effectively provides a short circuit for the alternative current (AC) and allows the DC current passing to the load which is placed at the output port to measure the output power.

A key performance indicator for a rectenna is the rectenna energy conversion efficiency which is not only of practical importance but also a useful tool to analyze the system performance. This efficiency (η) is a product of the antenna and the rectifier efficiency η_a and η_r, that is,

$$\eta = \eta_a \eta_r \tag{1}$$

where η_a is the antenna total efficiency and η_r is the conversion efficiency of the rectifier. This term includes both circuit and rectifying diode losses.

Previously, a range of rectennas have been investigated such as dipole rectennas [4]-[6], patch rectennas [7], spiral rectennas [8], and slot rectennas [9]. Most of these rectenna designs are for narrowband operations (typically around the 2.5 GHz ISM band) for relatively high input power density (>0.5 mW/cm^2) applications and very high conversion efficiencies (over 60%) were achieved. However, they produce very low conversion efficiency for low-power density levels and not suitable for RF energy harvesting applications. It is very important to note that most of the power in air in the RF spectrum is in GSM, 3G and Wi-Fi frequency bands [10]. Since a wideband rectenna can capture more wireless energy in air than a narrowband rectenna, a wideband rectenna covering 1.7 to 2.4 GHz with relatively high conversion efficiency for low-input energy harvesting is of great interest.

In this paper, a new wideband integrated rectenna is proposed to achieve a good performance from 1.7 to 2.5 GHz to cover most of the GSM, 3G, LTE (4G) and Wi-Fi services. The design is particularly suitable for applications with a low-input power density (<0.2 mW/cm^2). This paper is organized as follows: Section 2 briefly discusses the design of a wideband rectenna system and demonstrates the performance of the wideband cross dipole antenna, the microwave low-pass filter and the rectifier. A wideband dipole is optimized to achieve the required bandwidth. A microwave low-pass filter is designed to prevent higher order harmonics and the voltage doubling rectifying circuit is characterized for a low input power. In Section 3, the wideband rectenna is implemented and measured in anechoic chamber and outdoor environments. Moreover, the results of rectenna conversion efficiency and the DC output voltage are presented and analyzed. Finally, the conclusion is drawn in Section 4.

2. The Rectenna Design and Analysis

A major problem in the rectenna design is how to deal with the frequency-dependent impedance of the antenna and the rectifier. For the maximum power transfer, the impedance of an antenna must match with the impedance of the rectifying circuit. In this Section, a wideband rectenna for wireless energy harvesting is designed and investigated. A rectenna system can be separated into three parts which are a receiving antenna, a microwave low-pass filter and a rectifying circuit as shown in **Figure 1**. Furthermore, higher order harmonics

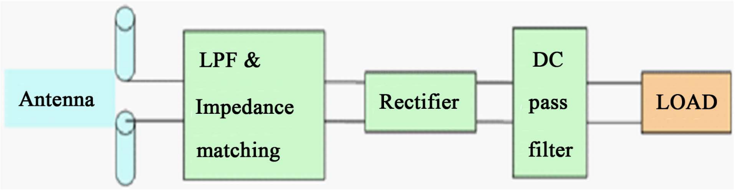

Figure 1. The block diagram of the rectenna device.

produced by the rectifying circuit should be removed by the microwave low-pass filter. In addition, the rectifier is optimized to match the antenna with a filter.

2.1. The Antenna Design

The first step in the rectenna design is to design a wideband antenna with a low-pass filter (to be discussed late) which matches to 50 Ω. A cross-dipole antenna shown in **Figure 2(a)** was proposed as the receiving antenna for the following reasons. Firstly, the antenna structure with parallel feed lines is convenient to connect with a low-pass filter. Secondly, the antenna has wide bandwidth to receive signals at our frequency region of interest, and thirdly it can receive signals of different polarization which is relatively random in reality.

The proposed antenna is fed by a parallel strip line which is on the surface of a dielectric sheet. The detailed characteristics of the parallel strip line can be found in [11]. All the dimensions of the antenna are optimized using CST software package in order to produce the best performance over the frequency range of interest. The length of the proposed and optimized antenna is around 80 mm which is 0.48 λ_0 at 1.8 GHz and 0.8 λ_0 at 3 GHz. The final dimensions are also given in **Figure 2(a)**. The antenna is built using a low-cost FR4 PCB (printed circuit board) as shown in **Figure 2(b)**. The measured and simulated reflection coefficients of the cross dipole antenna are shown in **Figure 3**. A good agreement is demonstrated between them, where S11 is less than −10 dB for the frequency band covers from 1.8 GHz to 2.9 GHz. The difference between them around the higher frequencies is mainly due to the cable effects in the measurement. Thus the design requirement is met.

2.2. The Antenna with a Microwave Low-Pass Filter (AMLPF)

A microwave low-pass filter is designed and also shown in **Figure 2(a)**. The low-pass filter has a cutoff frequency around 4 GHz, as shown in **Figure 4**, to pass signals from 1.8 to 3 GHz and to reject signals above 3.6 GHz, which are the higher order harmonics of operation frequencies.

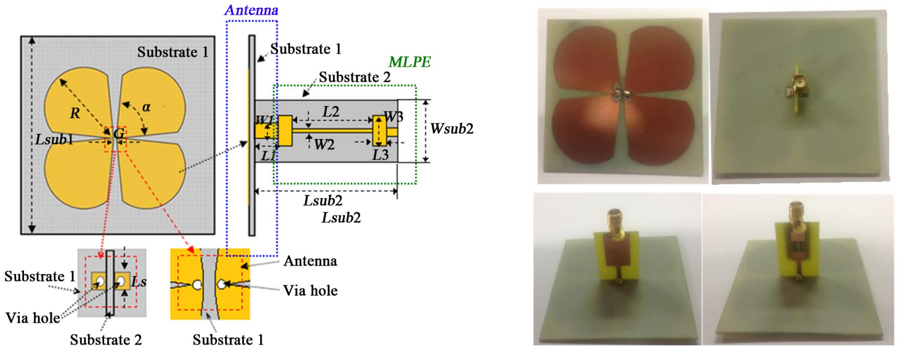

(a) (b)

Figure 2. (a) The layout of the rectenna with a low-pass filter. The parameters are: $\acute{a} = 75^{\circ}$, $R = 40$ mm, $G = 1.5$ mm, $Ls = 5$ mm, $Lsub1 = 10$ cm, $W1 = 2$ mm, $L1 = 10$ mm, $W3 = 6.7$ mm, $L3 = 5.1$ mm, $W2 = 0.3$ mm, $L2 = 6$ mm, $Lsub2 = 30.2$ mm, $Wsub2 = 20$ mm. (b) The photograph of the antenna with a low-pass filter.

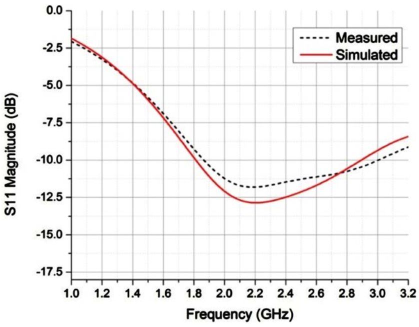

Figure 3. The measured and simulated S11 results of the proposed antenna.

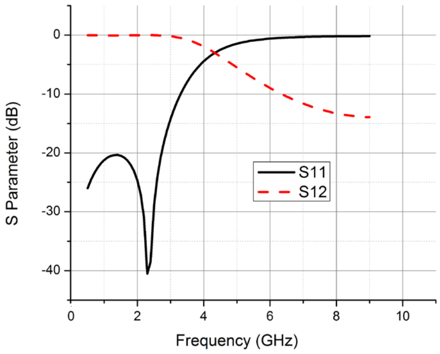

Figure 4. The simulated frequency response of the proposed microwave low-pass filter.

In addition, the microwave low-pass filter connected to the end of the antenna can tune the input impedance of AMLPF. The performance of AMLPF is optimized by using CST software to make sure that it matches with 50 Ω. The layout and photograph of the antenna with a low-pass filter structure are also shown in **Figure 2**. A SMA connector is connected to the end of the low-pass filter to measure the AMLPF performance. The simulated and measured results are shown in **Figure 5(a)** and compared to full-wave simulations, a reflection coefficient lower than −8 dB has been measured in frequency range from 1.7 GHz to 2.5 GHz. The discrepancy may be caused by the inaccuracy of the fabrication and the measurement result. It can be seen from the results, which are shown up to 9 GHz to include the frequencies of the 2nd and 3rd higher harmonics, this antenna with a low-pass filter should guarantee a good level of matching over the operation bandwidth while the at second and third harmonics should be well rejected. In addition, the Input impedance of the AMLPF is shown in **Figure 5(b)** and the desired performance is obtained over the frequency of the interest.

2.3. The Voltage Doubling Rectifying Circuit

As shown in Equation (1), the total conversion efficiency of a rectenna is basically determined by the antenna and the rectifying circuit performance. The goal of this Section is to introduce the voltage doubling rectifying circuit and to investigate its characteristics with the aid of a computer software package, Agilent ADS.

The equivalent circuit of a voltage doubling rectifying circuit is shown in **Figure 6**. The half-wave voltage doubling rectifying circuit consists of two sections; each comprises a diode and a capacitor for rectification. It functions as follows: on the negative half cycle of the input voltage, capacitor C1 is charged while on the positive half-cycle, the input voltage, in series with the voltage of C1 charges capacitor C2 to the desired output voltage. Capacitor C1, which aides charging capacitor C2, sees alternating current ("AC Cap") while C2 sees only direct current ("DC Cap"). Thus, the voltage on output capacitor is roughly two times the peak voltage of the diode. The conversion efficiency of a rectifying circuit strongly depends on the converted output voltage. Thus a voltage doubling rectifier circuit has the potential to achieve higher conversion efficiency than a single diode.

To validate such a voltage doubling rectifying circuit, the Schottky HSMS-2852 diode package is chosen because it has a low built-in voltage with a fast switching response and a high cut-off frequency. The voltage doubling rectifying circuit is built on a low-cost double-sided FR4 PCB with the thickness of 1.44 mm and relative permittivity ε_r of 4.4. Components in the rectifying circuit are connected by a micro-strip transmission line and placed on the top layer of the PCB as shown in **Figure 7**. In order to match the impedance of rectifying circuit to the AMLPF, all dimensions shown in **Figure 7** are optimized by using ADS over the target frequency range and input power levels. The back layer of board is the ground plane. Additionally, the output load is soldered on the edge of output port and two via holes are connected to the ground. The Harmonic Balance simulation has been used in ADS to simulate the rectifying circuit performance. The HSMS 2852 diode package model is imported from the ADS components library. Moreover, the power conversion efficiency η_r for the rectifying circuit is defined as:

$$\eta_r = \frac{P_{out}}{P_{in}} = \frac{V_{dc}^2 / R}{P_{in}} \tag{2}$$

where P_{in} is the input power on the input port, V_{dc} is the output DC voltage on the output load R.

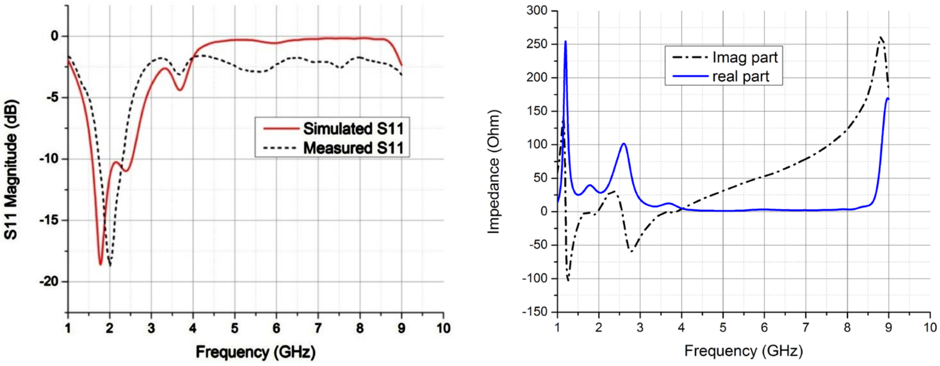

(a) (b)

Figure 5. (a) The measured and simulated frequency response of the AMLPF; (b) The simulated input impedance of the AMLPF.

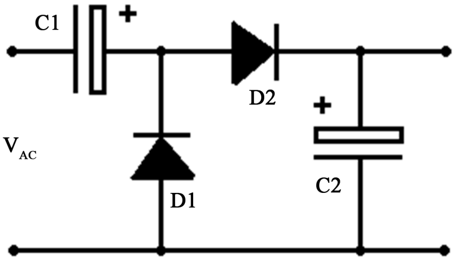

Figure 6. The equivalent circuit of a voltage doubling circuit.

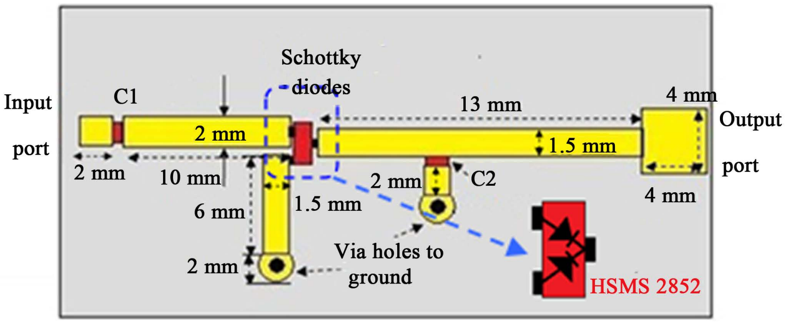

Figure 7. The layout of the voltage doubling rectifying circuit.

According to the IEEE standard for safety levels in general public area, the maximum power density in public area is 0.2 mW/cm^2 over the frequency range from 100 MHz to 3 GHz. This power density is low and of great interest to this study. Thus, the power levels from −5 to 10 dBm are the power range of interest, which can be obtained using the conversion equation below:

$$P_{in} = S_d A_{eff} \tag{3}$$

where the S_d is the power density and A_{eff} is the effective area of the proposed AMLP.

Because of the characteristics of the diode, the input impedance of the doubling circuit is affected by the frequency, the input power level and the output load. The voltage doubling rectifying circuit is optimized by using ADS to achieve the best performance over the power range from −5 to 10 dBm. For our investigation we first fix the input power at 2 dBm, the calculated conversion efficiency as a function of output load is shown in **Figure 8(a)** for three cases: 1.7, 2.0 and 2.4 GHz, respectively. These three frequencies are important in real applications thus selected in our investigation. For the source impedance, the value of the AMLPF impedance is used in the simulation.

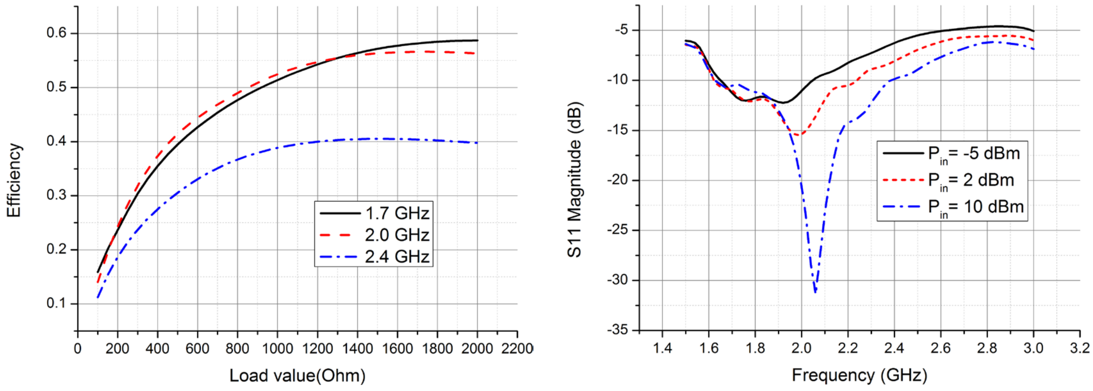

(a) (b)

Figure 8. (a) The conversion efficiency for the rectifying circuit as a function of the output load resistance; (b) The simulated reflection coefficient of the voltage doubling circuit for various input power.

It should be noted that the conversion efficiency reaches the maximum value of 40% when the load is around 1400 Ω at 2.4 GHz. When the output load is around 1800 Ω the conversion efficiency has a peak value of 57% at 2.0 GHz. Moreover, at 1.7 GHz the conversion efficiency gradually increases with the load value. The peak efficiency is a function of the load as well as the frequency. **Figure 8(b)** gives the reflection coefficient of the proposed circuit with the AMLPF impedance. It is clear that the impedance band for S11 < −10 dB for the doubling rectifying circuit is from 1.63 to 2.24 GHz for the input power of 2 dBm. Hence, the conversion efficiency at 2.4 GHz is lower than the conversion efficiency at 1.7 and 2 GHz due to the impedance mismatch. Furthermore, it is clear that the proposed rectifier can roughly meet our required bandwidth over the power range of −5 to 10 dBm. It is also shown that the lower the input power the narrower the bandwidth which is one of the challenges to develop a broadband rectenna.

When we fixed the load value at 1100 Ω, the conversion efficiency including the voltage doubling rectifying circuit as a function of the input power for various frequencies is plotted in **Figure 9**. It is obvious that efficiencies increase with input power and reach the peak with the input power around 8.5 dBm due to the characteristic of the rectifying diodes [12]. In addition, a part of the incident power is lost due to the source to the doubling rectifier impedance mismatching and the RF to DC rectifier conversion. Thus, choosing suitable antenna impedance will have a great impact on the rectifier power conversion performance.

3. The Rectenna Implementation and Measurements

The rectenna is formed using the proposed wideband dipole antenna with the microwave low-pass filter and the voltage doubling rectifying circuit as shown in **Figure 10(a)**. Although, in practice, the rectenna is used to harvest the RF power radiated by ambient sources with power levels fluctuating across the target frequency range, a laboratory measurement system shown in **Figure 10(b)** is selected to characterize the rectenna over the frequency range of interest, thus we have full control of the measurement for the frequency, incident power level and wave polarization. The source for the measurement was an Agilent signal generator connected a 30 dB power amplifier; thus far field measurements could be performed for a range of incident power levels using the integrated rectenna at different incidence. A broadband double ridged horn antenna was used for transmitting power from 1.6 to 2.8 GHz. Furthermore, the proposed rectenna was placed in a fixed position and the output voltage V_{dc} over the load R = 1100 Ω was measured using an oscilloscope. The rectenna was placed at a distance of 0.6 m from the horn to the rectenna under test. The total conversion efficiency of the rectenna can be obtained using

$$\eta = \frac{P_{out}}{P_{in}} = \frac{V_{dc}^2 / R}{S_d A_{eff}} \tag{4}$$

The effective area of the antenna A_{eff} can be calculated using

$$A_{eff} = \frac{G_r \lambda_0^2}{4\pi} \tag{5}$$

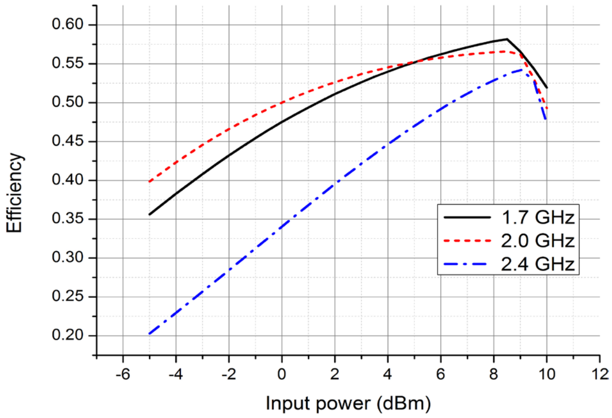

Figure 9. The conversion efficiency of the rectifying circuit as a function of output load with various input power at 2.4 GHz.

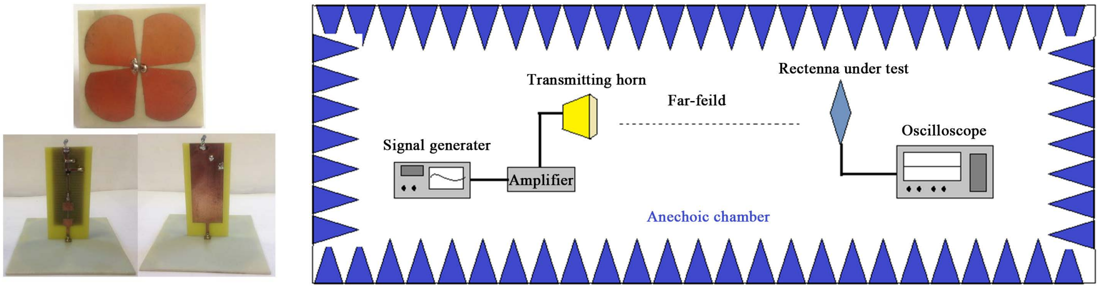

(a) (b)

Figure 10. (a) Photographs of the proposed rectenna under test. (b) The measurement system in anechoic chamber.

where S_d is the power density, G_r is the gain of the proposed antenna and λ_0 is the wavelength for the operation frequency. Before the measurement, the measurement system has been calibrated to remove the attenuation caused by cables and connections.

Firstly, the polarization of the rectenna is investigated. **Figure 11** shows the measured conversion efficiency of the rectenna as a function of the frequency for three polarization angles. It can be seen that a similar conversion efficiency is obtained for two orthogonal positions of the antenna with the polarization angles $\beta = 45^\circ$ and $\beta = 135^\circ$. The differences are most likely resulted from the fabrication error and asymmetric feeding structure. It is confirmed that the proposed rectenna is a dual-polarized rectenna. In addition, the rectenna has highest conversion efficiency over the frequency range with the polarization angle $\beta = 90^\circ$.

Secondly, the rectenna is studied for the polarization angle $\beta = 90^\circ$. Measured and simulated results of the integrated dipole rectenna are presented in **Figure 12**. It is clear that the rectenna has the conversion efficiency from 28% to 57% with incident power densities varying from 5 μW/cm^2 to 200 μW/cm^2 (around −5 dBm to 10 dBm in terms of the power) at 1.7 GHz. In the same power density range, conversion efficiencies change from 16% to 40% and 5% to 24% at 2.0 and 2.4 GHz, respectively. There are good agreements between the simulated and measured results. The measured efficiency was slightly lower than the simulated due to the inaccuracy of the fabrication, the connector and the material (FR4) losses on the rectifier at higher frequencies. Moreover, since the measured efficiency follows closely with the simulated efficiency, higher order harmonics re-radiation at those frequencies is quite low which is expected from the reflection coefficient plot of the antenna integrated with the low-pass filter in **Figure 5**.

The measured conversion efficiency and output voltage of the rectenna for different input power densities and frequencies are depicted in **Figure 13(a)** and **Figure 13(b)**. It is seen that the maximum conversion efficiency of 57% is achieved with the input power density of 120 μW/cm^2 at 1.7 GHz. From the power conversion point of view, due to relatively high power conversion efficiency (>20%) the optimum operational frequency range of the proposed rectenna is found to be 1.6 to 2.4 GHz at a power density of 200 μW/cm^2. Indeed, the rectenna op-

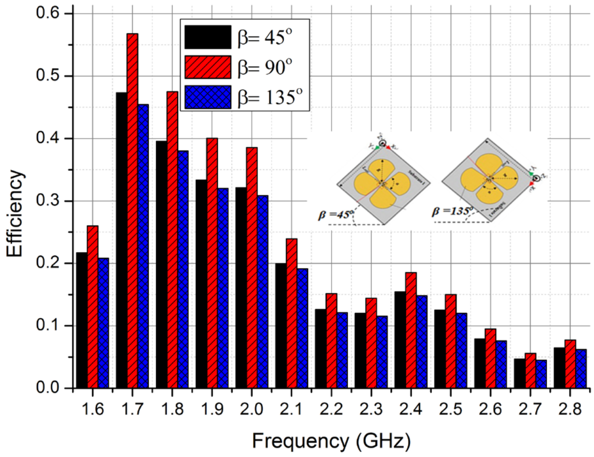

Figure 11. The Measured conversion efficiency for the proposed rectenna for the power density of 120 μW/cm^2.

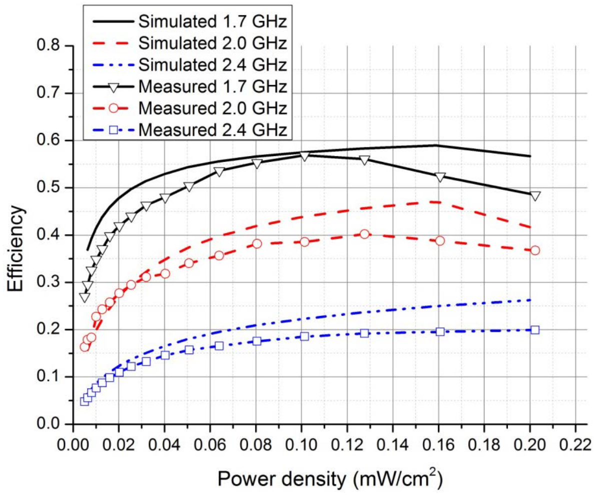

Figure 12. Measured conversion efficiency as a function of power density at 1.7, 2.0 and 2.4 GHz, respectively.

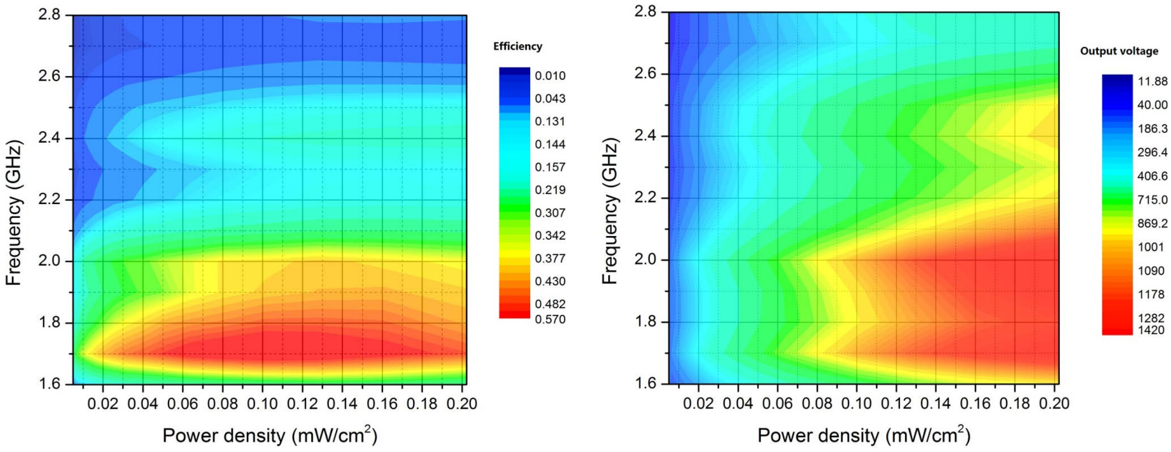

(a) (b)

Figure 13. (a) Measured conversion efficiency for wideband rectenna for different power densities with R = 1100 Ω; (b) Measured output voltage (in mV) for wideband rectenna at broadside for power densities from 5 to 200 μW/cm^2.

erational bandwidth approximately matches with the bandwidth of the doubling rectifying circuit and the bandwidth of the AMLPF. Furthermore, it can noticed that the conversion efficiencies shapely decrease over a higher frequency range (2.5 GHz - 2.8 GHz), which mainly result from the mismatch of the antenna and the rectifier and the performance of the AMLPF. **Figure 13(b)** shows the measured DC output voltage of the rectenna as a function of the frequency and power density. Over the frequency band of interest, the output voltage stays above 400 mV over the range of power densities of 30 to 200 μW/cm^2.

In addition, a measurement was taken in the University of Liverpool for ambient wireless energy harvesting. The rectenna was tested in an outdoor environment which is so empty that no buildings around shield the visual field. Power sources were a cell phone base station and Wi-Fi systems. The distance between the rectenna and the base station was around 500 m. The output voltage was measured by using a DC meter. The measured output voltage of the rectenna was around 100 mV most of the time and it could also reach more than 120 mV sometimes due to the variation of ambient RF energy. It confirmed that, in practice, the proposed rectenna also can indeed convert the wireless power generated by various systems

4. Conclusion

A dual-polarized wideband rectenna with a good conversion performance has been demonstrated through both simulation and experiment in this paper. The wideband antenna has been designed and optimized for energy harvesting. In addition, an integrated AMLPF has been demonstrated to block higher order harmonics. For maximum power transfer, the voltage doubling rectifying circuit has been designed to match with the AMLPF over power levels of interest. The study has confirmed that the impedance matching between the AMLPF and the rectifying circuit can effectively improve the rectenna conversion efficiency. Moreover, the proposed rectenna has achieved maximum conversion efficiency around 57% at 1.7 GHz and overall conversion efficiency over 20% has been achieved over the frequency range from 1.6 to 2.5 GHz. The new rectenna not only covers a wide operation bandwidth but also has a good performance for the low input power density (<200 μW/cm^2). The performances are confirmed through both simulated and measured results which demonstrate that the proposed rectenna has potential to harvest the wireless energy over a wide bandwidth.

References

[1] Monti, G., Corchia, L. and Tarricone, L. (2013) UHF Wearable Rectenna on Textile Materials. *IEEE Transactions on Antennas and Propagation*, **61**, 3869-3873. http://dx.doi.org/10.1109/TAP.2013.2254693

[2] Georgiadis, A., Collado, A., Via, S. and Meneses, C. (2011) Flexible Hybrid Solar/EM Energy Harvester for Autonomous Sensors. *Microwave Symposium Digest (MTT)*, *IEEE MTT-S International*, Baltimore, 5-10 June 2011, 1-4.

[3] Chin, C.H.K., Xue, Q. and Chan, C.H. (2005) Design of a 5.8-GHz Rectenna Incorporating a New Patch Antenna. *IEEE Antennas and Wireless Propagation Letters*, **4**, 175-178. http://dx.doi.org/10.1109/LAWP.2005.846434

[4] Sun, H., Guo, Y., He, M. and Zhong, Z. (2012) Design of a High-Efficiency 2.45-GHz Rectenna for Low-Input-Power Energy Harvesting. *IEEE Antennas and Wireless Propagation Letters*, **11**, 929-932. http://dx.doi.org/10.1109/LAWP.2012.2212232

[5] Young-Ho, S. and Kai, C. (2002) A Novel Dual Frequency Rectenna for High Efficiency Wireless Power Transmission at 2.45 and 5.8 GHz. *IEEE Microwave Symposium Digest*, **2**, 1297-1300.

[6] Brown, W.C. and Triner, J.F. (1982) Experimental Thin-Film, Etched-Circuit Rectenna. *IEEE Microwave Symposium Digest*, Dallas, 15-17 June 1982, 185-187.

[7] Falkenstein, E., Roberg, M. and Popovic, Z. (2012) Low-Power Wireless Power Delivery. *IEEE Transactions on Microwave Theory and Techniques*, **60**, 2277-2286. http://dx.doi.org/10.1109/TMTT.2012.2193594

[8] Hagerty, J.A., Helmbrecht, F.B., McCalpin, W.H., Zane, R. and Popovic, Z.B. (2004) Recycling Ambient Microwave Energy with Broad-Band Rectenna Arrays. *IEEE Transactions on Microwave Theory and Techniques*, **52**, 1014-1024. http://dx.doi.org/10.1109/TMTT.2004.823585

[9] Monti, G., Tarricone, L. and Spartano, M. (2011) X-Band Planar Rectenna. *IEEE Antennas and Wireless Propagation Letters*, **10**, 1116-1119. http://dx.doi.org/10.1109/LAWP.2011.2171029

[10] Pinuela, M., Mitcheson, P.D. and Lucyszyn, S. (2013) Ambient RF Energy Harvesting in Urban and Semi-Urban Environments. *IEEE Transactions on Microwave Theory and Techniques*, **61**, 2715-2726. http://dx.doi.org/10.1109/TMTT.2013.2262687

[11] Wheeler, H.A. (1964) Transmission-Line Properties of Parallel Strips Separated by a Dielectric Sheet. *IEEE Transactions on Microwave Theory and Techniques*, **13**, 173-185.

[12] Yoo, T.W. and Kai, C. (1992) Theoretical and Experimental Development of 10 and 35 GHz Rectennas. *IEEE Transactions on Microwave Theory and Techniques*, **40**, 1259-1266. http://dx.doi.org/10.1109/22.141359

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.

Advances in Bioscience and Biotechnology

American Journal of Analytical Chemistry

American Journal of Plant Sciences
Special Issue on Photosynthesis
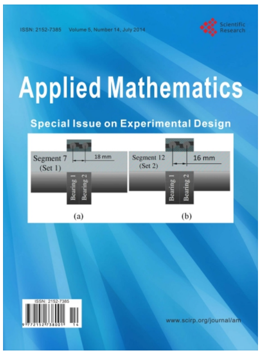
Applied Mathematics
Special Issue on Experimental Design

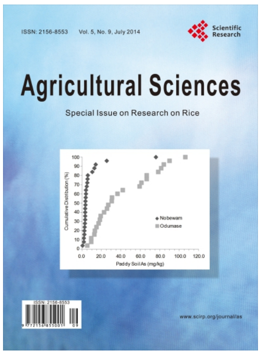
Agricultural Sciences
Special Issue on Research on Rice
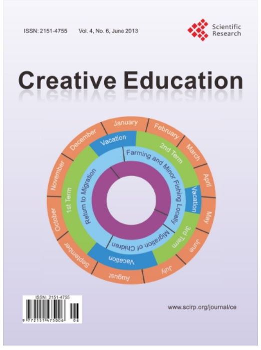
Creative Education
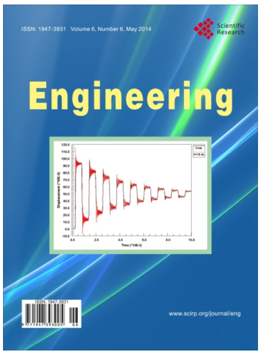
Engineering

Food and Nutrition Sciences

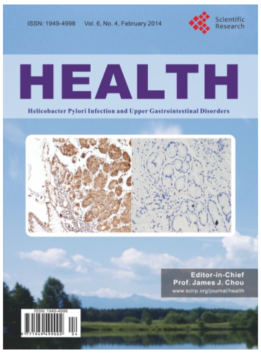
HEALTH
Helicobacter Pylori Infection and Upper Gastrointestinal Disorders
Editor-in-Chief
Prof. James J. Chou

Journal of Computer and Communications

Journal of Cancer Therapy
Special Issue on Bone Tumor Research
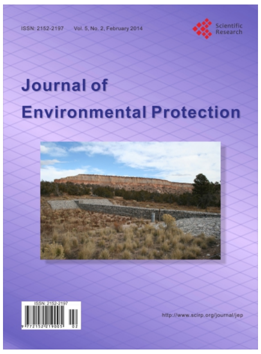
Journal of Environmental Protection

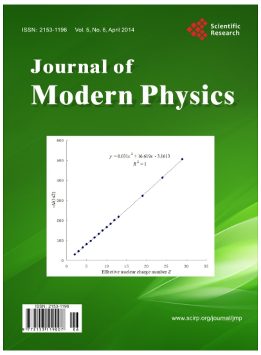
Journal of Modern Physics

Modern Economy
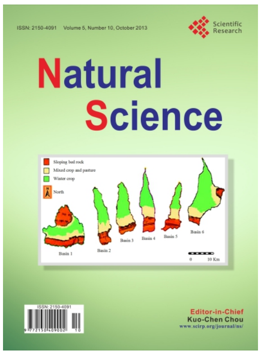
Natural Science
Editor-in-Chief
Kuo-Chen Chou
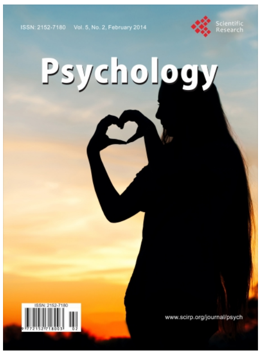
Psychology

PUBLISHED BY

INTECH
open science | open minds

World's largest Science,
Technology & Medicine
Open Access book publisher

3,250+
OPEN ACCESS BOOKS

106,000+
INTERNATIONAL
AUTHORS AND EDITORS

112+ MILLION
DOWNLOADS

BOOKS
DELIVERED TO
151 COUNTRIES

AUTHORS AMONG
TOP 1%
MOST CITED SCIENTIST

12.2%
AUTHORS AND EDITORS
FROM TOP 500 UNIVERSITIES

BOOK CITATION INDEX
WEB OF SCIENCE™

Selection of our books indexed in the
Book Citation Index in Web of Science™
Core Collection (BKCI)

Chapter from the book *Advances in Satellite Communications*

Downloaded from: http://www.intechopen.com/books/advances-in-satellite-communications

Interested in publishing with InTechOpen?
Contact us at book.department@intechopen.com

7

New Antenna Array Architectures for Satellite Communications

Miguel A. Salas Natera et al.*
Universidad Politécnica de Madrid,
Spain

1. Introduction

Ground stations which integrate the control segment of a satellite mission have as a common feature, the use of large reflector antennas for space communication. Apart from many advantages, large dishes pose a number of impairments regarding their mechanical complexity, low flexibility, and high operation and maintenance costs. hus, reflector antennas are expensive and require the installation of a complex mechanical system to track only one satellite at the same time reducing the efficiency of the segment (Torre et al., 2006).
With the increase of new satellite launches, as well as new satellites and constellation of low earth orbit (LEO), medium earth orbit (MEO), and geostationary earth orbit (GEO), the data download capacity will be saturated for some satellite communication systems and applications. Thus, the feasibility of other antenna technologies must be evaluated to improve the performance of traditional earth stations to serve as the gateway for satellite tracking, telemetry and command (TT&C) operation, payload and payload message or data routing (Tomasic et al., 2002). One alternative is the use of antenna arrays with smaller radiating elements combined with signal processing and beamforming (Godara, 1997).
Main advantages of antenna arrays over large reflectors are the higher flexibility, lower production and maintenance cost, modularity and a more efficient use of the spectrum. Moreover, multi-mission stations can be designed to track different satellites simultaneously by dividing the array in sub-arrays with simultaneous beamforming processes. However, some issues must be considered during the design and implementation of a ground station antenna array: first of all, the architecture (geometry, number of antenna elements) and the beamforming process (optimization criteria, algorithm) must be selected according to the specifications of the system: gain requirements, interference cancellation capabilities, reference signal, complexity, etc. During implementation, deviations will appear as compared to the design due to the manufacturing process: sensor location deviation and sensor gain and phase errors (Martínez & Salas, 2010). In an antenna array, the computation of a close approach of the direction of arrival (DoA) and the correct performance of the beamformer depends on the calibration procedure implemented.

---

* Andrés García-Aguilar, Jonathan Mora-Cuevas, José-Manuel Fernández González, Pablo Padilla de la Torre, Javier García-Gasco Trujillo, Ramón Martínez Rodríguez-Osorio, Manuel Sierra Pérez, Leandro de Haro Ariet and Manuel Sierra Castañer.
Universidad Politécnica de Madrid, Spain

This chapter is organized with the following sections. Section 2, introduces the relationship between applications and antenna design architectures. Section 3, introduces the new antenna array architectures for satellite communication including motivation and explains experimental examples. Section 4, explains adaptive antenna array and receiver architectures for adaptive antennas systems considering the beamforming with synchronization algorithms. Finally, Section 5 explains the A3TB concept.

2. Applications and antenna design architectures

In recent effort, new antenna array architectures have been under analysis and development. In (Tomasic et al., 2002) a highly effective, multi-function, low cost spherical phased array antenna design that provides hemispherical coverage is analyzed. This kind of novel architecture design, as the geodesic dome phased array antenna (GDPAA) presented in (Tomasic et al., 2002) preserves all the advantages of spherical phased array antennas while the fabrication is based on well-developed, easily manufacturable, and affordable planar array technology (Liu et al., 2006; Tomasic, 1998). This antenna architecture consists of a number of planar phased sub-arrays arranged in an icosahedral geodesic dome configuration.

In contrast to the about 10 *m* diameters dome of the GDPAA, there is the geodesic dome array (GEODA) (Sierra et al., 2007) with 5 *m* diameters dome. This antenna, presented in Fig. 1, has two geometrical structure parts. The first one, is based on a cylinder conformed by 30 triangular planar active arrays, and the second is a half dodecahedron geodesic dome conformed by 30 triangular planar active arrays. The GEODA is specified in a first version for satellite tracking at 1.7 GHz, including multi-mission and multi-beam scenarios (Martínez & Salas, 2010). Subsequently, the system of the GEODA has been upgraded also for transmission (Arias et al., 2010).

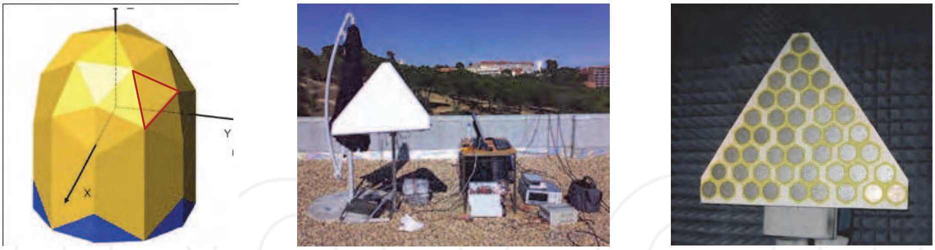

a b c

Fig. 1. a) The GEODA, b) The active sub-array demonstration, and c) The 45 elements planar active sub-array.

The antenna arrays technology in the user segment for satellite communications will substitute reflectors providing a more compact and easy to install antenna system, which is an interesting solution e.g. for satellite on the move (SOTM) system. There is a great diversity of solutions for fixed and mobile satellite communication systems including a large number of applications. Inmarsat broadband global area network (Inmarsat-BGAN) (Franchi et al., 2000) is the most representative example among mobile satellite systems (MSS), which gives land, maritime and aeronautical high speed voice and data services with global coverage using GEO satellites at L-band.

MSS services are divided into two groups, those that offer a regional coverage usually with GEO satellites, and those which offer a global coverage based on LEO or MEO satellite constellations. Depending on the coverage, there are some examples for MSS with regional coverage as the mobile satellite system (MSAT) in EEUU, Canada and South America, Optus in Australia, N-Star in Japan, Asia cellular satellite (ACeS) in Asia or Thuraya in the Middle East and in the North of Africa. While for MSS of global coverage there are some examples as Iridium, ICO Global Communications, Globalstar, Teledesic, etc. (Evans, 2009; Wu, 1994). Most of the MSSs work at L and S band, new applications on satellite to mobile terminal links work at X, Ku and Ka band, and satellite to base station connections work at L, S and C band. A number of applications is broad and lead terrestrial telecommunications market to offer a wider coverage: high speed voice and data (internet access, SMS, VoIP), digital video broadcasting by satellite 2 (DVB-S2) and digital video broadcasting satellite services to handhelds (DVB-SH), global position system (GPS) and Galileo, security, control and machinery monitoring on ships and aircrafts, teleeducation or telemedicine.

These modern satellite communications systems require new antenna solutions for base stations, aeronautical applications or personal communications services (PCS) on-the-move (Fujimoto & James, 2001). Within these applications, antenna array systems are potentially the best choice due to, as discussed above, its capability to perform electronically steering or beamforming, increase the antenna gain, and conform over curved or multifaceted surfaces the radiating elements. Portable antennas for PCS must be easy to install and mechanically robust, besides compact and lightweight (García et al., 2010) as the antenna array presented in Fig. 4.a. The design of antenna systems to provide high data rates for reliable PCS boarded on ships is not so strict in term of the geometrical requirements because it does not have space limitations (Geissler et al., 2010). However, in the case of land or airborne vehicles, geometrical and mechanical constraints are more severe. Antennas for terrestrial vehicles must be low profile, and for airborne vehicles aerodynamic shapes must be considered (Baggen et al., 2007; Vaccaro et al., 2010). Moreover, for the civil market conformal antenna arrays (Schippers, 2008; Kanno et al., 1996), or multi-surface arrays (Khalifa & Vaughan, 2007) are suitable choices to deal with the system aesthetic partiality.

Technological challenges have been faced during the implementation of satellite communication systems in the last decades. The design of a Test-Bed flexible and modular for testing or debugging beamforming algorithms and receiver architectures is an invaluable contribution in the educational, research and development area on satellite communication systems. The adaptive antenna array Test-Bed (A3TB) concept is based on the use of antenna arrays with beamforming capability to receive signals from LEO satellites (Salas et al., 2008). The scope of the A3TB is to probe the concept of antenna arrays applied to ground stations instead of reflectors for different applications, such as telemetry data downloading. It is also a good chance for Universities and Research Centers aiming to have their own ground station sited in their installations.

The A3TB ground station relies on the use of an antenna array to smartly combine the received signals from the satellite thanks to the implementation based on software defined radio (SDR) technology. The advantages of the SDR implementation is that A3TB architecture can be used to process any received signal from LEO satellites in the band imposed by the radio frequency (RF) circuits. Moreover, most of the processing is performed in software, so that appropriate routines can be used to process any received signal. The A3TB can be used to analyze the feasibility of different receivers and beamformer algorithms, regarding the capability to switch the receiver architecture in terms of the synchronizer algorithm configuration (Salas et al., 2007).

The current version has been developed to track The National Oceanic and Atmospheric Administration (NOAA) satellites in the very high frequency (VHF) band, in particular, the automated picture transmission (APT) channel (Salas et al., 2008). Previous versions of A3TB dealt with low rate picture transmission (LRPT) signals from the meteorological operational satellite-A (MetOp-A), where a complete receiver with beamforming and synchronization stages has been implemented (Salas et al., 2007; Martínez et al., 2007).

3. Antenna arrays for satellite communications

Satellite applications require compactness, lightweight and low cost antenna systems to be mounted on a terrestrial vehicle, an aircraft or a ship, or as a portable man-pack or a handset, and to be competitive against ground systems. Its major advantage is the possibility of getting a wider or even a global coverage. For such purposes, antenna arrays offer the technology to get a directive system whose steering direction can be electronically and/or mechanically controlled. However, planar arrays usually cannot steer more than 60^o-70^o from the normal direction of the antenna (Mailloux, 2005). Thus, when a wider angular coverage is required conformal arrays are an appropriate option (Josefsson & Persson, 2006). Arrays can approximate conformal shapes, such as spheres or cylinders, using several planar arrays, simplifying fabrication of active components (Sierra et al., 2007).

Since the low cost and low weight specifications are of importance, micro-strip antennas are mostly used, due to its capacity to be printed over a dielectric substrate with photolithography techniques. Low cost and low permittivity substrates are usually used such as FR4 or PTFE with different quantities of glass or ceramic impurities. For more demanding applications, ceramics, like alumina or high/low temperature co-fired ceramics (HTCC/LTTC) allow the use of smaller components thanks to its high permittivity, and give robustness against mechanical stresses and high temperatures.

3.1 Geodesic antenna array for satellite tracking in ground station

The aim of using a single antenna for tracking many satellites at the same time avoiding mechanical movements as well as its inexpensive cost make these antennas an alternative to be considered (Salas et al., 2008). Multi-beam ability and interference rejection are facilitated thanks to the electronic control system of such antennas that improves the versatility of the ground stations.

The GEODA is a conformal adaptive antenna array designed for MetOp satellite communications with specifications shown in Table 1. This antenna was conceived to receive signals in single circular polarization (Montesinos et al., 2009). Subsequently, in recent efforts the system has been upgraded also for transmission and double circular polarization (Arias et al., 2010). Hence, operating at 1.7 GHz with double circular polarization it can communicate with several LEO satellites at once in Downlink and Uplink. Current structure is the result of a comprehensive study that valued the ability to cover a given spatial range considering conformal shape surface and a given beamwidth (Montesinos et al., 2009). As Fig. 1 shows, GEODA structure consists of a hemispherical dome placed on a cylinder of 1.5 meters height. Both cylinder and dome are conformed by 30 similar triangular planar arrays (panels). Each panel consists of 15 sub-arrays of 3 elements (cells). The radiating element consists of 2 stacked circular patches with their own

RF circuits. The principal patch is fed in quadrature in 2 points separated 90º in order to obtain circular polarization. The upper coupled patch is used in the aim of improving the bandwidth.

Each panel is able to work itself as an antenna since they have a complete receiver that drives the 1.7 GHz signal to an analog to digital converter (ADC). In order to adapt the signal power to the ADC, it is mandatory to implement a complete intermediate frequency (IF) receiver consisting of heterodyne receiver with an automatic gain control block. Hence, each triangular array has active pointing direction control and leads the signal to a digital receiver through an RF conversion and filtering process. To follow the signal from the satellite, the main beam direction has to be able to sweep an angle of 60º. In this way, it is needed a phase shift in the feeding currents of the single radiating element. Previous calculations have demonstrated that 6 steps of 60 degrees are needed to achieve the required sweeping angle. An adaptive digital system allows the adequate signal combination from several triangular antennas. The control system is explained in (Salas et al., 2010).

Parameter	Specification	Parameter	Specification
Frequency range [GHz] **Tx:** **Rx:**	1.65 to 1.75 1.65 to 1.75	**Isolation between Tx and Rx [dB]**	>20
Polarization	Dual circular for Tx and Rx bands	**VSWR**	1.2:1
G/T [dB/K] **For elevation >30º** **For elevation 5º**	3 6	**SLL [dB]**	-11
EIRP [dBW]	36	**Size [m]**	1.5x1.5x3
3dB beamwidth [deg.]	5	**Accuracy steering [deg.]**	±1.4
Maximum gain [dBi]	29	**Coverage [deg.]:** **Azimuth** **Elevation**	360º >5º
Efficiency [%]	50		

Table 1. Main specifications for GEODA antenna.

3.1.1 Cell radiation pattern

Based on the study presented in (Sierra et al., 2007), the single radiating element is a double stacked circular patch that works at 1.7 GHz with 100 MHz bandwidth. In order to obtain circular polarization, the lower patch, which has 90 mm diameter, is fed by 2 coaxial cables in quadrature. Both coaxial cables connect the patch with a hybrid coupler to transmit and receive signals with both, right and left, circular polarizations. The upper patch is a circular plate with 78.8 mm diameter, and it is coupled to the lower patch increasing the bandwidth by overlapping both resonant frequencies tuning the substrate thickness and the patch diameter size. Fig. 2.a shows the radiating element scheme and main features of the layer structure are specified in (Montesinos et al., 2009).

A cell sub-array of 3 radiating elements shown in Fig. 2.b is considered the basic module to build the planar triangular arrays. The whole cell fulfills radiation requirements since it has a good polar to crosspolar ratio and a very low axial ratio. Likewise, as it is presented in Fig. 2.c, the radiation pattern shows symmetry and low side lobes for full azimuth.

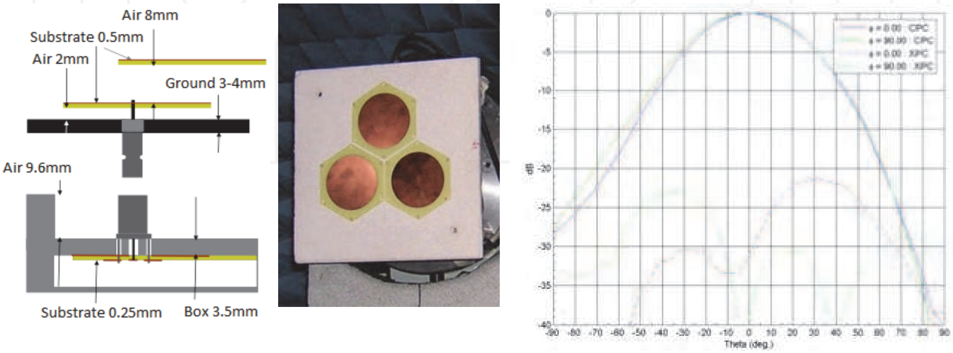

a b c

Fig. 2. a) Assembly of the single radiating element, b) Cell scheme, and c) Cell radiation pattern.

3.1.2 Transmission and Reception (T/R) module and cell distribution

Different T/R module configurations have been considered, providing either single or double polarization (Arias et al., 2010). T/R module allows amplifying and controlling the phase shift between signals, received and transmitted, providing an adaptive beam and steering direction controller in the whole working pointing range. As Fig. 3 shows, the design implemented contains a hybrid coupler, enabling double circular polarization; a double pole double throw (DPDT) switch, selecting polarization associated with transmission and reception way; 2 low noise amplifiers (LNAs), which amplify the signal received or transmitted; a single pole double throw (SPDT) switch, choosing transmission or reception way; and phase shifters, introducing multiples of 22.5º relative shift phases to form the desired beam. These surface mount devices have been chosen in order to reduce space and simplify the design.

Signals transmitted/received by the 3 T/R modules placed in a cell are divided/combined thanks to a divider/combiner circuit composed of 3 hybrid couplers that leads the signal to a general T/R module where signal is amplified. Due to transmission and reception duality, 2 SPDT switches are used to select the amplification way. Furthermore, each T/R module has associated a -25dB directional coupler that is used to test T/R modules in the transmission mode. Additionally, reception mode is tested by measuring signal in the divider/combiner circuit. A single pole 6 throw (SP6T) switch selects the path that is tested.

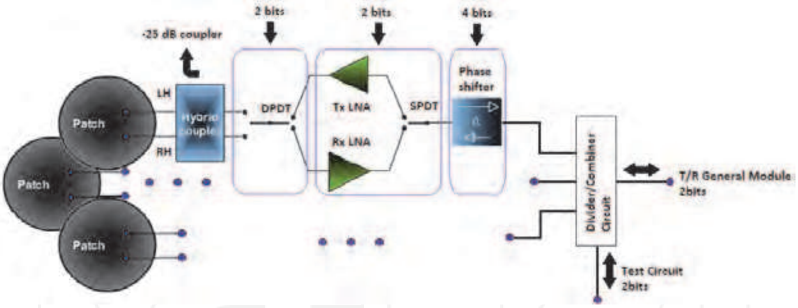
2 bits
2 bits
4 bits
-25 dB coupler
LH
RH
Hybrid coupler
DPDT
Tx LNA
Rx LNA
SPDT
Phase shifter
Patch
Patch
Patch
Divider/Combiner Circuit
T/R General Module 2bits
Test Circuit 2bits

Fig. 3. Cell sub-array and RF circuit.

3.1.3 Control system

The control system has two main parts (Salas et al., 2010), the hardware structure and the control software. The two level hardware structure has the lowest possible number of elements, making the control simpler in contrast to the previous in (Salas et al., 2010). Finally, an inter-integrated circuit (I2C) expander is used to govern T/R modules individually, and one more cover cell needs (LNA of call and test). A multipoint serial standard RS-485 is used to connect the computer with the panels.

3.2 Portable antenna for personal satellite services

New fix and mobile satellite systems (Evans, 2000) require antenna systems which can be portable, low profile and low weight. Planar antennas are perfect candidates to fulfill these specifications. Usually slots (Sierra-Castañer et al., 2005) and printed elements (García et al., 2010) are most used as radiating elements.

3.2.1 Antenna system structure

In this subsection it is introduced a printed antenna for personal satellite communications at X band, in Fig. 4. Table 2 shows main antenna characteristics.

Parameter	Specification	Parameter	Specification
Frequency range[GHz] **Tx:** **Rx:**	7.9 to 8.4 7.25 to 7.75	**Efficiency [%]**	50
Polarization	Dual circular polarization for Tx and Rx bands	**Isolation between Tx and Rx [dB]**	>17
G/T [dB/K]	7	**VSWR**	1.4:1
EIRP [dBW]	32	**SLL [dB]**	-11
3dB beamwidth [deg.]	5	**Size [m]**	40x40x2.5
Maximum gain [dBi]	25	**Weight [Kg]**	2

Table 2. Portable antenna specifications.

This is a planar, compact, modular, low loss and dual circular polarized antenna, for Tx and Rx bands, simultaneously. It is made up by a square planar array of 16x16 double stacked micro-strip patches, fed by two coaxial probes. A hybrid circuit allows the dual circular polarization (Garg et al., 2001). Elements are divided in 16 sub-arrays excited by a global power distribution network of very low losses, minimizing the losses due to the feeding network and maximizing the antenna efficiency. In order to reduce side lobe levels (SLL), the signal distribution decreases from the centre to the antenna edges, keeping symmetry with respect to the main antenna axes. The antenna works at X band from 7.25 up to 8.4 GHz with a 14.7% relative bandwidth for a 1.4:1 VSWR and a maximum gain of 25 dBi.

3.2.2 Sub-array configuration

The sub-array configuration can be seen in Fig. 4.a. It makes possible to separate the fabrication of these sub-arrays from the global distribution network, simplifying the corporative network and getting a modular structure suitable for a serial fabrication process. Each sub-array is a unique multilayer board, where PTFE-Glass substrate of very low losses has been used as base material. The power distribution network is connected to each sub-array through (SMP-type) coaxial connectors.

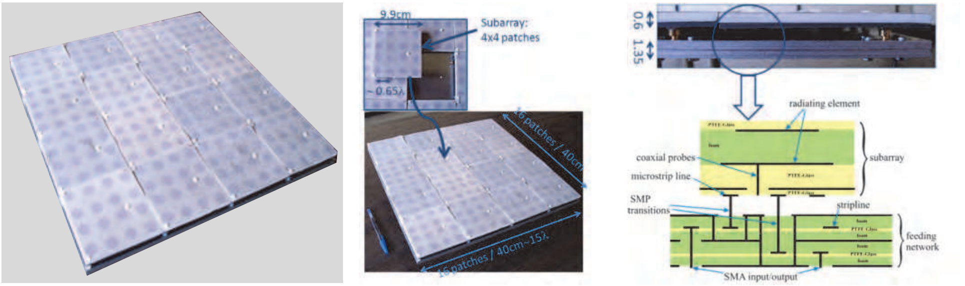

a b c

Fig. 4. a) Dual polarized portable printed antenna for satellite communication at X band, b) Sub-array perspective view, and c) Side view and multilayer scheme.

Fig. 5.a and Fig. 5.b show the sub-array unit cell. In order to obtain better polarization purity, each element is rotated 90° and excited by a 90° phase-shifted signal. Moreover, in Fig. 5.c is showed a miniaturized branch-line coupler (BLC) of three branches working as a wide band hybrid circuit (García et al., 2010; Tang & Chen, 2007).

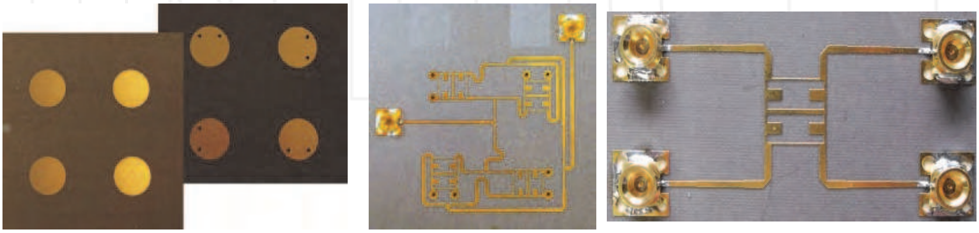

a b c

Fig. 5. Unit cell test board, a) Unit cell test board 2x2 stacked patches, b) Micro-strip feeding network, and c) Miniaturized BLC Prototype.

A conventional configuration takes up an area of 13.3 cm^2 which is big compared to the radiating element and the sub-array subsystem size. Therefore, a miniaturization of the BLC is needed using the equivalence between a λ/4 transmission line and a line with an open-ended shunt stub. An area reduction about 35% is achieved and the hybrid circuit behaves like a conventional BLC. In Fig. 6.b and Fig. 6.c measurement results for the BLC in Fig. 5.c are shown compared with simulations.

Fig. 7 depicts some sub-array measurements. The copular to crosspolar ratio is better than 25 dB and axial ratio is under 0.9 dB in the whole bandwidth.

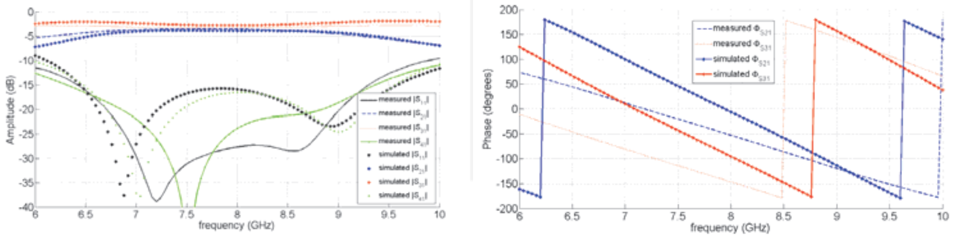

a b

Fig. 6. Miniaturized BLC, Measured and simulated S-parameters in: a) Amplitude, and b) Phase.

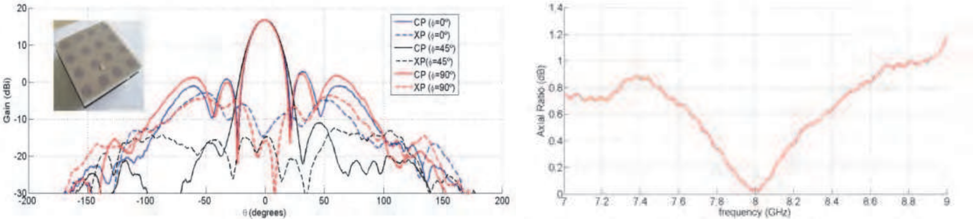

a b

Fig. 7. 4x4 patch sub-array measurements, a) Radiation pattern at 7.75 GHz, and c) Axial ratio for right-handed circular polarization.

3.2.3 Low losses power distribution network

The global feeding network presented in Fig. 8.a is a protected strip-line, where foam sheets of high thickness are used to get low losses. Such a kind of feeding network allows keeping a trade-off between the simplicity of exciting the radiating elements using printed circuits and the loss reduction when the distribution network is separated in a designed structure to have low losses. Losses in the structure are around 0.6 dB/m which yields to 0.3 dB of losses in the line. Two global inputs/outputs using SMA-type connectors, one for each polarization, excite the strip-line networks.

Vertical transitions have to be treated carefully and must be protected to avoid undesired higher order mode excitation. Thereby, it has been design a short-ended pseudo-waveguide, adding some extra losses about 0.3 dB, for two kinds of vertical transitions, as can be seen in Fig. 8.b and Fig. 8.c.

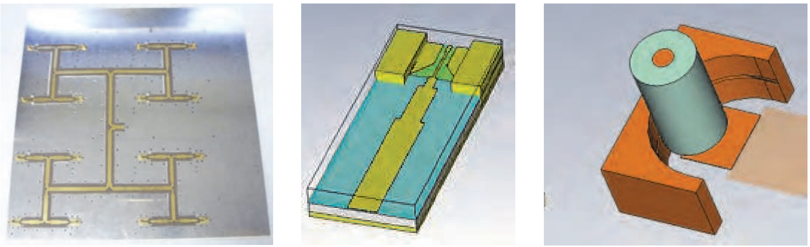

a b c

Fig. 8. a) Protected strip-line global corporative network for one polarization, b) Transitions from strip-line to SMA-type connector, and c) Transitions from strip-line to SMP-type connector.

3.2.4 Antenna performance

Fig. 9 depicts measured radiation pattern at 7.75 GHz, gain and axial ratio for the antenna system. It is shown a maximum gain of 25 dBi in the lower band and about 22 dBi in the upper band, and a SLL around 11 dB. Copolar to crosspolar ratio is better than 30 dB and axial ratio is under 0.7 dB. Total losses are about 4 dB in the working band.

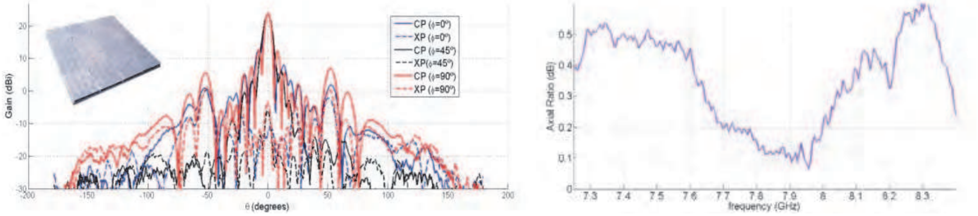

a b

Fig. 9. Antenna measurements results, a) Radiation pattern at 7.75 GHz, and c) Axial ratio for right-handed circular polarization.

3.3 Electronically steerable antennas for mobile and fixed portable systems

At present, two types of electric steerable antenna systems can be used to access the satellite communication services (Bialkwoski et al., 1996). These are: fixed position portable systems and mobile systems such as those installed on a land vehicle. The fixed portable antenna system is relatively easy to be accomplished by the antenna designer. The design involves standard procedures that concern the operational bandwidth, polarization and moderate gain (García et al., 2010). One drawback of the fixed position portable system is that they require the user to be stationary with respect to the ground. This inconvenience can be overcome with the mobile antenna system. A mobile user complicates the scenario since the ground mobile antenna needs to track the satellite (Alonso et al., 1996). The design of such a system is more challenging as new features associated with the mobility of the system have to be incorporated (Fernández et al., 2009). The requirement leads to a narrow beamwidth, for which satellite tracking is required as the vehicle moves around. Electronically steerable antennas enable the development of reconfigurable antennas for satellite applications.

3.3.1 Steerable antenna for fixed position portable systems

This antenna is a fixed satellite communication system with high gain at X band, consisting of an antenna array that integrates 32 2x2 sub-array modules in the complete antenna, as shown in Fig. 10.a. It is a planar and dual circular polarized antenna for Tx and Rx bands simultaneously. It is made up by a planar array of double stacked circular micro-strip patches, fed by 2 coaxial probes to generate circular polarization. A hybrid circuit allows the dual circular polarization as shown in Fig. 10.b.

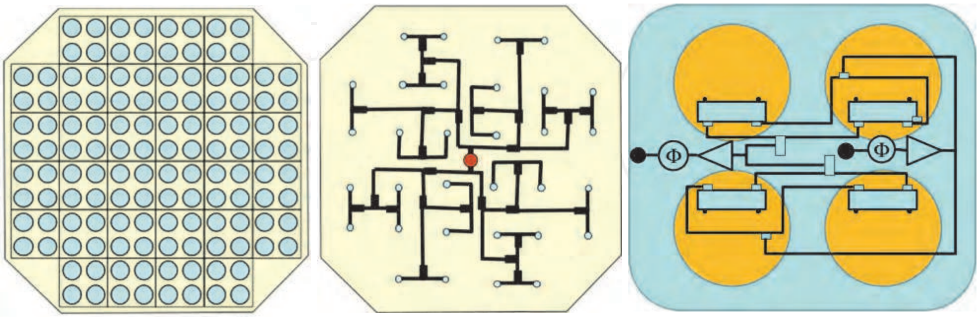

a b c

Fig. 10. Active multi-beam antenna, a) Top view, b) Feeding network of the complete antenna, and c) Beamforming network of the 2x2 sub-array module

The antenna has the same design parameters, structure and configuration as the antenna explained in Section 3.2 but with a different feeding network, as previously shown. In this case, the beamforming network requires changes in the feeding phase in the 2x2 sub-arrays, which can be achieved by phase shifters (ϕ) associated with different sub-arrays (Fig. 10.c). All these sub-arrays are connected to a feeding network, in Fig. 10.b, formed by transmission lines with low losses in strip-line. General specifications of the steerable antenna for fixed position portable systems are provided in Table 3.(a).

3.3.2 Automatic steerable antenna for mobile systems

A broadband circularly polarized antenna for satellite communication in X band is presented in Fig. 11 and specified in Table 3.(b). The arrangement features and compactness are required for highly integrated antenna arrays. It is desired to get a low-gain antenna for mobile satellite communications with low speed of transmission. In this system, the antennas are formed by 5 planar 4x4 arrays of antennas, which form a truncated pyramid with a pointing capability in a wide angular range, so that among the 5 planar arrays the complete antenna can cover any of the relative positions between the mobile system and the satellite in a practical way. The scheme of the active antenna can be seen in Fig. 11.

As it can be observed in Fig. 11.a, the antenna terminal is a multi-beam printed antenna shaped as a trunk pyramid capable of directing a main beam in the direction of the satellite. The antenna steering system consists of a multi-beam feeding structure with switches that lets combine the feed of each 4x4 arrays to form multiple beams. Switching the different 4x4 arrays, it is achieved different multiple beams and the variation of the steering direction.

The complete antenna consists of a Tx and Rx module that works independently in the 2 frequency bands.

The antenna has multiple beams covering the entire space to capture the satellite signal without moving the antenna. The signal detected in each of the beams is connected to a switch, which, by comparison, is chosen the most appropriate 4x4 array. The steering direction of the 4x4 array can vary between a range of directions that covers a cone angle range of 90º. To obtain the required gain and cover the indicated range, it is required around 15 beams, which can be obtained by integrating the beamforming networks with switches in the design as presented in (Fernández et al., 2009).

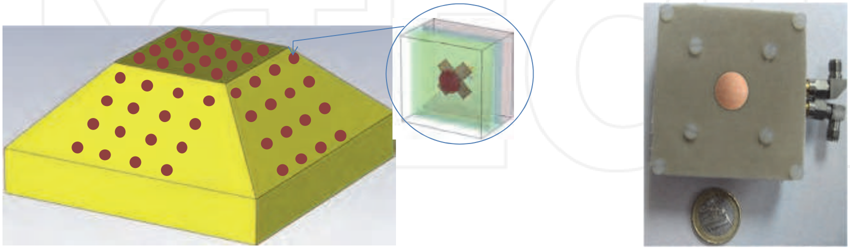
a b

Fig. 11. Complete antenna structure, a) Radiating element of the 4x4 arrays, and b) Prototype top view.

The radiating element of the 4x4 array is one 2 crossed dipoles with a stacked circular patch as shown in Fig. 11.a and Fig. 11.b. In Fig. 12 the cross-section of the radiating element structure is presented.

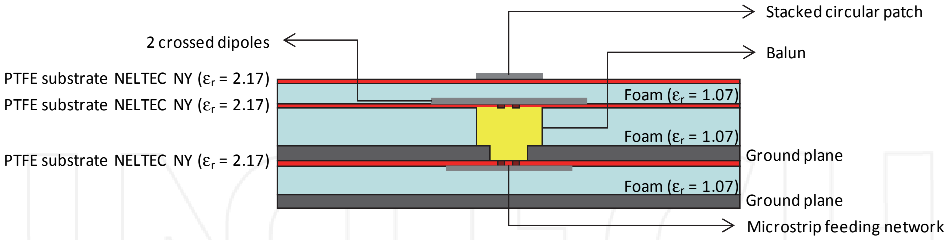
Stacked circular patch
2 crossed dipoles
Balun
PTFE substrate NELTEC NY (εr = 2.17)
PTFE substrate NELTEC NY (εr = 2.17)
Foam (εr = 1.07)
Foam (εr = 1.07)
PTFE substrate NELTEC NY (εr = 2.17)
Ground plane
Foam (εr = 1.07)
Ground plane
Microstrip feeding network

Fig. 12. Cross-section scheme of the radiating element.

The key element of the radiating element feeding structure (Fig. 14.b) is a resonant micro-strip feed ring that has been implemented, as well as a micro-strip 90º branch-line coupler to obtain the desired right hand or left hand circular polarizations (RHCP or LHCP) which ensures adequate port coupling isolation. The S-parameters in amplitude and phase of the micro-strip feeding structure are shown in Fig. 13.a and Fig. 13.b.

Fig. 14.a depicts the S-parameters of the radiating element with the micro-strip feed structure and they fulfill the specification, in Table 3.(b). In Fig. 14.c, the radiation pattern of the radiating element at 7.825 GHz is shown and in Fig. 14.d the radiation pattern of the 4x4 arrays is presented. It is shown a maximum gain of 19.4 dBi at the center frequency band (7.825 GHz). Copolar (CP) to crosspolar (XP) ratio is better than 17 dB and the axial ratio is under -3dB.

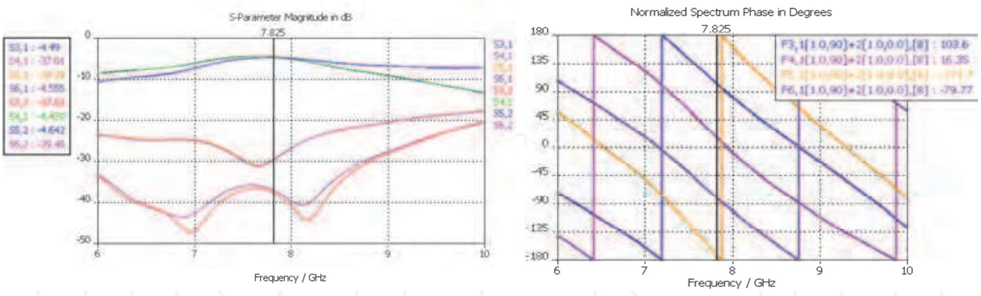

a b

Fig. 13. Micro-strip feeding structure, a) Amplitude of S-parameters, and b) Phase of S-parameters.

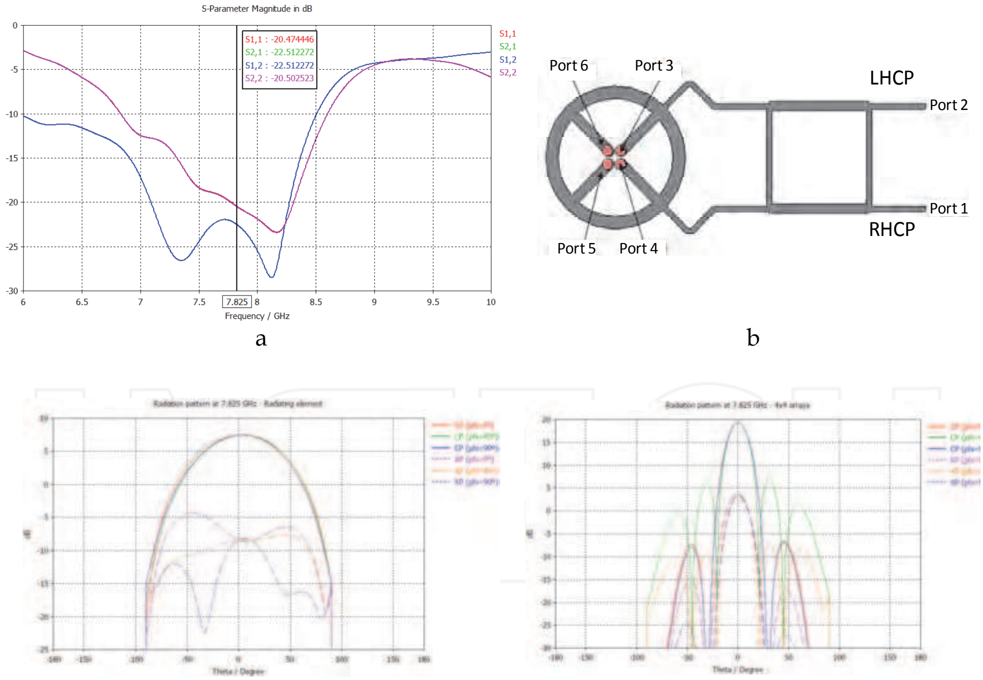

a b

c d

Fig. 14. a) S-parameters, b) Resonant ring + 90º branch-line coupler, c) radiation pattern at 7.825 GHz, and d) 4x4 array radiation pattern.

Parameter	Value (a)	Value (b)	Comments
Freq. range [GHz] **Rx** **Tx**	7.25 - 7.75 7.9 - 8.4	7.25 - 7.75 7.9 - 8.4	Microwave applications.
G/T (in Rx) [dB/K]	7	7	
EIRP (in Tx) [dBW]	32	32	
Beamwidth at -3dB [deg.]	4	20	
Polarization	circular	circular	In both, reception and transmission.
Gain [dBi]	>28	>15	
Axial ratio [dB]	< 1	<3	**(a)** Between ±50°. **(b)** Between ±45°.
VSWR	< 1.4:1 (-15.6 dB)	< 1.5:1 (-13.9 dB)	
Isolation between ports [dB]	< -17	< -15	
Radiation pattern [deg.]	±35	±90	Steering direction tilt.
Dimensions [cm]	40x40x4	20x20x15	

Table 3. (a) General specifications of the steerable antenna for fixed position portable systems , and (b) General features of the automatic steerable antenna for mobile systems.

3.4 Transmit-array-type lens antenna for terrestrial and on board receivers

Technology in satellite communications has revealed an increasing interest in novel smart antenna designs. Phased-array based designs are basic in electronically reconfigurable devices for satellite applications, which are more and more demanding. The strict requirements in terms of architecture, shape and robustness are important constraints for the development of planar lens-type devices. Regarding the usage and location, lens-type devices are useful for either terrestrial or on board receivers, in vehicular technology. Some clear examples are satellite communications for aircrafts preserving the fuselage aerodynamics or for some other kind of vehicles such as trains, etc.

3.4.1 Introduction to lens-type structures

In a general view, in lens-type a particular signal is received (in our case, an electromagnetic wave with specific features in terms of frequency, wave-front, etc.), it is processed (either complex signal processing techniques or only phase correction tasks can be considered in this interface), and finally, the processed signal is retransmitted.

Regarding the lens configuration, a transmit-array lens consists of three well distinguished interfaces: the first one for signal reception, one interface for signal processing, and the last one for processed signal re-radiation, as depicted in Fig. 15.

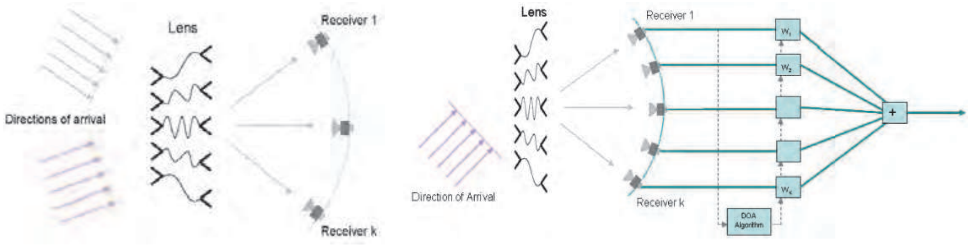

a b

Fig. 15. a) Multi-user scheme with different receivers and transmitters, and b) Adaptive scheme with DoA determination.

These structures are intimately related to reflect-array ones, where the reception and transmission interfaces are turned to be the same interface, with a reflection-type behavior (Encinar & Zornoza, 2001). Although in an equal output phase configuration a transmit-array device behavior would be similar to the one obtained with a reflect-array, the transmit-array offers the advantage of removing the feed blockage.

In a transmission scheme, depending on the transmitter position regarding the lens, a different steering direction is achieved and a different user is pointed. In the case of reception, the situation is the same: the user position configures the direction of arrival, which determines the receiver position around the lens (Padilla et al., 2010a). In adaptive schemes, applying the proper processing algorithm to the signal received in the different receivers around the lens, it is possible to develop an adaptive steering vector, in terms of the desired direction of arrival.

3.4.2 Transmit-array lens architecture and design

Lens-type structures provide two fundamental advantages. First, phase error correction due to spherical wave front coming from the feeding antenna. Fig. 16.a shows this effect. Second, new radiation patterns configuration. Fig. 16.b depicts this fact.

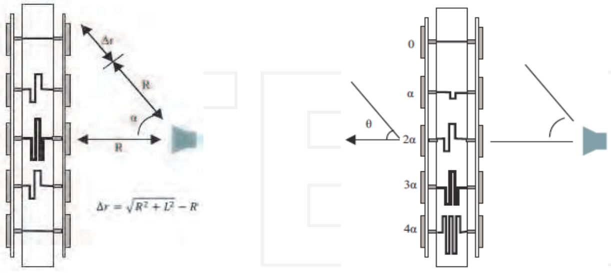

a b

Fig. 16. a) Phase error correction, and b) Radiation pattern reconfiguration.

3.4.3 Electronically reconfigurable devices for active transmit-array lenses

The addition of reconfigurability on transmit-array devices requires the possibility of controlling the phase response of the transmitted signal at each cell of the lens. Electronic control of phase signal may be added in two different ways: First, electronic tuning of the

radiating element phase response (Padilla et al., 2010a): Modifications in the radiating element circuital behavior lead to changes in phase response (arg[S_{21}]). Fig. 17 shows an electronically reconfigurable microwave patch antenna for this purpose, along with the equivalent circuit and prototype outcomes in terms of phase.
Second, electronic tuning of phase shifters in transmission lines (Padilla et al., 2010c): Modifications in the phase response of the phase shifters lead to corresponding changes in phase response. Some options are applied for these devices, such as hybrid couplers, etc. Fig. 18 shows a microwave phase shifter prototype for this purpose, along with the working scheme and its outcomes in phase.

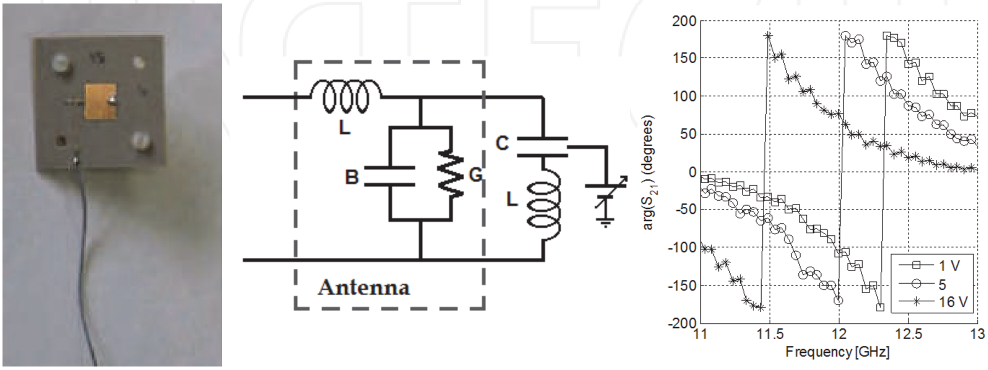

a b c

Fig. 17. Electronically reconfigurable antenna, a) Patch antenna prototypes, b) Equivalent circuit, and c) Phase behavior in frequency.

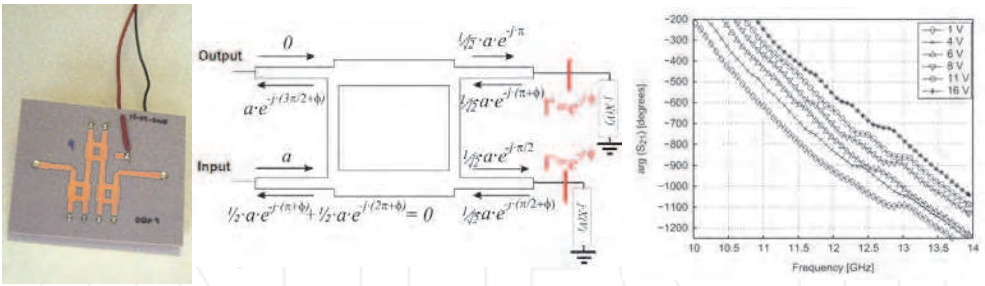

a b c

Fig. 18. Electronically reconfigurable phase shifter, a) Phase shifter prototype, b) Working scheme, and c) Phase behavior in frequency.

3.4.4 Electronically reconfigurable active transmit-array prototype

One electronically reconfigurable prototype is presented in Fig. 19 and detailed in this section. The prototype design implies the use of microwave phase shifters according to the design specified in section 3.4.3. This transmit-array lens prototype operates at 12 GHz. Main specifications are provided in Table 4.

Parameter	Value	Comments
Frequency range [GHz]	12 ± 0.5	Microwave applications.
Polarization	Linear	In both, reception and transmission.
Directivity [dBi]	>21	
Axial ratio [dB]	< 1	Between ±50° elevation.
S_{11} [dB]	< -20	
Radiation pattern [deg.]	±30	Steering direction tilt, for both H and V planes.
Feeding antenna [mm]	120	Corrugated horn linearly polarized
Phase shifters [deg.]	360	Full phase range variation.
Transmit-array elements	36	6x6 array topology.
Separation between elements	$0.7\lambda_0$	Related to the wavelength

Table 4. Main features of the electronically reconfigurable transmit-array prototype.

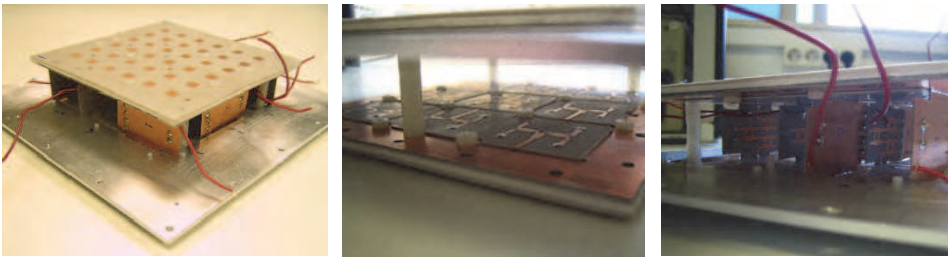

a b c

Fig. 19. Transmit-array core, a) Transmit-array prototype, b) Distribution networks, and c) Phase shifter integration.

The electronically controllable steering capabilities are tested and assured for a range of ± 30°in each main axis. An example of radiation pattern is provided in Fig. 20, for 9° tilt in one of the main axes.

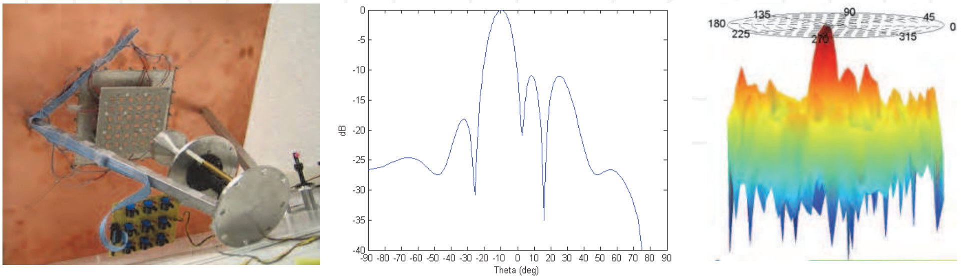

a b c

Fig. 20. a) Complete transmit-array with feeder and control circuits; and transmit-array measurement results for 9° tilt in one axis, b) H plane, and c) 3D plot.

4. Adaptive antenna array

Adaptive antennas can be described as systems usually based on three main parts: the antenna array, the receiver architecture and the beamforming scheme. Thus, adaptive antennas have those advantages owing to those three main parts. The system capabilities increase as complexity and development cost do. Furthermore, since signal processing is the basement of the adaptive antenna concept it is important to analyze the design challenges in terms of hardware architecture and components such as processors and embedded systems.

The antenna array provides the capability of performing the antenna pattern meeting the environment requirement under study. Besides, receiver architectures have some interesting advantages depending on the implemented receiver arraying technique such as signal to noise ratio (SNR) and bit error rate (BER) performance enhancement. Furthermore, symbol synchronization and carrier recovery can be used increasing the receiver complexity but providing higher performances. Finally, beamforming schemes use multiple antennas in order to maximize the strength of the signals being sent and received while eliminating, or at least reducing, interference as discussed in Section 4.3.

Adaptive antenna arrays are often called Smart Antennas because they have some key benefits over traditional antennas, by adjusting traffic patterns, space diversity or using multiple access techniques. The main four key benefits are: First, enhanced coverage through range extension by increasing the gain and steering capability of the ground station antenna; Second, enhanced signal quality through multi-target capability and reduction of interferences; finally, adaptive antennas improve the data download capacity in the ground segment of satellite communication by increasing the coverage range (Martínez et al., 2007).

4.1 Design and architecture based on software defined radio

For design there is the well known waterfall life cyclic model (Royce, 1970) that can be used to manage main aspects of the design of architectures. Thus, some tasks must be fulfilled subsequently as follow in Fig. 21.a.

Fig. 21.b shows the design schemes resulting of the requirement analysis stage corresponding software and hardware system specifications. In the depicted scheme, there are some system components such as the radiating element and RF circuits that are often designed under iterative prototyping model.

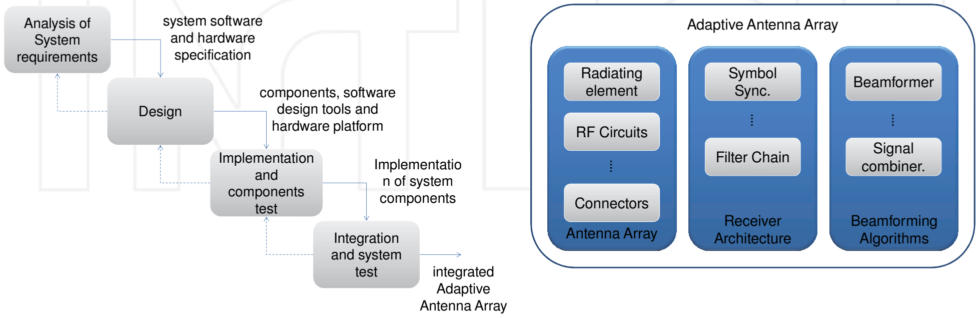
Analysis of System requirements
system software and hardware specification
Design
components, software design tools and hardware platform
Implementation and components test
Implementation of system components
Integration and system test
integrated Adaptive Antenna Array
Adaptive Antenna Array
Radiating element
RF Circuits
Connectors
Antenna Array
Symbol Sync.
Filter Chain
Receiver Architecture
Beamformer
Signal combiner.
Beamforming Algorithms

a b

Fig. 21. a) Water life cyclic model of the adaptive antenna array design, and b) Simplified design scheme of adaptive antenna arrays.

Regarding the hardware implementation, tables presented in (Martínez et al., 2007) show the hardware resource consumption in the field programmable gate array (FPGA) Virtex-4 for the least mean squared (LMS) beamforming algorithm with full spectrum combining (FSC) receiver architecture and SIMPLE beamforming algorithm with symbol combining (SC) receiver architecture. Both scheme designs have an antenna array of 2 elements. The algorithm based on correlation requires less hardware. The main difference can be appreciated in the amount of digital signal processing oriented component (DSP48) resources, typically used for filtering applications (Martínez et al., 2007).

4.2 Receiver architectures based on algorithms type

Several receiver architectures can be implemented, and they are frequently based on the type of the beamforming algorithm used. When training signals are available in the transmitted frame, a time-based reference algorithm can be used. However, this solution is only valid when the earth station is capable of demodulating the received training sequence. Other algorithms used in deep space communications are based on signal correlation and they avoid performing the demodulating process. This kind of algorithms are blind techniques that do not require any additional signal demodulation before applying some beamforming technique and work better in low SNR conditions than time-based algorithms. Several receiver architectures can be implemented exploiting the processing capabilities of the SDR, such as FPGA, application-specific integrated circuits (ASICS), and digital signal processing (DSPs). The design of the receiver architecture fundamentally depends on the selection of beamforming algorithms. An example of beamforming technique is the LMS algorithm whose estimation of coefficients or weights requires a temporal reference and is implemented through SC receiver architecture (Fig. 22.a). In the other hand, the SIMPLE algorithm (Rogstad, 1997) constitutes a beamforming technique that is implemented using FSC receiver architecture (Fig. 22.b) in order to perform the calculation of weights.

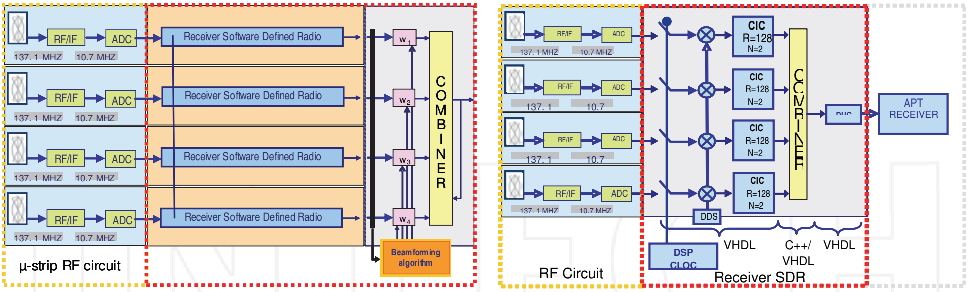

a b

Fig. 22. Comparison of receiver architectures. a) Symbol Combining (SC), and b) Full Spectrum Combining (FSC).

The SC architecture can be divided into two more sub-classes which work on a phase-recovery basis. The complex symbol combining (CSC) recovers the phase information with regard to a reference element using feed-forward and feedback algorithms. One of the advantages of this scheme is that the rate of data sent to the combining module has a rate slightly higher than the symbol rate. For most applications, the symbol rate is relatively low and is a multiple of the data rate. In this kind of schemes, there is an important cost consideration in real-time applications and the requirements of instrumental phase stability are very severe (Rogstad et al., 2003). Other type of SC architecture is the stream symbol combining (SSC). In this kind of scheme, data are sent to the combining module at a rate equal to the symbol rate. The symbol rate depends on the coding scheme and for most applications is relatively modest. Also, the requirements of instrumental phase stability are no severe, as in the case of CSC scheme. The disadvantage of the SSC is the additional hardware required for each antenna.

Furthermore, there are the baseband combining (BC) and carrier arraying (CA) architectures discussed in (Rogstad et al., 2003). In BC architectures the signal from each antenna is carrier locked and combining in baseband for further demodulation and synchronization. In effect, the carrier signal from the spacecraft is used as a phase reference so that locking to the carrier eliminates the radio-frequency phase differences between antennas imposed by the propagation medium. Besides, in CA architectures, one individual carrier-tracking loop is implemented on each array element. Then, the elements branches are coupled in order to increase the carrier-to-noise ratio (CNR), but losses of radio channel are far compensated (Rogstad et al., 2003).

In general, the selection of the beamforming algorithms is determined by the following aspects: Hardware and computational resources; Speed of convergence and residual error of adaptive algorithms; Calibration requirements and auto-compensation ability; and system signal-transmission characteristics.

4.3 Beamforming techniques for satellite tracking

Some satellites transmit useful information inside its frames for synchronization and tracking purposes. The gathering of satellite data requires the tracking operation along its earth orbit. To accomplish this goal with adaptive array architectures, some beamforming techniques should be implemented. Fig. 23 illustrates a simple example of a narrowband linear adaptive beamformer system.

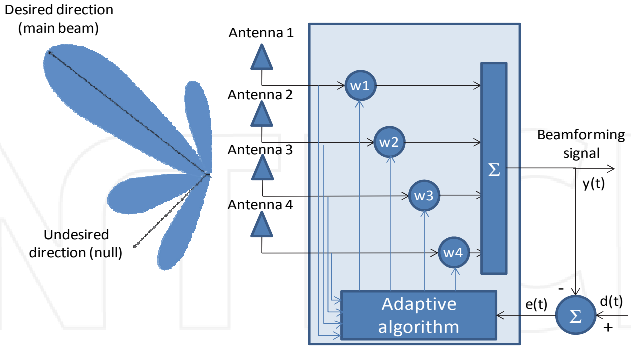

Fig. 23. Adaptive antenna system.

A linear beamformer combines signals according to some weights w_i, to produce a desired radiation pattern. The mathematical expression of a linear beamformer at the array output in vector notation can be expressed as $y = w^H x$, where x is the received signal vector to be combined, w are the weights computed by the beamforming algorithm and ^{H} denotes transposition and conjugate of $(\cdot)$.

In adaptive antennas design, weights are dynamically calculated with a certain algorithm in order to optimize some signal parameter like signal to interference-plus-noise ratio (SINR), SNR, or BER. An extended variety of algorithms exist in the literature for beamforming purpose and the most appropriated selection is done depending on the signal characteristics of the received signal.

4.3.1 Blind techniques

Blind beamformers make use of an inherent property of the received signal, such as the ciclo-stationarity of the constant modulus. In the latter, the algorithm eliminates the fluctuation of the signal amplitude and computes the weights to minimize the effect produced by those variations. The algorithms that make use of these methods are denoted as Constant Modulus Algorithms (CMA) (Biedka, 2001).

CMA algorithms present an important disadvantage: as the phase information is not considered, the constellation of quadrature phase shift keying (QPSK) signals commonly used in satellite communications appears rotated after beamforming, which imposes the need of an additional phase recovery subsystem in the array output.

4.3.2 Temporal-reference algorithms

Algorithms based on a temporal reference require a known reference included in the frame of the signal, such as training sequences, unique word (UW) or pilot bits. Thus, these schemes are normally used for digital signals. The aim of these beamformers is the minimization of the energy of an error signal integrated by interferences and noise. In order to reduce the order of the problem, the weight calculation is usually done iteratively.

The most popular adaptive filters are the LMS and Recursive Least Squares (RLS) algorithms (Haykin, 2002). Briefly, the main differences lie in the method to calculate and the final convergence behavior: while LMS has a linear complexity order with the number of antennas in the array, RLS makes use of matrix operation, so that the complexity order is quadratic, but the convergence is faster.

An interesting alternative to the LMS is the Normalized LMS (NLMS), which normalizes the adaptive step to avoid variation during the convergence process. The counterpart is the more intensive processing requirements to calculate signal power and normalization operation.

4.3.3 Correlation-based algorithm

In contrast to beamformers based on temporal reference, schemes based on signal correlation do not require the demodulation of any signal. These techniques are the most popular to extract the spatial information for beamforming, and we have focused on the use of the SIMPLE algorithm (Rogstad, 1997). This algorithm has been used by the Deep Space Network (DSN) of National Aeronautics and Space Administration (NASA) to combine the signals received from spatial probes in radio telescopes located in different sites around the Earth surface. The main disadvantage of correlation based schemes is the lack of ability to cancel interference signals.

4.4 Performance comparison

Some simulation comparisons between spatial and blind algorithms are presented to show benefits and drawbacks. Four algorithms have been selected with a 4-element uniform linear array (ULA). The spatial algorithms simulated are post-beamformer interference canceller – orthogonal interference beamformer (PIC-OIB) (Godara, 2004) and minimum power distortionless response (MPDR) (Van Trees, 2002). On the other hand, the blind algorithms are the matrix-free EIGEN and the SUMPLE (Rogstad, 1997). The convergence process is compared as a function of the input SNR as depicted in Fig. 24.

As it can be observed from the above results, spatial algorithms outperform blind ones at low SNR, and vice versa. On the other hand, with medium-low SNR and low or absence of interferences, the behavior of all algorithms is quite similar.

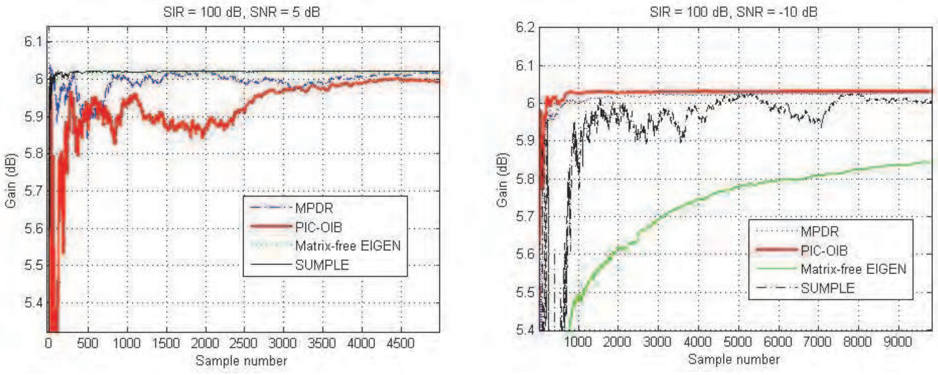

a b

Fig. 24. Convergence behavior of spatial versus blind algorithms in the absence of interferences with several input SNR. a) SNR = 5 dB, and b) SNR = -10 dB.

5. Experimental Test-Bed based on SDR platform

This section presents a test platform known as Adaptive Antenna Array Test-Bed - A3TB, where a comparative study of several beamforming algorithms can be performed and modularity of the architecture is a well proved advantage. The test bed is based on SDR technology and uses a novel architecture that can be used with both blind and spatial-based beamforming algorithms. The A3TB concept can be applied to a number of scenarios as the current version is independent of the signal properties. Simulation results using the A3TB with the APT channel from NOAA satellites show the performance of the concept and the feasibility of the proposed implementation.

The scope of the system development was is to prove the concept of antenna arrays applied to ground stations instead of reflectors for different applications, such as telemetry data downloading or end-user in mobile applications as discussed in the introduction section. In contrast to reflector antennas, antenna arrays offer the possibility of electronic beam-steering avoiding the use of complex mechanical parts and therefore reducing the cost of the antenna. It is also a good chance for Universities and Research Centers aiming to have their own ground station sited in their installations.

5.1 A3TB concept

The A3TB can be defined as *a software-defined radio beamformer applied to a ground station for tracking LEO satellites*. The novelty relies on the use of an antenna array to smartly combine the received signals from the satellite and its implementation based on SDR technology. The reason to use an antenna array instead of a single antenna is to electronically steer the beam in the direction of the satellite along its orbit without requiring a mechanical system for tracking. In addition to the advantages of the use of SDR technology and antenna array, it is the modularity and flexible architecture implemented in the A3TB. Fig. 25 shows the A3TB architecture where it is evident the feasibility to update or change during operation any of the main blocks. It is possible to change during operation the beamforming algorithm and to include new beamforming modules to the system. Furthermore, changes on the BENADC are possible to implement not during operation, but new receiver architecture at off-line such as those options discussed at follow.

In (Salas et al., 2007), the block diagram represents the software system implementation of the first version of the test-bed prototype and most of it is based on VHDL. Depending on the firmware, three options could be installed into the FPGA Virtex4. The option A is implemented with the signal processing on the PC, so the SIMPLE beamforming is done in the module developed in C++. The option B is implemented completely on VHDL and this option need to export the beamforming weights just to draw the array pattern diagram. Finally, in contrast to the option B, the option C is implemented for the LMS beamforming algorithm.

With the first version of the Test-Bed, the modularity on the selection of firmwares was proved switching between A, B or C receiver architectures, and an important result of the Test-Bed development is the hardware resources occupation presented in (Salas et al., 2007).

The advantage of the SDR implementation is that A3TB architecture can be used to process any received signal from a LEO satellite in the appropriate band imposed by the RF stages. Moreover, most of the processing tasks are performed on software, using appropriate routines to process any receive signal. There are 2 main schemes to implement the beamforming stage: SC and FSC [41]. Both schemes are compared in Section 4.2.

The current version of the A3TB in Fig. 25.a was updated to track NOAA satellites in the VHF band, in particular the APT channel. Previous versions of A3TB dealt with LRPT signals from MetOp-A, where a complete receiver with beamforming and synchronization stages has been implemented(Salas et al., 2007; Martínes et al., 2007).

5.2 Implementation of the A3TB

The A3TB prototype consists of 4 main parts as shown in Fig. 25.a. The first part is the antenna array, which has 4 crossed-dipole antennas as depicted in Fig. 25.b. The second part consists of RF-IF circuits which amplify and down convert to IF incoming signals. Furthermore, an automatic gain control (AGC) was implemented using two steps of variable attenuators in the IF domain.

The third part is the SDR platform which consists of the beamforming algorithms implemented on C++ and the FPGA firmware on VHDL, PC and BENADC blocks show in Fig. 25, respectively. The hardware resources occupation for this Test-Bed implementation is similar to one presented in (Martínes et al., 2007). The last part is the software from weather satellite signal to image decoder (WXtoImg) on the PC using the sound card output/input in order to get the weather satellite image.

Since the implemented architecture is FSC the demodulation is not required and the IF signal is digitized. For the signal processing hardware design the BenADC-v4 has been chosen. This solution includes a FPGA Xilinx Virtex4-SX55 with four 12-bit analog inputs at

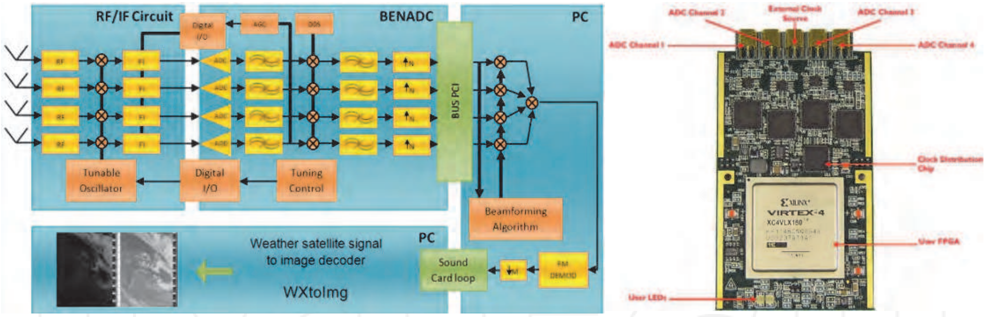
RF/IF Circuit
BENADC
PC
Digital I/O
AGC
DDS
RF
FI
ADC
BUS PCI
Tunable Oscillator
Digital I/O
Tuning Control
Beamforming Algorithm
PC
Weather satellite signal to image decoder
WXtoImg
Sound Card loop
FM DEMOD
ADC Channel 2
External Clock Source
ADC Channel 3
ADC Channel 1
ADC Channel 4
Clock Distribution Chip
XILINX VIRTEX-4
User FPGA
User LEDs

a b

Fig. 25. a) Block diagram of the A3TB, and b) BenADC – Virtex 4-sx55.

250 Msps (Martínes et al., 2007). Digital samples are transferred to the PC where beamforming and subsequent APT demodulation of the array output are performed using C++ routines. This implementation design offers higher flexibility for testing different beamforming schemes. Finally, demodulated APR frames are sent to the WXtoImg software to show meteorological maps.

The A3TB is controlled by the PC for simulations and field trials. The graphical user interface allows presented in (Salas et al., 2008) the user to choose the beamforming algorithm and set all the parameters of the LEO satellite for tracking such as the number of antennas of the array, distance between the elements, direction of arrival and IF frequency. The C++ routine calculates the beamforming weights and plots the synthesized array factor. Subsequently, the reception of meteorological images has real time system requirements. Thus, it is necessary a data transfer from the FPGA to the C++ module to process the samples continuously, and give APT frames to the audio output of the PC. Since, the meteorological satellites often have a low baud rate, in the case of study with NOAA satellites the data transfer is made using two buffers controlled by a thread.

It is important to mention that the A3TB with SDR architecture can evaluate different beamforming algorithms and receiver schemes. The update of A3TB for larger arrays is immediate, as the basis for algorithms is independent of the number of elements in the array. The architecture of a new ground station concept to track LEO satellites based on software defined radio and antenna arraying as Test-Bed is a well proved choice to evaluate future antenna array architectures for satellite communication and benchmark features of the proposed system. As the A3TB VHF version is based on FSC scheme, the concept can be applied to a number of satellite tracing scenarios.

6. Conclusions

The performance analysis of different beamforming algorithms is an important issue in the new generation antenna array development and research. Thus, A3TB helps to analyze beamforming algorithms paving the way for testing and debugging for posteriori use in larger arrays, such as GEODA. Results obtained in real scenarios with A3TB state, for example, that spatial reference algorithms such as MPDR should be used in the absence of interferences, whereas blind algorithms are appropriate for low SNR conditions. Finally, the A3TB can also serve to validate the performance of calibration procedures.

In future work, the A3TB will deal with the system combining of full modularity with the capability of change firmwares based on the first version design of the Test-Bed, plus the flexible architecture of the current design of the Test-Bed based on VHDL, C++ and Antenna Arraying. Furthermore, the addition of more modules to increase the number of antenna array elements is evident in next generations.

7. Acknowledgment

Authors wish to thank MICINN (Ministerio de Ciencia e Innovación) for grants and CROCANTE project (ref: TEC2008-06736/TEC), INSA (Ingeniería y servicios Aeroespaciales) and Antenas Moyano S.L., for the partial funding of this work. Simulations in this work have been done using CST Studio Suite 2011 under a cooperation agreement between Computer Simulation Technology and Universidad Politécnica de Madrid. Substrates used in prototypes were kindly given by NELCO S.A.

8. References

Torre, A.; Gonzalo, J.; Pulido, M., & Martínez Rodríguez-Osorio, R. (2006). New generation Ground Segment Architecture for LEO satellites. *57th International Astronautical Congress*, pp. 221-226. Valencia, Spain, October 2006.

Tomasic, B.; Turtle, J. & Liu, S. (2002). A Geodesic Sphere Phased Array Antenna for satellite control and communication. *XXVII General Assembly of the International Union of Radio Science,* Maastricht, August 2002.

Godara, L. C. (1997). Application of Antenna Arrays to Mobile Communication, Part II: Beamforming and Direction of Arrival Considerations. *Proc. IEEE* , vol.85, No.8, (August 1997), pp. 1195-1245.

Martínez, R. & Salas Natera, M. A. (2010). On the use of Ground Antenna Arrays for Satellite Tracking: Architecture, Beamforming, Calibration and Measurements. *61st International Astronautical Congress*, pp. 1-7. Prague, 2010.

Liu, S.; Tomasic, B.; Hwang, S. & Turtle, J. (2006). The Geodesic Dome Phased Array Antenna (GDPAA) for Satellite Operations Support. *IEEE 18th International Conference on Applied Electromagnetics and Communications*, pp. 1-1, Dubrovnik, April 2006.

Tomasic, B. (1998). Analysis and Design Trade-Offs of Candidate Phased Array Architectures for AFSCN Application. *Second AFSCN Phased Array Antenna Workshop,* Hanscom, April 1998.

Sierra Pérez, M.; Torre, A.; Masa Campos, J. L.; Ktorza, D. & Montesinos, I. (2007). GEODA: Adaptive Antenna Array for Metop Satellite Signal Reception. *4th ESA International Workshop on Tracking, Telemetry and Command System for Space Application*, pp. 1-4, Darmstadt, Germany, September 2007.

Arias Campo, M.; Montesinos Ortego, I.; Fernández Jambrina, J. & Sierra Pérez, M. (2010). T/R Module Design for GEODA Antenna. *4th European Conference on Antenna and Propagation,* pp. 115-116, Barcelona, Spain, April 2010.

Franchi, A.; Howell, A. & Sengupta, J. (2000). Broadband mobile via satellite: Inmarsat BGAN. *IEEE Seminar on the Critical Success Factors - Technology, Services and Markets*, pp. 23-30, October 2000.

Evans, J. V. (1998). Satellite systems for personal communications. *Proc. IEEE* , Vol.86, No.7, (July 1998), pp. 1325-1341.

Wu, W. W. (1994). Mobile satellite communications. *Proc. IEEE* , Vol.82, No.9, (September 1994), pp. 1431-1448.

Fujimoto, K. & James, J. (2001). *Mobile Antenna Systems Handbook.* Artech House.

Garcia Aguilar, A.; Inclán Alonso, J. M.; Vigil Herrero, L.; Fernández González, J. M. & Sierra Pérez, M. (2010). Printed antenna for satellite communications. *IEEE International Symposium on Phased Array Systems and Technology*, pp. 529-535, Boston, USA, October 2010.

Geissler, M.; Woetzel, F.; Bottcher, M.; Korthoff, S.; Lauer, A. & Eube, M., (2010). L-band phased array for maritime satcom. *IEEE International Symposium on Phased Array Systems and Technology*, pp. 518-523, Boston, USA, October 2010.

Baggen, R.; Vaccaro, S. & Del Río, D. (2007). Design Considerations for Compact Mobile Ku-Band Satellite Terminals. *EuCAP The Second European Conference on Antennas and Propagation*, pp. 1-5. Edimburg, Scothland, September 2007.

Vaccaro, S.; Tiezzi, F.; Rúa, M. & De Oro, C. (2010). Ku-Band Low-Profile Rx-only and Tx-Rx antennas for Mobile Satellite Communications. *IEEE International Symposium on Phased Array Systems and Technology*, pp. 536-542, Boston, USA, October 2010.

Schippers, H. (2008). Broadband Conformal Phased Array with Optical Beam Forming for Airborne Satellite Communication. *IEEE Aerospace Conference*, pp. 1-17, September 2008.

Kanno, M.; Hashimura, T. T.; Sato, M.; Fukutani, K. & Suzuki, A. (1996). Digital beam forming for conformal active array antenna. *IEEE International Symposium on Phased Array Systems and Technology*, pp. 37-40, October 1996.

Khalifa, I. & Vaughan, R. (2007). Optimal Configuration of Multi-Faceted Phased Arrays for Wide Angle Coverage. *IEEE 65th Vehicular Technology Conference*, pp. 304-308, Boltimore, USA, April 2007.

Salas Natera, M. A.; Martínez Rodríguez-Osorio, R.; Antón Sánchez, A.; García-Rojo, I. & Cuellar, L. (2008). A3TB: Adaptive Antenna Array test-bed for tracking LEO satellites based on software defined radio. *59th International Astronautical Congress*, pp. 313-317, Glasgow. September 2008.

Salas Natera, M. A.; Martínez Rodríguez-Osorio, R. & García-Rojo López, I. (2007). Design of an Adaptive Antenna Array Test-Bed based on Software Radio for Tracking LEO Satellites. *IEEE EuCAP.* Edinburgh, Scotchland, Npvember 2007.

Martínez, R.; Salas Natera, M.; Bravo, A.; García-Rojo, I.; de Haro, L.; Mateo, M. & Gómez, M. (2007). VHF Ground Station with increased angular coverage for reception of meteorological satellites with electronic beamforming. *4th ESA International Workshop on Tracking, Telemetry and Command Systems for Space Application.* Darmstadt, Germany, 2007.

Mailloux, R. (2005). *Phased Array Antenna Handbook,* Artech House, Norwood, Massachusetts, USA.

Josefsson, L. & Persson, P. (2006). *Conformal Array Antenna. Theory and Design*, John Wiley & Sons , Hoboken, New Jersey, USA.

Montesinos, I.; Sierra Pérez, M.; Fernández, J. L.; Martínez, R. & Masa, J. L. (2009). GEODA: Adaptive Antenna of Multiple Planar Arrays for Satellite Communications. *European Conference on Antenna and Propagation.* Berlin, Germany, 2009.

Salas Natera, M. A.; Martínez, R.; De Haro Ariet, L. & Fernández Jambrina, J. (2010). Automated System for Measurement and Characterization of Planar Active Arrays. *IEEE Internationa Symposium on Phase Array Systems and Technology*, pp. 1-6, Boston USA, October 2010.

Sierra-Castañer, M.; Vera-Isasa, M.; Sierra-Pérez, M. & Fernández-Jambrina, J. (2005). Double-Beam Parallel Plate Slot Antenna. *IEEE Transactions on Antennas and Propagation*, vol.53, No.3, (2005) pp. 977-984.

Garg, R.; Bhartia, P.; Bahl, I. & Ittipiboon, A. (2001). *Microstrip Antennas Design Handbook,* Artech House, Norwood, Massachusetts, USA.

Tang, C. & Chen, M. (2007). Synthesizing Microstrip Branch-Line Couplers With Predetermined Compact Size and Bandwidth. *IEEE Transactions on Microwave Theory and Techniques, vol. 55, No.9,* (September 2007), pp. 1926-1934.

Bialkowski, M.; Jellett, S. & Varnes, R. (1996). Electronically Steered Antenna System for the Australian Mobilesat. *IEEE Transaction on Antennas ans Propagation*, vol. 143, No. 4, (August 1996), pp. 347-352.

Alonso, J.; Blas, J.; Garcia, L.; Ramos, J.; Pablos, J. & Grajal, J. (1996). Low Cost Electronically Steered Antenna and Receiver System for Mobile Saltellite Communications. *Trans. IEEE MTT*, vol. 44, No. 12, (December 1996), pp. 2438-2449.

Fernández, J. M.; Rizzo, C. & Sierra-Pérez, M. (2009). Antena Impresa Multihaz con Polarización Circular Derechas/Izquierdas Para Comunicaciones por Satélite en Banda X. *Proccedings of XXIV Simposium Nacional de la Unión Científica Internacional de Radio (URSI).* Santander, Spain, September 2009.

Encinar, J. A. & Zornoza, J. (2003). Broadband design of three- layer printed reflectarrays. *IEEE Transactions on Antennas and Propagation*, vol. 51, No. 7, (July 2003), pp. 1662-1664.

Padilla, P.; Muñoz-Acevedo, A.; Sierra-Castañer, M. & Sierra-Pérez, M. (2010). Electronically reconfigurable transmitarray at Ku band for microwave applications. *IEEE Transactions on Antennas and Propagation*, vol. 58, No. 8, (August 2010), pp. 2571-2579.

Padilla, P.; Muñoz-Acevedo, A. & Sierra-Castañer, M. (2010). Low Loss 360º Ku Band electronically Reconfigurable Phase Shifter. *International Journal of Electronics and Communications,* vol. 64, No. 11, (November 2010), pp. 1100-1104.

Royce, W. W. (1970). Managing the Development of Large Software Systems. *IEEE WESCON*, pp. 328 – 338, 1970.

Rogstad, D. H. (1997). *The SUMPLE Algorithm for Aligning Arrays of Receiving Radio Antennas: Coherence Achivied with Less Hardware and Lower Combining Loss.* TDA Progress Report, Jet Propulsion Laboratory.

Rogstad, D. H.; Mileant, A. & Pham, T. (2003). *Antenna Arraying Techniques in the Deep Space Network.* Deep Space Communication and Navegation Series, JPL California Institute of Technology, Ref. 03-001, Pasadena, USA.

Biedka, T. E. (2001). *Analysis and Development of Blind Adaptive Beamforming Algorithms.* PdD Thesis, Faculty of Virginia Polytechnique Institute and State University, Virginia.

Haykin, S. (2002). *Adaptive Filter Theory* (4th ed.). Prentice-Hall.

Godara, L. C. (2004). *Smart Antennas* (1st ed.). CRC Press..

Van Trees, H. L. (2002). *Optimum Array Processing. Part IV of Detection, Estimation, and Modulation Theory*. Wiley.

INTECH

INTECH

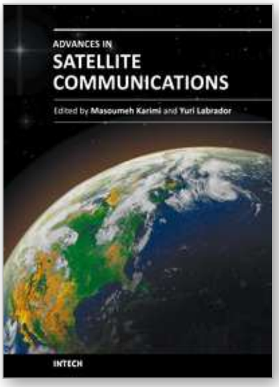

Advances in Satellite Communications

Edited by Dr. Masoumeh Karimi

ISBN 978-953-307-562-4
Hard cover, 194 pages
Publisher InTech
Published online 27, July, 2011
Published in print edition July, 2011

Satellite communication systems are now a major part of most telecommunications networks as well as our everyday lives through mobile personal communication systems and broadcast television. A sound understanding of such systems is therefore important for a wide range of system designers, engineers and users. This book provides a comprehensive review of some applications that have driven this growth. It analyzes various aspects of Satellite Communications from Antenna design, Real Time applications, Quality of Service (QoS), Atmospheric effects, Hybrid Satellite-Terrestrial Networks, Sensor Networks and High Capacity Satellite Links. It is the desire of the authors that the topics selected for the book can give the reader an overview of the current trends in Satellite Systems, and also an in depth analysis of the technical aspects of each one of them.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Miguel Alejandro Salas Natera, Andrés García Aguilar, Jonathan Mora Cueva, José Manuel Fernández, Pablo Padilla De La Torre, Javier García-Gasco Trujillo, Ramón Martínez Rodríguez-Osorio, Manuel Sierra-Perez, Leandro De Haro Ariet and Manuel Sierra Castañer (2011). New Antenna Array Architectures for Satellite Communications, Advances in Satellite Communications, Dr. Masoumeh Karimi (Ed.), ISBN: 978-953-307-562-4, InTech, Available from: http://www.intechopen.com/books/advances-in-satellite-communications/new-antenna-array-architectures-for-satellite-communications

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

Tran *et al. Micro and Nano Syst Lett (2017) 5:14*
DOI 10.1186/s40486-017-0051-0

Micro and Nano Systems Letters

REVIEW **Open Access**

RF power harvesting: a review on designing methodologies and applications

Le-Giang Tran[1], Hyouk-Kyu Cha[2] and Woo-Tae Park[1,3*]

Abstract

Wireless power transmission was conceptualized nearly a century ago. Certain achievements made to date have made power harvesting a reality, capable of providing alternative sources of energy. This review provides a summary of radio frequency (RF) power harvesting technologies in order to serve as a guide for the design of RF energy harvesting units. Since energy harvesting circuits are designed to operate with relatively small voltages and currents, they rely on state-of-the-art electrical technology for obtaining high efficiency. Thus, comprehensive analysis and discussions of various designs and their tradeoffs are included. Finally, recent applications of RF power harvesting are outlined.

Keywords: Wireless power harvesting, Radio frequency, Impedance matching network, Voltage multiplier, Antenna, Wireless sensor networks

Background

Motivation for wireless energy harvesting

An early definition of a wireless power transmission system portrays a unit that emits electrical power from one place and captures it at another place in the Earth's atmosphere without the use of wires or any other supporting medium [1]. The history of RF power scavenging in free space originated in the late 1950s with a microwave-powered helicopter system [2]. Later, the concept of power harvesting or energy scavenging was explained as a technique for reaping energy from the external environment using different methods including thermoelectric conversion, vibrational excitation, solar energy conversion, and pressure gradients. This technique promises tremendous scope for the replacement of small batteries in low power electrical devices and systems. Figure 1 introduces the structure of an RF energy harvesting system and factors that contribute to the performance of the whole system.

RF wireless power harvesting (WPH) holds vast potential for replacing batteries or increasing their lifespans. Currently, batteries power a majority of low-power remote sensor devices and embedded equipment. In fact, batteries have finite lifespans and require periodic replacements. By applying power harvesting technologies, devices and equipment can become self-sustaining with respect to the energy required for operation, thereby obtaining an unlimited operating lifespan. Thus, the demand for power maintenance will become negligible.

The sources for WPH are available in many forms, such as solar power [3–5], wind energy [6–8], thermal energy [9–11], electromagnetic energy [12–14], kinetic energy [15–17], etc. Among them, electromagnetic energy is abundant in space and can be retrieved without limit. Electromagnetic waves come from a variety of sources such as satellite stations, wireless internet, radio stations, and digital multimedia broadcasting. A radio frequency power harvesting system can capture and convert electromagnetic energy into a usable direct current (DC) voltage. The key units of an RF power harvesting system are the antenna and rectifier circuit that allows the RF power or alternating current (AC) to be converted into DC energy.

The processing of battery wastes is a critical problem. A majority of batteries end up in landfills, leading to the pollution of the land and water underneath. The most effective solution for reducing battery wastes is to

---

*Correspondence: wtpark@seoultech.ac.kr
[1] Convergence Institute of Biomedical Engineering and Biomaterials, Seoul National University of Science and Technology, Seoul, South Korea
Full list of author information is available at the end of the article

© The Author(s) 2017. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

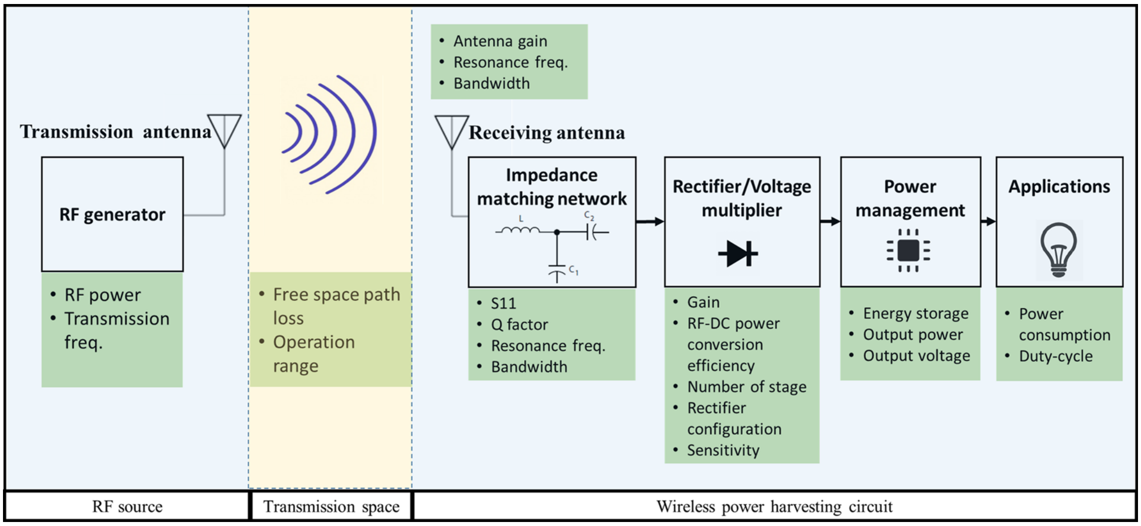
Transmission antenna
RF generator
RF power
Transmission freq.
Free space path loss
Operation range
Antenna gain
Resonance freq.
Bandwidth
Receiving antenna
Impedance matching network
L
C2
C1
S11
Q factor
Resonance freq.
Bandwidth
Rectifier/Voltage multiplier
Gain
RF-DC power conversion efficiency
Number of stage
Rectifier configuration
Sensitivity
Power management
Energy storage
Output power
Output voltage
Applications
Power consumption
Duty-cycle
RF source
Transmission space
Wireless power harvesting circuit

Fig. 1 Conceptual block diagram of an RF power harvesting system

avoid using them. Applying WPH technology will help to reduce the dependency on batteries, which will ultimately have a positive impact on the environment. Moreover, the process of harnessing electromagnetic energy will not generate wastes as it is a clean energy source.

A summary of the present sources of energy available for power harvesting is shown in Table 1. The data in Table 1 was collected from references [18, 19]. In comparison with thermal or kinetic energy, electromagnetic energy is not limited by space or time. The RF wave is available both indoors and outdoors, in rural and urban areas, throughout the day. Despite its low power density in the environment, an intentional source can be added for more efficient power transmission and a boosting circuit can be built to suit the requirements of the load application. This feature promotes research to realize RF WPH technology through applications such as wireless sensor networks (WSNs) and Internet of Things (IoT).

Fundamentals of RF transmission

Understanding electromagnetic waves is important to design an RF WPH system. The behavior of electromagnetic waves varies according to the distance, frequency, and conducting environment. Depending on the requirements of the application, the designer needs to select suitable parameters for electromagnetic waves to obtain best results.

The loss of power in space can be characterized by free space path loss (FSPL), which is the loss of signal power during propagation in free space. Calculating FSPL requires information about the antenna gain, frequency of transmitting wave, and distance between the transmitter and receiver.

The behavior of electromagnetic waves depends upon the distance from the transmitting antenna. These characteristics are categorized into two segments: far-field and near-field [20]. While the electromagnetic wave pattern at the far-field is relatively uniform, in the near-field,

Table 1 Overview of alternative sources of energy to replace batteries

Source	Power density	Harvesting tech.	Advantages	Disadvantages
Solar	Indoor: 10 $\mu W/cm^2$ Outdoor: 10 mW/cm^2	Photovoltaic	High power density Mature	Not always available Required exposure to light (not implantable) Expensive
Vibration	Human: 4 $\mu W/cm^2$ Industrial: 100 $\mu W/cm^2$	Piezoelectric Electrostatic Electromagnetic	Implantable High efficiency	Not always available Material physical limitation
Thermal	Human: 30 $\mu W/cm^2$ Industrial: 1–10 mW/cm^2	Thermoelectric Pyroelectric	High power density Implantable	Not always available Excess heat
RF	GSM: 0.1 $\mu W/cm^2$ WI-FI: 1 mW/cm^2	Antenna	Always available Implantable	Low density Efficiency inversely proportional to distance

the electric and magnetic components are very strong and independent such that one component can dominate the other. The near-field region is considered as the space that lies within the Fraunhofer's distance and the far-field region lies outside the Fraunhofer's distance. The distribution of the near and far field regions is illustrated in Fig. 2. The Fraunhofer's distance is defined as

$$d_f = \frac{2D^2}{\lambda}, \tag{1}$$

where d_f is the Fraunhofer's distance, D is the maximum dimension of the radiator (or diameter of the antenna), and λ is the wavelength of the electromagnetic wave. Although the Fraunhofer creates a boundary region, the actual transition between regions is not distinct. Inside the near-field, the space from the antenna to a distance of $\sqrt[3]{\frac{D}{2\lambda}}$ is a non-radiative/reactive near-field region, where E and H fields are not in phase, creating energy distortion. As we propagate forward in this region towards the far-field, we encounter a radiative near-field or Fresnel region, where the reactive behavior of electromagnetic waves is not dominant but the phases of E and H fields still vary with distance.

For a transmitter–receiver antenna in the far-field free space, the power propagation at the receiver antenna can be expressed as

$$P_R = \frac{P_T G_T G_R \lambda^2}{(4\pi R)^2}, \tag{2}$$

where P_R is power at the receiver antenna; G_R is receiver antenna gain relative to the isotropic source (dBi); λ is the wavelength of the electromagnetic signal, which is equal to the speed of light in vacuum divided by the signal frequency, $\lambda = \frac{c}{f}$; and $k = 2\pi/\lambda$ is the wave number. From the above formula, the FSPL, P_L for far-field can be inferred as

$$P_L = \frac{P_T}{P_R} = \frac{(4\pi R)^2}{G_T G_R \lambda^2} = \frac{(4\pi f R)^2}{G_T G_R c^2} = \frac{4}{G_T G_R}(kR)^2 \tag{3}$$

or

$$\begin{aligned} P_L(dB) = {} & 20\log_{10}(f) + 20\log_{10}(R) \\ & + 20\log_{10}\left(\frac{4\pi}{c}\right) - G_T - G_R \end{aligned} \tag{4}$$

In case f is measured in MHz, distance R is measured in km, and gain G_T and G_R are measured in dBi, the above function becomes

$$P_L(dB) = 20\log_{10}(f) + 20\log_{10}(R) + 32.44 - G_T - G_R \tag{5}$$

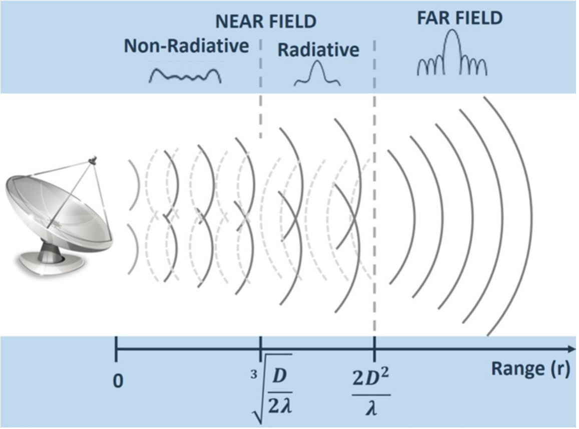

Fig. 2 Distribution of near field and far field regions in space

By using the path loss equation, it is possible to indicate the signal power at the far field region. However, this equation does not present all the factors that affect the propagation process such as reflection, diffraction, absorption, etc.

In contrast to the uniform far-field wave, the electric and magnetic fields behave unusually in the near-field. The relationship between electric and magnetic waves in the near-field varies unpredictably with time and space and is too complicated to predict. All these factors make estimating the power density at this range problematic.

The FSPL estimation process is critical to design the energy-harvesting unit. Being informed about the amount of power that the system is required to handle helps the designer in choosing the right technology and method.

Since the electromagnetic field (EMF) is widely used for a variety of purposes, there must be a standard to ensure the safe operation of EMF devices. For this purpose, the International Electrotechnical Commission [21] established international standards and frequency limits for EMFs. For safety measures, there must be a separate emission and immunity standard for each class of application including broadcast, communication, military, medical, and household appliances.

On the other hand, radiation exposure must be considered carefully, especially in biomedical applications. EMFs at different spectrums affect the human body differently. Between 1 MHz to 10 GHz frequency, EMF possibly penetrates through tissues and produce heat because of energy absorption. EMFs over 10 GHz are strongly blocked by the skin, and then the heat created causes damage such as eye cataracts or skin burns if the field density is above 1000 W/m^2. Hoping to prevent negative health effects when utilizing EMF applications, the International Commission on Non-Ionizing Radiation Protection (ICNIRP) developed general guidelines for EMF exposure limits recognized under the World

Health Organization (WHO). To date, there have been no consistent reports corroborating that lasting exposure to EMF shortens one's life span or induces adverse health conditions. However, for safety issues, RF emission power and frequency must be considered carefully to avoid unwanted damages.

Wireless power harvesting evaluation metrics

There are various parameters that need to be evaluated, which decide the performance of a WPH design. Evaluation merits change depending on different applications. Nevertheless, critical values such as efficiency, sensitivity, operation distance, output power are defined as standards to make comparisons. However, tradeoffs exist between these values such as operation distance and overall efficiency. Moreover, besides this merit, other manufacturing auxiliary factors like low cost, maturity of fabrication process, and bulk manufacturing availability are also paramount.

Operation range

Operation distance is mostly related to operation frequency. In fact, transmission at high frequencies is attenuated by atmospheric conditions more than at low frequencies while low frequency penetrates deeper through matter. Therefore, if the application of WPH is for implantable devices, then the transmitting frequency should not exceed the megahertz range.

RF-DC power conversion efficiency (PCE)

This is the ratio between the amount of power applied to the load and that retrieved by the antenna. Generally, the RF-DC PCE covers the efficiency of the rectifier, voltage multiplier, and storage elements. PCE can be simply calculated as the ratio of power delivered to the load to the retrieved power. RF transmission loss in space is not considered in this term.

$$\eta_{PCE} = \frac{P_{load}}{P_{retrived}}, \tag{6}$$

where P_{load} is power delivered to the load and $P_{retrieved}$ is harvested power at the antenna. Factors that determine the value of PCE include parasitic effects, leakage in the circuits, design topologies, and nonlinear thresholds of electrical components.

Resonator Q factor

The Q factor is generally defined as a dimensionless value that describes how strong the resonance is and the bandwidth of resonance [22]. In electric circuits, the Q factor represents how much the peak voltage increases when the system resonates at resonant frequency. From this concept, the equation of Q factor can be express as the formula below [23]:

$$Q = 2\pi \frac{Energy\,stored}{Energy\,dissipated\,per\,cycle}\,or\;Q = \frac{f_c}{\Delta f} \tag{7}$$

From the equation above, it is inferred that a high Q factor comes with a narrow resonant bandwidth, but high voltage gain is obtained at resonance. Further, this equation also indicates that the Q factor is inversely proportional to the energy dissipated per cycle. The specific Q factor for capacitor C and inductor L at frequency ω is given by

$$\text{Capacitor Q factor:}\, Q_C = \frac{1}{\omega R_c C} = \frac{X_c}{R_c}, \tag{8}$$

$$\text{Inductor Q factor:}\, Q_L = \frac{\omega L}{R_L} = \frac{X_L}{R_L}, \tag{9}$$

where R_C is the series resistance of capacitor and R_L is the series resistance of inductor. From the equations of Q_C and Q_L, we note that resistive component is the cause of power dissipation. Additionally, in an electrical circuit, energy loss can be reduced by adding reactive components such as capacitors or inductors and minimizing resistive components. Since obtaining a high Q factor is a usual consideration in designing WPH systems, this feature is typically included in the impedance matching network design.

Sensitivity

Sensitivity of a WPH system is defined as the minimum limit of incident power that is required to trigger the operation of the system. Sensitivity is quantified as follows:

$$Sensitivity\,(dBm) = 10\log_{10}\left(\frac{P}{1mW}\right), \tag{10}$$

where P is the minimum power the system requires to perform a task.

The threshold voltage of CMOS technology affects sensitivity. A lower threshold CMOS is more sensitive but comes with more leakage current, which reduces the overall efficiency.

Output power

Usually, the outcome of a WPH system is DC power, which is characterized by load voltage V_{DD} and current I_{DD}. Measuring open-load voltage demonstrates the performance of WPH in general since V_{DD} and I_{DD} depend on load impedance. If the load is a sensor, V_{DD} is more important than I_{DD} while in applications like electrolysis or LED, current is the dominant parameter.

Recent advances in wireless power harvesting

Antenna/rectenna design

Gain, resonance frequency, bandwidth are parameters that characterize the performance of an antenna. Assuming an unobstructed space and isotropic transmitting source, the spreading of the waves in all directions is uniform. Thus, the power per unit area at a distance from the source is inversely proportional to the square of the distance:

$$S_{isotropic} = \frac{P_T}{4\pi R^2}, \tag{11}$$

where R is the distance from the source, S is power per unit area at the distance R, and P_T is the transmitted power.

However, it is important to note that antennas do not always transmit power spherically (isotropic antenna), they also transmit energy in some specific directions according to their designs. The ratio between the maximum power density of an antenna at a given distance to the power density of an optimal omnidirectional or isotropic antenna at the same distance, radiating the same power, is known as antenna gain (G). This parameter represents the directivity of an antenna. Since power density (S) is non-directional, the definition of S was developed as a function of direction i.e. S (θ, φ). Consequently, the gain of the antenna is a directional function as well. It is defined as the ratio of the antenna's power density in a given direction to the power density of an isotropic antenna as formula below

$$G(\theta, \varphi) = \frac{S(\theta, \varphi)}{S_{isotropic}} = \frac{4\pi R^2 S(\theta, \varphi)}{P_t}, \tag{12}$$

Thus, the power density at a distance R from an antenna in general is given by

$$S = \frac{P_T G_T}{4\pi R^2}, \tag{13}$$

where P_T is the transmitted power by the antenna and G_T is the transmitting antenna gain. It is known that an ideal isotropic antenna has $G_T = 0$ dBi. The aforementioned formula is also applicable to receiving antennas. Specifically, for power harvesting applications, a receiving antenna, which constitutes a rectifier, is called a rectenna.

Preference for a high gain antenna depends on the application requirements. In case of RF transmission, if the positions of source and receiving antennas are known, then a high gain rectenna is advantageous. On the other hand, if the positions of the source and receiving antenna are relatively uncertain, a low gain antenna is preferable in order to collect signals from various directions simultaneously.

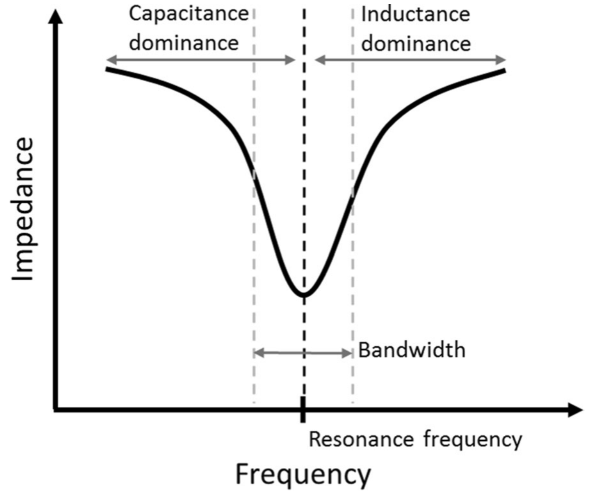

Fig. 3 Correlation between impedance and resonance frequency, and bandwidth of an antenna

Every antenna has its own optimal operation frequency known as resonance frequency (Fig. 3). The resonance frequency is determined by the capacitance and inductance characteristics of the antenna. As frequency increases, inductive behavior becomes dominant and capacitance decreases. The frequency at which the inductance and capacitance nullify each other, minimizing the impedance of the antenna, is called the resonance frequency.

The capacitance and inductance are functions of the frequency and physical size of the antenna. The larger the dimensions of the antenna, the lower the resonance frequency. Therefore, transmitting and receiving low frequency waves requires a large aperture, which is not suitable for small device applications. The antenna designed in [24] has a sensitivity as low as −35 dBm (0.32 μW), but the tradeoff is a large size aperture measuring up to 64.68 cm^3 (7 × 7 × 1.32 cm).

The bandwidth of an antenna is the range of frequencies in which the antenna can operate efficiently. A wide bandwidth antenna can collect signals from a wider range of frequencies than a narrow bandwidth antenna. Hence, a wide bandwidth antenna is advantageous in retrieving incident energy but presents a high risk of interference by noise through unwanted frequencies (Table 2).

Communication antennas have been studied for decades. However, power-harvesting antennas are currently in the developmental stage. At first, antenna classification was based on design characteristics and applications. It originally included wire antennas, aperture antennas, printed planar antennas, and reflector antennas. An illustration of some examples of antennas is shown in Fig. 4. To date, the growth in technology has paved the way for

Table 2 Comparison of published work regarding power-harvesting antennas

Ref.	Type	Gain (dBi)	Freq. (GHz)	Dimension (mm)	Sensitivity	RF-DC PCE
[25]	Air-substrate patch	7	2.45	261 × 5	–	30% @ 1 μW/cm^2
[26]	Polarization patch	–	5.1–5.8 5.8–6.1	90 × 160	–	23.8%–31.9% 22.7%–24.5%
[27]	Patch	–	2.45	100 × 70	–	73.9% @ 207 μW/cm^2
[28]	Dual-linearly polarized patch	7.45–7.63	2.45	70 × 47.5	–	78% @ 295.3 μW/cm^2
[29]	Microstrip	5.3	160	–	–	8.5% @ −2.46 dBm
[30]	Patch	–	1.9–3.2	40 × 40 × 0.8	20 μW/cm^2	70% @ 50 μW/cm^2
[31]	Patch	−20.20	0.915	Π × 6 × 1.584		~ 60%
[32]	Patch, metal strip	4.33 6.64	4.9 5.9	68 × 34	–	65.2% 64.8%
[33]	Stack differential	–	5.8	120 × 40	–	44.1% @ 0.041 W/m^2
[24]	Cross dipole	–	1.8–2.5	70 × 70 × 13.2	−35 dBm	55% @ −10 dBm
[34]	Dual-polarized patch	–	2.45	100 × 100 × 3.8	–	82.3% @ 22 dBm
[35]	1 × 4 quasi-Yagi	10.9 13.3	1.8 2.2	300 × 300 × 1.6	–	40% @ 455 μW/cm^2
[36]	Dipole	–	0.915 2.45	60 × 60 × 60	1 μW/cm^2	48% @ 0 dBm 39% @ 0 dBm
[37]	Microstrip	8.6	2.45	–		83%
[38]	Patch	4	2.45	–	–	70%
[39]	Patch	2.19 3.6	2.45 5.8	40 × 43	–	65% @ 10 mW/cm^2 46% @ 10 mW/cm^2

a variety of antenna design and fabrication methods for making it more compact and mature.

The plate antennas are popular and have many applications [27, 34, 38]; on-chip antennas are preferred for small and compact applications. Recently, many publications addressed wide-band and multi-band antennas. It has been proven that narrow-band antennas offer high energy conversion efficiencies but can only retrieve a limited amount of energy. On the other hand, wide-band or multi-band frequency antennas can retrieve more RF energy in space. However, the tradeoffs are low overall efficiency and large aperture. In [32], antennas with a resonance frequency of 4.9 and 5.9 GHz were designed with PCEs of 65.2 and 64.8%, respectively. Further work by Lu et al. [26] on polarization antennas supports the assertion that expanding the bandwidth of an antenna leads to increasing the amount of power harvested. In this work, the demonstration of broadband polarization antennas with three separate modes allows the antenna to operate in a wider range of frequencies. One common mechanism in the aforementioned works is the control of the antenna configuration by switching the diodes on and off, thus altering its resonant frequencies. However, since it uses separate modes for different frequencies, this antenna is not able to simultaneously resonate at two frequencies. On the other hand, the antenna presented in [39] is capable of operating at 2.45 and 5.8 GHz, simultaneously, providing 2.6 V output with a PCE of 65% and power density of 10 mW/cm^2.

The main purpose of aligning antennas in arrays is to enhance the antenna gain and obtain high voltage/current. Array antennas are preferred over large aperture antennas because they do not require large breakdown voltage diodes to operate. Antenna arrays can be connected before or after rectification. The first configuration enhances the retrieved power at the main beam while the second configuration expands the ability to retrieve power from various angles away from the main beam [40]. In case the RF waves are combined before rectification, the rectifier requires a large breakdown diode. If RF waves are combined after rectification, combining DC current becomes problematic. Antenna arrays can be connected in series or parallel to obtain high voltage or large current. Nonetheless, expanding the arrays yields better output but this might cause a deduction in conversion efficiency [33].

For demonstrating antenna arrays, Sun et al. [35] invented a T-junction to connect four quasi-Yagi antennas together. The advancement in this work was that the T-junction was flexible in changing from 1 × 4 array to 2 × 2 array topologies. Consequently, the system was able to operate at an ambient power level as low as 455 μW/cm^2 while obtaining 40% PCE.

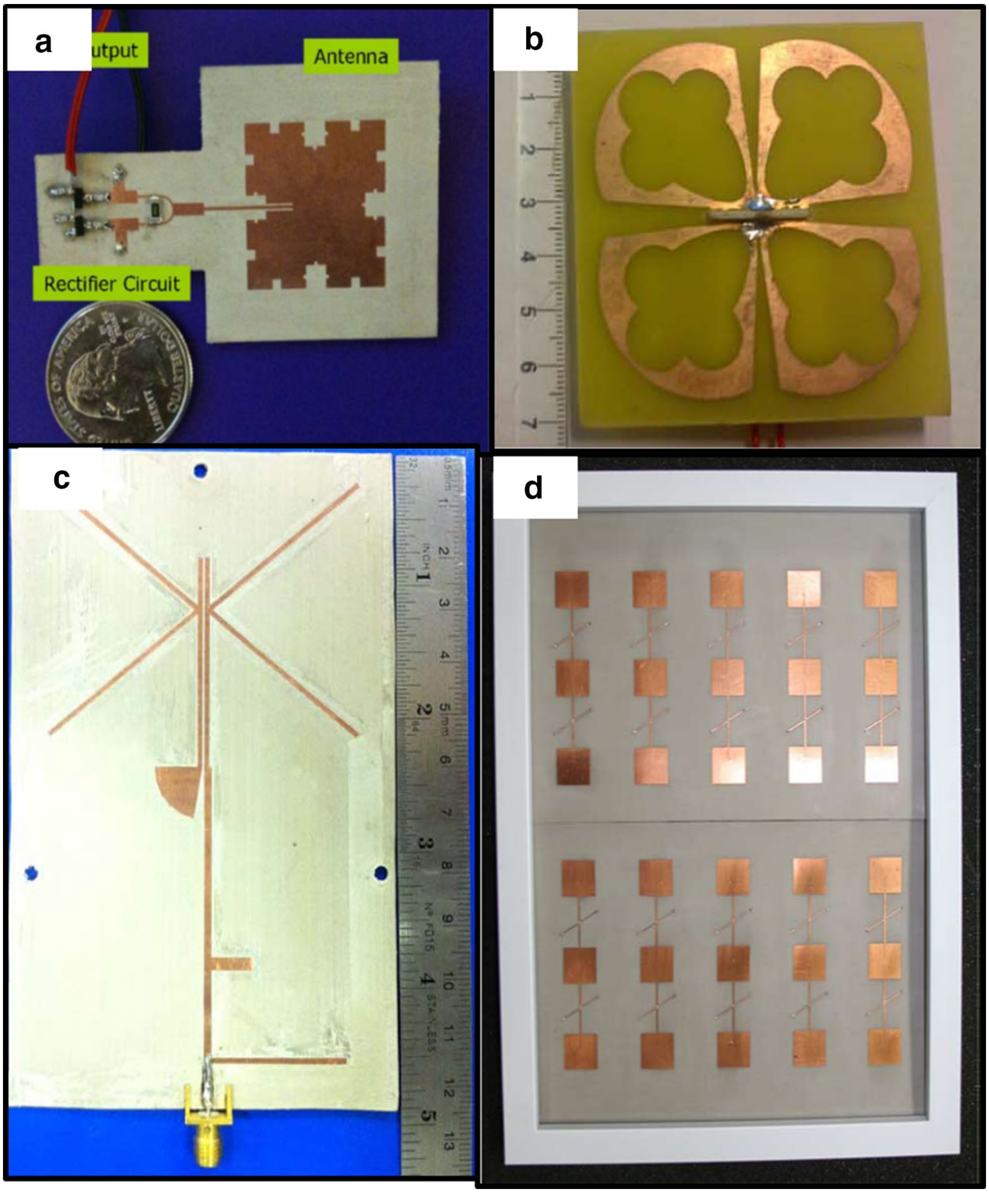

Fig. 4 Types of antennas **a** 2.45 GHz patch antenna with rectifier [38]. ©2010 IEEE. All rights reserved; **b** Planar dual-polarized antenna [24]. ©2015 IEEE. All rights reserved; **c** Microstrip antenna [37]. ©2012 IEEE. All rights reserved; **d** Array of stacked differential patch antenna [33]. ©2015 IEEE. All rights reserved

Impedance matching network

In low-power consumption electrical systems, power leakage during transmission may lead to energy insufficiency. In these circumstances, adding an impedance matching network (IMN) ensures that the maximum power transfers between the RF source and load. For WPH applications, the receiving antenna is considered as the source while the rectifier/voltage multiplier is considered as the load. It is acknowledged that in DC, power transfer is optimum when the resistances of the source and load are indistinguishable. In an RF circuit, the impedance is referred to instead of resistance. An impedance mismatch between the source and load creates reflected power flow in the circuit that lowers the efficiency of the system. As its name indicates, the IMN ensures that the impedance of the source and load are identical by adding reactive components in between.

There are three basics matching configurations i.e. L, T and π matching networks (Fig. 5). The L matching is commonly used since it typically has two components,

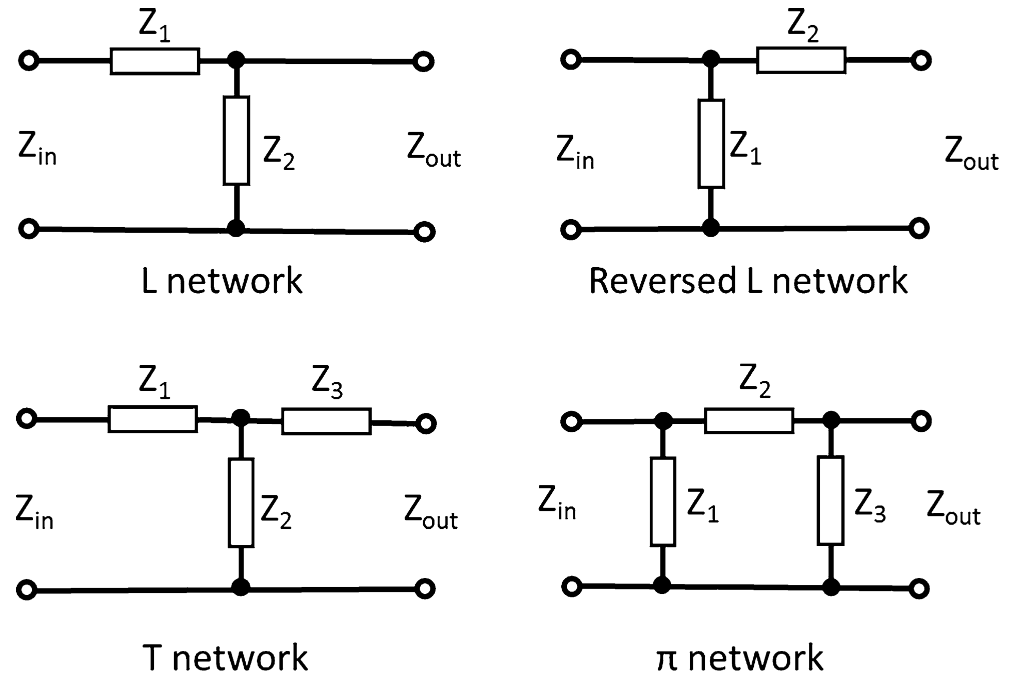

Fig. 5 Configuration of common impedance matching networks

which simplifies the designing and controlling process. Additionally, the L matching networks do not alter the quality factor (Q) of the circuit.

The T and π matching configurations are more complex than the L network. Furthermore, organizing the T and π configurations into multiple stages will retain the final matching results but will change the Q factor. This strategy is useful in improving voltage boost.

There are tradeoffs between the attributes of an IMN, which include frequency, bandwidth, adjustability, and complexity. For instant, in [41], the suboptimal impedance matching and multiport ladder matching methods were introduced to enhance the matching performance and harvested power of antennas, respectively. However, the tradeoff was that implementation of these configurations required more components than traditional matching networks, thus, escalating the circuit's complexity.

Etor et al. [42] designed an IMN for THz frequencies applications using transmission lines and self-designed metal–insulator-metal diodes instead of lumped components. Moreover, fixed IMN and tunable IMN [43–45] were introduced as a technique for better matching with wide-band and multi-band antennas.

Rectifier/voltage multiplier

RF energy extracted from free space usually possesses low power density since the electric field power density decreases at the rate of $1/d^2$, where d is the distance from the RF source [46]. Therefore, a power amplifier circuit is required that yields enough DC energy from the electromagnetic waves to drive the loads. This gives rise to two possibilities, if the power consumption of the load is lower than the average power harvesting, the electronic devices at the load may work continuously; otherwise, if the load consumes more energy than the power harvesting circuit can generate, the devices cannot work continuously [47].

Rectifying is the most popular application of diodes, which refers to the conversion of AC current to DC current. In terms of power harvesting application, the RF signal retrieved in the antenna has a sinusoidal waveform. The signal after transformation through IMN would be rectified and boosted to meet the power requirements of the application.

The most fundamental topology of the rectifier is the half-wave rectifier that comprises of a single diode D1 (Fig. 6a). When AC voltage transfers through D1, only the positive cycle remains and the negative cycle is cutoff; thus, it diminishes half of the AC power. Moreover, the output V_{out} is discontinuous since the negative cycle is cutoff. Despite its simplicity, a half-wave rectifier is usually inadequate for common applications. Hence, a full-wave rectifier is more preferable. The circuit design of the full-wave rectifier is shown in Fig. 6b. During the first negative cycle of AC input, diode D1 is conductive and capacitor C1 is charged to the corresponding energy

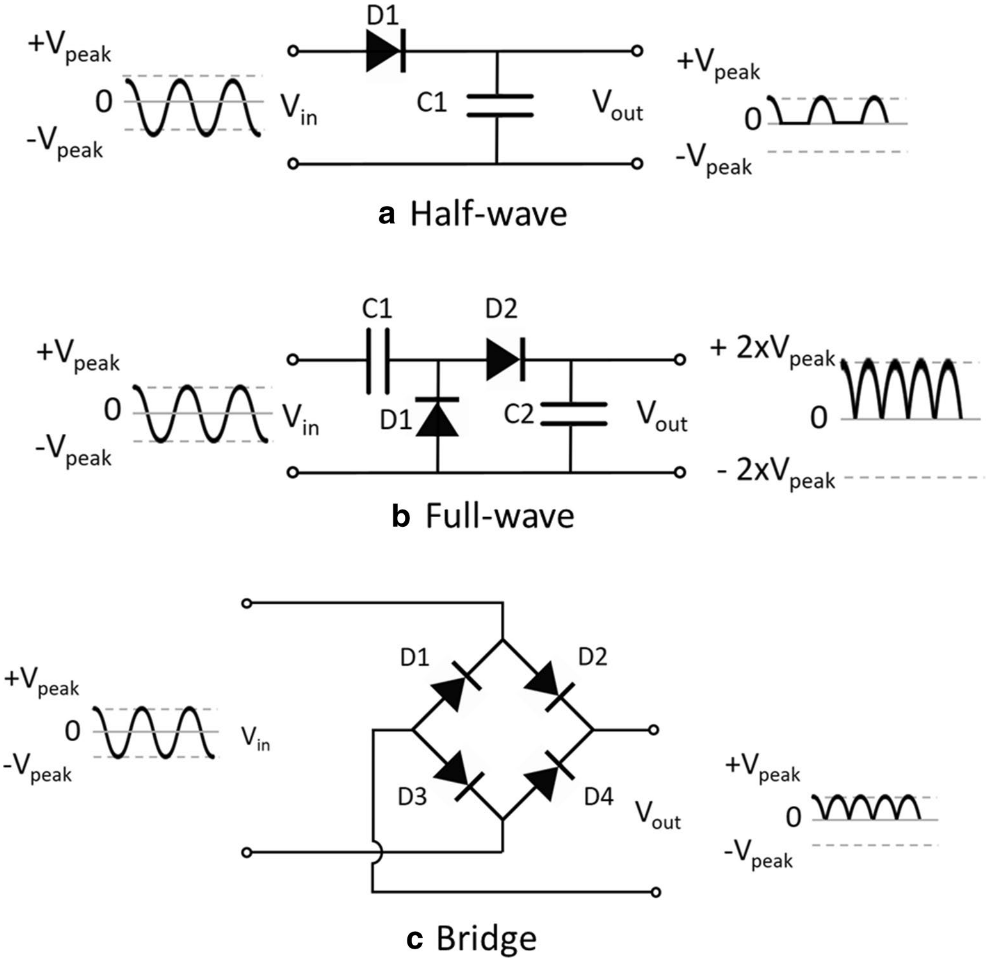

Fig. 6 Some common topologies of a rectifier

level of V_{peak} of the input. Then, at the next positive cycle, diode D1 is blocked, diode D2 is conductive so that capacitor C2 is also charged. In consequence, the output V_{out} would see two capacitors in series (each one is storing a voltage of $V_{peak)}$. Thus, V_{out} is twice V_{peak}. Therefore, this topology is more stable and efficient than the half-wave rectifier. There is also a bridge rectifier that rectifies both positive and negative cycles of the AC input but retains $V_{out} = V_{peak}$ by alternatively blocking pairs of diodes D1, D4 and D2, D3 (Fig. 6c).

Voltage multiplier is a special type of rectifier circuit that converts and boosts AC input to DC output. In some case where the rectified power is inadequate for the application, there is a need for boosting the output DC by stacking single rectifiers into series, forming the voltage multiplier [48]. Several configurations of the voltage multiplier are shown in Fig. 7. The most fundamental configuration is the Cockcroft–Walton voltage multiplier (Fig. 7a). This circuit's operational principle is similar to the full-wave rectifier (Fig. 6b) but has more stages for higher voltage gain. The Dickson multiplier in Fig. 7b is a modification of Cockcroft–Walton's configuration with stage capacitors being shunted to reduce parasitic effects. Thus, the Dickson multiplier is preferable for small voltage applications. However, it is challenging to obtain high PCE due to the high threshold voltage among diodes creating leakage current, thus reducing the overall efficiency. Additionally, for high resistance loads, output voltage drops drastically leading to low current supply to the load. A summary of recent works related to voltage multiplier is shown in Table 3.

The MOSFET (metal–oxide–semiconductor field effect transistor) technology is overcoming the limitations of diodes and becoming an alternate solution for rectifying and boosting. Owing to MOSFET technology, Dickson multiplier can be integrated together in integrate circuits

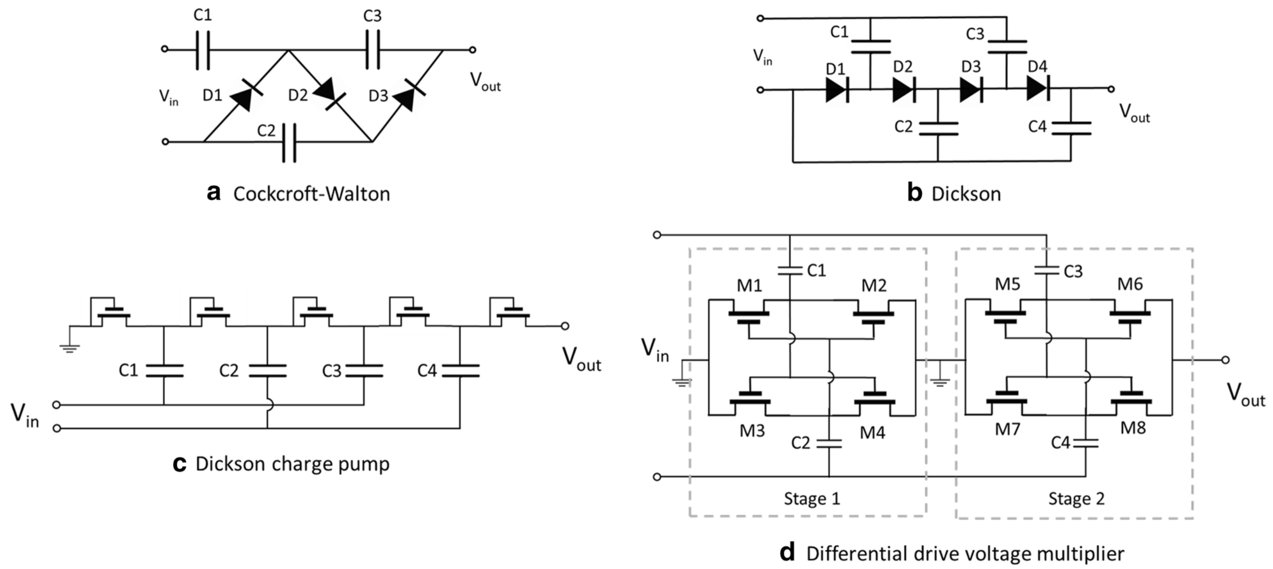

Fig. 7 Common voltage multiplier configurations: **a** Three stages Cockcroft–Walton voltage multiplier, **b** Four stages Dickson voltage multiplier, **c** Four stage Dickson voltage multiplier using CMOS technology, **d** Two stages voltage multiplier comprised of differential drive unit

Table 3 Published work comparison regarding voltage multiplier

Ref.	Rectifier topology	No. of stage	Freq. (MHz)	Tech.	Range (dBm)	Maximum PCE	Size
[51]	Differential	2	433	0.18 µm CMOS	0–20	74% @ 2 dBm	–
[52]	PMOS transistors	7	900	40 nm CMOS	–	44% @ > 390 mV	0.04 mm^2
[53]	Half-wave	4	900	0.18 µm CMOS	–	37.42% @ 390 mV	–
[54]	Comparator-based/active-diode	3	13.56	0.18 µm CMOS	8–15	67.9% @ 12.8 dBm	–
[55]	Dickson	3	13.56	250 nm CMOS	–	72%	0.13 mm^2
[56]	–	–	915	0.13 µm CMOS	−21.6	22.6% @ −16.8 dBm	0.186 mm^2
[57]	–	–	900, 2400	130 nm CMOS	–	41% @ −20.6 dBm	–

(IC) by replacing diodes with NMOS as shown in Fig. 7c. Relatively low threshold voltages and high PCEs are features of this design. Moreover, differential drive voltage multiplier (Fig. 7d) is widely used because of its low leakage current and potential for further modification for specific applications. A detail explanation and analysis is presented in [49, 50].

There is a strong relationship between the number of stages of a voltage multiplier and its sensitivity and efficiency. As the number of stages increase, there is more loss across each added stage. However, the tradeoff is higher voltage multiplication and small threshold voltage at the first stage. On the other hand, a voltage multiplier with a few stages has less voltage drop between its stages but requires higher threshold voltage for all stages to work simultaneously. For this reason, a voltage multiplier becomes more sensitive when a large number of stages are present and becomes more efficient at fewer stages. This tradeoff feature was analyzed in several researches [53, 58]. Therefore, the optimal number of stages should be considered depending upon the application targets.

Pros and cons of Schottky diode and CMOS technology

Rectifying elements such as diodes or transistors are indispensable components of the rectifier/voltage multiplier block, which determine operation frequency and power-conversion efficiency. Traditionally, a Schottky diode was used because of its low threshold voltage. Moreover, there are also Esaki (tunnel) diodes, metal–insulator–metal (MIM) diodes, and spin diodes with recent technology enhancements that make it more mature. Particularly, the Esaki diode can operate at very high frequency with fast response because of its low parasitic elements. Moreover, the MIM diodes technology makes integration with CMOS process possible, which is a limitation of Schottky diodes. Spin diodes offer lower threshold voltages than Schottky diodes.

There are several parameters that need to be considered when choosing a diode for designing a voltage multiplier. These include RF to DC power conversion efficiency, parasitic efficiency, and matching efficiency. When it comes to microwatt applications, Schottky diodes face some limitation in RF-DC conversion efficiency because of their high zero-bias junction resistance [59]. However, owing to recent developments in technology, if the retrieved power is less than −30 dBm (1 μW), then the low-barrier SMS7630 and VDI W-Band ZBD diodes are recommended due to their low parasitic power losses and low power dissipation in the matching network [60].

Recently, many publications have moved toward CMOS processes, since they not only assist in custom designing in electronics but are also more sensitive to low operation voltages than traditional Schottky diodes. Many researches focus on modifying present rectifier and voltage multiplier topologies to achieve higher gain, sensitivity, efficiency [52, 55–57, 61]. For instance, the voltage multiplier achieved maximum 11% PCE at −24 dBm (4 μW) in [62] and 41% PCE at −20.6 dBm (8.7 μW) input power in [57]. While the majority of publications utilized 0.18 μm CMOS technology in their work, [52] applied commercial 40 nm CMOS process to build a low-voltage operation voltage multiplier that reached 44% PCE at 390 mV input. Hwang et al. [54] applied reducing reverse loss to minimize the reverse leakage loss in the rectifier, thus increasing the overall efficiency.

Similar to Schottky diodes, the number of stages directly dictates the performance efficiency in the task of designing a custom CMOS voltage multiplier. A basic assumption when using transistors is that the retrieved RF signal from the antenna must be high enough to trigger the transistors on and off. Different voltage multiplier topologies possess different features and characteristics. However, as long as transistors are used in the design, there is considerable amount of voltage drop among transistors. It has been observed that the major cause of rectified voltage drop is the threshold loss. If the number of stages increases, the threshold voltage at the transistors also adds up, requiring the higher amplitude RF signal to trigger the transistors [62]. Therefore, the number of stages in the majority of publications usually does not exceed 10.

Designing and optimizing the WPH system

The efficiencies of individual modules and the integration of all modules together constitute the total efficiency of a WPH system. As a consequence, the only way to optimize the total efficiency is to maximize efficiencies of each module and consolidate them in harmony. Commonly, several tests and adjustments need to be done to obtain an optimized WPH system. The work flow of designing a WPH system is described in Fig. 8.

Firstly, choosing the appropriate operation frequency to the application's requirement critically dictates the efficiency of the WPH system. Secondly, the operation range also need to be specified. Long range (far-field) power harvesting or transmission requires high frequency such as 2.45, 5.24 GHz, whereas short range (near-field) applications requires electromagnetic waves at frequencies in the megahertz range. Additionally, for power harvesting in dense environment different from air, very low frequency (~kHz) is preferable. Besides the operation frequency and distance, the required output power and voltage determines the suitable topology of the voltage multiplier.

Based on the operation and application, the antenna's design has to match the gain, frequency and size. Choosing the right rectifying element is also critical. For rectifier and voltage multiplier modules in general, the dependence of RF-DC PCE on frequency and output load was discussed in [50]. In addition, it is obligated to have a correct-tuned IMN for maximum power transfer inside the WPH circuit.

Detailed features of individual modules of the WPH system are explained and discussed in the previous sections. The problem of optimizing the WPH system becomes the game of defining the goals and integrating individual modules in harmony to archive the goal.

Applications

RF power harvesting in medical and healthcare

In order to deploy WPH in the real world, power consumption rate of the device should be less than the harvested power. Since retrieved power is unstable and difficult to predict, an energy storage module is highly recommended to enhance its consistency. If the WPH is

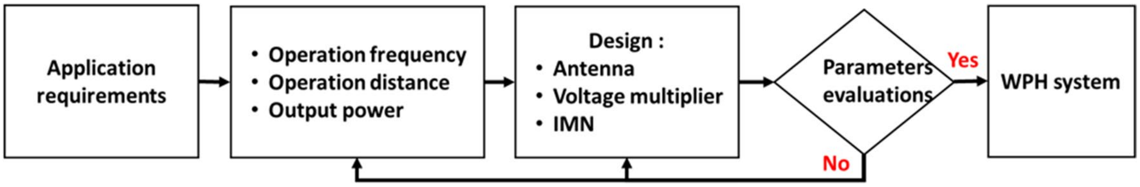

Fig. 8 WPH system's designing protocol

integrated to the device as a single system on chip, the total size of the system would be reduced significantly.

The first wireless battery-free bio-signal processing system on chip [63] was introduced by [64]. This system was able to monitor various bio-signals via electrocardiogram (ECG), electromyogram (EMG), and electroencephalogram (EEG). The total size of the chip was 8.25 mm^2. This chip only consumed 19 µW to measure heart rate. The power module of this system comprised an RF WPH that was supported by a thermoelectric generator, which together supplied a voltage of 1.35 V. Table 4 presents the performance of several health monitoring system on chips (SoCs). The majority of chips have the advantages of small size, low voltage supply and low power consumption which is applicable to be integrated with WPH technology.

It is a fact that a considerable amount of TV broadcast signals never reach the TV. Fortunately, WPH technology can harness this wasted energy without causing any effect on the TV broadcasting quality. In this context, Nishimoto et al. [71] have explored this potential source of energy to remove batteries from temperature monitoring WSNs. Since the power density of TV broadcasts is relatively weak, an array antenna was designed to collect multi-frequency waves. Compared to the antenna array designed in [72], the antenna rectifying efficiency obtained was higher by 50% in the frequencies between 15 and 800 MHz. This design is reported to successfully measure and transmit data every 5 s.

In some research experiments where free movement is unavoidable, stimulation signal and response data has to be delivered and collected wirelessly. In these circumstances, RF WPH is a superior choice in comparison with inductive coupling transmission technology to power the wireless system. Cobo et al. designed an implantable micro-pump that is powered by inductive coupling coils. Even though this work has obtained high efficiency power transmission, the operation range is limited to a few centimeters from the source. On the other hand, in [73], the RF electromagnetic waves were deployed to power optogenetic microneedles to be wirelessly injected into mice. The design of WPH in this work was able to retrieve 4.08 mW power at a distance of 1 m from the 7.9 W source at a frequency of 910 MHz. Another rectenna was built for mounting on the mice' heads for deep brain stimulation (DBS) devices [31]. In this system, the power-harvesting module retrieves 0.254 mW DC power at a maximum operation distance of 20 cm. With this available power, the DBS device generates 200 µA pulses at 130 Hz. The above results show that RF WPH technology offers a better operation range than traditional inductive coupling power transmission technology. It is now possible to harvest sufficient power to run low-power consumption devices in the range of meters. The noninvasive bio-telemetry system for monitoring and investigating various health indexes is a promising field for WPH technology. Cheng et al. [74] introduced a power harvesting design that included a loop antenna to be attached to the pig eye with 31% PCE and output power up to 2.10 mW. Additionally, in vivo experiments, using WPH technology to supply medical devices for implantation in rapid eyes was demonstrated without recorded side effects [63, 75].

Wireless power harvesting network (IoT/WSN)

Along with the development of the micro-electromechanical systems (MEMS) technology, wireless sensor networks have gained widespread popularity in recent years. This field has achieved several milestones and is still developing further. WSNs applications widespread from smart house, healthcare to industry and military. In addition to the issues of sensing reliability, communication protocol, and network services, energy is a critical concern in WSNs [76]. Sensor nodes of WSNs are low power devices, and minimization of energy consumption and application of WPH to replenish power storage elements are two concurrent schemes to maximize the network lifespan. Recently, several researches were done to

Table 4 Power consumption of published SoC devices

Ref.	Sensors	Operation voltage (V)	Chip power (µW)	Size (mm)
[21]	Intraocular pressure (IOP)	–	0.0053	0.5 × 1.5 × 2.5
[65]	ECG, Bio-impedance	1.2	345–374	7 × 7
[66]	ECG	1.2	32	5 × 4.7
[64]	ECG, EMG, EEG	1.2	19	2.5 × 3.3
[67]	ECG, thoracic impedance variance (TIV)	1.2	2400	5 × 5
[68]	EEG	1	77.1	2.5 × 2.5
[69]	Temperature, pressure	0.4/0.5	7.7	2.2 × 2.7
[70]	Neural, ECG, EMG, EEG	1	500	2.5 × 1

find optimized approaches to power WSNs including RF power harvesting [77–81].

Energy is the greatest challenge in deploying WSNs [82]. Applying RF energy to recharge the batteries is one approach to enhance the lifespan of WSNs [83, 84]. Lee et al. [85] is an example of an RF charger utilizing WPH technology. The system was able to work with a minimum input power of −10 dBm (0.1 mW) in the frequency range from 1.96 to 1.98 GHz to charge a 5 V super-capacitor. Maximum RF to DC power conversion efficiency is 81% at 6 dBm (3.98 mW). Gudan et al. [86] also designed a WPH system to charge a NiMH battery using either ambient Wi-Fi or Bluetooth signals. With the sensitivity down to −20 dBm, this system was capable of charging a battery to 5.8 μJ after 1 h.

Table 5 lists some low-power consumption sensor applications that can be integrated with a WPH module. The majority of these sensors were powered by battery with limited lifetime. Once WPH technology is implemented with these applications, it could be used in as a wireless sensor node. In particular, the work presented in [81] utilized RF power harvesting into WSNs for building's structure monitoring. In this work, far-field RF powered wireless sensors were developed with novel data transmission approach to detect humidity, temperature and light inside a building. Results show at a distance of 1 m from 3 W source, the sensor node received 3.14 mW power through air, 1.53 mW of power through 2 inches of brick, 2.88 mW through wood and 0.7 mW through 2 inches of steel. This amount of power is sufficient for the sensors presented in [88–91]. This result indicates the applicability of RF WPH toward WSNs.

Furthermore, placement of power transmitters and sensor nodes in WSNs also play a critical role in the efficiency of the system [94–96]. To enhance directivity of RF transmissions, Shao et al. [97] proposed a beam-steered antenna array to direct its main beam toward energy harvesting units. By controlling the length of transmission line to each antenna unit in an array, the phase radiating difference between units created shift in the main beam. By this way, the harvesting unit remains fed with energy even it was misaligned with the source by 35°. Using the harvesting unit presented in [98] as a demonstration, within 1 min, this system was able to harvest 68.7 μJ at 0° and 13.6 μJ at 35° at a distance of 2 m from the beam-steered source antenna array.

Conclusion

This paper summarized the state-of-the-art of RF power harvesting technology in recent years. This technology will play a key role in replacing batteries in the near future. Some applications of RF power harvesting have been practically realized. A basic RF power-harvesting unit includes three main modules: the antenna, IMN, and voltage multiplier. The total efficiency of the system is dictated by the harmony in the integration of all the modules. The designs and principles of each module were also discussed in the paper.

The RF electromagnetic waves are harmless, abundant in space, and is able to penetrate through soft tissues. Those are properties that make RF electromagnetic waves an alternative source of energy to replace batteries in many applications. Particularly, RF power harvesting supports low-power medical and healthcare devices and facilitates the development of WSNs and IoT by providing mobility of use. Additionally, the progress in integrating RF power harvesting circuits into CMOS technology creates a completely wireless SoC.

Besides progressive accomplishment in recent years, there are still a variety of rooms to further optimize the RF power harvesting technology such as increasing operation range, reducing transmission loss, optimizing PCE, and minimizing system dimensions are the targets of RF power harvesting research. Furthermore, research focusing on harsh working environments for RF WPH such as implantation conditions or underwater zones is attracting much attention to extend the capabilities of this technology. A fine manufacturing and packing process with cost-effective production is necessary to make this technology mature.

Table 5 State-of-the-art published sensors for WSNs

Ref.	Sensors	Total power consumption (mW)	Supplied voltage (V)	Power source	Applications
[87]	Pressure	12.69		Battery	Tire pressure monitoring
[88]	Pressure	0.03		Battery	Blood pulses monitoring, small vibration recording
[89]	Temperature	0.0068	1.5-2		RFID, implantable devices
[90]	Temperature	0.065	2.5–5.5		Room temperature monitoring
[91]	Temperature	0.000071	1.2	Battery	Environment monitoring
[92]	Gas	24		DC power source	C_2H_5OH detection
[93]	Gas	2	5		Smart phone gas sensor, air monitor

In short, the RF WPH technology is gradually becoming a reality. Although this technology still faces many issues, overcoming these challenges can lead the power industry into a new era of clean and sustainable energy.

Authors' contributions
LGT carried out literature search and wrote the manuscript. HKC and WTP reviewed/edited the manuscript. All authors read and approved the final manuscript.

Author details
[1] Convergence Institute of Biomedical Engineering and Biomaterials, Seoul National University of Science and Technology, Seoul, South Korea. [2] Department of Electrical and Information Engineering, Seoul National University of Science and Technology, Seoul, South Korea. [3] Department of Mechanical and Automotive Engineering, Seoul National University of Science and Technology, Seoul, South Korea.

Competing interests
The authors declare that they have no competing interests.

Funding
This work was supported by Grant No. 10060065 from the Industrial Source Technology Development Programs of the MOTIE (Ministry Of Trade, Industry and Energy), Korea. This work was also supported by the Radiation Technology R&D program through the National Research Foundation of Korea funded by the Ministry of Science, ICT & Future Planning (NRF-2013M2A2A9043274).

Received: 8 November 2016   Accepted: 17 February 2017
Published online: 28 February 2017

References

1. Brown WC (1996) The history of wireless power transmission. Solar Energy 56:3–21
2. Brown WC (1969) Experiments involving a microwave beam to power and position a helicopter. IEEE Trans Aerosp Electron Syst AES-5:692–702
3. Raghunathan V, Kansal A, Hsu J, Friedman J, Srivastava M (2005) Design considerations for solar energy harvesting wireless embedded systems. In: Proceedings of the 4th international symposium on Information processing in sensor networks, p 64
4. Brunelli D, Benini L, Moser C, Thiele L (2008) An efficient solar energy harvester for wireless sensor nodes. In: 2008 design, automation and test in europe, pp 104–109
5. Abdin Z, Alim MA, Saidur R, Islam MR, Rashmi W, Mekhilef S et al (2013) Solar energy harvesting with the application of nanotechnology. Renew Sustain Energy Rev 26:837–852
6. Ackermann T, Söder L (2000) Wind energy technology and current status: a review. Renew Sustain Energy Rev 4:315–374
7. GM Joselin Herbert, S. Iniyan, E. Sreevalsan, and S. Rajapandian, "A review of wind energy technologies," *Renewable and Sustainable Energy Reviews,* vol. 11, pp. 1117-1145, 8//2007
8. Şahin AD (2004) Progress and recent trends in wind energy. Prog Energy Combust Sci 30:501–543
9. Xin L, Shuang-Hua Y (2010) Thermal energy harvesting for WSNs. In: 2010 IEEE international conference on systems man and cybernetics (SMC), pp 3045–3052
10. Dalola S, Ferrari V, Marioli D (2010) Pyroelectric effect in PZT thick films for thermal energy harvesting in low-power sensors. Procedia Eng 5:685–688
11. Cuadras A, Gasulla M, Ferrari V (2010) Thermal energy harvesting through pyroelectricity. Sens Actuators A Phys 158:132–139
12. Cao X, Chiang WJ, King YC, Lee YK (2007) Electromagnetic energy harvesting circuit with feedforward and feedback DC–DC PWM boost converter for vibration power generator system. IEEE Trans Power Electron 22:679–685
13. Beeby SP, Torah RN, Tudor MJ, Glynne-Jones P, Donnell TO, Saha CR et al (2007) A micro electromagnetic generator for vibration energy harvesting. J Micromech Microeng 17:1257
14. Yang B, Lee C, Xiang W, Xie J, He JH, Kotlanka RK, Low SP, Feng H (2009) Electromagnetic energy harvesting from vibrations of multiple frequencies. J Micromech Microeng 19:035001
15. Beeby SP, Tudor MJ, White NM (2006) Energy harvesting vibration sources for microsystems applications. Meas Sci Technol 17:R175
16. Challa VR, Prasad M, Shi Y, Fisher FT (2008) A vibration energy harvesting device with bidirectional resonance frequency tunability. Smart Mater Struct 17:015035
17. Khaligh A, Zeng P, Zheng C (2010) Kinetic energy harvesting using piezoelectric and electromagnetic technologies—state of the art. IEEE Trans Ind Electron 57:850–860
18. Vullers RJM, van Schaijk R, Doms I, Van Hoof C, Mertens R (2009) Micropower energy harvesting. Solid-State Electron 53:684–693
19. Akhtar F, Rehmani MH (2015) Energy replenishment using renewable and traditional energy resources for sustainable wireless sensor networks: a review. Renew Sustain Energy Rev 45:769–784
20. Yaghjian A (1986) An overview of near-field antenna measurements. IEEE Trans Antennas Propag 34:30–45
21. Chen G, Ghaed H, Haque RU, Wieckowski M, Kim Y, Kim G et al (2011) A cubic-millimeter energy-autonomous wireless intraocular pressure monitor. In: 2011 IEEE international solid-state circuits conference, 2011, pp 310–312
22. Harlow JH (2004) Electric power transformer engineering. CRC Press, Boca Raton
23. Lee TH (2004) The design of cmos radio-frequency integrated circuits. Commun Eng 2:47
24. Song C, Huang Y, Zhou J, Zhang J, Yuan S, Carter P (2015) A high-efficiency broadband rectenna for ambient wireless energy harvesting. IEEE Trans Antennas Propag 63:3486–3495
25. Momenroodaki P, Fernandes RD, Popovi Z (2016) Air-substrate compact high gain rectennas for low RF power harvesting. In: 2016 10th European conference on antennas and propagation (EuCAP), pp 1–4
26. Lu P, Yang XS, Li JL, Wang BZ (2016) Polarization reconfigurable broadband rectenna with tunable matching network for microwave power transmission. IEEE Trans Antennas Propag 64:1136–1141
27. Sun H (2016) An enhanced rectenna using differentially-fed rectifier for wireless power transmission. IEEE Antennas Wirel Propag Lett 15:32–35
28. Sun H, Geyi W (2016) A new rectenna with all-polarization-receiving capability for wireless power transmission. IEEE Antennas Wirel Propag Lett 15:814–817
29. Zhu P, Ma Z, Vandenbosch GAE, Gielen G (2015) 160 GHz harmonic-rejecting antenna with CMOS rectifier for millimeter-wave wireless power transmission. In: 2015 9th European conference on antennas and propagation (EuCAP), pp 1–5
30. Zhang J, Wu ZP, Liu CG, Zhang BH, Zhang B (2015) A double-sided rectenna design for RF energy harvesting. In: 2015 IEEE international wireless symposium (IWS), pp 1–4
31. Hosain MK, Kouzani AZ, Samad MF, Tye SJ (2015) A miniature energy harvesting rectenna for operating a head-mountable deep brain stimulation device. IEEE Access 3:223–234
32. Lu P, Yang XS, Li JL, Wang BZ (2015) A compact frequency reconfigurable rectenna for 5.2- and 5.8-GHz wireless power transmission. IEEE Trans Power Electron 30:6006–6010
33. Matsunaga T, Nishiyama E, Toyoda I (2015) 5.8-GHz stacked differential rectenna suitable for large-scale rectenna arrays with DC connection. IEEE Trans Antennas Propag 63:5944–5949
34. Chou JH, Lin DB, Weng KL, Li HJ (2014) All polarization receiving rectenna with harmonic rejection property for wireless power transmission. IEEE Trans Antennas Propag 62:5242–5249
35. Sun H, Guo Y, He M, Zhong Z (2013) A dual-band rectenna using broadband yagi antenna array for ambient RF power harvesting. IEEE Antennas Wirel Propag Lett 12:918–921
36. Niotaki K, Kim S, Jeong S, Collado A, Georgiadis A, Tentzeris MM (2013) A compact dual-band rectenna using slot-loaded dual band folded dipole antenna. IEEE Antennas Wirel Propag Lett 12:1634–1637
37. Hucheng S, Yong-Xin G, Miao H, Zheng Z (2012) Design of a high-efficiency 2.45-GHz rectenna for low-input-power energy harvesting. IEEE Antennas Wirel Propag Lett 11:929–932

38. Olgun U, Chen CC, Volakis JL (2010) Wireless power harvesting with planar rectennas for 2.45 GHz RFIDs. In: 2010 URSI international symposium on electromagnetic theory (EMTS), pp 329–331
39. Ren YJ, Farooqui MF, Chang K (2007) A compact dual-frequency rectifying antenna with high-orders harmonic-rejection. IEEE Trans Antennas Propag 55:2110–2113
40. Olgun U, Chen CC, Volakis JL (2011) Investigation of rectenna array configurations for enhanced RF power harvesting. IEEE Antennas Wirel Propag Lett 10:262–265
41. Shen S, Murch RD (2016) Impedance matching for compact multiple antenna systems in random RF fields. IEEE Trans Antennas Propag 64:820–825
42. Etor D, Dodd LE, Wood D, Balocco C (2015) Impedance matching at THz frequencies: optimizing power transfer in rectennas. In: 2015 40th international conference on infrared, millimeter, and terahertz waves (IRMMW-THz), pp 1–2
43. Hoarau C, Corrao N, Arnould JD, Ferrari P, Xavier P (2008) Complete design and measurement methodology for a tunable RF impedance-matching network. IEEE Trans Microw Theory Tech 56:2620–2627
44. Marrocco G (2008) The art of UHF RFID antenna design: impedance-matching and size-reduction techniques. IEEE Antennas Propag Mag 50:66–79
45. Mingo JD, Valdovinos A, Crespo A, Navarro D, Garcia P (2004) An RF electronically controlled impedance tuning network design and its application to an antenna input impedance automatic matching system. IEEE Trans Microw Theory Tech 52:489–497
46. Hatay M (1980) Empirical formula for propagation loss in land mobile radio services. IEEE Trans Veh Technol 29:317–325
47. Radiom S, Vandenbosch G, Gielen G (2008) Impact of antenna type and scaling on scavenged voltage in passive RFID tags. In: International workshop on antenna technology: small antennas and novel metamaterials, 2008. iWAT 2008, pp 442–445
48. Gosset G, Flandre D (2011) Fully-automated and portable design methodology for optimal sizing of energy-efficient CMOS voltage rectifiers. IEEE J Emerg Sel Top Circuits Syst 1:141–149
49. Facen A, Boni A (2007) CMOS power retriever for UHF RFID tags. Electron Lett 43:1424
50. Kotani K, Sasaki A, Ito T (2009) High-efficiency differential-drive CMOS rectifier for UHF RFIDs. IEEE J Solid-State Circuits 44:3011–3018
51. Chouhan SS, Nurmi M, Halonen K (2016) Efficiency enhanced voltage multiplier circuit for RF energy harvesting. Microelectron J 48:95–102
52. Wang W, Xiangjie C, Wong H (2015) Analysis and design of CMOS full-wave rectifying charge pump for RF energy harvesting applications. In: 2015 IEEE Region 10 conference TENCON 2015, pp 1–4
53. Rodriguez AN, Cruz FRG, Ramos RZ (2015) Design of 900 Mhz AC to DC converter using native Cmos device of TSMC 0.18 micron technology for RF energy harvest application. Univers J Electr Electron Eng 3:7
54. Hwang YS, Lei CC, Yang YW, Chen JJ, Yu CC (2014) A 13.56-MHz low-voltage and low-control-loss RF-DC rectifier utilizing a reducing reverse loss technique. IEEE Trans Power Electron 29:6544–6554
55. Haddad PA, Gosset G, Raskin JP, Flandre D (2014) Efficient ultra low power rectification at 13.56 MHz for a 10 μA load current. In: 2014 SOI-3D-subthreshold microelectronics technology unified conference (S3S), pp 1–2
56. Hameed Z, Moez K (2014) Hybird forward and backward threshold-compensated RF-DC power converter for RF energy harvesting. IEEE J Eng Sel Top Circuits Syst 4:9
57. Karolak D, Taris T, Deval Y, Béguéret JB et al (2012) Design comparison of low-power rectifiers dedicated to RF energy harvesting. In: 2012 19th IEEE international conference on electronics, circuits and systems (ICECS), pp 524–527
58. Kadupitiya JCS, Abeythunga TN, Ranathunga PDMT, De Silva DS (2015) Optimizing RF energy harvester design for low power applications by integrating multi stage voltage doubler on patch antenna. In: 2015 8th international conference on Ubi-Media computing (UMEDIA), pp 335–338
59. Hemour S, Zhao Y, Lorenz CHP, Houssameddine D, Gui Y, Hu CM et al (2014) Towards low-power high-efficiency RF and microwave energy harvesting. IEEE Trans Microw Theory Tech 62:965–976
60. Lorenz CHP, Hemour S, Wu K (2016) Physical mechanism and theoretical foundation of ambient RF power harvesting using zero-bias diodes. IEEE Trans Microw Theory Tech 64:2146–2158
61. Sun H, Xu G (2015) A differentially-driven rectifier for enhanced RF power harvesting. In: 2015 IEEE MTT-S international microwave workshop series on advanced materials and processes for RF and THz applications (IMWS-AMP), pp 1–3
62. Papotto G, Carrara F, Palmisano G (2011) A 90-nm CMOS threshold-compensated RF energy harvester. IEEE J Solid-State Circuits 46:1985–1997
63. Lingley AR, Ali M, Liao Y, Mirjalili R, Klonner M, Sopanen M et al (2011) A single-pixel wireless contact lens display. J Micromech Microeng 21:125014
64. Zhang Y, Zhang F, Shakhsheer Y, Silver JD, Klinefelter A, Nagaraju M et al (2013) A batteryless 19 W MICS/ISM-band energy harvesting body sensor node SoC for ExG applications. IEEE J Solid-State Circuits 48:199–213
65. Helleputte NV, Konijnenburg M, Pettine J, Jee DW, Kim H, Morgado A et al (2015) A 345 uW multi-sensor biomedical SoC with bio-impedance, 3-channel ECG, motion artifact reduction, and integrated DSP. IEEE J Solid-State Circuits 50:230–244
66. Kim H, Kim S, Helleputte NV, Artes A, Konijnenburg M, Huisken J et al (2014) A configurable and low-power mixed signal SoC for portable ECG monitoring applications. IEEE Trans Biomed Circuits Syst 8:257–267
67. Yan L, Bae J, Lee S, Roh T, Song K, Yoo HJ (2011) A 3.9 mW 25-electrode reconfigured sensor for wearable cardiac monitoring system. IEEE J Solid-State Circuits 46:353–364
68. Verma N, Shoeb A, Bohorquez J, Dawson J, Guttag J, Chandrakasan AP (2010) A micro-power EEG acquisition SoC with integrated feature extraction processor for a chronic seizure detection system. IEEE J Solid-State Circuits 45(4):804–816
69. Chen G, Fojtik M, Kim D, Fick D, Park J, Seok M et al. (2010) Millimeter-scale nearly perpetual sensor system with stacked battery and solar cells. In: 2010 IEEE international solid-state circuits conference—(ISSCC), pp 288–289
70. Rai S, Holleman J, Pandey JN, Zhang F, Otis B (2009) A 500 μW neural tag with 2 μVrms AFE and frequency-multiplying MICS/ISM FSK transmitter. In: 2009 IEEE international solid-state circuits conference—digest of technical papers, pp 212–213
71. Nishimoto H, Kawahara Y, Asami T (2010) Prototype implementation of ambient RF energy harvesting wireless sensor networks. In: Sensors, 2010 IEEE, pp 1282–1287
72. Shinohara N, Kawasaki S (2009) Recent wireless power transmission technologies in Japan for space solar power station/satellite. In: 2009 IEEE radio and wireless symposium, pp 13–15
73. Kim T-I, McCall JG, Jung YH, Huang X, Siuda ER, Li Y et al (2013) Injectable, cellular-scale optoelectronics with applications for wireless optogenetics. Science 340:211–216
74. Cheng HW, Yu TC, Huang HY, Ting SH, Huang TH, Chiou JC et al (2014) Design of miniaturized antenna and power harvester circuit on the enucleated porcine eyes. IEEE Antennas Wirel Propag Lett 13:1156–1159
75. Chow EY, Yang CL, Ouyang Y, Chlebowski AL, Irazoqui PP, Chappell WJ (2011) Wireless powering and the study of RF propagation through ocular tissue for development of implantable sensors. IEEE Trans Antennas Propag 59:2379–2387
76. Yick J, Mukherjee B, Ghosal D (2008) Wireless sensor network survey. Comput Netw 52:2292–2330
77. Correia R, Carvalho NB, Kawasaki S (2016) Continuously power delivering for passive backscatter wireless sensor networks. IEEE Trans Microw Theory Tech 64:3723–3731
78. Praveen MP, Mehta NB (2016) Trade-offs in analog sensing and communication in RF energy harvesting wireless sensor networks. In: 2016 IEEE international conference on communications (ICC), pp 1–6
79. Collado A, Georgiadis A (2014) Optimal waveforms for efficient wireless power transmission. IEEE Microwave Wirel Compon Lett 24:354–356
80. Zhao Y, Chen B, Zhang R (2013) Optimal power allocation for an energy harvesting estimation system. In: 2013 IEEE international conference on acoustics, speech and signal processing, pp 4549–4553
81. Ruisi G, Hong P, Zhibin L, Na G, Jinhui W, Xiaowei C (2016) RF-powered battery-less wireless sensor network in structural monitoring. In: 2016 IEEE international conference on electro information technology (EIT), pp 0547–0552
82. Seah WKG, Eu ZA, Tan HP (2009) Wireless sensor networks powered by ambient energy harvesting (WSN-HEAP)—survey and challenges. In: 2009 1st international conference on wireless communication, vehicular

technology, information theory and aerospace and electronic systems technology, pp 1–5
83. Jabbar H, Song YS, Jeong TT (2010) RF energy harvesting system and circuits for charging of mobile devices. IEEE Trans Consum Electron 56:247–253
84. Che W, Chen W, Meng D, Wang X, Tan X, Yan N et al (2010) Power management unit for battery assisted passive RFID tag. Electron Lett 46:589–590
85. Lee JH, Jung WJ, Jung JW, Jang JE, Park JS (2015) A matched RF charger for wireless RF power harvesting system. Microw Opt Technol Lett 57:1622–1625
86. Gudan K, Chemishkian S, Hull JJ, Thomas SJ, Ensworth J, Reynolds MS (2014) A 2.4 GHz ambient RF energy harvesting system with −20 dBm minimum input power and NiMH battery storage. In: RFID technology and applications conference (RFID-TA), 2014 IEEE, pp 7–12
87. Nagaraju MB, Lingley AR, Sridharan S, Gu J, Ruby R, Otis BP (2015) A 0.8 mm^3 ± 0.68 psi single-chip wireless pressure sensor for TPMS applications. In: 2015 IEEE international solid-state circuits conference—(ISSCC) digest of technical papers, pp 1–3
88. Gong S, Schwalb W, Wang Y, Chen Y, Tang Y, Si J et al (2014) A wearable and highly sensitive pressure sensor with ultrathin gold nanowires. Nat Commun 5:3132
89. Souri K, Chae Y, Makinwa KAA (2013) A CMOS temperature sensor with a voltage-calibrated inaccuracy of ±0.15 °C (3σ) from −55 °C to 125 °C. IEEE J Solid-State Circuits 48:292–301
90. Aita AL, Pertijs MAP, Makinwa KAA, Huijsing JH, Meijer GCM (2013) Low-power CMOS smart temperature sensor with a batch-calibrated inaccuracy of ±0.25 °C (±3σ) from −70 °C to 130 °C. IEEE Sens J 13:1840–1848
91. Jeong S, Foo Z, Lee Y, Sim JY, Blaauw D, Sylvester D (2014) A fully-integrated 71 nW CMOS temperature sensor for low power wireless sensor nodes. IEEE J Solid-State Circuits 49:1682–1693
92. Moon SE, Lee HK, Choi NJ, Kang HT, Lee J, Ahn SD et al (2015) Low power consumption micro C2H5OH gas sensor based on micro-heater and ink jetting technique. Sens Actuators B Chem 217:146–150
93. Zhou Q, Sussman A, Chang J, Dong J, Zettl A, Mickelson W (2015) Fast response integrated MEMS microheaters for ultra low power gas detection. Sens Actuators A Phys 223:67–75
94. Erol-Kantarci M, Mouftah HT (2012) Mission-aware placement of RF-based power transmitters in wireless sensor networks. In: 2012 IEEE symposium on computers and communications (ISCC), pp 000012–000017
95. He S, Chen J, Jiang F, Yau DKY, Xing G, Sun Y (2013) Energy provisioning in wireless rechargeable sensor networks. IEEE Trans Mob Comput 12:1931–1942
96. Li Y, Fu L, Chen M, Chi K, Zhu YH (2015) RF-based charger placement for duty cycle guarantee in battery-free sensor networks. IEEE Commun Lett 19:1802–1805
97. Shao S, Gudan K, Hull JJ (2016) A mechanically beam-steered phased array antenna for power-harvesting applications [Antenna Applications Corner]. IEEE Antennas Propag Mag 58:58–64
98. Gudan K, Shao S, Hull JJ, Ensworth J, Reynolds MS (2015) Ultra-low power 2.4 GHz RF energy harvesting and storage system with −25 dBm sensitivity. In: 2015 IEEE international conference on RFID (RFID), pp 40–46

Submit your manuscript to a SpringerOpen journal and benefit from:

► Convenient online submission
► Rigorous peer review
► Immediate publication on acceptance
► Open access: articles freely available online
► High visibility within the field
► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com

United States Patent [19]
Meyer

[11] **Patent Number:** **4,826,581**
[45] **Date of Patent:** **May 2, 1989**

[54] CONTROLLED PROCESS FOR THE PRODUCTION OF THERMAL ENERGY FROM GASES AND APPARATUS USEFUL THEREFORE

[76] Inventor: **Stanley A. Meyer**, 3792 Broadway, Grove City, Ohio 43123

[21] Appl. No.: **81,859**

[22] Filed: **Aug. 5, 1987**

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 835,564, Mar. 3, 1986, abandoned.

[51] **Int. Cl.**^{4} **C07G 13/00**
[52] **U.S. Cl.** **204/157.41**; 204/164
[58] **Field of Search** 204/164, 157.41, 157.44

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,233,109	11/1980	Nishizawa	204/164 X
4,406,765	9/1983	Higashi et al.	204/164
4,687,753	8/1987	Fiato et al.	204/157.41 X
4,695,357	9/1987	Boussert	204/157.41

Primary Examiner—Stephen J. Kalafut
Attorney, Agent, or Firm—Porter, Wright, Morris & Arthur

[57] ABSTRACT

A method of and apparatus for obtaining the release of energy from a gas mixture including hydrogen and oxygen in which charged ions are stimulated to an activated state, and then passed through a resonant cavity, where successively increasing energy levels are achieved, and finally passed to an outlet orifice to produce thermal explosive energy.

2 Claims, 7 Drawing Sheets

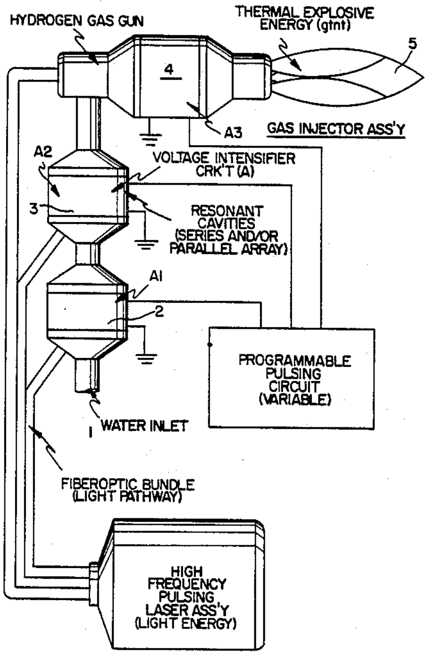
HYDROGEN GAS GUN
THERMAL EXPLOSIVE ENERGY (gtnt)
5
4
A3
GAS INJECTOR ASS'Y
A2
VOLTAGE INTENSIFIER CRK'T (A)
3
RESONANT CAVITIES (SERIES AND/OR PARALLEL ARRAY)
AI
2
PROGRAMMABLE PULSING CIRCUIT (VARIABLE)
WATER INLET
1
FIBEROPTIC BUNDLE (LIGHT PATHWAY)
HIGH FREQUENCY PULSING LASER ASS'Y (LIGHT ENERGY)

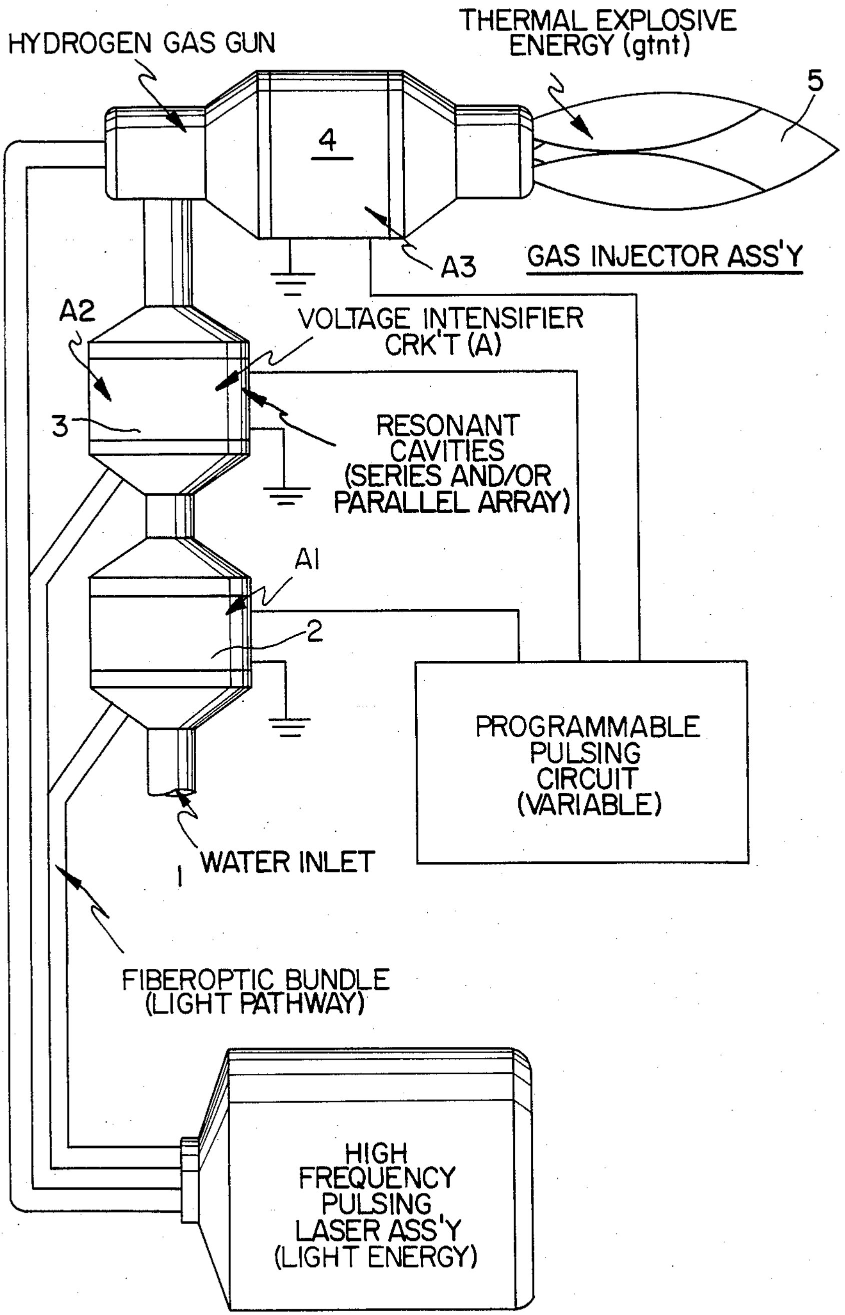
HYDROGEN GAS GUN
THERMAL EXPLOSIVE ENERGY (gtnt)
5
4
A3
GAS INJECTOR ASS'Y
A2
VOLTAGE INTENSIFIER CRK'T (A)
3
RESONANT CAVITIES (SERIES AND/OR PARALLEL ARRAY)
A1
2
PROGRAMMABLE PULSING CIRCUIT (VARIABLE)
WATER INLET
1
FIBEROPTIC BUNDLE (LIGHT PATHWAY)
HIGH FREQUENCY PULSING LASER ASS'Y (LIGHT ENERGY)

FIG. 1

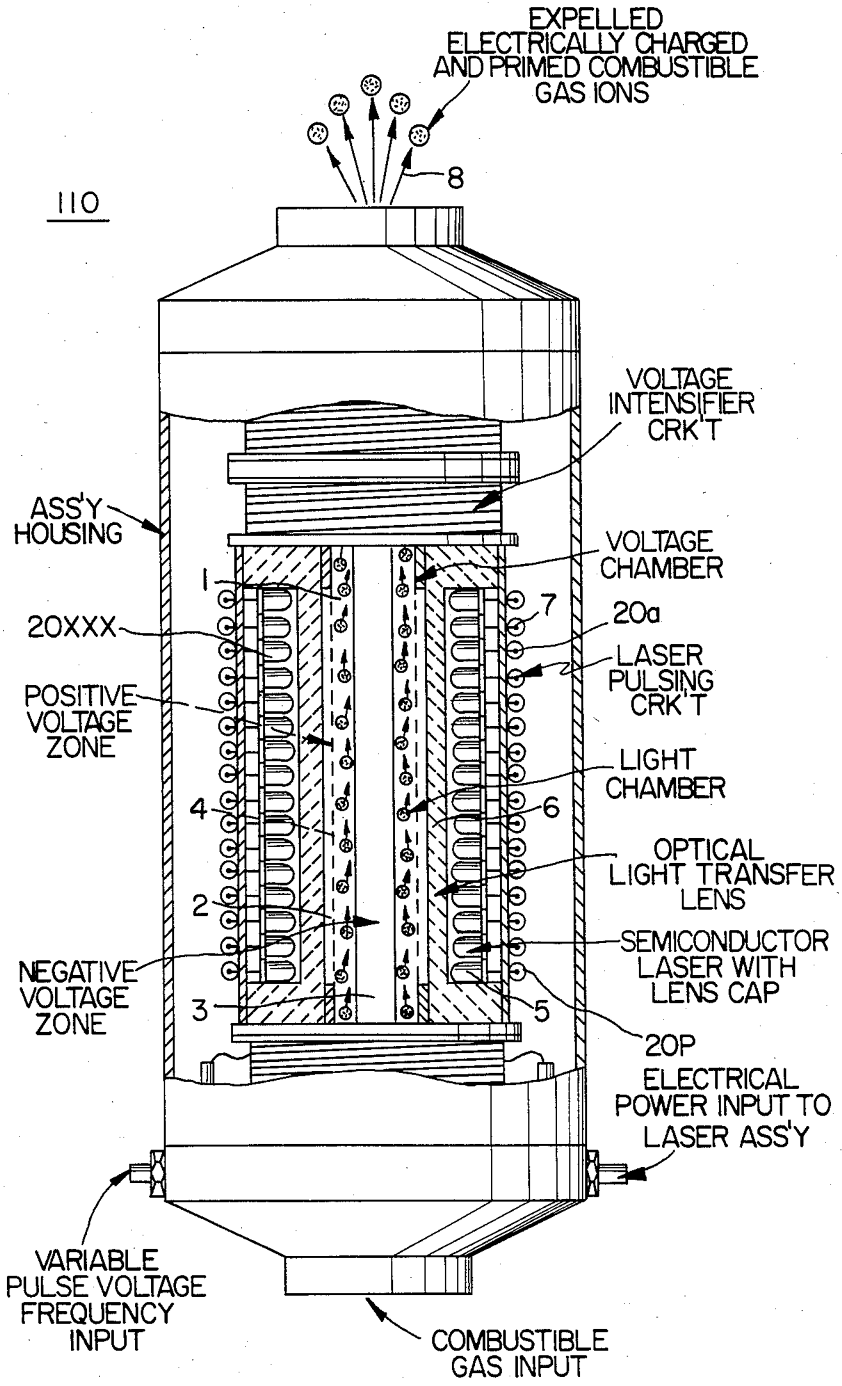
EXPELLED ELECTRICALLY CHARGED AND PRIMED COMBUSTIBLE GAS IONS
8
110
VOLTAGE INTENSIFIER CRK'T
ASS'Y HOUSING
VOLTAGE CHAMBER
1
7
20a
20XXX
LASER PULSING CRK'T
POSITIVE VOLTAGE ZONE
LIGHT CHAMBER
4
6
OPTICAL LIGHT TRANSFER LENS
2
SEMICONDUCTOR LASER WITH LENS CAP
NEGATIVE VOLTAGE ZONE
3
5
20P
ELECTRICAL POWER INPUT TO LASER ASS'Y
VARIABLE PULSE VOLTAGE FREQUENCY INPUT
COMBUSTIBLE GAS INPUT

FIG. 2

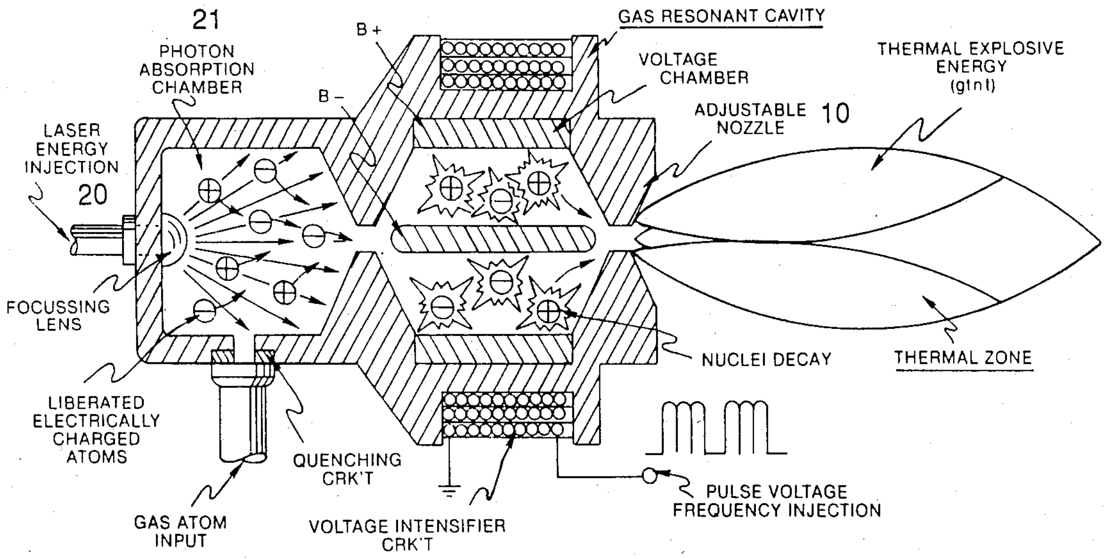
21
PHOTON ABSORPTION CHAMBER
B+
B−
GAS RESONANT CAVITY
VOLTAGE CHAMBER
THERMAL EXPLOSIVE ENERGY (gtnt)
ADJUSTABLE NOZZLE
10
LASER ENERGY INJECTION
20
FOCUSSING LENS
LIBERATED ELECTRICALLY CHARGED ATOMS
GAS ATOM INPUT
QUENCHING CRK'T
VOLTAGE INTENSIFIER CRK'T
NUCLEI DECAY
THERMAL ZONE
PULSE VOLTAGE FREQUENCY INJECTION

FIG. 3

FIG. 4

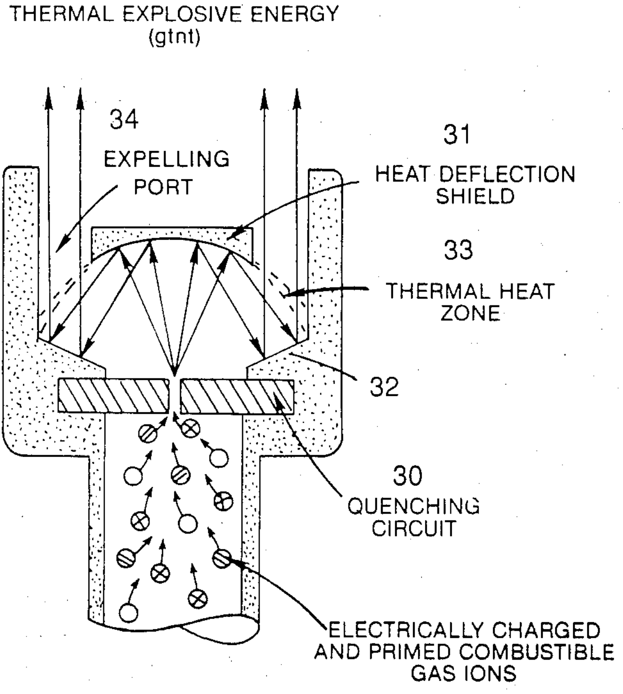
THERMAL EXPLOSIVE ENERGY (gtnt)
34
EXPELLING PORT
31
HEAT DEFLECTION SHIELD
33
THERMAL HEAT ZONE
32
30
QUENCHING CIRCUIT
ELECTRICALLY CHARGED AND PRIMED COMBUSTIBLE GAS IONS

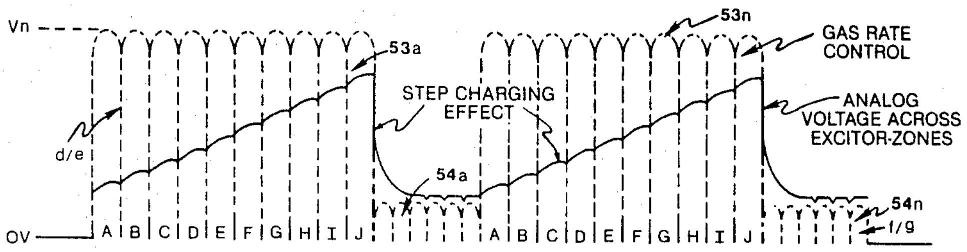

SECTION AA: APPLIED VOLTAGE PULSE TRAIN

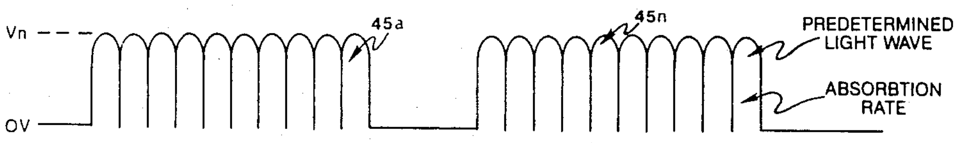

SECTION BB: APPLIED LASER OR PHOTON PULSE TRAIN

FIG. 5

FIG. 6

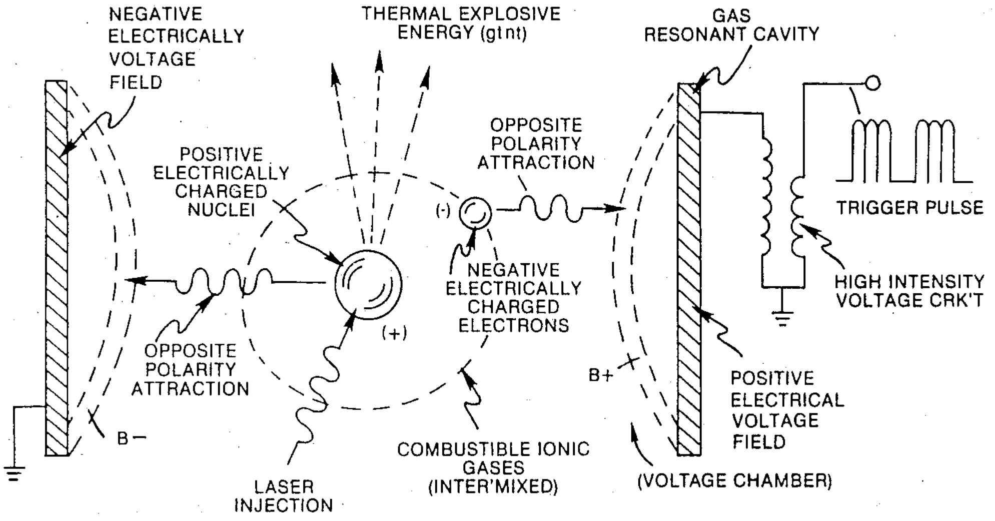
NEGATIVE ELECTRICALLY VOLTAGE FIELD
THERMAL EXPLOSIVE ENERGY (gtnt)
GAS RESONANT CAVITY
POSITIVE ELECTRICALLY CHARGED NUCLEI
OPPOSITE POLARITY ATTRACTION
(-)
(+)
NEGATIVE ELECTRICALLY CHARGED ELECTRONS
TRIGGER PULSE
HIGH INTENSITY VOLTAGE CRK'T
OPPOSITE POLARITY ATTRACTION
B+
POSITIVE ELECTRICAL VOLTAGE FIELD
B−
LASER INJECTION
COMBUSTIBLE IONIC GASES (INTER'MIXED)
(VOLTAGE CHAMBER)

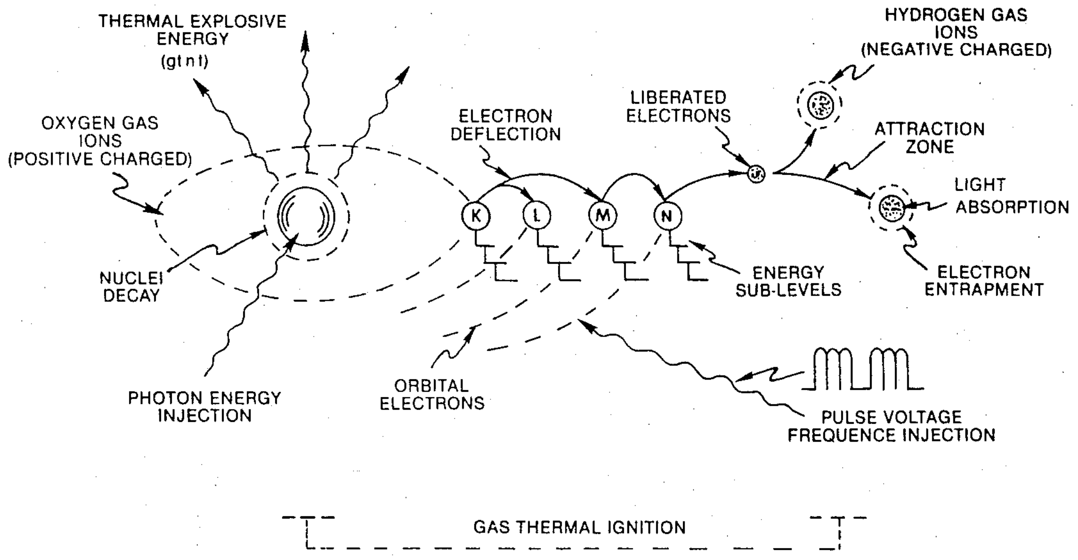
THERMAL EXPLOSIVE ENERGY (gtnt)
OXYGEN GAS IONS (POSITIVE CHARGED)
NUCLEI DECAY
PHOTON ENERGY INJECTION
ELECTRON DEFLECTION
K
L
M
N
ORBITAL ELECTRONS
LIBERATED ELECTRONS
HYDROGEN GAS IONS (NEGATIVE CHARGED)
ATTRACTION ZONE
LIGHT ABSORPTION
ELECTRON ENTRAPMENT
ENERGY SUB-LEVELS
PULSE VOLTAGE FREQUENCE INJECTION
GAS THERMAL IGNITION

FIG. 7

FIG. 8

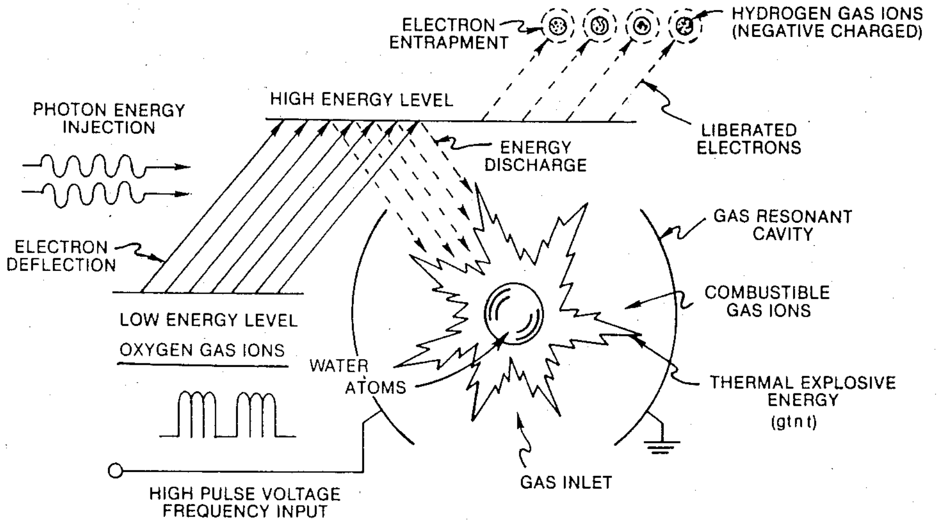
ELECTRON ENTRAPMENT
HYDROGEN GAS IONS (NEGATIVE CHARGED)
PHOTON ENERGY INJECTION
HIGH ENERGY LEVEL
LIBERATED ELECTRONS
ENERGY DISCHARGE
GAS RESONANT CAVITY
ELECTRON DEFLECTION
COMBUSTIBLE GAS IONS
LOW ENERGY LEVEL
OXYGEN GAS IONS
WATER ATOMS
THERMAL EXPLOSIVE ENERGY (gtnt)
GAS INLET
HIGH PULSE VOLTAGE FREQUENCY INPUT

FIG. 9

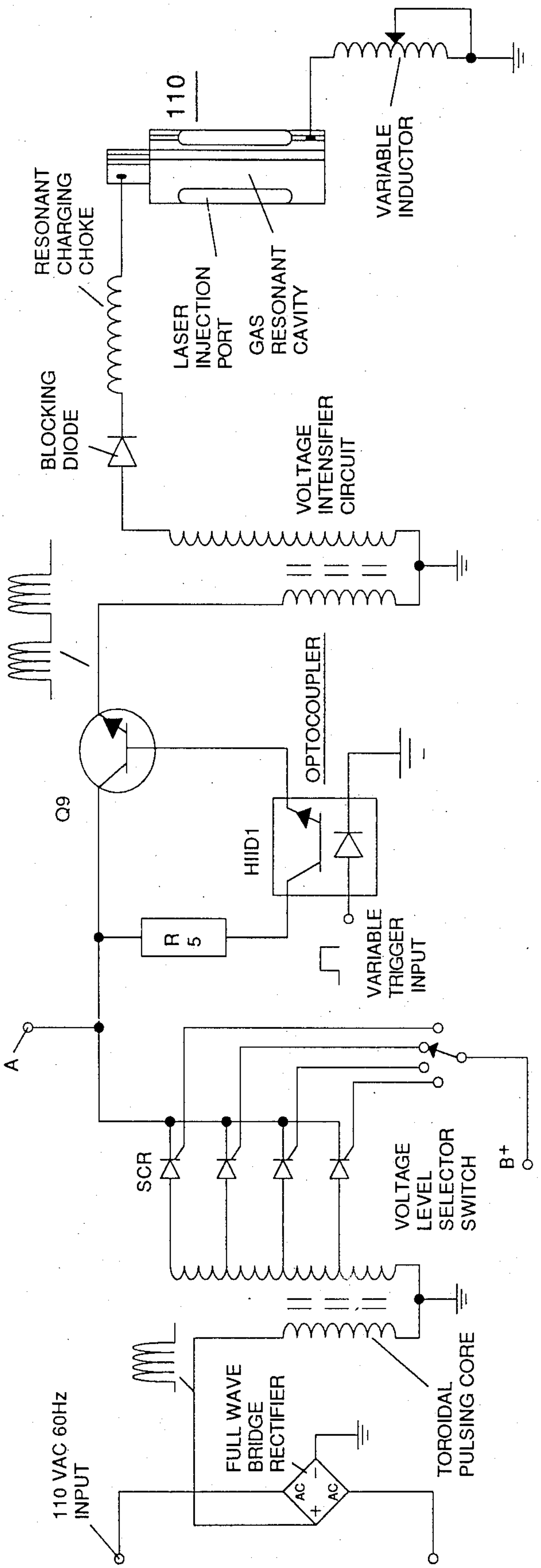
110 VAC 60Hz INPUT
FULL WAVE BRIDGE RECTIFIER
AC
AC
+
−
TOROIDAL PULSING CORE
SCR
VOLTAGE LEVEL SELECTOR SWITCH
B+
A
R 5
VARIABLE TRIGGER INPUT
HIID1
OPTOCOUPLER
Q9
VOLTAGE INTENSIFIER CIRCUIT
BLOCKING DIODE
RESONANT CHARGING CHOKE
LASER INJECTION PORT
GAS RESONANT CAVITY
110
VARIABLE INDUCTOR

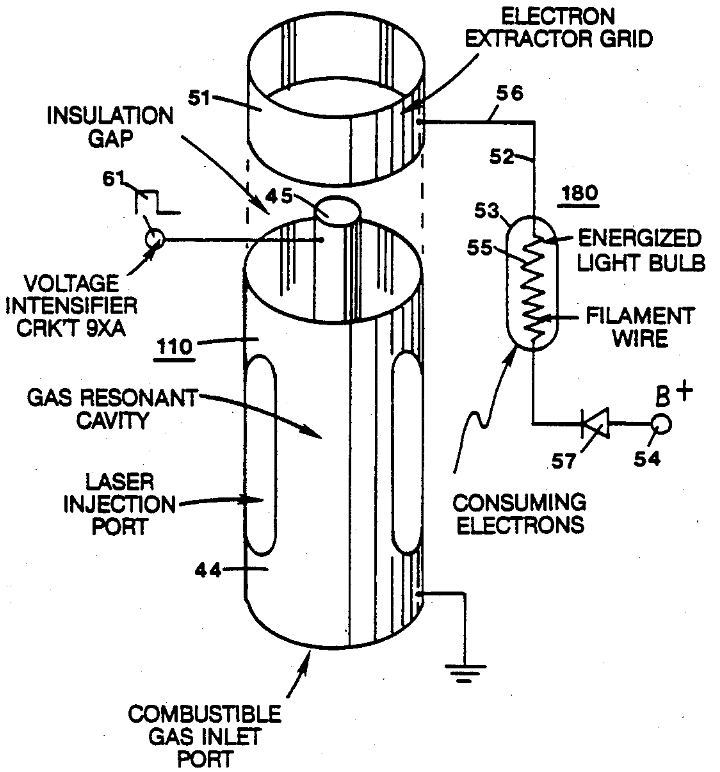
ELECTRON
EXTRACTOR GRID
51
56
INSULATION
GAP
52
180
61
45
53
55
ENERGIZED
LIGHT BULB
VOLTAGE
INTENSIFIER
CRK'T 9XA
FILAMENT
WIRE
110
GAS RESONANT
CAVITY
B+
57
54
LASER
INJECTION
PORT
CONSUMING
ELECTRONS
44
COMBUSTIBLE
GAS INLET
PORT

FIG. 10

FIG. 11

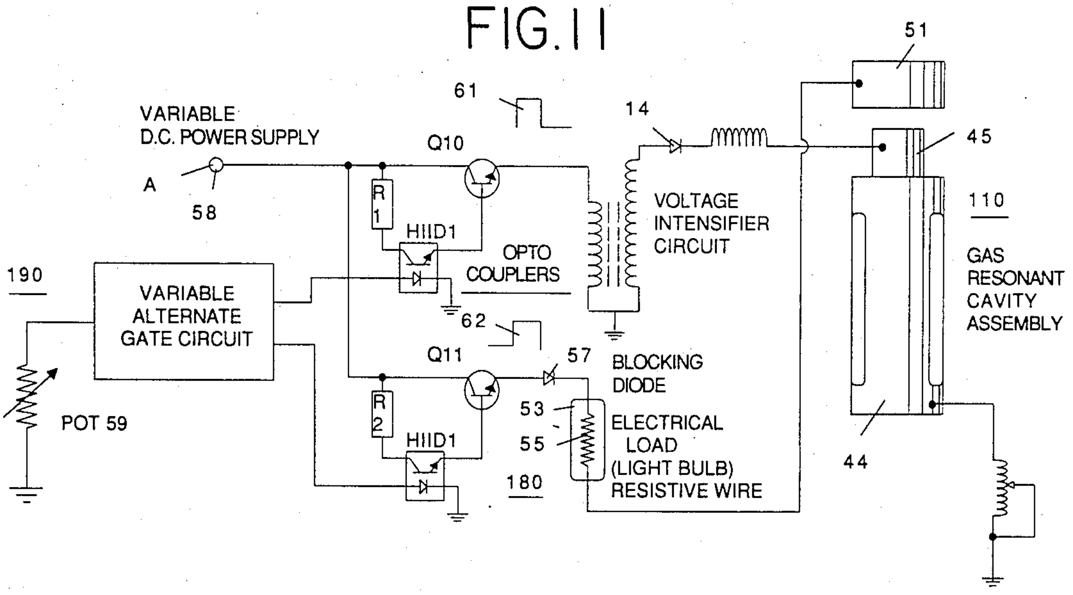
51
61
VARIABLE
D.C. POWER SUPPLY
A
58
Q10
14
45
R
1
HIID1
VOLTAGE
INTENSIFIER
CIRCUIT
110
OPTO
COUPLERS
GAS
RESONANT
CAVITY
ASSEMBLY
190
VARIABLE
ALTERNATE
GATE CIRCUIT
62
Q11
57
BLOCKING
DIODE
POT 59
R
2
53
55
ELECTRICAL
LOAD
(LIGHT BULB)
RESISTIVE WIRE
HIID1
180
44

CONTROLLED PROCESS FOR THE PRODUCTION OF THERMAL ENERGY FROM GASES AND APPARATUS USEFUL THEREFORE

RELATED APPLICATION

This is a continuation-in-part of my co-pending application Ser. No. 835,564, now abandoned.

FIELD OF THE INVENTION

This invention relates to a method of and apparatus for obtaining the release of energy from a gas mixture including hydrogen and oxygen in which charged ions are stimulated to an actived state, and then passed through a resonant cavity, where successively increasing energy levels are achieved, and finally passed to an outlet orifice to produce thermal explosive energy.

BACKGROUND OF THE PRIOR ART

Processes have been proposed for many years in which controlled energy producing reactions of atomic particles are expected to occur under "cold" conditions. [See. e.q.. Rafelski, J. and Jones, S.E., "Cold Nuclear Fusion," *Scientific American,* July, 1987, page 84]. The process and apparatus described herein are considered variations to and improvements in processes by which energy is derived from excited atomic components in a controllable manner.

OBJECTS OF THE INVENTION

It is an object of the invention to realize significant energy-yield from water atoms. Molecules of water are broken down into hydrogen and oxygen gases. Electrically charged gas ions of opposite electrical polarity are activated By Express Mail No. 26224690 on August 5, 1987 by electromagnetic wave energy and exposed to a high temperature thermal zone. Significant amounts of thermal energy with explosive force beyond the gas burning stage are released.

An explosive thermal energy under a controlled state is produced. The process and apparatus provide a heat energy source useful for power generation, aircraft, rocket engines, or space stations.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 illustrates a staged arrangement of apparatus useful in the process, beginning with a water inlet and culminating in the production of thermal explosive energy.

FIG. 2A shows a cross-section of a circular gas resonant cavity used in the final stage assembly of FIG. 1.

FIG. 2B shows an alternative final stage injection system useful in the apparatus of FIG. 1.

FIG. 2C shows an optical thermal lens assembly for use either final stage of FIG. 2A or FIG. 2B.

FIGS. 3A, 3B, 3C and 3D are illustrations depicting various theoretical bases for atomic phenomena expected to occur during operation of the invention herein.

FIG. 4 is an electrical schematic of the voltage source for the gas resonant cavity.

FIGS. 5A and 5B, respectively, show (A) an electron extractor grid used in the injector assemblies of FIG. 2A and FIG. 2B, and (B) the electronic control circuit for the extractor grid.

DESCRIPTION OF THE PREFERRED EMBODIMENT

The hydrogen fracturing process, follows the sequence of steps shown in the following Table I in which beginning with water molecules, the molecule is subjected to successively increasing electrical, wave energy and thermal forces. In the succession of forces, radomly oriented water molecules are aligned with respect to molecular polar orientation and are themselves polarized and "elongated" by the application of an electric potential to the extent that covalent bonding of the water molecule is so weakened that the atoms disassociate and the molecule breaks down into hydrogen and oxygen elemental components. The released atomic gases are next ionized and electrically charged in a vessel while being subjected to a further energy source that promotes inter-particle impact in the gas at an increased overall energy level. Finally, the atomic particles in the excited gas, having achieved successively higher energy levels, are subjected to a laser or electromagnetic wave energy source that produces atomic destabilization and the final release of thermal explosive energy. Engineering design parameters based on known theoretical principles of atomic physics determine the incremental levels of electrical and wave energy input required to produce resonance in each stage of the system. Instead of a dampening effect, a resonant energization of the molecule, atom or ion provides a compounding energy interaction resulting in the final energy release.

TABLE I

PROCESS STEPS LEADING TO IGNITION	
RELATIVE STATE OF WATER MOLECULE AND/OR HYDROGEN/OXYGEN/OTHER ATOMS	
RANDOM	1st Stage
ALIGNMENT	Water to Gas
POLARIZATION	2nd Stage
MOLECULAR ELONGATION	Gas Ionization
ATOM LIBERATION	3rd Stage
LIQUID TO GAS IONIZATION	Priming
ELECTRICAL CHARGING EFFECT	Final Stage
PARTICLE IMPACT	Ignition
ELECTROMAGNETIC WAVE, LASER OR PHOTON INJECTION	
ELECTRON EXTRACTION	
ATOMIC DESTABILIZATION	
THERMAL IGNITION	

After the first stage in which water is broken down into its atomic components in a mixture of hydrogen, oxygen and formerly dissolved entrapped gasses, the gas atoms become elongated during electron removal as the atoms are ionized. Laser, or light wave energy of a predetermined frequency is injected into a containment vessel in a gas ionization process. The light energy absorbed by voltage stimulated gas nuclei causes destabilization of gas ions still further. The absorbed laser energy causes the gas nuclei to increase in energy state, which, in turn, causes electron deflection to a higher orbital shell.

The electrically charged and laser primed combustible gas ions from a gas resonant cavity may be directed into an optical thermal lens assembly for triggering. Before entry into the optimal thermal lens, however, electrons are stripped from the ions and the atom is destabilized. The destabilized gas ions which are electrically and mass unbalanced atoms having highly energized nuclei are pressurized during spark ignition. The unbalanced, destablized atomic components thermally interact; the energized and unstable hydrogen gas nuclei collide with highly energized and unstable oxygen gas nuclei, causing and producing thermal explosive energy beyond the gas burning stage. The ambient air gas components in the initial mixture aid the thermal explosive process under a controlled state.

In the process, the point of optimum energy-yield is reached when the electron deficient oxygen atoms (having less than a normal number of electrons) lock onto and capture a hydrogen atom electron prior to or during thermal combustion of the hydrogen/oxygen mixture. Atomic decay results in the release of energy.

In a general outline of the method, a first gas mixture including at least a portion of hydrogen and oxygen gases is provided. The gas mixture is subjected to a pulsating, polar electric field whereby electrons of the gas atoms are distended in their orbital fields by reason of their subjection to electrical polar forces. The polar pulsating frequency applied is such that the pulsating electric field induces a resonance with respect to an election of the gas atom. A cascade effect results and the energy level of specific resonating electron is increased in cascading, incremental steps.

Next, the gas atoms are ionized and subjected to electro-magnetic wave energy having a predetermined frequency to induce a further election resonance in the ion, whereby the energy level of the election is successively increased. Electrons are extracted from the resonating ions while such ions are in an increased energy state to destabilize the nuclear electron configuration of said ions; and the gas mixture of destabilized ions is thermally ignited.

In the apparatus shown in FIG. **1**, water is introduced at inlet **1** into a first stage water fracturing module **2** in which water molecules are broken down into hydrogen, oxygen and released entrapped gas components by an electrical disassociation process and apparatus such as shown in my co-pending application Ser. No. 835,564, filed March 3, 1986, which is incorporated herein by reference. The released atomic gases and other gas components formerly entrapped as dissolved gases in water may be introduced to a successive stage **3** or other number of like resonant cavities, which are arranged in either a series or parallel combined array. The successive energization of the gas atoms provides a cascading effect, successively increasing the voltage stimulation level of the released gasses as they sequentially pass through cavities **2**, **3**, etc. In a final stage, an injector system **4**, of a configuration of the type shown in FIGS. **2A** or **2B**, receives energized atomic and gas particles where the particles are subjected to further energy input, electrical excitation and thermal stimulation, whereby thermal explosive energy results **5**, which may be directed thru a lens assembly of the type shown in FIG. **2C** to provide a controlled thermal energy output.

Electromagnetic wave activated and electrically charged gas ions of hydrogen and oxygen (of opposite polarity) are expelled from the cascaded cells **2**, **3**, etc. The effect of cascading successively increases the voltage stimulation level of the released gases, which then are directed to the final injector assembly **4**. In the injector assembly, gas ions are stimulated to a yet higher energy level. The gases are continually exposed to a pulsating laser or other electromagnetic wave energy source together with a high intensity oscillating voltage field that occurs within the cell between electrodes or conductive plates of opposite electrical polarity. A preferred construction material for the plates is a stainless steel T-304 which is non-chemically reactive with water, hydrogen, or oxygen. An electrically conductive material which is inert in the fluid environment is a desirable material of construction for the electrical field producing plates, through which field the gas stream of activated particles passes. Gas ions of opposite electrical charges reach and maintain a critical energy level state. The gas ions are oppositely electrically charged and subjected to oscillating voltage fields of opposite polarity and are also subjected to a pulsating electromagnetic wave energy source. Immediately after reaching critical energy, the excited gas ions are exposed to a high temperature thermal zone in the injection cell, **4**, that causes the excited gas ions to undergo gas combustion. The gas ignition triggers atomic decay and releases thermal energy, **5**, with explosive force.

Once triggered, the thermal explosive energy output is controllable by the attenuation of operational parameters. With reference to FIG. **4A**, for example, once the frequency of resonance is identified, by varying applied pulse voltage to the initial water fuel cell assemblies, **2**, **3**, the ultimate explosive energy output is likewise varied. By varying the pulse shape and/or amplitude or pulse train sequence of the electromagnetic wave energy source, final output is varied. Attenuation of the voltage field frequency in the form of *OFF* and *ON* pulses likewise affects output of the staged apparatus. Each control mechanism can be used separately, grouped in sections, or systematically arranged in a sequential manner.

The overall apparatus thus includes means for providing a first gas mixture consisting of at least a portion of hydrogen and oxygen gas. The gases may be obtained by disassociation of the water molecule. An electrical circuit of the type shown in FIG. **4** provides a pulsating, polar electric field to the gas mixture as illustrated in FIG. **3A**, whereby electrons of the gas atoms are distended in their orbital fields by reason of their subjection to electrical polar forces, changing from the state conceptually illustrated by FIG. **3B** to that of FIG. **3C**, at a frequency such that the pulsating electric field induces a resonance with respect to electrons of the gas atoms. The energy level of the resonant electrons is thereby increased in cascading, incremental steps. A further electric field to ionize said gas atoms is applied and an electromagnetic wave energy source for subjecting the ionized gas atoms to wave energy of a predetermined frequency to induce a further electron resonance in the ion, whereby the energy level of the election is successively increased is an additional element of the apparatus as shown in FIG. **3D**.

An electron sink, which may be in the form of the grid element shown in FIG. **5A**, extracts further electrons from the resonating ions while such ions are in an increased energy state and destabilizes the nuclear electron configuration of the ions. The "extraction" of electrons by the sink means is coordinated with the pulsating electrical field of the resonant cavity produced by the circuit of FIG. **4**, by means of an interconnected synchronization circuit, such as shown in FIG. **5B**. A nozzle, **10** in FIG. **2B**, or thermal lens assembly, FIG. **2C**, provides the directing means in which the destabilized ions are finally thermally ignited.

As previously noted, to reach and trigger the ultimate atomic decay of the fuel cell gases at the final stage, sequential steps are taken. First, water molecules are split into component atomic elements (hydrogen and oxygen gases) by a voltage stimulation process called the electrical polarization process which also releases dissolved gases entrapped in the water (See my co-pending application for letters patent, Ser. No. 835, 564, supra). In the injector assembly, a laser produced light wave or other form of coherent electromagnetic wave energy capable of stimulating a resonance within the atomic components is absorbed by the mixture of gases (hydrogen/oxygen/ambient air gases) released by the polarization process. At this point, as shown in FIG. 3B, the individual atoms are subjected to an electric field to begin an ionization process.

The laser or electromagnetic wave energy is absorbed and causes gas atoms to lose electrons and form positively charged gas ions. The energized hydrogen atoms which, as ionized, are positively charged, now accept electrons liberated from the heavier gases and attract other negatively charged gas ions as conceptually illustrated in FIG. 3C. Positively and negatively charged gas ions are re-exposed to further pulsating energy sources to maintain random distribution of ionized atomic gas particles.

The gas ions within the wave energy chamber are subjected to an oscillating high intensity voltage field in a chamber **11** in FIGS. 2A and 2B formed within electrodes **12** and **13** in FIGS. 2A and 2B of opposite electrical polarity to produce a resonant cavity. The gas ions reach a critical energy state at a resonant state.

At this point, within the chamber, additional electrons are attracted to said positive electrode; whereas, positively charged ions or atomic nuclei are attracted to the negative electrode. The positive and negative attraction forces are co-ordinate and operate on said gas ions simultaneously; the attraction forces are non-reversible. The gas ions experience atomic component deflection approaching the point of electron separation. At this point electrons are extracted from the chamber by a grid system such as shown in FIG. 5A. The extracted electrons are consumed and prevented from re-entering the chamber by a circuit such as shown in FIG. 5B. The elongated gas ions are subjected to a thermal heat zone to cause gas ignition, releasing thermal energy with explosive force. During ionic gas combustion, highly energized and stimulated atoms and atom nuclei collide and explode during thermal excitation. The hydrogen fracturing process occurring sustains and maintains a thermal zone, at a temperature in excess of normal hydrogen/oxygen combustion temperature, to wit, in excess of 2500° F. To cause and maintain atomic elongation depicted in FIG. 3C before gas ignition, a voltage intensifier circuit such as shown in FIG. 4 is utilized as a current restricting voltage source to provide the excitation voltage applied to the resonant cavity. At the same time the interconnected eletron extractor circuit, FIG. 5B, prevents the reintroduction of electrons back into the system. Depending on calculated design parameters, a predetermined voltage and frequency range may be designed for any particular application or physical configuration of the apparatus.

In the operation of the assembly, the pulse train source for the gas resonant cavity shown at **2** and **3** in FIG. **1** may be derived from a circuit such as shown in FIG. **4**. It is necessary in the final electron extraction that the frequency with which electrons are removed from the system by sequenced and synchronized with the pulsing of the gas resonant cavity In the circuit of FIG. 5B, the coordination or synchronization of the circuit with the circuit of FIG. **4** may be achieved by interconnecting point "A" of the gate circuit of FIG. 5B to coordinate point "A" of the pulsing circuit of FIG. **4**.

Together the hydrogen injector assembly **4** and the resonant cavity assemblies **2**, **3** form a gas injector fuel cell which is compact, light in weight and design variable. For example, the hydrogen injector system is suited for automobiles and jet engines. Industrial applications require larger systems. For rocket engine applications, the hydrogen gas injector system is positioned at the top of each resonant cavity arranged in a parallel cluster array. If resonant cavities are sequentially combined in a parallel/series array, the hydrogen injection assembly is positioned after the exits of said resonant cavities are combined.

From the outline of physical phenomena associated with the process described in Table **1**, the theoretical basis of the invention considers the respective states of molecules, gases and ions derived from liquid water. Before voltage stimulation, water molecules are randomly dispersed throughout water within a container. When a unipolar voltage pulse train such as shown in FIG. 3A (**53***a* xxx **53***n*) is applied, an increasing voltage potential is induced in the molecules, gases and/or ions in a linear, step-like charging effect. The electrical field of the particles within a chamber including the electrical field plates increases from a low energy state (A) to a high energy state (J) in a step manner following each pulse-train as illustrated in FIG. 3A. The increasing voltage potential is always positive in direct relationship to negative ground potential during each pulse. The voltage polarity on the plates which create the voltage fields remains constant. Positive and negative voltage "zones" are thus formed simultaneously.

In the first stage of the process described in Table **1**, because the water molecule naturally exhibits opposite electrical fields in a relatively polar configuration (the two hydrogen atoms are positively electrically charged relative to the negative electrically charged oxgen atom), the voltage pulse causes initially randomly oriented water molecules in the liquid state to spin and orient themselves with reference to positive and negative poles of the voltage fields applied. The positive electrically charged hydrogen atoms of said water molecule are attracted to a negative voltage field; while, at the same time, the negative electrically charged oxygen atoms of the same water molecule an attracted to a positive voltage field. Even a slight potential difference applied to the inert, conductive plates of a containment chamber will initiate polar atomic orientation within the water molecule based on polarity differences.

When the potential difference applied causes the orientated water molecules to align themselves between the conductive plates, pulsing causes the voltage field intensity to be increased in accordance with FIG. 3A. As further molecular alignment occurs, molecular movement is hindered. Because the positively charged hydrogen atoms of said aligned molecules are attracted in a direction opposite to the negatively charged oxygen atoms, a polar charge alignment or distribution occurs within the molecules between said voltage zones, as shown in FIG. 3B. And as the energy level of the atoms subjected to resonant pulsing increases, the stationary water molecules become elongated as shown in FIG. 3C. Electrically charged nuclei and electrons are attracted toward opposite electrically charged voltage zones—disrupting the mass equilibium of the water molecule.

In the first stage, as the water molecule is further exposed to a potential difference, the electrical force of attraction of the atoms within the molecule to the electrodes of the chamber also increases in intensity. As a result, the covalent bonding between said atoms which forms the molecule is weakened and ultimately terminated. The negatively charged electron is attracted toward the positively charged hydrogen atoms, while at the same time, the negatively charged oxygen atoms repel electrons.

Once the applied resonant energy caused by pulsation of the electrical field in the cavities reaches a threshold level, the disassociated water molecules, now in the form of liberated hydrogen, oxygen, and ambient air gases begin to ionize and lose or gain electrons during the final stage in the injector assembly. Atom destablization occurs and the electrical and mass equilibrium of the atoms is disrupted. Again, the positive field produced within the chamber or cavity that encompasses the gas stream attracts negatively charged ions while the positively charged ions (and/or hydrogen nuclei) are attracted to the negative field. Atom stabilization does not occur because the pulsating voltage applied is repetitive without polarity change. A potential of approximately several thousand volts triggers the ionization state.

As the ionized particles accumulate within said chamber, the electrical charging effect is again an incremental stepping effect that produces an accumlative increased potential while, at the same time, resonance occurs. The components of the atom begin to "vibrate" at a resonant frequency such that an atomic instability is created. As shown in FIG. 3D, a high energy level is achieved, which then collapses resulting in the release of thermal explosive energy. Particle impact occurs when liberated ions in a gas are subjected to further voltage. A longitudinal cross section of a gas resonant cavity is shown in FIG. 2A. To promote gas ionization, electromagnetic wave energy such as a laser or photon energy source of a predetermined wave length and pulse-intensity is directed to and absorbed by the ions forming said gas. In the device of FIG. 2A, semiconductor optical lasers **20***a*–**20***p*, **20***xxx* surround the gas flow path. In the device of FIG. 2B, photon energy **20** is injected into a separate absorption chamber **21**. The incremental stimulation of nuclei to a more highly energized state by electromagnetic wave energy causes electron deflection to a higher orbital state. The Pulse rate as well as intensity of the electromagnetic wave source is varied to match the absorption rate of ionized particles to produce the stepped incremental increase in energy. A single laser coupled by means of fiber optic light guides is an alternative to the plurality of lasers shown in FIG. 2B. Continued exposure of the gas ions to different forms of wave energy during voltage stimulation maintains individual atoms in a destabilized state and prevents atomic stabilization.

The highly energized gas ions are thermally ignited when said combustible gas ions pass from injector **4** and enter into and pass through a nozzle, **10** in FIG. 2B, or an optical thermal lens assembly such as shown in FIG. 2C. In FIG. 2C, the combustible gas ions are expelled through and beyond a quenching circuit, **30**, and reflected by lenses, **31** and **32**, back and forth through a thermal heat zone, **33**, prior to atomic breakdown beyond exiting through a final port, **34**. A quenching circuit is a restricted orifice through which the particle stream passes such that flashback does not occur. (See my application Ser. No. 835, 564, supra.) The deflection shield or lens, **31**, superheats beyond 3,000° F. and the combustible gas ions passing through said exiting-ports are regulated to allow a gas pressure to form inside said thermal zone. The energy yield is controlled by varying the applied voltage, or Pulse-train since said thermal-lens assembly is self-adjusting to the flow-rate of said ionized and primed gases. The combustible ionic gas mixture is composed of hydrogen, oxygen, and ambient air gases. The hydrogen gas provides the thermal explosive force, the oxygen atoms aid the gas thermal ignition, and the ambient air gases retard the gas thermal ignition process to a controllable state. As the combustible gas mixture is exposed to a voltage pulse train, the stepped increasing voltage potential causes said moving gas atoms to become ionized (losing or gaining electrons) and changes the electrical and mass equilibrium of said atoms. Gases that do not undergo the gas ionization process may accept the liberated electrons (electron entrapment) when exposed to light or photon stimulation. The electron extractor grid circuit, FIGS. 5A and 5B, is applied to the assembly of FIG. 2A or FIG. 2B, and restricts electron replacement. The extractor grid, **56**, is applied adjacent to electric field producing members, **44** and **45**, within the resonant cavity. The gas ions incrementally reach a critical-state which occurs after a high energy resonant state. At this point the atoms no longer tolerate the missing electrons, the unbalanced electrical field, and the energy stored in the nucleus. Immediate collapse of the system occurs and energy is released as the atoms decay into thermal explosive energy.

The repetitive application of a voltage pulse train (A through J of FIG. 3A) incrementally achieves the critical state of said gas ions. As the gas atoms or ions (la xxx ln) shown in FIG. 3C become elongated during electron removal, electromagnetic wave energy of a predetermined frequency and intensity is injected. The wave energy absorbed by the stimulated gas nuclei and electrons causes further destabilization of the ionic gas. The absorbed energy from all sources causes the gas nuclei to increase in energy state, and induces the ejection of electrons from the nuclei.

To further stimulate the electron entrapment process beyond the atomic level (capturing the liberated electrons during the hydrogen fracturing process) the electron extractor grid (as shown in FIG. 5A) is placed in spaced relationship to the gas resonant cavity structure shown in FIG. 2A. The electron extractor grid is attached to an electrical circuit (such as shown in FIG. 5B) that allows electrons to flow to an electrical load, **55**, when a positive electrical potential is placed on the opposite side of said electrical load. The electrical load may be a typical power consuming device such as a light bulb or resistive heat producing device. As the positive electrical potential is switched on or pulse-applied, the negative charged electrons liberated in the gas resonant cavity are drawn away and enter into resistive load where they are consumed and released as heat or light energy. The consuming electrical circuit can be directly connected to the gas resonant cavity positive electrical voltage zone. The incoming positive wave form applied to resonant cavity voltage zone through a blocking diode is synchronized with the pulse train applied to the gas resonant cavity by the circuit of FIG. **4** via alternate gate circuit. As one pulse train is gated

"ON," the other pulse train is switched "OFF." A blocking diode directs the electron flow to said electrical load while resistive wire prevents voltage leakage during pulse train "ON" time.

The electron extraction process is maintained during gas flow-rate change by varying the trigger pulse rate in relationship to applied voltage. The electron extraction process also prevents spark-ignition of the combustible gases traveling through the gas resonant cavity because electron build-up and potential sparking is prevented.

In an optical thermal lens assembly or thrust-nozzle, such as shown in FIG. 2C, destablized gas ions (electrically and mass unbalanced gas atoms having highly energized nuclei) can be pressurized during spark-ignition. During thermal interaction, the highly energized and unstable hydrogen gas nuclei collide with the highly energized and unstable oxygen gas nuclei and produce thermal explosive energy beyond the gas burning stage. Other ambient air gases and ions not otherwise consumed limit the thermal explosive process.

Variations of the process and apparatus may be evident to those skilled in the art.

What is claimed is:

1. A method of obtaining the release of energy from a gas mixture including hydrogen and oxygen consisting of:

(A) providing a first gas mixture including at least a portion of hydrogen and oxygen gases;

(B) subjecting the gas mixture to a pulsating, polar electric field whereby electrons of the gas atoms are distended in their orbital fields by reason of their subjection to electrical polar forces, at a frequency such that the pulsating electric field induces a resonance with respect to an electron of the gas atom;

(C) cascading said gas atoms with respect to the pulsating electric field such that the energy level of the resonant electron is increased in cascading incremental steps;

(D) ionizing said gas atoms;

(E) subjecting the ionized gas atoms to electromagnetic wave energy having a predetermined frequency to induce a further election resonance in the ion, whereby the energy level of the electron is successively increased;

(F) extracting further electrons from the resonating ions while such ions are in an increased energy state to destabilize the nuclear and electron configuration of said ions; and

(G) subjecting the destabilized ions to thermal ignition.

2. An apparatus for obtaining the release of energy from a gas mixture including hydrogen and oxygen consisting of successively interconnected:

(A) first means for providing a first gas mixture including at least a portion of hydrogen and oxygen gas;

(B) second means for providing a pulsating, polar electric field to the gas mixture, whereby electrons of the gas atoms are distended in their orbital fields by reason of their subjection to electrical polar forces, at a frequency such that the pulsating electric field induces a resonance with respect to an electron of the gas atom; and the energy level of the resonant electron is increased in cascading, incremental steps;

(C) third means for providing a further electric field to ionize said gas atoms;

(D) an electromagnetic wave energy source for subjecting the ionized gas atoms to wave energy of a predetermined frequency to induce a further election resonance in the ion, whereby the energy level of the electron is further successively increased;

(E) an electron sink for extracting electrons from the resonating ions while such ions are in an increased energy state to destabilize the nuclear and electron configuration of said ions;

(F) fourth means for directing particle flow in a continuous manner through the electric fields, wave energy source and electron sink to a final orifice at which the destabilized ions are thermally ignited; and

(G) a final orifice at which the mixture initially provided by the first means, after having passed through and been processed by the preceeding means of the apparatus, is thermally ignited.

* * * * *

United States Patent [19]
Meyer

[11] **Patent Number: 4,936,961**
[45] **Date of Patent: Jun. 26, 1990**

[54] METHOD FOR THE PRODUCTION OF A FUEL GAS

[76] Inventor: **Stanley A. Meyer,** 3792 Broadway, Grove City, Ohio 43123

[21] Appl. No.: **207,730**

[22] Filed: **Jun. 16, 1988**

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 81,859, Aug. 5, 1987, Pat. No. 4,826,581.

[51] **Int. Cl.**^{5} **C07G 13/00**
[52] **U.S. Cl.** **204/157.5;** 204/157.52
[58] **Field of Search** 204/157.5, 157.52, 183.1

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,740,283	4/1988	Laas et al.	204/183.1 X
4,511,450	4/1985	Neefe	204/152.5 X
4,696,809	9/1987	Vialoron et al.	204/157.52 X

OTHER PUBLICATIONS

Julius Grant, ed., *Hachh's Chemical Dictionary,* 4th ed., McGraw-Hill Book Co., 1969, p. 282.

Primary Examiner—Stephen J. Kalafut
Attorney, Agent, or Firm—Porter, Wright, Morris & Arthur

[57] **ABSTRACT**

A method for obtaining the release of a fuel gas mixture including hydrogen and oxygen from water in which the water is processed as a dielectric medium in an electrical resonant circuit.

2 Claims, 3 Drawing Sheets

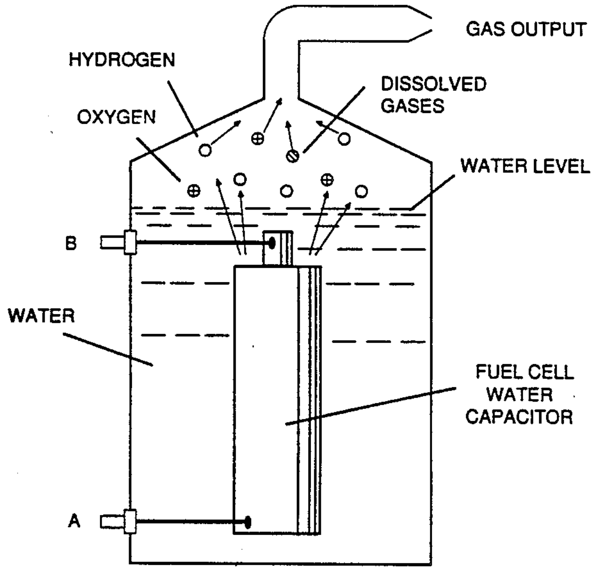

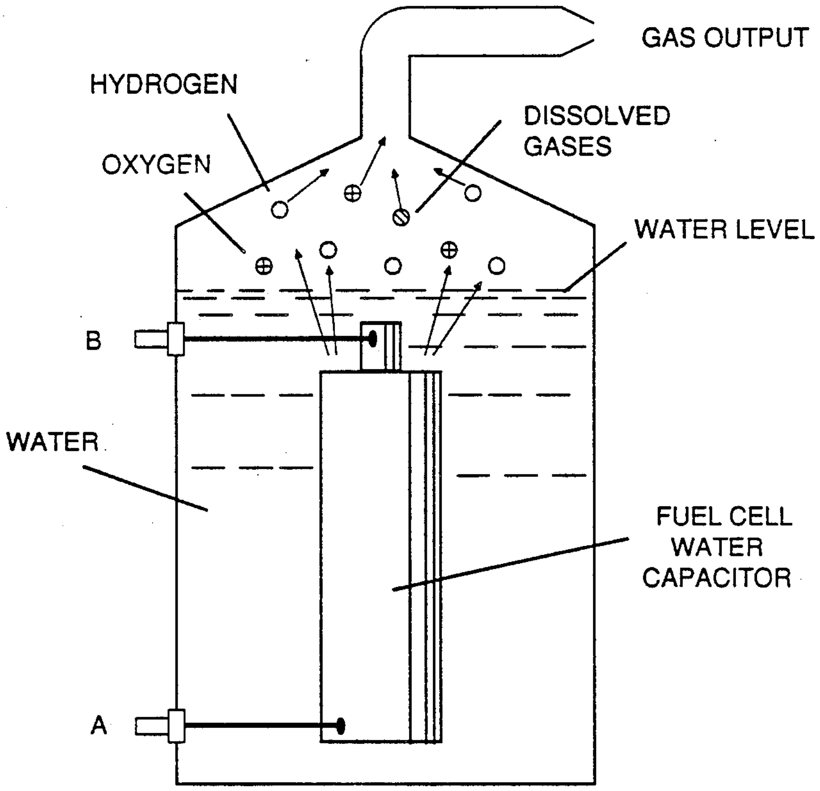
GAS OUTPUT
HYDROGEN
DISSOLVED GASES
OXYGEN
WATER LEVEL
B
WATER
FUEL CELL WATER CAPACITOR
A

FIGURE 2

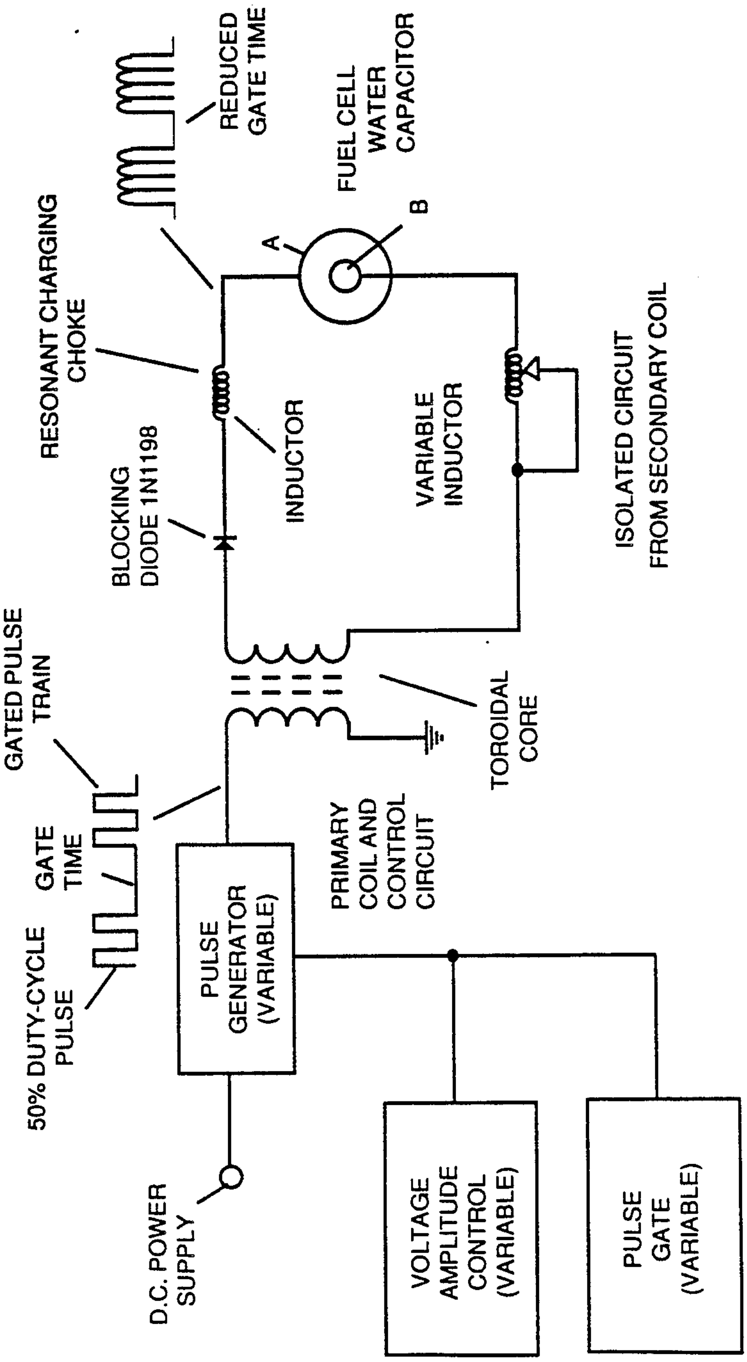
D.C. POWER SUPPLY
50% DUTY-CYCLE PULSE
GATE TIME
GATED PULSE TRAIN
PULSE GENERATOR (VARIABLE)
PRIMARY COIL AND CONTROL CIRCUIT
VOLTAGE AMPLITUDE CONTROL (VARIABLE)
PULSE GATE (VARIABLE)
TOROIDAL CORE
BLOCKING DIODE 1N1198
INDUCTOR
RESONANT CHARGING CHOKE
VARIABLE INDUCTOR
ISOLATED CIRCUIT FROM SECONDARY COIL
A
B
FUEL CELL WATER CAPACITOR
REDUCED GATE TIME

FIGURE 1

FIG.

3A

3B

3C

3D

3E

3F

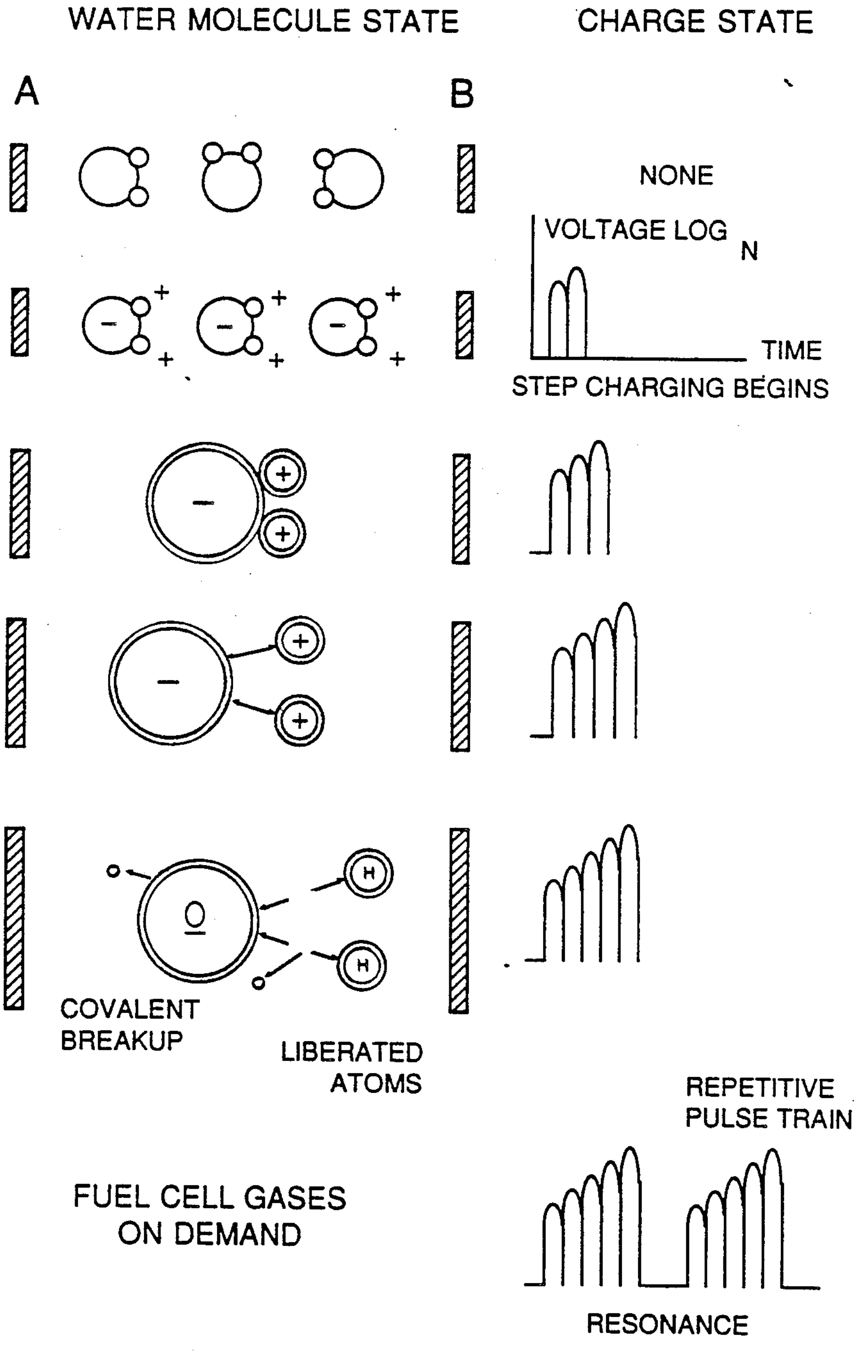
WATER MOLECULE STATE
CHARGE STATE
A
B
NONE
VOLTAGE LOG
N
TIME
STEP CHARGING BEGINS
−
+
O
H
COVALENT BREAKUP
LIBERATED ATOMS
REPETITIVE PULSE TRAIN
FUEL CELL GASES ON DEMAND
RESONANCE

METHOD FOR THE PRODUCTION OF A FUEL GAS

Related Application

This is a continuation-in-part of my co-pending application Ser. No. 081,859, filed 8/5/87, now U.S. Pat. No. 4,826,581.

Field of Invention

This invention relates to a method of and apparatus for obtaining the release of a fuel gas mixture including hydrogen and oxygen from water.

BACKGROUND OF THE PRIOR ART

Numerous processes have been proposed for separating a water molecule into its elemental hydrogen and oxygen components. Electrolysis is one such process. Other processes are described in United States patents such as 4,344,831; 4,184,931; 4,023,545; 3,980,053; and Patent Cooperation Treaty application No. PCT/US80/1362, published 30 April, 1981.

OBJECTS OF THE INVENTION

It is an object of the invention to provide a fuel cell and a process in which molecules of water are broken down into hydrogen and oxygen gases, and a fuel gas mixture including hydrogen, oxygen and other gasses formerly dissolved within the water is produced. As used herein the term "fuel cell" refers to a single unit of the invention comprising a water capacitor cell, as hereinafter explained, that produces the fuel gas in accordance with the method of the invention. Brief Description of the Drawings

FIG. 1 illustrates a circuit useful in the process.

FIG. 2 shows a perspective of a "water capacitor''element used in the fuel cell circuit.

FIGS. 3A through 3F are illustrations depicting the theoretical bases for phenomena encountered during operation of the invention herein.

Description of the Preferred Embodiment

In brief, the invention is a method of obtaining the release of a gas mixture including hydrogen and oxygen and other dissolved gases formerly entrapped in water, from water consisting of: (A) providing a capacitor, in which the water is included as a dielectric liquid between capacitor plates, in a resonant charging choke circuit that includes an inductance in series with the capacitor; (B) subjecting the capacitor to a pulsating, unipolar electric voltage field in which the polarity does not pass beyond an arbitrary ground, whereby the water molecules within the capacitor are subjected to a charge of the same polarity and the water molecules are distended by their subjection to electrical polar forces; (C) further subjecting the water in said capacitor to said pulsating electric field to achieve a pulse frequency such that the Pulsating electric field induces a resonance within the water molecule; (D) continuing the application of the pulsing frequency to the capacitor cell after resonance occurs so that the energy level within the molecule is increased in cascading incremental steps in proportion to the number of pulses; (E) maintaining the charge of said capacitor during the application of the pulsing field, whereby the co-valent electrical bonding of the hydrogen and oxygen atoms within said molecules is destabilized such that the force of the electrical field applied, as the force is effective within the molecule, exceeds the bonding force of the molecule, and hydrogen and oxygen atoms are liberated from the molecule as elemental gases; and (F) collecting said hydrogen and oxygen gases, and any other gases that were formerly dissolved within the water, and discharging the collected gases as a fuel gas mixture.

The process follows the sequence of steps shown in the following Table I in which water molecules are subjected to increasing electrical forces. In an ambient state, randomly oriented water molecules are aligned with respect to a molecular polar orientation. They are next, themselves polarized and "elongated" by the application of an electric Potential to the extent that covalent bonding of the water molecule is so weakened that the atoms disassociate and the molecule breaks down into hydrogen and oxygen elemental components. Engineering design parameters based on known theoretical principles of electrical circuits determine the incremental levels of electrical and wave energy input required to produce resonance in the system whereby the fuel gas comprised of a mixture of hydrogen, oxygen, and the other gases such as air test were formerly dissolved within the water, is produced.

TABLE I

	PROCESS STEPS THE SEQUENCE OF THE RELATIVE STATE OF THE WATER MOLECULE AND/OR HYDROGEN/OXYGEN/OTHER ATOMS:
A.	(AMBIENT STATE) RANDOM
B.	ALIGNMENT OF POLAR FIELDS
C.	POLARIZATION OF MOLECULE
D.	MOLECULAR ELONGATION
E.	ATOM LIBERATION BY BREAKDOWN OF COVALENT BOND
F.	RELEASE OF GASES

In the process, the point of optimum gas release is reached at a circuit resonance. Water in the fuel cell is subjected to a pulsating, polar electric field produced by the electrical circuit whereby the water molecules are distended by reason of their subjection to electrical polar forces of the capacitor plates. The polar pulsating frequency applied is such that the pulsating electric field induces a resonance in the molecule. A cascade effect occurs and the overall energy level of specific water molecules is increased in cascading, incremental steps. The hydrogen and oxygen atomic gases, and other gas components formerly entrapped as dissolved gases in water, are released when the resonant energy exceeds the co-valent bonding force of the water molecule. A preferred construction material for the capacitor plates is a stainless steel T-304 which is non-chemically reactive with water, hydrogen, or oxygen. An electrically conductive material which is inert in the fluid environment is a desirable material of construction for the electrical field plates of the "water capacitor" employed in the circuit.

Once triggered, the gas output is controllable by the attenuation of operational parameters. Thus, once the frequency of resonance is identified, by varying the applied pulse voltage to the water fuel cell assembly, gas output is varied. By varying the pulse shape and/or amplitude or pulse train sequence of the initial pulsing wave source, final gas output is varied. Attenuation of the voltage field frequency in the form of OFF and ON pulses likewise affects output.

The overall apparatus thus includes an electrical circuit in which a water capacitor having a known dielectric property is an element. The fuel gases are obtained from the water by the disassociation of the water molecule. The water molecules are split into component atomic elements (hydrogen and oxygen gases) by a voltage stimulation process called the electrical polarization process which also releases dissolved gases entrapped in the water.

From the outline of physical phenomena associated with the process described in Table 1, the theoretical basis of the invention considers the respective states of molecules and gases and ions derived from liquid water. Before voltage stimulation, water molecules are randomly dispersed throughout water within a container. When a unipolar voltage pulse train such as shown in FIGS. 3B through 3F is applied to positive and negative capacitor plates, an increasing voltage potential is induced in the molecules in a linear, step-like charging effect. The electrical field of the particles within a volume of water including the electrical field plates increases from a low energy state to a high energy state successively in a step manner following each pulse-train as illustrated figuratively in the depictions of FIG. 3A through 3F. The increasing voltage potential is always positive in direct relationship to negative ground potential during each pulse. The voltage polarity on the plates which create the voltage fields remains constant although the voltage charge increases. Positive and negative voltage "zones" are thus formed simultaneously in the electrical field of the capacitor plates.

In the first stage of the process described in Table 1, because the water molecule naturally exhibits opposite electrical fields in a relatively polar configuration (the two hydrogen atoms are positively electrically charged relative to the negative electrically charged oxygen atom), the voltage pulse causes initially randomly oriented water molecules in the liquid state to spin and orient themselves with reference to positive and negative poles of the voltage fields applied. The positive electrically charged hydrogen atoms of said water molecule are attracted to a negative voltage field; while, at the same time, the negative electrically charged oxygen atoms of the same water molecule are attracted to a positive voltage field. Even a slight potential difference applied to inert, conductive plates of a containment chamber which forms a capacitor will initiate polar atomic orientation within the water molecule based on polarity differences.

When the potential difference applied causes the orientated water molecules to align themselves between the conductive plates, pulsing causes the voltage field intensity to be increased in accordance with FIG. 3B. As further molecular alignment occurs, molecular movement is hindered. Because the positively charged hydrogen atoms of said aligned molecules are attracted in a direction opposite to the negatively charged oxygen atoms, a polar charge alignment or distribution occurs within the molecules between said voltage zones, as shown in FIG. 3B. And as the energy level of the atoms subjected to resonant pulsing increases, the stationary water molecules become elongated as shown in FIGS. 3C and 3D. Electrically charged nuclei and electrons are attracted toward opposite electrically charged voltage zones —disrupting the mass and charge equilibrium of the water molecule.

As the water molecule is further exposed to an increasing potential difference resulting from the step charging of the capacitor, the electrical force of attraction of the atoms within the molecule to the capacitor plates of the chamber also increases in strength. As a result, the co-valent bonding between atoms which form the molecule is weakened —and ultimately terminated. The negatively charged electron is attracted toward the positively charged hydrogen atoms, while at the same time, the negatively charged oxygen atoms repel electrons.

In a more specific explanation of the "sub-atomic" action that occurs in the water fuel cell, it is known that natural water is a liquid which has a dielectric constant of 78.54 at 20° C. and 1 atm pressure. [*Handbook of Chemistry and Physics*, 68th ed., CRC Press (Boca Raton, Florida (1987-88)), Section E-50. H_2O (water)].

When a volume of water is isolated and electrically conductive plates, that are chemically inert in water and are separated by a distance, are immersed in the water, a capacitor is formed, having a capacitance determined by the surface area of the plates, the distance of their separation and the dielectric constant of water.

When water molecules are exposed to voltage at a restricted current, water takes on an electrical charge. By the laws of electrical attraction, molecules align according to positive and negative polarity fields of the molecule and the alignment field. The plates of a capacitor constitute such an alignment field when a voltage is applied.

When a charge is applied to a capacitor, the electrical charge of the capacitor equals the applied voltage charge; in a water capacitor, the dielectric property of water resists the flow of amps in the circuit, and the water molecule itself, because it has polarity fields formed by the relationship of hydrogen and oxygen in the covalent bond, and an intrinsic dielectric property, becomes part of the electrical circuit, analogous to a "microcapacitor" within the capacitor defined by the plates.

In the Example of a fuel cell circuit of FIG. 1, a water capacitor is included. The step-up coil is formed on a conventional torroidal core formed of a compressed ferromagnetic powdered material that will not itself become permanently magnetized, such as the trademarked "Ferramic 06# "Permag" powder as described in *Siemens Ferrites Catalog*, CG-2000-002-121, (Cleveland, Ohio) No. F626-1205. The core is 1.50 inch in diameter and 0.25 inch in thickness. A primary coil of 200 turns of 24 gauge copper wire is provided and a coil of 600 turns of 36 gauge wire comprises the secondary winding.

In the circuit of FIG. 1, the diode is a lN1198 diode which acts as a blocking diode and an electric switch that allows voltage flow in one direction only. Thus, the capacitor is never subjected to a pulse of reverse polarity.

The primary coil of the torroid is subject to a 50% duty cycle pulse. The torroidal pulsing coil provides a voltage step-up from the pulse generator in excess of five times, although the relative amount of step-up is determined by pre-selected criteria for a particular application. As the stepped-up pulse enters first inductor (formed from 100 turns of 24 gauge wire 1 inch in diameter), an electromagnetic field is formed around the inductor, voltage is switched off when the pulse ends, and the field collapses and produces another pulse of the same polarity; i.e., another positive pulse is formed where the 50% duty cycle was terminated. Thus, a double pulse frequency is produced; however, in a pulse train of unipolar pulses, there is a brief time when pulses are not present.

By being so subjected to electrical pulses in the circuit of FIG. 1, water confined in the volume that includes the capacitor plates takes on an electrical charge that is increased by a step charging phenomenon occurring in the water capacitor. Voltage continually increases (to about 1000 volts and more) and the water molecule starts to elongate.

The pulse train is then switched off; the voltage across the water capacitor drops to the amount of charge that the water molecules have taken on, i.e. voltage is maintained across the charged capacitor. The pulse train is then reapplied.

Because a voltage potential applied to a capacitor can perform work, the higher the voltage potential, the more work is performed by a given capacitor. In an optimum capacitor that is wholly non-conductive, zero (0) current flow will occur across the capacitor. Thus, in view of an idealized capacitor circuit, the object of the water capacitor circuit is to prevent electron flow through the circuit, i.e. such as occurs by electron flow or leakage through a resistive element that produces heat. Electrical leakage in water will occur, however, because of some residual conductivity and impurities or ions that may be otherwise present in the water. Thus, the water capacitor is preferably chemically inert. An electrolyte is not added to the water.

In the isolated water bath, the water molecule takes on charge, and the charge increases. The object of the process is to switch off the co-valent bonding of the water molecule and interrupt the sub-atomic force, i.e. the electrical force or electromagnetic force, that binds the hydrogen and oxygen atoms to form a molecule so that the hydrogen and oxygen separate.

Because an electron will only occupy a certain electron shell (the shells are well known) the voltage applied to the capacitor affects the electrical forces inherent in the co-valent bond. As a result of the charge applied by the plates, the applied force becomes greater than the force of the co-valent bonds between the atom of the water molecule; and the water molecule becomes elongated. When this happens, the time share ratio of the electrons between the atoms and the electron shells is modified.

In the process, electrons are extracted from the water bath; electrons are not consumed nor are electrons introduced into the water bath by the circuit as electrons are conventionally introduced in an electrolysis process. There may nevertheless occur a leakage current through the water. Those hydrogen atoms missing electrons become neutralized; and atoms are liberated from the water. The charged atoms and electrons are attracted to opposite polarity voltage zones created between the capacitor plates. The electrons formerly shared by atoms in the water co-valent bond are reallocated such that neutral elemental gases are liberated.

In the process, the electrical resonance may be reached at all levels of voltage potential. The overall circuit is characterized as a "resonant charging choke" circuit which is an inductor in series with a capacitor that produces a resonant circuit. [*SAMS Modern Dictionary of Electronics*, Rudolff Garff, © 1984, Howard W. Sams & Co. (Indianapolis, Ind.), page 859.]Such a resonant charging choke is on each side of the capacitor. In the circuit, the diode acts as a switch that allows the magnetic field produced in the inductor to collapse, thereby doubling the pulse frequency and preventing the capacitor from discharging. In this manner a continuous voltage is produced across the capacitor plates in the water bath; and the capacitor does not discharge. The water molecules are thus subjected to a continuously charged field until the breakdown of the co-valent bond occurs.

As noted initially, the capacitance depends on the dielectric properties of the water and the size and separation of the conductive elements forming the water capacitor.

EXAMPLE I

In an example of the circuit of FIG. 1 (in which other circuit element specifications are provided above), two concentric cylinders 4 inches long formed the water capacitor of the fuel cell in the volume of water. The outside cylinder was 0.75 inch in outside diameter; the inner cylinder was 0.5 inch in outside diameter. Spacing from the outside of the inner cylinder to the inner surface of the outside cylinder was 0.0625 inch. Resonance in the circuit was achieved at a 26 volt applied pulse to the primary coil of the torroid at $0KH_Z$, and the water molecules disassociated into elemental hydrogen and oxygen and the gas released from the fuel cell comprised a mixture of hydrogen, oxygen from the water molecule, and gases formerly dissolved in the water such as the atmospheric gases or oxygen, nitrogen, and argon.

In achieving resonance in any circuit, as the pulse frequency is adjusted, the flow of amps is minimized and voltage is maximized to a peak. Calculation of the resonance frequency of an overall circuit is determined by known means; different cavities have a different frequencY of resonance dependent on parameters of the water dielectric, plate size, configuration and distance, circuit inductors, and the like. Control of the production of fuel gas is determined by variation of the period of time between a train of pulses, pulse amplitude and capacitor plate size and configuration, with corresponding value adjustments to other circuit components.

The wiper arm on the second inductor tunes the circuit and accommodates to contaminants in water so that the charge is always applied to the capacitor. The voltage applied determines the rate of breakdown of the molecule into its atomic components. As water in the cell is consumed, it is replaced by any appropriate means or control system.

Variations of the process and apparatus may be evident to those skilled in the art.

What is claimed is:

1. A method of obtaining the release of a gas mixture including hydrogen and oxygen and other dissolved gases formerly entrapped in water, from water, consisting of:

(A) providing a capacitor in which water is included as a dielectric between capacitor plates, in a resonant charging choke circuit that includes an inductance in series with the capacitor;

(B) subjecting the capacitor to a pulsating, unipolar electric charging voltage in which the polarity does not pass beyond an arbitrary ground, whereby the water molecules within the capacitor are subjected to the electric field between the capacitor plates;

(C) further subjecting the water in said capacitor to a pulsating electric field resulting from the subjection of the capacitor to the charging voltage such that the pulsating electric field induces a resonance within the water molecules;

(D) continuing the application of the pulsating charging voltage to the capacitor after resonance occurs so that the energy level within the molecules is increased in cascading incremental steps in proportion to the number of pulses;

(E) maintaining the charge of said capacitor during the application of the pulsating charging voltage, whereby the co-valent electrical bonding of the hydrogen and oxygen atoms within said molecules is destabilized, such that the force of the electrical field applied to the molecules exceeds the bonding force within the molecules, and hydrogen and oxygen atoms are liberated from the molecules as elemental gases.

2. The method of claim 1 including the further steps of collecting said liberated hydrogen and oxygen gases, and any other gases that were formerly dissolved with in the water and discharging said collected gases as a fuel gas mixture.

* * * * *

US005149407A

United States Patent [19]

Meyer

[11] **Patent Number: 5,149,407**

[45] **Date of Patent: * Sep. 22, 1992**

[54] **PROCESS AND APPARATUS FOR THE PRODUCTION OF FUEL GAS AND THE ENHANCED RELEASE OF THERMAL ENERGY FROM SUCH GAS**

[76] Inventor: **Stanley A. Meyer**, 3792 Broadway, Grove City, Ohio 43123

[*] Notice: The portion of the term of this patent subsequent to Jun. 26, 2007 has been disclaimed.

[21] Appl. No.: **460,859**

[22] PCT Filed: **Jun. 15, 1989**

[86] PCT No.: **PCT/US89/02622**

§ 371 Date: **Feb. 13, 1990**

§ 102(e) Date: **Feb. 13, 1990**

[87] PCT Pub. No.: **WO89/12704**

PCT Pub. Date: **Dec. 28, 1989**

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 207,730, Jun. 16, 1988, Pat. No. 4,936,961, which is a continuation-in-part of Ser. No. 81,859, Aug. 5, 1987, Pat. No. 4,826,581.

[51] **Int. Cl.**^{5} **C07G 13/00**

[52] **U.S. Cl.** **204/157.41**; 204/157.5; 204/157.52; 204/193

[58] **Field of Search** 204/157.41, 157.5, 157.52, 204/164, 193

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,772,180	11/1973	Prestidge	204/305
4,511,450	4/1985	Neefe	204/278
4,696,809	9/1987	Vialaron et al.	204/157.52 X
4,740,283	4/1988	Laas et al.	204/182.8
4,826,581	5/1989	Meyer	204/157.41
4,936,961	6/1990	Meyer	204/157.5

OTHER PUBLICATIONS

Chemical Abstracts vol. 88, NO. 52526n, 1978 (Hasebe).
Chemical Abstracts vol. 77, No. 156261m, 1972 (Zubrilov et al.).

Primary Examiner—Stephen J. Kalafut
Attorney, Agent, or Firm—Porter, Wright, Morris & Arthur

[57] **ABSTRACT**

Water molecules are broken down into hydrogen and oxygen gas atoms in a capacitive cell by a polarization and resonance process dependent upon the dielectric properties of water and water molecules. The gas atoms are thereafter ionized or otherwise energized and thermally combusted to release a degree of energy greater than that of combustion of the gas in ambient air.

12 Claims, 15 Drawing Sheets

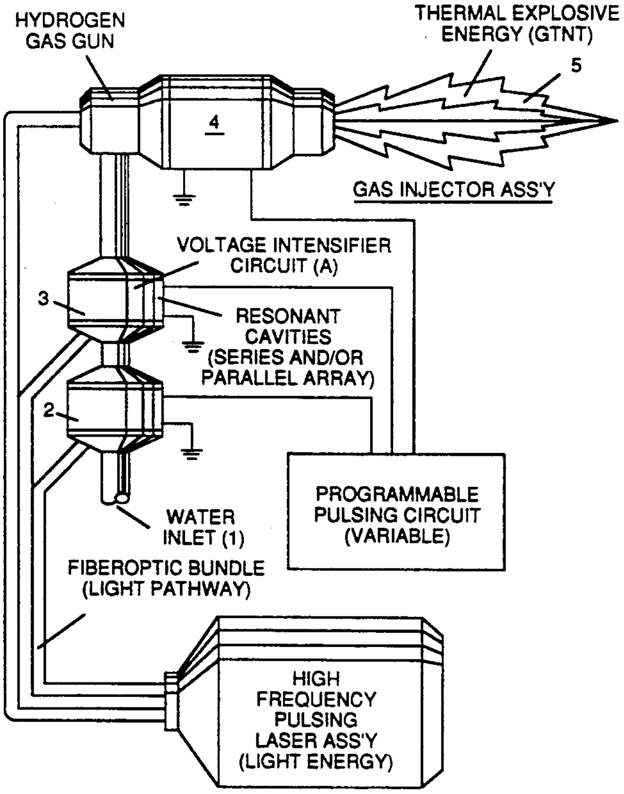
HYDROGEN GAS GUN
THERMAL EXPLOSIVE ENERGY (GTNT)
5
4
GAS INJECTOR ASS'Y
VOLTAGE INTENSIFIER CIRCUIT (A)
3
RESONANT CAVITIES (SERIES AND/OR PARALLEL ARRAY)
2
PROGRAMMABLE PULSING CIRCUIT (VARIABLE)
WATER INLET (1)
FIBEROPTIC BUNDLE (LIGHT PATHWAY)
HIGH FREQUENCY PULSING LASER ASS'Y (LIGHT ENERGY)

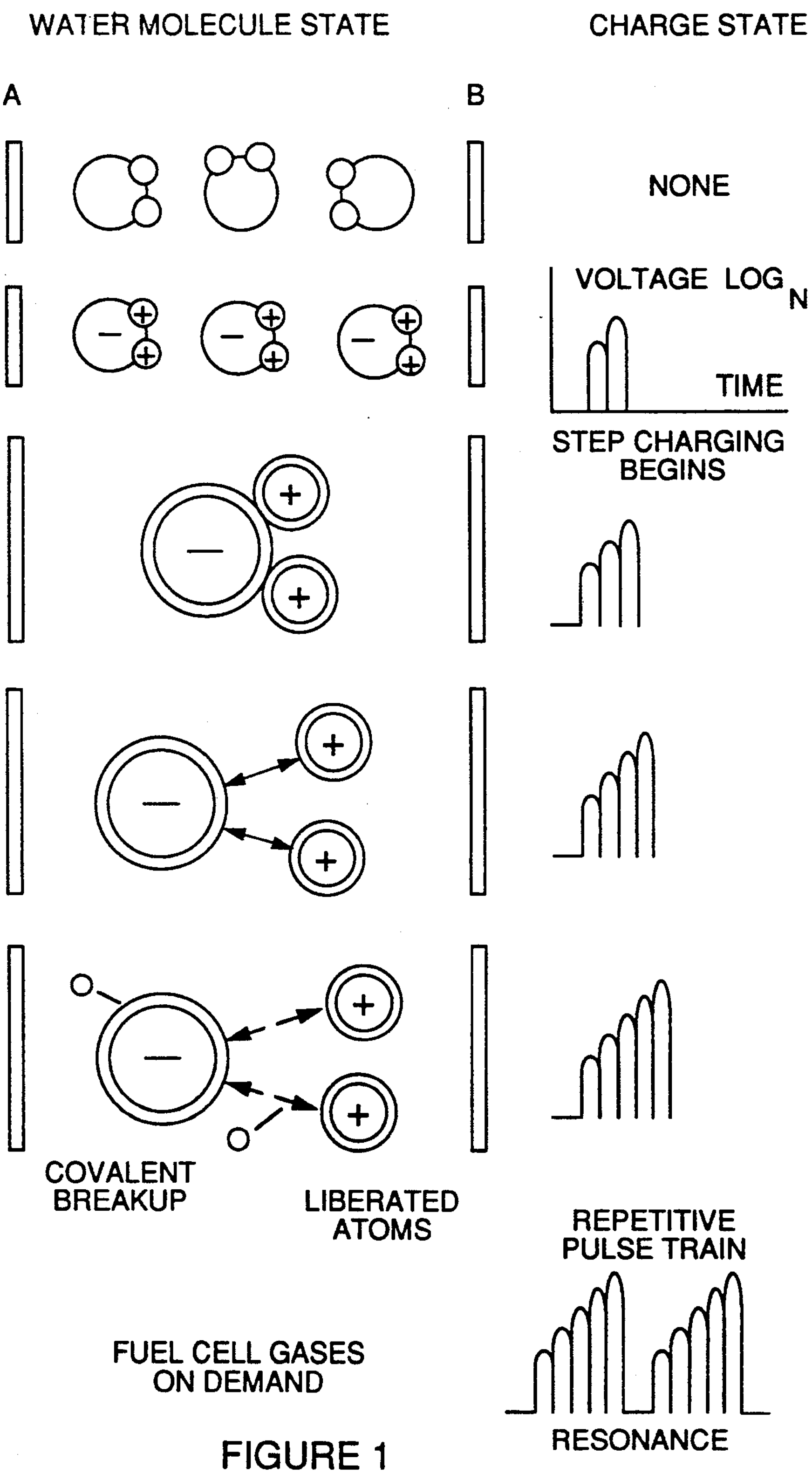

FIG. 1A

FIG. 1B

FIG. 1C

FIG.1D

FIG. 1E

FIG. 1F

WATER MOLECULE STATE
CHARGE STATE
A
B
NONE
VOLTAGE LOG N
TIME
STEP CHARGING BEGINS
COVALENT BREAKUP
LIBERATED ATOMS
REPETITIVE PULSE TRAIN
FUEL CELL GASES ON DEMAND
RESONANCE

FIGURE 1

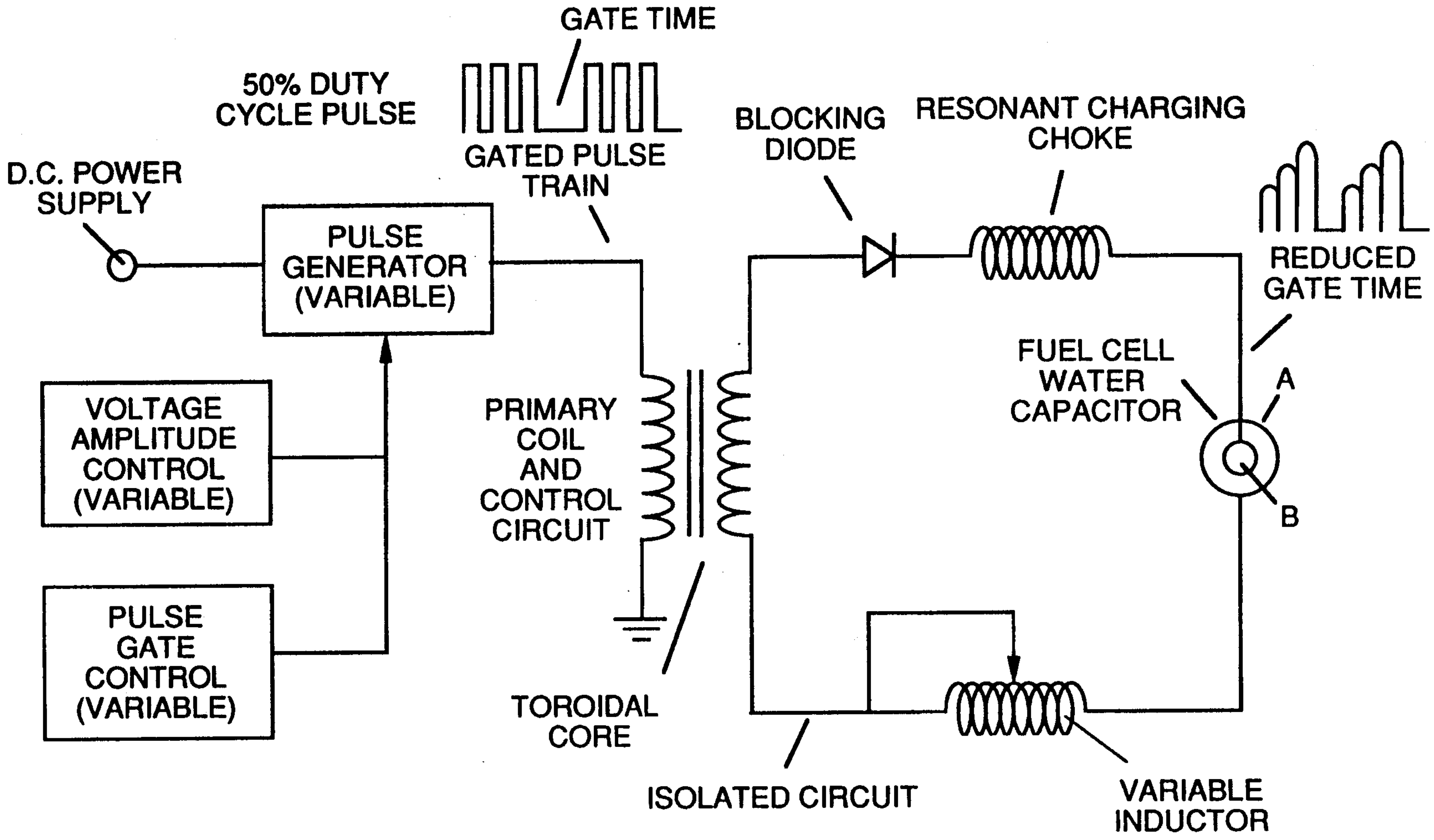
GATE TIME
50% DUTY CYCLE PULSE
GATED PULSE TRAIN
D.C. POWER SUPPLY
PULSE GENERATOR (VARIABLE)
BLOCKING DIODE
RESONANT CHARGING CHOKE
REDUCED GATE TIME
FUEL CELL WATER CAPACITOR
A
B
VOLTAGE AMPLITUDE CONTROL (VARIABLE)
PRIMARY COIL AND CONTROL CIRCUIT
PULSE GATE CONTROL (VARIABLE)
TOROIDAL CORE
ISOLATED CIRCUIT
VARIABLE INDUCTOR

FIGURE 2

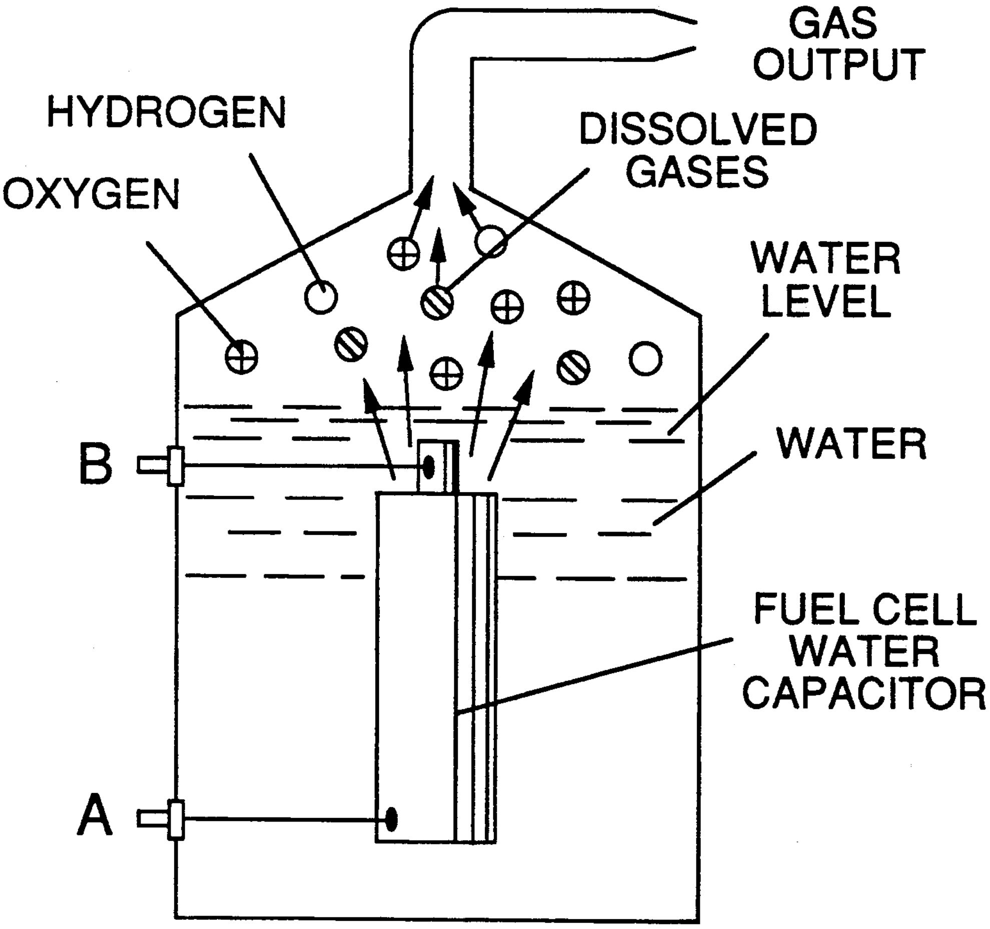
GAS OUTPUT
HYDROGEN
DISSOLVED GASES
OXYGEN
WATER LEVEL
WATER
B
FUEL CELL WATER CAPACITOR
A

FIGURE 3

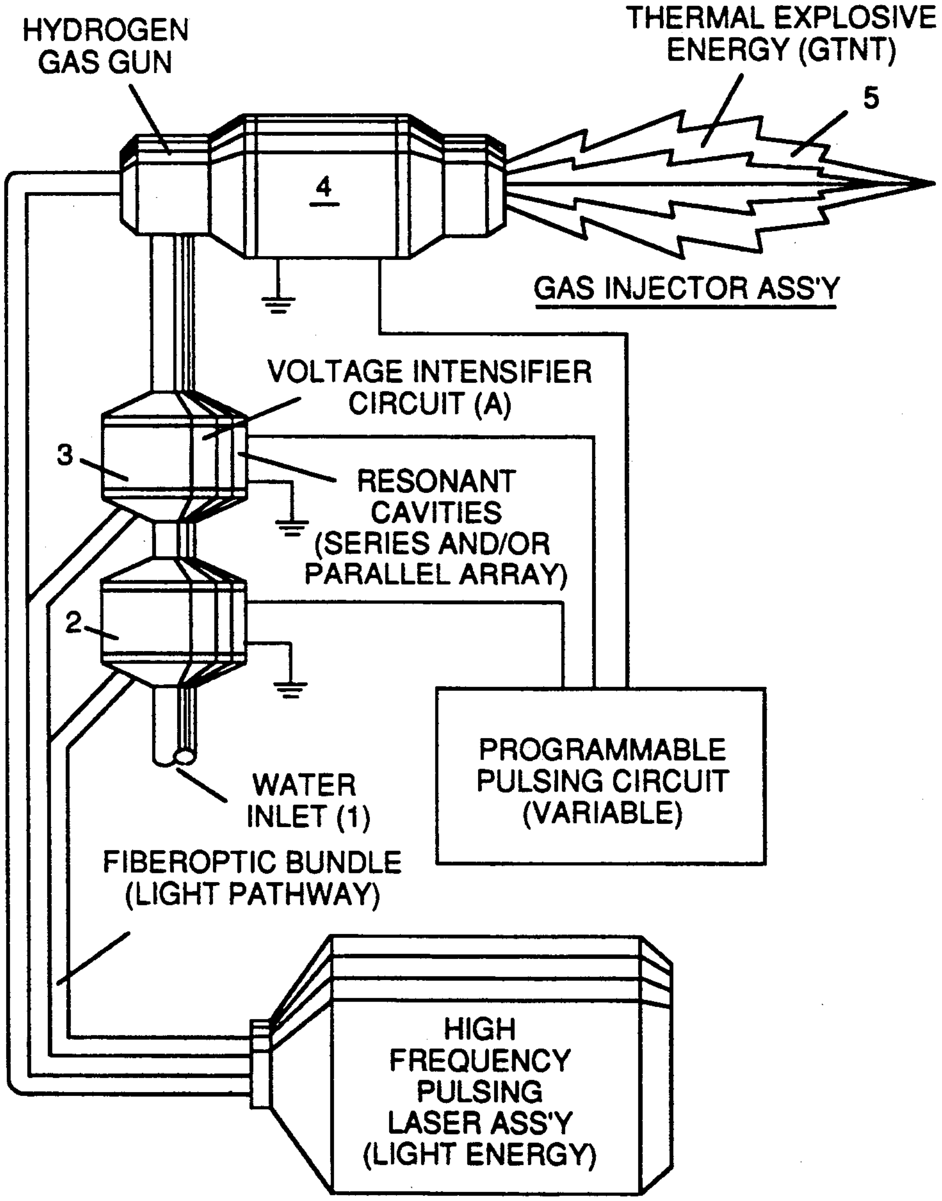
HYDROGEN GAS GUN
THERMAL EXPLOSIVE ENERGY (GTNT)
5
4
GAS INJECTOR ASS'Y
VOLTAGE INTENSIFIER CIRCUIT (A)
3
RESONANT CAVITIES (SERIES AND/OR PARALLEL ARRAY)
2
PROGRAMMABLE PULSING CIRCUIT (VARIABLE)
WATER INLET (1)
FIBEROPTIC BUNDLE (LIGHT PATHWAY)
HIGH FREQUENCY PULSING LASER ASS'Y (LIGHT ENERGY)

FIGURE 4

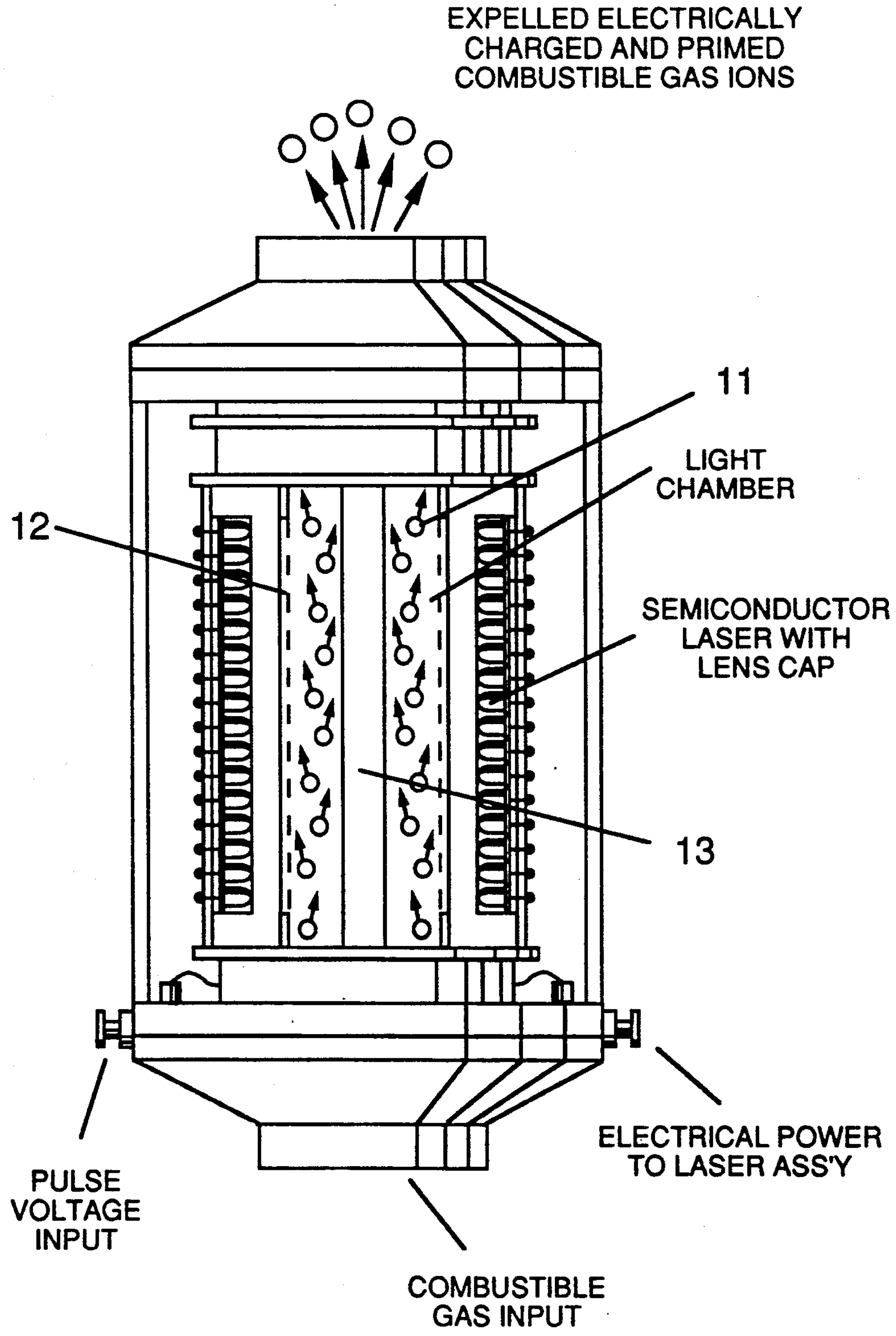
EXPELLED ELECTRICALLY
CHARGED AND PRIMED
COMBUSTIBLE GAS IONS
11
LIGHT
CHAMBER
12
SEMICONDUCTOR
LASER WITH
LENS CAP
13
PULSE
VOLTAGE
INPUT
ELECTRICAL POWER
TO LASER ASS'Y
COMBUSTIBLE
GAS INPUT

FIGURE 5A

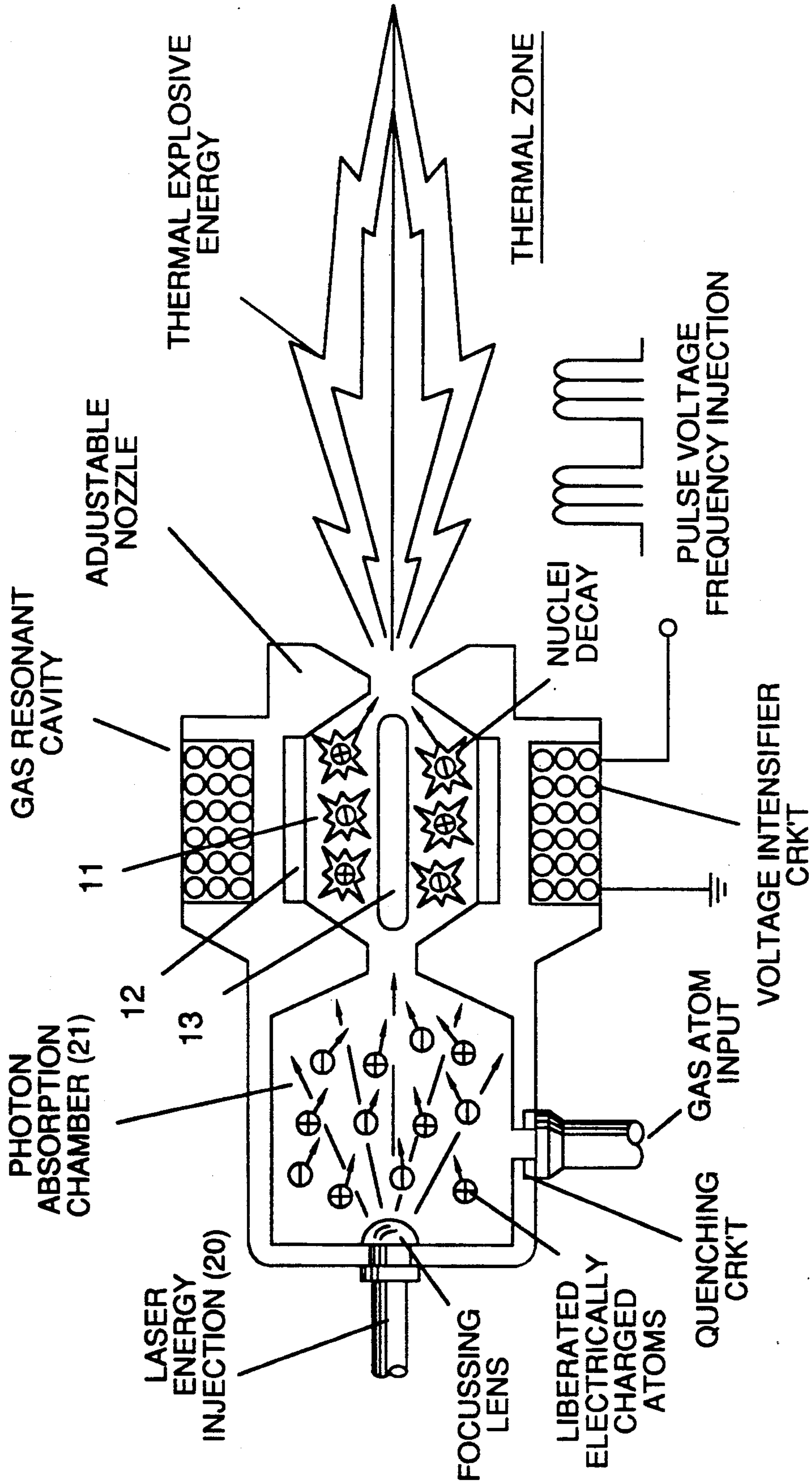
LASER ENERGY INJECTION (20)
PHOTON ABSORPTION CHAMBER (21)
GAS RESONANT CAVITY
11
12
13
ADJUSTABLE NOZZLE
THERMAL EXPLOSIVE ENERGY
THERMAL ZONE
FOCUSSING LENS
LIBERATED ELECTRICALLY CHARGED ATOMS
QUENCHING CRK'T
GAS ATOM INPUT
NUCLEI DECAY
PULSE VOLTAGE FREQUENCY INJECTION
VOLTAGE INTENSIFIER CRK'T

FIGURE 5B

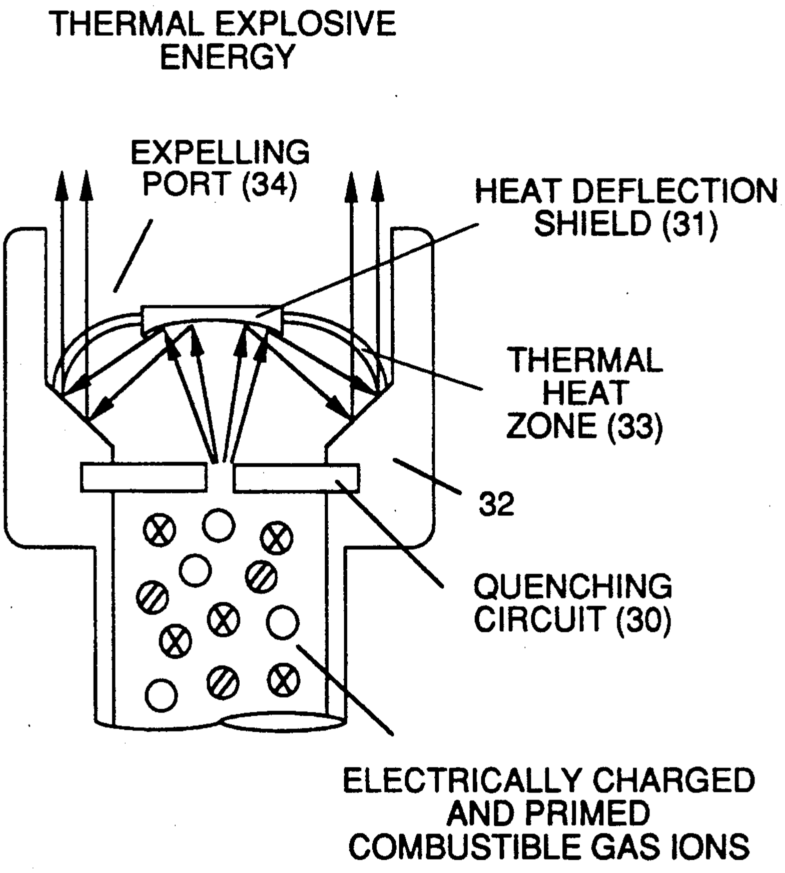
THERMAL EXPLOSIVE ENERGY
EXPELLING PORT (34)
HEAT DEFLECTION SHIELD (31)
THERMAL HEAT ZONE (33)
32
QUENCHING CIRCUIT (30)
ELECTRICALLY CHARGED AND PRIMED COMBUSTIBLE GAS IONS

FIGURE 5C

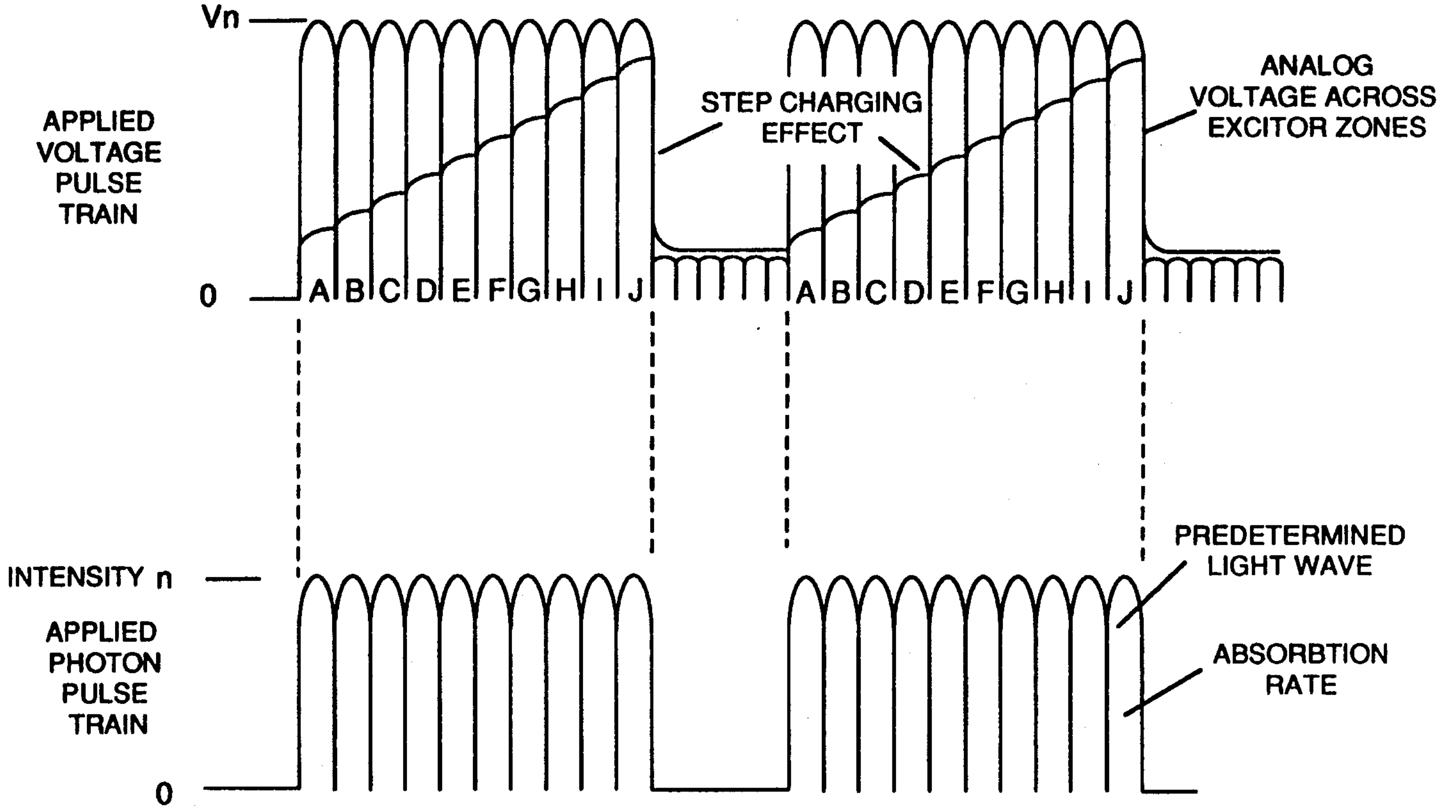
Vn
0
APPLIED VOLTAGE PULSE TRAIN
A B C D E F G H I J
A B C D E F G H I J
STEP CHARGING EFFECT
ANALOG VOLTAGE ACROSS EXCITOR ZONES
INTENSITY n
0
APPLIED PHOTON PULSE TRAIN
PREDETERMINED LIGHT WAVE
ABSORBTION RATE

FIGURE 6A

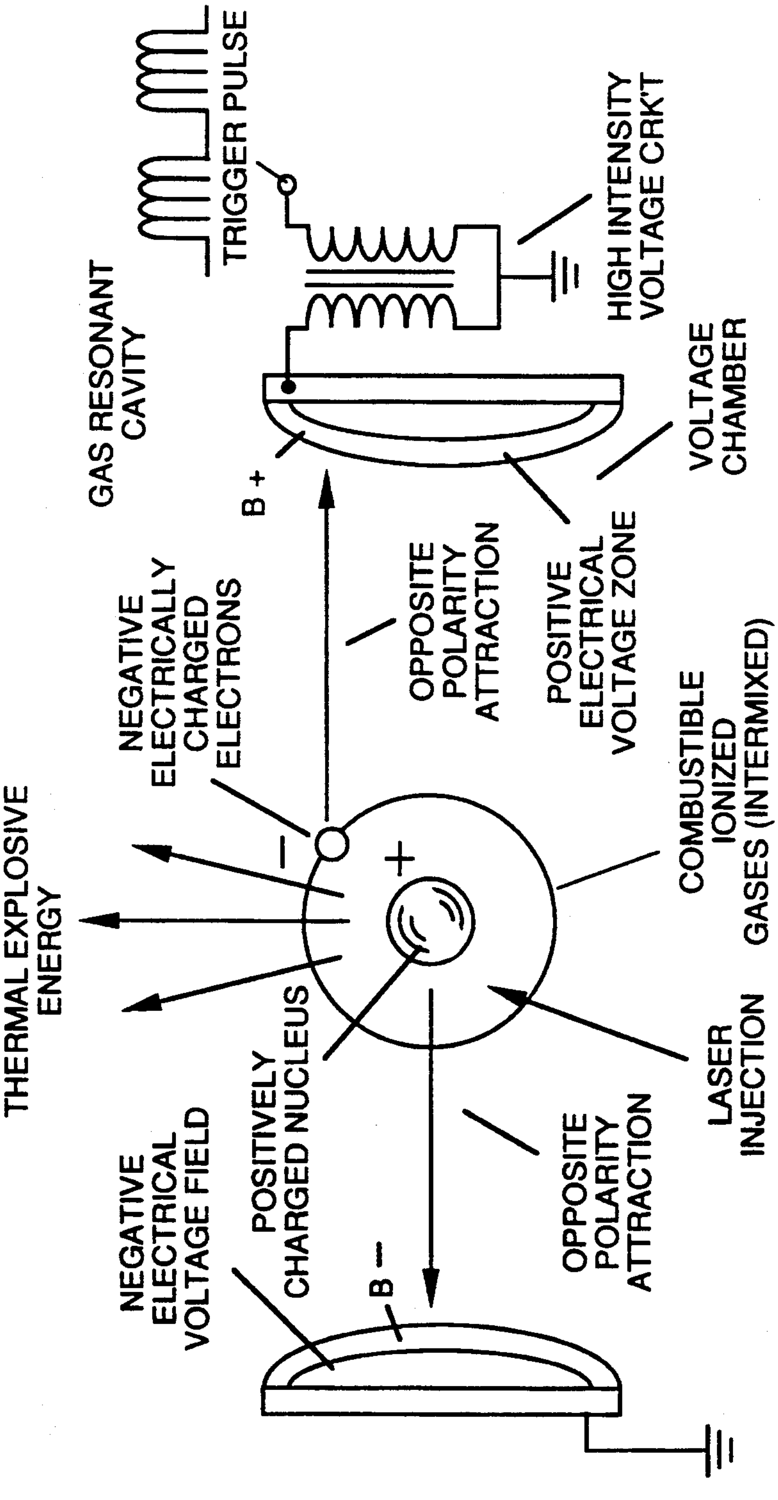
THERMAL EXPLOSIVE ENERGY
NEGATIVE ELECTRICAL VOLTAGE FIELD
POSITIVELY CHARGED NUCLEUS
B −
−
+
NEGATIVE ELECTRICALLY CHARGED ELECTRONS
GAS RESONANT CAVITY
TRIGGER PULSE
B +
OPPOSITE POLARITY ATTRACTION
OPPOSITE POLARITY ATTRACTION
POSITIVE ELECTRICAL VOLTAGE ZONE
HIGH INTENSITY VOLTAGE CRK'T
VOLTAGE CHAMBER
COMBUSTIBLE IONIZED GASES (INTERMIXED)
LASER INJECTION

FIGURE 6B

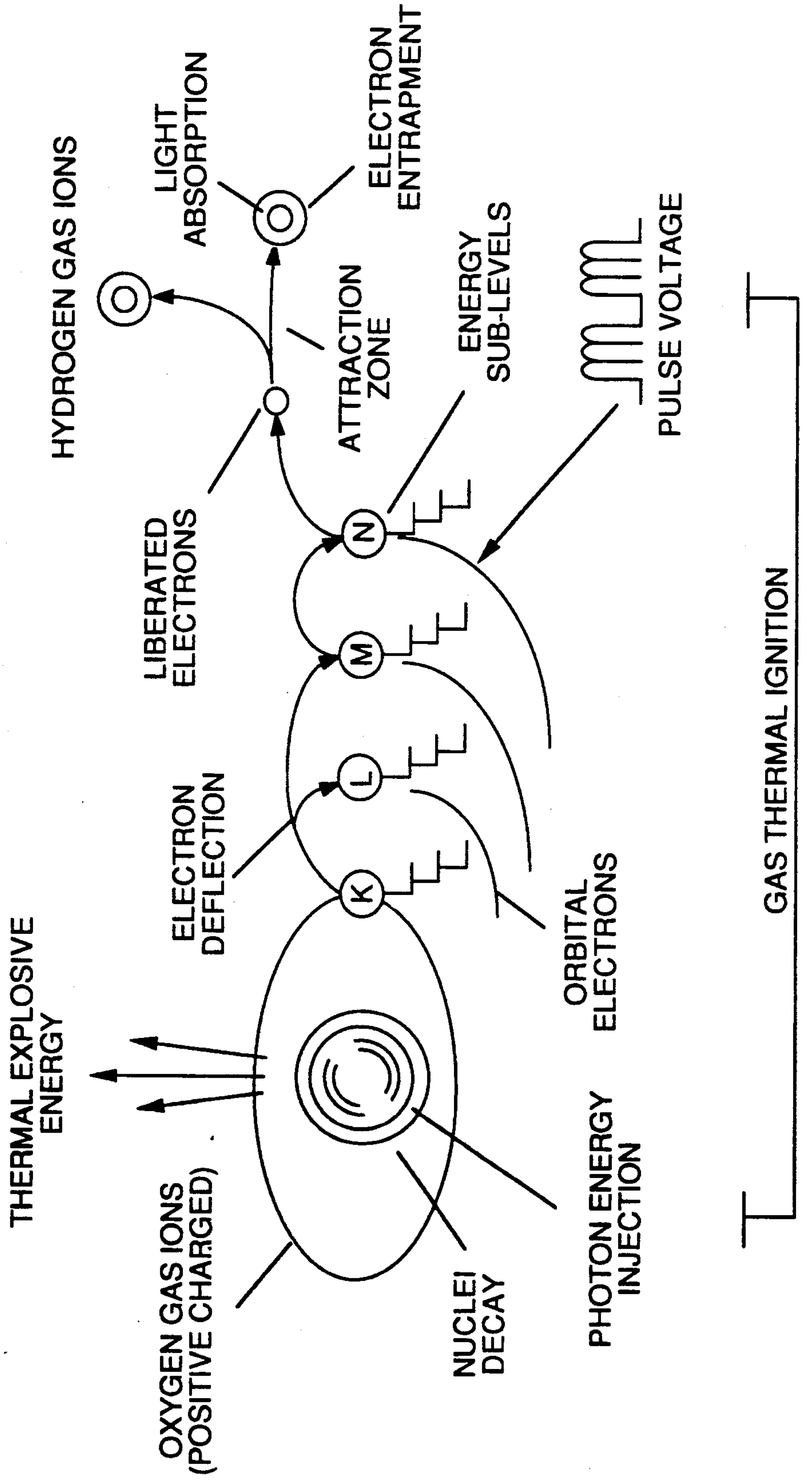
THERMAL EXPLOSIVE ENERGY
OXYGEN GAS IONS (POSITIVE CHARGED)
NUCLEI DECAY
PHOTON ENERGY INJECTION
ELECTRON DEFLECTION
ORBITAL ELECTRONS
K
L
M
N
LIBERATED ELECTRONS
ENERGY SUB-LEVELS
HYDROGEN GAS IONS
ATTRACTION ZONE
LIGHT ABSORPTION
ELECTRON ENTRAPMENT
PULSE VOLTAGE
GAS THERMAL IGNITION

FIGURE 6C

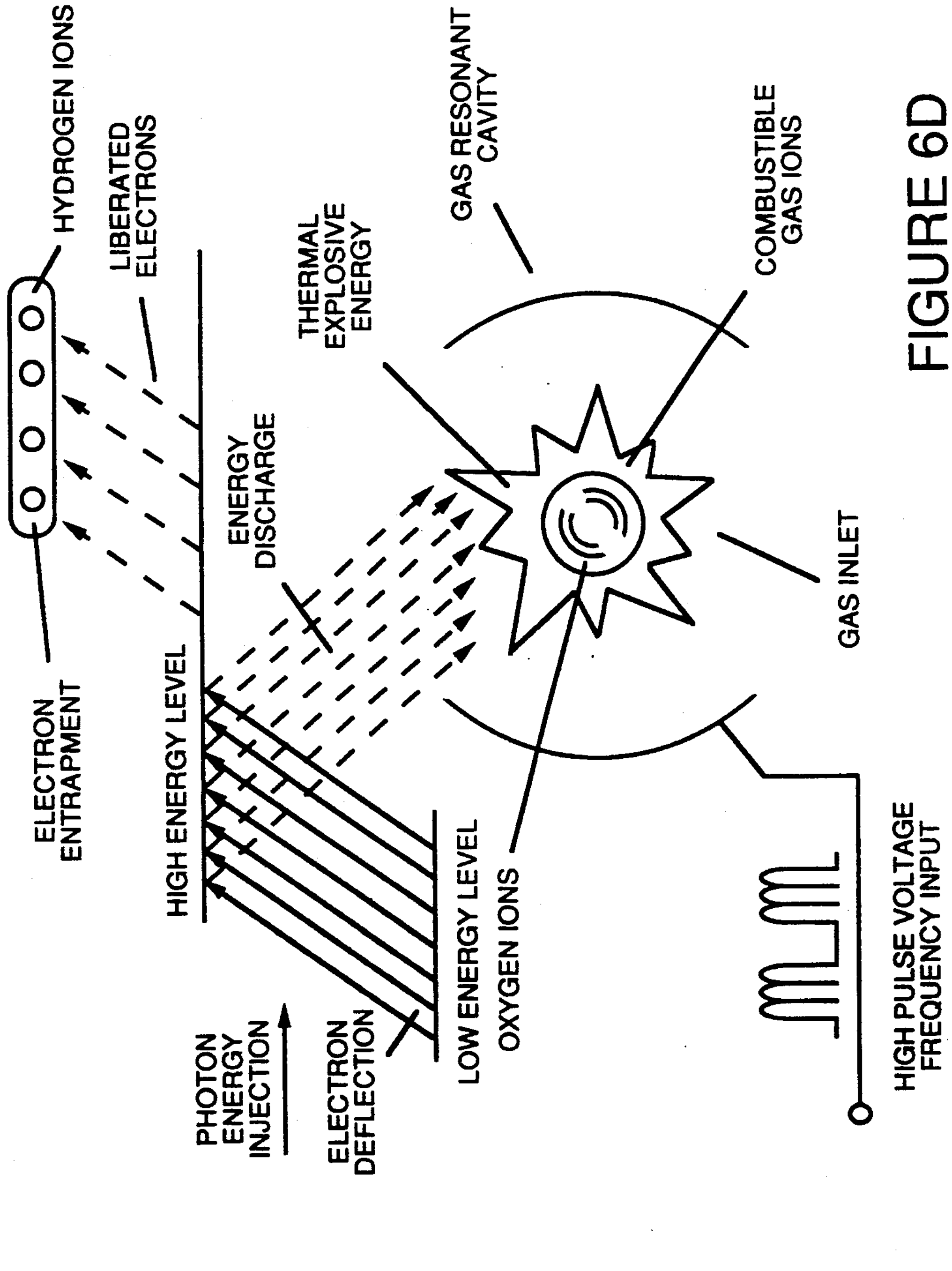
HYDROGEN IONS
ELECTRON ENTRAPMENT
LIBERATED ELECTRONS
ENERGY DISCHARGE
HIGH ENERGY LEVEL
PHOTON ENERGY INJECTION
ELECTRON DEFLECTION
LOW ENERGY LEVEL
THERMAL EXPLOSIVE ENERGY
GAS RESONANT CAVITY
COMBUSTIBLE GAS IONS
GAS INLET
OXYGEN IONS
HIGH PULSE VOLTAGE FREQUENCY INPUT

FIGURE 6D

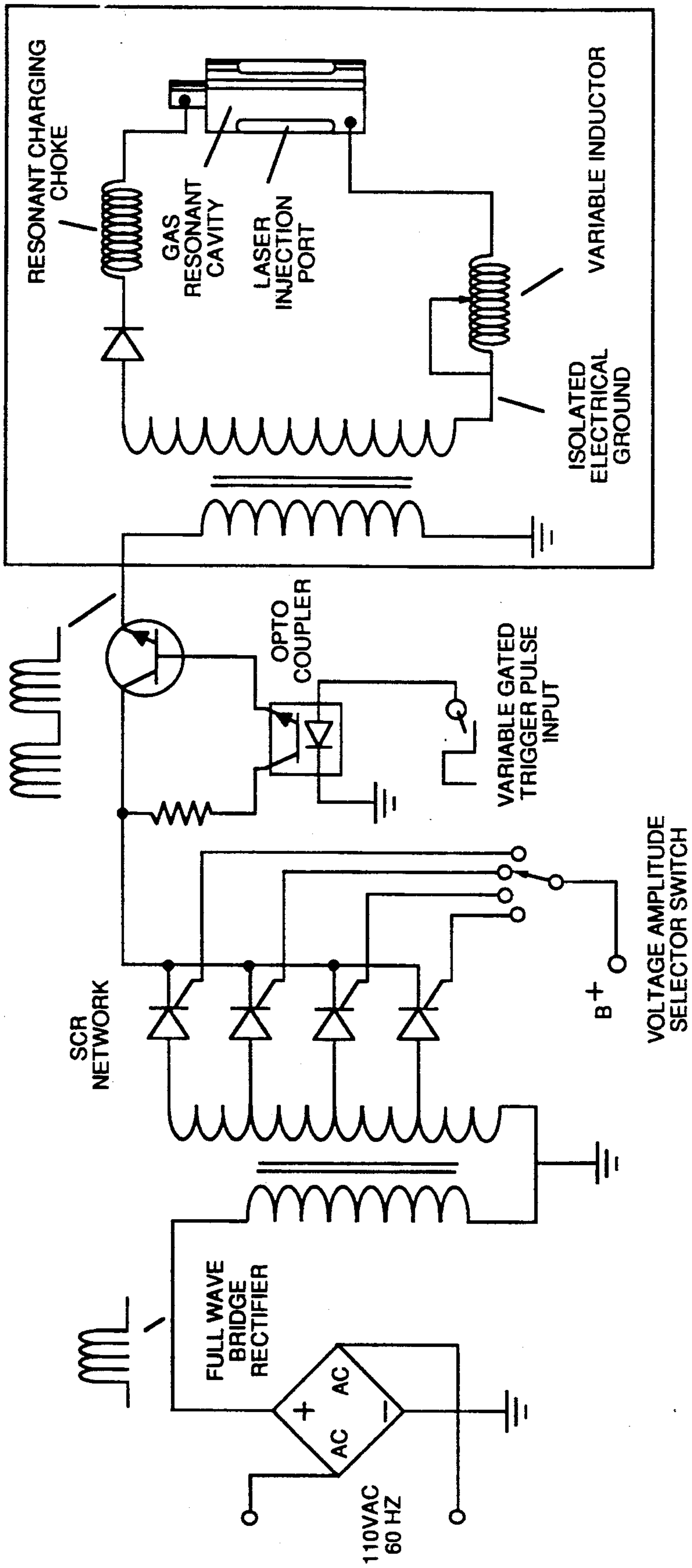
VOLTAGE INTENSIFIER CIRCUIT
RESONANT CHARGING CHOKE
GAS RESONANT CAVITY
LASER INJECTION PORT
VARIABLE INDUCTOR
ISOLATED ELECTRICAL GROUND
OPTO COUPLER
VARIABLE GATED TRIGGER PULSE INPUT
VOLTAGE AMPLITUDE SELECTOR SWITCH
B+
SCR NETWORK
FULL WAVE BRIDGE RECTIFIER
+
AC
AC
−
110VAC 60 HZ

FIGURE 7

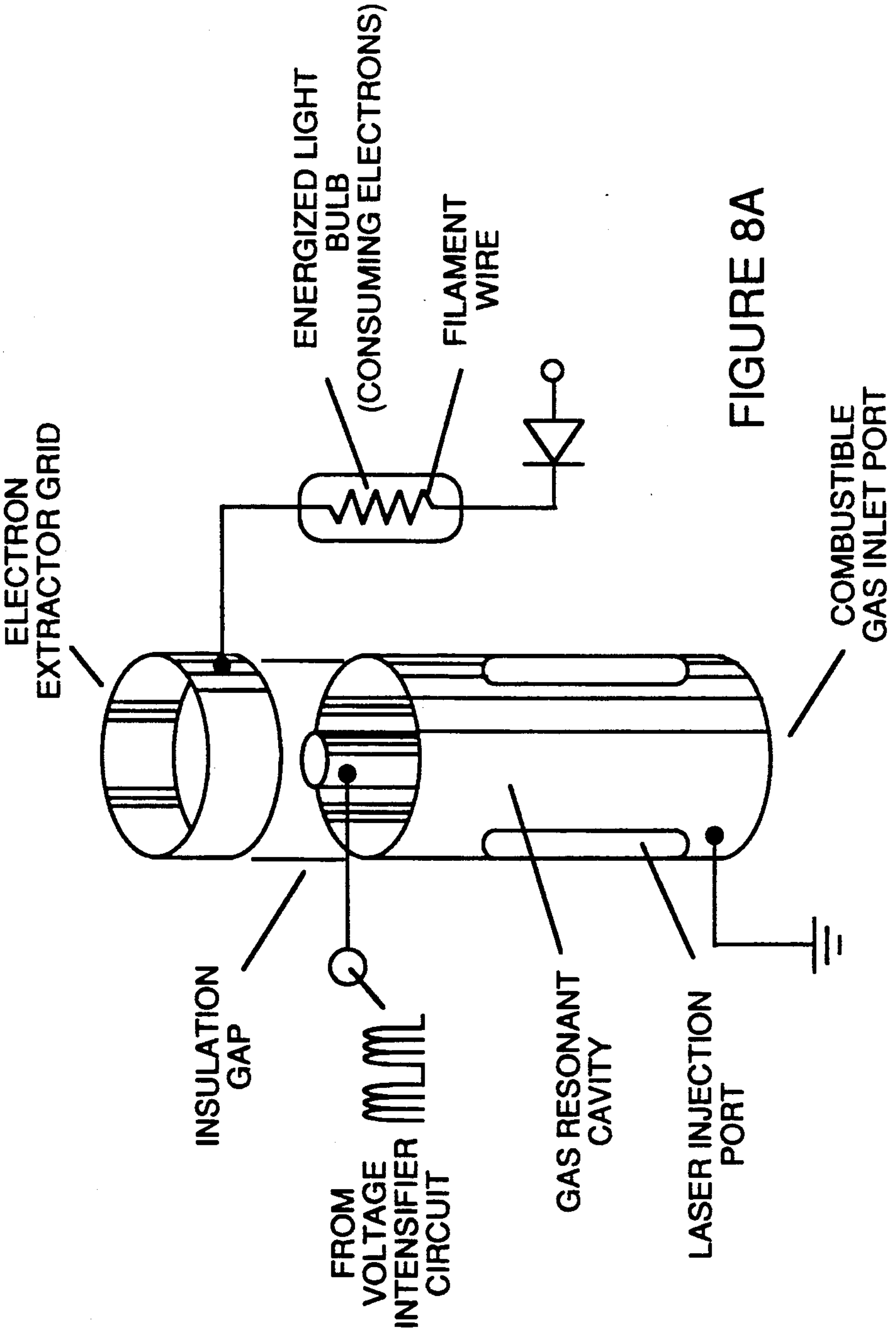
ELECTRON EXTRACTOR GRID
ENERGIZED LIGHT BULB (CONSUMING ELECTRONS)
FILAMENT WIRE
COMBUSTIBLE GAS INLET PORT
INSULATION GAP
FROM VOLTAGE INTENSIFIER CIRCUIT
GAS RESONANT CAVITY
LASER INJECTION PORT

FIGURE 8A

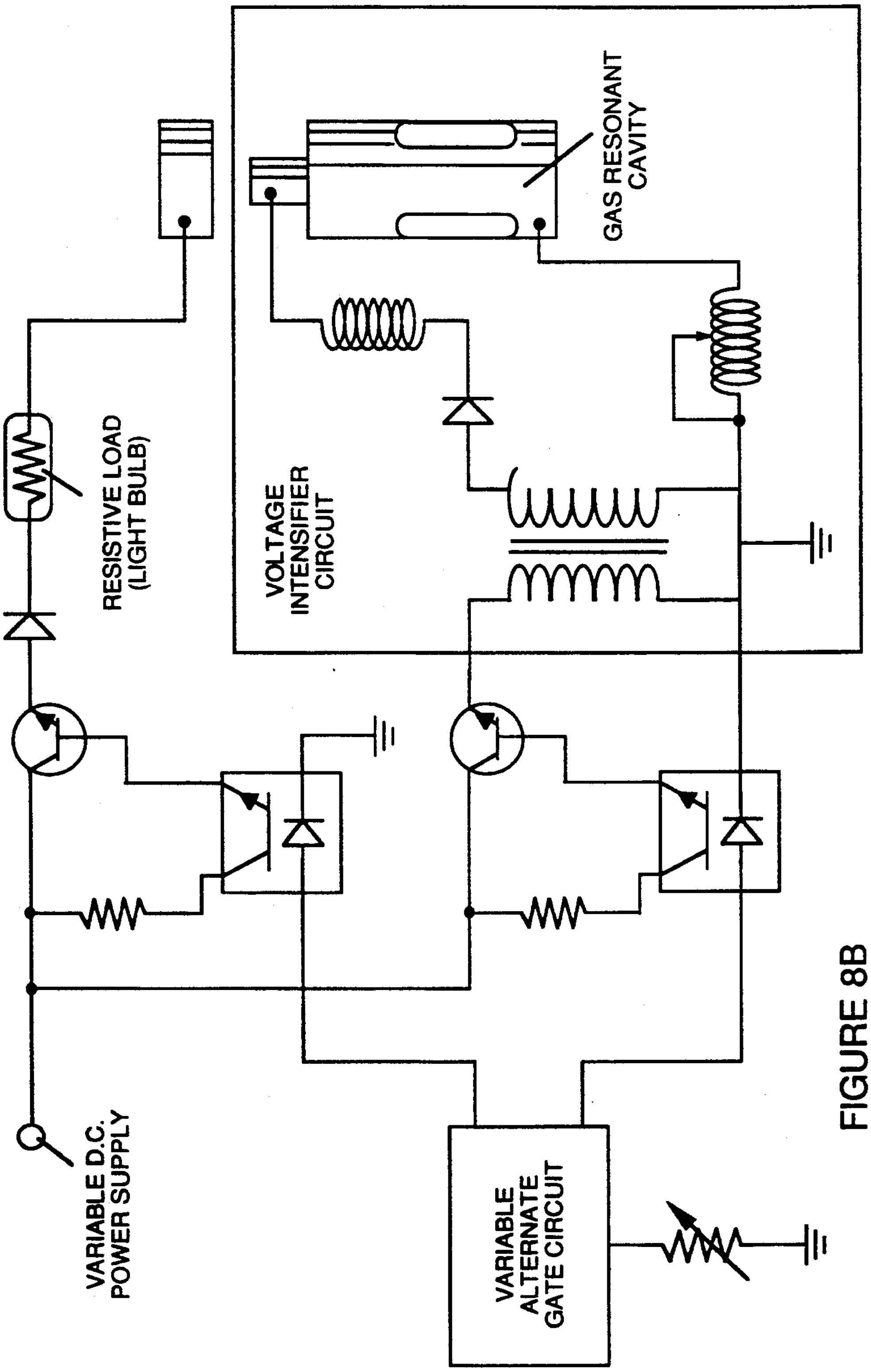
VARIABLE D.C. POWER SUPPLY
RESISTIVE LOAD (LIGHT BULB)
VOLTAGE INTENSIFIER CIRCUIT
GAS RESONANT CAVITY
VARIABLE ALTERNATE GATE CIRCUIT

FIGURE 8B

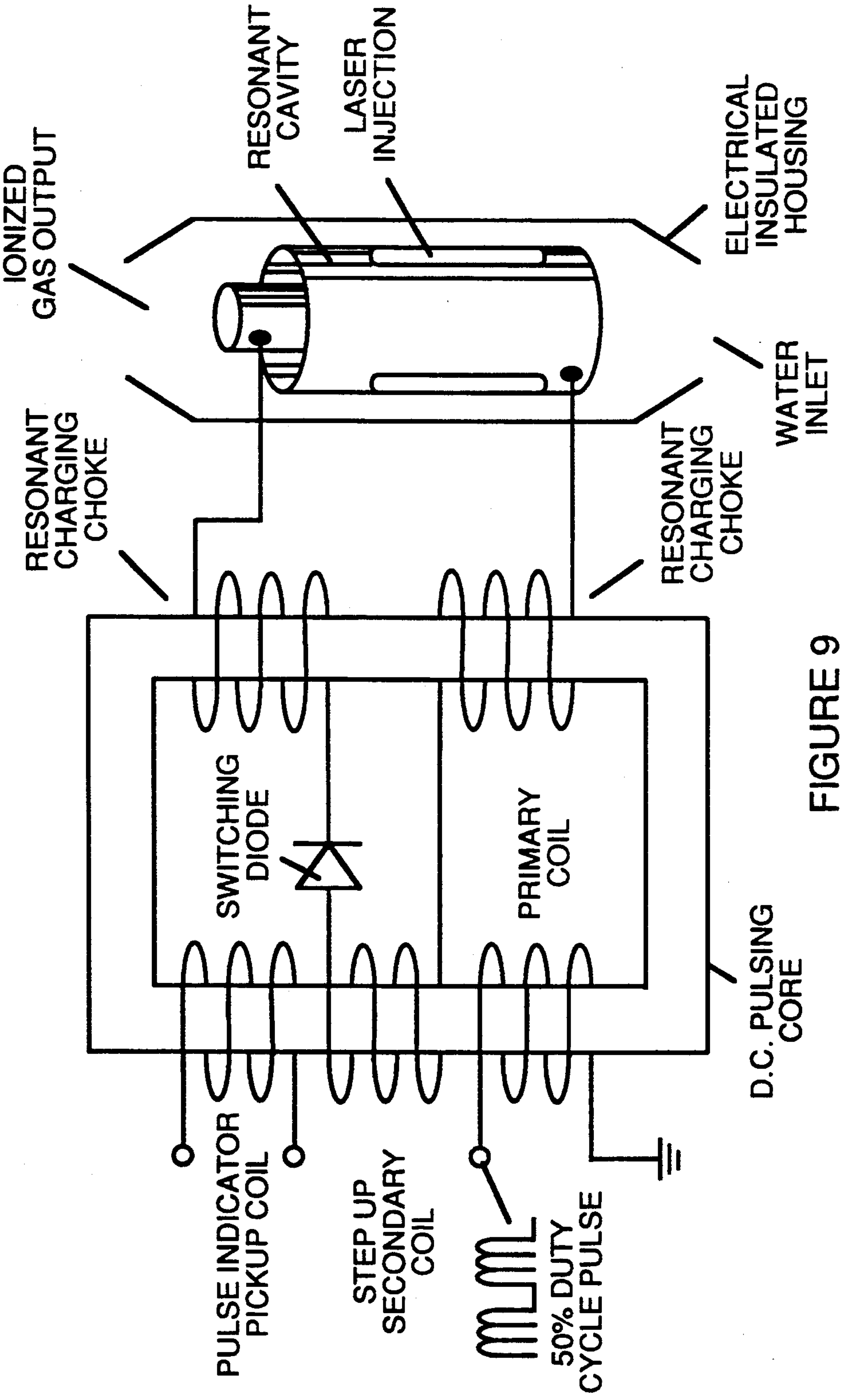
IONIZED GAS OUTPUT
RESONANT CAVITY
LASER INJECTION
ELECTRICAL INSULATED HOUSING
WATER INLET
RESONANT CHARGING CHOKE
RESONANT CHARGING CHOKE
SWITCHING DIODE
PRIMARY COIL
D.C. PULSING CORE
PULSE INDICATOR PICKUP COIL
STEP UP SECONDARY COIL
50% DUTY CYCLE PULSE

FIGURE 9

PROCESS AND APPARATUS FOR THE PRODUCTION OF FUEL GAS AND THE ENHANCED RELEASE OF THERMAL ENERGY FROM SUCH GAS

RELATED APPLICATIONS

This is a continuation-in-part of my co-pending application Ser. No. 207,730 filed Jun. 6, 1988, now U.S. Pat. No. 4,936,961, which in turn was a continuation in part of Ser. No. 081,859, filed Aug. 5, 1987, now U.S. Pat. No. 4,826,581.

FIELD OF THE INVENTION

This invention relates to a method of and apparatus for obtaining the release of a fuel gas mixture including hydrogen and oxygen from water and to a method of and apparatus for obtaining the further release of energy from such a fuel gas mixture. Charged ions derived from the fuel gas are stimulated to an activated state, and then passed through a resonant cavity, where successively increasing energy levels are achieved, and finally passed to an outlet orifice to produce thermal explosive energy.

BACKGROUND OF THE PRIOR ART

Numerous processes have been proposed for separating a water molecule into its elemental hydrogen and oxygen components. Electrolysis is one such process. Other processes are described in U.S. patents such as U.S. Pat. Nos. 4,344,831; 4,184,931; 4,023,545; 3,980,053; and Patent Cooperation Treaty Application No. PCT/US80/1362, published Apr. 30, 1981. Other processes have been proposed for many years in which controlled energy producing reactions of atomic particles are expected to occur under "cold" conditions. [See, e.g., Rafelski, J. and Jones, S. E., "Cold Nuclear Fusion," *Scientific American*, July, 1987, page 84].

Further processes are also described in U.S. Pat. Nos. 4,233,109; 4,406,765; 4,687,753 and 4,695,357. The process and apparatus described herein are considered variations to and improvements in fuel sources and processes by which energy is derived from fuel gas components in a controllable manner.

OBJECTS OF THE INVENTION

A first object of the invention is to provide a fuel cell and a process in which molecules of water are broken down into hydrogen and oxygen gases, and a fuel gas mixture including hydrogen, oxygen and other gasses formerly dissolved within the water is produced. A further object of the invention is to realize significant energy-yield from a fuel gas derived from water (H_2O) molecules. Molecules of water are broken down into hydrogen and oxygen gases. Electrically charged hydrogen and oxygen ions of opposite electrical polarity are activated by electromagnetic wave energy and exposed to a high temperature thermal zone. Significant amounts of thermal energy with explosive force beyond the gas burning stage are released.

An explosive thermal energy under a controlled state is produced. The process and apparatus provide a heat energy source useful for power generation, aircraft, rocket engines, or space stations.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. **1A** through **1F** are illustrations depicting the theoretical bases for phenomena encountered during operation of the fuel gas production stage of the invention herein.

FIG. **2** illustrates a circuit useful in the fuel gas generation process.

FIG. **3** shows a perspective of a "water capacitor" element used in the fuel cell circuit.

FIG. **4** illustrates a staged arrangement of apparatus useful in the process, beginning with a water inlet and culminating in the production of thermal explosive energy.

FIG. **5A** shows a cross-section of a circular gas resonant cavity used in the final stage assembly of FIG. **4**.

FIG. **5B** shows an alternative final stage injection system useful in the apparatus of FIG. **4**.

FIG. **5C** shows an optical thermal lens assembly for use with either final stage of FIG. **5A** or FIG. **5B**.

FIGS. **6A**, **6B**, **6C** and **6D** are illustrations depicting various theoretical bases for atomic phenomena expected to occur during operation of the invention herein.

FIG. **7** is an electrical schematic of the voltage source for the gas resonant cavity.

FIGS. **8A** and **8B**, respectively, show (A) an electron extractor grid used in the injector assemblies of FIG. **5A** and FIG. **5B**, and (B) the electronic control circuit for the extractor grid.

FIG. **9** shows an alternate electrical circuit useful in providing a pulsating waveform to the apparatus.

DESCRIPTION OF THE PREFERRED EMBODIMENT

A fuel gas is produced by a hydrogen fracturing process that follows the sequence of steps shown in the following Table I. Beginning with water molecules, the molecule is subjected to successively increasing electrical, wave energy and thermal forces. In the succession of forces, radomly oriented water molecules are aligned with respect to molecular polar orientation and themselves polarized and "elongated" by the application of an electric potential to the extent that covalent bonding of the water molecule is so weakened that the atoms disassociate and the molecule breaks down into hydrogen and oxygen elemental components. The released atomic gases are next ionized and electrically charged in a vessel while being subjected to a further energy source that promotes inter-particle impact in the gas at an increased overall energy level. Finally, the atomic particles in the excited gas, having achieved successively higher energy levels, are subjected to a laser or electromagnetic wave energy source that produces atomic destabilization and the final release of thermal explosive energy. Engineering design parameters based on known theoretical principles of atomic physics determine the incremental levels of electrical and wave energy input required to produce resonance in each stage of the system. Instead of a dampening effect, a resonant energization of the molecule, atom or ion provides a compounding energy interaction resulting in the final energy release.

TABLE I

PROCESS STEPS LEADING TO IGNITION

RELATIVE STATE OF WATER MOLECULE AND/OR HYDROGEN/OXYGEN/OTHER ATOMS	1st Stage
RANDOM (AMBIENT STATE)	Water to Gas
ALIGNMENT OF POLAR FIELDS	
POLARIZATION OF MOLECULES	

TABLE I-continued

PROCESS STEPS LEADING TO IGNITION	
MOLECULAR ELONGATION	
ATOM LIBERATION BY BREAKDOWN OF COVALENT BOND	
	2nd Stage
RELEASE OF GASES	Gas Ionization
LIQUID TO GAS IONIZATION	
ELECTRICAL CHARGING EFFECT	
PARTICLE IMPACT	
	3rd Stage
ELECTROMAGNETIC WAVE, LASER OR PHOTON INJECTION	Priming
ELECTRON EXTRACTION	
ATOMIC DESTABILIZATION	
	Final Stage
THERMAL IGNITION	Ignition

In brief, in the first stage a gas mixture including hydrogen and oxygen and other dissolved gases formerly entrapped in water is obtained, from water. In general, the method used in the first stage consists of: (A) providing a capacitor, in which the water is included as a dielectric liquid between capacitor plates, in a resonant charging choke circuit that includes an inductance in series with the capacitor; (B) subjecting the capacitor to a pulsating, unipolar electric voltage field in which the polarity does not pass beyond an arbitrary ground, whereby the water molecules within the capacitor are subjected to a charge of the same polarity and the water molecules are distended by their subjection to electrical polar forces; (C) further subjecting the water in said capacitor to said pulsating electric field to achieve a pulse frequency such that the pulsating electric field induces a resonance within the water molecule; (D) continuing the application of the pulsing frequency to the capacitor cell after resonance occurs so that the energy level within the molecule is increased in cascading incremental steps in proportion to the number of pulses; (E) maintaining the charge of said capacitor during the application of the pulsing field, whereby the co-valent electrical bonding of the hydrogen and oxygen atoms within said molecules is destabilized such that the force of the electrical field applied, as the force is effective within the molecule, exceeds the bonding force of the molecule, and hydrogen and oxygen atoms are liberated from the molecule as elemental gases; and (F) collecting said hydrogen and oxygen gases, and any other gases that were formerly dissolved within the water, and discharging the collected gases as a fuel gas mixture.

The water molecules are subjected to increasing electrical forces. In an ambient state, randomly oriented water molecules are aligned with respect to a molecular polar orientation. They are next, themselves polarized and "elongated" by the application of an electric potential to the extent that covalent bonding of the water molecule is so weakened that the atoms disassociate and the molecule breaks down into hydrogen and oxygen elemental components. In the process, the point of optimum gas release is reached at a circuit resonance. Water in the fuel cell is subjected to a pulsating, polar electric field produced by the electrical circuit whereby the water molecules are distended by reason of their subjection to electrical polar forces of the capacitor plates. The polar pulsating frequency applied is such that the pulsating electric field induces a resonance in the molecule. A cascade effect occurs and the overall energy level of specific water molecules is increased in cascading, incremental steps. The hydrogen and oxygen atomic gases, and other gas components formerly entrapped as dissolved gases in water, are released when the resonant energy exceeds the co-valent bonding force of the water molecule. A preferred construction material for the capacitor plates is a stainless steel T-304 which is non-chemically reactive with water, hydrogen, or oxygen. An electrically conductive material which is inert in the fluid environment is a desirable material of construction for the electric field plates of the "water capacitor" employed in the circuit.

Once triggered, the gas output is controllable by the attenuation of operational parameters. Thus, once the frequency of resonance is identified, by varying the applied pulse voltage to the water fuel cell assembly, gas output is varied. By varying the pulse shape and/or amplitude or pulse train sequence of the initial pulsing wave source, final gas output is varied. Attenuation of the voltage field frequency in the form of OFF and ON pulses likewise affects output.

The overall apparatus thus includes an electrical circuit in which a water capacitor having a known dielectric property is an element. The fuel gases are obtained from the water by the disassociation of the water molecule. The water molecules are split into component atomic elements (hydrogen and oxygen gases) by a voltage stimulation process called the electrical polarization process which also releases dissolved gases entrapped in the water.

From the outline of physical phenomena associated with the first stage of the process described in Table 1, the theoretical basis of the invention considers the respective states of molecules and gases and ions derived from liquid water. Before voltage stimulation, water molecules are randomly dispersed throughout water within a container. When a unipolar voltage pulse train such as shown in FIGS. **1B** through **1F** is applied to positive and negative capacitor plates, an increasing voltage potential is induced in the molecules in a linear, step-like charging effect. The electrical field of the particles within a volume of water including the electrical field plates increases from a low energy state to a high energy state successively in a step manner following each pulse-train as illustrated figuratively in the depictions of FIGS. **1A** through **1F**. The increasing voltage potential is always positive in direct relationship to negative ground potential during each pulse. The voltage polarity on the plates which create the voltage fields remains constant although the voltage charge increases. Positive and negative voltage "zones" are thus formed simultaneously in the electrical field of the capacitor plates.

In the first stage of the process describe in Table 1, because the water molecule naturally exhibits opposite electrical fields in a relatively polar configuration (the two hydrogen atoms are positively electrically charged relative to the negative electrically charged oxygen atom), the voltage pulse causes initially randomly oriented water molecules in the liquid state to spin and orient themselves with reference to positive and negative poles of the voltage fields applied. The positive electrically charged hydrogen atoms of said water molecule are attracted to a negative voltage field; while, at the same time, the negative electrically charged oxygen atoms of the same water molecule are attracted to a positive voltage field. Even a slight potential difference applied to inert, conductive plates of a containment chamber which forms a capacitor will initiate polar atomic orientation within the water molecule based on polarity differences.

When the potential difference applied causes the orientated water molecules to align themselves between the conductive plates, pulsing causes the voltage field intensity to be increased in accordance with FIG. **1B**. As further molecular alignment occurs, molecular movement is hindered. Because the positively charged hydrogen atoms of said aligned molecules are attracted in a direction opposite to the negatively charged oxygen atoms, a polar charge alignment or distribution occurs within the molecules between said voltage zones, as shown in FIG. **1B**. And as the energy level of the atoms subjected to resonant pulsing increases, the stationary water molecules become elongated as shown in FIGS. **1C** and **1D**. Electrically charged nuclei and electrons are attracted toward opposite electrically charged voltage zones—disrupting the mass and charge equilibrium of the water molecule.

As the water molecule is further exposed to an increasing potential difference resulting from the step charging of the capacitor, the electrical force of attraction of the atoms within the molecule to the capacitor plates of the chamber also increases in strength. As a result, the co-valent bonding between atoms which form the molecule is weakened—and ultimately terminated. The negatively charged electron is attracted toward the positively charged hydrogen atoms, while at the same time, the negatively charged oxygen atoms repel electrons.

In a more specific explanation of the "sub-atomic" action that occurs in the water cell that provides a fuel gas for the subsequent stages, it is known that natural water is a liquid which has a dielectric constant of 78.54 at 20° C. and 1 atm pressure. [*Handbook of Chemistry and Physics*, 68th ed., CRC Press (Boca Raton, Fla. (1987–88)), *Section E*-50, H_2O (water)].

When a volume of water is isolated and electrically conductive plates, that are chemically inert in water and are separated by a distance, are immersed in the water, a capacitor is formed, having a capacitance determined by the surface area of the plates, the distance of their separation and the dielectric constant of water.

When water molecules are exposed to voltage at a restricted current, water takes on an electrical charge. By the laws of electrical attraction, molecules align according to positive and negative polarity fields of the molecule and the alignment field. The plates of a capacitor constitute such an alignment field when a voltage is applied.

When a charge is applied to a capacitor, the electrical charge of the capacitor equals the applied voltage charge; in a water capacitor, the dielectric property of water resists the flow of amps in the circuit, and the water molecule itself, because it has polarity fields formed by the relationship of hydrogen and oxygen in the covalent bond, and an intrinsic dielectric property, becomes part of the electrical circuit, analogous to a "microcapacitor" within the capacitor defined by the plates.

In the Example of a fuel cell circuit of FIG. **2**, a water capacitor is included. The step-up coil is formed on a conventional torroidal core formed of a compressed ferromagnetic powdered material that will not itself become permanently magnetized, such as the trademarked "Ferramic 06# "Permag" powder as described in *Siemens Ferrites Catalog*, CG-2000-002-121, (Cleveland, Ohio) No. F626-1205. The core is 1.50 inch in diameter and 0.25 inch in thickness. A primary coil of 200 turns of 24 gauge cooper wire is provided and a coil of 600 turns of 36 gauge wire comprises the secondary winding. Other primary/secondary coil winding ratios may be conventionally determined.

An alternate coil arrangement using a conventional M27 iron transformer core is shown in FIG. **9**. The coil wrap is always in one direction only.

In the circuit of FIG. **2**, the diode is a 1N1198 diode which acts as a blocking diode and an electric switch that allows voltage flow in one direction only. Thus, the capacitor is never subjected to a pulse of reverse polarity.

The primary coil of the torroid is subject to a 50% duty cycle pulse. The torroidal pulsing coil provides a voltage step-up from the pulse generator in excess of five times, although the relative amount of step-up is determined by pre-selected criteria for a particular application. As the stepped-up pulse enters first inductor (formed from 100 turns of 24 gauge wire 1 inch in diameter), an electromagnetic field is formed around the inductor, voltage is switched off when the pulse ends, and the field collapses and produces another pulse of the same polarity; i.e., another positive pulse is formed where the 50% duty cycle was terminated. Thus, a double pulse frequency is produced; however, in a pulse train of unipolar pulses, there is a brief time when pulses are not present.

By being so subjected to electrical pulses in the circuit of FIG. **2**, water confined in the volume that includes the capacitor plates takes on an electrical charge that is increased by a step charging phenomenon occurring in the water capacitor. Voltage continually increases (to about 1000 volts and more) and the water molecules starts to elongate.

The pulse train is then switched off; the voltage across the water capacitor drops to the amount of charge that the water molecules have taken on, i.e. voltage is maintained across the charged capacitor. The pulse train is then reapplied.

Because a voltage potential applied to a capacitor can perform work, the higher the voltage potential, the more work is performed by a given capacitor. In an optimum capacitor that is wholly non-conductive, zero (0) current flow will occur across the capacitor. Thus, in view of an idealized capacitor circuit, the object of the water capacitor circuit is to prevent electron flow through the circuit, i.e. such as occurs by electron flow or leakage through a resistive element that produces heat. Electrical leakage in water will occur, however, because of some residual conductivity and impurities or ions that may be otherwise present in the water. Thus, the water capacitor is preferably chemically inert. An electrolyte is not added to the water.

In the isolated water bath, the water molecule takes on charge, and the charge increases. The object of the process is to switch off the co-valent bonding of the water molecule and interrupt the sub-atomic force, i.e. the electrical force or electromagnetic force, that binds the hydrogen and oxygen atoms to form a molecule so that the hydrogen and oxygen separate.

Because an electron will only occupy a certain electron shell (the shells are well known) the voltage applied to the capacitor affects the electrical forces inherent in the co-valent bond. As a result of the charge applied by the plates, the applied force becomes greater than the force of the co-valent bonds between the atoms of the water molecule; and the water molecule becomes elongated. When this happens, the time share ratio of the electrons between the atoms and the electron shells is modified.

In the process, electrons are extracted from the water bath; electrons are not consumed nor are electrons introduced into the water bath by the circuit as electrons are conventionally introduced in an electrolysis process. There may nevertheless occur a leakage current through the water. Those hydrogen atoms missing electrons become neutralized; and atoms are liberated from the water. The charged atoms and electrons are attracted to opposite polarity voltage zones created between the capacitor plates. The electrons formerly shared by atoms in the water co-valent bond are reallocated such that neutral elemental gases are liberated.

In the process, the electrical resonance may be reached at all levels of voltage potential. The overall circuit is characterized as a "resonant charging choke" circuit which is an inductor in series with a capacitor that produces a resonant circuit. [*SAMS Modern Dictionary of Electronics*, Rudolff Garff, ©1984, Howard W. Sams & Co. (Indianapolis, Ind.), p. 859.] Such a resonant charging choke is on each side of the capacitor. In the circuit, the diode acts as a switch that allows the magnetic field produced in the inductor to collapse, thereby doubling the pulse frequency and preventing the capacitor from discharging. In this manner a continuous voltage is produced across the capacitor plates in the water bath; and the capacitor does not discharge. The water molecules are thus subjected to a continuously charged field until the breakdown of the co-valent bond occurs.

As noted initially, the capacitance depends on the dielectric properties of the water and the size and separation of the conductive elements forming the water capacitor.

EXAMPLE I

In an example of the circuit of FIG. 2 (in which other circuit element specifications are provided above), two concentric cylinders 4 inches long formed the water capacitor of the fuel cell in the volume of water. The outside cylinder was 0.75 inch in outside diameter; the inner cylinder was 0.5 inch in outside diameter. Spacing from the outside of the inner cylinder to the inner surface of the outside cylinder was 0.0625 inch. Reasonance in the circuit was achieved at a 26 volt applied pulse to the primary coil of the torroid at 10 KH_z, and the water molecules disassociated into elemental hydrogen and oxygen and the gas released from the fuel cell comprised a mixture of hydrogen, oxygen from the water molecule, and gases formerly dissolved in the water such as the atmospheric gases or oxygen, nitrogen, and argon.

In achieving resonance in any circuit, as the pulse frequency is adjusted, the flow of amps is minimized and voltage is maximized to a peak. Calculation of the resonance frequency of an overall circuit is determined by known means; different cavities have a different frequency of resonance dependent on parameters of the water dielectric, plate size, configuration and distance, circuit inductors, and the like. Control of the production of fuel gas is determined by variation of the period of time between a train of pulses, pulse amplitude and capacitor plate size and configuration, with corresponding value adjustments to other circuit components.

The wiper arm on the second inductor tunes the circuit and accommodates to contaminants in water so that the charge is always applied to the capacitor. The voltage applied determines the rate of breakdown of the molecule into its atomic components. As water in the cell is consumed, it is replaced by any appropriate means or control system.

Thus in the first stage, which is of itself independently useful, a fuel gas mixture is produced having, in general, the components of elemental hydrogen and oxygen as well as formerly dissolved entrapped atmospheric gases such as nitrogen, argon, and the like. The fuel gas is itself combustible in a conventional manner.

After the first stage the gas atoms become elongated during electron removal as the atoms are ionized. Laser, or light wave energy of a predetermined frequency is injected into a containment vessel in a gas ionization process. The light energy absorbed by voltage stimulated gas nuclei causes destabilization of gas ions still further. The absorbed laser energy causes the gas nuclei to increase in energy state, which, in turn, causes electron deflection to a higher orbital shell.

The electrically charged and laser primed combustible gas ions from a gas resonant cavity may be directed into an optical thermal lens assembly for triggering. Before entry into the optimal thermal lens, however, electrons are stripped from the ions and the atom is destabilized. The destabilized gas ions which are electrically and mass unbalanced atoms having highly energized nuclei are pressurized during spark ignition. The unbalanced, destablized atomic components thermally interact; the energized and unstable hydrogen gas nuclei collide with highly energized and unstable oxygen gas nuclei, causing and producing thermal explosive energy beyond the gas burning stage. The ambient air gas components in the initial mixture aid the thermal explosive process under a controlled state.

In the process, the point of optimum energy-yield is reached when the electron deficient oxygen atoms (having less than a normal number of electrons) lock onto and capture a hydrogen atom electron prior to or during thermal combustion of the hydrogen/oxygen mixture. Atomic decay results in the release of energy.

After the first stage, the gas mixture is subjected to a pulsating, polar electric field whereby electrons of the gas atoms are distended in their orbital fields by reason of their subjection to electrical polar forces The polar pulsating frequency applied is such that the pulsating electric field induces a resonance with respect to an electron of the gas atom. A cascade effect results and the energy level of specific resonating electron is increased in cascading, incremental steps.

Next, the gas atoms are ionized and subjected to electro-magnetic wave energy having a predetermined frequency to induce a further election resonance in the ion, whereby the energy level of the election is successively increased. Electrons are extracted from the resonating ions while such ions are in an increased energy state to destabilize the nuclear electron configuration of said ions; and the gas mixture of destabilized ions is thermally ignited.

In the apparatus shown in FIG. 4, water is introduced at inlet 1 into a first stage water fracturing module 2, such as the water fuel cell described above, in which water molecules are broken down into hydrogen, oxygen and released entrapped gas components. The released atomic gases and other gas components formerly entrapped as dissolved gases in water may be introduced to a successive stage 3 or other number of like resonant cavities, which are arranged in either a series of parallel combined array. The successive energization of the gas atoms provides a cascading effect, successively increasing the voltage stimulation level of the released gasses as they sequentially pass through cavities **2**, **3**, etc. In a final stage, an injector system **4**, of a configuration of the type shown in FIGS. **5**A or **5**B, receives energized atomic and gas particles where the particles are subjected to further energy input, electrical excitation and thermal stimulation, whereby thermal explosive energy results **5**, which may be directed thru a lens assembly of the type shown in FIG. **5**C to provide a controlled thermal energy output.

A single cell, or a battery of cells such as shown in FIG. 3, provides a fuel gas source for stages after the first stage. The fuel gas is activated by electromagnetic waves, and electrically charged gas ions of hydrogen and oxygen (of opposite polarity) are expelled from the cascaded cells **2**, **3**, etc. shown in FIG. **4**. The circuit of FIG. **9** may be utilized as a source of ionizing energy for the gases. The effect of cascading successively increases the voltage stimulation level of the released gases, which then are directed to the final injector assembly **4**. In the injector assembly, gas ions are stimulated to a yet higher energy level. The gases are continually exposed to a pulsating laser or other electromagnetic wave energy source together with a high intensity oscillating voltage field that occurs within the cell between electrodes or conductive plates of opposite electrical polarity. A preferred construction material for the plates is a stainless steel T-304 which is non-chemically reactive with water, hydrogen, or oxygen. An electrically conductive material which is insert in the fluid environment is a desirable material of construction for the electrical field producing plates, through which field the gas stream of activated particles passes. Gas ions of opposite electrical charges reach and maintain a critical energy level state. The gas ions are oppositely electrically charged and subjected to oscillating voltage fields of opposite polarity and are also subjected to a pulsating electromagnetic wave energy source. Immediately after reaching critical energy, the excited gas ions are exposed to a high temperature thermal zone in the injection cell, **4**, that causes the excited gas ions to undergo gas combustion. The gas ignition triggers atomic decay and releases thermal energy, **5**, with explosive force.

Once triggered, the thermal explosive energy output is controllable by the attenuation of operational parameters. With reference to FIG. **6**A, for example, once the frequency of resonance is identified, by varying applied pulse voltage to the initial water fuel cell assemblies, **2**, **3**, the ultimate explosive energy output is likewise varied. By varying the pulse shape and/or amplitude or pulse train sequence of the electromagnetic wave energy source, final output is varied. Attenuation of the voltage field frequency in the form of OFF and ON pulses likewise affects output of the staged apparatus. Each control mechanism can be used separately, grouped in sections, or systematically arranged in a sequential manner.

A complete system in accordance with the present application thus includes a water fuel cell for providing a first fuel gas mixture consisting of at least a portion of hydrogen and oxygen gas. An electrical circuit of the type shown in FIG. **7** provides a pulsating, polar electric field to the gas mixture as illustrated in FIG. **6**A, whereby electrons of the gas atoms are distended in their orbital fields by reason of their subjection to electrical polar forces, changing from the state conceptually illustrated by FIG. **6**B to that of FIG. **6**C, at a frequency such that the pulsating electric field induces a resonance with respect to electrons of the gas atoms. The energy level of the resonant electrons is thereby increased in cascading, incremental steps. A further electric field to ionize said gas atoms is applied and an electromagnetic wave energy source for subjecting the ionized gas atoms to wave energy of a predetermined frequency to induce a further electron resonance in the ion, whereby the energy level of the election is successively increased is an additional element of the apparatus as shown in FIG. **6**D.

An electron sink, which may be in the form of the grid element shown in FIG. **8**A, extracts further electrons from the resonating ions while such ions are in an increased energy state and destabilizes the nuclear electron configuration of the ions. The "extraction" of electrons by the sink means is coordinated with the pulsating electrical field of the resonant cavity produced by the circuit of FIG. **7**, by means of an interconnected synchronization circuit, such as shown in FIG. **8**B. A nozzle, **10** in FIG. **5**B, or thermal lens assembly, FIG. **5**C, provides the directing means in which the destabilized ions are finally thermally ignited.

As previously noted, to reach and trigger the ultimate atomic decay of the fuel cell gases at the final stage, sequential steps are taken. First, water molecules are slit into component atomic elements (hydrogen and oxygen gases) by a voltage stimulation process which also releases dissolve gases entrapped in the water. In the injector assembly, a laser produced light wave or other form of coherent electromagnetic wave energy capable of stimulating a resonance within the atomic components is absorbed by the mixture of gases (hydrogen/oxygen/ambient air gases) released by the polarization process. At this point, as shown in FIG. **6**B, the individual atoms are subjected to an electric field to begin an ionization process.

The laser or electromagnetic wave energy is absorbed and causes gas atoms to lose electrons and form positively charged gas ions. The energized hydrogen atoms which, as ionized, are positively charged, now accept electrons liberated from the heavier gases and attract other negatively charged gas ions as conceptually illustrated in FIG. **6**C. Positively and negatively charged gas ions are re-exposed to further pulsating energy sources to maintain random distribution of ionized atomic gas particles.

The gas ions within the wave energy chamber are subjected to an oscillating high intensity voltage field in a chamber **11** in FIGS. **5**A and **5**B formed within electrodes **12** and **13** in FIGS. **5**A and **5**B of opposite electrical polarity to produce a resonant cavity. The gas ions reach a critical energy state at a resonant state.

At this point, within the chamber, additional electrons are attracted to said positive electrode; whereas, positively charged ions or atomic nuclei are attracted to the negative electrode. The positive and negative attraction forces are co-ordinate and operate on said gas ions simultaneously; the attraction forces are non-reversible. The gas ions experience atomic component deflection approaching the point of electron separation. At this point electrons are extracted from the chamber by a grid system such as shown in FIG. **5**A. The extracted electrons are consumed and prevented from re-entering the chamber by a circuit such as shown in FIG. **8**B. The elongated gas ions are subjected to a thermal heat zone to cause gas ignition, releasing thermal energy with explosive force. During ionic gas combustion, highly energized and stimulated atoms and atom nuclei collide and explode during thermal excitation. The hydrogen fracturing process occurring sustains and maintains a thermal zone, at a temperature in excess of normal hydrogen/oxygen combustion temperature, to wit, in excess of 2500° F. To cause and maintain atomic elongation depicted in FIG. **6**C before gas ignition, a voltage intensifier circuit such as shown in FIG. **7** is utilized as a current restricting voltage source to provide the excitation voltage applied to the resonant cavity. At the same time the interconnected electron extractor circuit, FIG. **8**B, prevents the reintroduction of electrons back into the system. Depending on calculated design parameters, a predetermined voltage and frequency range may be designed for any particular application or physical configuration of the apparatus.

In the operation of the assembly, the pulse train source for the gas resonant cavity shown at **2** and **3** in FIG. **4** may be derived from a circuit such as shown in FIGS. **2**, **7** or **9**, and such cavity circuits may be in sequence to provide a cascading energy input. It is necessary in the final electron extraction that the frequency with which electrons are removed from the system by sequenced and synchronized with the pulsing of the gas resonant cavity. In the circuit of FIG. **8**B, the coordination of synchronization of the circuit with the circuit of FIG. **7** may be achieved by interconnecting point "A" of the gate circuit of FIG. **8**B to coordinate point "A" of the pulsing circuit of FIG. **7**.

The circuit shown in FIG. **9** enhances the voltage potential across the resonant charging choke coils during pulsing operations and restricts amp flow by allowing an external electromagnetic pulsing field, F, derived from the primary coil A being energized to transverse the coil windings D and E being energized by the incoming pulse train Ha xxx Hn, through switching diode G. The external pulse field, F and the incoming pulse-train Ha xxx Hn, are sequentially the same, allowing resonant action to occur, restricting amp flow while allowing voltage intensity to increase to stimulate the electrical polarization process, the gas ionization process and the electron extraction process. The voltage intensifier circuit of FIG. **9** prevents electrons from entering into those processes.

Together, the hydrogen injector assembly **4** and the resonant cavity assemblies **2**, **3** form a gas injector fuel cell which is compact, light in weight and design variable. For example, the hydrogen injector system is suited for automobiles and jet engines. Industrial applications require larger systems. For rocket engine applications, the hydrogen gas injector system is positioned at the top of each resonant cavity arranged in a parallel cluster array. If resonant cavities are sequentially combined in a parallel/series array, the hydrogen injection assembly is positioned after the exits of said resonant cavities are combined.

From the outline of physical phenomena associated with the process described in Table 1, the theoretical basis of the invention considers the respective states of molecules, gases and ions derived from liquid water. Before voltage stimulation, water molecules are randomly dispersed throughout water within a container. When a unipolar voltage pulse train such as shown in FIG. **6**A (**53***a* xxx **53***n*) is applied, an increasing voltage potential is induced in the molecules, gases and/or ions in a linear, step-like charging effect. The electrical field of the particles within a chamber including the electrical field plates increases from a low energy state (A) to a high energy state (J) in a step manner following each pulse-train as illustrated in FIG. **6**A. The increasing voltage potential is always positive in direct relationship to negative ground potential during each pulse. The voltage polarity on the plates which create the voltage fields remains constant. Positive and negative voltage "zones" are thus formed simultaneously.

In the first stage of the process described in Table 1, because the water molecule naturally exhibits opposite electrical fields in a relatively polar configuration (the two hydrogen atoms are positively electrically charged relative to the negative electrically charged oxygen atom), the voltage pulse causes initially randomly oriented water molecules in the liquid state to spin and orient themselves with reference to positive and negative poles of the voltage fields applied. The positive electrically charged hydrogen atoms of said water molecule are attracted to a negative voltage field; while, at the same time, the negative electrically charged oxygen atoms of the same water molecule are attracted to a positive voltage field. Even a slight potential difference applied to the inert, conductive plates of a containment chamber will initiate polar atomic orientation within the water molecule based on polarity differences.

When the potential difference applied causes the orientated water molecules to align themselves between the conductive plates, pulsing causes the voltage field intensity to be increased in accordance with FIG. **6**A. As further molecular alignment occurs, molecular movement is hindered. Because the positively charged hydrogen atoms of said aligned molecules are attracted in a direction opposite to the negatively charged oxygen atoms, a polar charge alignment or distribution occurs within the molecules between said voltage zones, as shown in FIG. **6**B. And as the energy level of the atoms subjected to resonant pulsing increases, the stationary water molecules become elongated as shown in FIG. **6**C. Electrically charged nuclei and electrons are attracted toward opposite electrically charged voltages zones—disrupting the mass equilibrium of the water molecule.

In the first stage, as the water molecule is further exposed to a potential difference, the electrical force of attraction of the atoms within the molecule to the electrodes of the chamber also increases in intensity. As a result, the covalent bonding between said atoms which forms the molecule is weakened and ultimately terminated. The negatively charged electron is attracted toward the positively charged hydrogen atoms, while at the same time, the negatively charged oxygen atoms repel electrons.

Once the applied resonant energy caused by pulsation of the electrical field in the cavities reaches a threshold level, the disassociated water molecules, now in the form of liberated hydrogen, oxygen, and ambient air gases begin to ionize and lose or gain electrons during the final stage in the injector assembly. Atom destabilization occurs and the electrical and mass equilibrium of the atoms is disrupted. Again, the positive field produced within the chamber or cavity that encompasses the gas stream attracts negatively charged ions while the positively charged ions (and/or hydrogen nuclei) are attracted to the negative field. Atom stabilization does not occur because the pulsating voltage applied is repetitive without polarity change. A potential of approximately several thousand volts triggers the ionization state.

As the ionized particles accumulate within said chamber, the electrical charging effect is again an incremental stepping effect that produces an accumulative increased potential while, at the same time, resonance occurs. The components of the atom begin to "vibrate" at a resonant frequency such that an atomic instability is created. As shown in k FIG. **6D**, a high energy level is achieved, which then collapses resulting in the release of thermal explosive energy. Particle impact occurs when liberated ions in a gas are subjected to further voltage. A longitudinal cross section of a gas resonant cavity is shown in FIG. **5A**. To promote gas ionization, electromagnetic wave energy such as a laser or photon energy source of a predetermined wave length and pulse-intensity is directed to and absorbed by the ions forming said gas. In the device of FIG. **5A**, semiconductor optical lasers **20***a*-**20***p*, **20**xxx surround the gas flow path. In the device of FIG. **5B**, photo energy **20** is injected into a separate absorption chamber **21**. The incremental stimulation of nuclei to a more highly energized state by electromagnetic wave energy causes electron deflection to a higher orbital state. The pulse rate as well as intensity of the electromagnetic wage source is varied to match the absorption rate of ionized particles to produce the stepped incremental increase in energy. A single laser coupled by means of fiber optic light guides is an alternative to the plurality of lasers shown in FIG. **5B**. Continued exposure of the gas ions to different forms of wave energy during voltage stimulation maintains individual atoms in a destabilized state and prevents atomic stabilization.

The highly energized gas ions are thermally ignited when said combustible gas ions pass from injector **4** and enter into and pass through a nozzle, **10** in FIG. **5B**, or an optical thermal lens assembly such as shown in FIG. **5C**. In FIG. **5C**, the combustible gas ions are expelled through and beyond a quenching circuit, **30**, and reflected by lenses, **31** and **32**, back and forth through a thermal heat zone, **33**, prior to atomic breakdown beyond exiting through a final port, **34**. A quenching circuit is a restricted orifice through which the particle stream passes such that flashback does not occur. The deflection shield or lens, **31**, superheats beyond 3,000° F. and the combustible gas ions passing through said exiting-ports are regulated to allow a gas pressure to form inside said thermal zone. The energy yield is controlled by varying the applied voltage, or pulse-train since said thermal-lens assembly is self-adjusting to the flow-rate of said ionized and primed gases. The combustible ionic gas mixture is composed of hydrogen, oxygen, and ambient air gases. The hydrogen gas provides the thermal explosive force, the oxygen atoms aid the gas thermal ignition, and the ambient air gases retard the gas thermal ignition process to a controllable state. As the combustible gas mixture is exposed to a voltage pulse train, the stepped increasing voltage potential cases said moving gas atoms to become ionized (losing or gaining electrons) and changes the electrical and mass equilibrium of said atoms. Gases that do not undergo the gas ionization process may accept the liberated electrons (electron entrapment) when exposed to light or photon stimulation. The electron extractor grid circuit, FIGS. **8A** and **8B**, is applied to the assembly of FIG. **5A** or FIG. **5B**, and restricts electron replacement. The extractor grid, **56**, is applied adjacent to electric field producing members, **44** and **45**, within the resonant cavity. The gas ions incrementally reach a critical-state which occurs after a high energy resonant state. At this point the atoms no longer tolerate the missing electrons, the unbalanced electrical field, and the energy stored in the nucleus. Immediate collapse of the system occurs and energy is released as the atoms decay into thermal explosive energy.

The repetitive application of a voltage pulse train (A through J of FIG. **6A**) incrementally achieves the critical state of said gas ions. As the gas atoms or ions (**1***a* xxx **1***n*) shown in FIG. **6C** become elongated during electron removal, electromagnetic wave energy of a predetermined frequency and intensity is injected. The wave energy absorbed by the stimulated gas nuclei and electrons causes further destabilization of the ionic gas. The absorbed energy from all sources causes the gas nuclei to increase in energy state, and induces the ejection of electrons from the nuclei.

To further stimulate the electron entrapment process beyond the atomic level (capturing the liberated electrons during the hydrogen fracturing process) the electron extractor grid (as shown in FIG. **8A**) is placed in spaced relationship to the gas resonant cavity structure shown in FIG. **5A**. The electron extractor grid is attached to an electrical circuit (such as shown in FIG. **8B**) that allows electrons to flow to an electrical load, **55**, when a positive electrical potential is placed on the opposite side of said electrical load. The electrical load may be a typical power consuming device such as a light bulb or resistive heat producing device. As the positive electrical potential is switched on or pulse-applied, the negative charged electrons liberated in the gas resonant cavity are drawn away and enter into resistive load where they are consumed and released at heat or light energy. The consuming electrical circuit can be directly connected to the gas resonant cavity positive electrical voltage zone. The incoming positive wave form applied to resonant cavity voltage zone through a blocking diode is synchronized with the pulse train applied to the gas resonant cavity by the circuit of FIG. **7** via alternate gate circuit. As one pulse train is gated "ON", the other pulse train is switched "OFF." A blocking diode directs the electron flow to said electrical load while resistive wire prevents voltage leakage during pulse train "ON" time.

The electron extraction process is maintained during gas flow-rate change by varying the trigger pulse rate in relationship to applied voltage. The electron extraction process also prevents spark-ignition of the combustible gases traveling through the gas resonant cavity because electron build-up and potential sparking is prevented.

In an optical thermal lens assembly or thrust-nozzle, such as shown in FIG. **5C**, destabilized gas ions (electrically and mass unbalanced gas atoms having highly energized nuclei) can be pressurized during spark-ignition. During thermal interaction, the highly energized and unstable hydrogen gas nuclei collide with the highly energized and unstable oxygen gas nuclei and produce thermal explosive energy beyond the gas burning stage. Other ambient air gases and ions not otherwise consumed limit the thermal explosive process.

Variations of the process and apparatus may be evident to those skilled in the art.

What is claimed is:

1. In an apparatus for obtaining the release of a gas mixture including hydrogen and other dissolved gases entrapped in water, from water, the improvement consisting of a resonant electronic circuit in operative relationship with the water in which the dielectric property of water determines the resonance of the circuit.

2. The apparatus of claim **1** in which the resonant circuit includes a resonant charging choke.

3. The apparatus of claim **1** in which water is included as a dielectric between conductive members that form a capacitor in the resonant circuit.

4. A method of obtaining thermal energy from water, consisting of:

(A) providing a capacitor, within which water is included as a dielectric, in a resonant charging choke circuit that includes an inductance in series with the capacitor;

(B) subjecting the capacitor to a pulsating, unipolar electric field in which the polarity does not pass beyond an arbitrary ground, whereby the water molecules within the capacitor are subjected to a charge of the same polarity;

(C) further subjecting the water in said capacitor to said pulsating electric field to achieve a pulse frequency such that the pulsating electric field induces a resonance within the water molecule;

(D) continuing the application of the pulsing frequency to the capacitor after resonance occurs so that the energy level within the molecule is increased in cascading incremental steps in proportion to the number of pulses;

(E) maintaining the charge of said capacitor during the application of the pulsing field, whereby the co-valent electrical bonding of the hydrogen and oxygen atoms within said molecules is destabilized, such that the force of the electrical field applied within the molecule exceeds the bonding force of the molecule, and hydrogen and oxygen atoms are liberated from the molecule as elemental gases;

(F) collecting said hydrogen and oxygen gases, and any other gases that were formerly dissolved within the water

(G) subjecting the collected gas mixture to a pulsating, polar electric field whereby electrons of the gas atoms are distended in their orbital fields by reason of their subjection to electrical polar forces, at a frequency such that the pulsating electric field includes a resonance with respect to an electron of the gas atom;

(H) cascading said gas atoms with respect to the pulsating electric field such that the energy level of the resonant electron is increased in cascading incremental steps;

(I) ionizing said gas atoms;

(J) subjecting the ionized gas atoms to electromagnetic wave energy having a predetermined frequency to induce a further election resonance in the ion, whereby the energy level of the electron is successively increased;

(K) extracting further electrons from the resonating ions while such ions are in an increased energy state to destabilize the nuclear and electron configuration of said ions; and

(L) subjecting the destabilized ions to thermal ignition.

5. The apparatus of claim **2** in which water is included as a dielectric between conductive members that form a capacitor in the resonant circuit.

6. An apparatus in accordance with claim **1** or claim **2** or claim **3** or claim **5** in which the gas mixture released is collected and utilized as a source of thermal energy, and including further successively interconnected:

(A) means for collecting the gas mixture released in the apparatus;

(B) means for providing a pulsating, polar electric field to the gas mixture, whereby electrons of the gas atoms are distended in their orbital fields by reason of their subjection to electrical polar forces, at a frequency such that the pulsating electric field induces a resonance with respect to an electron of the gas atom; and the energy level of the resonant electron is increased in cascading, incremental steps; and

(C) means for providing a further electric field to ionize said gas atoms;

said further means being connected to an electromagnetic wave energy source for subjecting the ionized gas atoms to wave energy of a predetermined frequency to induce a further election resonance in the ion, whereby the energy level of the electron is further successively increased; and

(D) an electron sink for extracting electrons from the resonanting ions while such ions are in an increased energy state to destabilize the nuclear and electron configuration of said ions;

(E) a control means for directing particle flow in a continuous manner through the electric fields, wave energy source and electron sink to a final orifice at which the destabilized ions exit from the apparatus; and

(F) a terminal orifice at which the said ions exit from the apparatus.

7. An apparatus in accordance with claim **1** or claim **2** or claim **3** or claim **5** including means for collecting the gas mixture released from the water.

8. An apparatus in accordance with claim **7** including means for directing the gas mixture to a combustion locus.

9. The apparatus of claim **8** in which a flame is produced at the combustion locus.

10. The apparatus of claim **8** in which an explosion is produced at the combustion locus.

11. The apparatus of claim **8** in which the combustion locus is in an automobile.

12. The apparatus of claim **8** in which the combustion locus is in a jet engine.

* * * * *

[rexresearch.com](http://www.rexresearch.com)

Paul BECKER

Anti-Aging Machine

---

it's
stock
shutterstock

Ad
PayPal
starts with BUSINESS in a BOX
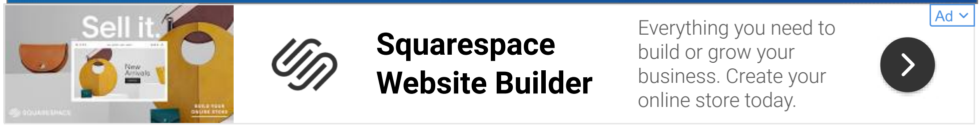

Squarespace Website Builder
Everything you need to build or grow your business. Create your online store today.

---

http://earthpulse.net/anti-aging.htm
http://www.earthpulsetechnologies.com

EarthPulse™ Anti-Aging Machine: MoreATP - the Mitochondrial Theory of Aging... in Reverse!

by

Paul F. Becker

EARTHPULSE

The Anti-Aging machine technology described herein is known as the EarthPulse™ sleep & performance enhancement system [**http://www.earthpulsetechnologies.com**]. It represents nothing less than a breakthrough in the science of anti-aging. This novel anti-aging machine provides methods that expose the entire body to night-long sessions of very-weak, pulsed DC electromagnetic fields, mimicking those fields measurable within the bottom of the Schumann scale, but particularly at 9.6 Hz that have been shown to promote deep sleep while greatly enhancing levels of naturally derived ATP.
More-ATP!

Curiously the same psychoactive frequencies that entrain humans to sleep via well established process known as brain-wave entrainment, have been shown in several separate physiological and neurological studies to stimulate cell mitochondria, subsequently resulting in enhanced cell respiration / oxydative phosphorilation including the Krebs cycle (citric acid cycle) whereby increased enzyme levels known to be associated with oxygen metabolism provide more ATPwith less oxidative waste known as reactive oxygen species (ROS). After weeks of chronic nightly stimulation, cells are so well detoxified and producing such high levels of ATP (known by the body to be associated with youthful levels), that the cells actually revert from mature to developmental DNA signature as proven by NASA (see citation link below).

This study from 2012 at 10 Hz shows that we're more than likely "right on the money"-> Mechanism of functional recovery after repetitive transcranial magnetic stimulation (rTMS at 10 Hz) in the subacute cerebral ischemic rat model: neural plasticity or anti-apoptosis? It turns out 10 Hz is anti-apoptopic.

Aerobic metabolism becomes so efficient after night long stimulation at frequencies between 2 & 10 Hz that during intense athletic performance, it prolongs aerobic metabolism and minimizes (or eradicates) time spent in anaerobic metabolism; thereby minimizing lactic acid production, lost performance & exercise induced pain.

Enhancing sleep while synergistically turbo-charging cell respiration over months and years is a concept I refer to as More-ATP. More-ATP is quite literally the Mitochondrial Theory of Aging...but in reverse. It won't completely erase cross-linking but it'll slow it considerably while it reverses many of the known aging-markers.

The Anti-Aging Machine vs. the Mitochondrial Process of Aging

The Mitochondrial Theory of Aging is Denham Harman's 1970's sequel to his popular and generally accepted Free Radical Theory of Agingwhich first dates back to the 1950's. It can be deduced from Harman's work that if one could stop free-radical production, then aging could be slowed (or possibly halted). If one could restore mitochondrial integrity simultaneously with a reduced oxidative load, then true age reversal would be possible. In 2003 NASA proved 10 Hz resulted in cellular genetic signature switching from mature to developmental at 17 to 20 days exposure. See Key 10 Hz research below.

The anti-aging product we call EarthPulse™ Sleep on Command™ sleep machine, literally 'tunes' down thought processes, decreasing sleep onset and prolonging Delta-rhythm sleep. Increases in Delta rhythm sleep should provide linear increases in energy conservation during nighttime. Simultaneously your cell mitochondria are being turbocharged so cells produce more energy with far-less oxidative waste. During deeper sleep the synergistic gain in cellular energy from both conservation and mitochondrial output is immediately sequestered for improved memory consolidation, hormone synthesis, cellular repair and immune function. Over weeks, months and years our standardized client feedback forms show synergistic changes in wellbeing and athletic performance in hundreds and hundreds of clients. More or less they match those identified personally by the author and can be described as nothing less than true age-reversal.

In citations below, 10 Hz is described to produce 4x baseline tissue regeneration & cell DNA switching from mature to developmental in experimental tissue culture at between 17 and 21 days (5 Hz and 15 Hz produced 2x baseline regeneration and no DNA changes); increase of 3x baseline oxygen metabolizing enzymes at 10 Hz, 10 Hrs per day, at 30 days; and 2x baseline increase mitochondrial density & 3x baseline ATP levels at nerve synapse within 7 minutes of 10 Hz stimulation. Knowing this, the reader shouldn't be surprised to hear that the same 10 Hz frequency produced the most efficient and repeatable healing effects as reported in 5 decades of peer-reviewed pulsed electromagnetic field (PEMF) research. Soviet research focused on 10 Hz for years leading up to the end of the Soviet Union.

The Holy Grail of ANY medical paradigm is a way to directly boost integrity of the mitochondria, thereby enhancing cell membrane potential and reducing oxidative stress. Oxidative stress damages mitochondrial DNA (mtDNA), resulting in detrimental effect on mitochondrial integrity long before detrimental effects are felt by the cell or cell DNA. The device and methods described herein as Sleep on Command™ is the aforementioned Holy Grail.

SearchPubMed.gov or Google scholar for "motochondrial dysfunction" (in quotes) AND (uppercase) plus any disease term; in 9 out of 10 (or more) disease states you will find relatively convincing evidence of coexisting mitochondrial dysfunction. Take a moment and try it with some diseases that touch your life and see that I am correct. Mitochondrial dysfunction causes these disease states in the tissue at the site of the dysfunction be it endocrine, cardio-pulmonary, neurological, physiological or psychological.

We disclose herein, a very simple technique that systemically upregulates mitochondrial integrity locally and systemically, thereby increasing excess ATP for repair and detoxification, subsequently lowering oxidative stress and restoring much, if not all of the cells' function over weeks, months and years.

We propose that once the anti-aging process is complete (this is a years long process), the natural aging process again proceeds but at greatly reduced rates. Standardized feedback forms from hundreds of clients collected over period of 6 years, and 9+ years of personal experiences (upon which the technology is continually developed) confirms; long duration, nightly exposure to a spectrum of DC pulsed electromagnetic field (PEMF) at under 10 Hz enhances cell respiration mechanisms and lower free radical waste production sufficiently for the organism to exhibit performance mimicking those of himself at 1/2 his age. According to NASA-Goodwin, only 10 Hz field provides the genetic effects described.

Anyone can measure improved or cellular respiration, we set up a relatively standardized test called the RBH or Resting-Breath-Hold Test to nearly-objectively measure the length of time the body can sustain itself on the oxygen contained in one deep breath of air.

The test is done for baseline, then once per week for the first month and once per month for remaining two months during the 90 day trial period. Test is performed total of 7 times only so little effect is due to practice. RBH times increase 20% to 100% in all clients who return the questionnaire in 90 days. Clients' resting-breath-hold statistics. Failure rate 2011 down to 3% overall.

This is a very simple principal that is not dissimilar to tuning an internal combustion engine. The engines of our cells (hundreds to thousands of trillions of mitochondria) can literally be 'tuned' like a mechanical engine via external stimulation to produce better mileage, more power, reduced toxic emissions and prolonged engine life. Think of it as an Electronic-Tune-Up for every cell in your body. Reducing the amount of free-radical oxidative stress and increasing ATP production is the preposition More-ATP, i.e. Harman's Mitochondrial Theory of Aging in REVERSE!

I offer video evidence of Anti-Aging Machine's reverse-aging effect w/ Gizmo a 16 year old Yorkie-Maltese (17 y.o. 14 Sept '11) that has been on EP for 9.5 years who barks at and plays with her food like a puppy. 9 years ago this dog had cloudy cornea that comes in old dogs. After 3 years of EP (and currently) her eye's are nearly clear-black. The video includes the author at 53 y.o. explaining some of the dramatic effects of long-term nightly use since the first prototype May 2002. We're trying to attain video evidence of an 18 year old cat on EarthPulse™ 4 years every night sleeping next to its master. The owner says he's had to take the cat toys down from the attic because the cat is clawing the furniture again. People think she's 5 years old, where at 12 years old she was starting to get a bit slow.

http://www.youtube.com/watch?v=FPY9hpIw3a4&feature=player_embedded

UPDATE on Gizmo: Gizmo nearly died March 2011 as a result of a cortisone-antibiotic shot given by the veterinarian that resulted in kidney failure w/subsequent congestive heart failure. After 3 weeks of IV therapy she was sent home with a 25% chance of survival w/kidney function that had not been restored above 35%. Gizmo is alive and apparently well now (no doubt EP helped and continues her rehabilitation) although the cortisone critically deteriorated two disks in her spine. She is on high dose GLC3000 for her disks and owner places EP magnet in dogs bedding and places directly against the spine for as much of the day as possible. She improves daily but her jump-height has been permanently affected because of the discs. Recent 17 y.o. video of her walking briskly on the beach in Florida shows a hump in her back she didn't have in last years video (where a reasonable person would swear she was between 4 and 5 years old.

Corticosteroid-Associated Congestive Heart Failure in 12 Cats Stephanie A. Smith, DVM, MS, DACVIM (Internal Medicine)a Anthony H. Tobias, BVSc, PhD, DACVIM (Cardiology) Deborah M. Fine, DVM, MS, DACVIM (Cardiology)b Kristin A. Jacob, DVM, DACVIM (Cardiology) Trasida Ployngam, DVM:

While the science of "longevity" and anti-aging fans world-wide are searching for the next miracle phytochemical, or nutricutical anti-aging miracle-molecule, or stem cell therapy, or the safest, most-affordable injectable HGH or hormone replacement therapy, you've just stumbled across a very 'simple' Fountain of Youth right here.

Anti-Aging Machine's More-ATP based upon the simple preposition that if one were able to sustain the body longer on a fixed amount of oxygen (one full inhale), then body would be extracting more usable energy out of that breath of air with less oxidative waste. It is well established that reducing oxidative stress can slow aging. Enhanced cell respiration is the ONLY plausible explanation for the amount of change we see in the RBH-test (20% to 100% increase in 90 days).

Here I provide a cohesive method to radically extend healthy human and animal lifespan & explain precisely why 5 decades of peer reviewed pulsed electromagnetic field research (utilizing frequencies between 1 Hz and 15 Hz have shown such reliable / repeatable healing and regenerative effects with no negative side- effects. By default, More-ATP provides a simple explanation for much of the documented effect of repetitive transcranial magnetic stimulation (TMS / rTMS) as it is nearly always applied under 20 Hz.

This range of frequencies provide roughly the same effect regardless of whether the stimulation is magnetic or electric in nature. However this paper will deal with pulsed electromagnetic fields as they are easier to control and since electric currents limit effectiveness by following the path of least resistance, such that homogeneous and systemic electrical stimulation of mitochondria is virtually impossible.

The More-ATP preposition remains cohesive at frequencies as low as 1-2 Hz and as high as 20 Hz, regardless of tissue type (even in plants), and particularly where tissue is damaged or debilitated and suffering underlying mitochondrial dysfunction; which i propose is the rule rather than exception in most disorders.

The FDA has refused approvals for novel uses of pulsed electromagnetic fields hiding behind heretofore vague explanations to FDA regarding how the efficacy is achieved. For instance "raising cell membrane potential". Raising cell membrane potential is the end result of raising mitochondrial performance!

In high frequency therapeutics however, the raised cell membrane potential is unexplainable but for some Alternative Cellular Energy (ACE) pathway, whereby cells 'pick up energy' by some undisclosed mechanism as the pulses pass through the cells. This paper rectifies the dilemma by making it relatively easy to explain to the FDA in the minutest detail (at the mitochondrial level) how PEMF stimulation between 1 Hz and 20 Hz upregulate cell mitochondria and raise cell membrane potential organically, from within. More-ATP relies on organically produced excess ATP!

This alone should kick-open the door to many new PEMF approvals by FDA and similar agencies worldwide should the patent holders of the several effective devices in this frequency range be wise enough to study this information and use it. Although these several patented technologies are effective, none are designed for long duration sleep-enhancement sessions. Shorter sessions during the daytime with any of these will never cumulatively have the same effect as regular, night-long sessions nor will it have a statistically significant anti-aging effect as duration of stimulation is simply not long enough. It took 8 hours stimulation per day for 30 days to peak the level of enzymes responsible for oxygen metabolism, 24 hours per day for 21 days for NASA to change DNA signature; yet only minutes for 10 Hz to triple nerve synapse energy.

I've chosen the name of this 10 Hz stimulation mechanism More-ATP because nearly all the effects of PEMF at this frequency range are due very simply to excess production of organic ATP that is synthesized naturally in the cell mitochondria via the cell respiration process albeit at amplified or accelerated levels. It must be so efficient, that NASA scientist Goodwin (see below) found over 175 Genes associated with cell maturity switched -OFF- and 150 genes associated with cell development switched -ON-.

This is not theory at all, but applied - evidence based science as data has been collected on hundreds of clients showing vast improvements in oxygen metabolism within 7 days. We test oxygen metabolism through a very low-tech, yet accurate and objective means called the resting-breath-hold (RBH) test. Free divers call it the "static" breath-hold. Very simply, it is how long you can hold your breath at rest;....providing a very reliable objective method to measure oxygen consumption rates. Draw oxygen from the blood slower and saturated blood oxygen (SAO2) goes up.

We've seen SAO2 go up 3% over 3 days in a 68 year old woman and 5% in a 55 year old distance runner living at 6000 feet altitude who had never previously achieved 99% SAO2 at altitude. Holding breath longer at rest, means your mitochondria are burning the O2 contained in that breath more efficiently than it had previously. As the breath-hold goes up, cellular energy is going up in (at least) linear fashion. One could arguably attain the same effect through years of meditation, yoga or practice; albeit with infinitely more effort. This takes none. An interesting and important point is this effect is amplified under-load. Two Olympic swimmers with 100% baseline SAO2 increased their RBH test by 19% & 20% respectively yet swam 49% and 50% further underwater on one breath. Three years ago one of these swimmer went undefeated during FINA World Cup at 3 distances (50M, 100M, 200M) in 7 cities. That is 21 Gold medals out of 21 races. He was nearing retirement age and only one of two swimmers to ever complete the World Cup undefeated. He almost retired due to lost performance prior to use of EP. I should also add here that he never competed in 200M distance since college ("because I suck at 200M") and only began training for it at my suggestion in pre-season training the year he went 21:21 in FINA World Tour Championships. Power lifters see between 10% and 20% peak strength (in range of anabolic steroids) w/no weight gain in a matter of weeks. See our page on Ergogenic Athletic Performance via Pulsed Electromagnetic Fields.

Very simply, exposure to 10 Hz (or close enough; between 5 and 15 Hz) has an immediate and direct effect on the mitochondria as demonstrated in the studies immediately following. While 5 Hz and 15 Hz have been found to produce DOUBLE the rates of tissue regeneration, 10 Hz produced tissue regeneration at FOUR times baseline according to NASA. At power frequency (50-60 Hz) and higher, you are crippling your mitochondria. At 20 Hz or more, stress hormones are produced so better to not use PEMF therapies under influence of higher than 15 Hz for prolonged periods. The Anti-Aging-Product known as Sleep on Command™ operates at 14.1 Hz maximum frequency in ALERT-MODE or during wake-up phase from 4 modality programs.

Key 10 Hz Pulsed Magnetic Anti-Aging Research:

In 1989 David Hood found chronic 10 Hz stimulation (30+ days 10 hours per day) TRIPLED two CRITICAL enzymes used during final stages of cell respiration. Citrate-synthase the regulator of the Krebs cycle and cytochrome-c-oxidase, the last enzyme in the electron transport chain. Hood concluded this was due to similarly increased mitochondrial ribonucleic acid (mtRNA). Hood's finding is consistent with our finding of improved oxygen use in every one of our satisfied clients. Testing systemic oxygen consumption via RBH test since 2005 we have found routine improvements in 94% of our clients (96% since 2008 after last technical upgrade). Since these enzymes control critical steps in oxygen metabolism it would make sense that oxygen molecules are being utilized more efficiently.

In 2003 (nearly a year AFTER EarthPulse™ began commercial sales utilizing DC pulsed electromagnetic field @ 9.6 Hz) NASA-Goodwin found a 10 Hz pulsed electromagnetic field caused neural tissue regeneration @ 4x baseline w/ improved 3-D orientation (pg 17); while causing cellular DNA to revert from mature to developmental (more than 175 maturation genes switched -OFF- and 150 developmental genes switched -ON- pgs.15-18). The implications on longevity are staggering!

NASA didn't dig deep enough to explain why cell regeneration QUADRUPLED under 10 Hz pulsed electromagnetic field (PEMF) or why the genetic effects occurred ONLY at 10 Hz pulsed electromagnetic field (PEMF). This effect can only be due to immense amounts of energy being generated by the cell. See Tong immediately below. Very plainly the cell is detoxified with all this extra energy and literally "tricked" into believing it is young again. It is not outside the realm of possibility that it will be found genetic and chromosomal repair, and lengthening of telomere is accomplished also. The longevity effect on the human/animal organism is staggering. We feel this effect since 24th May 2002.

In 2007 James Tong DOUBLED mitochondrial density at nerve synapse junctions but TRIPLED nerve synapse energy utilizing 10 Hz stimulation. These synergistic changes took only 5 minutes to manifest.It is prudent to consider Tong first stimulated at 1/2 Hz for just two minutes and both mitochondrial density and nerve junction energy parameters fell by 20%. However within 5 minutes at 10 Hz both measurements soared and remained static at increased levels even after subjected to 1/2 Hz again at end of the study.

In the chart linked above, compare mitochondrial DENSITY 'unpotentiated' mitochondria (white circles) w/ 10 Hz 'potentiated' mitochondria (black circles);

compare TOTAL ENERGY at synapse junction 'unpotentiated' (white circles) w/ 10 Hz 'potentiated' (black circles). Mitochondrial density change = 2x+ & total Synapse Energy change = 3x+ so each mitochondria increased its energy production by nearly 50% within 5 minutes. Also keep in mind that Hood's samples took 30 days, 8 hours per day 10 Hz stimulation to peak enzyme production and NASA took 17-21 days and we find in human subjects improvement in performance (ROB) takes YEARS to peak even if measured just once every year or two.

We have relied upon 9.6 Hz in our Sleep on Command™ pulsed-magnetic sleep-machine since 2001 because it was reported (in an obscure journal circa 1972) by Bob Beck to emanate from the molten core. It rises perpendicular to the horizon (to be distinguished from Schumann resonance which flows parallel to horizon). Beck discovered this 9.6 Hz resonation in the early to mid 1970's by driving metal stakes 20 feet into the earth, attaching an amplifier and measuring the frequency of the signal. Although Beck continually worked on the BT (Brain Tuner) series for 20 years until his death, he neglected to mention that experiment and tuned his BT to Schumann 7.83 Hz. I've always found this very curious.

Historically speaking, if we were bedded-down on the ground every night w/ no radiofrequency microwave telecommunication we would be entrained predominantly by the naturally occurring 9.6 Hz and what probably was a well designed biological law on the order of biblical creator or evolutionary design depending on your viewpoint; explaining why 10 Hz stimulation works as well as it does. Probably why it is so widely known that people sleep well on camping trips too. It is NOT the fresh air and exercise.

While our neurological system is known to be "paced" by the Schumann waves it is clear that the mitochondria are tuned by the Earth's core resonation at or around 9.6 Hz; particularly while lying horizontal and preferably when sleeping in contact with the ground. This is the primary reason camping is known to produce such remarkable sleep and why earthing technologies are gaining popularity. EarthPulse™ is earthing on steroids!

In Gizmo's Summer 2010 video clip (where she's 80+ human years old) she moves as fast and quick as a 5 year old dog. Her mind state is definitely puppy-like. She's currently on the device for more than 1/2 her lifetime. During the day she spends hours in her favorite perch (on the top of the couch arm-rest) where she sleeps for hours (1/2 inch above the 200 G inductor at 9.6 Hz) plus nighttime exposure at under 1 Gauss with the inductor under her master's mattress. She still stalks her food like a puppy; and if you turn up the volume in the 8 second video #2, you'll hear her digging her nails DEEP into the carpet. No arthritis or joint pain in her whatsoever. Please take the time to check out some of the other life changing videos while at our YouTube channel.

In 2001 at 43 years old I built my first device for a fellow with Parkinson's disease. I discovered the sleep effect the very first afternoon it was completed. I quickly found that regular nighttime use of the EarthPulse™ DC pulsed electromagnetic fields at 9.6 Hz quickly and reliably enhanced sleep and daytime mental & physical performance. Recovery even from old sports injuries was phenomenal. 9 years later my destroyed left rotator cuff is still 99% perfect...

It took 7 full years for my RBH to double with no effort, no practice or aerobic physical conditioning but for sporadic surfing and rounds of golf. It is my propositon that 9.6 Hz every night could have resulted in taking 7 full years for a complete turnover of cells in the body, that older cells had revived and lingered for years. Based upon that fact I summize a doubling of lifespan with potential to reach 200 years or more if the modality is started in early 20's. The longevity possibilites are staggering; especially if user combines 9.6 Hz while sleeping with other known anti-aging therapies and supplement programs.

These pulsed magnetic field therapy bibliographies contain several hundred peer reviewed pulsed electromagnetic therapy study abstracts spanning nearly 6 decades, organized by disorder, most linked directly to the U.S. National Institutes of Health & National Libraries of Medicine. You will find many, if not most of the successful pulsed electro-magnetic therapy research used frequencies under 20 Hz.

Robert O. Becker MD and the Electromedical Revolution:

Robert O. Becker MD proved optimum cellular repair at stimulation periods 8 hours - ON - / 8 hours - OFF -; using minute, microampere DC electric current. His mouth to my ears... "The cell can only absorb so-much energy at a time and create only-so-much change over 24 hours. Your 8 hours during the nighttime is the next best thing." This being the classic example of ACE pathway (alternative cellular energy), and the one explanation that the FDA finds inadequate enough to HIDE behind, even in the case of fresh fracture healing at 200% - 300% faster rate, spinal cord repair or stroke rehabilitation. Click here to see a synopsis of Robert O. Becker's work.

At the time I had been unable to discuss mitochondrial effects with Bob as I hadn't been aware of the NASA research until late 2005 and the Tong and Hood studies until late 2008, and although feeling better "winded" I had not noticed my RBH increase as I had not checked it for some years. My several conversations with Bob prior to his death were relegated to ACE pathways discussion as I was yet unable to discuss mitochondrial stimulation. Other topics of our converstions were DC "microcurrents" with or without conductive silver nylon fabrics; the electrochemical treatment of cancer tumors using silver electrodes (rather than tungsten or platinum); and of course EarthPulse™ PEMF entrainment for sleep. We did discuss the incrontravertable 10 Hz pattern in the peer reviewed research, the fact his long-time friend (and one-time student) Andrew Marino, PhD had done research with 10 Hz and the discovery of the 9.6 Hz "resonation" discovered by Bob C. Beck. However without knowledge of the NASA, Tong or Hood studies, I was unable to discuss these effects with Becker.

This link to Becker & Marino's Electromagnetism and Life is probably the most well documented essay in existence on the detrimental effects of power and microwave telecommunications frequencies as well as some of the earlier works using electric and magnetic stimulation techniques. Far, far ahead of their time, but very relevant today.

The end effect of nightlong sessions with enhanced ATP production and reduced oxidative stress allow cells to detoxify and repair over weeks and months to become healthy normal functioning cells again. Old injuries heal remarkably well during this several months process. Cells that are healthy now have extended life-time. Over time, the technique literally stops aging, reverses it for perhaps as much as a decade; once repair process is complete and 'normal' aging process resumes, it does so at much slower cell turnover rate.

Not just age reversal on one current cell generation but over weeks and months you realize those new cells then went on to produce new cells next to it and so on, in attempt to reverse all age related degeneration. Over time you realize old nagging problems like knees or hips or shoulders haven't been bothering nearly as bad as before, knees, hips, lower back, shoulders and neck feel looser, digestion is better, energy levels much higher and more stable, and you begin to feel in most cases healthier than ever in lifetime. "Wind" under-load changes noticably going-up steps or in athletic endevours within the first week or two. Age related dis-repair takes several months or more to improve as much as possible while oxygen metabolism can improve apparently over years. Once that process is complete, it then freezes user in that condition for proportionately as long as cell turnover rate is slowed (barring any health catastrophies)...

Reactive Oxygen Species and Anti-Aging:

Check this paper particularly the comment by Aubrie de Grey about reduced ROS production and its affect on cell longevity. See this study by YH Wei et al. and then go back to the 1998 David Hood study where enzyme activity responsible for oxygen molecule metabolism increased 3 fold (by 10 hrs per day for 35 days stimulation). It would figure there would be less electron (oxygen) 'leakage' when there are 3x's more enzyme responsible for oxygen molecule metabolism.

In summary: more ATP induces anti-aging effects through more organically produced ATP; resulting in upregulation of cell membrane potential; increased nerve synapse potential; improved ion transport; reduced ROS production; improved ROS neutralization; improved cell detoxification; reduced cell inflammation; thereby synergistically enhancing cell survival and longevity at the mitochondrial (potentially mtDNA) level.

Get our Sleep on Command Anti-Aging-Machine at EarthPulseTechnologies.com

---

http://www.rebprotocol.net/November2007/Robert%20O.%20Becker%20and%20Andrew%20A.%20Marino%201982%20Electromagnetism%20and%20
http://www.ortho.lsuhsc.edu/Faculty/Marino/EL/ELTOC.html

Electromagnetism & Life: Drs Robert O. Becker and Andrew Marino
(full book pdf)

http://www.microondes.files.wordpress.com/2011/01/becker_the_body_electric-full.pdf

The Body Electric: the classic by Robert O. Becker, MD
(full book pdf)

http://www.sld.cu/galerias/pdf/sitios/rehabilitacion-fis/biomagnetic_healing..pdf

Garry Null - BioMagnetic Healing Book

METHOD AND APPARATUS FOR THE TREATMENT OF PHYSICAL AND MENTAL DISORDERS WITH LOW FREQUENCY, LOW FLUX DENSITY MAGNETIC FIELDS

US7988613

A method and apparatus for generating electromagnetic fields for healing. A device preferably includes a microcontroller and associated memory, a wire coil in electrical communication with a driving circuit that is controlled by the microcontroller in accordance with a program stored in the associated memory, wherein the driving circuit is effective to produce a pulsed DC output having a frequency in the range of about 0-45 Hz, more preferably in the range of 0.5-14.1 Hz and most preferably around 9.6 Hz. A user interface is provided for selecting one of a plurality of modes of operation and a port (e.g., a USB port) is provided to allow the program stored in the associated memory to be modified by way of a computer, memory card or the Internet.; In another embodiment, the apparatus takes the form of a medallion that can be worn around a user's neck or strategically placed on a user's body or embedded in other user hardware such as a combat or racing helmet.

This application is a divisional application of U.S. patent application Ser. No. 11/095,612, filed Apr. 1, 2005 now U.S. Pat. No. 7,819,794, which is a continuation-in-part of U.S. patent application Ser. No. 10/927,840, filed Aug. 27, 2004 now U.S. Pat. No. 7,276,020, which is a continuation-in-part of U.S. patent application Ser. No. 10/278,109, filed Oct. 21, 2002 now U.S. Pat. No. 6,899,667, the contents of each of which are hereby incorporated by reference herein in their entireties.

BACKGROUND

1. Field of the Invention

This invention relates generally to a method and apparatus for the treatment of physical and mental disorders, and, more particularly, to a portable device capable of being operated safely and effectively by patients which produces a time varying, magnetic field having a low frequency and low flux density effective in the treatment of a wide variety of physical and mental disorders.

2. Background of the Invention

Magnetic fields have long been used for the treatment of physical injuries and chronic pain. Early magnetic therapy involved the use of static magnetic fields produced by permanent magnets incorporated into items such as bracelets, belts, back pads, mattress pads and mattresses. It is believed that static magnetic fields have some efficacy in the treatment of broken bones and soft tissue injuries, and tend to promote the circulation of blood as well as relieve stiffness in muscles. The effectiveness of such treatments in human and veterinary applications has been the subject of debate.

More recent attempts to employ the therapeutic effects of magnetic fields have focused on devices which generate an electromagnetic field, and the methods of treatment employing such devices. Although a variety of designs have been proposed in the prior art, electromagnetic devices generally comprise a power supply coupled to a circuit capable of producing an AC or DC output which is transmitted to an inductor coil. One form of inductor coil consists of a number of wire windings wrapped about a coil body with an open or air center, or, alternatively, a ferrous core wrapped with wire windings. In response to the output from the circuit, an electromagnetic field is generated by the inductor coil which is then directed toward the area(s) of the body of a patient to be treated.

In many instances, the circuit of electromagnetic devices produces a pulsed or time-varying output in the shape of a square wave, sine wave, triangular wave or the like. Such output can be at essentially any selected frequency and voltage. A pulsed output from the circuit results in the production of a time-varying or pulsed magnetic field by the inductor coil. If the circuit emits an AC signal, the position of the north and south poles of the resulting magnetic field from the inductor coil changes with each cycle, whereas a DC output produces an electromagnetic field in which the position of the magnetic poles remains constant.

The application of the general concepts of the formation of electromagnetic fields noted above to the treatment of physical and mental disorders has resulted in a widely varying array of devices and treatment methods. Prior art devices operate at completely different ends of the spectrum in terms of field strength and frequency. The predominant approach appears to follow the adage that "more is better." U.S. Pat. Nos. 6,425,852; 6,132,361; 5,813,970 and 5,769,778, for example, teach electromagnetic devices which produce a magnetic field having a flux density in range of up to 10,000 to 20,000 gauss. Devices of this type are used for therapies such as transcranial magnetic brain stimulation for the treatment of neurological and mental disorders. On the other end of the spectrum, devices have been developed for the treatment of various conditions using a magnetic field having a flux density in the range of 10 nanogauss to 10 milligauss, applied at frequencies in the range of 0 to 1000 Hz. See, for example, U.S. Pat. Nos. 6,099,459 and 5,496,258.

There appears to be no consensus whatsoever as to what flux density levels or frequencies should be employed in electromagnetic therapy. Although proposed as a non-invasive alternative to pharmacological and nutritional solutions, it is believed that electromagnetic therapy conducted at the high flux density and/or high frequency levels noted above may, in fact, be harmful whereas treatment at the lower end of the spectrum as suggested in U.S. Pat. No. 6,099,459 will have little, if any, therapeutic effect without extensive technical expertise. None of these treatment methods are reflective of the magnetic field density levels and frequencies which occur naturally within a patient, or are produced naturally within the ionosphere and by the earth.

Other significant limitations of many prior art therapeutic electromagnetic devices is their lack of portability, their complexity and the need for relatively skilled medical personnel to operate them effectively. For example, U.S. Pat. Nos. 6,280,376; 6,099,459; 6,210,317 and application US 2002/0103411 disclose devices which are not portable and require a skilled technician or physician to operate. In order to receive treatment, patients must undertake the time and expense of traveling to the office where the machine is located during normal business hours. Other devices, while they may be more portable, permit a relatively wide range of adjustment of field strength and/or frequency. Allowing patients and practitioners to control these parameters, even with prior instruction, can lead to ineffective or potentially harmful treatment.

BRIEF SUMMARY OF THE INVENTION

It is therefore among the objectives of this invention to provide a method and apparatus for the treatment of physical and mental disorders with electromagnetic therapy which does not require skilled personnel to administer, which is portable, which operates at naturally occurring magnetic field strengths and frequencies, which limits the extent of operating adjustments permitted on the part of a patient or practitioner and which is capable of treating a wide variety of physical and mental disorders in human or animal subjects.

These objectives are accomplished in the instant method and apparatus comprising a circuit adapted to be coupled to a power supply which produces a pulsed DC output, and a magnetic field generating coil coupled to the output of the circuit which is effective to produce a time varying magnetic field having a flux density in the range of about 0.0001 to 90 gauss, and greater, at only specific frequencies which occur naturally within the patient or are naturally occurring terrestrially in the range of about 0 to 45 Hz, and most preferably to no more than about 20 Hz. The coil is positioned at or near the site on the body of the patient to be treated, or, alternatively, beneath the patient's sleeping surface, for a period ranging from about one-half hour to several hours depending upon the condition or method used for treatment.

An important aspect of this invention is predicated upon the concept that naturally occurring magnetic fields, both in terms of flux density and frequency, have beneficial therapeutic effects on a wide variety of physical and mental conditions. It is known that the geomagnetic field strength of the earth is about 0.3 to 0.6 gauss, and that such magnetic field exists terrestrially in the atmosphere and geomagnetically from the earth itself. What has not been recognized prior to the present invention is the beneficial effects of electromagnetic therapy applied at flux density and frequency levels which closely approximate those occurring naturally; further, that these field densities and frequencies when applied nocturnally while sleeping have a cumulative and synergistic effect on many health aspects of the body. It has been found that psychiatric and neurological disorders, central nervous system disorders, tissue damage, orthopedic conditions, wounds, muscle stiffness and a variety of conditions which cause pain and chronic pain all can be safely and effectively treated with a DC, time varying electromagnetic field whose flux density and frequency approximate that which are naturally occurring.

In one embodiment, the device of the instant invention is portable and can be readily and safely operated by patients at home or work without the presence of medical personnel. Such a portable embodiment comprises a housing which contains a circuit operable to produce a pulsed, DC output having a wave form such as square, triangular, sine or the like. In one implementation, the housing is provided with an on/off switch and the frequency of the output is fixed at one level by the circuitry at for instance 9.6 Hz, as this frequency is particularly effective overall. Alternatively, the housing includes an adjustment knob or the like connected to a potentiometer in the circuit for adjustment of the frequency of the output from the circuit. In still another embodiment, a microprocessor is incorporated in the circuitry which is programmable to sequentially vary the output frequency of the circuit to selected frequencies in the range of 0.5 to 45 Hz, and most preferably up to not more than 20.1 Hz, the second most predominant Schumann wave, as the use of higher frequencies may, in some instances, cause stress on the neurological system.

In all embodiments, an inductor coil, e.g., a magnetic field generating coil, is coupled to the output of the circuit. The coil can be in the form of a wire winding about a coil body having an open or air core, or a hard ferrous core, about which the wire is wound. As described below, the two coils are used for different types of treatments in accordance with the method of this invention.

In a preferred implementation, the device is digitally controlled and implemented as a substantially integral "system on a chip" (SOC). Such a design allows for, for example, connectivity to a computer or network via, e.g., a USB or mini USB connector. This connectivity permits a user to download the latest control software for generating the magnetic fields described herein. The SOC design of the present invention also, preferably, allows for more user-defined control through a user interface comprising an LCD display and a plurality of user controls on the device. The user interface may control, for instance, the mode of operation of the device.

One of the operation modes, for instance, is a program designed specifically for deeper more restful sleep. This mode comprises sweeping predetermined frequencies to promote deeper and more restful sleep.

In one implementation, the device incorporates memory for storing music files, which can be played back through a speaker or headphones, in coordination with the type of electromagnetic energy being generated, or independently thereof.

In still another implementation, multiple electromagnets can be connected to a single device such that the magnetic fields generated by the electromagnets are synchronized. Such multiple electromagnets can be use in lieu of two separate devices that might not necessarily generate magnetic fields in synchronicity with each other.

In a slightly different embodiment, the device of the present invention can be configured as a "medallion" to be worn, for example, around one's neck. A device configured in this way preferably includes a rechargeable battery as a power source. A medallion device like this may also be incorporated into clothing, belts and helmets to promote healing, increased metabolism, bone stimulation, brain stimulation, and overall well-being, when positioned in appropriate locations on one's body.

The device of the present invention can also be configured to plug directly into a conventional AC electrical wall socket, or to be mounted, among other places, in a home or vehicle to fill a predetermined space with the therapeutic magnetic fields described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

The structure, operation and advantages of the presently preferred embodiment of this invention will become further apparent upon consideration of the following description, taken in conjunction with the accompanying drawings, wherein:

FIG. 1 is schematic illustration of the electromagnetic device of this invention;

FIG. 2 is a diagram of the circuit of the device in FIG. 1;

FIG. 3 is a schematic view of an open center coil employed in the subject device; and

FIG. 4 is a view similar to FIG. 3 of an alternative coil with a hard ferrous center.

FIG. 5 is a cross sectional view, taken along line 5-5 of FIG. 4;

FIG. 6 is a schematic view of the magnetic lines of flux produced by the magnetic flux generating coil of FIG. 4;

FIG. 7 is a view of a digitally-controlled with user interface device in accordance with the present invention;

FIG. 8 is a schematic representation of a device in accordance with the present invention; and

FIG. 9 is a schematic representation of the user interface of the device in accordance with the present invention.

FIGS. 10A-C show the several components of an exemplary modular electromagnet in accordance with the present invention.

FIGS. 11A-D show the several components of an exemplary medallion in accordance with the present invention.

FIGS. 12A-D show different views of a pouch configured to hold the medallion and be easily attached to the body via a belt, in accordance with the present invention.

FIG. 13 shows the pouch attached to a portion of a conventional back-brace, in accordance with the present invention.

FIGS. 14A-B show how the medallion can be placed into the interior padding of a helmet or onto the interior webbing of a helmet, in accordance with the present invention.

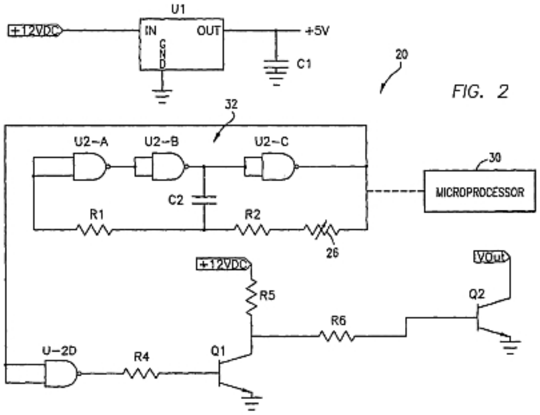
U1
+12VDC
IN
OUT
GND
+5V
C1
20
32
U2-A
U2-B
U2-C
C2
R1
R2
26
30
MICROPROCESSOR
+12VDC
R5
VOut
Q2
R6
U-2D
R4
Q1

FIG. 2

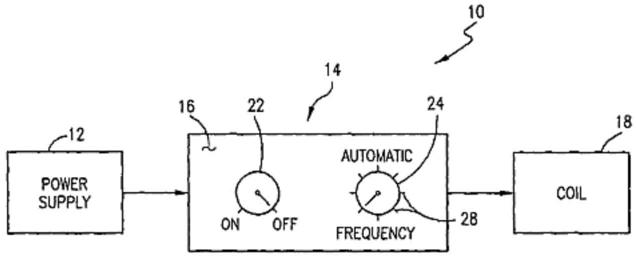
10
14
16
22
24
12
18
POWER SUPPLY
AUTOMATIC
ON
OFF
FREQUENCY
28
COIL

FIG. 1

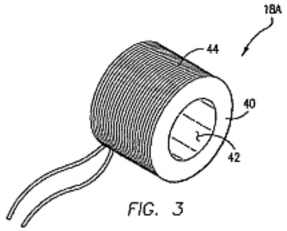

FIG. 3

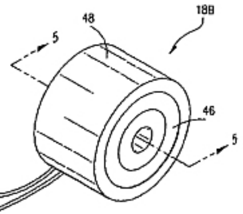

FIG. 4

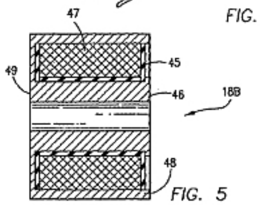
47
49
45
46
18B
48

FIG. 5

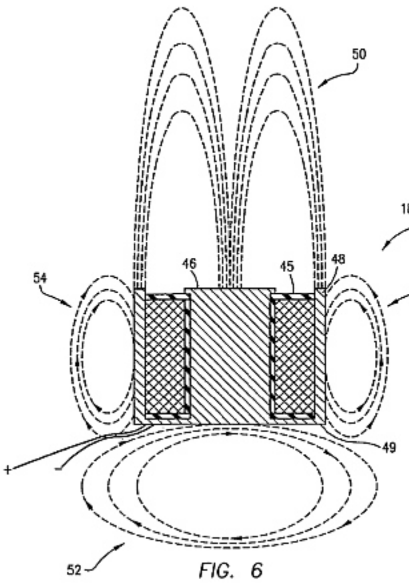
50
18B
54
46
45
48
54
+
49
52

FIG. 6

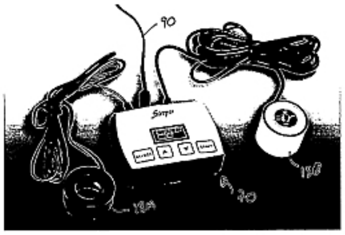
90
18A
70
18B

FIG. 7

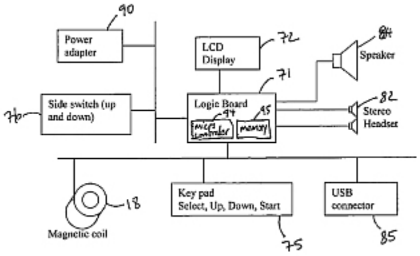
90
Power adapter
LCD Display
72
84
Speaker
71
Side switch (up and down)
70
Logic Board
94
95
micro controller
memory
82
Stereo Headset
Magnetic coil
18
Key pad Select, Up, Down, Start
75
USB connector
85

FIG. 8

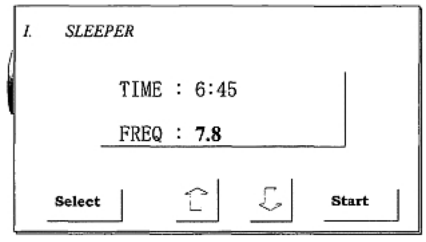
I. SLEEPER
TIME : 6:45
FREQ : 7.8
Select
Start

FIG. 9

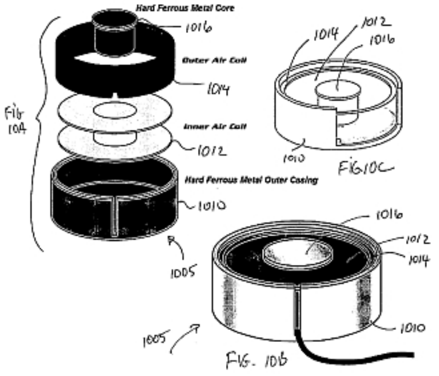
Hard Ferrous Metal Core
1016
Outer Air Coil
1014
Inner Air Coil
1012
Hard Ferrous Metal Outer Casing
1010
Fig 10A
1005
1014
1012
1016
1010
FIG 10C
1016
1012
1014
1010
1005
FIG. 10B

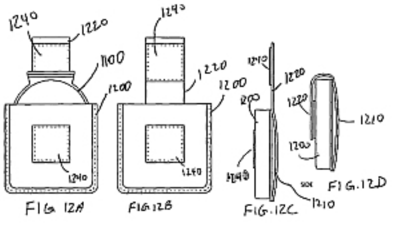
1240
1220
1100
1200
1240
FIG 12A
1240
1220
1200
1240
FIG 12B
1242
1200
1220
1248
FIG. 12C
1220
1210
1200
FIG. 12D
1210

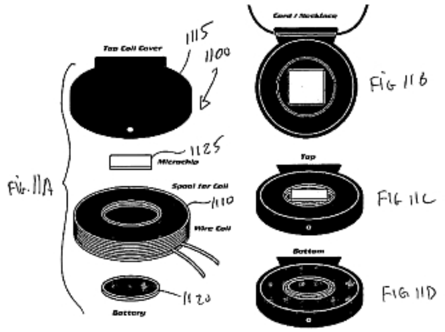
1115
1100
Top Coil Cover
1125
Microchip
Spool for Coil
1110
Wire Coil
Battery
1120
Fig. 11A
Cord / Necklace
FIG 11B
Top
FIG 11C
Bottom
FIG 11D

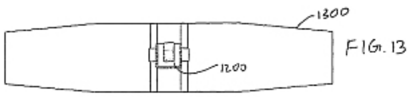
1300
FIG. 13
1200

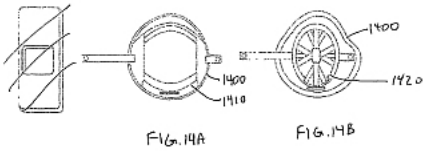
1400
1410
FIG. 14A
1400
1420
FIG. 14B

DETAILED DESCRIPTION OF THE INVENTION

Referring now to the drawings, FIG. 1 is a diagrammatic depiction of an electromagnetic device 10 of this invention which comprises a power supply 12, and housing 14 having a control panel 16 and a magnetic field generating or inductor coil 18. The power supply 12 is shown as being coupled to the housing 14, but it should be understood that the power supply could be incorporated within the housing 14 in the form of a battery pack using preferably a rechargeable battery (not shown). Alternatively, the power supply 12 can take the form of a transformer which would plug into a wall socket and step down 120/240 volt supply to voltage for use in the circuit 20 described below in connection with a discussion of FIG. 2.

The housing 14 is shown with a control panel 16 for purposes of illustration. In one embodiment of this invention, the control panel 16 includes only an on/off switch 22 which would turn on the circuit 20 allowing the coil 18 to produce a magnetic field at a predetermined, fixed flux density and frequency. Alternatively, the control panel 16 is provided with a control knob 24 coupled to a potentiometer 26 included within the circuit 20, as described below in connection with a discussion of FIG. 2, to permit variation of the output frequency of the circuit 20, and, hence, the frequency of the magnetic field produced by the coil 18. The control knob 24 may be adjusted manually to selected frequency settings, represented by the radial lines 28 on the control panel 16, or, alternatively, to an "automatic" setting in which a programmable microprocessor 30 within the circuit 20 is activated to sequentially vary the output frequency of the circuit 20, as described below.

Referring now to FIG. 2, the circuit 20 contained within the housing 14 is shown coupled to the power supply 12 which includes an IC voltage regulator U1 and a filter capacitor C1. The power supply 12 provides a voltage output of 5 volts to an astable multivibrator 32 consisting of NAND gates U2-A, U2-B, U2-C, resistors R1 and R2, capacitor C2 and the potentiometer 26. The operating frequency of the astable multivibrator 32 is determined by the values of R1, R2, potentiometer 26 and capacitor C2, which can be varied over a range of 0.5 Hz to 45 Hz (preferably not beyond 20.1 Hz) by operation of the potentiometer 26.

As schematically depicted with a phantom line in FIG. 2, the circuit 20 may optionally include a microprocessor. As noted above, the microprocessor 20 is operative to sequentially vary the frequency output of the astable multivibrator 32. The selected frequencies over which the output is varied are discussed below in connection with the description of a particular treatment method in accordance with this invention.

The signal from the astable multivibrator 32 is input to the NAND gate U2-D which is configured as an inverter. U2-D is connected through resistor R4 to and NPN bipolar junction transistor Q1 configured as an emitter follower to serve as a level shifter, e.g. to convert the signal from 5 volts to the output voltage used in one embodiment or another. Q2 is an NPN bipolar junction transistor which is coupled to Q1 through resistors R5 and R6. It functions to invert the signal from Q1, thus producing a pulsed or time varying DC output signal in the range of 0.5 to 45 Hz, preferably to no more than 20.1 Hz. When the output signal is coupled to the coil 18, a pulsed magnetic field is produced having a flux density in the range of 0.0001 to 90 gauss depending upon the embodiment of the device 10 and the size of the coil 18 which is coupled to the device at a frequency of 0.5 to 45 Hz, and preferably to no more than 20.1 Hz. Since a DC output signal is provided to the coil 18 by the circuit 20, the north and south poles of the resulting magnetic field do not vary in position relative to the coil 18.

With reference to FIGS. 3-6, two basic types of a magnetic flux generating coil 18 are shown which are employed in the device 10 of this invention. The magnetic flux generating coil 18A of FIG. 3 includes a coil body 40 with an open center or core 42 around which is wound a wire winding 44. The coil body 40 is schematically depicted in FIG. 3 with a cylindrical shape, but it may take the form of a flat disc in the general shape of a donut. The diameter of the coil body 40 and its thickness can vary, and is chosen to accommodate a particular treatment therapy, as described below.

The magnetic flux generating coil 18B depicted in FIGS. 4-6 is an electromagnet commercially available from A.P.W. Co. Inc. of Rockaway, N.J. It comprises a core 46 which is made of a hard ferrous material surrounded by a plastic sleeve 45 about which is wrapped a wire winding 47. The wire winding 47, in turn, is received within a metal jacket 48 having an end wall 49. "Hard" ferrous metals are known to possess magnetic memory, that, once energized, are capable of producing a static magnetic field for moments at a time absent any live current through the coil.

The magnetic flux generating coil 18B, in response to the pulsed DC output signal from circuit 20, produces both a time varying magnetic field and a lower amplitude static magnetic field. Because of the magnetic memory possessed by the "hard" ferrous metal, the static magnetic field remains in between the "live" pulsed DC output signal. In response to pulsed DC output signal from circuit 20, coil 18B produces a substantially time varying magnetic field of "north" polarity and simultaneously generates a static magnetic field of "north" polarity in the narrow core which remains after the pulsed DC output signal switches to off; and generates a substantially static "south" polarity magnetic field with a pulsed component out of the opposite pole mounted to the "hard" ferrous jacket. As schematically depicted in FIG. 6, the magnetic lines of flux 50 from the time varying magnetic field with the north polarity project from one end of the hard ferrous core 46. Two magnetic lines of flux 52 and 54 are produced from the substantially static magnetic field with the south polarity. The magnetic lines of flux 52 project from the opposite end of the hard ferrous core 46, and the south magnetic field also travels along the metal jacket 48 forming magnetic lines of flux 54 which project from the same end of the "hard" ferrous core 46 as the time varying, north magnetic field with static component, but in a pattern generally concentric thereto.

The magnetic field with the south polarity is characterized herein as a "substantially static" magnetic field with a time varying "south" polarity component, because the magnetic memory of the comparatively large metal jacket 48, and the frequency of the DC pulses from circuit 20, dissipate into the "hard" ferrous jacket which produce a magnetic field of south polarity having an essentially constant field strength but containing "some" time varying magnetic field of similar "south" polarity. The "time varying" magnetic field projecting from the hard ferrous core 46, on the other hand, has a field strength which increases and decreases with the DC pulse supplied from the circuit 20. Due to the magnetic memory of the hard ferrous core, the time varying magnetic field with a north polarity does not become "zero" in between pulses, but exhibits a static characteristic while the coil 18B is operating, albeit at a lesser field strength than the pulsed field. Consequently, both a pulsed north magnetic field and a lower field strength, static north magnetic field, whose lines of flux are not shown, project from the hard ferrous core 46 of the coil 18B.

The combination of a time varying north magnetic field, a substantially static north magnetic field and a substantially static south magnetic field, all acting in the same direction but with the south magnetic field projecting in concentric relation to the north magnetic fields, provides a beneficial, therapeutic effect when used in accordance with the methods of treatment of this invention. The device 10 of this invention has been successfully employed in treating a number of different conditions. The configuration of the device 10, method of treatment and related case studies for different types of disorders or conditions are discussed in separate sections below.

Psychiatric, Neurological and Central Nervous System Disorders

In one embodiment of the method of this invention, the device 10 is used to apply repetitive transcranial magnetic stimulation ("rTMS") to the back, frontal portion or other areas of the head. It is believed that the application of a pulsed, DC electromagnetic field, at the flux density and frequency ranges noted above and discussed in more detail below, is a novel approach to brain wave entrainment which has the attendant benefits of reducing errant brain wave activity and stimulating certain neurological processes. When applied to the frontal brain regions, rTMS therapy performed with the device 10 herein has an upmodulatory effect on both the mesolimbic and mesostriatal dopaminergic systems. An increase in dopaminergic neurotransmission may contribute to the beneficial effects of rTMS in the treatment of psychiatric and neurological disorders. Further, rTMS performed according to this invention synchronizes the circadian rhythm and neuroendocrine functions of the pineal, hypothalamus and pituitary glands.

The pineal gland, which has been found to be particularly sensitive to magnetic fields, regulates the function of the pituitary, thyroid and adrenal glands through the production of several neurochemical agents. It also affects the central nervous system and immune system via the production of melatonin. Melatonin has been found to be a potent antioxidant and free radical scavenger with anti-aging, antimutagenic and oncostatic properties. In pathologies wherein high production of free radicals is a primary cause of disease, melatonin is protective of mitochondrial damage due to oxidative stress, thereby protecting against impaired mitochondrial production of adenosine tri-phosphate ("ATP"), the fuel that fires all cellular processes. Melatonin is also useful in combating oxidative neurotoxicity which is associated with several acute and chronic neurodegenerative diseases. It has been shown to possess anti-inflammatory effects and reduces tissue destruction during inflammatory reaction. Melatonin attenuates transendothelial cell migration and edema which contribute to tissue damage by reducing the regulation of a variety of proinflammatory cytokines.

In order to perform therapeutic treatment on a selected area of the head, the device 10 is employed with an air core coil 18A positioned proximate the head, e.g. about 5 to 10 centimeters away from the head, beneath a pillow or other padding for comfort. The patient can be placed in a prone or reclined seated position, or lying flat on a surface. In most instances, it is preferable to place the coil 18A at the back of the head, although it may be located proximate any specific area exhibiting symptoms to be treated. The pulsed, DC output from circuit 20 results in a time varying magnetic field from the coil 18A whose north and south poles remain in the same position relative to the coil 18A. With a flux density level at the coil 18A of in the range of about 30 gauss, and the coil 18A positioned at the back of the head beneath a pillow, the flux density at the back of the patient's head is typically on the order of about 1-2 gauss. The flux density at the subthalmic region of the brain is significantly less.

It has also been discovered that certain symptoms respond more favorably to utilization of the solid core "hard" ferrous electromagnet for applying rTMS. Specifically, a majority of Parkinsonians appear to respond more favorably to this "hard" ferrous electromagnet as do those who suffer from acute migraine headache.

In the method of treatment employing rTMS according to this invention, the device 10 is operated using the control knob 24 to vary frequency settings to approximate those of the delta, theta, alpha and beta frequencies of the brain. In the manual mode of operation, the control knob 24 is turned to an initial setting, represented by radial lines 28 on the control panel 16 of housing 14, and then sequentially moved to different settings every 1 to 2 minutes or so. If only one or two of the brain wave segments is to be treated, the control knob 24 is operated accordingly. Alternatively, a microprocessor may be set to the automatic mode of operation in which case the microprocessor 30 functions to sequentially shift between the delta, theta, alpha and beta frequency segments at timed intervals.

It is generally accepted that the frequency segments of the brain range in frequency from about 0.5 to 20 Hz, unless in a stressed or anxious condition, with the delta segment being in the range of about 0 to 3 Hz, the theta segment about 3 to 8 Hz, the alpha segment about 8 to 12 Hz and the beta segment about 12 to 20 Hz and higher. Respective settings or radial lines 28 on the control panel of housing 14 correspond to specific frequencies within each of these frequency ranges so that each of the delta, theta, alpha and beta segments of the brain can be treated by sequential movement of the control knob 24 to different settings or by operation of the circuit 20 in the microprocessor "automatic" mode. In the presently preferred embodiment, the setting corresponding to the alpha segment frequency range of about 8 to 12 Hz is specifically set at 9.6 Hz.

Solar particles trapped in the ionosphere of the earth are found to resonate at frequencies of about 7.83, 14.1, 20.3, 26.4, 31.32, 39 and 45 Hz, known as the Schumann wave frequencies. A principle objective of this invention is to provide a therapeutic magnetic field at a flux density and frequencies substantially the same as those occurring in nature. Accordingly, the microprocessor 30 may also be programmed to cause the circuit 20 to produce an output signal which varies in frequency corresponding to the Schumann wave frequencies. The magnetic field produced by coils 18A and 18B at such frequencies is believed to be beneficial for the treatment of pain and other conditions while the lower three frequencies double as rTMS brain wave entrainment frequencies.

Research has shown that rTMS performed in accordance with the method of this invention is effective at reducing symptoms of central nervous system disorders, including, without limitation, Alzheimer's disease, epilepsy, seizure disorders in general, multiple sclerosis, depression, Parkinson's disease, schizophrenia and dementias of various etiologies, migraine headache, cluster headache, recurrent headache syndromes in general, severe pre-menstrual and pre-menopausal syndromes, attention deficit disorders, age-related cognitive and motor deficits, optic nerve atrophy and degenerative diseases of the retina. Improvements will also be realized in cognitive and motor functions including spatial orientations, sense of balance, improved mobility, memory, alertness, organizational skills and problem solving abilities; improved mood or decreased depression and anxiety; increased endurance, strength and stamina; and improved hand-eye coordination.

A number of subjects have been treated with the device 10 employing the method described herein, as follows.

Case 1: A 74 year old fully medicated male was diagnosed 15 years ago with Parkinson's disease, and experienced mild consistent tremor in the right hand, erratic sleep with many sleepless nights, weakness in the arms and legs and moderate to pronounced balance deficit. One 45 minute rTMS session with the inductor coil 18A set to deliver magnetic flux density of about 0.6 gauss at 9.6 Hz applied to back of head resulted in dramatic improvement in balance, speaking volume, strength, motor coordination and sleep patterns. The tremor was reduced significantly to nearly nonexistent. The subject continued the same self-administered treatment, two to three days a week, and over a three month period the subject slowly lost about 25% of the initial benefit at 9.6 Hz. At this point, rTMS sessions employing the coil 18A producing a flux density of 0.6 gauss but sequentially set at delta, theta, alpha and beta frequencies was initiated and applied during rest in bed with the coil 18A placed under 2 pillows behind the back of head delivering approximately 0.2 gauss to the subthalmic region of the brain. Such treatments applied 2-3 times weekly have produced steady improvement in all aspects of motor and cognitive deficit to a new baseline high.

Case 2: A 76 year old, unmedicated male was diagnosed with Parkinson's disease in 1985. The subject experienced severe tremor in both hands, nighttime anxiety requiring several over-the-counter or prescription sleep aids nightly for the past three years, pain in neck, arms, legs and hands, and weakness in the hands, arms and legs. After one 45 minute rTMS session using the coil 18A of this invention operated to produced a 0.6 gauss magnetic flux density at 9.6 Hz applied to back of head under two pillows, the tremor was reduced by about 50%, the subject slept well with no sleeping pills, anxiety was reduced to near nonexistent, mobility in the hands increased dramatically, overall pain was reduced to 25% of baseline and strength in the hips and legs improved by about 30%. This same therapy was continued three times per week for eight weeks, and the subject continued to sleep well with no sleep medication, the tremor in both hands was reduced to about 15% of baseline and strength in the legs and hands further improved. At week 9, rTMS brain wave sessions using the coil 18A operated to produce a 0.6 gauss flux density at delta, theta, alpha and beta frequencies were initiated but applied during rest in bed with the coil 18A placed under the mattress in alignment with the patient's head and yielding approximately 0.1 gauss to the subthalmic region of the brain. Self-administered magnetic brain-wave entrainment for 30 minutes each evening has resulted in further tremor reduction to 5-10% of baseline and reduced incidence of tremor episodes, better mood, clarity of thinking and further improvement in strength and further reduction of overall pain.

Case 3: A 32 year old female, diagnosed with multiple sclerosis in 1992, exhibited symptoms including an erratic signature, poor mood with mild depression, very little energy and the need to walk with a walker. After three, one hour rTMS sessions using the coil 18A operated to produce a 0.7 gauss flux density magnetic field at delta, theta, alpha and beta frequencies, applied to the back of the head, the subject's mood improved, her facial expressions were more animated with better color in the face, her sleep improved with no need for sleep aids, her signature has improved to nearly normal and her leg strength and stamina have improved about 20%.

Case 4: One of two 40 year old identical twin sisters was administered a 45 minute, rTMS session using the coil 18A operated to produce a 0.7 gauss flux density magnetic field at delta, theta, alpha and beta frequencies applied to back of head under two pillows twice weekly. Although both sisters have been diagnosed as border-line schizophrenics, the twin receiving treatment experienced improvements in mood, motivation and well-being, even while engaged in rigorous academic endeavors which, in the past, had exacerbated her symptoms. Their mother has noted a significant improvement in the treated twin, i.e., the schizophrenic episodes have diminished significantly despite her rigorous academic endeavors, while the other twin remains the same.

Case 5: A 62 year old male had been experiencing balance deficit. Two 45 minute sessions using the coil 18A operated to produce a magnetic field with a flux density of 0.6 gauss at 9.6 Hz, located at the back of the head, corrected the balance deficit by about 80%. The subject has continued self administered therapy once per week and reports balance deficit as nearly nonexistent. The subject also reports sleeping sounder and feeling better overall.

Case 6: A 72 year old female had been experiencing balance deficit when ballroom dancing. One 45 minute rTMS session using the coil 18A operated to produce a 0.7 gauss flux density magnetic field set at 9.6 Hz, and applied through one pillow located at the back of head, improved balance significantly while making the subject sleep recognizably better.

Case 6a: A 73 year old male had been experiencing balance deficit when ballroom dancing. One 45 minute rTMS session using the coil 18A operated to produce a magnetic field having a flux density of about 0.7 gauss at 9.6 Hz, and applied through one pillow located at the back of head, improved balance significantly.

Case 7: A 44 year old male experienced balance deficit while walking the seawall behind his property, and while surfing. He also noticed difficulty driving in that he found himself weaving within lanes especially when taking his eyes off of the road even for a couple of seconds. One 45 minute rTMS session using the coil 18A operated to produce a magnetic field having a flux density of about 0.7 gauss at 9.6 Hz, applied to the back of the head, completely eradicated the balance deficit while restoring his driving proficiency especially in regard to not weaving within highway traffic lane lines. After two additional 45 minute rTMS sessions of the same type noted above, the subject's physical endurance/stamina unexpectedly improved by about 25% as measured by specific exercise movements performed to failure.

Case 8: A 52 year old male noticed cognitive deficit relating to balance, coordination and decision making process. One 45 minute rTMS session using the coil 18A operated to produce a 0.6 gauss flux density magnetic field at 9.6 Hz, applied to back of head through a pillow, resulted in significant improvement in balance, hand-eye coordination and thought/decision making processes.

Pain Management and Treatment of Underlying Conditions Causing Pain

It has been found that the device 10 of this invention is useful in the management of pain and the treatment of a variety of conditions which create pain and inflammation, including, without limitation, bone non-union, bone fractures, joint abnormalities and degenerative disorders, spinal or degenerative disk disorders, neuralgia, neuropathy, chronic pelvic pain syndromes, temporomandibular joint disorder ("TMJ"), post-traumatic maxillofacial defects, fibromyalgia, sciatica, spinal cord injuries and resulting neuromuscular deficits, acute pain and inflammation due to arthritis, periarthritis, osteochondrosis, osteoporosis, trauma, chronic ulcers, diabetic neuropathy or circulatory problems, muscle spasms, muscle soreness, muscle stiffness and others. Two types of therapies are performed with the device 10 to treat these conditions, including one with the coil 18A of FIG. 3 and the other with the coil 18B of FIG. 4.

In one presently preferred embodiment, one or more coils 18A are placed beneath the sleeping surface of a patient in general alignment with the head or other area of the body which is experiencing pain or other symptoms to be treated. The coil 18A can be placed under the mattress pad proximal to the injury or under the entire mattress of the sleeping surface. The time varying magnetic field produced by the coil 18A is approximately 30 gauss in the immediate area of the coil 18A, and preferably in the range of about 0.05 to 0.1 gauss at the interface of the patient and the sleeping surface directly above the coil 18A, depending upon the thickness of the mattress pad or mattress. Preferably, the patient is exposed continuously to the magnetic field from the coil 18A for a normal sleeping period, e.g. 6 to 8 hours per treatment. The treatment is appropriate for rTMS or to stimulate a wide variety of healing mechanisms in damaged tissues.

Sleeping in the DC, time varying magnetic field generated by the device 10 of this invention is believed to induce the repair and regeneration of many types of physical and neurological injuries, to enhance the efficiency of cellular processes including promoting the body's synthesis of adenosine tri-phosphate and to enhance the blood's ability to carry oxygen. This increases the supply of oxygen and nutrients via the vascular system, improves the removal of waste through the lymphatic system and helps to rebalance the distribution of ions across the cell membrane. Healthy cells in tissue have a membrane potential difference between the inner and outer membrane causing a steady flow of ions through its pores. In a damaged cell, the potential is raised and an increased sodium flow occurs. As a result, interstitial fluid is attracted to the area, resulting in swelling and edema. The application of DC pulsed magnetic fields to damaged cells in accordance with the method of this invention accelerates the reestablishment of normal potentials thus tending to increase the rate of healing while reducing swelling, bruising and inflammation.

The coil 18A, or any array of coils 18A, or in cases of acute injury where coil 18B is employed at night proximal to injury, produce homogenous magnetic waves which expand and collapse radially therefrom with little or no eddy currents to the subject being treated. If two persons sleeping in the same bed are to be treated simultaneously, the coils are preferably located at least 3 feet apart and beneath each of the two sleeping subjects to avoid or reduce the intensity of the intersection of the magnetic field produced by each coil.

Case 1: A 44 year old male has been rehabilitating rotator cuff and ligaments in and around hid left shoulder for over 15 years since dislocating the shoulder 3 times during a period of 10 years in sports-related incidents. A coil 18A was placed between the mattress and box spring of the subject's bed, in substantial alignment with the shoulder area, while sleeping. The coil 18A was operated to produce a magnetic field having a flux density of about 0.7 gauss in the area of the coil 18A, and about 0.01 to 0.05 gauss as measured in the area where the subject laid on the bed, at a frequency of 9.6 Hz. After 4 to 5 nights, the constant pain in the shoulder previously experienced by the subject had ended and the subject slept better without being awakened by shoulder pain during the night. By the tenth day, the subject was able to lift weights without the ringing and shooting pain, or the clicking and crunching noises, which he had previously experienced in the left rotator cuff and shoulder ligaments especially while the shoulder was under structural load. Significant improvements in strength and stamina of his upper body during such workouts were evident. The subject has subsequently surfed a number of times and noticed a reduction of all post-session shoulder pain, back pain and neck to nonexistent levels.

Case 2: A 54 year old female complained of attention deficit, a sleep disorder which caused her to awaken at 6 a.m. regardless of what time she went to bed, and persistent weaving in and out of highway lanes while driving. The subject arranged a coil 18A between the mattress and box spring of her bed in general alignment with the back of the head, neck and shoulder area. The coil 18A was operated to produce a magnetic field having a flux density of about 0.7 gauss in the area of the coil 18A, and about 0.01 to 0.05 gauss as measured in the area where the subject's head and neck rested on the bed, at a frequency of 9.6 Hz. After two successive nights of all night therapy, during which time the subject slept deeply for 7 hours, the patient felt noticeably calmer, more focused and she slept better and for longer periods of time. The subject is more comfortable driving and can now take her eyes off of road momentarily without swerving nor is she weaving within lanes. The subject has also in the past suffered from 1-2 migraine headache episodes per month which generally lasted for 2-3 days. The subject reports that upon continued use of the device 10 for 1-3 nights per week, the incidence and severity of migraines have been reduced by at least 80%.

Case 3: A 52 year old male has experienced arthritic pain, rotator cuff, neck and back pain for nearly 10 years. Surfing and spear-fishing significantly exacerbates those symptoms. The subject arranged a coil 18A between the mattress and box spring of his bed in general alignment with the back of the head, neck and shoulder area. The coil 18A was operated to produce a magnetic field having a flux density of about 0.7 gauss in the area of the coil 18A, and about 0.01 to 0.05 gauss as measured in the area where the subject's head and neck rested on the bed, at a frequency of 9.6 Hz. After two all night sessions, the subject experienced approximately a 98% decrease in overall pain which lasted for nearly a week with no additional therapy. Surfing sessions increased from 1 hour to up to 3 hours yet pain remained nearly completely relieved for about one week. Subsequently, the subject reported three weeks later that his overall pain was still only 25% of baseline.

Case 4: A husband suffering from lower back pain, who had 4 back operations in which pins and plates were implanted and who, at time, would rely on a cane for assistance walking, and his wife suffering from tendonitis/bursitis of the elbow and previously wearing an elbow support constantly for 6 months, have been sleeping on a mattress under which a coil 18A was positioned in approximately the center of the bed under the husband's side. The coil 18A was operated to produce a magnetic field having a flux density of about 0.7 gauss in the area of the coil 18A, and about 0.01 to 0.05 gauss as measured in the area where the subjects laid on the bed, at a frequency of 9.6 Hz. They report excellent sleep, the wife remains pain free while the husband has been able to take less-potent, non-habit forming pain medication with further reduction of pain, resulting in increased mobility and no further need for his cane. Husband/wife report wife's migraine episodes have been nearly completely ameliorated since night-time therapy was initiated 6 weeks ago.

Case 5: A husband and wife each 40 years old, initiated sleeping upon coil 18A two days prior to onset of the wife's menstrual cycle. The wife had experienced severe menstrual symptoms for 20 years, and over the previous 18 months such menstrual symptoms worsened due presumably to the onset of menopause. At times, should would be completely debilitated with cramps and migraine headache requiring her for the past several months to take 1-2 days off from work at a time. Her menstrual cycle passed with no migraine headache and minor cramping for the first time in over 20 years.

Case 6: A 62 year old woman experienced knee pain subsequent to knee replacement surgery, resulting in restless sleep. Therapy was initiated by sleeping upon coil 18A for two nights after which time the subject reported better sleep with a 10% reduction in pain. The woman's 10 year old dog which sleeps with her on the bed ceased limping from its recently mended, broken leg.

An alternative method of treatment according to this invention for the management of pain and the treatment of the acute underlying conditions which generate pain, employs the coil 18B depicted in FIG. 4 as part of the device 10. In this embodiment, the coil 18B is placed immediately adjacent the affected areas of the body, i.e., at the location where the pain originates, where the pain is referred, and/or where the condition to be treated is located. Additionally, as noted in case study #15 below, the coil 18B may be located under a sleeping surface as in the treatment method employing the open core coil 18A described above.

Preferably, the patient lies or sits in a comfortable position, with the coil 18B positioned as noted above. Care must be taken to align the center of the electromagnet (North pole) directly with the damaged tissue or pain emanation point so that outer ring (South pole) will be surrounding the adjacent area. The time varying magnetic field produced by the coil 18B has a flux density on the order of about 90 gauss at the surface at frequencies preferably in the range of 8-11 Hz, and most preferably at 9.6 which is the pulsed component of the geomagnetic field. Coil 18B creates a strong bi-polar field gradient in the area of application with the damaged tissues being polarized to North and the surrounding, healthy tissue polarized to South. Coil 18B also is composed of hard ferrous metal so that the area is simultaneously exposed to an oscillating and a static magnetic field. The combination of these three factors-frequency, concentric circle relationship of North and South poles and simultaneous exposure to pulsed and static magnetic fields-results in far more profound healing than any method and system available in the prior art. The duration of the treatment is normally 30 to 60 minutes per session, with the number and frequency of the sessions depending upon the severity of the pain or condition, and the results experienced by the patient. The coil 18B may be moved during the session, as desired, e.g., from the area where the pain is being produced to the area where the pain is referred and visa versa. In most instances, it is preferable to position the coil 18B directly over the pain emanation point initially for about 20-45 minutes, and then move the coil 18B at 5-10 minute intervals to other affected areas such as surrounding nerves, muscles or ligaments.

Case 1: A 39 year old female experienced pain in the hip and thigh area due to severe bone/tissue trauma to her left leg in a motorcycle accident 5 years prior. The pain has been recurrent and worsens during changes in weather. The coil 18B was placed in contact with the body over the pain emanation point(s), for one session of about 30 minutes in duration, and operated to produce a magnetic field having a flux density of about 0.7 gauss at 9.6 Hz. The pain was completely eradicated with no further treatment necessary, even though several frontal systems visited the area during the subsequent 4 weeks.

Case 2: A 25 year old male suffered severe trauma to the neck and spine due to a motorcycle accident, resulting in swelling of the spinal cord and loss of all sensation from the chest down. The subject has been confined to a wheelchair with no pain syndromes. The coil 18B was placed in direct contact with the neck, back, hips, knees and feet for approximately 10 minutes each at approximately equal intervals over a one hour period, and operated to produce a magnetic field having a flux density of about 0.9 gauss at 9.6 Hz. During therapy, the subject reported sensations similar to bugs crawling up and down his legs, which was the first sensation the subject had felt below his chest since his accident. Four days post therapy, the patient reported that sensation had partially returned to left calf muscle and hip. Four weeks later the subject reported that he can still feel the touch of a finger to his right calf and hip.

Case 3: A 29 year old female experienced severe TMJ for 3 years involving her neck and muscles of the face. On the morning of the first therapy session, the subject advised she was considering surgery because the pain was unbearable and nothing had relieved or stopped its progression for nearly 3 years. The coil 18B was placed in direct contact with the right side of the patient's neck, first while laying on her back and then her side for approximately one hour, and operated to produce a magnetic field having a flux density of about 0.9 gauss at 9.6 Hz. The subject fell asleep during the therapy on the neck and when she awoke, she was able to open her mouth wide for the first time in over 8 months. Mobility in the neck area improved 90% on her right side, about 70% on her left side and overall pain was reduced by 85-90% from pre-therapy levels.

Case 4: A 38 year old female has been diagnosed with fibromyalgia, which during flare ups, prevent the subject from raising her right elbow above shoulder level. During a recent flare-up while lying prone on her back, the coil 18B was placed in contact with the intersection of the right trapezius muscle and shoulder for 45 minutes, centered over the spine between the shoulders for 10 minutes and then placed at the lumbar region of the spine for 10 minutes. The coil 18B was operated to produce a magnetic field having a flux density of about 0.9 gauss at 9.6 Hz. Upon sitting upright, the subject was able to make full rotation of right arm with no pain, mobility in neck improved by 80-90% and pain in the neck decreased about 85%. The subject's massage therapist confirmed spasms in neck were reduced by 50% immediately after completion of therapy.

Case 5: A 68 year old female was treated complaining of chronic sciatica in right leg. After 15 minutes of therapy with the coil 18B placed at the point where the pain emanated, the discomfort was reduced by about 60%. Massage therapist confirmed spasms in the gluteus maximus muscle were reduced by 30%. The coil 18B was operated to produce a magnetic field having a flux density of about 0.9 gauss at 9.6 Hz.

Case 6: A 28 year old male athlete has experienced lower back pain for several years due to sports activities, and the condition was exacerbated prior to therapy from to three hours of strenuous surfing in sizable waves and strong currents. The coil 18B was placed in direct contact with the back over a one hour period, and operated to produce a magnetic field having a flux density of about 0.9 gauss at 9.6 Hz. The treatment resulted in nearly complete resolution of pain and associated stiffness. On the following morning, the subject played tennis for 2 hours and reported that the back felt much better upon waking, and the relief remained even after playing tennis. Nearly four weeks post-therapy, the subject reports his back pain is still reduced despite surfing several hours twice weekly and playing tennis at least twice weekly.

Case 7: A 54 year old female complained of severe pain in her feet due to ill fitting shoes worn while traveling and hiking through Alaska. The pain syndrome lasted 8 weeks even after taking cortisone shots beginning at week 7. One eight hour therapy session was performed during sleep with the coil 18B placed in the region of the feet under a mattress pad approximately 8-15 cm from soles of her feet. The coil was operated to produce a magnetic field having a flux density of about 0.9 gauss at 9.6 Hz. A resolution of the pain was obtained with the treatment which did not return after at least 12 weeks subsequent to the therapy.

Case 8: A 42 year old male with a history of several back operations which included metal implants in the lumbar region complained of severe lower back pain for several years requiring the use of a cane at times and the constant use of habit forming prescription pain medications on a daily basis. After an initial 2 hour therapy session with the coil 18B placed in direct contact against the lower back and operating to produce a magnetic field having a flux density of 0.9 gauss set at 9.6 Hz, the subject reported significant relief After 4 therapy sessions of the same duration during a 2 week period, the subject reported tremendous reduction of pain and a significant reduction in his need for pain medications. (See Case Study 4 in previous section).

Case 9: A 40 year old female reported bursitis/tendonitis in her right elbow requiring a brace to be worn most days to reduce pain for several months. After a single 1 hour therapy session in which the coil 18B was placed against the elbow and operated to produce a magnetic field of 0.9 gauss at 9.6 Hz, the subject was pain free for 6 days. Repeated therapy sessions of the same duration for two weeks kept the subject pain free and without need for elbow brace. (See Case Study 4 in previous section).

Case 10: A 40 year old male reported severe lower back pain due to a work related incident. The coil 18B was placed in direct contact with the back over a one hour period, and operated to produce a magnetic field having a flux density of about 0.9 gauss to 9.6 Hz. The subject reported a reduction in pain for 5 days to levels less than experienced in the previous 12 years. He treats himself once per week for several hours while lying on a couch in the same manner noted above, and the pain remains about 70-75% reduced.

Case 11: A 45 year old male reported severe neck pain due to an auto accident, resulting in migraine-like, muscle tension headaches lasting days at a time. During a recent headache episode, the coil 18B was placed in direct contact with the neck over a one hour period, and operated to produce a magnetic field having a flux density of about 0.9 gauss to 9.6 Hz. A complete resolution of the headache episode was obtained.

Case 12: A 60 year old female with a degenerated hip and a candidate for hip replacement experienced moderate back and neck pain with restless sleep. The coil 18B was placed in direct contact with the neck for 45 minutes and lower back, over a 45 minutes during a 90 minute period, and operated to produce a magnetic field having a flux density of about 0.9 gauss at 9.6 Hz. The treatment resulted in nearly complete resolution of pain in neck, back and hip, allowing the patient to sleep. The patient was remarkably pain and medication free for 10 days, and sleep improved to near normal for two weeks.

Case 13: A 49 year old male reported an injury to his hip which occurred during birth, resulting in a crooked walk and in severe hip and knee pain throughout his life. The coil 18B was placed in direct contact with the knee and hip over a 15 minute period, and operated to produce a magnetic field having a flux density of about 0.9 gauss to 9.6 Hz. An 80% reduction of pain was obtained which lasted for 8 days. The same therapy was repeated on two other occasions resulting in a reduction in pain of 75% each time the therapy was repeated. Peak pain during a four week period when therapy was not available was still 50% less than normal.

Case 14: A 70 year old candidate for lumbar fusion, reported severe pain in the lumbar region. The coil 18B was placed in direct contact over the lumbar region for a 45 minute period, and operated to produce a magnetic field having a flux density of about 0.9 gauss to 9.6 Hz. The subject's pain was reduced by about 50% for 48 hours. Two subsequent therapy sessions reduced pain for two days at a time to 50% of baseline.

Case 15: A 59 year old female with a neurodegenerative disease experienced daily moderate to severe pain and weakness in the legs. The coil 18B was placed in direct contact with the legs, feet and knees over a period of three hours, and operated to produce a magnetic field having a flux density of about 0.9 gauss at 9.6 Hz. After a 3 hour self therapy session, the subject experienced a reduction in pain of about 70% and the weakness in her legs was reduced by about 25%. The relief lasted for several days.

Case 16: Two sisters 68 and 69 years old were diagnosed with rheumatoid arthritis over two decades ago, each with progressive inflammatory joint destruction resulting in deformity in the ankles, feet and hands. Each received a single hour of therapy to the base of the spine/lower back. Immediately following therapy, the 69 year old reported overall reduction of pain to 30% of baseline, and the 68 year old reported overall pain reduced to 25% of baseline. Ambulation in both significantly improved.

Case 17: 40 year old male diagnosed with moderate to severe spinal stenosis at C2/C4 utilized the solid core electromagnet through a thin pad directly upon the area for 6 months about 1 hour per day while watching television recently received new MRI which was diagnosed recently as mild to moderate spinal stenosis. Patient is in substantially less pain and is far more active than he has been in years.

Case 18: 41 year old male who suffered a Vioxx-related stroke and was left post event with severe auditory stutter which lasted for two weeks, utilized coil 18B at 90 gauss and 0.5-14.1 Hz upon all quadrants of the head for 30 minutes 2-4 times a day, resulting in complete eradication of the speech deficit within 4 days of initiating treatment.

Although the therapies performed with the coils 18A and 18B have been described separately above, it is contemplated that both may be used by patients as part of an overall treatment regime. For example, a patient may undergo a 30 to 60 minute session with the coil 18B applied directly to pain emanation points followed by treatment with the coil 18A or 18B under the mattress for several hours while sleeping. The beneficial results from such combined therapy are cumulative and synergistic. Effects have been observed including improved quality of sleep that benefits the person's appearance and well being.

Further improved therapeutic results are attainable with an embodiment of the invention as depicted in FIGS. 7-11D. These embodiments exploit the benefits of miniaturization available with present day electronics. Here, substantially all of the functionality described previously, in addition to other features, are all provided as a single "system on a chip" (SOC), which is packaged as shown, for example, in FIG. 7. Such miniaturization results in, among other things, improved energy efficiency (little heat generation) and extremely accurate frequency control (up to 0.01 Hz).

This embodiment, a schematic drawing of which is shown in FIG. 8, comprises, generally, logic board 71 for coil excitation and user control, LCD display 72, four-button function control 75, two side-of-case buttons 76 for power and menu selection, jack for stereo headset 82, integrated speaker 84, USB interface 85, power adapter 90, and software stored within memory on logic board 71.

As will be explained more fully below, this embodiment includes enhanced magnetic field control, enhanced overall control, and enhanced user benefits, as compared to the device 10. For example, as one distinct difference, device 70 preferably includes a display that displays menu/mode choices, field frequency and Time of therapy. As shown more clearly in FIG. 9, four-button function control 75 includes buttons for selecting one of the following exemplary items that would be displayed:

Brain Entrainment
Nap
Sleep
Sleep enhancing recovery and athletic performance
Sleep enhancing REM
Local Recuperation

User interface buttons are also preferably provided for starting operations or for selecting up or down functions for increasing/decreasing, e.g., the volume of audio output through stereo headset 82.

USB connector 85 is preferably used to connect to a computer, portable memory card or, e.g., the Internet, to enable field upgrading of programs and for downloading audio programs. In the preferred implementation, software, upgradeable via USB connector 85 and running on logic board 71, controls the timing for coil excitation.
More detailed descriptions of the features of device 70 are set forth below.

Logic Board

Referring again to FIG. 8, logic board 71 controls the operation of device 70 and can be considered the heart of the system. It preferably comprises a microcontroller 94 along with flash memory 95. Microcontroller 94 monitors the status of the control buttons and accordingly controls the frequency of pulses for excitation of coils. Logic board 71 is also preferably configured to play audio as selected by the user. One possible variation could be to add a SD/MMC (Secure Digital/Multimedia Card) slot, so that, in addition to USB interface/connector 85, a user without a computer system can use SD/MMC for the purposes of downloading program updates and music.

LCD Display

Display 72 displays menu choices, frequency and Time of therapy as shown in FIG. 9. LCD display 72 may include graphics, color, or may be a relatively simple LED display. Menu options preferably include Mode/time (e.g., Elapsed Time/Time left) and audio.

Front Buttons for Control

Buttons 75 are used to select, or to start operations, to select up or down functions, or to increase/decrease audio volume. Buttons 75 are preferably used to select a menu option, scroll through options upward or downward, and start the selected operation. An alternative implementation might include a touch screen, with or without a stylus.

Side Buttons

Side buttons 76 may similarly be used to select up or down functions or to increase/decrease volume. Side buttons may also be used to scroll through a menu. A jog wheel, or similar mechanical device, is preferably employed for this purpose.

Modular Designed Electromagnet

An air core magnet like 18A exhibits relatively clean on-off characteristics, which are different from hard-ferrous solid core magnets with or without an outer jacket, like electromagnet 18B. The modular design of the electromagnetic coil of the present invention, another embodiment of which is shown in FIGS. 10A-C, allows a user to convert a simple air core electromagnet, to a solid-core of hard ferrous metal, resulting in a component with magnetic memory. In another embodiment a "hard" ferrous jacket may also be installed which causes the "south" polarity magnetic field to wrap from the reverse side of the magnet to a given treatment surface, thereby enhancing field gradient in the tissues being treated.

Referring to FIGS. 10A-C, there is shown a modular electromagnet that comprises a hard ferrous metal outer casing 1010 into which is disposed an inner air coil (wire coil is not shown). An outer air coil 1014 is disposed surrounding inner air coil 1012 inside outer casing 1010. Finally, a hard ferrous metal core 1016 is provided in a cavity within inner air coil 1012. FIGS. 10B and 10C show a perspective view and a partial cutaway view of the assembled modular electromagnet.

Stereo Handset

Stereo headset 82 is used for listening to stereo audio programs in privacy. Stereo headset 82 preferably emits either soothing music that contributes to a calm environment, meditation or to provide learning programs during therapy.

Speaker

Speaker 84 can be used in lieu of a headset. The stereo headset jack can also be connected to a home stereo/theatre system.

USB Interface

USB interface or connector 85 is preferably used for connecting to a computer, a memory card, or the Internet for field upgrading of programs and for downloading audio. In one implementation, USB connector 85 can be used to program device 70 to generate synchronized binaural beats and/or light therapy to the pulsed electromagnetic field. USB connector 85 may also be used for remote device diagnostics, in the event a malfunction occurs. USB connector 85 can also be used to upload user characteristics so that patient compliance issues can be addressed by the practitioner or manufacturer. Such diagnostics could occur over the Internet through a predetermined web site. Other connectivity means such as Blue Tooth, Zigbee, Wi-Fi, and the like can also serve to provide connectivity to device 70.

Software

Device 70 is preferably digitally controlled via software loaded in memory 95 on logic board 71. This software controls the timing for coil excitation, playing of music and, as previously mentioned, can be upgraded in the field.

As shown in FIGS. 7 and 8, logic board 71 is powered from electrical mains via power adapter 90. However, device 70 may also be operated from batteries, especially in view of the much lower power consumption resulting from an integrated package. Coil 18A/18B or a modular coil as described above, is/are connected to respective coil jack(s). Headset 82 may be connected to a headset jack.

In a preferred embodiment, the software ensures that device 70 will run a predetermined frequency sweep but over a shorter-or longer duration of time depending on a given setting on device 70. In one implementation, the unit can be set from 15 minutes to 12 hours in steps of 15 minutes. Consequently, a user has the ability to initiate a program, e.g., within 15 minutes of his desired time of wake-up. Thus, if set for 1 hour, the device cycles through a program in 1 hour; if set for 12 hours for, e.g., prolonged sleep, the device preferably cycles through a given program in 12 hours.

An exemplary frequency program is as follows. If the time is set for 1 hour, the Frequency sweeps from 9.6 Hz to 0.5 Hz within 3.75 minutes in steps of 0.1 Hz. Then the frequency alternates between 0.5 Hz and 5 Hz every 3.75 minutes in steps of 0.1 Hz until total elapsed time is 45 minutes (75% complete). Then, the frequency sweeps from 5 Hz to 14.1 Hz in steps of 0.1 Hz. It is of note that 14.1 Hz is one of the predominant Schumann wave frequencies and is an alert active thought process brain wave rhythm which should allow the user to awaken in an alert, refreshed state of mind.

Of course, other programming methodologies are possible, especially in view of the ability to upgrade and modify the software in the field.

More specifically, the digitally-controlled system of the present invention might be programmed to operate at night to balance the electromagnetic fields in the body and brain. Through the concept of brain entrainment or, alternatively, through frequency specific blocking, device 70 can be programmed to sweep through a range of sleep frequencies during the night, smoothing out electromagnetic anomalies while a person is unconsciously sleeping, and thereby subsequently further "upregulating" physical and mental capabilities during the following day. Device 70 can be, for example, tuned to meditation, relaxation, or to put someone to sleep for a nap or all night, in accordance with the mode selected. Alternatively, device 70 may be tuned to "alert" frequencies for daytime use to enhance mental thought process and keep a person in active thought process mode without excess coffee or other nutritive stimulants.

Additionally, it is believed that that even faster healing occurs when a frequency, particularly 9.6 Hz (identified as creating profound analgesic and healing effects), is modulated from 9.1 to 10.1 Hz continually during a treatment session for physiological healing effects.

Studies show that brainwave anomalies in Theta rhythm occur in central nervous system disorders including ADD, migraine headaches, and like disorders. The system of the present invention preferably sweeps the theta region continually all night in 0.1 Hz steps during a nap or while sleeping cycling through REM mode several times during the night.

In each mode, e.g., brain entrainment, nap, sleep, alert, recuperation, device 70 outputs a waveform consistent with pre-programmed timings. These timings and modes are modifiable in the field through company website or memory card.

In a preferred embodiment, device 70/logic board 71 has a built-in security code to ensure that only an authorized website or memory card can modify the timings. On the other hand, audio programs are preferably downloadable without restriction, as long as the format is compatible with device 70. The MP3 format is one such format.

As mentioned previously, device 70 is preferably configured to play music. In one implementation, the software operates to play music in coordination with the particular pulse frequency. This can further enhance the therapeutic effects of the system, especially in connection with sleep and/or resting modes. However, coordinated music is not a requirement of the invention. In the case where the device is connected through the Internet, both music and associated pulse patterns can be downloaded and selected for "playback" by a user.

Continuing research with coils 18A and 18B has resulted in the discovery that these coils can produce peak electromagnetic force (PEMF) up to 30 Gauss and 90 Gauss, respectively, when not penetrating through a barrier, such as a standard 6-10 inch thick mattress.

In still further research, it has been discovered that sleeping with a solid core electromagnet generating 90 gauss and attached to a headboard of a bed 1-2 inches directly above the 'crown' of a user's head results in remarkably deep and restful sleep. Further, sleep, strength, stamina, and general focus all appear to further improve with peak EMF at levels as high as 1000 gauss. These higher levels of EMF can be attenuated as desired through, e.g., a mattress, to produce PEMF in the range substantially of 0 Gauss to 25 Gauss.

Other Embodiments

Battery Powered Medallion

Due to the size reduction and energy efficiency resulting from the SOC technology described herein, device 70 can, alternatively, be fashioned into a wearable power source and coil in the form of a medallion, as shown in FIGS. 11A-D. As shown, medallion 1100 comprises a wire coil wound around form 1110 and encased in a cover 1115. A battery 1120 and microchip 1125 are preferably secured to opposite sides of the coil, or where space is available. Medallion 1100 is preferably configured to include a rechargeable battery that could be recharged at night. The device itself can then be worn all day, using, e.g., a necklace, to neutralize the detrimental effects of man-made EMF and to help the wearer maintain an alert, focused state and reduce psychological and environmental stressors by producing a close field source of VLF magnetic fields.

Sweeping across frequencies between 10 and 14.1. Hz, the wearer's alertness is maintained. At the same time, the wearer's biological system is shielded from stressors. In particular, such frequencies may protect the wearer from detrimental effects of man-made electromagnetic pollution from, e.g., cell phones and wireless networks which create havoc to the neurological and physiological system, by creating a close source of friendly electromagnetic energy of natural endogenous and exogeneous frequencies conducive to active thought process state (e.g., low beta and high alpha ranges), thereby drowning out the electromagnetic pollution of considerably higher frequencies including power frequency, radio and microwaves.

The instant medallion embodiment can also be worn over the heart for stimulation of the heart muscle, or over liver, kidneys, bone disorder, fracture or other tissue, which is failing in health. Medallion 1100 can be placed on the body in the appropriate position by fastening with tape, Velcro or straps. The device can be worn by a user for up to 8-16 hours per day due to variable frequency and very low amplitude.

More specifically, as shown in FIGS. 12A-D, medallion 1100 can be secured in a pouch 1200 that preferably includes a belt loop 1210 through which a belt can be passed and then fastened to the user. FIG. 12A is a front view with medallion 1100 being inserted into pouch 1200 and FIG. 12B is a front view of the pouch. FIG. 12C shows belt loop 1210 and a flap 1220 that can be secured via a hook and loop fastener pair 1240. A belt could be a weight-lifting type belt, or a back-brace 1300, as shown in FIG. 13, to have dual advantage of reducing weight upon the lower back structure and maintaining better disk alignment and posture, while PEMF goes to work enhancing blood flow, relaxing muscles, regenerating damaged bone, cartilage, ligament and nerve tissue.

Far Infrared/Pulsed far Infrared

The features of the present invention may also be combined with far infrared heat for synergistic effects. Research has determined that addition of radiant or far infrared heating pad to physiological therapy applications synergistically affect the therapeutic outcome, therefore an apparatus for generating continuous or pulsed infrared heat could be combined with the weight-lifting type fixation belt and PEMF concentric circle coil described above. This combination is proving to be particularly effective with respect to sever lower back pain symptoms.

Medallion For Helmet Retrofit

More stable and reliable neurological entrainment and stimulation can be achieved by adapting the medallion style device to be incorporated into a helmet, as shown in FIGS. 14A and 14B, for use in military, auto racing or any sport/activity which normally requires high degree of concentration over long periods of time, especially where fatigue and stress can be problematic. In the case of FIG. 14A, the medallion style device is inserted into padding 1410 of helmet 1400, whereas in FIG. 14B, a pouch-like device, like pouch 1200, is used to attach the device to internal webbing 1420 of helmet 1400. In a military application, troops in the field will be able to maintain focus and reduce stress in face of combat fatigue by utilizing the device in high alpha-low beta range during day. When the time arrives for napping or sleeping, the device can be tuned to promote relaxation, which is more conducive to sleeping and physical recuperation. More importantly, while the user sleeps, the body is in more recuperative state than normal so that one can maintain effectiveness on substantially less sleep and repair normal wear and tear by substantially more effective means.

As enhanced strength, stamina and motor coordination are all routinely upregulated by the device, soldiers wearing the device will be more effective in situations of hand to hand combat and others where strength to weight ratio are important.

Medallion for Bone Stimulation and Tumor Treatment

A medallion device consistent with the present invention can also be positioned in proximity to a fracture or bone non-union on the outside of a cast or fixation device.

Further, as it is being discovered more and more that cancer cells do not propagate well within a North Pole field (because they themselves carry a South Pole field) the medallion or the other configurations described herein can be worn in proximity to a user's known cancerous tumors.

Smart Home Application

Devices in accordance with the present invention can also be configured to be plugged directly into any electrical wall outlet or even hardwired into a home to create a home network whereby the devices are controlled by the home's computer or set individually in each room according to a user's preferences. Preferably the device creates a frequency 'zone' in any room that overshadows any other EMF present in that room. As an example, a bedroom can be tuned to a relaxed state at night, and an alert state in the morning whereas an office can be tuned to alert frequencies 24 hours per day.

Automobile Application

As the detrimental effects of EMF become more well-known, it will be found that one of the most detrimental EMF spaces are those in an automobile. These EMF's are concentrated forward of the passenger compartment putting the driver and his front seat passenger in proximity to the largest fields. The device of the present invention can be configured to be powered from a cigarette lighter and to direct North Pole energy into the passenger compartment. The device may also be programmed to scan the 'alert' state frequencies whenever plugged in, and thereby help to maintain a driver's alert state.

Agriculture

The frequencies and EMF magnitudes described herein have been used successfully to improve the growth rate and taste of various fruits and vegetables in a small garden setting with electromagnetic field generating coils spaced throughout a small plot of vegetables and fruits. With only a negligible increase in electric consumption, the average home gardener may experience an increased yield of up to 20% per square meter of garden which will yield tastier and perhaps more nutritious fruits and vegetables.

Aquaculture

In a similar way to the agriculture application, exposing hatchlings to the electromagnetic energy described herein may also be used to enhance growth rate, thereby. Hatchlings could be in tanks or small ponds into which the EMF can be directed significantly reducing the time required before harvesting.

PEMF Used Synergistically with Silver Ion Wound Dressings

It is now well established that a silver ion wound dressing is bacteriostatically superior to all other wound care therapies. What is proposed herein is the novel synergistic use of extremely low frequency, pulsed North pole/South pole concentric circle application of electromagnetic energy at 100-1000 gauss, using an 8 hours on/8 hours off cycle, although other timing cycles or electromagnet configurations such as air-core or medallion embodiments could be employed. This energy is preferably applied through moistened silver nylon bandages with or without use of microampere negative D.C. electric currents.
While the invention has been described with reference to a preferred embodiment, it should be understood by those skilled in the art that various changes may be made and equivalents substituted for elements thereof without departing from the scope of the invention. In addition, many modifications may be made to adapt a particular situation or material to the teachings of the invention without departing from the essential scope thereof. Therefore, it is intended that the invention not be limited to the particular embodiment disclosed as the best mode contemplated for carrying out this invention, but that the invention will include all embodiments falling within the scope of the appended claims.

Method and apparatus for the treatment of physical and mental disorders with low frequency, low flux density magnetic fields
US7276020

---

http://ston.jsc.nasa.gov/collections/TRS/_techrep/TP-2003-212054.pdf
September 2003
NASA/TP-2003-212054

PHYSIOLOGICAL AND MOLECULAR GENETIC EFFECTS OF TIME-VARYING ELECTROMAGNETIC FIELDS ON HUMAN NEURONAL CELLS

by

Thomas J. Goodwin, Ph.D.

---

http://www.ncbi.nlm.nih.gov/pubmed/2537205
Eur J Biochem. 1989 Feb 1;179(2):275-80.

Chronic stimulation of rat skeletal muscle induces coordinate increases in mitochondrial and nuclear mRNAs of cytochrome-c-oxidase subunits.

by

Hood DA, Zak R, Pette D.

Abstract

Fast-twitch tibialis anterior muscle of the rat was subjected to chronic low-frequency (10 Hz, 10 h daily) nerve stimulation in order to investigate the time course of changes in cytochrome-c-oxidase activity, as well as in tissue levels of specific mitochondrially and nuclear-encoded, cytochrome-c-oxidase-subunit mRNAs. Chronic stimulation induced a progressive increase in cytochrome-c-oxidase activity which was threefold elevated after 35 days. A similar increase was recorded for citrate-synthase activity. Glyceraldehyde-3-phosphate dehydrogenase, which was studied as a glycolytic reference enzyme, moderately decreased, as did the tissue level of its corresponding mRNA. There was a parallel increase in the tissue levels of the two cytochrome-c-oxidase-subunit mRNAs over the entire stimulation time course. The extent of increase (stimulated/control) was 2.4 +/- 0.3 and 1.8 +/- 0.2 (means +/- SEM) for the mitochondrial and nuclear subunit mRNAs, respectively. This parallel increase suggested a coordinate regulation of the two subunits. The increase in cytochrome-c-oxidase activity initially corresponded to the changes at the mRNA level. However, with longer stimulation times (beyond 14 days), the increase in cytochrome-c-oxidase activity clearly exceeded that of the two mRNAs. This divergence was progressive and was interpreted to indicate that the increase in cytochrome-c-oxidase content was brought about not only by changes in the levels of the specific mRNAs, but also by alterations at the level of translation.

---

http://www.earthpulse.net/index.htm#Pulsed_Magnetic_Therapy_Research

Pulsed magnetic field therapy bibliographies

---

http://www.biolbull.org/cgi/content-nw/full/212/2/169
Biol. Bull. 212: 169-175. (April 2007)

Mitochondrial Delivery Is Essential for Synaptic Potentiation

James Jiayuan Tong

Abstract -- Mitochondria, as portable generators that power synaptic function,regulate the ATP supply and calcium homeostasis in the neuron.As molecular interactions within the synapses before and afterthe potentiation are beginning to be elucidated, the decidingmoment during the tetanic stimulation that gives rise to thestrengthening of the synapse remains a mystery. Here, I recordedelectrically from an intact Drosophila nervous system, whilesimultaneously using time-lapse confocal microscopy to visualize mitochondria labeled with green fluorescent protein. I showthat tetanic stimulation triggers a fast delivery of mitochondriato the synapse, which facilitates synaptic potentiation. Rotenone,an inhibitor of mitochondrial electron transport chain complexI, suppresses mitochondrial transport and abolishes the potentiationof the synapse. Expression of neurofibromin, which improvesmitochondrial ATP synthesis in the neuron, enhances the movementsof mitochondria to the synapse and promotes post-tetanic potentiation.These findings provide unprecedented evidence that the mitochondrialdelivery to the synapse is critical for cellular learning.

---

http://www.ncbi.nlm.nih.gov/pubmed?term=Wei YH 259-70
Zhonghua Yi Xue Za Zhi (Taipei). 2001 May;64(5):259-70.

Mitochondrial theory of aging matures--roles of mtDNA mutation and oxidative stress in human aging.

Wei YH, Ma YS, Lee HC, Lee CF, Lu CY.

Abstract

Mitochondrial theory of aging, a variant of free radical theory of aging, proposes that accumulation of damage to mitochondria and mitochondrial DNA (mtDNA) leads to aging of humans and animals. It has been supported by the observation that mitochondrial function declines and mtDNA mutation increases in tissue cells in an age-dependent manner. Age-related impairment in the respiratory enzymes not only decreases ATP synthesis but also enhances production of reactive oxygen species (ROS) through increased electron leakage in the respiratory chain. Human mtDNA, which is not protected by histones and yet is exposed to high levels of ROS and free radicals in the matrix of mitochondria, is susceptible to oxidative damage and mutation in tissue cells. In the past decade, more than one hundred mtDNA mutations have been found in patients with mitochondrial disease, and some of them also occur in aging human tissues. The incidence and abundance of these mutant mtDNAs are increased with age, particularly in tissues with great demand for energy. On the other hand, recent studies have revealed that the ability of the human cell to cope with oxidative stress is compromised in aging. Comparative analysis of gene expression by microarray technology has shown that a number of genes related to oxidative stress response are altered in aging animals. We discovered that the transcripts of early growth response protein-1, growth arrest and DNA damage-inducible proteins and glutathione S-transferase genes are increased in response to oxidative stress in human skin fibroblasts. Moreover, the activities of Cu,Zn-SOD, catalase and glutathione peroxidase decrease with age, whereas Mn-SOD activity increases with age up to 65 years and slightly declines thereafter in skin fibroblasts. Such an imbalance in the function of antioxidant enzymes may result in excess production of damaging ROS in the cell. This notion is supported by the observation that intracellular levels of H2O2 and oxidative damage to DNA and lipids are significantly increased with age of the fibroblast donor. Furthermore, the mitochondrial pool of reduced glutathione declines and DNA damage is enhanced in aging tissues. Taken together, these observations and our previous findings that mtDNA mutations and oxidative damage are increased in aging human tissues suggest that mitochondrial theory of aging is mature.

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

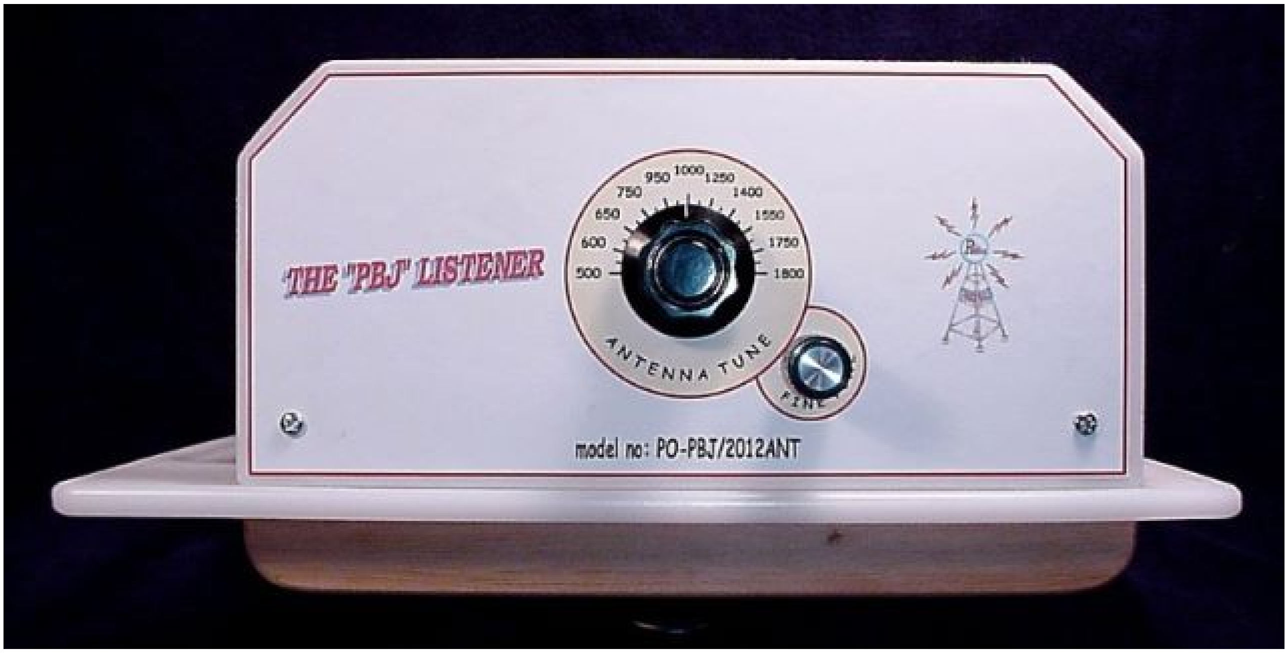
THE "PBJ" LISTENER
500
600
650
750
950
1000
1250
1400
1550
1750
1800
ANTENNA TUNE
FINE
model no: PO-PBJ/2012ANT

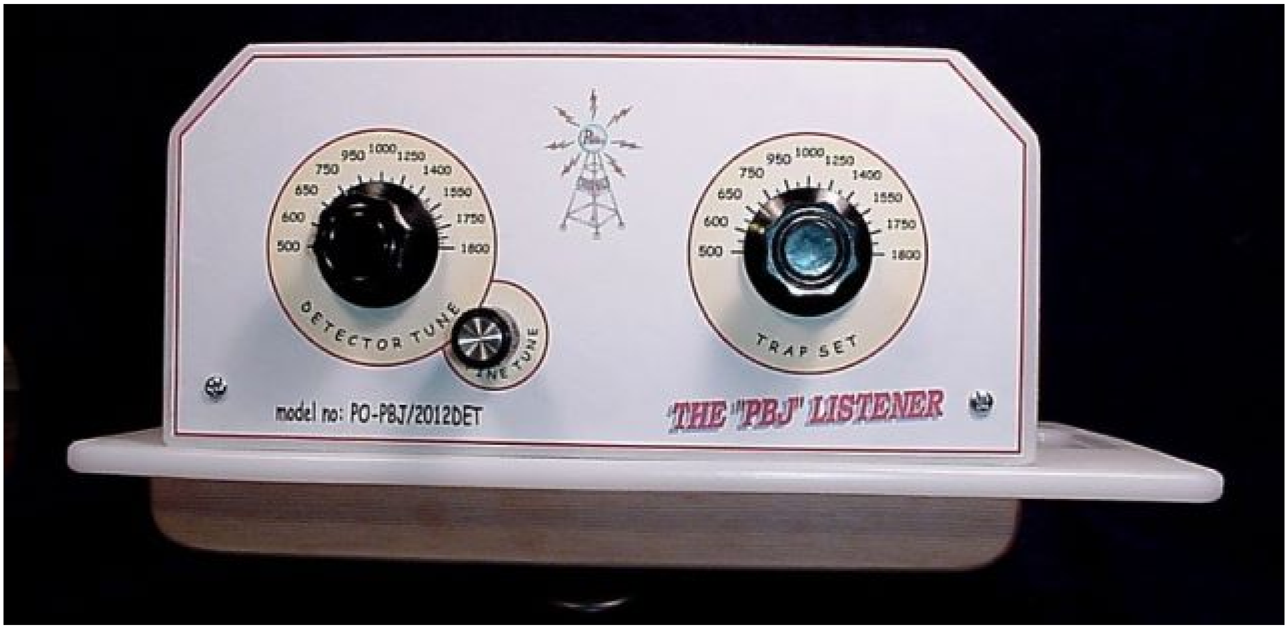
500
600
650
750
950
1000
1250
1400
1550
1750
1800
DETECTOR TUNE
FINE TUNE
TRAP SET
model no: PO-PBJ/2012DET
THE "PBJ" LISTENER

A- (javascript:changeFontSize(-1)) A+ (javascript:changeFontSize(1)) (#)

The main objectives of this project are, to make a Double-Tuner Crystal Radio Set with: 1) Careful and well thought-out construction practices; 2) As low-cost as possible; 3) Using most-available parts and materials, and 4) Keeping design as low-loss as possible, as well as High Performance, with all the aforementioned.

Note: The pictures of the prototype set, has some component/hardware positioning a little different from from the illustrations and drawings. As usual, the prototype showed me a few places where positioning can be improved, and cleaned-up. The illustrations will prove to be best way to construct the set. The prototype and drawings, are electrically/electronically the same. The drawings are just a little cleaner.

The set is built on two separate, 8-7/8” x 13-7/8” x 3/8” (approx.) Cutting Boards, made of HDPE plastics. Found in any store that provides a nice variety of household products. The Coils are made on 5” x 5” (approx.) Plastic, Generic Peanut Butter Jars. The coils are wound with 22ga, PVC-coated, solid hook-up wire that I purchased from my local Radio-Shack Store. The Front Panels/Sub-Panels, are made of 1/8” HDPE Plastic, that can be obtained in most scrap barrels of most Plastics Suppliers. The rest of the parts are readily available from this website, Peebles Originals (../index.php), your local hardware store, and other, common sources.

The main-tuning, Variable Capacitors P/N 15-007 (../catalog/15.php) were used for two reasons: 1) They are “Frequency Linear” (offset), for easier tuning on the higher-end of the BC Band, and 2) By paralleling the two sections with an Alligator Clip-Lead, then the coil only needs to be 140uHy. Thus, the BC-Band is split into a “hi-Band”, and “Lo-Band” mode. Note that Alligator Clip-Leads are used in this project, rather than switches to keep losses minimal, project expenses low.

The Fine-Tuning control is comprised of: a 1/16” x 3” disk, attached to the shoulder/shaft of the non-metallic shaft-coupler. The drive mechanism is a ¼-Dowel, two ¼” I.D., 5/16” O.D. Rubber Grommets. Jumbo Paper Clips are used to help maintain tension while rotation is smooth and positive. See pictures for details of the Fine Tuning Mechanism.

The proto-type set has a built-in TRAP-Circuit, using a 410pf Single-Gang Variable Capacitor P/N 15-001 (../catalog/15.php), and a coil of 230uHy. The Coil is 28ga Magnet Wire, wound on the far-end of the DET COIL, to trap those overwhelming stations you may have, near-by. I have two: 860/1550, and the trap is most necessary, here. If your area is not plagued with stations like this, the TRAP may not be necessary.

The circuitry for both ATU, and DET/AUDIO sections are very conventional, and using commonly-available parts components, parts and hardware. My initial tests of the prototype have proven to be very impressive. The set-up is very selective, and sensitive, and was able to pull-in a lot of nice DX.

The set, and its usage is a learning-curve, so please be patient with yourself, as you will find, with practice, the total capabilities of the set’s performance.

If you have an interest in building the set, or any variations there-of, please, contact me for a custom-made kit of parts, suited to your needs. I will give some package discounts, for any orders exceeding $65.00, and I will then calculate a package discount (5% to 10%), depending on how much your needs will be. E-mail me. (../order.php)

Additional to aforementioned items, I can supply:

#78-014 (../catalog/78.php), F0-215 FO-215 Diode
#78-015 (../catalog/78.php) Ten FO-215 Diodes
#72-725 (../catalog/72.php), Bogen Transformer
Also, Chokes, Resistors, Capacitors, etc. . .Contact me for your needs.

Hope this project proves to be interesting, fun, and very satisfactory for all you fine Radio Hobbyists.

Will try to come-up with new, and interesting projects, as time goes-on.

Mike Peebles

BTW: I am offering the Proto-Type of the PBJ Listener for: $150.00 + Shipping.

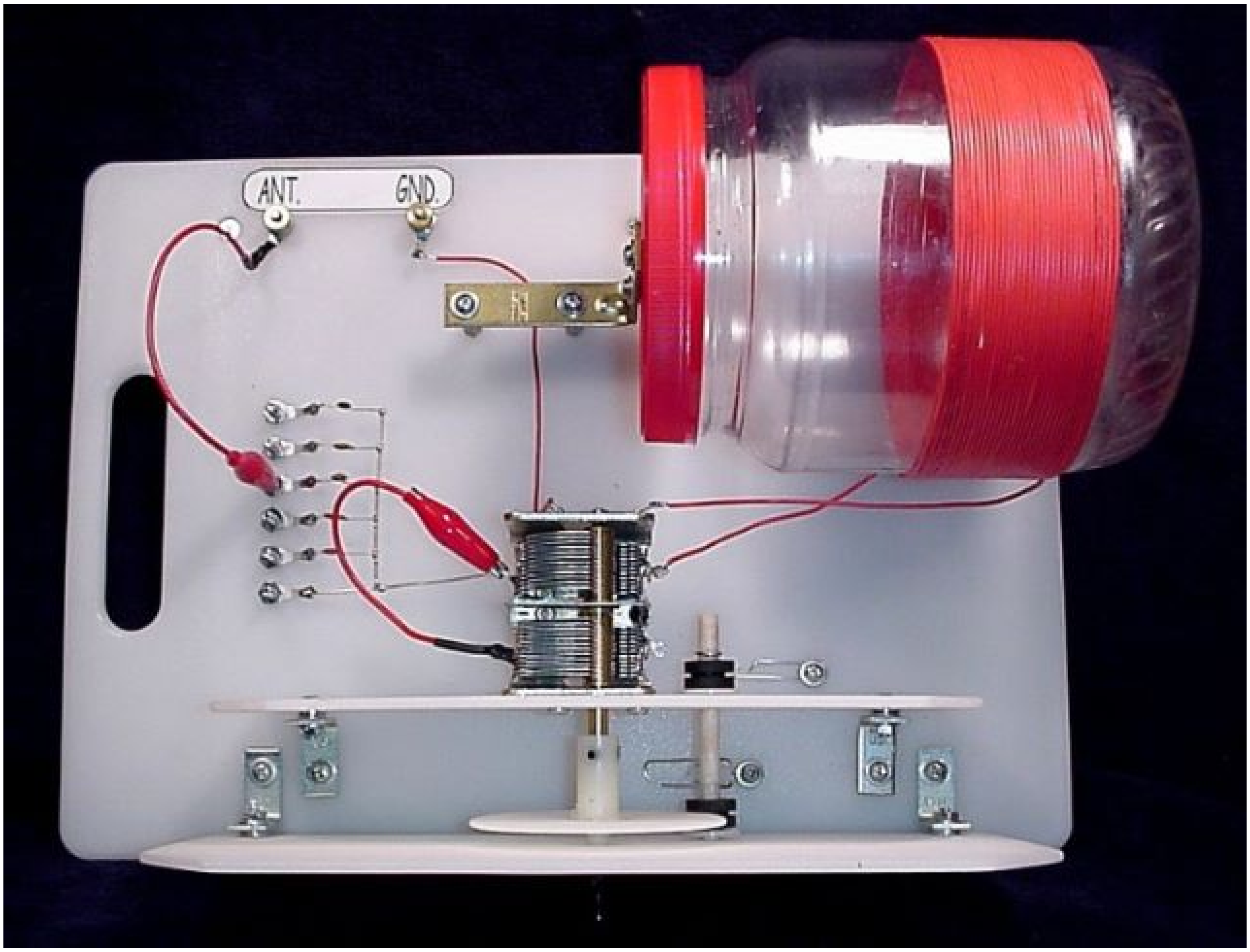

Antenna Tuning Unit, Top View

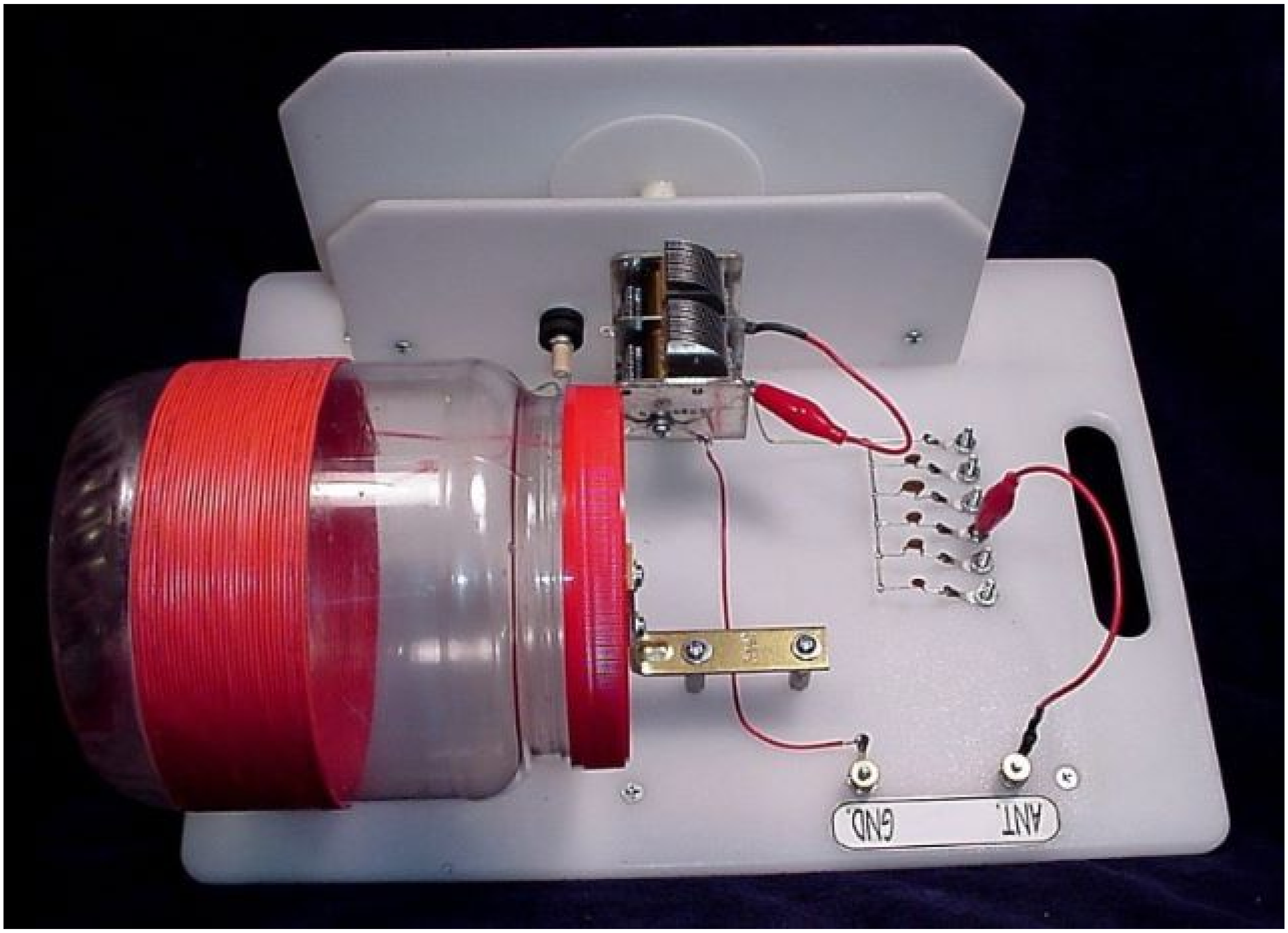

Antenna Tuning Unit, Side View

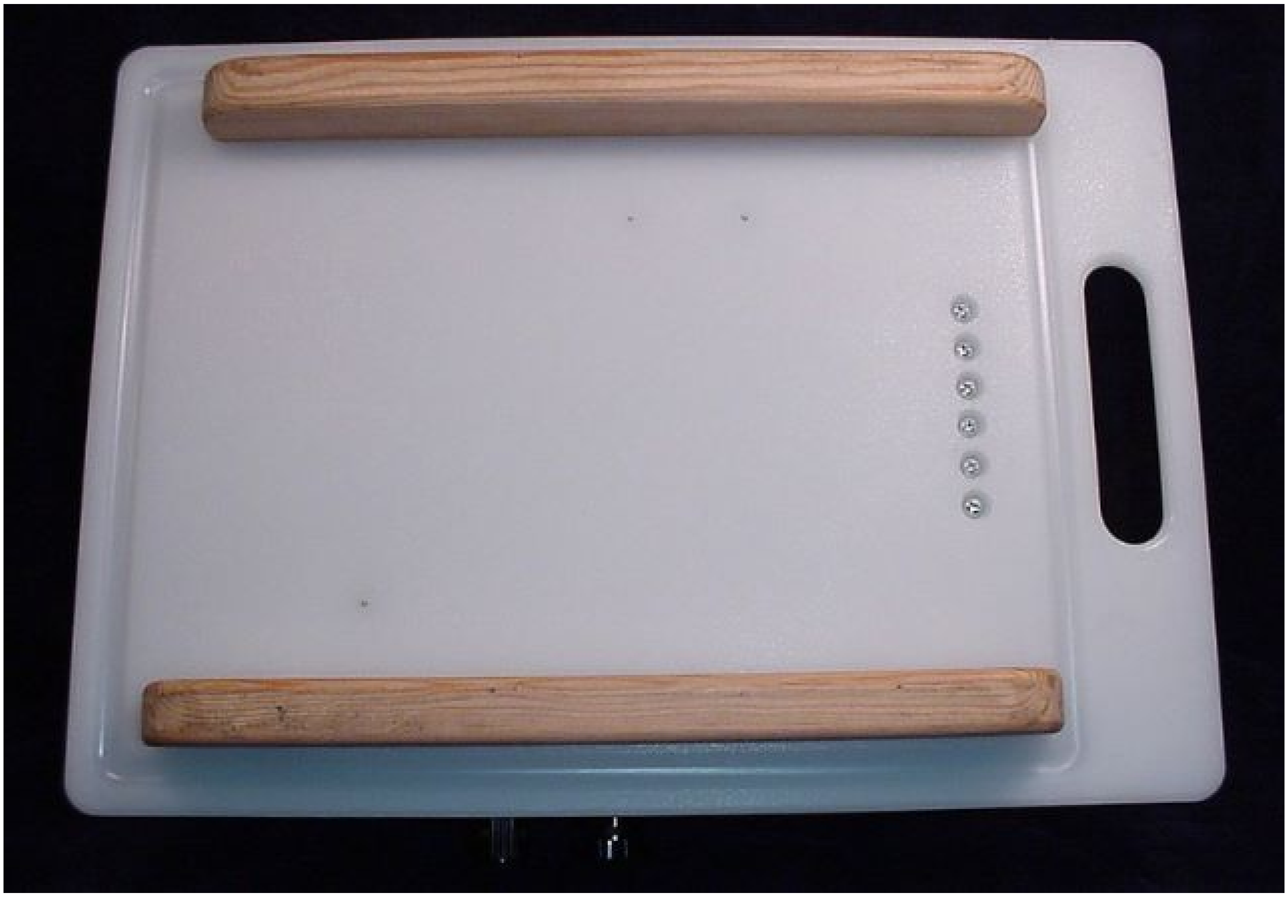

Antenna Tuning Unit, Bottom View

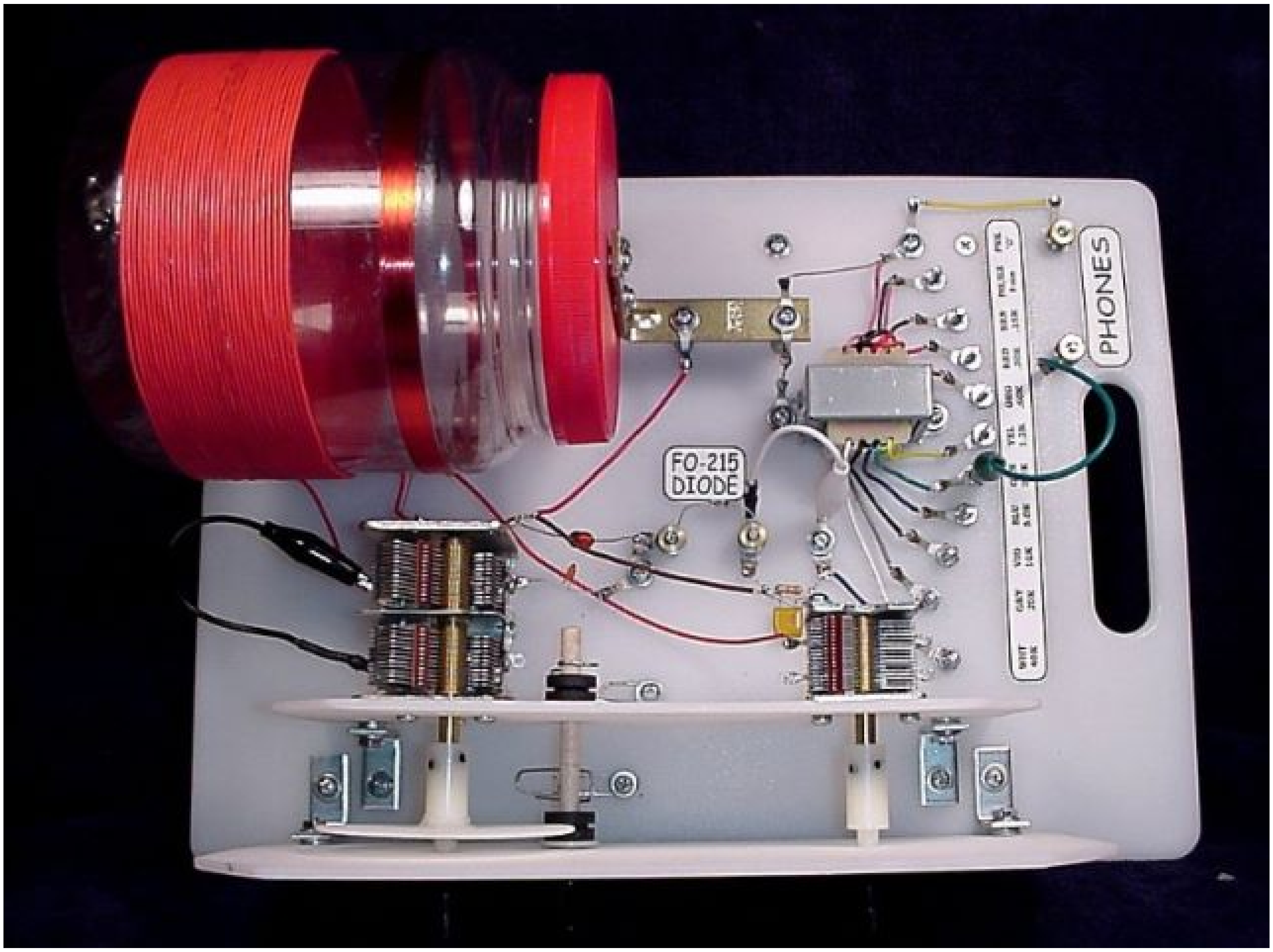
FO-215
DIODE
PHONES

Detector Unit, Top View

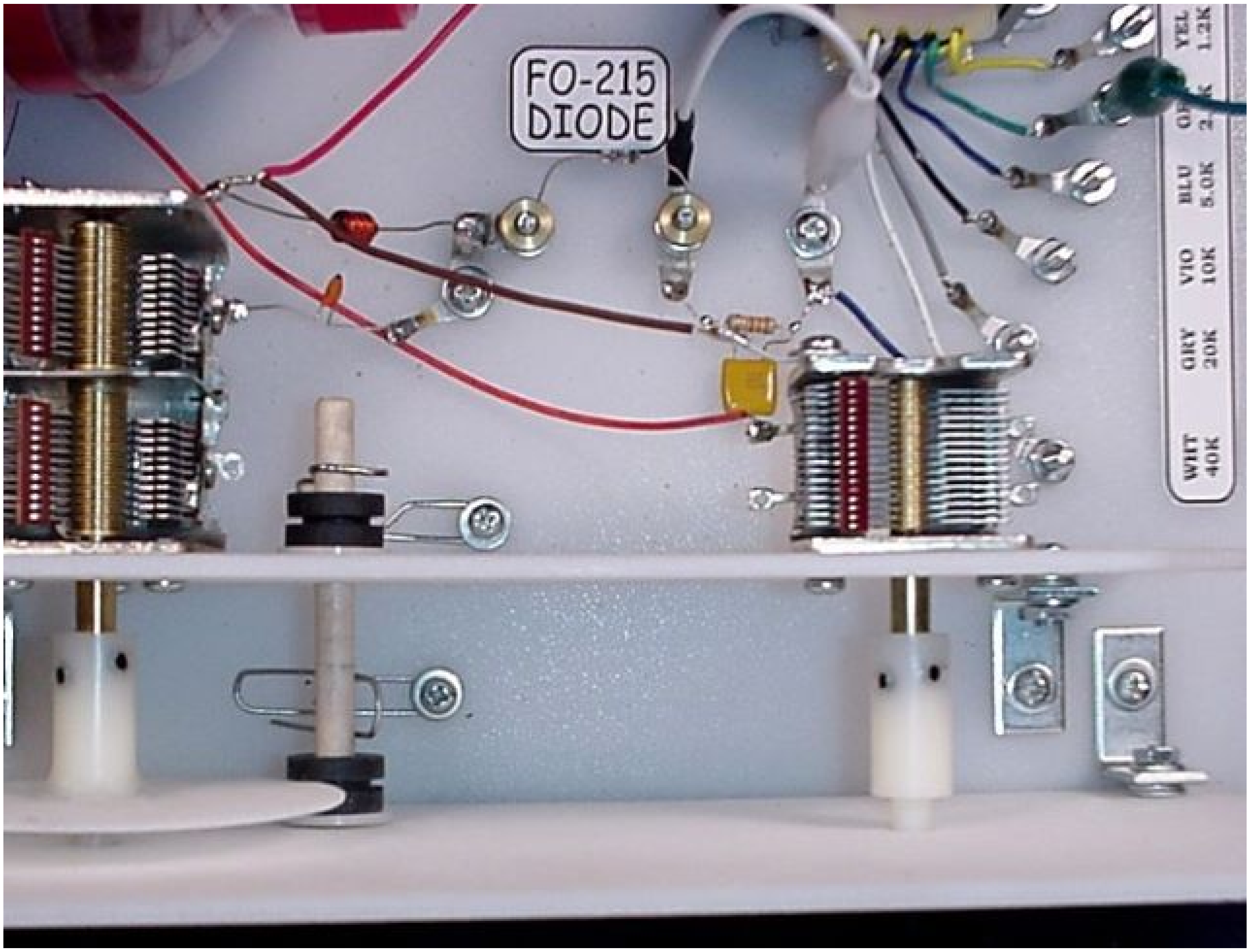
FO-215
DIODE
YEL 1.2K
BLU 5.0K
VIO 10K
GRY 20K
WHT 40K

Detector Unit, Detail View

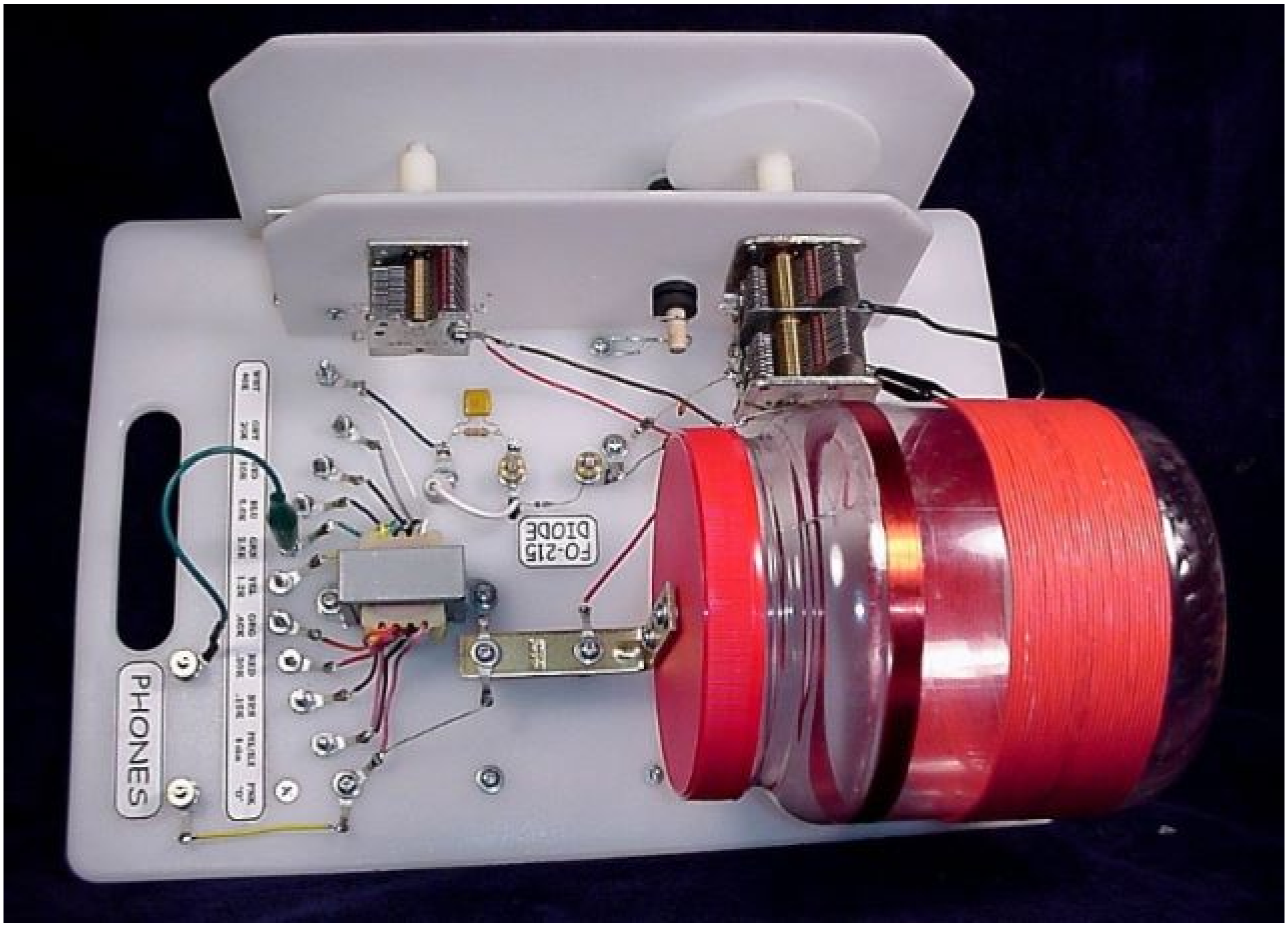
PHONES
FO-215
DIODE

Detector Unit, Back View

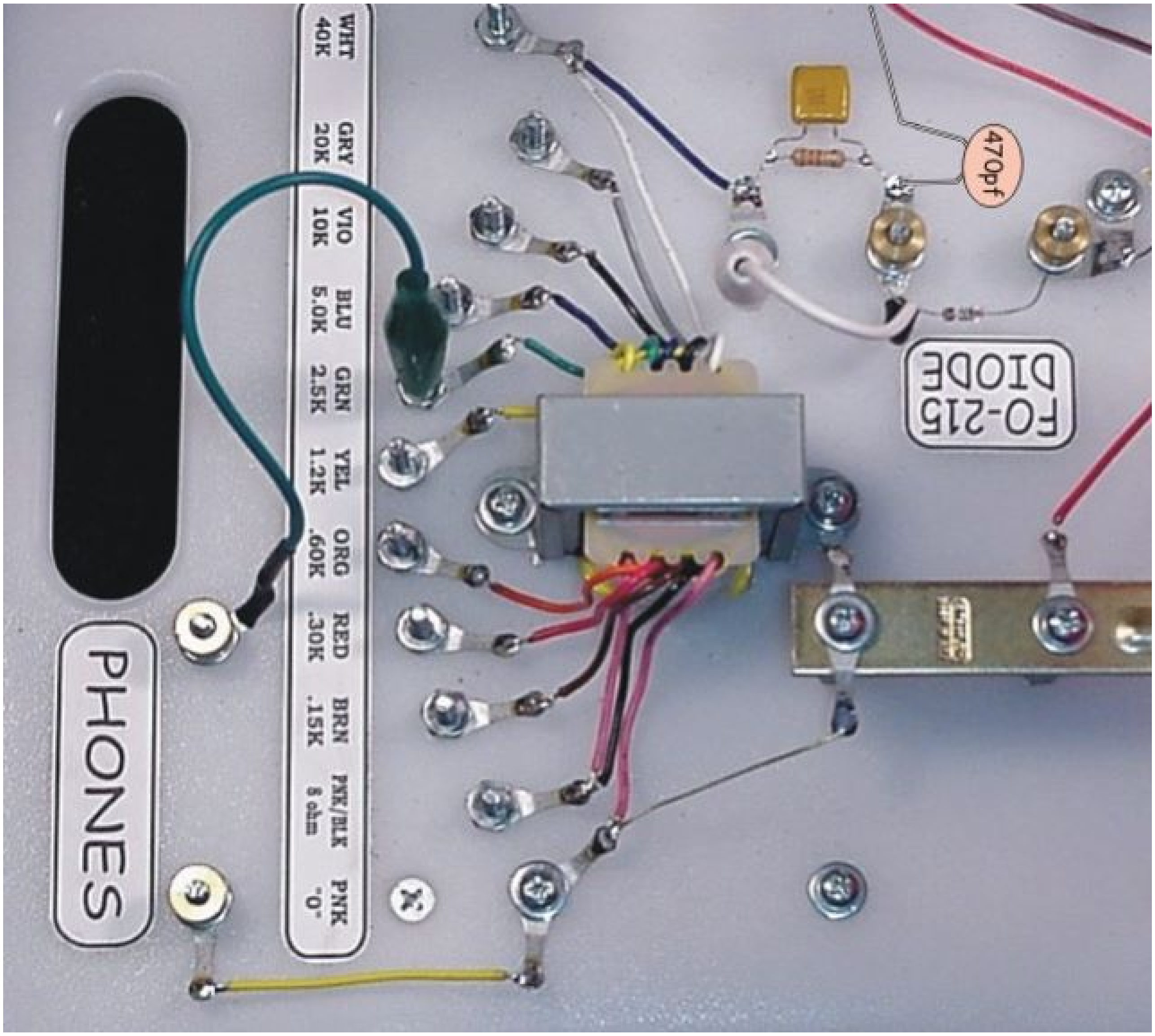
WHT 40K
GRY 20K
VIO 10K
BLU 5.0K
GRN 2.5K
YEL 1.2K
ORG .60K
RED .30K
BRN .15K
PNK/BLK 8 ohm
PNK "0"
PHONES
470pf
FO-215 DIODE

Detector Unit, Matching Transformer Detail

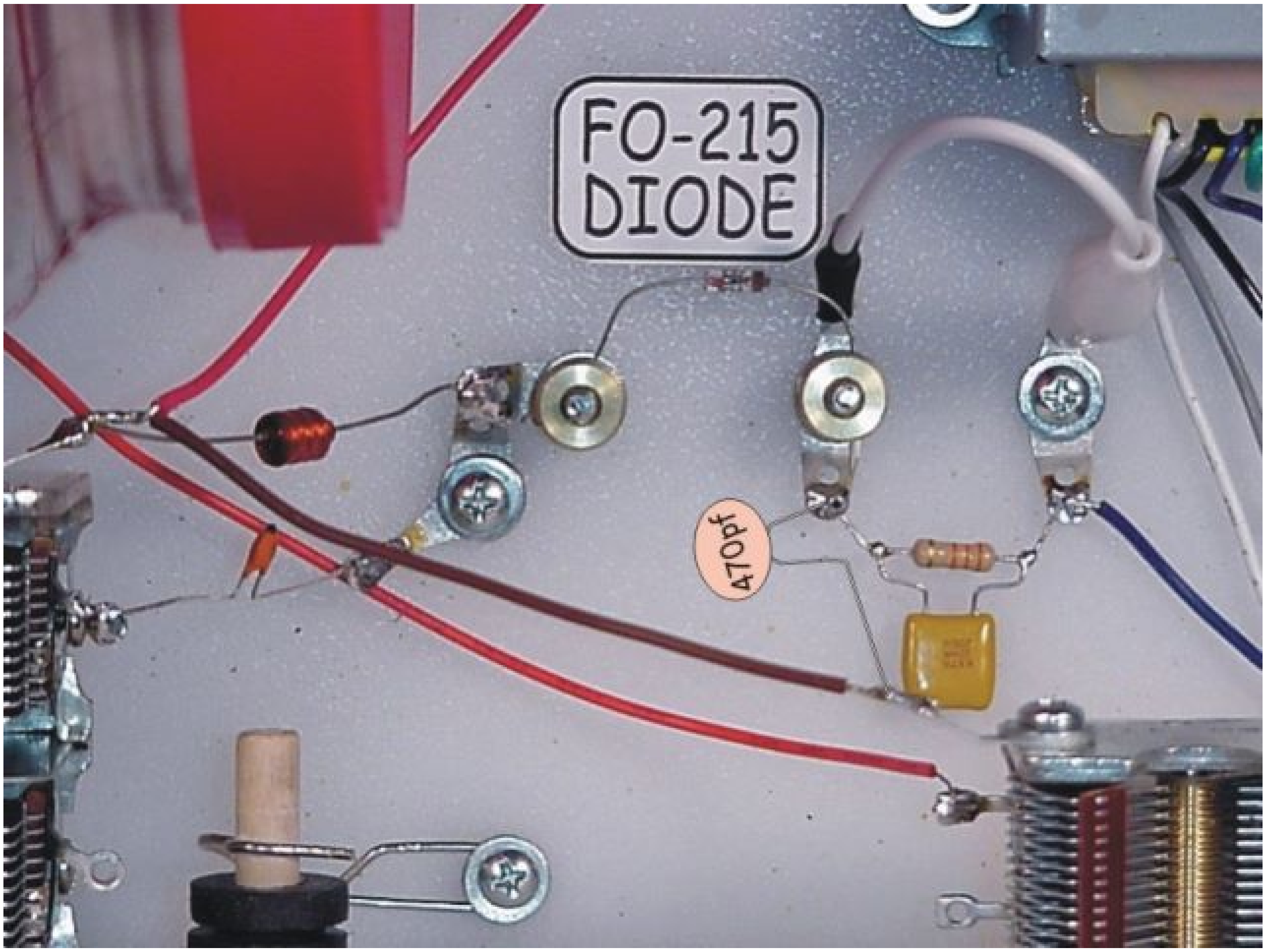
FO-215
DIODE
470pf

Detector Unit, Detail View

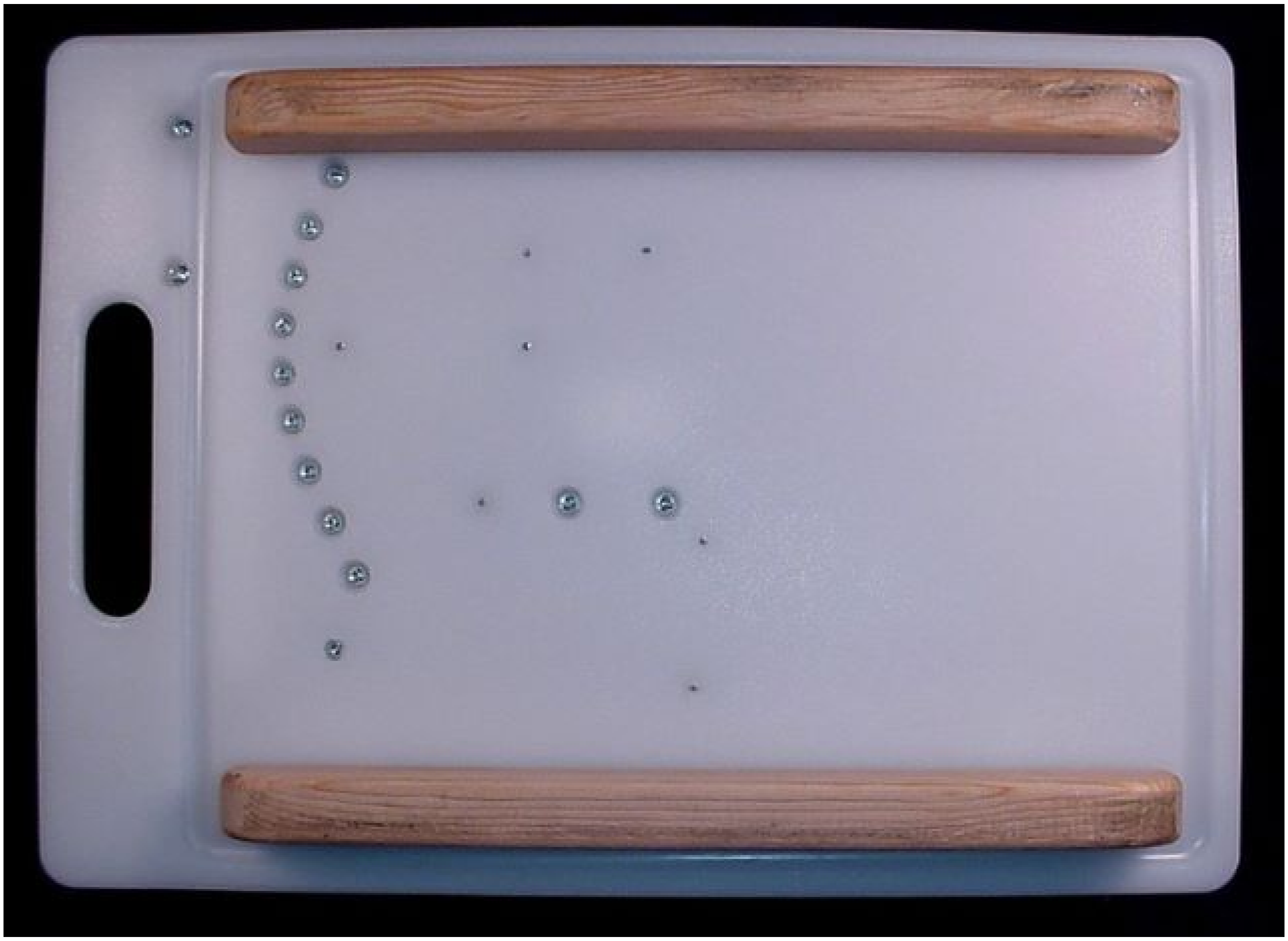

Detector Unit, Bottom View

Template Section – Click For Large View

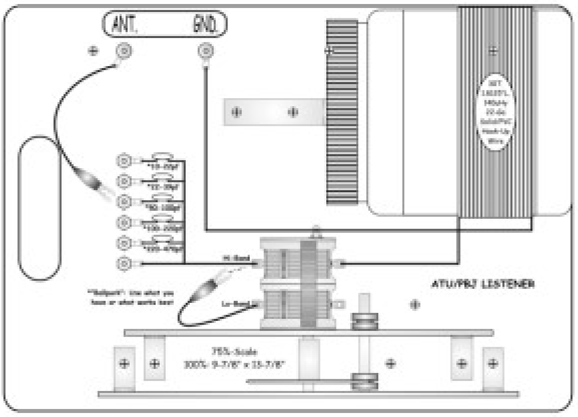
ANT.
GND.
ATU/PBJ LISTENER
Hi-Band
Lo-Band
75%-Scale
100%: 9-7/8" x 13-7/8"

[(img/pbj1a.jpg)](img/pbj1a.jpg)

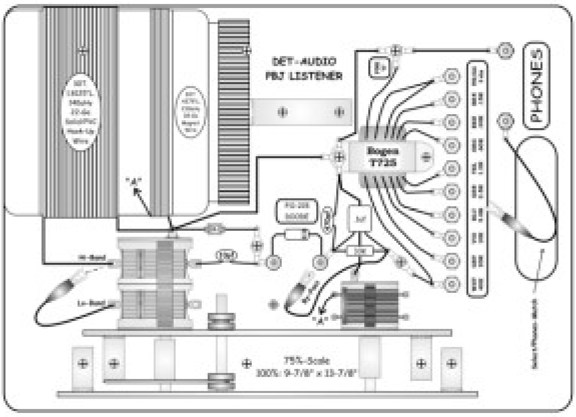
DET-AUDIO
PBJ LISTENER
PHONES
Bogen
T725
"A"
"A"
Hi-Band
Lo-Band
75%-Scale
100%: 9-7/8" x 13-7/8"

[(img/pbj1b.jpg)](img/pbj1b.jpg)

[Antenna Tuner Layout (img/pbj1a.jpg)](img/pbj1a.jpg)

[Detector Layout (img/pbj1b.jpg)](img/pbj1b.jpg)

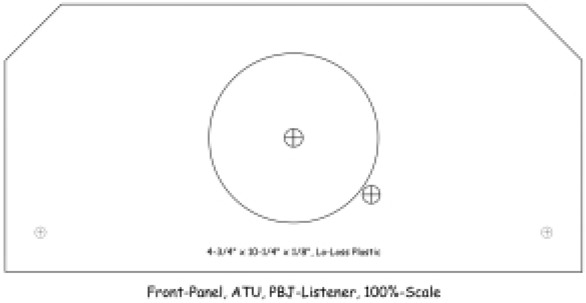
4-3/4" x 10-1/4" x 1/8", Lo-Loss Plastic
Front-Panel, ATU, PBJ-Listener, 100%-Scale

(img/pbj1c.jpg)

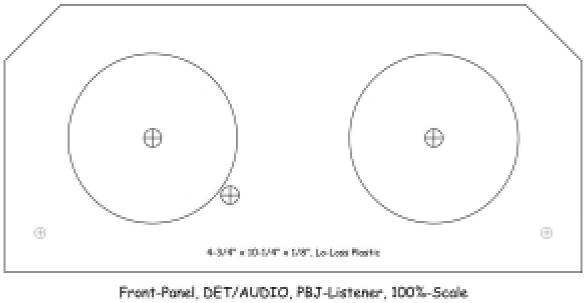
4-3/4" x 10-1/4" x 1/8", Lo-Loss Plastic
Front-Panel, DET/AUDIO, PBJ-Listener, 100%-Scale

(img/pbj1d.jpg)

Antenna Tuner Panel Layout (img/pbj1c.jpg)

Detector Panel Layout (img/pbj1d.jpg)

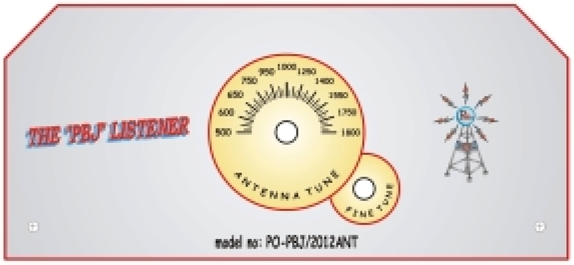
THE "PBJ" LISTENER
500
600
650
750
950
1000
1250
1400
1550
1750
1800
ANTENNA TUNE
FINE TUNE
model no: PO-PBJ/2012ANT

(img/pbj1g.jpg)

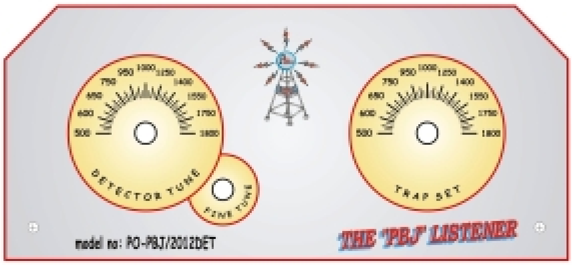
500
600
650
750
950
1000
1250
1400
1550
1750
1800
DETECTOR TUNE
FINE TUNE
TRAP SET
model no: PO-PBJ/2012DET
THE "PBJ" LISTENER

(img/pbj1f.jpg)

Antenna Tuner Front Panel Overlay (img/pbj1g.jpg)

Detector Front Panel Overlay (img/pbj1f.jpg)

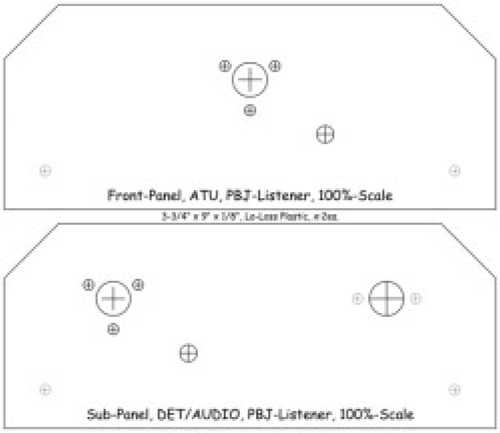
Front-Panel, ATU, PBJ-Listener, 100%-Scale
3-3/4" x 9" x 1/8", Lo-Loss Plastic, x 2ea.
Sub-Panel, DET/AUDIO, PBJ-Listener, 100%-Scale

(img/pbj1e.jpg)

ATU & DET Sub Panels (img/pbj1e.jpg)

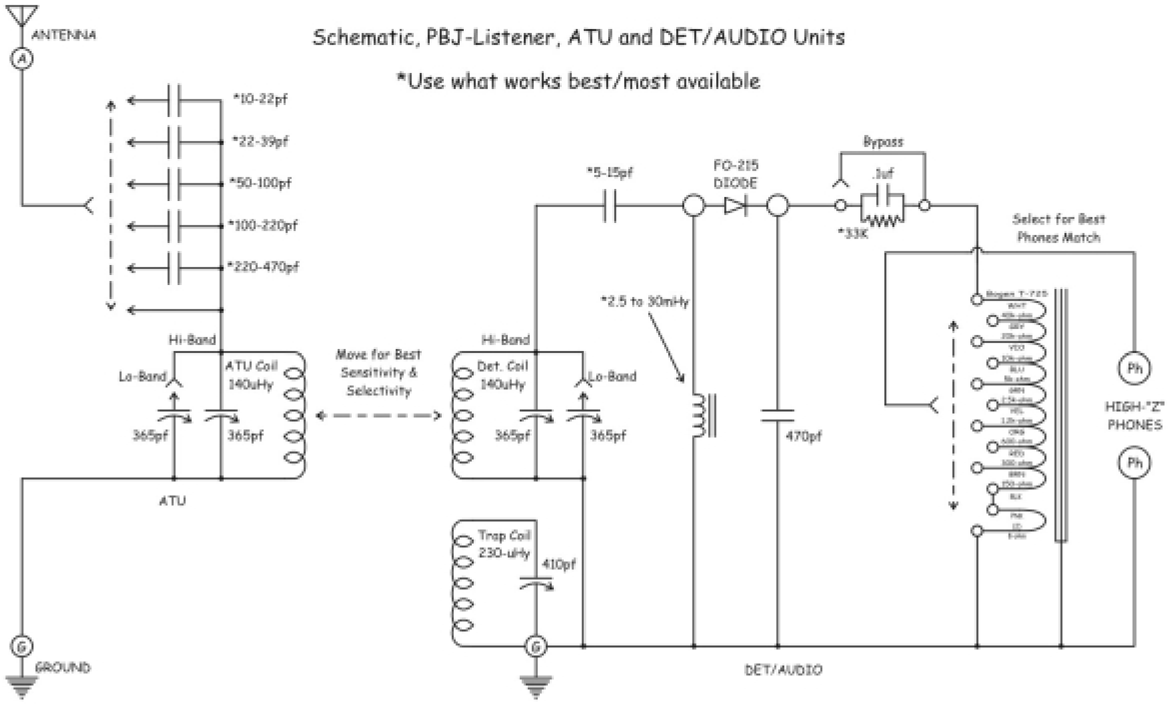
Schematic, PBJ-Listener, ATU and DET/AUDIO Units
*Use what works best/most available
ANTENNA
A
*10-22pf
*22-39pf
*50-100pf
*100-220pf
*220-470pf
Hi-Band
Lo-Band
ATU Coil
140uHy
365pf
365pf
ATU
Move for Best
Sensitivity &
Selectivity
Hi-Band
Det. Coil
140uHy
Lo-Band
365pf
365pf
*5-15pf
FO-215
DIODE
*2.5 to 30mHy
470pf
Bypass
.1uf
*33K
Select for Best
Phones Match
Bogen T-725
Ph
HIGH-"Z"
PHONES
Ph
Trap Coil
230-uHy
410pf
G
GROUND
G
DET/AUDIO

[(img/pbj1h.jpg)](img/pbj1h.jpg)

[Schematic (img/pbj1h.jpg)](img/pbj1h.jpg)

W. I. PENNOCK.
CAPTIVE BALLOON.
APPLICATION FILED AUG. 9, 1910.

1,005,871. Patented Oct. 17, 1911.

Fig. 1.

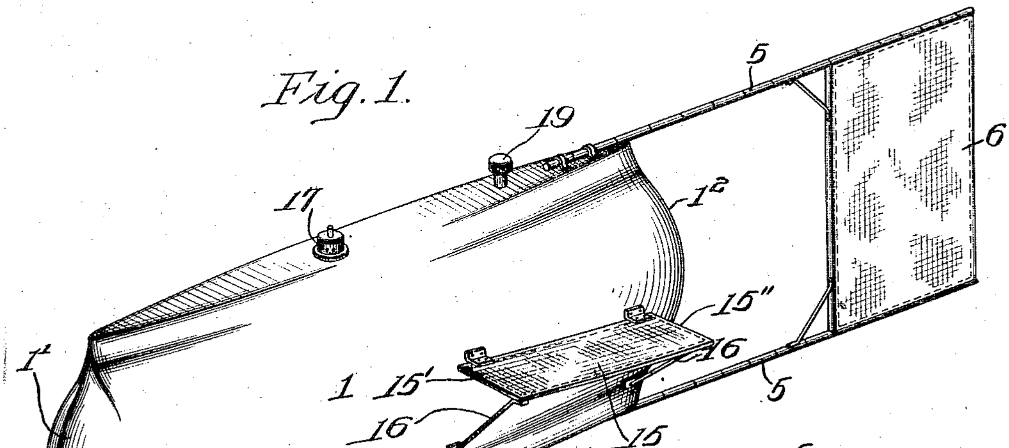
5
19
6
17
1²
15″
1¹
16
1
15′
5
16
15
12′
14′
13′
14
12
11
13
8
18
7

Fig. 2.

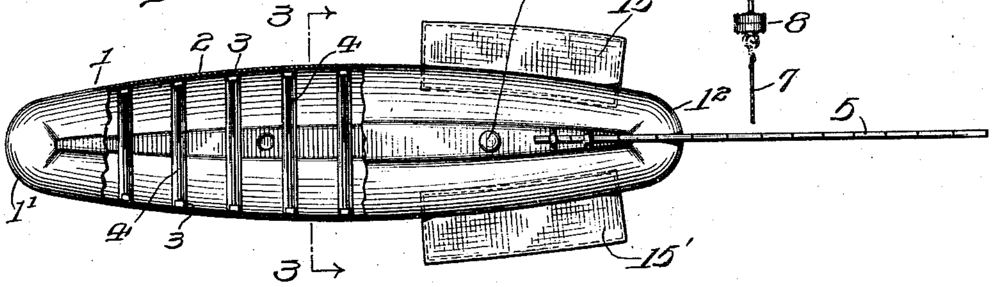
3
19
15
1
2
3
4
1²
5
1¹
4
3
15′
3

Fig. 3.

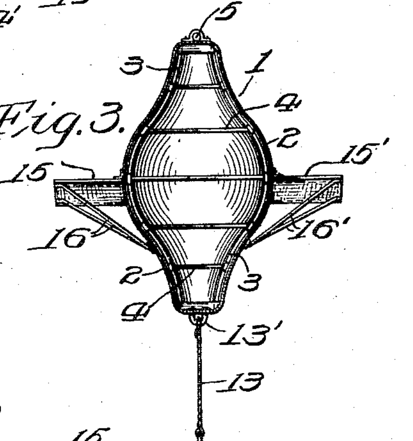
5
3
1
4
2
15
15′
16
16′
2
3
4
13′
13
8
7

Fig. 4.

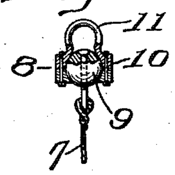

WITNESSES
F. G. Hartman.
Alston B. Moulton

INVENTOR
Walter I. Pennock.

BY

ATTORNEY

UNITED STATES PATENT OFFICE.

WALTER I. PENNOCK, OF PHILADELPHIA, PENNSYLVANIA.

CAPTIVE BALLOON.

1,005,871. Specification of Letters Patent. Patented Oct. 17, 1911.

Application filed August 9, 1910. Serial No. 576,414.

To all whom it may concern:

Be it known that I, Walter I. Pennock, a citizen of the United States, residing at Philadelphia, county of Philadelphia, and State of Pennsylvania, have invented certain new and useful Improvements in Captive Balloons, of which the following is a full, clear, and exact disclosure.

My invention relates to captive balloons, and particularly to such balloons as may be used for elevating collectors of atmospheric electricity from the high altitudes of the earth's atmosphere, and it relates to the construction and arrangement of the parts and shape of the balloon body or envelop.

The objects of my invention are, the providing of the balloon with an envelop or gas container, which is elongated in shape and which will freely turn in the air as it may be acted upon by the wind or air currents; to keep the end of the balloon or envelop which is of least transverse cross sectional area toward the direction from which the wind is blowing; to provide the balloon or gas container with a stationary vane or rudder, by means of which the forward and smaller end of the balloon is kept against the wind; to provide the envelop with side wings or planes, so arranged as to assist in elevating the balloon to high altitudes, and to utilize the force of the wind or air currents in lifting the balloon; to provide a means whereby the elongated envelop or gas container is maintained in a substantially fixed plane, with the forward end elevated a little above the level of the rear end, so that the side wings or planes may have a kite effect and lift the rear end of the gas container or envelop, should it tend to drop; to attach the anchor line or stay to the bottom of the balloon or gas container in such a manner that the forward end of the balloon will always be kept at a little higher level than the rear end, and to further provide means whereby the point of the virtual attachment of the stay or anchor line to the bottom of the balloon may be shifted forwardly or rearwardly whenever the balloon substantially deviates from its normal position; to swivel the balloon to the anchor line or stay so that the balloon will turn as the wind may change without being required to twist the whole length of the anchor line; and to make the envelop or gas container of a suitable, thin, metallic material, so that it may act as a collector of atmospheric electricity and to conduct the electricity which may collect upon the envelop or gas container to the earth, to be used for any useful purpose.

Other objects of my invention will appear in the specification and claim below.

Referring to the drawings forming a part of this specification, in which the same reference numerals are used to designate the same part throughout the various views, Figure 1 is a perspective view of one embodiment of my invention; Fig. 2 a top plan view of the same, the forward part of the envelop being broken away to show the interior construction; Fig. 3 is a vertical transverse cross sectional view on the line 3—3 of Fig. 2; and Fig. 4 is a detail of the swivel joint, by means of which the balloon is attached to the stay or anchor line.

The balloon body or gas container or envelop, 1, is elongated in shape, having its largest dimensions substantially midway between its ends, from which point the body or envelop gradually tapers forwardly and rearwardly to form relatively narrow ends of small transverse area. The forward and rear ends 1′ and 1^2, respectively, are preferably slightly bulged and rounded, so as not to present an absolutely sharp edge to the wind, to give space for the gas to be contained in the envelop, and to prevent the end of the envelop from being flattened or bent inwardly by the action of the wind. I also prefer to make the bottom and top edges of the balloon narrow, so that the sides of the balloon widen out from the top and bottom as they approach the horizontal, longitudinal, diametrical plane of the balloon or gas container, at which point the width of the envelop is greatest.

The body or envelop 1 may be constructed of any suitable material, but I prefer to make the shell 2 of very thin, tough steel, stiffened as may be necessary by substantially vertically arranged angle iron ribs 3, and suitable cross bars 4 reaching from side to side of the envelop and engaging with the ribs 3. At the rear ends of the top and bottom sides of the body or gas envelop 1, I secure light rods 5—5, of bamboo or other suitable light strong material, the said rods extending outwardly from the rear end of the body 1 for such a distance as may be found necessary, and I stretch a light tough fabric between the said rods 5—5 to form a vane or rudder 6.

The end of the stay or anchor line 7 is provided with a suitable swiveled joint 8, which may consist, and preferably does, of a ball and socket construction. In the form shown in Fig. 4 of the drawings, the ball 9 is secured to the stay or anchor line 7, and surrounding and inclosing the ball 9 is a socket 10 provided with a loop or eye 11. In using this swiveled joint 8 in connection with my improved captive balloon, the main cord or line 12 is attached to the loop or eye 11 at its lower end, and to the bottom of the balloon or gas container 1 at the point 12′. This point 12′ is so positioned with respect to the whole balloon, that in a still atmosphere the forward end 1′ of the gas container is at a little higher level than the rear end 1^2. To do this the point 12′ would generally be slightly to the rear of the middle of the gas container; at least it is to the rear of a vertical plane passing through the center of buoyancy of the balloon when the balloon is in a horizontal position. From the eye 11 also run secondary or supplemental cords 13 and 14 attached to the bottom of the gas container as at points 13′ and 14′, the former forwardly of, and the latter to the rear of the point 12′. The length of the cords 13 and 14 is such that when the balloon or gas container is in its normal position, the cords 13 and 14 will be slack. When, therefore, one end of the balloon, as, for instance, the rear end, tends to swing downwardly about the point 12′ as a center, tension will immediately be put upon the cord 13, thus virtually making the point of attachment of the stay or anchor line 7 to the bottom of the balloon or gas container intermediate of the points 12′ and 13′. As these points 13′ and 14′ are on opposite sides of a line passing through the center of gravity or the center of buoyancy of the balloon, the gas in the rear end of the balloon will immediately tend to lift that end about this new point of virtual attachment of the stay or anchor line 7 with the bottom of the balloon or gas container, and the balloon will immediately come to its normal position. The converse is true whenever the forward end of the balloon dips downwardly. Tension is immediately put upon the cord 14; the point of the virtual attachment of the cord 7 immediately shifts to a point somewhere between 12′ and 14′ and the gas in the forward end of the balloon will lift that end until the gas container reassumes its normal position. To further steady the balloon body or gas container 1, and to assist in elevating the same, I provide side or laterally extending wings or planes 15 and 15′, stationary with respect to the envelop or gas container 1, and rigidly secured to the sides thereof, preferably to the ribs 3 through the shell 2. These wings or planes 15 and 15′ are suitably braced by braces 16 and 16′, and are set relatively to the gas container or envelop 1 so that when said envelop is in its normal position, the forward ends of the wings or planes 15 and 15′ will be higher than the rear ends of the same. When the wind is blowing against the forward end of the balloon, as it passes the body it strikes upon the end sides of the wings or planes 15 and 15′, and exerts a lifting effect upon the whole balloon in the same manner as a kite exerts a lifting pull when similarly inclined.

From the above description it will be plain that the whole balloon or gas envelop, with its attached vanes and wings, will be maintained in the air with its forward end slightly higher than its rear end, and with its forward end always turned toward the direction from which the wind is blowing, by reason of the action of the stationary rudder or vane 6 and the swiveled connection 8 between the body or gas envelop 1 and the stay or anchor line 7. It will also be seen that the body will tend to be always maintained in this slightly inclined position, and that should one end or the other end of the gas container or body 1 drop or lift so as to turn the body in a vertical plane around the point 12′ as a center, tension will immediately be put on the cord 13 or 14, depending upon which direction the balloon turns, and the balloon will be righted and brought to its normal position by a shifting of the virtual point of attachment of the cord 7 with the bottom of the balloon.

Any suitable metal may be used in constructing the balloon, although thin steel appears to me to be the best material for the same, inasmuch as it can be made very thin and its tensile strength is very high. The ribs and cross bars 3 and 4, respectively, may be made of thin steel, aluminum or wood, and the rudder 6 and wings 15 and 15′ may be made of thin, tough silk, oiled or otherwise treated so that the same will not absorb moisture.

Inasmuch as the pressure of the gas contained within the body or envelop 1 will expand as the atmosphere in which the balloon is suspended grows rarer, I provide a safety valve 17 of any approved type which may relieve the interior of the balloon of pressure when the pressure within the gas container or envelop 1 becomes so great as to be liable to burst the container or envelop.

In order to fill the envelop or gas container 1 with gas, I prefer to fill the envelop first with water and to then displace the water with hydrogen gas, as, for instance, by allowing the water to escape through the valve 18, the water as it runs out drawing in the hydrogen gas, as through the valve 19 at the top of the envelop.

When the envelop is made of a suitable metal, such as steel or copper, the same may be used as a collector of atmospheric electricity, and to conduct the electricity so collected upon the envelop 1 to the earth, where it may be used as desired, I prefer to either incorporate into the anchor line 7 and the cord or line 12 a copper wire, or to wrap said wire around said anchor line 7 and cord 12, and to secure the upper end of said wire to the metallic covering of the balloon or gas container 1. This wire is indicated diagrammatically by the line 20.

Although I have described but one embodiment of my invention, it is to be understood that other forms and modifications are fully contemplated by me, so long as the structure thereof falls within the scope of the appended claim.

Having thus described my invention, what I claim and desire to protect by Letters Patent of the United States, is:

The combination of an elongated longitudinally rigid balloon, and a stationary rudder secured to the rear end thereof, with an anchor line secured to said balloon wholly to the rear of the center of buoyancy thereof, laterally projecting rearwardly positioned and angularly elevated planes secured to said balloon, and a pair of slack auxiliary lines secured to said anchor line and to said balloon on opposite sides of said anchor line, the whole coöperating to increase the vertical stability of the balloon.

In witness whereof, I have hereunto set my hand this 6th day of August, A. D., 1910.

WALTER I. PENNOCK.

Witnesses:

H. N. Dority,

Laura Evans Lalor.

W. I. PENNOCK.

APPARATUS FOR COLLECTING ELECTRICAL ENERGY.

APPLICATION FILED JAN. 4, 1911.

1,014,719. Patented Jan. 16, 1912.

2 SHEETS—SHEET 1.

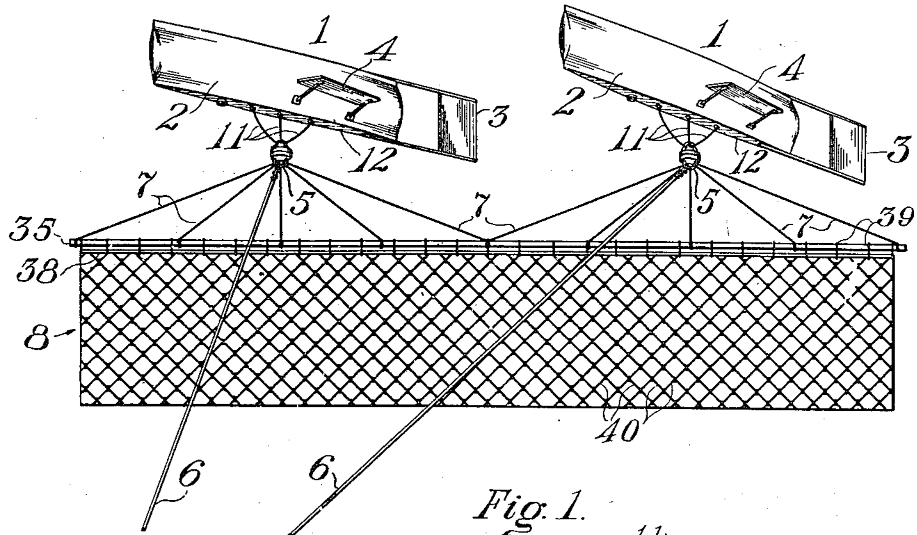
1
4
2
11
12
3
1
4
2
11
12
3
7
5
7
5
7
39
35
38
8
40
6
6

Fig. 1.

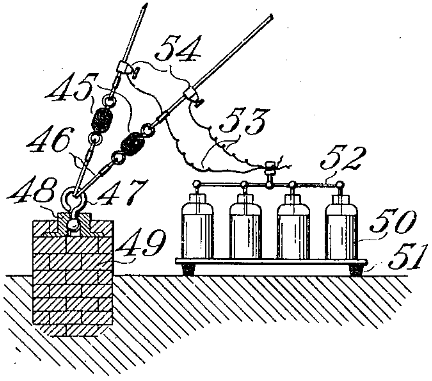
54
45
53
46
52
48
47
49
50
51

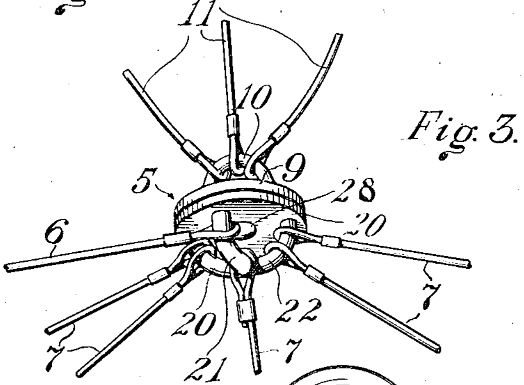
11
10
9
5
28
20
6
7
20
22
21
7
7

Fig. 3.

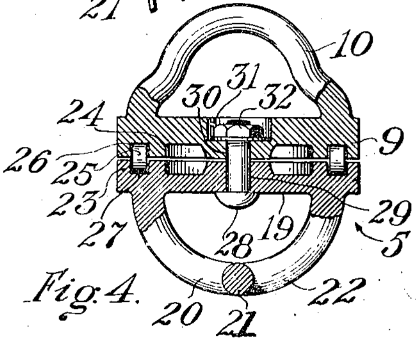
10
31
30
32
24
9
26
25
23
29
27
28
19
5
22
20
21

Fig. 4.

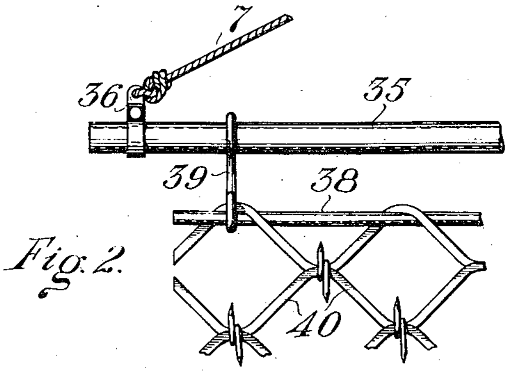
7
36
35
39
38
40

Fig. 2.

INVENTOR
Walter I. Pennock.

WITNESSES
Elizabeth L. Smith
J. Stuart Freeman.

BY

ATTORNEY

W. I. PENNOCK.

APPARATUS FOR COLLECTING ELECTRICAL ENERGY.

APPLICATION FILED JAN. 4, 1911.

1,014,719. Patented Jan. 16, 1912.

2 SHEETS—SHEET 2.

Fig. 5.

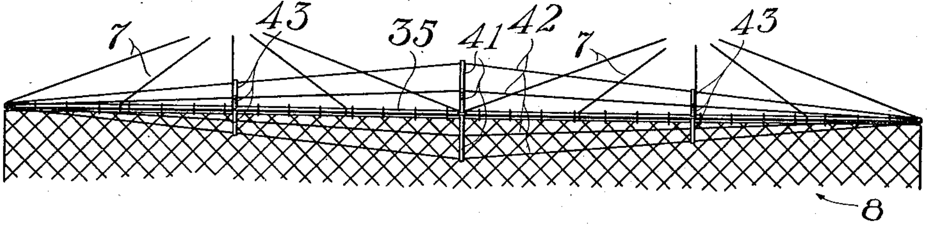
7
43
35
41
42
7
43
8

Fig. 6.

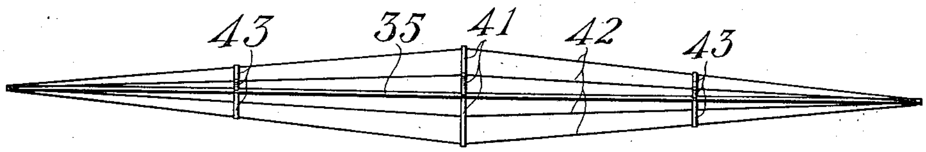
43
35
41
42
43

Fig. 7.

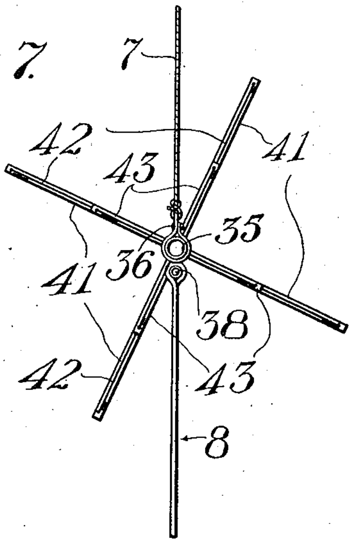
7
41
42
43
35
41
36
38
42
43
8

INVENTOR

Walter I. Pennock.

WITNESSES

Elizabeth L. Smith,

J. Stuart Freeman.

BY

ATTORNEY

UNITED STATES PATENT OFFICE.

WALTER I. PENNOCK, OF PHILADELPHIA, PENNSYLVANIA.

APPARATUS FOR COLLECTING ELECTRICAL ENERGY.

1,014,719. Specification of Letters Patent. **Patented Jan. 16, 1912.**

Application filed January 4, 1911. Serial No. 600,777.

To all whom it may concern:

Be it known that I, WALTER I. PENNOCK, a citizen of the United States, residing at Philadelphia, county of Philadelphia, and State of Pennsylvania, have invented certain new and useful Improvements in Apparatus for Collecting Electrical Energy, of which the following is a full, clear, and exact disclosure.

The present invention relates to an improved means for collecting the charges of electricity from the upper atmosphere and more particularly to that form which consists in one or more captive balloons from which is suspended a suitable form of metallic conductor.

The principal objects of the device are: to provide a collector for atmospheric and static electricity, which when in operative position will present a large surface to currents of air, but which will offer comparatively little resistance thereto, to provide a collector of such material and construction as will be more efficient in its operation than any previously constructed for a similar purpose, to provide a means for maintaining such a collector suspended in the air and at right angles to opposing currents thereof, and to provide a suitable anchorage for holding said means captive.

With these principal objects in view, the present invention consists in further advantages which are brought out in the following specification and accompanying drawings, in both of which like numerals refer to like parts, and in which drawings—

Figure 1 is a perspective view of the complete device in operative position, Fig. 2 is an enlarged detail view of the wire mesh and the manner in which it is attached to the supporting balloons, Fig. 3 is a detail of the manner of securing the collector-supporting and anchor cables to the balloons, Fig. 4 is an enlarged cross section of the swivel connection shown in Fig. 3, Fig. 5 is top plan view of the reinforcing braces on the screen, Fig. 6 is an elevation of the same and Fig. 7 is an end view of the structure shown in Figs. 5 and 6.

Referring to the drawings, in Fig. 1 thereof, a plurality of balloons 1 of any suitable type is shown, each of which embodies hollow metallic elongated gas tanks 2, extending from the rear of which are single, rigidly affixed rudders 3, while on the sides of the tanks are secured stationary lifting planes 4.

To the bottom and slightly to the rear of the center of the tanks 2 is secured a suitable swivel 5, by which the anchor ropes 6 and the suspension ropes 7 for the metallic conductor 8 are secured to the balloons 1. A suitable form of swivel joint is illustrated in Figs. 3 and 4, but any type can be used that embodies the essential features shown therein.

The swivel joint illustrated consists in the base plate 9 having a looped portion 10 integral therewith and projecting from the upper face thereof. Secured to the loop 10 is a set of three light electrically conductive supporting ropes or cables 11 which extend upwardly and are secured at intervals to the bottom 12 of the balloon above.

The lower or revoluble member 19 of the swivel joint preferably comprises three upwardly directed curved arms, 20, 21, and 22, respectively, forming at their junction a T-shape as shown, said arms at their upper extremities being integral with the plate 23. The member 19 is revoluble below and concentric with the plate 9, and the two members are lightened in weight by opposed concentric grooves as shown at 24. Contact between said members is made through the roller bearings 25, which are carried in the opposed concentric grooves 26 and 27 of the respective upper and lower plates. Furthermore, said plates are maintained in coöperative relation with each other by means of a bolt 28 passing through centrally drilled holes 29 and 30 in the respective lower and upper plates, the drilled hole 30 opening upwardly into an enlarged recess 31, in which is sunken the nut 32 on the bolt 28. Two of the arms 20 and 22 of the lower revoluble member 19 extend in diametrical alinement, while from the central point thereof extends the third arm 21 at right angles thereto, and upward to the plate 23.

Suspended below the plurality of balloons is a hollow rod 35, of any suitable material, connected at regular intervals such as at points 36, by metallic ropes 7, to the alined arms 20 and 22 of the swivel joint 5. Below and parallel to the rod 35 is a similar but smaller rod 38 suspended therefrom by means of suitable couplings 39. From the rod 38 hangs a wire mesh 40 of any suitable form, such as can be extended over a considerable area and which will offer comparatively little resistance to the passage of air therethrough. The metal in such a screen is preferably rough, sharp or jagged and a convenient form embodying these characteristics is expanded sheet metal such as is used for lathing and for reinforcing concrete construction. Instead of the single hollow rod 35 alone, applicant contemplates the use of reinforcing means to prevent the buckling of the rod in a stiff wind, such, for instance, as the arrangement of wire bracing shown in Figs. 5, 6 and 7 in which cross rods 41, secured to the rod 35 at the center thereof are attached to the ends thereof by means of stretched wires 42. Between the rods 41 and the ends of the rod 35 are secured sets of cross rods 43 which support the wires 42 and strengthen the rod 35 at as many points as they may be placed.

For the purpose of maintaining the balloons and apparatus suspended therefrom captive, the light metallic cables 6 terminating downwardly in insulators 45 are employed. These insulators in turn are connected by means of short sections of rope 46 to the eye 47 of a suitable swivel 48, embedded in the anchorage 49.

When the balloons with the metallic screen suspended therefrom are allowed to rise into one of the higher altitudes, the entire apparatus being of metallic construction and uninsulated will become energized by contact with the surrounding natural charges of electricity. From the above description it is evident then that, while the screen 8, on account of its great extent, will be the greatest collecting agent, it will be seen that the balloons themselves and the suspending wires will also coöperate as one large collector, since no parts of which are insulated from any of the neighboring parts thereof. Consequently, when the apparatus described has reached an altitude or strata of the atmosphere abounding in static charges of electricity, an amount of the said charges proportionate to the surface area of the metal exposed will collect upon the apparatus as a whole and will be conducted downwardly toward the earth by means of the various anchor ropes 6, but will not pass into the ground on account of the interposition of the insulators 45.

To use the electrical charges thus acquired, a plurality of Leyden jars 50, or other suitable collectors are supported above the surface of the earth and insulated therefrom by any suitable means as represented by the blocks 51. Either the inner or outer conducting surfaces of the jars may be connected together and energized by the accumulated charge. In the present instance, the inner surfaces of the accumulators are shown to be connected, and the connecting means 52 is in turn connected to the ropes 6 by means of wires 53. These wires are secured to the ropes mentioned by means of suitable binding posts 54.

In the device described the anchor ropes 6 are of substantially the same length, and when the apparatus is raised to the desired altitude and is being blown by the currents of air, the balloons are turned by means of the vanes 3 to parallel relation with each other, and furthermore, from the manner in which the device is held captive and the collecting net 40 is suspended from the balloons, it is obvious that said net will at all times readily swing into a plane substantially perpendicular to any current of air acting upon the balloons above. The purpose of the swivel joint shown in Figs. 3 and 4 is principally for allowing the balloons to readily aline themselves with any new direction of the wind before the apparatus, including the suspended screen, can swing about the swivel 48, and said joint will also prevent the twisting of the ropes 7, when any rapid shifting of the air currents may occur.

While applicant has shown a set of Leyden jars as the accumulators in the accompanying drawings, it is obvious that any other suitable form may be used, and furthermore, that although not illustrated, any suitable apparatus may be run thereby, such as for instance, wireless telegraphic instruments.

Furthermore, although but one embodiment of the invention has been described, it is to be understood that various modifications may be made therein, and in fact several are contemplated by applicant that are of such structure as fall well within the scope of the appended claim.

Having thus described my invention, what I claim and desire to protect by Letters Patent of the United States, is:

A collector for charges of electricity, comprising a plurality of supporting means, a metallic gauze sustained thereby to lie in a plane and substantially equidistant from each of said supporting means, anchoring means emanating from a common point to each of said supporting means, means operative to maintain said supporting means in their normal positions, and means between each of said supporting means and said gauze to permit each of said supporting means to readily and independently aline itself to accord with any alteration in the direction of opposing air currents.

In witness whereof I have hereunto set my hand this 28 day of December, A. D. 1910.

WALTER I. PENNOCK.

Witnesses:
MILDRED S. TEMPLE,
E. EUGENIA PENNOCK.

US 20070007393A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2007/0007393 A1**

Pinto (43) **Pub. Date: Jan. 11, 2007**

(54) **METHOD FOR PROPULSION**

(75) Inventor: **Fabrizio Pinto**, Monrovia, CA (US)

Correspondence Address:
DEMONT & BREYER, LLC
100 COMMONS WAY
HOLMDEL, NJ 07733 (US)

(73) Assignee: **InterStellar Technologies Corporation**, Monrovia, CA

(21) Appl. No.: **11/197,697**

(22) Filed: **Aug. 4, 2005**

Related U.S. Application Data

(60) Provisional application No. 60/598,658, filed on Aug. 4, 2004.

Publication Classification

(51) **Int. Cl.**
B64G 1/40 (2006.01)

(52) **U.S. Cl.** **244/171.5**

(57) **ABSTRACT**

A method for propulsion that does not consume fuel. The method involves the modification of the dispersion force (i.e., van der Waals) that arises between particles, such as neutral atoms. The method comprises generating a lifting force by subjecting a plurality of confined particles to a trigger acceleration, exposing the particles to an amount of electromagnetic radiation that is sufficient to induce the lifting force to (1) exhibit relatively long-range interactions and (2) increase the momentum of the particles, and then transferring at least a portion of the increase in momentum to a vehicle.

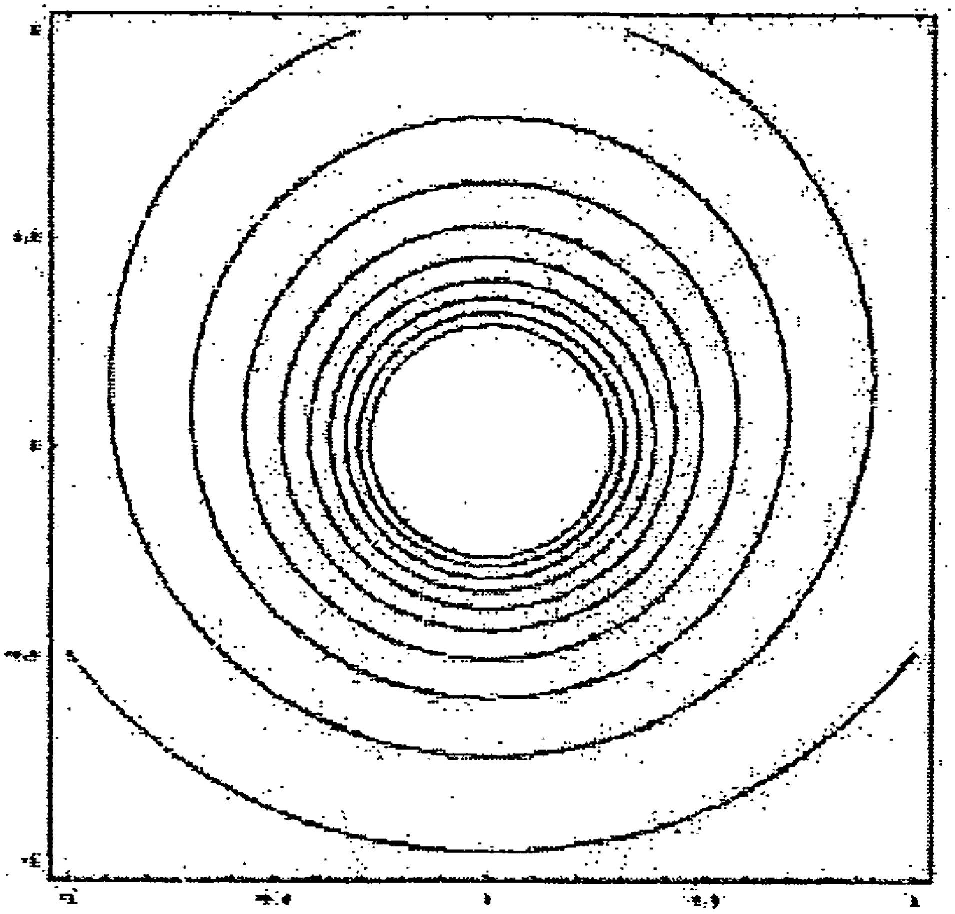

Figure 1

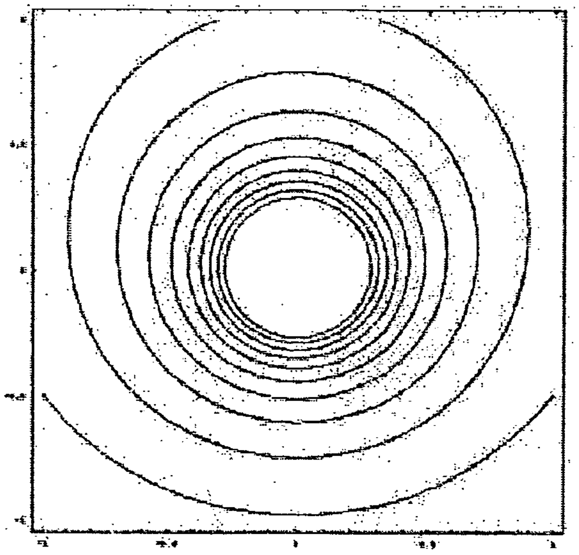

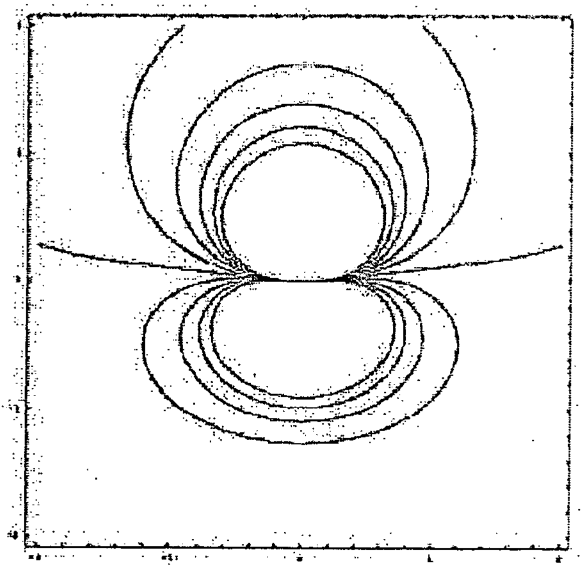

Figure 2

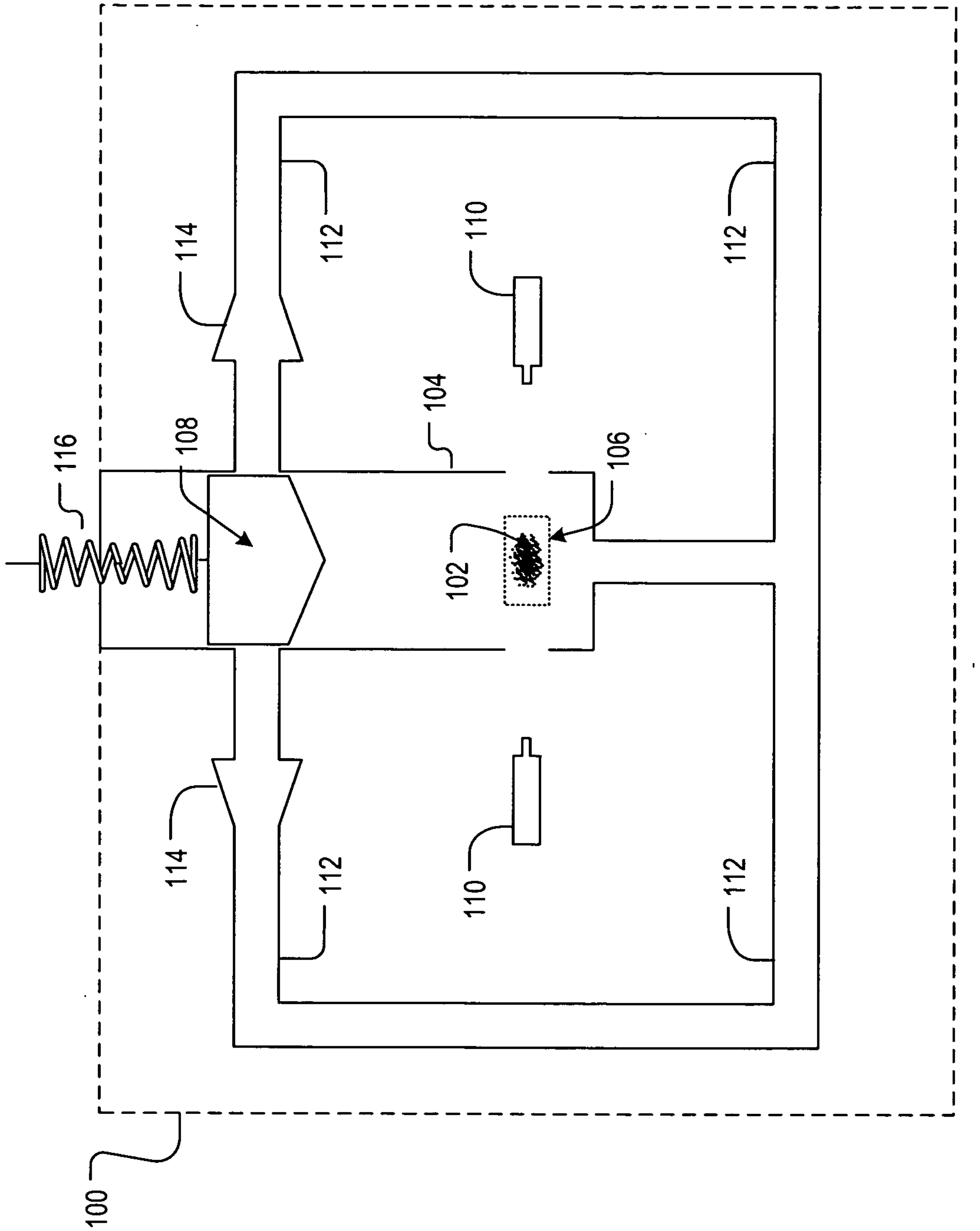
116
108
114
112
110
112
104
106
102
114
112
110
112
100

Figure 3

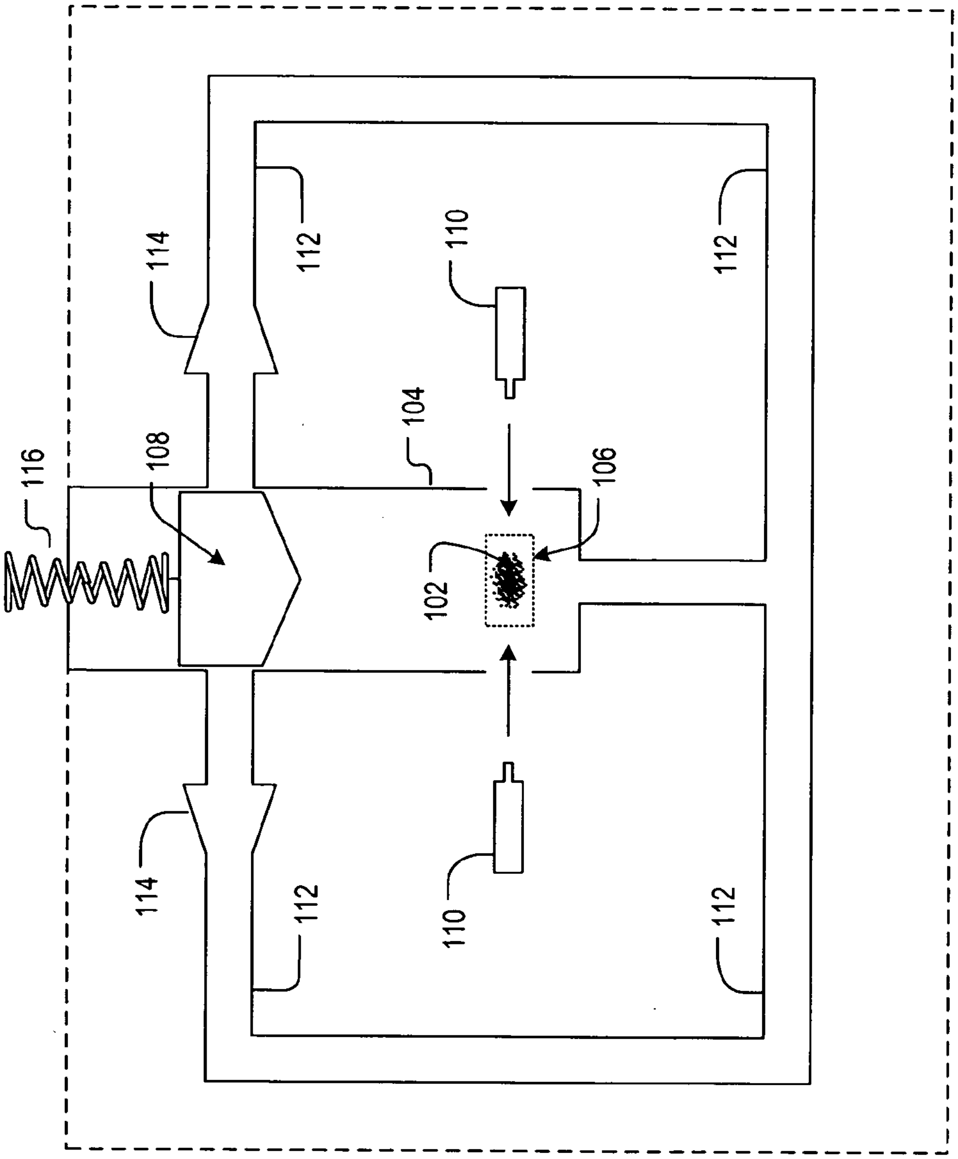
116
108
114
112
110
104
106
102
114
112
110
112
112

Figure 4

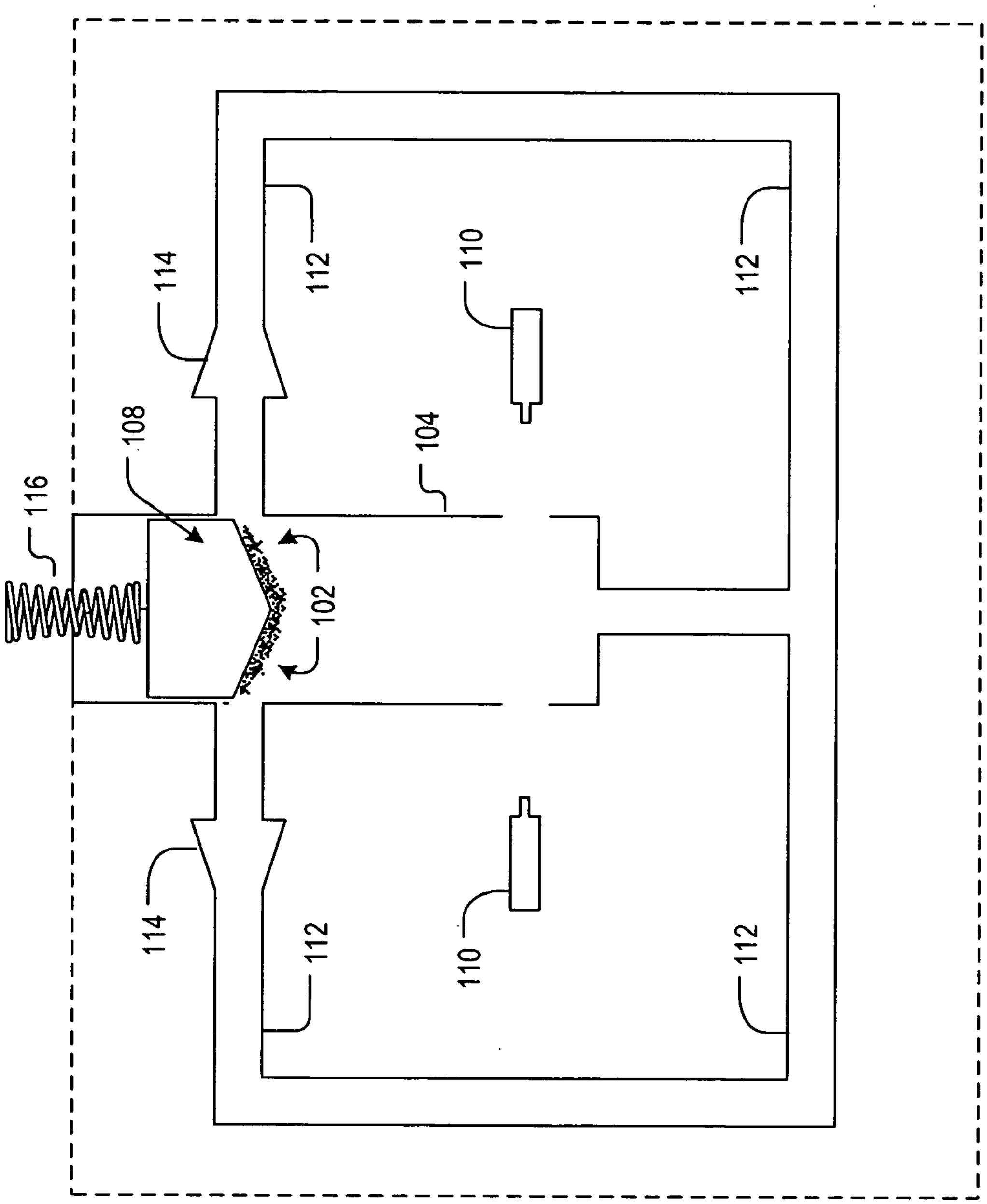
116
108
104
102
114
112
110
112
114
112
110
112

Figure 5

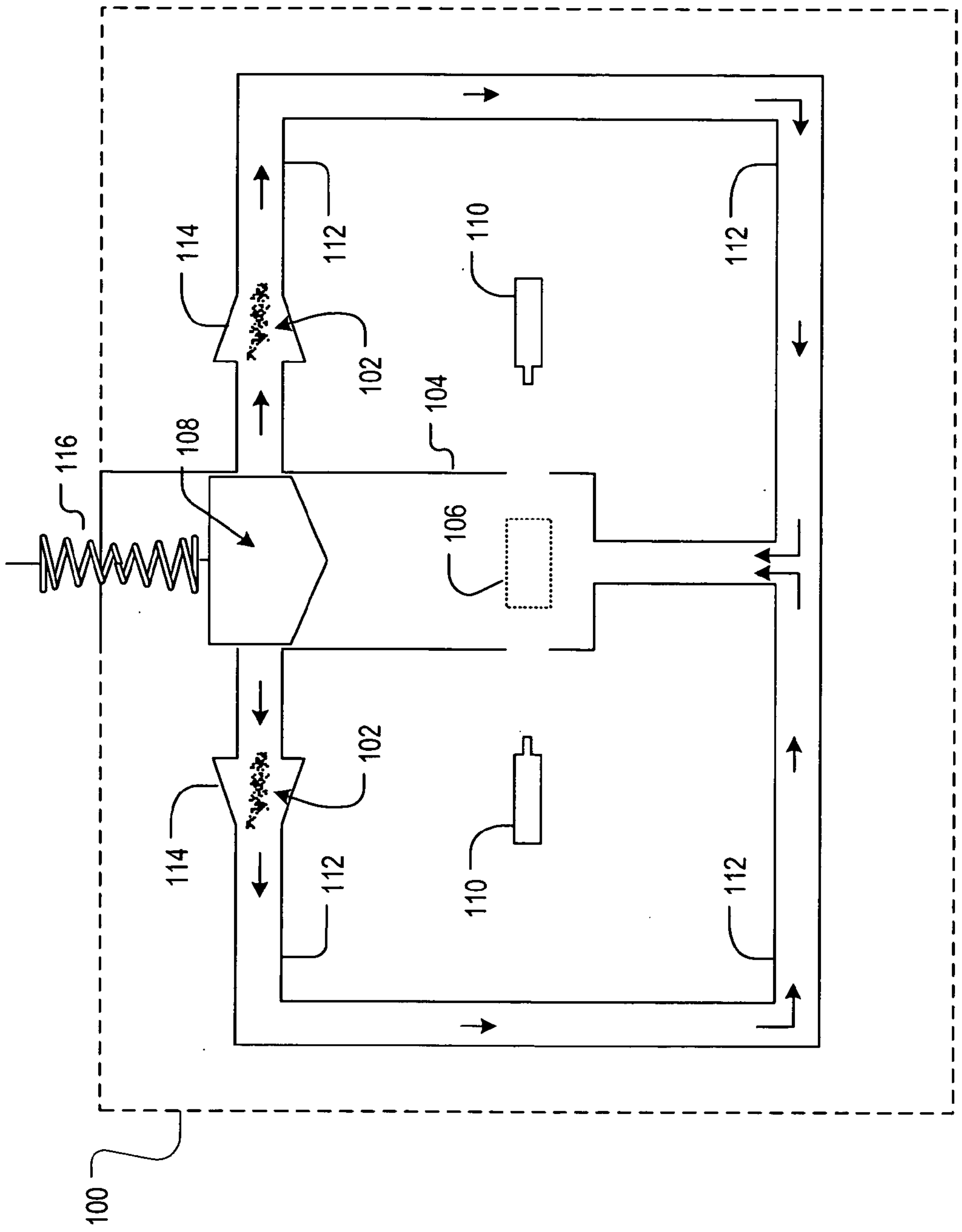
116
108
114
102
112
110
104
106
112
102
114
112
110
112
100

Figure 6

Figure 7

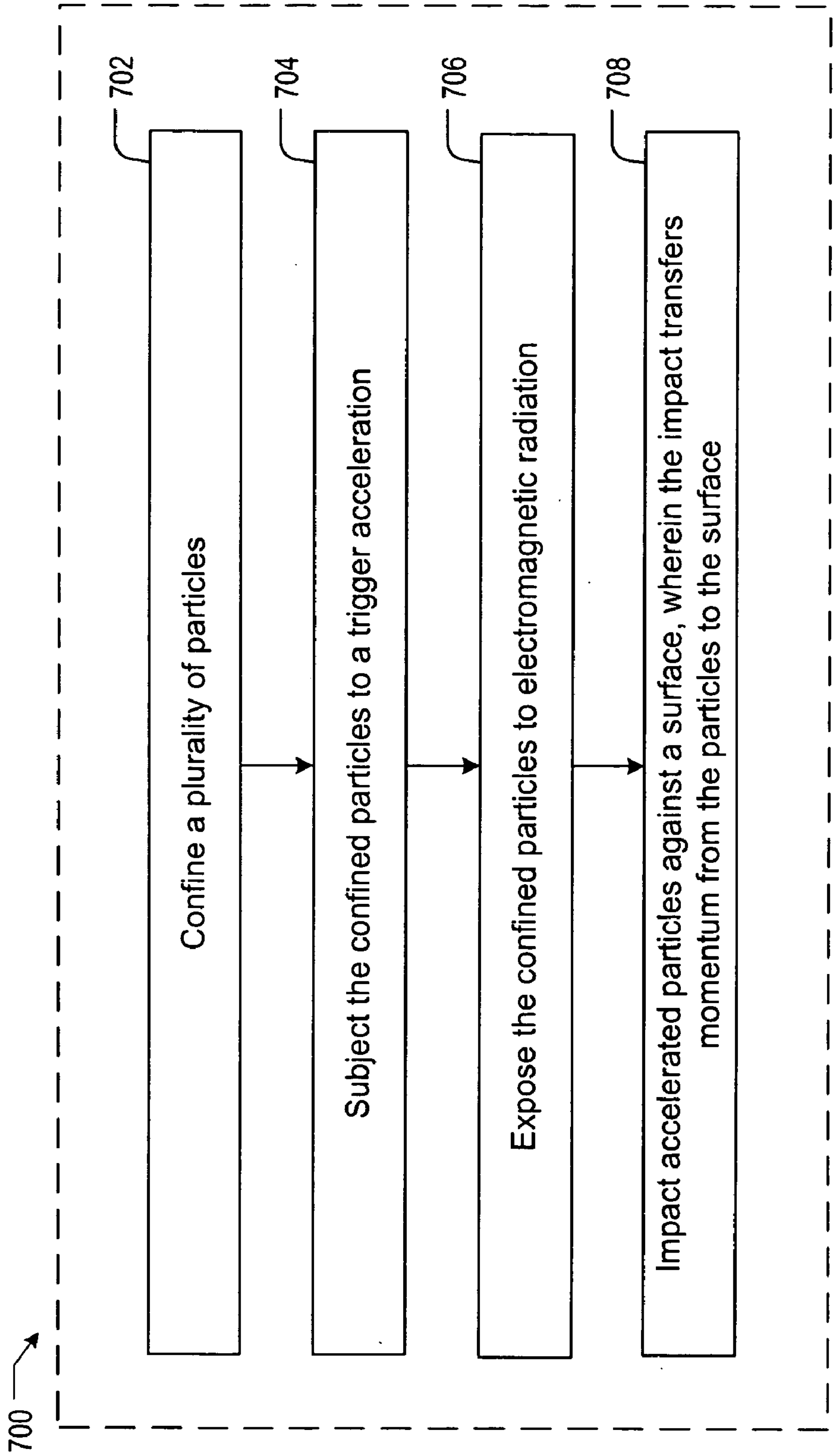
700
702
Confine a plurality of particles
704
Subject the confined particles to a trigger acceleration
706
Expose the confined particles to electromagnetic radiation
708
Impact accelerated particles against a surface, wherein the impact transfers momentum from the particles to the surface

Figure 8

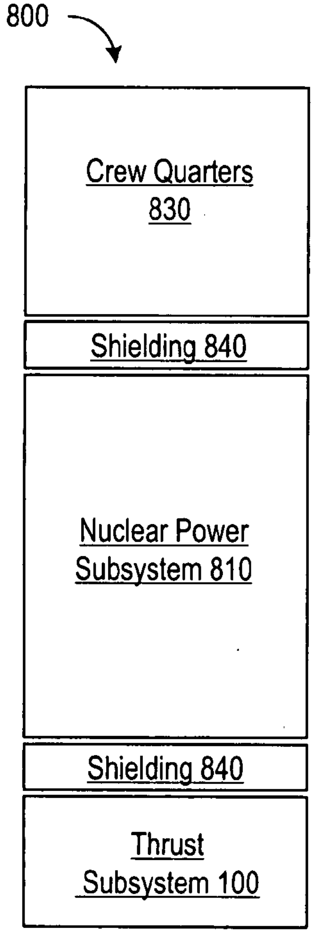
800
Crew Quarters 830
Shielding 840
Nuclear Power Subsystem 810
Shielding 840
Thrust Subsystem 100

Figure 9

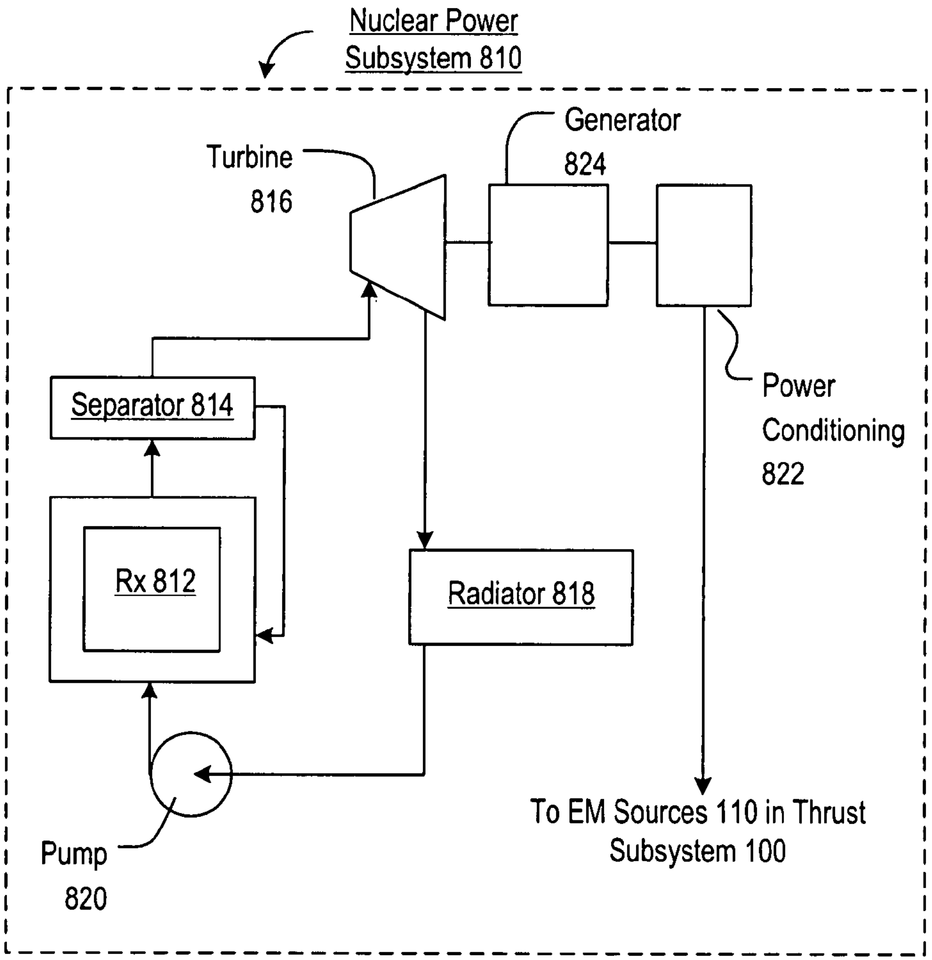
Nuclear Power Subsystem 810
Turbine 816
Generator 824
Power Conditioning 822
Separator 814
Rx 812
Radiator 818
Pump 820
To EM Sources 110 in Thrust Subsystem 100

METHOD FOR PROPULSION

STATEMENT OF RELATED APPLICATION

[0001] This application claims priority of U.S. Provisional Patent Application 60/598,658, which was filed on Aug. 4, 2004 and is incorporated by reference herein.

FIELD OF THE INVENTION

[0002] The present invention relates generally to propulsion systems.

BACKGROUND OF THE INVENTION

[0003] Sixty-six years after the Wright brothers made their first, sustained powered flight, Neil Armstrong walked on the Moon. Incredible progress to be sure, but can this pace of innovation be sustained? Will we soon visit neighboring planets or the nearest stars? Can we reach these destinations with the technology that got us to the Moon? If we can't, what propulsion technologies might be able to take us to these unthinkably remote places?

[0004] Current propulsion technology is based on an action-reaction principle, whereby a gas is expelled at high-speed to propel a payload in the opposite direction. This technology is typically embodied as a chemical rocket engine. While a payload can be rapidly accelerated using a chemical rocket, fuel is quickly consumed to develop the required thrust. To illustrate the problem, consider that if a spacecraft could be powered to achieve a constant acceleration of only 1 g, the trip from Earth to Mars would require about 2-4 days. In fact, modern chemical rocket engines can achieve accelerations much greater than 1 g. But even at 1 g, the fuel would be exhausted within minutes. As a consequence, the trip to Mars from Earth via chemical rocket takes about six months.

[0005] With current chemical-rocket technology, most of the weight at launch is fuel. For example, a typical choice for a mission to Mars would involve the Boeing Delta II 7925 or 7925H rocket stages. In its common configuration, the RS-27A engine of the Delta II first stage, along with an additional nine strap-on sold rocket motors, will have a mass of about 285,000 kilograms at launch. But of this mass, only slightly more than 1000 kilograms will reach Mars.

[0006] As noted above, the delivery of several tons of a payload to Mars via chemical rockets is contemplated to take about six months, with total mission duration of about two to three years. For the majority of transit time, astronauts will be weightless, which is known to adversely affect the human body. Furthermore, the astronauts will be subject to exposure from harmful radiation. Additionally, the prospects of mounting a rescue or recovering from a serious malfunction are slim due to the transit times involved.

[0007] And the distances involved in interstellar travel are so large that with this technology, a trip to even the nearest star systems would take hundreds to thousands of years.

[0008] It seems clear that current technology does not provide a means to manned exploration of the Solar System or beyond. That being the case, can technological approaches be conceived that will send spacecraft from the Earth to destinations within the Solar System in a matter of days or weeks, as opposed to years or decades? Any such approach will face a daunting technological requirement. Namely, in order to drastically reduce travel time to "neighboring" planets and "nearby" stars, exceedingly large velocities must be achieved—velocities that are on the order of a significant fraction of the speed of light.

[0009] Proposals that meet this mission time requirement will, therefore, typically require what can only be described as "fantastic" technologies. From a feasibility perspective, perhaps the most "promising" of those technologies that have been proposed is the matter-antimatter drive. When combined, matter and antimatter will completely annihilate, releasing unfathomable quantities of energy. But even if we were able to develop a matter-antimatter drive, its use should be proscribed. The reason is that if antimatter were to leak from its containment chamber while in the vicinity of Earth, there is a distinct possibility that the resulting energy release would destroy Earth or at least cause the extinction of all life thereon.

[0010] Another exotic propulsion technology is the "solar sail." Although solar sails can produce momentum by reflecting a portion of the light that they receive from the sun, this approach, on its own, does not offer a solution to the problem of achieving interstellar or even interplanetary travel. More specifically, in order to deliver a space probe to a nearby star in less than a century, the sail must be driven by laser light aimed at it throughout the trip. The power requirement for the laser, which would be located on Earth, is on the order of hundreds of thousands of terawatts. For the sake of comparison, the current planetwide consumption of electricity is on the order of about 1 terawatt. And this approach has a further complication. Namely, the craft must be slowed from a non-trivial fraction of the speed of light to orbital velocity at its final destination using light that is coming from earth. This would require the coordination of very complex maneuvers that, if not carried out correctly, might result in the destruction of the ecosystem of the destination planet.

[0011] In the late 1990s, NASA established and funded a program, now defunct, called the "Breakthrough Propulsion Program." The program's charter was to evaluate entirely new propulsive principles that would enable interstellar or at least interplanetary travel. Technologies under consideration included the Schlicher thruster, Deep Dirac Energy, Podkletnov gravity shielding, Podkletnov force-beam, transient inertia, coupling between electromagnetism and spacetime, gravity modification schemes, anomalous heat effect, Biefeld-Brown effect, warp drives, wormholes, high-frequency gravitational waves, superluminal tunneling, the Slepian Drive, and the quantum vacuum (e.g., dynamical Casimir effect, etc.).

[0012] Unfortunately, none of these approaches were deemed to be promising. For example, one study pertaining to the quantum vacuum concluded that the acceleration of a spacecraft propelled by the dynamical Casimir effect would, after ten years under acceleration, be traveling at 0.1 meters per second!

[0013] In light of the foregoing, it seems likely that an as yet unidentified propulsion technology will be required to make routine, manned interplanetary and interstellar travel a reality.

SUMMARY

[0014] The illustrative embodiment of the present invention is a system and method for propulsion that avoids some of the drawbacks of the prior art. Unlike conventional propulsion technology, the propulsion system described in this specification does not consume fuel (although there is an energy requirement). In fact, the system does not even use fuel, as the term is commonly used.

[0015] The illustrative embodiment is grounded in accepted physics principles, albeit leading-edge theoretical, experimental and applied physics. The propulsion method does not violate basic physical "laws," such as the conversation of momentum. The equations on which the propulsion system is based are clearly established in the art, although extended to a domain of applicability and mode of use that has not been previously contemplated.

[0016] The propulsion system operates by modifying the dispersion force (i.e., van der Waals) that arises between particles, such as neutral atoms. The following two discoveries by the inventor enable the propulsion system:

[0017] (1) The dispersion force interaction between any two neutral atoms is affected by the presence of an external gravitational field in a way that results in a repulsive force upon the atomic pair.

[0018] (2) The distortion of the dispersion interaction described above, which is usually quite small and is proportional to the number of atoms present, can be magnified by many orders of magnitude. This requires that (a) the atom-atom interaction to be transformed from a relatively shorter-range interaction to a relatively longer-range interaction; and (b) a very large number of atoms are present for mutual interaction.

[0019] A method in accordance with the illustrative embodiment of the present invention comprises:

[0020] generating a lifting force by subjecting a plurality of confined particles to a trigger acceleration; and

[0021] exposing the particles to an amount of electromagnetic radiation that is sufficient to induce the lifting force to:

[0022] (i) exhibit relatively long-range interactions; and

[0023] (ii) increase the momentum of the particles; and

[0024] transferring at least a portion of the increase in momentum to a vehicle.

[0025] To begin the propulsion cycle, the particles must be subjected to acceleration, that is, the "trigger" acceleration. This can be accomplished, for example, by supporting a craft that contains the propulsion system in a gravitational field (i.e., the craft cannot be in free fall). Or, acceleration can be kinematic, such as by rotating the craft, or using a conventional propulsion system to accelerate the craft. The force is referred to as a "lifting" force because it's direction is opposite to the weight of the particles.

[0026] In the illustrative embodiment, the particles are neutral atoms, the atoms are confined in an atomic trap, and a laser provides the electromagnetic energy that is required to transform the atom-atom interaction from a short-range to a long-range interaction and to increase the momentum of the atoms.

[0027] Some of the increase in momentum of the atoms is transferred to a vehicle, such as a spacecraft. This transfer of momentum propels the vehicle. In the illustrative embodiment, this is accomplished by letting the atoms work against a piston, which in turn impacts against a part of the vehicle. Atoms that hit the piston are recycled to the atomic trap for the next propulsion cycle.

[0028] Several points of explanation or definition will be useful in understanding the illustrative embodiment of the present invention and its underlying principles.

[0029] The term "long range" interaction or force usually describes a force that decays with distance as I/R^n, where n is a positive number. The term "short range" interaction or force usually describes a force that decays with distance as $\exp[-R]$. Those conventions are not followed in this disclosure. Rather, for the purpose of this disclosure and the appended claims, terms that "short range" and "long range" are comparative or relative terms. For example, the language "inducing the lifting force to exhibit relatively long range interactions" means that the lifting force is induced to exhibit a relatively longer-range interaction than is normally the case.

[0030] Since the momentum that is donated by the particles propels the vehicle, there might be a tendency to characterize the particles as "fuel." But the particles are not "fuel" in any conventional sense of that term. They simply serve as a "momentum exchanging element."

[0031] It is important to recognize that the particles are not accelerated by the wall of the spacecraft but, rather, by their mutual dipole fields (van der Waals force) as distorted by the craft's acceleration (gravitational or otherwise). In other words, this method does not violate the action-reaction law, since the motion of the particles is not due to action of vehicle upon them.

[0032] An example of "action of the vehicle" on the particles is if the atoms were accelerated due to an explosion in the trap. In that case, the net of all internal forces on the system would be zero and, at the end of the process, the craft would not gain any momentum. In accordance with the illustrative method, however, the atoms accelerate towards the piston, etc., independently of the vehicle and do transfer a net amount of momentum to it during impact.

[0033] The lifting force that is "generated" by the method is not a new force. Rather, it is simply the vertical component of a known intermolecular force; in particular, the van der Waals force. This vertical component arises from an asymmetry in the van der Waals force that results from the introduction of a gravitational field (or acceleration). There would be no such asymmetry, nor vertical component of force, in flat space-time.

[0034] This result—that the interaction potential between two neutral atoms in their ground state depends upon the position of the atoms in a gravitational field—is new. Previous studies pertained to the distortion of the field of two point charges, not two dipoles (atoms). While this "new" force is measurable with presently existing technology, harnessing it, such as to lift an object, is not feasible, since this force amounts to an exceedingly small correction to the total weight of each atom.

[0035] It is useful to note that using many such atom pairs does not improve this situation, because that does not result in a larger force per atom. The reason for this is that the inter-atomic energy is a function, to a large power, of the reciprocal of distance. In other words, the atom-atom dispersion-force interaction is a realtively short-range force. As a consequence, the total "lifting" force on a large number of atoms is increased only minutely with respect to the lifting force acting on just one pair of atoms. That is, if the number of atoms is N, there will be about N pairs (for N>>1) to consider, but the mass of the system also goes up as N. So, there is no gain realized by adding atoms.

[0036] Critical to the present invention is the inventor's recognition that if this newly discovered "effective force" could be transformed from a relatively short-range interaction into a relatively long-range interaction, the lifting force that is available would be greatly increased. In particular, if particles could be made to interact over the long range, then the total energy of the system results from the interaction of every particle with all other particles present. For large groups of particles (N>>1), the interaction grows as N^2, while the total mass of the system is only growing proportionally to N. As a consequence, a large gain in energy can be realized by using large groups of particles.

[0037] A mechanism for transforming relatively short-range interactions into relatively long-range interactions was theoretically discovered several years ago and has been re-evaluated more recently as a way to introduce unusual behaviors in a cloud of trapped atoms. See, Kurizki et al., "New Regimes in Cold Gases Via Laser-Induced Long-Range Interactions," . . . The method involves isotropic illumination of atoms by lasers. That technique, with several modifications, is utilized in conjunction with the illustrative embodiment.

[0038] As previously noted, the present propulsion system and method overcomes a key drawback of chemical engines; namely, the fact that at some point, the fuel is expended. In accordance with the illustrative embodiment of the present invention, it is possible to maintain acceleration without expelling high-speed gases. In other words, the propulsion system does not require fuel. Alternatively, if the "particles" are considered to be "fuel," then there is no consumption of fuel due to the process.

[0039] As previously mentioned, the illustrative propulsion systems and methods described herein are not energy free. In particular, to achieve the required transformations, an intense radiation field, such as can be generated by powerful lasers, must be developed throughout the region in which the particles are trapped. In the case of a craft destined for extremely long interplanetary or interstellar flights, the energy required to power the lasers is obtained, for example, from an on-board nuclear reactor, akin to the reactors powering some submarines.

[0040] The propulsion system described herein has many applications. In particular, in addition to its use as a propulsion system for spacecraft, it can be used to deliver a payload into low earth orbit without requiring orbital speeds. Furthermore, a small version of the propulsion system could be attached to literally any item (e.g., a pallet of goods, a railroad car, etc.) so that the item could be readily moved (e.g., in a warehouse, loaded onto a cargo ship, etc.) as needed. The propulsion system can, of course, also be used in conventional aircraft.

[0041] Additionally, the present propulsion system can be used to supplement a main, conventional propulsion system. In fact, this would facilitate phase-in to replace conventional technologies. For example, a propulsion system in accordance with the illustrative embodiment that is not sufficiently powered to bring a craft to a hover could be used to effectively reduce the mass of the craft, thereby improving the fuel consumption of the main propulsion system. Alternatively, it could be used as a supplemental system for emergencies.

BRIEF DESCRIPTION OF THE DRAWINGS

[0042] FIG. **1** depicts distortion in the spherical symmetry of the field of a simple charge, as caused by the presence of a gravitational field.

[0043] FIG. **2** depicts distortion in the cylindrical symmetry of the field of a classical dipole, as caused by the presence of a gravitational field.

[0044] FIG. **3** depicts a propulsion system in accordance with the illustrative embodiment of the present invention.

[0045] FIG. **4** depicts the propulsion system of FIG. **3**, wherein lasers are illuminating confined particles.

[0046] FIG. **5** depicts the propulsion system of FIG. **3**, wherein particles impact against an elastically-bound piston.

[0047] FIG. **6** depicts the propulsion system of FIG. **3**, wherein particles are pumped back to a reservoir for use in a subsequent propulsion cycle.

[0048] FIG. **7** depicts a method in accordance for propulsion in accordance with the illustrative embodiment of the present invention.

[0049] FIG. **8** depicts a schematic of a vehicle that incorporates the propulsion system of FIG. **3**.

[0050] FIG. **9** depicts a schematic of the nuclear power subsystem of the vehicle of FIG. **8**.

DETAILED DESCRIPTION

[0051] This Detailed Description proceeds with Section 1.1, which provides a description of propulsion system **100** and a method for propulsion in accordance with the illustrative embodiment of the present invention. Section 1.2 discloses a vehicle that incorporates propulsion system **100**. The remaining sections, which include Sections 2.1-2.4 and 4 provide a theoretical development for propulsion system **100** and performance estimates.

1.1 Propulsion System 100

[0052] FIG. **3** depicts propulsion system **100** in accordance with the illustrative embodiment of the invention. Propulsion system **100** includes particles **102**, chamber **104**, piston **108**, source(s) of electromagnetic radiation **110**, return line(s) **112**, and pump(s) **114**, interrelated as shown.

[0053] FIG. **7** depicts method **700** for propulsion, which can be used in conjunction with propulsion system **100**. In accordance with operation **702** of method **700**, a plurality of particles are confined.

[0054] FIG. **3** depicts propulsion system **100** at the beginning phase of the propulsion cycle. Particles **102**, which in some embodiments are ground-state atoms, are confined in particle trap **106** of chamber **104** in known fashion and in accordance with operation **702**.

[0055] In operation **704**, particles **102** are subjected to a trigger acceleration. This can be accomplished, for example, by supporting propulsion system **100** in a gravitational field. Assuming propulsion system **100** is in a vehicle, such support is provided, for example, if the vehicle is at rest on the surface of the Earth or in flight, as long the vehicle is not in free fall. In some alternative embodiments, the trigger acceleration can be kinematic, such as by rotating the craft, or by using a conventional propulsion system to accelerate the craft.

[0056] At operation **706**, the particles are exposed to an amount of electromagnetic radiation that is sufficient to:

[0057] 1. induce an effective inter-particle force that arises between said particles to exhibit long-range interactions; and

[0058] 2. increase the momentum of the particles.

This operation is depicted in FIG. **4**, wherein sources of electromagnetic radiation **110** ("EM sources 110") are activated and directed toward particles **102**. In some embodiments, the EM sources are high-power lasers. In the FIG. **4**, two EM sources **110** are shown. Depending upon the power required for a given embodiment, far more EM sources might be required. Power requirements for driving the propulsion system are described in

[0059] The EM radiation causes an upward acceleration of particles **102** with respect to a vehicle, etc., that houses propulsion system **100**. In the case of an ideal propulsion system, particles **102** remain trapped in place (in particle trap **106**) as they mutually interact and the craft is accelerated upward by the reaction of the atoms themselves against whatever forces are used to keep them in trap **106**.

[0060] It is possible, if not likely, that once particles **102** are accelerated by conducting operations **704** and **706**, they will escape from particle trap **106**. This is a non-ideal situation, which yields less than the ideal momentum. But, if particles are allowed to escape, this relaxes the constraints on particle traps **106**. That is, suitable traps can be readily constructed with existing technologies. See, e.g., H. J. Metcalf and P. van der Straten, *Laser Cooling and Trapping* (Springer, N.Y., 1999); http://ww.rle.mit.edu/cua/research/project02/project 02.vandp.htm. These references describe techniques for trapping atoms at extremely low temperatures.

[0061] Although it is necessary for the generation of the lifting force itself, it is desirable to trap particles at very low temperatures because the thermal speed of, for example, atoms, even at room temperature, is comparable to the maximum speeds that can be obtained by this method. It is, therefore, "easier" to illuminate the atoms for an appropriate length of time if they are not moving at very high thermal speeds.

[0062] Continuing with the description of method **700**, operation **708** recites impacting the accelerated particles against a surface, thereby transferring some of the momentum of the particles to the surface. This operation is depicted in FIG. **5**, wherein particles **102** impact piston **108** at high speed.

[0063] Piston **108** functions as a shock absorber. That is, the piston provides an area against which particles **102** can impact and which can transfer momentum to the vehicle non-destructively at every forward stroke. In propulsion system **300**, non-destructive momentum transfer is indicated by spring **116**, which elastically couples piston **108** to a vehicle. Shock absorber technology for aerospace applications is well developed within the context of pyrotechnic release technology. See, e.g., N. Butterfield, *Pyrotechnic Release Devices, in Space Vehicle Mechanisms,* P. Conley, Ed. (Wiley, N.Y., 1998).

[0064] As illustrated in FIG. **5**, in propulsion system **100**, as piston **108** moves upward, return lines **112** are accessed. The return lines provide a route back to particle trap **106**. As particles move away from piston **108**, the piston drops back to a seated position against chamber **104**.

[0065] FIG. **6** depicts particles **102** in return lines **112**, being pumped via pumps **114** toward a gas reservoir (not depicted) for reuse in a subsequent propulsion cycle. The propulsion cycles occur at a rapid and substantially continuous pace.

[0066] It is very important to recognize that in propulsion system **100**, and in accordance with method **700**, particles **102** are NOT being accelerated by the walls of the vehicle or by the chamber in which they reside. Rather, they are accelerated by their mutual dipole fields, as distorted by vehicle acceleration (i.e., gravitational or otherwise). There is, therefore, no concern that this scheme violates the action-reaction law, since the motion of particles **102** is NOT due to an action upon them by the vehicle.

[0067] An example of a situation in which the walls of the vehicle are acting on particles **102** is if the particles were accelerated by an explosion in chamber **104**. In such a case, the net sum of all internal forces on the system would be zero and, at the end of the process, the vehicle would not gain any net momentum. But using the methods and apparatus described herein, particles **102** are accelerated toward piston **108** INDEPENDENTLY of the vehicle and, on impact, transfer a net amount of momentum to it.

1.2 Vehicle Incorporating Propulsion System 100

[0068] FIG. **8** depicts vehicle **800**, which incorporates propulsion system **100** in accordance with the illustrative embodiment of the present invention. As depicted in FIG. **8**, vehicle **800** includes propulsion subsystem **100**, nuclear power subsystem **810**, crew quarters **830**, and shielding **840**, arranged as shown.

[0069] The presence of nuclear power subsystem **810** requires the use of shielding **840** to protect crew quarters **830** and propulsion subsystem **100**. Those skilled in the art will be capable of designing and building shielding suitable for this purpose.

[0070] The purpose for nuclear power subsystem **810** is to generate electricity to power EM source(s) **110**. Nuclear power is used as a power source due to the ability of a nuclear reactor to provide continuous power for extended periods of time (e.g., several years, etc.). Operation of nuclear power subsystem **810** is described in more detail below in conjunction with FIG. **9**.

[0071] Propulsion subsystem **100** couples to shielding **840**, which receives momentum transferred from piston **108** (see FIGS. **3-7** and the accompanying description). This substantially continuous transfer of momentum from particles **102** to piston **108** to vehicle **800** (e.g., shielding **840**) drives the vehicle.

[0072] In the embodiment that is depicted in FIG. **8**, propulsion subsystem **100** is disposed an end of vehicle **800**. This location draws maximum advantage from a rotational trigger acceleration while providing the crew, in crew quarters **830**, with appropriate gravity-like conditions.

[0073] FIG. **9** depicts an embodiment of nuclear subsystem **810** suitable for use to provide electricity to drive EM sources **110** (e.g., lasers, etc.) in propulsion subsystem **100**. The embodiment that is depicted in FIG. **9** is a direct Rankine cycle continuous power system. (See, e.g., M. W. Edenburn, "Models for Multimegawatt Space Power Systems," Sandia Report SAND86-2742 (June 1990). This type of system is suitable for use with vehicle **800** due to its ability to provide continuous power for several years.

[0074] Nuclear subsystem **810** includes nuclear reactor **812**, separator **814**, turbine **816**, radiator **818**, pump **820**, generator **822**, and power conditioning unit **824**.

[0075] Reactor **812**, which is liquid metal cooled, boils potassium and sends the saturate vapor to turbine **816** for power generation. Since the fluid leaving the "hot" end of reactor **812** is unlikely to be pure vapor, separator **814** is used to separate the saturated vapor from its accompanying liquid. The liquid is recirculated to the "cold" end of reactor **812**.

[0076] Waste heat is rejected by space radiator **818**. Since the system rejects heat from a condensing working fluid, the radiator operates nearly isothermally and radiates a relatively large amount of heat per unit area. Condensed liquid is returned to the "cold" end of reactor **812** via pump **820**.

[0077] Electricity that is produced by generator **822** is appropriately conditioned in power conditioning unit **822** to provide EM sources **110** with a suitable supply of electrical power.

2.1 Distorted Dipole-Dipole Potential

[0078] It is already a well-known fact that a gravitational field can introduce novel forces acting on a single charge or on a dipole. An example is the self-interaction of a point charge in a Schwarzschild geometry [8], ultimately due to the term Linet [9] discovered has to be added to the Copson potential [10] in order to satisfy the appropriate asymptotic boundary conditions for this problem. Commenting about the very extreme conditions nearby a miniblack-hole, Smith and Will wrote that "[I]t is amusing to note that . . . the test particle's electrostatic self-force would suffice to support it against the hole's gravity, without the help of any external force."

[0079] Unfortunately, however, such fascinating conclusion is undermined by the fact that, as pointed out by these authors, "it is meaningless to talk of an electron being held fixed at, say, 10^{-13} cm from a miniblack hole, when the Compton wavelength of an electron is two orders of magnitude larger than this. "Following this approach, the self-interaction of a static dipole has also been calculated [11], but Parker has shown that this force has no effect on the Hamiltonian of a neutral atom in free-fall [12].

[0080] Another example is the "electrostatic levitation of a dipole," predicted on the basis of the distortion caused by a uniform gravitational field [13]. This author found that "one is unlikely to witness such levitation," which could only be observed in a fixed classical dipole whose electron charge separation is 1.4×10^{-15} m. The outlook for detection of these field distortion phenomena was effectively summarized by Boyer, who stated that " Clearly our example may be instructive from a theoretical point of view, but it does not lend itself to easy experimental measurement."[14]

[0081] In this section, we consider the effect of a weak gravitational field upon intermolecular forces. In particular, the effect of gravitation on the van der Waals hydrogenic interatomic potential in the unretarded regime is discussed within non-relativistic first-order perturbation theory. The quantitative conclusion of these computations is that the system proposed herein shows extreme promise for direct experimental verification although the effect is certainly too small to be of any practical engineering use in the field of propulsion.

[0082] The first step to obtain the distorted dipole-dipole potential is the calculation of the electrostatic potential, and thus of the electric field, of a point dipole in the presence of gravitation. For this purpose, let us start by considering the potential field of a single point charge q located at a position $r_0=x_0^i$ in the quasi-homogeneous gravitational field caused by a relatively distant spherically symmetrical mass distribution M located at a radial distance R from the dipole. This has been the subject of several investigations, starting with the pioneering work of Whittaker [15].

[0083] Since we are considering a charge in a gravitational field g antiparallel to the z-axis and located at a position other than the origin, we transform the unprimed Rindler coordinates defined by the usual metric

$$ds^2=\left(1+\frac{gz}{c^2}\right)^2c^2dt^2-(dx^2+dy^2+dz^2), \qquad (1)$$

by introducing new primed coordinates given by $t=(1+gz_0/c^2)^{-1}t'$, $x^i=x_0^i+x'^i$. By substituting these definitions into Eq. (1), it is simple to show that the metric in the new coordinates is:

$$ds^2=\left(1+\frac{G_Rz'}{c^2}\right)^2c^2dt'^2-(dx'^2+dy'^2+dz'^2), \qquad (2)$$

where $G_Rg/(1+gz_0/c^2)$. With this result, we can transform Whittaker's expression for the electrostatic potential in Kottler-Whittaker coordinates into our transformed Rindler frame (for simplicity of notation, we neglect to write the primes in what follows):

$$V(r;r_0)=\frac{q}{|r-r_0|}\frac{\left[1+\frac{g(z+z_0)}{c^2}+\frac{g^2}{2c^4}(|\rho-\rho_0|^2+(z^2+z_0^2)\right]}{\left(1+\frac{gz_0}{c^2}\right)\sqrt{1+\frac{g}{c^2}(z+z_0)+\frac{g^2}{4c^4}|r-r_0|^4}}\tag{3}$$

where $|r-r_0|^2=(x-x_0)^2+(y-y_0)^2+(z-z_0)^2$, $|\rho-\rho_0|^2=(x-x_0)^2+(y-y_0)^2$. This expression of course approaches the Coulomb potential as gx^i/c^2, $gx_0^i/c^2 \rightarrow 0$. In this paper, we shall neglect the Linet term [9] responsible for the self-interaction discussed above since this contribution will be shown to be negligible with respect to the effects treated herein.

[0084] In order to write down the electrostatic dipole field in the presence of a gravitational field, we use the formulation of Leaute and Linet, originally designed to calculate the self-interaction of an electric dipole [11]. For a point dipole A of moment $d_A=d_A^i$, with k=1 . . . 3, located at r_A, the result found by these authors, neglecting the self-interaction term [12], can be written as:

$$U_{dip,W}(x^k,x_A^k)=d_A^i\frac{\partial V_C(x^k,x_0^k)}{\partial x_0^i}\Bigg|_{r_A},\tag{4}$$

where V_C is the Copson potential, which, in the quasi-homogeneous field limit, coincides with our solution above.

[0085] Since, to the best of this author's knowledge, this potential has never been graphically represented, it is shown, along with the corresponding electric field lines, in FIG. **3** for a point dipole $d_A^i=d_Ak$.

[0086] By computing the electric field as usual, after some very lengthy algebra [13] one obtains the general expression for the interaction potential energy $W_{dd}(r, r_0; d_A^i, d_B^i)$ of two point dipoles of moments d_A^i and d_B^i, placed at positions r_0 and r, respectively. In order to illustrate the physical meaning of this result, let us write it to second order in gR/c^2 for the case of dipole A placed at the origin and dipole B placed at r=(R, 0, 0):

$$W_{dd}\approx\frac{q^2}{R^3}\left[(x_Ax_B+y_Ay_B-2z_Az_B)+\frac{1}{2}\left(\frac{gR}{c^2}\right)\left(1-\frac{gR}{c^2}\right)(x_Ax_B+y_Ay_B)\right],\tag{5}$$

which again yields the usual Minkowski space result [17] if the gravitational field is absent. Interestingly, the asymmetry due to the presence of the gravitational field causes a net vertical force upon each dipole in this geometry where there would of course be none in flat space-time. For instance, in the case of two antiparallel dipoles $d_A^i=-d_B^i=d_Ak$, we find that this force $F_{dd}=-\partial W_{dd}/\partial Z$ is, again to second order:

$$F_{dd,z}=\frac{3}{2}\frac{q^2}{R^3}\frac{g}{c^2}z_A^2,\tag{6}$$

which is the dipole-dipole analogy of the levitation force upon a single dipole mentioned at the beginning of this section [15].

[0087] The problem of a single hydrogenic atom either held fixed or in free-fall within various assigned metrics has been discussed extensively by starting from generally covariant expressions of the Dirac equation in curved space-time [18]. The present author has discussed realistic astrophysical settings for the observation of the perturbative effects of gravitational fields on freely-falling atoms both in the static case and within the framework of possible remote gravitational wave detection [19].

[0088] Tourrenc and Grossiord, in their treatment of a hydrogen atom held fixed in a Schwarzschild geometry, have shown that the most significant contribution to the perturbative Hamiltonian by far derives from what can be interpreted as the classical weight of the electron in the gravitational field. However, since the corresponding energy shift for a ground-state atom is found to be $\Delta E\sim2(GMm^e/R^2)a_0\approx6\times10^{-21}$ eV, we shall neglect it here and use the unperturbed hydrogenic wavefunctions to evaluate the interatomic gravitational self-force.

[0089] As in the undistorted van der Waals case, the general expression for the potential $W_{dd}(r, r_0; d_A^i, d_B^i)$ contains only bilinear forms of the type $x_A^ix_B^j$ and thus yields no first-order contribution in the case of symmetrical hydrogenic nS. The second order correction on the other hand is [20]:

$$\Delta E_{vdW}^{(2)}={\sum}'\frac{|<\phi_{n,l,m}^A;\phi_{n',l',m'}^B|W_{dd}(r,r_0;d_A^i,d_B^i)|\phi_{1,0,0}^A;\phi_{1,0,0}^B>|^2}{-2E_I-E_n-E_{n'}},\tag{7}$$

where $|\phi_{n,l,m}{}^A>$ are the unperturbed eigenfunctions of energy E_n, the total energy of the unperturbed atomic pair is $-2E_I$, and the primed summation indicates that the $|\phi_{1,0,0}{}^A;\phi_{1,0,0}{}^B>$ term is excluded.

[0090] By inspecting Eq. (5), it is evident that, in the gravitational case, the second order inter-molecular potential, to second order in gx^i/c^2, takes on the general form:

$$\Delta E_{vdW}^{(2)}=\frac{C_0}{R^6}+\left(\frac{g}{c^2}\right)\frac{C_1}{R^5}+\left(\frac{g}{c^2}\right)^2\frac{C_2}{R^4},\tag{8}$$

where C_0, C_1, and C_2 are appropriate dimensional constants, which in principle depend on the variables r, r_0, d_A^i, and d_B^i. For instance, again in the geometry used in the examples above, one can quickly recover the well-known result that $C^0\approx-6e^2a_0^5$ as well as show that $C_1\approx2e^2a_0^5$, and $C_2\approx+3/2e^2a_0^5$, where a_0 is the Bohr radius.

[0091] In our case, however, we are not here interested in the modification of the intermolecular forces due to the presence of gravitation, but rather we want to pursue isolating the vertical component of the gravitational self-force due to the dipole-dipole interaction of two hydrogenic atoms in the $|1,0,0>$ eigenstate located at $r_0=0$ and r=(R, 0, 0), respectively:

$$F_{sf,z}^{(2)}=-\frac{\partial}{\partial z}\Delta E_{vdW}^{(2)}\approx\tag{9}$$

-continued

$$-\frac{|<\phi^A_{1,0,0};\phi^B_{1,0,0}|\frac{\partial}{\partial z}W_{dd}(r,r_0;d^i_A,d^i_B)|\phi^A_{1,0,0};\phi^B_{1,0,0}>|^2}{-2E_I},$$

Explicit evaluation [13] yields the following result to first order in gx^i/c^2:

$$F^{(2)}_{sf,z}\approx\frac{e^4}{2E_IR^6}\frac{g}{c^2}<\phi^A_{1,0,0};\phi^B_{1,0,0}|(4x_A^2x_B^2-4x_Ax_By_Ay_B+y_A^2y_B^2-3z_A^2x_B^2-14x_Ax_Bz_Az_B+4y_Ay_Bz_Az_B-3x_A^2z_B^2+3z_A^2z_B^2)|\phi^A_{1,0,0};\phi^B_{1,0,0}>. \tag{10}$$

By again making use of the fact that the cross terms vanish in the is state and that $<\phi_{1,0,0}{}^A|x_A{}^2|\phi_{1,0,0}>=<\phi_{1,0,0}{}^A|\tfrac{1}{3}r_A{}^2|\phi_{1,0,0}{}^A>$ and equally for all the other squared terms, we finally find:

$$F^{(2)}_{sf,z}\approx+\frac{e^4}{2E_IR^6}\frac{g}{c^2}\times2\left|<\phi^A_{1,0,0}\left|\frac{1}{3}r_A^2\right|\phi^A_{1,0,0}>\right|^2=+2\frac{e^2a_0^5}{R^6}\frac{g}{c^2}=+\frac{1}{16}\hbar\omega_0\frac{\alpha_0^2}{R^6}\frac{g}{c^2}, \tag{11}$$

where the last equality was obtained by writing the ionization energy and the polarizability as $E_I=e^2/2a_0$ and $\alpha_0=(2a_0)^3$, respectively. Notice that this result does not coincide with what one might expect from naively calculating one half of the negative contribution to the gravitational mass of the London binding energy of the pair, since in that case the coefficient would be ⅜ and not 1⁄16.

[0092] An explicit estimate of this result in the case of two hydrogen atoms in their ground state at, for instance, 20 a_0, yields a relative acceleration $\alpha_{lift,H}$~4×10^{-13} cm/s^2. The situation improves dramatically if one considers two positronium (Ps) atoms, in which case we find $\alpha_{lift,\ Ps}$~8×10^{-10} cm/s^2, which is in principle detectable via atomic interferometry, at the price of dealing with the added complications of interferometry of atoms with a finite lifetime.

2.2 Revelance of the Above Results

[0093] The above results show that the distortion of the classical dipole-dipole field caused by gravity results in a net force upon each dipole, which is anti-parallel to the weight of each atom. This phenomenon has been known for some time, although its implications at the quantum level have not been fully explored and engineering implications do not appear to have attracted their due attention. Interestingly, the warping of the Coulomb field due to gravitation renders their mutual interaction non-central, which in turn implies that Newton's action-reaction law will not be satisfied by this system. Although there is no study of this exotic problem in the literature, it is clear that, if the problem were to be considered from the standpoint of full quantum electrodynamics (QED), it would result that the virtual photon field responsible for the charge-charge interaction is perturbed by the gravitational acceleration so as to carry a net momentum flux, part of which is simply transferred to the dipole, or dipoles, thus resulting in their upward motion.

[0094] Importantly, this phenomenon makes the interaction energy between two hydrogen atoms in their ground state dependent upon their position within the gravitational field, which results in a force upon the pair even in the quantum case. The existence of this additional force acting on an atomic pair is by itself a new result, since previous studies had concentrated on the distortion of the field of two point charges, and not of two dipoles. Despite the fact that this new and additional force is measurable with presently existing technology, its use in an actual lifting device is unlikely, since it amounts to a very small correction of the total weight of each atom.

[0095] The important consideration, critical to the present invention, is that adding many atomic pairs does not result in a larger force per atom. The reason for this is that the interatomic energy is a function of a large power of the reciprocal of the distance—which results in the atom-atom dispersion interaction being a short-range force. Therefore, the total lifting force on a large number of atoms is increased only very slightly with respect to that acting on just one pair. Therefore, if the number of atoms is N, there will be ~N pairs to consider but the mass of the of the system also went up as N. Thus no gain is made.

[0096] The critical, and non-obvious, point of the present invention is to transform the new force outlined above from a short-range force into a long-range force, as we consider in the next section. In principle, if point-like particles interact via a long-range force, the total energy of the system results form the interaction of every atom with all the others, something which is not possible in the short-range. For instance, if a system is made up of three particles, the total energy will result from the interaction of particle 1 with 2, particle 1 with 3, and particle 2 with 3. However, if the number of particle is now ten, we will have to consider the interaction of particle 1 with 2, 3 . . . , 10, and so fort, which results, in the case of a large number of particles. For large N (N>>1), in this case the number of interaction grows as N^2, while the total mass of the system is still only growing proportionally to N.

[0097] This point would only be of philosophical importance if we did not have available a mechanism to indeed transform atom-atom interactions from short-range into long-range. Such mechanism was discovered theoretically several years ago and has been reproposed recently as a way to introduce unusual behaviors in a cloud of trapped atoms. Atomic traps have become one of the hottest subjects of scientific research in recent years. The method proposed by the researchers to cause the atom-atom interaction to become a long-range force can of course also be used in our case to leverage the presence of a large number of atoms so as to make the lifting force due to the gravitational distortion we have seen above much larger by many orders of magnitude. The technological price to pay is that, in order for this transformation to occur, powerful lasers must be pumping an intense radiation field throughout the region where the atoms are trapped.

2.3 Trapped Gases in Curved Space-Time: The Effect of Radiation Fields

[0098] It is well-known that an intense directional radiation field, such as that produced by a laser, alters the nature of intermolecular forces [21]. For instance, in the near zone region, where $k_L R<<1$ and k_L is the laser light wavenumber, the power dependence of the force becomes $\alpha-1/R^3$. Importantly, if a molecular pair is allowed to "tumble" with equal probability in all directions with respect to the radiation field, the unretarded force averages out and the only term left is that due to the retarded part of the Hamiltonian. The resulting contribution, which can be produced by appropriate laser beams [22] and is still attractive, is [21]:

$$\Delta E \simeq -\frac{44}{15}\left(\frac{I\pi k_L^2}{c}\right)\frac{\alpha^A(k_L)\alpha^B(k_L)}{R}, \quad k_L R \ll 1, \tag{12}$$

where $\alpha^{A,B}(k_L)$ are the dynamic polarizabilities of the two atoms, I is the intensity of the beams, and the domain of validity of this result is everywhere in space except where the atom-atom exchange interactions become important. It is simple to see numerically that this energy is much smaller than the usual van der Waals force at near range but, as originally pointed out by Thirunamachandran [21], the long-range nature of the force offers the potential to actually achieve remarkable effects.

[0099] Let us now consider the distortion of Thirunamachandran's long-range interaction due to an external gravitational field, such as that present in a ground-based laboratory. One is fully justified by both our results in the unretarded case and by dimensional considerations to assume that a molecular pair interacting through the gravity-like attractive long-range force at Eq. (12) will also undergo a lifting force of the type:

$$F_{lift,z}^{AB} \sim \left(\frac{Ik_L^2}{c}\right)\frac{\alpha^A(k_L)\alpha^B(k_L)}{R}\frac{g}{c^2}. \tag{13}$$

[0100] For the purpose of order of magnitude estimation, let us write the total potential energy of N atoms contained in a spherical volume of diameter D interacting through a mean field determined by Eq. (12) simply as $U_{gas}\sim-N^2 (Ik_L^2/c)\ \alpha_A(k_L)\alpha^B(k_L)/D$. Therefore, the total lifting force acting upon the center-of-mass of the trapped gas is $F_{lift,z}^{gas}\sim N^2 F_{lift,z}^{AB}$, assuming all atoms to be of species A, and the corresponding acceleration is:

$$\frac{a_{lift,z}^{gas}}{g} = \frac{F_{lift,z}^{gas}}{Nm_A g} = N\left(\frac{Ik_L^2}{c}\right)\frac{[\alpha^A(k_L)]^2}{D}\frac{1}{m_A c^2}. \tag{14}$$

The relevant figures of merit to judge the feasibility to bring such an atomic cluster to a hover are the number of atoms in the trap and its size, the intensity and wavelength of the laser light, and the average intermolecular distance. Consider $N=10^2$ atoms in a trap with $D\approx 2\times10^{-7}$ cm, which yields an intermolecular distance $R\sim D/N^{1/3}\approx 10a_0$. Now let us have 18 (six triads [22]) high-power lasers each outputting 2.5 kW into a 0.5 cm diameter beam at a wavelength $\lambda_L\approx 1000$ Å. Appropriately focused onto the trap size, this would yield an intensity $I\approx 1.3\times10^{17}$ W/cm^2. By moderate off-resonance detuning, it is possible to obtain dramatic increases in atomic polarizability over its static value [23] (another approach may consist of using a cold gas of highly excited Rydberg atoms, since, in this case, the polarizability is proportional to n^7, where n is the principal quantum number). By adopting $\alpha^A(k_L)\approx 3\times10^5\alpha_0$, and by substituting the above numerical values into Eq. (14) in c. g. s. units, we find $a_{lift,z}^{gas}/g\approx+1.5$, that is, the system will hover unsupported in the gravitational field of the earth or accelerate upward.

2.4 Relevance of the Above Results

[0101] The mechanism outlined above represents the first novel and realistic proposal to achieve lift in the history of flight since the great inventions of the airplane and of the rocket in the 20th century. As we shall see in the detailed numerical estimates below, more atoms than just 10^2 must be present in the trap for this mechanism to be technologically convenient, although it is possible to trade off a higher laser power for a lower number of atoms in the trap. What is important to stress at this time is that this mechanism represents a non-obvious use of well understood quantum laws in the presence of gravitation for the purpose of creating fuel-free propulsion. Of course fuel-free propulsion does not imply energy-free propulsion. In other words, a source of energy is still needed to achieve the needed thrust. In the case of extremely long interplanetary and interstellar flights it is expected that on-board nuclear energy production will continue to grow in engineering importance, given the fact that no other source has been able to achieve similarly convenient power outputs.

[0102] The above invention, however, removes the greatest problem in the way of achieving fuel-free propulsion, that is, the fact that, independently of the energy source used, at some point the fuel available on board is exhausted. For instance, if one could power a spacecraft so as to achieve a constant acceleration of 1 g, the trip from Earth to Mars would require approximately 2-4 days. Although the acceleration of 1 g is smaller than the larger accelerations achievable by present-day rocket technologies, it is absolutely impossible to maintain those accelerations for times longer than minutes at the most, simply because of quick fuel exhaustion. With the scheme outlined in this invention, on the other hand, it is possible to maintain the needed acceleration without any need to expel high speed gases, provided that the required laser illumination is constantly at work transforming the atom-atom interactions from short-range into long-range ones.

[0103] It is important to stress another characteristic of the propulsive method of this invention. All the calculations carried out so far imply that the atoms are "at rest," that is, not in "free-fall." This is extremely important, since it is only if the atoms are somehow supported by an external force against the gravitational field that a relative acceleration due to that field can affect their mutual interactions. Such is not the case if the atoms are freely falling. In that case, in fact, the acceleration felt by a freely falling object is rigorously zero, because of the Principle of Equivalence at the foundation of the General Relativity theory. To use a well-known popular example, if two atoms are at rest with respect to the walls of a freely falling elevator, they will not feel the presence of any gravitational field—the acceleration of the elevator exactly cancels exactly the gravitational acceleration. In other words, locally, there is no distortion of the dipole-dipole field and thus no change to the van der Waals force (see below for further subtle clarifications on this point).

[0104] The Principle of Equivalence can be stated by saying that, locally, there exists no experiment that can indicate the difference between the acceleration due to the presence of a gravitational field and that due to the kinematics of the system. For instance, once could simulate the presence of a gravitational field in the same elevator travelling through outer space by simply accelerating it "upwards" at the same rate as the free-fall acceleration it would have in the gravitational field to simulate. In fact, our Eq. (1) above was obtained exactly by making this assumption. Therefore, the atoms will undergo the lifting force at the basis of this invention whether or not there is a gravitational field against which to lift—their behavior is due to their being within an accelerated reference frame no matter what the reason for the existence of such frame.

[0105] In principle, this represents an operational limitation of the present invention, in the sense that, if a spacecraft were to be left to freely-fall in the gravitational field of a massive body, the lifting mechanism would not be operating. For this purpose, the vehicle must be provided with an initial acceleration through other means in order for the thrust cycle described below to commence. For instance, this happens if the craft is at ground level initially, or somehow hovering under the action of an external force. Once the cycle starts, it is only necessary to coordinate the laser illumination of the (n+1)-th cycle to occur during the transfer of momentum due to the atoms that were accelerated during the n-th cycle. The dipole-dipole field during that time will behave as though under the effect of a gravitational field of that acceleration, because of the Principle of Equivalence.

[0106] From the practical standpoint, it is appropriate to stress that every propulsive or lifting system has an appropriate envelope of performance which, is exceeded or not met, will result in insufficient or abnormal behavior. For instance, the lifting force due to the wings of an airplane will cease to be effective if the airflow detaches from the wings because of a stall condition. Therefore, pilots are trained to operate so as to remain well clear of the conditions that might lead to a stall of the airfoils, such as, for instance, excessively low speed for a fixed wing aircraft. Similarly, it is expected that the propulsive system of this invention, in its simplest embodiment, must be operated under appropriate conditions of initial acceleration, in both magnitude and direction. This is much less complicated than it may appear. For instance, in outer space, an initial acceleration can be imparted by causing the entire spacecraft to rotate around an axis by means of a reaction wheel. Once the entire vehicle is rotating like a rigid disk, from the standpoint of the atoms in the propulsive subsystem the rotational acceleration will be indistinguishable from that caused by gravity. Once the thrust cycle is successfully started, the spacecraft can then be despun while the engine provides its own acceleration. In a fixed wing application, it might be possible to make use of the lift provided by the wings themselves in order to start the process.

[0107] Finally, it is also important to stress that the Principle of Equivalence only rigorously applies to a volume of space that is infinitely small. Therefore, there exists some distortion of the dipole-dipole field even in the case of freely-falling atoms, although this distortion is far smaller than that of supported atoms, in the sense described above. The effects of this distortion were studied by this author in several papers (see for instance [19]) and it is therefore possible that lift might be obtained in some embodiments even if the craft is initially in free-fall.

4. Performance

[0108] Estimating the performance of an aerospace propulsion system based upon entirely novel physical principles naturally presents some difficulties. For instance, the typical concept of specific impulse [29] is undefined in the case in which thrust is obtained without the ejection of high speed gases. At the same time, since the approach calls for the use of high power lasers to engineer the atom-atom interactions into a long-range force, it is of interest to determine whether the thrust thus obtained is in fact larger than that which would be obtained if the laser power utilized were, for instance, projected from the spacecraft into a particular direction in space or whether a laser beam of the same power were to be aimed at a hypothetical laser sail on the spacecraft [3]. In the following subsections we obtain some important order of magnitude estimates both in equation and in graphic form of a few important quantities in order to gain a more realistic understanding of the potential capabilities of a vehicle propelled by means of the physical principle of the present invention.

[0109] The conclusions below will clearly establish that the propulsion concept of this invention offers great potential from the standpoint of realistic engineering applications, although such parameters as the exact laser wavelength and power, trap size, atomic mass, and number of atoms of course will have to be optimized according to both accurate theoretical modeling and prototype testing.

[0110] In what follows, in order to make a firm connection between the theoretical treatment, which was developed here in the c. g. s. system (centimeter-gram-second), and the more typical engineering M. K. S. units (meter-kilogram-second), the cgs or MKS subscripts will be appended as appropriate. If no subscript is used, the quantity should be assumed as expressed in the cgs system. No use is made if English units throughout (such as, for instance, lbf for thrust). Also, for improved legibility, all order of magnitude signs will be replaced by equal signs.

4.1 Fundamental Equations

4.1.1 Atomic Physics of Trapped Atoms in the Accelerated Propulsive System

[0111] Let us consider a gas of N_A identical atoms of mass m_A, polarizability $\alpha_A^2(k_L)$, confined within an appropriate trap of such dimensions as to correspond to an average interatomic distance $\bar{R}$. In what follows, we shall assume that the number of atoms, N_A, the size of the trap, D, and the average interatomic distance, $\bar{R}$, are related simply as $D \sim RN_A^{1/3}$. In addition, Thirunamachandran's theory of dispersion forces under the effect of illumination also requires the constraint that $\lambda_L >> \bar{R}$[21].

[0112] The polarizability $\alpha_A^2(k_L)$ can be made several orders of magnitude larger than its static value, $\alpha_0=(2a_0)^3$, where Bohr's radius is $\alpha_0=h^2/\mu_e e^2$ and μ_e is the reduced electron mass ($\mu \approx m_e$), by choosing an appropriate near-resonance wavelength. Without getting into the details of atomic physics calculations, in this section we shall rely upon the well-established theoretical and experimental fact that such near-resonance condition can be satisfied, that is, the polarizability can be made larger than the static value by a factor, α_{nr}, which can be as large as $\alpha_{nr} \sim 10^5$ [23].

[0113] Another strategy to produce values of the polarizability that are vastly larger than the static value is to use Rydberg atoms. A qualitative argument in favor of this choice is that, as we have seen above, the static polarizability is $\alpha_0=(2a_0)^3$, where a_0 is Bohr's radius. If the atom is in a Rydberg state, that is, in an excited state with relatively large principal quantum number n>>1, the atomic radius can be replaced by $a_n=a_0n^2$. On the strength of this argument alone, the atomic polarizability of a Rydberg state appears proportional to n^6. In fact, if one accounts for all states with different orbital quantum number l, the static polarizability of a Rydberg atom results proportional to n^7. For instance, the atomic radius of atoms in one-electron Rydberg states with $n \sim 10^2$, which are routinely created in the laboratory and are present in interstellar space, is $\sim 10^4 a_0$—similar to the size of an Ebola virus! Under these circumstances, the static polarizability is a stunning $\sim 10^{14}$ times larger than its static value [19, 30, and Refs. therein].

[0114] From the practical standpoint, it is important to notice that Rydberg atoms gases have already been "frozen" and trapped in order to study, among others, the very dipole-dipole interactions we discussed at the very beginning of this disclosure [31]. The atoms themselves are prepared by causing them to absorb laser light of wavelength appropriate to induce a radiative transition to the desired excited state. Noticeably, the radiative lifetime of Rydberg atoms can be quite long, even compared to the relatively short crossing time within the propulsive system. The choice of atoms in states other than the ground state, such as Rydberg atoms, imposes an additional constraint upon the interatomic distance, since the radius of these atoms can be macroscopic. It is therefore important to require that the interatomic distance be much larger than the Rydberg atom radius, which in turn affects the size of the entire propulsive system.

[0115] It is appropriate at this time to again stress the difference between the atom-atom force, which is related to the dependence of the dispersion energy on the interatomic distance and is a central force, and the vertical lifting force acting upon all atoms as a consequence of the modification of their interaction potential in an accelerated reference frame. For instance, let us again consider the potential energy of a pair of atoms (Eq. (12)) and let us rewrite it to highlight the close similarity to the gravitational case:

$$\Delta E = -G_{QED}\frac{\alpha_A(k_L)\alpha_B(k_L)}{R},\ \ k_LR \ll 1, \tag{12a}$$

where the equivalent quantum-electro-dynamical "gravitational constant" is, by definition:

$$G_{QED} \equiv \frac{176\pi^3}{15c}\frac{I}{\lambda_L^2}, \tag{15}$$

and the dynamic polarizabilities play the role of the "gravitational mass." The total potential energy can then be written as usual:

$$U_{gas} = -\frac{1}{2}G_{QED}\sum_{i\neq j}\frac{\alpha_i(k_L)\alpha_j(k_L)}{R_{ij}}, \tag{16}$$

where the summation is meant over all pairs and R_{ij} is the magnitude of the interparticle distance vector, R_{ij}.

[0116] Every atom in the gas is acted upon by a gravitational-like self-force and thus undergoes an acceleration gQED towards the center of the cloud (assumed approximately spherical) similar to a typical gravitational acceleration. This is in complete analogy to the collapse of a "cold" gas sphere taking place whenever the gravitational pressure is vastly larger than any opposing gas pressure gradient. The order of magnitude of the time required for the entire cloud to collapse to its center is an important characteristic time in stellar evolution and in stellar dynamics, and it is referred to as the free-fall time [32-33]. In order to generalize this quantity to our case, let us consider the equation of motion of a particle (atom) in the above potential at a distance r from the center of the cloud with $\alpha_A(k_L)=\alpha_B(k_L)$ and N_A>>1. In this case it is well-known from elementary mechanics that Gauss' Theorem allows us to only consider the force exerted by the atoms inside a sphere of radius equal to r and we find:

$$\ddot{r} = -G_{QED}\frac{\alpha_A^2(k_L)}{r^2}\frac{1}{m_A}. \tag{17}$$

Since we are only estimating an order of magnitude, let us assume, as done typically, that the acceleration is approximately constant during the free-fall process. Thus:

$$\frac{1}{2}\ddot{r}t_{ff}^2 \sim D,$$

which yields, for r~D

$$t_{ff} \sim \sqrt{\frac{2D}{\ddot{r}}} \sim \sqrt{\frac{2N_A\overline{R}^3}{G_{QED}\alpha_A^2(k_L)m_A}}\ . \tag{19}$$

Finally, by substituting Eq. (15) into this result, we obtain:

$$t_{ff} \sim \sqrt{\frac{2N_A\overline{R}^3}{\alpha_A^2(k_L)}\frac{15c}{176\pi^3}\frac{\lambda_L^2}{I}m_A}\ . \tag{18}$$

The importance of this result lies with the fact that the physics of the system after such free-fall time must be expected to be substantially different than in its initial state. For instance, in the case of atoms in their Rydberg states, a drastic evolution of the system towards a higher density configuration can be expected to result in the transformation of the gas into a neutral plasma, with consequent complete loss of thrust.

[0117] The condition to be required so that the atoms do not have the time to evolve into an extremely different, and technologically useless, state is that the free-fall time above be much longer than the time the atoms spend in the trap before the lifting force causes them to be ejected, Δt_A. Evidently, if the gas evolves into, for instance, a plasma upon ejection, that is of no consequence to the momentum it will transfer to the vehicle. By using the result below at Eq. (40), we can write this requirement as:

$$t_{ff} \gg \Delta t_A, \tag{20}$$

or

$$\sqrt{\frac{2N_A\overline{R}^3}{\alpha_A^2(k_L)}\frac{15c}{176\pi^3}\frac{\lambda_L^2}{I}m_A} \gg \sqrt{\frac{\overline{R}N_A^{1/3}}{\left[\frac{4\pi^2}{c^3}N_A\left(\frac{I}{\lambda_L^2}\right)\frac{\alpha_A^2(k_L)}{\overline{R}N_A^{1/3}m_A}-1\right]g}}, \tag{21}$$

If the lift acceleration is much larger than g, the term in square brackets in the denominator of the right-hand-side will become much larger than unity so that term can be neglected and we easily find the limiting condition:

$$N_A\frac{Dg}{c^2} \gg \frac{22}{15}\pi. \tag{22}$$

This condition, which of course results also by requiring that $\alpha_A \gg r$, can be satisfied by realistic values of the geometry of the propulsive system, although it clearly points to the usefulness of employing larger traps, for which $N_A \gg 1$.

[0118] That the physical state of the gas can in fact be extreme after a few free-fall times if this is not accomplished, can be seen by writing the condition that the gas be in equilibrium under the action of this gravitylike interaction. As well-known, gravitationally bound systems do not display what can be properly referred to as equilibrium configurations in the thermodynamical sense. This is well illustrated by an appropriate similitude between the cold gas in the propulsive system of this invention and a globular star cluster—a spherical system in which thousands to hundreds of thousands of stars are bound by their mutual gravitational interaction—or, alternatively, a star [22].

[0119] In a globular cluster (or in a star), the distribution of velocities of the constituent particles "relaxes" to a quasi-Maxwellian velocity distribution after a time properly referred to as the "relaxation" time of the object. However, a Maxwellian velocity distribution contemplates a finite number of particles whose speeds at any given time are higher than the escape velocity from the system. Therefore there occurs a process of constant evaporation, which clearly forbids the existence of any equilibrium configuration [32-35].

[0120] However, it is possible to define a condition of quasi-equilibrium, in which a gravitationally bound system does not change drastically over many relaxation times. Under these conditions, the object obeys a general theorem referred to as the virial theorem, which links its total average kinetic and potential energies. This connection is very powerful, as it allows one to obtain estimates of the average speeds of its particles, whether they be stars or atoms [35-36]. For the purposes of our estimates here, this very general theorem can be written as

$$2\langle K_{gas}\rangle+\langle U_{gas}\rangle=0, \tag{23}$$

where K_{gas} is the total kinetic energy and the triangular brackets indicate the time-average. By writing the total kinetic energy as $K_{gas}=3/2N_Ak_BT_{gas}=½N_Am_A\langle v_A^2\rangle$, and by approximating $\langle 1/D\rangle=1/\langle D\rangle$, the virial theorem yields:

$$T_{gas}=\frac{8\pi^2}{3k_Bc}N_A\frac{I}{\lambda_L^2}\frac{\alpha_A^2(k_L)}{D}, \tag{24}$$

and

$$v_A=\sqrt{\frac{8\pi^2}{3c}\frac{N_A}{m_A}\frac{I}{\lambda_L^2}\frac{\alpha_A^2(k_L)}{D}}, \tag{25}$$

The physical significance of these equations can be found by imposing the condition that the average kinetic energy be equal to the ionization potential of the atoms E_I in the gas so that the atoms may become ionized as they collide, in analogy to the Saha and Boltzmann equations of stellar astrophysics [33]:

$$\frac{3}{2}k_BT_{gas}\sim\frac{E_0}{n^2}, \tag{26}$$

where the right-hand-side of the above equation contains the ionization potential of a hydrogen atom in its n-th state and E_0 is the ionization energy of the ground state. By using Eq. (24), this yields:

$$\frac{4\pi^2}{c}N_A\frac{I}{\lambda_L^2}\frac{\alpha_A^2(k_L)}{D}\sim\frac{E_0}{n^2}, \tag{27}$$

which can be expressed in even more fundamental terms in the case of Rydberg atoms as:

$$\frac{4\pi^2}{c}N_A\frac{I}{\lambda_L^2}\frac{(2a_0)^6n^{14}}{D}\sim\frac{e^2}{2a_0n^2}, \tag{27}$$

$$\frac{4\pi^2}{e^2c}N_A\frac{I}{\lambda_L^2}\frac{(2a_0)^7n^{16}}{D}\sim 1. \tag{28}$$

In the near-resonance case we can instead write:

$$\frac{4\pi^2}{E_0c}N_A\frac{I}{\lambda_L^2}\frac{(2a_0)^6\alpha_{nr}}{D}\sim 1, \tag{29}$$

Replacing the numerical values we shall produce below in a few realistic examples, it is immediate to conclude that the almost immediate ionization of the entire atomic population is highly likely in all technologically meaningful cases, and that, as already pointed out in this section, it is imperative to choose parameters that ensure the expulsion of the gas from the trap well before this process is completed.

4.1.2 Validity of the Present Approach

[0121] In this subsection we concern ourselves with the possible limitations of our treatment. Let us first of all notice that we have so far dealt with atoms in the trap as "classical" objects, that is, as material particles whose positions and velocities can be likened to those of stars in a cluster, for instance. This is approximately correct only if the de Broglie wavelength λ_A of the atoms is much smaller than the interatomic distance, that is, if:

$$\lambda_A \equiv \frac{h}{p_A} \approx \frac{h}{m_A v_A} \ll \bar{R}, \tag{30}$$

where p_A is the momentum of the atoms and the middle step is warranted for non-relativistic speeds. This condition may be violated for extremely low temperatures and for high densities, such as those of artificial "white dwarfs," and in these cases the gas must be treated as a quantum gas of the appropriate statistics as is the case, for instance, in white dwarf stars [22, 33]. Although such extreme conditions are only marginally important to the present work, a well-established theoretical framework exists in the literature to describe them.

[0122] Secondly, the basic result at Eq. (12) is rigorously valid only within fourth-order perturbation theory [21]. An order of magnitude of the regime within which our results are certainly warranted can thus be found by requiring that the energy shift per atom due to the total intermolecular force of the gas cloud be much smaller than the energy of the n-th unperturbed state occupied by the atoms, that is:

$$|E_n|>>N_A|\Delta E|, \tag{31}$$

or, for hydrogen atoms:

$$\begin{aligned} \frac{e^2}{2a_n} &\gg \frac{176\pi^3}{15c} N_A \frac{I}{\lambda_L^2} \frac{\alpha_A^2(\lambda_L)}{D} \\ &= \frac{176\pi^3}{15c} N_A \frac{W}{18 N_A^{2/3} \bar{R}^2} \frac{1}{\lambda_L^2} \frac{\alpha_A^2(\lambda_L)}{N_A^{1/3} \bar{R}} \\ &= \frac{176\pi^3}{270c} \frac{W}{\bar{R}^3} \frac{1}{\lambda_L^2} \alpha_A^2(\lambda_L), \end{aligned} \tag{32}$$

where the relationship between total power and intensity at Eq. (48) was used. Let us consider only the s-state polarizability as $\alpha_n(\lambda_L)=\alpha_{nr}(2a_n)^3$ and let us write the average intermolecular distance in terms of the atomic radius as $\bar{R}=\bar{r}a_n$. Finally, by recalling that $a_n=a_0 n^2$, we find:

$$\frac{e^2}{2a_0 n^2} \gg \frac{1.2\times 10^3}{c} \frac{W}{\bar{r}^3 a_0^3 n^6} \frac{1}{\lambda_L^2} \alpha_{nr}^2(\lambda_L) a_0^6 n^{12}, \tag{33}$$

or

-continued

$$\frac{W}{\bar{r}^3} \frac{\alpha_{nr}^2(\lambda_L)}{\lambda_L^2} n^8 \ll \frac{e^2}{2a_0^4} \frac{c}{1.2\times 10^3} \sim 4\times 10^{21}. \tag{34}$$

The dynamical consequences of this constraint can be seen by recalling our basic Eq. (14) written by making use of the first two terms of Eq. (32):

$$\frac{a_{lift,z}}{g} = \frac{4\pi}{c} N_A \left(\frac{I}{\lambda_L^2}\right) \frac{\alpha_A^2(\lambda_L)}{D} \frac{1}{m_A c^2} \ll \frac{60}{176\pi^2} \frac{e^2}{2a_n} \frac{1}{m_A c^2}. \tag{35}$$

In order to extract a physical meaning from this requirement, let us consider the problem of bringing the gas cloud to a hover from a different standpoint. The condition of balance between weight and lift of one atom is given by Eq. (14) above:

$$\frac{a_{lift,z}}{g} \sim \frac{F_{lift,z}^{gas}}{N_A m_A g} = N_A \left(\frac{I k_L^2}{c}\right) \frac{[\alpha^A(k_L)]^2}{D} \frac{1}{m_A c^2}. \tag{14}$$

By using the expression for the potential of one pair at Eq. (12), we can express this condition by introducing the total intermolecular energy of one atom due to all the others, $N_A \Delta E$:

$$\frac{a_{lift,z}}{g} \sim \frac{\Delta E}{m_A c^2}. \tag{36}$$

With the usual strong caveats of using our non-quantum "intuition" to interpret results in the realm of quantum-electro-dynamics (QED) in curved space-time, a possible qualitative view of this hovering condition is that this requirement corresponds to having the total (negative) intermolecular potential of the gas cloud cancel its total gravitational mass, thus resulting in atoms which are, for all practical purposes, "weightless." By going further, it is possible to consider intermolecular potentials which are negative and larger in magnitude than the gravitational mass of the atoms, thus causing the total energy of the cloud to become effectively negative. According to Newton's Law of Gravitation, this would require the atoms to be "repelled by gravity," instead of being attracted. Although this is an interesting and useful image—which has in fact been pursued by Boyer in the past [16]—one has to be extremely careful to adopt it as an "explanation."

[0123] In the scheme of the present invention, the upward lifting force does not result from such view, but from the well-understood distortion of the Coulomb potential due to the presence of the gravitational field. Although Boyer's work proved these two points of view to be equivalent [16], his study dealt with completely classical dipoles within the realm of special relativity and very weak gravity, and not with fully quantized atoms with the framework of QED in curved space-time. Since one possible interpretation of dispersion forces between molecules involves a modification of the zero-point-energy of the quantum vacuum [37], such simple "semi-classical" explanations should be looked at only as useful mental pictures, since the "system" under consideration is actually an open system.

[0124] In this context, two possible objections must be addressed. The first is whether it is at all correct to use perturbation theory in this case. After all, in order to bring a cloud of hydrogen atoms to a hover, the intermolecular potential must contribute an energy approximately equal to the rest-mass of a hydrogen atoms, or ≈1 GeV, whereas the ionization potential of a hydrogen atom is 13.59 eV if the atom is in its ground state and much less if it is in a Rydberg state. The intermolecular potential will be vastly larger if we want to achieve an upward acceleration. The answer rests with the key fact that, in our case, the processes we are studying take place on relatively "short" time-scales. At some time after the cloud of atoms has been cooled and trapped, a very intense beam of radiation is turned on within a total time that will be assumed to be much shorter than that of any possible atomic transition. For instance, the radiative lifetimes of very high n Rydberg atoms can be even fractions of a second, whereas lasers can be turned on in fractions of a nanosecond (10^{-9} s) and that radiation will require a similar time to cross the entire trap at the speed of light.

[0125] A well-known feature of the perturbation theory of "sudden" interactions is that the perturbation does not have to be small for theory to be used, unlike most other applications of perturbation theory, as stated, for instance in [38] (see also [39-41]): "The transition probabilities in instantaneous perturbations can also be found in cases where the perturbation is not small . . . If the change in the Hamiltonian occurs instantaneously (i.e. in a time short compared with the periods $1\omega_{if}$ of transitions from the given state i to other states), then the wavefunction of the system is "unable" to vary and remains the same as before the perturbation. It will no longer, however, be an eigenfuction of the new Hamiltonian $\hat{H}$ of the system . . . "

[0126] By making further use of this same mental picture, we can say that in the description of the fundamental physical process at the basis of the present invention we assume that the phase of upward acceleration of the atoms takes place in the very early stages immediately following the application of the laser radiation field and before the eventual modification of the wavefunctions intervenes. In other words, there exists a relatively brief time span immediately following the laser turn-on time, during which the atomic dipoles still behave as such, although, eventually, the gas evolves towards a very different state, such as a plasma for instance. Part of the refining work of the present invention will be to consider in detail the dynamics of the evolution of the atoms in the intense radiation field. However, the references quoted show that there is a solid logical foundation in the use of perturbation theory in the present case due to the instantaneous application of the radiation field.

[0127] The second objection to consider is whether it is against any fundamental law of physics to even consider the existence of a volume of space where the total energy is negative, as appears to be necessary in order to obtain a gravitational "repulsion" of the atoms. We have already provided a first hint above that one answer to this objection is that it is not at all necessary to look at this process as being due to a "negative" energy. In fact, one alternative and much more satisfactory way to explain the upward motion of the atoms is to appeal to the distortion of the dipole-dipole field due to the acceleration in the presence of a gravitational field. The distortion of the field lines due to the presence of gravitational appears to be a much firmer concept than that of a "negative" energy.

[0128] However, even if one wants to engage in the widespread debate concerning the existence of negative energy, it is interesting to point out that dispersion forces in general and Casimir forces in particular are in fact commonly used as examples to the affirmative. In other words, whereas there exist strict quantum limits, similar to the uncertainty principle, as to the length of time a very negative energy density can be imposed upon a volume of space, dispersion forces clearly afford an example where the energy density can remain negative indefinitely (although the same limitations apply if one attempts to decrease the energy even further). The important point to the present invention is that there is no fundamental physical reason that dispersion forces cannot be made negative for the brief time needed to accelerate the atoms upward as required to provide thrust. On the other hand, much more exotic and fantastic applications of negative energy to "warp-drives,""time-machines," and "faster-than-light" travel, which now appear forbidden by fundamental laws of quantum mechanics in curved space-time [42-45] are not involved in the physical processes of this invention.

4.1.3 Thrust

[0129] By using Eq. (13) above, we can write the total lifting force on the trapped atoms in the case that $\alpha_A(k_L)=\alpha_B(k_L)$, by assuming that the average of the interatomic distance between any given atom and all others is ~D, as:

$$F_{gas}=N_A^2F_z^{AB}=\frac{4\pi^2}{c^3}N_A^2\left(\frac{I}{\lambda_L^2}\right)\frac{\alpha_A^2(k_L)}{D}g. \tag{37}$$

Let us now consider the total momentum acquired by the gas as it flows out of the atomic trap. Without getting into the details of the dynamics of such process, which certainly deserve further consideration, let us consider the acceleration of the atoms under the action of the lifting force, F_{gas}. In an inertial reference frame, the total acceleration undergone by the atoms would be F_{gas}/N_Am_A, but, in this non-inertial frame, the acceleration will be $(F_{gas}-N_Am_Ag)/N_Am_A=(F_{gas}/N_Am_A)-g$. For instance, if the lifting force is exactly equal to N_Am_Ag, the atoms will not accelerate with respect to the vehicle, but only hover, as we have seen in the above theoretical treatment. Therefore:

$$a_A=\frac{F_{gas}}{N_Am_A}-g=\left[\frac{4\pi^2}{c^3}N_A\left(\frac{I}{\lambda_L^2}\right)\frac{\alpha_A^2(k_L)}{(\bar{R}N_A^{1/3})m_A}-1\right]g. \tag{38}$$

At this rate, the average final velocity of the atoms will be $v_{A,\,fin}=\sqrt{2a_A(D/2)}$:

$$v_{A,fin}=\sqrt{\left(\frac{F_{gas}}{N_A m_A}-g\right)\overline{R}N_A^{1/3}}=\sqrt{\left[\frac{4\pi^2}{c^3}N_A\left(\frac{I}{\lambda_L^2}\right)\frac{\alpha_A^2(k_L)}{(\overline{R}N_A^{1/3})m_A}-1\right]g\overline{R}N_A^{1/3}}, \tag{39}$$

where, of course, the application considered in this invention considers only circumstances in which the radical is real.

[0130] It is also of interest to find an order of magnitude for the time required to leave the atomic trap defined as $\frac{1}{2}D=\frac{1}{2}a_A(\Delta t_A)^2$, which sets a lower minimum to the "reload" time of the thrust cycle:

$$\Delta t_A=\sqrt{\frac{D}{\alpha_A}}=\sqrt{\overline{R}N_A^{1/3}\frac{1}{\frac{F_{gas}}{N_A m_A}-g}}=\sqrt{\frac{\overline{R}N_A^{1/3}}{\left[\frac{4\pi^2}{c^3}N_A\left(\frac{I}{\lambda_L^2}\right)\frac{\alpha_A^2(k_L)}{(\overline{R}N_A^{1/3})m_A}-1\right]g}}, \tag{40}$$

[0131] From Eq. (39) we can immediately write the total momentum of the gas, ΔP_{gas}, for each cycle at it approaches the shock absorber:

$$\Delta P_{gas}=m_{gas}v_{A,fin}=N_A m_A\sqrt{\left(\frac{F_{gas}}{N_A m_A}-g\right)\overline{R}N_A^{1/3}}, \tag{41}$$

or

$$\Delta P_{gas}=N_A m_A\sqrt{\left[\frac{4\pi^2}{c^3}N_A\left(\frac{I}{\lambda_L^2}\right)\frac{\alpha_A^2(k_L)}{\overline{R}N_A^{1/3}m_A}-1\right]g\overline{R}N_A^{1/3}}, \tag{42}$$

Finally, let us estimate the total momentum transferred to the vehicle in the assumption that the impact of the gas against it is dissipative, that is, by assuming the complete conservation of momentum. In this case, the final speed change of the (gas+craft) system at the end of the n-th cycle will be given by $m_{gas}v_{A,fin}=(m_{gas}+M_{craft})\Delta v_n$. That is, by solving for Δv_n,

$$\Delta v_n=\frac{N_A m_A}{N_A m_A+M_{craft}}\sqrt{\left[\frac{4\pi^2}{c^3}N_A\left(\frac{I}{\lambda_L^2}\right)\frac{\alpha_A^2(k_L)}{\overline{R}N_A^{1/3}m_A}-1\right]g\overline{R}N_A^{1/3}}. \tag{43}$$

Although the acceleration of the vehicle varies over time during the cycle, it is useful to introduce an average acceleration over the cycle itself, we the understanding that the acceleration experienced by the atoms during their phase of acceleration may be different, and usually higher, than this value because of the cycle synchronization we discussed earlier. We write:

$$a_{craft}=\frac{\Delta v_n}{\Delta t_A+\Delta t_{Reload}}, \tag{44}$$

where Δt_{Reload} is the time required to prepare the new atomic cloud in the trap for the following cycle. This definition allows us to also define a nominal thrust for the engine as:

$$F_{engine}=(N_A m_A+M_{craft})a_{craft} \tag{45}$$

[0132] In the limit in which the reload time vanishes ($\Delta t_{Reload}\rightarrow 0$), of course we find, as expected:

$$a_{craft}\rightarrow\frac{\Delta v_n}{\Delta t_A}=\frac{N_A m_A}{N_A m_A+M_{craft}}a_A, \tag{46}$$

and

$$F_{engine}\rightarrow F_{gas}. \tag{47}$$

In this ideal limit, the hovering condition becomes that $F_{gas}=(N_A m_A+M_{craft})g$, that is, the total force on the atomic cloud must be equal to the weight of the entire vehicle. This is also the thrust that could be obtained if the atomic cloud could be trapped permanently within the chamber even as it acts on the vehicle via the action-reaction law.

4.1.4 Energy Considerations and Thruster Efficiency

[0133] The total radiation power utilized will be estimated as $W=18\ ID^2$, which corresponds to six laser triads (this is an upper estimate of the power needed as it is possible to induce gravitation-like behavior in a particular direction by using less power than this maximum). This yields:

$$W=18\ I\ N_A^{2/3}\overline{R}^2. \tag{48}$$

By solving this expression for the radiation flux I, we find:

$$I=\frac{W}{18N_A^{2/3}\overline{R}^2}. \tag{49}$$

This result allows us to obtain a realistic estimate of the relationship between the dynamics of the process and the total amount of energy needed. For instance, by substituting it into Eq. (38), we obtain the following useful expression:

$$a_A=\left[\frac{2\pi^2}{9c^3}\left(\frac{W\alpha_A^2(k_L)}{\overline{R}^3\lambda_L^2 m_A}\right)-1\right]g, \tag{50}$$

which exposes the interesting fact that, in the present approximation, the atomic acceleration as a function of the total laser power irradiated is independent of the number of atoms in the trap. By means of this equation, it is possible to write the total power needed to be focused onto the trap to achieve a hover, that is, a vanishing acceleration of the atoms with respect to the vehicle.

This corresponds to:

$$W_{hover} = \frac{9c^3}{2\pi^2}\frac{\overline{R}^3\lambda_L^2 m_A}{\alpha_A^2(k_L)}, \tag{51}$$

independently of g. For practical reasons, let us rewrite this result in units of Megawatts (MW) in terms of the wavelength in micrometers, λ(μm), of the average interatomic distance in units of Bohr radii, $\overline{R}/a_0$, and by involving the dimensionless polarizability factor as $\alpha_A(k_L)=\alpha_{nr}(k_L)\alpha_0$:

$$W_{hover} = 2.18\times10^9\frac{(\overline{R}/a_0)^3\lambda_L^2(\mu m)}{\alpha_{nr}^2(k_L)}\text{MW}, \tag{52}$$

Of course this result should be interpreted as allowing the atoms to be brought to a hover for a time no longer than the free-fall time given at Eq. (19) above unless an independent trapping approach is employed to hold the atoms at constant intermolecular distances, such as in an optical crystal.

Similarly for Eqs. (17), (18), (20), and (21):

$$v_{A,fin} = \sqrt{\left[\frac{2\pi^2}{9c^3}\left(\frac{W\alpha_A^2(k_L)}{\overline{R}^3\lambda_L^2 m_A}\right) - 1\right]g\overline{R}N_A^{1/3}}\ . \tag{39a}$$

$$\Delta t_A = \sqrt{\frac{\overline{R}N_a^{1/3}}{\frac{2\pi^2}{9c^3}\left(\frac{W\alpha_A^2(k_L)}{\overline{R}^3\lambda_L^2 m_A}\right) - 1}\frac{1}{g}}\ , \tag{40a}$$

$$\Delta P_{gas} = N_A m_A\sqrt{\left[\frac{2\pi^2}{9c^3}\left(\frac{W\alpha_A^2(k_L)}{\overline{R}^3\lambda_L^2 m_A}\right) - 1\right]g\overline{R}N_a^{1/3}}\ , \tag{42a}$$

$$\Delta v_n = \frac{N_A m_A}{N_A m_A + M_{craft}}\sqrt{\left[\frac{2\pi^2}{9c^3}\left(\frac{W\alpha_A^2(k_L)}{\overline{R}^3\lambda_L^2 m_A}\right) - 1\right]g\overline{R}N_a^{1/3}}\ , \tag{43a}$$

with the same restrictions on the duration of the impulse.

[0134] The ratio of thrust to radiation pressure: This dimensionless quantity compares the thrust obtained from the present thrust mechanism to the radiation pressure one would obtain by simply radiating away the power used instead to alter the intermolecular potential. Clearly one expects this quantity to be much larger than unity in order for the approach of this invention to be of practical use. Should that not be so, the case could be made that focusing the radiation onto an appropriately large sail would be a more efficient use of that energy. We have, in the ideal case in which the reload time vanishes,

$$\frac{F_{gas}}{P_{\text{rad}}} = \frac{1}{W/c}N_A m_A\left[\frac{2\pi^2}{9c^3}\left(\frac{W\alpha_A^2(k_L)}{\overline{R}^3\lambda_L^2 m_A}\right) - 1\right]g. \tag{53}$$

In the limit in which $a_A \gg g$, we have, simply:

$$\frac{F_{gas}}{P_{rad}} \to N_A\frac{2\pi^2}{9c^3}\left(\frac{\alpha_A^2(k_L)}{\overline{R}^3\lambda_L^2}\right)g. \tag{54}$$

Typical estimates of this quantity in interesting cases indeed indicate values much larger than unity for this quantity.

[0135] As in any other system that does not make use of the traditional expulsion of high speed gases to generate momentum (e.g. light sails), one must introduce new figures of merit. For instance, the typical concept of the specific impulse (the ratio of engine thrust to the weight of the material ejected in the unit time) [29], requires special attention in this case. Let us consider the case of a thrust system in which all radiation emitted by the high power lasers is permanently lost and radiated away from the spacecraft. Because of the mass-energy equivalency, this will correspond to a net mass-loss of the vehicle at a rate $M_{craft}=-W/c^2$. In the literature, the "photon rocket," which annihilates matter and antimatter to eject the corresponding radiation in a particular direction, is defined by theoreticians as the "perfect" rocket engine, because it yields the highest terminal speed at burnout for the same final to initial mass ratio [46]. However, whereas in that case the thrust is directly due to the reaction to the emitted photons, in our case the presence of the radiation in the chamber where the atoms are trapped is only a catalyst to create the thrust upon the vehicle. Thus, since the origin of the thrust in our case is not the lost radiation one could just as convincingly argue that the specific impulse becomes undefined in this case.

[0136] From the quantitative standpoint, therein lies the great interest of the approach of the present invention. That is, for the same amount of energy expended, the thrust derived is much higher than that obtainable in the photon rocket [46]. At the same time, this high thrust does not come at the price of an impulsive propulsive system as in chemical engines, but one that can actually provide high accelerations over almost indefinite periods of time [47].

[0137] Thruster efficiency: This is the ratio of the kinetic energy of the atomic cloud as it emerges from the trap to the total energy used during the acceleration of the gas. By using the above expression for the final velocity of the atoms, this quantity can be written as:

$$\begin{aligned}\frac{\frac{1}{2}N_A m_A v_{A,fin}^2}{\Delta t_A}\frac{1}{W} &= \frac{N_A m_A}{\sqrt{2}}\alpha_A^{3/2}D^{1/2}\frac{1}{W}\\ &= \frac{N_A m_A}{\sqrt{2}}\left[\frac{2\pi^2}{9c^3}\left(\frac{W\alpha_A^2(k_L)}{\overline{R}^3\lambda_L^2 m_A}\right) - 1\right]^{3/2}\end{aligned} \tag{55}$$

-continued

$$\frac{g^{3/2}\overline{R}^{1/2}N_A^{1/6}}{W}.$$

In the same limit $a_A >> g$ we find:

$$\frac{\frac{1}{2}N_A m_A v_{A,fin}^2}{\Delta t_A}\frac{1}{W} \to \left(\frac{W}{2m_A}\right)^{1/2}\frac{1}{\overline{R}^4}\left[\left(\frac{2\pi^2}{9c^3}\frac{\alpha_A^2(k_L)}{\lambda_L^2}\right)\right]^{3/2} g^{3/2}N_A^{7/6} \tag{56}$$

Special care must be exercised in interpreting the results obtained by extrapolating this equation to values of the efficiency that exceed unity, since this may be an indication that other phenomena are becoming important. Furthermore, we have here neglected other forms of energy that are also required, such as, for instance, the radiation required to excite the atoms if Rydberg states are used and the energy used for the initial trapping. As shall become clear in the examples below, the present invention represents a useful and revolutionary technological innovation even in regimes where the thruster efficiency is below unity.

4.2 Start-Up, Maneuvers, Cruise, and Turn-Off

[0138] The fundamental, basic physics principle of this invention is the distortion of the field of a dipole in an accelerated reference frame and the resulting "lifting" dipole-dipole force. In order for this principle to operate, an acceleration must be induced upon the dipole system before the thrust on it can appear. In other words, it is not possible to turn on the thrusting system of this invention from a cruise phase (approximately at constant speed). Because of the Principle of Equivalence, such initial acceleration can be provided by means of any combination of two different, but equivalent mechanisms. In the former, the acceleration is due to gravitation, which provided the initial motivation for the present invention.

[0139] Because of the Principle of Equivalence, the effect of a gravitational field is indistinguishable from that a uniformly accelerated reference frame (if we neglect the Linet term mentioned in our comment of Eq. (3) and in Ref. [9] and quantitatively unimportant to this invention). Therefore, the trigger acceleration can be provided by the gravitational field as a supported (not freely falling system) rests, for instance on the ground. Because of the Principle of Equivalence, if the spacecraft is freely falling its acceleration exactly cancels that of the gravitational field and it becomes indistinguishable from a laboratory at a large distance from any other mass, to the extent that its size is relatively small.

[0140] Because of the mechanism of this invention, if the spacecraft is resting on the ground, upon laser turn-on, the intermolecular forces will be distorted in such a way as to cause an upward acceleration of the atoms, which can be made, for instance, as large as needed for the spacecraft to hover. On the other hand, if the spacecraft is in outer space at a large distance from any other celestial objects, such as in interplanetary flight, the initial acceleration must be provided by independent means. This can be achieved in a variety of ways. For instance: (1) an initial forward acceleration can be produced by a traditional thrusting system until the first few cycles of the present engine are produced—in this case the traditional thruster becomes an "ignition system" for the present invention; (2) the entire craft can be put into rotation around an appropriate axis, thus reproducing Einstein's rotating disk [46]; (3) on re-entry, a spacecraft undergoes an aerodynamic deceleration that can be used as the trigger.

[0141] The same above sample triggers can be also used for maneuvering, that is, to create an initial acceleration in a direction different than that of the present motion. This can be very useful to transform the present invention in a vehicle for motion near the surface of the earth. A slight acceleration parallel to the ground could be maintained with relatively little radiative energy while a larger amount of radiation would keep the vehicle hovering safely. A slow decrease in the power of the lasers, or any other change in the parameters determining the thrust, such as intermolecular distance or total atom number, would result in a decreased acceleration of the atoms and, thus, of the vehicle. In order to bring the vehicle to a cruise (constant speed) the only necessary action is to turn off the high power lasers, or, alternatively, to empty the atomic trap system of atoms. The acceleration upon the vehicle would then immediately stop along with the dipole-dipole field distortion, thus bringing the craft to constant velocity. Finally, a trigger acceleration can be exercised in a direction opposite to the instantaneous velocity to start an opposite thrust that brings the vehicle to a deceleration and to a new hover.

[0142] Although the notion of flying by exploiting an initial acceleration may appear counterintuitive, it must be said that a similar transition also challenges some pilots at the beginning of their training. In fact, flying through air requires speed to create lift from the airfoils and a slower and slower aircraft is unable to keep altitude. This concept requires constant training in beginning pilots today to fight the instinct to "pitch-up" to regain lost altitude on landing, as that can result in a stall. Similarly, the lifting mechanism of this invention requires the user to become sensitive to acceleration—as opposed to speed—and to the notion that an accelerating vehicle can maintain thrust, whereas one at constant speed will loose thrust.

4.3 Numerical Examples

[0143] The following numerical examples were generated by making use of a Mathematica notebook [13] which encoded the equations discussed in this document. The design of any of the missions below could (and does) take many years of study, but the goal of this section is to provide the information necessary to appreciate the fact that the physics of this invention leads to realistic engineering demands. In other words, whereas relativistic travel is usually assumed to require harnessing absolutely fantastic amounts of energy, the method below yields numbers that can be realistically contemplated to be put into feasible design with presently existing technologies.

[0144] Any space mission must start with some requirements, which are to some extent arbitrary or descend from other constraints. In the examples below, we assumed that all travel takes place at a given spacecraft acceleration $a/g<1$. This is to contrast the approach to space travel of this invention with typical designs, which either impose high (several g) accelerations for a short periods of time (chemical rockets) or provide very low thrust over very long periods of time, such in the case of ion propulsion, in which the author of this invention has been directly involved [47]. Let us recall that the present method requires that the intermolecular potential be distorted by the acceleration of the vehicle (or by gravitation if the vehicle is held at rest, as required by the Principle of Equivalence). Therefore, the acceleration of the vehicle due to the impact of a gas cloud into the plate determines the efficiency of the atomic acceleration in the following cycle.

[0145] According to the present method, space travel is most efficient when it is significantly accelerated. This is not a drawback, since, for instance, it is now understood that the dangers to human health of very long periods of weightlessness are significant. The data presented herein were thus produced by assuming a spacecraft acceleration a=g and then by demanding self-consistency, that is, by requiring that the spacecraft acceleration produced by any given gas cloud also be equal to g. Even with this requirement, the choices shows below represent only one of many possibilities chosen for their being realistic from the engineering standpoint or for their being representative of a contrast between the present invention and present-day technology. For simplicity, the reload time was everywhere assumed to be negligible.

[0146] 4.2.1 A Robotic Low-Thrust Delivery System to the planet Mars

PROPULSIVE SYSTEM SPECIFICATIONS

$W = 35.0$ MW $= 3.5 \times 10^{14}$ erg/s
$\alpha_{nr} = 1 \times 10^{6}$ (quasi-resonant response)
$n = 1$ (ground state)
$\overline{R} = 5\ a_0$
$\lambda_L = 1000$ Å (quasi-Lyman-α transition)
$M_{craft} = 10^{7}$ g = 10 metric tons
$N_A = 4.66 \times 10^{26}$
$m_{gas} = 7.86 \times 10^{3}$ g = 0.786 kg
Chamber Size D = 20.3 cm
$W_{hover}(a_{craft} = 1\ g) = 2.7$ kW
$a_A = 1.28 \times 10^{5}$ cm/s^2
$v_{A,fin} = 2.29 \times 10^{3}$ cm/s
$\Delta t_A = 1.79 \times 10^{-2}$ s
$t_{ff} = 60.6 \times 10$ s
Efficiency $= 3.3 \times 10^{-4}$
Δv_{craft}/cycle = 2.62 cm/s
$a_{craft} = 10$ cm/s^2 (self-consistent)
Thrust = 10.0 kN
(Thrust/Total Radiation Pressure) $= 8.57 \times 10^{4}$
Earth-Mars distance at opposition (Aug. 27, 2003) 55.758×10^{11} cm [48]
Total Travel time $= 1.52 \times 10^{6}$ s = 17 d 13 h 17 min 46.7 s
Maximum Speed (reached after $= 6 \times 10^{5}$ s) $= 6.0 \times 10^{6}$ cm/s = 60 km/s
Maximum Kinetic Energy (reached after 6×10^{5} s) $= 1.8 \times 10^{20}$ erg
Total Energy Radiated (up to maximum speed) $= 2.1 \times 10^{20}$ erg
Time to Mach 1 (Speed of Sound $v_{sound} \simeq 331$ m/s) $= 3.31 \times 10^{3}$ s
Time to Clear Low-Earth-Orbit (400 km) $= 2.8 \times 10^{3}$ s = 47.1 min
Time to Clear the Earth-Moon System (4×10^{5} km) $= 8.94 \times 10^{4}$ s = 1 d 50.7 min

Comments

[0147] In order to obtain the above estimates for the transfer to Mars, we greatly simplified the problem by neglecting the gravitational force of all objects involved, including the Sun, the Earth, and Mars. Of course, the gravitational force of the Earth is actually an integral part of the engine start-up mechanism of the present invention and in that sense its effect is accounted for here. However, since the gravitational field of all near-by objects is cleared in a matter of minutes, its influence on the dynamics of flight was neglected in these order-of-magnitude calculations.

[0148] It is useful to make come comparisons between the specifications of the system above and present-day delivery systems, since that helps elucidate the great technological relevance of the present invention. A typical choice for such recent space missions to Mars as the ones of Spirit and Opportunity is the Boeing Delta II 7925 or 7925H (the letter H indicates the more powerful high performance version) [49]. In its common configuration, the RS-27A engine of the Delta II first stage, along with the additional nine strap-on solid rocket motors, generates approximately 8.9×10^{5} Newtons of thrust, which are necessary to lift the total "wet" (fueled up) vehicle mass of 285,228 kg off the launch pad. This thrust is almost one order of magnitude larger than that of the engine described in this document until the Main Engine Cut Off (MECO), approximately 265 s after lift-off. The thrust of the following stages is smaller, with the thrust of the third stage at approximately 6.6×10^{4} N.

[0149] Since the initial phase of ascent takes place under the thrust of a chemical rocket, not surprisingly we see that both in the case of Spirit and Opportunity the Delta II vehicle passes through Mach 1 in a much shorter time than the one propelled by the engine of this invention (32.4 s and 29.6 s, respectively) [50]. However, very importantly, by a very good approximation, the entire launch mass of 10 metric tons (10^{4} kg) is propelled towards Mars in the present case whereas, in the traditional approach, only 1.070×10^{3} kg out of the initial 285,228 kg represent the useful remaining payload. Of this surviving mass, only 533 kg actually lands on Mars in the traditional case, since approximately 250 kg are allocated to the cruise stage alone.

[0150] Finally, the entire vehicle propelled by the engine of this invention arrives at Mars in a matter of less than twenty days, whereas both Spirit and Opportunity extremely reduced in mass, arrive approximately six months later. It is very significant that the gravitational field of Mars, added to the deceleration of the spacecraft, makes the process of landing much slower and completely safe, in contrast with what NASA/JPL itself defines as the "six minutes of terror," during which the vehicle must be slowed down from 12,000 miles/hr to 0 miles/hr corresponding to average accelerations a_{craft}~1.5 g, although during the last drop the acceleration can reach ~40 g [51]! In contrast, since the present vehicle is approaching Mars decelerating at $a_{craft}=10$ cm/s^2, the same process can be executed within over a week-long time and distributed over a path at uniform acceleration that concludes with the entire vehicle safely hovering at a desired height above the ground of the planet.

[0151] An obvious question when carrying out a comparison between traditional propulsion technology and the mechanism described in this document concerns the feasibility of generating ~10^{-3}-10^{-2} s laser pulses requiring over ~30 MWe (that is, MW of electric power) on board of a space vehicle. Is this possible? Has this possibility ever been carefully considered in the past? The answer to such legitimate question is that the study of high power, high efficiency, low-mass nuclear reactors for use in space applications is actually extremely advanced, although political and public opinion considerations tend to hide the enormous amount of available information away from the non-technical readership. A small selection of such literature, which cannot be cited here even partially because of its sheer size, can be found at [52] and References therein. General historic motivations behind this technical effort are to be found in the large energy needs of any hypothetical "Star Wars" defense system, in research into the possibility of interstellar travel, and, finally, in the recently renewed commitment to the human colonization of space made by the President of the United States and NASA. What is important to the evaluation of the engineering feasibility of the present invention is that, typically, reactor power densities in the order of ~1-5 kW/kg are surely possible, with very wide variations in either direction of that estimate depending on the specifics of design and shielding requirements. However, the figure above is sufficient to make the case that, by means of presently existing technology, it is absolutely appropriate to consider a 10 ton-vehicle carrying a reactor able to produce pulses in the 30 MWe range in space.

[0152] The implications for human flight to Mars by means of the present invention are very significant as well. At present, the delivery of several tens of tons of payload to Mars is contemplated to take approximately 180 days, with a typical mission duration for the crew of 2-3 years. The overwhelming majority of the mission duration would be spent in complete weightless conditions during transit and possibly under exposure from harmful radiation. In addition, the prospects of recovery in case of an even minor malfunction are dire not to speak of a major accident of the type that occurred on Apollo 13. Being able to reduce the travel time to a matter of days, while transporting the crew under at least partial gravitational conditions completely changes the prospects for successful colonization of the Red Planet as well as the potential for a rescue mission should that be necessary [53].

[0153] 4.2.4 A Thrust System for Safe, Low-Speed, Near-Earth Human Transportation

PROPULSIVE SYSTEM SPECIFICATIONS

$a_A = 6.29 \times 10^6$ cm/s^2
$v_{A,fin} = 1.61 \times 10^4$ cm/s
$\Delta t_A = 2.55 \times 10^{-3}$ s
$t_{ff} = 60.6 \times 10$ s
Efficiency = 11.2%
Δv_{craft}/cycle = 18.35 cm/s
$a_{craft} = g/2$ (self-consistent)
Thrust = 0.491 MN
(Thrust/Total Radiation Pressure) = 4.20×10^6
Earth-to-LEO distance = 4×10^7 cm
Total Travel time = 4.97×10^2 s = 8 min 17.5 s
Maximum Speed (reached at midpoint) = 1.22×10^5 cm/s = 1.22 km/s
Maximum Kinetic Energy (reached at midpoint) = 7.44×10^{16} erg
Total Energy Radiated (at midpoint) = 1.74×10^{17} erg
Time to Mach 1 (Speed of Sound $v_{sound} \simeq 331$ m/s) = 67.5 s

Note:
The quantities not repeated are unchanged with respect to the previous example.

Comments

[0154] The interest of this particular case lies not in the acceleration, which is less favorable than by means of already available technology, but in its ability to deliver the entire payload to high altitude at rest. This allows us to consider an entirely new philosophy or air transportation and space travel around the Earth (or other planets). Whereas the key objective to reach extreme altitudes with ordinary technologies must be the achievement of high speeds, as that is the only strategy which allows the vehicle to be injected into a permanent orbit, in the case of the present invention it is possible to deliver a payload to a high hovering altitude without requiring orbital speeds.

[0155] Similarly, the descent maneuver of our vehicle does not require the fiery but unavoidable re-entry of typical deorbiting, thus avoiding the accompanying extreme heating and grave dangers to the crew, as in the recent Columbia tragedy. In fact, the vehicle of this example would not reach speeds higher than Mach 4 before decelerating to its hovering point, as opposed to the orbital re-entry speed of the Shuttle of approximately Mach 24. This achievement would represent nothing less than a revolution in aerospace technology. Interestingly, the present approach also lends itself to being phased-in as it replaces traditional propulsion technologies. In other words, it is conceivable that a system of lower thrust, unable by itself to attain a complete hover, could be placed into service for the only purpose to provide additional breaking in an emergency at those speeds that make parachute deployment am impossible option.

I claim:

1. A method comprising:

generating a force by subjecting a plurality of confined particles to a trigger acceleration, wherein said force exhibits short-range interactions; and

exposing said particles to an amount of electromagnetic radiation that is sufficient to induce said force to:

(iii) exhibit long-range interactions; and

(iv) cause said particles to either accelerate or hover.

2. The method of claim 1 wherein said particles are characterized by a polarizability, and wherein the method further comprises increasing said polarizability of said confined particles.

3. The method of claim 2 wherein said electromagnetic radiation has a near-resonance wavelength.

4. The method of claim 2 wherein said method further comprises creating Rydberg atoms from said particles.

5. The method of claim 1 wherein said electromagnetic radiation is delivered to said particles as a plurality of pulses, wherein each pulse lasts for an amount of time that is shorter than any atomic transition that said particles undergo.

6. The method of claim 1 further comprising transferring at least some of the momentum of said particles to a vehicle.

7. The method of claim 1 further comprising impacting said particles against a surface, wherein, during impact, at least some of the momentum of said particles is transferred to said surface.

8. The method of claim 1 further comprising impacting said particles against a piston.

9. The method of claim 1 wherein a minimum sufficient amount of electromagnetic radiation is given by the expression:

$$W = 2.18 \times 10^9 \frac{(\bar{R}/a_0)^3 \lambda_L^2}{\alpha_{nr}^2(k_L)}$$

wherein: W :is power, in Megawatts;

$\bar{R}/a_0$: is the average interatomic distance, in Bohr radii;

λ_L: is the wavelength of the electromagnetic radiation, in micrometers;

α_{nr}: is a factor (dimensionless) by which the static polarizability of a particle is increased at near resonance; and

k_L: is the laser-light wave number (dimensionless).

10. The method of claim 1 wherein said trigger acceleration is a gravitational field.

11. The method of claim 6 wherein said trigger acceleration is provided a method selected from the group consisting of: supporting said vehicle in a gravitational field, firing a chemical rocket, rotating said vehicle.

12. The method of claim 1 wherein said confined particles are confined in an atomic trap.

13. The method of claim 12 wherein said atomic trap is aboard a vehicle.

14. A method comprising:

confining a plurality of particles;

subjecting said particles to a trigger acceleration;

exposing said particles to an amount of electromagnetic radiation sufficient to:

(i) induce an inter-particle force that arises between said particles to exhibit long-range interactions; and

(ii) cause said particles to either accelerate or hover.

15. The method of claim 14 wherein the operation of confining further comprises confining said particles in an atomic trap.

16. The method of claim 14 wherein the operation of confining further comprises confining said particles aboard a vehicle.

17. The method of claim 16 further comprising impacting said particles against a surface of said vehicle.

18. The method of claim 14 further comprising increasing the polarizability of said particles above a static value.

19. The method of claim 14 further comprising transferring at least a portion of a momentum possessed by said particles to a vehicle.

20. A method comprising generating a lifting force by modifying the interaction potential of particles in an accelerated reference frame.

21. The method of claim 20 wherein the operation of modifying the interaction potential comprises causing the interaction potential to exhibit long-range interactions.

22. The method of claim 20 wherein the operation of modifying the interaction potential comprises exposing said particles to an amount of electromagnetic radiation that is sufficient to cause said particles to either accelerate or hover.

23. A method comprising developing thrust to propel a vehicle by:

(1) confining a plurality of particles within said vehicle;

(2) exposing said particles to an amount of electromagnetic radiation that is sufficient to:

(i) induce an inter-particle force that arises between said particles to exhibit long-range interactions; and

(ii) cause said particles to accelerate in a first direction due to said force; and

(3) impacting the accelerating particles against a surface that is coupled to said vehicle, wherein:

said thrust that is developed propels said vehicle in said first direction.

24. The method of claim 23 wherein the operation of impacting further comprises moving a piston, and wherein said piston is coupled to said vehicle.

25. The method of claim 24 wherein, before being exposed to said electromagnetic radiation, said particles are confined within a trap, and wherein the operation of impacting further comprises recycling to said trap particles that have moved said piston.

26. A method comprising developing thrust to propel a vehicle by:

generating asymmetry in a plurality of electric fields, wherein said electric fields surround a plurality of confined particles;

exposing said confined particles to real photons, wherein said real photons deliver an amount of energy to said confined particles that is sufficient to cause an inter-particle force that results from said asymmetry to:

exhibit long-range interactions; and

cause said particles to either accelerate.

* * * * *

June 9, 1925. **1,540,998**

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 1

Fig. 1.

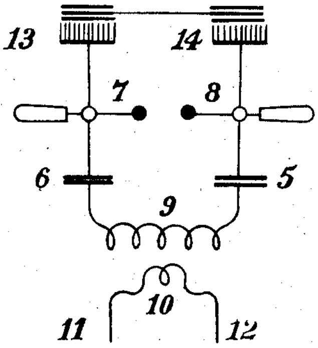
13
14
7
8
6
5
9
10
11
12

Fig. 2.

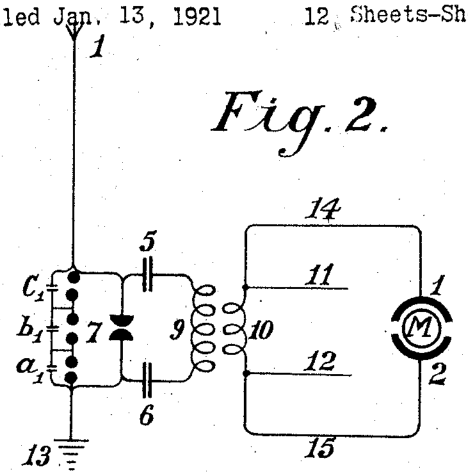
1
14
5
11
C_1
b_1
7
9
10
1
a_1
12
2
6
15
13

Fig. 3.

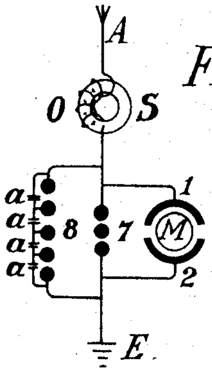
A
O
S
1
a
a
8
7
a
a
2
E.

Fig. 4.

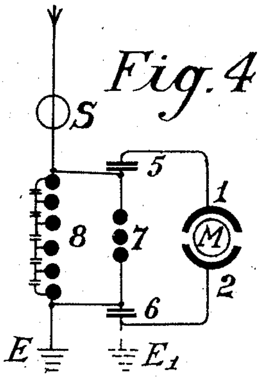
S
5
1
8
7
2
6
E
E_1

Fig. 5.

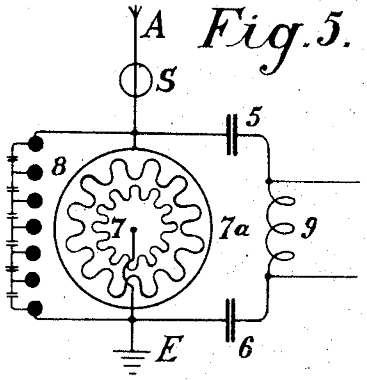
A
S
5
8
7
7a
9
E
6

Fig. 6

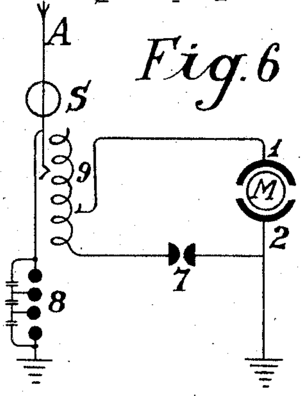

Inventor
Hermann Plauson

by Knight Bros
attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 2

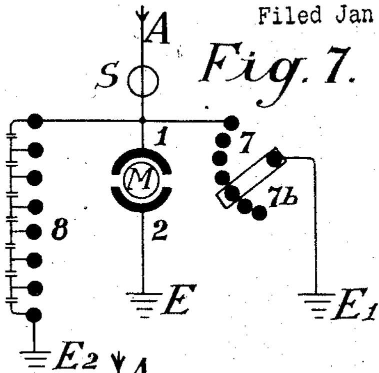
A
S
1
M
2
7
7b
8
E
E1
E2

Fig. 7.

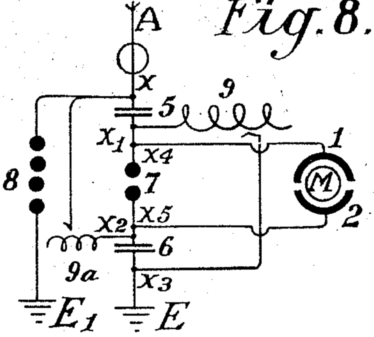
A
X
5
9
X1
X4
1
M
7
8
X5
2
X2
6
9a
X3
E1
E

Fig. 8.

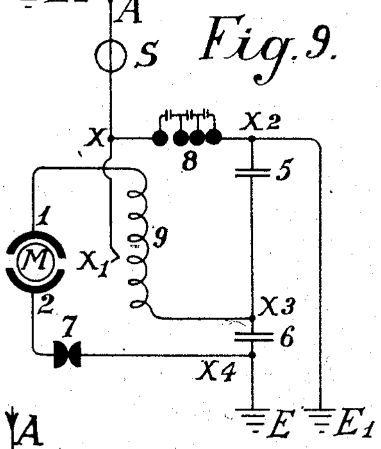
A
S
X
8
X2
5
1
M
X1
9
2
X3
7
6
X4
E
E1

Fig. 9.

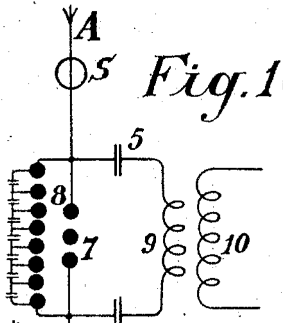

Fig. 10.

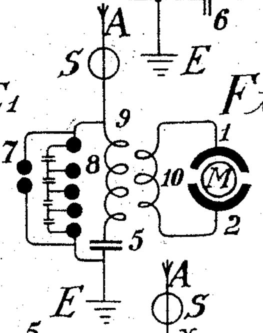
A
S
E
9
7
8
10
1
M
2
5
E

Fig. 11.

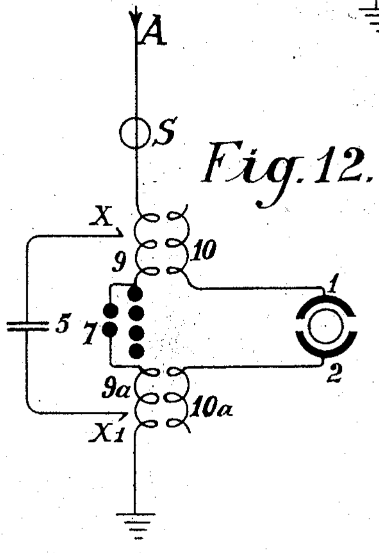
A
S
X
9
10
1
5
7
2
9a
10a
X1

Fig. 12.

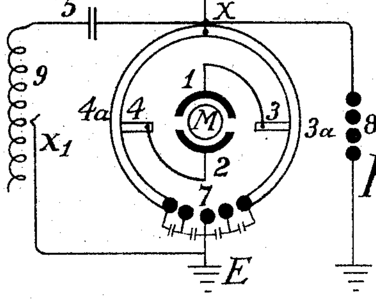
A
S
X
5
9
1
M
4a
4
3
3a
8
X1
2
7
E

Fig. 13.

Inventor

Normann Plauson

by Knight Bros

attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 3

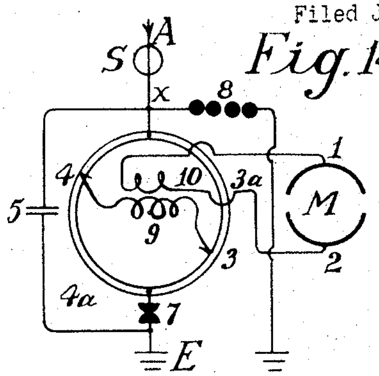
Fig. 14.
S
A
X
8
1
M
2
4
10
3a
5
9
3
4a
7
E

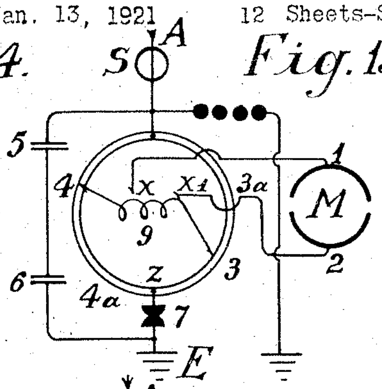
Fig. 15.
S
A
5
1
M
4
X
X1
3a
9
2
6
Z
3
4a
7
E

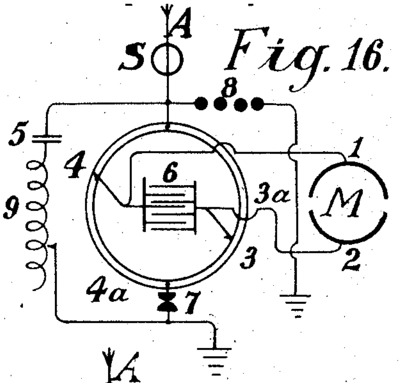
Fig. 16.
S
A
8
5
1
4
6
3a
M
9
2
3
4a
7

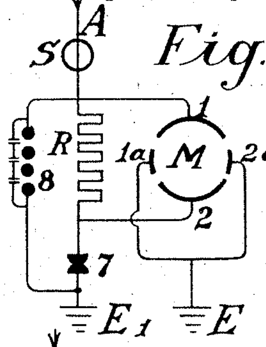
Fig. 17.
S
A
1
R
1a
M
2a
8
2
7
E1
E

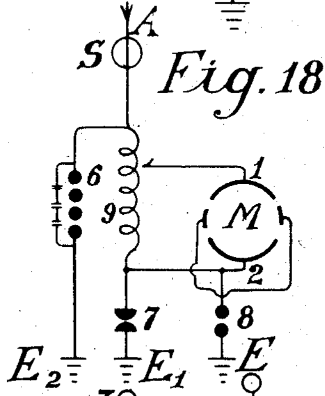
Fig. 18.
S
A
1
6
9
M
2
7
8
E2
E1
E

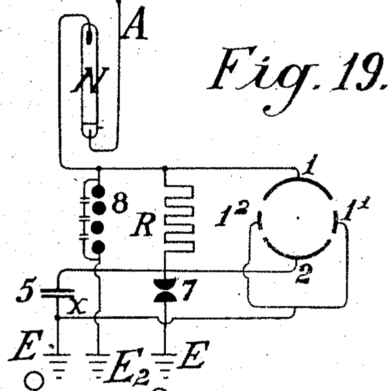
Fig. 19.
A
N
1
8
R
1²
1¹
2
5
X
7
E
E2
E

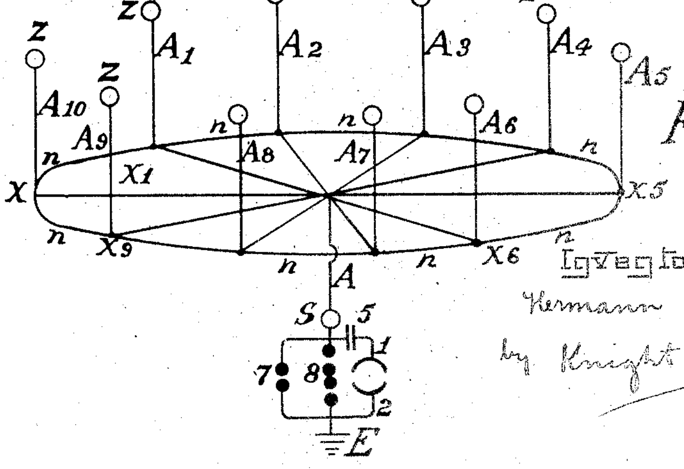
Fig. 20.
Z
A1
A2
A3
A4
A5
Z
A10
Z
A9
n
A8
n
A7
A6
n
X1
X
X5
n
X9
n
A
n
X6
S
5
1
7
8
2
E

Inventor
Hermann Plauson
by Knight Bros
Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 4

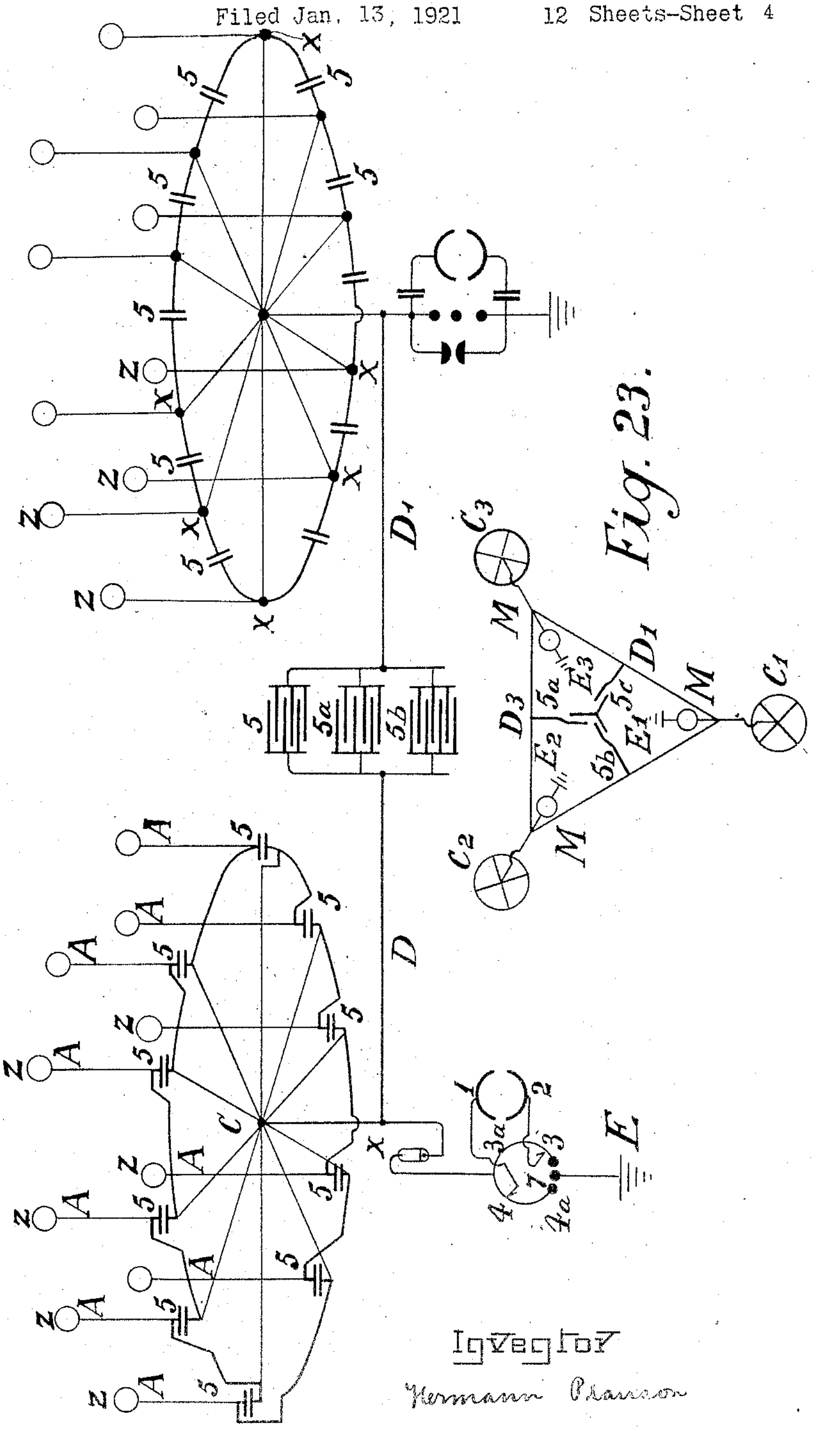
5
x
Z
A
C
D
D1
5a
5b
5c
D3
E1
E2
E3
M
C1
C2
C3
1
2
3
3a
4
4a
7
E

Fig. 22.

Fig. 23.

Fig. 21.

Inventor

Hermann Plauson

by Knight Bros

Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 5

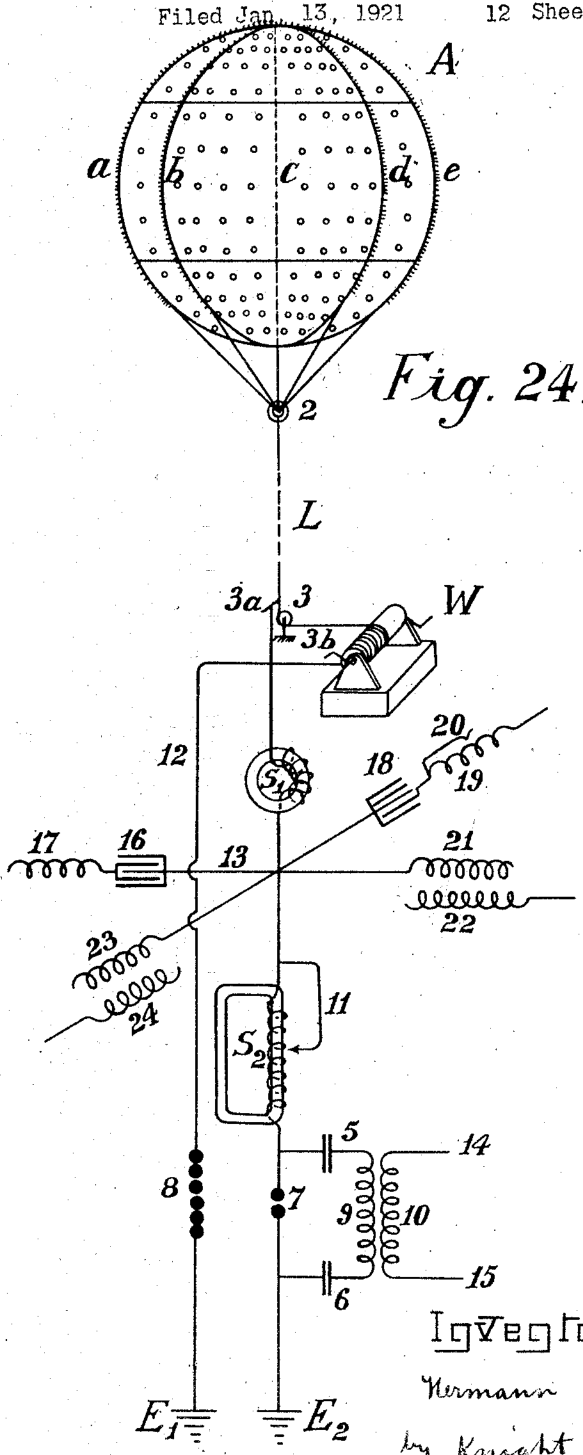
A
a
b
c
d
e
2
L
3a
3
3b
W
12
S1
18
20
19
17
16
13
21
22
23
24
11
S2
5
14
8
7
9
10
15
6
E1
E2

Fig. 24.

Inventor

Hermann Plauson

by Knight Bros

Attorneys

June 9, 1925. **1,540,998**

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 6

Fig. 25.

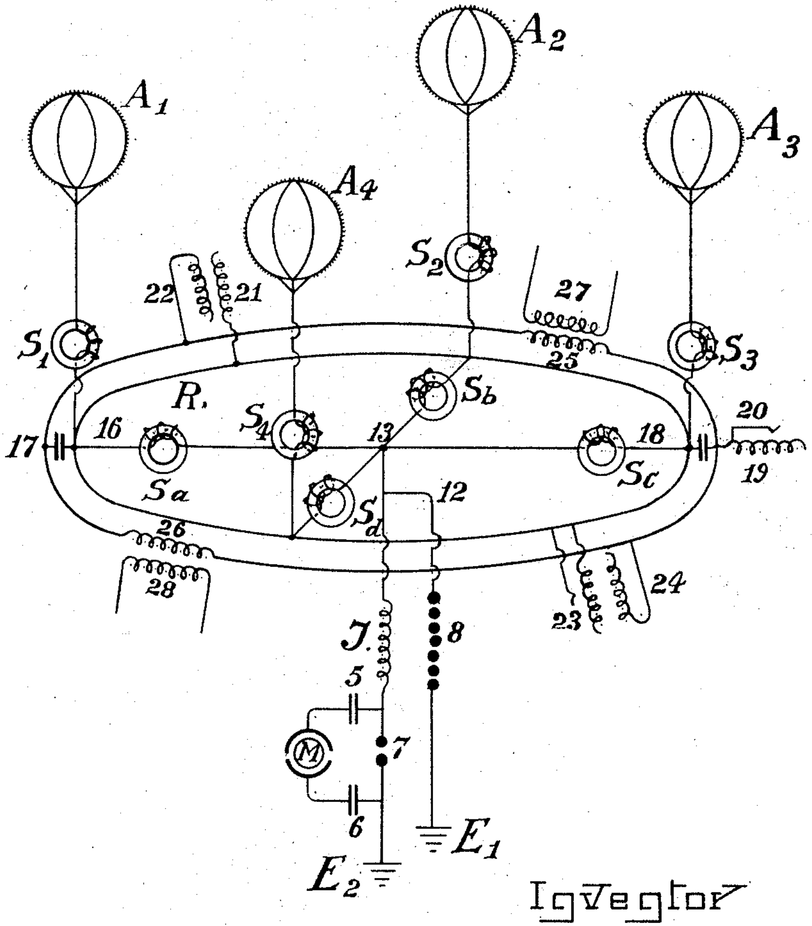
A1
A2
A3
A4
S1
S2
S3
S4
Sa
Sb
Sc
Sd
R.
12
13
16
17
18
19
20
21
22
23
24
25
26
27
28
J.
5
6
7
8
M
E1
E2

Inventor

Hermann Plauson

by

Knight Bros

attorneys

June 9, 1925. **1,540,998**

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 7

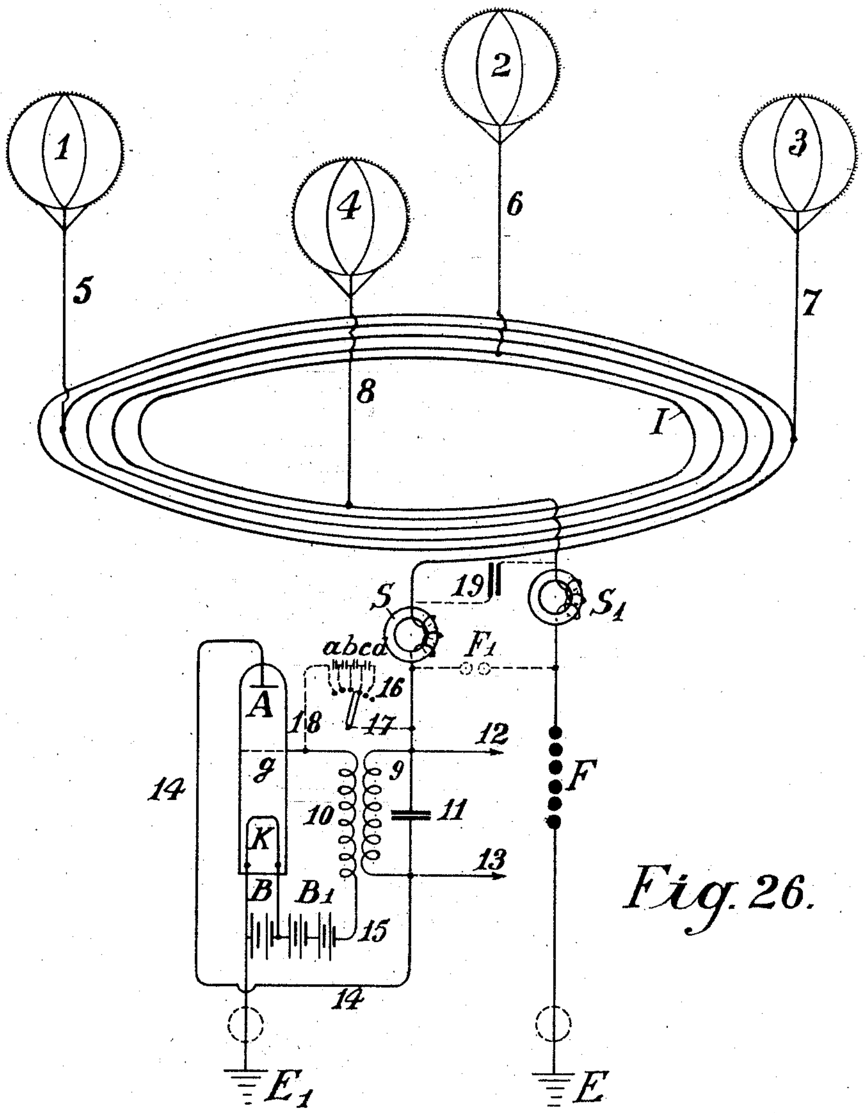
2
1
3
4
6
5
7
8
I
S
19
S1
abcd
F1
16
A
18
17
12
14
g
9
F
10
11
K
13
B
B1
15
14
E1
E

Fig. 26.

Inventor
Hermann Plauson
by Knight Bros
Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921. 12 Sheets-Sheet 8

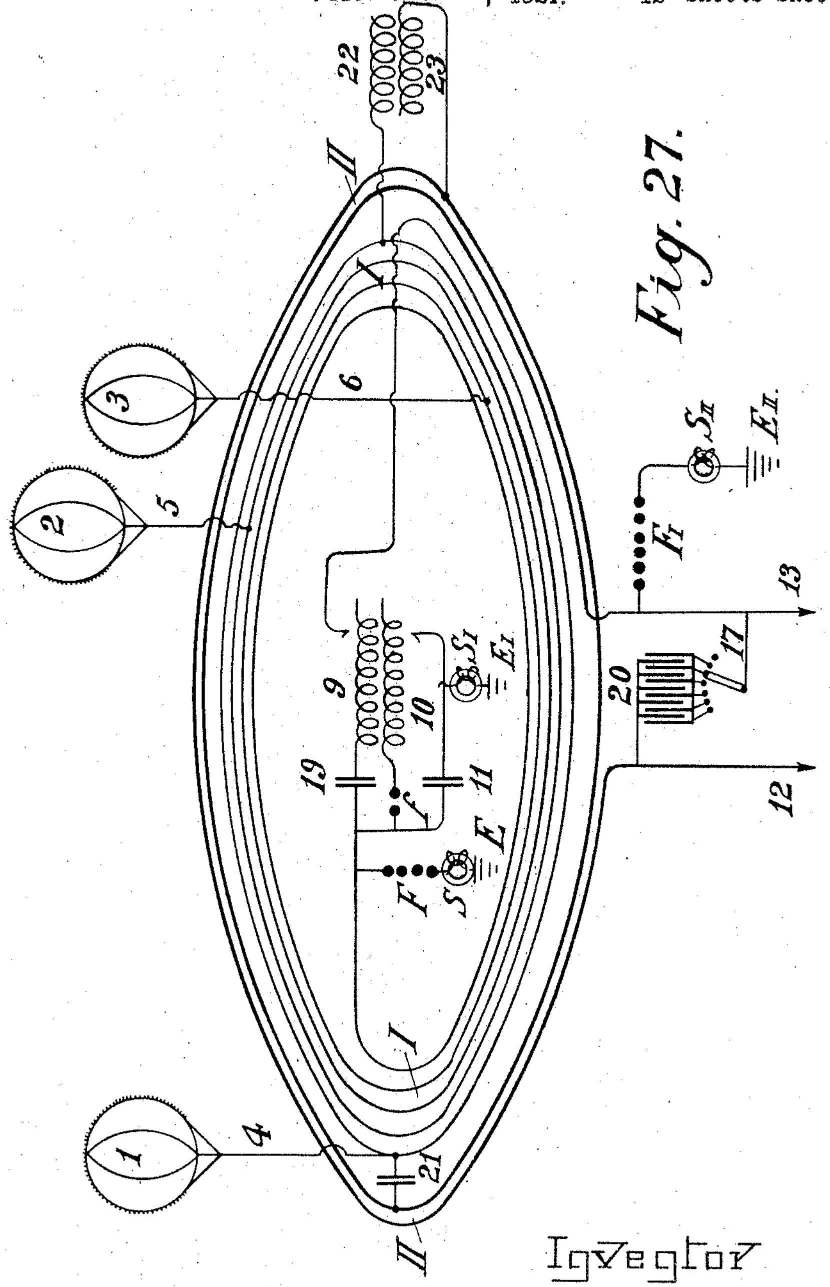
22
23
II
I
6
3
2
5
9
10
S_I
E_I
19
11
F
S
E
20
17
13
12
F_I
S_{II}
E_{II}
1
4
21
II
I

Fig. 27.

Inventor

Hermann Plauson

by Knight Bros

Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 9

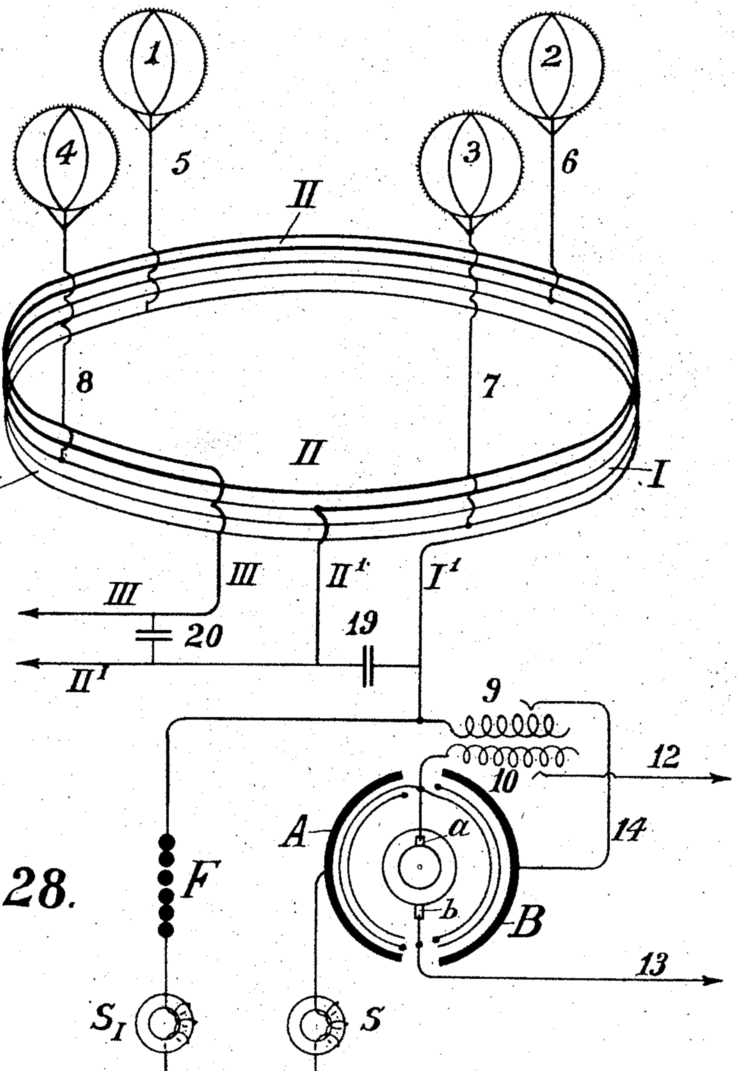
1
2
4
5
3
6
II
8
7
II
I
I
III
II^1
I^1
III
20
19
II^1
9
12
10
A
a
14
B
b
13
F
S_I
S
E_I
E

Fig. 28.

Inventor
Hermann Plauson
by
Knight Bros
attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921. 12 Sheets-Sheet 10

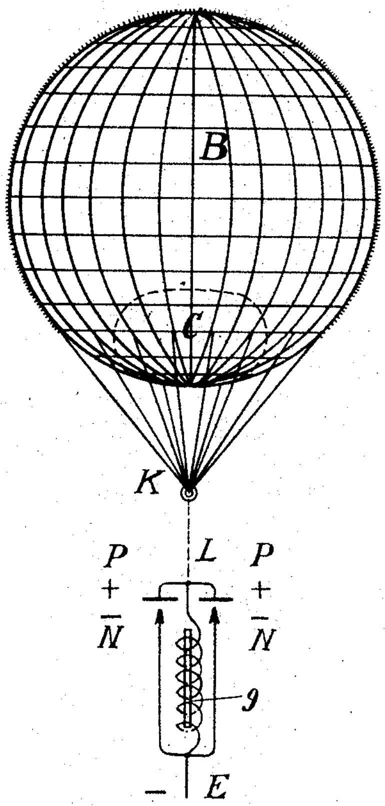
B
C
K
P
L
P
+
+
N
N
9
−
E

Fig. 29.

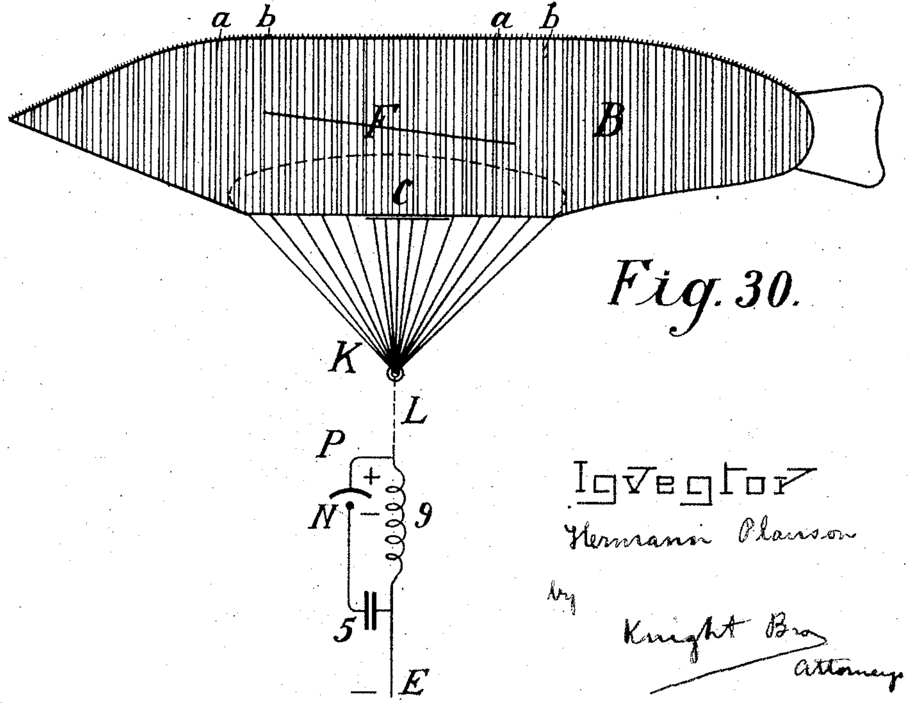
a
b
a
b
F
B
C
K
L
P
+
N
−
9
5
−
E

Fig. 30.

Inventor
Hermann Plauson
by
Knight Bros
Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 11

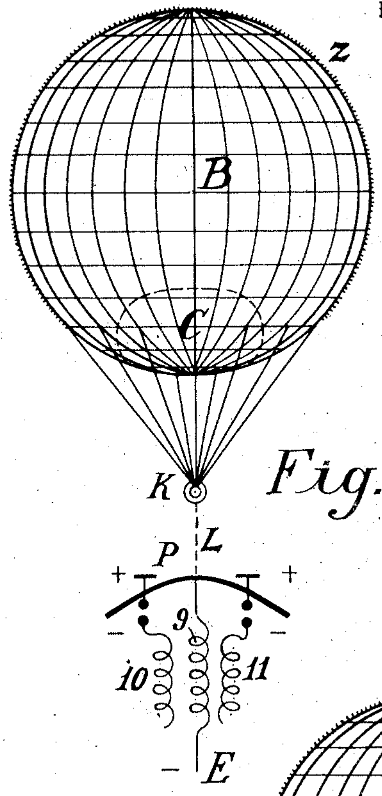
z
B
C
K
P
L
+
+
–
–
9
10
11
–
E

Fig. 31.

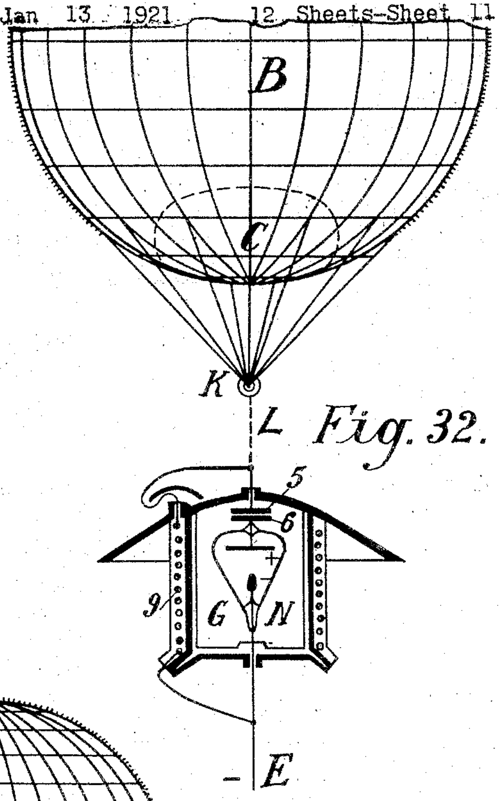
z
B
C
K
L
5
6
+
–
9
G
N
–
E

Fig. 32.

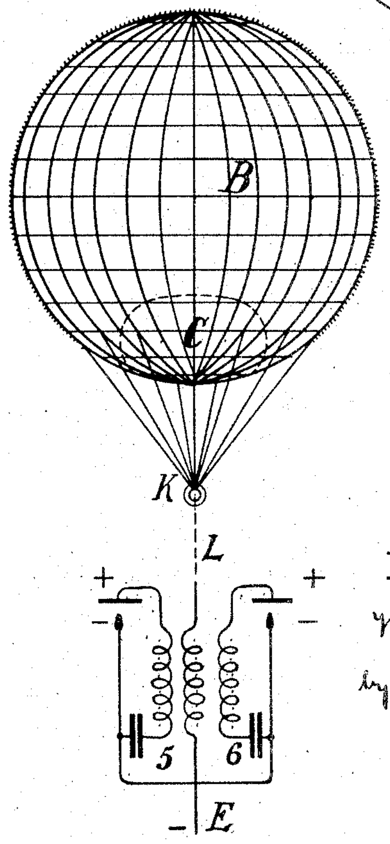

Fig. 33.

Inventor
Hermann Plauson
by Knight Bros
attorneys

June 9, 1925.

1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921

12 Sheets-Sheet 12

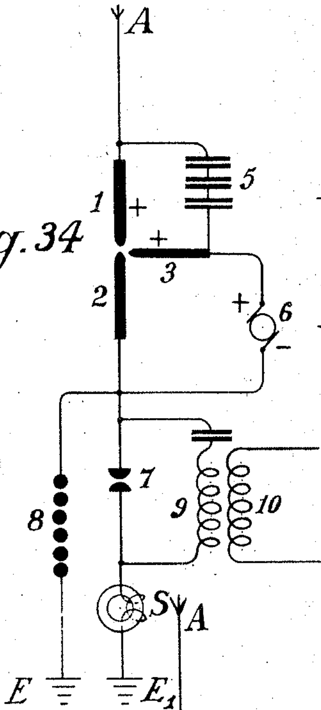
A
1
+
5
+
3
2
+
6
–
8
7
9
10
S
E
E1

Fig. 34

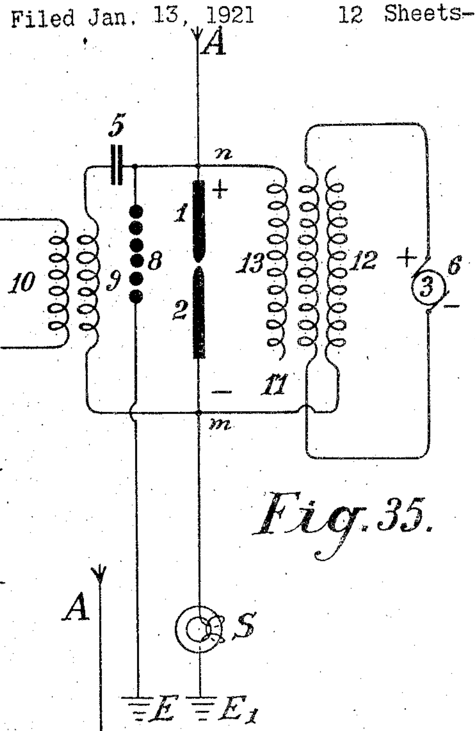
A
5
n
+
1
8
10
9
13
12
+
3
6
–
2
–
11
m
S
E
E1

Fig. 35.

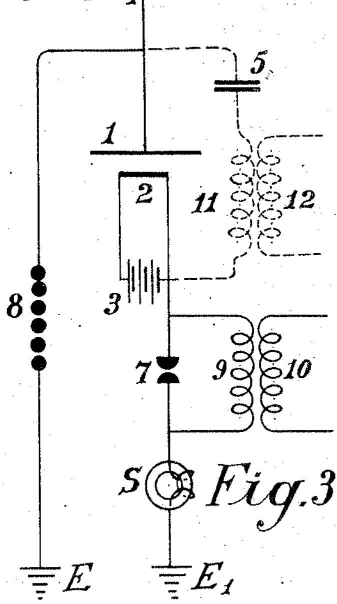
A
5
1
2
11
12
3
8
7
9
10
S
E
E1

Fig. 36.

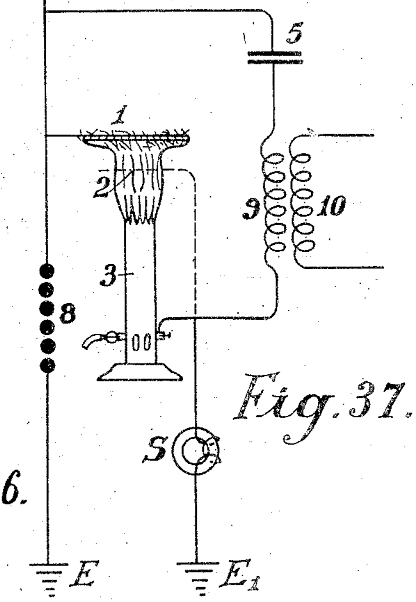
A
5
1
2
9
10
3
8
S
E
E1

Fig. 37.

Inventor

Hermann Plauson

by Knight Bros

Attorneys

Patented June 9, 1925. **1,540,998**

UNITED STATES PATENT OFFICE.

HERMANN PLAUSON, OF HAMBURG, GERMANY.

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY.

Application filed January 13, 1921. Serial No. 437,107.

To all whom it may concern:

Be it known that I, HERMANN PLAUSON, Esthonian subject, residing in Hamburg, Germany, have invented certain new and useful Improvements in the Conversion of Atmospheric Electric Energy, of which the following is a specification.

Methods of obtaining atmospheric electricity by means of metallic nettings set with spikes which are held by means of ordinary or anchored kite balloons made of fabrics and filled with hydrogen, are in theory already known. Atmospheric electricity obtained in this way has been suggested to be used in the form of direct current for the charging of accumulators. This knowledge however is at present only theoretical as the conversion in practice has hitherto been a failure. No means are known of protecting the apparatus from destruction by lightning. The balloons used for collecting the charge must also be made of very large size in order to be able to support the weight of the metallic netting and the heavy cable connections.

Instead of using heavy metallic netting as collectors attached to single air balloons of non-conducting materials which are liable to be torn and are permeable to the gas, it is proposed to use metallic balloon collectors which have the following important advantages—

(*a*) The metallic cases are impenetrable to helium and hydrogen; they also represent large metallic weather-proof collecting surfaces.

(*b*) Radio active means and the like may be easily applied internally or externally; whereby the ionization is considerably increased and therewith also the quantity of atmospheric electricity capable of being collected.

(*c*) Such balloon collectors of light metal do not require to be of large size as they have to carry only their own moderate weight, and that of the conducting cable or wire.

(*d*) The entire system therefore offers little surface for the action of storm and wind and is resistant and stable.

(*e*) Each balloon can be easily raised and lowered by means of a winch so that all repairs, recharging and the like can be carried out without danger during the operation.

It is further proposed to use a collecting aerial network of several separate collectors spread out in the air above the earth, which collectors are interconnected by electrical conductors.

According to this invention charges of atmospheric electricity are not directly converted into mechanical energy, and this forms the main difference from previous inventions, but the static electricity which runs to earth through aerial conductors in the form of direct current of very high voltage and low current strength is converted into electro-dynamic energy in the form of high frequency vibrations. Many advantages are thereby obtained and all disadvantages avoided.

The very high voltage of static electricity of a low current strength can be converted by this invention to voltages more suitable for technical purposes and of greater current strength. By the use of closed oscillatory circuits it is possible to obtain electromagnetic waves of various amplitude and thereby to increase the degree of resonance of such current. Such resonance allows various values of inductance to be chosen whereby again the governing of the starting and stopping of machines driven thereby by simply tuning the resonance between coils of the machine and the transformer circuit forming the resonance can easily be obtained. Further, such currents have the property of being directly available for various uses, even without employing them for driving motors, of which there may be particularly mentioned, lighting, production of heat and use in electro-chemistry.

Further, with such currents a series of apparatus may be fed without direct current supply through conductors and also the electro-magnetic high frequency currents may be converted by means of special motors adapted for electro-magnetic oscillations into mechanical energy, or finally converted by special machines into alternating current of low frequency or even into direct current of high potential.

The invention is more particularly described with reference to the accompanying diagrams in which:—

Figure 1 is an explanatory figure.

Figure 2 is a diagrammatic view of the simplest form.

Figure 3 shows a method of converting atmospheric electrical energy for use with motors.

Figure 4 is a diagram showing the use of protective means.

Figure 5 is a diagram of an arrangement for converting large current strengths.

Figure 6 is a diagram of an arrangement including controlling means.

Figure 7 shows means whereby the spark gap length can be adjusted.

Figure 8 shows a unipolar connection for the motor.

Figure 9 shows a weak coupled system suitable for use with small power motors.

Figures 10, 11 and 12 show modified arrangements.

Figure 13 shows a form of inductive coupling for the motor circuit.

Figure 14 is a modified form of Figure 13 with inductive coupling.

Figure 15 is an arrangement with non-inductive motor.

Figure 16 is an arrangement with coupling by condenser.

Figures 17, 18 and 19 are diagrams of further modifications.

Figure 20 shows a simple form in which the aerial network is combined with special collectors.

Figure 21 shows diagrammatically an arrangement suitable for collecting large quantities of energy.

Figure 22 is a modified arrangement having two rings of collectors.

Figure 23 shows the connections for three rings of collectors.

Figure 24 shows a collecting balloon and diagram of its connection of condenser batteries.

Figures 25 and 26 show modified collector balloon arrangements.

Figure 27 shows a second method of connecting conductor for the balloon aerials.

Figure 28 shows an auto-transformer method of connection.

Figure 29 shows the simplest form of construction with incandescent cathode.

Figure 30 shows a form with cigar shaped balloon.

Figure 31 is a modified arrangement.

Figure 32 shows a form with cathode and electrode enclosed in a vacuum chamber.

Figure 33 is a modified form of Figure 32.

Figure 34 shows an arc light collector.

Figure 35 shows such an arrangement for alternating current.

Figure 36 shows an incandescent collector with Nernst lamp.

Figure 37 shows a form with a gas flame.

Figure 1 illustrates a simple diagram for converting static electricity into dynamic energy of a high number of oscillations. For the sake of clearness in the drawings an influence machine is assumed to be employed and not an aerial antenna. 13 and 14 are combs for collecting the static electricity of the influence machine. 7 and 8 are spark discharging electrodes, 6 and 5 condensers, 9 an inductive primary coil, 10 secondary coil, 11 and 12 ends of conductors of the secondary coil 10. When the disc of the static influence machine is rotated by mechanical means, the combs collect the electric charges one the positive and the other the negative, and charge the condensers 5 and 6 until such a high potential is formed across the spark gap 7—8, that the spark gap is jumped. As the spark gap 7—8 forms a closed circuit with condensers 6 and 5, and inductive resistance 9, as is well known, waves of high frequency electromagnetic oscillations will pass in this circuit.

The high frequency of the oscillations produced in the primary circuit induces waves of the same periodicity in the secondary circuit. Thus in the primary circuit electromagnetic oscillations are formed by the passage of the spark over the spark gap and these waves are maintained by fresh charges of static electricity.

By suitably selecting the ratio between the number of the coils in the primary and secondary circuits with regard to a correct application of the co-efficients of resonance (capacity, inductance, and resistance) the high voltage of the primary circuit may be suitably converted into low voltage and high current strength.

When the oscillatory discharges in the primary circuit becomes weaker or entirely cease, the condensers are charged again by the static electricity until the accumulated charge again breaks down the spark gap. All this is repeated as long as electricity is produced by the static machine by employing mechanical energy.

An elementary form of the invention is shown in Figure 2 in which two spark gaps in parallel are used one of which may be termed the working gap 7 in Figure 2, whilst the second serves as a safety device for excess voltage and consists of a larger number of spark gaps than the working section, which gaps are arranged in series and are bridged by very small capacities as is illustrated in a_1, b_1, c_1, Figure 2 which allow of uniform sparking in the safety section.

In Figure 2 A is the aerial antenna for collecting charges of atmospheric electricity, 13 is the earth connection of the second part of the spark gap, 5 and 6 are condensers, 9 a primary coil. Now when through the aerial A the positive atmospheric electricity seeks to combine with the negative charge to earth, this is prevented by (the air gap between) the spark gaps. The resistance of the spark gap 7 is, as shown in the drawings, lower than that of the other safety section which consists of three spark gaps connected in

series, and consequently a three times greater air resistance is offered by the latter.

So long therefore, as the resistance of the spark gap 7 is not overloaded, so that the other spark gaps have an equal resistance with it the discharges take place only over spark gap 7. Should however the voltage be increased by any influences so that it might be dangerous for charging the condensers 5 and 6 or for the coil insulation 9 and 10 in consequence of break down, by a correct regulation of this spark gap the second spark gap can discharge free from inductive effects direct to earth without endangering the machine.

Without this second spark gap, arranged in parallel having a higher resistance than the working spark gap it is impossible to collect and render available large quantities of electrical energy.

The action of this closed oscillation circuit consisting of spark gap 7, two condensers 5 and 6, primary coil 9, and also secondary coil 10 is exactly the same as the one described in Figure 1 with the arrangement of the static induction machine with the only difference that here the second spark gap is provided. The electromagnetic high frequency alternating current obtained can be tapped off from the conductors 11 and 12 for lighting and heating purposes. Special kinds of motors adapted for working with these peculiar electrical charges may be connected at 14 and 15 which can work with static electricity charges or with high frequency oscillations.

In addition to the use of spark gaps in parallel a second measure of security is also necessary for taking off the current. This precaution consists according to this invention, in the introduction of and method of connecting certain protective electromagnets or choking coils in the aerial circuit as shown by S in Figure 3.

A single electromagnet only having a core of the thinnest possible separate laminations is connected with the aerial.

In the case of high voltages in the aerial network or at places where there are frequent thunder storms, several such magnets may however be connected in series.

In the case of large units or plants several electromagnets can be employed in parallel or in series parallel.

The windings of these electromagnets may be simply connected in series with the aerials. In this case the winding preferably consists of several thin parallel wires, which make up together, the necessary section.

The winding may be made of primary and secondary windings in the form of a transformer. The primary winding will be then connected in series with the aerial network, and the secondary winding more or less short-circuited over a regulating resistance or an induction coil. In the latter case it is possible to regulate to a certain extent the effect of the choking coils. In the further description of the connecting and constructional diagrams the aerial electromagnet choke coil is indicated by a simple ring S.

Figure 3 shows the simplest way of converting atmospheric electricity into electromagnetic wave energy by the use of special motors adapted for high oscillatory currents or static charges of electrical energy. Recent improvements in motors for working with static charges and motors working by resonance, that is to say, having groups of tuned electromagnetic cooperating circuits render this possible but such do not form part of the present invention.

A motor adapted to operate with static charges will for the sake of simplicity be diagrammatically indicated by two semicircles 1 and 2 and the rotor of the motor by a ring M. (Figure 3.) A is a vertical aerial or aerial network. S the safety choke or electromagnet with coil O as may be seen is connected with the aerial A. Adjacent the electromagnet S the aerial conductor is divided into three circuits, the circuit 8 giving the safety spark gap, the circuit 7 with the working spark gap, and then a circuit including the stator terminal 1, the rotor and stator terminal 2 at which a connection is made to the earth wire. The two spark gaps are also connected metallically with the earth wire. The method of working these diagrams is as follows:

The positive atmospheric electric charge collected tends to combine with the negative electricity (or earth electricity) connected with the earth wire. It travels along the aerial A through the electromagnet S without being checked as it flows in the same direction as the direct current. Further, its progress is arrested by two sparks gaps placed in the way and the stator condenser surfaces. The stator condenser surfaces are charged until the charge is greater than the resistance of the spark gap 7, whereupon a spark springs over the spark gap 7 and an oscillatory charge is obtained as by means of the motor M, stator surfaces 1 and 2, and spark gap 7, a closed oscillation circuit is obtained for producing the electromagnetic oscillations. The motor here forms the capacity and the necessary inductance and resistance, which, as is well known, are necessary for converting static electricity into electromagnetic wave energy.

The discharges formed are converted into mechanical energy in special motors and can not reach the aerial network by reason of the electromagnet or choke. If, however, when a spark springs over the spark gap 7 a greater quantity of atmospheric electricity tends to flow to earth, a counter voltage is

induced in the electromagnet, which is greater the more rapidly and strongly the flow of current direct to the earth is. By the formation of this opposing voltage a sufficiently high resistance is offered to the flow of atmospheric electricity direct to earth to prevent a short circuit with the earth.

The circuit containing spark gap 8 having a different wave length which is not in resonance with the natural frequency of the motor, does not endanger the motor and serves as security against excess voltage, which, as practical experiments have shown, may still arise in certain cases, but can be conducted direct to earth through this spark gap.

In the diagram illustrated in Figure 4 the spark gap 7 is shunted across condensers 5 and 6 from the motor M. This construction affords mainly a better insulation of the motor against excess voltage and a uniform excitation through the spark gap 7.

In Figure 5 a diagram is illustrated for transforming large current strengths which may be employed direct without motors, for example, for lighting or heating purposes. The main difference is that here the spark gap consists of a star shaped disc 7 which can rotate on its own axis and is rotated by a motor opposite similarly fitted electrodes 7^a. When separate points of stars face one another, discharges take place, thus forming an oscillation circuit over condensers 5 and 6 and inductance 9 for oscillatory discharges. It is evident that a motor may also be directly connected to the ends of the spiral 9.

The construction of the diagram shown in Figure 6 permits of the oscillation circuit of the motor being connected with an induction coil. Here a regulating inductive resistance is introduced for counter-acting excess voltages in the motor. By cutting the separate coils 9 (coupled inductively to the aerial) in or out the inductive action on the motor may be more or less increased or variable aerial action may be exerted on the oscillation circuit.

In Figure 7 the oscillation circuit is closed through the earth (E and E_1). The spark gap 7 may be prolonged or shortened by more or fewer spark gaps being successively connected by means of a contact arm 7^b.

Diagram 8 shows a unipolar connection of the motor with the aerial network. Here two oscillation circuits are closed through the same motor. The first oscillation circuit passes from aerial A through electromagnet S, point x, inductance 9^a to the earth condenser 6 and further, over spark gap 7 to the aerial condenser 5 and back to x. The second oscillation circuit starts from the aerial condenser 5 at the point x^1 over the inductance 9 to the earth condenser 6 at the point x^3 and through the condenser 6 over the spark gap 7 back to x^1. The motor itself is inserted between the two points of the spark gap 7. From this arrangement slightly damped oscillation wave currents are produced.

In the diagram illustrated in Figure 9 a loosely coupled system of connections is illustrated which is assumed to be for small motors for measuring purposes. A indicates the aerial conductor, S the electromagnet in the aerial conductor, 9 the inductance, 7 the spark gap, 5 and 6 condensers, E the earth, M the motor, and 1 and 2 stator connections of the motor. The motor is directly metallically connected with the oscillation circuit.

In Figure 10 a purely inductive coupling is employed for the motor circuit. The motor is connected with the secondary wire 10 as may be seen in Figure 11 in a somewhat modified diagram connection. The same applies to the diagram of Figure 12.

The diagrams hitherto described preferably allow of motors of small and medium strength to be operated. For large aggregates, however, they are too inconvenient as the construction of two or more oscillation circuits for large amounts of energy is difficult; the governing is still more difficult and the danger in switching on or off is greater.

A means of overcoming such difficulties is shown in Figure 13. The oscillation circuit here runs starting from the point x over condenser 5, variable inductance 9, spark gap 7 and the two segments (3^a and 4^a) forming arms of a Wheatstone bridge, back to x. If the motor is connected by brushes 3 and 4 transversely to the two arms of the bridge as shown in the drawings, electromagnetic oscillations of equal sign are induced in the stator surfaces 1 and 2 and the motor does not revolve. If however, the brushes 3 and 4 are moved in common with the conducting wires 1 and 2 which connect the brushes with the stator poles a certain alteration or displacement of the polarity is obtained and the motor commences to revolve.

The maximum action will result if one brush 3 comes on the central sparking contact 7 and the other brush 4 on the part x. They are however, usually in practice not brought on to the central contact 7 but only held in the path of the bridge segments 4^a and 3^a in order not to connect the spark gaps with the motor oscillation circuit.

As however, the entire oscillation energy can thereby not act on the motor it is better to carry out the same system according to the diagram 14. The diagram 14 differs from the foregoing only by the motor not being directly metallically connected with the segments of the commutator, but only a primary coil 9 which induces in a secondary coil 10, current which feeds the motor M and takes the place of the rotor. By this

arrangement a good transforming action is obtained, a loose coupling and also an oscillation circuit without a spark gap.

In Figure 15 the motor is not purely inductively as in 14, but directly metallically branched off from the primary coil (at x and x^1) after the principle of the auto-transformer.

In Figure 16 instead of an inductance a condenser 6 is in similar manner, and for the same object inserted between the segments 3^a and 4^a. This has the advantage that the segments 3^a and 4^a need not be made of solid metal but may consist of spiral coils whereby a more exact regulation is possible and further motors of high inductance may be employed.

The arrangements of Figures 17, 18 and 19 may be employed for use with resonance and particularly with induction condenser motors; between the large stator induction condenser surfaces, small reversing pole condensers are connected, which, as may be seen from Figures 17, 18 and 19 are led together to earth. Such reversing poles have the advantage that with large quantities of electrical energy the spark formation between the separate oscillation circuits ceases.

Figure 19 shows a further method which prevents electromagnetic oscillations of high number of alternations formed in the oscillation circuit striking back to the aerial conductor. It is based on the well known principle that a mercury lamp, one electrode of which is formed of mercury, the other of solid metal such as steel allows an electric charge to pass in only one direction from the mercury to the steel and not vice versa. The mercury electrode of the vacuum tube N is therefore connected with the aerial conductor and the steel electrode with the oscillation circuit. From this it results that charges can pass only from the aerial through the vacuum tube to the oscillation circuit, but not vice versa. Oscillations which are formed on being transformed in the oscillation circuit cannot pass to the aerial conductor.

In practice these vacuum tubes must be connected behind an electromagnet as the latter alone affords no protection against the danger of lightning.

As regards the use of spark gaps, all arrangements as used for wireless telegraphy may be used. Of course the spark gaps in large machines must have a sufficiently large surface. In very large stations they are cooled in liquid carbonic acid or better still in liquid nitrogen or hydrogen; in most cases the cooling may also take place by means of liquefied low homologues of the metal series or by means of hydrocarbons the freezing point of which lies at between −90° C. and −40° C. The spark gap casing must also be insulated and be of sufficient strength to be able to resist any pressure which may arise. Any undesirable excess super-pressure which may be formed must be automatically let off. I have employed with very good results mercury electrodes which were frozen in liquid carbonic acid, the cooling being maintained during the operation from the outside through the walls.

Figure 20 is one of the simplest forms of construction of an aerial network in combination with collectors, transformers and the like illustrated diagrammatically. E is here the earth wire, 8 the safety spark gap, 7 the working spark gap, 1 and 2 the stator surfaces of the motor, 5 a condenser battery, S the protective magnet which is connected with the coil in the aerial conductor, A^1 to A^{10} aerial antennæ with collecting balloons, N horizontal collecting or connecting wires from which, to the centre a number of connections run.

The actual collectors consist of metal sheaths preferably made of an aluminium magnesium alloy, and are filled with hydrogen or helium and are attached to copper plated steel wires. The size of the balloon is selected so that the actual weight of the balloon and the weight of the conducting wire is supported thereby. On the top of the balloon aluminium spikes, made and gilded in a special manner hereinafter described, are arranged in order to produce a conductor action. Small quantities of radium preparations, more particularly polonium-ionium or mesothorium preparations considerably increase the ionization, and therewith the action of these collectors.

In addition to metal balloons, fabric balloons which are superficially metal coated according to Schoop's metal spraying process, may however also be employed. A metallic surface may also be produced by lacquering with metallic bronzes, preferably according to Schoop's spraying process or lacquering with metallic bronze powders in two electrical series of widely different metals, because thereby the collecting effect is considerably increased.

Instead of the ordinary round balloons, elongated cigar shaped ones may be employed. In order also to utilize the frictional energy of the wind, patches or strips of non-conducting substances which produce electricity by friction, may be attached to the metallized balloon surfaces. The wind will impart a portion of its energy in the form of frictional electricity, to the balloon casing, and thereby the collecting effect is substantially increased.

In practice however, very high towers (up to 300 metres is fully admissible) may be employed as antennæ. In these towers copper tubes rise freely further above the top of the tower. A gas lamp secured

against the wind is then lit at the point of the copper tube and a netting is secured to the copper tube over the flame of this lamp to form a collector. The gas is conveyed through the interior of the tube up to the summit. The copper tube must be absolutely protected from moisture at the place at which it enters the tower and also rain must be prevented running down the walls of the tower which might lead to a bad catastrophe. This is done by bell shaped enlargements which expand downwards, being arranged in the tower in the form of high voltage insulators of Siamese pagodas.

Special attention must be devoted to the foundations of such towers. They must be well insulated from the ground, which may be obtained by first embedding a layer of concrete in a box form to a sufficient depth in the ground and inserting in this an asphalt lining and then glass bricks cast about 1 or 2 metres in thickness. Over this in turn there is a ferro-concrete layer in which alone the metal foot of the tube is secured. This concrete block must be at least 2 metres from the ground and be fully protected at the sides by a wooden covering, from moisture. In the lower part of the tower a wood or glass house for the large condenser batteries or for the motors may be constructed. In order to lead the earth connection to the ground water, a well insulated pit constructed of vitreous bricks, must be provided. Several such towers are erected at equal distances apart and connected with a horizontal conductor. The horizontal connecting wires may either run directly from tower to tower or be carried on bell shaped insulators similar to those in use for high voltage conductors. The width of the network may be of any suitable size and the connection of the motors can take place at any suitable places.

In order to collect large quantities of electricity with few aerials it is well to provide the aerial conductor with batteries of condensers as shown in two methods of construction in Figures 21 and 22. In Figure 21 the batteries of condensers 5 are connected on the one hand with the aerial electricity collectors Z by the aerial conductor A, and on the other hand interconnected in series with an annular conductor from which horizontal conductors run to the connecting points C to which the earth wire is connected.

Figure 22 shows a similar arrangement. Should two such series of antennæ rings be shown by a voltmeter to have a large difference of potential (for example, one in the mountains and one in the plain) or even of different polarity these differences may be compensated for by connecting sufficiently large condenser batteries (5, 5^a, 5^b) by means of Maji star conductors D and D^1. In Figure 23 a connection of three such rings of collectors to form a triangle with a central condenser battery is illustrated.

The condenser batteries of such large installations must be embedded in liquefied gases or in liquids freezing at very low temperatures. In such cases a portion of the atmospheric energy must be employed for liquefying these gases. It is also preferable to employ pressure. By this means the condenser surfaces may be diminished, and still allow for large quantities of energy to be stored, secure against breakdown. For smaller installations the immersing of the condensers in well insulated oil or the like, suffices. Solid substances on the other hand cannot be employed as insulators.

The arrangement in the diagrams hitherto described was always such that the condenser batteries were connected with both poles directly to the aerial conductors. An improved diagram of the connections for obtaining atmospheric electricity for the condenser batteries has however, been found to be very advantageous, this arrangement consists in that they are connected by only one pole (unipolar) to the collecting network. Such a method of arrangement is very important, as by means of it a constant current and an increase of the normal working pressure or voltage is obtained. If for example a collecting balloon aerial which is allowed to rise to a height of 300 metres, shows 40,000 volts above earth voltage, in practice it has been found that the working voltage (with a withdrawal of the power according to the method hereinbefore described by means of oscillating spark gaps and the like) is only about 400 volts. If however, the capacity of the condenser surfaces be increased, which capacity in the above mentioned case was equal to that of the collecting surface of the balloon aerials, to double the amount, by connecting the condenser batteries with only one pole, the voltage rises under an equal withdrawal of current up to and beyond 500 volts. This can only be ascribed to the favourable action of the connecting method.

In addition to this substantial improvement it has also been found preferable to insert double inductances with electromagnets and to place the capacities preferably between two such electromagnets. It has also been found that the useful action of such condensers can be further increased if an induction coil be connected as inductive resistance to the unconnected pole of the condenser, or still better if the condenser itself be made as an induction condenser. Such a condenser may be compared with a spring which when compressed carries in itself accumulated force, which it again gives off when released. In charging, a charge with reversed sign is formed at the

other free condenser pole, and if through the spark gap a short circuit results, the accumulated energy is again given back since now new quantities of energy are induced at the condenser pole connected with the conductor network, which in fact charges with opposite signs to that at the free condenser pole. The new induced charges have of course the same sign as the collector network. The whole voltage energy in the aerial is thereby however increased. In the same space of time larger quantities of energy are accumulated than is the case without such inserted condenser batteries.

In Figures 24 and 25 two different diagrams of connections are more exactly illustrated, Figure 24 shows a collecting balloon and the diagram of the connections to earth. Figure 25 four collecting balloons and the parallel connection of the condenser batteries belonging thereto.

A is the collecting balloon made of an aluminium magnesium alloy (electron metal, magnalium) of a specific gravity of 1.8 and a thickness of plate 0.1 to 0.2 mm. Inside there are eight strong vertical ribs of T shaped section about 10 to 20 mm. in height and about 3 mm. in thickness with the projecting part directed inwards (indicated by *a*, *b*, *c*, *d* and so forth); they are riveted together to form a firm skeleton and are stiffened in a horizontal direction by two cross ribs. The ribs are further connected with one another internally and transversely by means of thin steel wires, whereby the balloon obtains great power of resistance and elasticity. Rolled plates of 0.1 to 0.2 mm. in thickness made of magnalium alloy are then either soldered or riveted on this skeleton so that a fully metallic casing with smooth external surface is obtained. Well silvered or coppered aluminium plated steel wires run from each rib to the fastening ring 2. Further, the coppered steel hawser L preferably twisted out of separate thin wires (shown in dotted lines in Figure 24) and which must be long enough to allow the balloon to rise in the desired height, leads to a metal roller or pulley 3 and from thence to a winch W, well insulated from the earth. By means of this winch, the balloon, which is filled with hydrogen, or helium, can be allowed to rise to a suitable height (300 to 5,000 metres) and brought to the ground for recharging or repairs.

The actual current is taken directly through a friction contact from the metal roller 3 or from the wire, or even from the winch or simultaneously from all three by means of brushes (3, 3^a and 3^b). Beyond the brushes the conductor is divided, the paths being:—firstly over 12 to the safety spark gap 8, from thence to the earth conductor E^1, and secondly over electromagnet S^1, point 13, to a second loose electromagnet having an adjustable coil S^2, then to the spark gap 7 and to the second earth conductor E^2. The actual working circuit is formed through the spark gap 7, condensers 5 and 6, and through the primary coil 9; here the static electricity formed by oscillatory discharges is accumulated and converted into high frequency electromagnetic oscillations. Between the electromagnets S^1 and S^2 at the crossing point 13, four condenser batteries are introduced which are only indicated diagrammatically in the drawings each by one condenser. Two of these batteries (16 and 18) are made as plate condensers and prolonged by regulating induction coils or spirals 17 and 19 while the two others (21 and 23) are induction condensers. As may be seen from the drawings each of the four condenser batteries 16, 18, 21, 23 is connected only by one pole to the aerial or to the collector conductor. The second poles 17, 19, 22, 24 are open. In the case of plate condensers having no inductive resistance an induction coil is inserted. The object of such a spiral or coil is the displacement of phase of the induction current by ¼ periods, whilst the charging current of the condenser poles which lie free in the air, works back to the collector aerial. The consequence of this is that in discharges in the collector aerial the back inductive action of the free poles allows a higher voltage to be maintained in the aerial collecting conductor than would otherwise be the case. It has also been found that such a back action has an extremely favourable effect on the wear of the contacts. Of course the inductive effect may be regulated at will within the limits of the size of the induction coil, the length of the coil in action being adjustable by means of wire connection without induction (see Fig. 24, No. 20).

S^1 and S^2 may also be provided with such regulating devices in the case of S^2 (illustrated by 11). If excess voltage be formed it is conducted to earth through the wire 12 and spark gap 8 or through any other suitable apparatus, since this formation would be dangerous for the other apparatus.

The action of these condenser batteries has already been hereinbefore described.

The small circles on the collector balloon indicate places at which zinc amalgam or gold amalgam or other photoelectric acting metals in the form of small patches in extremely thin layers (.01 to .05 mm. in thickness) are applied to the balloon casing of light metal. Such metallic patches may also be applied to the entire balloon as well as in greater thickness to the conducting network. The capacity of the collector is thereby considerably strengthened at the surface. The greatest possible effect in collecting may be obtained by polonium amalgams and the like. On the surface of the

collector balloon metal points or spikes are also fixed along the ribs, which spikes serve particularly for collecting the collector charge. Since it is well known that the resistance of the spikes is less the sharper the spike is, for this purpose it is therefore extremely important to employ as sharp spikes as possible. Experiments made as regards these have shown that the formation of the body of the spike or point also plays a large part. for example, spikes made of bars or rollers with smooth surfaces, have a many times greater point resistance as collector accumulator spikes than those with rough surfaces. Various kinds of spike bodies have been experimented with for the collector balloons hereinbefore mentioned. The best results were given by spikes which were made in the following way. Fine points made of steel, copper, nickel, or copper and nickel alloys, were fastened together in bundles and then placed as anode with the points in a suitable electrolyte (preferably in hydrochloric acid or muriate of iron solutions) and so treated with weak current at 2 to 3 volts pressure. After 2 to 3 hours according to the thickness of the spikes or pins the points become extremely sharp and the bodies of the spikes have a rough surface. The bundle can then be removed and the acid washed off with water. The spikes are then placed as cathode in a bath consisting of solution of gold, platinum, iridium, paladium or wolfram salts or their compounds and coated at the cathode galvanically with a thin layer of precious metal, which must however be sufficiently firm to protect them from atmospheric oxidation.

Such spikes act at a 20 fold lower voltage almost as well as the best and finest points made by mechanical means. Still better results are obtained if polonium or radium salts are added to the galvanic bath when forming the protective layer or coating. Such pins have a low resistance at their points and even at one volt and still lower pressures have an excellent collector action.

In Figure 24 the three unconnected poles are not connected with one another in parallel. That is quite possible in practice without altering the principle of the free pole. It is also preferable to interconnect in parallel to a common collector network, a series of collecting aerials.

Figure 25 shows a diagram for such an installation. A^1, A^2, A^3, A^4 are four metal collector balloons with gold or platinum coated spikes which are electrolytically made in the presence of polonium emanations or radium salts, which spikes or needles are connected over four electro-magnets S^1, S^2, S^3, S^4, through an annular conductor R. From this annular conductor four wires run over four further electromagnets S^a, S^b, S^c, S^d, to the connecting point 13. There the conductor is divided, one branch passing over 12 and the safety spark gap 8 to the earth at E^1, the other over inductive resistance J and working spark gap 7 to the earth at E^2. The working circuit, consisting of the condenser 5 and 6 and a resonance motor or a condenser motor M, such as hereinbefore described, is connected in proximity round the sparking gap section 7.

Instead of directly connecting the condenser motor of course the primary circuit for high frequency oscillatory current may also be inserted.

The condenser batteries are connected by one pole to the annular conductor R and can be either inductionless (16 and 18) or made as induction condensers as shown by 21 and 23. The free poles of the inductionless condensers are indicated by 17 and 19, those of the induction condensers by 22 and 24. As may be seen from the drawings all these poles 17, 22, 19, 24 may be interconnected in parallel through a second annular conductor without any fear that thereby the principle of the free pole connection will be injured. In addition to the advantages already set forth the parallel connection also allows of an equalization of the working pressure in the entire collector network. Suitably constructed and calculated induction coils 25 and 26 may also be inserted in the annular conductor of the free poles, by means of which a circuit may be formed in the secondary coils 27 and 28 which allows current produced in this annular conductor by fluctuations of the charges or the like appearances to be measured or otherwise utilized.

According to what has been hereinbefore stated separate collector balloons may be connected at equidistant stations distributed over the entire country, either connected directly with one another metallically or by means of intermediate suitably connected condenser batteries through high voltage conductors insulated from earth. The static electricity is converted through a spark gap into dynamic energy of a high number of oscillations and may in such form be coupled as a source of energy by means of a suitable method of connecting, various precautions being observed and with special regulations. The wires leading from the collector balloons have hitherto been connected through an annular conductor without this endless connection, which can be regarded as an endless induction coil, being able to exert any action on the whole conductor system.

It has now been found that if the network conductor connecting the aerial collector balloons with one another is not made as a simple annular conductor, but preferably short circuited in the form of coils over a

condenser battery or spark gap or through thermionic tubes or valves or audions, then the total collecting network exhibits quite new properties. The collection of atmospheric electricity is thereby not only increased but an alternating field may be easily produced in the collector network. Further, the atmospheric electrical forces showing themselves in the higher regions may also be directly obtained by induction. In Figures 26 and 28 a form of construction is shown on the basis of which the further foundations of the method will be more particularly explained.

In Figure 26 1, 2, 3, 4 are metal collector balloons, 5, 6, 7, 8 their metallic aerial conductors and I the actual collector network. This consists of five coils and is mounted on high voltage insulators in the air, on high voltage masts (or with a suitable construction of cable embedded in the earth). One coil has a diameter of 1 to 100 km. or more. S and S^1 are two protective electromagnets, F the second safety section against excess voltage, E its earth conductor and E^1 the earth conductor of the working section. When an absorption of static atmospheric electricity is effected through the four balloon collectors, the current in order to reach the earth connection E^1 must flow spirally through the collector network over the electromagnet S, primary induction coil 9, conductor 14, anode A of the audion tube, incandescent cathode K, as the way over the electromagnet and safety spark gap F offers considerably greater resistance. Owing to the fact that the accumulated current flows in one direction, an electromagnetic alternating field is produced in the interior of the collector network coil, whereby the whole free electrons are directed more or less into the interior of the coil. An increased ionization of the atmosphere is therefore produced. In consequence of this the points mounted on the collector balloon show a considerably reduced resistance and therefore increased static charges between the points on the balloon and the surrounding atmosphere are produced. The result of this is a considerably increased collector effect.

A second effect which could not be obtained otherwise is obtained by the electromagnetic alternating field which running parallel to the earth surface, acts more or less with a diminishing or increasing effect on the earth magnetic field, whereby in the case of fluctuations in the current a return induction current of reversed sign is always produced in the collector coil by earth magnetism. Now if, however, a constantly pulsating continuous alternating field is produced as stated in the above collector network I, an alternating current of the same periodicity is produced also in the collecting network coil. As the same alternating field is further transmitted to the aerial balloon, the resistance of its points is thereby considerably reduced, whilst the collector action is considerably increased. A further advantage is that positive electrons which collect on the metal surfaces during the conversion into dynamic current produce a so-called drop of potential of the collector area. As an alternating field is present, the negative ions surrounding the collector surfaces, when discharge of the collector surfaces takes place produce by the law of induction, an induction of reversed sign on the collector surface and so forth (that is to say again a positive charge). In addition to the advantages hereinbefore set forth, the construction of connecting conductors in coil form when of sufficiently large diameter, allows of a utilization of energy arising in higher regions also in the simplest way. As is well known electric discharges frequently take place at very great elevations which may be observed, such as St. Elmo's fires or northern lights. These energy quantities have not been able to be utilized up to now. By this invention all these kinds of energy, as they are of an electromagnetic nature and the direction of the axis of the collector coils stands at right angles to the earth's surface, can be more or less absorbed in the same way as a receiver in wireless telegraphy absorbs waves coming from a far distance. With a large diameter of the spiral it is possible to connect large surfaces and thereby to take up also large quantities of energy.

It is well known that large wireless stations in the summer months, and also in the tropics are very frequently unable to receive the signals in consequence of interruptions which are caused by atmospheric electricity, and this takes place with vertical coils of only 40 to 100 m. diameter. If on the contrary horizontal coils of 1 to 100 km. diameter be employed very strong currents may be obtained through discharges which are constantly taking place in the atmosphere. Particularly in the tropics or still better in the polar regions where the northern lights are constantly present, large quantities of energy may probably be obtained in this way. A coil with several windings should act the best. In similar manner any alteration of the earth magnetism should act inductively on such a coil.

It is not at all unlikely that earthquakes and spots on the sun will also produce an induction in such collector coils of sufficient size. In similar manner this collector conductor will react on earth currents more particularly when they are near the surface of the earth or even embedded in the earth. By combining the previous kind of current collectors so far as they are adapted for

the improved system with the improved possibilities of obtaining current the quantities of free natural energy which are to be obtained in the form of electricity are considerably increased.

In order to produce in the improved collector coil uniform current oscillations of an undamped nature so-called audion high vacuum or thermionic tubes of suitable connection are employed instead of the previously known spark gaps (Fig. 26, Nos. 9–18). The main aerial current flows through electromagnet S (which in the case of a high number of alternations is not connected here but in the earth conductor E^1) and may be conveyed over the primary coils in the induction winding through wire 14 to the anode A of the high vacuum grid tube. Parallel with the induction resistance 9 a regulating capacity of suitable size, such as condenser 11 is inserted. In the lower part of the vacuum grid tube is arranged the incandescent filament or the cathode K which is fed through a battery B. From the battery B two branches run, one to the earth conductor E^1 and the other through battery B^1 and secondary coil 10 to the grid anode *g* in the vacuum tube. By the method of connections shown in dotted lines, a desired voltage at the grid electrode *g* may also be produced through the wire 17 which is branched off from the main current conductor through switches 16 and some small condensers (*a*, *b*, *c*, *d*) connected in series, and conductor 18, without the battery B^1 being required.

The action of the entire system is somewhat as follows:—

On the connecting conductor of the aerial collector network being short circuited to earth, the condenser pole 11 is charged and slightly damped oscillations are formed in the short circuited existing oscillation circuit formed of the condenser 11 and self inductance 9. In consequence of the coupling through coil 10, fluctuations of voltage take place in the grid circuit 15 with the same frequency, which, fluctuations in turn influence the strength of the electrode current passing through the high vacuum amplyfying tube and thus produce current fluctuations of the same frequency in the anode circuit. A permanent supply of energy to the oscillation circuits 9 and 10 consequently takes place, until a condition of balance is set up, in which the consumed oscillation energy is equal to that absorbed. Thereby constant undamped oscillations are now produced in the oscillation circuits 9–11.

For regular working of such oscillation producers high vacuum strengthening tubes are necessary and it is also necessary that the grid and anode voltages shall have a phase difference of 180° so that if the grid is negatively charged, then the anode is positively charged and vice versa. This necessary difference of phase may be obtained by most varied connections, for example, by placing the oscillation circuit in the grid circuit or by separating the oscillation circuit and inductive coupling from the anodes and the grid circuit and so forth.

A second important factor in this way of converting static atmospheric electricity into undamped oscillations is that care must be taken that the grid and anode voltages have a certain relation to one another; the latter may be obtained by altering the coupling and a suitable selection of the self induction in the grid circuit, or as shown by dotted lines 18, 17, 16 by means of a larger or smaller number of condensers of suitable size connected in series; in this case the battery B^1 may be omitted. With a suitable selection of the grid potential a glow discharge takes place between the grid *g* and the anode A, and accordingly at the grid there is a cathode drop and a dark space is formed. The size of this cathode drop is influenced by the ions which are emitted in the lower space in consequence of shock ionization of the incandescent cathodes K and pass through the grid in the upper space. On the other hand the number of the ions passing through the grid is dependent on the voltage between the grid and the cathode. Thus if the grid voltage undergos periodic fluctuations (as in the present case) the amount of the cathode drop at the grid fluctuates and consequently the internal resistance of the tube correspondingly fluctuates, so that when a back coupling of the feed circuit with the grid circuit takes place, the necessary means are afforded for producing undamped oscillations and of taking current, according to requirements from the collecting conductor.

The frequency of the undamped oscillations produced is with a suitably loose coupling equal to the self frequency of the oscillation circuits 9 and 10. By a suitable selection of the self induction of the coil 9 and capacity 11 it is possible to extend from frequencies which produce electromagnetic oscillations of only a few metres wave length down to the lowest practical alternating current frequency. For large installations a suitable number of frequency producing tubes in the form of the well known high vacuum transmission tubes of .5 to 2 kw. in size may be connected in parallel so that in this respect no difficulty exists.

The use of such tubes for producing undamped oscillations, and also the construction and method of inserting such transmission tubes in an accumulator or dynamo circuit is known and also that such oscillation producing tubes only work well at voltages of 1,000 up to 4,000 volts, so that on the contrary their use at lower voltages is

considerably more difficult. By the use of high voltage static electricity this method of producing undamped oscillations as compared with that through spark gaps must be regarded as an ideal solution particularly for small installations of outputs of from 1 to 100 kw.

By the application of safety spark gaps, with interpolation of electro-magnets, not only is short circuiting avoided but also the taking up of current is regulated. Oscillation producers inserted in the above way form a constantly acting electromagnetic alternating field in the collector coil, whereby as already stated, a considerable accumulating effect takes place. The withdrawal wire or working wire is connected at 12 and 13, but current may be taken by means of a secondary coil which is firmly or movably mounted in any suitable way inside the large collector coil, i. e. in its electromagnetic alternating field, so long as the direction of its axis runs parallel with that of the main current collecting coil.

In producing undamped oscillations of a high frequency (50,000 per second and more) in the oscillation circuits 9 and 11, electromagnets S and S^1 must be inserted if the high frequency oscillations are not to penetrate the collector coil, between the oscillation producers and the collector coil. In all other cases they are connected shortly before the earthing (as in Figs. 27 and 28).

In Figure 27 a second method of construction of the connecting conductor of the balloon aerials is illustrated in the form of a coil. The main difference consists in that in addition to the connecting conductor I another annular conductor II is inserted parallel to the former on the high voltage masts in the air (or embedded as a cable in the earth) but both in the form of a coil. The connecting wire of the balloon aerials is indicated as a primary conductor and also as a current producing network; the other is the consumption network and is not in unipolar connection with the current producing network.

In Figure 27 the current producing network I is shown with three balloon collectors 1, 2, 3 and aerial conductors 4, 5, 6; it is short circuited through condenser 19 and inductance 9. The oscillation forming circuit consists in this diagram of spark gap *f*, inductance 10, and condenser 11; the earth wire E, is connected to earth over electromagnet S^1. F is the safety spark gap which is also connected to earth through a second electromagnet S at E. On connecting up the condenser circuit 11 this is charged over the spark gap *f* whereby an oscillatory discharge is formed. This discharging current acts through inductance 10 on the inductively coupled secondary 9, whereby in the producing network a modification of the potential of the condenser 19 is produced. The consequence of this is that oscillations arise in the coil shaped producer network. These oscillations induce a current in the secondary circuit II, which has a smaller number of windings and a less resistance, the voltage of which, according to the proportion of the number of windings and of the ohmic resistance, is considerably lower whilst the current strength is greater.

In order to convert the current thus obtained into current of an undamped character, and to tune its wave lengths, a sufficiently large regulatable capacity 20 is inserted between the ends 12 and 13 of the secondary conductor II. Here also current may be taken without an earth conductor, but it is advisable to insert a safety spark gap E^1 and to connect this with the earth over an electromagnet S^2.

The producer network may be connected with the working network II over an inductionless condenser 21 or over an induction condenser 22, 23. In this case the secondary conductor is unipolarly connected with the energy conductor.

In Figure 28 the connecting conductor between the separate accumulator balloons is carried out according to the autotransformer principle. The collecting coil connects four aerial balloons 1, 2, 3, 4, the windings of which are not made side by side but one above the other. In Figure 28 the collector coil I is shown with a thin line, the metallically connected prolongation coils II with a thick line. Between the ends I^1 and II^1 of the energy network I a regulating capacity 19 is inserted. The wire I^1 is connected with the output wire and with the spark gap F.

As transformer of the atmospheric electricity an arrangement is employed which consists in using rotary pairs of condensers in which the one stator surface B is connected with the main current, whilst the other A is connected with the earth pole. Between these pairs of short circuited condensers are caused to rotate from which the converted current can be taken by means of two collector rings and brushes, in the form of an alternating current, the frequency of which is dependent on the number of balloons and the revolutions of the rotor. As the alternating current formed in the rotor can act, in this improved method of connection described in this invention, through coils 10 on the inductance 9, an increase or diminution of the feed current in I can be obtained according to the direction of the current by back induction. Current oscillations of uniform rhythm thereby result in the coil shaped windings of the producer net-work.

As the ends of this conductor are short cir-

cuited through the regulatable condenser 19 these rhythms produce short circuited undamped oscillations in the energy conductor, the periodicity and wave lengths of which oscillations can be adjusted according to desire by altering the capacity 19 to a given wave length and therewith also to a given frequency. These currents may also be employed in this form directly as working current through the conductors II^1 and III. By inserting the condenser 20 a connection between these conductors may also be made, whereby harmonic oscillations of desired wave length are formed. By this means quite new effects as regards current distribution are obtained. The withdrawal of current can even take place without direct wire connection if, at a suitable point in the interior of the producing network (quite immaterially whether this has a diameter of 1 or 100 km.) a coil tuned to these wave lengths and of the desired capacity is firmly or movably mounted in the aerial conductor in such a way that its axial direction is in parallel with that of the collector coil. In this case a current is induced in the producing network, the size of which is dependent on the total capacity and resistance and also on the periodicity employed. A possibility is thereby afforded in future, of taking energy from the producer network by wireless means. As thereby in addition to atmospheric electricity also magnetic earth currents and the energy from the higher atmosphere (at least partially) may be simultaneously obtained, this last system for collecting the atmospheric energy is of particular importance for the future.

Of course everywhere instead of spark gaps suitable grid vacuum tubes may be employed as producers for undamped oscillations. The separate coils of the producer net-work with large diameters may be connected with one another through separate conductors all in parallel or all in series or in groups in series. By regulating the number of oscillations and also the extent of the voltage more or less large collector coils of this kind may be employed. The coils may also be divided spirally over the entire section. The coils may be carried out in annular form or also in triangular, quadrangular, hexagonal or octagonal form.

Of course wires may be carried from a suitable place to the centre or also laterally which serve the current waves as guides. This is necessary when the currents have to be conducted over mountains and valleys and so forth. In all these cases the current must be converted into a current of suitable periodicity.

As already hereinbefore mentioned separate collecting balloons may be directly metallically interconnected at equidistant stations distributed over the entire country or may be connected by interpolation of suitable condenser batteries by means of high voltage conductors. The static electricity is converted through a spark gap into dynamic energy of a high number of oscillations, and could then in such form, with a suitable arrangement of the connections, observing various measures of precaution, be employed as source of energy after separate or special regulation.

According to this invention in order to increase the collecting effect of the balloon in the aerial collector conductor or in the earth wire, radiating collectors are employed. These consist either of incandescent metal or oxide electrodes in the form of vacuum grid tubes, or electric arcs (mercury and the like electrodes) Nernst lamps, or finally flames of various kinds may be simply connected with the respective conductor.

It is well known that energy can be drawn off from a cathode consisting of an incandescent body opposite an anode charged with positive electricity (vacuum grid tube). Hitherto however, a cathode was always first directly placed opposite an anode, and secondly the system always consisted of a closed circuit.

Now if we dispense with the ordinary ideas in forming light or flame arcs in which a cathode must always stand directly opposite an anode, and if we place an incandescent cathode opposite an anode charged to a high potential or another body freely floating in the air, or regard the incandescent cathode only as a source of unipolar discharge (which represent group and point discharges in electro-static machines similar to unipolar discharges), it may be ascertained that incandescent cathodes and less perfectly all incandescent radiators, flames and the like admit of relatively large current densities and allow large quantities of electric energy to radiate into the open space in the form of electron streams as transmitters.

The object of this invention is as described below, if such incandescent oxide electrodes or other incandescent radiators or flames are not freely suspended in space but connected metallically with the earth so that they can be charged with negative terrestrial electricity, these radiators possess the property of absorbing the free positive electrical charges contained in the air space surrounding them (that is to say of collecting them and conducting them to earth). They can therefore, serve as collectors and have, in comparison to the action of the spikes, or points, a very large radius of action R; the effective capacity of these collectors is much greater than the geometrical capacity (R_0–) calculated in an electro-static sense.

Now as our earth is surrounded as is well

known, with an electro-static field and the difference of potential

$$\frac{\delta V}{\delta h}$$

of the earth field according to the latest investigations, is in summer about 60 to 100 volts and in winter 300 to 500 volts per metre of difference in height (δh), a simple calculation gives the result that when such a radiation collector or flame collector is arranged for example on the ground, and a second one is mounted vertically over it at a distance of 2,000 metres and both are connected by a conducting cable, there is a difference of potential in summer of about 2,000,000 volts and in winter even of 6,000,000 volts and more.

According to Stefan Boltzmann's law of radiation, the quantity of energy which an incandescent surface (temperature T) of 1 sq. cm. radiates in a unit of time into the open air (temperature T_o) is expressed by the following formula:

$$S=\sigma\,(T^4-T^4{}_o)\ \text{watt/sq. cm.}$$

and the universal radiation constant σ is according to the latest researches of Ferry (Annales de Chimie et de physique 17 page 267 (1909)) equal to 6.30×10^{-12} watt/sq. cm.

Now if an incandescent surface of 1 sq. cm. shows, as compared with the surrounding space a periodic fall of potential δV it radiates (independent of the current direction, that is to say of the sign) in accordance with the above formula, for example at a temperature of 3725° C. an energy of 1.6 kw. per sq. cm. per second. As for the radiation the same value can be calculated for the collection of energy, but reversed. Now as carbon electrodes at the temperature of the electric arc support on the current basis a current density up to from 60 to 65 amperes per sq. cm. no difficulties will result in this direction in employing radiating collectors as accumulators.

If the earth be regarded as a cosmically insulated condenser in the sense of geometrical electro-statics *x* there results from the geometric (compare Ewald Rasch, "das elektrische Bogenlicht" (The electric arc light) page 169) capacity of the earth according to Chwolson:

For negative charging 1.3×10^6 Coulomb

For negative potential $V=10\times10^8$ volts.

From this there results however, $EJT\cong24.7\times10^{24}$ watt/Sec. Now if it is desired to make a theoretic short circuit through an earthed flame collector this would represent an electric total work of about 79,500 10^{10} kilowatt years. As the earth must be regarded as a rotating mechanism which is thermo-dynamically, electromagnetically, and also kinematically coupled with the sun and stars system by cosmic radiations and gravitation a diminution of the electric energy of the earth field is not to be feared. The energies which the incandescent collectors would withdraw from the earth field can only cause by the withdrawal of motor work a lowering of the earth temperature (temperature $T_E=300$) and reduce this to that of the world space ($T=O$) by using the entire energy. This is however not the case as the earth does not represent a cosmically entirely insulated system. On the contrary there is conveyed to the same according to the recent value corrected by Ferry for the solar constants through the radiation from the sun an energy of $18{,}500\times10^{10}$ kw. Accordingly any lowering of the earth temperature (T_E) without a simultaneous lowering of the sun's temperature (T_S) would contradict Stefan Boltzmann's law of radiation.

$$S=\sigma\,(T_S{}^4-T_E{}^4).$$

From this it must be concluded that if the earth temperature (T_E) sinks, the total radiation S absorbed by the earth increases, and further also that the secular speed of cooling of the earth is directly dependent on that of the sun and the other radiators cosmically coupled with the sun and is connected most closely with these.

The incandescent radiation collectors may, according to this invention, be employed for collecting atmospheric electricity if they (1) are charged with the negative earth electricity (that is to say when they are directly connected by means of a metallic conductor with the earth) and (2) if large capacities (metal surfaces) charged with electricity are mounted opposite them as positive poles in the air. This is regarded as the main feature of the present invention as without these inventive ideas it would not be possible to collect with an incandescent collector, sufficiently large quantities of the electrical charges contained in the atmosphere as technology requires; the radius of action of the flame collectors would also be too small, especially if it be considered that the very small surface density (energy density) (σ about$=2\times7\ .\ 10^9$ St. E. per sq. cm.) does not allow of large quantities of charge being absorbed from the atmosphere.

x) Calculated according to Poisson's calculation;

$\Delta V=-4\pi\delta$; as here the alteration of the potential or potential gradients only takes place in the direction of the normal, this calculation assumes the simple form

$$\delta=\frac{1}{4\pi}\times\frac{\delta^2V}{\delta n^2}$$

It has indeed already been proposed to employ flame collectors for collecting atmospheric electricity and it is known that

their collecting effect is substantially greater opposite the points. It is however, not known that the quantities of current which could hitherto be obtained are too small for technical purposes. According to my experiments the reason for this is to be found in the too small capacities of the collector conductor poles. If such flame or radiating collectors have no or only small positive surfaces, their radius of action for large technical purposes is too small. If the incandescent collectors be constantly kept in movement in the air they may collect more according to the speed of the movement, but this is again not capable of being carried out in practice.

By this invention the collector effect is considerably increased by a body charged with a positive potential and of the best possible capacity being also held floating (without direct earth connection) opposite such an incandescent collector which is held floating in the air at a desired height. If for example, a collecting balloon of sheet metal or of metalized balloon fabric be caused to mount to 300 up to 3,000 metres in the air and as positive pole it is brought opposite such a radiating collector connected by a conductor to earth, quite different results are obtained.

The metallic balloon shell (with a large surface) is charged to a high potential by the atmospheric electricity. This potential is greater the higher the collecting balloon is above the incandescent collector. The positive electricity acts concentratedly on the anode floating in the air as it is attracted through the radiation shock ionization, proceeding from the incandescent cathode. The consequence of this is that the radius of action of the incandescent cathode collector is considerably increased and thereby also the collecting effect of the collecting balloon surface. Further the large capacity of the anode floating in the air plays therefore an important part because it allows of the taking of large charges, and thereby a more uniform current is obtained even when there is a large consumption: this cannot be the case with small surfaces.

In the present case the metallic collecting balloon is a positive anode floating in the air and the end of the earth conductor of this balloon serves as positive pole surface opposite the surface of the radiating incandescent cathode, which in turn is charged with negative earth electricity being conductingly connected to earth.

The process may be carried out by two such contacts (negative incandescent cathode and anode end of a capacity floating in the air) a condenser and an inductive resistance being switched on in parallel, whereby simultaneously undamped oscillations may be formed.

In very large installations it is advisable to connect two such radiating collectors in series. Thus an arc light incandescent cathode may be placed below on the open ground and an incandescent cathode which is heated by special electro-magnetic currents be located high in the air. Of course for this the special vacuum Liebig tubes with or without grids may also be employed. An ordinary arc lamp with oxide electrodes may be introduced on the ground and the positive pole is not directly connected with the collecting balloon, but through the upper incandescent cathode or over a condenser. The method of connecting the incandescent cathode floating in the air may be seen in Figs. 29–33.

B is the air balloon, K a Cardan ring (connection with the hawser) C the balloon, L a good conducting cable, P a positive pole, N negative incandescent cathode, and E earth conductor.

Fig. 29 represents the simplest form of construction. If electric oscillations are produced below on the ground by means of a carbon arc lamp or in other suitable way a considerably greater electric resistance is opposed to that in the direct way by inserting an electrical inductive resistance 9. Consequently between P and N a voltage is formed, and as, over N and P only an inductionless ohmic resistance is present, a spark will spring over so long as the separate induction co-efficients and the like are correctly calculated. The consequence of this is that the oxide electrode (carbon or the like) is rendered incandescent and then shows as incandescent cathode an increased collecting effect. The positive poles must be substantially larger than the negative in order that they may not also become incandescent. As they are further connected with the large balloon area which has a large capacity and is charged at high voltage, an incandescent body which is held floating in the air and a positive pole which can collect large capacities is thereby obtained in the simplest way. The incandescent cathode is first caused to become incandescent by means of separate energy produced on the earth, and then maintained by the energy collected from the atmosphere.

Fig. 30 only shows the difference that instead of a round balloon a cigar shaped one (of metal or metalized fabric) may be employed and also a condenser 5 is inserted between the incandescent cathode and the earth conductor so that a short circuited oscillation circuit over P. N 5 and 9 is obtained. This has the advantage that quite small quantities of electricity cause the cathode to become incandescent and much larger cathode bodies may be rendered incandescent.

In this form of construction both the in-

candescent cathode and also the positive electrode may be enclosed in a vacuum chamber as may be seen in Fig. 32. A cable L is carried well insulated through the cover of a vessel and ends in a condenser disc 5. The cover is arched in order to keep off the rain. The vessel is entirely or partially made of magnetic metal and well insulated inside and outside. Opposite the disc 5 another disc 6 and on this again a metallic positive pole of the vacuum tube *g* with the incandescent cathode (oxide electrode) N is arranged. The negative electrode is on the one hand connected with the earth conductor E, and on the other hand with the inductive resistance 9 which is also connected with the cable L with the positive pole and wound round the vessel in coils. The action is exactly the same as that in Fig. 29 only instead of an open incandescent cathode one enclosed in vacuo is employed. As in such collectors only small bodies can be brought to incandescence in large installations a plurality of such vacuum tubes must be inserted in proximity to one another. According to the previous constructions Figs. 31 and 33 are quite self evident without further explanations.

Figs. 34–37 represent further diagrams of connections over radiating and flame collectors, and in fact, how they are to be arranged on the ground.

Fig. 34 shows an arc light collector with oxide electrodes for direct current and its connection; Fig. 35 a similar one for alternating current, Fig. 36 an incandescent collector with a Nernst lamp and Fig. 37 a similar one with a gas flame.

The positive pole 1 of the radiating collectors is always directly connected to the aerial collecting conductor A. In Fig. 34 this is further connected over the condenser battery 5 with a second positive electrode 3. The direct current dynamo *b* produces current which flows over between the electrodes 3 and 2 as an arc light. On the formation of an arc the negative incandescent electrode 2 absorbs electricity from the positive poles standing opposite it and highly charged with atmospheric electricity and conveys the same to the working circuit. The spark gap 7, inductive resistance 9 and induction coil 10 are like the ones previously described. The protective electromagnet S guards the installation against earth circuiting, the safety spark gap 8 from excess voltage or overcharging.

In Fig. 35 the connection is so far altered that the alternating current dynamo feeds the exciting coil 11 of the induction condenser. 12 is its negative and 13 its positive pole; if the coil 3 on the magnet core of the dynamo is correctly calculated and the periodicity of the alternating current is sufficiently high an arc light can be formed between the two poles 1 and 2. As the cathode 2 is connected with the negatively charged earth, and therefore always acts as a negative pole, a form of rectification of the alternating current produced by the dynamo 3 is obtained, the second half of the period is always suppressed. The working circuit may be carried out in the same way as in Fig. 34; the working spark gap 7 may however be dispensed with, and instead thereof between the points *n* and *m* a condenser 5 and an induction resistance 9 may be inserted from which the current is taken inductively.

Fig. 36 represents a form of construction similar to Fig. 34 only that here instead of an arc lamp a Nernst incandescent body is employed. The Nernst lamp is fed through the battery 3. The working section is connected with the negative pole, the safety spark gap with the + poles. The working spark gap 7 may also be dispensed with and the current for it taken at 12 over the oscillation circuit 5, 11 (shown in dotted lines).

Flame collectors (Fig. 37) may also be employed according to this invention. The wire network 1 is connected with the aerial collector conductor A and the burner with the earth. At the upper end of the latter, long points are provided which project into the flame. The positive electrode is connected with the negative over a condenser 5 and the induction coil 9 with the earth.

The novelty in this invention is firstly, the use of incandescent cathodes opposite positive poles which are connected with large metallic capacities as automatic collecting surfaces, (2) the connection of the incandescent cathodes with the earth whereby, in addition to the electricity conveyed to them from the battery or machine which causes the incandescing, also the negative charge of the earth potential is conveyed, and (3) the connection of the positive and negative poles of the radiating collectors over a condenser circuit alone or with the introduction of a suitable inductive resistance, whereby simultaneously an oscillatory oscillation circuit may be obtained. The collecting effect is by these methods quite considerably increased.

I declare that what I claim is:—

1. An electrical energy generating system, comprising a conducting surface for static charges, means to support same at a distance above the earth, a conductor leading to the earth level, a spark gap associated with said conductor to convert electrostatic charges into electromagnetic high frequency oscillations means to supply said electromagnetic energy to a net work, and a spark gap of greatly increased relative resistance in parallel therewith.

2. An electrical energy generating system

comprising a conductor, means to support same above the earth level, an inductance therein, a spark gap associated with said conductor, a second spark gap of much higher relative resistance in parallel therewith and an energy receiving circuit coupled with the spark gap of lesser resistance.

3. An electrical energy generating system comprising a collecting surface, means to support same above the earth level, a conductor connecting said collecting surface with the earth level, a choke in said conductor, an electromagnetic resistance converting electrostatic energy to electromagnetic energy, a safety higher resistance in parallel therewith and a net work coupled with the conversion resistance of lesser value.

4. An electrical energy generating system comprising electric conductors spaced above the earth to form electromagnetic oscillating circuits, conductors connecting to earth level, electrostatic to electromagnetic energy conversion means therein, a safety high electrostatic resistance in parallel therewith and means to alter the electromagnetic characteristic of the circuits.

5. An electrical energy generating system comprising in combination a static collecting surface arranged above the earth, conductors connecting to earth level, a pair of spark gaps in parallel of different electrostatic resistance, a utilization net work shunted across the spark gap of lesser resistance and an electromagnetic choke in said conductors.

6. An electrical energy generating system comprising an open circuit energy collecting aerial, a pair of sparking gaps in parallel of widely different resistance, connected thereto and a closed electric oscillation circuit in shunt across the gap of lesser resistance.

7. An electrical energy generating system comprising an open circuit energy collecting aerial, a pair of sparking gaps in parallel of widely different resistance connected thereto, a closed electric oscillation circuit in shunt across the gap of lesser resistance, a plurality of electrostatic collecting surfaces, means to connect said collecting surfaces in parallel in groups and means to connect said groups symmetrically with said aerial.

In witness whereof, I have hereunto signed my name this 30 day of Dec., 1920, in the presence of two subscribing witnesses.

HERMANN PLAUSON.

Witnesses:
H. F. ARMSTRONG,
W. H. BEESTON.

Please note that this is a re-worded excerpt from this patent. It describes in considerable detail, different methods for abstracting useable electrical power from passive aerial systems. He describes a system with 100 kilowatt output as a “small” system.

US Patent 1,540,998 9th June 1925 Inventor: Hermann Plauson

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Be it known that I, Hermann Plauson, Estonian subject, residing in Hamburg, Germany, have invented certain new and useful improvements in the Conversion of atmospheric Electric Energy, of which the following is a specification.

According to this invention, charges of atmospheric electricity are not directly converted into mechanical energy, and this forms the main difference from previous inventions, but the static electricity which runs to earth through aerial conductors in the form of direct current of very high voltage and low current strength is converted into electro-dynamic energy in the form of high frequency vibrations. Many advantages are thereby obtained and all disadvantages avoided.

The very high voltage of static electricity of a low current strength can be converted by this invention to voltages more suitable for technical purposes and of greater current strength. By the use of closed oscillatory circuits it is possible to obtain electromagnetic waves of various amplitudes and thereby to increase the degree of resonance of such current. Such resonance allows various values of inductance to be chosen which, by tuning the resonance between a motor and the transformer circuit, allows the control of machines driven by this system. Further, such currents have the property of being directly available for various uses, other than driving motors, including lighting, heating and use in electro-chemistry.

Further, with such currents, a series of apparatus may be fed without a direct current supply through conductors and the electro-magnetic high frequency currents may be converted by means of special motors, adapted for electro-magnetic oscillations, into alternating current of low frequency or even into high voltage direct current.

DESCRIPTION OF THE DRAWINGS

Fig. 1.

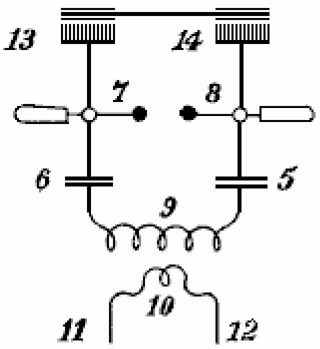

Fig.1 is an explanatory figure

Fig. 2.

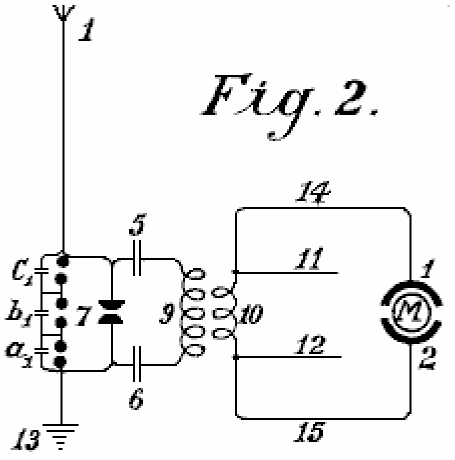
1
14
5
11
C1
b1
7
9
10
M
1
a1
12
2
6
13
15

Fig.2 is a diagrammatic view of the most simple form.

Fig. 3.

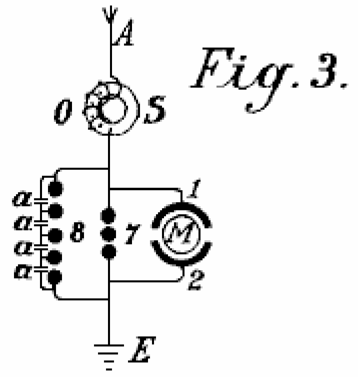
A
0
S
1
a
a
a
a
8
7
M
2
E

Fig.3 shows a method of converting atmospheric electrical energy into a form suitable for use with motors.

Fig. 4.

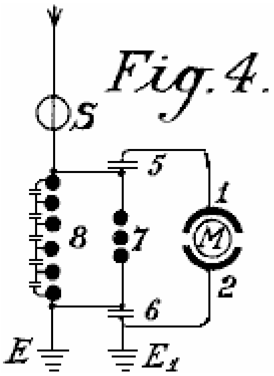
S
5
1
8
7
M
2
6
E
E1

Fig.4 is a diagram showing the protective circuitry.

Fig. 5.

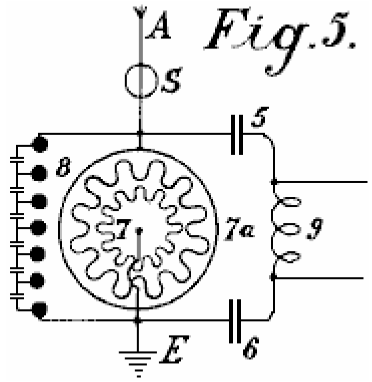
A
S
5
8
7
7a
9
E
6

Fig.5 is a diagram of an arrangement for providing control

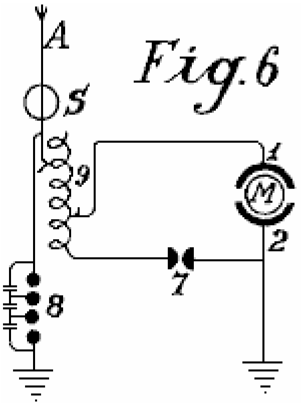
A
Fig.6
S
9
1
M
2
7
8

Fig.6 is an arrangement including a method of control

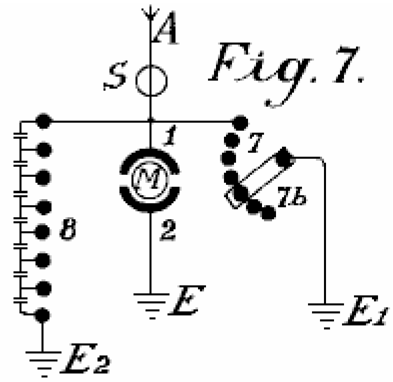
A
Fig. 7.
S
1
M
7
7b
8
2
E
E1
E2

Fig.7 shows how the spark gap can be adjusted

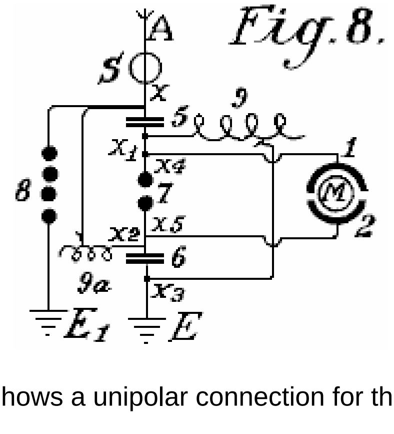
A
Fig.8.
S
X
5
9
X1
X4
1
M
7
8
X5
2
X2
6
9a
X3
E1
E

Fig.8 shows a unipolar connection for the motor

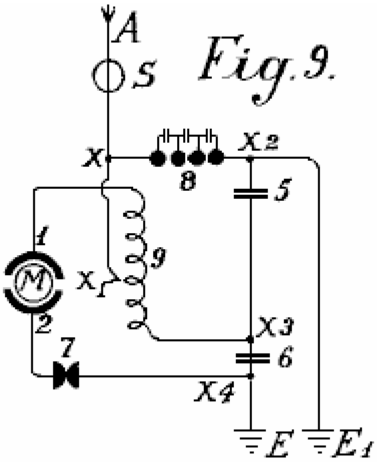
A
S
Fig.9.
X
X2
8
5
1
M
9
X1
2
7
X3
6
X4
E
E1

Fig.9 shows a weak coupled system suitable for use with small power motors

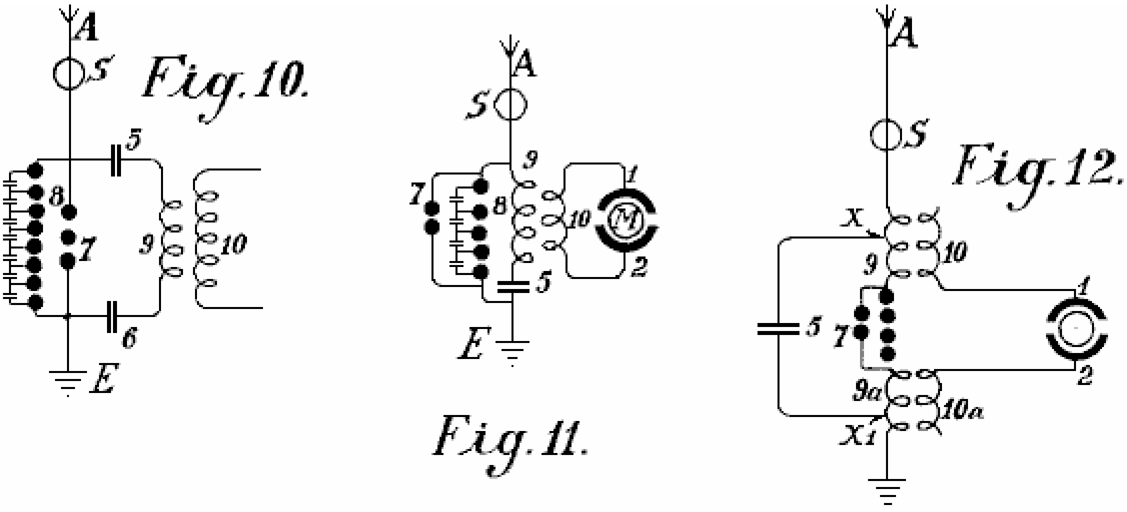
A
S
Fig.10.
5
8
7
9
10
6
E
A
S
9
7
8
10
1
M
2
5
E
Fig.11.
A
S
Fig.12.
X
10
9
5
7
1
2
9a
10a
X1

Fig.10, Fig.11 and Fig.12 show modified arrangements

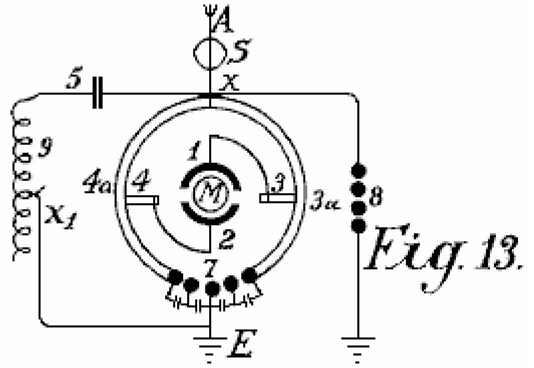
A
S
5
X
9
1
4a
4
M
3
3a
8
X1
2
7
Fig.13.
E

Fig.13 shows a form of inductive coupling for the motor circuit

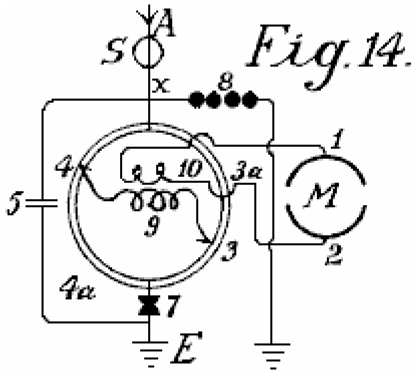
A
S
Fig. 14.
x
8
1
4
10
3a
5
9
M
3
2
4a
7
E

Fig.14 is a modified form of Fig.13 with inductive coupling.

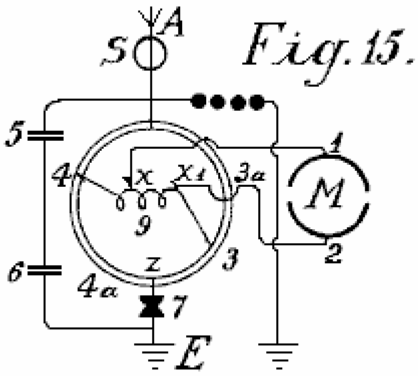
A
S
Fig. 15.
5
1
4
x
x1
3a
M
9
2
6
z
3
4a
7
E

Fig.15 is an arrangement with non-inductive motor

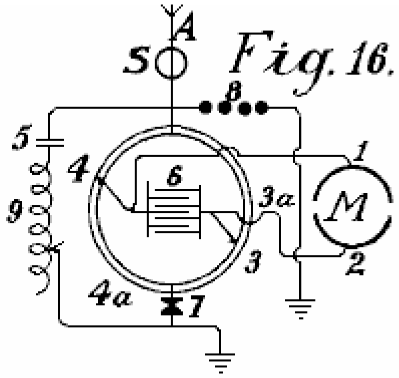
A
S
Fig. 16.
8
5
1
4
6
3a
9
M
2
3
4a
7

Fig.16 is an arrangement with coupling by capacitor.

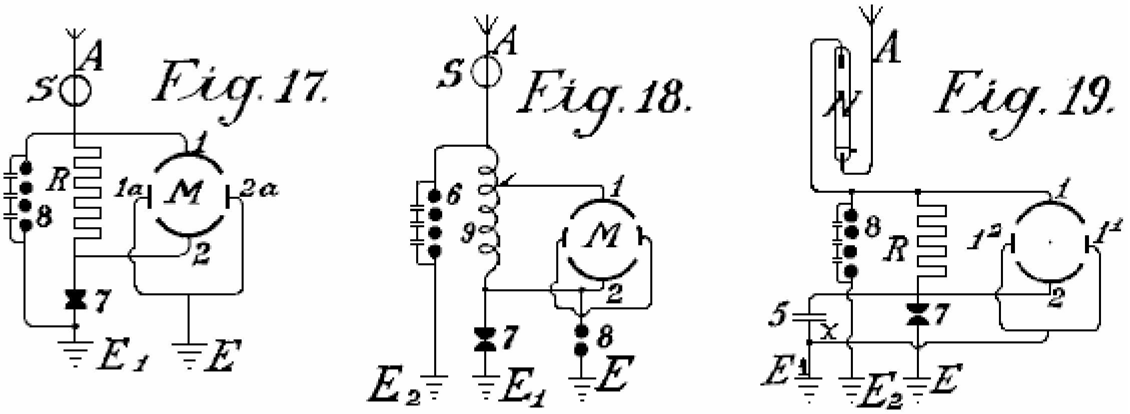
A
S
Fig. 17.
1
R
1a
M
2a
8
2
7
E1
E
A
S
Fig. 18.
6
1
9
M
2
7
8
E2
E1
E
A
N
Fig. 19.
1
8
R
1²
1¹
2
5
x
7
E1
E2
E

Fig.17, **Fig.18** and **Fig.19** are diagrams showing further modifications

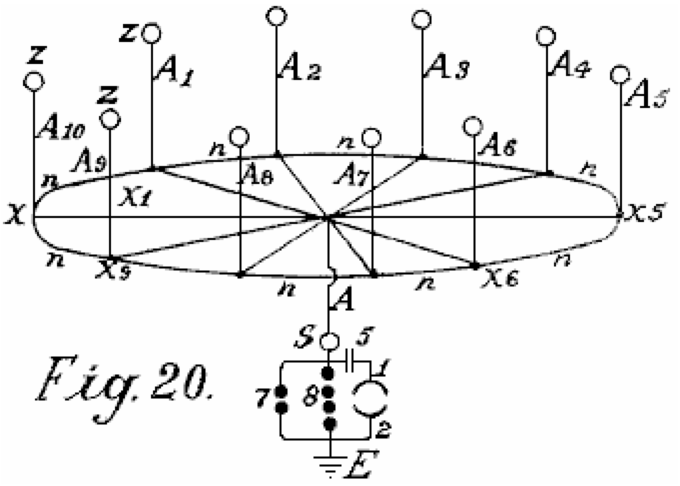

Fig. 20.

z
z
z
A1
A2
A3
A4
A5
A10
A9
A8
A7
A6
n
X
X1
X5
X9
X6
A
S
5
1
7
8
2
E

Fig.20 shows a simple form in which the aerial network is combined with special collectors

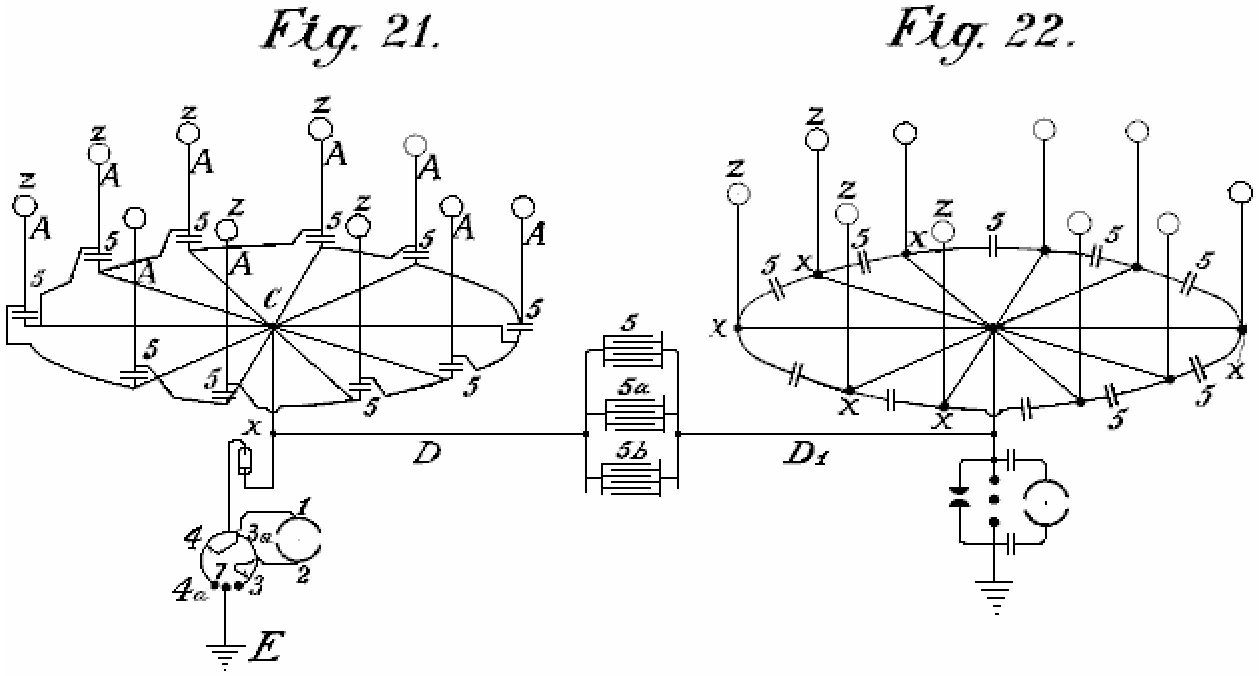

Fig. 21.

Fig. 22.

z
A
5
C
x
D
5a
5b
D1
1
2
3a
4
4a
7
3
E

Fig.21 shows diagramatically, an arrangement suitable for collecting large quantities of energy.
Fig.22 is a modified arrangement having two rings of collectors

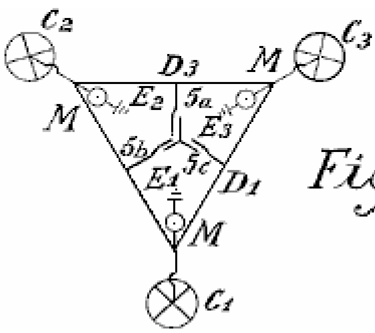

Fig. 23.

C2
C3
D3
M
E2
5a
E3
5b
5c
E1
D1
C1

Fig.23 shows the connections for three rings of collectors

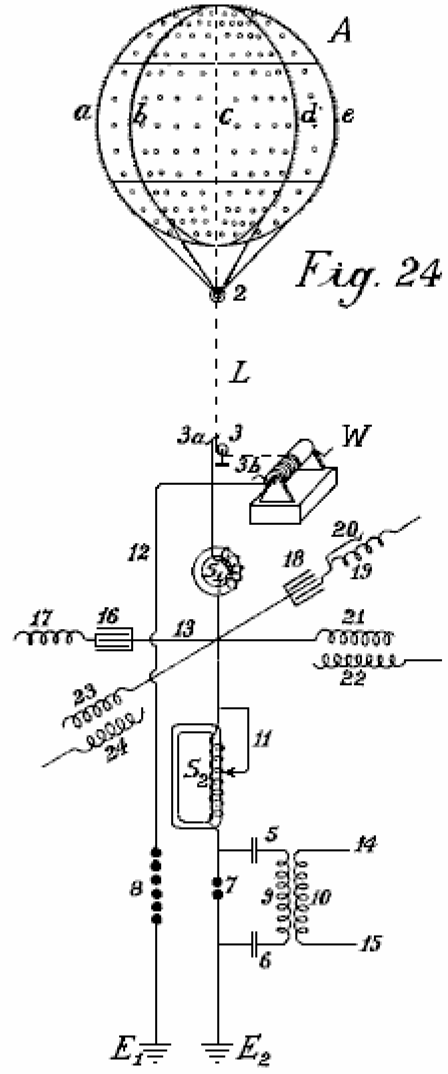

Fig. 24

A
a
b
c
d
e
2
L
3a
3
W
3b
12
18
20
19
S_1
17
16
13
21
22
23
24
11
S_2
5
14
8
7
9
10
6
15
E_1
E_2

Fig.24 shows a collecting balloon and diagram of its battery of capacitors

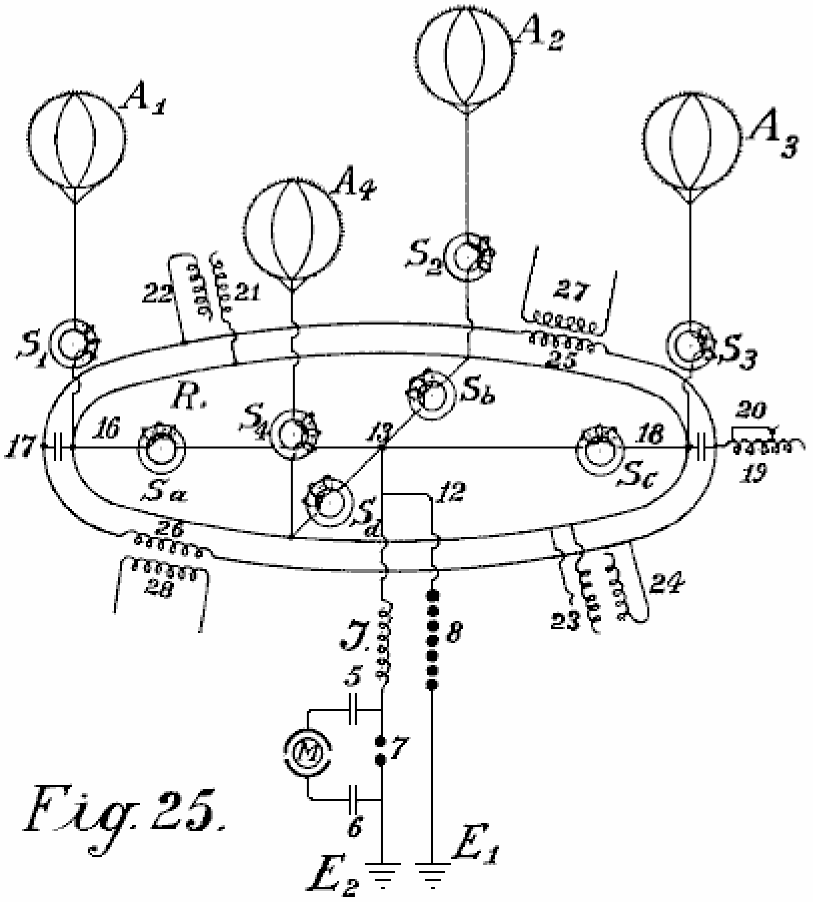
A1
A2
A4
A3
S2
27
22
21
S1
25
S3
R.
Sb
S4
20
17
16
13
18
Sa
Sc
19
Sd
12
26
28
23
24
J.
8
5
M
7
6
E2
E1

Fig. 25.

Fig.25 and Fig.26 show modified collector balloon arrangements.

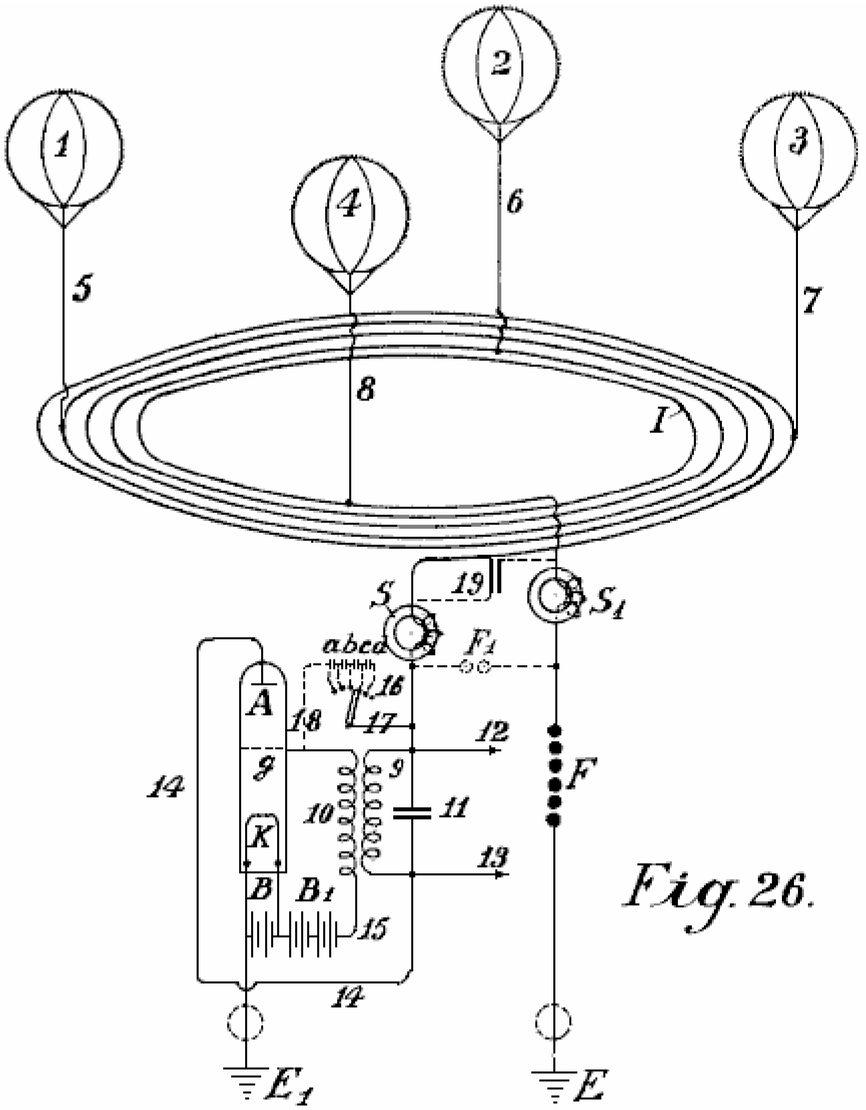
1
2
3
4
5
6
7
8
I
S
19
S1
a b c d
F1
A
16
18
17
12
g
9
14
10
11
K
13
F
B
B1
15
14
E1
E

Fig. 26.

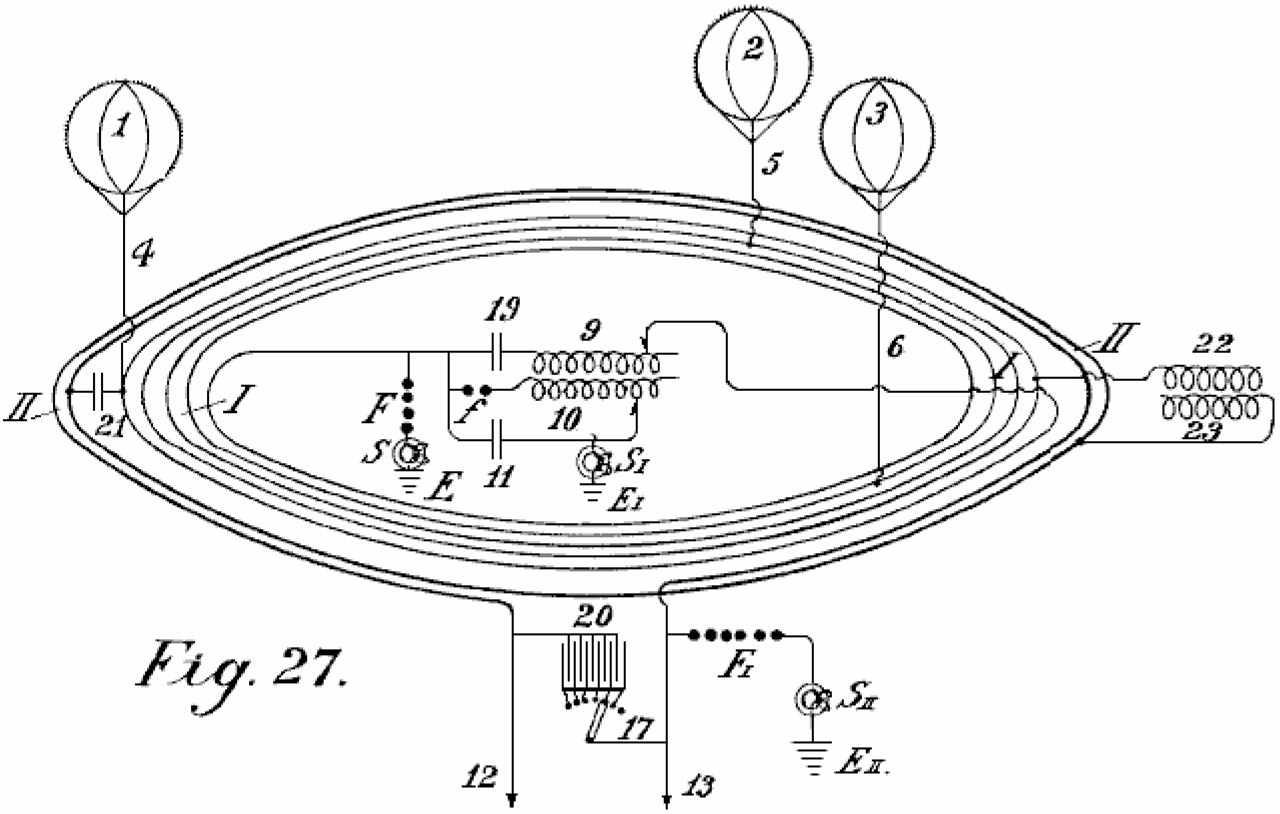
1
2
3
4
5
6
19
9
II
22
II
I
I
21
F
f
10
23
S
E
11
S1
E1
20
F1
SII
17
EII.
12
13

Fig. 27.

Fig.27 shows a second method of connecting conductors for the balloon aerials.

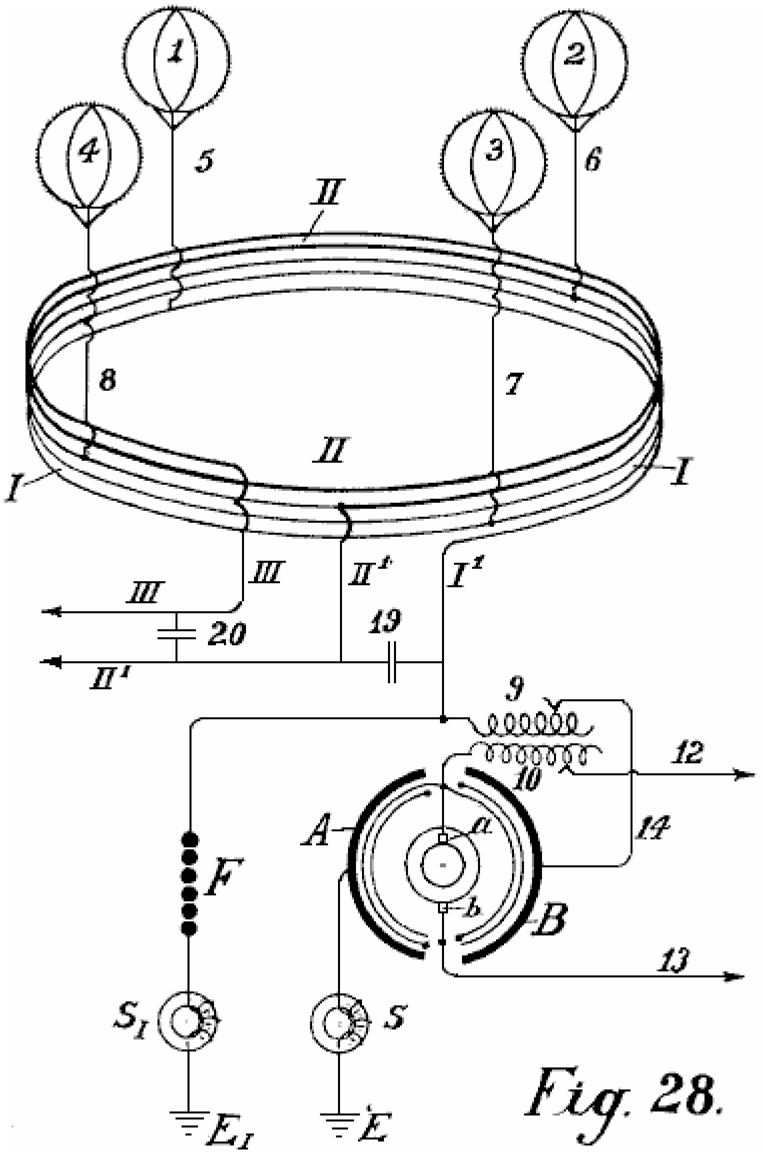
1
2
4
3
5
6
II
8
7
II
I
I
III
II¹
I¹
III
20
19
II¹
9
12
10
A
a
14
F
b
B
13
S₁
S
E₁
E
Fig. 28.

Fig.28 shows an auto-transformer method of connection.

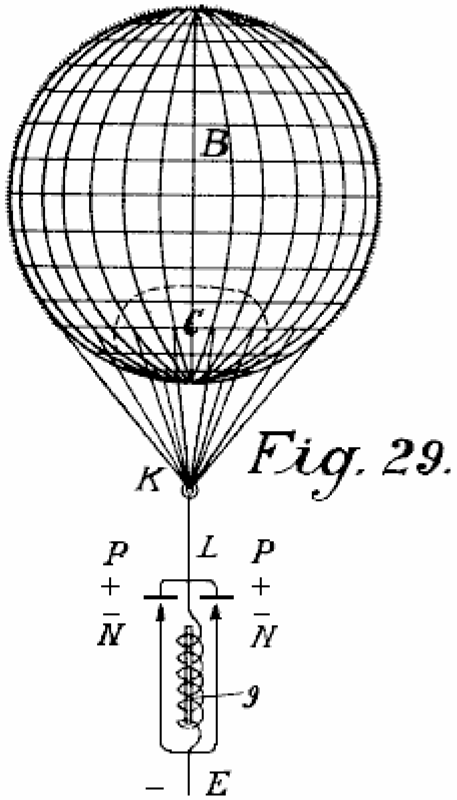
B
C
Fig. 29.
K
L
P
+
–
N
P
+
–
N
9
–
E

Fig.29 shows the simplest form of construction with incandescent cathode.

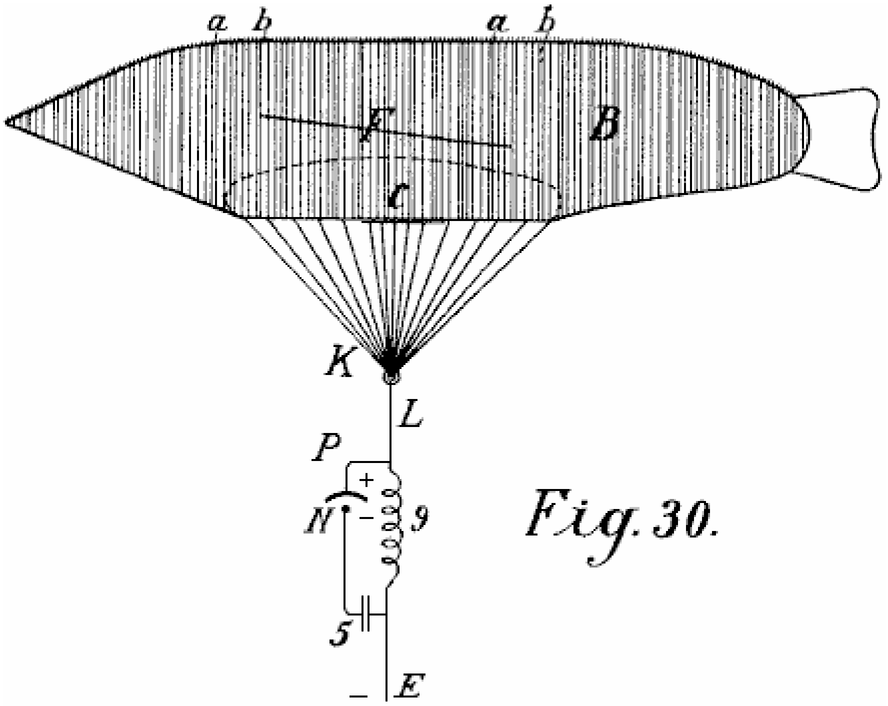
a
b
a
b
F
B
C
K
L
P
+
N
−
9
Fig. 30.
5
−
E

Fig.30 shows a form with a cigar-shaped balloon.

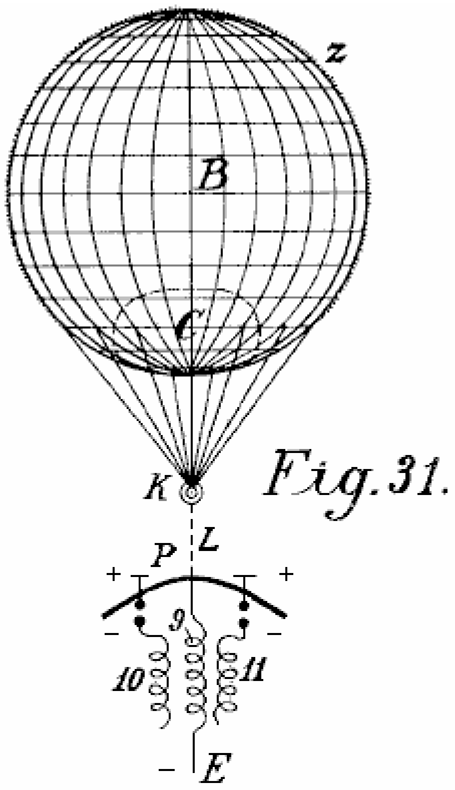
z
B
C
K
Fig. 31.
P
L
+
+
−
−
9
10
11
−
E

Fig.31 is a modified arrangement.

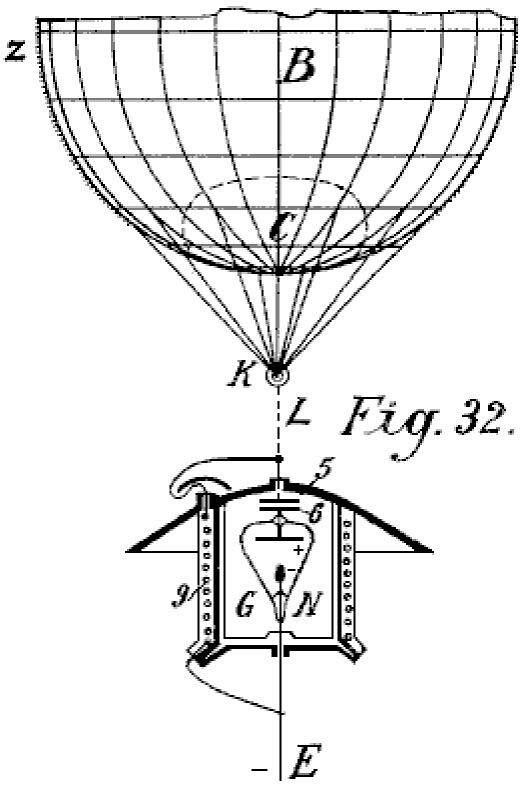
z
B
C
K
L
5
6
+
−
9
G
N
−
E

Fig.32 shows a form with cathode and electrode enclosed in a vacuum chamber.

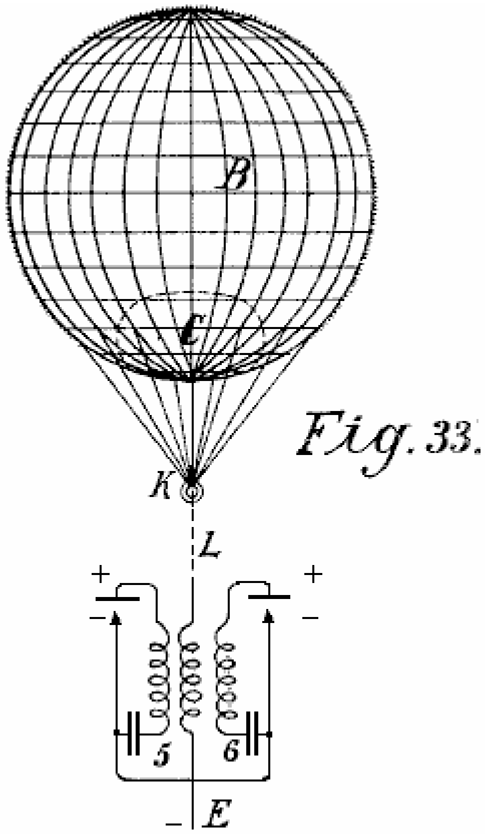

Fig.33 is a modified form of Fig.32

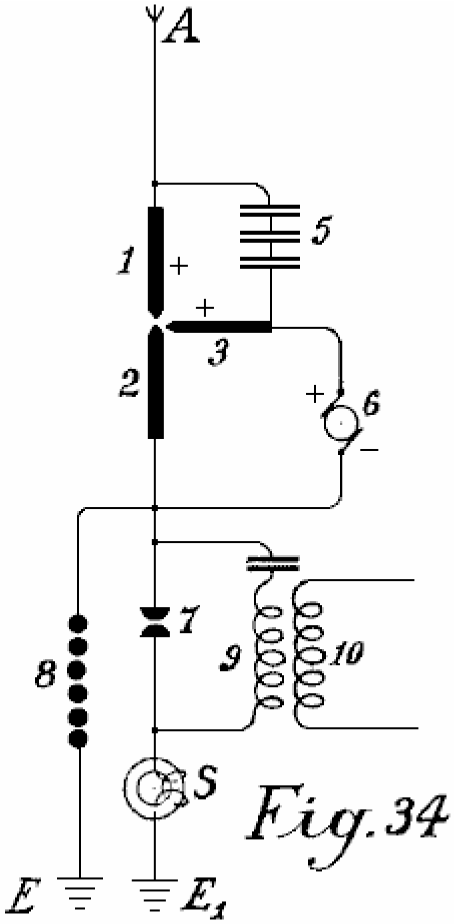
A
1
+
+
5
3
2
+
6
−
8
7
9
10
S
E
E1

Fig.34

Fig.34 shows an arc light collector.

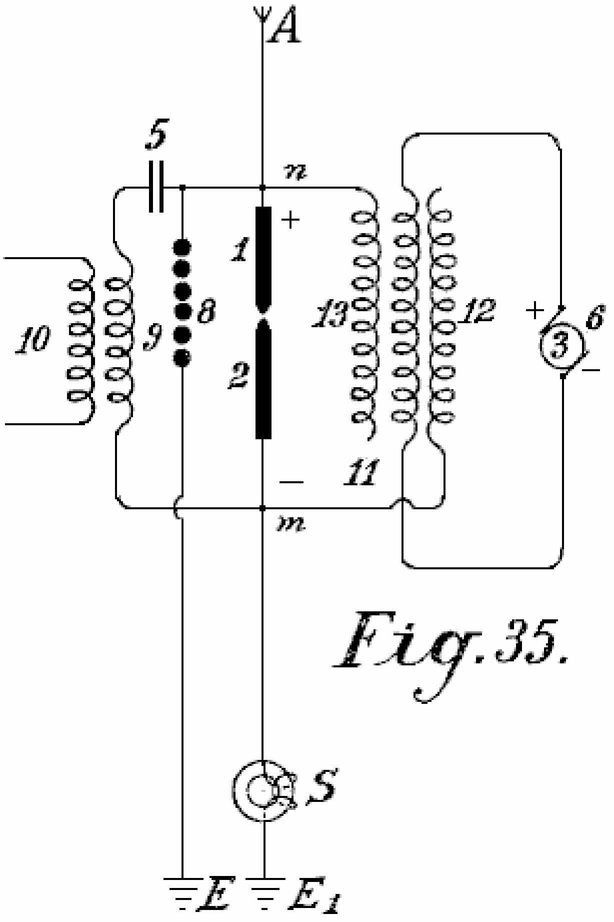
A
5
n
+
1
8
10
9
13
12
+
6
3
−
2
−
11
m
S
E
E1

Fig.35.

Fig.35 shows such an arrangement for alternating current

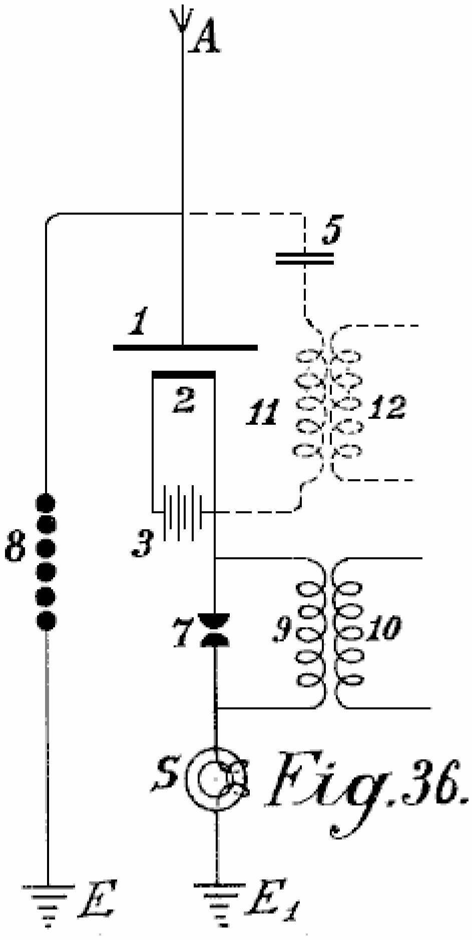
A
5
1
2
11
12
3
8
7
9
10
S
E
E1

Fig.36.

Fig.36 shows an incandescent collector with Nernst lamp

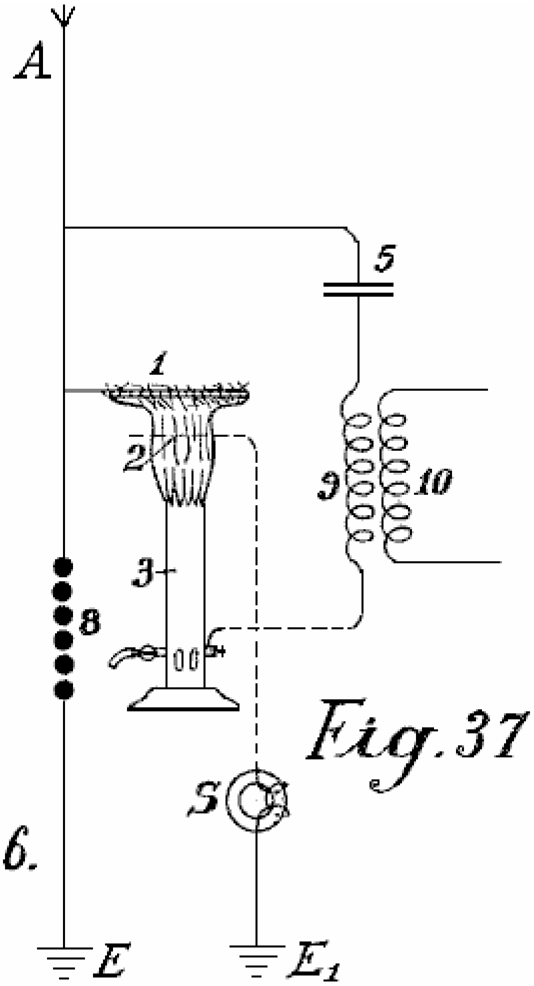
A
5
1
2
9
10
3
8
S
6.
E
E1

Fig.37

Fig.37 shows a form with a gas flame.

Fig. 1.

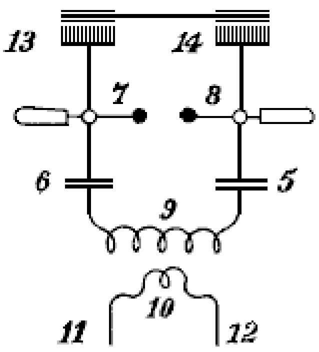

Fig.1 illustrates a simple diagram for converting static electricity into dynamic energy of a high number of oscillations. For the sake of clarity, a Wimshurst machine is assumed to be employed and not an aerial antenna. Items **13** and **14** are combs for collecting the static electricity of the influence machine. Items **7** and **8** are spark-discharging electrodes. Items **5** and **6** are capacitors, **9** is the primary winding of an inductive coil, **10** is the secondary winding whose ends are **11** and **12**. When the disc of the static influence machine is rotated by mechanical means, the combs collect the electric charges, one being positive and one negative and these charge the capacitors **5** and **6** until such a high voltage is developed across the spark gap **7-- 8** that the spark gap is jumped. As the spark gap forms a closed circuit with capacitors **5** and **6**, and inductive resistance **9**, as is well known, waves of high frequency electromagnetic oscillations will pass in this circuit.

The high frequency of the oscillations produced in the primary circuit induces waves of the same frequency in the secondary circuit. Thus, in the primary circuit, electromagnetic oscillations are formed by the spark and these oscillations are maintained by fresh charges of static electricity.

By suitably selecting the ratio between the number of turns in the primary and secondary windings, with regard to a correct application of the coefficients of resonance (capacitance, inductance and resistance) the high voltage of the primary circuit may be suitably converted into a low voltage high current output.

When the oscillatory discharges in the primary circuit become weaker or cease entirely, the capacitors are charged again by the static electricity until the accumulated charge again breaks down across the spark gap. All this is repeated as long as electricity is produced by the static machine through the application of mechanical energy to it.

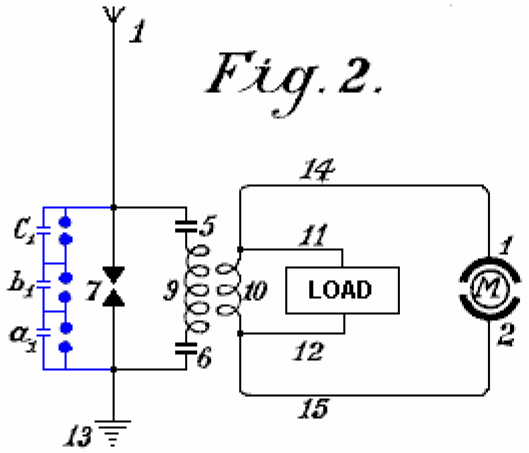

An elementary form of the invention is shown in **Fig.2** in which two spark gaps in parallel are used, one of which may be termed the working gap **7** while the second serves as a safety device for excess voltage and consists of a larger number of spark gaps than the working section, the gaps being arranged in series and which are bridged by very small capacitors a_1, b_1, c_1, which allow uniform sparking in the safety section.

1 is the aerial antenna for collecting charges of atmospheric electricity, **13** is the earth connection of the second part of the spark gap, **5** and **6** are capacitors and **9** is the primary coil winding. When the positive atmospheric electricity seeks to combine with the negative earth charge via aerial **1**, this is prevented by the air gap between the spark gaps. The resistance of spark gap **7** is lower than that of the safety spark gap set of three spark gaps connected in series a which consequently has three times greater air resistance.

Therefore, so long as the resistance of spark gap **7** is not overloaded, discharges take place only through it. However, if the voltage is increased by any influence to such a level that it might be dangerous for charging the capacitors **5** and **6**, or for the coil insulation of windings 9 and 10, the safety spark gap set will, if correctly set, discharge the voltage directly to earth without endangering the machine. Without this second spark gap arrangement, it is impossible to collect and render available large quantities of electrical energy.

The action of this closed oscillation circuit consisting of spark gap **7**, two capacitors **5** and **6**, primary coil **9** and secondary coil **10**, is exactly the same as that of **Fig.1** which uses a Wimshurst machine, the only difference being the provision of the safety spark gap. The high frequency electromagnetic alternating current can be tapped off through the conductors **11** and **12** for lighting and heating purposes. Special motors adapted for working with static electricity or high frequency oscillations may be connected at **14** and **15**.

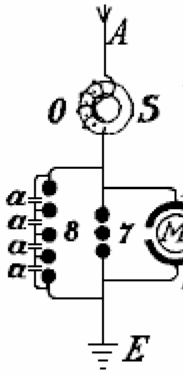

Fig. 3.

In addition to the use of spark gaps in parallel, a second measure of security is also necessary for taking the current from this circuit. This is the introduction of protective electromagnets or choking coils in the aerial circuit as shown by **S** in **Fig.3**. A single electromagnet having a core of the thinnest possible separate laminations is connected with the aerial. In the case of high voltages in the aerial network or at places where there are frequent thunderstorms, several such magnets may be connected in series.

In the case of large units, several such magnets can be employed in parallel or in series parallel. The windings of these electromagnets may be simply connected in series with the aerials. In this case, the winding preferably consists of several thin parallel wires, which together, make up the necessary cross-sectional area of wire. The winding may be made of primary and secondary windings in the form of a transformer. The primary winding will then be connected in series with the aerial network, and the secondary winding more or less short-circuited through a regulating resistor or an induction coil. In the latter case it is possible to regulate, to a certain extent, the effect of the choking coils. In the following circuit and constructional diagrams , the aerial electromagnet choke coil is indicated by a simple ring **S**.

Fig.3 shows the most simple way of converting atmospheric electricity into electromagnetic wave energy by the use of special motors adapted for high oscillatory currents or static charges of electrical energy. Recent improvements in motors for working with static energy and motors working by resonance, that is to say, having groups of tuned electromagnetic co-operating circuits render this possible but such do not form part of the present invention.

A motor adapted to operate with static charges, will for the sake of simplicity, be shown in the diagrams as two semi-circles **1** and **2** and the rotor of the motor by a ring **M** (**Fig.3**). **A** is a vertical aerial or aerial network. **S** is the safety choke or electromagnet with coil **O** as may be seen is connected with the aerial **A**. Adjacent to the electromagnet **S**, the aerial conductor is divided into three circuits, circuit **8** containing the safety spark gap, circuit **7** containing the working spark gap, and then a circuit containing the stator terminal **1**, the rotor and stator terminal **2** at which a connection is made to the earth wire. The two spark gaps are also connected metallically with the earth wire. The method of working in these diagrams is as follows:

The positive atmospheric electric charge collected tends to combine with the negative electricity (or earth electricity) connected via the earth wire. It travels along the aerial **A** through the electromagnet **S** without being checked as it flows in the same direction as the direct current. Further, its progress is arrested by two spark gaps placed in the way and the stator capacitors. These capacitors charge until their voltage exceeds that needed to jump the spark gap **7** when a spark occurs and an oscillatory charge is obtained via the closed oscillation circuit containing motor **M**. The motor here forms the capacity and the necessary inductance and resistance, which as is well known, are necessary for converting static electricity into electromagnetic wave energy.

The discharges are converted into mechanical energy in special motors and cannot reach the aerial network because of the electromagnet or choke. If, however, when a spark occurs at spark gap **7**, a greater quantity of atmospheric electricity tends to flow to earth, then a counter voltage is induced in the electromagnet, which is greater the more rapidly and strongly the flow of current direct to earth is. This opposing voltage causes the circuit to exhibit a sufficiently high resistance to prevent a short circuit between the atmospheric electricity and the earth.

The circuit containing spark gap **8**, having a different wave length which is not in resonance with the natural frequency of the motor, does not endanger the motor and serves as security against excess voltage, which, as practical experiments have shown, may still arise in certain cases.

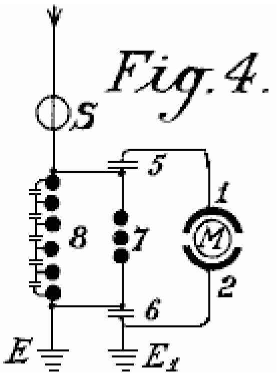
Fig.4.

In **Fig.4**, spark gap **7** is shunted across capacitors **5** and **6** from the motor **M**. This arrangement provides improved over-voltage protection for the motor and it gives a uniform excitation through the spark gap **7**.

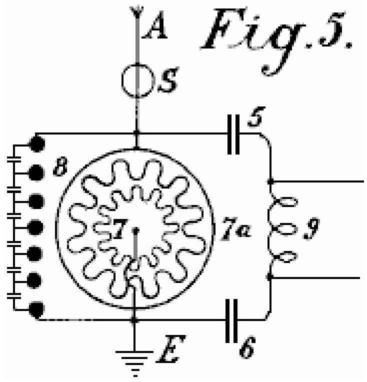
Fig.5.

Fig.5 shows an arrangement for producing large currents which can be used direct without motors, to provide heating and lighting. The main difference here is that the spark gap consists of a star-shaped disc **7** which can rotate on its own axis and is rotated by a motor opposite similarly fitted electrodes **7a**. When separate points of starts face one another, discharges take place, thus forming an oscillation circuit with capacitors **5** and **6** and inductor **9**. It is evident that a motor may also be connected directly to the ends of inductor **9**.

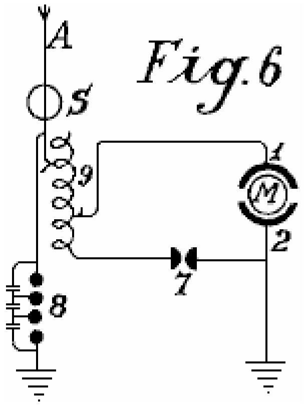
Fig.6

Fig.6 shows how the oscillation circuit may have a motor connected via a variable inductor which opposes any excess voltages which might be applied to the motor. By cutting the separate coils **9** (coupled inductively to the aerial) in or out, the inductive action on the motor may be more or less increased, or variable aerial action may be exerted on the oscillation circuit.

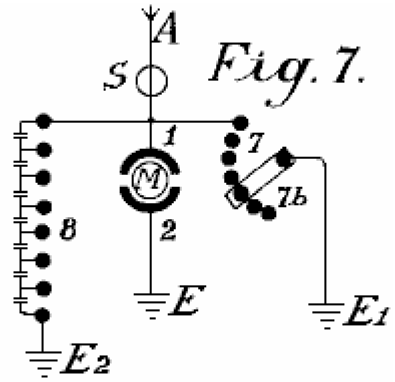

In **Fig.7** the oscillation circuit is closed through the earth (**E** and **E$_1$**). The spark gap **7** may be increased or reduced by means of a contact arm **7b**.

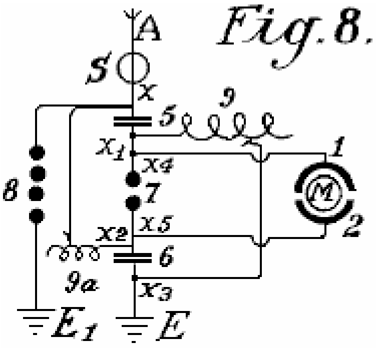

Fig.8 shows a unipolar connection of the motor with the aerial network. Here, two oscillation circuits are closed through the same motor. The first oscillation circuit passes from aerial **A** through electromagnet **S**, point **x**, inductance **9a** to the earth capacitor **6**, across spark gap **7** to the aerial capacitor **5** and back to point **x**. The second oscillation circuit starts from the aerial **5** at the point **x1** through inductor **9** to the earth capacitor **6** at the point **x3**, through capacitor **6**, across spark gap **7** back to point **x1**. The motor itself, is inserted between the two points of spark gap **7**. This arrangement produces slightly dampened oscillation wave currents.

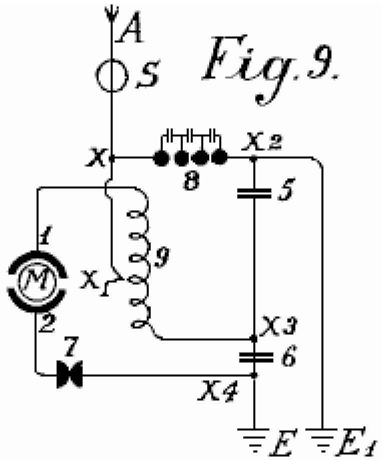

Fig.9 shows a loosely coupled system intended for small motors for measuring purposes. **A** is the serial, **S** is the electromagnet or aerial inductor, **9** the inductor, **7** the spark gap, **5** and **6** capacitors, **E** the earth, **M** the motor, and **1** and **2** the stator connections of the motor which is directly connected to the oscillator circuit.

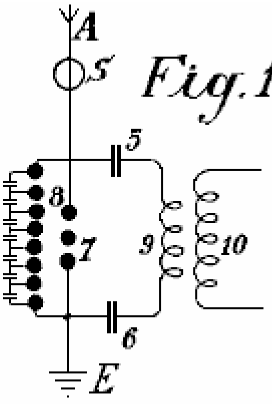

Fig.10.

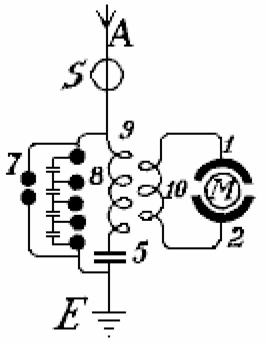

Fig.11.

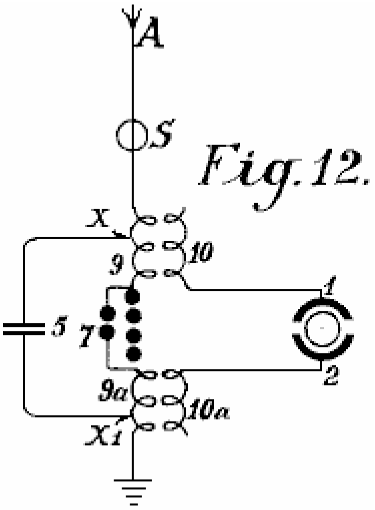

Fig.12.

Fig.10 shows a motor circuit with purely inductive coupling. The motor is connected with the secondary wire **10** as may be seen in **Fig.11** in a somewhat modified circuit. The same applies to the circuit of **Fig.12**.

The circuit diagrams shown so far, allow motors of small to medium strength to be operated. For large aggregates, however, they are too inconvenient as the construction of two or more oscillation circuits for large amounts of energy is difficult; the governing is still more difficult and the danger in switching on or off is greater.

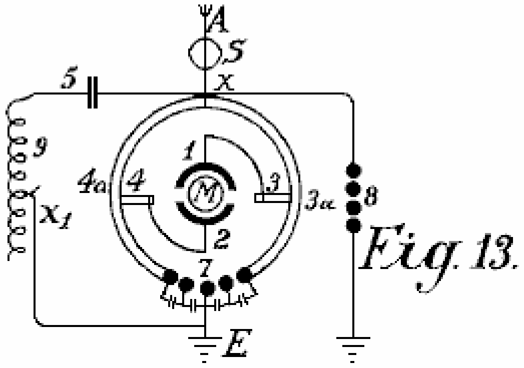

Fig. 13.

A means for overcoming such difficulties is shown in **Fig.13**. The oscillation circuit shown here, runs from point **x** over capacitor **5**, variable inductor **9**, spark gap **7** and the two segments **3a** and **3b** forming arms of a Wheatstone bridge, back to **x**. If the motor is connected by brushes **3** and **4** transversely to the two arms of the bridge as shown in the drawing, electromagnetic oscillations of equal sign are induced in the stator surfaces **1** and **2** and the motor does not revolve. If however, the brushes **3** and **4** are moved in common with the conducting wires **1** and **2** which connect the brushes with the stator poles, a certain alteration or displacement of the polarity is obtained and the motor commences to revolve.

The maximum action will result if one brush **3** comes on the central sparking contact **7** and the other brush **4** on the part **x**. In practice however, they are usually brought on to the central contact **7** but only held in the path of the bridge segments **4a** and **3a** in order to avoid connecting the spark gaps with the motor oscillation circuit.

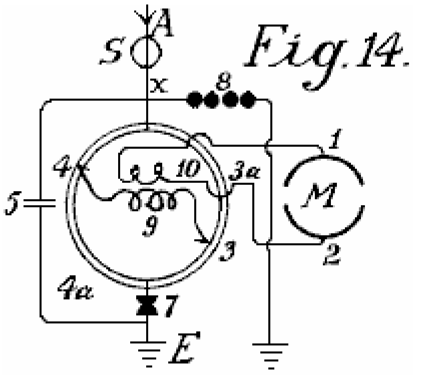

Fig.14.

As this prevents the whole of the oscillation energy acting on the motor, it is better to adopt the modification shown in **Fig.14**. The only difference here is that the motor is not wired directly to the segments of the commutator, but instead it is wired to secondary coil **10** which receives induced current from primary coil **9**. This arrangement provides a good transforming action, a loose coupling and an oscillation circuit without a spark gap.

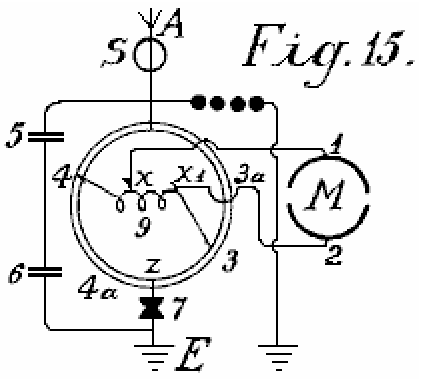
Fig. 15.

In **Fig.15**, the motor is wired directly to the primary coil at **x** and **x1** after the principle of the auto-transformer. In **Fig.16**, instead of an inductor, capacitor **6** replaces the inductance and is inserted between the segments **3a** and **4a**. This has the advantage that the segments **3a** and **4a** need not be made of solid metal, but may consist of spiral coils which allow a more exact regulation, and high inductance motors may be used.

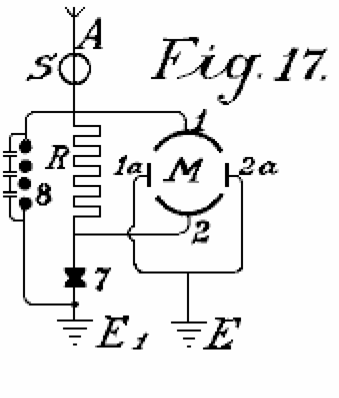
Fig. 17.

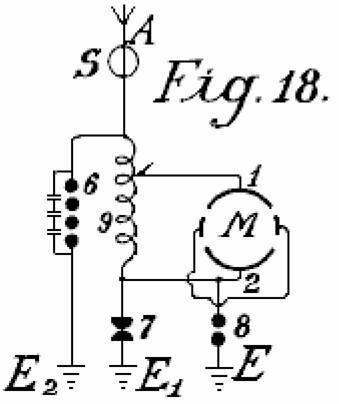
Fig. 18.

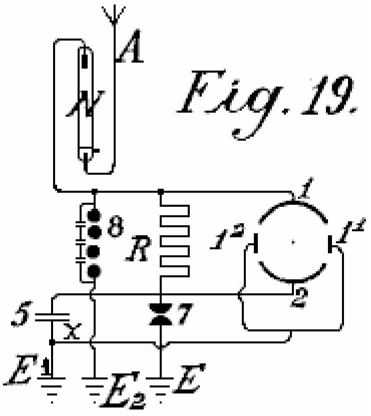
Fig. 19.

The circuits shown in **Fig.17**, **Fig.18** and **Fig.19** may be used with resonance and particularly with induction capacitor motors; between the large stator induction capacitor surfaces, small reversing pole capacitors are connected which are lead together to earth. Such reversing poles have the advantage that, with large quantities of electrical energy, the spark formation between the separate oscillation circuits ceases.

Fig.19 shows another method which prevents high frequency electromagnetic oscillations formed in the oscillation circuit, feeding back to the aerial. It is based on the well known principle that a mercury lamp, one electrode of which is formed of mercury, the other of solid metal such as steel, allows an electric charge to pass in only one direction: from the mercury to the steel and not vice versa. The mercury electrode of the vacuum tube **N** is therefore connected with the aerial conductor and the steel electrode with the oscillation circuit. Charges can then only pass from the aerial through the vacuum tube to the oscillation circuit and no flow occurs in the opposite direction. In practice, these vacuum tubes must be connected behind an electromagnet as the latter alone provides no protection against the danger of lightning.

As regards the use of spark gaps, all arrangements as used for wireless telegraphy may be used. Of course, the spark gaps in large machines must have a sufficiently large surface. In very large stations they are cooled in liquid carbonic acid or better still, in liquid nitrogen or hydrogen; in most cases the cooling may also take place by means of liquefied low homologues of the metal series or by means of hydrocarbons, the freezing point of which lies between -90^{0}C and -40^{0}C. The spark gap casing must also be insulated and be of sufficient strength to be able to resist any pressure which may arise. Any undesirable excess super-pressure which may be formed must be let off automatically. I have employed with very good results, mercury electrodes which were frozen in liquid carbonic acid, the cooling being maintained during the operation from the outside, through the walls.

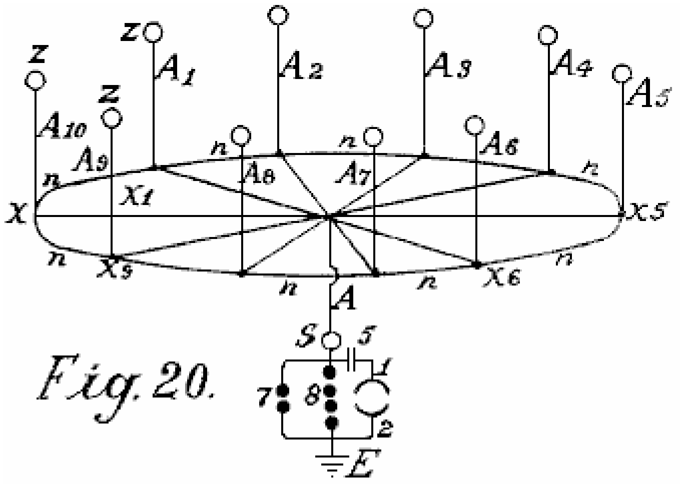

Fig. 20.

Fig.20 shows one of the most simple forms of construction of an aerial network in combination with collectors, transformers and the like. **E** is the earth wire, **8** the safety spark gap, **7** the working spark gap, **1** and **2** the stator surfaces of the motor, **5** a capacitor battery, **S** the protective magnet which is connected with the coil in the aerial conductor, **A**^{1} to **A**^{10} aerial antennae with collecting balloons, **N** horizontal collecting or connecting wires, from which, a number of connections run to the centre.

The actual collectors consist of metal sheaths, preferably made of an aluminium magnesium alloy, and are filled with hydrogen or helium, and are attached to copper-plated steel wires. The size of the balloon is selected so that the actual weight of the balloon and its conducting wire is supported by it. Aluminium spikes, made and gilded as described below, are arranged on top of the balloons in order to produce a conductor action. Small quantities of radium preparations, more particularly, polonium-ionium or mesothorium preparations, considerably increase the ionisation, and the performance of these collectors.

In addition to metal balloons, fabric balloons which are sprayed with a metallic coating according to Schoop's metal-spraying process may also be used. A metallic surface may also be produced by lacquering with metallic bronzes, preferably according to Schoop's spraying process, or lacquering with metallic bronze powders in two electrical series of widely different metals, because this produces a considerably increased collecting effect.

Instead of the ordinary round balloons, elongated cigar-shaped ones may be employed. In order also to utilise the frictional energy of the wind, patches or strips of non-conducting substances which produce electricity by friction, may be attached to the metallised balloon surfaces. The wind will impart a portion of its energy in the form of frictional electricity, to the balloon casing, thus substantially increasing the collection effect.

In practice however, very high towers of up to 300 metres may be employed as antennae. In these towers, copper tubes rise freely further above the top of the tower. A gas lamp secured against the wind is then lit at the point of the copper tube and a netting is secured to the copper tube over the flame of this lamp to form a collector. The gas is conveyed through the interior of the tube, up to the summit. The copper tube must be absolutely protected from moisture at the place where it enters the tower, and rain must be prevented from running down the walls of the tower, which might lead to a bad catastrophe. This is done by bell-shaped enlargements which expand downwards, being arranged in the tower in the form of high voltage insulators of Siamese pagodas.

Special attention must be devoted to the foundations of such towers. They must be well insulated from the ground, which may be achieved by first embedding a layer of concrete in a box form to a sufficient depth in the ground, and inserting in this, an asphalt lining and then glass bricks cast about 1 or 2 metres in thickness. Over this in turn, there is a ferro-concrete layer in which alone the metal foot of the tube is secured. This concrete block must be at least 2 metres from the ground and at the sides, be fully protected from moisture by a wooden covering. In the lower part of the tower, a wood or glass housing should be constructed to protect the capacitors and/or motors. In order to ensure that the ground lead connects to the water-table, a well insulated pit lined with vitreous bricks must be provided. Several such towers are erected at equal distances apart and connected with a horizontal conductor. The horizontal connecting wires may either run directly from tower to tower or be carried on bell-shaped insulators similar to those in use for high voltage electricity transmission lines. The width of the aerial tower network may be of any suitable size and the connection of the motors can take place at any convenient location.

Fig. 21. Fig. 22.

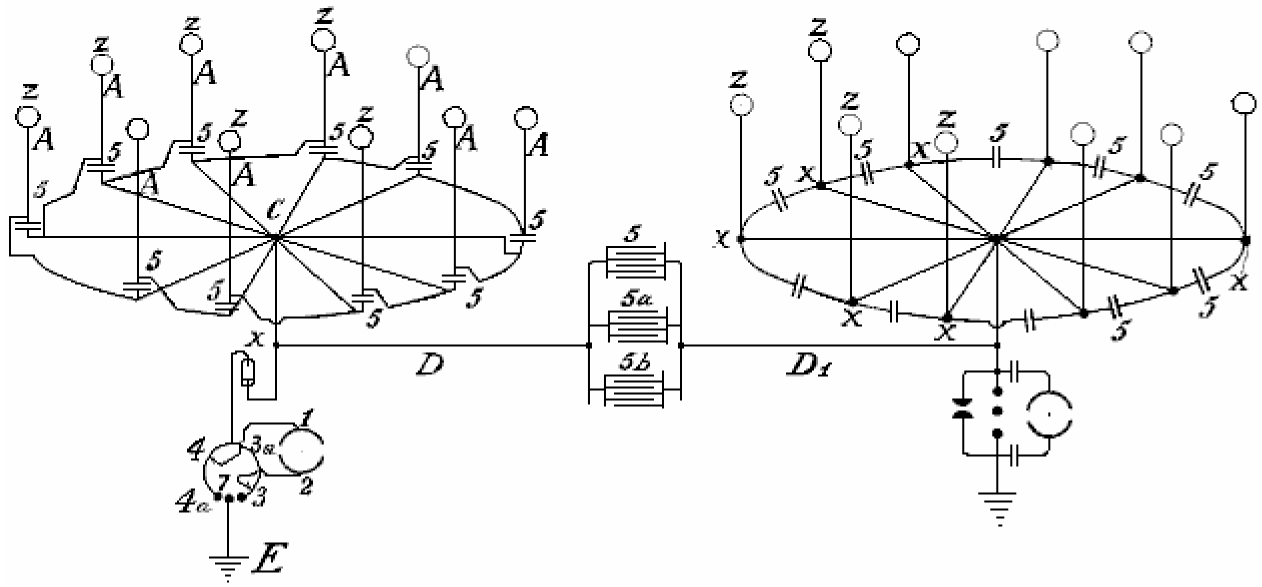

In order to collect large quantities of electricity with few aerials, it is as well to provide the aerial conductor with sets of capacitors as shown in the two methods of construction illustrated in **Fig.21** and **Fig.22**. In **Fig.21** the set of capacitors **5** is connected between the aerials **Z** via lead **A** and an annular conductor from which horizontal run to the connecting points **C** to which the earth wire is connected. **Fig.22** shows a similar arrangement.

Should two such series of antenna rings be shown by a voltmeter to have a large voltage difference (for example, one in the mountains and one on the plain) or even of a different polarity, these differences may be compensated for by connecting sufficiently large capacitor sets (**5, 5a, 5b**) by means of Maji star conductors **D** and **D^1**. **Fig.23**, shows a connection of three such rings of collectors are positioned in a triangle with a central set of capacitors.

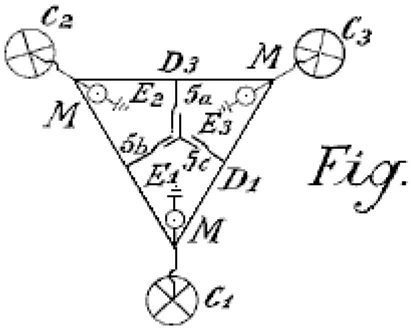

Fig. 23.

The capacitor sets of such large installations must be embedded in liquefied gasses or in liquids freezing at very low temperatures. In such cases, a portion of the atmospheric energy must be employed for liquefying these gasses. It is also preferable to employ pressure. By this means, the capacitor surfaces may be reduced in area and still allow the storage of large quantities of energy to be stored, secure against breakdown. For the smaller installations, the immersing of the capacitors in well insulated oil or the like, is sufficient. Solid substances, on the other hand, cannot be employed as insulators.

The arrangement in the diagrams shown earlier has always shown both poles of the capacitors connected to the aerial conductors. An improved method of connection has been found to be very advantageous. In this method, only one pole of each capacitor is connected to the collecting network. Such a method of connection is very important, as by means of it, a constant current and an increase in the normal working voltage is obtained. If, for example, a collecting balloon aerial which is allowed to rise to a height of 300 metres, shows 40,000 volts above earth voltage, in practice it has been found that the working voltage (with a withdrawal of the power as described earlier by means of oscillating spark gaps and the like) is only about 400 volts. If however, the capacity of the capacitor surfaces be increased, which capacity in the above mentioned case was equal to that of the collecting surface of the balloon aerials, to double the amount, by connecting the capacitors with only one pole, the voltage rises under an equal withdrawal of current up to and beyond 500 volts. This can only be ascribed to the favourable action of the connecting method.

In addition to this substantial improvement it has also been found preferable to insert double inductances with electromagnets and to place the capacitors preferably between two such electromagnets. It has also been found that the useful action of such capacitors can be further increased if an induction coil is connected as an inductive resistance to the unconnected pole of the capacitor, or still better if the capacitor itself be made as an induction capacitor. Such a capacitor may be compared to a spring, which when compressed, carries in itself accumulated force, which it gives off again when released. In charging, a charge with reversed sign is formed at the other free capacitor pole, and if a short circuit occurs through the spark gap, the accumulated energy is again given back since now new quantities of energy are induced at the capacitor pole connected to the conductor network, which in fact, charges with opposite sign to that at the free capacitor pole. The new induced charges have of course, the same sign as the collector network. The whole voltage energy in the aerial is thereby increased. In the same time interval, larger quantities of energy are accumulated than is the case without such capacitor sets being inserted.

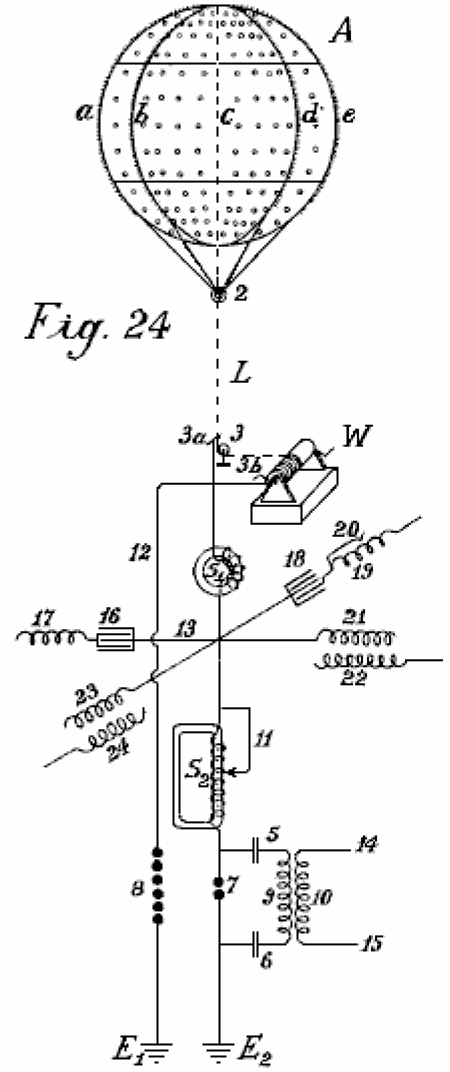

Fig. 24

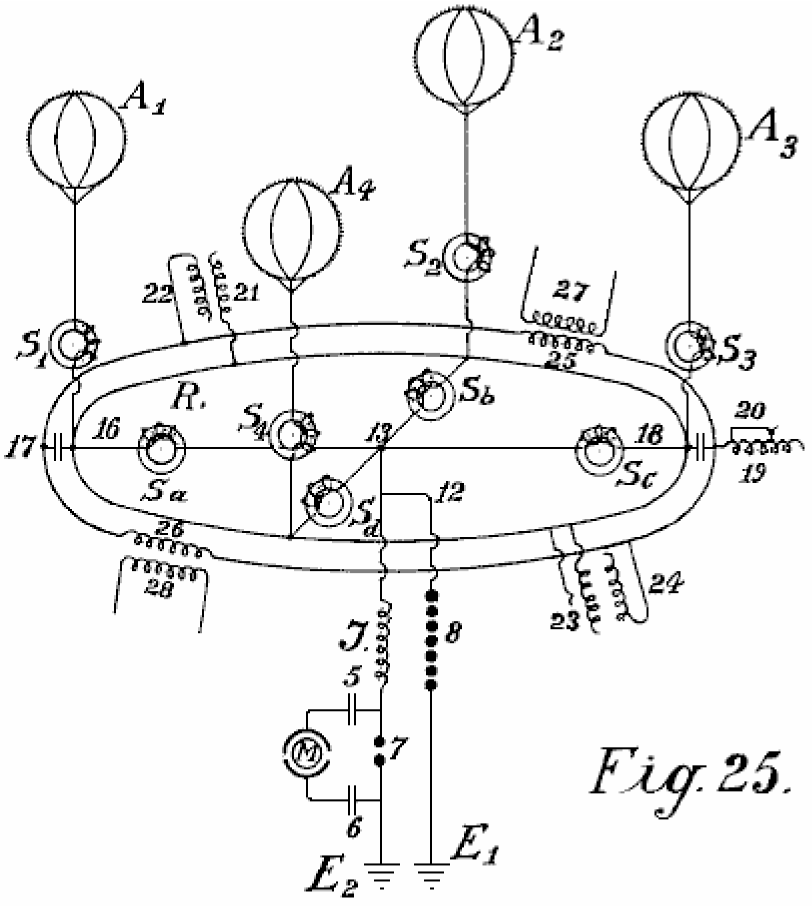

Fig. 25.

In **Fig.24** and **Fig.25**, two different connection diagrams are illustrated in more detail. **Fig.24** shows a collecting balloon along with its earth connections. **Fig.25** shows four collecting balloons and the parallel connection of their capacitor sets.

A is the collecting balloon made of an aluminium magnesium alloy (electron metal magnalium) of a specific gravity of 1.8 and a plate thickness of 0.1 mm to 0.2 mm. Inside, there are eight strong vertical ribs of T-shaped section of about 10 mm to 20 mm in height and about 3 mm in thickness, with the projecting part directed inwards (indicated by **a, b, c, d** and so forth). They are riveted together to form a firm skeleton and are stiffened in a horizontal direction by two cross ribs. The ribs are further connected to one another internally and transversely by means of thin steel wires, whereby the balloon obtains great strength and elasticity. Rolled plates of 0.1 mm to 0.2 mm in thickness made of magnalium alloy are then either soldered or riveted on to this skeleton so that a fully metallic casing with a smooth external surface is created. Well silvered or coppered aluminium plated steel wires run from each rib to the fastening ring **2**. Further, the coppered steel hawser **L**, preferably twisted out of separate thin wires (shown as dotted lines in **Fig.24**) and which must be long enough to allow the balloon to rise to the desired height, leads to a metal roller or pulley **3** and on to a winch **W**, which must be well insulated from the earth. By means of this winch, the balloon which is filled with hydrogen or helium, can be allowed to rise to a suitable height of 300 to 5,000 metres, and brought to the ground for recharging or repairs.

The actual current is taken directly through a friction contact from the metal roller **3** or from the wire or even from the winch, or simultaneously from all three by means of brushes (**3**, **3a** and **3b**). Beyond the brushes, the conductor is divided, the paths being:- firstly, over **12** to the safety spark gap **8**, on to the earth conductor **E**^{1}, and secondly over electromagnet **S**^{1}, point **13**, to a second loose electromagnet having an adjustable coil **S**^{2}, then to the spark gap **7** and to the second earth conductor **E**^{2}. The actual working circuit is formed through the spark gap **7**, capacitors **5** and **6**, and through the primary coil **9**; here the static electricity formed by oscillatory discharges is accumulated and converted into high frequency electromagnetic oscillations. Between the electromagnets **S**^{1} and **S**^{2} at the crossing point **13**, four capacitor sets are introduced which are only indicated diagramatically in the drawings by a single capacitor. Two of these sets of capacitors (**16** and **18**) are made as plate capacitors and prolonged by regulating induction coils or spirals **17** and **19** while the two others (**21** and **23**) are induction capacitors. As may be seen from the drawings, each of the four capacitor sets, **16**, **18**, **21** and **23** is connected by only one pole to either the aerial or to the collector conductor. The second poles **17**, **19**, **22** and **24** are open. In the case of plate capacitors having no inductive resistance, an induction coil is inserted. The object of such a spiral or coil is the displacement of phase of the induction current by $^1/_4$ periods, whilst the charging current of the capacitor poles which lie free in the air, works back to the collector aerial. The consequence of this is that in discharges in the collector aerial, the back-inductive action of the free poles allows a higher voltage to be maintained in the aerial collecting conductor than would otherwise be the case. It has also been found that such a back action has an extremely favourable effect on the wear of the contacts. Of course, the inductive effect may be regulated at will within the limits of the size of the induction coil, the length of the coil in action being adjustable by means of wire connection without induction (see **Fig.24** No. **20**).

S^1 and **S^2** may also be provided with such regulating devices, in the case of **S^2** illustrated by **11**. If excess voltage be formed, it is conducted to earth through wire **12** and spark gap **8**, or through any other suitable apparatus, since this voltage would be dangerous for the other components. The action of these capacitor sets has already been described.

The small circles on the collector balloon indicate places where small patches of extremely thin layers (0.01 to 0.05 mm thick) of zinc amalgam, gold amalgam or other photoelectric acting metals, are applied to the balloon casing of light metal. Such metallic patches may also be applied to the entire balloon as well as in greater thickness to the conducting network. The capacity of the collector is thereby considerably strengthened at the surface. The greatest possible effect in collecting may be obtained by polonium amalgams and the like. On the surface of the collector balloon, metal points or spikes are also fixed along the ribs. These spikes enhance the charge collection operation. Since it is well known that the sharper the spikes, the less the resistance of the spikes, it is therefore extremely important to use spikes which are as sharp as possible. Experiments have shown that the formation of the body of the spike or point also play a large part, for example, spikes made of bars or rollers with smooth surfaces, have point resistance many times greater than those with rough surfaces. Various kinds of spike bodies have been experimented with for the collector balloons and the best results were given with spikes which were made in the following way: Fine points made of steel, copper, nickel or copper and nickel alloys, were fastened together in bundles and then placed as anode with the points placed in a suitable electrolyte (preferably in hydrochloric acid or muriate of iron solutions) and so treated with weak current driven by 2 to 3 volts. After 2 to 3 hours, according to the thickness of the spikes, the points become extremely sharp and the bodies of the spikes have a rough surface. The bundle can then be removed and the acid washed off with water. The spikes are then placed as cathode in a bath containing a solution of gold, platinum, iridium, palladium or wolfram salts or their compounds, and coated at the cathode galvanically with a thin layer of precious metal, which mush however be sufficiently firm to protect them from atmospheric oxidation.

Such spikes act at a 20 fold lower voltage almost as well as the best and finest points made by mechanical means. Still better results are obtained if polonium or radium salts are added to the galvanic bath when forming the protective layer or coating. Such pins have low resistance at their points and have excellent collector action even at one volt or lower.

In **Fig.24**, the three unconnected poles are not connected with one another in parallel. That is quite possible in practice without altering the principle of the free pole. It is also preferable to interconnect a series of collecting aerials in parallel to a common collector network. **Fig.25** shows such an arrangement. **A^1, A^2, A^3, A^4** are four metal collector balloons with gold or platinum coated spikes which are electrolytically mad in the presence of polonium emanations or radium salts, the spikes being connected over four electromagnets **S^1, S^2, S^3, S^4**, through an annular conductor **R**. From this annular conductor, four wires run over four further electromagnets **S^a, S^b, S^c, S^d**, to the connecting point **13**. There, the conductor is divided, one branch passing over **12** and the safety spark gap **7** to the earth at **E^1**, the other over inductive resistance **J** and working spark gap **7** to the earth at **E^2**. The working circuit, consisting of the capacitors **5** and **6** and a resonance motor or a capacitor motor **M**, such as already described, is connected in proximity around the sparking gap section **7**. Of course, instead of connecting the capacitor motor directly, the primary circuit for high frequency oscillatory current may also be inserted.

The capacitor sets are connected by one pole to the annular conductor **R** and can be either inductionless (**16** and **18**) or made as induction capacitors as shown by **21** and **23**. The free poles of the inductionless capacitors are indicated by **17** and **19**, and those of the induction capacitors by **22** and **24**. As may be seen from the drawings, all of these poles **17, 22, 19** and **24** may be interconnected in parallel through a second annular conductor without any fear that thereby the principle of the free pole connection will be lost. In addition to the advantages already mentioned, the parallel connection also allows an equalisation of the working voltage in the entire collector network. Suitably calculated and constructed induction coils **25** and **26** may also be inserted in the annular conductor of the free poles, by means of which, a circuit may be formed in the secondary coils **27** and **28** which allows current produced in this annular conductor by fluctuations of the charges, to be measured or otherwise utilised.

According to what has already been stated, separate collector balloons may be connected at equidistant stations distributed over the whole country, either connected directly with one another metallically or by means of intermediate suitably connected capacitor sets through high voltage conductors insulated from earth. The static electricity is converted through a spark gap, into high frequency dynamic electricity which may be utilised as a source of energy by means of a suitable connection method, various precautions being observed, and with special regulations. The wires leading from the collector balloons, have up to now been connected through an annular conductor without this endless connection, which can be regarded as an endless induction coil, being able to exert any action on the whole conductor system.

It has now been found that if the network conductor connecting the aerial collector balloons with one another, is not made as a simple annular conductor, but preferably short-circuited in the form of coils over a capacitor set or spark gap or through thermionic valves, then the total collecting network exhibits quite new properties. The collection of atmospheric electricity is thereby not only increased but an alternating field may easily be produced in the collector network. Further, the atmospheric electrical forces showing themselves in the higher regions, may also be obtained directly by induction. In **Fig.26** and **Fig.28**, a form of construction is shown, on the basis of which, the further foundations of the method will be explained in more detail.

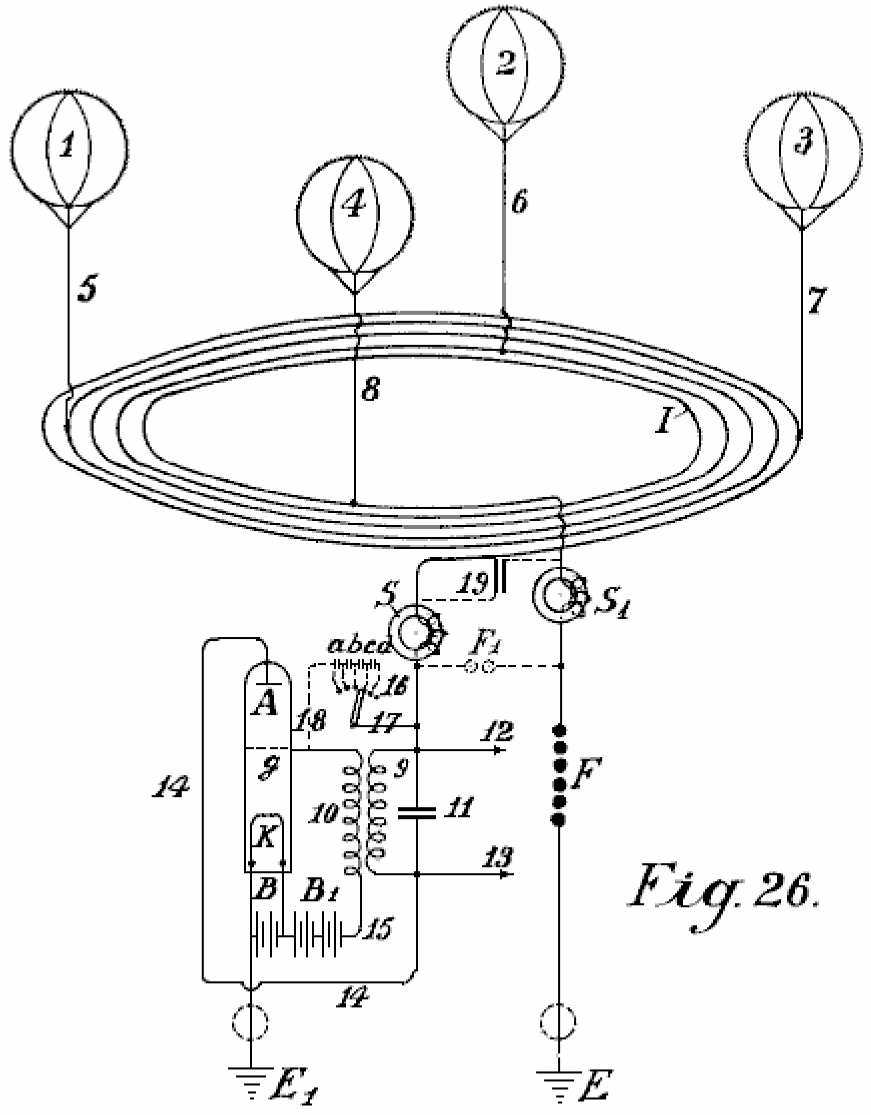

Fig. 26.

In **Fig.26, 1,2,3** and **4** are metallic collector balloons, with **5, 6, 7** and **8** their metallic aerial conductors and **I** the actual collector network. This consists of five coils and is mounted on high voltage insulators in the air, on high voltage masts (or with a suitable construction of cable, embedded in the earth). One coil has a diameter of 1 to 100 km. or more. **S** and **S**^{1} are two protective electromagnets, **F** is the second safety section against excess voltage, **E** its earth conductor and **E**^{1} the earth conductor of the working section. When an absorption of static atmospheric electricity is effected through the four balloon collectors, in order to reach the earth connection **E**^{1}, the current must flow spirally through the collector network, over the electromagnet **S**, primary induction coil **9**, conductor **14**, anode **A** of the audion tube, incandescent cathode **K**, as the way over the electromagnet and safety spark gap **F** offers considerably greater resistance. Owing to the fact that the accumulated current flows in one direction, an electromagnetic alternating field is produced in the interior of the collector network coil, whereby all of the free electrons are directed more or less into the interior of the coil. An increased ionisation of the atmosphere is therefore produced. Consequently, the points mounted on the collector balloon, show a considerably reduced resistance and therefore increased static charges are produced between the points on the balloon and the surrounding atmosphere. This results in a considerably increased collector effect.

A second effect, which could not be achieved in any other way, is obtained by the alternating electromagnetic field running parallel to the earth's surface, which acts more or less with a diminishing or increasing effect on the earth's magnetic field, whereby in the case of fluctuations in the current, a return induction current of reversed sign is always produced in the collector coil by earth magnetism. Now if a constantly pulsating, continuous alternating field is produced as stated in the collector network **I**, an alternating current of the same frequency is also produced in the collecting network coil. As the same alternating field is further transmitted to the aerial balloon, the resistance of its points is thereby considerably reduced, while the collector action is considerably increased. A further advantage is that positive charges which collect on the metal surfaces during the conversion into dynamic current, produce a so-called voltage drop in the collector area. As an alternating field is present, when discharge of the collector surfaces takes place, the negative ions surrounding the collector surfaces produce, by the law of induction, an induction of reversed sign on the collector surface - that is, a positive charge. In addition to the advantages already stated, the construction of connecting conductors in coil form, when of sufficiently large diameter, allows a utilisation of energy arising in higher regions, also in the most simple way. As is well known, electric discharges frequently take place at very great elevations which may be observed, such as 'St. Elmo's fires' or 'northern lights'. These energy quantities have not been able to have been utilised before now. By this invention, all of these kinds of energy, as they are of electromagnetic nature and since the axis of the collector coils is at right angles to the earth's surface, can be absorbed in the same way as a radio absorbs distant radio signals. With a large diameter of the spiral, it is possible to connect large surfaces and thereby take up large quantities of energy.

It is well known that in the summer months and in the tropics, large radio stations are very frequently unable to receive signals due to interruptions caused by atmospheric electricity, and this takes place with vertical coils of only 40 to 100 metres in diameter. If, on the contrary, horizontal coils of 1 to 100 kilometres in diameter are used, very strong currents may be obtained through discharges which are constantly taking place in the atmosphere. Particularly in the tropics, or still better in the polar regions where the northern lights are constantly present, large quantities of energy may probably be obtained in this way. A coil with several windings should perform the best. In a similar manner, any alteration of the earth's magnetic field should act inductively on such a coil.

It is not at all unlikely that earthquakes and sunspots will also produce an induction in collector coils of that size. In similar manner, this collector conductor will react to earth currents more particularly when they are near the surface of the earth or even embedded in the earth. By combining the previous kind of current collectors, so far as they are adapted for the improved system with the improved possibilities of obtaining current, the quantities of free natural energy which are to be obtained in the form of electricity are considerably increased.

In order to produce uniform undamped current oscillations in the improved collector coil, so-called audion high vacuum or thermionic valves are used instead of the previous described spark gaps (**Fig.26, 9-18**). The main aerial current flows through electromagnet **S** (which in the case of a high number of alternations is not connected here but in the earth conductor **E**^{1}) and may be conveyed over the primary coils in the induction winding through wire **14** to the anode **A** of the high vacuum grid valve. Parallel with the induction resistance **9**, a regulating capacity of suitable size, such as capacitor **11**, is inserted. In the lower part of the vacuum grid valve is the incandescent filament cathode **K** which is fed through a battery **B**. From the battery, two branches run, one to the earth conductor **E**^{1} and the other through battery **B**^{1} and secondary coil **10** to the grid anode **g** of the vacuum tube. By the method of connections shown in dotted lines, a desired voltage may also be produced at the grid electrode **g** through wire **17** which is branched off from the main current conductor through switches **16** and some small capacitors (**a, b, c, d**) connected in series, and conductor **18**, without the battery **B**^{1} being required. The action of the whole system is somewhat as follows:-

On the connecting conductor of the aerial collector network being short-circuited to earth, the capacitor pole **11** is charged, and slightly dampened oscillations are formed in the short-circuited oscillation circuit formed by capacitor **11** and self inductance **9**. Because of the coupling through coil **10**, voltage fluctuations of the same frequency take place in the grid circuit **15** and in turn, these fluctuations influence the strength of the electrode current passing through the high vacuum amplifying valve and thus produce current fluctuations of the same frequency in the anode circuit. A permanent supply of energy. Consequently, a permanent supply of energy is supplied to the oscillation circuits **9** and **10** takes place, until a balance is achieved where the oscillation energy consumed exactly matches the energy absorbed. This produces constant undamped oscillations in the oscillation circuits **9 - 11**.

For regular working of such oscillation producers, high vacuum strengthening tubes are necessary and it is also necessary that the grid and anode voltages shall have a phase difference of 180^0 so that if the grid is negatively charged, then the anode is positively charged and vice versa. This necessary difference of phase may be obtained by most varied connections, for example, by placing the oscillating circuit in the grid circuit or by separating the oscillation circuit and inductive coupling from the anodes and the grid circuit, and so forth.

A second important factor is that care must be taken that the grid and anode voltages have a certain relation to one another; the latter may be obtained by altering the coupling and a suitable selection of the self induction in the grid circuit, or as shown by the dotted lines **18, 17, 16** by means of a larger or smaller number of capacitors of suitable size connected in series; in this case, the battery **B**^{1} may be omitted. With a suitable selection of the grid potential, a glow discharge takes place between the grid **g** and the anode **A**, and accordingly at the grid there is a cathode drop and a dark space is formed. The size of this cathode drop is influenced by the ions which are emitted in the lower space in consequence of shock ionisation of the incandescent cathodes **K** and pass through the grid in the upper space. On the other hand, the number of the ions passing through the grid is dependent on the voltage between the grid and the cathode. Thus, if the grid voltage undergoes periodic fluctuations (as in the present case), the amount of the cathode drop at the grid fluctuates, and consequently, the internal resistance of the valve fluctuates correspondingly, so that when a back-coupling of the feed circuit with the grid circuit takes place, the necessary means are in place for producing undamped oscillations and of taking current as required, from the collecting conductor.

With a suitably loose coupling, the frequency of the undamped oscillations produced is equal to the self-frequency of the oscillation circuits **9** and **10**. By selecting a suitable self-induction for coil **9** and capacitor **11**, it is possible to extend operation from frequencies which produce electromagnetic oscillations with a wavelength of only a few metres, down to the lowest practical alternating current frequency. For large installations, a suitable number of frequency producing tubes in the form of the well known high vacuum transmission tubes of 0.5 kW to 2 kW in size may be connected in parallel so that in this respect, no difficulty exists.

The use of such tubes for producing undamped oscillations, and the construction and method of inserting such transmission tubes in an accumulator or dynamo circuit is known, also, such oscillation producing tubes only work well at voltages of 1,000 volts up to 4,000 volts, so that on the contrary, their use at lower voltages is considerably more difficult. By the use of high voltage static electricity, this method of producing undamped oscillations as compared with that through spark gaps, must be regarded as an ideal solution, particularly for small installations with outputs from 1 kW to 100 kW.

By the application of safety spark gaps, with interpolation of electromagnets, not only is short-circuiting avoided but also the taking up of current is regulated. Oscillation producers inserted in the above way, form a constantly acting alternating electromagnetic field in the collector coil, whereby, as already stated, a considerable accumulating effect takes place. The withdrawal or 'working' wire is connected at **12** and **13**, but current may be taken by means of a secondary coil which is firmly or moveably mounted in any suitable way inside the large collector coil, i.e. in its alternating electromagnetic field, so long as the direction of its axis is parallel to that of the main current collecting coil.

In producing undamped oscillations of a high frequency (50 KHz and more) in the oscillation circuits **9** and **11**, electromagnets **S** and **S^1** must be inserted if the high frequency oscillations are not to penetrate the collector coil, between the oscillation producers and the collector coil. In all other cases they are connected shortly before the earthing (as in **Fig.27** and **Fig.28**).

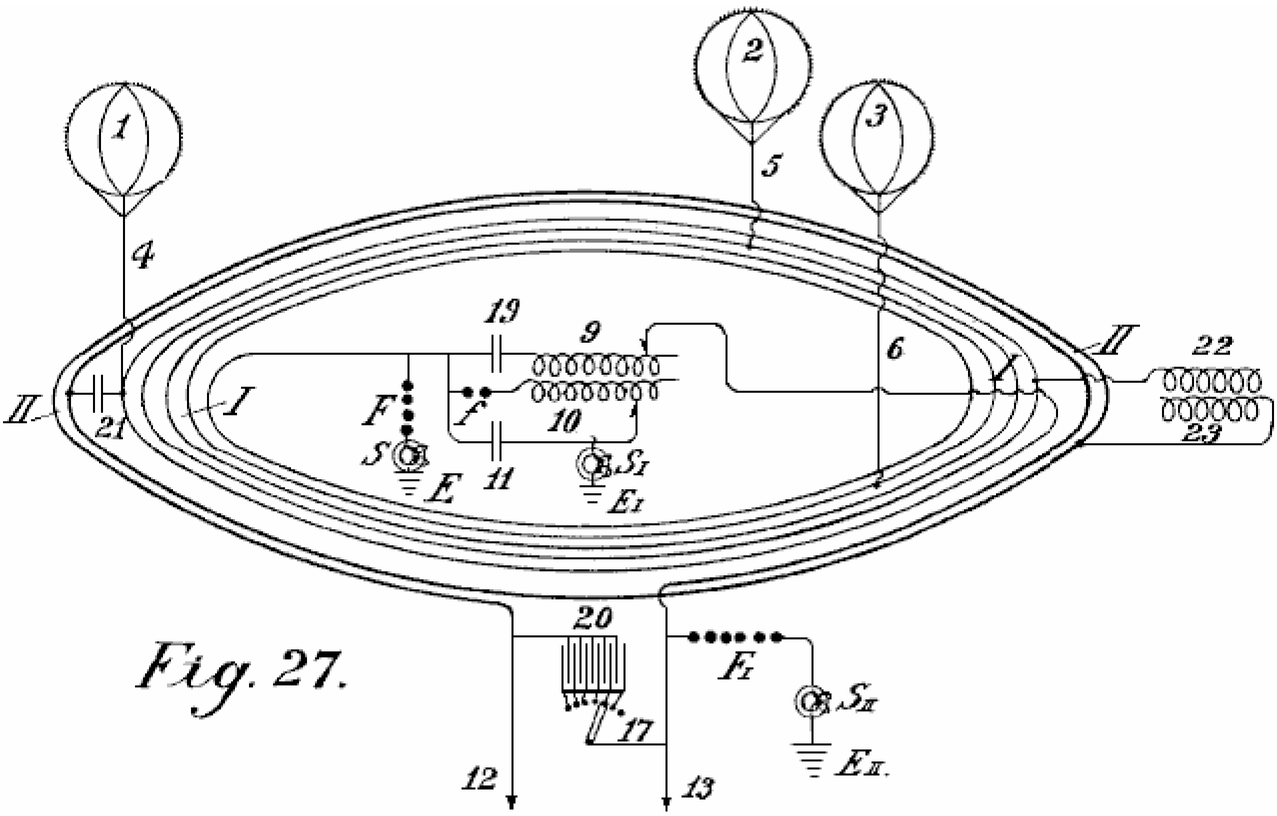

Fig. 27.

In **Fig.27** a second method of construction of the connecting conductor of the balloon aerials is illustrated in the form of a coil. The main difference is that in addition to the connecting conductor **I** another annular conductor **II** is inserted parallel to the former on the high voltage masts in the air (or embedded as a cable in the earth) but both in the form of a coil. The connecting wire of the balloon aerials is both a primary conductor and a current producing network while the coil is the consumption network and is not in unipolar connection with the current producing network.

In **Fig.27** the current producing network **I** is shown with three balloon collectors **1, 2, 3** and aerial conductors **4, 5, 6**; it is short-circuited through capacitor **19** and inductor **9**. The oscillation forming circuit consists of spark gap **f**, inductor **10** and capacitor **11**. The earth wire **E** is connected to earth through electromagnet **S^1**. **F_I** is the safety spark gap which is also connected to earth through a second electromagnet **S_{II}** at **E_{II}**. On connecting up the capacitor circuit **11** it is charged over the spark gap **f** and an oscillatory discharge is formed. This discharging current acts through inductor **10** on the inductively coupled secondary **9**, which causes a change in the producing network, by modifying the voltage on capacitor **19**. This causes oscillations in the coil-shaped producer network. These oscillations induce a current in the secondary circuit **II**, which has a smaller number of windings and lower resistance, consequently, this produces a lower voltage and higher current in it.

In order to convert the current thus obtained, into current of an undamped character, and to tune its wavelengths, a sufficiently large regulatable capacitor **20** is inserted between the ends **12** and **13** of the secondary conductor **II**. Here also, current may be taken without an earth conductor, but it is advisable to insert a safety spark gap **E**^{1} and to connect this with the earth via electromagnet **S**^{2}. The producer network may be connected with the working network **II** over an inductionless capacitor **21** or over an induction capacitor **22, 23**. In this case, the secondary conductor is unipolarly connected with the energy conductor.

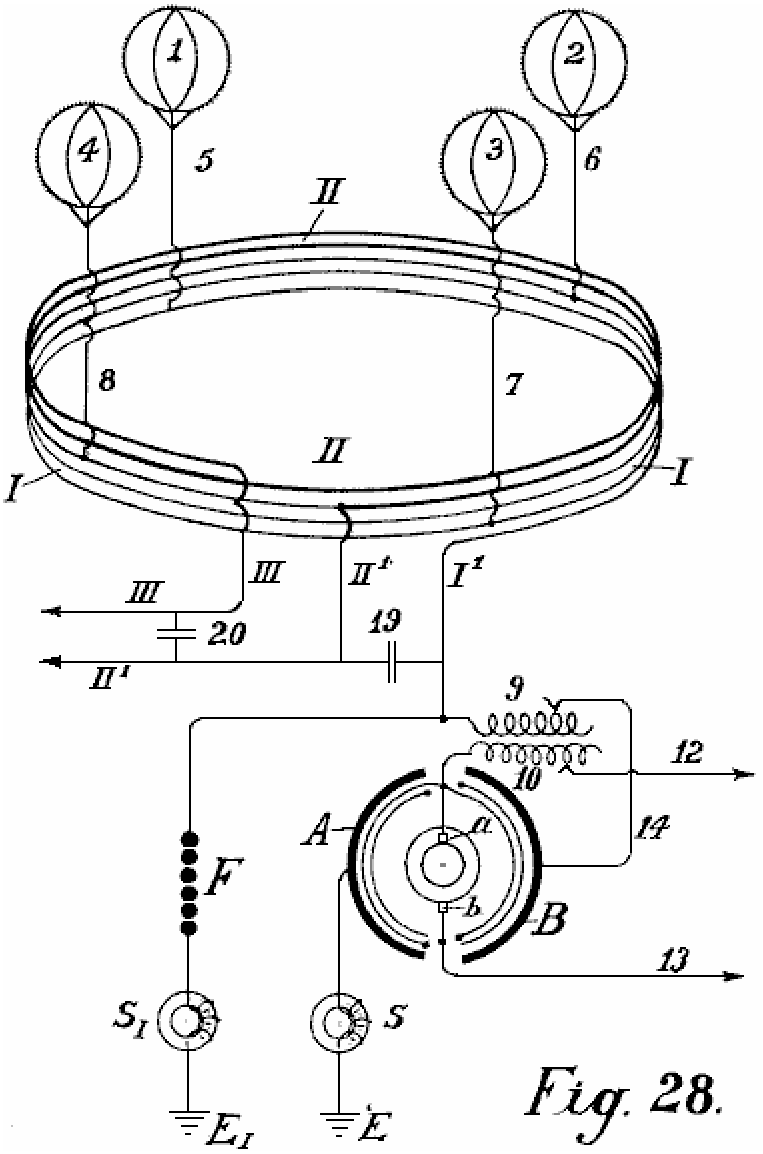

Fig. 28.

In **Fig.28**, the connecting conductor between the separate collecting balloons is carried out according to the autotransformer principle. The collecting coil connects four aerial balloons **1, 2, 3, 4**, the windings of which are not made side-by-side but one above the other. In **Fig.28**, the collector coil **I** is shown with a thin line and the metallically connected prolongation coils **II** with a thick line. Between the ends **I**^{1} and **II**^{1} of the energy network **I**, a regulating capacitor **19** is inserted. The wire **I**^{1} is connected with the output wire and with the spark gap **F**.

As transformer of the atmospheric electricity, an arrangement is employed which consists of using rotary pairs of capacitors in which the stator surface **B** is connected with the main current, while the other **A** is connected to the earth pole. These pairs of short-circuited capacitors are caused to rotate and the converted current can be taken from them via two collector rings and brushes. This current is alternating current with a frequency dependent on the number of balloons and the rate of revolutions of the rotor. As the alternating current formed in the rotor can act through coils **10** on the inductor **9**, an increase or decrease of the feed current in **I** can be obtained according to the direction of the current by back-induction. Current oscillations of uniform rhythm are produced in the coil-shaped windings of the producer network.

As the ends of this conductor are short-circuited through the regulatable capacitor **19**, these rhythms produce short-circuited undamped oscillations in the energy conductor. The frequency of these oscillations can be altered at will by adjusting the capacitance of capacitor **19**. These currents may also be used as working current via the conductors **II**^{1} and **III**. By inserting capacitor **20**, a connection between these conductors may also be made, whereby harmonic oscillations of desired wavelength are formed. By this means, quite new effects as regards current distribution are obtained. The withdrawal of current can even take place without direct wire connection if, at a suitable point in the interior of the producing network (quite immaterially whether this has a diameter of 1 or 100 km) a coil tuned to these wavelength and of the desired capacity, is firmly or moveably mounted in the aerial conductor in such a way that its axis is parallel with the axis of the collector coil. In this case, a current is induced in the producing network, the size of which is dependent on the total capacity and resistance and on the frequency selected. A future possibility is taking energy from the producer network by radio signals as in addition to atmospheric electricity, magnetic earth currents and energy from the upper atmosphere may be tapped.

Of course, vacuum tubes may be used to produce undamped oscillations anywhere spark gaps are shown in the circuits. The separate large-diameter coils of the producer network may be connected to one another through separate conductors all in parallel or all in series or in groups in series. By regulating the number of oscillations and the magnitude of the voltage, more or fewer large collector coils of this kind may be used. The coils may also be divided spirally over the entire section. The coils may be carried out in annular form or in triangular, quadrangular, hexagonal or octagonal form.

Of course, wires which form guides for the current waves, may be carried from a suitable place to the centre or also laterally. This is necessary when the currents have to be conducted over mountains and valleys and so forth. In all these cases, the current must be converted into a current of suitable frequency.

As already mentioned, separate collecting balloons may be directly metallically interconnected a equidistant stations distributed over the entire country, or may be connected by interpolation of suitable capacitor sets by means of high voltage conductors. The static electricity is converted through a spark gap into dynamic energy of high frequency and could then in that form be used as an energy source after special regulation.

According to this invention, in order to increase the collecting effect of the balloon in the aerial collector conductor or in the earth wire, radiating collectors are used. These consist of either incandescent metal or oxide electrodes in the form of vacuum grid valves, or electric arcs (mercury or similar electrodes), Nernst lamps, or flames of various kinds maybe simply connected with the respective conductor.

It is well known that energy can be drawn off from a cathode consisting of an incandescent body opposite an anode charged with positive electricity (vacuum grid tube). Hitherto however, a cathode was always first directly placed opposite an anode, and secondly, the system always consisted of a closed circuit.

Now if we dispense with the ordinary ideas in forming light or flame arcs in which a cathode must always stand directly opposite an anode charged to a high voltage or another body freely floating in the air, or consider the incandescent cathode to be only a source of unipolar discharge, (which represents group and point discharges in electro-static machines similar to unipolar discharges), it may be ascertained that incandescent cathodes and less perfectly, all incandescent radiators, flames and the like, have relatively large current densities and allow large quantities of electric energy to radiate into open space in the form of electron streams as transmitters.

The object of this invention is as described below, if such incandescent oxide electrodes or other incandescent radiators or flames are not freely suspended in space but instead are connected metallically with the earth so that they can be charged with negative terrestrial electricity, these radiators possess the property of absorbing the free positive electrical charges contained in the air space surrounding them (that is to say, of collecting them and conducting them to earth). They can therefore serve as collectors and have in comparison to the action of the spikes, a very large radius of action R; the effective capacity of these collectors is much greater than the geometrical capacity (R_0) calculated in an electro-static sense.

As is well known, our earth is surrounded with an electro-static field and the difference of potential dV/dh of the earth field according to the latest investigations, is in summer about 60 to 100 volts, and in winter, 300 to 500 volts per metre difference in height, a simple calculation gives the result that when such a radiation collector or flame collector is arranged, for example, on the ground, and a second one is mounted vertically over it at a distance of 2,000 metres and both are connected by a conducting cable, there is a voltage difference in summer of about 2,000,000 volts and in winter 6,000,000 volts or more.

According to Stefan Boltzmann's law of radiation, the quantity of energy which an incandescent surface (temperature T) of 1 sq. cm. radiates in a unit of time into the open air (temperature T_0) is expressed by the following formula:

$$S = R\,(T^4 - T_0^4) \text{ watts per square centimetre}$$

and the universal radiation constant R, according to the latest researches of Ferry, is equal to 6.30 x 10^{-12} watts per square centimetre.

Now, if an incandescent surface of 1 sq. cm., as compared to the surrounding space, shows a periodic fall of potential dV, it radiates (independent of the direction of the current) in accordance with the above formula, for example at a temperature of 3715^0 C. an energy of 1.6 kW per square centimetre. As for the radiation, the same value can be calculated for the collection of energy, but reversed. Now, as carbon electrodes at the temperature of the electric arc, support a current density up to 60 to 65 amps per sq. cm., no difficulties will result in this direction in employing radiating collectors as accumulators.

If the earth be regarded as a cosmically insulated capacitor in the sense of geometrical electro-statics *x*, according to Chwolson, there results from the geometric capacity of the earth:

For negative charging 1.3 x 10^6 Coulomb For negative potential V = 10 x 10^8 volts.

It follows from this that EJT is approximately equal to 24.7 x 10^{24} watts/sec. Now if it is desired to make a theoretical short circuit through an earthed flame collector, this would represent an electrical total work of about 79,500 x 10^{10} kilowatt years. As the earth must be regarded as a rotating mechanism which is thermo-dynamically, electromagnetically and kinematically coupled with the sun and star system by cosmic radiation and gravitation, a reduction in the electric energy of the earth field is not to be feared. The energies which the incandescent collectors could withdraw from the earth field can only cause a lowering of the earth temperature. This is however, not the case as the earth does not represent a cosmically entirely insulated system. On the contrary, there is conveyed from the sun to the earth an energy of 18,500 x 10^{10} kilowatts. Accordingly, any lowering of the earth temperature without a simultaneous lowering of the sun's temperature would contradict Stefan Boltzmann's law of radiation.

From this it must be concluded that if the earth temperature sinks, the total radiation absorbed by the earth increases, and further, the rate of cooling of the earth is directly dependent on that of the sun and the other radiators cosmically coupled with the sun.

The incandescent radiation collectors may, according to this invention, be used for collecting atmospheric electricity if they (1) are charged with the negative earth electricity (that is to say, when they are directly connected to the earth by means of a metallic conductor) and (2) if large capacities (metal surfaces) charged with electricity are mounted opposite them as positive poles in the air. This is regarded as the main feature of the present invention as without these inventive ideas it would not be possible to collect with an incandescent collector, sufficiently large quantities of the electrical charges contained in the atmosphere as technology requires; the radius of action of the flame collectors would also be too small, especially if it be considered that the very small surface density does not allow of large quantities of charge being absorbed from the atmosphere.

It has already been proposed to employ flame collectors for collecting atmospheric electricity and it is known that their collecting effect is substantially greater opposite the points. It is however, not known that the quantities of current which hitherto be obtained are too small for technical purposes. According to my experiments, the reason for this is to be found in the inadequate capacities of the collector conductor poles. If such flame or radiating collectors have no or only small positive surfaces, their radius of action for large technical purposes is too small. If the incandescent collectors be constantly kept in movement in the air, they may collect more according to the speed of the movement, but this is again not capable of being carried out in practice.

By this invention, the collector effect is considerably increased by a body charged with a positive potential and of the best possible capacity, being also held floating (without direct earth connection) opposite such an incandescent collector which is held floating in the air at a desired height. If, for example, a collecting balloon of sheet metal or metallised fabric, be caused to mount to 300 to 3,000 metres in the air, and as a positive pole it is brought opposite such a radiating collector connected by a conductor to earth, quite different results are obtained.

The metallic balloon shell which has a large surface area is charged to a high potential by atmospheric electricity. This potential is greater the higher the collecting balloon is above the incandescent collector. The positive electricity acts concentratedly on the anode floating in the air as it is attracted through the radiation shock ionisation, proceeding from the incandescent cathode. The consequence of this is that the radius of action of the incandescent cathode collector is considerably increased and so is the collecting effect of the balloon surface. Further, the large capacity of the anode floating in the air, plays therefore an important part because it allows the collection of large charges resulting in a more uniform current even when there is substantial current withdrawal - this cannot be the case with small surfaces.

In the present case, the metallic collecting balloon is a positive anode floating in the air and the end of the earth conductor of this balloon serves as positive pole surface opposite the surface of the radiating incandescent cathode, which in turn is charged with negative earth electricity as it is connected to the earth by a conductor. The process may be carried out by two such contacts (negative incandescent cathode and anode end of a capacity floating in the air) a capacitor and an inductive resistance being switched on in parallel, whereby simultaneously undamped oscillations may be formed.

In very large installations it is advisable to connect two such radiating collectors in series. Thus an arc light incandescent cathode may be placed below on the open ground and an incandescent cathode which is heated by special electro-magnetic currents, be located high in the air. Of course for this, the special vacuum Liebig tubes with or without grids may also be used. An ordinary arc lamp with oxide electrodes may be introduced on the ground and the positive pole is not directly connected with the collecting balloon, but through the upper incandescent cathode or over a capacitor. The method of connecting the incandescent cathode floating in the air may be seen in **Figs.29-33**.

B is the air balloon, **K** a Cardan ring (connection with the hawser) **C** the balloon, **L** a good conducting cable, **P** a positive pole, **N** negative incandescent cathode and **E** the earth conductor.

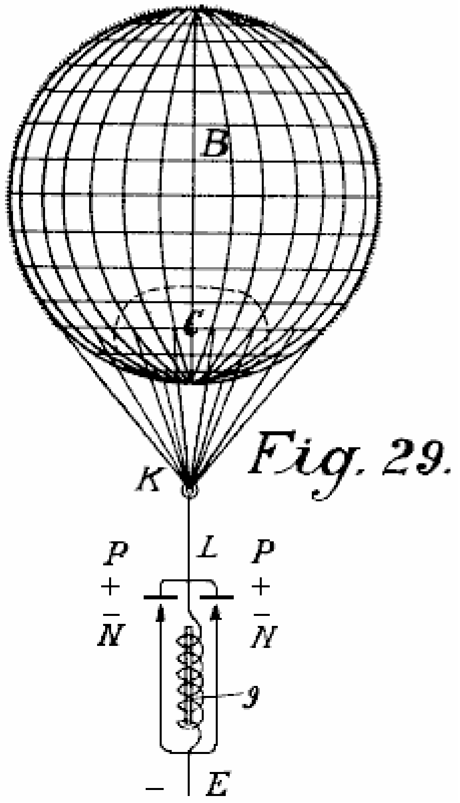

Fig.29 represents the simplest form of construction. If electric oscillations are produced below on the ground by means of a carbon arc lamp or in any other suitable way, a considerably greater electric resistance is opposed to that in the direct way by inserting an electrical inductive resistance **9**. Consequently, between **P** and **N**, a voltage is formed, and as, over **N** and **P** only an inductionless ohmic resistance is present, a spark will spring over so long as the separate induction coefficients and the like are correctly calculated. The consequence of this is that the oxide electrode (carbon or the like) is rendered incandescent and then shows as incandescent cathode, an increased collecting effect. The positive poles must be substantially larger than the negative in order that they may not also become incandescent. As they are further connected with the large balloon area which has a large capacity and is charged at high voltage, an incandescent body which is held floating in the air and a positive pole which can collect large capacities is thereby obtained in the simplest way. The incandescent cathode is first caused to become incandescent by means of separate energy produced on the earth, and then maintained by the energy collected from the atmosphere.

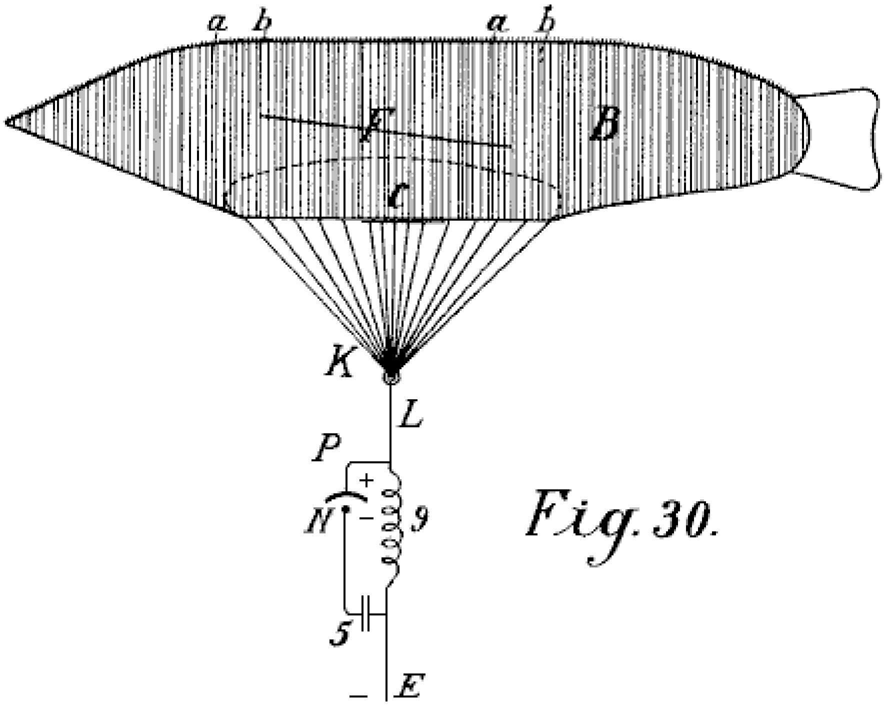

Fig.30 only shows the difference that instead of a round balloon, a cigar-shaped one may be used, also, a capacitor **5** is inserted between the incandescent cathode and the earth conductor so that a short-circuited oscillation circuit over **P N 5**

and **9** is obtained. This has the advantage that quite small quantities of electricity cause the cathode to become incandescent and much larger cathode bodies may be made incandescent.

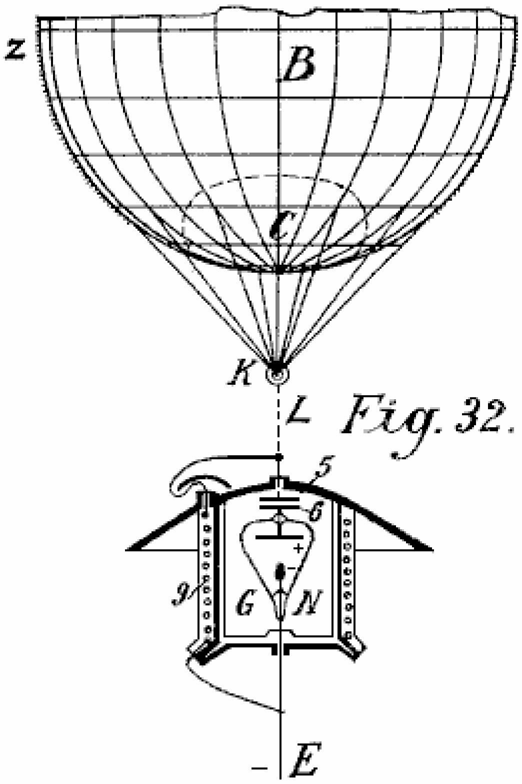
Fig. 32.

In this form of construction, both the incandescent cathode and the positive electrode may be enclosed in a vacuum chamber as shown in **Fig.32**. A cable **L** is carried well insulated through the cover of a vessel and ends in a capacitor disc **5**. The cover is arched in order to keep the rain off. The vessel is entirely or partially made of magnetic metal and well insulated inside and outside. Opposite disc **5** another disc **6** and on this again a metallic positive pole of the vacuum tube **g** with the incandescent cathode (oxide electrode) **N** is arranged. The negative electrode is on the one hand connected to the earth conductor **E**, and on the other hand with the inductive resistance **9** which is also connected with the cable **L** with the positive pole and wound around the vessel in coils. The action is exactly the same as that in **Fig.29** only instead of an open incandescent cathode, one enclosed in vacuo is used. As in such collectors, only small bodies be brought to incandescence, in large installations a plurality of such vacuum tubes must be inserted in proximity to one another. According to the previous constructions **Fig.31** and **Fig.33** are quite self evident without further explanations.

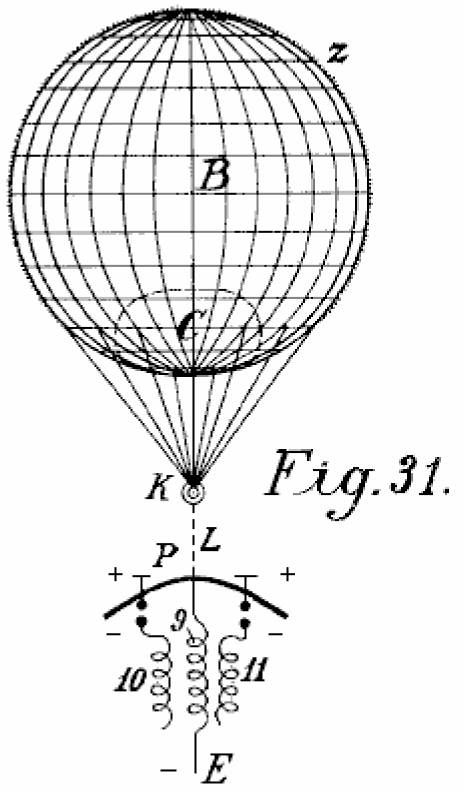
Fig. 31.

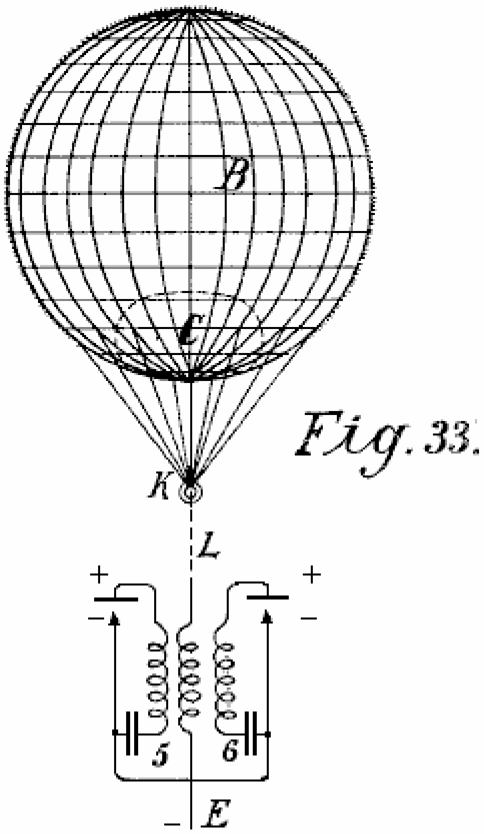
Fig. 33.

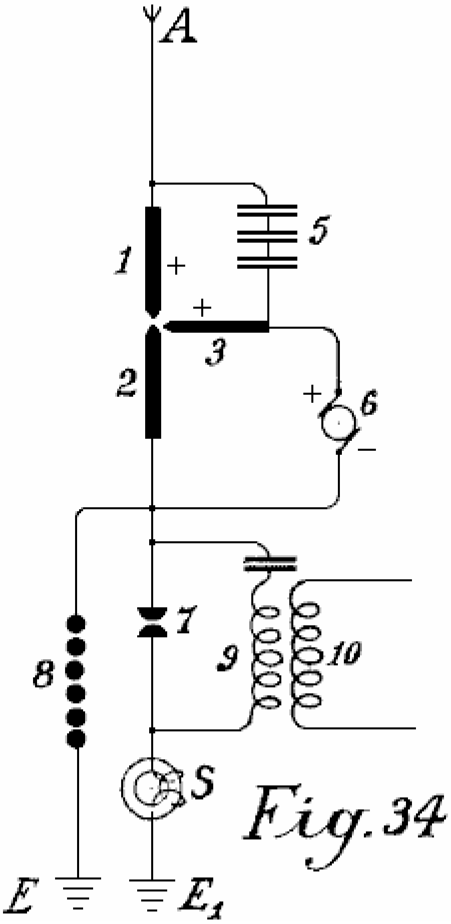
A
1
+
5
+
3
2
+
6
−
7
8
9
10
S
E
E_1

Fig.34

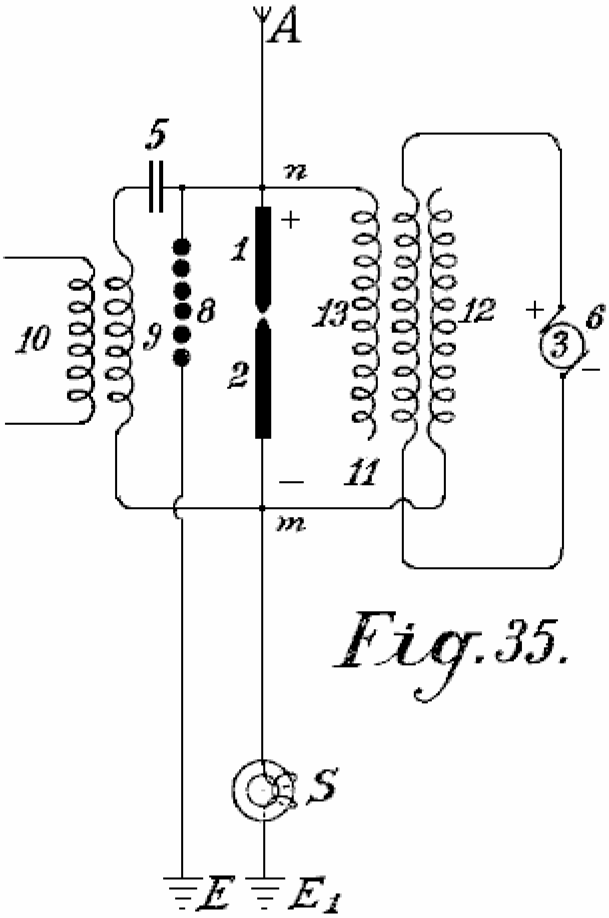
A
5
n
+
1
8
9
10
13
12
+
6
3
−
2
−
11
m
S
E
E_1

Fig.35.

Figs.34-37 represent further diagrams of connections over radiating and flame collectors, and in fact, how they are to be arranged on the ground. **Fig.34** shows an arc light collector with oxide electrodes for direct current and its connection. **Fig.35** shows a similar one for alternating current. **Fig.36** an incandescent collector with a Nernst lamp and **Fig.37** a similar one with a gas flame.

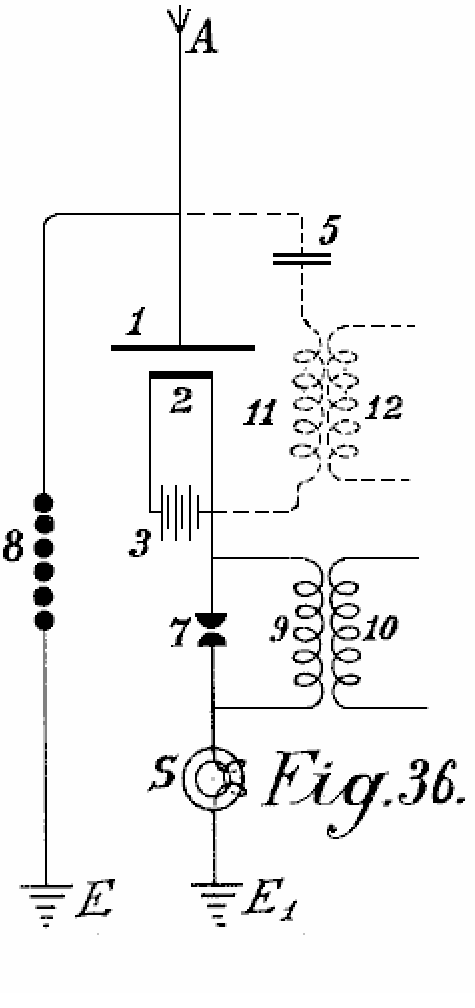
A
5
1
2
11
12
3
8
7
9
10
S
E
E_1

Fig.36.

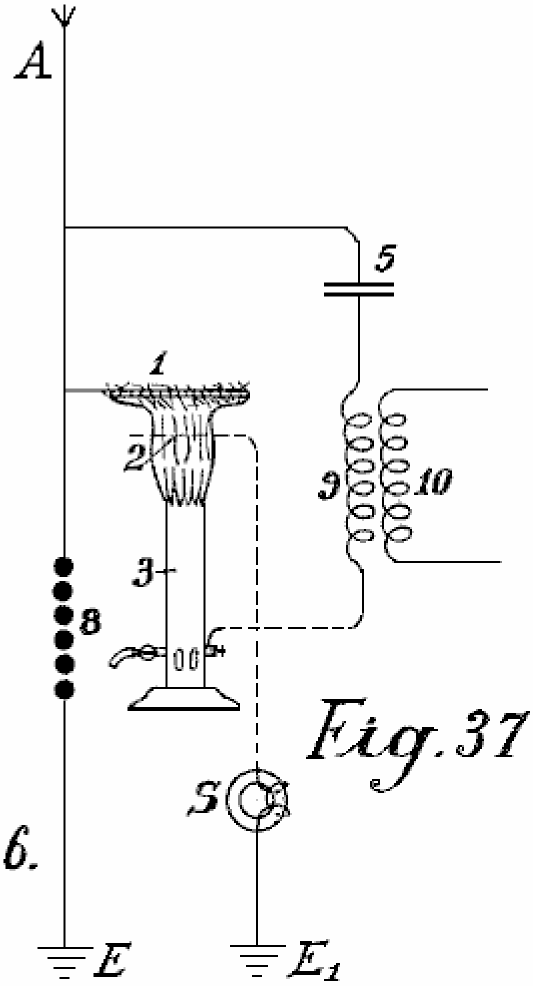
A
5
1
2
9
10
3
8
S
6.
E
E_1

Fig.37

The positive pole **1** of the radiating collectors is always directly connected to the aerial collecting conductor **A**. In **Fig.34**, this is further connected over the capacitor set **5** with a second positive electrode **3**. The direct current dynamo **b** produces current which flows over between the electrodes **3** and **2** as an arc light. On the formation of an arc, the negative incandescent electrode **2** absorbs electricity from the positive poles standing opposite it and highly charged with atmospheric electricity which it conveys to the working circuit. The spark gap 7, inductive resistance **9** and induction coil **10** are like the ones previously described. The protective electromagnet **S** protects the installation from earth circuiting and the safety spark gap **8** from excess voltage or overcharging.

In **Fig.35**, the connection is so far altered that the alternating current dynamo feeds the excitation coil **11** of the induction capacitor. **12** is its negative and **13** its positive pole. If the coil **3** on the magnet core of the dynamo is correctly calculated and the frequency of the alternating current sufficiently high, then an arc light can be formed between poles **1** and **2**. As the cathode **2** is connected to the negatively charged earth, and therefore always acts as a negative pole, a form of rectification of the alternating current produced by the dynamo **3** is obtained, since the second half of the period is always suppressed. The working circuit may be carried out in the same way as in **Fig.34**; the working spark gap **7** may however be dispensed with, and instead of it, between the points **n** and **m**, a capacitor **5** and an induction resistance **9** may be inserted, from which, a current is taken inductively.

Fig.36 represents a form of construction similar to that shown in **Fig.34** except that here instead of an arc lamp, a Nernst incandescent body is used. The Nernst lamp is fed through the battery **3**. The working section is connected with the negative pole, the safety spark gap with the positive poles. The working spark gap **7** may also be dispensed with and the current for it taken at **12** over the oscillation circuit **5, 11** (shown in dotted lines).

Flame collectors (**Fig.37**) may also be employed according to this invention. The wire network **1** is connected with the aerial collector conductor **A** and the burner with the earth. At the upper end of the burner, long points are provided which project into the flame. The positive electrode is connected with the negative over a capacitor **5** and the induction coil **9** with the earth.

The novelty in this invention is:

(1) The use of incandescent cathodes opposite positive poles which are connected to large metallic capacities as automatic collecting surfaces.
(2) The connection of the incandescent cathodes to the earth whereby, in addition to the electricity conveyed to them from the battery of machine which causes the incandescing, also the negative charge of the earth potential is conveyed, and
(3) The connection of the positive and negative poles of the radiating collectors over a capacitor circuit alone or with the introduction of a suitable inductive resistance, whereby simultaneously an oscillatory oscillation circuit may be obtained. The collecting effect is by these methods quite considerably increased.

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 1

Fig. 1.

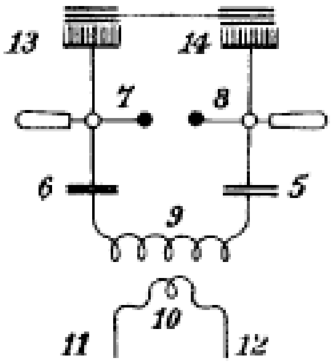
13
14
7
8
6
5
9
10
11
12

Fig. 2.

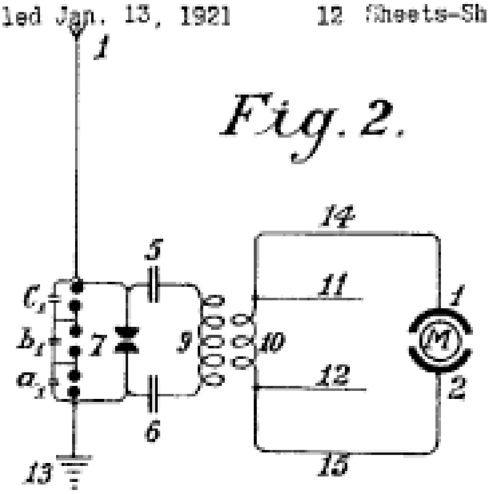
1
14
5
11
c1
b1
7
9
10
1
M
a1
12
2
6
15
13

Fig. 3.

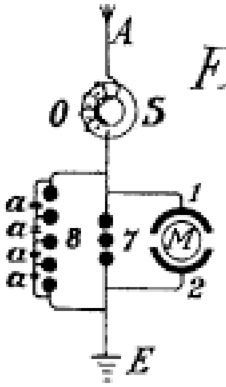
A
0
S
a
a
a
a
8
7
1
M
2
E

Fig. 4.

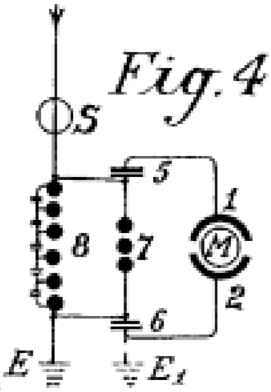
S
5
1
8
7
M
2
6
E
E1

Fig. 5.

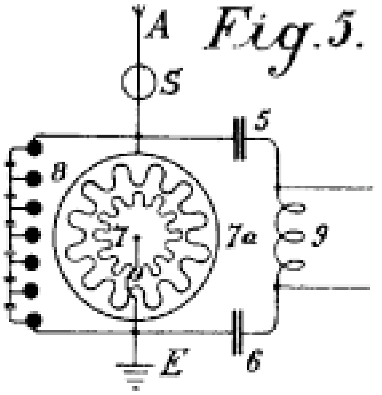
A
S
5
8
7
7a
9
E
6

Fig. 6

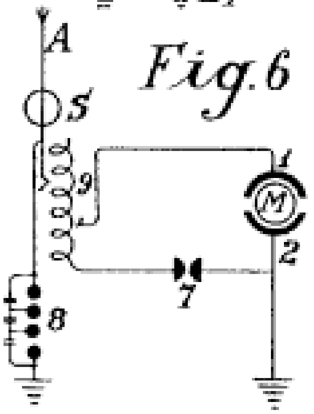

Inventor
Hermann Plauson

by Knight Bros
attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 2

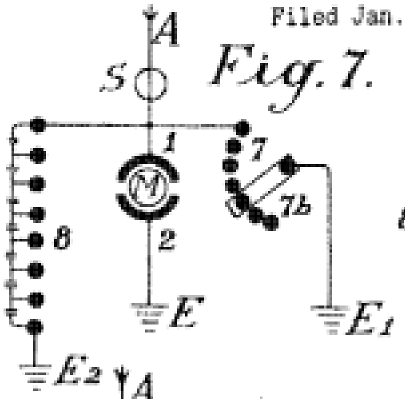
A
S
1
M
2
7
7b
8
E
E1
E2

Fig. 7.

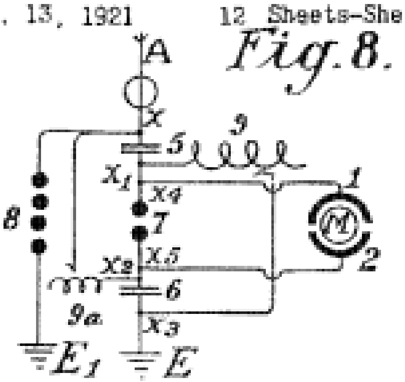
A
X
5
9
X1
X4
1
M
7
X5
2
8
X2
9a
6
X3
E1
E

Fig. 8.

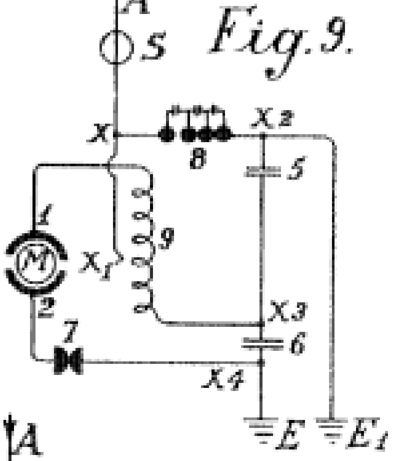
A
S
X
8
X2
5
1
M
X1
9
2
X3
6
7
X4
E
E1

Fig. 9.

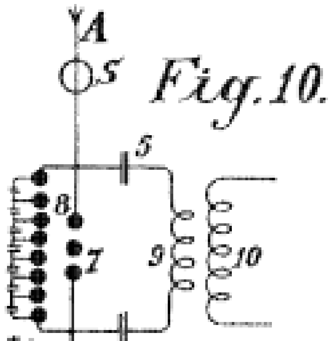
A
S
5
8
7
9
10
6
E

Fig. 10.

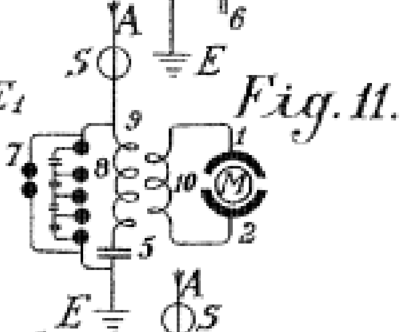
A
S
9
1
7
8
10
M
5
2
E

Fig. 11.

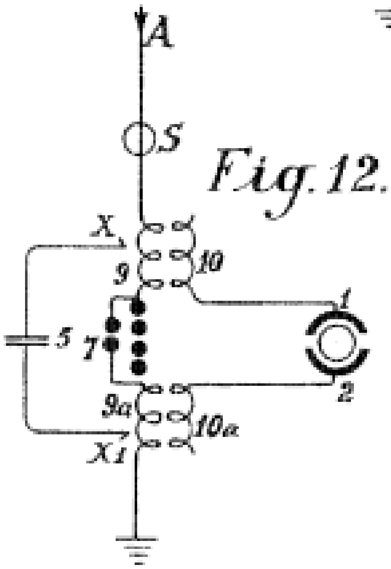
A
S
X
9
10
1
5
7
2
9a
10a
X1

Fig. 12.

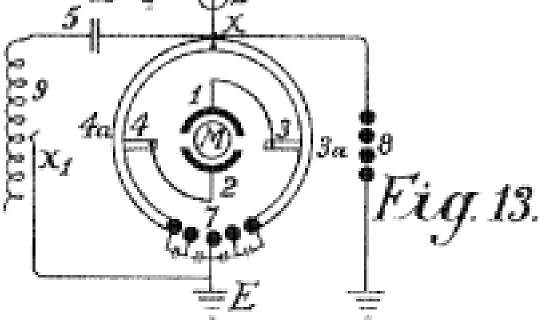
A
S
5
X
9
1
4a
4
M
3
3a
8
X1
2
7
E

Fig. 13.

Inventor

Hermann Plauson

by Knight Bros

attorneys

June 9, 1925. **1,540,998**

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 3

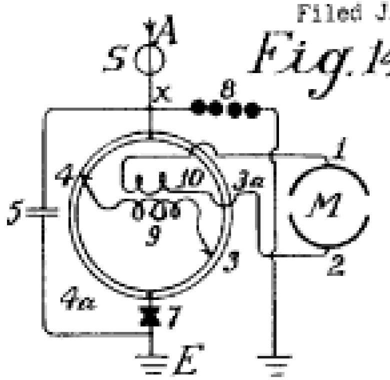
S
A
x
8
1
M
2
4
10
3a
5
9
3
4a
7
E

Fig. 14.

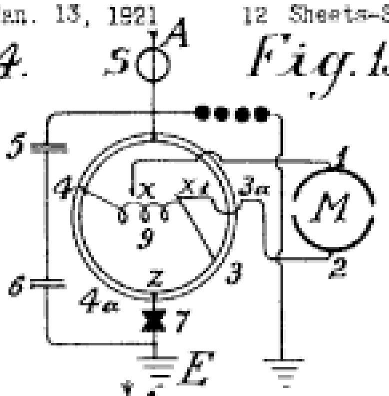
S
A
5
4
X
X1
3a
1
M
2
9
3
6
Z
4a
7
E

Fig. 15.

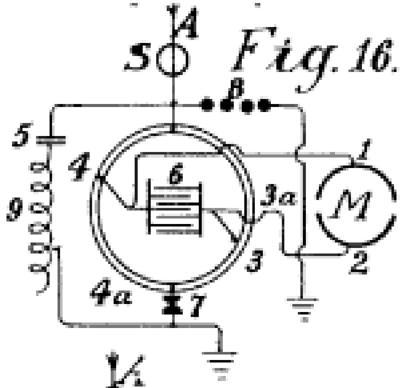
S
A
8
5
4
6
3a
1
M
9
2
3
4a
7

Fig. 16.

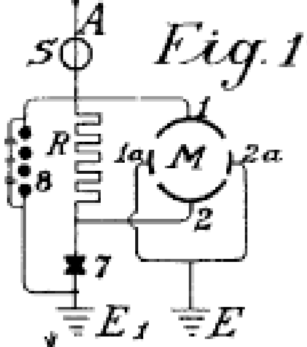
S
A
1
R
1a
M
2a
8
2
7
E1
E

Fig. 17.

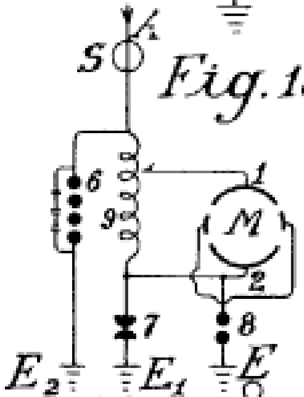
S
A
6
9
1
M
2
7
8
E2
E1
E

Fig. 18.

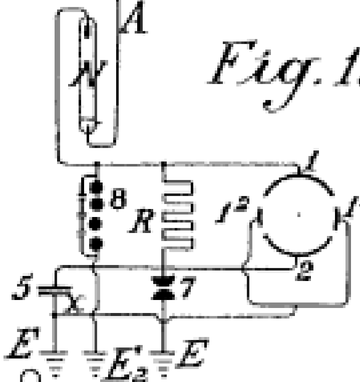
A
N
8
R
1
1²
1¹
2
5
x
7
E
E2
E

Fig. 19.

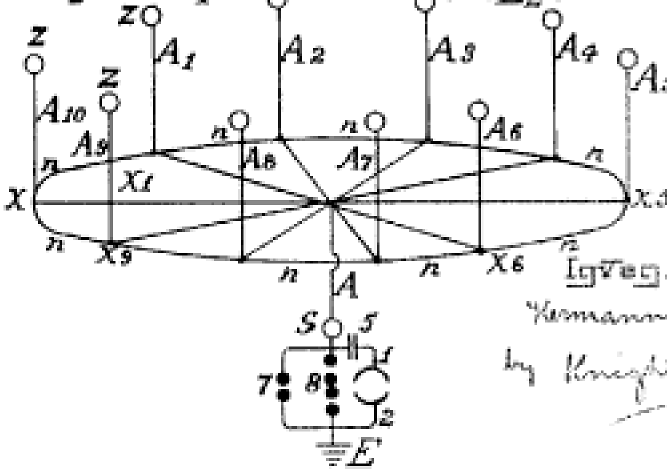
Z
A1
A2
A3
A4
A5
A10
A9
A8
A7
A6
n
X
X1
X5
X9
X6
A
S
5
1
7
8
2
E

Fig. 20.

Inventor
Hermann Plauson
by Knight Bros
Attorneys

June 9, 1925. **1,540,998**

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 4

Fig. 22.

Fig. 21.

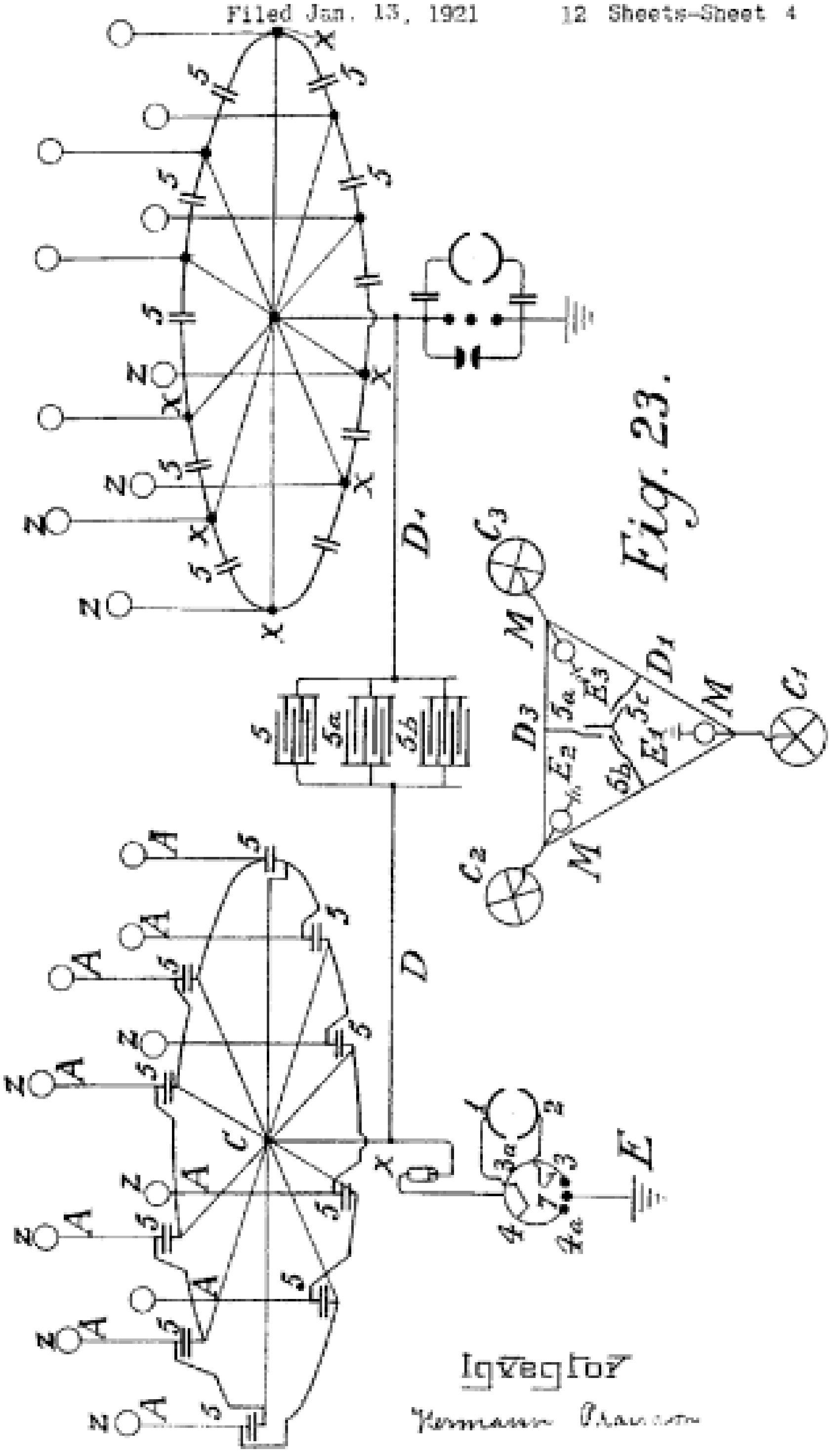
Fig. 23.
5
5a
5b
5c
X
Z
N
A
C
D
D.
D1
D3
E
E1
E2
E3
M
C1
C2
C3
1
2
3
3a
4
4a
7

Inventor
Hermann Plauson

by Knight Bros
Attorneys

June 9, 1925. **1,540,998**

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 5

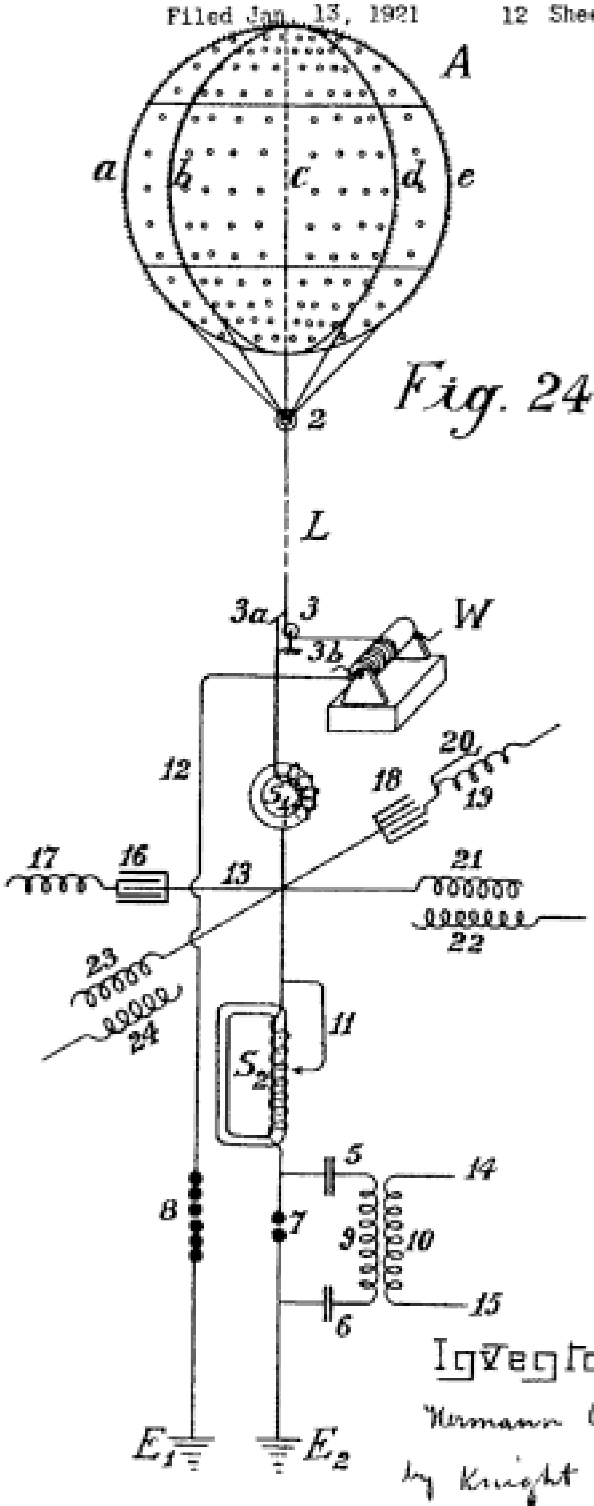
A
a
b
c
d
e
2
L
3a
3
3b
W
12
20
18
19
S_1
17
16
13
21
22
23
24
11
S_2
5
14
8
7
9
10
15
6
E_1
E_2

Fig. 24.

Inventor
Hermann Plauson
by Knight Bros
Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 6

Fig. 25.

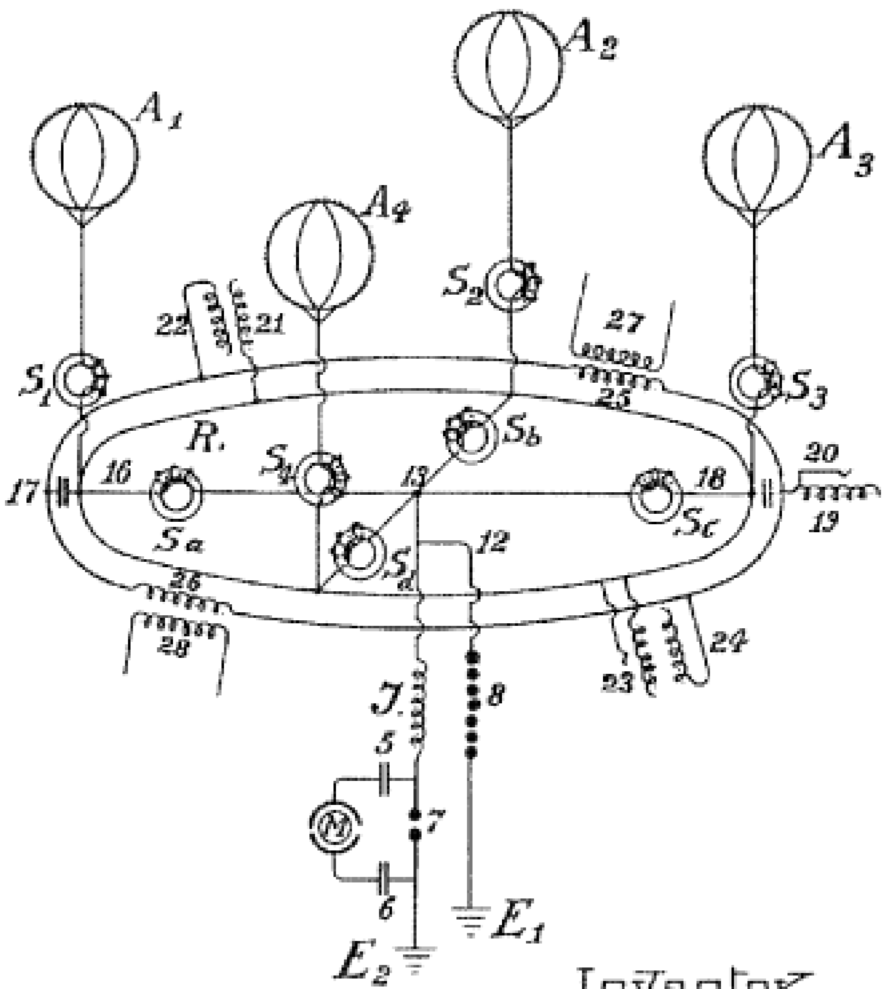

A_1
A_2
A_3
A_4
S_2
27
22
21
S_1
25
S_3
R.
S_b
20
17
16
S_4
13
18
19
S_a
S_d
12
S_c
26
28
24
23
J.
8
5
7
M
6
E_1
E_2

Inventor

Hermann Plauson

by

Knight Bros

Attorneys

June 9, 1925. **1,540,998**

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 7

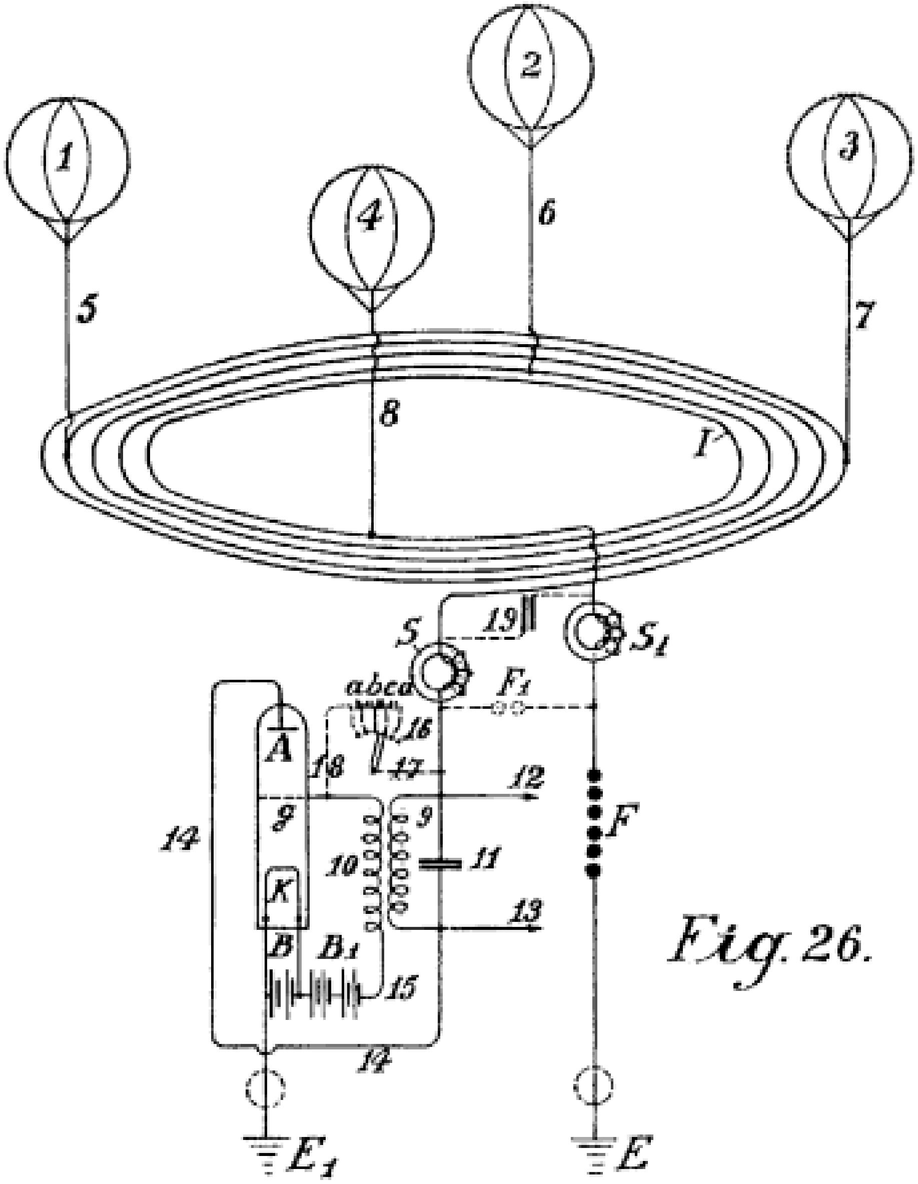

1
2
3
4
5
6
7
8
I'
S
19
S1
abcd
F1
16
A
18
17
12
9
9
14
10
11
F
K
13
B
B1
15
14
E1
E

Fig. 26.

Inventor
Hermann Plauson
by Knight Bros
Attorneys

June 9, 1925. **1,540,998**

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921. 12 Sheets-Sheet 8

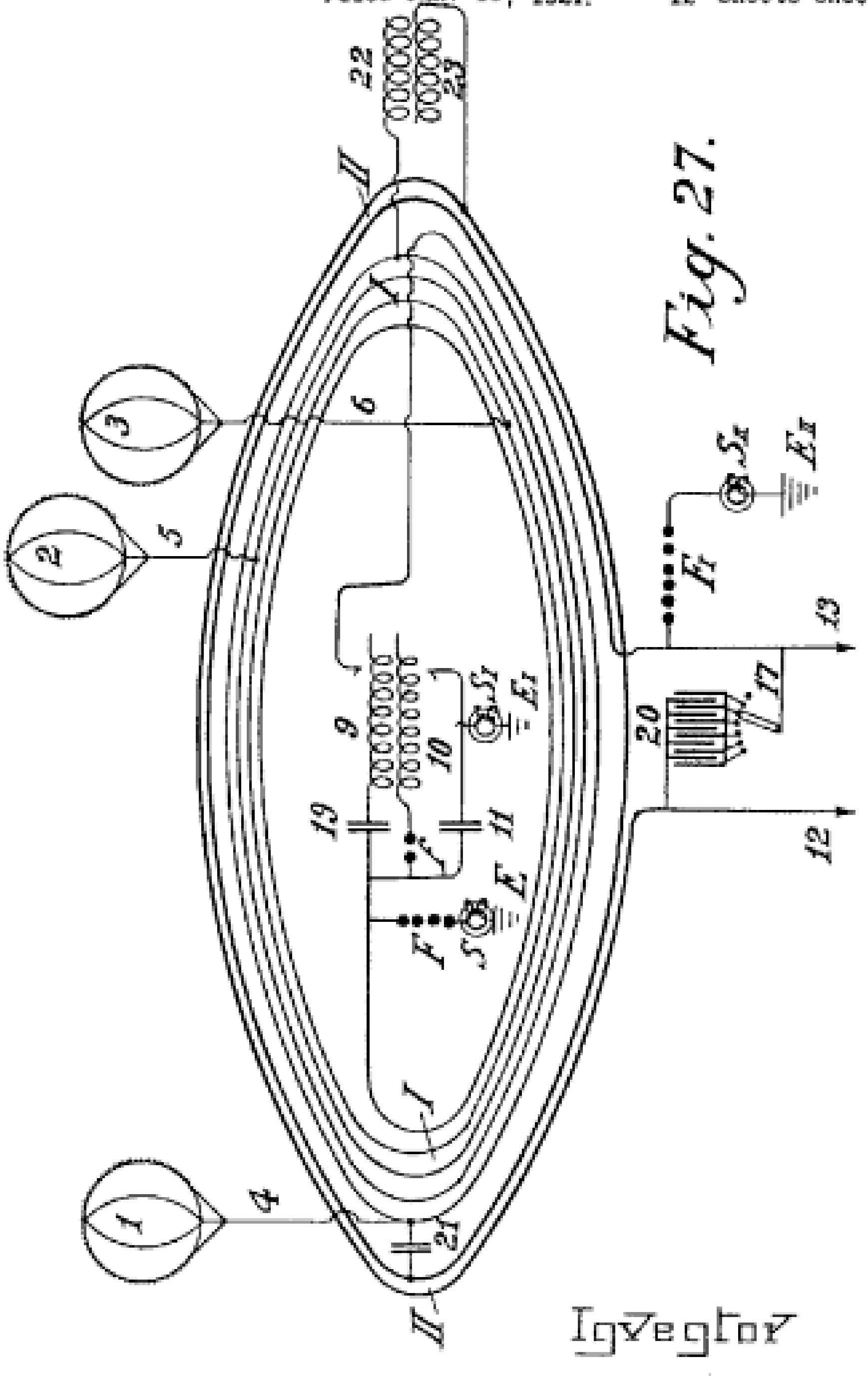
22
23
II
I
6
3
2
5
9
10
1
S_I
E_I
19
11
F
S
E
I
4
1
21
II
20
17
13
12
F_I
S_{II}
E_{II}

Fig. 27.

Inventor

Hermann Plauson

by Knight Bros

Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 9

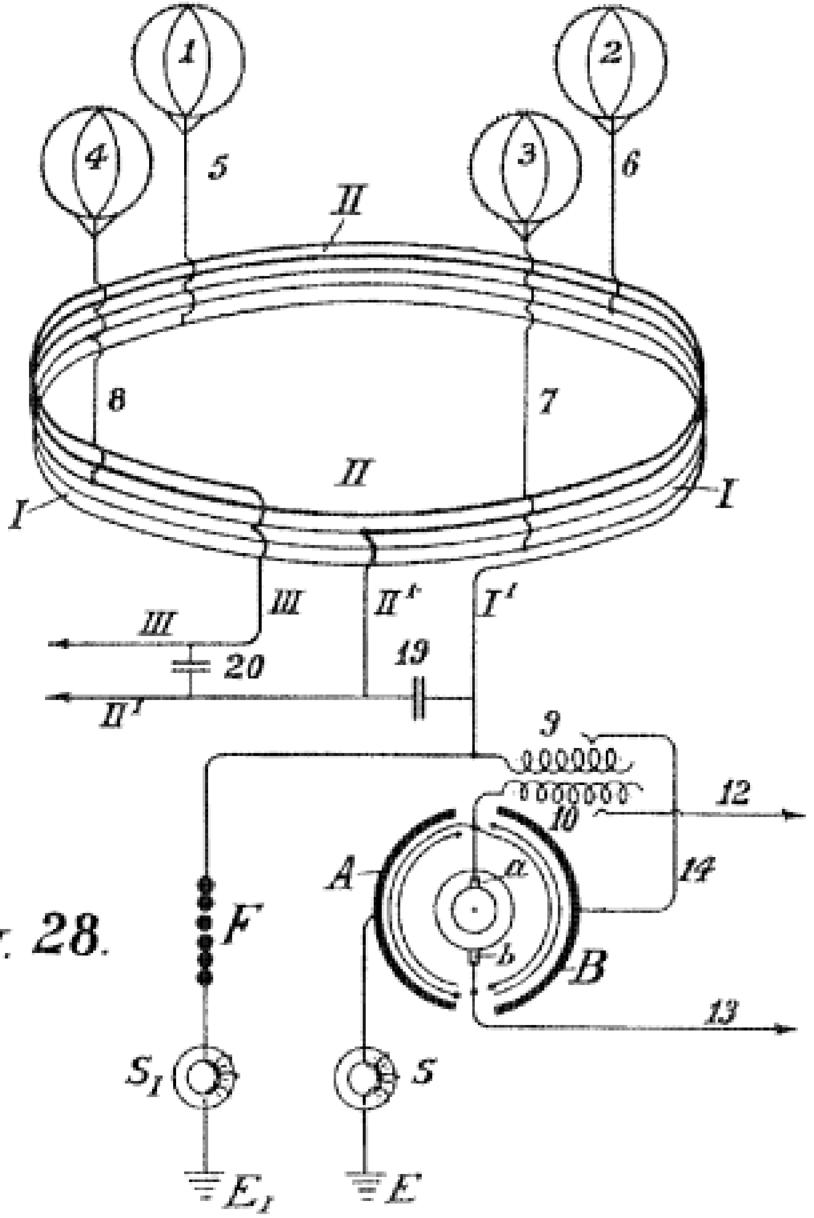
1
2
4
5
3
6
II
8
7
II
I
I
III
II¹
I¹
III
20
19
II¹
9
12
10
14
A
a
F
b
B
13
S₁
S
E₁
E

Fig. 28.

Inventor

Hermann Plauson

by

Knight Bros

attorneys

June 9, 1925. **1,540,998**

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921. 12 Sheets-Sheet 10

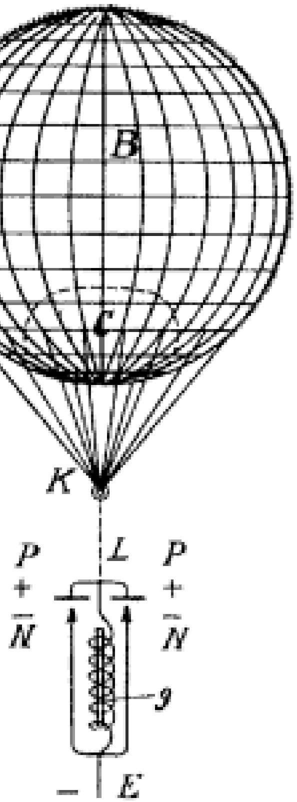

B
C
K
P
+
N
L
P
+
N
9
−
E

Fig. 29.

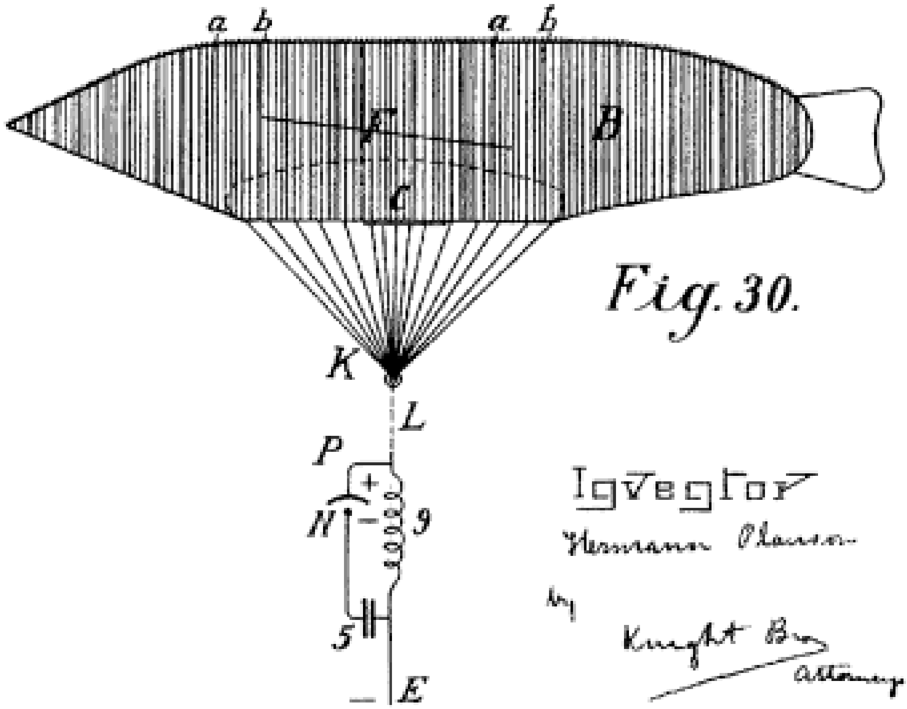

a
b
a
b
F
B
C
K
L
P
+
N
−
9
5
−
E

Fig. 30.

Inventor
Hermann Plauson
by
Knight Bros
Attorneys

June 9, 1925. **1,540,998**

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan 13 1921 12 Sheets-Sheet 11

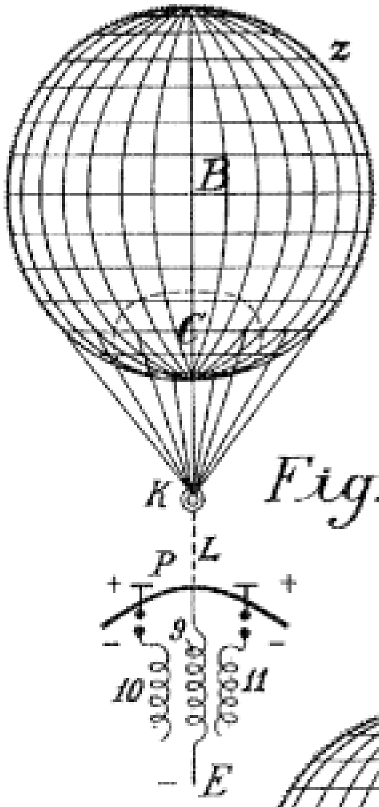
z
B
C
K
L
P
+
+
−
−
9
10
11
−
E

Fig. 31.

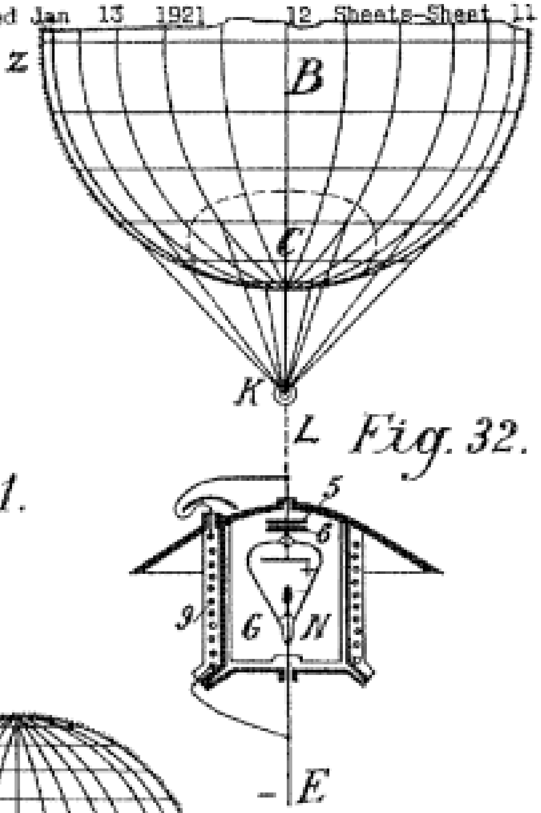
z
B
C
K
L
5
6
9
G
N
+
−
E

Fig. 32.

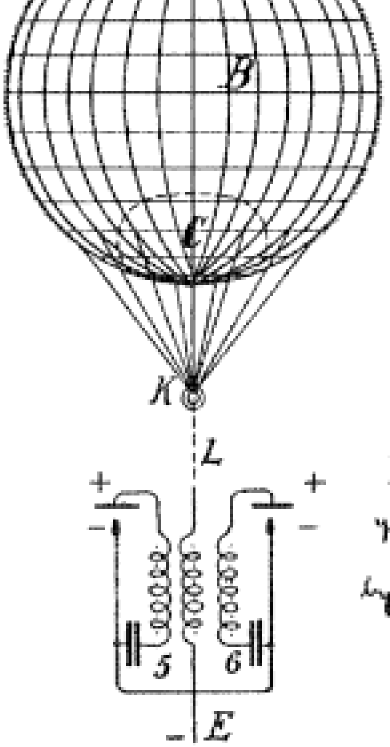

Fig. 33.

Inventor
Hermann Plauson
by Knight Bros
Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 12

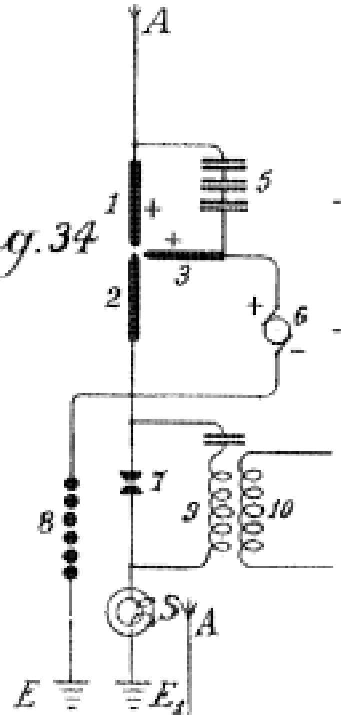
A
1
+
+
5
3
2
+
6
−
7
8
9
10
S
E
E1

Fig. 34

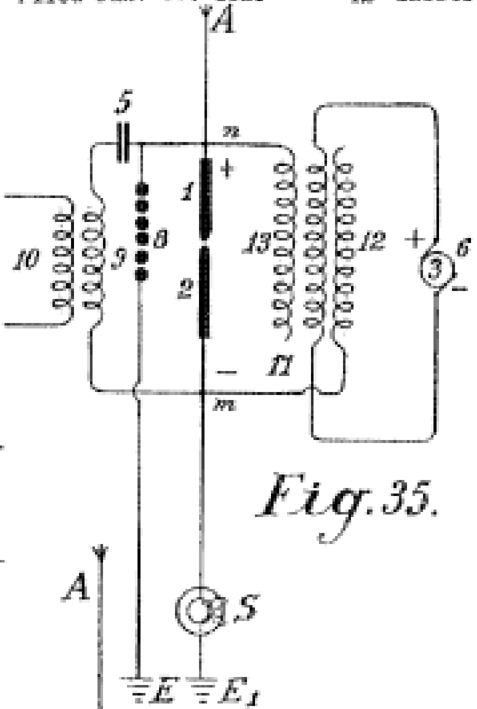
A
5
m
+
1
8
10
9
2
13
12
+
6
−
11
m
S
E
E1

Fig. 35.

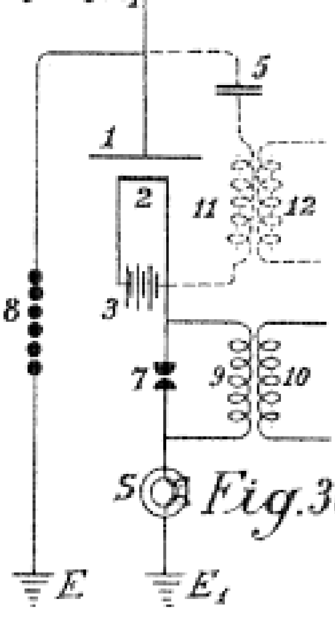
A
5
1
2
11
12
8
3
7
9
10
S
E
E1

Fig. 36.

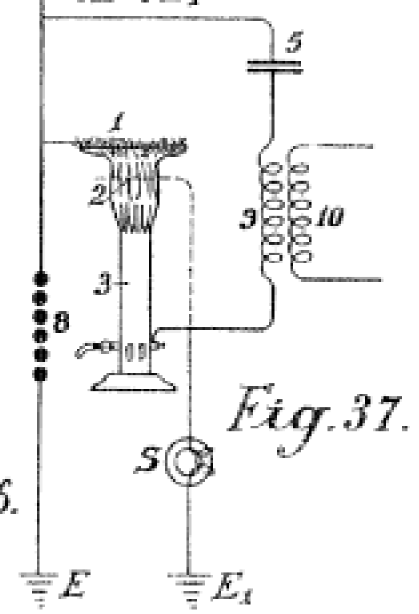
A
5
1
2
9
10
3
8
S
E
E1

Fig. 37.

Inventor

Hermann Plauson

by Knight Bros

Attorneys

Patented June 9, 1925. **1,540,998**

UNITED STATES PATENT OFFICE.

HERMANN PLAUSON, OF HAMBURG, GERMANY.

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY.

Application filed January 13, 1921. Serial No. 437,107.

To all whom it may concern:

Be it known that I, HERMANN PLAUSON, Esthonian subject, residing in Hamburg, Germany, have invented certain new and useful Improvements in the Conversion of Atmospheric Electric Energy, of which the following is a specification.

Methods of obtaining atmospheric electricity by means of metallic nettings set with spikes which are held by means of ordinary or anchored kite balloons made of fabrics and filled with hydrogen, are in theory already known. Atmospheric electricity obtained in this way has been suggested to be used in the form of direct current for the charging of accumulators. This knowledge however is at present only theoretical as the conversion in practice has hitherto been a failure. No means are known of protecting the apparatus from destruction by lightning. The balloons used for collecting the charge must also be made of very large size in order to be able to support the weight of the metallic netting and the heavy cable connections.

Instead of using heavy metallic netting as collectors attached to single air balloons of non-conducting materials which are liable to be torn and are permeable to the gas, it is proposed to use metallic balloon collectors which have the following important advantages—

(*a*) The metallic cases are impenetrable to helium and hydrogen; they also represent large metallic weather-proof collecting surfaces.

(*b*) Radio active means and the like may be easily applied internally or externally; whereby the ionization is considerably increased and therewith also the quantity of atmospheric electricity capable of being collected.

(*c*) Such balloon collectors of light metal do not require to be of large size as they have to carry only their own moderate weight, and that of the conducting cable or wire.

(*d*) The entire system therefore offers little surface for the action of storm and wind and is resistant and stable.

(*e*) Each balloon can be easily raised and lowered by means of a winch so that all repairs, recharging and the like can be carried out without danger during the operation.

It is further proposed to use a collecting aerial network of several separate collectors spread out in the air above the earth, which collectors are interconnected by electrical conductors.

According to this invention charges of atmospheric electricity are not directly converted into mechanical energy, and this forms the main difference from previous inventions, but the static electricity which runs to earth through aerial conductors in the form of direct current of very high voltage and low current strength is converted into electro-dynamic energy in the form of high frequency vibrations. Many advantages are thereby obtained and all disadvantages avoided.

The very high voltage of static electricity of a low current strength can be converted by this invention to voltages more suitable for technical purposes and of greater current strength. By the use of closed oscillatory circuits it is possible to obtain electromagnetic waves of various amplitude and thereby to increase the degree of resonance of such current. Such resonance allows various values of inductance to be chosen whereby again the governing of the starting and stopping of machines driven thereby by simply tuning the resonance between coils of the machine and the transformer circuit forming the resonance can easily be obtained. Further, such currents have the property of being directly available for various uses, even without employing them for driving motors, of which there may be particularly mentioned, lighting, production of heat and use in electro-chemistry.

Further, with such currents a series of apparatus may be fed without direct current supply through conductors and also the electro-magnetic high frequency currents may be converted by means of special motors adapted for electro-magnetic oscillations into mechanical energy, or finally converted by special machines into alternating current of low frequency or even into direct current of high potential.

The invention is more particularly described with reference to the accompanying diagrams in which:—

Figure 1 is an explanatory figure.

Figure 2 is a diagrammatic view of the simplest form.

Figure 3 shows a method of converting atmospheric electrical energy for use with motors.

Figure 4 is a diagram showing the use of protective means.

Figure 5 is a diagram of an arrangement for converting large current strengths.

Figure 6 is a diagram of an arrangement including controlling means.

Figure 7 shows means whereby the spark gap length can be adjusted.

Figure 8 shows a unipolar connection for the motor.

Figure 9 shows a weak coupled system suitable for use with small power motors.

Figures 10, 11 and 12 show modified arrangements.

Figure 13 shows a form of inductive coupling for the motor circuit.

Figure 14 is a modified form of Figure 13 with inductive coupling.

Figure 15 is an arrangement with non-inductive motor.

Figure 16 is an arrangement with coupling by condenser.

Figures 17, 18 and 19 are diagrams of further modifications.

Figure 20 shows a simple form in which the aerial network is combined with special collectors.

Figure 21 shows diagrammatically an arrangement suitable for collecting large quantities of energy.

Figure 22 is a modified arrangement having two rings of collectors.

Figure 23 shows the connections for three rings of collectors.

Figure 24 shows a collecting balloon and diagram of its connection of condenser batteries.

Figures 25 and 26 show modified collector balloon arrangements.

Figure 27 shows a second method of connecting conductor for the balloon aerials.

Figure 28 shows an auto-transformer method of connection.

Figure 29 shows the simplest form of construction with incandescent cathode.

Figure 30 shows a form with cigar shaped balloon.

Figure 31 is a modified arrangement.

Figure 32 shows a form with cathode and electrode enclosed in a vacuum chamber.

Figure 33 is a modified form of Figure 32.

Figure 34 shows an arc light collector.

Figure 35 shows such an arrangement for alternating current.

Figure 36 shows an incandescent collector with Nernst lamp.

Figure 37 shows a form with a gas flame.

Figure 1 illustrates a simple diagram for converting static electricity into dynamic energy of a high number of oscillations. For the sake of clearness in the drawings an influence machine is assumed to be employed and not an aerial antenna. 13 and 14 are combs for collecting the static electricity of the influence machine. 7 and 8 are spark discharging electrodes, 6 and 5 condensers, 9 an inductive primary coil, 10 secondary coil, 11 and 12 ends of conductors of the secondary coil 10. When the disc of the static influence machine is rotated by mechanical means, the combs collect the electric charges one the positive and the other the negative, and charge the condensers 5 and 6 until such a high potential is formed across the spark gap 7—8, that the spark gap is jumped. As the spark gap 7—8 forms a closed circuit with condensers 6 and 5, and inductive resistance 9, as is well known, waves of high frequency electromagnetic oscillations will pass in this circuit.

The high frequency of the oscillations produced in the primary circuit induces waves of the same periodicity in the secondary circuit. Thus in the primary circuit electromagnetic oscillations are formed by the passage of the spark over the spark gap and these waves are maintained by fresh charges of static electricity.

By suitably selecting the ratio between the number of the coils in the primary and secondary circuits with regard to a correct application of the co-efficients of resonance (capacity, inductance, and resistance) the high voltage of the primary circuit may be suitably converted into low voltage and high current strength.

When the oscillatory discharges in the primary circuit becomes weaker or entirely cease, the condensers are charged again by the static electricity until the accumulated charge again breaks down the spark gap. All this is repeated as long as electricity is produced by the static machine by employing mechanical energy.

An elementary form of the invention is shown in Figure 2 in which two spark gaps in parallel are used one of which may be termed the working gap 7 in Figure 2, whilst the second serves as a safety device for excess voltage and consists of a larger number of spark gaps than the working section, which gaps are arranged in series and are bridged by very small capacities as is illustrated in a_1, b_1, c_1, Figure 2 which allow of uniform sparking in the safety section.

In Figure 2 A is the aerial antenna for collecting charges of atmospheric electricity. 13 is the earth connection of the second part of the spark gap, 5 and 6 are condensers, 9 a primary coil. Now when through the aerial A the positive atmospheric electricity seeks to combine with the negative charge to earth, this is prevented by (the air gap between) the spark gaps. The resistance of the spark gap 7 is, as shown in the drawings, lower than that of the other safety section which consists of three spark gaps connected in

series, and consequently a three times greater air resistance is offered by the latter.

So long therefore, as the resistance of the spark gap 7 is not overloaded, so that the other spark gaps have an equal resistance with it the discharges take place only over spark gap 7. Should however the voltage be increased by any influences so that it might be dangerous for charging the condensers 5 and 6 or for the coil insulation 9 and 10 in consequence of break down, by a correct regulation of this spark gap the second spark gap can discharge free from inductive effects direct to earth without endangering the machine.

Without this second spark gap, arranged in parallel having a higher resistance than the working spark gap it is impossible to collect and render available large quantities of electrical energy.

The action of this closed oscillation circuit consisting of spark gap 7, two condensers 5 and 6, primary coil 9, and also secondary coil 10 is exactly the same as the one described in Figure 1 with the arrangement of the static induction machine with the only difference that here the second spark gap is provided. The electromagnetic high frequency alternating current obtained can be tapped off from the conductors 11 and 12 for lighting and heating purposes. Special kinds of motors adapted for working with these peculiar electrical charges may be connected at 14 and 15 which can work with static electricity charges or with high frequency oscillations.

In addition to the use of spark gaps in parallel a second measure of security is also necessary for taking off the current. This precaution consists according to this invention, in the introduction of and method of connecting certain protective electromagnets or choking coils in the aerial circuit as shown by S in Figure 3.

A single electromagnet only having a core of the thinnest possible separate laminations is connected with the aerial.

In the case of high voltages in the aerial network or at places where there are frequent thunder storms, several such magnets may however be connected in series.

In the case of large units or plants several electromagnets can be employed in parallel or in series parallel.

The windings of these electromagnets may be simply connected in series with the aerials. In this case the winding preferably consists of several thin parallel wires, which make up together, the necessary section.

The winding may be made of primary and secondary windings in the form of a transformer. The primary winding will be then connected in series with the aerial network, and the secondary winding more or less short-circuited over a regulating resistance or an induction coil. In the latter case it is possible to regulate to a certain extent the effect of the choking coils. In the further description of the connecting and constructional diagrams the aerial electromagnet choke coil is indicated by a simple ring S.

Figure 3 shows the simplest way of converting atmospheric electricity into electromagnetic wave energy by the use of special motors adapted for high oscillatory currents or static charges of electrical energy. Recent improvements in motors for working with static charges and motors working by resonance, that is to say, having groups of tuned electromagnetic cooperating circuits render this possible but such do not form part of the present invention.

A motor adapted to operate with static charges will for the sake of simplicity be diagrammatically indicated by two semicircles 1 and 2 and the rotor of the motor by a ring M. (Figure 3.) A is a vertical aerial or aerial network. S the safety choke or electromagnet with coil O as may be seen is connected with the aerial A. Adjacent the electromagnet S the aerial conductor is divided into three circuits, the circuit 8 giving the safety spark gap, the circuit 7 with the working spark gap, and then a circuit including the stator terminal 1, the rotor and stator terminal 2 at which a connection is made to the earth wire. The two spark gaps are also connected metallically with the earth wire. The method of working these diagrams is as follows:

The positive atmospheric electric charge collected tends to combine with the negative electricity (or earth electricity) connected with the earth wire. It travels along the aerial A through the electromagnet S without being checked as it flows in the same direction as the direct current. Further, its progress is arrested by two sparks gaps placed in the way and the stator condenser surfaces. The stator condenser surfaces are charged until the charge is greater than the resistance of the spark gap 7, whereupon a spark springs over the spark gap 7 and an oscillatory charge is obtained as by means of the motor M, stator surfaces 1 and 2, and spark gap 7, a closed oscillation circuit is obtained for producing the electromagnetic oscillations. The motor here forms the capacity and the necessary inductance and resistance, which, as is well known, are necessary for converting static electricity into electromagnetic wave energy.

The discharges formed are converted into mechanical energy in special motors and can not reach the aerial network by reason of the electromagnet or choke. If, however, when a spark springs over the spark gap 7 a greater quantity of atmospheric electricity tends to flow to earth, a counter voltage is

induced in the electromagnet, which is greater the more rapidly and strongly the flow of current direct to the earth is. By the formation of this opposing voltage a sufficiently high resistance is offered to the flow of atmospheric electricity direct to earth to prevent a short circuit with the earth.

The circuit containing spark gap 8 having a different wave length which is not in resonance with the natural frequency of the motor, does not endanger the motor and serves as security against excess voltage, which, as practical experiments have shown, may still arise in certain cases, but can be conducted direct to earth through this spark gap.

In the diagram illustrated in Figure 4 the spark gap 7 is shunted across condensers 5 and 6 from the motor M. This construction affords mainly a better insulation of the motor against excess voltage and a uniform excitation through the spark gap 7.

In Figure 5 a diagram is illustrated for transforming large current strengths which may be employed direct without motors, for example, for lighting or heating purposes. The main difference is that here the spark gap consists of a star shaped disc 7 which can rotate on its own axis and is rotated by a motor opposite similarly fitted electrodes 7^a. When separate points of stars face one another, discharges take place, thus forming an oscillation circuit over condensers 5 and 6 and inductance 9 for oscillatory discharges. It is evident that a motor may also be directly connected to the ends of the spiral 9.

The construction of the diagram shown in Figure 6 permits of the oscillation circuit of the motor being connected with an induction coil. Here a regulating inductive resistance is introduced for counter-acting excess voltages in the motor. By cutting the separate coils 9 (coupled inductively to the aerial) in or out the inductive action on the motor may be more or less increased or variable aerial action may be exerted on the oscillation circuit.

In Figure 7 the oscillation circuit is closed through the earth (E and E_1). The spark gap 7 may be prolonged or shortened by more or fewer spark gaps being successively connected by means of a contact arm 7^b.

Diagram 8 shows a unipolar connection of the motor with the aerial network. Here two oscillation circuits are closed through the same motor. The first oscillation circuit passes from aerial A through electromagnet S, point x, inductance 9^a to the earth condenser 6 and further, over spark gap 7 to the aerial condenser 5 and back to x. The second oscillation circuit starts from the aerial condenser 5 at the point x^1 over the inductance 9 to the earth condenser 6 at the point x^2 and through the condenser 6 over the spark gap 7 back to x^1. The motor itself is inserted between the two points of the spark gap 7. From this arrangement slightly damped oscillation wave currents are produced.

In the diagram illustrated in Figure 9 a loosely coupled system of connections is illustrated which is assumed to be for small motors for measuring purposes. A indicates the aerial conductor, S the electromagnet in the aerial conductor, 9 the inductance, 7 the spark gap, 5 and 6 condensers, E the earth, M the motor, and 1 and 2 stator connections of the motor. The motor is directly metallically connected with the oscillation circuit.

In Figure 10 a purely inductive coupling is employed for the motor circuit. The motor is connected with the secondary wire 10 as may be seen in Figure 11 in a somewhat modified diagram connection. The same applies to the diagram of Figure 12.

The diagrams hitherto described preferably allow of motors of small and medium strength to be operated. For large aggregates, however, they are too inconvenient as the construction of two or more oscillation circuits for large amounts of energy is difficult; the governing is still more difficult and the danger in switching on or off is greater.

A means of overcoming such difficulties is shown in Figure 13. The oscillation circuit here runs starting from the point x over condenser 5, variable inductance 9, spark gap 7 and the two segments (3^a and 4^a) forming arms of a Wheatstone bridge, back to x. If the motor is connected by brushes 3 and 4 transversely to the two arms of the bridge as shown in the drawings, electromagnetic oscillations of equal sign are induced in the stator surfaces 1 and 2 and the motor does not revolve. If however, the brushes 3 and 4 are moved in common with the conducting wires 1 and 2 which connect the brushes with the stator poles a certain alteration or displacement of the polarity is obtained and the motor commences to revolve.

The maximum action will result if one brush 3 comes on the central sparking contact 7 and the other brush 4 on the part x. They are however, usually in practice not brought on to the central contact 7 but only held in the path of the bridge segments 4^a and 3^a in order not to connect the spark gaps with the motor oscillation circuit.

As however, the entire oscillation energy can thereby not act on the motor it is better to carry out the same system according to the diagram 14. The diagram 14 differs from the foregoing only by the motor not being directly metallically connected with the segments of the commutator, but only a primary coil 9 which induces in a secondary coil 10, current which feeds the motor M and takes the place of the rotor. By this

arrangement a good transforming action is obtained, a loose coupling and also an oscillation circuit without a spark gap.

In Figure 15 the motor is not purely inductively as in 14, but directly metallically branched off from the primary coil (at x and x^1) after the principle of the auto-transformer.

In Figure 16 instead of an inductance a condenser 6 is in similar manner, and for the same object inserted between the segments 3^a and 4^a. This has the advantage that the segments 3^a and 4^a need not be made of solid metal but may consist of spiral coils whereby a more exact regulation is possible and further motors of high inductance may be employed.

The arrangements of Figures 17, 18 and 19 may be employed for use with resonance and particularly with induction condenser motors; between the large stator induction condenser surfaces, small reversing pole condensers are connected, which, as may be seen from Figures 17, 18 and 19 are led together to earth. Such reversing poles have the advantage that with large quantities of electrical energy the spark formation between the separate oscillation circuits ceases.

Figure 19 shows a further method which prevents electromagnetic oscillations of high number of alternations formed in the oscillation circuit striking back to the aerial conductor. It is based on the well known principle that a mercury lamp, one electrode of which is formed of mercury, the other of solid metal such as steel allows an electric charge to pass in only one direction from the mercury to the steel and not vice versa. The mercury electrode of the vacuum tube N is therefore connected with the aerial conductor and the steel electrode with the oscillation circuit. From this it results that charges can pass only from the aerial through the vacuum tube to the oscillation circuit, but not vice versa. Oscillations which are formed on being transformed in the oscillation circuit cannot pass to the aerial conductor.

In practice these vacuum tubes must be connected behind an electromagnet as the latter alone affords no protection against the danger of lightning.

As regards the use of spark gaps, all arrangements as used for wireless telegraphy may be used. Of course the spark gaps in large machines must have a sufficiently large surface. In very large stations they are cooled in liquid carbonic acid or better still in liquid nitrogen or hydrogen; in most cases the cooling may also take place by means of liquefied low homologues of the metal series or by means of hydrocarbons the freezing point of which lies at between −90° C. and −40° C. The spark gap casing must also be insulated and be of sufficient strength to be able to resist any pressure which may arise. Any undesirable excess super-pressure which may be formed must be automatically let off. I have employed with very good results mercury electrodes which were frozen in liquid carbonic acid, the cooling being maintained during the operation from the outside through the walls.

Figure 20 is one of the simplest forms of construction of an aerial network in combination with collectors, transformers and the like illustrated diagrammatically. E is here the earth wire, 8 the safety spark gap, 7 the working spark gap, 1 and 2 the stator surfaces of the motor, 5 a condenser battery, S the protective magnet which is connected with the coil in the aerial conductor, A^1 to A^{10} aerial antennæ with collecting balloons, N horizontal collecting or connecting wires from which, to the centre a number of connections run.

The actual collectors consist of metal sheaths preferably made of an aluminium magnesium alloy, and are filled with hydrogen or helium and are attached to copper plated steel wires. The size of the balloon is selected so that the actual weight of the balloon and the weight of the conducting wire is supported thereby. On the top of the balloon aluminium spikes, made and gilded in a special manner hereinafter described, are arranged in order to produce a conductor action. Small quantities of radium preparations, more particularly polonium-ionium or mesothorium preparations considerably increase the ionization, and therewith the action of these collectors.

In addition to metal balloons, fabric balloons which are superficially metal coated according to Schoop's metal spraying process, may however also be employed. A metallic surface may also be produced by lacquering with metallic bronzes, preferably according to Schoop's spraying process or lacquering with metallic bronze powders in two electrical series of widely different metals, because thereby the collecting effect is considerably increased.

Instead of the ordinary round balloons, elongated cigar shaped ones may be employed. In order also to utilize the frictional energy of the wind, patches or strips of non-conducting substances which produce electricity by friction, may be attached to the metallized balloon surfaces. The wind will impart a portion of its energy in the form of frictional electricity, to the balloon casing, and thereby the collecting effect is substantially increased.

In practice however, very high towers (up to 300 metres is fully admissible) may be employed as antennæ. In these towers copper tubes rise freely further above the top of the tower. A gas lamp secured

against the wind is then lit at the point of the copper tube and a netting is secured to the copper tube over the flame of this lamp to form a collector. The gas is conveyed through the interior of the tube up to the summit. The copper tube must be absolutely protected from moisture at the place at which it enters the tower and also rain must be prevented running down the walls of the tower which might lead to a bad catastrophe. This is done by bell shaped enlargements which expand downwards, being arranged in the tower in the form of high voltage insulators of Siamese pagodas.

Special attention must be devoted to the foundations of such towers. They must be well insulated from the ground, which may be obtained by first embedding a layer of concrete in a box form to a sufficient depth in the ground and inserting in this an asphalt lining and then glass bricks cast about 1 or 2 metres in thickness. Over this in turn there is a ferro-concrete layer in which alone the metal foot of the tube is secured. This concrete block must be at least 2 metres from the ground and be fully protected at the sides by a wooden covering, from moisture. In the lower part of the tower a wood or glass house for the large condenser batteries or for the motors may be constructed. In order to lead the earth connection to the ground water, a well insulated pit constructed of vitreous bricks, must be provided. Several such towers are erected at equal distances apart and connected with a horizontal conductor. The horizontal connecting wires may either run directly from tower to tower or be carried on bell shaped insulators similar to those in use for high voltage conductors. The width of the network may be of any suitable size and the connection of the motors can take place at any suitable places.

In order to collect large quantities of electricity with few aerials it is well to provide the aerial conductor with batteries of condensers as shown in two methods of construction in Figures 21 and 22. In Figure 21 the batteries of condensers 5 are connected on the one hand with the aerial electricity collectors Z by the aerial conductor A, and on the other hand interconnected in series with an annular conductor from which horizontal conductors run to the connecting points C to which the earth wire is connected.

Figure 22 shows a similar arrangement. Should two such series of antennæ rings be shown by a voltmeter to have a large difference of potential (for example, one in the mountains and one in the plain) or even of different polarity these differences may be compensated for by connecting sufficiently large condenser batteries (5, 5^a, 5^b) by means of Maji star conductors D and D^1. In Figure 23 a connection of three such rings of collectors to form a triangle with a central condenser battery is illustrated.

The condenser batteries of such large installations must be embedded in liquefied gases or in liquids freezing at very low temperatures. In such cases a portion of the atmospheric energy must be employed for liquefying these gases. It is also preferable to employ pressure. By this means the condenser surfaces may be diminished, and still allow for large quantities of energy to be stored, secure against breakdown. For smaller installations the immersing of the condensers in well insulated oil or the like, suffices. Solid substances on the other hand cannot be employed as insulators.

The arrangement in the diagrams hitherto described was always such that the condenser batteries were connected with both poles directly to the aerial conductors. An improved diagram of the connections for obtaining atmospheric electricity for the condenser batteries has however, been found to be very advantageous, this arrangement consists in that they are connected by only one pole (unipolar) to the collecting network. Such a method of arrangement is very important, as by means of it a constant current and an increase of the normal working pressure or voltage is obtained. If for example a collecting balloon aerial which is allowed to rise to a height of 300 metres, shows 40,000 volts above earth voltage, in practice it has been found that the working voltage (with a withdrawal of the power according to the method hereinbefore described by means of oscillating spark gaps and the like) is only about 400 volts. If however, the capacity of the condenser surfaces be increased, which capacity in the above mentioned case was equal to that of the collecting surface of the balloon aerials, to double the amount, by connecting the condenser batteries with only one pole, the voltage rises under an equal withdrawal of current up to and beyond 500 volts. This can only be ascribed to the favourable action of the connecting method.

In addition to this substantial improvement it has also been found preferable to insert double inductances with electromagnets and to place the capacities preferably between two such electromagnets. It has also been found that the useful action of such condensers can be further increased if an induction coil be connected as inductive resistance to the unconnected pole of the condenser, or still better if the condenser itself be made as an induction condenser. Such a condenser may be compared with a spring which when compressed carries in itself accumulated force, which it again gives off when released. In charging, a charge with reversed sign is formed at the

other free condenser pole, and if through the spark gap a short circuit results, the accumulated energy is again given back since now new quantities of energy are induced at the condenser pole connected with the conductor network, which in fact charges with opposite signs to that at the free condenser pole. The new induced charges have of course the same sign as the collector network. The whole voltage energy in the aerial is thereby however increased. In the same space of time larger quantities of energy are accumulated than is the case without such inserted condenser batteries.

In Figures 24 and 25 two different diagrams of connections are more exactly illustrated, Figure 24 shows a collecting balloon and the diagram of the connections to earth. Figure 25 four collecting balloons and the parallel connection of the condenser batteries belonging thereto.

A is the collecting balloon made of an aluminium magnesium alloy (electron metal, magnalium) of a specific gravity of 1.8 and a thickness of plate 0.1 to 0.2 mm. Inside there are eight strong vertical ribs of T shaped section about 10 to 20 mm. in height and about 3 mm. in thickness with the projecting part directed inwards (indicated by *a*, *b*, *c*, *d* and so forth); they are riveted together to form a firm skeleton and are stiffened in a horizontal direction by two cross ribs. The ribs are further connected with one another internally and transversely by means of thin steel wires, whereby the balloon obtains great power of resistance and elasticity. Rolled plates of 0.1 to 0.2 mm. in thickness made of magnalium alloy are then either soldered or riveted on this skeleton so that a fully metallic casing with smooth external surface is obtained. Well silvered or coppered aluminium plated steel wires run from each rib to the fastening ring 2. Further, the coppered steel hawser L preferably twisted out of separate thin wires (shown in dotted lines in Figure 24) and which must be long enough to allow the balloon to rise in the desired height, leads to a metal roller or pulley 3 and from thence to a winch W, well insulated from the earth. By means of this winch, the balloon, which is filled with hydrogen, or helium, can be allowed to rise to a suitable height (300 to 5,000 metres) and brought to the ground for recharging or repairs.

The actual current is taken directly through a friction contact from the metal roller 3 or from the wire, or even from the winch or simultaneously from all three by means of brushes (3, 3^a and 3^b). Beyond the brushes the conductor is divided, the paths being:—firstly over 12 to the safety spark gap 8, from thence to the earth conductor E^1, and secondly over electromagnet S^1, point 13, to a second loose electromagnet having an adjustable coil S^2, then to the spark gap 7 and to the second earth conductor E^2. The actual working circuit is formed through the spark gap 7, condensers 5 and 6, and through the primary coil 9; here the static electricity formed by oscillatory discharges is accumulated and converted into high frequency electromagnetic oscillations. Between the electromagnets S^1 and S^2 at the crossing point 13, four condenser batteries are introduced which are only indicated diagrammatically in the drawings each by one condenser. Two of these batteries (16 and 18) are made as plate condensers and prolonged by regulating induction coils or spirals 17 and 19 while the two others (21 and 23) are induction condensers. As may be seen from the drawings each of the four condenser batteries 16, 18, 21, 23 is connected only by one pole to the aerial or to the collector conductor. The second poles 17, 19, 22, 24 are open. In the case of plate condensers having no inductive resistance an induction coil is inserted. The object of such a spiral or coil is the displacement of phase of the induction current by ¼ periods, whilst the charging current of the condenser poles which lie free in the air, works back to the collector aerial. The consequence of this is that in discharges in the collector aerial the back inductive action of the free poles allows a higher voltage to be maintained in the aerial collecting conductor than would otherwise be the case. It has also been found that such a back action has an extremely favourable effect on the wear of the contacts. Of course the inductive effect may be regulated at will within the limits of the size of the induction coil, the length of the coil in action being adjustable by means of wire connection without induction (see Fig. 24, No. 20).

S^1 and S^2 may also be provided with such regulating devices in the case of S^2 (illustrated by 11). If excess voltage be formed it is conducted to earth through the wire 12 and spark gap 8 or through any other suitable apparatus, since this formation would be dangerous for the other apparatus.

The action of these condenser batteries has already been hereinbefore described.

The small circles on the collector balloon indicate places at which zinc amalgam or gold amalgam or other photoelectric acting metals in the form of small patches in extremely thin layers (.01 to .05 mm. in thickness) are applied to the balloon casing of light metal. Such metallic patches may also be applied to the entire balloon as well as in greater thickness to the conducting network. The capacity of the collector is thereby considerably strengthened at the surface. The greatest possible effect in collecting may be obtained by polonium amalgams and the like. On the surface of the

collector balloon metal points or spikes are also fixed along the ribs, which spikes serve particularly for collecting the collector charge. Since it is well known that the resistance of the spikes is less the sharper the spike is, for this purpose it is therefore extremely important to employ as sharp spikes as possible. Experiments made as regards these have shown that the formation of the body of the spike or point also plays a large part, for example, spikes made of bars or rollers with smooth surfaces, have a many times greater point resistance as collector accumulator spikes than those with rough surfaces. Various kinds of spike bodies have been experimented with for the collector balloons hereinbefore mentioned. The best results were given by spikes which were made in the following way. Fine points made of steel, copper, nickel, or copper and nickel alloys, were fastened together in bundles and then placed as anode with the points in a suitable electrolyte (preferably in hydrochloric acid or muriate of iron solutions) and so treated with weak current at 2 to 3 volts pressure. After 2 to 3 hours according to the thickness of the spikes or pins the points become extremely sharp and the bodies of the spikes have a rough surface. The bundle can then be removed and the acid washed off with water. The spikes are then placed as cathode in a bath consisting of solution of gold, platinum, iridium, paladium or wolfram salts or their compounds and coated at the cathode galvanically with a thin layer of precious metal, which must however be sufficiently firm to protect them from atmospheric oxidation.

Such spikes act at a 20 fold lower voltage almost as well as the best and finest points made by mechanical means. Still better results are obtained if polonium or radium salts are added to the galvanic bath when forming the protective layer or coating. Such pins have a low resistance at their points and even at one volt and still lower pressures have an excellent collector action.

In Figure 24 the three unconnected poles are not connected with one another in parallel. That is quite possible in practice without altering the principle of the free pole. It is also preferable to interconnect in parallel to a common collector network, a series of collecting aerials.

Figure 25 shows a diagram for such an installation. A^1, A^2, A^3, A^4 are four metal collector balloons with gold or platinum coated spikes which are electrolytically made in the presence of polonium emanations or radium salts, which spikes or needles are connected over four electro-magnets S^1, S^2, S^3, S^4, through an annular conductor R. From this annular conductor four wires run over four further electromagnets S^5, S^6, S^7, S^8, to the connecting point 13. There the conductor is divided, one branch passing over 12 and the safety spark gap 8 to the earth at E^1, the other over inductive resistance J and working spark gap 7 to the earth at E^2. The working circuit, consisting of the condenser 5 and 6 and a resonance motor or a condenser motor M, such as hereinbefore described, is connected in proximity round the sparking gap section 7.

Instead of directly connecting the condenser motor of course the primary circuit for high frequency oscillatory current may also be inserted.

The condenser batteries are connected by one pole to the annular conductor R and can be either inductionless (16 and 18) or made as induction condensers as shown by 21 and 23. The free poles of the inductionless condensers are indicated by 17 and 19, those of the induction condensers by 22 and 24. As may be seen from the drawings all these poles 17, 22, 19, 24 may be interconnected in parallel through a second annular conductor without any fear that thereby the principle of the free pole connection will be injured. In addition to the advantages already set forth the parallel connection also allows of an equalization of the working pressure in the entire collector network. Suitably constructed and calculated induction coils 25 and 26 may also be inserted in the annular conductor of the free poles, by means of which a circuit may be formed in the secondary coils 27 and 28 which allows current produced in this annular conductor by fluctuations of the charges or the like appearances to be measured or otherwise utilized.

According to what has been hereinbefore stated separate collector balloons may be connected at equidistant stations distributed over the entire country, either connected directly with one another metallically or by means of intermediate suitably connected condenser batteries through high voltage conductors insulated from earth. The static electricity is converted through a spark gap into dynamic energy of a high number of oscillations and may in such form be coupled as a source of energy by means of a suitable method of connecting, various precautions being observed and with special regulations. The wires leading from the collector balloons have hitherto been connected through an annular conductor without this endless connection, which can be regarded as an endless induction coil, being able to exert any action on the whole conductor system.

It has now been found that if the network conductor connecting the aerial collector balloons with one another is not made as a simple annular conductor, but preferably short circuited in the form of coils over a

condenser battery or spark gap or through thermionic tubes or valves or audions, then the total collecting network exhibits quite new properties. The collection of atmospheric electricity is thereby not only increased but an alternating field may be easily produced in the collector network. Further, the atmospheric electrical forces showing themselves in the higher regions may also be directly obtained by induction. In Figures 26 and 28 a form of construction is shown on the basis of which the further foundations of the method will be more particularly explained.

In Figure 26 1, 2, 3, 4 are metal collector balloons, 5, 6, 7, 8 their metallic aerial conductors and I the actual collector network. This consists of five coils and is mounted on high voltage insulators in the air, on high voltage masts (or with a suitable construction of cable embedded in the earth). One coil has a diameter of 1 to 100 km. or more. S and S^1 are two protective electromagnets, F the second safety section against excess voltage, E its earth conductor and E^1 the earth conductor of the working section. When an absorption of static atmospheric electricity is effected through the four balloon collectors, the current in order to reach the earth connection E^1 must flow spirally through the collector network over the electromagnet S, primary induction coil 9, conductor 14, anode A of the audion tube, incandescent cathode K, as the way over the electromagnet and safety spark gap F offers considerably greater resistance. Owing to the fact that the accumulated current flows in one direction, an electromagnetic alternating field is produced in the interior of the collector network coil, whereby the whole free electrons are directed more or less into the interior of the coil. An increased ionization of the atmosphere is therefore produced. In consequence of this the points mounted on the collector balloon show a considerably reduced resistance and therefore increased static charges between the points on the balloon and the surrounding atmosphere are produced. The result of this is a considerably increased collector effect.

A second effect which could not be obtained otherwise is obtained by the electromagnetic alternating field which running parallel to the earth surface, acts more or less with a diminishing or increasing effect on the earth magnetic field, whereby in the case of fluctuations in the current a return induction current of reversed sign is always produced in the collector coil by earth magnetism. Now if, however, a constantly pulsating continuous alternating field is produced as stated in the above collector network I, an alternating current of the same periodicity is produced also in the collecting network coil. As the same alternating field is further transmitted to the aerial balloon, the resistance of its points is thereby considerably reduced, whilst the collector action is considerably increased. A further advantage is that positive electrons which collect on the metal surfaces during the conversion into dynamic current produce a so-called drop of potential of the collector area. As an alternating field is present, the negative ions surrounding the collector surfaces, when discharge of the collector surfaces takes place produce by the law of induction, an induction of reversed sign on the collector surface and so forth (that is to say again a positive charge). In addition to the advantages hereinbefore set forth, the construction of connecting conductors in coil form when of sufficiently large diameter, allows of a utilization of energy arising in higher regions also in the simplest way. As is well known electric discharges frequently take place at very great elevations which may be observed, such as St. Elmo's fires or northern lights. These energy quantities have not been able to be utilized up to now. By this invention all these kinds of energy, as they are of an electromagnetic nature and the direction of the axis of the collector coils stands at right angles to the earth's surface, can be more or less absorbed in the same way as a receiver in wireless telegraphy absorbs waves coming from a far distance. With a large diameter of the spiral it is possible to connect large surfaces and thereby to take up also large quantities of energy.

It is well known that large wireless stations in the summer months, and also in the tropics are very frequently unable to receive the signals in consequence of interruptions which are caused by atmospheric electricity, and this takes place with vertical coils of only 40 to 100 m. diameter. If on the contrary horizontal coils of 1 to 100 km. diameter be employed very strong currents may be obtained through discharges which are constantly taking place in the atmosphere. Particularly in the tropics or still better in the polar regions where the northern lights are constantly present, large quantities of energy may probably be obtained in this way. A coil with several windings should act the best. In similar manner any alteration of the earth magnetism should act inductively on such a coil.

It is not at all unlikely that earthquakes and spots on the sun will also produce an induction in such collector coils of sufficient size. In similar manner this collector conductor will react on earth currents more particularly when they are near the surface of the earth or even embedded in the earth. By combining the previous kind of current collectors so far as they are adapted for

the improved system with the improved possibilities of obtaining current the quantities of free natural energy which are to be obtained in the form of electricity are considerably increased.

In order to produce in the improved collector coil uniform current oscillations of an undamped nature so-called audion high vacuum or thermionic tubes of suitable connection are employed instead of the previously known spark gaps (Fig. 26, Nos. 9–18). The main aerial current flows through electromagnet S (which in the case of a high number of alternations is not connected here but in the earth conductor E^1) and may be conveyed over the primary coils in the induction winding through wire 14 to the anode A of the high vacuum grid tube. Parallel with the induction resistance 9 a regulating capacity of suitable size, such as condenser 11 is inserted. In the lower part of the vacuum grid tube is arranged the incandescent filament or the cathode K which is fed through a battery B. From the battery B two branches run, one to the earth conductor E^1 and the other through battery B^1 and secondary coil 10 to the grid anode *g* in the vacuum tube. By the method of connections shown in dotted lines, a desired voltage at the grid electrode *g* may also be produced through the wire 17 which is branched off from the main current conductor through switches 16 and some small condensers (*a*, *b*, *c*, *d*) connected in series, and conductor 18, without the battery B^1 being required.

The action of the entire system is somewhat as follows:—

On the connecting conductor of the aerial collector network being short circuited to earth, the condenser pole 11 is charged and slightly damped oscillations are formed in the short circuited existing oscillation circuit formed of the condenser 11 and self inductance 9. In consequence of the coupling through coil 10, fluctuations of voltage take place in the grid circuit 15 with the same frequency, which, fluctuations in turn influence the strength of the electrode current passing through the high vacuum amplyfying tube and thus produce current fluctuations of the same frequency in the anode circuit. A permanent supply of energy to the oscillation circuits 9 and 10 consequently takes place, until a condition of balance is set up, in which the consumed oscillation energy is equal to that absorbed. Thereby constant undamped oscillations are now produced in the oscillation circuits 9–11.

For regular working of such oscillation producers high vacuum strengthening tubes are necessary and it is also necessary that the grid and anode voltages shall have a phase difference of 180° so that if the grid is negatively charged, then the anode is positively charged and vice versa. This necessary difference of phase may be obtained by most varied connections, for example, by placing the oscillation circuit in the grid circuit or by separating the oscillation circuit and inductive coupling from the anodes and the grid circuit and so forth.

A second important factor in this way of converting static atmospheric electricity into undamped oscillations is that care must be taken that the grid and anode voltages have a certain relation to one another; the latter may be obtained by altering the coupling and a suitable selection of the self induction in the grid circuit, or as shown by dotted lines 18, 17, 16 by means of a larger or smaller number of condensers of suitable size connected in series; in this case the battery B^1 may be omitted. With a suitable selection of the grid potential a glow discharge takes place between the grid *g* and the anode A, and accordingly at the grid there is a cathode drop and a dark space is formed. The size of this cathode drop is influenced by the ions which are emitted in the lower space in consequence of shock ionization of the incandescent cathodes K and pass through the grid in the upper space. On the other hand the number of the ions passing through the grid is dependent on the voltage between the grid and the cathode. Thus if the grid voltage undergos periodic fluctuations (as in the present case) the amount of the cathode drop at the grid fluctuates and consequently the internal resistance of the tube correspondingly fluctuates, so that when a back coupling of the feed circuit with the grid circuit takes place, the necessary means are afforded for producing undamped oscillations and of taking current, according to requirements from the collecting conductor.

The frequency of the undamped oscillations produced is with a suitably loose coupling equal to the self frequency of the oscillation circuits 9 and 10. By a suitable selection of the self induction of the coil 9 and capacity 11 it is possible to extend from frequencies which produce electromagnetic oscillations of only a few metres wave length down to the lowest practical alternating current frequency. For large installations a suitable number of frequency producing tubes in the form of the well known high vacuum transmission tubes of .5 to 2 kw. in size may be connected in parallel so that in this respect no difficulty exists.

The use of such tubes for producing undamped oscillations, and also the construction and method of inserting such transmission tubes in an accumulator or dynamo circuit is known and also that such oscillation producing tubes only work well at voltages of 1,000 up to 4,000 volts, so that on the contrary their use at lower voltages is

considerably more difficult. By the use of high voltage static electricity this method of producing undamped oscillations as compared with that through spark gaps must be regarded as an ideal solution particularly for small installations of outputs of from 1 to 100 kw.

By the application of safety spark gaps, with interpolation of electro-magnets, not only is short circuiting avoided but also the taking up of current is regulated. Oscillation producers inserted in the above way form a constantly acting electromagnetic alternating field in the collector coil, whereby as already stated, a considerable accumulating effect takes place. The withdrawal wire or working wire is connected at 12 and 13, but current may be taken by means of a secondary coil which is firmly or movably mounted in any suitable way inside the large collector coil, i. e. in its electromagnetic alternating field, so long as the direction of its axis runs parallel with that of the main current collecting coil.

In producing undamped oscillations of a high frequency (50,000 per second and more) in the oscillation circuits 9 and 11, electromagnets S and S^1 must be inserted if the high frequency oscillations are not to penetrate the collector coil, between the oscillation producers and the collector coil. In all other cases they are connected shortly before the earthing (as in Figs. 27 and 28).

In Figure 27 a second method of construction of the connecting conductor of the balloon aerials is illustrated in the form of a coil. The main difference consists in that in addition to the connecting conductor I another annular conductor II is inserted parallel to the former on the high voltage masts in the air (or embedded as a cable in the earth) but both in the form of a coil. The connecting wire of the balloon aerials is indicated as a primary conductor and also as a current producing network; the other is the consumption network and is not in unipolar connection with the current producing network.

In Figure 27 the current producing network I is shown with three balloon collectors 1, 2, 3 and aerial conductors 4, 5, 6; it is short circuited through condenser 19 and inductance 9. The oscillation forming circuit consists in this diagram of spark gap *f*, inductance 10, and condenser 11; the earth wire E, is connected to earth over electromagnet S^1. F is the safety spark gap which is also connected to earth through a second electromagnet S at E. On connecting up the condenser circuit 11 this is charged over the spark gap *f* whereby an oscillatory discharge is formed. This discharging current acts through inductance 10 on the inductively coupled secondary 9, whereby in the producing network a modification of the potential of the condenser 19 is produced. The consequence of this is that oscillations arise in the coil shaped producer network. These oscillations induce a current in the secondary circuit II, which has a smaller number of windings and a less resistance, the voltage of which, according to the proportion of the number of windings and of the ohmic resistance, is considerably lower whilst the current strength is greater.

In order to convert the current thus obtained into current of an undamped character, and to tune its wave lengths, a sufficiently large regulatable capacity 20 is inserted between the ends 12 and 13 of the secondary conductor II. Here also current may be taken without an earth conductor, but it is advisable to insert a safety spark gap E^2 and to connect this with the earth over an electromagnet S^2.

The producer network may be connected with the working network II over an inductionless condenser 21 or over an induction condenser 22, 23. In this case the secondary conductor is unipolarly connected with the energy conductor.

In Figure 28 the connecting conductor between the separate accumulator balloons is carried out according to the autotransformer principle. The collecting coil connects four aerial balloons 1, 2, 3, 4, the windings of which are not made side by side but one above the other. In Figure 28 the collector coil I is shown with a thin line, the metallically connected prolongation coils II with a thick line. Between the ends I^1 and II^1 of the energy network I a regulating capacity 19 is inserted. The wire I^2 is connected with the output wire and with the spark gap F.

As transformer of the atmospheric electricity an arrangement is employed which consists in using rotary pairs of condensers in which the one stator surface B is connected with the main current, whilst the other A is connected with the earth pole. Between these pairs of short circuited condensers are caused to rotate from which the converted current can be taken by means of two collector rings and brushes, in the form of an alternating current, the frequency of which is dependent on the number of balloons and the revolutions of the rotor. As the alternating current formed in the rotor can act, in this improved method of connection described in this invention, through coils 10 on the inductance 9, an increase or diminution of the feed current in I can be obtained according to the direction of the current by back induction. Current oscillations of uniform rhythm thereby result in the coil shaped windings of the producer net-work.

As the ends of this conductor are short cir-

cuited through the regulatable condenser 19 these rhythms produce short circuited undamped oscillations in the energy conductor, the periodicity and wave lengths of which oscillations can be adjusted according to desire by altering the capacity 19 to a given wave length and therewith also to a given frequency. These currents may also be employed in this form directly as working current through the conductors II^1 and III. By inserting the condenser 20 a connection between these conductors may also be made, whereby harmonic oscillations of desired wave length are formed. By this means quite new effects as regards current distribution are obtained. The withdrawal of current can even take place without direct wire connection if, at a suitable point in the interior of the producing network (quite immaterially whether this has a diameter of 1 or 100 km.) a coil tuned to these wave lengths and of the desired capacity is firmly or movably mounted in the aerial conductor in such a way that its axial direction is in parallel with that of the collector coil. In this case a current is induced in the producing network, the size of which is dependent on the total capacity and resistance and also on the periodicity employed. A possibility is thereby afforded in future, of taking energy from the producer network by wireless means. As thereby in addition to atmospheric electricity also magnetic earth currents and the energy from the higher atmosphere (at least partially) may be simultaneously obtained, this last system for collecting the atmospheric energy is of particular importance for the future.

Of course everywhere instead of spark gaps suitable grid vacuum tubes may be employed as producers for undamped oscillations. The separate coils of the producer net-work with large diameters may be connected with one another through separate conductors all in parallel or all in series or in groups in series. By regulating the number of oscillations and also the extent of the voltage more or less large collector coils of this kind may be employed. The coils may also be divided spirally over the entire section. The coils may be carried out in annular form or also in triangular, quadrangular, hexagonal or octagonal form.

Of course wires may be carried from a suitable place to the centre or also laterally which serve the current waves as guides. This is necessary when the currents have to be conducted over mountains and valleys and so forth. In all these cases the current must be converted into a current of suitable periodicity.

As already hereinbefore mentioned separate collecting balloons may be directly metallically interconnected at equidistant stations distributed over the entire country or may be connected by interpolation of suitable condenser batteries by means of high voltage conductors. The static electricity is converted through a spark gap into dynamic energy of a high number of oscillations, and could then in such form, with a suitable arrangement of the connections, observing various measures of precaution, be employed as source of energy after separate or special regulation.

According to this invention in order to increase the collecting effect of the balloon in the aerial collector conductor or in the earth wire, radiating collectors are employed. These consist either of incandescent metal or oxide electrodes in the form of vacuum grid tubes, or electric arcs (mercury and the like electrodes) Nernst lamps, or finally flames of various kinds may be simply connected with the respective conductor.

It is well known that energy can be drawn off from a cathode consisting of an incandescent body opposite an anode charged with positive electricity (vacuum grid tube). Hitherto however, a cathode was always first directly placed opposite an anode, and secondly the system always consisted of a closed circuit.

Now if we dispense with the ordinary ideas in forming light or flame arcs in which a cathode must always stand directly opposite an anode, and if we place an incandescent cathode opposite an anode charged to a high potential or another body freely floating in the air, or regard the incandescent cathode only as a source of unipolar discharge (which represent group and point discharges in electro-static machines similar to unipolar discharges), it may be ascertained that incandescent cathodes and less perfectly all incandescent radiators, flames and the like admit of relatively large current densities and allow large quantities of electric energy to radiate into the open space in the form of electron streams as transmitters.

The object of this invention is as described below, if such incandescent oxide electrodes or other incandescent radiators or flames are not freely suspended in space but connected metallically with the earth so that they can be charged with negative terrestrial electricity, these radiators possess the property of absorbing the free positive electrical charges contained in the air space surrounding them (that is to say of collecting them and conducting them to earth). They can therefore, serve as collectors and have, in comparison to the action of the spikes, or points, a very large radius of action R; the effective capacity of these collectors is much greater than the geometrical capacity (R_0-) calculated in an electro-static sense.

Now as our earth is surrounded as is well

known, with an electro-static field and the difference of potential

$$\frac{\delta V}{\delta h}$$

of the earth field according to the latest investigations, is in summer about 60 to 100 volts and in winter 300 to 500 volts per metre of difference in height (δh), a simple calculation gives the result that when such a radiation collector or flame collector is arranged for example on the ground, and a second one is mounted vertically over it at a distance of 2,000 metres and both are connected by a conducting cable, there is a difference of potential in summer of about 2,000,000 volts and in winter even of 6,000,000 volts and more.

According to Stefan Boltzmann's law of radiation, the quantity of energy which an incandescent surface (temperature T) of 1 sq. cm. radiates in a unit of time into the open air (temperature T_o) is expressed by the following formula:

$$S=\sigma\ (T^4-T^4_0)\ \text{watt/sq. cm.}$$

and the universal radiation constant σ is according to the latest researches of Ferry (Annales de Chimie et de physique 17 page 267 (1909)) equal to 6.30×10^{-12} watt/sq. cm.

Now if an incandescent surface of 1 sq. cm. shows, as compared with the surrounding space a periodic fall of potential δV it radiates (independent of the current direction, that is to say of the sign) in accordance with the above formula, for example at a temperature of 3725° C. an energy of 1.6 kw. per sq. cm. per second. As for the radiation the same value can be calculated for the collection of energy, but reversed. Now as carbon electrodes at the temperature of the electric arc support on the current basis a current density up to from 60 to 65 amperes per sq. cm. no difficulties will result in this direction in employing radiating collectors as accumulators.

If the earth be regarded as a cosmically insulated condenser in the sense of geometrical electro-statics x there results from the geometric (compare Ewald Rasch, "das elektrische Bogenlicht" (The electric arc light) page 169) capacity of the earth according to Chwolson:

For negative charging 1.3×10^{6} Coulomb

For negative potential $V=10\times10^{8}$ volts.

From this there results however, $EJT\cong24.7\times10^{24}$ watt/Sec. Now if it is desired to make a theoretic short circuit through an earthed flame collector this would represent an electric total work of about 70,500 10^{18} kilowatt years. As the earth must be regarded as a rotating mechanism which is thermo-dynamically, electromagnetically, and also kinematically coupled with the sun and stars system by cosmic radiations and gravitation a diminution of the electric energy of the earth field is not to be feared. The energies which the incandescent collectors would withdraw from the earth field can only cause by the withdrawal of motor work a lowering of the earth temperature (temperature $T_E=300$) and reduce this to that of the world space ($T=O$) by using the entire energy. This is however not the case as the earth does not represent a cosmically entirely insulated system. On the contrary there is conveyed to the same according to the recent value corrected by Ferry for the solar constants through the radiation from the sun an energy of 18,500 $\times10^{10}$ kw. Accordingly any lowering of the earth temperature (T_E) without a simultaneous lowering of the sun's temperature (T_S) would contradict Stefan Boltzmann's law of radiation.

$$S=\sigma\ (T_s^4-T_E^4).$$

From this it must be concluded that if the earth temperature (T_E) sinks, the total radiation S absorbed by the earth increases, and further also that the secular speed of cooling of the earth is directly dependent on that of the sun and the other radiators cosmically coupled with the sun and is connected most closely with these.

The incandescent radiation collectors may, according to this invention, be employed for collecting atmospheric electricity if they (1) are charged with the negative earth electricity (that is to say when they are directly connected by means of a metallic conductor with the earth) and (2) if large capacities (metal surfaces) charged with electricity are mounted opposite them as positive poles in the air. This is regarded as the main feature of the present invention as without these inventive ideas it would not be possible to collect with an incandescent collector, sufficiently large quantities of the electrical charges contained in the atmosphere as technology requires; the radius of action of the flame collectors would also be too small, especially if it be considered that the very small surface density (energy density) (σ about$=2\times7\ .\ 10^{8}$ St. E. per sq. cm.) does not allow of large quantities of charge being absorbed from the atmosphere.

x) Calculated according to Poisson's calculation;

$\Delta V=-4\pi\delta$; as here the alteration of the potential or potential gradients only takes place in the direction of the normal, this calculation assumes the simple form

$$\delta=\frac{1}{4\pi}\times\frac{\delta^2V}{\delta n^2}$$

It has indeed already been proposed to employ flame collectors for collecting atmospheric electricity and it is known that

their collecting effect is substantially greater opposite the points. It is however, not known that the quantities of current which could hitherto be obtained are too small for technical purposes. According to my experiments the reason for this is to be found in the too small capacities of the collector conductor poles. If such flame or radiating collectors have no or only small positive surfaces, their radius of action for large technical purposes is too small. If the incandescent collectors be constantly kept in movement in the air they may collect more according to the speed of the movement, but this is again not capable of being carried out in practice.

By this invention the collector effect is considerably increased by a body charged with a positive potential and of the best possible capacity being also held floating (without direct earth connection) opposite such an incandescent collector which is held floating in the air at a desired height. If for example, a collecting balloon of sheet metal or of metalized balloon fabric be caused to mount to 300 up to 3,000 metres in the air and as positive pole it is brought opposite such a radiating collector connected by a conductor to earth, quite different results are obtained.

The metallic balloon shell (with a large surface) is charged to a high potential by the atmospheric electricity. This potential is greater the higher the collecting balloon is above the incandescent collector. The positive electricity acts concentratedly on the anode floating in the air as it is attracted through the radiation shock ionization, proceeding from the incandescent cathode. The consequence of this is that the radius of action of the incandescent cathode collector is considerably increased and thereby also the collecting effect of the collecting balloon surface. Further the large capacity of the anode floating in the air plays therefore an important part because it allows of the taking of large charges, and thereby a more uniform current is obtained even when there is a large consumption; this cannot be the case with small surfaces.

In the present case the metallic collecting balloon is a positive anode floating in the air and the end of the earth conductor of this balloon serves as positive pole surface opposite the surface of the radiating incandescent cathode, which in turn is charged with negative earth electricity being conductingly connected to earth.

The process may be carried out by two such contacts (negative incandescent cathode and anode end of a capacity floating in the air) a condenser and an inductive resistance being switched on in parallel, whereby simultaneously undamped oscillations may be formed.

In very large installations it is advisable to connect two such radiating collectors in series. Thus an arc light incandescent cathode may be placed below on the open ground and an incandescent cathode which is heated by special electro-magnetic currents be located high in the air. Of course for this the special vacuum Liebig tubes with or without grids may also be employed. An ordinary arc lamp with oxide electrodes may be introduced on the ground and the positive pole is not directly connected with the collecting balloon, but through the upper incandescent cathode or over a condenser. The method of connecting the incandescent cathode floating in the air may be seen in Figs. 29–33.

B is the air balloon, K a Cardan ring (connection with the hawser) C the balloon, L a good conducting cable, P a positive pole, N negative incandescent cathode, and E earth conductor.

Fig. 29 represents the simplest form of construction. If electric oscillations are produced below on the ground by means of a carbon arc lamp or in other suitable way a considerably greater electric resistance is opposed to that in the direct way by inserting an electrical inductive resistance 9. Consequently between P and N a voltage is formed, and as, over N and P only an inductionless ohmic resistance is present, a spark will spring over so long as the separate induction co-efficients and the like are correctly calculated. The consequence of this is that the oxide electrode (carbon or the like) is rendered incandescent and then shows as incandescent cathode an increased collecting effect. The positive poles must be substantially larger than the negative in order that they may not also become incandescent. As they are further connected with the large balloon area which has a large capacity and is charged at high voltage, an incandescent body which is held floating in the air and a positive pole which can collect large capacities is thereby obtained in the simplest way. The incandescent cathode is first caused to become incandescent by means of separate energy produced on the earth, and then maintained by the energy collected from the atmosphere.

Fig. 30 only shows the difference that instead of a round balloon a cigar shaped one (of metal or metalized fabric) may be employed and also a condenser 5 is inserted between the incandescent cathode and the earth conductor so that a short circuited oscillation circuit over P. N 5 and 9 is obtained. This has the advantage that quite small quantities of electricity cause the cathode to become incandescent and much larger cathode bodies may be rendered incandescent.

In this form of construction both the in-

candescent cathode and also the positive electrode may be enclosed in a vacuum chamber as may be seen in Fig. 32. A cable L is carried well insulated through the cover of a vessel and ends in a condenser disc 5. The cover is arched in order to keep off the rain. The vessel is entirely or partially made of magnetic metal and well insulated inside and outside. Opposite the disc 5 another disc 6 and on this again a metallic positive pole of the vacuum tube *g* with the incandescent cathode (oxide electrode) N is arranged. The negative electrode is on the one hand connected with the earth conductor E, and on the other hand with the inductive resistance 9 which is also connected with the cable L with the positive pole and wound round the vessel in coils. The action is exactly the same as that in Fig. 29 only instead of an open incandescent cathode one enclosed in vacuo is employed. As in such collectors only small bodies can be brought to incandescence in large installations a plurality of such vacuum tubes must be inserted in proximity to one another. According to the previous constructions Figs. 31 and 33 are quite self evident without further explanations.

Figs. 34–37 represent further diagrams of connections over radiating and flame collectors, and in fact, how they are to be arranged on the ground.

Fig. 34 shows an arc light collector with oxide electrodes for direct current and its connection; Fig. 35 a similar one for alternating current, Fig. 36 an incandescent collector with a Nernst lamp and Fig. 37 a similar one with a gas flame.

The positive pole 1 of the radiating collectors is always directly connected to the aerial collecting conductor A. In Fig. 34 this is further connected over the condenser battery 5 with a second positive electrode 3. The direct current dynamo 5 produces current which flows over between the electrodes 3 and 2 as an arc light. On the formation of an arc the negative incandescent electrode 2 absorbs electricity from the positive poles standing opposite it and highly charged with atmospheric electricity and conveys the same to the working circuit. The spark gap 7, inductive resistance 9 and induction coil 10 are like the ones previously described. The protective electromagnet S guards the installation against earth circuiting, the safety spark gap 8 from excess voltage or overcharging.

In Fig. 35 the connection is so far altered that the alternating current dynamo feeds the exciting coil 11 of the induction condenser. 12 is its negative and 13 its positive pole; if the coil 3 on the magnet core of the dynamo is correctly calculated and the periodicity of the alternating current is sufficiently high an arc light can be formed between the two poles 1 and 2. As the cathode 2 is connected with the negatively charged earth, and therefore always acts as a negative pole, a form of rectification of the alternating current produced by the dynamo 3 is obtained, the second half of the period is always suppressed. The working circuit may be carried out in the same way as in Fig. 34; the working spark gap 7 may however be dispensed with, and instead thereof between the points *n* and *m* a condenser 5 and an induction resistance 9 may be inserted from which the current is taken inductively.

Fig. 36 represents a form of construction similar to Fig. 34 only that here instead of an arc lamp a Nernst incandescent body is employed. The Nernst lamp is fed through the battery 3. The working section is connected with the negative pole, the safety spark gap with the + poles. The working spark gap 7 may also be dispensed with and the current for it taken at 12 over the oscillation circuit 5, 11 (shown in dotted lines).

Flame collectors (Fig. 37) may also be employed according to this invention. The wire network 1 is connected with the aerial collector conductor A and the burner with the earth. At the upper end of the latter, long points are provided which project into the flame. The positive electrode is connected with the negative over a condenser 5 and the induction coil 9 with the earth.

The novelty in this invention is firstly, the use of incandescent cathodes opposite positive poles which are connected with large metallic capacities as automatic collecting surfaces. (2) the connection of the incandescent cathodes with the earth whereby, in addition to the electricity conveyed to them from the battery or machine which causes the incandescing, also the negative charge of the earth potential is conveyed, and (3) the connection of the positive and negative poles of the radiating collectors over a condenser circuit alone or with the introduction of a suitable inductive resistance, whereby simultaneously an oscillatory oscillation circuit may be obtained. The collecting effect is by these methods quite considerably increased.

I declare that what I claim is:—

1. An electrical energy generating system, comprising a conducting surface for static charges, means to support same at a distance above the earth, a conductor leading to the earth level, a spark gap associated with said conductor to convert electrostatic charges into electromagnetic high frequency oscillations means to supply said electromagnetic energy to a net work, and a spark gap of greatly increased relative resistance in parallel therewith.

2. An electrical energy generating system

comprising a conductor, means to support same above the earth level, an inductance therein, a spark gap associated with said conductor, a second spark gap of much higher relative resistance in parallel therewith and an energy receiving circuit coupled with the spark gap of lesser resistance.

3. An electrical energy generating system comprising a collecting surface, means to support same above the earth level, a conductor connecting said collecting surface with the earth level, a choke in said conductor, an electromagnetic resistance converting electrostatic energy to electromagnetic energy, a safety higher resistance in parallel therewith and a net work coupled with the conversion resistance of lesser value.

4. An electrical energy generating system comprising electric conductors spaced above the earth to form electromagnetic oscillating circuits, conductors connecting to earth level, electrostatic to electromagnetic energy conversion means therein, a safety high electrostatic resistance in parallel therewith and means to alter the electromagnetic characteristic of the circuits.

5. An electrical energy generating system comprising in combination a static collecting surface arranged above the earth, conductors connecting to earth level, a pair of spark gaps in parallel of different electrostatic resistance, a utilization net work shunted across the spark gap of lesser resistance and an electromagnetic choke in said conductors.

6. An electrical energy generating system comprising an open circuit energy collecting aerial, a pair of sparking gaps in parallel of widely different resistance, connected thereto and a closed electric oscillation circuit in shunt across the gap of lesser resistance

7. An electrical energy generating system comprising an open circuit energy collecting aerial, a pair of sparking gaps in parallel of widely different resistance connected thereto, a closed electric oscillation circuit in shunt across the gap of lesser resistance, a plurality of electrostatic collecting surfaces, means to connect said collecting surfaces in parallel in groups and means to connect said groups symmetrically with said aerial.

In witness whereof, I have hereunto signed my name this 30 day of Dec., 1920, in the presence of two subscribing witnesses.

HERMANN PLAUSON.

Witnesses:
H. F. Armstrong,
W. H. Beeston.

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 1

Fig. 1.

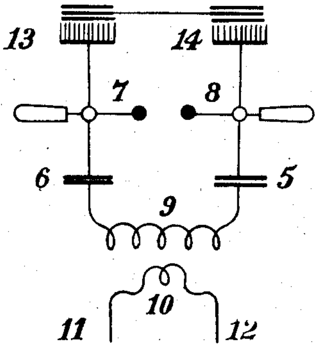
13
14
7
8
6
5
9
10
11
12

Fig. 2.

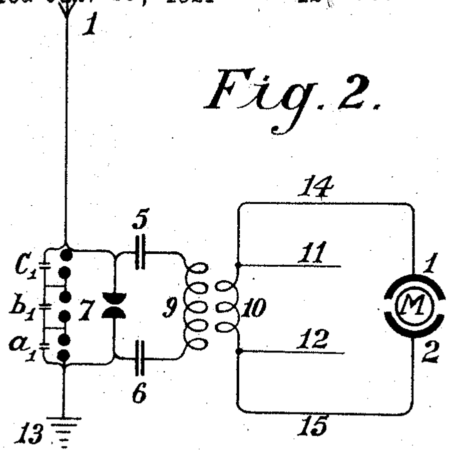
1
C1
b1
a1
7
5
6
9
10
14
11
12
15
1
2
M
13

Fig. 3.

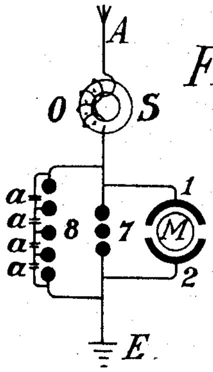
A
0
S
a
a
a
a
8
7
1
M
2
E.

Fig. 4.

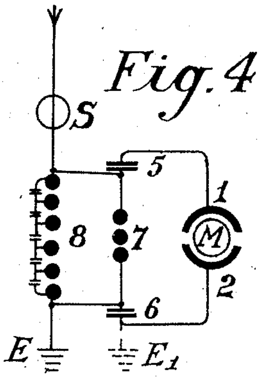
S
5
8
7
1
M
2
6
E
E1

Fig. 5.

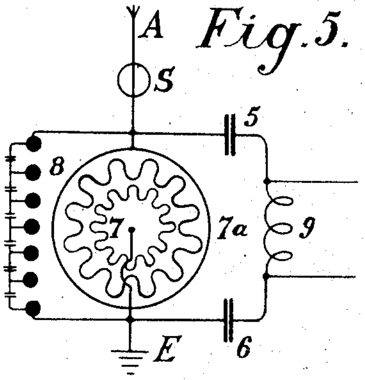
A
S
5
8
7
7a
9
6
E

Fig 6

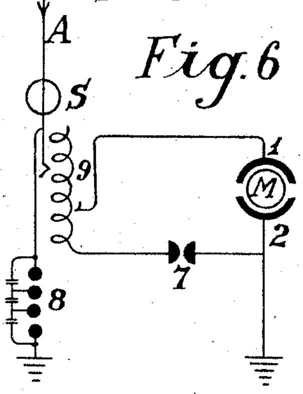

Inventor

Hermann Plauson

by Knight Bros

attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 2

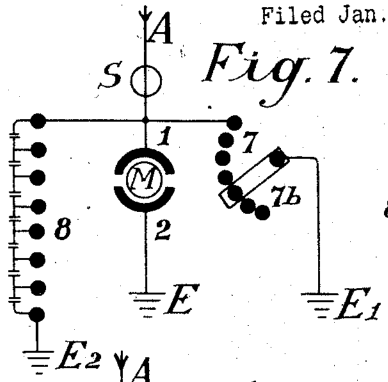
Fig. 7.
A
S
1
M
2
7
7b
8
E
E1
E2

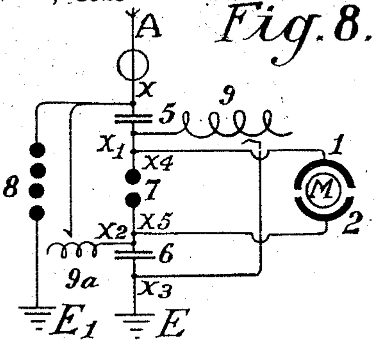
Fig. 8.
A
x
5
9
1
x1
x4
7
8
x5
M
2
x2
6
9a
x3
E1
E

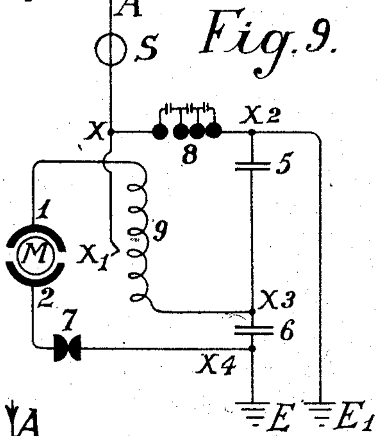
Fig. 9.
A
S
X
8
X2
5
1
9
M
X1
2
X3
7
6
X4
E
E1

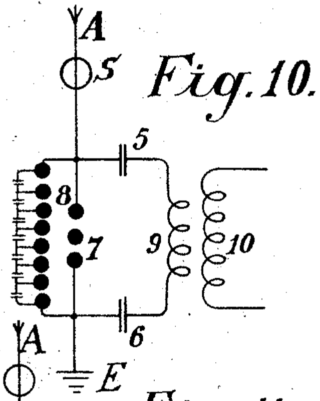
Fig. 10.
A
S
5
8
7
9
10
6
E

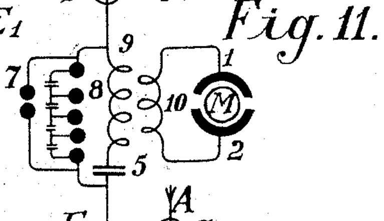
Fig. 11.
A
S
9
7
8
10
1
M
2
5
E

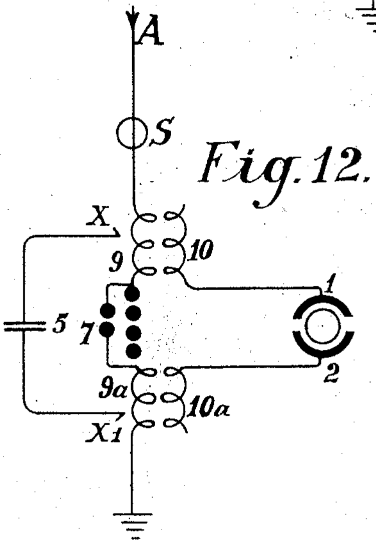
Fig. 12.
A
S
X
9
10
1
5
7
2
9a
10a
X1

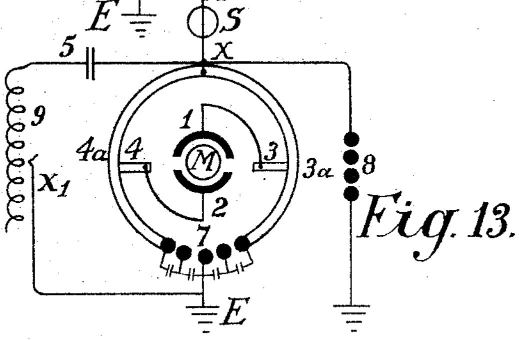
Fig. 13.
A
S
X
5
9
1
4a
4
M
3
3a
8
X1
2
7
E

Inventor

Normann Plauson

by Knight Bros

attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 3

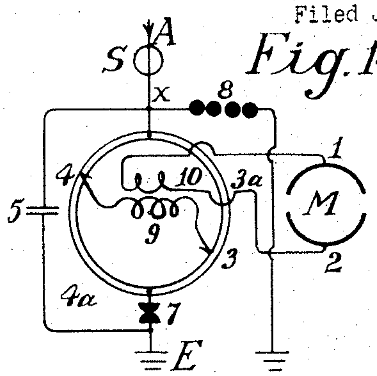
Fig. 14.
S
A
x
8
1
M
2
4
10
3a
5
9
3
4a
7
E

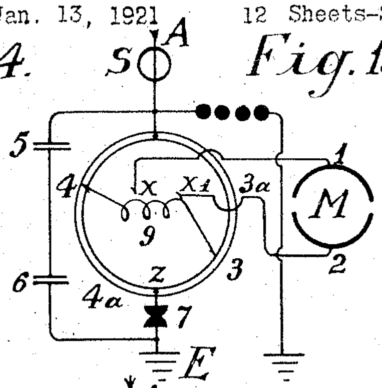
Fig. 15.
S
A
5
1
M
2
4
x
x1
3a
9
3
6
z
4a
7
E

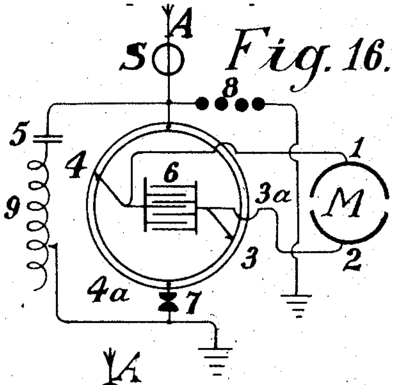
Fig. 16.
S
A
8
5
1
M
2
4
6
3a
9
3
4a
7

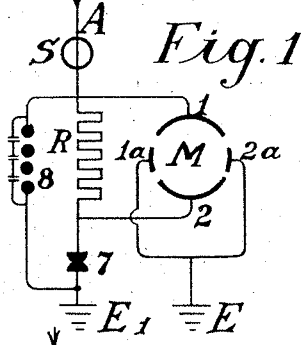
Fig. 17.
S
A
1
R
1a
M
2a
8
2
7
E1
E

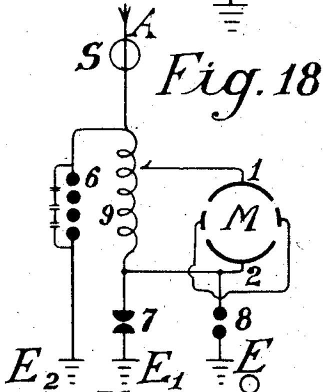
Fig. 18.
S
A
6
9
1
M
2
7
8
E2
E1
E

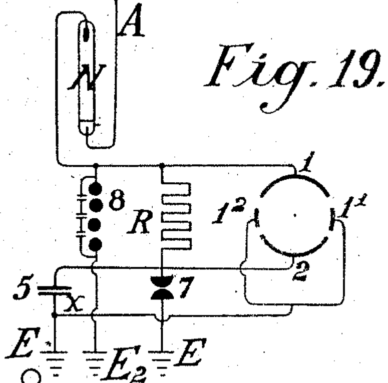
Fig. 19.
A
N
1
8
R
1²
1¹
2
5
x
7
E
E2
E

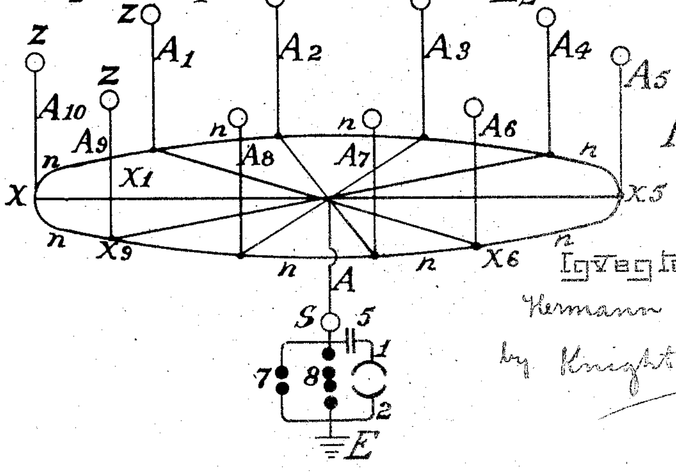
Fig. 20.
z
z
z
A1
A2
A3
A4
A5
A10
A9
A8
A7
A6
n
x
x1
x5
x9
x6
A
S
5
1
7
8
2
E

Inventor
Hermann Plauson
by Knight Bros
Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 4

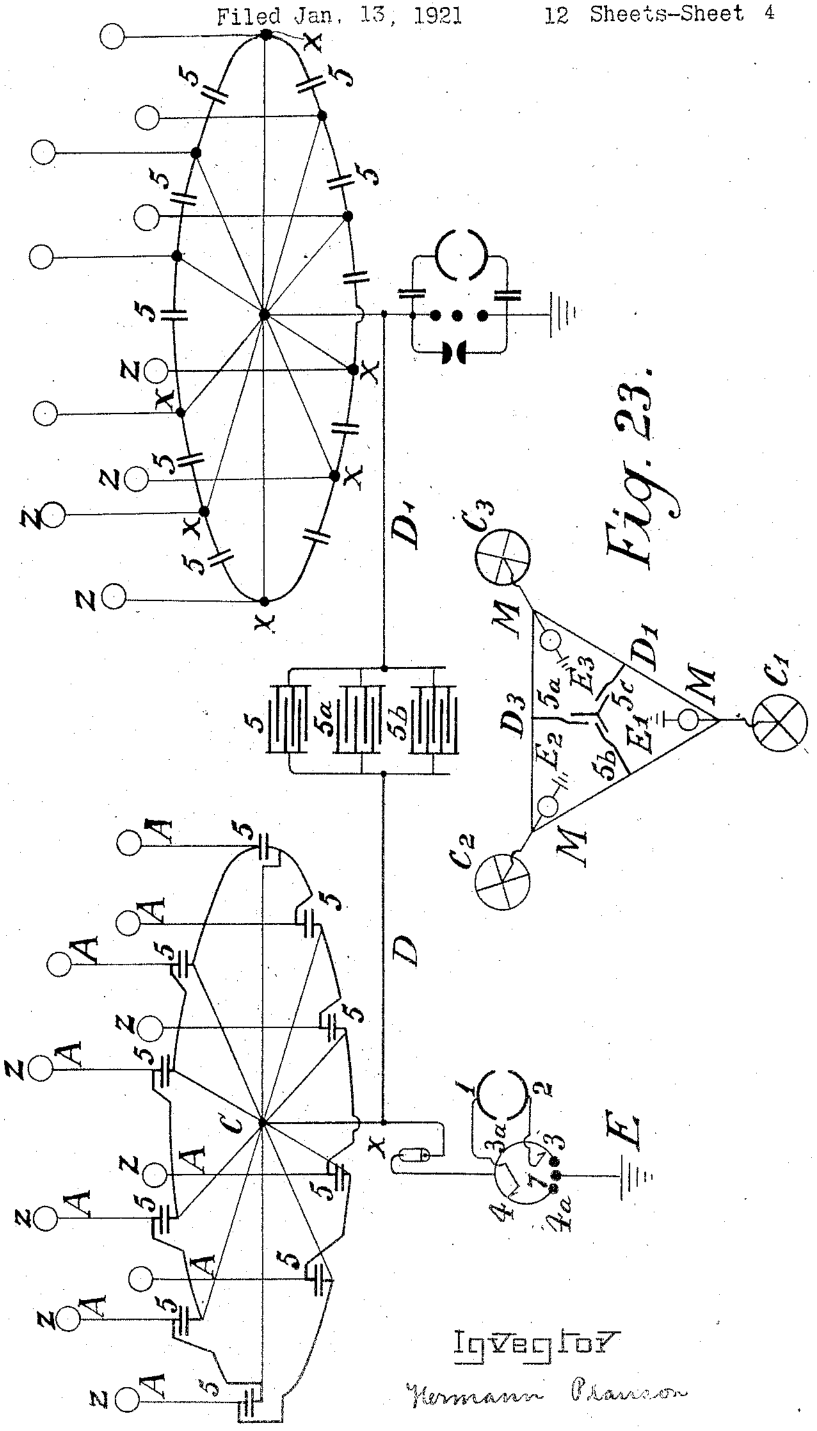
5
x
Z
A
C
D
D1
5a
5b
5c
1
2
3
3a
4
4a
7
E
M
C1
C2
C3
D3
E1
E2
E3

Fig. 22.

Fig. 21.

Fig. 23.

Inventor
Hermann Plauson
by Knight Bros
Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 5

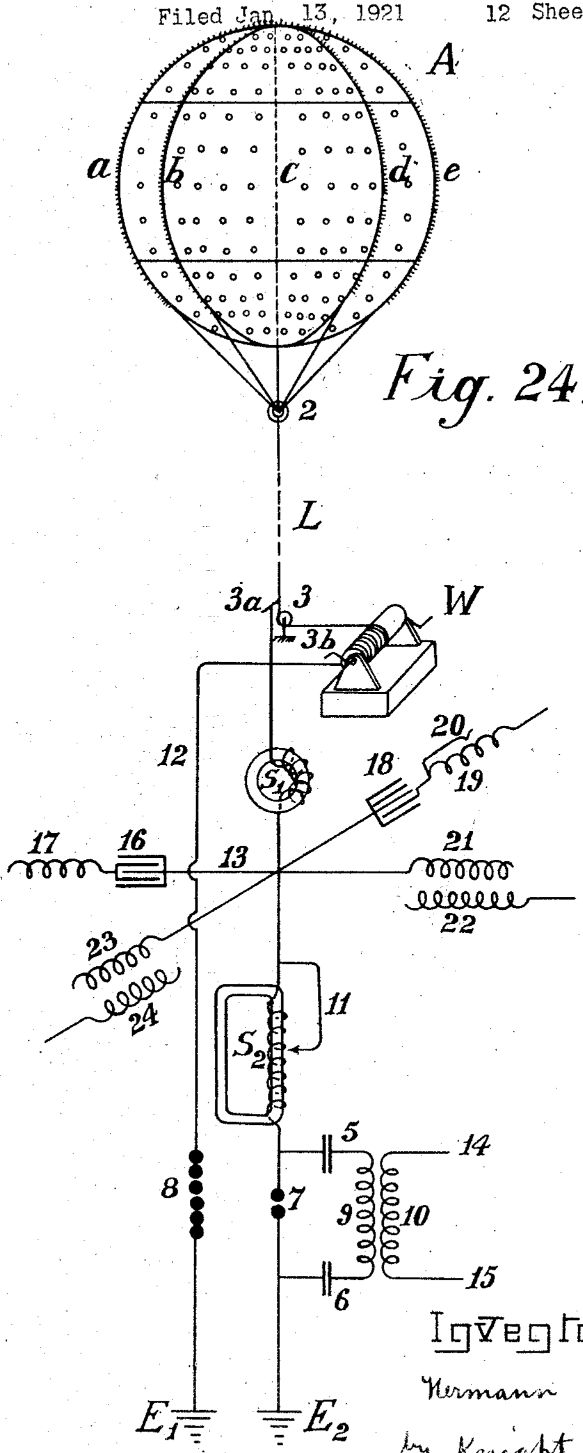
A
a
b
c
d
e
2
L
3a
3
3b
W
12
S_1
18
20
19
17
16
13
21
22
23
24
11
S_2
5
14
7
8
9
10
15
6
E_1
E_2

Fig. 24.

Inventor

Hermann Plauson

by Knight Bros

Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 6

Fig. 25.

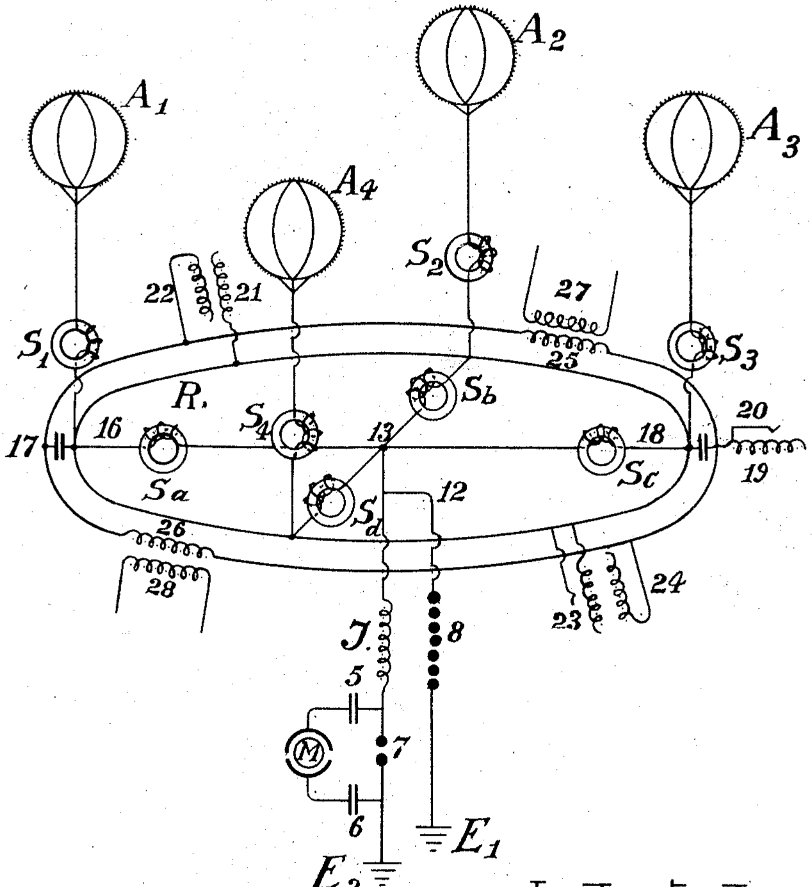
A1
A2
A3
A4
S2
27
22
21
S1
25
S3
R.
Sb
20
17
16
S4
13
18
19
Sa
Sc
12
Sd
26
28
24
23
J.
8
5
M
7
6
E1
E2

Inventor

Hermann Plauson

by

Knight Bros

attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 7

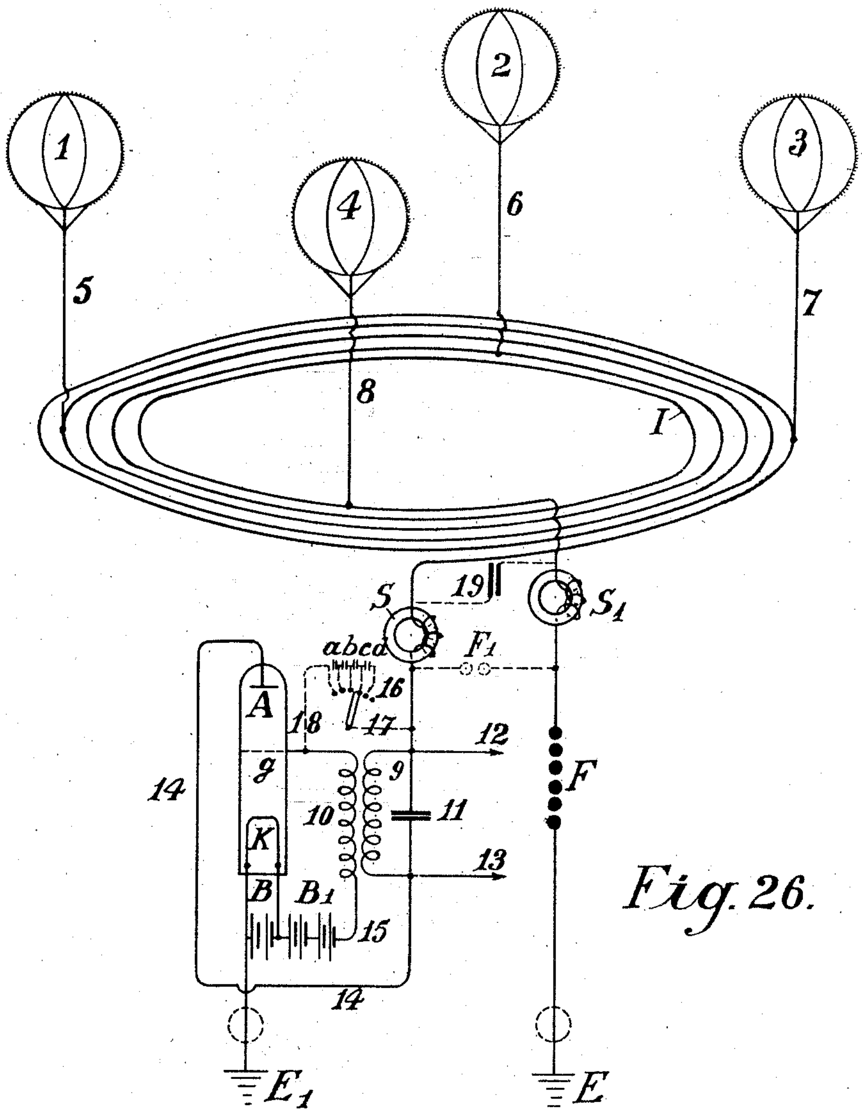
1
2
3
4
5
6
7
8
I
19
S
S1
abcd
F1
16
A
18
17
12
g
9
14
10
11
K
13
B
B1
15
14
F
E1
E

Fig. 26.

Inventor

Hermann Plauson

by Knight Bros

Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921. 12 Sheets-Sheet 8

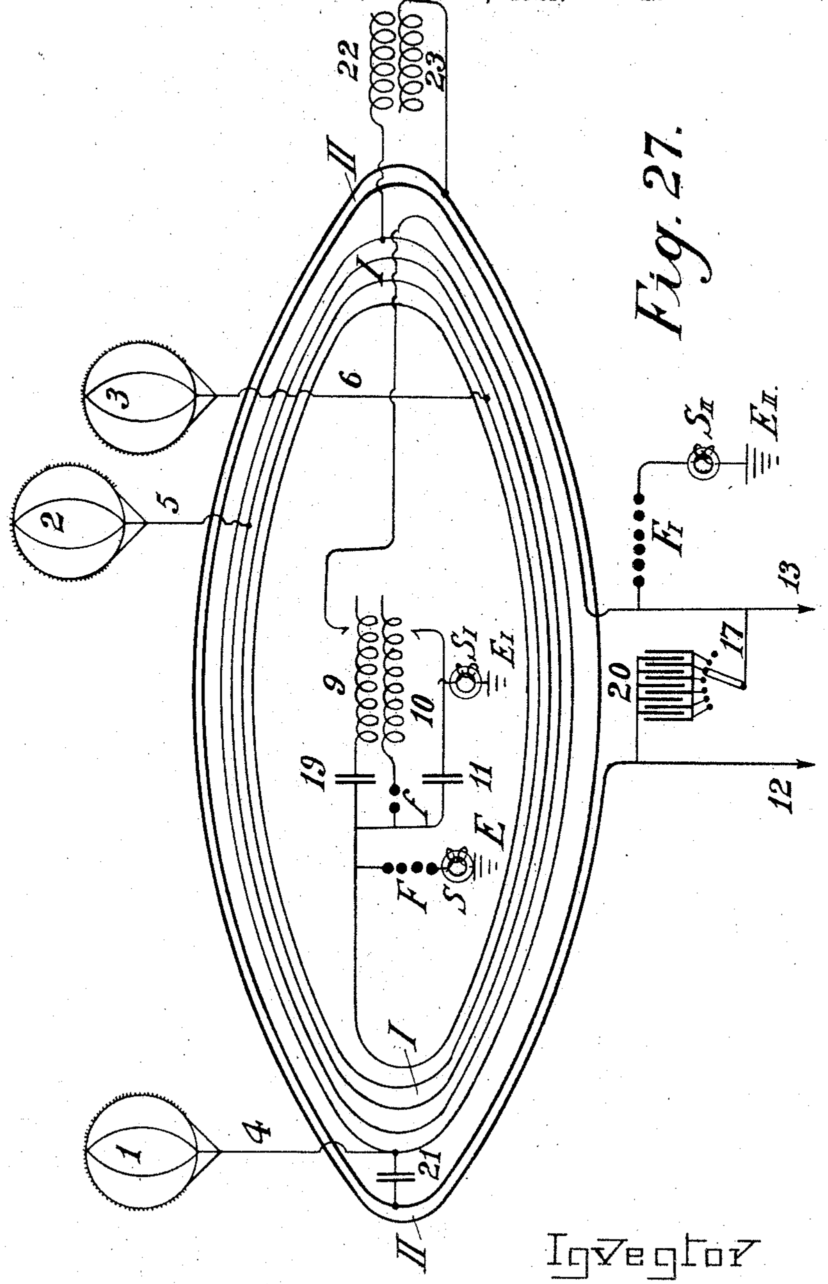
22
23
II
I
6
3
2
5
S_{II}
E_{II}
F_I
13
17
20
9
10
S_I
E_I
19
11
F
E
S
12
I
4
1
21
II

Fig. 27.

Inventor
Hermann Plauson
by Knight Bros
Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 9

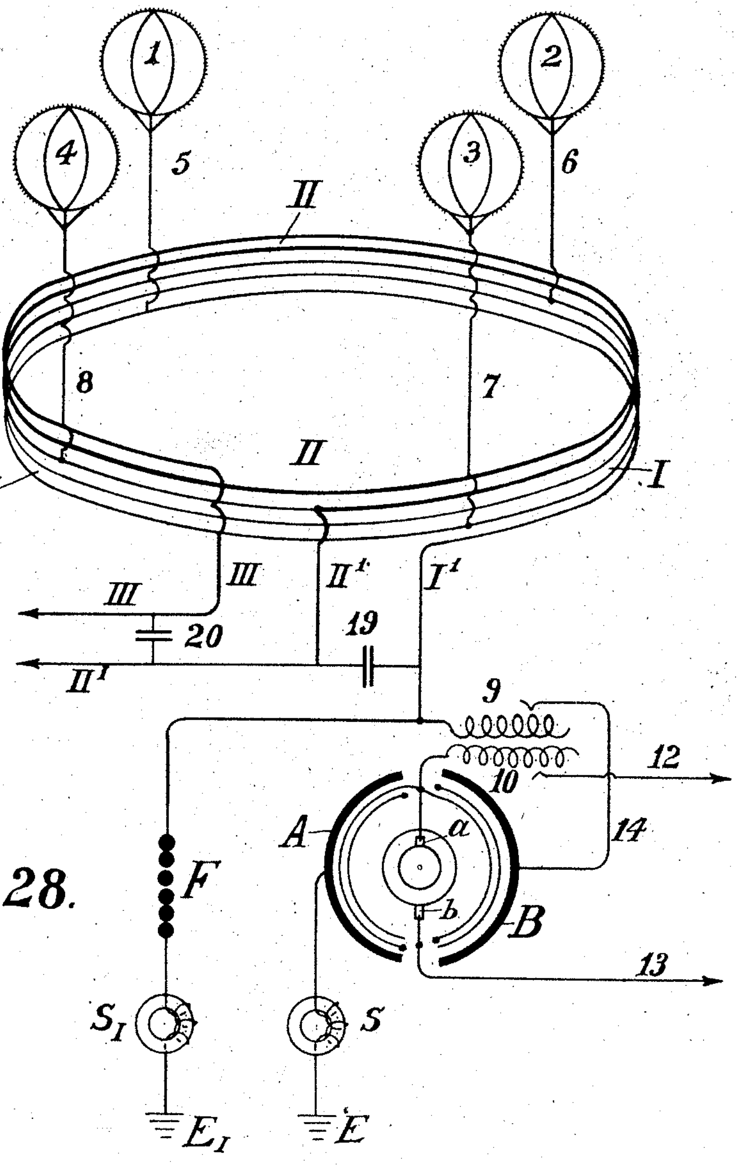
1
2
3
4
5
6
II
8
7
II
I
I
III
II1
I1
III
20
19
II1
9
12
10
A
a
14
B
b
F
13
S1
S
E1
E

Fig. 28.

Inventor

Hermann Plauson

by

Knight Bros

attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921. 12 Sheets-Sheet 10

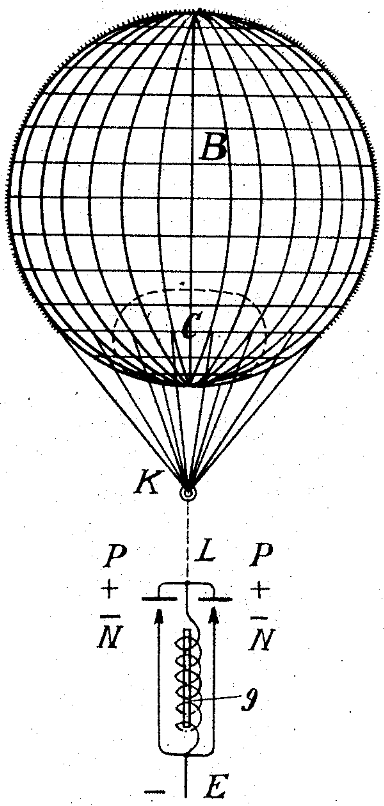
B
C
K
P
L
P
+
+
N
N
9
−
E

Fig. 29.

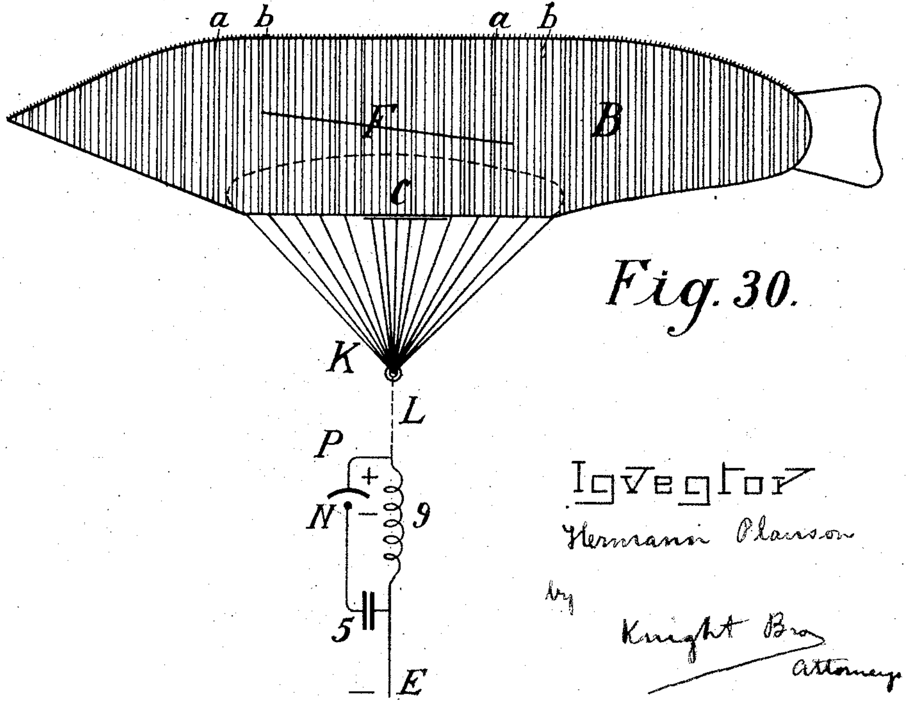
a
b
a
b
F
B
C
K
L
P
+
N
−
9
5
−
E

Fig. 30.

Inventor
Hermann Plauson
by
Knight Bros
Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 11

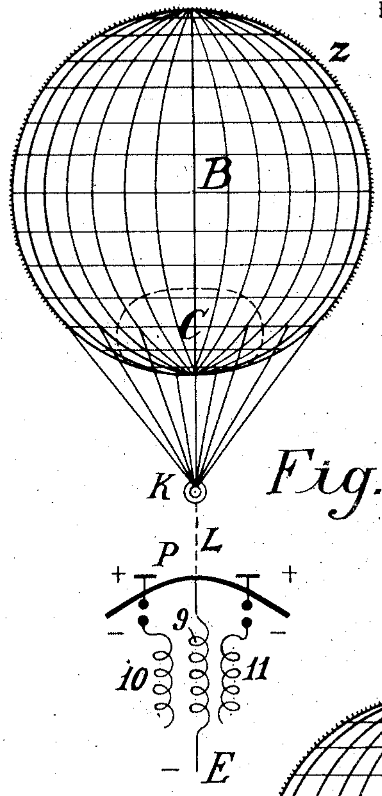
z
B
C
K
P
L
+
+
−
−
9
10
11
−
E

Fig. 31.

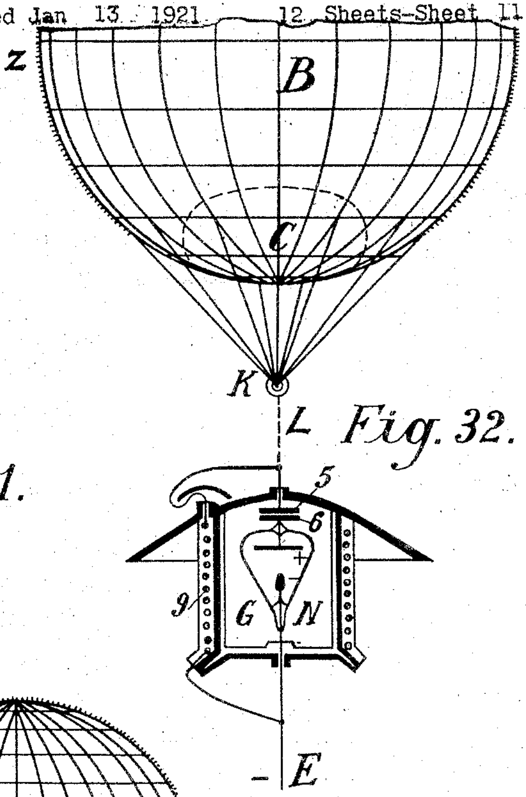
z
B
C
K
L
5
6
+
−
9
G
N
−
E

Fig. 32.

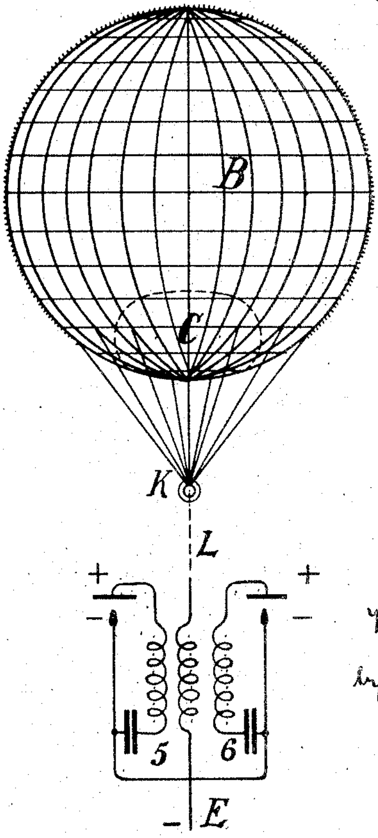

Fig. 33.

Inventor
Hermann Plauson
by Knight Bros
attorneys

June 9, 1925. **1,540,998**

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12 Sheets-Sheet 12

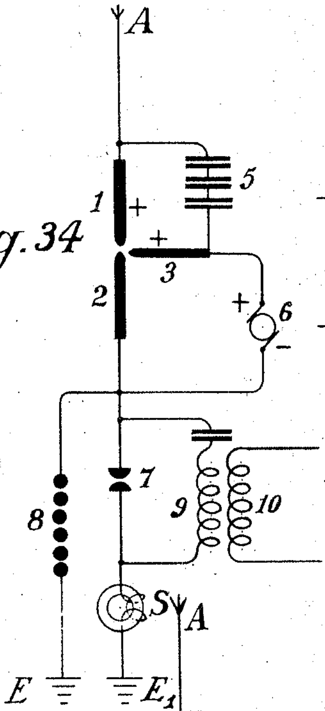
A
1
+
5
+
3
2
+
6
–
7
8
9
10
S
E
E1

Fig. 34

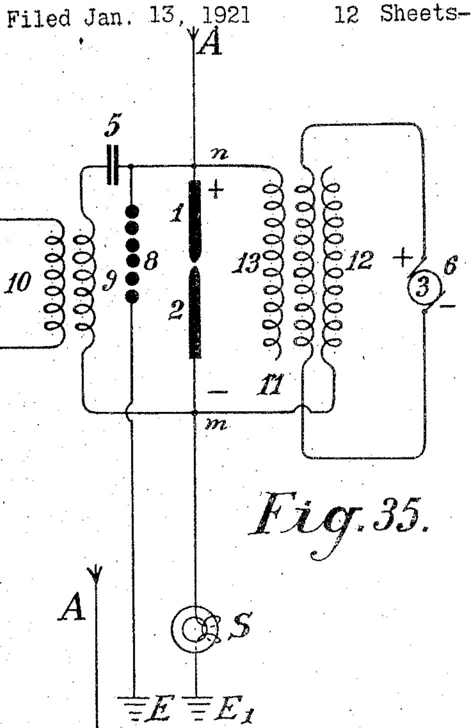
A
5
n
+
1
8
10
9
13
12
+
6
–
2
–
11
m
S
E
E1

Fig. 35.

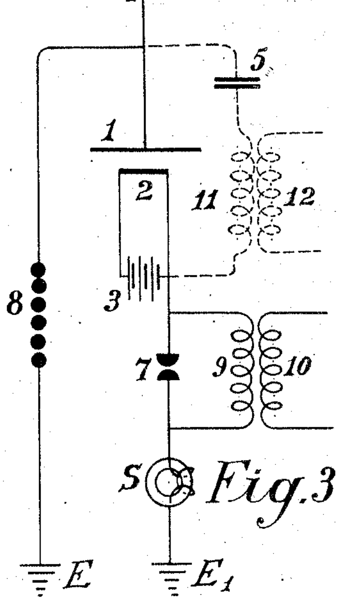
5
1
2
11
12
8
3
7
9
10
S
E
E1

Fig. 36.

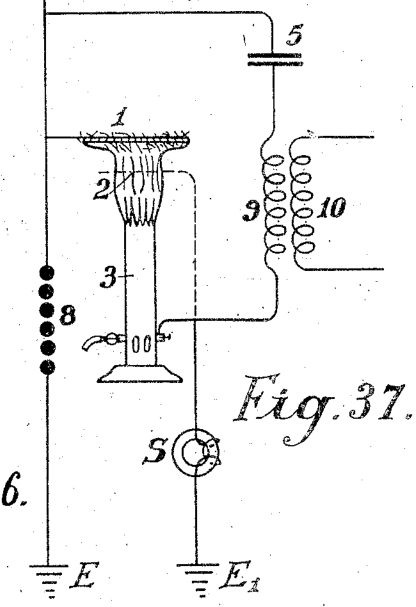
5
1
2
9
10
3
8
S
E
E1

Fig. 37.

Inventor

Hermann Plauson

by Knight Bros

attorneys

Patented June 9, 1925. **1,540,998**

UNITED STATES PATENT OFFICE.

HERMANN PLAUSON, OF HAMBURG, GERMANY.

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY.

Application filed January 13, 1921. Serial No. 437,107.

To all whom it may concern:

Be it known that I, HERMANN PLAUSON, Esthonian subject, residing in Hamburg, Germany, have invented certain new and useful Improvements in the Conversion of Atmospheric Electric Energy, of which the following is a specification.

Methods of obtaining atmospheric electricity by means of metallic nettings set with spikes which are held by means of ordinary or anchored kite balloons made of fabrics and filled with hydrogen, are in theory already known. Atmospheric electricity obtained in this way has been suggested to be used in the form of direct current for the charging of accumulators. This knowledge however is at present only theoretical as the conversion in practice has hitherto been a failure. No means are known of protecting the apparatus from destruction by lightning. The balloons used for collecting the charge must also be made of very large size in order to be able to support the weight of the metallic netting and the heavy cable connections.

Instead of using heavy metallic netting as collectors attached to single air balloons of non-conducting materials which are liable to be torn and are permeable to the gas, it is proposed to use metallic balloon collectors which have the following important advantages—

(*a*) The metallic cases are impenetrable to helium and hydrogen; they also represent large metallic weather-proof collecting surfaces.

(*b*) Radio active means and the like may be easily applied internally or externally; whereby the ionization is considerably increased and therewith also the quantity of atmospheric electricity capable of being collected.

(*c*) Such balloon collectors of light metal do not require to be of large size as they have to carry only their own moderate weight, and that of the conducting cable or wire.

(*d*) The entire system therefore offers little surface for the action of storm and wind and is resistant and stable.

(*e*) Each balloon can be easily raised and lowered by means of a winch so that all repairs, recharging and the like can be carried out without danger during the operation.

It is further proposed to use a collecting aerial network of several separate collectors spread out in the air above the earth, which collectors are interconnected by electrical conductors.

According to this invention charges of atmospheric electricity are not directly converted into mechanical energy, and this forms the main difference from previous inventions, but the static electricity which runs to earth through aerial conductors in the form of direct current of very high voltage and low current strength is converted into electro-dynamic energy in the form of high frequency vibrations. Many advantages are thereby obtained and all disadvantages avoided.

The very high voltage of static electricity of a low current strength can be converted by this invention to voltages more suitable for technical purposes and of greater current strength. By the use of closed oscillatory circuits it is possible to obtain electromagnetic waves of various amplitude and thereby to increase the degree of resonance of such current. Such resonance allows various values of inductance to be chosen whereby again the governing of the starting and stopping of machines driven thereby by simply tuning the resonance between coils of the machine and the transformer circuit forming the resonance can easily be obtained. Further, such currents have the property of being directly available for various uses, even without employing them for driving motors, of which there may be particularly mentioned, lighting, production of heat and use in electro-chemistry.

Further, with such currents a series of apparatus may be fed without direct current supply through conductors and also the electro-magnetic high frequency currents may be converted by means of special motors adapted for electro-magnetic oscillations into mechanical energy, or finally converted by special machines into alternating current of low frequency or even into direct current of high potential.

The invention is more particularly described with reference to the accompanying diagrams in which:—

Figure 1 is an explanatory figure.

Figure 2 is a diagrammatic view of the simplest form.

Figure 3 shows a method of converting atmospheric electrical energy for use with motors.

Figure 4 is a diagram showing the use of protective means.

Figure 5 is a diagram of an arrangement for converting large current strengths.

Figure 6 is a diagram of an arrangement including controlling means.

Figure 7 shows means whereby the spark gap length can be adjusted.

Figure 8 shows a unipolar connection for the motor.

Figure 9 shows a weak coupled system suitable for use with small power motors.

Figures 10, 11 and 12 show modified arrangements.

Figure 13 shows a form of inductive coupling for the motor circuit.

Figure 14 is a modified form of Figure 13 with inductive coupling.

Figure 15 is an arrangement with non-inductive motor.

Figure 16 is an arrangement with coupling by condenser.

Figures 17, 18 and 19 are diagrams of further modifications.

Figure 20 shows a simple form in which the aerial network is combined with special collectors.

Figure 21 shows diagrammatically an arrangement suitable for collecting large quantities of energy.

Figure 22 is a modified arrangement having two rings of collectors.

Figure 23 shows the connections for three rings of collectors.

Figure 24 shows a collecting balloon and diagram of its connection of condenser batteries.

Figures 25 and 26 show modified collector balloon arrangements.

Figure 27 shows a second method of connecting conductor for the balloon aerials.

Figure 28 shows an auto-transformer method of connection.

Figure 29 shows the simplest form of construction with incandescent cathode.

Figure 30 shows a form with cigar shaped balloon.

Figure 31 is a modified arrangement.

Figure 32 shows a form with cathode and electrode enclosed in a vacuum chamber.

Figure 33 is a modified form of Figure 32.

Figure 34 shows an arc light collector.

Figure 35 shows such an arrangement for alternating current.

Figure 36 shows an incandescent collector with Nernst lamp.

Figure 37 shows a form with a gas flame.

Figure 1 illustrates a simple diagram for converting static electricity into dynamic energy of a high number of oscillations. For the sake of clearness in the drawings an influence machine is assumed to be employed and not an aerial antenna. 13 and 14 are combs for collecting the static electricity of the influence machine. 7 and 8 are spark discharging electrodes, 6 and 5 condensers, 9 an inductive primary coil, 10 secondary coil, 11 and 12 ends of conductors of the secondary coil 10. When the disc of the static influence machine is rotated by mechanical means, the combs collect the electric charges one the positive and the other the negative, and charge the condensers 5 and 6 until such a high potential is formed across the spark gap 7—8, that the spark gap is jumped. As the spark gap 7—8 forms a closed circuit with condensers 6 and 5, and inductive resistance 9, as is well known, waves of high frequency electromagnetic oscillations will pass in this circuit.

The high frequency of the oscillations produced in the primary circuit induces waves of the same periodicity in the secondary circuit. Thus in the primary circuit electromagnetic oscillations are formed by the passage of the spark over the spark gap and these waves are maintained by fresh charges of static electricity.

By suitably selecting the ratio between the number of the coils in the primary and secondary circuits with regard to a correct application of the co-efficients of resonance (capacity, inductance, and resistance) the high voltage of the primary circuit may be suitably converted into low voltage and high current strength.

When the oscillatory discharges in the primary circuit becomes weaker or entirely cease, the condensers are charged again by the static electricity until the accumulated charge again breaks down the spark gap. All this is repeated as long as electricity is produced by the static machine by employing mechanical energy.

An elementary form of the invention is shown in Figure 2 in which two spark gaps in parallel are used one of which may be termed the working gap 7 in Figure 2, whilst the second serves as a safety device for excess voltage and consists of a larger number of spark gaps than the working section, which gaps are arranged in series and are bridged by very small capacities as is illustrated in a_1, b_1, c_1, Figure 2 which allow of uniform sparking in the safety section.

In Figure 2 A is the aerial antenna for collecting charges of atmospheric electricity, 13 is the earth connection of the second part of the spark gap, 5 and 6 are condensers, 9 a primary coil. Now when through the aerial A the positive atmospheric electricity seeks to combine with the negative charge to earth, this is prevented by (the air gap between) the spark gaps. The resistance of the spark gap 7 is, as shown in the drawings, lower than that of the other safety section which consists of three spark gaps connected in

series, and consequently a three times greater air resistance is offered by the latter.

So long therefore, as the resistance of the spark gap 7 is not overloaded, so that the other spark gaps have an equal resistance with it the discharges take place only over spark gap 7. Should however the voltage be increased by any influences so that it might be dangerous for charging the condensers 5 and 6 or for the coil insulation 9 and 10 in consequence of break down, by a correct regulation of this spark gap the second spark gap can discharge free from inductive effects direct to earth without endangering the machine.

Without this second spark gap, arranged in parallel having a higher resistance than the working spark gap it is impossible to collect and render available large quantities of electrical energy.

The action of this closed oscillation circuit consisting of spark gap 7, two condensers 5 and 6, primary coil 9, and also secondary coil 10 is exactly the same as the one described in Figure 1 with the arrangement of the static induction machine with the only difference that here the second spark gap is provided. The electromagnetic high frequency alternating current obtained can be tapped off from the conductors 11 and 12 for lighting and heating purposes. Special kinds of motors adapted for working with these peculiar electrical charges may be connected at 14 and 15 which can work with static electricity charges or with high frequency oscillations.

In addition to the use of spark gaps in parallel a second measure of security is also necessary for taking off the current. This precaution consists according to this invention, in the introduction of and method of connecting certain protective electromagnets or choking coils in the aerial circuit as shown by S in Figure 3.

A single electromagnet only having a core of the thinnest possible separate laminations is connected with the aerial.

In the case of high voltages in the aerial network or at places where there are frequent thunder storms, several such magnets may however be connected in series.

In the case of large units or plants several electromagnets can be employed in parallel or in series parallel.

The windings of these electromagnets may be simply connected in series with the aerials. In this case the winding preferably consists of several thin parallel wires, which make up together, the necessary section.

The winding may be made of primary and secondary windings in the form of a transformer. The primary winding will be then connected in series with the aerial network, and the secondary winding more or less short-circuited over a regulating resistance or an induction coil. In the latter case it is possible to regulate to a certain extent the effect of the choking coils. In the further description of the connecting and constructional diagrams the aerial electromagnet choke coil is indicated by a simple ring S.

Figure 3 shows the simplest way of converting atmospheric electricity into electromagnetic wave energy by the use of special motors adapted for high oscillatory currents or static charges of electrical energy. Recent improvements in motors for working with static charges and motors working by resonance, that is to say, having groups of tuned electromagnetic cooperating circuits render this possible but such do not form part of the present invention.

A motor adapted to operate with static charges will for the sake of simplicity be diagrammatically indicated by two semicircles 1 and 2 and the rotor of the motor by a ring M. (Figure 3.) A is a vertical aerial or aerial network. S the safety choke or electromagnet with coil O as may be seen is connected with the aerial A. Adjacent the electromagnet S the aerial conductor is divided into three circuits, the circuit 8 giving the safety spark gap, the circuit 7 with the working spark gap, and then a circuit including the stator terminal 1, the rotor and stator terminal 2 at which a connection is made to the earth wire. The two spark gaps are also connected metallically with the earth wire. The method of working these diagrams is as follows:

The positive atmospheric electric charge collected tends to combine with the negative electricity (or earth electricity) connected with the earth wire. It travels along the aerial A through the electromagnet S without being checked as it flows in the same direction as the direct current. Further, its progress is arrested by two sparks gaps placed in the way and the stator condenser surfaces. The stator condenser surfaces are charged until the charge is greater than the resistance of the spark gap 7, whereupon a spark springs over the spark gap 7 and an oscillatory charge is obtained as by means of the motor M, stator surfaces 1 and 2, and spark gap 7, a closed oscillation circuit is obtained for producing the electromagnetic oscillations. The motor here forms the capacity and the necessary inductance and resistance, which, as is well known, are necessary for converting static electricity into electromagnetic wave energy.

The discharges formed are converted into mechanical energy in special motors and can not reach the aerial network by reason of the electromagnet or choke. If, however, when a spark springs over the spark gap 7 a greater quantity of atmospheric electricity tends to flow to earth, a counter voltage is

induced in the electromagnet, which is greater the more rapidly and strongly the flow of current direct to the earth is. By the formation of this opposing voltage a sufficiently high resistance is offered to the flow of atmospheric electricity direct to earth to prevent a short circuit with the earth.

The circuit containing spark gap 8 having a different wave length which is not in resonance with the natural frequency of the motor, does not endanger the motor and serves as security against excess voltage, which, as practical experiments have shown, may still arise in certain cases, but can be conducted direct to earth through this spark gap.

In the diagram illustrated in Figure 4 the spark gap 7 is shunted across condensers 5 and 6 from the motor M. This construction affords mainly a better insulation of the motor against excess voltage and a uniform excitation through the spark gap 7.

In Figure 5 a diagram is illustrated for transforming large current strengths which may be employed direct without motors, for example, for lighting or heating purposes. The main difference is that here the spark gap consists of a star shaped disc 7 which can rotate on its own axis and is rotated by a motor opposite similarly fitted electrodes 7^a. When separate points of stars face one another, discharges take place, thus forming an oscillation circuit over condensers 5 and 6 and inductance 9 for oscillatory discharges. It is evident that a motor may also be directly connected to the ends of the spiral 9.

The construction of the diagram shown in Figure 6 permits of the oscillation circuit of the motor being connected with an induction coil. Here a regulating inductive resistance is introduced for counter-acting excess voltages in the motor. By cutting the separate coils 9 (coupled inductively to the aerial) in or out the inductive action on the motor may be more or less increased or variable aerial action may be exerted on the oscillation circuit.

In Figure 7 the oscillation circuit is closed through the earth (E and E_1). The spark gap 7 may be prolonged or shortened by more or fewer spark gaps being successively connected by means of a contact arm 7^b.

Diagram 8 shows a unipolar connection of the motor with the aerial network. Here two oscillation circuits are closed through the same motor. The first oscillation circuit passes from aerial A through electromagnet S, point x, inductance 9^a to the earth condenser 6 and further, over spark gap 7 to the aerial condenser 5 and back to x. The second oscillation circuit starts from the aerial condenser 5 at the point x^1 over the inductance 9 to the earth condenser 6 at the point x^3 and through the condenser 6 over the spark gap 7 back to x^1. The motor itself is inserted between the two points of the spark gap 7. From this arrangement slightly damped oscillation wave currents are produced.

In the diagram illustrated in Figure 9 a loosely coupled system of connections is illustrated which is assumed to be for small motors for measuring purposes. A indicates the aerial conductor, S the electromagnet in the aerial conductor, 9 the inductance, 7 the spark gap, 5 and 6 condensers, E the earth, M the motor, and 1 and 2 stator connections of the motor. The motor is directly metallically connected with the oscillation circuit.

In Figure 10 a purely inductive coupling is employed for the motor circuit. The motor is connected with the secondary wire 10 as may be seen in Figure 11 in a somewhat modified diagram connection. The same applies to the diagram of Figure 12.

The diagrams hitherto described preferably allow of motors of small and medium strength to be operated. For large aggregates, however, they are too inconvenient as the construction of two or more oscillation circuits for large amounts of energy is difficult; the governing is still more difficult and the danger in switching on or off is greater.

A means of overcoming such difficulties is shown in Figure 13. The oscillation circuit here runs starting from the point x over condenser 5, variable inductance 9, spark gap 7 and the two segments (3^a and 4^a) forming arms of a Wheatstone bridge, back to x. If the motor is connected by brushes 3 and 4 transversely to the two arms of the bridge as shown in the drawings, electromagnetic oscillations of equal sign are induced in the stator surfaces 1 and 2 and the motor does not revolve. If however, the brushes 3 and 4 are moved in common with the conducting wires 1 and 2 which connect the brushes with the stator poles a certain alteration or displacement of the polarity is obtained and the motor commences to revolve.

The maximum action will result if one brush 3 comes on the central sparking contact 7 and the other brush 4 on the part x. They are however, usually in practice not brought on to the central contact 7 but only held in the path of the bridge segments 4^a and 3^a in order not to connect the spark gaps with the motor oscillation circuit.

As however, the entire oscillation energy can thereby not act on the motor it is better to carry out the same system according to the diagram 14. The diagram 14 differs from the foregoing only by the motor not being directly metallically connected with the segments of the commutator, but only a primary coil 9 which induces in a secondary coil 10, current which feeds the motor M and takes the place of the rotor. By this

arrangement a good transforming action is obtained, a loose coupling and also an oscillation circuit without a spark gap.

In Figure 15 the motor is not purely inductively as in 14, but directly metallically branched off from the primary coil (at x and x^1) after the principle of the auto-transformer.

In Figure 16 instead of an inductance a condenser 6 is in similar manner, and for the same object inserted between the segments 3^a and 4^a. This has the advantage that the segments 3^a and 4^a need not be made of solid metal but may consist of spiral coils whereby a more exact regulation is possible and further motors of high inductance may be employed.

The arrangements of Figures 17, 18 and 19 may be employed for use with resonance and particularly with induction condenser motors; between the large stator induction condenser surfaces, small reversing pole condensers are connected, which, as may be seen from Figures 17, 18 and 19 are led together to earth. Such reversing poles have the advantage that with large quantities of electrical energy the spark formation between the separate oscillation circuits ceases.

Figure 19 shows a further method which prevents electromagnetic oscillations of high number of alternations formed in the oscillation circuit striking back to the aerial conductor. It is based on the well known principle that a mercury lamp, one electrode of which is formed of mercury, the other of solid metal such as steel allows an electric charge to pass in only one direction from the mercury to the steel and not vice versa. The mercury electrode of the vacuum tube N is therefore connected with the aerial conductor and the steel electrode with the oscillation circuit. From this it results that charges can pass only from the aerial through the vacuum tube to the oscillation circuit, but not vice versa. Oscillations which are formed on being transformed in the oscillation circuit cannot pass to the aerial conductor.

In practice these vacuum tubes must be connected behind an electromagnet as the latter alone affords no protection against the danger of lightning.

As regards the use of spark gaps, all arrangements as used for wireless telegraphy may be used. Of course the spark gaps in large machines must have a sufficiently large surface. In very large stations they are cooled in liquid carbonic acid or better still in liquid nitrogen or hydrogen; in most cases the cooling may also take place by means of liquefied low homologues of the metal series or by means of hydrocarbons the freezing point of which lies at between −90° C. and −40° C. The spark gap casing must also be insulated and be of sufficient strength to be able to resist any pressure which may arise. Any undesirable excess super-pressure which may be formed must be automatically let off. I have employed with very good results mercury electrodes which were frozen in liquid carbonic acid, the cooling being maintained during the operation from the outside through the walls.

Figure 20 is one of the simplest forms of construction of an aerial network in combination with collectors, transformers and the like illustrated diagrammatically. E is here the earth wire, 8 the safety spark gap, 7 the working spark gap, 1 and 2 the stator surfaces of the motor, 5 a condenser battery, S the protective magnet which is connected with the coil in the aerial conductor, A^1 to A^{10} aerial antennæ with collecting balloons, N horizontal collecting or connecting wires from which, to the centre a number of connections run.

The actual collectors consist of metal sheaths preferably made of an aluminium magnesium alloy, and are filled with hydrogen or helium and are attached to copper plated steel wires. The size of the balloon is selected so that the actual weight of the balloon and the weight of the conducting wire is supported thereby. On the top of the balloon aluminium spikes, made and gilded in a special manner hereinafter described, are arranged in order to produce a conductor action. Small quantities of radium preparations, more particularly polonium-ionium or mesothorium preparations considerably increase the ionization, and therewith the action of these collectors.

In addition to metal balloons, fabric balloons which are superficially metal coated according to Schoop's metal spraying process, may however also be employed. A metallic surface may also be produced by lacquering with metallic bronzes, preferably according to Schoop's spraying process or lacquering with metallic bronze powders in two electrical series of widely different metals, because thereby the collecting effect is considerably increased.

Instead of the ordinary round balloons, elongated cigar shaped ones may be employed. In order also to utilize the frictional energy of the wind, patches or strips of non-conducting substances which produce electricity by friction, may be attached to the metallized balloon surfaces. The wind will impart a portion of its energy in the form of frictional electricity, to the balloon casing, and thereby the collecting effect is substantially increased.

In practice however, very high towers (up to 300 metres is fully admissible) may be employed as antennæ. In these towers copper tubes rise freely further above the top of the tower. A gas lamp secured

against the wind is then lit at the point of the copper tube and a netting is secured to the copper tube over the flame of this lamp to form a collector. The gas is conveyed through the interior of the tube up to the summit. The copper tube must be absolutely protected from moisture at the place at which it enters the tower and also rain must be prevented running down the walls of the tower which might lead to a bad catastrophe. This is done by bell shaped enlargements which expand downwards, being arranged in the tower in the form of high voltage insulators of Siamese pagodas.

Special attention must be devoted to the foundations of such towers. They must be well insulated from the ground, which may be obtained by first embedding a layer of concrete in a box form to a sufficient depth in the ground and inserting in this an asphalt lining and then glass bricks cast about 1 or 2 metres in thickness. Over this in turn there is a ferro-concrete layer in which alone the metal foot of the tube is secured. This concrete block must be at least 2 metres from the ground and be fully protected at the sides by a wooden covering, from moisture. In the lower part of the tower a wood or glass house for the large condenser batteries or for the motors may be constructed. In order to lead the earth connection to the ground water, a well insulated pit constructed of vitreous bricks, must be provided. Several such towers are erected at equal distances apart and connected with a horizontal conductor. The horizontal connecting wires may either run directly from tower to tower or be carried on bell shaped insulators similar to those in use for high voltage conductors. The width of the network may be of any suitable size and the connection of the motors can take place at any suitable places.

In order to collect large quantities of electricity with few aerials it is well to provide the aerial conductor with batteries of condensers as shown in two methods of construction in Figures 21 and 22. In Figure 21 the batteries of condensers 5 are connected on the one hand with the aerial electricity collectors Z by the aerial conductor A, and on the other hand interconnected in series with an annular conductor from which horizontal conductors run to the connecting points C to which the earth wire is connected.

Figure 22 shows a similar arrangement. Should two such series of antennæ rings be shown by a voltmeter to have a large difference of potential (for example, one in the mountains and one in the plain) or even of different polarity these differences may be compensated for by connecting sufficiently large condenser batteries ($5, 5^a, 5^b$) by means of Maji star conductors D and D^1. In Figure 23 a connection of three such rings of collectors to form a triangle with a central condenser battery is illustrated.

The condenser batteries of such large installations must be embedded in liquefied gases or in liquids freezing at very low temperatures. In such cases a portion of the atmospheric energy must be employed for liquefying these gases. It is also preferable to employ pressure. By this means the condenser surfaces may be diminished, and still allow for large quantities of energy to be stored, secure against breakdown. For smaller installations the immersing of the condensers in well insulated oil or the like, suffices. Solid substances on the other hand cannot be employed as insulators.

The arrangement in the diagrams hitherto described was always such that the condenser batteries were connected with both poles directly to the aerial conductors. An improved diagram of the connections for obtaining atmospheric electricity for the condenser batteries has however, been found to be very advantageous, this arrangement consists in that they are connected by only one pole (unipolar) to the collecting network. Such a method of arrangement is very important, as by means of it a constant current and an increase of the normal working pressure or voltage is obtained. If for example a collecting balloon aerial which is allowed to rise to a height of 300 metres, shows 40,000 volts above earth voltage, in practice it has been found that the working voltage (with a withdrawal of the power according to the method hereinbefore described by means of oscillating spark gaps and the like) is only about 400 volts. If however, the capacity of the condenser surfaces be increased, which capacity in the above mentioned case was equal to that of the collecting surface of the balloon aerials, to double the amount, by connecting the condenser batteries with only one pole, the voltage rises under an equal withdrawal of current up to and beyond 500 volts. This can only be ascribed to the favourable action of the connecting method.

In addition to this substantial improvement it has also been found preferable to insert double inductances with electromagnets and to place the capacities preferably between two such electromagnets. It has also been found that the useful action of such condensers can be further increased if an induction coil be connected as inductive resistance to the unconnected pole of the condenser, or still better if the condenser itself be made as an induction condenser. Such a condenser may be compared with a spring which when compressed carries in itself accumulated force, which it again gives off when released. In charging, a charge with reversed sign is formed at the

other free condenser pole, and if through the spark gap a short circuit results, the accumulated energy is again given back since now new quantities of energy are induced at the condenser pole connected with the conductor network, which in fact charges with opposite signs to that at the free condenser pole. The new induced charges have of course the same sign as the collector network. The whole voltage energy in the aerial is thereby however increased. In the same space of time larger quantities of energy are accumulated than is the case without such inserted condenser batteries.

In Figures 24 and 25 two different diagrams of connections are more exactly illustrated, Figure 24 shows a collecting balloon and the diagram of the connections to earth. Figure 25 four collecting balloons and the parallel connection of the condenser batteries belonging thereto.

A is the collecting balloon made of an aluminium magnesium alloy (electron metal, magnalium) of a specific gravity of 1.8 and a thickness of plate 0.1 to 0.2 mm. Inside there are eight strong vertical ribs of T shaped section about 10 to 20 mm. in height and about 3 mm. in thickness with the projecting part directed inwards (indicated by *a*, *b*, *c*, *d* and so forth); they are riveted together to form a firm skeleton and are stiffened in a horizontal direction by two cross ribs. The ribs are further connected with one another internally and transversely by means of thin steel wires, whereby the balloon obtains great power of resistance and elasticity. Rolled plates of 0.1 to 0.2 mm. in thickness made of magnalium alloy are then either soldered or riveted on this skeleton so that a fully metallic casing with smooth external surface is obtained. Well silvered or coppered aluminium plated steel wires run from each rib to the fastening ring 2. Further, the coppered steel hawser L preferably twisted out of separate thin wires (shown in dotted lines in Figure 24) and which must be long enough to allow the balloon to rise in the desired height, leads to a metal roller or pulley 3 and from thence to a winch W, well insulated from the earth. By means of this winch, the balloon, which is filled with hydrogen, or helium, can be allowed to rise to a suitable height (300 to 5,000 metres) and brought to the ground for recharging or repairs.

The actual current is taken directly through a friction contact from the metal roller 3 or from the wire, or even from the winch or simultaneously from all three by means of brushes (3, 3^a and 3^b). Beyond the brushes the conductor is divided, the paths being:—firstly over 12 to the safety spark gap 8, from thence to the earth conductor E^1, and secondly over electromagnet S^1, point 13, to a second loose electromagnet having an adjustable coil S^2, then to the spark gap 7 and to the second earth conductor E^2. The actual working circuit is formed through the spark gap 7, condensers 5 and 6, and through the primary coil 9; here the static electricity formed by oscillatory discharges is accumulated and converted into high frequency electromagnetic oscillations. Between the electromagnets S^1 and S^2 at the crossing point 13, four condenser batteries are introduced which are only indicated diagrammatically in the drawings each by one condenser. Two of these batteries (16 and 18) are made as plate condensers and prolonged by regulating induction coils or spirals 17 and 19 while the two others (21 and 23) are induction condensers. As may be seen from the drawings each of the four condenser batteries 16, 18, 21, 23 is connected only by one pole to the aerial or to the collector conductor. The second poles 17, 19, 22, 24 are open. In the case of plate condensers having no inductive resistance an induction coil is inserted. The object of such a spiral or coil is the displacement of phase of the induction current by ¼ periods, whilst the charging current of the condenser poles which lie free in the air, works back to the collector aerial. The consequence of this is that in discharges in the collector aerial the back inductive action of the free poles allows a higher voltage to be maintained in the aerial collecting conductor than would otherwise be the case. It has also been found that such a back action has an extremely favourable effect on the wear of the contacts. Of course the inductive effect may be regulated at will within the limits of the size of the induction coil, the length of the coil in action being adjustable by means of wire connection without induction (see Fig. 24, No. 20).

S^1 and S^2 may also be provided with such regulating devices in the case of S^2 (illustrated by 11). If excess voltage be formed it is conducted to earth through the wire 12 and spark gap 8 or through any other suitable apparatus, since this formation would be dangerous for the other apparatus.

The action of these condenser batteries has already been hereinbefore described.

The small circles on the collector balloon indicate places at which zinc amalgam or gold amalgam or other photoelectric acting metals in the form of small patches in extremely thin layers (.01 to .05 mm. in thickness) are applied to the balloon casing of light metal. Such metallic patches may also be applied to the entire balloon as well as in greater thickness to the conducting network. The capacity of the collector is thereby considerably strengthened at the surface. The greatest possible effect in collecting may be obtained by polonium amalgams and the like. On the surface of the

collector balloon metal points or spikes are also fixed along the ribs, which spikes serve particularly for collecting the collector charge. Since it is well known that the resistance of the spikes is less the sharper the spike is, for this purpose it is therefore extremely important to employ as sharp spikes as possible. Experiments made as regards these have shown that the formation of the body of the spike or point also plays a large part. for example, spikes made of bars or rollers with smooth surfaces, have a many times greater point resistance as collector accumulator spikes than those with rough surfaces. Various kinds of spike bodies have been experimented with for the collector balloons hereinbefore mentioned. The best results were given by spikes which were made in the following way. Fine points made of steel, copper, nickel, or copper and nickel alloys, were fastened together in bundles and then placed as anode with the points in a suitable electrolyte (preferably in hydrochloric acid or muriate of iron solutions) and so treated with weak current at 2 to 3 volts pressure. After 2 to 3 hours according to the thickness of the spikes or pins the points become extremely sharp and the bodies of the spikes have a rough surface. The bundle can then be removed and the acid washed off with water. The spikes are then placed as cathode in a bath consisting of solution of gold, platinum, iridium, paladium or wolfram salts or their compounds and coated at the cathode galvanically with a thin layer of precious metal, which must however be sufficiently firm to protect them from atmospheric oxidation.

Such spikes act at a 20 fold lower voltage almost as well as the best and finest points made by mechanical means. Still better results are obtained if polonium or radium salts are added to the galvanic bath when forming the protective layer or coating. Such pins have a low resistance at their points and even at one volt and still lower pressures have an excellent collector action.

In Figure 24 the three unconnected poles are not connected with one another in parallel. That is quite possible in practice without altering the principle of the free pole. It is also preferable to interconnect in parallel to a common collector network, a series of collecting aerials.

Figure 25 shows a diagram for such an installation. A^1, A^2, A^3, A^4 are four metal collector balloons with gold or platinum coated spikes which are electrolytically made in the presence of polonium emanations or radium salts, which spikes or needles are connected over four electro-magnets S^1, S^2, S^3, S^4, through an annular conductor R. From this annular conductor four wires run over four further electromagnets S^a, S^b, S^c, S^d, to the connecting point 13. There the conductor is divided, one branch passing over 12 and the safety spark gap 8 to the earth at E^1, the other over inductive resistance J and working spark gap 7 to the earth at E^2. The working circuit, consisting of the condenser 5 and 6 and a resonance motor or a condenser motor M, such as hereinbefore described, is connected in proximity round the sparking gap section 7.

Instead of directly connecting the condenser motor of course the primary circuit for high frequency oscillatory current may also be inserted.

The condenser batteries are connected by one pole to the annular conductor R and can be either inductionless (16 and 18) or made as induction condensers as shown by 21 and 23. The free poles of the inductionless condensers are indicated by 17 and 19, those of the induction condensers by 22 and 24. As may be seen from the drawings all these poles 17, 22, 19, 24 may be interconnected in parallel through a second annular conductor without any fear that thereby the principle of the free pole connection will be injured. In addition to the advantages already set forth the parallel connection also allows of an equalization of the working pressure in the entire collector network. Suitably constructed and calculated induction coils 25 and 26 may also be inserted in the annular conductor of the free poles, by means of which a circuit may be formed in the secondary coils 27 and 28 which allows current produced in this annular conductor by fluctuations of the charges or the like appearances to be measured or otherwise utilized.

According to what has been hereinbefore stated separate collector balloons may be connected at equidistant stations distributed over the entire country, either connected directly with one another metallically or by means of intermediate suitably connected condenser batteries through high voltage conductors insulated from earth. The static electricity is converted through a spark gap into dynamic energy of a high number of oscillations and may in such form be coupled as a source of energy by means of a suitable method of connecting, various precautions being observed and with special regulations. The wires leading from the collector balloons have hitherto been connected through an annular conductor without this endless connection, which can be regarded as an endless induction coil, being able to exert any action on the whole conductor system.

It has now been found that if the network conductor connecting the aerial collector balloons with one another is not made as a simple annular conductor, but preferably short circuited in the form of coils over a

condenser battery or spark gap or through thermionic tubes or valves or audions, then the total collecting network exhibits quite new properties. The collection of atmospheric electricity is thereby not only increased but an alternating field may be easily produced in the collector network. Further, the atmospheric electrical forces showing themselves in the higher regions may also be directly obtained by induction. In Figures 26 and 28 a form of construction is shown on the basis of which the further foundations of the method will be more particularly explained.

In Figure 26 1, 2, 3, 4 are metal collector balloons, 5, 6, 7, 8 their metallic aerial conductors and I the actual collector network. This consists of five coils and is mounted on high voltage insulators in the air, on high voltage masts (or with a suitable construction of cable embedded in the earth). One coil has a diameter of 1 to 100 km. or more. S and S^1 are two protective electromagnets, F the second safety section against excess voltage, E its earth conductor and E^1 the earth conductor of the working section. When an absorption of static atmospheric electricity is effected through the four balloon collectors, the current in order to reach the earth connection E^1 must flow spirally through the collector network over the electromagnet S, primary induction coil 9, conductor 14, anode A of the audion tube, incandescent cathode K, as the way over the electromagnet and safety spark gap F offers considerably greater resistance. Owing to the fact that the accumulated current flows in one direction, an electromagnetic alternating field is produced in the interior of the collector network coil, whereby the whole free electrons are directed more or less into the interior of the coil. An increased ionization of the atmosphere is therefore produced. In consequence of this the points mounted on the collector balloon show a considerably reduced resistance and therefore increased static charges between the points on the balloon and the surrounding atmosphere are produced. The result of this is a considerably increased collector effect.

A second effect which could not be obtained otherwise is obtained by the electromagnetic alternating field which running parallel to the earth surface, acts more or less with a diminishing or increasing effect on the earth magnetic field, whereby in the case of fluctuations in the current a return induction current of reversed sign is always produced in the collector coil by earth magnetism. Now if, however, a constantly pulsating continuous alternating field is produced as stated in the above collector network I, an alternating current of the same periodicity is produced also in the collecting network coil. As the same alternating field is further transmitted to the aerial balloon, the resistance of its points is thereby considerably reduced, whilst the collector action is considerably increased. A further advantage is that positive electrons which collect on the metal surfaces during the conversion into dynamic current produce a so-called drop of potential of the collector area. As an alternating field is present, the negative ions surrounding the collector surfaces, when discharge of the collector surfaces takes place produce by the law of induction, an induction of reversed sign on the collector surface and so forth (that is to say again a positive charge). In addition to the advantages hereinbefore set forth, the construction of connecting conductors in coil form when of sufficiently large diameter, allows of a utilization of energy arising in higher regions also in the simplest way. As is well known electric discharges frequently take place at very great elevations which may be observed, such as St. Elmo's fires or northern lights. These energy quantities have not been able to be utilized up to now. By this invention all these kinds of energy, as they are of an electromagnetic nature and the direction of the axis of the collector coils stands at right angles to the earth's surface, can be more or less absorbed in the same way as a receiver in wireless telegraphy absorbs waves coming from a far distance. With a large diameter of the spiral it is possible to connect large surfaces and thereby to take up also large quantities of energy.

It is well known that large wireless stations in the summer months, and also in the tropics are very frequently unable to receive the signals in consequence of interruptions which are caused by atmospheric electricity, and this takes place with vertical coils of only 40 to 100 m. diameter. If on the contrary horizontal coils of 1 to 100 km. diameter be employed very strong currents may be obtained through discharges which are constantly taking place in the atmosphere. Particularly in the tropics or still better in the polar regions where the northern lights are constantly present, large quantities of energy may probably be obtained in this way. A coil with several windings should act the best. In similar manner any alteration of the earth magnetism should act inductively on such a coil.

It is not at all unlikely that earthquakes and spots on the sun will also produce an induction in such collector coils of sufficient size. In similar manner this collector conductor will react on earth currents more particularly when they are near the surface of the earth or even embedded in the earth. By combining the previous kind of current collectors so far as they are adapted for

the improved system with the improved possibilities of obtaining current the quantities of free natural energy which are to be obtained in the form of electricity are considerably increased.

In order to produce in the improved collector coil uniform current oscillations of an undamped nature so-called audion high vacuum or thermionic tubes of suitable connection are employed instead of the previously known spark gaps (Fig. 26, Nos. 9–18). The main aerial current flows through electromagnet S (which in the case of a high number of alternations is not connected here but in the earth conductor E^1) and may be conveyed over the primary coils in the induction winding through wire 14 to the anode A of the high vacuum grid tube. Parallel with the induction resistance 9 a regulating capacity of suitable size, such as condenser 11 is inserted. In the lower part of the vacuum grid tube is arranged the incandescent filament or the cathode K which is fed through a battery B. From the battery B two branches run, one to the earth conductor E^1 and the other through battery B^1 and secondary coil 10 to the grid anode *g* in the vacuum tube. By the method of connections shown in dotted lines, a desired voltage at the grid electrode *g* may also be produced through the wire 17 which is branched off from the main current conductor through switches 16 and some small condensers (*a*, *b*, *c*, *d*) connected in series, and conductor 18, without the battery B^1 being required.

The action of the entire system is somewhat as follows:—

On the connecting conductor of the aerial collector network being short circuited to earth, the condenser pole 11 is charged and slightly damped oscillations are formed in the short circuited existing oscillation circuit formed of the condenser 11 and self inductance 9. In consequence of the coupling through coil 10, fluctuations of voltage take place in the grid circuit 15 with the same frequency, which, fluctuations in turn influence the strength of the electrode current passing through the high vacuum amplyfying tube and thus produce current fluctuations of the same frequency in the anode circuit. A permanent supply of energy to the oscillation circuits 9 and 10 consequently takes place, until a condition of balance is set up, in which the consumed oscillation energy is equal to that absorbed. Thereby constant undamped oscillations are now produced in the oscillation circuits 9–11.

For regular working of such oscillation producers high vacuum strengthening tubes are necessary and it is also necessary that the grid and anode voltages shall have a phase difference of 180° so that if the grid is negatively charged, then the anode is positively charged and vice versa. This necessary difference of phase may be obtained by most varied connections, for example, by placing the oscillation circuit in the grid circuit or by separating the oscillation circuit and inductive coupling from the anodes and the grid circuit and so forth.

A second important factor in this way of converting static atmospheric electricity into undamped oscillations is that care must be taken that the grid and anode voltages have a certain relation to one another; the latter may be obtained by altering the coupling and a suitable selection of the self induction in the grid circuit, or as shown by dotted lines 18, 17, 16 by means of a larger or smaller number of condensers of suitable size connected in series; in this case the battery B^1 may be omitted. With a suitable selection of the grid potential a glow discharge takes place between the grid *g* and the anode A, and accordingly at the grid there is a cathode drop and a dark space is formed. The size of this cathode drop is influenced by the ions which are emitted in the lower space in consequence of shock ionization of the incandescent cathodes K and pass through the grid in the upper space. On the other hand the number of the ions passing through the grid is dependent on the voltage between the grid and the cathode. Thus if the grid voltage undergos periodic fluctuations (as in the present case) the amount of the cathode drop at the grid fluctuates and consequently the internal resistance of the tube correspondingly fluctuates, so that when a back coupling of the feed circuit with the grid circuit takes place, the necessary means are afforded for producing undamped oscillations and of taking current, according to requirements from the collecting conductor.

The frequency of the undamped oscillations produced is with a suitably loose coupling equal to the self frequency of the oscillation circuits 9 and 10. By a suitable selection of the self induction of the coil 9 and capacity 11 it is possible to extend from frequencies which produce electromagnetic oscillations of only a few metres wave length down to the lowest practical alternating current frequency. For large installations a suitable number of frequency producing tubes in the form of the well known high vacuum transmission tubes of .5 to 2 kw. in size may be connected in parallel so that in this respect no difficulty exists.

The use of such tubes for producing undamped oscillations, and also the construction and method of inserting such transmission tubes in an accumulator or dynamo circuit is known and also that such oscillation producing tubes only work well at voltages of 1,000 up to 4,000 volts, so that on the contrary their use at lower voltages is

considerably more difficult. By the use of high voltage static electricity this method of producing undamped oscillations as compared with that through spark gaps must be regarded as an ideal solution particularly for small installations of outputs of from 1 to 100 kw.

By the application of safety spark gaps, with interpolation of electro-magnets, not only is short circuiting avoided but also the taking up of current is regulated. Oscillation producers inserted in the above way form a constantly acting electromagnetic alternating field in the collector coil, whereby as already stated, a considerable accumulating effect takes place. The withdrawal wire or working wire is connected at 12 and 13, but current may be taken by means of a secondary coil which is firmly or movably mounted in any suitable way inside the large collector coil, i. e. in its electromagnetic alternating field, so long as the direction of its axis runs parallel with that of the main current collecting coil.

In producing undamped oscillations of a high frequency (50,000 per second and more) in the oscillation circuits 9 and 11, electromagnets S and S¹ must be inserted if the high frequency oscillations are not to penetrate the collector coil, between the oscillation producers and the collector coil. In all other cases they are connected shortly before the earthing (as in Figs. 27 and 28).

In Figure 27 a second method of construction of the connecting conductor of the balloon aerials is illustrated in the form of a coil. The main difference consists in that in addition to the connecting conductor I another annular conductor II is inserted parallel to the former on the high voltage masts in the air (or embedded as a cable in the earth) but both in the form of a coil. The connecting wire of the balloon aerials is indicated as a primary conductor and also as a current producing network; the other is the consumption network and is not in unipolar connection with the current producing network.

In Figure 27 the current producing network I is shown with three balloon collectors 1, 2, 3 and aerial conductors 4, 5, 6; it is short circuited through condenser 19 and inductance 9. The oscillation forming circuit consists in this diagram of spark gap *f*, inductance 10, and condenser 11; the earth wire E, is connected to earth over electromagnet S¹. F is the safety spark gap which is also connected to earth through a second electromagnet S at E. On connecting up the condenser circuit 11 this is charged over the spark gap *f* whereby an oscillatory discharge is formed. This discharging current acts through inductance 10 on the inductively coupled secondary 9, whereby in the producing network a modification of the potential of the condenser 19 is produced. The consequence of this is that oscillations arise in the coil shaped producer network. These oscillations induce a current in the secondary circuit II, which has a smaller number of windings and a less resistance, the voltage of which, according to the proportion of the number of windings and of the ohmic resistance, is considerably lower whilst the current strength is greater.

In order to convert the current thus obtained into current of an undamped character, and to tune its wave lengths, a sufficiently large regulatable capacity 20 is inserted between the ends 12 and 13 of the secondary conductor II. Here also current may be taken without an earth conductor, but it is advisable to insert a safety spark gap E¹ and to connect this with the earth over an electromagnet S².

The producer network may be connected with the working network II over an inductionless condenser 21 or over an induction condenser 22, 23. In this case the secondary conductor is unipolarly connected with the energy conductor.

In Figure 28 the connecting conductor between the separate accumulator balloons is carried out according to the autotransformer principle. The collecting coil connects four aerial balloons 1, 2, 3, 4, the windings of which are not made side by side but one above the other. In Figure 28 the collector coil I is shown with a thin line, the metallically connected prolongation coils II with a thick line. Between the ends I¹ and II¹ of the energy network I a regulating capacity 19 is inserted. The wire I¹ is connected with the output wire and with the spark gap F.

As transformer of the atmospheric electricity an arrangement is employed which consists in using rotary pairs of condensers in which the one stator surface B is connected with the main current, whilst the other A is connected with the earth pole. Between these pairs of short circuited condensers are caused to rotate from which the converted current can be taken by means of two collector rings and brushes, in the form of an alternating current, the frequency of which is dependent on the number of balloons and the revolutions of the rotor. As the alternating current formed in the rotor can act, in this improved method of connection described in this invention, through coils 10 on the inductance 9, an increase or diminution of the feed current in I can be obtained according to the direction of the current by back induction. Current oscillations of uniform rhythm thereby result in the coil shaped windings of the producer net-work.

As the ends of this conductor are short cir-

cuited through the regulatable condenser 19 these rhythms produce short circuited undamped oscillations in the energy conductor, the periodicity and wave lengths of which oscillations can be adjusted according to desire by altering the capacity 19 to a given wave length and therewith also to a given frequency. These currents may also be employed in this form directly as working current through the conductors II^1 and III. By inserting the condenser 20 a connection between these conductors may also be made, whereby harmonic oscillations of desired wave length are formed. By this means quite new effects as regards current distribution are obtained. The withdrawal of current can even take place without direct wire connection if, at a suitable point in the interior of the producing network (quite immaterially whether this has a diameter of 1 or 100 km.) a coil tuned to these wave lengths and of the desired capacity is firmly or movably mounted in the aerial conductor in such a way that its axial direction is in parallel with that of the collector coil. In this case a current is induced in the producing network, the size of which is dependent on the total capacity and resistance and also on the periodicity employed. A possibility is thereby afforded in future, of taking energy from the producer network by wireless means. As thereby in addition to atmospheric electricity also magnetic earth currents and the energy from the higher atmosphere (at least partially) may be simultaneously obtained, this last system for collecting the atmospheric energy is of particular importance for the future.

Of course everywhere instead of spark gaps suitable grid vacuum tubes may be employed as producers for undamped oscillations. The separate coils of the producer net-work with large diameters may be connected with one another through separate conductors all in parallel or all in series or in groups in series. By regulating the number of oscillations and also the extent of the voltage more or less large collector coils of this kind may be employed. The coils may also be divided spirally over the entire section. The coils may be carried out in annular form or also in triangular, quadrangular, hexagonal or octagonal form.

Of course wires may be carried from a suitable place to the centre or also laterally which serve the current waves as guides. This is necessary when the currents have to be conducted over mountains and valleys and so forth. In all these cases the current must be converted into a current of suitable periodicity.

As already hereinbefore mentioned separate collecting balloons may be directly metallically interconnected at equidistant stations distributed over the entire country or may be connected by interpolation of suitable condenser batteries by means of high voltage conductors. The static electricity is converted through a spark gap into dynamic energy of a high number of oscillations, and could then in such form, with a suitable arrangement of the connections, observing various measures of precaution, be employed as source of energy after separate or special regulation.

According to this invention in order to increase the collecting effect of the balloon in the aerial collector conductor or in the earth wire, radiating collectors are employed. These consist either of incandescent metal or oxide electrodes in the form of vacuum grid tubes, or electric arcs (mercury and the like electrodes) Nernst lamps, or finally flames of various kinds may be simply connected with the respective conductor.

It is well known that energy can be drawn off from a cathode consisting of an incandescent body opposite an anode charged with positive electricity (vacuum grid tube). Hitherto however, a cathode was always first directly placed opposite an anode, and secondly the system always consisted of a closed circuit.

Now if we dispense with the ordinary ideas in forming light or flame arcs in which a cathode must always stand directly opposite an anode, and if we place an incandescent cathode opposite an anode charged to a high potential or another body freely floating in the air, or regard the incandescent cathode only as a source of unipolar discharge (which represent group and point discharges in electro-static machines similar to unipolar discharges), it may be ascertained that incandescent cathodes and less perfectly all incandescent radiators, flames and the like admit of relatively large current densities and allow large quantities of electric energy to radiate into the open space in the form of electron streams as transmitters.

The object of this invention is as described below, if such incandescent oxide electrodes or other incandescent radiators or flames are not freely suspended in space but connected metallically with the earth so that they can be charged with negative terrestrial electricity, these radiators possess the property of absorbing the free positive electrical charges contained in the air space surrounding them (that is to say of collecting them and conducting them to earth). They can therefore, serve as collectors and have, in comparison to the action of the spikes, or points, a very large radius of action R; the effective capacity of these collectors is much greater than the geometrical capacity (R_0–) calculated in an electro-static sense.

Now as our earth is surrounded as is well

known, with an electro-static field and the difference of potential

$$\frac{\delta V}{\delta h}$$

of the earth field according to the latest investigations, is in summer about 60 to 100 volts and in winter 300 to 500 volts per metre of difference in height (δh), a simple calculation gives the result that when such a radiation collector or flame collector is arranged for example on the ground, and a second one is mounted vertically over it at a distance of 2,000 metres and both are connected by a conducting cable, there is a difference of potential in summer of about 2,000,000 volts and in winter even of 6,000,000 volts and more.

According to Stefan Boltzmann's law of radiation, the quantity of energy which an incandescent surface (temperature T) of 1 sq. cm. radiates in a unit of time into the open air (temperature T_o) is expressed by the following formula:

$$S=\sigma\,(T^4-T^4_0)\ \text{watt/sq. cm.}$$

and the universal radiation constant σ is according to the latest researches of Ferry (Annales de Chimie et de physique 17 page 267 (1909)) equal to 6.30×10^{-12} watt/sq. cm.

Now if an incandescent surface of 1 sq. cm. shows, as compared with the surrounding space a periodic fall of potential δV it radiates (independent of the current direction, that is to say of the sign) in accordance with the above formula, for example at a temperature of 3725° C. an energy of 1.6 kw. per sq. cm. per second. As for the radiation the same value can be calculated for the collection of energy, but reversed. Now as carbon electrodes at the temperature of the electric arc support on the current basis a current density up to from 60 to 65 amperes per sq. cm. no difficulties will result in this direction in employing radiating collectors as accumulators.

If the earth be regarded as a cosmically insulated condenser in the sense of geometrical electro-statics *x* there results from the geometric (compare Ewald Rasch, "das elektrische Bogenlicht" (The electric arc light) page 169) capacity of the earth according to Chwolson:

For negative charging 1.3×10^6 Coulomb

For negative potential $V=10\times10^8$ volts.

From this there results however, $EJT\cong24.7\times10^{24}$ watt/Sec. Now if it is desired to make a theoretic short circuit through an earthed flame collector this would represent an electric total work of about 79,500 10^{10} kilowatt years. As the earth must be regarded as a rotating mechanism which is thermo-dynamically, electromagnetically, and also kinematically coupled with the sun and stars system by cosmic radiations and gravitation a diminution of the electric energy of the earth field is not to be feared. The energies which the incandescent collectors would withdraw from the earth field can only cause by the withdrawal of motor work a lowering of the earth temperature (temperature $T_E=300$) and reduce this to that of the world space ($T=O$) by using the entire energy. This is however not the case as the earth does not represent a cosmically entirely insulated system. On the contrary there is conveyed to the same according to the recent value corrected by Ferry for the solar constants through the radiation from the sun an energy of $18{,}500\times10^{10}$ kw. Accordingly any lowering of the earth temperature (T_E) without a simultaneous lowering of the sun's temperature (T_S) would contradict Stefan Boltzmann's law of radiation.

$$S=\sigma\,(T_S^4-T_E^4).$$

From this it must be concluded that if the earth temperature (T_E) sinks, the total radiation S absorbed by the earth increases, and further also that the secular speed of cooling of the earth is directly dependent on that of the sun and the other radiators cosmically coupled with the sun and is connected most closely with these.

The incandescent radiation collectors may, according to this invention, be employed for collecting atmospheric electricity if they (1) are charged with the negative earth electricity (that is to say when they are directly connected by means of a metallic conductor with the earth) and (2) if large capacities (metal surfaces) charged with electricity are mounted opposite them as positive poles in the air. This is regarded as the main feature of the present invention as without these inventive ideas it would not be possible to collect with an incandescent collector, sufficiently large quantities of the electrical charges contained in the atmosphere as technology requires; the radius of action of the flame collectors would also be too small, especially if it be considered that the very small surface density (energy density) (σ about$=2\times7\ .\ 10^9$ St. E. per sq. cm.) does not allow of large quantities of charge being absorbed from the atmosphere.

x) Calculated according to Poisson's calculation;

$\Delta V=-4\pi\delta$; as here the alteration of the potential or potential gradients only takes place in the direction of the normal, this calculation assumes the simple form

$$\delta=\frac{1}{4\pi}\times\frac{\delta^2V}{\delta n^2}$$

It has indeed already been proposed to employ flame collectors for collecting atmospheric electricity and it is known that

their collecting effect is substantially greater opposite the points. It is however, not known that the quantities of current which could hitherto be obtained are too small for technical purposes. According to my experiments the reason for this is to be found in the too small capacities of the collector conductor poles. If such flame or radiating collectors have no or only small positive surfaces, their radius of action for large technical purposes is too small. If the incandescent collectors be constantly kept in movement in the air they may collect more according to the speed of the movement, but this is again not capable of being carried out in practice.

By this invention the collector effect is considerably increased by a body charged with a positive potential and of the best possible capacity being also held floating (without direct earth connection) opposite such an incandescent collector which is held floating in the air at a desired height. If for example, a collecting balloon of sheet metal or of metalized balloon fabric be caused to mount to 300 up to 3,000 metres in the air and as positive pole it is brought opposite such a radiating collector connected by a conductor to earth, quite different results are obtained.

The metallic balloon shell (with a large surface) is charged to a high potential by the atmospheric electricity. This potential is greater the higher the collecting balloon is above the incandescent collector. The positive electricity acts concentratedly on the anode floating in the air as it is attracted through the radiation shock ionization, proceeding from the incandescent cathode. The consequence of this is that the radius of action of the incandescent cathode collector is considerably increased and thereby also the collecting effect of the collecting balloon surface. Further the large capacity of the anode floating in the air plays therefore an important part because it allows of the taking of large charges, and thereby a more uniform current is obtained even when there is a large consumption; this cannot be the case with small surfaces.

In the present case the metallic collecting balloon is a positive anode floating in the air and the end of the earth conductor of this balloon serves as positive pole surface opposite the surface of the radiating incandescent cathode, which in turn is charged with negative earth electricity being conductingly connected to earth.

The process may be carried out by two such contacts (negative incandescent cathode and anode end of a capacity floating in the air) a condenser and an inductive resistance being switched on in parallel, whereby simultaneously undamped oscillations may be formed.

In very large installations it is advisable to connect two such radiating collectors in series. Thus an arc light incandescent cathode may be placed below on the open ground and an incandescent cathode which is heated by special electro-magnetic currents be located high in the air. Of course for this the special vacuum Liebig tubes with or without grids may also be employed. An ordinary arc lamp with oxide electrodes may be introduced on the ground and the positive pole is not directly connected with the collecting balloon, but through the upper incandescent cathode or over a condenser. The method of connecting the incandescent cathode floating in the air may be seen in Figs. 29–33.

B is the air balloon, K a Cardan ring (connection with the hawser) C the balloon, L a good conducting cable, P a positive pole, N negative incandescent cathode, and E earth conductor.

Fig. 29 represents the simplest form of construction. If electric oscillations are produced below on the ground by means of a carbon arc lamp or in other suitable way a considerably greater electric resistance is opposed to that in the direct way by inserting an electrical inductive resistance 9. Consequently between P and N a voltage is formed, and as, over N and P only an inductionless ohmic resistance is present, a spark will spring over so long as the separate induction co-efficients and the like are correctly calculated. The consequence of this is that the oxide electrode (carbon or the like) is rendered incandescent and then shows as incandescent cathode an increased collecting effect. The positive poles must be substantially larger than the negative in order that they may not also become incandescent. As they are further connected with the large balloon area which has a large capacity and is charged at high voltage, an incandescent body which is held floating in the air and a positive pole which can collect large capacities is thereby obtained in the simplest way. The incandescent cathode is first caused to become incandescent by means of separate energy produced on the earth, and then maintained by the energy collected from the atmosphere.

Fig. 30 only shows the difference that instead of a round balloon a cigar shaped one (of metal or metalized fabric) may be employed and also a condenser 5 is inserted between the incandescent cathode and the earth conductor so that a short circuited oscillation circuit over P. N 5 and 9 is obtained. This has the advantage that quite small quantities of electricity cause the cathode to become incandescent and much larger cathode bodies may be rendered incandescent.

In this form of construction both the in-

candescent cathode and also the positive electrode may be enclosed in a vacuum chamber as may be seen in Fig. 32. A cable L is carried well insulated through the cover of a vessel and ends in a condenser disc 5. The cover is arched in order to keep off the rain. The vessel is entirely or partially made of magnetic metal and well insulated inside and outside. Opposite the disc 5 another disc 6 and on this again a metallic positive pole of the vacuum tube *g* with the incandescent cathode (oxide electrode) N is arranged. The negative electrode is on the one hand connected with the earth conductor E, and on the other hand with the inductive resistance 9 which is also connected with the cable L with the positive pole and wound round the vessel in coils. The action is exactly the same as that in Fig. 29 only instead of an open incandescent cathode one enclosed in vacuo is employed. As in such collectors only small bodies can be brought to incandescence in large installations a plurality of such vacuum tubes must be inserted in proximity to one another. According to the previous constructions Figs. 31 and 33 are quite self evident without further explanations.

Figs. 34–37 represent further diagrams of connections over radiating and flame collectors, and in fact, how they are to be arranged on the ground.

Fig. 34 shows an arc light collector with oxide electrodes for direct current and its connection; Fig. 35 a similar one for alternating current, Fig. 36 an incandescent collector with a Nernst lamp and Fig. 37 a similar one with a gas flame.

The positive pole 1 of the radiating collectors is always directly connected to the aerial collecting conductor A. In Fig. 34 this is further connected over the condenser battery 5 with a second positive electrode 3. The direct current dynamo *b* produces current which flows over between the electrodes 3 and 2 as an arc light. On the formation of an arc the negative incandescent electrode 2 absorbs electricity from the positive poles standing opposite it and highly charged with atmospheric electricity and conveys the same to the working circuit. The spark gap 7, inductive resistance 9 and induction coil 10 are like the ones previously described. The protective electromagnet S guards the installation against earth circuiting, the safety spark gap 8 from excess voltage or overcharging.

In Fig. 35 the connection is so far altered that the alternating current dynamo feeds the exciting coil 11 of the induction condenser. 12 is its negative and 13 its positive pole; if the coil 3 on the magnet core of the dynamo is correctly calculated and the periodicity of the alternating current is sufficiently high an arc light can be formed between the two poles 1 and 2. As the cathode 2 is connected with the negatively charged earth, and therefore always acts as a negative pole, a form of rectification of the alternating current produced by the dynamo 3 is obtained, the second half of the period is always suppressed. The working circuit may be carried out in the same way as in Fig. 34; the working spark gap 7 may however be dispensed with, and instead thereof between the points *n* and *m* a condenser 5 and an induction resistance 9 may be inserted from which the current is taken inductively.

Fig. 36 represents a form of construction similar to Fig. 34 only that here instead of an arc lamp a Nernst incandescent body is employed. The Nernst lamp is fed through the battery 3. The working section is connected with the negative pole, the safety spark gap with the + poles. The working spark gap 7 may also be dispensed with and the current for it taken at 12 over the oscillation circuit 5, 11 (shown in dotted lines).

Flame collectors (Fig. 37) may also be employed according to this invention. The wire network 1 is connected with the aerial collector conductor A and the burner with the earth. At the upper end of the latter, long points are provided which project into the flame. The positive electrode is connected with the negative over a condenser 5 and the induction coil 9 with the earth.

The novelty in this invention is firstly, the use of incandescent cathodes opposite positive poles which are connected with large metallic capacities as automatic collecting surfaces, (2) the connection of the incandescent cathodes with the earth whereby, in addition to the electricity conveyed to them from the battery or machine which causes the incandescing, also the negative charge of the earth potential is conveyed, and (3) the connection of the positive and negative poles of the radiating collectors over a condenser circuit alone or with the introduction of a suitable inductive resistance, whereby simultaneously an oscillatory oscillation circuit may be obtained. The collecting effect is by these methods quite considerably increased.

I declare that what I claim is:—

1. An electrical energy generating system, comprising a conducting surface for static charges, means to support same at a distance above the earth, a conductor leading to the earth level, a spark gap associated with said conductor to convert electrostatic charges into electromagnetic high frequency oscillations means to supply said electromagnetic energy to a net work, and a spark gap of greatly increased relative resistance in parallel therewith.

2. An electrical energy generating system

comprising a conductor, means to support same above the earth level, an inductance therein, a spark gap associated with said conductor, a second spark gap of much higher relative resistance in parallel therewith and an energy receiving circuit coupled with the spark gap of lesser resistance.

3. An electrical energy generating system comprising a collecting surface, means to support same above the earth level, a conductor connecting said collecting surface with the earth level, a choke in said conductor, an electromagnetic resistance converting electrostatic energy to electromagnetic energy, a safety higher resistance in parallel therewith and a net work coupled with the conversion resistance of lesser value.

4. An electrical energy generating system comprising electric conductors spaced above the earth to form electromagnetic oscillating circuits, conductors connecting to earth level, electrostatic to electromagnetic energy conversion means therein, a safety high electrostatic resistance in parallel therewith and means to alter the electromagnetic characteristic of the circuits.

5. An electrical energy generating system comprising in combination a static collecting surface arranged above the earth, conductors connecting to earth level, a pair of spark gaps in parallel of different electrostatic resistance, a utilization net work shunted across the spark gap of lesser resistance and an electromagnetic choke in said conductors.

6. An electrical energy generating system comprising an open circuit energy collecting aerial, a pair of sparking gaps in parallel of widely different resistance, connected thereto and a closed electric oscillation circuit in shunt across the gap of lesser resistance.

7. An electrical energy generating system comprising an open circuit energy collecting aerial, a pair of sparking gaps in parallel of widely different resistance connected thereto, a closed electric oscillation circuit in shunt across the gap of lesser resistance, a plurality of electrostatic collecting surfaces, means to connect said collecting surfaces in parallel in groups and means to connect said groups symmetrically with said aerial.

In witness whereof, I have hereunto signed my name this 30 day of Dec., 1920, in the presence of two subscribing witnesses.

HERMANN PLAUSON.

Witnesses:
H. F. ARMSTRONG,
W. H. BEESTON.

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12, Sheets-Sheet 1

Fig. 1.

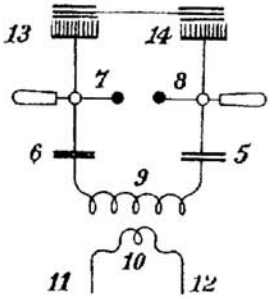
13
14
7
8
6
5
9
10
11
12

Fig. 2.

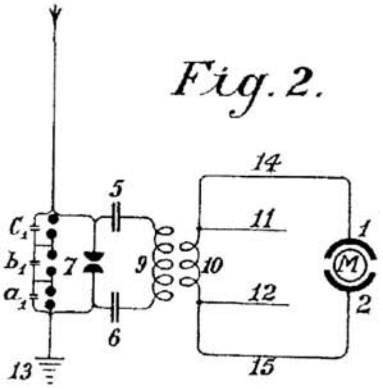
5
14
11
1
C1
b1
7
9
10
a1
12
2
6
13
15
M

Fig. 3.

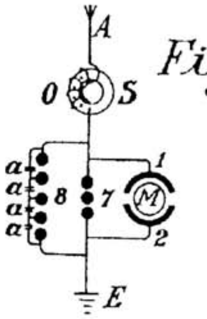
A
0
S
a
a
a
a
a
8
7
1
M
2
E

Fig. 4.

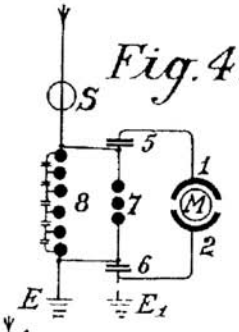
S
5
1
8
7
M
2
6
E
E1

Fig. 5.

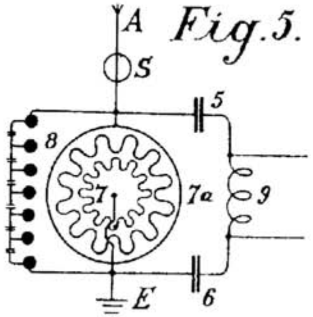
A
S
5
8
7
7a
9
E
6

Fig. 6

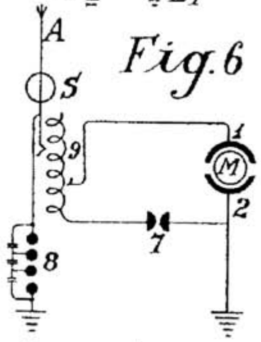

Inventor

Hermann Plauson

by Knight Bros

Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12, Sheets-Sheet 2

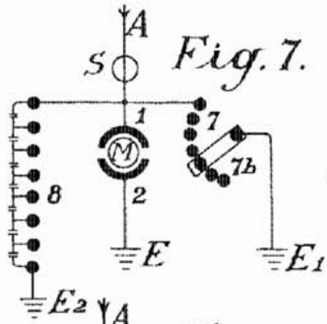
Fig. 7.
A
S
1
M
2
7
7b
8
E
E1
E2

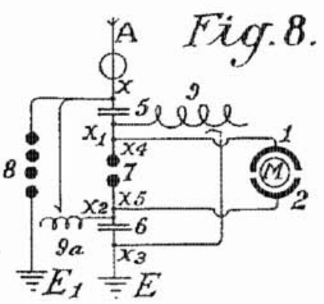
Fig. 8.
A
X
5
9
X1
X4
7
1
M
2
8
X2
X5
6
9a
X3
E1
E

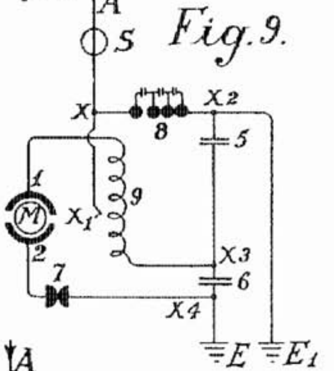
Fig. 9.
A
S
X
8
X2
5
1
M
X1
9
2
X3
6
7
X4
E
E1

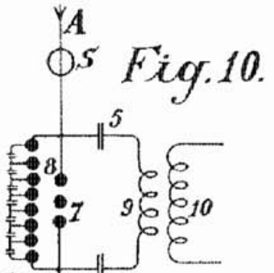
Fig. 10.

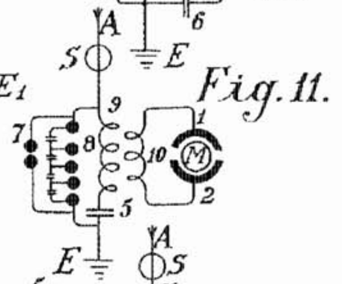
Fig. 11.
A
S
9
1
M
7
8
10
2
5
E

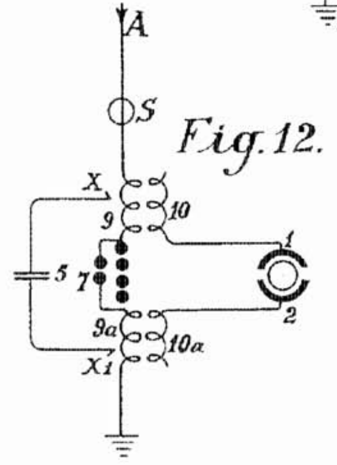
Fig. 12.
A
S
X
9
10
1
5
7
2
9a
10a
X1

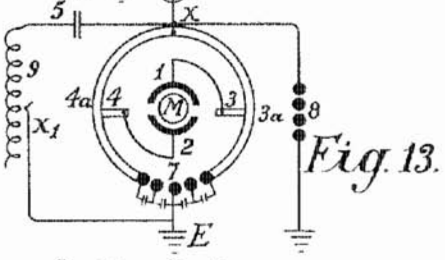
Fig. 13.
5
X
9
1
M
4a
4
3
3a
8
X1
2
7
E

Inventor

Normann Plauson

by Knight Bros

attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12, Sheets-Sheet 3

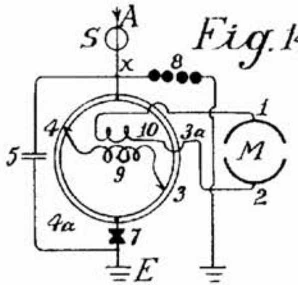
Fig. 14.
S
A
x
8
1
4
10
3a
5
9
M
3
2
4a
7
E

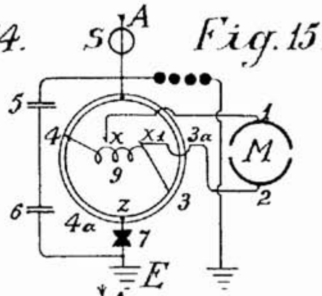
Fig. 15.
S
A
5
1
4
x
x1
3a
9
M
2
z
3
6
4a
7
E

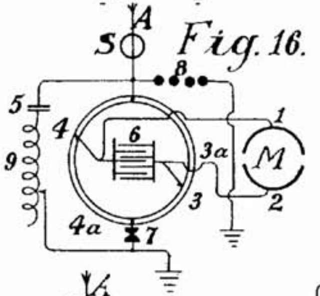
Fig. 16.
S
A
8
5
1
4
6
9
3a
M
3
2
4a
7

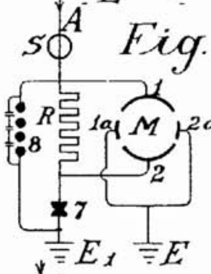
Fig. 17.
S
A
1
R
1a
M
2a
8
2
7
E1
E

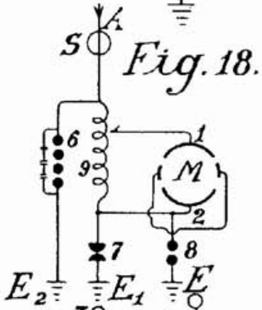
Fig. 18.
S
A
1
6
9
M
2
7
8
E2
E1
E

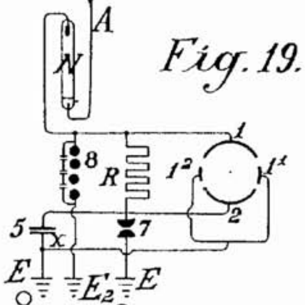
Fig. 19.
A
N
1
8
1²
1¹
R
2
5
7
x
E
E2'
E

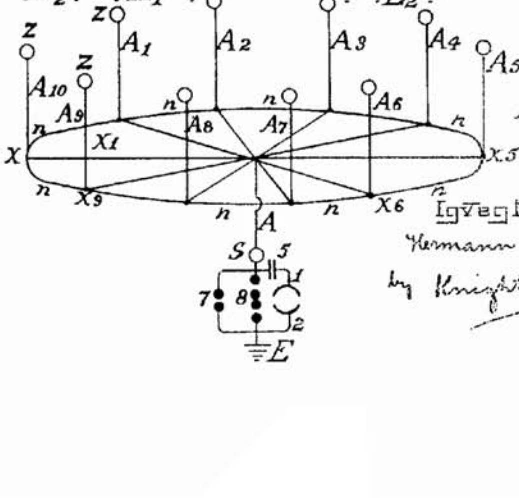
Fig. 20.
z
A1
A2
A3
A4
A5
z
z
A10
A9
n
A8
n
A7
A6
n
X
X1
X5
n
X9
n
n
X6
n
A
S
5
1
7
8
2
E

Inventor
Hermann Plauson
by Knight Bros
Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12, Sheets-Sheet 4

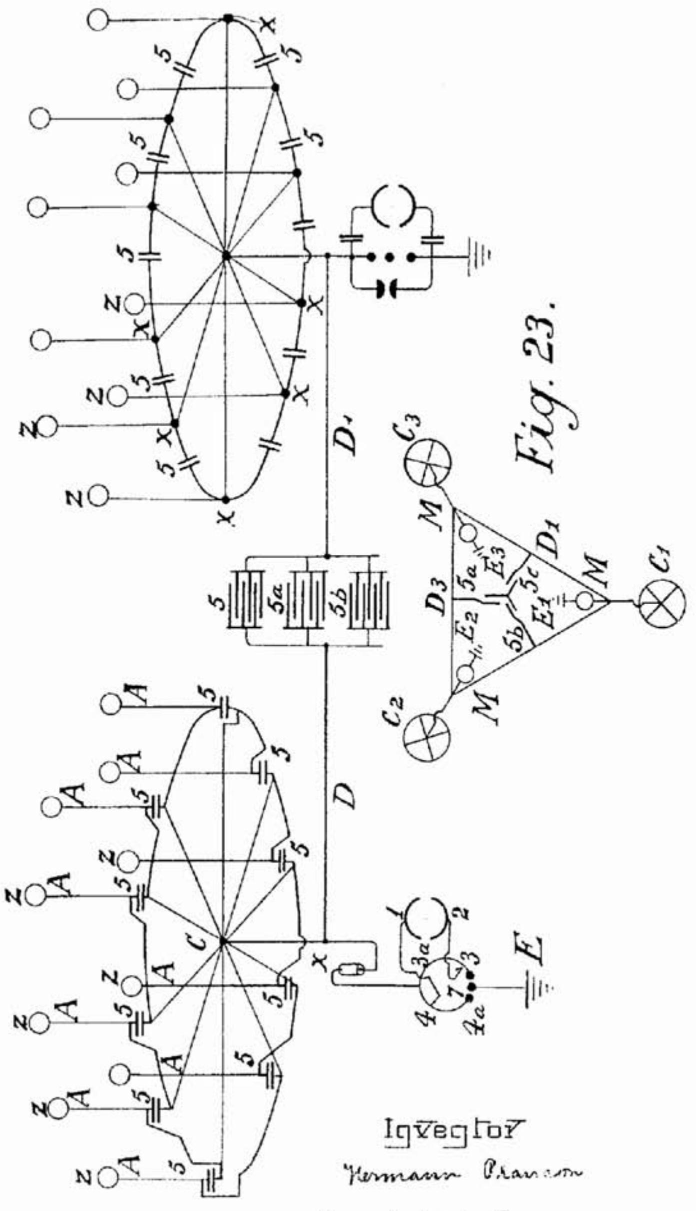
Fig. 22.

Fig. 21.

Fig. 23.

Z
A
5
5a
5b
5c
x
C
D
D'
1
2
3
3a
4
4a
7
E
C1
C2
C3
M
D1
D3
E1
E2
E3

Inventor

Hermann Plauson

by Knight Bro

Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12, Sheets-Sheet 5

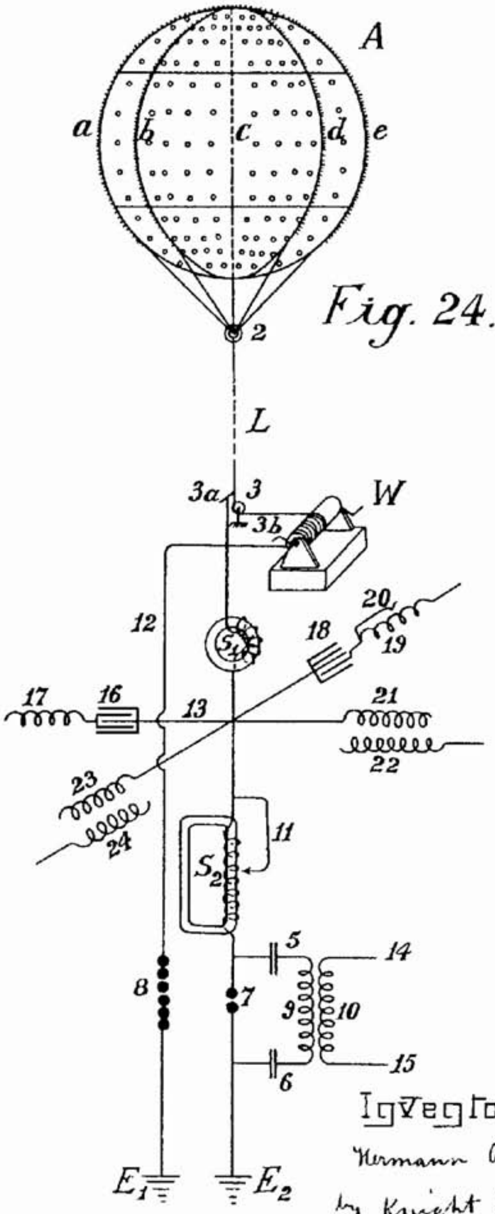
A
a
b
c
d
e
2
L
3a
3
W
3b
12
S_1
20
19
18
17
16
13
21
22
23
24
11
S_2
5
14
8
7
9
10
6
15
E_1
E_2

Fig. 24.

Inventor

Hermann Plauson

by Knight Bros

Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12, Sheets-Sheet 6

Fig. 25.

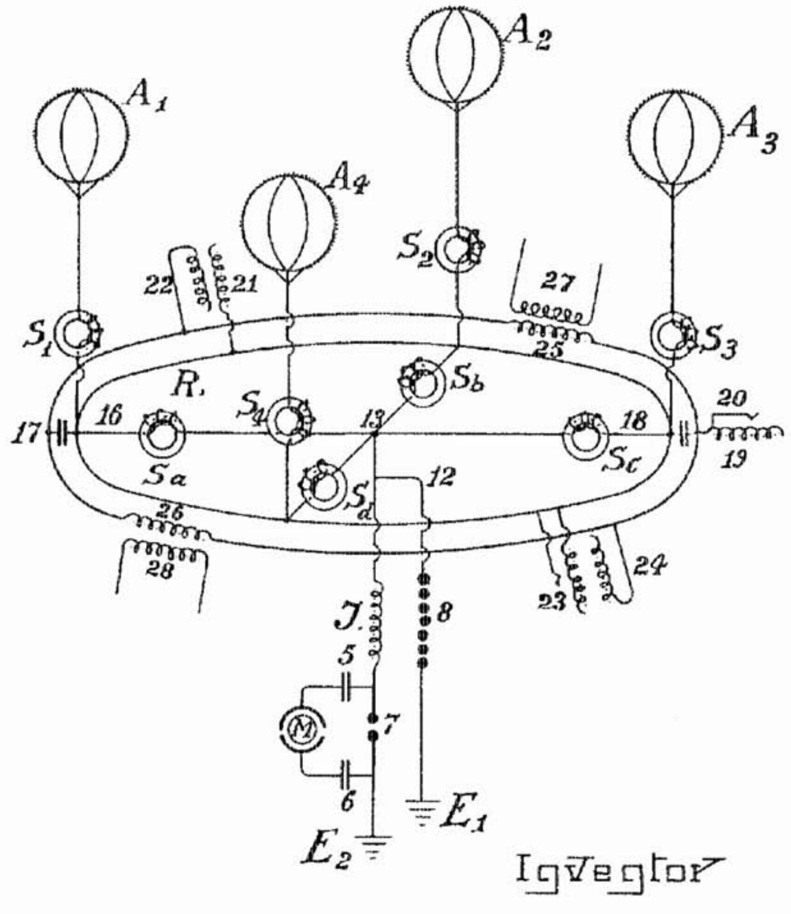
A1
A2
A3
A4
S2
27
22
21
S1
S3
25
R.
S4
Sb
13
20
17
16
18
19
Sa
Sc
12
Sd
26
28
24
23
J.
8
5
7
M
6
E1
E2

Inventor

Hermann Plauson

by Knight Bros attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12, Sheets-Sheet 7

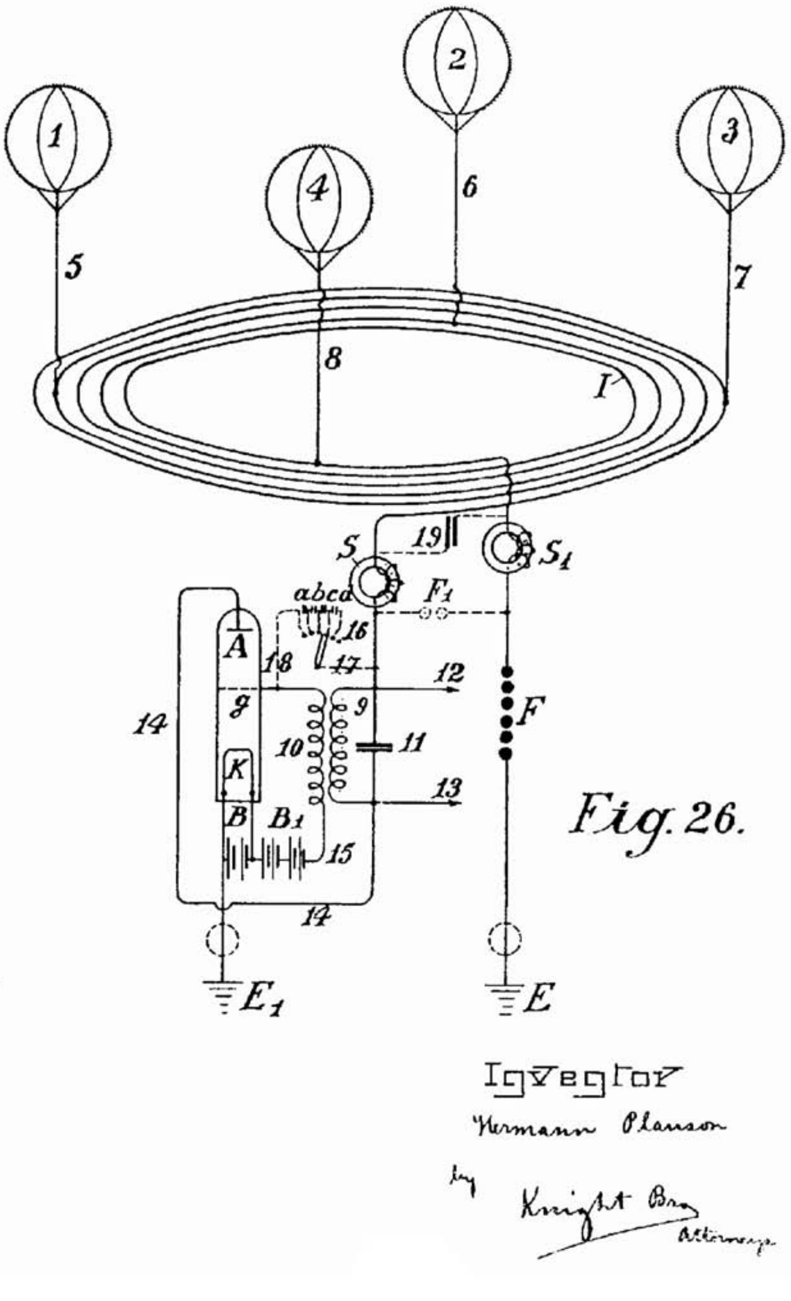
1
2
3
4
5
6
7
8
I
S
19
S1
a b c d
F1
16
A
17
18
12
g
9
14
10
11
K
13
F
B
B1
15
14
E1
E

Fig. 26.

Inventor

Hermann Plauson

by Knight Bros Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY
Filed Jan. 13, 1921 12, Sheets-Sheet 8

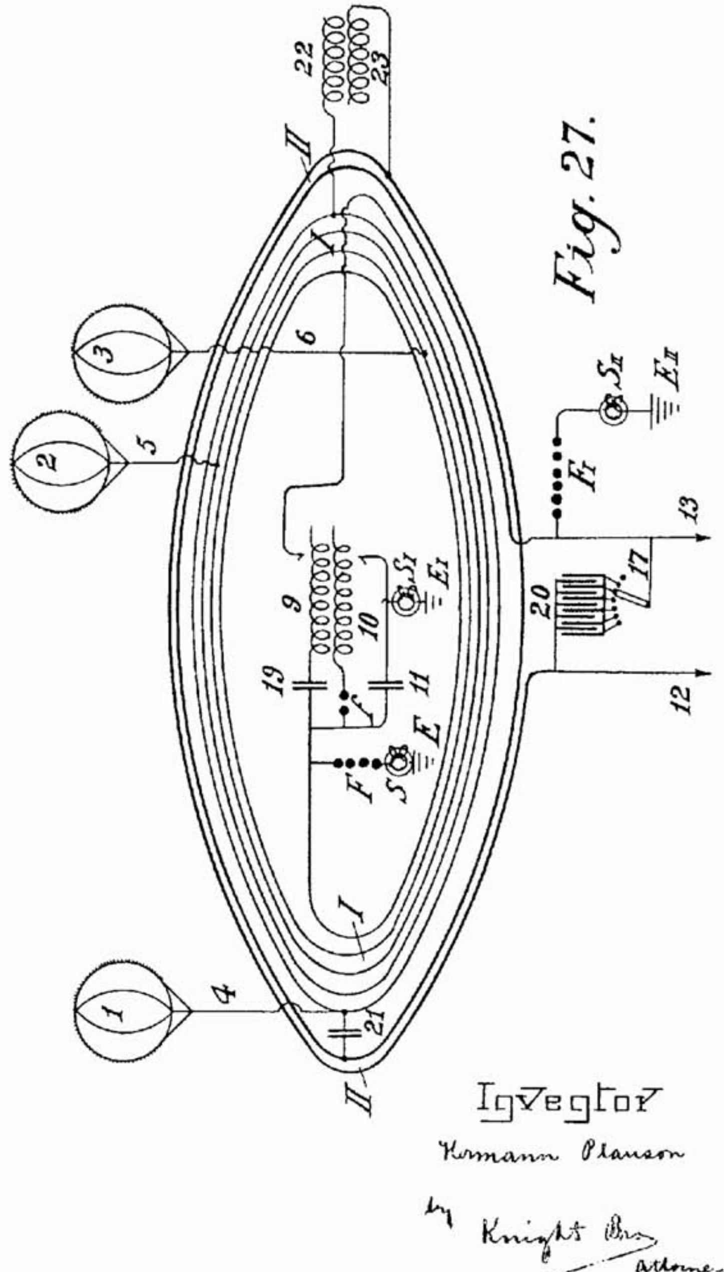
22
23
II
3
6
2
5
4
S_{II}
E_{II}
F_I
13
17
20
12
9
10
S_I
E_I
19
11
F
S
E
I
1
21
II

Fig. 27.

Inventor
Hermann Plauson

by Knight Bros
Attorneys

June 9, 1925.
1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12, Sheets-Sheet 9

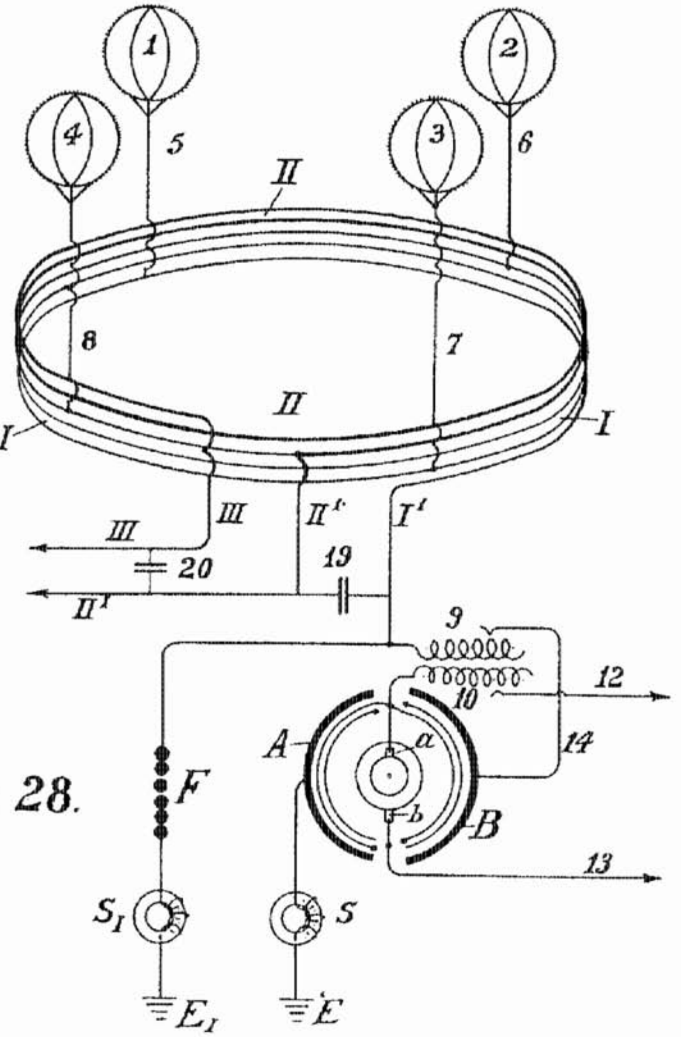
1
2
4
5
3
6
II
8
7
II
I
I
III
II'
I'
III
20
19
II'
9
12
10
A
a
14
F
b
B
13
S1
S
E1
E

Fig. 28.

Inventor
Hermann Plauson
by
Knight Bros
attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY
Filed Jan. 13, 1921 12, Sheets-Sheet 10

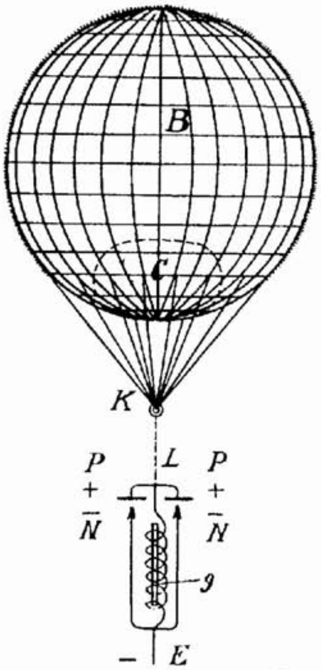
B
C
K
P
L
P
+
+
N
N
9
–
E

Fig. 29.

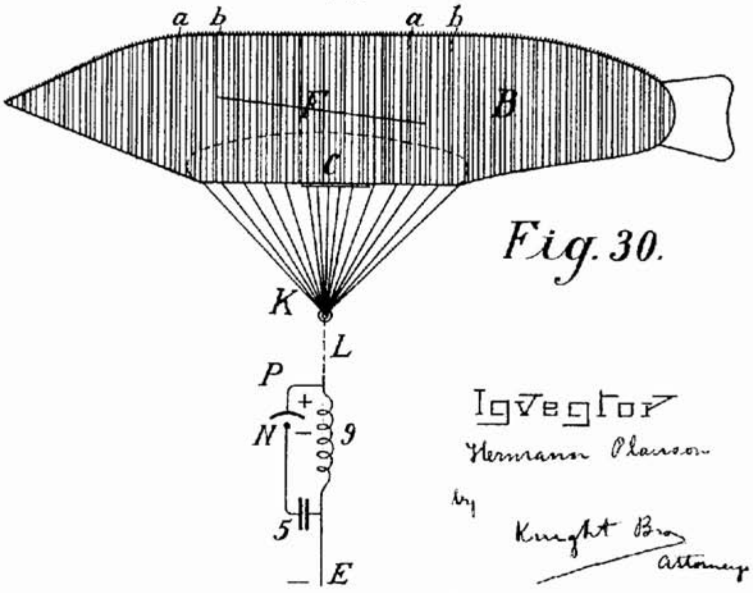
a
b
a
b
F
B
C
K
L
P
+
N
–
9
5
–
E

Fig. 30.

Inventor
Hermann Plauson
by
Knight Bros
attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12, Sheets-Sheet 11

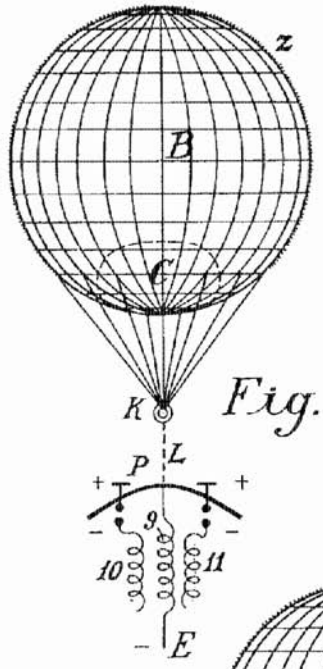
z
B
C
K
L
P
+
+
−
−
9
10
11
−
E

Fig. 31.

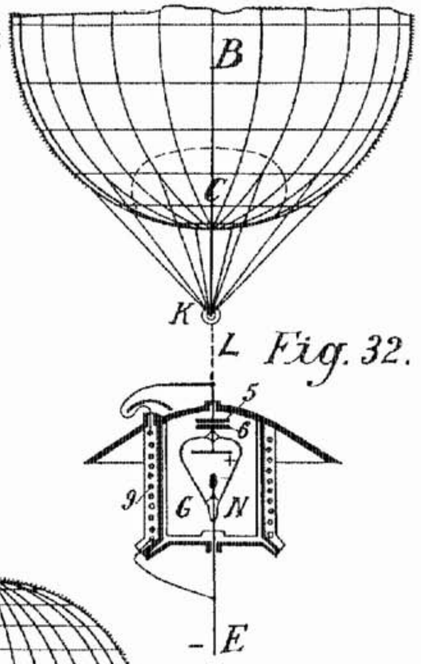
z
B
C
K
L
5
6
+
−
9
G
N
−
E

Fig. 32.

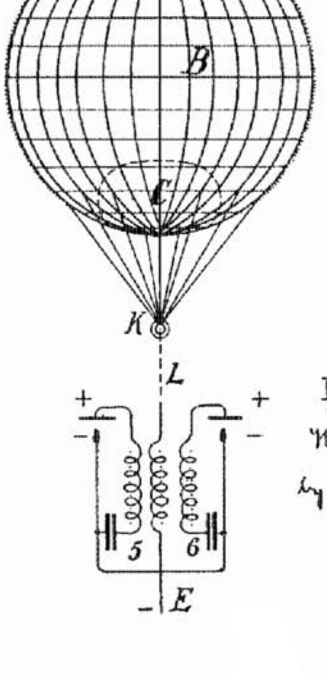

Fig. 33.

Inventor
Hermann Plauson
by Knight Bros
Attorneys

June 9, 1925. 1,540,998

H. PLAUSON

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY

Filed Jan. 13, 1921 12, Sheets-Sheet 12

Fig. 34

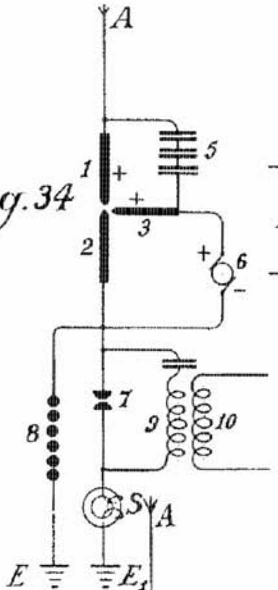
A
1
+
5
+
3
2
+
6
−
7
9
10
8
S
E
E1

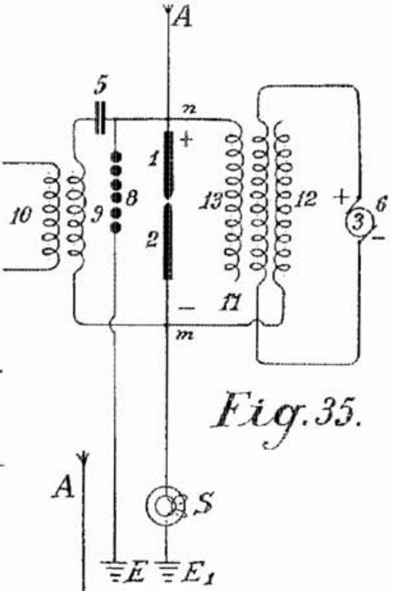
A
5
n
+
1
8
10
9
2
13
12
+
3
6
−
−
11
m
S
E
E1

Fig. 35.

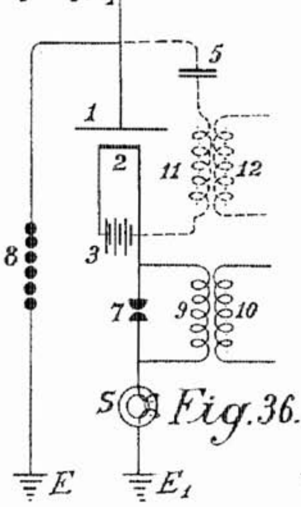
A
5
1
2
11
12
8
3
7
9
10
S
E
E1

S Fig. 36.

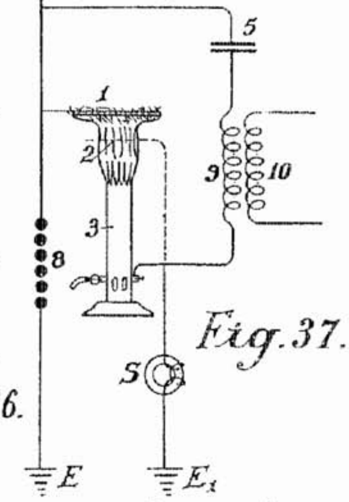
A
5
1
2
9
10
3
8
S
E
E1

Fig. 37.

Inventor

Hermann Plauson

by Knight Bros

Attorneys

Patented June 9, 1925 1,540,998

UNITED STATES PATENT OFFICE.

HERMANN PLAUSON, OF HAMBURG, GERMANY.

CONVERSION OF ATMOSPHERIC ELECTRIC ENERGY.

Application filed January 13, 1921. Serial No. 437,107

To all whom it may concern:

Be it known that I, Hermann Plauson, Estonian subject, residing in Hamburg, Germany, have invented certain new and useful improvements in the Conversion of Atmospheric Electric Energy, of which the following specification.

Methods of obtaining atmospheric electricity by means of metallic nettings set with spikes which are held by means of ordinary or anchored kite balloons made of fabrics and filled with hydrogen, are in theory already known. Atmospheric electricity obtained in this way has been suggested to be used in the form of direct current for the charging of accumulators. This knowledge however is at present only theoretical as the conversion practice has hitherto been a failure. No means are known for protecting the apparatus from destruction by lightening. The balloons used for collecting the charge must also be made of a very large size in order to be able to support the weight of the metallic netting and the heavy cable connections.

Instead of using heavy metallic netting as collectors attached to single air balloons of non-conducting material which are liable to be torn and are permeable to the gas, it is proposed to use metallic balloon collectors which have to following important advantages -

(a) The metallic cases are impenetrable to helium and hydrogen; they also represent large metallic weather-proof collecting surfaces.

(b) Radio active means and the like may be easily applied internally or externally; whereby the ionisation is considerably increased and therewith also the quantity of atmospheric electricity capable of being collected.

(c) Such balloon collectors of light metal do not require to be of large size as they have to carry only their own moderate weight, and that of the conducting cable or wire.

(d) the entire system therefore offers little surface for the action of storm and wind and is resistant and stable.

(e) Each balloon can be easily raised and lowered by means of a winch so that all repairs, recharging and the like can be carried out without danger during the operation.

It is further proposed to use a collecting aerial network of several separate collectors spread out in the air above the earth, which collectors are interconnected by electrical conductors.

According to this invention charges of atmospheric electricity are not directly converted into mechanical energy, and this forms the main differences from previous inventions, but the static electricity which runs to earth through aerial conductors in the form of direct current of very high voltage and low current strength is converted into electro-dynamic energy in the form of high frequency vibrations. Many advantages are thereby obtained and disadvantages avoided.

The very high voltage of static electricity of a low current strength can be converted by this invention to voltages more suitable for technical purposes and of greater current strength. By the use of close oscillatory circuits it is possible to obtain electromagnetic waves of various amplitude and thereby to increase the degree of resonance of such current. Such resonance allows various values of inductance to be chosen whereby again the governing of the starting and stopping of machine driven thereby by simply tuning the resonance between coils of the machine and the transformer circuit forming the resonance can easily be obtained. Further, such currents have the property of being directly available for various uses, even without employing them for driving motors, of which there may be particularly mentioned, lighting, production, of heat and use in electro-chemistry.

Further, with such current a series of apparatus may be fed without direct current supply through conductors and also the electromagnetic oscillations into mechanical energy, or finally converted by special machines into alternating current of low frequency or even into direct current of high potential.

The invention is more particularly described with reference to the accompanying diagrams in which:-

Figure 1 is an explanatory figure.

Figure 2 is a diagrammatic view of the simplest form.

Figure 3 shows a method of converting atmospheric electrical energy for use with motors.

Figure 4 is a diagram showing the use of protective means.

Figure 5 is a diagram of an arrangement for converting large current strengths.

Figure 6 is a diagram of an arrangement including controlling means.

Figure 7 shows means whereby the spark gap length can be adjusted.

Figure 8 shows a unipolar connection for the motor.

Figure 9 shows a weak coupled system suitable for use with small power motors.

Figures 10, 11 and 12 show modified arrangements.

Figure 13 shows a form of inductive coupling for the motor circuit.

Figure 14 is a modified form of Figure 13 with inductive coupling.

Figure 15 is an arrangement with non-conductive motor.

Figure 16 is an arrangement with coupling by condenser.

Figure 17, 18 and 19 are diagrams of further modifications.

Figure 20 shows a simple form in which the aerial network is combined with special collectors.

Figure 21 show diagrammatically an arrangement suitable for collecting large quantities of energy.

Figure 22 is a modified arrangement having two rings of collectors.

Figure 23 shows the connections for three rings of collectors.

Figure 24 shows a collecting balloon and diagram of its connection of condenser batteries.

Figure 25 and 26 show modified collector balloon arrangements.

Figure 27 shows a second method of connecting conductor for the balloon aerials.

Figure 28 shows an auto-transformer method of connection.

Figure 29 shows the simplest form of construction with incandescent cathode.

Figure 30 shows a form with cigar shape balloon.

Figure 31 is a modified arrangement.

Figure 32 shows a form with cathode and electrode enclosed in vacuum chamber.

Figure 33 is a modified form of Figure 32.

Figure 34 shows an arc light collector

Figure 35 shows such an arrangement for alternating current.

Figure 36 shows incandescent collector with Nernst lamp.

Figure 37 shows a form with a gas flame.

Figure 1 illustrates a simple diagram for converting static electricity into dynamic energy of a high number of oscillations. For the sake of clearness in the drawings an influence machine is assumed to be employed and not an aerial antenna. 13 and 14 are combs for collecting the static electricity of the influence machine. 7 and 8 are spark discharging electrodes, 6 and 5 condensers, 9 an inductive primary coil, 10 secondary coil, 11 and 12 ends of conductors of he secondary coil 10. When the disc of the static influence machine is rotated by mechanical means, the combs collect the electric charges one the positive and the other the negative, and charge the condensers 5 and 6 until such a high potential is formed across the spark gap 7-8, that the spark gap is jumped. As the spark gap 7-8 forms a closed circuit with condensers 6 and 5, and inductive resistance 9, as is well know, waves of high frequency electromagnetic oscillations will pass in this circuit.

The high frequency of the oscillations produced in this primary circuit induces waves of the same periodicity in the second circuit. Thus in the primary circuit electromagnetic oscillations are formed by the passage of the spark over the spark gap and these waves are maintained by fresh charges of static electricity.

By suitably selecting the ratio between the number of coils in the primary and secondary circuits with regard to a correct application of the co-efficients of resonance (capacity, inductance, and resistance) the high voltage of the primary circuit may be suitably converted into low voltage and high current strength.

When the oscillatory discharges in the primary circuit becomes weaker or entirely cease, the condensers are charged again by the static electricity until the accumulated charge again breaks down the spark gap. All this is repeated as long as electricity is produced by the static machine by employing mechanical energy.

An elementary form of the invention is shown in Figure 2 in which two spark gaps in parallel are used one of which may be termed the working gap 7 in Figure 2, whilst the second serves as a safety device for excess voltage and consists of a large number of spark gaps than the working section, which gaps are arranged in series and are bridged by very small capacities as is illustrated in a_1, b_1, c_1, Figure 2 which allow of uniform sparking in the safety section.

In Figure 2 A is the aerial antenna for collecting charges of atmospheric electricity. 13 is the earth connection of the second part of the spark gap, 5 and 6 are condensers, 9 a primary coil. Now when through the aerial A the positive atmospheric electricity seeks to combine with the negative charge to earth, this is prevented by (the air gap between) the spark gaps. The resistance of the spark gap 7 is, as shown in the drawings, lower than that of the other safety section which consists of the three spark gaps connected in

series, and consequently a three times greater air resistance is offered by the latter.

So long therefore, as the resistance of the spark gap 7 is not overloaded, so that the other spark gaps have an equal resistance with it the dischargers take place only over spark gap 7. Should however the voltage be increased by any influences so it might dangerous for charging the condensers 5 and 6 or for the coils insulation 9 and 10 in consequence of break down, by a correct regulation of this spark gap the second spark gap can discharge free from inductive effects direct to earth without endangering the machine.

Without this second spark gap, arranged in parallel having a higher resistance than the working spark gap it is impossible to collect and render available large quantities of electrical energy.

The action of this closed oscillation circuit consisting of spark gap 7, two condensers 5 and 6, primary coil 9, and also secondary coil 10 is exactly the same as the one described in Figure 1 with the arrangement of the static induction machine with the only difference that here the second spark gap is provided. The electromagnetic high frequency alternating current obtained can be tapped off from the conductors 11 and 12 for lighting and heating purposes. Special kinds of motors adapted for working with these peculiar electrical charges may be connected at 14 and 15 which can work with static electricity charges or with high frequency oscillations.

In addition to the use of spark gaps in parallel a second measure of security is also necessary for taking off the current. This precaution consists according to this invention, in the introduction of and method of connecting certain protective electromagnets or choking coils in the aerial circuit as show by S in Figure 3.

A single electromagnet only having a core of the thinnest possible separate laminations is connected with the aerial.

In the case of high voltages in the aerial network or at places where there are frequent thunderstorms, several such magnets may however be connected in series.

In the case or large units or plants several electromagnets can be employed in parallel or in series parallel.

The windings of these electromagnets maybe simply connected in series with the aerials. In this case the winding preferably consists of several thin parallel wires, which make up together the necessary section.

The winding may be made of primary and secondary windings in the form of a transformer. The primary winding will be then connected in series with the aerial network, and the secondary winding more or less short circuited over a regulating resistance or an induction coil. In the latter case it is possible to regulate to a certain extent the effect of the choking coils. In the further description of the connecting and constructional diagrams the aerial electromagnetic choke coil is indicated by a simple ring S.

Figure 3 shows the simplest way of converting atmospheric electricity into electromagnetic wave energy by the use of special motors adapted for high oscillatory currents or static charges of electrical energy. Recent improvements in motors for working with static charges and motors working by resonance, that is to say having groups of tuned electromagnetic cooperating circuits render this possible but such do not form part of the present invention.

A motor adapted to operate with static charges will for the sake of simplicity be diagrammatically indicated by two semi-circles 1 and 2 and the rotor of the motor by a ring M. (Figure 3) A is a vertical aerial or aerial network. S the safety choke or electromagnet with coil O as may be seen is connected with the aerial A. Adjacent the electromagnet S the aerial conductor is divided into three circuits, the circuit 8 giving safety spark gap, the circuit 7 with the working spark gap and then a circuit including the stator terminal 1, the rotor and stator terminal 2 at which a connection is made to the earth wire. The two spark gaps are also connected metallically with the earth wire. The method of working these diagrams are as follows:

The positive atmospheric electric charge collected tends to combine with the negative electricity (or earth electricity) connected with the earth wire. It travels along the aerial A through the electromagnet S without being checked as it flows in the same direction as the direct current. Further, its progress is arrested by two sparks gaps placed in the way and the stator condenser surfaces. The stator condenser surfaces are charged until the charge is greater than the resistance of the spark gap 7 and an oscillatory charge is obtained as by means of motor M, stator surfaces 1 and2, and spark gap 7, a closed oscillation circuit is obtained for producing the electromagnetic oscillations. The motor here form the capacity and the necessary inductance and resistance, which, as is well know, are necessary for converting static electricity into electromagnet wave energy.

The discharges formed are converted into mechanical energy in special motors and cannot reach the aerial network by reason of the electro-magnet or choke. If, however, when a spark springs over the spark gap 7 a greater quantity of atmospheric electricity tends to flow to earth, a counter voltage is induced in the electro- magnet , which is greater the more rapidly and strongly the flow of current direct to earth is. By the formation of this opposing voltage a sufficiently high resistance is offered to the flow of atmospheric electricity direct to earth to prevent a short circuit with the earth.

The circuit containing spark gap 8 having a different wave length which is not in resonance with the nature frequency of the motor, does not endanger the motor and serves as security against excess voltage, which, as practical experiments have shown, may still arise in certain cases but can be conducted direct to earth through this spark gap.

In the diagram illustrated in Figure 4 the spark gap 7 is shunted across condensers 5 and 6 from the motor M. This construction affords mainly a better insulation of the motor against excess voltage and a uniform excitation through the spark gap 7.

In Figure 5 a diagram is illustrated for transforming large current strengths which may be employed direct without motors, for example, for lighting or heating purposes. The main difference is that here the spark gap consists of a star shaped disc 7 which can rotate on its own axis and is rotated by a motor opposite similarly fitted electrodes 7^a. When separate points of stars face one another, discharging takes place, thus forming an oscillation circuit over condensers 5 and 6 and inductance 9 for oscillatory discharges. It is evident that a motor may also be directly connected to the ends of the spiral 9.

The construction of the diagram shown in Figure 6 permits of the oscillation circuit of the motor being connected with an induction coil. Here a regulating inductive resistance is introduced for counter-acting excess voltages in the motor. By cutting the separate coils 9 (coupled inductively to the aerial) in or out the inductive action on the oscillation circuit.

In Figure 7 the oscillation circuit is closed through the earth (E an E_1). Then spark gap 7 may be prolonged or shortened by more or fewer spark gaps being successively connected by means of a contact arm 7^b.

Diagram 8 shows a unipolar connection of the motor with the aerial network. Here two oscillation circuits are closed through the same motor. The first oscillation circuit passes from aerial A through electromagnet S, point x, inductance 9^a to the earth condenser 6 and further, over spark gap 7 to aerial condenser 5 and to x. The second oscillation circuit starts from the aerial condenser 5 at the point x^1 over the inductance 9 to earth condenser 6 at the point x^3 and through the condenser 6 over the spark gap 7 back to x^1. The motor itself is inserted between two points of the spark gap 7. From this arrangement slightly damped oscillation wave currents are produced.

In the diagram illustrated in Figure 9 a loosely coupled system of connections is illustrated which is assumed to be for small motors for measuring purposes. A indicates the aerial conductor, S the electromagnet in the aerial conductor, 9 the inductance, 7 the spark gap, 5 and 6 condensers, E the earth, M the motor, and 1 and 2 stator connections of the motor. The motor is directly metallically connected with the oscillation circuit.

In Figure 10 a purely inductive coupling is employed for the motor circuit. the motor is connected with the secondary wire 10 as can be seen in Figure 11 in a somewhat modified diagram connection. The same applies to the diagram of figure 12.

The diagrams hitherto described preferably allow of motors of small and medium strength to be operated. For large aggregates, however, they are too inconvenient as the construction of two or more oscillation circuits for large amounts of energy is difficult; the governing is still more difficult and the danger in switching on or off is greater.

A means of overcoming such difficulties is shown in Figure 13. The oscillation circuit here runs starting from the point x over condenser 5, variable inductance 9, spark gap 7 and the two segments (3^a and 4^a) forming arms of a Wheatstone bridge, back to x. If the motor is connected by brushes 3 and 4 transversely to the two arms of the bridge as shown in the drawings, electromagnetic oscillations of equal sign are induced in the stator surfaces 1 and 2 and the motor does not revolve. If however, the brushes 3 and 4 are moved in common with the conducting wore 1 and 2 which connect the brushes with the stator poles a certain alteration or displacement of the polarity is obtained and the motor commences to revolve.

The maximum action will result if one brush 3 comes onto the central sparking contact 7 and the other brush 4 on the part x. They are however, usually in practice not brought on to the central contact 7 but only held in the path of the bridge segments 4^a and 3^a in order not to connect the spark gaps with the motor oscillation circuit.

As however, the entire oscillation energy can thereby not act on the motor it is better to carry out the same system according to the diagram Figure 14. The diagram 14 differs from the foregoing only by the motor not being directly metallically connected with segments of the communicator, but only a primary coil 9 which induces in a secondary coil 10, current which feeds the motor M and takes the place of the rotor. By this arrangement a good transforming action is obtained, a loose coupling and also an oscillation circuit without spark gap.

In Figure 15 the motor is not purely inductive as in 14, but directly metallically branched off from the primary coil (at x and x^1) after the principle of the auto-transformer.

In Figure 16 instead of an inductance a condenser 6 is in similar manner, and for the same object inserted between segments 3^a and 4^a. This has the advantage that the segments 3^a and 4^a need not be made of solid metal but may consist of spiral coils whereby a more exact regulation is possible and further motors of high inductance may be employed.

The arrangement of Figure 17, 18 and 19 may be employed for use with resonance and particularly with induction condenser motors; between the large stator induction condenser surfaces small reversing pole condensers are connected, which, as may be seen from Figure 17, 18 and 19 are led together to earth. Such reversing poles have the advantage that with large quantities of electrical energy the spark formation between the separate oscillation circuits ceases.

Figure 19 shows a further method which prevents electromagnetic oscillations of high number of alterations formed in the oscillation circuit striking back to the aerial conductor. It is based on the well known principle that a mercury lamp, one electrode of which is formed of mercury, the other of solid metal such as steel allows an electric charge to pass in only one direction from the mercury to the steel and not visa versa. The mercury electrode of the vacuum tube N is therefore connected with the aerial conductor and the steel electrode with the oscillation circuit cannot pass to the aerial conductor.

In practice these vacuum tubes must be connected behind the electromagnet as the latter alone affords no protection against the danger of lightening.

As regards the use of spark gaps, all arrangements as used for wireless telegraphy may be used. Of course the spark gaps in large machines must have a sufficiently large surface. In very large stations they are cooled in liquid carbonic acid or better still in liquid nitrogen or hydrogen; in most cases the cooling may also take place by means of liquefied low homologues of the metal series or by means of hydrocarbons the freezing point of which lies at between -90° C and -40° C. The spark gap casing must also be insulated and be of sufficient strength to be able to resist any pressure which may arise. Any undesirable excess super-pressure which may be formed must be automatically let off. I have employed with very good results mercury electrodes which were frozen in liquid carbonic acid, the cooling being maintained during the operation from the outside through the walls.

Figure 20 is one of the simplest forms of construction of an aerial network in combination with collectors, transformers and the like illustrated diagrammatically. E is here the earth wire, 8 the safety spark gap, 7 the working spark gap, 1 and 2 the stator surfaces of the motor, 5 a condenser battery, S the protective magnet which is connected with the coil in the aerial conductor, A^1 to A^{10} aerial antenna with collecting balloons, N horizontal collecting or connecting wires from which, to the centre a number of connections run.

The actual collectors consist of metal sheaths preferably made of an aluminium magnesium alloy, and are filled with hydrogen or helium and are attached to copper plated steel wires. The size of the balloon is selected so that the actual weight of the balloon and the weight of the conducting wire is supported thereby. On the top of the balloon and the weight of the conducting wire is supported thereby. On the top of the balloon aluminium spikes, made and gilded in a special manner hereinafter described, are arranged in order to produce a conductor action. Small quantities of radium preparations, more particularly polonium-ionium or mesothorium preparations considerably increase the ionisation, and therewith the action of these collectors.

In addition to metal balloon, fabric balloons which are superficially metal coated according to Schoop's metal spraying process, may however also be employed. A metallic surface may also be produced by lacquering with metal bronzes, preferably according to Schoop's metal spraying process or lacquering with metallic bronze powders in two electrical series of widely different metals, because thereby the collecting effect is considerably increased.

Instead of the ordinary round balloons, elongated cigar shaped ones may be employed. In order also to utilise the frictional energy of the wind, patches or strips of non-conducting substances which produce electricity by friction may be attached to the metallised balloon surfaces. The wind will impart a portion of its energy in the form of frictional electricity, to the balloon asing, and thereby the collecting effect is substantially increased.

In practice however, very high towers (up to 300 metres is fully admissible) may be employed as antenna. In these towers copper tubes rise freely further above the top of the tower. A gas lamp secured against the wind is then lit at the point of the copper tube and a netting is secured to the copper tube over the flame of this lamp to form a collector. The gas is conveyed through the interior of the tube up to the summit the copper tube must be absolutely protected from moisture at the place at which it enters the tower and also rain must be prevented running down the walls of the tower which may lead to a bad catastrophe. This is done by bell shaped enlargements which expand downwards, being arranged in the tower in the form of high voltage insulators of Siamese pagodas.

Special attention must be devoted to the foundations of such towers. They must be well insulated from the ground, which may be obtained by first embedding a layer of concrete in a box form to a sufficient depth in the ground and inserting in this an asphalt lining and then glass bricks cash about 1 or 2 metres in thickness. Over this in turn there is a ferro-concrete layer in which alone the metal foot of the tube is secured. This concrete block must be at least 2 metres from the ground and be fully protected at the sides by a wooden covering, from moisture. In the lower part of the tower a or glass house for the large condenser batteries or for the motors may be constructed. In order to lead the earth connection to the ground water, a well insulated pit constructed of vitreous bricks must be provided. Several such towers are erected at equal distances apart and connected with a horizontal conductor. The horizontal connecting wires may either run directly from tower to tower or be carried on bell shaped insulators similar to those in use for high voltage conductors. The width of the network may be of any suitable size and the connection of the motors can take place at any suitable places.

In order to collect large quantities of electricity with few aerials it is well to provide the aerial conductor with batteries of condensers as shown in two methods of construction in Figures 21 and 22. In Figure 21 the batteries of condenses 5 are connected on the one hand with the aerial collectors Z by the aerial conductor A, and the other hand interconnected in series with an annular conductor from which horizontal conductors run to the connecting points C to which the earth wire is connected.

Figure 22 shows a similar arrangement. Should two such series of antenna rings be shown by a voltmeter to have a large difference of potential (for example, one in the mountains and one in the plain) or even of different polarity these differences may be compensated for by connecting sufficiently large condenser batteries (5, 5^a, 5^b) by means of Maji star conductors D and D^1, In Figure 23 a connection of three such rings of collectors to form a triangle with central condenser battery is illustrated.

The condenser batteries of such large insulations must be embedded in liquefied gases or in liquids freezing at very low temperatures. In such cases a portion of the atmospheric energy must be employed for liquefying these gases. It is also preferable to employ pressure. By this means the condenser surfaces may be de diminished, and still allow for large quantities of energy to be stored, secure against breakdown. For smaller installations the immersing of the condensers in well insulated oil or the like, suffices. Solid substances on the other hand cannot be employed as insulators.

The arrangements in the diagrams hitherto described was always such that the condenser batteries were connected with both poles directly to the aerial conductor. An improved diagram of the connections for obtaining atmospheric electricity for the condenser batteries has however, been found to be very advantageous, this arrangement consists in that they are connected by only one pole (unipolar) to the collecting network. Such a method of arrangement is very important, as by means of it a constant current and an increase of the normal working pressure or voltage is obtained. If for example a collecting balloon aerial which is allowed to rise to a height of 300 metres, shows 40,000 volts above earth voltage, in practice it has been found that the working voltage (with a withdrawal of the power according to the method hereinbefore described by means of oscillation spark gaps and the like) is only about 400 volts. If however, the capacity of the condenser surface be increased, which capacity in the above mentioned case was equal to that of the collecting surface of the balloon aerials, to double the amount, by connecting the condenser batteries with only one pole, the voltage rises under an equal withdrawal of current up to and beyond 500 volts. This can only be ascribed to the favourable action of the connection method.

In addition to this substantial improvement it has also been found preferable to insert double inductances with electromagnets. It has also been found that the useful action of such condensers can be further increased if an induction coil be connected as inductive resistance to the unconnected pole of the condenser, or still better if the condenser itself be made as an induction condenser. Such a condenser may be compared with a spring which when compressed carries in itself accumulated force, which it again gives off when released. In charging a charge with reversed sign is formed at the

other free condenser pole, and if through the spark gap a short circuit results, the accumulated energy is again given back since now new quantities of energy are induced at the condenser pole connected with the conductor network, which in fact charges with opposite signs to that at the free condenser pole. The new induced charges have of course the same sign as the collector network. The whole voltage energy in the aerial is thereby however increased. In the same space of time larger quantities of energy are accumulated than is the case without such inserted condenser batteries.

In Figures 24 and2 5 two different diagrams of connections are more exactly illustrated. Figure 24 show a collecting balloon and the diagram of the connections to earth. Figure 25 four collecting balloons and the parallel connection of the condenser batteries belonging thereto.

A is the collecting balloon made of an aluminium magnesium alloy (electron metal, magnalium) of a specific gravity of 1.8 and in thickness of plate 0.1 to 0.2 mm. Inside there are eight strong vertical ribs of T shaped section about 10 to 20 mm in height and about 3 mm in thickness with the projecting part directed inwards (indicated by a, b, c, d and so forth); they are riveted together to form a firm skeleton and are stiffened in a horizontal direction by two cross ribs. The ribs are further connected with one another internally and transversely by means of thin steel wires, whereby the balloon obtains great power of resistance and elasticity. Rolled plates of 0.1 to 0.2 mm in thickness made of magnalium alloy are then either soldered or riveted on this skeleton so that fully metallic casing with smooth external surface is obtained. Well silvered or coppered aluminium plated steel wires run form each rib to the fastening ring 2. Further, the coppered steel hawser L preferably twisted out of separate thin wires (shown in dotted lines in Figure 24) and which must be long enough to allow the balloon to rise in the desired height, leads to a metal roller or pulley 3 and from thence a winch W, well insulated from the earth. By means of this winch, the balloon, which is filed with hydrogen, or helium, can be allowed to rise to suitable height (300 to 5,000 metres) and brought to the ground for recharging or repairs.

The actual current is taken directly through a friction contact form the metal roller 3 or from the wore, or even from the winch or simultaneously form all thereby means of brush (3, 3^a and 3^b). Beyond the brushes the conductor is divided, the paths being:- firstly over 2 to the safety spark gap 8, from thence to the earth conductor E^1, and secondly over electromagnet S^1, point 13, to a second loose electromagnet having and adjustable coil S^2, then to the spark gap 7 and to the second earth conductor E^2. The actual working circuit is formed through spark gap 7, condensers 5 and 6, and through the primary coil 9; here the static electricity formed by oscillatory discharges is accumulated and converted into high frequency electromagnet oscillations. Between the electromagnets S^1 and S^2 at the crossing point 13, four condenser batteries are introduced which are only indicated diagrammatically in the drawings each by one condenser. Two of these batteries (16 and 18) are made as plate condensers and prolonged by regulating induction coils or spiral 17 and 19 while the two other (21 and 23) are induction condensers. As may be seen from the drawings each of the four condenser batteries 16, 18, 21 and 23 is connected only by one pole to the aerial or to the collector conductor. The second poles 17, 19, 22 and 24 are open. In the case of plate condensers having no inductive resistance and induction coil is inserted. The object of such a spiral or coil is the displacement of phase of the induction current by ¼ periods, whilst the charging current of the condenser poles which lie free in the air, works back to the collector aerial. The consequence of this is that in discharges in the collector aerial the back inductive action of the free pole allows a higher voltage to be maintained in the aerial collecting conductor than would otherwise be the case. It has also been found that such a back action has an extremely favourable effect on the wear of the contacts. Of course the inductive effect may be regulated at will within the limits of the size of the induction coil, the length of the coil in action being adjustable by means of wire connection without induction (see Fig. 24, No. 20).

S^1 and S^2 may also be provided with such regulating devices in the case of S^2 (illustrated by 11). If excess voltage be formed it is conducted to earth through the wire 12 and spark gap 8 or through any other suitable apparatus, since this formation would be dangerous for the other apparatus.

The action of these condenser batteries has already been hereinbefore described.

The small circles on the collector balloon indicate places at which zinc amalgam or gold amalgam or other photoelectric acting metals in the form of small patches in extremely thin layers (.01 to .05 mm in thickness) are applied to the balloon casing of light metal. Such metallic patches may also be applied to the entire balloon as well as in greater thickness to the conducting network. the capacity of the collector is thereby considerably strengthened at the surface. The greatest possible effect in collecting may be obtained by polonium amalgams and the like. On the surface of the collector balloon metal points or spikes are also fixed along the ribs, which spikes serve particularly for collecting the collector charge. Since it is well known that the resistance of the spikes is less the sharper the spike is, for this purpose it is therefore extremely important to employ as sharp spikes as possible. Experiments made as regards these have shown that the formation of the body of the spike or point also plays a large part for example spikes made of bars or rollers with smooth surfaces, have many times greater point resistance as collector accumulator spikes than those with rough surfaces. Various kinds of spike bodies have been experimental with for the collector balloons hereinbeforementioned. The best results were given by spikes which were made in the following way. Fine points made of steel, copper, nickel, or copper and nickel alloys, were fastened together in bundles and then placed as anode with the points in a suitable electrolyte (preferably in hydrochloric acid or muriate of iron solutions) and so treated with weak current at 2 and 3 volts pressure. After 2 to 3 hours according to the thickness of the spikes or pins the points become extremely sharp and the bodies of the spikes have rough surface. The bundle can then be removed and the acid washed off with water. The spikes are then placed as cathode in a bath consisting of solution of gold, platinum, iridium, paladium or wolfram salts or their compounds and coated at the cathode galvanically with a thin layer of precious metal, which must however be sufficiently firm to protect them from atmospheric oxidation.

Such spikes act at a 20 fold lower voltage almost as well as the best and finest points made by mechanical means. Still better results are obtained if polonium or radium salts are added to the galvanic bath when forming the protective layer or coating. Such pins have a low resistance at their points and even at one volt and still lower pressures have an excellent collector action.

In Figure 24 the three unconnected poles are not connected with one another in parallel. That is quite possible in practice without altering the principle of the free pole. It is also preferable to interconnect in parallel to a common collector network, a series of collecting aerials.

Figure 25 shows a diagram for such an installation. A^1, A^2, A^3, A^4 are four metal collector balloons with gold or platinum coated spikes which electromagnetically made in the presence of polonium emanations or radium salts, which spikes or needles are connected over four electromagnets S^1, S^2, S^3, S^4, through an annular conductor R. From this annular conductor four wires run over four further electromagnets S^a, S^b, S^c, S^d, to the connecting point 13. There the conductor is divided, one branch passing over 12 and the safety spark gap 8 to the earth at E^1, the other over inductive resistance J and working spark gap 7 to the earth at E^2. The working circuit, consisting of the condenser 5 and 6 and the resonance motor or condenser motor M, such as hereinbefore described, is connected in proximity round the parking gap section 7.

Instead of directly connecting the condenser motor of course the primary circuit for high frequency oscillatory current may also be inserted.

The condenser batteries are connected b one pole to the annular conductor R and can be either inductionless (16 an 18) or made as induction condensers as shown by 21 and 23. The free poles of the inductionless condensers are indicated by 17 and 19, those of the induction condensers by 22 and 24. As may be seen from the drawings all these pole 17, 22, 19, 24 may be interconnected in parallel through a second annular conductor without fear that thereby the principle of the free pole connection will be injured. In addition to the advantages already set forth the parallel connection also allows of an equalisation of the working pressure in the entire collector network. Suitably constructed and calculated induction coils 25 and 26 may also be inserted in the annular conductor of the free poles, by means of which a circuit may be formed in the secondary coils 27 and 28 which allows current produced in this annular conductor by fluctuations of the charges or the like appearances to be measured or otherwise utilised.

According to what has been hereinbefore stated separate collector balloons may be connected at equidistant stations distributed over the entire country, either connected directly with one another metallically or by means of intermediate suitably connected condenser batteries through high voltage conductors insulated form earth. The static electricity is converted through a spark gap into dynamic energy of a high number of oscillations and may in such form be coupled as a source of energy by means of a suitable method of connecting, various precautions being observed, and special regulations. The wires leading from the collector balloons have hitherto been connected through an annular conductor without this endless connection, which can be regarded as an endless induction coil, being able to exert any action on the whole conductor system.

It has now been found that if the network conductor connecting the aerial collector balloons with one another is not made as a simple annular conductor but preferably short circuited in the form of coils of a condenser battery or spark gap or through thermionic tubes or valves or audions, then the total collecting network exhibits quite new properties. The collection of atmospheric electricity is thereby not only increased but an alternating field may be easily produced in the collector network. Further, the atmospheric electrical forces showing themselves in the higher regions may also be directly obtained by induction. In Figure 26 and 28 a form of construction is shown on the basis of which the further foundation of the method will be more particularly explained.

In Figure 26 1, 2, 3, 4 are metallic collector balloons 5, 6, 7, 8 their metallic aerial conductors and I the actual collector network. This consists of five coils and is mounted on high voltage insulators in the air, on high voltage insulators in the air, on high voltage masts (or with suitable construction cable embedded in the earth). One coil has a diameter of 1 to 100 km or more. S and S' are two protective electromagnets. F the second safety section against excess voltage, E its earth conductor and E' the earth conductor of the working section. When an absorption of static atmospheric electricity is effected through the four balloon collectors, the current in order to reach the earth connection E1 must flow spirally through the collector network over the electromagnet S, primary induction coil 9, conductor 14, anode A of the audion tube, incandescent cathode K, as the way over the electromagnet and safety spark gap F offers considerably greater resistance. Owing to the fact that the accumulated current flows in one direction, an electromagnetic alternating field is produced in the interior of the collector network coil, whereby the whole free electrons are directed more or less into the interior of the coil. An increased ionisation of the atmosphere is therefore produced. In consequence of this the points mounted on the collector balloon show a considerably reduced resistance and therefore increased static charges between the points on the balloon and the surrounding atmosphere are produced. The result of this is a considerably increased collector effect.

A second effect which could not be obtained otherwise is obtained by the electromagnetic alternating field which running parallel to the earth surface, act more or less with a diminishing or increasing effect on the earth magnetic field, whereby in the case of fluctuations in the current a return induction current of reversed sign is always is always produced in the collector coil by earth magnetism. Now if, however, a constantly pulsating continuous alternating field is produced as stated in the above collector network I, an alternating current of the same periodicity is produced also in the collecting network coil. As the same alternating field is further transmitted to the aerial balloon, the resistance of its points is thereby considerably reduced, whilst the collector action is considerably increased. A further advantage is that positive electrons which collect on the metal surfaces during the conversion into dynamic current produce a so called drop potential of the collector area. As an alternating field is present, the negative ions surrounding the collector surfaces, when discharge of the collector surfaces takes place produce by the law of induction, an induction of reversed sign on the collector surface and so forth (that is to say again a positive charge). In addition to the advantages hereinbefore set forth, the construction of connecting conductors in coil form when sufficiently large diameter, allows of a utilisation of energy arising in higher regions also in the simplest way. As is well know electric discharges frequently take place at very great elevations which may be observed, such as St. Elmo's fires or northern lights. These energy quantities have not been able to be utilised up to now. By this invention all these kinds of energy, as they are of an electromagnetic nature and the directions of the axis of the collector coils stands at right angles to the earth's surface, can be more or less absorbed in the same way as a receiver in wireless telegraphy absorbs waves coming from a far distance. With a large diameter of the spiral it is possible to connect large surfaces and thereby to take up also large quantities of energy.

It is well know that large wireless stations in the summer months, and also in the tropics are very frequently unable to receive the signals in consequence of interruptions which are caused by atmospheric electricity and this takes place with vertical coils of only 40 to 100 m diameter. If on the contrary horizontal coils of 1 to 100 km diameter be employed very strong currents may be obtained through discharges which are constantly taking place in the atmosphere. Particularly in the tropics or still better in the polar regions where the northern lights are constantly present, large quantities of energy may probably be obtained in this way. A coil with several windings should act the best. In similar manner any alteration of the earth magnetism should act inductively on such a coil.

It is not at all unlikely that earthquakes and spots on the sun will also produce an induction in such collector coils of sufficient size. In similar manner this collector conductor will react on earth currents more particularly when they are near the surface of the earth or even embedded in the earth. By combining the previous kind of current collectors so far as they are adopted for improved system with the improved possibilities of obtaining current the quantities of free natural energy which are to be obtained in the form of electricity are considerably increased.

In order to produce in the improved collector coil uniform current oscillations of an undamped nature so-called audion high vacuum or thermionic tubes of suitable connection are employed instead of the previously know spark gaps (Fig. 26, Nos. 9-18). The main aerial current flows through electromagnet S (which in the case of a high number of alterations is not connected here but in the earth conductor E^1) and may be conveyed over the primary coils in the induction winding through wire 14 to the anode A of the high vacuum grid tube. Parallel with the induction resistance 9 a regulating capacity of suitable size, such as condenser 11 is inserted. In the lower part of the vacuum grid tube is arranged the incandescent filament or cathode K which is fed through a battery B. From the battery B two branches run, one to the earth conductor E^1 and the other through battery B^1 and secondary coil 10 to the grid anode g in the vacuum tube. By the method of connections shown in dotted lines, a desired voltage at the grid electrode g may also be produced through the wire 17 which is branched off from the main current conductor through switches 16 and some small condensers (a, b, c, d) connected in series, and conductor 18, without the battery B^1 being required.

The action of the entire system is somewhat as follows:-

On the connecting conductor of the aerial collector network being short circuited to earth, the condenser pole 11 is charged ans slightly damped oscillations are formed in the short circuited existing oscillation circuit formed of the condenser 11 and self inductance 9. In consequence of the coupling through coil 10, fluctuations of voltage take place in the grid circuit 15 with the same frequency, which, fluctuations in turn influence the strength of the electrode current passing through the high vacuum amplyfying tube and thus produce current fluctuations of the same frequency in the anode circuit. A permanent supply of energy to the oscillation circuits 9 and 10 consequently takes place, until a condition of balance is set up, in which the consumed oscillation energy is equal to that absorbed. Thereby constant undamped oscillations are now produced in the oscillation circuits 9-11.

For regular working of such oscillation producers high vacuum strengthening tubes are necessary and it is also necessary that the grid and anode voltages shall have a phase difference of 180° so that if the grid is negatively charged, then the anode is positively charged and vice versa. This necessary difference of phase may be obtained by most varied connections, for example, by placing the oscillation circuit in the grid circuit or by separating the oscillation circuit and inductive coupling from the anodes and the grid circuit and so forth.

A second important factor in this way of converting static atmospheric electricity into undamped oscillations is that care must be taken that the grid and anode voltages have a certain relation to one another; the latter may be obtained be altering the coupling and a suitable selection of the self induction in the grid circuit, or as shown by dotted lines 18, 17, 16 by means of a larger or smaller number of condensers of suitable size connected in series; in this case the battery B^1 may be omitted. With a suitable selection of the grid potential a glow discharge takes place between the grid g and the anode A, and accordingly at the grid there is a cathode drop and a dark space is formed. The size of this cathode drop is influenced by the ions which are emitted in the lower space in consequence of shock ionization of the incandescent cathodes K and pass through the grid in the upper space. On the other hand the number of the ions passing through the grid is dependent on the voltage between the grid and the cathode. Thus if the grid voltage undergos periodic fluctuations (as in the present case) the amount of the cathode drop at the grid fluctuates and consequently the internal resistance of the tube correspondingly fluctuates, so that when a back coupling of the feed circuit with the grid circuit takes place, the necessary means are afforded for producing undamped oscillations and of taking current, according to requirements from the collecting conductor.

The frequency of the undamped oscillations produced is with a suitably loose coupling equal to the self frequency of the oscillation circuits 9 and 10. By a suitable selection of the self induction of the coil 9 and capacity 11 it is possible to extend from frequencies which produce electromagnetic oscillations of only a few metres wave length down to the lowest practical alternating current frequency. For large installations a suitable number of frequency producing tubes in the form of the well known high vacuum transmission tubes of .5 to 2kw. in size may be connected in parallel so that in this respect no difficulty exists.

The use of such tubes for producing undamped oscillations, and also the construction and method of inserting such transmission tubes in an accumulator or dynamo circuit is known and also that such oscillation producing tubes only work well at voltages of 1,000 up to 4,000 volts, so that on the contrary their use at lower voltages is

considerably more difficult. By use of high voltage static electricity this method of producing undamped oscillations as captured with that through spark gaps must be regarded as an ideal solution particularly for small installations of outputs of from 1 to 100 kw.

By the application of safety spark gaps, with interpolation of electromagnets, not only is short circuiting avoided by also the taking up of current is regulated. Oscillation producers inserted in the above way form a constantly acting electromagnetic alternating field in the collector coil, whereby as already stated, a considerable accumulated affect takes place. The withdrawal wire is connected at 12 and 13, but current may be taken by means of a secondary coil which is firmly or movably mounted in any suitable way inside the large collector coil, i.e. in its electromagnetic alternating field, so long as the direction of its axis runs parallel with that of the main current collecting coil.

In producing undamped oscillations of a high frequency (50,000 per second and more) in the oscillation circuits 9 and 11, electromagnets S and S^1 must be inserted if the high frequency oscillations are not to penetrate the collector coil. In all other cases they are connected shortly before earthing (as in Figs 27 and 28).

Figure 27 a second method of construction of the connecting conductor I another annular conductor II is inserted parallel to the former on the high voltage mast in the air (or embedded as a cable in the earth) but both n the form of a coil. The connecting wire of the balloon aerials is indicated as a primary conductor and also as a current producing network; the other is the consumption network and is not in the unipolar connection with current producing network.

In Figure 27 the current producing network I is shown with three balloon collectors 1, 2, 3 and aerial conductors 4, 5, 6; it is short circuited through condenser 19 and inductance 9. The oscillation forming circuit consists in this diagram of spark gap f, inductance 10, and condenser 11; the earth wire E, is connected to earth over electromagnet S^1. F is safety spark gap which is also connected to earth through a second electromagnet S and E. On connecting up the condenser circuit 11 this is charged over the spark gap f whereby an oscillatory discharge formed. This discharging current act through inductance 10 on the inductively coupled secondary 9, whereby in the producing network a modification of the potential of the condenser is produced. The consequence of this is that oscillations arise in the coil shaped producer network. These oscillations induce a current in the secondary circuit II, which has a smaller number of windings an a less resistance, the voltage of which, according to the proportion of the number of windings and of the ohmic resistance, is considerably lower whilst the current strength is greater.

In order to convert the current thus obtained into current of an undamped character, and to time its wave lengths, a sufficiently large regulatable capacity 20 is inserted between the ends 12 and 13 of the secondary conductor II. Here also current may bet taken without an earth conductor, but it is advisable to inert a safety spark gap E^1 and to connect this with the earth over electromagnet S^2.

The producer network may be connected with the working network II over an inductionless condenser 21 or over an induction condenser 22, 23. In this case the secondary conductor is unipolarly connected with the energy conductor.

In Figure 28 the connecting conductor between the separate accumulator balloons is carried out according to the autotransformer principle. The collecting coil connects four aerial balloons 1, 2, 3, 4, the windings of which are not made side by side but one above the other. In Figure 28 the collector coil I is shown with a thin line, the metallically connected prolongation coils II with a thick line. Between the ends I^1 and II^1 of the energy network I a regulating capacity 19 is inserted. The wire I^1 is connected with the output wire and with the spark gap F.

As transformer of the atmospheric electricity an arrangement is employed which consists in using rotary pair of condenser in which the one stator surface B is connected with the main current, whilst the other A is connected with the earth pole. Between these pairs of short circuited condensers are caused to rotate from which the converted current can be taken by means of two collector rings and brushes, in he form of an alternating current, the frequency of which is dependent on the number of balloons and the revolutions of the rotor. As the alternating current formed in the rotor can act, in this improved method of connection described in this invention, through coils 10 on the inductance 9, an increase or diminution of the feed current in I can be obtained according to the direction of the current by back induction. Current oscillations of uniform rhythm thereby result in the coil shaped winding of the producer network.

As the ends of this conductor are short circuited through the regulatable condenser 19 these rhythms short circuited undamped oscillations in the energy conductor, the periodicity and wave lengths of which oscillations can be adjusted according to desire by alternating the capacity 19 to a given wave length and therewith also to a given frequency. These currents may also be employed in this form directly as working current through conductors II^1 and III. By inserting the condenser 20 a connection between these conductors may also be made, whereby harmonic oscillations of desired wave length are formed. By this means quite new effects as regards current distribution are obtained. The withdrawal of current can even take place without direct wire connection if, at a suitable point in the interior of the producing network (quite immaterially whether this has a diameter of 1 or 100 km) a coil tuned to these wave lengths and of the desired capacity is firmly or movably mounted in the aerial conductor in such a way that is axial direction is in parallel with tht of the collector coil. In this case a current is induced in the producing network, the size of which is dependent on the total capacity and resistance and also on the periodicity employed. A possibility is thereby afforded in future, of taking energy form the producer network by wireless means. As thereby in addition to atmospheric electricity also magnetic earth currents and the energy from the higher atmosphere (at least partially) may be simultaneously obtained, this last system for collecting the atmospheric energy is of particular importance for the future.

Of course everywhere instead of spark gaps suitable grid vacuum tubes may be employed as producers for undamped oscillations. The separate coils of the producer network with large diameters may be connected with one another through separate conductors all in parallel or all in series or in groups in series. By regulating the number of oscillations and also the extent of the voltage more or less large collector coils of this kind may be employed. The coils may also be divided spirally over the entire section. The coils may be carried out in annular form or also in triangular, quadrangular, hexagonal or octagonal form.

Of course wires may be carried from a suitable place to the centre or also laterally which serve the currents waves as guides. This is necessary when the currents have to be conducted over mountains and valleys and so forth. In all these cases the current must be converted into a current of suitable periodicity.

As already hereinbefore mentioned separate collecting balloons may be directly metallically interconnected at equidistant stations distributed over the entire country or may be connected by interpolation of suitable condenser batteries by means of high voltage conductors. The static electricity is converted through a spark gap into dynamic energy of a high number of oscillations, and could then in such form, with a suitable arrangement of the connections, observing various measures of precaution, be employed as source of energy after separate or special regulation

According to this invention in order to increase the collecting effect of the balloon in the aerial collector conductor or in the earth wire, radiating collectors are employed. These consist either of incandescent metal or oxide electrodes in the form of vacuum grid tubes, or electric arcs (mercury and the like electrodes) Nernst lamps or finally flames of various kinds may be simply connected with the respective conductor.

It is well known that energy can be drawn off from a cathode consisting of an incandescent body opposite an anode charged with positive electricity (vacuum grid tube). Hitherto however, a cathode was always first directly placed opposite an anode, and secondly the system always consisted of closed circuit.

Now if we dispense with the ordinary ideas in forming light or flame arcs in which a cathode must always stand directly opposite an anode, and if we place an incandescent cathode opposite an anode charged to a high potential or another body freely floating in the air, or regard the incandescent cathode only as a source of unipolar discharge (which represent group and point discharges in electrostatic machines similar to unipolar discharges), it may be ascertained that incandescent cathodes and less perfectly all incandescent radiators, flames and the like admit of relatively large current densities and allow large quantities of electric energy to radiate into the open space in the form of electron streams as transmitters.

The object of this invention is as described below, if such incandescent radiators or flames are not freely suspended in space but connected metallically with the earth so that they can be charged with negative terrestrial electricity, these radiators possess the property of absorbing the free positive electrical charges contained in the air space surrounding them (that is to say of collecting them and conducting them to earth). They can therefore, serve as collectors and have, in comparison to the action of the spikes, or points, a very large radius of action R; the effective capacity of these collectors is much greater than the geometrical capacity (R_0-) calculated in an electrostatic sense.

Now as our earth is surrounded as is well known, with an electrostatic field and the difference of potential

$$\frac{\delta V}{\delta h}$$

of the earth field according to the latest investigations, is in summer about 60 to 100 volts and in winter 300 to 500 volts per metre of difference in height (δh), a simple calculation gives the result that when such a radiation collector or flame collector is arranged for example on the ground, and a second one is mounted vertically over it at a distance of 2,000 metres and both are connected by a conducting cable, there is a difference of potential in summer of about 2,000,000 volts and in winter 6,000,000 volt or more.

According to Stefan Boltzmann's law of radiation, the quantity of energy which an incandescent surface (temperature T) of 1 sq cm radiates in a unit of time into the open air (Temperature T_O) is expressed by the following formula:

$$S = \delta\,(T^4 - T^4_O) \text{ watt/sq cm}$$

and the universal radiation constant 6 is according to the latest researches of Ferry (Annales de Chimie et de physique 17 page 267 (1909)) equals to 6.30×10^{-22} watt/sq cm.

Now if an incandescent surface of 1 sq cm shows, as compared with the surrounding space a periodic fall of potential δV it radiates (independent of the current direction, that is to say of the sign) in accordance with the above formula, for example at a temperature of 3725° C an energy of 1.6 kw per sq cm per second. As for the radiation the same value can be calculated for the collection of energy, but reversed. Now as carbon electrodes at the temperature at the temperature of the electric arc support on the current basis density up to from 60 to 65 amperes per sq cm no difficulties will result in the direction in employing radiating collectors as accumulators.

If the earth be regarded as a cosmically insulated condenser in the sense of geometrical electrostatics x there results from the geometric (compare Ewald Rasch, "das elcktrische Bogenlicht" (The electric arc light) page169) capacity of the earth according to Chwolson:

For negative charging 1.3×10^5 Coulomb. For negative potential $V = 10 \times 10^8$ volts. From this there results however, $EJT \cong 24.7 10^{24}$ watt/Sec. Now if it is desired to make a theoretic short circuit through an earthed flame collector this would represent an electric total work of about $79{,}500\ 10^{10}$ kilowatt years as the earth must be regarded as a rotating mechanism which is thermo-dynamically, electromagnetically, and also kinematically coupled with the sun and stars by cosmic radiations and gravitation a diminution of the electric energy of the earth field is not to be feared. The energies which the incandescent collectors would withdraw from the earth field can only cause by the withdrawal of motor work a lowering of the earth temperature (temperature T_E=300) and reduce this to that of the world space (T=O) by using the entire energy. This is however not the case as the earth does not represent a cosmically entirely insulated system. On the contrary there is conveyed to the same according to the recent value corrected by Ferry for the solar constants through the radiation from the sun an energy of $18{,}500 \times 10^{10}$ kw. Accordingly any lowering of the earth temperature (T_E) without simultaneous lowering of the sun's temperature (T_S) would contradict Stefan Boltzmann's law of radiation.

$$S = \delta\,(T_S^4 - T_E^4).$$

From this it must be concluded that if the earth temperature (T_E) sinks, the total radiation S absorbed by the earth increases, and further also that the secular speed of cooling of the earth is directly dependant on that of the sun and the other radiators cosmically coupled with the sun and is connected most closely with these.

The incandescent radiation collectors may, according to this invention, be employed for collecting atmospheric electricity if they (1) are charged with the negative earth electricity (that is to say when they are directly connected by means of a metallic conductor with earth) and (2) if large capacities (metal surfaces) charged with electricity are mounted opposite them as positive poles in the air. This is regarded as the main feature of the present invention as without these inventive ideas it would not be possible to collect with an incandescent collector, sufficiently large quantities of the electrical charges contained in the atmosphere as technology requires; the radius action of the flame collectors would also be too small, especially if it be considered that the very small surface density (energy density) (δ about=$2 \times 7 \,.\, 10^9$ St. E per sq cm) does not allow of large quantities of charge being absorbed form the atmosphere.

x) Calculated according to Poisson;s calculation;

$\Delta V = -4\pi\delta$; as here the alteration of potential or potential gradients only takes place in the direction of the normal, this calculation assumes the simple form

$$\delta \equiv \frac{1}{4\pi} \times \frac{\delta^2 V}{\delta n^3}$$

It has indeed already been proposed to employ flame collectors for collecting atmospheric electricity and it is known that

n collecting effect is substantially greater opposite the points. It is however, not known that the quantities of current which could hitherto be obtained are too small for technical purposes. According to my experiments the reason for this is to be found in the too small capacities of the collector conductor poles. If such flame or radiating collectors have no or only small positive surfaces, their radius of action for large technical purposes is too small. If the incandescent collectors be constantly kept in movement in the air they may collect more according to the speed of the movement, but this again not capable of being carried out in practice.

By this invention the collector effect is considerably increased by a body charged with a positive potential and of the best possible capacity being also held floating (without direct earth connection) opposite such an incandescent collector which is held floating in the air at a desired height. if for example, a collecting balloon of sheet metal or metallized balloon fabric be caused to mount to 300 up to 3000 metres in the air and as positive pole it is brought opposite such a radiating collector connected by a conductor to earth, quite different results are obtained.

The metallic balloon shell (with a large surface) is charged to a high potential by the atmospheric electricity. This potential is greater the higher the collecting balloon is above the incandescent collector. The positive electricity acts concentratedly on the anode floating in the air as it is attracted through the radiation shock ionization, proceeding from the incandescent cathode. The consequence of this is that the radius of action of the incandescent cathode collector is considerably increased and thereby also the collecting effect of the collecting balloon surface. Further the large capacity of the anode floating in the air plays therefore an important part because it allows of the taking of large charges, and thereby a more uniform current is obtained even when there is a large consumption; this cannot be the case with small surfaces.

In the present case the metallic collecting balloon is a positive anode floating in the air and the end of the earth conductor of this balloon serves as positive pole surface opposite the surface of the radiating incandescent cathode, which in turn is charged with negative earth electricity being conductingly connected to earth

The process may be carried out by two such contacts (negative incandescent cathode and anode end of a capacity floating in the air) a condenser and an inductive resistance being switched on in parallel, whereby simultaneously undamped oscillations may be formed.

In very large installations it is advisable to connect two such radiating collectors in series. Thus an are light incandescent cathode may be placed below on the open ground and an incandescent cathode which is heated be special electromagnetic currents be located high in the air. Of course for this the special vacuum Liebig tubes with or without grids may also be employed. An ordinary lamp with oxide electrodes may be introduced on the ground and the positive pole is not directly connected with the collecting balloon, but through the upper incandescent cathode or over a condenser. The method of connecting the incandescent cathode floating in the air may be seen in Fis. 29-33.

B is the air balloon, K a Cardan ring (connection with the hawser) C the balloon L a good conducting cable. P a positive pole, N negative incandescent cathode, and E earth conductor.

Fig. 29 represents the simplest form of construction. If electric oscillations are produced below on the ground by means of a carbon lamp or in other suitable way a considerably greater electric resistance is opposed to that in the direct way by inserting an electrical inductive resistance 9. Consequently between P and N a voltage is formed and as, over N and P only an inductionless ohmic resistance is present, a spark will spring over so long as the separate induction co-efficients and the like are correctly calculated. The consequence of this is that the oxide electrode (carbon or the like) is rendered incandescent and then shows as incandescent cathode an increased collecting effect. The positive poles must be substantially larger than the negative in order that they may not also become incandescent. As they are further connected with the large balloon area which has a large capacity and is charged at high voltage, an incandescent body which is held floating in the air and a positive pole which can collect large capacities is thereby obtained in the simplest way. The incandescent cathode is first caused to become incandescent by means of separate energy produced on the earth, and then maintained by the energy collected from the atmosphere.

Fig. 30 only shows the difference that instead of n round balloon a cigar shaped one (of metal or metalized fabric) may be employed and also a condenser 5 is inserted between the incandescent cathode and the earth conductor so that a short circuited oscillation circuit P. N 5 and 9 is obtained. This has the advantage that quite small quantities of electricity cause the cathode to become incandescent and much larger cathode bodies may be rendered incandescent.

In this form of construction both the in candescent cathode and also the positive electrode may be enclosed in a vacuum chamber as may be seen in Fig 32. A cable L is carried well insulated through the cover of a vessel and ends in a condenser disc 5. The cover is arched in order to keep off the rain. The vessel is entirely or partially made of magnetic metal and well insulated inside and outside. Opposite the metallic positive pole of the vacuum tube g with the incandescent cathode (oxide electrode) N is arranged. The negative electrode is on the one hand connected with the earth conductor E, and on the other hand with the inductive resistance 9 which is also connected with the cable L with the positive pole and wound round the vessel in coils. The action is exactly the same as that in Fig. 29 only instead of an open incandescent cathode one enclosed on vacuo is employed. As in such collectors only small bodies can be brought to incandescence in large installations a plurality of such vacuum tubes must be inserted in proximity to one another. According to the previous constructions Figs. 31 and 33 are quite self evident without further explanations

Figs 34-37 represent further diagrams of connections over radiating the flame collectors, and in fact, how they are to be arranged on the ground.

Fig. 34 shows n arc light collector with oxide electrodes for direct current and its connection; Fig. 35 a similar one for alternating current, Fig. 36 an incandescent collector with a Nernst lamp and Fig. 37 a similar one with a gas flame.

The positive pole 1 of the radiating collectors is always directly connected to the aerial collecting conductor A. In Fig. 34 this is further connected over the condenser battery 5 with a second positive electrode 3. The direct current dynamo b produces current which flows over between the electrodes 3 and 2 as an arc light. On the formation of arc the negative incandescent electrode 2 absorbs electricity from the positive poles standing opposite it and highly charged with atmospheric electricity and conveys the same to the working circuit. The spark gap 7, inductive resistance 9 and induction coil 10 are like the ones previously described. The protective electromagnet S guards the installation against earth circuiting, the safely spark gap 8 from excess voltage and overcharging.

In Fig. 35 the connection is so far altered that the alternating current dynamo feeds the exciting coil 11 of the induction condenser. 12 is its negative and 13 its positive pole; if the coil 3 on the magnet core of the dynamo is correctly calculated and the periodicity of the alternating current is sufficiently high an arc light can be formed between the two pole 1 and 2. As the cathode 2 is connected with the negatively charged earth, and therefore always acts as a negative pole, a form of rectification of the alternating current produced by the dynamo 3 is obtained, the second half of the period is always suppressed. The working circuit may be carried out in the same way as Fig. 34; the working spark gap 7 may however be dispensed with, and instead thereof between the points n and m a condenser 5 and a induction resistance 9 may be inserted from which the current is taken inductively.

Fig. 36 represents a form of construction similar to Fig. 34 only that here instead of an arc lamp a Nernst incandescent body is employed. The Nernst lamp is fed through the battery 3. the working section is connected with the negative pole, the safety spark gap with the + poles. The working spark gap 7 may also be dispensed with the current for it taken at 12 over the oscillation circuit 5, 11 (shown in dotted lines).

Flame collectors (Fig. 37) may also be employed according to this invention. The wire network 1 is connected with the aerial collector conductor A and the burner with the earth. At the upper end of the latter, long points are provided which project the flame. The positive electrode is connected with the negative over a condenser 5 and the induction coil 9 with the earth.

The novelty in this invention is firstly the use of incandescent cathodes opposite positive poles which are connected with large metallic capacities as automatic collecting surfaces, (2) the connection to the incandescent cathodes with the earth whereby, in addition to the electricity conveyed to them from the battery or machine which causes the incandescing, also the negative charge of the earth potential is conveyed, and (3) the connection of the positive and negative poles of the radiating collectors over a condenser circuit alone or with the introduction of a suitable inductive resistance, whereby simultaneously an oscillatory oscillation circuit by obtained. The collecting effect is by these methods quite considerably increased.

I declare that what I claim is:-

1. An electrical generating system, comprising a conducting surface for static charges, means to support same at a distance above the earth, a conductor leading to the earth level, n spark gap associated with said conductor to convert electrostatic charges into electromagnetic high frequency oscillations means to supply said electromagnetic energy to a net work, and a spark gap of greatly increased relative resistance in parallel therewith.

2. An electrical energy generating system

comprising a conductor, means to support same above the earth level, an inductance therein, a spark gap associated with said conductor, a second spark gap of much higher relative resistance in parallel therewith and an energy receiving circuit coupled with the spark gap of lesser resistance.

3. An electrical generating system comprising a collecting surface, means to support same above the earth level, a conductor connecting said connecting surfaces with the earth level, a choke in said conductor, an electromagnetic resistance converting electrostatic energy to electromagnetic energy, a safety higher resistance in parallel therewith and a net work coupled with the conversion resistance of lesser value,

4. An electrical energy generating system comprising electric conductors spaced above the earth to form electromagnetic oscillating circuits, conductors connecting to earth level, electrostatic to electromagnetic energy conversion means therein, a safety high electrostatic resistance in parallel therewith and means to alter the electromagnetic characteristics of he circuits

5. An electrical energy generating system comprising in combination a static collecting surface arranged above the earth, conductors connecting to earth level, a pair of spark gaps in parallel of different electrostatic resistance, a utilization net work shunted across the spark gap of lesser resistance and an electromagnetic choke in said conductors.

6. An electrical energy generating system comprising an open circuit energy collecting aerial, a pair sparking gaps in parallel of widely different resistance, connected thereto and a closed electric oscillation circuit in shunt across the gap of lesser resistance.

7. An electrical energy generating system comprising an open circuit energy collecting aerial, a pair of sparking gaps in parallel of widely different resistance connected thereto, a closed electric oscillation circuit in shunt across the gap of lesser resistance, a plurality of electrostatic collecting surfaces, means to connect said collecting surfaces in parallel in groups and means to connect said groups symmetrically with said aerial.

In witness whereof, I have hereunto signed my name this 30 day of Dec., 1920, in the presence to two subscribing witnesses.

HERMAN PLAUSON.

Witnesses:
H.F. Armstrong
W.H. Beeston

Poor Man's Electronics Web Page

Crystal Radio Project Links:

- Back to main page
- How to build it
- Using the radio
- How it works
- Computer program

Crystal radio

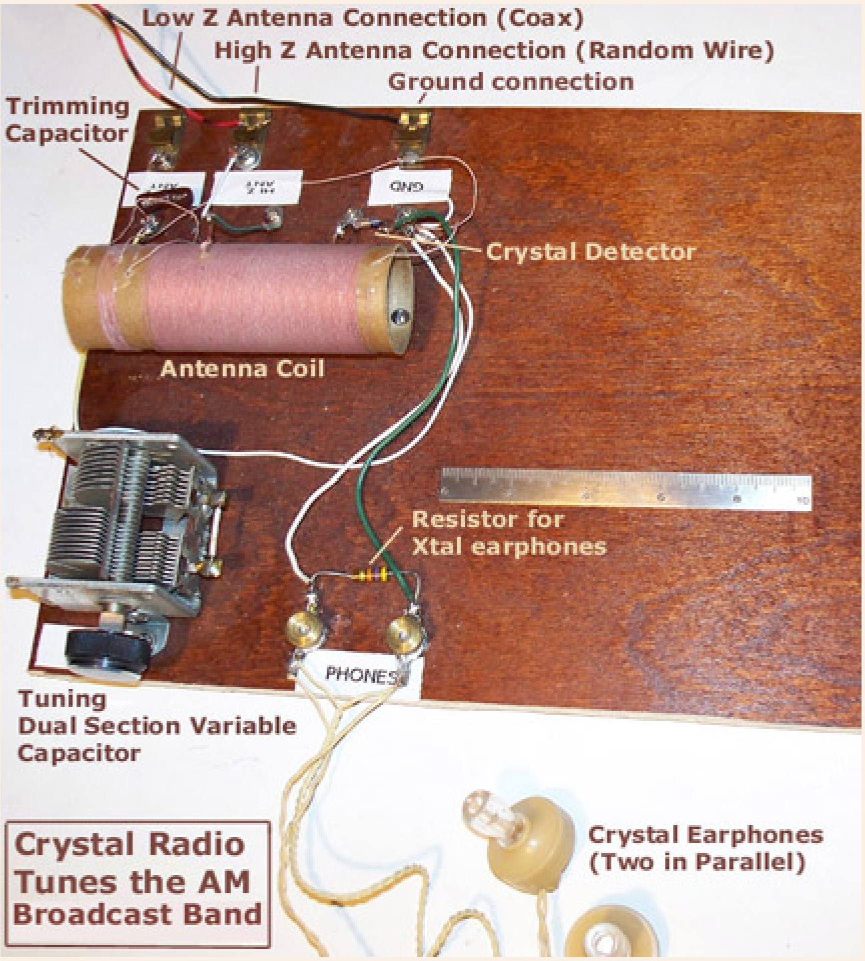

Low Z Antenna Connection (Coax)
High Z Antenna Connection (Random Wire)
Ground connection
Trimming Capacitor
LoZ ANT
Hi Z ANT
GND
Crystal Detector
Antenna Coil
Resistor for Xtal earphones
PHONES
Tuning Dual Section Variable Capacitor
Crystal Radio Tunes the AM Broadcast Band
Crystal Earphones (Two in Parallel)

Build this Crystal Radio

A crystal radio is a radio that does not require a power source other than the energy contained in the signal being received. With a good antenna, like the random wire you can build by following the instructions on this web site, strong local stations will be very discernable during the day and at night even a few distant "skip" stations may be picked up.

Here's how to make the crystal radio:

First, you must gather the material. Here is a list of what you will need including a possible *poor man's* source: (If you can't locate a PME source, then try Radio Shack, E-Bay, or Mouser Electronics.)

- A scrap piece of wood about 8 inches square to build the radio on. I used a piece of 1/4" stained panel.
- Sensitive headphones or earphones. The best is the relative high Z (2000 - 4000 ohm) military style headphones. This type is hard to find and may be expensive, so you may use the very inexpensive ceramic earphones instead. Walkman type headphones or earphones will **not** work. I use two of the ceramic types connected in parallel, one for each ear, for increased volume.
- Variable capacitor (or condenser, which is another name for a capacitor.) You can find this device in an old tube or transistor radios. The air dielectric type is best for this project. That means the plates of the capacitor are separated by an air gap. The small variable capacitors found in pocket-sized transistor radios usually have a plastic dielectric. This type will work OK. Most of the variable capacitors found in radios will be dual section; that is, two capacitors on the same shaft. We will tie these two in parallel for increased capacitance -- shooting for about 450pF (picoFarads.) Find a nice knob that will fit on the shaft of the variable capacitor.
- Detector diode. A 1N34 or 1N60 will work great. You can find them in most transistor radios. They look like a small glass resistor. If you have to buy one, they are cheap -- less than a dollar apiece.

1N34 Diode.

- Terminations. You may use Fahnstock clips (I used these on the antenna and ground connections) or binding posts (I used these to connect the earphones.) Binding posts can be salvaged from an old telephone cell. In a pinch, use small double-nutted machine screws.

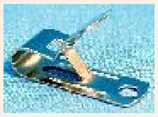
Fahnstock clip.

- Coil form. I used a 1-1/4" x 3-1/2" cardboard tube that was the center spool in a printing calculator paper roll. If you want to use a different diameter form, then use the computer program that you can download and run. It will calulate the number of turns of wire required based on varying coil form diameters and lengths.
- You will need about 50 feet of magnet wire. I used some Litz wire I had. It is 40/44 Litz which means 40 strands of #44 size. This turns out to be about #24 US Standard solid wire size. Litz wire is great for radio frequency coils. You can find Litz wire in CRT monitor or TV picture tube deflection coils. If you can't find a poor man's electronic free source, then this is one of the supplies that a PME guy should buy. You can purchase Litz wire at the Hobby Electronic Parts page. However, if you use solid enameled wire, the radio will work OK. Solid enameled type of wire can be salvaged from old relays, transformers, chokes, and solenoids.

The following pictorial diagram shows how to hook up the simple crystal radio. I provided connections for both a high impedance antenna such as the random wire or a low impedance antenna such as a coax fed dipole.

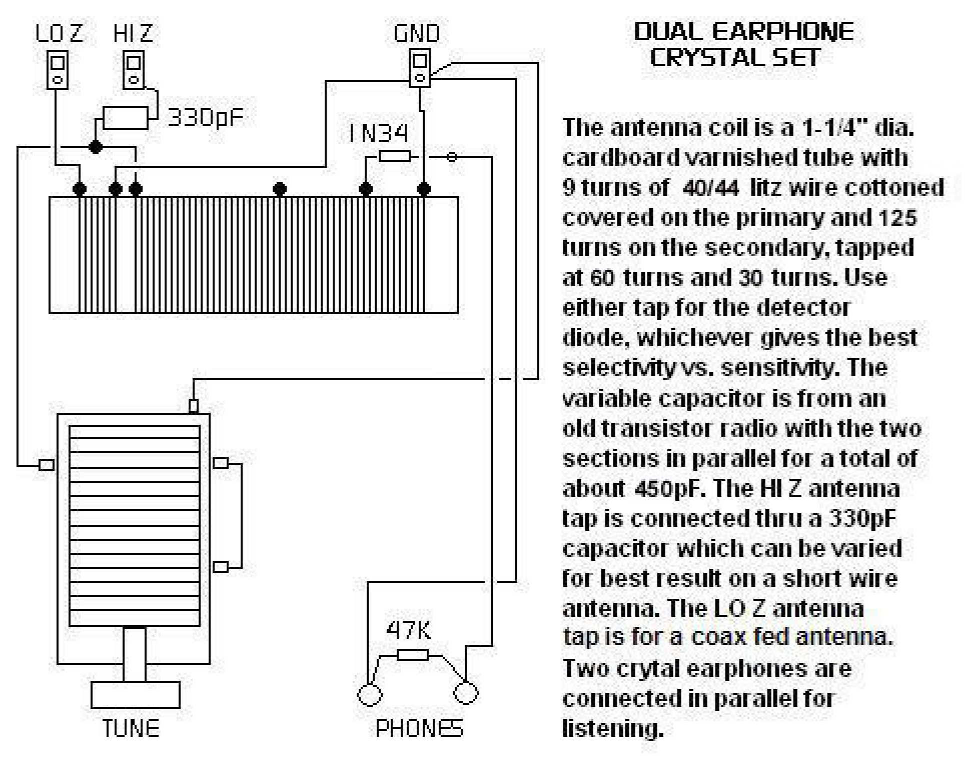

DUAL EARPHONE CRYSTAL SET

The antenna coil is a 1-1/4" dia. cardboard varnished tube with 9 turns of 40/44 litz wire cottoned covered on the primary and 125 turns on the secondary, tapped at 60 turns and 30 turns. Use either tap for the detector diode, whichever gives the best selectivity vs. sensitivity. The variable capacitor is from an old transistor radio with the two sections in parallel for a total of about 450pF. The HI Z antenna tap is connected thru a 330pF capacitor which can be varied for best result on a short wire antenna. The LO Z antenna tap is for a coax fed antenna. Two crytal earphones are connected in parallel for listening.

The step by step procedure follows:

- Wind the coil first. Coat the 1-1/4" x 3-1/2" form with 3 layers of spray polyurathane varnish. Let each coat dry before applying the next one. Punch a pair of holes using a small pin in the side of the coil form about 3/8" from the edge. Feed the #24 solid enamel wire or 40/44 Litz wire through these two holes in such a way as to lock the wire in place. Close wind 9 turns for the primary coil and finish by locking the wire into another pair of holes just like the first pair. Both ends of the wire should be well secured and have 3 inches of lead wire.
- Lock the start of the secondary coil in place as you did before about 3/8" from the end of the primary coil. Close wind a total of 125 turns. At 30 turns and 60 turns, put a twist in the wire that sticks out about a 1/4"; these will be coil taps. Finish off the secondary as you did the primary. You should be about 3/8" from the end of the form when done and again have 3 inches of lead wire on the each end.
- Once again, spray the finished coils with three coats of polyurathane varnish. Put a drop of Elmer's glue on each of the ends of the coils and around the taps to provide additional securing.
- Mount the completed antenna coil assembly, variable capacitor, and terminals as per the drawing above. You will

mount the coil by drilling holes in the board and in the coil form into which to feed some small 6-32 machine screws. (Make sure that when you are drilling holes that you do so on the ends where there is no wire.) I used two small micarta bushings between the coil form and the board so that the bottom of the coil will not touch the board. If you don't have any bushings, you can raise the coil form off of the board by triple nutting the screws. Feed the machine screw up from the bottom of the board, hold the screw into place with a nut on the top side of the board; then place another nut on the screw close to it's top, pass the remaining small piece of screw through the form and nut it again. Do this on both ends of the form. Plastic screws would be great for mounting the coil if you can find them.

- If you use an air variable capacitor, it will have 5 connections. For the first capacitor of the dual section unit, there will be one connection on each side of the stator plates (the ones that do not move.) Please note that these two connections are the same; they are NOT separate. Both are connected to the same set of non-rotating plates. The other section will have two similar connections. The frame of the whole unit is the ground side of both sections. Thus, the rotors are already tied together; the ones that move as you turn the knob. To connect the two capacitors in parallel, simply connect the stator plates together a shown in the wiring diagram. Don't worry if the two sections are different in size, in fact, they probably will be. Also, you may find a variable capacitor with three sections. In that case, connect only two of them together.
- I mounted my air variable by using some small "L" brackets fixed to the board with small wood screws and fixed to the capacitor with very short machine screws. Another way is to put a small length of wire from each of the four stator connection tabs that go down and under wood screws in the board.
- If you use the plastic insulated variable capacitor such as is found in small transistor radio radios, it will be a much smaller unit than the air type and will probably only have three connection tabs. The center tab is the rotor or ground connection, and the outer ones are the stator connections which will be tied together. To mount this type of capacitor, you can epoxy the back side to the board. This means the shaft will point up.
- Mount the knob on the shaft of the variable capacitor. There are several ways that knobs are fitted to shafts. One way is with a small slotted or Allen screw in the side of the knob that is tightened down on the shaft. Another way, used when the shaft is notched into a "D" shape, is to use a knob with a matching "D" shaped hole and a spring clip -- no screw is needed in this case. Another way, common to small transistor type variable caps, is to use a very small machine screw that

passes through the center of the knob into a tapped hole in the end of the shaft.

- The terminations for the Lo-Z, Hi-Z, and Ground connections are either Fahnstock clips or machine screws with double nuts. If you use double-nutted screws, then the wires connecting into the circuit will be under the first nut. Make a small loop in the wire(s) then tightened the nut securely down on them. (You may want to add a washer.) Then another nut can be used to make the connection to the antenna lead-in wire and similarly to the ground wire. You may used the same double nutted machine screw scheme for the earphones.
- To make connections to the coil, scrape off the wires to expose the copper after cutting then to length. Tin them with solder. For the taps, scrape and tin them and then solder extension wires cut to the proper length. Start by using the 30 turn tap. Later you can try the 60 turn tap and see if it works better.
- Wire up the circuit as shown in the wiring diagram. This includes wiring in the 330 uuFd (micro-micro farad or *picofarad*) capacitor, the detector diode, and the 47K resistor. Be careful when soldering the detector diode because too much heat can damage it.
- Finish off the project by putting four stick-on rubber feet on the bottom of the board in each of the corners. In a pinch, epoxy on pencil erasers for the non-slip feet.

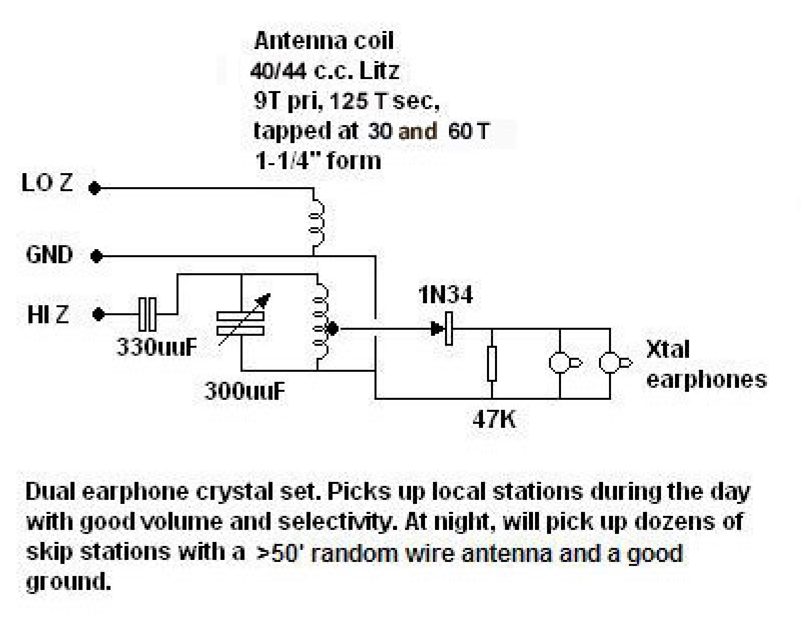

Dual earphone crystal set. Picks up local stations during the day with good volume and selectivity. At night, will pick up dozens of skip stations with a >50' random wire antenna and a good ground.

Schematic of the crystal radio

- Learn to read schematic diagrams. Eventually you will learn to wire up projects from the schematic without the need of a wiring pictorial.

Copyright © 2007 - Prof. W.S. Walker, West Virginia State University - All Rights Reserved.

Contact me

rexresearch.com

Pyramid Energy Patents (II)

shutterstock
images | videos | music
THINK YOU KNOW STOCK?
Think again.

See also : Pyramid Patents (I)

RO121993
PYRAMID-TYPE CONSTRUCTION FOR RELAXATION

KR20040076310
STRUCTURE OF SWEATING BATHROOM USING THE PRINCIPLE OF GATHERING ENERGY IN THE FORM OF A PYRAMID AND IMPROVING ENDURANCE

RO125394
SQUARE PYRAMID-SHAPED METAL CONSTRUCTION WITH QUARTZ CRYSTALS IN THE APEX

DE202004010557
Flat sheet material blank for construction of hollow tetrahedral pyramid...

KR20140011732
CLOCK WISE COIL ON QUARTZ INSTALLED PYRAMID SHAPE DEVICE FOR SUBTLE ENERGY

USD527302
Pyramid for harmonizing energy

UA18114
ENERGY PYRAMID "LABIRINT"

UA18109
ENERGY PYRAMID

RU2184574
METHOD FOR TREATING PATIENTS WITH PYRAMID ENERGY

KR20070122080

APPARATUS FOR GENERATING PYRAMID ENERGY

JPS6084139
PYRAMID ENERGY GENERATING APPARATUS

JPS5786366
PYRAMID-POWER DEVICE FITTED AROUND SCROTUM IN ORDER TO REINFORCE ENERGY

JPS5424180
MILK CONTAINER SERVING ALSO AS PYRAMID ENERGY GENERATOR

JPS548055
BED APPLYING PYRAMID ENERGY GENERATOR

JPS548058
PILLOW APPLYING PYRAMID ENERGY GENERATOR

JPS55157456
REGENERATIVE DEVICE AND METHOD FOR CUTLERY UTILIZING PYRAMID ENERGY COMBINED WITH MAGNETIC FORCE

JPS59187300
PYRAMID ENERGY GENERATOR

JPS5495851
PYRAMID ENERGY GENERATOR

WO2004058339
INSTRUMENT AND PROCEDURE FOR THE USE OF PYRAMID ENERGY

EP0302192
Device for the treatment of living beings and plants with pyramid energy.

DE202006015477
Energy twin-pyramid to measure cosmic and terrestrial energy has mountain crystal energy condenser

DE202006004778
Pyramid with square base containing special energy level...

DE102005052831
Gravitational energy conversion device, has pyramid produced from cuboidal stones

DE29724691
Energy pyramid for medical use

DE19717053
Energy pyramid for medical use

CN1617188
Pyramid energy focusing model and its making method

Ad
75% troubleshoot their networks manually. Why?
NetBrain

RO121993
PYRAMID-TYPE CONSTRUCTION FOR RELAXATION

Inventor(s): BECIU NICULINA

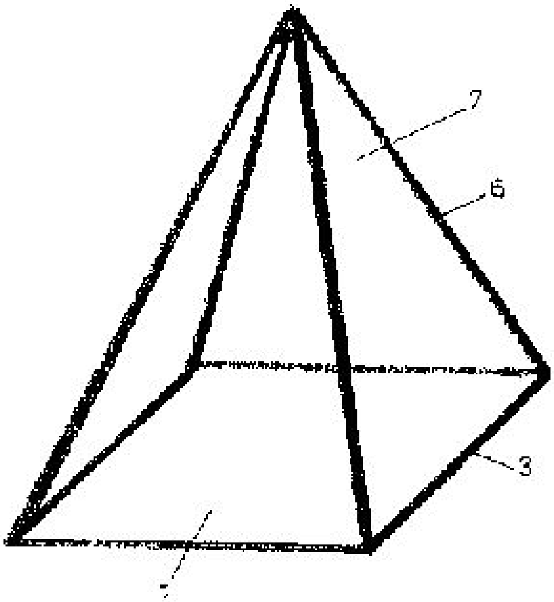

Fig. 1

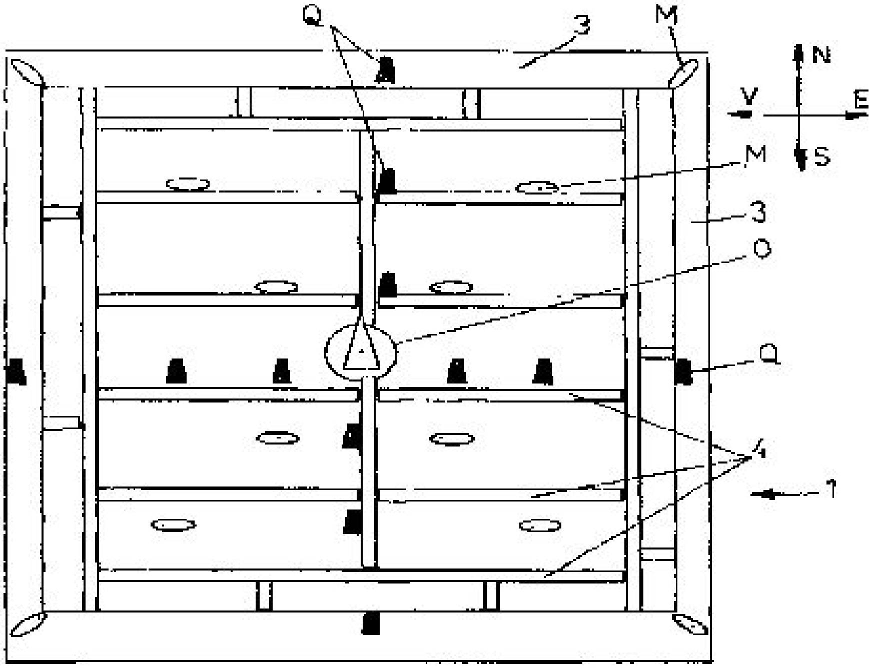

Fig. 2

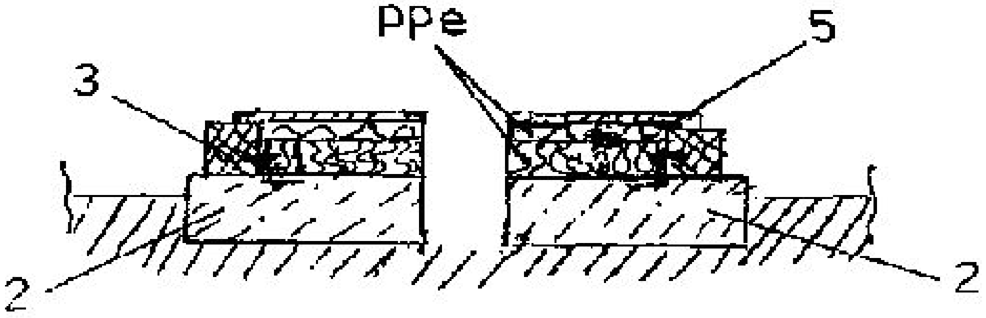

Fig. 3

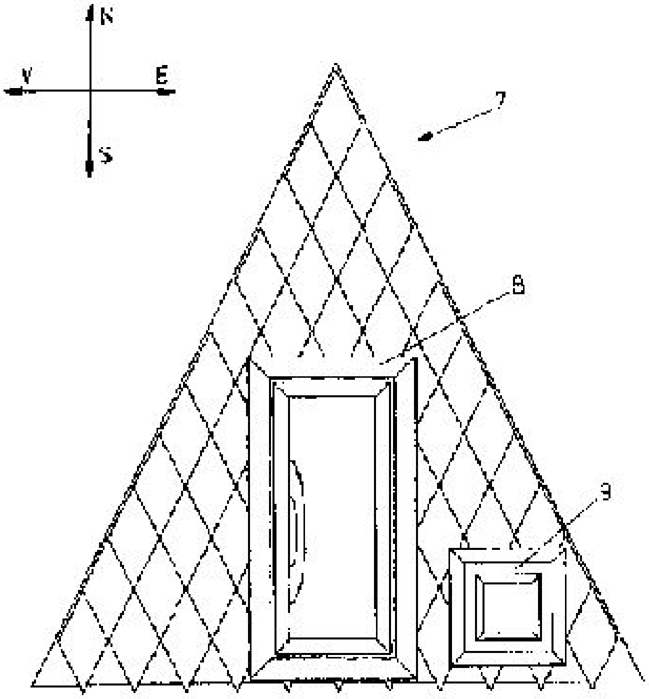

Fig. 4

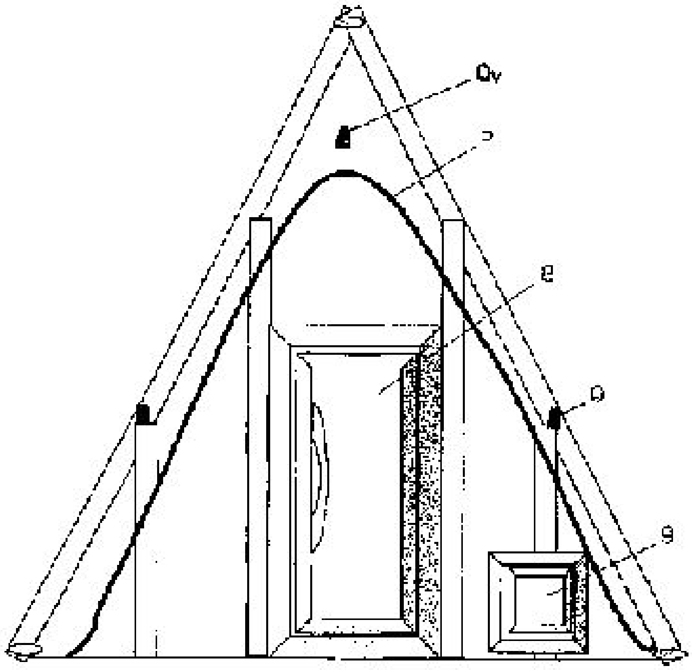

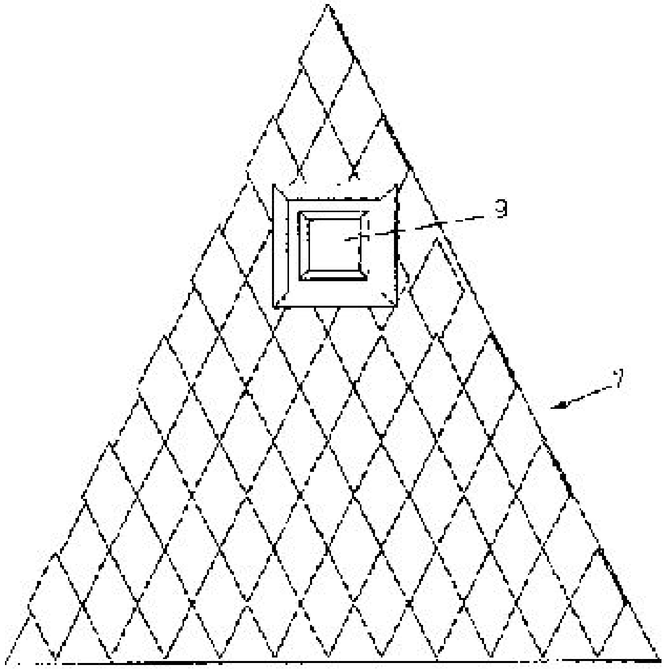

Fig. 6

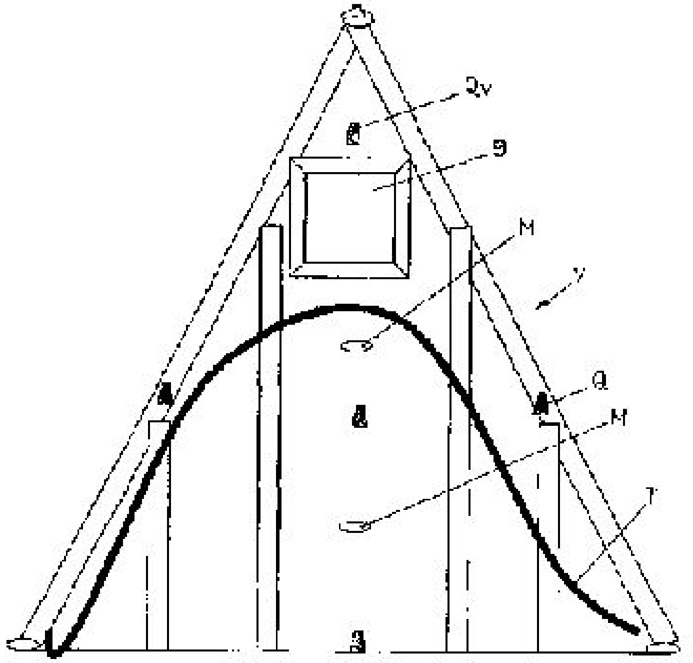

Fig. 7

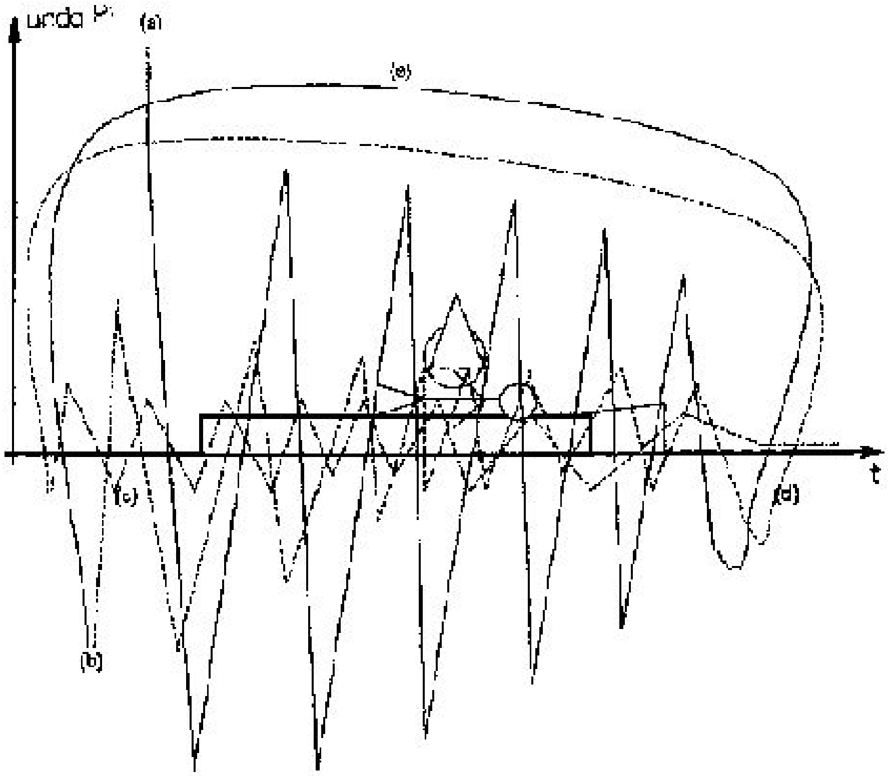

Fig. 8

The invention relates to a pyramid-type construction for body relaxation, meant to be used as therapeutical means in the field of complementary and alternative therapies. According to the invention, the pyramid-type construction for relaxation, comprising a square base and four triangular walls forming an enclosure provided with windows (9, 9', 10) and an access door (8), has multi-layered base and walls made of rigid plates with quartz crystals (Q) and permanent magnets (M) therebetween, in the base center there being placed an aluminium tube filled with consecutive layers of semi-precious stones, clay and sand, the sand layer embedding a hexagon-shaped quartz crystal with a pointed tip oriented towards the pyramid apex, where a quartz crystal (Qv) with a downwardly-oriented tip is located, while inside the walls there is placed a multi-layer water pipe (P) with aluminium insertion, having a continuous lay-out starting from a pyramid angle and returning thereto after forming in each wall a loop reaching a height of 1/3 up to the pyramid apex, while observing the side inclination angle generating the "pi" relation between the height and the base side.

KR20040076310
STRUCTURE OF SWEATING BATHROOM USING THE PRINCIPLE OF GATHERING ENERGY IN THE FORM OF A PYRAMID AND IMPROVING ENDURANCE

Inventor(s): EOM DONG IL

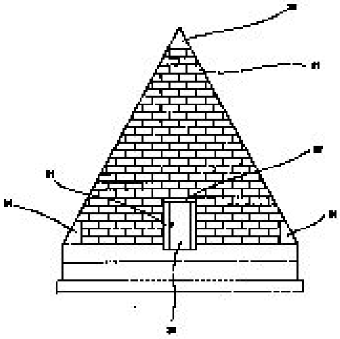

도면2

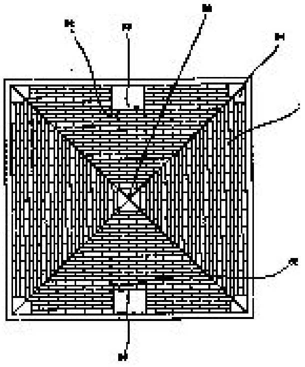

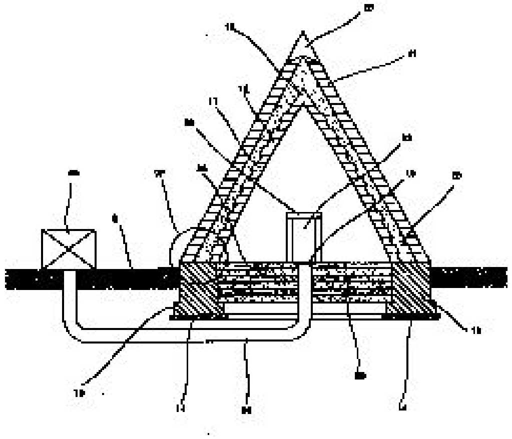

도면4

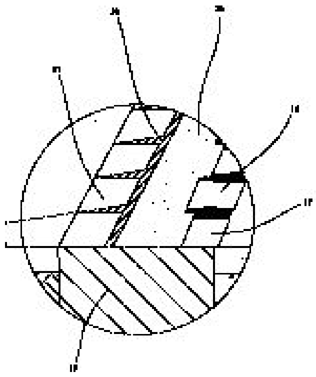

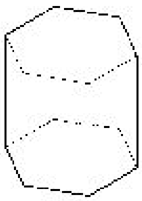

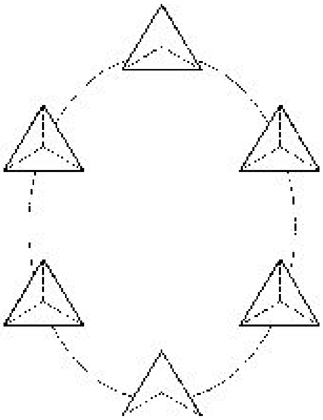

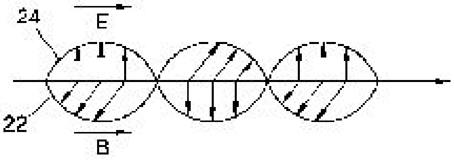

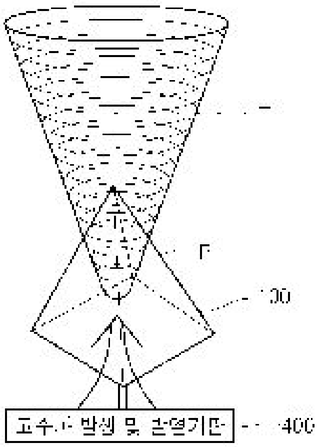

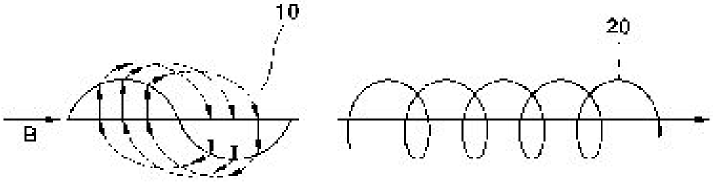

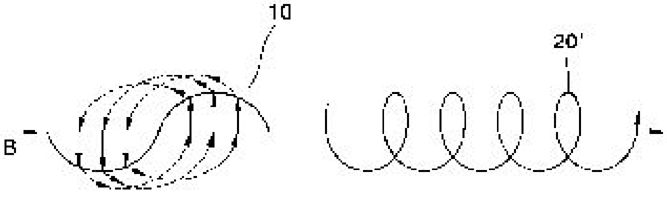

PURPOSE: Provided is a structure of a sweating bathroom to remove any inconvenience and uneasiness of a user staying in an internal space by applying the principle of gathering energy in the form of a pyramid and improving endurance; and to improve the user's health by using yellow soil in layering stone in the form of a pyramid. CONSTITUTION: The structure of a sweating bathroom where stone is layered in the form of a pyramid on a basis(10) and a high temperature is kept by the burning for a predetermined time comprises an inner layer block(11,12) layered for forming an inner space in the form of a pyramid; an outer layer block(21) layered for forming a predetermined space to the inner layer block(11,12); a filler(25) filled in the space in the state of combining the inner layer block(11,12) and made of yellow soil as a main material; and a floor material forming the bottom of the inside by layering a yellow soil layer(35) and a salt layer(36) alternately on the space formed by the basis(10).

RO125394

SQUARE PYRAMID-SHAPED METAL CONSTRUCTION WITH QUARTZ CRYSTALS IN THE APEX

Inventor(s): ANDRONOVICI LIVIU

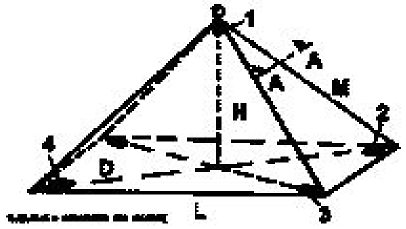

Fig. 1

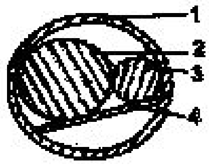

Fig. 2

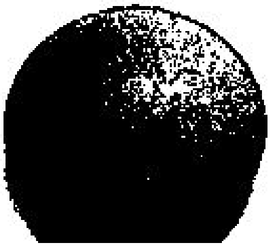

Fig. 3

Fig. 4

Fig. 5

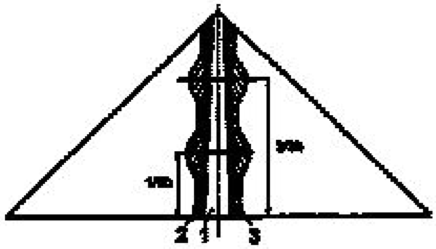

Fig. 6

Fig. 7

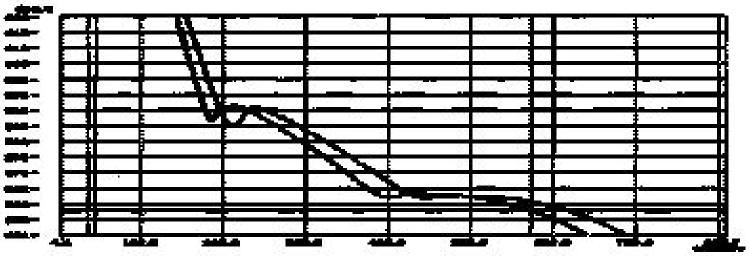

Fig. 8

DE202004010557
Flat sheet material blank for construction of hollow tetrahedral pyramid...

Fig. 1

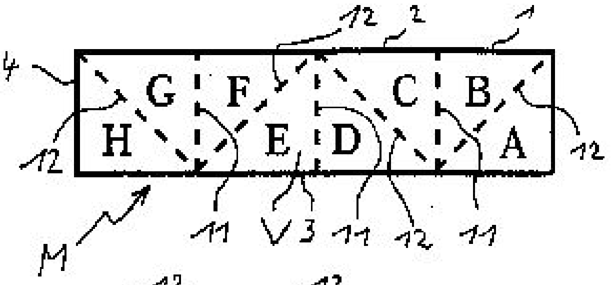
12
2
1
4
G
F
C
B
12
H
E
D
A
12
M
11
V
3
11
12
11

Fig. 2

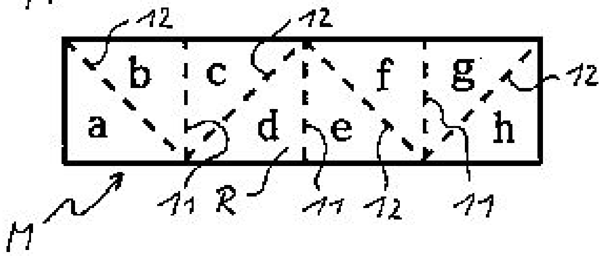
12
12
b
c
f
g
12
a
d
e
h
11
R
11
12
11
M

Fig. 3

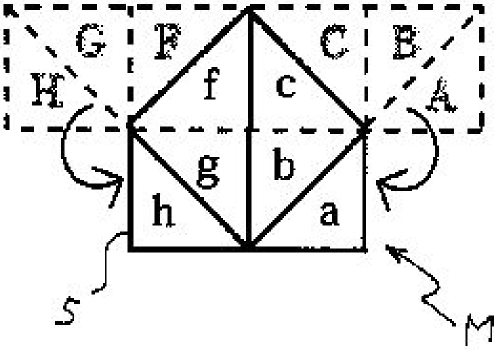
G
F
C
B
H
f
c
A
g
b
h
a
S
M

Fig. 4

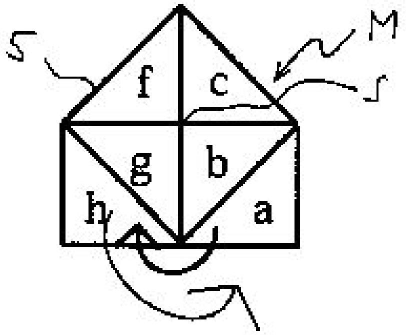

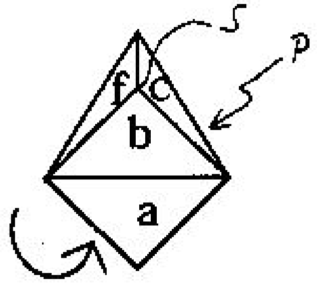

Fig. 5

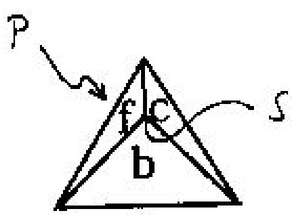

Fig. 6

Fig. 7

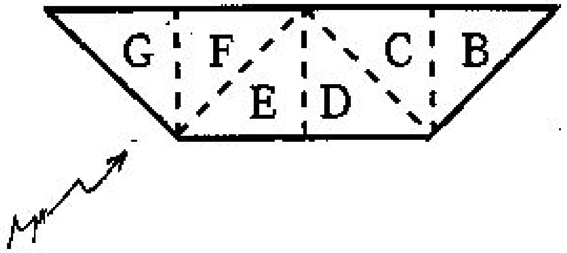

Fig. 8

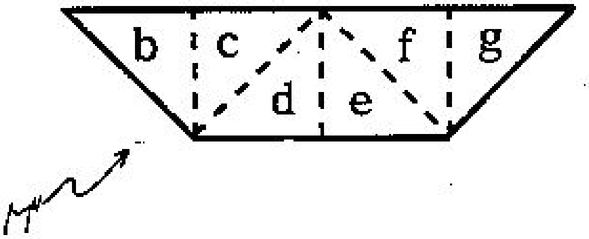

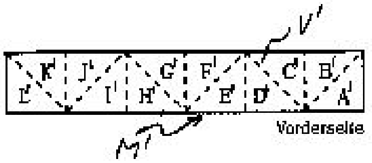
K'
J'
G'
F'
C'
B'
L'
I'
H'
E'
D'
A'
V'
M'
Vorderseite

Fig. 9

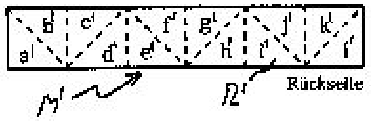
b'
c'
f'
g'
j'
k'
a'
d'
e'
h'
i'
l'
M'
R'
Rückseite

Fig. 10

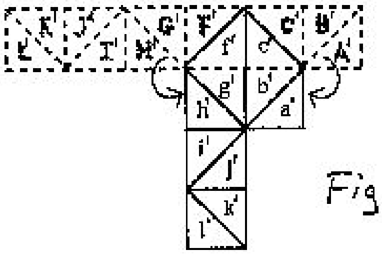
K'
J'
G'
F'
C'
B'
L'
I'
H'
A'
f'
c'
g'
b'
h'
a'
i'
j'
k'
l'

Fig. 11

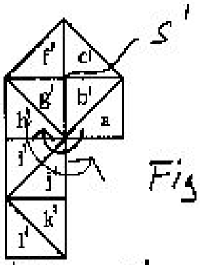
S'
f'
c'
g'
b'
h'
a
i'
j'
k'
l'

Fig. 12

b' auf g' schieben und h' unter a' legen.

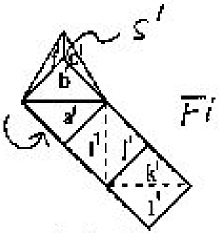

Fig. 13

a unter b klappen.

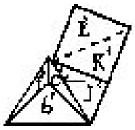

Fig. 14

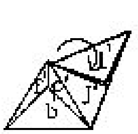

Fig. 15

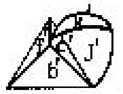

Fig. 16

Nachdem l auf K gefaltet wurde, schiebt man k unter f.

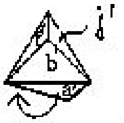

Fig. 17

a von innen nach außen klappen.

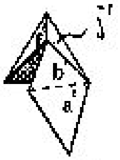

Fig. 18

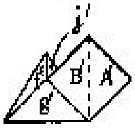

Fig. 19

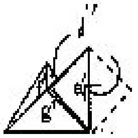

Fig. 20

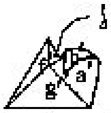

Fig. 21

b unter g schieben.

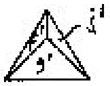

Fig. 22

Modulvarianten

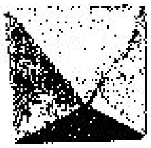

8er-Modul
Fig. 23

Flyer-Verkaufsbox 1
Fig. 24

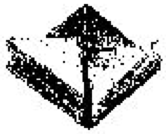

Flyer-Verkaufsbox
Fig. 25

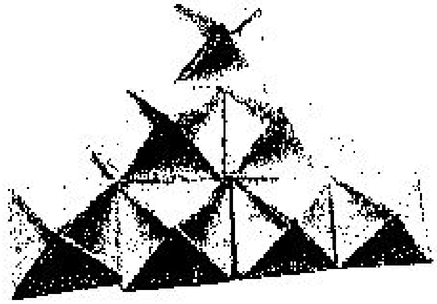

Pyramide aus 12 8er Modulen
Fig. 26

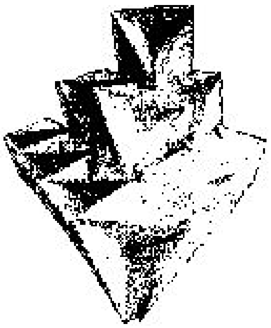

Rückansicht Pyramide
Fig. 27

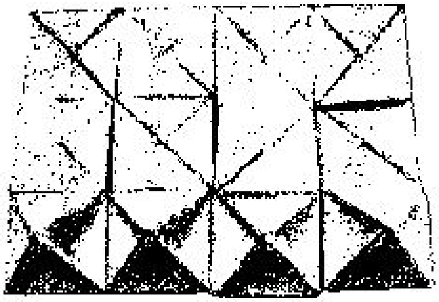

Modulwand
(untere Reihe 8erModul, obere 4er-Modul
Fig. 28

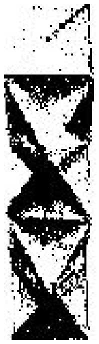

Säule aus 4 8erModulen
und 1 4erModul
Fig. 29

Inventor(s): PUPILLO ROSA MARIA et al

Some of the six triangular faces (A-H) defined on the blank (1) are folded over each other during the folding process to form a regular tetrahedron. Alternate cross-folds (11) and diagonal folds (12) define equilateral right-angle triangles with the short sides equal to the width (4) of the rectangle. The width of the rectangle is one quarter of the length (2).

KR20140011732

CLOCK WISE COIL ON QUARTZ INSTALLED PYRAMID SHAPE DEVICE FOR SUBTLE ENERGY

Inventor(s): KIM CHANG HYUN

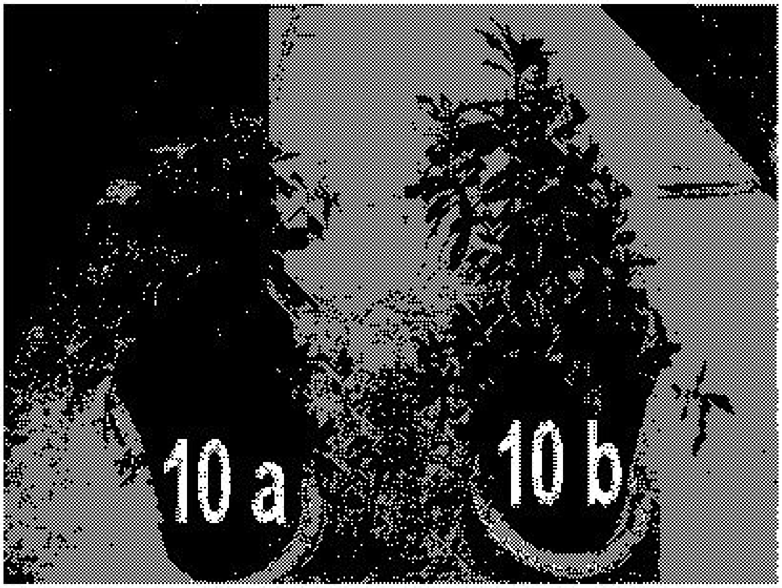

[illegible]

11A

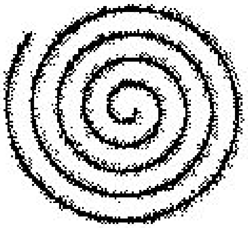

11B

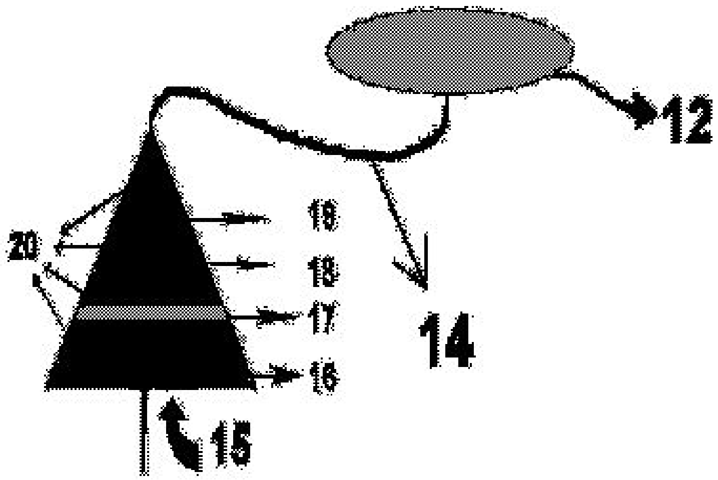

The present invention relates to a pyramid-shaped subtle energy generating device emitting scalar waves, which is characterized by winding a coil around quartz in a clockwise direction and putting the quartz into a pyramid-shaped body. The device of the present invention is configured to generate scalar waves using a scalar wave generating coil wound in the opposite direction with the same number of windings and right-handed torsion waves using a right-handed torsion coil wound clockwise starting at the center from free electrons present in a living space. The pyramid-shaped body for accumulating the free electrons is made by piling up a mixture of 50% minerals such as steel shavings, quartz shavings, brass shavings, copper shaving, etc. and 50% polyester resin into a pyramid shape to effectively accumulate the free electrons, and the accumulated free electrons pass through the coils wound the quartz to generate scalar waves and right-handed torsion waves. A drawn line is connected to the top of the pyramid-shaped body to effectively deliver the generated scalar and right-handed torsion waves to a human body. Therefore, the pyramid-shaped subtle energy generating device generates the scalar waves not using a man-made power source and provides energy beneficial to the human body.

USD527302
Pyramid for harmonizing energy

Inventor(s): VLADAMIROV STEPHAN GEORGIEV

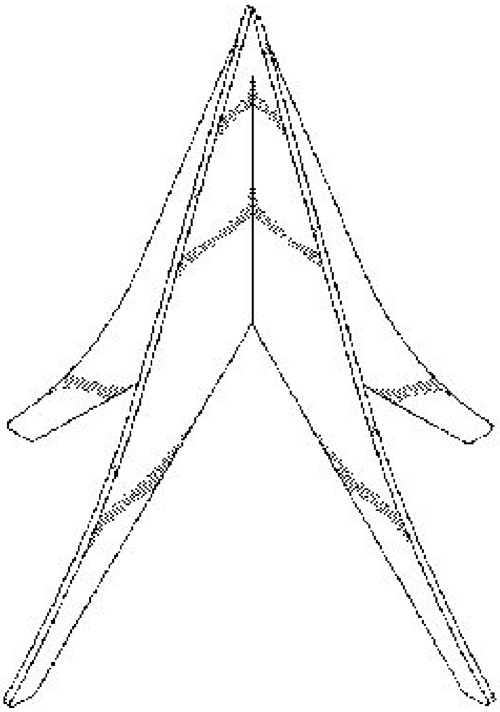

FIG. 1

UA18114
ENERGY PYRAMID "LABIRINT"

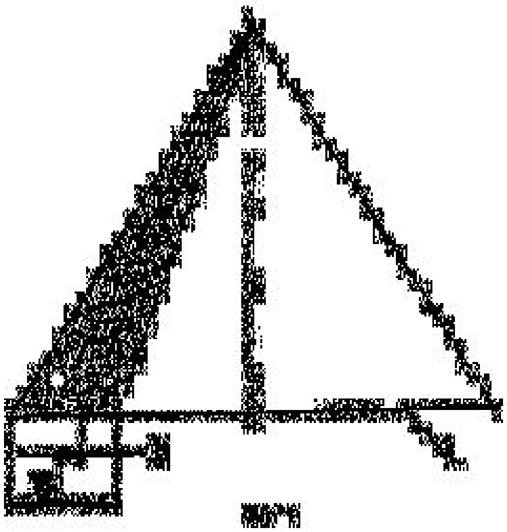

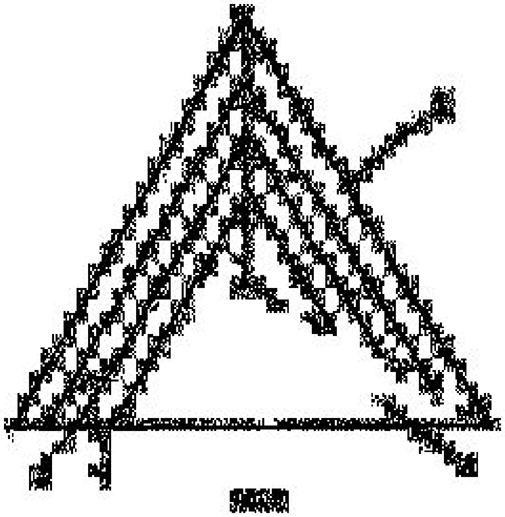

Inventor(s): YEVREIENKO YURII PAVLOVYCH [UA]; HURYN VASYL ARSENTIIOVYCH

The proposed energy pyramid has a central opening which is coupled with cavities that are separated by walls with variable profile.

UA18109
ENERGY PYRAMID

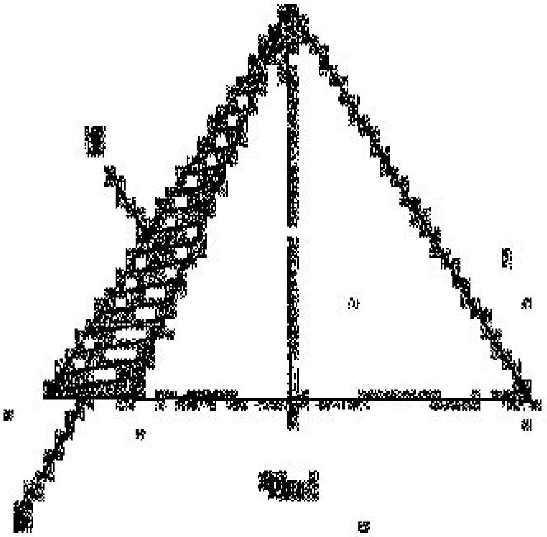

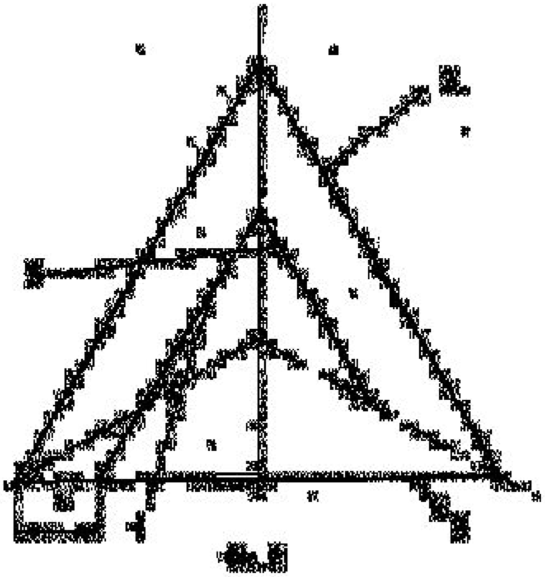

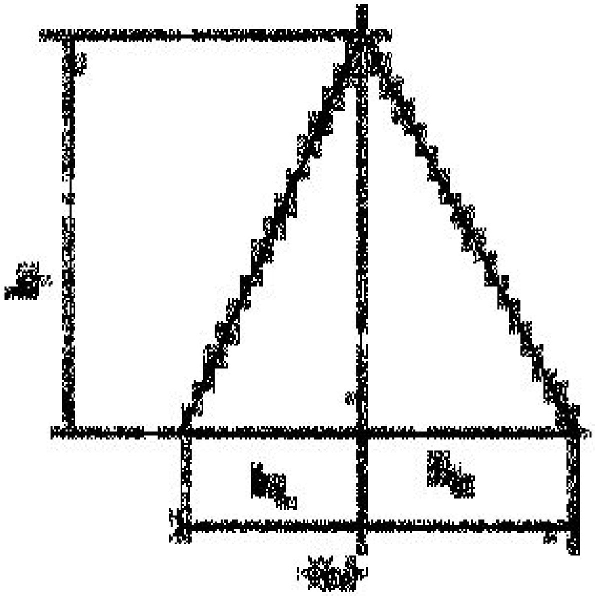

Inventor(s): YEVREIENKO YURII PAVLOVYCH [UA]; HURYN VASYL ARSENTIIOVYCH

The proposed energy pyramid has planar facets with spiral bases.

RU2184574
METHOD FOR TREATING PATIENTS WITH PYRAMID ENERGY

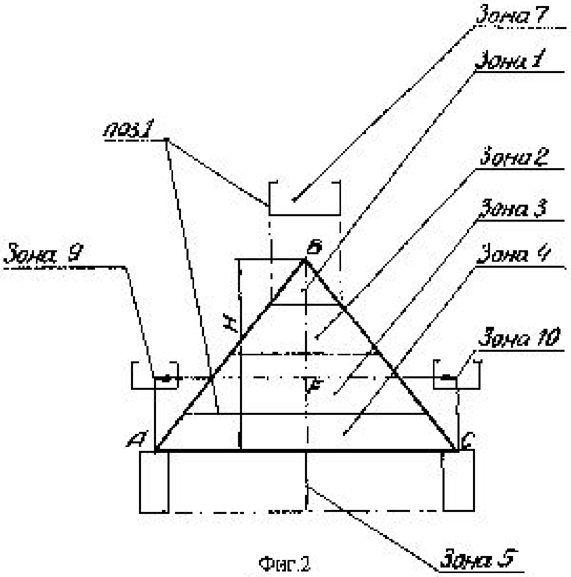

Фиг.2

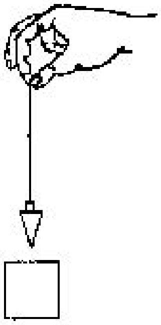

Фиг.3

Inventor(s): PEREPELKIN I A

SUBSTANCE: method involves placing water under negative temperature in various places inside, above and under the pyramid at different heights and apart from each of the borders. Water is hold until it freezes. Filament-shaped structures are observed in frozen water to determine energy distribution zones in the pyramid and beyond its borders according to their number and length. Then patient state is examined by applying biological indicator device. Energy zone is selected in the pyramid and beyond its borders matching the patient's state. The patient or means to be used for following internal or external application are placed in this zone. The means are hold there not less than 24 h. Water is usable as internal application means, foil, stones, copper and bed cloths are usable as external application means. EFFECT: enhanced effectiveness of catheter protection against moisture with no sticking effects. 4 cl, 3 dwg

---

KR20070122080
APPARATUS FOR GENERATING PYRAMID ENERGY USING PYRAMID STRUCTURE

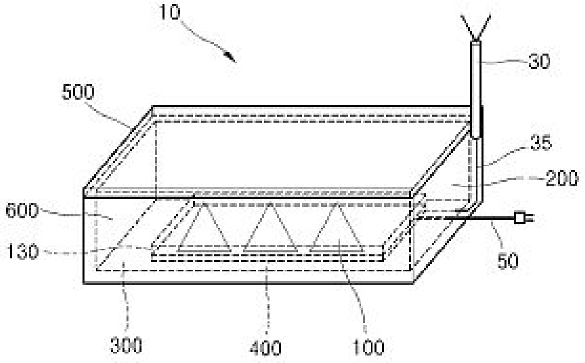

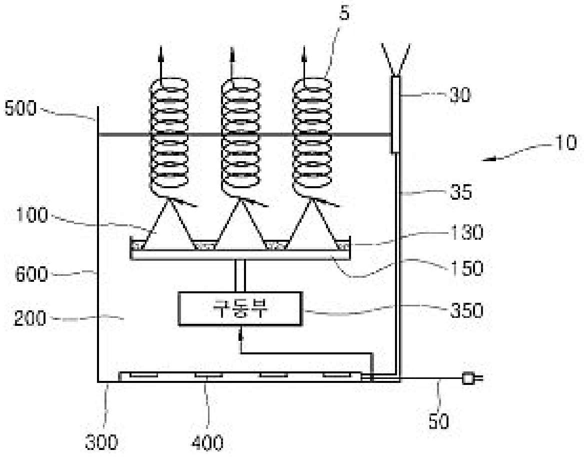
구동부

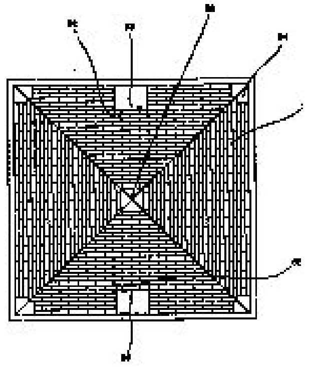

도면4

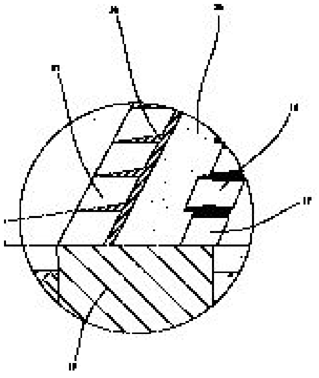

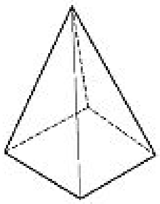

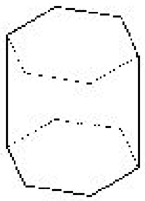

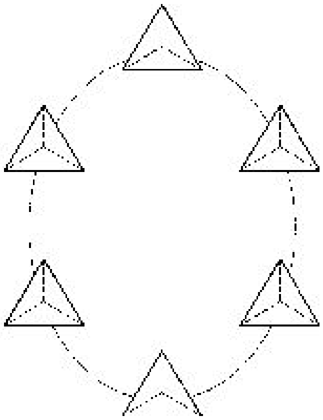

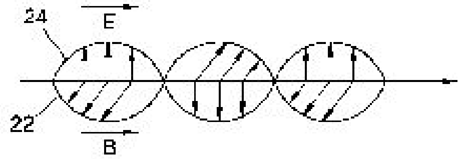

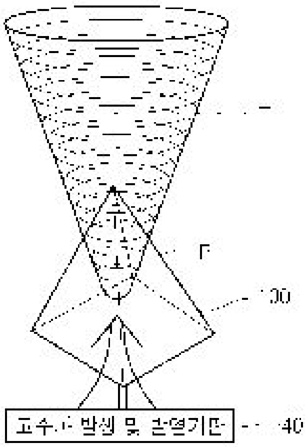

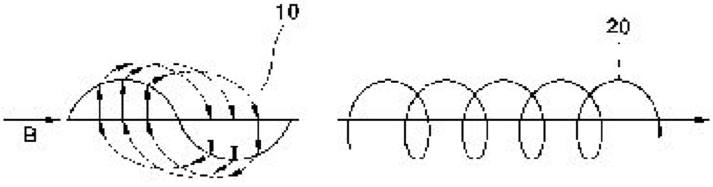

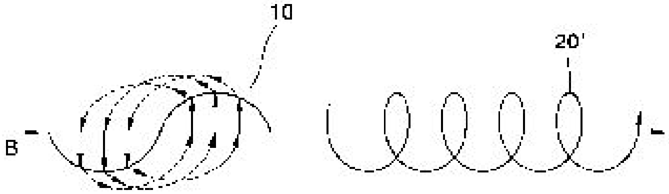

Inventor(s): I YEONG GYU [KR]; NRHO BYUNG

An energy generating device using a pyramid structure is provided to amplify pyramid energy of the pyramid structure by changing a magnetic field by applying high frequency to the pyramid structure. An energy generating device(10) generating pyramid energy in a pyramid structure(100) is composed of the pyramid energy generating pyramid structure; a case(200) containing the pyramid structure; a high frequency generating and heating substrate(400) installed at the inner bottom of the case and operated electrically or magnetically by receiving electricity from an external power source, in order to generate a high frequency signal and heat and amplify pyramid energy generated from the pyramid structure; and an object holder(500) formed at the outer upper part of the case to place an object and absorb pyramid energy generated from the pyramid structure stored in the case.

JPS6084139
PYRAMID ENERGY GENERATING APPARATUS

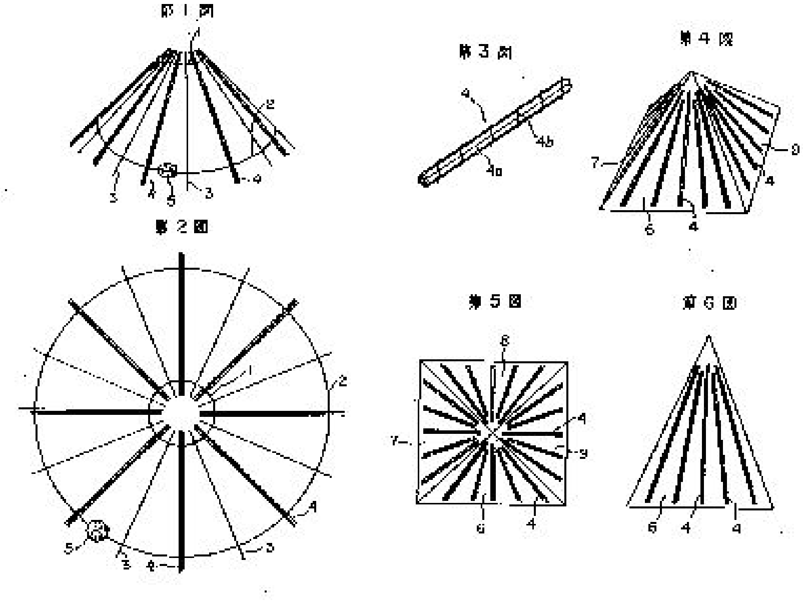

Inventor(s): WATANABE NAOMICHI

PURPOSE:To provide an apparatus for generating strong pyramid energy regardless of azimuth, obtained by arranging a wire member and a composite body of the wire members along a secondary curved surface. CONSTITUTION:A large number of single wire members 3 and composite bodies 4 of the wire members are connected to a ring shaped upper support member 1 and a ring shaped lower support member 2 and the single wire members 3 and the composite bodies 4 are arranged along a conical surface. Furthermore, a metal or bamboo material is appropriately used as each single wire member. In addition, when the pole of a magnet 5 is contacted with the lower support member 2, the radiation direction of pyramid energy can be bundled to a lower direction. This generation apparatus always generates a large amount of strong pyramid energy even if directed to any direction and excellent in dehydration effect or corrosion preventing effect.

JPS5786366
PYRAMID-POWER DEVICE FITTED AROUND SCROTUM IN ORDER TO REINFORCE ENERGY

Inventor(s): MAGATA YASUHEI

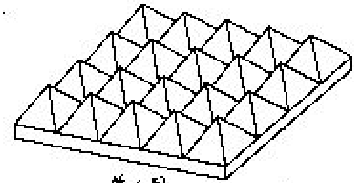

第1図

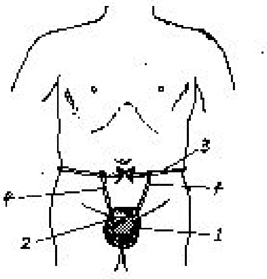

第2図

JPS5424180
MILK CONTAINER SERVING ALSO AS PYRAMID ENERGY GENERATOR

Inventor(s): SANO KATSUO

第1図

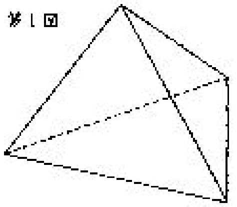

第2図

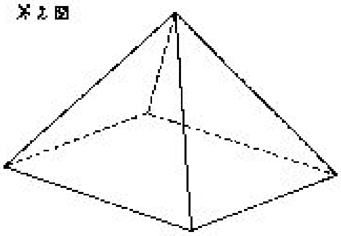

JPS548055
BED APPLYING PYRAMID ENERGY GENERATOR

Inventor(s): SANO KATSUO

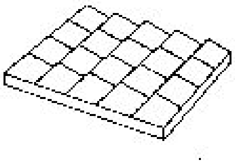

第2図

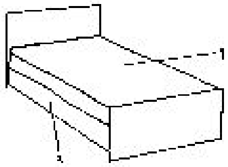

JPS548058
PILLOW APPLYING PYRAMID ENERGY GENERATOR

Inventor(s): SANO KATSUO

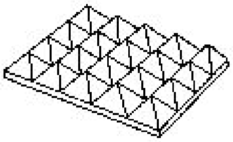

第2図

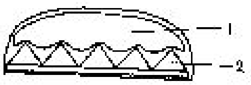

JPS55157456
REGENERATIVE DEVICE AND METHOD FOR CUTLERY UTILIZING PYRAMID ENERGY COMBINED WITH MAGNETIC FORCE

第1図

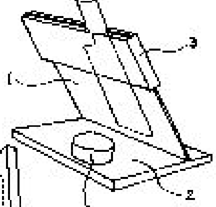

第2図

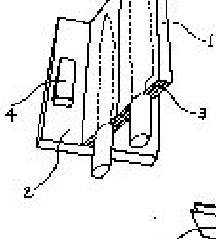

第3図

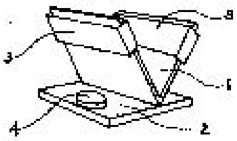

Inventor(s): OOHASHI TATSUNORI

PURPOSE:To increase pyramid energy power and regenerate cutlery in a shorter period, by fitting a cutlery holder with a certain angle of elevation to a base plate equipped with a magnet in the magnetic north. CONSTITUTION:A cutlery holder 1 is fixed with about 60 deg. of elevation on a base plate 2 and a magnet 4 is also fixed in the magnetic north on the plate 1 to increase the magnetic force in the cardinal direction of north. This can reinforce energy power working in other three cardinal directions and regenerate cutlery set to the holder 1 equipped with a supporting frame 3 in about a week.

JPS59187300
PYRAMID ENERGY GENERATOR

Inventor(s): WATANABE NAOMICHI

第1図

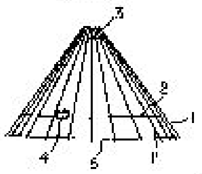

第4図

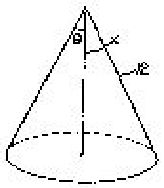

第2図

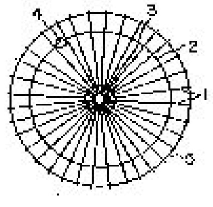

第5図

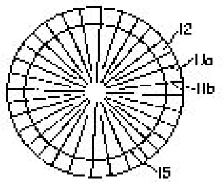
12
11a
11b
15

第3図

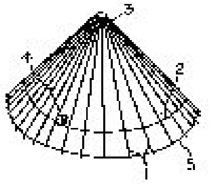

第6図

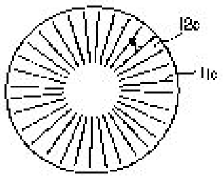

第7図

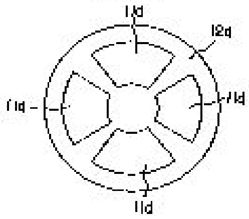
11d
12d
11d
11d
11d

第8図

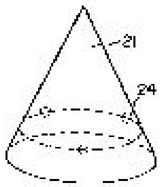

第9図

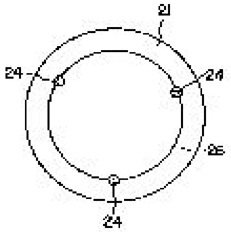
21
24
24
25
24

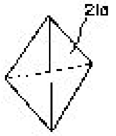

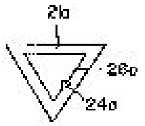

第12図

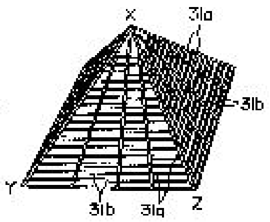
X
31a
31b
Y
31b
31a
Z

第13図

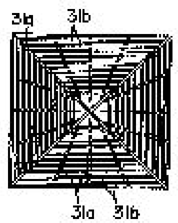
31a
31b
31a
31b

JPS5495851
PYRAMID ENERGY GENERATOR

Inventor(s): SANO KATSUO

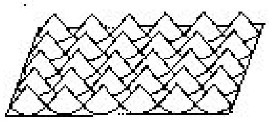

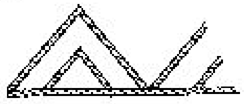

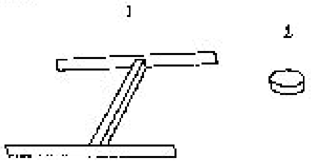

WO2004058339
INSTRUMENT AND PROCEDURE FOR THE USE OF PYRAMID ENERGY

Inventor(s): CSAPO GYOERGY [HU]; CSAPO GYOERGYNE

The subject of the invention is on the one hand an instrument for the use of pyramid energy, which consists of a carrier of any substance and square based pyramidal bodies arranged on the carrier, characterized by the spiral arrangement of the pyramidal bodies on the carrier, the multiple spiral formations arranged into a spiral system originating from multiple points close to each other on the carrier, and by the defined manner in which the size of the pyramidal bodies decreases within a spiral formation. The spiral formations of the instrument according to the invention follow the Fibonacci Principle. The subject of the invention is also a procedure using the device above for the use of pyramid energy, characterized by placing the pyramid energy-enhancing instrument above or below the substance to be treated - ideally water or substance containing water - in a way that the top vertices of the pyramidal bodies arranged on the carrier point towards the substance to be treated, and carrying on with the treatment for the desired time. The substance containing water may ideally be the human body, tea, fruits, fruit juices, etc.

The instrument and procedure for use of pyramid-energie The subject of the invention is an instrument and procedure for the use of pyramid energy.

The energy concentrating effect of pyramids has long been known.

Such is, for example, **Czechoslovakian patent Na 91304** which describes that a razor blade is repeatedly resharpened in a pyramid shaped housing and is reusable hundreds of times.

According to the currently accepted supposition the common resultant of cosmic space, terrestrial space, electric, magnetic, corpuscular, and other, thus far unknown particles and energies could be responsible for phenomena taking place in pyramid shaped spaces.

German patent Na 3525521 gives directions as to the formation and placement of the pyramid. The pyramid is fashioned of a material with a high dielectric constant to the reduced size of the Cheops pyramid.

German patent Na 19717053 describes an energy pyramid for therapeutic purposes. A number of pyramids according to the patent are placed on different points of the body to achieve the desired remedial effect. The pyramids may be 3-5 sided, with edge lengths between 1-10 cm, and may be made of aluminum or its alloys.

Japanese patent Na 9313624 proposes the use of reduced size Cheops pyramids with a plane inclination angle of 68-25 degrees for destroying infecting bacteria, viri and fungi, and curing joint and muscle pains.

According to the view currently held by researchers energy travel downwards in the pyramid in a spiral from the top vertex. According to suppositions this spiral movement lends special properties to pyramidal shapes. Spiral movements found in nature are characterized by the Fibonacci spiral model. (For example the expansion of an explosion, a spiraling galaxy also works according to the Fibonacci Principle.)
The aim during the elaboration of our invention was to allow for a much more widespread and general use of pyramid energy than before.

We have recognized that a great number of pyramidal shapes arranged according to the Fibonacci Principle allow for the use of the inner energy of pyramids with greater efficacy. In a formation thus constructed energies flowing upward also exercise their effects according to the Fibonacci spiral formation, their effect is multiplied (not simply added up).

Our further research has shown that if the pyramidal shapes we arranged in a spiral are not of equal size, rather they differ in size in an orderly manner along each spiral arm, the energy concentration that can be achieved is larger than before by up to two magnitudes.

The subject of the invention is on the one hand an instrument according to the above for the use of pyramid energy, which consists of a carrier of any substance and square based

pyramidal bodies arranged on the carrier, characterized by the spiral arrangement of the pyramidal bodies on the carrier, the multiple spiral formations arranged into a spiral system originating from multiple points close to each other on the carrier, and by the defined manner in which the size of the pyramidal bodies decreases within a spiral formation.

The spiral formations of the instrument according to the invention follow the Fibonacci Principle; the number of spirals on a carrier is between 3-21.

The instrument according to the invention may be made of any formable or afterhardening material, ideally rubber, plastic, metal, paper, concrete, gypsum or epoxy. Metal may ideally be copper.

The instrument may naturally be made with spirals formed clockwise or counter-clockwise, which correspond to carrying positive or negative energy/Ying and Yang/.

A ring shaped body made of metal and/or magnetic rods with high field intensity and/or stones (e. g. crystals) may be positioned inside the pyramidal bodies constituting the spiral arms.

Experience suggests that this also increases the energy level of the device.

The pyramidal bodies may be fixed to the carrier by gluing or using detachable jointing, or they may be formed together with the carrier by vacuum-forming plastic or machining metal.

The subject of the invention is also a procedure using the device above for the use of pyramid energy, characterized by placing the pyramid energy-enhancing instrument above or below the substance to be treated-ideally water or substance containing water-in a way that the top vertices of the pyramidal bodies arranged on the carrier point towards the substance to be treated, and carrying on with the treatment for the desired time.

While carrying out the procedure according to the invention it may be expedient to rotate the instrument around its center of symmetry.

The substance containing water may ideally be the human body, tea, fruits, fruit juices, etc.

Substance containing water thus treated may be used for treating human and animal sicknesses. The instrument and procedure according to the invention is also suited to set the satisfactory energy level of the human body.

The instrument according to the invention may be formulated so that it is set on a shoe sole shaped carrier.

The color of the instrument according to the invention may diverse/ideally according to the seven chakra colors/.

The formation of the instrument according to the invention is presented in Figures 1-9 as an example.

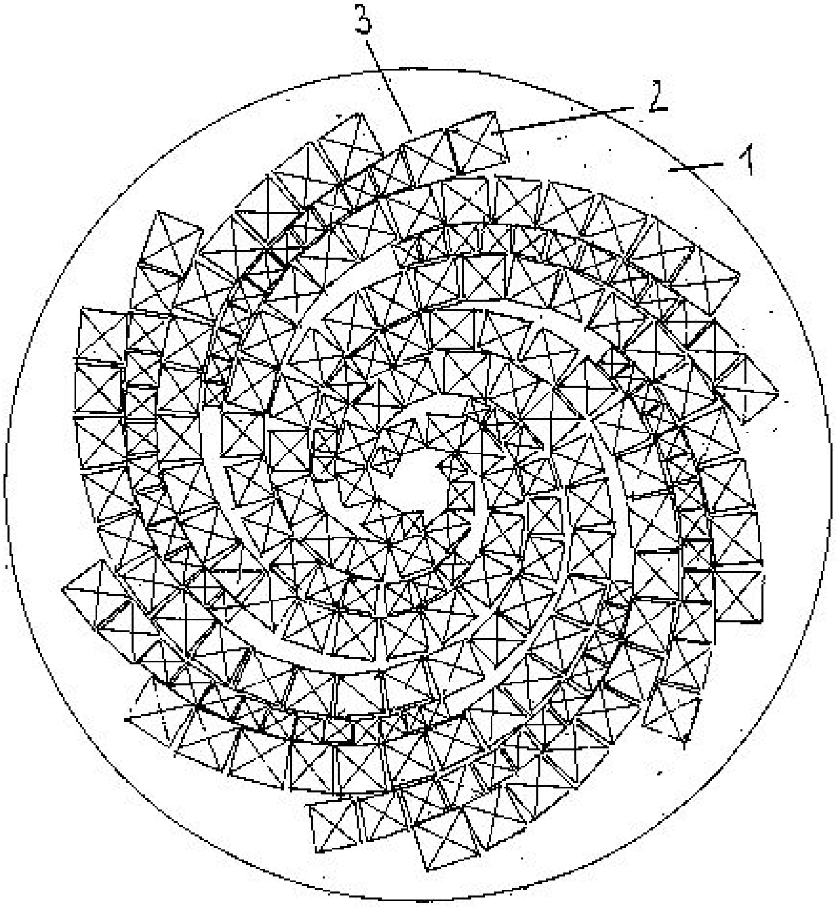

fig 1

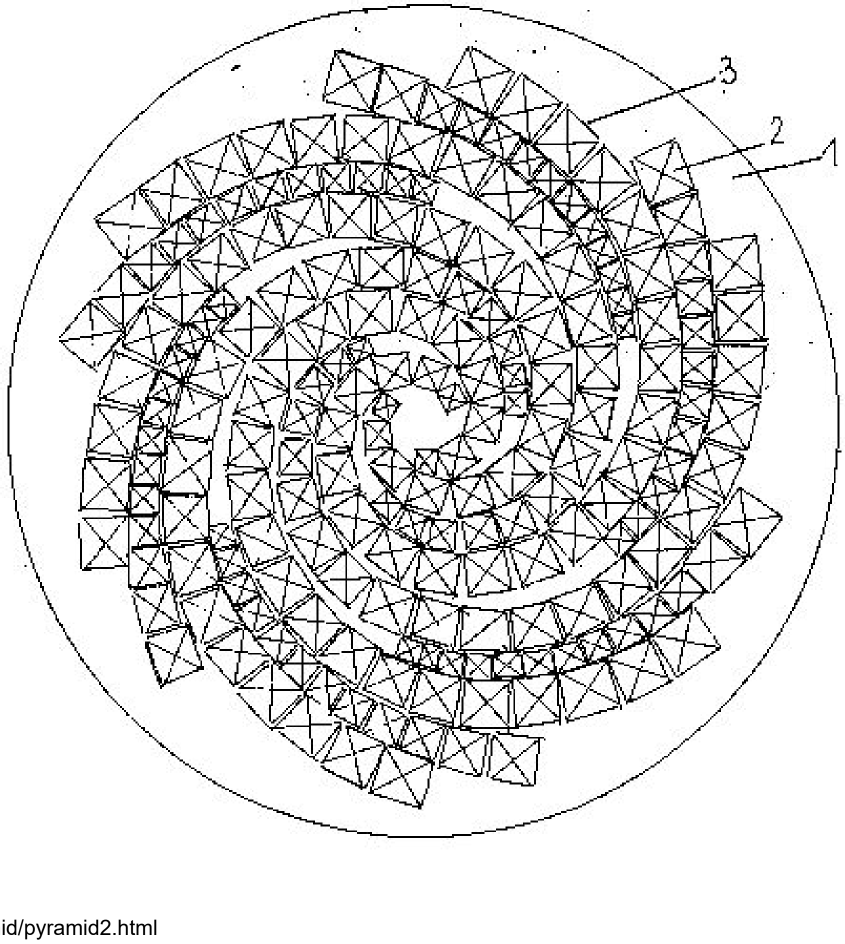

fig 2

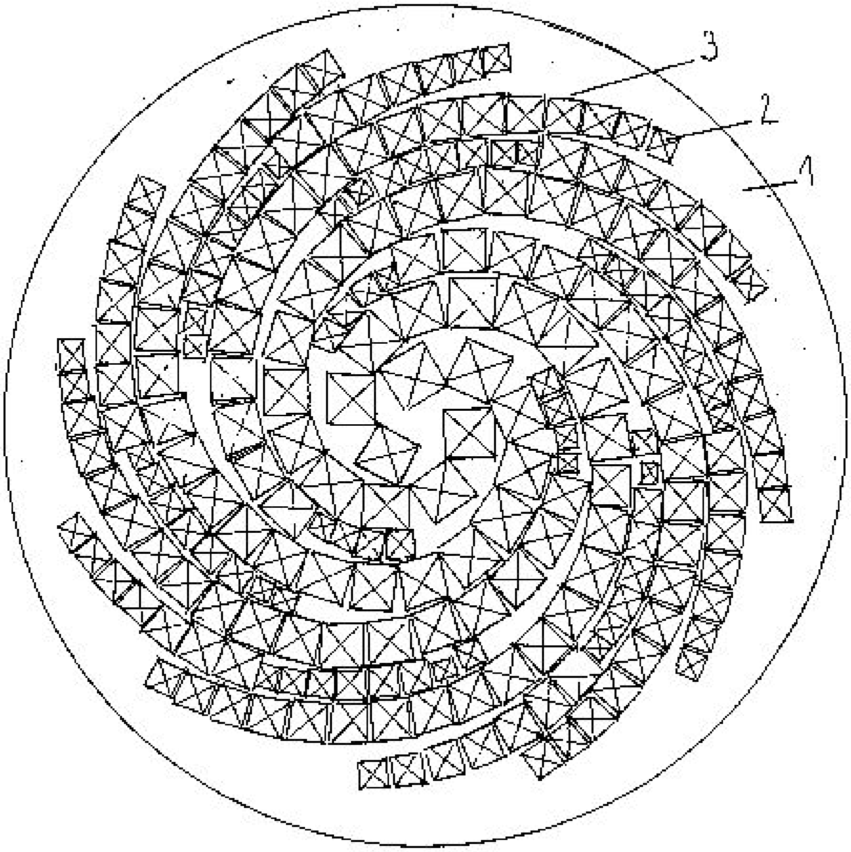

fig 3

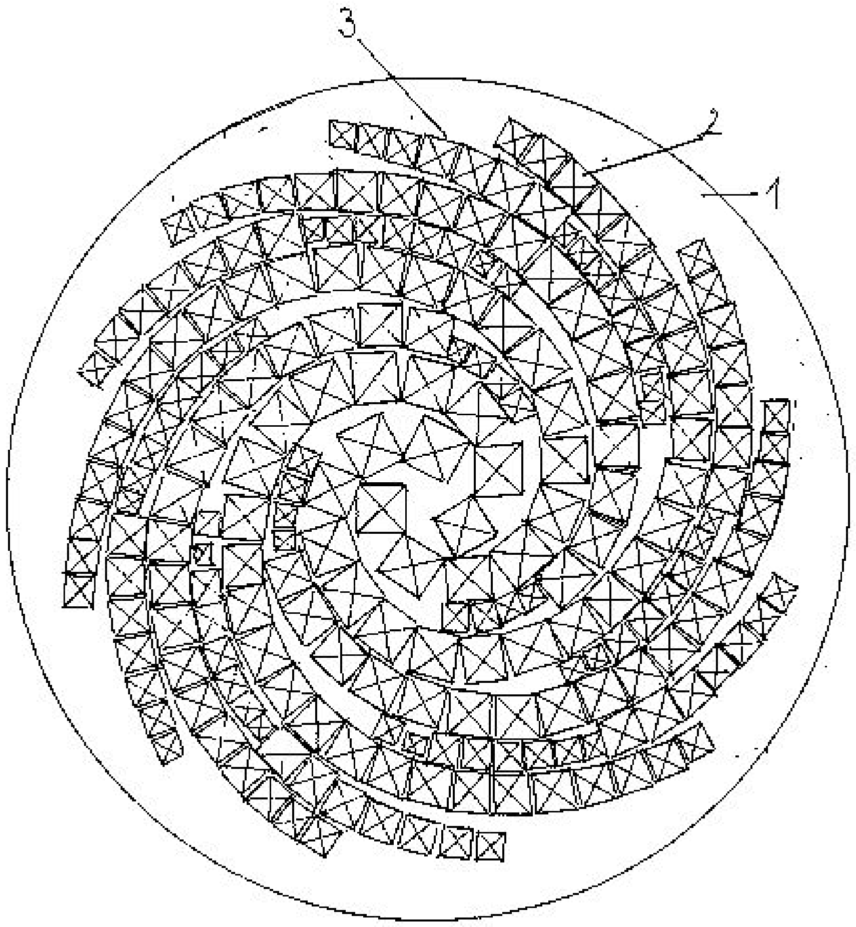

fig 4

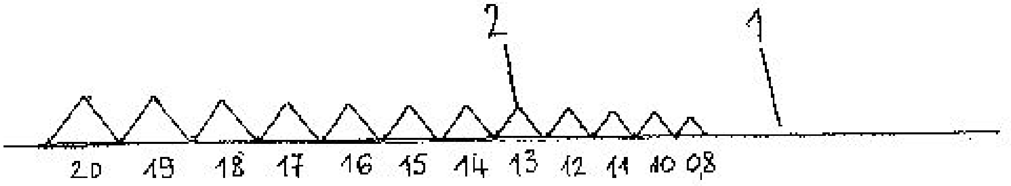
2
1
20
19
18
17
16
15
14
13
12
11
10
0,8

fig 5

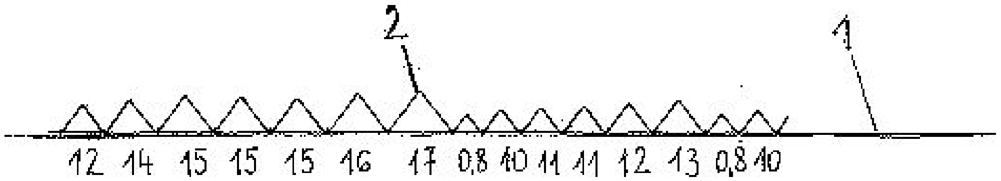
2
1
12
14
15
15
15
16
17
0,8
10
11
11
12
13
0,8
10

fig 6

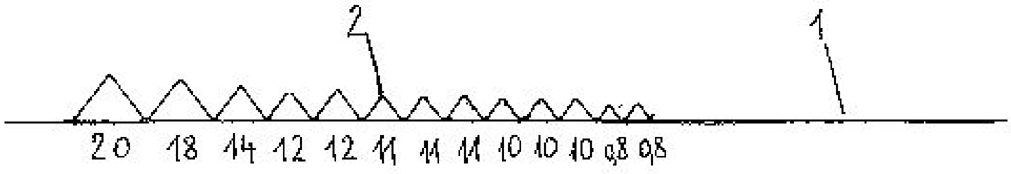
2
1
20
18
14
12
12
11
11
11
10
10
10
0,8
0,8

fig 7

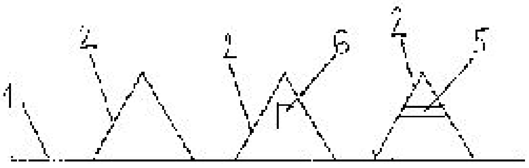

Fig 8

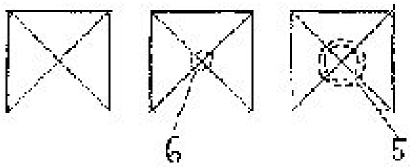

Fig 9.

Figure 1 shows the instrument according to the invention fashioned so that the pyramidal bodies/2/gradually increase in size from the originating point along the spiral arms 131.

- This is presented as a section in Figure 5.

Figure 2 shows the instrument in Figure 1 with an opposite direction of rotation.

Figure 3 shows a formation where the pyramidal bodies/2/in the spiral arms 131 alternate in increasing from a smaller size toward a larger size and decreasing from a larger size toward a smaller size.

- This is presented as a section in Figure 6.

Figure 4 shows a formation where they increase from a smaller size toward a larger size, then decrease from a larger size toward a smaller size, with the same repeating in the opposite direction on a neighboring spiral arm.

- This is presented as a section in Figure 7.

Figure 8 shows the ring made of metal 151 and the strong permanent magnet/6/to be placed inside the pyramid as a section.

- The same is shown in a top view in Figure 9.

EP0302192
Device for the treatment of living beings and plants with pyramid energy.

Inventor(s): OEHME RAINER

In a device for the treatment of living beings and plants with pyramidal energy, the energy is picked up on energy pick-up lines (12, 13), which are arranged electrically separate along the four edges of the pyramid (6). Since the emission of energy is particularly strong along the pyramid edges, a particularly large amount of energy can be picked up by energy pick-up lines (12, 13) of this type. Such energy pick-up lines (12, 13) are considerably cheaper to produce.

DE202006015477

Energy twin-pyramid to measure cosmic and terrestrial energy has mountain crystal energy condenser

Inventor(s): GROTHOFF HEINZ HUBERT

An energy double-pyramid with mountain rock crystal energy condenser has two discrete, independent pyramid bases (1, 2) without an electro-galvanic link.

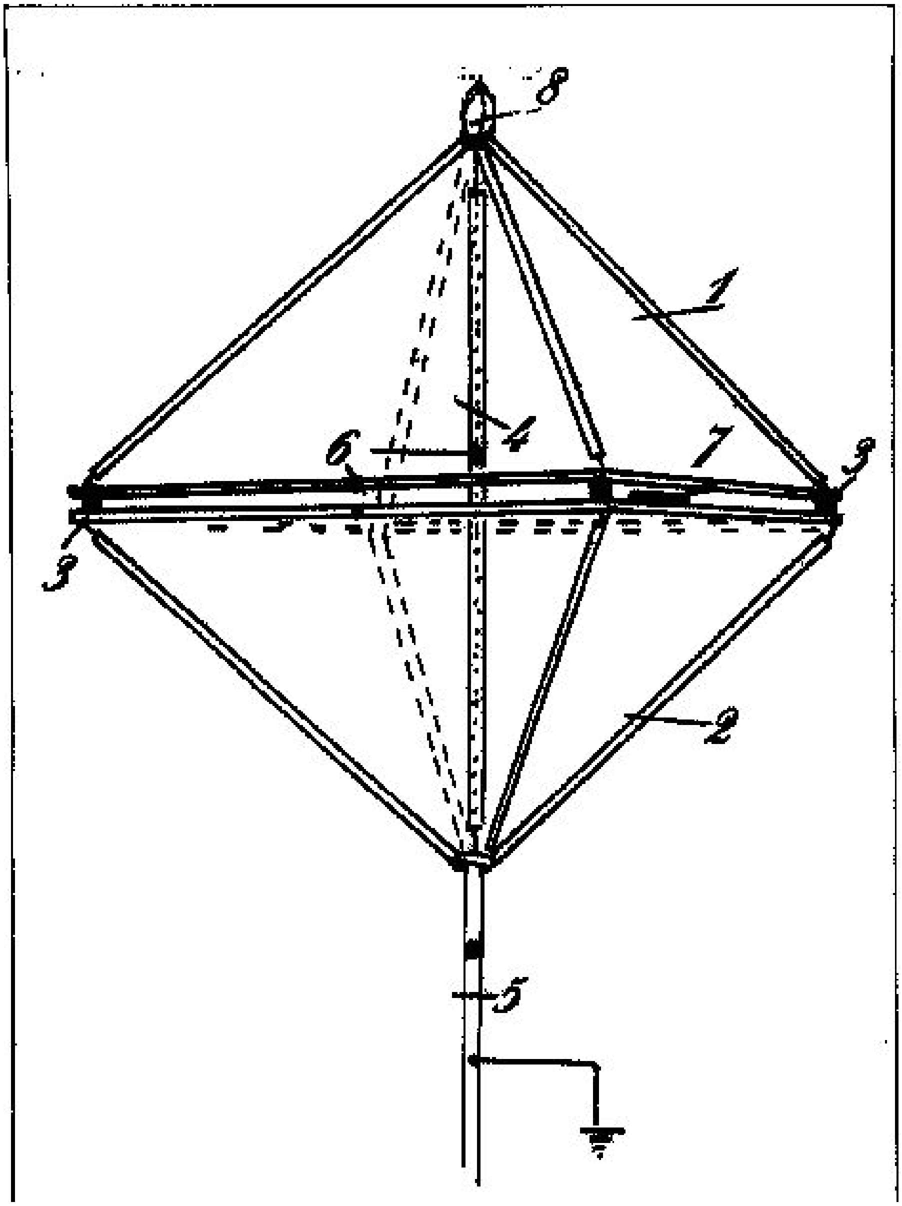

A separation and measurement of terrestrial and cosmic energy is therefore not possible.
The protection specified in claim 1 invention has the object to produce such a single measurement of cosmic and terrestrial energy on the creation of an energy-double pyramid of metals of all kinds and sizes.
This problem is solved with the features listed in the protection Claims 1 to 6.
With the invention ensures that a flow of energy in the mV range can be detected by the mechanical connection of the individual pyramids, but with electro-galvanic separation of the two basic structures of individual measuring points with plug and socket with electronic measuring instruments.
This both the kosmischenals on the terrestrial side of the two pyramids.
The finely-energy portion of the energy is dissipated in the so-called energy capacitor, with a brief break in rock crystals of 1-3 cm diameter filled plexiglass tube, this energy capacitor is mounted in the inner tips of the two pyramids.
Also on energy capacitor has a socket for tapping the energy is applied.
At the intersection of the two pyramids a plate for placing objects attached to energize.

The overall design is then on the top of the lower pyramid.
At the apex of the upper pyramid a natural quartz crystal is attached.
The bulk of the energy-double pyramid always correspond to a reduction corresponding to the masses of the Cheops pyramid.
Over the two sockets but also the electrical connection between the two pyramids is possible and the result is a big power cage.
The final energy double pyramid is shown in the protection claim 1 in the image and in the drawing.
1: upper pyramid base 1
2: lower pyramid base 2
3: 4 insulators
4: Rhinestone energy capacitor
5: grounded standpipe
6 sockets
7: Storage Plates
8: rock crystal tip

DE202006004778
Pyramid with square base containing special energy level...

Inventor(s): NOWACK REINHARD

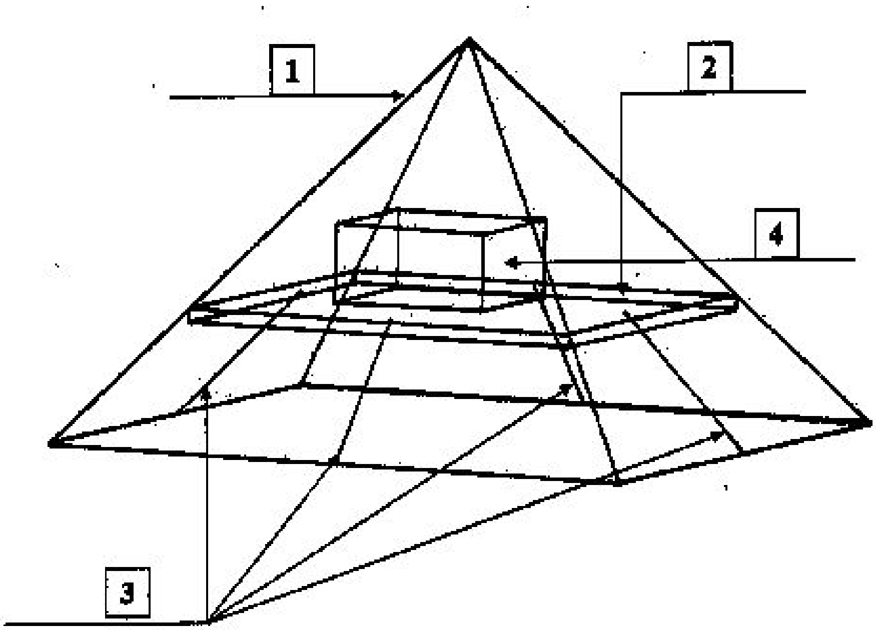

The focusing bars (3) serve as a holder for the energy platform (2), whilst simultaneously focusing energy on the object (4) to be energized.

DE102005052831
Gravitational energy conversion device, has pyramid produced from cuboidal stones...

Inventor(s): HUBER KLEMENS

The device has a pyramid used with a quadratic base surface of a side length and a vertical height (hy) in a center. The pyramid is produced from cuboidal stones, where a cuboidal-shaped hollow space is designed with volume (V1) in the pyramid. A well-shaped hollow body is designed with outer volume and inner volume in the pyramid. The hollow body is adjusted on the volume, such that the electrical energy that is provided in the hollow body is converted into gravitational energy.

The invention relates to a device with a pyramid according to the preamble of claim 1.
A variety of such pyramids there are on earth.
Their intended purpose and their age are controversial among archaeologists and researchers pyramid.

State of the art

The starting point of the present invention is to **DE 42 33 678 A1**.
There, the oscillation energy of an electron is equated with its electrostatic energy ECOUL.
Conversion factor is the Sommerfeld fine structure constant [alpha].
In the resonant energy product [omega] is 1 <2> V1 contain.
The part of the electron volume V1 is determined from the known angular frequency [omega] 1 and other known factors.
In the second equation, the product [omega] 2 <2> V1 appears; is the angular frequency [omega] 2 is calculated from known V1 and other factors.
The invention is now expected that applies to energy transfer between two bodies with the volume V3 and V4: [omega] 3 <2> V3 = [omega] 4 <2> V4
Harmonious relationships exist when V3 and V4 and [omega] 3 <2> and [omega] 4 <2> primes or prime products.
So there is a certain extent square harmony.
In contrast, the sound frequencies f of the music and its scales are linear or numbers
Number quotients dependent.

Task

It is the object of the invention to convert gravitational energy directly into electrical energy.
To achieve this objective the characterizing features of claim 1.
To explain is to serve subsequent calculation.
To achieve this objective the characterizing features of claim 1.
To explain is to serve subsequent calculation.
If the Earth rotates daily on its axis, acting on the electrons of a pyramid built on her a
Corriolisbeschleunigung acor the size acor = 2 wrel [omega] S
[Omega] S = angular velocity of the solar system sphere
<Img class = "EMIRef" id = "027964757-00020001" />
RES = mean distance Earth - Sun
RES = 1.495921 10 <11> m
wrel = relative velocity of the earth to the sun sphere
<Img class = "EMIRef" id = "027964757-00020002" />
From this:
<Img class = "EMIRef" id = "027964757-00020003" />
where for the determination of the rE relationship:
<Img class = "EMIRef" id = "027964757-00020004" />
And hy <3> according to claim 1: hy <3> = 9 X 11 X 29 X 37 X 299 521 X 10 <-4> m <3>
On the electrons then the moment acts:
Dcor = acor me RB
me = mass of the electron
= 9.10953 10 <-31> kg
RB = Bohr radius = distance of the electron from the proton
<Img class = "EMIRef" id = "027964757-00020005" />
From this:
Dcor = 1.476049 10 <-43> MN
Dcor <2> = 2.17872 10 <-86> (Nm) <2> ECOUL 10 <-68>
ECOUL = 2.17872 10 <-18> MN
= Electrostatic energy between proton and electron
The process of power generation is similar to known generators.
There, an armature rotates in a magnetic field and induces a sinusoidal alternating current by cutting the

lines of force of the magnetic field.

In the invention, the pyramid intersects on a daily rotation of the earth's gravitational field caused by the Sun-Earth and Corriolisbeschleunigung also a sinusoidal alternating current.

For this purpose, the pyramid height hy hy or <3> is matched in a particular way to a particular sphere volume of the earth with radius rE.

The advantage of the invention is a low-cost electricity without polluting the environment.

An energy conversion can be carried out according to the features of claim 4 in the reverse direction.

The occurring there factor [alpha] is an indication that electrical energy is converted into mechanical vibration energy.

It is

<Img class = "EMIRef" id = "027964757-00030001" />

Located in the trough-shaped hollow body a person with suitable Body and Brain volume, the energy of his brain waves is enhanced by conversion to gravitational energy.

His thoughts can be transmitted over long distances.

Embodiment

In the drawing the invention is shown schematically.

FIG. 1 middle section of the pyramid

FIG. 2 Larger view of the cavity with volume V1 and inside it hollow body with external volume V2.

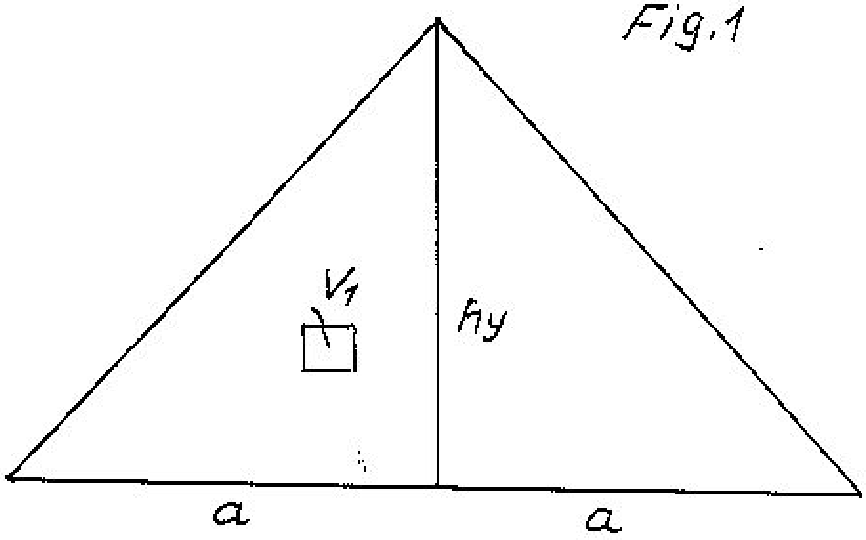

Fig.1

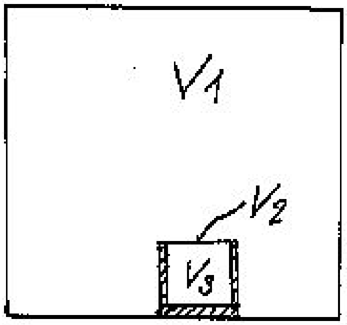

Fig.2

The internal volume V3 of the hollow body is half as large as the external volume V2.

It is important for the function of the device that the volumes in the construction of the pyramid are adhered to as closely as possible.

The pyramid must have so much mass that change the volumes at different daytime and nighttime temperatures do not markedly.

Seasonal changes in volume are kept within limits when the pyramid is built near the equator.

DE19717053 / DE29724691
Energy pyramid for medical use

Inventor(s): AUER ARMIN C

FiGUR (1)

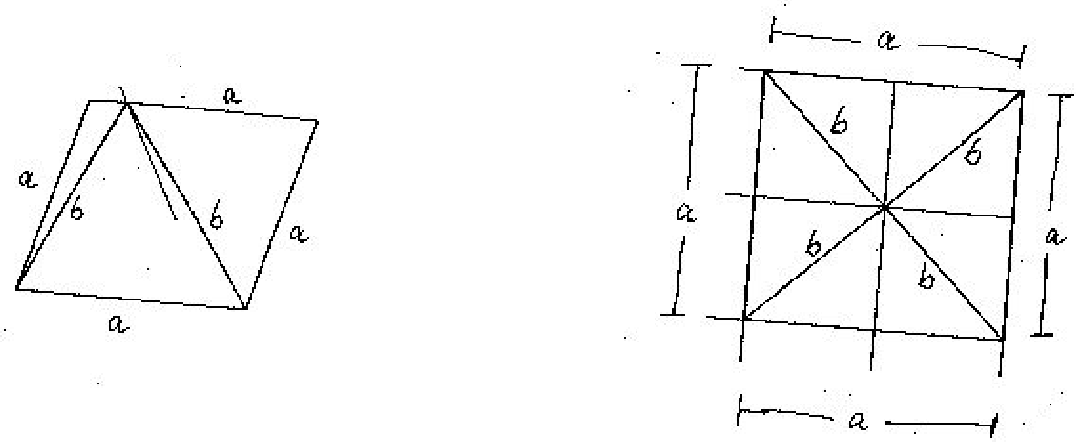

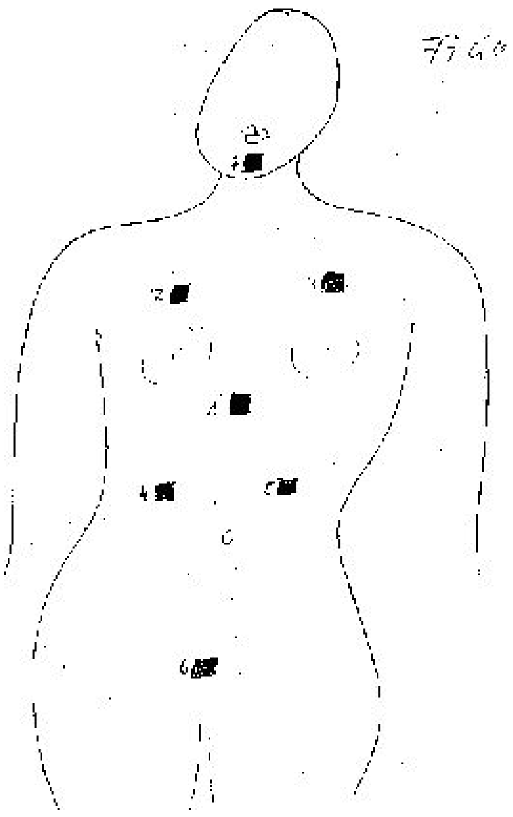

The aluminium or aluminium alloy pyramid has a base with 3-5 corners and equal or non-equal side lengths of 1-10 cm. Preferably the aluminium alloy has the composition 80-96, especially 87.4% Al, 5-15,

especially 9% Si, 1-6, especially 3% Cu, 0.08-1.0, especially 0.3% Mn and 0.08-1.0, especially 0.3% Mg.
Also claimed is an arrangement of several such pyramids.
Pyramids of specific size and alloy
Pyramid or Pyramids special size and special aluminum alloy for power charging
The use of pyramid energy is common and known.
The use of pyramid energy is not unknown especially in the medical field.
The achievement of specific effect has not been possible because the pyramids used hitherto produced only an insufficient energy.
The problem of sufficient energy accumulation is achieved by the features listed in the protection claim 1:
The dependent claims represent advantageous embodiments of the invention.
An arrangement according to claim 6 is particularly noteworthy.
With the invention is achieved that the pyramid energy specifically in metabolic diseases and especially in diseases with mental component (including Allergies) can be used.
The necessity of certain sizes and certain alloy was determined empirically.
During these investigations were during the run of these pyramids to specific points on the patient for skin.
As severe pain and other clearly identifiable reactions observed, which led to cures.
The representation of a particularly advantageous embodiment of the invention is shown in Figure 1.
An inventive arrangement of several pyramids according to the invention is the Figure 2 below.
Inverted pyramid with a square base with
= 2.26 cm
0.951 x
advantageously aluminum alloy according to claim 4
Description 2
Innovative arrangement according to the invention of several pyramids

---

CN1617188
Pyramid energy focusing model and its making method

Inventor(s): LI YUANLIANG [CN]; LI YUJIAN [CN]; ZOU QIN

The energy focusing pyramid model is made of plastic, three-ply board, fiber board, granite and other insulating material and is shaped as rectangular pyramid or triangular pyramid with each side plane is one isosceles triangle of apex angle of 52 deg and base angle of 64 deg. It may be set inside bed room, office and class room to result in body building, disease preventing and treating effect, and entering to one large model will result in even high effect. Food or beverage may be set inside the model for preserving even long period. The reason is that the model can well absorb cosmic wave energy.

Your Support Maintains This Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...

Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

rexresearch.com

Pyramid Energy Patents

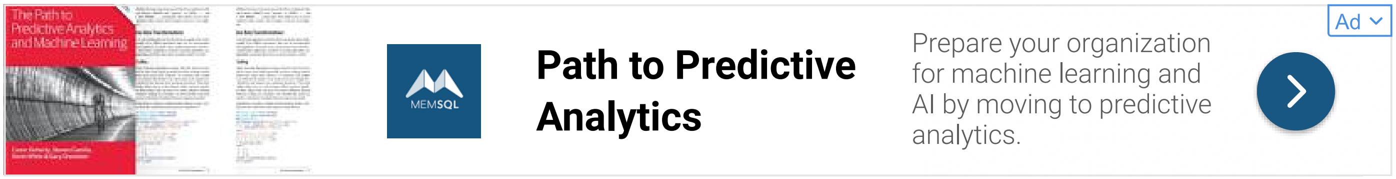
Path to Predictive Analytics

Prepare your organization for machine learning and AI by moving to predictive analytics.

Ad
The Path to Predictive Analytics and Machine Learning
MEMSQL

fetchingstock
shutterstock

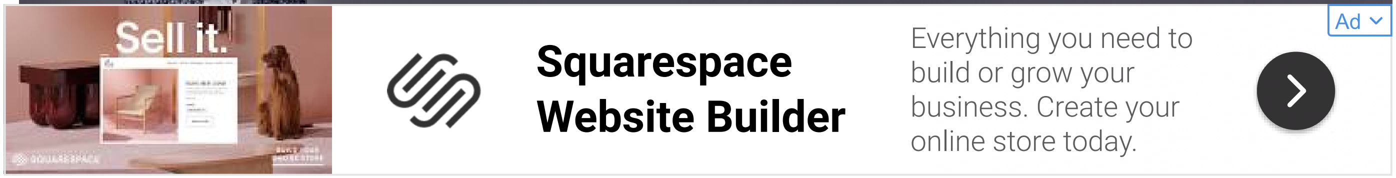
Squarespace Website Builder

Everything you need to build or grow your business. Create your online store today.

Ad
Sell it.
SQUARESPACE

CN101703442
Pyramid type magnetic field effect health-promoting and disease-curing functional bed

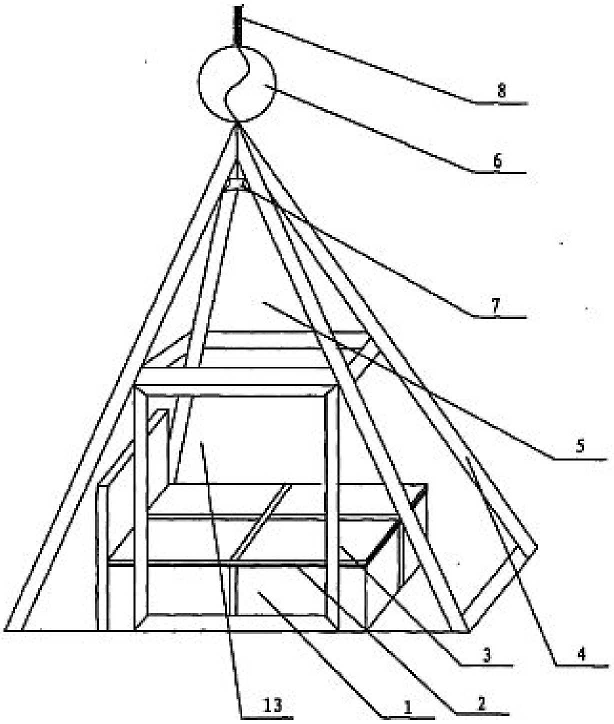
8
6
7
5
3
4
13
1
2

图 1

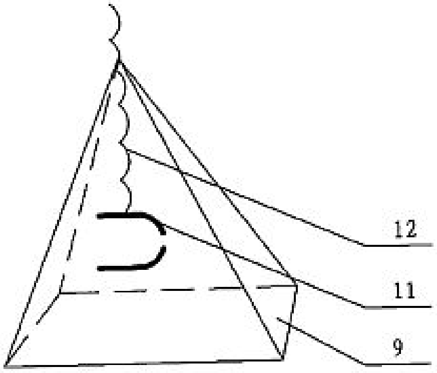

图 2

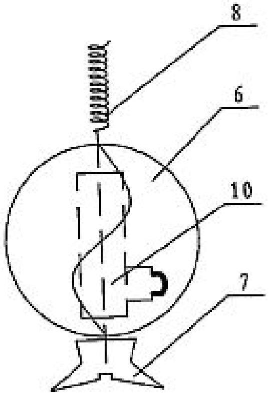

图 3

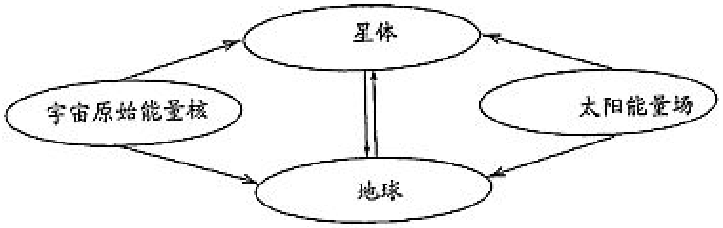
星体
宇宙原始能量核
太阳能量场
地球

图 10

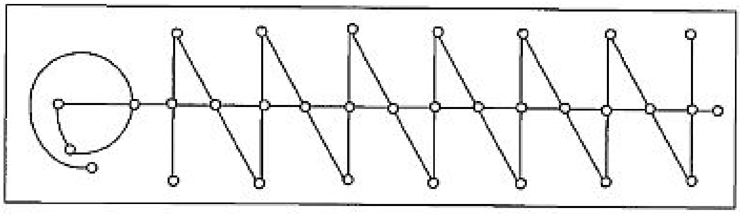

图 4

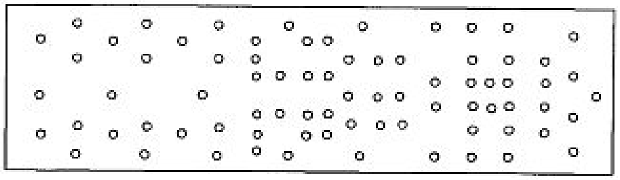

图5

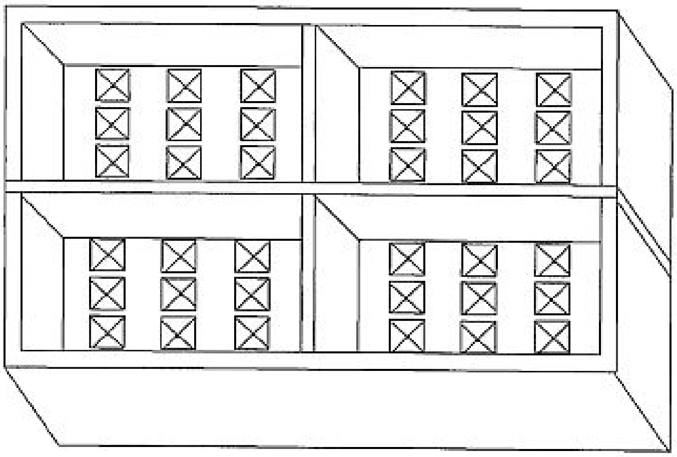

图6

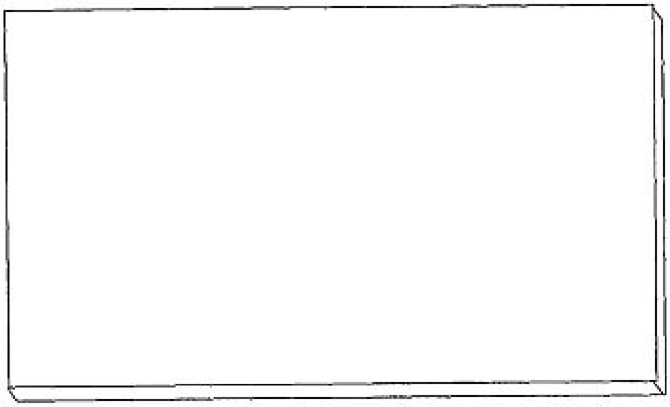

图7

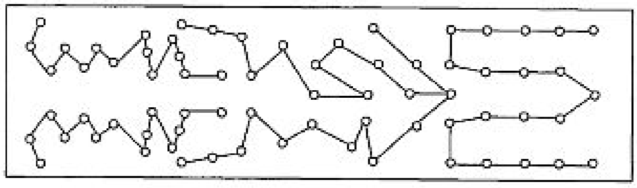

图8

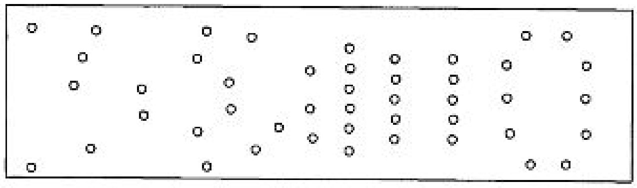

图9

The invention relates to a pyramid type magnetic field effect health-promoting and disease-curing functional bed, aiming to satisfy the demand that people increasingly pay attention to the quality of life, do not satisfy the current life expectancy and seek good health and a long life. A group of terrestrial magnetism energy concentrators are arranged in an existing bed body; magnets which are connected together through magnet-wire printed circuit boards are distributed on the cover board of a box body below the mattress according to the human body shape and acupuncture points; the bed body is arranged in a pyramid type magnetic field effect device; the device is a tower-shape frame with four equicrural triangle-shaped side faces which are made from four metal sections, transparent bodies are arranged in the tower-shape frame to form a functional cover, a spherical magnetic receiver and a magnetic energy releaser are arranged on the top of the cover; and an open inlet is arranged on the side of the functional cover. The characteristics and beneficial effect of the functional bed are as follows: by using the geometric energy field, geomagnetic field, astrology and star gravitational field theories of the Taiji pyramid building, the energy can be concentrated and act on the human body so as to have function of curing human diseases; and a plurality of clinical tests show that the functional bed has unique effect for curing various difficult and complicated cases.

KR20100005740
GI ENERGY SEX DEVICE

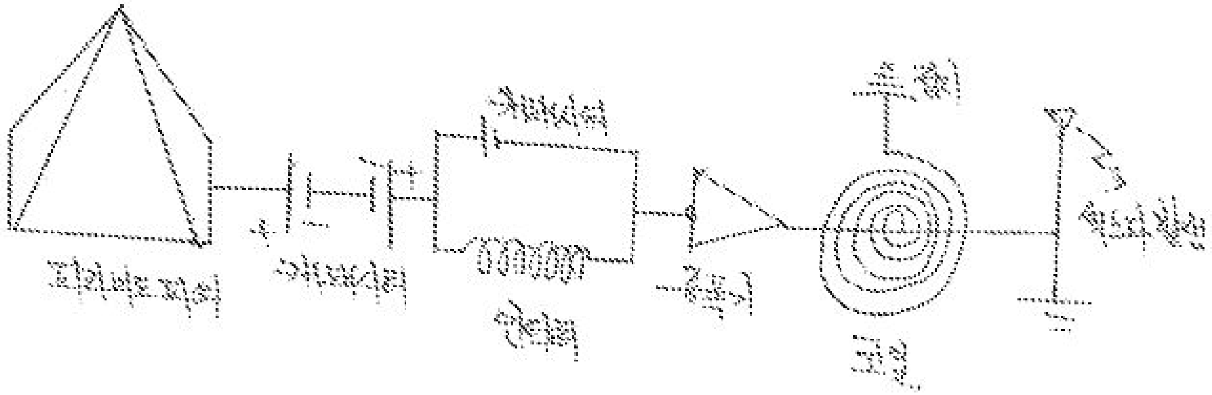

PURPOSE: A pyramid energy sex apparatus is provided, which makes penis erected in a few minutes using energy of the pyramid. CONSTITUTION: A pyramid energy sex apparatus comprises a pyramid; a copper wire connected to the bottom side of the pyramid; a first capacitor connected to the bottom of the pyramid in the plus direction; a second capacitor in which the plus direction is connected to the frequency tuned part; an inductor which is parallel-connected to the capacitor and performs frequency tuning; an amplifier amplifying the frequency tuned energy; and a coil which is parallel-connected to the amplifier, generates sex wave, and is conversely wound counterclockwise.

KR20100005739
GI ENERGY SLEEPING DEVICE

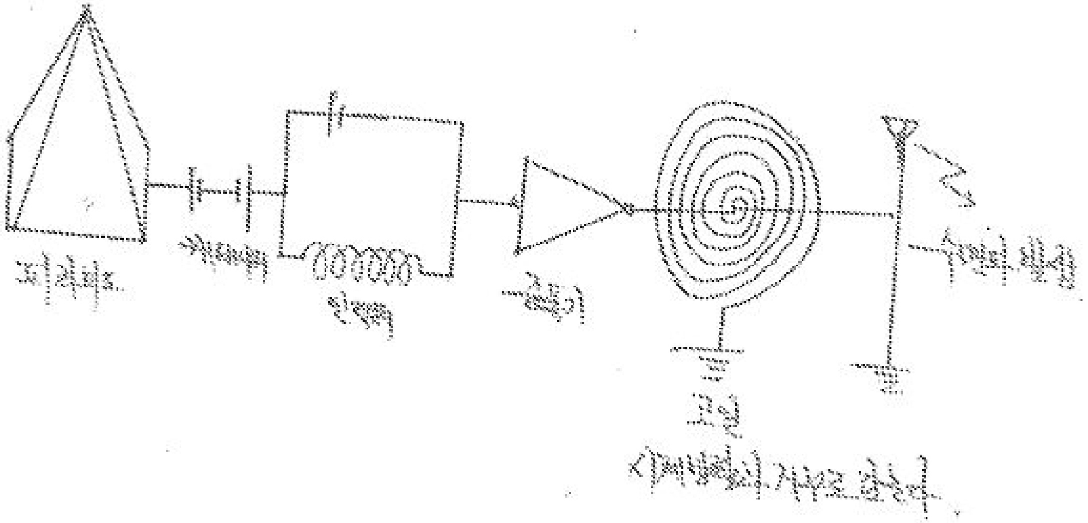

PURPOSE: An energy sleeping apparatus is provided, which makes a user sleep in a few minutes using energy of pyramid power. CONSTITUTION: An energy sleeping apparatus comprises a pyramid; a copper wire connected to the bottom side of the pyramid; a first capacitor connected to the bottom of the pyramid in the plus direction; a second capacitor; an inductor which is parallel-connected to the capacitor and performs frequency tuning; an amplifier amplifying the frequency tuned energy; and a coil which is parallel-connected to the amplifier and is conversely wound counterclockwise.

UA32342
METHOD FOR ENERGY-AND-INFORMATION TREATMENT OF FLUID

A method for energy-and-information treatment of the fluid comprises storing the fluid inside the container with the cover designed as the regular rectangular pyramid and providing the exposure of the fluid inside to the fields of the regular rectangular shape.

UA32341
DEVICE FOR ENERGY-AND-INFORMATION TREATMENT OF FLUID OR FLOWABLE MATERIAL

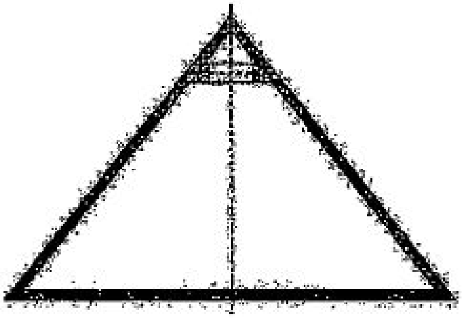

Фіг. 1

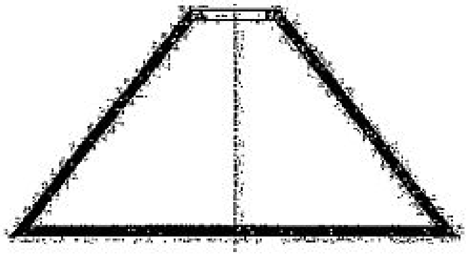

Фіг. 4

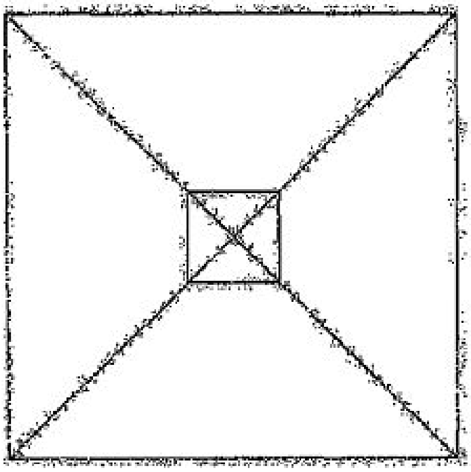

Фіг. 2

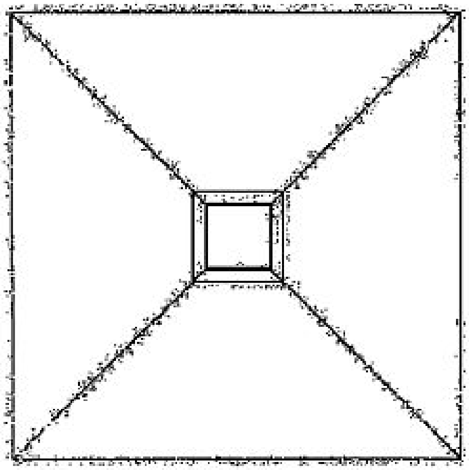

Фіг. 5

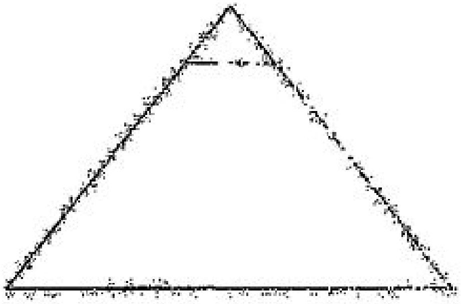

Фіг. 3

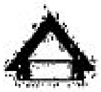

Фіг.7

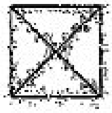

Фіг. 8

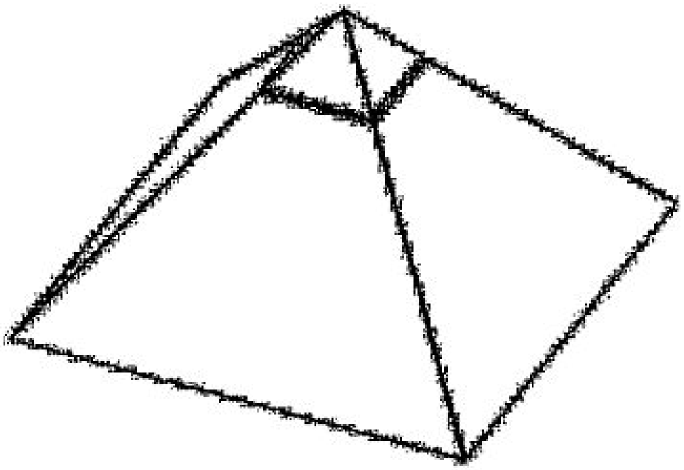

Фіг. 9

Device for energy-and-information treatment of the fluid or the flowable material contains the container with the cover in the shape of the regular rectangular pyramid for storing the material and providing energy-and-information treatment.

UA29223
TANK FOR IMPROVEMENT OF WATER ENERGY PROPERTIES

A tank for the improvement of water energy properties corresponds a housing made of dielectric material in the form of a nine-hedral regular truncated pyramid with an open upper base.

UA28144
FACET OF AN ENERGY-EMITTING PYRAMID "KHVYLIA"

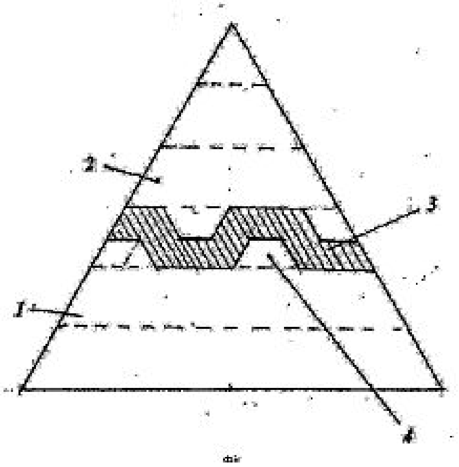

It is proposed the design of a facet of an energy pyramid with increased intensity of emitted energy.

UA28143
FACET OF AN ENERGY-EMITTING PYRAMID "STRILA"

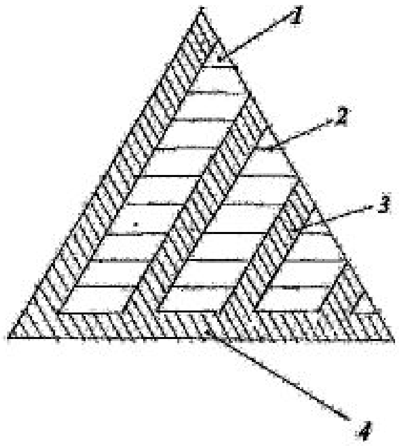

It is proposed the design of a facet of an energy pyramid with increased intensity of emitted energy.

UA28813
ENERGY FACE OF A PYRAMID "KATAMARAN"

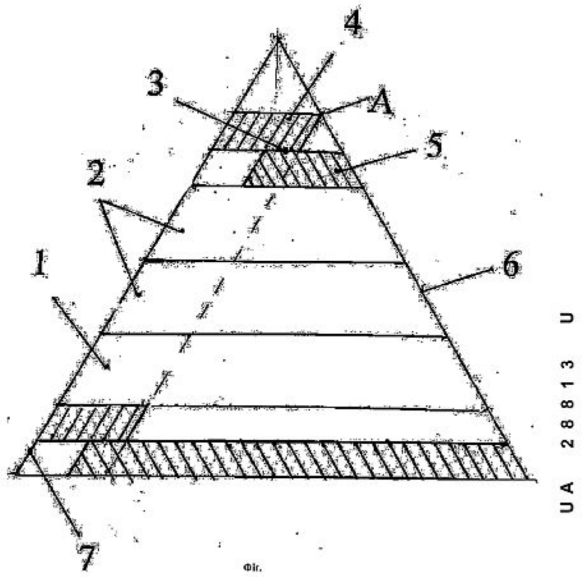

The proposed face of an energy pyramid contains parallel horizontal lines. Between the lines, a strip with metallic coating is arranged. The strip forms a projection that is arranged in parallel to the edge of the pyramid face.

RS943
PYRAMID ENERGY UNIT

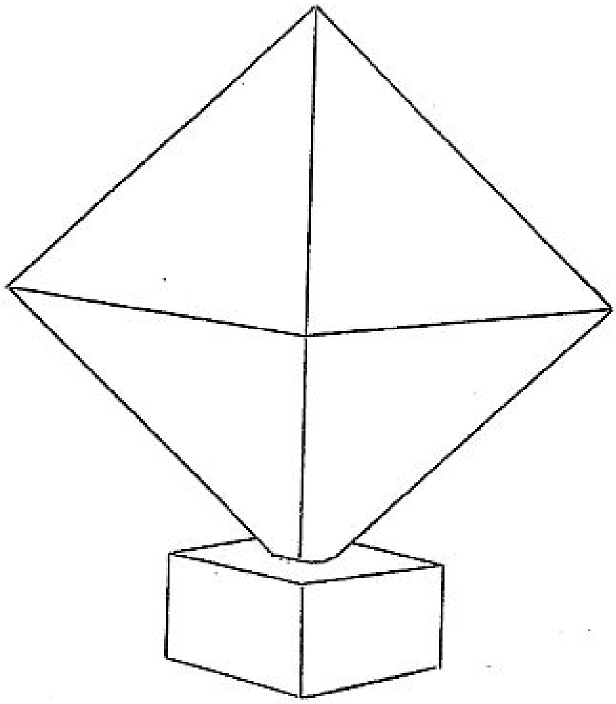

SLIKA 1

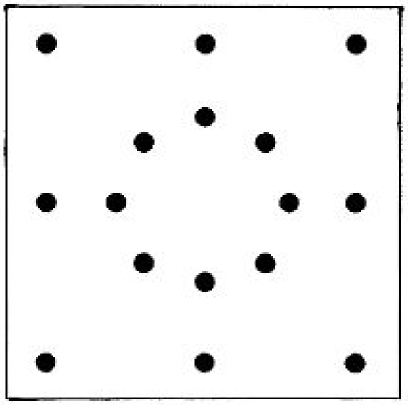

SLIKA 2

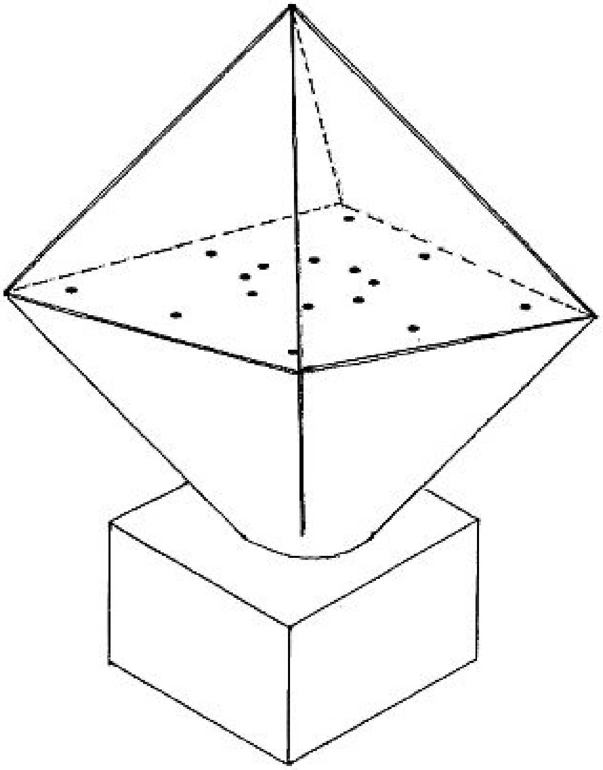

SLIKA 3

Abstract not available for RS943

UA27840
RESERVOIR FOR IMPROVING ENERGETIC PROPERTIES OF WATER USED FOR PLANT

IRRIGATION

A reservoir for improving the energetic properties of water used for plant irrigation is designed as nonahedron with the walls made of dielectric material. The reservoir represents the regular truncated pyramid with the open upper base. The ratio between the diameters of the circumcircles of the upper and lower bases is 0.19-0.22:1. The ratio between the height of the reservoir and the diameters of the circumcircle of the lower base is 0.7-0.9:1. The angle between facet of the nonahedron and its base is 61 DEGREE -67 DEGREE . The device is used for improving the energetic properties of the water used for plant irrigation due to the energy of the pyramidal construction.

WO2008094038

DEVICE FOR GENERATING AN ENERGY FIELD AND A SYSTEM OF SUCH A DEVICE AND ONE OR MORE TREATMENT INSTRUMENTS

The present invention relates to a device (1) for generating an energy field, comprising a first housing (10) in which one or more pyramid-shaped elements of electrically conductive material and one or more coils are arranged. The housing is provided with a connection for a voltage source (20) to the pyramid-shaped elements and/or the one or more coils. The housing is further provided with means (30, 110, 120, 130, 140, 150) for transferring the energy field to a living organism, for instance a human being. The invention also relates to a system (100) of a device (1) according to the invention and one or more treatment instruments (110, 120, 130, 140, 150) connected to the device, wherein each of the treatment instruments is adapted to transfer the energy field onto a living organism, for instance a human being.

AU2007200359

Pyramid energy devices

Invention

The energy concentrating effects of pyramids has long been known.

According to the currently accepted supposition the common resultant of cosmic space, terrestrial space, electric, magnetic, corpuscular and other, thus far unknown particles and energies could be responsible for phenomena taking place in pyramid shaped spaces.

The closest, to the offered invention is the device in a general concept, considering the principle of effective concentration and generation of pyramid energy and principle of effective distribution of this energy by spiral (INSTRUMENT AND PROCEDURE FOR THE USE OF PYRAMID ENERGY, application AU2002358920 (Publication date is 20 04-07-22) (WO2004058339 Application number: AU20020358920 20021231; Priority number WO2002HU00179 20021231, IPCI-7: A61M21/00, A61N1/16, A61N1/00).

In this general concept a separate device from system of the same devices in the specified application represents a figure of the pyramidal form with the flat bases. The description of separate devices of the pyramidal form, from the system does not contain data about concrete specific constructive attribute of realization of the device. Specification of the prototype is the spiral system of pyramidal devices, and not a separate device, and influence of system on bioobjects through saturation of water contained in them by pyramid energy.

The substance containing water may ideally be the human body, tea, fruit, vegetables, juices etc. Substance containing water thus treated may be used for treating human and animal sicknesses. However, influence of pyramid energy on different objects can be not only curing but also intimidating.

A method of protection against rodents and insects with its application Technical field of the invention

The invention relates to the field of devices of generated concentration of energy which can be used for correction of energy states of bioobjects in various industries, agriculture or domestic situations, in particular, the protection of premises (shops, warehouses, eating establishments) from rodents (rats, mice) and insects (ants, termites, cockroaches, flies) by means of detracting and/or destroying them.

Summary

Task of the invention is the expansion of functions of the device for generated concentration of pyramid energy by effective energy influence on different objects and including the application for effective protection of premises and their contents against destruction by rodents and insects.

00 To fulfill this task, the essential attribute of the offered device is the usage of r, rigidly connected pieces of electric wire for creation of the symmetric pyramidal form of the device, arrangement of the Reich's energy accumulator executed in the form of sphere inside the pyramidal form on the central vertical axis and arrangement of sphere of amplification and optimization of energy radiation on top of the pyramidal form connected with Reich's energy accumulator by electric conductor.

Variants (embodiments) of the pyramidal form the device can be executed:

c,1 In the form of a correct tetrahedral pyramid enclosed inside of an additional correct tetrahedral pyramid with the common plane of the bases and with the common top of both pyramids, and the attitude of the areas of the bases of external and internal pyramids make 2:1, and the center of sphere of Reich's accumulator is located on 1/3 height of pyramids from their bases;

The figure of the pyramidal form is formed by a ring of wire, forming the base, three pieces of wire have two parts, the first part represents the spatial spirals located on a surface of part of ellipsoid with a direction of coils counter-clockwise, and the second part represents flat spirals inside of the ring base, converging in its center, and points of interface of spatial and flat parts are located on the base ring on angular distance 120 degrees from each other, and angles of twisting of flat parts of spirals and the angles of twisting of projections of spatial parts on a plane of the base are equal to each other and made from 0.5 up to 1.5 angular turnovers around vertical axis, the center of the sphere of Reich's accumulator is located on half of the height from the top to a plane of the base.

By such embodiments of the device inventors made purposeful influences on objects: improving energy fields of people and foodstuff while dispersing small rodents and insects.

For realization of the method of protection against rodents and insects, the devices are kept on the protected premises whenever possible and at a convenient distance without interruption to the user. The necessary quantity of the devices depends on the size of the protected area with an approximate ratio of overall dimension of the base of the device of 10 cm for 3.5 m length of floor of an area.

Brief description of the drawings

FIG. 1. Shows a first embodiment of the device.

FIG. 2. Shows a second embodiment of the device.

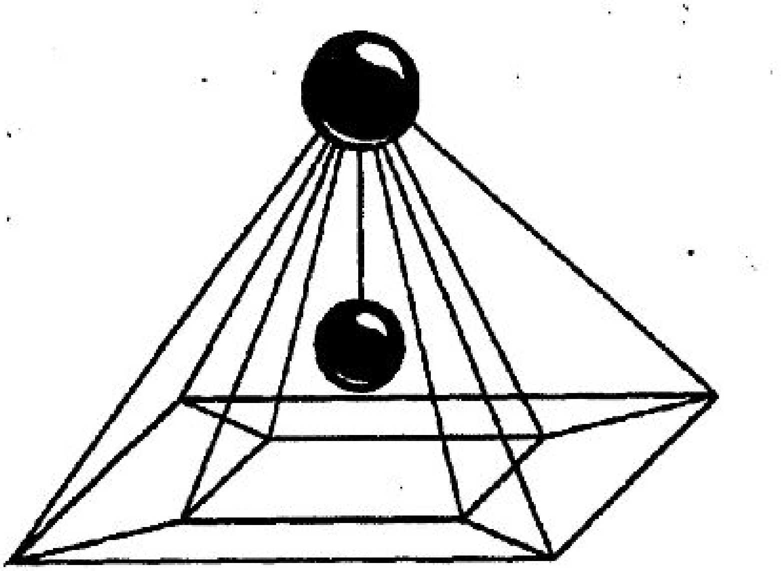

FIG. 1.

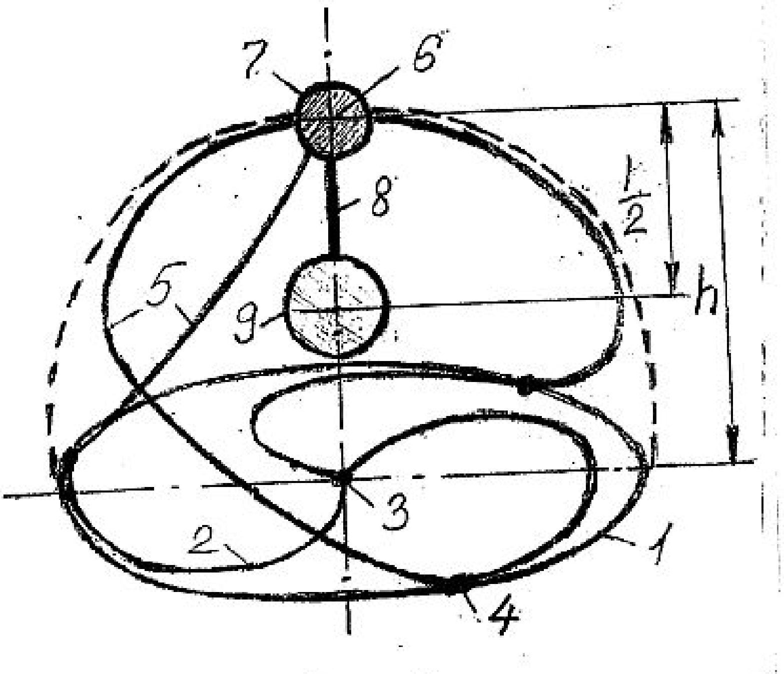

FIG. 2

Detailed description of the preferred embodiments:

00 Main principle of application of the offered devices is the definition of influence of a level of energy initially existing in object and demanded effect of a so-called resource level necessary for achievement. Further means of the t devices of the necessary size and quantity, the demanded resource level of energy is created for the decision of tasks in view.

ri Effective generation of pyramid energy is impossible without its effective concentration in the device.

Reich's accumulator is used for amplification Seffect of concentrated energy, and the sphere on top of the pyramidal form of N the device is used for optimizing the energy radiation. It is known, that Reich's accumulator acts in a role of the receiver-condenser of energy from the outside that through the sphere on top of the pyramidal form device forms the concentrated field around the device. Reich made metal and wood boxes cosmic orgone energy accumulators (in more common case accumulators with metal and dielectric layers). So in our case, Reich's accumulator represents multilayered structure of dielectric and metal layers, with an external dielectric layer (for example, a dielectric polymeric film) and an internal core from metal (a steel ball). Reliability of wide energy opportunities of Reich's accumulator proves to be true documentary, and its action continuum. The sphere of amplification and optimization of radiation of energy represents a ball from the metal foil, which has been wound up around the top. The effect of optimization of radiation of energy is represented in more harmonious tranquility radiated energy above and along the surface of the device.

The first embodiment of the device has the traditional form of correct pyramids with an inclusion of a smaller pyramid inside the larger one also for amplification effect of concentration and generation of pyramid energy. The second embodiment of the device with the smoothed pyramidal form of ellipsoid realizes the spiral principle considering close interrelation of Space and Time in existential thermographicaly objectificated, thus in additional influence on a person there is a strengthened component of adjustment (objective parameters of state), time rhythms of the person, in particular, his sleep as essential part of a daily regime.

Practice has shown that small rodents and insects do not remain in the invaded areas due to correcting surrounding space energy concentrated and generated by the offered devices, as the size and capacity of their own energy fields of rodents (rats, mice) and insects are much less than sizes of area and capacity of the energy generated by the offered device. For example, rats have field size of cm (mice and insects as essences of the smaller size, than rats, have the sizes of their fields naturally also much less) in comparison with approximately 25 m of the size of area of concentration of energy by the device with length of sides of the bases 70 cm.

O The method has many obvious advantages, especially important in the food 0processing industry: ecologically pure and safe, without application of harmful pesticides and radiations harmful to people; rodents and insects fail to build immunity to the influence of the devices; the task of installation of the offered systems in remote places do not require any technological process or electrical 00 installation. Proofs of efficiency of application of the offered prototype devices and the method are given in the following examples, which proved to be of positive benefit to the user:
a) After visiting a veterinary clinic with a "sick" gang gang, the diagnosis revealed an incurable virus found quite common amongst these birds.

Had no appetite, loss of feathers, beak breakage and very poor Scondition. After introducing the pyramid devices for two or three months, the bird's improvement was very noticeable by his condition, his "hungry" appetite, growth of feathers and mended beak.

b) A young couple in the midst of a marital break up mended their relationship with the introduction of the spiral device. It provided a much more relaxing atmosphere in the home, taking away the tension between them. They also noticed the disappearance of ants after several days that had been present earlier. All the above was achieved within one or two months.

c) The pyramid device was introduced to a food preparation and take away establishment. The proprietor was an experienced businessman and well established but disharmony within the establishment was evident. Within two or three months of the device being installed, the proprietor became much more mellow and staff increased productivity without any tension. Business takings increased, insects disappeared followed by a very pleasing event that never experienced by the proprietor before, and he received a cheque in the mail.

d) A friend installed the device due to irritability and "nothing going right". Within two or three months of the installation, he bought his first home for a good price, he sleeps better, his plants are growing very healthy, all visiting relatives staying overnight remark on their restful sleep, better fuel economy running his car and no insects.

e) A relative experiencing an on going health problem with his heart, stomach and poor circulation in the left leg installed a pyramid device to help with sleep. After several months, his cardiologist commented on his heart function being better than usual after a regular check up, although nothing was mentioned about the device. The left leg is functioning and looks better while the stomach is gradually improving. Food stores noticeably longer and better, very evident by the disappearance of flies and ants or other insects.

EA200602270

METHOD FOR COMBATING RODENTS, COCKROACHES AND DEVICE THEREFOR

Also published as: EA008673

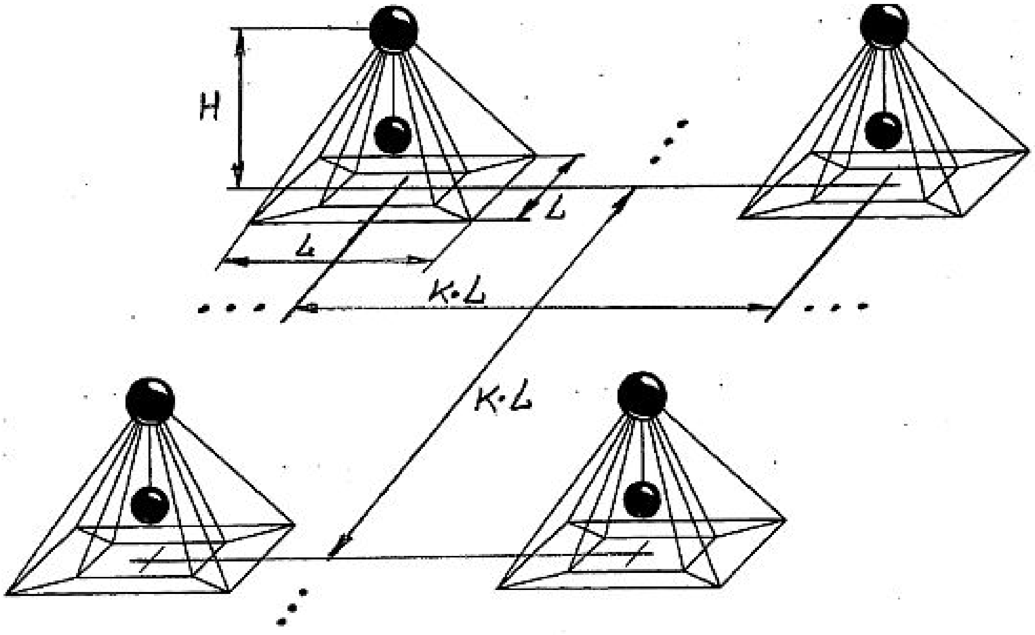

Фиг.

The invention relates to means of affecting upon rodents (rats, mice) and cockroaches for protection premises (shops, warehouses, eating establishments, rooms, etc) against them by scaring and/or killing. For better efficiency there proposed concentrated energy of electromagnetic field naturally induced in current-conducting material of formed wires. The inventive method and the device are aimed at providing an appropriate effectiveness of protection of food products and consumer goods stored in premises of any sizes against rodents and cockroaches. The method includes using electromagnetic field naturally induced in current-conducting material of formed wires cardinal oriented.; The method is characterized in that tetrahedral pyramids are preassembled from the formed wires, each of the pyramid has an additional tetrahedral pyramid of smaller size enclosed thereinto, both pyramids have common a base plane and a common pyramid vertex. A ball-shaped power amplifier is arranged in the pyramid vertex and coupled with a Reich storage battery made as a sphere and spaced about 1/3 of the pyramid height from the pyramid base to the center of the sphere of the Reich storage battery. The bases areas of the outer and inner pyramids are rated to about 2:1. The pyramids are cardinal oriented by their faces. A number of pyramids required for premises protection is calculated and constitutes 10 cm of the length of the outer pyramid base side per about 3.5 m of the premises floor length. The method is accompanied by a device therefore.

JP2008068056
ENERGY PATCH FOR REGULATING VITAL ENERGY AND BLOOD OF HUMAN BODY

Also published as: WO2008032890 (A1) US2008081940

【図7】

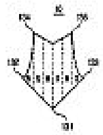

【図10】

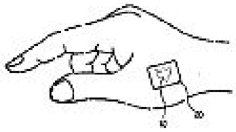

【図8】

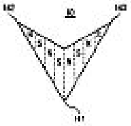

【図11a】

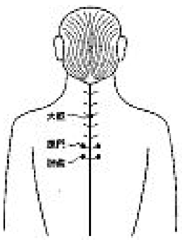

【図9】

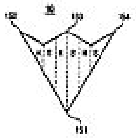

【図11b】

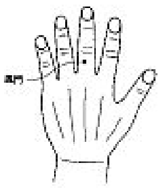

【図12a】

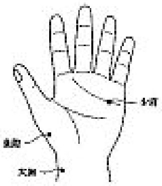

【図12c】

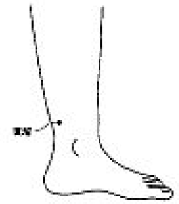

【図12b】

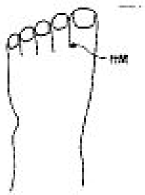

【図11c】

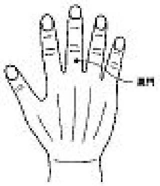

【図13】

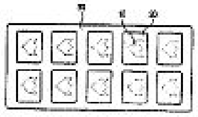

PROBLEM TO BE SOLVED: To provide an energy patch for regulating vital energy and blood, which is equipped with the concept of "Hosha" (a Chinese medical treatment for redeeming a sick person's physical strength and vigor and for expelling diseases from the human body) so that the energy patch can be applied to the meridian and meridian point regions of the human body and simply used. ; SOLUTION: This energy patch comprises: a rubber alloy magnet which is formed in such a planar shape as to have the characteristics of a permanent magnet equipped with a north pole and a south pole, which keeps a pointed corner directed downward, and at least a pointed corner upward; a color coating member which enables color printing to be applied to either of the sides of the rubber alloy magnet; and an adhesive member which enables the rubber alloy magnet to adhere to the skin of the human body. The energy patch can be applied to the understanding of the nature of the disease, the measurement and regulation of the excess and deficiency of the vital energy and blood, the prevention of the disease, etc.; and the pyramid effect of the shape of the energy patch brings about the effect of activating biological energy and treating the diseases. In the application of the energy patch, since the energy patch adheres to the specific region of the human body in consideration of the directional essence and color essence of life of the human body, the energy patch can bring about the effect of acupuncture concurrently with a magnetic treatment, so as to exert the effect of promoting the health of the human body by regulating the vital energy and blood.

KR20070122080
APPARATUS FOR GENERATING PYRAMID ENERGY USING PYRAMID STRUCTURE

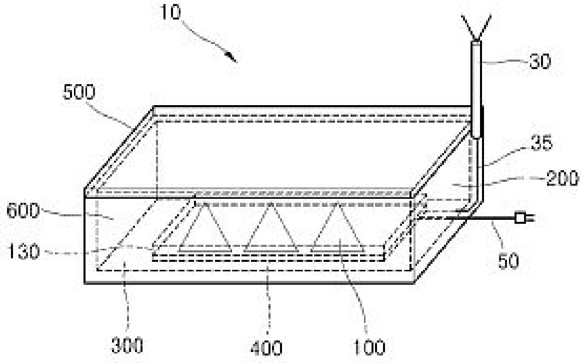

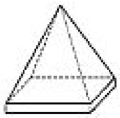

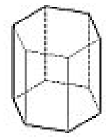

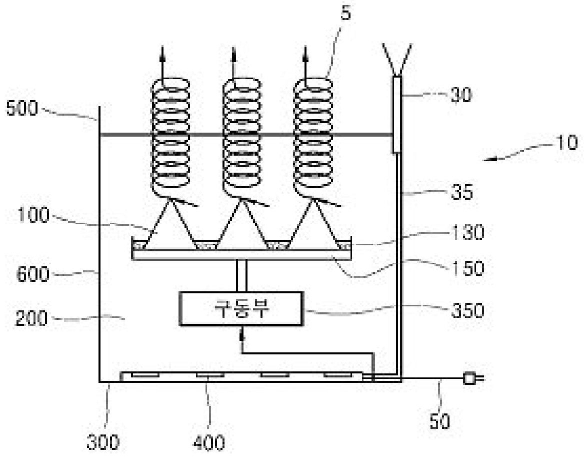
구동부

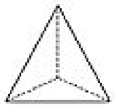

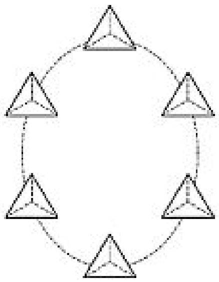

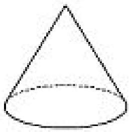

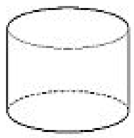

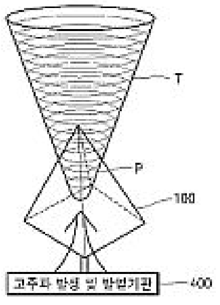
T
P
100
고주파 발생 및 발번기관
400

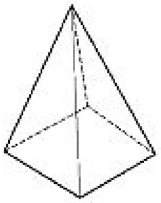

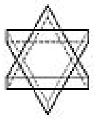

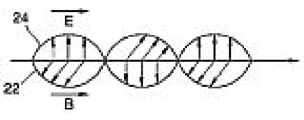
24
E
22
B

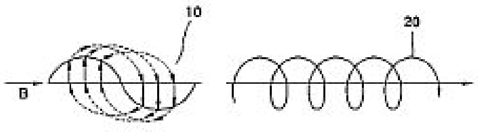
10
20
B

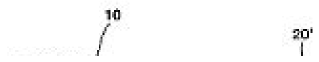
10
20'

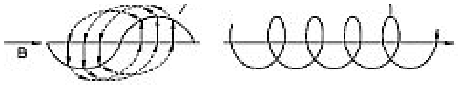

An energy generating device using a pyramid structure is provided to amplify pyramid energy of the pyramid structure by changing a magnetic field by applying high frequency to the pyramid structure. An energy generating device(10) generating pyramid energy in a pyramid structure(100) is composed of the pyramid energy generating pyramid structure; a case(200) containing the pyramid structure; a high frequency generating and heating substrate(400) installed at the inner bottom of the case and operated electrically or magnetically by receiving electricity from an external power source, in order to generate a high frequency signal and heat and amplify pyramid energy generated from the pyramid structure; and an object holder(500) formed at the outer upper part of the case to place an object and absorb pyramid energy generated from the pyramid structure stored in the case.

US2010031842

SPACE ENERGY IMPLOSION UNIT AND AN ENERGY AMPLIFICATION GENERATOR USING THE SAME

Also published as: WO2008041812 (A1) JP2010521415 (A) EP2086637 (A1) EP2086637

Abstract

Provided are a space energy implosion unit and an energy amplification generator using the same. The space energy implosion unit includes: a planar structure 1 having a regular pentagonal shape; a cubic structure 2 which is installed to be separated from an upper portion of the planar structure 1 and has a regular pentagonal pyramid shape; and a separation structure 3 which separates the planar structure 1 and the cubic structure 2 from each other and has a smaller area than areas of the planar structure 1 and the cubic structure 2. The energy amplification generator includes: a first geometrical structure in which vertices of each of five or seven pieces of planar structures 1 having a regular pentagonal shape contact one another; a second geometrical structure which is separated from an upper portion of the first geometrical structure and in which vertices of each of five or seven cubic structures 2 having a regular pentagonal pyramid shape contact one another; and a plurality of separation structures 3 which separate the first geometrical structure and the second geometrical structure from each other, are installed between the planar structure 1 and the cubic structure 2 and have a smaller area than areas of the planar structure 1 and the cubic structure 2.

TECHNICAL FIELD

[0001] The present invention relates to a space energy implosion unit which implodes a little energy that is full in the universe to be directly and indirectly used, and an energy amplification generator using the same.

BACKGROUND ART

[0002] Newton's Mechanics in which the nature of materials is defined as particles dominate classical scientific circles. However, as quantum mechanics and relativity theory are established before and after 1900, the nature of materials can be understood as particles and waves.

[0003] For example, it can be understood that an untouchable energy is changed into a touchable material by using equation (E=mc2). Also, as it is understood that, according to the Heisenberg uncertainty principle which is a basis for quantum mechanics, a particle position x and a momentum p cannot be determined simultaneously and are limited by mutual uncertainty ([Delta]x[Delta]p>=h 2[pi] (where, h is a Planck's constant), new thinking that particles as materials and waves as non-materials cannot be separated from one another has emerged. Thus, it is accepted to be obvious that the universe is comprised of materials and non-materials and materials can be changed into energies which are non-materials.

[0004] On the other hand, in early 1990, G. I. Shipov doctor (selected as one of 500 leaders who made the most affect on the 20thcentury, by American Biographic Institute, 1998) which is Russian scientist has found that an energy reacts with a particular geometrical shape. Such an energy field is referred to a third force which is not gravity or an electromagnetic force, i.e., a torsion field or a space energy. The torsion field (space energy) is mutual influence or a wide-meaning resonance effect that is originated from all materials, rotation of electricity or magnetism, and rotation of biomolecules.

[0005] Such a space energy is very little and is full in the universe. It is very little and is not measured using a conventional apparatus for measuring an electromagnetic field. However, the existence of the space energy is verified by a trace (an interference effect) that leaves in materials or an electromagnetic field.

[0006] The space energy is also applied to medicine. A new academic field such as quantum medicine which is a new medical treatment method has emerged. In advanced countries such as USA or France, enormous research expenses for studying an effect of a space energy on a living body have been supported.

[0007] In this way, study and development for using a space energy (a torsion field) generated in the universe filled with mutually-changed materials and energies by a particular geometrical shape in reality has proceeded in several countries.

DISCLOSURE OF INVENTION

Technical Problem

[0008] The present invention provides a space energy implosion unit in which a space energy full in the universe is effectively imploded using a particular geometrical shape so that the imploded space energy can be indirectly and directly used in reality, and an energy amplification generator using the same.

Technical Solution

[0009] According to an aspect of the present invention, there is provided a space energy implosion unit, the space energy implosion unit comprising: a planar structure 1 having a regular pentagonal shape; a cubic structure 2 which is installed to be separated from an upper portion of the planar structure 1 and has a regular pentagonal pyramid shape; and a separation structure 3 which separates the planar structure 1 and the cubic structure 2 from each other and has a smaller area than areas of the planar structure 1 and the cubic structure 2.

[0010] According to another aspect of the present invention, there is provided a space energy implosion unit, the space energy implosion unit comprising: a planar structure 6 having a regular heptangular shape; a cubic structure 7 which is installed to be separated from an upper portion of the planar structure 6 and has a regular heptangular pyramid shape; and a separation structure 8 which separates the planar structure 6 and the cubic structure 7 from each other and has a smaller area than areas of the planar structure 6 and the cubic structure 7.

[0011] According to another aspect of the present invention, there is provided an energy amplification generator, the energy amplification generator comprising: a first geometrical structure in which vertices of each of five or seven pieces of planar structures 1 having a regular pentagonal shape contact one another; a second geometrical structure which is separated from an upper portion of the first geometrical structure and in which vertices of each of five or seven cubic structures 2 having a regular pentagonal pyramid shape contact one another; and a plurality of separation structures 3 which separate the first geometrical structure and the second geometrical structure from each other, are installed between the planar structure 1 and the cubic structure 2 and have a smaller area than areas of the planar structure 1 and the cubic structure 2.

[0012] According to another aspect of the present invention, there is provided an energy amplification generator, the energy amplification generator comprising: a first geometrical structure in which an a first geometrical structure in which both-end vertices at two continuous sides of each of five or seven pieces of planar structures 6 having a regular heptangular shape contact one another; a second geometrical structure which is separated from an upper portion of the first geometrical structure and in which both-end vertices at two continuous sides of each of five or seven cubic structures 7 having a regular heptangular pyramid shape contact one another; and a plurality of separation structures 8 which separate the first geometrical structure and the second geometrical structure from each other, are installed between the planar structure 6 and the cubic structure 7 and have a smaller area than areas of the planar structure 6 and the cubic structure 7.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] FIG. 1 is an exploded perspective view of a space energy implosion unit according to an embodiment of the present invention;

[0014] FIG. 2 is a perspective view of a cubic structure having another shape of the space energy implosion unit of FIG. 1;

[0015] FIG. 3 is an exploded perspective view of a space energy implosion unit according to another embodiment of the present invention;

[0016] FIG. 4 is a perspective view of a cubic structure having another shape of the space energy implosion unit of FIG. 3;

[0017] FIG. 5 is an exploded perspective view of a space energy implosion unit according to another embodiment of the present invention;

[0018] FIG. 6 is a perspective view of a cubic structure having another shape of the space energy implosion unit of FIG. 5;

[0019] FIG. 7 is an exploded perspective view of a space energy implosion unit according to another embodiment of the present invention;

[0020] FIG. 8 is a perspective view of a cubic structure having another shape of the space energy implosion unit of FIG. 5;

[0021] FIG. 9 is an energy amplification generator according to an embodiment of the present invention;

[0022] FIG. 10 is a perspective view of a first geometrical structure of the energy amplification generator of FIG. 9;

[0023] FIG. 11 is a perspective view of a second geometrical structure of the energy amplification generator of FIG. 9;

[0024] FIG. 12 is a perspective view of another shape of the second geometrical structure of FIG. 10;

[0025] FIG. 13 illustrates a stack structure of a second geometrical structure, a first geometrical structure, a plate structure, and a separation structure of the energy amplification generator of FIG. 9;

[0026] FIG. 14 is an energy amplification generator according to another embodiment of the present invention;

[0027] FIG. 15 is a perspective view of a first geometrical structure of the energy amplification

generator of FIG. 14;

[0028] FIG. 16 is a perspective view of a second geometrical structure of the energy amplification generator of FIG. 14;

[0029] FIG. 17 is a perspective view of another shape of the second geometrical structure of FIG. 16;

[0030] FIG. 18 illustrates the structure in which a second geometrical structure, a first geometrical structure, a plate structure and separation structures are stacked, of the energy amplification generator of FIG. 9;

[0031] FIG. 19 is an energy amplification generator according to another embodiment of the present invention;

[0032] FIG. 20 is a perspective view of a first geometrical structure of the energy amplification generator of FIG. 19;

[0033] FIG. 21 is a perspective view of a second geometrical structure of the energy amplification generator of FIG. 19;

[0034] FIG. 22 is a perspective view of another shape of the second geometrical structure of FIG. 21;

[0035] FIG. 23 illustrates the structure in which a second geometrical structure, a first geometrical structure, a plate structure and separation structures are stacked, of the energy amplification generator of FIG. 19;

[0036] FIG. 24 is an energy amplification generator according to another embodiment of the present invention;

[0037] FIG. 25 is a perspective view of a first geometrical structure of the energy amplification generator of FIG. 24;

[0038] FIG. 26 is a perspective view of a second geometrical structure of the energy amplification generator of FIG. 24;

[0039] FIG. 27 is a perspective view of another shape of the second geometrical structure of FIG. 26;

[0040] FIG. 28 illustrates the structure in which a second geometrical structure, a first geometrical structure, a plate structure and separation structures are stacked, of the energy amplification generator of FIG. 25;

[0041] FIG. 29 is an energy amplification generator according to another embodiment of the present invention;

[0042] FIG. 30 is a perspective view of a first geometrical structure of the energy amplification generator of FIG. 29;

[0043] FIG. 31 is a perspective view of a second geometrical structure of the energy amplification generator of FIG. 29;

[0044] FIG. 32 is a perspective view of another shape of the second geometrical structure of FIG. 31;

[0045] FIG. 33 illustrates the structure in which a second geometrical structure, a first geometrical

structure, a plate structure and separation structures are stacked, of the energy amplification generator of FIG. 29;

[0046] FIG. 34 is an energy amplification generator according to another embodiment of the present invention;

[0047] FIG. 35 is a perspective view of a first geometrical structure of the energy amplification generator of FIG. 34;

[0048] FIG. 36 is a perspective view of a second geometrical structure of the energy amplification generator of FIG. 34;

[0049] FIG. 37 is a perspective view of another shape of the second geometrical structure of FIG. 36;

[0050] FIG. 38 illustrates the structure in which a second geometrical structure, a first geometrical structure, a plate structure and separation structures are stacked, of the energy amplification generator of FIG. 34;

[0051] FIG. 39 is an energy amplification generator according to another embodiment of the present invention;

[0052] FIG. 40 is a perspective view of a first geometrical structure of the energy amplification generator of FIG. 39;

[0053] FIG. 41 is a perspective view of a second geometrical structure of the energy amplification generator of FIG. 39;

[0054] FIG. 42 is a perspective view of another shape of the second geometrical structure of FIG. 41;

[0055] FIG. 43 illustrates the structure in which a second geometrical structure, a first geometrical structure, a plate structure and separation structures are stacked, of the energy amplification generator of FIG. 39;

[0056] FIG. 44 is an energy amplification generator according to another embodiment of the present invention;

[0057] FIG. 45 is a perspective view of a first geometrical structure of the energy amplification generator of FIG. 44;

[0058] FIG. 46 is a perspective view of a second geometrical structure of the energy amplification generator of FIG. 44;

[0059] FIG. 47 is a perspective view of another shape of the second geometrical structure of FIG. 46;

[0060] FIG. 48 illustrates the structure in which a second geometrical structure, a first geometrical structure, a plate structure and separation structures are stacked, of the energy amplification generator of FIG. 44;

[0061] FIG. 49 shows a graph of Table 1 and illustrates the number of subjects before taking a space energy processing water;

[0062] FIG. 50 shows a graph of Table 2 and illustrates the number of subjects that show Erythema symptoms after taking a space energy processing water;

[0063] FIG. 51 shows a graph of Table 2 and illustrates the number of subjects that show Edema Papulation symptoms after taking a space energy processing water;

[0064] FIG. 52 shows a graph of Table 2 and illustrates the number of Oozing subjects after taking a space energy processing water;

[0065] FIG. 53 shows a graph of Table 2 and illustrates the number of Excoriation subjects symptoms after taking a space energy processing water;

[0066] FIG. 54 shows a graph of Table 2 and illustrates the number of Lichenification subjects after taking a space energy processing water; and

[0067] FIG. 55 is a graph showing tumor volumes of white rats that have drunk a general drinking water and white rats that have drunk a space energy processing water.

[Fig. 1]

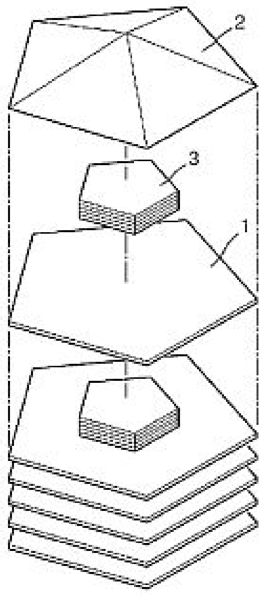

[Fig. 3]

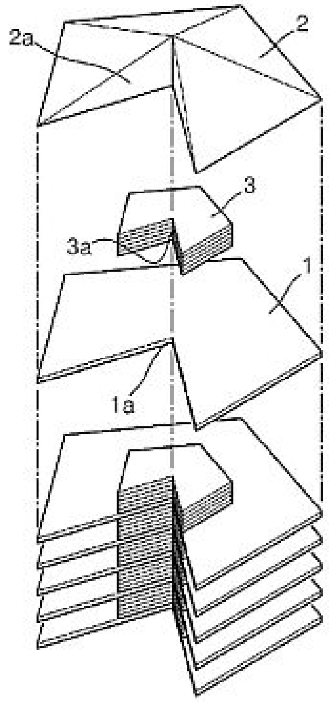

[Fig. 2]

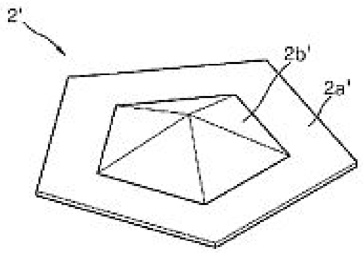

[Fig. 4]

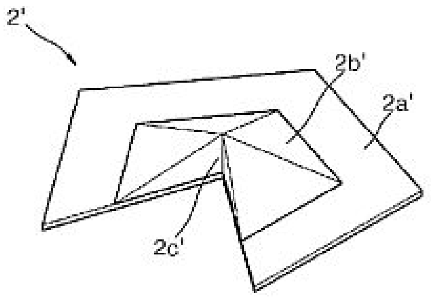

[Fig. 5]

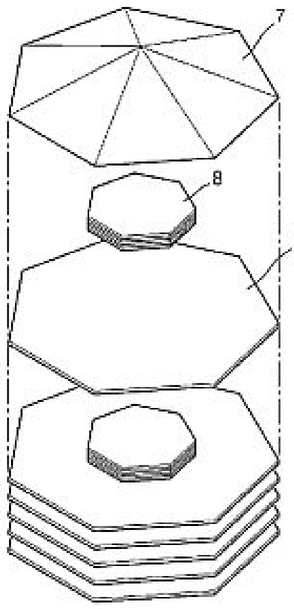

[Fig. 6]

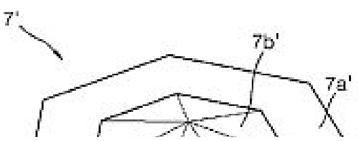

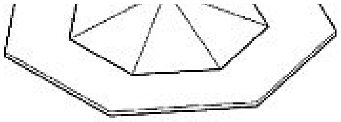

[Fig. 9]

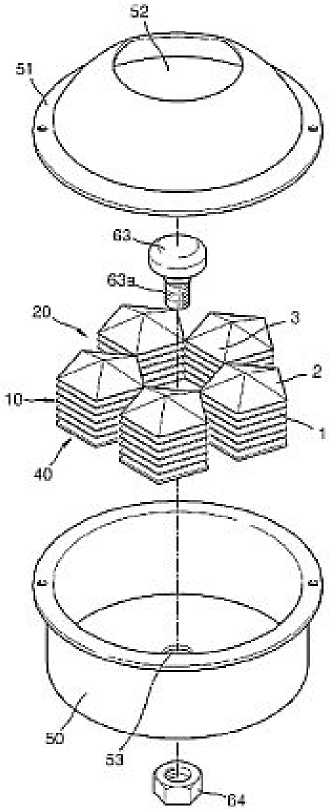
52
51
63
63a
20
3
2
10
1
40
50
53
64

[Fig. 7]

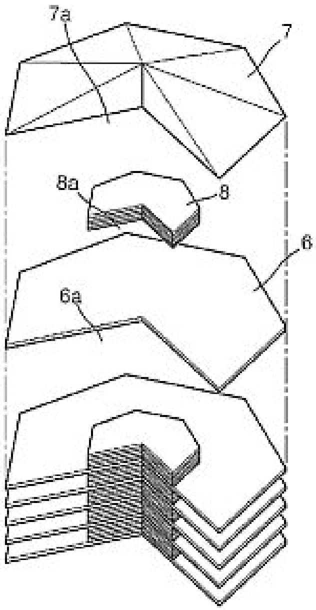

[Fig. 10]

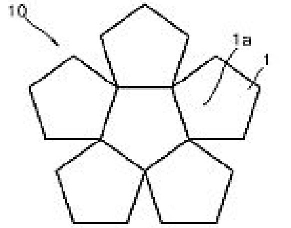

[Fig. 8]

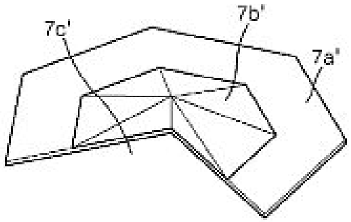

[Fig. 11]

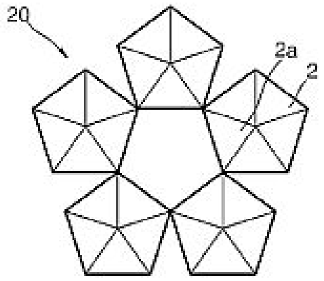

[Fig. 12]

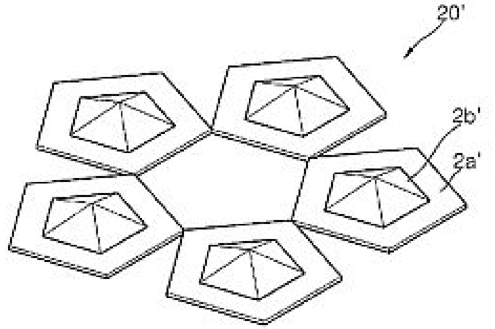

[Fig. 13]

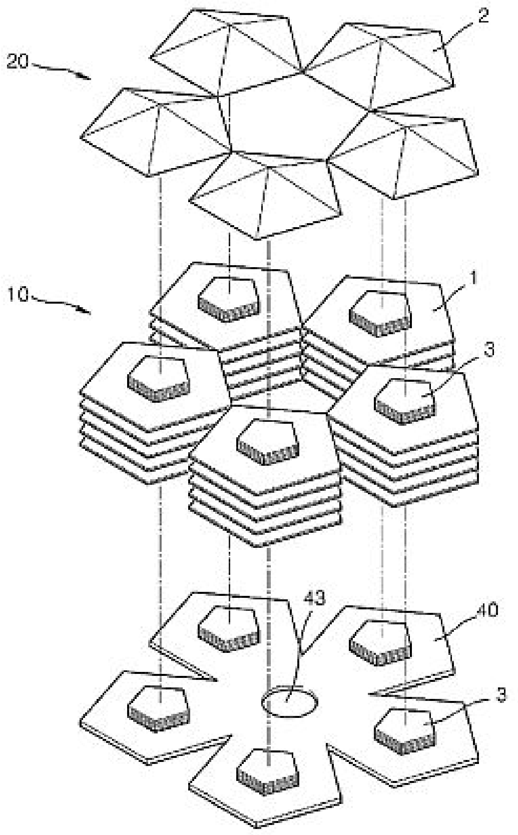
2
20
10
1
3
43
40
3

[Fig. 14]

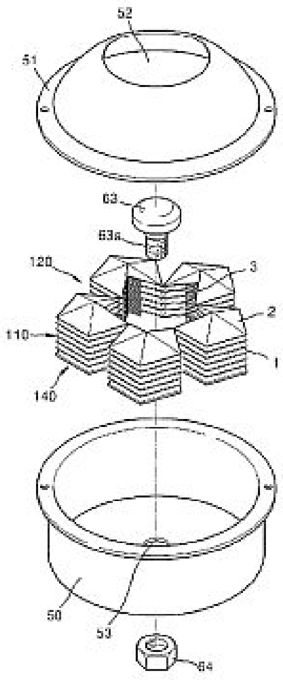
52
51
63
63a
120
3
2
110
1
140
50
53
64

[Fig. 15]

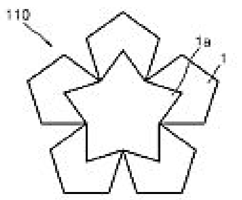

[Fig. 16]

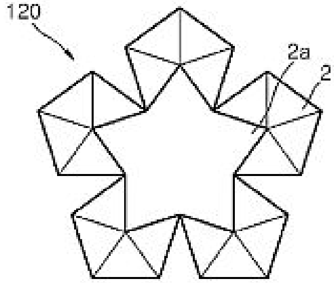

[Fig. 17]

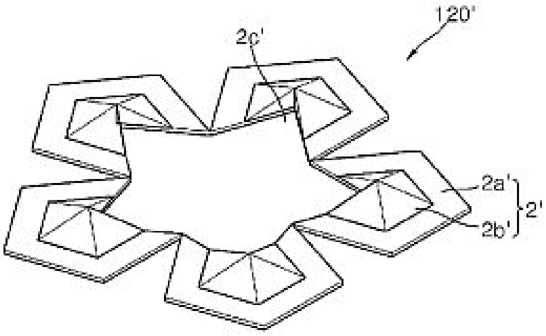
120'
2c'
2a'
2'
2b'

[Fig. 19]

[Fig. 18]

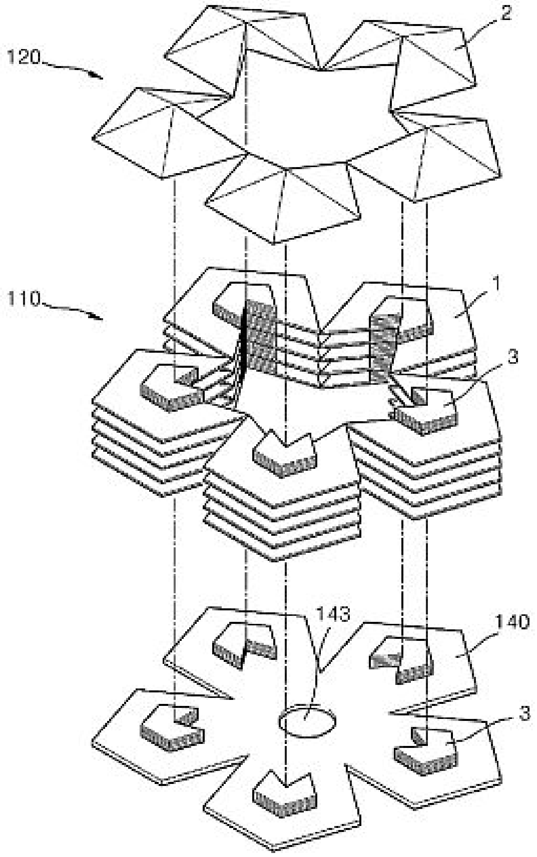
2
120
1
110
3
143
140
3

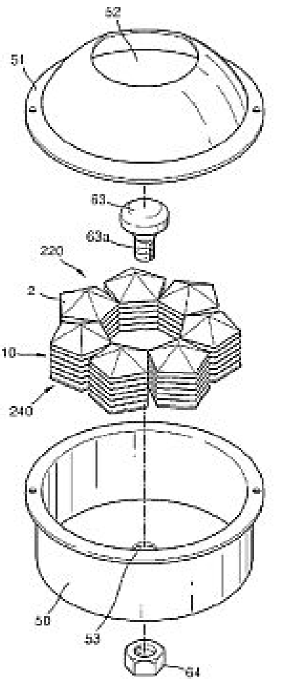
52
51
63
63a
220
2
210
240
50
53
64

[Fig. 20]

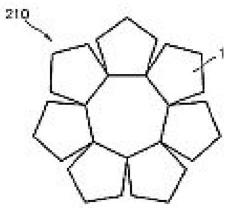

[Fig. 21]

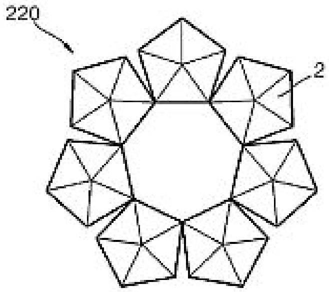

[Fig. 22]

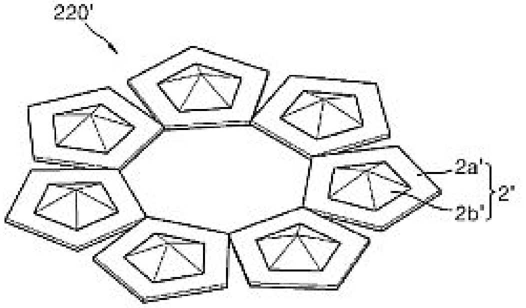

[Fig. 24]

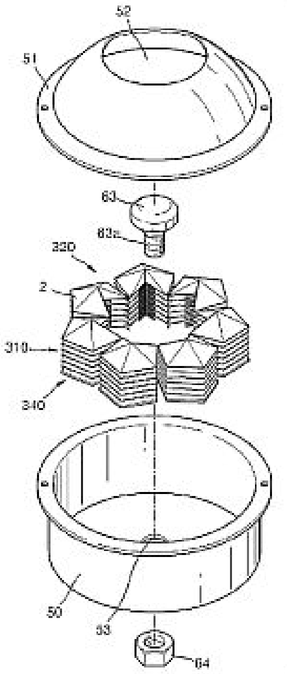
52
51
63
63a
320
2
310
340
50
33
64

[Fig. 23]

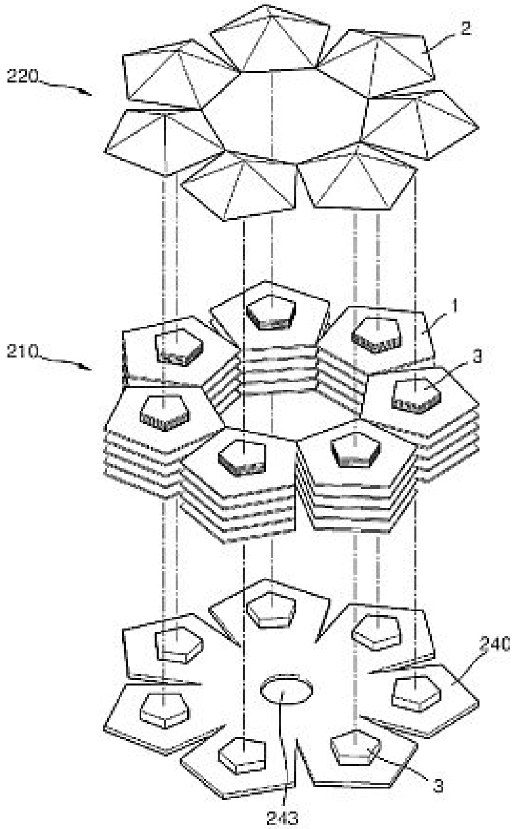
2
220
1
210
3
240
3
243

[Fig. 25]

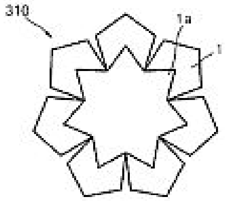

[Fig. 26]

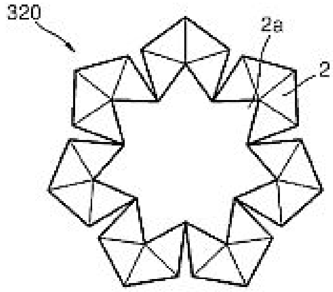

[Fig. 27]

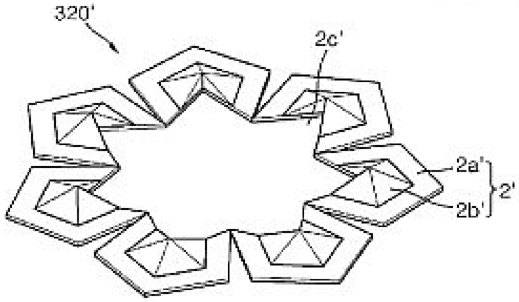
320'
2c'
2a'
2b'
2'

[Fig. 28]

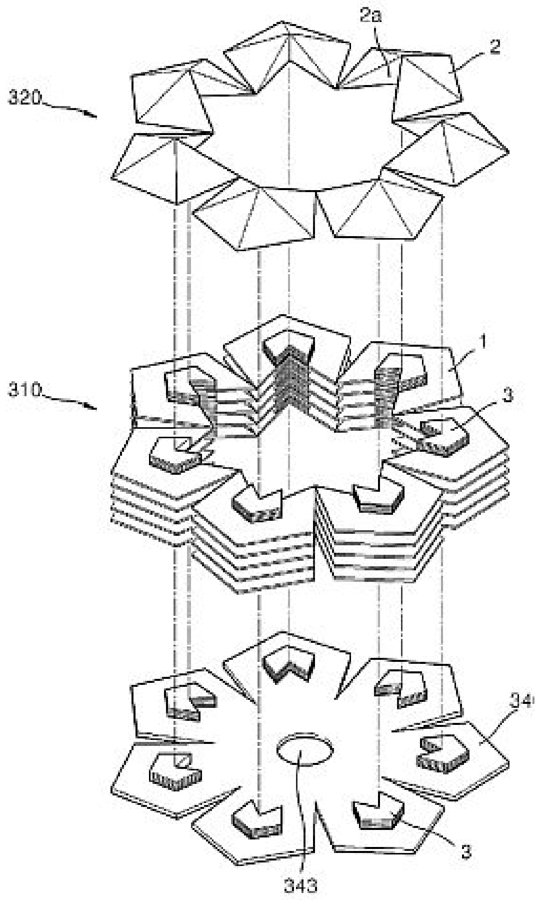
2a
2
320
1
310
3
34
3
343

[Fig. 29]

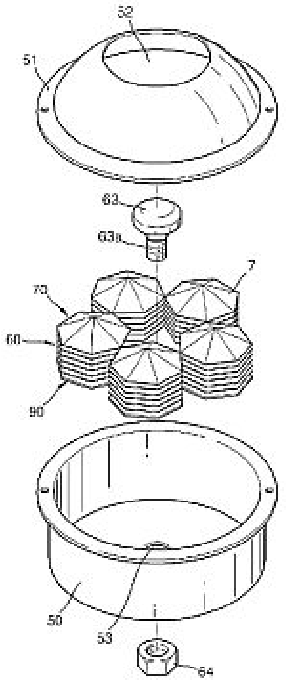
52
51
63
63a
7
70
60
90
50
53
64

[Fig. 30]

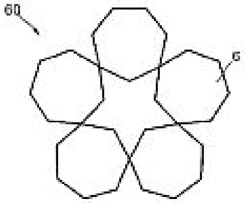

[Fig. 31]

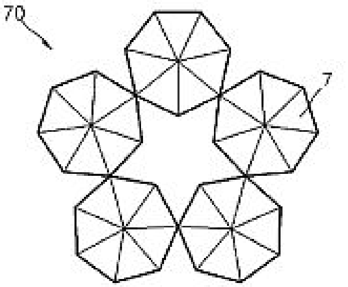

[Fig. 32]

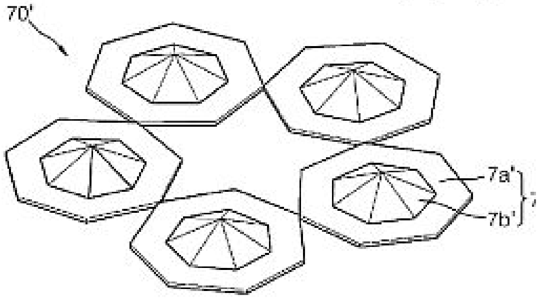

[Fig. 33]

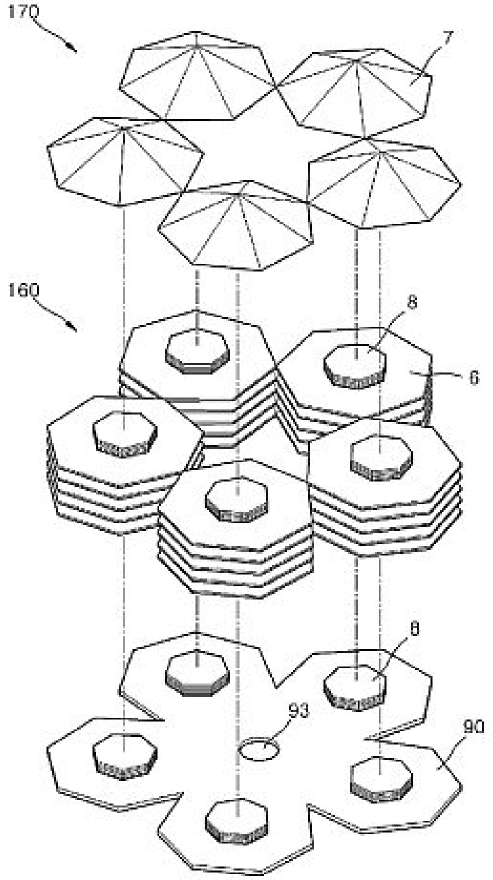
170
7
160
8
6
8
93
90

[Fig. 34]

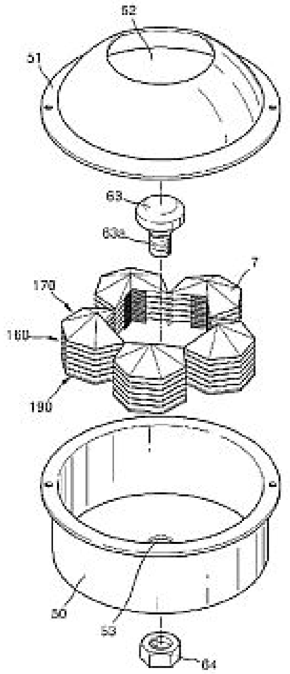
52
51
63
63a
170
7
160
190
50
53
64

[Fig. 35]

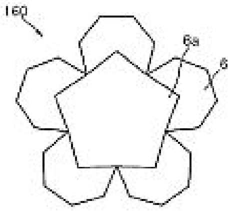

[Fig. 36]

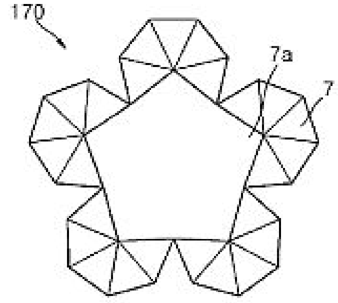

[Fig. 37]

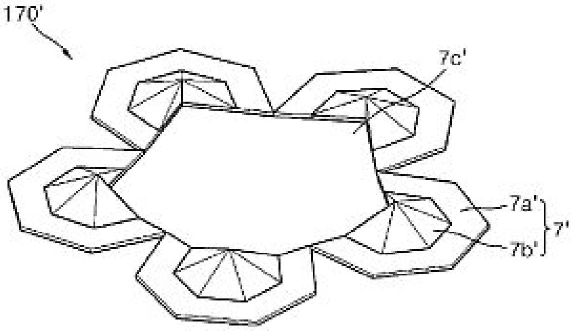
170'
7c'
7a'
7'
7b'

[Fig. 38]

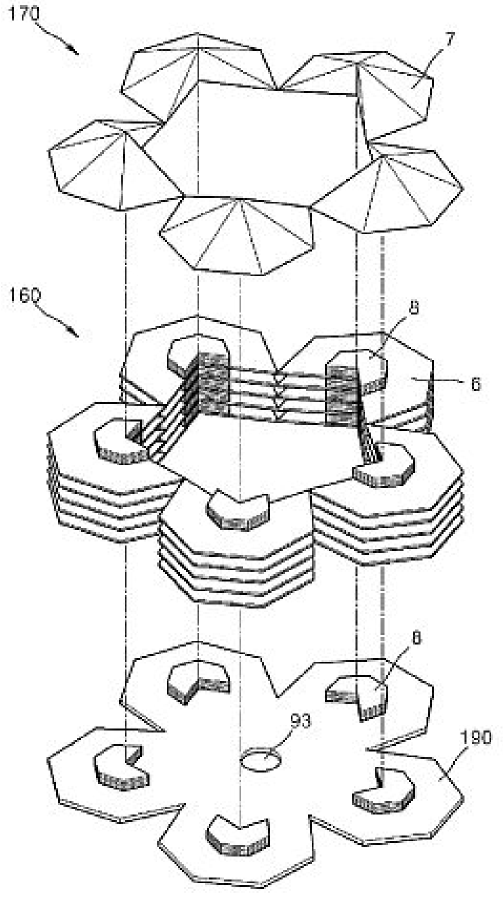
170
7
160
8
6
8
93
190

[Fig. 39]

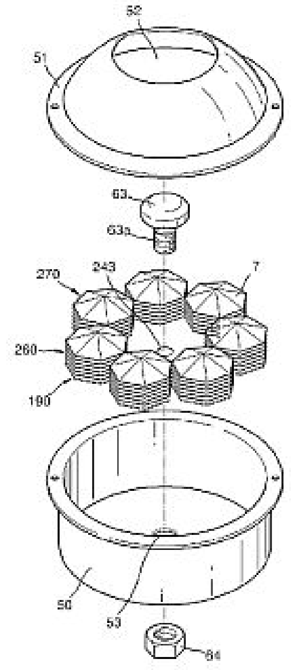
52
51
63
63a
270
243
7
260
190
50
53
64

[Fig. 40]

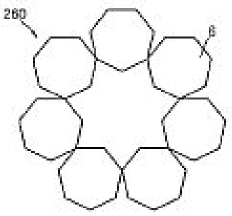

[Fig. 41]

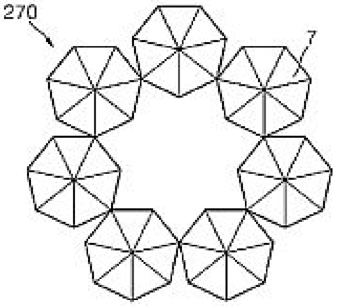

[Fig. 42]

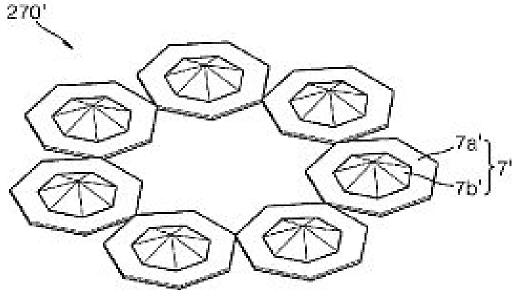

[Fig. 43]

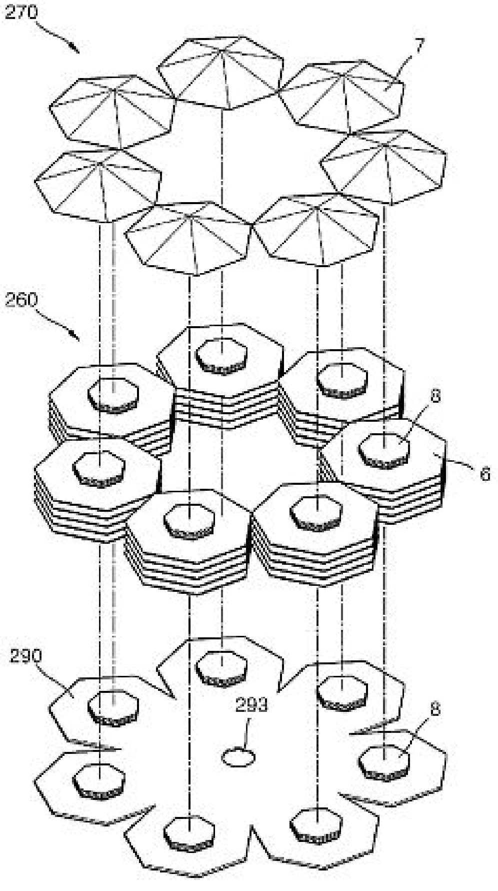
270
7
260
8
6
290
293
8

[Fig. 44]

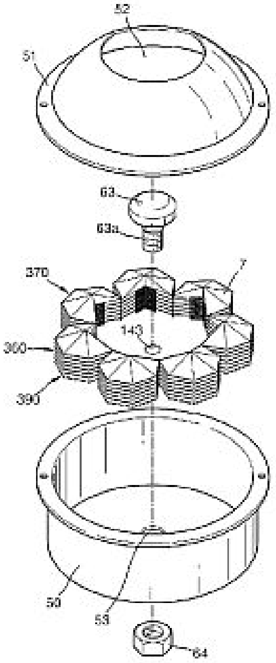
52
51
63
63a
370
7
143
360
390
50
53
64

[Fig. 45]

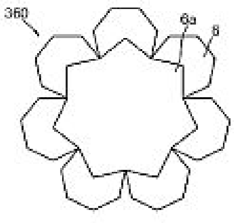

[Fig. 46]

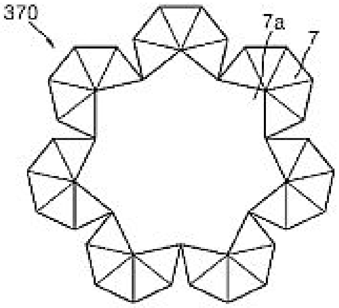

[Fig. 47]

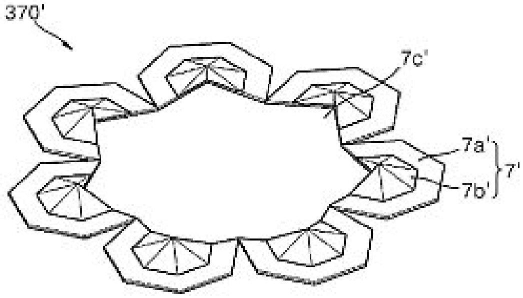
370'
7c'
7a'
7'
7b'

[Fig. 48]

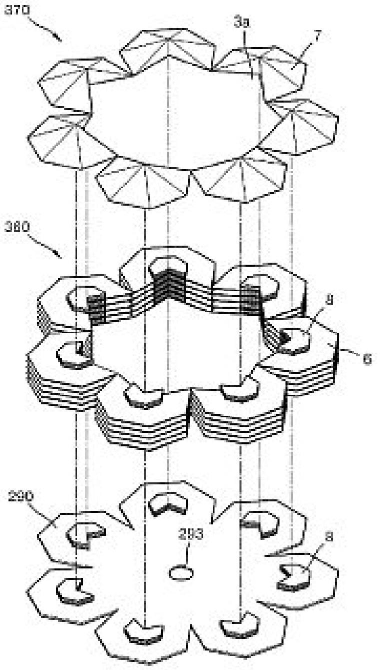

[Fig. 49]

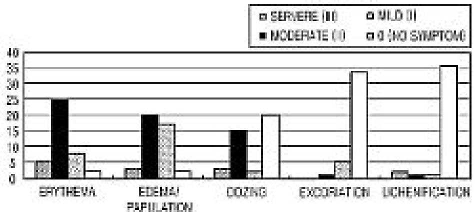

[Fig. 50]

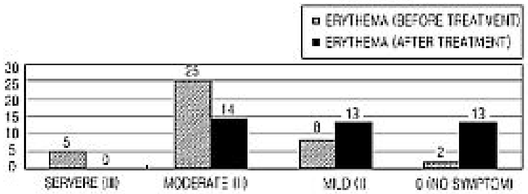

[Fig. 51]

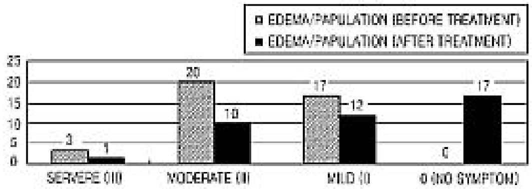

[Fig. 52]

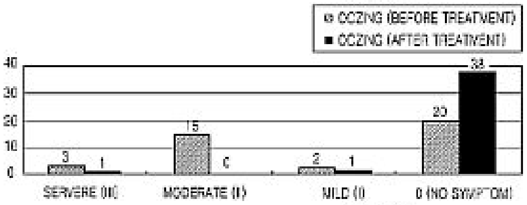

[Fig. 53]

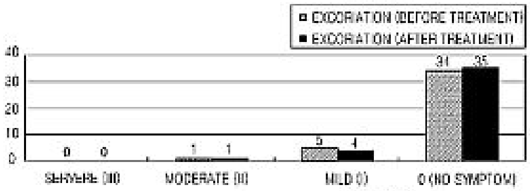

[Fig. 54]

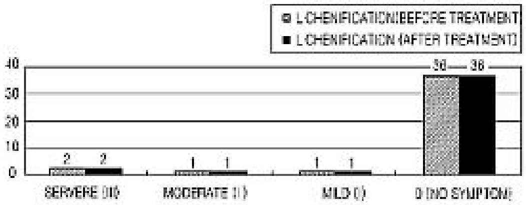

[Fig. 55]

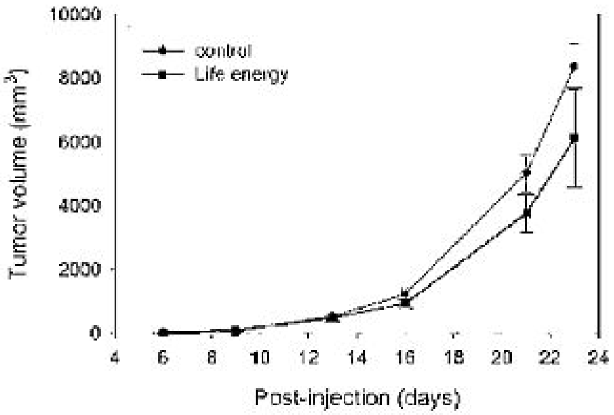

BEST MODE FOR CARRYING OUT THE INVENTION

[0068] The present invention will now be described more fully with reference to the accompanying drawings, in which exemplary embodiments of the invention are shown.

[0069] FIG. 1 is an exploded perspective view of a space energy implosion unit according to an embodiment of the present invention, and FIG. 2 is a perspective view of a cubic structure having another shape of the space energy implosion unit of FIG. 1.

[0070] Referring to FIG. 1, the space energy implosion unit according to an embodiment of the present invention includes a planar structure 1 having a regular pentagonal shape, a cubic structure 2 which is installed to be separated from an upper portion of the planar structure 1 and has a regular pentagonal pyramid shape, and a separation structure 3 which separates the planar structure 1 and the cubic structure 2 from each other and has a smaller area than areas of the planar structure 1 and the cubic structure 2. In this case, the separation structure 3 has a regular pentagonal shape. Here, the planar structure 1 is implemented as one or more pieces, and as the number of planar structures increases, the quantity of implosion of a space energy increases.

[0071] The cubic structure 2, the separation structure 3, and the planar structure 1 are stacked so that the arrangement angle of their respective sides is identical.

[0072] Here, a cubic structure 2' may also be applied in another shape. In other words, the cubic structure 2' may be implemented by forming a subcubic structure 2b' having a regular pentagonal pyramid shape having a smaller area than the area of a regular pentagonal planar plate 2a' on the planar plate 2a' as illustrated in FIG. 2.

[0073] FIG. 3 is an exploded perspective view of a space energy implosion unit according to another embodiment of the present invention, and FIG. 4 is a perspective view of a cubic structure having another shape of the space energy implosion unit of FIG. 3.

[0074] Referring to FIG. 3, the space energy implosion unit according to another embodiment of the present invention is similar to the space energy implosion unit shown in FIG. 1. The only difference therebetween is that, in FIG. 3, the space energy implosion unit further includes a planar cut groove 1a in which the planar structure 1 is cut to one side from the center and a cubic cut groove 2a in which the cubic structure 2 is cut to one side from the center. In this case, the separation structure 3 has a regular pentagonal shape but has a cut groove 3a in which the separation structure 3 is cut to one side from the center.

[0075] The cubic structure 2, the separation structure 3, and the planar structure 1 are stacked so that the arrangement angle of their respective sides is identical.

[0076] Here, a cubic structure 2' may also be applied in another shape. In other words, the cubic structure 2' may be implemented by forming a subcubic structure 2b' having a regular pentagonal pyramid shape having a smaller area than the area of a regular pentagonal planar plate 2a' on the planar plate 2a' as illustrated in FIG. 4. In this case, a cubic cut groove 2c in which the cubic structure 2 is cut to one side of the planar plate 2a from the center is formed in the cubic structure 2'.

[0077] FIG. 5 is an exploded perspective view of a space energy implosion unit according to another embodiment of the present invention, and FIG. 6 is a perspective view of a cubic structure having another shape of the space energy implosion unit of FIG. 5.

[0078] Referring to FIG. 5, the space energy implosion unit according to another embodiment of the present invention includes a planar structure 6 having a regular heptangular shape, a cubic structure 7 which is installed to be separated from an upper portion of the planar structure 6 and has a regular heptangular pyramid shape, and a separation structure 8 which separates the planar structure 6 and the cubic structure 7 from each other and has a smaller area than areas of the planar structure 6 and the cubic structure 7.

[0079] In this case, the separation structure 8 has a regular heptangular shape.

[0080] Here, the planar structure 6 is implemented as one or more pieces, and as the number of planar structures increases, the quantity of implosion of a space energy increases.

[0081] The cubic structure 7, the separation structure 8, and the planar structure 6 are stacked so that the arrangement angle of their respective sides is identical.

[0082] Here, a cubic structure 7' also be applied in another shape. In other words, the cubic structure 7' may be implemented by forming a subcubic structure 7b' having a regular heptangular pyramid shape having a smaller area than the area of a regular heptangular planar plate 7a' on the planar plate 7a' as illustrated in FIG. 6.

[0083] FIG. 7 is an exploded perspective view of a space energy implosion unit according to another embodiment of the present invention, and FIG. 8 is a perspective view of a cubic structure having another shape of the space energy implosion unit of FIG. 5.

[0084] Referring to FIG. 7, the space energy implosion unit according to another embodiment of the present invention is similar to the space energy implosion unit shown in FIG. 5. The only difference therebetween is that, in FIG. 7, the space energy implosion unit further includes a planar cut groove 6a in which the planar structure 6 is cut to two continuous sides from the center and a cubic cut groove 7a in which the cubic structure 7 is cut to two continuous sides from the center. In this case, the separation structure 8 has a regular pentagonal shape but has a cut groove 8a in which the separation structure 8 is cut to two continuous sides from the center.

[0085] The cubic structure 7, the separation structure 8, and the planar structure 6 are stacked so that the arrangement angle of their respective sides is identical.

[0086] Here, a cubic structure 7' may also be applied in another shape. In other words, the cubic structure 7' may be implemented by forming a subcubic structure 7b' having a regular heptangular pyramid shape having a smaller area than the area of a regular heptangular planar plate 7a' on the planar plate 7a' as illustrated in FIG. 8. In this case, a cubic cut groove 7c' in which the cubic structure 7' is cut to two continuous sides of the planar plate 7a' from the center is formed in the cubic structure 7'.

[0087] In the space energy implosion unit shown in FIGS. 1, 3, 5, and 7, the planar structure 1 implodes a space energy that is full in the universe due to its geometrical structure, and the cubic structure 2 concentrates the space energy imploded from the planar structure 1 due to its geometrical structure in a forward direction. When the cubic structure 2 is implemented in the shapes shown in FIGS. 2, 4, 6, and 8, the space energy that is full in the universe is imploded and simultaneously is concentrated in a forward direction.

[0088] An energy amplification generator using the space energy implosion unit according to exemplary embodiments of the present invention will now be described.

[0089] FIG. 9 is an energy amplification generator according to an embodiment of the present invention, FIG. 10 is a perspective view of a first geometrical structure of the energy amplification generator of FIG. 9, FIG. 11 is a perspective view of a second geometrical structure of the energy amplification generator of FIG. 9, FIG. 12 is a perspective view of another shape of the second geometrical structure of FIG. 10, and FIG. 13 illustrates a stack structure of a second geometrical structure, a first geometrical structure, a plate structure, and a separation structure of the energy amplification generator of FIG. 9. Here, like reference numerals in FIGS. 1 through 4 denote like elements having the same function.

[0090] Referring to FIG. 9, the energy amplification generator according to an embodiment of the present invention includes a first geometrical structure 10 in which vertices of each of five pieces of planar structures 1 having a regular pentagonal shape contact one another, a second geometrical structure 20 which is separated from an upper portion of the first geometrical structure 10 and in which vertices of each of five cubic structures 2 having a regular pentagonal pyramid shape contact one another, and a plurality of separation structures 3 which separate the first geometrical structure 10 and the second geometrical structure 20 from each other, are installed between the planar structure 1 and the cubic structure 2 and have a smaller area than areas of the planar structure 1 and the cubic structure 2.

[0091] The first geometrical structure 10 is one or two or more. The first geometrical structure 10 implodes a space energy that is full in the universe due to its geometrical structure to be indirectly and directly used.

[0092] The second geometrical structure 20 concentrates the space energy imploded by the first geometrical structure 10 in a forward direction.

[0093] Areas of the first geometrical structure 10 and the second geometrical structure 20 allow the quantity of the imploded space energy to be increased. In other words, in order to obtain a larger quantity of space energy, the number of stacked first geometrical structures 10 increases or the areas of the first and second geometrical structures 10 and 20 increase.

[0094] On the other hand, a second geometrical structure 20' may be modified in another shape. In other words, as illustrated in FIG. 12, in the second geometrical structure 20' a cubic structure 2 is implemented by forming a subcubic structure 2b' having a regular pentagonal pyramid shape having a smaller area than the area of a regular pentagonal planar plate 2a' on the planar plate 2a' Five cubic structures 2' are disposed on a plane so that vertices of each of five cubic structures 2' contact one another, thereby implementing the second geometrical structure 20' of the present application. The second geometrical structure 20' having the cubic structure 2' implodes the space energy and simultaneously concentrates the space energy in a forward direction.

[0095] A plate structure 40 supports the first geometrical structure 10 and has the same shape as that of the first geometrical structure 10. A plate hole 43 corresponding to a case hole 53 that will be described later is formed in the center of the plate structure 40.

[0096] The separation structures 3 may be installed among several first geometrical structures 10 so that the first geometrical structures 10 are separated from one another, or may be installed between the first geometrical structure 10 and the second geometrical structure 20 so that they are separated from each other, or may be installed between the first geometrical structure 10 and the plate structure 40 so that they are separated from each other. The separation structures 3 have a regular pentagonal shape and may be implemented to a predetermined thickness by stacking several plates or as a one product having a predetermined thickness.

[0097] The separation structures 3 allow the frequency of a radiated space energy to vary. In other words, the frequency of the space energy varies according to the thickness of the separation structures 3. The thickness of the separation structures 3 may be between 3 mm and 20 mm, and 7 mm thick separation structures 3 are used in the present embodiment.

[0098] A case 50 has a cover 51 in which a radiation sphere 52 through which a space energy is radiated in a forward direction is formed. The case hole 53 corresponding to the plate hole 43 is formed in the case 50.

[0099] The first geometrical structure 10, the separation structures 3, and the second geometrical structure 20 that will be stacked on the plate structure 40 are built in the case 50. In other words, the plate structure 40 on which the first and second geometrical structures 10 and 20 and the separation structures 3 are stacked, is fixed in the case 50 in such a way that a bolt portion 63a of a fixing member 63 perforates the plate hole 43 and the case hole 53 and then is engaged with a nut 64 in the rear of the case 50, as illustrated in FIG. 9.

[0100] An energy amplification generator according to another embodiment of the present invention will now be described.

[0101] FIG. 14 is an energy amplification generator according to another embodiment of the present invention, FIG. 15 is a perspective view of a first geometrical structure of the energy amplification generator of FIG. 14, FIG. 16 is a perspective view of a second geometrical structure of the energy amplification generator of FIG. 14, FIG. 17 is a perspective view of another shape of the second geometrical structure of FIG. 16, and FIG. 18 illustrates the structure in which a second geometrical structure, a first geometrical structure, a plate structure and separation structures are stacked, of the energy amplification generator of FIG. 9. Here, like reference numerals in FIG. 9 denote like elements having the same function.

[0102] Referring to FIG. 14, the energy amplification generator according to another embodiment of the present invention includes a first geometrical structure 110 in which vertices of each of five pieces of planar structures 1 having a regular pentagonal shape contact one another, a second geometrical structure 120 which is separated from an upper portion of the first geometrical structure 110 and in which vertices of each of five cubic structures 2 having a regular pentagonal pyramid shape contact one another, and a plurality of separation structures 3 which separate the first geometrical structure 110 and the second geometrical structure 120 from each other, are installed between the planar structure 1 and the cubic structure 2 and have a smaller area than areas of the planar structure 1 and the cubic structure 2.

[0103] The first geometrical structure 110 is one or two or more. The first geometrical structure 110 implodes a space energy that is full in the universe due to its geometrical structure to be indirectly and directly used.

[0104] The second geometrical structure 120 concentrates the space energy imploded by the first geometrical structure 110 in a forward direction.

[0105] Areas of the first geometrical structure 110 and the second geometrical structure 120 allow the quantity of the imploded space energy to be increased. In other words, in order to obtain a larger quantity of space energy, the number of stacked first geometrical structures 110 increases or the areas of the first and second geometrical structures 110 and 120 increase.

[0106] In this case, in the first geometrical structure 110, the planar structure 1 further includes a planar cut groove 1a in which the planar structure 1 is cut to one side between two contacting vertices from the center, and the cubic structure 2 further includes a cubic cut groove 2a in which the cubic structure 2 is cut to one side between two contacting vertices from the center.

[0107] In addition, the separation structures 3 have a regular pentagonal shape but also have a cut groove 3a in which each of the separation structures 3 is cut to one side from the center.

[0108] The cubic structure 2, the separation structures 3, and the planar structure 1 are stacked that the arrangement angle of their respective sides are identical.

[0109] On the other hand, a second geometrical structure 120' may be modified in another shape. In other words, as illustrated in FIG. 17, a cubic structure 2' of the second geometrical structure 120' is implemented by forming a subcubic structure 2b' having a regular pentagonal pyramid shape having a smaller area than the area of a regular pentagonal planar plate 2a on the planar plate 2a'. At this time, a cubic cut groove 2c' in which the cubic structure 2' is cut to one side between two contacting vertices of the planar plate 2a' from the center is formed in the cubic structure 2'. Five cubic structures 2' are disposed on a plane so that vertices of each of five cubic structures 2' contact one another, thereby implementing the second geometrical structure 120' of the present application.

[0110] A plate structure 140 supports the first geometrical structure 110 and has the same shape as that of the first geometrical structure 110. A plate hole 143 corresponding to a case hole 53 is formed in the center of the plate structure 140.

[0111] The case 50, the cover 51, the radiation sphere 52, the case hole 53, the fixing member 63, the bolt portion 63a, and the nut 64 are the same as in the energy amplification generator shown in FIG. 9 and thus, a detailed description thereof will be omitted.

[0112] An energy amplification generator according to another embodiment of the present invention will now be described.

[0113] FIG. 19 is an energy amplification generator according to another embodiment of the present invention, FIG. 20 is a perspective view of a first geometrical structure of the energy amplification generator of FIG. 19, FIG. 21 is a perspective view of a second geometrical structure of the energy amplification generator of FIG. 19, FIG. 22 is a perspective view of another shape of the second geometrical structure of FIG. 21, and FIG. 23 illustrates the structure in which a second geometrical structure, a first geometrical structure, a plate structure and separation structures are stacked, of the energy amplification generator of FIG. 19. Here, like reference numerals in FIGS. 9 and 14 denote like elements having the same function.

[0114] Referring to FIG. 19, the energy amplification generator according to another embodiment of the present invention includes a first geometrical structure 210 in which vertices of each of seven pieces of planar structures 1 having a regular pentagonal shape contact one another, a second geometrical structure 220 which is separated from an upper portion of the first geometrical structure 210 and in which vertices of each of seven cubic structures 2 having a regular pentagonal pyramid shape contact one another, and a plurality of separation structures 3 which separate the first geometrical structure 210 and the second geometrical structure 220 from each other, are installed between the planar structure 1 and the cubic structure 2 and have a smaller area than areas of the planar structure 1 and the cubic structure 2.

[0115] The first geometrical structure 210 is one or two or more. The first geometrical structure 210 implodes a space energy that is full in the universe due to its geometrical structure to be indirectly and directly used.

[0116] The second geometrical structure 220 concentrates the space energy imploded by the first geometrical structure 210 in a forward direction.

[0117] Areas of the first geometrical structure 210 and the second geometrical structure 220 allow the quantity of the imploded space energy to be increased. In other words, in order to obtain a larger quantity of space energy, the number of stacked first geometrical structures 210 increases or the areas of the first and second geometrical structures 210 and 220 increase.

[0118] The cubic structure 2, the separation structures 3, and the planar structure 1 are stacked that the arrangement angle of their respective sides are identical.

[0119] On the other hand, a second geometrical structure 220' may be modified in another shape. In other words, as illustrated in FIG. 22, a cubic structure 2' of the second geometrical structure 220' is implemented by forming a subcubic structure 2b' having a regular pentagonal pyramid shape having a smaller area than the area of a regular pentagonal planar plate 2a' on the planar plate 2a'. Seven cubic structures 2' are disposed on a plane so that vertices of each of seven cubic structures 2' contact one another, thereby implementing the second geometrical structure 220' of the present application.

[0120] A plate structure 240 supports the first geometrical structure 210 and has the same shape as that of the first geometrical structure 210. A plate hole 243 corresponding to a case hole 53 is formed in the center of the plate structure 240.

[0121] The case 50, the cover 51, the radiation sphere 52, the case hole 53, the fixing member 63, the bolt portion 63a, and the nut 64 are the same as in the energy amplification generator shown in FIGS. 9 and 14 and thus, a detailed description thereof will be omitted.

[0122] An energy amplification generator according to another embodiment of the present invention will now be described.

[0123] FIG. 24 is an energy amplification generator according to another embodiment of the present invention, FIG. 25 is a perspective view of a first geometrical structure of the energy amplification generator of FIG. 24, FIG. 26 is a perspective view of a second geometrical structure of the energy amplification generator of FIG. 24, FIG. 27 is a perspective view of another shape of the second geometrical structure of FIG. 26, and FIG. 28 illustrates the structure in which a second geometrical structure, a first geometrical structure, a plate structure and separation structures are stacked, of the energy amplification generator of FIG. 25. Here, like reference numerals in FIGS. 9, 14, and 19 denote like elements having the same function.

[0124] Referring to FIG. 24, the energy amplification generator according to another embodiment of the present invention includes a first geometrical structure 310 in which vertices of each of seven pieces of planar structures 1 having a regular pentagonal shape contact one another, a second geometrical structure 320 which is separated from an upper portion of the first geometrical structure 310 and in which vertices of each of seven cubic structures 2 having a regular pentagonal pyramid shape contact one another, and a plurality of separation structures 3 which separate the first geometrical structure 310 and the second geometrical structure 320 from each other, are installed between the planar structure 1 and the cubic structure 2 and have a smaller area than areas of the planar structure 1 and the cubic structure 2.

[0125] The first geometrical structure 310 is one or two or more. The first geometrical structure 310 implodes a space energy that is full in the universe due to its geometrical structure to be indirectly and directly used.

[0126] The second geometrical structure 320 concentrates the space energy imploded by the first geometrical structure 310 in a forward direction.

[0127] Areas of the first geometrical structure 310 and the second geometrical structure 320 allow the quantity of the imploded space energy to be increased. In other words, in order to obtain a larger quantity of space energy, the number of stacked first geometrical structures 310 increases or the areas of the first and second geometrical structures 310 and 320 increase.

[0128] In this case, in the first geometrical structure 310, the planar structure 1 further includes a planar cut groove 1a in which the planar structure 1 is cut to one side between two contacting vertices from the center, and the cubic structure 2 further includes a cubic cut groove 2a in which the cubic structure 2 is cut to one side between two contacting vertices from the center.

[0129] In addition, the separation structures 3 have a regular pentagonal shape but also have a cut groove 3a in which each of the separation structures 3 is cut to one side from the center.

[0130] The cubic structure 2, the separation structures 3, and the planar structure 1 are stacked that the arrangement angle of their respective sides are identical.

[0131] On the other hand, a second geometrical structure 320' may be modified in another shape. In other words, as illustrated in FIG. 27, a cubic structure 2' of the second geometrical structure 320' is implemented by forming a subcubic structure 2b' having a regular pentagonal pyramid shape having a smaller area than the area of a regular pentagonal planar plate 2a' on the planar plate 2a'. At this time, a cubic cut groove 2c' in which the cubic structure 2' is cut to one side between two contacting vertices of the planar plate 2a' from the center is formed in the cubic structure 2'. Seven cubic structures 2' are disposed on a plane so that vertices of each of seven cubic structures 2' contact one another, thereby implementing the second geometrical structure 320' of the present application.

[0132] A plate structure 340 supports the first geometrical structure 310 and has the same shape as that of the first geometrical structure 310. A plate hole 343 corresponding to a case hole 53 is formed in the center of the plate structure 340.

[0133] The case 50, the cover 51, the radiation sphere 52, the case hole 53, the fixing member 63, the bolt portion 63a, and the nut 64 are the same as in the energy amplification generator shown in FIGS. 9, 14, and 19, and thus, a detailed description thereof will be omitted.

[0134] An energy amplification generator according to another embodiment of the present invention will now be described.

[0135] FIG. 29 is an energy amplification generator according to another embodiment of the present invention, FIG. 30 is a perspective view of a first geometrical structure of the energy amplification generator of FIG. 29, FIG. 31 is a perspective view of a second geometrical structure of the energy amplification generator of FIG. 29, FIG. 32 is a perspective view of another shape of the second geometrical structure of FIG. 31, and FIG. 33 illustrates the structure in which a second geometrical structure, a first geometrical structure, a plate structure and separation structures are stacked, of the energy amplification generator of FIG. 29. Here, like reference numerals in FIGS. 5 through 8 denote like elements having the same function.

[0136] Referring to FIG. 29, the energy amplification generator according to another embodiment of the present invention includes a first geometrical structure 60 in which both-end vertices at two continuous sides of each of five pieces of planar structures 6 having a regular heptangular shape contact one another, a second geometrical structure 70 which is separated from an upper portion of the first geometrical structure 60 and in which both-end vertices at two continuous sides of each of five cubic structures 7 having a regular heptangular pyramid shape contact one another, and a plurality of separation structures 8 which separate the first geometrical structure 60 and the second geometrical structure 70 from each other, are installed between the planar structure 6 and the cubic structure 7 and have a smaller area than areas of the planar structure 6 and the cubic structure 7.

[0137] The first geometrical structure 60 is one or two or more. The first geometrical structure 60 implodes a space energy that is full in the universe due to its geometrical structure to be indirectly and directly used.

[0138] The second geometrical structure 70 concentrates the space energy imploded by the first geometrical structure 60 in a forward direction.

[0139] Areas of the first geometrical structure 60 and the second geometrical structure 70 allow the quantity of the imploded space energy to be increased. In other words, in order to obtain a larger quantity of space energy, the number of stacked first geometrical structures 60 increases or the areas of the first and second geometrical structures 60 and 70 increase.

[0140] On the other hand, a second geometrical structure 70' may be modified in another shape. In other words, as illustrated in FIG. 32, a cubic structure 7' of the second geometrical structure 70' is implemented by forming a subcubic structure 7b' having a regular heptangular pyramid shape having a smaller area than the area of a regular heptangular planar plate 7a' on the planar plate 7a'. Five cubic structures 7' are disposed on a plane so that both-end vertices at two continuous sides of each of five cubic structures 7' contact one another, thereby implementing the second geometrical structure 70' of the present application. The second geometrical structure 70' having the cubic structure 7' implodes a space energy and simultaneously, concentrates the space energy in a forward direction.

[0141] A plate structure 90 supports the first geometrical structure 60 and has the same shape as that of the first geometrical structure 60. A plate hole 93 corresponding to a case hole 53 that will be described later is formed in the center of the plate structure 90.

[0142] The separation structures 8 may be installed among several first geometrical structures 60 so that the first geometrical structures 60 are separated from one another, or may be installed between the first geometrical structure 60 and the second geometrical structure 70 so that they are separated from each other, or may be installed between the first geometrical structure 60 and the plate structure 90 so that they are separated from each other. The separation structures 8 have a regular heptangular shape and may be implemented to a predetermined thickness by stacking several plates or as a one product having a predetermined thickness.

[0143] The separation structures 8 allow the frequency of a radiated space energy to vary. In other words, the frequency of the space energy varies according to the thickness of the separation structures 8. The thickness of the separation structures 8 may be between 3 mm and 20 mm, and 7 mm thick separation structures 3 are used in the present embodiment.

[0144] A case 50 has a cover 51 in which a radiation sphere 52 through which a space energy is radiated in a forward direction is formed. The case hole 53 corresponding to the plate hole 93 is formed in the case

50.

[0145] The first geometrical structure 60, the separation structures 8, and the second geometrical structure 70 that will be stacked on the plate structure 90 are built in the case 50. In other words, the plate structure 90 on which the first and second geometrical structures 60 and 70 and the separation structures 8 are stacked, is fixed in the case 50 in such a way that a bolt portion 63a of a fixing member 63 perforates the plate hole 93 and the case hole 53 and then is engaged with a nut 64 in the rear of the case 50, as illustrated in FIG. 29.

[0146] An energy amplification generator according to another embodiment of the present invention will now be described.

[0147] FIG. 34 is an energy amplification generator according to another embodiment of the present invention, FIG. 35 is a perspective view of a first geometrical structure of the energy amplification generator of FIG. 34, FIG. 36 is a perspective view of a second geometrical structure of the energy amplification generator of FIG. 34, FIG. 37 is a perspective view of another shape of the second geometrical structure of FIG. 36, and FIG. 38 illustrates the structure in which a second geometrical structure, a first geometrical structure, a plate structure and separation structures are stacked, of the energy amplification generator of FIG. 34. Here, like reference numerals in FIG. 29 denote like elements having the same function.

[0148] Referring to FIG. 34, the energy amplification generator according to another embodiment of the present invention includes a first geometrical structure 160 in which both-end vertices at two continuous sides of each of five pieces of planar structures 6 having a regular heptangular shape contact one another, a second geometrical structure 170 which is separated from an upper portion of the first geometrical structure 160 and in which both-end vertices at two continuous sides of each of five cubic structures 7 having a regular heptangular pyramid shape contact one another, and a plurality of separation structures 8 which separate the first geometrical structure 160 and the second geometrical structure 170 from each other, are installed between the planar structure 6 and the cubic structure 7 and have a smaller area than areas of the planar structure 6 and the cubic structure 7.

[0149] The first geometrical structure 160 is one or two or more. The first geometrical structure 160 implodes a space energy that is full in the universe due to its geometrical structure to be indirectly and directly used.

[0150] The second geometrical structure 170 concentrates the space energy imploded by the first geometrical structure 160 in a forward direction.

[0151] Areas of the first geometrical structure 160 and the second geometrical structure 170 allow the quantity of the imploded space energy to be increased. In other words, in order to obtain a larger quantity of space energy, the number of stacked first geometrical structures 160 increases or the areas of the first and second geometrical structures 160 and 170 increase.

[0152] In this case, in the first geometrical structure 160, the planar structure 6 further includes a planar cut groove 6a in which the planar structure 6 is cut to two continuous sides between two contacting vertices from the center, and the cubic structure 7 further includes a cubic cut groove 7a in which the cubic structure 7 is cut to two continuous sides between two contacting vertices from the center. In addition, the separation structures 8 have a regular heptangular shape but also have a cut groove 8a in which each of the separation structures 8 is cut to two continuous sides from the center.

[0153] The cubic structure 7, the separation structures 8, and the planar structure 6 are stacked that the arrangement angle of their respective sides are identical.

[0154] On the other hand, a second geometrical structure 170' may be modified in another shape. In other words, as illustrated in FIG. 37, a cubic structure 7' of the second geometrical structure 170' is implemented by forming a subcubic structure 7b' having a regular heptangular pyramid shape having a smaller area than the area of a regular heptangular planar plate 7a' on the planar plate 7a'. At this time, a cubic cut groove 7c' in which the cubic structure 7' is cut to two continuous sides between two contacting vertices of the planar plate 7a' from the center is formed in the cubic structure 7'. Five cubic structures 7' are disposed on a plane so that both-end vertices at two continuous sides of each of five cubic structures 7' contact one another, thereby implementing the second geometrical structure 170' of the present application.

[0155] A plate structure 190 supports the first geometrical structure 160 and has the same shape as that of the first geometrical structure 160. A plate hole 193 corresponding to a case hole 53 that will be described later is formed in the center of the plate structure 190.

[0156] A case 50, a cover 51, a radiation sphere 52, a case hole 53, a fixing member 63, a bolt portion 63a, and a nut 64 are the same as in the energy amplification generator shown in FIG. 29 and thus, a detailed description thereof will be omitted.

[0157] An energy amplification generator according to another embodiment of the present invention will now be described.

[0158] FIG. 39 is an energy amplification generator according to another embodiment of the present invention, FIG. 40 is a perspective view of a first geometrical structure of the energy amplification generator of FIG. 39, FIG. 41 is a perspective view of a second geometrical structure of the energy amplification generator of FIG. 39, FIG. 42 is a perspective view of another shape of the second geometrical structure of FIG. 41, and FIG. 43 illustrates the structure in which a second geometrical structure, a first geometrical structure, a plate structure and separation structures are stacked, of the energy amplification generator of FIG. 39. Here, like reference numerals in FIGS. 29 and 34 denote like elements having the same function.

[0159] Referring to FIG. 39, the energy amplification generator according to another embodiment of the present invention includes a first geometrical structure 260 in which both-end vertices at two continuous sides of each of seven pieces of planar structures 6 having a regular heptangular shape contact one another, a second geometrical structure 270 which is separated from an upper portion of the first geometrical structure 260 and in which both-end vertices at two continuous sides of each of seven cubic structures 7 having a regular heptangular pyramid shape contact one another, and a plurality of separation structures 8 which separate the first geometrical structure 260 and the second geometrical structure 270 from each other, are installed between the planar structure 6 and the cubic structure 7 and have a smaller area than areas of the planar structure 6 and the cubic structure 7.

[0160] The first geometrical structure 260 is one or two or more. The first geometrical structure 260 implodes a space energy that is full in the universe due to its geometrical structure to be indirectly and directly used.

[0161] The second geometrical structure 270 concentrates the space energy imploded by the first geometrical structure 260 in a forward direction.

[0162] Areas of the first geometrical structure 260 and the second geometrical structure 270 allow the quantity of the imploded space energy to be increased. In other words, in order to obtain a larger quantity of space energy, the number of stacked first geometrical structures 260 increases or the areas of the first and second geometrical structures 260 and 270 increase.

[0163] The cubic structure 7, the separation structures 8, and the planar structure 6 are stacked that the arrangement angle of their respective sides are identical.

[0164] On the other hand, a second geometrical structure 270' may be modified in another shape. In other words, as illustrated in FIG. 42, a cubic structure 7' of the second geometrical structure 270' is implemented by forming a subcubic structure 7b' having a regular heptangular pyramid shape having a smaller area than the area of a regular heptangular planar plate 7a' on the planar plate 7a'. Seven cubic structures 7' are disposed on a plane so that both-end vertices at two continuous sides of each of seven cubic structures 7' contact one another, thereby implementing the second geometrical structure 270' of the present application.

[0165] A plate structure 290 supports the first geometrical structure 260 and has the same shape as that of the first geometrical structure 260. A plate hole 293 corresponding to a case hole 53 that will be described later is formed in the center of the plate structure 290.

[0166] A case 50, a cover 51, a radiation sphere 52, a case hole 53, a fixing member 63, a bolt portion 63a, and a nut 64 are the same as in the energy amplification generator shown in FIGS. 29 and 34 and thus, a detailed description thereof will be omitted.

[0167] An energy amplification generator according to another embodiment of the present invention will now be described.

[0168] FIG. 44 is an energy amplification generator according to another embodiment of the present invention, FIG. 45 is a perspective view of a first geometrical structure of the energy amplification generator of FIG. 44, FIG. 46 is a perspective view of a second geometrical structure of the energy amplification generator of FIG. 44, FIG. 47 is a perspective view of another shape of the second geometrical structure of FIG. 46, and FIG. 48 illustrates the structure in which a second geometrical structure, a first geometrical structure, a plate structure and separation structures are stacked, of the energy amplification generator of FIG. 44. Here, like reference numerals in FIGS. 29, 34, and 39 denote like elements having the same function.

[0169] Referring to FIG. 44, the energy amplification generator according to another embodiment of the present invention includes a first geometrical structure 360 in which both-end vertices at two continuous sides of each of seven pieces of planar structures 6 having a regular heptangular shape contact one another, a second geometrical structure 370 which is separated from an upper portion of the first geometrical structure 360 and in which both-end vertices at two continuous sides of each of seven cubic structures 7 having a regular heptangular pyramid shape contact one another, and a plurality of separation structures 8 which separate the first geometrical structure 360 and the second geometrical structure 370 from each other, are installed between the planar structure 6 and the cubic structure 7 and have a smaller area than areas of the planar structure 6 and the cubic structure 7. Here, the first geometrical structure 360 is one or two or more.

[0170] The first geometrical structure 360 implodes a space energy that is full in the universe due to its geometrical structure to be indirectly and directly used.

[0171] The second geometrical structure 370 concentrates the space energy imploded by the first geometrical structure 360 in a forward direction.

[0172] Areas of the first geometrical structure 360 and the second geometrical structure 370 allow the quantity of the imploded space energy to be increased. In other words, in order to obtain a larger quantity of space energy, the number of stacked first geometrical structures 360 increases or the areas of the first and second geometrical structures 360 and 370 increase.

[0173] In this case, in the first geometrical structure 310, the planar structure 6 further includes a planar cut groove 6a in which the planar structure 6 is cut to two continuous sides between two contacting vertices from the center, and the cubic structure 7 further includes a cubic cut groove 7a in which the cubic structure 7 is cut to two continuous sides between two contacting vertices from the center. In addition, the separation structures 8 have a regular heptangular shape but also have a cut groove 8a in which each of the separation structures 8 is cut to two continuous sides from the center.

[0174] The cubic structure 7, the separation structures 8, and the planar structure 6 are stacked that the arrangement angle of their respective sides are identical.

[0175] On the other hand, a second geometrical structure 370' may be modified in another shape. In other words, as illustrated in FIG. 47, a cubic structure 7' of the second geometrical structure 370' is implemented by forming a subcubic structure 7b' having a regular heptangular pyramid shape having a smaller area than the area of a regular heptangular planar plate 7a' on the planar plate 7a'. At this time, a cubic cut groove 7c' in which the cubic structure 7' is cut to two continuous sides between two contacting vertices of the planar plate 7a' from the center is formed in the cubic structure 7'. Seven cubic structures 7' are disposed on a plane so that both-end vertices at two continuous sides of each of seven cubic structures 7' contact one another, thereby implementing the second geometrical structure 370' of the present application.

[0176] A plate structure 390 supports the first geometrical structure 360 and has the same shape as that of the first geometrical structure 360. A plate hole 393 corresponding to a case hole 53 that will be described later is formed in the center of the plate structure 390.

[0177] A case 50, a cover 51, a radiation sphere 52, a case hole 53, a fixing member 63, a bolt portion 63a, and a nut 64 are the same as in the energy amplification generator shown in FIGS. 29, 34, and 39 and thus, a detailed description thereof will be omitted.

[0178] The space energy implosion unit shown in FIGS. 9, 14, 19, 24, 29, 34, 39, and 44, a space energy which is a kind of a torsion field, i.e., a small space energy that is full in the universe, is imploded and is converted into a usable energy to be directly or indirectly used due to the above-described geometrical structure.

[0179] An implosion mechanism of a space energy has not clearly been investigated scientifically yet. However, the present applicant has found a structure in which a space energy can be most effectively radiated, through many repetitive experiments. Such a space energy does not have the intensity that can be directly perceived by human's five senses but its existence can be verified through several experiments.

[0180] In other words, when a case where a space energy is radiated on a particular object and a case where a space energy is not radiated on a particular object are compared, if any difference therebetween occurs, it can be inductively known that a space energy exists. Hereinafter, various experimental examples for verifying the existence of a space energy will be described.

[0181] (1) Cancer Cell Growth Experiment

[0182] FIG. 55 is a graph showing tumor volumes of white rats that have drunk a general drinking water and white rats that have drunk a space energy processing water.

[0183] The present experiment has been carried out by professor Hyunwon Kim of Wonju medical college of the Yonsei University, April 2004 and his researchers.

[0184] In order to carry out the present experiment, after two chambers where experiment white rats C56BL/6 live were prepared, a water supply tube in which a processing water processed by radiating a space energy on a general drinking water was put was installed in one of the chambers, and a water supply tube in which a general drinking water was put was installed in the other one (a control group) so that the space energy processing water or the general drinking water can be freely taken by subjects from the water supply tubes that are put in each chamber. The same number of experiment white rats C56BL 6 as experiment white rats C56BL/6 to whose subcutaneous tissues malignant skin cancer cells B16 Melanoma are injected were put in each chamber and tumor volumes were observed according to days for 20 days.

[0185] In the graph of FIG. 55, a portion marked by control indicates tumor volumes of white rats that have drunk a general drinking water according to days, and a portion marked by life energy indicates tumor volumes of white rats that have drunk a space energy processing water according to days.

[0186] As a result of measuring tumor volumes for 20 days, as shown in the graph of FIG. 55, in case of white rats that have drunk a space energy processing water that is imploded by an energy amplification generator, tumor volumes are very smaller than those of the control group.

[0187] As known from the graph of FIG. 55, in case of white rats (life energy) that have drunk a space energy processing water, tumor volumes are very smaller than those of white rats (control) that have drunk a general drinking water. Thus, the space energy processing water makes an anticancer action on a living body of a white rat. That is, the existence of the space energy can be verified from the graph.

[0188] (2) Clinical Tests Relating to Whether a Processing Water that is Processed by a Space Energy Affects Infant Atopic Patients or not.

[0189] FIG. 49 shows a graph of Table 1 and illustrates the number of subjects before taking a space energy processing water. Also, FIG. 50 shows a graph of Table 2 and illustrates the number of subjects that show Erythema symptoms after taking a space energy processing water, FIG. 51 shows a graph of Table 2 and illustrates the number of subjects that show Edema Papulation symptoms after taking a space energy processing water, FIG. 52 shows a graph of Table 2 and illustrates the number of Oozing subjects after taking a space energy processing water, FIG. 53 shows a graph of Table 2 and illustrates the number of Excoriation subjects symptoms after taking a space energy processing water, and FIG. 54 shows a graph of Table 2 and illustrates the number of Lichenification subjects after taking a space energy processing water.

[0190] The clinical tests have been carried out at Chungnoi oriental hospital (principal Jaekyu Ryu) located in the third floor, Winplus commercial building, 507, Daechi-dong, Kangnam-gu, Seoul, Republic of Korea, in early 2005. The clinical tests are experiments for verifying whether, when infant atopic patients took a space processing water, the space processing water affects the infant atopic patients clinically or not.

[0191] In the clinical tests, D-company spring water (hereinafter, referred to as a space energy processing water) in which the space energy that is radiated from the energy amplification generator according to the present invention is investigated for 10 days, was taken by 40 infant atopic patients (male 15, female 25) of 0-4 years old who visited at the Chungnoi oriental hospital for atopic treatment.

[0192] In the clinical tests, five hypotheses that, when infant atopic patients took the space energy processing water, there may be a difference between before and after taking the space energy processing water were formed, and an atopic index ((SCORAD: SCORing Atopic Dermatitis) index)) before and after taking the space energy processing water was measured, thereby verifying a clinical effect of a space energy processing water.

[0193] Here, the atopic index (SCORAD index) has clinical significance only when the whole body skin status, partial skin photos and itching are simultaneously checked and a resultant value thereof is obtained. However, there was a difficulty for checking the whole body skin status and itching because subjects were infants. Thus, Erythema, Edema Papulation, Oozing, Excoriation, and Lichenification which are main atopic symptoms were scored and classified.

[0194] The symptoms of atopic patients before taking a space energy processing water are shown in Table 1 and FIG. 50.

[0000]

TABLE 1

Subjects Symptoms before taking Space Energy Processing Water

Status	Erythema	Edema Papulation	Oozing	Excoriation	Lichenification
Severe(III)	5	3	3	0	2
Moderate(II)	25	20	15	1	1

Mild(I)	8	17	2	5	1
0(No symptom)	2	0	20	34	36

[0195] Table 1 is expressed as a graph, as shown in FIG. 51.

[0196] Five hypotheses that, when infant atopic patients took a space energy processing water, there may be a difference between before and after taking the space energy processing water are as described below.

[0197] (a) Hypothesis 1

[0198] Erythema symptoms of infant atopic patients who took 200-300 cc of a space energy processing spring water for one month three times a day for 10 days may be different from those before taking the space energy processing spring water.

[0199] (b) Hypothesis 2

[0200] Edema Papulation symptoms of infant atopic patients who took 200-300 cc of a space energy processing spring water for one month three times a day for 10 days may be different from those before taking the space energy processing spring water.

[0201] (c) Hypothesis 3

[0202] Oozing symptoms of infant atopic patients who took 200-300 cc of a space energy processing spring water for one month three times a day for 10 days may be different from those before taking the space energy processing spring water.

[0203] (d) Hypothesis 4

[0204] Excoriation symptoms of infant atopic patients who took 200-300 cc of a space energy processing spring water for one month three times a day for 10 days may be different from those before taking the space energy processing spring water.

[0205] (e) Hypothesis 5

[0206] Lichenification symptoms of infant atopic patients who took 200-300 cc of a space energy processing spring water for one month three times a day for 10 days may be different from those before taking the space energy processing spring water.

[0207] The result of the clinical tests is shown in Table 2.

[0000]

TABLE 2

Subjects: Symptoms after taking Space Energy Processing Water.

Status	Erythema	Edema Papulation	Oozing	Excoriation	Lichenification
Severe(III)	0	1	1	0	2
Moderate(II)	14	10	0	1	1

Mild(I) 13 12 1 4 1

0(No 13 17 38 35 36

symptom)

[0208] (a) Result of Hypothesis 1

[0209] In Erythema symptoms of infant atopic patients, there was a significant difference between before taking 200-300 cc of the space energy processing water three times a day for 10 days and after taking it and treatment. In other words, as illustrated in FIG. 50, in Erythema symptoms of patients after treatment by taking the space energy processing water, the status of a number of patients who belong to 'severe' and 'moderate' of Table 2 is moved to the status of mild or '0(no symptom)', that is, symptoms have been improved.

[0210] (b) Result of Hypothesis 2

[0211] In Edema Papulation symptoms of infant atopic patients, there was a significant difference between before taking 200-300 cc of the space energy processing water three times a day for 10 days and after taking it and treatment. In other words, as illustrated in FIG. 51, in Edema Papulation symptoms of patients after treatment by taking the space energy processing water, the status of a number of patients is moved to the status of 0(no symptom), that is, symptoms have been improved.

[0212] (c) Result of Hypothesis 3

[0213] In Oozing symptoms of infant atopic patients, there was a significant difference between before taking 200-300 cc of the space energy processing water three times a day for 10 days and after taking it and treatment. In other words, as illustrated in FIG. 52, in Oozing symptoms of patients after treatment by taking the space energy processing water, the status of a number of patients is moved to the status of '0(no symptom)', that is, symptoms have been improved.

[0214] (d) Result of Hypothesis 4

[0215] In Excoriation symptoms of infant atopic patients, there was no significant difference between before taking 200-300 cc of the space energy processing water three times a day for 10 days and after taking it and treatment. In other words, as illustrated in FIG. 53, in Excoriation symptoms of patients after treatment by taking the space energy processing water, the Excoriation symptoms of most subjects were not severe, and thus, it was a difficulty for finding significance in comparison with the results before and after taking the space energy processing water. However, it was verified by naked eyes that the skin status after taking the space energy processing water is more smooth than before taking the space energy processing water.

[0216] (e) Result of Hypothesis 5

[0217] In Lichenification symptoms of infant atopic patients, there was no significant difference between before taking 200-300 cc of the space energy processing water three times a day for 10 days and after taking it and treatment. Since Lichenification symptoms occur in adult chronic serious-case patients having a long disease-contraction period, as illustrated in FIG. 54, many Lichenification symptoms have not been found in infants having a short disease-contraction period of less than 4 years old.

[0218] As verified in the above results, symptoms such as Erythema (hypothesis 1), Edema Papulation (hypothesis 2), Oozing (hypothesis 3), and Excoriation (hypothesis 4) have been improved. Although, in case of Lichenification (hypothesis 5) symptoms, there was no significant difference, the existence of a space energy can be verified according to the results of hypotheses 1, 2, 3, and 4.

[0219] The space energy radiated from the energy amplification generator according to the present invention allows a living body energy to be activated, as verified in the above-described cancer cell growth experiments or clinical tests. However, the space energy shows various effects in the following fields.

[0220] If the space energy radiated from the energy amplification generator is radiated on a plant, the growth of the plant is brought on, and chemical reactions such as decay etc. are suppressed. Thus, it can be understood from the results that one action of the space energy according to the present invention is a non-ionization action for suppressing ionic polarization. Such a phenomenon occurs even in a human body and is as described above.

[0221] In addition, when the space energy is radiated on milk at about 20[deg.] C. for about three weeks, milk is not decayed but can be eaten and is changed into yogurt in which lactobacillus that has also been used in a health food is contained.

[0222] In addition, when the space energy is radiated on a strawberry for two months, the strawberry is not decayed or acidified, is dried and dehydrated and is changed into a glutinous shape like raisins. This does not harm the human body when man takes the strawberry. This represents a processing method by which foods can be kept for a long time, comparing when strawberries are kept in a general environment, for more than three days, they begin to be deformed and be decayed after several days have elapsed.

[0223] In addition, when the space energy is radiated on a cold-storage apple for about two or more years, due to moisture evaporation, part of the epidermis is crumpled and its interior fleshiness is not greatly different from the fleshiness of an original apple. In this way, when the present invention is applied to living body objects around man, their living body energy can be preserved and reinforced.

[0224] While the present invention has been particularly shown and described with reference to exemplary embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the spirit and scope of the invention as defined by the following claims.

INDUSTRIAL APPLICABILITY

[0225] As described above, in the space energy implosion unit and the energy amplification generator using the same according to the present invention, a space energy which is a kind of a torsion field and is imploded in a particular geometrical shape, is more effectively imploded so that the space energy can affect the human body and animals and plants favorably and furthermore, affect inanimate objects favorably.

WO2009101466

ABSOLUTELY NEW TECHNOLOGY OF COLLECTING ENERGY AND GENERATING ELECTRICITY. THE NEW TYPE NUCLEAR POWER REACTOR

Abstract

The absolutely new technology of collecting energy and generating electricity by using huge Pyramid or a Cone (approximate size and angles see figure 7) as a nuclear power reactor, that should be made of big blocks of natural stones (limestones and granites stones) weighing 3-5 tons and larger that should be brought into a very tight contact with a cement, with ideally polished, smooth Pyramid's surface and special structure of the Pyramid's inner chambers and special chambers under the Pyramid's. The process of collecting energy : The Pyramid or Cone starts the process of a nuclear chain reaction that induce the gain of temperature inside and under the Pyramid. Supplying water into Pyramid's inner chambers and special chambers under the Pyramid and collecting steam under very high pressure and very high temperature for spinning a steam turbines. The approximate plan of the Pyramid's special structure see figure 7.

DESCRIPTION

I invented the absolutely new, yet unknown for the mankind super technology of collecting energy and generating electricity and absolutely new NUCLEA POWER REACTOR (NPR). My NUCLEAR POWER REACTOR works without any fuel! ! ! My NPR produces absolutely CHEAP ENERGY! ! ! Because my NPR does not need any fuel it takes only operating costs and outlay for its building. My NPR produces absolutely ENVIRONMENTAL FRIENDLY, SAFE, STABLE ENERGY WITHOUT LIMITS!!! The power of my NPR depends on its size and supply of water. The larger size of the reactor the more power the reactor produces, the more water Power station requires. As my NPR needs good supply of water it should be built near large enough source of water. Ideally it could be very big river or big enough freshwater lake. Before we start building NPR we need to chose .a natural earthquake-resistant platform (some plateau or outcrop of bedrock). The earthquake resistance of the platform is the most important factor. The natural weight of my NPR is very high (from aboutlO million tons and higher) so as a platform we should take some very solid monolith plateau. Another reason for this is that part of this base under the Pyramid is the very important part of the reactor. For water supply we need to create a special harbour from the river as close as possible to the Power station.

MY NUCLEAR POWER REACTOR AND MY TECHNOLOGY OF COLLECTING ENERGY AND PRODUCING ELECTRICITY IS UNIQUE. The reactor and technology that our civilization has not known yet. Without a doubt this is the best reactor and technology that have ever been created. My NPR is using the ideal way of collecting energy. Technology that mankind will use for producing electricity from now on and forever. Technology that will end up any other main ways of generating electricity, that will save our planet from pollution and poisoning and even will save our civilization in the case of some global catastrophe that already has happened on our planet in the past (global total flood), because my NPR and my technology works non-stop, forever without any fuel even deep under water surface, on the ocean's floor where people could hide and live during total flood or ice age. My technology will allow people to get energy even on other planets if this planets have enough water.

The reactor of my Power station is a PYRAMID(it can be a CONE as well, but PYRAMID is much better) that should be made from blocks of solid enough material(natural stones, cement...) It could be better to use granite and limestones. All body of the Pyramid should be made from huge blocks of limestones. Some parts of the interior design that should resist very high pressure and high temperature should be made from huge blocks of high quality granite. The type of the Pyramid is a four sided Pyramid(It could be three, five, six and much more sided Pyramid even as I have said before a Cone). Four sided Pyramid is much better. The approximate optimum size of the Pyramid see figure 1. The position of the Pyramid on Earth see figure 2 (the view from above for the reactors situated in the northern hemisphere). The angles of the Pyramid see figure 1. Inside the Pyramid and on its surface it is very important to minimize the thickness of the joints between stones. Ideally it should be monolith structure. All blocks of stones should be brought as close as just 0,001 inches, or ideally in very tight contact with cement between stones. The precision in the consistency of the lengths of the Pyramid's four sides at the base as well as its four sides from the capstone to the base, all angles, positioning the Pyramid to the real north, the real horizontal position of its base are the most important things! The power of my NPR depends on this precision and accuracy. The precision should be as good as today's technology can afford it. There should be large enough limestones weighting at about 3 - 5 tons and more to create firm enough structure. All this Pyramid should be made from such stones because we expect very high pressure of steam in special inner chambers of the Pyramid hence the structure should successfully resist high pressure and high temperature at the same time! The surface of the Pyramid should be perfect as a mirror. It should be made from ideally polished limestones that should be cat and fitted together with a cement with ideal precision, with joints almost invisible to the naked eye. All four sides of the Pyramid should look like a single, flawless, pyramid- shaped slab. As I have said before there should be perfect monolith structure inside the Pyramid as well. If we successfully fulfil all this conditions the Pyramid will start working. The Pyramid starts a nuclear chain reaction and starts to accumulate energy in its inner space and under the Pyramid. In this case to collect energy from the core of the Pyramid and from special chambers under the Pyramid we need water.

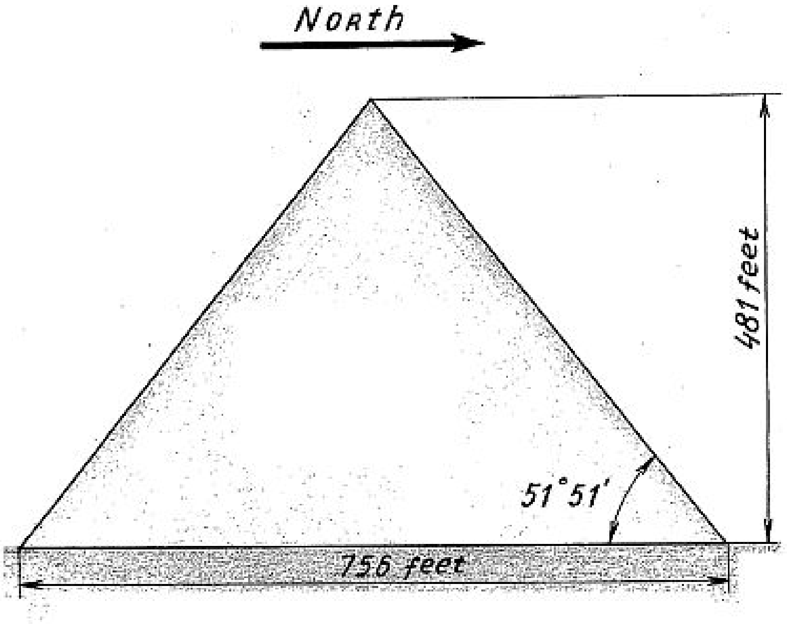
North
481 feet
51°51'
756 feet

Figure 1.

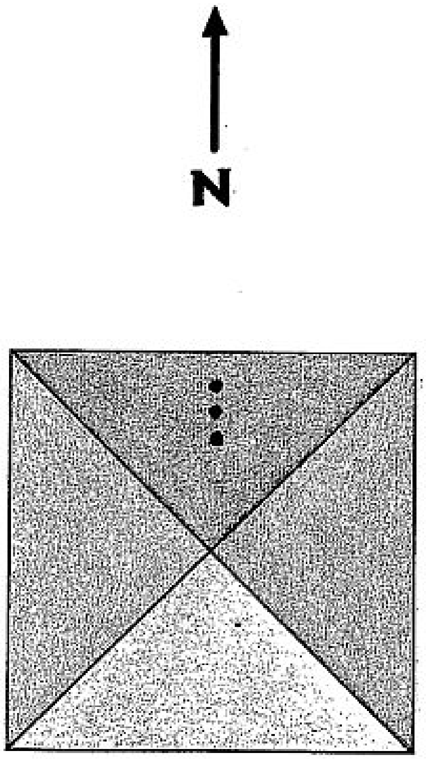

Figure 2.

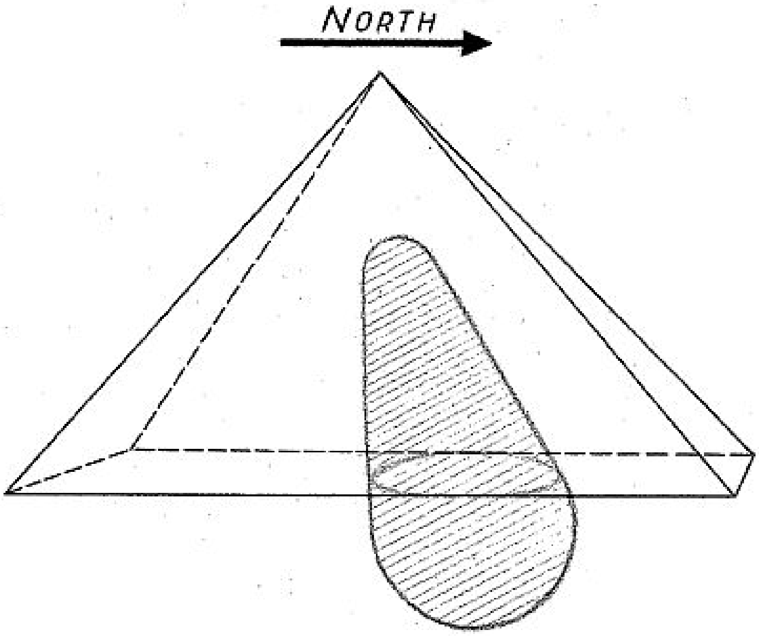

Figure 3.

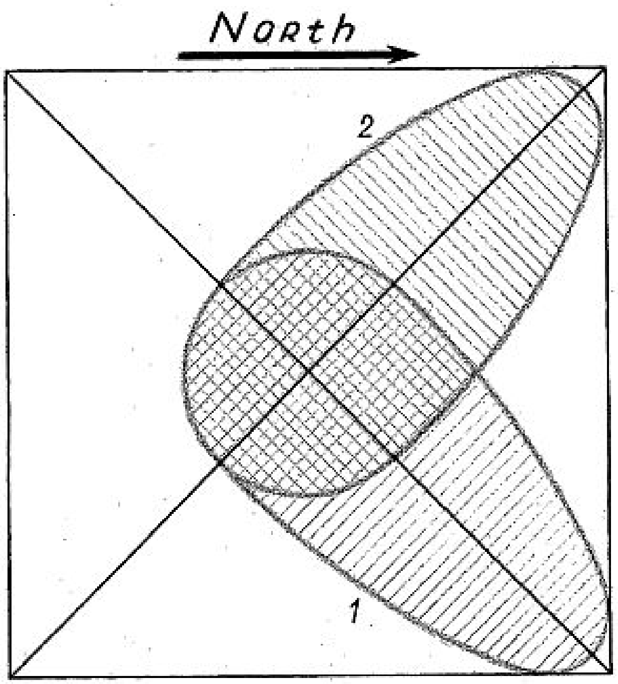

Figure 4.

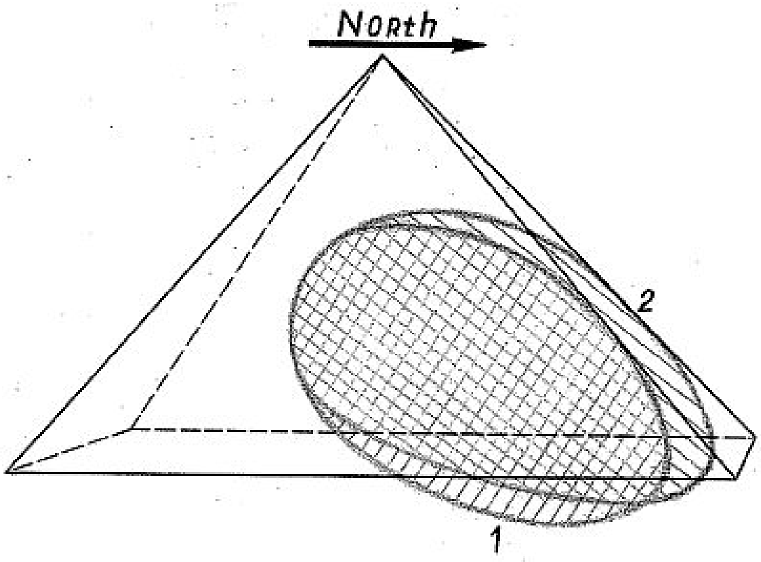

Figure 5.

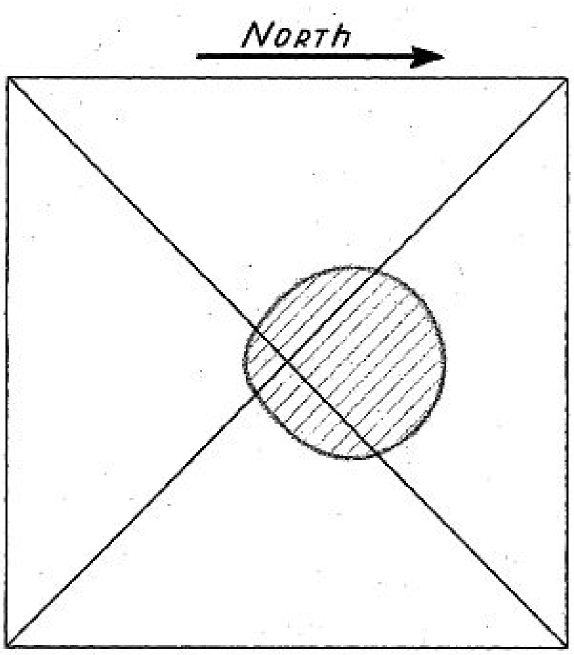

Figure 6.

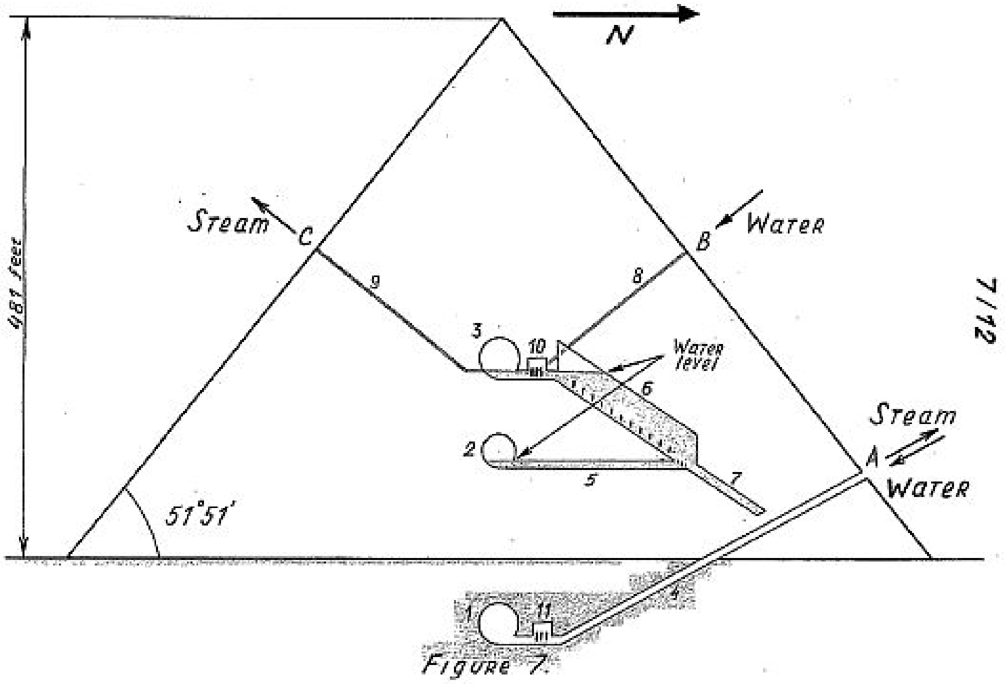
N
Steam
C
Water
B
9
8
3
10
Water level
6
2
5
7
Steam
A
Water
481 feet
51°51'
1
11
4
7/12

Figure 7.

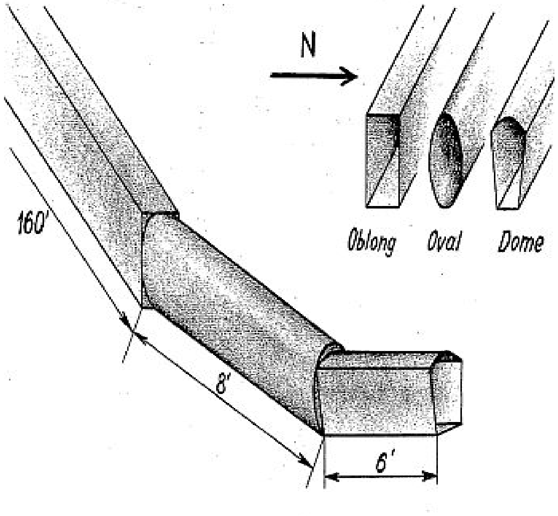
N
160'
8'
6'
Oblong
Oval
Dome

Figure 8

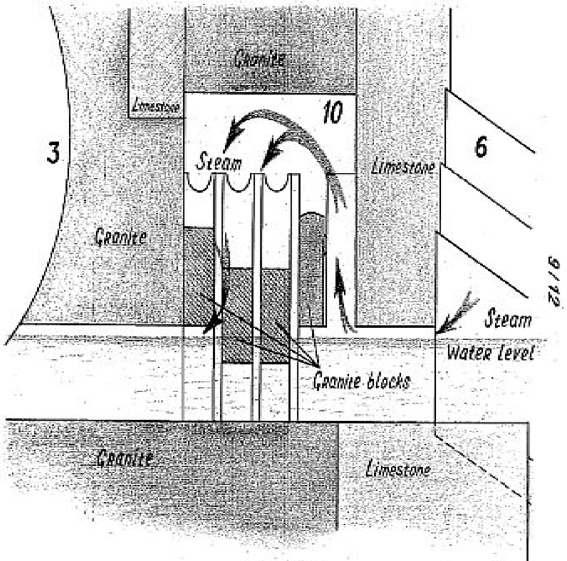
Granite
Limestone
10
3
Steam
Limestone
6
Granite
9/12
Steam
Water Level
Granite blocks
Granite
Limestone

Figure 9.

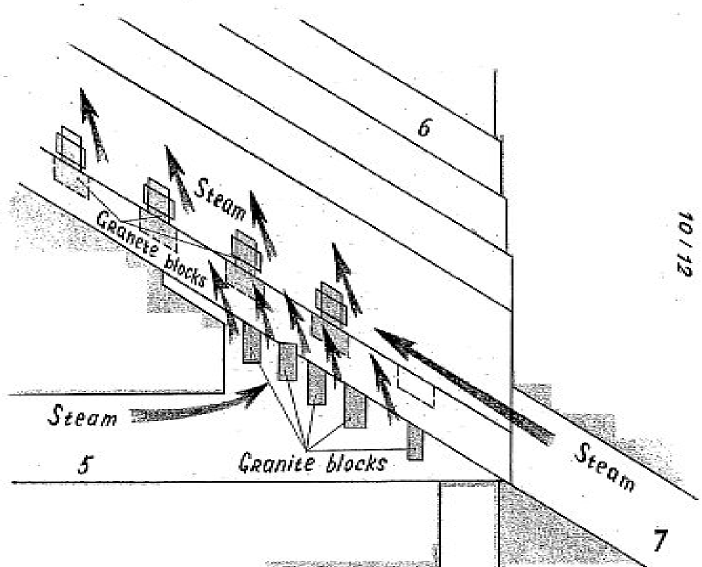
6
Steam
Granete blocks
10/12
Steam
5
Granite blocks
Steam
7

Figure 10.

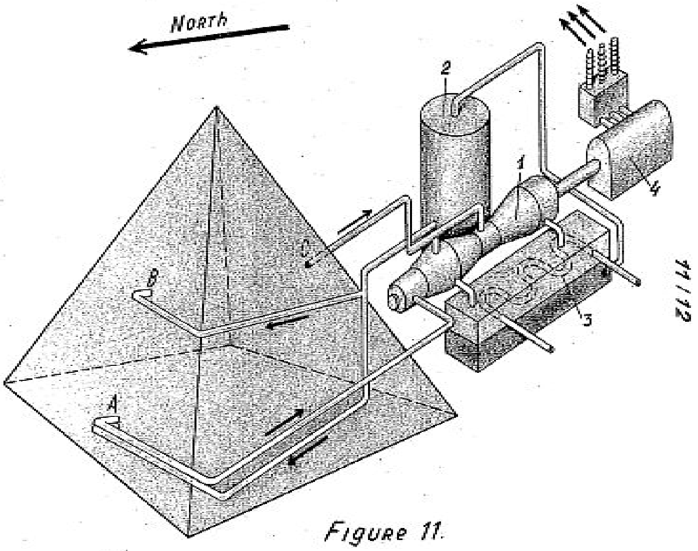
North
2
1
4
C
B
3
11/12
A

Figure 11.

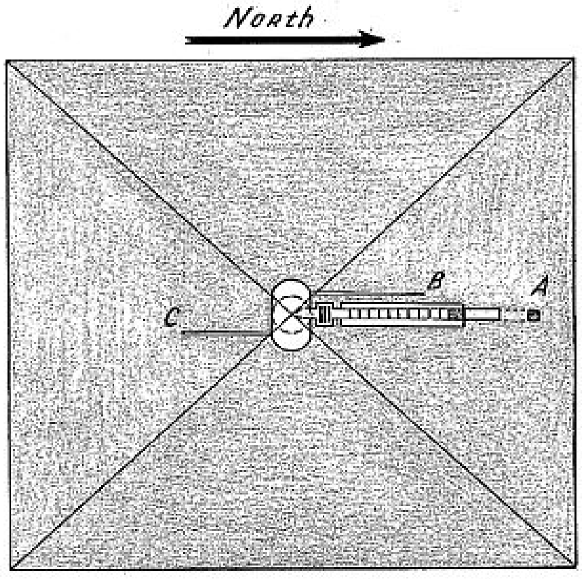

Figure 12

To achieve very precise horizontal position of the Pyramid's base we shall have to level the surrounding plateau by fitting it with a limestone pavement, carved to match the profile of the underlying bedrock. This pavement should be built from fine quality limestones and should be extended some 40 feet beyond the Pyramid's boundary on each of its four sides all around the Pyramid. Of course in this case we shall have to extend the precision of the Pyramid on the pavement as well. The mean thickness of the joints between limestones should be not more then <-> 0,001 inches or ideally it should be very tight contact with cement.

Before we start building the Pyramid's casing the whole Power station, the whole its infrastructure should be absolutely ready to work. Because having started building the Pyramid's casing we START THE REACTOR OF THE POWER STATION. It means that building of the Pyramid's casing is THE LAST STEP. During the process of building the Pyramid's casing we will start collecting first kilowatts of electricity!

And now it is high time to have a look inside and under the Pyramid. As I have said before we expect very high pressure of steam in its inner chambers and in special chambers under the Pyramid and we expect very high temperature as well. To resist both we need to create special structure of chambers and tunnels inside the Pyramid from huge blocks of high quality granite. All granite stones should be brought together in very tight contact with a cement. We need to carve special structure of chambers and tunnels in the base under the Pyramid. The positioning of the Pyramid's inner chambers and special chambers in the base under the Pyramid see figure 7. The view from above see figure 12. The process of collecting energy is quite simple: The Pyramid itself starts a nuclear chain reaction and starts to accumulate energy in its inner space and in the base under the Pyramid. The nuclear chain reaction induce the gain of temperature in the Pyramid's inner space and in the base under the Pyramid. Some parts of the Pyramid and the base will get very high temperature. In this case we need water as a moderator and as a coolant at the same time. Water takes heat from the core of the Pyramid and special chambers and runnel under the Pyramid and converts into steam. Eventually we will use steam to drive a steam turbines as normal in any power station.

And now about very important thing: Before we start planning the Pyramid's inner chambers and special chambers in the base under the Pyramid and building our first real huge Pyramid we need very carefully check the map of allocation of temperature of different parts of the Pyramid and how this map is changing during the day and during the year. It is very difficult to know about this without having made an experiment on a smaller model first. So we need to build the very precise smaller copy of our huge Pyramid first. We need to install a lot of very sensitive temperature sensors all over this smaller model and, that is very important, under the Pyramid's base as well to check temperature of different parts of this smaller model and under it. We shall have to wait for a one year so that we could check how temperature of different parts of our smaller model and under it is changing during the whole day and during the whole year. The approximate map of allocation of temperature inside the Pyramid and under the Pyramid is on figures 3 and 5. The view from above see figures 4 and 6. During the day the active zone of energy concentration is changing from position (2) on figure 5 (view from above of this position see figure 4.) to position that is shown on figure 3 (view from above of this position see figure 6.) and then to position (1) on figure 5 (view from above - position 1 on figure 4.). This process takes 12 hours. In the active zones of energy concentration we expect very high temperature. To be certain about safety we need to install radiation sensors all around outside our smaller model too. Of course without making experiments no one knows how the size enlargement of the Pyramid could affect the power. I suppose it is quit possible that enlargement twice as much of the size of the Pyramid could bring three or even four times more power! Who knows it could be even more. I think that the Pyramid as huge as volcano Vesuvius which is CONE could supply enough energy for the whole world!!! The problem is: we could not be able to supply enough water for so huge reactor. All about pyramids, cones and volcanoes as well we could know only making experiments and calculations on a smaller models. I have no doubts all experiments will pay for themselves more then enough.

It would be very advisable to built smaller copy of the Pyramid on the rotating platform so that we could know how power and allocation of temperature all over the Pyramid is changing if we wind the Pyramid around its axis.

Having know the allocation map of temperature inside the Pyramid and even deep under the Pyramid (which is very important as well), how this temperature is changing during the day and during the year we could eventually plan the disposition of the Pyramid's inner chambers and their design. We could plan the power of the reactor, the temperature and the pressure inside, how much water do we need and so on. In fact we could then compare the power of my reactor with powers of already existent other types of reactors and eventually plan the rest infrastructure of the power station which will be the same as in any other types of power stations (see figure 11.). The steam from exit (C) of the Pyramid is passed directly to the steam turbine (1) where it is joined by the steam from exit (A) of the Pyramid. After expanding outwards through the turbine, the steam has had all of its useful energy removed. It is then condensed by passing over pipes carrying cold water (3): this water in turn loses its heat in cooling towers. Condensed water is pumped into a cold water tank (2) and then into entrance (B) and entrance (A) of the Pyramid. The blades of the steam turbine (1) turn at 3600 rpm, spinning a shaft connected to a generator (4) producing 22 kV of three-phase electricity.

As I have said before the allocation map of temperature all over the Pyramid and under the Pyramid is not constant. It is changing during the day and during the year. It depends on positioning of the Pyramid on Earth - on the latitude on which the Pyramid lie. It means that during the day different chambers of the Pyramid and the base under the Pyramid work actively at different times. Although all chambers of the Pyramid and the base are generating enough power as a steam for the whole day, non-stop because of the Pyramid's capability to accumulate and keep enough energy as a heat for a long period of time. Now we have a look at figure 7. This is just approximate plan of the reactor's inner chambers that I have made without experiments on a smaller model. The disposition of the reactor's inner chambers depends on the Pyramid's positioning on Earth (latitude). The plan that we can see on figure 7 is for the reactors which should be situated in the northern hemisphere. The structure inside the Pyramid consist of two separately working parts. The first part consist of chamber (1), antechamber (11) and tunnel (4). The second part consist of chamber (2), tunnel (5), tunnel (7), chamber (6), chamber (3), antechamber (10), tunnels (8) and (9). During the day some of this two parts starts working actively first. Water is coming into chamber (1) through the entrance (A), tunnel (4) and antechamber (ll).Because of very high temperature all around chamber (1), antechamber (11) and tunnel (4), water starts boiling right in tunnel (4) converting into steam. The process of steam generation continues in chamber (1). Because of the very high temperature all around this chamber the temperature and pressure of steam in chamber (1) is growing. When the pressure of steam is high enough steam pushes water level down, comes out of the chamber and hit into granite blocks in antechamber (11). This granite blocks is especially meant to withstand the steam's blow from the chamber (1). Steam is coming out under big pressure through tunnel (4) and entrance (A) to a steam turbine. The approximate size of tunnel's (4) section is 3 feet 11 inches high by 3 feet 5 inches wide. The active process of growing temperature all around this part continues for about four hours, but the process of steam generation continues all day round non-stop. The second part of the Pyramid's inner structure works independent of the first part. Water is coming into this part through the entrance (B) and tunnel (8). Then from chamber (3) water is coming through antechamber (10) into chamber (6) and further into tunnel (7), through tunnel (5) into chamber (2). Water level in both (2) and (3) chambers see figure 7. The very high temperature all around this part of inner structure starts the process of steam generation in all chambers and tunnels. In chamber (2) steam under big pressure pushes down water level and comes out into tunnel (5). Big bubbles of steam on the way into chamber (6) hit into five granite blocks (see figure 10.) This granite blocks is especially meant to withstand the steam's blow from chamber (2). This granite blocks slice and disperse big steam bubbles into smaller fractions. And at the same time this granite blocks do not allow steam bobbles come into tunnel (7) because of their special overlap position. The bubbles of steam from tunnels (5) and (7) come into chamber (6). Here, in chamber (6) we can see special structure of granite blocks fixed on special ramp (see figure 10.). This granite blocks is especially meant to direct steam bubbles upwards on their way to chamber (3) and antechamber(l[theta]) so that steam bubbles do not get into antechamber (10) straight but come to the water surface in chamber (6). The space in chamber (6) over water surface accumulates steam. The pressure and temperature of steam in chamber (6) over water surface is growing. When the pressure is high enough steam pushes water level down, comes into antechamber (10) and hit into granite blocks (see figure 9.). This granite blocks is especially meant to withstand the steam's blow from chamber (6). Then steam comes into chamber (3). We expect very high temperature all around chamber (3). It raise the steam's temperature and pressure even higher. That is why all this chamber should be made of high quality granite blocks. Steam under big pressure and temperature pushes water level down and comes out of chamber (3) through tunnel (9) and exit (C) to a steam turbine. The active process of accumulating energy all around this part continues non-stop for about 12 hours, but the process of steam generation continues non-stop all day round. I should say that for at least 12 hours every day the Pyramid works very actively. This amount of time is far then enough for the Pyramid to accumulate enough energy for another 12 hours when the Pyramid is not accumulating energy in its inner space and under the Pyramid. Stones of the Pyramid keep enough heat for the next 12 hours till the Pyramid starts working very actively accumulating energy again in its inner space to keep on generating steam non- stop for the whole day, the whole year, the whole eternity! ! ! Yes . MY REACTOR WORKS NON-STOP, FOREVER, WITHOUT ANY FUEL ! ! !

Everything inside the Pyramid should be made very firm and reliable. Because any mistake could cost a lot of money. To change or repair something inside the reactor we shall have to stop the reactor. To stop the reactor means that we shall have to remove the upper layer of stones which is the Pyramid's casing. This is the very expensive task. Hence everything inside the Pyramid should be ready to work firmly and reliably for hundreds of years or even thousands of years nonstop. The system inside the Pyramid should be simple, without replaceable details.

Another problem is the configuration profile of the steam tunnel (9) . The steam under very high pressure and high temperature is coming out of the Pyramid with the very high speed. It means that resonance could make this tunnel very noisy. We need to choose the right profile for this tunnel to keep its prime resonant frequencies in the infra sound range. The approximate example of the steam tunnel's profile see figure 8.

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...

Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

QRP HomeBuilder - QRPHB -

Component level (mostly analog) electronics from microwave to AF. Jazz guitars and guitar amplifiers. Radio astronomy. I apply science and critical thinking in my daily life.

Showing posts with label **Inverter crystal oscillator tips**. Show all posts

Sunday, 17 January 2016

Inverter Crystal Oscillator

Greetings -- a short, first post for 2016 !

In numerous RF synthesizer chips lies an inverter with input and output pins for making a reference crystal oscillator clock. I built some discrete chip inverter xtal oscillators with 74HC series logic gates to better examine them. You'll quickly recognize the oft-used Pierce oscillator topology with 1 trimmer capacitor to tweak the fundamental frequency which might vary from factors like crystal aging and gate, crystal, crystal holder + board reactances.

I determined the 27 pF and trimmer cap values through experiments and measures.

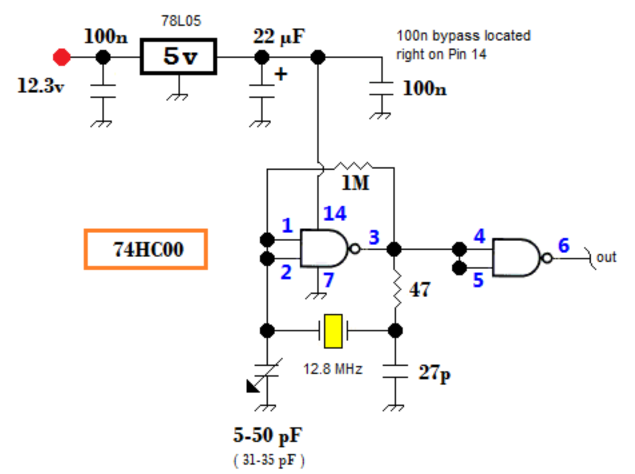

Above — A crystal reference oscillator + buffer with inverters built from NAND gates. The crystal is a good 1 — built in 2013; AT- cut; parallel 20 pF load capacitance; fundamental 12.8 MHz; a measured QuL of 265K and zero spurs during my test sweeps. Further, this crystal ages < 5 ppm per annum for at least 2 decades.

If I contrast this with some cheap xtals I bought and tested from eBay — it's night versus day. You might find such xtals in DDS and other low-cost synthesizers kits. They typically come in a HC-49S case, might suffer a QuL of 40-60K — and more alarmingly, those I measured often showed strong, close-in spurs to further trash the already compromised close-in phase noise of these low-cost synthesizers.

Translate

Select Language ▼

My Blog List

Jay's Astronomical Observing Blog

Observing and Sketches from December 19th through December 21st, 2017. - I got some observing in on December 18th and 19th, 2017 out at my dark site off of Forest Road 006 at Juniper Grove. A couple of images. A little snow...

3 hours ago

Ripples in the Ether

Arduino – Si5351 Powered Signal Generator - Arduino – Si5351 Powered Signal Generator Please enjoy this write-up of a nifty, compact signal generator project using the Etherkit Si5351A Breakout Board...

23 hours ago

VE7SL - Steve - Amateur Radio Blog

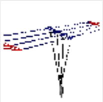

CLE 226 / CLE 227 NDB Listening Event Results - *CLE 226 and CLE 227*, both running during the Christmas to New Year period, produced some excellent propagation on the MF NDB band in most parts of the...

1 day ago

Victor Herrero Radio Astronomy Blog

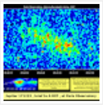

Sun Earth Jupiter 20171223 - ... *Many thanks to: Taxpayers of the United States of America, the Juno, STEREO, WIND Spacecraft Teams, the Panetary Data System Team, Prof. Dr. Kazumas...

2 weeks ago

The Signal Path

Rohde & Schwarz RTB2004 10-Bit, 2.5GS/s MSO Oscilloscope Review, Teardown & Experiments - In this episode Shahriar reviews the Rohde & Shwarz RTB2004 10-Bit oscilloscope. With its high-resolution touch-screen, intuitive and capable GUI as well a...

3 weeks ago

EA4EOZ, an amateur radio electronic enthusiast

Icom IC-R7100 and its clicking noise - Some time ago an Icom IC-R7100

Quoting Dr. Ulrich Rohde " [ALL] elements in a synthesizer contribute to noise. Two primary noise contributors are the reference and the VCO. Actually, the crystal oscillator or frequency standard is a high-Q version of the VCO" [Reference 1].

Although this post isn't about phase noise; in this era of poor quality, "cheapo" crystals, I think a low-noise reference is worth considering when synthesizing signals for specific applications that require low phase noise. Big thanks to Alexei Luk for sending me this 12.8 MHz gem.

I found a problem with my circuit as shown above: strong spikes on the positive and negative edges. My quest became finding ways to decrease these spikes and enhance the square waveform seen in my DSO

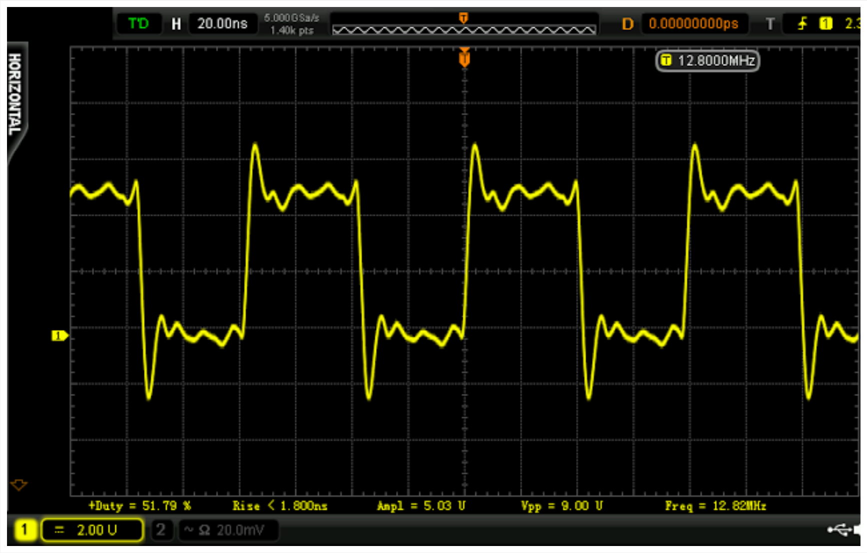

Above — My DSO screen capture of the NAND oscillator circuit show earlier. The edge peaks swings 9 Vpp. In 1 circuit they ran over 10.1 Vpp or double the rail to rail DC voltage. What's this all about?

I emailed my circuit and measures to Professor Ken Kuhn who gave me some excellent suggestions which I'll augment with experiments and apply in a circuit.

My favorite point from Ken: No matter how low the frequency you're working at, design and construct your circuits like you're operating them @ 1 GHz. Transistors don't know what frequency you're working at -- and many work well into 100's of MHz ! -- "If the circuit is built to work correctly at high frequency then it will work great at low frequency."

So from Ken's wisdom and a little of my thinking + experiments, here's what I did:

[1] Because we're operating a square wave oscillator, odd harmonics will run at high amplitude. The third harmonic at ~38.4 MHz was only 8-10 dB down from the fundamental in some of my frequency domain experiments. This means the power supply bypass cap should minimally bypass into lower VHF and go right on the DC power pin (14) with the shortest possible leads to keep its SRF as high as possible. The bypass cap should ideally offer high Q / low ESR at VHF.

[2] Apply compact construction to reduce stray C, L, — and to minimize distortion and start-up stabilization time. In particular, short ground lead lengths for the 27

arrived to my shack. It's a nice receiver and works surprisingly well but almost the very first time I powered it up an anno...

1 month ago

Circuit Salad

PWM Vibrato that sound Awesome! - Vibrato is the slight changing of frequency of a musical note. This can be accomplished by means of a varying time delay in a delay line or by phase shifti...

2 months ago

Amateur Radio Astronomy

Radio-Sky Spectrograph works with Radio-Jupiter Pro, - There are a number of ways that Radio-Sky programs can interact with one another. For example, RTL Bridge can send signal strength from a RTL dongle recei...

4 months ago

Davide Bucci's Blog

-

Pages

- Home
- Guitar Related Index

My Micro Updates

Working on guitar amp circuitry Dec 2017

My Support Pages

- My Email
- My Articles
- Old pops.net site downloads & info
- Javascript Tools
- My RF Resources
- Jupiter Modular Receiver Main Index
- Radio Astronomy Links
- Solar Ham Data and Graphs

My Links

- Ed Ting: Telescope Review site
- Etherkit
- Kuhne Microwave Components
- Luis Cupido: Deep Space Radio

Simple Post Index

- 7 MHz scratch synthesizer
- 90 Degree Phase Shift: AF - UHF
- AF power amp design
- Anton Chekhov
- Funster receiver Part 1
- Funster Receiver Part 2
- Funster receiver part 3
- Harmonic VXO
- High IIP3 mixers
- hybrid-cascode regen detector
- Inverter crystal oscillator tips
- Jupiter receiver notes. KISS mixer
- Line - in audio amplifier Part 1
- Line - in audio amplifier Part 2
- LM386 notes
- Measure Crystal Q
- Measure resonator Q at VHF
- Oscillator phase noise
- Rail to rail audio power amp

pF and trimmer caps proved important to reduce my edge spikes The shorter the circuit wires, the better rings TRUE here.

[3] Although in synthesizer chips, we only get 1 inverter with 2 pins, for off-chip oscillators a buffer proves useful. A small series resistor between the oscillator output and the buffer input pins serves to dampen any resonance in the output circuit (often in the 10s of MHz) from excitation that causes the spikes you see as the inverter switches on and off. It's affected by stray L and C around the IC.

[4] 10X scope probe capacitance lowers the resonant frequency and boosts RF energy. To minimize this loading, you might tack solder a 100 Ω (or so) resistor onto the buffer output pin and attach the probe clip to this resistor. Experiment with the resistor values in [3] and [4] to find out what works best on your bench.

[5] The 27 pF cap and trimmer cap grounds should lie as close as possible to the IC ground pin.

[6] Don't overdrive your crystal. I placed a 47 Ω resistor between the inverter output and the 27p capacitor/crystal lead and determined this R value experimentally by watching the trace and frequency counter in my DSO. Since the output resistance of the inverter driver is very low compared to the reactance of the capacitors and crystal (the crystal , trimmer and 27 pF cap form a complex impedance) the resistor isolates the output driver and also lowers the crystal drive level.

My crystal features high Q and the 74HC inverter drives it hard. Adding the resistor reduced the edge spikes slightly. Further I performed a test where I raised the DC supply voltage slightly and my signal appeared to distort and the frequency dropped slightly. The 47 Ω resistor removed this issue and stabilized my TTL inverter oscillator.

Here's the final schematic:

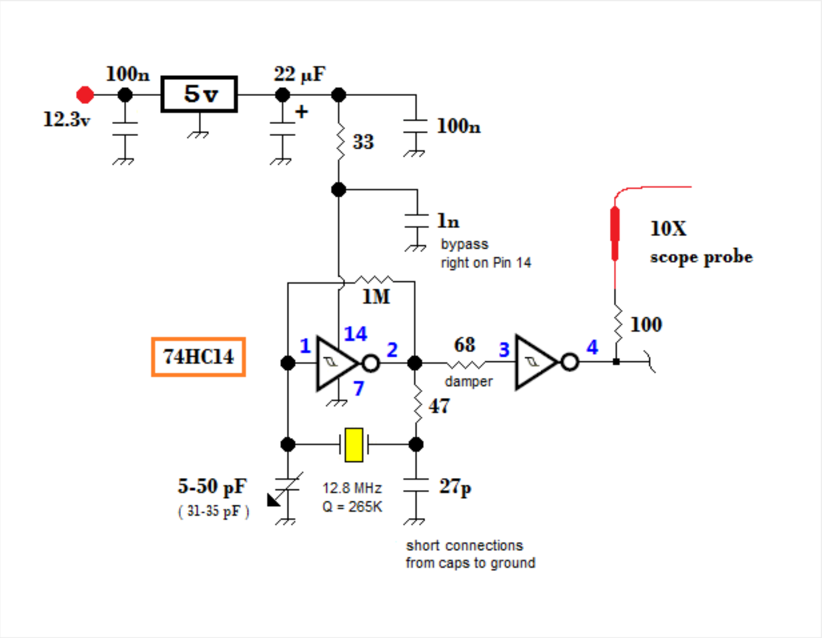

Above — My final TTL Pierce gate oscillator design with a 100 Ω resistor to isolate the 10X DSO probe during measurement. On the 5 VDC line. A 22 µF + 100nF then 33 Ω resistor plus 1 nF capacitor on Pin 14 form a pi-filter for wide band, low-pass DC filtration from AF to lower VHF. VHF bypass on pin 14 helps quench edge spikes.

- Matjaž, S53MV
- MIT Haystack
- Moonbouncers org
- My Twitter Page
- Paul, W1GHZ
- Radio Astronomy supplies
- Radio-Sky Publishing
- Reeves Radio Science
- Rigol Canada - Electro-Meters
- Society of Amateur Radio Astronomers
- Solid State Regenerative Receivers
- Test Equipment
- UK Radio Astronomy Association
- W7ZOI -- EMRFD
- YOUDAC-VA3IU

- Regen receiver #4
- Regenerative receiver #5
- SAW Resonator
- Simpler Single-Signal Direct Conversion Receivers
- Sine wave side tone
- Tone balance control
- Trixie 2 -- simple low current DC receiver
- Two tone mixer IIP3 Notes
- UHF Bypass and decoupling
- UHF RF Grounding
- VXO- based freq synthesizer
- Xtal parameters with spectrum analyzer and tracking generator

Blog Archive

- ► 2014 (13)
- ► 2015 (19)
- ► 2016 (13)
- ▼ 2017 (8)
  - ► January (3)
  - ► April (1)
  - ► November (1)
  - ▼ December (3)
    - Op-amps Make Life Better
    - Peerless Martin Taylor Maestro - a brief review
    - 1969 Fender Deluxe Reverb Restoration

The Professors

I learned another point from Ken about my earlier circuit with the NAND inputs shorted to make an inverter. It's often better to bias 1 NAND input gate to high and then use the other input for the oscillator feedback connections (inverter mode operation). This halves the input capacitance seen by the feedback network and may result in less effect tuning effect on the IC in some applications.

I tried this ---- and like when I connected the buffer inverter up, I had to adjust the trimmer cap to either re-establish high frequency oscillation, or set the desired oscillator frequency in my counter.

During my final experiments, I remembered that I purchased some 74HC14s in 2015 to build a simple HF sawtooth generator to externally quench a super regenerative receiver. The 74HC14 features 6 inverters with a Schmitt trigger input. Quickly, I built my oscillator around this chip. Further, I ordered 10 standard hex inverter 74HC04 chips for future projects. Logic ICs provide major fun!

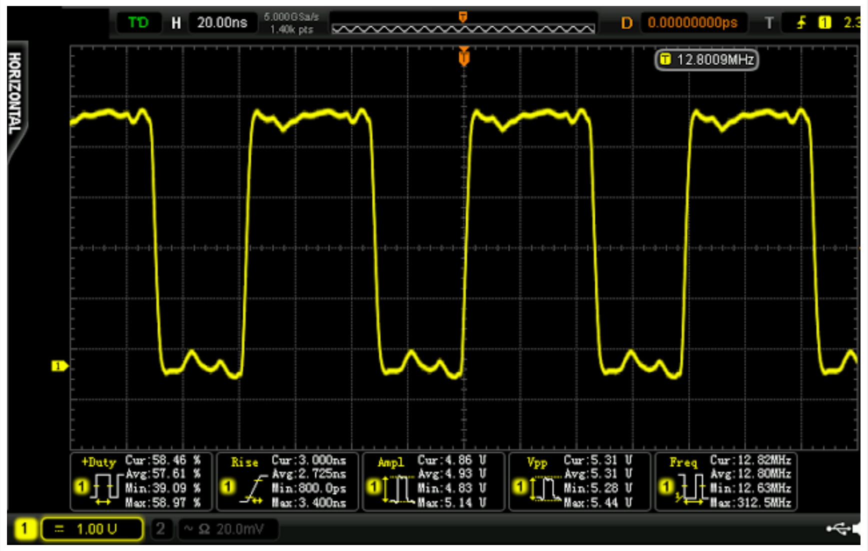
TD
H 20.00ns
5.000GSa/s
1.40k pts
D 0.00000000ps
T
1
2.3
HORIZONTAL
12.8009MHz
1
+Duty Cur:58.46 % Avg:57.61 % Min:39.09 % Max:58.97 %
Rise Cur:3.000ns Avg:2.725ns Min:800.0ps Max:3.400ns
Ampl Cur:4.86 V Avg:4.93 V Min:4.83 V Max:5.14 V
Vpp Cur:5.31 V Avg:5.31 V Min:5.28 V Max:5.44 V
Freq Cur:12.82MHz Avg:12.80MHz Min:12.63MHz Max:312.5MHz
1 = 1.00 V
2 ~ Ω 20.0mV

Above — DSO output of the improved inverter crystal oscillator. I'm quite happy with the output voltage(s) and oscillation + frequency stability now.

I also read about and performed some temperature compensation experiments. Nothing worth mentioning however. For reference purposes, here's a video of my uncompensated, board - on - bench 12.8 MHz clock into a HP, 10-digit, ovenized reference frequency counter.

Above — Video shot just after power up @ room temperature. The (temperature) frequency drift of my inverter crystal oscillator circuit appears good. This crystal will provide an excellent on-chip reference for an experimental PLL project I'm working on. The 12.8 MHz reference gets divided by 2048 in a PLL chip to achieve a tuning resolution of 6250 KHz.

Additional Bits and Pieces

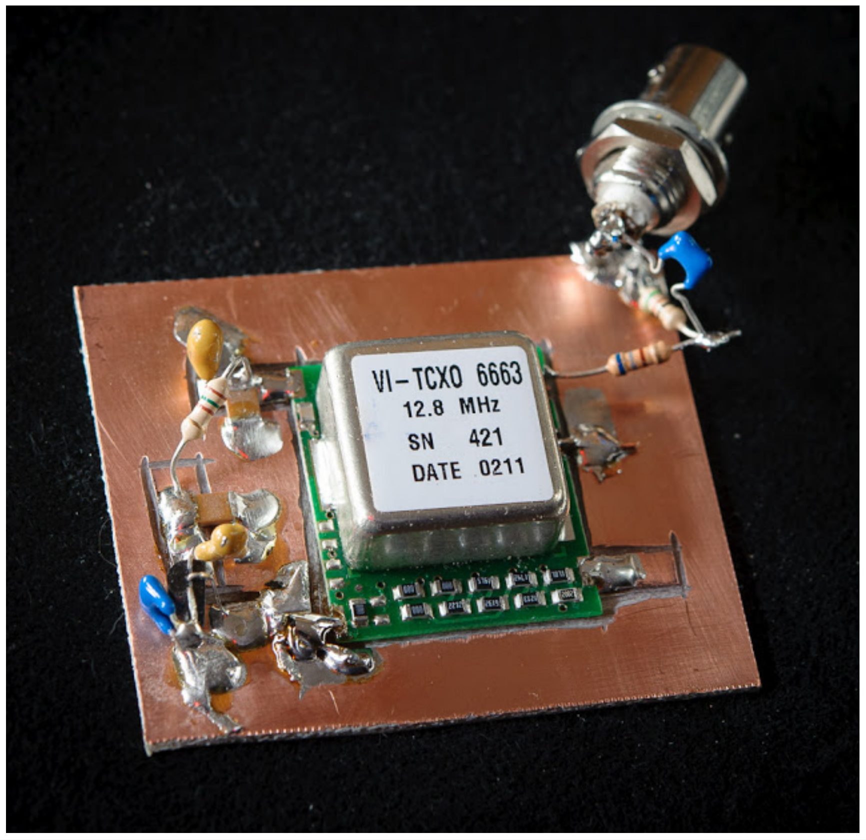

Above — Testing a commercial 12.8 MHz oscillator by Vectron International. Great engineering coupled with with a fabulous crystal results in a typical phase noise of -140 dBc/Hertz @ a 10 KHz offset — perfect for a UHF reference clock.

Above — I've always got music playing in and around my lab. Since 2006 my favorite singer = Julia Savicheva. Twitter. All of December to January I listened to Julia for creative inspiration while working on my PLL experiments. No Auto-Tune on her voice; amazing band; hard working. She sounds equally good live or recorded — how refreshing!

References

[1] Synthesizer Design for Microwave Applications. Some notes published online by Dr. Ulrich L. Rohde. Year unknown.

[2] Professor Ken Kuhn. Email correspondence Jan 2016. My sincere thanks to Ken.

[3] An Analysis of Inverter Crystal Oscillators", RF Design,Aug. 1989, Leonard L. Klein berg, pp. 28,29,3l,32.

[4] Negative Gain-Single Pole Oscillators, RF Design, Sep. 1990, Leonard L. Kleinberg, pp. 35, 36, & 38. Modern Communications

Posted by Vasily Ivanenko at 20:49 7 comments:
Labels: Inverter crystal oscillator tips

Home Older Posts

Subscribe to: Posts (Atom)

Simple theme. Powered by Blogger.

Fix the World

QEG

Quantum Energy Generator

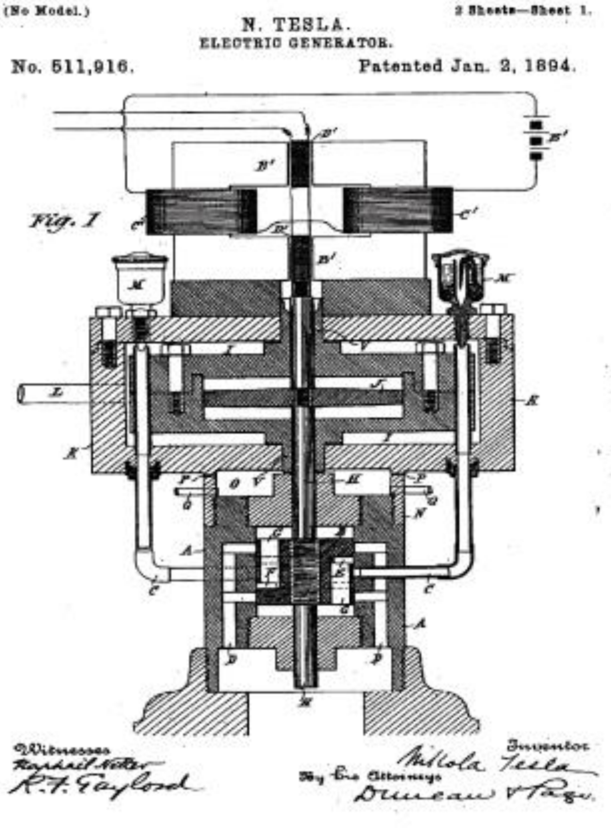

(No Model.)

2 Sheets—Sheet 1.

N. TESLA.

ELECTRIC GENERATOR.

No. 511,916.

Patented Jan. 2, 1894.

Fig. I

B'
B'
B'
B'
C'
C'
D'
M
M
L
I
V
J
I
K
K
P
O
V
H
P
Q
Q
N
G
E
A
F
E
C
C
G
A
D
D
H

Witnesses
Raphael Netter
R. F. Gaylord

Inventor
Nikola Tesla
By his Attorneys
Duncan & Page.

Nikola Tesla

QEG SYSTEM DESCRIPTION 3-25-2014

The Quantum Electric Generator system (QEG) is an adaptation of one of Nikola Tesla's many patented electrical generator / dynamo / alternator designs. The particular patent referenced is No. 511,916, titled simply "Electric Generator", and dated January 2, 1894 (see back of this manual). The adaptation is a conversion from a linear system, to a rotary system.

The QEG prototype is scaled to produce electrical power in the range of 10-15 kW (kilowatts) continuously, and can be set up to provide either 120 Volt or 230-240 Volt single phase output. We are also planning future designs to provide 3-phase power.

Service life of the device is limited only by certain replaceable components, such as bearings, v-belts, and capacitors. The basic machine should operate trouble-free (with minimal maintenance) for as long as any good quality electro-mechanical appliance, such as a quality washing machine or refrigerator. Heavy-duty mechanical components are used throughout for reliability.

The QEG is not a complicated device, as it is designed (like Tesla's other 'discoveries'), to work in harmony with natural laws, rather than with the power-wasting symmetric motor and generator designs used in today's mainstream industry.

An effective way to understand the operating principle of the QEG is to think of it as a high-powered, self-resonant oscillator (a power tank circuit), which generates high-voltage AC (15 to 25kV). These HV oscillations are then transformed into line voltage AC output, at current levels up to approximately 85 A. In today's alternative energy terminology, it would be called a type of resonance machine.

The circuitry that develops high power in this device is really based on an existing but under-utilized power oscillator configuration, however, the 'quantum' part of the design has to do with how the generator output is tuned for maximum power. Conventional alternators (AC generators) consume more input power than the output power they provide. For example, one brand of power take off (PTO) alternator uses 18,000 watts (24 horsepower) to develop 13,000 watts of output power. In the QEG, input power is used only to maintain resonance in the core, which uses a small fraction of the output power (under 1000 watts to produce 10,000 watts), and once running, the QEG provides this power to its own 1 horsepower motor. This is known as over-unity. Once the machine builds up to the resonant frequency, it powers itself (self-running).

In the QEG, the exciter coil is used to provide a conduction path through the quantum field (zero point) into the generator core. This has the effect of polarizing the core, which increases power output over time.

James M. Robitaille

NOTICE

TO BE READ AND UNDERSTOOD BY ALL QEG PROJECT CREWS

Fix the World (FTW) is not responsible for the actions of others. We can only tell you our experience. We have discovered it is essential that those wishing to build a QEG use careful thinking, patience, and consideration for the greater good.

The inhabitants of planet Earth are entering into a new paradigm and a new way of doing business. In honor of Nikola Tesla, the QEG is a gift freely given to the world, and FTW's involvement is strictly altruistic.

The QEG is an electromechanical device and as such, safety for the individual and end user should always be of prime concern. It is therefore essential that persons assembling the device are experienced in the field of electro-mechanical assembly. A considerable level of knowledge in quantum physics is also required.

IF YOU ARE TRAINED IN TRADITIONAL PHYSICS, AND HAVE NOT BEEN DOING THIS WORK FOR MANY YEARS, YOU MUST FIRST DO YOUR RESEARCH ON BASIC QUANTUM ENERGY DEVICES AND HOW THEY WORK (e.g. resonance and tuning).

Electrical / Mechanical devices are inherently dangerous. Electrical shock hazards can cause serious injury and in some cases death. Mechanical hazards can result in dismemberment and in some cases death.

Due diligence has been applied to ensure that the QEG instructions are complete and correct. All local and country-specific electrical and mechanical code implications, by which a QEG might be installed and operated, cannot possibly be known. Nor is it conceivable that any and all possible hazards and/or results of each procedure or method have been accounted for.

It is for these reasons that the QEG must be either directly installed or supervised by an experienced electromechanical engineer to ensure the installation is done safely and in accordance with local electrical code, however, the QEG is installed the same way as any commercial generator and does not violate any electrical codes. Anyone who uses the QEG installation instructions (including but not limited to any procedure or method of installation) must first satisfy themselves that neither their safety, nor the safety of the end user, will be endangered over the course of the installation and operation of the QEG.

It is imperative to understand YOU NEED PROFESSIONALS AND EXPERT ADVICE to build a QEG.

The installation instructions are designed to show how we have found the building of the device to be accomplished, and any negative outcomes that result are completely the responsibility of the person/company building it; FTW provides no guarantee for the successful installation of the QEG.

This notice serves the purpose of communicating the serious nature of building a quantum machine, as we are well aware that there have been severe restrictive agencies involved with their suppression. Quantum free energy isn't taught at University and most designers have heretofore been unsuccessful at mass distribution. It is YOUR RESPONSIBILITY therefore to make certain you are building the QEG with positive intentions for humanity, and lashing out legally or otherwise to FTW, HopeGirl and/or the designer and his family, is a violation of goodwill and will in no way be attended to. We know of no other way to do this but to go back to the "HONOR SYSTEM."

In reading this notice I agree that:

1) I WILL NOT ATTEMPT TO BUILD A QEG UNLESS I DO SO APPROPRIATELY WITH AN ELECTROMECHANCIAL ENGINEERING PROFESSIONAL.

2) I WILL NOT COMMISSION (TURN ON) OR INSTALL THE QEG WITHOUT AN ELECTROMECHANICAL ENGINEERING PROFESSIONAL.

3) UNDER THE ABOVE CONDITIONS, I MAY USE THE QEG INSTALLATION INSTRUCTIONS FOR PERSONAL USE, AND UNDERSTAND THE NEED FOR IMPECCABLE COMMITMENT TO THE BETTERMENT OF HUMANITY. IN THE BEST INTEREST OF THE PEOPLE OF PLANET EARTH, I WILL NOT ATTEMPT TO MISUSE OR MONOPOLIZE THE QEG INSTALLATION INSTRUCTIONS IN ANY CAPACITY, NOR WILL I ATTEMPT TO MAKE A HUGE PROFIT AT THE EXPENSE OF ANOTHER HUMAN BEING.

IMPORTANT – Please make certain that persons who are to use this equipment thoroughly read and understand these instructions and any additional instructions prior to construction, installation and operation. In addition, we require you to read this notice again when you are ready to assemble the core.

Letter from the Editor 3-25-2014

Dear Builder,
It is no easy task to build the QEG, and we want to encourage you by offering a short treatise on the importance of ‘consciousness’ in this endeavor. Many are becoming more familiar with Nikola Tesla, his desire for all people to have access to free energy, and his failed attempt to expose this technology to the world. Many have followed him with the same aspirations and, similar to Tesla’s plight, have also been prevented by powers beyond their control. The “free” energy movement is rife with horror stories ranging from government theft of patents, to reputations destroyed, to the murder of untold numbers of brilliant scientists/inventors.

We must all consciously and constantly rise above these tyrannical infringements, and create an environment for ourselves and our neighbors, and rediscover Nature’s Laws to be able to live and thrive. We must leave off responding in incredulity to what was “done to us” when we were told and believed we couldn’t create free energy. We are now awake to the point that we know they were lying. Who’s they? The elite for starters – follow the money (watch THRIVE); J.P. Morgan couldn’t put a meter on an energy plan for the world, and so destroyed any chance of that happening through several vicious attacks on Tesla’s reputation and livelihood – basically got Tesla’s ideas out of the way for his own profit and power – and maliciously destroyed the man (youtube: Tesla’s Autobiography). While the electric companies have told the people theirs is the only way to get electricity, and we are dependent on them, the truth is that we have been deprived of this alternative (quantum) energy source for close to 130 years. Morgan’s grip on the energy supply has not loosened one bit in all that time, in fact, you are probably paying more for electricity than ever before, all things being equal.

So how do we effect our future now, and free ourselves and the generations to come after us completely from energy tyranny? Building the QEG is one way. It is a journey that requires you think deeply about processes that will, in turn, expand your senses to enable you to receive information from the quantum field of consciousness, or God if you prefer this reference. We believe we had Divine inspiration and help, which began with a burning desire to “get off the grid” and also do something significant for humanity. The timing is wonderful because, as of the writing of this, the entire planet is in turmoil as never before, and people will need to not only be self-sufficient, we will all need to live according to what is good for all (Ubuntu), and help each other for our continued evolution as species/planet.

It is with great love that the QEG is offered to the world, and as you take on the task of building one, it is hoped that this becomes your path also: the mission of free energy for all! The next step we ask you to take on your journey, before and during building, is to listen to the discussion here: http://www.youtube.com/watch?v=3FqzTW7qh2U&feature=youtu.be with HopeGirl, Ralph and Marsha Ring, Fernando Vossa and the 3D Global Network.

I am, in service to Gaia and its inhabitants, deeply grateful for this technology, and the opportunity to share it!

Valerie Robitaille

IMPORTANT ADDITIONAL INFORMATION

We are not professional writers or photographers and didn't always have opportunities to document or photograph every step of development. Therefore, please take the level of engineer/electrical experience required to build a QEG very seriously as we are giving these to you under this premise. You will discover the advanced level of knowledge of mechanical/electrical processes needed quickly enough. The correct construction of the QEG requires patience and careful thought. We made several mistakes in development and have given here the steps that were successful. You will probably still make mistakes – and these will be your greatest learning opportunities as you gain more knowledge about this type of energy.

Before beginning to build, consider how much you would like to outsource to one of the cottage industry community units (CICUs) near you! In the US we recommend Polaris for the steel stator/rotor construction, and Torelco for toroidal winding. As FTW continues to roll out the distribution plan, and more connections across the world are made, we think CICUs will be commonplace and hence, QEG parts accessible (many people will be making them!)

When website URLs were available we provided links for the person reading this online. You may certainly use your own sources for materials but it is imperative you do not alter the instructions/parts herein if you are building a QEG. (We know with increased knowledge you will discover many applications for this technology.) When photographs can be shown to help you visualize a process, they are provided. Please remember, we are not professional manual writers. What we offer you here is free of charge and our gift to humanity – but it comes with great responsibility. Learn as much as you can, use discernment and wisdom, share freely, and you will be privileged to know the secrets of energy creation from the quantum field.

We would like to dedicate the success we've experienced to our first teacher, Sir Timothy Thrapp, and WITTS Ministries, without whose guidance none of this would be available so soon. We acknowledge and honor the work WITTS has done for over 200 years bringing technology forward, and hope that you will consider making a donation to the ministry for their great work.

We would also like to thank our greatest teacher and fellow humanitarian, Nikola Tesla. It is our most gratifying honor to present modern plans for a quantum energy generator to the world, based on Tesla's discoveries, especially at a time when we the people are being manipulated and controlled by a corrupt energy economy. Tesla wanted everyone on the planet to have energy. We continue to carry out his vision.

SCHEMATIC

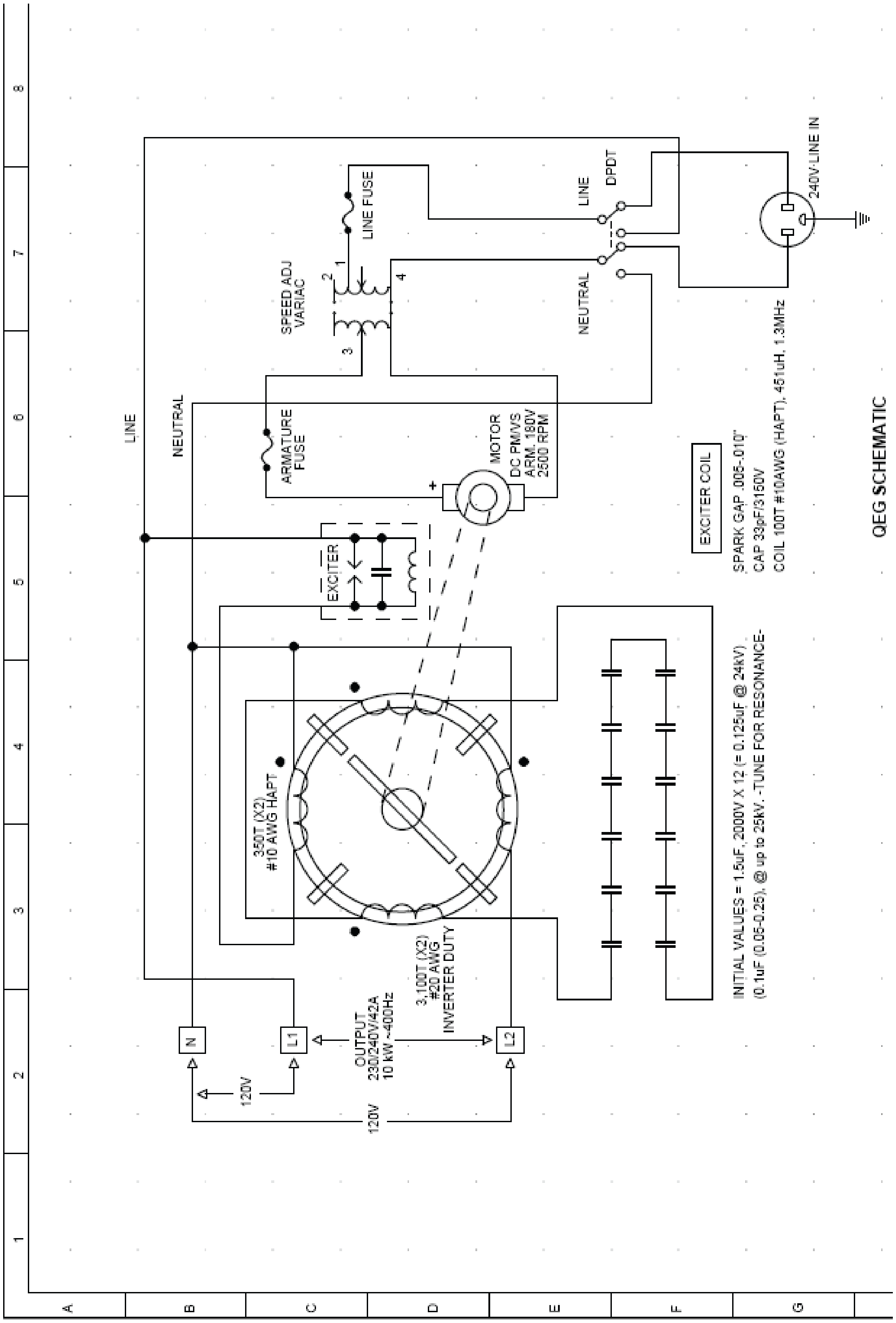
QEG SCHEMATIC
240V LINE IN
DPDT
LINE
NEUTRAL
LINE FUSE
SPEED ADJ
VARIAC
1
2
3
4
LINE
NEUTRAL
ARMATURE
FUSE
+
MOTOR
DC PM/VS
ARM. 180V
2500 RPM
EXCITER
EXCITER COIL
SPARK GAP .005-.010"
CAP 33pF/3150V
COIL 100T #10AWG (HAPT), 451uH, 1.3MHz
350T (X2)
#10 AWG HAPT
3,100T (X2)
#20 AWG
INVERTER DUTY
INITIAL VALUES = 1.5uF, 2000V X 12 (= 0.125uF @ 24kV)
(0.1uF (0.05-0.25), @ up to 25kV. -TUNE FOR RESONANCE-
N
L1
L2
OUTPUT
230/240V/42A
10 kW ~400Hz
120V
120V
1
2
3
4
5
6
7
8
A
B
C
D
E
F
G

Parts List

Part	Type, Model # or MFG PN	Quantity
Capacitors		
Ceramic Disc Capacitors	15pF 3150volts	2
Film Capacitors	2.5uF 2000V	12
End Plates and Shrouds		
Fiberglass reinforced epoxy laminate (for 2 end plates)	G10/FR4	(1) sheet ½ inch thick by 3 feet by 4 feet
Fiberglass reinforced epoxy laminate (shrouds)	G10/FR4	(2) 1/8” x 5.875” diameter
Exciter Coil		
Clear Cast Acrylic Tube	4-3/4” OD x 4-1/2” ID, 1’ Length	1
#10 Magnet wire (see generator core)		
V-Belts and Pulleys		
Goodyear 4L430 V-Belt	GDYR_4L430	1
1 Groove, 3” x 7/8” bore, type A Pulley (Motor)	AK30 x 7/8	1
1 Groove 2.50” Pulley 7/8” bore type A (Generator)	AK25X7/8	1
Drive Motor	DC PM Variable Speed, 2500 RPM, 180V armature, 7/8” shaft, with base	1
Generator Core		
Spacer Blocks 1-1/2” x 1-1/2” x 4-1/2”	Aluminum 6061-T6	16
Mica Tape 1.00” x 50YD	MICA77956X1X50	1
Bonder for shaft/rotor	LOCTITE 648	1
7387 Activator (use with bonder)		1.75oz
Mica Plate NEMA 6	36” x 36” x .030	1
7/8” Three Bolt Flange Bearings	SATRD205-14G	2
Magnet Wire #10 Round HPT or HAPT		~620’
Magnet Wire #20 Round Pulse Shield	HTAIHSD 6” SPL/060-Heavy MW35, 73, 36	~5200’

Teflon tubing	TFT20019 NA005 (Alpha Wire)	8 pieces (12" each)
Fiberglass sleeving w/PVC for #10 HAPT wire (tubing)	PF1308	8 pieces (12" each)
Tape, white, 1"fiberglass, hi-temp (outer wrap)	RG48 (Intertape)	2 rolls
Tape, black, 1" reinforced, high-strength	60020719 (Von Roll)	2 rolls
Nomex Corner Insulation	Torelco	16
Shafting 7/8" dia. x 11.0" long w/standard 3/16" x 3/32" keyway	C1045 TGP Trukey	7/8" dia. x 11"
8 inch Bolts, ¼ - 28 thread, Grade 8	1050095555 (Instock Fasteners)	8
Electrical Terminals		
Assorted ring, spade, and quick connect terminals		
Additional Parts		
Variac, 120/240V in, 0-280V out, 9.5 Amps	Type 1520 (STACO)	1
Console Box w/panel	1456FG4BKBU (Hammond Mfg.)	1
Plexiglass sheet for mounting 2.5uF capacitors	¼ inch thick by 12 inches sq.	1
Electrical box	4" x 4"	1
50 amp plug		1
50 amp receptacle		1
Switch, start/run	DPDT center off, 15 amp, 240V	1
Bridge Rectifier	600 volt, 25 Amp., quick connect terminals	1
Nuts	¼ - 28 Grade 8	8
Washers	¼ " flat	16
Frame and Base		
Angle aluminum	1 ½" x 1 ½" x 4 feet. 1/8" thk.	1

Suppliers and Parts/Service List

POLARIS LASER LAMINATIONS – Generator Core; Rotor
TORELCO – Toroidal winding service and complete core processing ready to ship
FASTENAL – Retaining (bonding) compound (Loctite 648: bonds rotor to shaft) with activator
EIS – Mica tape; 20 gauge Magnet wire
S & W – 10 gauge Magnet wire
INDUSTRIAL SENSORS AND CONTROLS (ebay) – Motor controller KBIC-240D variable speed DC motor controller; Resistor
MOUSER – Capacitors, Enclosures, Variac, Rectifiers, Start/Run switch
JDS (ebay) – V belt; pulleys
EMCO PLASTICS – End plates
ASHEVILLE-SCHOONMAKER MICA – Mica plates
DISCOUNT STEEL – Aluminum squares (spacers)
BRIGHTON BEST – 8 in. bolts
MCMASTER-CARR – Clear acrylic tube for exciter coil
LAKE CITY ELECTRIC (ebay) – Variable speed DC Motor (1 hp)
THE BIG BEARING STORE – 7/8” Three Bolt Flange Bearing w/set screws

Additional Parts

High-temperature fiberglass tape (outer wrap)
4” x 4”electrical box
50 amp plug
50 amp receptacle
Black fiberglass sleeving (pvc-coated)
Black mylar insulating tape
Wood or welded steel tubular frame for base
12 - TPC Thomson / AVX Medium Power Film Capacitors
Angle aluminum
Start/run switch

MAJOR GENERATOR COMPONENTS

- Stator
- Rotor
- Bearings
- V belts
- Capacitors
- External exciter coil
- End panels/plates
- Magnet wire
- Drive motor
- Control box
- Frame and packaging
- Variac
- Inverter
- Pulleys

THE STATOR, or generator core, is made using 140 laminations of 24 gauge M19 C5 electrical steel forming a stack of 3 - ½ inches, with a 4 pole configuration. Corresponding **ROTOR** with 2 poles. Both STATOR and ROTOR are tig welded in 4 places.

GENERATOR STATOR AND ROTOR

DRAWING # 7410-1

USE MATERIAL : 24GA / M19C5 (0.025")

STACK AND TIG WELD STATOR AND ROTOR TO : 3.500"+/-1 LAMINATION LENGTH

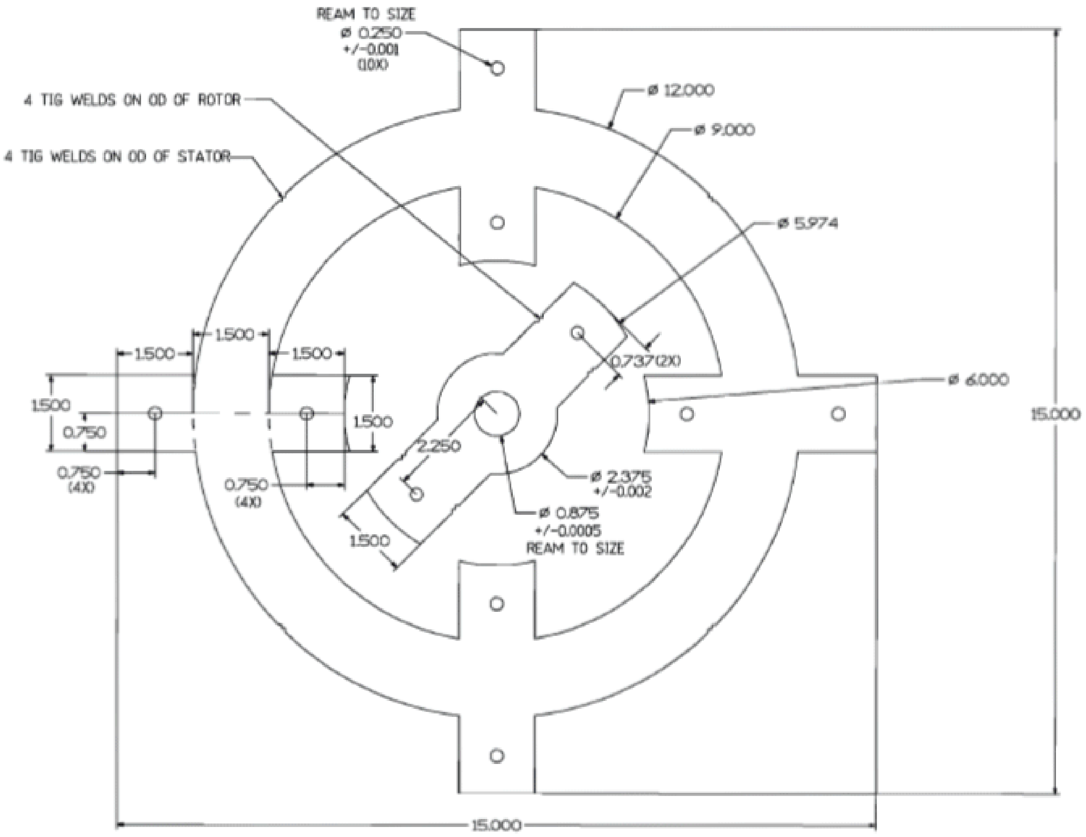
REAM TO SIZE
ø 0.250
+/-0.001
(10X)
4 TIG WELDS ON OD OF ROTOR
4 TIG WELDS ON OD OF STATOR
ø 12.000
ø 9.000
ø 5.974
0.737(2X)
ø 6.000
15.000
1.500
1.500
1.500
1.500
0.750
1.500
0.750
(4X)
0.750
(4X)
2.250
1.500
ø 2.375
+/-0.002
ø 0.875
+/-0.0005
REAM TO SIZE
15.000

Shaft

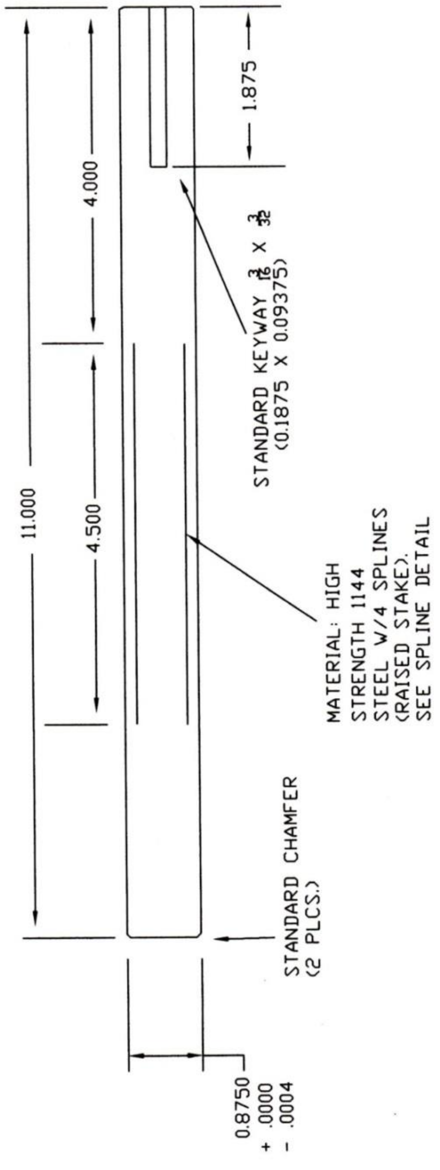
11.000
4.000
4.500
1.875
STANDARD KEYWAY 3/16 X 3/32
(0.1875 X 0.09375)
MATERIAL: HIGH
STRENGTH 1144
STEEL W/4 SPLINES
(RAISED STAKE).
SEE SPLINE DETAIL
STANDARD CHAMFER
(2 PLCS.)
0.8750
+ .0000
- .0004

Shaft cont'd*

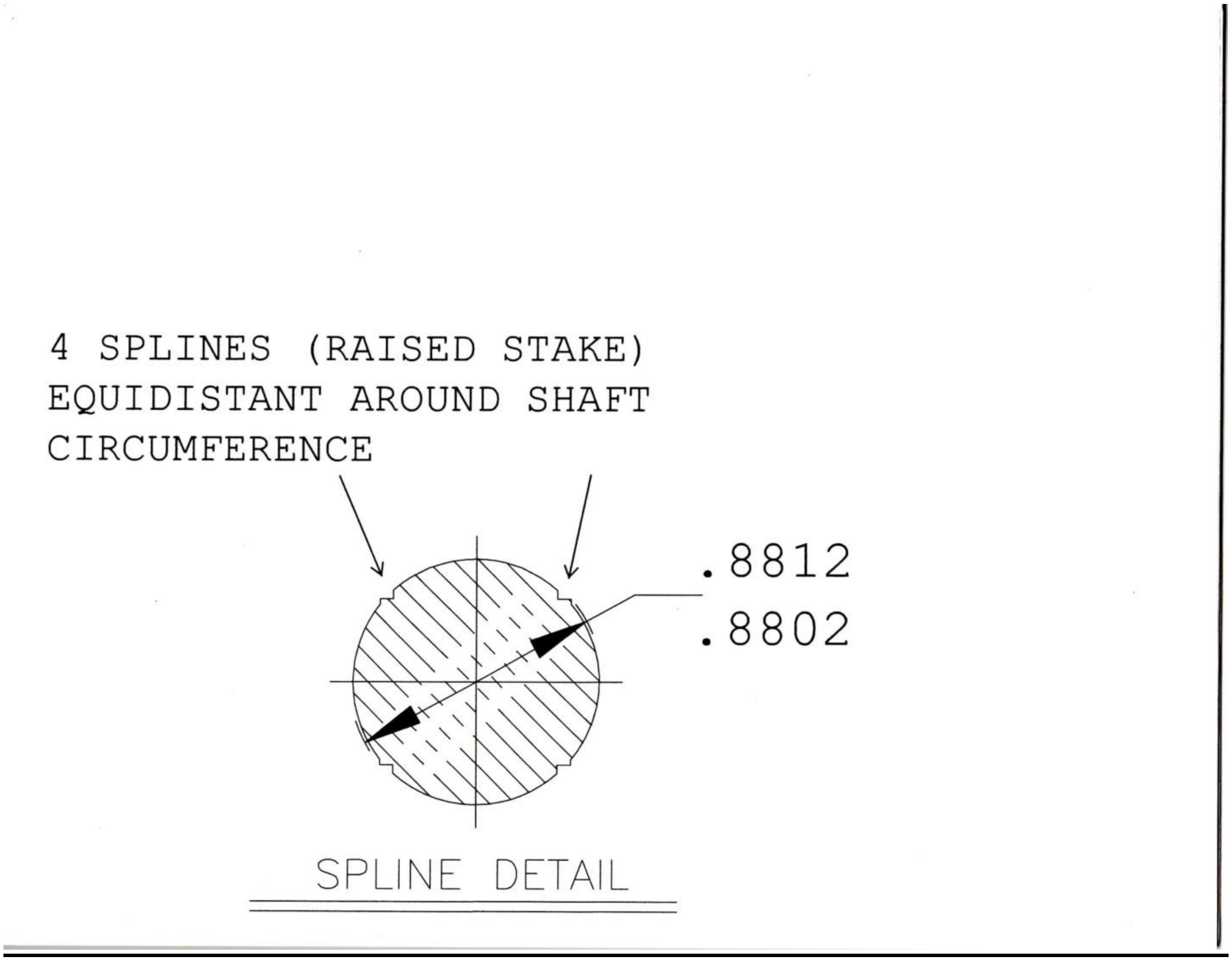
4 SPLINES (RAISED STAKE)
EQUIDISTANT AROUND SHAFT
CIRCUMFERENCE
.8812
.8802

SPLINE DETAIL

*** We didn't use the splines, rather, we used Loctite 648 retaining compound to bond the rotor to the shaft. This technique works very well with a close slip fit between the parts.**

End Plates

Fiberglass reinforced epoxy laminate (FR-4/G10) is used for end plate construction. End plates must be constructed of insulating material, but must also be structurally strong as they support all generator components, including bearings, shaft, rotor and stator. FR4 is the same material used to make circuit boards and is very strong, machinable, and dimensionally stable. Dimensions: End Plates: .500” Thk. G10/FR-4 15” X 16.5” with 15” radius and 2.450” center hole.

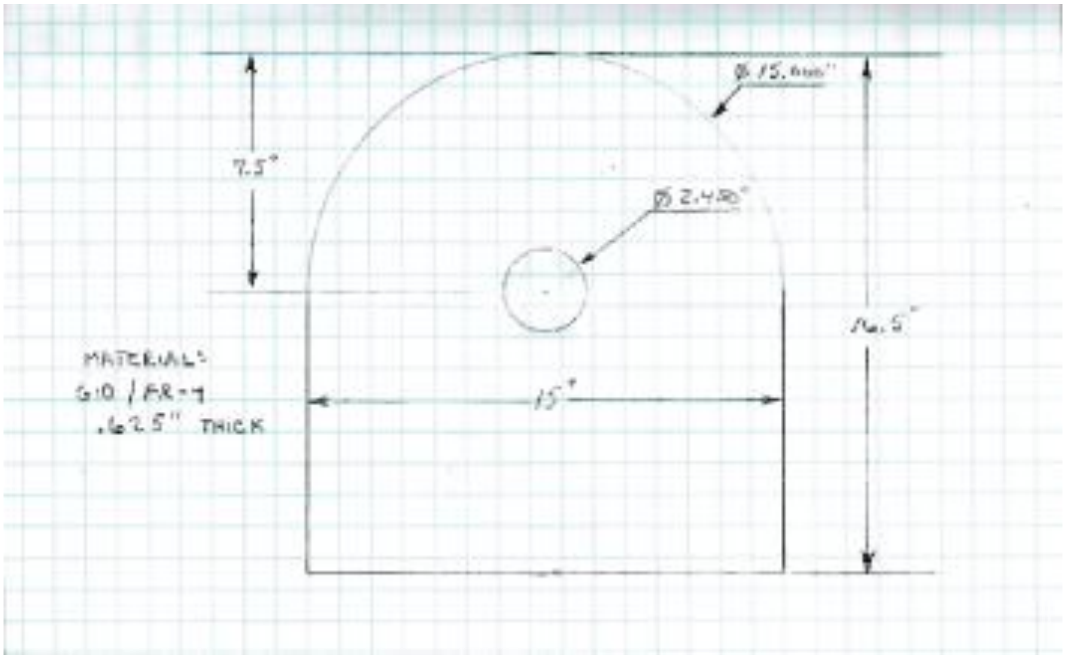

Bearings

The bearings should have a narrow inner ring with set screws for attaching to the shaft. Housing is cast iron with a grease zerk for re-lubing the bearing. We used a 3-bolt flange type mount, but 2-hole or 4-hole can also be used. Bearings are mounted on the *inside* of the end plates toward the rotor.

Capacitors

The capacitors are a critical part of the system. The initial configuration on our prototype uses 12 caps, 2.5uF (microfarad) each. Each cap is rated for 2000V. These capacitors are wired in series in order to be able to withstand up to 25,000V in the primary circuit. The value and quantity of these capacitors will be adjusted to tune the frequency of the generator.

Variac

The variac is used to control the drive motor speed which effectively controls the system power. It’s used during construction/tuning, and prior to self-running set-up when it can be replaced with the smaller, lighter electronic motor drive circuit board.

Drive motor control circuit board

The drive motor control circuit board is an industry standard SCR type DC motor drive manufactured by KB Electronics. The control board can be mounted in the console box we have included in the parts list, and is provided with a speed control potentiometer for adjusting motor speed.

End Plate Layout

We used the bare core as a template to drill all the core mounting holes in the proper locations on the end plates. After end plates are cut and finished, place one on a flat work surface that will support 100 lbs. Place the bare core over end plate, aligning the center bore of the core with the center hole in the end plate. Make sure the pole pieces are right to the edge of the radius at the top of the end plate. We used an extra long drill bit to drill the 8 mounting holes. Repeat this process for the other end plate. Alternately, a long pin could be used as a center punch to mark hole locations and drill the holes using a drill press. If using the core as a template be sure to make assembly marks on the core and the end plate so that final assembly will have all the parts in the same orientation and the mounting bolts will go through without binding. Be sure to mark the in-facing and out-facing sides of each panel.

CORE ASSEMBLY

This is the time to review the NOTICE and section on consciousness at the beginning.

We highly recommend ordering your generator set (stator and rotor) from an experienced professional lamination house. When your stator/rotor is welded and drilled, you are ready to bolt down the 8 spacers and wrap the core with 2 types of tape: Wrap 2 layers mica tape around the steel core (round part) followed by 1 layer of 1” reinforced, high-strength black tape. These 3 layers will bring the thickness needed for insulation to 17 mil (be very mindful at the corners of the pole pieces making certain there is no opening in the insulation for the wire to fall down into contact with the steel. If this happens, the coil will be short-circuited).

Mica Plate Cuts (16 pieces)

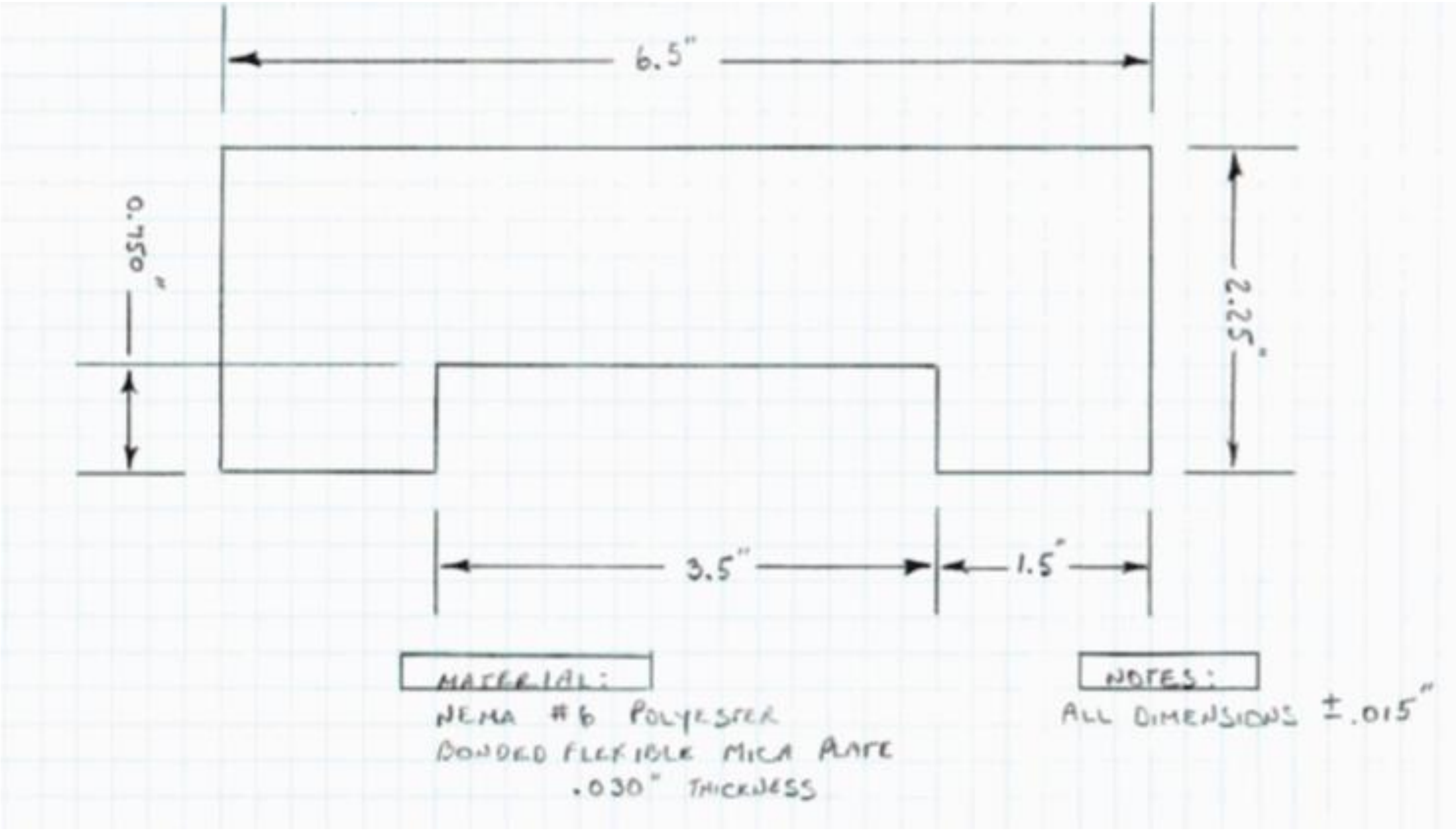
6.5"
0.750"
2.25"
3.5"
1.5"
MATERIAL:
NEMA #6 POLYESTER
BONDED FLEXIBLE MICA PLATE
.030" THICKNESS
NOTES:
ALL DIMENSIONS ±.015"

Installing Mica Plates

After you’ve cut 16 c-shaped mica plates, install them on the top and bottom of each pole piece (front and back). We used a small amount of contact cement to hold them in place for the rest of the processing, but they can also be taped in place with the reinforced black mylar tape. These are installed after core taping and before winding.

Wiring

You will need to commission a toroidal winding service. They might agree to process the entire core if you supply the materials (mica tapes and plates, corner insulation, aluminum spacers, bolts, outer taping, etc.). Proper winding is critical.

Teflon sleeving is installed on the first complete turn of each winding of the #20 wire, and fiberglass/PVC sleeving on the #10 wire. Two coils of 3100 turns each of #20 wire are wound o’n opposing sides (left and right), and 2 coils of 350 turns each of #10 wire on the other sides (top and bottom). Leave about 3 feet of wire at the start of each winding and also at the finish for lead wires. Use enough sleeving to make sure the lead wires are completely insulated where they come through the back end panel. Be sure to secure the finish leads of each coil so that they don’t unravel during handling.

Outer Wrap Taping

Wrap a single layer of 1” white fiberglass tape tightly and securely around each coil making sure that all wire is covered and tape is butted up against the 4 pole pieces.

Generator Assembly Steps

Rotor/Shaft/Shroud Assembly

Drawings are provided for the shaft showing an optional spline operation that can be used to mount the rotor to the shaft, if desired. We used Loctite 648 industrial adhesive (with activator), which is effective with close fitting parts.

Drill a 7/8” center hole, and two ¼” mounting holes into the shroud disks (mounting holes are lined up with the holes in the rotor). Slide one disk onto the shaft on each side of the rotor. Bolt both shrouds to the rotor using two 4” or 4-1/4” long ¼ - 28 through-bolts and nuts. These bolts should not be any longer than necessary or a rotor imbalance can occur. Shrouds are used to quiet the windage noise generated by the spinning rotor.

Bearings

Mount bearings to the *inside* of the front and rear end plates. Center each bearing on the 2.450” hole in the center of the plate. Drill the holes oversize for the mounting bolts. This is done to provide adjustability in the position of the shaft at final assembly. The bearings will have to be moved slightly to center the rotor in the bore of the generator. The gap between rotor and stator is very small (.010”) and the rotor will need to be positioned so it does not rub on the stator bore. Only tighten finger tight at this time.

We opted to bring the leads from the coils out directly through holes drilled in the rear end plate. You may decide to bring the leads out a different way. Here are the steps for our method:

1) Insert all 8 bolts into the rear end plate, then lay the plate down on a flat work surface with the bolts pointing up. The work surface should have a hole under the center hole in the end plate to provide clearance when the rotor is inserted. About 1 ½ inches of clearance is needed below the plate.

2) With an assistant or two, place the fully processed core (about 90 lbs.) down onto the bolts. Slide the core all the way down into contact with the end plate.

3) Insert the short end of the rotor/shaft/shroud assembly through the stator bore and into the rear bearing. Let the rotor assembly drop gently to the bottom and place the front end plate with bearing over the bolts and shaft end. Tap into place with rubber mallet if necessary. Once plate is in contact with stator assembly, install washers and nuts and tighten securely.

4) With assistance, place the assembly upright onto the raised portion of the base. We used 5 lag bolts across the bottom of the end plates on each side to mount the assembly to the wood base/frame. Other methods could be employed.

5) Mount the drive motor to the base/frame. We mounted the shaft end of the motor onto the aluminum angle on the front of the base with one bolt to provide adjustability for belt tightening. We built a sliding mount for the rear of the motor using sheet metal parts, but sliding motor bases are commercially available that provide adjustability for proper belt tensioning.

6) When the motor is mounted to the base, install 3” pulley on motor shaft using set screws.

7) At this point the rotor position should be adjusted so that it spins freely inside the core without rubbing. This is where you may need to adjust the bearing positions repeatedly until the rotor spins freely. (The gap between the rotor and stator is .010” making this step a little delicate.) however, once the rotor is tightened in position it does not tend to move. Place the 2 ½” pulley on the generator shaft at this time; it can be used to turn the rotor by hand while adjusting its position.

8) Place the V-belt over both pulleys and position pulleys as close to the motor and the generator as possible. Both pulleys should be positioned an equal distance from the faces of the motor and generator to assure that the belt runs true.

9) The variac can be mounted on the base at this time. We used two 1/4 – 20 x 1” bolts with nuts to mount the variac to the aluminum angle. After all the components are mounted on the base, wiring and testing will be performed using the variac. (After set-up and testing is completed, the variac can be replaced with the electronic motor control circuit board for less bulk and weight. The console box in the parts list can be used as an enclosure/control panel for the motor speed control board, and for mounting the DPDT start/run switch.)

10) With all components mounted on the base, wiring can begin. Please follow the schematic to make connections. We mounted a 4” x 4” electrical box on the base to support a large (50 Amp) receptacle to bring the power out of the generator.

Wiring Notes: The generator output can be wired in series (220, 230-240V), or parallel (110, 115, 120V). For the series connection shown on the schematic, the start leads from each coil are connected together. This connection provides the highest voltage output from the windings. If using a parallel connection for lower voltage/higher current, be careful to connect the four leads with polarity opposed (start lead of one coil connected to finish lead of other coil).

The variac we used can be wired for 120 or 240 volt input, and provides 0-280 volts output, at up to 9.5 amps. This is a versatile variac and can be used with either a 120 or 240 volt system. The output of the variac is connected to a 600 volt, 25 Amp full-wave bridge rectifier to power the variable speed DC drive motor.

Set-up and testing

Starting with the wiring setup as shown in the schematic, disconnect the primary coils from the series capacitor string on one end (disconnect capacitors). This will prevent resonance momentarily.

Connect input power to the variac. We started with a full 240 volt series wired system, but parallel 120 volt wiring can also be used.

Test mechanical assembly by spinning up the motor/rotor/belt and observing operation. Adjust variac voltage from zero to about ¾ through its range. The active rpm range is under 2500 rpm, so we don’t need to spin very fast. Assure there is no stack rub (rotor scrubbing on stator), or other mechanical issues that need to be corrected for smooth operation.

When proper mechanical operation is assured, re-connect the series capacitor bank. The initial configuration of 12 (twelve) 2.5 uF, 2000 volt capacitors gives us .208uF, that will withstand up to 24,000 volts. This initial value should be in the range to produce resonance.

As the machine spins up to resonance, the sound will change, and the rotor speed will lock into the resonant frequency. At this point any further increase of the motor speed control will change the speed only slightly, but the additional mechanical power (horsepower) will drive the core deeper into resonance, thereby increasing the power output. With a single control, the voltage and current (power) can be increased or decreased.

As previously mentioned, the exciter coil is used to provide a conduction path through the quantum field (zero point) into the generator core. This has the effect of polarizing the core, which increases power output over time. After the QEG is first built, the spark gap on the exciter coil should be adjusted (with power off) to between .005" and .010". Start the generator and let it spark for 2-3 seconds, and repeat this 4 or 5 times. Do this whenever starting the generator for the first few weeks of operation.

BEARINGS

SHROUDS

CAPACITORS

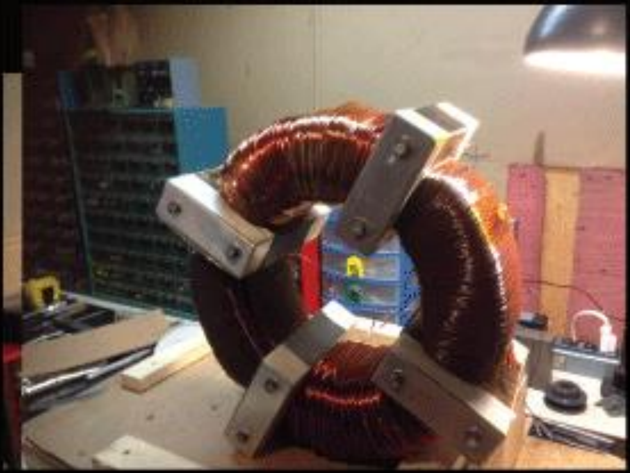

WOUND CORE

ACRYLIC TUBING & PLATE

EXCITER COIL

MICA TAPING

MOTOR

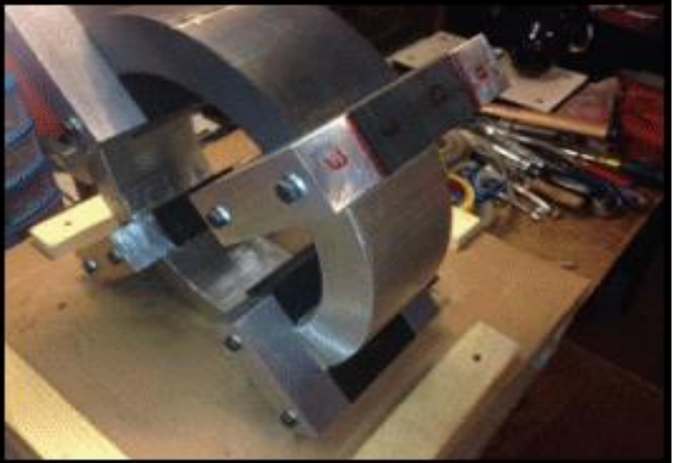

ASSEMBLY MARKS

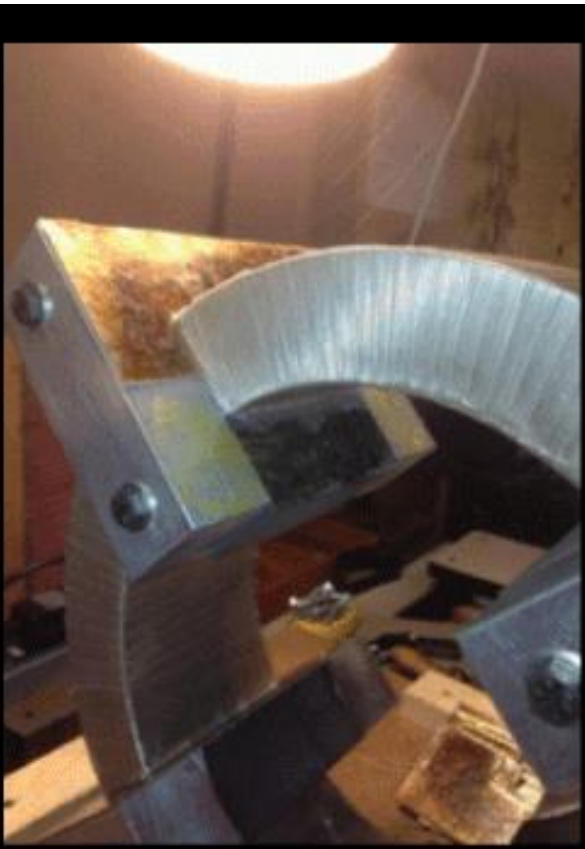

MICA PLATING TOP

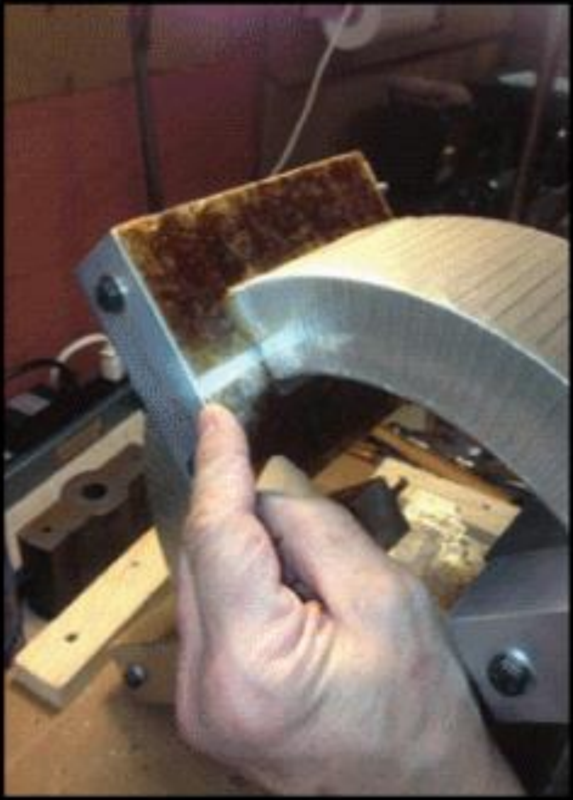

MICA PLATING BOTTOM

FIBERGLASS SLEEVING

MOTOR CONTROL BOX

ASSEMBLY MARKS

ROTOR & SHAFT

SHAFT

CAPACITORS

VARIAC

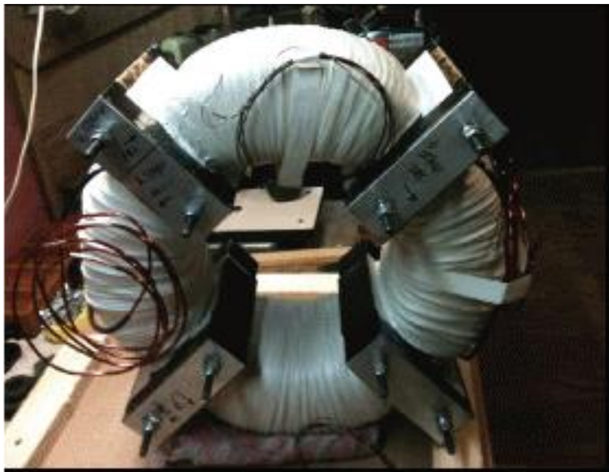

FULLY WRAPPED CORE

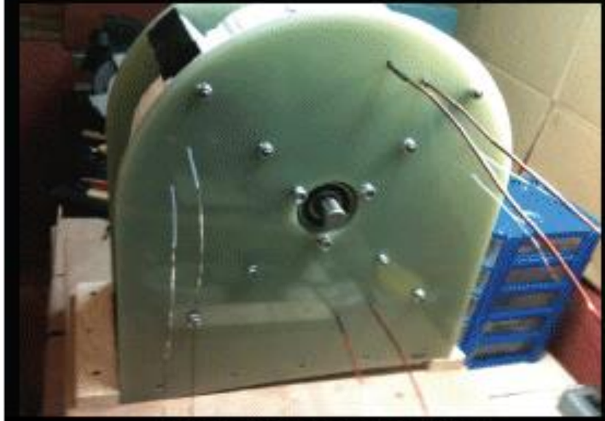

END PANEL REAR

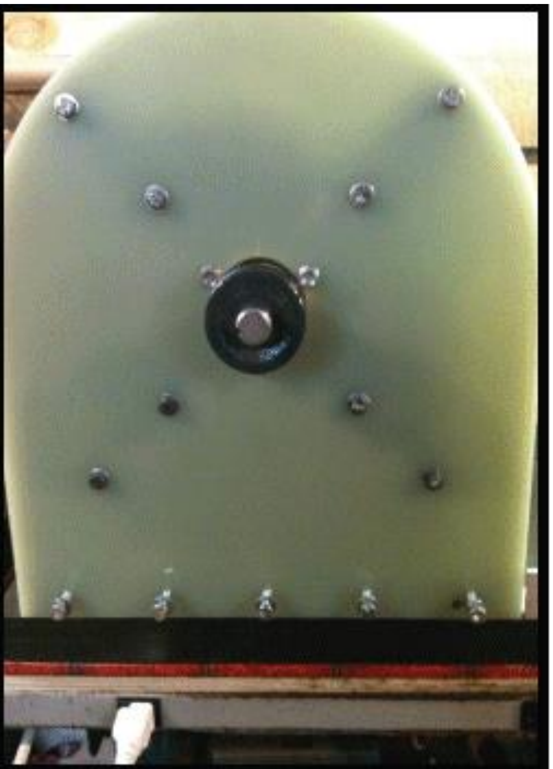

END PANEL FRONT

OUTER WRAP

TOROIDAL WINDER

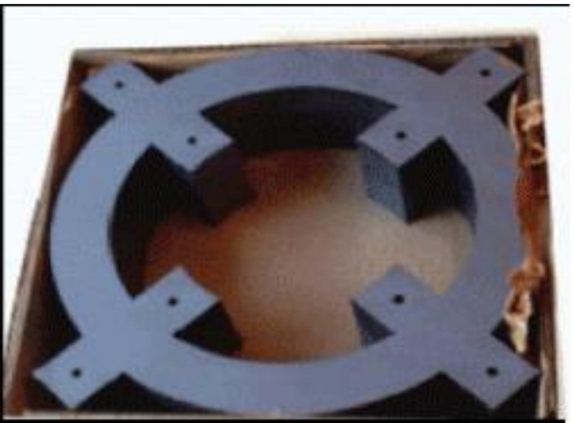

STATOR

ELECTRICAL BOX

END PLATES

WOOD FRAME

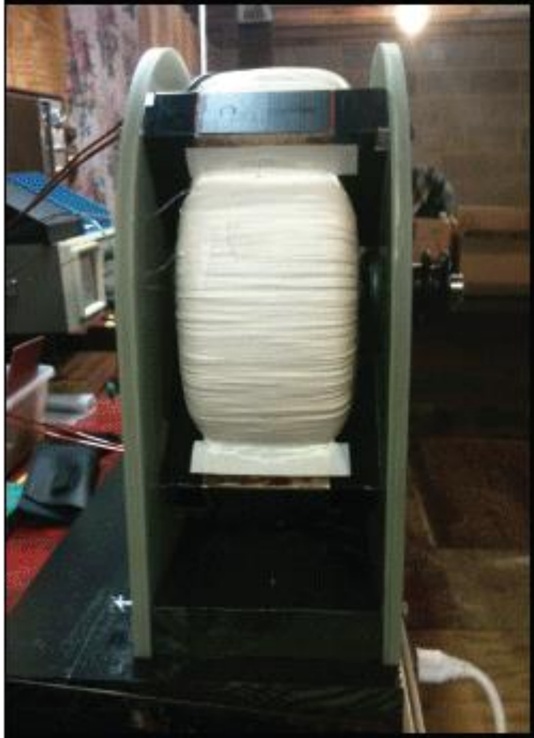

WRAPPED CORE/END PLATES

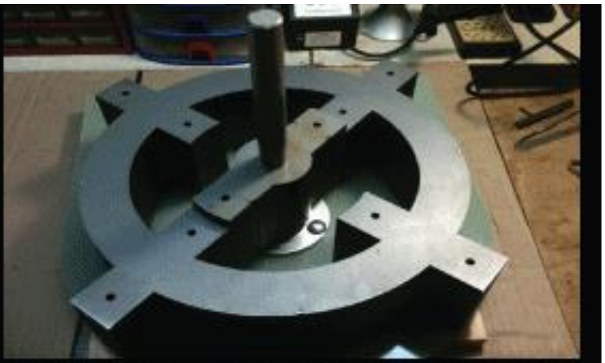

STATOR/ROTOR/SHAFT

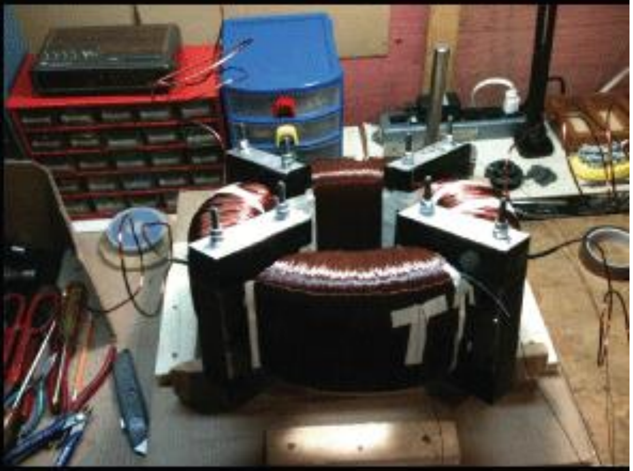

WIRED CORE

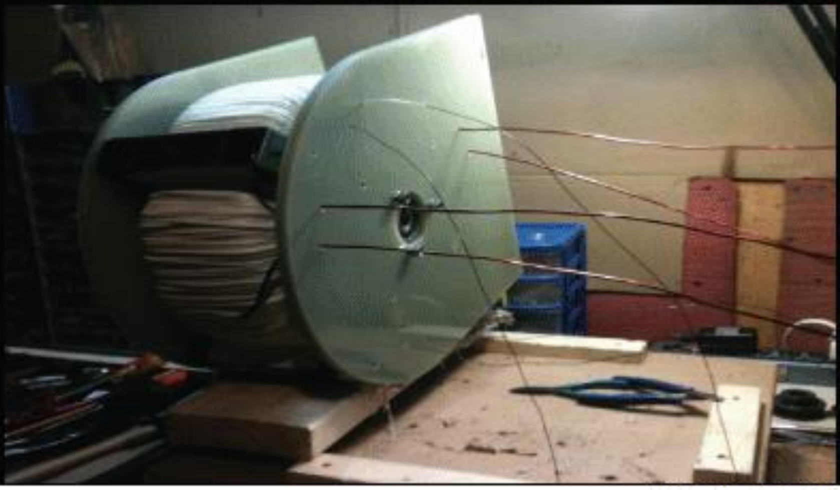

WOOD FRAME

MOTOR CONTROLLER

20 GAUGE MAGNET WIRE

S&W
WIRE
1-800-635-8467
UP

10 GAUGE MAGNET WIRE

SPACER BLOCKS

UNITED STATES PATENT OFFICE.

NIKOLA TESLA, OF NEW YORK, N. Y.

ELECTRIC GENERATOR.

SPECIFICATION forming part of Letters Patent No. 511,916, dated January 2, 1894.

Application filed August 19, 1893. Serial No. 483,562. (No model.)

To all whom it may concern:

Be it known that I, NIKOLA TESLA, a citizen of the United States, residing at New York, in the county and State of New York, have invented certain new and useful Improvements in Electric Generators, of which the following is a specification, reference being had to the drawings accompanying and forming a part of the same.

In an application of even date herewith, Serial No. 483,563, I have shown and described a form of engine invented by me, which, under the influence of an applied force such as the elastic tension of steam or a gas under pressure, yields an oscillation of constant period.

In order that my present invention may be more readily understood I will explain the conditions which are to be observed in order to secure this result.

It is a well known mechanical principle that if a spring possessing a sensible inertia be brought under tension, as by being stretched, and then freed, it will perform vibrations which are isochronous, and as to period, in the main, dependent upon the rigidity of the spring, and its own inertia or that of the system of which it may form an immediate part. This is known to be true in all cases where the force which tends to bring the spring or movable system into a given position is proportionate to the displacement.

In the construction of my engine above referred to I have followed and applied this principle, that is to say, I employ a cylinder and a piston which in any suitable manner I maintain in reciprocation by steam or gas under pressure. To the moving piston or to the cylinder, in case the latter reciprocate and the piston remain stationary, a spring is connected so as to be maintained in vibration thereby, and whatever may be the inertia of the piston or of the moving system and the rigidity of the spring relatively to each other, provided, the practical limits within which the law holds true that the forces which tend to bring the moving system to a given position are proportionate to the displacement, are not exceeded, the impulses of the power impelled piston and the natural vibrations of the spring will always correspond in direction and coincide in time. In the case of the engine referred to, the ports are so arranged that the movement of the piston within the cylinder in either direction ceases when the force tending to impel it and the momentum which it has acquired are counterbalanced by the increasing pressure of the steam or compressed air in that end of the cylinder toward which it is moving, and as in its movement the piston has shut off at a given point, the pressure that impelled it and established the pressure that tends to return it, it is then impelled in the opposite direction, and this action is continued as long as the requisite pressure is applied. The length of the stroke will vary with the pressure, but the rate or period of reciprocation is no more dependent upon the pressure applied to drive the piston, than would be the period of oscillation of a pendulum permanently maintained in vibration, upon the force which periodically impels it, the effect of variations in such force being merely to produce corresponding variations in the length of stroke or amplitude of vibration respectively.

In practice I have found that the best results are secured by the employment of an air spring, that is, a body of confined air or gas which is compressed and rarefied by the movements of the piston, and in order to secure a spring of constant rigidity I prefer to employ a separate chamber or cylinder containing air at the normal atmospheric pressure, although it might be at any other pressure, and in which works a plunger connected with or carried by the piston rod. The main reason why no engine heretofore has been capable of producing results of this nature is that it has been customary to connect with the reciprocating parts a heavy fly-wheel or some equivalent rotary system of relatively very great inertia, or in other cases where no rotary system was employed, as in certain reciprocating engines or tools, no regard has been paid to the obtainment of the conditions essential to the end which I have in view, nor would the pressure of such conditions in said devices appear to result in any special advantage.

Such an engine as I have described affords a means for accomplishing a result heretofore unattained, the continued production of electric currents of constant period, by imparting the movements of the piston to a core or coil in a magnetic field. It should be stated however, that in applying the engine for this purpose certain conditions are encountered which should be taken into consideration in order to satisfactorily secure the desired result. When a conductor is moved in a magnetic field and a current caused to circulate therein, the electro-magnetic reaction between it and the field, might disturb the mechanical oscillation to such an extent as to throw it out of isochronism. This, for instance, might occur when the electro-magnetic reaction is very great in comparison to the power of the engine, and there is a retardation of the current so that the electro-magnetic reaction might have an effect similar to that which would result from a variation of the tension of the spring, but if the circuit of the generator be so adjusted that the phases of the electromotive force and current coincide in time, that is to say, when the current is not retarded, then the generator driven by the engine acts merely as a frictional resistance and will not, as a rule, alter the period of the mechanical vibration, although it may vary its amplitude. This condition may be readily secured by properly proportioning the self induction and capacity of the circuit including the generator. I have, however, observed the further fact in connection with the use of such engines as a means for running a generator, that it is advantageous that the period of the engine and the natural period of electrical vibration of the generator should be the same, as in such case the best conditions for electrical resonance are established and the possibility of disturbing the period of mechanical vibrations is reduced to a minimum. I have found that even if the theoretical conditions necessary for maintaining a constant period in the engine itself are not exactly maintained, still the engine and generator combined will vibrate at a constant period. For example, if instead of using in the engine an independent cylinder and plunger, as an air spring of practically constant rigidity, I cause the piston to impinge upon air cushions at the ends of its own cylinder, although the rigidity of such cushions or springs might be considerably affected and varied by the variations of pressure within the cylinder, still by combining with such an engine a generator which has a period of its own approximately that of the engine, constant vibration may be maintained even through a considerable range of varying pressure, owing to the controlling action of the electro-magnetic system. I have even found that under certain conditions the influence of the electro-magnetic system may be made so great as to entirely control the period of the mechanical vibration within wide limits of varying pressure. This is likely to occur in those instances where the power of the engine while fully capable of maintaining a vibration once started, is not sufficient to change its rate. So, for the sake of illustration, if a pendulum is started in vibration, and a small force applied periodically in the proper direction to maintain it in motion, this force would have no substantial control over the period of the oscillation, unless the inertia of the pendulum be small in comparison to the impelling force, and this would be true no matter through what fraction of the period the force may be applied. In the case under consideration the engine is merely an agent for maintaining the vibration once started, although it will be understood that this does not preclude the performance of useful work which would simply result in a shortening of the stroke. My invention, therefore, involves the combination of a piston free to reciprocate under the influence of steam or a gas under pressure and the movable element of an electric generator which is in direct mechanical connection with the piston, and it is more especially the object of my invention to secure from such combination electric currents of a constant period. In the attainment of this object I have found it preferable to construct the engine so that it of itself controls the period, but as I have stated before, I may so modify the elements of the combination that the electro-magnetic system may exert a partial or even complete control of the period.

In illustration of the manner in which the invention is carried out I now refer to the accompanying drawings.

Figure 1 is a central sectional view of an engine and generator embodying the invention. Fig. 2 is a modification of the same.

Referring to Fig. 1 A is the main cylinder in which works a piston B. Inlet ports C C pass through the sides of the cylinder opening at the middle portion thereof and on opposite sides. Exhaust ports D D extend through the walls of the cylinder and are formed with branches that open into the interior of the cylinder on each side of the inlet ports and on opposite sides of the cylinder. The piston B is formed with two circumferential grooves E F which communicate through openings G in the piston with the cylinder on opposite sides of said piston respectively.

The particular construction of the cylinder, the piston and the ports controlling it may be very much varied, and is not in itself material, except that in the special case now under consideration it is desirable that all the ports, and more especially the exhaust ports should be made very much larger than is usually the case so that no force due to the action of the steam or compressed air will tend to retard or affect the return of the piston in either direction. The piston B is secured to a piston rod H which works in suitable stuffing boxes in the heads of the cylinder A. This rod is prolonged on one side and extends through bearings V in a cylinder I suitably mounted or supported in line with the first, and within which is a disk or plunger J carried by the rod H. The cylinder I is without ports of any kind and is air-tight except as a small leakage may occur through the bearings V, which experience has shown need not be fitted with any very considerable accuracy. The cylinder I is surrounded by a jacket K which leaves an open space or chamber around it. The bearings V in the cylinder I, extend through the jacket K to the outside air and the chamber between the cylinder and jacket is made steam or air-tight as by a suitable packing. The main supply pipe L for steam or compressed air leads into this chamber, and the two pipes that lead to the cylinder A run from the said chamber, oil cups M being conveniently arranged to deliver oil into the said pipes for lubricating the piston. In the particular form of engine shown, the jacket K which contains the cylinder I is provided with a flange N by which it is screwed to the end of the cylinder A. A small chamber O is thus formed which has air vents P in its sides and drip pipes Q leading out from it through which the oil which collects in it is carried off.

To explain now the operation of the engine described, in the position of the parts shown, or when the piston is at the middle point of its stroke, the plunger J is at the center of the cylinder I and the air on both sides of the same is at the normal pressure of the outside atmosphere. If a source of steam or compressed air be then connected to the inlet ports C C of the cylinder A and a movement be imparted to the piston as by a sudden blow, the latter is caused to reciprocate in a manner well understood. The movements of the piston compress and rarefy the air in the cylinder I at opposite ends of the same alternately. A forward stroke compresses the air ahead of the plunger J which acts as a spring to return it. Similarly on the back stroke the air is compressed on the opposite side of the plunger J and tends to drive it forward. The compressions of the air in the cylinder I and the consequent loss of energy due mainly to the imperfect elasticity of the air, give rise to a very considerable amount of heat. This heat I utilize by conducting the steam or compressed air to the engine cylinder through the chamber formed by the jacket surrounding the air-spring cylinder. The heat thus taken up and used to raise the temperature of the steam or air acting upon the piston is availed of to increase the efficiency of the engine. In any given engine of this kind the normal pressure will produce a stroke of determined length, and this will be increased or diminished according to the increase of pressure above or the reduction of pressure below the normal.

In constructing the apparatus proper allowance is made for a variation in the length of stroke by giving to the confining cylinder I of the air spring properly determined dimensions. The greater the pressure upon the piston, the higher the degree of compression of the air-spring, and the consequent counteracting force upon the plunger. The rate or period of reciprocation of the piston, however, is mainly determined as described above by the rigidity of the air spring and the inertia of the moving system, and any period of oscillation within very wide limits may be secured by properly portioning these factors, as by varying the dimensions of the air chamber which is equivalent to varying the rigidity of the spring, or by adjusting the weight of the moving parts. These conditions are all readily determinable, and an engine constructed as herein described may be made to follow the principle of operation above stated and maintain a perfectly uniform period through very wide limits of pressure.

The pressure of the air confined in the cylinder when the plunger I is in its central position will always be practically that of the surrounding atmosphere, for while the cylinder is so constructed as not to permit such sudden escape of air as to sensibly impair or modify the action of the air spring there will still be a slow leakage of air into or out of it around the piston rod according to the pressure therein, so that the pressure of the air on opposite sides of the plunger will always tend to remain at that of the outside atmosphere.

To the piston rod H is secured a conductor or coil of wire D' which by the movements of the piston is oscillated in the magnetic field produced by two magnets B' B' which may be permanent magnets or energized by coils C' C' connected with a source of continuous currents E'. The movement of the coil D' across the lines of force established by the magnets gives rise to alternating currents in the coil. These currents, if the period of mechanical oscillation be constant will be of constant period, and may be utilized for any purpose desired.

In the case under consideration it is assumed as a necessary condition that the inertia of the movable element of the generator and the electro-magnetic reaction which it exerts will not be of such character as to materially disturb the action of the engine.

Fig. 2 is an example of a combination in which the engine is not of itself capable of determining entirely the period of oscillation, but in which the generator contributes to this end. In this figure the engine is the same as in Fig. 1. The exterior air spring is however omitted and the air spaces at the ends of the cylinder A relied on for accomplishing the same purpose. As the pressure in these spaces is liable to variations from variations in the steam or gas used in impelling the piston they might affect the period of oscillation, and the conditions are not as stable and certain as in the case of an engine constructed as in Fig. 1. But if the natural period of vibration of the elastic system be made to approximately accord with the average period of the engine such tendencies to variation are very largely overcome and the engine will preserve its period even through a considerable range of variations of pressure. The generator in this case is composed of a magnetic casing F′ in which a laminated core G′ secured to the piston rod H is caused to vibrate. Surrounding the plunger are two exciting coils C′ C′, and one or more induced coils D′ D′. The coils C′ C′ are connected with a generator of continuous currents E′ and are wound to produce consequent poles in the core G′. Any movement of the latter will therefore shift the lines of force through coils D′ D′ and produce currents therein.

In the circuit of coils D′ is shown a condenser H′. It need only be said that by the use of a proper condenser the self induction of this circuit may be neutralized. Such a circuit will have a certain natural period of vibration, that is to say that when the electricity therein is disturbed in any way an electrical or electro-magnetic vibration of a certain period takes place, and as this depends upon the capacity and self induction, such period may be varied to approximately accord with the period of the engine.

In case the power of the engine be comparatively small, as when the pressure is applied through a very small fraction of the total stroke, the electrical vibration will tend to control the period, and it is clear that if the character of such vibration be not very widely different from the average period of vibration of the engine under ordinary working conditions such control may be entirely adequate to produce the desired results.

Having now described my invention, what I claim is—

1. The combination with the piston or equivalent element of an engine which is free to reciprocate under the action thereon of steam or a gas under pressure, of the moving conductor or element of an electric generator in direct mechanical connection therewith.

2. The combination with the piston or equivalent element of an engine which is free to reciprocate under the action of steam or a gas under pressure, of the moving conductor or element of an electric generator in direct mechanical connection therewith, the engine and generator being adapted by their relative adjustment with respect to period to produce currents of constant period, as set forth.

3. The combination with an engine comprising a piston which is free to reciprocate under the action of steam or a gas under pressure, and an electric generator having inducing and induced elements one of which is capable of oscillation in the field of force, the said movable element being carried by the piston rod of the engine, as set forth.

4. The combination with an engine operated by steam or a gas under pressure and having a constant period of reciprocation, of an electric generator, the moving element of which is carried by the reciprocating part of the engine, the generator and its circuit being so related to the engine with respect to the period of electrical vibration as not to disturb the period of the engine, as set forth.

5. The combination with a cylinder and a piston reciprocated by steam or a gas under pressure of a spring maintained in vibration by the movement of the piston, and an electric generator, the movable conductor or element of which is connected with the piston, these elements being constructed and adapted in the manner set forth for producing a current of constant period.

6. The method of producing electric currents of constant period herein described which consists in imparting the oscillations of an engine to the moving element of an electric generator and regulating the period of mechanical oscillation by an adjustment of the reaction of the electric generator, as herein set forth.

NIKOLA TESLA.

Witnesses:
PARKER W. PAGE,
R. F. GAYLORD.

(No Model.)

2 Sheets—Sheet 2.

N. TESLA.

ELECTRIC GENERATOR.

No. 511,916.

Patented Jan. 2, 1894.

Fig. 2

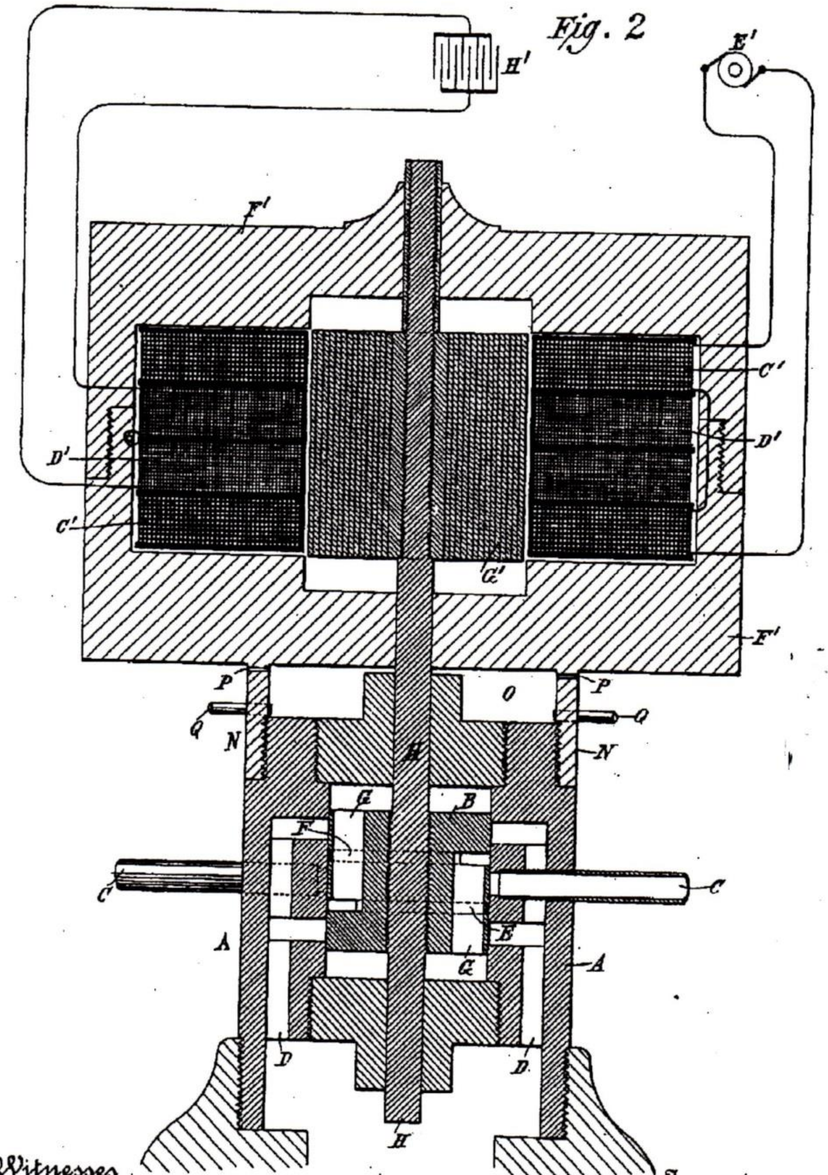
H'
E'
F'
C'
D'
D'
C'
G'
F'
P
P
O
Q
Q
N
N
H
G
B
F
C
C
E
A
A
G
D
D
H

Witnesses

Raphael Netter

R. F. Gaylord

Inventor

Nikola Tesla

By his Attorneys

Duncan & Page.

Frequently Asked Questions

- Where is the energy this device is using coming from ?
  The quantum field

- How much overall power does the QEG produce?
  10 KW (same design scaled up can produce 40 KW)

- If 'free' energy devices work, why aren't the electric companies using them?
  This should be obvious but in case you need a reference, please see:
  http://hopegirl2012.wordpress.com/

- How is the QEG started up if it doesn't require fuel?
  All that's needed is to spin the machine up to resonance. At that point it will run itself. It can be started using existing electrical power if available, or a crank mechanism, or a battery powered motor-start system. A battery start system could also keep its own batteries fully charged, by tapping some power from the generator.

- How long will the QEG run?
  Indefinitely (or until parts wear out)

- How did the QEG improve upon Tesla's work?
  This design adapts the linear, reciprocating element of the Patent, to a rotating element, and some electronics are employed for added stability; controlled amount of power and correct frequency range.

- Does the QEG slow down when more of a load is added?
  No – it's not that type of energy.

- Does the QEG emit radiation?
  No – it's not that type of energy.
- What form of energy does QEG use?
  Electromagnetic and atmospheric charge

rexresearch.com

Radiesthesia Patents

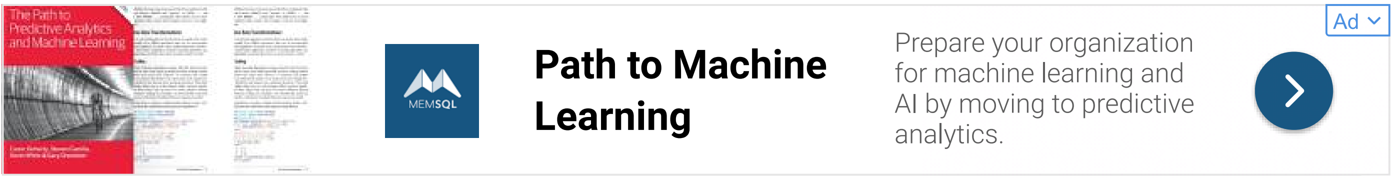
The Path to Predictive Analytics and Machine Learning
MEMSQL
Ad
Path to Machine Learning

Prepare your organization for machine learning and AI by moving to predictive analytics.

wildstock
shutterstock

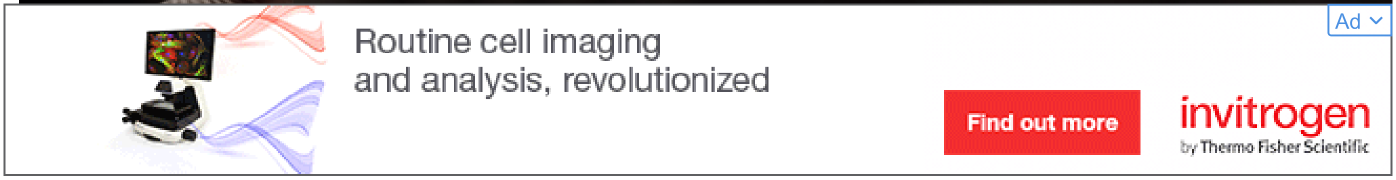

Routine cell imaging and analysis, revolutionized

Find out more

invitrogen by Thermo Fisher Scientific

European Patent Office Advanced Search :
http://ep.espacenet.com/advancedSearch?locale=en_EP

Classification : EC: A61N1/16

WO 9523628
DEVICE FOR FORMING AN ENERGY FIELD

Inventor: TOLSTOI VLADIMIR STEPANOVICH
Also published as: AU7865694

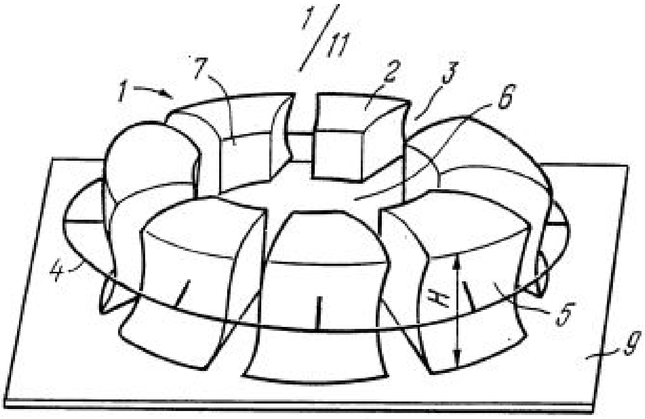

FIG.1

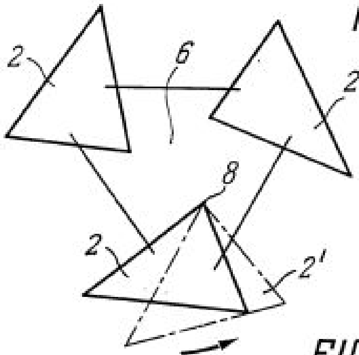

FIG.2

Abstract -- The proposed device for forming an energy field comprises at least one collector (1) made up of a plurality of energy sources (2) arranged along a closed line and separated by gaps, with a central open area (6). The energy sources (2) are interconnected and can move in relation to each other.

DE 3433292
Apparatus for the directed deflection or screening of so-called earth radiation - geopathogenic zones

Inventor: VERZICHT DES ERFINDERS AUF NENNUNG

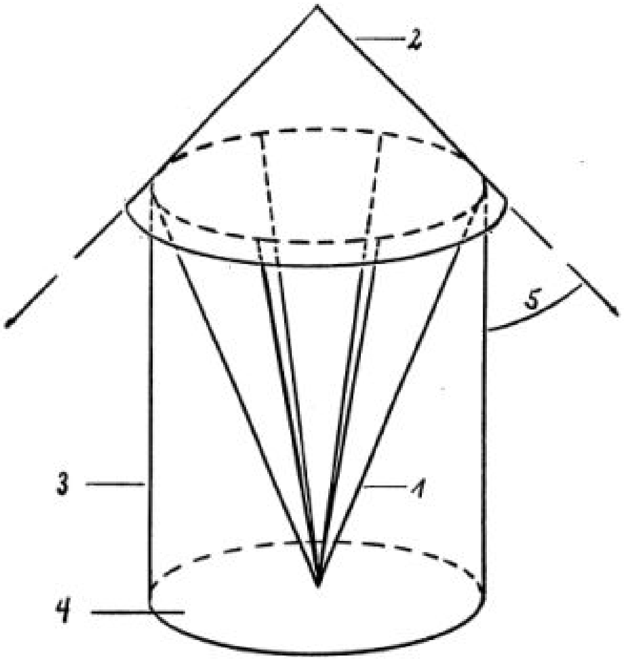

Abstract -- Antennas - bar antennas, V-shaped multiple-bar antennas, spirals, in connection with tubular sleeves which serve as wave guides and are or may be equipped with a ground reflector to increase their effect, emit all around a deflecting radiation or an interference radiation in the upright position in conjunction with a directional reflector mounted at the top, charged by a geopathic wave field. This radiation is capable in its effective range of drifting or neutralising other interfering stimuli/radiation caused by further interference zones (water veins, etc.). In the case of interference radiation/deflection radiation directed downwards due to a corresponding reflector shape there is no rod deflection above the erection site. In the case of interference radiation/deflection radiation directed upwards due to a corresponding reflector shape there is no rod deflection below the erection site. The effective range is influenced by the size of the antenna, the charge thereof (interference zone crossing) and by the shape of the reflector screen.

DE 3525521
Pyramidal magnetic field device for the regulation of energy disturbances in the human body or for the continuous material recovery

Inventor: KAUSCH OTTO

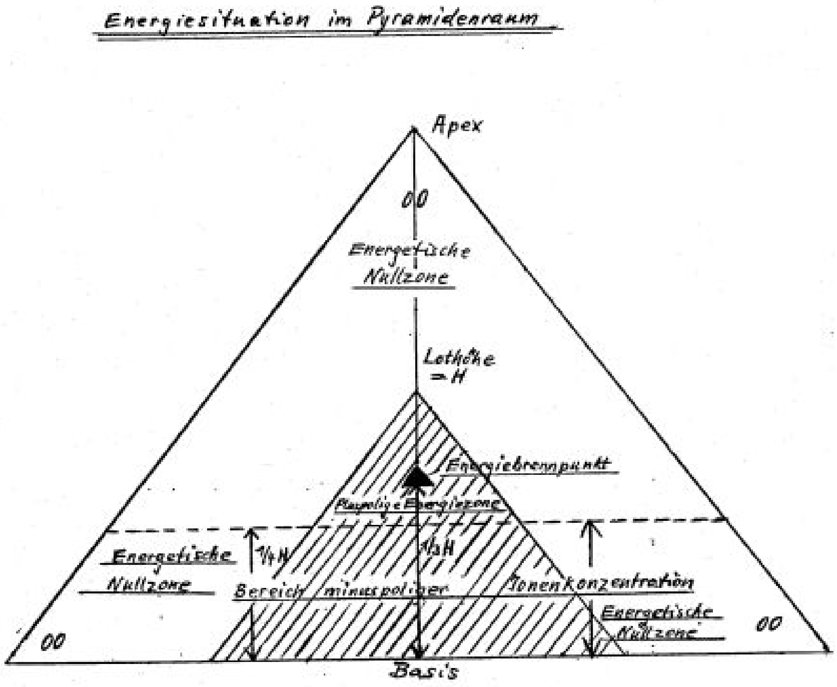
Energiesituation im Pyramidenraum
Apex
00
Energetische Nullzone
Lothöhe = H
Energiebrennpunkt
Pluspolige Energiezone
Energetische Nullzone
1/4 H
1/3 H
Bereich minuspoliger Ionenkonzentration
Energetische Nullzone
00
00
Basis

Abstract -- The device comprises, for example, a pyramid of dielectric material having the dimensions, as true to scale as possible, of the Cheops pyramid and in a size corresponding to the use to which it is put. The collector effect of the pyramid, one edge of which is to be aligned to magnetic north, bundles the positive pole energy of the magnetic earth field in the energy core of the device. As in a mirror image, "reflected" force fields with negative energy result above and below the pyramid. a) By appropriately positioning a person in, above or underneath the pyramid it is possible to supply the bundled energy resulting from the collector effect of the pyramid to the person in order to balance out energy disturbances of the body. b) If the subject-matter of the application is used for the continuous automatic material recovery a pyramid of appropriate size is to be fitted over the metal part (e.g. bearing) in such a way that the part is located within the energy core of the pyramid and the pyramid is aligned to the direction of magnetic north/south.

DE 3823178

Device for collecting and selectively transmitting programmed radiesthetically measurable radiation, in particular, to the human body

Inventor: NEMITZ MANFRIED

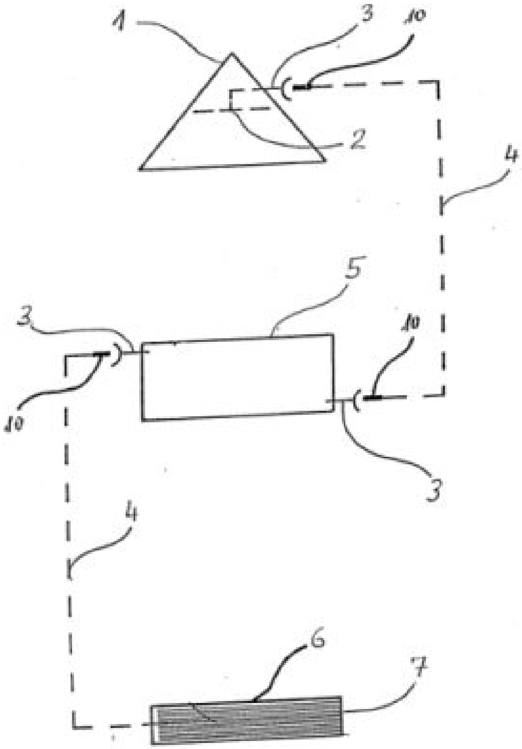

Abstract -- The invention relates to a device for collecting and selectively transmitting programmed radiesthetically measurable radiation, in particular, to the human body. The device comprises three or four functional elements which are connected to one another, namely a pyramid-shaped body as a radiation collector, a vessel serving as radiation transmitter and a tubular radiation injector and, if appropriate, a radiation adapter.

DE 8460071
Device for accruing cosmobiological energy

Inventor: RUDHARDT, ELFRIEDE, 8960 KEMPTEN, DE

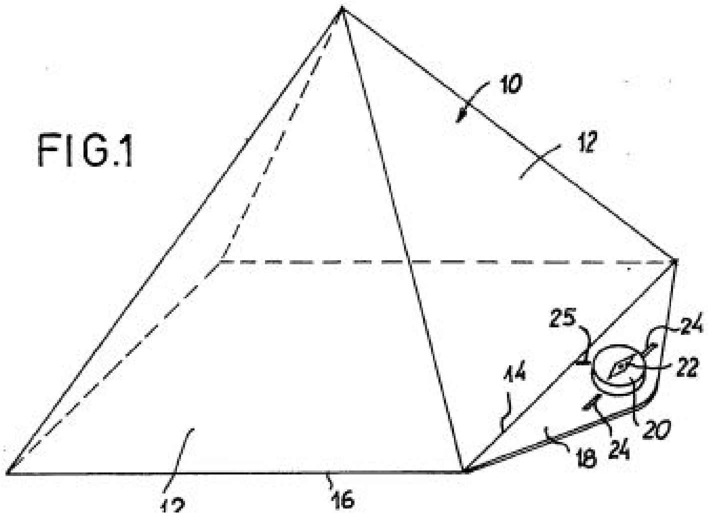
FIG.1

Abstract -- A hollow body (10) in the form of a regular pyramid with a square base has an outer attachment (18) along its one bottom edge (14), on which attachment a magnetic compass (20) is secured, a marking (24; 25), provided on the attachment (18), lying with the compass axis in a plane which is parallel to one pair of bottom edges (14) and at right angles to the other pair of bottom edges. The hollow body (10) can be moved quickly and accurately into a position in which two parallel bottom edges (14; 16) are aligned in the earth's

North-South direction. In this position, the hollow pyramid is able to accrue cosmobiological energy and to supply it to objects situated therein.

EP 0838208
DEVICE FOR TREATING AN ORGANISM WITH ENERGY

Inventor: DENISOV STANISLAV GEORGIEVICH [RU] ; ATAEV DZHAVANSHIR ISMAIL OGLY

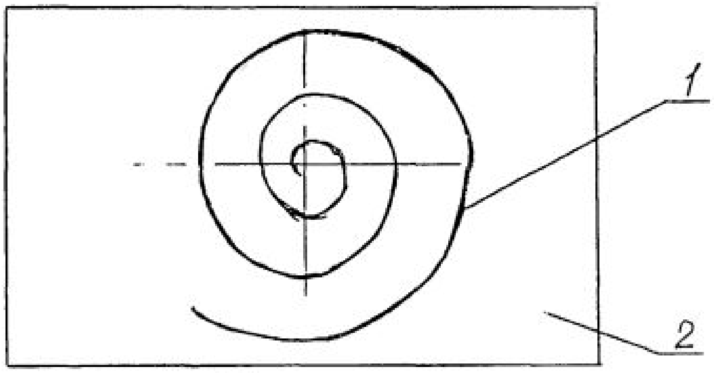

Fig.1

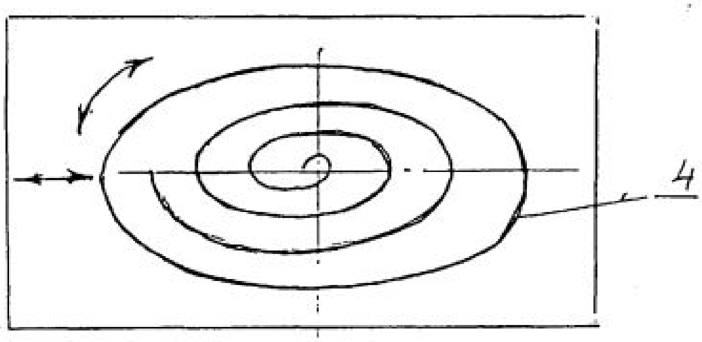

Fig.2

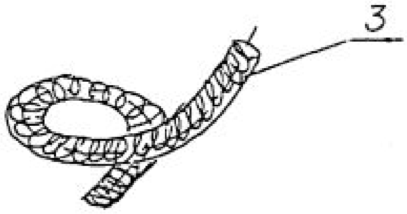

Fig.3

Abstract -- Device by energetics affect to bio-object contains wire which is fixed on a substrate as open-ended space spiraling curve, and wire, is representing by not crossing coils, is located in spiral form, and the device is supplied by the element for regulation of intensity by affect, this element executed as closed electro-conductivity contour, which is located with possibility to change its position concerning space spiral-graphic curve. Space spiral-graphic curve can be ellepes -graphic form, and closed contour - form curve, located with a opportunity of turn and linear moving concerning space spiral-graphic curve. The wire can be executed as electro-conductiving cover, which is make up dielectric material. The device can be supplied having by bio-energetics activity chemical elements such as tellurium, lanthanum, gadolinium, bismuth and etc., which are located in a zone of wire arrangement.; The way of valuation at efficiency device's work by energetic affect to bio-object is concluded to measure conductivity of biologic active points by Foll's method at ten main points on each foot and brush before and after affect, to determine for each points its deviation from given significance, which is accepted 50+2 scale points of the device, thus repeated measurements conductivity make after period of time, equal 1, 24 or 48 hours, and the efficiency by affect is evaluate at size of deviations conductivity is received by results its measurement after effect, do not exceeding 40% from size of deviations conductivity, received by results of its measurement before the affect.

EP 0673185
Device for improving the efficiency of engines

Inventor: BERGMANN HERMANN
EC: H05F7/00

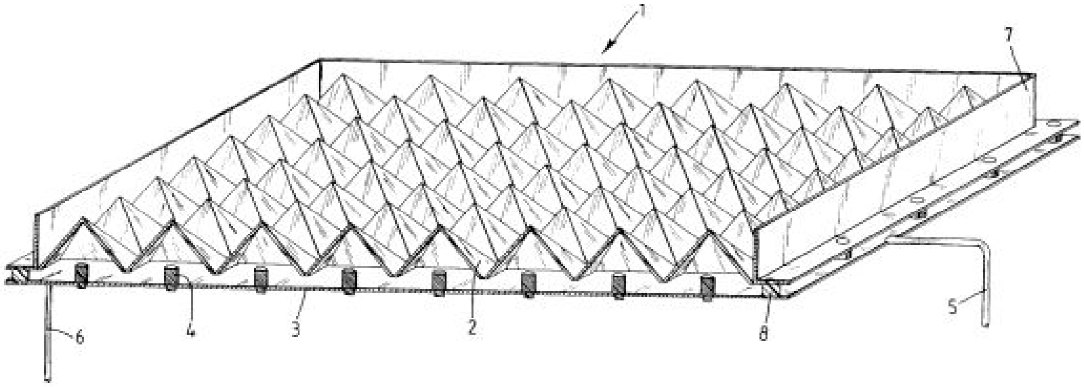

The device has a plate (1) provided with a number of pyramids (2), with a pin (4) positioned centrally under the peak of each pyramid, supported by an electrically-conductive collector plate (3). A pair of electrical leads (5,6) extend between the collector plate and the fuel feed for the i.c. engine, or the engine block, or between the collector plate and the electric motor leads, or the electric motor housing. Pref. the individual pyramids are made of Cu sheet and arranged in a matrix with eight pyramids in each direction. Magnets may be inserted in the leads between the collector plate and the i.c. engine or electric motor.

WO 9420702
DEVICE FOR CONVEYING DAMP OR SALTS

Inventor: MOHORN WILHELM

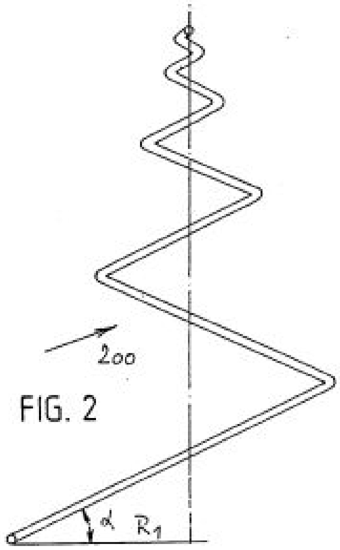

FIG. 2

Abstract -- In a device for conveying damp or salts, e.g., to dry masonry, with at least one electrical conductor wound into a coil (100) and fitted in a housing, the winding diameter of the coil (100) diminishes spirally from one end to the other. Here the distance between the windings of a spiral or wedge-spiral coil (100) and the coil axis in the inward direction is 40 to 60 % smaller on each full revolution than the previous distance. On average, therefore, this gives: for the spiral radius measured at 90 DEG : R'1 = R1-R2/4; at 180 DEG : R"1 = R1-R2/2; at 270 DEG : R"'1 = R1-R2x3/4; at 360 DEG : R2 = R1/2.

FR 2554354
Device sensing telluric and cosmic waves and inhibiting their effects

Inventor: DUPIN JACQUES

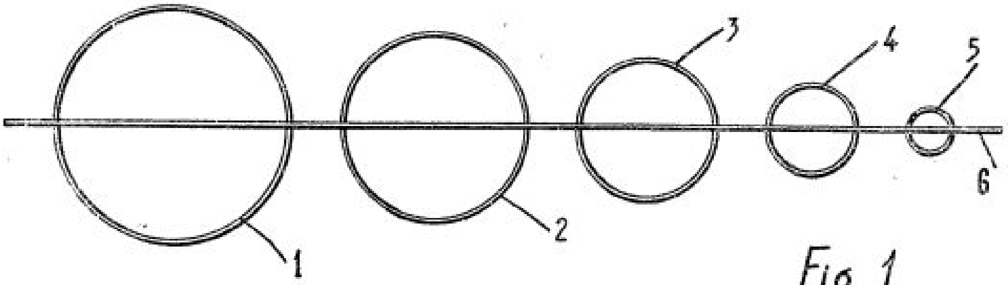

Fig. 1

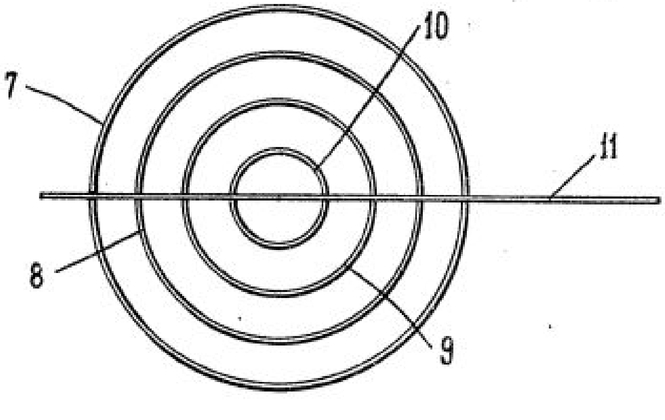

Fig. 2

Abstract -- Device ensuring physiological protection against the harmful effects of telluric and cosmic waves.

The invention is constituted by one or several oscillating circuits formed of rings associated with a bar parallel to their plane and extending on either side, the said device being placed horizontally oriented northwards. It is in this way that the device can comprise the rings 1 to 5 centred, side by side, on a single bar 6 or concentric according to 7 to 10 associated with the bar 11, their effect differing in power according to the disposition adopted. The invention can be used to increase physiological comfort and to improve health.

GB 1077765
Device for affecting the growth or properties of organic or mineral bodies

Inventor: ROBERT ANFOSSO

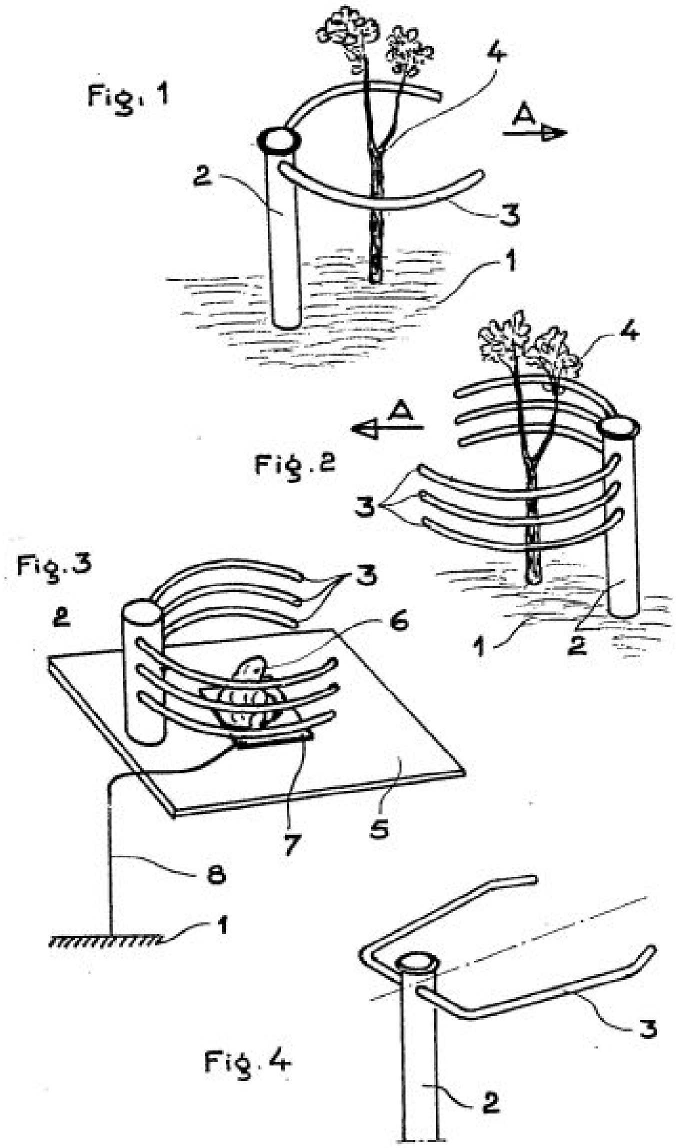

Abstract -- A device said to stimulate plants, or animal or mineral bodies, comprises one or more arcuate or bent member 3 of an electrically conductive material supported on an electrically insulating member 2 of ebonite or a synthetic plastics material. An animal or mineral body being treated may be positioned on an earthed plate. The effect is said to be improved if the member 3 is directed to the East.

GB 766886
A device for screening or neutralising dowsing rays or fields

Inventor: WALTER GISHFORD EDWARDS

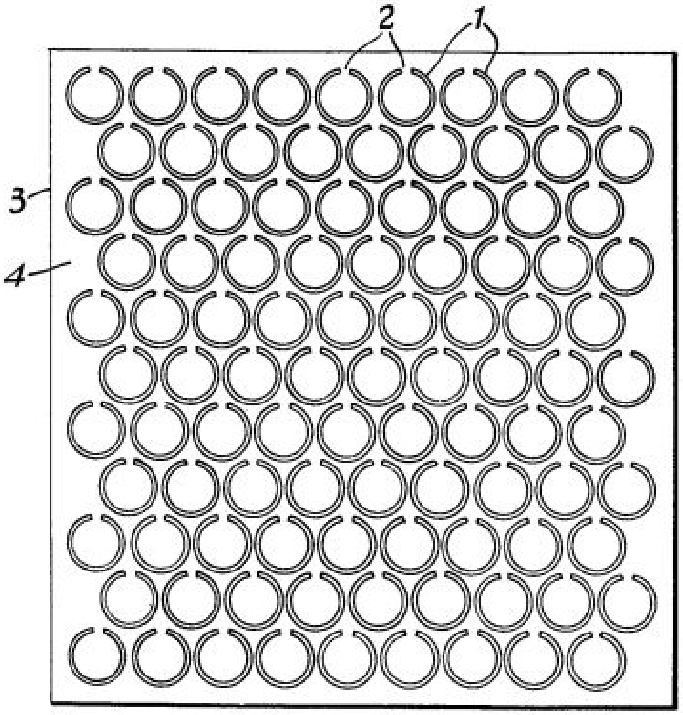

Abstract -- A device for screening the body from dowsing rays or fields comprises a foundation 4 of paper or other suitable material to which is secured rings 1 or loops of metal wire or foil arranged in a comtiion plane so that they do not touch one another. The rings 1 or loops may be made of copper wire or foil and each ring may be broken to form a gap 2.

EP 0098656
Earthing device

Inventor: GIERKINK ALPHONS JOHANNES

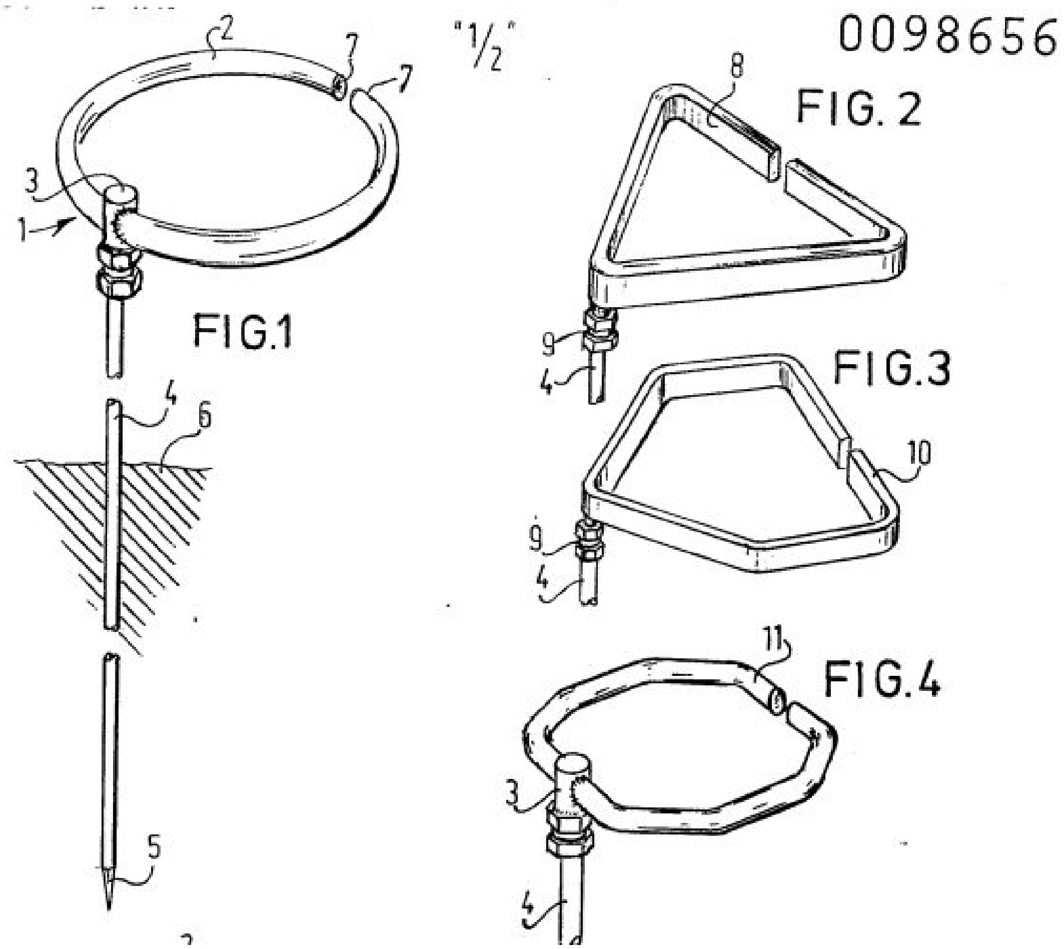

Abstract -- The invention has for its object to counteract the negative influence of biologically active fields on living organisms and provides for this purpose a device for neutralizing such fields, said device comprising an electrically conductive strip or wire having an earth connection at the middle, said strip or wire being located substantially in one plane and being constructed in the form of a ring symmetrical to the middle, whilst the ends of the strip or wire are spaced by a small distance.

WO 0240859
SYSTEM AND METHOD FOR GENERATING A RIGHT TORSION FIELD

Inventor: PAVLENKO ANATOLIY R [UA] ; PAVLENKO OLEXANDER

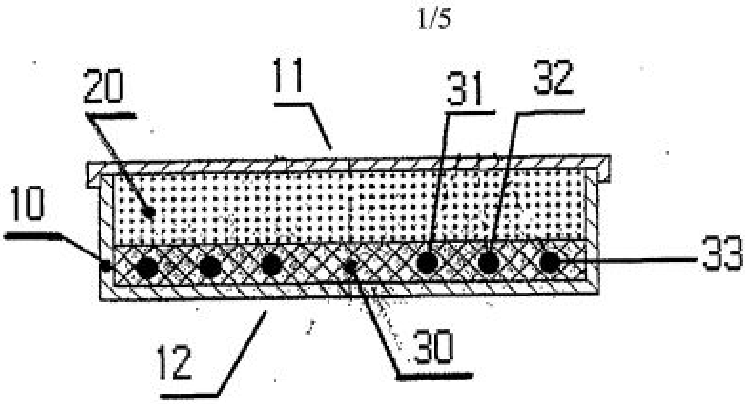

Abstract -- A device for generating a right torsion field having an outer body, a salt solution, and ring. The ring generates a first right torsion field. The ring and the salt solution are disposed within the outer such that a portion of the first right torsion field propagates through the salt solution.; A device for generating a right torsion field (43) having an outer body (10), a salt solution, and a ring (31, 32, 33). The ring (31, 32, 33) generates a first right torsion field (41). The ring (31, 32, 33) and the salt solution are disposed within the outer body (10) such that a portion of the first right torsion field (41) propagates through the salt solution.

FR 1260006
Emetteur de radiations calculées

Inventor: BIENBOIRE GUY-EDOUARD-GABRIEL

US 5449376
System and method for producing highly amplified radio signals for feedback into the human body

Inventor: CALLAHAN PHILIP

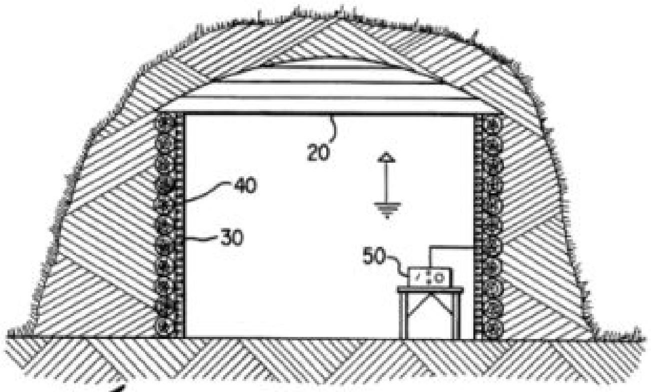

FIG. 1A

Abstract -- The present invention includes an apparatus and method for generating, duplicating, and transmitting biophotonic radio frequencies that propagate along the atmospheric boundary layer of the human skin and the earth's atmosphere. These biophotonic radio frequencies are applied to a human patient for therapeutic purposes. This function is realized with the use of a series of oscillators (i.e., electronic function generators) designed to produce modulated or unmodulated sine and square wave photons. The oscillators are further adapted to emit the modulated or unmodulated sine and square wave photons from a dielectric or metal antenna built as loops into the walls of a diamagnetic-paramagnetic chamber. When a person is seated in the chamber, the photons of energy from the waves generated along the loop antenna penetrate the human body and provide a variety of therapeutic benefits. Among these therapeutic benefits are relief of rheumatic muscular pain, slowing the metabolism, increasing the efficiency of the immune system, eliminating certain psychosomatic ailments, and suppressing cancer.

WO 9325270
DEVICE FOR PROTECTION AGAINST THE BIOLOGICAL EFFECTS OF ELECTROMAGNETIC NON-IONISING RADIATIONS, PARTICULARLY THOSE EMITTED BY VDU EQUIPMENT

Inventor: FILLION-ROBIN MAURICE SERGE

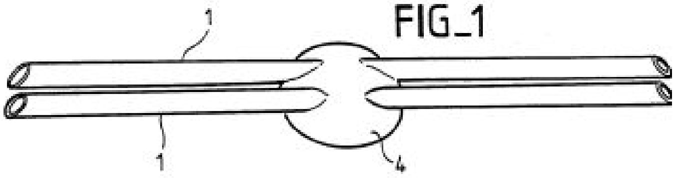

FIG_1

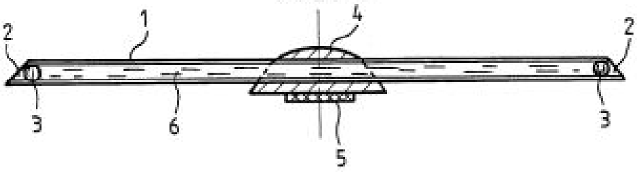

FIG_2

EP 0551668
Resonators eliminating the harmful effects on living beings from electromagnetic radiation

Inventor: SURBECK JACQUES
EC: H05F7/00

ES 2040624
Electromotive power generator for therapeutic uses

Inventor: RAMOS FERNANDEZ FELICISIMO

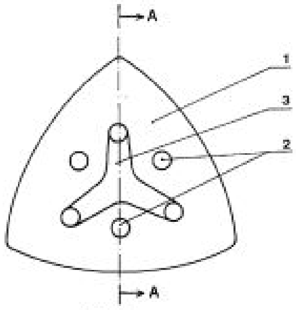

Fig. 1

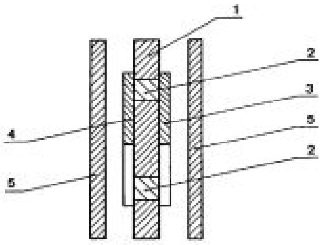

Fig. 2

EP 0302192
Device for the treatment of living beings and plants with pyramid energy

Inventor: OEHME RAINER

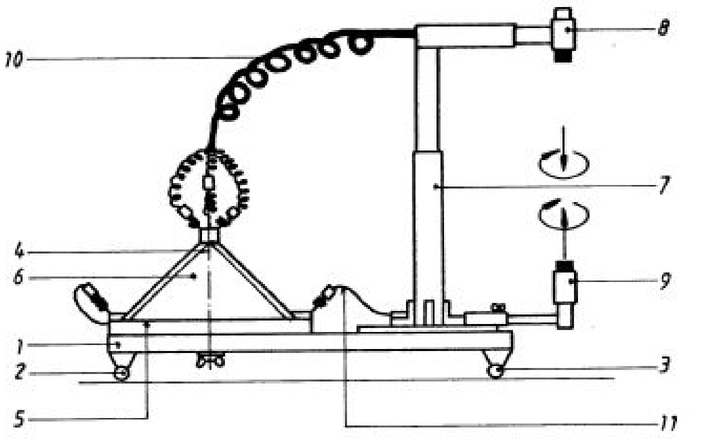

Fig. 1

Abstract -- In a device for the treatment of living beings and plants with pyramidal energy, the energy is picked up on energy pick-up lines (12, 13), which are arranged electrically separate along the four edges of the pyramid (6). Since the emission of energy is particularly strong along the pyramid edges, a particularly large amount of energy can be picked up by energy pick-up lines (12, 13) of this type. Such energy pick-up lines (12, 13) are considerably cheaper to produce.

FR 2618076
Device for emitting beneficial wave forms

Inventor: BOURGOUIN DOMINIQUE [FR] ; DACIER CHRISTOPHE

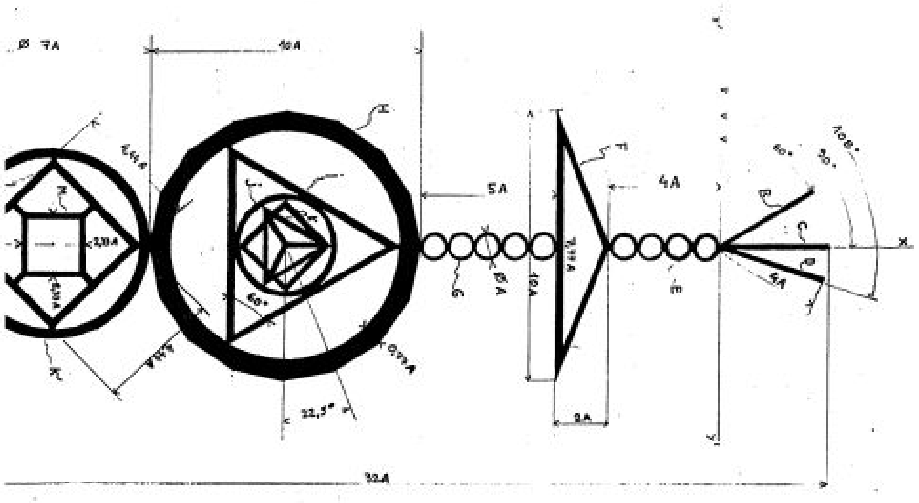

Abstract -- The invention relates to a device allowing the emission of wave forms which are beneficial for all living things. It consists of a set of geometrical shapes whose arrangement generates a symbolic application and gives rise to the appearance of powerful emissions which are beneficial for all living beings. The device according to the invention is in particular intended for creating a beneficial vibrational field.

Cited documents: FR2329089 // FR2554354 // FR1584210 // FR2421531

FR 2591494

Ambience harmoniser consisting of a cube surmounted by three hemispheres which can move around a vertical axis

Inventor: BABONNEAU BERNARD

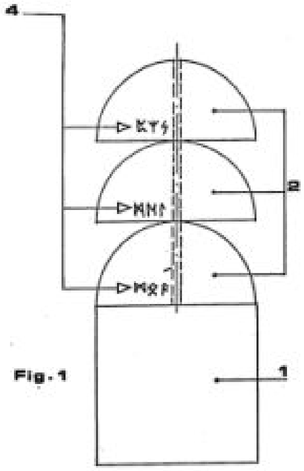

Fig. 1

VUE DE FACE ET DE COTÉS

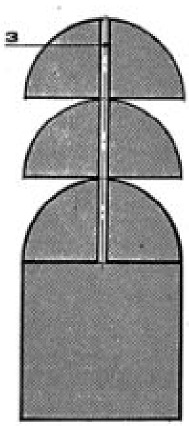

COUPE aa

Fig. 3

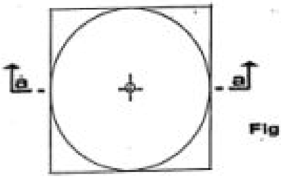

Fig. 2

VUE DE DESSUS

Abstract -- Device for harmonising a location. The invention relates to a device for obtaining or re-establishing maximum vibrational harmony of a location. It consists of a cube 1 surmounted by three hemispheres 2 which can move around a vertical axis 3. Twenty-four "runes" 4 are pyrographed on each hemisphere. The device according to the invention is intended in particular for general harmonisation working towards maintaining plants, animals and human beings in "good health" in the place in which they live.

FR 930466

Method and apparatus for the processing of fluid by radiation

Inventor: PEQUIGNOT MICHEL

FIG. 1

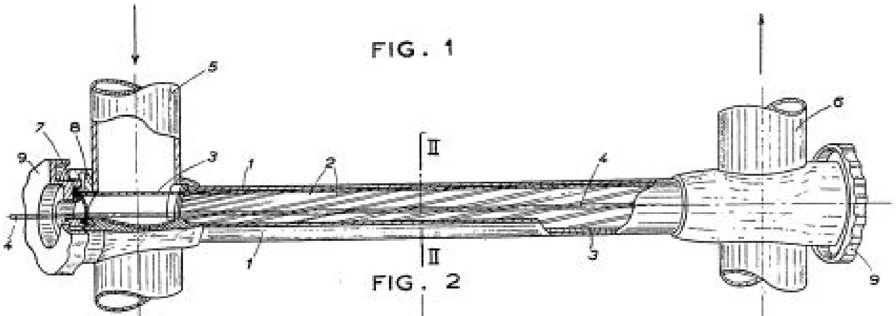

FIG. 2

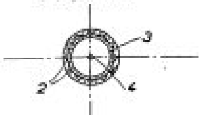

FR 1153623
Collecting and use of the cosmic rays

Inventor: LARONZE JEAN-ANDRE

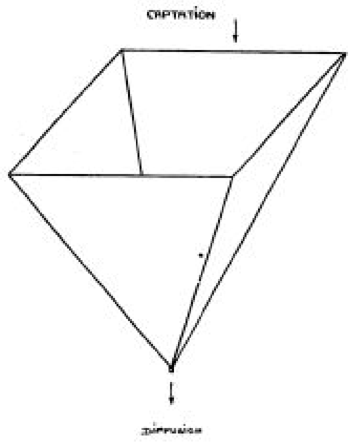

DE 2457792
TECHNISCHE ANORDNUNGEN

Inventor: HANSCHMANN HORST

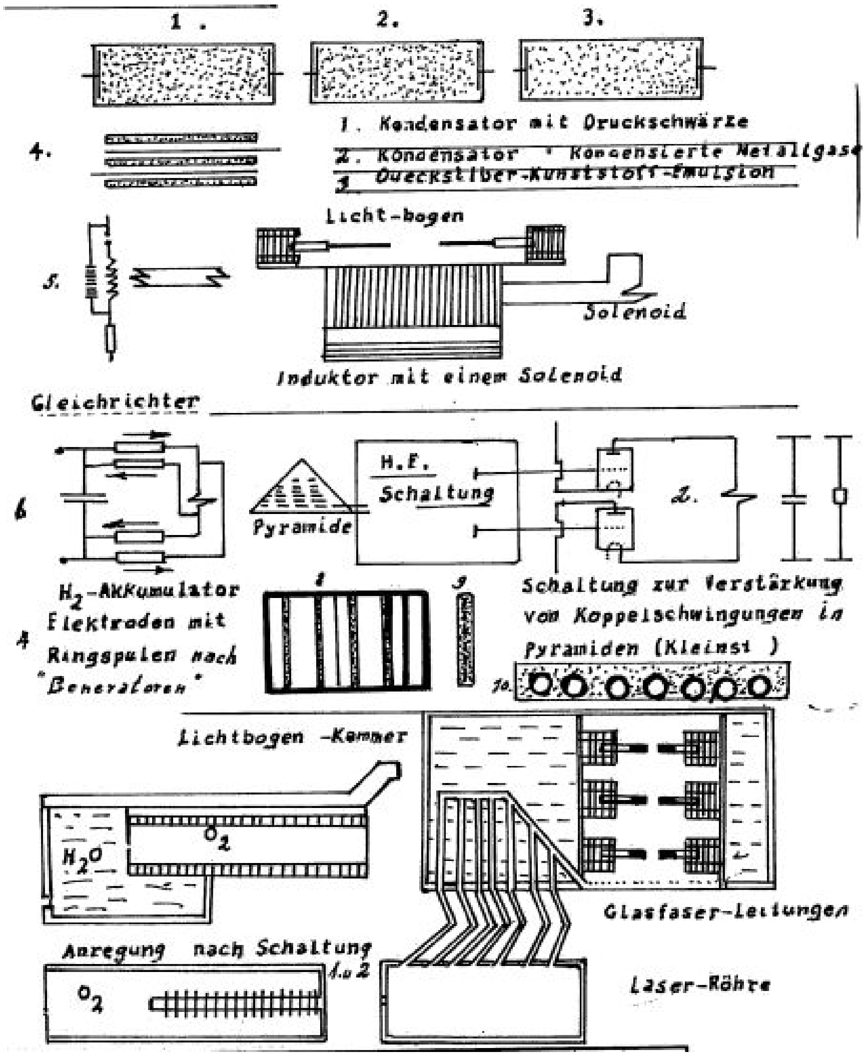
1.
2.
3.
4.
1. Kondensator mit Druckschwärze
2. Kondensator * Kondensierte Metallgase
3 Quecksilber-Kunststoff-Emulsion
5.
Licht-bogen
Solenoid
Induktor mit einem Solenoid
Gleichrichter
6
Pyramide
H.F.
Schaltung
2.
Schaltung zur Verstärkung von Koppelschwingungen in Pyramiden (Kleinst)
H_2-Akkumulator Elektroden mit Ringspulen nach "Generatoren"
4
8
9
10.
Lichtbogen -Kammer
O_2
H_2O
Glasfaser-Leitungen
Anregung nach Schaltung 1.u 2
O_2
Laser-Röhre

Construction of arrangements and technical apparatus among other things for the degradation of cancer cells.

[DE 8535780](http://www.rexresearch.com/radiesthpat/radiesthpat.htm)
Probe

Inventor: HANS BRUEGEMANN GMBH

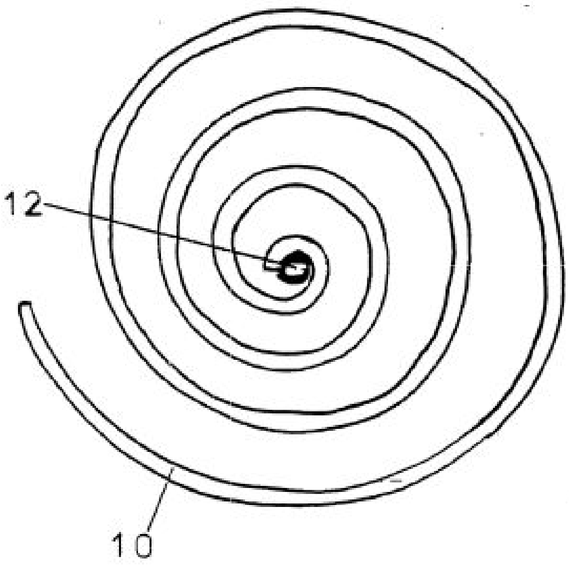

FIG. 1

Such probes with body electrode are known and come before all things for therapeutic purposes to the use. A known therapy procedure plans it to strengthen the bio alternating field in one, detected of the body of an patient with an acceptance probe, the probe downstream metre to diagnose and provide feeding back with the calculated complement either over the same acceptance probe or over an other particular

---

DE 2559383
Method & Apparatus for Programming of Cells

Inventor: HANSCHMANN HORST

GB 2117183
Protective apparatus

Inventor: ARENS CARL [DE] ; SCHMIDT PAUL

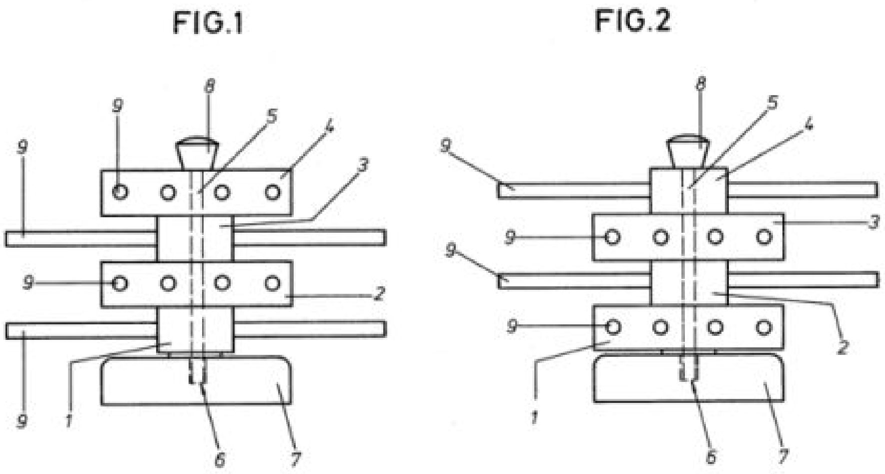

Abstract -- Apparatus for protecting living organisms against geo-biological radiation comprises a series of layers of spaced apart electrically conducting metal rods 9. The rods in each layer are parallel to each other and are held in an electrically insulating holder member 1-4 of plastics material. The members 1 to 4 are clamped together by a knob 8 on a rod 5 which is screwed into a pedestal 7. The rods in adjacent layers are shown at 90 degrees to each other, but this angle can be adjusted by loosening the rod 5, turning the holder members and then tightening the rod again.

Also published as: FR2523850 // DE8208300 // CH658790
Cited documents: DE3036085 // GB1077765

DE 7143603

Shielding Equipment

Inventor: ANGELE J

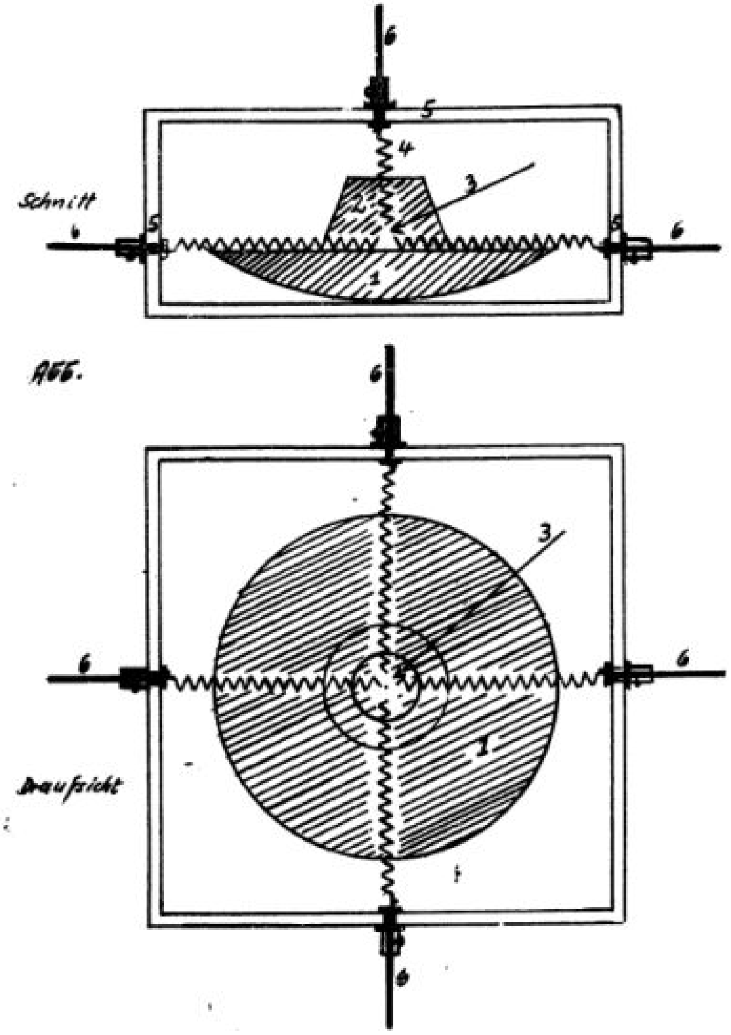

The invention relates to shielding equipment to weaken electrical radiation, caused from shafts by running water in the earth and variations of the earth's magnetic field, furthermore wave gate zones and other effects of other origins.

CH 145271

Mechanism to Collect Telluric Energy

Inventor: EDUARD HART

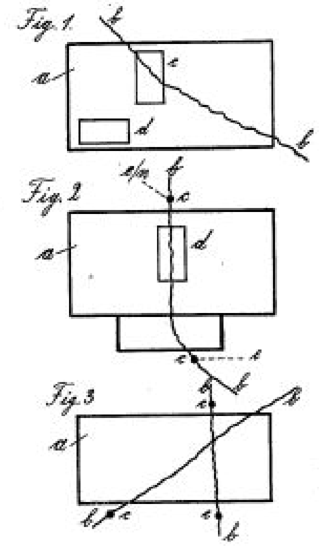

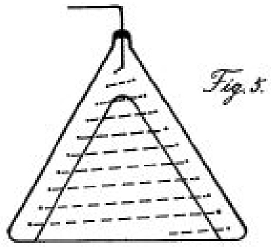

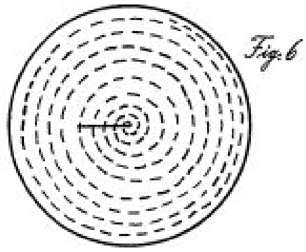

[EP0651595](#)
Process and associated device for the neutralisation of the effects of electromagnetic fields within open areas and/or volumes

Inventor: CICCOLO ENZA
Cited documents: GB385987 // DE3543765 // DE3433292 //DE4009003

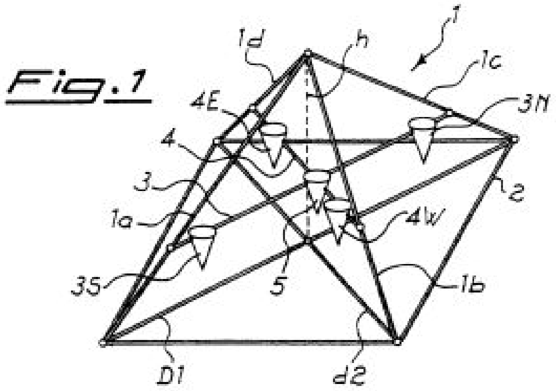

Abstract -- Device for neutralizing the effects of electromagnetic fields inside areas and/volumes in which said fields occur, consisting of a pyramid (1) having as its base a rhombus (2), the greater diagonal (D1) and smaller diagonal (d2) of which are respectively of a length such as to give rise to a so-called golden ratio D1/d2, and also of a height (h) such as to give rise to a ratio h/D1 also of the golden type, there also being provided at least one pair of arms (3) and (4) respectively parallel to the greater diagonal (D1) and to the smaller diagonal (d2) and respectively joining pairs of opposite edges (1a,1c,1b,1d) of the pyramid (1), said arms (3,4) supporting in the region of said edges respective containers (3N,3S,4E,4W) designed to contain a predetermined quantity of a liquid and intersecting in the region of a further, central, container (5) arranged along the height of the pyramid.

WO 9212761
DEVICE FOR ACCELERATING GROWTH

Inventor: KOLAR GABOR

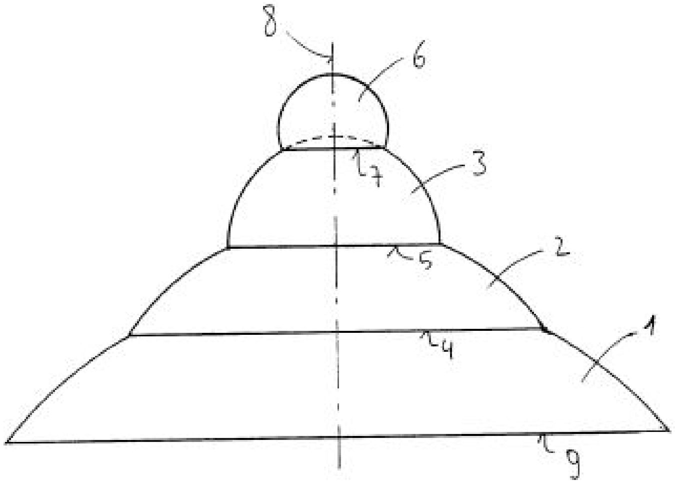

Fig. 1

Abstract -- The invention relates to a device for promoting and accelerating growth in human beings and animals and for accelerating the recovery of human beings and animals following illness and injury and to support healing processes. It consists of a combination of at least two of the items below, possibly with suitably selected intermediate components: a) pyramids with at least hexagonal bases; b) truncated pyramid with at least hexagonal base and summit surface; c) cylinders; d) tapered domes; e) truncated tapered domes; f) toroidal section; g) paraboloid region or cap; h) ellipsoid region or cap (9, 10); i) ridged band (11) on the surface of one of items a) to h), which may be of various shapes. Two items of the same group with differing dimensions may also be combined, with the proviso that, if only one of items a) to h) is combined with item i), there are at least six items i).

Also published as: JP5505332 // HU66203 // EP0522117 //AU1176492 // AT397618
Cited documents: EP0302192 // EP0229874 // FR2329089 // FR2572289 // FR2591494

WO 2006070213
DEVICE TO TRANSFER BIOLOGICAL ENERGY INTO BIOMECHANICAL ENERGY

Inventor: AKSZJONOVICS IGOR VASZILJEVICS
Also published as: RU2007129006 // HU0402686 // EP1861163

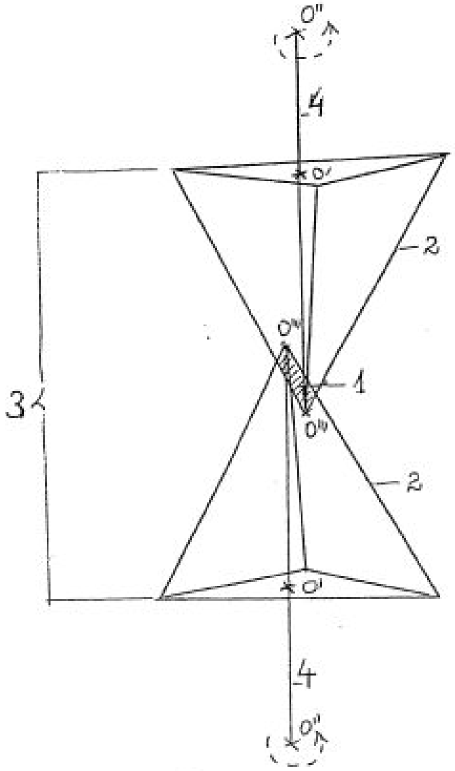

Fig. 1

Abstract -- The invention relates to a device realises the transformation of biological energy into biomechanical energy in the form of a geometric body. The characteristic feature of the invention is that it contains a number of pyramids or cones inserted into one another at their points or bases which form a unified geometric form, with a common space inside them, which may stand empty or be filled with organic or inorganic material, which may have any form, colour or extent, what is more the extent and form of the internal volume may be changed and regulated, the structure itself may be made of organic or inorganic material.

WO 9404221
DEVICE FOR ACCUMULATION OF BIOLOGICO-COSMIC ENERGY

Inventor: DODONOV BORIS PETROVICH [RU] ; DODONOV ROMAN BORISOVICH
AKTSIONERNOE OBSCHESTVO ZAKRYT [RU] ; DODONOV BORIS PETROVICH [RU] (+1)
Also published as: AU2517692

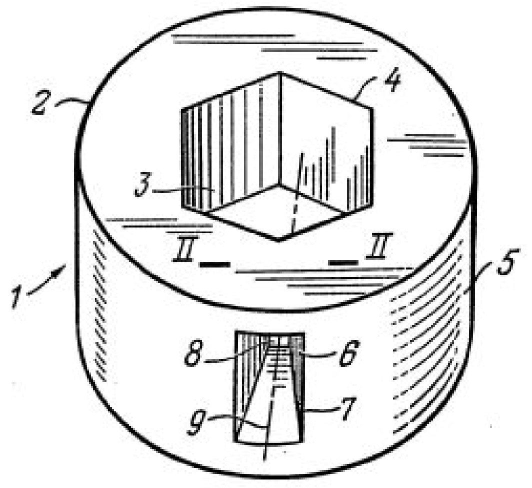

FIG. 1

Abstract -- The device (1) comprises a body (2) with a cavity in its central area having the form of a linear channel (3) and connecting with the surrounding space by an inlet opening (4). The lateral wall (5) of the body (2) is provided with at least one linear tangential channel (6) connecting by an outlet opening (8) with the linear channel (3) and by an inlet opening (7) with the surrounding space. The longitudinal axis of the tangential channel (6) is situated in a cross-sectional plane of the channel (3).

DE 3826890
Method for influencing the degree of order of an inorganic or organic system by using the change, occurring in the area of action of a hollow pyramid

Inventor: KEPPELER MANFRED
EC: B01D9/00

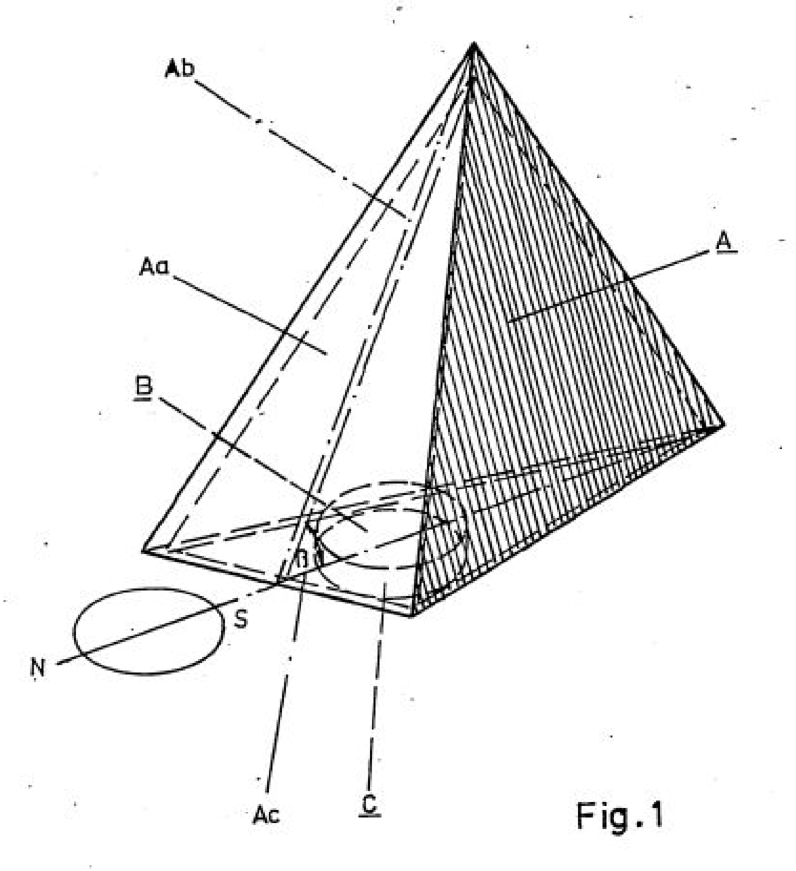

Fig.1

Abstract -- The invention includes a method for influencing the degree of order of an inorganic or organic system by using the change, occurring in the active region of a hollow pyramid, in the sub-region of the terrestial magnetic field otherwise prevailing there, by directing the terrestial magnetic subfield to at least one hollow pyramid face, which is inclined essentially at the local inclination angle, is aligned together with the inclination plane with the magnetic north-south direction and is formed from diamagnetic material, on to the system, which is preferably brought into the interior, below or in the immediate vicinity of the pyramid, and exposing the system to the influence of the pyramid in essence over the duration of its natural development time. In addition, the invention includes a hollow pyramid for carrying out this method, in which at least one pyramid face is inclined at the inclination angle which is decisive for the application site, and at least this pyramid face consists of diamagnetic material having a volumetric susceptibility below -15. The hollow pyramid preferably consists of brass, zinc, silver, bismuth, graphite or ytterbium with in each case a volumetric susceptibility of below -15, or else of an alloy or composition of these materials. The method according to the invention is suitable for improving the formation of crystals in solutions, cultivating microorganisms (microbes), improving the germination of seeds...

EP 0259769
Installation for pyramidal energy

Inventor: OEHME RAINER
Also published as: DE3630523

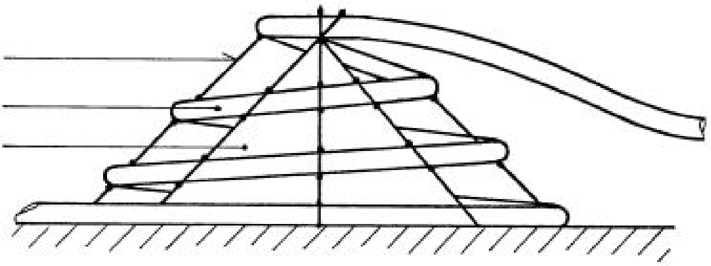

Fig. 1

Abstract -- A conductor (2) runs externally over a pyramid (1) in the form of a conical spiral. Its outgoing conductor part (4) is passed over the apex of the pyramid (1). The conductor (2) consists either of an electrical conductor or alternatively of a tubular body through which a medium to be treated can be passed. Using such a conductor (2) as an energy absorbing device, the complete vibration spectrum of the pyramid (1) can be detected.

EP 0287687
Device for the treatment of living beings and plants by pyramidal energy

Inventor: OEHME RAINER
EC: A23L3/26; A61N1/04

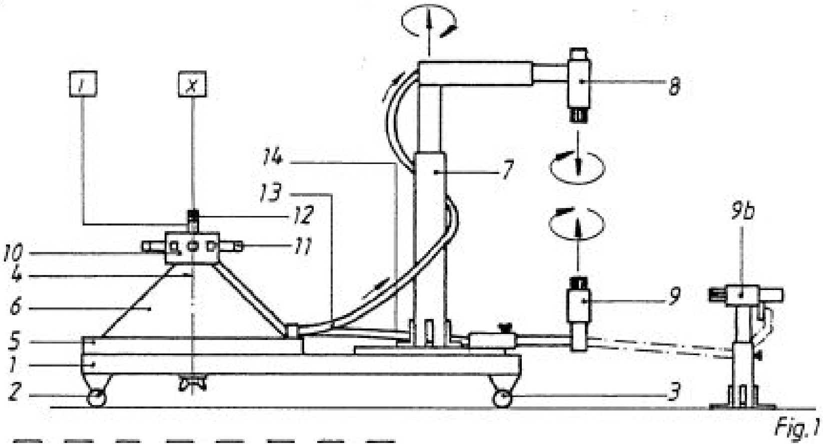

Fig. 1

Abstract -- Pyramidal energy is taken from a pyramid (6) by means of a vibration pick-up applied to it from above. Lines (13, 14) are guided down from this vibration pick-up (10) along an edge of the pyramid (6) and convey the picked-up energy to an upper and lower chakra electrode (8, 9), which are arranged on a support (7).

EP 0229874
Protection device

Inventor: PALM HUBERT DR MED

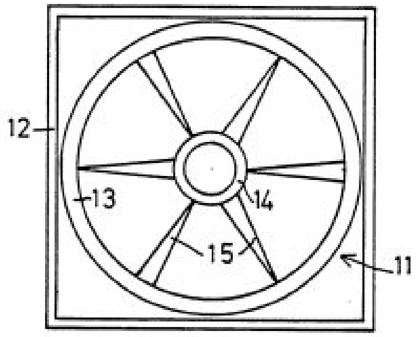
12
13
14
15
11

FIG. 1

Abstract -- The protective effect and healing action of the device for all life is based on its qualitative biological and structural equivalence to the harmonical system of the cosmos and especially, according to Kepler, of the solar system. Its construction starts from the assumption of all high religions and high cultures that the good part of the cosmos as a whole - as creation of God - and therefore with all its parts is an integral unit of life. Harmonic equipment that bears a qualitative structural resemblence to life and, in particular, is capable of electromagnetic vibrations, must therefore process biologically the same cosmic life forces as it has to transmit inductively, which experience shows occur everywhere in the cosmos, especially in the form of vibrations.

DE 3530841
Chakra electrode

Inventor: OEHME RAINER

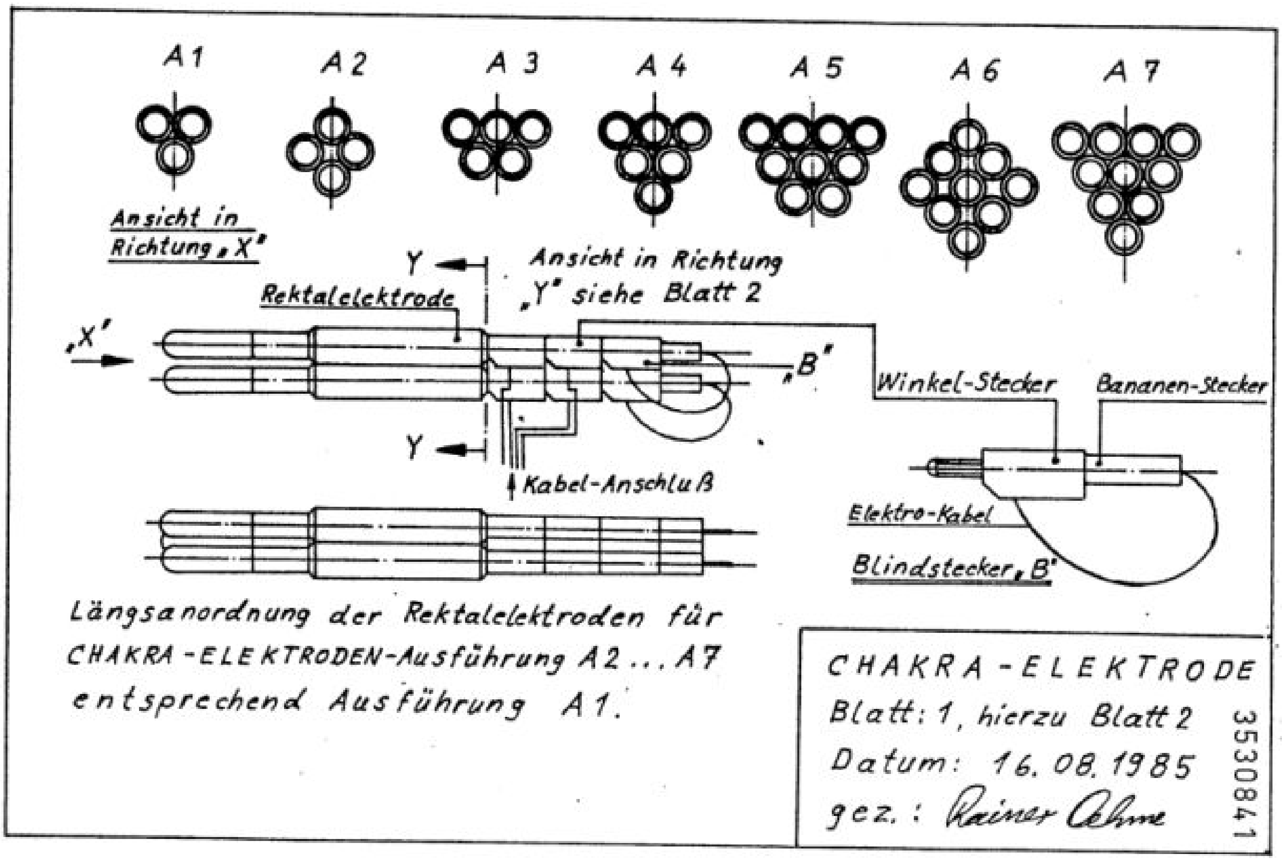
A 1
A 2
A 3
A 4
A 5
A 6
A 7
Ansicht in Richtung „X"
Y
Rektalelektrode
Ansicht in Richtung „Y" siehe Blatt 2
„X"
„B"
Winkel-Stecker
Bananen-Stecker
Y
Kabel-Anschluß
Elektro-Kabel
Blindstecker „B"
Längsanordnung der Rektalelektroden für CHAKRA-ELEKTRODEN-Ausführung A2 ... A7 entsprechend Ausführung A1.
CHAKRA-ELEKTRODE
Blatt: 1, hierzu Blatt 2
Datum: 16.08.1985
gez.: Rainer Oehme
3530841

FR 2580180
System for protection against the physiological effects of electromagnetic fields

RAZE ETIENNE [FR]

Abstract -- System intended for protecting living beings against the physiological effects of electromagnetic fields, characterised in that it is composed of at least two devices each consisting of a casing, on the bottom of which there are disposed at least a first bar made of a magnetisable material and at least one second bar made of a material which is non-magnetisable but which has a high coefficient of conductivity, these bars being disposed without touching each other in an open angular pattern, the remaining volume of the casing being filled with ferrite particles dispersed within a dry, inert microionised powder, and agglomerated using an inert binder, these devices being subjected before use to an activation treatment consisting in passing them through at least one pair of electromagnetic fields with inverted orientations,; the said devices together defining a protective screen situated in the space which separates them.

FR 2572289
Device having the ability to cure certain cancers

LEMBEYE CHARLES [FR]

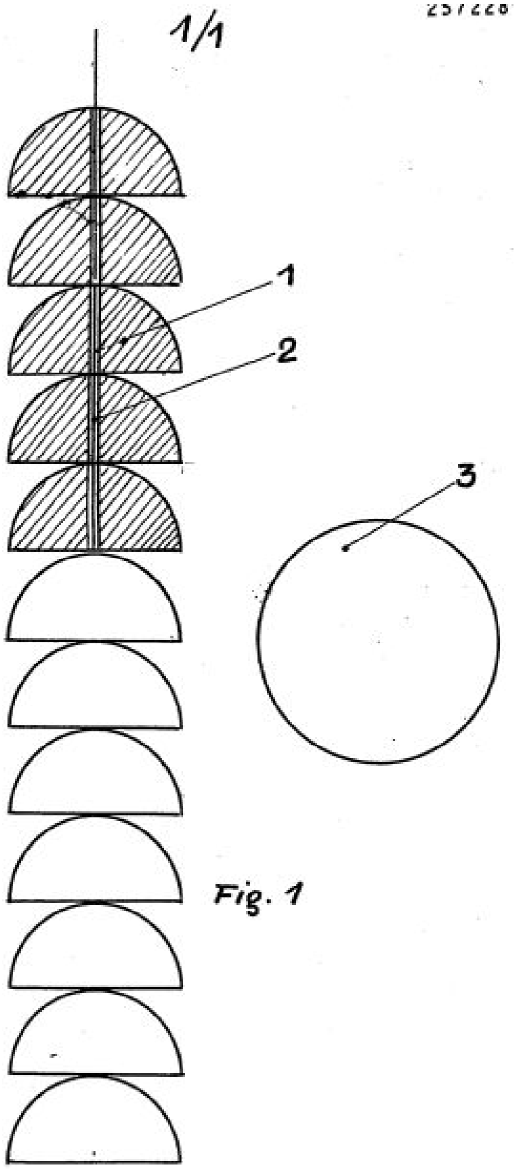

Fig. 1

Abstract -- The invention relates to a device which has the ability to transmit an energy flux which cures certain cancers. This device consists of twelve hemispheres made of hard wood pierced with a hole 1 and threaded on a leather thong 2. A leather washer 3 is placed under the stack thus constituted.

CA 2043936
Device for accumulation, amplification and local direction of the positive cosmic energy

Inventor: STOJANOVIC BRANISLAV
Also published as:
EP0460670 // AU7819491

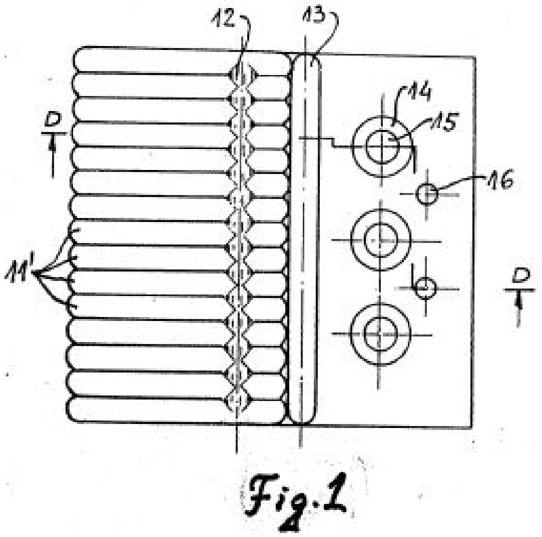

Abstract -- The novelty about the device for accumulation, amplification and local direction of the positive cosmic energy is that between the V-shaped arms (11) of the support (1) there is anothed V-shaped piece (2) which is fixed tightly by its arms (21) between the convexity (16) and a semi-cylindrical convexity (13) on the arms (11). At the end of the arms (11) there are cylindrical convexities (11') over which there is a convexity (12) at the angle of 90 degrees in relation to the longitudinal axis of the cylindrical convexities (11'). At the end of convexities (11') there is a cylindrical convexity (13).; On the outer surface of arms (11) there are three holes (15) on each side, respectively, and spherical convexities (14) around them, and two hemispherical convexities (16) on each side, respectively, next to them, whereas on arms (21) of the piece (2) there are three holes (24) on each side, lined with convexities (23) and there are two hemispherical convexities (22) on the outer surface of each side, as well as two hemispherical convexities (25) on the inner surface of the arms (21).

EP 0296142
Device to counterbalance the surrounding electromagnetic radiation

Inventor: DECKERS FRANCOIS

FIG. 1

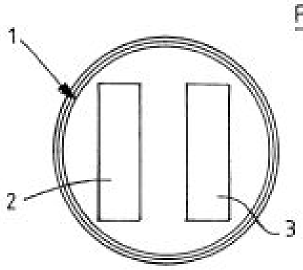

WO 9526215
Method of Immunomodulation of the non-specific reactivity of an animal organism

---

GB 1496582
An Energy Reabsorption Enhancement Chamber

Inventor(s): SACHER FRANCES VON

Also published as: NL7501808 // SE7412497 //JP50114888 // BR7500933 // DE2503622

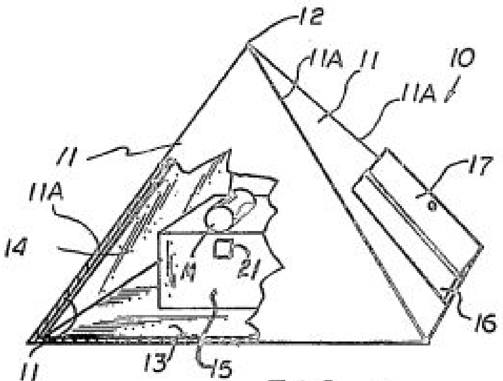
12
11A
11
10
11A
11
17
11A
14
M
21
16
11
13
15

FIG. 1

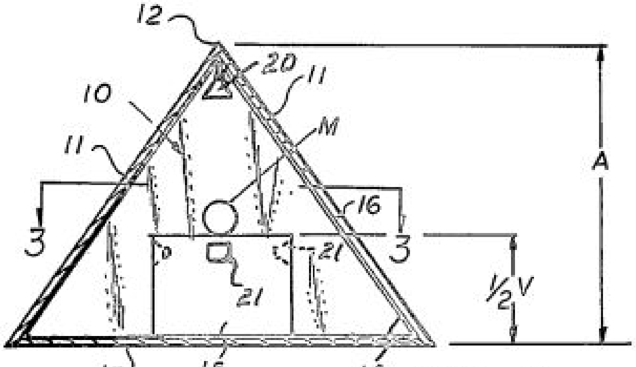
12
20
11
10
M
11
16
3
3
21
21
A
1/2 V
13
15
14

FIG. 2

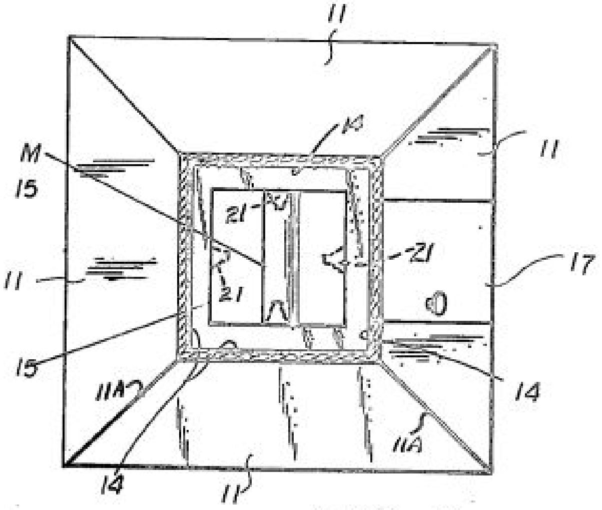
11
14
M
15
11
21
21
21
11
17
15
14
11A
11A
14
11

FIG. 3

FR 48135

Emetteurs d'ondes permettant de ranimer ou de transporter les ondes émises par les corps de la nature

Inventor: TURENNE LOUIS

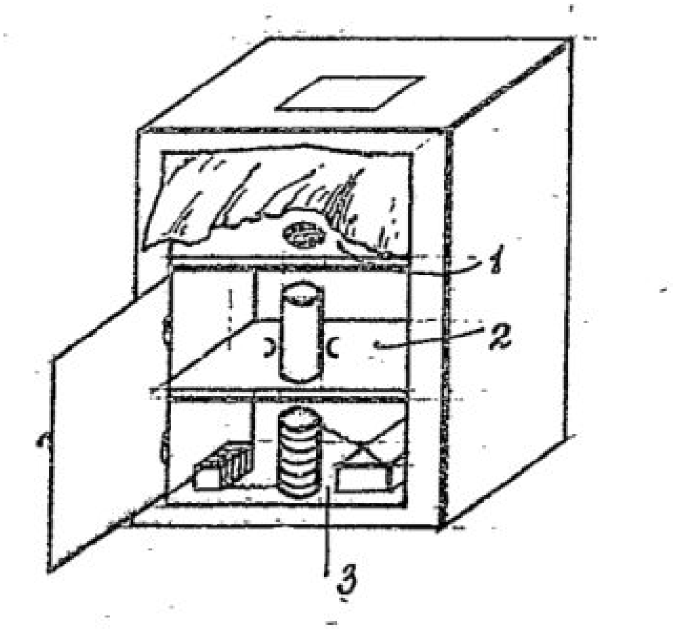

DE 1719878
ANORDNUNG ZUR ABSCHIRMUNG ODER ZUR NEUTRALISIERUNG VON FELDSTOERUNGEN

Inventor: JAKOB W F STAENGLE [DE]

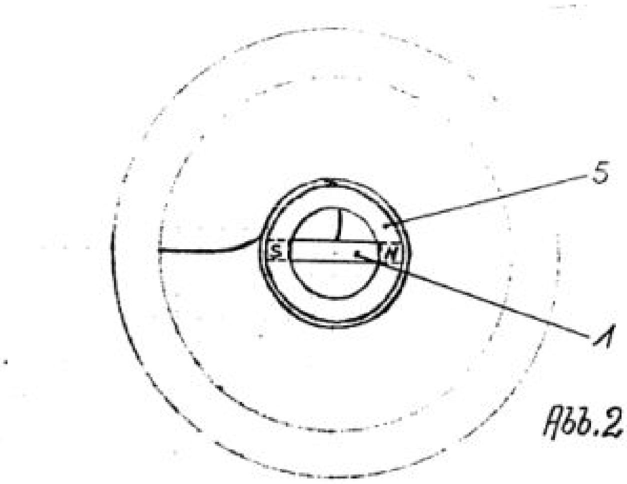

Abb.2

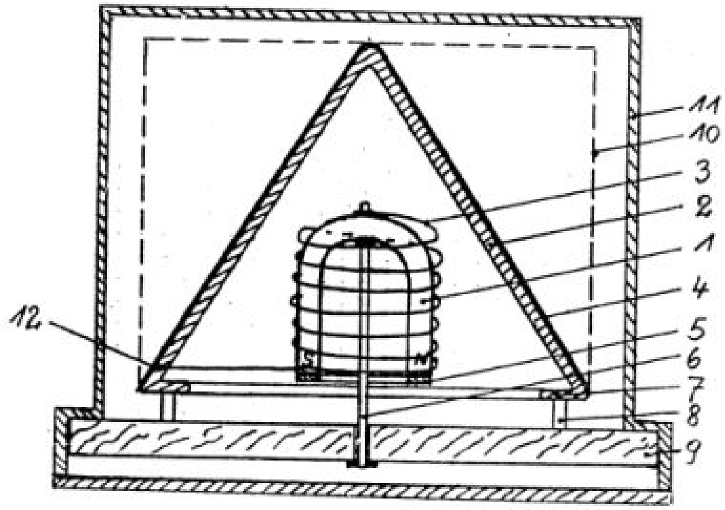
11
10
3
2
1
4
12
5
6
7
8
9
S
N

Abb.1

FR 485945
No title available

Fig 1

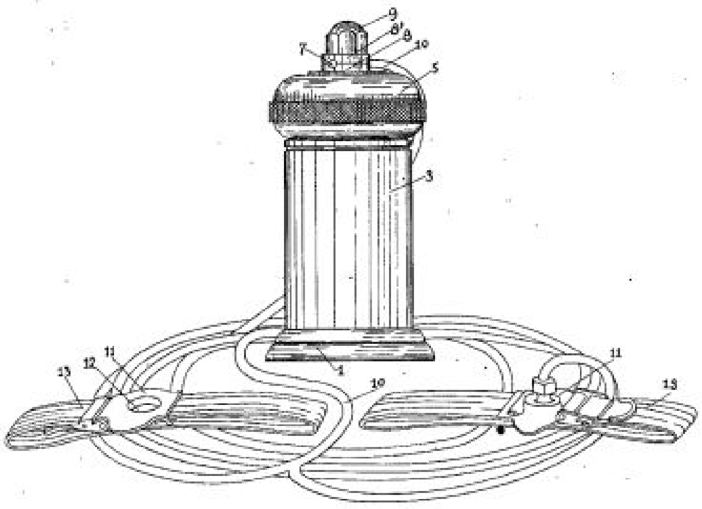
9
8'
9
10
7
5
3
1
10
11
12
13
11
13

Fig 2

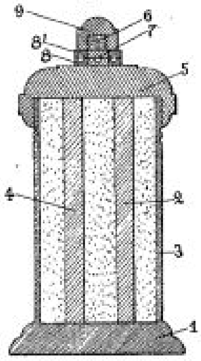
9
6
7
8'
8
5
4
2
3
1

Fig 3

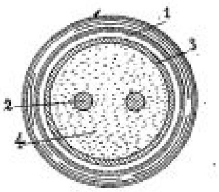

DE 3320510
Radiation switch for radiation deflection and irradiation

Inventor: VOELKNER KARL HEINZ [DE] Applicant: VOELKNER KARL HEINZ

16. Schaf

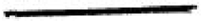

17. Widder

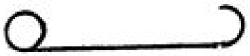

18. Wolf

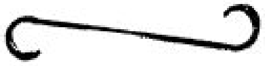

19. Hirte

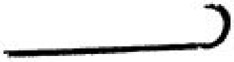

20. Trüffel

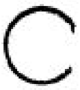

21. Pilz

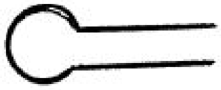

22. Embryo

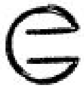

23. Klee

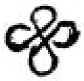

24. Baby

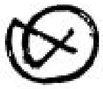

25. Schlange

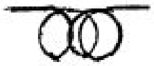

26. Spirale

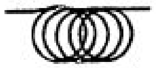

27. Spule

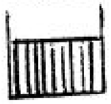

28. Saturnring, Sauger

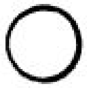

29. Van Allen Gürtel offen

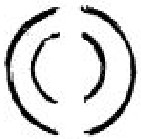

30. Van Allen Gürtel

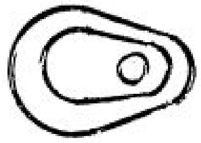

Abstract -- Radiation switch for radiation deflection and irradiation. b) Radiation surface suction probe for radiation deflection and irradiation. The patent specification for which a patent is applied is the combined result of my many years of research into environmentally harmful radiation returning to the sun. In fairly high concentration it is visible as livid-coloured light having a frequency of 40,000 mm. Livid radiation reflectors are produced and assembled by the industry over an uncontrolled, immense area. When used conscientiously and in a controlled manner, the switches and probes listed in the patent specification are suitable for preventing vegetation damage to man, animals, trees and plants. Measurable: with a 50 cm long golden chain on which a golden ring is threaded, shortened to 25 cm. The harmful radiation intensity starts at a pendulum deflection of 1 cm which leads absolutely to death of the

vegetation. Optically: photo/video: falsification of the spectral colours by superposition of the livid light. This livid radiation knows no obstacles. It can only be prevented by application and extension by experts of the law governing radiation protection.

WO 0218007
METHOD FOR PROTECTING A HUMAN BEING FROM THE ACTION OF A SPIN FIELD (RADIATION) HAVING LEFT POLARISATION OF SPACE

Inventor: ORLOV IGOR IVANOVICH [UA] ; BACHISHEV GORCHMA NURY-OGLY

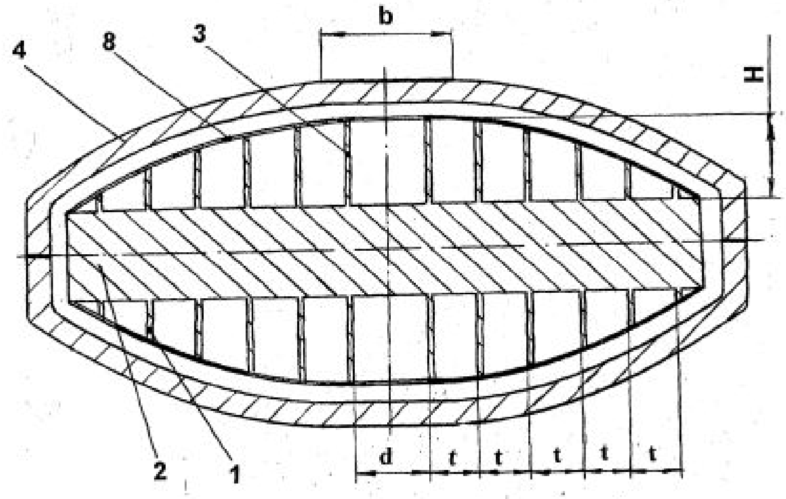

Abstract -- The invention relates to medical equipment which is directed at neutralising damaging effects produced by data display devices, more specifically terminal display devices, TV sets etc, and can be used for video devices, computers and power devices. Said invention can be also used for protecting against the negative effects of geopathogenic zones and geobiological systems. The inventive method is based on an additional field (radiation) having the opposite direction of a spin polarisation within the volume of the field. Said field is produced by the data display device. In addition, a supplementary compensating spin field having a right polarisation of environmental space is created with the aid of a physical vacuum polariser having a right polarisation which is based on the effect of prepolarisation of the polariser's substance. The inventive device comprises a compensating element embodied in the form of at least two coaxial sockets which are arranged on a common base. The internal socket has the lower of the dimensions d of the cross section thereof, which is equal to d> 0,5 H, where H represents the height of the socket. Said coaxial sockets have the cross section chosen from a range consisting of a polygon with the side number thereof m >= 3, a circle, a rotary body having an arbitrary shape and the combination of said figures. The walls of the sockets are embodied in such a way that they are sharp-edged on the side of the free ends thereof.

WO 9632192
DEVICE FOR INFLUENCING THE STRUCTURE OF MATERIAL OBJECTS

Inventor: UVAROV VLADIMIR VASILIEVICH
UVAROV VLADIMIR VASILIEVICH [RU]

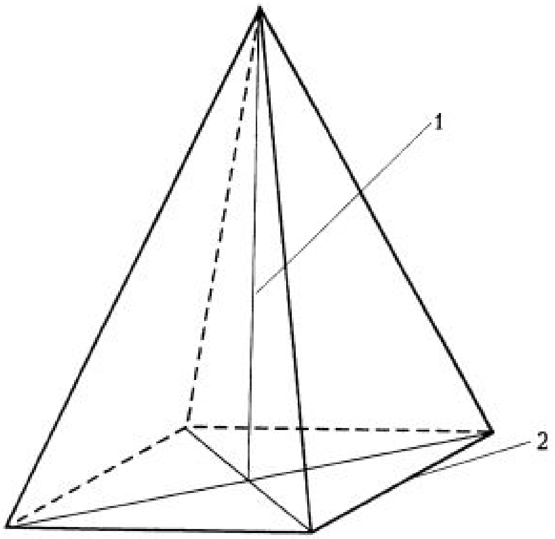

Fig. 1

Abstract --The technical problem addressed by the invention is that of enhancing an influence on material objects. To enhance a positive influence on material objects, regular rectangular pyramids are proposed, made from a dielectric material and set up in such a way that each of the side facets is aligned with the directions of the horizon, while the geometrical shape of the pyramid is determined by a ratio of the base sides to the pyramid height expressed as follows: A = 1.572302...H or A = 0.485868...H, wherein A is the base side of the regular rectangular pyramid and H is the height of the regular rectangular pyramid.

WO 9319579
METHOD AND DEVICE FOR INFLUENCING AN OBJECT AND APPARATUS USING SAID DEVICE

Inventor: KARP YULY SEMENOVICH
Also published as: WO9319579 // EP0646392 // EP0646392 // AU3098692

Abstract -- A method for influencing an object participating in a technological process consists in directing an external energy on an intermediate element which changes its functional state and influences the object participating in the technological process, and the influence is formed as a secondary radiation which is interconnected with the external energy and which is generated from the substance contained in the material of the intermediate element. The formation of the secondary radiation flow with desired parameters is effected by periodical variation of the vector of anisotropy of the substance of the intermediate element and the thus formed flow of the secondary radiation is directed on the object.; The device comprises an external energy source (2) and an intermediate element (3) capable of generating an influence in the form of secondary radiation matching with the external energy from the source (2) and resulting from the periodical change of the vector of anisotropy of the substance contained in the material of the intermediate element (3). The device is used as apparatus for programmed stimulation of self-regulation of biological objects for correction of their functional state.

DE 3320518
Device for changing radiesthetically measurable fields

Inventor: SCHWEITZER PAUL DR

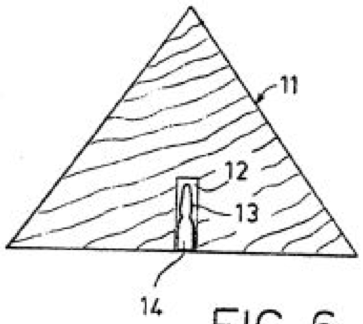

FIG. 6

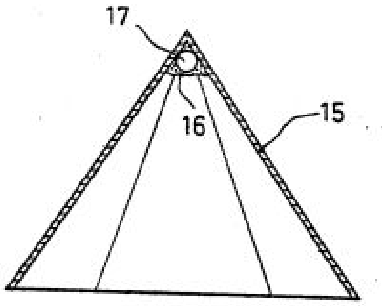

FIG. 7

Abstract -- The invention relates to a device for changing radiesthetically measurable fields with the object of avoiding the pathogenic effects of their biologically active regions, the so-called irritant zones. These fields are mainly caused by underground water veins. However, they are also found in space lattice-like structures, the so-called lattice systems incident from the atmosphere. All the measures and devices known to date for attenuating or eliminating the effect of the irritant zones have shortcomings or disadvantages. As a rule, they are ineffective with regard to the particularly dangerous crossing region of the irritant zones, or they detect only individual frequencies and must be adjusted with the risk of inefficacy due to maladjustment. The invention is based on the object of creating a device for changing radiesthetically measurable fields which avoids the aforesaid disadvantages. This object is achieved using a pyramidal body in that one or more elements, e.g. another pyramidal body, or an element which is resonant in the microwave range, e.g. a spiral, or a homeopathic preparation, is arranged within or directly underneath this body.

Your Support Maintains this Service

Buy

The Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

HACKER WAREHOUSE
HackerWarehouse.com
Your one-stop shop for all computer security needs from defense to offense.

FM Crystal radio

Article Featured Image

0 0 0 103

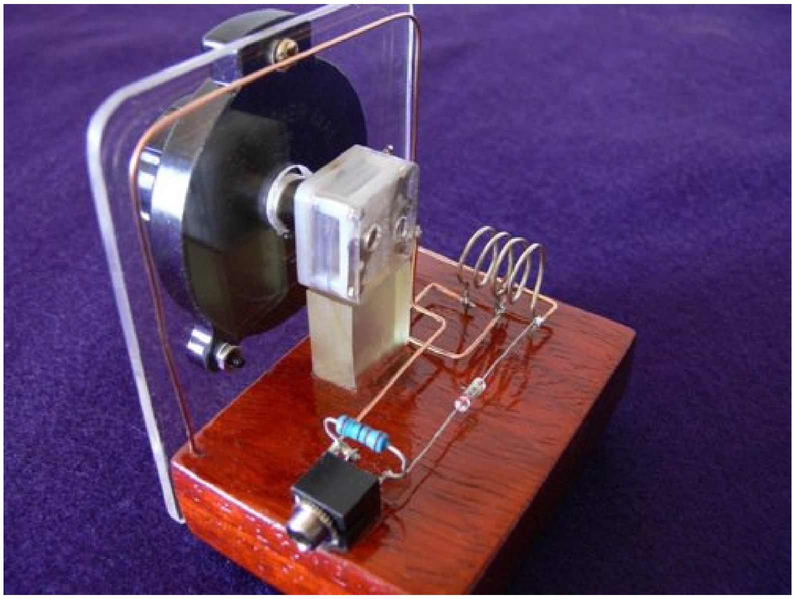

Larry writes –

> I have heard, even from a physicist, that it is impossible to build FM crystal radios. On the other hand some experimenters claim that they have built them. This argument intrigued me to try and build an FM crystal radio, which I have done successfully. To my surprise, the result is an astounding performer, pulling in four local stations in Tucson. When connected as a receiver to a good sound system the sound fidelity is as good or better than more expensive AM radios. In fact, it sounds “high-fidelity”.

FM Crystal Radio, thanks Pekar! – **Link.**

SPONSORED CONTENT

Beyond the Screen: Connected Devices Offer Increasing Amounts of Freedom

BY DEVRY UNIVERSITY

Related Stories from Make:

Lindsay's Technical Books

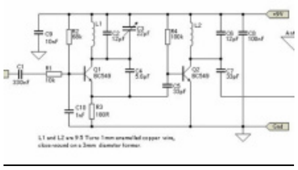

HOW TO - Make a low power FM transmitter

HOW TO - Make a Leopold bench

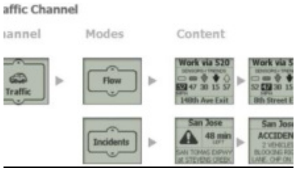

Traffic data, on your wrist watch...(A DIY Traffic Gauge?)

More Ideas From Make:

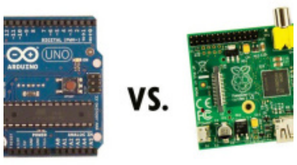

Raspberry Pi or Arduino Uno? One Simple Rule to Choose the Right Board

10 Classic Make: Projects

Build Heron's Fountain

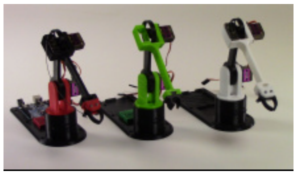

LittleArm Is a Little Robot with a Little Price Tag

Maker Shed Picks

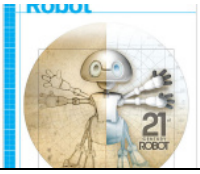

21st Century Robot, 1Ed

Getting Started with Adafruit FLORA

Getting Started with Adafruit FLORA

ComposiMold - Reusable Casting Mold

Arduino ADK Mega for Android

Powered by Contextly

ADVERTISEMENT

SHOW COMMENTS

By Phillip Torrone

September 18, 2007, 12:00 am PST

ADVERTISEMENT

Holiday Sale
30% Off
c6 CARBON6

RELATED TOPICS

DIY PROJECTS

TELECOMMUNICATIONS

ADVERTISEMENT

Holiday Sale
30% Off
c6 CARBON6

Let's Stay in Touch!

Enter your Email Sign Up

Learn More »

From Around the Web

How to Get Your Scissors Sharp Again

Quick & Dirty Tips

How to Get Rid of Rust on Your Car, Tools, Cast Iron, and More

Quick & Dirty Tips

15 Weight Loss Tips for Your New Year's Resolution

Quick & Dirty Tips

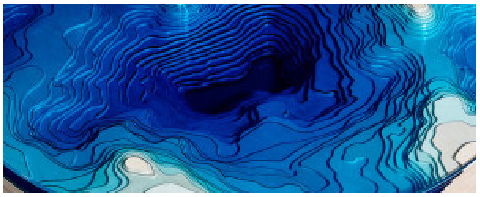

A Coffee Table That Lets You Stare into the Depths of the Ocean

Design Milk

powered by **PubExchange**

ADVERTISEMENT

Math Monday: Make a Flake

1 0 6

MATH MONDAY
NATIONAL MUSEUM OF MATHEMATICS

Today we revive **an old snowflake drawing program** by Evil Mad Scientist, and realize the designs in a modern meltingly sweet medium. **Here's an updated version** of that old program with the trigonometry corrected (cheers for applied math!) so that it will work with any number of sides to your snowflake. The program runs fine in the **latest version of Processing**; just download and install the development environment, and open the snowflake .pde file. Make sure to enter the number of sides you want your snowflake to have as the value of the variable "sides" on line 25 of the .pde file, overwriting the "6" that appears in the file as distributed. (If any reader would like to modify the program so that you can enter the number of sides in the running program without changing the code, please send the modification to mondays@momath.org and we'll post it here.) Then fire the program up by hitting the run button in Processing; it will look like this, shown with the number of sides set to seven.

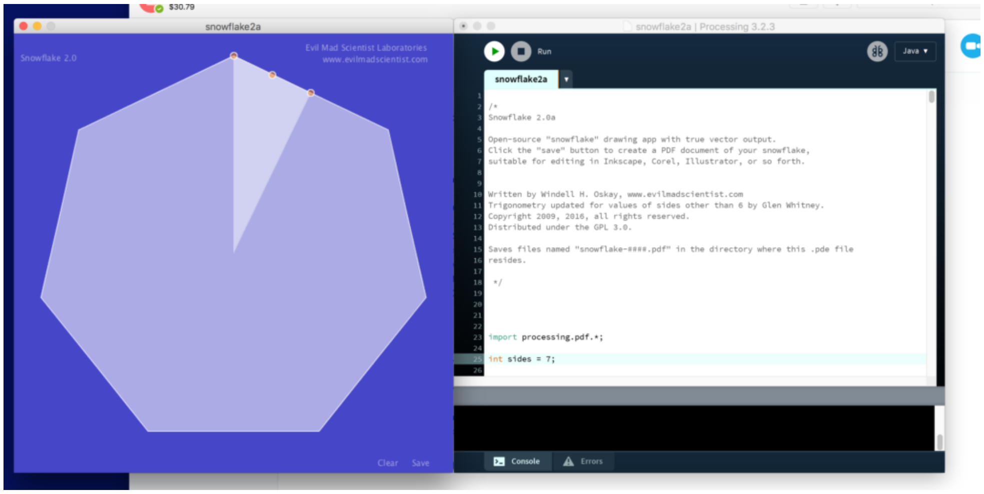
snowflake2a
Snowflake 2.0
Evil Mad Scientist Laboratories
www.evilmadscientist.com
Clear
Save
snowflake2a | Processing 3.2.3
Run
Java
snowflake2a
Console
Errors

```

/*
Snowflake 2.0a

Open-source "snowflake" drawing app with true vector output.
Click the "save" button to create a PDF document of your snowflake,
suitable for editing in Inkscape, Corel, Illustrator, or so forth.


Written by Windell H. Oskay, www.evilmadscientist.com
Trigonometry updated for values of sides other than 6 by Glen Whitney.
Copyright 2009, 2016, all rights reserved.
Distributed under the GPL 3.0.

Saves files named "snowflake-####.pdf" in the directory where this .pde file
resides.

 */




import processing.pdf.*;

int sides = 7;

```

You can drag any of the circular “handles” to alter the outline of the snowflake, and every time you move the middle handle of a segment, more handles appear so that you can make that outline more intricate. The program will automatically replicate your changes throughout the snowflake, keeping it symmetrical with the number of sides selected. It’s ok to cross the outline segments over each other, in which case holes will appear within the snowflake, potentially adding to the interest of the design. Here’s an example of a six-fold design in progress.

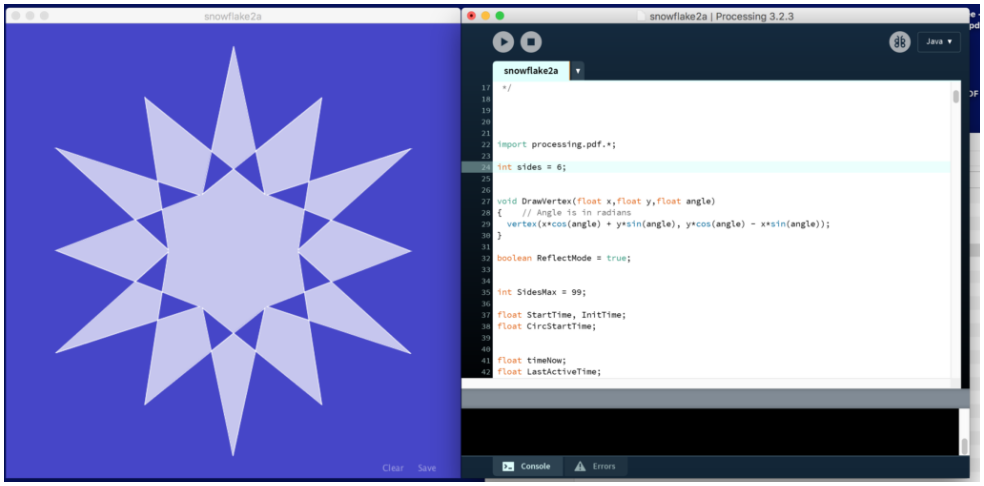
snowflake2a
Clear
Save
snowflake2a | Processing 3.2.3
Java
snowflake2a
Console
Errors

```
 */




import processing.pdf.*;

int sides = 6;


void DrawVertex(float x,float y,float angle)
{    // Angle is in radians
  vertex(x*cos(angle) + y*sin(angle), y*cos(angle) - x*sin(angle));
}

boolean ReflectMode = true;


int SidesMax = 99;

float StartTime, InitTime;
float CircStartTime;


float timeNow;
float LastActiveTime;
```

When you have reached a symmetric polygon, a.k.a. snowflake, that you like, hit the “Save” button on the snowflake screen to write a pdf file of your design. I also suggest saving a screenshot to catalogue the open and closed areas of the snowflake. For example, here’s the five-fold pattern used for the purposes of this column.

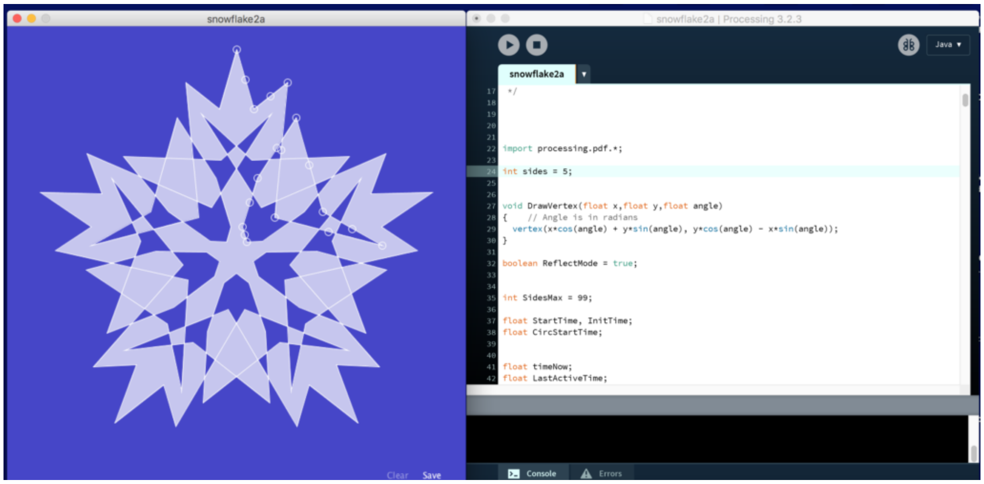
snowflake2a
Clear
Save
snowflake2a | Processing 3.2.3
Java
snowflake2a
Console
Errors

```
 */




import processing.pdf.*;

int sides = 5;


void DrawVertex(float x,float y,float angle)
{    // Angle is in radians
  vertex(x*cos(angle) + y*sin(angle), y*cos(angle) - x*sin(angle));
}

boolean ReflectMode = true;


int SidesMax = 99;

float StartTime, InitTime;
float CircStartTime;


float timeNow;
float LastActiveTime;
```

Using this program, three snowflake designs were created for today's making project: one **five-fold**, one **six-fold**, and one **seven-fold**. The idea is to render these designs as snowflake marshmallow cutouts to decorate festive winter-time drinks. Begin by printing out each design at dimensions just small enough to fit in the top of the cups or glasses you want to use.

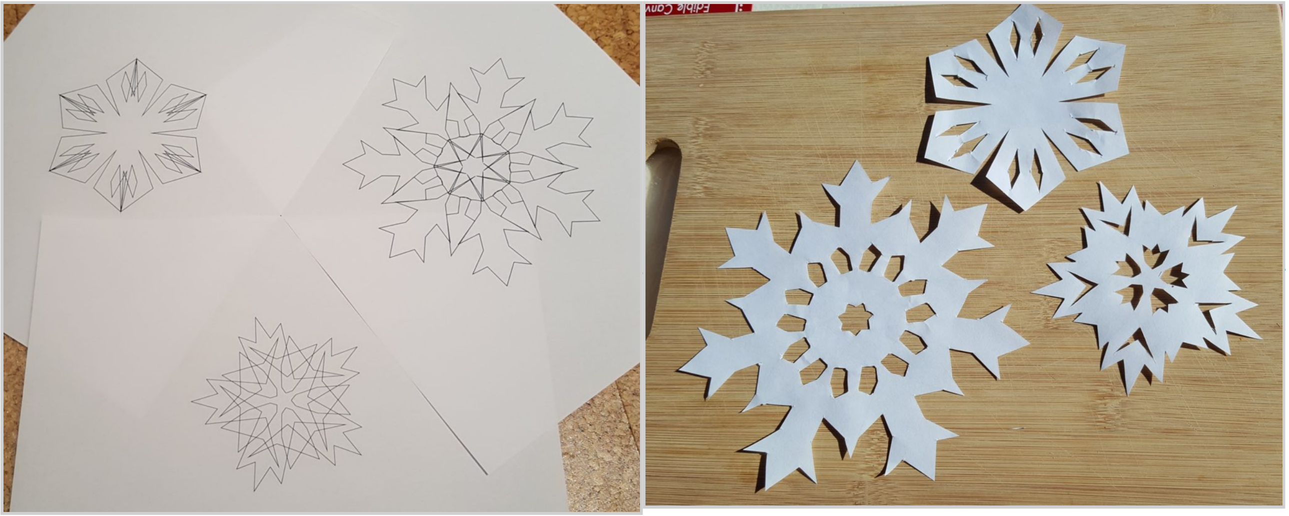

Then cut carefully around the outlines of the printouts, and select what areas of the snowflakes you want to cut out of the marshmallow snowflakes; carefully slice those portions out. This processes creates paper templates for the snowflakes.

MELVILLE CANDY
GIANT MARSHMALLOW
FULL SHEET
DIMENSIONS 15in x 11in x 1/2in
·IDEAL FOR·
Food Crafts
Cookie Cutters
Edible Canvas
Gluten Free Cakes
And Much Much
More!
1 pc NET WT 16oz (453g)

Next you need sheets of marshmallow, available in some craft or specialty food stores. Here's a sheet, one half inch thick. Line up one of the templates on the marshmallow sheet, and use thumbtacks through the paper and into the marshmallow to hold it in place during the cutting.

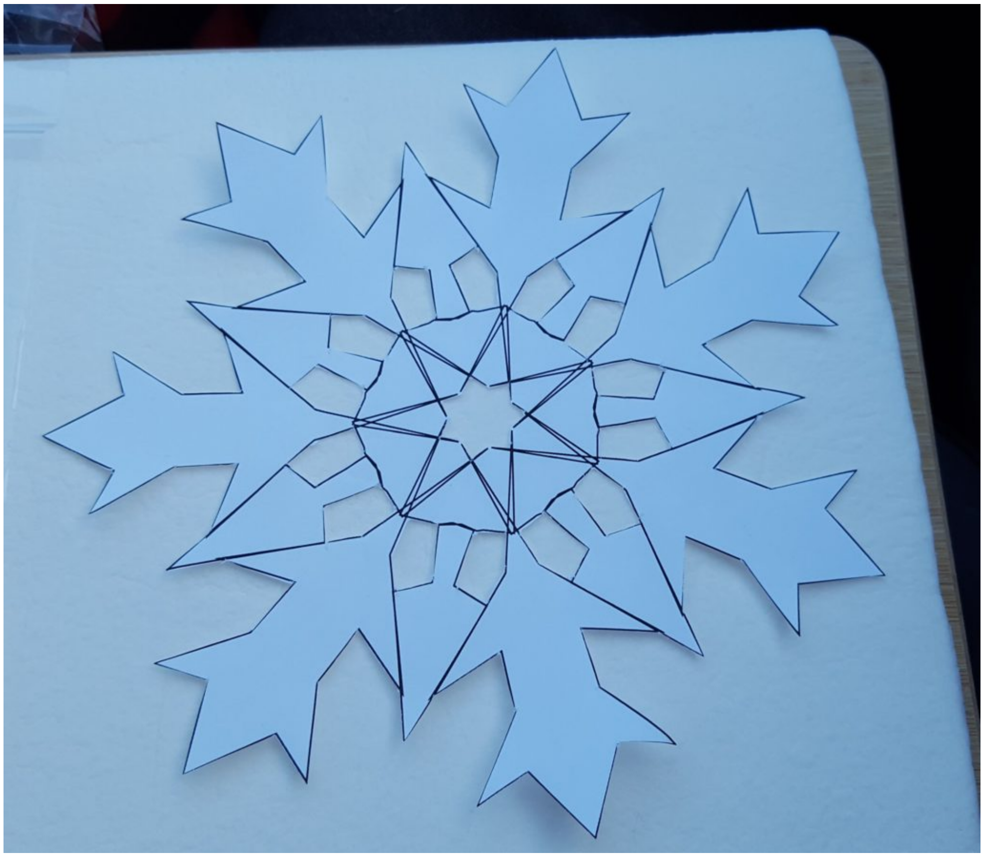

Then you can begin transferring the design to the marshmallow sheet. I recommend cutting it out with an X-acto or similar craft knife.

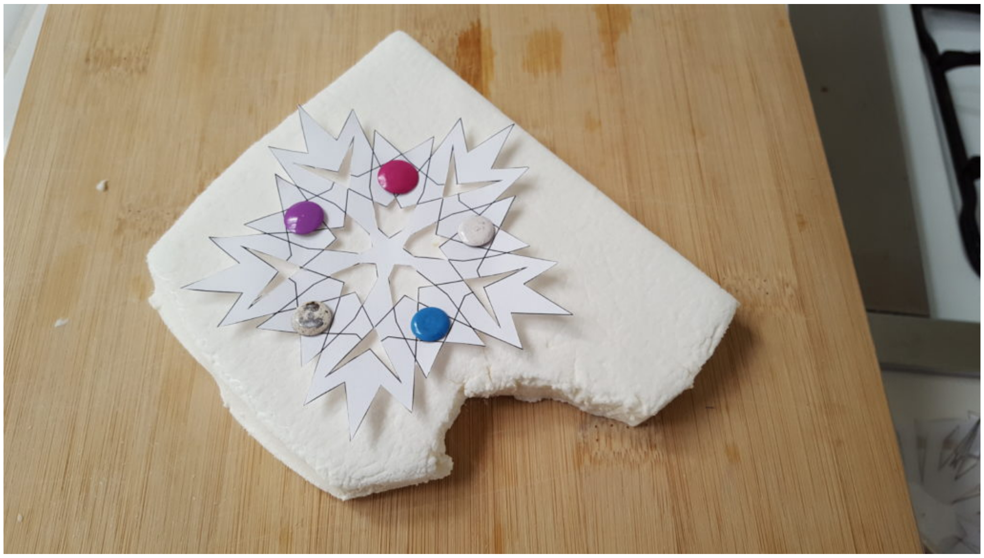

Here are some recommendations to make the cutting go more easily: first cut out roughly around the whole snowflake, but fairly close to its outline, because the cleanest cuts come from slicing along a segment of the design's perimeter all the way out to the edge of the marshmallow slab. Dip the knife blade in vegetable oil to prevent sticking, and keep it well oiled. Keep a damp cloth or paper towel handy to wipe off accumulated marshmallow stuff from the blade from time to time. To cut out interior holes in the pattern, cut around the entire outline of the hole once, go over it again lifting the marshmallow slab up to make sure the blade is going all the way through the marshmallow, and then push the cutout out from the front with the tip of a chopstick. Here's the seven-fold pattern just before and just after removing the template.

And here are all three different symmetries studied in airy sweetness:

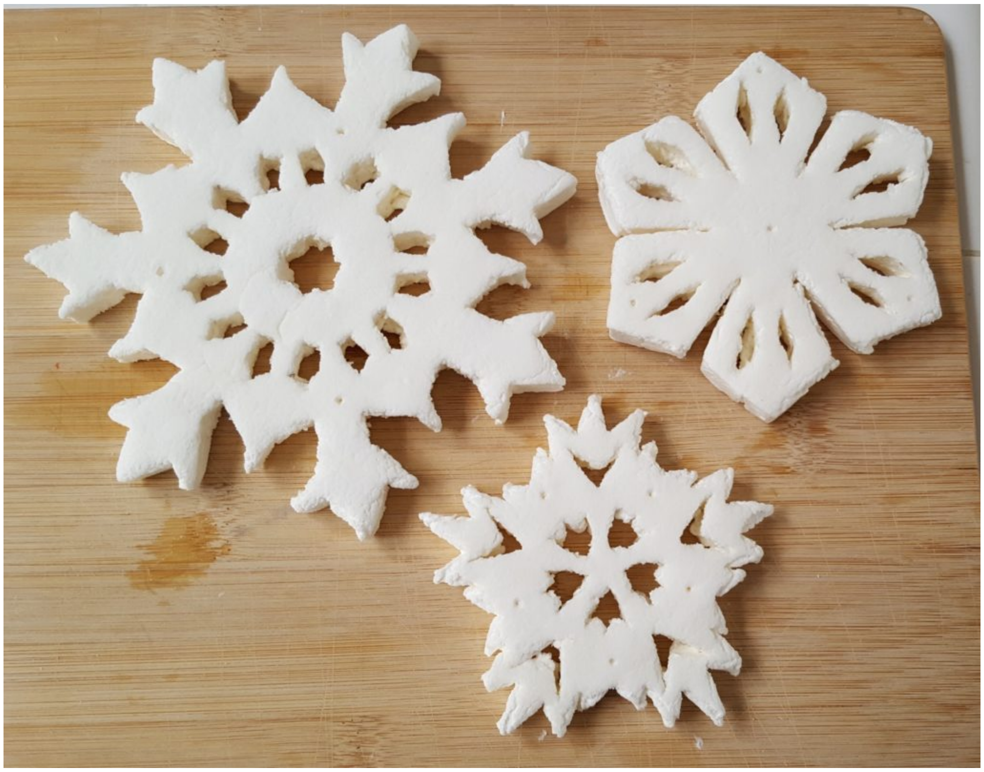

Of course, these geometric snowflakes cry out to serve as adornments for suitably wintry drinks — dark chocolate hot chocolate or eggnog, for example.

Happy New Year from Math Mondays!

SHOW COMMENTS

By Glen Whitney For The National Museum Of Mathematics (MoMath)

20 hours ago

RELATED TOPICS

FOOD MATH MONDAY

ADVERTISEMENT

ADVERTISEMENT

Radio a cristallo e dintorni - Le Radio di Sophie - Crystal radio sets

Ricevitore a cristallo con amplificatore

Un progetto di Luciano Loria

Si tratta di una classica radio a cristallo per O.M. con sintonia a variazione d'induttanza, (tramite il nucleo in ferrite che può scorrere all'interno della bobina di sintonia) con amplificazione del segnale audio rivelato.

Lo schema elettrico è suddiviso, idealmente, in due parti: la prima, sulla sinistra, comprende il circuito di sintonia e di rivelazione; la seconda, a destra, è un tipico amplificatore Darlington (Tr1 e Tr2 collegati a cascata).

L'amplificazione permette l'ascolto della stazione locale, in auricolare o in cuffia, anche se non si dispone di una lunga antenna filare, inoltre, nelle ore serali e con bel tempo, è possibile la ricezione d'altre stazioni radio, nazionali ed estere; il collegamento di terra, in presenza di segnali forti, potrebbe causare del ronzio; se ciò accadesse si può omettere.

Schema elettrico

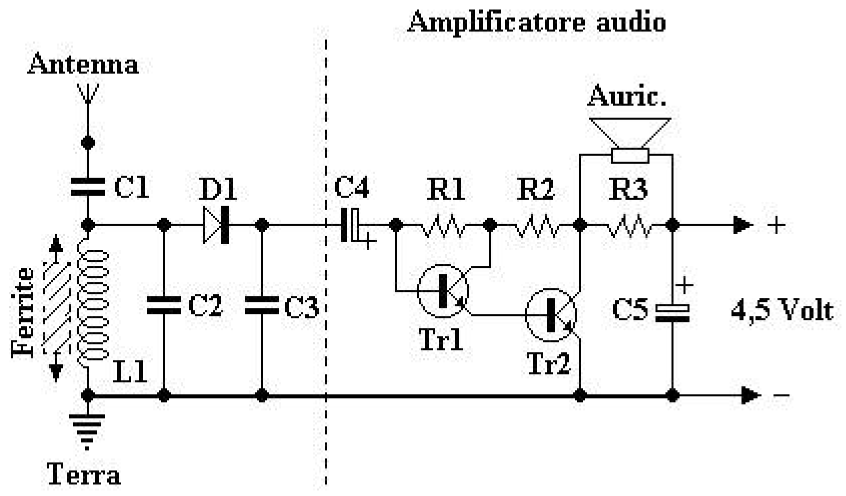

Componenti:
L1 = bobina 60/70 spire
C1 = condensatore 470 pF
C2 = condensatore 220pF
C3 = condensatore 4,7 nF
C4 = condensatore 1 uF elettr.
C5 = condensatore 47 uF elettr.
R1 = resistore 1 Mohm
R2 = resistore 4,7 Kohm
R3 = resistore 2,2 Kohm
D1 = diodo al germanio
Tr1 = transistor NPN BC547
Tr2 = transistor NPN BC547
Auricolare = Z da 100 Kohm
Ferrite = nucleo scorrevole di sintonia

Sullo schema elettrico c'è ben poco da aggiungere, si noti la linea tratteggiata che separa l'amplificatore audio dal rivelatore (diodo al germanio), il condensatore C4, oltre che da accoppiatore tra i due circuiti, funziona da filtro passa-basso, non permette, cioè, il passaggio di eventuali frequenze di A.F., presenti ancora dopo la rivelazione.

Il condensatore C1, posto tra l'antenna ed il circuito accordato di sintonia, funziona, invece, da filtro passa-alto, quindi, lascia passare l'alta frequenza e si oppone al passaggio della bassa frequenza, come quella di rete, che può disturbare la ricezione.

La sintonizzazione della stazione radio si ottiene facendo scorrere un pezzetto di ferrite all'interno della bobina di sintonia, l'induttanza della bobina è minima con il nucleo completamente all'esterno, è massima quando il nucleo è completamente all'interno.

Per ottenere risultati soddisfacenti è necessario provvedersi d’adeguata antenna filare, dai 10 ai 30 metri, stesa all’aperto ad almeno tre/quattro metri dal suolo e ben isolata dai supporti, in un punto qualsiasi collegare il filo di discesa all’apparecchio radio.

All’alimentazione provvede una comune pila piatta da 4,5 Volt che durerà tantissimo, l’assorbimento in corrente si aggira intorno ai 3-4 milliampere.

La costruzione dell’apparecchio non è assolutamente impegnativa ed è alla portata di chiunque abbia pazienza, un minimo d’abilità manuale e qualche arnese d’uso comune, presente, ormai, in ogni casa.

E’ indispensabile, comunque, per il montaggio, usare un saldatore elettrico da circa 30 W max. di potenza ed un rotolino di stagno per elettronica (lega Sn/60 Pb/40), occorrono anche un martellino ed un paio di forbici, molto utile può rivelarsi una pinza a becchi lunghi o, in alternativa, una pinzetta da manicure.

Tutti gli elementi della radio trovano posto su una tavoletta di legno di minime dimensioni e sono collegati, fra di loro, mediante saldatura ai punti d’ancoraggio che si predispongono piantando, in posizione opportuna sulla tavoletta, dei chiodini di ferro ottonato.

Montaggio

Per la buona riuscita della costruzione, seguendo i disegni e le foto, e per ottenere un effetto estetico discreto, date le dimensioni dei chiodini, occorre aiutarsi con delle pinzette per sostenere, perfettamente verticale, il chiodo mentre si va a battere col martello; lasciare la testa del chiodo fuori del piano per circa 3 o 4 mm.

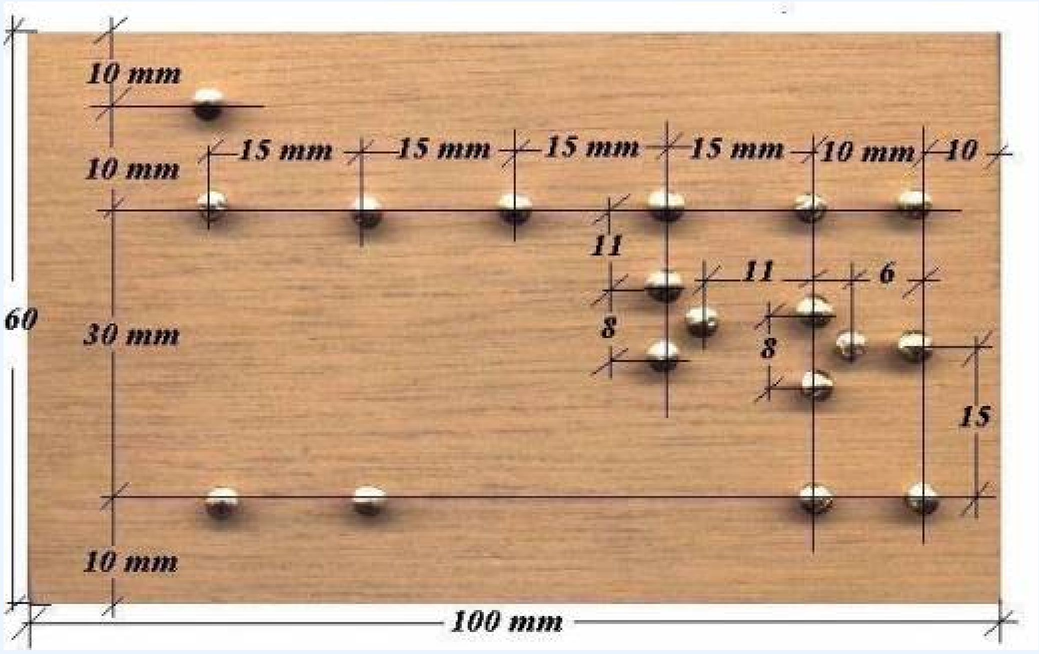

Si comincia proprio con chiodi e martello, le operazioni da specialisti le lasciamo ad un secondo tempo!

Nella foto sopra ecco come vanno disposti i chiodini che serviranno come punti di collegamento per i vari elementi; le dimensioni indicate non sono tassative, dato che resistenze e condensatori hanno i terminali lunghi, si potrebbero aumentare di un buon 50%, l’importante è rispettare la disposizione indicata, anche la tavoletta della foto può essere sostituita da altra più grande e costituita, per chi lo desidera, da legno pregiato.

L’operazione successiva consiste nella filatura: termine oramai desueto che prevede la disposizione dei fili di collegamento occorrenti da un ancoraggio all’altro; la filatura, nei moderni apparati elettronici è scomparsa quasi del tutto con l’uso dei circuiti stampati, dove le varie piste ramate, realizzate con un processo d’incisione, provvedono al collegamento elettrico dei vari elementi.

La fotografia successiva mostra la disposizione del filo di rame rigido così come occorre per il nostro montaggio, si può utilizzare anche un filo rigido isolato in plastica, volta per volta occorrerà spellare lo spezzone occorrente e provvedere alla saldatura, questa andrà eseguita, su ogni chiodino interessato, appena sotto la testa.

Sarà utile sagomare a forma d'anello aperto il terminale iniziale e quello finale del collegamento affinché, restando il filo in posizione, sia più agevole saldarlo al chiodino.

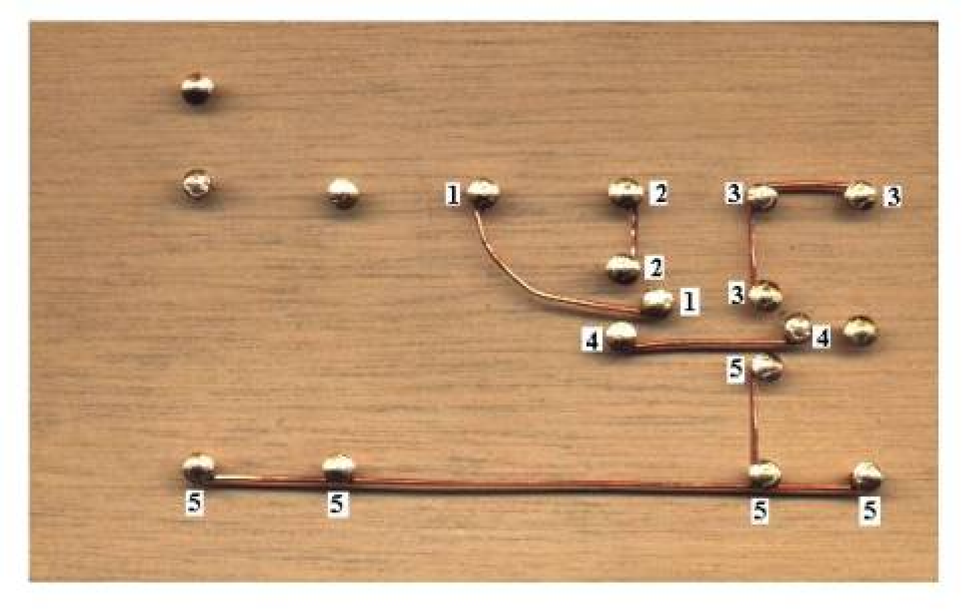

I collegamenti sono stati numerati, sistemare lo spezzone di filo e saldarlo in corrispondenza d'ogni numero.

A questo punto è possibile proseguire inserendo, al loro posto nell'ordine: resistenze, condensatori, bobina, diodo e transistor; naturalmente occorre identificare con assoluta certezza ogni valore per evitare errori che faranno perdere tempo o determinare un insuccesso.

Le spiegazioni successive che a molti parranno superflue sono destinate ai neofiti!

Le resistenze sono cilindriche con i terminali assiali e sono identificate da strisce colorate:

1 Megaohm = marrone-nero-verde-oro (il colore oro determina la tolleranza =5%)

4,7 Kiloohm = giallo-viola-rosso-oro;

2,2 Kiloohm = rosso-rosso-rosso-oro.

I condensatori ceramici hanno forma di pastiglia e riportano scritto sul corpo il loro valore:

470 picoFarad = 471 (le prime due cifre indicano il valore, la terza indica il numero di zeri da aggiungere: 47 + 0 = 470)

220 picoFarad = 221 (22 + 0 = 220)

4,7 nanoFarad = 4700 picoFarad= 472 (per cui: 47 + 2 zeri = 4700 pF).

I condensatori elettrolitici sono cilindretti d'alluminio ricoperti da una guaina plastica con sopra riportato il valore in microFarad e la tensione di lavoro, per il loro montaggio occorre rispettare la polarità indicata sullo schema, generalmente, sul corpo è segnato il terminale negativo:

1 microFarad = 1uF/25V (la tensione indicata può essere anche più elevata: 63 o 100 V);

47 microFarad = 47uF/25V (può essere anche da 100uF, il valore non ha influenza sul risultato).

Il diodo è costituito da un cilindretto in vetro con terminali assiali ed il catodo (o terminale positivo) è contraddistinto mediante striscia colorata, generalmente nera; anche se nello schema elettrico, e sulle fotografie è riportato un preciso senso d'inserzione, per il funzionamento della radio non è necessario rispettare tale senso, vorrà dire che anziché rivelare la semi-onda positiva del segnale di BF, sarà rivelata quella negativa.

I transistor sono gli unici ad avere tre terminali, quello centrale è la base, gli altri due ai lati sono il collettore e l'emettitore, per il loro montaggio (da eseguire come ultima operazione) sarà bene osservare attentamente figure e fotografie, infatti, un errore, in questo caso, comporta l'immediata distruzione del semiconduttore.

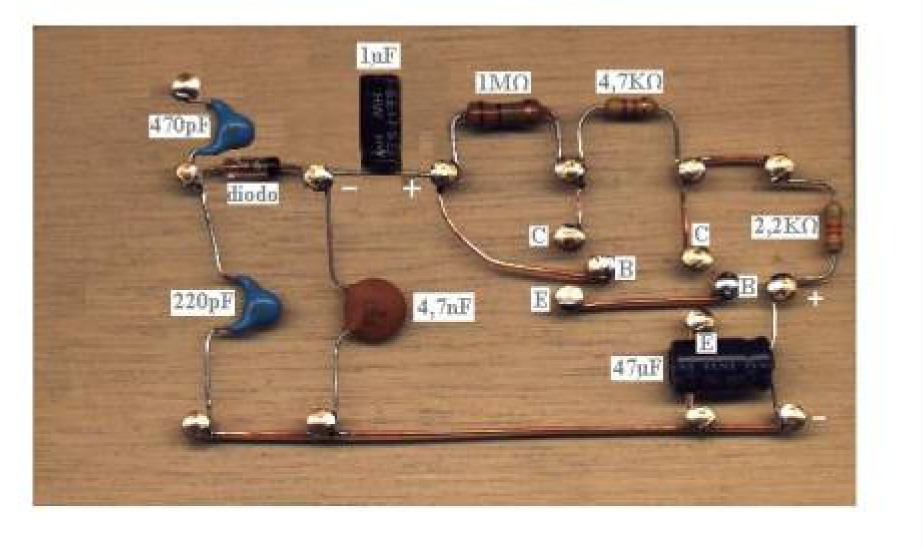

La fotografia qui sopra mostra come deve essere eseguito il montaggio, ci sono ancora da inserire: la bobina, i transistor ed i fili di collegamento alla pila ed all'auricolare; per proseguire è necessario costruire la bobina, questa sarà avvolta, usando del filo smaltato da 0,3 mm di diametro, su di un tubetto di PVC lungo 6 cm e di 16mm di diametro.

Iniziare ad avvolgere, a circa 1,5 cm dal bordo, fermando il capo iniziale con un pezzetto di nastro adesivo, non è necessario contare le spire, basta avvolgere il filo in modo da ottenere circa 3 cm di avvolgimento, il numero di spire sarà, all'incirca, di 60/70. Le spire devono essere affiancate molto bene senza lasciare alcuno spazio fra l'una e l'altra.

Terminato l'avvolgimento, perché questo non si svolga, sarà bene rivestirlo con un giro di nastro adesivo trasparente, i terminali vanno raschiati dallo smalto e stagnati, la bobina trova spazio sul lato sinistro della basetta e va collegata ai capi del condensatore da 220 pF, due gocce di colla saranno sufficienti a tenerla in posizione, anche la colla a caldo o il silicone sono una buona soluzione.

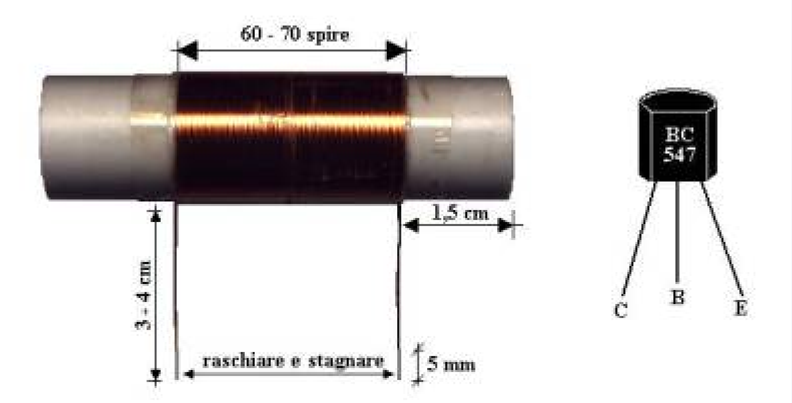

A questo punto si salderanno, al loro posto direttamente sulla testa dei chiodini, i due transistor (occorre fare attenzione per non sbagliare nell'identificare i terminali), i fili rosso-nero provvisti delle pinzette a coccodrillo per il collegamento alla pila, l'auricolare o la cuffia ai capi della resistenza da 2,2 Kohm, il filo d'antenna al terminale del condensatore da 470 pF, il filo per il collegamento di terra al negativo.

Il pezzetto di ferrite va infilato su un'estremità di un tubetto di PVC (lungo 10 cm circa e di diametro leggermente inferiore al diametro interno del tubetto della bobina) e fermato in posizione con colla, ovatta, nastro adesivo, elastici, collante a caldo, resina epossida, silicone,stecchini o quanto altro la fantasia può suggerire.

La radio è ora pronta, se non sono stati commessi errori clamorosi funzionerà certamente, collegare il filo rosso al positivo della pila ed il filo nero al negativo, infilare il tubetto col nucleo di ferrite all'interno della bobina e muoverlo lentamente avanti ed indietro, si troverà la posizione in cui la ricezione della stazione locale avrà un suo massimo; facilmente, a causa della scarsa selettività che contraddistingue le radio a cristallo, si udranno anche altre emittenti sovrapposte, soprattutto se le stazioni hanno frequenze di trasmissione vicine o multiple.

Certamente, nelle ore notturne, si riscontrerà un maggior affollamento di stazioni, sarà necessario muovere il tubetto, in maniera micrometrica, per l'esatta sintonizzazione.

In alcuni momenti, specialmente se ci si trova nelle vicinanze del ripetitore radio, si potrebbe udire un ronzio molto fastidioso, in tal caso provare a scollegare il filo di terra, il fenomeno dovrebbe scomparire o attenuarsi.

Chi vuole, può scollegare l'auricolare e collegare, sempre ai capi della resistenza da 2,2 Kohm, l'ingresso di un amplificatore per l'ascolto in altoparlante, ancora, sul sito è spiegato come utilizzare una normale cuffia stereo, più confortevole ad usarsi, dove occorre utilizzare, come in questo caso, un trasduttore ad alta impedenza.

Nell'ultima immagine qui sotto ecco come appare la radio terminata e funzionante, naturalmente, se s'incontrassero dei problemi, relativi al montaggio ed all'uso, inviare un e-mail al seguente indirizzo:

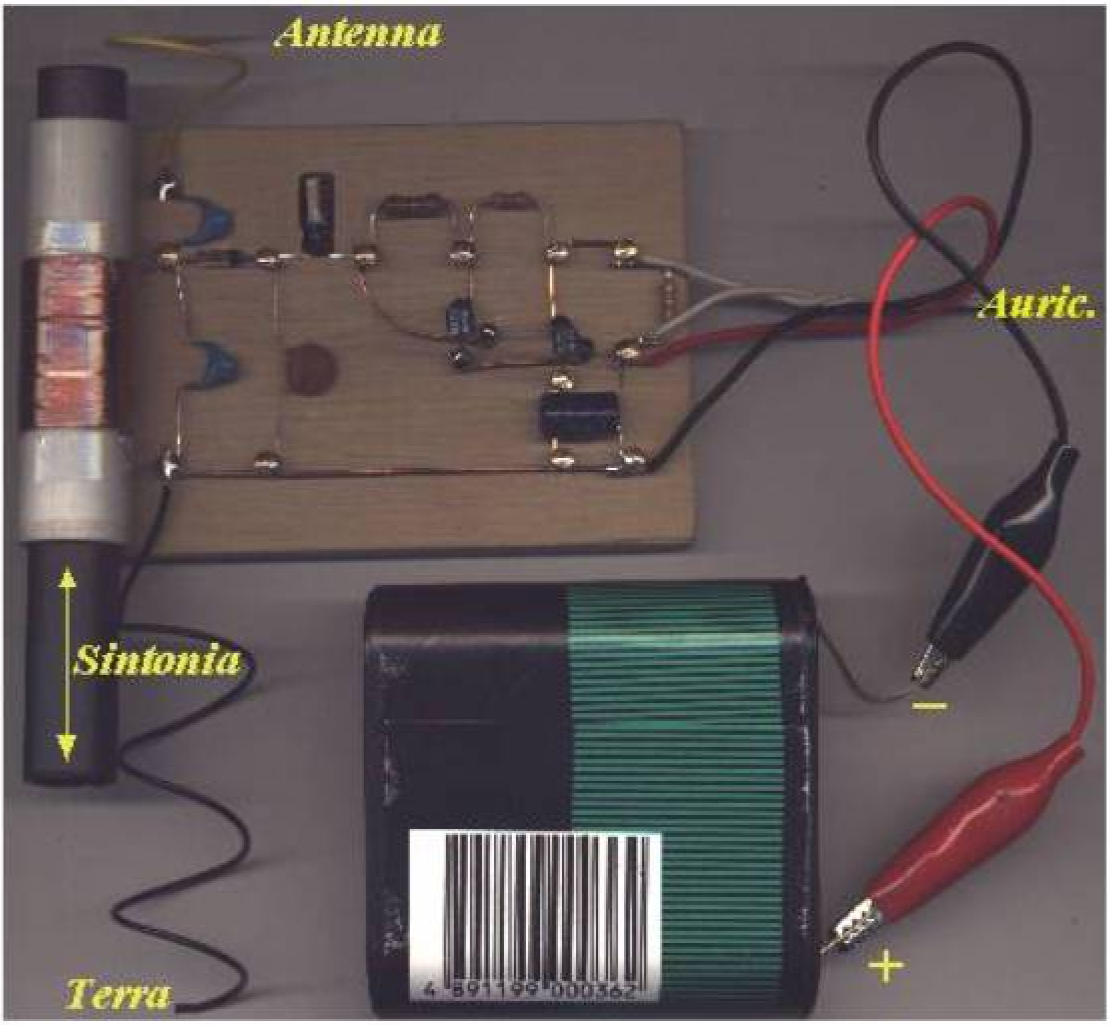

Newnes Radio and RF Engineering Pocket Book

Newnes Radio and RF Engineering Pocket Book

3rd edition

Steve Winder
Joe Carr

Newnes

OXFORD AMSTERDAM BOSTON LONDON NEW YORK PARIS
SAN DIEGO SAN FRANCISCO SINGAPORE SYDNEY TOKYO

Newnes
An imprint of Elsevier Science
Linacre House, Jordan Hill, Oxford OX2 8DP
225 Wildwood Avenue, Woburn, MA 01801-2041

First published 1994
Reprinted 2000, 2001
Second edition 2000
Third edition 2002

Copyright © 1994, 2000, 2002, Steve Winder. All rights reserved

The right of Steve Winder to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988

No part of this publication may be reproduced in any material form (including photocopying or storing in any medium by electronic means and whether or not transiently or incidentally to some other use of this publication) without the written permission of the copyright holder except in accordance with the provisions of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London, England W1T 4LP. Applications for the copyright holder's written permission to reproduce any part of this publication should be addressed to the publisher

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN 0 7506 5608 5

For information on all Newnes publications
visit our website at www.newnespress.com

Typeset by Laserwords Private Limited, Chennai, India.
Printed and bound in Great Britain

Contents

Preface to second edition xi

Preface to third edition xiii

1 Propagation of radio waves 1
  - 1.1 Frequency and wavelength 1
  - 1.2 The radio frequency spectrum 1
  - 1.3 The isotropic radiator 3
  - 1.4 Formation of radio waves 3
  - 1.5 Behaviour of radio waves 7
  - 1.6 Methods of propagation 13
  - 1.7 Other propagation topics 18
  - References 24

2 The decibel scale 25
  - 2.1 Decibels and the logarithmic scale 25
  - 2.2 Decibels referred to absolute values 25

3 Transmission lines 35
  - 3.1 General considerations 35
  - 3.2 Impedance matching 35
  - 3.3 Base band lines 36
  - 3.4 Balanced line hybrids 36
  - 3.5 Radio frequency lines 37
  - 3.6 Waveguides 45
  - 3.7 Other transmission line considerations 47
  - References 51

4 Antennas 52
  - 4.1 Antenna characteristics 52
  - 4.2 Antenna types 56
  - 4.3 VHF and UHF antennas 60
  - 4.4 Microwave antennas 69
  - 4.5 Loop antennas 73
  - References 78

5 Resonant circuits 79
  - 5.1 Series and parallel tuned circuits 79
  - 5.2 *Q* factor 81
  - 5.3 Coupled (band-pass) resonant circuits 81
  - References 84

6 Oscillators 85
- 6.1 Oscillator requirements 85
- 6.2 Tunable oscillators 85
- 6.3 Quartz crystal oscillators 87
- 6.4 Frequency synthesizers 89
- 6.5 Caesium and rubidium frequency standards 93
- References 94

7 Piezo-electric devices 95
- 7.1 Piezo-electric effect 95
- 7.2 Quartz crystal characteristics 97
- 7.3 Specifying quartz crystals 101
- 7.4 Filters 102
- 7.5 SAW filters and resonators 105
- References 109

8 Bandwidth requirements and modulation 110
- 8.1 Bandwidth of signals at base band 110
- 8.2 Modulation 112
- 8.3 Analogue modulation 113
- 8.4 Digital modulation 123
- 8.5 Spread spectrum transmission 129
- References 131

9 Frequency planning 132
- 9.1 International and regional planning 132
- 9.2 National planning 132
- 9.3 Designations of radio emissions 134
- 9.4 Bandwidth and frequency designations 135
- 9.5 General frequency allocations 135
- 9.6 Classes of radio stations 139
- 9.7 Radio wavebands 142
- Reference 142

10 Radio equipment 143
- 10.1 Transmitters 143
- 10.2 Receivers 148
- 10.3 Programmable equipment 157
- References 158

11 Microwave communication 159
- 11.1 Microwave usage 159
- 11.2 Propagation 159

11.3 *K* factor 161
11.4 Fresnel zones, reflections and multi-path fading 161
11.5 Performance criteria for analogue and digital links 164
11.6 Terminology 165
11.7 Link planning 165
11.8 Example of microwave link plan 165
Reference 166

12 Information privacy and encryption 167
12.1 Encryption principles 167
12.2 Speech encryption 168
12.3 Data encryption 169
12.4 Code division multiple access (CDMA) or spread spectrum 172
12.5 Classification of security 172
References 172

13 Multiplexing 173
13.1 Frequency division multiplex 173
13.2 Time division multiplex (TDM) 174
13.3 Code division multiple access (CDMA) 177
Reference 178

14 Speech digitization and synthesis 179
14.1 Pulse amplitude modulation 179
14.2 Pulse code modulation 179
14.3 ADPCM codecs 181
14.4 The G728 low delay CELP codec 181
14.5 The GSM codec 182
References 182

15 VHF and UHF mobile communication 183
15.1 Operating procedures 183
15.2 Control of base stations 186
15.3 Common base station (CBS) operation 186
15.4 Wide area coverage 187

16 Signalling 194
16.1 Sub-audio signalling 194
16.2 In-band tone and digital signalling 195
16.3 Digital signalling 197
16.4 Standard PSTN tones 198
References 199

17 Channel occupancy, availability and trunking 200
17.1 Channel occupancy and availability 200
17.2 Trunking 201
17.3 In-band interrupted scan (IBIS) trunking 203
17.4 Trunking to MPT 1327 specification 203
References 204

18 Mobile radio systems 205
18.1 Paging 205
18.2 Cordless telephones 206
18.3 Trunked radio 207
18.4 Analogue cellular radio-telephone networks 208
18.5 Global system mobile 209
18.6 Other digital mobile systems 211
18.7 Private mobile radio (PMR) 213
18.8 UK CB radio 213
References 213

19 Base station site management 214
19.1 Base station objectives 214
19.2 Site ownership or accommodation rental? 214
19.3 Choice of site 214
19.4 Masts and towers 215
19.5 Installation of electronic equipment 216
19.6 Earthing and protection against lightning 217
19.7 Erection of antennas 219
19.8 Interference 221
19.9 Antenna multi-coupling 225
19.10 Emergency power supplies 226
19.11 Approval and certification 227
References 227

20 Instrumentation 229
20.1 Accuracy, resolution and stability 229
20.2 Audio instruments 230
20.3 Radio frequency instruments 231
References 235

21 Batteries 236
21.1 Cell characteristics 236
21.2 Non-rechargeable, primary batteries 238
21.3 Rechargeable batteries 242

22 Satellite communications 246
22.1 Earth orbits 246
22.2 Communications by satellite link 248
22.3 Proposed satellite television formats 248
22.4 Global positioning system (GPS) 252
References 255

23 Connectors and interfaces 256
23.1 Audio and video connectors 256
23.2 Co-axial connector 258
23.3 Interfaces 268
Reference 280

24 Broadcasting 281
24.1 Standard frequency and time transmissions 281
24.2 Standard frequency formats 283
24.3 UK broadcasting bands 284
24.4 BBC VHF test tone transmissions 284
24.5 Engineering information about broadcast services 287
24.6 Characteristics of UHF terrestrial television systems 288
24.7 Terrestrial television channels 291
24.8 Terrestrial television aerial dimensions 294
24.9 AM broadcast station classes (USA) 295
24.10 FM broadcast frequencies and channel numbers (USA) 296
24.11 US television channel assignments 299
24.12 License-free bands 301
24.13 Calculating radio antenna great circle bearings 302

25 Abbreviations and symbols 307
25.1 Abbreviations 307
25.2 Letter symbols by unit name 313
25.3 Electric quantities 321

26 Miscellaneous data 323
26.1 Fundamental constants 323
26.2 Electrical relationships 323
26.3 Dimensions of physical properties 324
26.4 Fundamental units 324
26.5 Greek alphabet 325

26.6 Standard units 325
26.7 Decimal multipliers 327
26.8 Useful formulae 327
26.9 Colour codes 334
Index 337

Preface to second edition

This edition of the *Newnes Radio and RF Engineer's Pocket Book* is something special. It is a compendium of information of use to engineers and technologists who are engaged in radio and RF engineering. It has been updated to reflect the changing interests of those communities, and reflects a view of the technology like no other. It is packed with information!

This whole series of books is rather amazing with regard to the range and quality of the information they provide, and this book is no different. It covers topics as diverse as circuit symbols and the abbreviations used for transistors, as well as more complex things as satellite communications and television channels for multiple countries in the English speaking world. It is a truly amazing work.

We hope that you will refer to this book frequently, and will enjoy it as much as we did in preparing it.

John Davies
Joseph J. Carr

Acknowledgements

I gratefully acknowledge the ready assistance offered by the following organizations: Andrew Ltd, Aspen Electronics Ltd, BBC, British Telecommunications plc, Farnell Instruments Ltd, Independent Television Authority, International Quartz Devices Ltd, Jaybeam Ltd, MACOM Greenpar Ltd, Marconi Instruments Ltd, Panorama Antennas Ltd, Radiocommunications Agency, the Radio Authority, RTT Systems Ltd. A special thanks goes to my wife Dorothy for once again putting up with my months of seclusion during the book's preparation.

Preface to third edition

This, the third edition of the *Newnes Radio and RF Engineering Pocket Book* has been prepared with a tinge of sadness. Joe Carr, who edited the second edition, has died since the last edition was published. Although I did not know Joe personally, his prolific writing over recent years has impressed me. His was a hard act to follow.

I have updated this book to be more international. Thus the long tables giving details of British television transmitters have been removed (they are available on the Web). Details of the European E1 multiplexing system have been supplemented by a description of the US and Japanese T1 system. There are many more general updates included throughout.

Steve Winder

1 Propagation of radio waves

1.1 Frequency and wavelength

There is a fixed relationship between the frequency and the wavelength, which is the distance between identical points on two adjacent waves (*Figure 1.1*), of any type of wave: sound (pressure), electromagnetic (radio) and light. The type of wave and the speed at which the wavefront travels through the medium determines the relationship. The speed of propagation is slower in higher density media.

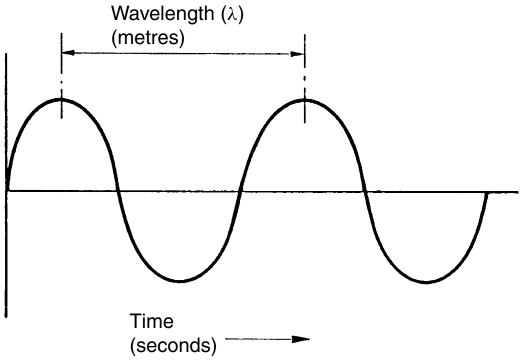

Figure 1.1 Frequency and wavelength

Sound waves travel more slowly than radio and light waves which, in free space, travel at the same speed, approximately 3×10^8 metres per second, and the relationship between the frequency and wavelength of a radio wave is given by:

$$\lambda = \frac{3 \times 10^8}{f} \text{ metres}$$

where λ is the wavelength and f is the frequency in hertz (Hz).

1.2 The radio frequency spectrum

The electromagnetic wave spectrum is shown in *Figure 1.2*: the part usable for radio communication ranges from below 10 kHz to over 100 GHz.

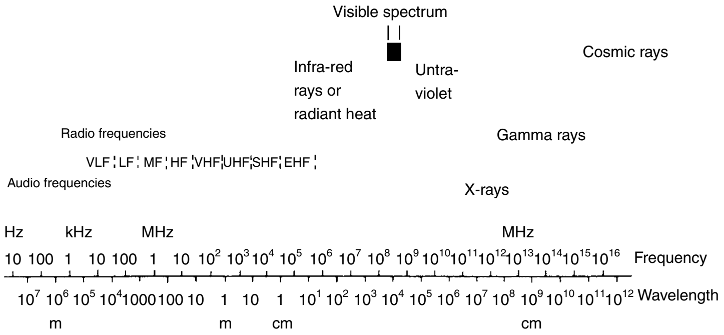
Visible spectrum
Infra-red rays or radiant heat
Untra-violet
Cosmic rays
Radio frequencies
Gamma rays
VLF ¦ LF ¦ MF ¦ HF ¦VHF¦UHF¦SHF¦ EHF ¦
Audio frequencies
X-rays
Hz
kHz
MHz
MHz
10 100 1 10 100 1 10 10^2 10^3 10^4 10^5 10^6 10^7 10^8 10^9 10^10 10^11 10^12 10^13 10^14 10^15 10^16 Frequency
10^7 10^6 10^5 10^4 1000 100 10 1 10 1 10^1 10^2 10^3 10^4 10^5 10^6 10^7 10^8 10^9 10^10 10^11 10^12 Wavelength
m
m
cm
cm

Figure 1.2 The electromagnetic wave spectrum

The radio spectrum is divided into bands and the designation of the bands, their principal use and method of propagation is shown in *Table 1.1*. Waves of different frequencies behave differently and this, along with the amount of spectrum available in terms of radio communication channels in each band, governs their use.

Table 1.1 Use of radio frequencies

Frequency band	*Designation, use and propagation*
3–30 kHz	Very low frequency (VLF). Worldwide and long distance communications. Navigation. Submarine communications. Surface wave.
30–300 kHz	Low frequency (LF). Long distance communications, time and frequency standard stations, long-wave broadcasting. Ground wave.
300–3000 kHz	Medium frequency (MF) or medium wave (MW). Medium-wave local and regional broadcasting. Marine communications. Ground wave.
3–30 MHz	High frequency (HF). 'Short-wave' bands. Long distance communications and short-wave broadcasting. Ionospheric sky wave.
30–300 MHz	Very high frequency (VHF). Short range and mobile communications, television and FM broadcasting. Sound broadcasting. Space wave.
300–3000 MHz	Ultra high frequency (UHF). Short range and mobile communications. Television broadcasting. Point-to-point links. Space wave. Note: The usual practice in the USA is to designate 300–1000 MHz as 'UHF' and above 1000 MHz as 'microwaves'.
3–30 GHz	Microwave or super high frequency (SHF). Point-to-point links, radar, satellite communications. Space wave.
Above 30 GHz	Extra high frequency (EHF). Inter-satellite and micro-cellular radio-telephone. Space wave.

1.3 The isotropic radiator

A starting point for considering the propagation of radio- or lightwaves is the isotropic radiator, an imaginary point source radiating equally in all directions in free space. Such a radiator placed at the centre of a sphere illuminates equally the complete surface of the sphere. As the surface area of a sphere is given by $4\pi r^2$ where r is the radius of the sphere, the brilliance of illumination at any point on the surface varies inversely with the distance from the radiator. In radio terms, the power density at distance from the source is given by:

$$P_\text{d} = \frac{P_\text{t}}{4\pi r^2}$$

where P_t = transmitted power.

1.4 Formation of radio waves

Radio waves are electromagnetic. They contain both electric and magnetic fields at right angles to each other and also at right angles to the direction of propagation. An alternating current flowing in a conductor produces an alternating magnetic field surrounding it and an alternating voltage gradient – an electric field – along the length of the conductor. The fields combine to radiate from the conductor as in *Figure 1.3*.

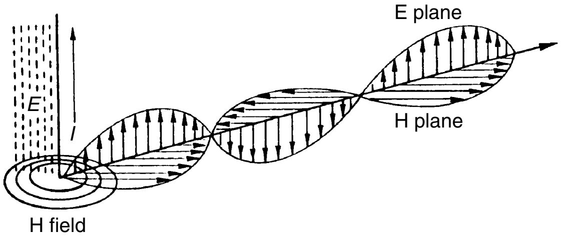

Figure 1.3 Formation of electromagnetic wave

The plane of the electric field is referred to as the E plane and that of the magnetic field as the H plane. The two fields are equivalent to the voltage and current in a wired circuit. They are measured in similar terms, volts per metre and amperes per metre, and the medium through which they propagate possesses an impedance. Where $E = ZI$ in a wired circuit, for an electromagnetic wave:

$$E = ZH$$

where
E = the RMS value of the electric field strength, V/metre
H = the RMS value of the magnetic field strength, A/metre
Z = characteristic impedance of the medium, ohms

The voltage is that which the wave, passing at the speed of light, would induce in a conductor one metre long.

The characteristic impedance of the medium depends on its permeability (equivalent of inductance) and permittivity (equivalent of capacitance). Taking the accepted figures for free space as:

$\mu = 4\pi \times 10^{-7}$ henrys (H) per metre (permeability) and

$\varepsilon = 1/36\pi \times 10^{9}$ farads (F) per metre (permittivity)

then the impedance of free space, Z, is given by:

$$\sqrt{\frac{\mu}{\varepsilon}} = 120\pi = 377\,\text{ohms}$$

The relationship between power, voltage and impedance is also the same for electromagnetic waves as for electrical circuits, $W = E^2/Z$.

The simplest practical radiator is the elementary doublet formed by opening out the ends of a pair of wires. For theoretical considerations the length of the radiating portions of the wires is made very short in relation to the wavelength of the applied current to ensure uniform current distribution throughout their length. For practical applications the length of the radiating elements is one half-wavelength ($\lambda/2$) and the doublet then becomes a dipole antenna (*Figure 1.4*).

When radiation occurs from a doublet the wave is polarized. The electric field lies along the length of the radiator (the E plane) and the magnetic field (the H plane) at right angles to it. If the E plane is vertical, the radiated field is said to vertically polarized. Reference to the E and H planes avoids confusion when discussing the polarization of an antenna.

Unlike the isotropic radiator, the dipole possesses directivity, concentrating the energy in the H plane at the expense of the E plane. It effectively provides a power gain in the direction of the H plane

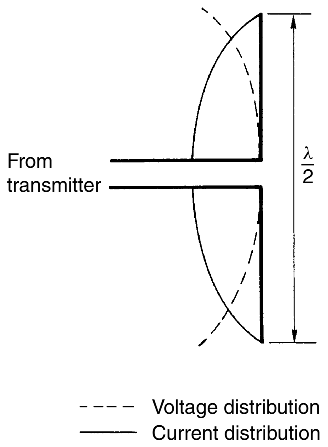
From
transmitter
$\frac{\lambda}{2}$
- - - - Voltage distribution
—— Current distribution

Figure 1.4 Doublet (dipole) antenna

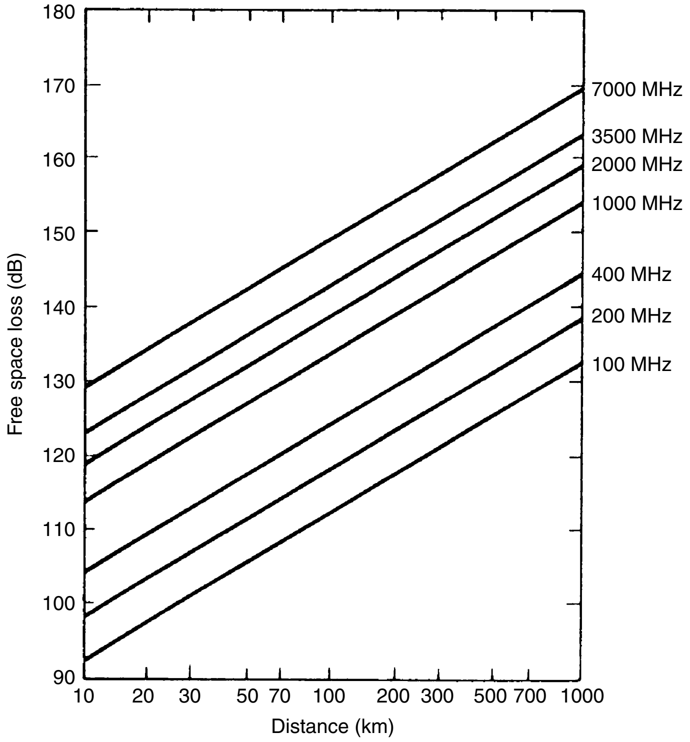

Figure 1.5 Free space loss vs. distance and frequency

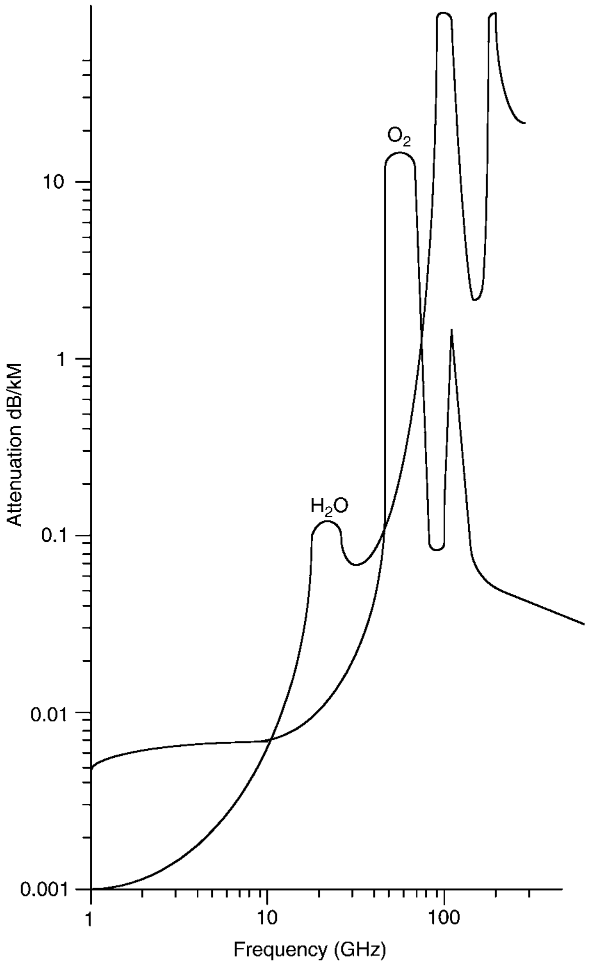

Figure 1.6 Major loss of microwave communications and radar systems due to atmospheric attenuation

compared with an isotropic radiator. This gain is 1.6 times or 2.15 dBi (dBi means dB relative to an isotropic radiator).

For a direct ray the power transfer between transmitting and receiving isotropic radiators is inversely proportional to the distance between them *in wavelengths*. The free space power loss is given by:

$$\text{Free space loss, dB} = 10\log_{10}\frac{(4\pi d)^2}{\lambda^2}$$

where d and λ are in metres, or:

$$\text{Free space loss (dB)} = 32.4 + 20 \times \log_{10} d + 20 \times \log_{10} f$$

where d = distance in km and f = frequency in MHz.

The free space power loss, therefore, increases as the square of the distance and the frequency. Examples are shown in *Figure 1.5*.

With practical antennas, the power gains of the transmitting and receiving antennas, in dBi, must be subtracted from the free space loss calculated as above. Alternatively, the loss may be calculated by:

$$\text{Free space loss (dB)} = 10 \log_{10} \left[\frac{(4\pi d)^2}{\lambda^2} \times \frac{1}{G_t \times G_r} \right]$$

where G_t and G_r are the respective actual gains, not in dB, of the transmitting and receiving antennas.

A major loss in microwave communications and radar systems is atmospheric attenuation (see *Figure 1.6*). The attenuation (in decibels per kilometre (dB/km)) is a function of frequency, with especial problems showing up at 22 GHz and 64 GHz. These spikes are caused by *water vapour* and *atmospheric oxygen* absorption of microwave energy, respectively. Operation of any microwave frequency requires consideration of atmospheric losses, but operation near the two principal spike frequencies poses special problems. At 22 GHz, for example, an additional 1 dB/km of loss must be calculated for the system.

1.5 Behaviour of radio waves

1.5.1 Physical effects

The physical properties of the medium through which a wave travels, and objects in or close to its path, affect the wave in various ways.

Absorption

In the atmosphere absorption occurs and energy is lost in heating the air molecules. Absorption caused by this is minimal at frequencies below about 10 GHz but absorption by foliage, particularly when wet, is severe at VHF and above.

Waves travelling along the earth's surface create currents in the earth causing ground absorption which increases with frequency. A horizontally polarized surface wave suffers more ground absorption than a vertically polarized wave because of the 'short-circuiting' by the ground of the electric field. Attenuation at a given frequency is least for propagation over water and greatest over dry ground for a vertically polarized wave.

Refraction and its effect on the radio horizon

As radio waves travel more slowly in dense media and the densest part of the atmosphere is normally the lowest, the upper parts of a wave usually travel faster than the lower. This refraction (*Figure 1.7*) has the effect of bending the wave to follow the curvature of the earth and progressively tilting the wavefront until eventually the wave becomes horizontally polarized and short-circuited by the earth's conductivity.

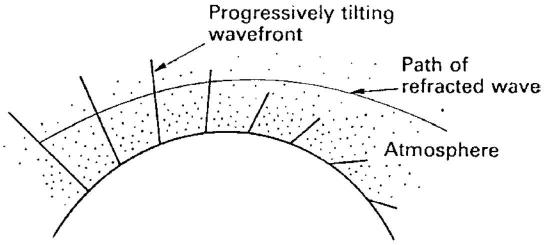

Figure 1.7 Effects of refraction

Waves travelling above the earth's surface (space waves) are usually refracted downwards, effectively increasing the radio horizon to greater than the visual.

The refractive index of the atmosphere is referred to as the K factor; a K factor of 1 indicates zero refraction. Most of the time K is positive at 1.33 and the wave is bent to follow the earth's curvature. The radio horizon is then 4/3 times the visual. However, the density of the atmosphere varies from time to time and in different parts of the world. Density inversions where higher density air is above a region of low density may also occur. Under these conditions the K factor is negative and radio waves are bent away from the earth's surface and are either lost or ducting occurs. A K factor of 0.7 is the worst expected case.

Ducting occurs when a wave becomes trapped between layers of differing density only to be returned at a great distance from its source, possibly creating interference.

Radio horizon distance at VHF/UHF

The radio horizon at VHF/UHF and up is approximately 15% further than the optical horizon. Several equations are used in calculating the distance. If D^- is the distance to the radio horizon, and H is the antenna height, then:

$$D = k\sqrt{H}$$

- When D is in statute miles (5280 feet) and H in feet, then $k = 1.42$.
- When D is in nautical miles (6000 feet) and H in feet, then $k = 1.23$.
- When D is in kilometres and H is in metres, then $k = 4.12$.

Repeating the calculation for the receiving station and adding the results gives the total path length.

Diffraction

When a wave passes over on the edge of an obstacle some of its energy is bent in the direction of the obstacle to provide a signal in what would otherwise be a shadow. The bending is most severe when the wave passes over a sharp edge (*Figure 1.8*).

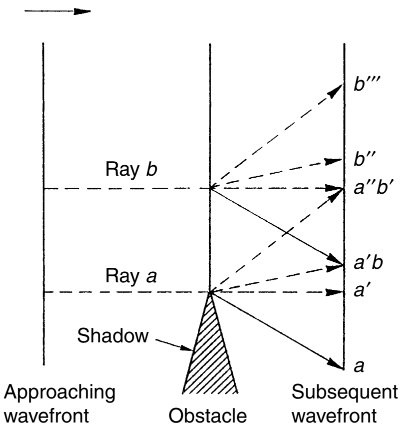

Figure 1.8 Effects of diffraction

As with light waves, the subsequent wavefront consists of wavelets produced from an infinite number of points on the wavefront, rays *a* and *b* in *Figure 1.8* (Huygens' principle). This produces a pattern of interfering waves of alternate addition and subtraction.

Reflection

Radio waves are reflected from surfaces lying in and along their path and also, effectively, from ionized layers in the ionosphere – although most of the reflections from the ionized layers are actually the products of refraction. The strength of truly reflected signals increases with frequency, and the conductivity and smoothness of the reflecting surface.

Multi-path propagation

Reflection, refraction and diffraction may provide signals in what would otherwise be areas of no signal, but they also produce interference.

Reflected – or diffracted – signals may arrive at the receiver in any phase relationship with the direct ray and with each other. The relative phasing of the signals depends on the differing lengths of their paths and the nature of the reflection.

When the direct and reflected rays have followed paths differing by an odd number of half-wavelengths they could be expected to arrive at the receiver in anti-phase with a cancelling effect. However, in the reflection process a further phase change normally takes place. If the reflecting surface had infinite conductivity, no losses would occur in the reflection, and the reflected wave would have exactly the same or opposite phase as the incident wave depending on the polarization in relation to the reflecting surface. In practice, the reflected wave is of smaller amplitude than the incident, and the phase relationships are also changed. The factors affecting the phasing are complex but most frequently, in practical situations, approximately 180° phase change occurs on reflection, so that reflected waves travelling an odd number of half-wavelengths arrive in phase with the direct wave while those travelling an even number arrive anti-phase.

As conditions in the path between transmitter and receiver change so does the strength and path length of reflected signals. This means that a receiver may be subjected to signal variations of almost twice the mean level and practically zero, giving rise to severe fading. This type of fading is frequency selective and occurs on troposcatter systems and in the mobile environment where it is more severe at higher frequencies. A mobile receiver travelling through an urban area can receive rapid signal fluctuations caused by additions and cancellations of the direct and reflected signals at half-wavelength intervals. Fading due to the multi-path environment is often referred to as Rayleigh fading and its effect is shown in *Figure 1.9*. Rayleigh fading, which can cause short signal dropouts, imposes severe restraints on mobile data transmission.

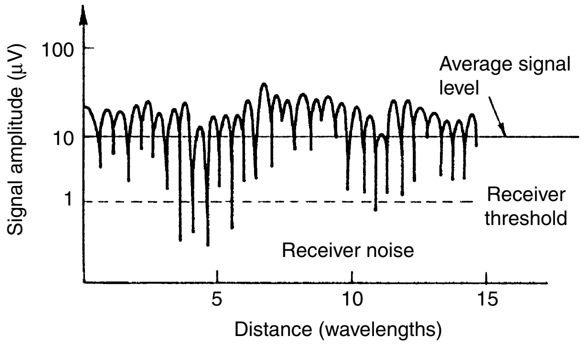

Figure 1.9 Multi-path fading

Noise

The quality of radio signals is not only degraded by the propagation losses: natural or manmade electrical noise is added to them, reducing their intelligibility.

Atmospheric noise includes static from thunderstorms which, unless very close, affects frequencies below about 30 MHz and noise from space is apparent at frequencies between about 8 MHz to 1.5 GHz.

A type of noise with which radio engineers are continually concerned is thermal. Every resistor produces noise spread across the whole frequency spectrum. Its magnitude depends upon the ohmic value of the resistor, its temperature and the bandwidth of the following circuits. The noise voltage produced by a resistor is given by:

$$E_n = \sqrt{4kTBR}$$

where
E_n = noise voltage, V(RMS)
k = Boltzmann's constant
= 1.38×10^{-23} joules/kelvin
T = temperature in degrees K
B = bandwidth of measurement, Hz
R = resistance in ohms

An antenna possesses resistance and its thermal noise, plus that of a receiver input circuit, is a limiting factor to receiver performance.

Noise is produced in every electronic component. Shot noise – it sounds like falling lead shot – caused by the random arrival of electrons at, say, the collector of a transistor, and the random division of electrons at junctions in devices, add to this noise.

Doppler effect

Doppler effect is an apparent shift of the transmitted frequency which occurs when either the receiver or transmitter is moving. It becomes significant in mobile radio applications towards the higher end of the UHF band and on digitally modulated systems.

When a mobile receiver travels directly towards the transmitter each successive cycle of the wave has less distance to travel before reaching the receiving antenna and, effectively, the received frequency is raised. If the mobile travels away from the transmitter, each successive cycle has a greater distance to travel and the frequency is lowered. The variation in frequency depends on the frequency of the wave, its propagation velocity and the velocity of the vehicle containing the receiver. In the situation where the velocity of the vehicle is small compared with the velocity of light, the frequency shift when moving directly towards, or away from, the transmitter is given to sufficient accuracy for most purposes by:

$$f_d = \frac{V}{C} f_t$$

where
f_d = frequency shift, Hz
f_t = transmitted frequency, Hz
V = velocity of vehicle, m/s
C = velocity of light, m/s

Examples are:

- 100 km/hr at 450 MHz, frequency shift = 41.6 Hz
- 100 km/hr at 1.8 GHz – personal communication network (PCN) frequencies – frequency shift = 166.5 Hz
- Train at 250 km/hr at 900 MHz – a requirement for the GSM pan-European radio-telephone – frequency shift = 208 Hz

When the vehicle is travelling at an angle to the transmitter the frequency shift is reduced. It is calculated as above and the result multiplied by the cosine of the angle of travel from the direct approach (*Figure 1.10*).

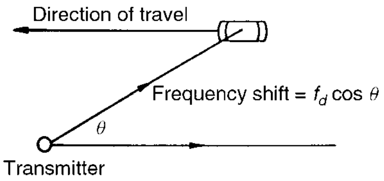

Direction of travel
Frequency shift = $f_d \cos\theta$
θ
Transmitter

Figure 1.10 Doppler frequency shift and angle to transmitter

In a radar situation Doppler effect occurs on the path to the target and also to the reflected signal so that the above formula is modified to:

$$f_{\mathrm{d}} = \frac{2V}{C} f_{\mathrm{t}}$$

where f_{d} is now the total frequency shift.

1.6 Methods of propagation

The effects of all of the above phenomena vary with frequency and are used in the selection of frequencies for specific purposes. The behaviour of waves of different frequencies gives rise to the principal types of wave propagation.

Ground wave propagation

Waves in the bands from very low frequencies (VLF, 3–30 kHz), low frequencies (LF, 30–300 kHz) and medium frequencies (MF, 300–3000 kHz) travel close to the earth's surface: the ground wave (*Figure 1.11*). Transmissions using the ground wave must be vertically polarized to avoid the conductivity of the earth short-circuiting the electric field.

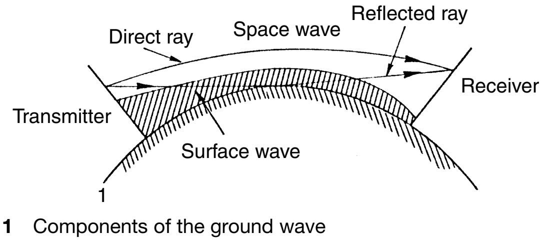

Direct ray
Space wave
Reflected ray
Receiver
Transmitter
Surface wave
1

Figure 1.11 Components of the ground wave

The ground wave consists of a surface wave and a space wave. The surface wave travels along the earth's surface, and is attenuated by ground absorption and the tilting of the wavefront due to diffraction.

The losses increase with frequency and thus VLF radio stations have a greater range than MF stations. The attenuation is partially offset by the replacement of energy from the upper part of the wave refracted by the atmosphere.

The calculation of the field strength of the surface wave at a distance from a transmitter is complex and affected by several variables. Under plane earth conditions and when the distance is sufficiently short that the earth's curvature can be neglected the field intensity is given by:

$$E_{\mathrm{su}} = \frac{2E_0}{d}A$$

where

E_{su} = field intensity in same units as E_0

d = distance in same units of distance as used in E_0

A = a factor calculated from the earth losses, taking frequency, dielectric constant and conductivity into account

E_0 = the free space field produced at unit distance from the transmitter. (With a short (compared with $\lambda/4$) vertical aerial, $2E_0 = 300\sqrt{P}$ mV/m at 1 km where P is the radiated power in kW.) (Terman, 1943)

For a radiated power of 1 kW and ground of average dampness, the distance at which a field of 1 mV/m will exist is given in *Table 1.2*.

Table 1.2 Distance at which a field of 1 mV/m will exist for a radiated power of 1 kW and ground of average dampness

Frequency	*Range (km)*
100 kHz	200
1 MHz	60
10 MHz	6
100 MHz	1.5

The direct and reflected components of the ground wave produce multi-path propagation and variations in received single strength will arise depending on the different path lengths taken by the two components. When the transmitting and receiving antennas are at ground level the components of the space wave cancel each other and the surface wave is dominant. When the antennas are elevated, the space wave becomes increasingly strong and a height is eventually reached where the surface wave has a negligible effect on the received signal strength.

Sky wave propagation

High frequency (HF) waves between 3 MHz and 30 MHz are effectively reflected by ionized layers in the ionosphere producing the sky wave. Medium frequency waves may also be reflected, but less reliably.

The ionosphere contains several layers of ionized air at varying altitudes (*Figure 1.12*). The heights and density of the layers vary diurnally, seasonally and with the incidence of sunspot activity. The E and F_2 layers are semi-permanent while the F_1 layer is normally only present during daytime.

Radio waves radiated at a high angle and reflected by these layers return to earth at a distance from the transmitter. The HF reflection

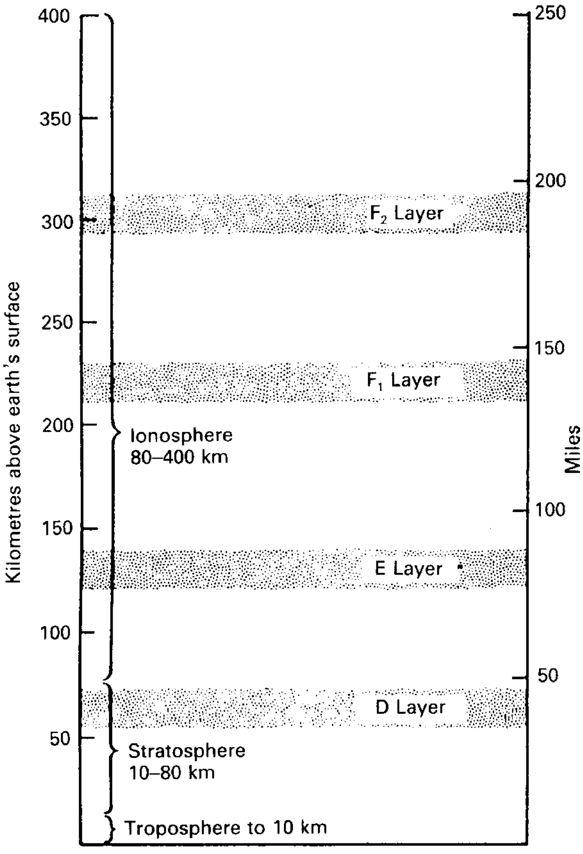

400
350
300
250
200
150
100
50
Kilometres above earth's surface
250
200
150
100
50
Miles
F_2 Layer
F_1 Layer
E Layer
D Layer
Ionosphere
80–400 km
Stratosphere
10–80 km
Troposphere to 10 km

Figure 1.12 The ionosphere

process is in reality one of refraction in layers possessing a greater free electron density than at heights above or below them. The speed of propagation is slowed on entering a layer and the wave is bent and, if of a suitable frequency and angle of incidence, returned to earth (*Figure 1.13*). The terms used are defined as follows:

- *Virtual height*. The height at which a true reflection of the incident wave would have occurred (*Figure 1.13*).
- *Critical frequency* (f_c). The highest frequency that would be returned to earth in a wave directed vertically at the layer.
- *Maximum usable frequency* (muf). The highest frequency that will be returned to earth for a given angle of incidence. If the angle of incidence to the normal is θ, the muf $= f_c / \cos\theta$.
- *Skip distance*. The minimum distance from the transmitter, along the surface of the earth, at which a wave above the critical frequency will be returned to earth (*Figure 1.12*). Depending on the frequency, the ground wave will exist at a short distance from the transmitter.
- *Sporadic E-layer reflections*. Reflections from the E layer at frequencies higher than those which would normally be returned to earth. They appear to be reflections from electron clouds having sharp boundaries and drifting through the layer. As the name implies the reflections are irregular but occur mostly in summer and at night.

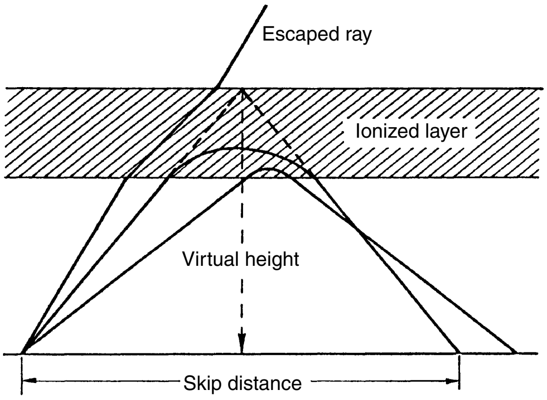

Figure 1.13 Sky wave propagation

Space wave propagation

The space wave travels through the troposphere (the atmosphere below the ionosphere) between the transmitter and the receiver. It contains

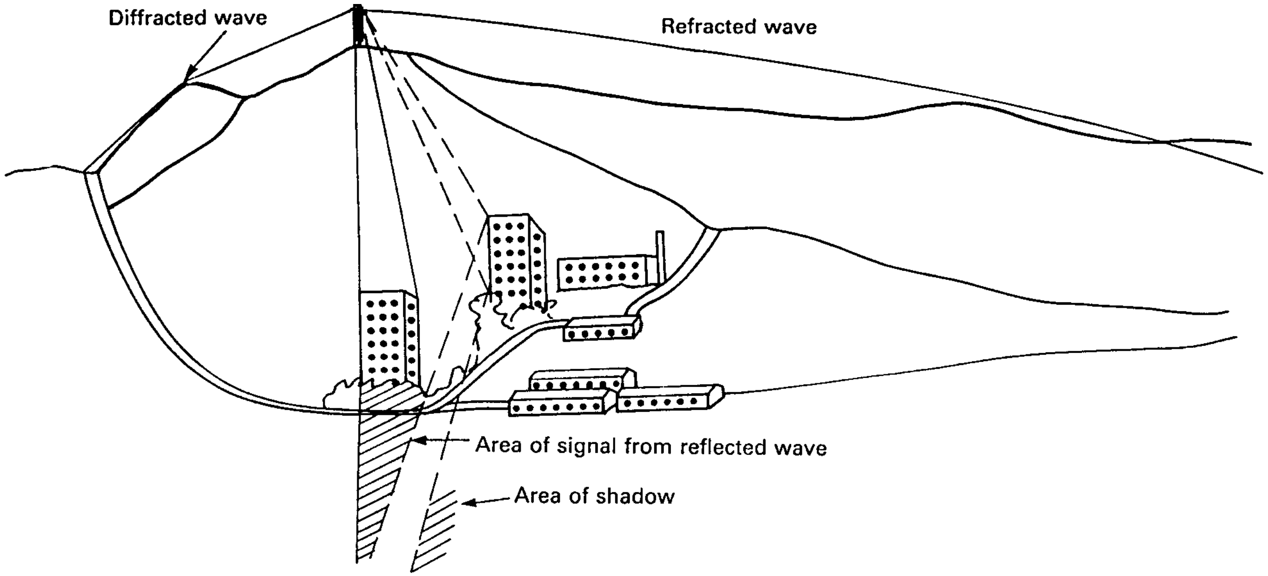
Diffracted wave
Refracted wave
Area of signal from reflected wave
Area of shadow

Figure 1.14 Pictorial representation of radio coverage from a base station

both direct and reflected components (see *Figure 1.11*), and is affected by refraction and diffraction. The importance of these effects varies with frequency, the nature of the terrain and of objects close to the direct path, and the type of communication, e.g. data. Apart from medium-wave broadcasting, space waves are used mainly for communications using frequencies of VHF and upwards.

The range of space waves is the radio horizon. However, places of little or no signal can arise in the lee of radio obstacles. Fortunately, they may be filled with either reflected or diffracted signals as depicted in *Figure 1.14*.

Tropospheric scatter

The tropospheric, or forward, scatter effect provides reliable, over the horizon, communication between fixed points at bands of ultra and super high frequencies. Usable bands are around 900, 2000 and 5000 MHz and path lengths of 300 to 500 km are typical.

The mechanism is not known with certainty but reflections from discontinuities in the dielectric constant of the atmosphere and scattering of the wave by clouds of particles are possibilities. It is an inefficient process, the scattered power being -60 to -90 dB relative to the incident power, so high transmitter powers are necessary. The phenomenon is regularly present but is subject to two types of fading. One occurs slowly and is due to variations of atmospheric conditions. The other is a form of Rayleigh fading and is rapid, deep and frequency selective. It is due to the scattering occurring at different points over a volume in the atmosphere producing multipath propagation conditions.

Troposcatter technique uses directional transmitting and receiving antennas aimed so that their beams intercept in the troposphere at the mid-distance point. To overcome the fading, diversity reception using multiple antennas spaced over 30 wavelengths apart is common.

1.7 Other propagation topics

Communications in the VHF through microwave regions normally takes place on a ‘line-of-sight’ basis where the radio horizon defines the limit of sight. In practice, however, the situation is not so neat and simple. There is a transition region between the HF and VHF where long distance ionospheric ‘skip’ occurs only occasionally. This effect is seen above 25 MHz, and is quite pronounced in the 50 MHz region. Sometimes the region behaves like line-of-sight VHF, and at others like HF shortwave.

1.7.1 Scatter

There are a number of scatter modes of propagation. These modes can extend the radio horizon a considerable amount. Where the radio horizon might be a few tens of kilometres, underscatter modes permit very much longer propagation. For example, a local FM broadcaster at 100 MHz might have a service area of about 40 miles, and might be heard 180 miles away during the summer months when *Sporadic-E* propagation occurs. One summer, a television station in Halifax, Nova Scotia, Canada, was routinely viewable in Washington, DC in the United States during the early morning hours for nearly a week.

Sporadic-E is believed to occur when a small region of the atmosphere becomes differentially ionized, and thereby becomes a species of 'radio mirror'. Ionospheric scatter propagation occurs when clouds of ions exist in the atmosphere. These clouds can exist in both the ionosphere and the troposphere, although the tropospheric model is more reliable for communications. A signal impinging this region may be scattered towards other terrestrial sites which may be a great distance away. The specific distance depends on the geometry of the scenario.

There are at least three different modes of scatter from ionized clouds: *back scatter, side scatter*, and *forward scatter*. The back scatter mode is a bit like radar, in that signal is returned back to the transmitter site, or in regions close to the transmitter. Forward scatter occurs when the reflected signal continues in the same azimuthal direction (with respect to the transmitter), but is redirected toward the Earth's surface. Side scatter is similar to forward scatter, but the azimuthal direction might change.

Unfortunately, there are often multiple reflections from the ionized cloud, and these are shown as 'multiple scatter' in *Figure 1.15*. When these reflections are able to reach the receiving site, the result is a rapid, fluttery fading that can be of quite profound depths.

Meteor scatter is used for communication in high latitude regions. When a meteor enters the Earth's atmosphere it leaves an ionized trail of air behind it. This trail might be thousands of kilometres long, but is very short lived. Radio signals impinging the tubular metre ion trail are reflected back towards Earth. If the density of meteors in the critical region is high, then more or less continuous communications can be achieved. This phenomenon is noted in the low VHF between 50 and about 150 MHz. It can easily be observed on the FM broadcast band if the receiver is tuned to distant stations that are barely audible. If the geometry of the scenario is right, abrupt but short-lived peaks in the signal strength will be noted.

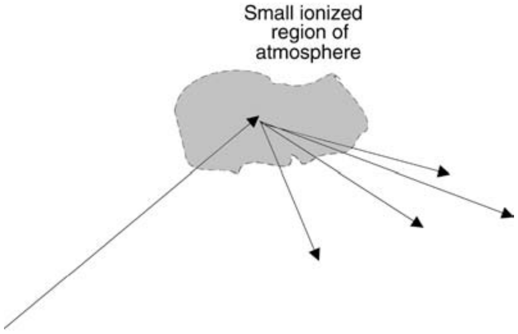

Figure 1.15 Multiple scatter

1.7.2 Refraction modes

Refraction is the mechanism for most tropospheric propagation phenomena. The dielectric properties of the air, which are set mostly by the moisture content, are a primary factor in tropospheric refraction. Refraction occurs in both light or radio wave systems when the wave passes between mediums of differing density. Under that situation, the wave path will bend an amount proportional to the difference in density of the two regions.

The general situation is typically found at UHF and microwave frequencies. Because air density normally decreases with altitude, the top of a beam of radio waves typically travels slightly faster than the lower portion of the beam. As a result, those signals refract a small amount. Such propagation provides slightly longer surface distances than are normally expected from calculating the distance to the radio horizon. This phenomenon is called *simple refraction*, and is described by the *K factor*.

Super refraction

A special case of refraction called super refraction occurs in areas of the world where warmed land air flows out over a cooler sea (*Figure 1.16*). Examples of such areas are deserts that are adjacent to a large body of water: the Gulf of Aden, the southern Mediterranean, and the Pacific Ocean off the coast of Baja, California. Frequent VHF/UHF/microwave communications up to 200 miles are reported in such areas, and up to 600 miles have reportedly been observed.

Ducting

Another form of refraction phenomenon is weather related. Called *ducting*, this form of propagation is actually a special case of super

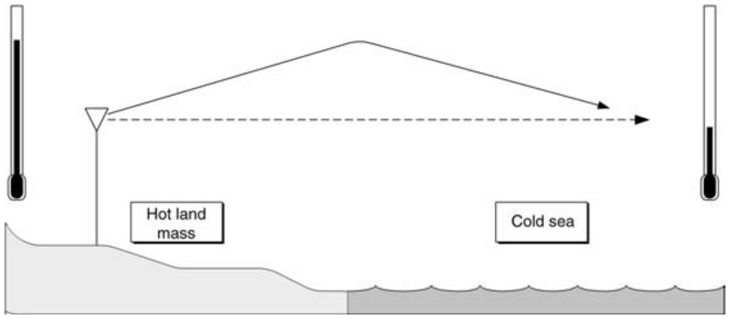

Figure 1.16 An example of super refraction

refraction. Evaporation of sea water causes temperature inversion regions to form in the atmosphere. That is, layered air masses in which the air temperature is greater than in the layers below it (note: air temperature normally decreases with altitude, but at the boundary with an inversion region, it begins to increase). The inversion layer forms a 'duct' that acts similarly to a waveguide. Ducting allows long distance communications from lower VHF through microwave frequencies; with 50 MHz being a lower practical limit, and 10 GHz being an ill-defined upper limit. Airborne operators of radio, radar, and other electronic equipment can sometimes note ducting at even higher microwave frequencies.

Antenna placement is critical for ducting propagation. Both the receiving and transmitting antennas must be either: (a) physically inside the duct (as in airborne cases), or (b) able to propagate at an angle such that the signal gets trapped inside the duct. The latter is a function of antenna radiation angle. Distances up to 2500 miles or so are possible through ducting.

Certain paths, where frequent ducting occurs, have been identified: in the United States, the Great Lakes region to the southeastern Atlantic seaboard; Newfoundland to the Canary Islands; across the Gulf of Mexico from Florida to Texas; Newfoundland to the Carolinas; California to Hawaii; and Ascension Island to Brazil.

Subrefraction

Another refractive condition is noted in the polar regions, where colder air from the land mass flows out over warmer seas (*Figure 1.17*). Called *subrefraction*, this phenomena bends EM waves away from the Earth's surface – thereby reducing the radio horizon by about 30 to 40%.

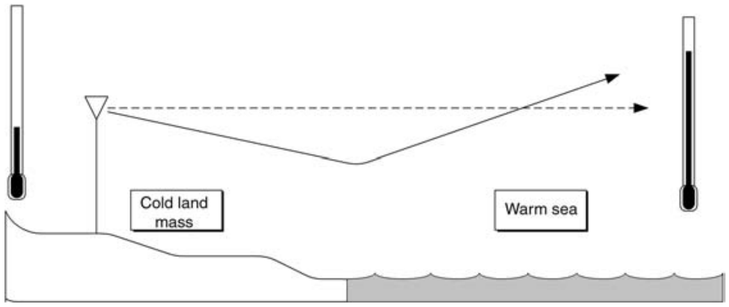

Figure 1.17 An example of subrefraction

All tropospheric propagation that depends upon air-mass temperatures and humidity shows diurnal (i.e. over the course of the day) variation caused by the local rising and setting of the sun. Distant signals may vary 20 dB in strength over a 24-hour period. These tropospheric phenomena explain how TV, FM broadcast, and other VHF signals can propagate great distances, especially along seacoast paths, sometimes weak and sometimes nonexistent.

1.7.3 Great circle paths

A great circle is the shortest line between two points on the surface of a sphere, such that it lays on a plane through the Earth's centre and includes the two points. When translated to 'radiospeak', a great circle is the shortest path on the surface of the Earth between two points. Navigators and radio operators use the great circle for similar, but different, reasons. Navigators use it in order to get from here to there, and radio operators use it to get a transmission path from here to there.

The heading of a directional antenna is normally aimed at the receiving station along its great circle path. Unfortunately, many people do not understand the concept well enough, for they typically aim the antenna in the wrong direction. For example, Washington, DC in the USA is on approximately the same latitude as Lisbon, Portugal. If you fly due east, you will have dinner in Lisbon, right? Wrong. If you head due east from Washington, DC, across the Atlantic, the first landfall would be West Africa, somewhere between Ghana and Angola. Why? Because the great circle bearing 90 degrees takes us far south. The geometry of spheres, not flat planes, governs the case.

Long path versus short path

The Earth is a sphere (or, more precisely, an 'oblique spheroid'), so from any given point to any other point there are two great circle paths: the long path (major arc) and the short path (minor arc). In general, the best reception occurs along the short path. In addition, short path propagation is more nearly 'textbook', compared with long path reception. However, there are times when long path is better, or is the only path that will deliver a signal to a specific location from the geographic location in question.

Grey line propagation

The Grey line is the twilight zone between the night and daytime halves of the earth. This zone is also called the *planetary terminator*. It varies up to +23 degrees either side of the north–south longitudinal lines, depending on the season of the year (it runs directly north–south only at the vernal and autumnal equinoxes). The D-layer of the ionosphere absorbs signals in the HF region. This layer disappears almost completely at night, but it builds up during the day. Along the grey line, the D-layer is rapidly decaying west of the line, and has not quite built up east of the line.

Brief periods of abnormal propagation occur along the grey line. Stations on either side of the line can be heard from regions, and at distances, that would otherwise be impossible on any given frequency. For this reason, radio operators often prefer to listen at dawn and dusk for this effect.

1.7.4 Scatter propagation modes

Auroral propagation

The auroral effect produces a luminescence in the upper atmosphere resulting from bursts of particles released from the sun 18 to 48 hours earlier. The light emitted is called the northern lights and the southern lights. The ionized regions of the atmosphere that create the lights form a radio reflection shield, especially at VHF and above, although 15 to 20 MHz effects are known. Auroral propagation effects are normally seen in the higher latitudes, although listeners in the southern tier of states in the USA and Europe are often treated to the reception of signals from the north being reflected from auroral clouds. Similar effects exist in the southern hemisphere.

Non-reciprocal direction

If you listen to the 40 metre (7–7.3 MHz) amateur radio band receiver on the East Coast of the United States, you will sometimes hear European stations – especially in the late afternoon. But when the

US amateur tries to work those European stations there is no reply whatsoever. The Europeans are unable to hear the US stations. This propagation anomaly causes the radio wave to travel different paths dependent on which direction it travels, i.e. an east–west signal is not necessarily the reciprocal of a west–east signal. This anomaly can occur when a radio signal travels through a heavily ionized medium in the presence of a magnetic field, which is exactly the situation when the signal travels through the ionosphere in the presence of the Earth's magnetic field.

References

Belcher, R. *et al*. (1989). *Newnes Mobile Radio Servicing Handbook*. Butterworth-Heinemann, Oxford.

Kennedy, G. (1977). *Electronic Communications Systems*. McGraw-Hill Kogashuka, Tokyo.

Terman, F.E. (1943). *Radio Engineers' Handbook*. McGraw-Hill, London.

2 The decibel scale

2.1 Decibels and the logarithmic scale

The range of powers, voltages and currents encountered in radio engineering is too wide to be expressed on linear scale. Consequently, a logarithmic scale based on the decibel (dB, one tenth of a bel) is used. The decibel does not specify a magnitude of a power, voltage or current but a ratio between two values of them. Gains and losses in circuits or radio paths are expressed in decibels.

The ratio between two powers is given by:

$$\text{Gain or loss, dB} = 10\log_{10}\frac{P_1}{P_2}$$

where P_1 and P_2 are the two powers.

As the power in a circuit varies with the square of the voltage or current, the logarithm of the ratio of these quantities must be multiplied by twenty instead of ten. To be accurate the two quantities under comparison must operate in identical impedances:

$$\text{Gain or loss, dB} = 20\log_{10}\frac{V_1}{V_2}$$

To avoid misunderstandings, it must be realized that a ratio of 6 dB is 6 dB regardless of whether it is power, voltage or current that is referred to: if it is power, the ratio for 6 dB is four times; if it is voltage or current, the ratio is two times (*Table 2.1*).

2.2 Decibels referred to absolute values

While the decibel scale expresses ratios only, if a reference value is added to the statement as a suffix it can be used to refer to absolute values. For example, a loss of 10 dB means a reduction in power to a level equal to one tenth of the original and if the statement is −10 dBm the level referred to is 1/10 of a milliwatt. Commonly used suffixes and, where applicable, their absolute reference levels are as follows. *Table 2.2* shows the relative levels in decibels at 50 ohms impedance.

Table 2.1 The decibel figures are in the centre column: figures to the left represent decibel loss, and those to the right decibel gain. The voltage and current figures are given on the assumption that there is no change in impedance

Voltage or current ratio	*Power ratio*	*dB* ← − + →	*Voltage or current ratio*	*Power ratio*
1.000	1.000	0	1.000	1.000
0.989	0.977	0.1	1.012	1.023
0.977	0.955	0.2	1.023	1.047
0.966	0.933	0.3	1.035	1.072
0.955	0.912	0.4	1.047	1.096
0.944	0.891	0.5	1.059	1.122
0.933	0.871	0.6	1.072	1.148
0.912	0.832	0.8	1.096	1.202
0.891	0.794	1.0	1.122	1.259
0.841	0.708	1.5	1.189	1.413
0.794	0.631	2.0	1.259	1.585
0.750	0.562	2.5	1.334	1.778
0.708	0.501	3.0	1.413	1.995
0.668	0.447	3.5	1.496	2.239
0.631	0.398	4.0	1.585	2.512
0.596	0.355	4.5	1.679	2.818
0.562	0.316	5.0	1.778	3.162
0.501	0.251	6.0	1.995	3.981
0.447	0.200	7.0	2.239	5.012
0.398	0.159	8.0	2.512	6.310
0.355	0.126	9.0	2.818	7.943
0.316	0.100	10	3.162	10.00
0.282	0.0794	11	3.55	12.6
0.251	0.0631	12	3.98	15.9
0.224	0.0501	13	4.47	20.0
0.200	0.0398	14	5.01	25.1
0.178	0.0316	15	5.62	31.6
0.159	0.0251	16	6.31	39.8
0.126	0.0159	18	7.94	63.1
0.100	0.0100	20	10.00	100.0
3.16×10^{-2}	10^{-3}	30	3.16×10	10^{3}
10^{-2}	10^{-4}	40	10^{2}	10^{4}
3.16×10^{-3}	10^{-5}	50	3.16×10^{2}	10^{5}
10^{-3}	10^{-6}	60	10^{3}	10^{6}
3.16×10^{-4}	10^{-7}	70	3.16×10^{3}	10^{7}
10^{-4}	10^{-8}	80	10^{4}	10^{8}
3.16×10^{-5}	10^{-9}	90	3.16×10^{4}	10^{9}
10^{-5}	10^{-10}	100	10^{5}	10^{10}
3.16×10^{-6}	10^{-11}	110	3.16×10^{5}	10^{11}
10^{-6}	10^{-12}	120	10^{6}	10^{12}

Table 2.2 Relative levels in decibels at 50 ohms impedance

dB μV	*Voltage*	*dBV*	*dBm*	*dBW*	*Power*
−20	100 nV	−140	−127	−157	200 aW
−19	115	−139	−126	−156	250
−18	125	−138	−125	−155	315
−17	140	−137	−124	−154	400
−16	160	−136	−123	−153	500
−15	180	−135	−122	−152	630
−14	200	−134	−121	−151	800
−13	225	−133	−120	−150	1 fW
−12	250	−132	−119	−149	1.25
−11	280	−131	−118	−148	1.6
−10	315	−130	−117	−147	2.0
−9	355	−129	−116	−146	2.5
−8	400	−128	−115	−145	3.15
−7	450	−127	−114	−144	4.0
−6	500	−126	−113	−143	5.0
−5	565	−125	−112	−142	6.3
−4	630	−124	−111	−141	8.0
−3	710	−123	−110	−140	10
−2	800	−122	−109	−139	12.5
−1	900	−121	−108	−138	16
0	1 μV	−120	−107	−137	20
1	1.15	−119	−106	−136	25
2	1.25	−118	−105	−135	31.5
3	1.4	−117	−104	−134	40
4	1.6	−116	−103	−133	50
5	1.8	−115	−102	−132	63
6	2.0	−114	−101	−131	80
7	2.25	−113	−100	−130	100
8	2.5	−112	−99	−129	125
9	2.8	−111	−98	−128	160
10	3.15	−110	−97	−127	200
11	3.55	−109	−96	−126	250
12	4.0	−108	−95	−125	315
13	4.5	−107	−94	−124	400
14	5.0	−106	−93	−123	500
15	5.65	−105	−92	−122	630
16	6.3	−104	−91	−121	800
17	7.1	−103	−90	−120	1 pW
18	8.0	−102	−89	−119	1.25
19	9.0	−101	−88	−118	1.6
20	10.0	−100	−87	−117	2.0
21	11.5	−99	−86	−116	2.5
22	12.5	−98	−85	−115	3.15
23	14	−97	−84	−114	4.0
24	16	−96	−83	−113	5.0
25	18	−95	−82	−112	6.3

(*continued overleaf*)

Table 2.2 (*continued*)

dB μV	*Voltage*	*dBV*	*dBm*	*dBW*	*Power*
26	20	−94	−81	−111	8.0
27	22.5	−93	−80	−110	10
28	25	−92	−79	−109	12.5
29	28	−91	−78	−108	16
30	31.5	−90	−77	−107	20
31	35.5	−89	−76	−106	25
32	40	−88	−75	−105	31.5
33	45	−87	−74	−104	40
34	50	−86	−73	−103	50
35	56.5	−85	−72	−102	63
36	63	−84	−71	−101	80
37	71	−83	−70	−100	100
38	80	−82	−69	−99	125
39	90	−81	−68	−98	160
40	100	−80	−67	−97	200
41	115	−79	−66	−96	250
42	125	−78	−65	−95	315
43	140	−77	−64	−94	400
44	160	−76	−63	−93	500
45	180	−75	−62	−92	630
46	200	−74	−61	−91	800
47	225	−73	−60	−90	1 nW
48	250	−72	−59	−89	1.25
49	280	−71	−58	−88	1.6
50	315	−70	−57	−87	2.0
51	355	−69	−56	−86	2.5
52	400	−68	−55	−85	3.15
53	450	−67	−54	−84	4.0
54	500	−66	−53	−83	5.0
55	565	−65	−52	−82	6.3
56	630	−64	−51	−81	8.0
57	710	−63	−50	−80	10
58	800	−62	−49	−79	12.5
59	900	−61	−48	−78	16
60	1 mV	−60	−47	−77	20
61	1.15	−59	−46	−76	25
62	1.25	−58	−45	−75	31.5
63	1.4	−57	−44	−74	40
64	1.6	−56	−43	−73	50
65	1.8	−55	−42	−72	63
66	2.0	−54	−41	−71	80
67	2.25	−53	−40	−70	100
68	2.5	−52	−39	−69	125
69	2.8	−51	−38	−68	160
70	3.15	−50	−37	−67	200
71	3.55	−49	−36	−66	250
72	4.0	−48	−35	−65	315

Table 2.2 (*continued*)

dB μV	*Voltage*	*dBV*	*dBm*	*dBW*	*Power*
73	4.5	−47	−34	−64	400
74	5.0	−46	−33	−63	500
75	5.65	−45	−32	−62	630
76	6.3	−44	−31	−61	800
77	7.1	−43	−30	−60	1 μW
78	8.0	−42	−29	−59	1.25
79	9.0	−41	−28	−58	1.6
80	10 mV	−40	−27	−57	2.0
81	11.5	−39	−26	−56	2.5
82	12.5	−38	−25	−55	3.15
83	14	−37	−24	−54	4.0
84	16	−36	−23	−53	5.0
85	18	−35	−22	−52	6.3
86	20	−34	−21	−51	8.0
87	22.5	−33	−20	−50	10
88	25	−32	−19	−49	12.5
89	28	−31	−18	−48	16
90	31.5	−30	−17	−47	20
91	35.5	−29	−16	−46	25
92	40	−28	−15	−45	31.5
93	45	−27	−14	−44	40
94	50	−26	−13	−43	50
95	56.5	−25	−12	−42	63
96	63	−24	−11	−41	80
97	71	−23	−10	−40	100
98	80	−22	−9	−39	125
99	90	−21	−8	−38	160
100	100	−20	−7	−37	200
101	115	−19	−6	−36	250
102	125	−18	−5	−35	315
103	140	−17	−4	−34	400
104	160	−16	−3	−33	500
105	180	−15	−2	−32	630
106	200	−14	−1	−31	800
107	225	−13	0	−30	1 mW
108	250	−12	1	−29	1.25
109	280	−11	2	−28	1.6
110	315	−10	3	−27	2.0
111	355	−9	4	−26	2.5
112	400	−8	5	−25	3.15
113	450	−7	6	−24	4.0
114	500	−6	7	−23	5.0
115	565	−5	8	−22	6.3
116	630	−4	9	−21	8.0
117	710	−3	10	−20	10
118	800	−2	11	−19	12.5

(*continued overleaf*)

Table 2.2 (*continued*)

dB μV	*Voltage*	*dBV*	*dBm*	*dBW*	*Power*
119	900	−1	12	−18	16
120	1 V	0	13	−17	20
121	1.15	1	14	−16	25
122	1.25	2	15	−15	31.5
123	1.4	3	16	−14	40
124	1.6	4	17	−13	50
125	1.8	5	18	−12	63
126	2.0	6	19	−11	80
127	2.25	7	20	−10	100
128	2.5	8	21	−9	125
129	2.8	9	22	−8	160
130	3.15	10	23	−7	200
131	3.55	11	24	−6	250
132	4.0	12	25	−5	315
133	4.5	13	26	−4	400
134	5.0	14	27	−3	500
135	5.65	15	28	−2	630
136	6.3	16	29	−1	800
137	7.1	17	30	0	1 W
138	8.0	18	31	1	1.25
139	9.0	19	32	2	1.6
140	10	20	33	3	2.0
141	11.5	21	34	4	2.5
142	12.5	22	35	5	3.15
143	14	23	36	6	4.0
144	16	24	37	7	5.0
145	18	25	38	8	6.3
146	20	26	39	9	8.0
147	22.5	27	40	10	10
148	25	28	41	11	12.5
149	28	29	42	12	16
150	31.5	30	43	13	20
151	35.5	31	44	14	25
152	40	32	45	15	31.5
153	45	33	46	16	40
154	50	34	47	17	50
155	56.5	35	48	18	63
156	63	36	49	19	80
157	71	37	50	20	100
158	80	38	51	21	125
159	90	39	52	22	160
160	100	40	53	23	200
161	115	41	54	24	250
162	125	42	55	25	315
163	140	43	56	26	400
164	160	44	57	27	500
165	180	45	58	28	630

Table 2.2 (*continued*)

dB μV	*Voltage*	*dBV*	*dBm*	*dBW*	*Power*
166	200	46	59	29	800
167	225	47	60	30	1 kW
168	250	48	61	31	1.25
169	280	49	62	32	1.6
170	315	50	63	33	2.0
171	355	51	64	34	2.5
172	400	52	65	35	3.15
173	450	53	66	36	4.0
174	500	54	67	37	5.0
175	565	55	68	38	6.3
176	630	56	69	39	8.0
177	710	57	70	40	10
178	800	58	71	41	12.5
179	900	59	72	42	16
180	1 kV	60	73	43	20

Decibel glossary

dBa	stands for dB 'adjusted'. This is a weighted circuit noise power referred to −85 dBm, which is 0 dBa. (Historically measured with a noise meter at the receiving end of a line. The meter is calibrated on a 1000 Hz tone such that 1 mW (0 dBm) gives a reading of +85 dBm. If the 1 mW is spread over the band 300–3400 Hz as random white noise, the meter will read +82 dBa.)
dBa0	circuit noise power in dBa referred to, or measured at, a point of zero relative transmission level (0 dBr). (A point of zero relative transmission level is a point arbitrarily established in a transmission circuit. All other levels are stated relative to this point.) It is preferable to convert circuit noise measurement values from dBa to dBa0 as this makes it unnecessary to know or to state the relative transmission level at the point of measurement.
dBd	used for expressing the gain of an antenna referred to a dipole.
dBi	used for expressing the gain of an antenna referred to an isotropic radiator.
dB μV	decibels relative to 1 microvolt.
dbm	decibels relative to 1 milliwatt. The term dBm was originally used for telephone and audio work and, when used in that context, implies an impedance of 600 Ω, the nominal impedance of a telephone line. When it is desired to define a relative transmission level in a circuit, dBr is preferred.

Table 2.3 Binary decibel values

Bits	*Max. value*	*Decibels (dB)*
1	2	6.02
2	4	12.04
3	8	18.06
4	16	24.08
5	32	30.10
6	64	36.12
7	128	42.14
8	256	48.16
9	512	54.19
10	1024	60.21
11	2048	66.23
12	4096	72.25
13	8192	78.27
14	16 384	84.29
15	32 768	90.31
16	65 536	96.33
17	131 072	102.35
18	262 144	108.37
19	524 288	114.39
20	1 048 576	120.41
21	2 097 152	126.43
22	4 194 304	132.45
23	8 388 608	138.47
24	16 777 216	144.49
25	33 554 432	150.51
26	67 108 864	156.54
27	134 217 728	162.56
28	268 435 456	168.58
29	536 870 912	174.60
30	1 073 741 824	180.62
31	2 147 483 648	186.64
32	4 294 967 296	192.66

dBm0	dBm referred to, or measured at, a point of zero transmission level.
dBmp	a unit of noise power in dBm, measured with psophometric weighting. dBmp $= 10\log_{10}$ pWp $- 90 =$ dBa $- 84 =$ dBm $- 2.5$ (for flat noise 300–3400 Hz). pWp = picowatts psophometrically weighted.
dBm0p	the abbreviation for absolute noise power in dBm referred to or measured at a point of zero relative transmission level, psophometrically weighted.
dBr	means dB ‘relative level’. Used to define transmission level at various points in a circuit or system referenced to the zero transmission level point.

dBrn	a weighted circuit noise power unit in dB referenced to 1 pW (−90 dBm) which is 0 dBrn.
dBrnc	weighted noise power in dBrn, measured by a noise measuring set with 'C-message' weighting.
dBrnc0	noise measured in dBrnc referred to zero transmission level point.
dBu	decibels relative to 0.775 V, the voltage developed by 1 mW when applied to 600 Ω. dBu is used in audio work when the impedance is not 600 Ω and no specific impedance is implied.
dbV	decibels relative to 1 volt.
dbW	decibels relative to 1 watt.

Note: To convert dBm to dB μV add 107 (e.g. $-20\,\text{dBm} = -20 + 107 = 87\,\text{dB}\,\mu\text{V}$.

The beauty of decibel notation is that system gains and losses can be computed using addition and subtraction rather than multiplication and division. For example, suppose a system consists of an antenna that delivers a −4.7 dBm signal at its terminals (we conveniently neglect the antenna gain by this ploy). The antenna is connected to a 40 dB low-noise amplifier (A1) at the head end, and then through a 370 metre long coaxial cable to a 20 dB gain amplifier (A2), with a loss (L1) of −48 dB. The amplifier is followed by a bandpass filter with a −2.8 dB insertion loss (L2), and a −10 dB attenuator (L3). How does the signal exist at the end of this cascade chain?

S1	−4.7 dBm
A1	40.0 dB
A2	20.0 dB
L1	−48.0 dB
L2	−2.8 dB
L3	−10.0 dB
Total:	−5.5 dBm

Converting dBm to watts

$$P = \frac{10^{\text{dBm}/10}}{1000}$$

Converting any dB to ratio

Power levels: $\dfrac{P1}{P2} = 10^{\text{dB}/10}$

Voltage levels: $\frac{V1}{V2} = 10^{\mathrm{dB}/20}$

Current levels: $\frac{I1}{I2} = 10^{\mathrm{dB}/20}$

Binary decibel values

Binary numbers are used in computer systems. With the digitization of RF systems it is necessary to understand the decibel values of binary numbers. These binary numbers might be from an analogue-to-digital converter (ADC or A/D) that digitizes the IF amplifier output, or a digital-to-analogue converter (DAC or D/A) used to generate the analogue signal in a direct digital synthesis (DDS) signal generator.

3 Transmission lines

3.1 General considerations

The purpose of any transmission line is to transfer power between a source and a load with the minimum of loss and distortion in either amplitude, frequency or phase angle.

Electrons travel more slowly in conductors than they do in free space and all transmission lines contain distributed components: resistance, inductance and capacitance. Consequently, lines possess an impedance which varies with frequency, and loss and distortion occur. Because the impedance is not constant over a wide frequency band the insertion loss will not be the same for all frequencies and frequency distortion will arise. A wavefront entering a line from a source takes a finite time to travel its length. This transit time, because of the distributed components, also varies with frequency and creates phase distortion.

3.2 Impedance matching

To transfer the maximum power from a generator into a load the impedance of the load and the internal impedance of the generator – and any intervening transmission line – must be equal.

Figure 3.1 illustrates the simplest case of a generator of internal impedance Z_s equal to 5 ohms and producing an e.m.f. of 20 volts.

When loads of varying impedance, Z_1, are connected the output voltage, V (p.d.) and the power in the load, P_1, varies as follows:

$Z_1 = 5\,\Omega \quad I = 20/10 = 2\,\text{A} \quad V = 10\,\text{V}$
$Z_1 = 3\,\Omega \quad I = 2.5\,\text{A} \quad V = 7.5\,\text{V}$
$Z_1 = 8.33\,\Omega \quad I = 1.5\,\text{A} \quad V = 12.5\,\text{V}$

$P_1 = V^2/\text{Z}_1 = 100/5 = 20\,\text{W}$
$P_1 = V^2/\text{Z}_1 = 56.25/3 = 18.75\,\text{W}$
$P_1 = V^2/\text{Z}_1 = 156.25/8.33 = 18.75\,\text{W}$

When dealing with alternating current and when transmission lines, particularly at radio frequencies, are interposed between the source and the load, other factors than the power transfer efficiency must also be considered.

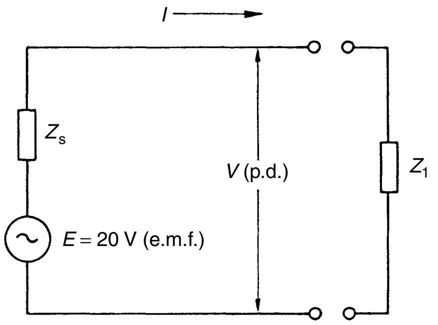

Figure 3.1 Impedance matching

3.3 Base band lines

These are the lines which generally operate at comparatively low frequencies carrying information at base band, e.g. speech, music, video or data. Generally provided by the telecommunications or telephone companies, usually on a rental basis, they are no longer likely to be hard wired, solid copper lines, although these may still be obtainable for lengths below about 10 km within one exchange area. Longer lines will probably be multiplexed, and comprised of radio and optical circuits over part of their length.

Baseband line impedance may vary between 450 Ω and 750 Ω. Nominal impedance is 600 Ω. Most line parameters are specified when measured between 600 Ω non-reactive impedance.

3.4 Balanced line hybrids

Radio transmitters and receivers are often controlled over a two-wire line. To facilitate this a balanced line hybrid circuit, consisting of two transformers connected back to back as in *Figure 3.2*, is inserted between the transmitter and receiver, and the line.

A signal from the receiver audio output is fed to winding L_1 of transformer T_1 which induces voltages across L_2 and L_3. The resultant line current also flows through L_4 and produces a voltage across L_6 which would appear as modulation on the transmitter but for the anti-phase voltage appearing across L_5. To ensure that the voltages cancel exactly a variable resistor, and often a capacitor to equalize the frequency response, is connected between L_2 and L_5.

A signal arriving via the line is applied to the transmitter as modulation; that it is also applied to the receiver poses no problem.

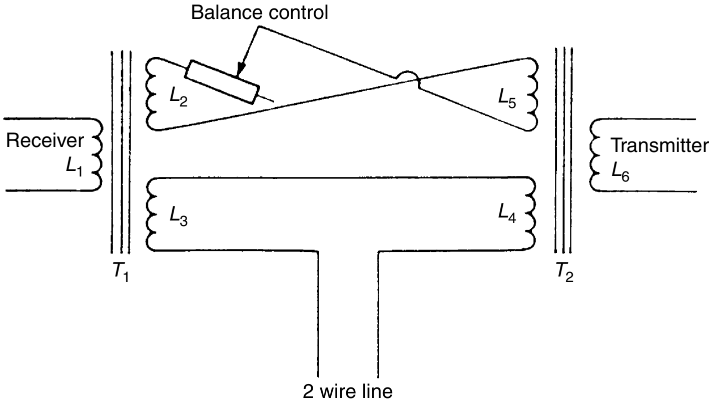

Figure 3.2 Balanced line hybrid

3.5 Radio frequency lines

Radio frequency transmission lines possess similar electrical characteristics to base band lines. However, they may be carrying large powers and the effects of a mismatched load are much more serious than a loss of transferred power. Three types of wire RF line are commonly used: a single wire with ground return for MF and LF broadcast transmission, an open pair of wires at HF and co-axial cable at higher frequencies. Waveguides are used at the higher microwave frequencies. RF lines exhibit an impedance characterized by their type and construction.

3.5.1 Characteristic impedance, Z_0

The physical dimensions of an RF transmission line, the spacing between the conductors, their diameters and the dielectric material between them, determine the characteristic impedance of the line, Z_0, which is calculated for the most commonly used types as follows.

Single wire with ground return (*Figure 3.3(a)*):

$$Z_0 = 138 \log_{10} \frac{4h}{d} \text{ ohms}$$

2-wire balanced, in air (*Figure 3.3(b)*):

$$Z_0 = \frac{276}{\sqrt{\varepsilon_p}} \log_{10} \frac{2s}{d} \text{ ohms}$$

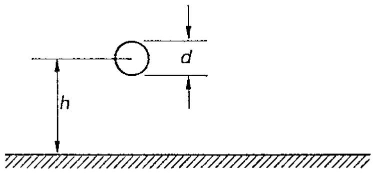

(a) Single wire, ground return line

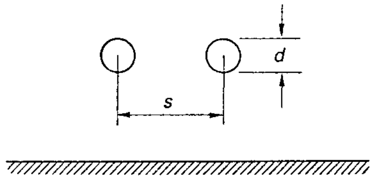

(b) 2-wire balanced air dielectric line

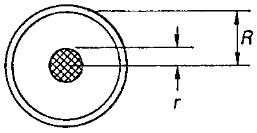

(c) Co-axial cable

Figure 3.3

Co-axial (*Figure 3.3(c)*):

$$Z_0 = \frac{138}{\sqrt{\varepsilon_r}} \log_{10} \frac{R}{r} \text{ ohms}$$

where dimensions are in mm and ε_r = relative permittivity of continuous dielectric.

3.5.2 Insertion loss

The loss in RF cables is quoted in specifications as attenuation in dB per unit length at specific frequencies, the attenuation increasing with frequency. The electrical specifications for cables having a braided outer conductor are given in *Tables 3.1* and *3.2*. Those for the now commonly used foam dielectric, solid outer conductor cables are provided in *Table 3.3*.

Table 3.1 RF cables USA RG and URM series

RG number	Nominal impedance Z_0 (ohms)	Overall diameter (inches)	Velocity factor	Attenuation (dB per 100 ft)					Capacity (pF/ft)	Maximum operating voltage RMS
				1 MHz	10 MHz	100 MHz	1000 MHz	3000 MHz		
RG-6A/U	75.0	0.332	0.659	0.21	0.78	2.9	11.2	21.0	20.0	2700
RG-11A/U	75.0	0.405	0.66	0.18	0.7	2.3	7.8	16.5	20.5	5000
RG-12A/U	75.0	0.475	0.659	0.18	0.66	2.3	8.0	16.5	20.5	4000
RG-16/U	52.0	0.630	0.670	0.1	0.4	1.2	6.7	16.0	29.5	6000
RG-34A/U	75.0	0.630	0.659	0.065	0.29	1.3	6.0	12.5	20.5	5200
RG-34B/U	75.0	0.630	0.66	–	0.3	1.4	5.8	–	21.5	6500
RG-35A/U	75.0	0.945	0.659	0.07	0.235	0.85	3.5	8.60	20.5	10 000
RG-54A/U	58.0	0.250	0.659	0.18	0.74	3.1	11.5	21.5	26.5	3000
RG-55/U	53.5	0.206	0.659	0.36	1.3	4.8	17.0	32.0	28.5	1900
RG-55A/U	50.0	0.216	0.659	0.36	1.3	4.8	17.0	32.0	29.5	1900
RG-58/U	53.5	0.195	0.659	0.33	1.25	4.65	17.5	37.5	28.5	1900

(*continued overleaf*)

Table 3.1 (*continued*)

				Attenuation (dB per 100 ft)						
RG number	*Nominal impedance Z_0 (ohms)*	*Overall diameter (inches)*	*Velocity factor*	*1 MHz*	*10 MHz*	*100 MHz*	*1000 MHz*	*3000 MHz*	*Capacity (pF/ft)*	*Maximum operating voltage RMS*
RG-58C/U	50.0	0.195	0.659	0.42	1.4	4.9	24.0	45.0	30.0	1900
RG-59A/U	75.0	0.242	0.659	0.34	1.10	3.40	12.0	26.0	20.5	2300
RG-59B/U	75	0.242	0.66	–	1.1	3.4	12.0	–	21	2300
RG-62A/U	93.0	0.242	0.84	0.25	0.85	2.70	8.6	18.5	13.5	750
RG-83/U	35.0	0.405	0.66	0.23	0.80	2.8	9.6	24.0	44.0	2000
RG-174A/U	50	0.1	0.66	1.83	3.35	11.0	32.0	64.0	30.0	1400
RG-188A/U	50	0.11	0.66	–	–	10.0	30.0	–	–	1200
[*]RG-213/U	50	0.405	0.66	0.16	0.6	1.9	8.0	–	29.5	5000
[†]RG-218/U	50	0.870	0.66	0.066	0.2	1.0	4.4	–	29.5	11 000
[‡]RG-220/U	50	1.120	0.66	0.04	0.2	0.7	3.6	–	29.5	14 000
URM43	50	0.116	0.66	0.396	1.25	3.96	–	–	30.0	2600
URM70	75	0.128	0.66	0.457	1.46	4.63	–	–	20.5	1800
URM76	50	0.116	0.66	0.457	1.46	4.72	–	–	30.0	2600

[*] Formerly RG8A/U.

[†] Formerly RG17A/U.

[‡] Formerly RG19A/U.

Table 3.2 British UR series

UR number	Nominal impedance Z_0 (ohms)	Overall diameter (inches)	Inner conductor (inches)	Capacity (pF/ft)	Maximum operating voltage RMS	Attenuation (dB per 100 ft) 10 MHz	100 MHz	300 MHz	1000 MHz	Nearest RG equivalent
43	52	0.195	0.032	29	2750	1.3	4.3	8.7	18.1	58/U
57	75	0.405	0.044	20.6	5000	0.6	1.9	3.5	7.1	11A/U
63*	75	0.855	0.175	14	4400	0.15	0.5	0.9	1.7	
67	50	0.405	7/0.029	30	4800	0.6	2.0	3.7	7.5	213/U
74	51	0.870	0.188	30.7	15 000	0.3	1.0	1.9	4.2	218/U
76	51	0.195	19/0.0066	29	1800	1.6	5.3	9.6	22.0	58C/U
77	75	0.870	0.104	20.5	125 000	0.3	1.0	1.9	4.2	164/U
79*	50	0.855	0.265	21	6000	0.16	0.5	0.9	1.8	
83*	50	0.555	0.168	21	2600	0.25	0.8	1.5	2.8	
85*	75	0.555	0.109	14	2600	0.2	0.7	1.3	2.5	
90	75	0.242	0.022	20	2500	1.1	3.5	6.3	12.3	59B/U

All the above cables have solid dielectric with a velocity factor of 0.66 with the exception of those marked with an asterisk, which are helical membrane and have a velocity factor of 0.96.

Table 3.3 Foam dielectric, solid outer conductor co-axial RF cables 50 ohms characteristic impedance

	Superflexible FS-J series		*LDF series*	
Cable type	FSJ1-50A 1/4″	FSJ4-50B 1/2″	LDF2-50 3/8″	LDF4-50A 1/2″
Min. bending radius, in. (mm)	1.0 (25)	1.25 (32)	3.75 (95)	5.0 (125)
Propagation velocity, %	84	81	88	88
Max. operating frequency, GHz	20.4	10.2	13.5	8.8
Att. dB/100 ft (dB/100 m)				
50 MHz	1.27 (4.17)	0.732 (2.40)	0.75 (2.46)	0.479 (1.57)
100 MHz	1.81 (5.94)	1.05 (3.44)	1.05 (3.44)	0.684 (2.24)
400 MHz	3.70 (12.1)	2.18 (7.14)	2.16 (7.09)	1.42 (4.66)
1000 MHz	6.00 (19.7)	3.58 (11.7)	3.50 (11.5)	2.34 (7.68)
5000 MHz	14.6 (47.9)	9.13 (30.0)	8.80 (28.9)	5.93 (19.5)
10 000 MHz	21.8 (71.5)	15.0 (49.3)	13.5 (44.3)	N/A N/A
Power rating, kW at 25°C (77°F)				
50 MHz	1.33	3.92	2.16	3.63
100 MHz	0.93	2.74	1.48	2.53
400 MHz	0.452	1.32	0.731	1.22
1000 MHz	0.276	0.796	0.445	0.744
5000 MHz	0.11	0.30	0.17	0.29
10 000 MHz	0.072	0.19	0.12	N/A

The claimed advantages of foam dielectric, solid outer conductor cables are:

1. Lower attenuation
2. Improved RF shielding, approximately 30 dB improvement
3. High average power ratings because of the improved thermal conductivity of the outer conductor and the lower attenuation.

A disadvantage is that they are not so easy to handle as braided cables.

3.5.3 Voltage standing wave ratio (VSWR)

When an RF cable is mismatched, i.e. connected to a load of a different impedance to that of the cable, not all the power supplied to the cable is absorbed by the load. That which does not enter the load is reflected back down the cable. This reflected power adds to the incident voltage when they are in phase with each other and subtracts from the incident voltage when the two are out of phase. The result is a series of voltage – and current – maxima and minima at half-wavelength intervals along the length of the line (*Figure 3.4*). The maxima are referred to as antinodes and the minima as nodes.

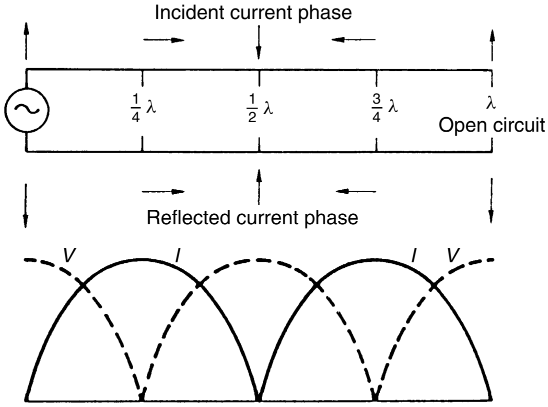

Figure 3.4 Formation of standing waves

The voltage standing wave ratio is the numerical ratio of the maximum voltage on the line to the minimum voltage: VSWR = $V_{\text{max}}/V_{\text{min}}$. It is also given by: VSWR = R_L/Z_0 or Z_0/R_L (depending on which is the larger so that the ratio is always greater than unity) where R_L = the load resistance.

The return loss is the power ratio, in dB, between the incident (forward) power and the reflected (reverse) power.

The reflection coefficient is the numerical ratio of the reflected voltage to the incident voltage.

The VSWR is 1, and there is no reflected power, whenever the load is purely resistive and its value equals the characteristic impedance of the line. When the load resistance does not equal the line impedance, or the load is reactive, the VSWR rises above unity.

A low VSWR is vital to avoid loss of radiated power, heating of the line due to high power loss, breakdown of the line caused by high voltage stress, and excessive radiation from the line. In practice, a VSWR of 1.5:1 is considered acceptable for an antenna system, higher ratios indicating a possible defect.

3.5.4 Transmission line filters, baluns and matching circuits

Use can be made of standing waves on sections of line to provide filters and RF transformers. When a line one-quarter wavelength long (a$\lambda/4$ stub) is open circuit at the load end, i.e. high impedance, an effective short-circuit is presented to the source at the resonant frequency of the section of line, producing an effective band stop filter. The same effect would be produced by a short-circuited $\lambda/2$ section. Unbalanced co-axial cables with an impedance of 50 Ω are commonly used to connect VHF and UHF base stations to their antennas although the antennas are often of a different impedance and balanced about ground. To match the antenna to the feeder and to provide a balance to unbalance transformation (known as a balun), sections of co-axial cable are built into the antenna support boom to act as both a balun and an RF transformer.

Balun

The sleeve balun consists of an outer conducting sleeve, one quarter-wavelength long at the operating frequency of the antenna, and connected to the outer conductor of the co-axial cable as in *Figure 3.5*. When viewed from point Y, the outer conductor of the feeder cable and the sleeve form a short-circuited quarter-wavelength stub at the operating frequency and the impedance between the two is very high. This effectively removes the connection to ground for RF, but not for DC, of the outer conductor of the feeder cable permitting the connection of the balanced antenna to the unbalanced cable without short-circuiting one element of the antenna to ground.

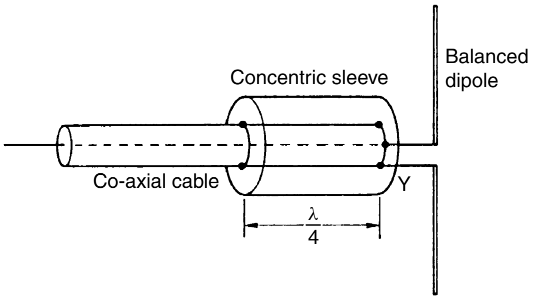

Figure 3.5 Sleeve balun

RF transformer

If a transmission line is mismatched to the load variations of voltage and current, and therefore impedance, occur along its length (standing waves). If the line is of the correct length an inversion of the load impedance appears at the input end. When a λ/4 line is terminated in other than its characteristic impedance an impedance transformation takes place. The impedance at the source is given by:

$$Z_s = \frac{{Z_0}^2}{Z_L}$$

where

Z_s = impedance at input to line

Z_0 = characteristic impedance of line

Z_L = impedance of load

By inserting a quarter-wavelength section of cable having the correct characteristic impedance in a transmission line an antenna of any impedance can be matched to a standard feeder cable for a particular design frequency. A common example is the matching of a folded dipole of 300 Ω impedance to a 50 Ω feeder cable.

Let $Z_s = Z_0$ of feeder cable and Z_0' = characteristic impedance of transformer section. Then:

$$\begin{aligned} Z_0 &= \frac{Z_0'^{\,2}}{Z_L} \\ Z_0' &= \sqrt{Z_0 Z_L} \\ &= \sqrt{300 \times 50} = 122\,\Omega \end{aligned}$$

3.6 Waveguides

At the higher microwave frequencies waveguides which conduct electromagnetic waves, not electric currents, are often used. Waveguides are conductive tubes, either of rectangular, circular or elliptical section which guide the wave along their length by reflections from the tube walls. The walls are not used as conducting elements but merely for containment of the wave. Waveguides are not normally used below about 3 GHz because their cross-sectional dimensions must be comparable to a wavelength at the operating frequency. The advantages of a waveguide over a co-axial cable are lower power loss, low VSWR and a higher operating frequency, but they are more expensive and difficult to install.

In a rectangular waveguide an electromagnetic wave is radiated from the source at an angle to the direction of propagation and is

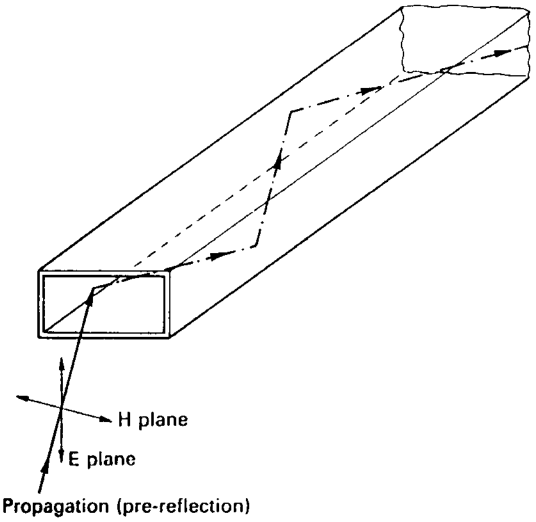

Figure 3.6 Propagation in rectangular waveguide

bounced off the walls (*Figure 3.6*). If the wave were transmitted directly along the length of the guide the electric field would be parallel to one of the walls and be short-circuited by it. Radiating the wave at an angle to the walls creates the maximum field at the centre of the guide and zero at the walls, if the dimensions of the guide are correct for the frequency. However, because the wave does not travel directly along the length of the guide, the speed of propagation is less than in space.

In an electro-magnetic wave the electric and magnetic fields, and the direction of propagation, are mutually perpendicular (see *Figure 1.3*) and such a wave may therefore be thought of as transverse electro-magnetic (TEM). In a waveguide though, because of the short-circuiting effect of the walls, a TEM wave cannot exist. A method of making the wave either transverse electric or transverse magnetic is needed.

When a wave is propagated by a reflection either the magnetic or the electric field is changed. The changed field will now contain the normal component perpendicular to the direction of propagation and a component in its direction, i.e. the wave is no longer wholly transverse. It must be either transverse electric or transverse magnetic. The terminology used to distinguish the type of wave differs: the American system uses the field which behaves as it would in free space to describe the type of wave, e.g. when there is no electric field in the direction of propagation the wave is called TE and the mode with no magnetic field in the direction of propagation, TM; the European system uses the field that is modified and an American TE wave is a European TM wave. In the European system H and E may also be used in lieu of (American) TE and TM.

The behaviour of waves in circular waveguides is similar to that in rectangular guides. Circular waveguides minimize feeder attenuation and are particularly suitable for long vertical runs. A single circular waveguide can carry two polarizations with a minimum isolation of 30 dB. Circular waveguides are recommended where attenuation is critical or where multi-band capability is needed.

Elliptical waveguides have the advantages of flexibility, the availability of long continuous runs and reduced cost.

3.7 Other transmission line considerations

3.7.1 Noise factor of coaxial cable transmission line

Any lossy electrical device, including coaxial cable, produces a noise level of its own. The noise factor of coaxial line is:

$$F_{\mathrm{N(COAX)}} = 1 + \frac{(L-1)T}{290}$$

where

$F_{\mathrm{N(COAX)}}$ is the noise factor of the coax

L is the loss expressed as a linear quantity

T is the physical temperature of the cable in Kelvins

The linear noise factor due to loss can be converted to the noise figure by $10\log(F_{\mathrm{N(COAX)}})$, which can be added to the system noise decibel for decibel.

The attenuation loss figure published in manufacturers' tables is called the *matched line loss* (L_{M}) because it refers to the situation where the load and characteristic impedance of the line are equal. But we also have to consider the *Total Line Loss* (TLL), which is:

$$\mathrm{TLL} = 10\log\left[\frac{B^2 - C^2}{B(1-C^2)}\right]$$

where

$B = \text{antilog } L_{\mathrm{M}}$

$C = (\mathrm{SWR_{LOAD}} - 1)/(\mathrm{SWR_{LOAD}} + 1)$

$\mathrm{SWR_{LOAD}}$ is the VSWR at the load end of the line

3.7.2 Types of coaxial cable

Coaxial cable consists of two cylindrical conductors sharing the same axis (hence 'co-axial') and separated by a dielectric. For low frequencies (in flexible cables) the dielectric may be polyethylene or polyethylene foam, but at higher frequencies Teflon® and other materials are used. Also used in some applications, notably high powered broadcasting transmitters, are dry air and dry nitrogen.

Several forms of coaxial line are available. Flexible coaxial cable discussed earlier in this chapter is perhaps the most common form. The outer conductor in such cable is made of either braided wire or foil. Again, television broadcast receiver antennas provide an example of such cable from common experience. Another form of flexible or semi-flexible coaxial line is *helical line* (*Figure 3.7*) in which the outer conductor is spiral wound. This type of coaxial cable is usually 2.5 or more centimetres in diameter.

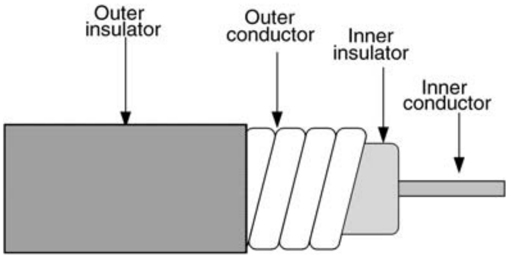
Outer insulator
Outer conductor
Inner insulator
Inner conductor

Figure 3.7 The helical line

Hardline is coaxial cable that uses a thin-walled pipe as the outer conductor. Some hardline coax used at microwave frequencies has a rigid outer conductor and a solid dielectric. *Gas-filled line* is a special case of hardline that is hollow (*Figure 3.8*), the centre conductor

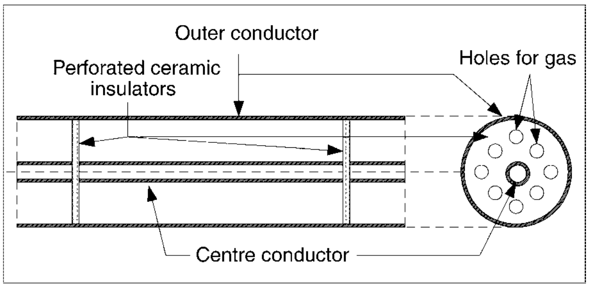
Outer conductor
Perforated ceramic insulators
Holes for gas
Centre conductor

Figure 3.8 Gas-filled line

being supported by a series of thin ceramic or Teflon insulators. The dielectric is either anhydrous (i.e. dry) nitrogen or some other inert gas.

Some flexible microwave coaxial cable uses a solid 'air-articulated' dielectric (*Figure 3.9*), in which the inner insulator is not continuous around the centre conductor, but rather is ridged. Reduced dielectric losses increase the usefulness of the cable at higher frequencies. Double shielded coaxial cable (*Figure 3.10*) provides an extra measure of protection against radiation from the line, and EMI from outside sources from getting into the system.

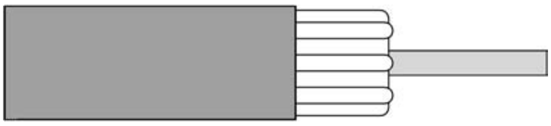

Figure 3.9 Solid 'air-articulated' dielectric

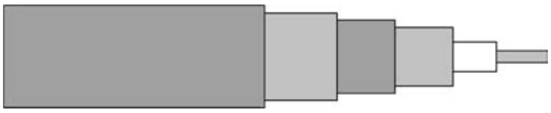

Figure 3.10 Double shielded coaxial cable

3.7.3 Transmission line noise

Transmission lines are capable of generating noise and spurious voltages that are seen by the system as valid signals. Several such sources exist. One source is coupling between noise currents flowing in the outer and inner conductors. Such currents are induced by nearby electromagnetic interference and other sources (e.g. connection to a noisy ground plane). Although coaxial design reduces noise pick-up compared with parallel line, the potential for EMI exists. Selection of high-grade line, with a high degree of shielding, reduces the problem.

Another source of noise is thermal noise in the resistances and conductances of the line. This type of noise is proportional to resistance and temperature.

There is also noise created by mechanical movement of the cable. One species results from movement of the dielectric against the two conductors. This form of noise is caused by electrostatic discharges in much the same manner as the spark created by rubbing a piece of plastic against woollen cloth.

A second species of mechanically generated noise is piezoelectricity in the dielectric. Although more common in cheap cables, one should be aware of it. Mechanical deformation of the dielectric causes electrical potentials to be generated.

Both species of mechanically generated noise can be reduced or eliminated by proper mounting of the cable. Although rarely a problem at lower frequencies, such noise can be significant at microwave frequencies when signals are low.

3.7.4 Coaxial cable capacitance

A coaxial transmission line possesses a certain capacitance per unit of length. This capacitance is defined by:

$$C = \frac{24\varepsilon}{\log(D/d)} \frac{\text{pF}}{\text{Metre}}$$

where

C is the capacitance

D is the outside conductor diameter

d is the inside conductor diameter

ε is the dielectric constant of the insulator.

A long run of coaxial cable can build up a large capacitance. For example, a common type of coax is rated at 65 pF/metre. A 150 metre roll thus has a capacitance of (65 pF/m) (150 m), or 9750 pF. When charged with a high voltage, as is done in performing breakdown voltage tests at the factory, the cable acts like a charged high voltage capacitor. Although rarely if ever lethal to humans, the stored voltage in new cable can deliver a nasty electrical shock and can irreparably damage electronic components.

3.7.5 Coaxial cable cut-off frequency (F_c)

The normal mode in which a coaxial cable propagates a signal is as a transverse electromagnetic (TEM) wave, but others are possible – and usually undesirable. There is a maximum frequency above which TEM propagation becomes a problem, and higher modes dominate. Coaxial cable should not be used above a frequency of:

$$F_{\text{CUT-OFF}} = \frac{1}{3.76(D + d)\sqrt{\varepsilon}}$$

where

F is the TEM mode cut-off frequency

D is the diameter of the outer conductor in mm

d is the diameter of the inner conductor in mm
ε is the dielectric constant

When maximum operating frequencies for cable are listed it is the TEM mode that is cited. Beware of attenuation, however, when making selections for microwave frequencies. A particular cable may have a sufficiently high TEM mode frequency, but still exhibit a high attenuation per unit length at X or Ku-bands.

References

Andrew Antennas, (1991). *Catalogue 35*, Illinois.
British Telecommunications (1992). *Connect Direct, private circuits*. BT, London.
Kennedy, G. (1977). *Electronic Communications Systems*. McGraw-Hill Kogashuka, Tokyo.
Terman, F.E. (1943), *Radio Engineers' Handbook*, McGraw-Hill, London.
Winder, S.W. (2001). *Newnes Telecommunications Pocket Book*. Butterworth-Heinemann, Oxford.

4 Antennas

4.1 Antenna characteristics

4.1.1 Bandwidth

Stated as a percentage of the nominal design frequency, the bandwidth of an antenna is the band of frequencies over which it is considered to perform acceptably. The limits of the bandwidth are characterized by unacceptable variations in the impedance which changes from resistive at resonance to reactive, the radiation pattern, and an increasing VSWR.

4.1.2 Beamwidth

In directional antennas the beamwidth, sometimes called half-power beamwidth (HPBW), is normally specified as the total width, in degrees, of the main radiation lobe at the angle where the radiated power has fallen by 3 dB below that on the centre line of the lobe (*Figure 4.1A*).

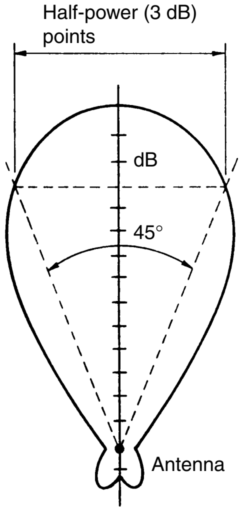

Figure 4.1A Half-power beamwidth

4.1.3 Directivity and forward gain

All practical antennas concentrate the radiated energy in some directions at the expense of others. They possess directivity but are completely passive; they cannot increase the power applied to them. Nevertheless, it is convenient to express the enhanced radiation in some directions as a power gain.

Antenna gain may be quoted with reference to either an isotropic radiator or the simplest of practical antennas, the dipole. There is a difference of 2.15 dB between the two figures. A gain quoted in dBi is with reference to an isotropic radiator and a gain quoted in dBd is with reference to a dipole. When gain is quoted in dBi, 2.15 dB must be subtracted to relate the gain to that of a dipole.

4.1.4 Effective height or length

The current flowing in an antenna varies along its length (see *Figure 1.4*). If the current were uniform along the length of an antenna it would produce a field appropriate to its physical length, and the effective height or length of the antenna would be its physical length. In practice, because the current is not uniform, the effective length is less than the physical length and is given by:

$$l_{\text{eff}} = \frac{l_{\text{phys}} \times I_{\text{mean}}}{I}$$

where
l_{eff} = effective length
l_{phys} = physical length
I = current at feed point

With an antenna which is short in comparison with a wavelength the current can be considered to vary linearly over its length and $I_{\text{mean}} = I/2$. Because the apparent length of a vertical radiator is twice that of its physical dimension due to the mirror image formed below the ground, the effective length of an electrically short vertical antenna may be approximated to be its physical length.

4.1.5 Effective radiated power (erp)

This is the power effectively radiated along the centre line of the main lobe. It is the power supplied to the antenna multiplied by the antenna gain with reference to a dipole.

4.1.6 Radiation resistance and efficiency

The power radiated by an antenna can conveniently be expressed in terms of the value of a resistor which would dissipate the same power that the antenna radiates. This value is referred to as the radiation resistance and is defined as the ratio of the power radiated to the square of the current at the feed point. The efficiency is the ratio of the power radiated to that lost in the antenna. It is given by:

$$\text{eff} = \frac{R_\mathrm{r}}{R_\mathrm{r} + R_\mathrm{L}} \times 100\%$$

where R_r is the radiation resistance and R_L represents the total loss resistance of the antenna. The sum of the two resistances is the total resistance of the antenna and, for a resonant antenna, is also the impedance.

4.1.7 Front-to-back ratio

The ratio, in dB, of the strength of the radiation (or received signal) in the forward (desired) direction to that in the reverse (unwanted) direction. The front-to-back ratio of the antenna shown in *Figure 4.1A* is 13 dB.

4.1.8 Impedance

The impedance of an antenna is that presented to the feeder cable connecting it to the transmitter or receiver. It is the result of the vectorial addition of the inductive, capacitive and resistive elements of the antenna. Each resonant antenna possesses an impedance characteristic of the type, and when an antenna operates at its resonant frequency the reactive elements cancel out and the impedance becomes resistive. The radiation resistance plus the losses in the antenna, i.e. the series resistance of the conductors, the shunt resistance of the base material and losses in nearby objects, form the resistive portion of the impedance.

4.1.9 Polarization

The radiated field from an antenna is considered to be polarized in the plane of the length of the conductors which is the plane of the electric field, the E plane. Confusion arises when reference is made to vertical or horizontal polarization and it is preferable when referring to polar diagrams to use the E and H plane references.

Circular polarization, produced by crossed dipoles or helical wound antennas, is occasionally used for point-to-point work at VHF and above to reduce multi-path propagation losses.

Cross polarization discrimination defines how effectively an antenna discriminates between a signal with the correct polarization, i.e. mounted with the elements in the same plane, and one operating at the same frequency with the opposite polarization. 20 dB is typical.

4.1.10 Radiation pattern

A plot of the directivity of an antenna showing a comparison of the power radiated over 360°. Two polar diagrams are required to show the radiation in the E and H planes. The polar diagrams may be calibrated in either linear (voltage) or logarithmic (decibel) forms.

4.1.11 Voltage standing wave ratio (VSWR)

Most VHF and UHF antennas contain an impedance matching device made up of lengths of co-axial cable. Thus the VSWR (see Chapter 3) of these types of antenna varies with the operating frequency, more so than the bandwidth of the antenna alone would produce. At the centre design frequency, the VSWR should, theoretically, be 1:1 but in practice a VSWR less than 1.5:1 is considered acceptable.

4.1.12 Receive aperture

Receiving antennas also possess a property called aperture, or capture area. This concept relates the amount of power that is delivered to a matched receiver to the power density (watts per square metre). The aperture is often larger than the physical area of the antenna, as in the case of the half-wavelength dipole (where the wire fronts a very small physical area), or less as in the case of a parabolic reflector used in microwave reception. *Figure 4.1B* shows the capture area of a half-wavelength (0.5λ) dipole. It consists of an ellipse with major axes of 0.51λ and 0.34λ. The relationship between gain and aperture is:

$$A_e = \frac{G\lambda^2}{4\pi n}$$

where

A_e is the effective aperture

G is the gain

λ is the wavelength of the signal

n is the aperture effectiveness ($n = 1$ for a perfect no-loss antenna, real values are typically 0.3 to 0.55

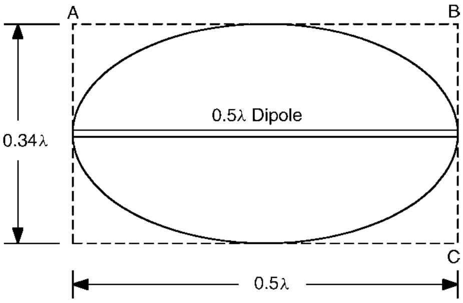
A
B
C
0.34λ
0.5λ Dipole
0.5λ

Figure 4.1B Capture area of half-wavelength (0.5λ) dipole

4.2 Antenna types

4.2.1 The dipole

The half-wavelength (λ/2) dipole as described in Section 1.4 is the antenna on which many others are based. *Figure 4.2* shows the relative radiation in the E and H planes of a dipole in free space.

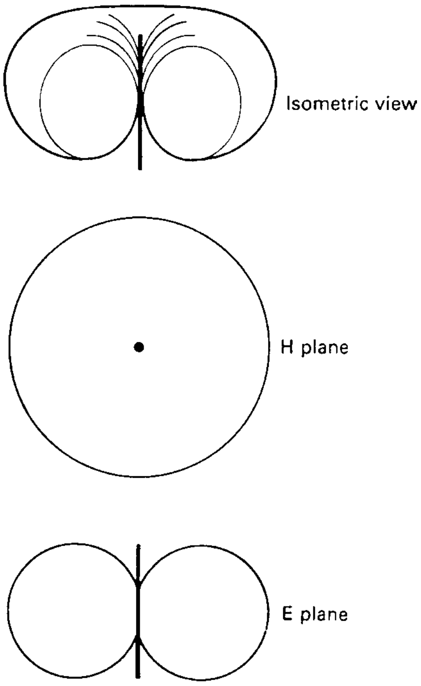
Isometric view
H plane
E plane

Figure 4.2 Radiation patterns of a half-wave dipole

The impedance of a half-wavelength dipole is 72 Ω; that of a full wavelength or folded dipole is 300 Ω.

4.2.2 *The quarter-wavelength vertical radiator*

The quarter-wavelength ($\lambda/4$) vertical radiator is a commonly used antenna for MF broadcasting and for VHF and UHF mobile radio applications. It is derived from the $\lambda/2$ dipole and it is assumed that a mirror image of the radiator is formed below the ground to complete the $\lambda/2$ structure of the dipole as in *Figure 4.3*. The radiation pattern of a $\lambda/4$ vertical radiator mounted close to a perfect earth shows a strong similarity to that of a dipole. The effect of reducing the size and conductivity of the ground plane raises the angle of radiation.

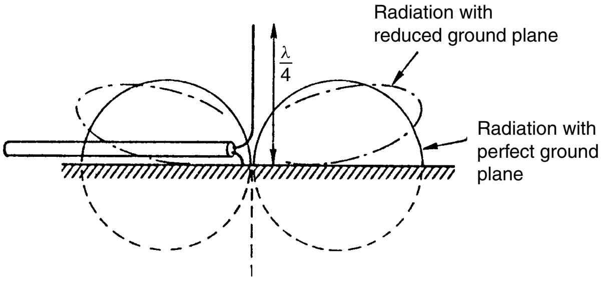
Radiation with reduced ground plane
$\frac{\lambda}{4}$
Radiation with perfect ground plane

Figure 4.3 Quarter-wavelength vertical radiator

The impedance of a perfect $\lambda/4$ vertical radiator is 36 Ω but reducing the effectiveness of the ground plane raises the impedance.

4.2.3 *LF, MF and HF antennas*

Because of the physical lengths involved, LF and MF antennas are usually non-resonant and their impedances do not conform to the resistive 70 Ω or 36 Ω of the basic resonant types. The impedance of a non-resonant antenna is usually higher and reactive so an antenna tuning or matching unit is used to couple the antenna efficiently to the transmission line and also act as filter to reduce out-of-band radiations. The matching unit comprises a tuned circuit with either a tap on the coil at the correct impedance point or a separate coupling coil to feed the antenna.

Obtaining an adequate length is always the problem with low frequency antennas and various methods have been used based mainly on the $\lambda/4$ radiator. The horizontal section of the inverted L (*Figure 4.4*) extends the effective length but, as the ground wave is much used at

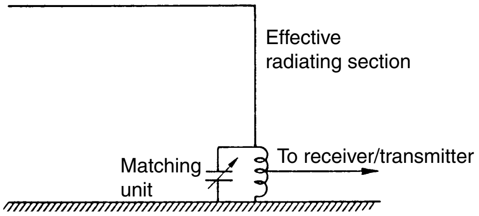

Figure 4.4 Inverted L antenna

the lower frequencies, these antennas are intended for vertical polarization and it is therefore only the down-lead which radiates, or receives, effectively. An alternative method of increasing the effective height of a vertical radiator is to provide a capacitance top where the system of horizontal conductors provides a high capacitance to ground. This prevents the current falling to zero at the top of the antenna, maintaining a higher mean current and so increasing the antenna's effective length.

Dipoles used at HF are mounted horizontally because of their length and have directivity in the horizontal (E) plane. Propagation is mainly by the sky wave and the omni-directional properties in the vertical (H) plane, modified by ground reflections, produce wide angle upwards radiation.

4.2.4 Directional arrays

Broadside array

A broadside array consists of several radiators spaced uniformly along a line, each carrying currents of the same phase. When each radiator has an omni-directional pattern, and the spacing between radiators is less than $3\lambda/4$, maximum radiation occurs at right angles to the line of the array. The power gain is proportional to the length of the array, provided that the length is greater than two wavelengths; this means, effectively, the number of radiators. *Figure 4.5* shows a typical H plane polar diagram for an array with vertically mounted radiators and a spacing of $\lambda/2$.

End-fire array

Physically an end-fire array is identical to a broadside except for the feeding arrangements and the spacing of the elements. In an end-fire array the radiators are fed with a phase difference between adjacent radiators equal in radians to the spacing between them in wavelengths.

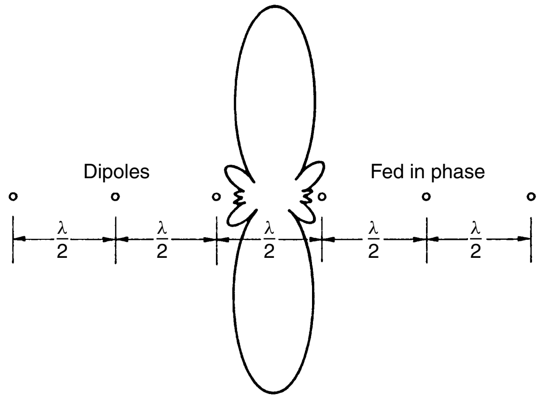
Dipoles
Fed in phase
$\frac{\lambda}{2}$
$\frac{\lambda}{2}$
$\frac{\lambda}{2}$
$\frac{\lambda}{2}$
$\frac{\lambda}{2}$

Figure 4.5 Broadside array

A spacing of λ/4 requires a phase shift of 90° between adjacent radiators. *Figure 4.6* shows a typical radiation pattern. An end-fire array concentrates the power in both the E and H planes and the maximum radiation is in the direction of the end of the array with the lagging phase.

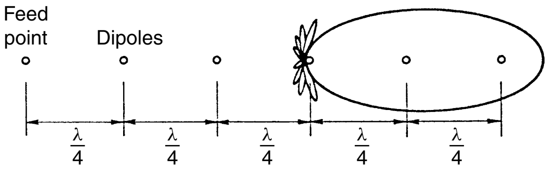
Feed point
Dipoles
$\frac{\lambda}{4}$
$\frac{\lambda}{4}$
$\frac{\lambda}{4}$
$\frac{\lambda}{4}$
$\frac{\lambda}{4}$

Figure 4.6 End-fire array

Rhombic antenna

A rhombic is a wide band, directional antenna comprised of four non-resonant wire antennas, each several wavelengths long, arranged as shown in *Figure 4.7(a)* which also shows the radiation pattern for each leg of the rhombus. The lobe angle θ can be varied by adjusting the length, in wavelengths, of each radiator. The antenna has greater directivity than a single wire and can be terminated by an appropriate value resistor to ensure non-resonance and a wide bandwidth. However, because it must be terminated in a resistance equal to the characteristic impedance of the conductors, it cannot be more than 50% efficient. It also exhibits considerable side lobes of radiated power. Rhombics are used for sky wave working at HF and more than one frequency is allocated to allow for varying propagation conditions. The conductors of a rhombic are normally horizontal and the horizontal directivity is determined by the tilt angle, β in *Figure 4.7(a)*. If the lobe angle θ is equal to $(90 - \beta)^\circ$ the radiation in the A lobes cancels, while that from the B lobes, which point in the same direction, is added. The resultant pattern in the horizontal plane is shown in *Figure 4.7(b)*. The vertical directivity is controlled by the height of the conductors above the ground.

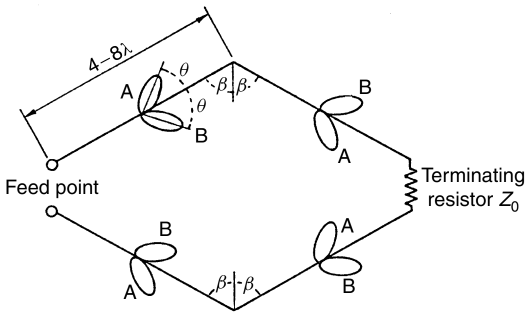

(a) Rhombic antenna

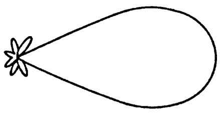

(b) Radiation pattern of rhombic

Figure 4.7

Log-periodic antenna

An alternative, usable from HF through UHF, to the rhombic for wide band operation is the log-periodic antenna. It is comprised of several dipoles of progressive lengths and spacings as in *Figure 4.8*, and is resonant over a wide frequency range and may be mounted with either polarization. The dipoles are fed via the support booms and this construction ensures that the resultant phasing of the dipoles is additive in the forward direction producing an end-fire effect. However, because at any one frequency only a few of the dipoles are close to resonance, the forward gain of the antenna is low considering the number of elements it contains.

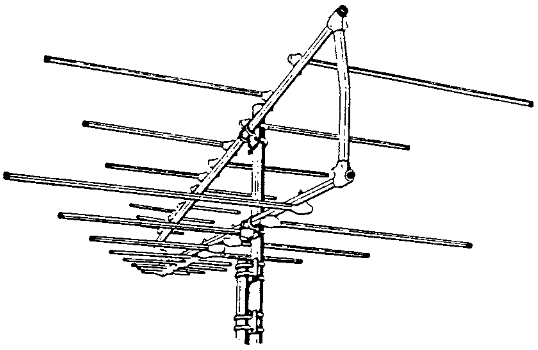

Figure 4.8 Log-periodic antenna at VHF frequency

4.3 VHF and UHF antennas

4.3.1 Base station antennas

Apart from entertainment broadcasting and mobile telephony, most VHF and UHF systems use vertical polarization and a dipole – or to prevent noise from rain static, the folded dipole – with the conductors mounted vertically is a frequently used antenna for VHF and UHF base station installations. Unfortunately it is often mounted on the side of the support structure in a manner which seriously affects its omnidirectional radiation pattern. Where practical, there should be a minimum spacing of one wavelength between the structure and the rearmost element of the antenna.

To obtain a good omni-directional pattern either an end-fed dipole (*Figure 4.9*) or a unipole antenna (*Figure 4.10*) protruding from the top of the mast or tower is the best option. A unipole is a variation of the vertical quarter-wave radiator and provides a low angle of radiation.

To reduce the likelihood of co-channel interference directional antennas are often necessary. The simplest of these is the combination of a $\lambda/2$ dipole and reflector shown in *Figure 4.11*. The reflector is slightly longer than the dipole and spaced one quarter-wavelength from it. The portion of the signal radiated by the dipole in the direction of the reflector is received and re-transmitted by the reflector, with a 180° phase change occurring in the process. The signal re-transmitted

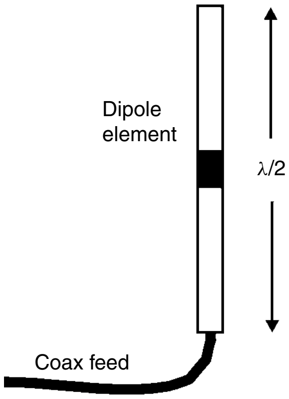

Figure 4.9 Type EDV end-fed antenna (by kind permission of C and S Antennas Ltd)

to the rear of the antenna – the direction of the reflector – cancels the signal from the dipole, that towards the front of the antenna adds to the signal from the dipole giving the radiation pattern shown. The power gain of a dipole and reflector, a two-element array, is 3 dBd.

Directivity can be increased by adding directors forward of the dipole, the result is a Yagi–Uda array. The limit to the number of radiators is set by physical constraints and the reduction of bandwidth produced by their addition. At low VHF, a 3-element array is about the practical maximum, while at 1500 MHz, 12-element arrays are commonplace. As a rule of thumb, doubling the number of elements in an array increases the forward gain by 3 dB. Where the maximum front-to-back ratio is essential the single rod reflector can be replaced by a corner reflector screen.

Co-linear antennas provide omni-directional characteristics and power gain in the H plane. A co-linear consists of a number of dipoles stacked vertically and, in the normal configuration, fed so that they radiate in phase and the maximum power is radiated horizontally. *Figure 4.12* shows alternative feeding arrangements. One advantage of the co-linear is that the horizontal angle of radiation can be tilted to about 15° downwards by changing the phasing of the elements. The gain of a co-linear is limited, because of the physical lengths involved and losses in the feeding arrangements to 3 dBd at VHF and 6 dBd at UHF.

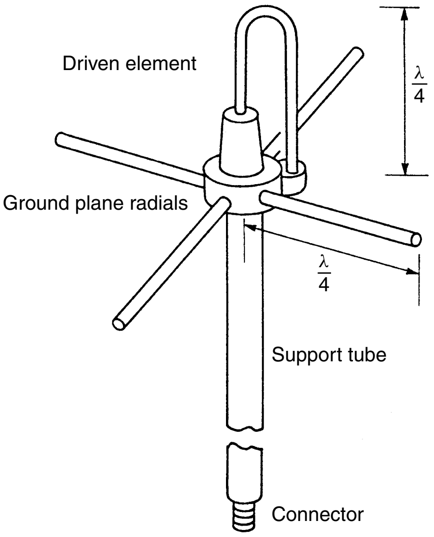
Driven element
$\frac{\lambda}{4}$
Ground plane radials
$\frac{\lambda}{4}$
Support tube
Connector

Figure 4.10 Folded unipole antenna

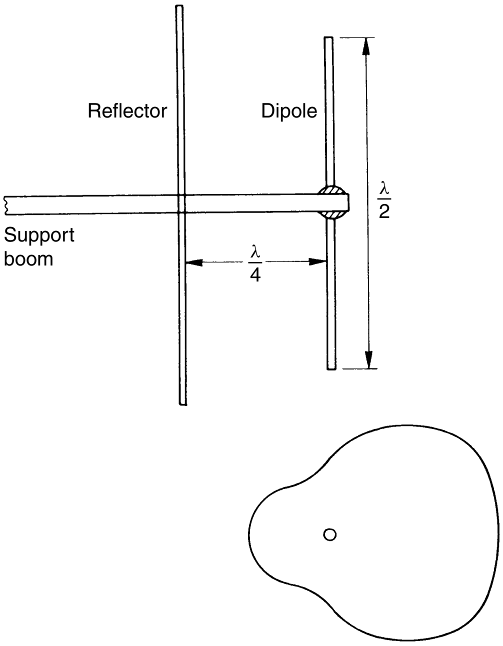
Reflector
Dipole
$\frac{\lambda}{2}$
Support boom
$\frac{\lambda}{4}$

Figure 4.11 Dipole and reflector

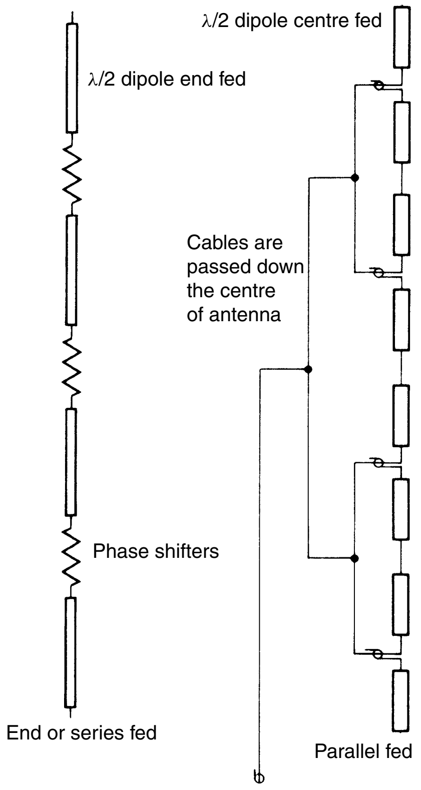
λ/2 dipole centre fed
λ/2 dipole end fed
Cables are passed down the centre of antenna
Phase shifters
End or series fed
Parallel fed

Figure 4.12 General construction of co-linears

Figure 4.13 shows a slot antenna cut into a flat metal sheet. Current (I) injected at the centre of the slot flows around the edge and creates a vertical electric field. The radiated field pattern is like a dipole.

The type of slot antenna typically used for mobile telephony base stations is a cylindrical waveguide with slots cut width-wise. Current flowing along the waveguide creates an electric field along the length of the cylinder. The radiation pattern produced by a slot antenna cut into a cylinder is directional, with the main beam perpendicular to the slot. Using two slot antennas side by side provides radio coverage over a 120° sector. Three pairs of slot antennas placed around a mast gives three sectors that can operate at different frequencies.

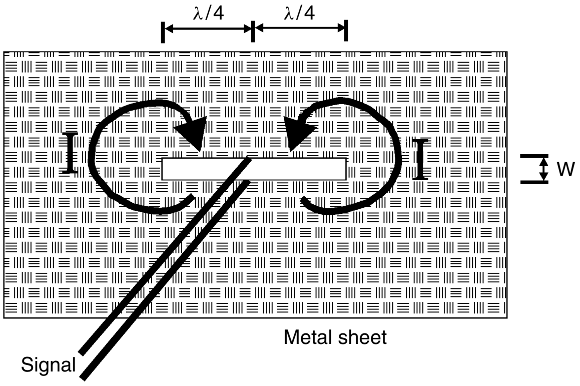

Figure 4.13 Slot antenna

A wide-band alternative to the log-periodic is the conical (discone) antenna (*Figure 4.14*). It provides unity gain, is omni-directional and has a bandwidth of approximately 3:1, depending on the designed frequency range. In practice there has been a tendency to expect these antennas to perform outside their specified bandwidths with unsatisfactory results.

Stacking and baying

A method of increasing an antenna's directivity is to mount two or more antennas vertically above one another (stacking) or side-by-side (baying), and to feed them so that they radiate in phase. Stacking two dipoles vertically increases the directivity in the E plane and baying them increases the directivity in the H plane, approximately halving the beamwidth in each case.

An array of two stacked plus two bayed antennas approximately halves the beamwidth in both planes.

4.3.2 Mobile antennas

The aerial is the least expensive, and most abused, component of a mobile radio installation. Frequently installed in a manner which does not produce optimum performance it can have a profound effect on the performance of the whole installation.

Most mobile antennas consist of a metal rod forming a quarter-wavelength radiator. The ideal mounting position is the centre of a

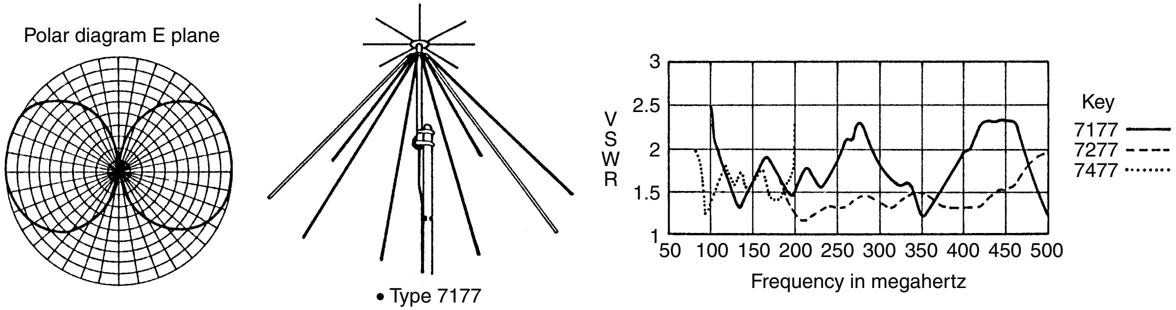

- Type 7177
  100 to 470 MHz
- Type 7277
  225 to 400 MHz
- Type 7477
  80 to 200 MHz

Figure 4.14 Discone wide band antenna (by kind permission of Jaybeam Ltd)

metallic roof, and as the area of the ground plane is reduced the radiation pattern changes and more of the energy is radiated upwards (not always a bad thing in inner city areas); also, the impedance rises.

The effect of the mounting position on the H plane radiation can be dramatic, resulting in ragged radiation patterns and, in some directions, negligible radiation. Advice on the installation of mobile antennas and the polar diagrams produced by typical installations are illustrated in MPT 1362, *Code of Practice for Installation of Mobile Radio Equipment in Land Based Vehicles*.

As the installation moves away from the ideal and the antenna impedance rises a mismatch is introduced between the antenna and the feeder with the consequent production of standing waves on the feeder. Under high VSWR conditions the cable is subject to higher voltage stresses and it also behaves as an aerial radiating some of the reflected power. This spurious radiation adds to the radiation from the antenna in some directions but subtracts from it in others giving rise to jagged radiation patterns or deep nulls in radiated signal.

Mobile antennas providing a small amount of gain, typically 3 dB and obtained by narrowing the radiation lobes, are on the market. These have a length of 5/8 wavelength and, because the extra length makes the impedance capacitive at the operational frequency, a loading coil is inserted at the lower end of the element to cancel the capacitive reactance. An adjustable metallic disk towards the base of the whip is often provided for tuning purposes. Note that gain figures quoted for mobile antennas are usually with reference to a quarter-wave whip.

Low profile antennas

Low profile antennas are available for use at UHF. They have a built-in ground plane approximately 150 mm in diameter and a height of some 30 mm and have obvious applications for use on high vehicles and, although not strictly covert, where a less obtrusive antenna is required. They are fitted with a tuning screw and when adjusted to resonance a VSWR of better than 1.2:1 is quoted by one maker and a bandwidth of 10 MHz at a VSWR of 2:1. *Figure 4.15* shows the radiation pattern for one type.

Motor-cycle antennas

The installations of antennas on motor cycles poses problems because of the absence of a satisfactory ground plane. One frequently used method is to employ a 5/8 wavelength whip and loading coil. Another

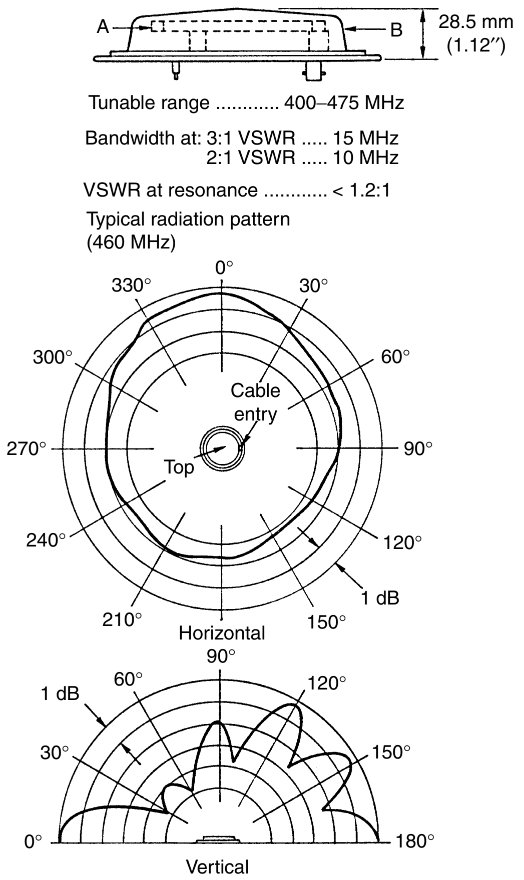
A
B
28.5 mm
(1.12″)
Tunable range .... 400–475 MHz
Bandwidth at: 3:1 VSWR ..... 15 MHz
2:1 VSWR ..... 10 MHz
VSWR at resonance .... < 1.2:1
Typical radiation pattern
(460 MHz)
0°
330°
30°
300°
60°
Cable
entry
270°
90°
Top
240°
120°
1 dB
210°
150°
Horizontal
90°
60°
120°
1 dB
30°
150°
0°
180°
Vertical

Antenna mounted on a 40 cm × 40 cm aluminium plate

Figure 4.15 Low profile UHF antenna (by kind permission of Panorama Antennas Ltd)

method uses a pair of grounded downwards-pointing rods to form the lower half of a dipole.

Hand-portable antennas

Again, because of the lack of a ground plane high performance antennas are difficult to provide for hand-portables, particularly at VHF. Body-worn sets may have an antenna incorporated in the microphone lead but the high current portion of the antenna must then be at a low height and in some directions the radiation must pass through the body, which is highly absorbent, to reach the base station. Helical antennas are frequently used on hand-held sets to reduce the physical length. Useful operating tips are to face the base station when using the radio in low signal areas, while placing the set on a nearby car roof effectively increases the performance of the antenna.

Safety

There are two safety aspects to consider when installing mobile antennas: physical and electrical. The physical considerations are that the antenna must be incapable of inflicting injury when it is in its correct position, and also when it has been bent or damaged. Rear wing mounted antennas need particular care in their positioning; a Band 111 aerial tip is just about eye height when bending over an open boot lid. The same considerations apply to hand-portables, helical antennas being perhaps safer than whips because they are thicker and thus more easily seen. They also have rounded tips.

The electrical dangers are from radiation affecting the body either directly – radiation from a hand-portable helical into the eye is a possible example – or indirectly by affecting electronic equipment. The danger of radiation affecting equipment in the vehicle is increased when the VSWR is high because of increased radiation from the feeder. Advice should be sought from the Radiological Protection Board.

4.4 Microwave antennas

The small antenna elements at microwaves facilitate the construction of highly directive, high gain antennas with high front-to-back ratios.

At frequencies below about 2 GHz, 12- to 24-element Yagi arrays, enclosed in plastic shrouds for weather protection, may be used. At higher frequencies, antennas with dish reflectors are the norm.

The aperture ratio (diameter/wavelength) of a dish governs both its power gain and beamwidth. The power gain of a parabolic dish is given to a close approximation by:

$$\text{Gain} = 10\log_{10} 6(D/\lambda)^2 \times N, \text{dBi}$$

where $D =$ dish diameter and $N =$ efficiency. Dimensions are in metres. The half-power beam width (HPBW) in degrees is approximately equal to $70\lambda/D$.

A microwave antenna with its dish reflector, or parasitic elements in the case of a Yagi type, is a large structure. Because of the very narrow beamwidths – typically 5° for a 1.8 m dish at 2 GHz – both the antenna mounting and its supporting structure must be rigid and able to withstand high twisting forces to avoid deflection of the beam in high winds. Smooth covers, radomes, fitted to dishes and the fibre-glass shrouds which are normally integral with Yagis designed for these applications considerably reduce the wind loading and, for some antenna types, increase the survival wind speed.

The electrical performance of a selection of microwave antennas is given in *Table 4.1* and the wind survival and deflection characteristics in *Table 4.2* (Andrew Antennas, 1991).

Table 4.1 2.1–2.2 GHz antennas – electrical characteristics

Type number	*Dia. (m)*	*Gain (dBi)*			*Beam width (deg.)*	*Cross pol. disc (dB)*	*F/B ratio (dB)*	*VSWR max.*
		Bottom	*Mid-band*	*Top*				
Ultra High Performance Antenna, F-Series Unpressurized – Radome Inc.								
Single polarized								
UHP8F-21	2.4	31.9	32.1	32.3	4.2	32	61	1.10
UHP10F-21	3.0	33.7	33.9	34.0	3.6	33	64	1.10
UHP12F-21	3.7	35.4	35.6	35.8	2.9	32	65	1.10
Dual polarized								
UHX8F-21	2.4	31.9	32.1	32.3	4.2	30	58	1.20
UHX10F-21A	3.0	33.8	34.0	34.2	3.6	32	62	1.20
UHX12F-21A	3.7	35.4	35.6	35.8	2.8	32	67	1.20
High Performance Antenna, F-Series Unpressurized – Radome Inc.								
Single polarized								
HP6F-21B	1.8	29.4	29.6	29.8	5.5	30	46	1.12
HP8F-21A	2.4	31.9	32.1	32.3	4.1	30	53	1.12
HP10F-21A	3.0	33.8	34.0	34.2	3.4	32	55	1.12
HP12F-21A	3.7	35.4	35.6	35.8	2.9	32	56	1.12
Standard Antenna, F-Series Unpressurized								
Single polarized								
P4F-21C	1.2	26.4	26.6	26.8	7.6	30	36	1.15
P6F-21C	1.8	29.8	30.0	30.2	4.9	30	39	1.12
P8F-21C	2.4	32.3	32.5	32.7	3.8	30	40	1.12
P10F-21C	3.0	34.0	34.2	34.4	3.4	30	44	1.12
Grid Antenna, F-Series Unpressurized								
Single polarized								
GP6F-21A	1.8	29.6	29.8	30.0	5.4	31	36	1.15
GP8F-21A	2.4	32.0	32.2	32.4	4.0	35	39	1.15
GP10F-21A	3.0	34.0	34.2	34.4	3.3	40	42	1.15
GP12F-21	3.7	35.5	35.7	35.9	2.8	40	44	1.15

With shrouded Yagis and some dishes low loss foam-filled cables are generally used up to about 2 GHz although special connectors may

Table 4.2 Wind survival and deflection characteristics

Antenna types	*Survival ratings*		*Max. deflection in 110 km wind (degrees)*
	Wind velocity (km/h)	*Radial ice (mm)*	
P4F Series			
Without Radome	160	12	0.1
With Radome	185	12	0.1
Standard Antennas (except P4F Series)			
Without Radome	200	25	0.1
With Standard Radome	200	25	0.1
UHX, UMX, UGX Antennas	200	25	0.1

be required. At higher frequencies, air-spaced or pressurized nitrogen-filled cables are frequently used with waveguides as an alternative.

4.4.1 Omnidirectional normal mode helix

The normal mode helix antenna shown in *Figure 4.16* produces an omnidirectional pattern when the antenna is mounted vertically. The diameter (D) of the helical coil should be one-tenth wavelength ($\lambda/10$), while the pitch (i.e. S, the distance between helix loops) is one-twentieth wavelength ($\lambda/20$). An example of the normal mode helix is the 'rubber ducky' antenna used on VHF/UHF two-way radios and scanners.

4.4.2 Axial mode helical antenna

An axial mode helical antenna is shown in *Figure 4.17*. This antenna fires 'off-the-end' in the direction shown by the arrow. The helix is mounted in the centre of a ground plane that is at least 0.8λ across. For some UHF frequencies some manufacturers have used aluminum pie pans for this purpose. The helix itself is made from either heavy copper wire (solid, not stranded) or copper or brass tubing. The copper tubing is a bit easier to work. The dimensions are:

$$D \approx \lambda/3$$

$$S \approx \lambda/4$$

$$\text{Length} \approx 1.44\lambda$$

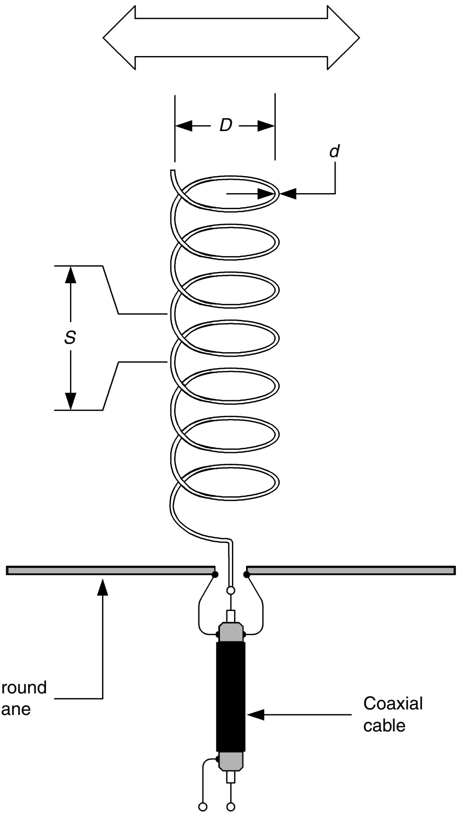

Figure 4.16 Normal mode helix antenna

A 'rule of thumb' for the circumference is that maximum gain is obtained when circumference C is:

$$C = 1.066 + [(N - 5) \times 0.003]$$

4.4.3 Small loop antennas

Small loop antennas are used mostly for receiving, although some designs are also used for transmitting. One application for the small loop antenna is radio direction finding (RDF). Another use is for providing a small footprint antenna for people who cannot erect a full sized receiving antenna. Perhaps the greatest use of the small loop antenna is for receiving stations on crowded radio bands. The small loop antenna has very deep nulls that make it easy to null out interfering co-channel and adjacent channel signals.

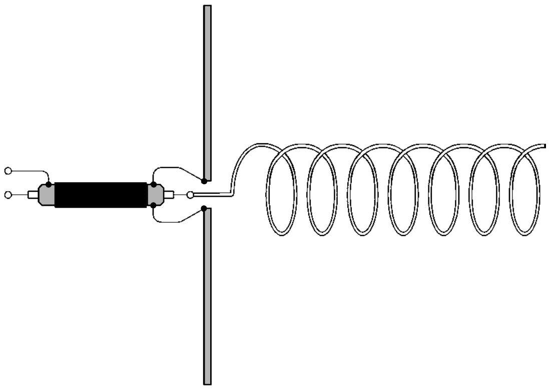

Figure 4.17 Axial mode helical antenna

4.5 Loop antennas

4.5.1 Small loop antennas defined

Large loop antennas are those with overall wire lengths of 0.5λ to more than 2λ. Small loop antennas, on the other hand, have an overall wire length that is much less than one wavelength (1λ). According to a Second World War US Navy training manual such antennas are those with an overall length of $\leq 0.22\lambda$. Jasik's classic 1961 text on radio antennas uses the figure $\leq 0.17\lambda$, while John Kraus (1950) used the figure $\leq 0.10\lambda$. An amateur radio source, *The ARRL Antenna Book*, recommends $\leq 0.085\lambda$ for small loop antennas. For the purposes of our discussion we will use Kraus's figure of $\leq 0.10\lambda$.

A defining characteristic of small loops versus large loops is seen in the current distribution. In the small loop antenna the current flowing in the loop is uniform in all portions of the loop. In the large loop, however, the current varies along the length of the conductor, i.e. there are current nodes and antinodes.

The small loop antenna also differs from the large loop in the manner of its response to the radio signal. A radio signal is a transverse electromagnetic (TEM) wave, in which magnetic and electrical fields alternate with each other along the direction of travel. The large loop, like most large wire antennas, respond primarily to the *electrical field* component of the TEM, while small loops respond mostly to the *magnetic field* component. The importance of this fact is that it means the small loop antenna is less sensitive to local electromagnetic interference sources such as power lines and appliances. Local EMI consists largely of electrical fields, while radio signals have both magnetic and electrical fields. With proper shielding, the electrical response can be reduced even further.

4.5.2 Small loop geometry

Small loop antennas can be built in any of several different shapes (*Figure 4.18*). Popular shapes include hexagonal (*Figure 4.18A*), octagonal (*Figure 4.18B*), triangular (*Figure 4.18C*), circular (*Figure 4.18D*) and square (*Figure 4.18E*).

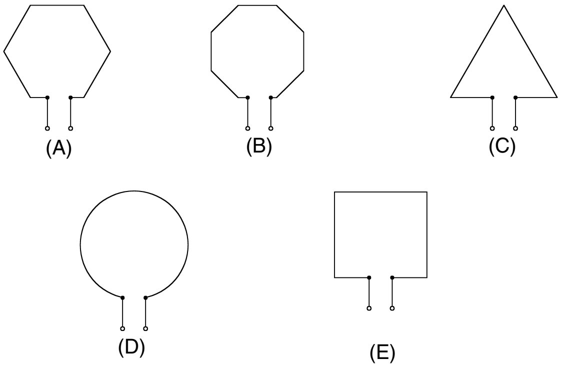

Figure 4.18 Small loop antennas: (A) hexagonal, (B) octagonal, (C) triangular, (D) circular and (E) square

The far-field performance of small loop antennas is approximately equal provided that $A^2 \leq \lambda/100$, where A is the loop area and λ is the wavelength of the desired frequency.

The 'standard' loop used in this discussion is the square loop depicted in *Figure 4.19*. There are two forms of winding used: *depth wound* and *spiral wound*. The difference is that the depth wound has its turns in different parallel planes, while in the spiral wound version all of the turns are in the same plane. The spiral wound loop theoretically has a deeper null than depth wound, but in practical terms there is usually little difference.

The length of each side of the loop is designated A, while the depth or width is designated B. These dimensions will be used in equations shortly. The constraint on the dimensions is that A should be $\leq 0.10\lambda/4$ and $B \leq A/5$.

The loop may be either tuned or untuned. The differences between tuned and untuned will be discussed shortly.

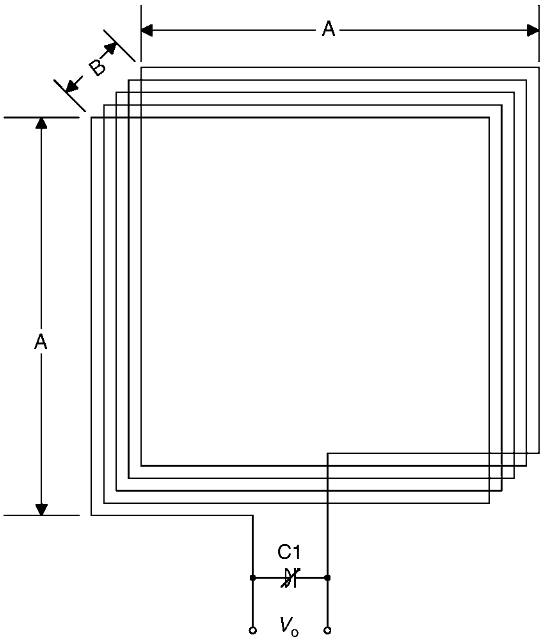

Figure 4.19 Square loop

4.5.3 *Small loop antenna patterns*

Small loop antennas have patterns opposite those of large loops. The minima, or 'nulls', are perpendicular to the plane of the loop, while the maxima are off the ends. *Figure 4.20* shows the directions of maximum and minimum response. The loop antenna is viewed from above. The nulls are orthogonal to the loop axis, while the maxima are along the loop axis.

The fact that the small loop pattern has nulls perpendicular to the loop axis, i.e. perpendicular to the plane of the loop, is counterintuitive to many people. The advancing radio wave produces alternating regions of high and low amplitude. A *potential difference* exists between any two points. When the loop is aligned such that its axis is parallel to the isopotential lines low signal levels are induced into the loop. If the turns of the loop cut several isopotential lines, a larger signal is induced from this direction.

4.5.4 *Signal voltage (V_o) developed by the loop*

The actual signal voltage (V_o) at the output of the terminals of an untuned loop is a function of the direction of arrival of the signal (α),

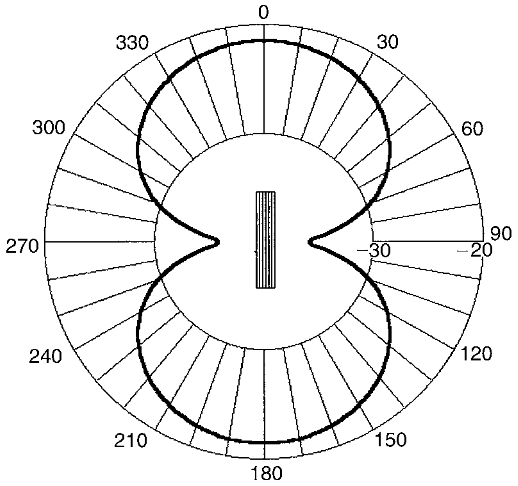

Figure 4.20 Small loop azimuth

as well as the strength of the arriving signal and the design of the loop. The angle α is formed between the loop axis and the advancing isopotential lines of the radio signal.

The output voltage of an untuned loop is:

$$V_o = \frac{2\pi A N E_f \cos\alpha}{\lambda}$$

While the output voltage of a tuned loop is:

$$V_o = \frac{2\pi A N Q E_f \cos\alpha}{\lambda}$$

where

V_o is the output voltage of an untuned loop in volts (V)

A is the length of one side of the loop

N is the number of turns in the loop

E_f is the strength of the incoming signal in volts per metre (V/m)

α is the angle between the advancing wavefront and the loop axis

λ is the wavelength of the radio signal in metres (m), i.e. the reciprocal of the frequency ($\lambda = 1/F$)

Q is the loaded Q (figure of merit) of the tuned circuit formed by C1 and the loop inductance (typically 10 to 100)

4.5.5 Effective height

Loop antennas are sometimes described in terms of the *effective height* (H_{eff}) of the antenna. This number is a theoretical construct that compares the output voltage of a small loop with a vertical piece of the same kind of wire that has a height of:

$$H_{\text{eff}} = \frac{2\pi NA}{\lambda}$$

where H_{eff} is the effective height in metres, and all other terms are as defined above.

Loop inductance

A loop antenna essentially forms a coil of wire (or other conductor), so will have inductance. There are several methods for calculating loop inductance, but the most common are the *Grover equation* and the *Patterson equation.*

Grover equation:

$$L_{\mu\text{H}} = \left[K_1 N^2 A \text{Ln}\left[\frac{K_2 AN}{(N+1)B}\right]\right] + K_3 + \left[\frac{K_4(N+1)B}{AN}\right]$$

where all terms are as previously defined, except K_1–K_4, which are defined in *Table 4.3*.

Table 4.3 *K* factors for calculating loop inductance

Shape	K_1	K_2	K_3	K_4
Triangle	0.006	1.1547	0.65533	0.1348
Square	0.008	1.4142	0.37942	0.3333
Hexagon	0.012	2.00	0.65533	0.1348
Octagon	0.016	2.613	0.75143	0.07153

Patterson equation:

$$L\mu\text{H} = (0.00508A) \times \left[2.303 \log\left(\frac{4A}{d}\right) - \phi\right]$$

where

d is the conductor diameter

ϕ is a factor found in *Table 4.4*

Table 4.4 ϕ Factor for calculating loop inductance

Shape	*Factor (ϕ)*
Circle	2.451
Octagon	2.561
Hexagon	2.66
Pentagon	2.712
Square	2.853
Triangle	3.197

Of these equations, most people will find that the Grover equation most accurately calculates the actual inductance realized when a practical loop is built.

References

Andrew Antennas (1991). *Catalog 35*. Illinois.

Belcher, R. *et al*. (1989). *Newnes Mobile Radio Servicing Handbook*. Butterworth-Heinemann, Oxford.

Davies, J. (1987). *Private Mobile Radio – A Practical Guide*. Heinemann, London.

Green, D.C. (1979). *Radio systems, Vols II and III*. Pitman, London.

Kennedy, G. (1977). *Electronic Communication Systems*. McGraw-Hill Kogashuka, Tokyo.

Terman, F.E. (1943). *Radio Engineers' Handbook*. McGraw-Hill, London.

5 Resonant circuits

5.1 Series and parallel tuned circuits

Tuned resonant circuits composed of inductance and capacitance are used to generate alternating voltages of a specific frequency and to select a wanted frequency or band of frequencies from the spectrum. *Figure 5.1* contains the diagrams of series and parallel resonant circuits including the resistances which account for the losses present in all circuits. In practice the greatest loss is in the resistance of the inductor, R_L, with a smaller loss, r_c, occurring in the dielectric of the capacitor.

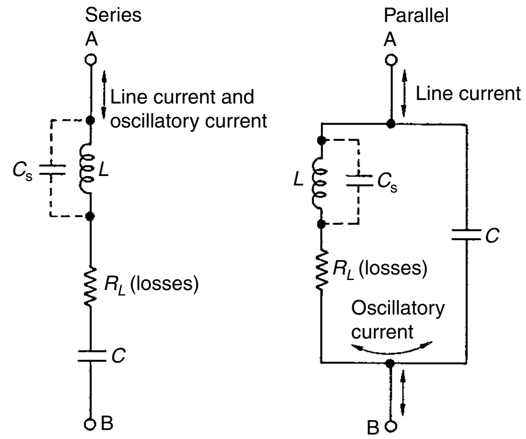
Series
A
Line current and oscillatory current
C_s
L
R_L (losses)
C
B
Parallel
A
Line current
L
C_s
C
R_L (losses)
Oscillatory current
B

Figure 5.1 Series and parallel resonant circuits

5.1.1 Series resonance

Off resonance, the series circuit exhibits a high impedance to a voltage applied across A and B. This impedance is formed by the vectorial addition of the reactances of the inductance and capacitance at the applied frequency plus the resistances. Ignoring the dielectric losses and the very small stray shunt capacitance C_s, the resonant impedance is given by:

$$Z = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}$$

where

R = resistance of components in ohms
$\omega = 2\pi \times$ frequency in hertz
L = inductance in henries
C = capacitance in farads

At resonance the reactances cancel out and the impedance falls to approximately the value of the resistance, R, and a maximum line current will flow. *Figure 5.2* shows the response curves of series and parallel circuits near resonance. The resonant frequency is given by:

$$f = \frac{1}{2\pi\sqrt{LC}}$$

where

f = resonant frequency in hertz
L = inductance in henries
C = capacitance in farads

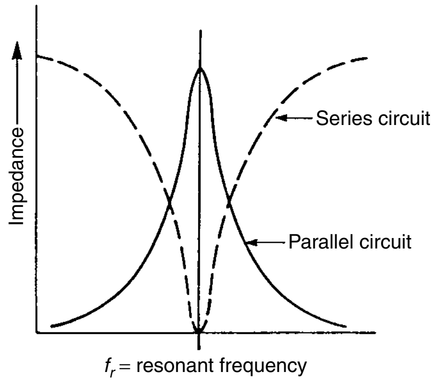

Figure 5.2 Variation of impedance around resonance with series and parallel tuned circuits

5.1.2 Parallel resonance

At resonance, calculated by using the same formula as for a series circuit, the impedance of a parallel tuned circuit is also resistive, and the circulatory current in the circuit is high producing the maximum voltage across the inductance and capacitance. Consequently, at resonance the minimum line current flows. The impedance at resonance or dynamic resistance of a parallel tuned circuit of moderate to high Q is given by:

$$R_{\mathrm{d}} = \frac{L}{CR}$$

where
R_{d} = dynamic resistance of circuit
R = resistance of components in ohms
L = inductance in henries
C = capacitance in farads

5.2 *Q* factor

The voltages across the inductor and capacitor in a circuit at resonance are substantially opposite in phase (the loss resistances affect the phasing slightly) and cancel each other. The voltage developed across the inductor, usually the lossiest component, compared with the voltage applied in series to the circuit, is a measure of the 'goodness' of the circuit. This ratio is often referred to as the magnification factor, Q, of the circuit. The Q factor is calculated from:

$$Q = \frac{\omega L}{R} \text{ or } \frac{1}{\omega CR}$$

where L and C are in henries and farads respectively.

5.3 Coupled (band-pass) resonant circuits

5.3.1 Methods of coupling

Radio signals carrying intelligence occupy a band of frequencies and circuits must be able to accept the whole of that band whilst rejecting all others. When tuned circuits are coupled together in the correct manner they form such a band-pass circuit. The coupling may be either by mutual inductance between the two inductances of the circuits as in *Figure 5.3* or through discrete electrical components as in *Figure 5.4*.

Mutual inductance can be defined in terms of the number of flux linkages in the second coil produced by unit current in the first coil. The relationship is:

$$M = \frac{\substack{\text{flux linkages in 2nd coil} \\ \text{produced by current in 1st coil}}}{\text{current in 1st coil}} \times 10^{-8}$$

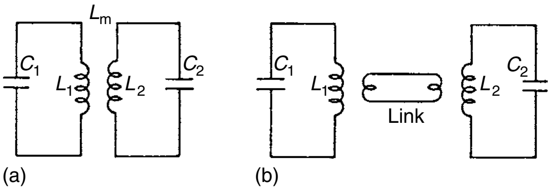

L_m
C_1
L_1
L_2
C_2
(a)
C_1
L_1
Link
L_2
C_2
(b)

Figure 5.3 Mutual inductance (low impedance) coupling

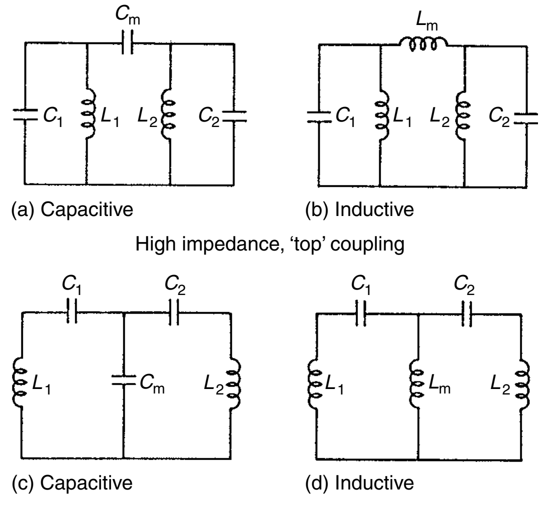

C_m
C_1
L_1
L_2
C_2
L_m
C_1
L_1
L_2
C_2
(a) Capacitive
(b) Inductive
High impedance, 'top' coupling
C_1
C_2
L_1
C_m
L_2
C_1
C_2
L_1
L_m
L_2
(c) Capacitive
(d) Inductive
Low impedance, 'bottom' coupling

Figure 5.4 Electrical coupling

where $M =$ mutual inductance in henries. The e.m.f. induced in the secondary is $e_2 = -j\omega M I_1$ where I_1 is the primary current.

The maximum value of mutual inductance that can exist is $\sqrt{L_1 L_2}$ and the ratio of the actual mutual inductance to the maximum is the coefficient of coupling, k:

$$k = \frac{M}{\sqrt{L_1 L_2}}$$

The maximum value of k is 1 and circuits with a k of 0.5 or greater are said to be close coupled. Loose coupling refers to a k of less than 0.5.

An advantage of coupling using discrete components is that the coupling coefficient is more easily determined. Approximations for k for the coupling methods shown in *Figure 5.4* are:

(a) $k = \dfrac{C_m}{\sqrt{C_1C_2}}$ where C_m is much smaller than (C_1C_2)

(b) $k = \dfrac{\sqrt{L_1L_2}}{L_m}$ where L_m is much larger than (L_1L_2)

(c) $k = \dfrac{\sqrt{C_1C_2}}{C_m}$ where C_m is much larger than (C_1C_2)

(d) $k = \dfrac{L_m}{\sqrt{L_1L_2}}$ where L_m is much smaller than (L_1L_2)

5.3.2 Response of coupled circuits

When two circuits tuned to the same frequency are coupled their mutual response curve takes the forms shown in *Figure 5.5*, the actual shape depending on the degree of coupling.

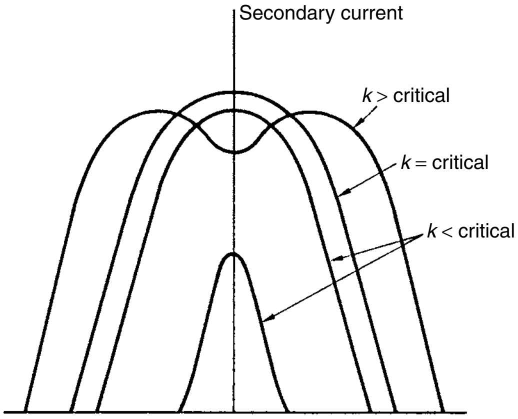

Figure 5.5 Effect of degree of coupling

When coupling is very loose the frequency response and the current in the primary circuit are very similar to that of the primary circuit alone. Under these conditions the secondary current is small and the secondary response curve approximates to the product of the responses of both circuits considered separately.

As coupling is increased the frequency response curves for both circuits widen and the secondary current increases. The degree of coupling where the secondary current attains its maximum possible value is called the critical coupling. At this point the curve of the primary circuit shows two peaks, and at higher coupling factors the secondary response also shows two peaks. The peaks become more prominent and further apart as coupling is increased, and the current at the centre frequency decreases.

References

Langford-Smith, F. (1955). *Radio Designer's Handbook*, Iliffe and Sons, London.
Terman, F.E. (1943). *Radio Engineers' Handbook*. McGraw-Hill, London.

6 Oscillators

6.1 Oscillator requirements

Oscillators generate the frequencies used in radio and electronic equipment. The performance of those which determine the operating frequencies of radio systems is tremendously important. Most oscillators must:

- Generate a precise frequency of high purity.
- Be highly stable, i.e. produce an output constant in frequency and level despite changes in temperature, supply voltage and load.
- Be tunable in frequency.
- Produce minimum noise and microphony (fluctuations in frequency with vibration).

These requirements conflict. A readily tunable oscillator cannot be precise and highly stable, and compromises must be made; either a less stringent specification must be accepted where permissible or the facility for tuning restricted.

Not all oscillators need to produce a pure output, devoid of spurious frequencies. The clock generators in digital circuitry, for instance, produce square waves, but a radio transmitter carrier generator and receiver local oscillator must produce a pure sine wave output if spurious radiations and receiver responses are to be avoided.

6.2 Tunable oscillators

An oscillator is an amplifier with a portion of the output fed back to the input. The feedback must be positive, i.e. in phase with the input, and the loop gain, input back to input via the feedback loop, must be sufficient to overcome the losses in the circuit.

Most radio frequency oscillators – and some audio ones – use inductance and capacitance (LC) tuned circuits as the frequency determining elements. *Figure 6.1* shows two commonly used basic circuits, the Hartley and the Colpitts.

The frequency is determined by the values of L and C_1 (the combined values of C_1 in the Colpitts circuit) and the amount of feedback by the collector choke and C_2. Such circuits produce a very pure output but, principally because the physical dimensions of the frequency

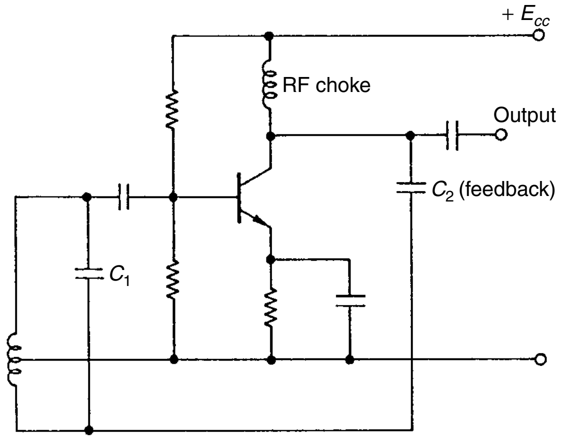

(a) Hartley circuit

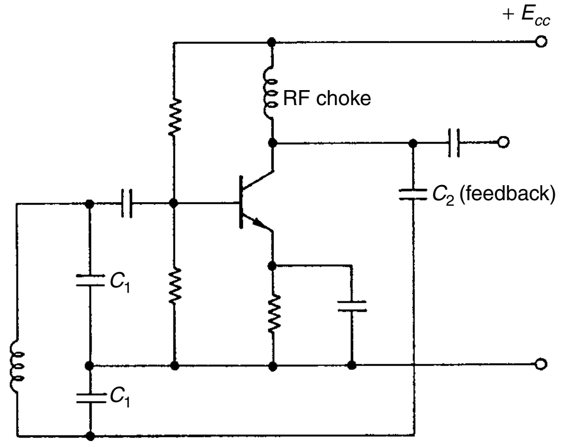

(b) Colpitts circuit

Figure 6.1 Hartley and Colpitts oscillators

determining components change with temperature, the accuracy of the set frequency is doubtful and is not very stable. Temperature compensation can be applied by selecting a capacitor for C_1 with the correct negative temperature coefficient (assuming that the inductance increases with temperature), inserting temperature compensation for the rise in collector current with temperature and stabilizing the supply voltage. In the design of equipment an oscillator should be built into an area of low temperature change.

6.3 Quartz crystal oscillators

The problems of frequency accuracy and stability are largely overcome by using a quartz crystal as the frequency determining element (see Chapter 7).

Figure 6.2(a) shows a circuit for an oscillator using the crystal's parallel resonant mode. In this circuit, the rising voltage developed across R_e on switch-on is applied via C_1 to the base accelerating the rise of current through the transistor. When saturation is reached the voltage across R_e becomes static and the voltage on the base falls, reducing the transistor current. The oscillations are only sustained

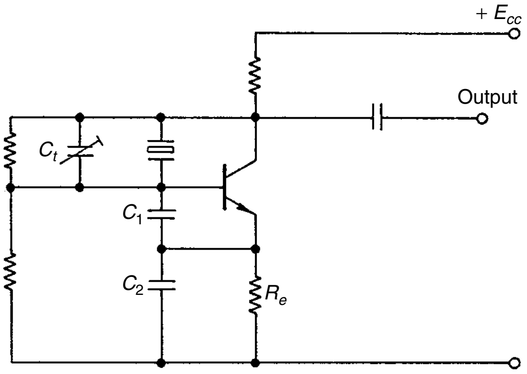

(a) Parallel resonant

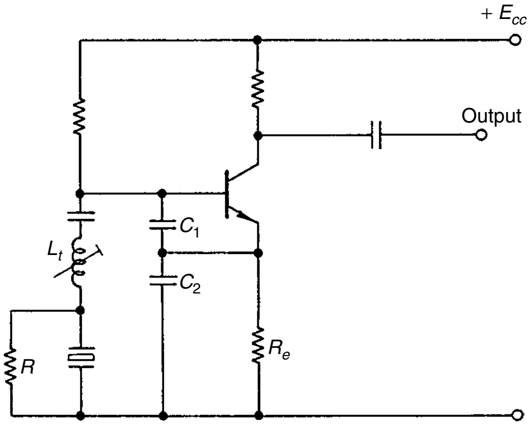

(b) Series resonant

Figure 6.2 Quartz crystal oscillators

at the parallel resonant frequency of the crystal where it presents a high impedance between base and collector. C_t enables the parallel resonance of the crystal to be adjusted to a precise frequency.

Figure 6.2(b) shows a series resonant crystal oscillator and here L_t is the tuning inductor of a Colpitts oscillator. The loop gain is adjusted so that the circuit will oscillate only at the series resonant frequency of the crystal where it presents a very low resistance. At other frequencies the crystal presents an increasing impedance in series with L_t, shunted by R which can be of a low value. When first setting the oscillator, L_t is adjusted, with the crystal short-circuited, for oscillation at a frequency close to that desired. The short-circuit is then removed and L_t used as a fine trimmer.

The same circuit will operate at the parallel resonant frequency of the crystal by making C_1 equal to the crystal load capacitance.

The maximum frequency error permitted by the British Radiocommunications Agency specification MPT 1326 for mobile radio equipment designed for 12.5 kHz channel separation in the band 100–300 MHz is plus or minus 1.5 MHz. This is an overall accuracy of 0.0005% over the temperature range $-10\,^\circ$C to $+55\,^\circ$C. Well-designed standard crystal oscillators meet this specification, but higher stability can be obtained by operating the crystal in an oven at a constant higher temperature.

Until recently equipment which was required to change operating frequency quickly was fitted with several crystals, one for each operating frequency, and a change of frequency was made by selecting the appropriate crystal. Frequency synthesizer circuits are now normally used for such applications.

6.3.1 Overtone oscillators

Piezoelectric crystals can oscillate at more than one frequency. The oscillations of a crystal slab are in the form of *bulk acoustic waves* (BAWs), and can occur at any frequency that produces an odd half-wavelength of the crystal's physical dimensions (e.g. $1\lambda/2$, $3\lambda/2$, $5\lambda/2$, $7\lambda/2$, $9\lambda/2$, where the fundamental mode is $1\lambda/2$). Note that these frequencies are not exact harmonics of the fundamental mode, but are actually valid oscillation modes for the crystal slab. The frequencies fall close to, but not directly on, some of the harmonics of the fundamental (which probably accounts for the confusion). The overtone frequency will be marked on the crystal, rather than the fundamental (it is rare to find fundamental mode crystals above 20 MHz or so, because their thinness makes them more likely to fracture at low values of power dissipation).

The problem to solve in an overtone oscillator is encouraging oscillation on the correct overtone, while squelching oscillations at the fundamental and undesired overtones. Crystal manufacturers can help with correct methods, but there is still a responsibility on the part of the oscillator designer. It is generally the case that overtone oscillators will contain at least one $L-C$ tuned circuit in the crystal network to force oscillations at the right frequency.

6.4 Frequency synthesizers

Frequency synthesizers offer the stability of a quartz crystal oscillator combined with the facility to change operating frequency rapidly. They are essential for equipment operating on trunked or cellular networks where the frequency of the mobiles is changed very rapidly on instructions from the network.

6.4.1 Voltage controlled oscillators

The advent of the variable capacitance diode (varicap), where the capacitance across the diode varies according to the applied DC voltage, made the frequency synthesizer a practicality.

When a varicap diode replaces the tuning capacitor in an oscillator the circuit becomes a DC-voltage-controlled oscillator (VCO). The two varicaps in *Figure 6.3* are used to minimize harmonic production and to obtain a greater capacitance change per volt.

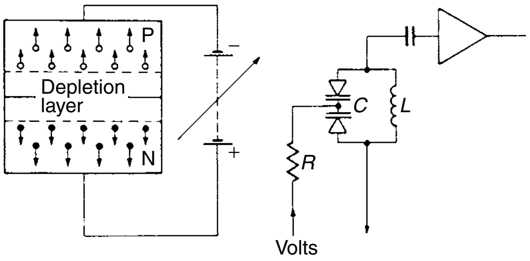

Figure 6.3 Voltage variable capacitance diode

The VCO is the circuit that directly generates the output frequency of a frequency synthesizer but, by itself, inherits the problem of frequency stability. To overcome this, the frequency and phase of the VCO are compared with those of a crystal-controlled high stability oscillator. Any frequency or phase difference between the two oscillators creates a DC voltage of the correct sense to change the frequency of the VCO to agree with that of the crystal oscillator.

While the stability of the crystal oscillator is transferred to the synthesizer output, additional noise is produced close to the operational frequency and the elimination of microphony requires careful physical design.

6.4.2 Phase-locked loops

Figure 6.4(a) and *(b)* are diagrams of a simple phase-locked loop (PLL). The outputs of both the crystal oscillator and the voltage-controlled oscillator are fed to the phase comparator which produces pulses whenever there is a frequency or phase difference between the two inputs. The pulses will be either positive or negative depending on the sense of the difference, and their width is dependent upon the magnitude of the differences. The pulses are then fed to a low pass loop filter which smooths them. If the time constant of the filter is sufficiently long it will completely remove the pulses and produce a DC output proportionate to the input pulse width which is applied to the VCO in the right sense to correct the frequency error. To enable the pulses to swing the VCO frequency in either direction, a small bias voltage of about 4 volts is applied to the varicap.

Identical frequencies have been selected for both oscillators in *Figure 6.4* but in practice this is seldom the case. More frequently, the

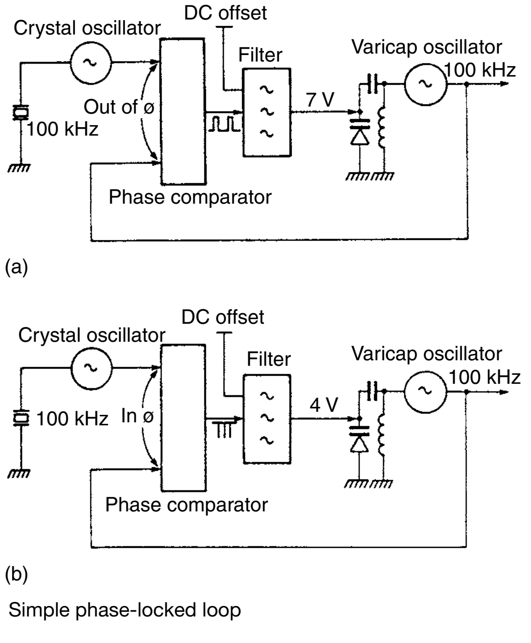
DC offset
Crystal oscillator
Filter
Varicap oscillator
100 kHz
Out of ø
100 kHz
7 V
Phase comparator
(a)
DC offset
Crystal oscillator
Filter
Varicap oscillator
100 kHz
100 kHz
In ø
4 V
Phase comparator
(b)

Figure 6.4 Simple phase-locked loop

VCO runs at a higher frequency than the crystal oscillator and a divider is used to equate the frequencies applied to the phase comparator (*Figure 6.5*). Changing the division ratio provides a convenient means of tuning the oscillator.

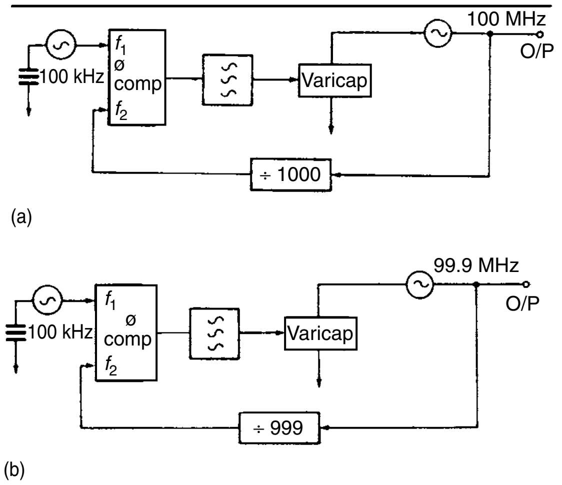

Figure 6.5 Frequency variable phase-locked loop

While the division ratio is 1000, the VCO will run at 100 MHz but if the division ratio is changed to 999 the comparator will produce pulses which, when converted to a DC voltage by the loop filter, will change the frequency of the VCO to 99.9 MHz, and the loop will lock at this new frequency.

The design of the loop filter is critical. Too long a time constant lengthens the settling time when changing frequency, yet if it is too short any deliberate frequency modulation will be removed. In practice, a relatively long time constant is chosen which is shortened by a 'speed up' circuit introduced whenever a channel change is called for.

The above values would enable a radio operating on a system with a channel separation of 100 kHz to change channel, but mobile radio channel separations are 25 kHz, 12.5 kHz or even 6.25 kHz at frequencies from 50 MHz to at least 900 MHz. To change channel at these frequencies a synthesizer must use a high division ratio. With a reference frequency applied to the comparator of 6.25 kHz and an operating frequency of, say, 450 MHz, the frequency select divider must have a ratio of 72 000, and be programmable in steps of 1 with a minimum operating speed of at least 900 MHz. A problem is then that the technology capable of meeting these requirements, emitter coupled logic (ECL), is power hungry, and the preferred LSI low power technology, CMOS, has a maximum operating speed of about 30 MHz. A simple ECL pre-scaler to bring the VCO frequency to about 30 MHz needs a ratio of 20 (500 MHz to 25 MHz). However, every change of 1 in the CMOS divider ratio then changes the total division ratio by 20. The solution is to use a dual modulus pre-scaler.

6.4.3 Dual modulus pre-scaler

The division ratio of the dual modulus pre-scaler (*Figure 6.6*) is programmable between two consecutive numbers, e.g. 50 and 51 (P and $P + 1$) and, in conjunction with two CMOS dividers, ÷A and ÷N, provides a fully programmable divider.

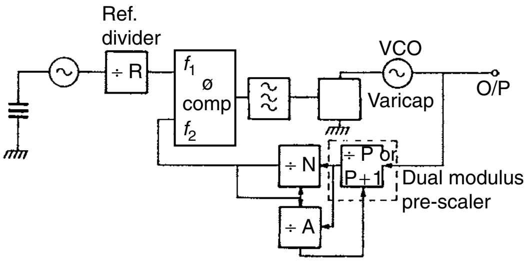

Figure 6.6 Programmable frequency synthesizer

The A and N dividers are pre-loaded counters. These count down and when the count value reaches zero they produce an output which changes the division ratio of the pre-scaler. The total division ratio, N_t, is decided by the initial programmed contents of the A and N counters and the setting of the pre-scaler. The initial content of the A counter must be less than that of the N counter.

Consider the pre-scaler set to divide by $P + 1$. For every count of $P + 1$, the contents of the A and N counters are reduced by 1 until the contents of the A counter are zero. The difference between the original contents of the A and N counters, $N - A$, remains in the N counter, and the total count, N_t, up to now, is $A(P + 1)$. At this point the division ratio of the pre-scaler is changed to P. Now, for every P count, the contents of the N counter are reduced by 1 until zero is reached. Under these conditions the total division ratio is given by:

$$\begin{aligned} N_t &= A(P + 1) + (N - A)P \\ &= AP + A + NP - AP \\ &= NP + A \end{aligned}$$

For example, let $P = 50$ so $P + 1 = 51$, let $N = 10$ and $A = 7$. Then:

$$N_t = 10 \times 50 + 7 = 507$$

Now, change A to 6:

$$N_t = 10 \times 50 + 6 = 506$$

a change of N_t by 1.

Programming a divider

Example:

$$\text{VCO frequency} = 455.6\,\text{MHz}$$

$$\text{Reference frequency} = 12.5\,\text{kHz}$$

Calculate N_t, and the numbers which must be programmed into the A and N counters, assuming $P = 50$:

1. Calculate $N_t = 455.6\,\text{MHz}/12.5\,\text{kHz} = 36\,448$.
2. Divide N_t by P: $36\,488/50 = 728.96$. Make $N = 728$.
3. For A, multiply fraction by P: $0.96 \times 50 = 48$.
4. Check $N_t = NP + A = 728 \times 50 + 48 = 36\,448$.
   Change A to 47: $NP + A = 728 \times 50 + 47 = 36\,447$.
   $36\,447 \times 12.5\,\text{kHz} = 455.5875\,\text{MHz}$, the adjacent 12.5 kHz channel.

6.4.4 Direct digital synthesis

A method of direct digital frequency synthesis replaces the voltage-controlled oscillator by a numerically controlled oscillator (NCO) where the function of the VCO is digitally synthesized.

The direct digital synthesizer generates an analogue sine wave from digital sine wave samples applied to a digital to analogue (D/A) converter. There are limitations to the method in terms of bandwidth and spectral purity.

6.5 Caesium and rubidium frequency standards

Where extra high stability is required for, say, laboratory standards or in quasi-synchronous wide area coverage systems, oscillators utilizing the atomic resonances of substances like caesium and rubidium, although expensive, may be employed.

Caesium oscillators are used to provide standard frequencies such as 1, 5 and 10 MHz with accuracies of $\pm 7 \times 10^{-12}$ over a temperature range of 0 to 50 °C with a long-term stability of 2×10^{-12}.

Rubidium oscillators are used to provide secondary standards and in some quasi-synchronous radio systems. Their accuracy is less than that of caesium, the long-term drift being of the order of 1×10^{-11} per month.

References

Belcher, R. *et al*. (1989). *Newnes Mobile Radio Servicing Handbook*. Butterworth-Heinemann, Oxford.

Winder, S.W. (2001). *Newnes Telecommunications Pocket Book*. Butterworth-Heinemann, Oxford.

7 Piezo-electric devices

7.1 Piezo-electric effect

When electrical stress is applied to one axis of a quartz crystal it exhibits the piezo-electric effect: a mechanical deflection occurs perpendicular to the electric field. Equally, a crystal will produce an e.m.f. across the electrical axis if mechanical stress is applied to the mechanical axis. If the stress is alternating – the movement of the diaphragm of a crystal microphone is an example – the e.m.f. produced will be alternating at the frequency of the movement. If the stress alternates at a frequency close to the mechanical resonance of the crystal as determined by its dimensions, then large amplitude vibrations result. Polycrystalline ceramics possess similar qualities.

Quartz crystals used for radio applications are slices cut from a large, artificially grown crystal. The slices are then ground to the appropriate size to vibrate at a desired frequency. The performance of an individual slice – the crystal as the end user knows it – depends upon the angle at which it was cut from the parent crystal.

Each crystal slice will resonate at several frequencies and if the frequency of the stimulus coincides with one of them the output, electrical or mechanical, will be very large.

The vibrations occur in both the longitudinal and shear modes, and at fundamental and harmonic frequencies determined by the crystal dimensions.

Figure 7.1A shows a typical natural quartz crystal. Actual crystals rarely have all of the planes and facets shown. There are three *optical axes* (X, Y and Z) in the crystal used to establish the geometry and locations of various cuts. The actual crystal segments used in RF circuits are sliced out of the main crystal. Some slices are taken along the optical axes, so are called Y-cut, X-cut and Z-cut slabs. Others are taken from various sections, and are given letter designations such as BT, BC, FT, AT and so forth.

7.1.1 Equivalent circuit of a quartz crystal

A quartz crystal behaves similarly to a very high Q tuned circuit and the equivalent circuit of a crystal is shown in *Figure 7.1B*.

C_1 and L_1 are equivalent to the inductance and capacitance of a conventional tuned circuit and R_1 represents the losses in the quartz

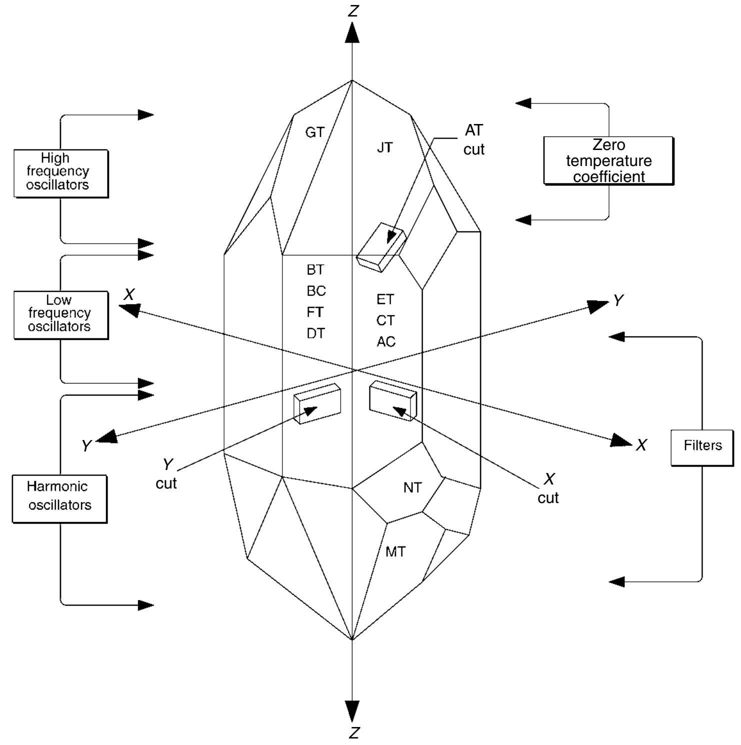
Z
GT
JT
AT
cut
Zero
temperature
coefficient
High
frequency
oscillators
Low
frequency
oscillators
Harmonic
oscillators
BT
BC
FT
DT
ET
CT
AC
X
Y
Y
X
Y
cut
X
cut
Filters
NT
MT
Z

Figure 7.1A Natural quartz crystal

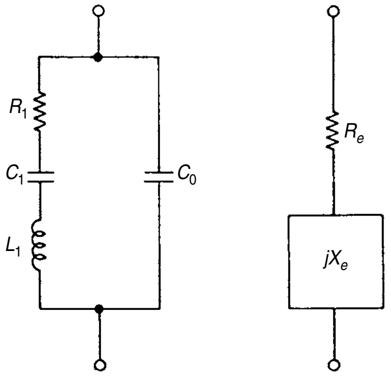
R_1
C_1
L_1
C_0
R_e
jX_e

Figure 7.1B Equivalent circuit of a crystal

and the mounting arrangements. C_0, typically 3–15 pfd, represents the shunt capacitance of the electrodes in parallel with the can capacitance. If the oscillatory current is considered, the resonant frequency is decided by the values of C_0 in series with C_1, L_1 and R_1, and all crystals basically resonate in a series mode. *Figure 7.2* illustrates the changes in impedance close to resonance. However when a high impedance, low capacitance, load is connected across the crystal terminals it behaves as a parallel tuned circuit exhibiting a high resistance at the resonant frequency. A crystal operating in the parallel mode oscillates at a higher frequency than that of series resonance.

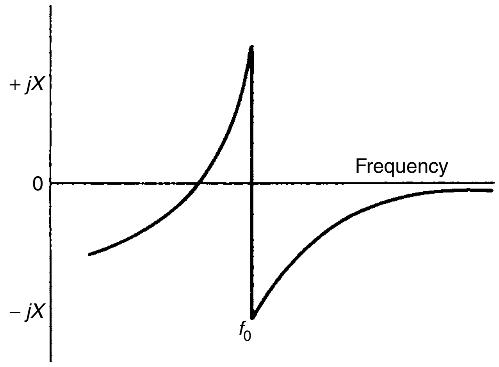

Figure 7.2 Crystal reactance close to resonance

A crystal will resonate at its fundamental frequency or at one or more of its harmonics. As the desired resonant frequency is increased, a crystal slice operating at its fundamental frequency becomes extremely thin and fragile. Consequently, overtone crystals are composed of larger slices of quartz operating close to, but not necessarily at, an exact harmonic of the fundamental frequency. Crystals operating at the 3rd, 5th and 7th harmonics are often employed at frequencies above approximately 25 MHz.

7.2 Quartz crystal characteristics

7.2.1 Resonant frequency

The resonant frequency is determined by the mass of the finished crystal which can be adjusted by grinding and the deposition of gold or other metal onto the crystal faces during manufacture. The adjustment is made to suit the intended method of operation, series or parallel, and at a specific temperature, usually 25°C. When parallel mode is specified, allowance is made for the load or circuit capacitance, usually 20–30 pfd, in parallel with C_0.

7.2.2 Frequency stability

Temperature coefficient

A crystal's resonant frequency varies with temperature and this temperature coefficient is determined by the angle at which the slice was cut from the parent crystal. Commonly used cuts are AT and BT. Because of its better performance AT is the most common.

Typical examples of the temperature coefficients for these are shown in *Figure 7.3*.

The temperature coefficient is specified, usually in parts per million (ppm) per degree C, or as a percentage, over a defined temperature range. The standard European temperature range is −10°C to +60°C. A crystal designed for a restricted temperature range has a better

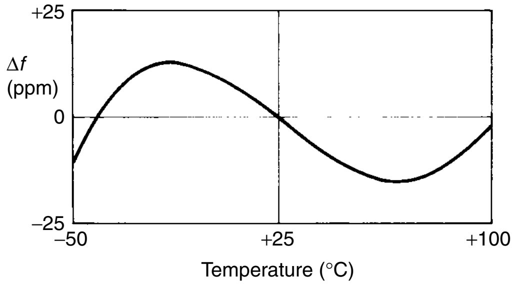

AT cut crystal

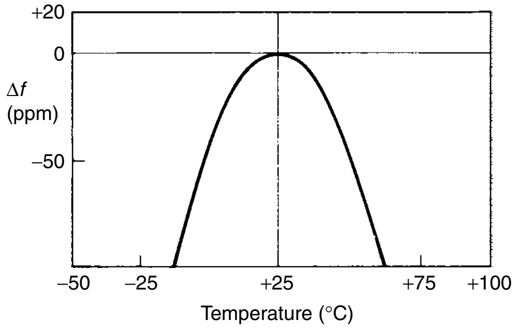

BT cut crystal

Figure 7.3 Frequency vs. temperature curves AT and BT cut crystals

stability over that range than one designed for operation over a wide temperature range will have when used over a restricted range.

For higher frequency stability crystals may be operated in a temperature-controlled oven operating at a more constant high temperature.

Common frequency tolerance specifications are ±0.005% or 0.0025% from −55°C to +105°C. These include the frequency errors from all sources, including the calibration tolerance; thus, the temperature coefficient is slightly better than these figures.

Ageing

The resonant frequency shifts with age from that set at production, following a curve similar to that in *Figure 7.4*. Initially the frequency shift for a given period of time is rapid but slows with age. The frequency may shift in either direction, and although it is possible to specify crystals ageing in one direction – high stability oscillators for quasi-synchronous transmission systems is an application – they are selected from a batch, not specifically manufactured. Once a crystal has been operated, a subsequent long period of inactivity can produce a glitch in the ageing curve followed by a higher rate of change for a short time.

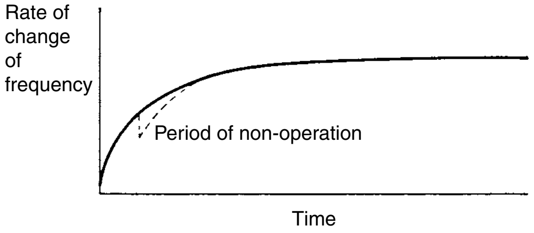

Figure 7.4 Effect of ageing

7.2.3 Load capacitance and pullability

When a crystal is operated in the parallel mode across a low capacitance load the results are a higher frequency and larger output voltage to the load. Increasing the load capacitance causes a reduction in frequency approaching that of series resonance.

The change in frequency that can be achieved by varying the load capacitance – a small trimmer capacitor is often connected across the crystal for this purpose – is the crystal's pullability. A typical pullability is from −1 ppm/pfd to −20 ppm/pfd for a total shunt capacitance of 40 pfd ($C_0 + C_{\text{load}}$).

7.2.4 Activity, effective series resistance (ESR) and Q

All these characteristics are interrelated. A crystal's activity, its vibrational response, can be quoted in terms of the effective series resistance. A higher effective series resistance implies lower activity, lower output and lower Q. The usual range of ESRs is from $20\,\Omega$ to $100\,\Omega$ although higher values occur in some low frequency crystals. Some manufacturers may quote activity levels for crystals for use in a parallel mode as effective parallel resistance (EPR). The EPR is the value of the resistor which, if connected in lieu of the crystal in an oscillator, would give the same output level as the crystal. The higher the EPR, the greater the crystal activity and Q.

7.2.5 Spurious responses

Crystals will resonate at frequencies other than those of the fundamental and harmonic modes for which they were designed; *Figure 7.5* shows the overtone (harmonic) and some typical spurious responses. The spurious responses of overtone crystals can occur with very little separation from the desired overtone frequency requiring very careful oscillator design if they are to be avoided.

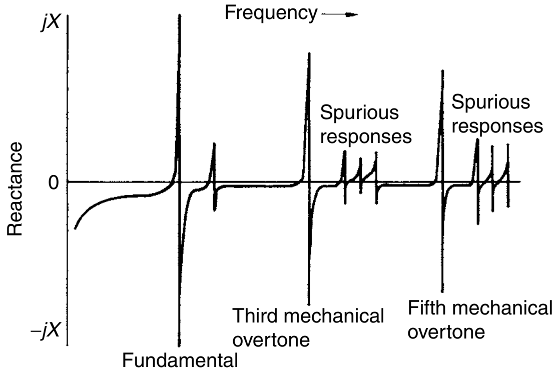

Figure 7.5 Overtone response of a quartz crystal

7.2.6 Case styles

A wide range of mounting styles is available. The American military nomenclature is widely used to describe them and *Figure 7.6* shows the outlines of some of the more popular styles.

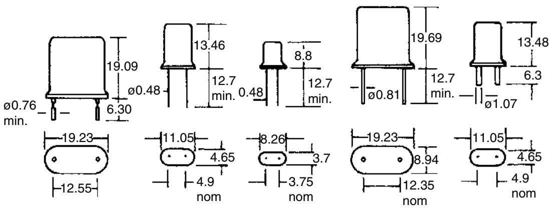
ø0.76
min.
19.09
ø0.48
6.30
19.23
12.55
13.46
12.7
min.
11.05
4.65
4.9
nom
0.48
8.8
12.7
min.
8.26
3.7
3.75
nom
19.69
12.7
min.
ø0.81
19.23
8.94
12.35
nom
13.48
6.3
ø1.07
11.05
4.65
4.9
nom

Solder seal	HC6/U AA	HC18/U J	– –	HC33/U BF	HC25/U CX	MIL IEC
Cold weld	HC36/U DNC	HC43/U DPC	HC45/U EBC	HC47/U AC	HC42/U DQ	MIL IEC
Glass	HC27/U DA	HC26/U CY	– –	– –	HC29/U CZ	MIL IEC
Res weld	HC48/U DNR	HC43/U DPR	HC45/U EBR	HC51/U RF	HC50/U RX	MIL IEC

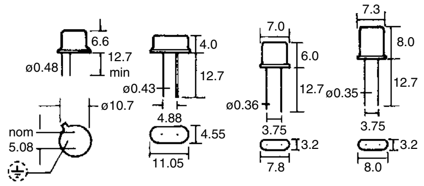
6.6
12.7
min
ø0.48
ø10.7
nom
5.08
4.0
12.7
ø0.43
4.88
4.55
11.05
7.0
6.0
12.7
ø0.36
3.75
3.2
7.8
7.3
8.0
12.7
ø0.35
3.75
3.2
8.0

Solder seal	– –	– –	– –	– –		MIL IEC
Cold weld	HC35/U DKC	– –	– –	– –		MIL IEC
Glass	– –	– –	– –	– –		MIL IEC
Res weld	HC35/U DKR	HC49/4H –	– UM-5	– UM-1		MIL IEC

Figure 7.6 Crystal case styles

7.3 Specifying quartz crystals

The details which must be specified when ordering crystals are:

1. *Frequency*. Normally specified in kHz up to 9999.999 kHz and in MHz from 10.0 MHz upwards except for integer values which are all specified in MHz. The frequency must be described to seven significant figures, otherwise any figure that might follow those given will be taken as zero.
2. *Mounting or holder style*.
3. *Frequency tolerance*. This is the cutting or calibration tolerance acceptable at 25°C. It should be borne in mind that cost rises with increased manufacturing accuracy and a slight adjustment (pullability) is possible in the circuit.

4. *Frequency stability*. Normally specified as a plus or minus value measured over a defined temperature range. A crystal designed for a restricted range has a better performance over that range than one designed for a wider range so it is important not to overspecify.
5. *Temperature range*. The range over which the crystal is required to operate and meet the performance specified in 4. Standard temperature ranges are:

   0 to 5°C
   −10 to 60°C
   −20 to 70°C
   −30 to 80°C
   −40 to 90°C
   −55 to 105°C
   −55 to 125°C

   It is sufficient when ordering from some manufacturers to quote only the lower temperature limit.

   For ovened operation the quoted figure, say 80°C, would denote the oven temperature.
6. *Circuit condition*. This specifies the shunt capacitance that the circuit will place across the crystal in parallel mode operation.
7. *Drive level*. The maximum power that the crystal can safely dissipate. 1 mW is a typical value for crystals used in radio transmitters and receivers.

A typical specification therefore reads:

16.66667 MHz	HC49	20	30	10	30
1	2	3	4	5	6

referring to the items listed above

When the crystal is for operation in series mode, it is usually sufficient to replace the last figure with ‘S’. The drive level is not normally specified in the ordering details.

7.4 Filters

Both quartz and ceramic materials are used in the production of radio frequency filters. Ceramic filters do not have the same performance as quartz but have the advantages of a lower cost. They are used at lower frequencies and where the higher stability and lower spurious responses of quartz are not essential.

Crystal filters are obtainable at frequencies up to about 45 MHz. Most of these use either a number of discrete crystals arranged in the form of a lattice or a monolithic structure. A single crystal will behave as an extremely narrow band filter and it is possible to use a crystal bar in this way down to a very few kilohertz.

The characteristics of filters can be divided into groups affecting the performance (*Figure 7.7*).

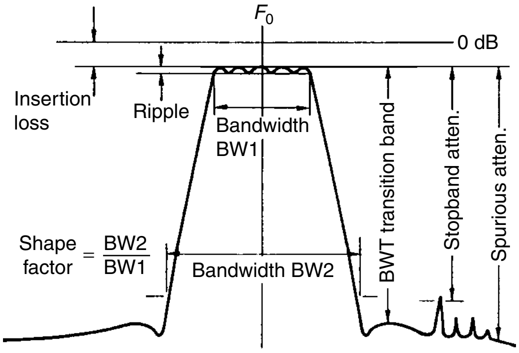

Figure 7.7 Filter characteristics

7.4.1 Passband performance

- *Insertion loss*. The loss at centre frequency, in dB, resulting from the insertion of the filter in a transmission system.
- *Flat loss*. The insertion loss at the frequency of minimum loss within the passband.
- *Attenuation*. The loss of a filter at a given frequency measured in dB.
- *Passband (bandwidth, BW_1)*. The range of frequencies attenuated less than a specified value, typically 3 or 6 dB.
- *Centre frequency* (f_0). The arithmetic mean of the passband limits.
- *Fractional bandwidth*. A specified frequency, typically the minimum loss point or F_0, from which all attenuation measurements are made.
- *Ripple*. The amplitude difference, in dB, between the maximum peak and minimum passband valley. Both the peak and the valley are defined by a surrounding change in slope, i.e. sign of the amplitude response. This is very important as a high ripple, particularly between a peak and the adjacent trough, produces rapid phase changes as the signal moves across the

passband resulting in audio distortion and corruption in digital signals.

7.4.2 Stopband performance

- *Attenuation*. The output of a filter at a given frequency relative to the defined insertion loss reference.
- *Stopband*. The range of frequencies attenuated by a greater amount than some specified minimum level of attenuation.
- *Transition band (bandwidth, BW_2)*. The range of frequencies differently attenuated between the passband and stopband limits.
- *Shape factor*. The ratio of the bandwidth at some point within the transition region, typically 60 dB, to the specified passband bandwidth. It is given by:

$$\text{Shape factor } \frac{BW_2}{BW_1}$$

- *Spurious attenuation*. The specified minimum level of attenuation received by all non-harmonic related resonances of each crystal resonator within the filter network.

7.4.3 Time domain performance

- *Insertion phase*. The phase shift at the output load (measured at the reference frequency) resulting from the insertion of the filter.
- *Differential phase*. The measurement of phase at a given frequency relative to the phase at the reference frequency.
- *Phase linearity*. The phase error in degrees between the phase points and a straight line drawn through the phase points.
- *Group delay*. The time by which a signal will be delayed before it appears at the filter output, i.e. the derivative of phase with respect to frequency.
- *Differential delay*. The measurement of delay at a given frequency relative to the reference frequency.

7.4.4 Source and load impedance

- *Source impedance*. The impedance of the circuit driving the filter, measured at the reference frequency.
- *Load impedance*. The impedance of the circuit terminating the filter at its output, measured at the reference frequency.

7.4.5 Non-linear effects

- *Maximum input level*. The driving point power, voltage or current level above which intolerable signal distortion or damage to the device will result.
- *Drive level stability*. The ability of the filter to return within a specified tolerance of its original insertion loss, at a specified drive level, after experiencing changing environmental and/or drive level conditions.
- *Drive level linearity*. The maximum permissible variation in insertion loss, per dB change in drive level, measured over a specific dynamic range.
- *Inband intermodulation distortion*. The attenuation, in dB, of third and higher order signal products, inband, relative to the power level of two signals placed within the passband.
- *Out-of-band intermodulation distortion*. The attenuation, in dB, of third and higher order signal products, inband, relative to the power level of two signals placed in the stopband, or one signal in the transition region and the other in the stopband.

A manufacturer's specifications for two stock 10.7 MHz filters are given in *Table 7.1*.

Table 7.1 Manufacturers' specifications for two stock 10.7 MHz filters

Centre freq.	*Passband width (plus/minus)*	*Attenuation bandwidth*	*Ripple (max)*	*Ins. loss (max)*	*Term. impedance ΩS/pfd*
10.7 MHz	3.75 kHz (6 dB)	8.75 kHz (45 dB) 12.5 kHz (60 dB)	2.0 dB	3.0 dB	1.5k/1
10.7 MHz	7.5 kHz (6 dB)	15.0 kHz (60 dB) 20.0 kHz (80 dB)	2.0 dB	4.0 dB	3k/1

7.5 SAW filters and resonators

The piezo-electric effect that some ceramic materials such as lithium niobate exhibit enables useful applications, such as pressure sensors and audible alarms. When piezo-electric material has an electric field applied across it, usually through metal electrodes, the material changes shape. This often takes the form of local material expansion or contraction (depending on the polarity of the applied voltage). In the case of the audible alarm, the material expansion and contraction modulates the air pressure to give an audible tone if the electric field is alternating at audio frequencies.

One form of the piezo-electric effect is the SAW (surface-acoustic-wave) phenomenon, in which signals travel due to an acoustic wave in the ceramic material. To see this effect the piezo-electric material has one side bonded to a metal plate. The upper surface has metal electrodes applied to it, to provide an electric energizing force. The surface wave is created when a material is forced to expand (by the piezo-electric effect) beneath the metal electrodes. Physical bonds between molecules then force adjacent material to expand and this expansion propagates through the material.

The piezo-electric effect is bi-directional, so that if the material between electrodes changes shape, an electric field is generated between the electrodes. This is the effect seen when a piezo-electric gas lighter is used: a sudden force across a piezo-electric element generates high voltages that force a spark to be produced. In the SAW device, the acoustic wave travels along the material and deforms the piezo-electric material between a second set of electrodes, which are located at the other end of the piezo-electric material. A voltage is then formed between these electrodes. By having two sets of electrodes, known as transducers, at opposite ends of a piece of piezo-electric material, a signal can be transmitted through the material by applying a voltage at one end and detecting it at the other.

The range of frequencies that propagate through the piezo-electric material can be controlled by suitable transmit and receive transducer spacing. The propagation frequencies can be further controlled by applying additional metallized areas (described later) between the transmit and receive transducers. Thus SAW devices could be used to make compact and low-cost filters.

SAW resonators use piezo-electric material that is free to vibrate in one direction. The speed of wave propagation and the dimensions of the material are such that the wave reflects back and forth, resonating at a certain frequency. This can replace quartz crystals in many oscillator circuits, where the frequency accuracy is not critical.

7.5.1 Fundamentals of SAW transversal filters

In its simplest form, a transversal SAW filter consists of two transducers with inter-digital arrays of thin metal electrodes deposited on a piezoelectric substrate, such as quartz or lithium niobate (see *Figure 7.8*). The electrodes that comprise these arrays are arranged to have alternate polarities, so that an RF signal voltage of the proper

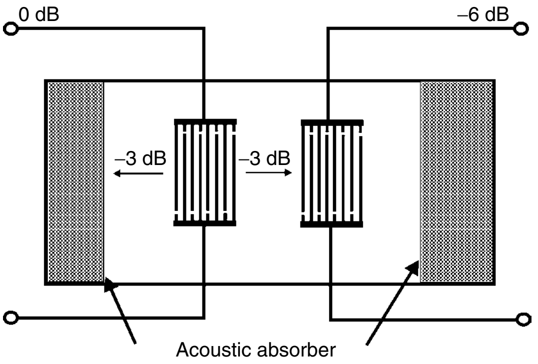

Figure 7.8 Simple transversal SAW filter configuration

frequency applied across them causes the surface of the crystal to alternately expand and contract along its length. This generates the Rayleigh wave, or surface wave, as it is more commonly called.

These inter-digital electrodes are generally spaced at half or quarter wavelength of the operating centre frequency. Since the surface wave or acoustic velocity is 10^5 slower than the speed of light, an acoustic wavelength is much smaller than its free space electromagnetic counterpart. For example, a CW signal at 1000 MHz with a free space wavelength of 300 mm would have a corresponding acoustic wavelength of about 3 μm. This wavelength compressing effect results in the SAW filter's unique ability to provide considerable signal processing or delay in a very small volume.

The wavelength compressing effect of the piezo-electric material has the effect of producing physical limitations at very high and low frequencies. At very high frequencies the electrodes become too narrow to fabricate with standard photo-lithographic techniques. At low frequencies the devices become impracticably large. Consequently, SAW devices are most typically used over the frequency range 10 MHz to about 3 GHz.

The basic SAW transducer is a bi-directional radiator. That is, half of the power (or -3 dB) is directed towards the output transducer while the other half is radiated towards the end of the crystal and is lost. By reciprocity, only half of the intercepted acoustic energy at the output is reconverted to electrical energy; hence, the inherent 6 dB loss associated with this structure (refer to *Figure 7.8*). Numerous second-order effects, such as coupling efficiency, resistive losses, and impedance mismatch, raise the insertion loss of practical filters to about 15 to 30 dB.

A new low-loss structure is the single-phase unidirectional transducer (SPUDT). Unlike the traditional three-phase unidirectional filters, the SPUDT devices are generally as straightforward to fabricate as ordinary bi-directional transducers. In addition, the required impedance matching network is greatly simplified, usually consisting of an L-C network on each port. Most SPUDT structures contain acoustic reflectors within the inter-digital pattern, illustrated schematically in *Figure 7.9* as wider electrodes. These internal reflectors serve to redirect most of the acoustic energy that is normally lost in a conventional bi-directional device. Second-order effects tend to limit the practical insertion loss to about 5 to 12 dB.

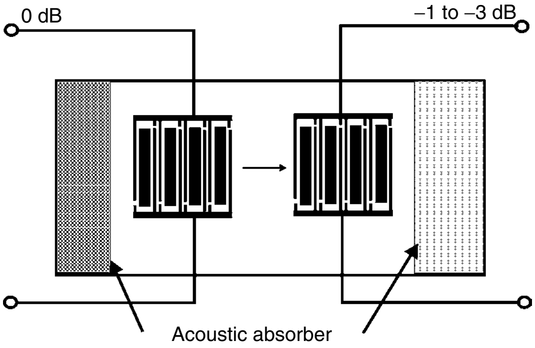

Figure 7.9 Single-phase unidirectional filter configuration

7.5.2 SAW device design

A better understanding of SAW device operation can be obtained by understanding how they are designed. SAW filters use finite impulse response design techniques, which are very similar to those used for digital filters. Hence, the Fourier transform is used to relate the time and frequency responses of the transducers and resultant filter. The desired total frequency response characteristics are used to find the impulse responses for the two transducers. These two impulse response shapes are then etched onto the surface of a metallized piezo-electric substrate to form transducers. When an impulse is applied to the transducer, the output wave shape is the same as the shape of the metallized transducer pattern.

To illustrate this relationship, the following examples may be valuable. First, consider the case of the uniform-overlap (or unweighted) transducer as shown in *Figure 7.10*. Since all fingers have the same

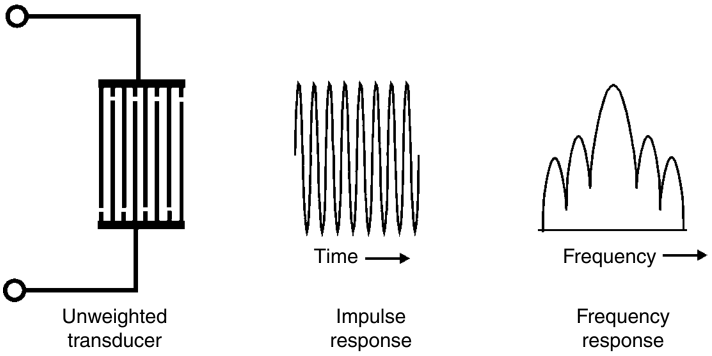

Figure 7.10 Uniform-overlap transducer

length, the impulse response envelope of this transducer is a rectangle. The Fourier transform of a rectangle is a $\sin(x)/x$ function with a 4 dB bandwidth roughly equal to the reciprocal of its time length. This illustrates the important trade-off between bandwidth and size, the narrower the bandwidth: the longer the transducer.

The $\sin(x)/x$ response has 13 dB first side-lobe attenuation, which results in 26 dB of rejection when two of these unweighted transducers are cascaded on the substrate. The 26 dB frequency rejection and poor skirt selectivity make the unweighted transducer suitable for delay lines, but unsuitable for most filter applications where more rejection is needed close to the passband edge.

According to Fourier transform theory, to obtain the rectangular frequency response, a transducer must be built that is infinitely long with electrode lengths that have $\sin(x)/x$ weighting. This is impossible, but the principle behind this example will always hold true: the steeper the skirts (the more rectangular the response) for a given bandwidth, the longer the device must be. The longest possible SAW device would exhibit a narrow bandwidth with a very steep roll-off. The transducer pattern must be limited in length but a gradual tailing off to zero length of the outer electrodes provides close to ideal performance (akin to the windowing function in digital FIR filters).

References

Crystal Product Data Book (1993). International Quartz Devices Ltd, Crewkerne, Somerset.

Winder, S.W. (2001). *Newnes Telecommunications Pocket Book*. Butterworth-Heinemann, Oxford.

8 Bandwidth requirements and modulation

8.1 Bandwidth of signals at base band

8.1.1 Analogue signals

The amount of information and the speed at which it is transmitted determines the base bandwidth occupied by a signal. For analogue signals, the base bandwidth is the range of frequencies contained in the signal; it is not the same as that occupied by a radio frequency carrier modulated by the signal. Examples of base bandwidths are given in *Table 8.1*.

Table 8.1 Base bandwidths

Application	*Frequency range (Hz)*
Speech	
High fidelity reproduction	15–15 000
Good fidelity	150–7000
Public address	200–5000
Restricted bass and treble	500–4000
Toll quality (good quality telephone line)	300–3400
Communications quality (radio communication)	300–3000
Mobile radio (12.5 kHz channel separation)	300–2700
Music (for FM broadcasting)	30–15 000
Video	60 Hz–4.2 MHz

8.1.2 Digital signals

Bit rate (b/s) and baud rate are terms used to specify the speed of transmitting digital information. Where the duration of all the signalling elements is identical the terms are synonymous, but not where the duration of the information bits differs.

As the term implies, the bit rate is the number of bits transmitted per second but the baud rate (after J.M.E. Baudot, the code's inventor) is the reciprocal of the length, in seconds, of the shortest duration signalling element. *Figure 8.1(a)* shows a binary code pattern where

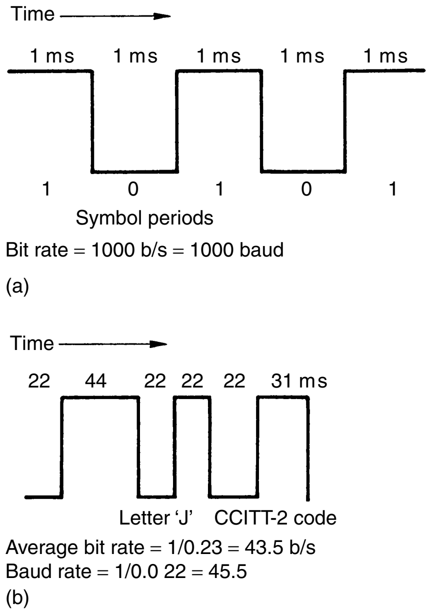
Time
1 ms 1 ms 1 ms 1 ms 1 ms
1 0 1 0 1
Symbol periods
Bit rate = 1000 b/s = 1000 baud
(a)
Time
22 44 22 22 22 31 ms
Letter 'J' CCITT-2 code
Average bit rate = 1/0.23 = 43.5 b/s
Baud rate = 1/0.0 22 = 45.5
(b)

Figure 8.1 Bit rate and baud rate

all the bits are of equal duration, in this case 1 millisecond; the bit rate is 1000 per second and the baud rate is $1/0.001 = 1000$ also.

On telegraphy systems all the bits may not be of the same duration and *Figure 8.1(b)* shows the pattern for the letter 'J' in the CCITT-2 code as used for teletype. In this code a letter is composed of 5 elements, each of 22 ms duration, but each letter is preceded by a space of 22 ms and followed by a mark of 31 ms. The duration of each character is 163 ms – the time for 7.5 elements – but is comprised of only 7 bits.

The baud rate is $1/0.022 = 45.5$ baud.

The average bit duration is $163/7 = 23.29$ ms.

The average bit rate is $1/0.023 = 43.5$ bits per second.

If the bandwidth were under consideration the baud rate, being faster, would be the figure to use.

A stream of binary coded information is composed of pulses where, say, a pulse (mark) represents digit 1 and the absence of a pulse (space) represents digit 0. The highest frequency contained in the information is determined by the bit – or baud – rate. Because a series of 1s or 0s may be consecutive in the data stream, the pulse repetition rate will vary throughout the message although the bit rate will be constant, and over a long period of time as many spaces will be sent as pulses. Transmitting either a stream of spaces or of marks requires no bandwidth; it is only when a change of state occurs that frequencies are produced. In the duration of one cycle (*Figure 8.1(a)*) two bits may be carried and the maximum *fundamental* frequency contained in the wave is one-half the number of bits per second, i.e. the channel capacity, bits/second, equals twice the bandwidth in hertz.

8.1.3 Channel capacity – Hartley – Shannon theorem

Channel capacity as stated by Hartley's law is, in the absence of noise:

$$C = 2\delta f \log_2 N$$

where
C = channel capacity, bits per second
δf = channel bandwidth, Hz
N = number of coding levels (2 in binary system)

When noise is present, the channel capacity calculated according to the Hartley–Shannon theorem is:

$$C = \delta f \log_2(1 + S/N)$$

where S/N = the ratio of total signal power to total noise power at the receiver input within the bandwidth, δf.

8.2 Modulation

For radio transmission, the low frequency information signal is carried on a radio frequency wave and it must change (modulate) that carrier. The modulation may change the amplitude, frequency or phase of the carrier. Modulation aims to achieve:

1. the transfer of information with the minimum distortion or corruption
2. the modulation of the carrier with the minimum loss of power
3. efficient use of the frequency spectrum.

8.3 Analogue modulation

8.3.1 Amplitude modulation (AM)

There are a number of methods of modulation where the amplitude of the carrier is varied by the information signal but the most commonly used is double sideband AM (DSB). *Figure 8.2* shows a radio frequency carrier modulated by a low frequency signal.

The amount or depth of modulation is expressed as percentage ratio, $m\%$, of the maximum to minimum amplitude:

$$\begin{aligned}\text{Mod. depth} &= m\% \\ &= \frac{\text{max. amplitude} - \text{min. amplitude}}{\text{max. amplitude} + \text{min. amplitude}} \times 100\%\end{aligned}$$

When the modulation is increased to the point where the minimum amplitude falls to zero, 100% modulation occurs. Any further increase in modulation produces spurious, out-of-band frequencies (AM splash), a source of interference for other radio users. For this reason, the depth of amplitude modulation is usually limited to 70%.

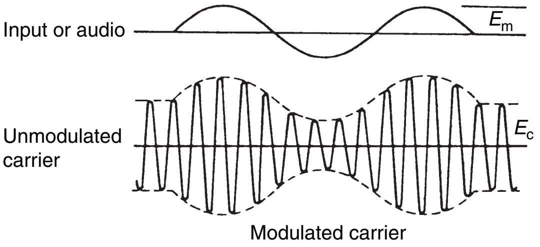

Figure 8.2 Amplitude modulation

An alternative expression for modulation depth is in terms of a modulation index from 0 to 1. The peak carrier voltage in *Figure 8.2* is E_c and the peak modulation voltage, E_m. The modulation index, m, is:

$$m = \frac{E_m}{E_c}$$

Amplitude modulation produces a band of frequencies above and below the carrier frequency – the upper and lower sidebands. The width of each sideband is equal to the highest modulating frequency so the bandwidth of an AM wave is 2 × the highest modulating frequency. To conserve spectrum, the range of modulating frequencies

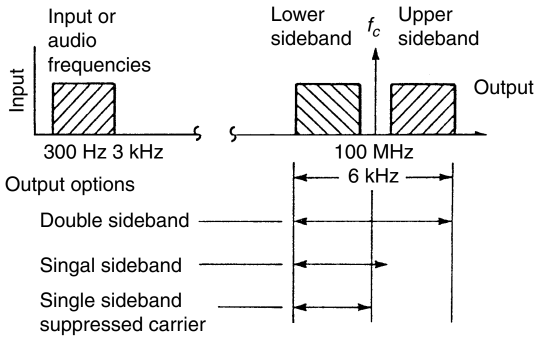
Input or audio frequencies
Input
300 Hz 3 kHz
Lower sideband
f_c
Upper sideband
Output
100 MHz
6 kHz
Output options
Double sideband
Singal sideband
Single sideband suppressed carrier

Figure 8.3 Amplitude modulation – sidebands

is restricted. For example, radio communication quality speech is limited to 300 Hz to 3000 Hz. The bandwidth occupied by a double sideband, amplitude modulated carrier for this service is 6 KHz (*Figure 8.3*).

Power relationships in an AM wave

The total power in an AM wave is the sum of powers of the carrier, the upper sideband and lower sideband:

$$P_t = \frac{E_{\text{carr}}{}^2}{R} + \frac{E_{\text{lsb}}{}^2}{R} + \frac{E_{\text{usb}}{}^2}{R}$$

where all values are RMS and R is the resistance in which the power is dissipated. From the peak voltages shown in *Figure 8.2* the power in the unmodulated carrier is:

$$P_c = \frac{E_c/\sqrt{2}^{\,2}}{R} = \frac{E_c^2}{2R}$$

The power in the sidebands

$$P_{\text{lsb}} = P_{\text{usb}} = \left(\frac{mE_c/2}{\sqrt{2}}\right)^2 \div R = \frac{m^2 E_c{}^2}{8R}$$

$$= \frac{m^2}{4}\frac{E_c{}^2}{2R}$$

The ratio of the total power in the wave to the carrier power is:

$$\frac{P_t}{P_c} = 1 + \frac{m^2}{2}$$

As m cannot exceed 1, the maximum RMS power in the wave is $P_t = 1.5P_c$; but if m reaches 1, the peak sum of E_c and E_m is $2E_c$ and so the instantaneous peak power is $2P_c$. Circuitry must be capable of handling this power level without distortion.

Double sideband amplitude modulation wastes power and spectrum. Two-thirds of the power is in the carrier which conveys no information and one sideband is discarded in the receiver. Also, the modulation must be accomplished either in the final power amplifier of the transmitter necessitating a high power modulator, or in an earlier, low power stage when all subsequent amplifiers must operate in a linear, but inefficient, mode.

8.3.2 Double sideband suppressed carrier (DSBSC)

In an amplitude modulated wave the carrier conveys no information yet contains 2/3 of the transmitted power. It is possible to remove the carrier by using a balanced modulator (*Figure 8.4*), and improve the power efficiency by this amount.

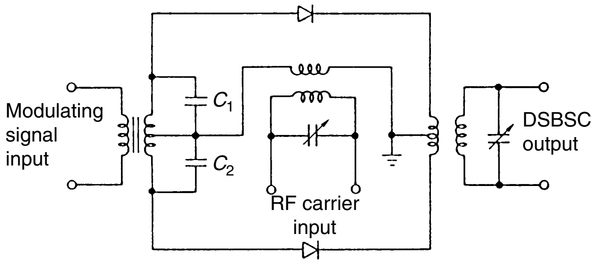

C_1 and C_2 are RF bypass capacitors

Figure 8.4 Balanced modulator

In a balanced modulator, the modulating voltage is fed in push–pull to a pair of matched diodes or amplifiers while the carrier is fed to them in parallel. The carrier components in the output cancel leaving the two sidebands. The result is a double sideband suppressed-carrier (DSBSC) wave, which is not sinusoidal, formed by the sum of the two sidebands. The carrier must be re-introduced in the receiver and its accuracy in both frequency and phase is critical.

8.3.3 Single sideband suppressed carrier (SSB or SSBSC)

The advantages of single sideband suppressed carrier transmission over double sideband AM are:

- removal of the carrier saves 2/3 of the total power
- removal of one sideband saves 50% of the remaining power
- an SSBSC transmitter only produces power when modulation is present
- the occupied bandwidth is halved; a spectrum saving
- the received signal-to-noise ratio is improved by 9 dB for a 100% modulated carrier. Halving the bandwidth accounts for 3 dB, the remainder from the improved sideband power to total power ratio. The S/N ratio improves further with lower modulation levels
- reduced susceptibility to selective fading and consequent distortion.

Two methods of generating a single sideband wave are in general use. One filters out the unwanted sideband after removal of the carrier by a balanced modulator. The other is a phase shift method (*Figure 8.5*). Here, the modulating signal is fed to two balanced modulators with a 90° phase difference. The output from both modulators contains only the sidebands but, while both upper sidebands lead the input carrier voltage by 90°, one of the lower sidebands leads it by 90° and the other lags it by 90°. When applied to the adder, the lower sidebands cancel each other while the upper sidebands add.

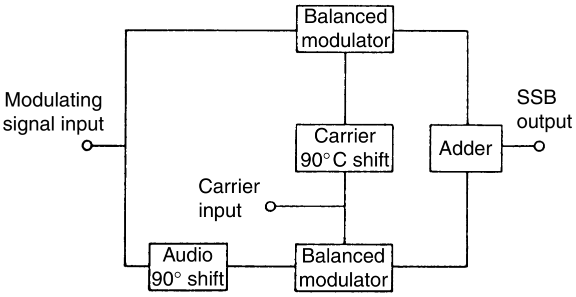

Figure 8.5 Phase shift production of single sideband

A single sideband AM wave modulated by a sinusoid consists of a constant amplitude signal whose frequency varies with the frequency of the modulating wave. Note that this is not the same as FM: the frequency in SSB does not swing to either side of the carrier. It is higher than the carrier frequency if the upper sideband is transmitted, and lower if the lower sideband is selected. The single sideband waveform is sinusoidal and, although the frequency of the reintroduced carrier must be highly accurate (±2 Hz), the phase is unimportant making a single sideband receiver less complex than one for DSBSC.

On some systems a pilot carrier is transmitted and the transmitter output power is then specified in terms of peak envelope power (pep), the power contained in a wave of amplitude equal to the pilot carrier and transmitted sideband power. Where no pilot carrier is transmitted, the power is specified as peak sideband power (psp).

8.3.4 Frequency modulation (FM)

Both frequency and phase modulation (both may be referred to as angle modulation) effectively vary the frequency of the carrier rather than its amplitude. Frequency modulation varies the carrier frequency directly but its amplitude remains constant regardless of the modulating voltage. Angle modulation is employed at VHF and above for both communications and broadcasting services.

When frequency modulated, a carrier frequency either increases or decreases when the modulation voltage is positive and varies in the opposite sense when the modulating voltage is negative (*Figure 8.6*). The amount of modulation, i.e. the 'deviation' of the carrier from its nominal frequency, is proportional to the amplitude, and not the frequency, of the modulating voltage. The modulation index, M, is defined as the deviation divided by the modulating frequency:

$$M = \frac{f_d}{f_m} \text{ or } \frac{\omega_d}{\omega_m}$$

where

f_d = deviation in hertz

f_m = modulating frequency in hertz

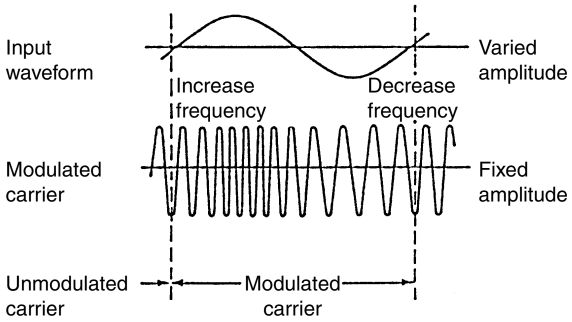

Figure 8.6 Frequency modulation

$\omega_d = 2\pi f_d$ = deviation in radians
$\omega_m = 2\pi f_m$ = modulating frequency in radians

The 'deviation ratio', D, is given by:

$$D = \frac{f_{d(\max)}}{f_{m(\max)}}$$

The peak deviation is the maximum amount of modulation occurring (equivalent to 100% amplitude modulation). With FM, this is limited only by the need to conserve spectrum; there is no technical limit where distortion occurs, as with 100% AM. The maximum permitted deviation for a service is determined by regulation.

The bandwidth of a frequency modulated signal is made up of the carrier and a series of sidebands, sometimes referred to as sidecurrents, spaced apart from each other at the modulating frequency. The number of sidebands is proportional to the modulation index and their amplitudes decrease with spacing from the carrier. It is generally considered satisfactory to transmit those sidebands $M + 1$ in number, with amplitudes greater than 10% of that of the carrier for that modulation index (*Figure 8.7*). The sidebands occur on both sides of the carrier and:

$$\text{1st order sidebands} = f_c \pm f_m$$

$$\text{2nd order sidebands} = f_c \pm 2f_m, \text{ etc.}$$

When the modulation index approaches 6 a good approximation of the bandwidth (δf) required for an FM transmission is $2(f_d + f_m)$ Hz. For example, speech 300–3000 Hz, max deviation 15 kHz (relevant to a VHF, 50 kHz channel spacing system):

$$\delta f = 2(f_d + f_m) = 2(15 + 3)\,\text{kHz} = 36\,\text{kHz}$$

$$M = 15/3 = 5$$

The bandwidth is also given by $\delta f = f_m \times$ highest needed sideband × 2. From Bessel functions (*Figure 8.7*), a modulation index of 5 requires the 6th order sideband to be transmitted ($M + 1$). Therefore:

$$\delta f = f_m \times 6 \times 2 = 3 \times 6 \times 2 = 36\,\text{kHz}$$

For specific values of M the carrier of an FM wave disappears. The successive disappearances and the modulation index are given in *Table 8.2*.

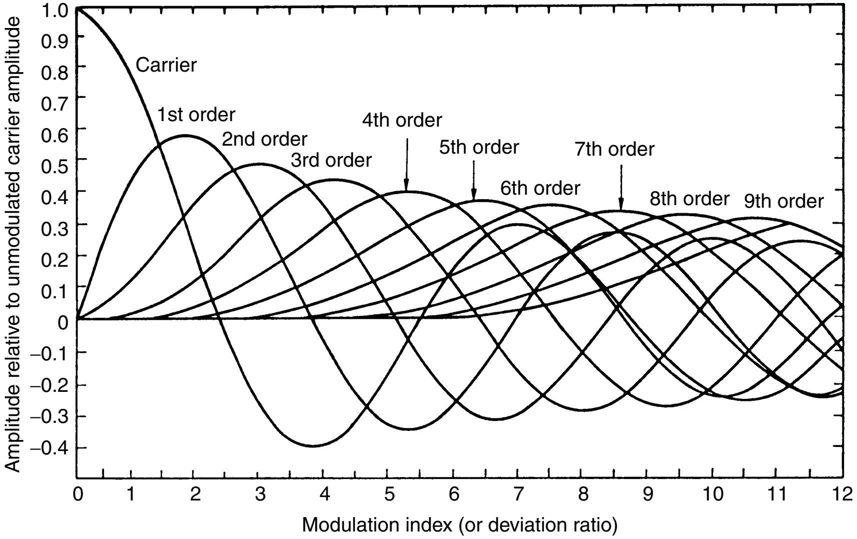

Figure 8.7 Amplitudes of components of FM wave with variation of modulation index

Table 8.2 Successive disappearances and modulation index

Order of disappearance	*Modulation index M*
1	2.40
2	5.52
3	8.65
4	11.79
5	14.93
6	18.07
$n(n > 6)$	$18.07 + \pi(n - 6)$

Systems where the modulation index exceeds $\pi/2$ are considered to be wide band FM (WBFM), those with a modulation index lower than $\pi/2$, narrow band FM (NBFM). The bandwidth of an NBFM signal is $2f_{\mathrm{m(max)}}$.

Pressure on spectrum necessitates narrower channel spacings for communications and, at VHF, 12.5 kHz is normal with ±2.5 kHz as the maximum permitted deviation. Lower standards of performance with a restricted modulation index and a highest modulating frequency of 3 kHz have had to be accepted.

For transmitters used on 12.5 kHz channel spaced systems the highest modulating frequency is, in practical terms, 2700 Hz because the specification requires the frequency response to fall above 2.55 kHz.

The modulation index on such systems, assuming a highest modulating frequency of 3 kHz, is $2.5/3 = 0.8333$ ($<\pi/2$ and so system is NBFM) and the amplitude of the 2nd order sideband is $<10\%$ of carrier amplitude (*Figure 8.7*), so the bandwidth $= 2f_{\mathrm{m(max)}} = 6\,\mathrm{kHz}$.

8.3.5 Phase modulation

The end result of phase modulation is frequency modulation, but the method of achieving it and the definition of the modulation index is different. Phase modulation is used in VHF and UHF transmitters where the carrier frequency is generated directly by a crystal oscillator. The frequency of a crystal oscillator can be varied by only a few radians but if the oscillator frequency is multiplied to produce the final carrier frequency the phase variation is also multiplied to produce a frequency deviation.

In frequency modulation deviation is proportional to the modulating voltage, but in phase modulation the frequency deviation is proportional to both the modulating voltage and frequency.

The phase modulation, ϕ_d radians, equals the modulation index so, for phase modulation, the modulation index is $\phi_d = \omega_d/\omega_m$.

Phase modulation has, therefore, a frequency response for the deviation that rises at 6 dB per octave of the modulating frequency. The flat frequency response of FM can be produced on a phase modulated transmitter by installing a filter with a response falling at 6 dB per octave in the audio amplifier, and the rising response of PM can be produced in an FM transmitter with a rising response filter.

8.3.6 Pre- and de-emphasis

Phase modulation, and FM modified to give a rising frequency response (pre-emphasis), offer an improved signal-to-noise ratio in the receiver. The +6 dB per octave response produced in the transmitter is restored to a flat response in the receiver by a −6 dB per octave filter in the audio circuitry which reduces both the enhanced levels of the higher speech frequencies and the high frequency noise (de-emphasis).

8.3.7 Merits of amplitude and frequency modulation

The advantages and disadvantages of AM and FM are given in *Table 8.3*.

Table 8.3 Advantages and disadvantages of AM and FM

	Advantages	*Disadvantages*
AM	Simple modulators and de-modulators	Susceptible to man-made noise
	Narrower bandwidth than wide-band FM	Audio strength falls with decreasing RF signal strength
		Inefficient power usage
		Limited dynamic range
		Transmitter output power not easily adjusted
FM	Less susceptible to noise	Wider bandwidth
	Constant audio level to almost the end of radio range	
	Capture effect in receiver	Capture effect may be undesirable, e.g. aviation communications
	More power-efficient	
	Transmitter output power easily adjustable	

8.3.8 Stereo FM radio

An important advantage of FM over AM radio broadcasting is the availability of stereo sound. This requires the transmission and reception of the left and right audio signals, to produce a stereo image. However, the introduction of stereo had to be achieved while still catering for large numbers of listeners who wanted to continue receiving in mono. The introduction of stereo FM radio had to go unnoticed by anyone using old mono radios. Without this condition, a simpler method could have been chosen.

The standard FM stereo system used by broadcasters around the world is illustrated in the block diagram of *Figure 8.8*. It uses frequency division multiplexing (FDM) to combine the two signals of the left and right channels. The signals are filtered to limit the bandwidth to 15 kHz. The left (L) and right (R) signals are then added to produce a sum signal and subtracted one from the other to produce a difference signal.

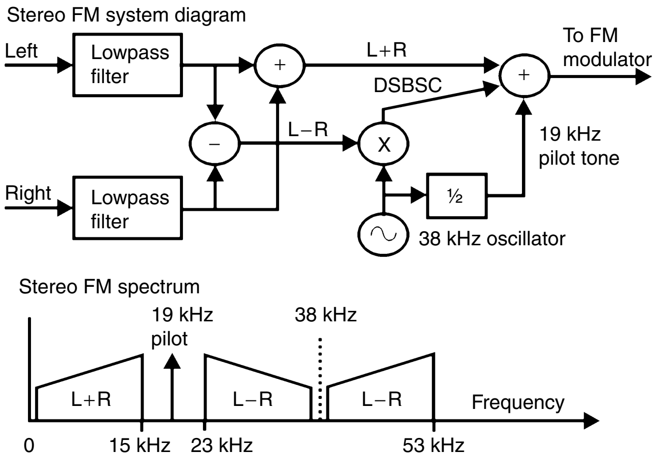
Stereo FM system diagram
Left
Lowpass filter
+
L+R
To FM modulator
DSBSC
+
−
L−R
X
19 kHz pilot tone
Right
Lowpass filter
½
38 kHz oscillator
Stereo FM spectrum
19 kHz pilot
38 kHz
L+R
L−R
L−R
Frequency
0
15 kHz
23 kHz
53 kHz

Figure 8.8 FM stereo transmitter

The sum signal provides a monophonic signal, which provides a baseband signal for the frequency modulator. This was the technique used in mono FM and thus was the obvious choice for stereo FM, to allow backward compatibility. An mono FM radio can receive this signal and recover the combined L and R channels, thereby satisfying the requirement for providing unchanged service to mono radios.

The difference signal is used to amplitude modulate a 38 kHz sinewave. By utilizing a balanced mixer, double sideband suppressed carrier (DSBSC) is generated. However, the modulation method must take into account the ease of demodulation. In particular, demodulating a DSBSC signal can be difficult. Both frequency and phase of the carrier are needed to perform faithful demodulation.

In the stereo system the DSBSC demodulation problem is dealt with by including a 19 kHz pilot tone in the broadcast. This tone is generated by a divide-by-two frequency converter circuit, which takes the 38 kHz carrier and produces the 19 kHz pilot tone. The 19 kHz pilot tone falls midway in the spectral region between the mono sum signal (up to 15 kHz) and below the DSBSC difference signal information. The DSBSC signal extends from 23 kHz to 53 kHz, since the input modulating signals are band limited to 15 kHz. The DSBSC output is added to the baseband (L and R sum) signal and the 19 kHz pilot tone before being sent to the FM modulator.

A mono FM receiver ignores the stereo information by using a filter after its FM demodulator to block everything above 15 kHz. It passes the combined L and R channel signal, which is monophonic.

A stereo receiver has an additional circuit after the FM demodulator to detect and demodulate the DSBSC signal. The stereo receiver detects a 19 kHz pilot tone and uses this to generate a 38 kHz signal. This is then used to demodulate the DSBSC signal that carries the L and R channel difference information. The stereo receiver then has both the sum and difference signals, which is all that is needed to recreate the separate left and right signals. Separation is achieved by adding and subtracting sum and difference signals.

The noise power spectral density of a demodulated FM signal tends to increase with the square of the modulation frequency. This is why pre-emphasis is used to boost the high frequency baseband signals for maintaining the signal-to-noise ratio of the transmitted signal. However, this means that there will be more noise in the 23 kHz to 53 kHz band used for the difference signal than for the 0–15 kHz band used for the sum signal. Consequently a significantly higher input signal level is required to receive a stereo transmission compared with a mono signal for the same output signal-to-noise ratio. Thus stereo reception requires far higher radio signal levels than for mono reception and is more susceptible to interference from other radio sources.

8.4 Digital modulation

8.4.1 Data processing

Filtering

A data pulse with a sharp rise- and fall-time produces harmonic frequencies and requires a wide bandwidth if it is to maintain its shape during transmission. Consequently, to transmit the data over a limited bandwidth the pulses must be shaped to reduce the harmonic content as much as possible without impairing the intelligibility of the signal. This is accomplished by the use of low-pass (Gaussian) filters of which the result is a string of smoother pulses, often referred to as 'tamed' (*Figure 8.9(b)*). Tamed FM permits high data rates within a limited channel bandwidth whilst maintaining acceptable adjacent channel interference levels.

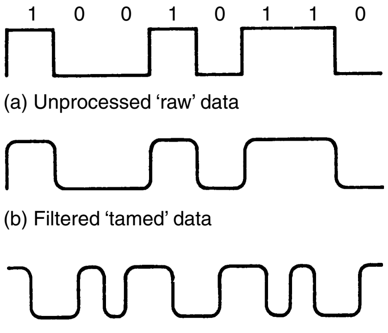

(a) Unprocessed 'raw' data

(b) Filtered 'tamed' data

(c) Manchester encoded, 'tamed' data

Figure 8.9 Data processing

Manchester encoding

A serious problem with the transmission of binary data is that unless the clocks in the transmitter and receiver are synchronous the digits become confused, particularly where a continuous string of 1s or 0s occurs. Manchester encoding makes 1s change state from 1 to 0, and 0s from 0 to 1 during each digit period (*Figure 8.9(c)*), facilitating the synchronization of the clocks and rendering the digits more easily recognizable.

Gray coding

For binary data the number of signalling levels is not restricted to two (mark and space). Multi-level signalling has the advantage that one signalling element carries the information for more than one information bit, thus reducing the bandwidth requirement.

If the number of levels is increased to m, where $m = 2^n$ (i.e. $n = \log_2 m$), the m-level data symbol is represented by n binary digits of 0 or 1 (see quaternary phase shift keying, Section 8.4.5). For example, in a quaternary data signal, $m = 4 = 2^2$ giving the binary sequences 00, 01, 10, 11 for each signalling level. This process is Gray coding where there is only one digit difference for each transition between adjacent levels. The number of levels is not restricted to four and 16 level quadrature amplitude modulation (QAM), where $n = 4$, is an efficient system.

In a Gray coded signal where each element contains n bits ($n = \log_2 m$) and the signalling element rate is B bauds (elements per second), the transmission rate is $B \log_2 m$ bits per second.

8.4.2 On/off and amplitude shift keying

On/off keying

The earliest modulation method. A continuous radio frequency wave (CW) is interrupted in a recognizable pattern (Morse code). To provide audibility the carrier is heterodyned with a beat frequency oscillator (BFO) in the receiver. The use of a modulated continuous wave (MCW) eliminates the need for a BFO but the bandwidth of the signal is increased. The problem with on/off keying is the lack of a reference level. If the signal strength temporarily falls below the sensitivity threshold of the receiver it appears to the operator as a series of spaces.

Binary amplitude shift keying (ASK or BASK)

This shifts the level of an audio frequency subcarrier which then modulates a radio frequency carrier.

Because the level of a subcarrier is changed, AM sidebands are produced. Also, because the keyed waveform is non-sinusoidal harmonics occur. The occupied sub-carrier bandwidth for ASK is:

$$\text{Bandwidth} = 2B$$

where B = bit repetition rate (bits/second).

When the RF carrier is modulated its bandwidth is $2(f_c + B)$ where f_c = subcarrier frequency.

8.4.3 Frequency shift keying (FSK)

Although used for conveying digital information, frequency shift keying in reality employs frequency modulation. In its original form, developed for HF transmission, FSK changes the carrier frequency to indicate a 1 or a 0 but retains the nominal carrier frequency as a reference and to represent a mark. A downwards shift of carrier frequency by 170 Hz represents a space in the HF radio system.

Table 8.4 Standard ITU-T frequencies for various bit rates

Bit rate (bps)	*Frequencies*		*Subcarrier frequency*
	0	*1*	
600	1700	1300	1500
1200	2100	1300	1700
up to 300	1180	980	1080
	1850	1650	1750

Modern FSK uses two different modulation frequencies to represent 1s and 0s. If intersymbol interference (ISI) is to be avoided the separation of the tones must be more than half the bit rate, and a factor of 0.7 is often used. Standard ITU-T frequencies for various bit rates are given in *Table 8.4*.

The base bandwidth requirement is:

$$\text{Bandwidth} = f_2 - f_1 + 2B$$

The bandwidth of a modulated carrier is:

$$\text{RF bandwidth} = 2(f_2 + B)(\text{narrow band FM})$$

Minimum shift keying (MSK) is a form of FSK where the frequency deviation is equal to half the bit rate.

Gaussian minimum phase shift keying (GMSK)

Similar to MSK, the Gaussian filters improve the adjacent channel performance against a small cost (approximately 1%) in ISI while achieving high data transmission rates.

8.4.4 Fast frequency shift keying (FFSK)

Fast frequency shift keying may either amplitude or frequency modulate the carrier. In binary FFSK, the data is changed in a modem to tones of 1800 Hz to represent binary 0 and 1200 Hz to represent binary 1. During transmission a binary 1 consists of 1 cycle of 1200 Hz, f_1, and a 0, $1\frac{1}{2}$ cycles of 1800 Hz, f_2, i.e. a bit rate of 1200 bps. For acceptable intersymbol interference the distance between the tones cannot be less than half the bit rate and the 600 Hz separation in FFSK represents the fastest signalling speed – hence the description – and minimum bandwidth. For this reason it is sometimes called minimum frequency shift keying (MFSK). The base bandwidth is the same as for FSK, but the RF bandwidth depends upon the system deviation. For example, on a 12.5 kHz channel spaced system carrying FFSK and complying with Radiocommunications Agency Code of Practice MPT 1317, deviation = 60% of system deviation = 1.5 kHz:

$$f_1 = 1200\,\text{Hz}$$

$$f_2 = 1800\,\text{Hz}$$

$$B = 1200\,\text{bps}$$

$f_{\text{max}} = f_2 + B = 3\,\text{kHz}$ and the mod. index $m = \text{max. dev.}/f_{\text{max}} < \pi/2$ so the system is NBFM and the bandwidth $= 2(f_2 + B) = 6\,\text{kHz}$.

Where more data states than binary are to be transmitted, multistate FSK (M-ary FSK) is also possible where M may be up to 32 states.

Both FFSK and M-ary FSK are well suited to radio transmission as the change of state occurs while the signals are passing through zero, avoiding sudden phase changes (*Figure 8.10(a)*).

The minimum distance between the tones used in FFSK of 0.5 times the bit rate is not ideal for immunity to intersymbol interference (ISI).

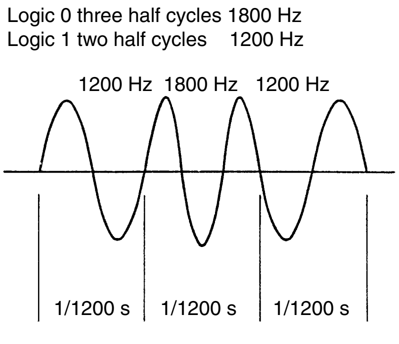
Logic 0 three half cycles 1800 Hz
Logic 1 two half cycles 1200 Hz
1200 Hz
1800 Hz
1200 Hz
1/1200 s
1/1200 s
1/1200 s

(a) FFSK modulation

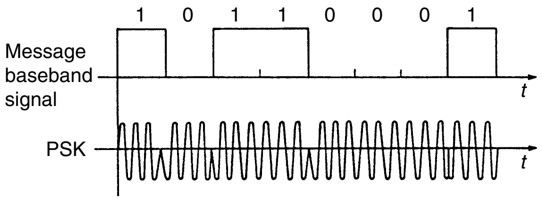
1 0 1 1 0 0 0 1
Message baseband signal
t
PSK
t

(b) PSK modulation

Figure 8.10 FFSK and PSK

A separation of 0.7 times the bit rate would improve the ISI but increase the bandwidth.

8.4.5 *Phase shift keying*

There are several variants of phase shift keying. Binary phase shift keying (BPSK or PSK) changes the phase of the carrier by 180° at the zero crossing point (*Figure 8.10(b)*). No carrier frequency is present with PSK as half the time the carrier is multiplied by $+1$ and the other half by -1 and cancels out, but the reference phase of the carrier must be re-inserted at the receiver. The bandwidths occupied are the same as for ASK, i.e.:

$$\text{Baseband} = 2B$$

$$\text{RF bandwidth (AM or NBFM)} = 2(f_c + B)$$

Differential phase shift keying (DPSK) advances the phase 90° or 270° at each change of logic state (*Figure 8.11*). Changing phase only at a

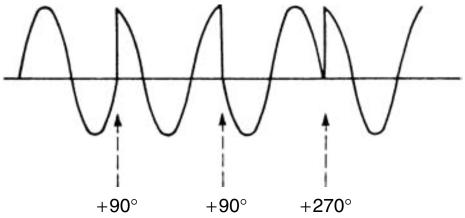

Data	Phase change
0	+90°
1	+270°

(a) DPSK modulation

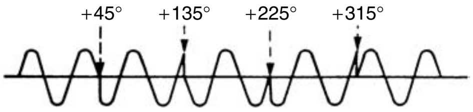

Dibit	Phase change
00	+45°
01	+135°
11	+225°
10	+315°

(b) QPSK modulation

Figure 8.11 DPSK and QPSK

change of logic state saves bandwidth which, for DPSK, is equal to the bit rate.

An important advantage of both FFSK and PSK over FSK is that because the moment of change is predefined it is possible to recover data more accurately. However, the transition between signalling states is not smooth requiring large and rapid phase shifts. Multilevel systems with less phase shift between elements are preferable.

Quaternary, or quadrature, phase shift keying (QPSK) is a four-level Gray-coded signalling method with 90° phase shift between adjacent signalling elements (*Figure 8.11(b)*). If the signal is considered as a vector the points at $\pi/4$ (45°), $3\pi/4$ (135°), $5\pi/4$ (225°) and $7\pi/4$ (315°) represent the transition points between states and the binary data.

8.5 Spread spectrum transmission

The spread spectrum technique spreads the carrier containing the information over a very wide bandwidth, typically 1.25 MHz, using pseudo-noise generation techniques as described in Chapter 12. The transmitter uses what is in effect a digital key to spread the bandwidth and the receiver is equipped with an identical key for despreading. A number of users with different keys can occupy the same band at the same time. The system operates well in poor signal-to-noise or high interference environments.

A continuous wave (CW) transmission concentrates all the radiated energy on a single frequency (*Figure 8.12*). Amplitude modulation and narrow band FM widen the radiated bandwidth, reducing the energy at the carrier frequency and per kHz. Wide band FM carries the process a stage further until with spread spectrum the band width is increased to the extent that the signal almost disappears into the noise floor.

Spreading of the bandwidth is achieved by multiplying the digitally modulated signal by a spreading code at a much higher bit rate (100–1000 times the signal bit rate). This is done by combining the signal with the output of a random code generator running at 2 or 3 orders of magnitude faster than the binary signal rate. *Figure 8.13* is a block diagram of a spread spectrum system. A clock running at the spreading rate R_c is used to drive both the spreading generator and, after frequency division, the data encoder. The carrier is first BPSK modulated by the encoded data and then in a balanced modulator (the spreading correlator) by the high rate code from the spreading generator. The resultant transmitted bits are referred to as chips to distinguish them from data bits. In the receiver, the clock pulses R_c are recovered and used to drive both the despreading generator and decoder.

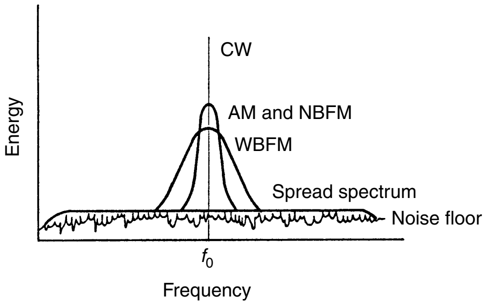

Figure 8.12 Comparative energy dispersal

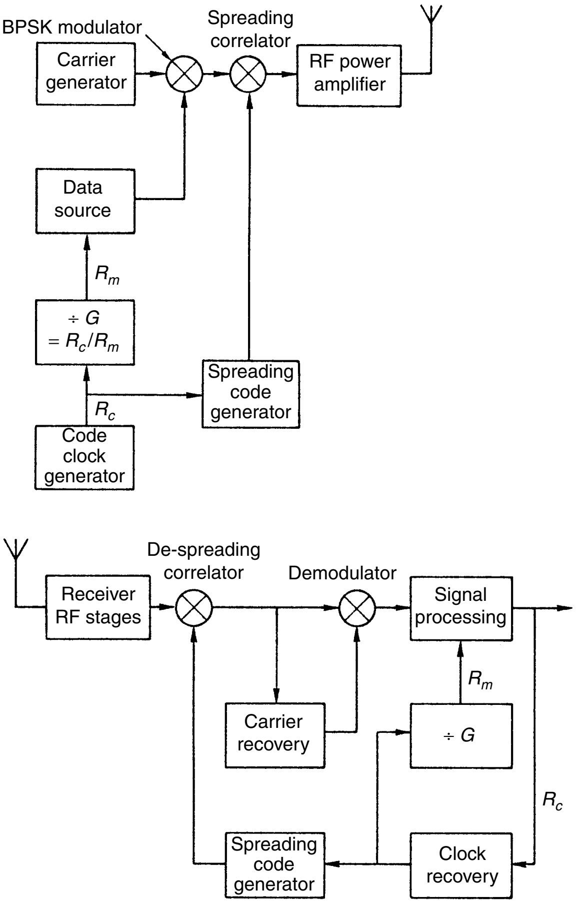
BPSK modulator
Spreading correlator
Carrier generator
RF power amplifier
Data source
R_m
÷ G = R_c/R_m
R_c
Spreading code generator
Code clock generator
De-spreading correlator
Demodulator
Receiver RF stages
Signal processing
Carrier recovery
R_m
÷ G
R_c
Spreading code generator
Clock recovery

Figure 8.13 CDMA (spread spectrum) system

Many users can be accommodated by allocating each a unique spreading code. It is common to use a pseudo-noise (PN) generator to multiply the bit rate and then to modulate the carrier with either FSK or PSK.

Although spread spectrum is a digital system, in quality of signal there are similarities with analogue:

Analogue	*Digital*
Signal/noise ratio	Energy per bit, E_b/N_o
Intelligibility, signal/ noise + distortion	Bit error rate

References

Belcher, R. *et al.* (1989). *Newnes Mobile Radio Servicing Handbook*. Butterworth-Heinemann, Oxford.

Kennedy, G. (1977). *Electronic Communications Systems*, McGraw-Hill Kogashuka, Tokyo.

Winder, S.W. (2001). *Newnes Telecommunications Pocket Book*. Butterworth-Heinemann, Oxford.

9 Frequency planning

9.1 International and regional planning

The International Telecommunications Union (ITU) administers the planning and regulation of the radio frequency spectrum on a worldwide basis through the World Administrative Radio Conferences.

For planning purposes the world is divided into three regions as shown in *Figure 9.1*. The boundaries are formed by geographical features suited to the purpose such as seas, high mountain ranges or uninhabited remote areas.

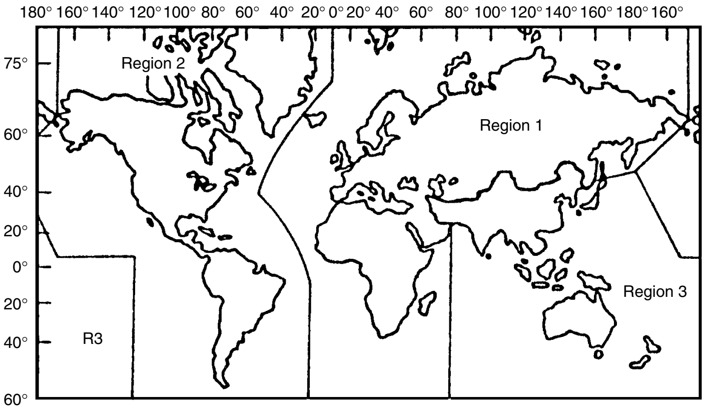

Figure 9.1 ITU defined regions. For purposes of international allocations of frequencies the world has been divided into three regions

The World Administrative Radio Conferences decide the use of blocks of the spectrum, e.g. sound broadcasting, television, marine communications, and the countries permitted to use the blocks for those allotted purposes.

9.2 National planning

Having been allocated blocks of frequencies for a particular type of use, the administration of each country determines the method of allocating frequency bands from each block to the user categories within their country. Every country has a radio regulatory department within its administration. In the UK this is the Radiocommunications Agency of the Department of Trade and Industry.

9.2.1 *The role of the Radiocommunications Agency*

The aims of the Radiocommunications Agency in the UK, and, in general, the radio regulatory bodies of other countries are:

1. To ensure that the radio frequency spectrum is used in ways which maximize its contribution to national social and economic welfare, having regard to safety of life factors.
2. To make the maximum amount of spectrum available for commercial use.
3. To provide an expert service to government as a whole in the field of radio regulation.

The first stage of national planning is the assignment of the radio frequencies (channels) within a geographic area. Both the allocation of the blocks to users and the geographic assignment of channels if not wisely carried out can result in spectrum pollution – intermodulation products are one source (see Chapter 19) – and unacceptable interference between services and users. The incorrect allocation of blocks may not only cause interference in the home country but, even at VHF and UHF, between adjacent countries. Incorrect assignment of channels causes a more local problem but, depending on the frequencies involved, the undesirable effects can spread over a wide area.

The second stage of national planning is the assignment of discrete channels for use on multi-user sites where the selection of incompatible frequencies causes interference, receiver de-sensitization and, possibly, blocking, and intermodulation products.

Not all channels are allocated directly by the regulatory body. Blocks of channels, usually comprised of two frequencies, may be issued to responsible user organizations: the Home Office for the police and fire services is an example. These organizations then become responsible for the frequency planning and allocation within their user group.

Additional to the allocation of frequencies, the Radiocommunications Agency through the licensing procedure regulates the use of base station sites, the maximum transmitter power, and antenna directivity for each service. It also prepares and publishes technical specifications with which all equipment must comply. A list of the current specifications, of which single copies can be obtained, is available from

The Information and Library Service, Radiocommunications Agency, Wyndham House, 189 Marsh Wall, London E14 9SX.

The Radio Investigation Service (RIS) is the branch of the Agency which, in addition to investigating interference, inspects all radio installations prior to commissioning and, if in order, issues an inspection certificate. No station is permitted to operate without a certificate and may not be modified subsequent to the issue of the certificate.

Current policy throughout Europe is leading towards the de-regulation of radio communications while safeguarding the protection from interference. Allocation of frequencies by pricing is also under consideration on the basis that a scarce resource, the spectrum, will be allocated to the users having the greatest need.

9.3 Designations of radio emissions

Radio emissions should be expressed in a three-symbol code form, which defines the exact nature of carrier, signal and transmitted information. The first symbol defines the carrier, the second symbol defines the signal, and the third symbol defines the information.

First symbol

A	Double-sideband amplitude-modulated
B	Independent sideband amplitude-modulated
C	Vestigial sideband amplitude-modulated
D	Amplitude- and angle-modulated simultaneously, or in a predefined sequence
F	Frequency modulated
G	Phase modulated
H	Single-sideband, full carrier
J	Single-sideband, suppressed carrier
K	Amplitude-modulated pulse sequence
L	Width-modulated pulse sequence
M	Position phase modulated pulse sequence
N	Unmodulated carrier
P	Unmodulated pulse sequence
Q	Pulse sequence in which carrier is angle-modulated during the pulse period
R	Single-sideband, reduced or variable level carrier
V	Pulse sequence with a combination of carrier modulations, or produced by other means
W	Carrier is modulated by two or more of angle, amplitude, and pulse modes, simultaneously or in a defined sequence
X	Other cases

Second symbol

0	No modulating signal
1	Digital signal without modulating sub-carrier
2	Digital signal with modulating sub-carrier
3	Analogue signal
7	Two or more channels with digital signals
8	Two or more channels with analogue signals
9	Composite system with one or more channels of digital signals and one or more channels of analogue signals
X	Other cases

Third symbol

A	Aural telegraph
B	Automatic telegraph
C	Facsimile
D	Data
E	Telephony (and sound broadcasting)
F	Television
N	No information transmitted
W	Combination of any of the above
X	Other cases

9.4 Bandwidth and frequency designations

A four symbol code should be used to express bandwidth and frequency to three significant figures. A letter to denote the unit of frequency is placed in the position of the decimal point, where the letters and bandwidths are:

Letter	Bandwidth
H	Below 1000 Hz
K	Between 1 and 999 kHz
M	Between 1 and 999 MHz
G	Between 1 and 999 GHz

So, a frequency of 120 Hz is 120H, while a frequency of 12 Hz is 12H0 etc.

9.5 General frequency allocations

VLF, LF, MF (frequency in kHz)

10.0	140.5	Fixed; maritime; navigation
140.5	283.5	Broadcast

255.0	526.5	Radio navigation; fixed
526.0	1606.5	Broadcast
1606.5	1800.0	Maritime and land mobile; fixed
1810.0	1850.0	Amateur (shared in UK)
1850	2000	Amateur
1850	2045	Fixed; mobile
2045	2173.5	Maritime mobile; fixed
2160	2170	Radiolocation
2173.5	2190.5	Mobile
2190.5	2194	Maritime
2194	2625	Fixed; mobile
2300	2498	Broadcast
2625	2650	Maritime mobile
2650	2850	Fixed; mobile
2850	3155	Aero mobile

HF (frequency in kHz)

3155	3400	Fixed; mobile
3200	3400	Broadcast
3400	3500	Aero mobile
3500	3800	Amateur, fixed; mobile
3800	4000	Amateur (region 2 only)
3800	3900	Fixed; mobile
3800	3950	Aero mobile
3950	4000	Fixed; broadcast
4000	4063	Fixed; maritime mobile
4063	4438	Maritime mobile
4438	4650	Fixed; mobile
4650	4750	Aero mobile
4750	5060	Fixed; mobile; broadcast
5060	5480	Fixed; mobile
5450	5730	Aero mobile
5730	5950	Fixed; mobile
5950	6200	Broadcast
6200	6525	Maritime mobile
6525	6765	Aero mobile
6765	7000	Fixed; mobile
7000	7100	Amateur
7100	7300	Amateur (region 2 only)
7100	7300	Broadcast (regions 1 and 3)
7300	8195	Fixed
8100	8815	Maritime mobile
8815	9040	Aero mobile

9040	9500	Fixed
9500	10 000	Broadcast
10 000	10 100	Aero mobile
10 100	11 175	Fixed
10 100	10 150	Amateur
11 175	11 400	Aero mobile
11 400	11 650	Fixed
11 650	12 050	Broadcast
12 050	12 230	Fixed
12 230	13 200	Maritime mobile
13 200	13 360	Aero mobile
13 360	13 600	Fixed
13 600	13 800	Broadcast
13 800	14 000	Fixed
14 000	14 350	Amateur
14 350	15 000	Fixed
15 000	15 100	Aero mobile
15 100	15 600	Broadcast
15 600	16 360	Fixed
16 360	17 410	Maritime mobile
17 410	17 550	Fixed
17 550	17 900	Broadcast
17 900	18 030	Aero mobile
18 030	18 068	Fixed
18 068	18 168	Amateur
18 168	18 780	Fixed
18 780	18 900	Maritime mobile
18 900	19 680	Fixed
19 680	19 800	Maritime mobile
19 800	21 000	Fixed
21 000	21 450	Amateur
21 450	21 850	Broadcast
21 850	21 870	Fixed
21 870	22 000	Aero mobile
22 000	22 855	Maritime mobile
22 855	23 200	Fixed; mobile
23 200	23 350	Aero mobile
23 350	24 890	Fixed; mobile
24 890	24 990	Amateur
25 010	25 070	Fixed; mobile
25 070	25 210	Maritime mobile
21 210	25 550	Fixed; mobile

25 550	25 670	Radio astronomy
25 670	26 100	Broadcast
26 100	26 175	Maritime mobile
26 175	28 000	Fixed; mobile
28 000	29 700	Amateur
29 700	30 000	Fixed; mobile

VHF, UHF (frequencies in MHz)

30.0	50.0	Fixed; mobile
47.0	68.0	Broadcast (TV)
50.0	52.0	Amateur (UK)
50.0	54.0	Amateur (regions 2 and 3)
68.0	74.8	Fixed; mobile
70.0	70.5	Amateur (UK)
74.8	75.2	Aero navigation
75.2	87.5	Fixed; mobile
87.5	108	Broadcast (FM)
108	118	Aero navigation
118	137	Aero mobile
137	138	Spacecraft; satellites
138	144	Aero mobile; space research
144	146	Amateur
146	148	Amateur (regions 2 and 3 only)
146	174	Fixed; mobile
156	174	Maritime mobile
174	230	Broadcast (TV)
220	225	Amateur (USA)
230	328.6	Fixed; mobile
328.6	335.4	Aero navigation
335.4	400	Fixed; mobile
400	410	Space research; meteorology
410	430	Fixed; mobile
430	440	Amateur; radiolocation
440	470	Fixed; mobile
470	855	Broadcast (TV)
855	1300	Fixed; mobile
902	928	Amateur (USA)
934	935	Citizens band (UK)
1240	1325	Amateur
1300	1350	Aero navigation
1350	1400	Fixed; mobile
1400	1429	Space (uplink); fixed
1429	1525	Fixed; mobile

1525	1600	Space (downlink)
1600	1670	Space (uplink)
1670	1710	Space (downlink)
1710	2290	Fixed; mobile
2290	2300	Space (downlink); fixed
2300	2450	Amateur; fixed
2310	2450	Amateur (UK)
2300	2500	Fixed; mobile
2500	2700	Fixed; space (downlink)
2700	3300	Radar
3300	3400	Radiolocation; amateur
3400	3600	Fixed; space (uplink)
3600	4200	Fixed; space (downlink)
4200	4400	Aero navigation
4400	4500	Fixed; mobile
4500	4800	Fixed; space (downlink)
4800	5000	Fixed; mobile
5000	5850	Radio navigation; radar
5650	5850	Amateur
5850	7250	Fixed; space (uplink)
7250	7900	Fixed; space (downlink)
7900	8500	Fixed; mobile; space
8500	10 500	Radar; navigation
10 000	10 500	Amateur
10 700	12 700	Space (downlink); fixed
12 700	15 400	Space (uplink); fixed
17 700	20 000	Space (up/down); fixed
24 000	24 250	Amateur

9.6 Classes of radio stations

AL	Aeronautical radionavigation land station
AM	Aeronautical radionavigation mobile station
AT	Amateur station
AX	Aeronautical fixed station
BC	Broadcasting station, sound
BT	Broadcasting station, television
CA	Cargo ship
CO	Station open to official correspondence exclusively
CP	Station open to public correspondence
CR	Station open to limited public correspondence
CV	Station open exclusively to correspondence of a private agency

DR	Directive antenna provided with a reflector
EA	Space station in the amateur-satellite service
EB	Space station in the broadcasting-satellite service (sound broadcasting)
EC	Space station in the fixed-satellite service
ED	Space telecommand space station
EE	Space station in the standard frequency-satellite service
EF	Space station in the radiodetermination-satellite service
EG	Space station in the maritime mobile-satellite service
EH	Space research space station
EJ	Space station in the aeronautical mobile-satellite service
EK	Space tracking space station
EM	Meteorological-satellite space station
EN	Radionavigation-satellite space station
EO	Space station in the aeronautical radionavigational-satellite service
EQ	Space station in the maritime radionavigation-satellite service
ER	Space telemetering space station
ES	Station in the intersatellite service
EU	Space station in the land mobile-satellite service
EV	Space station in the broadcasting-satellite service (television)
EW	Space station in the earth exploration-satellite service
EX	Experimental station
EY	Space station in the time signal-satellite service
FA	Aeronautical station
FB	Base station
FC	Coast station
FL	Land station
FP	Port station
FR	Receiving station only, connected with the general network of telecommunication channels
FS	Land station established solely for the safety of life
FX	Fixed station
GS	Station on board a warship or a military or naval aircraft
LR	Radiolocation land station
MA	Aircraft station
ME	Space station
ML	Land mobile station
MO	Mobile station
MR	Radiolocation mobile station

MS	Ship station
ND	Non-directional antenna
NL	Maritime radionavigation land station
OD	Oceanographic data station
OE	Oceanographic data interrogating station
OT	Station open exclusively to operational traffic of the service concerned
PA	Passenger ship
RA	Radio astronomy station
RC	Non-directional radio beacon
RD	Directional radio beacon
RG	Radio direction-finding station
RM	Maritime radionavigation mobile station
RT	Revolving radio beacon
SM	Meteorological aids station
SS	Standard frequency and time signal station
TA	Space operation earth station in the amateur-satellite service
TB	Fixed earth station in the aeronautical mobile-satellite service
TC	Earth station in the fixed-satellite service
TD	Space telecommand earth station
TE	Transmitting earth station
TF	Fixed earth station in the radiodetermination-satellite service
TG	Mobile earth station in the maritime mobile-satellite service
TH	Earth station in the space research service
TI	Earth station in the maritime mobile-satellite service at a specified fixed point
TJ	Mobile earth station in the aeronautical mobile-satellite service
TK	Space tracking earth station
TL	Mobile earth station in the radiodetermination-satellite service
TM	Earth station in the meteorological-satellite service
TN	Earth station in the radionavigation-satellite service
TO	Mobile earth station in the aeronautical radionavigation-satellite service
TP	Receiving earth station
TQ	Mobile earth station in the maritime radionavigation-satellite service
TR	Space telemetering earth station
TS	Television, sound channel
TT	Earth station in the space operation service

TU	Mobile earth station in the land mobile-satellite service
TV	Television, vision channel
TW	Earth station in the earth exploration-satellite service
TX	Fixed earth station in the maritime radionavigation-satellite service
TY	Fixed earth station in the land mobile-satellite service
TZ	Fixed earth station in the aeronautical radionavigation-satellite service

9.7 Radio wavebands

Frequency band	*Frequency*	*Wavelength*	*Waveband definition*
VLF	3 to 30 kHz	1 00 000 to 10 000 m	myriametric
LF	30 to 300 kHz	10 000 to 1000 m	kilometric
MF	300 to 3000 kHz	1000 to 100 m	hectometric
HF	3 to 30 MHz	100 to 10 m	decametric
VHF	30 to 300 MHz	10 to 1 m	metric
UHF	300 to 3000 MHz	1 to 0.1 m	decimetric
SHF	3 to 30 GHz	10 to 1 cm	centimetric
EHF	30 to 300 GHz	1 to 0.1 cm	millimetric
EHF	300 to 3000 GHz	0.1 to 0.01 cm	decimillimetric

Reference

Pannell, W.M. (1979). *Frequency Engineering in Mobile Radio Bands.* Granta Technical Editions, Cambridge.

10 Radio equipment

10.1 Transmitters

10.1.1 Transmitter functions

The functions of all transmitters and the terminology used to describe them, irrespective of the modulation method, are:

1. To generate the radio frequency carrier and amplify it to an appropriate power level; the RF power output.
2. To modulate the carrier with the intelligence to the pre-determined level; the modulation depth for AM, the deviation for FM or PM. The process must introduce the minimum noise and distortion, and prevent the modulation from exceeding the permitted level.
3. Radiate the minimum signals at frequencies outside the permitted bandwidth. Out-of-band or spurious radiation is strictly controlled by the Radiocommunications Agency MPT specifications.

10.1.2 Amplitude-modulated transmitters

Figure 10.1 is a block diagram of an amplitude modulated transmitter; in this case a quartz crystal oscillator generates the carrier frequency, although a frequency synthesizer could equally well be used. The carrier frequency in an AM transmitter is usually generated either at the transmitted frequency or one of its subharmonics; 2nd, 3rd or 6th subharmonic frequencies are typical choices depending on the final frequency.

The output of the oscillator is amplified to the level of the specified power output and if the oscillator runs at a subharmonic a frequency multiplier stage will be included, as in *Figure 10.1*, in the amplifier chain before the final stage, the power amplifier (PA). A tuned filter in the aerial circuit removes from the output unwanted frequencies which might cause interference with other users. A matching circuit correctly matches the impedance of the filter circuit to that of the aerial to ensure maximum power transfer.

In the audio circuits, the speech input from the microphone is processed by controlling the range of frequencies it contains and limiting its amplitude. This eliminates the risk of over-modulation and the production of out-of-band frequencies. Over-modulation produces

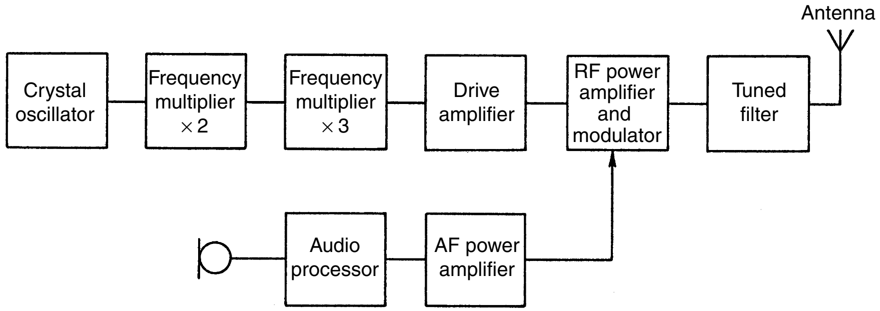
Antenna
Crystal oscillator
Frequency multiplier × 2
Frequency multiplier × 3
Drive amplifier
RF power amplifier and modulator
Tuned filter
Audio processor
AF power amplifier

Figure 10.1 Crystal controlled AM transmitter

out-of-band frequencies in any transmitter, but with AM once 100% modulation is exceeded frequencies are produced across a wide range of the spectrum: a disastrous situation for other users. After processing, the audio is amplified in an AM transmitter to a high level and applied to the RF power amplifier to vary its output in the form illustrated in *Figure 8.2*.

AM transmitters are power inefficient. The RF PA cannot operate in class C. It must be linear so as not to distort the speech, and the audio, when it is the power amplifier which is modulated, must be amplified to a high power. If the modulation is applied at a lower power level (in an early stage of the amplifier chain) all the subsequent amplifiers must operate in a linear, but inefficient, manner.

The power output of an AM transmitter is normally specified in terms of the RMS value of the carrier power but the average and peak powers will depend on the depth of modulation, 100% modulation producing a peak power of twice the carrier peak power (see Chapter 8).

10.1.3 Angle modulated transmitters

Figure 10.2 is a block diagram of a frequency modulated transmitter using a frequency synthesizer for carrier generation. The frequency is generated at the final frequency and RF amplifiers raise the power level to that specified for the transmitter output. When a crystal oscillator is used for carrier generation, it must operate at a very low frequency because a quartz crystal oscillator can be frequency modulated by a few radians only. Several stages of the RF amplifier chain then operate as frequency multipliers. Similar filter and aerial matching circuits to those of an AM transmitter are necessary in the output arrangements of the transmitter.

The audio processing circuitry is similar to that for AM transmitters but the modulation (deviation) is applied to a stage operating at a very low RF power level (directly to the VCO in a synthesizer-equipped transmitter and immediately following the crystal oscillator in a direct crystal controlled one); high power audio is not necessary. An additional simple circuit in an FM transmitter may be included to enhance the higher audio frequencies at a rate of 6 dB/octave. This is pre-emphasis and has the merit of improving the level of speech to noise at the receiver (signal-to-noise ratio). The audio level must still be limited because, while the effect of over-deviation is not as disastrous as over-modulation in an AM transmitter, increasing deviation produces a steadily increasing range of frequencies, known as side-currents, outside the permitted bandwidth.

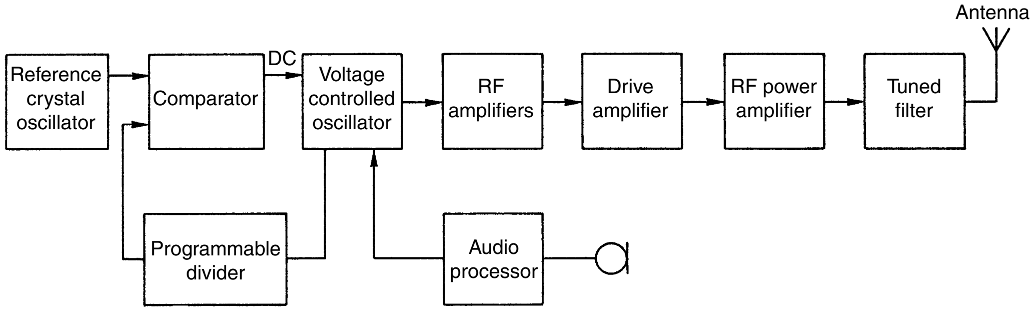
Antenna
Reference crystal oscillator
Comparator
DC
Voltage controlled oscillator
RF amplifiers
Drive amplifier
RF power amplifier
Tuned filter
Programmable divider
Audio processor

Figure 10.2 Synthesizer controlled FM transmitter

Angle modulated transmitters are more power efficient than amplitude modulated because the modulation is applied at a low power level and also, as no audio frequencies are directly present in the RF amplifier and PA stages, these can operate efficiently in class C.

Phase modulated (PM) transmitters, because the phase shift with modulation is very small, generate the carrier at a very low frequency and use direct crystal control. The frequency is then multiplied many times, thirty-two to thirty-six is common, up to the final frequency. After multiplication, the phase modulation which was originally a few radians has effectively become frequency modulation. The phase modulation process produces pre-emphasis inherently. As far as the user is concerned there is no practical difference between phase modulation and frequency modulation with added pre-emphasis.

10.1.4 Transmitter specifications

In the UK, the Radiocommunications Agency issues specifications with which all equipment must comply. Other countries have their own regulatory bodies, such as the FCC in the USA. These specifications are concerned principally with the prevention of interference and obtaining the maximum use of the frequency spectrum. The characteristics defined in the MPT specifications and other features which affect the user, apart from the physical dimensions, are:

- *Supply voltage.*
- *Operational frequency band.*
- *Modulation method.*
- *Channel separation.*
- *RF power output and impedance.* Output ranges from about 0.5 W to 5 W for hand-portables and 5 W to 25 W for mobiles. The maximum power permitted on a system will be specified in the Licence. Output impedance is commonly 50 Ω.
- *Spurious emissions.* The level of these is critical for the prevention of interference with other users on different frequencies. The limit for VHF and UHF is a maximum of 0.25 μW.
- *Residual noise.* Not always quoted by manufacturers, it is the noise level existing on an unmodulated carrier. A typical figure is better than −40 dB referred to full deviation.
- *Audio frequency distortion.* Typically <3% and usually measured with a modulating frequency of 1 kHz at 60% modulation.
- *Audio frequency response.* This is the variation of modulation level over the audio frequency spectrum. Typically within +1 dB to −3 dB over a frequency range of 300 to 3000 Hz (2.55 kHz for 12.5 kHz

channel spacing equipment). It may be quoted with reference to a pre-emphasis curve.

- *Switching bandwidth*. This is the frequency range over which the transmitter will operate without retuning and without degradation of performance. Much equipment is now specified to cover a complete frequency band, e.g. 146–174 MHz, without retuning.

10.2 Receivers

10.2.1 Receiver functions

A receiver's functions are:

- Detect a weak signal; the minimum level, which may be as low as 0.25 microvolts, defines the receiver sensitivity.
- Amplify a received signal and maintain the information contained in a minimum strength signal at a minimum of 12 dB above the electrical noise level (signal-to-noise ratio). If the audio distortion produced in the receiver is also taken into account the above figure becomes the signal-to-noise + distortion (Sinad) ratio. As the signal is increased, the ultimate Sinad should attain 50 to 55 dB.
- Separate the wanted signal from any unwanted ones which may be very close in frequency (the adjacent channel may be 12.5 kHz away at UHF); the selectivity.
- Recover the information from the carrier; demodulation.
- Amplify the audio information to a level suitable for operating a loudspeaker; the audio power output. The audio amplification must introduce the minimum distortion.
- Disenable the audio amplifiers in the absence of signal to cut out the electrical noise. This is done by the mute or squelch circuit.

10.2.2 Types of receiver

It is possible to amplify directly the incoming RF signal to a level suitable for demodulation. This is done in a tuned radio frequency (TRF) receiver, but these are seldom used today because of the problems of obtaining sufficient selectivity and gain at one radio frequency, and the difficulty of retuning a number of RF stages to change frequency. Almost all receivers designed for analogue communications now operate on the superheterodyne principle where the incoming radio frequency is converted to a lower, more manageable, intermediate frequency (IF). The fixed IF means that only the oscillator and, possibly, one RF amplifier stage need retuning for a change of channel.

The lower frequency of the IF facilitates the acquisition of adequate gain with stability and selectivity.

There is little difference in the layout of receivers for AM and FM except that the circuits perform their functions differently. *Figure 10.3* is a block diagram of a typical FM receiver with a crystal controlled local oscillator.

When two frequencies are applied to a non-linear circuit such as the mixer, they combine to produce other frequencies, their sum and difference being the strongest. The superheterodyne mixes a locally generated frequency with the received signal to produce a, usually lower, frequency retaining the modulation of the received signal. Commonly used values for this intermediate frequency (IF) are 465 kHz for MF and HF receivers and 10.7 MHz for VHF and UHF. At these fixed frequencies the necessary high amplification with low noise and stability and the required selectivity are easier to obtain. The local oscillator (injection) frequency = signal frequency ± IF frequency.

Superheterodyne receivers are susceptible to a particular form of interference. Assume a local oscillator frequency of 149.3 MHz is mixed with a wanted signal of 160 MHz to produce the IF frequency of 10.7 MHz. A signal of 138.6 MHz would also combine with the local oscillator to produce 10.7 MHz. The frequency of this spurious response is the image, or second channel frequency. The IF amplifier cannot discriminate against it so some degree of selectivity must also be provided in the RF amplifier and input circuitry. Intermediate frequencies are chosen which are a compromise between ease of obtaining adjacent channel discrimination and image frequency rejection.

The IF amplifier contains a block filter, either crystal or ceramic (see *Figure 7.7*), necessary to discriminate between channels adjacent in frequency. The design of the filter is crucial. It must be wide enough to accommodate the band of frequencies present in the modulation plus an allowance for frequency drift and its response over this band must be uniform with minimal ripple, particularly if data is to be received, yet its response must be of the order of -100 dB at the frequency of the adjacent channel.

Apart from the demodulator the main difference between AM and FM receivers lies in the operation of the IF amplifier. The IF stages in an AM receiver are linear, although their gain is variable. Part of the IF amplifier output is rectified and used to control the gain to provide automatic gain control (AGC). An increase of signal above a predetermined level, with delayed AGC, causes a reduction in IF gain maintaining a sensibly constant audio output level.

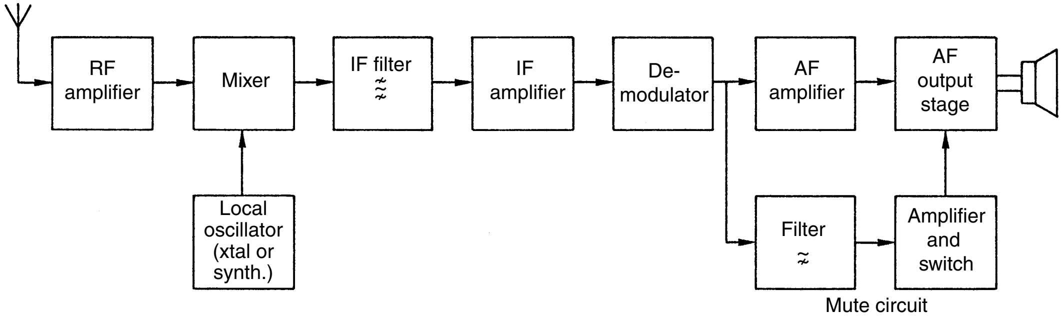
RF amplifier
Mixer
IF filter
IF amplifier
De-modulator
AF amplifier
AF output stage
Local oscillator (xtal or synth.)
Filter
Amplifier and switch
Mute circuit

Figure 10.3 FM single superheterodyne receiver

The IF amplifier in a FM receiver possesses a very high gain, some 100 dB, and is non-linear. On receipt of a signal, or even with only the receiver noise, it runs into limitation, cutting off both positive and negative peaks of the signal or noise. This gives FM its constant level audio output over a wide range of signal levels. It also produces the capture effect where a strong signal, fully limiting, completely removes a weaker signal. A signal difference of some 6 dB is required to provide effective capture.

Most receivers employ only one change of frequency, single superheterodynes, but double superheterodynes are occasionally used at VHF and above. A double superheterodyne changes the frequency twice, perhaps to 10.7 MHz for the first IF and then, using a fixed frequency crystal second local oscillator, to a lower, often 1.2 MHz or thereabouts, second IF. The result is greater gain and selectivity but the incorporation of a second oscillator and mixer increases the number of possible spurious responses.

A variation of a very old receiver circuit, the Autodyne, forgotten in about 1914, is now finding favour in receivers for digitally modulated signals. Under its new names of Homodyne or zero-IF receiver it lends itself to the purpose. In this type of receiver the local oscillator runs at the same frequency as the incoming signal, hence the zero-IF. Frequency or phase shifts of the incoming signal representing the data emerge from the mixer at base band and are applied directly to the processing circuits.

In a communications receiver the noise generated in the aerial and RF stages in the absence of a signal is amplified to what may be, when demodulated, an unacceptable level. To eliminate the annoyance the loudspeaker is switched off during no-signal periods by a squelch or mute circuit. The mute circuit rectifies the high frequency noise at the demodulator and uses it to switch off the audio amplifier. When a signal is received the limiting action of the IF amplifier of an FM receiver depresses the noise in favour of the signal. Some AM mobile receivers use additional FM circuitry to provide improved mute action.

10.2.3 Noise figure

An ideal receiver would generate no noise and the signal-to-noise ratio, in a receiver of given bandwidth, would be determined by the level of the signal at the base of the antenna compared with the noise produced in the antenna. The noise factor of the ideal receiver is the number of times the signal power must exceed the antenna noise power to produce a 1:1 ratio at the receiver. It is given by (see Section 1.5.1):

$$\frac{e^2_{\text{antenna}}(\text{e.m.f.})}{4kTBR}$$

When the receiver input impedance is matched to the antenna impedance, half the power is dissipated in the antenna and the noise factor is 2 (3 dB). In a practical receiver, the noise generated in the RF amplifier is the most significant, and the noise figure is the sum of the RF amplifier noise plus all the preceding losses. Figures between 4 and 6 dB are common and the higher the noise figure, the worse the receiver sensitivity.

Signal-to-noise ratio (SNR, S/N or S_N)

Receivers are evaluated for quality on the basis of *signal-to-noise ratio* (S/N or 'SNR'), sometimes denoted S_N. The goal of the designer is to enhance the SNR as much as possible. Ultimately, the minimum signal level detectable at the output of an amplifier or radio receiver is that level which appears just above the noise floor level. Therefore, the lower the system noise floor, the smaller the *minimum allowable signal*.

Noise factor, noise figure and noise temperature

The noise performance of a receiver or amplifier can be defined in three different, but related, ways: *noise factor* (F_N), *noise figure* (NF) and *equivalent noise temperature* (T_E); these properties are definable as a simple ratio, decibel ratio or Kelvin temperature, respectively.

Noise factor (F_N). For components such as resistors, the noise factor is the ratio of the noise produced by a real resistor to the simple thermal noise of an ideal resistor.

The noise factor of a radio receiver (or any system) is the ratio of output noise power (P_{NO}) to input noise power (P_{NI}):

$$F_N = \left[\frac{P_{NO}}{P_{NI}}\right]_{T=290/\text{K}}$$

In order to make comparisons easier the noise factor is usually measured at the standard temperature (T_o) of 290 K (standardized room temperature); although in some countries 299 K or 300 K are commonly used (the differences are negligible). It is also possible to define noise factor F_N in terms of the output and input signal-to-noise ratios:

$$F_N = \frac{S_{NI}}{S_{NO}}$$

where
S_{NI} is the input signal-to-noise ratio
S_{NO} is the output signal-to-noise ratio

Noise figure (NF). The *noise figure* is the frequency used to measure the receiver's 'goodness', i.e. its departure from 'idealness'. Thus, it is a *figure of merit*. The noise figure is the noise factor converted to decibel notation:

$$\mathrm{NF} = 10 \log(F_N)$$

where
NF is the noise figure in decibels (dB)
F_N is the noise factor
log refers to the system of base-10 logarithms

Noise temperature (T_e). The noise 'temperature' is a means for specifying noise in terms of an equivalent temperature. That is, the noise level that would be produced by a resistor at that temperature (expressed in degrees Kelvin). Evaluating the noise equations shows that the noise power is directly proportional to temperature in degrees Kelvin, and also that noise power collapses to zero at the temperature of Absolute Zero (0 K).

Note that the equivalent noise temperature T_e is *not* the physical temperature of the amplifier, but rather a theoretical construct that is an *equivalent* temperature that produces that amount of noise power in a resistor. The noise temperature is related to the noise factor by:

$$T_e = (F_N - 1)T_o$$

and to noise figure by

$$T_e = KT_o \log^{-1}\left[\frac{\mathrm{NF}}{10}\right] - 1$$

Noise temperature is often specified for receivers and amplifiers in combination with, or in lieu of, the noise figure.

Noise in cascade amplifiers

A noise signal is seen by any amplifier following the noise source as a valid input signal. Each stage in the cascade chain amplifies both signals and noise from previous stages, and also contributes some additional noise of its own. Thus, in a cascade amplifier the final stage sees an input signal that consists of the original signal and noise amplified by each successive stage plus the noise contributed by earlier stages. The overall noise factor for a cascade amplifier can be calculated from *Friis' noise equation*:

$$F_N = F_1 + \frac{F_2 - 1}{G1} + \frac{F_3 - 1}{G1G2} + \cdots + \frac{F_N - 1}{G1G2 \cdots G_{n-1}}$$

where

F_N is the overall noise factor of N stages in cascade
F_1 is the noise factor of stage-1
F_2 is the noise factor of stage-2
F_N is the noise factor of the *n*th stage
$G1$ is the gain of stage-1
$G2$ is the gain of stage-2
G_{n-1} is the gain of stage $(n - 1)$.

As you can see from Friis' equation, the noise factor of the entire cascade chain is dominated by the noise contribution of the first stage or two. High gain, low noise radio astronomy RF amplifiers typically use low noise amplifier (LNA) circuits for the first stage or two in the cascade chain. Thus, you will find an LNA at the feedpoint of a satellite receiver's dish antenna, and possibly another one at the input of the receiver module itself, but other amplifiers in the chain might be more modest (although their noise contribution cannot be ignored at radio astronomy signal levels).

The matter of signal-to-noise ratio (S/N) is sometimes treated in different ways that each attempts to crank some reality into the process. The signal-plus-noise-to-noise ratio (S + N/N) is found quite often. As the ratios get higher, the S/N and S + N/N converge (only about 0.5 dB difference at ratios as little as 10 dB). Still another variant is the SINAD (signal-plus-noise-plus-distortion-to-noise) ratio. The SINAD measurement takes into account most of the factors that can deteriorate reception.

10.2.4 Signal-to-noise ratio and bandwidth

To obtain the maximum signal-to-noise ratio the bandwidth of every circuit must be designed to admit its operational band of frequencies only. The wider its bandwidth the more noise a circuit admits, and the more the bandwidth exceeds that needed, the worse becomes the signal-to-noise ratio. There is a linear ratio between bandwidth and noise power admitted: doubling the bandwidth doubles the noise power.

Improvements in local oscillator crystal frequency stability has resulted in improved signal-to-noise ratios by reducing the necessary width of the IF filter.

Demodulation, the recovery of the audio from the IF bandwidth affects the signal-to-noise ratio. The relationship is complex, but consider two examples. First, for 12.5 kHz channel spacing FM:

> The audio frequency range is 300 Hz to 3000 Hz
> Bandwidth, $b = 2700\,\text{Hz}$
> The modulation index $M = f_{\text{d(max)}}/f_{\text{m(max)}}$
> Signal/noise out $= 3M^2/2b\times$ signal/noise in
> For a 12.5 kHz PMR channel:
> $M = 2500/3000 = 0.83$
> $b = 2.7\,\text{kHz}$
> $3M^2/2b = 0.38$
>
> A 0.38 times reduction in power is -4.8 dB. Demodulation in this case worsens the signal-to-noise ratio by some 5 dB, and if a signal-to-noise ratio of 12 dB is required at the loudspeaker, 17 dB is needed at the input to the demodulator.

When channel separations were 25 kHz, and deviation 5 kHz, the situation was:

> Audio frequency range, 300 Hz to 3000 Hz
> Bandwidth, $b = 2700\,\text{Hz}$
> For a 50 kHz channel:
> $M = 5000/3000 = 1.66$
> $b = 2.7\,\text{kHz}$
> $3M^2/2b = 1.54$
>
> A 1.54 times gain in power is $+1.8$ dB. In this case the demodulation improved the signal-to-noise ratio slightly. Reducing the bandwidth would have the same effect.

For AM, the signal-to-noise ratio is dependent on the modulation depth. The demodulation process reduces the signal-to-noise ratio by

6 dB but the recovered audio is less than the IF bandwidth by a factor of 3:1 which compensates for the demodulation loss. Reducing the modulation depth degrades the signal-to-noise ratio by 6 dB for every halving of the modulation depth.

10.2.5 Receiver specifications

The important features of receiver specifications are:

- *Sensitivity*. The minimum signal to which a receiver will respond. For an AM receiver, the generally accepted standard is the signal (30% modulated with sinusoidal tone, either 400 Hz or 1 kHz) required to provide an audio output of 50 mW. For an FM receiver, the standard is the unmodulated signal required to produce a 20 dB reduction in noise. A typical figure is 0.25 μV (p.d.) for 20 dB quieting. However, sensitivity is often quoted in terms of either the signal-to-noise ratio or Sinad so, in modern parlance, sensitivity and signal-to-noise ratio are sometimes considered to be synonymous.
- *Signal-to-noise ratio (may be quoted as Sinad, signal-to-noise and distortion)*. Typically 0.3 μV (p.d.) for 12 dB Sinad.
- *Spurious response attenuation*. Typically better than 80 dB.
- *Adjacent channel selectivity*. Better than 65 dB at 12.5 kHz channel spacing.
- *Cross modulation*. The modulation, in the receiver, of a wanted signal by a stronger, unwanted signal. It is usually caused by non-linearity in the receiver RF stages.
- *Blocking and de-sensitization*. The reduction in sensitivity of a receiver due to overloading of the RF stages when a strong signal is applied. A blocked receiver may take an appreciable time to recover.
- *Audio frequency response*. Typically within +1 dB to −3 dB of a 6 dB/octave de-emphasis curve from 300 Hz to 3000 Hz (2.55 kHz for 12.5 kHz channel spacing, above which the response falls more rapidly).
- *Audio output*. Typically 3–4 W. For hand-portables, 100–500 mW. (Distortion may also be quoted, typically better than 5%.)
- *Switching bandwidth*. As for transmitter.
- *Duplex separation*. For a receiver/transmitter combination, the minimum separation between the receiving and transmitting frequencies which will permit duplex operation with minimal degradation of the receiver performance.

10.3 Programmable equipment

The current trend is for many of the functions of radio equipment – even down to the control buttons on the front panel – to be software controlled. In one manufacturer's equipment the only screwdriver-adjustable control is that for setting maximum deviation; this is insisted upon by the Radiocommunications Agency for use in the UK.

Software control enables radio sets to be cloned so that once a single piece of equipment is programmed with its frequencies, power output, selective calling, etc., the whole of a fleet of mobiles can be identically programmed in a short time. Some of the functions can be programmed by the user, but not all, and different manufacturers permit different degrees of programming.

The functions are allocated degrees of priority which determine which functions can be changed by each class of person. Again, some of the priorities may be altered by the user but the highest priority functions are installed in a programmable read only memory (PROM) by the manufacturer and can only be altered by replacing the PROM, in most instances a job for the manufacturer.

The essential equipment for major reprogramming is an IBM or equivalent computer and the equipment manufacturer's software which is supplied as a package complete with hardware interface and instructions.

Cloning may be carried out using an inter-connecting cable between the equipment, or restricted reprogramming via a unit supplied by the equipment manufacturer for the purpose.

The functions which may be programmed from a computer are:

- Channel frequencies.
- Transmitter power and deviation.
- Receiver squelch setting (referred to Sinad ratio). Selective calling details including encode or decode only, type of signalling, extension of first tone and call sign.
- Transmission time out timer.
- Channel busy light. Transmitter inhibit on busy channel.
- Alert tones. Tone decoder indicator.
- Low battery indicator. Battery saver.

Restricted reprogramming may include:

- Channel frequencies, including any offsets.
- Lock out of any channels.

- Channel spacing.
- Channel search priority.
- Frequency stability.
- Signalling type.
- Control functions.
- Synthesizer reference crystal (the frequency can be trimmed to compensate for ageing).
- Transmitter power.
- CTCSS (encode and decode).
- Timers.
- Alerts.

Pagers are also programmable. The functions which can be changed include:

- Code number.
- Alert tone/vibrator/repeat.
- Urgency.
- Out of range warning.
- Printer on/off.
- Language of display.

References

Belcher, R. *et al*. (1989). *Newnes Mobile Radio Servicing Handbook*. Butterworth-Heinemann, Oxford.
Langford-Smith, F. (1955). *Radio Designer's Handbook*, Iliffe, London.

11 Microwave communication

11.1 Microwave usage

Microwaves are loosely considered to be those at frequencies between 1 GHz (30 cm) and 100 GHz (0.3 cm). Principal ground communications usage is for point-to-point links carrying information and control signals on systems such as multiplexed communication – including data networks, telemetry and mobile radio. The propagation of the higher frequencies, 30–100 GHz, is being studied for possible future use in micro-cellular radio-telephone systems. These higher frequencies are also used for satellite communications and radar applications. Microwave frequencies up to 3 GHz have now been reserved for mobile use.

11.2 Propagation

The path loss is higher at microwave frequencies than at VHF and UHF: *Figure 1.5* charts the free space loss between isotropic radiators for microwaves at frequencies of 1, 2, 3.5 and 7 GHz over distances of 10 to 1000 km. The free-space loss between isotropic radiators is given by

$$\text{Free space loss, d}B = 32.4 + 20\log_{10} d + 20\log_{10} f$$

where d is in km and f in MHz.

The free-space loss between practical antennas is given by:

$$\text{Loss, dB} = 10\log_{10}\left(\frac{(4\pi d)^2}{\lambda^2}\frac{1}{G_\text{t}G_\text{r}}\right)$$

where
d = path length, metres
λ = wavelength, metres
G_t = power gain of transmitting antenna
G_r = power gain of receiving antenna

The antenna gains are expressed relative to an isotropic radiator (not in dB).

Absorption varies with atmospheric humidity, and as energy is also absorbed by the ground the path height therefore has an effect on the losses. Absorption by rain is a factor at the higher frequencies but, although considered to be insignificant below about 3 GHz, it has an indirect effect. Wet foliage, for example, produces considerable absorption at frequencies as low as 450 MHz.

Waves of millimetric lengths are special cases and narrow bands of very high absorption due to resonance effects exist at 22 and 183 GHz for water vapour, and 60 and 119 GHz for oxygen. Non-resonant attenuation occurs due to scatter from rain, hail and snow. The attenuation increases with frequency as the wavelength approaches the dimensions of a raindrop. Bands of very low absorption, 'atmospheric windows', where the water vapour and oxygen attenuations are very low occur at 37, 97, 137 and 210 GHz.

Objects close to a path may severely affect the received signal and a proposed path must be examined for the likely effects of these at the planning stage. The effects may be due either to diffraction bending the wave away from the line of sight between the antennas, or reflection causing multi-path signals.

The additional losses in a microwave path caused by objects either intruding into the first Fresnel zone or close to it, and which exhibit 'knife-edge' or 'smooth-sphere' characteristics, are shown in *Figure 11.1*. A negative F/F_1 indicates an intrusion. When the path clearance exceeds 0.6 times the first Fresnel zone radius, the free space loss is achieved.

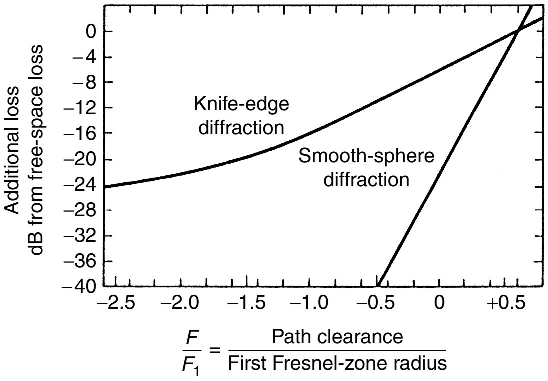

Figure 11.1 Knife-edge and smooth-sphere diffraction

Refraction also affects microwave propagation and temperature inversions may cause ducting resulting in a loss of signal and, possibly, interference. When ducting occurs, a layer of air of low refractive index is formed between two highly refractive layers and a wave may be trapped between them. Under these conditions a signal can be carried an abnormal distance before it can return to Earth resulting in loss of signal at the intended receiver and possibly interference at a distant one on the same frequency.

11.3 *K* factor

The degree and sense of refraction of wave is related to a factor known as K. Refraction normally bends the wave downwards, extending the range, but the refractive index of the atmosphere varies from place to place and with time and height. At times, during an atmospheric inversion for instance, and in some places on the earth the effect is reversed:

> When no bending occurs, $K = 1$.
> When $K > 1$, the bending is downwards, effectively increasing the Earth's radius.
> When $K < 1$, the bending is upwards, effectively reducing the Earth's radius.

For most of the time $K > 1$, 1.33 being accepted as the normal factor, but for small periods K may be less than one.

For a point-to-point link, the path between the transmitting and receiving sites should be a clear line of sight, although by making allowances for refraction, the Earth's radius has, up to now, generally been considered to be effectively increased by the factor of 1.33. However, in the interests of reliability a decrease to 0.7 times the radius (the minimum K factor considered likely to occur) is now often used in link planning. When a link has been planned using a higher K factor, a temporary reduction of K not only reduces the radio horizon but effectively raises objects close to the path, possibly to the point where they become significantly close to the first Fresnel zone.

It is usual when planning a link to plot the profile of a path on paper with curved horizontal graduations to represent the amended Earth's radius. The radio beam can then be drawn as a straight line between the antenna locations. *Figure 11.2* is an example.

11.4 Fresnel zones, reflections and multi-path fading

Signals which arrive at the receiver by more than one path as the results of reflection or diffraction may arrive in any phase relationship

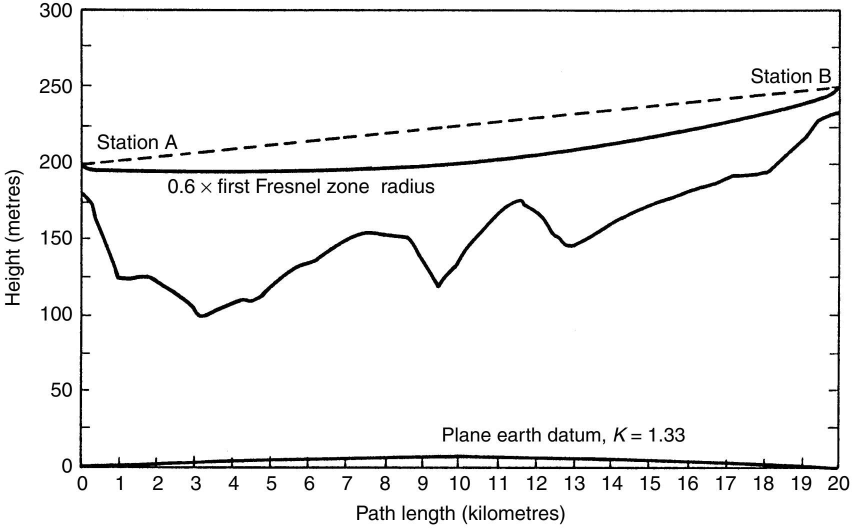

Figure 11.2 Example of path profile

to the direct wave. When they arrive anti-phase to the direct wave, cancellations result. The intensity and phase of the spurious signal may not be constant, thus providing random multi-path fading.

Where a carefully drawn profile of a link path shows there to be a clear line of sight, the effect of waves reflected or diffracted from objects close to the line of the direct wave must then be considered. The effect of these indirect waves can be predicted by calculating where the reflection occurs in relation to a series of ellipsoids which can be drawn around the line-of-sight path between the transmitting and receiving antennas. These ellipsoids, known as the Fresnel zones, contain the points where reflected waves will follow a path of constant length, as shown in *Figure 11.3*.

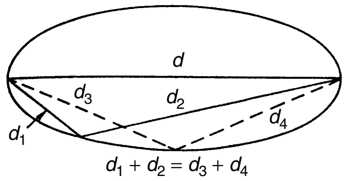

Figure 11.3 Fresnel zone: reflected path lengths

Waves reflected at the odd-numbered Fresnel zones will travel an odd number of half-wavelengths further than the direct wave but, because a 180° phase change usually occurs in the reflection process, will arrive at the receiver in phase with the direct wave. Waves reflected at even-numbered zones will arrive anti-phase to the direct wave with a cancelling effect. The effect of reflected waves diminishes with reflections from the higher order zones. The radius of a Fresnel zone in metres at the point of intrusion is given by:

First zone, F_1:

$$F_1 = 31.6\sqrt{\frac{\lambda d_1 d_2}{d}}$$

or

$$F_1 = 548\sqrt{\frac{d_1 d_2}{fd}}$$

where

F_1 = radius in metres at point of intrusion
$d_1 + d_2 = d$ (path length in km)
λ = wavelength in metres
f = frequency in MHz

Second zone:

$$F_2 = \sqrt{2} \times F_1$$

Third zone:

$$F_3 = \sqrt{3} \times F_1$$

and so on.

The degree of reflection from an object depends on its nature, the greatest reflection occurring from smooth flat ground or water. Where a path lies over the sea, variations in the path length of a reflected wave due to tides may render a path unusable. When the height of the antenna closest to the sea is varied, the *effect* of the reflected wave passes through a series of minima and maxima and adjustment of the height of that antenna can reduce or, occasionally, overcome the effect.

Atmospheric conditions change giving rise to fading and variations of the multi-path effects. The reliability of a link may be crucial to the success of a complete system and, where a critical path in terms of performance exists, long-term tests are advisable to ensure that variations of propagation do not reduce the reliability to an unacceptable level. Paths which contain obstacles in the line of sight which will cause additional losses are obviously suspect. So are those where objects or large stretches of water or flat ground which might produce diffraction or reflections of the wave lie close to the line of sight.

11.5 Performance criteria for analogue and digital links

The transmission quality for analogue modulated systems is based on the signal-to-noise ratio. The noise is specified relative to a standard test tone level and is commonly expressed as either picowatts psophometrically weighted (pWp) or decibels (dB) of C-message weighted noise above a reference noise level of −90 dBm (defined as 0 dBrnc0). Typical objectives range from 28 dBrnc0 for long-haul routes to 34 dBrnc0 for short-haul routes. When the signal fades the noise increases until the threshold noise level is reached. When the threshold (typically 55 to 58 dBrnc0) is exceeded, the transmission quality is considered unacceptable.

With digitally modulated systems the bit error rate (BER) is the measure of transmission quality. The bit error rate is the number of bit errors per total received averaged over a period of time. If the transmission rate is 10 Mbits per second and 100 bit errors occur over a 100 second period, the BER is 10^{-7}, an average of 1 error in 10^{6} bits. The acceptable level of transmission is determined by the type of traffic.

For PCM voice traffic, bit errors manifest themselves as clicks and a threshold of 10^{-6} (1 click approximately every 15 seconds) is usually considered acceptable. At this threshold the speech is intelligible, but beyond it the clicks become annoying and intelligibility falls rapidly.

For data with error correction a higher BER of 10^{-8} is normally acceptable (Communications International, 1989).

11.6 Terminology

A number of different units are used worldwide to define the performance and transmission levels of a radio relay system (for decibel definitions see Chapter 2). Important international definitions and units are:

- *Zero transmission reference point*. This is a point arbitrarily established in a transmission circuit, with all other levels in the circuit being stated with reference to this point. Its relative level is 0 dBr.
- *Standard test tone*. The standard test tone for use at audio circuit points is defined as a power of 1 milliwatt (0 dBm) at a frequency of 1000 Hz applied at the zero transmission reference level point.

11.7 Link planning

Planning a link involves producing a profile of the path and calculating the net loss in the system to arrive at a transmitter output power which will produce the designed signal-to-noise ratio at the receiver. A simple example using the profile of *Figure 11.2* is shown in Section 11.8.

Transmission lines and waveguides are discussed in Chapter 3 and microwave antennas in Chapter 4.

11.8 Example of microwave link plan

Frequency: 2000 MHz		
Antenna type, station A: P6F-17C	height agl.	20 m
Antenna type, station B: P6F-17C	height agl.	20 m
Feeder type, station A: LDF5P50A	loss, dB/100 m	6.46
Total length, antenna to equipment		30 m
Feeder type, station B: LDF5P50A	loss, dB/100 m	6.46
Total length, antenna to equipment		30 m

Performance	
Path length: 20 km, therefore clear path loss	124.0 dBi
Obstruction loss	0.0 dB
Feeder loss, station A:	1.9 dB
Feeder loss, station B:	1.9 dB
Feeder tail loss, total for link	1.5 dB
Connector loss, total for link	1.5 dB
Total loss	**130.8 dB**
Gain, antenna A	28.6 dB
Gain, antenna B	28.6 dB
Total gain	**57.2 dB**
Nett loss (total loss − total gain)	**73.6 dB**
Receiver threshold for max. signal/noise	−125.0 dBW
Design fade margin	+30.0 dB
Design receiver input level (threshold − fade margin)	**−95.0 dB**
Transmitter output power (receiver input − nett loss)	**−21.4 dBW**

Reference

Krzyczkowski, M. (1989). *Communications International*, August.

12 Information privacy and encryption

12.1 Encryption principles

Radio communication was never secret, but since the advent of fast frequency scanning receivers the ability to overhear, even on cellular radio telephones, is within easy reach of anyone. Privacy systems are, however, available which will deter the casual listener and gain time against the determined eavesdropper.

Messages, which may be either speech or data, are encrypted to prevent both eavesdropping and the injection of spurious information. The aim is to make the encryption and decryption as easy and inexpensive as possible for authorized users and time-consuming, difficult and costly for the eavesdropper (cryptanalyst). *Figure 12.1* shows a standard cryptographic channel.

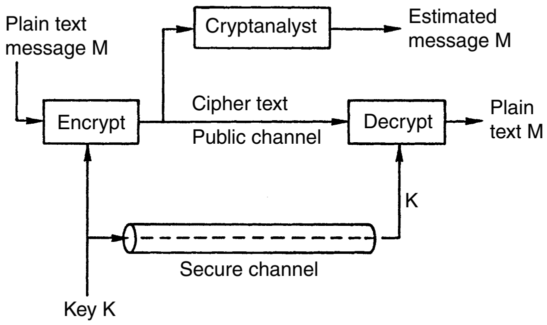

Figure 12.1 Cryptographic channel

A plain text message M (speech, written or digital) is encrypted by mixing with a key K to produce a cipher text. The cipher text may be transmitted over a channel which is accessible to the public and hence to the cryptanalyst. The key is issued via a secure channel to the authorized recipient who uses it to decipher the message. The cryptanalyst without access to the key attempts to derive the maximum information from the cipher text to enable him or her to estimate the content of the message.

One key may be used continuously or for long periods or, to increase the cryptanalyst's confusion, the key may be changed frequently, perhaps even for each character of the message. A sequence of key changes which repeats after a fixed number of characters produces what is known as periodic encryption.

Encryption may be either symmetrical or asymmetrical. Symmetrical encryption uses the same key for both encryption and decryption. Asymmetrical encryption uses a different key for each process, thus providing for different levels of authorization. Encryption keys may be supplied to many persons who are authorized to transmit encrypted messages but decryption keys may be issued to only a few authorized recipients.

12.2 Speech encryption

The encryption of speech offers fewer possibilities than does the encryption of written or digital data messages. The simplest method of encrypting speech is scrambling by inverting the speech frequencies; *Figure 12.2* shows this process.

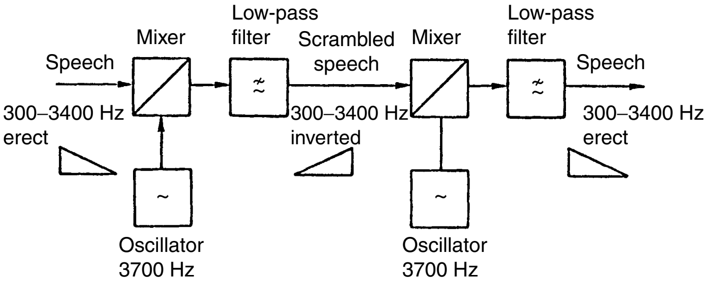

Figure 12.2 Speech inversion

The speech, contained in the band 300–3400 Hz, is mixed with a key frequency of 3700 Hz producing an erect, upper side band from 4000–7100 Hz, and an inverted, lower side band where the 300 Hz components of speech have become 3400 Hz and the 3400 Hz have been inverted to 300 Hz. The upper side band is rejected by a lowpass filter and the inverted lower side band is transmitted. In the receiver the scrambled speech is mixed again with 3700 Hz to produce an erect side band – the original non-inverted speech message.

Simple inverted speech is easily unscrambled. There is little choice of key frequency and if the eavesdropper uses a slightly different frequency the pitch is changed but the speech is readable. Also, if the centre frequencies only of the inverted band are selected by means of a band-pass filter, inverted speech becomes intelligible.

A sophistication which renders the speech more secure divides the speech band into sections and transmits them separately using a different key frequency for each band (audio frequency hopping). The divisions of the speech band may also be treated as blocks and transposed in time according to a user-programmable pattern to create further confusion in the mind of the eavesdropper.

The most up-to-date methods of speech scrambling convert the speech into digital form by either pulse code modulation (PCM) or some other method. The digits corresponding to the speech may then be either transmitted as frequency modulation, e.g. FFSK, on an analogue radio system or, possibly after further encryption, transmitted directly on a digital system. Digitized speech creates improved security not only by the digitization itself but by offering the higher encryption capabilities of data.

The price to be paid for security with analogue encryption is a degradation of the received signal-to-noise ratio by 9 dB, effectively reducing the range of a transmitter by approximately 40%.

12.3 Data encryption

Digital data may be encrypted by changing the digits, perhaps adding superfluous digits, and transmitting the resultant cipher message either in blocks of a fixed size or as a stream.

Block encryption treats the blocks in the same way as in the encryption of speech, with different keys being used for each block – or each character – and the blocks re-distributed in time.

Stream encryption has no fixed block size and each plain text bit, M_i, is encrypted with the ith element, K_i, of a pseudo-random, sometimes called pseudo-noise (PN), key.

Figure 12.3 shows two methods of generating pseudo-random keys or pseudo-noise. The first, *Figure 12.3(a)*, operates as follows. At each clock pulse the contents of the pre-loaded four-stage shift register are stepped forward from left to right. Immediately after the shift, the output bit is fed back into the 1st and 2nd stages. It introduces a new bit into stage 1, and is added by modulo 2 addition to the new content of stage 2, producing a new set of contents. The initial loading of 1100 emerges as a pseudo-random 15 bit sequence which then repeats. The periodicity of the sequence is given by:

$$\text{Sequence length, characters} = 2^N - 1$$

where N = the number of stages in the shift register.

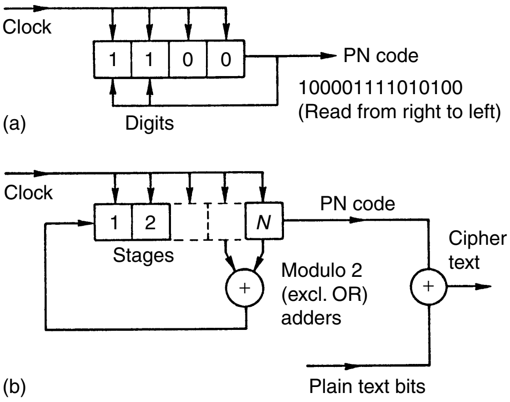

Code length = $2^N - 1$
e.g. $N = 5$, length = 31 characters

Figure 12.3 Pseudo-noise (PN) generation

For a four-stage register, therefore, the sequence repeats after 15 bits and is shown in *Figure 12.3(a).*

A more commonly used method combines the outputs of two or more of the earlier stages in a modulo 2 adder and feeds the result back to the input of the register as in *Figure 12.3(b).*

To form the cipher text the resultant pseudo-random key is mixed with the original data message in a second modulo 2 adder. If the clock rate for the shift register is the same as the bit rate of the plain text message, the plain text bits are exchanged for those of the modulo 2 sum, but if the shift register runs faster than the plain text bit rate, additional bits are added into the cipher text. This is more common and extends the time taken by a cryptanalyst to estimate the message. The price to be paid for the improved security is either a slower effective bit rate for the message or a higher overall bit rate and hence an increased bandwidth requirement.

Mixing the cipher text with the output from an identical PN generator in the receiver recovers the original text.

Modulo 2 addition

A modulo 2 adder is an exclusive-OR gate which produces a logic 1 output whenever either of the inputs is at logic 1 and the other is at logic 0. When both inputs are identical, the exclusive-OR gate produces a logic 0 output.

The truth table for an exclusive-OR gate is:

Input *A*	 *B*	*Output* *Y*
0	0	0
0	1	1
1	0	1
1	1	0

Modulo addition is not limited to two inputs. Any quantity of binary numbers may be added: if there is an odd number of logic 1s in a column, the adder produces a logic 1 output, if an even number, i.e. no remainder in the binary addition, the output is logic 0.

Synchronous encryption

The key is generated independently of the message from a previously loaded register. If a character is lost during transmission of a synchronous text, resynchronization of transmitter and receiver key generators is necessary before transmission can continue.

Self-synchronous encryption

The key for each character is derived from a fixed number of previous characters of the plain text message. For example, the shift register is pre-loaded with the plain text characters so that in a four-stage register the key used for encrypting the 4th character will be the 4th previous message character. If a self-synchronized transmission loses a character, the system automatically re-synchronizes the same number of characters (in this case four) later.

Written messages may be encrypted using one of the classical mechanical methods of rearrangement of the letters before digital encryption.

When sufficient RF channels exist frequency hopping is a further possibility, and the spread spectrum technique, where the signal energy is spread over a very wide band of frequencies, not only offers very high security but also makes detection of the signal difficult. The shift register techniques described above are also used for generation of the frequency hopping sequence and the spreading of the base band frequencies.

12.4 Code division multiple access (CDMA) or spread spectrum

The extension of pseudo-random key or noise generation is code division multiple access or spread spectrum transmission, described in Chapter 8. The spread spectrum technique provides an extremely high level of security by reducing the radiated energy at any one frequency to very little above the ambient noise level by spreading the transmission over a very wide band. The transmitter uses what is in effect an extended digital key to spread the bandwidth and the receiver is equipped with an identical key for de-spreading. The transmission almost disappears into the noise and, without the appropriate key, the existence of a spread spectrum signal is very difficult to detect.

12.5 Classification of security

Unconditionally secure

Those systems where the cryptanalyst has insufficient information to estimate the content of the cipher regardless of the amount of time and computation facilities available. This is only realistic when using what is known as a one-time pad where the key is used once and once only.

Computationally secure

Encryption systems are specified in terms of the amount of time taken by a cryptanalyst to estimate the cipher's contents using the state of the art techniques. Unless an extremely long periodicity is used for a progressive key – months and even years in some instances – requiring many stages in the shift register, it is possible for a cryptanalyst who knows, or can estimate, a small part of the message to calculate all the parameters necessary to decipher the message. However, stream encryption with a pseudo-random key approaches perfect secrecy for a finite number of messages.

References

Chambers, W.G. (1985). *Basics of Communications and Coding*. Oxford University Press, Oxford.

Sklar, B. (1988). *Digital Communications*. Prentice Hall, Englewood Cliffs, NJ.

13 Multiplexing

Multiplexing enables several information (speech or data) channels to be carried simultaneously over one bearer, a wide band, single frequency microwave radio link for example. Both frequency and time division multiplex are common methods. On trunked radio systems where channels are allocated to users on demand the multiplexing is referred to as frequency- or time-division multiple access (FDMA, TDMA).

Frequency division multiplexing (FDM) used to be a common technique for sharing analogue trunk circuits between 12 (or multiples of 12) separate channels. This is an analogue technique that been superseded by the use of digital trunk circuits, where digital multiplexing is employed. FDM is still used on some satellite and microwave links, although most links are now using digital techniques. Some systems such as cellular telephones use separate transmit and receive frequencies, which is known as frequency division duplexing (FDD).

13.1 Frequency division multiplex

Frequency division multiplex divides a broad band of frequencies into slots, each wide enough to accommodate an information channel. This is achieved by amplitude modulating a higher frequency subcarrier with each speech signal to form groups of channels.

Each speech channel contains frequencies between 300 and 3400 Hz plus, in some systems, an out-of-band signalling tone of 3825 Hz and a guard band. Each channel modulates a base band subcarrier spaced at 4 kHz intervals upwards from 64 kHz. This produces an upper and lower side band from each channel (*Figure 13.1*). The carrier and upper side band are removed by filters and the lower (inverted) side band is transmitted. At the receiver, the base band frequencies are again mixed with the same subcarrier frequency to restore the original speech. The subcarrier frequencies are maintained to an accuracy of ±1 Hz which creates the ±2 Hz frequency translation error quoted in some telephone line specifications.

Twelve such channels form a ITU-T basic group B occupying the band between 60 and 108 kHz. This basic group B may now be

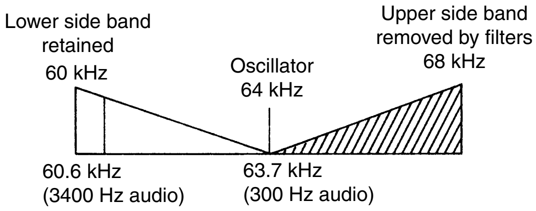
Lower side band
retained
60 kHz
Oscillator
64 kHz
Upper side band
removed by filters
68 kHz
60.6 kHz
(3400 Hz audio)
63.7 kHz
(300 Hz audio)

(a) Production of side bands in frequency division multiplex

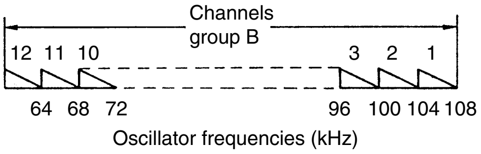
Channels
group B
12 11 10
3 2 1
64 68 72
96 100 104 108
Oscillator frequencies (kHz)

(b) Production of basic group B

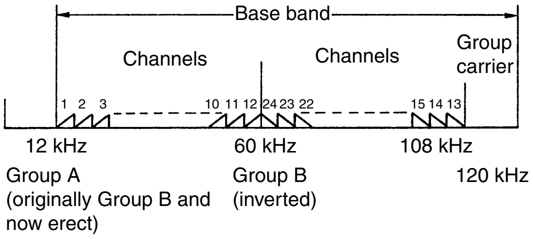
Base band
Channels
Channels
Group
carrier
1 2 3
10 11 12
24 23 22
15 14 13
12 kHz
60 kHz
108 kHz
Group A
(originally Group B and
now erect)
Group B
(inverted)
120 kHz

(c) Production of groups A and B

Figure 13.1 Frequency division multiplex

mixed with 120 kHz to produce a lower side band of 12 to 60 kHz, now basic group A. Filters leave 60 to 108 kHz free for a new basic group B.

The process may be repeated by using five basic groups to modulate still higher-frequency carriers, to produce super- and hypergroups.

For FDM data communications, the bearer circuit bandwidth of 3000 Hz is divided into 12 channels each of 240 Hz bandwidth. Data is transmitted at 110 bits per second allowing a send-and-receive channel in each block of 240 Hz.

13.2 Time division multiplex (TDM)

A time division multiplex system conveys digital data, and speech must first be converted to data. TDM allocates short-duration time slots within a wider time frame to each information channel. For example, a continuous stream of data sent over a link at a rate of 2400 bit/s could convey the information contained in four 600 bit/s channels in short sequential bursts.

If the duration of one input bit is 1/600 s or 1.666 ms, a seven-bit character occupies 11.66 ms and 85 such characters can be sent per second. If the transmitted rate can be speeded up each bit sent at 2400 bit/s has a duration of 416 μs and 343 seven-bit characters can be sent per second. Such a system is shown in *Figure 13.2*. The data is stored in the buffers at the transmitter and the clock pulses are applied to each store/gate sequentially allowing one character from each data channel to be transmitted at a rate of 2400 bit/s. Perfect synchronization must be maintained between all channels and the transmitter and receiver to avoid data errors.

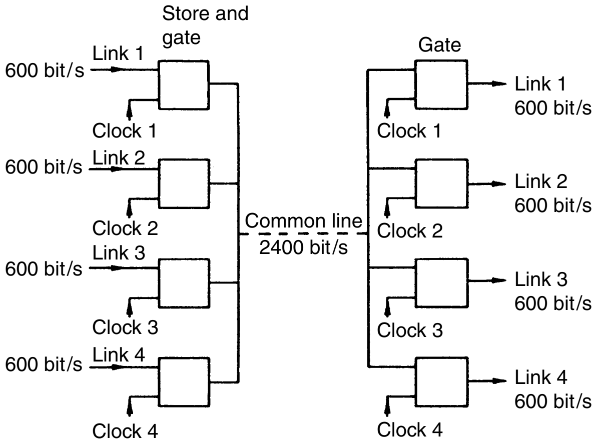
Store and gate
Link 1
600 bit/s
Clock 1
600 bit/s
Link 2
Clock 2
Link 3
600 bit/s
Clock 3
600 bit/s
Link 4
Clock 4
Common line
2400 bit/s
Gate
Link 1
600 bit/s
Clock 1
Link 2
600 bit/s
Clock 2
Link 3
600 bit/s
Clock 3
Link 4
600 bit/s
Clock 4

Figure 13.2 Time division multiplex

The principle is illustrated in *Figure 13.3* where a 1 s time frame at 2400 bit/s contains 343 time slots, each of 2.92 ms and containing a character from a specific information channel. Every second, therefore:

channel A would occupy slots 1, 5, 9, ..., 337

channel B would occupy slots 2, 6, 10, ..., 338

channel C would occupy slots 3, 7, 11, ..., 339

channel D would occupy slots 4, 8, 12, ..., 340

This leaves three blank slots; in practice slots are also allocated for preamble, address and synchronization purposes.

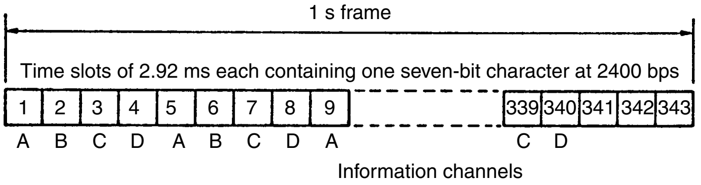
1 s frame
Time slots of 2.92 ms each containing one seven-bit character at 2400 bps
1
2
3
4
5
6
7
8
9
339
340
341
342
343
A B C D A B C D A
C D
Information channels

Figure 13.3 TDM time frame

13.2.1 European E1 multiplexing

The digital base band signal adopted in Europe operates at 64 kbit/s and the multiplexed signal at 2048 kbit/s. An eight-bit word or sample of a PCM voice channel (see Chapter 14) occupies 3.9 μs, and the interval between successive samples of a channel is 125 μs, the time frame duration. Therefore, the number of channels (time slots) that can be accommodated in one frame is $125/3.9 = 32$. Thirty of the slots are used for information channels and two for control purposes (*Figure 13.4*).

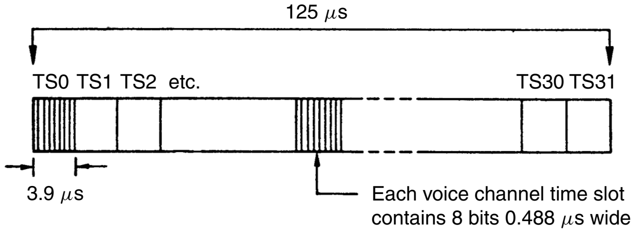
125 μs
TS0 TS1 TS2 etc.
TS30 TS31
3.9 μs
Each voice channel time slot contains 8 bits 0.488 μs wide

Figure 13.4 TS0, TS16 are used for signalling; TS1–TS15 and TS17–TS31 are used for voice channels

In the European E1 system, timeslots are numbered 0 to 32. Timeslot zero (TS0) is used for synchronization and TS16 is used for signalling. The signalling channel is therefore operating at 64 kbit/s. Digitized voice is carried in the remaining 30 timeslots. Voice channels 1 to 15 are carried over TS1–15. Voice channels 16–30 are carried over TS17–31.

13.2.2 American T1 (D1) multiplexing

The American T1 system is a 1.544 Mbit/s transmission system. The data format should be referred to as DS-1, but T1 is sometimes used instead. Timeslots (TS), or channels, in this system are numbered 1 to 24; there are 23 available to carry traffic and TS24 is used for synchronization. Each timeslot carries 8 bits, sampled at 8 kHz. An additional bit is sent at the end of each frame (known as the frame-bit or F-bit) so there are 24, 8-bit timeslots, plus 1 bit = 193 bits per frame.

The synchronization word sent in TS24 is coded as 1 0 1 1 1 Y R 0. The Y and R are dependent upon the data multiplexer and are for the manufacturer's use.

Signalling information is transmitted using 'stolen' bits from each channel, rather than using a separate timeslot as in the European E1 system. Bits are not stolen from timeslots in every frame, because this would degrade the quality of a voice channel. Instead, the least significant bit of every sixth frame is used for signalling. Thus the effective word length of the encoded voice is 7.833 bits, rather than 8 bits of the E1 system.

There are two virtual signalling channels (A and B) created by the use of stolen bits. Signalling Channel A is derived from bits stolen from timeslots during the sixth frame. Signalling Channel B is derived from bits stolen from timeslots during the twelfth frame. This is repeated for every superframe. Each channel has a data rate of 24 bits per 12 frames (1.5 ms), or 16 kbit/s, giving 32 kbit/s signalling in total. Although this data rate is half that of the E1 system, there are only 23 voice channels instead of the 30 used in the E1 system.

The F-bit is used to identify the frame, which is important when extracting the signalling information. Because only one in six frames are used for signalling, it is vital to know which ones carry the information and which ones carry voice traffic. Also there are two signalling channels and these must be separated at the receiver end of the system. A six bit data pattern of 1 0 0 0 1 1 is transmitted, followed by the inverse: 0 1 1 1 0 0. The complete 12-frame sequence is known as a superframe.

13.3 Code division multiple access (CDMA)

Spread spectrum transmission (described in Chapter 8) is a form of multiplexing. In addition to high security it permits multiple occupation of the wide – typically 1.25 MHz – frequency band. A number of users possessing keys of low correlation can occupy the same band at the same time. The system operates well in poor signal-to-noise or high interference environments.

Digital signals can be multiplexed by coding each source individually. Instead of transmitting logical 1's and 0's, a data pattern is assigned to each bit. The pattern for a logical 1 could be a ten-bit word (e.g. 1011010011); a separate pattern would represent a logical 0. Correlation (i.e. pattern matching) between a known signal and a signal received with a certain code gives an indication of whether the bit received is a 1 or a 0. This technique allows the occasional error in the data path without affecting the received bit. More often, it allows several users to transmit at the same frequency. Each user is allocated a different data pattern so that only the correct person will be able to decode the signal; all other signals will be ignored.

Reference

Winder, S.W. (2001). *Newnes Telecommunications Pocket Book*. Butterworth-Heinemann, Oxford.

14 Speech digitization and synthesis

14.1 Pulse amplitude modulation

The digitization of analogue waveforms by pulse code modulation is accomplished in two stages. First the waveform is sampled to produce pulse amplitude modulation (PAM). Short-duration samples are taken at regular intervals which are long compared with the sampling time but short in relation to the highest signal frequency. The result is a train of pulses whose amplitude envelope is the same as the analogue waveform. The envelope (*Figure 14.1*) will contain:

- clock frequency f_c, the sampling rate
- all the signal frequencies contained in the waveform from the lowest, f_1, to the highest, f_2
- upper and lower side bands, $(f_c - f_2)$ to $(f_c - f_1)$ and $(f_c + f_1)$ to $(f_c + f_2)$
- harmonics of f_c and the upper and lower side band frequencies
- a DC component equal to the mean value of the PAM waveform.

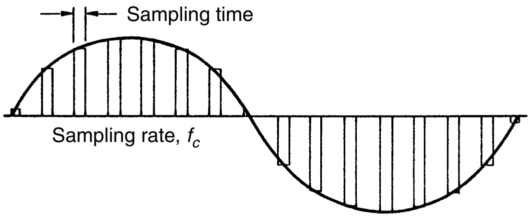

Figure 14.1 Pulse amplitude modulation

The envelope contains the original signal frequencies and can be demodulated by a low-pass filter which will pass f_2 but not the clock frequency. The clock frequency must therefore be higher than $2f_2$. For line communications in the UK a clock frequency of 8 kHz is used with a maximum modulating frequency of 3400 Hz.

14.2 Pulse code modulation

To overcome the susceptibility of PAM signals to corruption of the amplitude waveform by noise and interference, the waveform is processed further to produce pulse code modulation (PCM) before transmission.

In this process the magnitude of the PAM samples with respect to a fixed reference is quantized and converted to a digital (data) signal. Quantizing rounds off the instantaneous sample pulse amplitude to the nearest one of a number of adjacent voltage levels. *Figure 14.2* illustrates the process for an eight-level system. In the figure the amplitude at t_0 is 2, between 4 and 5 at t_1, between 5 and 6 at t_2, etc. After quantization the values would be 2 at t_0, 5 at t_1, 5 at t_2 and so on. The difference between the amplitude levels and the rounded-off values is the quantization noise or distortion.

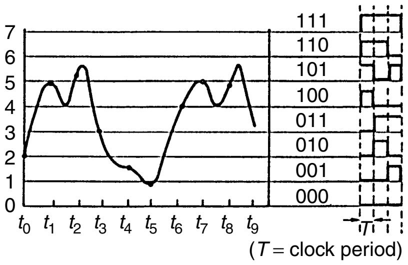

Figure 14.2 Eight-level pulse code modulation

The binary pulse train – leaving a one-bit synchronizing space between each number – for this example would be as in *Figure 14.3*.

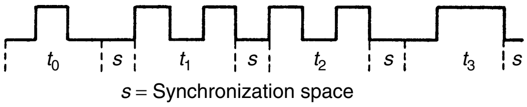

Figure 14.3 Binary pulse train (from *Figure 14.2*)

The number of quantizing levels is 2^n and the highest denary number represented is $(2^n - 1)$ where n is the number of bits used to represent each sample. If each train of pulses is accompanied by one synchronizing bit, the number of bits per sample is $(n + 1)$. If the sampling rate is f_c, the transmitted bit rate is $(n + 1)f_c$. For example, an eight-bit word – including the synch. bit – is commonly used to represent a sample, so, with a clock frequency of 8 kHz:

$$\text{number of quantizing levels} = 2^7 = 128$$

$$\text{sampling rate, } f_c = 8\,\text{kHz}$$

$$\text{therefore transmitted bit rate} = (n + 1)f_c$$

$$= 64\,\text{kb/s}$$

The maximum frequency of the pulses will be when transmitting alternate 1s and 0s, and the occupied bandwidth, $\frac{1}{2} \times$ bit rate, 32 kb/s.

Quantization as described above is linear, i.e. the spacing of the quantization levels is the same over the range of pulse amplitudes. This produces a poor signal-to-noise ratio for low-level signals which is improved by the use of more closely-spaced levels at small signal amplitudes than at large amplitudes. Where non-linear quantization is used the most significant bit of the sample character identifies the polarity of the signal.

If linear quantization is used then about 12 bits per sample are needed to give a good voice quality, which requires a bit rate of about 96 kbits/s. For coding speech using non-linear quantization 8 bits per sample is sufficient, which requires a bit rate of 64 kbits/s. Two slightly different non-linear PCM codecs were standardized in the 1960s. In America u-law coding is the standard, but in Europe A-law coding is used.

14.3 ADPCM codecs

Adaptive differential pulse code modulation (ADPCM) codecs are waveform codecs which quantize the difference between the speech signal and a prediction of the speech signal. If the prediction is accurate, then the difference between the real and predicted speech samples will be small. The signal will be accurately quantized with fewer bits than would be needed using standard PCM. At the decoder the quantized difference signal is added to the predicted signal to give the reconstructed speech signal.

The performance of the codec is aided by using adaptive prediction and quantization. The predictor and difference quantizer adapt themselves to the changing characteristics of the speech being coded.

The CCITT (now ITU-T) standardized a 32 kbits/s ADPCM, known as G721, in the 1980s. The quality of the reconstructed speech was almost as good as that from standard 64 kbits/s PCM codecs. Since then G726 and G727 codecs operating at 40, 32, 24 and 16 kbits/s were standardized.

14.4 The G728 low delay CELP codec

At bit rates of around 16 kbits/s and lower the quality of waveform codecs falls rapidly. Thus, at these rates, CELP codecs and their derivatives, tend to be used. CELP stands for Code Excited Linear Prediction. A code representing a voice pattern is transmitted, rather than coding short samples of voice signal. The code is calculated by looking at the long-term changes in the voice and then predicting how the voice pattern will change. Since the pattern takes some time to establish, CELP codecs produce a delay.

A backward adaptive CELP codec was standardized in 1992 as G728. This codec uses backward adaptation in its calculations. This means that, rather than buffer 20 ms or so of the input speech for its calculations, the pattern of previous speech is used. This means that the G728 codec has to analyse fewer samples than traditional CELP codecs. The G728 codec uses only 5 samples giving it a total delay of less than 2 ms.

14.5 The GSM codec

The 'Global System for Mobile communications' (GSM) is a digital mobile radio system which is extensively used throughout Europe, and also in many other parts of the world. The GSM codec provides good quality speech, although not as good as the slightly higher rate G728 codec.

The GSM full rate speech codec operates at 13 kbits/s and uses a regular pulse excited (RPE) codec. This codec has short-term and long-term predictors to estimate how the voice pattern will change over time.

The input speech is split up into 20 ms long sections, called frames, and for each frame a set of eight filter coefficients are calculated for the short-term predictor. Each frame is then divided into four 5 ms sub-frames, and for each sub-frame the encoder calculates the delay and gain needed for the codec's long term predictor. Finally, after both short and long term prediction has been applied, the residual signal is quantized for each sub-frame. The filter coefficients for the short term predictor, the delay and gain of the long-term predictor and the coded residual signal are all transmitted. The receiver uses all these values to reconstruct the signal.

References

Kennedy, G. (1977). *Electronic Communications Systems*, McGraw-Hill Kogashuka, Tokyo.

Winder, S.W. (2001). *Newnes Telecommunications Pocket Book*, Butterworth-Heinemann, Oxford.

15 VHF and UHF mobile communication

Mobile communication operating throughout the VHF and UHF bands is expanding rapidly and, although the fastest growth has been in the radio-telephone field, interest in private mobile radio (PMR) is undiminished. Most of the procedures described for PMR are applicable to many other branches of radio communication, so traditional PMR is considered here. Other systems are discussed in later chapters.

A private mobile radio system, comprising a base station and mobiles, is one that is effectively owned by the user and, under the conditions of the licence, may only be operated by the user's own staff for his or her own business. Airtime cannot be leased to other persons.

15.1 Operating procedures

Frequencies are normally allocated in pairs, one for the up-link to the base station and one for the down-link to the mobile (not to be confused with a radio link used for control purposes). Such a pair of frequencies, spaced sufficiently apart to permit simultaneous transmission and reception by a station, comprises the radio channel. Occasionally, for special purposes and for small, low-power, and possibly temporary, operations a single frequency only may be allocated. The mountain rescue teams are an example.

The methods of operating are as follows.

Single-frequency simplex

This method uses a common frequency for transmission and reception by all stations operating on the system (*Figure 15.1*). Transmission

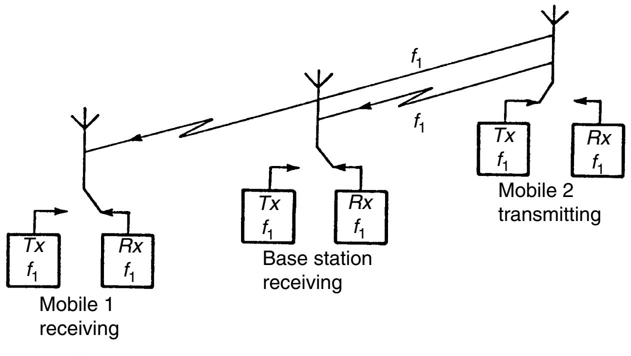

Figure 15.1 Single-frequency simplex

and reception cannot take place simultaneously at a station, and a receiver is switched off whenever the transmit switch is operated. This prevents blocking of the receiver by the transmitter and acoustic feedback occurring. The method allows all stations within range to hear both sides of a conversation and to relay messages to more distant stations; an obvious advantage for mountain rescue.

At the end of a transmission an operator must say 'over' and switch off the transmitter to hear the reply. A conversation must end with the word 'out' so that other stations are aware that the system is unoccupied.

Two-frequency simplex

Separate frequencies are used for transmission and reception but whilst a station is transmitting its receiver is still switched off (*Figure 15.2*). Mobiles hear only the base station and, therefore, the relaying of messages is not possible. A further disadvantage is that because mobiles hear only the base station, they may be unaware of the system occupancy and transmit, interfering with an existing conversation. The advantage of two-frequency simplex is the avoidance of receiver blocking or de-sensitization, not only from the associated transmitter but also, at base stations where several channels within the same band are located, from nearby transmitters.

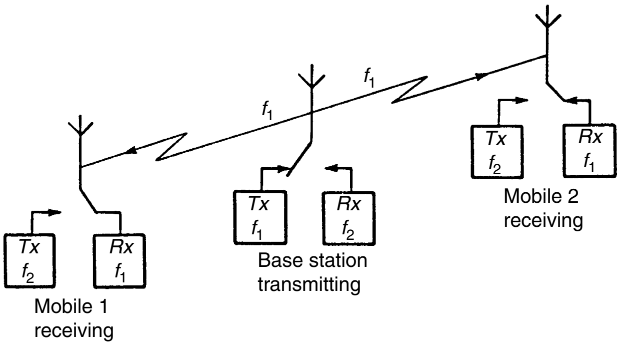

Figure 15.2 Two-frequency simplex

Duplex and semi-duplex

Separate frequencies are used for transmission and reception and, in full duplex, all stations can transmit and receive simultaneously as in a two-way telephone conversation (*Figure 15.3*). While a station is transmitting, its receiver audio output is switched from the loudspeaker

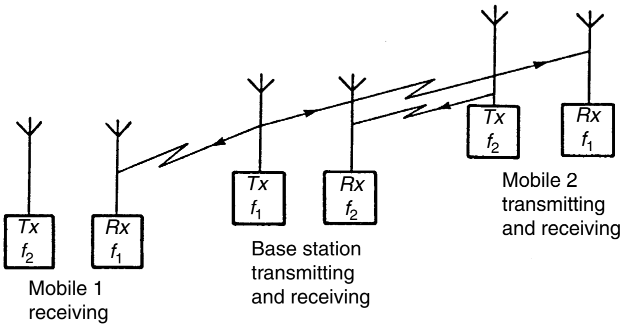

Figure 15.3 Duplex

to an earpiece to prevent acoustic feedback. A mobile cannot receive other mobiles directly but full duplex enables all stations to break in on a conversation in an emergency or to query part of a message; it also facilitates the use of talk-through where mobiles can speak to each other via the base station. To maintain awareness of system occupancy the base station may transmit a series of pips as an engaged signal during pauses in the despatcher's speech.

Many systems operate semi-duplex where only the base station operates a duplex procedure and the mobiles use a simplex procedure. This avoids the higher cost of duplex mobiles and offers most of the facilities of duplex, except that a despatcher cannot break in on a transmitting mobile.

Open channel/'All informed'

All mobiles hear all the calls from control, i.e. no selective calling is in operation.

Selective calling

Mobile receivers remain quiescent until specifically addressed; the opposite of open channel working. Individual mobiles, groups and a whole fleet may be addressed.

Auto-acknowledgement

When selectively called a mobile automatically transponds, sending the code for its address and, possibly, its status information. Mobiles can only acknowledge when individually addressed; auto-acknowledgement on group and fleet calls is prevented to avoid mobiles transponding simultaneously.

Status updating

The transmission, automatically or manually, of the data denoting the mobile's current status.

Call stacking

The storage of calls from mobiles and their presentation in call order to the despatcher. Arrangements are usually made to raise urgent calls to the top of the stack with an enhanced display.

15.2 Control of base stations

Where adequate radio signals over the desired service area of the system can be provided by one base station sited at the control point, it can be easily controlled, either directly from the front panel or over a multi-way cable.

In many instances, to obtain adequate coverage, the base station must be sited on high ground remote from the control point. Then, either a land line or a radio link, both of which will today probably be digital and the link microwave, must be used for control. A land line will most likely be rented but radio links are favoured by those users who insist on the complete system being under their direct control.

15.3 Common base station (CBS) operation

An economic method of providing mobile communication for users with a small quantity of mobiles and light traffic is a common base station or community repeater system. Although still referred to as private mobile radio, the base station is shared by several users who pay a fixed subscription for the service irrespective of airtime used. The station is controlled by radio using tone-controlled talk-through or what is sometimes called reverse frequency trigger (not normally permitted on a true PMR system). *Figure 15.4* shows the layout of a single-station CBS system. Each participant's office contains a fixed transmitter/receiver operating on the same frequencies as a mobile and equipped with CTCSS. Different CTCSS tones are assigned to each user to ensure privacy. All office stations must use a directional antenna so that they access only the base station to which they subscribe.

When a user transmits, from either the office station or a mobile, the base station, on receipt of a signal containing a valid CTCSS tone,

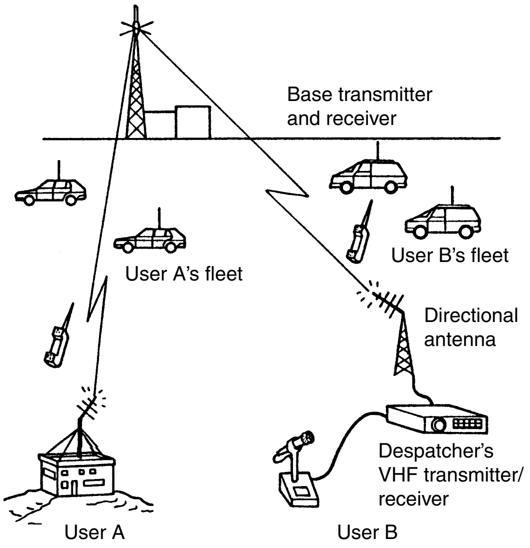

Figure 15.4 Common base station system

enters a talk-through mode. The caller's CTCSS tone is retransmitted and, in turn, opens the mute on all that user's mobiles allowing communication via talk-through. Mobiles are equipped with 'busy' lamps to indicate whenever the system is engaged, and transmission time-out timers to prevent excessively long calls excluding other users.

15.4 Wide area coverage

Where the desired service area for one radio channel is larger than can be covered by a single base station the way in which the base stations are to be controlled requires special consideration. Crucial aspects are the presentation of the best received signal from a mobile to the control operator or despatcher and the selection of the transmitter most likely to provide the best signal at the location of the mobile.

Receiver voting

Well-proven receiver voting circuits which present the best received signal to the despatcher have been used for many years. These circuits sample the signals received from a mobile at each base station and, by means of coded information–which may be either digital or in the form of continuous tones–enable equipment at the control centre to automatically select the best. The selection may be made by comparing either the signal-to-noise ratio or the signal strength of the received signal. If the information is to be used solely to select the best signal in terms of readability, the signal-to-noise ratio is probably the better characteristic to use, but if the information is also to be used to select a base transmitter, the signal strength could be considered more satisfactory. Some systems utilize both types of information.

A typical 3-station voting system is shown in *Figure 15.5* where signal sampling and vote encoding occur at the receiving sites and the coded information is passed over the base station control system to the control centre. This method is necessary when the selection is made on a signal strength basis, but where the signal-to-noise ratio is used for the selection the sampling and encoding can be done at the control centre taking into account the noise occurring in the control network. Receiver voting systems operate very quickly, and changes

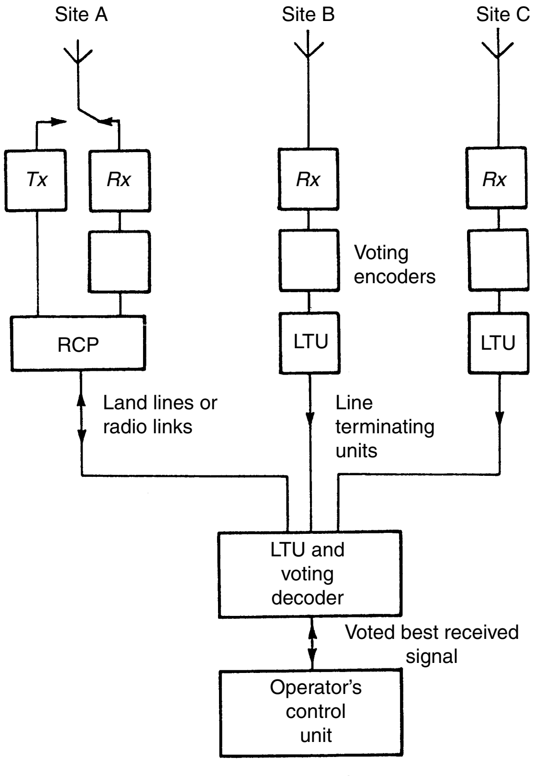
Site A
Site B
Site C
Tx
Rx
Rx
Rx
Voting encoders
RCP
LTU
LTU
Land lines or radio links
Line terminating units
LTU and voting decoder
Voted best received signal
Operator's control unit

Figure 15.5 Radio scheme with one transmit/receive plus two receiver-only stations with voting

of the selected receiving site may occur several times during a message without the despatcher's awareness.

The broadcast transmit and receive paths are not always reciprocal, for instance when low power hand-portables are integrated with higher power vehicular mobiles. In these circumstances the use of additional receiver-only fill-in stations is an economic and satisfactory proposition.

Base transmitter control poses a problem much more difficult to resolve. The selection of a base transmitter to communicate with a mobile whose precise whereabouts may be unknown, the broadcasting of messages to all, or groups of, mobiles whose locations may be widespread, and the provision of talk-through between mobiles are all facilities required on major systems, and difficult to provide satisfactorily. Apart from trunking, which can be economic only on very large networks, there are three traditional methods of operating the base transmitters on wide area schemes: manual selection, automatic selection and simultaneous transmission from more than one transmitter.

Manual transmitter selection

On many systems the despatchers select the transmitters manually. It is the simplest and least expensive method but has serious disadvantages:

1. Making the selection may entail trying a number of transmitters before sending the message, increasing the operator's work load and wasting air time.
2. Mobiles outside the service area of the selected transmitter may call, interrupting an existing conversation, either because they are unaware that the system is engaged or have received a poor signal that they believe may have been intended for them. Transmitting bursts of 'engaged' pips sequentially over the unselected transmitters during pauses in the despatcher's speech alleviates the first situation; the use of selective calling overcomes the second.
3. Broadcast messages must be transmitted on each transmitter in turn and talk-through between mobiles which are not in the service area of the same transmitter is not practicable.

Automatic transmitter selection

Selecting the transmitters automatically, or semi-automatically, is an improvement over manual selection. On a system where the mobiles are not equipped with selective calling and automatic acknowledgement of a call, automatic selection of transmitters can only occur on receipt of a call from a mobile. The transmitter through which to reply is then selected by the receiver voting system at the same time as it selects the signal to present to the despatcher. The selection is made at the start of the call and, because of the switching times involved, the transmitter selected is retained for the duration of the call. On these open channel systems, a calling despatcher must initially manually select a transmitter.

On systems where selective calling and auto-acknowledgement of a call are provided, the system can be made fully automatic by transmitting the data corresponding to a mobile's call sign from each transmitter in turn until a satisfactory acknowledgement is received. The successful transmitter is then retained for the duration of the conversation and, at its conclusion, is usually the one used to commence another call.

Such a system overcomes the disadvantage of the need to manually select the transmitter but the problems with mobiles outside the service area of the selected transmitter, and of broadcasting and talk-through, remain.

Simultaneous transmission

Operationally, simultaneous transmission from all sites is ideal and, under various names such as Spaced Carrier, Simulcast and Quasi-synch., has been around since the mid-1940s. Its operational value is proven but systems require special care in their planning, adjustment and subsequent maintenance.

An early form of simultaneous transmission was the amplitude modulated spaced carrier system. Used very successfully at VHF on systems using 25 kHz channel separation, the transmitter carrier frequencies were separated by 7 kHz – above the mobile receivers' audio pass-band. With the reduction of channel spacings to 12.5 kHz, spaced carrier operation on this basis was no longer possible and alternatives are to either synchronize, or very nearly synchronize, the carrier frequencies. There are, however, undesirable effects of synchronous and quasi-synchronous transmission but, with care, these can be reduced to an acceptable level. They are:

1. The beat note between transmitters being audible in the mobile receiver.
2. Variations in signal level due to interference patterns between signals from more than one transmitter.
3. Distortion due to audio phase differences and differing modulation levels when signals of comparable strength are received from more than one transmitter.

The beat note is easily dealt with. It is rendered unobjectionable either by placing it outside the mobile receiver audio pass-band – which is the usual method – or by synchronizing the transmitter carrier frequencies so that no beat note is produced. Synchronization, however, raises other problems, which are particularly severe at VHF but less so at UHF.

The spaced carrier system placed the beat note above the receiver pass-band, but modern systems place it below. This means that the beat note between the lowest and highest carrier frequencies must be below about 150 Hz if it is to be unobtrusive. Tests have shown the optimum carrier separation to be from 0.5 Hz to 4 Hz between any two transmitters using amplitude modulation, and from 5 Hz to 40 Hz between adjacent frequency transmitters using angle modulation.

Two transmitters on the same or closely-spaced frequencies produce deep nulls in the received signal in the areas where they provide almost equal strength signals. This is a natural phenomenon and the effect can only be reduced by the correct siting of stations and antenna configurations. Because of the longer wavelength the effect is more detrimental at VHF than UHF. Where the frequencies are quasi-synchronous the interference pattern is continually moving, but with synchronized carriers the pattern is virtually stationary and at low band VHF wavelengths it is possible to stop a vehicle in a place of semi-permanent zero signal. While moving slowly in an area of equal signal strengths from two transmitters, the cancellations become very objectionable. At 450 MHz the distance between the nulls is so short that it is almost impossible to remain in one, and while moving they are unnoticed. Strong signals in the overlap areas minimize the time that the signals fall below the receiver noise threshold at each cancellation. They are the key to reducing the annoyance from the cancellations. Amplitude modulated systems have an advantage in that the receivers produce less severe bursts of noise during the signal nulls.

The audio distortion, provided the equipment does not introduce significant additional harmonic distortion, is attributable to audio phase differences in the signals received from more than one transmitter. For the distortion to be severe, the received signals must differ by less than 6 dB on an angle modulated system; capture effect in the receiver removes the audible effects at greater differences.

The phase differences arise from differing audio characteristics in all the circuits in the path including land lines, radio links and control equipment. Common to all systems are the phase differences due to the different path lengths from the control centre to a receiving mobile, but where land lines or multiplexed circuits are used for the control of the base stations, variable bulk and group delays and frequency translation errors can present serious difficulties. Tests have shown AM to be slightly more tolerant than angle modulation in respect of this type of distortion, phase delays of 100 μs being acceptable with AM compared with 70 μs with FM.

Differing path lengths between the control centre to the transmitting sites can be equalized by installing audio delay circuits at the transmitters. The path lengths from each transmitting site to all the places where equal signals occur must then be less than about 21 km (equivalent to 70 μs) for acceptable quality. Modern techniques enable delays to be dynamically equalized to compensate for variations in the path.

Figure 15.6 shows the layout of a multi-station scheme including the audio delays. However, signals do not confine themselves to neat

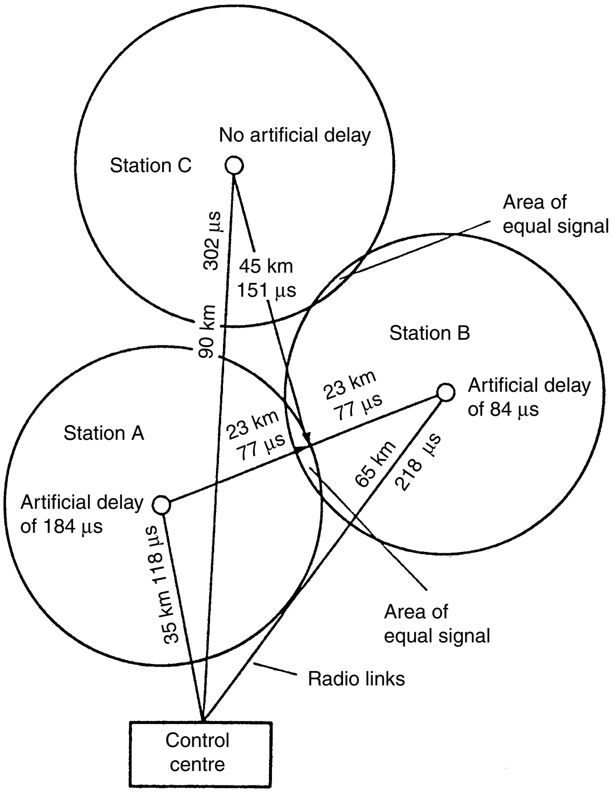
No artificial delay
Station C
302 μs
90 km
45 km
151 μs
Area of
equal signal
Station B
Station A
23 km
77 μs
23 km
77 μs
Artificial delay
of 84 μs
Artificial delay
of 184 μs
65 km
218 μs
35 km 118 μs
Area of
equal signal
Radio links
Control
centre

Figure 15.6 Quasi-synchronous transmitter system

circles and the worst situation is where the signals from two transmitters arrive at a mobile more or less equal in strength and phase. In this situation, apart from the ripple caused by the carrier frequency offset of the transmitters, a satisfactory signal would still be received but, occasionally, the presence of a weaker signal from a third, distant station (C in *Figure 15.6*) intrudes during the cancellation periods and, because of its long path length, introduces severe distortion. The area where this situation occurs is often small and the areas of overlap can usually be moved slightly by adjustment of either the transmitter power or antenna directivity.

16 Signalling

Many radio receivers are required to respond to instructions sent to them over the radio channel to which they are tuned. Several methods of signalling these instructions are in use, from the continuous tones used for controlling the receiver mute operation to the rapid data to initiate channel changes and transmitter power control on cellular radio-telephones.

16.1 Sub-audio signalling

The slowest signalling system uses continuous sub-audio frequency tones. Known as continuous tone controlled signalling (previously squelch) system (CTCSS) its performance, in the UK, is specified in Radiocommunications Agency specification MPT 1306. The most common use for the system is to control receiver mute opening. Permitting a mute to open only on receipt of an authorized signal, its use enables privacy between users to be maintained on shared systems, common base station systems for example, and reduces the annoyance factor from interference in the absence of a signal. Thirty-two tones are permitted and assigned by the Radiocommunications Agency (*Table 16.1*).

Table 16.1 CTCSS tones and modulation levels

67.0	110.9	146.2	192.8
71.9	114.8	151.4	203.5
77.0	118.8	156.7	210.7
82.5	123.0	162.2	218.1
88.5	127.3	167.9	225.7
94.8	131.8	173.8	233.6
103.5	136.5	179.9	241.8
107.2	141.3	186.2	250.3

The tones are transmitted at very low modulation levels (*Table 16.2*).

Table 16.2 Modulation levels

System channel spacing	*Amplitude modulation Modulation depth (%)*	*Angle modulation Peak deviation ± Hz*
25 kHz	10 to 20	400 to 800
12.5	10 to 20	200 to 400

16.2 In-band tone and digital signalling

16.2.1 Dual tone multi-frequency (DTMF)

Developed for telephone work, dual tone multi-frequency signalling has crept into some radio systems but is slow to operate compared with other in-band systems. Two tones are transmitted simultaneously to represent each digit. In the UK they conform to the MF4 code (*Table 16.3*).

Table 16.3 MF4 code of DTMF tones

Digit	*Frequencies (Hz)*	
1	697	1209
2	697	1336
3	697	1477
4	770	1209
5	770	1336
6	770	1477
7	852	1209
8	852	1336
9	852	1477
0	941	1336
*	941	1209
#	941	1477

Some push button telephones signal to the local exchange by generating a combination of two frequencies. Each row and each column of the push button keypad is connected to an oscillator of set frequency. When any push button is pressed, tones corresponding to its row and column frequencies are therefore generated. The row and column oscillator frequencies are as shown below:

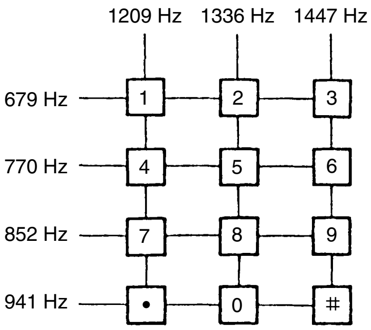

16.2.2 5-tone signalling

So-called 5-tone signalling uses a series of tones – usually five but other quantities are used in some systems. The system is popular for selective calling and addressing purposes (*Table 16.4*). There are a range of tones established by various organizations but, although the tones in each series may be assigned to each digit, equipment manufacturers sometimes amend the code and the usage. The relevant Radiocommunications Agency document is MPT 1316, *Code of Practice, Selective Signalling for use in the Private Mobile Radio Services*.

Table 16.4 5-tone codes

Standard	*ZVEI*	*DZVEI*	*EIA*	*CCIR/ EEA*	*CCITT*	*EURO*
Tone duration (ms)	70	70	33	100/ 40	100	100
Pause duration (ms)	0	0	0	0	0	0
Code/digit	Tone frequency (Hz)					
0	2400	970	600	1981	400	980
1	1060	1060	741	1124	697	903
2	1160	1160	882	1197	770	833
3	1270	1270	1023	1275	852	767
4	1400	1400	1164	1358	941	707
5	1530	1530	1305	1446	1209	652
6	1670	1670	1446	1540	1335	601
7	1830	1830	1587	1640	1477	554
8	2000	2000	1728	1747	1633	511
9	2200	2200	1869	1860	1800	471
Repeat	2600	2400	459	2110	2300	1063
Alarm	2800	2600		2400		
Free-tone						1153
Group tone	970	825	2151	1055		

ZVEI	Zuverein der Electronisches Industrie. Designed to operate on 20 kHz channel spaced systems. On 12.5 kHz channel spaced systems transmission of 2800 Hz creates difficulty and depressed ZVEI (DZVEI) was adopted and recommended in MPT 1316.
EIA	Electrical Industries Association.
CCIR	Committee Consultatif International Radio Communication. The longer tone duration offers robustness against corruption but is slow to operate. Originally designed for marine use, each digit was transmitted twice to ensure reliability.
EEA	Electronic Engineering Association. Recommended in MPT 1316.
CCITT	International Telegraph and Telephone Consultative Committee.

16.3 Digital signalling

FFSK

Many analogue modulated systems use digital signalling in the form of 1200 bit/s fast frequency shift keying which is faster than 5-tone but less robust.

DCS

Digitally coded squelch (DCS) is a digital, but slower, alternative to CTCSS. It is not in current use in the UK. DCS continuously repeats the constant binary bit pattern of a 23 bit word. Golay error coding reduces the number of data bits to 12 of which 3 initiate the Golay sequence leaving 9 information bits, i.e. a 9 bit word. With a 9 bit word, $2^9 = 512$ codes are possible but, because of problems with the frequencies produced, the codes with 7 or more 1s or 0s in succession are discarded leaving 104 ‘clean codes’. The transmitted rate is 134 bit/s and the highest fundamental frequency is therefore 67 Hz. The lowest fundamental frequency is produced when 6 zeros follow 6 ones = $1/6$th of the fundamental frequency = 11.7 Hz. The important harmonics are:

$$\begin{aligned} \text{third} &= 201\,\text{Hz} \\ \text{fifth} &= 335\,\text{Hz} \\ \text{seventh} &= 469\,\text{Hz} \end{aligned}$$

The seventh harmonic is well within the audio pass band of most receivers but at 3 octaves from the fundamental is relatively weak.

16.4 Standard PSTN tones

	Signal			
	UK		*USA*	
Tone	*Frequency*	*Sequence*	*Frequency*	*Sequence*
Number unobtainable	400 Hz	Continuous	500 + 600 Hz	Interrupted every second
Number busy	400 Hz	0.75 s on, 0.75 s off	480 + 620 Hz	Interrupted once per second
System busy	400 Hz	0.4 s on, 0.35 s off, 0.225 s on, 0.525 s off	480 + 620 Hz	Interrupted twice per second
Dial tone	50 Hz 33.3 Hz		350 + 440 Hz	
Ringing tone	400 Hz, or 400 + 450 Hz, modulated by 17, 25, or 50 Hz	0.4 s on, 0.2 s off, 0.4 s on, 2 s off	440 + 480 Hz	Interrupted once per second
Pay tone	400 Hz	0.125 s on, 0.125 s off		
Time tones	900 Hz	0.15 s on, 0.85 s off. Three times every three minutes		
Recorder warning			1400 Hz	One burst, every 15 s

References

Belcher, R. *et al*. (1989). *Newnes Mobile Radio Servicing Handbook*. Butterworth-Heinemann, Oxford.

Winder, S.W. (2001). *Newnes Telecommunications Pocket Book*. Butterworth-Heinemann, Oxford.

17 Channel occupancy, availability and trunking

17.1 Channel occupancy and availability

The usefulness of any communications network depends on successful calls. Assuming that no calls fail for technical reasons, the success rate depends on the amount of traffic on the channel and how well it is managed. Channel occupancy is a measure of the amount of traffic carried and is affected by:

1. the number of subscribers (mobiles on a radio network)
2. the number of calls per subscriber in a given time
3. the average duration of each message.

The occupancy of any system varies hourly, daily and possibly seasonally. To be successful the network must be able to cope with the traffic during the periods of highest occupancy and, when considering occupancy it is usual to examine the conditions over one hour, the 'busy hour'. The occupancy of a channel can be expressed as a percentage:

$$\text{Channel occupancy} = \frac{nct}{36}\%$$

where

n = number of subscribers (on a mobile radio system, mobiles on watch)
c = call rate (number of calls per subscriber in 1 hour)
t = average message duration in seconds

Occupancy may also be expressed in erlangs where continuous traffic, i.e. 100% occupancy over a specified period, is 1 erlang. Less than 100% occupancy is expressed as a decimal fraction, e.g. 30% occupancy is 0.3 erlang. Demand may also be expressed in traffic units, A:

$$A = \frac{Ct}{T} \text{ erlangs}$$

where C is the number of calls in time T and t is the mean call duration.

The amount of congestion occurring on a system can be defined by the grade of service it offers. Grade of service is expressed as a percentage in terms of the probability of meeting either congestion or waiting more than a specified delay before a call can be established. Erlang's formula is:

$$\text{Grade of service} = \frac{A^N/N!}{1 + A + A^2/2! + A^3/3! + \cdots + A/N!}$$

where A is the number of traffic units (erlangs) and N is the number of outlets (trunked channels).

The probability that delayed calls will have to wait more than a specified time ($W > t$) is:

$$P(W > t) = \exp\left(-\frac{(C - A)t}{H}\right)$$

where
A = traffic units (erlangs)
C = number of trunked channels
H = mean call duration in busy hour

The above assumes that:

1. the number of users is infinite
2. the intervals between calls are random
3. call durations are random
4. call set-up times are negligible
5. delayed calls are queued so that they are dealt with on a first come, first served basis.

The number of potential radio users is continuously growing and any technique that makes more effective use of the finite radio spectrum is welcome. Whenever technology has permitted, higher frequencies and reduced channel spacings have been employed in the attempt to meet the demand. A method of making more effective use of the available spectrum is trunking.

17.2 Trunking

The generally accepted term for the automatic allocation of communication channels to subscribers on demand is trunking, and a 'trunked radio system' is one in which many users share the use of a common pool of radio channels. Channels from the pool are allocated to users on demand and as they become unoccupied; no channels are allocated to specific users or groups of users.

When making a telephone call, provided the speech quality is good it is immaterial which route the call takes. The same applies to radio; providing the message reaches the right person quickly, the radio frequency which carries it does not matter.

Most trunked radio systems use a calling channel on which calls are requested. When a call is established and accepted a working channel from the pool is allocated for the duration of the call and at its termination the channel is returned to the pool for re-allocation to another user. The important principle is that any user has access to any free channel. This has the advantage that more mobiles per channel can be accommodated, or a better grade of service, i.e. less waiting time for access, can be provided. Basic trunked systems are unsuitable where everyone using the system needs to be a party to every call but the latest trunked mobile radio system arrange, on receipt of a call, to assign a working channel to a particular group of mobiles where they operate within the service area of one base station. They then operate on an 'all informed' basis.

Trunking provides the most benefit when the number of channels required is greater than the number assigned to the system. The effective gain in the number of mobiles that can be accommodated on a trunked system is illustrated by the following example:

> *A single radio channel with an efficient signalling system offering a 30% grade of service can accommodate 90 mobiles with a mean waiting time of 20 seconds if each mobile makes 1 call. Twenty non-trunked channels cater for 1800 mobiles. A 20 channel trunked system under the same conditions can accommodate 3430 mobiles with a mean waiting time of 16.4 seconds, a trunking gain of over 170 mobiles per channel.*

The mobile transmitter/receivers used on trunked systems are more complex than those used on systems where channels are permanently allocated because the mobiles must either scan the channels searching for an available one or one containing a call directed to them, or listen continuously on a channel designated for calling. In the latter case when a call is initiated, a complex 'handshake' procedure occurs and the mobile is automatically switched to the channel allocated for that conversation. The band III public access mobile radio (PAMR) networks use a calling channel and many common base station (CBS) systems use channel scanning methods with no calling channel.

Working channels may be assigned for the duration of conversation, which is message trunking, or for the duration of a transmission, which is transmission trunking. With either type of trunking the channel is not returned to the pool immediately the signal from the mobile is lost; some waiting time is essential to cater for short-term signal dropouts but with message trunking the channel is retained for a longer period of time to allow the complete conversation to take place. On a very busy system this is wasteful and, on a system with very rapid signalling, transmission trunking may be used. Then the use of the working channel is lost very soon after the cessation of transmission and a new working channel assigned on its resumption. Some networks automatically migrate from message trunking to transmission as traffic increases.

17.3 In-band interrupted scan (IBIS) trunking

IBIS trunking is a method which is finding favour for common base station applications. All channels are used for communication with no dedicated calling channel, thus making the most effective use of the channels available.

On a CBS system all users access the base station, which operates talk-through, by radio. Each user is allocated his own continuously coded squelch system (CTCSS) tone to access the base station and to prevent overhearing by other users. A trunked common base station contains the transmitters and receivers for a number of radio channels and when the system is quiescent transmits bursts of an in-band (speech band) free-tone – 1953 Hz is one that is used – on one channel only. The mobiles, which are fitted with decoders to recognize both the free-tone and their CTCSS tone, scan all channels looking for one containing free-tone onto which they lock. Should the mobile operator press the transmit button the transmitter will only be enabled when the mobile receiver is on a channel containing either the free-tone or its CTCSS code. As soon as the channel becomes occupied the base station commences to radiate free-tone on another channel and mobiles of other user groups re-lock onto the new free channel. When either a mobile or the control station, which is a fixed mobile, makes a call it is on the channel radiating free-tone and the transmission contains the CTCSS code allocated to that user group. The repeater re-transmits the CTCSS code which activates the calling mobile and all other mobiles within that group. Mobiles of other user groups sensing the alien CTCSS tone leave the channel and continue scanning for a free channel. If all channels are engaged, the caller hears a ‘busy’ tone and must try again later.

17.4 Trunking to MPT 1327 specification

Specification MPT 1327 produced by the Radiocommunications Agency lays down a signalling standard for trunked private land mobile radio systems. The system uses fast frequency shift keying (FFSK) at 1200 bit/s with time division random access. The mobiles operate two-frequency semi-duplex and the base stations or trunking system controller (TSC) operate full duplex. A calling channel is used for call establishment but may be either dedicated or nondedicated. After a call is established messages are transmitted over a traffic (working) channel. The basic components of the signalling format, in sequence, are:

1. The link establishment time (LET), a period of transmission of undefined modulation.
2. A preamble comprising a sequence of bit reversals 1010. . .10, minimum 16 bits and ending with a binary zero for the receiver data demodulator to achieve synchronization.
3. A message. A contiguous transmission of a synchronization sequence, address codeword and, where appropriate, data codewords.
4. A hangover bit, H, of either binary zero or binary one, added to the last transmitted message identifying the end of signalling transmission.

The message synchronization sequence for both control and message channels consists of 16 bits, the first bit being a binary one on a control channel and binary zero on a traffic channel.

Messages, whether address or data, are transmitted in 64 bit codewords. The first bit, bit 1, indicates the status of the codeword. Binary one denotes an address word, binary zero, a message word. Bits 2–48 are the information field and bits 49–64 are check bits.

References

Macario, R.C.V. (1991). *Personal and Mobile Radio Systems*. Peter Peregrinus, London.

Parsons, J.D. and Gardiner, J.G. (1989). *Mobile Communication Systems*. Blackie, London and Glasgow.

Radiocommunications Agency (1991). MPT 1327, *A Signalling Standard for Trunked Private Land Mobile Radio Systems*.

Radiocommunications Agency (1993). MPT 1318, *Engineering Memorandum, Trunked Systems in the Land Mobile Radio Service*.

Winder, S.W. (2001). *Newnes Telecommunications Pocket Book*. Butterworth-Heinemann, Oxford.

18 Mobile radio systems

18.1 Paging

Where an immediate reply to a message is not essential, paging systems, which are basically a one-way path from controller to mobile, are an economic option. Paging is also spectrally economic, only one frequency being required and messages being of very short duration. The recipient of a call may be merely alerted by means of a tone emitted by the paging receiver or, if silent alert is required, a vibration. Alternatively, pagers displaying, and storing, alpha-numeric messages are readily available. Message handling services are also offered on some paging systems.

On-site paging systems are usually owned or rented by the user and consist of either a radio base station or, in the case of a non-radio system, an inductive loop around the perimeter of the service area. The details of on-site paging systems and the Radiocommunications Agency specifications governing them are:

RF frequencies:

on-site, 26 MHz and 31 MHz bands, MPT 1365
49.0–49.5 MHz, MPT 1335
160–161 MHz
458–459 MHz, MPT 1305

Channel spacings:

12.5 kHz in the VHF bands and 25 kHz in UHF band

Modulation:

FM, FFSK data

Wide area, national or international systems are generally owned by a network provider to whom the user subscribes for the service. There are frequencies allocated in the high VHF and UHF bands. The appropriate specifications are MPT 1308 for receivers and MPT 1325 for transmitters.

The Post Office Code Standardization Group (POCSAG) standard which has been adopted by the CCITT allows the transmission of messages at a rate of 10 calls per second and a capacity of 1 million pagers with up to four alternative addresses per pager. The recommended bit speed is 512 bit/s with direct FSK deviating the transmitter frequency by ±4.5 kHz. Positive deviation indicates a binary zero and negative deviation a binary one.

In America the Golay sequential code is widely used for paging addressing. The format for tone-only paging consists of 12 address bits followed by 11 parity bits for each 23 bit word. Two bit rates are used, 300 bit/s for address codewords and 600 bit/s for message codewords.

There are European paging and message systems in operation including Euromessage and the frequency agile European Radio Message Service (ERMES).

18.2 Cordless telephones

CT1

The first generation of cordless telephone, still very popular in the UK, is designed to serve the domestic environment with a range of 100 m. The base station transmits on one of eight frequencies between 1.6 MHz and 1.8 MHz, and the handset on one of eight frequencies in the 47 MHz band. Frequency modulation is used with a deviation of ±4 kHz at the base station and 2.5 kHz at the handset. MPT 1322 applies and permits an erp of 10 mW at both the base station and the handset.

CT2

The second generation of cordless telephone. The services, under various names, telepoint, phonezone, etc. were planned to provide for the general public a lower cost alternative to the cellular radiotelephone networks which were seen at the time as a businessman's preserve. However, the low market uptake caused their demise in the UK although services are thriving in some European and Far Eastern countries.

One hand-portable transmitter/receiver can operate to a local base station installed in the home or office, or through one of a number of multi-channel base stations with a range of approximately 200 m installed in public places. While away from the local base station the subscriber must initiate a call: calls cannot be made from the PSTN to a CT2 subscriber's handset.

The operational frequencies are in the band 864.1 to 868.1 MHz and employ time division multiple access. Speech is digitized at 32 kbit/s, stored, and then transmitted at 64 kbit/s in 1 ms slots. This leaves the alternate 1 ms slots available for the digitized and stored speech of a reply. Duplex operation is achieved in this way on a single radio frequency.

Inter-operability between networks was to be ensured by specification MPT 1375 *Common Air Interface*.

Digital European cordless telephone (DECT)

A pan-European system complying with an ETSI standard, DECT operates in the 1880 to 1900 MHz band. It offers data handling facilities and the ability for a subscriber to receive calls while away from the local base station. The techniques are similar to those used for GSM although, because the mobile is virtually stationary, the constraints on data transmission are less severe and no hand-off is required. The 20 MHz RF bandwidth is divided into 13 carriers spaced at 1.7 MHz intervals, each carrier containing 12 TDMA channels with GMSK modulation.

CT3 is developed from DECT and operates in the band 800 to 900 MHz. Each 8 MHz section of that band is divided into 1 MHz blocks, each containing sixteen 1 ms time slots.

18.3 Trunked radio

Bank III trunked radio

So-called because it occupies part of the now redundant television band III, 174 to 225 MHz, and often referred to as private mobile radio, this is a subscriber access system. The network has almost nationwide coverage in the UK, arranged on both local and regional bases. The network permits radio communication between a fixed office station and mobiles, mobile to mobile, and limited access from mobiles to the PSTN.

The relevant Radiocommunications Agency specifications are:

MPT 1323	*Angle-modulated Radio Equipment for Use at Base and Mobile Stations Private Mobile Radio Service Operating in the Frequency Band 174–225 MHz.*
MPT 1327	*A Signalling Standard for Trunked Private Land Mobile Radio Systems.*
MPT 1343	*System Interface Specification for Radio Units to be Used with Commercial Trunked Networks Operating in Band III sub-band 2.*
MPT 1347	*Radio Interference Specification for Commercial Trunked Networks Operating in Band III sub-band 2.*

MPT 1352 *Test Schedule. For Approval of Radio Units to be Used with Commercial Trunked Networks Operating in Band III sub-band 2.*

Frequency modulation is used with FFSK signalling, a channel spacing of 12.5 kHz and TDMA techniques (Section 17.4).

Trans-European trunked radio (TETRA)

A new European standard for digital trunking. The system is aimed at PMR users and will employ TDMA with four users sharing a 25 kHz radio channel with the option of two users on a 12.5 kHz channel. Speech will be digitally encoded at 4.8 kbit/s and transmitted at 36 kbit/s (gross bit rate including error correction overhead) using $\pi/4$ differential quaternary phase shift keying (DQPSK).

Digital short range radio (DSRR)

This utilizes trunking techniques but with no infra-structure of control or switching centres; mobiles work to mobiles via a locally installed repeater. It offers great potential for short-term hire for local events. The band 933 to 935 MHz has been allocated to the service and accommodates 76 traffic and 2 control channels separated by 25 kHz. The operating procedure may be either single-frequency simplex or two-frequency semi-duplex. Mobiles remain on standby waiting to receive their selective signalling code (SSC) on one of the two control channels; mobiles with even serial numbers normally listen on control channel 26 and those with odd serial numbers on control channel 27. A mobile wishing to initiate a call first scans for a free traffic channel and, having obtained one, selectively addresses the intended recipients of the call over the relevant control channel. The recipient mobiles are then instructed automatically by the system to go to the appropriate traffic channel.

The modulation is GMSK at 4 kbit/s for addresses and 16 kbit/s for speech and data messages. The maximum transmitter erp is 4 W.

18.4 Analogue cellular radio-telephone networks

Analogue networks are becoming obsolete and are being replaced by digital cellular networks. There are no analogue networks operating in the UK and the frequency bands that they used to occupy are now reassigned to GSM operators.

However some countries still have an analogue service based on AMPS or TACS. These provide national coverage with roaming facility and hand-off from cell to cell as a mobile travels across the network, calling for more complex signalling than local trunking.

The control system must know at all times the location of all operational mobiles. It does this by continuously monitoring signal strength and instructing the mobile to change channel as it crosses cell boundaries.

To permit the maximum re-use of frequencies mobiles are instructed by the system controller to reduce transmitted power to the minimum necessary for acceptable communications in either eight or five steps, depending on the mobile classification, to 0.01 W. Mobiles are classified by their maximum effective radiated power output:

Class 1	10 W erp, vehicle mounted
Class 2	4 W erp, transportable
Class 3	1.6 W erp, transportable
Class 4	0.6 W erp, hand-portable

The channel separation is 25 kHz with a peak speech deviation of ±9.5 kHz. Signalling is by Manchester encoded PSK at 8 kbit/s with 6.4 kHz deviation.

To discriminate between stations transmitting on the same RF channel an FM supervisory audio tone (SAT) is transmitted concurrently with speech at ±1.7 kHz deviation. There are three such tones: 5970, 6000 and 6030 Hz.

An FM signalling tone (ST) of 8 kHz transmitted for 1.8 seconds (deviation = ±6.4 kHz) indicates the clear-down of a conventional telephone call.

The advanced mobile phone service (AMPS) is the American cellular system designed for 30 kHz channel spacing with a 10 kbit/s signalling rate. TACS is based on the AMPS system but modified for reduced channel spacing. Narrower channel spacing systems (NAMPS and NTACS) are also found in some countries.

18.5 Global system mobile

Global system mobile (GSM), formerly known as Groupe Speciale Mobile, is the pan-European digital cellular radio-telephone service. The operational requirements for GSM are severe, e.g. it must operate satisfactorily to a person walking, or in a slowly moving vehicle, in a street where much of the furnishings will introduce multi-path fading, and operate to a train travelling at 250 km/hr where Doppler frequency shift becomes significant. To reduce the corruption, a high degree of error detection and correction must be applied which increases the occupied bandwidth. To compensate, to some extent, for this, the system takes full advantage of the redundancy in speech to reduce the bandwidth during synthesis.

GSM operates full duplex in the band 890–915 MHz, up-link and 935–960 MHz, down-link.

A combination of FDMA and TDMA is employed. Each allocated band of 25 MHz is divided into 125 carriers spaced 200 kHz apart. The subdivision of each transmitted bit stream into 8 TDM time slots of 540 μs gives 8 channels per carrier and 1000 channels overall. Modulation is Gaussian minimum shift keying (GMSK).

Speech synthesis is by speech codec using linear predictive coding. It produces toll quality speech at 13 kbit/s.

18.5.1 Emergency locator service

The FCC in the US has placed a requirement on mobile phone operators to provide an emergency location service, referred to as E911. The aim is to locate people who need emergency services, but are not able to indicate their location either through injury or through lack of local knowledge. The locator service may also deter hoax calls (or help to catch people making hoax calls) and could have commercial applications in providing directions to businesses such as garages and restaurants.

There are two options: handset based solutions or network-based solutions. The handset-based solutions generally rely on signals picked up from the global positioning system (GPS). Network-based solutions rely on the timing difference in the arrival of signals from neighbouring cell sites. In network based solutions, the E911 requirement is for the mobile phone to be located within 100 metres at least 67% of the time and to within 300 metres in 95% of cases. In handset-based solutions, the location distances are 50 metres and 100 metres, respectively.

18.5.2 Data over GSM

Standard GSM data service allows a maximum of 9600 baud data rate transmission. An enhanced data rate of 14.4 kbps is also possible with some terminals.

The data link can be either synchronous or asynchronous and the data is conveyed through the network using an ISDN-based data protocol, the ITU-T standard V.110 rate adaptation scheme. Asynchronous transmission requires start and stop bits to be added, so the net data rate is reduced to 7680 bits per second (bps). Synchronous transmission potentially allows a faster data rate, but error checking is employed that requests re-transmission of data when errors are identified and this can produce a lower overall data throughput.

18.5.3 High-speed circuit switched data (HSCSD)

HSCSD allows higher data rates by combining up to four 14.4 kbps data channels together. A maximum data rate of 57.6 kbps is possible.

18.5.4 General packet radio service (GPRS)

GPRS combines up to eight 14.4 kbps channels to give data rates of up to 115.2 kbps. At these rates, Internet browsing is possible. GPRS supports X.25 and IP based transmission, which includes wireless application protocol (WAP) data.

18.6 Other digital mobile systems

Second generation mobile systems are digital systems. The GSM system has been described in Section 18.5. This section describes the GSM1800 system, the personal communications system (PCS1900, sometimes known as personal communications network or PCN) and third generation systems.

18.6.1 GSM1800 (DCS1800)

The GSM1800 system operates in the 1840–1880 MHz band, with a carrier spacing of 200 kHz. Each carrier gives multiple access through time division multiple access (TDMA). It is a lower power version of GSM900, which operates in the 900 MHz band. A GSM1800 handset has a maximum power output level of 1 watt, rather than the 2 watts of a GSM900 handset.

Voice coding can be either: half-rate using VSELP at 5.6 kbps, full-rate using RPE-LTP at 13 kbps, or enhanced full rate (EFR) using ACELP at 12.2 kbps. Convolutional coding is used at a rate of $\frac{1}{2}$ to reduce the effects of noise and errors on the radio channel. The radio carrier modulation is Gaussian minimum shift keying (GMSK).

18.6.2 PCS1900 (IS-136, or PCN)

The personal communications system (PCS) is based on GSM and used time division multiplexing to provide separate channels over each radio frequency. Convolutional coding is used at a rate of $\frac{1}{2}$ to reduce the effects of noise and errors on the radio channel. Each carrier is spaced 30 kHz apart and offset quadrature phase-shift keying (OQPSK) is used to modulate the carrier.

The voice-coder (vocoder) used by PCS1900 is either full-rate using VSELP at 7.95 kbps or enhanced full rate (EFR) that uses ACELP at 7.4 kbps. The EFR scheme was developed for use in PCS1900 and has since been adopted by both GSM900 and GSM1800 networks because it improves the audio quality. Previously, GSM used either a half-rate (HR) or full-rate (FR) vocoder.

18.6.3 Third generation systems

The so-called 3G systems have been described using a number of terms: UMTS, IMT-2000, cdma-2000, w-cdma, etc. There are variations between the proposed standards, but in general they all use code division multiple access (CDMA), convolutional coding, quadrature phase shift keying and time or frequency division duplexing.

Third generation systems will operate in the 1885–2025 MHz band and in the 2110–2200 MHz band. Initially the lower band between 1980 and 2010 MHz will not be allocated, nor will the upper band between 2170 and 2200 MHz. However, the full bands will be available when existing services in these bands can be moved (such as PCS1900, PHS and DECT).

The code division multiple access (CDMA) modulation scheme allows several channels to occupy the same frequency band. CDMA is a form of spread spectrum encoding. The digitized speech signals are coded using a unique code, known as a spreading code, so that one bit of speech code produces many encoded bits. When demodulated and mixed with the same spreading code, the original data is recovered. Receivers that use a different spreading code are unable to recover this data, but they are able to recover data intended for them.

18.7 Private mobile radio (PMR)

These systems are privately owned or rented and may be used only by the owner's own staff for the purpose of the owner's business. The leasing of air-time is forbidden. Common base station or community repeater systems where several users subscribe to, and use, the station is often also referred to as PMR.

18.8 UK CB radio

27 MHz band: 27.60125 to 27.99125 MHz

40 channels at 10 kHz spacing.
Max. erp, 2 W; max. transmitter output power, 4 W.
Antenna: single rod or wire, 1.5 m overall length, base loaded.
If mounted higher than 7 m transmitter output to be reduced at least 10 dB.
Modulation: FM only, deviation ±2.5 kHz max.

934 MHz band: 934.025 to 934.975 MHz

20 channels at 50 kHz spacing (may be reduced later to 25 kHz).
If synthesizer used spacing may be 25 kHz on precise channel frequencies specified.
Max. erp, 25 W; max. transmitter output power 8 W. If antenna integral, max. erp, 3 W.
Antenna: may have up to four elements, none exceeding 17 cm.
If mounted higher than 10 m, transmitter output to be reduced at least 10 dB.
Modulation: FM only, deviation ±5.0 kHz max.

Spurious emissions: for both bands, not exceeding 0.25 μW except for specified frequency bands where the limit is 50 nW.

For latest full specifications see publications MPT 1320 (27 MHz band) and MPT 1321 (934 MHz band) obtainable from the Information and Library Service of the Radiocommunications Agency.

References

Belcher, R. *et al.* (1989). *Newnes Mobile Radio Servicing Handbook*. Butterworth-Heinemann, Oxford.
Macario, R.C.V. (1991). *Personal and Mobile Radio Systems*. Peter Peregrinus, London.
Winder, S.W. (2001). *Newnes Telecommunications Pocket Book*. Butterworth-Heinemann, Oxford.

19 Base station site management

The information in this chapter applies mainly to sites for stations operating at VHF and above. Some conditions will be different for stations working at lower frequencies due to the generally physically larger transmitters and antennas.

19.1 Base station objectives

The objectives of a radio base station are:

1. To provide an adequate service over the whole of the desired service area in a well-defined manner. Special consideration must be given to those areas where, because of their location, weak signals or a high ambient electrical noise level might be expected.
2. To create the minimum interference with other radio users and be designed to receive the minimum from other users.
3. To be electromagnetically compatible with neighbouring installations, possibly on the same site.
4. To create the minimum impact on the environment.

19.2 Site ownership or accommodation rental?

The question of whether to privately develop a base station site or rent space on someone else's site may be resolved by the planning authorities rather than the developer. Before planning consent will be given the developer must normally prove that no existing site will meet the requirements. However, assuming that land is available and planning consent might be forthcoming, other factors to consider are costs, both initial and recurring, and access for services, materials during building and tower erection and for maintenance during bad weather (*Table 19.1*).

19.3 Choice of site

The first stage in selecting a site is to look for an existing site which might appear to cover the service area. Information may be available from the operators of such sites regarding their coverage, and coverage tests may be made or computer predictions obtained. For a

Table 19.1 Factors affecting the decision to own or rent

	Ownership	*Rental*
Initial costs	Legal charges Land clearance Provision of services Erection of buildings and mast or tower Installation of equipment Provision of standby power	Legal charges Connection fee Installation of equipment Provision of standby power
Recurring costs	Maintenance of site Power Insurance (other than radio equipment)	Rental Normally included in rental Normally included in rental
Other considerations	Very long commissioning time Site under user's control Space may be leased to other users creating revenue	Very short commissioning time Site under landlord's control Accommodation will be shared with other users (install equipment in locked cabinet to preserve security)

receiving site the ambient noise level at the base station must also be considered. It should be realized that a site, or an antenna, can be too high, particularly when coverage is required at very short range or if interference from other users is likely to occur. Antennas radiate the minimum signal along the line of their elements so that vertically polarized antennas, i.e. with the E plane vertical, provide very little, if any, signal immediately below them. This creates a hole in the coverage, 'the Polo mint effect', close to the antenna.

Where there is capacity on an existing site the possibilities are:

1. Sharing an existing lightly used radio channel, e.g. common base station.
2. Sharing equipment space in an existing building or providing own building or extension on a communal site.
3. Sharing an antenna by means of multi-coupling or erecting own antenna on existing structure.

19.4 Masts and towers

Apart from short, unguyed poles mounted on the side of buildings, antenna supports fall into two types: guyed masts and self-supporting towers.

A guyed mast may be a single pole, or a square or triangular section lattice structure. The range of single-pole masts includes those of about 4 m maximum height mounted on chimney stacks and secured by wire lashings, and poles supported on the ground or the roof of a building. The latter types may be up to 30 m high. They are usually constructed from steel or aluminium hollow tubular sections – the bottom section is invariably steel – with a set of guys at about every 5 m. Mast erection is specialist work but to estimate the approximate height that a site can accommodate, a four-guy, single-pole mast requires a minimum spacing for the anchorage points in the form of a square with sides equal to half the height of the mast. The guys may be of steel, possibly plastic-coated, or more likely of synthetic fibre. Steel guys may affect the directivity of antennas in close proximity, and if the guys corrode at their fixing points, intermodulation interference may result.

Self-supporting towers come in all heights up to many hundreds of feet and may be made of a variety of materials from steel sections to concrete.

The factors determining the type of structure to be erected are the number and types of antenna, and the site conditions.

Antenna considerations:

- Physical dimensions and space requirements.
- Weight.
- Wind loading.
- Directivity. For example, the beam width of a 6.5 GHz, 3.0 m dia. microwave dish is 0.9 degrees. A tower supporting such an antenna must twist, therefore, less than about 0.3 degrees in the highest wind likely to be experienced at the site.
- Access for riggers to antenna mounting positions.

Site conditions:

- Availability of space for tower footings, and guys and anchorages if a mast is contemplated.
- Stability and type of ground.
- Weather conditions. High winds. Accumulation of ice and snow. Build-up of ice increases wind loading of antenna – and of mast or tower – considerably. There is no point in using grid microwave antennas where icing is likely to occur.
- Aesthetic and planning consent considerations.

19.5 Installation of electronic equipment

For radio equipment, safety is one of the first priorities and in addition to the electrical considerations, physical hazards must be avoided.

Adequate working space must be provided around the equipment racks. Much equipment is now wall-mounted but tall, floor-standing cabinets must be firmly bolted down to prevent toppling when units are partially withdrawn for servicing and cables need to be routed safely. Overhead cable trays are generally considered the most satisfactory method but, to avoid interaction, it is recommended that cables for the various functions be segregated to reduce cross-talk and interference. Where cabling enters a cabinet at floor level, mounting the cabinet on a hollow plinth and running the cables through the plinth avoids damage to the cables and persons tripping over them.

For low voltage installations (see classifications below) the IEE wiring regulations apply. For high voltage installations, recommendations are given in BS 6701: part 1: 1990 which also covers accommodation, lighting levels and access arrangements.

Classification of installations:

- *Extra low voltage*. Normally not exceeding 50 V AC or 120 V DC.
- *Low voltage*. Exceeding 50 V AC or 120 V DC but not exceeding 1000 V AC or 1500 V DC between conductors, or 600 V AC or 900 V DC between conductors and earth.
- *High voltage*. Exceeding 1000 V AC or 1500 V DC between conductors, or 600 V AC or 900 V DC between conductors and earth.

19.6 Earthing and protection against lightning

Recommendations for earthing are given in BS 7430: 1991. This document covers the earthing of equipment and the principles of earthing for protection against lightning. All equipment metalwork must be bonded together and connected to the electricity supply earth point, the main earth terminal. In addition a connection must be made to the earthing system provided for protection against the effects of lightning.

Antenna systems by their nature are vulnerable to lightning strikes. Nearby taller structures may reduce the risk but precautions must still be taken. The zone of protection – a cone with its apex at the tip of the protecting structure and its base on the ground forming an angle of 45° to the perpendicular – does not necessarily protect structures above 20 m high. BS 6651: 1985 is the *Code of Practice for the Protection of Structures against Lightning*. Lightning protection begins at the top of the mast or tower where, ideally, the highest item should be a finial bonded to the tower. However, a finial mounted alongside an omnidirectional antenna will affect is radiation pattern. Grounded antennas are preferred and the outer conductor of each feeder cable must be bonded to the tower at the top and bottom of its vertical run. Grounding kits for the purpose are obtainable. The feeder cables must also be bonded to an earthing bar at their point of entry to the equipment building. The codes of practice should be consulted regarding the routes to be taken by earthing tape and the methods of jointing. An important point is that neither very sharp bends nor 'U' bends are created.

The legs of the tower must be bonded together with earthing tape and each leg connected to an earthing system which may consist of a buried plate or rods. Several rods may be necessary for each subsystem to attain the specified earth resistance of not more than 10 ohms. The rods should, except in rock, be driven into ground which is not made up or backfilled or likely to dry out. Where several rods are necessary to achieve the specified earth resistance these should be spaced well apart, the reduction in earth resistance being small with parallel rods closely spaced. Joints to facilitate testing of each subsystem separately must be provided either above ground or in a purpose-built pit or chamber. *Figure 19.1* shows a typical earthing system.

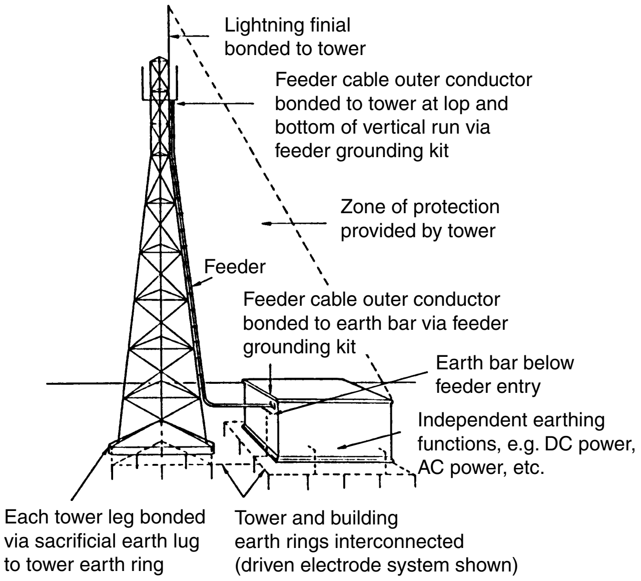
Lightning finial bonded to tower
Feeder cable outer conductor bonded to tower at lop and bottom of vertical run via feeder grounding kit
Zone of protection provided by tower
Feeder
Feeder cable outer conductor bonded to earth bar via feeder grounding kit
Earth bar below feeder entry
Independent earthing functions, e.g. DC power, AC power, etc.
Each tower leg bonded via sacrificial earth lug to tower earth ring
Tower and building earth rings interconnected (driven electrode system shown)

Figure 19.1 Typical example of good earthing practice (© Crown Copyright 1991, Radiocommunications Agency)

The system should be tested regularly, at least annually, and the results recorded in a lightning protection system log book. The test comprises:

1. Measurement of the resistance to earth of each termination network and each earth electrode.
2. Visual inspection of all conductors, bonds and joints or their measured electrical continuity.

BS 7430 recommends the method of testing.

Other vulnerable systems are overhead cables for either the electricity supply or telephone and control purposes.

19.7 Erection of antennas

The installation of the antenna probably influences the performance of a radio system more than any other part. The heights involved render supervision of the installation difficult and, although antennas are relatively inexpensive, labour charges raise the cost of their replacement, and bad weather may delay rigging work. It is therefore important that good quality antennas are specified and that the initial installation is correct. In addition to the physical mounting arrangements, the dispositions of antennas on the structure are important. Some aspects are detailed below and MPT 1331 contains more information.

19.7.1 Directivity

The radiation pattern diagrams provided in the antenna manufacturers' catalogues are produced by an isolated antenna. They can be greatly affected by the proximity of the support structure, e.g. if a dipole is mounted close to the leg of a tower, the leg behaves as a reflector and the radiation pattern of the antenna approaches that of two-element Yagi. The less directional the antenna, the worse the effect. Omni-directional antennas should not be mounted on the side of a mast or tower if it can be avoided. If an omni-directional antenna must be so mounted it should be spaced two wavelengths from it. This is difficult to achieve at the lower frequencies considering the antenna dimensions, weight and windloading. The rear element of a Yagi antenna must be one wavelength clear of the structure. To obtain a nominally omni-directional pattern a better, albeit expensive, solution is to mount a number of antennas, phased correctly, around the tower.

All antennas should point directly away from the structure, i.e. at right angles to the side of the tower.

The proximity of metalwork also adversely affects the VSWR of an antenna, causing standing waves with their additional voltage stresses on the feeder and possibly producing interference.

19.7.2 *Practical aspects of antenna installation*

All installation components must be made from compatible materials, electrically dissimilar metals, e.g. steel and brass, brass and aluminium, being avoided at all costs; they will corrode and cause intermodulation. Preferred materials are aluminium, and galvanized or stainless steel. Mounting brackets must be secure – remembering that antennas and feeders vibrate – but not overtightened so as to distort and weaken antenna support booms.

Connectors of the correct impedance, and preferably with a captive centre pin, are vital. Type N connectors are available at either 50 or 70 ohms impedance, are robust and can be used at frequencies up to 10 GHz. UHF 83 series connectors, although very robust, are of imprecise impedance and, from first-hand experience of interference aggravated by a UHF 83 used to connect an antenna feeder to a receiver tail, should not be used on base station installations. Assembly instructions for a range of connectors are shown in Chapter 23. All joints must be waterproofed, preferably by first wrapping with a self-amalgamating tape for two inches either side of the joint, then a layer of PVC tape covered by a further layer of Sylglass or Denso tape. When installing feeder cables the required length should be loosely coiled and taken up the tower and paid out as it is secured, working from top to bottom. It must not be dragged from a drum on the ground. Feeders must be cleated at the intervals specified by the manufacturer. Route cables where they are easily accessible but least likely to suffer physical damage.

Health and safety

Apart from the physical risks to riggers, there are radiation hazards. Large amounts of RF power may be radiated from antenna systems and research is continuing into its effects on health. The most recent recommendations of the National Radiological Protection Board and the Department of Trade and Industry should be sought and followed. Riggers should be aware of the power levels present on a structure before climbing, and equipment switched off where levels are considered unsafe.

19.7.3 *Antenna checking and fault finding*

The majority of VHF and UHF antennas present a short-circuit across the antenna and to ground as a means of reducing their vulnerability to static charges. An ohmmeter applied to the lower end of the feeder will indicate a circuit through the feeder and antenna but may also indicate a short-circuit or low resistance fault, water in the feeder for instance. Most antenna system faults increase the VSWR present on the feeder cable. A good antenna, operating within its design bandwidth, exhibits a VSWR of 1:1. When connected to a length of co-axial cable this may be raised and a VSWR of 1.5:1 for the system is generally considered acceptable. However, faults can be masked by conditions occurring on the feeder, and a measurement taken with a standing wave, or forward and reflected power meter at the bottom of the feeder, may not indicate the true condition of the antenna. A measurement taken with an accurate dummy resistive load connected to the top of the feeder will prove the cable, and a measurement of the VSWR should also be taken directly at the antenna. A record of the VSWR at installation and any subsequent measurements is helpful.

19.8 Interference

There are two sources of radio interference induced by antenna systems. One occurs from strong signals radiated by either the system's own transmitter antenna or from a co-sited system operating on a close neighbouring frequency. The other source is intermodulation, 'rusty bolt effect', the mixing of two or more signals to produce the interfering frequency.

19.8.1 Antenna isolation

One solution to the problem of direct radiation is to space the antennas so that there is sufficient isolation between them. Because minimum radiation is present immediately below and above the ends of the elements of vertically polarized antennas, maximum isolation occurs when such antennas are mounted in a vertical line. Maximum radiation, and hence minimum isolation, occurs when antennas are mounted side-by-side. The degree of isolation depends on the spacing but a figure of 40–45 dB between a transmitter and a receiver antenna, and 20–25 dB between transmitting antennas, should be the target. To achieve 25 dB isolation requires a vertical separation of 0.9 wavelengths between the centres of vertically polarized dipoles; 45 dB requires a vertical separation of 3 wavelengths. The spacing for horizontal separation of vertically polarized antennas is 2.5 wavelengths for 25 dB isolation and 25 wavelengths for 45 dB. To preserve this isolation the feeder cables should also be separated over their routes.

Antenna multi-coupling

A precise method of controlling the RF isolation between mutually sited systems where the frequencies are moderately spaced is for several systems to share one antenna by using multi-coupling techniques (see Section 19.9).

19.8.2 Intermodulation

Any two or more RF signals applied to a non-linear device intermodulate, that is, they combine to form additional frequencies. *Table 19.2* lists the combinations for two input frequencies. It must be remembered that the side frequencies produced by modulation of the original carriers will also be present.

Table 19.2 Low order intermodulation products

2nd	$A + B$	5th	$3A + 2B$	7th	$4A + 3B$
	$A - B$		$3B + 2A$		$4B + 3A$
			$3A - 2B^{*}$		$4A - 3B^{*}$
3rd	$2A + B$		$3B - 2A^{*}$		$4B - 3A^{*}$
	$2A - B^{*}$		$4B + A$		$5A + 2B$
	$2B + A$		$4A + B$		$5B + 2A$
	$2B - A^{*}$		$4A - B$		$5A - 2B$
			$4B - A$		$5B - 2A$
4th	$2A + 2B$				$6A + B$
	$2A - 2B$	6th	$5A + B$		$6B + A$
	$2B + 2A$		$5B + A$		$6A - B$
	$2B - 2A$		$5A - B$		$6B - A$
	$3A + B$		$5B - A$		
	$3A - B$		$4B + 2A$		
	$3B + A$		$4B + 2B$		
	$3B - A$		$4B - 2A$		
			$4B - 2B$		
			$3A + 3B$		
			$3A - 3B$		
			$3B - 3A$		

*indicates in-band products. No eighth-order products fall in-band but ninth-order in-band products are produced by:

$5A - 4B$

$5B - 4A$

19.8.3 Control of intermodulation

A clue to the source of interference will be obtained from examination of the relationship of all frequencies produced on site, and also those produced nearby. That the side frequencies produced by modulation must be considered was proved in an intermodulation situation where the interference was only received when a nearby band III television transmitter radiated a peak white signal.

Corroded joints in metalwork close to the site, receiver RF stages subjected to very large signals and transmitter output stages are all non-linear devices and can create intermodulation.

Regular maintenance of the metalwork around the site reduces the chances of intermodulation from this source, but it is difficult to see and eliminate all corrosion on an inspection. Should interference from this cause be suspected after elimination of the other sources, a successful method of locating the offending joint requires a receiver tuned to the interfering signal and fitted with a whip antenna. While receiving the interference, a very sharp null in signal strength occurs when the tip of the whip points directly at the source of the signal.

Receiver RF amplifiers are usually designed to cope with small signals. When a strong signal radiated from a nearby transmitter is received the stage may well be overloaded and driven into non-linearity, or even into a blocked state where the receiver is effectively dead and may take some seconds to recover after the signal is removed. In the non-linear state the receiver is in an ideal condition to create intermodulation. If this situation is suspected a method of identifying it is to install a variable RF attenuator in the antenna feed to the suspect receiver. With zero attenuation the interference will be received but, if the receiver is the cause, increasing attenuation will produce little reduction of the interference until a point is reached where the interference suddenly disappears – the receiver at this point has re-entered a linear state. If the interference reduces gradually to zero with increased attenuation, the source is elsewhere.

Most amplifiers are designed to be linear, that is, the output signal level will follow that of the input signal. However, with a sufficiently high input signal overloading occurs, the amplifier becomes non-linear and compression of the signal results (*Figure 19.2*). The point where 1 dB of compression occurs is a commonly referred to amplifier parameter. If the input level is increased further the gain of the amplifier is reduced until saturation is reached when the output level can no longer increase.

At the onset of non-linearity harmonics and intermodulation, if any other frequencies are present, are produced. The strength of these rises rapidly with an increase of input, and the point where an extension of the almost linear portion of their curve crosses an extension of the linear gain line of the amplifier is the third-order intercept – usually it is only the third-order products which are considered for this purpose.

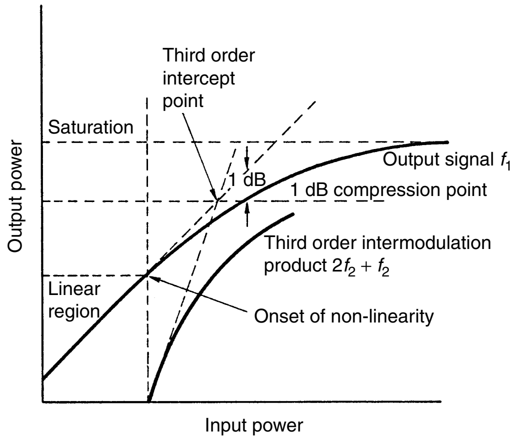

Figure 19.2 Production of intermodulation

The class C operated output stages of transmitters are by design non-linear. Strong signals from a neighbouring transmitter or antenna applied via the antenna feed will mix with the transmitted frequency to create intermodulation products. Increased isolation between antennas and feeders may be the simplest remedy. Alternatively, a circulator or filters may be connected to the output of the transmitter,

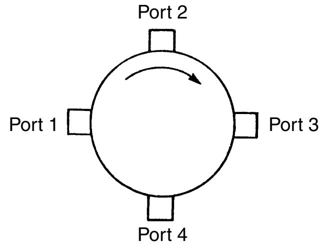
Port 2
Port 1
Port 3
Port 4

(a) RF circulator

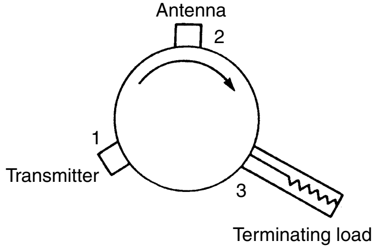
Antenna
2
1
Transmitter
3
Terminating load

(b) RF isolator

Figure 19.3 RF circulator and isolator

and the possibility of direct radiation from transmitters must not be ignored.

A circulator is a uni-directional device with either three or four ports as shown in *Figure 19.3(a)*. A signal entering at port 1 will leave at port 2 but not at ports 3 or 4. Similarly, a signal entering at port 2 will leave at port 3 but not ports 1 or 4. Circulators are used for combining the outputs of transmitters for application to a single antenna but if a three-port circulator is connected as in *Figure 19.3(b)* it forms an isolator. Typical isolation over a 1.5% bandwidth is 20–30 dB with an insertion loss of 0.7 dB.

19.9 Antenna multi-coupling

The sharing of antennas and feeders between a number of co-sited services is not only a precise method of controlling frequency isolation, it is an economic solution to the problems of antenna and tower management. An antenna may be shared between a number of receivers or transmitters, or shared simultaneously by both receivers and transmitters. Receiver sharing requires a splitter, i.e. a filter, to separate the frequencies for each receiver and an amplifier to make up the filter losses. Transmitters can share an antenna through circulator/isolators and filters. To obtain the required selectivity, stability

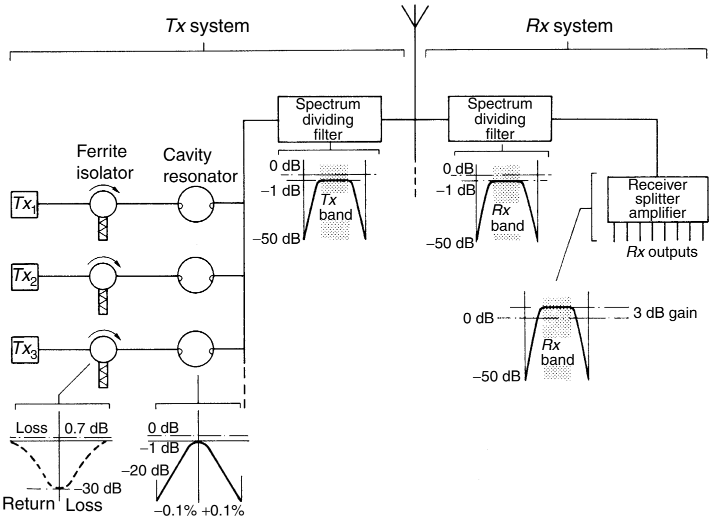

Tx system
Rx system
Spectrum dividing filter
Spectrum dividing filter
Ferrite isolator
Cavity resonator
*Tx*1
*Tx*2
*Tx*3
0 dB
−1 dB
−50 dB
Tx band
0 dB
−1 dB
−50 dB
Rx band
Receiver splitter amplifier
Rx outputs
0 dB
3 dB gain
−50 dB
Rx band
Loss
0.7 dB
−30 dB
Return Loss
0 dB
−1 dB
−20 dB
−0.1% +0.1%

Figure 19.4 Typical sub-band transmitter/receiver sharing (© Crown Copyright 1991, Radiocommunications Agency)

and power handling capability the filters are often solidly constructed cavity resonators with an insertion loss of 1 dB and a bandwidth of 0.1% to the −20 dB points. *Figure 19.4* shows the layout of a transmitter/receiver sharing system.

19.10 Emergency power supplies

The installation of equipment to provide power for all the equipment on a site during a failure of the public supplies has declined in recent years. But although lengthy power failures seem to be a thing of the past, they could recur, and electricity supplies do still fail for short periods. To provide a complete supply for large station necessitates a standby generator. On smaller stations batteries may be capable of supplying the demands of the electronic equipment and emergency lighting for one or two hours. The trend on multi-user shared sites appears to be for the individual user to provide the emergency power, derived from batteries, for his own equipment. Factors to consider are:

1. Standby motor generators are expensive. They need housing, ventilation and fuel storage. They take time to start up, 45 seconds being typical.
2. Uninterruptible power supplies (UPS). Where no interruptions can be tolerated the most efficient supply is provided by batteries alone but the equipment must be designed to operate directly from a DC, low voltage supply.

   To supply power at 220/240 V AC an inverter is the accepted method. Double inverters permanently charge a battery from the AC supply and the battery drives an oscillator to reproduce the AC supply voltage, a very inefficient process. Also, should any item of the inverter break down the supply is lost. This is overcome in some inverters by a switch which transfers the load to the public supply when a failure of the inverter occurs. Single inverters charge a battery and supply the equipment directly from the public supply until a failure occurs. When this happens the supply changes over to the battery and oscillator. Single inverters are more efficient but there is a short break in supply while the change-over occurs. Adequate stabilization and filtering to avoid disturbances on the supply must be provided. One possible hazard with all inverter supplies is that where more than one inverter is used, the outputs of the inverters are unlikely to be in phase and a 480 V difference could exist between the supplies for two items of equipment.

For all battery-operated supplies correct charging of the battery is crucial. Nickel–cadmium batteries require constant-current charging with a fall to the trickle rate when fully charged and sophisticated circuitry is needed to identify the precise point of full charge (see Section 21.3.2).

Lead acid batteries produce hydrogen if over-charged so their storage must be designed so that no danger, either to personnel or of explosion, exists. For reliability, natural ventilation is preferred to forced. Recommendations for the accommodation of lead acid batteries other than sealed are given in BS 6133: 1985. Batteries which gas also need more frequent maintenance. Because of the reduction in performance of batteries at low temperatures the ambient temperature of the battery storage room should not fall below 4°C, and to reduce water evaporation lead acid batteries should not be run continuously above 40°C.

The period of time over which a lead acid battery must be fully recharged determines the capacity of the battery more than the period of time over which it must supply power. Because of the need to reduce the charge rate – to avoid gassing – to a very low level as the battery approaches full charge, the last 15% of capacity takes a very long time to acquire (see Section 21.3).

19.11 Approval and certification

No base radio station may be operated in the UK prior to inspection of the installation by the Radio Investigation Service of the Radiocommunications Agency and the issue of their certificate of approval. The installation may not subsequently be modified without re-approval and certification.

References

Belcher R. *et al.* (1989). *Newnes Mobile Radio Servicing Handbook.* Butterworth-Heinemann, Oxford.

BS 6651: 1985. *Code of Practice for the Protection of Structures against Lightning.*

BS 6701: Part 1:1990. *Code of Practice for Installation of Apparatus Intended for Connection to Certain Telecommunications Systems.*

BS 7430: 1991. *Code of Practice for Earthing. Earthing of Telecommunications Installations.* International Telecommunications Union.

MPT 1331. *Code of Practice For Radio Site Engineering.*

MPT 1351. *Code of Practice For Repeater Operation at Communal Sites.*

MPT 1367. *Guidance Note to Legal Requirements Covering the Installation and use of Radio Apparatus and other Apparatus Generating Radio Frequency Emissions.*

MPT 1368. *Code of Practice for the Inspection of a Land Mobile Radio System for Conformity with the Wireless Telegraphy Act Licence and Performance Specification.*

Note: MPT documents are available from the Radiocommunications Agency (UK).

20 Instrumentation

20.1 Accuracy, resolution and stability

All measurements are subject to error and two instruments applied to the same piece of equipment under test may give a different answer. Tolerances must therefore be accepted. The errors arise from the following sources:

1. Human error, e.g. precision in reading a scale, use of incorrect instrument or range setting for the purpose.
2. The accuracy to which the instrument is able to display the result of a measurement or, in the case of a generator, the frequency or output level.
3. Accuracy of calibration.
4. Tolerances in the components used in the construction of the instrument. Variations in the load applied to the instrument.
5. Variations caused by long-term drift in the values of components.
6. Variations due to temperature and supply voltage fluctuations, and the warm-up time required by some instruments.
7. An effect on the circuit under test by the connection of the instrument.

There is an important difference between the accuracy and the resolution of an instrument. The accuracy is a statement of the maximum errors which may occur from the causes in statements 3 to 6 above. In instrument specifications, stability defined by statements 5 and 6 is usually quoted separately.

The accuracy of analogue measuring instruments is normally quoted as a percentage of full scale deflection (FSD). This is the accuracy of the instrument movement and components plus the scale calibration. The scale graduations, though, may not permit the user to determine the reading to the accuracy of the instrument perhaps because the graduations are cramped or parallax reading error occurs. These factors decide the resolution or precision of reading.

The accuracy of instruments with digital displays is usually quoted as a percentage of the reading plus or minus one count or one digit. While digital instruments are generally more accurate than their analogue counterparts, the fact that the least significant figure may be in error affects the resolution. *Figure 20.1* shows how this can arise. Most digital instruments use a gating process to switch the input to

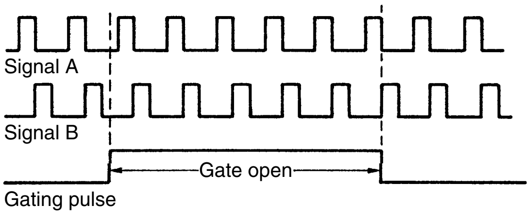

Figure 20.1 Gating error introduced by signal phasing

the measuring circuitry for the appropriate period of time. The gating time itself may vary and affect the accuracy, but even with a perfectly accurate and stable gating time, the phase of the input signal at the time of switching affects the number of pulses passing through the gate and thus the resolution.

Performance figures taken from a number of manufacturer's catalogues are listed below. They show only the more important features of the specifications and are typical of those for high quality instruments used in radio work.

20.2 Audio instruments

20.2.1 Output power meters

Those for radio applications measure RMS audio power and are usually calibrated in watts and dBW or dBm. They usually contain a dummy load resistor of adjustable impedance as a substitute for the receiver loudspeaker when in use.

20.2.2 Distortion factor and Sinad meters

These may be separate instruments or combined, and may also be incorporated within an audio output power meter. Both operate on the same principle. An audio input amplifier, with automatic gain control to give a fixed reference level at its output, applies the output from the unit under test, first to a notch filter to remove the test frequency of 1000 Hz, then to an AC voltmeter. The voltmeter is calibrated in percentage distortion or dB Sinad.

In testing a receiver Sinad ratio, the instrument is connected across the loudspeaker or equivalent dummy load. An RF signal generator connected to the antenna socket is adjusted to apply a moderately high level signal modulated with a 1 kHz tone – which must be pure and match the notch filter in the Sinad meter – at 60% of the system peak modulation. To be correct, the receiver volume control is adjusted to give the rated audio output but if only Sinad sensitivity is being measured a lower level is acceptable. The Sinad meter will now display the combined level of all the frequencies except 1000 Hz present in the receiver's output. These are the noise and distortion products. With the high level input applied these will be low and the meter deflection will be small indicating a high Sinad ratio. The RF input level is now reduced until the meter reads 12 dB (the standard reference) and the RF level, µV or dBm, is the Sinad sensitivity for the receiver.

The oscillator to produce the test tone may be included in the meter and filters may be provided to comply with various weightings.

20.2.3 Audio signal generators

The instrument provides an output for synchronizing to an external standard and a TTL compatible output. A high power version capable of delivering 3 watts into a 3 Ω load in the audio band is available.

20.3 Radio frequency instruments

20.3.1 RF power meters

Direct reading RF power meters either contain a non-reactive load or use an external load and may be calibrated in watts or dBm. RF calorimeters convert the RF energy into heat and measure the temperature of the heated element. At low powers, 'dry' calorimeters are used but their long thermal time constant inhibits their use at high power levels. To measure high powers 'flow' calorimeters, where a fluid flows around a closed system and the output temperature of the fluid is measured, are used. Power can be determined from:

$$P = F(T_{\text{out}} - T_{\text{in}})c(T)$$

where

P = power

F = mass flow rate of the fluid

$c(T)$ = the fluid's specific heat

T_{in} = temperature of the fluid entering the load

T_{out} = temperature of the fluid after being heated by the load

Thruline type instruments require an external load which may be either an antenna system or a load resistor. This type of meter reads forward and reflected power and, in some instruments, VSWR directly.

20.3.2 RF signal generators

The range of instruments designed for use on specific systems, e.g. cellular and digitally modulated, is so wide that manufacturers' catalogues must be consulted for each application. In addition to the accuracy of the carrier frequency and output level, RF leakage, spectral purity and modulation noise levels are important. The output level may be calibrated in μV, dBV or dBm, and may refer to either an unterminated instrument (p.d.) or terminated in a load equal to the output impedance of the generator (e.m.f.). If the instrument is calibrated in p.d., the output voltage must be halved when the instrument is terminated in an equal impedance.

The 50 ohm output impedance of many signal generators matches the input impedance of most VHF and UHF receivers directly. To simulate the impedance of antennas at HF and below a dummy antenna is usually inserted between the generator and the receiver under test. *Figure 20.2* shows the circuit of a standard dummy antenna.

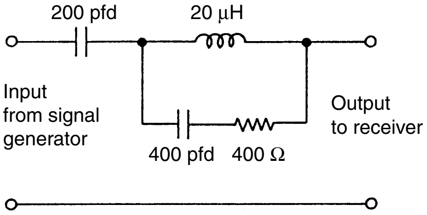

Figure 20.2 Standard dummy antenna

20.3.3 Frequency counters

Instruments are necessary to measure not only the precise frequencies of transmitters and receiver local oscillators but also those of low frequency signalling tones. Consequently they have a wide operating range.

The output of a transmitter must never be connected directly into a counter. The signal should either be obtained off-air or through a 'sniffer', appropriate for the frequency range, which siphons off a fragment of the transmitter output. It is also essential to ensure that the frequency displayed is the fundamental and not a harmonic or some other spurious frequency.

20.3.4 Modulation meters

The remarks in the above section relating to the connection of a transmitter and the selection of the fundamental frequency apply equally to modulation meters.

20.3.5 Spectrum analysers

Today's tight regulation of spectrum usage makes a spectrum analyser an essential tool for the precise alignment of radio equipment and for the measurement of modulation products and the noise content of signals.

A spectrum analyser is essentially a superheterodyne receiver with an adjustable IF bandwidth which sweeps across a portion of the frequency spectrum in synchronism with the horizontal trace of a cathode ray tube display. A signal at any frequency within the swept band will, while it remains within the IF passband, appear as a vertical displacement of the display trace. The design of the filter is crucial but must be a compromise. An analyser using multiple filters permits fast measurement but low resolution; a single filter offers high resolution but slower response time. Also, the shape factor of a wide filter inhibits the display of low level signals close to the centre frequency of the filter and a narrow filter, while permitting the display of these signals, may fail to display transients. The bandwidth determines the sweep speed, narrower bandwidths requiring slower speeds.

The possibility exists of spurious signals, e.g. intermodulation, being produced within the analyser with large input signals. A method of reducing the possibility is to insert a circulator and a notch filter

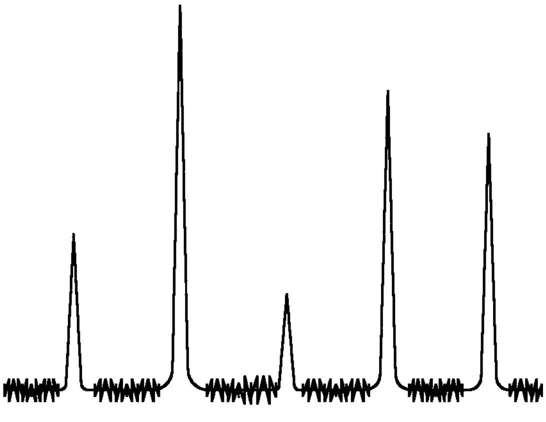

Figure 20.3 Amplitude-vs-frequency spectrum analyser display

between the unit under test and the analyser input to reduce the fundamental frequency amplitude. The circulator must be installed between the filter and the unit under test to absorb the reverse power produced by the impedance variations of the filter.

Many instruments are programmable and suitable for incorporation in automatic testing (ATE) systems, and instruments with similar features to those described are also offered in a combination test set form.

20.3.6 Network analysers

A network analysers examines *incident, reflected* and *transmitted* signals through a circuit or device, and displays the magnitude and phase of these signals. A spectrum analyser, on the other hand, measures only one channel, and displays magnitude and frequency.

Figure 20.4 shows an optical analogy to the network analyser. A 'device' of different optical density than ambient is in the path of an incident ray of light (R). When the light hits the surface, part of it is reflected (A) and part is transmitted through the 'devices' (even though refracted a bit).

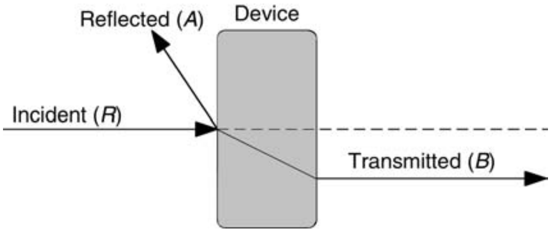

Figure 20.4 Optical analogy to network analyser

A network analyser consists of a three-channel RF receiver and a display. The incident signal is considered the *reference signal*, so is designated the R-channel. The other two channels receive the reflected signal on the A-channel, and the transmitted signal on the B-channel.

The uses of scalar and vector network analysers differ from the uses of spectrum analysers. The spectrum analyser measures external signals of unknown frequency and modulation type. Even when a tracking generator is added, to allow the spectrum analyser to perform stimulus-response tests, the spectrum analyser cannot do the job of the network analyser. The network analyser, by contrast, contains a known signal source, and is capable of sweeping a range of frequencies and power output levels. It can also perform ratio measurements.

References

Belcher, R. *et al*. (1989). *Newnes Mobile Radio Servicing Handbook*. Butterworth-Heinemann, Oxford.

Bird Electronic Corp., Catalog GC-92.

Farnell Instruments Ltd, Catalogue, 1993.

Marconi Instruments Ltd, Test and Measurement Instrument Systems 1993/94.

RTT Systems Ltd, Catalogue, 1993.

Winder, S.W. (2001). *Newnes Telecommunications Pocket Book*. Butterworth-Heinemann, Oxford.

21 Batteries

21.1 Cell characteristics

Batteries are composed of cells which exhibit characteristics peculiar to their chemical constituents and construction. Common features, important to users, are their ability to store energy within a small space and with the least weight, and to release it at an adequate rate for the purpose under consideration.

21.1.1 Capacity

The amount of energy a battery can store is measured in ampere hours (Ah) at a specified discharge rate. For large cells this is usually the 10 hour rate but American practice, which is now almost universal, at least for smaller cells, is to use the 1 hour rate. It is important, therefore, to be certain which rate is referred to. The capacity reduces as the rate of discharge increases. Thus a battery of 60 ampere hours capacity at the 10 hour rate will provide 6 amps for ten hours before reaching the point at which it is considered to be discharged. If a current of 12 amps is taken, the battery will become discharged in less than 5 hours, and if the current is 3 amps, it will last longer than 20 hours. The rate of discharge is often referred to in terms of the C-rate which may be expressed in several ways. 1C, C or C_1 are numerically the same as the rated capacity, e.g. a 500 mA Nicad cell supplying 500 mA, which may be expressed as C, 1C or C_1 continuously will be discharged in approximately 1 hour. If the cell supplies current at a reduced rate, 0.5 C, 0.5 C_1, or C/2, i.e. 250 mA, it will last approximately 125 minutes. The subscript, e.g. C_5, indicates the hourly discharge rate.

The terminal voltage at which a cell is considered discharged also varies with the discharge rate; a lead acid cell discharged at 1C is considered to be discharged when the terminal voltage falls to 1.75 V. At C_{10} the cell is considered to be discharged at 1.85 V.

The capacity of a battery or cell may also be specified at a given ambient temperature, usually 20°C. Lower temperatures reduce the effective capacity and maximum current off-take, higher temperatures increase them slightly.

For radio use battery duration may be quoted in terms of standby, receive and talk time. The duty cycle obviously varies from user to user but a useful standard for a radio-telephone is 90% standby, 5% receive and 5% transmit. On an open channel PMR system, 80%/15%/5% is more typical. Measuring the current drains during these activities enables the battery requirements to be calculated.

21.1.2 Internal resistance

The maximum instantaneous current which a battery can deliver is determined by its internal resistance. In this respect a battery behaves like any other generator (see Section 21.3.2), where increasing load currents produce an increasing voltage drop across the internal resistance. The internal resistance of a battery is seldom specified, but for a battery in good condition it is extremely low (one quoted figure is 15 mΩ for a fully charged 500 mAh Nicad cell) and the voltage drop in the connecting leads will govern the maximum withdrawable current. Equally, the resistance of the meter used will affect the measurement of charge or discharge currents. If using an analogue meter, the older low resistance types are preferable (I keep a model 40 Avometer, 0.03 Ω int. res. on 12 amp range, for the purpose). Low current-rating fuses also present a resistance higher than that of a battery (4 Ω measured for a 250 mA fuse, 0.5 Ω for a 1 A) and, probably, the connecting leads.

21.1.3 Power:weight and volume ratios

It is the battery that now limits the size to which radio equipment can be reduced. Recent developments have increased the power to weight and power to volume ratios which are possible. Typical ratios are referred to in the sections dealing with individual battery types.

21.1.4 Recharging conditions

The initial charge rate is the current flowing through a discharged battery to replace the charge in a specified time. Unless supplied from a constant current charger the current will fall as the battery voltage rises, but as full charge is approached the charge rate is usually reduced to a trickle or finishing charge rate.

Trickle charging maintains the cells in a fully charged condition by passing a very small current through them sufficient merely to replace any self-discharge losses through leakage.

Finishing charge is a rate to which the charging current is reduced when a battery reaches about 85% of its full capacity. It is a rate at which gassing is unlikely to occur.

Float charging maintains the cell voltage at its nominal while it is supplying continuous and variable loads.

21.2 Non-rechargeable, primary batteries

While rechargeable batteries are the obvious choice for use in equipment which is in use continuously, disposable types have economic and logistic advantages for some applications – the batteries are more expensive long term but no charger is needed. Also primary batteries have up to 4 times the available capacity of their nickel cadmium equivalents with less weight. Details of disposable batteries with some standard nickel-cadmium rechargeable equivalents are given in *Table 21.1*.

Lithium primary batteries are now available and have the following qualities:

1. A high cell voltage of 3.6 V on load. The voltage is constant – after an initial fall when load is first applied – until discharged when a rapid fall occurs.
2. An operating temperature range of −55°C to +75°C is possible.
3. Energy densities up to 630 mWh/g. A standard AA size cell has a power/weight ratio of 340 mWh/g.
4. Long shelf life. Ten years at room temperature for a 10% fall in capacity is envisaged.

They have a disadvantage in that fire, explosion or severe burns may result if the batteries are mistreated. They must not be recharged, disassembled, heated above 100°C, incinerated or the contents exposed to water.

One use of lithium batteries is as a power source for memories but protection against charging is needed. *Figure 21.1* shows a typical circuit.

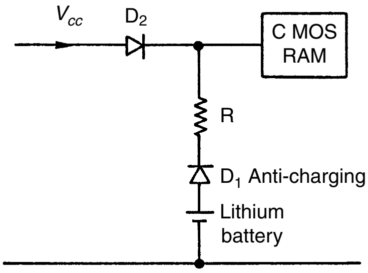

Figure 21.1 Typical lithium battery application

Table 21.1 Cells and batteries

					Maximum dimensions (mm)					
	USA size	Nominal voltage (V)	Type‡	IEC equivalent	Length (or diameter)	Width	Height	Contacts	Current (mA)	Weight (g)
Zine carbon	N	1.5	D23	R1	12	–	00.1	Cap and base	1–5	7
	AAA		HP 16	RO3	10.5	–	45	Cap and base	0–10 000	8.5
	AA		HP7	R6	14.5	–	50.5	Cap and base	0–75	16.5
	AA		C7	R6	14.5	–	50.5	Cap and base	0–75	16.5
	C		SP11	R14	26.2	–	50	Cap and base	20–60	45
	C		HP11	R14	26.2	–	50	Cap and base	0–1000	45
	C		C11	R14	26.2	–	50	Cap and base	0–5	45
	D		SP2	R20	34.2	–	61.8	Cap and base	25–100	90
	D		HP2	R20	34.2	–	61.8	Cap and base	0–2000	90
		4.5	AD28	3R25	101.6	34.9	105	Socket	30–300	453.6
			1289	3R12	62	22	67	Flat springs	0–300	113
		6.0	PP8	4-F100-4	65.1	51.5	200.8	Press studs	20–151	1100
			PJ996	4-R25	67	67	102	Spiral springs	30–300	581
			991		135.7	72.2	125.4	Two screws	30–500	1470
		9.0	PP3-P	6-F22	26.5	17.5	48.5	Press studs	0–50	39
			PP3-C	6-F22	26.5	17.5	48.5	Press studs	0–50	39

(*continued overleaf*)

Table 21.1 (*continued*)

	USA size	Nominal voltage (V)	Type‡	IEC equivalent	Maximum dimensions (mm)			Contacts	Current (mA)	Weight (g)
					Length (or diameter)	Width	Height			
			PP3	6-F22	26.5	17.5	48.5	Press studs	0–10	38
			PP4	6-F20	26.5	–	50	Press studs	0–10	51
			PP6	6-F50-2	36	34.5	70	Press studs	2.5–15	142
			PP7	6-F90	46	45	61.9	Press studs	5–20	198
			PP9	6-F100	66	52	81	Press studs	5–50	425
			PP10	6-F100-3	66	52	225	Socket	15–150	1250
		15.0	B154	10-F15	16	15	35	End contacts	0.1–0.5	14.2
			B121	10-F20	27	16	37	End contacts	0.1–0.1	21
		22.5	B155	15-F15	16	15	51	End contacts	0.1–0.5	20
			B122	15-F20	27	16	51	End contacts	0.1–1.0	32
Manganese alkaline	ED	1.5	MN1300*	LR20	34.2	–	61.5	Cap and base	10.00†	125
	C		MN1400*	LR14	26.2	–	50	Cap and base	5.50†	65
	AA		MN1500*	LR6	14.5	–	50.5	Cap and base	1.80†	23
	AAA		MN2400*	LR03	10.5	–	44.5	Cap and base	0.80†	13
	N		MN9100*	LR1	12	–	30.2	Cap and base	0.65†	9.6
Mercuric oxide		1.35/1.4	RM675H	NR07	11.6	–	5.4	Cap and base (button)	0.21†	2.6
			RM625N	MR9	15.6	–	6.2	Cap and base (button)	0.25†	4.3

			RM575H	NR08	11.6	–	3.5	Cap and base (button)	0.12[†]	1.4
			RM1H	NR50	16.4	–	16.8	Cap and base (button)	1.00[†]	12.0
Silver oxide		1.5	10L14	5R44	11.56	–	5.33	Cap and base (button)	0.13[†]	2.2
			10L124	5R43	11.56	–	4.19	Cap and base (button)	0.13[†]	1.7
			10L123	5R48	7.75	–	5.33	Cap and base (button)	0.08[†]	1.0
			10L125	5R41	7.75	–	3.58	Cap and base (button)	0.04[†]	0.8
Nickel cadmium		1.25	NC828	–				Button	0.28[†]	16.5
	AA		NCC50	–	See HP7			Button	0.60[†]	30.0
	C		NCC200	–	See HP11			Button	2.00[†]	78.8
	D		NCC400	–	See HP2			Button	4.00[†]	170.0
		10.0	NC828/8	–				Button stack	0.28[†]	125.0
		12.0	10/2250K	–				Button stack	0.225[†]	135.0
		9.0	TR7/8	(DEAC)	See PP3			Press studs	0.07[†]	45.0
	AA	1.25	501RS	(DEAC)	See HP7			Press studs	0.50[†]	30.0
	C		RS1.8	(DEAC)	See HP11			Press studs	1.80[†]	65.0
	D		RS4	(DEAC)	See HP2			Press studs	4.00[†]	150.0

[†] Capacity in ampere hours.

[‡] BEREC types unless otherwise indicated.

[*] Also Duracell (Mallory).

21.3 Rechargeable batteries

21.3.1 Lead acid batteries

Lead acid batteries, whether of free electrolyte or low-maintenance sealed construction, have a nominal terminal voltage on load of 2.0 V per cell. This voltage falls on load in a gradual curve, shown in *Figure 21.2*, until the discharged voltage of between 1.75 and 1.85 volts per cell is reached.

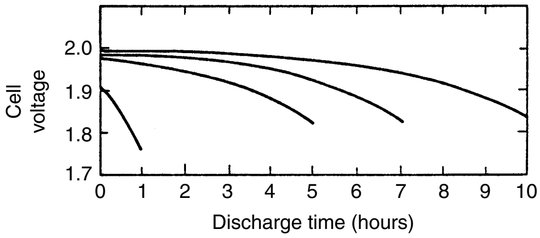

Figure 21.2 Typical lead acid cell discharge characteristics

A discharged battery will, because of its inefficiencies, require a recharge equal to the amperes × hours discharged +11%, e.g. a cell discharged at 5 amps for 10 hours will require a recharge of 55.5 ampere hours with a constant current at the 10 hour rate. Because of the reducing current as full charge is approached a recharge time of 1.4 to 1.5 times the capacity to be restored is more practical. The final on-charge cell voltage can increase to approximately 2.7 volts. Gassing occurs and hydrogen is liberated when the cell voltage reaches 23 V but provided the charging current is sufficiently low above this point gassing will be avoided. This lower charge rate, the 'finishing rate' can be applied by maintaining the charging voltage at about 2.4 volts when the battery will automatically limit the charging current. The specific gravity of the electrolyte in a fully charged cell is between 1.205 and 1.215.

The trickle charge current must be low enough to avoid gassing. A current of 7% of the 10 hour capacity is typical.

Float charging should maintain a cell voltage of approximately 2.2 volts.

The power:weight ratio of lead acid batteries is poor, a small (4 Ah), 6.0 V, sealed lead acid battery having a power:weight ratio of 26 mWh/g.

21.3.2 Nickel-cadmium (Nicad) batteries

At present the Nicad is probably the most commonly used rechargeable battery for portable applications. A standard size AA cell has a power:weight ratio of 27 mWh/g. The on-load Nicad cell voltage, after an initial fall from the on-charge voltage of between 1.3 and 1.4 V, remains substantially constant at about 1.2 V until the discharged voltage of 0.9 to 1.1 V is reached. Thereafter the voltage falls rapidly. This is illustrated in *Figure 21.3*. While the constant voltage is ideal during discharge it poses a problem in that it is difficult to reliably measure the intermediate state of charge which created difficulties with recharging. A Nicad battery which is repeatedly partially discharged and then recharged may, after many cycles, behave as though it were fully discharged when the repeated recharge condition is reached (the memory effect) and, with a fixed time charger, the possibility of over-charging is present. One solution fully discharged all batteries after use to a predetermined level, typically 1.1 V per cell, and then recharged them at a constant current for a fixed period of time. Unfortunately, this procedure shortened the life of the batteries; *Figure 21.4* shows the life expectancy of a cell with repeated discharges. The present solution is to charge the batteries automatically to the fully charged state and then reduce the current to the trickle charge level. Batteries which are subjected to repeated partial discharge may then be occasionally fully discharged to obviate the memory effect.

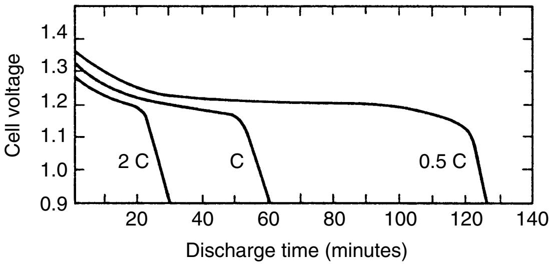

Figure 21.3 Typical Nicad cell discharge characteristics

Constant current, automatic charging is recommended. Chargers vary in complexity, some detecting the end-of-charge point by sensing a variation of voltage. At end of charge the cell voltage first rises and then falls slightly as in *Figure 21.5*. More sophisticated chargers also sense the cell case temperature which rises during charge. Batteries are available for standard charging at the ten hour rate where 14 to

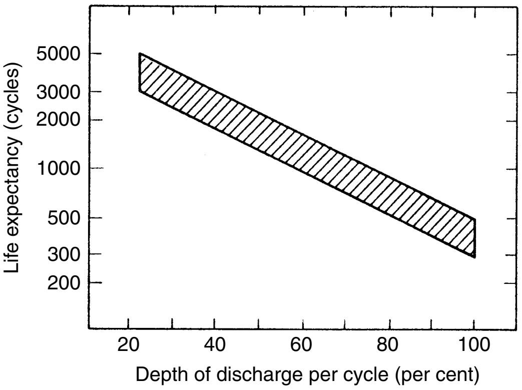

Figure 21.4 Nicad cell: effect of repeated discharge vs. cell life

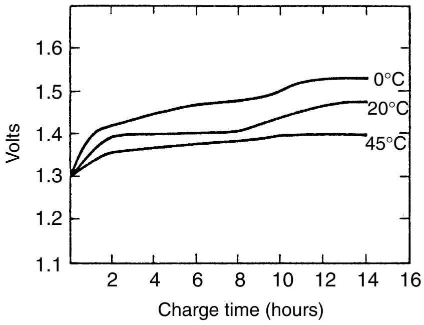

(a) Standard charge, C/10 × 15 hrs

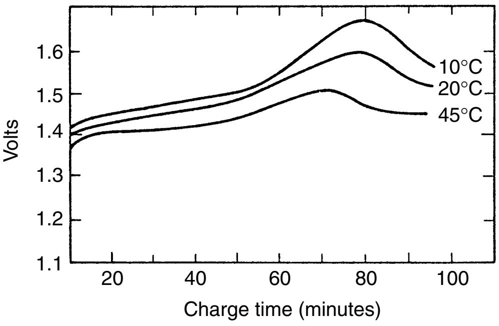

(b) Rapid charge, 1 C × 90 mins

Figure 21.5 Nicad cell: discharging characteristics

15 hours will be required to recharge a fully discharged battery, fast charging at 5 C and rapid charging at C rate.

Cells may be fitted with a re-sealing one-way vent which opens at about 200 psi and closes at about 175 psi to relieve any excess internal pressure caused by a fault or abuse.

22 Satellite communications

22.1 Earth orbits

Communications satellites are required to illuminate the earth with radio signals and their orbits are chosen according to the size and location of the part of the earth's surface they must light up.

A satellite orbiting the earth is continuously pulled by a centripetal force, in this case gravity, towards the centre of the earth. It is also pulled by centrifugal force to leave its orbit at a tangent. When these opposing forces are equal in magnitude the satellite is in a stable orbit. There is, then, for a given height (the radius of the path minus the radius of the earth, 6378 km), a velocity at which the conditions for stable orbit apply, and which determines the orbiting time.

22.1.1 Geostationary orbits

Satellites relay information from ground stations, either fixed or mobile, or between satellites. It is an advantage for some purposes, therefore, to use satellites with an orbit time identical to that of the earth so that no tracking from the ground stations is needed. Geostationary satellites have the same angular velocity as the earth making them appear to be stationary. Their height is 35 788 km and four such satellites cover the earth from latitude 81.3°N to 81.3°S as in *Figure 22.1*.

The disadvantages of geostationary satellites are that they are in a high earth orbit (HEO) resulting in a signal delay of 240 ms for the complete go and return path. Also they are in an equatorial orbit so that signals to the higher latitudes travel at a shallow angle to the earth's surface rendering them unsuitable for mobile use where communications must be achieved at street level in cities.

As each satellite covers a large portion of the earth, the design of their antennas to permit repeated frequency re-use is important and antennas with small, steerable footprints have been developed for this purpose. An advantage, in addition to the lack of tracking, is that the shadowing by the earth is minimal so that solar power cells receive almost continuous illumination.

22.1.2 Elliptical orbits

A polar orbit, where the satellite follows a North/South track, provides the opportunity to survey the earth in a series of strips. The satellites

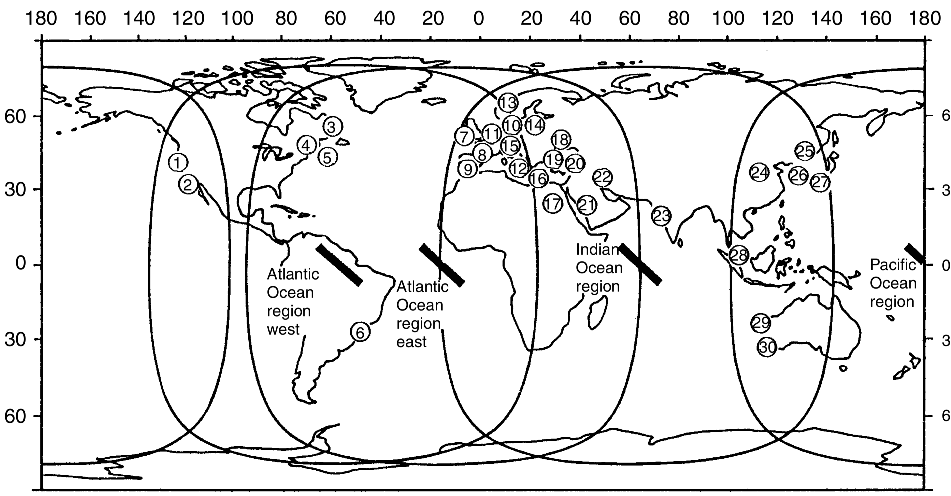
180 160 140 120 100 80 60 40 20 0 20 40 60 80 100 120 140 160 180
60
30
0
30
60
Atlantic Ocean region west
Atlantic Ocean region east
Indian Ocean region
Pacific Ocean region
①
②
③
④
⑤
⑥
⑦
⑧
⑨
⑩
⑪
⑫
⑬
⑭
⑮
⑯
⑰
⑱
⑲
⑳
㉑
㉒
㉓
㉔
㉕
㉖
㉗
㉘
㉙
㉚

Land earth station services	Inmarsat-A	Inmarsat-C	Inmarsat-Aero
1. Niles Canyon	●		●
2. Santa Paula	●	●	●
3. Laurentides			●
4. Southbury	●	●	●
5. Staten Island	●		
6. Tangua	●		
7. Goonhilly	●	●	●
8. Pleumeur Bodou	●	●	
9. Ausseguel			●
10. Blavand		●	
11. Burum	●	●	
12. Fucino	●		
13. Elk	●	●	●
14. Psary	●		
15. Raisting	●	●	
16. Thermopylae	●		
17. Maadi	●		
18. Odessa	●		
19. Anatolia	●		
20. Ata	●		
21. Jeddah	●		
22. Boumehen	●		
23. Arvi	●		
24. Beijing	●		
25. Nakhodka	●		
26. Kumsan	●		
27. Yamaguchi	●		●
28. Singapore	●	●	●
29. Perth	●	●	●
30. Ganagara	●		

Figure 22.1 Immarsat coverage and GES locations

used for this purpose are in low each orbit (LEO) and consequently have a high velocity. Their orbit time is approximately 1.5 hours and between successive orbits the earth has rotated 22.5°. Sixteen orbits are therefore needed to scan the earth's surface. Until recently polar orbits were used only for optical surveillance but now several projects for radio communication are either in the early stages of installation or under development. These use a number of satellites so that for mobile communications, for instance, there is a satellite continuously in view. Tracking of these extremely fast satellites by the ground stations, which may themselves be moving, along with Doppler effect has been a major obstacle which is now being overcome.

22.2 Communications by satellite link

Satellites are radio links and receive signals from the ground and other satellites which they must re-transmit. The signals from ground stations are comparatively weak and require high power amplification for onward transmission. As the satellite's receive and transmit antennas must be close together, the possibility of RF instability prohibits on-frequency repetition. The up-link, from ground station to satellite, must therefore be converted to another before re-transmission. The up-link frequency is normally higher than the down-link and the frequency converter is referred to as a 'down-converter'.

The frequency bands used for communications purposes are listed in *Table 22.1*.

Satellite television

The broadcasting of television may be via either communications satellites or Direct Broadcasting by Satellite (DBS) satellites. The positions of non-DBS satellites relative to the UK are shown in *Figure 22.2* and *Table 22.2* lists the European channels. *Figure 22.3* shows the world allocations of DBS satellites and *Tables 22.3* and *22.4* list the channel frequencies and national allocations. The frequency plan for the Astra satellite is in *Table 22.5*.

22.3 Proposed satellite television formats

Most current European satellite television programmes (non-DBS) are broadcast as fairly standard PAL signals, FM modulated into the satellite channel. DBS transmissions, and those from Astra satellite, will probably be of a multiplexed analogue component (MAC) format. In MAC, data corresponding to sound tracks and subtitles, etc., an

Table 22.1 Communications satellite frequencies

Frequency band (GHz)	*Link*	*European telecom links*	*International telecom links*
1.5–1.6	Down	Mobile	
1.6–1.7	Up	Mobile	
3.4–4.2	Down	Fixed	Fixed
4.5–4.8	Down	Fixed	Fixed
5.9–7.0	Up	Fixed	Fixed
10.7–11.7	Down	Fixed (+ Non-DBS television)	
11.7–12.5	Down	Fixed (+ television)	
12.5–12.75	Down	Fixed (private links)	
12.75–13.75	Up	Fixed (private links)	
14.0–14.8	Up	Fixed	
17.3–18.3	Up	Fixed	
17.7–20.2	Down	Fixed	
20.2–21.2	Down	Mobile	
27.0–30.0	Up	Fixed	
30.0–31.0	Up	Mobile	
22.5–23.55		Allocated for intersatellite links	
32.0–33.0		Allocated for intersatellite links	
54.25–58.2		Allocated for intersatellite links	
59.0–64.0		Allocated for intersatellite links	

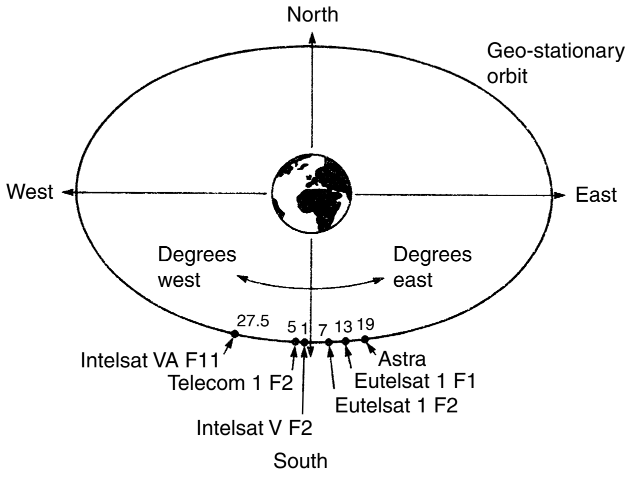
North
Geo-stationary orbit
West
East
Degrees west
Degrees east
27.5
5
1
7
13
19
Intelsat VA F11
Telecom 1 F2
Intelsat V F2
Astra
Eutelsat 1 F1
Eutelsat 1 F2
South

Figure 22.2 Non-DBS satellite television: positions of the main non-DBS satellites relevant to the UK

Table 22.2 European satellite television channels broadcast via communications satellites

Channel	Frequency (GHz)	Polarization	Audio sub-carrier	Video standard	Satellite
3Sat	11.175	H	6.65	PAL	Eutelsat 1 F1
Anglovision	11.515	V	6.6	PAL	Intelsat VA-F11
Arts Channel	11.135	H	6.6	PAL	Intelsat VA-F11
BBC1/2	11.175	H	6.65	PAL	Intelsat VA-F11
CanalJ	12.564	V	5.8	PAL	Telecom 1 F2
Children's Channel	11.015	H	6.6	PAL	Intelsat VA-F11
CNN	11.155	V	6.65	PAL	Intelsat VA-F11
Filmnet	11.140	V	6.6	PAL	Eutelsat 1 F1
Moscow 1	3.675	–	7.0	SECAM	Gorizont-12
Moscow 2	3.0	–	7.0	SECAM	Gorizont-7
Infofilm & Video	11.015	H	6.6	PAL	Intelsat F2
La Cinq	12.606	V	5.8	SECAM	Telecom 1 F2
Lifestyle	11.135	H	6.6	PAL	Intelsat VA-F11
M6	12.648	V	5.8	SECAM	Telecom 1 F2
MTV	10.975	H	6.65	PAL	Intelsat VA-F11
Norsk Rikskringkasti	11.644	H	Digital	C-MAC	Eutelsat 1 F2
Premiere	11.015	H	6.6	PAL	Intelsat VA-F11
RAI-Uno	11.007	H	6.6	PAL	Eutelsat 1 F1
RTL-Plus	11.091	V	6.65	PAL	Eutelsat 1 F1
SAT1	11.507	V	6.65	PAL	Eutelsat 1 F1
Satellite Information	11.575	H	Digital	B-MAC	Intelsat VA-F11
Screensport	11.135	H	6.6	PAL	Intelsat VA-F11
Skychannel	11.650	H	6.65	PAL	Eutelsat 1 F1
Superchannel	10.674	V	6.65	PAL	Eutelsat 1 F1
SVT-2	11.178	H	Digital	C-MAC	Intelsat F2
SVT-2	11.133	H	Digital	C-MAC	Intelsat F2
Teleclub	11.987	V	6.5	PAL	Eutelsat 1 F1
TV5	11.472	H	6.65	PAL	Eutelsat 1 F1
Worldnet	11.512	H	6.65	PAL	Eutelsat 1 F1
Worldnet	11.591	H	6.6	SECAM	Eutelsat 1 F1
Worldnet	12.732	V	5.8	NTSC	Telecom 1 F2

analogue signal corresponding to chrominance and an analogue signal corresponding to luminance are transmitted separately in each broadcast line of the picture.

In order to achieve multiplexing of the three parts of the format, time compression of chrominance and luminance signals occurs before

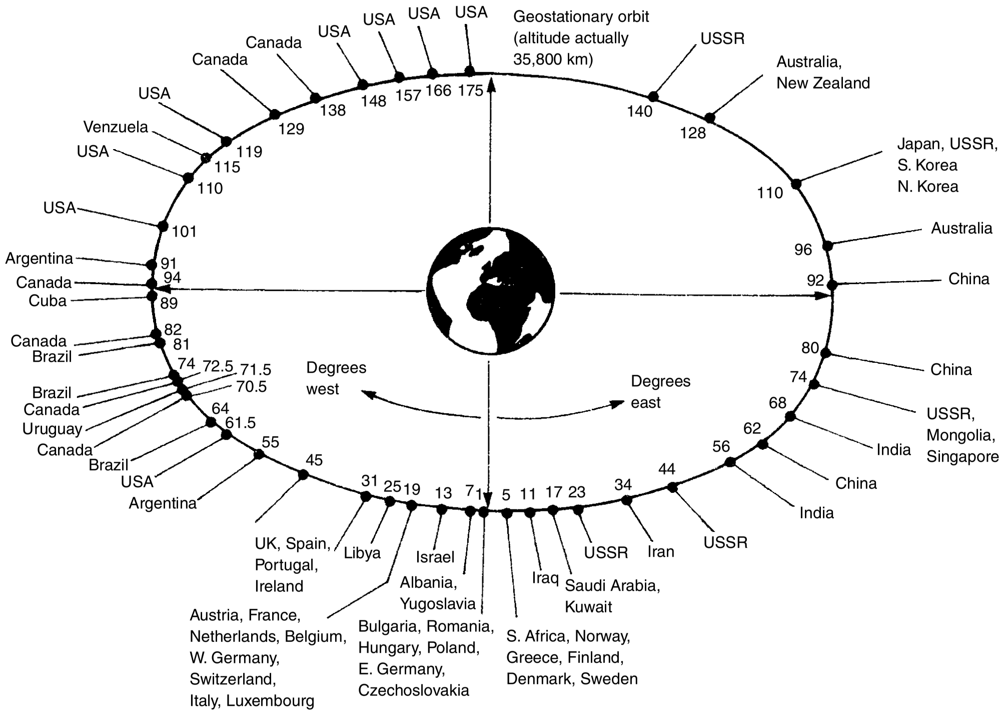
USA
USA
USA
USA
Canada
Canada
USA
Venzuela
USA
USA
Argentina
Canada
Cuba
Canada
Brazil
Brazil
Canada
Uruguay
Canada
Brazil
USA
Argentina
175
166
157
148
138
129
119
115
110
101
91
94
89
82
81
74
72.5
71.5
70.5
64
61.5
55
45
31
25
19
13
7
1
5
11
17
23
34
44
56
62
68
74
80
92
96
110
128
140
Geostationary orbit
(altitude actually
35,800 km)
USSR
Australia,
New Zealand
Japan, USSR,
S. Korea
N. Korea
Australia
China
China
USSR,
Mongolia,
Singapore
India
China
India
USSR
Iran
USSR
Saudi Arabia,
Kuwait
Iraq
S. Africa, Norway,
Greece, Finland,
Denmark, Sweden
Albania,
Yugoslavia
Bulgaria, Romania,
Hungary, Poland,
E. Germany,
Czechoslovakia
Israel
Libya
UK, Spain,
Portugal,
Ireland
Austria, France,
Netherlands, Belgium,
W. Germany,
Switzerland,
Italy, Luxembourg
Degrees
west
Degrees
east

Figure 22.3 World allocations of DBS satellite positions

Table 22.3 DBS television channels

Channel number	Frequency (GHz)		
1	11.72748	21	12.11108
2	11.74666	22	12.13026
3	11.76584	23	12.14944
4	11.78502	24	12.16862
5	11.80420	25	12.18780
6	11.82338	26	12.20698
7	11.84256	27	12.22616
8	11.86174	28	12.24534
9	11.88092	29	12.26452
10	11.90010	30	12.28370
11	11.91928	31	12.30288
12	11.93846	32	12.32206
13	11.95764	33	12.34124
14	11.97682	34	12.36042
15	11.99600	35	12.37960
16	12.01518	36	12.39878
17	12.03436	37	12.41796
18	12.05354	38	12.43714
19	12.07272	39	12.45632
20	12.09190	40	12.47550

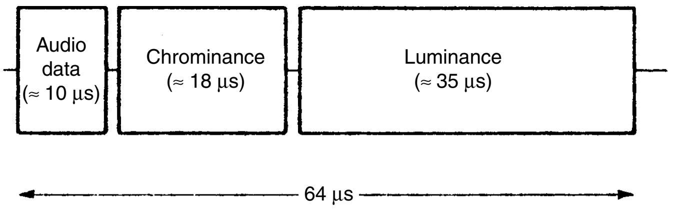

transmission, and they must be reconstituted at the receiver. Two main variations of the MAC format have been selected for European broadcasters, D-MAC and D2-MAC. They differ basically in the number of data channels, and hence the overall bandwidth required. Both modulate video signals in FM, and data signals in duobinary FM.

22.4 Global positioning system (GPS)

This system uses an American government satellite network comprised of 24 satellites in three circular orbits at a height of 20 000 km and with twelve-hour periods. At least six satellites should be visible from any point around the earth at any one time.

Table 22.4 Direct broadcast by satellite (DBS) television. European allocation of satellite positions, channel and polarizations

		Orbital position		
Band	*5° E*	*19° W*	*31° W*	*37° W*
11.7–12.1 GHz RH polarized	*Turkey:* CH. 1, 5, 9, 13, 17 *Greece:* CH. 3, 7, 11, 15, 19	*France:* CH. 1, 5, 9, 13, 17 *Luxembourg:* CH. 3, 7, 11, 15, 19	*Eire:* CH. 2, 6, 10, 14, 18 *UK:* CH. 4, 8, 12, 16, 20	*San Marino:* CH. 1, 5, 9, 13, 17 *Lichtenstein:* CH. 3, 7, 11, 15, 19
11.7–12.1 GHz LH polarized	*Finland:* CH. 2, 6, 10 *Norway:* CH. 14, 18 *Sweden* CH. 4, 8 *Denmark:* CH. 12, 16, 20	*Germany:* CH. 2, 6, 10, 14, 18 *Austria:* CH. 4, 8, 12, 16, 20	*Portugal:* CH. 3, 7, 11, 15, 19	*Andorra:* CH. 3, 8, 12, 16, 20
12.1–12.5 GHz RH polarized	*Cyprus:* CH. 21, 25, 29, 33, 37 *Iceland, etc.:* CH. 23, 27, 31, 35, 39	*Belgium:* CH. 21, 25, 29, 33, 37 *Netherlands:* CH. 23, 27, 31, 35, 39		*Monaco:* CH. 21, 25, 29, 33, 37 *Vatican* CH. 23, 27, 31, 35, 39
12.1–12.5 GHz LH polarized	*Nordic group** CH. 22, 24, 26, 28 30, 32, 36, 40 *Sweden:* CH. 34 *Norway:* CH. 38	*Switzerland:* CH. 22, 26, 30, 34, 38 *Italy:* CH. 24, 28, 32, 36, 40	*Iceland:* CH. 21, 25, 29, 33, 37 *Spain:* CH. 23, 27, 31, 35, 39	

*Wide beam channels: Denmark, Finland, Norway, Sweden.

Table 22.5 Astra television channels

Channel	*Astra 1-A (GHz)*	*Channel*	*Astra 1-B (GHz)*	*Channel*	*Astra 1-C (GHz)*	*Channel*	*Astra 1-D (GHz)*
1	11.21425	17	11.46425	33	10.96425	49	10.71425
2	11.22900	18	11.47900	34	10.97900	50	10.72900
3	11.24375	19	11.49375	35	10.99375	51	10.74375
4	11.25850	20	11.50850	36	10.00850	52	10.75850
5	11.27325	21	11.52325	37	10.02325	53	10.77325
6	11.28800	22	11.53800	38	10.03800	54	10.78800
7	11.30275	23	11.55275	39	10.05275	55	10.80275
8	11.31750	24	11.56750	40	10.06750	56	10.81750
9	11.33225	25	11.58225	41	10.08225	57	10.83225
10	11.34700	26	11.59700	42	10.09700	58	10.84700
11	11.36175	27	11.61175	43	10.11175	59	10.86175
12	11.37650	28	11.62650	44	10.12650	60	10.87650
13	11.39125	29	11.64125	45	10.14125	61	10.89125
14	11.40600	30	11.65600	46	10.15600	62	10.90600
15	11.42075	31	11.67075	47	10.17075	63	10.92075
16	11.43550	32	11.68550	48	10.18550	64	10.83550

Each satellite transmits continuously updated information about its orbit on two frequencies, 1227 MHz and 1575 MHz. One radio channel carries two pseudo-random codes, one very long and the other very short. The second channel is modulated only with the short code. The codes enable the satellite to be positively identified, and the distance from a receiver on, or close to, the earth to be calculated.

A receiver's position, in three-dimensional space, is identified by measuring and calculating the distance from three satellites. The short codes provide the initial fix but increased precision is obtained from the long codes. There is an error due to time variations arising from various sources of which satellite speed is one, but taking a measurement from a fourth satellite enables a correction factor to be applied.

Receivers which utilize only the short PN codes provide a resolution accurate to about 100 m. Those which can process the long codes provide a fix accurate to about 45 m.

References

Lewis, G.E. (1988). *Communication Services via Satellite*. BSP Professional Books, Oxford.

Long, M. *The 1993/1994 World Satellite Annual*. Mark Long Enterprises Inc., Ft Lauderdale, USA.

23 Connectors and interfaces

23.1 Audio and video connectors

Audio connectors

The DIN standards devised by the German Industrial Standards Board are widely used for the connection of audio equipment. The connectors are shown below. The 3-way and 5-way 45 are the most common and connections for those are listed.

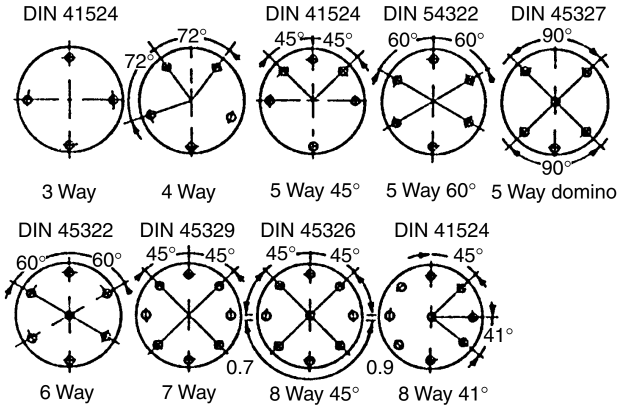
DIN 41524
72°
72°
DIN 41524
45°
45°
DIN 54322
60°
60°
DIN 45327
90°
90°
3 Way
4 Way
5 Way 45°
5 Way 60°
5 Way domino
DIN 45322
60°
60°
DIN 45329
45°
45°
DIN 45326
45°
45°
DIN 41524
45°
41°
0.7
0.9
6 Way
7 Way
8 Way 45°
8 Way 41°

	Mono	*Stereo*
Microphone	Input 1 O V 2 Pin 3 available for polarizing voltage	Input LH 1 Input RH 4 O V 2 Pins 3 and 5 available for polarizing voltage
Tape recorder inputs and monitor outputs	Input 1 O V 2 Output 3	Input LH 1 Input R4 4 O V 2 Output LH 3 Output RH 5

	Mono	*Stereo*
Tape recorder replay output	Output, low Z1 O V 2 Output, high Z3	Output LH, low Z1 Output RH, low Z4 O V 2 Output LH, high Z3 Output RH, high Z5
Amplifiers	Output to tape 1 O V 2 Input from tape 3	Output LH 1 Output RH 1 O V 2 Input LH 3 Input RH 5

Variations on the above exist between different manufacturers.

Videorecorder/televisions/camera connectors

Standard pin configurations for videorecorders, televisions and videocameras are shown below. Many follow standard DIN connector pinouts, but videocamera and SCART connectors differ significantly.

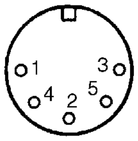

(a) 5-pin DIN

1,4 Audio in
2 Chassis
3,5 Audio out or stereo
1 L audio in
2 Chassis
3 L audio out
4 R audio in
5 R audio out

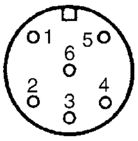

(b) 6-pin DIN

1 AV select, For VCR high = all outputs, low = all inputs, Opposite for TV set
2 Video in/out
3 Chassis
4 L audio in/out
5 12 V
6 R audio in/out

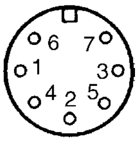

(c) 7-pin DIN

1 L audio in
2 Chassis
3 L audio out
4 R audio in
5 R audio out
6 Remote control data
7 Chassis

SCART (BS 6552)

SCART connectors, also known as Peritelevision or Euroconnector connectors, feature two control systems which allow remote control over the television's or videorecorder's functions.

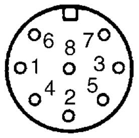

(d) 8-pin DIN

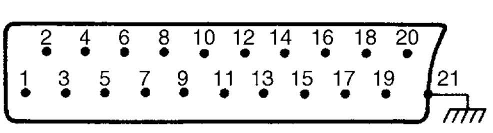

(e) SCART

1 L audio in
2 Remote control data
3 R audio in
4 Chassis (audio)
5 Chassis (remote control)
6 Chassis (video)
7 Chassis (audio)
8 Video in

1 R audio out
2 R audio in
3 L audio out
4 Chassis (audio)
5 Chassis (B video)
6 L video in
7 B video in
8 Source switching
9 Chassis (G video)
10 Data bus
11 G video in
12 Data bus
13 Chassis (data bus)
14 Chassis (data bus)
15 R video in
16 Fast video blanking
17 Chassis (composite video)
18 Chassis (fast video blanking)
19 Composite video out
20 Composite video in

The simplest is a source switching input (pin 8) in which an external source (videorecorder, computer, etc.) can, by issuing a 12 volt signal, cause the television to switch to baseband inputs.

A more complex control system, called *domestic data bus* (D^2B), is given through pins 10 and 12, in which serial data can be passed between controlling microprocessors in the television and external equipment. No standard yet exists for D^2B.

23.2 Co-axial connector

The most commonly used connectors for RF cables are:

Type	*Impedance, Z_0 (ohms)*	*Max. VSWR to (frequency GHz)*	*Maximum proof RF voltage*	*Notes*
N	50 or 75	1.30 (4)	1.5 kV (5 MHz at sea level)	Screw together

Type	*Impedance, Z_0 (ohms)*	*Max. VSWR to (frequency GHz)*	*Maximum proof RF voltage*	*Notes*
BNC	50 or 75	1.01 (1) 1.30 (4)	1 kV (5 MHz at sea level)	Miniature. Bayonet fitting
TNC	50 or 75	As BNC	As BNC	Miniature. Robust screw together
SMB	50	1.46 (4)	1.5 kV (At sea level. Cable dependent)	Sub-miniature. Snap together
SMC	50	1.41 (4) 1.69 (10)	As SMB	Sub-miniature. Screw together
SMD	50	As SMB	As SMB	Sub-miniature. Push together
BT 43	75			Developed from SMB range for use in telecomms and data transmission
7–16	50	1.3 (5 GHz)	2.7 (connector)	Suitable for medium to high power applications in the cellular and broadcast industries. Screw together
PL259/ SO239/ UHF 83	50	–	500 (pk)	Non-constant imp. High VSWR makes it unsuitable for use above 144 MHz and for extending RF cables. Very robust. Screw together
C	75	–	–	Bayonet fitting

Type	*Impedance, Z_0 (ohms)*	*Max. VSWR to (frequency GHz)*	*Maximum proof RF voltage*	*Notes*
F	50	–	–	American CCTV connector used on some 144 MHz hand-portable transceivers. Plugs use inner conductor of cable as centre pin
Belling Lee	50	–	–	British TV antenna connector. Aluminium versions may corrode when used outdoors
GR	50	1 MHz max. frequency	–	Constant imp. sexless connector
Phono	–	–	–	American connector for audio use

Examples of assembly instructions for co-axial RF connectors (by kind permission of M/ACOM Greenpar Ltd.)

Measuring instrument – a rule is shown, but better results are obtained by using a Vernier gauge

Stout trimming blade, suitable for cutting copper wire braid

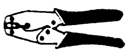

Crimping tool

Soldering iron

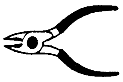

Side cutters, also for trimming braid

Spanner, of the relevant size for the connector

Small screwdriver

Hacksaw, sometimes appropriate for semi-rigid cable, although for repetitive operations a power trimmer should be considered

Assembly instructions Type N

Cable types:

50 ohm: PSF1/4M (BBC), RG 8A/U, RG 213/U, URM 67
75 ohm: RG 11A/U, RG 63B/U, RG 114A/U, URM 64

1 Side clamp nut and plain gasket over cable and trim outer sheath from cable as shown.

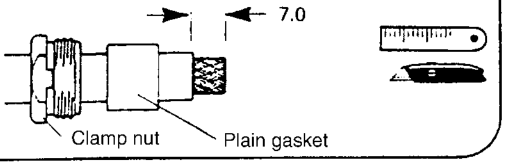

2 Fold back braid and push ferrule over dielectric to trap braid between outer sheath and ferrule. Trim off surplus braid.

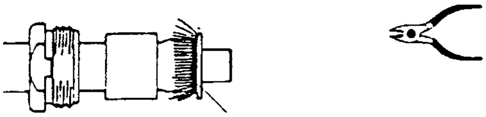

3 Trim back dielectric and check the length of the protruding centre conductor.

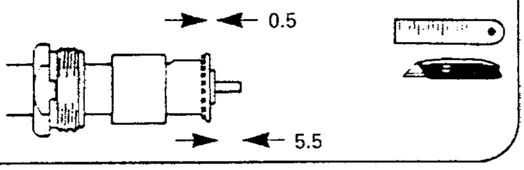

4 Tin centre conductor, then slide rear insulator over dielectric, to butt against ferrule.

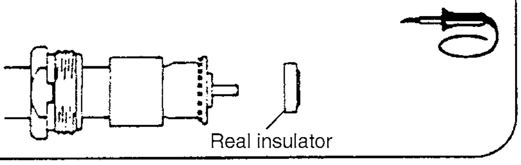

5 Fit contact (male for plugs, female for jacks) onto centre conductor. Hold cable and contact tightly together and solder.

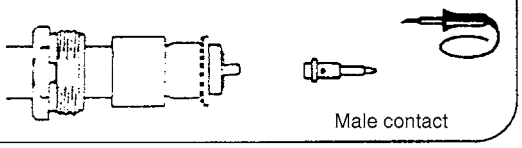

6 Slide plain gasket and clamp nut up to ferrule trapping braid. Fit front insulator over contact to butt against rear insulator and press sub-assembly into body as far as possible.

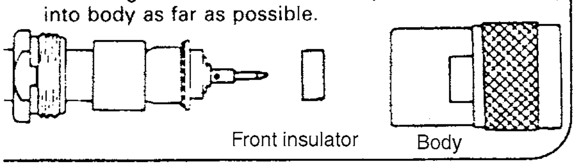

7 Engage and tighten clamp nut.

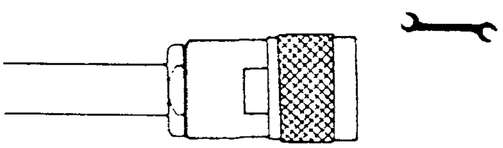

1 Slide crimp sleeve over cable

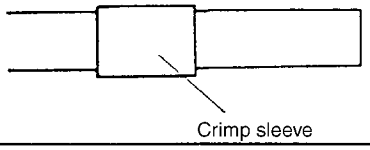

2 Trim back outer sheath and braid to dimensions shown. Any foils adjacent to the dielectric should be left in place but trimmed 1.5 mm from the face of the dielectric

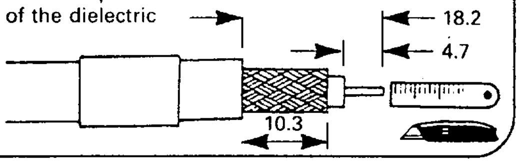

3 Fit contact (male for plugs, female for jacks) over centre conductor to butt against dielectric and crimp.

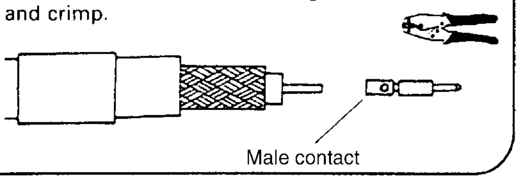

4 Push fit sub-assembly into body until contact clicks into insulator. In all cases, braids (and any intermediate foils) slide over ferrule while any foil adjacent to the dielectric should enter the ferrule bore.

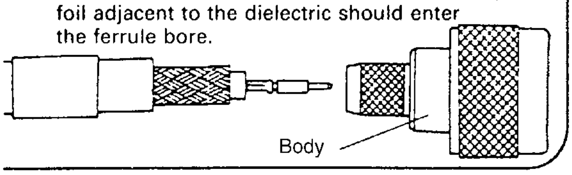

5 Slide crimp sleeve forward over braid and ferrule to butt against rear of the connector and crimp.

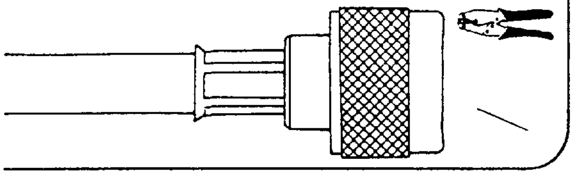

Assembly instructions Type BNC

Cable types:
RG 58C/U, RG 141A/U, URM 43, URM 76

1 Slide clamp nut and plain gasket over cable and trim outer sheath from cable, as shown.

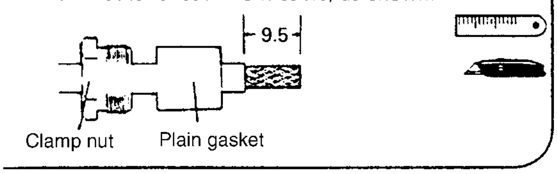

2 Fold back braid and push ferrule over dielectric to trap braid between outer sheath and ferrule. Trim off surplus braid.

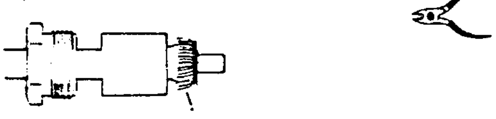

3 Trim back dielectric and check the length of the protruding centre conductor.

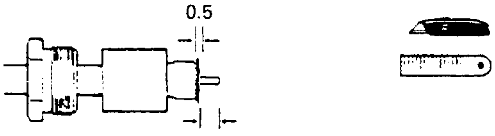

4 Tin centre conductor, then slide rear insulator over dielectric, to butt against ferrule.

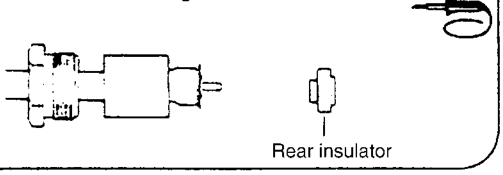

5 Fit contact (male for plugs, female for jacks) onto centre conductor, with collar pressed into recess in rear insulator. Hold cable and contact tightly together, and solder.

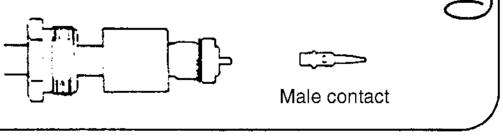

6 Slide plain gasket and clamp nut up to ferrule, trapping braid. Fit front insulator over contact to butt against rear insulator and press sub-assembly into body as far as possible.

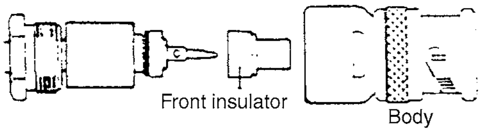
Front insulator
Body

7 Engage and tighten clamp nut

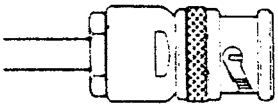

1 Slide metal crimp sleeve over cable, trim outer sheath from cable as shown.

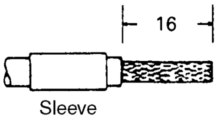

2 Trim back braid and dielectric to the dimensions shown.

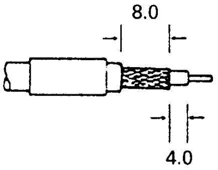

3 Fit contact over centre conductor to butt against the dielectric, then crimp.

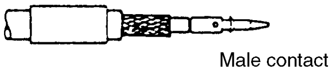
Male contact

4 Press sub-assembly into body, until contact clicks into place and ensuring that the knurled ferrule is inserted between the dielectric and braid.

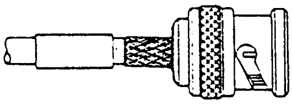

5 Slide the sleeve along the cable, until it butts against the body sub-assembly. Crimp, using the tool listed below.

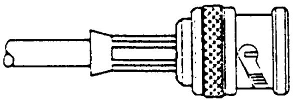

Note: A plug is shown, but these instructions are relevant to both plugs and jacks. The shape of contacts and insulators may also vary from the drawings shown.

Assembly instructions Types SMB/SMC/SMD

Cable types:
TM 3306, RG174A/U, RG188A/U, RG316/U
TM 3263, RG178B/U, RG196A/U, URM 110

1 Slide clamp nut, a washer, a gasket and the other washer over the cable trim outer sheath to dimension shown.

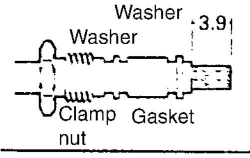

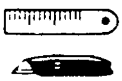

2 Fold back braid. Push ferrule over dielectric to trap braid between outer sheath and ferrule.

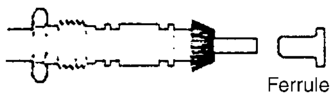

3 Trim off surplus braid. Trim dielectric flush with ferrule and check length of centre conductor. Tin centre conductor.

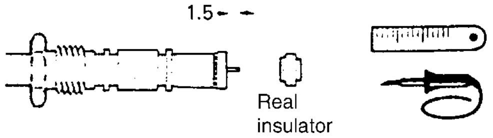

4 Side rear insulator over centre conductor until it butts against ferrule.

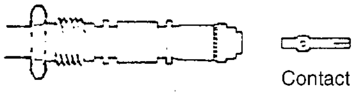

5 Fit contact onto centre conductor until it butts against rear insulator. Hold cable and contact tightly together and solder.

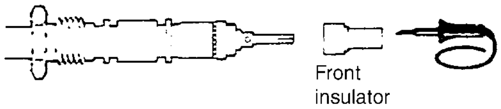

6 Fit front insulator over contact until it butts against internal shoulder.

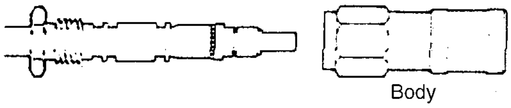

7 Press sub-assembly into body as far as possible. Engage and tighten clamp nut.

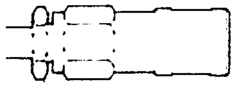

1 Place crimp sleeve over sheat and trim cable to the dimensions shown. Ensure that the centre conductor is not damaged.

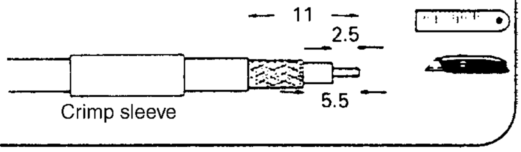

2 Place contact on centre conductor and crimp.

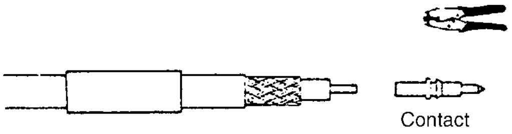

3 Push contact into body sub-assembly. Ensure the contact 'clicks' into rear insulator with the ferrule between the braid and the dielectric.

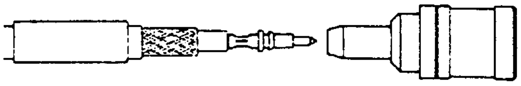

4 Slide crimp sleeve over the braid and crimp.

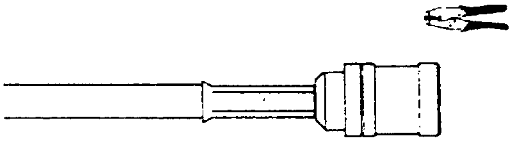

23.3 Interfaces

23.3.1 Connectors and connections

Data interchange by modems

When transmitting and receiving data across telephone or other circuits, the equipment which actually generates and uses the data (e.g. a computer or VDU terminal) is known as *data terminating equipment* (DTE). The equipment which terminates the telephone line and converts the basic data signals into signals which can be transmitted is known as *data circuit-terminating equipment* (DCE). As far as the user is concerned the interface between DTE and DCE is the most important. ITU-T recommendation V24 defines the signal interchanges and functions between DTE and DCE; these are commonly known as the 100 series interchanges circuits:

Interchange circuit		*Data*		*Control*		*Timing*	
Number	*Name*	*From DCE*	*To DCE*	*From DCE*	*To DCE*	*From DCE*	*To DCE*
101	Protective ground or earth						
102	Signal ground or common return						
103	Transmitted data		●				
104	Received data	●					
105	Request to send				●		
106	Ready for sending			●			
107	Data set ready			●			
108/1	Connect data set to line				●		
108/2	Data terminal ready				●		
109	Data channel received line signal detector			●			
110	Signal quality detector			●			
111	Data signalling rate selector (DTE)				●		
112	Data signalling rate selector (DCE)			●			
113	Transmitter signal element timing (DTE)						●

Interchange circuit		*Data*		*Control*		*Timing*	
Number	*Name*	*From DCE*	*To DCE*	*From DCE*	*To DCE*	*From DCE*	*To DCE*
114	Transmitter signal element timing (DCE)					•	
115	Receiver signal element timing (DCE)					•	
116	Select stand by				•		
117	Standby indicator			•			
118	Transmitted backward channel data		•				
119	Received backward channel data	•					
120	Transmit backward channel line signal				•		
121	Backward channel ready			•			
122	Backward channel received line signal detector			•			
123	Backward channel single quality detector			•			
124	Select frequency groups				•		
125	Calling indicator			•			
126	Select transmit frequency				•		
127	Select receive frequency				•		

Interchange circuit		*Data*		*Control*		*Timing*	
Number	*Name*	*From DCE*	*To DCE*	*From DCE*	*To DCE*	*From DCE*	*To DCE*
128	Receiver signal element timing (DTE)						●
129	Request to receive				●		
130	Transmit backward tone				●		
131	Received character timing					●	
132	Return to non-data mode				●		
133	Ready for receiving				●		
134	Received data present			●			
191	Transmitted voice answer				●		
192	Received voice answer			●			

Modem connector pin numbers

The connectors used with 100 series interchange circuits and its pin assignments are defined by international standard ISO 2110 and are (for modems following the ITU-T recommendations V21, V23, V26, V26bis, V27 and V27bis) as follows:

	Interchange circuit numbers			
Pin number	*V21*	*V23*	*V26/ V27*	*Pin Name*
1	*1	*1	*1	Ground
2	103	103	103	TXD
3	104	104	104	RXD
4	105	105	105	RTS
5	106	106	106	RFS
6	107	107	107	DSR
7	102	102	102	Signal return
8	109	109	109	Signal DET
9	*N	*N	*N	
10	*N	*N	*N	
11	126	*N	*N	STF
12	*F	122	122	
13	*F	121	121	
14	*F	118	118	
15	*F	*2	114	
16	*F	119	119	
17	*F	*2	115	
18	141	141	141	
19	*F	120	120	
20	108/1-2	108/1-2	108/1-2	DTR
21	140	140	140	
22	125	125	125	Call ind.
23	*N	111	111	
24	*N	*N	113	
25	142	142	142	

Notes:

*1 Pin 1 is assigned for connecting the shields between tandem sections of shielded cables. It may be connected to protective ground or signal ground.

*F Reserved for future use.

*N Reserved for national use.

9-Pin connector

Pin number	*Pin name*	*Description*
1	DCD	Data carrier detect
2	RXD	Receive data
3	TXD	Transmit data
4	DTR	Data terminal ready
5	GND	Ground
6	DSR	Data set ready
7	RTS	Ready to send
8	CTS	Clear to send
9	RI	Ring indicator

Automatic calling

A similar series of interchange circuits is defined in ITU-T recommendation V25 for automatic calling answering between modems over the telephone network. This is the 200 series interchange circuits:

Interchange circuit

Number	*Name*	*From DCE*	*To DCE*
201	Signal ground	●	●
202	Call request		●
203	Data line occupied	●	
204	Distant station connected	●	
205	Abandon call	●	
206	Digit signal (2^0)		●
207	Digit signal (2^1)		●
208	Digit signal (2^2)		●
209	Digit signal (2^3)		●
210	Present next digit	●	
211	Digit present		●
213	Power indication	●	

RS 232C

The EIA equivalent of CCITT V24 interface is the RS 232C specification, which similarly defines the electrical interface between DTE and DCE. Although the two have different designations, they are to all practical purposes equivalent. The RS 232C interchange circuits are:

Interchange circuit		*Data*		*Control*		*Timing*	
Mnemonic	*Name*	*From DCE*	*To DCE*	*From DCE*	*To DCE*	*From DCE*	*To DCE*
AA	Protective ground						
AB	Signal ground/common return						
BA	Transmitted data		•				
BB	Received data	•					
CA	Request to send				•		
CB	Clear to send			•			
CC	Data set ready			•			
CD	Data terminal ready				•		
CE	Ring indicator			•			
CF	Received line signal detector			•			
CG	Signal quality detector			•			
CH	Data signal rate selector (DTE)				•		
CI	Data signal rate selector (DCE)			•			
DA	Transmitter signal element timing (DTE)						•
DB	Transmitter signal element timing (DCE)					•	
DD	Receiver signal element timing (DCE)					•	

Interchange circuit		*Data*		*Control*		*Timing*	
Mnemonic	*Name*	*From DCE*	*To DCE*	*From DCE*	*To DCE*	*From DCE*	*To DCE*
SBA	Secondary transmitted data		●				
SBB	Secondary received data	●					
SCA	Secondary request to send				●		
SCB	Secondary clear to send			●			
SCF	Secondary received line signal detector			●			

RS 449

The EIA RS 232C standard, although the most common, is by no means perfect. One of its main limitations is the maximum data rate −18.2 K baud. Various improved interchange circuits (RS 422, RS 423) have been developed. The RS 449 standard is capable of very fast data rates (up to 2 Mbaud):

Interchange circuit		*Data*		*Control*		*Timing*	
Mnemonic	*Name*	*From DCE*	*To DCE*	*From DCE*	*To DCE*	*From DCE*	*To DCE*
SG	Single ground						
SC	Send common						
RC	Receive common						
IS	Terminal in service				●		
IC	Incoming call			●			
TR	Terminal ready				●		

Interchange circuit			*Data*		*Control*		*Timing*	
	Mnemonic	*Name*	*From DCE*	*To DCE*	*From DCE*	*To DCE*	*From DCE*	*To DCE*
Primary channel	DM	Data mode			•			
	SD	Send data		•				
	RD	Receive data	•					
	TT	Terminal timing						•
	ST	Send timing					•	
	RT	Receive timing					•	
	RS	Request to send				•		
	CS	Clear to send			•			
	RR	Receiver ready			•			
	SQ	Signal quality			•			
	NS	News signal				•		
	SF	Select frequency				•		
	SR	Signalling rate selector				•		
	SI	Signalling rate indicator			•			
Secondary channel	SSD	Secondary send data		•				
	SRD	Secondary receive data	•					
	SRS	Secondary request to send				•		
	SCS	Secondary clear to send			•			
	SRR	Secondary receiver ready			•			
	LL	Local loopback				•		
	RL	Remote loopback				•		
	TM	Test mode			•			
	SS	Select standby				•		
	SB	Standby indicator			•			

Centronics interface

Most personal computers use the Centronics parallel data transfer to a printer. The pin connections of the connector, abbreviations and signal descriptions are shown.

All signals are standard TTL, although not all signals necessarily exist in any given interface.

Pin number	*Abbreviation*	*Signal description*
1	STROBE	Strobe
2	DATA1	Data line 1
3	DATA2	Data line 2
4	DATA3	Data line 3
5	DATA4	Data line 4
6	DATA5	Data line 5
7	DATA6	Data line 6
8	DATA7	Data line 7
9	DATA8	Data line 8
10	ACKNLG	Acknowledge data
11	BUSY	Busy
12	PE	Paper end
13	SLCT	Select printer
14	AUTO FEED XT	Automatic line feed at end of line
15	NC	No connection
16	OV	Logic ground
17	CHASSIS GND	Printer chassis (not necessarily the same as logic ground)
18	NC	No connection
19 to 30	GND	Single ground
31	INIT	Initialize
32	ERROR	Error
33	GND	Signal ground
34	NC	No connection
35	Logic 1	Logic 1
36	SLCT IN	Select input to printer

RS 232C/RS 449/V24 comparison

RS 232C		*RS 449*		*V24*	
AB	Signal ground	SG	Signal ground	102	Signal ground
		SC	Send common	102a	DTE common
		RC	Receive common	102b	DCE common
		IS	Terminal in service		
CE	Ring indicator	IC	Incoming call	125	Calling indicator
CD	Data terminal ready	TR	Terminal ready	108/2	Data terminal ready
CC	Data set ready	DM	Data mode	107	Data set ready
BA	Transmitted data	SD	Send data	103	Transmitted data
BB	Receive data	RD	Receive data	104	Received data
DA	Transmitter signal element Timing (DTE source)	TT	Terminal timing	113	Transmitter signal element Timing (DTE source)
DB	Transmitter signal element Timing (DCE source)	ST	Send timing	114	Transmitter signal element Timing (DCE source)
DD	Receiver signal element timing	RT	Receive timing	115	Receiver signal element timing (DCE source)
CA	Request to send	RS	Request to send	105	Request to send
CB	Clear to send	CS	Clear to send	106	Ready for sending
CF	Received line single detector	RR	Receiver ready	109	Data channel received line signal detector

CG	Signal quality detector	SQ	Signal quality	110	Data signal quality detector
		NS	New signal		
		SF	Select frequency	126	Select transmit frequency
CH	Data signal rate selector (DTE source)	SR	Signalling rate selector	111	Data signalling rate selector (DTE source)
CI	Data signal rate selector (DCE source)	SI	Signalling rate indicator	112	Data signalling rate selector (DCE source)
SBA	Secondary transmitted data	SSD	Secondary send data	118	Transmitted backward channel data
SBB	Secondary received data	SRD	Secondary receive data	119	Received backward channel data
SGA	Secondary request to send	SRS	Secondary request to send	120	Transmit backward channel line signal
SCB	Secondary clear to send	SCS	Secondary clear to send	121	Backward channel ready
SCF	Secondary received line signal detector	SRR	Secondary receiver ready	122	Backward channel received line signal detector
		LL	Local loopback	141	Local loopback
		RL	Remote loopback	140	Remote loopback
		TM	Test mode	143	Test indicator
		SS	Select standby	116	Select standby
		SB	Standby indicator	117	Standby indicator

Reference

M/ACOM Greenpar Ltd, catalogue 1993.

24 Broadcasting

24.1 Standard frequency and time transmissions

Frequency (mHz)	*Wavelength (m)*	*Code*	*Station location*	*Country*	*Power (kW)*
60 kHz	5000	MSF	Rugby	England	–
–	–	WWVB	Colorado	USA	–
75 kHz	4000	HBG	–	Switzerland	–
77.5 kHz	3871	DCF77	Mainflingen	DDR	–
1.5	200	HD210A	Guayaquil	Ecuador	–
2.5	120	MSF	Rugby	England	0.5
–	–	WWV	Fort Collins	USA	2.5
–	–	WWVH	Kekaha	Hawaii	5
–	–	ZLF	Wellington	New Zealand	–
–	–	RCH	Tashkent	USSR	1
–	–	JJY	–	Japan	–
–	–	ZUO	Olifantsfontein	South Africa	–
3.33	90.09	CHU	Ottawa	Canada	3
3.81	78.7	HD201A	Guayaquil	Ecuador	–
4.5	66.67	VNG	Victoria	Australia	–
4.996	60.05	RWM	Moscow	USSR	5
5	60	MSF	Rugby	England	0.5
–	–	WWVB	Fort Collins	USA	10
–	–	WWVH	Kekaha	Hawaii	10
–	–	ATA	New Delhi	India	–
–	–	LOL	Buenos Aires	Argentina	2
–	–	IBF	Turin	Italy	5
–	–	RCH	Tashkent	USSR	1
–	–	JJY	–	Japan	–
–	–	ZUO	Olifantsfontein	South Africa	–
5.004	59.95	RID	Irkutsk	USSR	1
6.10	49.2	YVTO	Caracas	Venezuela	–
7.335	40.9	CHU	Ottawa	Canada	10
7.5	40	VNG	Lyndhurst	Australia	5

Frequency (mHz)	*Wavelength (m)*	*Code*	*Station location*	*Country*	*Power (kW)*
7.6	39.4	HD210A	Guayaquil	Ecuador	–
8	37.5	JJY	–	Japan	–
8.1675	36.73	LQB9	Buenos Aires	Argentina	5
9996	30.01	RWM	Moscow	USSR	–
10	30	MSF	Rugby	England	0.5
–	–	WWVB	Fort Collins	USA	10
–	–	WWVH	Kekaha	Hawaii	10
–	–	BPM	Xian	China	–
–	–	ATA	New Delhi	India	–
–	–	JJY	–	Japan	–
–	–	LOL	Buenos Aires	Argentina	2
–	–	RTA	Novosibirsk	USSR	5
–	–	RCH	Tashkent	USSR	1
10.004	29.99	RID	Irkutsk	USSR	1
12	25	VNG	Lyndhurst	Australia	10
14.67	20.45	CHU	Ottawa	Canada	3
14.996	20.01	RWM	Moscow	USSR	8
15	20	WWVB	Fort Collins	USA	10
–	–	WWVH	Kekaha	Hawaii	10
–	–	LOL	Buenos Aires	Argentina	2
–	–	RTA	Novosibirsk	USSR	5
–	–	BPM	Xian	China	–
–	–	ATA	New Delhi	India	–
–	–	JJY	–	Japan	–
15.004	19.99	RID	Irkutsk	USSR	1
16.384	18.31	–	Allouis	France	2000
15.55	17.09	LQC20	Buenos Aires	Argentina	5
20	15	WWVB	Fort Collins	USA	2.5
100	3	ZUO	Olifantsfontein	South Africa	–

24.2 Standard frequency formats

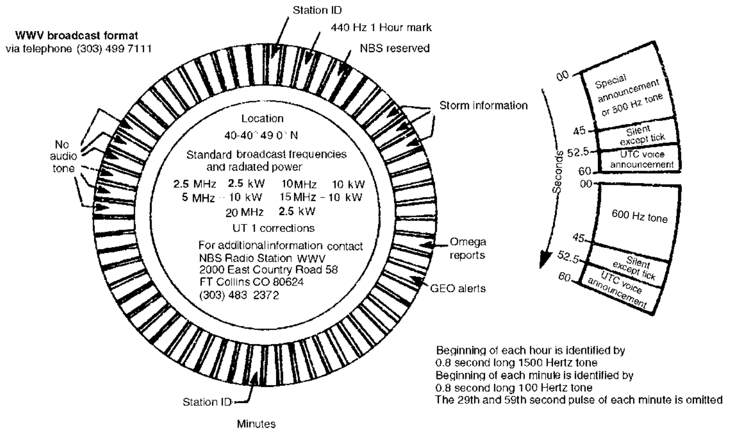
WWV broadcast format
via telephone (303) 499 7111
Station ID
440 Hz 1 Hour mark
NBS reserved
Storm information
No audio tone
Location
40·40° 49 0' N
Standard broadcast frequencies and radiated power
2.5 MHz 2.5 kW 10MHz 10 kW
5 MHz 10 kW 15MHz - 10 kW
20 MHz 2.5 kW
UT 1 corrections
For additional information contact
NBS Radio Station WWV
2000 East Country Road 58
FT Collins CO 80624
(303) 483 2372
Omega reports
GEO alerts
Station ID
Minutes
00
Special announcement or 500 Hz tone
45
Silent except tick
52.5
UTC voice announcement
60
00
600 Hz tone
45
52.5
Silent except tick
60
UTC voice announcement
Seconds
Beginning of each hour is identified by 0.8 second long 1500 Hertz tone
Beginning of each minute is identified by 0.8 second long 100 Hertz tone
The 29th and 59th second pulse of each minute is omitted

MSF Rugby

Time is inserted in the 60 kHz transmission in two ways, illustrated below.

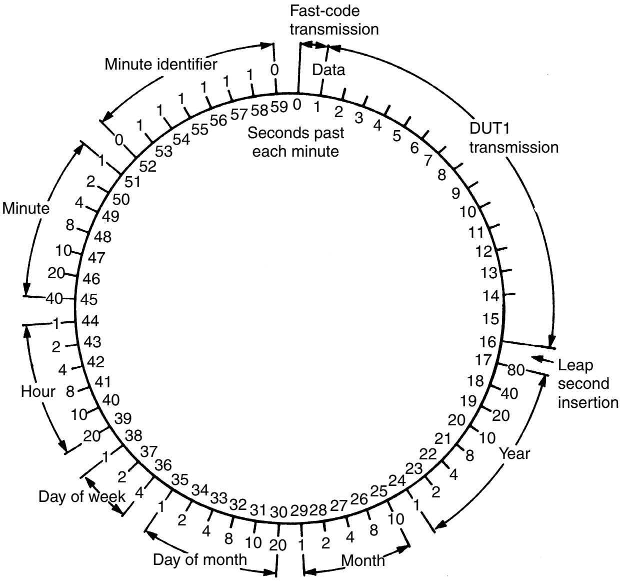
Fast-code transmission
Data
Minute identifier
DUT1 transmission
Seconds past each minute
Minute
Hour
Day of week
Day of month
Month
Year
Leap second insertion

Slow code time and date information is transmitted between the 17th and 59th seconds of the minute-long cycle, in normal BCD coding. *Fast code* time and date BCD coded information is inserted into a 500 ms window in the first second of each minute-long cycle, as illustrated below.

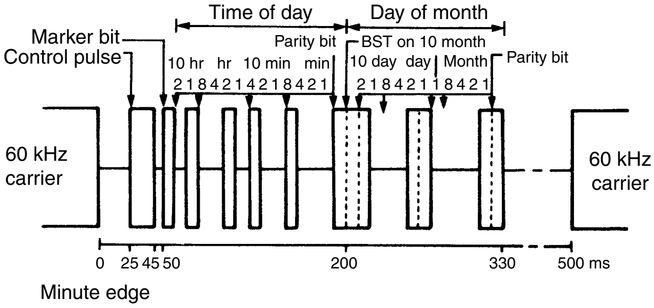
Time of day
Day of month
Marker bit
Control pulse
Parity bit
BST on 10 month
10 hr hr 10 min min
10 day day Month
Parity bit
2 1 8 4 2 1 4 2 1 8 4 2 1
2 1 8 4 2 1 1 8 4 2 1
60 kHz carrier
60 kHz carrier
0 25 45 50
200
330
500 ms
Minute edge

24.3 UK broadcasting bands

Frequency band	Frequency	Use
Long wave	150–285 kHz (2000–1053 m)	AM radio
Medium wave	525–1605 kHz (571–187 m)	AM radio
Band II (VHF)	88–108 MHz	FM radio
Band IV (UHF)	470–582 MHz (channels 21 to 34)	TV
Band V (UHF)	614–854 MHz (channels 39 to 68)	TV
Band VI (SHF)	11.7–12.5 GHz (channels 1 to 40)	Satellite TV

24.4 BBC VHF test tone transmissions

Transmission starts about 4 minutes after the end of Radio 3 programmes on Mondays and Saturdays.

Time min.	*Left channel*	*Right channel*	*Purpose*
–	250 Hz at zero level	440 Hz at zero level	Identification of left and right channels and setting of reference level
2	900 Hz at +7 dB	900 Hz at +7 dB, antiphase to left channel	Adjustment of phase of regenerated subcarrier (see Note 4) and check of distortion with L−R signal only
6	900 Hz at +7 dB	900 Hz +7 dB, in phase with left channel	Check of distortion with L + R signal only
7	900 Hz at +7 dB	No modulation	Check of L to R cross-talk
8	No modulation	900 Hz at +7 dB	Check of R to L cross-talk
9	Tone sequence at −4 dB: 40 Hz 6·3 kHz 100 Hz 10 kHz 500 Hz 12.5 kHz 1000 Hz 14 kHz This sequence is repeated	No modulation	Check of L-channel frequency response and L to R cross-talk at high and low frequencies
11′40″	No modulation	Tone sequence as for left channel	Check of R-channel frequency response and R to L cross-talk at high and low frequencies

Time min.	*Left channel*	*Right channel*	*Purpose*
14′20″	No modulation	No modulation	Check of noise level in the presence of pilot
15′20″	End of test transmissions		

Notes:

1. This schedule is subject to variation or cancellation to accord with programme requirements and essential transmission tests.
2. The zero level reference corresponds to 40% of the maximum level of modulation applied to either stereophonic channel before pre-emphasis. All tests are transmitted with pre-emphasis.
3. Periods of tone lasting several minutes are interrupted momentarily at one-minute intervals.
4. With receivers having separate controls of subcarrier phase and crosstalk, the correct order of alignment is to adjust first the subcarrier phase to produce maximum output from either the L or the R channel and then to adjust the crosstalk (or 'separation') control for minimum crosstalk between channels.
5. With receivers in which the only control of crosstalk is by adjustment of subcarrier phase, this adjustment of subcarrier phase, this adjustment should be made on the crosstalk checks.
6. Adjustment of the balance control to produce equal loudness from the L and R loudspeakers is best carried out when listening to the announcements during a stereophonic transmission, which are made from a centre-stage position. If this adjustment is attempted during the tone transmissions, the results may be confused because of the occurrence of standing-wave patterns in the listening room.
7. The outputs of most receivers include significant levels of the 19-kHz tone and its harmonics, which may affect signal-level meters. It is important, therefore, to provide filters with adequate loss at these frequencies if instruments are to be used for the above tests.

24.5 Engineering information about broadcast services

Information about all BBC services as well as advice on how best to receive transmissions (including television) can be obtained from:

British Broadcasting Corporation
Engineering Liaison
White City
201 Wood Lane
London W12 7TS

Telephone: 020 8752 5040

Transmitter service maps for most main transmitters can also be supplied, but requests for maps should be accompanied by a stamped addressed A4 sized envelope.

Similarly, information about all IBA broadcast services can be obtained from:

Radio:
Radiocommunications Agency
Wyndham House
189 Marsh Wall
London E14 9SX

Telephone: 020 7211 0211

Television:
The Independent Television Commission
Kings Worthy Court
Kings Worthy
Winchester
Hants SO23 7QA

Telephone: 01962 848647

24.6 Characteristics of UHF terrestrial television systems

24.6.1 World systems

System	Number of lines	Channel width (MHz)	Vision bandwidth (MHz)	Vision/sound separation (MHz)	Vestigial side-band (MHz)	Vision modulation	Sound modulation	Field frequency
A	625	8	5.5	+6	1.25	Neg.	FM	50
B	625	8	5.0	+5.5	1.25	Neg.	FM	50
C	625	8	5.0	+5.5	0.75	Neg.	FM	50
D	625	8	6.0	+6.5	1.25	Pos.	AM	50
E	625	8	6.0	+6.5	0.75	Neg.	FM	50
F	525	6	4.2	+4.5	1.25	Neg.	FM	60

A – UK and Eire
B – Eastern Europe
C – Most of Western Europe, Australia, New Zealand
D – France
E – Russia and Eastern Europe
F – USA, most of Central and South America, Japan

24.6.2 European systems

Country	System	Colour
Austria	C	PAL
Belgium	B	PAL
Bulgaria	No UHF system	
Cyprus	C	PAL
Czechoslovakia	E	SECAM
Denmark	No UHF system	PAL
Finland	C	PAL
France	D	SECAM
Germany	C	PAL

Country	*System*	*Colour*
German DR	C	SECAM
Greece	B	PAL
Holland	C	PAL
Hungary	E	SECAM
Iceland	No UHF system	
Ireland	A	PAL
Italy	C	PAL
Luxembourg	D	SECAM
Malta	B	PAL
Monaco	D	SECAM
Norway	C	PAL
Poland	E	SECAM
Portugal	C	PAL
Romania	E	PAL
Spain	C	PAL
Sweden	C	PAL
Switzerland	C	PAL
Turkey	No UHF system	
UK	A	PAL
USSR	E	SECAM
Yugoslavia	B	PAL

24.6.3 UK 625-line television system specification

Channel bandwidth	8 MHz
Upper sideband (vision signal)	5.5 MHz
Lower sideband (vision signal)	1.25 MHz
Vision modulation	AM negative
Sound modulation	FM
Sound deviation (max.)	±50 kHz
Sound pre-emphasis	50 μs
Sound carrier relative to vision carrier	+
Aspect ratio	4:3
Blanking and black level	76%
White level	20% peak
Sync. level	100% peak
Video bandwidth	5.5 MHz

Field frequency	50 Hz
Line frequency	15 625 Hz
Field sync. signal	5 equalizing then 5 broad pulses, followed by 5 equalizing pulses in 7.5 line periods
Field sync. and flyback intervals	2 × 25 line periods
Line period (approx.)	64 μs
Line syn. pulses (approx.)	4.7 μs
Line blanking (approx.)	12 μs
Field sync. pules (broad)	27.3 μs
Field sync. pulses (equalizing)	2.3 μs
Colour subcarrier frequency	4.43361875 MHz
Burst duration	2.25 μs
Burst amplitude	equal to sync
Burst phase	180° ± 45°

24.6.4 UK 625-line television system field blanking details

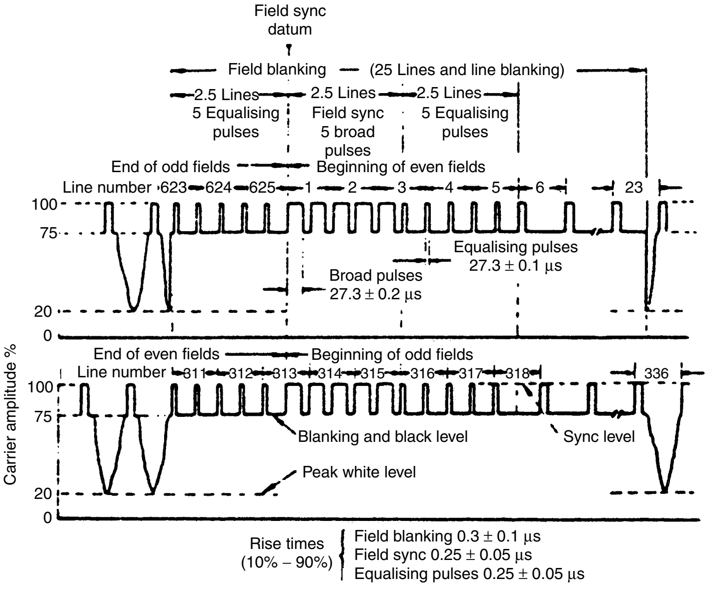

Field sync datum
Field blanking — (25 Lines and line blanking)
2.5 Lines
5 Equalising pulses
2.5 Lines
Field sync 5 broad pulses
2.5 Lines
5 Equalising pulses
End of odd fields
Beginning of even fields
Line number 623 624 625 1 2 3 4 5 6 23
100
75
20
0
Equalising pulses 27.3 ± 0.1 μs
Broad pulses 27.3 ± 0.2 μs
Carrier amplitude %
End of even fields
Beginning of odd fields
Line number 311 312 313 314 315 316 317 318 336
Blanking and black level
Sync level
Peak white level
Rise times (10% – 90%)
Field blanking 0.3 ± 0.1 μs
Field sync 0.25 ± 0.05 μs
Equalising pulses 0.25 ± 0.05 μs

24.7 Terrestrial television channels

UK

	Frequency (MHz)			Frequency (MHz)	
Channel	Vision	Sound	Channel	Vision	Sound
39	615.25	621.25	54	735.25	741.25
40	623.25	629.25	55	743.25	749.25
41	631.25	637.25	56	751.25	757.25
42	639.25	645.25	57	759.25	765.25
43	647.25	653.25	58	767.25	773.25
44	655.25	661.25	59	775.25	781.25
45	663.25	669.25	60	783.25	789.25
46	671.25	677.25	61	791.25	797.25
47	679.25	685.25	62	799.25	805.25
48	687.25	693.25	63	807.25	813.25
49	695.25	701.25	64	815.25	821.25
50	703.25	709.25	65	823.25	829.25
51	711.25	717.25	66	831.25	837.25
52	719.25	725.25	67	839.25	845.25
53	727.25	733.25	68	847.25	853.25

Republic of Ireland

	Frequency (MHz)			Frequency (MHz)	
Channel	Vision	Sound	Channel	Vision	Sound
IA	45.75	51.75	IF	191.25	197.25
IB	53.75	59.75	IG	199.25	205.25
IC	61.75	67.75	IH	207.25	213.25
ID	175.25	81.25	IJ	215.25	221.25
IE	183.25	189.25			

South Africa

	Frequency (MHz)			Frequency (MHz)	
Channel	Vision	Sound	Channel	Vision	Sound
4	175.25	181.25	9	215.25	221.25
5	183.25	189.25	10	223.25	229.25
6	191.25	197.25	11	231.25	237.25
7	199.25	205.25	13	247.43	253.43
8	207.25	213.25			

Australia

	Frequency (MHz)			Frequency (MHz)	
Channel	Vision	Sound	Channel	Vision	Sound
0	46.25	51.75	6	175.25	180.75
1	57.25	62.75	7	182.25	187.75
2	64.25	69.75	8	189.25	194.75
3	86.25	91.75	9	196.25	201.75
4	95.25	100.75	10	209.25	214.75
5	102.25	107.75	11	216.25	221.75
5A	138.25	143.75			

New Zealand

	Frequency (MHz)			Frequency (MHz)	
Channel	Vision	Sound	Channel	Vision	Sound
1	45.25	50.75	6	189.25	194.75
2	55.25	60.75	7	196.25	201.75
3	62.25	67.75	8	203.25	208.75
4	175.25	180.75	9	210.25	215.75
5	182.25	187.75			

USA

	Frequency (MHz)			Frequency (MHz)	
Channel	*Vision*	*Sound*	*Channel*	*Vision*	*Sound*
2	55.25	59.75	43	645.25	649.75
3	61.25	65.75	44	651.25	655.75
4	67.25	71.75	45	657.25	661.75
5	77.25	81.75	46	663.25	667.75
6	83.25	87.75	47	669.25	673.75
7	175.25	179.75	48	675.25	679.75
8	181.25	185.75	49	681.25	685.75
9	187.25	191.75	50	687.25	691.75
10	193.25	197.75	51	693.25	697.75
11	199.25	203.75	52	699.25	703.75
12	205.25	209.75	53	705.25	709.75
13	211.25	215.75	54	711.25	715.75
14	471.25	475.75	55	717.25	721.75
15	477.25	481.75	56	723.25	727.75
16	483.25	487.75	57	729.25	733.75
17	489.25	493.75	58	735.25	739.75
18	495.25	499.75	59	741.25	745.75
19	501.25	505.75	60	747.25	751.75
20	507.25	511.75	61	753.25	757.75
21	513.25	517.75	62	759.25	763.75
22	519.25	523.75	63	765.25	769.75
23	525.25	529.75	64	771.25	775.75
24	531.25	535.75	65	777.25	781.75
25	537.25	541.75	66	783.25	787.75
26	543.25	547.75	67	789.25	793.75
27	549.25	553.75	68	795.25	799.75
28	555.25	559.75	69	801.25	805.75
29	561.25	565.75	70	807.25	811.75
30	567.25	571.75	71	813.25	817.75
31	573.25	577.75	72	819.25	823.75
32	579.25	583.75	73	825.25	829.75
33	585.25	589.75	74	831.25	835.75
34	591.25	595.75	75	837.25	841.75
35	597.25	601.75	76	843.25	847.75
36	603.25	607.75	77	849.25	853.75
37	609.25	613.75	78	855.25	859.75
38	615.25	619.75	79	861.25	865.75
39	621.25	625.75	80	867.25	871.75
40	627.25	631.75	81	873.25	877.75
41	633.25	637.75	82	879.25	883.75
42	639.25	643.75	83	885.25	889.75

24.8 Terrestrial television aerial dimensions

Channel	*Dimensions in cm*													
	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*a*	*b*	*c*	Channels covered in the UHF groups are:		
UHF Groups												*Group letter*	*Colour Code*	*Channels*
A	30.1	30	24.1	23	22.8	21.1	20.4	19.9	10.3	10.3	1.8	A	Red	21–34
B	26.5	21.7	18.9	18	17.8	16.5	16	15.5	8.9	8.9	1.8	B	Yellow	39–51
C	23.2	18.2	16	15.3	15	14	13.3	12.2	7.5	7.5	1.8	C	Green	50–66
D	26.1	23.5	18.4	16	15.5	14.8	13.8	13	7.6	7.6	1.8	D	Blue	49–68
E	27	26.5	21.1	18.6	17.9	17.6	16	15.8	15.8	15.8	1.8	E	Brown	39–68

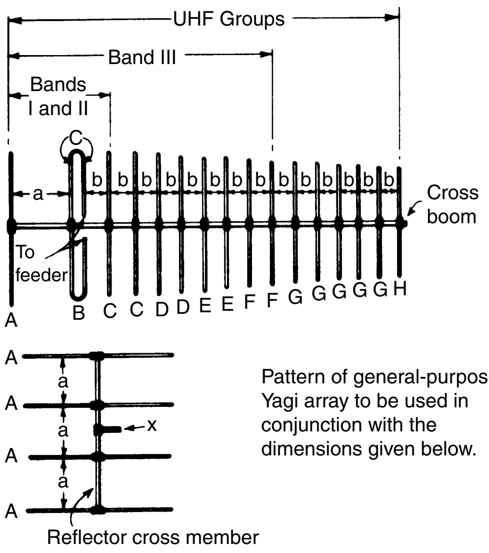

Pattern of general-purpose Yagi array to be used in conjunction with the dimensions given below.

24.9 AM broadcast station classes (USA)

The US AM broadcast band is 540 kHz to 1700 kHz, with 10 kHz channel spacings with centre frequencies divisible by ten (e.g. 780 kHz or 1540 kHz). Other countries in the western hemisphere operate on either 10 kHz channel spacings with frequencies ending in ‘5’, or 9 kHz spacing. The Domestic Class is generally the class of station defined in 47 CFR Section 73.21. The Region 2 Class is generally the class of station as defined in the Region 2 [Western Hemisphere] AM Agreement. This class also corresponds to the class in the 1984 US–Canadian AM Agreement and the 1986 US–Mexican Agreement.

24.9.1 Class A station

A Class A station is an unlimited time station (that is, it can broadcast 24 hours per day) that operates on a clear channel. The operating power shall not be less than 10 kilowatts (kW) or more than 50 kW.

24.9.2 Class B station

A Class B station is an unlimited time station. Class B stations are authorized to operate with a minimum power of 250 watts and a maximum power of 50 kW. (If a Class B station operates with less than 250 W, the RMS must be equal to or greater than 141 mV/m at 1 km for the actual power.) If the station is authorized to operate in the expanded band (1610 to 1700 kHz), the maximum power is 10 kW.

24.9.3 Class C station

A Class C station is an unlimited time station that operates on a local channel. The power shall not be less than 250 W nor more than 1 kW. Class C stations that are licensed to operate with 100 W may continue to operate as licensed.

24.9.4 Class D station

A Class D station operates either daytime, limited time, or unlimited time with a night-time power less than 250 W and an equivalent RMS antenna field less than 141 mV/m at 1 km for the actual power. Class D stations shall operate with daytime powers not less than 0.250 kW nor more than 50 kW. *Note*: If a station is an existing daytime-only station, its class will be Class D.

24.10 FM broadcast frequencies and channel numbers (USA)

From US Code USC 47 CFR 73. The FM broadcast band consists of that portion of the radio frequency spectrum between 88 MHz and 108 MHz. It is divided into 100 channels of 200 kHz each. For convenience, the frequencies available for FM broadcasting (including those assigned to non-commercial educational broadcasting) are given numerical designations which are shown in the table below:

Frequency (MHz)	*Channel No.*
88.1	201
88.3	202
88.5	203
88.7	204
88.9	205
89.1	206
89.3	207
89.5	208
89.7	209
89.9	210
90.1	211
90.3	212
90.5	213

Frequency (MHz)	*Channel No.*
90.7	214
90.9	215
91.1	216
91.3	217
91.5	218
91.7	219
91.9	220
92.1	221
92.3	222
92.5	223
92.7	224
92.9	225
93.1	226
93.3	227
93.5	228
93.7	229
93.9	230
94.1	231
94.3	232
94.5	233
94.7	234
94.9	235
95.1	236
95.3	237
95.5	238
95.7	239
95.9	240
96.1	241
96.3	242
96.5	243
96.7	244
96.9	245
97.1	246
97.3	247
97.5	248
97.7	249
97.9	250
98.1	251
98.3	252
98.5	253

Frequency (MHz)	*Channel No.*
98.7	254
98.9	255
99.1	256
99.3	257
99.5	258
99.7	259
99.9	260
100.1	261
100.3	262
100.5	263
100.7	264
100.9	265
101.1	266
101.3	267
101.5	268
101.7	269
101.9	270
102.1	271
102.3	272
102.5	273
102.7	274
102.9	275
103.1	276
103.3	277
103.5	278
103.7	279
103.9	280
104.1	281
104.3	282
104.5	283
104.7	284
104.9	285
105.1	286
105.3	287
105.5	288
105.7	289
105.9	290
106.1	291
106.3	292
106.5	293

Frequency (MHz)	*Channel No.*
106.7	294
106.9	295
107.1	296
107.3	297
107.5	298
107.7	299
107.9	300

24.11 US television channel assignments

Channel No.	*Frequency (MHz)*
1	(No Ch. 1 assigned)
2	54–60
3	60–66
4	66–72
5	76–82
6	82–88
7	174–180
8	180–186
9	186–192
10	192–198
11	198–204
12	204–210
13	210–216
14	470–476
15	476–482
16	482–488
17	488–494
18	494–500
19	500–506
20	506–512
21	512–518
22	518–524
23	524–530
24	530–536
25	536–542
26	542–548
27	548–554

Channel No.	*Frequency (MHz)*
28	554–560
29	560–566
30	566–572
31	572–578
32	578–584
33	584–590
34	590–596
35	596–602
36	602–608
37	608–614
38	614–620
39	620–626
40	626–632
41	632–638
42	638–644
43	644–650
44	650–656
45	656–662
46	662–668
47	668–674
48	674–680
49	680–686
50	686–692
51	692–698
52	698–704
53	704–710
54	710–716
55	716–722
56	722–728
57	728–734
58	734–740
59	740–746
60	746–752
61	752–758
62	758–764
63	764–770
64	770–776
65	776–782
66	782–788
67	788–794

Channel No.	*Frequency (MHz)*
68	794–800
69	800–806

Notes:

1. In Alaska, television broadcast stations operating on Channel 5 (76–82 MHz) and on Channel 6 (82–88 MHz) shall not cause harmful interference to and must accept interference from non-Government fixed operations authorized prior to 1 January 1982.
2. Channel 37, 608–614 MHz is reserved exclusively for the radio astronomy service.
3. In Hawaii, the frequency band 488–494 MHz is allocated for non-broadcast use. This frequency band (Channel 17) will not be assigned in Hawaii for use by television broadcast stations.

24.12 License-free bands

In Europe there are a number of license-free bands that can be used by anyone. However, there are restrictions on the use of these bands, both in terms of the application and the transmitted power. In particular, the manufacturer of transmitters operating in these bands must certify the equipment as meeting the required emission limits.

49.82 to 49.98 MHz – general purpose
173.2 to 173.35 MHz – telemetry. Note: 173.225 MHz is for short range alarms only
433.05 to 434.79 MHz – telemetry and vehicle security
458.5 to 458.95 MHz – industrial or commercial telemetry
868 to 870 MHz – general purpose.

Worldwide agreement has resulted in two bands, at 2.4 GHz and 5 GHz, being allocated to wireless LANs. Both of these bands are for industrial, scientific and medical (ISM) use.

The 2.4 GHz allocation covers 2.4 to 2.4835 GHz. Restrictions on the use of this ISM band include the use of carrier frequency hopping in step multiples of 1 MHz. This band is used by wireless LANs operating under IEEE802.11B and Bluetooth.

The IEEE802.11B standard uses 5 MHz carrier frequency steps, with some international variations. Europe allows the use of 13 channels (channel 1 to 13, with centre frequencies 2412 MHz to 2472 MHz), but the USA only allows the use of 11 channels (channel 1 to 11, with centre frequencies 2412 MHz to 2462 MHz). In Japan only channel 14 is allowed (no hopping), using a carrier centre frequency of 2477 MHz.

Bluetooth uses 79 hop frequencies, with multiples of a 1 MHz step size. The lowest carrier frequency is at 2402 MHz (channel 0) and the highest is 2480 MHz (channel 78). The frequency hopping pattern is psuedo-random, with 1600 hops per second.

The 5 GHz allocation is actually two sub-bands, one covering 5.15 to 5.35 GHz and the other covering 5.47 GHz to 5.725 GHz. The higher frequency ISM band allows 1 watt radiated power, rather than the 200 mW limit of the lower frequency band. As with the 2.4 GHz ISM band, sources in the 5 GHz band must use carrier frequency hopping with 5 MHz steps.

24.13 Calculating radio antenna great circle bearings

Aiming radio antennas to target a particular area of the world requires calculation of the *great circle bearing* between your location and the other stations' location. That bearing is calculated from some simple spherical trigonometry using a hand-held calculator or a computer program. Before talking about the maths, however, we need to establish a frame of reference that makes the system work.

24.13.1 Latitude and longitude

The need for navigation on the surface of the Earth caused the creation of a grid system uniquely to locate points on the surface of our globe. *Longitude* lines run from the north pole to the south pole, i.e. from north to south.

The reference point (longitude zero), called the *prime meridian*, runs through Greenwich, England. The longitude of the prime meridian is 0 degrees. Longitudes west of the prime meridian are given a plus sign (+), while longitudes east of the prime are given a minus (−) sign. If you continue the prime meridian through the poles to the other side of the Earth it has a longitude of 180 degrees. Thus, the longitude values run from −180 degrees to +180 degrees, with ±180 degrees being the same line.

The observatory at Greenwich is also the point against which relative time is measured. Every 15 degree change of longitude is equivalent to a one hour difference with the Greenwich time. To the west, subtract one hour for each 15 degrees and to the east add one

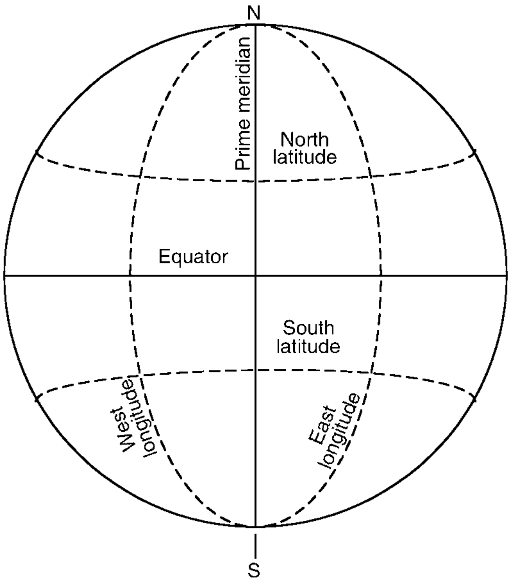

Figure 24.1 Lines of longitude and latitude

hour for each 15 degrees. Thus, the time on the East Coast of the United States is −5 hours relative to Greenwich time. At one time, we called time along the prime meridian *Greenwich mean time* (GMT), also called *Zulu time* to simplify matters for CW operators.

Latitude lines are measured against the equator, with distances north of the equator being taken as positive, and distances south of the equator being negative. The equator is 0 degrees latitude, while the north pole is +90 degrees latitude and the south pole is −90 degrees latitude.

Long ago navigators learned that the latitude can be measured by ‘shooting’ the stars and consulting a special atlas to compare the angle of certain stars with tables that translate to latitude numbers. The longitude measurement, however, is a bit different. For centuries sailors could measure latitude, but had to guess longitude (often with tragic results). In the early eighteenth century, the British government offered a large cash prize to anyone who could design a chronometer that could be taken to sea. By keeping the chronometer set accurately to Greenwich mean time, and comparing GMT against local time (i.e. at a time like high noon when the position of the sun is easy to judge), the longitude could be calculated. If you are interested in this subject, then most decent libraries have books on celestial navigation.

24.13.2 The great circle

On the surface of a globe, a curved line called a *great circle path* is the shortest distance between two points.

Consider two points on a globe: 'A' is your location, while 'B' is the other station's location. The distance 'D' is the great circle path between 'A' and 'B'.

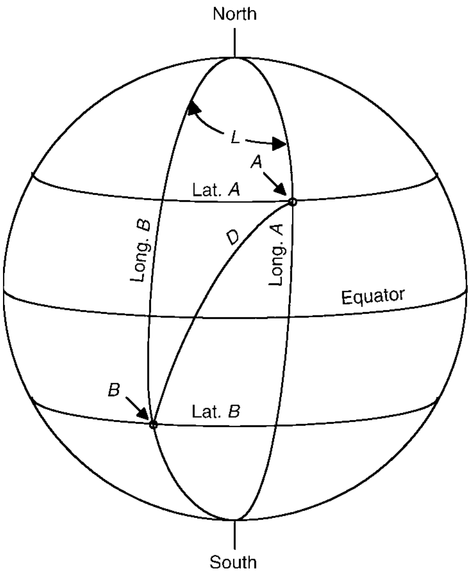

Figure 24.2 Great circle path

The great circle path length can be expressed in either degrees or distance (e.g. miles, nautical miles or kilometres). To calculate the distance, it is necessary to find the difference in longitude (L) between your longitude (LA) and the other station's longitude (LB): $L = LA - LB$. Keep the signs straight. For example, if your longitude (LA) is 40 degrees, and the other station's longitude (LB) is -120 degrees, then $L = 40 - (-120) = 40 + 120 = 160$. The equation for distance (D) is:

$$\cos D = (\sin A \times \sin B) + (\cos A \times \cos B \times \cos L)$$

where:

D is the angular great circle distance
A is your latitude
B is the other station's latitude.

To find the actual angle, take the arccos of the above equation, i.e.

$$D = \arccos(\cos D)$$

In the next equation you will want to use D in angular measure, but later on will want to convert D to miles. To do this, multiply D in degrees by 69.4. Or, if you prefer metric measures, then $D \times 111.2$ yields kilometres. This is the approximate distance in statute miles between ‘A’ and ‘B’.

To find the bearing from true north, work the equation below:

$$C = \arccos\left[\frac{\sin B - (\sin A \times \cos D)}{(\cos A \times \sin D)}\right]$$

However, this equation won’t always give you the right answer unless you make some corrections.

The first problem is the ‘same longitude error’, i.e. when both stations are on the same longitude line. In this case, $L = LA - LB = 0$. If LAT $A >$ LAT B, then $C = 180$ degrees, but if LAT $A <$ LAT B, then $C = 0$ degrees. If LAT $A =$ LAT B, then what’s the point of all these calculations?

The next problem is found when the condition $-180^\circ \leq L \leq +180^\circ$ is not met, i.e. when the absolute value of L is greater than 180°, $\text{ABS}(L) > 180^\circ$. In this case, either add or subtract 360 in order to make the value between $\pm 180^\circ$:

$$\text{If } L > +180, \text{ then } L = L - 360$$

$$\text{If } L < -180, \text{ then } L = L + 360$$

One problem seen while calculating these values on a computer or hand calculator is the fact that the $\sin(X)$ and $\cos(X)$ cover different ranges. The $\sin(X)$ function returns values from 0° to 360°, while the $\cos(X)$ function returns values only over 0° to 180°. If L is positive, then the result bearing C is accurate, but if L is negative then the actual value of $C = 360 - C$. The following test is necessary:

```
If $L < 0$ then
$L = 360 - L$
Else $L = L$
End if
```

Another problem is seen whenever either station is in a high latitude near either pole (±90°), or where both locations are very close together, or where the two locations are antipodal (i.e. on opposite points on the Earth's surface). The best way to handle these problems is to use a different equation that multiplies by the cosecant of D (i.e. cosec(D)), rather than dividing by sine of D (i.e. sin(D)).

25 Abbreviations and symbols

25.1 Abbreviations

Many abbreviations are found as either capital or lower case letters, depending on publishers' styles. Symbols should generally be standard, as shown.

A	Ampere or anode
ABR	Auxiliary bass radiator
a.c.	Alternating current
Ack	Acknowledgement
A/D	Analogue to digital
ADC	Analogue to digital converter
Ae	Aerial
a.f.	Audio frequency
a.f.c.	Automatic frequency control
a.g.c.	Automatic gain control
a.m.	Amplitude modulation
AMPS	Advanced mobile phone system
ANSI	American National Standards Institute
ASA	Acoustical Society of America
ASCII	American Standard Code for Information Interchange
a.t.u.	Aerial tuning unit
AUX	Auxiliary
a.v.c.	Automatic volume control
AWG	American Wine Gauge
b	Base of transistor
BAF	Bonded acetate fibre
B & S	Brown & Sharpe (U.S.) wire gauge
b.p.s.	Bits per second
BR	Bass reflex
BSI	British Standards Institution
BW	Bandwidth
C	Capacitor, cathode, centigrade, coulomb
c	Collector of transistor, speed of light
CB	Citizen's band
CCD	Charge coupled device
CCIR	International Radio Consultative Committee

CCITT	International Telegraph and Telephone Consultative Committee
CCTV	Closed circuit television
CDMA	Code division multiple access
chps	Characters per second
CISPR	Comité International Special Des Peturbations (radio interference standards body)
CLK	Clock signal
CMOS	Complementary metal oxide semiconductor
CrO_2	Chromium dioxide
CPU	Central processor unit
CTCSS	Continuous tone controlled signalled system
CTD	Charge transfer device
c.w.	Continuous wave
D	Diode
d	Drain of an f.e.t.
D/A	Digital to analogue
DAC	Digital to analogue converter
dB	Decibel
d.c.	Direct current
DCC	Double cotton covered
DCE	Data circuit-terminating equipment
DCS	Digital Communications Service
DF	Direction finding
DIL	Dual-in-line
DIN	German standards institute
DMA	Direct memory access
DPDT	Double pole, double throw
DPSK	Differential phase shift keying
DPST	Double pole, single throw
dsb	or dsbam. Double sideband amplitude modulation
DSRR	Digital short range radio
DTE	Data terminal equipment
DTL	Diode-transistor logic
DTMF	Dual tone multi-frequency
DX	Long distance reception
e	Emitter of transistor
EAROM	Electrically alterable read only memory
ECL	Emitter coupled logic
EDGE	Enhanced Data for GSM Evolution
e.h.t.	Extremely high tension (voltage)

EIRP	Effective Isotropic Radiated Power
e.m.f.	Electromotive force
en	Enamelled
EPROM	Erasable programmable read only memory
EQ	Equalization
EROM	Erasable read only memory
ERP	Effective radiated power
ETACS	Extended total access communications system
F	Farad, fahrenheit or force
f	Frequency
FDD	Frequency Division Duplex
FDM	Frequency division multiplex
FDMA	Frequency division multiple access
Fe	Ferrous
FeCr	Ferri-chrome
f.e.t.	Field effect transistor
FFSK	Fast frequency shift keying
f.m.	Frequency modulation
f.r.	Frequency response or range
f.s.d.	Full-scale deflection
FSK	Frequency shift keying
G	Giga (10^9)
g	Grid, gravitational constant
GMSK	Gaussian minimum shift keying
GSM	Global system mobile
H	Henry
h.f.	High frequency
Hz	Hertz (cycles per second)
I	Current
IB	Infinite baffle
i.c.	Integrated circuit
IF	Intermediate frequency
IHF	Institute of High Fidelity (U.S.)
I^2L(HL)	Integrated injection logic
i.m.d.	Intermodulation distortion
i/p	Input
i.p.s.	Inches per second
K	Kilo, in computing terms ($= 2^{10} = 1024$), or degrees Kelvin
k	Kilo (10^3) or cathode
L	Inductance or lumens
l.e.d.	Light emitting diode

l.f.	Low frequency
LIN	Linear
LOG	Logarithmic
LS	Loudspeaker
LSI	Large scale integration
l.w.	Long wave (approx. 1100–2000 m)
M	Mega (10^6)
m	Milli (10^{-3}) or metres
m.c.	Moving coil
MHz	Megahertz
mic	Microphone
MOS	Metal oxide semiconductor
MPU	Microprocessor unit
MPX	Multiplex
MSC	Mobile Switching Centre
MSK	Minimum shift keying
m.w.	Medium wave (approx. 185–560 m)
n	Nano (10^9)
NAB	National Association of Broadcasters
Ni-Cad	Nickel-cadmium
n/c	Not connected; normally closed
n/o	Normally open
NMOS	Negative channel metal oxide semiconductor
NRZ	Non-return to zero
o/c	Open channel; open circuit
o/p	Output
op-amp	Operational amplifier
p	Pico (10^{-12})
PA	Public address
PABX	Private automatic branch exchange
PAL	Phase alternation, line
p.a.m.	Pulse amplitude modulation
PCB	Printed circuit board
PCM	Pulse code modulation
PCN	Personal communications network
PCS	Personal communication system
PLA	Programmable logic array
PLL	Phase locked loop
pm	Phase modulation
PMOS	Positive channel metal oxide semiconductor
P.P.M.	Peak programme meter
p.r.f.	Pulse repetition frequency

PROM	Programmable read only memory
PSK	Phase shift keying
PSS	Packet SwitchStream
PSTN	Public Switched Telephone Network
PSU	Power supply unit
PTFE	Polytetrafluoroethylene
PU	Pickup
PUJT	Programmable unijunction transistor
Q	Quality factor; efficiency of tuned circuit, charge
QAM	Quadrature (or quaternary) amplitude modulation
QPSK	Quadrature (or quaternary) phase shift keying
R	Resistance
RAM	Random access memory
RCF	Recommended crossover frequency
r.f.	Radio frequency
r.f.c.	Radio frequency choke (coil)
RFI	Radio frequency interference
RIAA	Record Industry Association of America
r.m.s.	Root mean square
ROM	Read only memory
RTL	Resistor transistor logic
R/W	Read/write
RX	Receiver
S	Siemens
s	Source of an f.e.t.
s/c	Short circuit
SCR	Silicon-controlled rectifier
s.h.f.	Super high frequency
SI	International system of units
S/N	Signal-to-noise.
SPL	Sound pressure level
SPDT	Single pole, double throw
SPST	Single pole, single throw
ssb	Single sideband amplitude modulation
ssbdc	Single sideband diminished carrier
ssbsc	Single sideband suppressed carrier
SSI	Small scale integration
s.w.	Short wave (approx. 10–60 m)
s.w.g.	Standard wire gauge
s.w.r.	Standing wave ratio
T	Tesla

TACS	Total access communications system
TDD	Time division duplex
TDM	Time division multiplex
TDMA	Time division multiple access
t.h.d.	Total harmonic distortion
t.i.d.	Transient intermodulation distortion
TR	Transformer
t.r.f.	Tuned radio frequency
TRS	Transmitter repeater station
TTL	Transistor transistor logic
TTY	Teletype unit
TVI	Television interface; television interference
TX	Transmitter
UART	Universal asynchronous receiver transmitter
u.h.f.	Ultra high frequency (approx. 470–854 MHz)
u.j.t.	Unijunction transistor
ULA	Uncommitted logic array
V	Volts
VA	Volt-amps
v.c.a.	Voltage controlled amplifier
v.c.o.	Voltage controlled oscillator
VCT	Voltage to current transactor
v.h.f.	Very high frequency (approx. 88–216 MHz)
v.l.f.	Very low frequency
VSB	Vestigial side band
VSWR	Voltage standing wave ratio
VU	Volume unit
W	Watts
WAP	Wireless application protocol
Wb	Weber
W/F	Wow and flutter
WML	Wireless mark-up language
w.p.m.	Words per minute
X	Reactance
Xtal	Crystal
Z	Impedance
ZD	Zener diode

25.2 Letter symbols by unit name

Unit	*Symbol*	*Notes*
ampere	A	SI unit of electric current.
ampere (turn)	At	SI unit of magnetomotive force.
ampere-hour	Ah	
ampere per metre	$\mathrm{A\,m^{-1}}$	SI unit of magnetic field strength.
angstrom	Å	$1\ \text{Å} = 10^{-10}\,\mathrm{m}$.
apostilb	asb	$1\,\mathrm{asb}(1/\pi)\mathrm{cd\ m^{-2}}$ A unit of luminance. The SI unit, candela per square metre, is preferred.
atmosphere:		
standard atmosphere	atm	$1\,\mathrm{atm} = 101\,325\,\mathrm{N\ m^{-2}}$.
technical atmosphere	at	$1\ \mathrm{at} = 1\,\mathrm{kgf\ cm^{-2}}$.
atomic mass unit (unified)	u	The (unified) atomic mass unit is defined as one-twelfth of the mass of an atom of the ^{12}C nuclide. Use of the old atomic mass unit (amu), defined by reference to oxygen, is deprecated.
bam	b	$1\,\mathrm{b}10^{-28}\,\mathrm{m}^2$.
bar	bar	$1\,\mathrm{bar} = 100\,000\,\mathrm{N\ m^{-2}}$.
baud	Bd	Unit of signalling speed equal to one element per second.
becquerel	Bq	$1\ \mathrm{Bq} = 1\,\mathrm{s}^{-1}$. SI unit of radioactivity.
bel	B	
bit	b	
British thermal unit	Btu	

Unit	*Symbol*	*Notes*
calorie (International Table calorie)	$\mathrm{cal}_{\mathrm{IT}}$	1 cal°— = 4.1868 J. The 9th Conférence Générale des Poids et Mesures adopted the joule as the unit of heat, avoiding the use of the calorie as far as possible.
calorie (thermochemical calorie)	cal	1 cal = 4.1840 J. (See note for International Table calorie.)
candela	cd	SI unit of luminous intensity.
candela per square inch	$\mathrm{cd\ in^{-2}}$	Use of the SI unit, candela per square metre, is preferred.
candela per square metre	$\mathrm{cd\ m^{-2}}$	SI unit of luminance. The name nit has been used.
candle		The unit of luminous intensity has been given the name *candela*; use of the word *candle* for this purpose is deprecated.
centimetre	cm	
circular mil	cmil	$1\,\mathrm{cmil} = (\pi/4)10^{-6}\ \mathrm{in}^2$.
coulomb	C	SI unit of electrical charge.
cubic centimetre	cm^3	
cubic foot	ft^3	
cubic foot per minute	$\mathrm{ft^3\ min^{-1}}$	
cubic foot per second	$\mathrm{ft^3\ s^{-1}}$	
cubic inch	in^3	
cubic metre	m^3	
cubic metre per second	$\mathrm{m^3\ s^{-1}}$	
cubic yard	yd^3	
curie	Ci	Unit of activity in the field of radiation dosimetry.
cycle	c	
cycle per second	$\mathrm{c\,s^{-1}}$	Deprecated. Use hertz.
decibel	dB	
degree (plane angle)	°	

Unit	*Symbol*	*Notes*
degree (temperature):		Note that there is no space between the symbol ° and the letter. The use of the word *centigrade* of the Celsius temperature scale was abandoned by the Conférence Générale des Poids et Mesures in 1948.
degree Celsius	°C	
degree Fahrenheit	°F	
degree Kelvin		See Kelvin.
degree Rankine	°R	
dyne	dyn	
electronvolt	eV	
erg	erg	
erlang	E	Unit of telephone traffic.
farad	F	SI unit of capacitance.
foot	ft	
footcandle	fc	Use of the SI unit of illuminance, the lux (lumen per square metre), is preferred.
footlambert	fL	Use of the SI unit, the candela per square metre, is preferred.
foot per minute	$\mathrm{ft\ min^{-1}}$	
foot per second	$\mathrm{ft\ s^{-1}}$	
foot per second squared	$\mathrm{ft\ s^{-2}}$	
foot pound-force	$\mathrm{ft\ lb_f}$	
gal	Gal	$1\,\mathrm{Gal} = 1\,\mathrm{cm\ s^{-2}}$.
gallon	gal	The gallon, quart, and pint differ in the US and the UK, and their use is deprecated.
gauss	G	The gauss is the electromagnetic CGS (Centimetre Gram Second) unit of magnetic flux density. The SI unit, tesla, is preferred.
gigaelectronvolt	GeV	

Unit	*Symbol*	*Notes*
gigahertz	GHz	
gilbert	Gb	The gilbert is the electromagnetic CGS (Centimetre Gram Second) unit of magnetomotive force. Use of the SI unit, the ampere (or ampere-turn), is preferred.
grain	gr	
gram	g	
gray	Gy	$1\ \mathrm{Gy} = 1\,\mathrm{J\ kg}^{-1}$. SI unit of absorbed dose.
henry	H	
hertz	Hz	SI unit of frequency.
horsepower	hp	Use of the SI unit, the watt, is preferred.
hour	h	Time may be designated as in the following example; $9^{h}46^{m}30^{s}$.
inch	in	
inch per second	$\mathrm{in\ s}^{-1}$	
joule	J	SI unit of energy.
joule per Kelvin	$\mathrm{J\,K}^{-1}$	SI unit of heat capacity and entropy.
Kelvin	K	SI unit of temperature (formerly called *degree Kelvin*). The symbol K is now used without the symbol°.
kiloelectronvolt	KeV	
kilogauss	kG	
kilogram	kg	SI unit of mass.
kilogram-force	kg_{f}	In some countries the name *kilopond* (kp) has been adopted for this unit.
kilohertz	kHz	
kilohm	kΩ	
kilojoule	kJ	
kilometre	km	

Unit	*Symbol*	*Notes*
kilometre per hour	km h^{-1}	
kilopond	kp	See kilogram-force.
kilovar	kvar	
kilovolt	kV	
kilovoltampere	kVA	
kilowatt	kW	
kilowatthour	kWh	
knot	kn	1 kn = 1 nmi h^{-1}.
lambert	L	The lambert is the CGS (Centimetre Gram Second) unit of luminance. The SI unit, candela per square metre, is preferred.
litre	l	
litre per second	l s^{-1}	
lumen	lm	SI unit of luminous flux.
lumen per square foot	lm ft^{-2}	Use of the SI unit, the lumen per square metre, is preferred.
lumen per square metre	lm m^{-2}	SI unit of luminous excitance.
lumen per watt	lm W^{-1}	SI unit of luminous efficacy.
lumen second	lm s	SI unit of quantity of light.
lux	lx	1 lx = 1 lm m^{-2}. SI unit of illuminance.
maxwell	Mx	The maxwell is the electromagnetic CGS (Centimetre Gram Second) unit of magnetic flux. Use of the SI unit, the weber, is preferred.
megaelectronvolt	MeV	
megahertz	MHz	
megavolt	MV	
megawatt	MW	
megohm	MΩ	
metre	m	SI unit of length.
mho	mho	1 mho = 1 Ω^{-1} = 1 S.
microampere	μA	
microbar	μbar	

Unit	*Symbol*	*Notes*
microfarad	μF	
microgram	μg	
microhenry	μH	
micrometre	μm	
micron		The name *micrometre* (μm) is preferred.
microsecond	μS	
microwatt	μW	
mil	mil	1 mil = −0.001 in.
mile		
nautical	nmi	
statute	mi	
mile per hour	mi h^{-1}	
milliampere	mA	
millibar	mbar	mb may be used.
milligal	mGal	
milligram	mg	
millihenry	mH	
millilitre	ml	
millimetre	mm	
conventional millimetre of mercury	mm Hg	1 mm Hg = 133.322 N m^{-2}.
millimicron		The name *nanometre* (nm) is preferred.
millisecond	ms	
millivolt	mV	
milliwatt	mW	
minute (plane angle)	$\ldots'$	
minute (time)	min	Time may be designated as in the following example: $9^h46^m30^s$.
mole	mol	SI unit of amount of substance.
nanoampere	nA	
nanofarad	nF	
nanometre	nm	
nanosecond	ns	
nanowatt	nW	
nautical mile	nmi	

Unit	*Symbol*	*Notes*
neper	Np	
newton	N	SI unit of force.
newton metre	Nm	
newton per square metre	$\mathrm{N\,m^{-2}}$	See pascal.
nit	nt	$1\,\mathrm{nt} = 1\,\mathrm{cd\ m^{-2}}$. See candela per square metre.
oersted	Oe	The oersted is the electromagnetic CGS (Centimetre Gram Second) unit of magnetic field strength. Use of the SI unit, the ampere per metre, is preferred.
ohm	Ω	SI unit of electrical resistance.
ounce (avoirdupois)	oz	
pascal	Pa	SI unit of pressure or stress. $1\,\mathrm{Pa} = 1\,\mathrm{N\ m^{-2}}$.
picoampere	pA	
picofarad	pF	
picosecond	ps	
picowatt	pW	
pint	pt	The gallon, quart, and pint differ in the US and the UK, and their use is deprecated.
pound	lb	
poundal	pdl	
pound-force	$\mathrm{lb_f}$	
pound-force feet	$\mathrm{lb_f\ ft}$	
pound-force per square inch	$\mathrm{lb_f\ in^{-2}}$	
pound per square inch		Although use of the abbreviation psi is common, it is not recommended. See pound-force per square inch.

Unit	*Symbol*	*Notes*
Quart	qt	The gallon, quart, and pint differ in the US and the UK, and their use is deprecated.
rad	rd	Unit of absorbed dose in the field of radiation dosimetry.
revolution per minute	$\mathrm{r\ min^{-1}}$	Although use of the abbreviation rpm is common, it is not recommended.
revolution per second	$\mathrm{r\ s^{-1}}$	
roentgen	R	Unit of exposure in the field of radiation dosimetry.
second (plane angle)	$\ldots''$	
second (time)	s	SI unit of time. Time may be designated as in the following example: $9^{\mathrm{h}}46^{\mathrm{m}}30^{\mathrm{s}}$.
siemens	S	SI unit of conductance. $1\,\mathrm{S} = 1\ \Omega^{-1}$.
square foot	ft^2	
square inch	in^2	
square metre	m^2	
square yard	yd^2	
steradian	sr	SI unit of solid angle.
stilb	sb	$1\,\mathrm{sb} = 1\,\mathrm{cd\ cm^{-2}}$. A CGS unit of luminance. Use of the SI unit, the candela per square metre, is preferred.
tesla	T	SI unit of magnetic flux density. $1\,\mathrm{T} = 1\,\mathrm{Wb\ m^{-2}}$.
tonne	t	$1\,\mathrm{t} = 1000\,\mathrm{kg}$.
(unified) atomic mass unit	u	See atomic mass unit (unified).
var	var	Unit of reactive power.
volt	V	SI unit of electromotive force.
voltampere	VA	SI unit of apparent power.
watt	W	SI unit of power.
watthour	Wh	

Unit	*Symbol*	*Notes*
watt per steradian	W sr^{-1}	SI unit of radiant intensity.
watt per steradian square metre	W (sr $m^2)^{-1}$	SI unit of radiance.
weber	Wb	SI unit of magnetic flux. 1 Wb = 1 V s.
yard	yd	

25.3 Electric quantities

Quantity	*Symbol*	*Unit*	*Symbol*
Admittance	Y	siemens	S
Angular frequency	ω	hertz	Hz
Apparent power	S	watt	W
Capacitance	C	farad	F
Charge	Q	coulomb	C
Charge density	ρ	coulomb per square metre	$C\,m^{-2}$
Conductance	G	siemens	S
Conductivity	κ, γ, σ	siemens per metre	$S\,m^{-1}$
Current	I	ampere	A
Current density	j, J	ampere per square metre	$A\,m^{-2}$
Displacement	D	coulomb per square metre	$C\,m^{-2}$
Electromotive force	E	volt	V
Energy	E	joule	J
Faraday constant	F	coloumb per mole	$C\,mol^{-1}$
Field strength	E	volt per metre	$V\,m^{-1}$
Flux	ψ	coulomb	C
Frequency	v, f	hertz	Hz
Impedance	Z	ohm	Ω
Light, velocity of in a vacuum	c	metre per second	$m\,s^{-1}$
Period	T	second	s
Permeability	μ	henry per metre	$H\,m^{-1}$
Permeability of space	μ_o	henry per metre	$H\,m^{-1}$
Permeance	Λ	henry	H

Quantity	*Symbol*	*Unit*	*Symbol*
Permittivity	ε	farad per metre	$\mathrm{F\,m^{-1}}$
Permittivity of space	ε_o	farad per metre	$\mathrm{F\,m^{-1}}$
Phase	ϕ	–	–
Potential	V, U	volt	V
Power	P	watt	W
Quality factor	Q	–	–
Reactance	X	ohm	Ω
Reactive power	Q	watt	W
Relative permeability	μ_r	–	–
Relative permittivity	ε_r	–	–
Relaxation time	τ	second	s
Reluctance	R	reciprocal henry	$\mathrm{H^{-1}}$
Resistance	R	ohm	Ω
Resistivity	ρ	ohm metre	Ω m
Susceptance	B	siemens	S
Thermodynamic temperature	T	kelvin	K
Time constant	τ	second	s
Wavelength	λ	metre	m

26 Miscellaneous data

26.1 Fundamental constants

Constant	*Symbol*	*Value*
Boltzmann constant	k	1.38062×10^{-23} J K^{-1}
Electron charge, proton charge	e	$\pm 1.60219 \times 10^{-19}$ C
Electron charge-to-mass ratio	e/m	1.7588×10^{11} C kg^{-1}
Electron mass	m_e	9.10956×10^{-31} kg
Electron radius	r_e	2.81794×10^{-15} m
Faraday constant	F	9.64867×10^{4} C mol^{-1}
Neutron mass	m_n	1.67492×10^{-27} kg
Permeability of space	μ_o	$4\pi \times 10^{-7}$ H m^{-1}
Permittivity of space	ε_o	8.85419×10^{-12} F m^{-1}
Planck constant	h	6.6262×10^{-34} J s
Proton mass	m_p	1.67251×10^{-27} kg
Velocity of light	c	2.99793×10^{8} m s^{-1}

26.2 Electrical relationships

Amperes × ohms = **volts**
Volts ÷ amperes = **ohms**
Volts ÷ ohms = **amperes**
Amperes × volts = **watts**
(Amperes)^2 × ohms = **watts**
(Volts)^2 ÷ ohms = **watts**
Joules per second = **watts**
Coulombs per second = **amperes**
Amperes × seconds = **coulombs**
Farads × volts = **coulombs**
Coulombs ÷ volts = **farads**
Coulombs ÷ farads = **volts**
Volts × coulombs = **joules**
Farads × (volts)^2 = **joules**

26.3 Dimensions of physical properties

Length: metre [L]. Mass: kilogram [M]. Time: second [T]. Quantity of electricity: coulomb [Q]. Area: square metre $[L^2]$. Volume: cubic metre $[L^3]$.

Velocity: metre per second	$[LT^{-1}]$
Acceleration: metre per second^2	$[LT^{-2}]$
Force: newton	$[MLT^{-2}]$
Work: joule	$[ML^2T^{-2}]$
Power: watt	$[ML^2T^{-3}]$
Electric current: ampere	$[QT^{-1}]$
Voltage: volt	$[ML^2T^{-2}Q^{-1}]$
Electric resistance: ohm	$[ML^2T^{-1}Q^{-2}]$
Electric conductance: siemens	$[M^{-1}L^{-2}TQ^2]$
Inductance: henry	$[ML^2Q^{-2}]$
Capacitance: farad	$[M^{-1}L^{-2}T^2Q^2]$
Current density: ampere per metre^2	$[L^{-2}T^{-1}Q]$
Electric field strength: volt per metre	$[MLT^{-2}Q^{-1}]$
Magnetic flux: weber	$[MLT^2T^{-1}Q^{-1}]$
Magnetic flux density: tesla	$[MT^{-1}Q^{-1}]$
Energy: joule	$[ML^2T^{-2}]$
Frequency: hertz	$[T^{-1}]$
Pressure: pascal	$[ML^{-1}T^{-2}]$

26.4 Fundamental units

Quantity	*Unit*	*Symbol*
Amount of a substance	mole	mol
Charge	coulomb	C
Length	metre	m
Luminous intensity	candela	cd
Mass	kilogram	kg
Plane angle	radian	rad
Solid angle	steradian	sr
Thermodynamic temperature	kelvin	K
Time	second	s

26.5 Greek alphabet

Capital letters	*Small letters*	*Greek name*	*English equivalent*	*Capital letters*	*Small letters*	*Greek name*	*English equivalent*
A	α	*Alpha*	a	N	ν	*Nu*	n
B	β	*Beta*	b	Ξ	ξ	*Xi*	x
Γ	γ	*Gamma*	g	O	o	*Omicron*	ŏ
Δ	δ	*Delta*	d	Π	π	*Pi*	p
E	ε	*Epsilon*	e	P	ρ	*Rho*	r
Z	ζ	*Zeta*	z	Σ	σ	*Sigma*	s
H	η	*Eta*	é	T	τ	*Tau*	t
Θ	θ	*Theta*	th	Y	υ	*Upsilon*	u
I	ι	*Iota*	i	Φ	ϕ	*Phi*	ph
K	κ	*Kappa*	k	X	χ	*Chi*	ch
Λ	λ	*Lambda*	l	Ψ	ψ	*Psi*	ps
M	μ	*Mu*	m	Ω	ω	*Omega*	ö

26.6 Standard units

Ampere Unit of electric current, the constant current which, if maintained in two straight parallel conductors of infinite length of negligible circular cross-section and placed one metre apart in a vacuum, will produce between them a force equal to 2×10^{-7} newton per metre length.

Ampere-hour Unit of quantity of electricity equal to 3 600 coulombs. One unit is represented by one ampere flowing for one hour.

Candela Unit of luminous intensity. It is the luminous intensity, in the perpendicular direction, of a surface of $1/600\,000\,\mathrm{m}^{-2}$ of a full radiator at the temperature of freezing platinum under a pressure of 101 325 newtons m^{-2}.

Coulomb Unit of electric charge, the quantity of electricity transported in one second by one ampere.

Decibel (dB) Unit of acoustical or electrical power ratio. Although the bel is officially the unit, this is usually regarded as being too large, so the decibel is preferred. The difference between two power levels is P_1 and P_2, is given as

$$10 \log_{10} \frac{P_1}{P_2} \text{ decibels}$$

Farad Unit of electric capacitance. The capacitance of a capacitor between the plates of which there appears a difference of potential of one volt when it is charged by one coulomb of electricity. Practical units are the microfarad (10^{-6} farad), the nanofarad (10^{-9} farad) and the picofarad (10^{-12} farad).

Henry Unit of electrical inductance. The inductance of a closed circuit in which an electromotive force of one volt is produced when the electric current in the circuit varies uniformly at the rate of one ampere per second. Practical units are the microhenry (10^{-6} henry) and the millihenry (10^{-3} henry).

Hertz Unit of frequency. The number of repetitions of a regular occurrence in one second.

Joule Unit of energy, including work and quantity of heat. The work done when the point of application of a force of one newton is displaced through a distance of one metre in the direction of the force.

Kilovolt-ampere 1000 volt-amperes.

Kilowatt 1000 watts.

Light, velocity of Light waves travel at 300 000 kilometres per second (approximately). Also the velocity of radio waves.

Lumen m^{-2}, lux Unit of illuminance of a surface.

Mho Unit of conductance, see Siemens.

Newton Unit of force. That force which, applied to a mass of one kilogram, gives it an acceleration of one metre per second per second.

Ohm Unit of electric resistance. The resistance between two points of a conductor when a constant difference of potential of one volt, applied between these two points, produces in the conductor a current of one ampere.

Pascal Unit of sound pressure. Pressure is usually quoted as the root mean square pressure for a pure sinusoidal wave.

Siemens Unit of conductance, the reciprocal of the ohm. A body having a resistance of 4 ohms would have a conductance of 0.25 siemens.

Sound, velocity of Sound waves travel at 332 metres per second in air (approximately) at sea level.

Tesla Unit of magnetic flux density, equal to one weber per square metre of circuit area.

Volt Unit of electric potential. The difference of electric potential between two points of a conducting wire carrying a constant current of one ampere, when the power dissipated between these points is equal to one watt.

Volt-ampere The product of the root-mean-square volts and root-mean-square amperes.

Watt Unit of power, equal to one joule per second. Volts times amperes equals watts.

Weber Unit of magnetic flux. The magnetic flux which, linking a circuit of one turn, produces in it an electromotive force of one volt as it is reduced to zero at a uniform rate in one second.

26.7 Decimal multipliers

Prefix	*Symbol*	*Multiplier*	*Prefix*	*Symbol*	*Multiplier*
tera	T	10^{12}	centi	c	10^{-2}
giga	G	10^{9}	milli	m	10^{-3}
mega	M	10^{6}	micro	μ	10^{-6}
kilo	k	10^{3}	nano	n	10^{-9}
hecto	h	10^{2}	pico	p	10^{-12}
deka	da	10	femto	f	10^{-15}
deci	d	10^{-1}	atto	a	10^{-18}

26.8 Useful formulae

Boolean Algebra (laws of)

Absorption: $A + (A.B) = A$

$A.(A + B) = A$

Annulment:
$$A + 1 = 1$$
$$A.0 = 0$$

Association:
$$(A + B) + C = A + (B + C)$$
$$(A.B).C = A.(B.C)$$

Commutation:
$$A + B = B + A$$
$$A.B = B.A$$

Complements:
$$A + \bar{A} = 1$$
$$A.\bar{A} = 0$$

De Morgan's:
$$\overline{(A + B)} = \bar{A}.\bar{B}$$
$$(\overline{A.B}) = \bar{A} + \bar{B}$$

Distributive:
$$A.(B + C) = (A.B) + (A.C)$$
$$A + (B.C) = (A + B).(A + C)$$

Double negation:
$$\bar{\bar{A}} = A$$

Identity:
$$A + O = A$$
$$A.1 = A$$

Tautology:
$$A.A = A$$
$$A + A = A$$

Capacitance

The capacitance of a parallel plate capacitor can be found from

$$C = \frac{0.885\,KA}{d}$$

C is in picofarads, K is the dielectric constant (air = 1). A is the area of the plate in square cm and d the thickness of the dielectric.

Calculation of overall capacitance with:
Parallel capacitors $-C = C_1 + C_2 + \cdots$
Series capacitors $-\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \cdots$

Characteristic impedance

$$\text{(open wire) } Z = 276 \log \frac{2D}{d} \text{ohms}$$

where

$$\left.\begin{aligned} D &= \text{wire spacing} \\ d &= \text{wire diameter} \end{aligned}\right\} \text{in same units.}$$

$$\text{(coaxial) } Z = \frac{138}{\sqrt{(K)}} \log \frac{d_o}{d_i} \text{ ohms}$$

where K = dielectric constant, d_o = outside diameter of inner conductor, d_i = inside diameter of outer conductor.

Dynamic resistance

In a parallel-tuned circuit at resonance the dynamic resistance is

$$R_d = \frac{L}{Cr} = Q\omega L = \frac{Q}{\omega C} \text{ ohms}$$

where L = inductance (henries), C = capacitance (farads), r = effective series resistance (ohms). Q = Q-value of coil, and $\omega = 2\pi \times$ frequency (hertz).

Frequency – wavelength – velocity

(See also Resonance)
The velocity of propagation of a wave is

$$v = f\lambda \text{ metres per second}$$

where f = frequency (hertz) and λ = wavelength (metres).

For electromagnetic waves in free space the velocity of propagation v is approximately 3×10^8 m/sec, and if f is expressed in kilohertz and λ in metres

$$f = \frac{300\,000}{\lambda} \text{ kilohertz} \qquad f = \frac{300}{\lambda} \text{ megahertz}$$

or

$$\lambda = \frac{300\,000}{f} \text{ metres} \qquad \lambda = \frac{300}{f} \text{ metres}$$

f in kilohertz $\qquad$ f in megahertz

Impedance

The impedance of a circuit comprising inductance, capacitance and resistance in series is

$$Z = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}$$

where R = resistance (ohms), $\omega = 2\pi \times$ frequency (hertz). L = inductance (henries), and C = capacitance (farads).

Inductance

Single layer coils

$$L(\text{in microhenries}) = \frac{a^2N^2}{9a + 10l} \text{ approximately}$$

If the desired inductance is known, the number of turns required may be determined by the formula

$$N = \frac{5L}{na^2}\left[1 + \sqrt{\left(1 + \frac{0.36n^2a^3}{L}\right)}\right]$$

where N = number of turns, a = radius of coil in inches, n = number of turns per inch. L = inductance in microhenries (μH) and l = length of coil in inches.

Calculation of overall inductance with:
Series inductors $-\mathrm{L} = \mathrm{L}_1 + \mathrm{L}_2 + \cdots$
Parallel inductors $-\frac{1}{\mathrm{L}} = \frac{1}{\mathrm{L}_1} + \frac{1}{\mathrm{L}_2} + \cdots$

Meter conversions

Increasing range of ammeters or milliammeters
Current range of meter can be increased by connecting a shunt resistance across meter terminals. If R_m is the resistance of the meter; R_s the value of the shunt resistance and n the number of times it is wished to multiply the scale reading, then

$$R_\mathrm{s} = \frac{R_\mathrm{m}}{(n-1)}$$

Increasing range of voltmeters
Voltage range of meter can be increased by connecting resistance in series with it. If this series resistance is R_s and R_m and n as before, then $R_\mathrm{s} = R_\mathrm{m} \times (n-1)$.

Negative feedback

Voltage feedback

$$\text{Gain with feedback} = \frac{A}{1 + Ab}$$

where A is the original gain of the amplifier section over which feedback is applied (including the output transformer if included) and b is the fraction of the output voltage fed back.

$$\text{Distortion with feedback} = \frac{d}{1 + Ab} \text{ approximately}$$

where d is the original distortion of the amplifier.

Ohm's Law

$$I = \frac{V}{R} V = IR \; R = \frac{V}{I}$$

where I = current (amperes), V = voltage (volts), and R = resistance (ohms).

Power

In a d.c. circuit the power developed is given by

$$W = VI = \frac{V^2}{R} = I^2 R \text{ watts}$$

where V = voltage (volts), I = current (amperes), and R = resistance (ohms).

Power ratio

$$P = 10 \log \frac{P_1}{P_2}$$

where P = ratio in decibels, P_1 and P_2 are the two power levels.

Q

The Q value of an inductance is given by

$$Q = \frac{\omega L}{R}$$

Radio horizon distance

The radio horizon at VHF/UHF and up is approximately 15% further than the optical horizon. Several equations are used in calculating the distance. If D is the distance to the radio horizon, and H is the antenna height, then:

$$D = k\sqrt{H}$$

1. When D is in statute miles (5280 feet) and H in feet, then $K = 1.42$.
2. When D is in nautical miles (6000 feet) and H in feet, then $K = 1.23$.
3. When D is in kilometres and H is in metres, then $K = 4.12$.

Reactance

The reactance of an inductor and a capacitor respectively is given by

$$X_{\mathrm{L}} = \omega L \text{ ohms} \qquad X_{\mathrm{C}} = \frac{1}{\omega C} \text{ ohms}$$

where $\omega = 2\pi \times \text{frequency}$ (hertz), $L = \text{inductance}$ (henries), and $C = \text{capacitance}$ (farads).

The total resistance of an inductance and a capacitance in series is $X_{\mathrm{L}} - X_{\mathrm{C}}$.

Resistance

Calculation of overall resistance with:

Series resistors $- \mathrm{R} = \mathrm{R}_1 + \mathrm{R}_2 + \cdots$

Parallel resistors $- \dfrac{1}{\mathrm{R}} = \dfrac{1}{\mathrm{R}_1} + \dfrac{1}{\mathrm{R}_2} + \cdots$

Resonance

The resonant frequency of a tuned circuit is given by

$$f = \frac{1}{2\pi\sqrt{LC}} \text{ hertz}$$

where $L = \text{inductance}$ (henries), and $C = \text{capacitance}$ (farads). If L is in microhenries (μH) and C is picofarads, this becomes

$$f = \frac{10^6}{2\pi\sqrt{LC}} \text{ kilohertz}$$

The basic formula can be rearranged

$$L = \frac{1}{4\pi^2 f^2 C} \text{ henries} \qquad C = \frac{1}{4\pi^2 f L} \text{ farads}$$

Since $2\pi f$ is commonly represented by ω, these expressions can be written

$$L = \frac{1}{\omega^2 C} \text{ henries} \qquad C = \frac{1}{\omega^2 L} \text{ farads}$$

Time constant

For a combination of inductance and resistance in series the time constant (i.e. the time required for the current to reach 63% of its final value) is given by

$$\tau = \frac{L}{R} \text{ seconds}$$

where L = inductance (henries), and R = resistance (ohms).

For a combination of capacitance and resistance in series the time constant (i.e. the time required for the voltage across the capacitance to reach 63% of its final value) is given by

$$\tau = CR \text{ seconds}$$

where C = capacitance (farads), and R = resistance (ohms).

Transformer ratios

The ratio of a transformer refers to the ratio of the number of turns in one winding to the number of turns in the other winding. To avoid confusion it is always desirable to state in which sense the ratio is being expressed: e.g. the 'primary-to-secondary' ratio $n_{\mathrm{p}}/n_{\mathrm{s}}$. The turns ratio is related to the impedance ratio thus

$$\frac{n_{\mathrm{p}}}{n_{\mathrm{s}}}\sqrt{\frac{Z_{\mathrm{p}}}{Z_{\mathrm{s}}}}$$

where n_{p} = number of primary turns, n_{s} = number of secondary turns, Z_{p} = impedance of primary (ohms), and Z_{s} = impedance of secondary (ohms).

Wattage rating

If resistance and current values are known,

$$W = I^2 R \text{ when } I \text{ is in amperes}$$

or

$$W = \frac{\text{milliamps}^2}{1\ 000\ 000} \times R$$

If wattage rating and value of resistance are known, the safe current for the resistor can be calculated from

$$\text{milliampers} = 1.000 \times \sqrt{\frac{\text{watts}}{\text{ohms}}}$$

Wavelength of tuned circuit

Formula for the wavelength in metres of a tuned oscillatory circuit is: $1885\sqrt{LC}$, where L = inductance in microhenries and C = capacitance in microfarads.

26.9 Colour codes

26.9.1 Resistor and capacitor colour coding

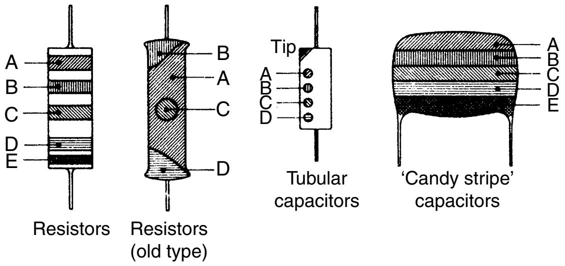

Resistors Resistors (old type)

Tubular capacitors

'Candy stripe' capacitors

Tantalum capacitors

	1	2	3	4
Black	–	0	× 1	10 V
Brown	1	1	× 10	
Red	2	2	× 100	
Orange	3	3	–	
Yellow	4	4	–	6.3 V
Green	5	5	–	16 V
Blue	6	6	–	20 V
Violet	7	7	–	
Grey	8	8	× 0.01	25 V
White	9	9	× 0.1	3 V
				(Pink 35 V)

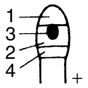

Resistor and capacitor colour coding

			Band C (multiplier)		*Band D (tolerance)*			*Band E*	
Colour	*Brand A*	*Brand B*	*Resistors*	*Capacitors*	*Resistors*	*Capacitors Up to 10 pF*	*Over 10 pF*	*Resistors*	*Polyester capacitors*
Black	–	0	1	1	–	2 pF	±20%	–	–
Brown	1	1	10	10	±1%	0.1 pF	±1%	–	–
Red	2	2	100	100	±2%	–	±2%	–	250 v.w.
Orange	3	3	1 000	1 000	–	–	±2.5%	–	–
Yellow	4	4	10 000	10 000	–	–	–	–	–
Green	5	5	100 000	–	–	0.5 pF	±5%	–	–
Blue	6	6	1 000 000	–	–	–	–	–	–
Violet	7	7	10 000 000	–	–	–	–	–	–
Grey	8	8	10^8	0.01 μF	–	0.25 pF	–	–	–
White	9	9	10^9	0^1 μF	–	1 pF	±10%	–	–
Silver	–	–	0.01	–	±10%	–	–	–	–
Gold	–	–	0.1	–	±5%	–	–	–	–
Pink	–	–	–	–	–	–	–	Hi-Stab.	–
None	–	–	–	–	±20%	–	–	–	–

Note that adjacent bands may be of the same colour unseparated.

Preferred values

E12 Series

1.0 1.2 1.5 1.8 2.2 2.7
3.3 3.9 4.7 5.6 6.8 8.2
and their decades

E24 Series

1.0 1.1 1.2 1.3 1.5 1.6
1.8 2.0 2.2 2.4 2.7 3.0
3.3 3.6 3.9 4.3 4.7 5.1
5.6 6.2 6.8 7.5 8.2 9.1
and their decades

26.9.2 Resistor and capacitor letter and digit code table (BS 1852)

Resistor values are indicated as follows:

0.47 Ω	marked	R47
1 Ω		1R0
4.7 Ω		4R7
47 Ω		47R
100 Ω		100R
1 kΩ		1K0
10 kΩ		10K
10 MΩ		10M

A letter following the value shows the tolerance.

F = ±1%; G = ±2%; J = ±5%; K = ±10%; M = ±20%; R33M = 0.33Ω ± 20%; 6K8F = 6.8 kΩ ± 1%.

Capacitor values are indicated as:

0.68 pF	marked	p68
6.8 pf		6p8
1000 pF		1n0
6.8 nf		6n8
1000 nF		1μ0
6.8 μF		6μ8

Tolerance is indicated by letters as for resistors. Values up to 999 pF are marked in pF, from 1000 pf to 999 000 pF (= 999 nF) as nF (1000 pF = InF) and from 1000 nF (= 1 μF) upwards as μF.

Some capacitors are marked with a code denoting the value in pF (first two figures) followed by a multiplier as a power of ten (3 = 10^3). Letters denote tolerance as for resistors but C = ±0.25 pf. E.g. 123 J = 12 pF × 10^3 ± 5% = 12 000 pF (or 0.12 μF).

Index

Abbreviations, 307
Absorption, 7, 159
Activity, 100
Adaptive Differential Pulse Code Modulation (ADPCM), 181
Adjacent channel selectivity, 156
Ageing, 99
AM broadcast station classes (USA), 295
AM splash, 114
AM transmitter, 143
Ampere-hour, 325
Ampere's rule, 325
Ampere's theorem, 325
Amplitude Modulation (AM), 113
Amplitude modulation transmitters, 143
Amplitude modulation, 113
Amplitude Shift Keying (ASK), 125
Analogue modulation, 113
Analogue signals, 110
Analogue to Digital Converter (ADC), 179
Angle modulated transmitters, 145
Angle modulation, 117
Antenna characteristics, 52
Antenna gain, 52, 67, 159
Antenna placement, 66, 214
Antenna radiation angle, 52
Antenna types, 52
Antenna, dummy, 232
Antennas, 52
Aperture, 55
Assembly of connectors, 260
AT-cut crystal, 95
Atmospheric attenuation, 6
Atmospheric conditions, 164
Atmospheric losses, 6
Atmospheric noise, 11
Attenuation, 103
Audible frequency range, 110
Audio connectors, 256
Audio frequency response, 156
Audio output, 156
Aurora propagation, 23
Auto-acknowledgement, 185
Automatic Gain Control (AGC), 149
Axial mode helix, 71

Back-scatter, 19
Balanced line hybrids, 36
Balanced modulator, 116, 122
BALUN, 44
Bandpass circuit, 81
Bandwidth (BW), 52, 110, 135
Bandwidth requirements, 110, 125
Base band lines, 36
Base band, 110
Base bandwidth, 110
Base station antennas, 61, 215
Batteries, 236
Baud rate, 110
Baying antennas, 65
BBC VHF test tone transmissions, 284
Beam-width, 65
Bessel functions, 118
Binary amplitude shift keying (BASK), 125
Binary decibel values, 34
Binary FFSK, 126
Binary phase shift keying, 128
Binary signal rate, 110
Bit error rate, 164
Bit rate, 110
Block encryption, 169
Blocking, 156
Bluetooth, 301
Boolean algebra, 327
Broadcasting, 281
Broadside array, 58
BT-cut crystal, 95
Bulk acoustic wave, 105
Bulk delay, 105

Caesium frequency standard, 93
Camera connectors, 257
Candela, 325
Capacitance, 58, 79, 328

Capacitance, coaxial cable, 50
Capacitive reactance, 79
Capacitor colour code, 334
Capacitor letter code, 336
Capacity, cell, 236
Capture area, 55
CCITT (see ITU)
Cell characteristics, 239
Cellular telephones, 208
Centre frequency, 103
Centronics interface, 276
Ceramic filter, 102
Ceramic resonator, 105
Channel capacity, 112, 200
Characteristic impedance, 37, 328
Characteristics of UHF TV, 288
Cipher text, 167
Circuit condition, 102
Circular polarization, 54
Circulator, 224
Class-A stations, 295
Class-B stations, 295
Class-C stations, 295
Class-D stations, 295
Classes of radio station, 139
Coaxial cable capacitance, 50
Coaxial cable cut-off frequencies, 50
Coaxial cable, 48
Coaxial cable, types, 39, 48
Coaxial connector, 258
Co-channel interference, 61
Code Division Multiple Access (CDMA), 129, 172, 177, 212
Code Excited Linear Prediction (CELP), 181
Co-linear antennas, 62
Colour codes, 334
Colpitt's oscillators, 85
Common base-station control, 186
Communications by satellite link, 248
Conical antenna, 65
Connectors, 256
Constant-current automatic charging, 243
Continuous wave, 129
Cordless Telephony, 206
Coulomb, 325
Coulomb's law, 325
Coupled bandpass circuits, 81
Critical coupling, 83
Critical frequency, 16
Cross modulation, 156
Cross polarization, 54
Cryptographic channel, 167
Crystal case styles, 101
Crystal filter, 102
Crystal oscillator, 87, 90

Data Communications Equipment (DCE), 269
Data encryption, 169
Data over GSM, 210
Data Terminating Equipment (DTE), 268
dB to any ratio conversion, 33
DB, 25
dBa, 31
dBa0, 31
dBd, 31
dBi, 31
dBm to watts conversion, 33
dBm, 31
dBm0, 32
dBm0p, 32
dBmp, 32
dBr, 32
dBrn, 33
dBrnc, 33
dBrnc0, 33
DBS transmission, 248
dBu, 33
dBuV, 31
dBV, 33
dBW, 33
DCE, 269
Decibel glossary, 31
Decibel scale, 25
Decibel, 325
Decibels referred to absolute values, 25
Decimal multipliers, 327
De-emphasis, 121
Density inversion, 160
Depth of modulation (AM), 113
Desensitisation, 156
Designation of radio emissions, 134
Deviation (FM), 117
Deviation ratio, 118
Differential delay, 104
Differential phase shift keying, 128
Differential phase, 104

Diffraction, 9, 160
Digital European Cordless Telephone (DECT), 207
Digital modulation, 123
Digital multiplexing, 174
Digital signalling, 195, 197
Digital signals, 110
Digitally coded squelch, 197
DIN standards, 256
Dipole antenna, 5, 55
Dipole, 53, 56, 63
Direct Broadcasting Satellite, 248
Direct digital synthesis, 93
Direct wave, 163
Direction of propagation, 46
Directional array, 58
Directivity, 52, 62
Directors, 62
Discone antenna, 65
Distortion, 230
Doppler effect, 12
Doppler frequency shift, 12
Double sideband amplitude modulation (DSB), 113
Double sideband suppressed carrier (DSBSC), 115, 122
Doublet antenna, 5
Drive level linearity, 105
Drive level stability, 105
Drive level, 102
DTE, 268
Dual modulus pre-scaler, 92
Ducting, 20, 160
Duplex separation, 156, 173
Duplex, 173, 184
Dynamic resistance, 81, 329

E1 multiplex, 176
Earth conductivity, 218
Earth orbits, 246
Effective height, 53, 58, 77
Effective length, 53
Effective parallel resistance, 100
Effective radiated power, 53
Effective series resistance, 100
Efficiency, 53
EIA, 274
Electric field, 3, 54, 106
Electrical relationships, 323
Electromagnetic wave, 2
Elliptical orbit, 246
Encryption key, 169
Encryption, 167
End-fed dipole, 61
End-fire array, 58
E-plane, 3, 54
Equivalent circuit (crystal), 95
Erlangs, 200

Farad, 80, 326
Far-field, 74
Fast frequency shift keying (FFSK), 126, 197
Field strength, 73
Filter, 102
Five-tone signalling, 196
Flat loss, 103
FM broadcast frequencies (USA), 296
FM transmitter, 145
FM, 117
Folded dipole, 61
Forward gain, 69
Forward scatter, 19
Fractional bandwidth, 103
Fr-cut crystal, 95
Free space loss, 6
Frequency band table, 2
Frequency designations, 2, 134
Frequency Division Duplex (FDD), 173
Frequency Division Multiplex (FDM), 122, 173
Frequency modulation deviation, 117
Frequency modulation, 117
Frequency planning, 132
Frequency response, 109
Frequency shift keying, 125
Frequency stability, 89, 98, 102
Frequency synthesizer, 89
Frequency tolerance, 102
Frequency, 1, 102, 232, 329
Fresnel zone, 160
Friis noise equation, 154
Front-to-back ratio, 54, 69
Fundamental constants, 323
Fundamental frequency, 87, 112
Fundamental units, 324

Gain, 4, 25
Gas-filled line, 48
Gaussian minimum phase shift keying (GMSK), 126
General frequency allocations, 135
Geostationary orbits, 246
Global positioning system (GPS), 252
Global System for Mobile communications (GSM), 102, 209
Grade of service, 201
Gray coding, 124
Great Circle Bearings, 302
Great circle path, 22
Great circle, 302
Greek alphabet, 325
Grey line propagation, 23
Ground plane antenna, 57
Ground wave, 13
Group delay, 104
Grover equation, 77

Half-power beam width, 52
Hand-portable antenna, 68
Hard-line, 48
Hartley oscillators, 85
Hartley-Shannon theorem, 112
Helical antenna, 68
Helical line, 48
Henry, 80, 326
HF band, 2, 136
High earth orbit (HEO), 246
High frequency, 2
Holder style, 102
Homodyne, 151
H-plane, 3, 54
Human voices, 156
Huygen's principle, 9

IEEE802.11b, 301
IF amplifier, 151
IF bandwidth, 149
Image frequency, 149
Impedance at resonance, 330
Impedance matching, 35
Impedance of free space, 4
Impedance, 54, 104, 329
Impulse response, 109
In band intermodulation distortion, 105
In band signalling, 195
Inductance, 79, 329
Industrial Scientific and Medical (ISM) band, 301
Insertion loss, 38, 103
Insertion phase, 104
Instrumentation, 229
Interfaces, 256, 268
Interfering waves, 113, 221
Intermediate frequency (IF), 149
Intermodulation, 222
Internal resistance, 237
International planning, 132
International Telecommunications Union (ITU), 132
Intersymbol interference, 127
Inverted speech, 168
Inverted-L antenna, 57
Ionised layers, 15
Ionosphere, 15
Isolator, 224
Isotropic radiator, 3

J.M.E. Baudot, 110
Joule, 326
Joule's law, 326

K-factor, 20, 82, 161
Kilovolt-ampere, 326
Kilowatt, 326

Lead acid batteries, 242
LF, 2, 135
License-free bands, 301
Light, velocity, 323, 327
Link planning, 165
Lithium battery, 238
Load capacitance, 99
Load impedance, 104
Location, 255
Log-periodic antenna, 60
Loop antennas, 73
Loop filter, 90
Loop inductance, 77
Loss, 25

Low profile antennas, 67
Lower sideband, 113
Low-pass (Gaussian) filter, 124
Lumen, 326

MAC format, 248
Magnetic field, 3
Manchester encoding, 124
Mark, 111
M-ary FSK, 127
Matched line loss, 47
Maximum input level, 105
Maximum usable frequency (MUF), 16
Medium frequency (MF), 2, 135
Memory effect (NiCad batteries), 243
Meteor ion trail, 19
Methods of coupling, 83
Metre conversion, 330
Mho, 326
Microwave antennas, 69
Microwave communications, 159
Microwave frequencies, 159
Minimum frequency shift keying (MFSK), 126
Mismatch, 35
Mobile antennas, 65
Mobile data transmission, 210
Mobile radio, 183, 205
MODEM connector, 271
Modulation depth, 113
Modulation index, 117
Modulation, 112
Motorcycle antenna, 67
Mounting, 102
MPT-1317, 127
MPT-1326, 88
MPT-1362, 67
MSF Rugby, 283
Multi-coupling, 225
Multipath propagation, 10
Multiplexing (analogue), 173
Multiplexing (digital), 174
Mutual inductance, 81

Narrow band FM, 127
National planning, 132
Near-field, 74
Negative feedback, 330
Newton, 326
Ni-cad, 243
Nickel cadmium batteries, 243
Noise factor, 47, 152
Noise figure, 47, 151
Noise in cascade amplifiers, 154
Noise temperature, 152
Noise voltage, 11
Noise, 11
Non-chargeable batteries, 238
Non-reciprocal direction, 23
Number code, 336
Numerically controlled oscillator, 93

Ohm, 327
Ohm's law, 326
Omni-directional normal mode helix, 71
Omni-directional radiation pattern, 61
Omni-directional, 71
On-off keying, 125
Oscillator frequency, 85
Oscillator requirements, 85
Oscillators, 85
Out-of-band intermodulation distortion, 105
Overtone crystals, 88
Overtone frequency, 88, 100
Overtone oscillator, 88

Paging, 205
PAL signals, 248
Parallel resonance, 80
Parallel resonant circuits, 80, 96
Pascal, 326
Passband, 103
Patterson equation, 77
Peak envelope power, 117
Personal Communication System, 212
Phase comparator, 90
Phase linearity, 104
Phase locked loops, 90
Phase modulated transmitters, 120
Phase modulation, 120
Phase shift keying (PSK), 128
Physical properties, 324

Piezoelectric devices, 95, 105
Piezoelectric effect, 95, 105
Pilot carrier, 117
Pilot tone, 122
Plain text, 167
Polarisation, 4, 54
Polarised, 4
Potential difference, 75
Power in unmodulated carrier, 114
Power output, 143
Power ratio, 237, 331
Power relationships AM wave, 114
Power supplies, 226, 236
Power, 114, 231, 331
Power-Volume ratio, 237
Power-Weight ratio, 237
Pre-emphasis, 121
Pre-scaler, 92
Primary batteries, 238
Privacy, 167
Private Mobile Radio (PMR), 183, 212
Programmable equipment, 91, 157
Programmable read only memory (PROM), 157
Propagation velocity, 1, 106, 323
Propagation, 3, 159
Propagation, direction of, 3
Propagation, methods of, 13
Pseudo-noise generation, 131
Public Access Mobile Radio (PAMR), 202
Pull-ability, 99
Pulse Amplitude Modulation (PAM), 179
Pulse Code Modulation (PCM), 179

Q (figure of merit), 81, 100, 331
Q factor, 81
Quadrature phase shift keying (QPSK), 116
Quarter wave section, 57
Quartz crystal characteristics, 97
Quartz crystal oscillator, 87
Quartz crystal, 87, 95
Quartz filter, 102

Radiation lobe, 52, 67
Radiation pattern, 55, 67
Radiation resistance, 53
Radiation, 3, 67
Radiator, 56–63
Radio equipment, 143
Radio frequency lines, 37
Radio frequency spectrum, 1, 139
Radio horizon, 8, 331
Radio Investigation Service, 134
Radio station classes, 139
Radio waves, formation, 3
Radio-communications Agency (RA), 133
Reactance of capacitors, 332
Reactance of inductors, 332
Reactance, 332
Receive aperture, 55
Receiver functions, 148
Receiver specifications, 156
Receiver voting, 187
Receiver, types of, 148
Rechargeable batteries, 226, 242
Recharging conditions, 237
Reflected component, 163
Reflected wave, 161
Reflection, 9, 160
Reflectors, 62
Refraction modes, 20
Refraction, 8
Refractive index, 8
Regional planning, 132
Regulation of radio, 132
Resistor colour code, 334
Resistor digit code, 336
Resistor letter code, 336
Resonance, 332
Resonant circuits, 79
Resonant frequency, 80, 97
Response of coupled circuits, 83
RF cable, 39
RF transformer, 44
Rhombic antenna, 59
Ripple, 103
RMS, 230
RS-232C, 274
RS-449 275
Rubidium frequency standard, 93

Safety, 69, 217
Sampling rate, 179
Satellite communications, 246

Satellite television formats, 248
Satellite television, 248
SAW filter, 105
SCART (BS-6552), 258
Scatter propagation modes, 19, 23
Scatter, 19
Scrambling, 168
Selective calling, 185
Sensitivity, 156
Series resonant (crystal), 87
Series resonant circuits, 79
Shape factor, 104
Side scatter, 19
Sideband, 113, 118
Signalling channel, 176–7, 194
Signal-to-noise ratio, 152–6
Simplex, 183
SINAD, 230
Single sideband AM wave, 114
Single sideband suppressed carrier, 114, 115–7
Skip distance, 16
Sky wave propagation, 15
Slot antenna, 64
Small loop antenna patterns, 75
Small loop antenna, 72
Small loop geometry, 74
Sound velocity, 327
Source impedance, 104
Space (digital signal), 111
Space wave propagation, 16
Space wave, 16
Specification of quartz crystals, 101
Spectrum, 2, 233
Speech encryption, 168
Speed of light, 327
Sporadic E-layer reflections, 16
Spread spectrum transmission, 129
Spread spectrum, 172
Spreading code, 177
Spurious attenuation, 104
Spurious response attenuation, 156
Spurious responses, 100
Squelch, 197
Stability, 98
Stacking antennas, 65
Standard frequency formats, 283
Standard frequency transmissions, 281
Standard test tone, 165
Standard time transmissions, 281
Standard units, 325
Stereo radio modulation scheme, 121
Stopband performance, 104
Stopband, 103
Stream encryption, 169
Sub-audio signalling, 194
Sub-refraction, 21
Super refraction, 20
Switching bandwidth, 156
Symbols, 313
Synchronisation, 176–7, 180
Synchronous encryption, 171

T1 multiplex, 176
Television channels (USA), 299
Television connectors, 257
Temperature coefficient, 86, 98
Temperature compensation, 86
Temperature range, 102
Tesla, 327
Thermal noise, 11
Time constant, 333
Time division multiplex (TDM), 174
Total line loss, 47
Transducer, 108
Trans-European Trunked Radio (TETRA), 208
Transformer ratios, 333
Transition band, 104
Transmission line considerations, 47
Transmission line filters, 44
Transmission line noise, 49
Transmission lines, 35
Transmission quality, 164
Transmitter specifications, 147
Transmitter, FM, 146
Transmitters, angle modulated, 145
Transmitters, phase modulated, 147
Transmitters143
Transverse electric, 46
Transverse electromagnetic wave, 46
Transverse magnetic, 47
Trickle charge current, 242
Trickle charging, 242
Tropospheric scatter, 18
Trunked radio, 202, 207
Trunking, 201
Tuned radio frequency (TRF), 148
Tuned resonant circuits, 79
TV channels (Australia), 292

TV channels (New Zealand), 292
TV channels (Republic of Ireland), 291
TV channels (South Africa), 292
TV channels (UK), 291
TV channels (USA), 293

UHF, 2, 61, 67, 138, 183
UK, 625-line TV system, 289
UK broadcasting band, 284
Unipole antenna, 61
Unweighted transducer, 109
Upper sideband, 113

Varicap diode, 89
Velocity of sound, 329
Vertical (H) plane, 4
Very low frequency, 2
VHF, 2, 18, 61, 68, 138, 183
Video recorder connectors, 257
Virtual height, 16
VLF, 2, 135
Volt, 327
Voltage controlled oscillators (VCO), 89
Voltage standing wave ratio (VSWR), 42, 55, 221
Volt-ampere, 327

Watt, 327
Wattage rating, 333
Wave-guide, 45
Wavelength of tuned circuit, 334
Wavelength, 1, 107, 329
Weber, 327
Wide-area coverage, 187
World Administrative Radio Conference, 132

X-cut crystal, 95

Yagi array, 62, 69, 294
Y-cut crystal, 95

Z-cut crystal, 95
Zero transmission reference point, 165
Zero-IF receiver, 151
Zulu time, 303

RADIO
AM•FM

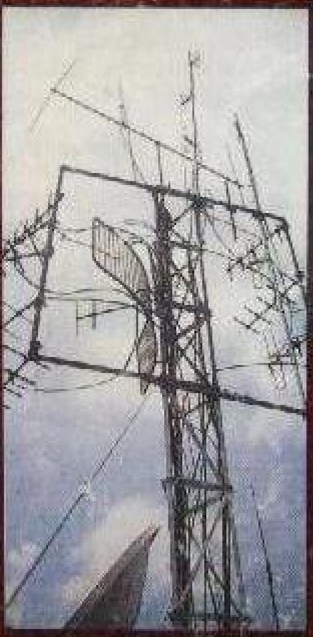

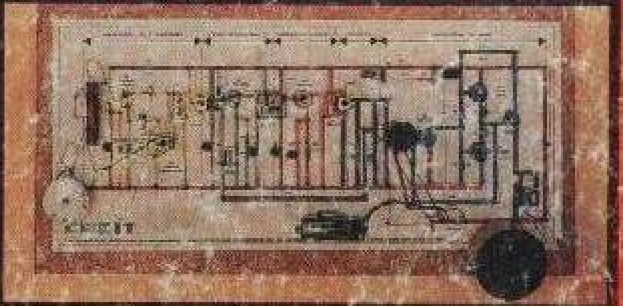

PROLOGO

Estimado lector :

Después del rotundo éxito de "Tu futuro es la electrónica" , y el "Curso práctico de ELECTRONICA DIGITAL y Circuitos integrados" , CEKIT S.A. presenta una nueva obra en el campo de la electrónica, el Curso de Radio AM, FM, Banda Ciudadana y Radioafición.

En esta obra estudiaremos una rama muy importante de la electrónica, la parte que se refiere a las ondas de radio. Conoceremos qué son, cómo se producen, transmiten y reciben en un lugar remoto, los aparatos receptores, transmisores y sus componentes.

De esta manera entraremos en el campo de la radio, las comunicaciones y la radioafición, uno de los hobbies o pasatiempos más difundidos en todo el mundo y que no sólo ha tenido una gran influencia en el desarrollo técnico y científico de la electrónica,sino que ha colaborado exitosamente con el hombre en momentos de grandes dificultades.

Con este curso se pueden complementar y profundizar los conocimientos adquiridos con la obra anterior,pero también está diseñado para que se pueda tomar como un curso de iniciación sin que para hacerlo se requiera de ningún curso anterior.

De esta manera es también el curso ideal para aquellas personas que se quieran iniciar en el maravilloso mundo de la electrónica, la radio y la radioafición.¿Quién,en algún momento,no ha sentido deseos de armar su propio radio o su propio transmisor, así las señales no lleguen muy lejos ?.

El curso de Radio AM, FM, Banda Ciudadana y Radioafición es una obra amena, clara, concisa y sencilla, tratada con un lenguaje comprensible y agradable, con muchas ilustraciones y ejemplos que ayudan a asimilar los conceptos teóricos y prácticos tratados.

Todas las leyes y teorías son explicadas en párrafos cortos, fáciles de leer y entender, apoyadas por medio de ejemplos y comparaciones sencillas, mostradas de una manera muy clara, sin caer en la simplicidad o superficialidad.

El estudio de la teoría y práctica de la radio es muy importante dentro de la electrónica en general,ya que si nos remontamos a los orígenes de la electrónica descubrimos que todo empezó con los primeros experimentos acerca de las ondas de radio u ondas electromagnéticas. Los pioneros de estos descubrimientos buscaron la forma de transmitir una onda o señal de un sitio a otro,sin necesidad de tener alambres conectados entre las dos estaciones.

Es entonces el conocimiento de estos principios lo que ha llevado a esta ciencia y tecnología hasta el lugar que ocupa actualmente,en donde casi ninguna actividad humana está libre de utilizar los aparatos electrónicos con el fin de que se pueda realizar mejor.

La importancia de la radio en la historia es muy grande, ya que originó después de su descubrimiento,una verdadera revolución de las comunicaciones entre los hombres. Después de ella, éstas no han sido las mismas.

Los principios de transmisión y recepción de las ondas de radio,utilizados inicialmente para transmitir la voz, luego la música y por último la televisión con sus imágenes a todo color desde y hasta cualquier lugar del mundo,han acercado a los hombres de tal manera que ya nadie es ajeno a los acontecimientos de todo el planeta.

Los grandes avances tecnológicos,como los viajes espaciales y muchos otros,han dependido en gran parte del desarrollo en las comunicaciones. La radio además de llevar información,también ha contribuido al esparcimiento y entretensión del hombre,irradiando música y cultura de todo tipo, que se puede seleccionar fácilmente, de acuerdo a las preferencias, moviendo un simple botón.

La masificación de la industria electrónica,con la llegada del transistor y los circuitos integrados, ha permitido la fabricación de miles de millones de receptores de radio a precios muy bajos, lo que hace casi imposible que alguna persona no tenga un radioreceptor disponible para su utilización diaria y permanente. Diríamos sin temor a equivocarnos que el radio es el aparato electrónico más numeroso sobre la tierra.

Es entonces muy importante para los aficionados y las personas dedicadas a la electrónica como profesión, conocer qué es un radio, cómo está fabricado, cómo funciona, cómo podemos ensamblarlo nosotros mismos y cómo podemos repararlo cuando se dañe.

El curso tendrá como característica principal la metodología teórico-práctica en donde se irán tratando los temas y principios teóricos de una manera muy sencilla pero profunda a la vez,y se van complementando con las prácticas de ensamble de un radioreceptor transistorizado de alta calidad que,al final del curso,quedará completamente armado y funcionando y que usted exhibirá orgulloso como una de sus grandes realizaciones.

Inicialmente en el curso haremos una introducción general al tema de la radio y sus características principales. Luego haremos un repaso de su historia, un tema fascinante y muy agradable lleno de anécdotas, sueños y decepciones. La lucha de los pioneros fué grande, casi todos pasaban por locos y nadie creía en sus ideas o teorías.

Después conoceremos el tema de las ondas electromagnéticas en sus aspectos generales para ver cómo se aplican en la teoría de la radio. Con el estudio de las ondas pasamos a ver cómo se transmite una señal sonora de un punto a otro,tan distante como se quiera,analizando en palabras comunes el proceso de transmisión y recepción de señales de radio.

Una vez que conozcamos los aspectos generales de la teoría que se va a estudiar, entramos en materia con las lecciones del curso,que se van presentando en forma sencilla y muy didáctica.

Es necesario primero conocer los aspectos fundamentales de la estructura interna de la materia que nos rodea,con el fin de entender de dónde sale la corriente eléctrica y para ello estudiamos las propiedades y constitución de la materia, tocando el tema de las moléculas, los átomos y los electrones.

El conocimiento de la materia es muy importante para entender muchos fenómenos,que se producen en los diferentes materiales que se utilizan para fabricar los componentes electrónicos,que a su vez,se utilizan en todos los aparatos electrónicos que se van a estudiar en este curso.

Una vez conocida la materia y los electrones,veremos cómo se produce una corriente eléctrica, veremos cómo hay varios tipos de corrientes y cómo se aplican en los diferentes circuitos para hacerlos funcionar correctamente.

Después estudiaremos cuáles son los componentes básicos y los diferentes tipos que se utilizan para formar los aparatos de radio y comunicaciones,que son el tema central del curso.

Iremos conociendo primero los componentes más sencillos,llamados " componentes pasivos," como son las resistencias, los condensadores y las bobinas;para luego pasar a los componentes llamados " componentes activos," que son los que realizan los trabajos más importantes en los circuitos y aparatos electrónicos,como son los diodos, los transistores y los circuitos integrados.

Pero como un componente solo no realiza ningún trabajo real, se estudia la forma como en la electrónica se unen dos o más componentes para formar un circuito y según las conexiones se forman circuitos diferentes. Además estudiaremos cómo, uniendo varios circuitos, se forma un aparato completo; en este caso veremos cómo está compuesto un radio de AM; o un radio de FM o frecuencia modulada;y cómo se construye y funciona un transmisor de radio,y, además,cómo opera una emisora de radio en sus diferentes funciones.

Simultáneamente con las lecciones teóricas se irán desarrollando las actividades de práctica,que consisten fundamentalmente en las instrucciones detalladas y muy ilustradas para ensamblar un radio completo y superdidáctico de AM con seis transistores. Esto es lo que hace realmente diferente este curso. A medida que se va estudiando la teoría vamos practicando y comprobando que lo que se ha conocido es verdad y realmente funciona.

Este radio se va armando por etapas, que están claramente definidas en cuanto a los componentes que la forman y a su función específica dentro del aparato total de radio. Así veremos cómo llega una determinada señal a la antena, cómo se separa de las otras señales y cómo se amplifica hasta quedar convertida,casi inmediatamente,en una señal sonora igual a la que se había originado en una distante emisora.

Para armar el radio se entregan varias láminas o ilustraciones con toda la información para guiar al alumno sobre la forma de ensamblar correctamente cada etapa y de cómo distinguir fácilmente los componentes para su correcta instalación. Para estas prácticas solamente se requieren unas pocas y económicas herramientas, que además le serán de gran utilidad al experimentador y al aficionado para continuar con su afición o trabajo en una forma mucho más eficiente.

Dentro del curso aprenderemos a fabricar algunos instrumentos muy útiles en el ajuste y reparación de todo tipo de radios y otros aparatos electrónicos,relacionados con el tema,para lo cual se entrega una sección dedicada a esta labor.

Estamos seguros que al final de este curso estará muy satisfecho ya que habrá aprendido muchas cosas que le aumentarán su cultura electrónica y que le serán útiles en sus momentos de esparcimiento o trabajo en esta rama de la electrónica, la radio y la radioafición.

Aspectos básicos de la teoría de Radio

Qué es la radio o el radio

La palabra radio se utiliza para diferentes cosas: En química para referirse a un elemento que es radioactivo o sea que emite radiaciones; en geometría radio se refiere a una línea que une el centro del círculo con un punto de su circunferencia; también existe en el cuerpo humano un hueso del brazo llamado radio.

En electrónica, que es lo que nos interesa, el término radio se utiliza como una abreviatura para radiocomunicación y se refiere principalmente a la técnica tanto teórica como práctica y a los aparatos que nos permiten enviar y escuchar mensajes, música o información en general de un sitio a otro, y ya sea lejano o cercano.

Radiocomunicación

Radiocomunicación es la transmisión por medio de un proceso radioeléctrico o de ondas electromagnéticas, de textos, signos, imágenes o sonidos de toda naturaleza de un lugar a otro.

Antes de entrar a explicar cómo se produce una radiocomunicación, vamos a estudiar qué es la radiación y qué son las ondas, pues debemos tener estos conocimientos básicos para lograr entender cómo opera este proceso.

Radiación y ondas

Cuando estamos cerca de una llama, no sólo la vemos sino que sentimos el calor que produce. Ese calor es la energía que se genera en la combustión y se llama energía calorífica. La energía que nos permite ver las cosas se llama luz visible.

La propagación o transmisión del sonido se efectúa en forma de una energía invisible pero que se puede sentir.

Estos fenómenos y muchos otros corresponden a formas de radiación de energía que se realiza por medio de ondas. La palabra onda se refiere a algo que sube y baja alternadamente y está relacionada con el movimiento.

Para entender más fácilmente qué es una onda vamos a traer dos ejemplos muy comunes y que se presentan en la vida diaria.

El primer ejemplo es el de las "ondas" u ondulaciones que se producen en la superficie de un lago cuando arrojamos en él una piedra, figura 1.

Ondas en el agua

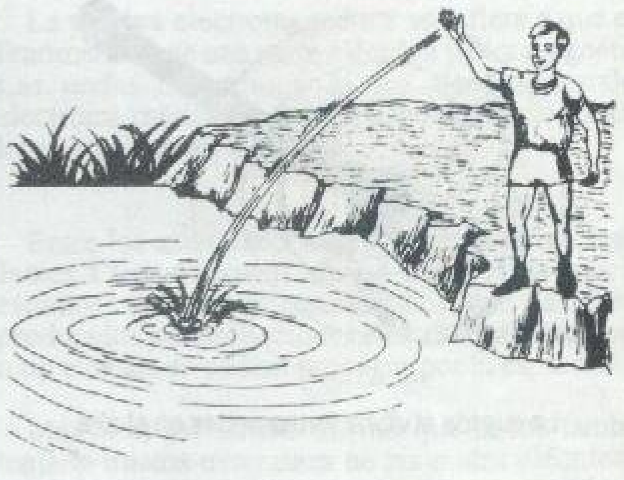

La piedra al caer forma ondas en el agua

Fig. 1

Podrá observarse que a partir del punto en que cayó la misma y sobre la superficie del elemento líquido comienzan a formarse una gran cantidad de ondas concéntricas que, aparentemente, se van alejando del punto en que cayó la piedra y terminan por desvanecerse a la distancia, una vez extinguida la energía que las originó.

Pareciera que el agua se traslada y que al hacerlo se lleva consigo las ondas; mas en realidad no es así. Prueba de ello es que si ubicamos sobre las ondulaciones un objeto flotante, como una pelota pequeña notaremos que él mismo asciende y desciende sobre las ondas, pero que no se aleja del lugar en que ha sido puesto, tal como da la impresión que lo hicieran ellas.

Queda así comprobado que la masa de agua no se desplaza y que sólo asciende o desciende con respecto a su nivel normal. Como resultado de esta observación cabe deducir que lo que en realidad se desplaza no es otra cosa que "energía".

Lo mismo que ocurre con las ondas de energía que se desplazan en el agua, provocando ondas, ocurre también con las ondas sonoras, con la diferencia que éstas se trasladan de un lugar a otro a través del aire.

Al pulsar una cuerda de guitarra, ella vibra con rapidez y se desplaza hacia uno y otro lado; al hacerlo comprime o empuja el aire o las moléculas del mismo durante todo el tiempo que tarda la vibración u oscilación de la cuerda en extinguirse. (figura 2).

Onda de sonido

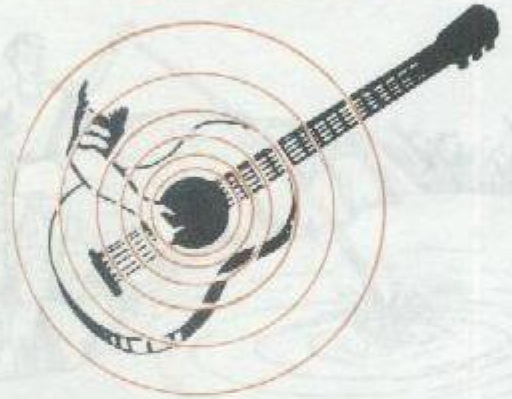

La cuerda al vibrar forma ondas en el aire

Fig. 2

Pese a la pequeñez de la oscilación de la cuerda, la presión originada por su desplazamiento se traslada o propaga bajo la forma de ondas de aire, tal cual ocurre en el agua con las ondas de energía originadas por la caída de la piedra.

Las variaciones descritas respecto a la presión del aire, al llegar a nuestros oídos, provocan idénticas vibraciones en los tímpanos, dándonos una sensación que llamamos sonido, por lo tanto las ondas sonoras no las podemos ver sino oir o sentir.

Características de las ondas

Existen diferentes tipos de ondas según su naturaleza y características lo que hace que tengan manifestaciones distintas en cada caso.

El sonido, la luz, el calor, los rayos X y las ondas de radio son todos ondas. Pero por qué unas ondas las podemos sentir y son invisibles y otras las podemos ver pero no sentir ?. Y por qué unas ondas pueden viajar miles o millones de kilómetros y otras sólo se pueden enviar unos cuantos metros más adelante ?

Todo ésto se puede explicar conociendo las características de cada una de ellas. Las principales características de una onda son su *frecuencia*, su tamaño o *amplitud de onda* y su forma.

Frecuencia

Denominamos frecuencia de una onda la *cantidad de veces* que se produce el fenómeno completo de irradiación en un segundo. Llamamos fenómeno completo a la acción de una onda cuando parte de cero, toma su valor máximo, va a su valor mínimo y luego vuelve a cero. Esto se llama también un *ciclo*. (figura 3).

Ciclo de una onda

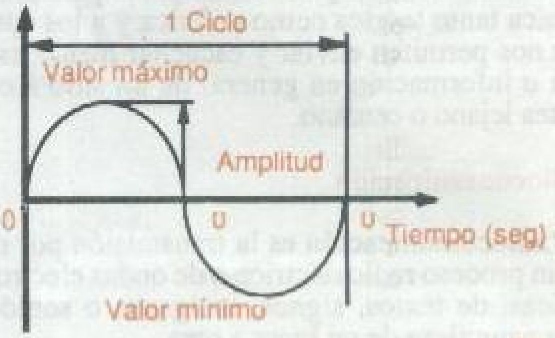

Fig. 3

Es así como la *frecuencia* de una onda se mide en *ciclos por segundo*.

La *amplitud* de la onda es la distancia o tamaño existente entre el valor estable o cero y el punto máximo de vibración que alcanza la onda.

Dependiendo de la frecuencia y de la amplitud que tenga una onda se puede oír, ver o puede viajar distancias muy grandes desde su sitio de origen o creación.

Las ondas sonoras o sonidos se encuentran entre las frecuencias de 20 a 20.000 ciclos por segundo. Esto se debe a que el oído humano está hecho para que responda a ese tipo de ondas o señales. A un grupo de frecuencias que tienen un límite bajo y uno alto se le llama una *banda*.

En cuanto a la *amplitud* de una onda sonora se relaciona con la que llamamos *volumen* o intensidad de sonido. Mientras más amplitud tenga un sonido, más fuerte escuchamos su efecto en nuestros oídos. Si un sonido tiene una amplitud muy baja no podemos escucharlo ya que los tímpanos no alcanzan a vibrar con ese débil estímulo.

Si la frecuencia de una onda sonora se pasa de 20.000 ciclos por segundo ya no podemos escuchar ese sonido y a ese tipo de onda se les llama *ultrasonidos*. Algunos animales, como el perro, tienen la capacidad de escuchar los ultrasonidos.

Otro tipo de ondas con las cuales estamos en contacto permanente son las ondas luminosas o radiación visible. Las ondas luminosas tienen una frecuencia muy alta y hay diferentes tipos o bandas según esta frecuencia.

La luz de más baja frecuencia es llamada *infraroja* y la de más alta frecuencia es la *ultravioleta*. Entre la luz infraroja y la luz ultravioleta tenemos la banda de la luz visible formada por los diferentes colores o frecuencias. Así el rojo tiene la frecuencia más baja dentro de la banda de la luz.

Podemos "ver" estas ondas porque el ojo humano está hecho de tal manera que pueda captar este tipo de radiaciones.

Después de estas ondas tenemos los *rayos X* , los rayos *gamma* y los *rayos cósmicos*. Como todos sabemos los rayos X se utilizan para tomar radiografías o fotografías de la parte interna del cuerpo humano.

Entre las ondas sonoras o sonidos y las ondas luminosas tenemos un grupo de ondas que es el que nos interesa para este curso y que se llaman *ondas de radio*. Las ondas de radio empiezan más o menos desde los 30.000 ciclos por segundo y llegan hasta los 300.000 millones de ciclos por segundo.

Las ondas sonoras sólo pueden viajar algunos metros debido a que el aire va absorviendo su energía y al final ésta desaparece totalmente.

Las ondas de radio tienen la característica muy especial de que pueden viajar grandes distancias sin perder su energía y al descubrirlas el hombre encontró la mejor manera para poder comunicarse de un sitio a otro fácilmente.

Antes de descubrir y aplicar las ondas de radio, se utilizaban las transmisiones por teléfono y telégrafo con hilos en donde los dos lugares a comunicar tenían que estar unidos por medio de cables eléctricos para transportar la señal.

Ondas electromagnéticas

Hay una diferencia muy importante entre las ondas sonoras, las ondas de radio y la luz. Esta diferencia se refiere no sólo a su frecuencia sino a su comportamiento, la forma en que se generan o producen y el medio de propagación.

Las ondas sonoras se producen por medios mecánicos o sea por vibración de algún elemento físico; además este tipo de ondas necesitan de un medio de propagación que en este caso es el aire. En el vacío no podemos transmitir o escuchar ningún sonido.

Cuando por algún medio se hacen vibrar los electrones de un átomo, la vibración es transmitida en cadena a los otros átomos, produciéndose una transmisión de vibraciones de un lugar a otro; a este fenómeno se le conoce con el nombre de "onda electromagnética".

La palabra electromagnética se refiere a que este fenómeno tiene una parte eléctrica y otra magnética. Las ondas de radio y la luz tienen naturaleza electromagnética y para las dos se aplican muchas leyes y principios similares.

Entre los años 1865 a 1870 el gran físico inglés James Clerk Maxwell afirmó después de muchos estudios teóricos y matemáticos que las ondas producidas por oscilaciones eléctricas de frecuencia muy elevada se podían propagar por el espacio.

Maxwell, así mismo afirmó que la luz también tenía la misma naturaleza de las ondas eléctricas y magnéticas y a partir de ese momento se estableció el término *electromagnético*. Maxwell estableció que estas ondas electromagnéticas y la luz se propagan a una velocidad de 300.000 kilómetros por segundo aproximadamente.

Nacimiento de la radio

Breve historia de los primeros años

Después de las afirmaciones de Maxwell se comprobó esta teoría por medio de los experimentos realizados por el físico alemán Enrique Rodolfo Hertz en los años 1887 y 1888.

Hertz observó que si hacía que saltara una chispa a través del aire en un circuito eléctrico, saltaba otra chispa a través del aire en un segundo circuito colocado cerca, pero que no estaba en contacto con el primero. (figura 4).

Por lo tanto, se había transmitido energía en alguna forma misteriosa a través del espacio y en ese momento crucial para el hombre había nacido la radiocomunicación por medio de ondas electromag-

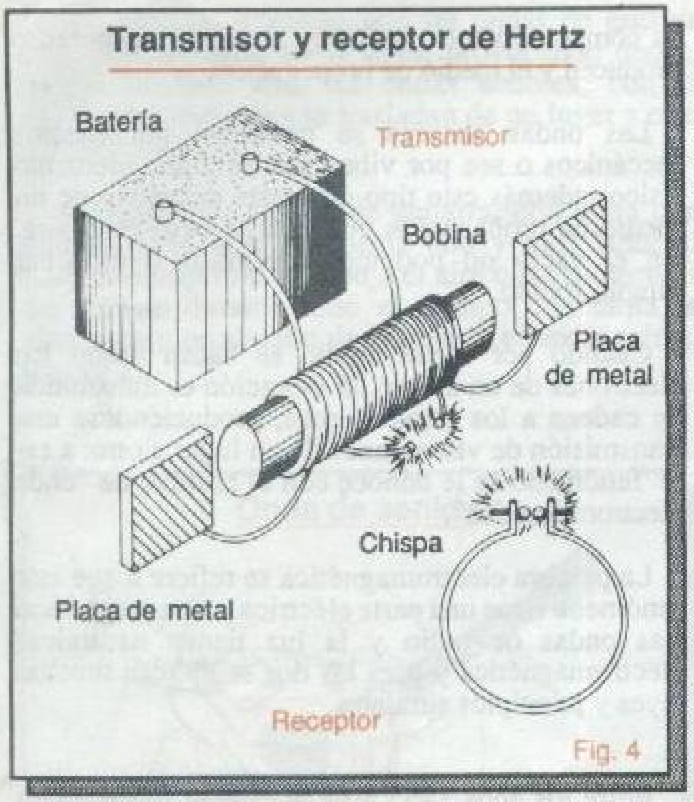

Transmisor y receptor de Hertz

Fig. 4

néticas que tanto le ha servido para su desarrollo y progreso.

Durante mucho tiempo estas ondas de energía producidas por Hertz fueron conocidas como ondas Hertzianas y actualmente se les llama ondas de radio.

En honor a Hertz y a su descubrimiento se ha asignado el nombre de *Hertzio* (Hz) para la unidad de medida de la frecuencia, en cambio de ciclos por segundo. Por lo tanto, para una determinada señal se puede decir que tiene una frecuencia de 100.000 *Hertzios* por ejemplo.

Después del descubrimiento de Hertz fué el italiano Guglielmo Marconi, nacido en Bologna en el año 1874, quien continuó trabajando en el desarrollo de las radiocomunicaciones hasta lograr avances muy significativos.

Marconi se interesó en el fenómeno de la radiación electromagnética al asistir a varias conferencias del profesor Augusto Righi, autoridad de la materia en esa época en Italia y a los 20 años ya había leído mucho sobre el tema. El momento definitivo en la vida de Marconi fué cuando leyó sobre el descubrimiento de Hertz.

Inmediatamente quedó sumergido en la idea de trabajar en ese mismo tipo de experimentos.

El empezó fabricando un aparato similar al de Hertz y después de muchos intentos fallidos tuvo éxito y logró tener un aparato que produjece una chispa más grande cada vez y a mayor distancia en el receptor.

Uno de los primeros avances significativos y originales de Marconi fué la invención de la antena hecha por medio de una placa metálica suspendida en un mástil y el otro terminal del transmisor conectado a la tierra. En el receptor la disposición era la misma, una placa metálica elevada y el otro terminal del receptor conectado a la tierra.

La distancia lograda con este método fué mucho mayor y Marconi tuvo que sacar sus aparatos fuera de su laboratorio a campo abierto para continuar con sus experimentos.

Marconi logró por medio de mucho trabajo y experimentación transmitir señales en clave o código Morse, utilizado en telegrafía, enviando puntos y rayas que logró imprimir en un papel en el receptor que estaba ahora a una distancia de una milla en el año de 1895.

En ese año Marconi le ofreció su invento al Ministerio Italiano de Correos y Telégrafos que lo rechazó por no considerarlo de utilidad. El se estableció entonces en Inglaterra durante un tiempo al tener allí familiares y mostrarse algún interés por parte de algunas entidades oficiales de ese país y mejorando los equipos logró comunicaciones hasta de ocho millas.

Después Marconi desarrolló y patentó en el año de 1900 el sistema de selección de frecuencias, tanto de transmisión como de recepción por medio de los circuitos resonantes, lo que permitió la escogencia de diferentes transmisiones en el mismo receptor, lo que evitaba la interferencia con otras señales. Este fué uno de los avances definitivos y más importantes en el desarrollo de las radiocomunicaciones.

En 1901 Marconi se propuso transmitir señales de radio a través del Océano Atlántico y luego de muchas dificultades e intentos fallidos logró transmitir varias veces la letra S en código Morse (La letra S en el código Morse está formada por tres puntos ...)

El intento empezó el 9 de Diciembre y finalmente se logró la primera recepción de la señal el 12 de Diciembre a la 12:30 PM a una distancia de 2000 millas aproximadamente.

La maravilla de las comunicaciones a larga distancia sin alambres era una realidad y nadie se imaginaba en ese momento hasta donde se iría a llegar con el avance de la radio y luego de la electrónica.

Las ondas de radio

Las ondas descubiertas por Hertz y expreimentadas por Marconi son por lo tanto ondas electromagnéticas *irradiadas* desde la antena de un transmisor. Como las ondas de radio no utilizan ningún medio para desplazarse, su pérdida de energía es sumamente pequeña y pueden recorrer distancias muy grandes.

Las ondas de radio pueden atravesar obstáculos sólidos como paredes, muros, vidrios, etc y pueden llevar información como voz, música e imágenes como en el caso de la televisión.

Las ondas de radio están comprendidas entre las frecuencias de 30KHz y 300 GHz (30 Kilohertzios y 300 Gigahertzios) y a su vez se han dividido en diferentes *bandas* o grupos según su frecuencia.

A estas bandas se les ha dado un nombre y tienen diferentes aplicaciones ya que su comportamiento es diferente. En la tabla 1 tenemos estas bandas con sus frecuencias, símbolo y nombre.

Espectro de las radiofrecuencias

Nombre	Símbolo	Frecuencia
Baja frecuencia	BF (LF)	30 ~ 300KHz
Media frecuencia	MF	300 ~ 3MHz
Frecuencia elevada	FE (HF)	3 ~ 30MHz
Frecuencia muy elevada	FME (VHF)	30 ~ 300MHz
Frecuencia ultra-elevada	FUE (UHF)	300 ~ 3GHz
Frecuencia super-elevada	FSE (SHF)	3 ~ 30GHz
Frecuencia extremadamente elevada	FEE (EHF)	30 ~ 300GHz

Tabla 1

Longitud de onda

A menudo se refiere a las ondas de radio no por su frecuencia en Hertzios sino por su *longitud de onda* medida en metros, centímetros o milímetros según sea el caso.

La longitud de onda es el espacio recorrido durante la realización completa del ciclo de la onda electromagnética, figura 5.

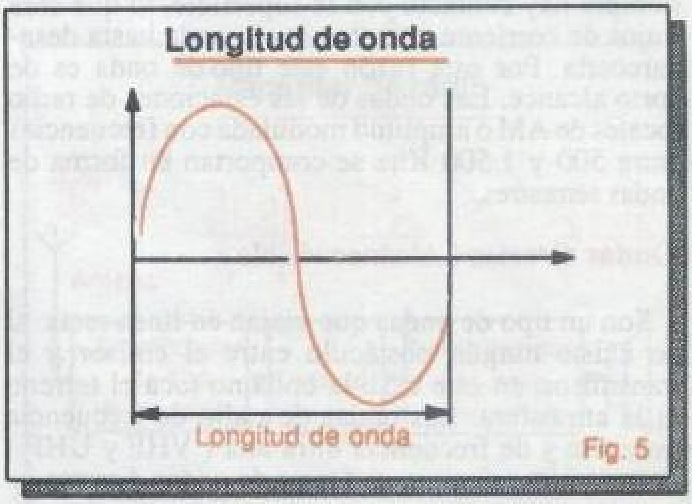
Longitud de onda

Fig. 5

La longitud de onda, que se indica con la letra λ(lambda), se obtiene dividiendo la velocidad de propagación que es de 300.000.000 metros por segundo por la frecuencia en Hertzios de la señal.

$$(\lambda) = \frac{300.000.000}{F\ (Hz)}$$

Por ejemplo para una señal de radio de 3MHz o 3.000.000 de Hertzios tenemos:

$$\lambda = \frac{300.000.000}{3.000.000} = 100 \text{ Metros}$$

A medida que aumenta la frecuencia disminuye la longitud de onda y se pueden tener ondas con longitudes de onda desde metros hasta milímetros.

El tamaño de las antenas depende de la longitud de onda de la señal que van a transmitir o a recibir.

Tomando la longitud de onda se mencionan comunmente las bandas de radio. Por ejemplo, la banda en onda corta de 49 metros, la banda de 11 metros o banda ciudadana para radioteléfonos, la banda de 2 metros para radioaficionados, etc.

Tipos de ondas de radio

Según su frecuencia o longitud de onda las señales de radio se comportan en forma diferente. En cuanto a su propagación o forma de viajar se refiere, los principales tipos son: ondas terrestres, ondas directas y las ondas espaciales o reflejadas.

Ondas terrestres

Este tipo de ondas viaja alrededor de la superficie curva de la tierra sin penetrar en la atmósfera; siempre hay contacto con la superficie, lo que crea flujos de corriente que atenúan la onda hasta desaparecerla. Por esta razón este tipo de onda es de corto alcance. Las ondas de las estaciones de radio locales de AM o amplitud modulada con frecuencias entre 500 y 1.500 Khz se comportan en forma de ondas terrestres.

Ondas directas (Alcance visible)

Son un tipo de ondas que viajan en línea recta si no existe ningún obstáculo entre el emisor y el transmisor, en este caso la onda no toca el terreno ni la atmósfera. Las ondas de radio de frecuencia muy alta y de frecuencia ultra alta (VHF y UHF) se transmiten siempre en forma de ondas directas.

Este tipo de radiación se utiliza principalmente en las transmisiones de televisión, emisoras de FM (frecuencia modulada), radioteléfonos para servicios públicos y radioaficionados y transmisiones por satélite. Su alcance está limitado por la topografía del terreno y por la altura entre las dos antenas, la transmisora y la receptora. (figura 6).

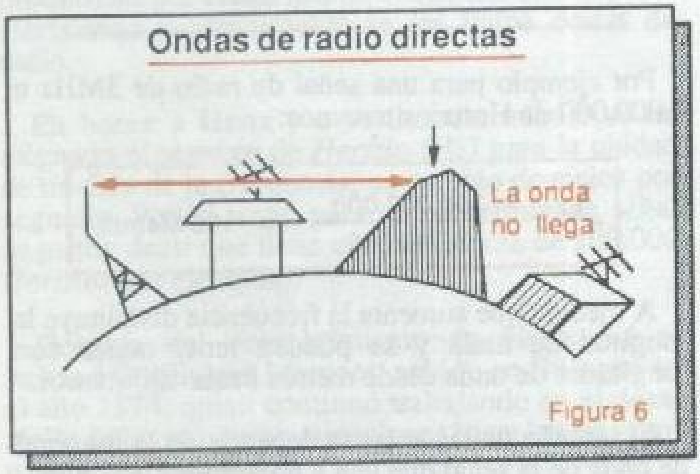

Figura 6

Ondas espaciales o reflejadas

En la atmósfera existe una capa ionizada compuesta por gases conocida con el nombre de ionosfera. Esta capa refleja las ondas de radio de ciertas frecuencias.

Debido a que una antena transmisora irradia ondas electromagnéticas en todas las direcciones, algunas se elevan y chocan contra la ionosfera, siendo reflejadas de nuevo a la tierra. (figura 7).

De esta manera las ondas de radio pueden llegar a un sitio más lejano obteniéndose un mayor alcance.

En las horas de la noche la intensidad de las capas de la ionosfera es mayor, lo cual produce un mayor reflejo de las ondas; por esta razón podemos escuchar en la noche estaciones de radio muy lejanas.

No todas las ondas son reflejadas por la ionosfera, sólo es posible para señales dentro de un rango de frecuencias altas, conocidas también como *ondas cortas*. La banda de las ondas cortas se extiende desde 3 hasta 30 Mhz y tienen su principal utilización en emisoras internacionales de tipo institucional y en la comunicación entre radioaficionados, tema que estaremos tratando ampliamente en este curso.

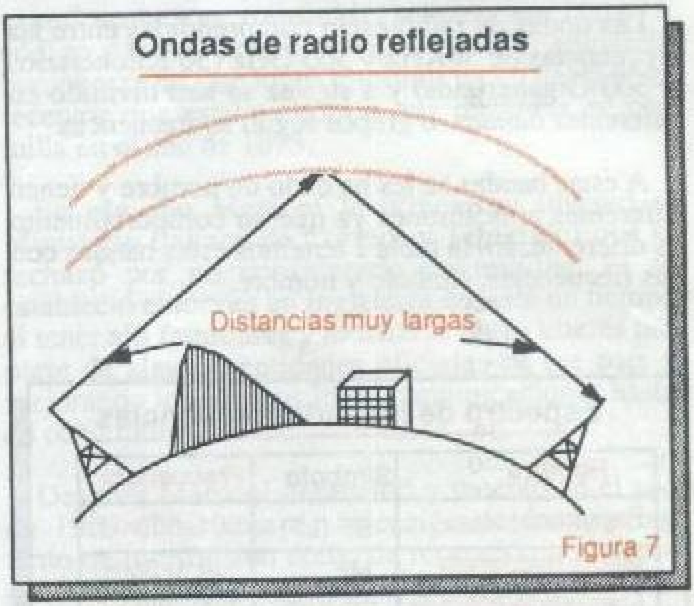

Figura 7

Transmisión y recepción

Las ondas de radio por sí solas no tienen ninguna utilidad práctica. Su descubrimiento aporta el medio o vehículo por medio del cual se podía transportar información o *inteligencia* desde un sitio a otro.

Esta información que inicialmente fué el código Morse, luego sonidos, imágenes fijas, imágenes en movimiento, datos de computador y muchos otros tipos se debe *montar* o *pegar* a las ondas de radio para que viaje con ellas. Las dos señales, una de alta frecuencia y otra de baja frecuencia forman la señal de radio completa. A la primera se le llama la *portadora* y a la segunda la *moduladora*.

El proceso de unión de las dos señales se le llama *modulación* y existen dos métodos principales para hacerlo: modulación en amplitud o AM y modulación en frecuencia o FM. Trataremos cada uno de estos sistemas ampliamente en próximas lecciones de este curso.

De todo lo anterior resumimos que para que haya una radiocomunicación se requiere de un transmisor que produzca las ondas de radio y un receptor que las capte. (figura 8).

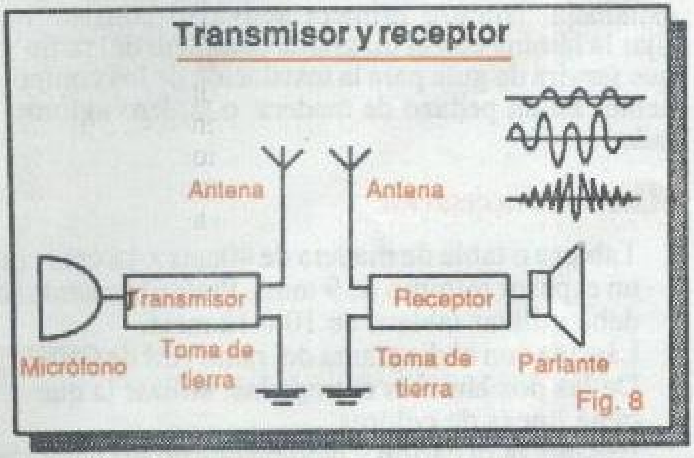
Transmisor y receptor
Antena
Antena
Transmisor
Receptor
Micrófono
Toma de tierra
Toma de tierra
Parlante

Fig. 8

Al transmisor debe estar conectado un micrófono que recibe las ondas sonoras; allí se mezclan con las ondas de radio que salen por la antena y se propagan por el espacio. El receptor tiene otra antena que recoge esas ondas mixtas y obtiene de ellas los sonidos que deben ser una réplica lo más exacta posible a la información original.

Mirando un poco más en el interior de los bloques básicos del transmisor y del receptor de radio tenemos que cada uno tiene a su vez varios elementos fundamentales.

El transmisor tiene cinco componentes principales que son: fuente de sonido, un amplificador para esos sonidos, un oscilador o generador de señales de alta frecuencia, un modulador o mezclador con amplificador y una antena. (figura 9).

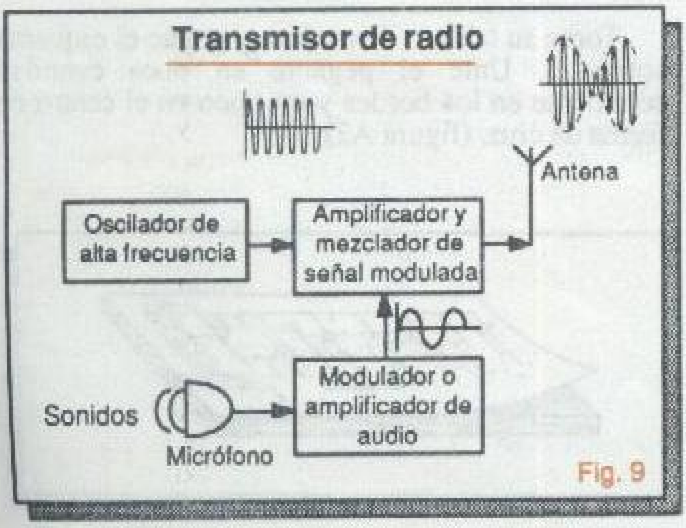
Transmisor de radio
Antena
Oscilador de alta frecuencia
Amplificador y mezclador de señal modulada
Modulador o amplificador de audio
Sonidos
Micrófono

Fig. 9

Los elementos del receptor realizan las funciones inversas al transmisor con el fin de *recuperar* la información transmitida. (figura 10).

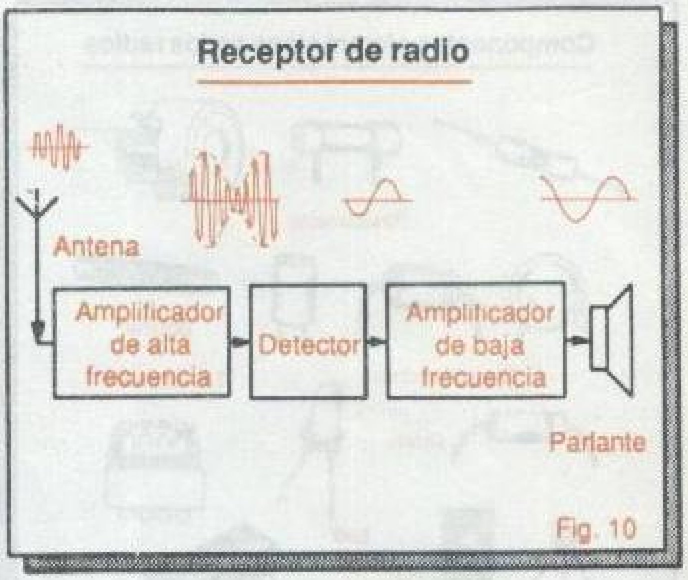
Receptor de radio
Antena
Amplificador de alta frecuencia
Detector
Amplificador de baja frecuencia
Parlante

Fig. 10

En próximas lecciones estudiaremos cómo son los circuitos electrónicos de cada una de estas etapas, sus componentes y cómo funcionan.

Principales componentes electrónicos utilizados en los aparatos de radio

Presentaremos en forma rápida inicialmente los componentes electrónicos que se utilizan para la construcción de los transmisores y de los receptores de radio con el fin de que el lector se vaya familiarizando con ellos.

Más adelante haremos un estudio más profundo de cada uno de ellos con sus diferentes tipos, características, modos para probarlos y toda la información indispensable para este curso.

Conectando estos componentes en configuraciones diferentes se obtienen los distintos circuitos de los aparatos de radio.

Los principales componentes utilizados en los circuitos de radio son: las resistencias, los potenciómetros o resistencias variables, los condensadores en sus tres tipos, de cerámica, electrolíticos y variables, las bobinas, los transformadores, los diodos, los transistores, los circuitos integrados, los micrófonos y los parlantes.

Podemos observar estos componentes en la figura 11.

Para citar un ejemplo sencillo mostramos en la figura 12 un circuito de un receptor muy simple formado por muy pocos componentes. Este receptor recibe el nombre de " radio de galena " y estaremos experimentando próximamente con él.

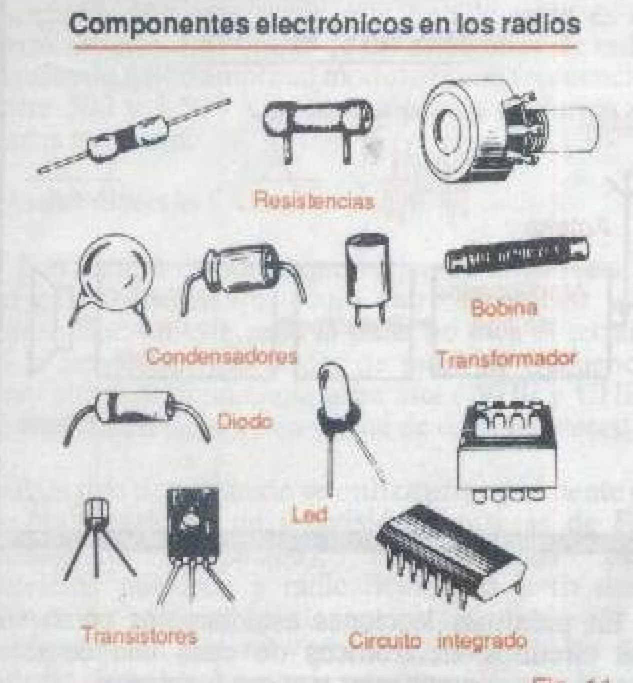
Componentes electrónicos en los radios
Resistencias
Condensadores
Bobina
Transformador
Diodo
Led
Transistores
Circuito integrado

Fig. 11

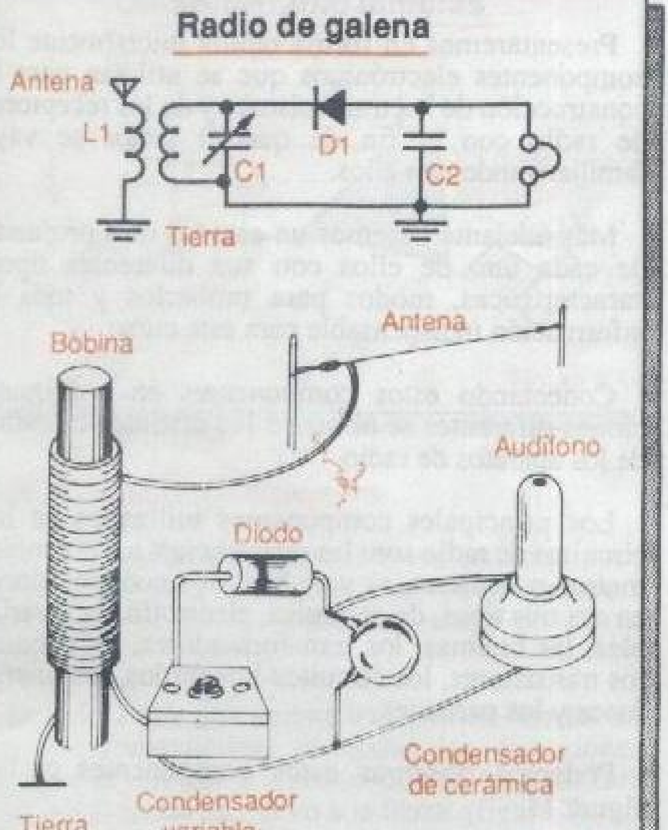
Radio de galena
Antena
L1
C1
D1
C2
Tierra
Antena
Bobina
Audífono
Diodo
Condensador de cerámica
Condensador variable
Tierra

Fig 12

Actividad práctica Nº 1

Durante este curso de radio y comunicaciones iremos ensamblando un radio de AM con seis transistores con el fin de practicar y comprobar la teoría estudiada. Nuestra primera actividad consiste en fijar la lámina con el diagrama completo del radio y que servirá de guía para la instalación de los componentes en un pedazo de madera o tablero aglomerado .

Materiales necesarios:

1 Tablero o tabla de madera de 40cms x 18 cms con un espesor mínimo de 9 mms. Preferiblemente se debe utilizar tablero de 10 o 12 mms.

1 Lámina con el diagrama del radio AM de CEKIT. De las dos láminas entregadas, utilizar la que tiene líneas de colores.

Pegante para cartón y madera que no sea demasiado líquido.

Paso 1. Cómo preparar el diagrama

Asegúrese de que el tamaño del tablero sea igual al diagrama o plano esquemático y recorte en este último, el papel sobrante.(figura A1).

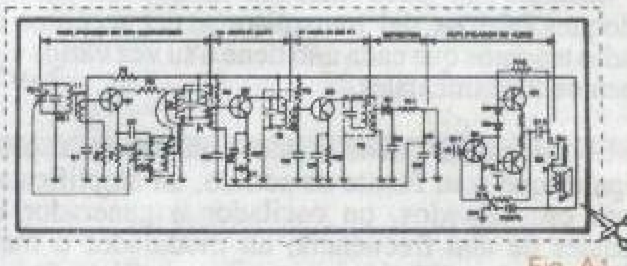

Fig. A1

Paso 2. Pegar la lámina sobre el tablero

Tome su tablero de madera y pegue el esquema sobre él. Unte el pegante en poca cantidad solamente en los bordes y un poco en el centro en forma de cruz. (figura A2).

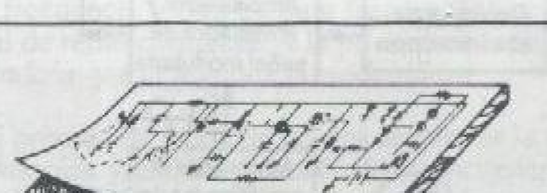

Fig. A2

La materia y los electrones

En la introducción anterior estudiamos en forma general el fenómeno de las ondas, especialmente las ondas de radio y su aplicación en las radiocomunicaciones. Estudiamos también la estructura básica de un transmisor y de un receptor de radio,enumerando los componentes electrónicos que se utilizan en su fabricación.

Se mencionó en esta introducción que las ondas de radio se producen al aplicar *corriente eléctrica* a un circuito, que éste las entrega a una *antena* y que por ella salen al espacio exterior propagándose de diferentes formas según su *frecuencia* .

Antes de continuar con el estudio de la teoría de los aparatos de radio es necesario conocer qué es la corriente eléctrica, de dónde sale, cuál es su comportamiento, cuál es su forma y otros aspectos importantes,ya que ella interviene y hace posible todos los procesos de las radiocomunicaciones.

Para entender mejor qué es la corriente eléctrica o *electricidad* , debemos estudiar qué es la materia y cuál es su estructura interna.

Qué es la materia

Todo lo que se puede ver, sentir o usar es materia. En realidad, la materia es cualquier cosa que tenga peso y que ocupe espacio. Esta puede encontrarse en forma de un sólido, un líquido o un gas. (figura 13).

Formas de la materia

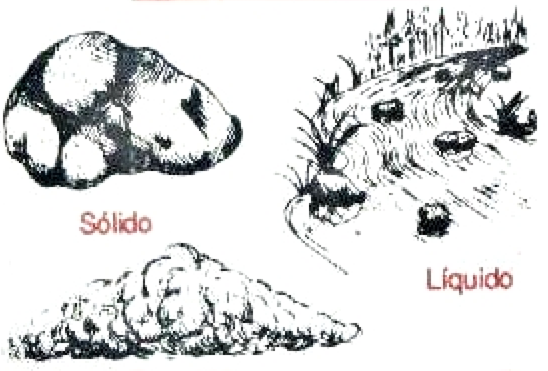

Fig. 13

Las rocas, la madera y el metal son formas de materia sólida; el agua, el alcohol, o la gasolina son líquidos y el oxígeno, el hidrógeno y el bióxido de carbono son gases.

Los elementos

Los elementos son la materia prima básica que constituye toda la materia y se encuentran en forma natural en la tierra. El oxígeno y el hidrógeno son elementos, lo mismo que el aluminio, el cobre, la plata, el oro y el mercurio. En efecto, existen poco más de 103 elementos conocidos, 92 de los cuales son naturales y los demás son artificiales, o hechos por el hombre.

Los compuestos o materiales

Los elementos básicos pueden combinarse para producir materiales cuyas características son totalmente distintas de las que tienen los elementos en su estado natural. Desde luégo, hay mucho más materiales que elementos.

El agua, por ejemplo, es un compuesto formado por los elementos hidrógeno y oxígeno; la sal de mesa ordinaria está formada por los elementos sodio y cloro.

Cabe anotar que el hidrógeno y el oxígeno, aunque son gases, pueden producir el agua que es un líquido.

La molécula

La molécula es la partícula más pequeña a la que puede reducirse un material, conservando sus características propias como tal,antes de descomponerse en sus elementos básicos.

Por ejemplo, si se tomara un grano de sal de mesa y se la dividiera sucesivamente en dos, hasta obtener el trocito más pequeño posible y que siga siendo sal, este trocito sería una molécula de sal. Si logramos descomponer la molécula por algún método, desaparece la sal como tal y aparecen sus elementos básicos, el sodio y el cloro.

El átomo

El átomo es la partícula más pequeña a la que se puede reducir un elemento,conservando las propiedades de ese elemento.

Si una gota de agua se reduce a su tamaño mínimo, se producirá una molécula de agua.

Pero si la molécula se redujera aún más, aparecerían átomos de hidrógeno y oxígeno. (figura 14).

Moléculas y átomos

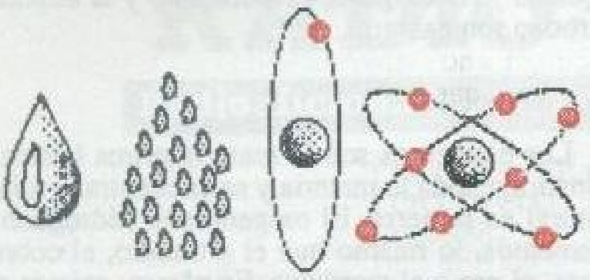

Fig. 14

Estructura del átomo

Ahora bien, si el átomo de un elemento se divide aún más, este elemento como tal deja de existir y aparece formado por varios tipos de partículas. Estas partículas más pequeñas que el átomo y que resultan de su división, se encuentran presentes en todos los átomos de los diferentes elementos.

El átomo de un elemento difiere del átomo de otro elemento, sólo en virtud de que los dos contienen números diferentes de estas pequeñísimas partículas.

Básicamente, un átomo está formado por tres tipos de partículas subatómicas que son de interés en el estudio de la electricidad: electrones, protones y neutrones.

Los protones y neutrones se localizan en el centro, formando el núcleo del átomo y los electrones giran en órbitas alrededor del núcleo. Esta disposición de los átomos es similar al sistema de los planetas que giran alrededor del sol en nuestro sistema solar, como se puede apreciar en la figura 15 .

El núcleo

El núcleo es la parte central de un átomo y está formado por dos tipos de partículas : los protones y los neutrones. El número de protones en el núcleo determina la forma en que el átomo de un elemento difiere de otro.

Por ejemplo, el núcleo de un átomo de hidrógeno contiene un protón, el del oxígeno contiene 8, el del cobre 29, el de la plata 47 y el del oro 79. De hecho, ésta es la forma en que se identifican los diferentes elementos, es decir, por sus números atómicos. Así que el *número atómico* de un elemento es el número de protones que contiene cada átomo en su núcleo.

Estructura del átomo

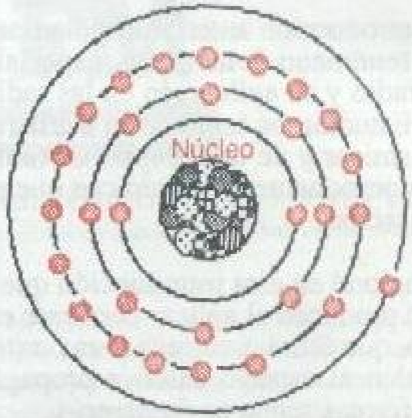

Fig. 15

El protón es muy pequeño. Se estima que tiene 1.778 trillonésimas de milímetro de diámetro; el protón mide la tercera parte del diámetro de un electrón, pero tiene casi 1.840 veces su masa ; es decir, el protón es casi 1.840 veces más pesado que el electrón.

Es muy difícil desalojar el protón del núcleo de un átomo. Por lo tanto, en la teoría eléctrica, se considera que los protones son partes permanentes del núcleo.

Los protones no toman parte activa en el flujo o transferencia de energía eléctrica. *El protón tiene una carga eléctrica positiva.*

El electrón

Según se ha explicado anteriormente, el electrón tiene un diámetro tres veces mayor que el del protón, o sea, aproximadamente 5.588 trillonésimas de milímetro; pero es 1.840 veces más liviano que el protón.

Los electrones son más fáciles de mover y son las partículas que participan activamente en el flujo o transferencia de energía eléctrica.

Los electrones giran en órbitas alrededor del núcleo de un átomo y tienen cargas eléctricas negativas.

Ley de las cargas eléctricas

La carga negativa de un electrón es igual, pero opuesta, a la carga positiva de un protón.

Las cargas de un electrón y un protón se llaman cargas electrostáticas. Las líneas de fuerza asociadas con cada partícula producen campos electrostáticos. Debido a la forma en que interactúan estos campos, las partículas cargadas pueden atraerse o repelerse entre sí. (figura 16).

Ley de las cargas eléctricas

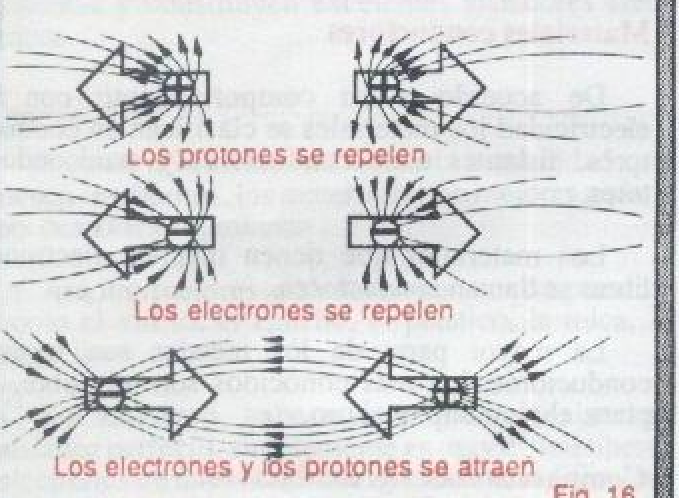
Los protones se repelen
Los electrones se repelen
Los electrones y los protones se atraen

Fig. 16

La ley de las cargas eléctricas dice que las partículas que tienen cargas del mismo tipo se repelen, y las que tienen cargas diferentes, se atraen.

Un protón (+) repele a otro protón (+).
Un electrón (-) repele a otro electrón (-)
Un protón (+) atrae a un electrón (-).

Cargas atómicas

Normalmente, un átomo contiene el mismo número de electrones y protones de manera que las cargas iguales y opuestas, es decir, las negativas y positivas,se equilibran entre sí y hacen que el átomo sea eléctricamente neutro.

Ahora bien, según ya se explicó, lo que le da al átomo de un elemento sus propiedades características,es el número de los protones que tiene en el núcleo; pero el número de electrones en diferentes tipos de átomos puede variar.

Si un átomo contiene menos electrones que protones, tendrá una carga positiva. Si tiene más electrones que protones tendrá una carga negativa. Los átomos cargados reciben el nombre de iones.

Materiales cargados eléctricamente

Cuando en un trozo de material eléctricamente neutro,muchos átomos pierden o ganan electrones, el material quedará cargado.

Hay muchas maneras de producir estos cambios en los átomos, según se explica más adelante. *El método que descubrieron los antiguos griegos fue el de la fricción*. Por ejemplo, si se frota una varilla de vidrio con un trozo de seda, la varilla de vidrio le donará algunos electrones a la seda. La varilla de vidrio se cargará positivamente y la seda quedará cargada negativamente. *A este fenómeno se le ha llamado electricidad estática.* (figura 17).

Cargas eléctricas en un material

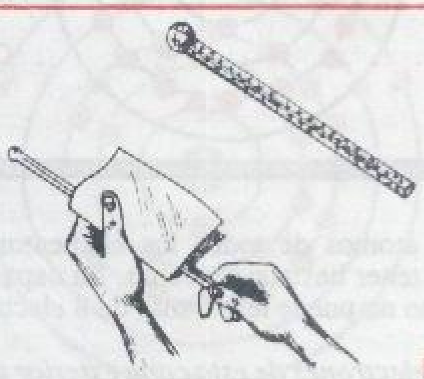

Fig. 17

Orbitas de los electrones

Según se ha visto, la electricidad se produce cuando los electrones salen de sus átomos.

Para entender los distintos métodos usados para lograrlo, se necesita saber algo más acerca de la naturaleza de las diferentes órbitas electrónicas que rodean el núcleo de un átomo.

Los electrones giran en sus órbitas a gran velocidad alrededor del núcleo del átomo. Debido a la gran velocidad del electrón, la fuerza centrífuga tiende a sacar al electrón de su órbita.

Por otra parte, la atracción positiva del núcleo impide que se escape el electrón.

Sin embargo, si se aplica suficiente fuerza externa para ayudar a la fuerza centrífuga, puede *liberarse* el electrón.

Capas orbitales

Los electrones que giran cerca del núcleo son difíciles de liberar debido a su proximidad a la fuerza positiva que los sujeta. Cuanto más lejos se encuentren los electrones del núcleo, más débil será la fuerza positiva que ejerce el núcleo.

Como ya lo habrán notado en algunas figuras anteriores, mientras más electrones tiene un átomo, mayor será el número de sus órbitas. Estas trayectorias orbitales comúnmente se llaman capas. (figura 18).

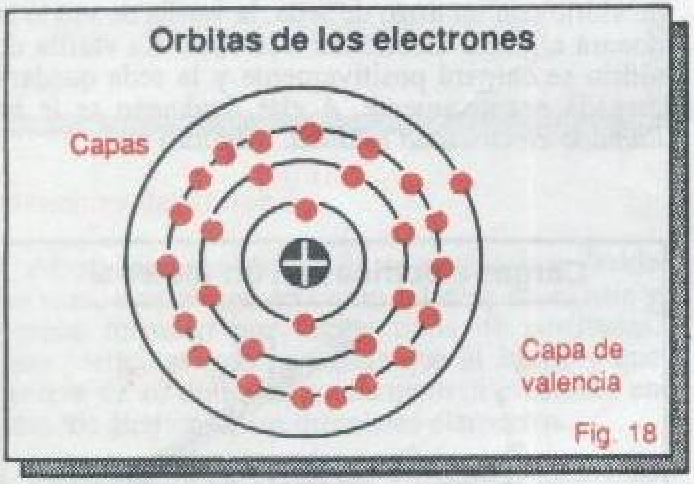

Fig. 18

Los átomos de todos los elementos conocidos pueden tener hasta siete capas. La capa exterior de un átomo no puede tener más de 8 electrones.

Los electrones de esta capa exterior de un átomo reciben el nombre de electrones de valencia .

El número de electrones en la capa de valencia de un átomo es muy importante en la electricidad, como se verá más adelante.

Energía del electrón

Aunque todo electrón tiene la misma carga negativa, no todos los electrones tienen el mismo nivel de energía.

Los electrones cuya órbita está próxima al núcleo contienen menos energía que los que se encuentran en órbitas externas. Cuanto más lejanas estén las órbitas electrónicas del núcleo, mayor será su energía.

Electrones libres

Si se añade suficiente energía a un electrón, saldrá fuera de su órbita, hacia la órbita inmediatamente superior. Y, si se aplica suficiente energía a un electrón de valencia, el electrón se desligará de su átomo, ya que no existe una órbita inmediata superior. *A estos electrones se les llama electrones libres.*

Cuándo se produce la electricidad

La electricidad se produce cuando los electrones se liberan de sus átomos y quedan libres. Puesto que los electrones de valencia son los más alejados de la fuerza atractiva del núcleo y además tienen el nivel de energía más alto, son los que pueden liberarse más fácilmente.

A este tipo de electricidad se le llama electricidad estable o electricidad estática.

La electricidad estática se manifiesta,por ejemplo, en la carga que toman las nubes con su movimiento y que se descarga por medio de los rayos.

Materiales conductores

De acuerdo a su comportamiento con la electricidad los materiales se clasifican en conductores, aislantes o no conductores y semiconductores.

Los materiales que tienen muchos electrones libres se llaman *conductores.*

La mayor parte de los metales son buenos conductores; los más conocidos son: el cobre, la plata, el aluminio y el oro.

Comparación de los conductores

Algunos metales son mejores conductores que otros y aunque los átomos de cobre, plata y oro tienen, cada uno, un solo electrón de valencia que puede liberarse fácilmente, la plata es el mejor conductor.

Le sigue el cobre y luégo el oro. Esto se debe al hecho de que en una misma cantidad de material, la plata tiene más atomos que los demás metales y por consiguiente, se dispone de un mayor número de electrones libres.

Sin-embargo, como el cobre es más barato que la plata,es el que más se utiliza en todo tipo de aplicaciones eléctricas.

Aislantes

Los aislantes o aisladores son materiales que no dejan que sus electrones se liberen fácilmente. Los átomos de los aisladores tienen capas de valencia que están llenas con 8 electrones o,bien llenas,o más de la mitad.

Cualquier energía que se aplique a uno de estos átomos, se distribuirá entre un número de electrones relativamente grande.

Además, estos átomos se resisten a desprenderse de sus electrones, debido a un fenómeno que se conoce como estabilidad química.

Un átomo es completamente estable cuando su capa exterior está completamente saturada, o cuando tiene 8 electrones de valencia.

Así pues, no solamente es difícil liberar a sus electrones, sino que los átomos de los aisladores también se opondrán a la producción de electricidad, debido a su tendencia a atrapar a todos los electrones que puedan ser liberados.

Los átomos con siete electrones de valencia, son los que tratan más activamente de llenar la capa de valencia y constituyen excelentes aisladores eléctricos.

Compuestos aislantes

Debido a la tendencia hacia la estabilidad que tienen los átomos, los materiales son mejores aislantes cuando se combinan.

Así, los mejores aisladores son compuestos, como el vidrio, el caucho, el plástico, la mica, la baquelita y otros.

Sin embargo, cabe anotar que no existe un aislador perfecto, simplemente es muy difícil liberar electrones de tales materiales, y en unos la dificultad es mayor que en otros, convirtiéndose en mejores aislantes.

Los aislantes se utilizan para recubrir los conductores de cobre, con el fin de que no se tenga contacto directo con la corriente eléctrica en el caso de altos voltajes.

Estos materiales se utilizan con mucha frecuencia para fabricar una gran cantidad de componentes electrónicos, como suiches, bornes de conexión, perillas para aparatos electrónicos, y, como soportes para las resistencias, los condensadores, las bobinas, los transistores, etc.

En la figura 19 podemos observar la estructura atómica de un conductor y de un aislante.

Los semiconductores

Puesto que los conductores tienen sus capas de valencia llenas hasta la mitad, y los aisladores tienen las suyas llenas a más de la mitad, las sustancias que tienen átomos con cuatro electrones de valencia reciben el nombre de *semiconductores*. Estos conducen mejor que los aislantes, pero no tan bien como los conductores.

Conductores y aislantes

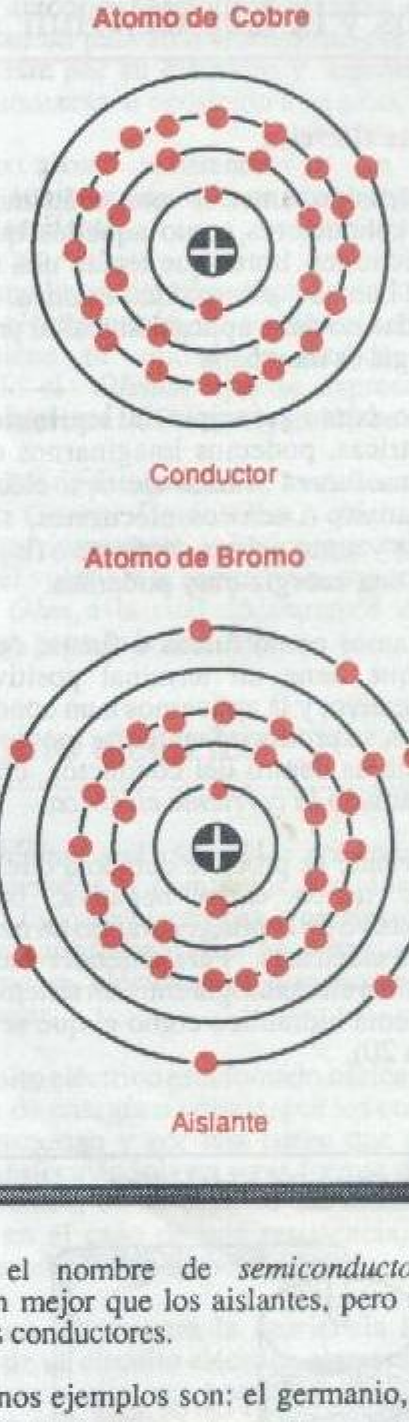

Fig. 19

Algunos ejemplos son: el germanio, el silicio y el selenio.

Los semiconductores son de mucha importancia en la electrónica, ya que con ellos se fabrican los diodos, los transistores y los circuitos integrados, que son los componentes que han revolucionado toda la práctica de la electrónica, debido a su pequeño tamaño, bajo costo y gran confiabilidad.

Estaremos estudiando los semiconductores, los componentes fabricados con ellos y sus aplicaciones en los circuitos de radio, en próximas lecciones.

La corriente eléctrica, los circuitos y la Ley de Ohm

La corriente eléctrica

En la lección anterior nos referimos a los materiales conductores, como aquellos que tenían muchos electrones libres que tenían una forma de electricidad llamada *electricidad estática.* Este tipo de electricidad no tiene aplicación real ni práctica,ya que su energía es muy baja.

Tomando como principio la ley básica de las cargas eléctricas, podemos imaginarnos que, si se aplica alguna fuerza similar de tipo eléctrico, las cargas negativas o sea los electrones, se pueden hacer mover y crear así una corriente o flujo de electrones, con una energía muy poderosa.

Si utilizamos como fuerza o fuente de energía una pila, que tiene un terminal positivo y un terminal negativo, y la aplicamos a un conductor en sus extremos, se producirá un fuerte movimiento de electrones libres dentro del conductor, creando lo que se ha llamado *la corriente eléctrica*.

Esta corriente se produce debido a que los electrones, que tienen carga negativa, buscan el terminal positivo de la pila,generando un movimiento de cargas eléctricas. Para entender mejor esto, estableceremos una analogía entre un sistema eléctrico y un sistema hidráulico como el que se muestra en la (figura 20).

Circuito hidráulico

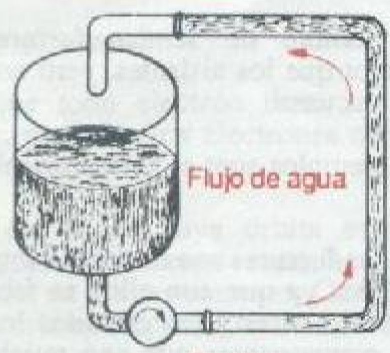

Fig. 20

Para que el agua pueda fluír por la tubería, se requiere que haya una fuerza aplicada por una bomba. En un circuito eléctrico,para que la corriente pueda fluír por los conductores,se requiere un voltaje o tensión que empuje los electrones; la acción del voltaje sobre el circuito eléctrico es similar a la acción de la bomba en el circuito hidráulico.

La unidad de medida de la corriente eléctrica es el *amperio.* Este nombre se ha dado en honor del físico y matemático francés André Marie Ampere, quién vivió entre 1775 y 1836 . Sus experimentos sobre la corriente eléctrica y el magnetismo,fueron uno de los primeros trabajos que se hicieron al respecto, y que ayudaron a conocer la verdadera naturaleza de estos importantes fenómenos.

Cuando decimos que por un conductor circula una corriente de un amperio, significa que por dicho conductor pasan 6.300.000.000.000.000.000 de electrones libres durante un segundo.

Esta forma de medir la corriente es comparable a la medida de "litros por segundo", que se utiliza para medir la cantidad de agua que circula por una tubería en ese período de tiempo. La corriente eléctrica se representa en los diagramas con la letra I, ya que también recibe el nombre de *Intensidad de corriente*.

El voltaje

Se habla de terminal positivo de una pila porque tiene un valor mayor de carga que el terminal negativo. Esta diferencia de cargas se ha llamado fuerza electromotriz, tensión, diferencia de potencial o *voltaje,* que es el término más común utilizado actualmente. La unidad de medida del voltaje es el *Voltio.* Esta unidad se ha llamado así en honor al científico italiano Alessandro Volta,quién inventó la primera pila eléctrica. Existen diferentes formas de producir o *generar* un voltaje; mencionaremos aquí las más utilizadas:

a) Por un proceso químico, introduciendo dos electrodos de metales diferentes en una solución química. Este es el caso de los diferentes tipos de pilas que existen en el mercado.

b) Mecánicamente, utilizando el proceso de inducción magnética por medio de los generadores o plantas eléctricas.

c) El método de la fricción, para producir electricidad estática,haciendo frotar un material como la ebonita contra un paño de lana.

d) Por medio del efecto piezoeléctrico, utilizando la propiedad que tienen algunos cristales como el cuarzo, que producen un voltaje cuando se someten a una presión mecánica.

e) Mediante el fenómeno fotovoltaico, caso de las pilas solares, en donde se produce un voltaje cuando le llega luz a un material especial como el selenio.

Los métodos más utilizados en la práctica para generar un voltaje utilizable,son los dos primeros y los estaremos estudiando próximamente, cuando tratemos el tema de la corriente continua y de la corriente alterna.

Además de la corriente y el voltaje, encontramos en los sistemas eléctricos un fenómeno llamado *resistencia*.

La resistencia

La *resistencia* es cierta dificultad que encuentran los electrones para circular libremente por los conductores. (figura 21). Esta dificultad está relacionada con el diámetro de los conductores, la longitud y el material del cual están conformados,como ya habíamos visto al estudiar la estructura atómica de los conductores.

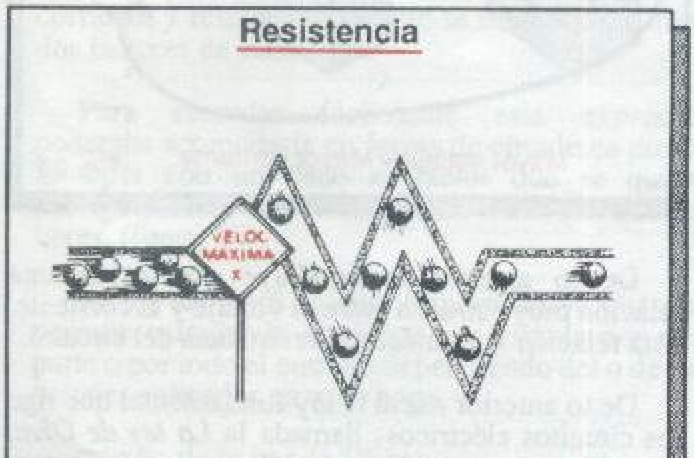

Fig. 21

Podemos comparar este fenómeno con un grupo de vehículos que fluyen por una carretera.

Si la carretera es muy angosta, los vehículos encuentran mucha dificultad en fluir por ella, lo mismo si tiene muchas curvas o si se encuentra en muy mal estado. Una carretera en esas condiciones ofrece resistencia al flujo de vehículos.

En un circuito hidráulico, el agua encuentra cierta dificultad para fluir libremente por los tubos, especialmente por su diámetro y curvas, los que ofrecen *resistencia* u oposición a su paso.

Por lo tanto, resistencia es en términos eléctricos, la oposición que un cuerpo presenta al paso de la corriente.

Así como la corriente se expresa en amperios y la fuerza electromotriz en voltios, la resistencia también tiene su unidad de medida y se ha denominado el *Ohmio,* que se expresa generalmente con la letra Ω (Omega) del alfabeto griego.

El nombre de esta unidad se ha dado en honor al gran físico *George Simón Ohm,* quien descubrió una ley muy importante y básica que relaciona la corriente, el voltaje y la resistencia, conocida como la *Ley de Ohm,* a la cual dedicaremos un estudio completo un poco más adelante.

El circuito eléctrico

Un circuito está formado por uno o más caminos cerrados o completos, por donde puede circular un flujo de electrones o corriente eléctrica para realizar un trabajo. Sin un circuito no puede haber corriente eléctrica, ya que no tendría por dónde circular.

El circuito eléctrico está fomado básicamente por una fuente de energía o voltaje, por los conductores que la transportan y por una *carga* que utiliza esa energía,transformándola en otras formas de energía, como mecánica, en el caso de un motor, energía calorífica en el caso de una resistencia,o energía luminosa en el caso de un bombillo.

La figura 22 muestra la apariencia física y el diagrama de un circuito eléctrico elemental conformado por una pila, un bombillo pequeño y los conductores o alambres.

La pila es la fuente de energía, los conductores, llevan la corriente y el bombillo es la carga.

Si se interrumpe cualquiera de los conductores se suspende la circulación de la corriente por el circuito.

Debemos tener en cuenta que en todos los circuitos eléctricos y electrónicos, sin importar su complejidad, tienen tres factores asociados a ellos. Estos son: Corriente, Voltaje y Resistencia.

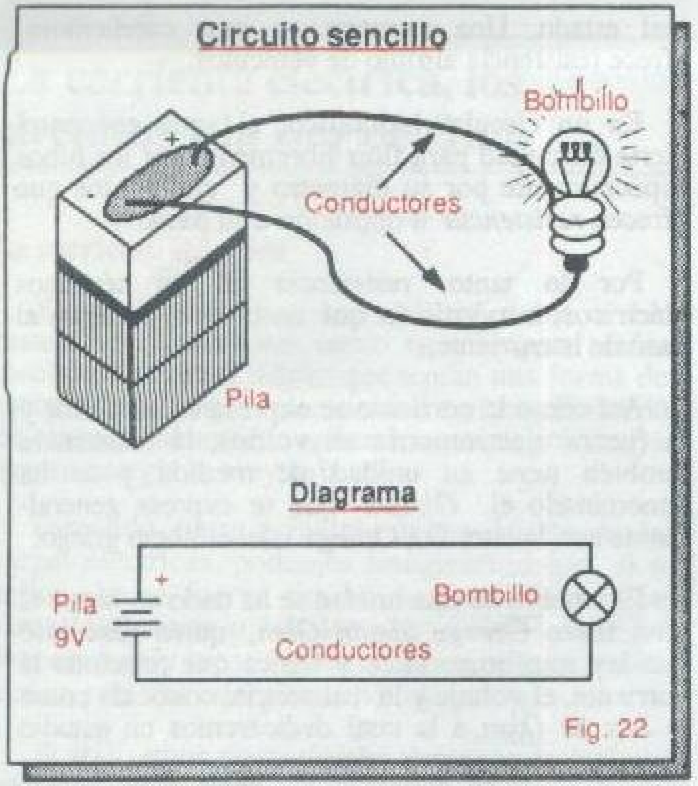

Fig. 22

La corriente o electricidad es la cantidad de electrones que fluyen a través de los conductores, el voltaje es la fuerza que hace que fluya esta corriente, y la resistencia se opone al paso de la corriente.

De lo anterior se deduce que en un circuito sometido a una tensión o voltaje constante,se puede lograr que fluyan distintos valores de corriente, modificando en cada caso el valor de la resistencia de la carga.

Si la resistencia es alta,la oposición es mayor y por tanto fluye menos corriente, y ante una resistencia baja,la corriente es alta.

El circuito que se describió anteriormente es el más elemental de todos. Existen circuitos muchísimo más complejos, con una enorme variedad de elementos que crean numerosos caminos para el flujo de la corriente.

Estos circuitos se forman con los componentes enumerados en la lección anterior, como son:las resistencias, los condensadores, las bobinas, los diodos, los transistores y otros, en diferentes configuraciones o conexiones, como son los circuitos en serie, en paralelo y mixtos,o sea en serie-paralelo.

A su vez todos los aparatos electrónicos como un radio, un televisor, un equipo de sonido, un computador, un satélite, etc., están constituidos por varios circuitos con diferentes formas y funciones. Un circuito eléctrico por muy complejo que sea, siempre estará basado en otros circuitos más sencillos.

La ley de Ohm

Hasta ahora se ha hecho mención a los tres elementos básicos en todo circuito eléctrico: Voltaje, Corriente y Resistencia, pero no hemos visto cómo se relacionan entre sí.

En los circuitos eléctricos es muy fácil modificar el valor de los elementos básicos,como la corriente que circula por él. Esto lo podemos lograr variando el voltaje o la resistencia del circuito.

Si hacemos por ejemplo que el voltaje aumente, ya sea agregando una pila más, obtendremos una mayor corriente circulante. Por otra parte, si dejamos el voltaje constante, pero de algún modo aumentamos la resistencia del circuito, notaremos que la corriente disminuye. (figura 23)

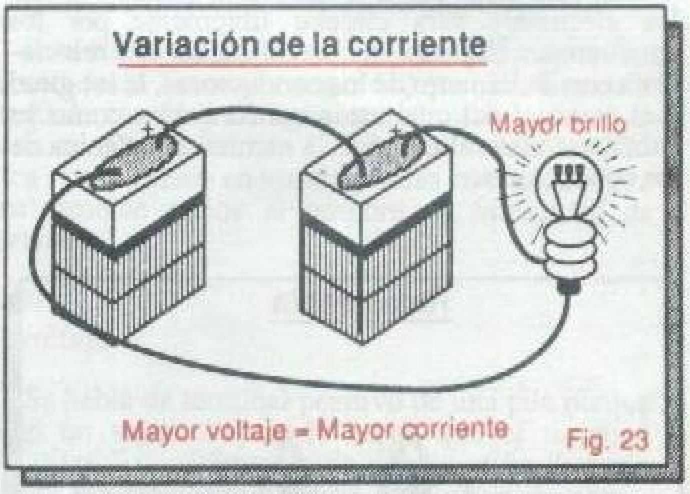

Fig. 23

De lo anterior se concluye que existe una relación muy estrecha entre el voltaje y la corriente. Esta relación la establece la resistencia del circuito.

De lo anterior nació la ley fundamental que rige los circuitos eléctricos, llamada la *La ley de Ohm*, descubierta y enunciada por el físico alemán George Simon Ohm, quien relacionó en una sencilla fórmula matemática a los tres elementos fundamentales en todo circuito eléctrico.

La ley de Ohm permite encontrar uno de los tres factores que no se conoce si tenemos el valor de los otros dos.Así, si en un circuito conocemos el voltaje dado en voltios y la resistencia en ohmios, podemos hallar la corriente en amperios.

La Ley de Ohm dice:

En todo el circuito electrónico, la corriente que circula por él, es directamenente proporcional al voltaje aplicado, e inversamente proporcional a la resistencia propia del circuito.

Su expresión matemática es la siguiente:

$$V = I \times R$$

En donde:

V: Voltaje o tensión expresado en Voltios

I: Corriente expresada en Amperios

R: Resistencia expresada en Ohmios (Ω).

En otras palabras:

VOLTIOS = AMPERIOS x OHMIOS

De esta fórmula general podemos obtener las fórmulas para calcular los otros dos valores, el de corriente y resistencia, cuando se conocen los otros dos factores en cada caso.

Para recordar fácilmente esta expresión podemos acomodarla en forma de círculo, en donde al tapar con un dedo el factor que se quiere averiguar, basta con visualizar las letras que quedan libres. (figura 24).

La ley de Ohm es muy importante, ya que nos permite calcular la corriente que circula por una parte o por todo el circuito, dependiendo del o de los voltajes aplicados en cada parte.

Utilización de la ley de Ohm

Una de las principales aplicaciones para la ley de Ohm es determinar cuánta corriente circula por los conductores de un circuito, con el fin de utilizar el diámetro correcto para este conductor.

Si no se hace ésto, y se utiliza un conductor o cable más delgado, éste introduce una resistencia adicional al circuito y la corriente disminuye generando pérdidas de energía.

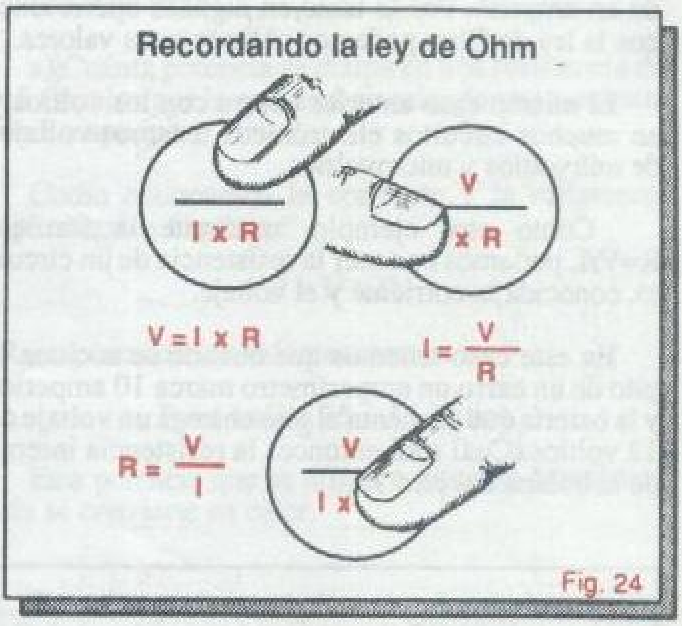

Fig. 24

La ley de Ohm nos permite calcular con facilidad en un circuito el voltaje, la corriente y la resistencia, usando las respectivas fórmulas así:

Tomaremos como primer ejemplo el circuito sencillo de la figura 22. Si tenemos conectada una pila de 9 voltios y el bombillo tiene una resistencia en su filamento de 30 Ohmios, podemos averiguar cuánta corriente en amperios circula por el circuito. Utilizando la ley de Ohm tenemos :

$$I = \frac{V}{R}$$

$$I = \frac{9 \text{ Voltios}}{30 \text{ Ohmios}} = 0.3 \text{ Amperios}$$

Por lo tanto, la corriente que circula por el circuito es de 300 miliamperios

Observando el ejemplo anterior vemos que se ha utilizado la palabra *miliamperios*.

Esto se debe a que en los circuitos y sobre todo en electrónica, se tienen corrientes muy pequeñas que son mucho menores que la unidad del Amperio. Estas unidades se llaman submúltiplos y las más empleadas son los miliamperios y los microamperios. Un miliamperio es la milésima parte de un amperio, y un microamperio es la millonésima parte de un amperio. Por lo tanto, en algunas operaciones con la ley de Ohm podemos obtener estos valores.

El mismo caso anterior ocurre con los voltios, y, en muchos circuitos electrónicos tenemos voltajes de milivoltios y microvoltios.

Como otro ejemplo: mediante la fórmula R=V/I, podemos conocer la resistencia de un circuito, conocida la corriente y el voltaje.

En este caso tenemos que cuando se acciona el pito de un carro, un amperímetro marca 10 amperios y la batería que alimenta al pito entrega un voltaje de 12 voltios. ¿Cuál será entonces la resistencia interna de la bobina del pito ?

$$R = \frac{V}{I}$$

$$R = \frac{12 \text{ Voltios}}{10 \text{ Amperios}}$$

$$R = 1.2 \text{ Ohmios}$$

Debido a que esta resistencia es muy baja, la corriente que consume el pito es alta, y, si se deja accionado en forma permanente, se puede descargar rápidamente la batería.

En otro ejemplo: podemos hallar también el voltaje con que se está alimentando un radio portátil, si conocemos que su consumo de corriente es de 200 miliamperios ó 0.2 Amperios y que la resistencia interna es de 45 Ohmios.

$$V = I \times R$$

$$V = 0.2 \text{ Amperios} \times 45\ \Omega$$

$$V = 9 \text{ Voltios}$$

Potencia eléctrica

La potencia es el factor que resume el trabajo total que realiza un circuito; por ejemplo, producir luz, irradiar calor, emitir sonidos, producir movimientos, etc.

La potencia o trabajo en un circuito eléctrico, resulta de la acción de una corriente eléctrica ante la presencia de una fuerza electromotriz o voltaje. Una batería sola no realiza ningún trabajo, se requiere que tenga conectada una carga para que haya flujo de corriente. La potencia eléctrica se mide en *Vatios*

De esta manera se obtiene la energía total que consume o que transforma un aparato. En términos de energía sabemos que ésta no se produce, ni se crea ni se destruye, solamente se transforma. En cualquier circuito eléctrico o electrónico, por sencillo o complejo que sea, lo que se realiza es una transformación de un tipo de energía en otro.

Así se menciona, por ejemplo, que un bombillo es de 100 vatios. Lo que se está diciendo es que toma esa energía de una fuente de electricidad y la transforma, parte en energía luminosa y parte en energía calorífica.

Un motor por ejemplo, toma energía eléctrica y la transforma en energía mecánica o movimiento.

En un circuito de un radio, éste toma energía de unas pilas o de un tomacorriente y la transforma en energía sonora, captando las ondas de radio y tomando la información de los sonidos que ellas traen.

Para determinar el valor de la potencia en un circuito se multiplica la corriente que circula por el voltaje aplicado a ese circuito, así :

$$P = V \times I$$

Donde:

P = Potencia en vatios

V = Voltaje en voltios

I = Corriente en amperios

Con frecuencia se utilizan múltiplos y submúltiplos para expresar la potencia.

Por ejemplo: el MILIVATIO que equivale a una milésima parte de un Vatio (1/1000); o el KILOVATIO que equivale a mil vatios (1000 Vatios). A menudo se utiliza la letra W para indicar los Vatios.

Veamos un ejemplo que nos ayude a entender mejor: Calculemos la potencia de un bombillo de un automóvil que requiere 5 amperios y se alimenta con un voltaje de 12 voltios.

$$P = V \times I$$

$$P = 12 \text{ voltios} \times 5 \text{ amperios}$$

$$P = 60 \text{ Vatios}$$

Lo anterior quiere decir que el reflector disipa 60 vatios de potencia.

Existe otra forma de calcular la potencia, partiendo del voltaje y la resistencia,o de la corriente y la resistencia. Esto se logra reemplazando cada factor por su equivalente,tomado de la ley de Ohm; así tenemos:

$$P = V \times I$$

De la ley de Ohm tenemos:

$$V = I \times R \quad \text{e} \quad I = \frac{V}{R}$$

Reemplazando:

1) $P = I \times R \times I = I^2 \times R$

$\boxed{P = I^2 \times R}$ Potencia en función de la corriente y la resistencia

2) $P = V \times \frac{V}{R} = \frac{V^2}{R}$

$\boxed{P = \frac{V^2}{R}}$ Potencia en función del voltaje y la resistencia

Con lo anterior podemos averiguar el trabajo realizado por un circuito,conociendo su resistencia y la corriente o el voltaje. Ejemplo:

Un equipo de sonido o amplificador de audio entrega en su salida una señal que tiene una tensión o voltaje de 20 Voltios.

¿Cuánta potencia entregará en sonido a un parlante de 8 Ω conectado en su salida ?

La potencia que entrega el parlante es de 50 W (vatios).

Otros ejemplos

a)¿Cuánta potencia se disipa en una resistencia de 15 Ohmios por la que está circulando una corriente de 1 amperio?

Como conocemos la corriente y la resistencia aplicamos la fórmula :

$$P = I^2 \times R$$

Reemplazando los valores tenemos :

$$P = (1 \times 1) \times 15 = 15 \text{ Vatios}$$

Esta potencia que se *disipa* o gasta en la resistencia se convierte en calor.

También podemos hallar el valor de la resistencia, la corriente y el voltaje conociendo la potencia y otro factor.

b)¿Cuál es la resistencia de un bombillo de 100 vatios que está conectado a 110 Voltios ?

Sabemos que $P = V^2/R$ de donde $R = V^2/P$

Reemplazando tenemos :

$$R = (110)^2 / 100 = (12.100) / 100 = 121\ \Omega$$

El bombillo convierte los 100 vatios que recibe de energía eléctrica, en energía calorífica una parte, y la otra en energía luminosa.

Formas de la corriente eléctrica

La corriente eléctrica en un circuito circula de muchas maneras,dependiendo del tipo de corriente que suministre la fuente que alimenta a un circuito determinado.

Hay corrientes que se mantienen constantes, otras están fluctuando o variando en el transcurso del tiempo, otras invierten su sentido de circulación cada tiempo determinado,cambiando periódicamente su polaridad, otras corrientes suben y caen bruscamente. Todo lo anterior de diferentes formas y magnitudes.

Las formas más conocidas de corriente son:

Corriente continua (CC) o Corriente directa (CD)

Corriente pulsante

Corriente alterna (CA)

Corriente continua

Cuando tenemos una fuente que suministra una corriente constante que siempre fluye en un solo sentido, es decir el polo positivo siempre será positivo y el polo negativo siempre será negativo, estamos ante la presencia de una Corriente Continua (CC). Las pilas y las baterías suministran corriente continua.

Cuando los electrones libres se mueven en una dirección, ellos sólo viajan una distancia muy corta. O sea, cada electrón se mueve una fracción de distancia muy pequeña, pero el efecto total es como si un electrón se moviera a través de todo el conductor. Esto trabaja más o menos así, como se muestra en la figura 25.

Bolas de ping-pong

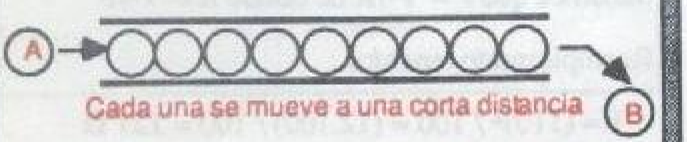

Fig. 25

Suponga que hay un tubo lleno de pelotas de ping-pong. Si usted mete otra pelota -A-,una sale por el otro extremo -B-. En el interior del tubo cada pelota sólo se mueve una corta distancia, pero el efecto se siente en todo el tubo.

Esto es lo que pasa en un circuito eléctrico. Las bolas de ping-pong son los *electrones*, cada cual empujando a otro, aunque cada electrón se mueve sólo a una corta distancia. (figura 26).

En el circuito que se muestra en la figura 27 (y en todos los circuitos eléctricos) cuando se cierra un suiche, los electrones se mueven en la dirección indicada. Esto pasa en todo el circuito al mismo tiempo.

Por cada electrón que sale del terminal negativo de la pila, llega uno inmediatamente al terminal positivo de la misma.

La corriente es, para términos prácticos, *instántanea* en todo el circuito.

Actividad práctica Nº 2

En esta segunda actividad vamos a instalar en el tablero del radio AM las puntillas de cobre o soportes que servirán para sostener y soldar los diferentes componentes del radio.

Materiales y herramientas necesarios

150 Puntillas de cobre calibre 18 x 5/8 ". Estas deben ser de cobre para facilitar la soldadura sobre ellas.

1 Martillo

Paso 1

Tome las puntillas y el martillo. Clave las puntillas en cada punto marcado con un círculo sombreado en el diagrama esquemático, figura A3.

Haga este trabajo cuidadosamente. Las puntillas deben estar en ángulo recto con la superficie de la tabla y deben entrar por lo menos siete (7) milímetros en la tabla para que queden bien sostenidas. Si no es suficientemente gruesa la tabla, las puntillas saldrán al otro lado o no se sostendrán firmemente.

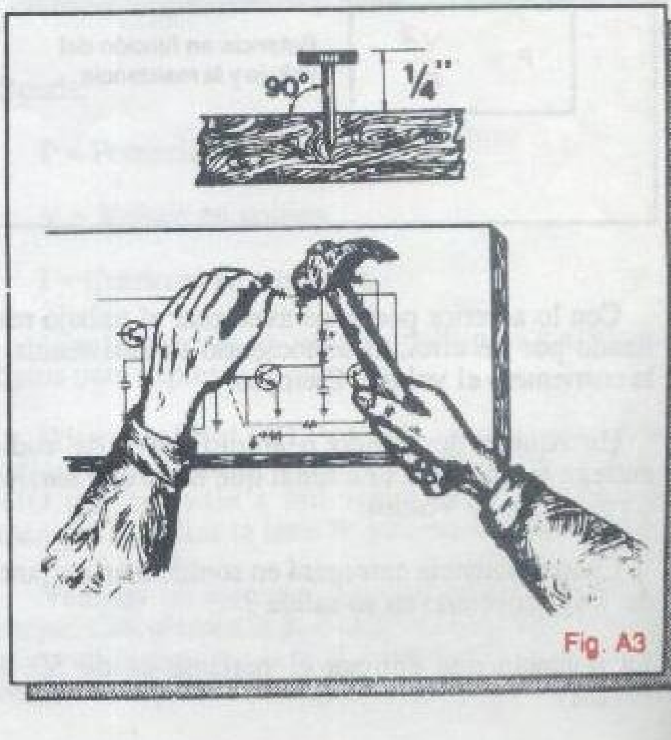

Fig. A3

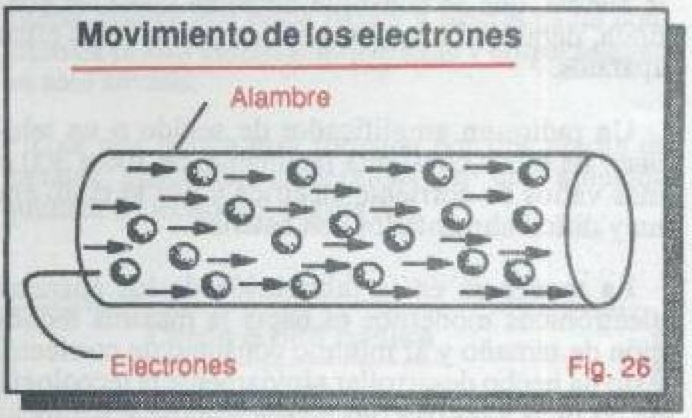
Movimiento de los electrones

Fig. 26

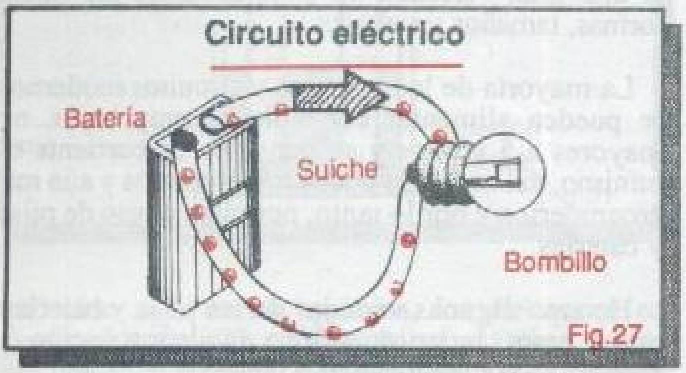
Circuito eléctrico

Fig.27

Necesidad de corriente continua

La corriente continua es aquella que fluye en una sola dirección y que no cambia con el tiempo; a la corriente continua también se le llama corriente directa y se representa con las siglas CC y CD.

En la figura 28 tenemos la representación gráfica de dos tipos de corriente continua.

Corriente continua

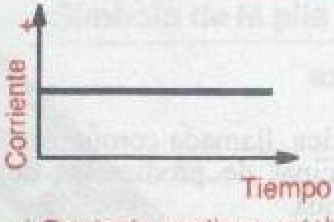

a) Corriente continua estable

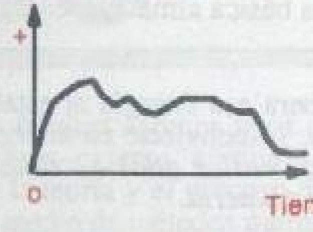

b) Corriente continua variable

Fig 28

En la forma de onda a, la corriente es positiva y siempre tiene el mismo valor.

En la forma de onda b, la corriente sigue siendo positiva pero cambia de valor con el tiempo sin llegar a ser negativa.

La necesidad de tener corriente continua es muy amplia. Si usted mira alrededor y piensa en la gran variedad de aparatos electrónicos que se utilizan actualmente en los hogares, en las oficinas y en la industria, quizás no se imagina que *todos* trabajan internamente con corriente continua.

Por lo tanto, todos los equipos electrónicos necesitan alimentación de corriente continua (CC).

Métodos de producción de CC

Hay dos métodos de producción o suministros principales de corriente continua: las pilas y las fuentes de poder.

Fuentes de poder

Las fuentes de poder o convertidores son aparatos electrónicos que convierten la corriente alterna (CA) de 110 o 220 voltios,que se encuentra en los tomacorrientes en corriente continua (CC), con voltajes que van desde 1 hasta cientos o miles de voltios. (figura 29).

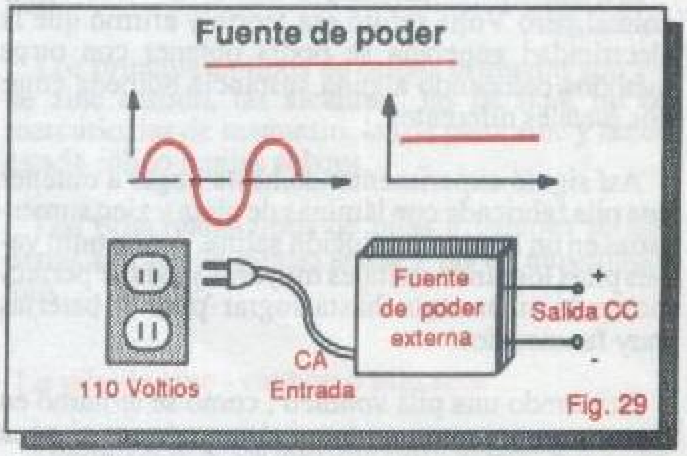
Fuente de poder

Fig. 29

Si no se necesita que un aparato sea portátil y liviano, se usan las fuentes de poder, ya sean internas o externas. La gran mayoría de los aparatos electrónicos tienen una fuente de poder interna. (figura 30).

Estaremos estudiando el tema de las fuentes de poder,tanto internas como externas,cuando conozcamos los componentes que se utilizan en ellas como son los transformadores, los condensadores, los semiconductores y otros.

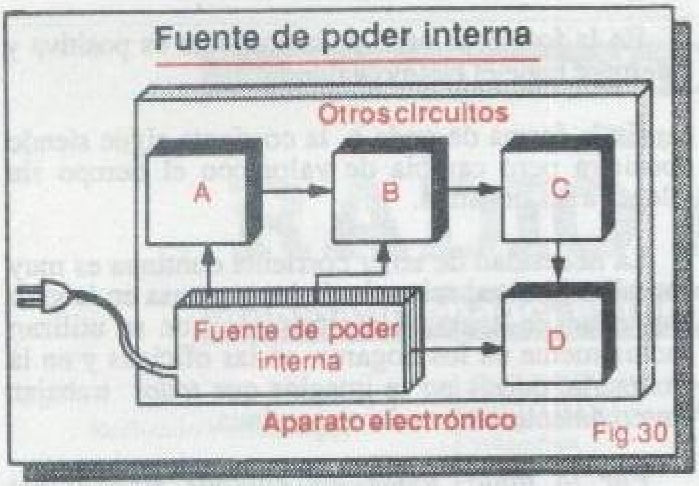

Fuente de poder interna

Fig.30

Pilas y Baterías

Los primeros experimentos y desarrollos de la radio y la electrónica se hicieron con corriente continua, utilizando las primeras pilas y baterías que se inventaron.

La invención de la pila se le debe a Alessandro Volta , físico italiano que vivió entre 1745 y 1827. Su invento lo anunció en el año de 1800 ante la Royal Society de Inglaterra.

Volta se basó en el principio descubierto por Galvani, otro científico italiano quien experimentó con tejidos animales, observando que al ponerlos en contacto con dos metales diferentes se contraían.

Inicialmente se pensó que existía la electricidad animal, pero Volta refutó esa teoría y afirmó que la electricidad generada se podía obtener con otros métodos, colocando alguna sustancia húmeda entre dos metales diferentes.

Así siguió experimentando, hasta llegar a obtener una pila fabricada con láminas de plata y zinc, sumergidas en un líquido o solución salina. Volta unió varias pilas, logrando voltajes mayores y así fue perfeccionando su invento hasta lograr pilas y baterías muy funcionales.

Utilizando una pila *voltaica* , como se le llamó en honor a su inventor, se logró descomponer el agua en sus dos elementos, el hidrógeno y el oxígeno, por medio de un proceso llamado *electrólisis*.

Durante muchos años las pilas fueron el principal método de producción de corriente eléctrica continua, hasta que se utilizó el fenómeno de la inducción magnética para fabricar los primeros generadores, tanto de corriente continua como de corriente alterna.

Luego, en la época de los tubos de radio o válvulas electrónicas, se utilizó principalmente la corriente alterna, que se convertía luego en corriente continua, debido al alto consumo de potencia de estos aparatos.

Un radio, un amplificador de sonido o un televisor de tubos, consumía fácilmente de 100 a 300 o más vatios de corriente eléctrica y por lo tanto era muy difícil alimentarlos con baterías.

La tendencia en la fabricación de los aparatos electrónicos modernos es hacia la máxima reducción de tamaño y al mínimo consumo de corriente. Esto ha hecho desarrollar rápidamente la tecnología de las pilas y las baterías, produciéndose actualmente una gran variedad de ellas en todos los tipos, formas, tamaños y voltajes.

La mayoría de los aparatos y circuitos modernos se pueden alimentar con voltajes muy bajos, no mayores a 5 voltios y su consumo de corriente es mínimo, del orden de pocos miliamperios y aún microamperios y por lo tanto, permiten el uso de pilas y baterías.

He aquí algunas ventajas de las pilas y baterías: poco peso, larga duración, fácil instalación y reemplazo y alta eficiencia.

Estas ventajas y otras no enumeradas han hecho muy popular el uso de pilas y baterías en los equipos electrónicos actuales. Así vemos, por ejemplo, una gran variedad de calculadoras, relojes, radios, equipos de sonido portátiles tipo *Walkman*, televisores, equipos de comunicaciones y computadores personales alimentados con baterías.

Aunque su uso esté creciendo rápidamente, son una fuente de energía costosa, comparándola con la energía que tenemos en los tomacorrientes comunes. Sin embargo el desarrollo de baterías de gran duración y recargables hace que su costo se haga cada vez más bajo.

La celda básica

La celda básica, llamada comúnmente pila, es el elemento principal de producción de voltaje de corriente continua.

La pila o celda básica almacena energía en forma química.

Cuando se libera esa energía al establecerse un circuito cerrado, se convierte en energía eléctrica para nuestro uso. *Varias pilas o celdas se pueden unir y forman una batería.*

Por lo tanto, aunque se hace con mucha frecuencia, no debemos confundir estos dos términos; pila es una celda básica y batería es un conjunto de celdas unidas internamente y empacadas en un solo envase.

Una pila básica está formada por dos placas de metales diferentes, sumergidas en una solución química o electrólito. (figura 31).

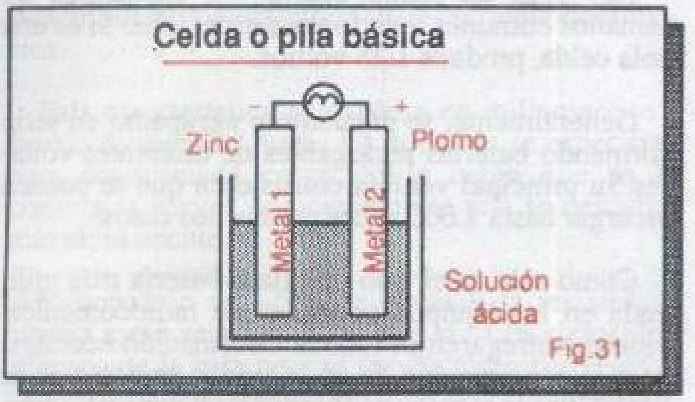
Celda o pila básica

Fig.31

Debido a la acción química, los electrones en exceso se acumulan en el terminal negativo, mientras que en el terminal positivo hacen falta electrones.

De esta manera se crea una diferencia de cargas o voltaje. Este voltaje es el que utilizamos para producir la corriente que alimenta los circuitos electrónicos.

En la figura 32 se muestra el símbolo utilizado para las pilas en los diagramas de los aparatos eléctricos y electrónicos. La línea más grande indica el polo positivo y la línea más corta es el polo negativo.

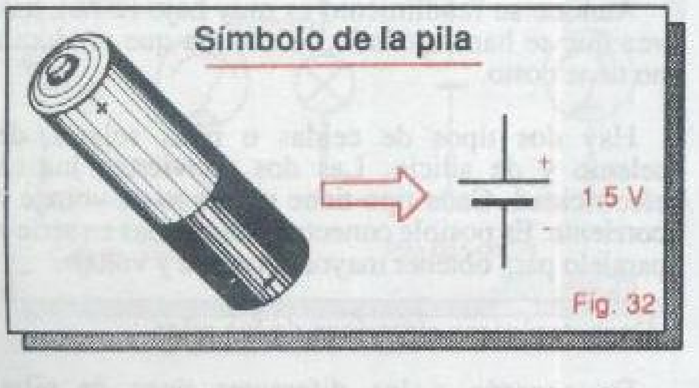
Símbolo de la pila

Fig. 32

Podemos fabricar algunos tipos de celdas básicas con elementos caseros comunes, con el fin de comprobar la teoría y el proceso de producción de voltaje por medio de métodos químicos.

Uno de estos métodos es la pila de limón. Podemos fabricarla con un limón que suministra el ácido y dos objetos de metal diferentes, una puntilla de hierro y un alambre grueso de cobre. (figura 33). Esta pila puede suministrar hasta 1 voltio,que se puede medir con un voltímetro digital sensible.

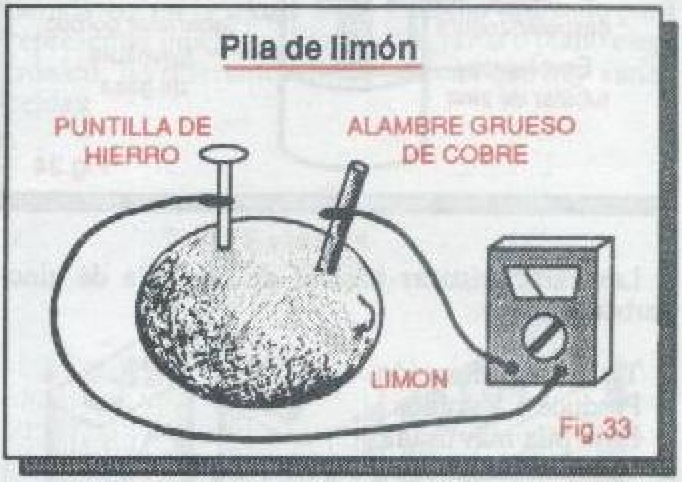
Pila de limón

Fig.33

Principales tipos de pilas

Las pilas se dividen en dos grandes grupos: las pilas y baterías primarias y las pilas y baterías secundarias.

Las pilas primarias no se pueden recargar y se utilizan en forma de descarga simple,y las secundarias se pueden recargar,o sea que se utilizan en forma de descargas sucesivas o de reciclaje repetido.

Los principales tipos de celdas primarias son los de zinc carbón, las alcalinas, las de litio, las de mercurio, las de magnesio, las de plata zinc y las de estado sólido o pilas solares.

Los principales tipos de pilas y baterías secundarias son las de plomo ácido y las de níquel cadmio.

La pila de zinc - carbón o pila seca

La pila de zinc-carbón está formada por una placa negativa de zinc, una placa positiva de carbón y un compuesto químico o electrólito de sal de amoníaco y cloruro de zinc. Estos elementos se distribuyen en un empaque cilíndrico como se muestra en la figura 34.

Se llama también pila seca debido a que el electrólito es seco y está impregnado en un medio absorbente,y por lo tanto no hay líquidos dentro de ellas. Este tipo de pila es el que se fabrica y distribuye.en mayor cantidad en el mundo y por su bajo costo y abundancia es el de mayor utilización.

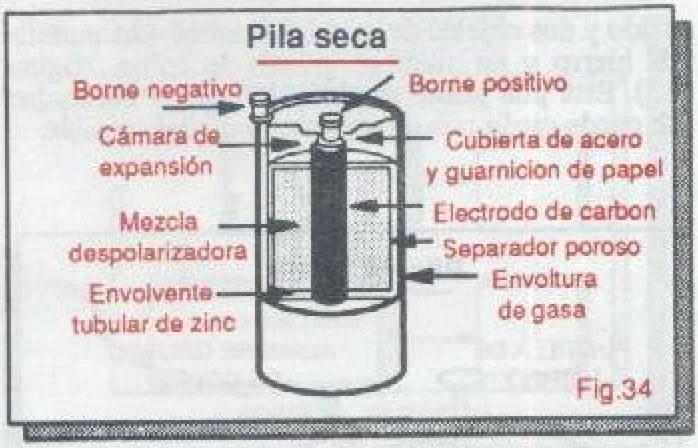

Fig.34

Las características básicas de una pila de zinc-carbón son: .

- Tiene larga duración.
- Produce 1.5 voltios.
- Es la pila más usada.
- Viene en tamaños AAA, AA, C y D.
- Es una celda primaria.
- Mientras más grande la pila, produce más corriente en un período de tiempo.

La pilas de zinc-carbón se encuentran en juguetes, radios, equipos portátiles que requieren una corriente moderada, etc. *Hay que recordar que por ser de tipo primario no se pueden recargar.*

La pila alcalina

La pila alcalina es similar a la de zinc-carbón y puede ser intercambiada en la mayoría de los casos. Esta pila también es conocida como de zinc-bióxido de manganeso.

La principal ventaja de la pila alcalina es que produce mayor corriente en tiempo largo. Aunque es más costosa que la de zinc-carbón, su duración compensa este valor.

Además, algunos tipos de pilas alcalinas son recargables.

Las pilas alcalinas son ideales para alimentar lámparas de destello en fotografía, juguetes con motores, linternas y otros equipos que requieren una corriente más o menos alta.

La pila de mercurio

La pila de mercurio está formada por los siguientes elementos: una placa negativa de óxido de mercurio, una placa positiva de zinc y un ácido o electrólito de hidróxido de potasio.

Las características básicas de una pila de mercurio son:

- Entrega aproximadamente de 1.33 a 1.4 voltios con una rata de descarga muy constante.
- Es muy pequeña, algunas son tan delgadas como una aspirina.
- Tiene alta eficiencia.
- Son primarias, o sea, no se recargan.

La pila de níquel - cadmio

Las pilas de níquel-cadmio se encuentran en tamaños comunes y en formas especiales. Si es una sola celda, produce 1.25 voltios.

Generalmente, se encuentran agrupadas en serie formando baterías recargables de diferentes voltajes.Su principal ventaja consiste en que se pueden recargar hasta 1.000 veces en muchos casos.

Como este es el tipo de pila o batería más utilizada en los equipos portátiles de radiocomunicaciones, entregaremos toda la información necesaria para la construcción de un cargador de este tipo de baterías en la sección de Banda Ciudadana y Radioafición.

Una pila de níquel-cadmio está formada por una placa negativa de cadmio metálico, una placa positiva de hidróxido de níquel y un ácido o electrólito de hidróxido de potasio.

Las pilas o baterías de níquel-cadmio se usan cuando se necesitan recargas y una larga vida, especialmente en equipos portátiles de radiocomunicación, linternas recargables y otros aparatos.

Las pilas solares

Hay un aumento creciente en el uso de tas pilas y las baterías solares. Estas son pilas que convierten energía luminosa en energía eléctrica directamente.

Aunque su rendimiento es muy bajo (20%), una vez que se han instalado, la energía que producen no tiene costo.

Hay dos tipos de celdas o pilas solares, de selenio y de silicio. Las dos convierten luz en electricidad. Cada tipo tiene un rango de voltaje y corriente. Es posible conectar estas celdas en serie y paralelo para obtener mayor corriente y voltaje.

Características eléctricas de las pilas

De acuerdo a los diferentes tipos de pilas estudiadas anteriormente,se producen en ellas características eléctricas distintas en cada una.

Las principales características eléctricas que se deben tener en cuenta en las pilas son voltaje en vacío o tensión nominal, capacidad de suministro de energía, voltaje o tensión con carga y su resistencia interna.

El voltaje o tensión nominal es aquel que está marcado en el empaque exterior de la pila,y depende de su fabricación y composición química. El voltaje nominal más común es de 1.5 voltios.

La capacidad de suministro de energía se refiere, principalmente,al tiempo que puede suministrar una pila,una determinada cantidad de corriente en amperios.

Esta característica se expresa en miliamperios hora o Amperios / hora. La duración se especifica para determinada carga. Si la carga que se le conecta es mayor, lógicamente la pila se descarga más rápidamente.

El voltaje o tensión con carga y la resistencia interna están relacionados entre sí. Los elementos internos de la pila no son conductores perfectos y tienen cierta resistencia al paso de la corriente,lo que se denomina *resistencia interna*. Esta resistencia aumenta con el uso, con el tiempo de almacenaje y con la disminución de la temperatura ambiente en la cual está la pila.

Esta resistencia interna produce una rebaja del voltaje que entrega la pila al circuito externo. Por lo tanto, cuando se conecta alguna carga, el voltaje real sobre esa carga es ligeramente menor que el voltaje nominal. (figura 35).

Voltaje real de una pila

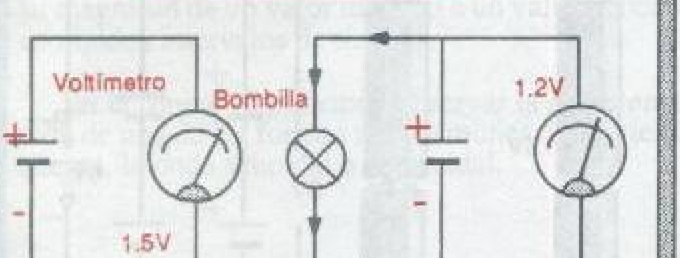

Fig.35

Baterías

Una celda química o pila básica produce un voltaje muy bajo y por lo tanto se deben agrupar para obtener voltajes mayores,que son necesarios para alimentar los circuitos de los aparatos electrónicos.

Varias celdas básicas o pilas se pueden conectar en serie y de esta forma se obtiene una batería.

Las baterías más comunes en el mercado son:la batería de 9 voltios y las baterías de plomo-ácido de 6, 12 y 24 voltios,que se utilizan en las motos, automóviles y similares. (figura 36). También podemos ver en la figura 36 el símbolo utilizado para representar una batería en un diagrama o plano electrónico; las diferentes líneas indican que hay varias celdas.

Baterías

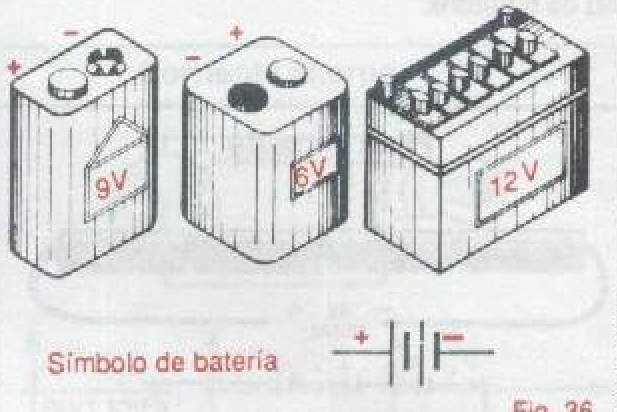

Fig. 36

Baterías de plomo-ácido

Uno de los tipos de batería más utilizados como suministro de corriente continua son las de plomo-ácido, también llamadas acumuladores.

Estas baterías son las que utilizan los automóviles y se pueden recargar muchas veces.

Se fabrican con voltajes de 6, 12 y 24 voltios y su característica más importante es la alta corriente que pueden suministrar.

En electrónica se utilizan para alimentar equipos de sonido portátiles, radioteléfonos, sistemas de iluminación de emergencia, plantas telefónicas, etc.

Conexiones de las pilas y las baterías

Las pilas y las baterías se pueden conectar o unir entre sí para obtener voltajes o corrientes de alimentación,mayores que los que pueden producir por sí solas.

Las conexiones que se pueden hacer se llaman conexión en *serie*, en *paralelo* y en *serie paralelo*.

Conexión en serie

Las pilas y las baterías se pueden conectar en serie,con el fin de obtener un voltaje mayor entre los terminales del nuevo conjunto.

Una conexión en serie significa que el polo negativo de la primera pila se conecta al polo positivo de la siguiente,y así sucesivamente hasta terminar con todas las pilas del circuito.

El voltaje de las pilas en serie se *suma,* obteniéndose así un voltaje mayor.

En la figura 37 podemos observar la conexión y los voltajes de un conjunto de cuatro pilas de 1.5 voltios conectadas en serie,obteniendo un voltaje total de 6 voltios.

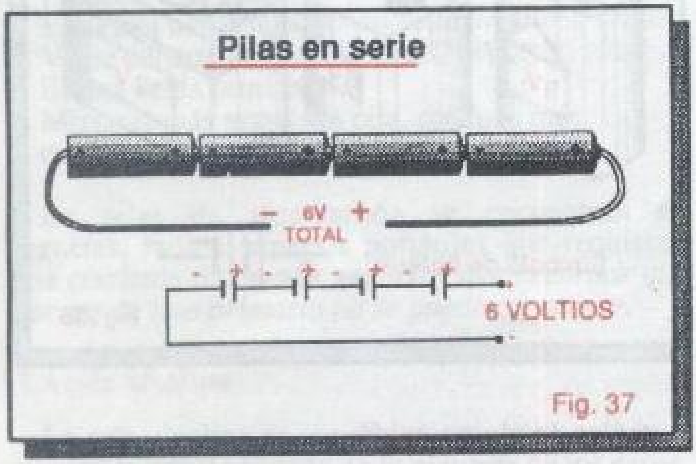

Fig. 37

Conexión en paralelo

Se llama conexión en paralelo cuando se unen todos los terminales o polos positivos de varias pilas o baterías a una línea común,que es el terminal positivo de todo el conjunto,y todos los terminales negativos a otra línea común que es el terminal negativo del conjunto. (figura 38).

El voltaje de salida de esta conexión es igual al de una sola pila o batería,pero la corriente en amperios que puede entregar esta nueva configuración es igual a la suma de las corrientes de todas las pilas.

De esta manera un conjunto de este tipo puede alimentar circuitos con más consumo de corriente,o proporcionar una mayor duración en el mismo circuito alimentado por una sola pila.

Conexión serie - paralelo

Se pueden combinar los dos tipos anteriores de conexiones de pilas con el fin de obtener las dos características de un mayor voltaje y una mayor corriente en un solo conjunto.

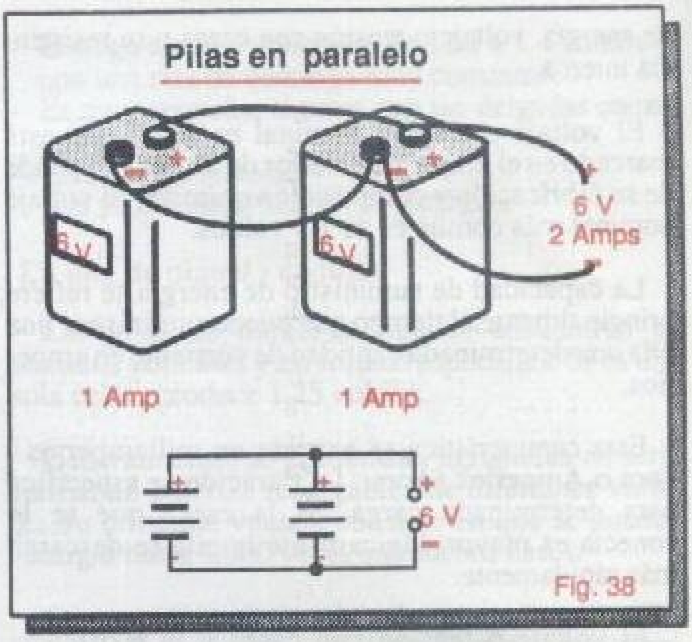

Fig. 38

En la figura 39 podemos observar este tipo de conexión con un ejemplo del voltaje y la corriente resultantes.

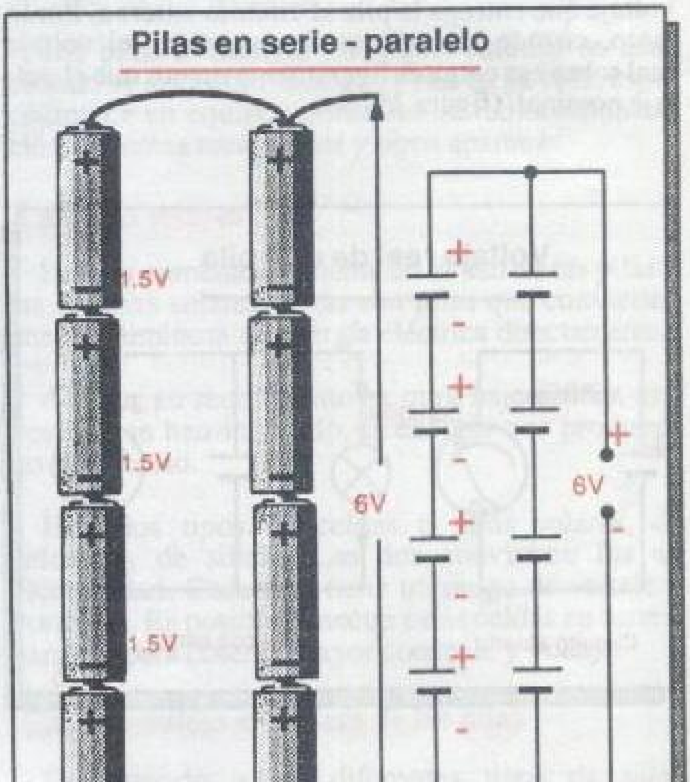

Fig. 39

Corriente alterna (CA o AC)

El estudio de la corriente alterna, la otra forma de energía eléctrica, es de mucha importancia en este curso de radio,ya que todas las señales que se utilizan en los circuitos de radio y comunicaciones son señales de corriente alterna.

Utilizamos el término *señales* para indicar una forma de energía eléctrica,con un voltaje y corriente determinados.

Las ondas de radio que se emiten en un transmisor,son producidas por señales de corriente alterna que llegan a la antena. Los sonidos que se transmiten por radio, son también señales de corriente alterna y, la energía que se utiliza para alimentar muchos aparatos electrónicos,es corriente alterna.

Hemos llegado a uno de los temas más importantes y apasionantes dentro de todo este curso, ya que la corriente alterna se encuentra en todas partes y los fenómenos que involucra son muy variados y especiales.

Haremos especial énfasis en el estudio del comportamiento de los diferentes componentes de los radios,como son los condensadores, las bobinas y los semiconductores cuando tienen aplicada a sus terminales una o varias señales de corriente alterna.

La corriente alterna, que se escribe a menudo abreviadamente con las letras CA o AC (Alternate Current), se puede considerar como una corriente que invierte periódicamente su polaridad, variando su magnitud de un valor mínimo a un valor máximo en rápidos intervalos de tiempo.

En la figura 40 podemos observar la representación de una de las formas más comunes de corriente alterna, la onda senoidal o sinusoidal.

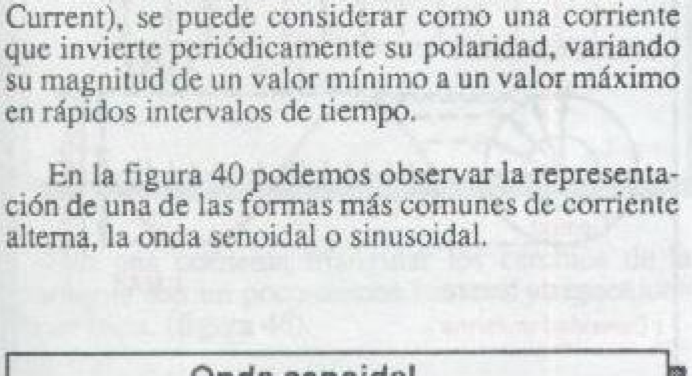

Fig. 40

Una fuente de corriente alterna no tiene borne positivo y borne negativo,ya que un borne es negativo en un instante y se vuelve positivo en otro instante, es decir la corriente circula en un sentido y luego en sentido opuesto,en un movimiento de ir y venir que se repite indefinidamente.

En la corriente alterna, el flujo de electrones va primero en una dirección, se detiene y luego circula en dirección contraria. (figura 41).

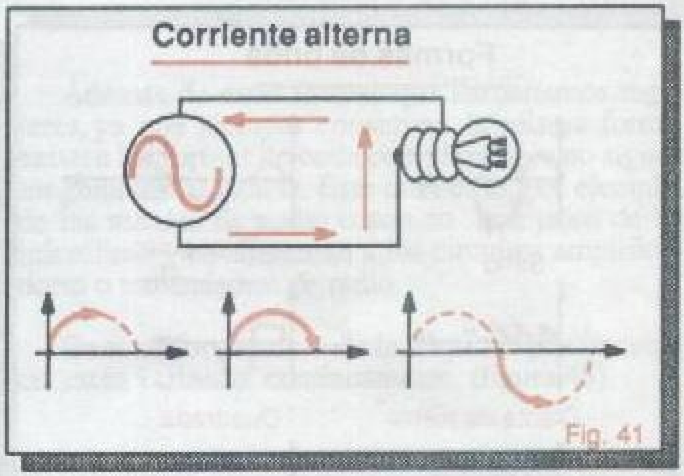

Fig. 41

Características de la corriente alterna

Una energía o señal de corriente alterna tiene diferentes características que las hace distinguir unas de otras.

Cuando hablamos de las ondas de radio, que son señales de corriente alterna, adelantamos algunas características de este tipo de corriente.

Las principales características de una señal de corriente alterna son :

a) La forma de onda
b) La amplitud o voltaje
c) La frecuencia

Cuando existen dos señales de corriente alterna en un mismo circuito, se tiene además otro factor importante que se llama *relación de fase*.

La forma de onda

La corriente alterna se produce por diferentes métodos. Los más utilizados son los generadores electromagnéticos o plantas eléctricas y los circuitos osciladores electrónicos.

Una señal de corriente alterna varía en su amplitud o tamaño con respecto al tiempo de varias maneras, dependiendo del sistema utilizado por el generador o la fuente de voltaje.

Si representamos esta variación por medio de una figura de amplitud y tiempo se producen gráficas con diferentes formas geométricas. Estas gráficas nos indican la forma de onda con las cuales se identifica la señal.

Las formas de onda más conocidas son: la senoidal, la triangular, la de diente de sierra, la onda cuadrada y los pulsos. (figura 42).

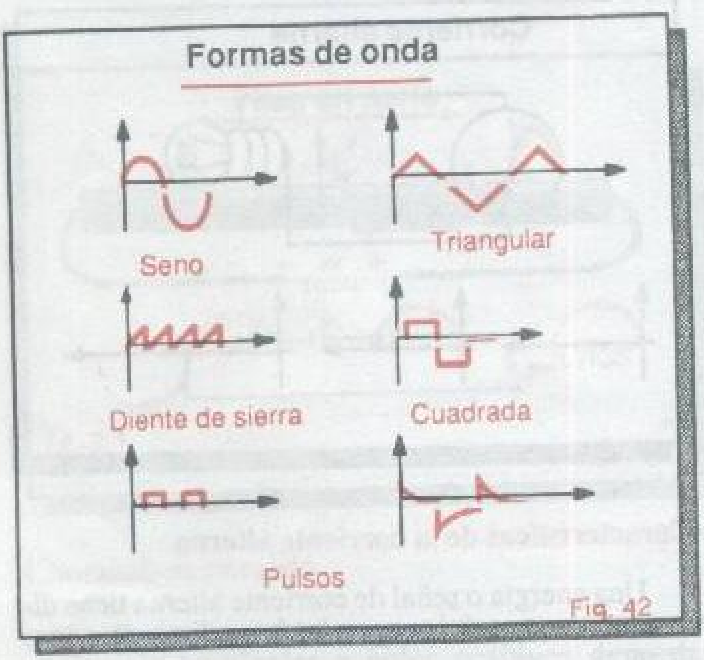

Fig. 42

La onda seno

La forma más generalizada de corriente alterna en los circuitos de radio y comunicaciones es la onda seno o senoidal.

Para entender mejor de dónde sale la forma de onda seno, debemos saber primero que una circunferencia se puede dividir en 360 segmentos o pedazos llamados grados.

Así, un cuarto de circunferencia son 90º, media circunferencia son 180º y así hasta completar el círculo completo con los 360º, llegando al mismo punto de partida o sea 0º. Como su nombre lo indica, su valor se toma de la función trigonométrica *seno*. Para obtener la gráfica se va tomando el valor del ángulo de rotación de un eje en una circunferencia, o sea 360º. Para cada ángulo se toma el valor del seno y se va colocando en el eje X. Uniendo los puntos obtenemos la gráfica para un ciclo o vuelta completa. (figura 43).

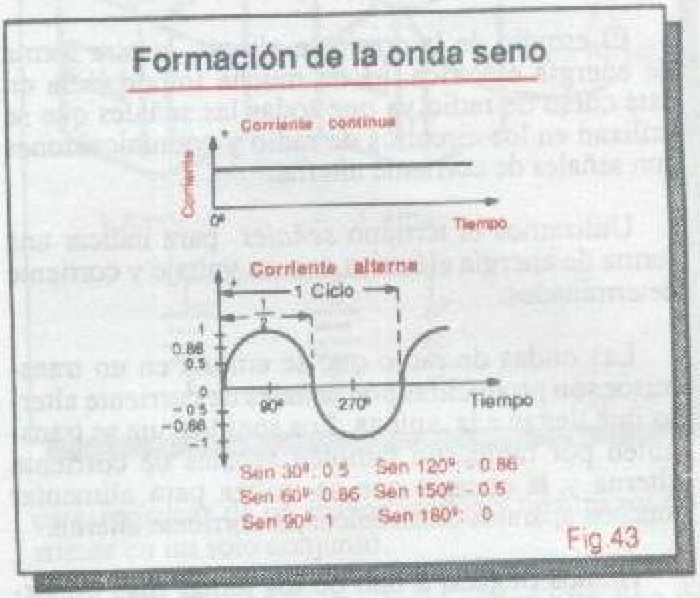

Fig.43

Esta forma de onda se deriva del método de producción obtenido en los generadores electromagnéticos rotativos, en donde una bobina gira dentro de un campo magnético obteniéndose así mayor o menor voltaje, dependiendo de la posición de la bobina. (figura 44).

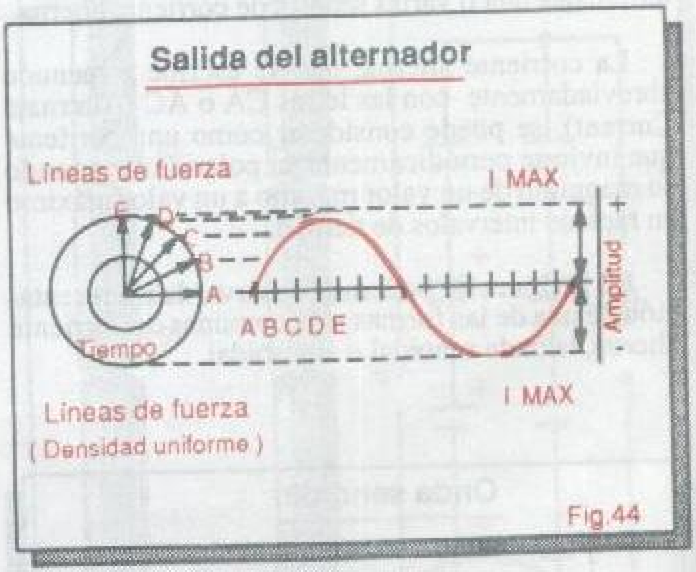

Fig.44

Todos los valores por encima del eje horizontal (eje X) son positivos y por debajo del eje X son negativos. Como notamos en la gráfica, la corriente alterna senoidal no cambia bruscamente.

En un tiempo inicial el voltaje es cero, luego empieza a aumentar progresivamente hasta llegar a su máximo valor positivo, a partir de este momento empieza a decrecer hasta llegar a cero; en este momento se invierte empezando a hacerse cada vez más negativo, alcanzando su máximo valor negativo, subiendo nuevamente a cero, retornando a su valor positivo, repitiéndose indefinidamente el ciclo.

Otras formas de onda

Otra de las formas más comunes de corriente alterna es la onda cuadrada en la cual los cambios de magnitud de la corriente son bruscos, van desde cero hasta su valor máximo en un período de tiempo muy corto. En la onda cuadrada el semiciclo positivo es igual en duración al semiciclo negativo.

Como una variación de la onda cuadrada se pueden tener señales llamadas pulsos, en donde la duración del semiciclo positivo es menor o mayor que el semiciclo negativo. En la figura 45 se ilustran formas de onda de estos dos tipos.

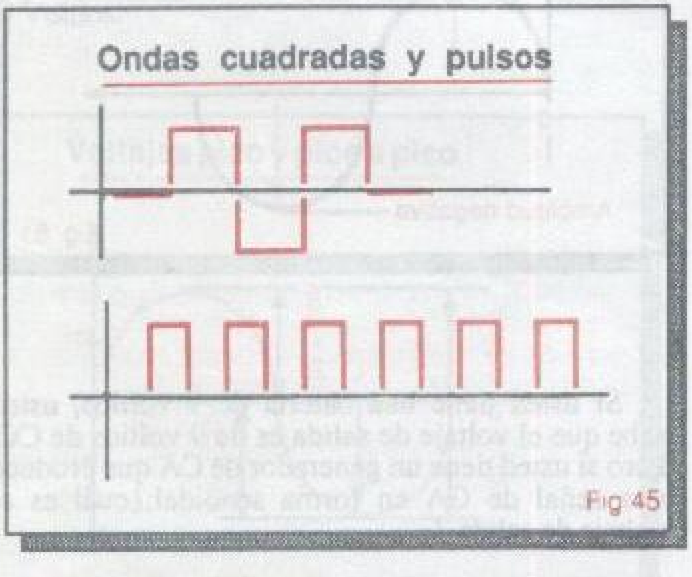
Ondas cuadradas y pulsos

Fig 45

En una corriente triangular los cambios de la corriente son un poco menos bruscos y siguen una línea recta. (figura 46).

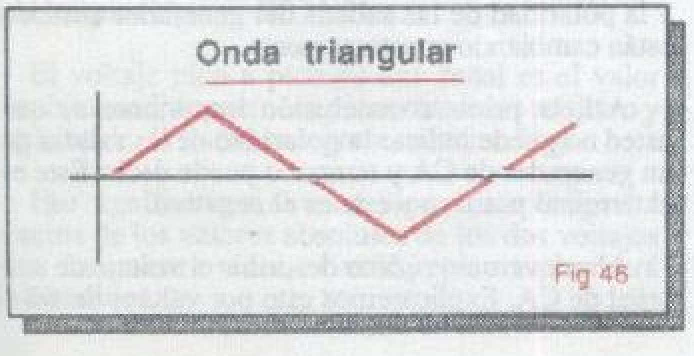
Onda triangular

Fig 46

Otra forma de onda muy común en los circuitos de radio y comunicaciones es la onda llamada de diente de sierra, y que se puede presentar en diferentes formas. (figura 47).

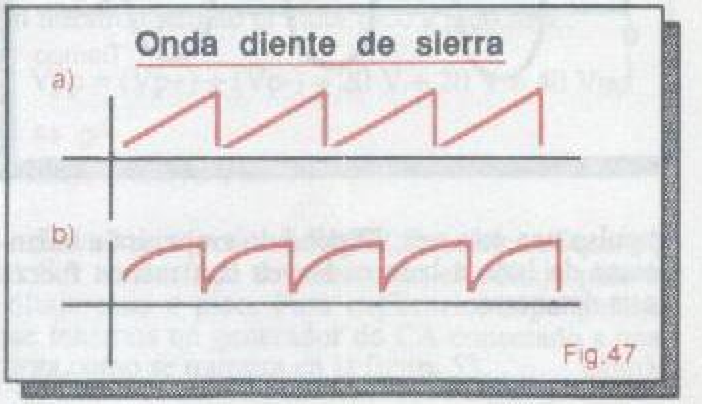
Onda diente de sierra

Fig.47

Además de estas formas que llamaríamos regulares, ya que siempre conservan la misma forma, existen las formas de onda complejas que no siguen ninguna forma exacta. Este es el caso, por ejemplo, de las señales de audio o sonido que salen de un micrófono y se alimentan a los circuitos amplificadores o transmisores de radio.

En estas formas de onda la amplitud y la frecuencia están variando continuamente. (figura 48).

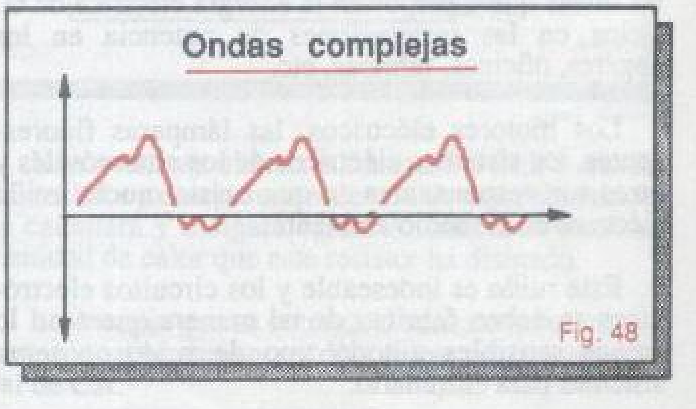
Ondas complejas

Fig. 48

Corriente alterna amortiguada

Existe una forma especial de corriente alterna llamada onda amortiguada.

Su polaridad se invierte de la misma forma que en cualquier corriente alterna, pero la diferencia radica en que su magnitud decrece cada vez más con el tiempo, hasta que desaparece totalmente. (figura 49).

Podemos pensar en esta onda como el resultado del movimiento de un péndulo al cual se le aplica un

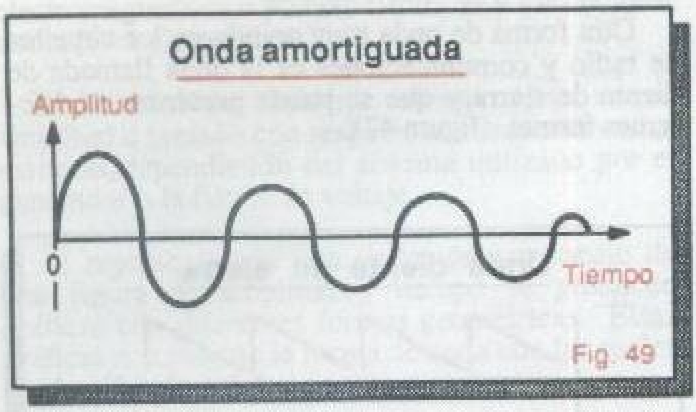
Onda amortiguada

Fig. 49

impulso una sola vez. El péndulo empezará a balancearse de lado a lado, cada vez con menos fuerza, hasta detenerse.

Ruido

El ruido es una forma de onda irregular, que se presenta en todos los circuitos electrónicos por diferentes causas.

Los componentes electrónicos generan ruido internamente, las señales de radio toman ruido de las descargas eléctricas y de otros fenómenos atmosféricos, llevándolo a los aparatos de radio.

También, se produce una gran cantidad de ruido por el fenómeno de inducción magnética sobre los cables que llevan la corriente y las señales en los aparatos electrónicos. Esta inducción proviene de las líneas que transportan la energía eléctrica de 60 ciclos, en las instalaciones de potencia en los hogares, oficinas, fábricas, etc.

Los motores eléctricos, las lámparas fluorescentes, los sistemas eléctricos de los automóviles y otros, son responsables de que exista mucho ruido eléctrico en el medio ambiente.

Este ruido es indeseable y los circuitos electrónicos se deben fabricar, de tal manera, que sean lo menos sensibles a todo tipo de ruidos o tener sistemas para eliminarlo.

Aunque este ruido es muy débil, se puede volver muy grande si se amplifica junto con las señales de un micrófono, por ejemplo, o con la señal muy débil de una emisora lejana de radio en onda corta. (figura 50).

Amplitud

La amplitud de una señal de corriente alterna es el tamaño o nivel que alcanza la onda en su punto o valor máximo con respecto al valor cero. (figura 51). La amplitud puede ser positiva o negativa, según el valor de la onda en un momento dado.

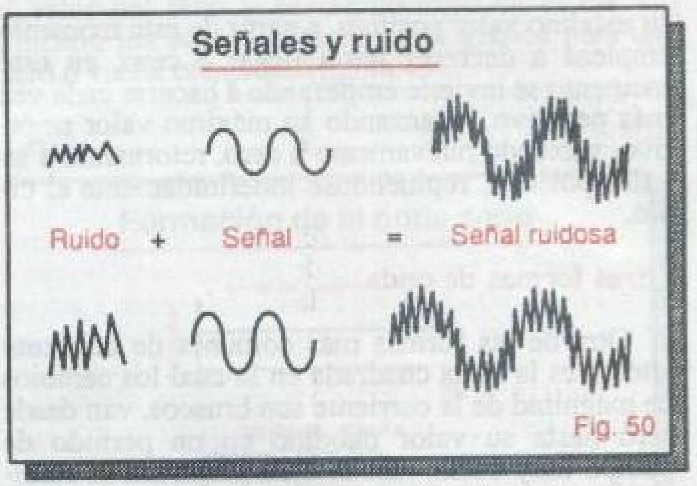
Señales y ruido

Fig. 50

La amplitud de una señal de corriente alterna se relaciona directamente con el voltaje y existen diferentes formas de expresarla o medirla .

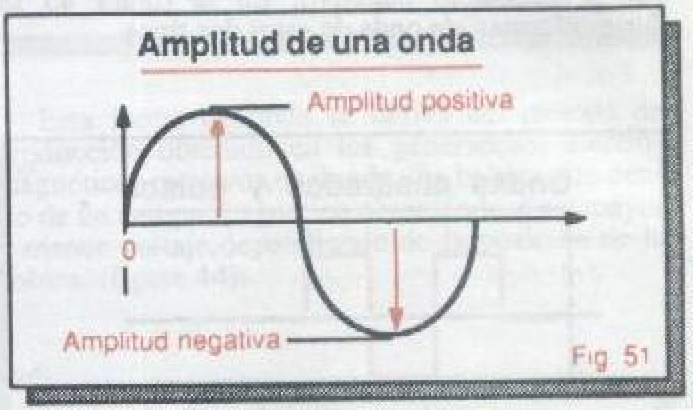
Amplitud de una onda

Fig. 51

Si usted tiene una batería de 9 voltios, usted sabe que el voltaje de salida es de 9 voltios de CC. Pero si usted tiene un generador de CA que produce una señal de CA en forma senoidal; ¿cuál es el voltaje de salida ?

Usted puede decir: No sé, porque el voltaje está constantemente cambiando y la polaridad de las salidas también.

Si usted dice esto, es correcto, porque el voltaje de una señal de CA está constantemente cambiando, y la polaridad de las salidas del generador también están cambiando constantemente.

Así, la primera conclusión importante es que usted no puede indicar la polaridad de las salidas de un generador de CA y tampoco puede decir: Este es el terminal positivo y este es el negativo.

Ahora veremos cómo describir el voltaje de una señal de CA. Explicaremos esto por valores de voltaje. Pero esta explicación es también válida en CA para corriente y potencia.

Existen diferentes formas de expresar el valor o voltaje de una señal de CA,dependiendo de la parte de la onda que se mida.

En radio y en electrónica, en general, es muy importante determinar cuál de estos valores se está expresando en un momento dado.

Los principales valores para una señal de CA son el voltaje pico, el voltaje pico a pico, el voltaje efectivo o RMS y el voltaje promedio.

Valores de voltaje pico y pico a pico

El valor pico de una señal de corriente alterna,es el valor máximo de amplitud que alcanza una señal en un momento dado. Este valor puede ser positivo o negativo.

El voltaje pico positivo de una señal CA es el valor más positivo de voltaje,y se representa como Vp. Por ejemplo, en la señal de CA que se muestra en la figura 52, el voltaje pico positivo es : Vp+ = 20 Voltios.

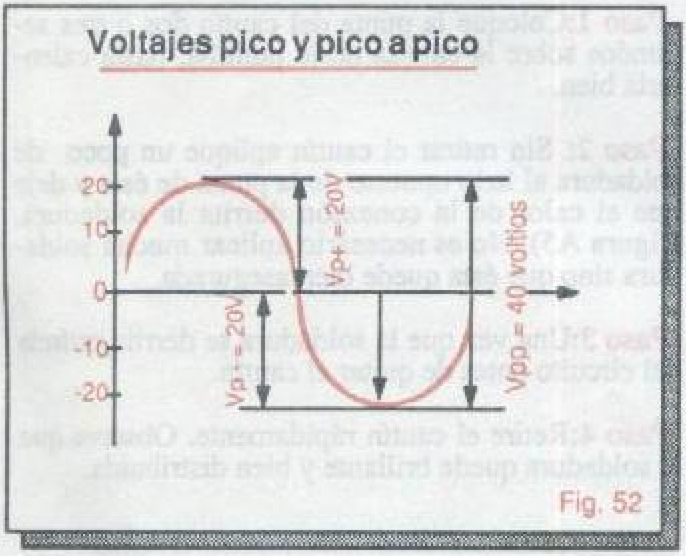
Voltajes pico y pico a pico

Fig. 52

En cambio, el voltaje pico negativo de una señal de CA es el valor más negativo del voltaje. En nuestro ejemplo: Vp - = - 20 Voltios.

El voltaje pico a pico de una señal es el valor absoluto,medido entre los picos máximos positivo y negativo de la onda; se representa como Vpp.

Eso significa que el valor pico a pico es igual a la suma de los valores absolutos de los dos voltajes pico. En nuestra señal, el voltaje pico a pico será de 40 voltios.

Valor absoluto es aquel que sólo tiene en cuenta la magnitud de la señal pero no su signo o polaridad. Por ejemplo, para un voltaje pico negativo de -20Voltios, el valor absoluto será de 20 Voltios. El signo menos desaparece y en una cantidad positiva el valor absoluto es la misma cantidad .

En nuestro ejemplo el valor pico a pico será :

Vpp = (Vp+) + (Vp-) = 20 V + 20 V = 40 Vpp

Voltaje efectivo o RMS

El voltaje efectivo o RMS (Root Mean Square) de CA no es fácil de ver,como el voltaje pico o voltaje pico a pico. Para explicarlo, supongamos que tenemos un generador de CA conectado a una carga,como se muestra en la figura 53.

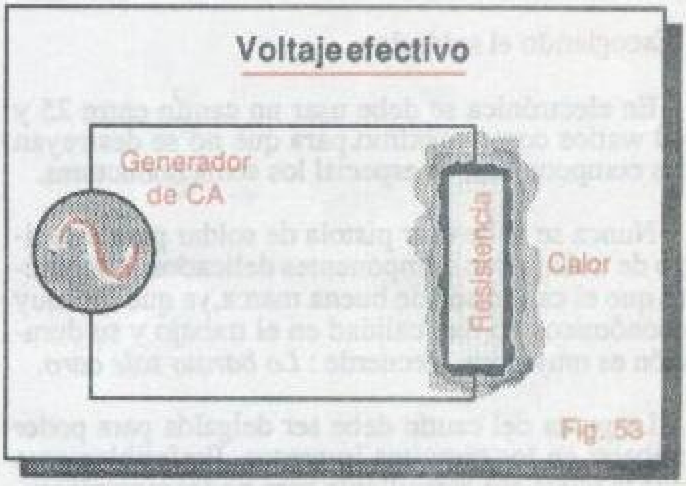
Voltaje efectivo

Fig. 53

Una corriente alterna comenzará a fluir a través del resistor y como consecuencia de esto, el resistor se calentará y disipará calor. Ahora,mediremos la cantidad de calor que este resistor ha disipado.

La energía que se entrega al resistor no es constante sino que aumenta y disminuye por ser una señal de CA.

Por esta razón el calor trata de subir y bajar, siguiendo las variaciones de la señal de voltaje.

Estas variaciones hacen que la energía *efectiva* de calor sea menor que si se aplicara un voltaje constante o de CC,con valor igual al valor pico o máximo de la señal de CA.

El valor del voltaje efectivo con esta señal de corriente alterna es igual al valor del voltaje en corriente continua, que aplicado por un minuto a este circuito, producirá la misma disipación de calor que la corriente alterna.

Actividad práctica Nº 3

Para realizar esta práctica usted debe tener todas las puntillas correctamente aseguradas en el tablero del radio según la actividad anterior.

Herramientas y materiales necesarios

1 Cautín de 25 o 40 W para uso en electrónica

1 pedazo de soldadura de estaño de 60/40

La soldadura en la práctica electrónica

Es indispensable para todas las personas que trabajen con electrónica a nivel experimental, de estudio o aficionados, saber soldar correctamente. De esta habilidad depende en gran parte el éxito, el correcto funcionamiento de los circuitos y aparatos que se armen.

Escogiendo el soldador

En electrónica se debe usar un cautín entre 25 y 40 watios como máximo, para que no se destruyan los componentes, en especial los semiconductores.

Nunca se debe usar pistola de soldar para este tipo de trabajo con componentes delicados. Asegúrese que el cautín sea de buena marca, ya que los muy económicos no dan calidad en el trabajo y su duración es muy poca. Recuerde : *Lo barato sale caro.*

La punta del cautín debe ser delgalda para poder trabajar en los circuitos impresos. Preferiblemente que la punta sea atornillable, para poder reemplazarla cuando se desgaste. (figura A4).

La soldadura es un material compuesto por una aleación de estaño y plomo que se funde fácilmente y hace muy buen contacto eléctrico. *La soldadura más recomendada es aquella que tiene una composición o aleación de 60 % de estaño y 40 % de plomo, con un espesor de 1/2 o 1 milímetro.*

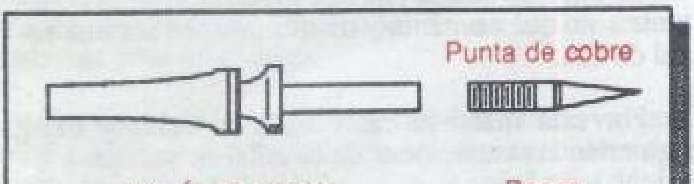

Fig. A4

La función del cautín es calentar la conexión. Luego el calor de la conexión debe fundir la soldadura. Es *obligatorio* que las superficies que se van a soldar estén perfectamente limpias. Para soldar correctamente se deben seguir dos pasos simples en orden que son:

1º-Caliente primero muy bien las superficies antes de aplicar la soldadura. *Este es el paso más importante en el proceso de soldar. Caliente bien y la soldadura se pega sola.*

2º- Luego derrita la soldadura sobre cada una de las partes a soldar, hasta que esta y las partes formen una sola pieza.

Vamos ahora a realizar nuestra primera práctica de soldadura, estañando las cabezas de las puntillas de cobre .

Antes de soldar limpie muy bien las cabezas de las puntillas para que queden libres de grasa, polvo u óxido. Caliente el cautín por lo menos unos 15 minutos antes de empezar a soldar y realice los siguientes pasos para cada una de las puntillas del tablero del radio:

Paso 1: Coloque la punta del cautín dos o tres segundos sobre la cabeza de la puntilla, hasta calentarla bien.

Paso 2: Sin retirar el cautín aplique un poco de soldadura al lado opuesto de la punta de éste y deje que el calor de la conexión derrita la soldadura. (figura A5). No es necesario aplicar mucha soldadura sino que ésta quede bien asegurada.

Paso 3: Una vez que la soldadura se derrita, retírela del circuito antes de quitar el cautín.

Paso 4: Retire el cautín rápidamente. Observe que la soldadura quede brillante y bien distribuida.

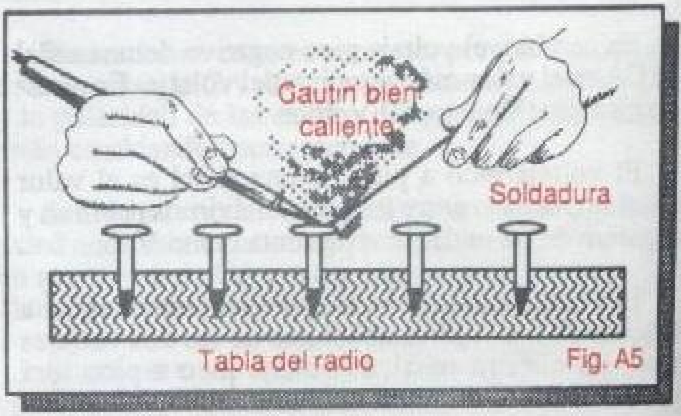

Fig. A5

Podemos definir el voltaje efectivo o RMS de una señal de CA como el valor causado por la misma fuerza al ser disipada en un resistor, como un voltaje de CC equivalente. El voltaje RMS es, en otras palabras, el voltaje útil y aprovechable. El valor del voltaje efectivo, depende de la forma de onda de la señal. No podemos dar una fórmula para calcular el valor del voltaje efectivo de todas las formas de onda. Como la forma de onda senoidal es la más utilizada, daremos la fórmula para calcular el voltaje RMS o efectivo para este tipo de señal.

$$\text{Vrms} = \frac{\text{Vp}}{\sqrt{2}} \qquad \text{Como} \sqrt{2} = 1.41$$

$$\text{Vrms} = \frac{\text{Vp}}{1.41} = \text{Vrms} = 0.707 \times \text{Vp}$$

$$\text{Así:} \quad \text{Vrms} = 0.707 \ \text{Vp}$$

$$\text{Vp} = 1.41 \times \text{Vrms}$$

Valor de voltaje promedio

Existe otra forma de expresar el voltaje de una señal de corriente alterna que se llama *valor promedio*.

Este valor se obtiene dividiendo el valor del área del semiciclo por su base. (figura 54).

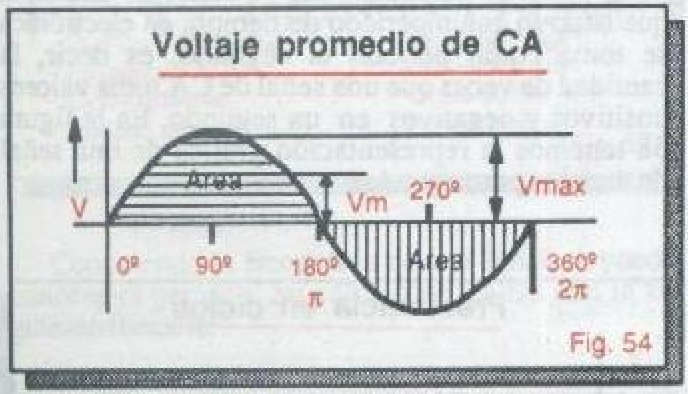

Fig. 54

Haciendo varios cálculos matemáticos se obtiene que el valor del voltaje promedio, que se expresa como Vm, será :

$$\text{Vm} = 0.637 \times \text{Vp}$$

Ahora, usted se pregunta: ¿Cuando medimos un voltaje de CA con el multímetro, qué voltaje de la señal mediremos?. El Vp, el Vpp, el voltaje efectivo o el valor promedio ?

El multímetro corriente o voltímetro, mide el voltaje efectivo o RMS de la señal.

El siguiente ejemplo le ayudará a entender mejor la relación entre las diferentes formas de medir un voltaje de CA, el Vrms, el Vp y el Vpp.

En la salida de un tomacorriente de su casa debe tener una señal de CA de 117 voltios. Si toma su multímetro y mide el voltaje, leerá en el aparato los 117 voltios de CA aproximadamente.

Ya que el multímetro mide voltaje efectivo o RMS y la forma de onda de la CA que usted tiene es una onda senoidal, usted puede calcular el voltaje pico de ésta, como sigue:

$$\text{Vp} = 1.41 \times \text{Vrms} = 1.41 \times 117 = 164.97 \text{ V}$$

También usted puede calcular el valor del voltaje pico a pico :

$$\text{Vpp} = 2 \times \text{Vp} = 2 \times 164.97 = 329.94 \text{ Voltios pico a pico}$$

Por eso la representación del voltaje de la señal de CA que usted tiene en la salida de pared será como muestra la figura 55.

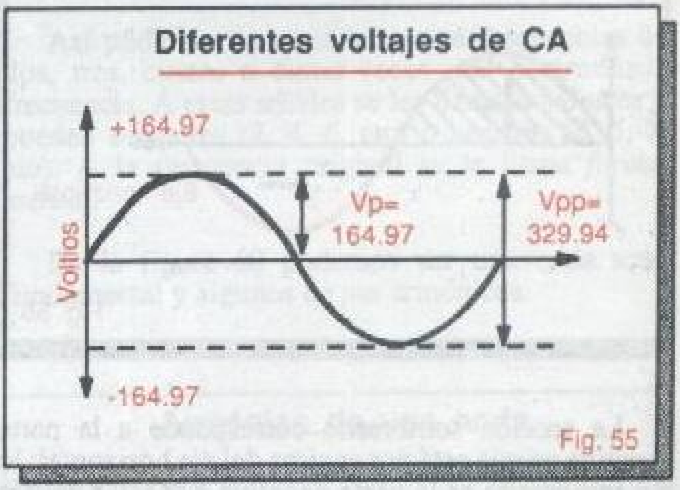

Fig. 55

Los instrumentos eléctricos y electrónicos están fabricados para medir el voltaje o la corriente de una determinada forma ya sea pico, pico a pico o RMS y para una forma de onda específica.

Por lo general los voltímetros y amperímetros comunes de CA están hechos para medir el voltaje efectivo o RMS en ondas senoidales.

Para medir voltajes pico, o pico a pico en ondas senoidales o en otras formas de onda, se utiliza el *osciloscopio*, instrumento que estudiaremos en una próxima lección.

Existen voltímetros especiales para medir voltaje RMS en formas de onda que no son senoidales.

Antes de estudiar las características de frecuencia y relación de fase en señales de CA, estudiaremos algunos términos necesarios para entender mejor estas características.

El ciclo

En términos generales un ciclo es una serie de fenómenos que suceden en un tiempo determinado. Los meses del año constituyen un ejemplo de la repetición cíclica.

El ciclo se comienza en Enero, se continúa a través de los meses restantes y termina en Diciembre que es el último mes del año. Tan pronto se cierra un ciclo anual, de inmediato se abre el siguiente y así sucesivamente.

Ahora analicemos el período cíclico de una onda senoidal. Como ya sabemos una onda senoidal está conformada por dos curvas iguales contrapuestas, la cual se repite indefinidamente, (figura 56).

Forma de un ciclo

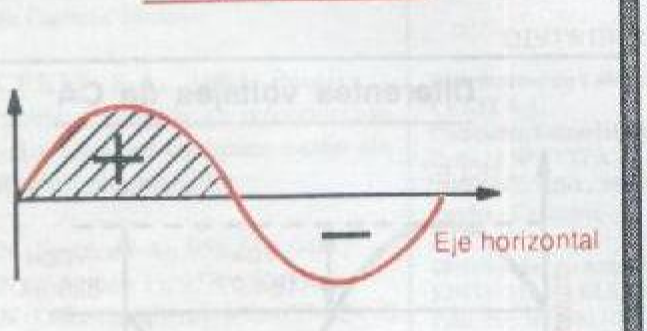

Fig. 56

La sección sombreada corresponde a la parte positiva ya que está por encima del eje horizontal; lo no sombreado es la parte negativa por estar debajo del eje horizontal. Si partimos del punto A y recorremos la onda hasta llegar al punto B, hemos completado un *ciclo*.

Un ciclo está compuesto por dos mitades de ciclos llamadas semiciclo positivo y semiciclo negativo. Hasta ahora hemos dicho que sobre el eje horizontal representamos el tiempo, el cual lo hemos dividido en milisegundos debido a que los cambios que se suceden en una onda ocurren en milésimas de segundo y en tiempos aún más cortos.

Expresando el ciclo en grados

La longitud del ciclo se mide en segundos, pero también se puede medir en grados de rotación. La longitud de un ciclo es 360º. (figura 57).

Ciclo en grados

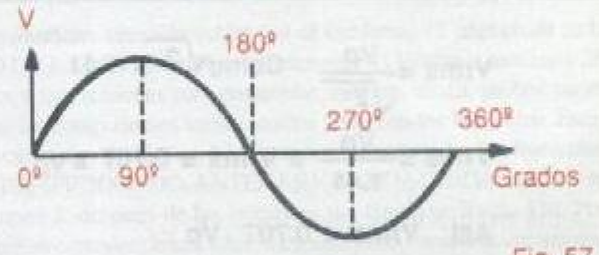

Fig. 57

La medida en grados se deriva de su relación con el giro mecánico de los generadores de corriente alterna. Cada que un alternador da un giro de 360º eléctricos, genera un ciclo completo de CA.

Si el alternador gira más rápido, se producirá mayor cantidad de ciclos, o sea el tiempo de duración de cada ciclo es más corto.

Frecuencia

Frecuencia es la cantidad de ciclos de una onda que ocurren en un período de tiempo; en electrónica se toma como período el segundo, es decir, la cantidad de veces que una señal de CA toma valores positivos y negativos en un segundo. En la figura 58 tenemos la representación gráfica de una señal de 4 ciclos por segundo.

Frecuencia en ciclos

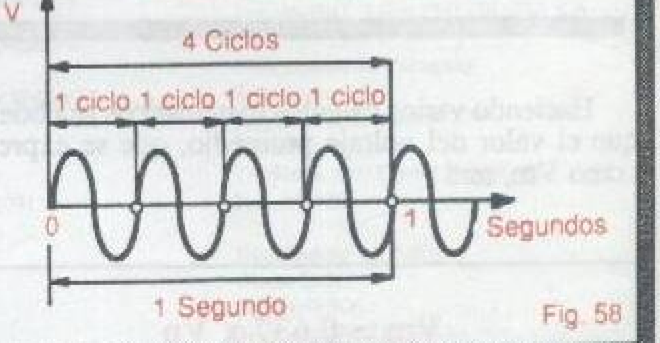

Fig. 58

La frecuencia de la tensión generada por un alternador tiene estrecha relación con la velocidad con que éste gira y con el número de polos o imanes con que esté fabricado.

En electrónica se expresa la frecuencia en Hertzios. Un Hertzio equivale a un ciclo por segundo.

Para cantidades grandes se utilizan los Kilohertzios y los Megahertzios y se representan por las abreviaturas KHz y MHz.

Un KHz es igual a 1.000 ciclos por segundo y 1 MHz es igual a 1.000.000 de ciclos por segundo.

Período

Es el tiempo necesario para que ocurra un ciclo completo, o sea el tiempo que transcurre desde que se inicia hasta que finaliza el ciclo.

El período se mide en segundos, milisegundos o microsegundos según sea la frecuencia y se representa con la letra T. Mientras más grande sea la frecuencia de una señal más corto es el período, ya que un ciclo tarda menos tiempo en suceder. (figura 59).

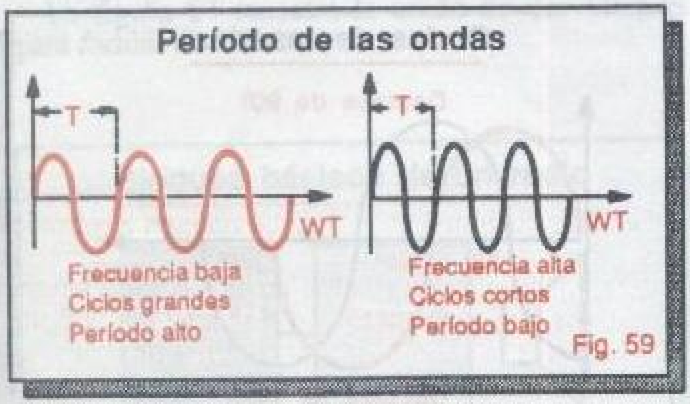

Fig. 59

Conociendo la frecuencia de una señal, se puede conocer el período. Su cálculo se realiza con la siguiente fórmula:

$$T = \frac{1}{F}$$

En donde:

T = Período de la onda dado en segundos.

F = Frecuencia de la onda dada en ciclos por segundo.

Ejemplo

Un alternador genera 60 ciclos por segundo. ¿Cuál es el período de la onda generada ?

$$T = \frac{1}{F}$$

$$T = \frac{1}{60} = 0.0167 \text{ Segundo}$$

Lo anterior quiere decir, que nuestro alternador tarda 0.0167 segundos o 16.7 milisegundos para generar un ciclo completo. Esto nos da una ligera idea de lo rápido que cambia una onda de CA.

Armónicos

En los circuitos de CA, sobre todo en los de radio, es muy común encontrar señales que tienen frecuencias con valores que son múltiplos enteros de una frecuencia original.

Así podemos tener señales con frecuencias de dos, tres, cuatro o cinco veces una determinada frecuencia. A estas señales se les llama *armónicos* y pueden ser pares (2, 4, 6, etc) o impares (3, 5, 7, etc). A la frecuencia original se le llama *fundamental*.

En la figura 60 podemos ver una onda seno fundamental y algunos de sus armónicos.

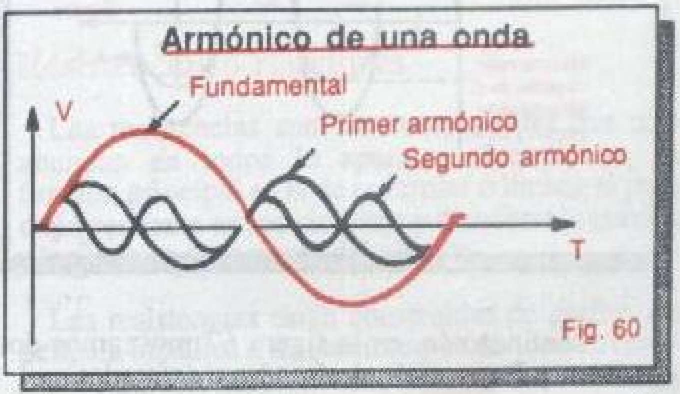

Fig. 60

Relación de fase o Fase

La relación de fase es el tiempo que separa dos hechos que están relacionados entre sí.

Por ejemplo, en una etapa contra reloj en la cual toman parte dos ciclistas llamados A y B, cuando parte el ciclista A ocurre un hecho: su partida; transcurrido un período de tiempo o espera,ocurre otro hecho relacionado: la partida del corredor B.

De lo anterior se puede decir que el ciclista A está adelantado en tiempo al ciclista B. Si los dos ciclistas hubieran partido al mismo tiempo y corrieran a la misma velocidad se diría que están en *fase*. En realidad los dos ciclistas están *desfasados o defasados*.

Refiriéndonos a ondas eléctricas, la fase tiene que ver con el adelanto o atraso de una señal u onda con respecto a otra.

Estas señales bien pueden ser corriente o voltaje. Dos señales están en fase si cumplen las siguientes condiciones:

- Los ciclos o frecuencia son iguales.
- Los valores máximos y mínimos ocurren en el mismo tiempo o ángulo.

La figura 61 muestra dos ondas en fase. Observemos que en 0º el valor de las ondas A y B es cero, para 90º las ondas A y B tienen su máximo valor, para 180º las ondas A y Btienen unvalor de cero, para 270º las ondas tienen su máximo valor negativo. En otras palabras las ondas A y B viajan juntas y alternan simultánenamente.

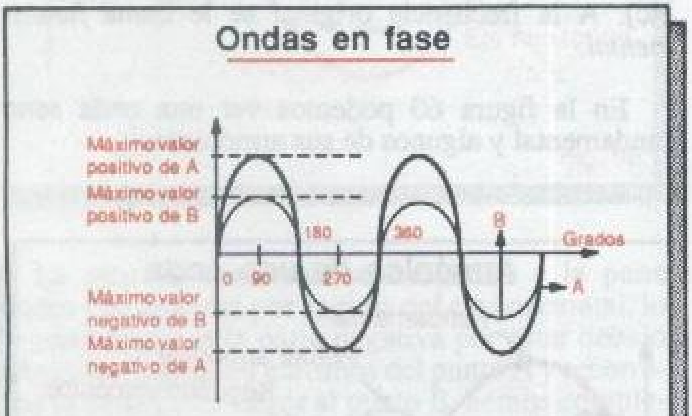
Máximo valor positivo de A
Máximo valor positivo de B
180
360
B
Grados
0
90
270
A
Máximo valor negativo de B
Máximo valor negativo de A

Ondas en fase

Fig. 61

A continuación, en la figura 62 mostramos dos ondas A y B que están desfasadas o defasadas.

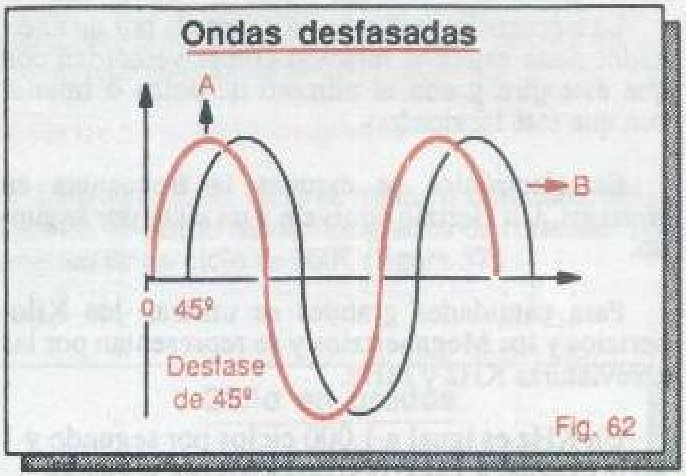

Ondas desfasadas

Fig. 62

La onda A está adelantada 45º con respecto a la onda B,o lo mismo, B está atrasada 45º con respecto a A.

Observemos que las dos ondas toman sus valores máximos y mínimos,cada una en tiempos o ángulos distintos, es decir no cambian al mismo tiempo.

La figura 63 muestra dos ondas desfasadas. En la figura A las ondas tienen un desfase de 90º y en la figura B el desfase es de 180º.

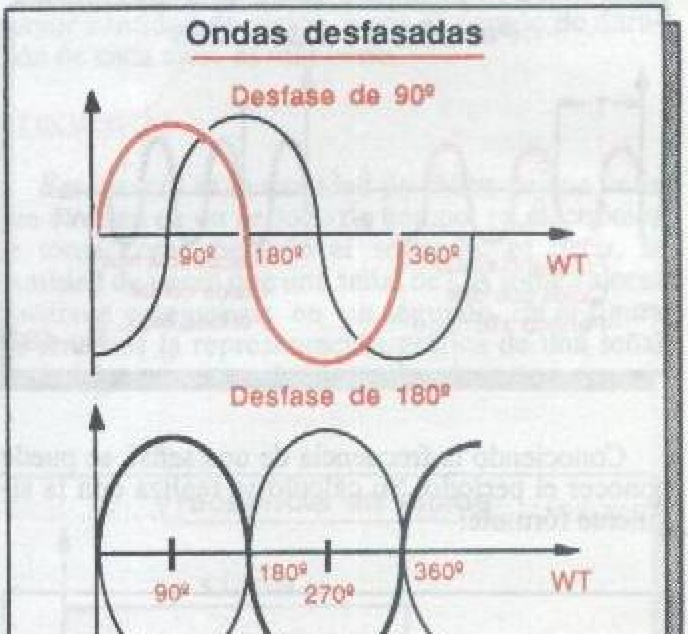

Ondas desfasadas

Fig. 63

En los circuitos de radio y electrónica donde se aplica corriente alterna a ciertos componentes como los condensadores y las bobinas, ocurren fenómenos de desfase entre la corriente y el voltaje que estaremos estudiando en una próxima lección.

Los componentes electrónicos

Cuando abrimos un televisor, un betamax, un equipo de sonido, un radio o cualquier aparato electrónico,nos sorprendemos al observar en su interior una gran cantidad de elementos extraños con múltiples formas, colores y tamaños. Son los componentes electrónicos.

Esto nos causa frustración y pensamos que la electrónica es muy difícil y no es para nosotros.

Pero se trata de una falsa apreciación, en verdad todo esto es muy fácil y está al alcance de todos.

Como se dará cuenta,un equipo electrónico tiene muchos componentes o partes pero éstos son de pocos tipos o grupos; es decir,hay muchos elementos repetidos, además están agrupados en bloques o circuitos que se identifican fácilmente, ya que cada uno cumple una función específica.

Un conjunto de elementos conectados adecuadamente, forman un circuito, uno o más circuitos forman un bloque y varios bloques forman un aparato electrónico.

La figura 64 muestra la unión de dos bloques para formar un radio.

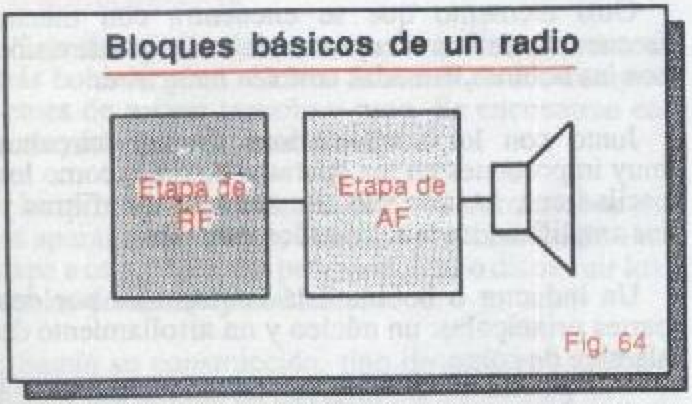

Fig. 64

La primera etapa llamada de radiofrecuencia o RF está compuesta por circuitos que se encargan de captar las señales del aire, dejar pasar una sola señal, separar esa señal y obtener los sonidos débiles para entregarla al bloque siguiente,llamado amplificador de audio.

De otra parte el bloque de audiofrecuencia o AF se encarga de recibir la señal del bloque de RF y amplificarla a un nivel adecuado para que pueda ser entregada al parlante.

Los circuitos electrónicos o bloques más utilizados en los aparatos de radio y comunicaciones son: amplificadores de radiofrecuencia, amplificadores de audio, osciladores, sintonizadores, detectores, filtros, controles automáticos de ganancia, rectificadores y reguladores de voltaje.

Todos estos circuitos se forman con los mismos componentes. La diferencia entre cada uno está en la forma como se conectan esos componentes entre sí,y en las señales que se le aplican.

Después de conocer los componentes electrónicos, estudiaremos en detalle cada uno de estos circuitos.

Incialmente conoceremos los componentes en forma individual, su funcionamiento básico , sus diferentes tipos y su símbolo.

Empecemos entonces por estudiar los componentes más utilizados en los circuitos electrónicos de los aparatos de radio. Los principales son:

- Las resistencias o resistores
- Los condensadores o capacitores
- Las bobinas y los transformadores
- Los diodos o rectificadores
- Los transistores
- Los circuitos integrados
- Los suiches o interruptores
- Los parlantes

Ahora conoceremos brevemente cada uno de estos componentes,con el fin de familiarizarnos con ellos para entender mejor las próximas lecciones.

Más adelante, en este curso, tendremos lecciones completas sobre los componentes más importantes en la teoría del radio,como son los condensadores, las bobinas y transformadores, los diodos, los transistores y los circuitos integrados.

Resistencias o resistores

Las resistencias son los componentes que más abundan en todos lo aparatos electrónicos, su función principal es la de controlar o limitar el paso de la corriente en los circuitos y distribuir la tensión adecuada en cada punto del circuito.

Las resistencias están construidas de grafito, de película metálica o de alambre y se dividen en resistencias fijas y resistencias variables. Las resistencias variables tienen un eje o tornillo que permite cambiar su valor. (figura 65).

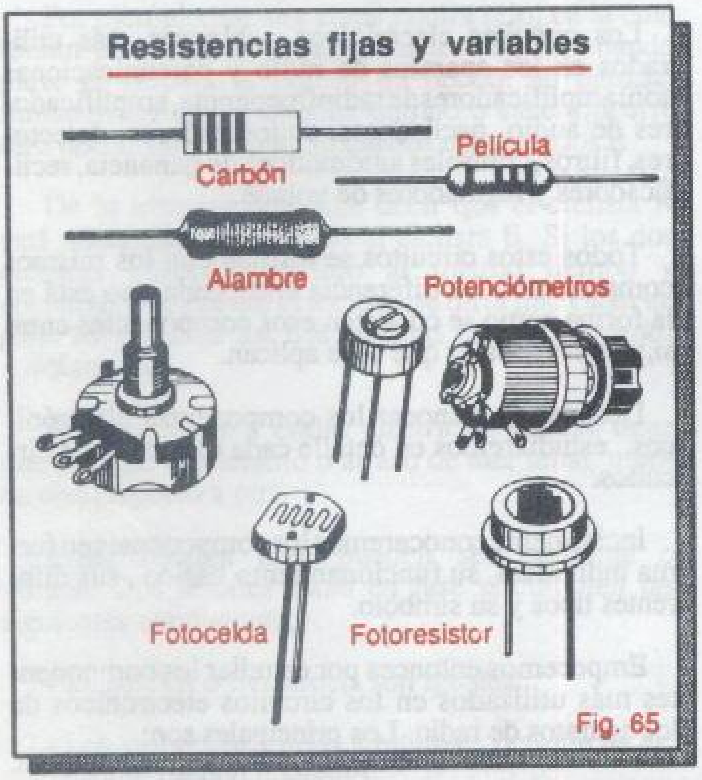

Fig. 65

Condensadores o capacitores

Los condensadores o capacitores son uno de los conjuntos de componentes más importantes que nunca faltan en todos los aparatos de radio y comunicaciones; se les encuentra en una gran variedad de formas y tamaños.

Un condensador está construido por dos láminas metálicas,separadas por un material aislante llamado dieléctrico.

La función principal de un condensador es la de almacenar energía en forma temporal, para luego devolverla al circuito.

Los condensadores se dividen en fijos y variables. Los condensadores fijos se dividen a su vez en condensadores polarizados o electrolíticos y los no polarizados.

Dentro del grupo de los condensadores fijos no polarizados están: los condensadores de poliéster, los condensadores de mica, los condensadores de cerámica y los condensadores de papel. (figura 66).

Cada tipo de condensador tiene su aplicación especial,dependiendo del voltaje, la frecuencia y el tipo de señal a manejar.

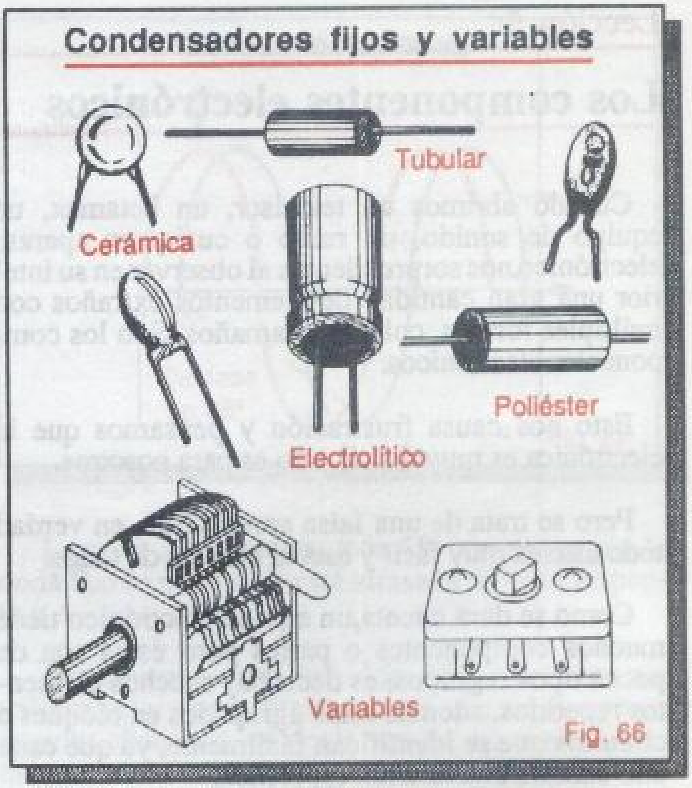

Fig. 66

Por ejemplo, los condensadores variables se utilizan en los circuitos de sintonía de las emisoras de los radios. Los condensadores cerámicos se utilizan para acoplar dos etapas o bloques, y los condensadores electrolíticos se utilizan en las fuentes de alimentación.

Bobinas

Otro elemento que se encuentra con mucha frecuencia en los receptores de radio y televisión son las bobinas,llamadas también *inductores*.

Junto con los condensadores forman circuitos muy importantes en los aparatos de radio,como los osciladores, los circuitos de sintonía, los filtros y los amplificadores sintonizados entre otros.

Un inductor o bobina está conformada por dos partes principales: un núcleo y un arrollamiento de alambre de cobre.

De hecho esto origina una extensa variedad de tipos, formas y tamaños de bobinas, cada una con una aplicación muy concreta,dependiendo del tipo de señal a manejar, la frecuencia y la magnitud de la corriente.

Tipos de bobinas

De acuerdo con el material utilizado como núcleo, estas pueden ser: con núcleo de hierro, con núcleo de ferrita,o con núcleo de aire.

Por su valor las bobinas pueden ser fijas o variables; una bobina variable tiene un núcleo ajustable que permite que éste esté más cerca o más alejado del arrollamiento de alambre,obteniéndose diferentes valores en su inductancia. (figura 67).

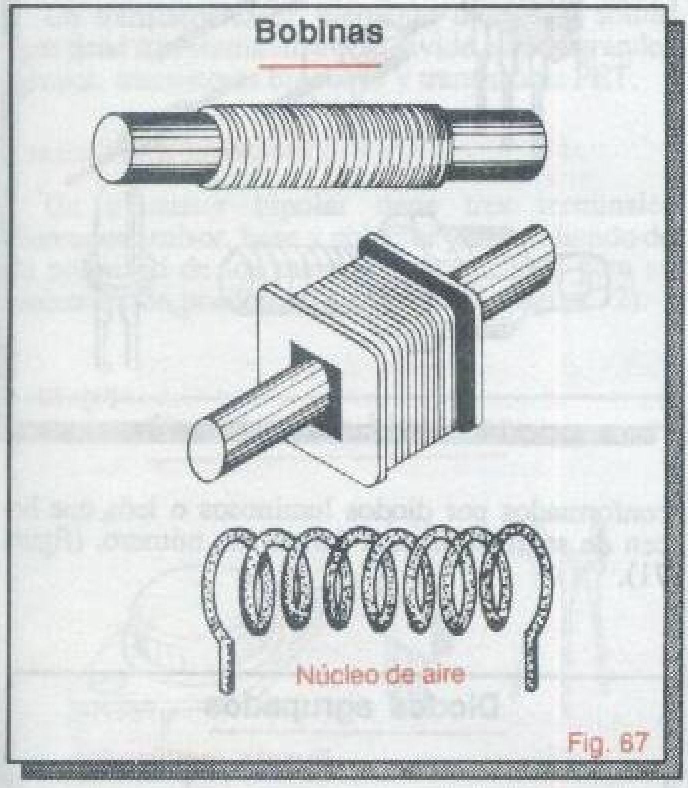

Fig. 67

Transformadores

Los transformadores están formados por dos o más bobinas y un núcleo. Son uno de los componentes de mayor tamaño y peso. Se encuentran en la mayoría de los aparatos de radio.

La principal función de los transformadores en los aparatos de radio es la de acoplar la señal de una etapa a otra,y también para aumentar o disminuir los niveles de voltaje y corriente.

Según su construcción, tipo de núcleo y aplicación,se encuentran en diferentes tipos. (figura 68).

En un transformador existe una bobina que recibe la señal llamada *primario* y una o más bobinas que entrega la señal (salidas), llamadas*bobinas secundarias* o secundario. (figura 69).

Los Semiconductores

Uno de los más grandes avances de la electrónica en las últimas décadas se debe al descubrimiento de los materiales semiconductores.

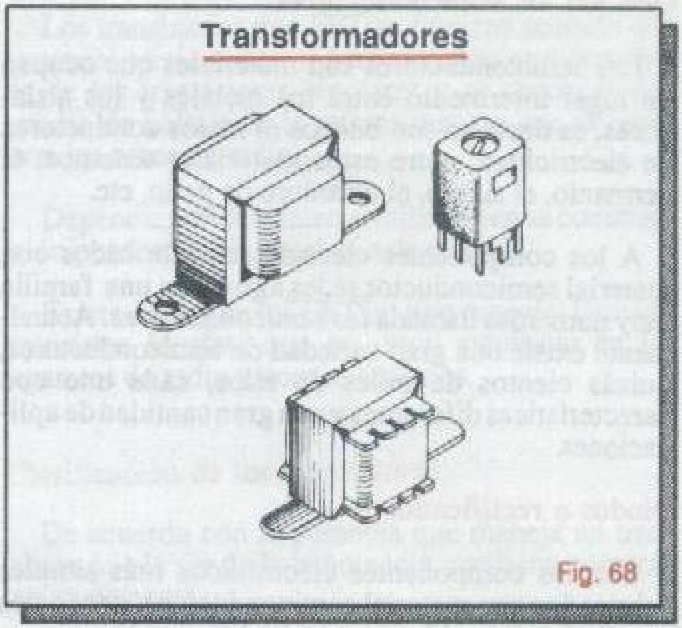

Fig. 68

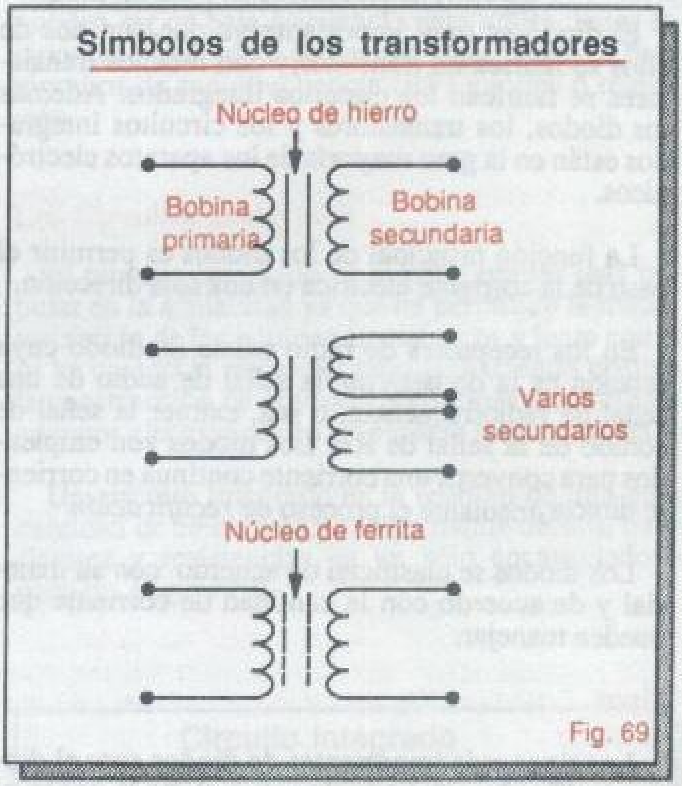

Fig. 69

Este descubrimiento hizo posible la creación de pequeños elementos como los diodos, los transistores y los circuitos integrados,los cuales son más compactos, consumen menos energía, cumplen con exactitud múltiples funciones y duran mucho más que sus antecesores, los tubos de radio.

Utilizando estos componentes como reemplazo de los tubos de radio o tubos de vacío, se han podido construir transmisores y receptores de radio y televisión mucho más avanzados,provocando una verdadera revolución en las comunicaciones humanas.

Qué son los semiconductores

Los semiconductores son materiales que ocupan un lugar intermedio entre los metales y los aisladores, es decir no son buenos ni malos conductores de electricidad; entre estos materiales tenemos: el germanio, el silicio, el arseniuro de galio, etc.

A los componentes electrónicos fabricados con material semiconductor,se les agrupa en una familia muy numerosa llamada *los semiconductores*. Actualmente existe una gran variedad de semiconductores, quizás cientos de miles de ellos, cada uno con características diferentes y una gran cantidad de aplicaciones.

Diodos o rectificadores

Son los componentes electrónicos más simples, fabricados con material semiconductor y tienen dos terminales conocidos como cátodo y ánodo.

El diodo es muy importante ya que,con dos de ellos se fabrica un transistor,y con muchos transistores se fabrican los circuitos integrados. Además los diodos, los transistores y los circuitos integrados están en la gran mayoría de los aparatos electrónicos.

La función principal de los diodos es permitir el paso de la corriente eléctrica en una sola dirección.

En los receptores de radio existe un diodo cuya función es la de *detectar* la señal de audio de una señal de radiofrecuencia, o sea, extraer la señal de sonido de la señal de RF. Los diodos son empleados para convertir una corriente continua en corriente directa,mediante el proceso de *rectificación*.

Los diodos se clasifican de acuerdo con su material y de acuerdo con la cantidad de corriente que pueden manejar.

Tipos

Los tipos más importantes de diodos son: el diodo rectificador, el diodo detector, el diodo zener, el diodo led o diodo emisor de luz, el fotodiodo y el SCR o rectificador controlado de silicio. (figura 70).

Cada uno de estos tiene una tarea específica,la cual explicaremos en su lección correspondiente.

Es frecuente encontrar en un solo empaque varios diodos; es el caso de los puentes rectificadores que traen internamente 2 a 4 diodos.

Igualmente sucede con los displays visualizadores numéricos de las calculadores, éstos están conformados por diodos luminosos o leds, que hacen de segmentos para formar un número. (figura 71).

Diodos

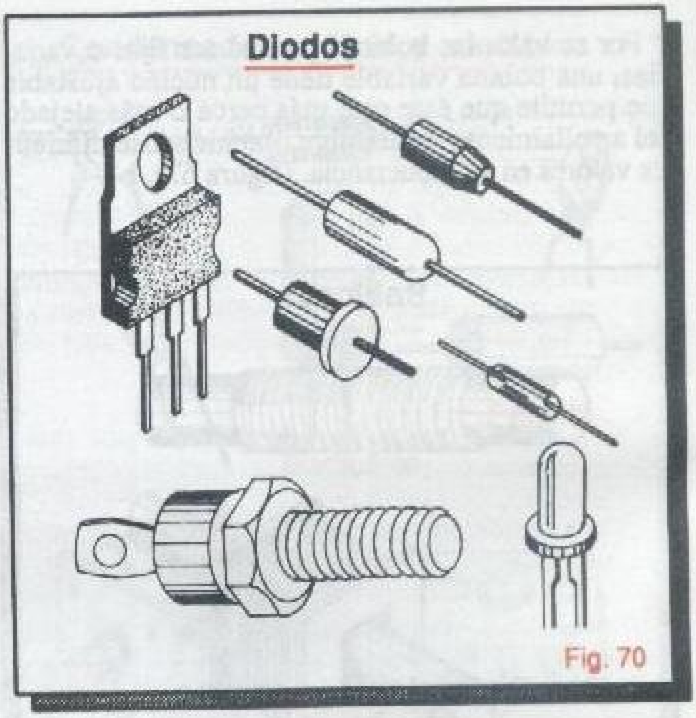

Fig. 70

Diodos agrupados

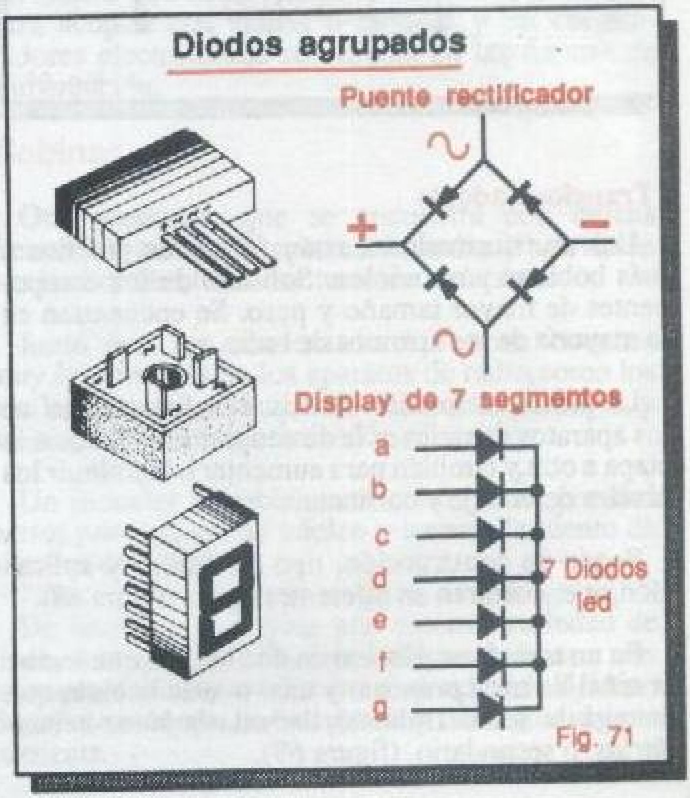

Fig. 71

Los Transistores

El transistor es el reemplazo de los tubos de vacío y fue el que marcó el inicio de la revolución de la electrónica.

El transistor consume poca energía, tiene una larga duración, sólida construcción y resiste las vibraciones. Tiene la cualidad de oscilar, operar como mezclador, modulador y suiche. Además, opera a altas temperaturas, resiste los choques, es pequeño y liviano.

Un transistor es un elemento de estado sólido que tiene tres terminales y se divide en dos grandes grupos: transistores bipolares y transistores FET.

Transistores bipolares

Un transistor bipolar tiene tres terminales llamados emisor, base y colector y, dependiendo de la polaridad de los materiales empleados para su construcción, pueden ser NPN o PNP. (figura 72).

Transistores bipolares

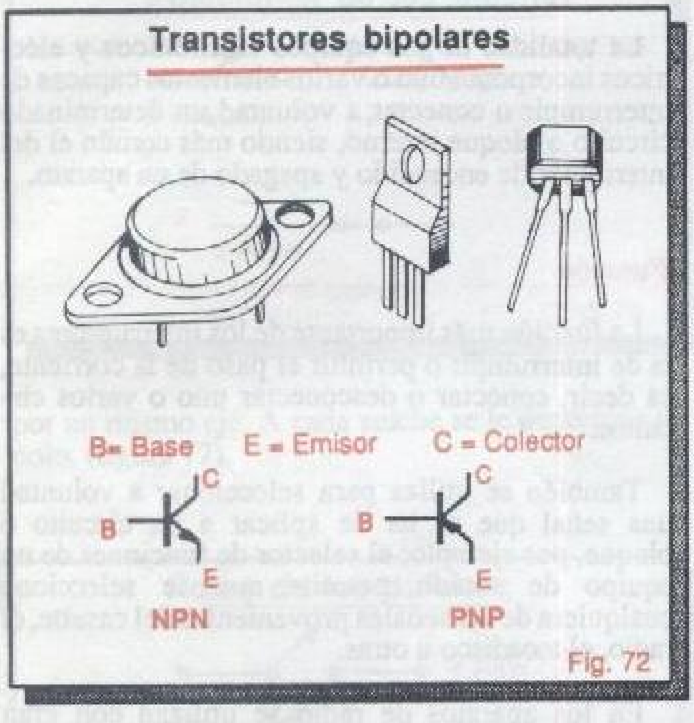

Fig. 72

Función de los transistores

A partir de una pequeña corriente de control que se aplica a la base, se obtiene una corriente amplificada en su salida.

Esta corriente permite que el transistor amplifique una señal, actúe como suiche, o pueda oscilar en conjunto con una bobina y un condensador para producir señales de radio o de audiofrecuencia.

Transistores FET (Field Effect Transitor)

Llamados también transistores de efecto de campo, tienen tres terminales llamados: Drenador (Drain), fuente (Source) y compuerta (Gate).

Los transistores tipo FET se fabrican también con material semiconductor, pero su forma es diferente a la de los transistores bipolares y se controlan por medio de voltaje, de la misma manera que se hacía con los tubos de radio.

Dependiendo del material utilizado en su construcción, pueden ser de canal N, o de canal P.

Existe dentro de los FET un tipo especial llamado transistor *Mosfet,* que es muy empleado en los aparatos de radio y comunicaciones.

Clasificación de los transistores

De acuerdo con la potencia que maneja un transistor puede ser de baja potencia, mediana potencia y alta potencia.

De acuerdo con la frecuencia y el tipo de señales, pueden ser de baja frecuencia para aplicaciones en circuitos de sonido o audio, y de alta frecuencia para circuitos de radiofrecuencia (RF) y video o televisión.

Los Circuitos Integrados

El circuito integrado es el componente más popular en la actualidad ya que ha permitido la miniaturización de los equipos electrónicos, y junto con él se ha creado la microelectrónica, haciendo posible la construcción de novedosos sistemas y complejos aparatos electrónicos en un espacio muy reducido.

Un circuito integrado es la reunión de una gran cantidad de elementos, principalmente diodos, transistores y resistencias, en un sólo encapsulado o paquete. (figura 73).

Circuito integrado

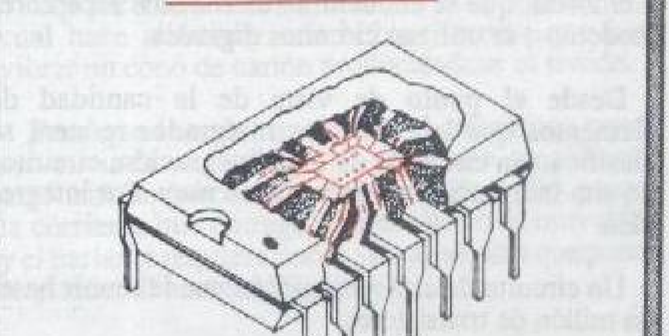

Fig. 73

Con un circuito integrado se pueden conformar uno o más bloques completos de un aparato de radio, sonido o televisión,razón por la cual ha reducido enormemente el tamaño de los equipos actuales.

Los circuitos integrados han facilitado enormemente la fabricación y reparación de los aparatos electrónicos.

Función

Las funciones que desempeñan los circuitos integrados no tienen límites, ya que se fabrican integrados para cada función específica que se necesite.

Existen, por ejemplo, circuitos integrados que conforman un circuito amplificador de sonido completo, existen otros que forman un oscilador, otros reguladores de voltaje, controles de temperatura, memorias de computadores, microprocesadores, etc.

Tipos

Desde el punto de vista de su aplicación existen dos grandes grupos de circuitos integrados: los circuitos integrados digitales y los circuitos integrados análogos o lineales.

Un circuito integrado digital maneja señales digitales, es decir, sólo admite una señal que toma los dos valores posibles: uno y cero, o alto y bajo.

Un circuito integrado análogo o lineal maneja señales que varían en el tiempo,tomando diferentes valores. Las ondas de radio que llegan a la antena son análogas,lo mismo que el sonido que entrega un parlante.

La mayoría de los circuitos de radio utilizan este tipo de circuitos integrados lineales, aunque para las funciones de sintonía y programación digital con memorias, que se encuentran en muchos receptores modernos, se utilizan circuitos digitales.

Desde el punto de vista de la cantidad de elementos que los circuitos integrados reúnen, se clasifican en circuitos de baja integración, circuitos de alta integración y circuitos de muy alta integración.

Un circuito de alta integración puede reunir hasta un millón de transistores.

En la figura 74 podemos observar algunos tipos de circuitos integrados,utilizados en aparatos electrónicos.

Tipos de Circuitos Integrados

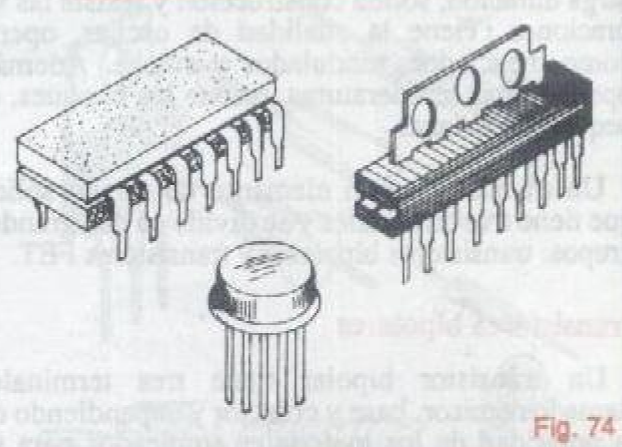

Fig. 74

Suiches o interruptores

La totalidad de los equipos electrónicos y eléctricos incorporan uno o varios elementos capaces de interrumpir o conectar, a voluntad, un determinado circuito o bloque interno, siendo más común el del interruptor de encendido y apagado de un aparato.

Función

La función más importante de los interruptores es la de interrumpir o permitir el paso de la corriente, es decir, conectar o desconectar uno o varios circuitos.

También se utiliza para seleccionar a voluntad una señal que se ha de aplicar a un circuito o bloque, por ejemplo: el selector de funciones de un equipo de sonido permite que se seleccione cualquiera de las señales provenientes del casette, el radio, el tocadisco u otras.

En los aparatos de radio se utilizan con gran frecuencia los suiches,en el caso de selección de las diferentes bandas y otras funciones.

Tipos

Los principales tipos de suiches son: deslizantes, de palanca, de codillo, pulsadores, rotatorios, electromagnéticos o relés, de mercurio, de imán y de estado sólido. (figura 75).

Respecto al número de contactos pueden ser de una posición, dos posiciones, tres posiciones, cuatro posiciones o más. (figura 76).

Pueden existir suiches independientes eléctricamente pero que son accionados simultáneamente

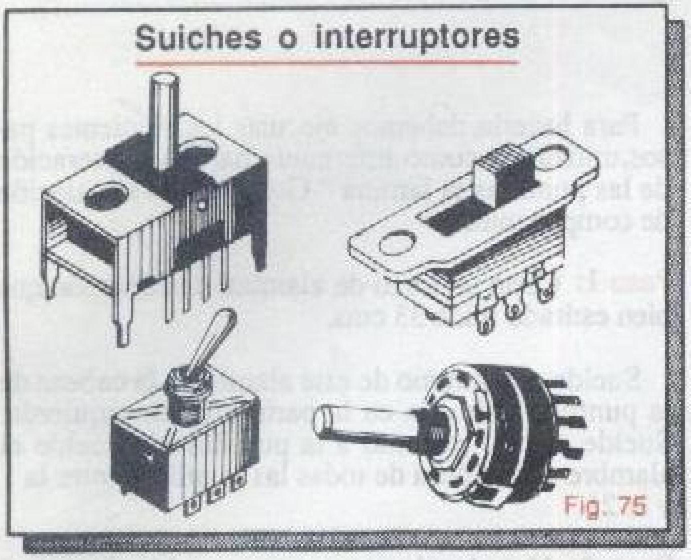

Suiches o interruptores

Fig. 75

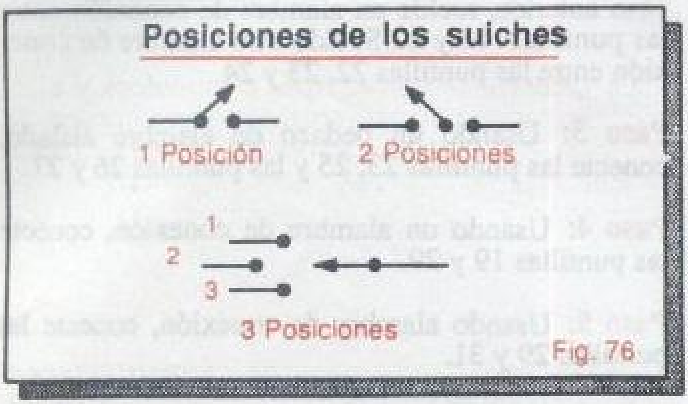

Posiciones de los suiches

Fig. 76

por un mismo eje. A cada suiche se le denomina un polo. (figura 77).

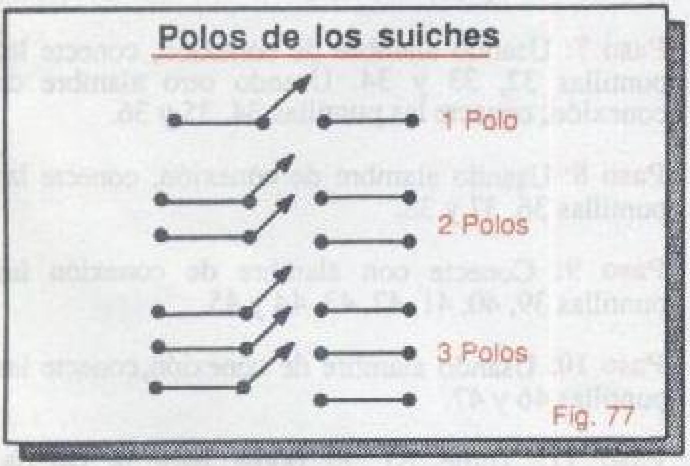

Polos de los suiches

Fig. 77

Un suiche puede ser de varios polos y varias posiciones en el caso de los suiches rotatorios o deslizantes muy utilizados en los aparatos de radio. Al grupo de suiches llamados pulsadores corresponden los teclados de las calculadoras y los computadores. Un pulsador puede ser normalmente abierto o normalmente cerrado. (figura 78).

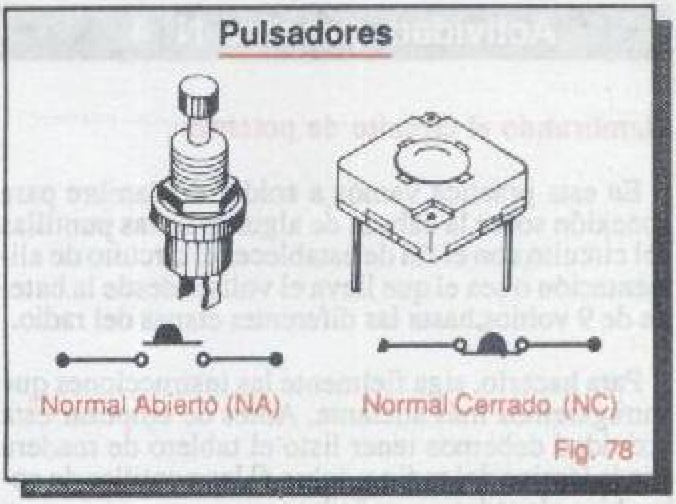

Pulsadores

Fig. 78

Micrófonos y parlantes

Con la invención del micrófono y el parlante se inventó el teléfono y posteriormente el amplificador de audio.

Los micrófonos y los parlantes son elementos trasductores que convierten un tipo de energía en otro.

En el caso del micrófono, éste convierte energía sonora o de movimiento de aire en energía eléctrica.

El parlante convierte energía eléctrica en energía sonora. Los parlantes y micrófonos realizan tareas opuestas.

Función

Los micrófonos captan las vibraciones de las ondas sonoras y las convierte en corriente eléctrica que oscila y varía en la misma forma que el sonido. En las emisoras de radio es de gran importancia la calidad de los micrófonos utilizados,ya que de ellos depende en gran parte la pureza del sonido emitido al aire.

Los parlantes reciben una corriente fluctuante la cual hace alimentar una bobina interna que hace vibrar un cono de cartón produciéndose el sonido.

Para que la señal que capta el micrófono pueda ser enviada al parlante, es necesario amplificarla mediante circuitos amplificadores de audio, ya que la corriente que entrega un micrófono es muy débil y el parlante requiere corrientes altas para que pueda responder al sonido. (figura 79).

Tipos

Los principales tipos son: de carbón, dinámicos o magnéticos, de condensador y de cristal.

Actividad práctica Nº 4

Alambrando el circuito de potencia

En esta práctica vamos a soldar el alambre para conexión sobre la cabeza de algunas de las puntillas del circuito,con el fin de establecer el circuito de alimentación o sea el que lleva el voltaje desde la batería de 9 voltios,hasta las diferentes etapas del radio.

Para hacerlo, siga fielmente las instrucciones que entregaremos más adelante. Antes de empezar esta actividad debemos tener listo el tablero de madera con la lámina del radio,y, sobre él,las puntillas de cobre bien clavadas y estañadas.

Si usted nunca ha soldado antes, estudie o repase la lección sobre soldadura que entregamos en la práctica anterior.

Herramientas y materiales necesarios

Cautín de 25 ó 40 vatios
Cortador de alambre o corta-frío
Pinzas de punta plana y alicate
3 Metros de alambre sólido,sin aislar,para conexiones
2 Pedazos de alambre aislado de 15 cms
1 Conector para batería de 9 voltios
1 Trozo de soldadura de estaño de 60/40

Alambrando el sistema de alimentación de potencia

En esta actividad conectaremos todos los alambres y partes mostradas en color rojo sobre el diagrama, menos los condensadores C10 y C14 y la resistencia R12,que se instalarán en una próxima actividad.

Importante: El alambre debe estirarse lo más que se pueda para lograr una buena presentación y asegurar buena calidad en el ensamble del radio.

En la figura A6 podemos observar el aspecto correcto de cómo deben quedar los alambres de conexión.

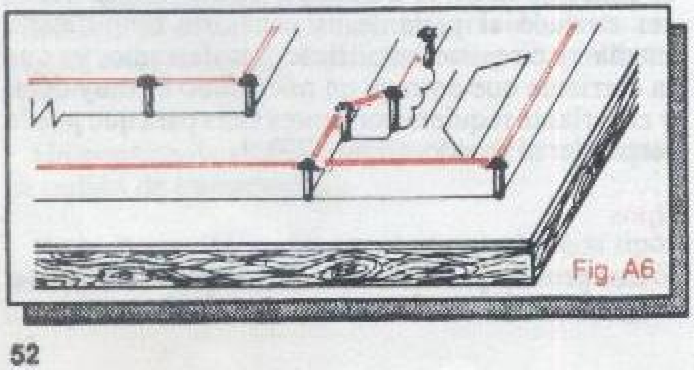

Fig. A6

Para hacerlo, debemos ejecutar los siguientes pasos,utilizando como referencia para la numeración de las puntillas la lámina " Guía para la instalación de componentes ":

Paso 1: Corte un trozo de alambre de conexión,que bien estirado mida 33 cms.

Suelde un extremo de este alambre a la cabeza de la puntilla 1, situada en la parte inferior izquierda. Suelde el otro extremo a la puntilla 21. Suelde el alambre a la cabeza de todas las puntillas entre la 1 y la 21.

Paso 2: Usando el mismo proceso descrito en el paso anterior, suelde un alambre de conexión entre las puntillas 18 y 22.Suelde otro alambre de conexión entre las puntillas 22, 23 y 24.

Paso 3: Usando un pedazo de alambre aislado, conecte las puntillas 23, 25 y las puntillas 26 y 27.

Paso 4: Usando un alambre de conexión, conecte las puntillas 19 y 29.

Paso 5: Usando alambre de conexión, conecte las puntillas 29 y 31.

Paso 6: Usando alambre de conexión, conecte las puntillas 30 y 31.

Nota: En este punto usted ya ha conectado la línea negativa de su radio y ahora se conectará la línea de alimentación positiva.

Paso 7: Usando alambre de conexión, conecte las puntillas 32, 33 y 34. Usando otro alambre de conexión, conecte las puntillas 34, 35 y 36.

Paso 8: Usando alambre de conexión, conecte las puntillas 36, 37 y 38.

Paso 9: Conecte con alambre de conexión las puntillas 39, 40, 41, 42, 43, 44 y 45.

Paso 10: Usando alambre de conexión,conecte las puntillas 46 y 47.

Paso 11: Tome B1, el borne para la batería. Conecte el alambre negro en la puntilla 28 y el rojo a la puntilla 32.

Usted puede pegar la batería a la tabla con un pedazo de cinta de doble cara, o unir la batería a la tabla, usando una abrazadera o soporte para batería. Si no está provista se puede hacer de una pequeña faja de lámina de acero inoxidable.

Amplificador para micrófono

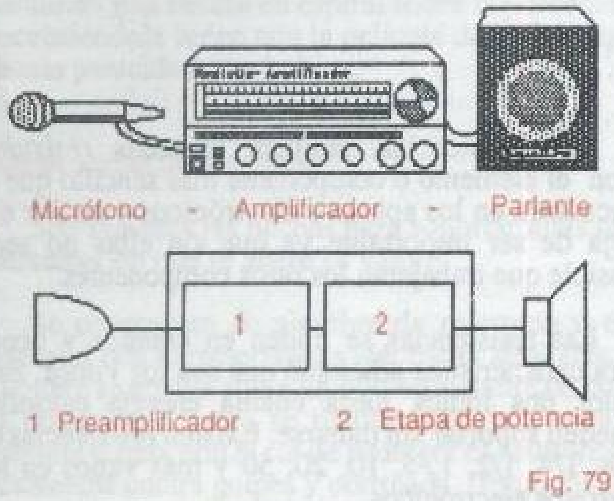

Fig. 79

Respecto a los parlantes los principales tipos son: el magnético y el de cristal.

Existen tipos de parlantes especiales, como los audífonos, que están construidos para entregar baja potencia y alta calidad de sonido.

Respecto a la frecuencia o tonos que los parlantes manejan se clasifican en tres grupos: parlantes para bajos, parlantes para medios, y parlantes para sonidos altos o *tweeters*. (figura 80).

Parlantes y audífonos

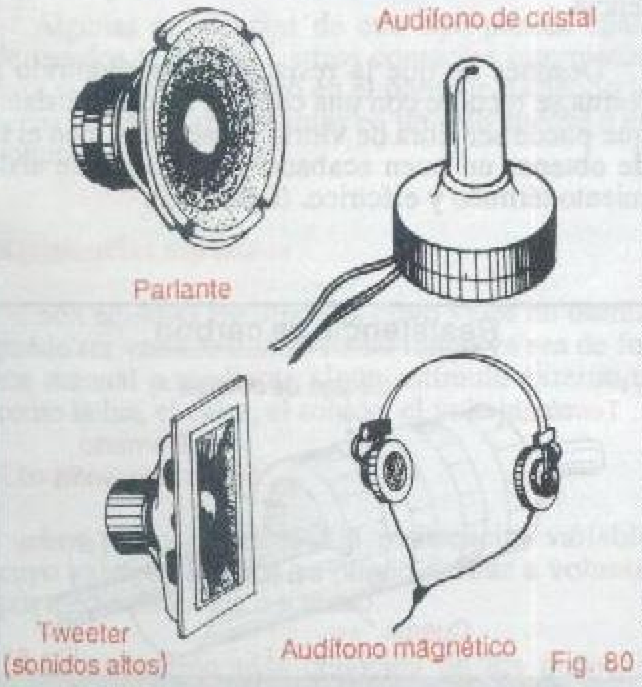

Fig. 80

Los parlantes también se clasifican por su potencia en parlantes de baja, media y alta potencia, expresada en vatios.

El tamaño de los parlantes depende de la frecuencia de trabajo y de la potencia sonora que pueden transmitir.

Otra característica de un parlante es la *impedancia* o resistencia de la bobina .

Esta puede ser de 4 Ω, 8 Ω, 16 Ω o 32 Ω. La potencia nada tiene que ver en este caso con la impedancia.

Para lograr un mejor sonido y buena reproducción de los tonos, los parlantes se deben encerrar en cajas de madera especiales llamadas *baffles*. (figura 81).

Es posible mejorar notablemente el sonido de cualquier aparato de radio solamente conectando un parlante externo con su bafle adecuado.

Para esta conexión se debe tener en cuenta que el radio suministre la potencia mínima para alimentar ese parlante.

Cajas acústicas o baffles

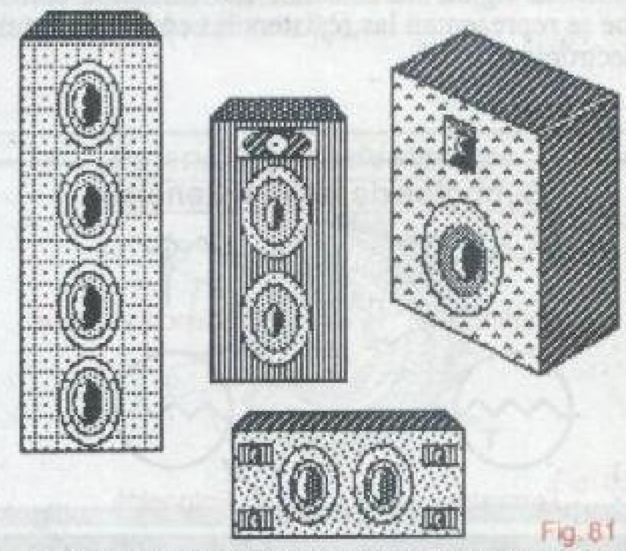

Fig. 81

El tamaño de la caja acústica depende de la potencia total de los parlantes, que están instalados en ella, y también de la frecuencia mínima de trabajo.

Para frecuencias bajas, los bafles son más grandes que para frecuencias medias o altas. Cuando hay varios parlantes para distintas frecuencias se utiliza un filtro para separar cada banda de sonidos.

Lección 6

Las Resistencias

Tipos y características

Como se sabe, la resistencia considerada en un sentido general es la oposición que presenta un material al paso de la corriente.

Tomado en un sentido particular y común la palabra resistencia se refiere a un elemento o componente muy utilizado en electricidad y electrónica. Este elemento está construido con un material resistivo y adopta diversas formas y tamaños dependiendo de su uso.

En electricidad las resistencias se utilizan para producir calor. Es el caso de las estufas, hornos, planchas y calentadores eléctricos.

En electrónica las resistencias tienen formas y tamaños más reducidos,y su utilidad está en la propiedad que éstas tienen de limitar y controlar el paso de la corriente,distribuyendo los niveles de voltaje diferentes que requiere cada parte de un circuito.

Existen dos grupos principales de resistencias, las resistencias *fijas* y las resistencias *variables.*

En la figura 82 tenemos los símbolos con los que se representan las resistencias en los diagramas electrónicos.

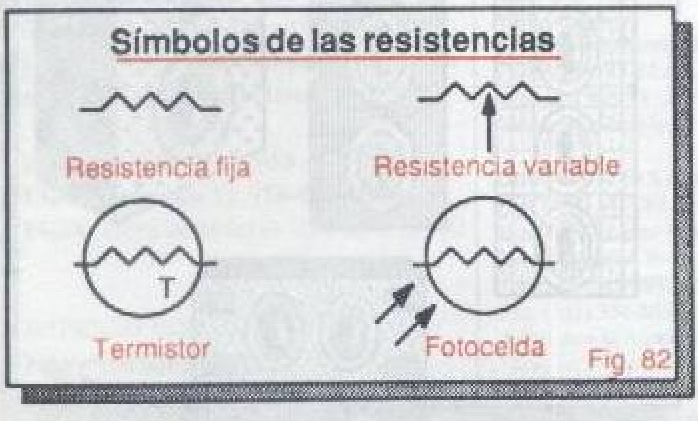

Fig. 82

Al observar los circuitos internos de un radio, un televisor, una grabadora o un betamax, notaremos que siempre habrá una buena cantidad de resistencias interconectadas a los demás elementos del aparato.

Ya habrá notado que hay una gran cantidad de ellas en el plano del radio que usted está armando, como complemento de este curso.

Las resistencias, también llamadas *resistores,* son el elemento o componente más sencillo que se encuentra en los aparatos electrónicos y no por eso deja de ser importante, ya que sin ellas no sería posible que trabajaran los otros componentes.

Las resistencias se miden en *Ohmios* y tienen otra característica adicional que son los *Vatios.* Este valor nos indica hasta cuánta energía calorífica pueden soportar sin dañarse. Existen resistencias de 1/8, 1/4, 1/2, 1, 5, 10, 20, 50 y más vatios en los aparatos electrónicos.

Resistencias fijas

A este grupo pertenecen todas las resistencias que presentan un mismo valor, sin tener la posibilidad de modificarlo a voluntad.

De acuerdo con su material de construcción, las resistencias fijas se clasifican en dos grandes grupos principales: de carbón y de alambre o bobinadas.

Resistencias de carbón

Estas resistencias se fabrican de grafito (carbón casi puro) o de carbón puro pulverizado. El grafito se mezcla con talco y con un material adhesivo, y se comprime con una prensa para darle forma y consistencia.

Después de que la resistencia ha adquirido su forma, se recubre con una capa de material aislante, que puede ser fibra de vidrio o baquelita, con el fin de obtener un buen acabado y proveerla de aislamiento térmico y eléctrico. (figura 83).

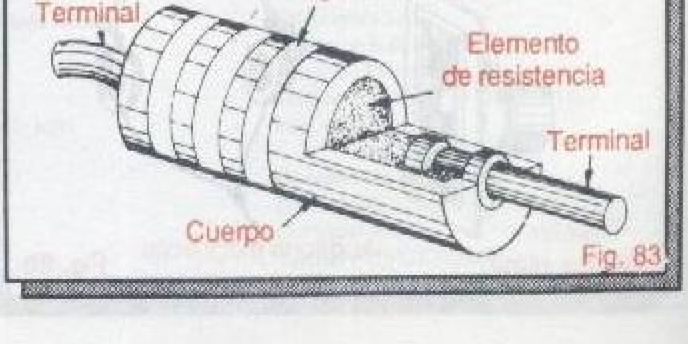

Fig. 83

Existe otro método de fabricación de las resistencias de carbón que consiste en recubrir un tubo o cilindro de porcelana con una película de carbón, o haciendo una ranura en espiral sobre la porcelana y recubriéndola luégo con la película de carbón, quedando parecida a una bobina.

Resistencias de alambre

Son resistencias hechas para soportar altas temperaturas sin que se altere su valor.

Se construyen de alambre de micromo u otro material con características eléctricas similares. El alambre se arrolla sobre un núcleo aislante,tal como la cerámica, y luégo se recubre con una capa de esmalte vítreo, con el fin de proteger el alambre y la resistencia contra golpes y corrosión. (figura 84).

Resistencia de alambre

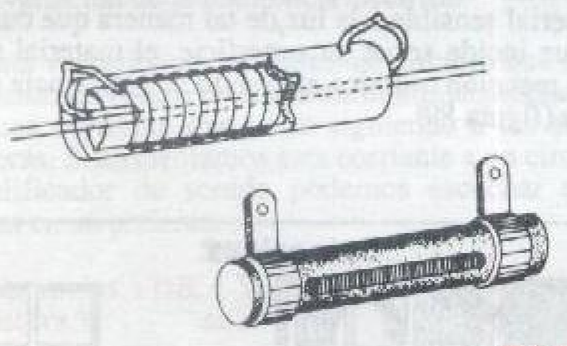

Fig. 84

Algunas resistencias de este tipo poseen aparte de sus dos terminales, otros contactos intermedios para facilitar un cambio en el ohmiaje,haciendo las veces de varias resistencias en un sólo cuerpo o envoltura.

Resistencias variables

Son aquellas resistencias cuyo valor en ohmios puede ser variado dentro de un rango,ya sea de forma manual o mediante algún estímulo externo,tal como la luz, el calor, el sonido, el voltaje, etc.

Los potenciómetros

Los potenciómetros son resistencias variables, cuyo valor en ohmios se puede ajustar a voluntad por medio de un eje o tornillo.

La aplicación más conocida de los potenciómetros la tenemos en los controles de volumen y tono en los aparatos de radio y de sonido, en el control de brillo y contraste en los televisores y para fines especiales en algunos instrumentos electrónicos.

Los potenciómetros se fabrican depositando una capa de carbón sobre una sección circular o rectangular de fibra o material compacto y aislante. Un eje en el centro permite que un contacto móvil se deslice a través de la sección resistiva. (figura 85).

Potenciómetro de carbón

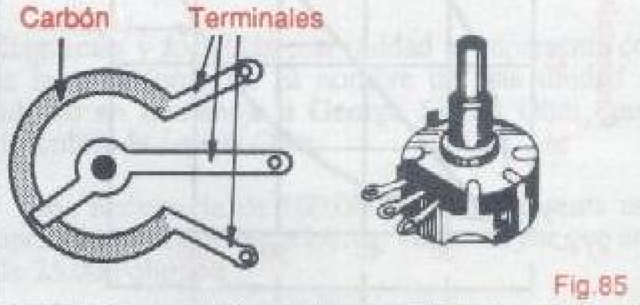

Fig.85

Existen también potenciómetros de alambre, generalmente de bajo valor en ohmios. Estos son costosos pero son más exactos y seguros. En un potenciómetro de este tipo,el alambre resistivo está sobre la superficie aisladora,de tal manera que el contacto móvil se desliza sobre la superficie recubierta de alambre.(figura 86).

Potenciómetro de alambre

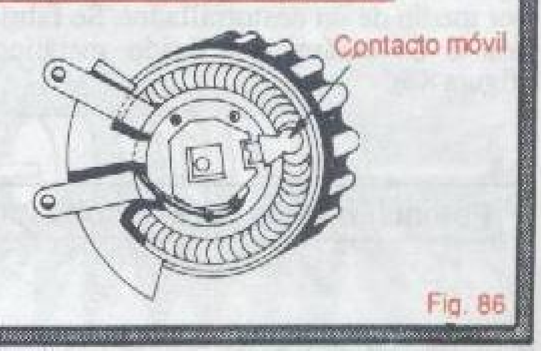

Fig. 86

Un potenciómetro puede ser lineal, logarítmico o antilogarítmico. Un potenciómetro lineal es aquel cuya variación es constante durante el giro del eje o cursor. Por ejemplo, si se gira 15º la resistencia aumenta 1.000 Ω, y si se gira 30º la resistencia aumenta 2.000Ω.

En un potenciómetro logarítmico o antilogarítmico no ocurre esto, se obtiene menos variación al principio y mayor variación al final del giro.

En la figura 87 se pueden observar los diferentes comportamientos o curvas de resistencia para cada tipo.

Curvas de potenciómetros

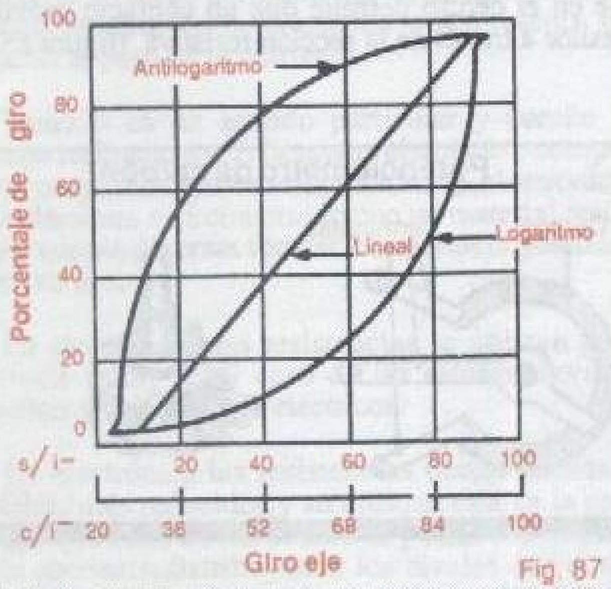

Fig. 87

Existe un tipo de potenciómetro que se fabrica especialmente para ser montado en los *circuitos impresos,* que son esas tabletas o tarjetas de baquelita en donde se montan los componentes de un aparato y se realizan todas las soldaduras para sus conexiones.

Estos potenciómetros se utilizan para ajustar voltajes o corrientes en algunos circuitos y se mueven por medio de un destornillador. Se fabrican de carbón o de cerámica y óxido metálico (cermet). (figura 88).

Potenciómetros para circuito impreso

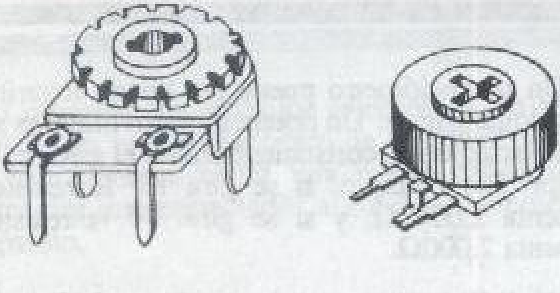

Fig. 88

Termistores

Un termistor es una resistencia especial que varía con la temperatura y puede ser de características directas o inversas.

Un termistor con características directas es aquel cuyo valor en Ohmios aumenta ante un incremento de la temperatura, mientras que un termistor con características inversas aumenta su valor ante una disminución de la temperatura.

Los termistores se utilizan como protección de sobrecorriente en algunas etapas de audio,o como sensores de calor en controles electrónicos de temperatura.

La fotocelda

Una fotocelda es una resistencia,cuyo valor en ohmios varía ante las variaciones de la luz.

Estas resistencias están construidas con un material sensible a la luz,de tal manera que cuando la luz incide sobre su superficie, el material sufre una reacción química,alterando su resistencia eléctrica.(figura 89).

Fotoceldas

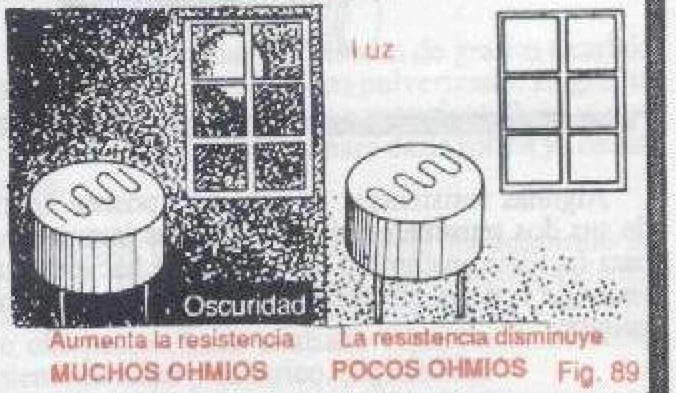

Fig. 89

Una fotocelda presenta un bajo valor de su resistencia ante la presencia de luz,y,un alto valor de resistencia ante la ausencia de luz.

La fotocelda se emplea para controlar el encendido automático del alumbrado público. También se utiliza ampliamente en circuitos contadores electrónicos de objetos y personas, en alarmas, etc.

Micrófono de carbón

Un micrófono de carbón es una resistencia que cambia su valor en ohmios ante las variaciones de sonido. Estos micrófonos están construidos por una membrana metálica que vibra al ser golpeada por la ondas sonoras. (figura 90).

Micrófono de carbón

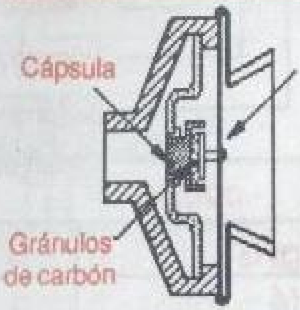

Fig.90

Las vibraciones de sonido hacen que se compriman con mayor o menor intensidad los gránulos de carbón que están aprisionados entre las dos placas, la membrana o diafragma y la caja, obteniéndose una variación de la resistencia eléctrica.

Esta variación se utiliza en un circuito que tiene aplicado un voltaje de corriente continua,produciendo una corriente que varía siguiendo a las ondas sonoras. Si alimentamos esta corriente a un circuito amplificador de sonido podemos escuchar estas ondas en un parlante.

Resistencias VDR (Voltage dependent resistors)

Son otro tipo especial de resistencias llamadas también varistores. Su valor en ohmios disminuye ante un aumento de la tensión aplicada.

Los varistores se emplean en circuitos de protección contra sobretensiones o picos abruptos de voltaje,que podrían dañar los delicados circuitos electrónicos.

Cuando el voltaje aplicado a sus terminales se pasa del valor para el cual está fabricado, su resistencia baja notablemente,haciendo que el pico de voltaje se reduzca. Cuando el pico de voltaje es de mucha duración, se ponen en cortocircuito y no dejan pasar ese pico hacia el resto del circuito, protegiéndolo.

En la figura 91 tenemos el símbolo, la conexión y un circuito típico de aplicación para un varistor.

Unidad de medida

Toda resistencia o resistor tiene cierta cantidad de *ohmios*, que es su unidad de medida. Para los diagramas y fórmulas esta unidad se representa con la letra Ω (omega). El nombre de esta unidad se adoptó en homenaje a George Simon Ohm,quien descubrió la *Ley de Ohm*.

Varistor

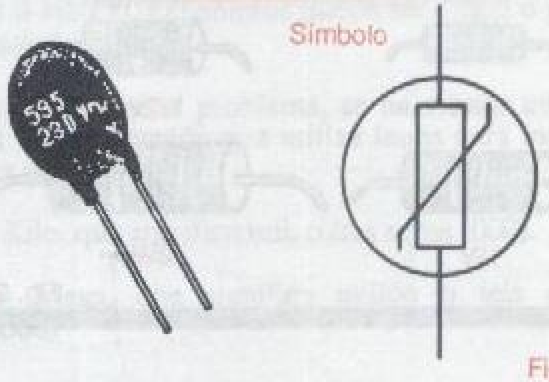

Fig.91

Una resistencia de 100.000 ohmios presenta una oposición a la corriente cuatro veces mayor que una de 25.000 ohmios.

Otras características

Además de su tipo, y su valor en ohmios, las resistencias tienen una característica adicional. Se llama *vatiaje* y representa cuánta energía se disipa en ellas, ya que cuando circula corriente por una resistencia, se produce calor a su alrededor. (figura 92).

Una resistencia disipa calor

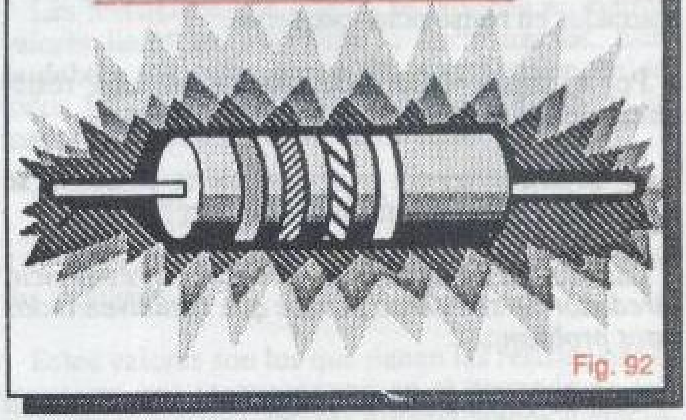

Fig. 92

Este vatiaje determina el tamaño físico de las resistencias,siendo las más grandes las de mayor vatiaje.

Las resistencias de vatiajes altos como 5, 10, 20, 50, 100 W, etc se fabrican de alambre y las de vatiajes bajos para los circuitos electrónicos se fabrican de carbón o de película de carbón en tamaños de 1/8, 1/4, 1/2, 1 y 2 vatios. (figura 93)

Tamaño de las resistencias

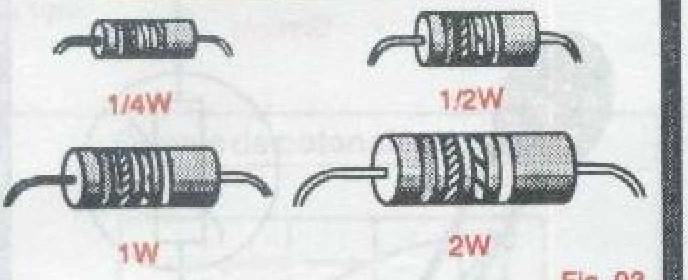

Fig. 93

El código de colores

El valor en ohmios de las resistencias se puede indicar de dos formas. Una de ellas es marcando este número en su cuerpo o utilizando el código de colores.

Para las resistencias de alambre o de carbón de 1 vatio en adelante,es fácil utilizar el primer método pero, para las resistencias más pequeñas es muy difícil hacerlo,ya que su tamaño lo impide.

Para las resistencias pequeñas de carbón y película de carbón, que son las más utilizadas en los circuitos de radio y comunicaciones, se inventaron un código de colores utilizando tres, cuatro o cinco líneas de colores que se pintan alrededor del cuerpo de la resistencia para indicar su valor en Ohmios,y el rango de tolerancia o precisión.

El sistema de las líneas de colores resuelve dos problemas principalmente:

Uno: Sería demasiado difícil ver números grandes marcados en resistencias pequeñas.

Por ejemplo: 1.000.000 ohmios en una resistencia de 1/4 de vatio no se vería muy bien.

Dos: Si la resistencia queda en cierta posición, se taparía este número y no se podría leer su valor.

Las bandas de colores que tiene toda resistencia alrededor de su cuerpo,parece que resuelven todos estos problemas.

Cuando se va a leer el código de colores de una resistencia, se debe colocar de esta forma: la línea o banda de color que está más cerca del borde es el primer número, la segunda línea es el segundo número, la tercera línea es el número de ceros o multiplicador,y la cuarta línea o banda es la tolerancia.

Además,cada color corresponde a un número en particular. (figura 94).

Código de colores

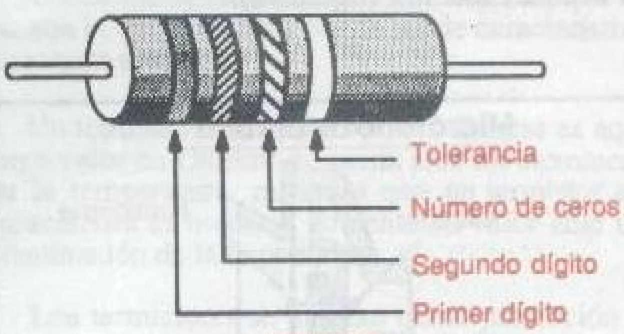

CODIGO

Negro	0
Café	1
Rojo	2
Naranja	3
Amarillo	4
Verde	5
Azul	6
Violeta	7
Gris	8
Blanco	9

Cada color representa un número

Fig. 94

Cuando leemos el código de colores debemos recordar:

1. La primera banda representa la primera cifra.

2. La segunda banda representa la segunda cifra.

3. La tercera banda representa el número de ceros que siguen a los dos primeros números. (Si la tercera banda es negra,no hay ceros en el número,y si esta banda es plateada,se divide por 10).

4. La cuarta banda representa la tolerancia. Esta es usualmente dorada que representa un 5%, plateada que es del 10% y,si no tiene banda,es del 20%.

En la figura 95 tenemos dos ejemplos de utilización del código de colores.

Es muy importante practicar mucho con este código hasta que se aprenda de memoria,ya que las resistencias que lo utilizan se encuentran en gran número en todo tipo de circuitos,y, si tenemos que consultar un libro o manual,cada vez que tengamos

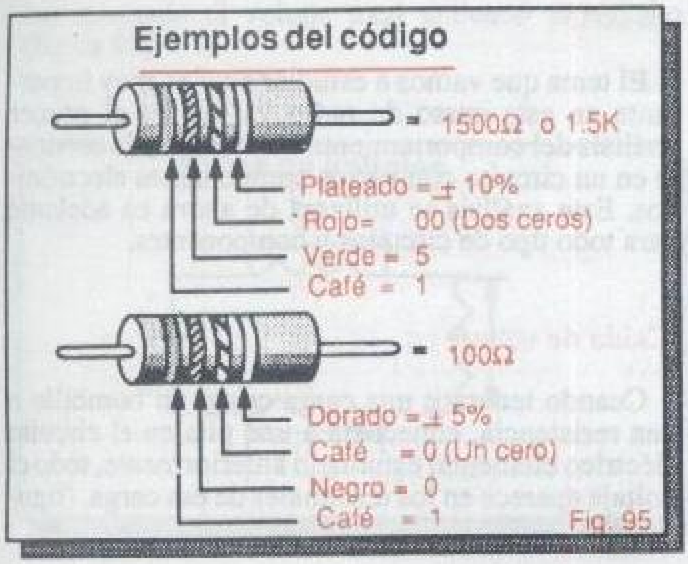

Fig. 95

que identificar una resistencia , vamos a perder mucho tiempo.

Después de algún tiempo de trabajar en electrónica, este código se hace tan familiar que ya se identifica una resistencia con sólo mirar brevemente su combinación de colores.

Tolerancia

Se ha mencionado que la cuarta banda indica la tolerancia de la resistencia. Esta tolerancia significa que el valor *real* no es necesariamente el mismo que indica el código. Un 10% de tolerancia significa que el valor real puede ser un 10% mayor o menor que el valor que indica el código.

Esta tolerancia se debe a la precisión del proceso de fabricación de esas resistencias,debido a que las máquinas depositan una capa ligeramente mayor o menor del compuesto resistivo.

Se fabrican resistencias con tolerancias del 20 %, 10 %, 5 % (que son las más comunes), 2 %, 1%, 0.5 %, 0.1 % y más.

El valor de las resistencias aumenta considerablemente a medida que su precisión aumenta,o sea que la tolerancia es menor. Debemos utilizar por lo tanto las resistencias más económicas posibles,pero que no alteren la operación del circuito.

Para una resistencia de 10.000 ohmios con una tolerancia del 5 %,se puede tener cualquier valor real entre 9.500 y 10.500 ohmios.

Muchos ceros

Hay otra cosa importante. A menudo encontramos que no es fácil escribir todos los ceros que aparecen en un número. Escribir 220.000 ohmios ó 10.000.000 ohmios puede ser difícil o gasta demasiado espacio.

Para resolver el problema, se ha creado un sistema de abreviación que utiliza letras para indicar Miles y Millones.

K = Kilo, que significa mil, o tres ceros (000).

M = Mega, que significa millón o seis ceros (000000)

Ejemplos

En vez de escribir 22.000 ohmios escribimos Kohmios o 22KΩ.

En vez de escribir 11.000.000 ohmios escribimos 11MΩ.

47KΩ sería 47.000 Ω (47 Kiloohmios).

10MΩ sería 10.000.000Ω (10 Megaohmios).

¿Cómo descifraría una resistencia de 4.7KΩ ?

Es sencillo, corremos el punto decimal tres puestos quedando 4.700Ω.

Una resistencia de 3.3MΩ sería entonces de 3.300.000Ω

Valores preferidos o normalizados para las resistencias

Las resistencias de carbón se fabrican en ciertos valores llamados *preferidos* o normalizados. Esto se debe a que sería imposible tener resistencias en todos los valores posibles, y no se justifica en la mayoría de los circuitos electrónicos.

Los valores preferidos son 1, 1.1, 1.2, 1.3, 1.5, 1.6, 1.8, 2, 2.2, 2.4, 2.7, 3, 3.3, 3.6, 3,9, 4.3, 4.7, 5.1, 5.6, 6.2, 6.8, 7.5, 8.2 y 9.1 y en todos sus múltiplos.

Estos valores son los que tienen las resistencias o resistores que se encuentran en el mercado, en los almacenes o distribuidores especializados y que se utilizan para toda clase de circuitos electrónicos.

Así tenemos resistencias de 1 KΩ, 10 KΩ, 430 KΩ, 82 KΩ, 33 Ω, etc. Para conseguir valores que no se encuentran en esta tabla, se pueden conectar resistencias en *serie* o en *paralelo*.

Las resistencias o resistores se pueden conectar en serie o en paralelo con el fin de obtener valores preferidos para reemplazar resistencias en un circuito que no tengamos a la mano,o también para lograr resistencias de mayor vatiaje.

Resistencias en serie

En una conexión en *serie* se conectan dos o más resistencias,uniendo un terminal de una de ellas con un terminal de la siguiente,formando una cadena. Este conjunto de resistencias forman una nueva resistencia con características diferentes, llamada *resistencia equivalente* (figura 96).

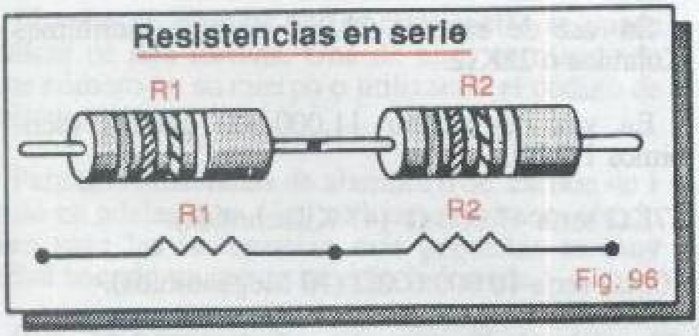

Fig. 96

El valor de la resistencia equivalente para resistencias en serie,es igual a la *suma* de todas las resistencias que intervienen en la cadena.

Req o Rt = R1 + R2 + R3 + . . + Rn

Como ejemplo para entender mejor cómo se determina el valor de la resistencia equivalente de una conexión o circuito de resistencias en serie, estudiemos el siguiente caso:

Tenemos un circuito con tres resistencias conectadas en serie con valores de 2.2 KΩ, 33 KΩ y 100 Ω¿Cuál será su resistencia equivalente?.

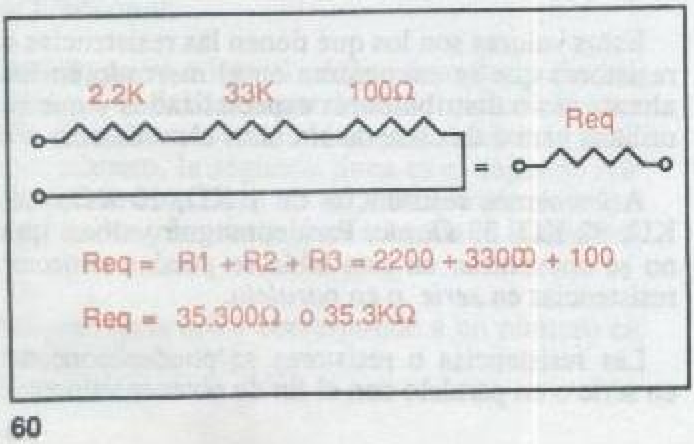

Req = R1 + R2 + R3 = 2200 + 33000 + 100

Req = 35.300Ω o 35.3KΩ

Corriente y voltaje en un circuito de resistencias en serie

El tema que vamos a estudiar aquí es muy importante en este curso de radio, ya que es el primer análisis del comportamiento del voltaje y la corriente en un circuito con varios componentes electrónicos. Este análisis se utilizará de ahora en adelante para todo tipo de circuitos y componentes.

Caída de voltaje

Cuando tenemos una carga como un bombillo o una resistencia, conectada a una pila en el circuito eléctrico elemental estudiado anteriormente, todo el voltaje aparece en los terminales de esa carga. (figura 97).

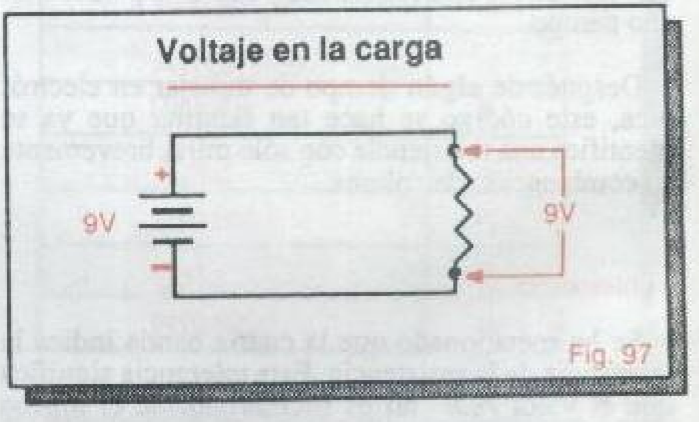

Fig. 97

Pero si tenemos dos o más resistencias en serie en un circuito electrónico sencillo,ocurre un fenómeno llamado *caída de voltaje* . Esta caída de voltaje nos indica que en cada resistencia se aplica una parte del voltaje total del circuito,dependiendo de su valor en ohmios. (figura 98).

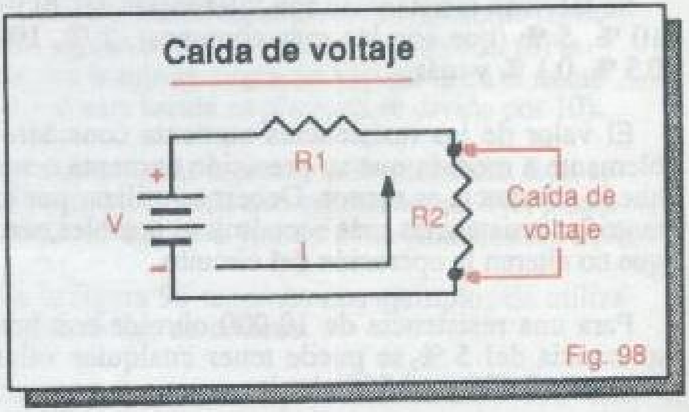

Fig. 98

Este principio se complementa con una ley que dice que si se suman todos los voltajes de los elementos de un circuito en serie siempre debe dar como resultado el voltaje total aplicado al circuito. (figura 99).

Suma de voltajes

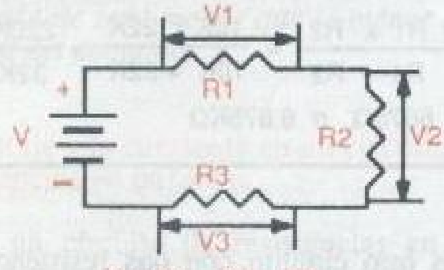

$V = V1 + V2 + V3$

Fig 99

Otro principio fundamental y muy importante dice que : *En todo circuito en serie la corriente que circula es la misma en todos los elementos de ese circuito*.

Para calcular la caída de voltaje en cada elemento del circuito, se utiliza la ley de Ohm, multiplicando el valor de la corriente que circula, por el valor en ohmios de cada uno de los componentes; en este caso, de las resistencias.

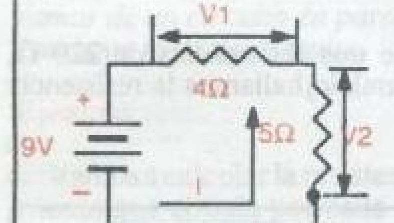

$$Req = R1 + R2$$

$$Req = 4\Omega + 5\Omega$$

$$I = \frac{V}{Req} = \frac{9V}{9\Omega} = 1Amp$$

$$V1 = R1 \times I = 4\Omega \times 1Am = 4 \text{ Voltios}$$

$$V2 = R2 \times I = 5\Omega \times 1Am = 5 \text{ Voltios}$$

Divisores de voltaje

Utilizando el principio de la caída de voltaje se pueden formar circuitos llamados *divisores de voltaje*, que permiten obtener varios voltajes de una fuente de valor fijo. Por ejemplo, con una pila de 9 voltios podemos obtener un voltaje de 6 voltios por medio de dos resistencias. (figura 100).

Si utilizamos una resistencia variable o potenciómetro, podemos obtener un voltaje variable de una pila u otra fuente de voltaje de corriente continua. (figura 101).

Divisor de voltaje

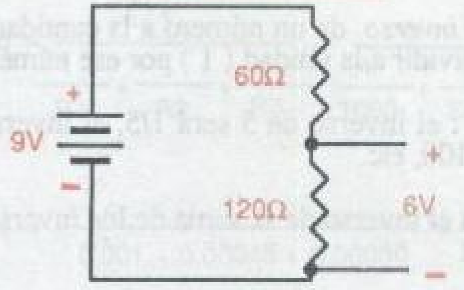

Fig. 100

Divisor variable de voltaje

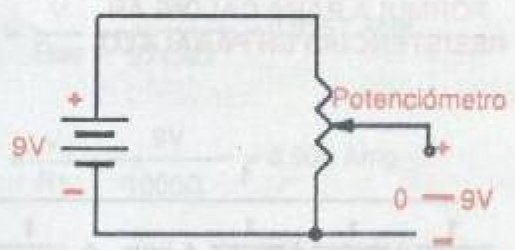

Fig. 101

Estos principios se utilizan mucho en el estudio y análisis de los diferentes circuitos de radio, que estaremos estudiando en próximas lecciones, y por lo tanto los debemos tener muy en cuenta.

Resistencias en paralelo

En la conexión de resistencias en paralelo se conectan todos los terminales de un lado de las resistencias a un punto común, y todos los terminales del otro lado a otro punto común. (figura 102).

Resistencias en paralelo

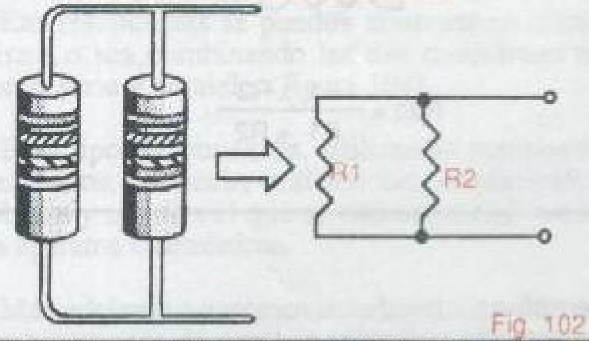

Fig. 102

El valor de la resistencia equivalente para un conjunto de resistencias en paralelo, no es tan fácil de obtener como en las resistencias en serie. *Su va-*

lor es igual al inverso de la suma de los inversos de cada resistencia del conjunto.

Se llama *inverso* de un número a la cantidad que resulta de dividir a la unidad (1) por ese número.

Ejemplo : el inverso de 5 será 1/5, el inverso de 100 será 1/100, etc.

Entonces el inverso de la suma de los inversos se escribe así :

FORMULA PARA CALCULAR RESISTENCIAS EN PARALELO

$$Req = \frac{1}{\frac{1}{R1} + \frac{1}{R2} + \frac{1}{R3} + \cdots + \frac{1}{Rn}}$$

Cuando tenemos solamente dos resistencias en paralelo, la fórmula se puede simplificar de la siguiente manera :

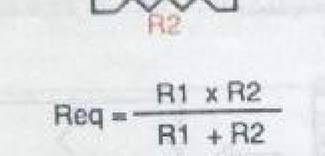

$$Req = \frac{R1 \times R2}{R1 + R2}$$

Ejemplos :

a) Tenemos un circuito con dos resistencias en paralelo de 10 KΩ y 22 KΩ, y encontraremos el valor de la resistencia equivalente o total. Aplicando la fórmula tenemos :

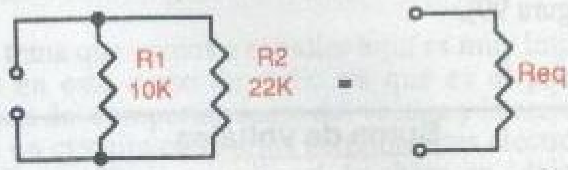

$$Req = \frac{R1 \times R2}{R1 + R2} = \frac{10K \times 22K}{10K + 22K} = \frac{220K^2}{32K}$$

$$Req = 6875\Omega \quad o \quad 6{,}875K\Omega$$

b) Tenemos otro circuito con dos resistencias en paralelo, de 1 MΩ cada una. Hallamos el valor de la resistencia equivalente :

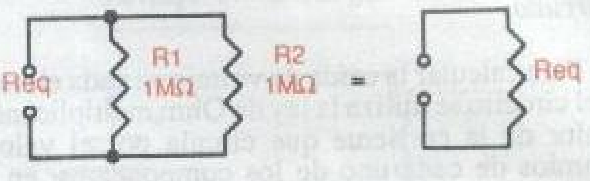

$$Req = \frac{1M\Omega}{2} = 500.000\Omega \quad o \quad 500K\Omega$$

c) En otro circuito de tres resistencias de 220 Ω, 470 Ω y 6.8 KΩ en paralelo, hallamos la resistencia equivalente :

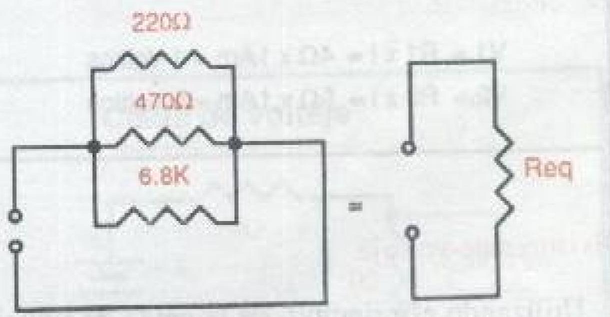

$$Req = \frac{1}{\frac{1}{R1} + \frac{1}{R2} + \frac{1}{R3}} = \frac{1}{\frac{1}{220} + \frac{1}{470} + \frac{1}{6800}}$$

$$Req = \frac{1}{0.0045 + 0.0021 + 0.00014} = \frac{1}{0.00674}$$

$$Req = 148.36\Omega$$

Como conclusión debemos tener en cuenta que, *si conectamos dos resistencias iguales en paralelo, el valor de la resistencia equivalente será la mitad del valor de una de estas resistencias.*

Otro principio importante es el que dice que, *en un circuito de resistencias en paralelo, la resistencia equivalente será menor que la menor de las resistencias del conjunto.*

El voltaje y la corriente en un circuito de resistencias en paralelo.

En un circuito de resistencias en paralelo, el voltaje para cada una de ellas será igual y es el mismo de la fuente de alimentación del circuito.

Sin embargo la corriente se *divide* o distribuye por cada una de las resistencias o ramales, de acuerdo al valor en ohmios de cada resistencia.

Debemos observar entonces que en el circuito en paralelo, ocurre lo contrario que en el circuito en serie, en cuanto a voltaje y corriente se refiere.

Para calcular cómo se divide la corriente por cada uno de los circuitos, se utiliza la ley de Ohm, dividiendo el voltaje total, que es el mismo, por el valor de cada resistencia.

En este tipo de configuración se establece una ley que dice : *La suma de las corrientes de todas las ramas de un circuito en paralelo, debe ser igual a la corriente total del circuito.*

Ejemplo:

Vamos a calcular la resistencia equivalente y la corriente que circula por cada una de las resistencias del circuito de la figura 103.

División de la corriente

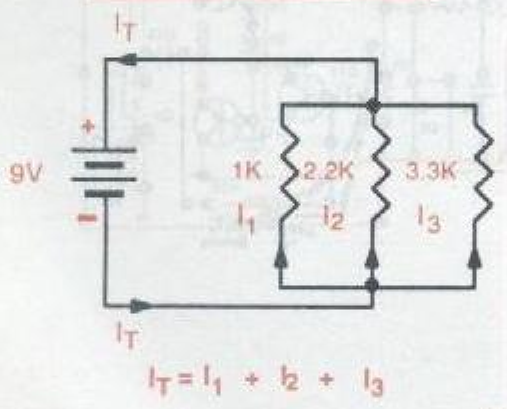

Fig. 103

$$Req = \frac{1}{\frac{1}{R1} + \frac{1}{R2} + \frac{1}{R3}} = \frac{1}{\frac{1}{1000} + \frac{1}{2200} + \frac{1}{3300}}$$

$$Req = \frac{1}{0.001 + 0.00045 + 0.00030} = \frac{1}{0.00175}$$

$$Req = 571.4\Omega$$

$$I_T = \frac{V}{Req} = \frac{9V}{571.4\Omega} = 0.0157 \text{ Amp}$$

$$I_1 = \frac{V}{R1} = \frac{9V}{1000\Omega} = 0.009 \text{ Amp}$$

$$I_2 = \frac{V}{R2} = \frac{9V}{2200\Omega} = 0.0040 \text{ Amp}$$

$$I_3 = \frac{V}{R3} = \frac{9V}{3300\Omega} = 0.0027 \text{ Amp}$$

$$I_T = I_1 + I_2 + I_3 = 0.009 + 0.004 + 0.0027$$

$$I_T = 0.0157 \text{ Amp}$$

Resistencias en serie - paralelo

Las resistencias se pueden conectar en circuitos *mixtos*, o sea, combinando las dos conexiones anteriores, serie y paralelo. (figura 104).

Este tipo de conexión, utilizando componentes mezclados, es decir, resistencias, condensadores, bobinas y otros, es el que se encuentra realmente en los aparatos electrónicos.

Más adelante estaremos estudiando las diferentes combinaciones de componentes, llamadas circuitos RC, RL, LC y RLC.

La letra R indica resistencia, la letra C indica condensador y la letra L indica bobina.

Actividad práctica Nº 5

En esta actividad vamos a instalar en el tablero del radio las resistencias correspondientes al circuito amplificador de audio o,sea la etapa marcada con el color azul en su lámina del diagrama del radio.

Para hacerlo, siga cuidadosamente las instrucciones siguientes,utilizando como ayuda la lámina " Guía para la instalación de componentes". Recuerde las instrucciones para soldar correctamente, entregadas en las prácticas anteriores.

Componentes y herramientas necesarios

1 Resistencia de 680 Ω (Azul-Gris-Café)
2 Resistencias de 0.5 Ω (Verde-Negro-Plateado)
1 Resistencia de 150 KΩ (Café-Verde-Amarillo)
1 Resistencia de 560 Ω (Verde-Azul-Café)

Todas las resistencias de 1/4 de vatio y 5% de tolerancia, por lo tanto la cuarta banda debe ser dorada.

1 Cautín o soldador de 25 vatios
1 Pinzas de punta
1 Cortador de alambre o cortafrío
1 Pedazo de soldadura

Paso 1: Localice la resistencia R13 de 680 ohmios (Azul-Gris-Café-Dorado) e instálela entre las puntillas A9 y A25, en la parte superior derecha del radio. Suelde primero un terminal en la puntilla A9 fijándoseque la resistencia que de sentrada entre las dos puntillas. Luégo suelde el otro terminal en la puntilla A25, fijándose que la resistencia que de bien recta. Y recorte los terminales sobrantes.

Paso 2 : Localice las dos resistencias R14 y R15 de 0.5 ohmios (Verde-Negro-Plateado-Dorado). Instale una entre las puntillas A19 y A20 y la otra entre las puntillas A20 y A21. Colóquelas en el centro de las puntillas, suelde bien sus terminales y recorte los alambres sobrantes.

Paso 3 : Localice la resistencia R16 de 150 KΩ (Café-Verde-Amarillo-Dorado) e instálela entre las puntillas A7 y A17, soldando bien sus dos terminales.

Paso 4 : Localice la resistencia R12 de 560 Ω (Verde-Azul-Café-Dorado), e instálela entre las puntillas 39 y 38 .

Después de pegar estas cinco resistencias su radio debe quedar como se muestra en la figura A7.

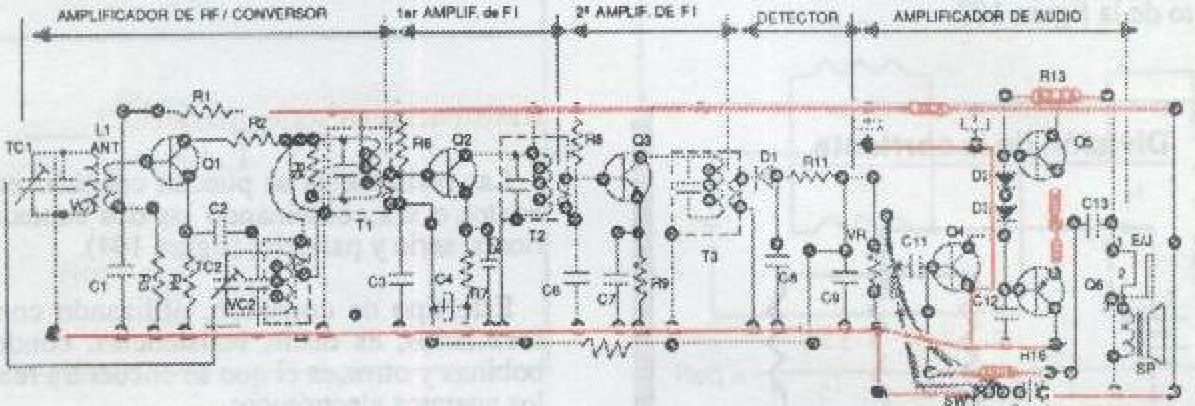
AMPLIFICADOR DE RF / CONVERSOR
1er AMPLIF. de F I
2ª AMPLIF. DE F I
DETECTOR
AMPLIFICADOR DE AUDIO
R1
R2
L1
ANT
TC1
Q1
VC1
C2
C1
TC2
VC2
T1
R6
Q2
C3
C4
R7
T2
R8
Q3
C6
C7
R9
T3
D1
R11
C8
C9
VR
C11
Q4
C12
R13
D2
D3
Q5
C13
Q6
EJ
R16
SP
SW1
Bateria

Fig. A7

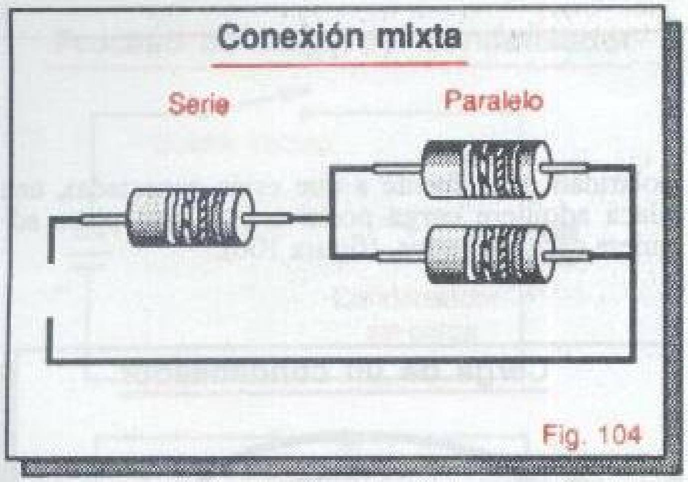
Serie
Paralelo

Conexión mixta

Fig. 104

Para aplicar la ley de Ohm a los circuitos mixtos o complejos como el de la figura 104 debemos primero encontrar la resistencia equivalente de todo el circuito. Para hacerlo, debemos ir resolviendo cada grupo hasta llevarlo a un circuito más simple y al final poder calcular fácilmente el valor de esa resistencia.

Con el fin de aclarar esta afirmación, vamos a realizar un ejemplo con un circuito real. Supongamos que tenemos un circuito con diferente combinación de resistencias en serie y paralelo y se debe hallar su resistencia equivalente.

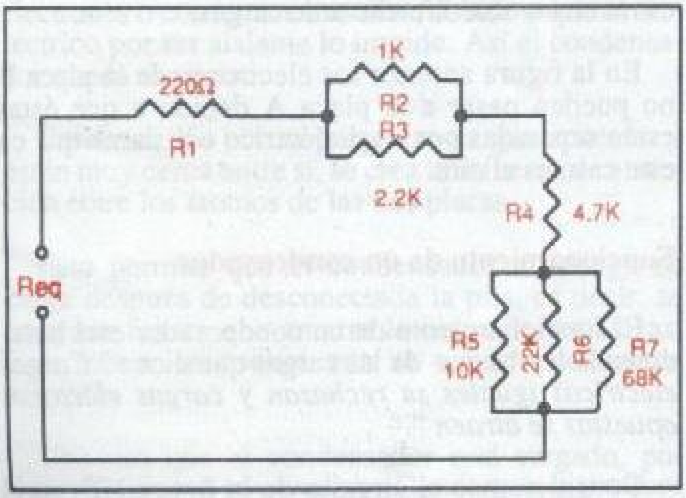
1K
220Ω
R2
R3
R1
2.2K
R4
4.7K
Req
R5
10K
22K
R6
R7
68K

Solución:

Para hallar la resistencia equivalente en el circuito debemos resolver primero los dos grupos de resistencias en paralelo.

Después reemplazando estos dos valores en el circuito quedará como muestra la gráfica.

Y aplicando la fórmula para resistencias en serie se obtiene el resultado.

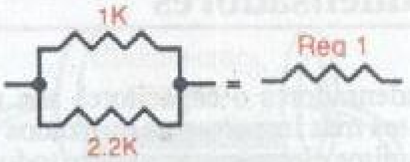

$$\text{Req 1} = \frac{R2 \times R3}{R2 + R3} = \frac{1K \times 2.2K}{1K + 2.2K} = \frac{2.2K^2}{3.2K}$$

Req 1 = 0.6875K o 687.5Ω

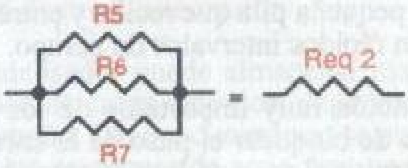
R5
R6
R7
Req 2

$$\text{Req 2} = \frac{1}{\frac{1}{10K} + \frac{1}{22K} + \frac{1}{68K}}$$

$$\text{Req 2} = \frac{1}{0.0001 + 0.000045 + 0.000014} = \frac{1}{0.000159}$$

Req 2 = 6.28K

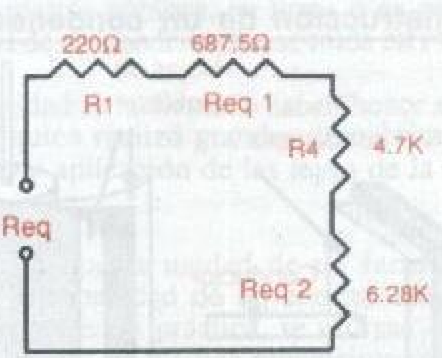
220Ω
687.5Ω
R1
Req 1
R4
4.7K
Req
Req 2
6.28K

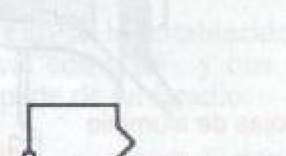

Req = 220Ω + 687,5Ω + 4700Ω + 6.280Ω

Req = 11887 Ω o 11.88KΩ

Los condensadores

Los condensadores o capacitores son uno de los componentes más importantes utilizados en casi todos los circuitos electrónicos y sobre todo en los circuitos de radio.

Sus aplicaciones son muy variadas y tienen como función principal almacenar energía temporalmente.

Prácticamente puede considerarse al condensador como una pequeña pila que recibe y entrega energía eléctrica en rápidos intervalos de tiempo.

Otra función muy importante de los condensadores es la de bloquear el paso de la corriente continua y permitir el paso de la corriente alterna de un circuito a otro.

Construcción de los condensadores

Un condensador está compuesto por dos superficies conductoras o placas metálicas, separadas por un material aislante llamado dieléctrico. (figura 105).

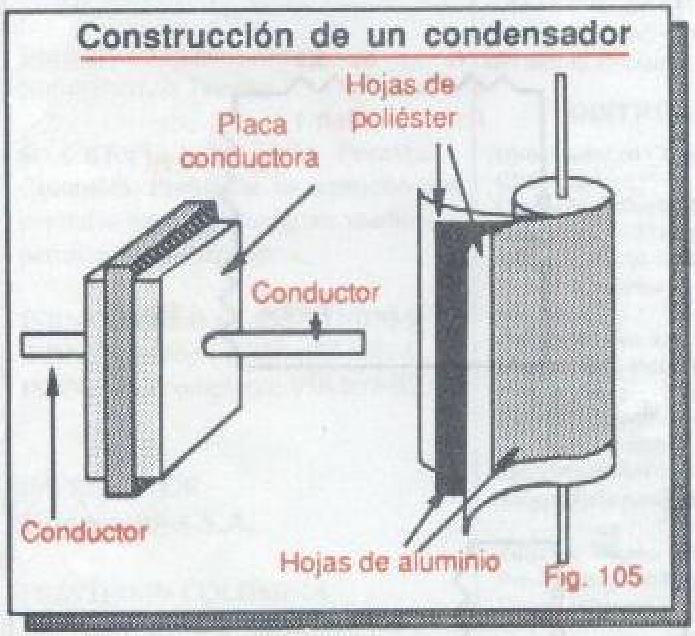
Construcción de un condensador

Fig. 105

El dieléctrico de un condensador puede ser de diferentes materiales. Entre los más utilizados tenemos el papel, la cerámica, el poliéster, la mica, el papel impregnado en aceite, el aire, etc.

Las placas de los condensadores tienen la función de almacenar los electrones. Dependiendo de la polaridad de la fuente a que estén conectadas, una placa adquiere carga positiva y la otra placa adquiere carga negativa. (figura 106).

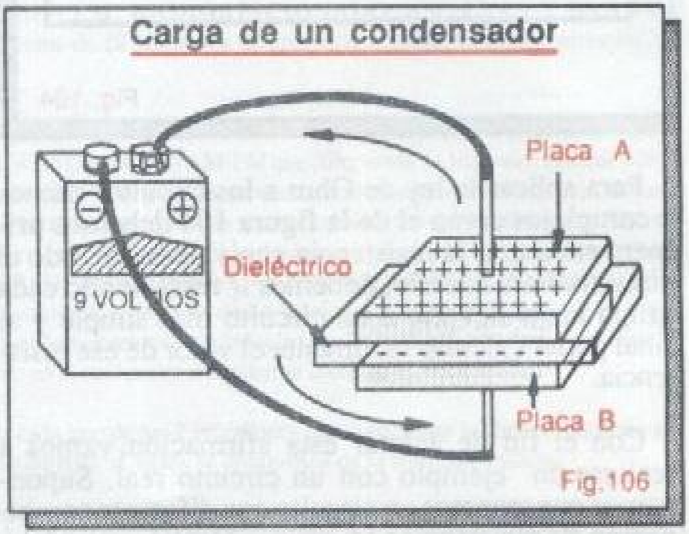
Carga de un condensador

Fig.106

Esta condición crea una diferencia de cargas entre las dos placas. La función del dieléctrico o aislante es mantener esa diferencia de cargas.

En la figura anterior los electrones de la placa B no pueden pasar a la placa A debido a que éstas están separadas por un dieléctrico o aislante que en este caso es el aire.

Funcionamiento de un condensador

El funcionamiento de un condensador está basado en la ley básica de las cargas que dice : "*Cargas eléctricas iguales se rechazan y cargas eléctricas opuestas se atraen*".

Consideremos el circuito de la figura 107, en el cual el suiche está abierto y no permite que fluya corriente hacia el condensador.

En estas condiciones, el condensador permanece inactivo por encontrarse aislado de la fuente y por lo tanto no se crea diferencia de carga entre sus placas. Veamos lo que sucede al cerrarse el suiche. (figura 107).

Al cerrarse el suiche, el exceso de electrones del borne negativo de la pila circula a través del hilo conductor depositándose en la placa A, cargándola negativamente.

Proceso de carga del condensador

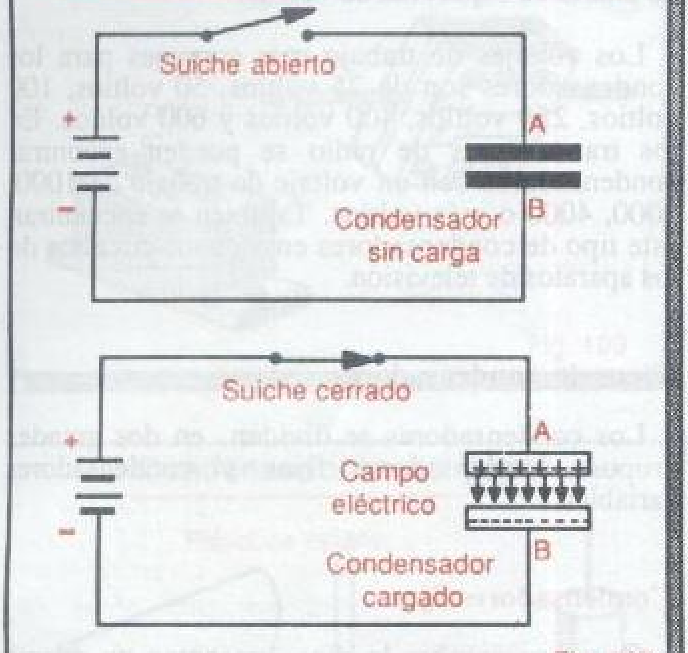

Fig. 107

El borne positivo de la fuente atrae electrones de la otra placa (B) dejándola con un faltante de electrones, en otras palabras, cargándola positivamente. Durante el proceso anterior se establece una diferencia de cargas entre las dos placas sin existir flujo de electrones o corriente entre ellas, debido a que el dieléctrico por ser aislante lo impide. Así el condensador queda cargado.

Como las dos placas presentan cargas opuestas y están muy cerca entre sí, se crea una fuerza de atracción entre los átomos de las dos placas.

Esto permite que el condensador mantenga su carga después de desconectada la pila, es decir, se crea una fuerza que amarra la carga de las placas. Esta fuerza se conoce con el nombre de *campo eléctrico*.

Una vez que el condensador está cargado, podríamos unir sus terminales, notando que se produce una pequeña chispa al hacerlo. (figura 108).

El condensador cargado actúa como una pequeña pila.

Los electrones sobrantes de la placa negativa se pasan a la placa positiva por medio de los conductores, estableciendo una corriente eléctrica.

Casi inmediatamente las dos placas quedan con igual cantidad de electrones libres, desapareciendo la diferencia de cargas con lo cual el condensador se considera descargado.

Descarga de un condensador

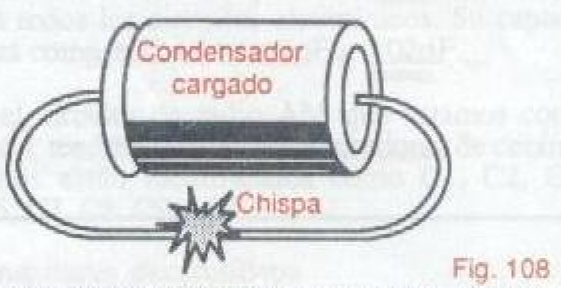

Fig. 108

Capacidad de un condensador

Un condensador puede almacenar mayor o menor cantidad de carga eléctrica, dependiendo de su forma y tamaño. Los condensadores se pueden comparar con los estanques de agua, hay unos con mayor capacidad de almacenamiento que otros.

Unidad de medida de la capacidad de los condensadores

A la cantidad de carga eléctrica que puede almacenar un condensador se le conoce como *Capacidad* y se mide por medio de una unidad llamada *Faradio*.

Al igual que la capacidad de un tanque de agua se mide en metros cúbicos, en litros o en galones, la capacidad de un condensador se mide en *Faradios*.

Esta unidad se ha llamado así en honor a Michael Faraday, quien realizó grandes descubrimientos en el estudio y aplicación de las leyes de la electricidad.

Debido a que la unidad de un faradio corresponde a la capacidad de un enorme condensador que no existe en la práctica, se utilizan frecuentemente unidades de medida muchísimo más pequeñas.

Para esto se ha establecido el *microfaradio*, que se abrevia como *µF* y que equivale a una millonésima parte de un faradio.

También se utilizan el *nanofaradio* que se ecribe como *nF* y el *picofaradio* que se escribe como *pF* y que equivale a una millonésima parte de un microfaradio.

1 Faradio = 1.000.000 µF
= 1.000.000.000.000 pF

$$1\mu F = \frac{1}{1000000}\ \text{Faradios} = 0.000001F = 1 \times 10^{-6}F$$

$$1pF = \frac{1}{1000000}\ \mu F = 0.000001\ \mu F = 1 \times 10^{-6}\ \mu F$$

$$1pF = \frac{1}{1000000000000}\ \text{Faradios} = 0.000000000001F = 1 \times 10^{12}$$

Es muy común encontrar en los aparatos electrónicos condensadores de 100 picofaradios, 10 microfaradios, 2200 microfaradios y que se expresan como 100 pF, 10 µF, 2200 µF, etc.

Factores que determinan la capacidad de un condensador

Existe una gran variedad de condensadores de diferentes valores, desde un picofaradio, hasta varios miles de microfaradios.

Hay dos factores físicos que determinan la capacidad de almacenamiento de electrones de un condensador. Estos dos aspectos: son el tamaño o superficie de las placas y la separación entre ellas o sea el espesor del dieléctrico.

A mayor tamaño o área de las placas, o a menor separación entre ellas, se obtiene una mayor capacidad en el condensador. Esta separación tiene cierto límite, ya que si se tienen las placas muy cerca entre sí puede saltar el voltaje de una placa a la otra quemando el condensador.

Por esta razón los condensadores de mayor capacidad son más grandes que otros de menor capacidad.

El voltaje de trabajo

Otro aspecto muy importante en los condensadores es el máximo voltaje que soportan conectados a un circuito. A esta característica se le llama *voltaje de trabajo.*

Dos condensadores pueden tener la misma capacidad en microfaradios, pero soportar diferentes voltajes. El voltaje de trabajo de un condensador está relacionado con el material del diélectrico utilizado y con la separación entre sus placas.

Un condensador cuyas placas están más cerca entre, sí soporta menor voltaje, rompiéndose con mayor facilidad el aislamiento del dieléctrico. Por esta razón los condensadores de mayor voltaje son más voluminosos, ya que requieren que sus placas estén más separadas y su superfice sea mayor para no alterar su capacidad de voltaje.

Los voltajes de trabajo más comunes para los condensadores son de 25 voltios, 50 voltios, 100 voltios, 250 voltios, 400 voltios y 600 voltios. En los transmisores de radio se pueden encontrar condensadores con un voltaje de trabajo de 1000, 2000, 4000 o más voltios. También se encuentran este tipo de condensadores en algunos circuitos de los aparatos de televisión.

Tipos de condensadores

Los condensadores se dividen en dos grandes grupos: condensadores fijos y condensadores variables.

Condensadores fijos

Como su nombre lo dice, presentan un mismo valor de capacidad. Hay diferentes tipos de condensadores fijos, clasificados principalmente por el tipo de dieléctrico utilizado en su fabricación.

Según el tipo de dieléctrico toman diferentes formas como veremos más adelante. Los principales tipos de condensadores fijos son: de mica, de cerámica, de papel y los electrolíticos.

Los condensadores de cerámica y los electrolíticos son los más utilizados en los actuales circuitos electrónicos y de radio, pero existen otros tipos de condensadores que se utilizan mucho, sobre todo en los aparatos de radio y televisión fabricados con tubos.

Condensadores de mica

Están formados por láminas metálicas delgadas, separadas entre sí por láminas de mica. Todo el conjunto está contenido dentro de una envoltura de plástico. Los condensadores de mica presentan bajas pérdidas de energía, soportan altos voltajes, son compactos y de buena precisión.(figura 109).

Condensadores de papel

Los condensadores de papel son baratos y su construcción es muy sencilla, lo cual ha hecho muy popular su empleo. Consiste en tiras metálicas muy delgadas, separadas entre sí por bandas de papel aislante impregnado en cera como dieléctrico. Las placas y el papel del condensador vienen enrollados en forma cilíndrica, con el fin de obtener una mayor área de las placas, logrando así una capacidad utilizable. (figura 110).

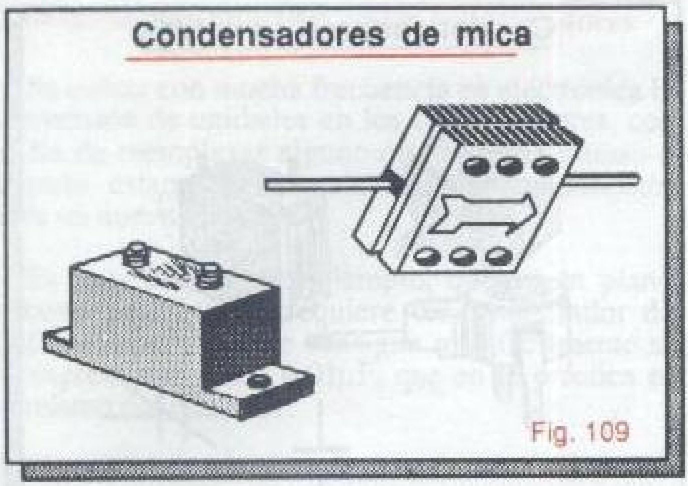
Condensadores de mica

Fig. 109

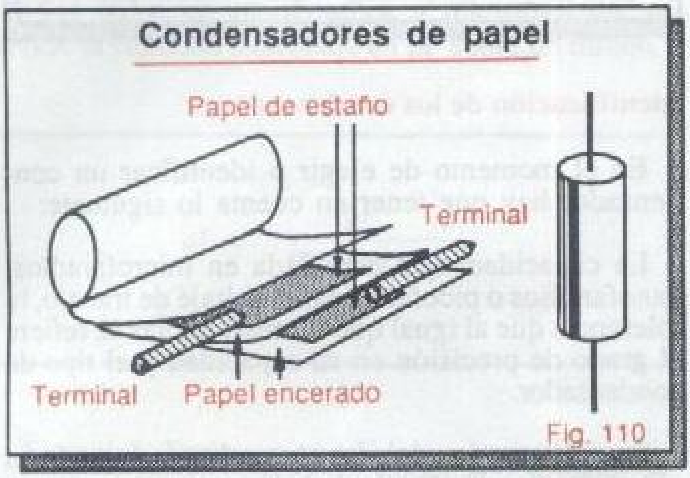
Condensadores de papel
Papel de estaño
Terminal
Terminal
Papel encerado

Fig. 110

Los voltajes de trabajo con que se fabrican estos condensadores son de 200, 400 y 600 voltios. Para voltajes mayores es necesario utilizar condensadores de papel con baño de aceite.

Condensadores de cerámica

Los condensadores de cerámica, llamados también condensadores de disco, son de tamaño reducido, baratos y versátiles. Estos condensadores están constituidos por dos delgadas placas o películas de plata en forma de disco y tienen cerámica como material dieléctrico. (figura 111).

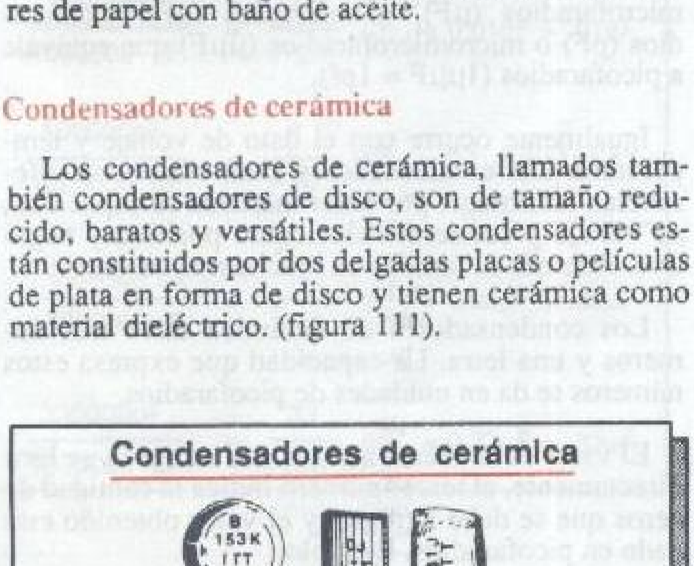
Condensadores de cerámica
B
153K
TTT
04

Fig. 111

Las formas más usuales de los condensadores cerámicos son la de disco y la tubular. Son condensadores relativamente pequeños y se utilizan mucho en todos los circuitos electrónicos. Su capacidad está comprendida entre 1pF y 0.02μF.

En el circuito de radio AM que estamos construyendo, tenemos varios condensadores de cerámica. Estos están identificados como C1, C2, C3, C5, C6, C7, C8, C9 y C12.

Condensadores electrolíticos

Los condensadores electrolíticos, conocidos también como condensadores polarizados, constan de una placa positiva y otra placa negativa, hechas generalmente de aluminio o tantalio poroso. La placa positiva está sumergida en un líquido, que hace las veces de electrolito en el caso de los condensadores electrolíticos húmedos. En el caso de los condensadores electrolíticos secos, las placas están en contacto con una gasa impregnada de electrólito.

La principal ventaja de los condensadores electrolíticos es la de presentar una capacidad muy alta para un tamaño relativamente pequeño, sin exceder su costo. Su capacidad puede alcanzar el orden de los miles de microfaradios.

En la figura 112 podemos observar el aspecto físico de este tipo de condensador, que es muy común también en todos los circuitos electrónicos.

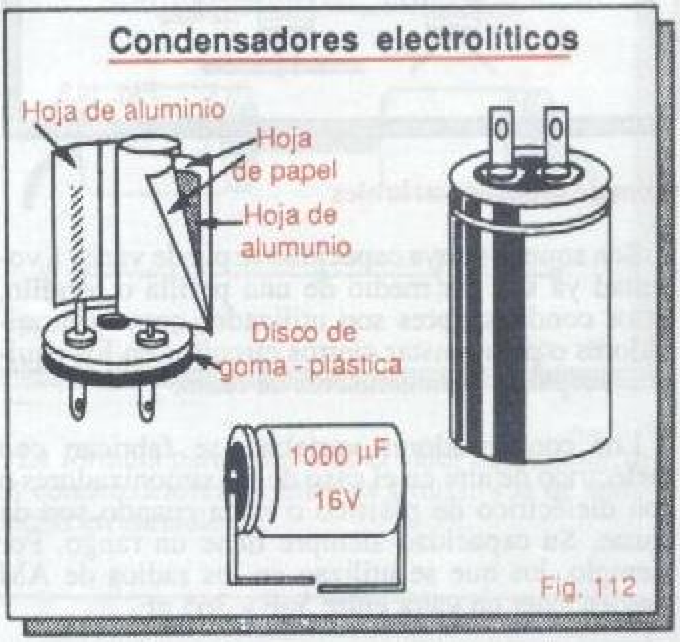
Condensadores electrolíticos
Hoja de aluminio
Hoja de papel
Hoja de alumunio
Disco de goma - plástica
1000 μF
16V

Fig. 112

Los valores más usuales son: 1μF, 10μF, 22μF, 100μF, 1000μF, 2200μF, 6800μF, 4700μF y 10000μF.

Debido a que los condensadores electrolíticos presentan polaridad, hay que ser cuidadoso al conectarlos, ya que una polaridad invertida lo puede destruir explotando violentamente, por tal razón estos condensadores sólo se utilizan en circuitos de corriente directa o continua.

Para conectar un condensador electrolítico deben conocerse los terminales positivo y negativo, tanto en el condensador como en el circuito.

La principal aplicación de los condensadores electrolíticos está en los circuitos de fuente de poder.

Podemos observar en el diagrama del radio que tenemos varios condensadores electrolíticos que son: C4, C10, C11, C13 Y C14.

Los condensadores electrolíticos se fabrican en dos formas: axial y radial. Los de tipo axial tienen los terminales en cada lado del condensador y se montan acostados. Los de tipo radial tienen los dos terminales por el mismo lado y se montan parados. (figura 113).

Montaje de condensadores electrolíticos

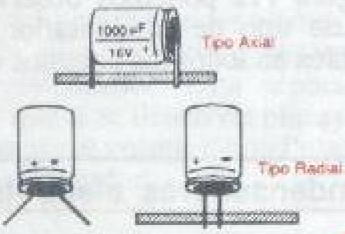

Fig. 113

Condensadores variables

Son aquellos cuya capacidad se puede variar a voluntad ya sea por medio de una perilla o tornillo. Estos condensadores son utilizados como sintonizadores o para ajustar ciertos circuitos en los equipos receptores y transmisores de radio.

Los condensadores variables se fabrican con dieléctrico de aire en el caso de los sintonizadores o con dieléctrico de plástico o mica cuando son de ajuste. Su capacidad siempre tiene un rango. Por ejemplo, los que se utilizan en los radios de AM pueden tener un valor entre 5pF y 365 pF.

En la figura 114 podemos observar los diferentes tipos de condensadores variables más utilizados en los circuitos de radio y los símbolos utilizados para representarlos en los diagramas.

Condensadores variables

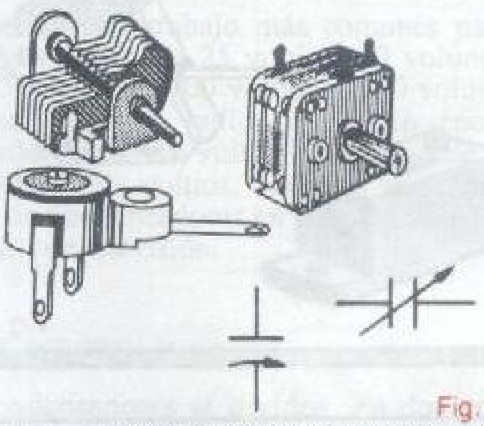

Fig. 114

Identificación de los condensadores

En el momento de elegir o identificar un condensador hay que tener en cuenta lo siguiente:

La capacidad, que está dada en microfaradios, nanofaradios o picofaradios, el voltaje de trabajo, la tolerancia que al igual que las resistencias se refiere al grado de precisión en su capacidad y el tipo de condensador.

Por último, en algunos tipos de condensadores electrolíticos viene estipulada la temperatura de trabajo y que se refiere al calor que éstos soportan.

Para su identificación, la mayoría de los condensadores traen directamente impresos sus valores en microfaradios (μF), nanofaradios (nF), picofaradios (pF) o micromicrofaradios (μμF) que equivale a picofaradios (1μμF = 1pf).

Igualmente ocurre con el dato de voltaje y temperatura. Los condensadores electrolíticos se diferencian fácilmente por tener marcada su capacidad, su voltaje y sus terminales como positivo y negativo. Además su tamaño es mucho mayor.

Los condensadores de cerámica traen tres números y una letra. La capacidad que expresa estos números se da en unidades de picofaradios.

El valor debe leerse así: los dos números se leen directamente, el tercer número indica la cantidad de ceros que se debe agregar. y el valor obtenido está dado en picofaradios. Ejemplo:

103 Z = 10 seguido de tres ceros, en pF = 10000pF

204 Z = 20 mas cuatro ceros, en pF = 200000pF

333 Z = 33 seguido de tres ceros, en pF = 33000pf

Conversión de unidades en los condensadores

Se utiliza con mucha frecuencia en electrónica la conversión de unidades en los condensadores, con el fin de reemplazar alguno que está defectuoso o cuando estamos consiguiendo los componentes para un nuevo circuito.

Es muy común, por ejemplo, que en un plano encontremos que se requiere un condensador de 10.000pF, pero éste se consigue más fácilmente si lo expresamos como 0.01µF, que en la práctica es el mismo condensador.

Conversión de picofaradios a microfaradios

Para pasar de picofaradios a microfaradios se divide la cantidad expresada en pF sobre un millón.

$$\mu F = \frac{pF}{1000000\ \mu F}$$

Ejemplo: Expresar en µF los valores de los siguientes condensadores: 10000pF, 200000pF, 20000pF y 330000pF.

$$\frac{10000pF}{1000000\ \mu F} = \frac{1}{100}\ \mu F = 0.01\mu F = .01\mu F$$

$$\frac{100000pF}{1000000\ \mu F} = \frac{1}{10}\ \mu F = 0.1\mu F = .1\mu F$$

$$\frac{20000pF}{1000000\ \mu F} = \frac{2}{100}\ \mu F = 0.02\mu F = .02\mu F$$

$$\frac{330000pF}{1000000\ \mu F} = \frac{33}{100}\ \mu F = 0.33\mu F = .33\mu F$$

Conversión de microfaradios a picofaradios

Para pasar de microfaradios a picofaradios se multiplica por un millón la cantidad a expresar:

$$1\mu F = 1000000pF$$

Ejemplo:

Expresar en picofaradios y en código de tres números los condensadores siguientes:

0.1 µF y 0.001 µF

$$0.1\ \mu F = 0.1\ \mu F \times 1000000\ pF = 100000\ pF$$
$$= 100000 = \mathbf{104}$$

$$0.001\ \mu F = 0.001\ \mu F \times 1000000pF = 1000pF$$
$$= 1000 = \mathbf{102}$$

Conexión de condensadores

Los condensadores, al igual que las resistencias, se pueden conectar en serie, en paralelo o en forma mixta.

Conexión de condensadores en serie

Al conectar condensadores en serie, la capacidad total disminuye.

Esto equivale a aumentar el espesor del dieléctrico quedando un sólo condensador cuyas placas están más separadas. (figura 115).

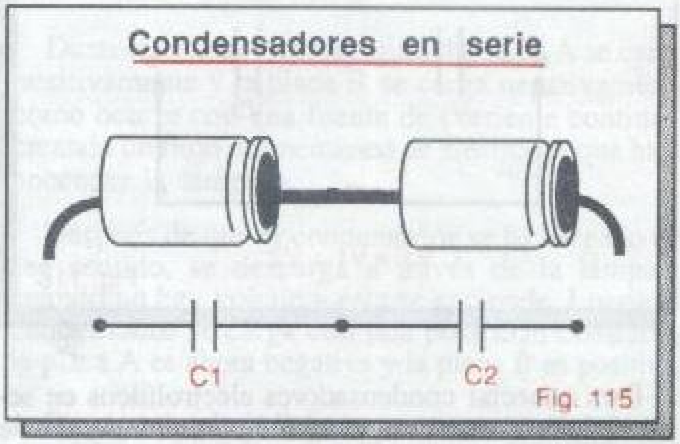

Fig. 115

La fórmula para encontrar el valor equivalente de los condensadores en serie es similar a la de resistencias en paralelo.

$$CT = \frac{1}{\frac{1}{C1} + \frac{1}{C2} + \frac{1}{C3} + \ldots\ldots}$$

Para dos condensadores diferentes:

$$CT = \frac{C1 \times C2}{C1 + C2}$$

Para condensadores iguales:

$$CT = \frac{C}{n}$$

CT = Capacidad Total
C = Capacidad de un condensador
n = Número de condensadores conectados

El máximo voltaje admitido por un grupo de condensadores conectados en serie, es igual a la suma de los voltajes de cada condensador. Por lo tanto al conectar condensadores en serie se aumenta el voltaje de trabajo pero se disminuye la capacidad. (figura 116).

Voltajes en condensadores en serie

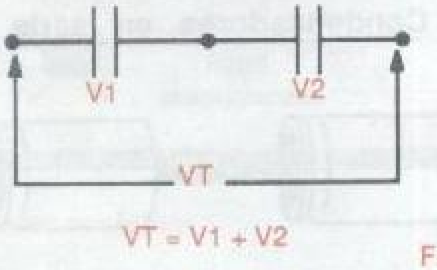

Fig. 116

Para conectar condensadores electrolíticos en serie se debe alternar su polaridad como en el caso de las pilas. (figura 117).

Condensadores electrolíticos en serie

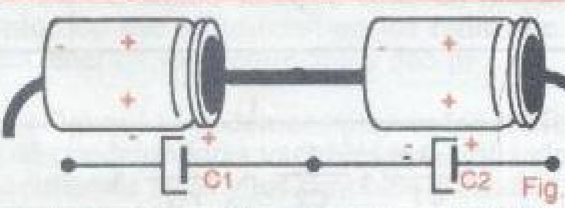

Fig. 117

Conexión de condensadores en paralelo

Cuando conectamos condensadores en paralelo, la superficie activa de las placas aumenta, obteniéndose un aumento de la capacidad. (figura 118).

Condensadores en paralelo

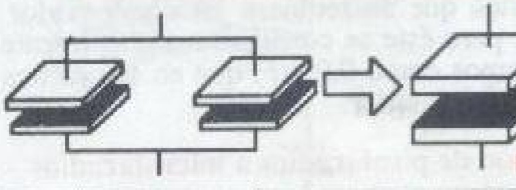

Fig. 118

Varios condensadores conectados en paralelo se pueden expresar como un solo condensador equivalente mediante la suma de la capacidad de cada uno:

$$C_T = C_1 + C_2 + C_3 + C_4 + \text{...........etc}$$

El voltaje aplicado al conjunto es el mismo para todos los condensadores y por lo tanto el máximo voltaje aplicado al circuito no debe ser superior al del condensador con menor voltaje de trabajo. Ejemplo:

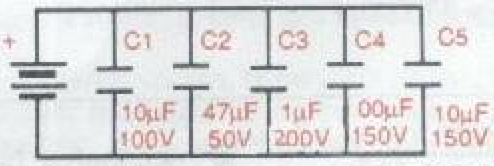

Hallar CT y voltaje máximo que se puede aplicar

SOLUCION:

CT = C1 + C2 + C3 + C4 + C5
CT = 10µF + 47µF + 1µF+ 100µF + 10 µF
CT = 168µF

Voltaje máximo permitido = Al menor voltaje de trabajo
Voltaje máximo = Vm = 50V

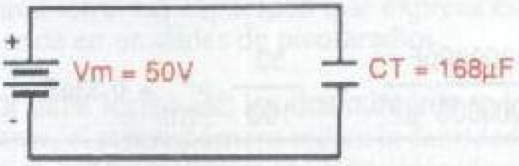

Observemos que la fórmula utilizada para condensadores en serie es la misma que para resistencias en paralelo, y la fórmula para condensadores en paralelo es la misma que para resistencias en serie.

Seguridad en el uso de condensadores

Cuando conectamos condensadores con voltajes superiores a 50 voltios y, sobre todo aquellos electrolíticos que vienen en los aparatos de sonido y televisión, debemos tener mucho cuidado de no tocar sus terminales. Estos condensadores permanecen cargados aún después de desconectarse el aparato.

Para evitar todo esto es preciso descargar los condensadores antes de manipular con estos circuitos. Una forma de hacerlo es poniendo momentáneamente en cortocircuito sus terminales.

Recomendaciones para elección de condensadores

Es recomendable que el voltaje de trabajo de un condensador sea de un 10% a 20% mayor, que el voltaje normal a que está conectado.

Hay que tener también en cuenta la temperatura de trabajo. En partes muy encerradas los condensadores se pueden sobrecalentar haciendo que éstos se sequen, derritan o se quemen. Se debe procurar que los condensadores, sobre todo los electrolíticos, no queden cerca de elementos que disipen mucho calor.

En este caso existen condensadores electrolíticos fabricados para diferentes temperaturas, por ejemplo: 85º C o 105º C.

Comportamiento de los condensadores con la corriente alterna (CA)

En los diferentes circuitos electrónicos de un radio, un televisor o un aparato de sonido de cualquier tipo, siempre encontramos corriente continua y corriente alterna trabajando juntas, lo que hace posible el funcionamiento total del aparato.

Los condensadores juegan un papel muy importante ante los dos tipos de corriente.

La acción de un condensador frente a la corriente continua sólo se manifiesta en los instantes en que se conecta y desconecta el voltaje, actuando después como un circuito abierto que no deja circular la corriente.

Al aplicar una corriente alterna (CA) a un condensador, éste actúa en forma muy diferente. Debido a que la corriente alterna invierte periódicamente su polaridad, las placas se hacen positivas y negativas alternadamente.

En la figura 119 observamos las flechas para lado y lado indicando el avance y retroceso permanente de la CA, pero en ningún momento la corriente atraviesa el dieléctrico del condensador. Sólo se trata de un efecto de carga y descarga sucesiva del condensador. Para entender bien todo lo anterior analizaremos lo que pasa en cada semiciclo de la corriente alterna. (figura 120).

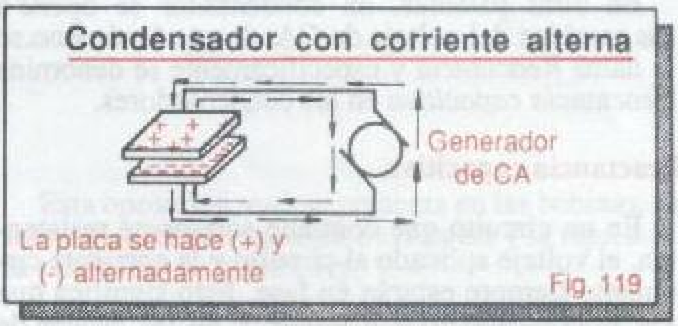

Fig. 119

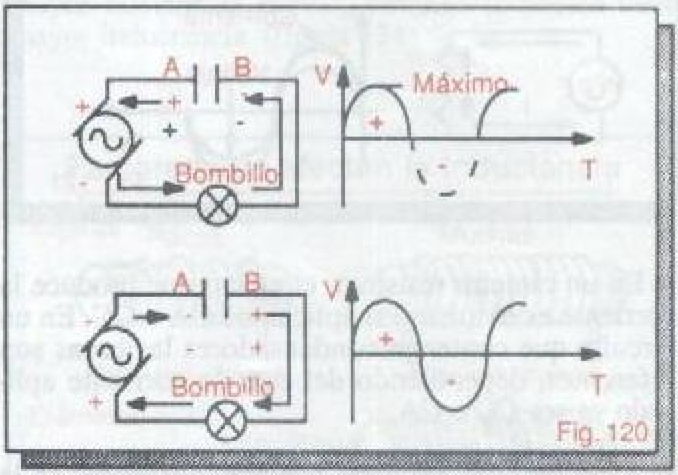

Fig. 120

Durante el semiciclo positivo, la placa A se carga positivamente y la placa B se carga negativamente como ocurre con una fuente de corriente continua, creando un flujo momentáneo de electrones que hace encender la lámpara.

Después de que el condensador se ha cargado en ese sentido, se descarga a través de la lámpara cuando no hay voltaje y ésta se enciende. Luego el condensador se carga con una polaridad contraria, la placa A es ahora negativa y la placa B es positiva.

Cuando el voltaje llega nuevamente a cero, el condensador se vuelve a descargar y la lámpara se enciende de nuevo.

Como los cambios de dirección de la corriente alterna (AC) se suceden muy rápidamente, la corriente fluirá por el circuito en forma permanente y por consiguiente la lámpara se mantendrá encendida.

El efecto real de un condensador en un circuito de CA es producir un retraso del voltaje con respecto a la corriente. La causa de este fenómeno es que el voltaje acumulado en las placas del condensador es de signo contrario al de la fuente.

En otras palabras, un condensador se opone a los cambios del voltaje de CA. A este fenómeno se le llama *Reactancia* y especificamente se denomina *Reactancia capacitiva* en los condensadores.

Reactancia capacitiva

En un circuito que contiene solamente resistencia, el voltaje aplicado al circuito y la corriente circulante siempre estarán en fase. Esto significa que los valores máximos y mínimos de las curvas de voltaje y corriente se suceden al mismo tiempo y siempre tienen el mismo signo o polaridad. (figura 121).

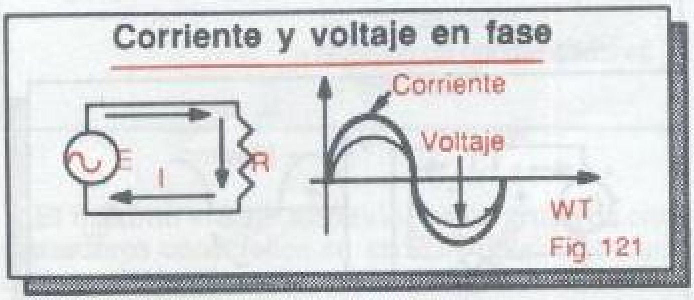
Corriente y voltaje en fase

Fig. 121

En un circuito resistivo, el efecto que produce la corriente es el mismo si aplicamos CA o CC. En un circuito que contenga condensadores las cosas son diferentes, dependiendo del tipo de corriente aplicado ya sea CC o CA.

A un circuito formado por condensadores se le llama un circuito capacitivo. En él el voltaje se atrasa con respecto a la corriente, o lo mismo, la corriente se adelanta al voltaje. (figura 122).

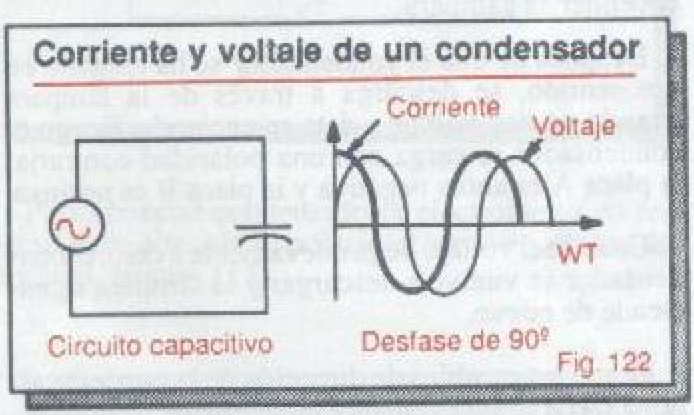
Corriente y voltaje de un condensador

Fig. 122

El voltaje en un circuito capacitivo sufre un desfase o atraso de 90º con respecto a la corriente. Así, cuando la corriente alcanza su máximo valor, el voltaje tiene su mínimo valor y viceversa. En la gráfica anterior observamos que para cero grados (0º) la corriente ha alcanzado su máximo valor y el voltaje apenas va a empezar a tomar valores.

El desfase que ocasiona un condensador crea una oposición al paso de la corriente. Es una especie de resistencia resultante del desfase entre la corriente y el voltaje. A esta resistencia tan especial se le llama *Reactancia Capacitiva*, se expresa como Xc y se mide en *ohmios*, al igual que las resistencias.

En otras palabras, un condensador se comporta como una resistencia variable para señales o voltajes de corriente alterna y cuyo valor en ohmios depende de la frecuencia de la señal.

Este comportamiento es importantísimo y es el que hace posible el funcionamiento de muchos de los circuitos de radio.

Cálculo de la reactancia capacitiva

La reactancia capacitiva de un condensador se puede calcular de forma muy sencilla mediante la siguiente fórmula:

$$Xc = \frac{1}{2\pi FC}$$ Donde

Xc = Reactancia capacitiva dada en ohmios

2π = Constante igual a 6.28 (2 x 3,14)

F = Frecuencia en ciclos por segundo del voltaje alterno aplicado

C = Capacidad en faradios

Hallar la reactancia capacitiva del siguiente circuito:

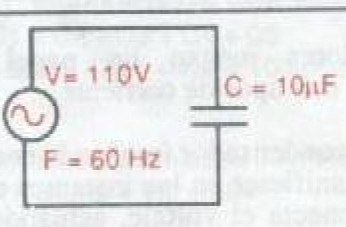

$$Xc = \frac{1}{2\pi FC}$$

Datos:

F = 60Hz = 60 Ciclos por segundo

C = 10µF se debe expresar en faradios, se divide por un millón ya que un Faradio es un millón de µF

$$C = \frac{10\mu F}{1000000} = 0.00001 \text{Faradios}$$

$$Xc = \frac{1}{6.28 \times 60 \times 0.00001} = \frac{1}{0.003768}$$

$$Xc = 265.39\Omega$$

Lección 8

Las bobinas

Las bobinas son uno de los componentes más utilizados en los circuitos de radio y comunicación tanto en los aparatos de transmisión como en los de recepción.

Una bobina es un componente formado por varias vueltas o *espiras* de alambre, enrolladas sobre una forma cilíndrica, cuadrada o rectangular. Las bobinas pueden tener un núcleo que generalmente es de hierro o ferrita. Cuando una bobina no tiene núcleo se dice que tiene *núcleo de aire*. (figura 123).

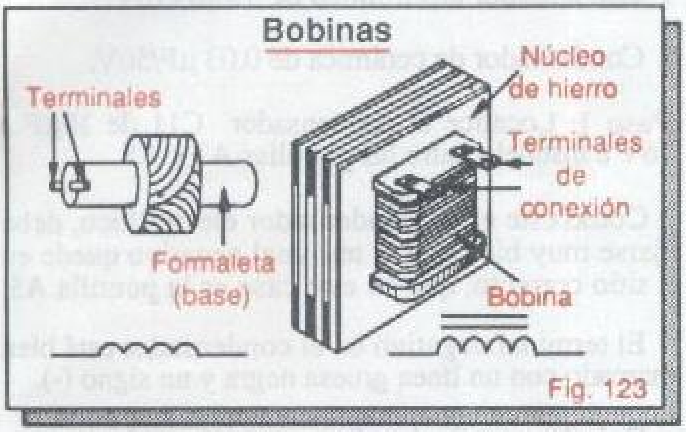
Bobinas

Fig. 123

Las bobinas también reciben el nombre de *inductores* o *choques*.

La palabra inductor viene de *inductancia* que es un fenómeno eléctrico que ocurre en las bobinas y que explicaremos más adelante, y la palabra choque se deriva del efecto de oposición a la variación de la corriente que se produce en una bobina.

Inductancia

En las bobinas ocurre un fenómeno de *oposición* a las *variaciones* de la corriente eléctrica.

En otras palabras, si la corriente en un circuito trata de subir o bajar y en ese circuito hay una bobina, ella se opone a que la corriente suba o baje.

Este comportamiento es comparable a la *inercia* en un movimiento mecánico.

Si un cuerpo con determinado peso se está moviendo y tratamos de aumentar o rebajar su velocidad, hay una fuerza que se opone a ese cambio y que se llama inercia.

Esta oposición, que se presenta en las bobinas, se llama *inductancia*, se mide en *Henrios* y se representa por medio de la letra Hy.

Por lo tanto, la unidad de medida para las bobinas son los Henrios o Hy.

En los circuitos electrónicos actuales se utilizan bobinas pequeñas con inductancias de milihenrios (mH) o microhenrios (µH). La inductancia de una bobina depende de la cantidad y del diámetro de las *espiras*. A mayor diámetro, mayor inductancia y a mayor número de espiras, también la bobina tiene mayor inductancia. (figura 124).

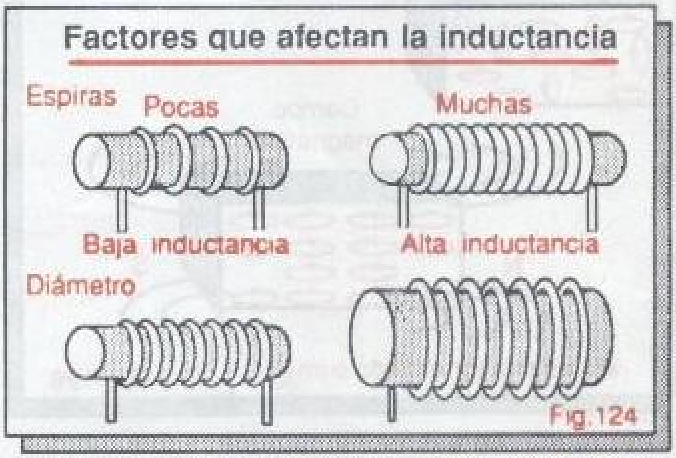
Factores que afectan la inductancia

Fig.124

La inductancia de una bobina también aumenta cuando tiene un núcleo de hierro u otro material magnético y es mayor que cuando tiene núcleo de aire. (figura 125).

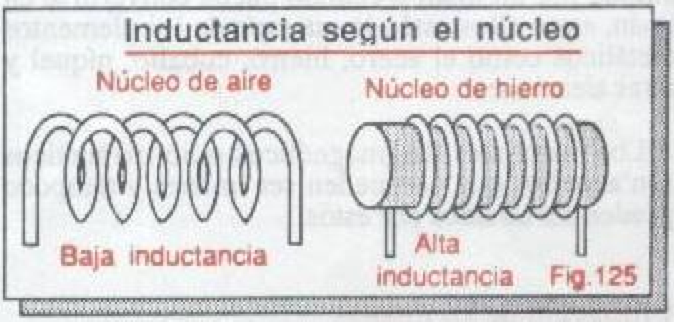
Inductancia según el núcleo

Fig.125

Debido a que una bobina solamente reacciona ante las variaciones de la corriente, no se opone a la corriente continua (CC) y sólo presenta una oposición a la corriente alterna (CA)

Magnetismo

El magnetismo es la propiedad de atracción que presentan ciertos materiales hacia otros.

Estos materiales son conocidos también como *imanes*.

El magnetismo es producido por la atracción de los electrones.

En los imanes, los átomos están orientados de tal manera que los electrones giran en un mismo sentido, formando dominios en los extremos de los imanes, conocidos como polo norte y polo sur. (figura 126)

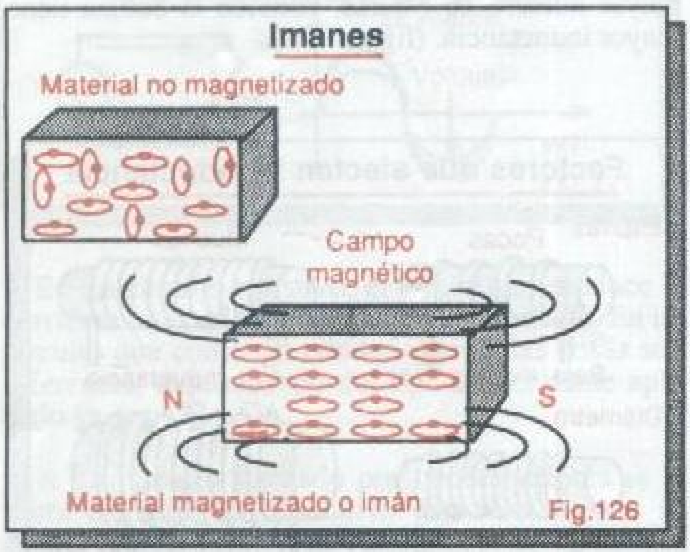

Fig.126

Un imán crea a su alrededor unas líneas de fuerza llamadas *campo magnético*.

Materiales magnéticos y diagmagnéticos

Un material es magnético cuando puede ser atraído por un imán o cuando puede convertirse en imán, entre ellos están la mayoría de los elementos metálicos como el acero, hierro, cobalto, níquel y otras aleaciones.

Los materiales diagmagnéticos o no magnéticos son aquellos que no pueden ser imanes y tampoco pueden ser atraídos por éstos.

Utilización de los imanes

Los imanes se utilizan en la construcción de motores eléctricos, generadores, parlantes, audífonos, micrófonos, instrumentos de medida y muchos otros componentes eléctricos y electrónicos.

Actividad práctica Nº 6

En esta actividad vamos a instalar en el tablero del radio los condensadores al circuito amplificador de audio o sea la etapa marcada con color azul en la lámina del diagrama del radio.

Recuerde las instrucciones para soldar correctamente, entregadas en las prácticas anteriores.

Componentes necesarios

1 Condensador electrolítico de 10 µF/10V o 10 µF/16V.

1 Condensador electrolítico de 47µF/10V.

1 Condensador electrolítico de 330µF/10V.

1 Condensador de cerámica de 0.03 µF/50V.

Paso 1: Localice el condensador C11 de 10µF a 16V e instálelo entre las puntillas A3 y A5.

Como este es un condensador electrolítico, debe fijarse muy bien que el terminal negativo quede en el sitio correcto, que en este caso es la puntilla A5.

El terminal negativo en el condensador está bien marcado con un línea gruesa negra y un signo (-).

Deje los terminales de este condensador con su tamaño original ya que se debe colocar alejado, con el fin de dejar espacio para el potenciómetro RV, (control de volumen).

Paso 2: Localice el condensador de cerámica de 0.03µF, éste puede estar marcado como .03, 303Z o 333Z.

Instale este condensador entre las puntillas A13 y a14, fijándose que quede en una posición recta.

Paso 3: Localice el condensador C13 de 47µF/10V e instálelo entre las puntillas A23 y A26.

Como éste es electrolítico, fíjese que el terminal negativo quede en el sitio correcto según el diagrama, o sea en la puntilla A23.

Paso 4: Localice el condensador C14 de 330µF/10V e instálelo entre las puntillas A31 y A38.

Fíjese bien en la conexión del terminal negativo que debe quedar en la puntilla A31.

Electromagnetismo

Cuando por un conductor circula corriente se establece un flujo o movimiento de electrones en un sentido determinado. Ese movimiento de electrones crea a su alrededor un campo magnético, haciendo que el conductor se comporte como un pequeño imán. A este efecto magnético producido por la corriente eléctrica se le denomina *electromagnetismo*. (figura 127).

Campo magnético en un conductor

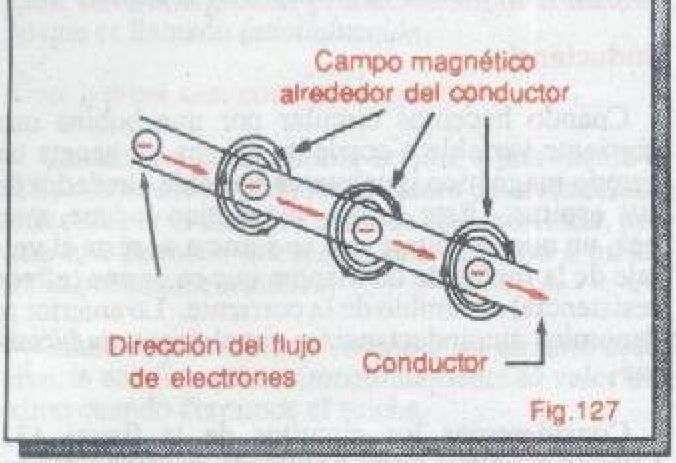

Fig.127

El electromagnetismo es muy importante en el funcionamiento de la mayoría de los aparatos eléctricos y electrónicos. Sin electromagnetismo no serían posible la radio, la televisión, los generadores de electricidad, los motores eléctricos y muchas otras maravillas.

Formación de una bobina

El campo magnético alrededor de un conductor o cable recto es muy débil y no tiene utilidad en la práctica.

Para concentrar el campo electromagnético, el alambre se enrolla formando un carrete o bobina. Así, los campos se suman, dando como resultado un campo electromagnético muy fuerte.

En otras palabras, todo el conjunto se comporta como un potente imán.

Una bobina es el arrollamiento de dos o más vueltas de alambre para producir efectos electromagnéticos especiales. Una vuelta de alambre también se denomina *espira*.

A continuación analizaremos lo que sucede en una espira o vuelta de alambre.

Al circular corriente por la espira, el campo alrededor del conductor de una parte de la espira se refuerza con el de la parte restante y el conjunto constituye un campo total de mayor intensidad. (figura 128).

Corriente en una espira

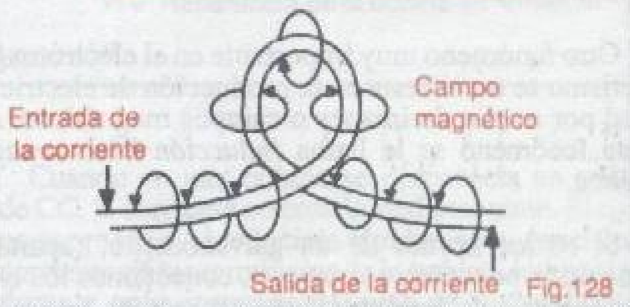

Fig.128

Si tenemos espiras o vueltas muy cercanas o unidas entre sí, el campo magnético resultante será muy fuerte. (figura 129).

Campo magnético en una bobina

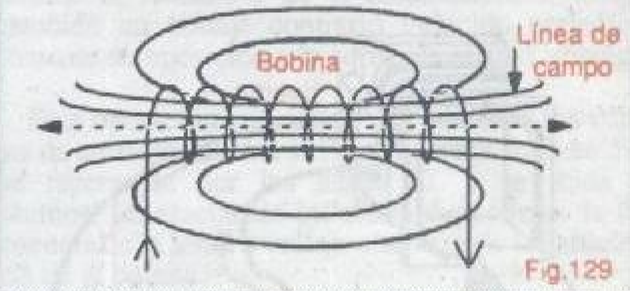

Fig.129

Efectos de un núcleo en la bobina

Al introducir una barra de hierro o algún material magnético dentro de una bobina, obtendremos un electroimán, el cual presentará propiedades magnéticas idénticas a los imanes simples mientras circula corriente por la bobina.

No todos los materiales son buenos conductores de las líneas magnéticas. Algunos tienen mejores propiedades que otros. Por ejemplo: el aire tiene una capacidad magnética muy baja, mientras que el hierro cuenta con una alta propiedad magnética.

Por tal razón, una bobina con núcleo de hierro presenta capacidades magnéticas de atracción y repulsión muy elevadas. En cambio, en una bobina con núcleo de aire estas propiedades son muy bajas.

Además del fenómeno del electromagnetismo que se produce en una bobina debemos tener en cuenta, para entender su funcionamiento, el fenómeno contrario llamado inducción.

Inducción

Producción de electricidad por medio del magnetismo

Otro fenómeno muy importante en el electromagnetismo se manifiesta en la producción de electricidad por medio de imanes o campos magnéticos. A este fenómeno se le llama *Inducción Electromagnética*.

Si en los bornes de un galvanómetro, (aparato que mide pequeñas corrientes), conectamos los extremos de una bobina y luego tomamos un imán fuerte y lo movemos de un lado u otro en las cercanías de ésta, tal como muestra la figura 130, la aguja del galvanómetro se moverá, indicando que sobre los extremos del conductor se está generando una corriente eléctrica.

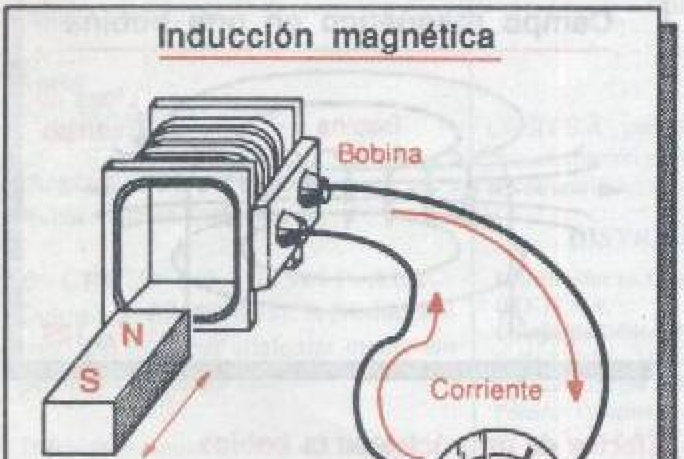
Inducción magnética

Fig.130

Cuando dejamos de mover el imán, la aguja del galvanómetro se detiene.

La deflexión de la aguja se hace más notoria cuando movemos el imán con mayor rapidez.

Para que haya generación de electricidad por este método, es necesario que se cumplan dos condiciones: primero, que el campo magnético esté en movimiento y, segundo, que las líneas de fuerza corten perpendicularmente al conductor.

Un voltaje que se produce en esta forma se denomina voltaje inducido, y, la corriente producida por el mismo, corriente inducida.

Resumiendo, podemos ver entonces que cuando aplicamos corriente a una bobina, ésta produce un campo magnético y que cuando aplicamos un campo magnético a una bobina, en ésta se produce corriente eléctrica.

La electricidad produce magnetismo y el magnetismo produce electricidad.

De esta manera, nos podemos explicar entonces el fenómeno de la inductancia que habíamos mencionado brevemente al principio de la lección.

Inductancia

Cuando hacemos circular por una bobina una corriente variable o corriente alterna, se genera un campo magnético igualmente variable alrededor de sus espiras. Este campo magnético induce, a su vez, un nuevo voltaje que se suma o se resta al voltaje de la fuente de tal manera que se opone (ofrece resistencia) al cambio de la corriente. Lo anterior se denomina autoinductancia o simplemente *inductancia*

Consideremos los circuitos de la figura 131. Cuando el voltaje alterno aplicado en estos circuitos va en ascenso, la corriente en la bobina aumenta: ésto produce un campo magnético que se extiende en torno a la bobina.

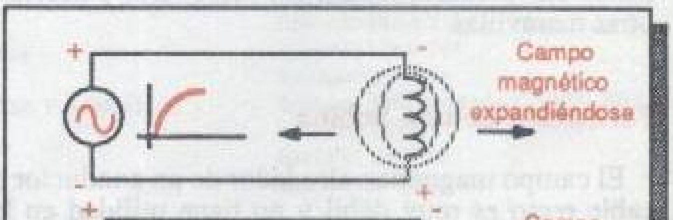

Fig.131

En ese momento, se induce un voltaje a través de la bobina de una polaridad contraria, restándose al voltaje de la fuente y oponiéndose así al cambio (aumento) de la corriente en la misma.

Cuando la corriente va en descenso, empieza a contraerse el campo magnético alrededor de la bobina. Como consecuencia, se invierte la polaridad del voltaje autoinducido en ella.

Este voltaje es ahora de una polaridad que se suma al voltaje de la fuente, tratando de impedir que la corriente disminuya.

Durante la segunda mitad del ciclo (ciclo negativo) ocurre lo mismo. Cuando la corriente aumenta negativamente, el voltaje autoinducido es tal que tiende a impedir el aumento de la corriente. Debido a todo lo anterior, al voltaje autoinducido se le llama contravoltaje o fuerza contra electromotriz (F.C.E.M).

Esta fuerza contraria u oposición es la que produce o genera la *inductancia.* Las bobinas se oponen a los cambios de corriente, debido a la inducción electromagnética producida sobre si mismas, lo que es llamado *autoinducción*.

Una bobina con corriente continua

Cuando conectamos una bobina a una fuente de CC, solamente se produce el efecto de la resistencia ofrecida por el alambre con que está fabricada, pero con una pequeña diferencia con respecto a un circuito puramente resistivo.

Cuando aplicamos el voltaje a un circuito resistivo, la corriente toma inmediatamente su valor máximo cuando cerramos el suiche.

Bobina con corriente continua

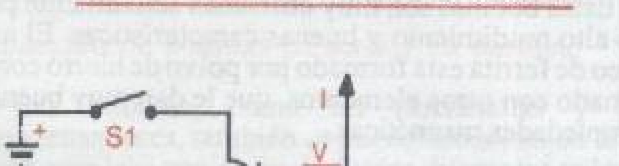

Fig.132

En cambio, en el circuito inductivo de la figura 132, (llamado así por estar formado de una inductor o bobina), la corriente se tarda un determinado tiempo para llegar al valor máximo.

A este tiempo se le llama *constante de tiempo* inductivo y depende de la inductancia en Henrios, de la bobina y de su resistencia.

Para calcularla, se utiliza la siguiente fórmula:

$$t = \frac{L}{R}$$

t = Constante de tiempo, en segundos

L = Inductancia de la bobina en Henrios

R = Resistencia de la bobina en Ohmios

Cuando en una bobina se desconecta un voltaje de CC, la corriente no cesa inmediatamente. El campo magnético, al reducirse rápidamente, genera una corriente en los terminales de la bobina, que aparece durante unos instantes después de suspenderse el voltaje. Por ésto, se dice que las bobinas *almacenan corriente*, así como los condensadores almacenan o guardan voltaje.

Una bobina con corriente alterna

Cuando aplicamos un voltaje de corriente alterna a una bobina, se producirá en ella un campo magnético que está variando continuamente. Por lo tanto, debido al fenómeno de la *autoinducción*, existirá también un voltaje contrario inducido permanentemente en oposición a la corriente alterna principal.

Esta oposición que ofrece una bobina a los voltajes de corriente alterna se llama *reactancia inductiva*, se representa por las letras X_L y se mide en ohmios. La reactancia inductiva depende de la frecuencia de la señal o voltaje alterno y de la inductancia de la bobina.

La fórmula para encontrar la reactancia inductiva es la siguiente:

$$X_L = 6.28 \times F \times L$$; donde:

X_L = Reactancia inductiva, en Ohmios

F = Frecuencia en Hertzios o ciclos por segundo

L = Inductancia de la bobina, en Henrios

Por ejemplo, podemos hallar la reactancia de una bobina de 10 Henrios que está trabajando a una frecuencia de 60Hz.

$X_L = 6.28 \times F \times L$
$X_L = 6.28 \times 60 \times 10$
$X_L = 3768$ Ohmios

El fenómeno de la reactancia inductiva y su dependencia de la frecuencia es fundamental para el funcionamiento de los circuitos de radio como veremos más adelante.

Desfase de la corriente continua y el voltaje en una bobina

En una bobina, al contrario de un condensador, la corriente está atrasada con respecto al voltaje. Este retraso se debe a que la corriente está aumentando desde cero a su valor máximo, y en ese aumento es cuando se produce una mayor inducción de voltaje, tomando éste su máximo valor. (figura 133).

Desfase de la corriente en una bobina

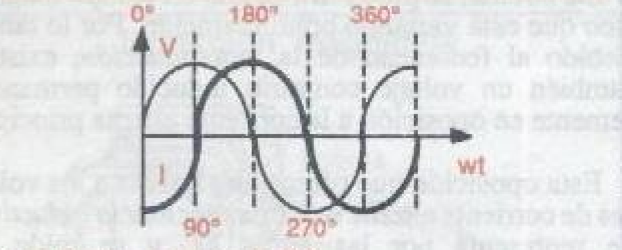

La corriente va atrasada 90° con respecto al voltaje

Fig. 133

Tipos de bobinas

Las bobinas se clasifican por distintos aspectos. La principal división es la de bobinas *fijas* y *variables*. Las bobinas fijas tienen una *inductancia* estable que está dada por sus características físicas.

En las bobinas variables se puede cambiar su inductancia por medio de diferentes métodos.

Las bobinas también se dividen en varios tipos según el núcleo utilizado. Los principales tipos son las de núcleo de hierro, las de núcleo de aire y las de núcleo de ferrita. (figura 134).

Bobinas con núcleo de hierro

Las bobinas con núcleo de hierro son las que tienen una inductancia mayor, ya que este tipo de núcleo aumenta considerablemente la inductancia.

Tipos de bobinas

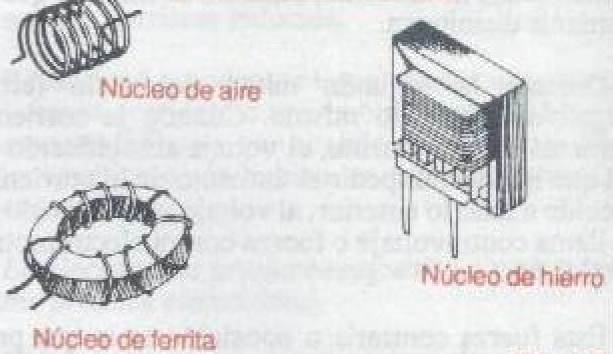

Fig. 134

Este núcleo se fabrica con láminas en forma de E e I, que se van intercalando dentro de la bobina. Estas bobinas también reciben el nombre de *choques* y se encuentran en circuitos de baja frecuencia.

Bobinas con núcleo de aire

Estas bobinas tienen una inductancia muy pequeña y se utilizan en circuitos de alta frecuencia, especialmente en los circuitos sintonizadores de los radios y transmisores de baja potencia. También se utilizan como *choques* de radiofrecuencia para eliminar ciertas señales indeseables en algunos circuitos.

Bobinas con núcleo de ferrita

Estas bobinas son muy utilizadas actualmente por su alto rendimiento y buenas características. El núcleo de ferrita está formado por polvo de hierro combinado con otros elementos, que le dan muy buenas propiedades magnéticas.

En la figura 135 podemos observar diferentes tipos de bobinas con núcleo de ferrita. Entre ellas tenemos la bobina de antena del radio CEKIT que estamos armando como práctica de este curso.

Bobinas con núcleo de ferrita

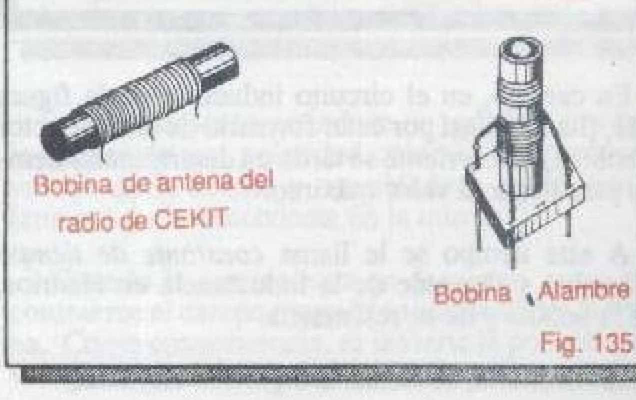
Bobina de antena del radio de CEKIT
Bobina
Alambre

Fig. 135

Un núcleo de ferrita puede aumentar la inductancia de una bobina de 1 milihenrio, hasta valores de 500 milihenrios.

Inductancias o bobinas variables

En los circuitos transmisores y receptores de radio se utilizan con mucha frecuencia bobinas cuya inductancia puede alterarse.

El principal método para producir esta variación es el de mover un núcleo de ferrita en forma cilíndrica dentro de la bobina. Este núcleo se fabrica en forma de tornillo para que se pueda ajustar fácilmente. En la figura 136 podemos observar diferentes tipos de bobinas variables.

Bobinas de inductancia variable

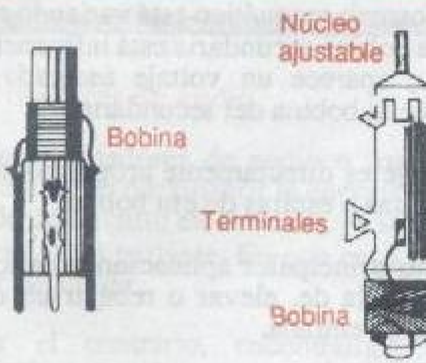

Fig. 136

Conexión de bobinas en serie y en paralelo

Las bobinas, como las resistencias y los condensadores, también se pueden conectar en serie y en paralelo con el fin de lograr diferentes valores de inductancia.

Cuando se conectan dos o más bobinas en serie, la inductancia total será la suma de las inductancias de las bobinas conectadas en el circuito. (figura 137).

Conexión de bobinas en serie

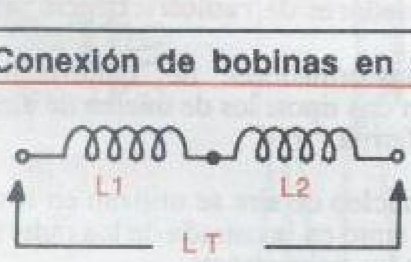

$$LT = L1 + L2 + ... + Ln$$

Fig. 137

Para bobinas en paralelo, la inductancia total se encuentra aplicando las mismas fórmulas que para resistencias en paralelo. (figura 138).

Conexión de bobinas en paralelo

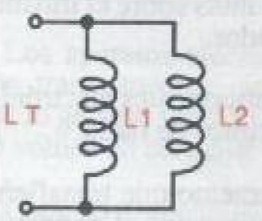

$$LT = \frac{1}{\frac{1}{L1} + \frac{1}{L2} + ... + \frac{1}{Ln}}$$

Fig. 138

Construcción de bobinas

Las bobinas son uno de los pocos componentes que podemos fabricar, ya que es muy fácil hacerlo. Además en la mayoría de proyectos que queramos realizar en forma experimental es necesario fabricar las bobinas, pues éstas no se consiguen en los almacenes especializados.

Estas bobinas se fabrican generalmente de alambre de cobre esmaltado, enrollado sobre una base o formaleta de material aislante. Generalmente, en los proyectos se especifica el número de espiras o vueltas, el diámetro y la longitud de la bobina.

Las bobinas se pueden fabricar a mano, enrollando cuidadosamente el alambre sobre una forma cilíndrica del diámetro indicado y asegurando los terminales del principio y del fin de la bobina para que ésta quede firme. Al alambre esmaltado se le debe quitar su aislante en los extremos para poder hacer las conexiones.

También podemos fabricar una herramienta muy útil, que nos puede ayudar para hacer bobinas de mejor calidad, por medio de un tornillo, un soporte y una manivela. (figura 139).

Fabricación de bobinas

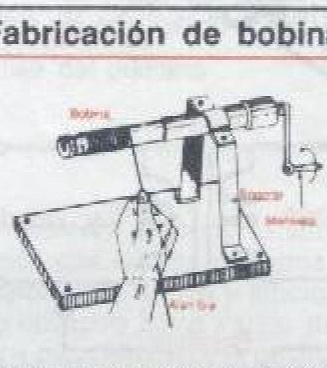

Fig. 139

Los transformadores

Qué es un transformador

Si enrollamos dos bobinas juntas sobre el mismo núcleo, tenemos un transformador.

Los transformadores son componentes electrónicos muy empleados en los circuitos de radio.

Un transformador es un elemento que transfiere energía de un circuito a otro, utilizando el principio de la inducción electromagnética.

Por lo tanto, los transformadores sólo trabajan con señales o voltajes de corriente alterna.

Elementos de un transformador

Un transformador está formado básicamente por dos bobinas y un núcleo.

Este puede ser de aire, de hierro o de ferrita como en las bobinas.

La bobina, por la cual entra la energía, se llama bobina primaria o *primario* simplemente, y la otra bobina se llama *secundario*.

En la figura 140 podemos observar un transformador real y el símbolo empleado en los diagramas para este componente.

El transformador

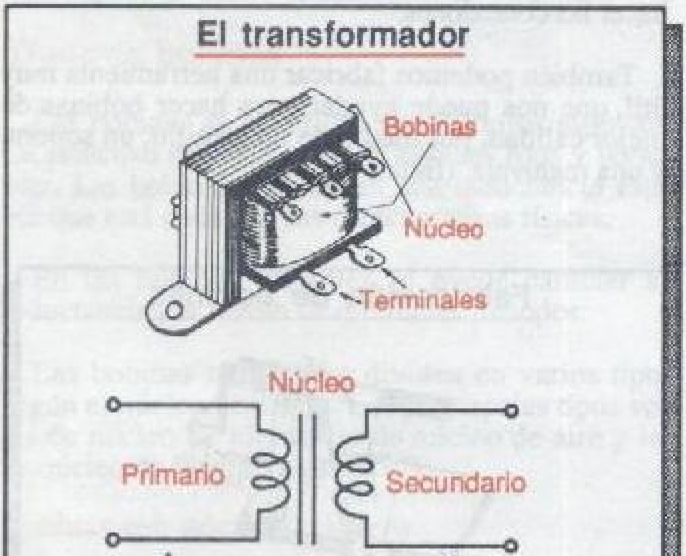

Fig. 140

Principio de funcionamiento

Durante el proceso de transferencia de energía en un transformador siempre se altera o *transforma* alguna característica de la señal de CA aplicada al *primario* o entrada.

Veamos ahora cómo se transfiere energía de una bobina a la otra sin estar conectadas físicamente.

Cuando se conecta una fuente de voltaje de corriente alterna al primario se produce un campo magnético variable que sigue la forma de la onda de CA.

Como el campo magnético está variando constantemente y la bobina secundaria está influenciada por este campo, aparece un voltaje *inducido* en los terminales de la bobina del secundario.

Este voltaje es directamente proporcional al número de vueltas o espiras de esa bobina.

Una de las principales aplicaciones de los transformadores es la de, elevar o rebajar un determinado voltaje de entrada.

Esto se logra enrollando una bobina secundaria con mas o menos vueltas que las de la bobina primaria.

Tipos de transformadores

Los transformadores se clasifican principalmente por su aplicación en los diferentes circuitos y aparatos electrónicos.

Así, tenemos que los principales tipos son los de radiofrecuencia, los de audio o sonido y los transformadores de poder o potencia.

Transformadores de radiofrecuencia

Los transformadores de radiofrecuencia se fabrican en dos tipos: los de núcleo de aire y los de núcleo de ferrita.

Los de núcleo de aire se utilizan en los circuitos de antena, tanto en la entrada de los radios como en la salida de los transmisores.

Estos se fabrican con dos o más bobinas enrolladas sobre una forma cilíndrica.

El tipo más conocido de núcleo de ferrita, es el llamado transformador de *frecuencia intermedia* o FI; tenemos tres en el radio de práctica de CEKIT y, están identificados como T1, T2 y T3. (figura 141).

Trasnsformadores de FI del radio CEKIT

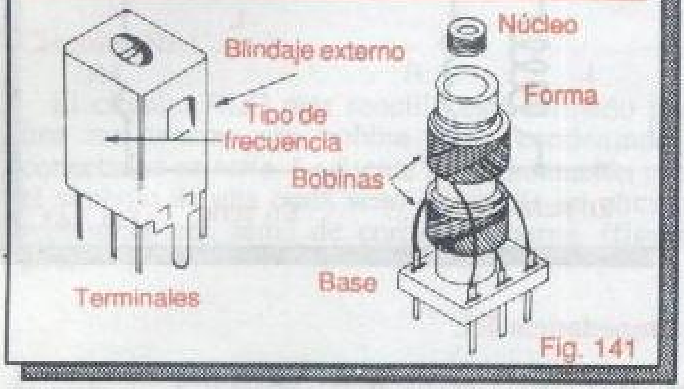

Fig. 141

Transformadores de audio

Los transformadores de audio o sonido, se utilizan para adaptar o transferir la energía de baja frecuencia de un circuito electrónico, ya sea de tubos o transistores, a un parlante. En este caso es un *transformador de salida.*

Si por el contrario, encontramos un transformador al cual se le conecta un micrófono en el primario y su secundario está conectado a un amplificador de sonido, es un *transformador de entrada* y está cumpliendo también la función de acoplar la energía del micrófono al amplificador.

Podemos observar su forma general de conexión y su aspecto físico en la figura 142.

Transformadores de audio

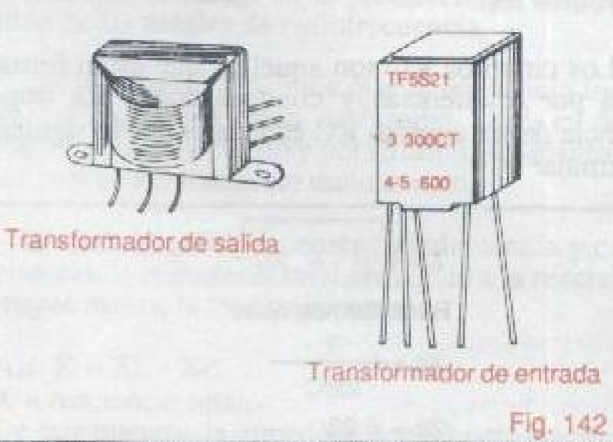

Fig. 142

Transformadores de poder

Los transformadores de poder, también llamados de alimentación o de potencia, son muy utilizados en casi todos los aparatos electrónicos y tienen la función de rebajar o aumentar el voltaje de entrada, ya sea de 110 ó 220 voltios, con el fin de alimentar los circuitos electrónicos internos de estos aparatos.

Los aparatos electrónicos modernos, fabricados con transistores y circuitos integrados, requieren por lo general voltajes bajos del orden de los 5 a los 20 voltios en corriente continua.

Como la alimentación que tenemos disponible en los tomacorrientes, es de corriente alterna a 110 ó 220 voltios, debemos rebajarla primero a bajo voltaje y luego convertirla a corriente continua.

Cuando el voltaje en el secundario es mayor que en el primario, se dice que es un transformador *elevador*, y cuado el voltaje en el secundario es menor, se dice que es un *reductor*.

Se puede tener el caso de transformadores de alimentación con varios secundarios en donde unos tienen voltaje mayor y otros menor.

Para que un transformador tenga más voltaje en el secundario, esta bobina debe tener más vueltas o espiras de alambre que el primario y si el transformador es reductor, el secundario debe tener menos espiras que el primario.

Esta relación de voltaje se llama *relación de transformación* o relación de espiras.

$$\text{Relación de espiras} = \frac{Ns}{Np} = \frac{Vs}{Vp}$$

Ns = Número de espiras del secundario

Np = Número de espiras del primario

Vs = Voltaje del secundario

Vp = Voltaje del primario

Los transformadores de poder o alimentación se fabrican también con núcleo en forma de láminas, de aleación especial de hierro y silicio, y en tamaños desde 1 centímetro hasta varios metros, como en el caso de los transformadores de distribución de energía de potencia que utilizan el mismo principio.

La Impedancia

La impedancia, en términos generales, se refiere a la resistencia total de un circuito que contiene bobinas, condensadores y resistencias . Es decir, es el efecto resultante de las resistencias y las reactancias inductiva y capacitiva.

En los circuitos de radiofrecuencia, como en los radios, transmisores y televisores, donde tenemos este tipo de componentes en gran cantidad y señales de corriente alterna, la impedancia juega un papel muy importante para determinar el comportamiento de estos circuitos con respecto a la corriente y al voltaje.

En los circuitos con este tipo de componentes y con señales de corriente alterna, se aplica la Ley de Ohm al trabajar con la impedancia. La impedancia se representa con la letra Z y también se mide en ohmios.

La impedancia se encuentra con la siguiente fórmula:

$$Z = \sqrt{R^2 + X^2}$$

Donde:

Z = Impedancia en ohmios

R = Resistencia en ohmios

X = Reactancia en ohmios

Circuitos combinados

Ya hemos estudiado en forma individual algunos de los componentes más importantes de los circuitos de radio como son las resistencias, los condensadores, las bobinas y los transformadores.

Ahora vamos a ver cómo se pueden conectar estos componentes para formar diferentes circuitos combinados, muy utilizados en este tipo de aparatos.

Estos circuitos reciben también el nombre de *circuitos sintonizados*.

Circuito RL

Los circuitos RL son aquellos que están compuestos por resistencias y bobinas conectadas en serie o en paralelo. (figura 143).

Circuitos RL

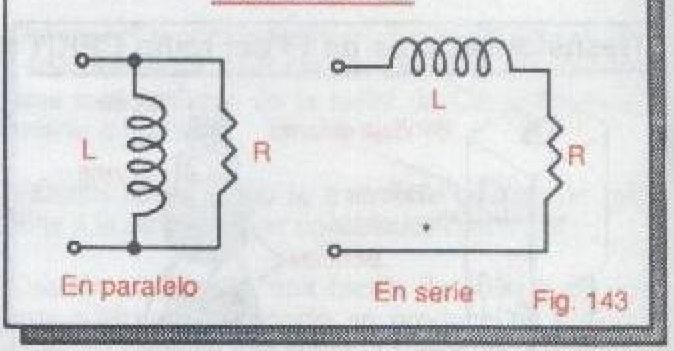

Fig. 143

Impedancia RL

Es la oposición total que ofrece a la corriente un circuito que contiene resistencias y bobinas.

$$Z = \sqrt{R^2 + XL^2}$$

Recordemos que la reactancia inductiva o XL = 2πFL. (2π = 6.28)

Para calcular la corriente en un circuito RL se aplica la ley de Ohm:

$$I = \frac{V}{Z}$$

Z = Impedancia

Circuito RC

Los circuitos RC son aquellos que están formados por resistencias y condensadores. La impedancia de un circuito RC está dada por la siguiente fórmula:

$$Z = \sqrt{R^2 + Xc^2}$$

Recordemos que:

$$Xc = \frac{1}{2\pi fC}$$

$2\pi = 6{,}28$

Hallando la impedancia, se puede aplicar la Ley de Ohm para calcular la corriente en el circuito.

Circuito RLC

Son circuitos que contienen bobinas, condensadores y resistencias; que conectadas adecuadamente, conforman los circuitos de sintonía de los receptores de radio, televisores y radares, entre otros.

Circuito RLC serie

El circuito RLC más sencillo está formado por una resistencia, una bobina y un condensador, conectados en serie. La fuente de alimentación con el símbolo de una onda seno dentro de un círculo representa una señal de corriente alterna. (figura 144).

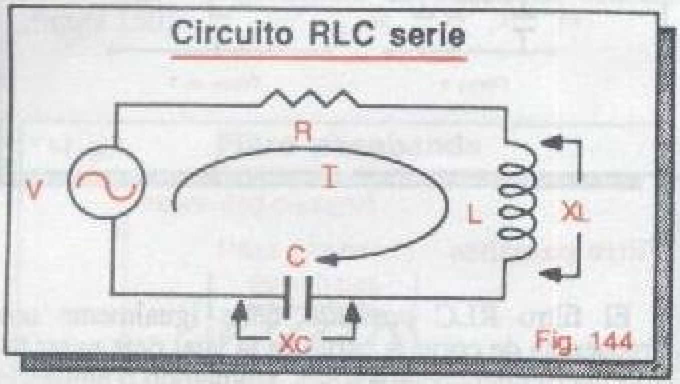

Fig. 144

Como habíamos dicho anteriormente, la reactancia capacitiva y la reactancia inductiva tienen efectos contrarios.

En el caso del condensador, la corriente se adelanta con respecto al voltaje y en la bobina la corriente se atrasa.

La combinación de estos dos efectos en un mismo circuito produce un fenómeno muy interesante, que es básico en la producción y amplificación de las señales de radiofrecuencia.

En la figura 145 vemos que la reactancia inductiva y capacitiva están desfasadas 180º, es decir, van en sentido opuesto y por lo tanto, sus efectos resistivos se contrarrestan mutuamente.

En un circuito que contenga inductancia y capacitancia, la reactancia total será igual a la reactancia mayor menos la reactancia menor.

Así $X = X_L - X_C$

X = reactancia total.

De esta manera, la impedancia total para un circuito RLC en serie, será:

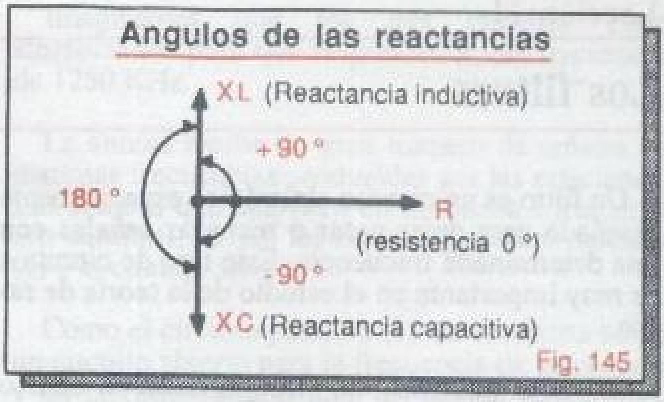
Angulos de las reactancias
XL (Reactancia inductiva)
+ 90 °
180 °
R
(resistencia 0 °)
- 90 °
XC (Reactancia capacitiva)

Fig. 145

$$Z = \sqrt{R^2 + (X_L - X_C)^2}$$

Ejemplo. Si tenemos el circuito siguiente:

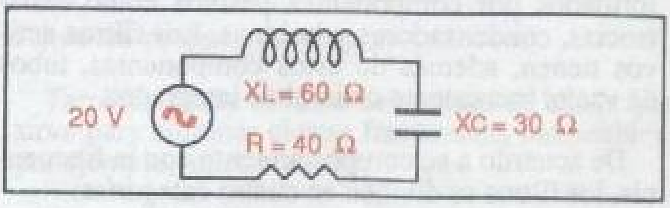

Cuál será la impedancia total ?. Aplicando la fórmula:

$$Z = \sqrt{R^2 + (X_L - X_C)^2}$$

$$Z = \sqrt{(40)^2 + (60 - 30)^2} = \sqrt{1600 + 900}$$

$$Z = \sqrt{2500} = 50\,\Omega$$

Cuando la reactancia capacitiva es igual a la reactancia inductiva, se dice que el circuito está en *resonancia*; tema del cual hablaremos ampliamente en una próxima lección.

Circuito RLC en paralelo

En este tipo de circuito, los tres componentes básicos, resistencia, condensador y bobina se conectan en paralelo y el voltaje que se aplica a cada uno de ellos es el mismo. De igual manera que en el circuito RLC en serie, la corriente sufre un desfase en la bobina y en el condensador.

La mayor corriente circula por el elemento que presenta menor resistencia. La corriente que circula por la bobina, lleva sentido contrario a la corriente que circula por el condensador y si son iguales, se anulan mutuamente. Esto sucede cuando la reactancia capacitiva es igual a la reactancia inductiva, es decir, cuando el circuito entra en *resonancia*.

Los filtros

Un filtro es un circuito electrónico especialmente diseñado para dejar pasar o rechazar señales con una determinada frecuencia. Este tipo de circuitos es muy importante en el estudio de la teoría de radio.

En otras palabras, un filtro selecciona cuál o cuáles señales pueden pasar de un punto a otro, en un circuito, dependiendo de su frecuencia en ciclos por segundo.

Hay dos grandes grupos de filtros: los filtros pasivos y los filtros activos.

Los filtros pasivos son aquellos que están formados por componentes pasivos como resistencias, condensadores y bobinas. Los filtros activos tienen, además de estos componentes, tubos de vacío, transistores o circuitos integrados.

De acuerdo a su comportamiento con la frecuencia, los filtros se dividen en cuatro categorías:

1) Filtros pasabajos
2) Filtros pasaaltos
3) Filtros pasabanda
4) Filtros de rechazo de banda

Filtro pasabajos

Un filtro *pasabajos* es un circuito que deja pasar todas las frecuencias por debajo de la llamada frecuencia de corte o frecuencia límite. Las frecuencias más altas las atenúa o rebaja y por lo tanto no pueden pasar. (figura 146).

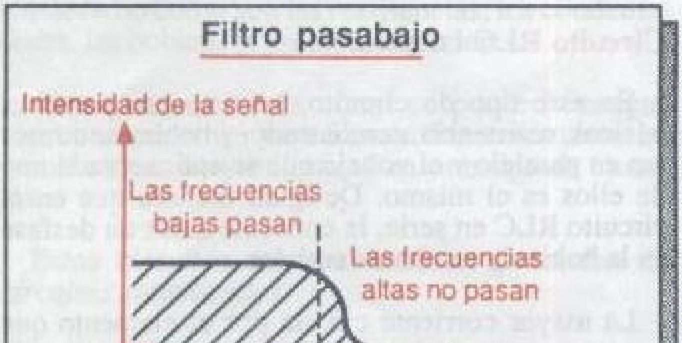
Filtro pasabajo

Fig. 146

En la figura 147 mostramos las configuraciones más utilizadas en receptores:

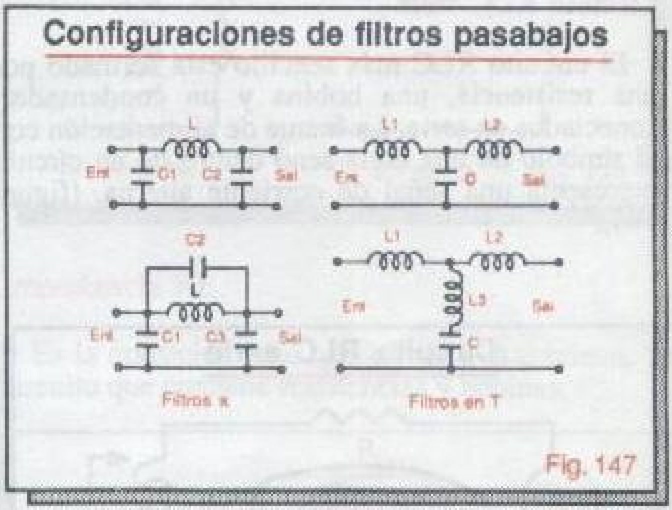
Configuraciones de filtros pasabajos

Fig. 147

Filtro pasaaltos

El filtro RLC *pasaalto* tiene igualmente una frecuencia de corte, a partir de la cual deja pasar todas las frecuencias mayores, atenuando o anulando todas las señales que tengan una frecuencia menor. (figura 148).

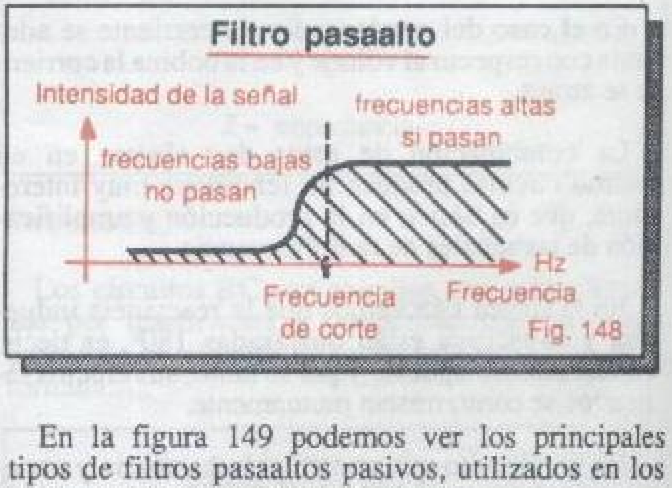
Filtro pasaalto

Fig. 148

En la figura 149 podemos ver los principales tipos de filtros pasaaltos pasivos, utilizados en los aparatos de radio y comunicaciones.

Filtro pasabanda

El filtro *pasabanda* es aquel que deja pasar un conjunto o banda de señales dentro de un rango de frecuencias.

Configuraciones de filtros pasaaltos

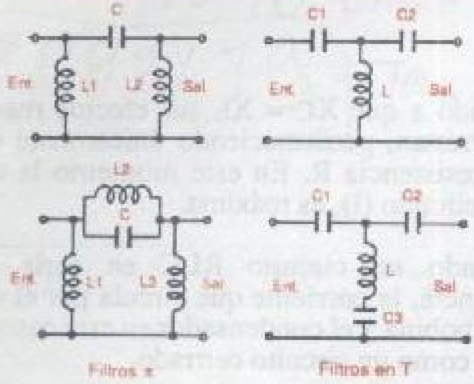

Fig. 149

También se dice que selecciona un *ancho de banda* determinado y rechaza las demás frecuencias. (figura 150).

Filtro pasabanda

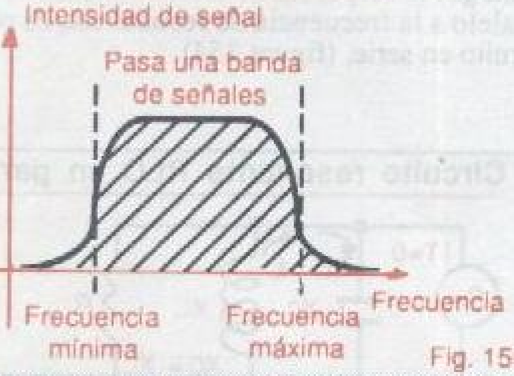

Fig. 150

La figura 151 muestra un filtro pasabanda, utilizado como etapa sintonizadora en un receptor de radio.

Filtro pasabanda en un radio

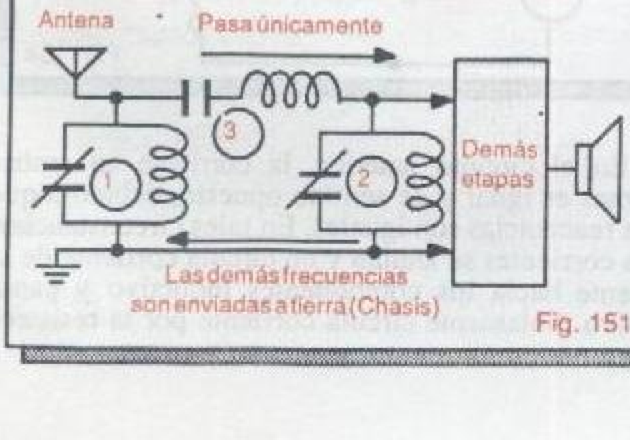

Fig. 151

Imaginemos que los tres circuitos están sintonizados para que respondan a una frecuencia de 1250 KHz.

La antena recibe un gran número de señales de distintas frecuencias producidas por las estaciones. Las señales que aparecen en la antena encuentran dos caminos: el que les ofrece el circuito paralelo (1) y el circuito serie (3).

Como el circuito paralelo (1) se comporta como un circuito abierto para la frecuencia de resonancia y un circuito cerrado para las demás frecuencias, todas las señales se van a tierra por este camino, excepto la señal de 1250 KHz (frecuencia de resonancia).

Debido a que el camino serie (3) se comporta como circuito cerrado para la señal de 1250 KHz, ésta puede circular por este camino llegando sin ningún problema a las demás etapas.

Filtro de rechazo de banda

También existe el filtro de rechazo de banda que sirve para eliminar ciertas frecuencias indeseables, por ejemplo, interferencias.

Este filtro deja pasar todo el espectro de frecuencias, excepto una pequeña franja de él. (figura 152).

Filtro de rechazo de banda

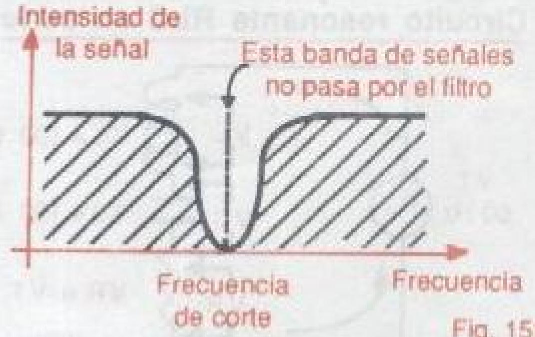

Fig. 152

En estas circunstancias la frecuencia resonante es atenuada.

Todos los sistemas de radio y comunicaciones están basados en sistemas de filtros de un tipo o de otro que, combinados adecuadamente, aceptan o rechazan ciertas señales.

Así podemos lograr que un receptor capte la señal que nosotros queremos ver o escuchar, de ahí su importancia en la teoría y práctica de radio.

Resonancia

Habíamos mencionado anteriormente que si en un circuito RLC en serie o en paralelo, la reactancia inductiva era igual a la reactancia capacitiva, su efecto se cancelaba y se producía un fenómeno llamado *resonancia.*

Los circuitos resonantes son los que hacen posible la sintonización y amplificación de una sola señal de radio, que se selecciona entre la gran cantidad que llegan a la antena de un receptor de radio en un momento dado.

Resonancia en serie

Un circuito RLC entra en resonancia solamente a una determinada frecuencia, cuando las dos reactancias son iguales.

A esta frecuencia se le llama *frecuencia de resonancia.*

Como ejemplo, consideremos el siguiente circuito, donde los valores de la bobina y el condensador son tales, que a una frecuencia de 60 Hz, sus reactancias son iguales. (figura 153).

$X_C = X_L = 100\Omega$

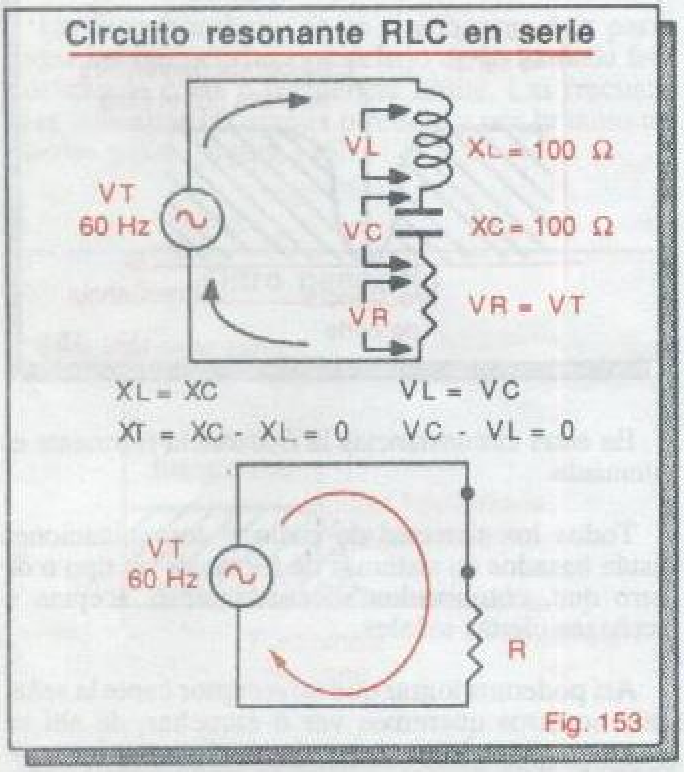
Circuito resonante RLC en serie

Fig. 153

Debido a que $X_C = X_L$, sus efectos reactivos se contrarrestan, permaneciendo únicamente el efecto de la resistencia R. En este momento la corriente por el circuito (I), es máxima.

Cuando un circuito RLC en serie está en resonancia, la corriente que circula por él es máxima; la bobina y el condensador en este caso se comportan como un circuito cerrado.

Si cambiamos la frecuencia de la fuente o señal, el circuito sale de resonancia (pierde sintonía) ya que se afectan las reactancias inductivas y capacitivas puesto que cada una responde diferente a las frecuencias, haciéndose X_C diferente a X_L.

Resonancia en paralelo

Al igual que los circuitos RLC serie, la resonancia en paralelo se presenta cuando $X_C = X_L$. Sin embargo, la respuesta de los circuitos resonantes en paralelo a la frecuencia de resonancia, es opuesta al circuito en serie. (figura 154).

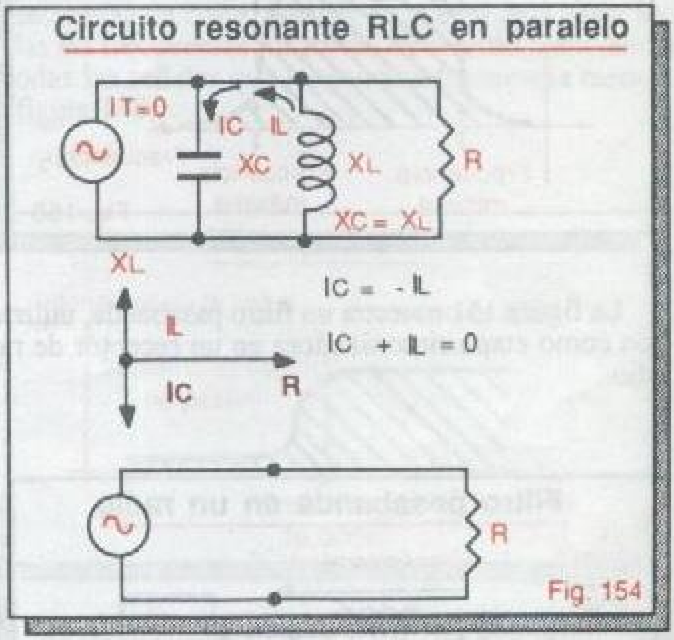
Circuito resonante RLC en paralelo

Fig. 154

En el circuito anterior, la corriente en ambas ramas es igual y de sentido opuesto, debido a que las reactancias son iguales. En tales circunstancias, las corrientes se anulan y no circula corriente de la fuente hacia los componentes inductivo y capacitivo. Solamente circula corriente por la resistencia.

Al suceder lo anterior, la bobina y el condensador en conjunto, se comportan como un circuito abierto, o sea, su impedancia es muy alta.

El intercambio de energía entre el condensador y la bobina, que genera la corriente por ellos circulante, se conserva permanentemente dentro del circuito resonante en paralelo y puede ser empleado para ciertos propósitos; por esta razón, a los circuitos resonantes en paralelo, se les denomina "circuitos tanque" debido a su poder de acumular energía.

Frecuencia de resonancia

Todo circuito que contenga bobinas y condensadores, resuena a una frecuencia determinada llamada condición de resonancia o frecuencia de resonancia, FR. Para que ésto ocurra, las dos reactancias deben ser iguales:

$$X_L = X_C$$

Como:

$$X_L = 6.28 \times F_R \times L$$

y

$$X_C = \frac{1}{6.28 \times F_R \times C}$$

Si reemplazamos XL y XC por sus respectivas expresiones tenemos:

$$X_L = X_C$$

$$6.28 \times F_R \times L = \frac{1}{6.28 \times F_R \times C}$$

$$(6.28 \times F_R \times L) \times (6.28 \times F_R \times C) = 1$$

$$(6.28)^2 \times F_R^2 \times LC = 1$$

De lo anterior, podemos encontrar la frecuencia a la cual llamamos frecuencia de resonancia.

$$F_R^2 = \frac{1}{(6.28)^2 \times LC}$$

$$F_R = \frac{1}{(6.28) \times \sqrt{LC}}$$

$$F_R = \frac{0.159}{\sqrt{LC}}$$

Ejemplo:

Supongamos que tenemos el circuito de sintonía de un radio, compuesto por un condensador variable y una bobina de antena conectados en paralelo.

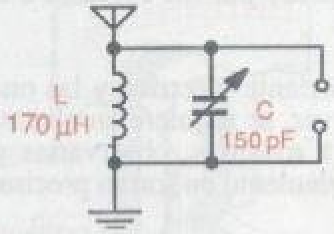

Qué frecuencia de resonancia, o en otras palabras, qué emisora se sintonizará, si el condensador variable queda en una posición cuya capacidad es de 150 pF (picofaradios).

Recordemos que: $1\mu H = 10^{-6} H$ y $1pF = 10^{-12} F$

Aplicando la fórmula tenemos:

$$F_R = \frac{0.159}{\sqrt{LC}} = \frac{0.159}{\sqrt{170 \times 10^{-6} \times 150 \times 10^{-12}}}$$

$$F_R = \frac{0.159}{1.59 \times 10^{-7}} = 1.000.000 \text{ Hz}$$

o FR = 1.000 KHz. Esto quiere decir que en esa posición del condensador variable se sintoniza la emisora que transmite a 1000 Kilohertz o 1000 Kilociclos. Más adelante, al avanzar el curso, estaremos utilizando este importante concepto de la frecuencia de resonancia al analizar las diferentes etapas del radio.

Los instrumentos de medida

Introducción

Hemos hablado, al iniciar este curso, de las ondas de radio, de la corriente eléctrica, del voltaje, de la resistencia y de los circuitos eléctricos. También se ha mencionado el tema de la corriente alterna, la corriente continua, la amplitud, la frecuencia y las formas de onda.

Estudiamos además los principales componentes electrónicos utilizados en los aparatos de transmisión y recepción de radio y también se estudiaron la impedancia, los filtros y la resonancia.

Todos estos fenómenos y componentes relacionados con los aparatos de radio, tienen un comportamiento establecido por los estudios e investigaciones.

Como la corriente eléctrica y las ondas de radio no se pueden ver, se requiere en electrónica de *instrumentos* para medirlas, observarlas y determinar así su comportamiento en forma precisa.

Simultáneamente, con el desarrollo de los componentes y los circuitos electrónicos, se han creado una gran cantidad de instrumentos de todo tipo que nos sirven para el diseño, reparación y mantenimiento de estos aparatos.

No se puede concebir un ingeniero, un técnico o un aficionado a la electrónica que no conozca y pueda utilizar los instrumentos electrónicos de medida.

Existen actualmente desde instrumentos muy sencillos, para realizar medidas simples, hasta complejos aparatos de gran precisión controlados por computador y que requieren de mucho estudio y entrenamiento para conocer su manejo y aplicación.

Siendo este tema fundamental en cualquier curso de electrónica, vamos a estudiar el manejo y aplicación de los instrumentos necesarios para ensamblar, ajustar o reparar los aparatos de recepción y transmisión de radio.

Tipos de instrumentos

Existen dos grandes grupos o familias de instrumentos: los análogos y los digitales.

Inicialmente, y hasta hace pocos años, todos los instrumentos eran del tipo análogo o analógico que utilizan una aguja que se mueve sobre una escala, para indicar el valor medido en ese momento.

Desde hace unos años, y haciéndose cada vez más notoria, hay una tendencia hacia el desarrollo y aplicación de instrumentos electrónicos digitales en los cuales la indicación es del tipo numérico, mostrando directamente la medida en forma de dígitos.

En la figura 155, podemos observar la diferencia fundamental entre estos dos tipos de instrumentos al medir una magnitud de 4.2 voltios, por ejemplo.

Tipos de instrumentos

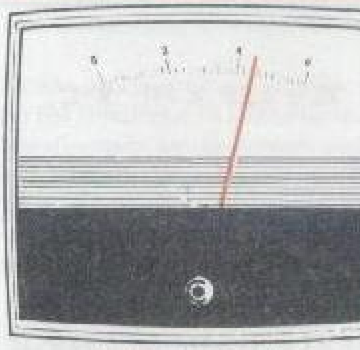

Voltímetro análogo

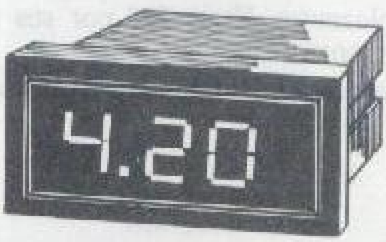

Voltímetro digital

Fig. 155

La mayoría de los instrumentos electrónicos, se utilizan para medir magnitudes eléctricas como el voltaje, la corriente y la resistencia.

Otros se utilizan para medir el valor o la calidad de los componentes como las resistencias, los condensadores, las bobinas, los diodos, los transistores, etc. A estos grupos se les llama instrumentos de medida.

Hay otro grupo de instrumentos que producen señales que se aplican a los circuitos para observar y verificar su funcionamiento. Estos se llaman los generadores de señal de baja y alta frecuencia y son muy utilizados en los circuitos de radio.

Otro instrumento muy común en electrónica es el *osciloscopio*, que permite visualizar las formas de onda de las señales presentes en los diferentes puntos de un aparato y que también es muy útil para la medida del voltaje, la corriente y la frecuencia de estas señales.

Los técnicos y aficionados pueden fabricar sus propios instrumentos que sean útiles para su trabajo. Durante este curso, suministraremos toda la información para la realización de algunos de estos proyectos.

Principio básico de funcionamiento

Los instrumentos electrónicos de medida se basan en su mayoría en la medida de la corriente eléctrica que circula por ellos.

Para derivar de la cantidad de corriente, una medida de su magnitud o tamaño, se utiliza el principio del *electromagnetismo,* o sea el uso del campo magnético que se produce alrededor de un conductor, cuando circula corriente eléctrica por él.

El primer trabajo que dió origen a los´instrumentos de aguja fue elaborado por A.D'Arsonval en 1882 al desarrollar un *galvanómetro* utilizado para medir corrientes eléctricas muy débiles, en los laboratorios de la época. En el año de 1888, E. Watson modificó el galvanómetro y lo adaptó para medir corrientes más grandes y redujo su tamaño considerablemente. Sin embargo, a los instrumentos modernos todavía se les conoce como del tipo D'Arsonval.

El medidor básico se conoce como de *bobina móvil* en donde una bobina, a la cual está conectada la aguja, se mueve dentro de un campo magnético generado por un imán permanente.

El desplazamiento de la bobina depende de la cantidad de corriente que circula por ella, y como la aguja se mueve simultáneamente con la bobina, nos muestra en una escala, debidamente trazada o calibrada, la magnitud o tamaño de esa corriente. En la figura 156 podemos observar la construcción y el funcionamiento básico de este instrumento.

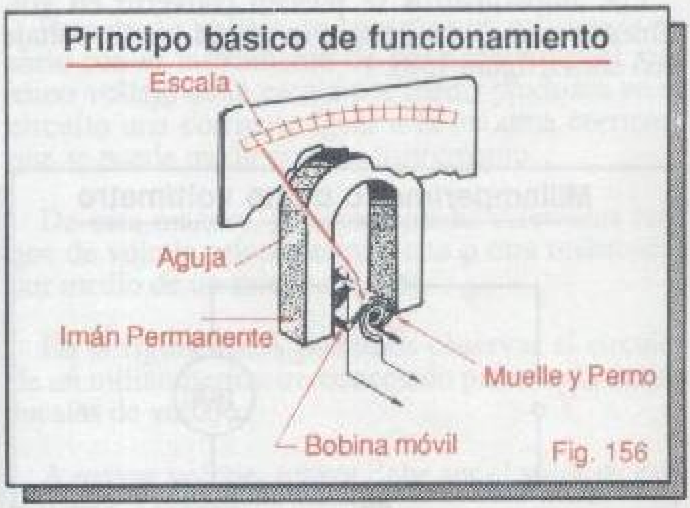
Principo básico de funcionamiento

Fig. 156

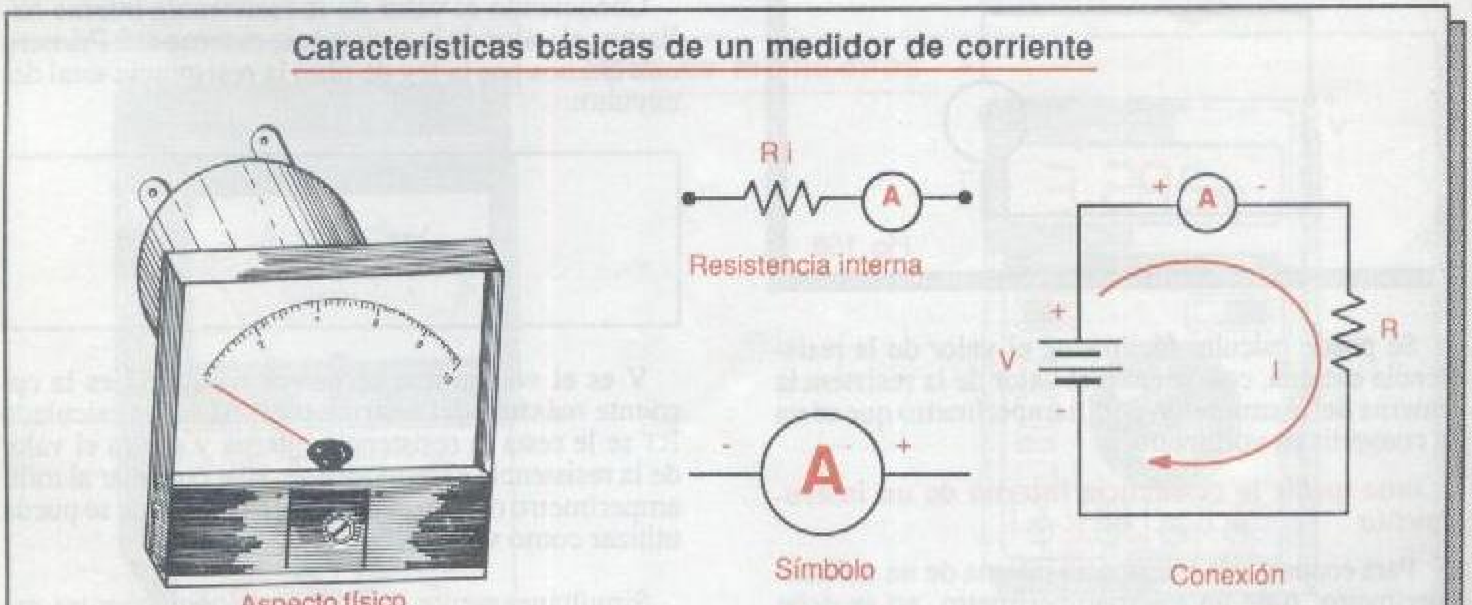
Características básicas de un medidor de corriente

Aspecto físico

Resistencia interna

Símbolo

Conexión

Fig. 157

Características de un instrumento

Los instrumentos básicos de bobina móvil están fabricados para medir corriente y se llaman microamperímetros, miliamperímetros o amperímetros y deben estar conectados en serie con el circuito.

La bobina del instrumento presenta una cierta resistencia que hace que circule una determinada corriente por ella. A esta resistencia se le llama *resistencia interna* y está directamente relacionada con la escala de corriente que se va a medir. Así tenemos medidores de 50 µA, 500 µA, 1 mA, 100 mA, 1 Amp, 5 Amp, etc. (figura 157).

La resistencia interna será menor mientras más grande sea la escala de corriente que mide el instrumento.

Los amperímetros se pueden convertir en voltímetros con el fin de que se pueda medir voltaje con ellos. (figura 158).

Miliamperímetro como voltímetro

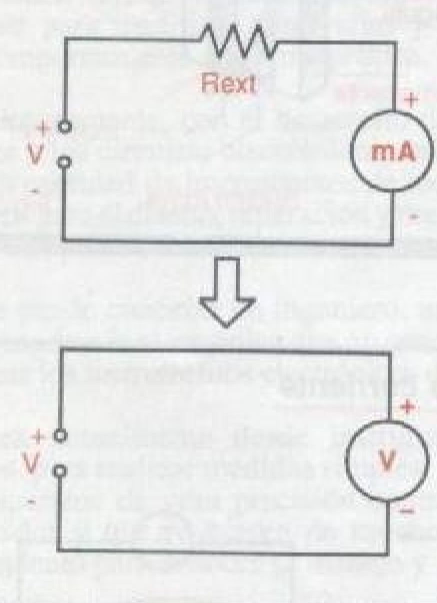

Fig. 158

Se puede calcular fácilmente el valor de la resistencia externa, conociendo el valor de la resistencia interna del instrumento o miliamperímetro que se va a convertir en voltímetro.

Cómo medir la resistencia interna de un instrumento

Para encontrar la resistencia interna de un miliamperímetro o de un microamperímetro, no se debe medir con un multímetro común, pues éste puede presentar un voltaje más alto del que soporta la bobina y se puede dañar.

Para hacerlo, se debe armar el circuito que se muestra en la figura 159.

Medida de la resistencia interna

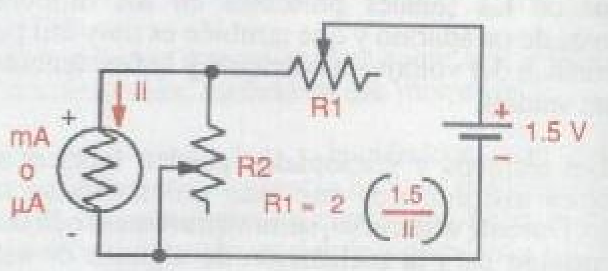

Fig. 159

Primero se conecta la pila, luego el potenciómetro R1, que debe ser de un valor del doble de ohmios que el necesario para lograr la deflexión a lo máximo de la escala y el instrumento, dejando R2 desconectado.

Se debe ajustar el potenciómetro R1 hasta que la aguja llegue al final de la escala. Luego conecte el potenciómetro R2 y gradúe el cursor hasta que el instrumento indique una lectura a la mitad de la escala.

En ese momento, el valor de R2 es el mismo que el valor de la resistencia interna del instrumento. Retírelo del circuito con cuidado para no mover el cursor y mida su valor con un multímetro común. En la siguiente lección se estudia la forma de hacerlo.

Conociendo el valor de la resistencia interna hallamos el valor de la resistencia externa así: Primero calculamos por la ley de ohm la resistencia total del circuito:

$$R_T = \frac{V}{I}$$

V es el voltaje que se quiere medir e I es la corriente máxima del instrumento. Al valor calculado R_T se le resta la resistencia interna y ese es el valor de la resistencia externa que se debe conectar al miliamperímetro o microamperímetro para que se pueda utilizar como voltímetro.

Simultáneamente, con algunos componentes externos, se puede armar un óhmetro, que es el instrumento utilizado para medir el valor en ohmios de las resistencias, como veremos en la siguiente lección.

Lección 14

El multímetro electrónico

Para los trabajos de los técnicos y aficionados existe un instrumento muy versátil que reune en un sólo aparato la posibilidad de medir corriente, voltaje y resistencia, en varias escalas o rangos. A este instumento se le llama multímetro o comunmente *tester*.

De los multímetros también hay dos grandes grupos: los multímetros análogos y los multímetros digitales. Estos últimos son los que más se utilizan actualmente. En la figura 160 tenemos un multímetro de cada tipo.

Internamente, un multímetro es un medidor de corriente muy sensible que tiene una escala graduada en varias unidades y tipos de medición.

Para medir diferentes magnitudes de corriente, se le conectan resistencias en paralelo al instrumento básico.

A esta resistencia se le conoce como "shunt" y su valor depende de la máxima corriente que se va a medir.

En la figura 161 tenemos un circuito de un miliamperímetro al cual se le han conectado varias resistencias en paralelo. En la lección 6, página 63 de este curso, vimos cómo la corriente se distribuía a través de las diferentes resistencias según su valor en ohmios. De acuerdo a ésto, podemos medir diferentes escalas de corriente con un sólo instrumento, seleccionando las diferentes resistencias mediante un suiche rotatorio.

En cada caso, la resistencia conectada absorve el mayor valor de corriente y por el instrumento o miliamperímetro siempre circulará la misma corriente, pero su aguja indicará un valor diferente de acuerdo a la escala.

Cómo se mide el voltaje

Para medir voltaje, se conecta una resistencia en serie con el instrumento de tal manera que el máximo voltaje de la escala por medir produzca en el circuito una corriente igual a la máxima corriente que se puede medir con ese instrumento.

De esta manera, podemos medir diferentes rangos de voltaje seleccionando una u otra resistencia por medio de un suiche selector.

En la figura 162A podemos observar el circuito de un miliamperímetro conectado para medir varias escalas de voltaje.

A mayor voltaje, mayor debe ser el valor de esta resistencia conectada en serie.

Tipos de multímetros

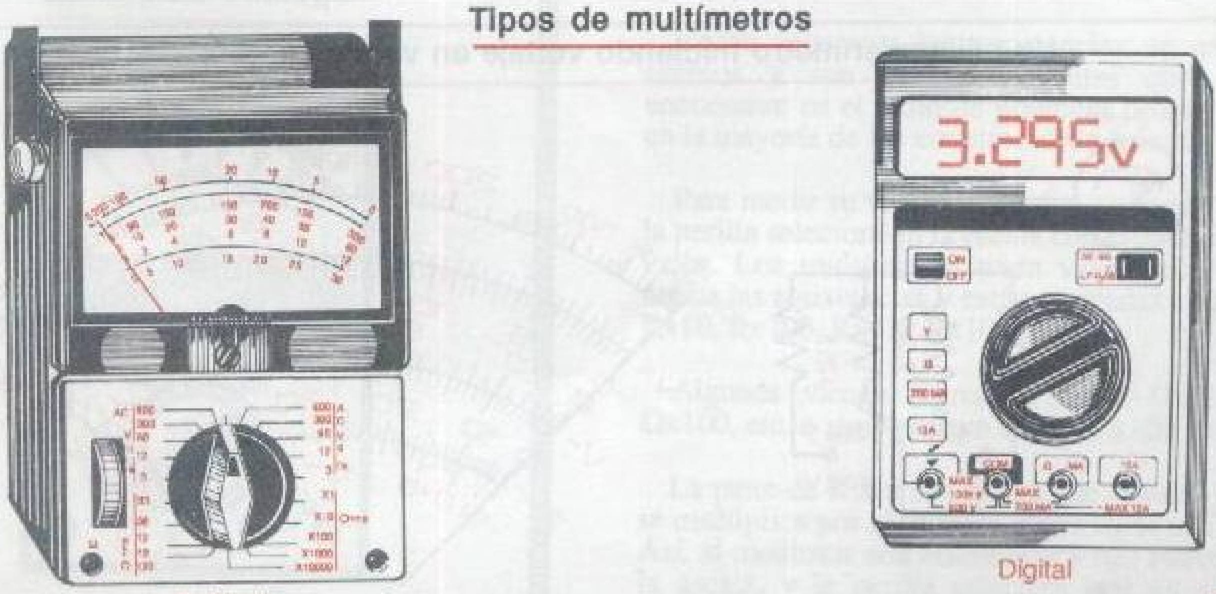

Análogo

Digital

Fig. 160

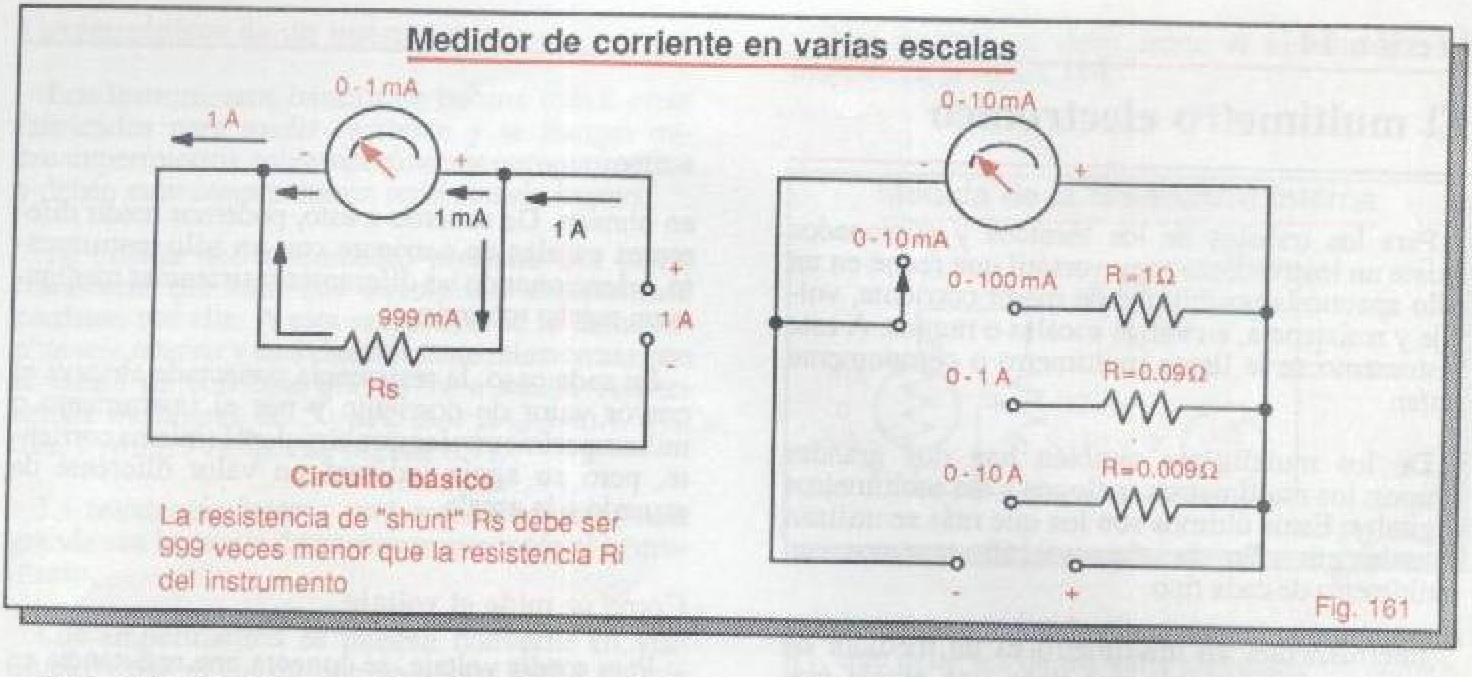
0-1 mA
1 A
-
+
1 mA
1 A
+
999 mA
1 A
-
Rs
0-10mA
-
+
0-10mA
0-100mA
R=1Ω
0-1 A
R=0.09Ω
0-10 A
R=0.009Ω
-
+

Medidor de corriente en varias escalas

Circuito básico

La resistencia de "shunt" Rs debe ser 999 veces menor que la resistencia Ri del instrumento

Fig. 161

Para cada una de las escalas seleccionadas, debe haber una escala marcada en el tablero del instrumento. (figura 162B).

Cómo se miden las resistencias

Para la medición de resistencias, que es la otra función de los multímetros, se utiliza una pila o batería interna para formar un circuito que al conectarse a la resistencia que se va a medir, varíe la corriente que circula por el instrumento básico.

Dependiendo del valor de esa resistencia, la deflexión o movimiento de la aguja marca un valor en ohmios en una parte de la escala especialmente señalada para la medida de las resistencias.

En la figura 163, se muestra el diagrama de uno de estos circuitos que se encuentra dentro de los multímetros y que recibe el nombre de óhmetro.

Para cambiar el instrumento como medidor de corriente a medidor de voltaje, o a medidor de los ohmios de las resistencias y, además, cada medida en varias escalas, éste tiene un suiche selector de varios circuitos y varias posiciones que nos selecciona las diferentes resistencias, que se conectan en cada caso al medidor básico de corriente.

Todos estos circuitos agregados al instrumento básico y a la perilla selectora de rangos forman el multímetro completo que se puede ver en uno de sus tantos tipos en la figura 164.

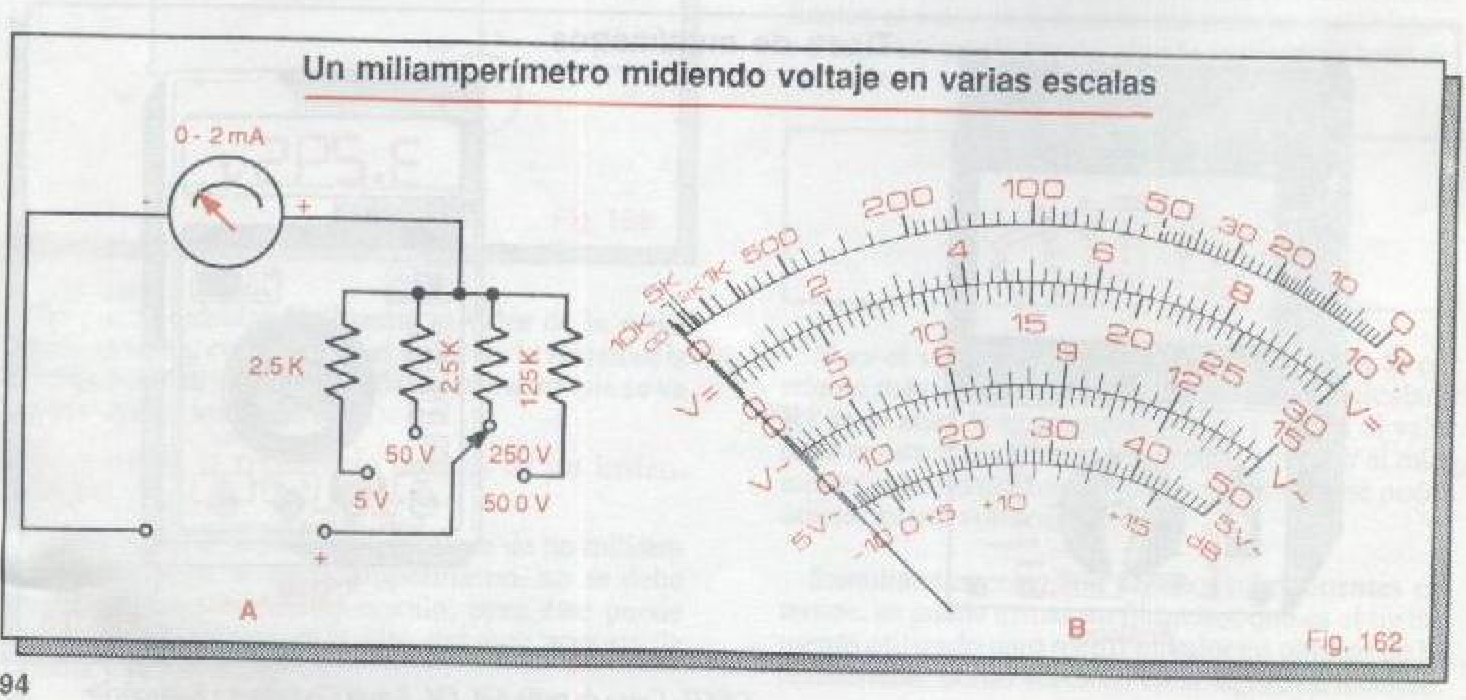
0-2mA
-
+
2.5K
2.5K
125K
50 V
250 V
5 V
500 V
+
A
200
100
50
30
20
10
0
500
1K
5K
10K
Ω
2
4
6
8
10
5
10
15
20
25
30
6
9
12
150
50
20
30
40
50
V=
V~
V=
V~
0
0
-10
0
+5
+10
+15
5V~
dB
B

Un miliamperímetro midiendo voltaje en varias escalas

Fig. 162

Circuito para medir ohmios

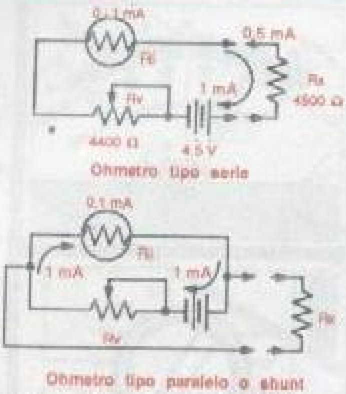

Fig. 163

Manejo del multímetro

El multímetro es el instrumento más utilizado en el estudio y práctica de la electrónica.

En este curso de radio lo vamos a utilizar mucho para probar los componentes y los circuitos que vamos ensamblando, hasta terminar nuestro radio de práctica.

Tales conocimientos serán muy útiles para la reparación y ensamble de todo tipo de aparatos electrónicos. Vamos a estudiar primero el manejo del multímetro análogo y después el manejo del multímetro digital.

En este momento del curso, le recomendamos que adquiera, si no lo tiene, un multímetro, que le será de gran utilidad con el fin de que vaya realizando las prácticas que se van mencionando.

Inicialmente, puede conseguir un multímetr sencillo y de bajo costo con el fin de estudiar s manejo y luego, cuando tenga buena experiencia puede conseguir uno más sofisticado y completo.

Cuando vaya a comprar el multímetro fíjese que tenga las siguientes características:

- Dos o tres escalas para medir ohmios.
- Posibilidad de medir voltajes en corriente continua y corriente alterna en varias escalas, 10V, 30V, 100V y 300V como mínimo.
- Capacidad de medir corriente en varias escalas, 10mA, 30 mA, 100 mA y 300 mA, o con valores similares.

Con estas posibilidades, usted ya tiene el instrumento necesario para seguir satisfactoriamente este curso y poder realizar las prácticas sobre instrumentos de medición.

Básicamente, el manejo del multímetro consiste en la ubicación correcta de la perilla selectora de rangos y funciones y en la instalación adecuada de las puntas de prueba en el componente, o en la parte del circuito que vamos a medir.

Inicialmente, estudiaremos la forma de medir las resistencias, después cómo se mide el voltaje en corriente continua y corriente alterna y por último, cómo se mide la intensidad de corriente en miliamperios o amperios.

Multímetro análogo

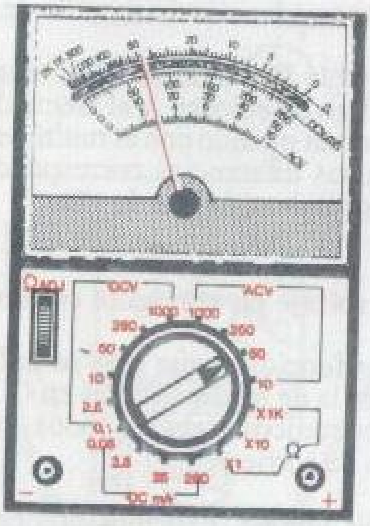

Fig. 164

Medición de resistencias

Como sabemos, las resistencias se miden en ohmios y son los componentes que más se encuentran en el radio de prácticas de este curso y en la mayoría de los circuitos electrónicos.

Para medir su valor en ohmios, debemos colocar la perilla selectora en la escala correspondiente a su valor. Los multímetros traen varias escalas para medir las resistencias y están marcadas como Rx1, Rx10, Rx100, Rx1K, Rx10K, etc.

Algunos vienen marcados como Ωx1, Ωx10, Ωx100, etc. o simplemente x1, x10, x100, etc.

La parte de R x n nos indica que el valor medido se multiplica por el número n que sigue a la escala. Así, si medimos una resistencia y nos marca 22 en la escala, y la perilla selectora está en el rango Rx1K, el valor de la resistencia es de 22KΩ, o 22.000 ohmios.

Si se va a medir una resistencia de muchos ohmios, se debe colocar la perilla en una escala alta como la de R x 10K. Si la aguja casi no se mueve, se debe pasar la perilla a la escala más baja siguiente, hasta que se pueda leer claramente el valor.

Para resistencias de bajo valor, se debe poner la perilla en la escala menor o sea R x 1 o Ω x 1.

Para medir una resistencia, lo primero que se debe hacer es ajustar la medida de *cero ohmios*, lo que se logra uniendo las dos puntas de prueba del multímetro y moviendo una perilla especialmente destinada a este propósito, colocándola en 0Ω, (figura 165).

Ajuste de cero del multímetro

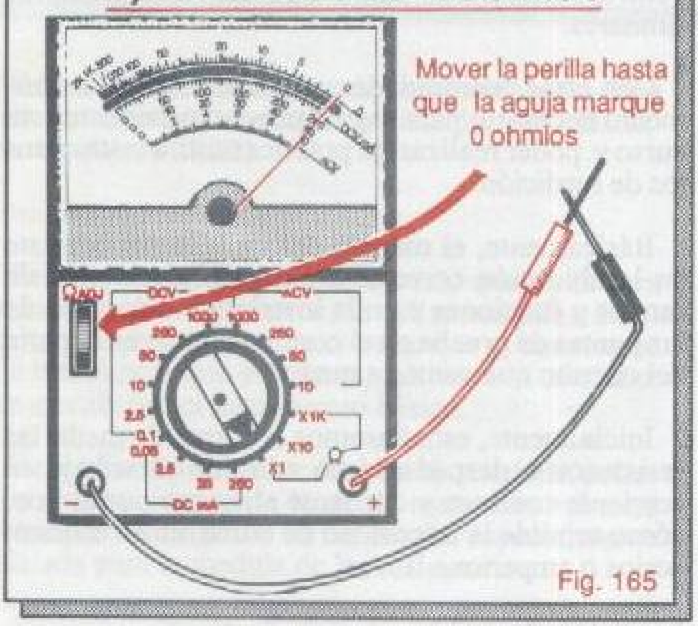

Fig. 165

Busque esta perilla de ajuste de cero que se puede encontrar en diferentes sitios del multímetro, según sea el modelo.

En algunos casos, está al lado derecho o izquierdo por uno de los bordes, o en otros está al frente del instrumento en la parte superior a la perilla selectora de rango y función.

Una vez que se hayan ajustado los 0 ohmios, conectamos firmemente las dos puntas de prueba del multímetro a los terminales de la resistencia por medir y leemos su valor en la escala, (figura 166).

Cuando vamos a medir una resistencia que está conectada en un circuito debemos estar seguros de que el aparato no tenga alimentación de voltaje.

Si la resistencia que vamos a medir está conectada, lo más recomendable es desconectar uno de sus terminales del circuito con el fin de que los demás componentes no alteren el valor de la medida, figura 167.

Medición de resistencias

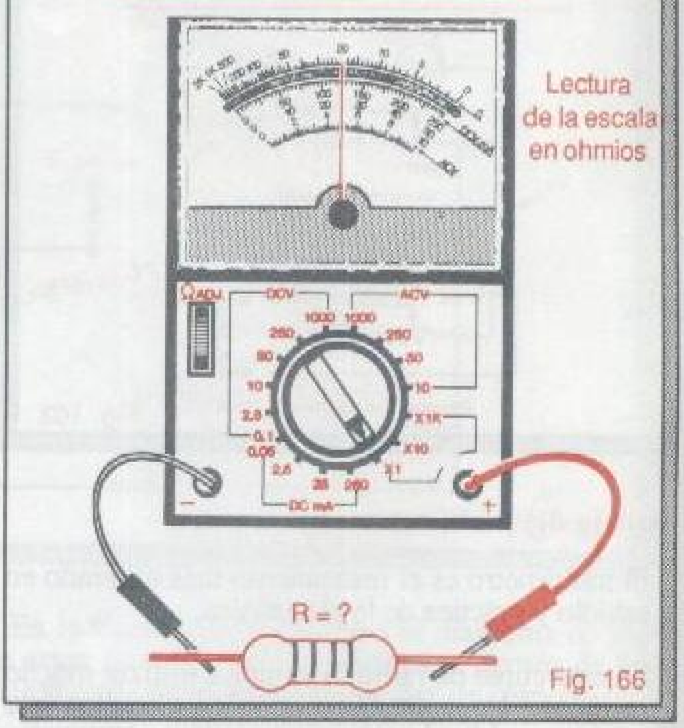

Fig. 166

Una precaución muy importante que debemos tener siempre en cuenta es la de no medir voltajes cuando el multímetro está en la escala de ohmios, pues el instrumento seguramente se quemará.

Este es uno de los daños más frecuentes que ocurren en estos aparatos y debemos fijarnos bien, al realizar una medida de voltaje, que la perilla selectora esté en la posición correcta.

Para familiarizarse con este proceso de medida, ensaye con las resistencias de su radio de práctica.

Haga una tabla anotando su valor teórico indicado por el código de colores, y al frente de cada una anote el valor real medido con el multímetro. Notará que hay algunas diferencias correspondientes a la tolerancia propia de cada una de ellas.

Medida de voltaje en corriente continua

Una de las funciones que más se utilizan con el multímetro es la de medir voltajes en CC ya que los circuitos electrónicos trabajan todos internamente con este tipo de corriente.

Para medir voltaje, debemos colocar la perilla selectora en un rango mayor al valor que vamos a medir. Por ejemplo, si vamos a medir un voltaje de

Medición de una resistencia conectada a un circuito

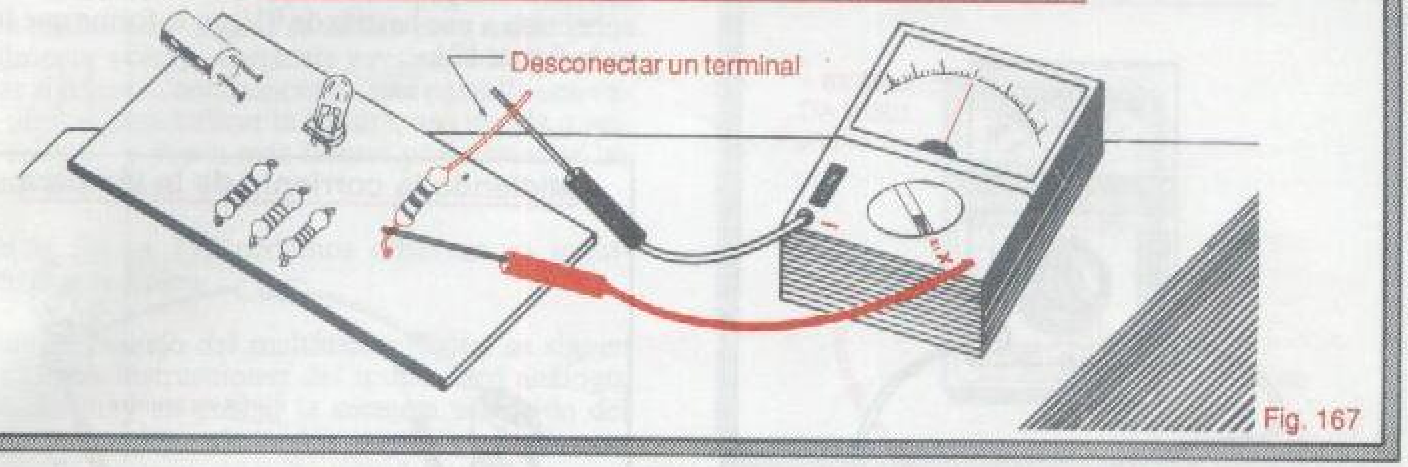

Fig. 167

9 voltios, la perilla selectora debe estar en la posición de 10 voltios CC.

Si no conocemos el voltaje, debemos empezar por la escala más alta e ir moviendo la perilla hacia un rango más bajo, hasta lograr una medida adecuada. La precisión de un multímetro análogo es mayor cuando la aguja está más cerca del extremo derecho de la escala.

Cuando se mide voltaje de CC, debemos tener en cuenta la polaridad en el circuito. La punta de prueba negativa (negra) del multímetro se conecta al terminal negativo del circuito, que generalmente se llama *tierra*. La punta de prueba positiva (roja) se debe conectar al circuito en su punta positiva, así como se muestra en la figura 168.

Medición de voltaje en CC

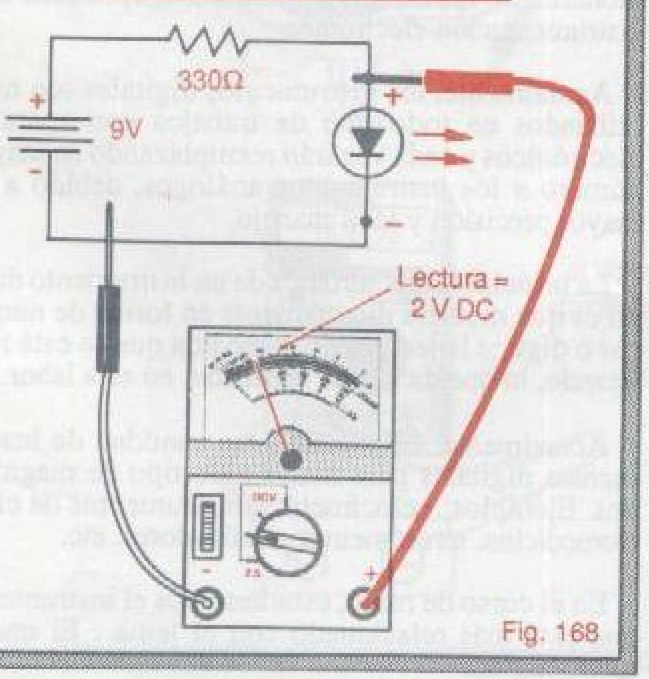

Fig. 168

Si en algún momento estamos midiendo un voltaje de CC y la aguja se desplaza rápidamente hacia la izquierda fuera de la escala, debemos invertir la posición de las puntas de prueba.

Medida de voltaje en corriente alterna

Para medir voltaje de CA como el de los tomacorrientes donde conectamos los aparatos eléctricos y electrónicos debemos colocar la perilla selectora en la posición para corriente alterna y en la escala correcta de acuerdo a su valor.

Por ejemplo, para medir los 110V que debe haber aproximadamente en un tomacorriente, colocamos la perilla en la escala de 250 voltios de CA.

Si conectamos el multímetro en una escala menor o en la función de ohmios o de corriente en miliamperios, éste probablemente se quemará y quedará inservible ya que su reparación es difícil debido a que en los multímetros se utilizan resistencias de precisión que no son fáciles de conseguir en el mercado corriente de los componentes electrónicos.

Para la medida de voltajes de CA, no interesa la posición o polaridad de las puntas de prueba cuando se conectan al circuito. (figura 169).

Medida de corriente en CC

Como un aparato electrónico está formado por varios circuitos y éstos a su vez por varios componentes, podemos medir la corriente que circula internamente en el aparato y sus circuitos con el fin de determinar si está funcionando bien.

Para medir corriente, primero debemos seleccionar la escala correcta con la perilla selectora, y luego conectar el multímetro *en serie* con el circuito o componente por medir. (figura 170).

Medición de voltaje en CA

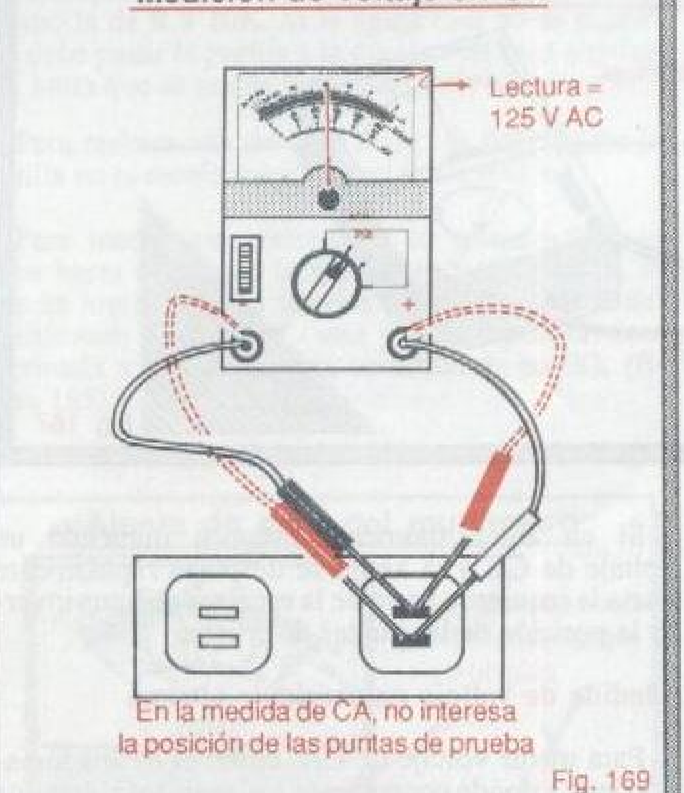
Lectura =
125 V AC
+
En la medida de CA, no interesa
la posición de las puntas de prueba

Fig. 169

Si la aguja se desplaza hacia la izquierda, debemos invertir la posición de las puntas de prueba.

Medición de corriente en CC

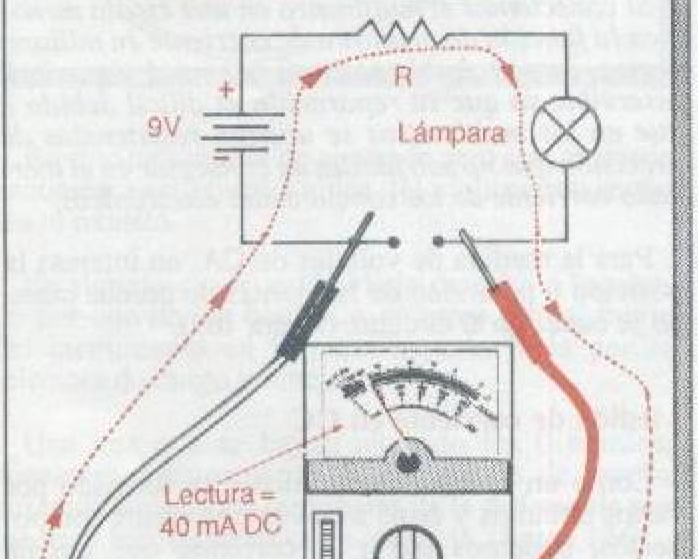

Fig. 170

Como práctica, podemos medir la corriente que consume una pequeña lámpara de corriente continua conectada a una batería de 9V, en la forma que indica la figura 171.

Midiendo la corriente de la lámpara

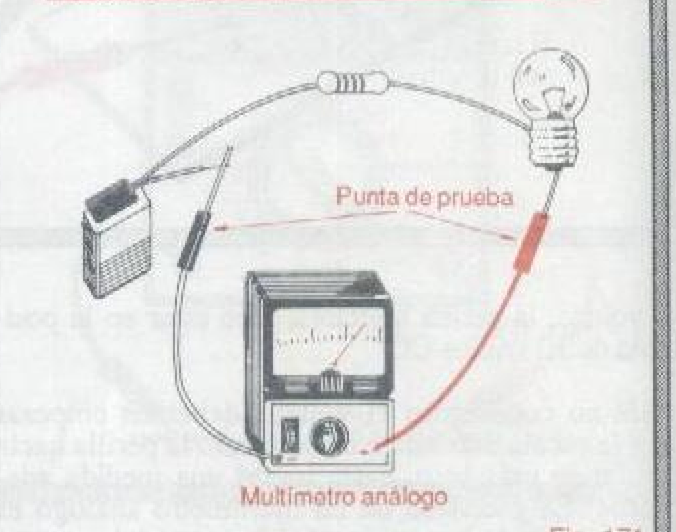

Fig. 171

Más adelante, iremos explicando detalladamente la aplicación del multímetro para medir el comportamiento de los diferentes circuitos y componentes del radio de práctica.

Instrumentos digitales

Los instrumentos digitales, son uno de los resultados más importantes del avance de la tecnología moderna de los circuitos integrados, aplicados a la instrumentación electrónica.

Actualmente, los instrumentos digitales son muy utilizados en todo tipo de trabajos con aparatos electrónicos y cada vez irán reemplazando en mayor número a los instrumentos análogos, debido a su mayor precisión y fácil manejo.

La principal característica de un instrumento digital es que muestra directamente en forma de números o dígitos la lectura de la medida que se está realizando, lo que da mayor precisión en esta labor.

Actualmente, existe una gran cantidad de instrumentos digitales para medir todo tipo de magnitudes. Ejemplos: velocímetros, instrumentos de electromedicina, termómetros, analizadores, etc.

En el curso de radio, estudiaremos el instrumento que está más relacionado con el tema : El multímetro digital.

El multímetro digital

Un multímetro digital se utiliza para medir principalmente voltaje, corriente y resistencia. Este tiene los siguientes componentes: una pantalla con varios dígitos para indicar la lectura, una perilla o suiche selector y dos o más bornes para conectar las puntas de prueba.

En la figura 172 podemos observar un multímetro digital típico.

Multímetro digital

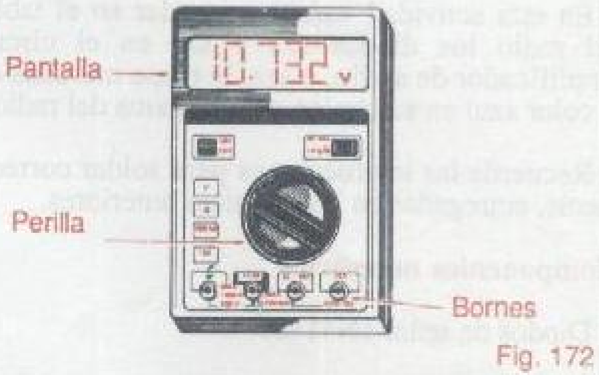
10.732 v
Pantalla
Perilla
Bornes

Fig. 172

Para el manejo del multímetro digital se siguen las mismas instrucciones del multímetro análogo, teniendo muy en cuenta la correcta selección del rango y función adecuados según el parámetro por medir, ya sea resistencia, voltaje o corriente.

En esto debemos tener un gran cuidado ya que los multímetros digitales son mucho más delicados, pues en su circuito interior tienen circuitos integrados y otros componentes muy sensibles.

Si usted posee un multímetro digital, practique cada uno de los tipos de medida que mencionamos en el multímetro análogo: resistencia, voltaje de CC y CA y corriente de CC.

Fíjese muy bien en la correcta selección de la escala para que no se vaya a quemar. En la figura 173 podemos observar cómo se ubica la perilla selectora para cada uno de estos casos.

Medidas electrónicas con el multímetro digital

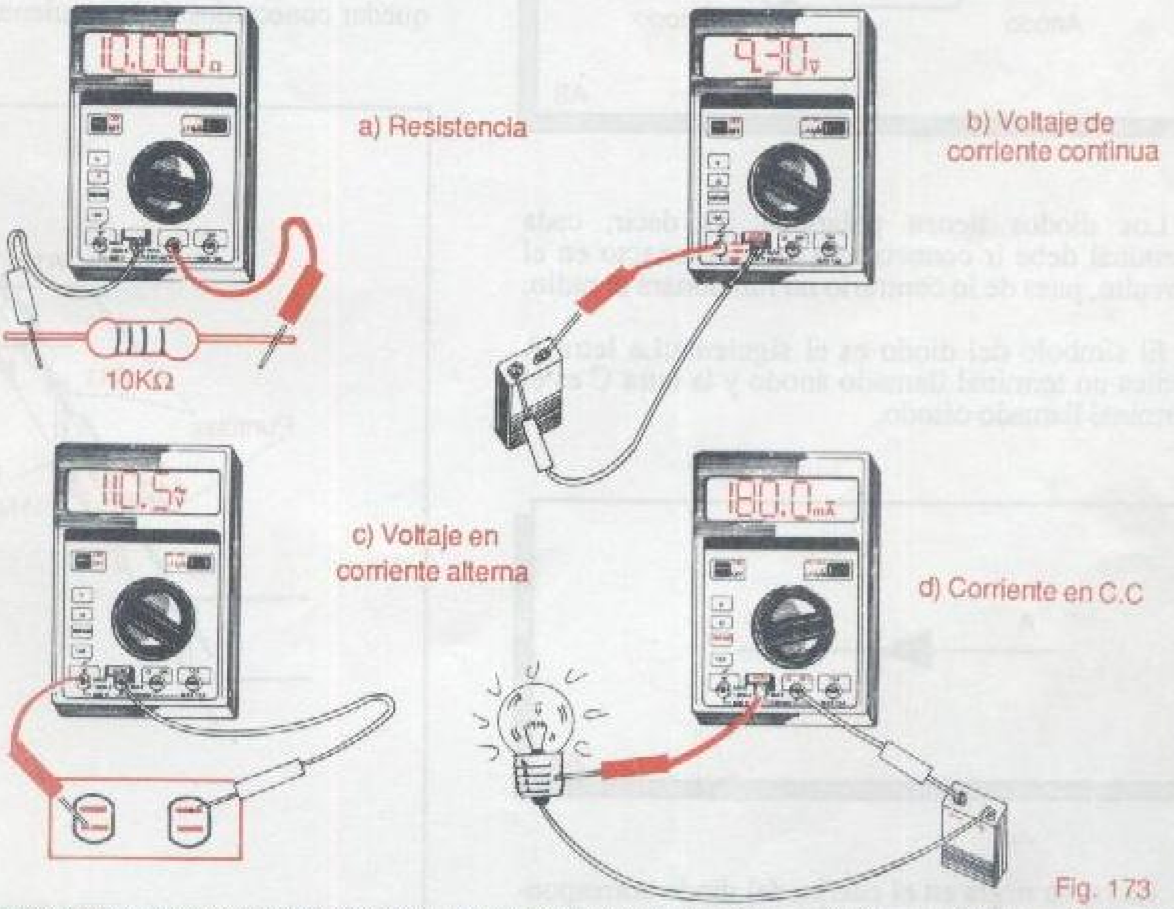
10.000 Ω
a) Resistencia
10KΩ
4.30 v
b) Voltaje de corriente continua
110.5 v
c) Voltaje en corriente alterna
180.0 mA
d) Corriente en C.C

Fig. 173

Actividad práctica Nº 7

En esta actividad vamos a instalar en el tablero del radio los diodos D2 y D3 en el circuito amplificador de audio, o sea, la etapa marcada con el color azul en su lámina del diagrama del radio.

Recuerde las instrucciones para soldar correctamente, entregadas en las prácticas anteriores.

Componentes necesarios

2 Diodos de señal 1N4148.

Los diodos, como se estudiará en la próxima lección, son unos componentes que dejan pasar la corriente en una sola dirección.

Paso 1: Identificación de los diodos.

Físicamente, el diodo 1N4148 tiene la siguiente forma:

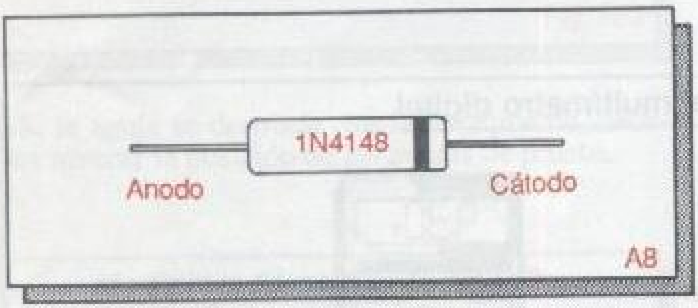

A8

Los diodos tienen polaridad, es decir, cada terminal debe ir conectado a un sitio exacto en el circuito, pues de lo contrario no funcionará el radio.

El símbolo del diodo es el siguiente:La letra A indica un terminal llamado ánodo y la letra C es el terminal llamado cátodo.

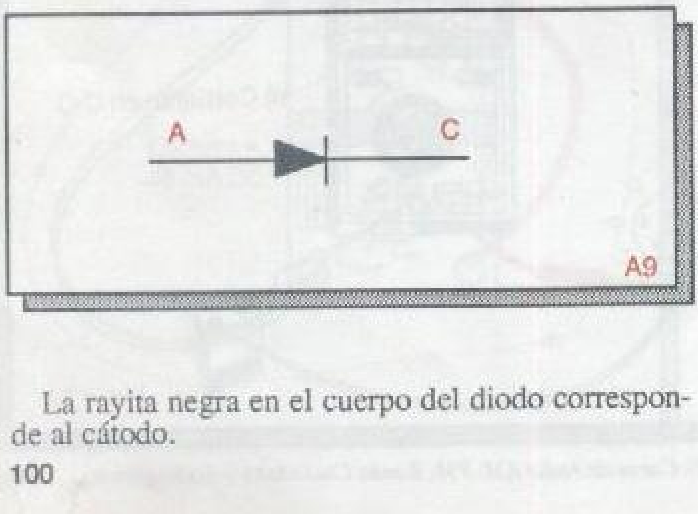

A9

La rayita negra en el cuerpo del diodo corresponde al cátodo.

Paso 2: Una vez identificado cada diodo y sus terminales, debe proceder a soldarlos en las puntillas correspondientes.

Instale y suelde uno de los diodos entre las puntillas A10 y A11. Este será el diodo D2. El cátodo del diodo, o sea el terminal más cercano a la rayita negra, debe quedar en la puntilla A11.

Paso 3: Instale y suelde el otro diodo, D3, entre las puntillas A11 y A12. El cátodo de este diodo debe ir conectado a la puntilla A12.

Precaución: Los diodos son unos componentes mucho más delicados que las resistencias y los condensadores, en cuanto a que se pueden dañar fácilmente si les aplicamos mucha temperatura al soldarlos.

Utilice para esta actividad un cautín pequeño, máximo de 25 W, y suelde rápidamente los terminales de los diodos. Para una mayor facilidad en el proceso de soldadura, las puntillas deben estar estañadas, es decir, deben tener un poco de soldadura en su superficie.

Al terminar esta actividad, los diodos deben quedar conectados de la siguiente manera:

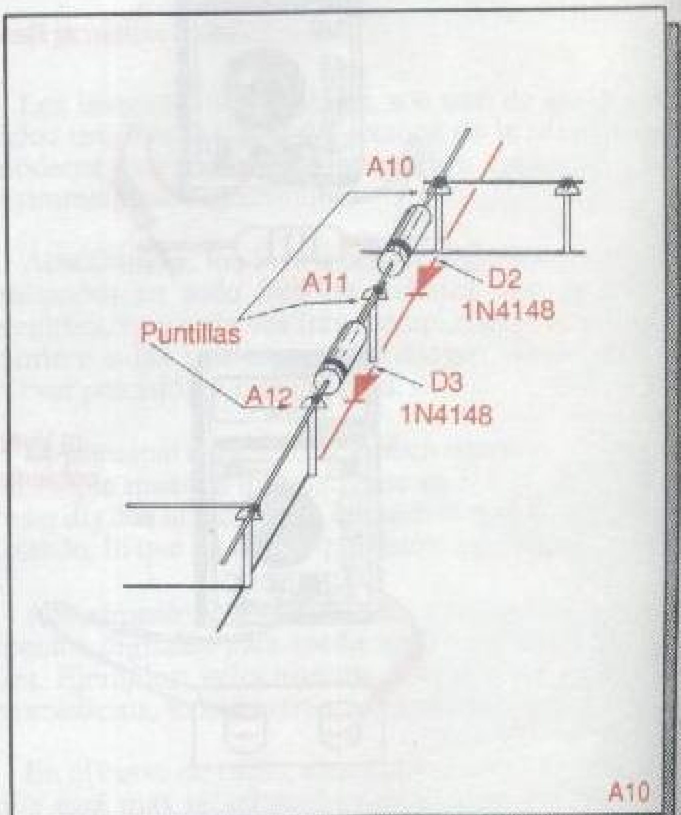

A10

Instrumentos para el ajuste y reparación de los aparatos de radio

Existen, además del multímetro, otros instrumentos electrónicos que se utilizan para el ajuste y la reparación de los aparatos de radio y comunicaciones en general.

Estos instrumentos permiten simular y visualizar, en una mesa de trabajo, las señales que se encuentran en estos aparatos.

Los principales instrumentos de este tipo y que vamos a conocer en este curso son: El generador de audio, el inyector de señales, el generador de radiofrecuencia, el seguidor o rastreador de señales y el osciloscopio.

El generador de audiofrecuencia

El generador de audiofrecuencia es un instrumento que genera o produce una señal de audio, es decir, que tiene una frecuencia audible, comprendida mas o menos entre 10 Hz y 30.000 Hz o 30 Khz.

Los generadores de audio se fabrican en muchos modelos pero todos tienen características comunes. Producen una señal con forma de onda seno y cuadrada como mínimo. Algunos modelos también generan señales del tipo triangular y pulsos.

En la figura 174 podemos observar dos tipos de generadores de audio, llamados también generadores de baja frecuencia.

Los generadores de audio se utilizan para aplicar una señal a un circuito amplificador con el fin de medir y observar su funcionamiento.

Este aparato debe tener, como mínimo, controles para variar la frecuencia en Hertz de la señal, para seleccionar la forma de onda y para regular su amplitud.

En la comunicación por radio, sabemos que una señal de audio o sonido, viaja en compañía de la señal de radiofrecuencia desde la antena del transmisor hasta la antena del receptor.

Internamente, en el aparato de radio, la señal de audio se separa, se amplifica y va al parlante.

El circuito que amplifica esta señal de un radio se llama el amplificador de audio.

Generadores de audio

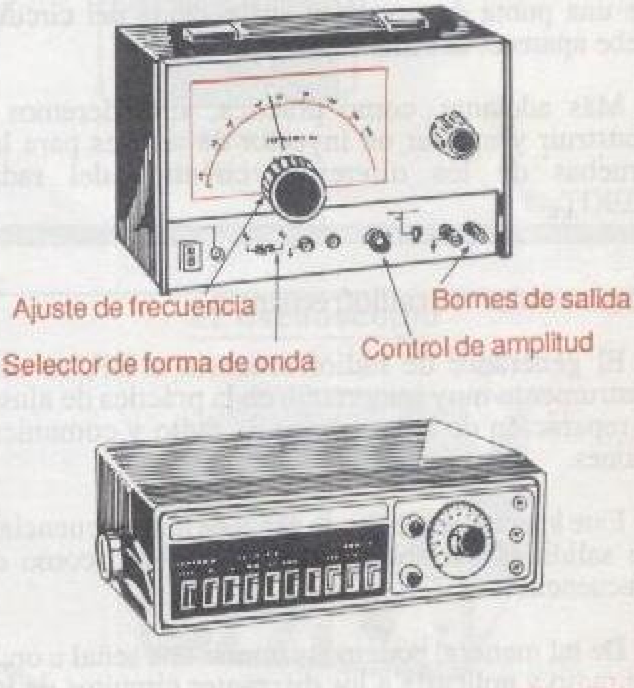

Fig. 174

En una próxima lección estudiaremos cómo se mide el nivel de amplificación de una etapa amplificadora de audio y cómo se observa la distorsión, utilizando el generador de audio en compañía de otro instrumento muy útil en electrónica, el osciloscopio.

El inyector de señales

Los generadores de audio son muy importantes para el trabajo en electrónica pero pueden ser costosos para el aficionado o estudiante.

Se pueden fabricar generadores de audio sencillos, con materiales comunes para ser utilizados en trabajos de fabricación y reparación de aparatos de radio en forma experimental.

Uno de los más sencillos, pero más eficientes instrumentos es el llamado *inyector de señales.*

Este es un circuito formado por pocos componentes (2 transistores, 4 resistencias y 3 condensadores) que se puede construir en un circuito impreso (kit CEKIT ref. K13) y alojar en una pequeña caja metálica o de plástico, alimentado por una pila o batería.

Este circuito produce una señal de onda cuadrada con una frecuencia aproximada de 1000 Hz en el rango audible .

Como su nombre lo indica, el inyector de señales *inyecta* su señal al circuito que deseemos por medio de una punta de prueba y en la salida del circuito debe aparecer la señal esperada.

Más adelante, como práctica, aprenderemos a construir y utilizar un inyector de señales para las pruebas de los diferentes circuitos del radio CEKIT.

El generador de radiofrecuencia

El generador de radiofrecuencia o RF, es otro instrumento muy importante en la práctica de ajuste y reparación de los aparatos de radio y comunicaciones.

Este aparato entrega una señal de alta frecuencia y su salida es variable, tanto en amplitud como en frecuencia.

De tal manera, podemos simular una señal u onda de radio y aplicarla a los diferentes circuitos de los aparatos, con el fin de determinar y lograr su correcto funcionamiento, (figura 175).

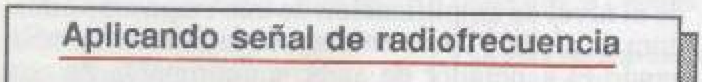

Fig. 175

Estos instrumentos también se fabrican en muchos modelos y calidades. En la figura 176 se muestra un generador de RF típico.

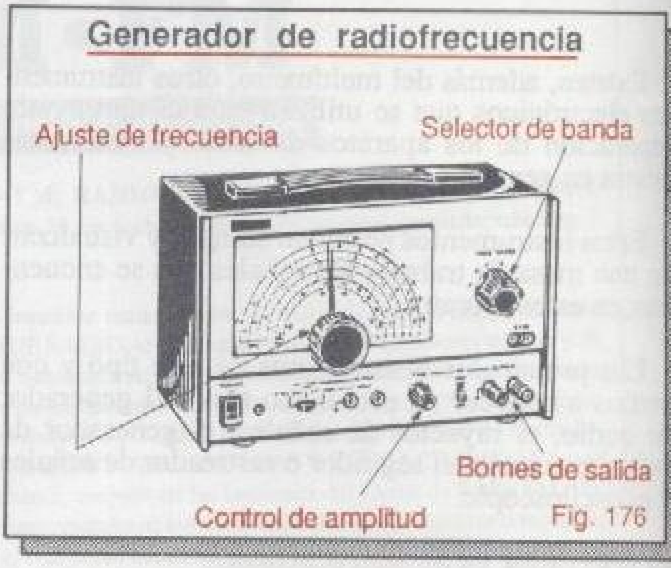

Fig. 176

Los principales controles de un generador de RF son: El selector de banda, que selecciona una banda o rango de frecuencias para la señal de salida, la perilla selectora de frecuencia dentro de la banda seleccionada y un control de amplitud que determina el tamaño de la señal que se va a aplicar al circuito.

Además de los controles, tiene dos bornes de salida y dos bornes para introducir la señal de un generador de audio y *modular* la señal de RF.

Veremos la aplicación práctica de este instrumento, en las lecciones correspondientes al ajuste y calibración de las etapas de radiofrecuencia, una vez las hayamos ensamblado en el radio CEKIT.

El seguidor o rastreador de señales

Este instrumento, también llamado analizador dinámico o *Signal Tracer*, en inglés, permite tomar la señal de audio presente en las diferentes etapas del radio lo que facilita la tarea de reparación de estos aparatos.

Básicamente, consiste en un amplificador de audio de poca potencia, un parlante, y una etapa detectora incorporada en una punta de prueba.

Trabajando simultáneamente con un generador de audio o con un inyector de señales, introduciendo su señal en la entrada de una etapa del radio y luego tomando la señal en la salida de esa misma etapa, con la punta de prueba, podemos verificar el buen funcionamiento de cada una de ellas.

Análisis de un radio con generador de audio y trazador de señales

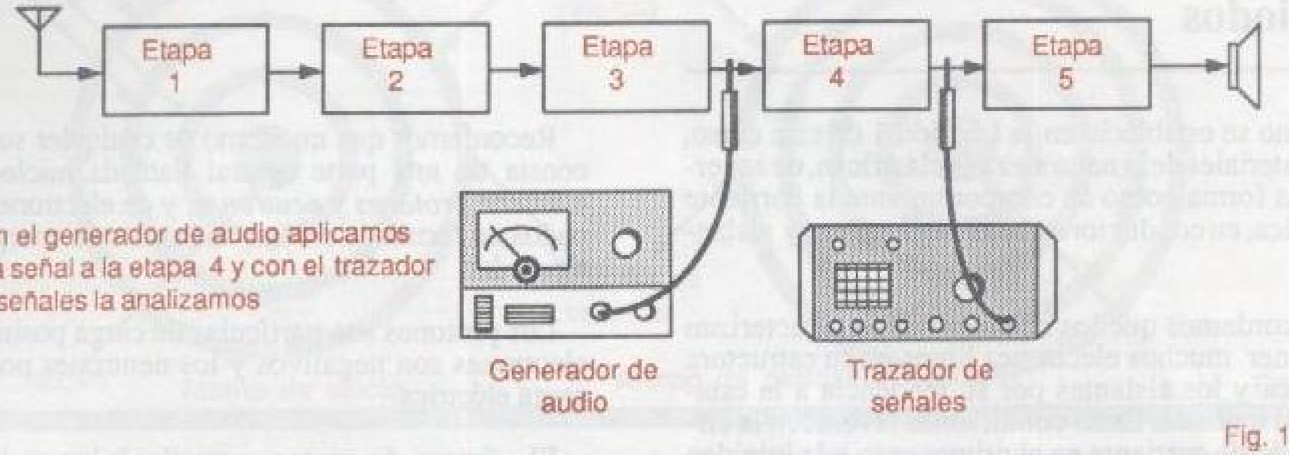

Con el generador de audio aplicamos una señal a la etapa 4 y con el trazador de señales la analizamos

Fig. 177

De esta manera, si un radio tiene 5 o 6 etapas, las podemos probar individualmente, encontrando fácilmente la falla en cualquiera de ellas.

En la figura 177 podemos observar la forma general de utilización de este instrumento.

El osciloscopio

El osciloscopio es uno de los instrumentos más importantes en el trabajo con aparatos de radio y comunicaciones y en toda la electrónica en general.

Por medio del osciloscopio podemos visualizar en una pantalla, la forma de onda, la amplitud y la frecuencia de una señal de corriente alterna. Esto hace que este instrumento nos suministre más información que cualquier otro.

También podemos medir con él, voltaje y corriente en corriente continua.

El osciloscopio utiliza una pantalla similar en su principio de funcionamiento, a la de un televisor.

Por medio de circuitos amplificadores y manejadores de rayos de electrones, las señales se pueden presentar en forma visible.

Para el técnico, el aficionado o el ingeniero, esto es muy importante pues permite tener una imagen real de las señales y voltajes dentro de un aparato.

Existen muchos tipos de osciloscopios dependiendo de sus características y calidad.

En la figura 178 podemos observar un osciloscopio común de los que se utilizan para los trabajos de ajuste y reparación en los aparatos de radio.

El osciloscopio

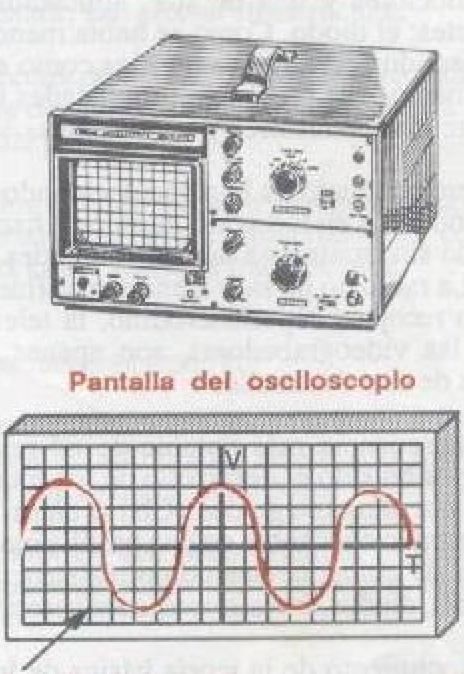

Visualización de la señal

Fig. 178

En próximas lecciones aprenderemos a utilizar el osciloscopio para la calibración o ajuste y análisis de las diferentes etapas del radio de prácticas de este curso.

Con el generador de audio y el osciloscopio, podemos probar el buen funcionamiento de una etapa en cuanto a su nivel de amplificación y a la distorsión que se produce a la señal original.

El generador de audio se conecta a la entrada del circuito que se va a probar y luego observamos en su salida, la forma y amplitud de la señal que se obtiene.

Teoría de Semiconductores y Diodos

Como se estableció en la Lección 1 de este curso, los materiales de la naturaleza se clasifican, de acuerdo a la forma como se comportan ante la corriente eléctrica, en conductores, semiconductores y aislantes.

Recordemos que los conductores se caracterizan por tener muchos electrones libres en su estructura atómica y los aislantes por su tendencia a la estabilidad química. Estas condiciones favorecen la circulación de corriente en el primer caso y la impiden en el otro.

En esta lección, vamos a estudiar los materiales semiconductores y una de sus aplicaciones más importantes: el diodo. Como se había mencionado, los semiconductores son sustancias como el silicio y el germanio, que presentan propiedades intermedias entre los conductores y los aislantes.

Los semiconductores han revolucionado la electrónica moderna en todos sus campos y han abierto un mundo sin fronteras a sus posibilidades de aplicación. La radio no ha sido ajena a su influencia. El moderno receptor superheterodino, la televisión a color y las videograbadoras, son apenas algunos ejemplos de este desarrollo.

La consecuencia más notable de tal revolución tecnológica ha sido el desarrollo de los transistores y de los circuitos integrados (CIs) o *chips*. Un circuito integrado concentra, en una pastilla muy delgada, miles de componentes como resistencias, condensadores, diodos, transistores, conductores, etc.

El conocimiento de la teoría básica de los semiconductores y del diodo, es indispensable para comprender cómo trabajan no solamente los transistores y los circuitos integrados sino también los FETs, los SCRs, los LEDs y otros dispositivos de la llamada *generación del silicio*.

Repaso de la teoría atómica

Como sabemos, los materiales están formados por elementos y los elementos por átomos. El átomo es la parte más pequeña e indivisible de la materia, que conserva las propiedades originales de un material.

Si dividimos el átomo, encontramos las partículas básicas.

Recordemos que un átomo de cualquier sustancia consta de una parte central llamada *núcleo* que contiene *protones y neutrones*, y de electrones agrupados en forma de *órbitas* o bandas de energía a su alrededor.

Los protones son partículas de carga positiva, los electrones son negativos y los neutrones no tienen carga eléctrica.

El número de protones en el núcleo es igual al número de electrones distribuidos en las órbitas. Esto determina el *número atómico* de un material. Un átomo de cobre, por ejemplo, difiere de uno de hierro en su número atómico. *El número atómico es el que establece la diferencia entre un elemento y otro.*

Los electrones están distribuidos en las órbitas, obedeciendo a ciertas leyes físicas y de acuerdo a su cantidad. Los electrones de la órbita más externa reciben el nombre de *electrones de valencia* o electrones libres y son muy importantes para la electrónica. El máximo número posible de electrones en la banda de valencia es 8.

El paso de una corriente eléctrica a través de un material depende de la movilidad de sus electrones de valencia. Entre más alejados estén estos electrones del núcleo, más fácil es producir en ese material una corriente eléctrica.

Además, el número de electrones de valencia determina el *grupo* al que pertenece un elemento en la tabla periódica.Un grupo en la tabla periódica asocia una serie de elementos con características y propiedades químicas afines (alcalinos, metales, gases, etc).

Los elementos del grupo 4, por ejemplo, tienen 4 electrones en su última capa. A este grupo pertenecen, entre otros, el carbono, el silicio y el germanio. De estos dos últimos, llamados los *semiconductores,* nos ocuparemos ahora.

Estructura atómica de los semiconductores

En la figura 179 se muestra la estructura atómica de un átomo de silicio y uno de germanio. El de silicio posee 14 electrones distribuidos en 3 órbitas y el de germanio 32, repartidos en 4 órbitas.

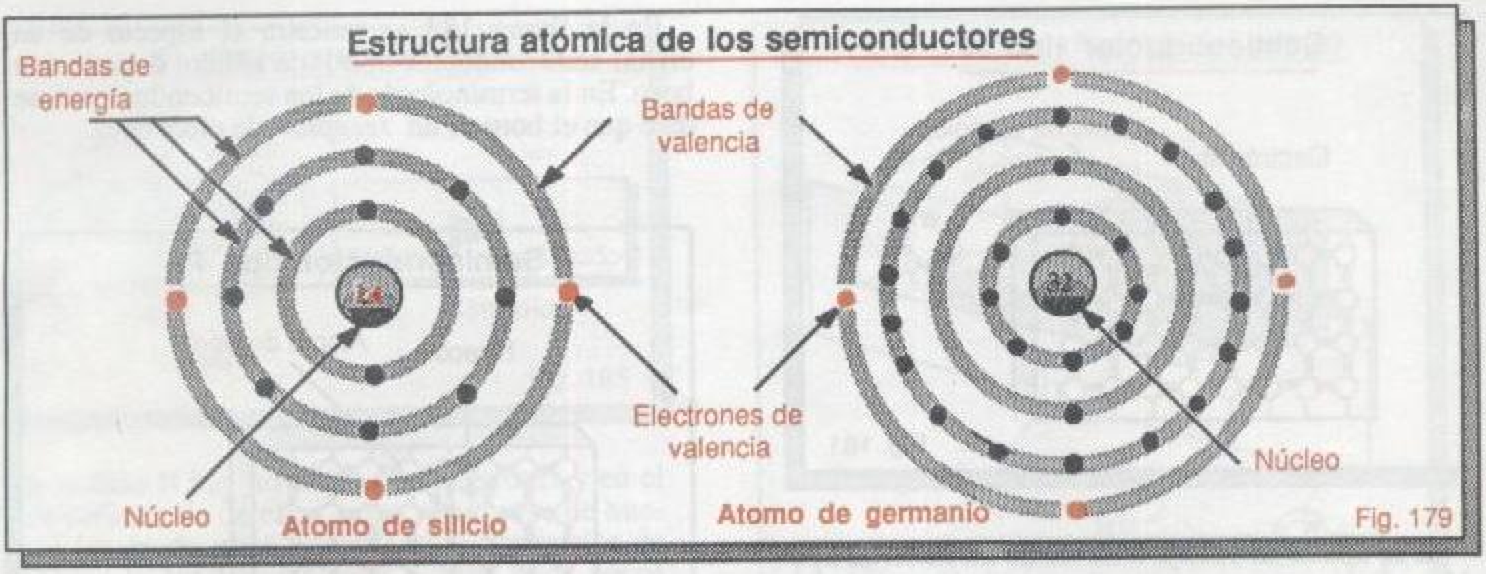
Estructura atómica de los semiconductores
Bandas de energía
Bandas de valencia
14
Electrones de valencia
Núcleo
Atomo de silicio
32
Atomo de germanio
Núcleo

Fig. 179

Ambos átomos comparten en común el hecho de tener cada uno 4 electrones de valencia.

Por ahora, sólo nos interesan el núcleo y los electrones de valencia. En un semiconductor puro, los átomos están perfectamente alineados, formando lo que se denomina una red cristalina o *cristal*. En la figura 180 se muestra el aspecto parcial de un cristal de silicio.

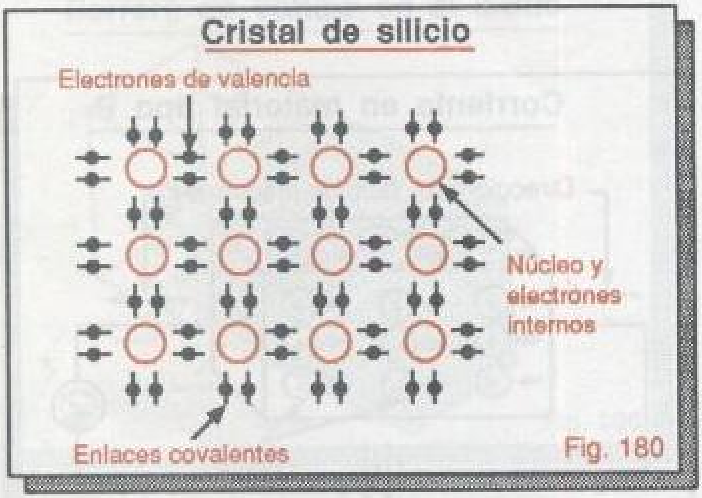
Cristal de silicio
Electrones de valencia
Núcleo y electrones internos
Enlaces covalentes

Fig. 180

En un cristal de silicio, cada átomo comparte sus 4 electrones de valencia con los átomos vecinos mediante *enlaces covalentes*. Estos enlaces son tan fuertes que es prácticamente imposible arrancar un electrón de valencia de su sitio, excepto con la aplicación de muy alta energía.

Por esta razón, un cristal semiconductor puro no favorece la circulación de corriente. Desde el punto de vista eléctrico, se comporta como un simple aislante, sin mayor utilidad práctica. Para que se convierta en conductor debe *doparse,* agregándole pequeñas cantidades de impurezas.

Las impurezas son átomos de otro material como el arsénico y el boro que, mediante el proceso de dopado, convierten un cristal semiconductor puro en un conductor, sin afectar su estructura.

De acuerdo al tipo de impureza inyectada, surgen dos tipos de semiconductores: el material tipo N y el material tipo P.

De ellos nos ocuparemos enseguida. La unión de un material tipo N con un material tipo P constituye un diodo, como veremos más adelante.

Semiconductores tipo N

Los semiconductores tipo N se obtienen inyectando átomos de antimonio, arsénico o fósforo a un cristal de silicio o germanio puro.

Estos átomos son pentavalentes, es decir, tienen 5 electrones de valencia.Su presencia en el cristal crea un exceso de electrones.

Al enlazarse un átomo de arsénico con uno de silicio, por ejemplo, queda sobrando un electrón en la estructura cristalina.

Este electrón, al no estar ligado a ningun átomo, es libre de moverse a través del cristal y, por lo tanto, es un portador potencial de corriente.

En la figura 181 se muestra el aspecto de un cristal semiconductor tipo N de silicio, dopado con arsénico.

En la terminología de los semiconductores, se dice que el arsénico es un *dador*.

Los *portadores* de corriente en un material tipo N son los *electrones*.

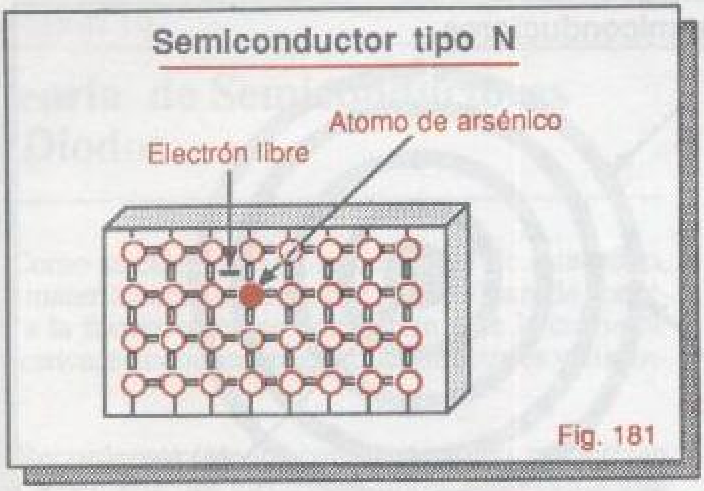
Semiconductor tipo N
Atomo de arsénico
Electrón libre
−

Fig. 181

Si se aplica un voltaje a través de un material tipo N, como se muestra en la figura 182, el resultado es un flujo de corriente a través del cristal. Los electrones libres son repelidos por el polo negativo de la batería y atraídos por el polo positivo de la misma.

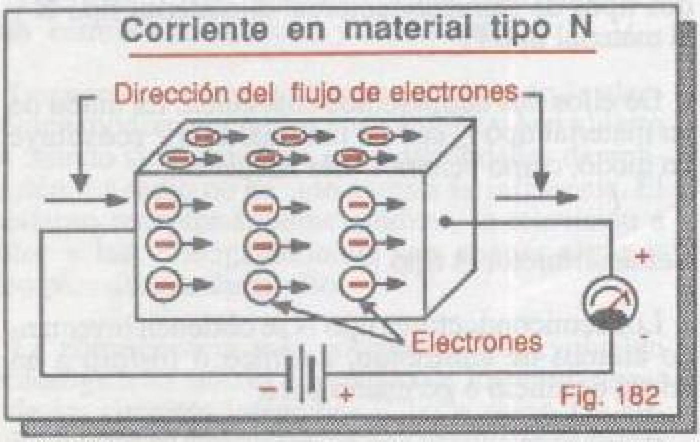
Corriente en material tipo N
Dirección del flujo de electrones
Electrones
+
−
−
+

Fig. 182

Recuerde: *los portadores de corriente en un material tipo N son los electrones libres, los cuales tienen carga negativa.*

Semiconductores tipo P

Los semiconductores tipo P se obtienen inyectando átomos de boro, galio o indio a un cristal de silicio o germanio puro.

Estos átomos son trivalentes, es decir, tienen 3 electrones de valencia. Su presencia en el cristal crea una deficiencia de electrones.

Al enlazarse un átomo de boro con uno de silicio, por ejemplo, queda faltando un electrón en la estructura cristalina. A ese espacio libre se le denomina *hueco*. Los huecos se comportan como cargas positivas porque atraen electrones y, por lo tanto, son portadores potenciales de corriente.

En la figura 183 se muestra el aspecto de un cristal semiconductor tipo P de silicio, dopado con boro. En la terminología de los semiconductores, se dice que el boro es un *receptor* de electrones.

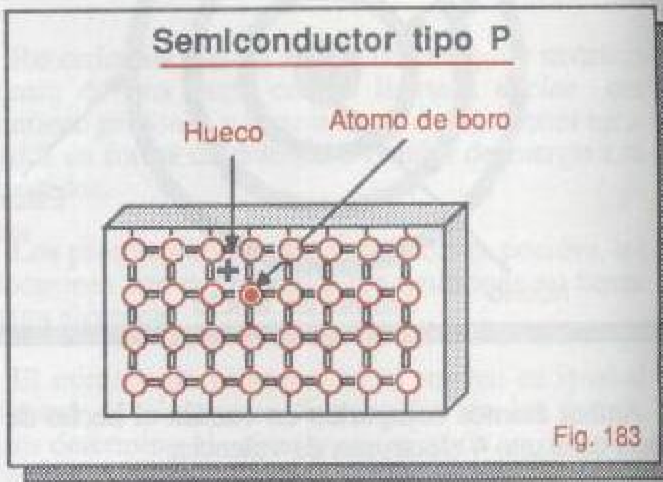
Semiconductor tipo P
Hueco
Atomo de boro
+

Fig. 183

Si se aplica un voltaje a través de un material tipo P, como se muestra en la figura 184, el resultado es un flujo de corriente a través del cristal. Los huecos son repelidos por el polo positivo de la batería y atraídos por el polo negativo. Esto equivale a una circulación de electrones en dirección contraria a la dirrección del material tipo N.

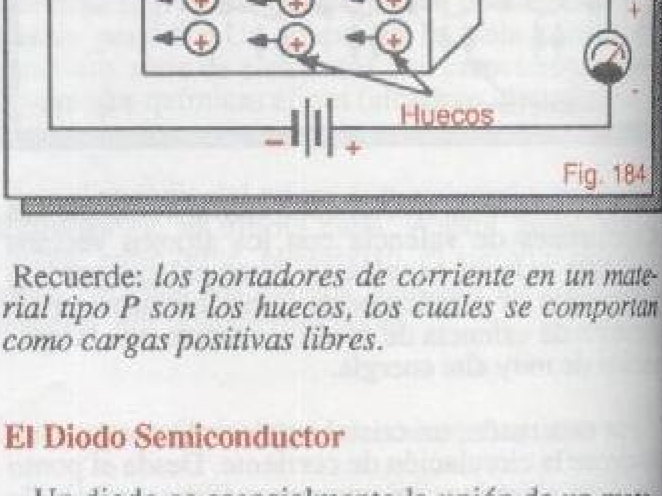
Corriente en material tipo P
Dirección del flujo de electrones
Huecos
+
-
−
+

Fig. 184

Recuerde: *los portadores de corriente en un material tipo P son los huecos, los cuales se comportan como cargas positivas libres.*

El Diodo Semiconductor

Un diodo es esencialmente la unión de un material tipo N con un material tipo P, como se muestra en la figura 185. El lado N recibe el nombre de *cátodo* y el lado P el de *ánodo*.

El diodo o unión PN

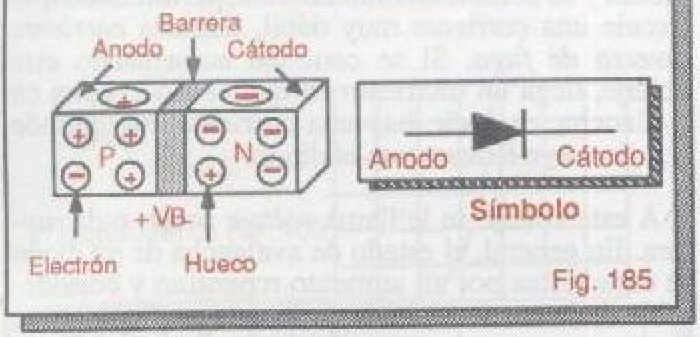

Fig. 185

En el lado N hay un exceso de electrones y en el P una deficiencia de ellos, o sea, un exceso de huecos. Además de estos *portadores mayoritarios* de corriente, en los lados N y P existen unos pocos *portadores minoritarios*, representados por algunos huecos en N y algunos electrones libres en P.

Los electrones del lado N tienden a pasar hacia el lado P, atraídos por los huecos o viceversa. Sin embargo, no pueden hacerlo porque en la unión o juntura de los dos materiales se forma una barrera eléctrica de voltaje que impide el paso de los portadores mayoritarios de corriente. (Figura 186).

Barrera de voltaje en el diodo

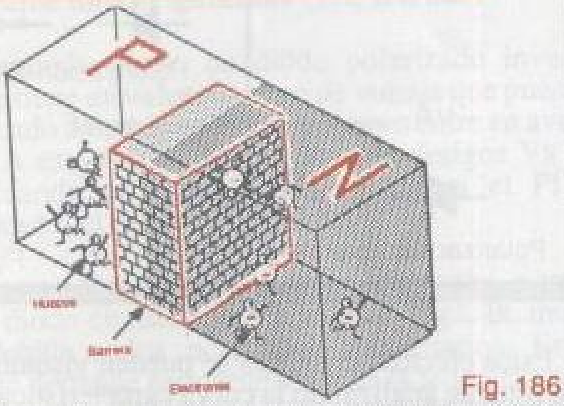

Fig. 186

A esta barrera de potencial la llamaremos unión o juntura PN y presenta propiedades muy interesantes para la corriente eléctrica.

Polarización directa de un diodo

Para que circulen los portadores mayoritarios de corriente de un lado a otro y pueda establecerse un flujo de corriente a través de la unión PN, la barrera de potencial debe ser eliminada.

La forma de conseguirlo es aplicando un voltaje externo de polarización, como se muestra en la figura 187.

Polarización directa del diodo

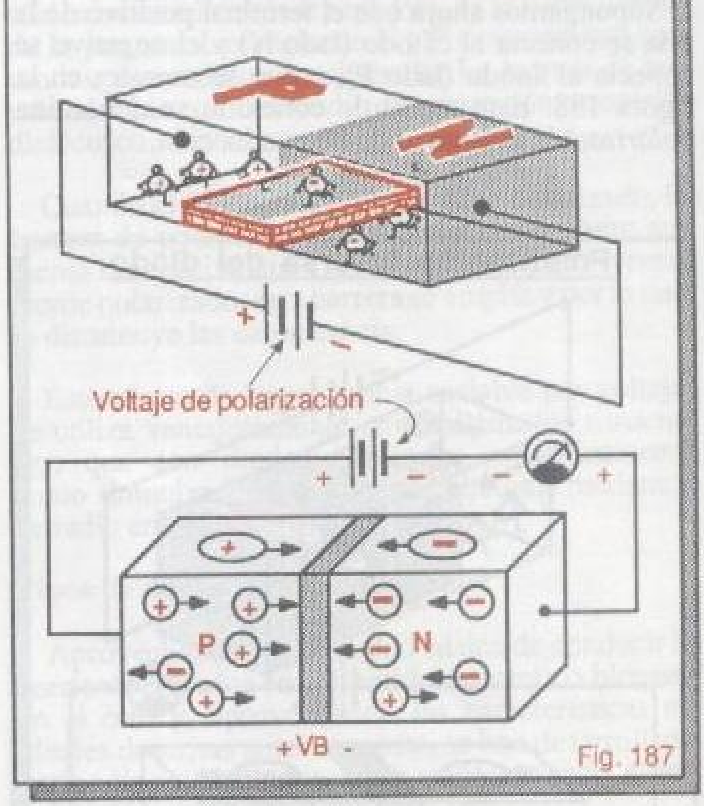

Fig. 187

En este modo de conexión, llamado *polarización directa*, el positivo de la fuente se conecta al electrodo del lado P o *ánodo* y el negativo al del lado N o *cátodo*. Conectado de este modo, el diodo conduce la corriente eléctrica. Por qué ?. Veámoslo:

En un diodo polarizado directamente, los electrones libres del material N son rechazados por el terminal negativo de la batería y emigran en dirección de la juntura.

Lo mismo sucede con los huecos del material P respecto al terminal positivo de la batería.

Como consecuencia de lo anterior, en la juntura PN se presenta una fuerte concentración de portadores de corriente.

Bajo la influencia del voltaje de la batería, los electrones y huecos atraviesan la barrera y se combinan mutuamente.

Por cada combinación de un electrón y un hueco, penetra un electrón por el terminal negativo y sale otro por el terminal positivo.

De este modo, fluye continuamente una corriente eléctrica a través del circuito externo.

. Podemos decir, entonces, que un diodo polarizado directamente actúa como un conductor, porque ofrece muy poca resistencia al paso de la corriente.

Polarización inversa de un diodo

Supongamos ahora que el terminal positivo de la pila se conecta al cátodo (lado N) y el negativo se conecta al ánodo (lado P), como se muestra en la figura 188. Este modo de conexión se denomina *polarización inversa* o de no conducción.

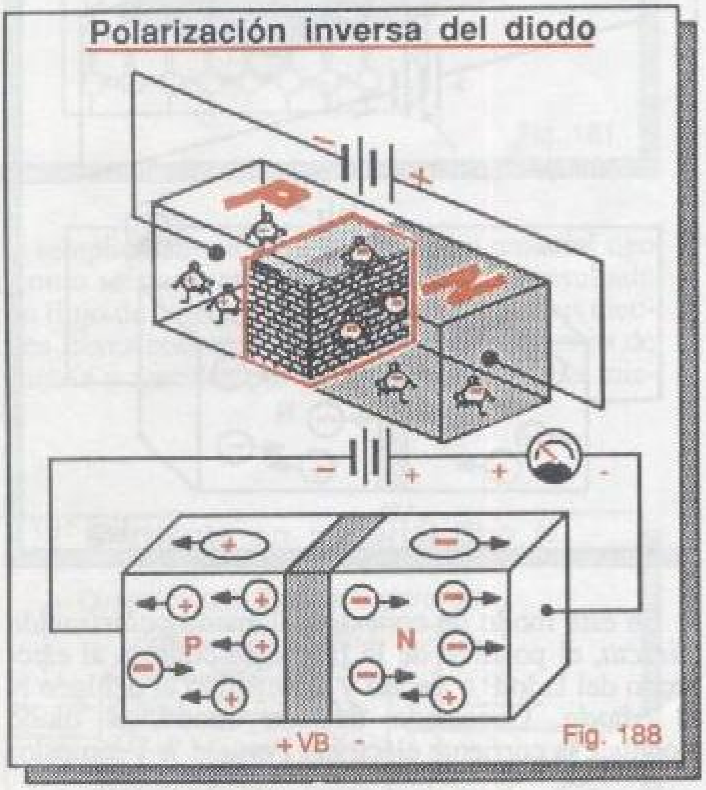

Fig. 188

En un diodo polarizado inversamente, los electrones libres del material N son atraídos por el terminal positivo, alejándolos de la juntura. Lo mismo sucede con los huecos de P respecto al terminal negativo.

Como consecuencia de lo anterior, en la juntura PN se presenta una drástica ausencia de portadores de corriente.

Bajo la influencia del voltaje de la batería, los electrones y huecos no pueden atravesar la barrera y, por lo tanto, no hay circulación de corriente a través del diodo.

Podemos decir, entonces, que un diodo polarizado inversamente actúa como un aislante, porque ofrece muy alta resistencia al paso de la corriente.

Recordemos que, además de portadores mayoritarios de corriente (electrones en N y huecos en P), en un diodo existen también unos pocos portadores minoritarios (huecos en N y electrones en P). Veamos los efectos de la presencia de estos portadores minoritarios en el cristal.

Bajo la influencia del voltaje inverso, los portadores minoritarios se concentran en la juntura, la atraviesan y se combinan mutuamente, permitiendo que circule una corriente muy débil, llamada *corriente inversa de fuga*. Si se continúa aumentando este voltaje, llega un momento en que el diodo entra en avalancha, es decir, hay una corriente muy grande que destruye físicamente el diodo.

A este voltaje se le llama *voltaje zener* o de ruptura. En general, el estado de avalancha de un diodo se caracteriza por un aumento repentino y considerable de la corriente inversa. Si esta corriente no se limita, se rompe la estructura cristalina y el diodo se destruye. En las características de fabricación de cada tipo de diodo se especifica cuál es el voltaje inverso máximo que soporta ese diodo sin dañarse.

Características de los diodos reales

Idealmente, un diodo debería comportarse como un interruptor perfecto, actuando como un cortocircuito en polarización directa y como un circuito abierto en polarización inversa (figura 189). Los diodos reales, sin embargo, exhiben ciertas características no acordes con esta idealización.

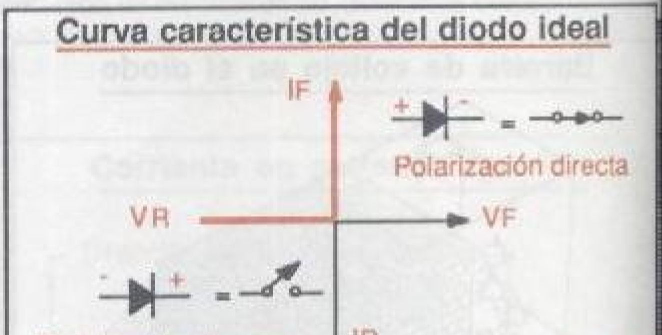

Fig. 189

Estos efectos no ideales se pueden visualizar más fácilmente analizando la curva característica real de un diodo, mostrada en la figura 190. Esta curva muestra la forma como se comportan la corriente y el voltaje a través de este componente.

En primer lugar, un diodo polarizado directamente necesita de un valor mínimo de voltaje para empezar a conducir. Este voltaje se designa como VF y es igual al potencial de barrera de la juntura PN. Los valores típicos de VF son 0.3 V para diodos de germanio y 0.6 a 0.7 V para diodos de silicio.

Esta caída directa de voltaje aparece siempre entre los terminales del diodo durante su operación normal, independientemente del voltaje externo aplicado.

Curva característica del diodo real

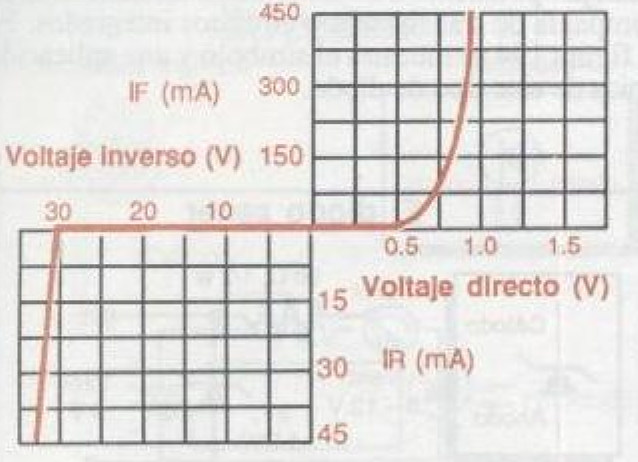

Fig. 190

Como consecuencia de la caída, al circular una corriente a través del dispositivo, éste disipa potencia en forma de calor.

Si tal disipación de potencia es excesiva, el diodo puede destruirse por sobrecalentamiento. Por esta razón, los fabricantes especifican siempre la máxima corriente directa admisible (1A, 2A, etc.).

En segundo lugar, un diodo polarizado inversamente tiene un valor máximo de voltaje que puede ser aplicado antes de que el dispositivo entre en avalancha. A este voltaje de ruptura se le designa VR o PIV. Cuando el voltaje aplicado supera el PIV especificado, el diodo se destruye.

Mientras no se llegue al estado avalancha, a través del diodo circula una corriente de fuga, IR, muy débil, debida a los portadores minoritarios. Esta corriente depende primariamente de la temperatura. En diodos de silicio, IR es del orden de los nanoamperios y se duplica cada 10ºC.

Cuando el diodo entra en avalancha, la corriente inversa crece súbitamente y puede ser fatal para el dispositivo. Si se continúa aumentando el voltaje inverso y se provee un medio de limitación de esta corriente, en los extremos del diodo aparece permanentemente un voltaje igual al de ruptura.

Los diodos comunes, por lo general, nunca se trabajan en la región de avalancha. Existen, sin embargo, diodos fabricados específicamente para este propósito. Son los llamados *diodos zener*, muy utilizados como reguladores de voltaje de corriente continua de baja potencia.

Otro fenómeno que se presenta en los diodos reales, y que no aparece explícito en la curva característica mostrada en la figura 190, es la capacitancia de la juntura. Las regiones N y P se comportan como las placas de un condensador. La barrera de potencial actúa como el medio de separación o material dieléctrico.

Cuando el diodo está directamente polarizado, la barrera de potencial se estrecha y por lo tanto aumenta la capacitancia. Cuando el diodo está inversamente polarizado, esta barrera se amplía y por lo tanto disminuye las capacitancia.

Este efecto de 'capacitancia variable por voltaje' se utiliza ventajosamente en los llamados *varactores*, que son diodos diseñados específicamente como sintonizadores en algunos circuitos modernos de radio en FM.

Tipos de diodos y sus aplicaciones

Aprovechando la propiedad básica de conducir la corriente eléctrica en un sentido mientras la bloquea en el otro y, aprovechando las características no ideales descritas anteriormente, se han desarrollado varios tipos de diodos, cada uno destinado a una aplicación específica.

Los diodos que se describen a continuación son los más representativos. Se han omitido algunos, altamente especializados o en desarrollo, como el láser, el piezoeléctrico, el pin, etc. pero se incluyen otros, igualmente interesantes, como el diodo túnel, el varicap, el diodo Schottky, etc.

Los *diodos de señal* son pequeños diodos de germanio que se utilizan como detectores de R.F. en receptores de radio. Se caracterizan por requerir de muy bajo voltaje para conducir. En la figura 191 se muestra un circuito típico de aplicación como detector de señales en un receptor de radio muy sencillo, llamado *radio de galena* o receptor de cristal.

Receptor de cristal con diodo

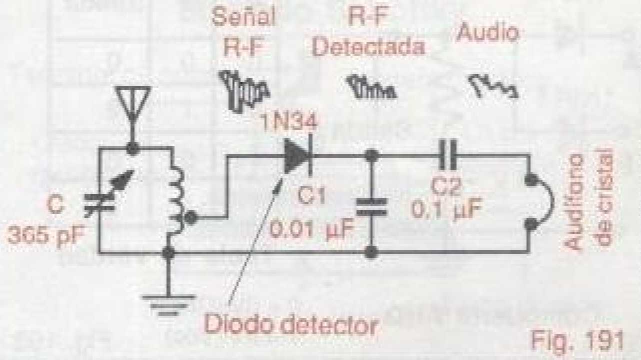

Fig. 191

Los *diodos rectificadores de potencia* son diodos de silicio de mayor tamaño que los diodos de señal y con especificaciones más amplias de corriente y voltaje, diseñados para ser usados en fuentes de poder y otras aplicaciones de potencia.

Esta es una de las aplicaciones más comunes para los diodos: la conversión de corriente alterna en corriente continua, un proceso que se encuentra en casi todos los aparatos electrónicos.

Los diodos rectificadores pueden manejar corrientes desde milésimas hasta cientos de amperios y pueden soportar tensiones inversas desde unos pocos hasta varios miles de voltios. En la figura 192 se muestra un circuito típico de aplicación de un rectificador de corriente.

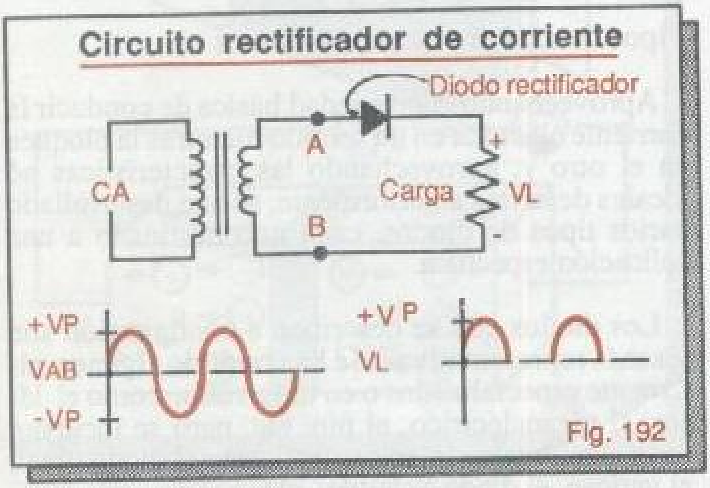
Circuito rectificador de corriente
Diodo rectificador
A
B
CA
Carga
VL
+
-
+VP
VAB
-VP
+V P
VL
Fig. 192

Los *diodos de conmutación* son también diodos de silicio, pero menos robustos que los diodos de potencia, diseñados para ser usados en circuitos digitales. Se caracterizan por su baja capacitancia de juntura, lo que los hace muy rápidos. En la figura 193 se muestra una aplicación típica.

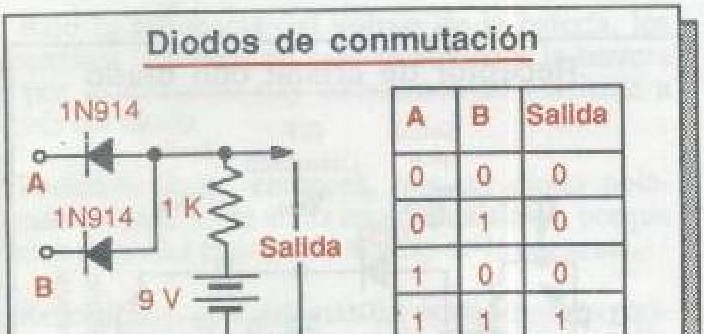
Diodos de conmutación
1N914
A
1N914
1 K
B
9 V
Salida
Compuerta AND

A	B	Salida
0	0	0
0	1	0
1	0	0
1	1	1

Tabla de verdad

0 = 0V (GND)
1 = 9V (Vcc)

Fig. 193

Los *diodos Zener* son dispositivos diseñados específicamente para ser utilizados como estabilizadores de voltaje. Pueden suministrar tensiones de referencia fijas hasta de varios cientos de voltios y manejar potencias de varias decenas de vatios en compañía de transistores o circuitos integrados. En la figura 194 se muestra el símbolo y una aplicación típica de este tipo de diodo.

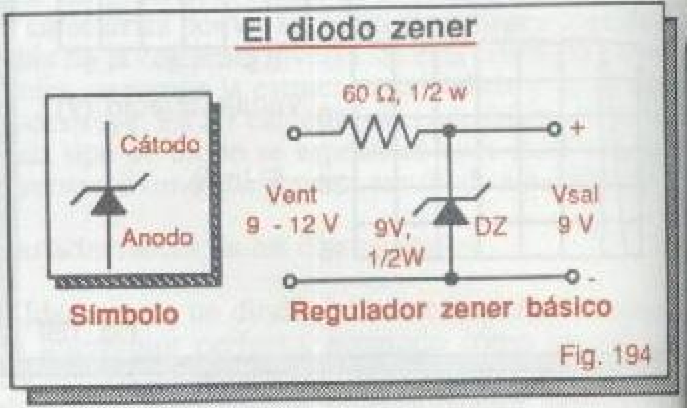
El diodo zener
Cátodo
Anodo
Símbolo
60 Ω, 1/2 w
+
Vent
9 - 12 V
9V,
1/2W
DZ
Vsal
9 V
-
Regulador zener básico
Fig. 194

Los *diodos emisores de luz* o LEDs son diodos que emiten una luz visible o infrarroja cuando se polarizan directamente. Se emplean en transmisores ópticos y como visualizadores o pilotos en múltiples aplicaciones. En la figura 195 se muestra un circuito típico de utilización.

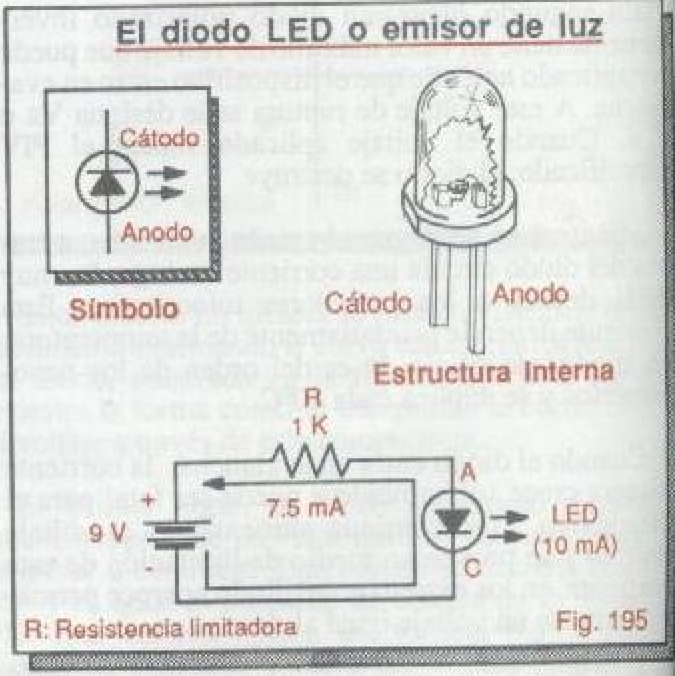
El diodo LED o emisor de luz
Cátodo
Anodo
Símbolo
Cátodo
Anodo
Estructura Interna
R
1 K
+
9 V
-
7.5 mA
A
C
LED
(10 mA)
R: Resistencia limitadora
Fig. 195

Los *fotodiodos* son diodos sensibles a la luz. Generan un voltaje de corriente continua proporcional a la cantidad de luz que incide sobre su superficie. Se utilizan como medidores y sensores de luz y en receptores ópticos de comunicaciones.

En la figura 196 se muestra un circuito típico de aplicación.

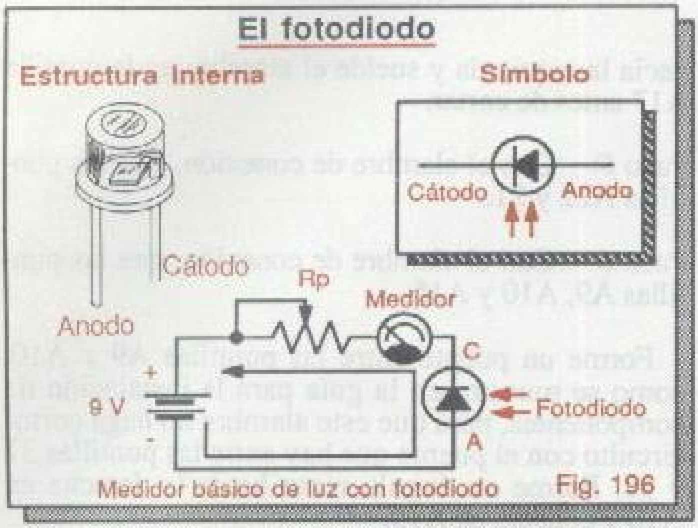
El fotodiodo
Estructura Interna
Símbolo
Cátodo
Anodo
Cátodo
Anodo
Rp
Medidor
+
9 V
C
Fotodiodo
A

Medidor básico de luz con fotodiodo Fig. 196

Los *diodos túnel* son dispositivos muy versátiles que pueden operar como detectores, amplificadores y osciladores.

Poseen una región de juntura extremadamente delgada y tienen *resistencia negativa*, esto es, la corriente disminuye a medida que aumenta el voltaje aplicado.

En un diodo túnel, los portadores de corriente atraviesan la juntura a la velocidad de la luz (3x10^8 m/seg). A los diodos túnel se les denomina también *diodos Esaki*, en honor de su inventor.

Los *varactores* son diodos de capacitancia variable que se utilizan como sintonizadores en sistemas de comunicaciones, especialmente en FM.

Poseen una región de juntura extremadamente flexible y ofrecen su máxima capacitancia con voltaje de polarización cero. En la figura 197 tenemos un circuito típico de aplicación. Estos diodos también reciben el nombre de diodos *Varicap*.

Los *rectificadores controlados de silicio* o SCRs son diodos de potencia que, además de ánodo y cátodo, tienen un tercer electrodo de control llamado *compuerta o Gate*. Cuando se polariza en directo, un SCR sólo conduce cuando recibe un pulso de disparo válido en la compuerta.

Una vez que entra en conducción, el SCR continúa conduciendo, incluso si se suspende la señal de la compuerta.

Esta característica lo hace muy útil como interruptor de estado sólido en aplicaciones de control y de potencia .

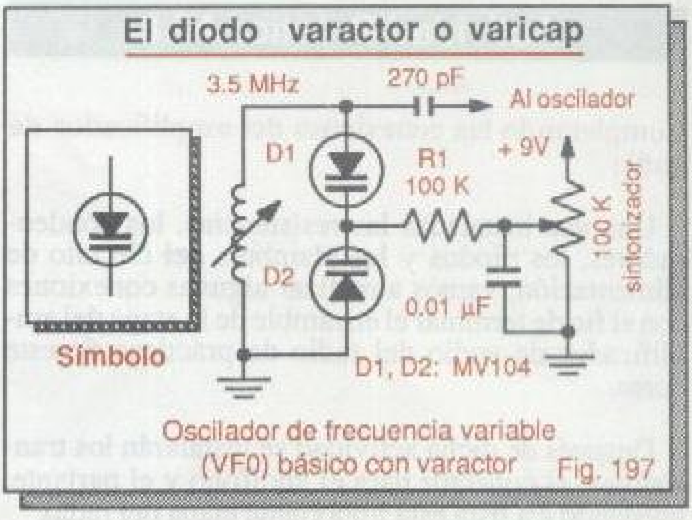
El diodo varactor o varicap
3.5 MHz
270 pF
Al oscilador
D1
R1
100 K
+ 9V
100 K
sintonizador
D2
0.01 µF
Símbolo
D1, D2: MV104

Oscilador de frecuencia variable (VF0) básico con varactor Fig. 197

En la figura 198 se muestra un circuito práctico de utilización con SCR.

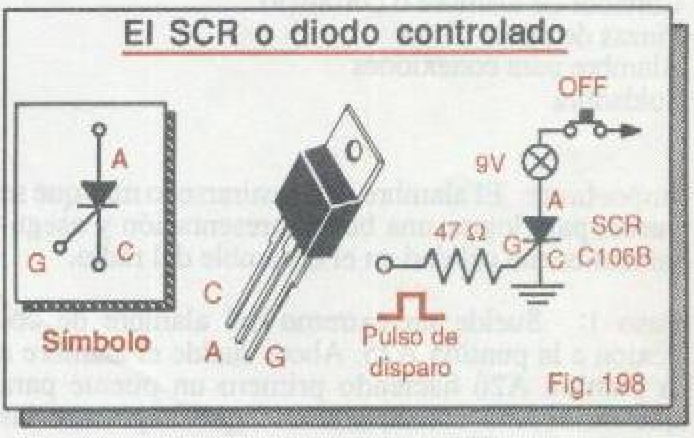
El SCR o diodo controlado
OFF
A
G
C
9V
A
SCR
47 Ω
G
C
C106B
C
Símbolo
A
G
Pulso de disparo

Fig. 198

Los *diodos Schottky* tienen una caída de voltaje directa (V_F) muy pequeña, del orden de 0.3 V o menos. Operan a muy altas velocidades y se utilizan en fuentes de potencia, circuitos de alta frecuencia y sistemas digitales. En la figura 199 se muestra su estructura interna. Reciben también el nombre de diodos de recuperación rápida (fast recovery diodes) o diodos de portadores calientes.

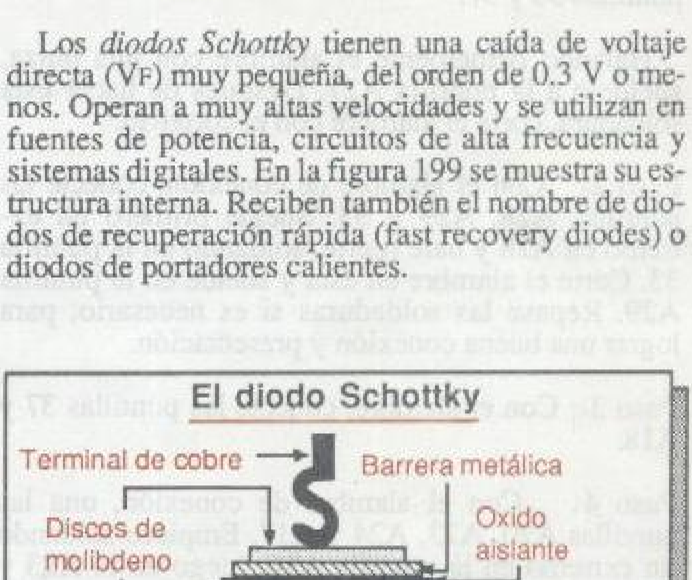
El diodo Schottky
Terminal de cobre
Barrera metálica
Discos de molibdeno
Oxido aislante
Base de cobre
Silicio dopado

Fig. 199

Actividad Práctica Nº 8

Completando las conexiones del amplificador de audio

Una vez instaladas las resistencias, los condensadores, los diodos y los alambres del circuito de alimentación, vamos a realizar algunas conexiones con el fin de terminar el ensamble de la etapa del amplificador de audio del radio de prácticas de este curso.

Después de dicha actividad se instalarán los transistores, el conector para el audífono y el parlante, quedando así lista esta importante etapa del radio.

Herramientas y materiales necesarios

Cautín de 25 o 40 vatios
Cortador de alambre o corta-frío
Pinzas de punta plana
Alambre para conexiones
Soldadura

Importante: El alambre debe estirarse lo más que se pueda, para lograr una buena presentación y asegurar una buena calidad en el ensamble del radio.

Paso 1: Suelde un extremo del alambre de conexión a la puntilla A25. Ahora suelde el alambre a la puntilla A26 haciendo primero un puente para que no se una con al alambre que hay entre las puntillas 36 y 37.

Halando firmemente el alambre con una pinza, suéldelo a la puntilla A27. Después de hacer esta soldadura, corte el alambre con el cortafrío.

Paso 2: Con el alambre de conexión, conecte las puntillas A28, A29 y 33. Suelde primero un extremo en A28 y hale fuerte, soldando en la puntilla 33. Corte el alambre en ésta y suelde en la puntilla A29. Repase las soldaduras si es necesario, para lograr una buena conexión y presentación.

Paso 3: Con el alambre, conecte las puntillas 37 y A18.

Paso 4: Con el alambre de conexión, una las puntillas A20, A23, A24 y A17. Empiece soldando un extremo en la puntilla A20, luego en la A23 y forme un ángulo recto en esa puntilla con el alambre hacia abajo.

Hale fuerte el alambre con la pinza y suelde el alambre en la puntilla A24. Haga otro ángulo recto hacia la izquierda y suelde el alambre en la puntilla A17 antes de cortar.

Paso 5: Con el alambre de conexión, una las puntillas A22 y 21.

Paso 6: Con el alambre de conexión, una las puntillas A9, A10 y A15.

Forme un puente entre las puntillas A9 y A10, como se muestra en la guía para la instalación de componentes, para que este alambre no haga cortocircuito con el puente que hay entre las puntillas 37 y 38. Forme un ángulo recto hacia la derecha en A10 y conecte en A15.

Paso 7: Con el alambre de conexión, una las puntillas A12, A13 y A16, formando ángulo recto hacia la derecha en A13.

Paso 8: Con el alambre de conexión, una las puntillas A14 y 20.

Paso 9: Con el alambre de conexión, una las puntillas A8 y A12, haciendo un puente para evitar la unión con el alambre que conecta las puntillas 29 y 31.

Paso 10: Con el alambre de conexión, una las puntillas A5, A6 y A7, haciendo un puente en el tramo A6 - A7 como se muestra en la guía.

Paso 11: Con el alambre de conexión, una las puntillas C7, A1 y A2, formando ángulo recto hacia abajo en la puntilla A1.

Paso 12: Con el alambre de conexión, una las puntillas A4 y 18.

Una vez terminada esta actividad, repase cada uno de los pasos aquí descritos, con el fin de verificar que hayan sido conectados correctamente todos los puntos indicados.

Repase las soldaduras que tengan un mal aspecto físico y rectifique las trayectorias de los alambres con el fin de que queden lo más rectos posibles.

De la buena calidad de estas conexiones depende en gran parte el buen funcionamiento del radio. Como ejercicio, puede verificar las conexiones con un multímetro ajustado, para la medida de ohmios, en la escala más baja.

Los Transistores

En esta lección continuamos nuestro estudio de los dispositivos semiconductores con uno de sus componentes más importantes: *el transistor*. Enfocaremos nuestra atención en los transistores bipolares (NPN y PNP) y en los de efecto de campo (JFET y MOSFET), haciendo especial énfasis en sus principios de operación, sus características y sus configuraciones básicas.

Qué es un transistor

La palabra **transistor** se ha derivado de **transfer resistor** (resistencia de transferencia) y designa, en forma genérica, a un dispositivo o componente electrónico de tres terminales, cuya resistencia interna es una función del nivel de corriente o voltaje aplicado a uno de sus terminales.

Aprovechando esta propiedad, el transistor se utiliza para amplificar señales, para producir ondas de corriente alterna (oscilaciones) o como suiche electrónico.

El transistor, inventado en 1948 en los laboratorios de la Bell Telephone en Estados Unidos es, sin duda, uno de los adelantos más significativos de nuestra era y uno de los componentes más versátiles e importantes de la electrónica moderna. De hecho, todos los circuitos integrados, que han revolucionado la ciencia moderna, se fabrican con transistores.

El transistor reemplazó, con grandes ventajas, a los tubos de vacío o válvulas en todos los circuitos electrónicos. Sus principales características son: tamaño reducido, bajo consumo de potencia, alta confiabilidad y costo mínimo.

Clases de transistores

Existen dos grandes grupos de transistores: los transistores ***bipolares*** y los transistores de efecto de campo o ***FETs***. Los bipolares son dispositivos controlados por corriente y los FET son dispositivos controlados por voltaje. Los FET se conocen también como transistores unipolares.

En los transistores bipolares, los tres terminales se denominan ***emisor*** (E), ***colector*** (C)y ***base*** (B). En los FET, se designan como ***fuente*** (S), ***drenador*** (D) y ***compuerta*** (G). La base y la compuerta actúan, respectivamente, como electrodos de control.

Los transistores bipolares pueden ser NPN o PNP. Los FET pueden ser de canal N o de canal P y de unión (JFET) o de compuerta aislada (MOSFET).

En la figura 200, se muestra un resumen de los símbolos utilizados en los diagramas electrónicos para referirse a los diferentes transistores de cada grupo.

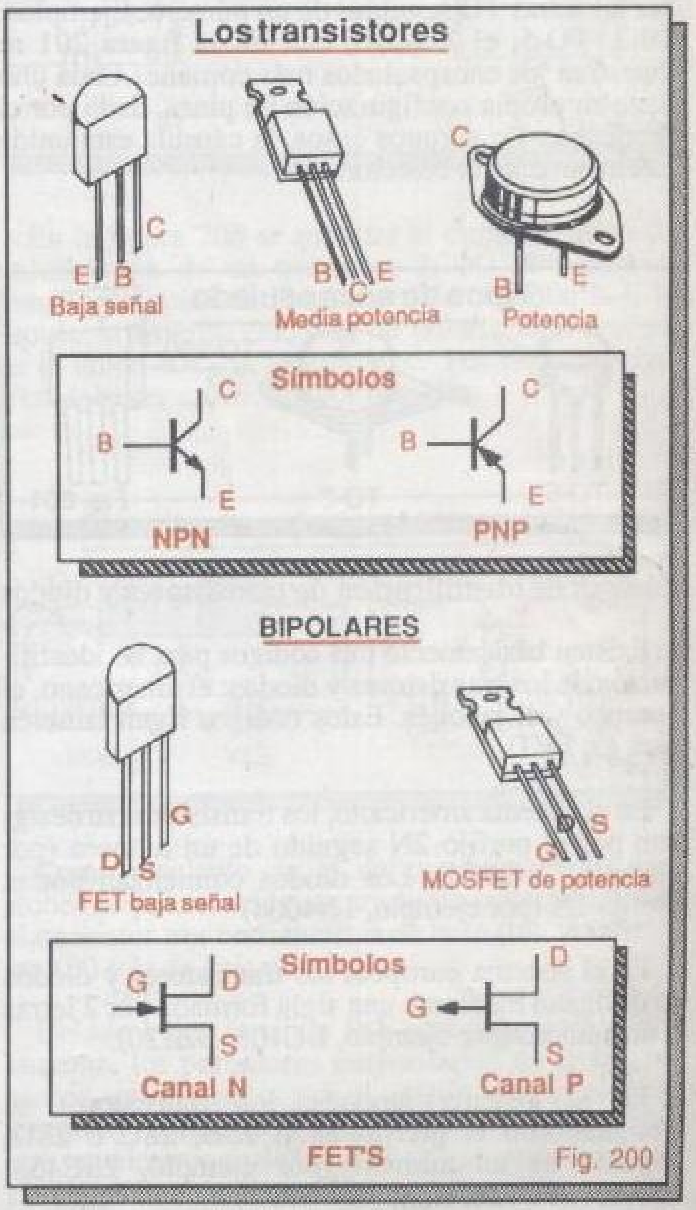

De acuerdo a su aplicación, existen transistores de pequeña señal, de conmutación, de potencia, de alta frecuencia, etc.

Tipos de encapsulados

Los transistores se fabrican en serie, formando simultáneamente varios cientos o millares de unidades sobre una oblea semiconductora de 38 a 50 mm de diámetro y luego cortándolos uno por uno. Las técnicas de fabricación más utilizadas son los procesos de aleación, difusión y crecimiento epitaxial.

Una vez construidos, los transistores se encierran en cápsulas plásticas o metálicas. La cápsula protege al dispositivo de la humedad y otros contaminantes, sirve como disipador de calor, aloja los pines del transistor, facilita su manipulación, etc.

El tipo de encapsulado se identifica generalmente por las letras TO seguidas de un número. Ejemplos: T0-3, TO-5, el TO-220, etc. En la figura 201 se muestran los encapsulados más comunes.Cada uno tiene su propia configuración de pines, dada por el fabricante. En algunos casos, la cápsula está unida eléctricamente al colector.

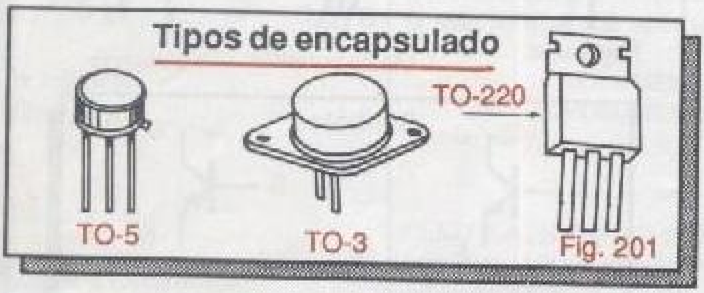
Tipos de encapsulado

Fig. 201

Códigos de identificación de transistores y diodos

Existen básicamente tres códigos para la identificación de los transistores y diodos: el americano, el europeo y el japonés. Estos códigos rigen también para los FETs.

En el sistema americano, los transistores se designan por el prefijo 2N seguido de un número (por ejemplo, 2N3904). Los diodos comienzan por el prefijo 1N (por ejemplo, 1N4004)

En el sistema europeo, los transistores y diodos se designan mediante una sigla formada por 2 letras y un número (por ejemplo, BC108, BZ120).

En el sistema japonés, los transistores se designan con el prefijo 2SA, 2SB, 2SC o 2SD, seguido de un número (por ejemplo, 2SC458, 2SA65, 2SD926, etc.).

Es importante señalar que, a pesar de esta normalización, algunos fabricantes insisten en utilizar sus propias referencias (por ejemplo, ECG123AP, CA3082, TIP31, etc).

Transistores bipolares

Un transistor bipolar es un dispositivo formado por tres capas de material semiconductor, organizadas en forma de "sandwich". Puede ser un material tipo N entre dos materiales tipo P o viceversa. De acuerdo a la forma como se alternen estas tres capas, se pueden obtener dos tipos de transistores: el tipo NPN y el tipo PNP.

En la figura 202 se muestra la estructura y el símbolo de un transistor **NPN**. Está formado por una región tipo P muy delgada colocada entre dos regiones tipo N ligeramente dopadas. La capa P del centro es la *base* (B) y las 2 capas N de los extremos son el *emisor* (E) y el *colector* (C), respectivamente.

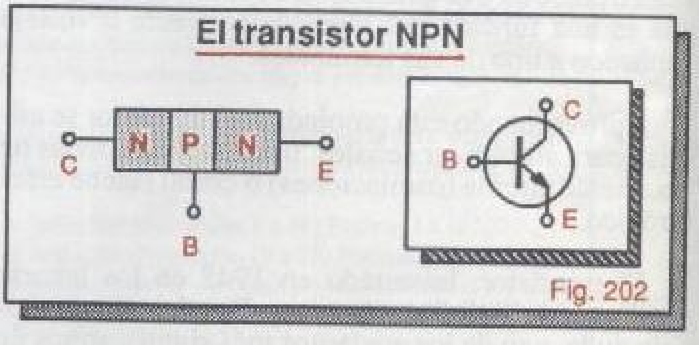
El transistor NPN

Fig. 202

En la figura 203 se muestra la estructura y el símbolo de un transistor **PNP**. Está formado por una región tipo N muy delgada, colocada entre dos regiones tipo P ligeramente dopadas. La capa N del centro es la *base* (B) y las 2 capas P de los extremos son el *emisor* (E) y el *colector* (C), respectivamente.

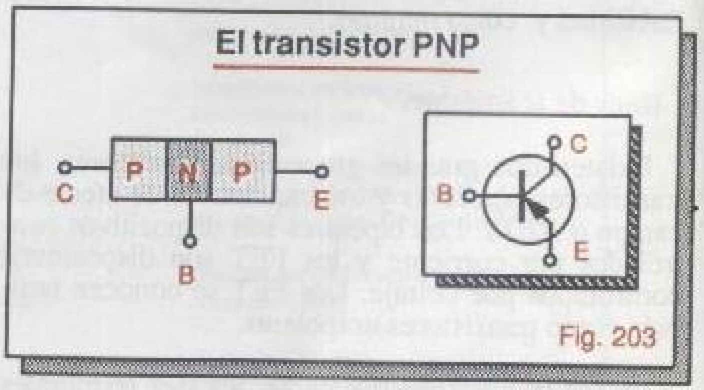
El transistor PNP

Fig. 203

Observe que la única diferencia entre los símbolos de los transistores NPN y PNP es la dirección en que apunta la flecha de emisor. En un PNP, la flecha entra ("**penetra**") y en un NPN sale ("**no** penetra").

Además, debido a la forma como se alternan las capas P y N, en un transistor bipolar se originan dos uniones o junturas PN: una entre emisor y base y otra entre colector y base. A estas zonas las denominaremos, respectivamente, unión B-E o E-B y unión B-C o C-B.

Características constructivas especiales de los transistores

En un transistor bipolar se han alterado ciertas características eléctricas y físicas de las regiones que conforman la base, el emisor y el colector, con el fin de conseguir las propiedades excepcionales que caracterizan al dispositivo. Veamos las más importantes.

En primer lugar, la base es una región sumamente estrecha y poco dopada en relación a las otras dos. En consecuencia, tiene una concentración muy baja de portadores de corriente.

La estrechez de la región de base facilita, además, el paso rápido de los portadores mayoritarios de corriente desde el emisor hasta el colector.

En segundo lugar, la región de emisor está fuertemente dopada y la concentración de portadores mayoritarios disponibles supera ampliamente la de la base.

En el caso de un transistor NPN, por ejemplo, lo anterior significa que la base no posee la suficiente cantidad de huecos para combinarse con todos los electrones que puede suministrar el emisor. En consecuencia, la corriente de polarización directa o de conducción que circula a través de la unión B-E es relativamente pequeña.

En tercer lugar, la región de colector es sumamente amplia y tiene una alta concentración de portadores minoritarios en relación a la base y muy pocos portadores mayoritarios en comparación con el emisor. Como consecuencia, la corriente de polarización inversa o de fuga que circula a través de la unión B-C es muy pequeña.

Funcionamiento del transistor NPN

La función básica y más importante de un transistor es la amplificación de corriente. Para que el dispositivo desarrolle correctamente esta misión, sus dos uniones o junturas PN deben polarizarse en forma adecuada. Específicamente, la unión E-B debe polarizarse en directo y la unión C-B en inverso, como si se tratara de dos diodos.

Lo anterior significa que, en un transistor NPN, la base (P) debe ser positiva con respecto al emisor (N), y negativa con respecto al colector (N).

Asimismo, en un transistor PNP, la base (N) debe ser negativa con respecto al emisor (P) y positiva con respecto al colector (P).

Estas condiciones se implementan en la práctica mediante dos voltajes externos de CC, aplicados simultáneamente. En la figura 204B se muestra la forma de polarizar directamente la unión B-E de un transistor NPN, y en la figura 204A la forma de polarizar inversamente la unión B-C del mismo.

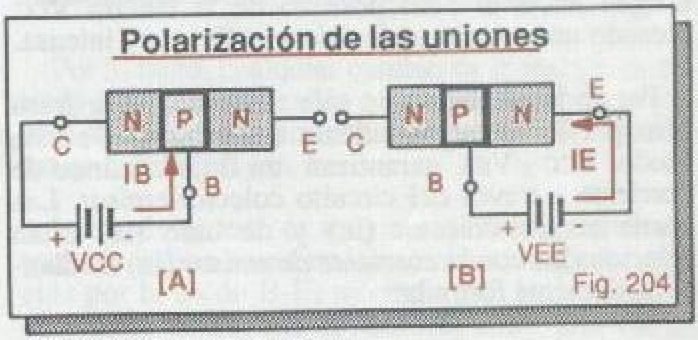

Fig. 204

En la figura 205 se muestra el circuito básico de polarización de un transistor NPN. El voltaje o tensión de polarización directa de la unión B-E la provee la batería VEE, y el de polarización inversa de la unión B-C, la batería Vcc. Por regla general, Vcc debe ser mucho mayor que VEE.

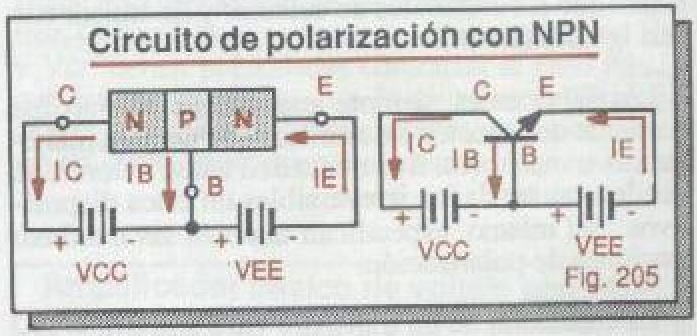

Fig. 205

Como resultado de la aplicación de estas tensiones de polarización, se producen internamente en el transistor tres corrientes: la de base (IB), la de emisor (IE) y la de colector (IC).

Debido a que la unión B-E está polarizada directamente, los portadores mayoritarios de ambas regiones son obligados por el voltaje VEE a cruzar la juntura y combinarse mutuamente. Esto significa que los electrones del emisor (N) se combinan con los pocos huecos disponibles en la base (P).

En realidad, sólo del 1 al 5% de los electrones suministrados por el emisor se combinan con huecos en la base, formándose así una *corriente de base* (IB) muy débil.

Por cada combinación electrón-hueco, penetra un electrón por el emisor y sale otro por la base. De este modo, VEE garantiza un flujo continuo de corriente a través del circuito emisor-base.

Los electrones restantes (del 95 al 99%) que invaden la base, son atraídos hacia el colector por la fuerte tensión inversa de polarización VCC de la unión B-C.

Estos electrones cruzan la unión B-C, pasan a través de la extensa región N de colector y se dirigen hacia el polo positivo de la batería VCC, creando una *corriente de colector* (IC) muy intensa.

Por cada electrón que sale del colector, ingresa otro por la base, procedente del emisor. De este modo, VCC y VEE garantizan un flujo continuo de corriente a través del circuito colector-emisor. Las corrientes de colector (IC) y de base (IB) están relacionadas con la *corriente de emisor* (IE) mediante la siguiente fórmula:

$$\boxed{I_E = I_B + I_C}$$

Es decir, la corriente de emisor es igual a la suma de las corrientes de base y de colector o, en otras palabras, la corriente de colector es igual a la corriente suministrada por el emisor menos la corriente que se pierde en la base. Para efectos prácticos, se considera que las corrientes de colector y de emisor son casi iguales ($I_C \approx I_E$).

Los fabricantes siempre especifican la máxima corriente de colector o de emisor que pueden manejar sus transistores. Si se exceden estos valores, se pueden causar daños irreversibles en estos dispositivos. Así mismo, especifican también las máximas tensiones de polarización.

La capacidad de un transistor para amplificar corriente se mide mediante dos parámetros muy importantes llamados *alfa* (α) y *beta* (β) que estudiaremos posteriormente.

Funcionamiento del transistor PNP

En la figura 206 se muestra el circuito básico de polarización de un transistor PNP. Su funcionamiento es exactamente igual al del transistor NPN descrito anteriormente con dos excepciones:

a) En un transistor PNP, los portadores mayoritarios de corriente son huecos y no electrones como en el transistor NPN. Por lo tanto, las corrientes IB, IC e IE tienen direcciones contrarias a las que llevan en el caso NPN.

Circuito de polarización con PNP

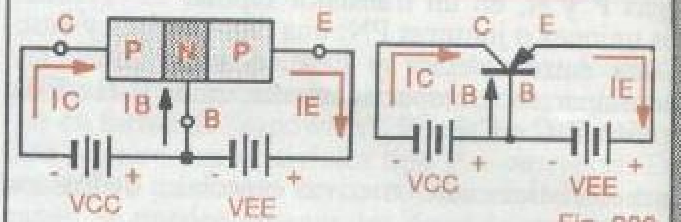

Fig. 206

b) Las polaridades de las tensiones de polarización directa (VEE) e inversa (VCC) son las opuestas al caso NPN.

Nota: Estrictamente hablando, una corriente es un flujo de electrones y no de huecos (los huecos no son partículas físicas). Por lo tanto, imagine siempre una corriente de huecos como una corriente de electrones que se desplaza en sentido contrario.

El transistor bipolar como amplificador

Una vez familiarizados con la acción básica que tiene lugar en el interior de un transistor, vamos a ver cómo puede emplearse un transistor para amplificar señales electrónicas.

En la figura 207 se muestra nuevamente el circuito de polarización de un transistor NPN, pero utilizando una tensión de polarización VEE variable.

Polarización con tensión variable

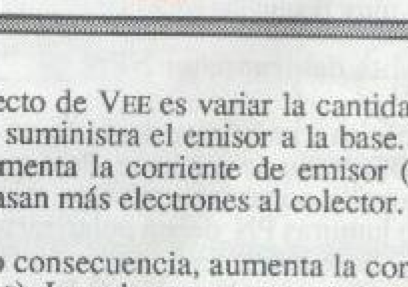

Fig. 207

El efecto de VEE es variar la cantidad de electrones que suministra el emisor a la base. Al aumentar VEE, aumenta la corriente de emisor (IE) y, por lo tanto, pasan más electrones al colector.

Como consecuencia, aumenta la corriente de colector (IC). Los electrones que no pasan al colector se combinan con los huecos de la base, aumentando la corriente de base (IB).

Al disminuir VEE, sucede el efecto contrario y las tres corrientes se reducen en la misma proporción.

Todo lo anterior es una consecuencia directa de la relación. $I_E = I_B + I_C$.

Para utilizar el circuito de la figura 207 como amplificador de corriente o de voltaje , el transistor debe ser capaz de aceptar a la entrada una señal pequeña y suministrar a la salida una señal de mayor intensidad o amplitud. *La señal de salida debe, por tanto, ser una réplica ampliada de la señal de entrada.*

Un circuito básico que permite utilizar el transistor (NPN, en este caso) como amplificador de voltaje se muestra en la figura 208. Observe que se ha intercalado una resistencia (RL) en el circuito colector-emisor y que se ha interrumpido el circuito emisor-base para inyectar la señal de entrada.

Amplificador básico de voltaje con NPN

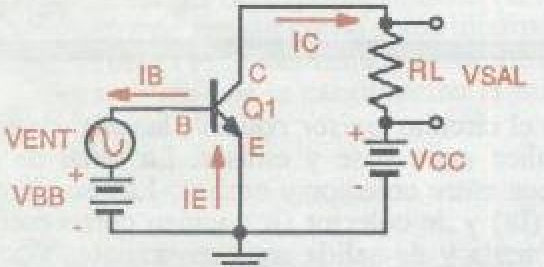

Fig. 208

La resistencia RL se denomina *resistencia de carga* y su propósito es permitir que entre sus terminales se desarrolle una tensión de salida (VSAL) cuando la atraviesa la corriente de colector.

El efecto de la señal de entrada (VENT) es superponerse a la tensión de polarización original (VBB) y variar el voltaje total de polarización entre base y emisor (VBE).

En la figura 209 se representan gráficamente estos 3 voltajes para mayor claridad. Se supone que VENT es un voltaje alterno y que VBE se mantiene siempre por encima de la caída directa de voltaje de la unión B-E (≈0.7V). VCQ es el voltaje de *reposo* entre colector y emisor, sin señal de entrada.

Si la señal de entrada tiene la misma polaridad de VBB, la tensión de polarización VBE aumenta y por lo tanto aumentan IB e IC. Al aumentar Ic, aumenta la tensión de salida.

Si la polaridad de la señal de entrada es contraria a la de VBB, la tensión de polarización VBE disminuye y, en consecuencia, disminuyen IB e IC. Como resultado, disminuye la tensión de salida.

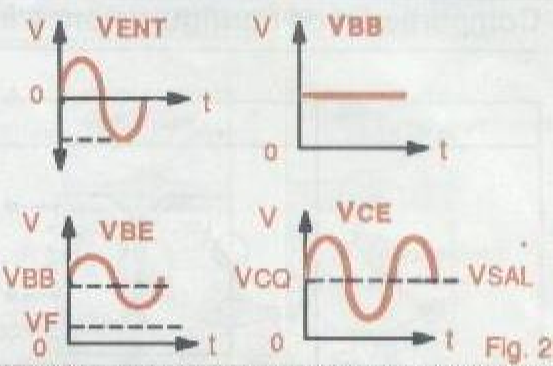

Fig. 209

Por lo tanto, cualquier cambio en la tensión de entrada produce un cambio correspondiente en la tensión de salida. Sin embargo, el cambio en esta última es mucho mayor.

La razón de esto es muy sencilla: la señal de entrada se aplica a una resistencia muy baja (la ofrecida por la unión B-E) mientras que la de salida se desarrolla sobre una resistencia muy alta (la de carga RL).

El circuito descrito anteriormente es básicamente un *amplificador de voltaje*. En la siguiente sección veremos cómo puede conectarse un transistor para proveer, además, amplificación de corriente y amplificación de potencia.

En la figura 210 se muestra la versión con transistor PNP del amplificador básico de voltaje anterior. Observe que las tensiones de polarización VBB y VCC tienen polaridades contrarias al caso NPN y que las corrientes IE, IB e IC circulan en las direcciones opuestas. El funcionamiento del circuito es, sin embargo, exactamente el mismo.

Amplificador básico de voltaje con PNP

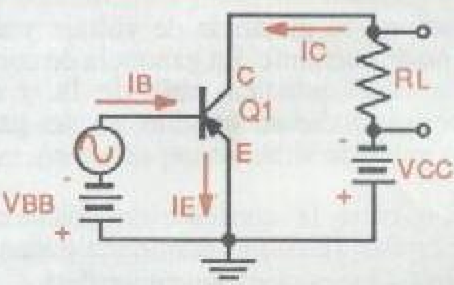

Fig. 210

Configuraciones básicas de circuitos amplificadores con transistores

Los transistores bipolares (NPN o PNP) se pueden conectar de tres formas diferentes para proveer la importante función de amplificación.

Comparación de configuraciones amplificadoras básicas con transistores bipolares

Parámetros	Base común	Emisor común	Colector común
Configuración básica	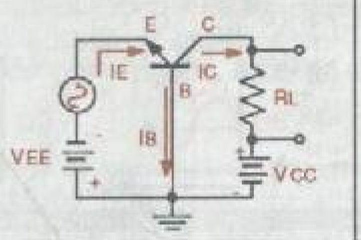	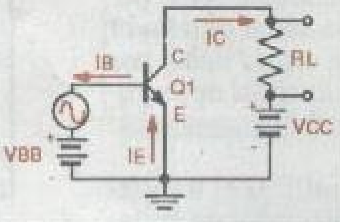	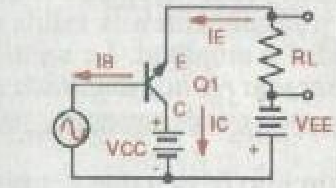
Ganancia de voltaje	Mayor de1	Mayor de 1	Menor de 1
Ganancia de corriente	Menor de 1 (α = 0.92 a 0.98)	Mayor de 1 (ß = 10 a 400)	Mayor de 1 (ß = 10 a 400)
Ganancia de potencia	Mayor de 1	Mayor de 1	Mayor de 1
Impedancia de entrada	Muy baja (30 Ω - 40 Ω)	Moderada (1 KΩ - 10 KΩ)	Muy alta (más de 100 KΩ)
Impedancia de salida	Muy alta (1 MΩ - 2 MΩ)	Moderada (10 KΩ - 50 KΩ)	Muy baja (menos de 100 Ω)
Aplicación típica	Amplificación de altas frecuencias	Amplificación de bajas frecuencias	Acoplamiento de impedancias

Fig. 211

Estas configuraciones se denominan *base común*, *emisor común* y *colector común*.

En la figura 211 se comparan cualitativamente estos tres montajes desde los siguientes puntos de vista: impedancia de entrada, impedancia de salida, ganancia de corriente, ganancia de voltaje, ganancia de potencia y aplicaciones típicas.

En el circuito *base común* , la señal de entrada se aplica entre emisor y base. La señal de salida aparece entre colector y base. Las corrientes de emisor (IE) y de colector (IC) actúan como corrientes de entrada y de salida respectivamente. VCC es la tensión de polarización inversa de la unión C-B y VEE la de polarización directa de la unión B-E.

El circuito provee ganancia de voltaje y de potencia pero no de corriente. La ganancia de corriente se determina en la práctica variando IE, y observando el correspondiente cambio en IC para un determinado valor de VCB.

La relación entre la corriente de colector y la corriente de emisor (IC/IE) se denomina ganancia de corriente *alfa* (α) y es siempre menor de 1.

Los fabricantes de transistores especifican en las hojas de datos de sus productos el valor de alfa a una determinada frecuencia, generalmente 1 KHz. Esto se debe a que, cuando se trabaja con señales de CA, este parámetro depende de la frecuencia. La frecuencia a la cual α se reduce al 70% de su valor nominal se denomina frecuencia de corte alfa (FAB).

En el circuito *emisor común* , la señal de entrada se aplica entre base y emisor. La señal de salida aparece entre colector y emisor. Las corrientes de base (IB) y de colector (IC) actúan como corrientes de entrada y de salida respectivamente. VCC es la tensión de polarización inversa de la unión C-B y VBB la de polarización directa de la unión B-E.

El circuito provee, simultáneamente, ganancia de corriente, de voltaje y de potencia. Es la configuración más utilizada en los circuitos comunes como radios, amplificadores de sonido y otros.

La ganancia de corriente del montaje emisor común se determina en la práctica variando IB, y observando el correspondiente cambio en IC para un determinado valor de VCE. La relación entre la corriente de colector y la corriente de base (IC/IB) se denomina ganancia de corriente ***beta*** (ß) o hFE. En general, el beta de un transistor puede tomar valores entre 10 y 400.

Los fabricantes de transistores especifican en las hojas de datos de sus productos, el valor de beta a una determinada frecuencia, generalmente 1 KHz. Esto se debe a que, cuando se trabaja con señales de CA, este parámetro (al igual que alfa), depende de la frecuencia. La frecuencia a la cual ß se reduce al 70% de su valor nominal se denomina frecuencia de corte beta (FAE).

En el circuito colector común, la señal de entrada se aplica entre base y colector. La señal de salida aparece entre emisor y colector.

Las corrientes de base (I_B) y de emisor (I_E) actúan como corrientes de entrada y de salida respectivamente: V_{CC} es la tensión de polarización inversa de la unión C-B y V_{BB} la de polarización directa de la unión B-E.

El circuito en colector común se denomina también *seguidor de emisor* y provee ganancia de corriente y de potencia pero no de voltaje. El voltaje de salida es prácticamente igual al voltaje de entrada, excepto por la caída de tensión directa en la unión base-emisor (≈0.7V).

La ganancia de corriente se determina en la práctica variando I_B, y observando el correspondiente cambio en I_E para un determinado valor de V_{CC}. Para efectos prácticos, se considera que las configuraciones emisor común y colector común tienen la misma ganancia de corriente (ß).

Los circuitos de colector común se emplean principalmente para acoplar fuentes de señal de alta resistencia con cargas de baja resistencia. Al utilizarse un seguidor de emisor como etapa intermedia, se evita que la carga exija demasiada corriente de la fuente y se garantiza que la carga reciba la misma tensión de la fuente.

Prueba de transistores bipolares

Existen varios instrumentos especializados para la prueba de transistores. Estos aparatos realizan, entre otras, las siguientes funciones:

- Diagnostican si un transistor particular es NPN o PNP.
- Determinan si una o ambas uniones están abiertas o en cortocircuito.
- Suministran el valor de sus parámetros α, ß, etc.
- Miden las corrientes de fuga.
- Visualizan en un osciloscopio la forma como se comporta el dispositivo ante diferentes valores de voltaje y de corriente.

No obstante, los problemas más comunes (circuitos abiertos y cortocircuitos) se pueden detectar fácilmente utilizando un multímetro en la función de óhmetro.

En la figura 212 se muestra gráficamente la forma de probar con un óhmetro un transistor NPN y un transistor PNP. Se supone que el terminal negro es de polaridad positiva y el rojo de polaridad negativa, como sucede en la mayoría de los casos. Se recomienda utilizar las escalas de Rx100 o Rx1K, por seguridad.

Prueba de transistores NPN y PNP

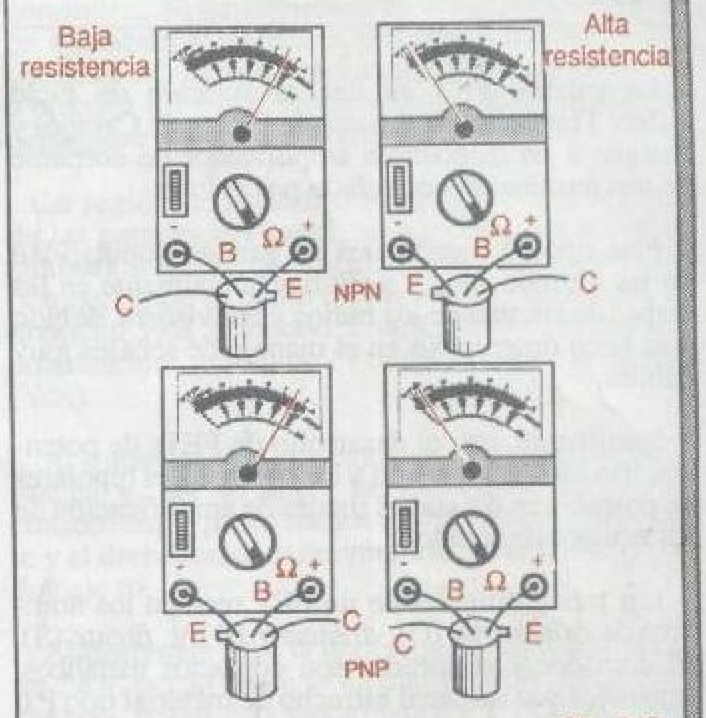

Fig. 212

El método consiste, básicamente, en probar la conducción o no conducción de las uniones B-E y B-C polarizándolas directa e inversamente. Recuerde que una unión PN presenta una baja resistencia en polarización directa y una alta resistencia en polarización inversa.

Un transistor NPN debe dar una baja resistencia cuando el terminal (-) se conecta a la base y el (+) al emisor o al colector y una alta resistencia en el caso contrario.

Un transistor PNP debe dar una baja resistencia cuando el terminal (+) se conecta a la base y el (-) al emisor o al colector y una alta resistencia en el caso contrario.

Un transistor NPN o PNP debe dar una alta resistencia entre colector y emisor, cualquiera que sea la conexión de las puntas de prueba.

Para transistores de baja y mediana potencia, la lectura de resistencia directa es de algunos cientos de Ω's y la de resistencia inversa de varios cientos de KΩ's. Los transistores de alta potencia dan lecturas menores. Sin embargo, las indicaciones de resistencia obtenidas no tienen significación real.

Para realizar la prueba de los transistores con el multímetro, se recomienda desconectarlo del circuito, con el fin de no medir otros componentes que están conectados a él.

Transistores de Efecto de Campo (FETs)

La palabra **FET** es una abreviación de Field Effect Transistor (Transistor de Efecto de Campo) y designa a un dispositivo amplificador de corriente de tres terminales, controlado por voltaje.

Este tipo de transistores ha ganado popularidad en los últimos años y se utiliza actualmente en las etapas de entrada de los radios y televisores, debido a su buen desempeño en el manejo de señales muy débiles.

Igualmente, con el desarrollo de FETs de potencia, han ido desplazando a los transistores bipolares de potencia en las etapas finales de amplificación de los equipos de sonido.

Los tres terminales de un FET reciben los nombres de *compuerta* (G), *drenador* (D) y *fuente* (S). El drenador y el surtidor son contactos metálicos, separados por un canal estrecho de material tipo P o N.

El flujo de corriente a través del canal lo regula el campo eléctrico producido al aplicar un voltaje de control a la compuerta.

La compuerta es también un contacto metálico y puede estar separada del canal por un material de tipo contrario o por una capa de óxido aislante. El primer caso corresponde a los FET de unión o JFET y el segundo a los FET de compuerta aislada o MOSFET. A los MOSFET se les denomina también IGFET.

Transistores FET de unión o JFET

En la figura 213 se muestra la estructura general de un FET de unión o JFET. Consta de un material de base tipo N o P, llamado *sustrato*, dentro del cual se forma una región de tipo opuesto, en forma de U, llamada *canal*. Sobre esta estructura se obtienen los contactos metálicos correspondientes a la compuerta, el drenador y la fuente.

La compuerta está conectada al sustrato y el drenador y la fuente a los extremos del canal. Por lo tanto, entre la compuerta y el canal se forma una unión PN. Sin embargo, el canal está levemente dopado en relación a la región de compuerta o sustrato.

En la mayoría de los casos, el diseño del canal es simétrico y, por lo tanto, cualquiera de los extremos se puede utilizar como drenador o como fuente. Sin embargo, existen casos especiales en los cuales el canal es asimétrico y, por consiguiente, no se pueden intercambiar estos terminales.

Estructura del FET de unión

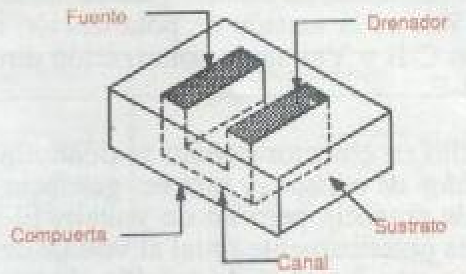

Fig. 213

Dependiendo de la naturaleza del canal, existen dos tipos de FET: el FET de canal N y el FET de canal P. En la figura 214 se muestran el símbolo y la estructura de un FET canal N y de un FET canal P. En el primer caso, el canal es de material N y el sustrato de material P. En el segundo, el canal es tipo P y el sustrato es tipo N.

FET de canal N y FET de canal P

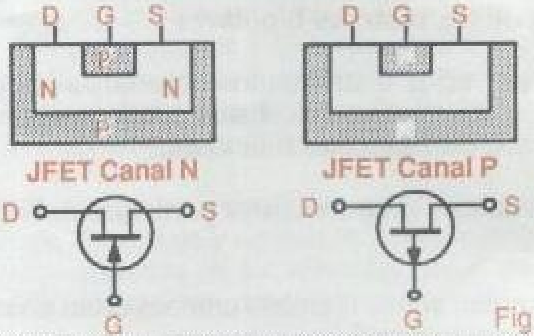

Fig. 214

Observe que la única diferencia entre los símbolos del FET canal N y del FET canal P es la dirección en que apunta la flecha de la compuerta. En el caso del FET canal N, la flecha se dirige hacia adentro y en el caso del FET canal P, se dirige hacia afuera.

Funcionamiento del FET de unión

Para operar correctamente, los FET de unión necesitan de dos tensiones o voltajes de polarización externas: una entre drenador y fuente y otra entre compuerta y fuente. La primera dirige el paso de los portadores de corriente por el canal y la segunda regula su cantidad.

En la figura 215 se muestra la forma correcta de polarizar un JFET canal N y un JFET canal P. La tensión entre drenador y fuente se designa por V_{DS}

Polarización del FET de unión

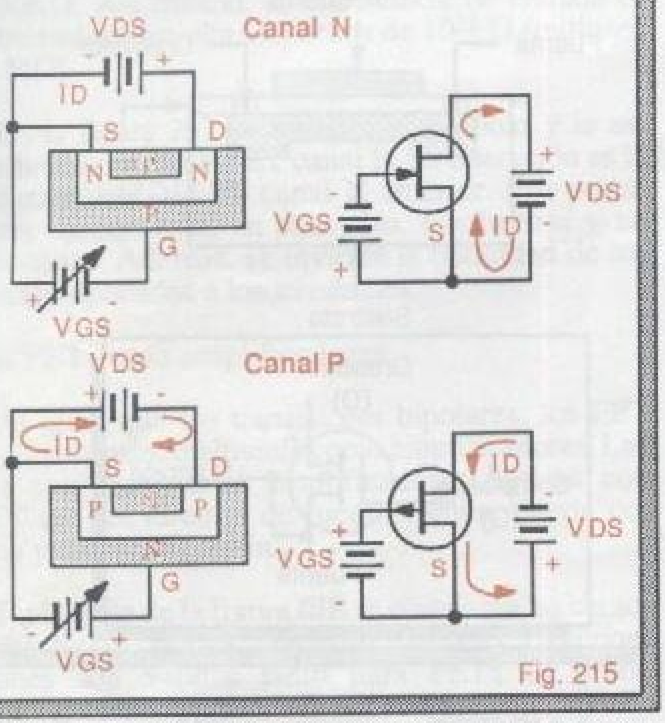

Fig. 215

y la tensión entre compuerta y fuente por V_{GS}. Esta última se conecta de modo que polarice inversamente la unión PN formada por el canal y el sustrato o compuerta.

Específicamente, en un JFET de canal N, la fuente debe ser *positiva* con respecto a la compuerta y *negativa* con respecto al drenador. Así mismo, en un JFET de canal P, la fuente debe ser *negativa* con respecto a la compuerta y *positiva* con respecto al drenador.

El efecto neto de cualquiera de estas disposiciones es la creación de una corriente entre drenador y fuente, la cual circula a lo largo del canal y se denomina ***corriente de drenaje*** (I_D).

Desde el punto de vista de la tensión V_{DS}, el canal se comporta simplemente como una resistencia variable, controlada por la tensión V_{GS}.

Analizaremos únicamente la operación del JFET de canal N. El funcionamiento del JFET de canal P es exactamente igual, teniendo en cuenta que los portadores de corriente son huecos y, por lo tanto, las tensiones de polarización V_{GS} y V_{DS} tienen polaridades contrarias.

En el caso del JFET canal N, al aplicar la tensión inversa de polarización V_{GS}, se crea en las proximidades de la unión PN sustrato-canal una zona de carga espacial, libre de electrones, llamada *región de agotamiento*.

Debido a que existe una fuerte concentración de huecos en el sustrato P o compuerta y una baja concentración de electrones en el canal N, la región de agotamiento se forma por completo dentro del canal. El espesor de esta zona determina el área útil o efectiva del canal y, por lo tanto, su capacidad de dejar pasar más o menos electrones.

La región de agotamiento se extiende a lo largo de las paredes del canal, siendo más amplia en el lado del drenador que en el de la fuente. Esto es así porque, desde el punto de vista de la compuerta, el drenador está sometido a una tensión inversa de polarización más alta (V_{DS}+V_{GS}) que la fuente (V_{GS}).

Al aumentar V_{GS} , la región de agotamiento se ensancha y, por lo tanto, se estrecha el canal. En consecuencia, pasan menos electrones entre la fuente y el drenador, disminuyéndose así la corriente de drenaje I_D.

Al disminuir V_{GS}, la región de agotamiento se estrecha y, por lo tanto, se amplía el canal . En consecuencia, pasan más electrones entre la fuente y el drenador, aumentándose así la corriente de drenaje I_D.

De este modo, V_{GS} varía la resistencia del canal y controla o "modula" la corriente de drenaje I_D. Aumentando V_{GS} disminuye I_D y viceversa.Puesto que V_{GS} polariza inversamente la unión G-S, el dispositivo presenta una resistencia de entrada extremadamente alta y, a través de la compuerta, sólo fluye una corriente de fuga muy débil.

Naturalmente, I_D puede hacerse circular por una resistencia de carga cualquiera y así obtener, ganancia de voltaje o de potencia.

La resistencia de entrada de un JFET es, típicamente, del orden de 10^9 Ω (miles de MΩ). Por lo tanto, no exige corriente de entrada. Esto lo hace ideal para amplificar señales extremadamente débiles. La resistencia de salida, sin embargo, es relativamente alta (de 50KΩ a 1MΩ).

La capacidad de amplificación de un FET se mide observando el efecto de la tensión de compuerta (V_{GS}) sobre la corriente de drenaje (I_D) para un determinado valor de V_{DS}. La relación entre ambas cantidades (I_D/V_{GS}) se denomina **transconductancia** y se designa por el símbolo **g_m**.

Los fabricantes de FETs siempre especifican el valor de g_m para sus productos. La transconductancia de los JFETs se encuentra típicamente en el rango de 4 a 10 mA/V. Los MOSFET tienden a tener valores mucho más altos, hasta de 20 mA/V inclusive.

Transistores FET de compuerta aislada o MOSFET

Generalidades

La palabra **MOSFET** es una abreviación de **M**etal **O**xide **S**emiconductor **F**ield **E**ffect **T**ransistor (Transistor de Efecto de Campo de Semiconductor de Oxido Metálico) y designa a un FET en el cual la compuerta está eléctricamente aislada del canal mediante una fina capa de óxido metálico. A los MOSFET también se les conoce como transistores IGFET o MOS.

A pesar de estar basado en el mismo principio (modulación de la conductividad del canal por voltaje), el MOSFET es marcadamente diferente del FET de unión. Por ejemplo, el JFET trabaja solamente con una polaridad de la tensión de compuerta. El MOSFET admite tanto voltajes positivos como negativos de compuerta.

Los circuitos con MOSFET son altamente inmunes al ruido, consumen muy poca potencia y son muy flexibles. Además, se prestan a la integración en gran escala. De hecho, los MOSFET han dado origen, entre otras, a las populares familias de circuitos integrados MOS, CMOS, NMOS Y PMOS.

Los transistores *MOSFET son particularmente* sensibles al daño por descarga electrostática (ESD).

Este fenómeno consiste en la creación, generalmente inconsciente, de altos voltajes entre los pines durante su manipulación o transporte. Estos voltajes perforan la capa de óxido metálico y destruyen el dispositivo.

Para proteger un MOSFET del daño por ESD, se recomienda mantener todos sus terminales al mismo potencial (cortocircuitados), hasta que sea instalado finalmente en el circuito. Existen varias formas sencillas de lograrlo: uniendo los pines mediante un alambre delgado, envolviendo los transistores en papel aluminio, etc.

Funcionamiento

En la figura 216 se muestra la estructura, el símbolo y el circuito básico de polarización de un FET de compuerta aislada o MOSFET de canal N. La compuerta está aislada eléctricamente del canal *mediante una fina capa de bióxido de silicio* (SiO_2), la cual le confiere características muy especiales al dispositivo, por ejemplo, una extremadamente alta impedancia de entrada.

Observe que el canal N se forma dentro del sustrato P como en el FET de unión. Sin embargo, en el MOSFET, el sustrato está conectado eléctricamente a la fuente y no a la compuerta.

El FET de compuerta aislada o MOSFET

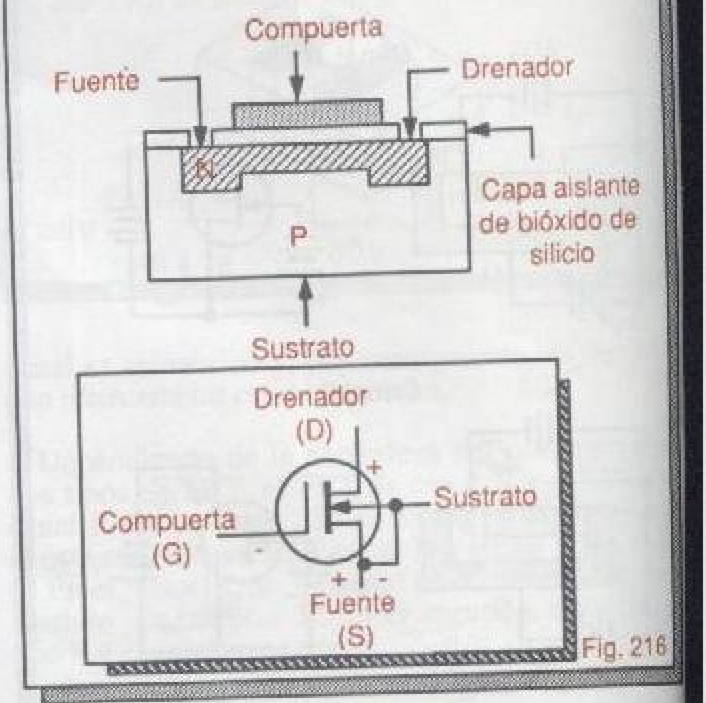

Fig. 216

Aunque la compuerta y el canal no forman una unión PN como en el JFET, la compuerta sigue siendo el terminal que controla la conductividad del canal.

El voltaje entre compuerta y fuente (V_{GS}) puede ser positivo o negativo y controla la concentración de portadores de corriente (electrones, en este caso) en el canal. Si el drenador es positivo con respecto a la fuente y VGS=0, fluye una corriente de drenaje a través del canal.

Cuando V_{GS} se hace negativo, los electrones del canal N son atraídos por los huecos del sustrato P, reduciéndose así la concentración de portadores de corriente dentro del canal. En consecuencia, aumenta la resistencia del canal y se reduce la corriente de drenaje. Se dice, entonces, que el MOSFET está operando en el *modo de agotamiento*, tal como sucede en un FET de unión.

Cuando V_{GS} se hace positivo, los electrones del canal N son rechazados por los huecos del sustrato P, aumentándose así la concentración de portadores *de corriente dentro del* canal. En consecuencia disminuye la resistencia *del canal* y *aumenta la* corriente de drenaje.

Se dice, entonces, que el MOSFET está operando en el *modo de realce*. Este modo de operación no existe en el FET de unión.

Los valores de transconductancia (gm) de los MOSFET son similares a los de los FET de unión. La resistencia de salida sin carga oscila entre 10KΩ y 50KΩ. Así mismo, la impedancia de entrada es extremadamente alta, del orden de 10^{12} Ω (millones de MΩ).

En la figura 217 se muestra el símbolo y la estructura de un MOSFET canal P. Su operación es la misma del MOSFET canal N anterior. Los portadores de corriente, en este caso, son huecos y no electrones. Además, se invierte la polaridad de los voltajes aplicados a los terminales.

Los FET como amplificadores

Al igual que los transistores bipolares, los FET se utilizan principalmente como amplificadores. Las tres configuraciones amplificadoras básicas con FET son los circuito de fuente común, puerta común y drenador común.

En la tabla de la figura 218 se comparan las características más notables de estos montajes. Las conexiones son válidas tanto para FETs de unión (JFET) como para FETs de compuerta aislada (MOSFET).

El MOSFET de canal P

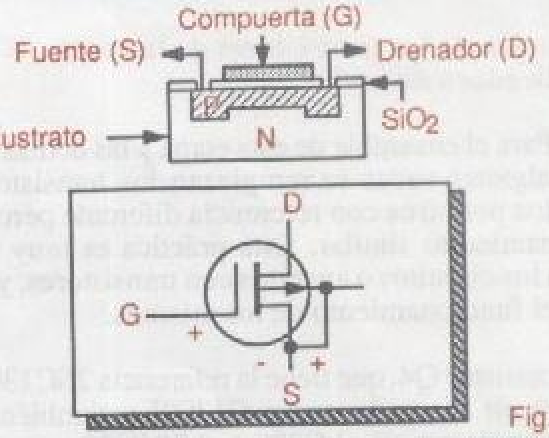

Fig. 217

La configuración fuente común es el montaje más extensamente empleado en circuitos con FETs. Además de proporcionar amplificación de tensión, presenta una resistencia de entrada muy alta y una resistencia de salida moderadamente baja.

Se utiliza en aquellas aplicaciones en las cuales deben conectarse varias entradas a la salida de un circuito sin afectar su funcionamiento.

Comparación de configuraciones amplificadoras básicas con FETs

Parámetros	Compuerta común	Fuente común	Drenador común
Configuración básica	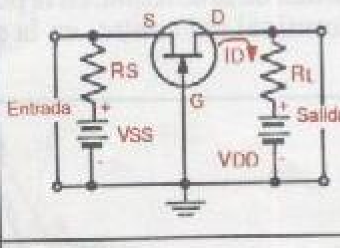 S D RS ID RL G Entrada VSS Salida VDO	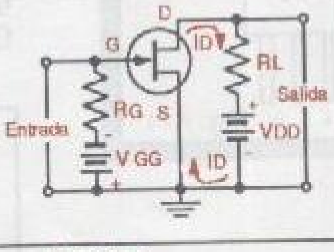 D G ID RL Salida RG S Entrada VDD V GG ID	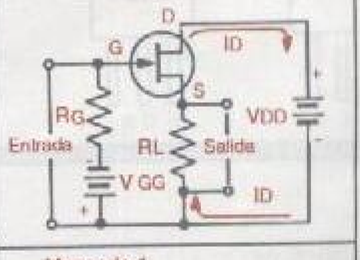 D G ID S RG VDD Entrada RL Salida V GG ID
Ganancia de voltaje	Mayor de1	Mayor de 1	Menor de 1
Ganancia de corriente	Menor de 1	Mayor de 1	Mayor de 1
Ganancia de potencia	Mayor de 1	Mayor de 1	Mayor de 1
Impedancia de entrada	Extremadamente alta	Extremadamente alta	Extremadamente alta
Impedancia de salida	Relativamente alta	Moderadamente alta	Relativamente baja
Aplicación típica	Amplificación de altas frecuencias	Amplificación de bajas frecuencias	Acoplamiento de impedancias

Fig. 218

Actividad práctica Nº 9

Instalación de los transistores de la etapa amplificadora de audio

Nota: Para el ensamble de esta etapa y las demás del radio, algunas veces se remplazan los transistores indicados por otros con referencia diferente pero de funcionamiemto similar. Esta práctica es muy común en los circuitos o aparatos con transistores, y no afecta el funcionamiento de los mismos.

El transistor Q4, que tiene la referencia 2SC1390I puede venir marcado como C1390I y también se puede remplazar por el C829 o el 2SC829 teniendo en cuenta la disposición de los pines que se muestra en la figura A11.

El transistor Q5, originalmente indicado como 2SC735, puede venir marcado como C735 y también se puede remplazar por el C828.

El transistor Q6, originalmente indicado como A564 puede venir marcado como 2SA564 y también se puede remplazar por el C1048G o el 2SA1015.

En la figura A11 se muestra la disposición de los pines de algunos de los transistores mencionados. Si se reciben los transistores originales, éstos se instalan como lo indica el diagrama pictórico.

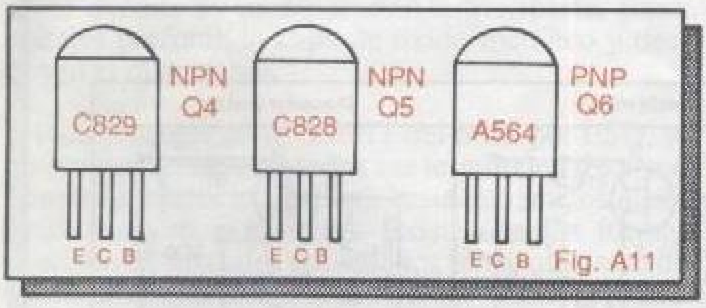

Fig. A11

Materiales necesarios :

1 Transistor NPN 2SC1390I o C1390I (Q4)
1 Transistor NPN 2SC735 o C735 (Q5)
1 Transistor PNP 2SA564 o A564 (Q6)

Paso 1: Tome el transistor Q4, con referencia C1390I, mirándolo de frente. Separe ligeremente los terminales de los extremos y fijándose en el diagrama pictórico, suelde con cuidado el emisor E, terminal derecho, en la puntilla 29. El terminal izquierdo o de la base B, se debe soldar en la puntilla A5 aplicando buena soldadura y fijándose que no se suelten los terminales del condensador C11 y el puente de alambre que va hacia la puntilla A6.

Sosteniendo el transistor con la mano, doble con mucho cuidado y con una pinza delgada, el terminal del centro o colector C y suéldelo en la puntilla A8. Aplique buena soldadura para que queden asegurados los dos terminales que van allí, el colector de Q4 y el puente de alambre que va de la puntilla A8 a la puntilla A12.

Paso 2: Tome el transistor NPN Q5, con referencia C735. Siga el mismo procedimiento del paso enterior conectando y soldando primero en la puntilla A19 el terminal del emisor E situado a la izquierda si miramos el transistor de frente.

Luego suelde la base B, terminal de la derecha, en la puntilla A15 y por último, el colector C, terminal del centro, en la puntilla A18.

Sosteniendo el transistor con la mano, arregle con una pinza la dirección de los terminales para que el transistor quede bien orientado físicamente. Los transistores deben quedar como se muestra en la figura A12.

Paso 3: Tome el transistor PNP Q6, con referencia A564. Siguiendo el mismo procedimiento anterior, instale y suelde sus terminales así: el emisor E, que va a la izquierda, en la puntilla A21; la base B, terminal de la derecha, en la puntilla A16 y el colector C, terminal del centro, en la puntilla A22.

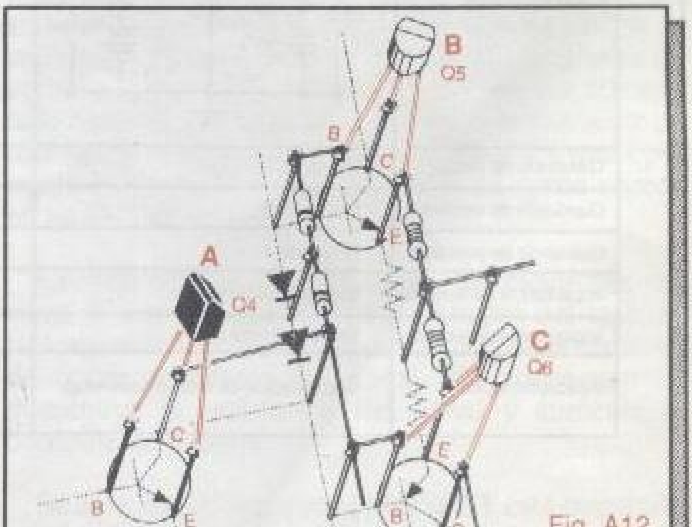

Fig. A12

Amplificadores de Audio

Qué es un amplificador

Un amplificador (figura 219) es un circuito electrónico que recibe una señal de entrada, la procesa internamente y entrega a la salida una señal que es una réplica ampliada de la señal de entrada. La necesidad de amplificar o convertir una señal débil en una señal fuerte, está prácticamente presente en todos los circuitos electrónicos.

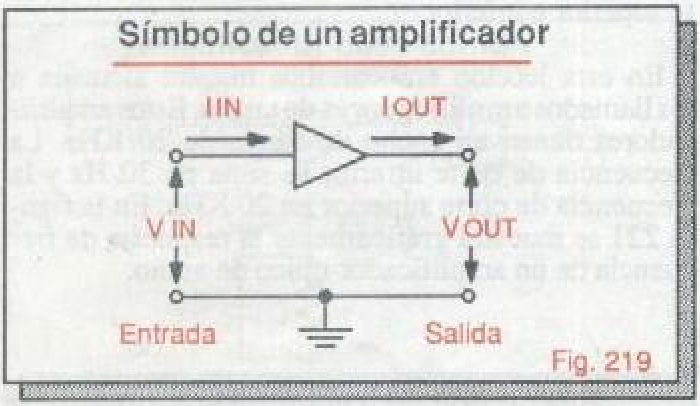
Símbolo de un amplificador

Fig. 219

Los amplificadores sólo manejan señales eléctricas (de corriente, voltaje o potencia) o representaciones eléctricas de cantidades físicas como voz, sonido, temperatura, luz, etc. La conversión de estas últimas en señales eléctricas, o viceversa, la efectúan *transductores* como parlantes, micrófonos, cámaras, termistores, etc.

Las señales eléctricas son amplificadas por dispositivos *activos* como transistores, FETs, circuitos integrados, válvulas, etc., asociados a componentes *pasivos* como resistencias, condensadores, bobinas, transformadores, etc.

Características generales de los amplificadores: Ganancia, impedancia de entrada, impedancia de salida y ancho de banda

La complejidad de un amplificador depende del tipo de señal a amplificar y de la cantidad de amplificación o *ganancia* que ésta requiera. Independientemente de su complejidad, todos los amplificadores tienen en común las siguientes características:

a) Todos utilizan, por lo menos, un dispositivo activo, ya sea transistor bipolar, FET, MOSFET, válvulas, circuito integrado, etc. Los amplificadores que se estudian en esta lección son, en su mayoría, basados en transistores bipolares y, en algunos casos, en FETs.

b) Todos requieren de una fuente de alimentación de CC, doble o sencilla. El amplificador de audio del radio AM CEKIT, por ejemplo, utiliza una batería de 9V.

c) Su operación se puede resumir especificando la ganancia, la impedancia de entrada, la impedancia de salida y la respuesta de frecuencia.

La ganancia es una medida de la cantidad de amplificación. Se define como la relación numérica entre la señal de salida y la señal de entrada. Se habla de ganancia de corriente (A_I), de ganancia de voltaje (A_V) y de ganancia de potencia (A_P). Esto es:

$$A_I = \frac{I_{OUT}}{I_{IN}} \qquad A_V = \frac{V_{OUT}}{V_{IN}} \qquad A_P = \frac{P_{OUT}}{P_{IN}}$$

La ganancia de un amplificador se puede también expresar en decibelios (dB) como sigue:

$$A_I\,(dB) = 20 \log \frac{I_{OUT}}{I_{IN}} \qquad A_V\,(dB) = 20 \log \frac{V_{OUT}}{V_{IN}}$$

$$A_P\,(dB) = 10 \log \frac{P_{OUT}}{P_{IN}}$$

Por ejemplo, si la potencia de entrada de un amplificador es de 0.5W y la potencia de salida es de 5 W, la ganancia en dB es de $10 \times \log_{10} (5/0.5) = 10$ dB. Una relación de potencias de 10 corresponde a un cambio de 10 dB, una relación de potencias de 100 corresponde a un cambio de 20 dB y así sucesivamente.

El oído humano responde logarítmicamente a la relación de potencias y no a su diferencia. Por esta razón, percibimos el mismo cambio de intensidad cuando la potencia de un sonido se incrementa de 0.1 W a 1W, que cuando lo hace de 1W a 10W. En ambos casos, la relación de cambio es de 10 dB.

La impedancia de entrada es la resistencia vista por la fuente que genera la señal que va a ser amplificada. Se define como la relación entre el voltaje y la corriente en los terminales de entrada del amplificador. Esto es:

$$Z_{IN} = \frac{V_{IN}}{I_{IN}}$$

La impedancia de entrada depende no solamente de la resistencia de entrada del elemento activo de amplificación sino también de la presencia de condensadores, resistencias, etc. en el circuito.

La **impedancia de salida** es la resistencia vista por la carga que recibe la señal amplificada. Se define como la relación entre el voltaje y la corriente en los terminales de salida del amplificador. Esto es:

$$Z_{OUT} = \frac{V_{OUT}}{I_{OUT}}$$

La existencia de una impedancia de salida finita causa una pérdida de voltaje en los terminales de salida de un amplificador, cuando se conecta entre los mismos una carga.

En la figura 220 se muestra el circuito equivalente de un amplificador de voltaje con una impedancia de entrada Z_{IN}, una impedancia de salida Z_{OUT} y una ganancia de voltaje A_V. R_{IN} es la resistencia interna de la fuente de señal y R_L el valor de la resistencia de carga.

Circuito equivalente de un amplificador

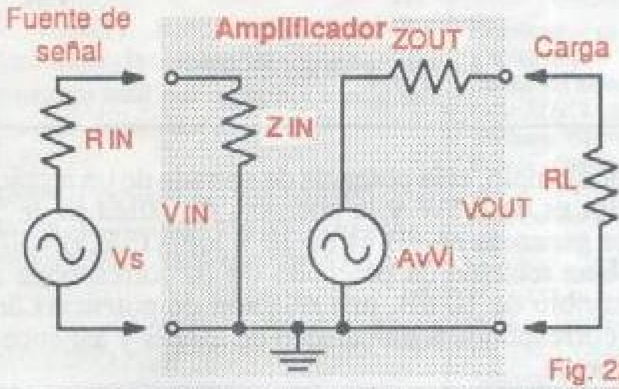

Fig. 220

Para garantizar la *máxima transferencia de voltaje* entre una fuente de señal y una carga, la resistencia de la carga debe ser grande comparada con la resistencia interna de la fuente. Como regla general, $Z_{IN} \geq 10R_{IN}$ y $R_L \geq 10Z_{OUT}$. ($\geq$: mayor o igual).

Así mismo, para garantizar la *máxima transferencia de potencia* entre una fuente de señal y una carga, la resistencia de la carga debe ser igual a la resistencia interna de la fuente. Esto implica $R_L = Z_{OUT}$ y $Z_{IN} = R_{IN}$.

La **respuesta de frecuencia** está íntimamente relacionada con la ganancia. Generalmente, la ganancia de un amplificador no es constante para todas las frecuencias de señal posibles. Esto es, dos o más señales de la misma amplitud y forma de onda, pero de frecuencias diferentes, no reciben necesariamente la misma cantidad de amplificación.

Por esto, cuando se especifica la ganancia de un amplificador, debe también especificarse el rango de frecuencias dentro del cual es válida esa ganancia. Este rango recibe el nombre de **ancho de banda** del amplificador y se expresa en unidades de frecuencia (Hz, KHz, MHz, etc).

El ancho de banda se puede también especificar en términos de los límites superior e inferior del mismo. Estos límites se denominan **frecuencias de corte** superior e inferior.

En esta lección enfocaremos nuestra atención a los llamados **amplificadores de audio**. Estos amplificadores tienen un ancho de banda de 20 KHz. La frecuencia de corte inferior se sitúa en 30 Hz y la frecuencia de corte superior en 20 KHz. En la figura 221 se muestra gráficamente la respuesta de frecuencia de un amplificador típico de audio.

Respuesta de frecuencia de un amplificador de audio

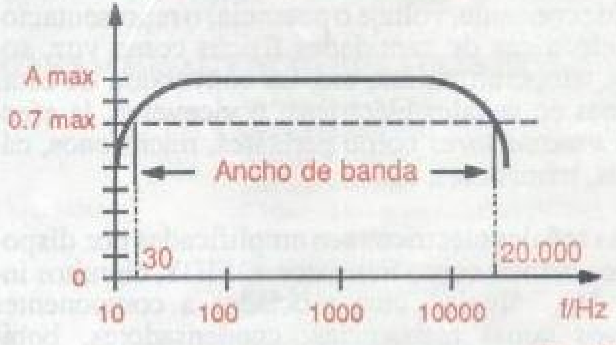

Fig. 221

La banda de 30 Hz a 20 KHz es el rango de frecuencias que puede ser percibido por el oído humano. Las frecuencias por encima de 20 KHz se denominan *ultrasónicas* y las frecuencias por debajo de 30 Hz se denominan *subsónicas*. Los ultrasonidos, por ejemplo, son percibidos por especies como ratas, plagas, mascotas, etc.

Los amplificadores de audio se utilizan en los equipos de sonido tanto domésticos como de alta fidelidad (Hi-Fi), en los estudios de grabación, en los receptores de radio y televisión, en los juegos de video, etc.

Amplificadores de voltaje, de corriente y de potencia.

De acuerdo a su función específica, los amplificadores se clasifican en amplificadores de voltaje, de corriente y de potencia. Las señales manejadas por estos circuitos pueden ser de CA (variables con el tiempo) o de CC (constantes).

Los amplificadores de voltaje elevan el nivel de señales de voltaje. Un ejemplo de amplificador de voltaje es el circuito base común mostrado en la figura 222. Como sabemos, este montaje provee ganancia de voltaje y de potencia pero no de corriente. Se caracteriza por su baja impedancia de entrada y su alta impedancia de salida.

Amplificador base común

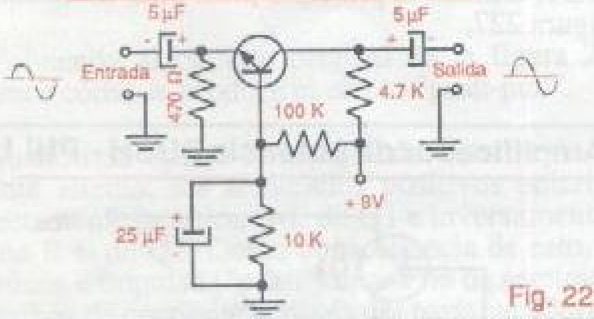

Fig. 222

Los amplificadores de corriente elevan el nivel de señales de corriente. Un ejemplo de amplificador de corriente es el circuito colector común mostrado en la figura 223. Como sabemos, este montaje provee ganancia de corriente y de potencia pero no de voltaje. Se caracteriza por su alta impedancia de entrada y su baja impedancia de salida.

Amplificador colector común

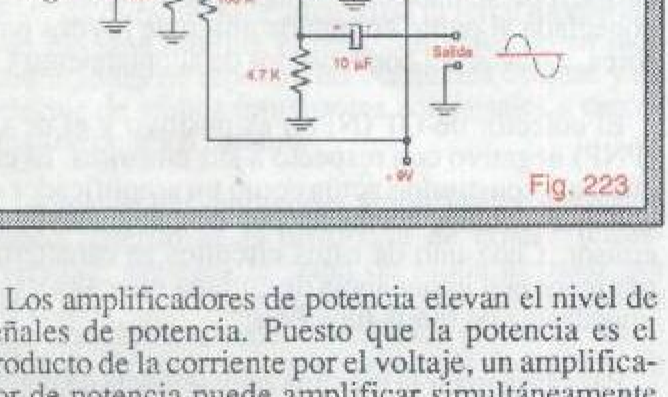

Fig. 223

Los amplificadores de potencia elevan el nivel de señales de potencia. Puesto que la potencia es el producto de la corriente por el voltaje, un amplificador de potencia puede amplificar simultáneamente voltaje y corriente o sólo voltaje o sólo corriente.

Un ejemplo de amplificador de potencia es el circuito emisor común mostrado en la figura 224. Como sabemos, este montaje provee al mismo tiempo ganancia de voltaje, de corriente y de potencia. Se caracteriza por tener una impedancia de entrada relativamente alta y una impedancia de salida relativamente baja.

Amplificador emisor común

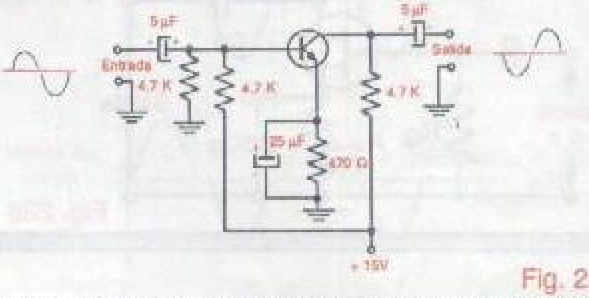

Fig. 224

Amplificadores de audio y de RF

De acuerdo al rango de frecuencias de operación, los amplificadores se clasifican en amplificadores de baja frecuencia y amplificadores de alta frecuencia.

Los de baja frecuencia se denominan **amplificadores de audio** y trabajan con señales desde 0 Hz (CC) hasta 20 KHz. Los de alta frecuencia se denominan **amplificadores de RF** (radio frecuencia) y operan a frecuencias por encima de 20 KHz.

En la figura 225 se muestra el circuito real de un amplificador de audio de dos etapas similar al utilizado en el radio AM CEKIT. La primera etapa o preamplificador, es un amplificador de voltaje y la segunda,es un amplificador de potencia.

Amplificador de audio típico

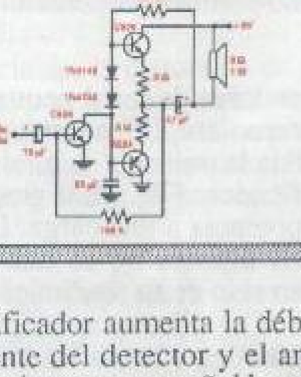

Fig. 225

El preamplificador aumenta la débil señal de entrada proveniente del detector y el amplificador de potencia la convierte en una señal lo suficientemente fuerte como para impulsar el parlante.

En la figura 226 se muestra el circuito práctico de un amplificador de RF con FET que responde a la configuración fuente común. Note la utilización de un circuito sintonizado LC como carga.

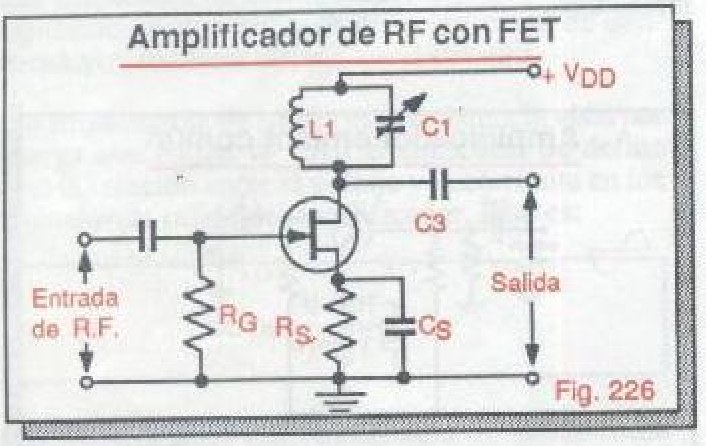

Amplificador de RF con FET

Fig. 226

Los amplificadores de RF son similares a los amplificadores de audio en muchos aspectos. Por ejemplo, deben tener una carga en su circuito de salida y necesitan estar correctamente polarizados.

Sin embargo, se diferencian por su selectividad (capacidad de amplificar sólo una banda estrecha de frecuencias) y otras consideraciones de diseño que generalmente se ignoran a bajas frecuencias. La selectividad, en este caso, la provee un circuito resonante LC.

Amplificadores de pequeña señal y amplificadores de señal grande o de potencia. Rendimiento.

De acuerdo a la magnitud de las señales involucradas, los amplificadores se clasifican en **amplificadores de señal pequeña** y **amplificadores de señal grande**. Estos últimos se denominan también **amplificadores de potencia.**

Los de señal pequeña operan en forma *lineal*, esto es, todas las señales dentro de su ancho de banda reciben la misma cantidad de amplificación o ganancia. En los de señal grande, la señal de entrada es tan fuerte que no permite la operación lineal del dispositivo.

Los amplificadores de señal pequeña se utilizan, básicamente, como amplificadores de voltaje y no es de importancia la potencia suministrada a la carga. Los amplificadores de señal grande se utilizan para entregar potencia a una carga. La ganancia de voltaje en estos últimos no es una consideración importante, pero si lo es su rendimiento.

El **rendimiento** de un amplificador de potencia se define como la relación entre la potencia de la señal de salida y la potencia que realmente le llega a la carga. Se expresa generalmente como un porcentaje. Esto es :

$$\text{Rendimiento (\%)} = \frac{\text{Potencia en la carga}}{\text{Potencia de salida}} \times 100$$

Los amplificadores de señal grande o de potencia se caracterizan por presentar diversos grados de **distorsión** en la forma de onda de la señal de salida. Esto es, en un amplificador de potencia, la forma de onda de la señal de salida no es una réplica fiel de la forma de onda de la señal de entrada.

El amplificador de potencia tipo "push-pull"

La mayoría de los amplificadores de potencia de audio, incluyendo el utilizado en el radio AM CEKIT, utilizan el principio *push-pull* mostrado en la figura 227.

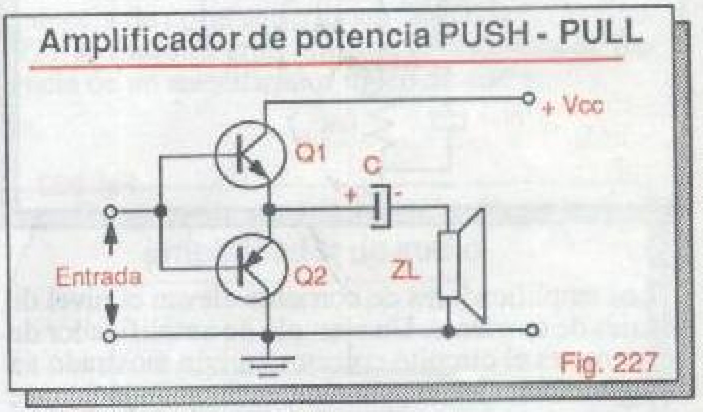

Amplificador de potencia PUSH - PULL

Fig. 227

A esta configuración se le denomina también de *simetría complementaria* debido a que utiliza dos transistores complementarios (NPN y PNP) de características idénticas. Otros nombres son los de etapa en contrafase o amplificador clase B.

Note que la señal de entrada se aplica simultáneamente a las bases de ambos transistores y que la carga, representada en este caso por el parlante, está conectada al punto común de unión de los dos emisores, a través del condensador de acoplamiento C.

El colector de Q1 (NPN) es positivo y el de Q2 (PNP) negativo con respecto a sus emisores. El circuito así constituído actúa como un amplificador de colector común, formado por dos seguidores de emisor. Cada uno de estos circuitos se caracteriza por tener una impedancia de entrada muy alta y una impedancia de salida muy baja.

Para garantizar la máxima transferencia de potencia entre la salida del circuito y la carga, la impedancia de salida del amplificador debe ser igual a la impedancia del parlante. Si este no es el caso, debe utilizarse entre ambos un transformador de acoplamiento, como se muestra en la figura 228.

Amplificador PUSH - PULL con transformador

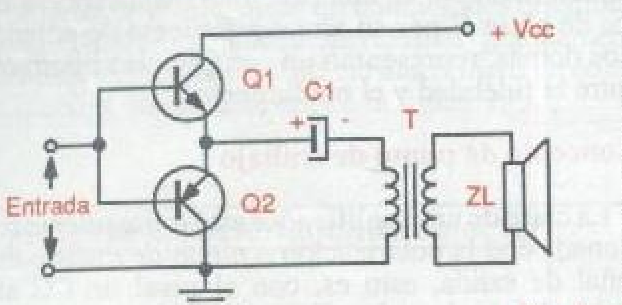

Fig. 228

Volviendo al circuito original de la figura 227, veamos cómo se produce el efecto "push-pull".

Cuando se aplica una señal de entrada de corriente alterna, los semiciclos positivos polarizan directamente la unión B-E de Q1 e inversamente la unión B-E de Q2. Como consecuencia de esto, Q1 conduce e impulsa ("push") una serie de semiciclos positivos de corriente a través del parlante. Durante estos semiperíodos, Q2 permanece bloqueado.

Durante los semiciclos negativos de la señal de entrada, la unión B-E de Q2 queda polarizada directamente y la unión B-E de Q1 queda polarizada inversamente. Como consecuencia de esto, Q2 conduce y arrastra ("pull") los semiciclos negativos de corriente restantes a través del parlante. Durante estos semiperíodos, Q1 permanece bloqueado.

De este modo, Q1 y Q2 manejan conjuntamente cada ciclo completo de la señal de entrada. En la figura 229 se muestran las formas de onda de las señales involucradas en el proceso.

Note que cada transistor sólo conduce corriente de salida durante un semiciclo de la señal de entrada y se bloquea durante el otro. Sin señal de entrada (condiciones de reposo), las corrientes de base y de colector de ambos transistores son iguales a cero y ninguno de ellos conduce.

La mayor desventaja del amplificador push-pull básico descrito es la *distorsión de cruce* (crossover) que presenta la señal de salida al pasar por cero. Este fenómeno se debe a que cada transistor sólo comienza a conducir cuando el voltaje de la señal de entrada supera el potencial de barrera de cada unión B-E (0.6V).

Formas de onda del proceso PUSH-PULL

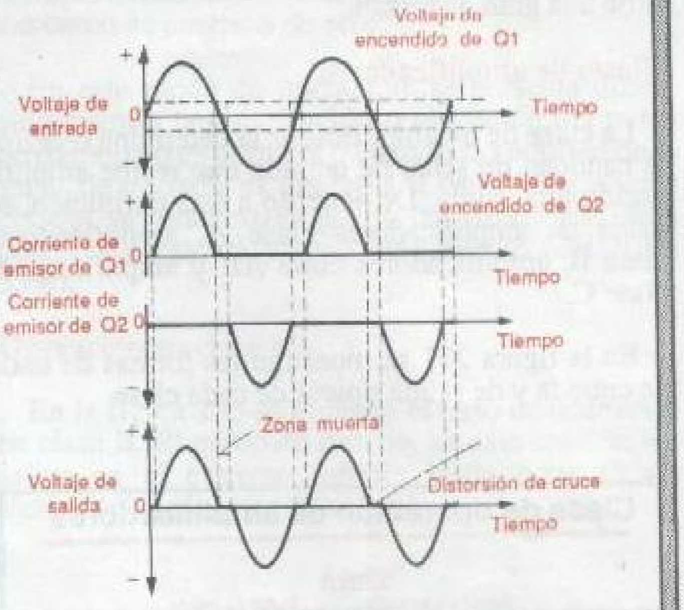

Fig. 229

La distorsión de cruce se puede corregir aplicando una ligera polarización directa a las uniones B-E de ambos transistores, mediante dos resistencias de valores apropiados, como se muestra en la figura 230. En condiciones de reposo (sin señal de entrada) las bases de Q1 y Q2 están ±0.6V por encima de sus emisores.

Amplificador PUSH - PULL polarizado

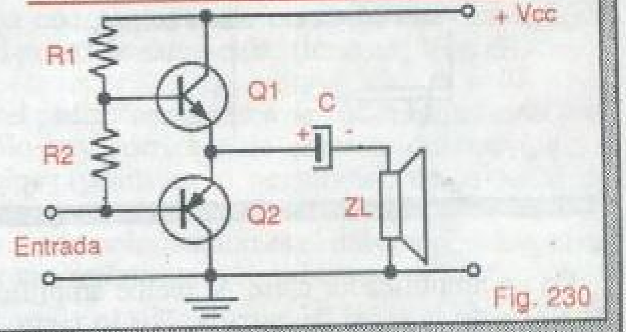

Fig. 230

Si R2 es variable, la distorsión de cruce se puede ajustar a su valor mínimo. En el caso del radio AM CEKIT, R1= 680 Ω y R2 está representada por la resistencia interna de dos diodos 1N4148 en serie. Esto se hace con el fin de proveer, además, estabilidad térmica.

El requisito más importante de un amplificador de potencia push-pull es que Q1 y Q2 sean una pareja complementaria. Esto es, Q1 y Q2 deben tener exactamente las mismas características eléctricas.

De otro modo, cada uno amplificaría de forma diferente el semiciclo que le corresponde y se causaría una gran distorsión.

Clases de amplificadores.

La **clase** de un amplificador la determina o define la cantidad de señal de entrada que recibe amplificación a la salida. De acuerdo a esta definición, se habla de amplificadores clase **A**, amplificadores clase **B**, amplificadores clase **AB** y amplificadores clase **C**.

En la figura 231 se muestran las formas de onda de entrada y de salida típicas de cada clase.

Clase de operación de amplificadores

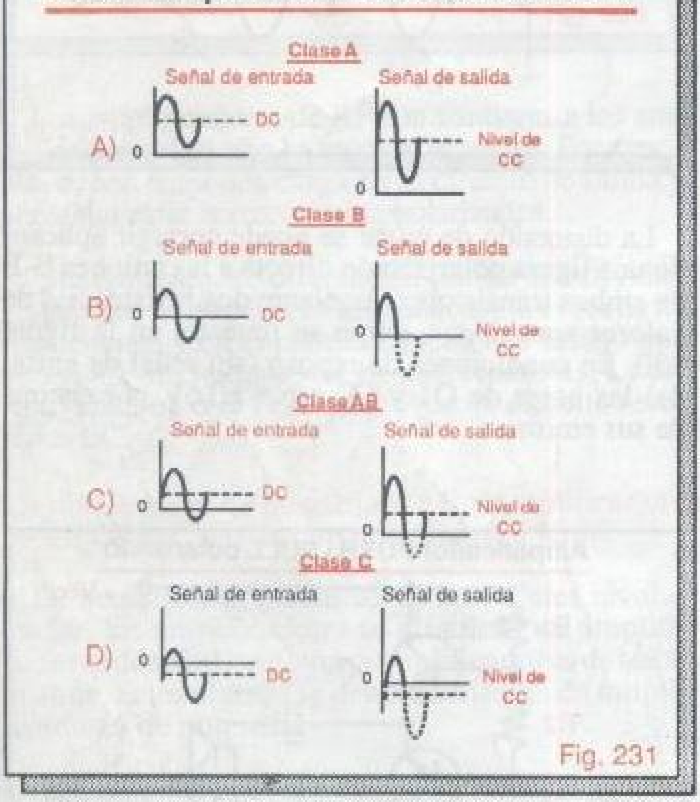

Fig. 231

En un amplificador clase **A**, recibe amplificación el 100% de la señal de entrada. Por lo tanto, la señal de salida está presente durante el ciclo completo de la señal de entrada. (Figura 231A).

En un amplificador clase B, recibe amplificación el 50% de la señal de entrada. Por lo tanto, la señal de salida está presente sólo durante los semiciclos positivos o negativos de la señal de entrada. (Figura 231B).

En un amplificador clase AB, recibe amplificación más del 50% de la señal de entrada. Por lo tanto, la señal de salida está presente durante más de un semiciclo de la señal de entrada. (Figura 231C).

En un amplificador clase C, recibe amplificación menos del 50% de la señal de entrada. Por lo tanto, la señal de salida está presente durante menos de un semiciclo de la señal de entrada. (Figura 231D).

Todos los amplificadores de señal pequeña (y algunos de potencia) operan en clase A y todos los de señal grande en clase B, AB o C. Los de clase A se caracterizan por su alta fidelidad (baja distorsión) y los de clase C por su alto rendimiento de potencia. Los demás, representan un compromiso intermedio entre la fidelidad y el rendimiento.

Concepto de punto de trabajo

La clase de un amplificador está íntimamente relacionada con la polarización o *punto de trabajo* de la señal de salida, esto es, con el nivel de CC alrededor del cual se desplaza esta última. Para comprender este concepto, consideremos el caso de un amplificador emisor común genérico, como el que se muestra en la figura 232.

Punto de trabajo de un amplificador emisor común

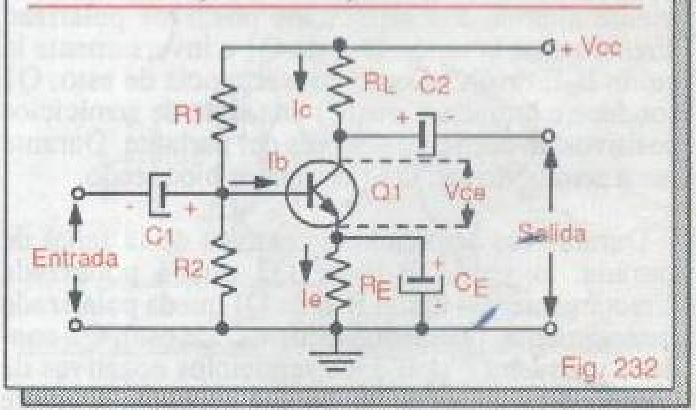

Fig. 232

Este circuito puede operar como amplificador de voltaje, de corriente o de potencia. La señal de entrada se aplica entre base y tierra a través de C1. La señal de salida aparece entre colector y tierra y se aplica a la carga mediante C2.

C_1 y C_2 son *condensadores de acoplamiento*. Su objetivo es eliminar el nivel de CC presente en ambas señales y transferir sólo su componente de CA. C_E es el condensador de emisor y su propósito es mantener constante el voltaje sobre R_E, enviando a tierra cualquier variación.

R_1, R_2, R_L y R_E polarizan correctamente las dos uniones del transistor y fijan el punto de trabajo del amplificador. Este último se refiere a los valores de corriente y de voltaje en el circuito cuando no existe señal de entrada. A estas condiciones se les denomina *estáticas* o de reposo y se identifican mediante el subíndice Q (I_{BQ}, I_{CQ}, V_{CEQ}, etc).

Las condiciones en presencia de señal se denominan *dinámicas*. Los valores de IB, IC, VCE, etc., en condiciones dinámicas, varían por encima y por debajo de sus respectivos valores de reposo, dependiendo de la amplitud de la señal de entrada y de la ubicación del punto de trabajo.

La forma como se comporta IC para diferentes valores de IB y VCE, tanto en condiciones dinámicas como estáticas, se puede representar gráficamente mediante un juego de *curvas características* de colector como el que se muestra en la figura 233.

Curva característica de colector

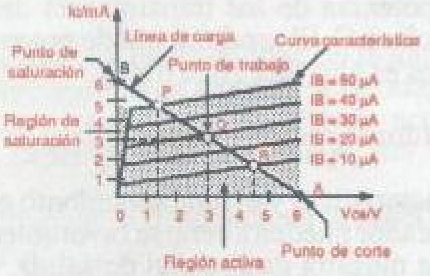

Fig. 233

Los puntos extremos de operación definidos por la *recta de carga* se denominan puntos de corte y de saturación, respectivamente. En el punto de saturación o de conducción, IC es máxima y VCE es cero. En el punto de corte o de no conducción, IC es cero y VCE es máximo e igual al voltaje de alimentación (VCC).

Operación en clase A

La posición del punto Q sobre la recta de carga determina la clase de operación (A, B, AB, C) del amplificador. En la figura 234 se muestra el caso de operación en clase A. El punto de trabajo se localiza en el centro de la recta de carga.

Operación en clase A

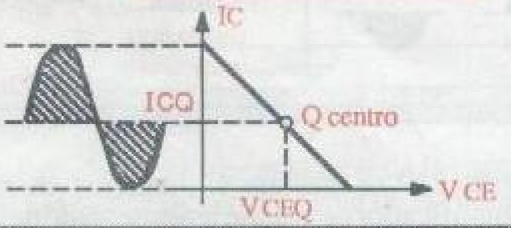

Fig. 234

En general, los amplificadores clase A a transistores trabajan dentro de la región activa o tramo lineal de las curvas características de colector.

Además, la etapa está polarizada de tal modo que siempre existe corriente de colector, tanto en presencia como en ausencia de señal.

En este modo de operación, se presenta distorsión cuando la polarización o ubicación del punto de trabajo no es correcta o cuando la amplitud de la señal de entrada es excesiva. Los amplificadores clase A se caracterizan por su baja distorsión, baja potencia de salida y bajo rendimiento (del 20 al 35%).

Operación en clase B

En la figura 235 se muestra el caso de operación en clase B. El punto de trabajo, en este caso, se localiza en el extremo inferior o de corte (IC=0, VCE=VCC) de la recta de carga.

Operación en clase B

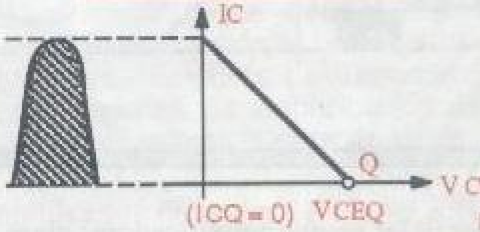

Fig. 235

En general, las etapas amplificadoras clase B se polarizan de modo que, sin señal de entrada, se obtenga IC=0 o VCE=0. Esto es, trabajan parcialmente en la zona activa, coincidiendo el punto de trabajo con el punto de corte (IC=0, VCE=máx) o con el punto de saturación (IC=máx, VCE=0).

En el primer caso, o sea de polarización en el corte, sólo hay corriente de colector durante los semiciclos, (positivos o negativos), de la señal de entrada que polarizan directamente la unión B-E. Este tipo de polarización es el más empleado y el de mejor rendimiento.

En el segundo caso, o de polarización en saturación, hay una corriente de colector muy intensa en ausencia de señal de entrada. Prácticamente, toda la tensión de alimentación aparece en los extremos de la resistencia de carga.

La tensión colector-emisor experimenta variaciones sólo durante los semiciclos (positivos o negativos) de la señal de entrada que disminuyen la polarización directa de la unión B-E. Este tipo de polarización no es muy empleado.

La operación en clase B se utiliza en amplificadores de potencia de baja frecuencia. Se caracteriza por su buen rendimiento (del 50 al 60%). Utilizando dos amplificadores clase B en contrafase, se pueden amplificar ambos ciclos de la señal de entrada, como sucede en el amplificador push-pull descrito en una sección anterior.

Operación en clase AB

En la figura 236 se muestra el caso de operación en clase AB. El punto de trabajo, en este caso, se localiza cerca del punto de corte pero podría localizarse también cerca del punto de saturación.

Operación en clase AB

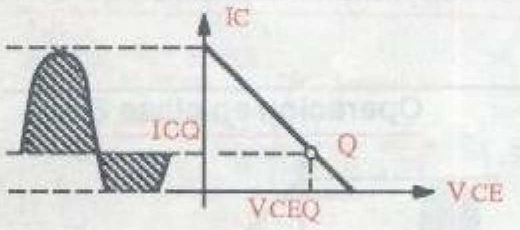

Fig. 236

En general, las etapas amplificadoras clase AB se polarizan de modo que, sin señal de entrada, I_C o V_{CE} se hagan iguales a cero durante menos de un semiciclo de la señal de entrada.

Esto es, los amplificadores clase AB trabajan parcialmente en la zona activa, pero durante mayor tiempo que los de clase B. En los demás aspectos, son similares. Por lo tanto, todo lo dicho anteriormente para clase B es válido también para clase AB. Este modo de operación es el que utiliza el amplificador de potencia del radio AM CEKIT.

Los amplificadores clase AB representan un compromiso intermedio entre la baja distorsión y el bajo rendimiento de los amplificadores clase A y el alto rendimiento y la alta distorsión de los amplificadores clase B.

Operación en clase C

En la figura 237 se muestra el caso de operación en clase C. El punto de trabajo, en este caso, está localizado más allá del punto de corte pero podría localizarse también más allá del punto de saturación.

En general, las etapas amplificadoras clase C se polarizan de modo que, sin señal de entrada, I_C o V_{CE} se hagan iguales a cero durante más de un semiciclo de la señal de entrada.

Operación en clase C

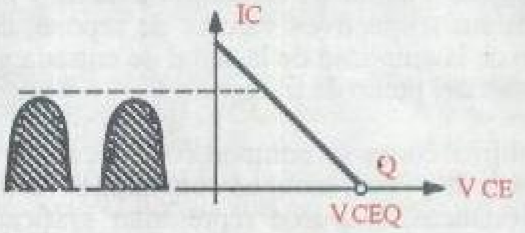

Fig. 237

Este modo de operación no se utiliza para la amplificación de bajas frecuencias, por la alta distorsión que genera. Casi todas las etapas amplificadoras de potencia de los transmisores de RF funcionan en clase C. El rendimiento de potencia de estos circuitos es alto (del 70 al 85%).

Amplificadores realimentados

Las características y el funcionamiento general de un amplificador pueden alterarse favorablemente tomando una muestra de la señal de salida y enviándola nuevamente a la entrada, junto con la señal de entrada original. A este proceso se le denomina *realimentación* o *feedback*.

La realimentación es *positiva* cuando la señal retornada está en fase con la señal de entrada y negativa cuando está fuera de fase. En el primer caso, las dos señales se suman y en el segundo se restan. Ambas situaciones se ilustran en la figura 238. La realimentación positiva debe evitarse porque produce inestabilidad y oscilaciones.

Realimentaciones positiva y negativa

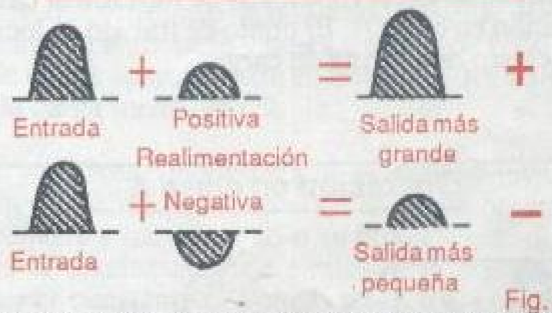

Fig. 238

En la figura 239 se muestra el diagrama en bloque de un amplificador de voltaje realimentado negativamente. Sin realimentación, la ganancia de voltaje del circuito es **A** y con realimentación es **A_o**

La ganancia con realimentación, A_o, es siempre inferior a la ganancia sin realimentación, A. A pesar

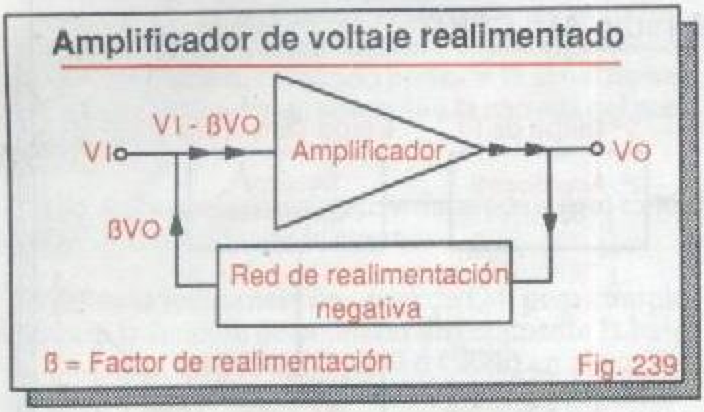
Amplificador de voltaje realimentado

Fig. 239

de reducir la ganancia, la realimentación negativa introduce varios efectos deseables. Estos son algunos de ellos:

a) Incrementa la estabilidad de la ganancia final esto es, la hace menos sensible a las variaciones en los parámetros de los dispositivos activos, a los cambios de temperatura, de voltaje, etc.

b) Disminuye la distorsión en la señal de salida.

c) Mejora el ancho de banda, esto es, se puede amplificar un rango más amplio de frecuencias con la misma ganancia.

d) Se pueden ajustar las impedancias de entrada y de salida a cualquier valor deseable.

En la figura 240 se muestra un ejemplo de amplificador realimentado con transistor NPN. El circuito utiliza realimentación negativa de voltaje. Esta última se realiza a través de RB.

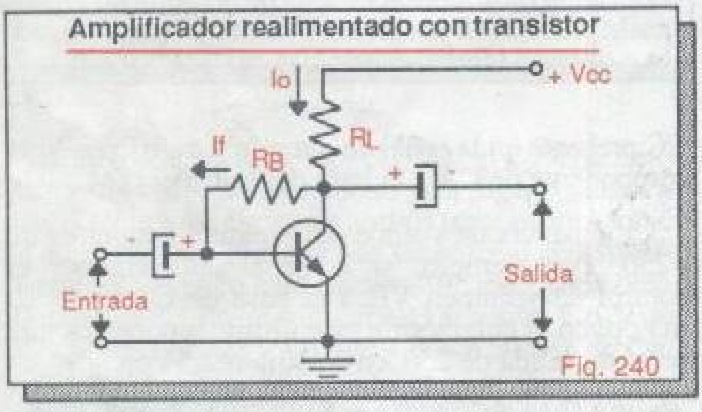
Amplificador realimentado con transistor

Fig. 240

Io es la componente alterna de la corriente de salida e IF la componente alterna de la corriente de realimentación. Esta última se superpone a la corriente de polarización directa de la base.

En la figura 241 se muestra un ejemplo de amplificador realimentado con JFET canal N. El circuito utiliza realimentación negativa de corriente.

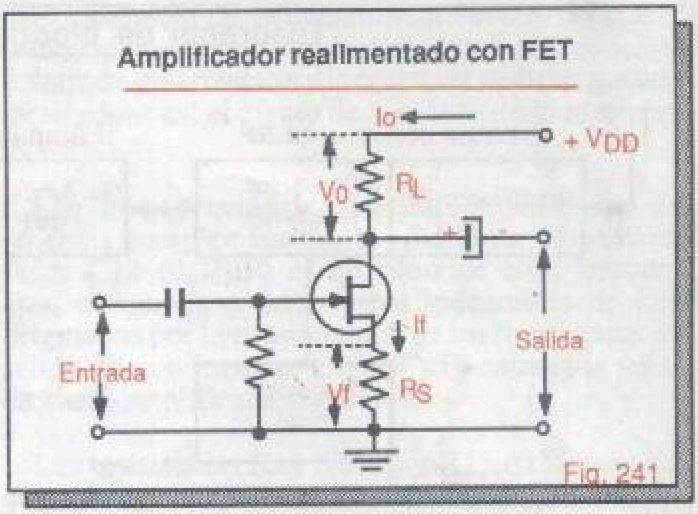
Amplificador realimentado con FET

Fig. 241

Esta última se realiza a través de Rs. Vo es la componente alterna del voltaje de salida y VF la componente alterna del voltaje de realimentación.

El amplificador de audio del radio AM CEKIT utiliza realimentación negativa con el fin de minimizar el fenómeno de distorsión de cruce (cross-over) que presenta la señal de salida. La distorsión de cruce es típica del esquema de amplificación de potencia (push-pull) utilizado en la etapa de salida.

El amplificador de audio del Radio AM CEKIT

El radio AM CEKIT utiliza amplificadores de RF en las etapas de conversión y de FI o frecuencia intermedia, y amplificadores de audio en su etapa final. El diagrama simplificado de bloques del sistema se muestra en la figura 242. La forma como se procesan las diferentes señales se explicará en lecciones posteriores.

En la figura 243 se muestra un diagrama en bloques del amplificador de audio utilizado por el radio AM CEKIT. Consta básicamente de dos etapas: un preamplificador y un amplificador de potencia.

El preamplificador es un amplificador de voltaje de señal pequeña que recibe la débil señal de audio proveniente del detector y la levanta a un nivel suficientemente alto como para excitar el parlante de la etapa de potencia. En este último, las variaciones de corriente se convierten en sonidos.

Otra función del preamplificador es la de reducir el nivel de ruido de la señal de entrada y minimizar la distorsión de la señal de salida. Esto gracias a la existencia de realimentación negativa entre la entrada y la salida del sistema completo.

El amplificador de potencia es un amplificador de señal grande push-pull que recibe la señal proveniente del preamplificador y la convierte en una señal lo suficientemente fuerte como para impulsar el

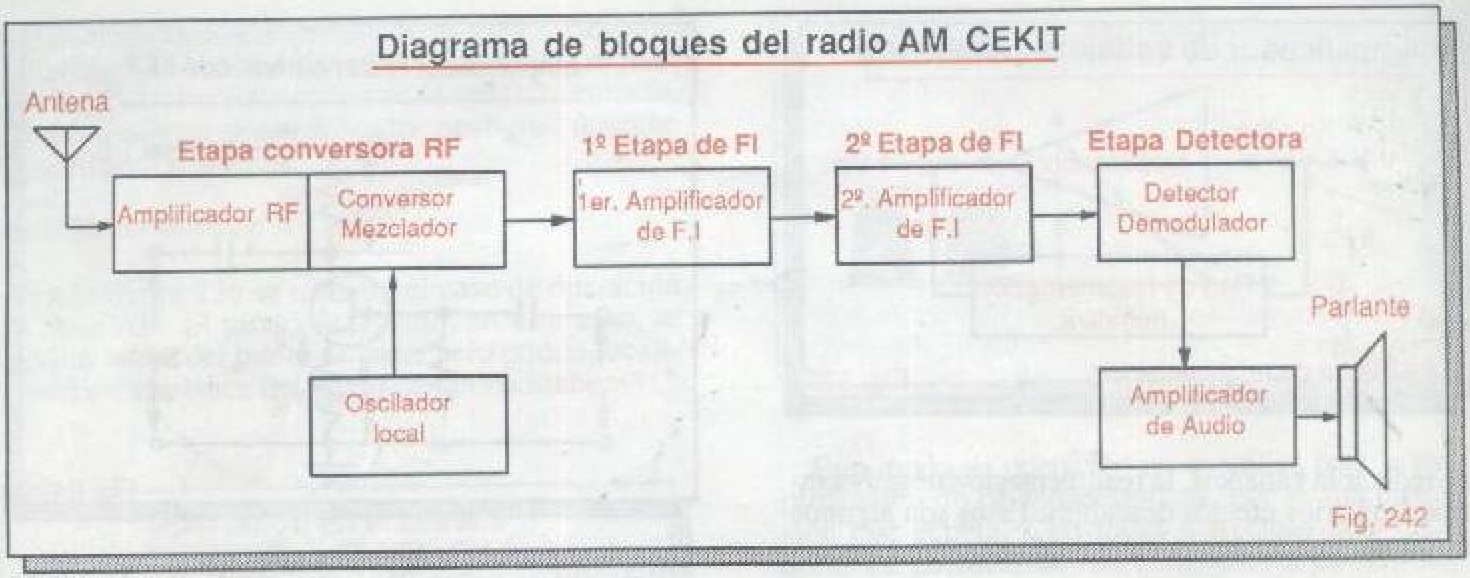
Diagrama de bloques del radio AM CEKIT

Antena
Etapa conversora RF
Amplificador RF
Conversor Mezclador
Oscilador local
1º Etapa de FI
1er. Amplificador de F.I
2º Etapa de FI
2º. Amplificador de F.I
Etapa Detectora
Detector Demodulador
Amplificador de Audio
Parlante

Fig. 242

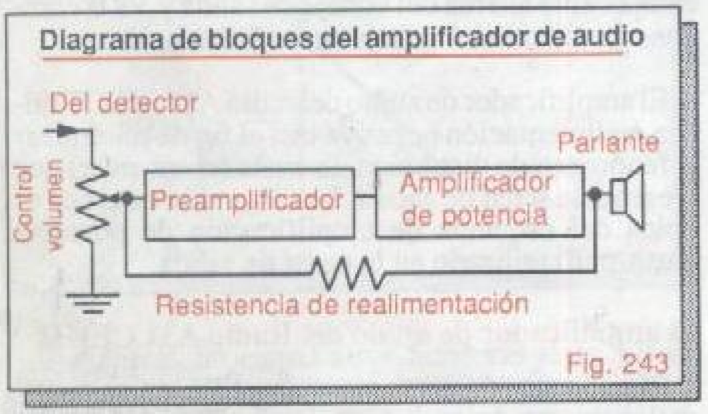
Diagrama de bloques del amplificador de audio

Del detector
Control volumen
Preamplificador
Amplificador de potencia
Parlante
Resistencia de realimentación

Fig. 243

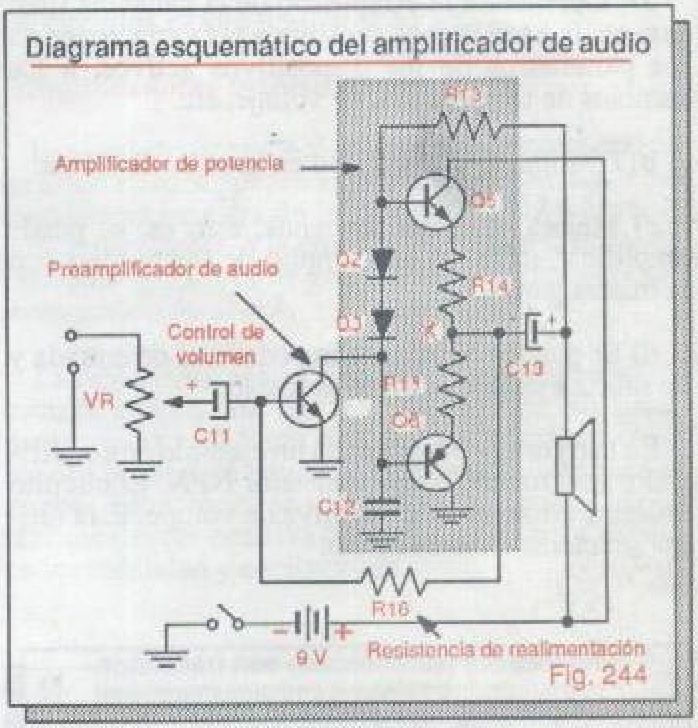
Diagrama esquemático del amplificador de audio

R13
Amplificador de potencia
Q5
D2
Preamplificador de audio
R14
D3
X
C13
Control de volumen
+
VR
C11
R11
Q6
C12
R16
-
+
9 V
Resistencia de realimentación

Fig. 244

parlante y convertir así la información eléctrica de audio en información sonora.

En la figura 244 se muestra el diagrama esquemático completo de nuestro amplificador de audio.

Se han delimitado las áreas correspondientes al preamplificador y al amplificador de potencia para mayor claridad.

La descripción de cada etapa se realiza en las siguientes secciones.

Operación del preamplificador

El preamplificador de audio del radio AM CEKIT es clase A y responde a la configuración emisor común. El elemento activo de esta etapa es el transistor Q4 (NPN, 2SC1390 o su reemplazo, el C829).

El preamplificador recibe, a través del condensador C11, una señal de audio, de 0 a 500 mVpp de amplitud. Esta señal es entregada por el potenciómetro VR (control de volumen) y éste, a su vez, la recibe del detector. (Vpp=Voltios pico a pico).

El condensador C11, de 10µF, actúa como condensador de acoplamiento, bloqueando el nivel de CC presente en la señal de entrada y transfiriendo la componente de CA a la base del transistor Q4.

La ganancia de voltaje del preamplificador es de ≈250. Por ejemplo, si la señal aplicada por el control de volumen VR a la base de Q4 es de 10 mVpp, en el colector de este último tendremos una señal de salida de 250 x 10mVpp = 2.5Vpp.

La tensión de polarización directa de la unión B-E de Q4, se obtiene de los semiciclos positivos de la señal de salida, (punto X), a través de R16 (150 KΩ).

La tensión de colector de Q4 se obtiene del polo positivo de la batería a través del parlante, R13 (680 Ω) y los diodos D1 y D2 (1N41480), polarizados en directo.
(2) (3)

Además de actuar como resistencia de polarización de base, R16 actúa también como resistencia de realimentación, enviando parte de la señal de salida del amplificador de potencia a la entrada del preamplificador.

Esta realimentación negativa tiene por objeto minimizar la distorsión de la etapa.

R13 es la resistencia de carga de Q4, pero cumple también la función de polarizar directamente la base de Q5 (NPN 2SC735 ó C1959 ó C828) en el amplificador de potencia. A su vez, D2 y D3 proporcionan el voltaje de polarización directo de Q6 (PNP, 2SCA1048G ó 2SA1015 ó A564) y proveen estabilidad térmica.

Operación del amplificador de potencia

El amplificador de potencia del radio AM CEKIT es del tipo colector común, clase AB y responde a la configuración push-pull o de simetría complementaria, explicada en una sección anterior. Los elementos activos de esta etapa son Q5 (NPN) y Q6 (PNP).

Cada uno de estos transistores actúa como un seguidor de emisor, presentando una alta impedancia de entrada a la señal proveniente del preamplificador y una baja impedancia de salida al parlante.

El resultado es una máxima transferencia de voltaje a la entrada y una máxima transferencia de potencia a la salida.

Cada transistor conduce sólo durante un semiciclo de la señal de entrada. En particular, Q5 suministra corriente de emisor durante los semiperíodos positivos y Q6 lo hace durante los negativos. Mientras Q5 conduce, Q6 está parcialmente cortado y viceversa.

La resistencia R13, de 680 Ω y D2-D3, referencia 1N4148, proporcionan una ligera polarización directa a las bases de ambos transistores con el fin de minimizar la distorsión de cruce o *crossover* , tal como se explicó para los amplificadores de potencia push-pull. En particular, R13 aplica 5.6 V de CC a la base de Q5 y D~~1~~2-D~~2~~3 4.2 V a la base de Q6.

Esto último es así porque a través de cada uno de estos diodos se presenta una caída de 0.7 V (4.2V=5.6V-1.4V). D2 y D3 proveen también la estabilización térmica de la etapa de salida.

Al aumentar la temperatura ambiental, las corrientes de colector-emisor de Q5 y Q6 tienden a aumentar, debido a que también aumentan las corrientes de fuga de las uniones B-C y B-E de los mismos.

El aumento de temperatura, a su vez, reduce la resistencia interna de ambos diodos. Esto reduce el voltaje de polarización y la corriente de base, contrarrestándose así el efecto de aumento de la corriente de colector.

Los demás componentes del amplificador de potencia cumplen funciones auxiliares. El condensador C12 (0.03µF) es un filtro de altas frecuencias, que envía a tierra picos indeseables de RF, originados por la conmutación de los transistores de salida. El condensador C13 (47µF) entrega la señal de audio amplificada al parlante.

Las resistencias R14 (0.5Ω) y R15 (0.5Ω) actúan como resistencias de emisor de Q5 y Q6, respectivamente. Su objetivo es mantener estable la polarización directa de ambos transistores. La resistencia R16 (150KΩ) actúa como resistencia de realimentación. Su propósito es minimizar la distorsión en la señal de salida.

La etapa de salida está diseñada para garantizar máxima transferencia de potencia cuando se utiliza como carga nominal un parlante de 8Ω. Si se utiliza un parlante de otra impedancia, debe utilizarse un transformador de acoplamiento. Con un parlante de 8Ω, la máxima potencia de salida del amplificador de audio es de ≈1.25W.

Nota:

La máxima potencia de salida que puede extraerse de un amplificador push-pull como el utilizado por el radio AM CEKIT se evalúa mediante la siguiente fórmula:

$$\boxed{P_{MAX} = \frac{V^2_{CC}}{8 Z_L}}$$

Vcc es el voltaje de la fuente de alimentación y ZL la impedancia del parlante. En nuestro caso, Vcc=9V y ZL=8 Ω. Por lo tanto, la máxima potencia de salida del mismo, en vatios, es:

$$\boxed{P_{MAX} = \frac{9^2}{8 \times 8} W = \frac{81}{64} W = 1.26\ W}$$

Si se reduce el voltaje de alimentación a 6V por ejemplo, la potencia de salida máxima desciende a ≈ 0.5W. Si en vez de un parlante de 8Ω se utiliza uno de 40Ω, esta potencia sería sólo de 0.25W.

Actividad práctica Nº 10

Construcción de un inyector de señales

En esta actividad daremos todas las instrucciones necesarias para la construcción de un instrumento muy sencillo pero muy útil para la prueba y reparación de los aparatos de radio, el inyector de señales.

Como lo habíamos mencionado, el inyector de señales produce un tren de ondas que se puede inyectar o introducir a un amplificador de audio o de radiofrecuencia para comprobar su funcionamiento.

De esta manera podemos verificar si lo hemos ensamblado bien o si hay algún componente defectuoso o también podemos reparar aparatos de radio o amplificadores de sonido con él.

Con unos pocos componentes ya muy conocidos, como resistencias, condensadores y transistores se forma el circuito inyector de señales, que es un *oscilador* o sea un generador de ondas. El tema completo de los osciladores se tratará próximamente en este curso.

El circuito completo para el inyector de señales se muestra en la figura A13.

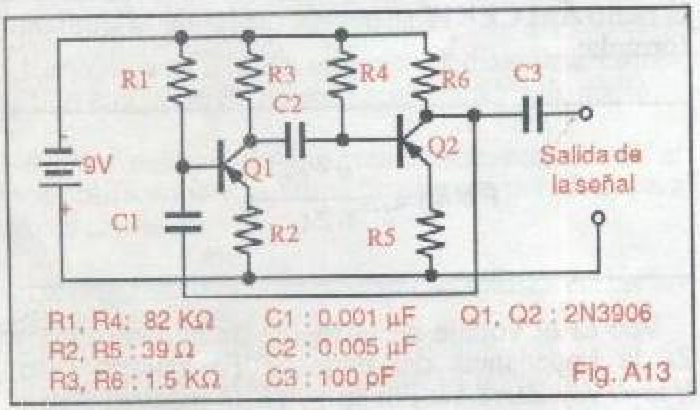

Fig. A13

Este circuito se puede ensamblar en un circuito impreso cuyo dibujo reproducimos en la figura A14 e instalando sus componentes como se indica en la figura A15. El circuito impreso se puede fabricar utilizando el Kit 550 de CEKIT que incluye la lámina de baquelita cobrizada, el lapicero especial y el ácido para rebajar el cobre.

Para fabricar el inyector de señales debe conseguir los materiales que se detallan en la siguiente lista.

Lista de materiales

2 Transistores PNP 2N3906
1 Condensador de cerámica de 0.001 µF
1 Condensador de cerámica de 0.005 µF
1 Condensador de cerámica de 100 pF
2 Resistencias de 82 KΩ 1/2 W
2 Resistencias de 39 Ω 1/2 W
2 Resistencias de 1.5 KΩ 1/2 W
1 Conector para batería de 9 voltios
1 Circuito impreso CEKIT K13

Todos estos materiales, inlcuyendo el circuito impreso y un completo manual de instrucciones se encuentran en el Kit K13 de CEKIT.

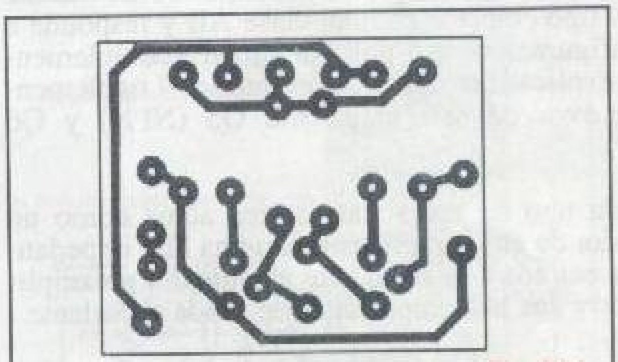

Fig. A14

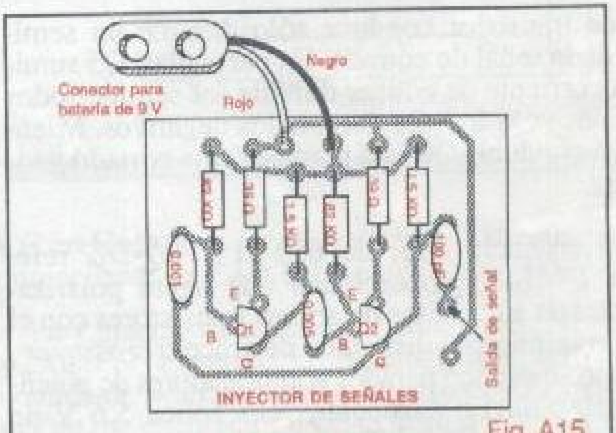

Fig. A15

Para darle una mejor presentación y funcionabilidad a este circuito, lo podemos alojar en una caja pequeña de metal o plástico y alimentarlo por medio de pilas o una batería de 9 voltios. En sus terminales de salida deben ir los cables para conectar la señal a los circuitos. Un cable con caimán para la tierra y una punta de prueba para la salida.

Lección 19

Cómo se produce la señal de una emisora

Fuentes de señal, osciladores y modulación

Hemos visto en la introducción de este curso, que un transmisor de radio mezcla una señal de alta frecuencia, o radiofrecuencia (RF), con una señal de baja frecuencia o audiofrecuencia (BF o AF) produciendo una señal *modulada* que se emite por la antena.

Qué es y cómo trabaja una emisora de radio

Con el fin de poder entender mas fácilmente cómo trabaja un receptor de radio, vamos a explicar qué es y cómo trabaja una emisora de radio.

En otras palabras, vamos a estudiar cómo se captan los sonidos de las palabras y la música en una emisora, cómo se convierten en corriente eléctrica para llevarlos a aparatos electrónicos y cómo estos aparatos producen las ondas electromagnéticas que pueden viajar por el espacio.

En la figura 245 podemos observar la estructura general de una emisora de radio.

Fuentes de señales de audio

Las principales fuentes de sonido o audio que se encuentran en las emisoras de radio son: los micrófonos, los tocadiscos (de Long Play y Compact Disc) y las grabadoras de cinta magnética (Cassete o Rollo).

Las palabras de los locutores son recogidas por uno o varios micrófonos, la música es captada por los tocadiscos o las grabadoras y los avisos comerciales pueden ser originados en forma directa por un locutor o estar grabadas en cintas.

Los micrófonos

Un micrófono es un dispositivo que convierte ondas sonoras (voz, música, lenguaje, etc.) en señales eléctricas de voltaje o de corriente equivalentes.

Se ha determinado experimentalmente que las frecuencias de la voz humana están comprendidas principalmente en el rango de 300 a 3500 Hz.

Características generales de los micrófonos

Los micrófonos poseen ciertas características básicas que determinan su calidad. Las más importantes son la respuesta de frecuencia, la direccionalidad, la impedancia, la sensibilidad y la inmunidad al ruido.

La *respuesta de frecuencia* de un micrófono es el ancho de banda dentro del cuál responde sin distorsión a las ondas de sonido.

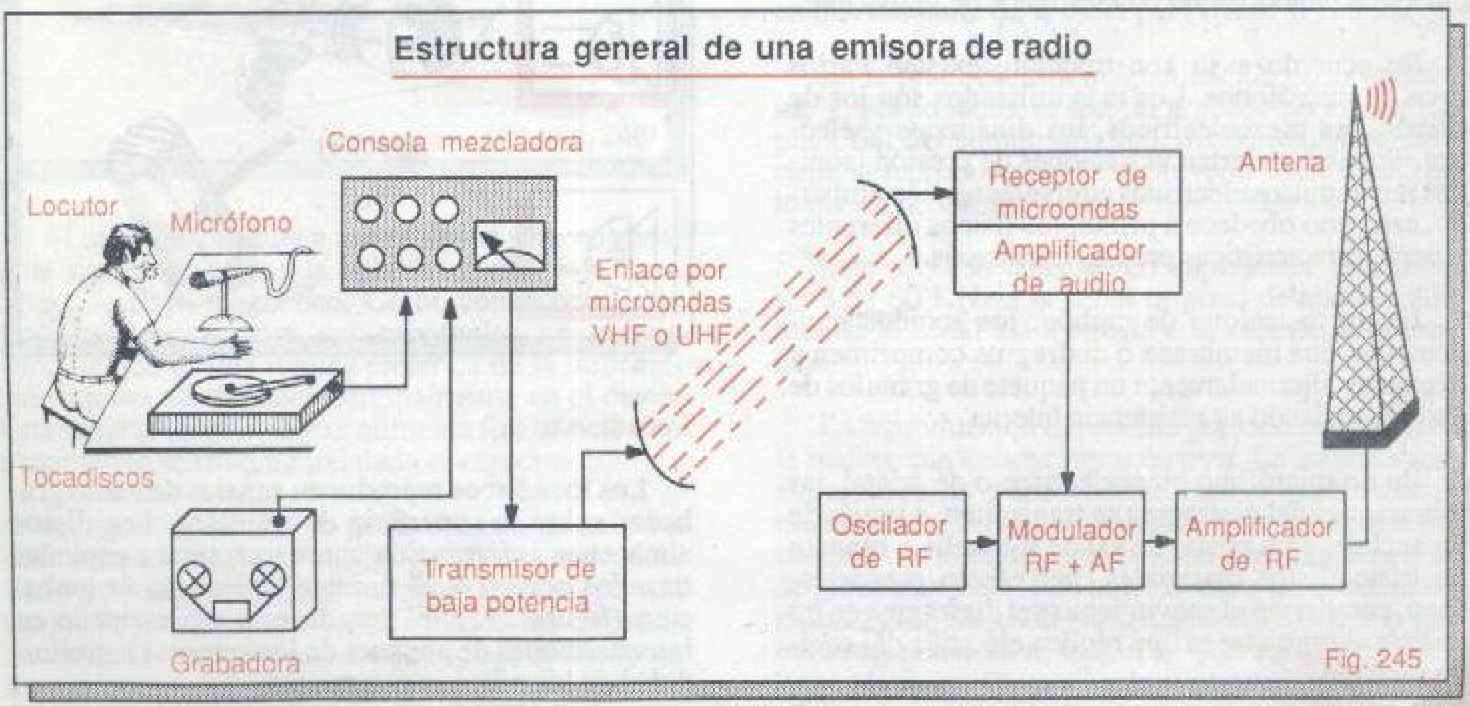
Estructura general de una emisora de radio

Fig. 245

La *direccionalidad* se refiere a la habilidad de un micrófono para captar sonidos provenientes de una u otra dirección. De acuerdo a su direccionalidad, los micrófonos se clasifican en omnidireccionales, bidireccionales y unidireccionales.

Los micrófonos omnidireccionales o panorámicos captan sonidos en todas las direcciones. Los bidireccionales captan sonidos en direcciones opuestas. Los unidireccionales o *cardioides* captan exclusivamente sonidos en una dirección.

La *impedancia* (Z) de un micrófono es el efecto de carga que éste le presenta al circuito que recibe la señal generada entre sus terminales. Existen micrófonos de alta impedancia y de baja impedancia. La impedancia de los primeros fluctúa típicamente entre 100 KΩ y 500 KΩ y la de los segundos puede ser de 50Ω, 150Ω ó 600 Ω.

La *sensibilidad* de un micrófono se especifica como el nivel de señal eléctrica que se obtiene del mismo bajo determinadas condiciones de referencia. La sensibilidad se expresa en voltios, vatios o dB y varía con la intensidad del sonido y la distancia de la fuente sonora.

Tipos de micrófonos

De acuerdo a su aplicación, los micrófonos pueden ser vocales, de uso general y profesionales. Los vocales reproducen frecuencias entre 150 y 4000 Hz y se utilizan en telefonía.

Los de uso general reproducen frecuencias entre 50 y 7500 Hz y se emplean en instalaciones de amplificación. Los profesionales reproducen frecuencias entre 20 y 20000 Hz y se utilizan en estaciones de radiodifusión o en estudios de grabación.

De acuerdo a su construcción, existen varios tipos de micrófonos. Los más utilizados son los de carbón, los piezoeléctricos, los dinámicos y electret. Todos convierten variaciones de presión (sonidos) en impulsos eléctricos equivalentes. Sin embargo, cada uno obedece a principios físicos diferentes y tiene características propias intrínsecas.

En un micrófono de carbón, los sonidos incidentes en una membrana o diafragma comprimen y expanden alternadamente un paquete de gránulos de carbón, variando su resistencia interna.

En un micrófono piezoeléctrico o de cristal, las vibraciones del diafragma se transmiten, a través de un anclaje, a un cristal de sal de Rochelle o titanato de bario. Estos materiales, por efecto piezoeléctrico, convierten el movimiento del diafragma en un voltaje alterno que es una réplica eléctrica del sonido recibido.

En un micrófono dinámico, las vibraciones del diafragma desplazan alternativamente una bobina liviana a través de un campo magnético estacionario, generando un voltaje variable entre sus terminales.

En un micrófono electret, las variaciones de presión del aire, producidas por el sonido incidente, cambian alternativamente la distancia entre dos placas cargadas eléctricamente de fábrica, alterando la capacitancia entre ellas y produciendo en los terminales del elemento un voltaje variable.

En la figura 246, se pueden observar los diferentes tipos de micrófonos y sus símbolos utilizados en los diagramas.

Los micrófonos

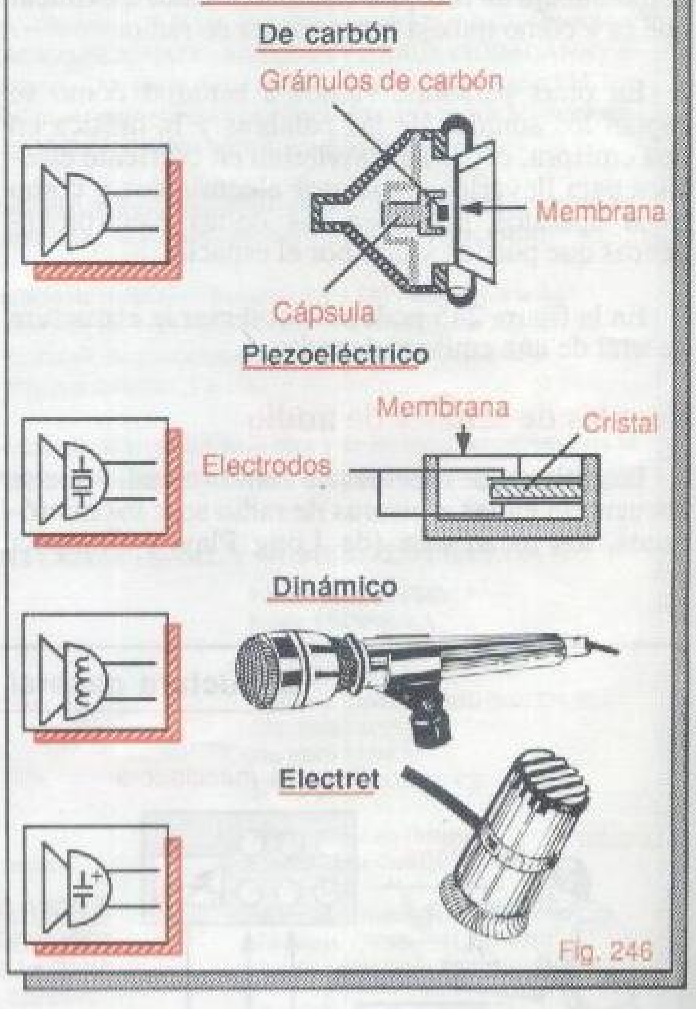

Fig. 246

Tocadiscos

Los tocadiscos reproducen señales de audio grabadas sobre la superficie de un disco. Los discos almacenan información sonora en surcos espirales trazados por un buril durante el proceso de grabación (figura 247). El sonido está representado en las variaciones de anchura de los surcos. La profundidad de los mismos es constante.

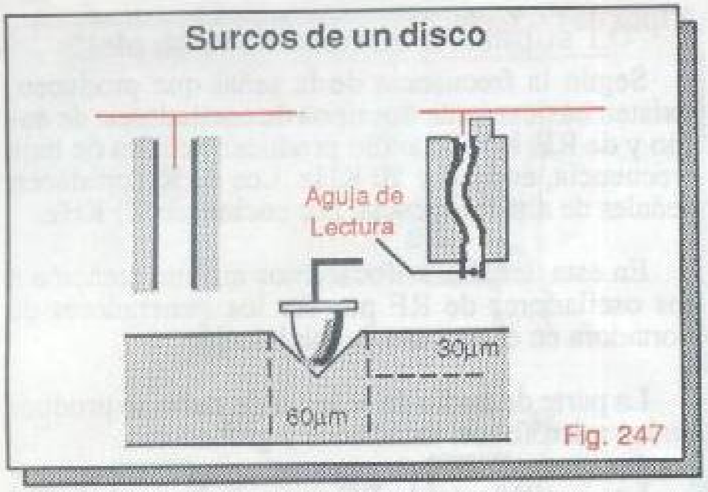

Fig. 247

La lectura de la señal grabada en un disco la realiza un dispositivo llamado pick up o cartucho que lleva una aguja en su extremo y un transductor magnético o piezoeléctrico en su interior.

La aguja recorre los surcos y transmite las vibraciones detectadas al transductor, el cual las convierte en señales eléctricas, equivalentes al sonido originalmente grabado. En la figura 248 se muestra el aspecto de un cartucho magnético.

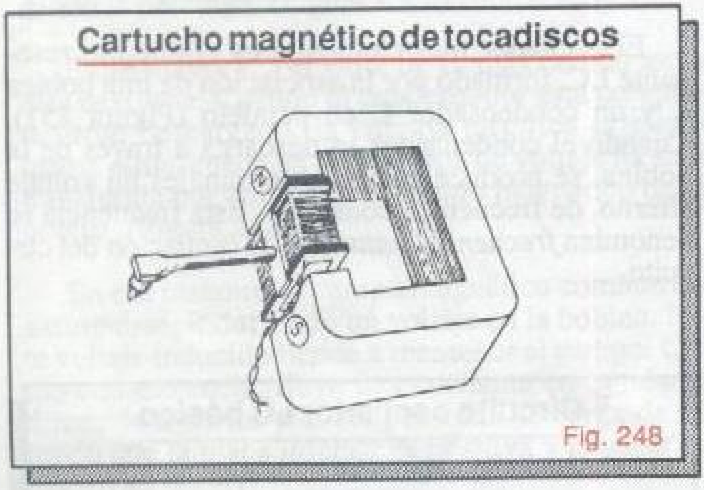

Fig. 248

Al moverse la aguja a través de los surcos, transmite sus vibraciones a la bobina móvil y ésta corta líneas de flujo magnético. Como consecuencia de esto, se genera entre sus terminales un voltaje variable que es una replica eléctrica de la información sonora almacenada originalmente en el disco. Esta señal se amplifica y se alimenta finalmente a un parlante y/o se irradia modulada al espacio.

La aguja de lectura es generalmente de zafiro o de diamante y su punta puede ser esférica o elíptica. Estos tipos de perfil y sus variantes le permiten a la aguja seguir con la máxima facilidad las tortuosidades de los surcos y evitan que toque el fondo de los mismos.

Grabadoras de sonido

El sonido se graba en cintas plásticas recubiertas con finas partículas de óxido de hierro (FeO) o de cromo (Cr_2O). Cada partícula contiene miles de pequeños imanes, cada uno de los cuales posee un polo norte y un polo sur. Normalmente, estos polos están al azar pero, bajo la influencia de un campo magnético externo, se puede lograr que apunten en una dirección determinada (alineamiento).

En la figura 249 se ilustra el principio básico de grabación del sonido. Cuando la señal de audio proveniente del micrófono se aplica a los devanados de la *cabeza de grabación*, se genera un campo magnético variable en el entrehierro de la misma.

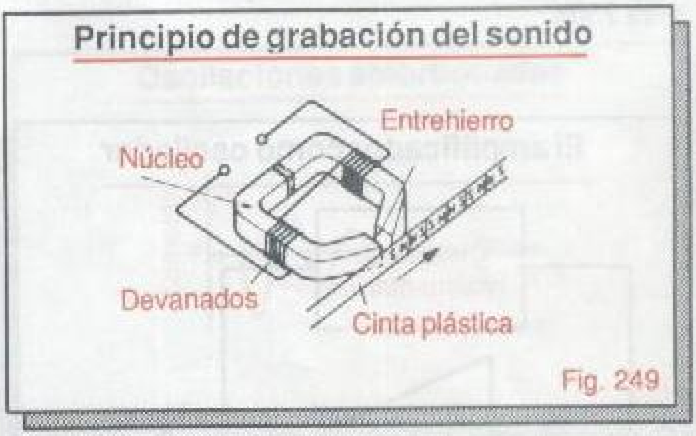

Fig. 249

A medida que pasa la cinta, sus partículas se magnetizan, creándose sobre ella un patrón magnético equivalente al sonido original. La *intensidad* de la magnetización en cualquier parte de la cinta depende del valor de la señal de audio en el momento en que esa parte de la cinta pasa frente al entrehierro.

En la práctica, esta magnetización no es uniforme y, por lo tanto, el sonido grabado no es una réplica fiel del sonido original, distorsionándose durante la reproducción. Para prevenir esto, se utiliza lo que se denomina *polarización* o *bias ac*.

Este proceso consiste en superponer una onda seno de 60 KHz a la señal original de audio y alimentar la señal así formada a la cabeza de grabación.

La reproducción del sonido grabado en una cinta la realiza una *cabeza reproductora*. La información almacenada se puede borrar total o parcialmente mediante una *cabeza de borrado*. Tanto la cabeza de borrado como la de reproducción son similares a la de grabación pero la de borrado tiene un entrehierro más amplio. La señal eléctrica producida por la cabeza reproductora se amplifica y se envía a un parlante y/o se irradia modulada al espacio.

Fuentes de señal de RF - Los Osciladores -

Teoría Básica. Qué es un oscilador

Como ya lo hemos mencionado, los circuitos osciladores se encuentran presentes en todos los aparatos de radio y televisión, tanto receptores como transmisores.

Un oscilador, es un circuito que genera por sí mismo una señal alterna de voltaje o de corriente, generalmente sinusoidal y de una frecuencia y amplitud específicas.

Se trata, en realidad, de un tipo especial de amplificador, caracterizado por tener una ganancia muy alta y por estar realimentado positivamente. (Figura 250).

El amplificador como oscilador

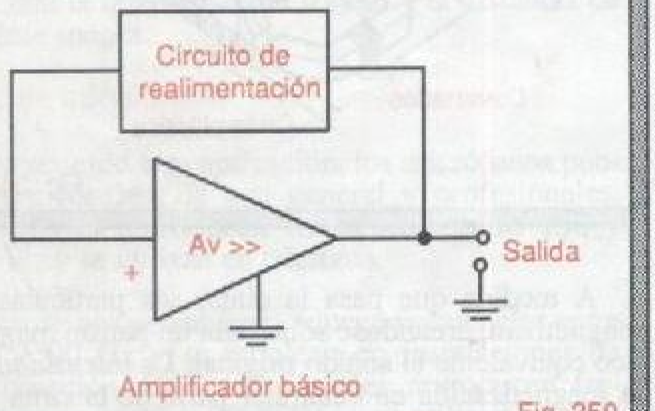

Fig. 250

Esto es, en un oscilador, el transistor o dispositivo activo usado amplifica la señal aplicada a su entrada como lo hace cualquier amplificador.

La diferencia está en que, en un oscilador, una parte de la señal de salida se realimenta a la entrada del amplificador, sumándose en fase con la señal aplicada.

Elementos que determinan la frecuencia de oscilación

El elemento activo o principal en un oscilador puede ser una válvula o tubo de vacío, un transistor, un FET o un circuito integrado.

La frecuencia de oscilación la determinan componentes pasivos como resistencias, condensadores, bobinas, cristales, diodos, etc. situados en el camino de realimentación.

Tipos de osciladores

Según la frecuencia de la señal que producen, existen básicamente dos tipos de osciladores: de audio y de RF. Los de audio producen señales de baja frecuencia, entre 0 y 20 KHz. Los de RF producen señales de alta frecuencia, por encima de 20 KHz.

En esta lección enfocaremos nuestra atención a los osciladores de RF por ser los generadores de portadora en el proceso de modulación.

La parte de audio de la señal de radio se produce en los micrófonos, tocadiscos y grabadoras.

Los osciladores de RF son, en su mayoría, del tipo LC, esto es, utilizan bobinas y condensadores como elementos determinantes de la frecuencia de oscilación. En muchos casos, estos elementos reactivos se simulan mediante cristales de cuarzo, logrando una mayor estabilidad en la frecuencia de la señal generada.

De acuerdo a su configuración, los osciladores de RF más conocidos son los tipos Colpitts, Clapp, Hartley y Pierce. De ellos nos ocuparemos en esta lección.

Circuito resonante LC como oscilador

El oscilador más elemental es el circuito resonante LC, formado por la asociación de una bobina L y un condensador C en paralelo (Figura 251). Cuando el condensador se descarga a través de la bobina, se produce entre sus terminales un voltaje alterno, de frecuencia constante. Esta frecuencia se denomina *frecuencia natural* o de oscilación del circuito.

Circuito oscilador LC básico

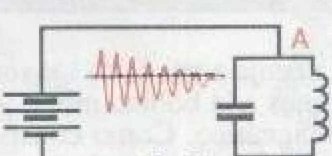

Fig. 251

El circuito LC mostrado, llamado también circuito tanque, con algunas variantes, es la base de los osciladores Hartley, Colpitts y otras configuraciones prácticas que se estudiarán más adelante. Su mayor ventaja radica en que se puede controlar la frecuencia de oscilación, variando el valor de sus componentes.

Ciclo de oscilación de un tanque LC

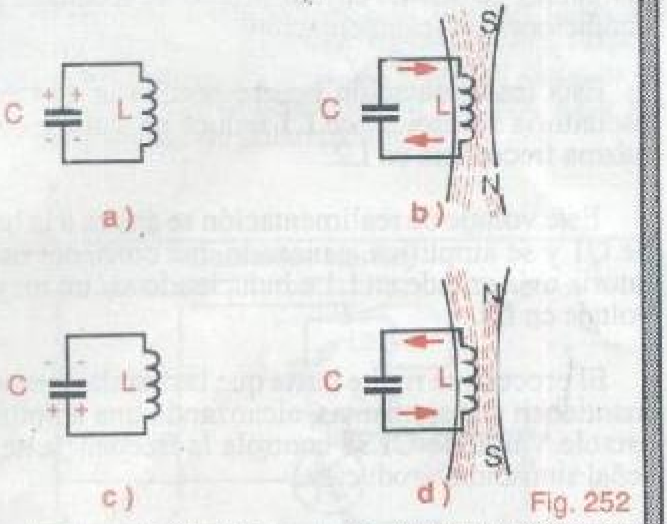

Fig. 252

En la figura 252 se ilustra gráficamente todo el proceso de oscilación de un circuito LC. A medida que el condensador comienza a descargarse, fluye una corriente a través del circuito y se crea un campo magnético variable, el cual induce un voltaje en la bobina. Este voltaje se opone a la corriente, retardando la descarga. (Figura 252a).

Entre mayor sea el valor de L, más tiempo toma esta descarga. Cuando el condensador está completamente descargado, la energía eléctrica originalmente almacenada en él en forma de voltaje ha sido transferida por completo al campo magnético de la bobina. (Figura 252b).

En ese instante, el campo magnético comienza a extinguirse, induciendo un voltaje en la bobina. Este voltaje inducido tiende a mantener el campo. Como consecuencia, fluye una corriente en la misma dirección anterior y el condensador se carga de tal modo que la placa inferior es positiva y la superior negativa.

Una vez que el campo magnético se ha extinguido por completo, la energía se almacena de nuevo en el condensador, en forma de voltaje (Figura 252c).

El condensador comienza entonces a descargarse a través de la bobina, pero la corriente ahora fluye en la dirección opuesta, creando un campo magnético de polaridad contraria al original. (Figura 252d).

Cuando este campo se ha extinguido por completo, el condensador se ha cargado de nuevo con su placa superior positiva, tal como lo estaba al comienzo del proceso. El ciclo se repite.

El resultado neto de esta operación ha sido la producción, por parte del circuito, de una oscilación eléctrica, esto es, de un ciclo de CA que se repite periódicamente a la frecuencia natural de resonancia del tanque LC.

La secuencia descrita anteriormente se repite varias veces, pero la energía se pierde gradualmente en el cicuito, en forma de calor y de radiación, debido a la naturaleza no ideal de los componentes involucrados.

Como resultado, las oscilaciones que se producen son amortiguadas, esto es, disminuyen progresivamente de amplitud y al cabo de un tiempo desaparecen. Su forma de onda se muestra en la figura 253.

Oscilaciones amortiguadas

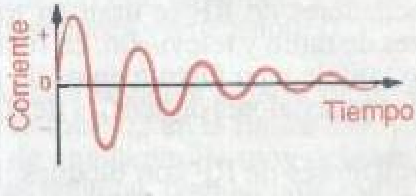

Fig. 253

Este sistema eléctrico, se puede comparar con el sistema mecánico de un péndulo. Este se mueve de un lado a otro, creando una oscilación o movimiento, pero el movimiento va perdiendo energía hasta que el péndulo se para totalmente. Para que un péndulo se mueva continuamente hay que suministrarle energía durante períodos fijos de tiempo.

Se puede demostrar matemáticamente que la frecuencia de oscilación, en Hz, de este circuito está dada por:

$$f = \frac{1}{2\pi\sqrt{(LC)}}$$

L está en Henrios y C en Faradios. Para generar oscilaciones de RF, L y C deben ser pequeñas. Por ejemplo, si L=100µH y C=100pF, entonces f = 1.6 MHz.

Para obtener oscilaciones no amortiguadas, de amplitud constante, como se muestra en la figura 254, debe alimentarse continuamente energía en el circuito LC, en la cantidad correcta y en el instante preciso.

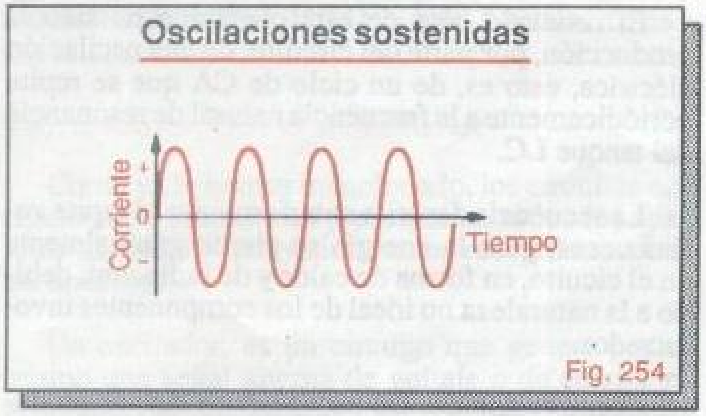
Oscilaciones sostenidas

Fig. 254

De este modo, se recuperan las pérdidas que son inevitables en este tipo de sistemas. Esta reinyección de energía se logra en la práctica mediante un dispositivo activo (por ejemplo, un transistor) asociado a una red pasiva de realimentación.

Osciladores de Radio-Frecuencia

Los osciladores de RF se usan en transmisores y receptores de radio y televisión, en fuentes de poder de alto voltaje y en instrumentos de prueba, tales como los generadores de señales de RF.

Los osciladores de RF son bloques básicos de todo equipo de radio. La primera etapa de casi todo transmisor es un oscilador de RF. Esta etapa determina la frecuencia particular en la cual va a operar el transmisor. De manera similar, el oscilador de RF utilizado en un receptor superheterodino determina la frecuencia a la cual está sintonizado.

Oscilador LC práctico de RF

En la figura 255 se muestra el circuito de un oscilador práctico de RF con tanque resonante LC. Se trata básicamente de un amplificador de voltaje, realimentado positivamente, construido alrededor del transistor Q1. R1, R2 y R3 son resistencias de polarización. C2 y C3 son condensadores de desacople de RF. C4 es un condensador de paso.

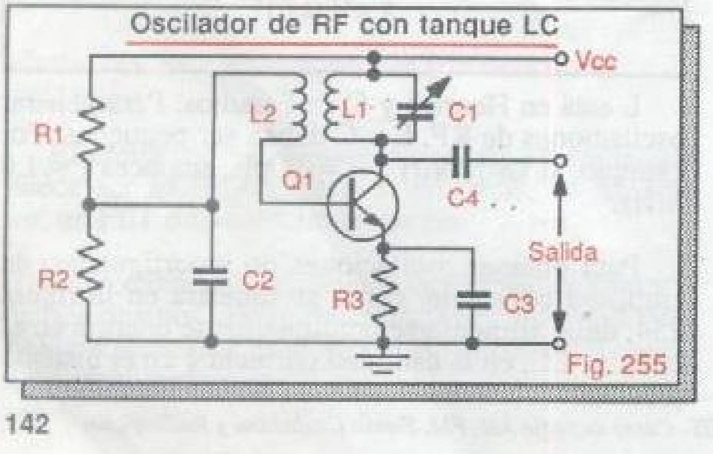
Oscilador de RF con tanque LC

Fig. 255

Al conectar la fuente de alimentación, se produce un pulso de corriente que carga el condensador C1 y origina automáticamente las oscilaciones, las cuales se sostienen tan pronto se alcanzan las condiciones de realimentación.

Esta realimentación ocurre porque la corriente oscilatoria sinusoidal en L1 induce un voltaje de la misma frecuencia en L2.

Este voltaje de realimentación se aplica a la base de Q1 y se amplifica, causando una corriente oscilatoria más grande en L1 e induciendo así un mayor voltaje en L2.

El proceso se repite hasta que las oscilaciones se mantienen por sí mismas, alcanzando una amplitud estable. Variando C1 se controla la frecuencia de la señal sinusoidal producida.

Osciladores Hartley y Colpitts

En la figura 256 se muestra la configuración general de estos dos populares osciladores de RF. El circuito de realimentación de ambos generadores es similar y lo constituyen las tres impedancias designadas como Z1, Z2 y Z3.

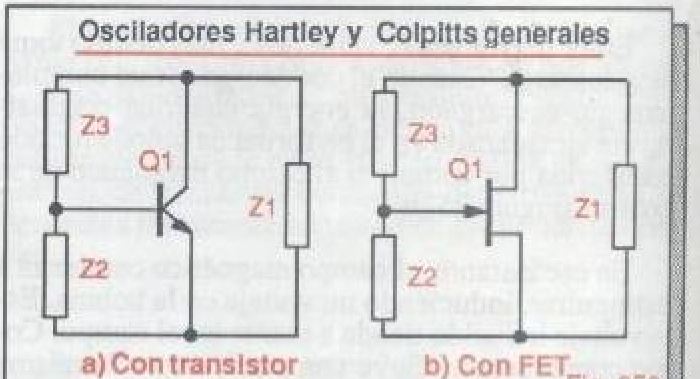
Osciladores Hartley y Colpitts generales

a) Con transistor
b) Con FET
Fig. 256

Se puede demostrar matemáticamente que para obtener realimentación positiva y oscilaciones sinusoidales, Z1 y Z2 deben ser ambas inductivas (o ambas capacitivas) y Z3 debe ser capacitiva (o inductiva).

El oscilador de RF utilizado en el radio AM CEKIT es del tipo Hartley. En este caso, la bobina osciladora L2 ("roja") provee las impedancias Z1 y Z2. La impedancia Z3 está representada por la asociación en paralelo de los condensadores VC2 (variable) y TC2 (trimmer).

Este último provee ajuste fino de frecuencia. La frecuencia de oscilación se puede variar entre 980 KHz y 2100 KHz, aproximadamente.

Oscilador Hartley

En la figura 257 se muestra el circuito práctico de un oscilador Hartley de RF. En este caso, las impedancias Z1 y Z2 están representadas, respectivamente, por las bobinas L0 y L1 del circuito de sintonía y Z3 por el condensador C1. R1, R2 y R3 son resistencias de polarización.

Oscilador Hartley

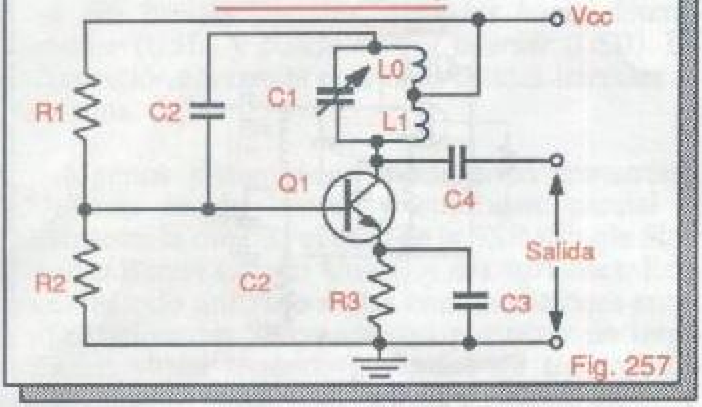

Fig. 257

C2 y C4 son condensadores de paso y C3 un condensador de desacople. Se supone que C2, C3 y C4 presentan una reactancia despreciable a la frecuencia de oscilación.

Oscilador Colpitts.

En la figura 258 se muestra el circuito práctico de un oscilador Colpitts. En este caso, las impedancias Z1 y Z2 están representadas, respectivamente, por los condensadores C0 y C1 del circuito de sintonía y Z3 por la bobina L1. R1, R2 y R3 son resistencias de polarización.

Oscilador Colpitts

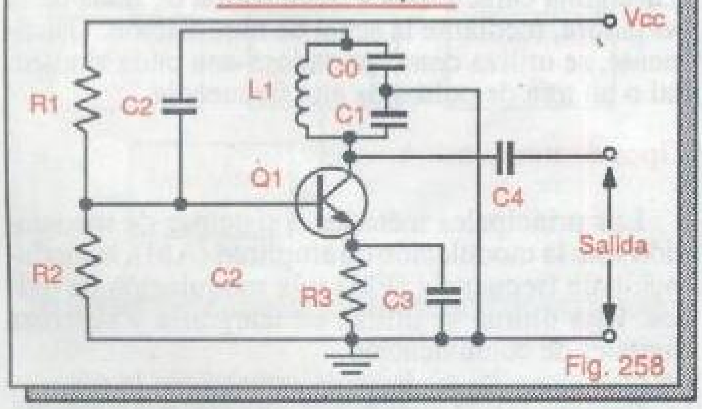

Fig. 258

C2 y C4 son condensadores de paso y C3 un condensador de desacople. Tanto C2 y C4 como C3 deben presentar una reactancia despreciable a la frecuencia de oscilación.

Osciladores a cristal

Este tipo de osciladores se caracterizan por utilizar en el circuito de sintonía un componente, llamado *cristal*, que exhibe propiedades *piezoeléctricas* y oscila por sí mismo, a una frecuencia determinada.

Los cristales más populares son los de cuarzo y sustituyen a los circuitos resonantes LC en la mayoría de aplicaciones de RF. La frecuencia a la cual un cristal de este tipo oscila depende de sus dimensiones físicas. Los osciladores a cristal se utilizan cuando se requiere generar una onda seno de frecuencia fija y altamente estable.

Teoría básica de cristales de cuarzo. El efecto piezoeléctrico.

El efecto piezoeléctrico consiste en la creación de oscilaciones eléctricas en las caras opuestas de un cristal cuando éste se somete a esfuerzos mecánicos (compresiones y expansiones).

Los cristales se tallan a partir de un cristal matriz, como se muestra en la figura 259, y su aspecto final es el de una laminilla muy delgada, de forma cuadrada, rectangular, circular o de anillo.

Obtención de cristales de cuarzo

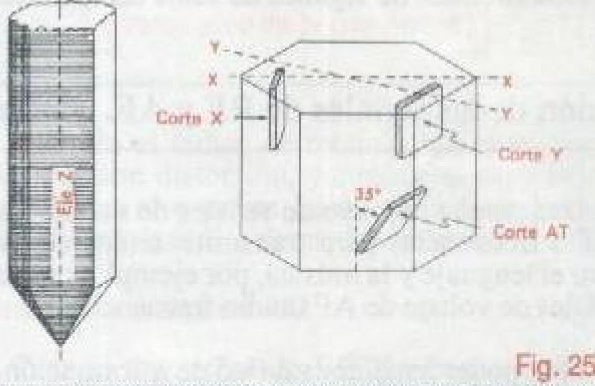

Fig. 259

Al ejercer una presión mecánica entre cualquier par de caras de un eje Y se produce una tensión eléctrica entre las aristas del eje X correspondiente. La polaridad de la tensión producida depende de si la presión aplicada al cristal se ejerce hacia adentro (compresión) o hacia afuera (expansión).

Así mismo, si se aplica una tensión eléctrica entre cualquier par de aristas X, las caras Y correspondientes vibrarán, expandiéndose y comprimiéndose alternadamente a una determinada frecuencia.

Los cristales para osciladores deben tallarse de acuerdo a dimensiones muy precisas pues éstas son las que determinan su frecuencia de operación. En general, entre más delgada sea la lámina, más alta será su frecuencia de oscilación y viceversa.

En la figura 260 se muestran la estructura interna, el aspecto externo y el símbolo de un cristal piezoeléctrico típico de cuarzo.

Cristal piezoeléctrico de cuarzo

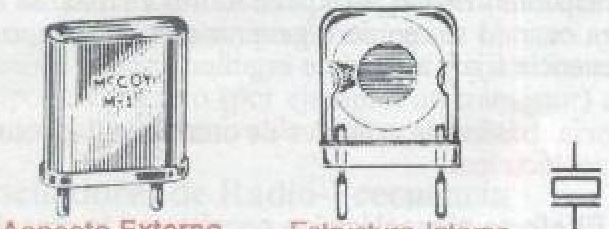

Aspecto Externo

Estructura Interna

Fig. 260

La cantidad de realimentación utilizada en un oscilador a cristal debe mantenerse a un nivel lo suficientemente bajo como para garantizar el arranque rápido de las oscilaciones y una operación confiable bajo condiciones de carga.

Son posibles numerosas configuraciones de osciladores controlados a cristal: Colpitts, Pierce, armónico, de sobretono, etc. En la figura 261 tenemos los circuitos de algunos de estos osciladores.

Osciladores controlados a cristal

Oscilador Colpitts

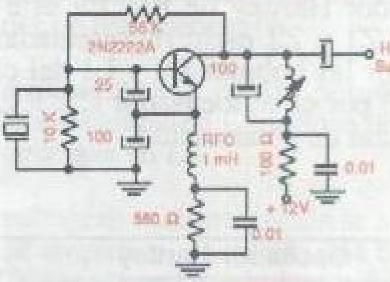

Oscilador Pierce

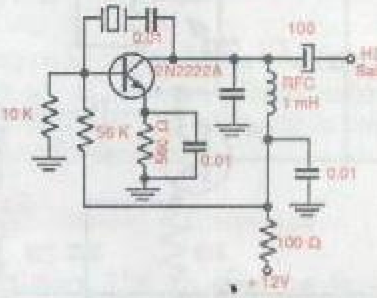

Oscilador Armónico de sobretono

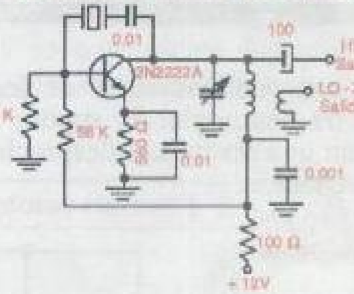

Fig. 261

Unión de las señales de RF y AF. Modulación.

Una antena requiere de señales de voltaje de RF (radio frecuencia) para transmitir ondas de radio, pero el lenguaje y la música, por ejemplo, producen señales de voltaje de AF (audio frecuencia).

Para poder irradiar este tipo de información, las señales de RF y de AF deben combinarse de tal modo que la primera transporte a la segunda. Este proceso se denomina *modulación* y lo realiza en el transmisor un circuito llamado modulador.

Las dos señales involucradas en el proceso reciben los nombres de onda *portadora* (RF) y onda *moduladora* (AF). Esta última se denomina también banda base y corresponde a la información o mensaje transmitido. El proceso contrario a la modulación se denomina *demodulación* y lo realiza en el receptor un circuito llamado demodulador o detector.

En general, la modulación es el proceso de variar alguna característica de la forma de onda de la portadora, mediante la señal de modulación. Usualmente, se utiliza como portadora una onda sinusoidal o un tren de pulsos de alta frecuencia.

Tipos de modulación

Los principales métodos o sistemas de modulación son: la modulación de amplitud (AM), la modulación de frecuencia (FM) y la modulación de pulsos. Esta última se utiliza en telegrafía y sistemas digitales de comunicación.

En esta lección, enfocaremos nuestra atención en los sistemas AM y FM. En AM, la información propiamente dicha está "codificada" en las variaciones de amplitud de la portadora y en FM lo está en las variaciones de frecuencia.

Concepto de Banda Lateral

Cualquiera que sea el esquema de modulación utilizado, siempre que se modula una portadora se producen *bandas laterales*. Las bandas laterales son bandas de frecuencia, a lado y lado de la portadora, que resultan de variar una característica (amplitud, frecuencia, etc.) de la portadora, mediante la señal de modulación. Las bandas laterales contienen la información inteligente transmitida.

El proceso de modulación propiamente dicho crea dos bandas laterales, llamadas *banda lateral superior* (USB) y *banda lateral inferior* (LSB). La información contenida en ambas bandas laterales es la misma.

Algunos sistemas de modulación transmiten sólo una banda lateral y suprimen parcial o totalmente la otra. Es el caso de la SSB (Single Side Band o Banda Lateral Unica) y sus variantes. Este es el método utilizado en las comunicaciones entre radioaficionados.Se caracteriza por tener un largo alcance y por requerir potencias de transmisión relativamente bajas.

Modulación de Amplitud (AM)

El método más común de modulación es la modulación de amplitud o AM. En la figura 262 se muestra el diagrama de bloques de un transmisor de AM típico. Veamos cómo opera en términos generales.

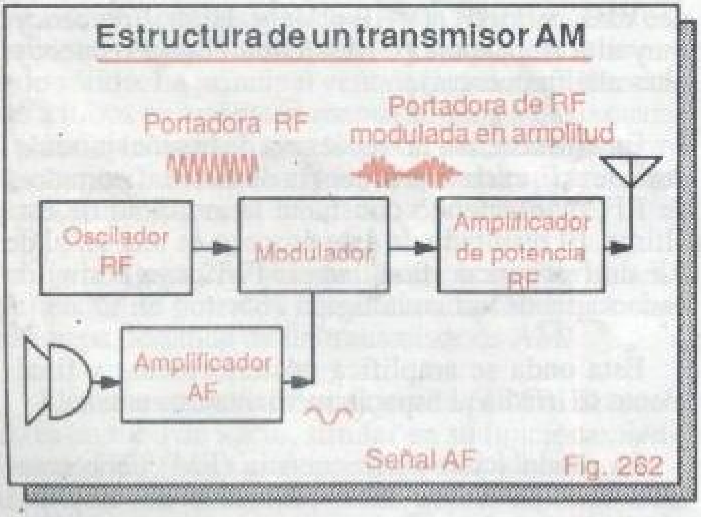
Estructura de un transmisor AM
Portadora RF
Portadora de RF modulada en amplitud
Oscilador RF
Modulador
Amplificador de potencia RF
Amplificador AF
Señal AF

Fig. 262

En el modulador, la señal de RF producida por el oscilador interno se combina con la señal de AF suministrada por el amplificador de audio y proveniente del micrófono. La frecuencia de la señal de AF (moduladora) varía la amplitud de la señal de RF (portadora).

El resultado es una señal de RF de amplitud modulada, que es amplificada y luego irradiada.

El proceso AM. Concepto de Indice de Modulación

En la figura 263 se muestra el efecto de combinar la señal de audio con la portadora de RF en el modulador de un transmisor de AM.

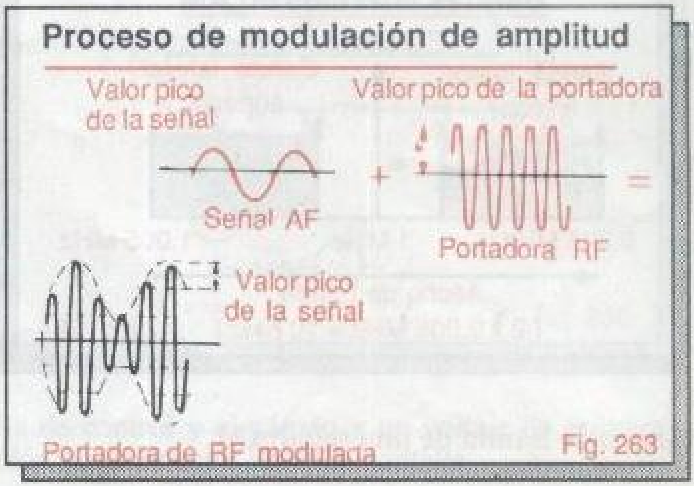
Proceso de modulación de amplitud
Valor pico de la señal
Valor pico de la portadora
+
=
Señal AF
Portadora RF
Valor pico de la señal
Portadora de RF modulada

Fig. 263

El factor más importante del proceso de modulación AM descrito es el *porcentaje o índice de modulación*. Este parámetro se designa como m y se define así:

$$m = \left[\frac{(\text{Valor pico de la señal})}{(\text{Valor pico de la portadora})} \right] \times 100$$

Cuando el índice de modulación es mayor del 100%, ocurre distorsión, y cuando es muy bajo, la recepción es muy pobre. La mayoría de emisoras de AM utilizan un índice de modulación del 85 al 95%.

Bandas Laterales en AM

Cuando una señal de RF, de frecuencia **fc**, se modula en amplitud con una señal de AF puramente sinusoidal, de frecuencia **fs**, la señal modulada de RF resultante está constituída por tres tipos de frecuencias: la frecuencia de la portadora (fc), la frecuencia lateral superior (fc+fs) y la frecuencia lateral inferior (fc-fs).

En la práctica, la información eléctrica de audio (música, voz, sonido, etc) que se envía sobre una portadora de RF, está formada por la combinación de varias señales puramente sinusoidales, cada una de las cuales produce su propio par de frecuencias laterales.

El resultado de este proceso de modulación es una gama o banda de frecuencias distribuidas simétricamente a lado y lado de la portadora de RF, como se muestra en la figura 264. A estas bandas de frecuencias se les denomina *bandas laterales* superior (USB) e inferior (LSB) y contienen toda la información de audio transmitida.

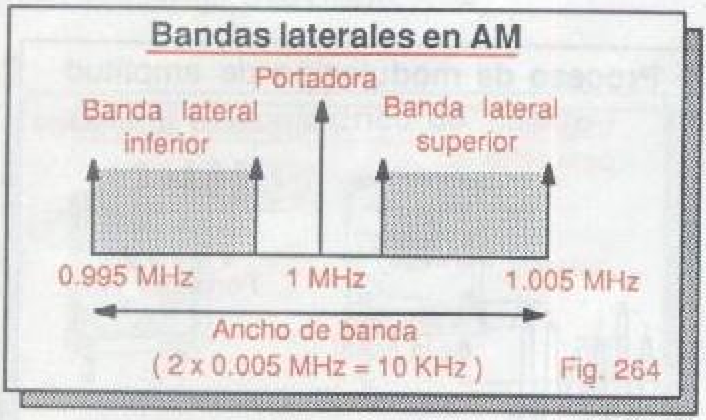

Fig. 264

Ancho de Banda de una señal AM

El ancho de banda **B** de la señal AM, generada bajo la circunstancia general descrita anteriormente, es la distancia entre las frecuencias laterales, superior e inferior, correspondientes a la señal de más alta frecuencia ($f_{smáx}$) presente en la información de audio. Esto es:

$$B_{AM} = (f_c + f_{smáx}) - (f_c - f_{smáx}) = 2f_{smáx}$$

Ejemplo: Calcule el ancho de banda mínimo requerido para transmitir, en AM, una información de audio que contiene frecuencias entre 30 Hz y 3 KHz, sobre una portadora de 1000 KHz.

Solución: La frecuencia de la portadora es fc=1000 KHz y la de la componente más alta en la información de audio es fsmáx=3 KHz. Por lo tanto:

$$B = 2 \times 3\text{KHz} = 6 \text{ KHz}$$

Al disponer de un ancho de banda de 6 KHz, la señal transmitida ocupará el espectro de frecuencias comprendido entre 997 KHz y 1003 KHz. Dentro de este rango no debe haber frecuencias laterales de otra estación porque se provocaría interferencia.

Transmisiones comerciales de AM

La banda de frecuencias asignada para las estaciones de radio locales que transmiten por el sistema AM estandar se extiende desde 530 KHz hasta 1605 KHz. A esta banda se le llama onda larga y en ella se encuentran emisoras dedicadas generalmente a la difusión de noticias, deportes, música, etc. Su alcance está limitado a una ciudad o región.

En el sistema AM también se han asignado frecuencias para transmisión internacional. Son las llamadas estaciones de onda corta y transmiten principalmente en las bandas de 60, 49, 31, 25 y 19 metros de longitud de onda. En estas bandas se encuentran generalmente emisoras estatales, dedicadas a la transmisión de noticias internacionales, música folclórica, música clásica, notas culturales, etc.

En AM, y en general en todo sistema organizado de comunicaciones, cada estación debe realizar sus transmisiones a una frecuencia de portadora específica, asignada por el organismo rector de las comunicaciones en cada país o estado. Así mismo, debe estar convenientemente espaciada de las estaciones vecinas, para minimizar la interferencia entre ellas.

La separación normal entre emisoras de AM que transmiten en la banda de 530 KHz a 1605 KHz es de 9 KHz. Este valor corresponde, realmente, al ancho de banda máximo de cada transmisión. En consecuencia, la máxima frecuencia de audio que puede reproducirse con toda fidelidad en AM es de 4.5 KHz. La frecuencia de la portadora no debe exceder de ±20 Hz con respecto a su valor nominal.

Modulación de Frecuencia (FM)

Otro método común de modulación es la modulación de frecuencia o FM, utilizada por estaciones comerciales de radio y televisión y en comunicaciones de VHF y UHF. (VHF= Very High Frecuency: muy alta frecuencia; UHF= Ultra High Frecuency: ultra alta frecuencia).

En este sistema, la frecuencia de la señal moduladora de AF varía la frecuencia de la señal portadora de RF, manteniendo constante la amplitud de esta última. El resultado de este proceso es una señal de RF de frecuencia modulada o FM cuya forma de onda se puede ver en la figura 265.

Esta onda se amplifica posteriormente y finalmente se irradia al espacio mediante una antena.

La modulación de frecuencia (FM) tiene notables ventajas sobre la modulación de amplitud (AM). Una de las más importantes es la inmunidad al ruido de interferencia generado por motores, alumbrado, sistemas de ignición de vehículos, tormentas eléctricas, etc.

Este ruido se induce en los receptores de AM porque estas señales tienden a ser moduladas en am-

Proceso de modulación de frecuencia

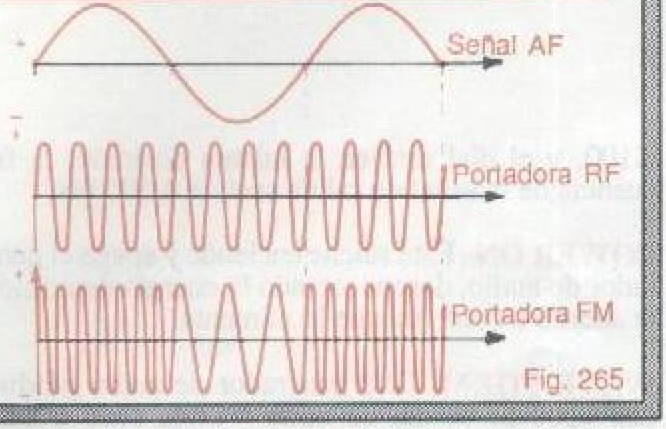

Fig. 265

plitud y no en frecuencia. En próximas lecciones estudiaremos el proceso completo de transmisión y recepción de señales de FM.

Las estaciones comerciales de FM transmiten en la banda de 88 a 108 MHz. El máximo ancho de banda concedido a cada estación es de 200 KHz. En este hecho radica la alta calidad de las transmisiones por este sistema.

Amplificación de la señal modulada de RF

Los amplificadores de RF se utilizan en las estaciones de radio para amplificar la señal de alta frecuencia modulada que se va a irradiar al espacio a través de la antena. Como sabemos, esta señal puede estar modulada en amplitud (AM) o en frecuencia (FM). Esta sección se refiere exclusivamente a transmisores de AM.

La mayoría de los circuitos modernos utilizados para amplificar la señal de RF son híbridos, esto es, emplean conjuntamente tubos y dispositivos de estado sólido. La principal ventaja de los amplificadores a tubos es que están menos sujetos a daño cuando manejan niveles excesivos de potencia, y soportan mejor las sobrecargas y los desacoplamientos que los transistores.

En la figura 266 se muestra el circuito de un amplificador de potencia clase C a tubos utilizado como etapa de salida de un transmisor de AM.

El elemento activo es un tetrodo 6146B. El tetrodo es un tubo de vacío, similar en su funcionamiento a un FET, que tiene un filamento y cuatro electrodos llamados cátodo, placa, rejilla de control y rejilla de pantalla.

Para operar correctamente, todos los amplificadores a tubos de RF requieren de un voltaje de CA para calentar el filamento, un voltaje positivo de alimentación de CC entre la placa y el cátodo, un voltaje de polarización negativo de CC entre la rejilla de control y el cátodo y un voltaje de polarización positivo entre la rejilla pantalla.

Etapa de salida de un transmisor AM

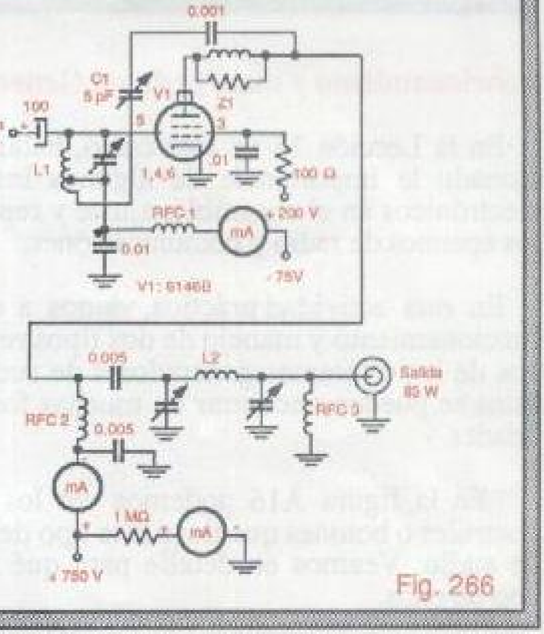

Fig. 266

Al calentarse el filamento, el cátodo emite electrones, debido a un fenómeno físico conocido como "emisión termoiónica". La rejilla de control actúa como una válvula que regula el paso de electrones entre el cátodo y la placa. Una pequeña variación en el voltaje de rejilla provoca una gran variación en la corriente de placa. La rejilla de pantalla actúa como un blindaje y su objetivo es acelerar el paso de electrones entre la rejilla de control y la placa.

En un transmisor AM como el mostrado en la figura 266, la operación en clase C del amplificador de salida es la que ofrece el máximo rendimiento (65 a 75%). Esta configuración puede llegar a suministrar entre 650 y 750 W de potencia de salida. Los medidores se localizan en puntos estratégicos del circuito para monitorear los voltajes y corrientes que influyen en la potencia de transmisión.

La señal de salida del amplificador de potencia se suministra al sistema de antena. Este último está constituido por la antena propiamente dicha, la línea de transmisión y los circuitos de acoplamiento. Estos circuitos son del tipo LC y se utilizan para transferir la potencia de RF del amplificador a la línea de transmisión y de ésta a la antena.

Además de garantizar la máxima transferencia de potencia, los circuitos de acoplamiento deben también minimizar la generación de armónicas, es decir, de frecuencias diferentes a la frecuencia de transmisión de la estación. A esta función se le llama filtraje.

Actividad práctica Nº 11

Funcionamiento y manejo de un Generador de Audio

En la Lección 15 de este curso, habíamos mencionado la importancia de algunos instrumentos electrónicos en el ensamble, ajuste y reparación de los aparatos de radio y comunicaciones.

En esta actividad práctica, vamos a estudiar el funcionamiento y manejo de dos tipos representativos de los diferentes generadores de audio, ya que estos se pueden encontrar en muchas formas y calidades.

En la figura A16 podemos ver los diferentes controles o botones que tiene este tipo de generador de audio. Veamos en detalle para qué sirve cada uno de ellos.

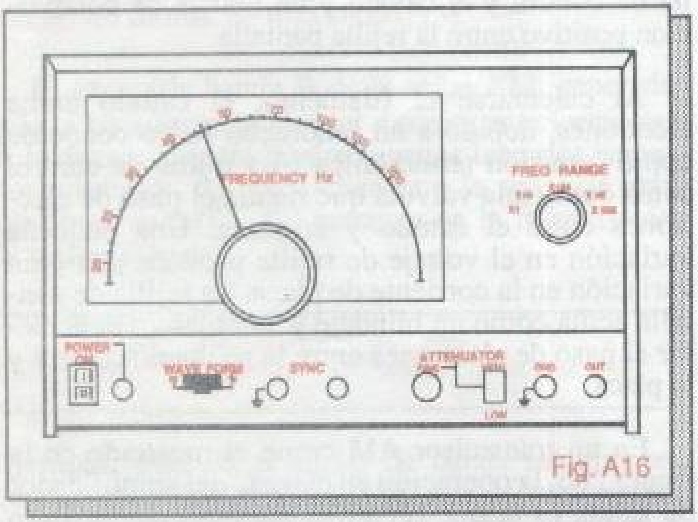

Fig. A16

FREQUENCY. Es un control de frecuencia cuya posición, marcada en la escala o dial, nos determina cuál será la frecuencia en Hertzios o ciclos por segundo, de la señal que se obtiene en los bornes de salida. Esta frecuencia depende también de la posición en que se encuentre el selector llamado FREQ. RANGE.

FREQ. RANGE . Este suiche selector rotatorio nos selecciona una determinada banda de frecuencias. Cuando está en la posición X1, por ejemplo, la frecuencia de salida será la que indica la escala. En el caso del dibujo, el generador estará entregando una señal de 45 Hz., aproximadamente.

En las otras posiciones, el suiche selector indica que en cada banda la frecuencia se multiplica por 10, por 100, por 1.000 (1K) y por 10.000 (10K). Así, por ejemplo, si este selector está en la posición X100, y el dial está en la misma posición, la frecuencia de la señal de salida será de 4.500 Hz.

POWER ON . Este suiche enciende y apaga el generador de audio, desconectando la energía de corriente alterna o continua que lo alimenta.

WAVE FORM. Este generador de audio produce dos tipos de forma de onda : Onda seno o sinusoidal y onda cuadrada. Este suiche, por lo tanto, selecciona cuál de éstas dos formas de onda tendrá la señal de salida del generador.

ATTENUATOR. Esta función la realizan dos elementos, un suiche y un potenciómetro de ajuste y se refiere a la " Atenuación". Esta función consiste en que se puede ajustar el nivel o amplitud de la señal de salida mediante estos dos controles.

En la posición " HIGH" o alto, el generador entregará una señal alta, entre 0 y 10 voltios pico a pico, aproximadamente. En la posición " LOW " o bajo, el generador entregará entre 0 y 1 voltio pico a pico.

Con el potenciómetro marcado como " FINE " o ajuste, podemos variar la amplitud de la señal de salida en forma lineal. Esto se requiere para ajustar o colocar el nivel de la salida de acuerdo con el circuito que vamos a probar o ensayar.

TERMINALES DE SALIDA. Estos terminales o bornes están marcados como " GND " y " OUT ". Por ellos sale la señal del generador de audio hacia el aparato o circuito de prueba. El terminal GND o tierra se debe conectar a la tierra, negativo o chasís del aparato que vamos a ensayar y el otro terminal, OUT, se conecta a la entrada del circuito.

Operación del generador

El procedimiento a seguir es el siguiente : Primero debemos conectar el generador al circuito que deseamos probar.

Antes de esto debemos estar seguros de que el suiche de " POWER " o energía esté apagado, que el suiche "ATTENUATOR" esté en la posición "LOW" y que el control del nivel de salida "FINE" esté al mínimo (todo a la izquierda).

Adicionalmente debemos seleccionar cuál forma de onda y que frecuencia de señal, vamos a utilizar. Generalmente, en circuitos de radio y comunicaciones, se utiliza la onda seno. Para el ajuste de la frecuencia seleccionamos primero la banda y luego el valor exacto.

Por ejemplo, para una señal de salida de 1.000 Hz o 1.000 ciclos por segundo, el selector de bandas o rangos debe estar en la posición X10 y el dial o escala debe estar en la posición 100.

Una vez que se ha seleccionado la frecuencia, debemos ajustar el nivel de salida adecuado según el circuito que vamos a probar. En el caso específico del generador de audio mostrado anteriormente, debemos determinar si el voltaje de salida que deseamos está entre 0 y 10 voltios o entre 0 y 1 voltio.

Según el rango, colocamos el suiche "ATTENUATOR" en la posición "HIGH o LOW". Luego debemos mover hacia la derecha la perilla "FINE" hasta que en la salida se obtenga la señal con el nivel esperado. Para medir este nivel de voltaje, en forma precisa, se requiere un multímetro digital o un osciloscopio.

Al realizar la medida, debemos tener en cuenta que al tomar una señal con un multímetro análogo o digital, se obtiene el voltaje RMS y para obtener el voltaje pico a pico o Vpp se debe multiplicar este valor por 2.82. Se hace este comentario en base a que en radio y comunicaciones se acostumbra trabajar con señales pico a pico.

En el osciloscopio se pueden obtener fácilmente los valores pico o pico a pico como veremos más adelante.

En la figura A17 podemos observar otro tipo común de generador de audio y que utilizaremos como referencia para el análisis y calibración del radio CEKIT que ensamblamos con este curso.

Generador de funciones

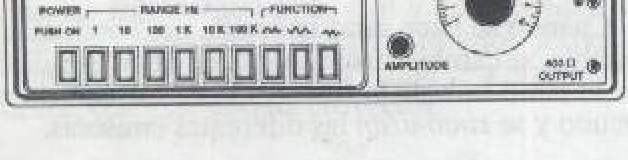

Fig. A17

POWER ON. Tiene la misma función que en el anterior para encender y apagar el instrumento.

RANGE Hz. Selecciona el rango de frecuencias multiplicando por 1,10,100,1.000,10.000 ó 100 mil el valor indicado por la perilla central "FREQUENCY".

FUNCTION. Selecciona la forma de onda de la señal de salida entre las opciones de onda cuadrada, seno y triangular, que se obtiene entre los bornes de tierra (⏚) y 600 Ω OUTPUT.

D.C. OFFSET. Permite colocar un nivel de CC entre + y - 5 voltios, sumado a la señal de salida en CA.

AMPLITUDE. Permite ajustar el nivel o voltaje de la señal de salida desde 0 hasta el máximo valor de 10 Voltios pico a pico que entrega el generador en el borne de salida de 600Ω.

FREQUENCY. Con esta perilla seleccionamos el valor de la frecuencia de salida en Hertzios con valores desde 0.1 hasta 10 y trabaja relacionada con el botón de rango.

Por ejemplo, si colocamos esta perilla en el punto 5, y se presiona el botón de rango 1K, tendremos en la salida una señal de 5.000 Hertz.

TTL. Es un borne que entrega una señal de onda cuadrada con nivel de salida fijo y frecuencia ajustable, con la perilla central y el botón de rango. Permite manejar hasta 20 cargas TTL.

600 Ω OUTPUT. Entre este borne y tierra se obtiene la señal de salida del generador con un nivel, forma de onda y frecuencia que ha sido seleccionado con los controles mencionados anteriormente.

Para su operación, se debe proceder de la misma forma que con el generador anterior.

Seleccionar la frecuencia, el rango, la forma de onda y el nivel del voltaje de salida.

Debemos tener siempre en cuenta que el nivel de salida de un generador de señal, se debe ajustar al valor apropiado máximo que acepta un circuito.

Si aplicamos una señal excesiva a un circuito con transistores o con circuitos integrados, estos podrían dañarse.

También debemos proteger el generador intercalando un condensador entre éste y el circuito, con el fin de bloquear la corriente continua de polarización que existe en los circuitos para que no pase hacia la parte interna del generador.

Receptores de radio

En la lección anterior estudiamos cómo se produce la señal de radio en una emisora. Sabemos que esta señal sale por la antena en forma de ondas electromagnéticas y que por su frecuencia se llaman específicamente ondas de radio.

También hemos estudiado cómo éstas viajan en diferentes formas según su frecuencia hasta llegar a lugares cercanos o muy distantes de su lugar de origen.

Vamos ahora a estudiar cómo se reciben estas ondas y cómo se obtiene de ellas nuevamente la información o mensaje que contienen. El aparato electrónico que realiza este trabajo se llama *receptor de radio* o, simplemente, *radio*.

El radio es quizás el aparato electrónico que más se encuentra sobre la Tierra y el más utilizado por el hombre. Con él escuchamos noticias, música, recibimos programas educativos y nos informamos de muchas otras actividades que han hecho que el hombre se comunique con sus semejantes a través de todo el mundo.

Para estudiar el radio lo haremos analizando cada una de sus etapas o partes en que está dividido. Aunque existen muchos tipos de radios, en su gran mayoría se fundamentan en los mismos principios básicos que estudiaremos y utilizan los componentes y leyes que hemos analizado en las lecciones anteriores.

Todos los receptores interceptan una señal de radio mediante la antena, la amplifican, la demodulan y luego la reproducen con la misma modulación (audio) con que fue enviada desde una estación remota. Los receptores difieren en la forma como procesan internamente la señal original y en los circuitos empleados para tales efectos.

Receptores sencillos de radio

Vamos ahora a estudiar el funcionamiento de receptores de radio muy sencillos con el fin de entender su operación y poco a poco iremos introduciendo circuitos más complejos, hasta llegar a un receptor completo tipo *superheterodino*, que es el sistema empleado por la mayoría de los receptores modernos.

En la figura 267 tenemos el diagrama de bloques o etapas de un receptor de radio muy sencillo.

Receptor de radio sencillo

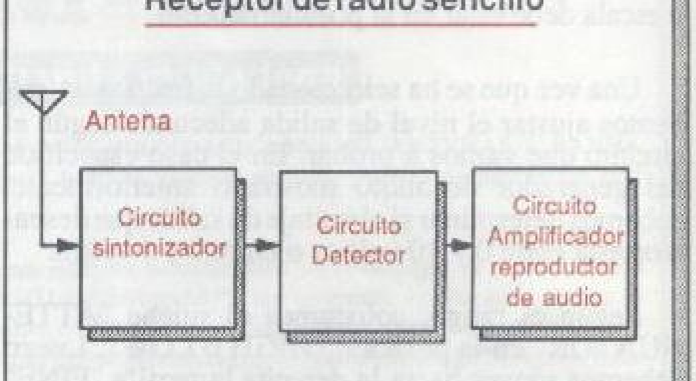

Fig. 267

Diríamos que es el radio mínimo que se puede construir.

Este circuito en la práctica se puede ensamblar como se muestra en las figuras 268a y 268b.

El receptor mostrado recibe el nombre de *receptor a cristal o galena* ya que anteriormente el diodo era construido con cristal de galena o carborundum, un material que permitía el paso de la corriente en una sola dirección.

Vamos ahora a analizar cada una de las etapas de este receptor a cristal. En el alambre de la antena están presentes una gran cantidad de emisoras de radio, siendo las señales más fuertes aquellas de las emisoras más cercanas o que transmiten con mayor potencia en kilovatios.

Al estar conectada la antena al primario de una bobina o transformador de antena, la energía de radiofrecuencia, por ser corriente alterna, induce un voltaje o señal en el secundario de esa bobina.

Este secundario está conectado en paralelo con un condensador variable, formando un circuito LC resonante, cuyo funcionamiento se estudió en la lección 12 de este curso.

Cuando se hace girar la perilla del condensador variable, se cambia su capacidad en microfaradios y por lo tanto cambia la frecuencia de resonancia del circuito y se *sintonizan* las diferentes emisoras.

A este circuito o etapa se le llama *circuito sintonizador* y su función básica es seleccionar cuál emisora se va a escuchar en un momento dado.

Receptor básico a cristal

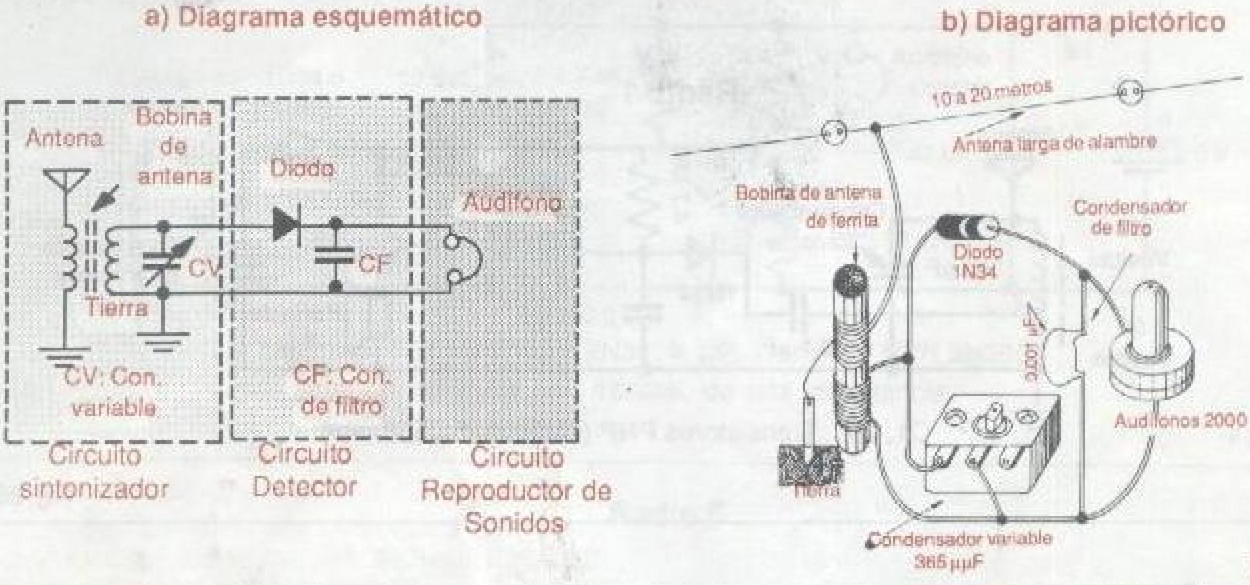
a) Diagrama esquemático

b) Diagrama pictórico

Antena
Bobina de antena
Tierra
CV
CV: Con. variable
Diodo
CF
CF: Con. de filtro
Audífono
Circuito sintonizador
Circuito Detector
Circuito Reproductor de Sonidos
10 a 20 metros
Antena larga de alambre
Bobina de antena de ferrita
Diodo 1N34
Condensador de filtro
0.001 μF
Audífonos 2000 Ω
Tierra
Condensador variable 365 μμF

Fig. 268

Esta señal de la emisora sintonizada, como vimos en la lección anterior, está formada por una onda de alta frecuencia cuya amplitud está *modulada* por la señal sonora o de audio.

Dicha señal se *rectifica* por medio del diodo, es decir, se elimina su parte negativa, quedando media onda de la original. Luego, con el condensador de filtro, se elimina la parte de alta frecuencia o *portadora*, quedando la señal de audio. A esta etapa se le llama *etapa detectora*.

Después, esta señal se lleva a un parlante, el cual convierte la corriente eléctrica de audio o baja frecuencia en sonidos que son escuchados por el oído humano.

Este circuito, por ser tan simple, tiene poca selectividad y baja sensibilidad pero es muy útil para comprender la teoría básica de un receptor de radio.

Adelante, estaremos entregando los detalles completos para el ensamble de un receptor a cristal más completo con el fin de que el estudiante o el aficionado puedan armar su primer radio, actividad emocionante y llena de satisfacción.

En las figuras 269 a, b, c, d y e, se pueden ver varios diagramas con diferentes circuitos derivados del anterior para que el lector empiece a practicar, si lo desea, con este tipo de proyectos.

Los componentes se pueden reemplazar por algunos similares sin alterar el funcionamiento de los aparatos. La bobina de antena se puede construir siguiendo las indicaciones suministradas en el dibujo.

Ahora describamos algunos aspectos prácticos relacionados con las características de la antena y la construcción de la bobina de ferrita. Estos conocimientos son importantes para obtener óptimos resultados en la captación de señales de radio mediante receptores a cristal.

Características del sistema de antena y de tierra

Para obtener buenos resultados con estos radios experimentales y maximizar el nivel de señal recibido, debe utilizarse una buena antena y un buen sistema de tierra.

La antena debe ser un tramo de alambre de cobre de unos 8 a 30 m de longitud, extendido horizontalmente entre dos postes tan altos como sea posible y en un área despejada. Entre los extremos del alambre de antena y los postes deben interponerse aisladores de plástico, vidrio o porcelana.

De este modo, desde el punto de vista eléctrico, la antena queda como si fuera un alambre suspendido en el aire. El alambre de bajada de la antena también debe estar aislado. En la figura 270 se muestra el aspecto final de una instalación típica de antena.

El sistema de tierra se puede obtener conectando un alambre de baja resistencia entre la tierra del circuito y una tubería o conducto de agua que penetre profundamente en el suelo. En algunos casos, se obtienen también buenos resultados conectando la tierra del circuito al retén metálico del disco de marcación de un telefóno rotatorio.

Diagramas de radios sencillos que se pueden construir

Radio 1

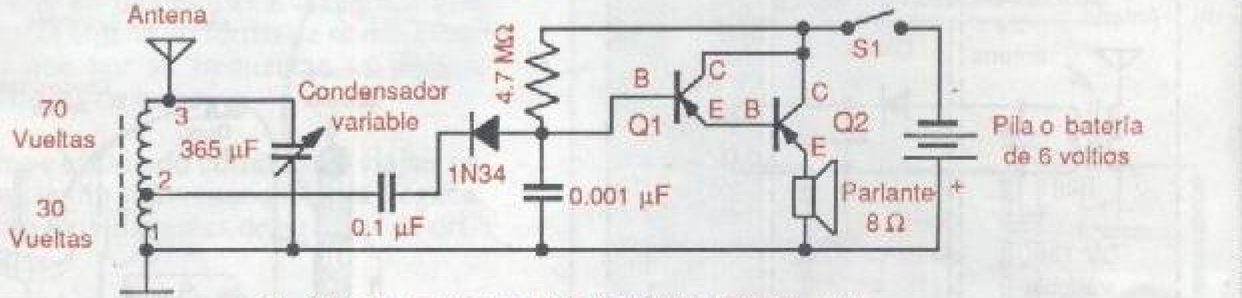
Antena
70
Vueltas
30
Vueltas
3
2
1
Condensador
variable
365 µF
0.1 µF
1N34
4.7 MΩ
0.001 µF
B
C
E
Q1
B
C
E
Q2
S1
Pila o batería
de 6 voltios
+
Parlante
8 Ω

Q1, Q2 : Transistores PNP (2N3906, BC558, etc)

Fig. 269 a

Radio 2

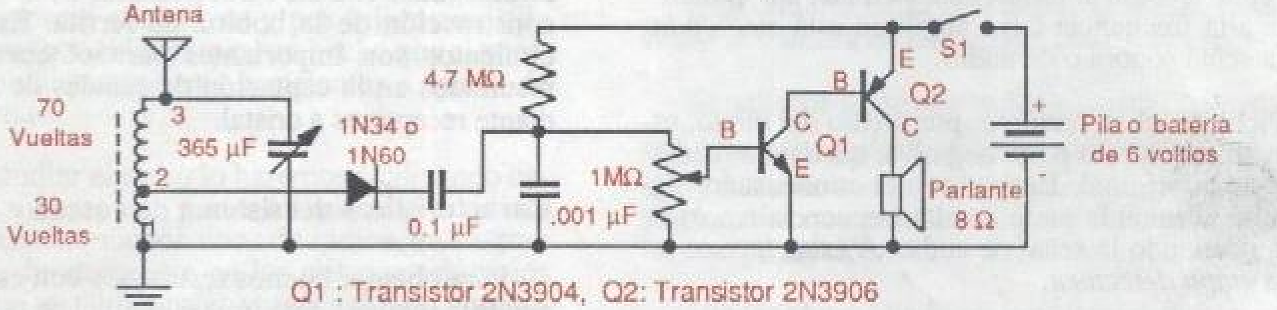
Antena
70
Vueltas
30
Vueltas
3
2
365 µF
1N34 o
1N60
0.1 µF
4.7 MΩ
.001 µF
1MΩ
B
C
E
Q1
B
E
C
Q2
S1
+
Pila o batería
de 6 voltios
Parlante
8 Ω

Q1 : Transistor 2N3904, Q2: Transistor 2N3906

Fig. 269 b

Radio 3

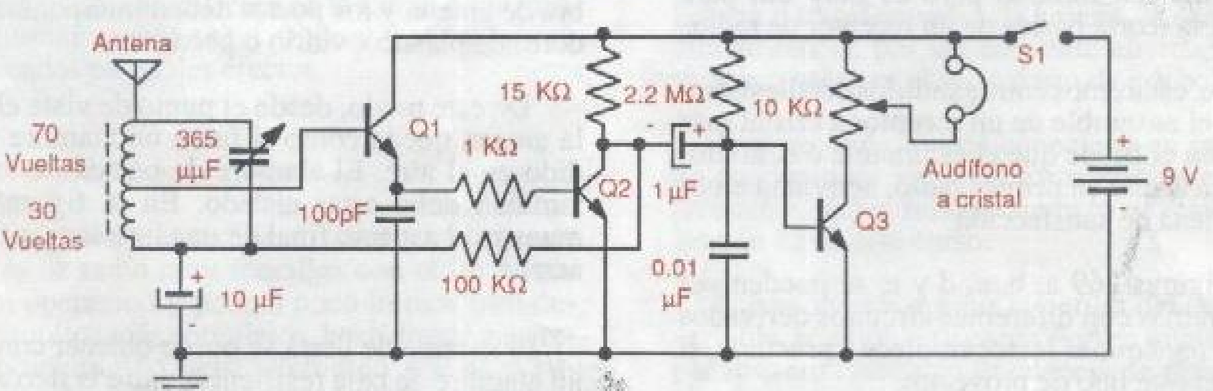
Antena
70
Vueltas
30
Vueltas
365
µµF
Q1
15 KΩ
1 KΩ
100pF
100 KΩ
2.2 MΩ
Q2
-
+
1 µF
0.01
µF
10 KΩ
Q3
Audífono
a cristal
S1
+
9 V
+
-
10 µF

Q1, Q2, Q3 : 2N3904

Fig. 269 c

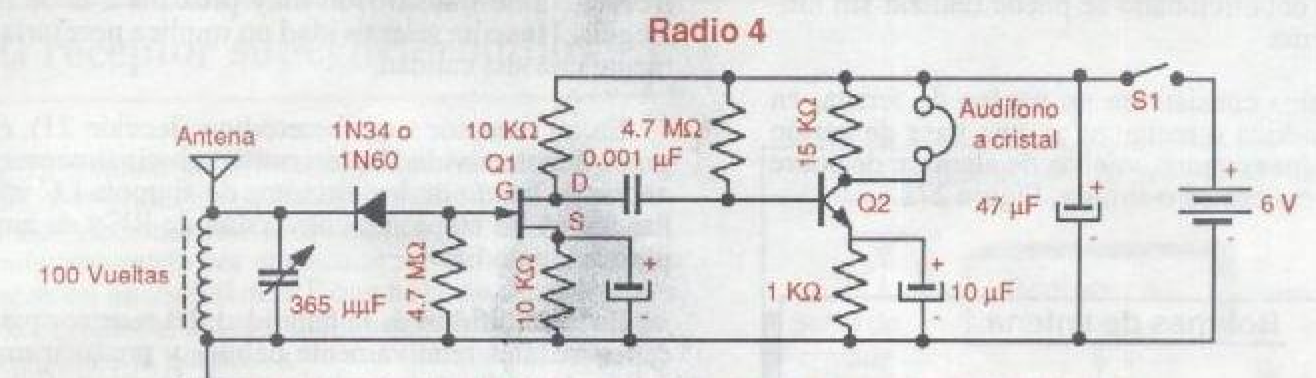
Radio 4
Antena
100 Vueltas
365 µµF
1N34 o
1N60
4.7 MΩ
10 KΩ
Q1
G
D
S
10 KΩ
4.7 MΩ
0.001 µF
15 KΩ
Audífono
a cristal
Q2
1 KΩ
10 µF
47 µF
S1
6 V

Q1: Transistor FET MPF102 o 2N3819, Q2: Transistor NPN 2N3904

El audífono debe ser a cristal, de alta impedancia

Fig. 269 d

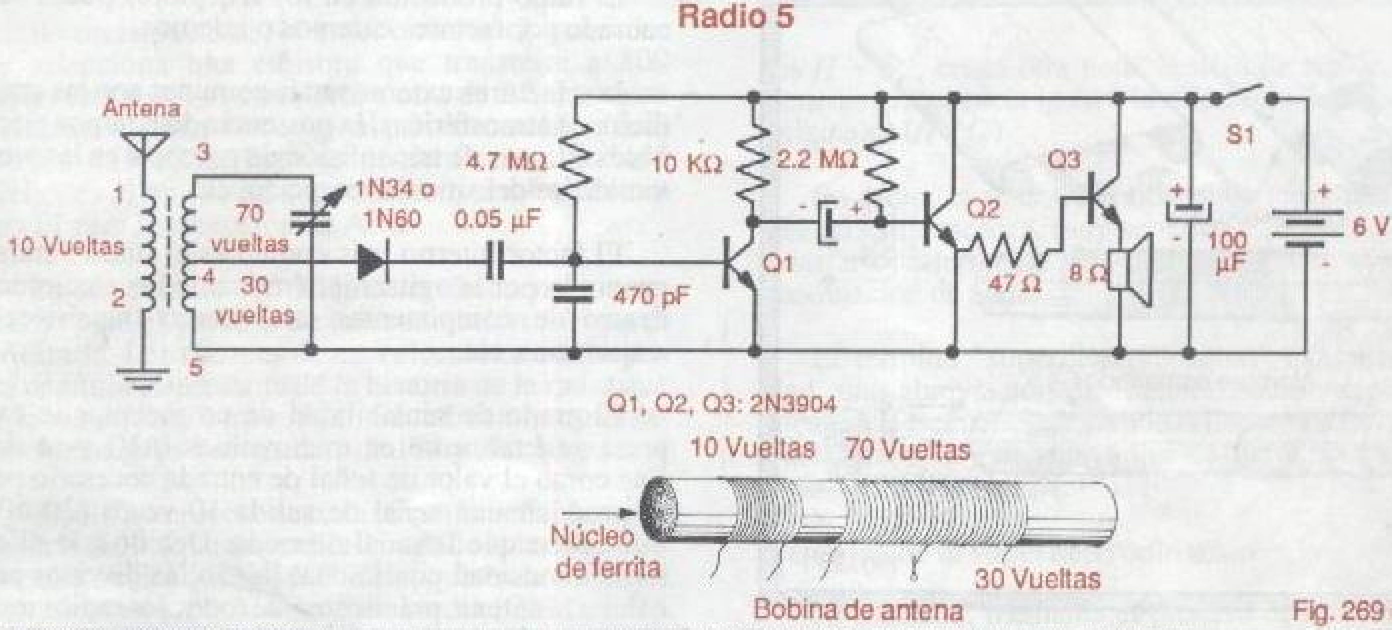
Radio 5
Antena
1
2
3
4
5
10 Vueltas
70
vueltas
30
vueltas
1N34 o
1N60
0.05 µF
470 pF
4.7 MΩ
10 KΩ
2.2 MΩ
Q1
Q2
Q3
47 Ω
8 Ω
100
µF
S1
6 V
10 Vueltas
70 Vueltas
Núcleo
de ferrita
30 Vueltas
Bobina de antena

Q1, Q2, Q3: 2N3904

Fig. 269 e

Si lo desea, Ud. mismo puede construir su propio sistema de tierra, utilizando una varilla gruesa de cobre de 2 ó 3 metros de larga, enterrada profundamente en el suelo.

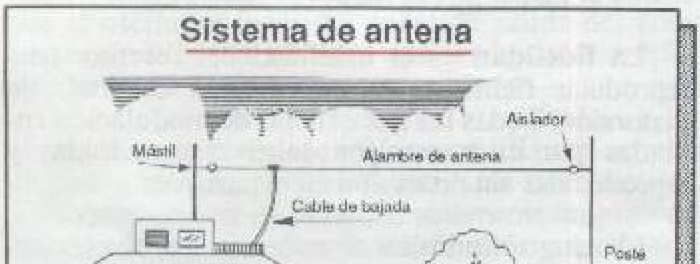
Sistema de antena
Aislador
Mástil
Alambre de antena
Cable de bajada
Poste

Fig. 270

Estas varillas se conocen comercialmente con el nombre de varillas Copperweld y poseen un extremo puntiagudo para facilitar la penetración, y, en el otro llevan una abrazadera metálica para realizar la conexión eléctrica.

La conductividad del suelo que rodea la varilla se puede mejorar regándolo con una solución líquida conductora de electricidad como aguasal, cloruro de cobre, etc. Estas sustancias se conocen con el nombre de *electrolitos* y se obtienen mezclando agua con una sal química cualquiera, incluyendo la de cocina (cloruro de sodio o NaCl).

Construcción de bobinas de antena con núcleo de ferrita

Las bobinas de antena con núcleo de ferrita son las más empleadas en los receptores actuales. Aunque se pueden conseguir fácilmente en los almacenes especializados, su construcción por parte del estudiante o del aficionado se puede realizar sin ningún problema.

La bobina consiste en un núcleo de ferrita, en forma cilíndrica o rectangular, una base de cartón aislante y unas cuantas vueltas de alambre de cobre esmaltado calibre 28 o similar. Figura 271.

Bobinas de antena

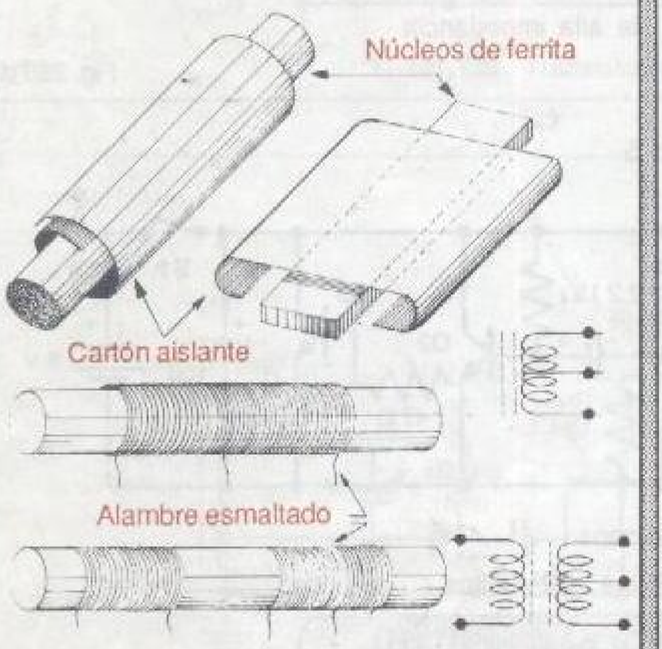

Fig. 271

Según las indicaciones de los diagramas, se deben enrollar firmemente, alrededor del cartón aislante que está sobre el núcleo de ferrita, una determinada cantidad de vueltas o espiras con el alambre de cobre.

A veces se construye una sola bobina con una derivación o "TAP" a cierto número de vueltas. Otras veces se requieren dos bobinas, una primaria y otra secundaria.

Características generales de los receptores de radio

Los receptores de radio, en general, deben reunir ciertas características con el fín de garantizar una recepción de óptima calidad. Las principales son la selectividad, la sensibilidad, la estabilidad, la fidelidad y el rango dinámico.

La **selectividad** es la habilidad del receptor para seleccionar una determinada estación de radio y rechazar las demás, incluso aquellas que tienen una frecuencia de transmisión muy próxima a la de la elegida. Una alta selectividad no implica necesariamente una alta calidad.

En un receptor superheterodino (lección 21), el grado de selectividad lo determina principalmente el *ancho de banda* de los circuitos de sintonía LC utilizados en las etapas de conversión de RF y de amplificación de FI.

La **sensibilidad** es la habilidad del receptor para captar señales relativamente débiles y producir una salida de audio con un bajo nivel de ruido. Entre más alta sea esta *figura de ruido*, más sensible es el receptor. El ruido enmascara las señales débiles y las hace difíciles de procesar.

El ruido producido en los receptores puede ser causado por factores externos o internos.

Los factores externos más comunes son las condiciones atmosféricas, la presencia de equipos eléctricos y líneas de transmisión de potencia en las proximidades del sitio de recepción, etc.

El factor interno más común es el ruido natural generado por la agitación térmica de los electrones dentro de componentes tales como transistores, resistencias, etc.

El grado de sensibilidad de un receptor se expresa generalmente en microvoltios (μV) y se define como el valor de señal de entrada necesario para producir una señal de salida 10 veces (20 dB) más fuerte que la señal de ruido. Debido a la diferente intensidad con la cual llegan las diversas señales a la antena, prácticamente todos los radios modernos están provistos de un circuito de control automático de ganancia (CAG).

La **estabilidad** es la habilidad del receptor para mantener sintonizada la estación elegida. La estabilidad puede ser afectada por factores tales como la temperatura, las variaciones del voltaje de alimentación, etc. La estabilidad la determina principalmente el diseño del oscilador local.

La **fidelidad** es la habilidad del receptor para reproducir fielmente la información original, sin distorsión. Todas las frecuencias de modulación enviadas por el transmisor deben ser recibidas y reproducidas sin distorsión en el parlante.

El **rango dinámico** se refiere a los niveles extremos de señal que puede manejar el receptor, manteniendo una alta sensibilidad y una alta figura de ruido. Se especifica en dB como la relación entre la más alta señal de entrada tolerable sin distorsión y la más baja señal discernible sin ruido.

El receptor superheterodino

El receptor *superheterodino* se llama así porque utiliza el principio de heterodinación o de *batido de señal*. Mediante este proceso, la señal de RF de entrada, entregada por el circuito sintonizador, se convierte en una señal de RF que tiene siempre la misma frecuencia, independientemente de la frecuencia original de la portadora.

A esta nueva frecuencia se le denomina frecuencia intermedia o FI. La señal modulada de FI lleva la misma información de audio de la señal de RF original.

En otras palabras, si en el circuito sintonizador se selecciona una emisora que transmite a 800 KHz, esta señal se convierte a otra de RF que tiene la misma modulación original pero con una frecuencia de 455 KHz. Este valor de frecuencia (455 KHz), es el que se ha adoptado universalmente como FI para la recepción de AM.

La elección de 455 KHz como frecuencia intermedia para el proceso de heterodinación AM no es arbitraria. Las razones de esta elección constituyen un capítulo interesante de la historia de la radio que le recomendamos investigar. Inicialmente se utilizaron frecuencias intermedias de 175 y 262 KHz.

El mismo proceso se realiza con la señal de cualquier emisora que transmita dentro de la banda comprendida entre 550 y 1600 KHz, que es la asignada oficialmente para AM de onda media (OM o MW). Lo que se hace, en síntesis, con la heterodinación es trasladar la modulación original de una portadora cualquiera a una portadora fija de 455 KHz, llamada frecuencia intermedia.

El proceso de heterodinación lo realiza en el receptor la etapa conversora de RF. Este circuito recibe la señal modulada de RF procedente de la antena, la amplifica y la mezcla con la señal producida por el oscilador local. La señal de salida del conversor de RF (455 KHz) se aplica al primer amplificador de FI. Figura 272.

Las etapas de frecuencia intermedia son amplificadores *sintonizados*, es decir, están configurados y ajustados de tal manera que solamente amplifican una banda estrecha de señales, dándole *selectividad* al receptor.

La heterodinación utilizada en radio es similar a la producción de batidos en sonido. En este último caso, dos notas de frecuencias ligeramente diferentes *f1* y *f2* , crean otra nota, la nota de batido, que tiene una frecuencia igual a la diferencia de las notas originales (*f1 - f2*).

El proceso de heterodinación

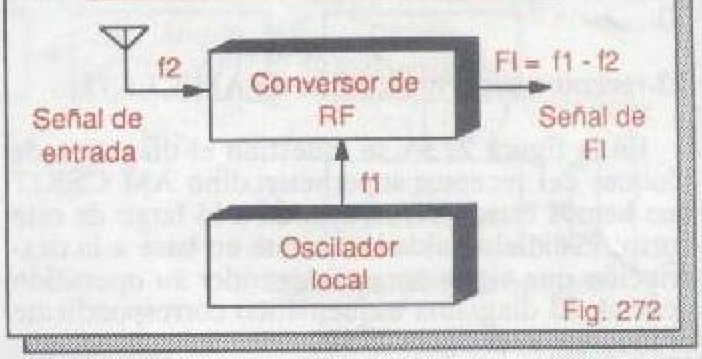

Fig. 272

En radio, la nota de batido debe tener una frecuencia supersónica, muy superior a 20 KHz. De lo contrario, interferiría con el sonido implícito en la modulación de audio.

El término "superheterodinación" es, en realidad, una abreviación de "heterodinación supersónica". A la frecuencia de batido la llamamos frecuencia intermedia y tiene un valor de 455 KHz en AM y de 10.7 MHz en FM.

Ventajas de la superheterodinación

Al trasladar, mediante el proceso de heterodinación, cualquier frecuencia de portadora captada en el circuito de antena a una portadora más baja, de frecuencia fija, que conserva la misma modulación original y a la cual ocurre la máxima ganancia y selectividad posibles, se obtienen, entre otras, las siguientes ventajas:

• Una muy alta estabilidad, porque la posibilidad de realimentación positiva y, en consecuencia, de oscilaciones indeseables de los amplificadores de alta frecuencia, es más baja cuanto más baja sea la frecuencia de operación.

• Una muy alta sensibilidad porque entre más amplificadores de FI se puedan utilizar (dos son generalmente suficientes), más grande es la ganancia que se puede obtener.

De hecho, la gran sensibilidad de los receptores modernos de radio es una consecuencia del elevado grado de amplificación que se consigue, sin dificultad, en las etapas de frecuencia intermedia.

• Una muy alta selectividad porque se pueden emplear cuantos circuitos sintonizados de FI se quieran, sin crear los problemas de seguimiento que se originan cuando se intenta ajustar varios condensadores variables al mismo tiempo.

Los términos estabilidad, sensibilidad y selectividad se discuten con mayor detalle en la lección 20.

El receptor superheterodino de AM CEKIT

En la figura 273A se muestran el diagrama de bloques del receptor superheterodino AM CEKIT que hemos estado construyendo a lo largo de este curso. Estúdielo cuidadosamente en base a la descripción que sigue para comprender su operación general. El diagrama esquemático correspondiente se muestra en la figura 273B.

Nuestro receptor superheterodino consta básicamente de las siguientes etapas:

- Un amplificador de radiofrecuencia (RF);
- Un mezclador o conversor;
- Un oscilador local;
- Dos amplificadores de frecuencia intermedia (FI);
- Un detector o demodulador y
- Un amplificador de audio.

El amplificador de RF incluye el circuito de antena y constituye junto con el mezclador la etapa conversora de RF, que es la responsable de la heterodinación. Veamos ahora cómo se realiza el flujo de las diferentes señales a lo largo del sistema y qué hace cada bloque.

El circuito de antena o de sintonía es un tanque resonante LC formado por un condensador variable (CV1) y una bobina de ferrita (L1). Su propósito es seleccionar una señal de entrada específica entre las muchas que llegan a la antena.

Ajustando el condensador variable, sintonizamos el tanque LC a una determinada frecuencia. Las señales moduladas de RF enviadas por todas las estaciones de radio llegan a la antena, pero el circuito de sintonía solo deja pasar aquella que tenga una frecuencia de portadora igual a su frecuencia natural de resonancia, digamos 950 KHz.

La señal captada por el circuito de sintonía se transfiere al amplificador de RF donde se amplifica a un nivel adecuado. La señal de salida del amplificador de RF se mezcla o heterodina con la señal producida por el oscilador local para producir una señal que tiene siempre una frecuencia intermedia (FI) de 455 KHz pero conserva la misma modulación de amplitud original.

La señal suministrada por el oscilador local es una onda sinusoidal pura, esto es, no contiene ningún tipo de modulación. Su frecuencia la determinan el condensador variable CV2 y la bobina osciladora L2 ("roja").

Los condensadores CV1 y CV2 trabajan en "tandem", es decir, varían su capacidad al mismo tiempo. De este modo, se garantiza que la diferencia de frecuencia entre las señales del oscilador local y la de antena sea siempre de 455 KHz. Es decir, la frecuencia del oscilador local es siempre 455 KHz más alta que la frecuencia captada. Esta es una condición necesaria para la heterodinación.

Tanto la amplificación de RF como la generación de la señal local y la heterodinación de las mismas las realiza el transistor Q1 en asocio con los demás componentes de la etapa conversora. La señal de salida del mezclador se aplica al primer amplificador de FI mediante el transformador T1.

Es importante destacar que cuando las señales de antena y del oscilador local se heterodinizan, el mezclador produce las frecuencias originales así como sus respectivas frecuencias suma y diferencia.

Sin embargo, la salida del mezclador se sintoniza para que entregue únicamente la *diferencia* entre las frecuencias local y de antena

Los dos amplificadores de FI realizan sucesivas amplificaciones de la señal modulada de FI suministrada por T1 hasta que ésta alcanza un nivel lo suficientemente alto como para excitar el detector. La señal de salida de la primera etapa de FI se suministra a la segunda a través del transformador T2.

La señal de salida de la segunda etapa de FI se suministra al detector a través del transformador T3. La amplificación en la primera etapa de FI la realiza el transistor Q2 y en la segunda la efectúa el transistor Q3.

La señal que ingresa al detector es una señal de alta frecuencia (455 KHz) modulada en amplitud por la información de audio que se desea reproducir. El detector elimina la componente de alta frecuencia y deja pasar únicamente la señal de audio que la modula. El componente principal del circuito de detección es el diodo D1.

La señal de salida del detector (audio) se aplica al amplificador de audio, alcanzando una potencia lo suficientemente alta como para impulsar el parlante. En este último, las variaciones de amplitud de la señal de entrada se convierten en sonidos que son una réplica de la información de audio (palabras, música, etc) originalmente enviada por la emisora.

Receptor CEKIT Superheterodino de 6 transistores

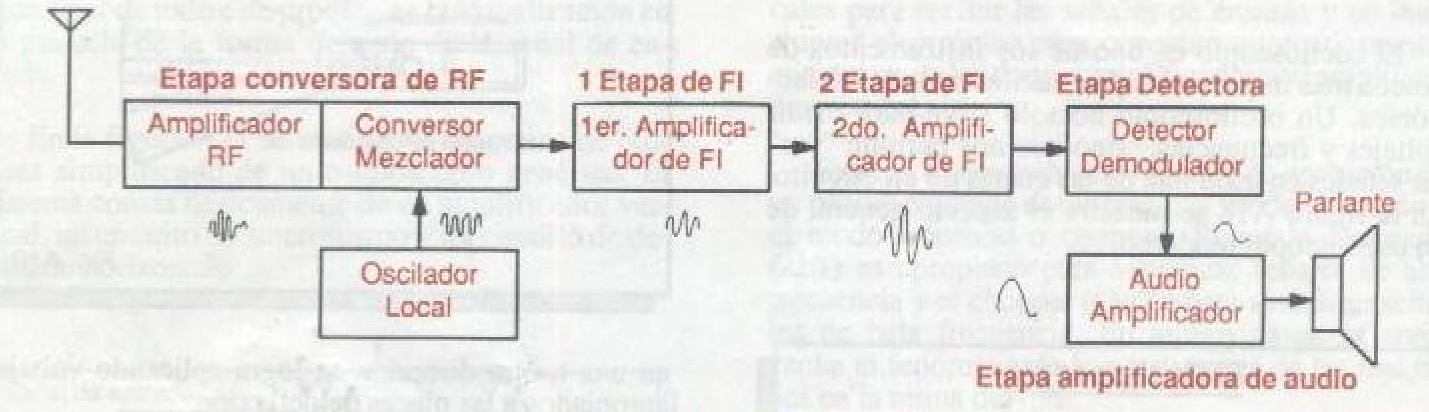
Etapa conversora de RF
Amplificador RF
Conversor Mezclador
Oscilador Local
1 Etapa de FI
1er. Amplificador de FI
2 Etapa de FI
2do. Amplificador de FI
Etapa Detectora
Detector Demodulador
Parlante
Audio Amplificador
Etapa amplificadora de audio

DIAGRAMA DE BLOQUES

Fig. 273 a

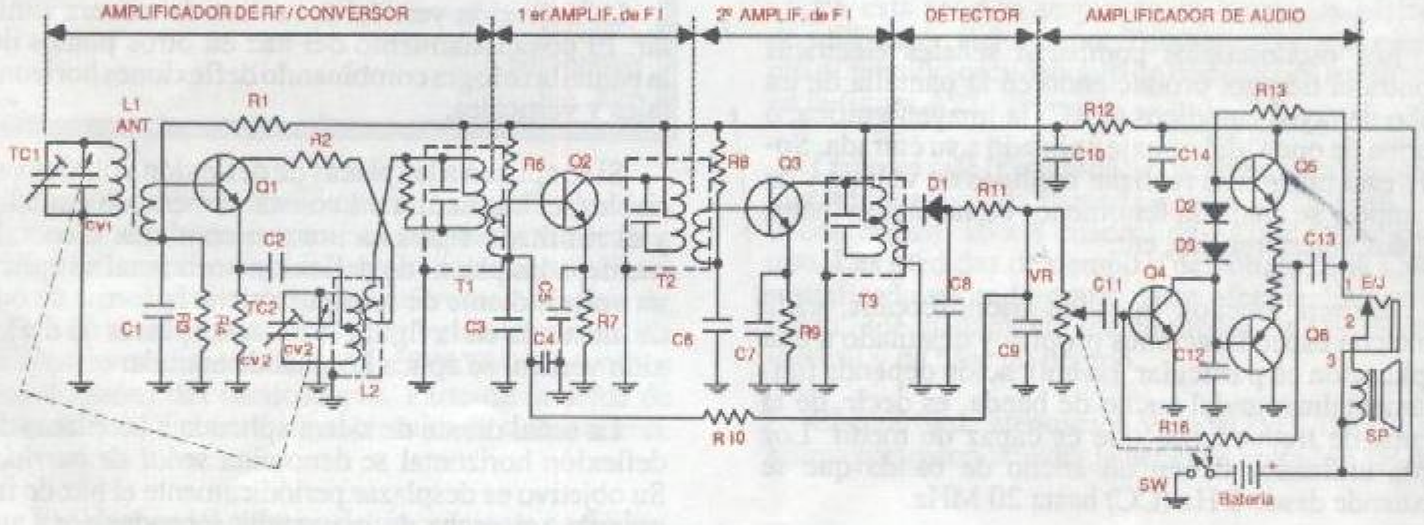
AMPLIFICADOR DE RF / CONVERSOR
1 er AMPLIF. de F I
2º AMPLIF. de F I
DETECTOR
AMPLIFICADOR DE AUDIO
TC1
CV1
L1
ANT
R1
R2
Q1
C1
C2
R3
R4
TC2
CV2
L2
T1
R5
R6
Q2
C3
C4
C5
R7
T2
R8
Q3
C6
C7
R9
R10
T3
D1
R11
C8
VR
C9
C10
R12
C11
Q4
C12
C14
R13
Q5
D2
D3
C13
Q6
R16
E/J
1
2
3
SP
SW
Batería

DIAGRAMA ESQUEMATICO

Fig. 273 b

Los componentes responsables de la amplificación de la señal de audio son los transistores Q4, Q5 y Q6. En particular, Q4 actúa como preamplificador mientras Q5 y Q6 lo hacen como amplificadores de potencia, dispuestos en una configuración llamada par complementario o "push-pull".

Resumiendo, el radio AM CEKIT es un receptor superheterodino de 6 transistores. El sistema capta transmisiones de AM enviadas por estaciones de radiodifusión que operan en la banda de 550 a 1600 KHz y utiliza una portadora interna estándar de 455 KHz llamada frecuencia intermedia (FI).

Actividad práctica Nº 12

Conocimiento y manejo del osciloscopio

El osciloscopio es uno de los instrumentos de prueba más importantes y versátiles de toda la electrónica. Un osciloscopio no solo sirve para medir voltajes y frecuencias, sino que nos permite "ver" las señales en cada una de las etapas de un circuito. En la figura A18 se muestra el aspecto general de un osciloscopio moderno.

El Osciloscopio

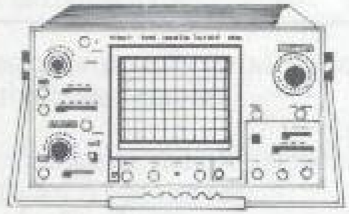

Fig. A18

Los osciloscopios comparan señales eléctricas contra el tiempo, produciendo en la pantalla de un tubo de rayos catódicos (TRC) la imagen gráfica, o forma de onda, del voltaje aplicado a su entrada. Sobre esta imagen se realizan medidas de voltaje y de tiempo y se analizan fenómenos como distorsiones, ruidos, interferencias, etc.

Existen muchos tipos de osciloscopios, cada uno con especificaciones propias y destinado a una aplicación en particular. La aplicación depende fundamentalmente del ancho de banda, es decir, de la gama de frecuencias que es capaz de medir. Los más utilizados tienen un ancho de banda que se extiende desde 0 Hz (CC) hasta 20 MHz.

Cómo trabaja un osciloscopio

El corazón de un osciloscopio es un tubo de rayos catódicos o TRC similar al de un televisor (figura A19). Un TRC típico consta de un cañon electrónico, una pantalla recubierta de fósforo y dos pares de placas de deflexión. Estas últimas se denominan placas de deflexión horizontal y placas de deflexión vertical, respectivamente.

Principio de operación del osciloscopio

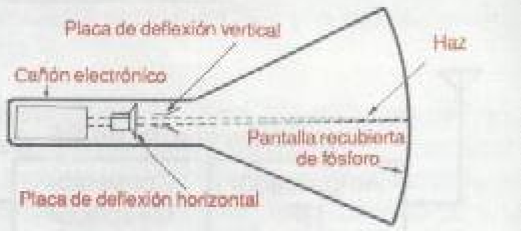
Placa de deflexión vertical
Haz
Cañón electrónico
Pantalla recubierta de fósforo
Placa de deflexión horizontal

Fig. A19

El cañon emite un haz o chorro de electrones. Este haz es desviado horizontal o verticalmente por las placas de deflexión y se estrella a gran velocidad contra el fósforo de la pantalla, provocando la aparición de un punto luminoso. La desviación del haz en una u otra dirección se logra aplicando voltajes apropiados a las placas de deflexión.

Si los voltajes netos aplicados a las placas de deflexión vertical y horizontal son iguales a cero, el haz no se desvía y el resultado es un punto en el centro de la pantalla. Si el voltaje de la placa derecha es más positivo que el de la placa izquierda o viceversa, el haz se desvía hacia la derecha o hacia la izquierda y el resultado es un punto a la derecha o a la izquierda del centro.

La deflexión vertical se realiza de manera similar. El posicionamiento del haz en otros puntos de la pantalla se logra combinando deflexiones horizontales y verticales.

Si se aplican a las placas de deflexión voltajes variables, el haz explorará constantemente la pantalla y el resultado será una imagen continua. Generalmente, a las placas de deflexión horizontal se aplica un voltaje diente de sierra que tiene la forma de onda mostrada en la figura A20. A las placas de deflexión vertical se aplica la señal de entrada.

La señal diente de sierra aplicada a las placas de deflexión horizontal se denomina *señal de barrido*. Su objetivo es desplazar periódicamente el haz de izquierda a derecha de la pantalla y producir, en ausencia de señal de entrada, un trazo horizontal visible llamado *barrido*.

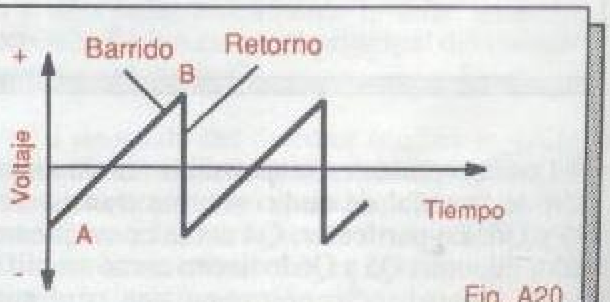

Fig. A20

El efecto de la señal aplicada a las placas de deflexión vertical es desviar el haz hacia arriba y hacia abajo a medida que se realiza el barrido. La cantidad y dirección de la desviación dependen del valor instantáneo y de la polaridad de la misma. El resultado final de todo este proceso es la visualización en la pantalla de la forma de onda de la señal de entrada.

En la figura A21 se muestra el diagrama de bloques simplificado de un osciloscopio genérico. El sistema consta básicamente de un amplificador vertical, un circuito de sincronismo y un circuito de deflexión horizontal.

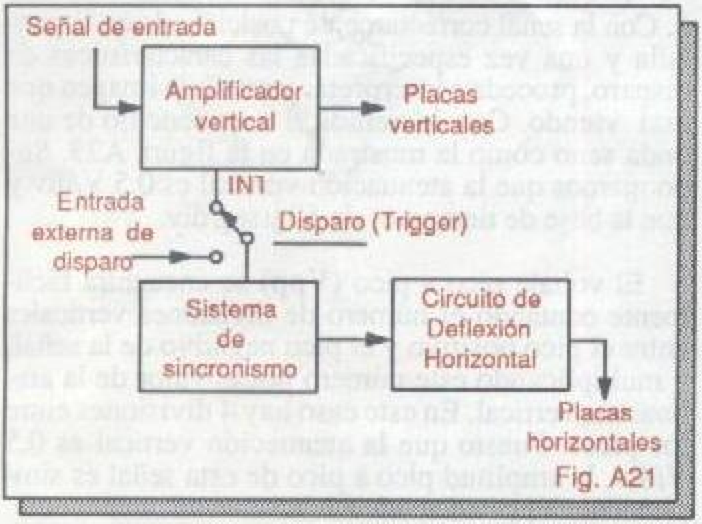
Señal de entrada
Amplificador vertical
Placas verticales
IN1
Entrada externa de disparo
Disparo (Trigger)
Sistema de sincronismo
Circuito de Deflexión Horizontal
Placas horizontales

Fig. A21

La señal de entrada se aplica al amplificador vertical y la salida de éste excita las placas verticales. El amplificador vertical incluye un atenuador calibrado que le permite manejar diferentes niveles de señal. El nivel de atenuación requerido lo selecciona el usuario mediante el control "VOLTS/DIV" (voltios por división) del osciloscopio. Parte de la señal de entrada se aplica también al circuito de sincronismo.

El objetivo del circuito de sincronismo es sincronizar la generación o *disparo* de la señal de barrido interna con la señal de entrada o con una señal externa, logrando que ambas señales, la de barrido y la de disparo, tengan la misma frecuencia y la misma fase. La fuente de la señal de disparo, interna o externa, se selecciona mediante el interruptor "SOURCE" (fuente) del osciloscopio.

La señal de barrido propiamente dicha la genera el circuito de deflexión horizontal. La salida de este último excita las placas horizontales. El período exacto de la señal de barrido lo programa el usuario mediante el *control de base de tiempo* del osciloscopio. Este control está usualmente etiquetado como "SEC/DIV" (segundos por división).

Osciloscopios de doble trazo o de dos canales

Los osciloscopios de doble trazo permiten visualizar simultáneamente dos imágenes diferentes en la misma pantalla. Utilizan dos preamplificadores verticales para recibir las señales de entrada y un interruptor electrónico para conectar automáticamente cualquiera de los dos preamplificadores al amplificador vertical.

Existen básicamente dos modos de operación de un osciloscopio de doble trazo: el modo alternado y el modo troceado o chopper. El modo alternado (ALT) es apropiado para visualizar señales de alta frecuencia y el chopper (CHOP) para visualizar señales de baja frecuencia. En ambos casos se aprovecha el fenómeno de la persistencia de las imágenes en la retina del ojo.

Manejo del osciloscopio paso a paso

El osciloscopio es uno de los instrumentos de medición y prueba más importantes que se pueden poseer para el trabajo en circuitos de radio y comunicaciones. Para lograr aprovechar al máximo sus posibilidades, es importante aprender a manejarlo correctamente.

En esta sección se presentan las reglas básicas de utilización del mismo. Tomaremos como ejemplo el sencillo osciloscopio de dos canales mostrado en la figura A22.

Observe que la pantalla se caracteriza por poseer una rejilla calibrada llamada *gratícula*. La gratícula la conforman 10 x 8 cuadros de 1 cm de lado cada uno. Las medidas de tiempo y de voltaje de la señal visualizada sobre la gratícula se efectúan interpretando adecuadamente los controles de atenuación vertical y de base de tiempo.

El control de atenuación vertical (VOLTS/DIV) se utiliza para acondicionar la señal de entrada a los re-

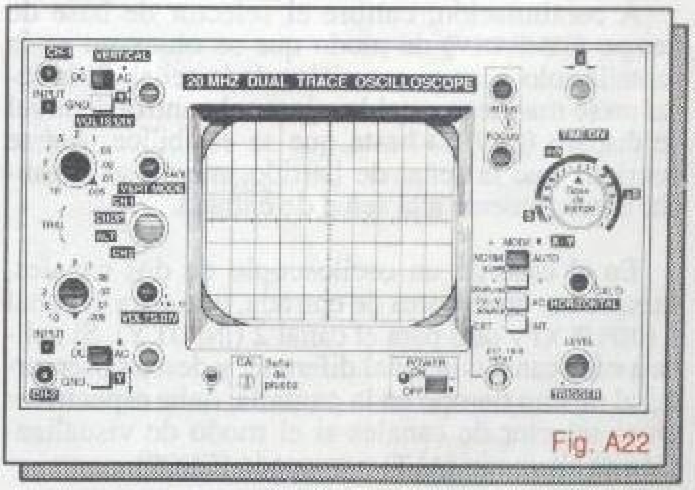
20 MHZ DUAL TRACE OSCILLOSCOPE
CH1
VERTICAL
INPUT
DC
AC
GND
VOLTS/DIV
VERT MODE
CH1
CHOP
ALT
CH2
VOLTS/DIV
INPUT
CH2
CAL
POWER
ON
OFF
INTEN
FOCUS
TIME/DIV
MODE
AUTO
NORM
HORIZONTAL
EXT
LEVEL
TRIGGER

Fig. A22

quisitos del amplificador vertical. El atenuador vertical posee generalmente un control auxiliar no calibrado (CAL'D) que se utiliza para alterar la amplitud original de la imagen en la pantalla.

El control de base de tiempo (SEC/DIV) se utiliza para establecer el período de la señal de barrido generada por el circuito de deflexión horizontal. Posee también un control auxiliar no calibrado (CAL'D) que se utiliza para alterar el período original.

Para la realización de medidas exactas de voltaje y de tiempo, los controles no calibrados se deben girar completamente hacia la derecha. Para proceder a la visualización de una determinada forma de onda en la pantalla, y realizar mediciones sobre esta imagen, siga estos pasos:

1. Conecte el osciloscopio a la red de CA y oprima el botón de encendido (POWER). El osciloscopio debe conectarse al voltaje especificado por el fabricante en el manual de operación, 110 ó 220 V.

2. Obtenga sobre la pantalla un trazo horizontal. Sitúe el selector de disparo (MODE) en la posición AUTO (automático), el de fuente (SOURCE) en la posición INTERNAL (interna), el de base de tiempo (TIME/DIV) en una posición intermedia y los controles de posición de trazo horizontal (X) y vertical (Y) a mitad de rango.

Ajuste entonces el control de intensidad (INTENSITY) y manipule los controles de posición X e Y hasta que aparezca el trazo. Utilice el control de enfoque (FOCUS) para mejorar la nitidez del trazo y mantenga la intensidad tan baja como sea posible.

3. Una vez obtenido el trazo, céntrelo en la pantalla mediante los controles de posición X y Y. Inyecte la señal por medir al conector de entrada (INPUT) y especifique si es continua (DC) o alterna (AC). Calibre el selector de atenuación vertical (VOLTS/DIV) de modo que la señal visualizada quede contenida verticalmente en la pantalla.

A continuación, calibre el selector de base de tiempo (TIME/DIV) de modo que se observen en la pantalla solo unos pocos ciclos de la señal. Si la señal no se mantiene estable, ajuste el control de nivel de disparo (LEVEL) hasta que se estabilice. Así se consigue que la señal de barrido interna se enganche correctamente a la señal de entrada.

En el caso de un osciloscopio de dos canales, existen dos conectores de entrada, uno para el canal 1 (INPUT X) y otro para el canal 2 (INPUT Y). Si aplica a cada canal una señal diferente y desea observarlas al mismo tiempo en la pantalla, debe especificar en el selector de canales si el modo de visualización es alternado (ALT) o troceado (CHOP).

El modo de visualización depende de la base de tiempo seleccionada. En general, el modo alternado (ALT) se utiliza con bases de tiempo inferiores a 1 mseg/div y el troceado o chopper (CHOP) con bases de tiempo iguales o superiores a ese valor.

4. Especifique las características de disparo de la señal de barrido. La mayoría de osciloscopios admiten dos modos de disparo, automático (AUTO) y NORMAL. El automático es el más utilizado.

El disparo automático puede ocurrir con los flancos de subida (+) o de bajada (-) de la señal de entrada, dependiendo de la posición del control de pendiente (SLOPE). El disparo manual puede ocurrir en cualquier punto específico de la forma de onda.

5. Con la señal correctamente posicionada en la pantalla y una vez especificadas las características de disparo, proceda a interpretar y medir la imagen que está viendo. Consideremos el caso sencillo de una onda seno como la mostrada en la figura A23. Supongamos que la atenuación vertical es 0.5 V/div y que la base de tiempo es de 50 µseg/div.

El voltaje pico a pico (**Vpp**) se encuentra fácilmente contando el número de divisiones verticales entre el pico positivo y el pico negativo de la señal, y multiplicando este número por el valor de la atenuación vertical. En este caso hay 4 divisiones entre los picos. Puesto que la atenuación vertical es 0.5 V/div, la amplitud pico a pico de esta señal es simplemente **Vpp** = 4 x 0.5 = **2 V**.

Para conocer la frecuencia (f) de esta señal, debe primero conocerse su período (T). El valor de T se encuentra fácilmente contando el número de divisiones horizontales entre dos picos consecutivos y multiplicando este número por el valor de la base de tiempo. En este caso, hay 5 divisiones entre los dos picos positivos.

Puesto que la base de tiempo es 50 µseg/div, el período de la señal es simplemente **T**= 5 x 50 = **250 µseg** y su frecuencia es **f**=1/T = **4 KHz** .

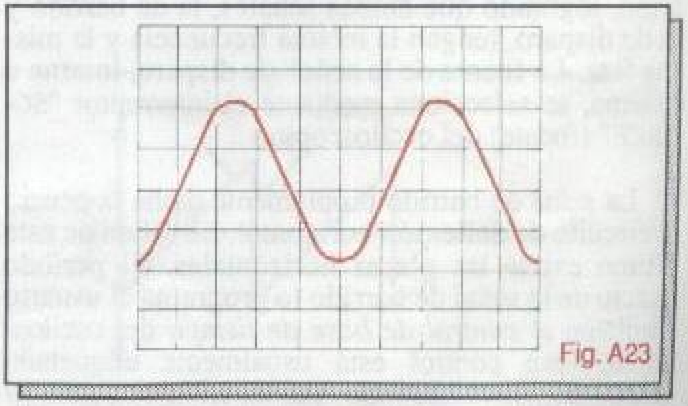

Fig. A23

La etapa detectora

Como usted lo habrá notado, el estudio y ensamble del radio CEKIT lo estamos realizando de derecha a izquierda en cuanto a sus etapas se refiere. Ya estudiamos y ensamblamos la etapa "Amplificador de audio" y ahora estudiaremos la etapa detectora. Al final tendremos el radio completo funcionando y conoceremos todos los detalles sobre su teoría y ensamble en ese orden.

Circuitos detectores

Los detectores o demoduladores (figura 274) son circuitos que recuperan la información de baja frecuencia (modulación o audio) de una señal modulada y eliminan la componente de alta frecuencia o portadora de la misma. La señal de salida suministrada por un detector conserva las mismas características de la señal moduladora original.

Función básica de un detector

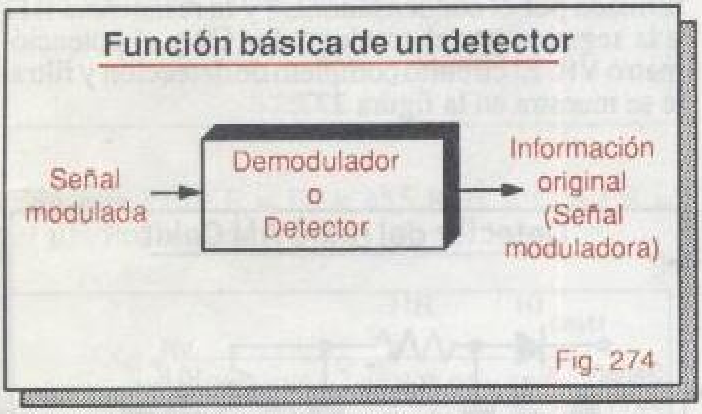

Fig. 274

Todos los detectores efectúan la misma función básica descrita anteriormente pero cada uno tiene características propias, dependiendo del proceso de modulación utilizado. Existen, por lo tanto, detectores de AM, de FM, de SSB, etc. En este capítulo nos referimos exclusivamente a los detectores de AM y, específicamente, al detector utilizado en el radio AM CEKIT.

Características de los detectores

Las características más notables de un detector son su linealidad, su sensibilidad, su selectividad y su capacidad.

La linealidad se refiere a la habilidad de reproducir fielmente, sin distorsión, la señal original. La sensibilidad es la capacidad de detectar señales pequeñas. La selectividad es la capacidad de responder únicamente a una determinada banda de frecuencias. La capacidad se refiere a la habilidad para manejar señales grandes, sin producir distorsión ni sobrecarga.

El proceso de detección AM

Como sabemos, la salida de la etapa amplificadora mezcladora de RF es una señal que tiene la misma información de audio (modulación) de la señal de RF originalmente enviada por la emisora sintonizada pero con una frecuencia de 455 KHz. A esta frecuencia se le denomina frecuencia intermedia o FI.

La señal de FI se amplifica en dos etapas con el fin de hacerla lo suficientemente fuerte como para ser detectada y superar así el voltaje de umbral impuesto por el diodo detector. La detección consiste en separar la señal moduladora de audio de la señal portadora de FI.

Como consecuencia de este doble proceso de amplificación de FI, a la salida de la segunda etapa de FI se tiene una señal de 455 KHz que tiene la misma información de audio de la señal de RF original.

Hasta el momento, la señal de RF originalmente sintonizada ha pasado por dos procesos: el de mezcla y el de amplificación de FI. Durante el primero, se cambió su frecuencia y durante el segundo se aumentó su amplitud. Sin embargo, en ambos casos, se ha conservado la información fundamental de audio (modulación). El proceso se describe gráficamente en la figura 275.

Procesos previos a la detección AM

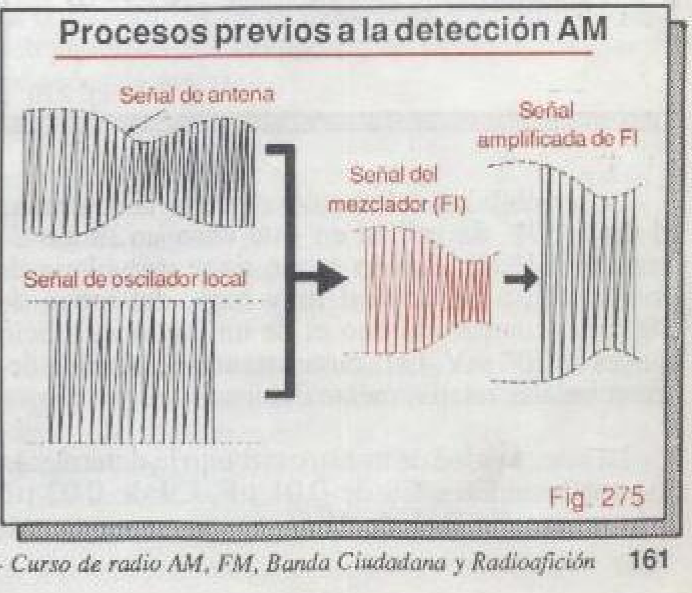

Fig. 275

La modulación de la señal de FI contiene la información de audio (palabras, música, etc.) que deseamos reproducir en el parlante de nuestro receptor de radio.

La componente de alta frecuencia de la señal de RF se usa solo como una *portadora* que lleva la señal de audio desde la antena transmisora hasta el receptor. La componente de alta frecuencia de la señal de FI se usa también como una portadora que transporta la señal de audio a través de las dos etapas amplificadoras de FI.

Una vez completado el proceso de amplificación de FI, la portadora de 455 KHz ya no es necesaria. La función del detector es precisamente detectar o extraer la información de audio (señal de modulación) y eliminar la portadora de FI.

El proceso de detección se realiza rectificando primero la señal compuesta de FI mediante el diodo D1 y filtrando a continuación la señal resultante mediante una red RC formada por la resistencia R11 y los condensadores C8 y C9. El proceso de detección se ilustra gráficamente en la figura 276.

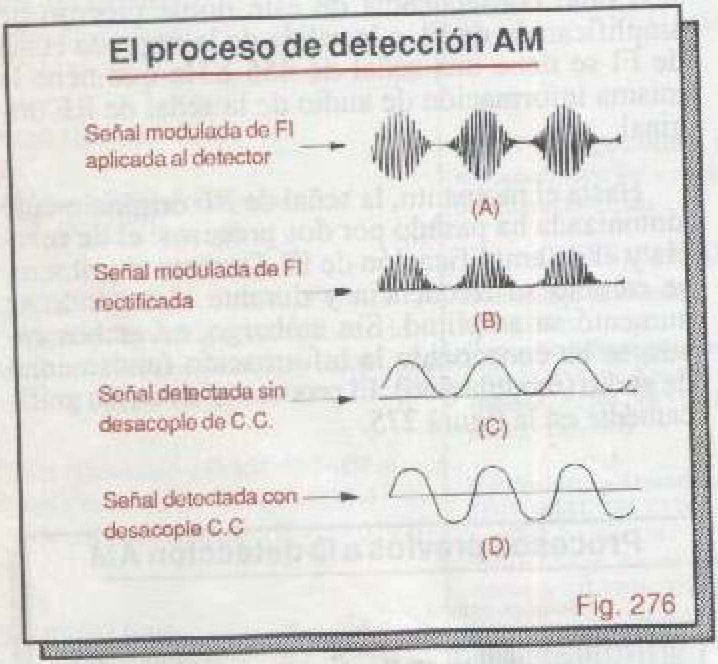
El proceso de detección AM
Señal modulada de FI aplicada al detector
(A)
Señal modulada de FI rectificada
(B)
Señal detectada sin desacople de C.C.
(C)
Señal detectada con desacople C.C
(D)

Fig. 276

La sensibilidad del circuito detector la determina el diodo D1. Se utiliza en este caso un diodo de germanio 1N60, debido a que tiene un voltaje de conducción o de umbral muy bajo, del orden de 300 mV, comparado con el de un diodo de silicio que es de 700 mV. Esta circunstancia le permite detectar señales relativamente débiles.

La selectividad de nuestro circuito la determinan los condensadores C8 de 0.01 µF, C9 de 0.02 µF y la resistencia R11 de 1KΩ. Como veremos más adelante, estos componentes configuran un filtro pasabajo que rechaza señales fuera de la banda de audio cuyas frecuencias están entre 20 Hz y 20 KHz.

Cómo opera la red filtradora de FI

Si no dispusiéramos de la red filtradora RC anterior, en el ánodo del diodo detector D1, observaríamos una señal como la que se muestra en la figura 276B, consistente en un tren de pulsos de RF de frecuencia constante y amplitud variable. Las variaciones de amplitud corresponden a la información de audio que deseamos recuperar.

El condensador C8, se carga al valor pico de cada pulso cuando el diodo conduce y se descarga parcialmente a través de la resistencia R11 y el potenciómetro VR cuando no conduce. El voltaje a través de C8, es un voltaje de CC variable que tiene la misma forma de onda de la modulación original de audio.

La red filtradora utilizada en el detector del radio AM CEKIT, es fundamentalmente un filtro RC pasa bajo de dos secciones. La primera sección está formada por el condensador C8 y la resistencia R11 y la segunda por el condensador C9 y el potenciómetro VR. El circuito completo de detección y filtraje se muestra en la figura 277.

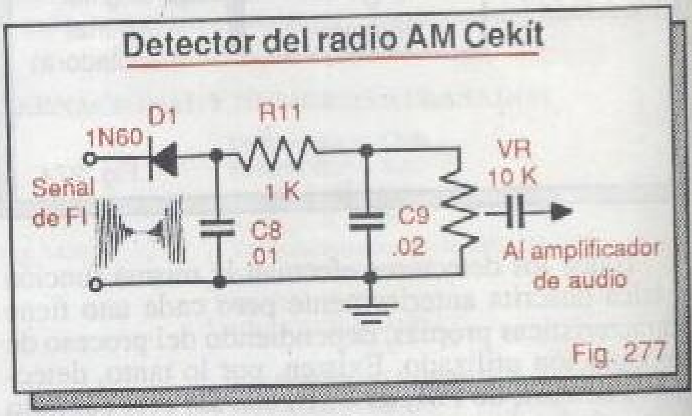
Detector del radio AM Cekit

Fig. 277

El condensador C8 presenta una reactancia capacitiva (X_C) muy baja, a la componente de alta frecuencia de la señal rectificada. Esta reactancia es del orden de 35 Ω, como se comprueba en el ejemplo 1. Este valor es muy pequeño comparado con el de la resistencia R11 de 1 KΩ.

Como consecuencia de lo anterior, la componente de alta frecuencia o portadora de FI, se deriva casi por completo a tierra a través de C8. La componente de baja frecuencia o modulación, encuentra en C8 una alta reactancia y por lo tanto no se deriva a tierra sino que continúa en dirección del amplificador de audio.

La reactancia capacitiva ofrecida por C8, a la componente de audio, es siempre mayor que el valor de R11.

Para una frecuencia de modulación de 1 KHz, por ejemplo, esta reactancia es del orden de 16 KΩ, como se comprueba en el ejemplo 2.

Recordemos que la máxima frecuencia de modulación que se puede enviar por AM es de 4.5 KHz.

Ejemplo 1. Calcular la impedancia o reactancia capacitiva a tierra que ofrece el condensador C8 de la red filtradora de FI, a la componente de alta frecuencia presente en el ánodo del detector de la figura 277.

Note que C8 = 0.01 µF y recuerde que FI=455 KHz.

Solución. La reactancia capacitiva u oposición que presenta un condensador al paso de una señal de corriente alterna se calcula de acuerdo a la fórmula :

$$X_C = \frac{1}{2\pi \times F \times C}$$

$$X_C = \frac{1}{6.28 \times F \times C}$$

En este caso, F = FI = 455 KHz y C = C8 = 0.01 µF. Por lo tanto:

$$X_C = \frac{1}{6.28 \times 455 \times 10^3 \times 0.01 \times 10^{-6}}$$

$$X_C = 34.9\ \Omega \approx 35\,\Omega$$

Ejemplo 2. Calcular la reactancia capacitiva que ofrece el condensador C8 de la red filtradora de FI a una señal de modulación de 1 KHz presente en la salida del detector de la figura 277.

Solución. La reactancia capacitiva la calculamos nuevamente mediante la fórmula:

$$X_C = \frac{1}{2\pi \times F \times C}$$

$$X_C = \frac{1}{6.28 \times F \times C}$$

En este caso, F = 1 KHz y C = C8 = 0.01 µF. Por lo tanto:

$$X_C = \frac{1}{6.28 \times 1 \times 10^3 \times 0.01 \times 10^{-6}}$$

$$X_C = 15923\Omega \approx 16K\Omega$$

Resumiendo, el condensador C8 presenta una muy baja oposición o reactancia al paso de la componente de alta frecuencia implícita en la señal modulada de FI, y, una muy alta al paso de la componente de audio de la misma. En consecuencia, C8 envía a tierra la portadora de FI y deja pasar la señal de audio. Figura 278.

Circuitos equivalentes del detector

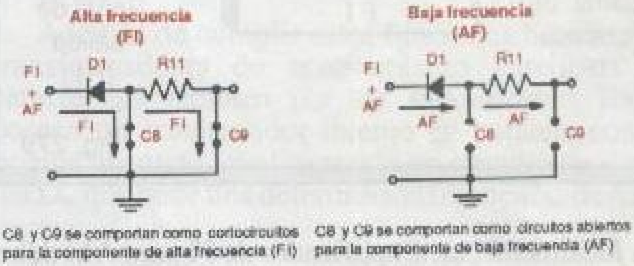

C8 y C9 se comportan como cortocircuitos para la componente de alta frecuencia (FI) C8 y C9 se comportan como circuitos abiertos para la componente de baja frecuencia (AF)

Fig. 278

El condensador C9 envía a tierra cualquier rastro de la portadora de FI que ho haya sido filtrado por C8. El potenciómetro VR (control de volumen) actúa como divisor de tensión de la señal modulada, entregando un determinado nivel de la misma al amplificador de audio.

La consideración de diseño más importante del detector es garantizar que la constante de tiempo del filtro RC tenga un valor intermedio entre la duración de un ciclo de FI (≈2 µseg) y la duración de un ciclo promedio de audio (1 mseg ó 1000 µseg). En este caso, RC ≈ (R11+VR) x C8 ≈ 110 µseg.

Si este valor es muy grande, C8 se cargará y descargará muy lentamente y si es muy pequeño, lo hará muy rápidamente. En el primer caso, sería incapaz de seguir fielmente la modulación de audio y, en el segundo, seguiría la señal de FI y no habría detección.

Etapas amplificadoras de frecuencia intermedia (FI)

Introducción

Un amplificador de frecuencia intermedia o FI (figura 279) es esencialmente un amplificador de RF o radiofrecuencia de banda estrecha o sintonizado, que amplifica una banda muy limitada de frecuencias alrededor de una misma frecuencia central. En nuestro caso, esta frecuencia central es de 455 KHz, la FI del proceso heterodino AM, y el ancho de banda no excede de 9 KHz.

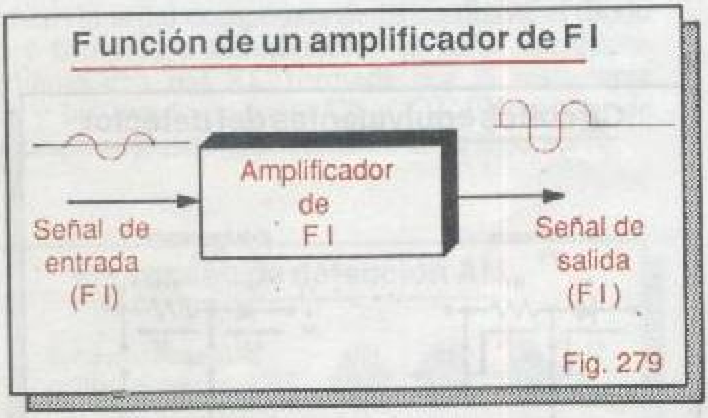
Función de un amplificador de FI

Fig. 279

Idealmente, un amplificador de FI para AM debería tener una respuesta de frecuencia como la que se muestra en la figura 280A. En la práctica, la respuesta de frecuencia es solo una aproximación a este modelo ideal (figura 280B).

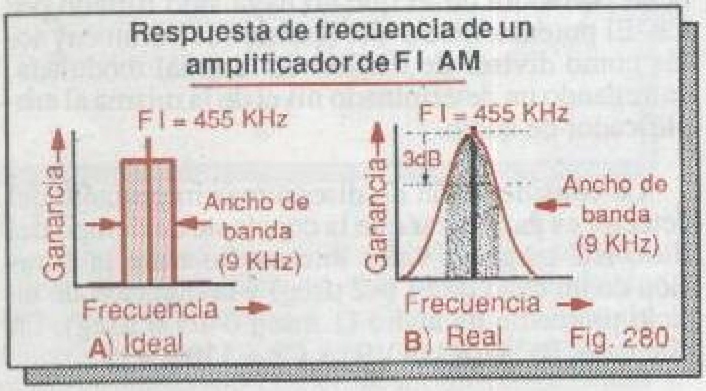
Respuesta de frecuencia de un amplificador de FI AM

A) Ideal
B) Real
Fig. 280

El nombre de FI se debe al hecho de que amplifica frecuencias intermedias o sea frecuencias que están en un punto intermedio entre la frecuencia de la portadora original de RF y la frecuencia de la señal de modulación.

Puesto que los amplificadores de FI son amplificadores sintonizados de RF, vamos a estudiar a continuación la teoría general de estos últimos. Posteriormente, se analizarán los amplificadores de FI del radio AM CEKIT, como una aplicación práctica de esta teoría.

Amplificadores sintonizados de RF

Los amplificadores de RF se utilizan para amplificar señales que tienen frecuencias por encima de 30 KHz. Son similares en muchos aspectos a los amplificadores de audio pero se caracterizan por ser selectivos, esto es, se diseñan para amplificar una banda estrecha de frecuencias alrededor de una frecuencia central para la cual están sintonizados.

En general, los amplificadores sintonizados de RF se componen de dos partes: el amplificador propiamente dicho y la red de sintonía.

La sintonización la proveen generalmente circuitos resonantes RLC como los estudiados en la Lección 12. En algunos casos especiales, se utilizan cristales de cuarzo para conseguir el mismo efecto y proveer mayor estabilidad. La amplificación de las señales de RF como tales, se realiza mediante transistores, FETs, etc.

En la figura 281A se muestra el diagrama simplificado de un amplificador de RF que utiliza un circuito sintonizado RLC paralelo como carga. A la frecuencia de resonancia, el circuito presenta su máxima impedancia posible y, en consecuencia, bajo esta circunstancia, la ganancia de voltaje del amplificador es máxima (figura 281B).

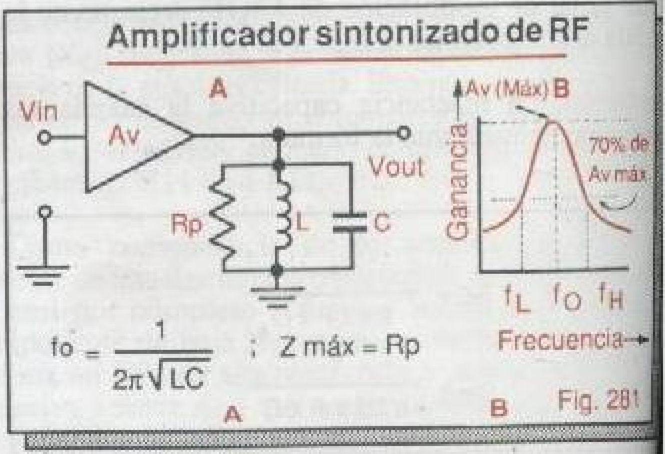
Amplificador sintonizado de RF

A
B
Fig. 281

Dependiendo de su aplicación, los amplificadores sintonizados pueden tener un circuito de sintonía a la entrada, a la salida, o tanto a la entrada como a la salida. Estos circuitos son los que determinan la frecuencia central y el ancho de banda del amplificador.

Los amplificadores de FI del radio AM CEKIT son sintonizados a la salida. La frecuencia central es de 455 KHz y el ancho de banda es de 9 KHz aproximadamente. Otras características importantes de los amplificadores sintonizados son su ganancia y su factor de ruido.

La ganancia de los amplificadores de FI es variable porque depende del nivel de señal captado a la entrada: las señales débiles necesitan más amplificación que las señales fuertes. La cantidad de amplificación requerida en cada caso la gobierna el control automático de ganancia (CAG), circuito que estudiaremos más adelante.

Los amplificadores de FI del radio AM CEKIT

Como se explicó en la Lección 21, a la salida del conversor de RF de nuestro receptor obtenemos una señal modulada de FI de 455 KHz. Esta señal lleva la misma información fundamental de audio (modulación) de la portadora original, pero es demasiado débil como para excitar el detector o demodulador.

El propósito de las etapas amplificadoras de frecuencia intermedia es reforzar la amplitud de la señal suministrada por el mezclador y entregarla con el nivel apropiado a la entrada del detector de audio. La señal que ingresa a los amplificadores de FI es del orden de los milivoltios y la que sale es de algunos voltios.

La amplificación de la señal modulada de FI se realiza en dos etapas acopladas magnéticamente, llamadas primera etapa de FI y segunda etapa de FI, respectivamente. En la figura 282 se muestra el circuito correspondiente.

El término *acoplamiento magnético* se refiere al hecho de que la transferencia de señales entre el mezclador y la primera etapa de FI, así como entre la primera y la segunda y entre ésta y el detector, se realiza a través de transformadores de acoplamiento.

Los transformadores de acoplamiento utilizados por las etapas amplificadoras de FI se han llamado T1, T2 y T3. El transformador T1, que se identifica por su núcleo amarillo, transfiere la señal del conversor a la primera etapa de FI. El transformador T2, con núcleo blanco, transfiere la señal de la primera etapa de FI a la segunda. El transformador T3, con núcleo negro, transfiere la señal de la segunda etapa de FI al detector.

Etapas de F I del radio AM Cekit

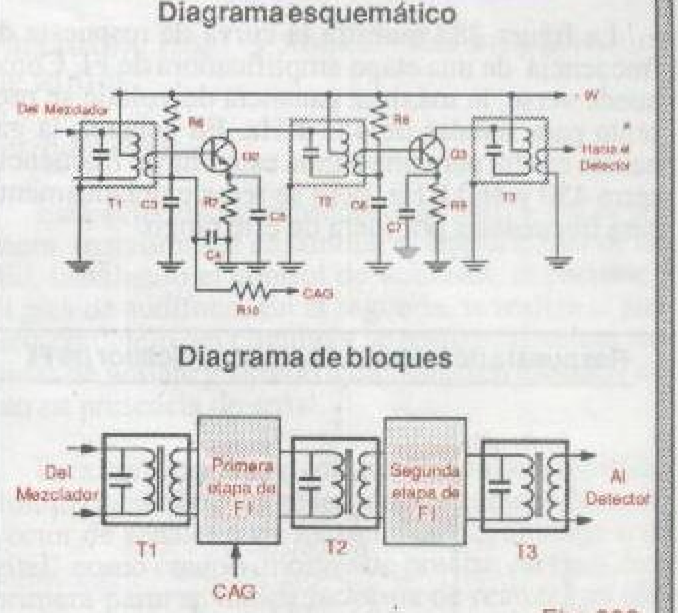

Diagrama esquemático

Diagrama de bloques

Fig. 282

Además de cumplir estas funciones básicas, los transformadores de acoplamiento anteriores se caracterizan también por ser *sintonizados*. Todos poseen un condensador interno en paralelo con la bobina primaria, configurando un circuito de sintonía LC que tiene una determinada frecuencia de resonancia.

La frecuencia de resonancia de estos circuitos se ajusta variando la posición del núcleo dentro del transformador. De este modo, se altera la inductancia L del tanque resonante. En nuestro caso, debemos calibrar los transformadores T1, T2 y T3 para que presenten una frecuencia de resonancia de 455 KHz, o sea igual al valor de la FI.

Bajo estas condiciones, la ganancia de voltaje de cada una de las etapas amplificadoras de FI será máxima para las señales que tengan una frecuencia igual a 455 KHz, y, muy baja para las señales de otras frecuencias.

Esto es así, porque un circuito resonante paralelo ofrece la máxima respuesta (máxima impedancia) a señales incidentes con frecuencias iguales a su frecuencia natural de resonancia, y atenúa señales de otras frecuencias.

Por consiguiente, si cada transformador está sintonizado a la FI, cada uno presentará muy alta impedancia a las señales de 455 KHz, y permitirá que las mismas desarrollen en los extremos de la bobina primaria una gran caída de voltaje. Entre mayor sea el voltaje desarrollado en el primario, mayor será el voltaje inducido en el secundario.

La figura 283 muestra la curva de respuesta de frecuencia de una etapa amplificadora de FI. Como puede verse, la máxima ganancia de voltaje se presenta para señales de 455 KHz. En general, la ganancia es alta para una banda estrecha de frecuencia entre 450 y 460 KHz, pero se reduce drásticamente para frecuencias por fuera de este rango.

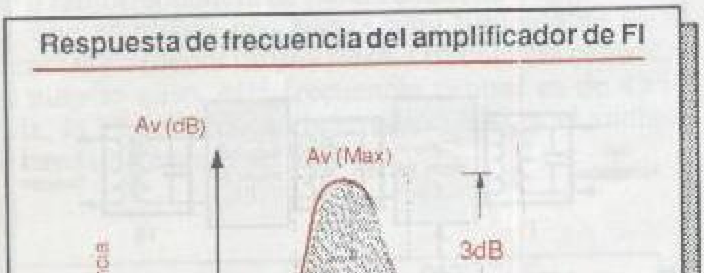

Fig. 283

Otra función importante de los transformadores de FI T1, T2 y T3, es la de acoplar la alta impedancia de colector de la etapa precedente con la baja impedancia de base de la etapa siguiente. El correcto acoplamiento de impedancias entre etapas es imprescindible para garantizar la máxima transferencia de señal.

En nuestro caso, se hace necesario acoplar entre sí las diferentes etapas, porque los transistores trabajan en la configuración emisor común y, en consecuencia, tienen una impedancia de salida relativamente alta, del orden de 50 KΩ, y una impedancia de entrada relativamente baja, del orden de 1 KΩ.

Operación de las etapas amplificadoras de FI

Como consecuencia del batido de frecuencia de las señales producidas por el oscilador local y la suministrada por el amplificador de RF, en el mezclador se produce toda una gama de señales, incluyendo las que tienen las mismas frecuencias originales, las que tienen una frecuencia igual a la suma de ambas y las que tienen una frecuencia igual a su diferencia.

Sin embargo, el circuito de salida del mezclador, constituido por el transformador T1, está sintonizado para responder a la señal de batido que tiene una frecuencia igual a la diferencia entre las frecuencias local y de antena. Esta frecuencia es precisamente la frecuencia intermedia o FI del proceso (455 KHz).

En consecuencia, en el secundario de T1 se induce únicamente la señal de FI. Esta señal, a pesar de tener una frecuencia diferente a la de la portadora original, conserva su misma modulación. Sin embargo, no es lo suficientemente fuerte como para ser detectada. Por consiguiente, debe ser amplificada.

Consideremos nuevamente el circuito de la figura 282. La débil señal de FI presente en el secundario de T1 se aplica a la base de Q2. Este transistor la amplifica y la envía a su circuito sintonizado LC de salida, formado por el primario de T2 y el condensador interno.

Este circuito, sintonizado a 455 KHz, actúa como carga del primer amplificador de FI. Su impedancia es máxima a esta frecuencia, lo mismo la ganancia de tensión de la etapa y el voltaje de la señal desarrollada en el primario de T1.

La señal proveniente de la primera etapa de FI se induce en el secundario de T2 y se aplica a la base de Q3, el segundo amplificador de FI. Nuevamente, este transistor amplifica la señal de entrada y la envía al transformador T3, el cual está también sintonizado a 455 KHz y ofrece la máxima respuesta a esta frecuencia.

La señal de salida de la segunda etapa de FI se induce en el secundario de T3 y se aplica al cátodo del diodo detector D1.

Polarización de los transistores de FI

El voltaje directo de polarización de la base de Q2 lo establece el divisor resistivo formado por R6 y R10. En particular, R6 está conectado al positivo de la fuente de alimentación y R10 a la salida del detector, actuando como control automático de ganancia (CAG). El CAG se estudia en detalle en otra lección de este curso.

El efecto de la resistencia del CAG, R10, es permitir que la ganancia de voltaje del transistor Q2 varíe en relación con la potencia de la señal de RF sintonizada, de tal modo que sea grande para señales débiles y se reduzca para señales fuertes.

De este modo, todas las señales captadas se escuchan al mismo volumen.

El voltaje inverso de polarización del colector de Q2 lo toma este transistor del positivo de la fuente a través de la línea central del primario de T2. Con se ha dicho, la carga de este transistor es el circuito sintonizado de 455 KHz formado por la bobina primaria de T2 y su condensador interno.

El transistor Q2 amplifica la señal de FI presente en la base y la envía a la bobina primaria de T2.

La señal presente en esta última se induce en el secundario de T2 y se aplica a la base de Q3, el segundo amplificador de FI.

El voltaje directo de polarización de la base de Q3 se obtiene del positivo de la fuente a través de R8.

El condensador C6 es un capacitor de derivación, que permite el paso de la corriente de señal mientras mantiene la polarización de corriente continua.

Por esta razón, la señal inducida de FI se desarrolla entre la base de Q3 y tierra.

La resistencia R9 entre el emisor de Q3 y tierra, tiene por finalidad limitar la corriente de colector de este transistor, a un valor seguro.

Además, contribuye a la estabilización de la polarización de la juntura base-emisor del mismo.

La corriente de colector produce una caída de voltaje a través de R9. El voltaje de polarización de la juntura BE de Q3, es la diferencia entre el voltaje de la base y la caída desarrollada a través de R9.

El condensador C7 es un capacitor de derivación que se comporta como un cortocircuito para las variaciones de señal, provenientes de la base de Q3.

De este modo, evita que el voltaje de polarización de la juntura BE siga las fluctuaciones de la señal de FI y se afecte el punto de trabajo del amplificador.

La carga del transistor Q3 es el circuito sintonizado formado por la bobina primaria de T3 y su capacitor interno.

Por lo tanto, Q3 provee máxima amplificación de la señal de FI presente en su base y la envía al primario de T3.

Esta señal modulada y amplificada de 455 KHz se induce en el secundario de T3 y se inyecta al cátodo del diodo detector D1.

En el análisis de los circuitos correspondientes a las etapas amplificadoras de FI, mediremos la ganancia de voltaje de cada etapa y los diferentes voltajes de polarización de cada transistor.

Actividad práctica Nº 13

Ensamble final y análisis del amplificador de audio

Introducción

Esta actividad comprende dos partes. En la primera, se termina de ensamblar el amplificador de audio, instalando el control de volumen, el parlante y el jack de audífono. En la segunda, se realiza el análisis completo del circuito y se realizan diversas medidas de voltaje y de corriente, tanto en ausencia como en presencia de señal.

Para verificar la correcta operación del amplificador, necesitará de un generador de audio o de un inyector de señales y de un multímetro, análogo o digital, como equipo mínimo de prueba. Al final de la primera parte se indica la forma de realizar un chequeo rápido y de localizar posibles fallas, utilizando estos instrumentos.

Durante el análisis se profundiza en estos conceptos y se realiza el seguimiento y medición de las diferentes condiciones de polarización y formas de onda presentes en el proceso, desde la entrada del preamplificador hasta la salida del amplificador de potencia. Para este estudio, es conveniente disponer también de un osciloscopio.

Terminando el ensamble del amplificador de audio

En esta actividad, vamos a conectar los componentes necesarios para terminar completamente la etapa amplificadora de audio.

Después de hacerlo y verificarlo, vamos a realizar una prueba y análisis de su funcionamiento en donde comprobaremos en la práctica, la teoría estudiada sobre este importante tema.

Conectaremos primero el potenciómetro de 10 KΩ que se utiliza como control de volumen y que está marcado en el diagrama como RV 10K y que se suelda con sus cinco terminales a la izquierda, sobre las puntillas 25, 26, A2 , A3 y A4 como se muestra en la figura A24.

Paso 1. Instalación y conexión del potenciómetro. Coloque cuidadosamente el potenciómetro, haciendo coincidir lo mejor posible sus terminales sobre las puntillas mencionadas, y vaya soldando cada uno de ellos hasta completar la instalación.

Para un mejor resultado, estañe ligeramente los terminales del potenciómetro por debajo, antes de instalarlo sobre las puntillas.Si se suelta alguno de

Instalación del control de volumen

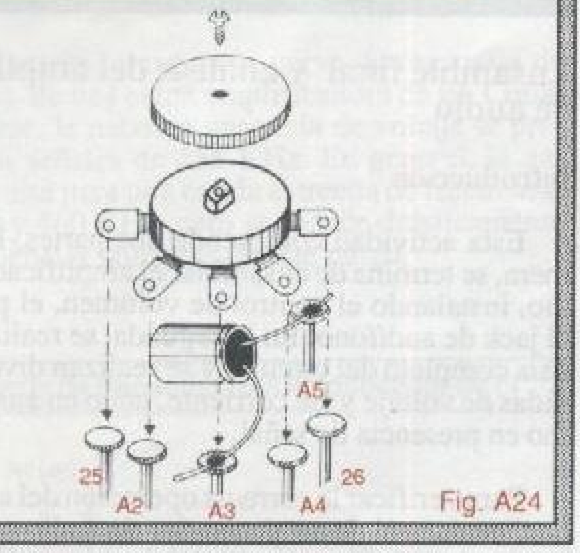

Fig. A24

los cables que estaban conectados, vuélvalos a instalar en su sitio.

Coloque ahora la perilla en el eje del potenciómetro y asegúrela con el tornillo central, haciendo coincidir la ranura especial en el eje.

Paso 2. Instalación y conexión del conector para el audífono. Conectaremos ahora el " jack " o conector para la salida de audífono, marcado en los diagramas como E/J.

Lo primero que se debe hacer es fijar bien con pegante, el conector para audífono sobre la superficie de la lámina. Antes de pegarlo, doble el terminal inferior izquierdo, completamente hacia abajo, de tal manera que al instalarlo, este terminal toque la puntilla A30.

Después de pegar el conector, suelde directamente sus terminales a las puntillas indicadas. Limpie cuidadosamente y caliente bien las dos superficies con el fin de que las soldaduras queden bien hechas.

Conecte ahora un cable entre el terminal inferior derecho del conector a la puntilla A28, fijándose que esta conexión no haga corto con el terminal inferior izquierdo.

Paso 3. Instalación y conexión del parlante. Antes de instalar el parlante suelde dos cables de 15 cms en su terminales. Fije con el pegante, el parlante en la parte superior derecha de la tabla del radio.

Conecte los cables de éste a las puntillas A30 y A31, uniendo con el mismo cable, esta última con la puntilla A29 según se muestra en la "Guía para la instalación de componentes". Así se termina el ensamble de la etapa "Amplificador de audio".

Una vez terminado el ensamble del amplificador de audio y antes de conectar la batería de 9 voltios, verifique la correcta instalación de cada uno de los componentes y los alambres de conexión, en cada una de las puntillas correspondientes a esta etapa.

Para hacer esta revisión, siga el diagrama esquemático que está pegado en la tabla y la "Guía para la instalación de componentes", que se ha entregado con este curso.

Revise especialmente los transistores, los diodos y los condensadores electrolíticos, ya que estos componentes son polarizados, y si se conectan mal se pueden dañar y no funciona esta etapa.

Prueba rápida del amplificador de audio

Una vez que se ha terminado el ensamble de esta etapa, debemos hacer una prueba de su correcto funcionamiento. Para hacerlo, se requiere de una fuente de señal que puede ser un generador de audio o un inyector de señales como el que se ensambló en la actividad Nº 11 de este curso.

Pasos para la prueba

Paso 1. Conecte la batería de 9 voltios al conector para la batería, B1.

Paso 2. Encienda el radio y deje la perilla en el punto inicial después de accionar el suiche. En este punto, al contrario de los controles de volumen tradicionales, éste se encuentra en la posición de máximo volumen. Si se gira completamente a la derecha, el volumen estará en el mínimo.

Paso 3. Conecte el terminal de tierra del generador de audio, si lo tiene, o del inyector de señal, a la línea negativa del radio, en la puntilla 18. El generador de audio debe estar ajustado para una salida de 1000 Hz.

Paso 4. Conecte la salida del generador de audio o del inyector de señales a la entrada de la etapa del amplificador de audio, en la puntilla A1. En este momento, si todo está bien, se debe escuchar un tono en el parlante.

Gire la perilla del control de volumen lentamente hacia la derecha. Si el tono rebaja de intensidad, la etapa amplificadora de audio está trabajando bien. Si no es así realice los siguientes chequeos en el circuito:

Mirando componentes defectuosos

Coloque su multímetro en el modo de voltios en CC, en la gama de 10 voltios.

Instale la batería y encienda el radio. Conecte la punta negra (negativo) de su multímetro a la línea negativa de tierra (puntilla 22). A continuación, toque con la punta roja (positivo) los siguientes puntos:

Puntilla 37: Colector de Q5. Deberá medir 9 voltios, aproximadamente. Si no es así, chequee la batería y el interruptor del potenciómetro VR.

Puntilla A10: Base de Q5. Deberá medir 5.6 voltios, aproximadamente. Si no es así, chequee el parlante, las conexiones del audífono y la resistencia R13.

Puntilla A12: Colector de Q4. Deberá medir 4.2 voltios, aproximadamente. Si no es así, chequee los diodos D2 y D3, el condensador C12 y el transistor Q4.

Puntilla A20: Punto de prueba N. Deberá medir 4.9 voltios, aproximadamente. Si no es así, chequee los transistores Q5 y Q6 y los resistores R14 y R15.

Análisis y prueba completa del amplificador de audio

El objetivo de esta actividad es analizar la operación general del amplificador de audio del radio AM CEKIT, desde el punto de vista de las corrientes y voltajes existentes en el circuito, tanto en condiciones estáticas como dinámicas.

Las condiciones estáticas o de reposo son las presentes en el circuito cuando no existe señal de entrada.

El efecto de la señal de entrada es variar las corrientes y voltajes del circuito por encima y por debajo de sus valores de reposo. A las nuevas condiciones se les denomina dinámicas.

Para realizar el análisis estático, necesitará únicamente de un multímetro, análogo o digital. Para realizar el análisis dinámico, requerirá también de un osciloscopio y un generador de funciones o generador de audio.

Si no posee este equipo adicional, las pruebas estáticas serán suficientes para determinar las condiciones de operación de su amplificador y diagnosticar posibles fallas.

Equipo de prueba necesario

Un multímetro SANWA TX-301 o similar.
Un generador de funciones B&K 3010 o similar.
Un osciloscopio LEADER LBO-514A o similar.

Análisis estático

1. Consumo de corriente del amplificador de audio

El objetivo de esta prueba es medir la corriente consumida por el amplificador de audio en condiciones de reposo, es decir, sin señal de entrada. Si la corriente medida es muy alta, lo más probable es que exista un corto en alguna parte del circuito. Una corriente muy baja evidencia la existencia de un circuito abierto.

Procedimiento

Paso 1. Sitúe el multímetro en el rango "30 mA DC" y conéctelo al circuito, como se muestra en la figura A25.

Medición del consumo estático de corriente

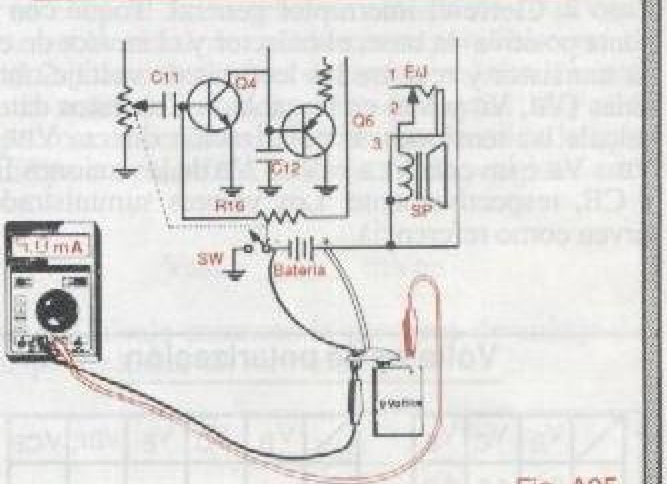

Fig. A25

Paso 2. Cierre el interruptor general (SW) y registre la lectura del medidor. Nosotros obtuvimos 5 mA.

ITQ = ______ mA

2. Voltajes de polarización

El objetivo de esta prueba es medir los voltajes de polarización (V_B, V_C, V_E, etc.) de los transistores Q4, Q5 y Q6. Verificaremos la polarización directa de las uniones BE y la polarización inversa de las uniones BC.

Procedimiento

Paso 1. Sitúe el multímetro en el rango "12 VDC" y conéctelo al circuito como se muestra en la figura A26. Coloque la punta negativa a tierra y utilice la punta positiva para recorrer los puntos de interés.

Medición de los voltajes de polarización

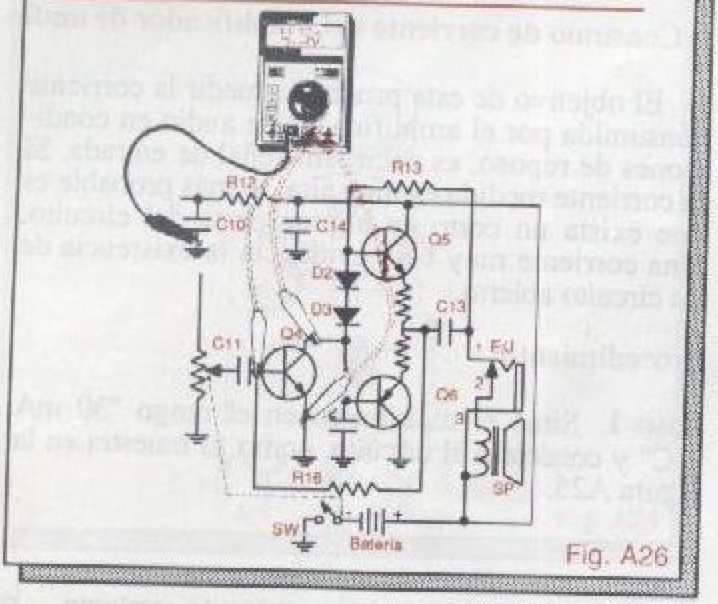

Fig. A26

Paso 2. Cierre el interruptor general. Toque con la punta positiva la base, el colector y el emisor de cada transistor y registre las lecturas de voltaje obtenidas (V_B, V_C y V_E) en la tabla 1. Con estos datos, calcule las tensiones de polarización directa $V_{BE} = V_B - V_E$ e inversa $V_{CB} = V_C - V_B$ de las uniones BE y CB, respectivamente. Los valores suministrados sirven como referencia.

Voltajes de polarización

	V_B	V_C	V_E
Q4	0.7	4.5	0.0
Q5	5.9	9.0	5.2
Q6	4.5	0.0	5.2

	V_B	V_C	V_E	V_{BE}	V_{CB}
Q4					
Q5					
Q6					

Nota: Valores en Voltios (V)

Tabla 1

Observe que, por tratarse de un transistor PNP, las tensiones de polarización V_{BE} y V_{CB} resultantes sobre Q6 deben ser negativas. Para Q4 y Q5 estos voltajes de polarización deben ser positivos, por tratarse de transistores NPN. El valor de V_{BE} para transistores de silicio está generalmente comprendido entre 0.6V y 0.7 V.

3. Corrientes de polarización

El objetivo de esta prueba es medir las corrientes de polarización (I_B e I_C) de los transistores Q4, Q5 y Q6 y determinar la ganancia estática de corriente ($h_{FE}=I_C/I_B$) de cada etapa. Las corrientes y voltajes de polarización definen el punto de trabajo y la clase de operación.

Procedimiento

Paso 1. Para medir la corriente de base (I_B) de cada transistor, despegue la base, sitúe el multímetro en el rango "60 µA DC" y conéctelo como se muestra en la figura A27. Al final de cada medición, suelde nuevamente la base a la puntilla correspondiente.

Corrientes de polarización

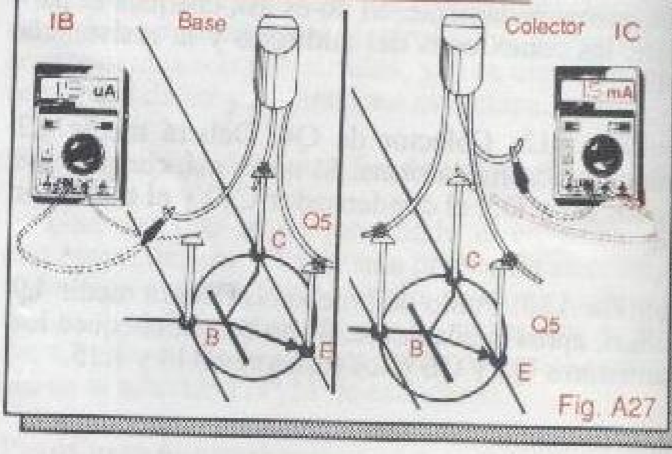

Fig. A27

Elija correctamente la polaridad de las puntas de prueba. En el caso de Q4 ó Q5, que son NPN, la punta positiva debe ir a la puntilla A5 ó A15 y la negativa a la base del transistor. En el caso de Q6, que es PNP, la punta positiva debe ir a la base y la negativa a la puntilla A16.

Consigne sus resultados en la tabla 2, columna "I_B". Los valores suministrados sirven como referencia.

Corrientes de polarización

	I_B (µA)	I_C (mA)
Q4	25	4.5
Q5	1.5	9.0
Q6	1.5	1.5

	I_B (µA)	I_C (mA)	hfe
Q4			
Q5			
Q6			

Tabla 2

Paso 2. Para medir la corriente de colector (I_C) de cada transistor, despegue el colector, sitúe el multímetro en el rango "6 mA DC" o similar, y conéctelo como se muestra en la figura A27B. Al final de cada medición, suelde nuevamente el colector a la puntilla respectiva.

Elija correctamente la polaridad de las puntas de prueba. En el caso de Q4 ó Q5, que son NPN, la punta positiva debe ir a la puntilla A8 ó A18 y la negativa al colector. En el caso de Q6, que es PNP, la punta positiva debe ir al colector y la negativa a la puntilla A22.

Consigne sus resultados en la tabla 2, columna "I_C". Los valores suministrados sirven como referencia.

Paso 3. Con los datos obtenidos, calcule la ganancia de corriente (h_{FE}), de cada transistor utilizando la fórmula $h_{FE} = I_C/I_B$. Consigne sus resultados en la tabla 2, columna "h_{FE}".

4. Caídas de voltaje a través de los diodos de polarización.

El objetivo de esta prueba es medir la caída de voltaje a través de los diodos de polarización D2 y D3 del amplificador de audio. Estos diodos proporcionan estabilidad térmica a la etapa de salida.

Procedimiento

Paso 1. Sitúe el multímetro en el rango "3 V DC". Para medir la caída de voltaje en cada diodo, conecte la punta positiva al ánodo y la negativa al cátodo. Para medir la caída total, conecte la punta positiva al ánodo de D2 y la negativa al cátodo de D1.

Paso 2. Cierre el interruptor general. Mida inicialmente la caída de voltaje sobre D2, a continuación la caída sobre D3 y, por último, la caída total entre D2 y D3. Registre sus resultados. Deberá obtener, aproximadamente, 0.65 V sobre cada diodo y 1.3V entre ambos.

V_{D2} = ______ V

V_{D3} = ______ V

V_{D2-3} = ______ V

Análisis dinámico

1. Ganancia de voltaje de la etapa preamplificadora

El objetivo de esta prueba es evaluar la ganancia de voltaje del preamplificador de audio y observar otras características interesantes de esta etapa. En particular, comprobaremos que la señal de salida está desfasada 180º con respecto a la señal de entrada, y que la ganancia depende estrechamente de la frecuencia.

Utilizaremos el generador de funciones como fuente de señal y el osciloscopio de dos canales como instrumento de visualización y medida. Necesitaremos también de un condensador de acople de 0.1 µF.

Procedimiento

Paso 1. Ajuste el generador para producir una onda seno de 100 mVpp de amplitud y 1 KHz de frecuencia. Conéctelo a la entrada del amplificador de audio (punto de prueba L) a través de un condensador de 0.1 µF como se muestra en la figura A28.

Paso 2. Cierre el interruptor general y gire el control de volumen completamente hacia la derecha. Conecte el canal 1 del osciloscopio a la base de Q4 y el canal 2 al colector, como se indica. Como referencia, ajuste la base de tiempo en 0.5 mseg/div, la atenuación del canal 1 en 5 mV/div y la del canal 2 en 200 mV/div.

Paso 3. Ajuste el control de volumen hasta visualizar en el canal 1 una señal de entrada de 10 mVpp. Observe la señal de salida correspondiente en el canal 2. Compare la fase de ambas señales. Notará que la señal de salida aparece desfasada 180º con respecto a la señal de entrada. Ahora mida la amplitud pico a pico de la señal de salida y registre este valor.

Vsal = ______ mVpp

Paso 4. Calcule entonces la ganancia de voltaje de la etapa mediante las fórmulas:

$$A_V = \frac{V_{sal}}{V_{ent}}$$

y

$$A_V\ (\mathrm{dB}) = 20 \log A_V$$

Consigne los resultados obtenidos. Nosotros registramos una ganancia de 136 (43 dB).

A_V = ______ = ______ dB

Recuerde que esta es la ganancia del preamplificador a 1 KHz. En realidad, la ganancia y, en menor grado, la diferencia de fase, dependen de la frecuencia. Evalúe la ganancia para otras frecuencias de la gama de audio (20Hz a 20 KHz). Los valores obtenidos configuran la respuesta de frecuencia del preamplificador.

2. Ganancia de potencia del amplificador de salida

El objetivo de esta prueba es determinar la ganancia de potencia del amplificador de salida para un determinado nivel de señal de entrada. Evaluaremos también la máxima potencia de salida posible.

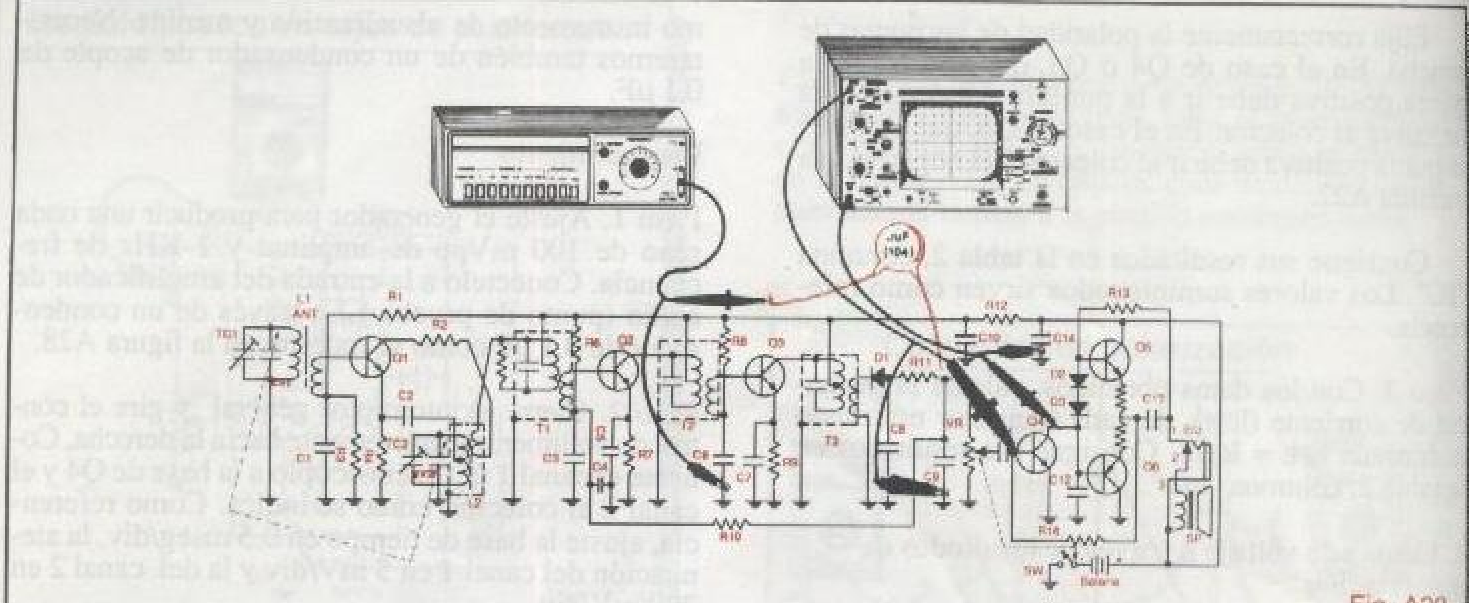

Fig. A28

Procedimiento

Paso 1. Aplique con el generador de funciones una señal de 100 mVpp, 1 KHz, al punto de prueba L. Utilice un condensador de desacople de 0.1 μF. Conecte entonces el canal 1 del osciloscopio a la base de Q4 y manipule el control de volumen hasta obtener una señal de 10 mVpp en ese punto.

Paso 2. Para determinar la potencia de entrada, debe medir el voltaje de entrada. Para efectos prácticos, la impedancia de entrada del amplificador puede considerarse igual a R13 (680Ω). La potencia se evalúa entonces utilizando la fórmula V^2/R, siendo V el voltaje rms de entrada y R=680Ω.

Para medir el voltaje rms de entrada, sitúe el multímetro en el rango " 3 VAC ". Conecte la punta negativa a tierra y toque con la punta positiva el colector de Q4. Utilice la salida "OUT" para desacoplar el nivel de CC de la señal. Registre el valor medido. Nosotros obtuvimos 1.2 Vrms.

Vent = ______ Vrms

Calcule entonces la potencia de entrada y registre el valor obtenido. En nuestro caso, la potencia de entrada es igual a $1.2^2/680 = 2$ mW.

Pent = ______ mW

Paso 3. Para determinar la potencia de salida, mida el voltaje de salida. La carga del amplificador es un parlante que tiene una impedancia de 8Ω. La potencia se evalúa entonces utilizando la fórmula V^2/R, siendo V el voltaje rms de salida y R = 8Ω .

Para medir el voltaje de salida, calibre el multímetro de la misma forma anterior pero toque con la punta positiva el punto de prueba O (puntilla A26). Registre el valor medido. Notará que no hay amplificación de voltaje. Nosotros obtuvimos 1.1 Vrms.

Vsal = ______ Vrms

Calcule entonces la potencia de salida y registre el valor obtenido. En nuestro caso, la potencia de salida es igual a $1.1^2/8 = 151$mW = 0.15 W.

Psal = ______ mW

Paso 4. Evalúe la ganancia de potencia utilizando las fórmulas:

$$AP = \frac{Psal}{Pent}$$

y

$$AP\ (dB) = 10 \text{ Log } AP$$

En nuestro caso, AP = 151/2 = 75 = 19 dB. Registre el valor obtenido por Ud.

AP = ______ = ______ dB

Paso 5. Para determinar la máxima potencia de salida, gire lentamente el control de volumen hasta obtener la máxima lectura de voltaje en el punto de prueba O. Calcule entonces la potencia mediante la fórmula V^2/R, siendo V el voltaje rms máximo de salida y R la impedancia del parlante (8Ω). Nosotros obtuvimos Vmáx=3.4 Vrms. En consecuencia, Pmáx = 1.5 W. Registre los valores obtenidos.

Vsal (máx) = ______ Vrms

Psal (máx) = ______ W

Continuación actividad práctica Nº 13

3. Operación del par complementario Q5-Q6

El objetivo de esta prueba es observar las formas de onda del proceso push-pull realizado en la etapa de potencia del amplificador de audio. Analizaremos también el fenómeno de la distorsión de amplitud y los efectos de la realimentación negativa.

Procedimiento

Paso 1. Retire la resistencia de 10 Ω utilizada como carga en las pruebas anteriores y conecte en su lugar el audífono (o el parlante, si lo posee). Conecte al punto de prueba L el generador de audio y el canal 1 del osciloscopio.

Calibre la atenuación vertical en 0.1 V/div y la base de tiempo en 0.5 mseg/div. Programe el generador para suministrar en el punto L una onda seno de 1 KHz y 300 mVpp de amplitud. Observe y mida la señal correspondiente en el osciloscopio.

Paso 2. Conecte ahora el canal 1 del osciloscopio al punto de prueba J. Calibre la atenuación vertical en 0.01 V/div. Con la batería nueva instalada, manipule lentamente el control de volumen (VR) hasta obtener en ese punto una señal de 10 mVpp. Observe y mida la señal correspondiente en el osciloscopio. Deberá escuchar un tono continuo en el audífono.

Paso 3. Desconecte la resistencia R16 de la puntilla A17 y conecte el extremo libre al polo positivo de la batería (puntilla 32). Notará que el tono escuchado en el audífono sube de volumen y observará en el osciloscopio que la señal de entrada aplicada al punto J aumenta ligeramente de amplitud. Conecte nuevamente R16 a la puntilla A17.

Al desconectar R16 de la puntilla A17 y conectarla a la puntilla 32 estamos eliminando la realimentación negativa que existe entre la salida y la entrada del amplificador de audio. El aumento en el nivel de la señal de entrada se interpreta como un aumento en la impedancia de entrada del amplificador.

Por consiguiente, hemos comprobado que *la realimentación negativa reduce la impedancia de entrada de un amplificador*.

Paso 4. Conecte ahora el canal 2 del osciloscopio a la puntilla A12 (salida del preamplificador y entrada del amplificador de potencia). Calibre la atenuación vertical en 1 V/div. Deberá observar una onda seno de cierta amplitud. Si la compara con la señal de entrada, notará que las dos señales están en oposición de fase, es decir, cuando la una es positiva, la otra es negativa y viceversa.(figura A29).

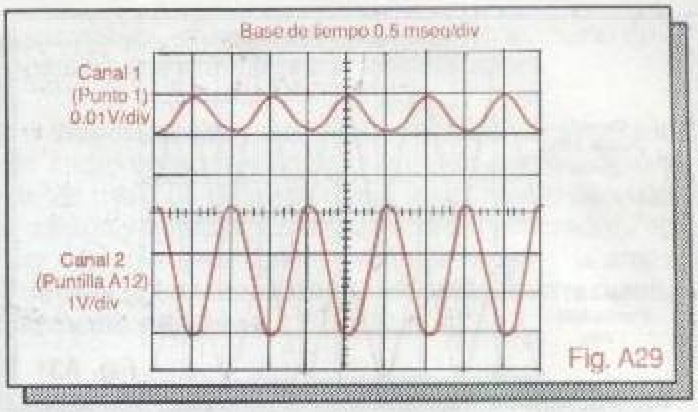

Fig. A29

Paso 5. Desconecte nuevamente R16 de la puntilla A17 y conéctela a la puntilla 32 (positivo de la batería). Observe la señal resultante en el osciloscopio y compárela con la señal original. Notará que aumenta de amplitud,pero experimenta una distorsión considerable (figura A30). Conecte otra vez R16 a la puntilla A17.

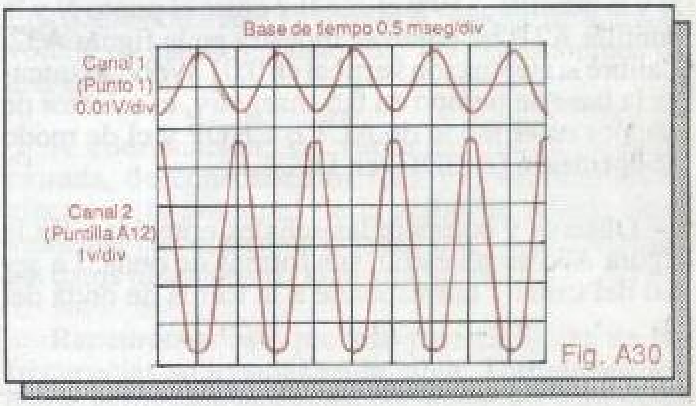

Fig. A30

El aumento en el nivel de la señal de salida, y el consiguiente aumento de la ganancia, al suspender la realimentación negativa, es una consecuencia directa del aumento de la impedancia de entrada observada en el paso 3. Sin embargo, la suspensión de la realimentación introduce distorsión.

Por consiguiente, hemos comprobado que *la realimentación negativa reduce la ganancia pero también reduce la distorsión en la señal de salida de un amplificador*.

Paso 6. Conecte ahora el canal 2 del osciloscopio a la puntilla A26 y el canal 1 a la puntilla A12. Calibre la atenuación vertical de ambos canales en 1 V/div. Deberá observar una onda seno de cierta amplitud. Si la compara con la señal de entrada aplicada a la puntilla A12, notará que las dos señales están en fase y tienen amplitudes similares (figura A31).

Por consiguiente, hemos comprobado que *la ganancia de voltaje del amplificador de potencia es*

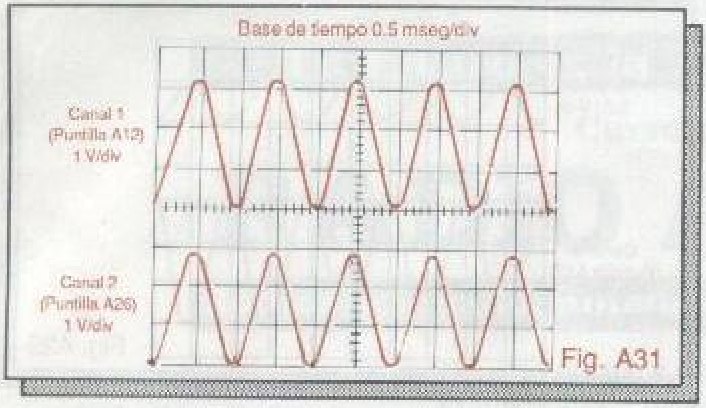

Fig. A31

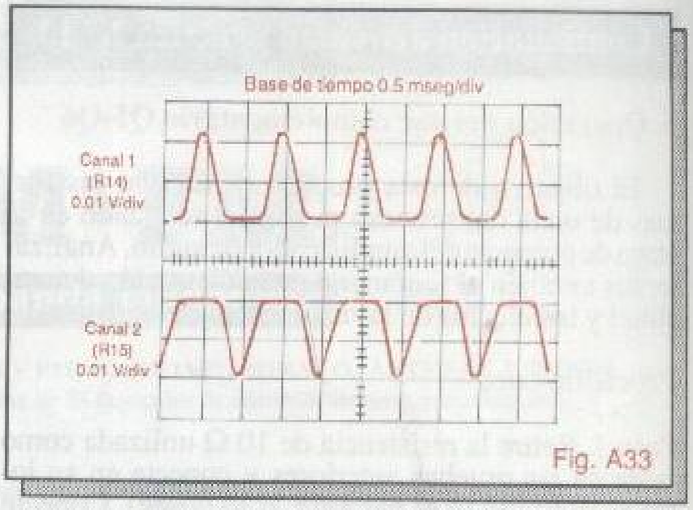

Fig. A33

aproximadamente igual a la unidad. Es decir, *el circuito no amplifica voltaje sino corriente.*

Paso 7. Repita el paso 5, con la atenuación vertical de ambos canales en 2V/div. Notará otra vez distorsión en la señal de salida. Recuerde reconectar R16.

Paso 8. Desconecte las puntas de prueba de ambos canales. Conecte el canal 1 entre el punto de prueba N y la puntilla A19 y el canal 2 entre el punto N y la puntilla A21, tal como se muestra en la figura A32. Calibre la atenuación vertical en 0.01 V/div. Mantenga la base de tiempo en 0.5 mseg/div, el selector de canales en el modo de ALT o CHOP y el de modo de operación (AC/DC) en DC.

Observe y compare las señales obtenidas. En la figura A33 se muestran sus formas de onda. La señal del canal 1 corresponde a la forma de onda del voltaje sobre R14, la resistencia de emisor de Q5, y la señal del canal 2 a la forma de onda del voltaje sobre R15, la resistencia de emisor de Q6.

Las formas de onda de las señales de corriente de emisor de Q5 y Q6 son idénticas a las de las señales de voltaje visualizadas porque, como sabemos, en una resistencia, el voltaje y la corriente están en fase. Los valores correspondientes de corriente se obtienen, de acuerdo a la Ley de Ohm, dividiendo los valores de voltaje en cada punto por el valor de R14 o R15 (0.5 Ω).

Observe en la figura A33 que Q5 suministra corriente de salida únicamente durante los semiciclos positivos de la señal de entrada y Q6 únicamente du-

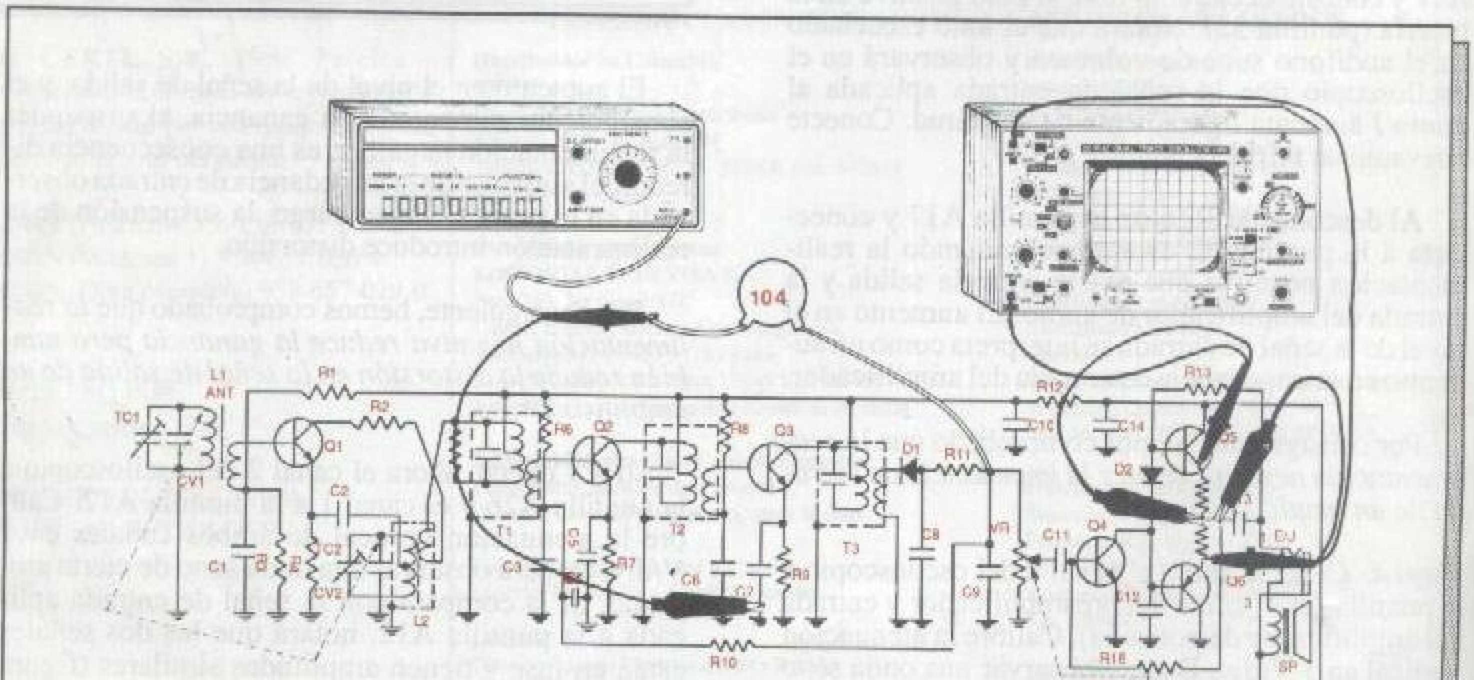

Fig. A32

rante los negativos. Estos impulsos de corriente se superponen en el punto N y reconstruyen en su totalidad la forma de onda de la corriente de salida aplicada a la carga, representada por el audífono.

Por consiguiente, hemos comprobado que *Q5 y Q6 efectivamente trabajan como un par complementario*. Es decir, *Q5 maneja la porción positiva de la señal de entrada del amplificador de potencia y Q6 la negativa* .

Para ratificarlo aún más, superponga los niveles de referencia de las señales observadas en el osciloscopio, manipulando los controles de posición vertical. Obtendrá una onda perfectamente sinusoidal, libre de distorsión, tal como se muestra en la figura A34.

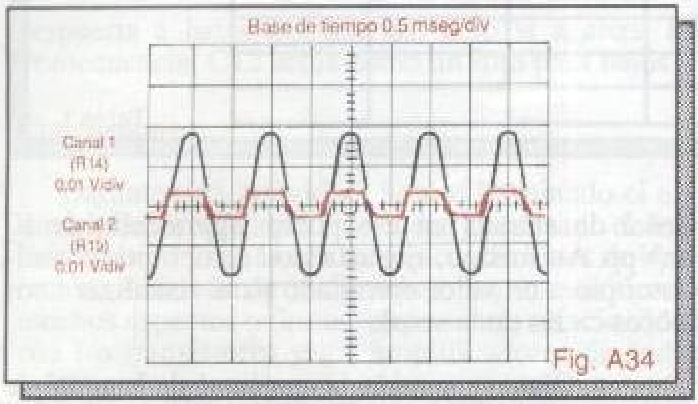

Fig. A34

4. Respuesta completa de frecuencia

El objetivo de esta prueba es determinar la respuesta completa de frecuencia del amplificador de audio, es decir, medir la ganancia de voltaje del circuito para distintos valores de frecuencias.

En particular, comprobaremos que la ganancia de voltaje depende estrechamente de la frecuencia y, lo más importante, que no es constante para todas las frecuencias posibles de señal.

Nos interesa, especialmente, lo que sucede entre 0 y 5 KHz, que es el ancho de banda ocupado por la información enviada desde una emisora AM. Los amplificadores de audio, en general, se diseñan para amplificar señales entre 0 y 20 KHz.

Realizaremos dos pruebas paralelas. La primera en condiciones normales de operación y la segunda desconectando el condensador C12. Esta última se hace con el fin de verificar la función de filtro de altas frecuencias que cumple C12.

El condensador C12 delimita el ancho de banda del circuito, evitando que amplifique también señales de alta frecuencia. Estas señales, además de ser indeseables, provocan un aumento innecesario de la potencia consumida por el amplificador.

Necesitaremos para esta práctica un generador de audio como fuente de señal y un osciloscopio de doble trazo (2 canales) como instrumento de visualización y medida. La frecuencia del generador debe ser variable dentro de un amplio rango y la amplitud de la señal suministrada debe mantenerse razonablemente estable.

Continuaremos utilizando el audífono como carga. El procedimiento general se detalla enseguida.

Mediante el generador, aplicaremos a la entrada del amplificador una onda seno de determinada frecuencia y la visualizaremos en uno de los canales del osciloscopio. En el otro canal, observaremos la señal de salida correspondiente.

Con el condensador C12 instalado, mediremos la amplitud de las dos señales y calcularemos la ganancia de voltaje "con C12" para esa frecuencia. Los datos obtenidos (frecuencia, voltaje de entrada, voltaje de salida y ganancia) los registraremos en una tabla.

A continuación, sin variar las condiciones de entrada, desconectaremos C12 y observaremos el efecto en la amplitud de la señal de salida. Nuevamente, calcularemos la ganancia y registraremos los datos obtenidos en la misma tabla.

Repetiremos este proceso para cada una de las frecuencias solicitadas en la tabla. Tomaremos los datos de frecuencia y ganancia obtenidos y los localizaremos en una gráfica en papel logarítmico.

Los valores de frecuencia se marcan horizontalmente y los de ganancia verticalmente. Cada valor de frecuencia y ganancia determina un punto sobre la gráfica. Uniendo entre sí estos puntos, obtendremos las curvas de respuesta de frecuencia buscadas, con C12 y sin C12, y procederemos a su comparación y análisis.

Procedimiento

Paso 1. Conecte el generador de audio al punto de prueba L a través de un condensador de 0.1 µF. Conecte a ese punto el canal 1 del osciloscopio. Calibre la atenuación vertical en 0.1 V/div y la base de tiempo en 0.5 mseg/div. Programe el generador para suministrar en el punto L una onda seno de 1 KHz y 300 mVpp de amplitud. Observe y mida la señal correspondiente en el osciloscopio.

Paso 2. Conecte ahora el canal 1 del osciloscopio al punto de prueba J. Calibre la atenuación vertical en 0.01 V/div. Con la batería instalada, manipule len-

Tabla de datos para respuesta de frecuencia

		Con C12					
Frecuencia	Entrada (Punto J)	Salida (Punto O)	Av	Av (dB)	Salida (Punto O)	Av	Av (dB)
1 K Hz	10 mVpp	760 mVpp	76	38 dB	640 mVpp	64	36 dB
200 Hz							
400 Hz							
600 Hz							
800 Hz							
1 KHz							
2 KHz							
3 KHz							
4 KHz							
5 KHz							
10 K Hz							
15 K Hz							
20 K Hz							
25 K Hz							
50 KHz							
100 K Hz							

Tabla 1

tamente el control de volumen (VR) hasta obtener en ese punto una señal de 10 mVpp. Observe y mida la señal correspondiente en el osciloscopio. Deberá escuchar un tono continuo en el audífono.

Paso 3. Conecte el canal 2 del osciloscopio al punto de prueba O. Calibre la atenuación vertical en 0.5 V/div. Observe la señal de salida y mida su correspondiente amplitud pico a pico. Calcule la ganancia de voltaje utilizando las siguientes fórmulas:

$$Av\ (1KHz) = \frac{Vsal}{Vent} = \frac{\qquad}{10\ mVpp} = \underline{\qquad}$$

$$Av_{dB}\ (1KHz) = 20\ Log \frac{Vsal}{Vent} = \underline{\qquad}\ dB$$

Registre los valores obtenidos en la tabla 1.

Paso 4. Desconecte momentáneamente el condensador C12 de la puntilla 20. Observe y mida la señal de salida resultante. Calcule la ganancia de voltaje correspondiente y registre los valores obtenidos en la tabla 1.

Paso 5. Ajuste el control de frecuencia del generador de funciones para suministrar una señal de 200 Hz en el punto de prueba J. Si es necesario, manipule el control de amplitud para garantizar que la señal de entrada en ese punto siga siendo de 10 mVpp. Así mismo, ajuste la base de tiempo del osciloscopio a un valor apropiado para visualizar unos pocos ciclos de la señal.

Paso 6. Observe y mida la amplitud de la señal de salida resultante en el canal 2 del osciloscopio. Si es necesario, ajuste la atenuación vertical a un valor apropiado para visualizar completamente la señal. Calcule la ganancia de voltaje correspondiente y registre los valores obtenidos en la tabla 1.

Paso 7. Repita el paso 4.

Paso 8. Repita los pasos 5, 6 y 7 para los demás valores de frecuencia solicitados en la tabla 1, hasta completarla.

Paso 9. Con los datos obtenidos, dibuje en papel logarítmico o semilogarítmico la curva de respuesta de frecuencia del amplificador, con C12 y sin C12.

Utilice el trazado de la figura A35. Localice horizontalmente todos los valores de frecuencia y verticalmente todos los valores de ganancia. Una los puntos resultantes en forma consecutiva. La gráfica resultante es la curva de respuesta de frecuencia buscada.

Notará que la ganancia no es constante para todas las frecuencias y que alcanza su valor máximo dentro de cierto rango, digamos, entre 1 y 2 KHz, y se reduce progresivamente para frecuencias por fuera de este rango. La ausencia de C12 no afecta la

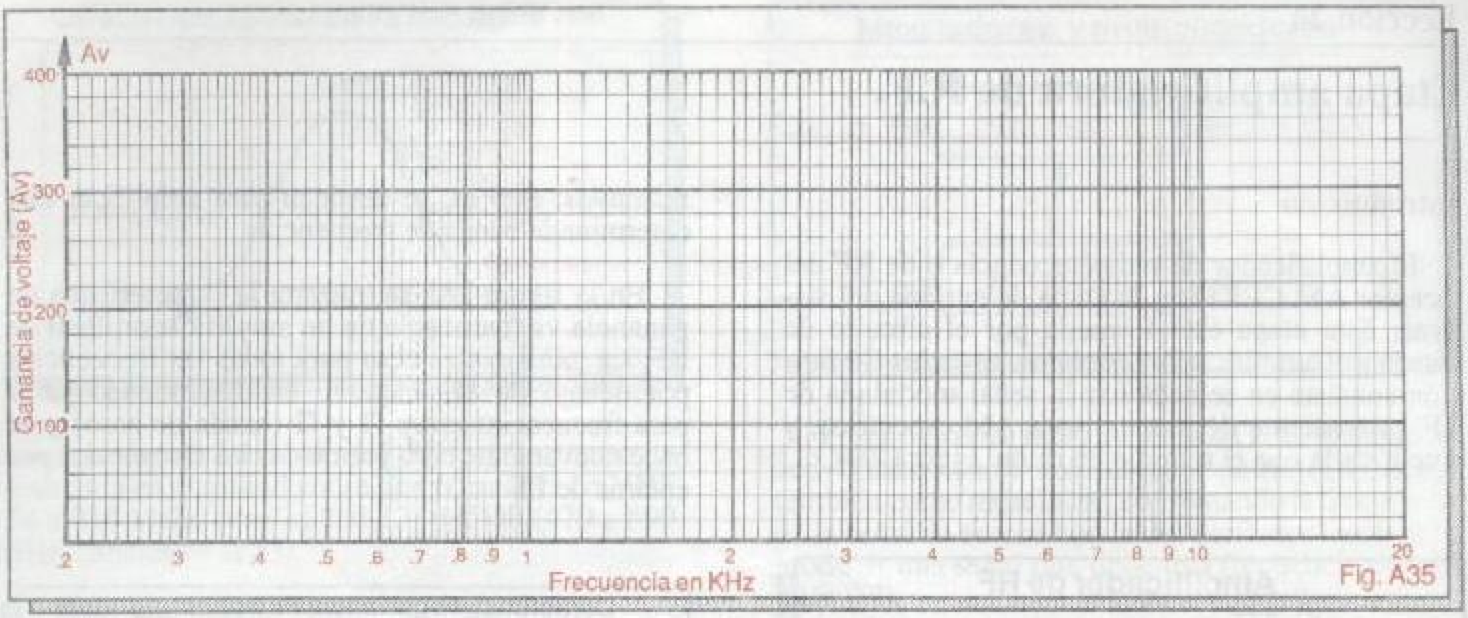

Fig. A35

respuesta a bajas frecuencias, pero sí a altas. En consecuencia, C12 actúa como un filto pasa bajos.

Conclusiones

Durante esta actividad, hemos terminado el ensamble del amplificador de audio y realizado el análisis estático y dinámico de sus condiciones de operación. A través de las pruebas, hemos verificado muchos aspectos teóricos importantes relacionados con los transistores y los amplificadores de audio en general.

Así mismo, nos hemos familiarizado más con el uso de los transistores y la manipulación de instrumentos de medida como el multímetro, el osciloscopio y el generador de señales. También, hemos adquirido ciertas destrezas y técnicas que nos servirán en el futuro para analizar circuitos similares.

Desde el punto de vista conceptual, hemos comprobado, entre otras cosas, que:

- El consumo de corriente del amplificador de audio es relativamente bajo en condiciones estáticas o de reposo (sin señal de entrada) y aumenta en condiciones dinámicas, dependiendo del nivel de señal aplicado.

- Todos los transistores del amplificador de audio deben estar correctamente polarizados para que el circuito funcione eficientemente. En particular, el voltaje de polarización directa de la unión BE de Q4 y Q5 debe ser positivo y el de Q6 negativo.

- Q4 es un amplificador de voltaje clase A mientras Q5 y Q6 son amplificadores de potencia complementarios clase AB. Q5 y Q6 suministran más ganancia de corriente que Q4 pero no amplifican voltaje.

- La señal de salida del preamplificador está desfasada 180º con respecto a su señal de entrada. La señal de salida del amplificador de potencia está en fase con su señal de entrada.

- Para una determinada frecuencia de señal, la ganancia de voltaje del preamplificador es prácticamente constante mientras la ganancia de potencia del amplificador de salida, depende del nivel de señal aplicado. En consecuencia, el preamplificador opera linealmente y el amplificador de potencia no.

- La máxima potencia de salida del amplificador de audio depende del valor de la fuente de alimentación y del valor de impedancia de la carga. La máxima potencia sin distorsión depende, además, del nivel de la señal de entrada.

- La resistencia R16 suministra realimentación negativa de voltaje desde la salida hasta la entrada del amplificador de audio. La realimentación negativa reduce la impedancia de entrada y la ganancia del amplificador, pero minimiza la distorsión.

- Q5 suministra corriente de salida sin distorsión durante los semiciclos positivos de la señal de entrada y Q6 durante los semiciclos negativos. En consecuencia, trabajan como un par complementario clase AB.

- Para un mismo nivel de señal de entrada, la ganancia de potencia del amplificador de audio depende estrechamente de la frecuencia. Es máxima para las frecuencias centrales de la banda de audio y disminuye gradualmente para frecuencias por fuera de esta banda.

- La ausencia de C12 no afecta la respuesta de frecuencia del amplificador de audio, a bajas frecuencias, pero sí a altas frecuencias.

Etapa amplificadora de R.F.

Introducción

El amplificador de radiofrecuencia o de RF del receptor AM CEKIT es la etapa de entrada del sistema. Esta etapa está formada por el circuito de antena y el amplificador propiamente dicho. Su función consiste en seleccionar la señal modulada de RF, proveniente de una emisora AM específica, y amplificarla con el mínimo de ruido. Figura 284.

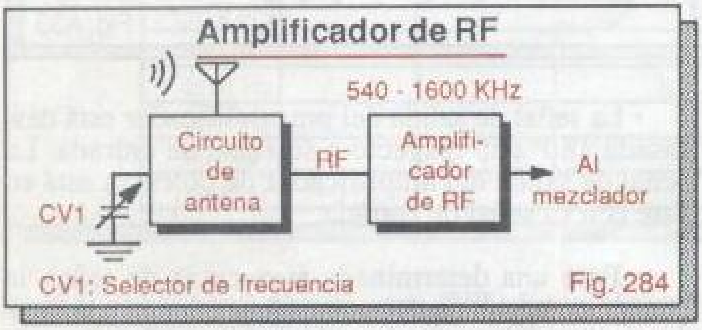

Fig. 284

El amplificador de RF debe tener un ancho de banda apropiado para seleccionar la portadora y sus bandas laterales y rechazar al mismo tiempo las señales provenientes de emisoras adyacentes. En el caso del radio AM CEKIT, este ancho de banda es de 9 KHz, aproximadamente.

Una vez amplificada, la señal de salida del amplificador de RF se mezcla o bate con la señal del oscilador local para producir la señal de FI. A este proceso, como sabemos, se le denomina heterodinación. La señal de FI tiene una frecuencia de 455 KHz y conserva la misma modulación original de la portadora de RF.

Nuestro amplificador de RF ha sido diseñado para operar en la banda de onda media (OM o MW) de AM, la cual se extiende desde 530 hasta 1600 KHz. Los amplificadores de RF, en general, se diseñan para operar en cualquier banda de frecuencias desde 30 KHz hasta más de 1 GHz (1000 MHz), incluyendo microondas. Las *microondas* son frecuencias por encima de 3 GHz.

Características generales de los amplificadores de RF

Los amplificadores de RF son similares a los amplificadores de audio en muchos aspectos. Deben tener una carga en su circuito de salida y necesitan estar correctamente polarizados. Sin embargo, son *selectivos*, esto es, se destinan para amplificar una determinada banda de frecuencias.

En la figura 285 se muestra la característica de ganancia vs frecuencia de un amplificador ideal de RF. La ganancia es cero para todas las frecuencias por debajo de **f1**, adquiere un determinado valor para frecuencias entre **f1** y **f2** (banda de paso) y se hace nuevamente cero para todas las frecuencias por encima de **f2**.

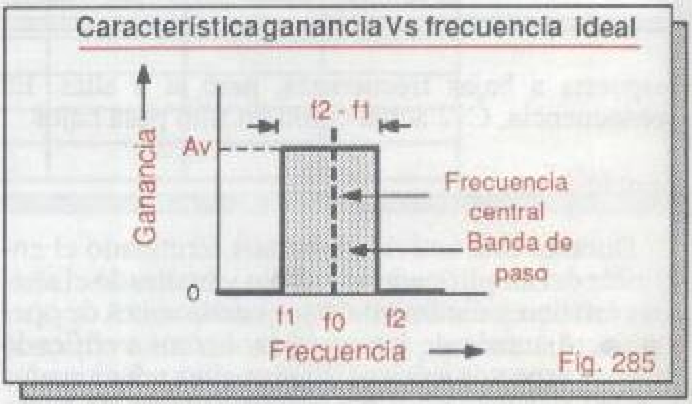

Fig. 285

Las frecuencias que definen la banda de paso (f2 y f1) se denominan *frecuencias de corte* superior e inferior, respectivamente, y su diferencia (f2-f1) define el ancho de banda (**B**) del amplificador. El punto medio **fo** de la banda de paso, equidistante de los puntos de corte, se denomina *frecuencia central.*

La característica de ganancia vs frecuencia de los amplificadores reales de RF (figura 286) es sólo una aproximación a la característica ideal de la figura 285. Por definición, los límites reales que definen la banda de paso son aquellos en los cuales la ganancia es 0.707 veces la ganancia máxima, obtenida a la frecuencia central.

De acuerdo a su ancho de banda, los amplificadores de RF pueden ser de banda ancha o de banda estrecha. Los de banda ancha tienen anchos de banda de varios MHz y se destinan, generalmente, para aplicaciones de video. Los de banda estrecha tienen anchos de banda de unos pocos Hz y se destinan, generalmente, para aplicaciones de audio.

Un ejemplo de amplificador de RF de banda ancha es el utilizado como etapa de entrada en un receptor de TV (figura 287). La señal de video ocu-

Característica ganancia Vs frecuencia real

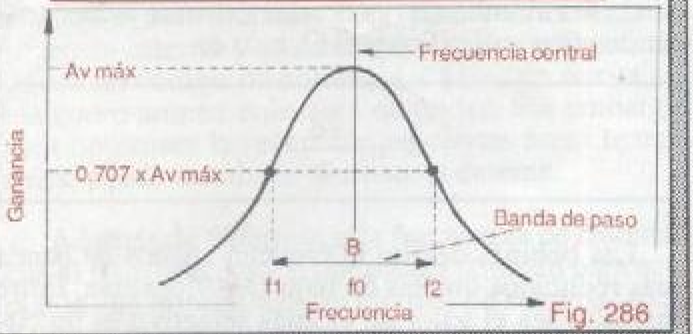

Fig. 286

pa un ancho de banda de 5.5 MHz y se transmite modulada en amplitud sobre una portadora de VHF (54 a 216 MHz, canales 2 al 13) o UHF (470 a 890 MHz, canales 14 al 83).

Circuitos de entrada de un TV

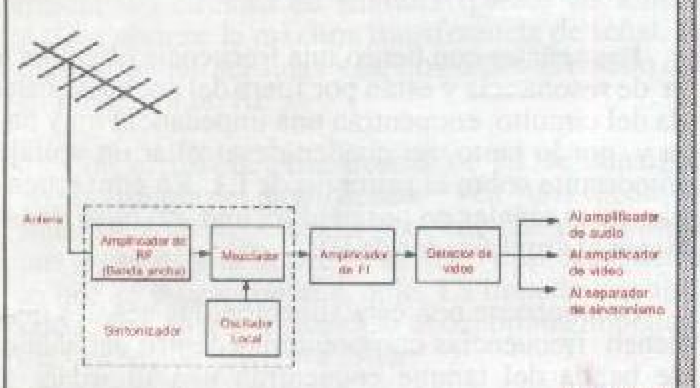

Fig. 287

Un ejemplo de amplificador de RF de banda estrecha o sintonizado es el que se utiliza como etapa de entrada en un receptor superheterodino de AM. En este caso, la señal de audio ocupa un ancho de banda de 9 KHz y se transmite modulada en amplitud sobre una portadora de 530 a 1600 KHz.

Los amplificadores de RF pueden encontrarse como etapas de entrada en receptores o como etapas de salida en transmisores. En el primer caso, trabajan como amplificadores de voltaje, manejando señales muy débiles. En el segundo, trabajan como amplificadores de potencia, manejando señales desde unos pocos vatios hasta cientos de kilovatios.

Los amplificadores de RF se utilizan también para realizar ciertas funciones especializadas, como la mezcla y la multiplicación de frecuencias (figura 288). Un mezclador combina dos señales de frecuencias diferentes. Un multiplicador entrega una frecuencia de salida que es 2, 3 ó 4 veces la frecuencia de entrada.

Mezcladores y multiplicadores

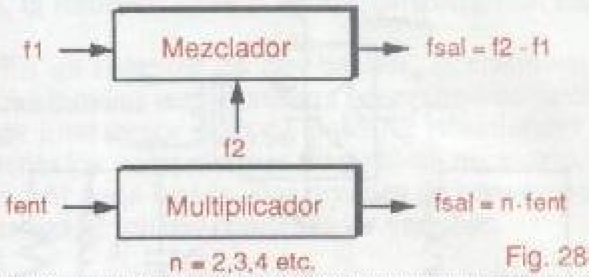

Fig. 288

En el caso del radio AM CEKIT, el dispositivo activo usado como amplificador de RF (Q1), actúa también como mezclador, combinando la frecuencia de la señal de antena con la del oscilador local, para producir una señal que tiene una frecuencia de 455 KHz (FI).

Los amplificadores de RF pueden estar formados por una o varias etapas. En este último caso, se conectan dos o más etapas sencillas en cascada, una a continuación de la otra, con el fin de aumentar gradualmente el nivel de la señal de entrada. El amplificador de RF del radio AM CEKIT es de una sola etapa y el de FI es de dos etapas.

Los amplificadores de RF se caracterizan por utilizar circuitos resonantes RLC como elementos de sintonía. Estos circuitos determinan el ancho de banda y la selectividad. La amplificación como tal la realizan transistores o FETs. Los FETs generan menos ruido y tienen una impedancia de entrada más alta que los transistores convencionales.

El amplificador de RF del radio AM CEKIT utiliza un tanque RLC, sintonizable a cualquier frecuencia entre 530 y 1600 KHz, como etapa de entrada y un tanque RLC, sintonizado a 455 KHz, como etapa de salida. El elemento activo de amplificación es un transistor bipolar NPN (Q1).

La etapa de antena. Concepto de factor de calidad o Q.

La etapa de antena del receptor AM CEKIT es un tanque o circuito resonante RLC paralelo, sintonizable a cualquier frecuencia entre 530 y 1600 KHz. Su función consiste en captar la señal de RF, emitida por una estación de radio AM específica dentro de esta banda, y suministrarla a la entrada del amplificador de RF.

Una vez amplificada, esta señal se mezcla con la del oscilador local para producir la señal de FI. En la figura 289 se muestra el circuito completo de la etapa de antena y su forma simplificada. La teoría sobre circuitos resonantes se estudia con detalle en la Lección 12 de este curso.

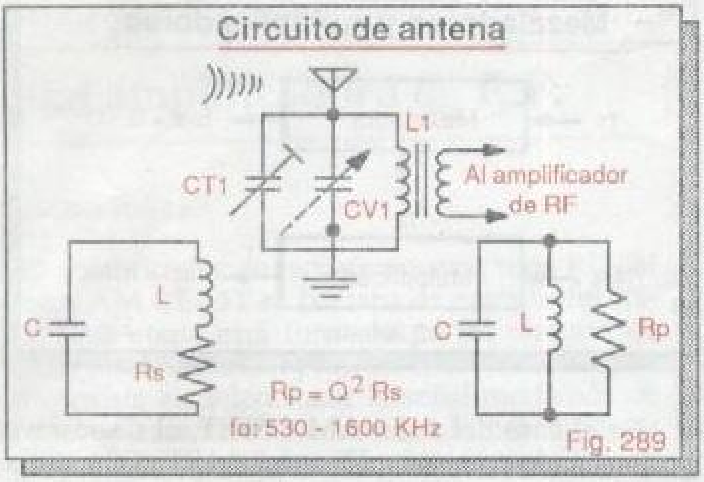

Fig. 289

La frecuencia de resonancia del circuito de antena está dada por:

$$Fant = \frac{1}{2\pi\sqrt{LC}}$$

Siendo 2π=6.28 una constante, L=L1 la inductancia del devanado primario en Henrios (H) y C = CV1+CT1 la capacitancia equivalente en Faradios (F). CV1 es el condensador de sintonía y CT1 el de compensación o "trimmer". Se utiliza para realizar ajustes finos de sintonía.

La resistencia RS representa las pérdidas asociadas a los elementos reactivos del circuito de sintonía (L y C) y al efecto de la impedancia de entrada del amplificador. Para efectos prácticos, se considera que las pérdidas se deben exclusivamente a la resistencia de la bobina L.

El efecto de estas pérdidas en la selectividad del circuito de sintonía y su ancho de banda, se puede evaluar mediante el *factor de calidad* o Q de la bobina, definido como:

$$Q = \frac{XL}{RS} = \frac{2\pi FoL}{RS}$$

Siendo 2π=6.28, FO la frecuencia de resonancia (Hz), L la inductancia de la bobina (H), XL la reactancia inductiva (Ω) y RS su resistencia serie o interna. Se dice que una bobina es de alto Q cuando el valor de Q es mayor o igual a 10 (diez). Es común encontrar, en aplicaciones de RF, bobinas con Q de 100 a 200 y más.

En términos del factor de calidad Q, el ancho de banda del circuito (B) es la relación entre la frecuencia de resonancia (Fo) y el Q, esto es:

$$B = \frac{Fo}{Q}$$

Las bobinas de alto Q proveen anchos de banda más reducidos que las de bajo Q y viceversa. Entre más alto sea el valor de Q, más selectivo es un circuito. El ancho de banda del circuito de antena debe ser compatible con el de las transmisiones comerciales de AM (10 KHz).

Las señales de radio, emitidas por las diversas estaciones, son interceptadas por la antena propiamente dicha y se dirigen al circuito de sintonía. Dependiendo de la posición de CV1, este circuito tendrá una determinada frecuencia de resonancia, digamos 750 KHz, y ofrecerá su máxima respuesta de voltaje a esa frecuencia.

Las señales que tienen una frecuencia diferente a la de resonancia y están por fuera del ancho de banda del circuito, encuentran una impedancia muy baja y, por lo tanto, no pueden desarrollar un voltaje importante sobre el primario de L1. En consecuencia, estas señales no pasan al secundario ni se transfieren al amplificador de RF.

En contraste con esta situación, las señales que tienen frecuencias comprendidas dentro del ancho de banda del tanque encuentran una impedancia muy alta y, por lo tanto, desarrollan voltajes apreciables sobre el primario de L1.

Estos voltajes, que son réplicas eléctricas de la señal electromagnética enviada por una emisora en particular, se inducen en el secundario y se transfieren al amplificador de RF.

La bobina o transformador del circuito de antena (L1) consiste de dos arrollamientos independientes, devanados sobre un núcleo común de ferrita y separados de éste por un cartón aislante (figura 290).

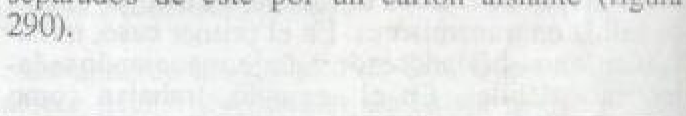

Fig. 290

Los arrollamientos se realizan con un tipo especial de alambre llamado "Litz" para minimizar su resistencia interna y sus extremos se identifican mediante un código de colores. La barra de ferrita actúa como antena colectora de ondas. Sin embargo, para optimizar la recepción en ciertas áreas marginales, puede instalarse una antena externa.

Además de determinar la frecuencia de resonancia, el transformador L1 se utiliza también para acoplar la alta impedancia del circuito de sintonía (≈150 KΩ) con la baja impedancia de entrada del amplificador de RF (≈1 KΩ).

Para lograr este acople, la relación de espiras (N) entre el secundario y el primario debe ser relativamente alta. En nuestro caso, el primario tiene 10 espiras y el secundario 120, aproximadamente. En consecuencia, la relación de espiras es N=12.

El acople consiste en igualar la impedancia de entrada del amplificador de RF (carga) con la impedancia del circuito de sintonía (fuente de señal). Así, se obtiene la máxima transferencia de señal, se minimizan las pérdidas y se simplifica el diseño del amplificador de RF.

Con el uso del transformador L1 se consigue que la entrada del amplificador " vea " una fuente de señal de baja impedancia y el circuito de sintonía una carga de alta impedancia. Esto es precisamente lo que se busca con el acople. La impedancia aparente vista en uno u otro caso se denomina impedancia " reflejada ". (figura 291).

Acoplamiento de impedancias

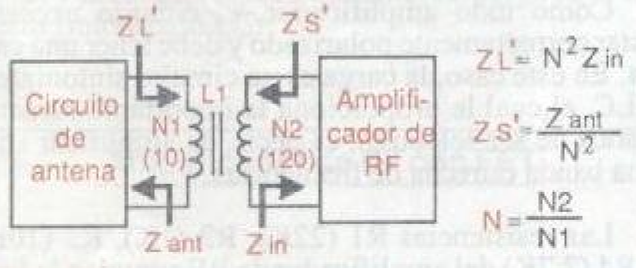

ZL', ZS' : Impedancias reflejadas

Fig. 291

Cómo se sintonizan otras bandas de AM

El circuito de sintonía del radio AM CEKIT provee únicamente recepción de las emisoras locales de onda media que transmiten en la banda comercial de AM (530 a 1600 KHz).

Muchos aparatos de radio captan también programas en onda larga (150 a 300 KHz) o en onda corta (5.5 a 30 MHz), emitidos por estaciones internacionales como Radio Monte Carlo, la BBC de Inglaterra, la Voz de los Estados Unidos de América, la Radiodifusora Nacional de Colombia, etc.

En un receptor de dos bandas, la conmutación de una banda a otra se realiza generalmente mediante un interruptor dpdt (2 polos, 2 posiciones) que conecta los componentes de sintonía necesarios para cubrir cada banda. Por razones de conveniencia, se utiliza el mismo condensador variable.

En la figura 292 se muestra un circuito de conmutación típico de onda larga/onda media utilizado en ciertos receptores de radio.

Circuito de conmutación de onda larga

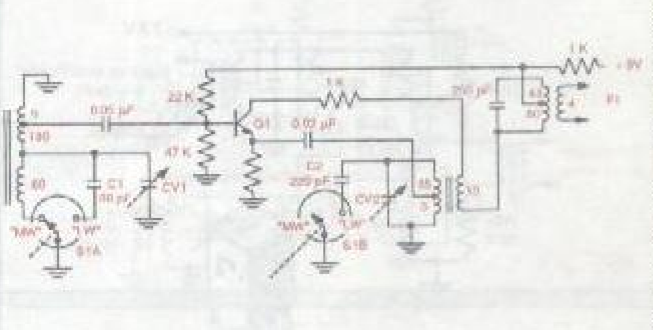

Fig. 292

En la posición "LW" (onda larga) el conmutador conecta el condensador C2 (220 pF) en paralelo con la sección osciladora del condensador variable (CV2) y el condensador C1 (88 pF) en paralelo con la sección de antena (CV1). No se utiliza el arrollamiento "60"de la bobina de sintonía y sólo interviene el arrollamiento "9/180". Estos números se refieren al número de espiras.

Con esta configuración la frecuencia de resonancia del circuito de antena se puede variar entre 150 y 300 KHz y la del oscilador local entre 605 y 755. Así, la frecuencia intermedia sigue siendo de 455 KHz.

En la posición "MW" (onda media o AM normal) el conmutador desconecta los condensadores C1 y C2 y conecta el arrollamiento "60" en paralelo con el arrollamiento "9/180".

Con esta configuración, la frecuencia de resonancia del circuito de antena se puede variar entre 540 y 1600 KHz y la del oscilador local entre 985y 2055 KHz. La frecuencia intermedia sigue siendo de 455 KHz. La operación del oscilador local se estudia en detalle en la Lección 25.

El amplificador de RF

El amplificador de RF propiamente dicho del radio AM CEKIT, recibe la débil señal captada por el circuito de antena, la amplifica, la mezcla con la señal del oscilador local y entrega la señal resultante al primer amplificador de FI. La salida del amplificador de RF es un tanque RLC, sintonizado a 455 KHZ.

En la figura 293 se muestra el circuito simplificado de nuestro amplificador de RF. Se han omitido los circuitos correspondientes a la etapa de antena y al oscilador local por comodidad. La forma como se realiza la mezcla de las señales de RF y del oscilador local se analiza en la Lección 26.

Circuito del amplificador de RF

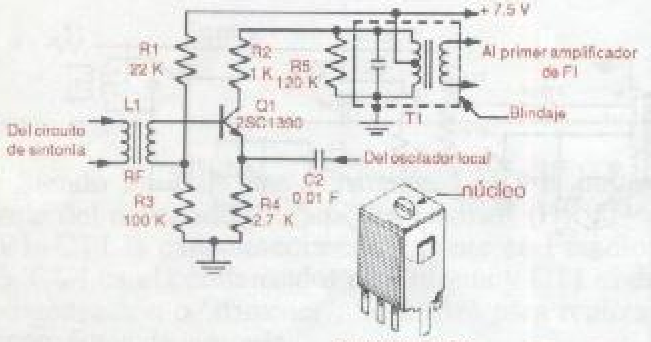

Fig. 293

El dispositivo activo del amplificador de RF es el transistor Q1. El circuito es básicamente un amplificador de voltaje, sintonizado a la salida, y responde a la configuración emisor común. Se acopla magnéticamente a la etapa siguiente mediante el primer amplificador de FI (T1, núcleo amarillo).

La tensión inversa de polarización del colector la extrae el circuito de la fuente de alimentación a través de la parte superior del devanado primario de T1. Este transformador es blindado magnéticamente para protegerlo de cualquier interferencia externa. El blindaje metálico se conecta a tierra.

El circuito sintonizado de salida lo constituyen R5 (120 KΩ), el primer transformador de FI (T1, núcleo amarillo) y el condensador interno de este último (≈500 pF). Este circuito actúa como carga del amplificador de RF y es el que determina su ganancia de voltaje.

La ganancia del amplificador de RF es máxima a 455 KHz porque, a esta frecuencia, la impedancia del tanque es también máxima y se puede desarrollar un alto voltaje de señal.

La señal de entrada del amplificador de RF proviene del secundario de la bobina de antena. Esta señal se amplifica a un nivel suficiente para minimizar el efecto del ruido y se mezcla con la del oscilador local.

La señal del oscilador local se aplica al emisor de Q1 a través del condensador C2 (0.01 μF). Esta señal tiene una amplitud relativamente alta, digamos 1 Vpp, y su frecuencia es siempre 455 KHz más alta que la captada en el circuito de antena.

Por ejemplo, si se sintoniza una emisora que transmite a 1000 KHz, el oscilador local generará automáticamente una señal de 1455 KHz. Para lograr este sincronismo, los condensadores variables del circuito de sintonía (CV1) y del oscilador local (CV2) están acoplados mecánicamente sobre el mismo eje, es decir, trabajan en "tandem".

Como resultado del proceso de mezcla, a la salida del amplificador de RF se tienen, además de las señales de 1000 KHz y 1455 KHz, la señal suma (2455 KHz), la señal diferencia (455 KHz) y todas las frecuencias *armónicas* o múltiplos (2000 KHz, 2910 KHz, 4910 Hz, 910 KHz, etc.).

Sin embargo, como la salida del amplificador de RF está sintonizada a 455 KHz, sólo la señal de esta frecuencia será amplificada. Todas las demás encontrarán una impedancia de salida muy baja y no recibirán amplificación. Naturalmente, se amplifica la señal de 455 KHz y sus bandas laterales, dentro de un ancho de banda de 9 KHz.

Repasemos brevemente la función de cada uno de los componentes que constituyen el amplificador de RF.

Como todo amplificador, el circuito necesita estar correctamente polarizado y debe tener una carga. En este caso, la carga es un circuito sintonizado RLC, el cual le proporciona la importante característica de ser selectivo, es decir, de amplificar sólo una banda estrecha de frecuencias.

Las resistencias R1 (22K), R2 (1K), R3 (10K) y R4 (2.7K) del amplificador de RF cumplen la función de polarizar adecuadamente el transistor Q1.

Específicamente, R1 y R3 establecen la corriente de reposo de la base (≈10 μA) y R2 el voltaje de reposo de colector (≈5.8 V). R4 estabiliza térmicamente el punto de trabajo, haciéndolo inmune a las variaciones de temperatura.

La resistencia R5 (120 KΩ) del circuito sintonizado de salida se utiliza para acomodar el ancho de banda a los requisitos del amplificador. Su efecto neto es reducir el Q del tanque sintonizado.

Como sabemos, al reducirse el Q aumenta el ancho de banda y disminuye la selectividad.

El transformador T1, sintonizado a 455 KHz, determina la ganancia del amplificador de RF. El circuito se sintoniza a 455 KHz, durante el proceso de alineación, desplazando el núcleo interno de ferrita hasta conseguir la máxima ganancia. Con esta operación se está variando, realmente, la inductancia del tanque.

Además de esta función, T1 acopla la alta impedancia de salida del amplificador de RF con la baja impedancia de entrada del primer amplificador de FI. Así mismo, T1 aplica la tensión de alimentación (≈+7.5 V) al colector de Q1. En este último caso se aprovecha la propiedad de una bobina de comportarse como un cortocircuito para la CC.

Amplificadores de RF con FET

Los transistores de efecto de campo (FETs y MOSFETs) se utilizan frecuentemente como amplificadores de RF, en sustitución de los transistores bipolares, por su bajo ruido y su alta impedancia de entrada.

Estas características mejoran la sensibilidad de la etapa de entrada del receptor, facilitando la captación de emisoras remotas, tales como las que transmiten en onda corta, y simplifican el acoplamiento entre etapas .

En la figura 294 se muestra un circuito básico de amplificador de RF con FET. Las resistencias R1, R2 y R3 y el condensador C3 establecen las condiciones de polarización. Los condensadores C1 y C2 proporcionan aislamiento de CC entre la fuente de señal y la carga, respectivamente. La carga es el circuito sintonizado LC.

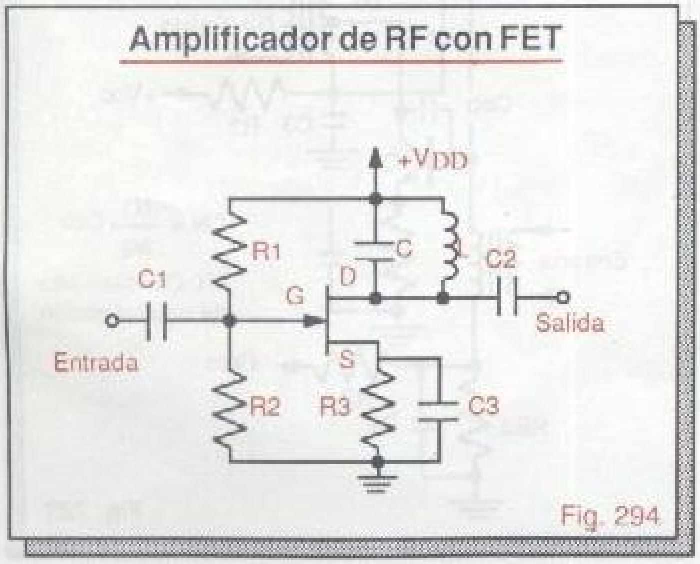

Fig. 294

Variando el valor de C, el circuito tanque se sintoniza a una determinada frecuencia de resonancia. A esta frecuencia, la impedancia de salida y la ganancia de voltaje son máximas. La señal amplificada de RF se acopla a la etapa siguiente a través del condensador C2.

En la figura 295 se muestra el circuito de un amplificador típico de RF con MOSFET de doble compuerta. Este tipo de amplificador se utiliza en receptores de FM que operan en la banda VHF (90 MHz a 110 MHz). La señal de entrada se aplica a la compuerta G1 y los voltajes de polarización a la compuerta G2.

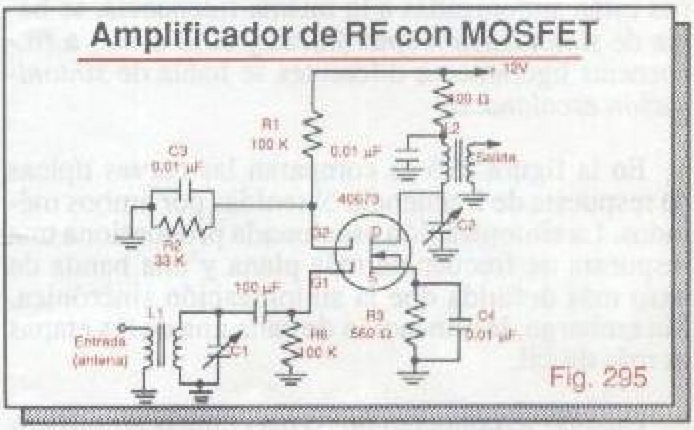

Fig. 295

La compuerta G2 actúa como una pantalla que neutraliza el efecto de la capacitancia parásita interna entre el drenador D y la compuerta G1. A frecuencias muy altas, esta capacitancia reduce la ganancia del amplificador y lo hace inestable, con tendencia a la oscilación. La importancia de la neutralización se analiza más adelante.

Note que tanto en la entrada como en la salida se utilizan circuitos sintonizados LC. El circuito de entrada está formado por L1 y C1 y el de salida por L2 y C2. Esta doble sintonía provee una alta selectividad. R1, R2 y R3 son resistencias de polarización. C3 y C4 son condensadores de desacople de RF.

Amplificadores de RF en cascada

Un amplificador de RF en cascada (figura 296) está formado por varias etapas individuales de amplificación conectadas en serie, es decir, una a continuación de la otra. Los amplificadores de RF en cascada se utilizan cuando la ganancia y la selectividad de una sola etapa son insuficientes para una aplicación dada.

Al conectar varios amplificadores de RF en cascada, la respuesta general de frecuencia depende de la sintonización de cada etapa. Si todas las eta-

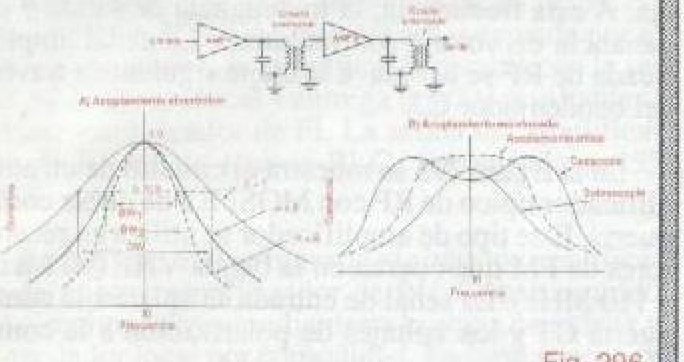

Amplificador de RF en cascada

Fig. 296

pas están sintonizadas a la misma frecuencia, se habla de *sintonización sincrónica*, y si lo están a frecuencias ligeramente diferentes se habla de *sintonización escalonada*.

En la figura 296 se comparan las curvas típicas de respuesta de frecuencia obtenidas por ambos métodos. La sintonización escalonada proporciona una respuesta de frecuencia más plana y una banda de paso más definida que la sintonización sincrónica. Sin embargo, la alineación de cada una de las etapas es más difícil.

Cuando se conectan dos o más etapas amplificadoras de RF en cascada, la impedancia de entrada de una etapa actúa como impedancia de carga de la etapa precedente. Para garantizar la máxima transferencia de señal entre etapas y mantener la selectividad del sistema, debe proveerse un medio apropiado de acoplamiento.

En amplificadores de RF con FETs, las etapas pueden acoplarse directamente debido a la alta impedancia de entrada de estos dispositivos. En amplificadores de RF con transistores las etapas deben acoplarse mediante un transformador u otro método que elimine o minimize el efecto de carga impuesto por la baja impedancia de entrada.

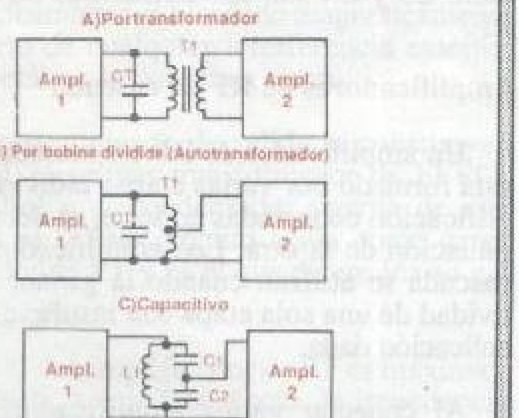

Métodos de acoplamiento

Fig. 297

En la figura 297 se muestran los 3 métodos básicos de acoplamiento entre amplificadores sintonizados de RF: por transformador, por bobina dividida y capacitivo. Los dos primeros son los más populares. El acoplamiento capacitivo se utiliza principalmente en VHF y UHF.

Neutralización de amplificadores de RF

Los amplificadores de RF, especialmente aquellos que trabajan a frecuencias muy altas, tienden a ser inestables, es decir, a oscilar. En el caso de amplificadores a transistores y FETs, la razón de esta inestabilidad es la realimentación positiva introducida por las capacitancias internas del dispositivo.

Los transistores y FETs se pueden utilizar eficientemente como amplificadores de RF a frecuencias muy altas si se neutralizan. La neutralización consiste en tomar una parte de la señal de salida y realimentarla negativamente a la entrada. De este modo, se cancela o minimiza la realimentación positiva que causa la oscilación.

En la figura 298 se muestra el circuito de un amplificador de RF neutralizado. La señal de realimentación se toma del secundario del transformador y se aplica a la base de Q1 a través del condensador de neutralización CN. El valor de CN debe elegirse de tal modo que cancele la realimentación interna de la capacitancia entre el colector y la base (Cob).

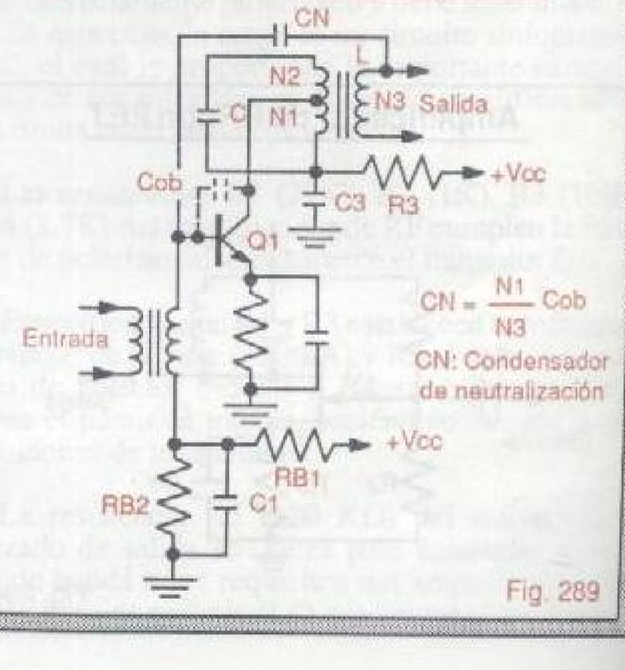

Amplificador de RF neutralizado

$$CN = \frac{N1}{N3}\,Cob$$

CN: Condensador de neutralización

Fig. 289

Circuitos osciladores

Generalidades

Como sabemos, los osciladores son circuitos que generan señales periódicas, de voltaje o corriente, de una determinada frecuencia, amplitud y forma de onda.

La teoría general de los osciladores se estudia en la Lección 19 de este curso, pero debido a que estos circuitos son bloques fundamentales de todo equipo o aparato de radio o comunicaciones, vamos a profundizar más en este tema en la presente lección.

La mayoría de los osciladores utilizados en los aparatos de radio y comunicaciones son amplificadores realimentados positivamente y generan formas de onda sinusoidales.

Para garantizar la permanencia de las oscilaciones, la señal de realimentación debe aplicarse en el momento preciso, con la fase correcta y en la cantidad apropiada. Una realimentación excesiva puede provocar distorsión, calentamiento e inestabilidad.

A pesar de ser esencialmente un amplificador, un oscilador no necesita de una señal de entrada para generar una señal de salida. Las únicas entradas externas de un oscilador son los terminales de la fuente de alimentación (figura 299).

El amplificador como oscilador

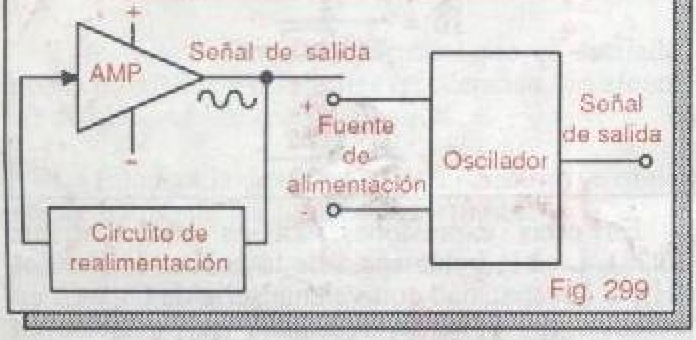

Fig. 299

Las oscilaciones se inician tan pronto se aplica el voltaje de la fuente y se sostienen por sí mismas. En consecuencia, los osciladores convierten la corriente continua de la fuente, en una corriente alterna que puede tener cualquier forma de onda, dependiendo del diseño.

En la figura 300 se muestran las formas de onda más comunes entregadas por los osciladores.

Formas de onda típicas de osciladores

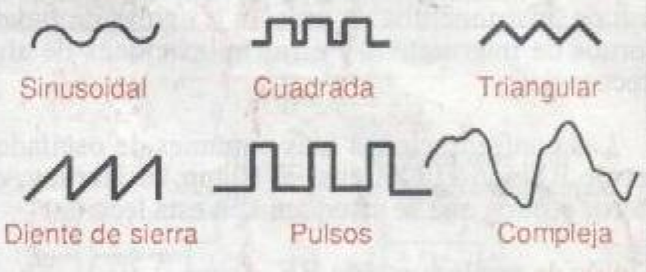

Fig. 300

Los circuitos que producen señales no sinusoidales se denominan comúnmente *osciladores de relajación*. Las formas de onda no sinusoidales, generadas por los osciladores de relajación, son ricas en armónicos de diversos órdenes, es decir, contienen frecuencias que son múltiplos enteros de la frecuencia fundamental de oscilación.

Los osciladores de relajación se utilizan frecuentemente como instrumentos de prueba de amplificadores de audio y de RF. Un ejemplo es el inyector de señales construido en la actividad Nº 10 de este curso. Este oscilador, en particular, produce un tren continuo de ondas cuadradas.

Los osciladores de relajación se emplean también en ciertas aplicaciones de VHF y UHF, para obtener señales sinusoidales puras a partir de señales no sinusoidales. Este tipo de circuitos se denominan *osciladores armónicos*.

Un oscilador armónico consiste de un oscilador de relajación de baja frecuencia, altamente estable, sintonizado a la salida a la frecuencia de un armónico específico. El armónico elegido, por ejemplo, el 5º, es una señal sinusoidal prácticamente pura (figura 301).

Osciladores armónicos

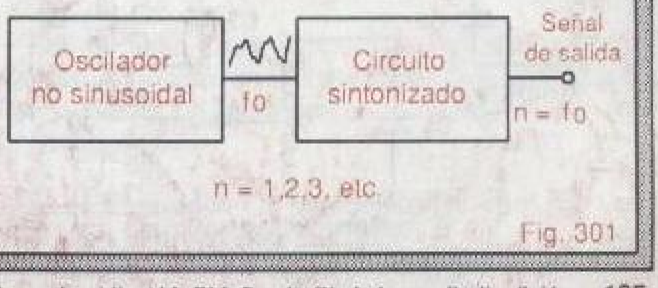

Fig. 301

En esta lección estudiaremos fundamentalmente los osciladores de RF utilizados como osciladores locales, en los receptores de radio. Este tipo de circuitos suministran señales de alta frecuencia, por encima de 30 KHz.

Los osciladores de RF se utilizan en los transmisores y receptores de radio y TV, fuentes de alto voltaje, instrumentos de prueba, sistemas de radar, hornos de microondas y otras aplicaciones de alta frecuencia.

Las configuraciones más comunes de osciladores de RF son el Colpitts, el Clapp, el Hartley, el Pierce y otros que se estudiarán en esta lección.

Clases de osciladores de RF

Prácticamente todos los osciladores de RF son realimentados y utilizan circuitos resonantes LC o cristales de cuarzo como elementos determinantes de la frecuencia de oscilación, y dispositivos activos como tubos, transistores, FETs o circuitos integrados como amplificadores básicos.

Los cristales y circuitos LC determinan, además, la estabilidad de frecuencia, el rango de sintonía y la amplitud y pureza de la señal de salida. Los osciladores más estables utilizan cristales. Sin embargo, cuando se requieren osciladores de frecuencia variable, se prefieren los circuitos LC.

La estabilidad de frecuencia, los osciladores de frecuencia variable (VFOs) y los osciladores a cristal son temas que se tratan más adelante, dentro de esta misma lección. Por ahora, estudiaremos los osciladores realimentados LC más comunes.

En la figura 302 se muestran las configuraciones de circuitos resonantes LC más utilizadas en osciladores de RF. En el caso de osciladores a transistores, el tanque LC se encuentra, por lo regular, en el circuito de colector y ofrece su máxima impedancia a la frecuencia de resonancia.

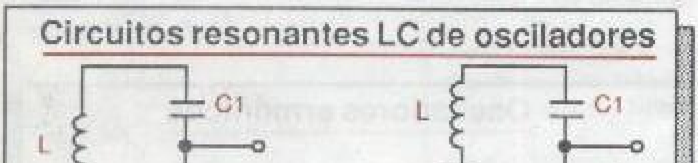
Circuitos resonantes LC de osciladores

Fig. 302

Los tanques resonantes LC anteriores son la base de los osciladores Colpitts, Clapp y Hartley, que se estudian a continuación. Estos mismos circuitos se pueden también implementar utilizando cristales. Desde el punto de vista eléctrico, los cristales se comportan como circuitos resonantes de alto Q.

Osciladores Colpitts

En un oscilador tipo Colpitts (figura 303), la realimentación positiva la realiza el divisor de voltaje capacitivo formado por los condensadores C1 y C2. El voltaje de salida, entre colector y tierra, se desarrolla sobre C1. El voltaje de realimentación aparece sobre C2.

$$fo = \frac{1}{2\pi \times \sqrt{L \times \frac{C1 \times C2}{C1+C2}}}$$

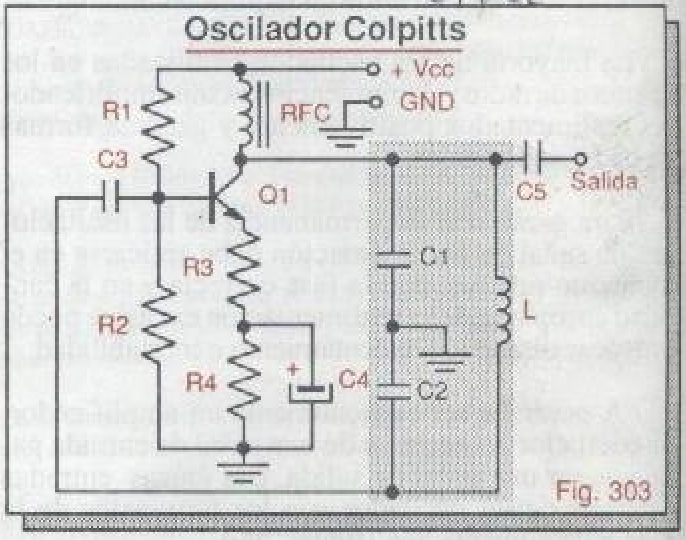
Oscilador Colpitts

Fig. 303

La frecuencia de oscilación del circuito se obtiene mediante las siguientes fórmulas:

$$F_0 = \frac{1}{2\pi\sqrt{LC}}$$

$$C = \frac{C1 \times C2}{C1 + C2}$$

En estas expresiones, 2π es una constante (6.28), L es la inductancia de la bobina en Henrios (H), C la capacidad equivalente serie de C1 y C2 en Faradios (F) y Fo la frecuencia de oscilación en Hertz (Hz). En general, C2 es siempre mayor que C1.

La relación C1/C2 se denomina *factor de realimentación* (ß) del oscilador y es siempre menor de 1. El factor de realimentación de un oscilador mide la proporción de señal de salida que se utiliza como señal de realimentación. Un ß de 0.1, por ejemplo, significa que el 10% de la señal de salida se utiliza como señal de realimentación.

La ganancia de voltaje propia del circuito (Av) debe ser siempre mayor o igual a **1/ß** (el inverso de ß) para que el sistema oscile. A esta condición se le denomina *criterio de Barkhausen* en la teoría de osciladores sinusoidales y es válido en todos los circuitos que siguen, definiendo apropiadamente el factor de realimentación ß.

Por ejemplo, si L=10 µH, C1=47 pF y C2=470 pF, entonces C≈43 pF, Fo≈7.65 MHz y ß=0.1. En consecuencia, la ganancia Av del amplificador debe ser mayor de 1/0.1=10 para que el circuito oscile.

Osciladores Clapp

El oscilador tipo Clapp (figura 304) es una versión mejorada del oscilador Colpitts descrito anteriormente. Este circuito también utiliza realimentación capacitiva, pero incorpora un tercer condensador en el circuito que determina la frecuencia (C3).

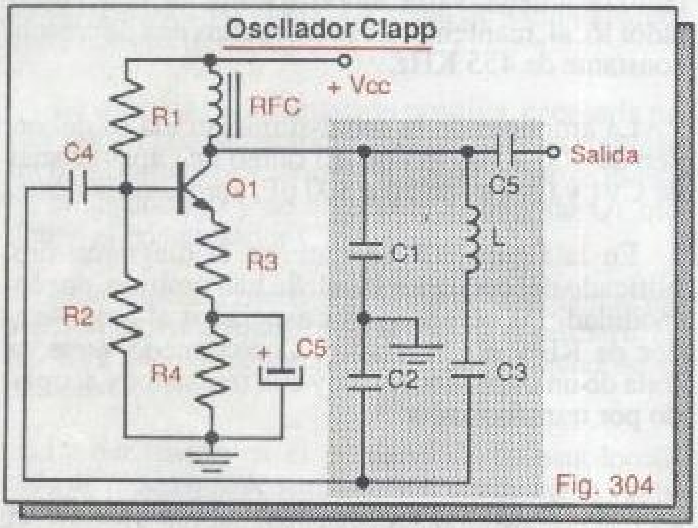

Fig. 304

Este nuevo elemento permite que el oscilador pueda trabajarse a diferentes frecuencias, sin afectar las condiciones de realimentación.

La frecuencia de oscilación del circuito se obtiene por medio de las siguientes fórmulas:

$$F_0 = \frac{1}{2\pi\sqrt{LC}}$$

$$\frac{1}{C} = \frac{1}{C1} + \frac{1}{C2} + \frac{1}{C3}$$

En estas expresiones, 2π=6.28, L es la inductancia de la bobina, C la capacidad equivalente serie de C1, C2 y C3 y Fo la frecuencia de oscilación. El valor de C3 es muy pequeño comparado con el de C1 y C2. El factor de realimentación es **ß=C1/C2**.

Osciladores Hartley

En un oscilador tipo Hartley (figura 305), la realimentación positiva la realiza el divisor de voltaje inductivo formado por las bobinas L1 y L2, acopladas magnéticamente.

$$f_0 = \frac{1}{2\pi \times \sqrt{(L_1 + L_2) \times C}}$$

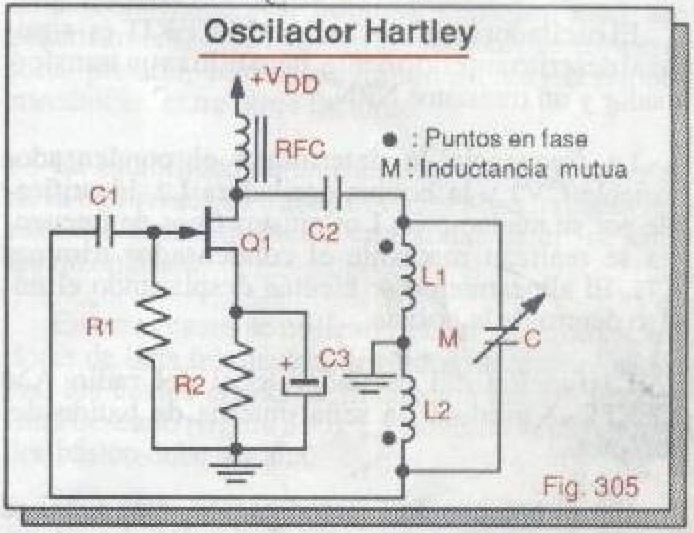

Fig. 305

El voltaje de salida se desarrolla sobre L1 y el de realimentación sobre L2. El condensador variable C permite sintonizar el oscilador en una amplia gama de frecuencias.

La frecuencia de oscilación del circuito se obtiene por medio de las siguiente fórmulas:

$$X_L = 2\pi \times f \times L$$

$$f = \frac{X_L}{2\pi \cdot L}$$

$$L = \frac{X_L}{2\pi \times f}$$

$$4\pi^2 \times f^2 + L \times C = 1$$

$$F_0 = \frac{1}{2\pi\sqrt{LC}} = \frac{0,159}{\sqrt{L \times C}}$$

$$L = \frac{1}{4\pi^2 \cdot f_0^2 \cdot C}$$

$$L = L1 + L2 + 2M$$

$$C = \frac{1}{4\pi^2 \times f_0^2 \cdot L}$$

En estas expresiones, 2π es una constante (6.28), L es la inductancia equivalente serie de L1 y L2, M es la inductancia mutua entre ambos devanados, C la capacidad del condensador de sintonía y Fo la frecuencia de oscilación. El factor de realimentación de este circuito es **ß=L2/L** y debe ser menor de 1.

Las bobinas L1 y L2 constituyen realmente un autotransformador en el cual L1 es el primario y L2 el secundario. La inductancia de L2 es siempre mayor que la de L1.

El condensador C1 suministra la señal de realimentación a la compuerta del FET y el condensador C2 desacopla el nivel de CC de la señal de salida.

La inductancia mutua M depende del grado de acoplamiento de los dos devanados y varía con la ubicación del núcleo. Al mover el núcleo, cambia la inductancia mutua y, en consecuencia, varía la frecuencia. Este es un procedimiento muy usual en la alineación o ajuste de los receptores de radio.

El oscilador local de RF del radio AM CEKIT

El oscilador local del radio AM CEKIT es similar al descrito anteriormente, pero utiliza un transformador y un transistor NPN.

La frecuencia la determinan el condensador variable CV2 y la bobina osciladora L2, identificable por su núcleo rojo. Los ajustes finos de frecuencia se realizan mediante el condensador trimmer CT2. El alineamiento se efectúa desplazando el núcleo dentro de la bobina.

La función del oscilador local del radio AM CEKIT es producir la señal interna de batido del receptor.

Por el proceso de heterodinación, esta señal se combina con la captada en el circuito de antena, produciéndose como resultado una señal de frecuencia intermedia (FI) a la salida del mezclador o conversor de RF (figura 306).

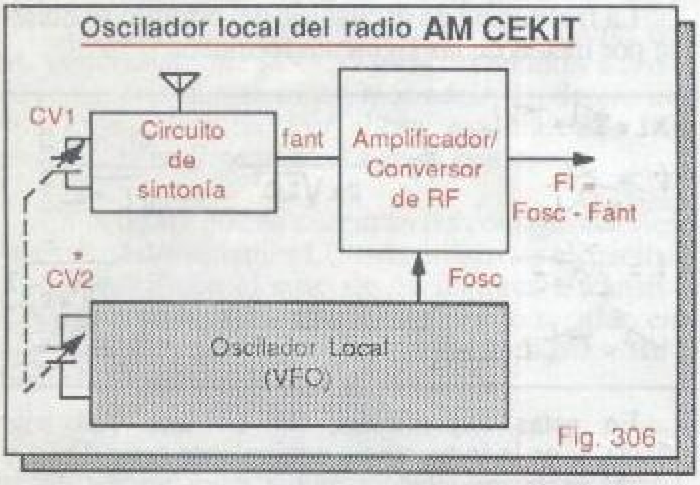
Oscilador local del radio AM CEKIT

Fig. 306

El oscilador local entrega una señal sinusoidal pura, sin modulación, de frecuencia variable y amplitud constante. Los osciladores con estas características se denominan VFOs (*osciladores de frecuencia variable*) y se utilizan ampliamente en los receptores de comunicaciones que trabajan en base al principio heterodino.

El oscilador local es sintonizable sobre el rango comprendido entre 995 y 2055 KHz y su frecuencia es siempre 455 KHz más alta que la de la portadora de RF captada en el circuito de sintonía.

Para lograr este sincronismo, los condensadores variables de ambos circuitos (CV1 y CV2) trabajan en "tandem", es decir, están acoplados mecánicamente al mismo eje (figura 307).

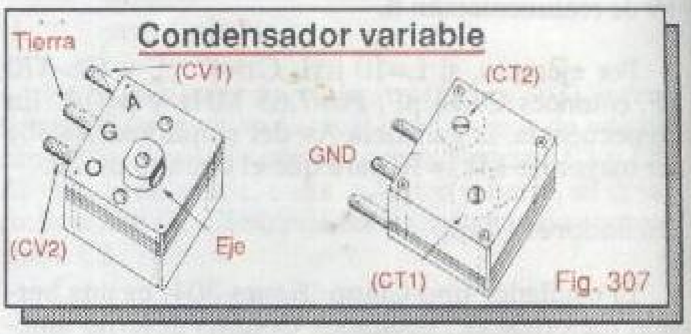
Condensador variable

Fig. 307

En el diagrama del radio este acoplamiento se representa por una línea punteada.

Así, al variar la frecuencia de resonancia del circuito de antena, varía automáticamente la del oscilador local, manteniéndose entre ellas una diferencia constante de 455 KHz.

La amplitud de la señal suministrada es del orden de 300 a 500 mVpp. El rango de capacitancias de CV1 y CV2 es de 30 a 300 pF, aproximadamente.

En la figura 308 se muestra el diagrama simplificado del oscilador local. Se han omitido, por comodidad, los componentes asociados al amplificador de RF y al mezclador. Como puede verse, se trata de un oscilador Hartley con transistor y acoplado por transformador.

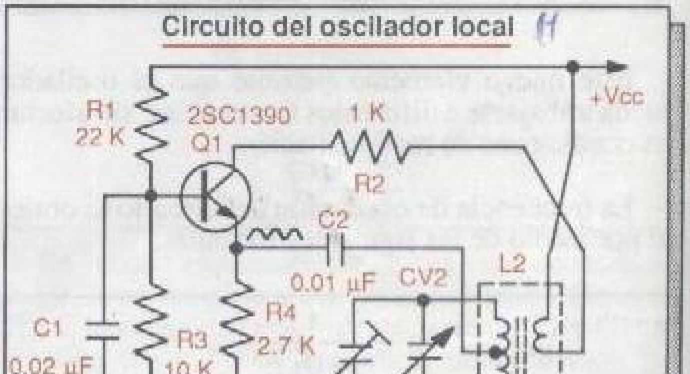
Circuito del oscilador local

Fig. 308

La frecuencia del oscilador local la determinan el condensador variable CV2, la bobina L2 y el condensador ajustable o "trimmer" CT2. Este último se utiliza para realizar ajustes finos de frecuencia.

La bobina L2 se denomina también "osciladora" y se identifica por su núcleo rojo móvil (figura 309). Variando la posición del núcleo, varía la inductancia y, en consecuencia, la frecuencia de oscilación.

Detalle de la bobina osciladora

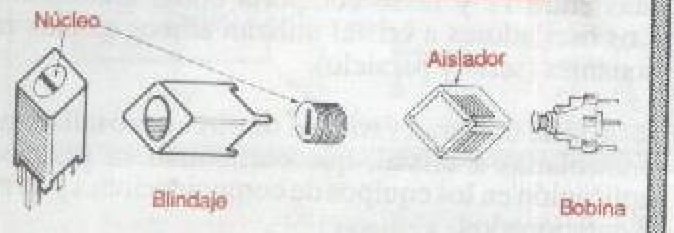

Fig. 309

El elemento activo básico de la etapa osciladora es el transistor NPN Q1. La señal de salida aparece entre colector y tierra.

La señal de realimentación positiva, necesaria para sostener las oscilaciones sinusoidales, la suministra el devanado primario de L2. Esta señal se induce en el secundario y se inyecta al emisor de Q1 mediante el condensador C2.

El transistor amplifica la señal de realimentación y suministra una muestra de la misma al primario de L2. Este la induce en el primario, repitiéndose el proceso.

La derivación en el primario de L2 está localizada de modo que se suministre únicamente la cantidad de realimentación necesaria para asegurar la operación confiable del circuito.

El condensador C1, conecta dinámicamente a tierra la base de Q1, permitiendo que la señal de realimentación quede aplicada entre emisor y tierra. Las resistencias R1, R2, R3 y R4 proveen la polarización de corriente continua de la etapa.

El mecanismo de oscilación de un circuito RLC resonante, como el utilizado en nuestro oscilador local, se explica en la Lección 19.

Al conectar la fuente de alimentación, se produce un pulso de corriente que carga el condensador CV2 e inicia automáticamente las oscilaciones.

Importancia de la estabilidad de frecuencia en los osciladores de RF

La consideración de diseño más importante de un oscilador de RF es la *estabilidad de frecuencia.*

La estabilidad es una medida de la capacidad del circuito de generar una señal de frecuencia constante y depende, principalmente, de la calidad de los componentes utilizados en su fabricación.

Los osciladores de RF más estables se construyen con componentes de precisión y se alimentan con fuentes reguladas. Los componentes de precisión (condensadores, bobinas, cristales, etc.) son prácticamente insensibles a los cambios de temperatura, presión, humedad, voltaje y a las vibraciones mecánicas, entre otros factores.

La estabilidad de un oscilador depende también de la frecuencia de trabajo. A muy altas frecuencias, resulta difícil mantener la estabilidad dentro de límites razonables.

En estos casos se utilizan frecuentemente osciladores de baja frecuencia, seguidos de multiplicadores, los cuales elevan la frecuencia inicial al valor final deseado (figura 310). La estabilidad del oscilador básico debe ser alta.

Obtención de muy altas frecuencias

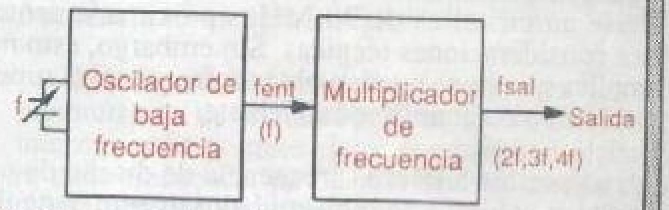

"La frecuencia de salida es 2, 3, 4, .. veces la frecuencia de entrada"

Fig. 310

La inestabilidad de un oscilador de RF puede ser también causa de prácticas de diseño y ensamble deficientes. Para mejorar la estabilidad, es útil observar las siguientes recomendaciones generales:

- Utilizar siempre baterías o fuentes de alimentación reguladas, con buen filtraje. Las variaciones de voltaje producen desplazamientos de frecuencia, alterando las condiciones de operación del oscilador.

- Utilizar bobinas de núcleo de aire o cerámico y condensadores de poliéster o de mica y plata en la red determinadora de frecuencia. Mientras sea posible, evitar el uso de bobinas de núcleo magnético, porque su inductancia tiende a cambiar con las variaciones de temperatura.

- Construir el oscilador sobre un chasis rígido y cerciorarse de que todos los componentes estén firmemente asegurados en su sitio. Si es posible, inmovilizar las espiras de la bobina osciladora con cemento plástico y encerrar el circuito dentro de un blindaje metálico.

En general, un oscilador de RF de buena calidad debe generar una señal estable, con un mínimo de ruido y espectralmente pura, es decir, sin distorsión ni espúreas. Las *espúreas* son oscilaciones parásitas, indeseables, que ocurren a frecuencias diferentes a la de diseño.

Una forma de minimizar las espúreas, es reducir el Q del circuito determinador de frecuencia, sacrificándose así la ganancia.

Osciladores a cristal

Los osciladores a cristal, se caracterizan por utilizar como elemento determinante de la frecuencia de oscilación un cristal piezoeléctrico, generalmente de cuarzo, que ha sido tallado a unas dimensiones específicas.

La frecuencia a la cual oscila el cristal depende, fundamentalmente, de su espesor y es muy estable, no siendo afectada por la temperatura, el voltaje de alimentación y otros factores.

La mayor frecuencia práctica a la cual puede tallarse un cristal es de 30 MHz aproximadamente, por consideraciones técnicas. Sin embargo, esto no implica que no se puedan obtener frecuencias superiores, como veremos más adelante.

Es posible alterar la frecuencia de diseño de un cristal si se abre y se lima cuidadosamente parte de su perfil, pero se corre el riesgo de dejarlo inservible.

El elemento activo de un oscilador a cristal puede ser una válvula o tubo, un transistor, un FET o un circuito integrado. Para efectos prácticos, un cristal de cuarzo es equivalente, en su comportamiento eléctrico, a un circuito RLC mixto como el que muestra la figura 311.

Circuito RLC equivalente de un cristal

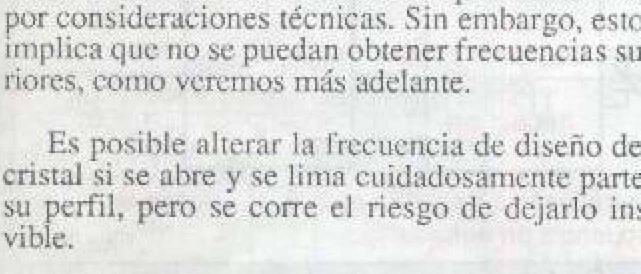

Fig. 311

Las propiedades reactivas del dispositivo pueden resumirse en los siguientes términos.

A la frecuencia f1, el cristal exhibe resonancia serie (Z muy baja). A la frecuencia f2, exhibe resonancia paralelo (Z muy alta).

A frecuencias por debajo de f1 o por encima de f2, se comporta como un condensador y a frecuencias entre f1 y f2 se comporta como una bobina. Los osciladores a cristal utilizan ambos efectos resonantes (serie y paralelo).

Existe una gran variedad de circuitos osciladores controlados a cristal, que encuentran su principal aplicación en los equipos de comunicaciones y de radioaficionados.

En esta lección discutiremos las configuraciones Colpitts, Pierce, armónica y de sobretono, que son las más representativas.

La mayoría de los osciladores a cristal, utilizan condensadores externos para producir la función de realimentación. Sin embargo, algunos circuitos no los necesitan. En estos últimos, se aprovechan las capacitancias internas del transistor o del dispositivo activo utilizado como amplificador.

En la figura 312 se muestra el circuito práctico de un oscilador a cristal de 7 MHz tipo Colpitts. El divisor capacitivo formado por C2 y C3 provee la realimentación positiva de voltaje necesaria para generar las oscilaciones.

El condensador C1 actúa como trimmer y se utiliza para ajustar la frecuencia del cristal al valor para el cual fue tallado.

Oscilador Colpitts a cristal de 7 MHz

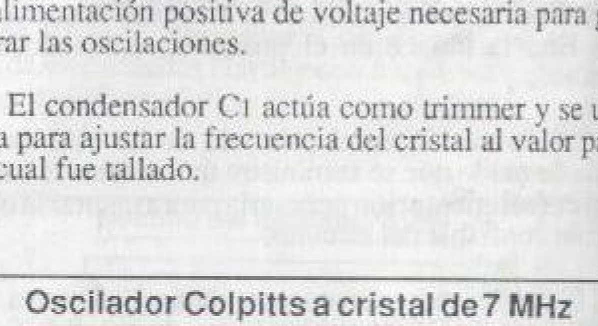

Fig. 312

En la figura 313 se muestra el circuito práctico de un oscilador a cristal de 3.5 MHz tipo Pierce. El circuito resonante serie formado por C1 y el cristal

Oscilador Pierce a cristal de 3.5MHz

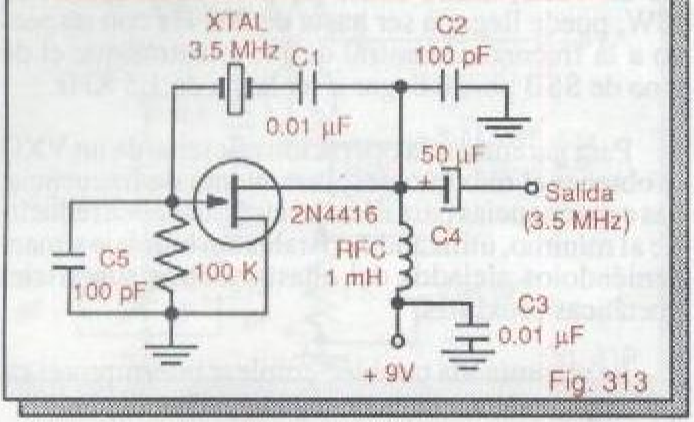
XTAL
3.5 MHz
C1
0.01 µF
C2
100 pF
50 µF
Salida
(3.5 MHz)
2N4416
C4
RFC
1 mH
C5
100 pF
100 K
C3
0.01 µF
+ 9V

Fig. 313

provee la realimentación positiva de corriente necesaria para generar las oscilaciones. El oscilador tipo Pierce utiliza el efecto resonante serie del cristal.

En la figura 314 se muestra el circuito práctico de un oscilador *armónico* o Miller. Este tipo de montaje se utiliza para obtener frecuencias que son múltiplos enteros de la frecuencia fundamental de oscilación del cristal.

Oscilador armónico de 14 MHz

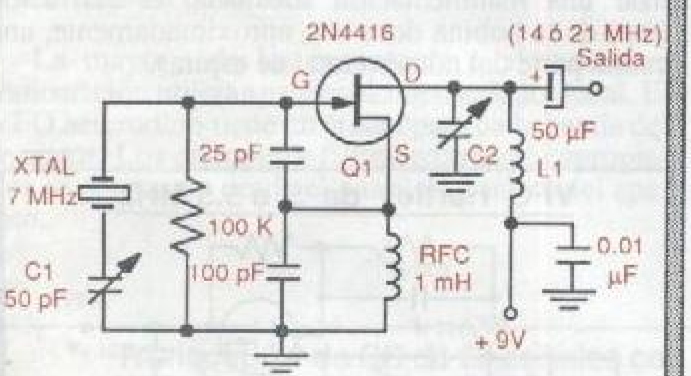
2N4416
(14 ó 21 MHz)
Salida
G
D
S
Q1
C2
50 µF
L1
XTAL
7 MHz
25 pF
100 K
100 pF
C1
50 pF
RFC
1 mH
0.01 µF
+ 9V

Fig. 314

La frecuencia armónica deseada, se sintoniza mediante el circuito resonante de salida, constituido por L1 y C2. Este oscilador utiliza el efecto resonante paralelo del cristal.

En la figura 315 se muestra el circuito práctico de un oscilador de *sobretono*. En este tipo de montaje, el cristal se talla para que oscile a una frecuencia llamada "sobretono", la cual es un múltiplo impar de su frecuencia fundamental de oscilación.

La frecuencia de sobretono deseada (3f, 5f, etc.), se programa mediante el circuito sintonizado de salida, constituido por L1 y C2.

Oscilador de tercer sobretono de 42 MHz

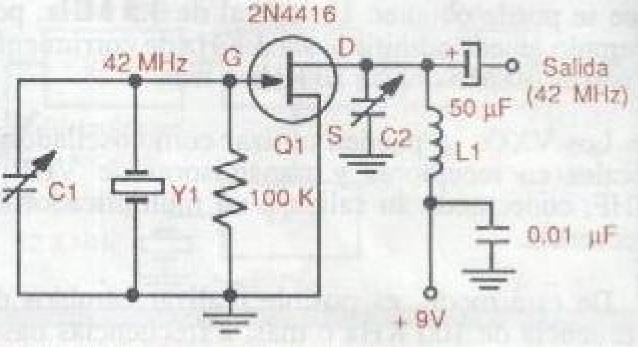
2N4416
42 MHz
G
D
Salida
(42 MHz)
50 µF
C1
Y1
100 K
Q1
S
C2
L1
0.01 µF
+ 9V

Y1: Cristal de sobretono de 42 MHz

Fig. 315

El oscilador de sobretono emplea el efecto resonante serie del cristal y se caracteriza por entregar una señal de alta pureza, sin espúreas. El estado actual de la tecnología permite obtener cristales que operan hasta en la 11ª frecuencia de sobretono.

En la cápsula de un cristal tallado para este propósito no se indica la frecuencia fundamental sino la de sobretono utilizable.

En ciertas aplicaciones, se requiere que el oscilador opere sólo a frecuencias específicas. Para cada frecuencia se necesita un cristal diferente.

En la figura 316 se muestra un circuito típico de conmutación de cristales para oscilador Colpitts. El interruptor S1 permite elegir el cristal apropiado para cada frecuencia de operación. Los diodos aislan eléctricamente los cristales no utilizados.

Conmutación de cristales mediante diodos

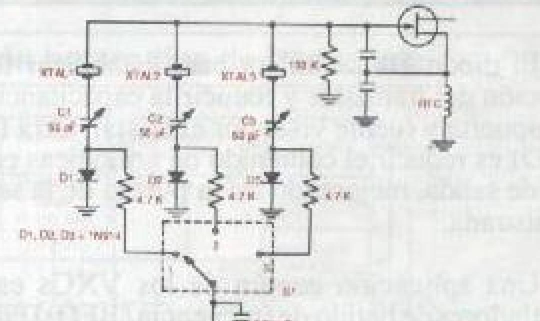
XTAL 1
XTAL 2
XTAL 3
C1
C2
C3
D1
D2
D3
4.7 K
RFC

Fig. 316

Osciladores a cristal de frecuencia variable (VXOs)

Los osciladores a cristal de frecuencia variable (VXOs) se utilizan en cambio de los osciladores de cristal convencionales cuando se necesita desplazar o correr la frecuencia del cristal unos pocos KHz respecto a su valor central.

Entre más alta sea la frecuencia fundamental del cristal, más amplio es el corrimiento de frecuencia que se puede obtener. Un cristal de 3.5 MHz, por ejemplo, puede admitir hasta 3 KHz de corrimiento y uno de 10 MHz hasta 10 KHz o más.

Los VXOs se pueden utilizar como osciladores locales en receptores y transmisores de VHF y UHF, conectando su salida a un multiplicador de frecuencia.

De este modo, es posible realizar cambios de frecuencia de 100 KHz o más, a frecuencias hasta de 140 MHz o superiores, utilizando, por ejemplo, cristales de 7 MHz.

En la figura 317 se muestra el diagrama de un VXO Colpitts a FET utilizando un cristal de cuarzo de 7 MHz. El desplazamiento de la frecuencia se obtiene ajustando la inductancia de la bobina L.

El circuito provee hasta 5 KHz de corrimiento por debajo de la frecuencia de tallado del cristal, permitiendo obtener oscilaciones sinusoidales entre 6995 y 7000 KHz.

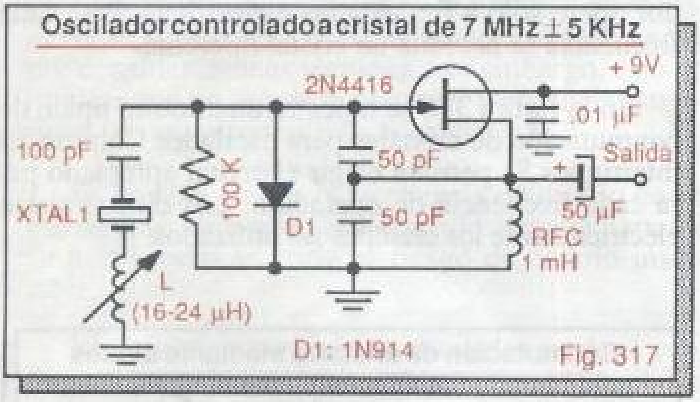
Osciladorcontroladoacristal de 7 MHz ± 5 KHz

Fig. 317

El diodo D1 se utiliza para estabilizar la polarización del transistor y reducir la capacitancia entre compuerta y fuente vista por el cristal. Otra función de D1 es reducir el contenido de armónicas en la señal de salida, mejorando así la pureza de la señal suministrada.

Una aplicación común de los VXOs es como osciladores de batido de frecuencia (BFOs) en receptores de banda lateral única (SSB) y de onda continua (CW) o telegrafía.

El BFO se encarga de crear internamente la portadora suprimida durante la transmisión y la reinserta en el modulador, para que así la señal captada pueda ser procesada.

Los receptores de banda lateral única y de telegrafía se discuten en el volumen de banda ciudadana y radioafición. El desplazamiento típico de un VXO, usado como BFO en la recepción de CW, puede llegar a ser hasta de 700 Hz con respecto a la frecuencia central de FI, mientras que el de uno de SSB puede llegar a ser hasta de 1.5 KHz.

Para garantizar la operación eficiente de un VXO y obtener el máximo desplazamiento de frecuencia, las capacitancias parásitas del circuito deben reducirse al mínimo, utilizando cristales sin blindaje y manteniéndolos alejados del chasís y otras superficies metálicas próximas.

Se recomienda también emplear interruptores rápidos y condensadores de baja capacitancia.

Osciladores de frecuencia variable (VFOs)

Los osciladores de frecuencia variable (VFOs) son similares en su funcionamiento a los VXOs descritos anteriormente. Sin embargo, los VFOs no utilizan cristales y tienen rangos de variación de frecuencia más amplios. Los VFOs prácticos pueden operar a frecuencias hasta de 10 MHz. Para frecuencias más altas se utilizan VFOs *heterodinos*.

En la figura 318 se muestra el circuito práctico de un VFO tipo Hartley con MOSFET. La frecuencia del oscilador la determinan la bobina L1, el condensador variable C1 y el trimmer C2. Para garantizar una realimentación adecuada, la derivación (tap) de la bobina debe ser, aproximadamente, una cuarta parte del número total de espiras.

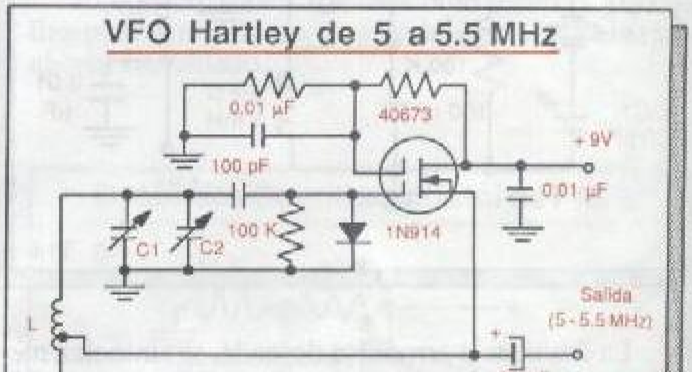
VFO Hartley de 5 a 5.5 MHz

Fig. 318

En la figura 319 se muestra el circuito práctico de un VFO tipo Colpitts con transistor bipolar. En el diagrama se indican los componentes determinantes de la frecuencia de oscilación. Este oscilador es muy estable si se utilizan condensadores con dieléctrico de poliéster o mylar.

Los circuitos mostrados anteriormente han sido diseñados para cubrir la banda de frecuencias entre

VFO Colpitts de 5 a 5.5 MHz

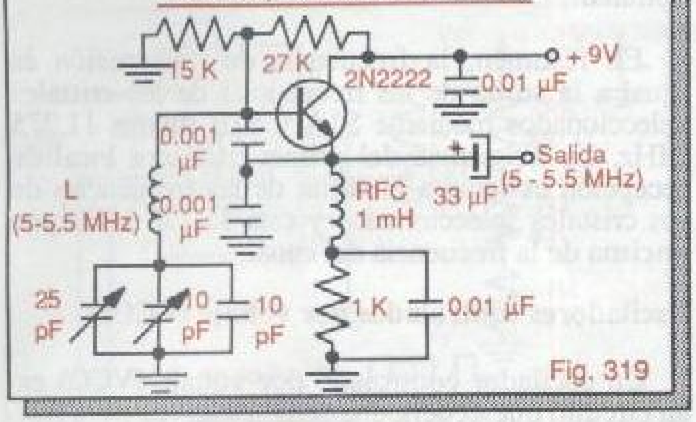
15 K
27 K
2N2222
+ 9V
0.01 µF
0.001 µF
L
(5-5.5 MHz)
0.001 µF
Salida
(5 - 5.5 MHz)
+
33 µF
RFC
1 mH
25 pF
10 pF
10 pF
1 K
0.01 µF

Fig. 319

5 y 5.5 MHz (SSB). Sin embargo, pueden adaptarse a otras frecuencias de operación, utilizando los valores dados como referencia.

Para hacerlo se calcula la reactancia de los condensadores a 5 MHz y se utiliza esta información para determinar los valores en pF requeridos para otras frecuencias.

En la figura 320 se muestra el principio de operación de un VFO heterodino. La salida de un VFO, como los descritos anteriormente, se heterodina en un mezclador con la salida de un VXO para suministrar una frecuencia resultante que es la suma o la diferencia de ambas.

La mayoría de los receptores modernos para radioafición utilizan este estilo de oscilador local. El VFO heterodino tiene un cristal para cada banda del receptor. Los cristales y filtros pasabanda apropiados se conmutan desde el panel de control del aparato.

VFO Heterodino

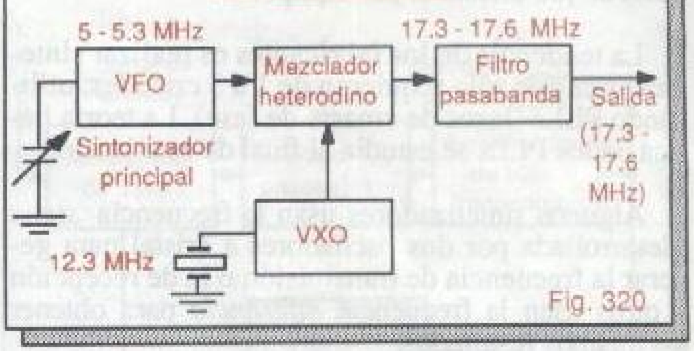
5 - 5.3 MHz
VFO
Sintonizador principal
Mezclador heterodino
17.3 - 17.6 MHz
Filtro pasabanda
Salida
(17.3 - 17.6 MHz)
VXO
12.3 MHz

Fig. 320

Una versión moderna de los VFOs heterodinos son los sintetizadores heterodinos de frecuencia que se estudian enseguida.

Este tipo de circuitos es muy popular en los transceptores de banda ciudadana y su propósito es reducir el número total de cristales necesarios para cubrir los 40 canales en que operan normalmente estos equipos.

Sintetizadores heterodinos de frecuencia para banda ciudadana

Los circuitos de transmisión y recepción de un transceptor de banda ciudadana (CB) utilizan cristales para asegurar el cumplimiento de los rigurosos requisitos de estabilidad y precisión de frecuencia impuestos por las autoridades que regulan este tipo de comunicaciones, como la FCC en los Estados Unidos.

La *sintetización de frecuencia* permite operar todos los canales de CB con sólo unos pocos crista-

Transceptor de CB de 23 canales con sintetización heterodina de frecuencia

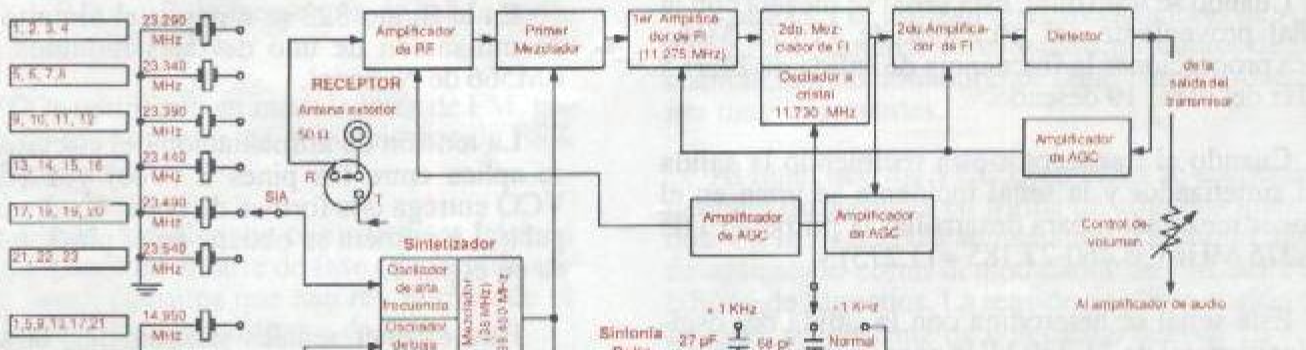
1, 2, 3, 4
23.290 MHz
5, 6, 7, 8
23.340 MHz
9, 10, 11, 12
23.390 MHz
13, 14, 15, 16
23.440 MHz
17, 18, 19, 20
23.490 MHz
21, 22, 23
23.540 MHz
SIA
1,5,9,13,17,21
14.950 MHz
2,6,10,14,18,22
14.960 MHz
3,7,11,15,19
14.970 MHz
4,8,12,16,20,23
14.980 MHz
SIB
Amplificador de RF
Primer Mezclador
RECEPTOR
Antena exterior
50 Ω
R
T
Sintetizador
Oscilador de alta frecuencia
Oscilador de baja frecuencia
Mezclador (38 MHz) 38.450 MHz
1er. Amplificador de FI (11.275 MHz)
2do. Mezclador de FI
Oscilador a cristal 11.730 MHz
2do. Amplificador de FI
Detector
de la salida del transmisor
Amplificador de AGC
Amplificador de AGC
Amplificador de AGC
Control de volumen
Al amplificador de audio
Sintonía Delta
+ 1 KHz
- 1 KHz
27 pF
68 pF
Normal
11.730 MHz
Nota: Transceptor sintonizado en el canal 19 (27.185 MHz)
Oscilador a cristal (11.275 MHz)
11.275 MHz
Mezclador (27 MHz)
Amplificador Buffer
Al Driver (Canal 19)
TRANSMISOR

Fig. 321

les, digamos 10 ó 12, consiguiéndose así un ahorro hasta de ¡36 cristales! por equipo.

La tendencia de los fabricantes es realizar sintetizadores que sólo requieran de 1 a 3 cristales, utilizando PLLs (lazos de amarre de fase). La teoría básica de los PLLs se estudia al final de esta lección.

Algunos sintetizadores usan la frecuencia *suma* desarrollada por dos osciladores a cristal para generar la frecuencia de transmisión o la de recepción y otros usan la frecuencia *diferencia* para obtener los mismos resultados.

Sin importar si la frecuencia desarrollada por el sintetizador es la de transmisión o la de recepción, la otra frecuencia se genera por batido de la frecuencia sintetizada con la frecuencia de salida de un tercer oscilador.

En la figura 321 se muestra el diagrama funcional de bloques de un transceptor de banda ciudadana que utiliza el principio de sintetización heterodina de frecuencia y emplea sólo 10 cristales para cubrir los 23 canales. Veamos cómo lo hace.

La salida del sintetizador, en este circuito está 11.275 MHz *por encima* de la frecuencia del canal y se usa como frecuencia de inyección para el primer mezclador del receptor. La frecuencia diferencia resultante (11.275 MHz) alimenta el primer amplificador de FI.

Cuando el transceptor está transmitiendo, el mezclador del sintetizador alimenta señales en un mezclador de 27 MHz, el cual genera frecuencias diferencia en el canal deseado.

En el canal 19, por ejemplo, las señales de los osciladores de 23.490 MHz y 14.970 MHz se suman en el mezclador del sintetizador para producir una salida de 38.460 MHz.

Cuando se transmite, esta señal se mezcla con la señal proveniente del oscilador de 11.275 MHz, para proporcionar la frecuencia de salida de 27.185 MHz del canal 19 deseado.

Cuando el transceptor está recibiendo la salida del sintetizador y la señal incidente se unen en el primer mezclador, para desarrollar la primera FI de 11.275 MHz (38.460-27.185 =11.275).

Esta señal se heterodina con la salida del oscilador a cristal de 11.730 MHz para desarrollar la segunda FI de 455 KHz, la cual se inyecta al detector.

Mediante el control de "sintonía delta" los condensadores fijos de 27 y 68 pF se pueden conectar en serie con el cristal de 11.730 MHz para desplazar su frecuencia ±1 KHz con respecto a su valor nominal.

En resumen, la frecuencia de transmisión es igual a la **suma** de las frecuencias de los cristales seleccionados mediante S1A y S1B **menos** 11.275 MHz. La frecuencia del primer oscilador local de recepción es igual a la **suma** de las frecuencias de los cristales seleccionados y está 11.275 MHz **por encima** de la frecuencia del canal.

Osciladores controlados por voltaje (VCOs)

Un oscilador controlado por voltaje (VCO) es un circuito que genera una señal de salida cuya frecuencia es proporcional al valor de un voltaje CC aplicado a su entrada de control (figura 322).

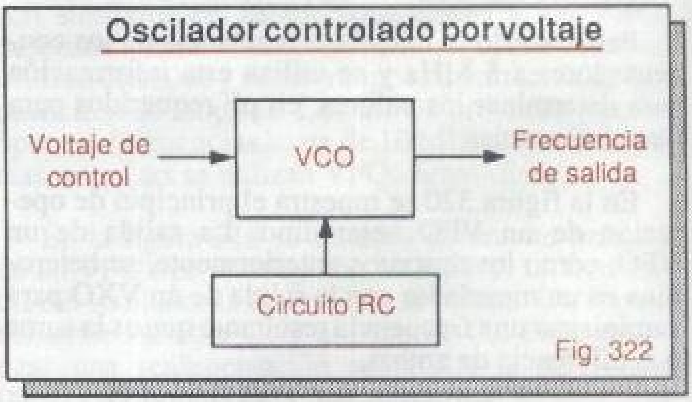

Fig. 322

En otras palabras, un VCO es un convertidor de voltaje a frecuencia. La forma de onda de la señal generada por un VCO es, generalmente, cuadrada o triangular.

Los VCOs se pueden construir a partir de componentes discretos, pero los más sencillos y económicos son los disponibles en forma de circuitos integrados.

En la figura 323 se muestra el circuito práctico de utilización de uno de tales módulos, el VCO LM566 de National.

La tensión de alimentación del circuito (+12V), se aplica entre los pines **8** (Vcc) y **1** (GND). El VCO entrega dos formas de onda: cuadrada y triangular. La primera se obtiene en el pin **3** y la segunda en el pin **4**.

Para obtener señales sinusoidales, la salida del VCO debe seguirse de un filtro o un amplificador sintonizado a la frecuencia deseada. Esta frecuencia debe ser un armónico de la del VCO.

Para efectos de acople la impedancia de salida es 50Ω. El voltaje de control se aplica al pin **5** y pro-

Circuito práctico de VCO con 566

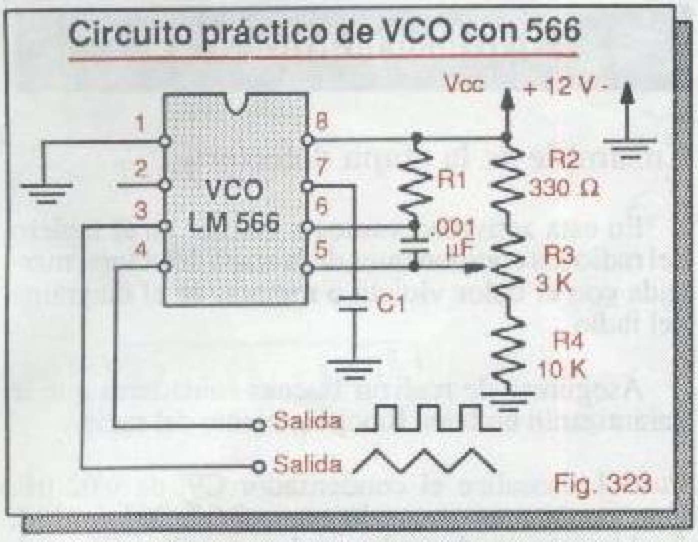

Fig. 323

viene del divisor de tensión resistivo formado por R2, R3 y R4. El condensador de 0.001 µF entre los pines **5** y **6** protege al circuito de oscilaciones parásitas.

El condensador C1, conectado entre el pin **7** y tierra, y la resistencia R1, conectada entre el pin **6** y Vcc, determinan, junto con el voltaje de control aplicado al pin 5, la frecuencia de la señal de salida así:

$$\mathrm{Fsal} = 2.4\ \frac{(\mathrm{Vcc} - \mathrm{V5})}{\mathrm{Vcc}\,(\mathrm{R1} \times \mathrm{C1})}$$

En esta expresión, V5 es el voltaje de control aplicado al pin 5 y debe estar entre 3/4 de Vcc y Vcc. El valor de R1 debe ser siempre de 2 a 20 KΩ.

La máxima frecuencia típica de operación es 1 MHz. Por ejemplo, si Vcc = 12V, V5 = 10V, R1= 2KΩ y C1 = 0.001 µF, la frecuencia de salida sería, entonces, **Fsal** = 200 KHz.

Los VCOs se utilizan en moduladores de FM, generadores de tonos y señales, transmisores de FSK y otras aplicaciones.

Son, además, los bloques constructivos básicos de los PLLs (lazos de amarre de fase o bucles de enganche de fase), circuitos que han revolucionado el diseño de los receptores modernos de radio.

Lazos de amarre de fase (PLLs)

Estructura de un PLL

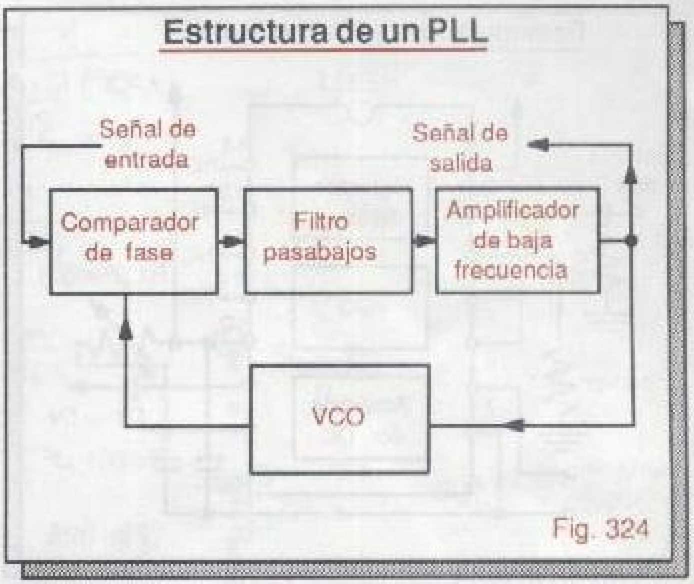

Fig. 324

Un lazo de amarre de fase o PLL (Phase-Locked Loop) es un circuito de control de frecuencia, constituido por un comparador de fase, un filtro pasabajos, un amplificador de baja frecuencia y un oscilador controlado por voltaje (VCO), interconectados tal como se muestra en la figura 324.

En condiciones normales, el VCO oscila a la misma frecuencia de la señal de entrada. Si la frecuencia de esta señal cambia, el comparador de fase produce un voltaje de error, proporcional a la diferencia de frecuencias.

Este voltaje se aplica al VCO y corrije su frecuencia, hasta sincronizarla o engancharla con la de la señal de entrada. El proceso se repite indefinidamente.

El PLL detecta cualquier cambio en la frecuencia de la señal de entrada y lo convierte en un voltaje de error que se suministra amplificado a la salida.

En consecuencia, un PLL es, intrínsecamente, un demodulador de FM. El filtro determina el ancho de banda del circuito y garantiza que sólo se amplifiquen variaciones de baja frecuencia, es decir, señales de audio.

Los PLLs se utilizan en demoduladores de AM, FM y FSK, sintetizadores y multiplicadores de frecuencia, acondicionadores de señal y otras aplicaciones muy interesantes.

Al igual que los VCOs, los PLLs se encuentran también disponibles en forma de circuitos integrados. En la figura 325 se muestra un ejemplo típico de aplicación como demodulador de FM del PLL NE565 de Signetics. La tensión de alimentación del circuito, proveniente de una fuente dual de ±5V, se aplica entre los pines **10** (+Vcc) y **1** (-Vcc).

La señal modulada de FI se aplica a una de las entradas de señal del detector de fase (pin **2**) a través del condensador de 0.01 µF. La señal de salida de audio se obtiene entre los pines **6** y **7**.

Demodulador de FM con PLL 565

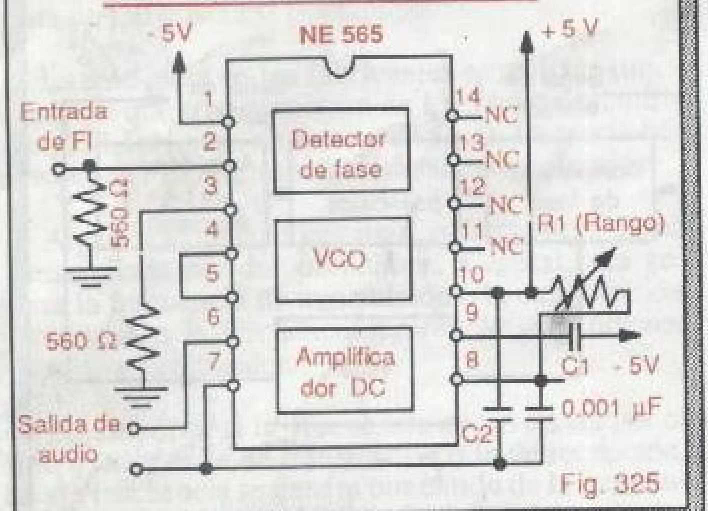

Fig. 325

El potenciómetro R1 y el condensador C1 determinan la frecuencia central (Fo) del VCO. La señal de entrada aplicada al pin 2 hace variar la frecuencia del VCO alrededor de su frecuencia central.

El PLL seguirá las variaciones de frecuencia de la señal de entrada mientras éstas se mantengan dentro de cierto margen llamado *rango de enganche* (ΔL).

El condensador C2 actúa como filtro pasabajos y determina el *rango de captura* del PLL, es decir, el rango de frecuencias alrededor de Fo dentro del cual el circuito puede engancharse con una señal de entrada inicialmente fuera del rango de enganche.

Los valores de F0, ΔL y ΔC se calculan mediante el siguiente grupo de fórmulas:

$$F_o = \frac{0.3}{R1 \times C1}$$

$$\Delta L = \pm 8 \times \frac{F_o}{V_{cc}}$$

$$\Delta C = \pm 0.21 \sqrt{\frac{\Delta L}{C2}}$$

Para que esta última expresión sea válida, ΔL debe expresarse en KHz y C2 en μF. De este modo, ΔC queda también expresado en KHz.

Por ejemplo, si Vcc=±5V (10V en total), R1=7.5KΩ, C1=0.0027μF y C2=0.0027μF, entonces Fo=14 KHz, ΔL=±11 KHz y ΔC= 13 KHz.

Actividad práctica Nº 14

Ensamble de la etapa detectora

En esta actividad vamos a instalar en el tablero del radio los componentes de la etapa detectora, marcada con el color violeta o morado en el diagrama del radio.

Asegúrese de realizar buenas soldaduras que le garantizarán un buen funcionamiento del radio.

Paso 1. Localice el condensador C9, de 0.02 μF. Este puede estar marcado como 0.02, 203 ó 223Z. Suelde este condensador en las puntillas C8 y 17. Recorte el sobrante de los terminales.

Paso 2. Localice el condensador C8, de 0.01 μF. Este puede estar marcado como 0.01, .01 ó 103Z. Suelde sus terminales en las puntillas C4 y 16.

Paso 3. Con alambre de conexión, una las puntillas C8, C5, C6 y E6, teniendo en cuenta no hacer corto con el cable que une las puntillas 16 y 17.

Paso 4. Con alambre de conexión, una las puntillas C7 y C8. Aparte, una las puntillas C3 y C4.

Paso 5. Localice la resistencia R11 de 1KΩ (café-negro-rojo-dorado) y suéldela en las puntillas C3 y C7.

Paso 6. Localice el diodo detector 1N60. Instálelo con cuidado en las puntillas C1 y C3. Fíjese muy bien que el cátodo, marcado por una banda roja o negra, quede en la puntilla C1.

Paso 7. Con alambre de conexión una las puntillas C2 y 15.

El aspecto final de la etapa detectora se muestra en la figura A36.

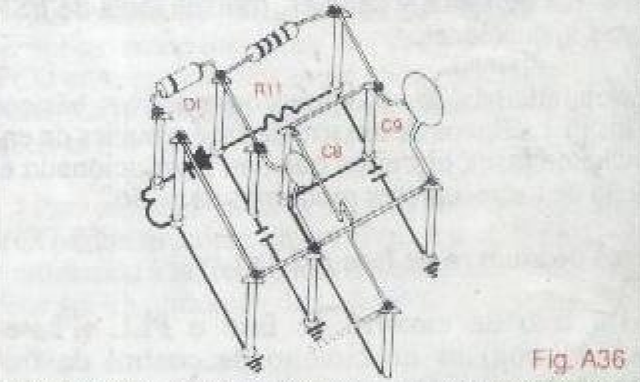

Fig. A36

Mezcladores o conversores de RF

Generalidades

Un mezclador o conversor es un circuito que convierte o traslada una frecuencia, o una banda de frecuencias, de un valor a otro. Los mezcladores son parte integral de todo receptor superheterodino de comunicaciones.

En la figura 326 se muestra el diagrama de bloques de un mezclador de frecuencia asociado a un filtro pasabanda. El mezclador propiamente dicho puede ser un transistor, un FET, una válvula, un circuito integrado, etc.

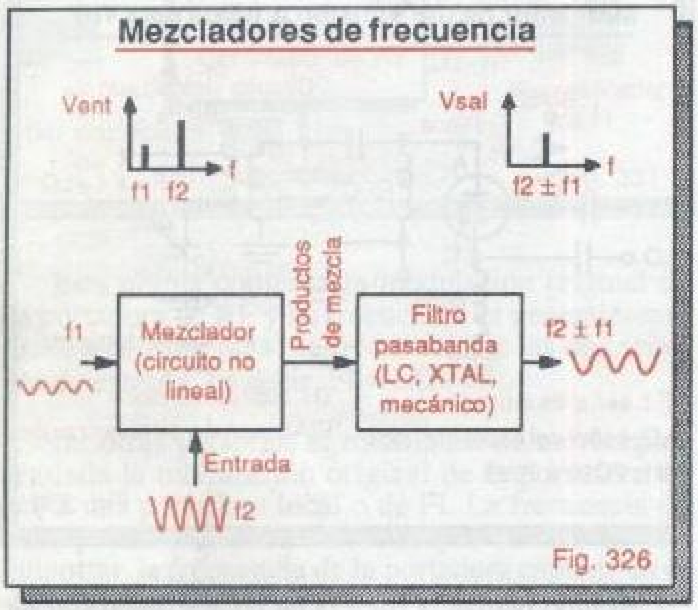
Mezcladores de frecuencia
Vent
f1 f2
f
Vsal
f2 ± f1
f
f1
Mezclador (circuito no lineal)
Productos de mezcla
Filtro pasabanda (LC, XTAL, mecánico)
f2 ± f1
Entrada
f2
Fig. 326

A las entradas del bloque mezclador se aplican dos señales de diferente frecuencia cuyos valores son **f1** y **f2**.

A la salida del filtro se obtiene una señal que tiene una frecuencia igual a la suma o a la diferencia de frecuencias de las señales de entrada (f2±f1). La señal f2 proviene generalmente de un oscilador local.

En el caso de un mezclador de transmisión, f1 proviene de las etapas de FI. La señal de salida (f1+f2) se aplica al amplificador de potencia de RF.

En el caso de un mezclador de recepción, f1 proviene del amplificador de RF o del circuito de sintonía. La señal de salida (f2-f1) se aplica a las etapas amplificadoras de FI.

El proceso de mezcla de las dos señales de entrada origina a la salida del mezclador una gran variedad de *productos de mezcla*, es decir, de señales cuyas frecuencias son una combinación de las frecuencias de entrada (figura 327).

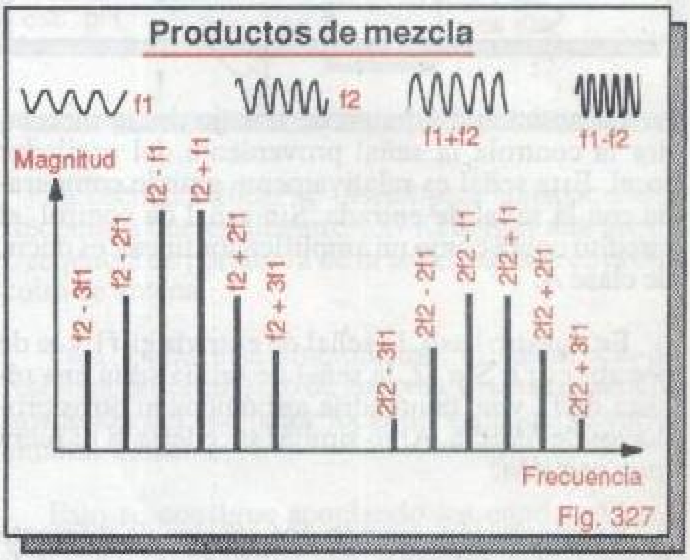
Productos de mezcla

Fig. 327

Los productos de mezcla más notables son:

- Las frecuencias originales f1 y f2
- La frecuencia suma (f1+f2)
- La frecuencia diferencia (f2-f1)

Por ejemplo, si f1=1 MHz y f2= 1.3 MHz, a la salida del mezclador obtendríamos, además de las señales de 1.5 MHz y de 1 MHz, una señal de 300 KHz (f2-f1) y una señal de 2.3 MHz (f1+f2).

El filtro selecciona únicamente uno de estos productos de mezcla, digamos la frecuencia diferencia, e ignora los demás, incluyendo sus armónicos, que también están presentes.

Los armónicos de una señal son señales cuya frecuencia es un múltiplo entero de la fundamental. Por ejemplo, los primeros armónicos de una señal de 1 MHz son 2 MHz (2^{do} armónico), 3 MHz (3^{er} armónico), 4 MHz (4º armónico), etc.

Desde otro punto de vista, un mezclador es, básicamente, un amplificador no lineal. En este tipo de circuitos, el punto de trabajo está cambiando permanentemente de posición y, en consecuencia, la señal de salida no es una réplica de la señal de entrada (figura 328).

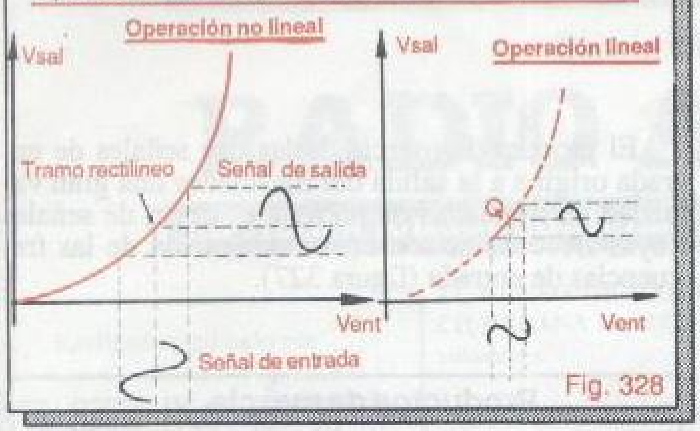

Operación lineal y no lineal de un amplificador

Fig. 328

La posición del punto de trabajo de un mezclador la controla la señal proveniente del oscilador local. Esta señal es relativamente grande comparada con la señal de entrada. Sin señal de control, el circuito opera como un amplificador lineal, es decir, de clase A.

En nuestro caso, la señal de entrada es f1 y la de control es f2. Sin f2, la señal de salida sería una réplica de f1 y no contendría armónicos ni otros productos de mezcla. Algo similar sucedería si f2 fuera muy pequeña.

En conclusión, cuando dos señales de diferente frecuencia se aplican a un amplificador no lineal, la señal de salida de este último contiene los armónicos de las señales de entrada y otras nuevas frecuencias derivadas de la mezcla.

Entre estas nuevas frecuencias o productos de mezcla están la frecuencia suma y la frecuencia diferencia.

Mezcladores de transmisión

En un transmisor superheterodino, el mezclador o modulador recibe la señal proveniente del amplificador de audio, la combina con la señal del oscilador local y produce a su salida una señal modulada de RF, la cual se aplica a un multiplicador que determina la frecuencia de transmisión (figura 329).

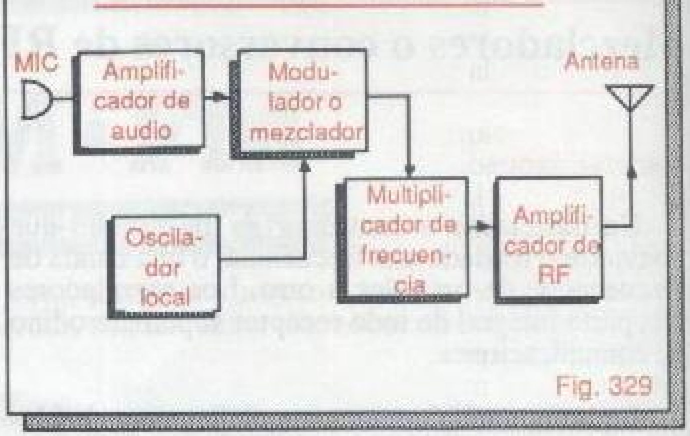

Transmisor superheterodino típico

Fig. 329

La señal modulada de RF conserva la modulación original de la señal del amplificador de audio.

Exceptuando los niveles de señal involucrados, los mezcladores de transmisión operan de manera similar a los utilizados en recepción. En la figura 330 se muestra el circuito de un mezclador de transmisión para VHF que utiliza un tubo de vacío como elemento activo.

La señal de entrada, proveniente de la última etapa de FI, se aplica a la rejilla de control. La señal interna, proveniente del oscilador local, se inyecta al cátodo.

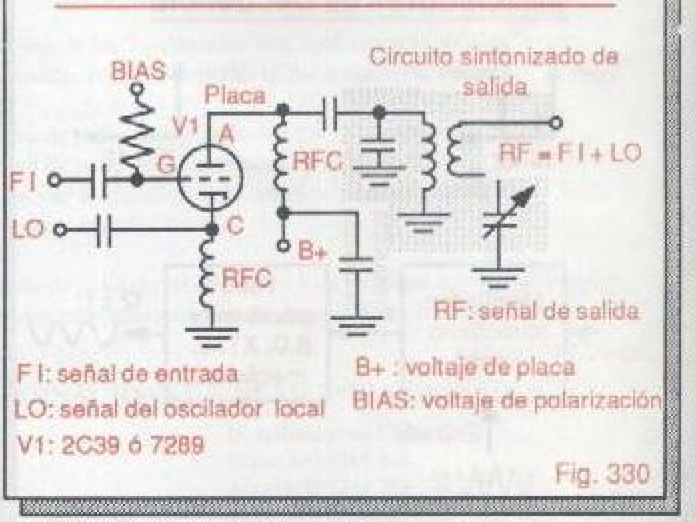

Mezclador de transmisión a tubos para VHF

RF: señal de salida
FI: señal de entrada
LO: señal del oscilador local
V1: 2C39 ó 7289
B+ : voltaje de placa
BIAS: voltaje de polarización

Fig. 330

El tubo realiza la mezcla de las dos señales y suministra en la placa todos los productos de mezcla posibles, incluyendo las frecuencias originales, su suma, su diferencia y los armónicos de todas ellas.

El circuito sintonizado de salida selecciona la señal correspondiente a la suma de frecuencias y rechaza las demás. La señal de RF resultante se amplifica hasta alcanzar un nivel específico de potencia y luego se irradia al espacio.

Los mezcladores a tubos, en general, manejan mayores niveles de potencia que los mezcladores a transistores y reducen el número de etapas de amplificación necesarias para alcanzar una potencia de transmisión específica.

Esta es su principal, y única, ventaja. Su mayor desventaja son los altos niveles de potencia requeridos a la entrada.

Un circuito como el de la figura 330 puede entregar, directamente, hasta 15 W de potencia de salida pero, en contraste, necesita cerca de 100 W de potencia de entrada para operar eficientemente.

La mayor parte de esta potencia de entrada proviene del oscilador local. La relación entre la potencia de salida y la potencia de entrada se denomina *ganancia de conversión*.

Mezcladores de recepción

En un receptor superheterodino, el mezclador recibe la señal modulada de RF, la combina con la señal del oscilador local y produce a su salida la señal de frecuencia intermedia (FI) (figura 331).

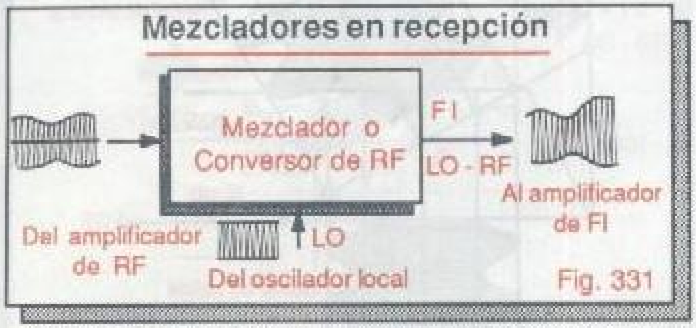

Fig. 331

Esta última conserva la modulación original de la portadora de RF y su frecuencia es generalmente igual a la diferencia de frecuencia de las dos señales de entrada. Es decir, FI = LO - RF.

En otras palabras, el mezclador de un receptor traslada la modulación original de la portadora de RF a una portadora local o de FI. La frecuencia de esta portadora interna es siempre la misma, sin importar la frecuencia de la portadora captada en el circuito de entrada.

En AM, la portadora puede tener cualquier valor entre 530 y 1600 KHz. Sin embargo, la FI tiene siempre un valor de 455 KHz.

Al ser la señal de FI de frecuencia más baja que la de RF, puede ser amplificada más fácil y eficientemente.

En la figura 332 se muestran las etapas de entrada de un receptor típico de AM.

La señal modulada de RF, captada en el circuito de antena, se inyecta amplificada a una de las entradas del mezclador o conversor de RF.

La otra entrada recibe la señal proveniente del oscilador local (LO).El mezclador combina las dos señales y produce la señal modulada de FI.

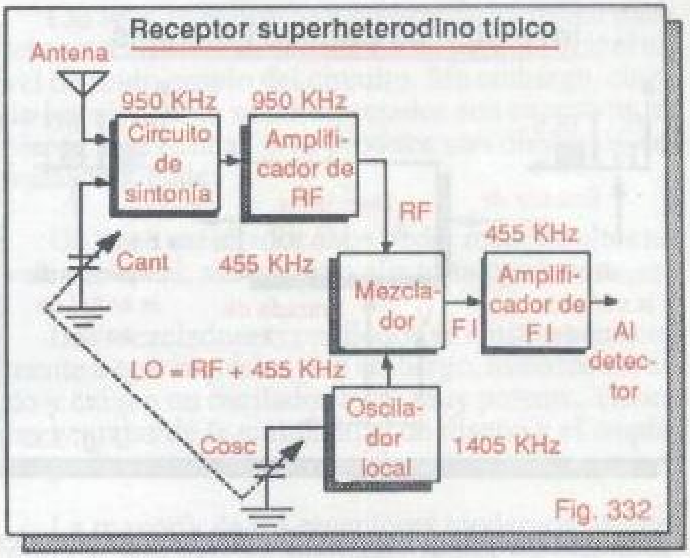

Fig. 332

El oscilador local se diseña para trabajar a una frecuencia que es siempre 455 KHz más alta que la frecuencia de portadora de la señal captada en el circuito de antena.

Para lograr este sincronismo, la frecuencia de resonancia del circuito de sintonía y la frecuencia de operación del oscilador local deben poderse variar simultáneamente.

Esto se consigue acoplando los condensadores variables de control de ambos circuitos en "tandem", es decir, sobre el mismo eje.

Así, al variar la frecuencia de resonancia del circuito de antena varía también la del oscilador local, y en la misma cantidad. Las líneas punteadas en el circuito de la figura 332 evidencian esta dependencia.

El proceso de mezclar, por ejemplo, una señal de 950 KHz proveniente del amplificador de RF con una de 1405 KHz proveniente del oscilador local, origina, además de las frecuencias de 950 y 1405 KHz, la frecuencia suma de 2355 KHz, la frecuencia diferencia de 455 KHz y todos los armónicos de estas 4 señales.

El mezclador entrega solamente la señal de 455 KHz, incluyendo sus bandas laterales, por poseer a su salida un tanque resonante sintonizado a esta frecuencia. Los demás productos de la mezcla son ignorados.

Una consecuencia interesante del proceso de mezcla es la inversión que experimentan las bandas laterales. Después de la conversión, la banda lateral inferior aparece como superior y viceversa.

En AM convencional, esta inversión no es importante para el posterior procesamiento de la señal

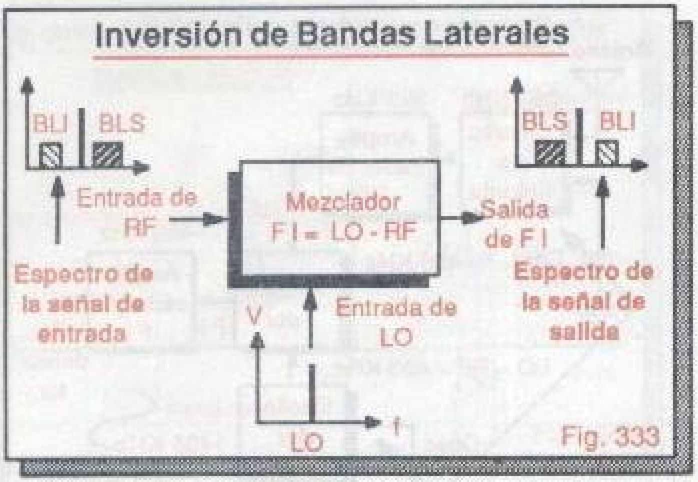
BLI
BLS
Entrada de RF
Mezclador
F I = LO - RF
Salida de F I
BLS
BLI
Espectro de la señal de entrada
V
Entrada de LO
Espectro de la señal de salida
f
LO

Inversión de Bandas Laterales

Fig. 333

por una razón muy sencilla: las dos bandas laterales contienen exactamente la misma información. En otros sistemas de comunicaciones (TV, SSB, etc.) sí es una consideración importante. La figura 333 ilustra gráficamente el fenómeno.

El conversor de RF del radio AM CEKIT

En la figura 334 se muestra el diagrama de bloques y el circuito completo del mezclador o conversor de RF del radio AM CEKIT.

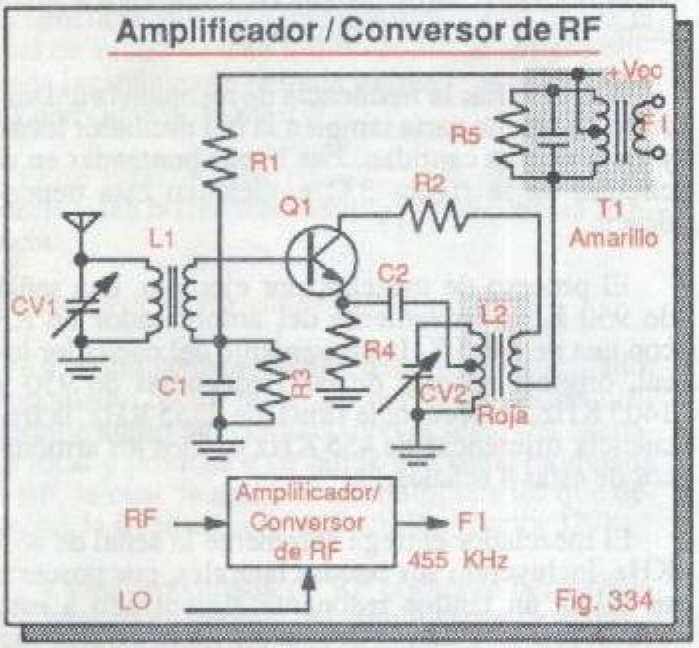
+Vcc
R5
F I
R1
R2
Q1
T1
Amarillo
L1
C2
CV1
L2
R4
C1
R3
CV2
Roja
RF
Amplificador/
Conversor
de RF
F I
455 KHz
LO

Amplificador / Conversor de RF

Fig. 334

La etapa recibe dos señales de entrada: una muy débil, proveniente del circuito de sintonía y otra muy fuerte, proveniente del oscilador local.

El elemento activo de conversión es el transistor Q1. Este dispositivo, además de mezclador, actúa también como oscilador y como amplificador de RF. La señal del oscilador local se inyecta al emisor y la de RF a la base. Los productos de la mezcla de las dos señales se obtienen en el colector.

El circuito de sintonía, constituido por el condensador CV1 y la bobina L1, selecciona una determinada señal de RF y la aplica a la entrada del mezclador.

Desde el punto de vista de esta señal de entrada, el transistor Q1 se comporta como un amplificador clase A, es decir, las variaciones de corriente de la base se suceden dentro de la parte lineal de su curva característica (figura 335).

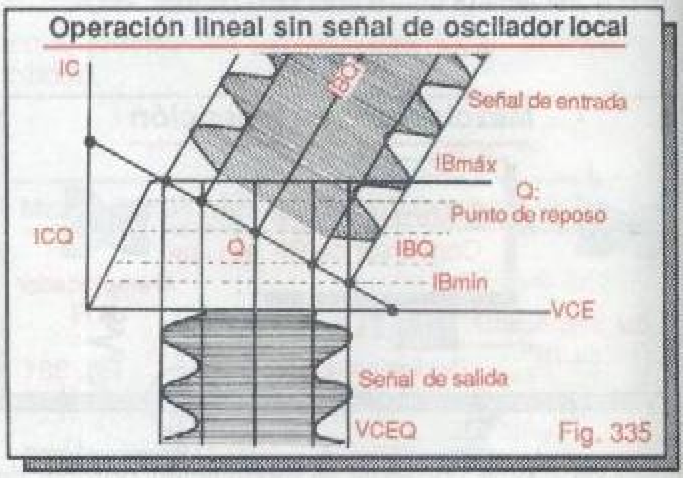

Operación lineal sin señal de oscilador local

Fig. 335

La operación de Q1 como amplificador de RF se describe con detalle en la Lección 24 de este curso. A pesar de la amplificación, la señal de RF sigue siendo muy débil comparada con la del oscilador local.

La señal del oscilador local, figura 336, debe ser espectralmente pura, sin armónicos, y relativamente grande. Esta última condición es importante para obligar el transistor Q1 a trabajar en la parte no lineal de sus curvas características. Como sabemos, la operación no lineal es básica para la mezcla.

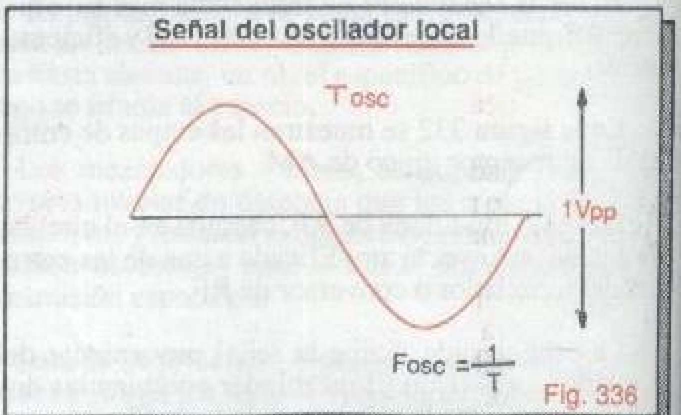

Señal del oscilador local

Fig. 336

Típicamente, la amplitud de la señal del oscilador local es del orden de 1 a 5 Vpp. Su frecuencia la determina el circuito tanque formado por la bobina L2 y el condensador CV2.

Como consecuencia de la mezcla de las dos señales, en el colector de Q1 obtendremos, además de las frecuencias originales de RF (radiofrecuencia) y de LO (oscilador local), la frecuencia suma (LO+RF) y la frecuencia diferencia (LO-RF). Esta última es la frecuencia intermedia (FI) de 455 KHz del proceso heterodino AM (figura 337).

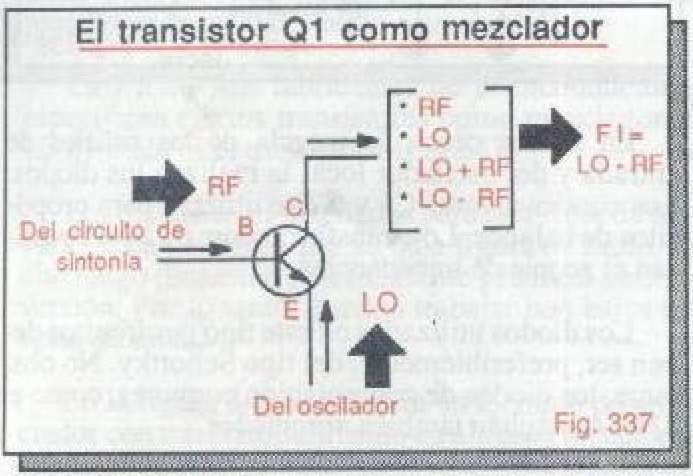
El transistor Q1 como mezclador

Fig. 337

En el colector de Q1 encontramos dos circuitos importantes: el primer transformador sintonizado de FI (T1, núcleo amarillo) y el primario de la bobina osciladora (L2, núcleo rojo).

El transformador T1 está sintonizado a 455 KHz. Su función consiste en seleccionar la señal de frecuencia intermedia (FI=LO-RF) derivada del proceso de mezcla y acoplarla a la primera etapa de FI. La señal de FI conserva la misma modulación de la portadora de RF captada en el circuito de antena.

La bobina osciladora L2 está sintonizada a la frecuencia local de oscilación (LO). Esta señal se reinyecta al emisor de Q1 con el fin de sostener las oscilaciones. La operación del oscilador local se explica con detalle en la Lección 25 de este curso.

Características generales de los mezcladores

El mezclador es una de las partes más importante de todo receptor de comunicaciones y el principal determinante de su rango dinámico.

El rango dinámico de un receptor se refiere a su habilidad para trabajar apropiadamente en presencia de señales débiles y fuertes, sin introducir distorsión ni ruido. La figura de ruido de un mezclador debe ser, idealmente, 0 dB.

Las señales aplicadas a un mezclador deben mantenerse a un nivel suficiente como para superar el nivel de ruido propio del circuito. Sin embargo, cuando los niveles de señal inyectados son excesivos, se pierde sensibilidad y se produce una distorsión demasiado grande.

Un buen mezclador debe poder manejar altos niveles de señal, sin que esto afecte su operación.

Los mezcladores con diodos se ajustan perfectamente a este requisito. Sin embargo, introducen ruido y exigen un oscilador local muy potente. Tienen las ventajas de la simplicidad de diseño y el amplio rango dinámico.

La mayoría de los receptores modernos utilizan FETs y transistores bipolares como mezcladores. Estos dispositivos son menos ruidosos que los diodos, proporcionan un buen rango dinámico y minimizan los requisitos de ganancia de las etapas subsiguientes.

Un buen mezclador también debe proveer un óptimo aislamiento entre el circuito de entrada y el oscilador local.

Un aislamiento deficiente puede provocar, por ejemplo, que la energía de RF del oscilador local se irradie al exterior, a través de la antena, causando interferencia en circuitos cercanos del mismo radio, o de otros aparatos.

En el caso de un mezclador con transistor bipolar, la señal de entrada, proveniente del amplificador de RF o del circuito de sintonía, se aplica generalmente a la base. La señal del oscilador local puede inyectarse a la base o al emisor.

La inyección por base exige menos niveles de señal pero no proporciona buen aislamiento. El mezclador del radio AM CEKIT utiliza inyección por emisor.

Para proveer un buen aislamiento entre la señal de entrada y la del oscilador local, resulta conveniente utilizar más de un dispositivo mezclador.

Este tipo de circuitos se denominan *mezcladores balanceados*. Los circuitos con un solo dispositivo mezclador se denominan *simplemente terminados*.

El balance evita que la energía inyectada en una de las entradas aparezca en la otra entrada o en la salida y se generen señales de mezcla indeseables. Esto es particularmente importante cuando se desea minimizar la distorsión de la señal de salida.

En un mezclador simplemente balanceado, únicamente el oscilador local está aislado. En un oscilador doblemente balanceado, todos los tres puertos del mezclador (entrada, oscilador local y salida) están aislados entre sí.

La mayoría de los receptores utilizan algún tipo de filtro pasabanda a la salida del mezclador, con el fin de establecer la selectividad general del sistema. Este filtro rechaza los productos de mezcla no deseados, garantizando que sólo pase la frecuencia intermedia.

Para garantizar el óptimo funcionamiento de un mezclador, la señal proveniente del oscilador local debe ser espectralmente pura, es decir, no debe contener ruido ni otras frecuencias distintas a la deseada.

Para prevenir esto, puede utilizarse un filtro pasabanda entre la salida del oscilador local y la entrada del mezclador (figura 338).

Filtrado de la señal del oscilador local

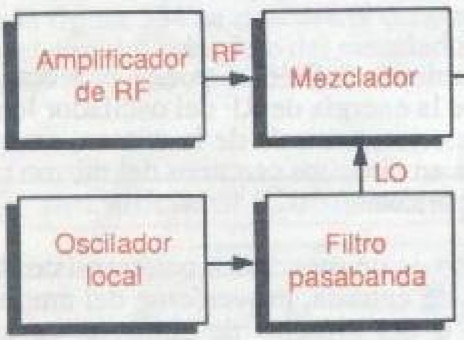

Fig. 338

Mezcladores con diodos

Los diodos resultan efectivos como mezcladores cuando se manejan señales fuertes. Los mezcladores con diodos ofrecen un amplio rango dinámico y son muy fáciles de implementar.

Sin embargo, este tipo de circuito introduce pérdidas y tiende a ser ruidoso. Para minimizar las pérdidas debe proveerse un buen acoplamiento de impedancias.

Los mezcladores con diodos son básicamente de dos tipos: simplemente balanceados y doblemente balanceados.

En la figura 339A se muestra un ejemplo de *mezclador simplemente balanceado* y en la figura 339B un ejemplo de *mezclador doblemente balanceado.*

Mezcladores con diodos

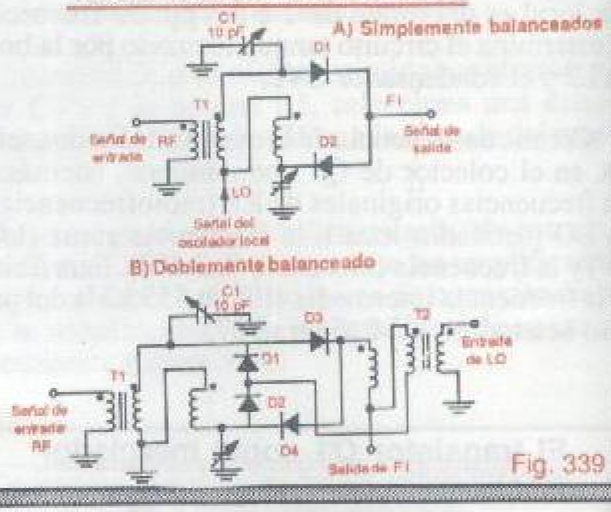

Fig. 339

En ambos casos, la mezcla de las señales de entrada y del oscilador local la realizan los diodos. Los condensadores C1 y C2 se utilizan para propósitos de balance. Los transformadores T1 y T2 realizan el acople de impedancias.

Los diodos utilizados en este tipo de circuitos deben ser, preferiblemente, del tipo Schottky. No obstante, los diodos de conmutación comunes, como el 1N914, resultan también apropiados.

Las razones por las cuales los diodos Schottky resultan efectivos como mezcladores de RF,. especialmente en aplicaciones de VHF y UHF, son, entre otras:

- Operación a muy altas frecuencias.
- Bajo voltaje de conducción.
- Bajo nivel de ruido.
- Alta eficiencia de conversión.
- Alto voltaje de ruptura.
- Baja corriente de fuga.
- Muy baja capacitancia interna.
- Buena tolerancia a las variaciones de temperaura.

Los transformadores de acoplamiento T1 y T2 ("baluns") se devanan, generalmente, sobre núcleos toroidales de ferrita. En la figura 340 se muestra su aspecto típico.

Mezcladores con transistores

Los transistores bipolares y los FETs se utilizan exitosamente como mezcladores en sustitución de los diodos Schottky o de conmutación cuando se manejan señales relativamente débiles.

En contraste con los diodos, que son elementos pasivos, los transistores, por ser activos, aseguran una ganancia de conversión. Esta característica mini-

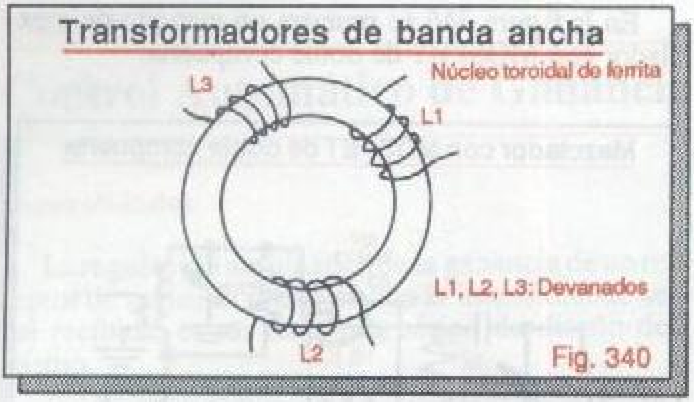
Transformadores de banda ancha
Núcleo toroidal de ferrita
L3
L1
L2
L1, L2, L3: Devanados

Fig. 340

miza el número de etapas de ganancia requeridas para llevar la señal de salida a un nivel dado.

Casi todos los fabricantes de semiconductores especifican ciertos transistores como mezcladores. Un ejemplo es el transistor MPS6547.

Los dispositivos diseñados para esta clase de servicio se caracterizan por su baja figura de ruido, su alto rango dinámico y su excelente ganancia de conversión. Por lo tanto, pueden trabajar con bajos niveles de señal.

En la figura 341 se muestra un ejemplo de mezclador con transistor utilizado en algunos receptores sencillos de AM.

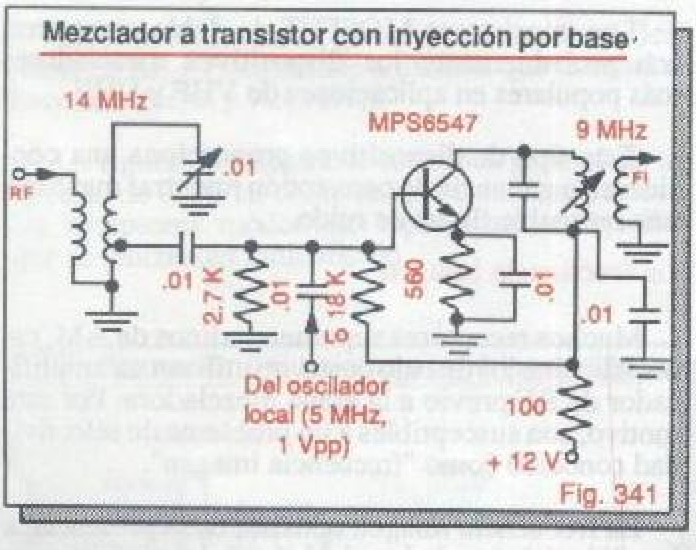
Mezclador a transistor con inyección por base
14 MHz
MPS6547
9 MHz
.01
RF
FI
.01
2.7 K
.01
18 K
LO
560
.01
.01
Del oscilador
local (5 MHz,
1 Vpp)
100
+ 12 V

Fig. 341

Para frecuencias por debajo de 7 MHz, este tipo de circuitos no requiere, generalmente, de amplificadores de RF previos. La señal del oscilador local se inyecta en la base.

La mayor limitación del circuito anterior es la falta de aislamiento efectivo entre la señal de entrada y la del oscilador local. El circuito de la figura 342 supera esta dificultad inyectando la señal del oscilador local al emisor.

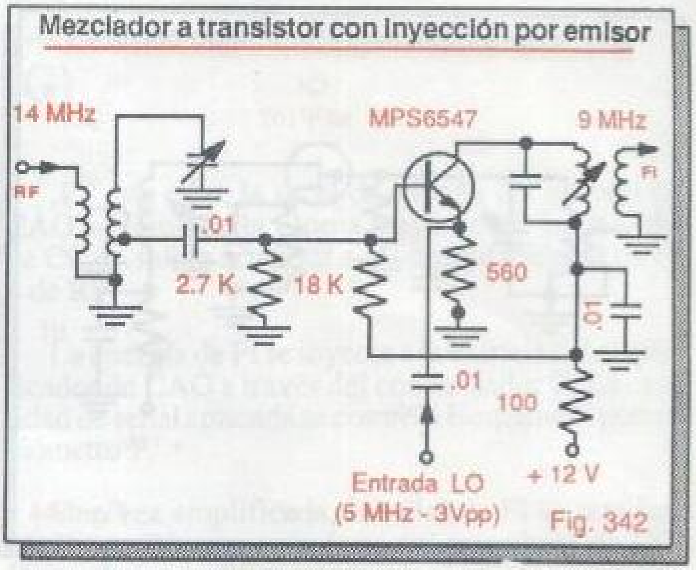
Mezclador a transistor con inyección por emisor
14 MHz
MPS6547
9 MHz
RF
FI
.01
2.7 K
18 K
560
.01
.01
100
Entrada LO
(5 MHz - 3Vpp)
+ 12 V

Fig. 342

Esta técnica requiere niveles ligeramente más altos de voltaje por parte del oscilador local, pero garantiza un mejor aislamiento con respecto a la entrada de RF.

En la figura 343 se muestra un ejemplo de mezclador simplemente balanceado con transistores. La señal del oscilador local se inyecta a la base. R1 se utiliza para efectos de balance.

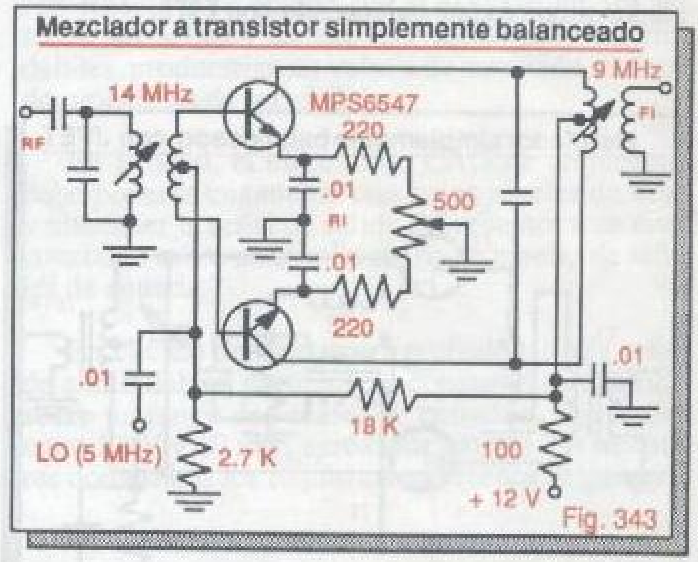
Mezclador a transistor simplemente balanceado
14 MHz
MPS6547
9 MHz
RF
FI
220
.01
R1
500
.01
220
.01
.01
18 K
LO (5 MHz)
2.7 K
100
+ 12 V

Fig. 343

Mezcladores con FET

Los mezcladores con FET se caracterizan por proporcionar un mejor rango dinámico que los mezcladores con diodos o con transistores.

En la figura 344 se muestra un ejemplo de mezclador simplemente terminado con FET.

Las señales de entrada y del oscilador local se aplican a la compuerta. El circuito se caracteriza por

Mezclador simplemente terminado con JFET

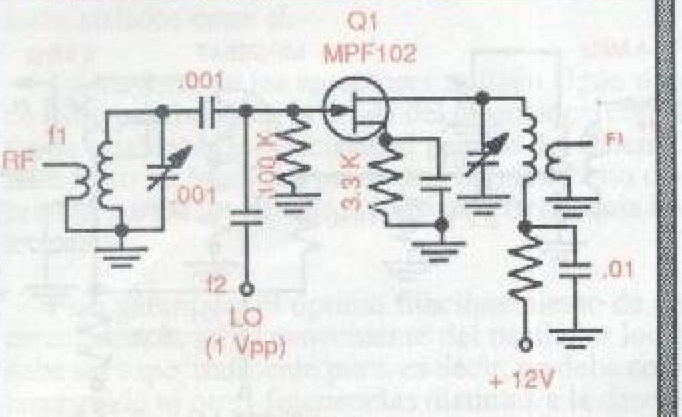

Fig. 344

su excelente ancho de banda. Su mayor desventaja es el aislamiento pobre entre la entrada de RF, la entrada del oscilador local y la salida de FI.

Para proveer un buen funcionamiento de este tipo de mezcladores, el voltaje de inyección del oscilador local debe ser del orden de 1Vpp. La ganancia de conversión (relación entre las señales de FI y de RF) es del orden de 10 dB.

En la figura 345 se muestra un ejemplo de mezclador simplemente balanceado con FET.

Mezclador simplemente balanceado con JFET

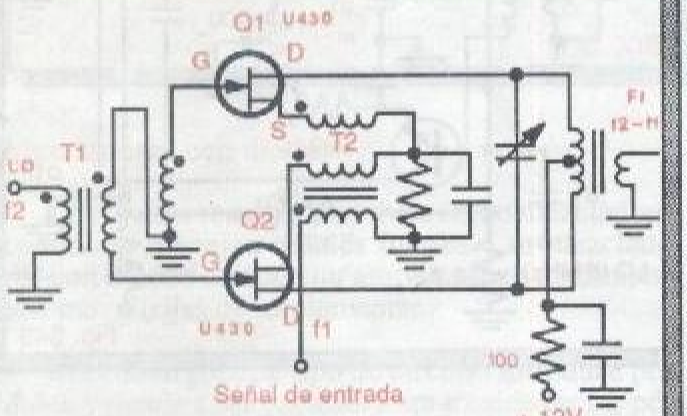

Fig. 345

El transformador de acoplamiento T1 suministra la señal del oscilador local a las compuertas (G) de Q1 y Q2.

La señal de entrada se inyecta a los surtidores de Q1 y Q2 a través del transformador T2. La salida de FI se obtiene de un circuito sintonizado.

En la figura 346 se muestra un ejemplo de mezclador con MOSFET de doble compuerta.

Mezclador con MOSFET de doble compuerta

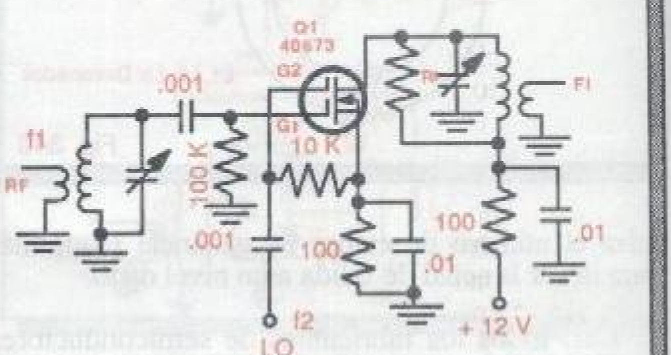

Fig. 346

La señal de entrada se aplica a la compuerta G1 y la del oscilador local a la compuerta G2. Esto garantiza un excelente aislamiento entre ambas señales.

La resistencia R1 se utiliza para disminuir la impedancia de carga y minimizar el efecto de las capacitancias internas del dispositivo. En los demás aspectos, el circuito es similar al de la figura 344.

Los transistores MOSFET de doble compuerta son probablemente los dispositivos mezcladores más populares en aplicaciones de VHF y UHF.

Este tipo de dispositivos proporciona una considerable ganancia de conversión mientras mantiene una razonable figura de ruido.

La frecuencia imagen

Muchos receptores superheterodinos de AM, especialmente los de bajo costo, no utilizan un amplificador de RF previo a la etapa mezcladora. Por este motivo, son susceptibles a un problema de selectividad conocido como "frecuencia imagen".

La frecuencia imagen consiste en la presencia, a la salida del mezclador, de dos señales simultáneas, provenientes de dos emisoras diferentes: la que se desea sintonizar y su imagen.Esta última tiene una frecuencia de portadora que está 455 KHz por encima de la frecuencia del oscilador local. Por ejemplo, si se sintoniza una estación de 700 KHz, su imagen estará localizada en 1610 KHz.

El problema de la frecuencia imagen se evita utilizando un amplificador de RF de banda estrecha o un circuito de sintonía muy selectivo.

Control Automático de Ganancia (CAG)

Generalidades

La regulación automática de la ganancia de un receptor de radio, en relación con la intensidad de señal recibida, es un requisito básico de diseño del mismo.

En otras palabras, el volumen de las emisoras en el parlante debe ser el mismo, independientemente de la intensidad de la señal con que cada emisora llega a la antena.

Esta función la realiza en el receptor un circuito especializado, llamado control automático de ganancia (CAG) o control automático de volumen (CAV).

El CAG mantiene constante el nivel de la señal de salida de audio del receptor, sin importar la intensidad de señal captada a la entrada.

El circuito de CAG utiliza el nivel de CC desarrollado por la señal de salida para variar la polarización de las etapas de RF y de FI.

Este voltaje es proporcional a la amplitud promedio de la señal.

En consecuencia, la ganancia de los amplificadores de RF y FI se reduce a medida que la señal se hace más fuerte y viceversa.

La figura 347 ilustra el concepto general de un circuito de CAG tal como se aplica en la mayoría de los receptores modernos, especialmente aquellos que se utilizan en radioafición.

Configuración de un sistema genérico de CAG

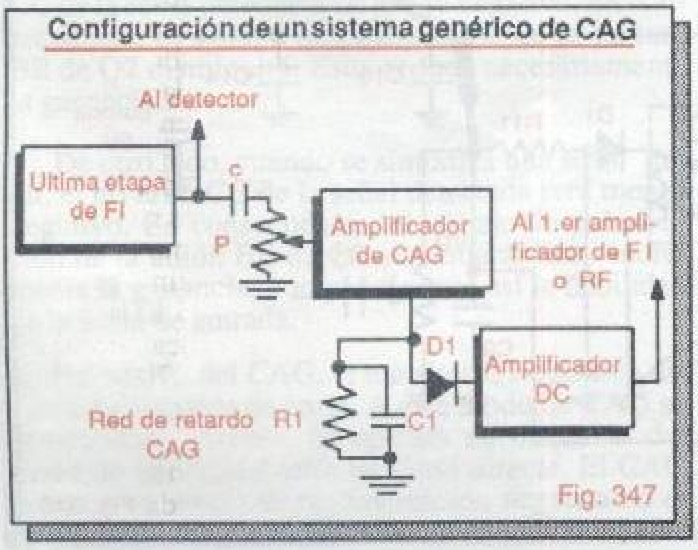

Fig. 347

En este caso, la señal de entrada del circuito de CAG se toma de la última etapa de FI. La tensión de CC de salida se aplica a los amplificadores de FI o de RF.

La energía de FI se inyecta a la entrada del amplificador de CAG a través del condensador C. La cantidad de señal aplicada se controla mediante el potenciómetro P.

Una vez amplificada, la señal de FI se rectifica mediante D1 y se convierte en un voltaje de CC. Este voltaje amplifica y se aplica al primer amplificador de FI y, en algunos casos, al amplificador de RF.

El circuito RC formado por R1 y C1 provee una función conocida como *retardo del CAG*. La función de este circuito de retardo es hacer que el CAG empiece a actuar únicamente después de que la señal de entrada alcanza un cierto nivel.

Las señales por debajo de este nivel recibirán plena amplificación y las que estén por encima quedarán sujetas a la acción del CAG. De no ser así, todas las señales de entrada, incluyendo las más débiles, producirian un voltaje de correción, causando problemas de sensibilidad.

En general, el circuito de CAG de un receptor debe poderse enganchar con bajos niveles de señal y mantener la señal de salida del receptor a un nivel constante sobre un amplio rango de niveles de señales de entrada.

En el caso de receptores profesionales la salida de audio deberá permanecer constante en amplitud sobre un rango de señales de entrada comprendido entre 1 µV y 10 mV, aproximadamente. En receptores domésticos los requisitos son menos exigentes.

El circuito de CAG del radio AM CEKIT

El propósito del control automático de ganancia (CAG) es mantener relativamente constante el nivel de señal de audio aplicado por el detector a la entrada del amplificador de audio, sin importar la intensidad de la señal de RF sintonizada en la etapa de RF.

La intensidad de la señal de RF captada en el circuito de antena varía con la potencia de la emisora, con su distancia del punto de recepción y con diversos factores atmosféricos y de propagación.

El CAG favorece la amplificación de las señales débiles y atenúa las fuertes. De este modo, todas se reproducen con la misma intensidad en el parlante.

Sin el CAG sería imposible una buena recepción porque la señal sintonizada no tendría siempre la misma fuerza (estaría subiendo y bajando de volumen). Todas estas variaciones son muy bien compensadas por el circuito de CAG.

La función de CAG en el radio AM CEKIT la provee la resistencia R10, la cual conecta la salida del detector con la entrada del primer amplificador de FI, tal como se muestra en la figura 348.

El CAG utiliza el nivel de CC desarrollado por la señal filtrada de FI a la salida del detector para controlar el nivel de polarización del transistor Q2.

Esta tensión se denomina *voltaje de CAG* y puede medirse fácilmente conectando un voltímetro de CC en bornes del condensador C9, como se muestra en la misma figura 348 anterior.

El voltaje de CAG es proporcional a la amplitud de la señal.

Por lo tanto, la ganancia del primer amplificador de FI se reduce a medida que la intensidad de la señal se hace más grande y viceversa.

En la figura 349 se comparan dos señales moduladas de RF de diferente intensidad y su efecto en diferentes partes del circuto.

Como puede verse, la señal filtrada presente en el ánodo del detector tiene un cierto nivel de CC.

Entre más fuerte sea la señal, más negativo es este voltaje.

El voltaje de polarización de la base del primer amplificador de FI (Q2) lo controla el divisor de tensión formado por R6 y R10.

La resistencia R11 y el condensador C9 del circuito de salida del detector forman un filtro que

Circuito de control automático de ganancia

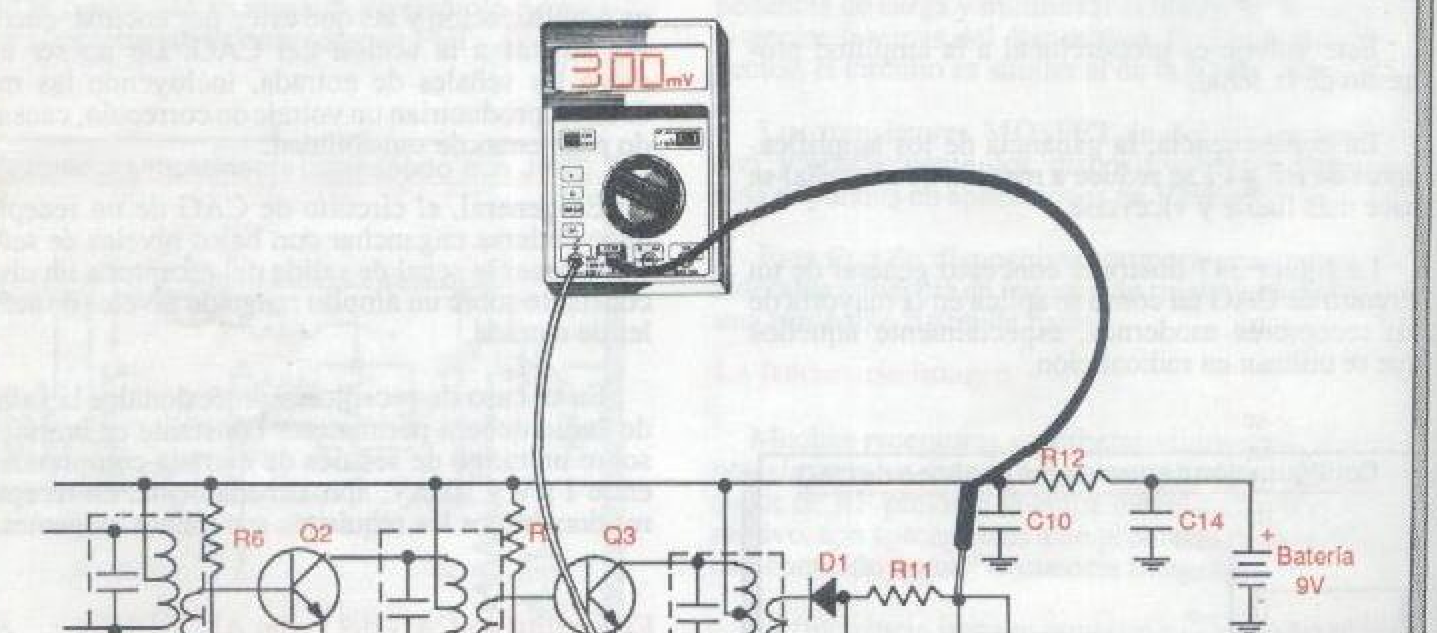

Fig. 348

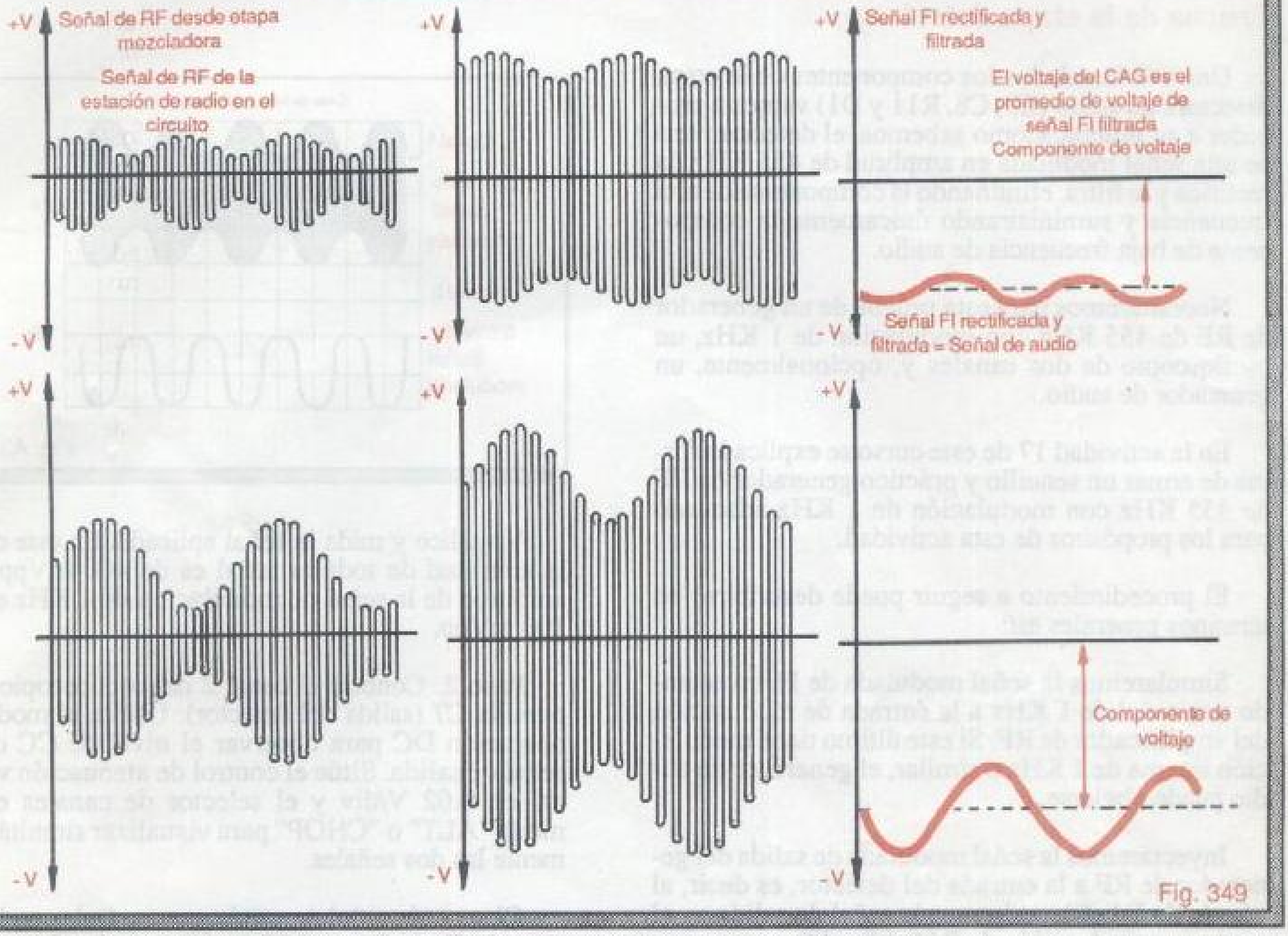

Acción del control automático de ganancia

+V
Señal de RF desde etapa mezcladora
Señal de RF de la estación de radio en el circuito
-V
+V
-V
+V
Señal FI rectificada y filtrada
El voltaje del CAG es el promedio de voltaje de señal FI filtrada
Componente de voltaje
Señal FI rectificada y filtrada = Señal de audio
-V
+V
-V
+V
-V
+V
Componente de voltaje
-V

Fig. 349

transfiere únicamente la componente de CC de la señal de salida. Este filtro contribuye también a la acción de retardo del CAG explicada al comienzo de esta lección.

Cuando se sintoniza una señal fuerte, el nivel de CC de la señal detectada será muy negativo. En consecuencia, el voltaje de polarización de la juntura BE de Q2 disminuirá. Esto reduce necesariamente la ganancia.

De otro lado, cuando se sintoniza una señal débil, el nivel de CC de la señal detectada será menos negativo. En consecuencia, el voltaje de polarización de la unión BE de Q2 aumentará. Esto incrementa la ganancia, compensándose así la debilidad de la señal de entrada.

Por acción del CAG, el transitor Q1 opera en dirección del punto de corte. A este modo de CAG se le denomina *inverso*. Sí operara en dirección del punto de saturación sería un CAG *directo*. El CAG provoca un efecto de realimentación negativa en el primer amplificador de FI.

Importancia del filtrado de la tensión de CAG

Son varias las razones por las cuales se requiere de una acción de filtrado o retardo de la tensión de CAG. La más importante es impedir que las variaciones instantáneas de la señal de audio se manifiesten a la entrada de los circuitos sometidos a la acción del CAG y provoquen la inestabilidad del punto de trabajo y otras perturbaciones.

En el sistema de la figura 347, el filtro de CAG está representado por R1 y C1. La principal función de este filtro es permitir que únicamente la componente de CC de la señal de salida del amplificador de CAG se transmita como tensión de polarización a la entrada de la primera etapa de FI.

El condensador C1, en particular, ofrece un camino fácil a tierra de las variaciones de alta y baja frecuencia de la señal de CAG. La constante de tiempo del filtro (R1xC1) se elige de modo que permita enmascarar los efectos del "fading" o desvanecimiento de la señal de entrada, sin ignorar la más baja de las frecuencias a reproducir.

Continuación Actividad Práctica Nº 14

Prueba de la etapa detectora

Una vez instalados los componentes de la etapa detectora de audio (C9, C8, R11 y D1) vamos a proceder a su prueba. Como sabemos, el detector recibe una señal modulada en amplitud de 455 KHz, la rectifica y la filtra, eliminando la componente de alta frecuencia y suministrando únicamente la componente de baja frecuencia de audio.

Necesitaremos para esta prueba de un generador de RF de 455 KHz con modulación de 1 KHz, un osciloscopio de dos canales y, opcionalmente, un generador de audio.

En la actividad 17 de este curso se explica la forma de armar un sencillo y práctico generador de RF de 455 KHz con modulación de 1 KHz adecuado para los propósitos de esta actividad.

El procedimiento a seguir puede describirse en términos generales así:

Simularemos la señal modulada de FI inyectando una señal de 1 KHz a la entrada de modulación del amplificador de RF. Si este último tiene modulación interna de 1 KHz o similar, el generador de audio puede obviarse.

Inyectaremos la señal modulada de salida del generador de RF a la entrada del detector, es decir, al cátodo de D1. Obtendremos la señal de salida en el punto común de unión de R11, C8 y C9.

Compararemos las dos señales y verificaremos que la forma de onda de la señal de audio observada en el canal 2 corresponde exactamente a la envolvente de la señal de RF observada en el canal 1.

Si esto es así, el detector de audio está operando correctamente. Variando la amplitud de la señal de RF podemos observar su efecto en el voltaje de CAG. Este voltaje es igual al nivel de CC de la señal de salida del detector.

Por cuestiones didácticas, las pruebas anteriores conviene realizarlas **sin batería** de alimentación. Si usted desea escuchar en el audífono la reproducción del tono de modulación de 1 KHz tendrá que conectar la batería. Notará que con la batería conectada la amplitud de la señal de audio a la salida del detector disminuye.

Paso 1. Obtenga con el generador de RF una señal modulada de 455 similar a la mostrada en la figura A37. Inyecte esta señal a la puntilla C1 (entrada del detector). Conecte a ese mismo punto el canal 1 del osciloscopio. Sitúe la base de tiempo en 0.5 µs/div, la atenuación vertical en 0.1 V/div y el selector de modo de operación en AC.

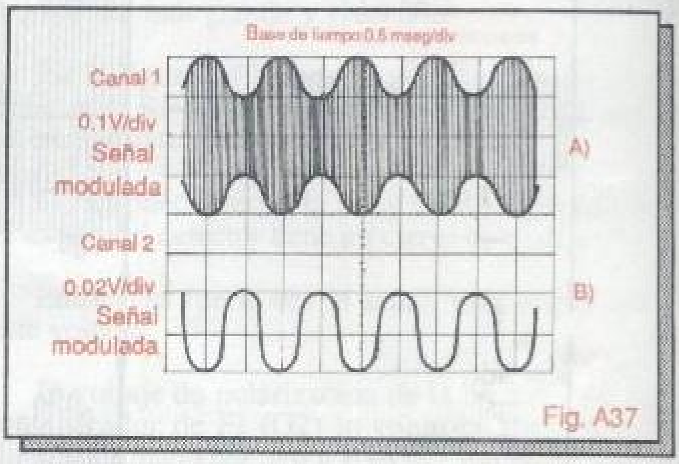

Fig. A37

Visualice y mida la señal aplicada. En este caso, la amplitud de toda la señal es de 400 mVpp. La amplitud de la señal de modulación de 1 KHz es de 100 mVpp.

Paso 2. Conecte el canal 2 del osciloscopio a la puntilla C7 (salida del detector). Utilice el modo de operación DC para observar el nivel de CC de la señal de salida. Sitúe el control de atenuación vertical en 0.02 V/div y el selector de canales en el modo "ALT" o "CHOP" para visualizar simultáneamente las dos señales.

Observe la señal de salida y compárela con la de entrada. En la figura A37 se muestran las formas de onda de las dos señales.

Notará que la señal de salida (canal 2) tiene una cierta amplitud, digamos 40 mVpp, y un cierto nivel negativo de CC, digamos 40 mV. Con respecto a la señal de entrada (canal 1), la forma de onda de la señal de salida es prácticamente igual a la forma de onda de la envolvente de la señal de entrada y no contiene ningún rastro de RF.

La envolvente corresponde a la información de audio transportada por la portadora de 455 KHz. En consecuencia, el detector ha extraído la señal de audio y eliminado la componente de RF. Así, hemos comprobado el proceso de detección AM.

Instalación del condensador filtro C10

Para terminar esta actividad, instale el condensador C10 de 47 µF entre las puntillas 39 (terminal positivo) y 30 (terminal negativo). La función de C10 es servir como filtro de radiofrecuencia de la fuente de alimentación. Después de soldado, corte el alambre restante.

Descripción general del receptor

Introducción

Hemos analizado hasta ahora todas las etapas de un radio AM típico, de derecha a izquierda, desde el parlante hasta la antena, empezando por el amplificador de audio y terminando en el circuito de sintonía. También se ha estudiado una gran cantidad de temas teóricos relacionados con cada etapa y con el campo de las comunicaciones en general.

Con el fin de aclarar las ideas principales, vamos a realizar a continuación un resumen en el sentido propio del flujo de las diferentes señales involucradas, de izquierda a derecha, es decir, desde la antena, donde llegan las señales de todas las emisoras, hasta el parlante, donde escuchamos la información de voz y sonido transmitida.

Con este resumen pretendemos que el lector o estudiante pueda visualizar, en una sola lección, cómo está estructurado un sistema de comunicaciones por radio desde el transmisor hasta el receptor y los circuitos especializados que realizan cada una de las funciones necesarias para el proceso.

Cómo se produce la señal AM. El transmisor

En la figura 350 se muestra la estructura simplificada de una emisora de AM.

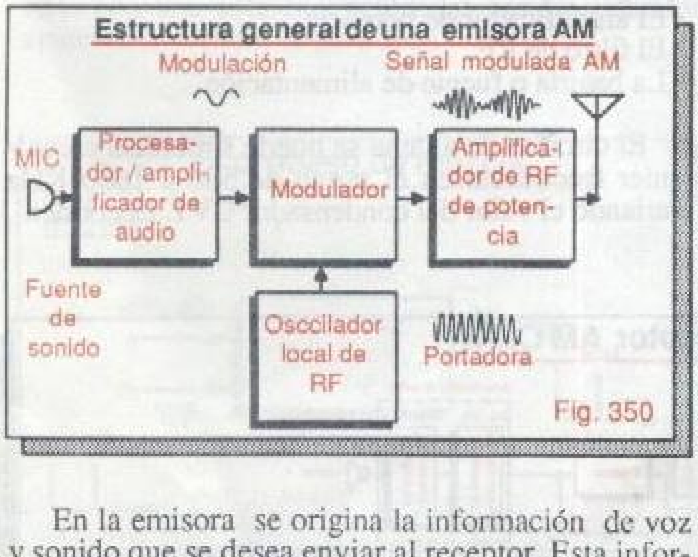
Estructura general de una emisora AM
Modulación
Señal modulada AM
MIC
Procesador / amplificador de audio
Modulador
Amplificador de RF de potencia
Fuente de sonido
Oscilador local de RF
Portadora

Fig. 350

En la emisora se origina la información de voz y sonido que se desea enviar al receptor. Esta información puede provenir de uno o más micrófonos o estar grabada en discos, cintas magnéticas u otros medios. Independientemente de su origen, debe convertirse en una representación eléctrica equivalente para ser procesada.

La conversión a señales eléctricas la realizan procesadores de sonido como tocadiscos, grabadoras de cassettes, reproductores de discos compactos, etc., asociados a amplificadores de audio. La señal eléctrica de salida del procesador de audio se aplica a un *modulador*.

El modulador combina la señal de baja frecuencia suministrada por el procesador de audio con una señal de alta frecuencia proveniente de un oscilador local y produce a su salida una señal de amplitud modulada (AM).

La señal de AM se inyecta a un amplificador de RF de potencia y se suministra amplificada a una antena, la cual la irradia al espacio. El enlace entre el amplificador de potencia de RF y la antena se establece a través de un cable llamado *línea de transmisión*.

Las emisoras de AM transmiten con frecuencias de portadora entre 530 y 1600 KHz. A cada estación se le asigna una frecuencia específica dentro de esta banda. La información de audio o *banda base* enviada contiene frecuencias entre 0 y 5 KHz. La banda base determina dos bandas laterales alrededor de la portadora (figura 351).

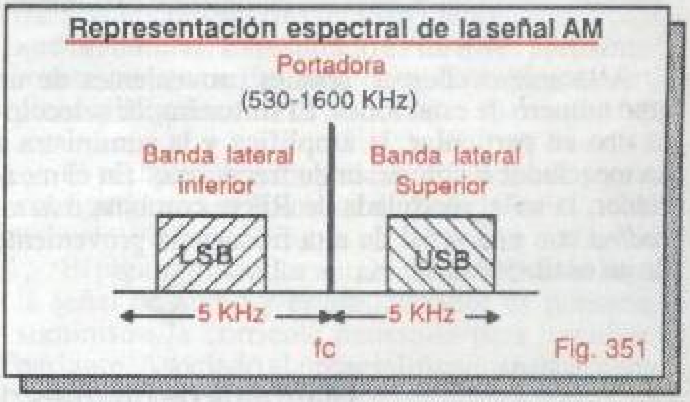

Fig. 351

La antena irradia la señal de AM en forma de una onda electromagnética. Esta onda se propaga en todas las direcciones y se dirige a la antena receptora siguiendo la curvatura de la tierra o reflejándose de ella (figura 352). Por esta razón, se dice que es una *onda de superficie* o terrestre.

Cómo se reproduce la señal AM. Recepción

En la figura 353 se muestra el diagrama de bloques simplificado de un receptor superheterodino

Propagación de la señal AM

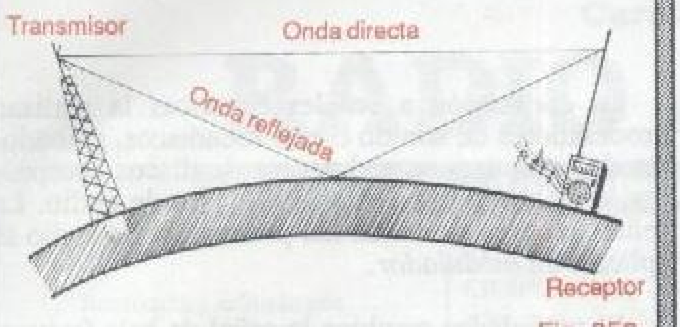
Transmisor
Onda directa
Onda reflejada
Receptor

Fig. 352

de AM. El receptor es el punto de llegada de la información originada en la emisora.

Estructura general de un receptor AM

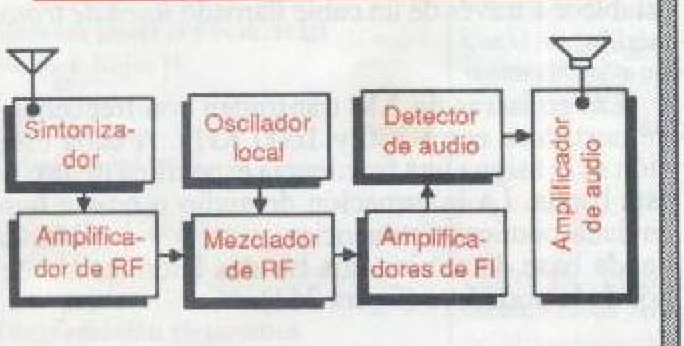
Sintonizador
Oscilador local
Detector de audio
Amplificador de audio
Amplificador de RF
Mezclador de RF
Amplificadores de FI

Fig. 353

A la antena llegan señales provenientes de un gran número de estaciones. El sintonizador selecciona una en particular, la amplifica y la suministra a un mezclador o conversor de frecuencia. En el mezclador, la señal modulada de RF se combina o *heterodina* con una señal de alta frecuencia proveniente de un oscilador local.

Como consecuencia del proceso de mezcla, la modulación original de la portadora se traslada a una portadora fija de 455 KHz llamada FI o *frecuencia intermedia*. La señal de FI conserva la misma modulación de la portadora de RF.

La señal de FI se amplifica y se inyecta al detector o demodulador. El detector separa la señal de modulación (audio) de la portadora de 455 KHz y la suministra al amplificador de audio.

Este amplificador eleva el nivel de la señal de audio y le suministra la suficiente potencia como para impulsar el parlante. En el parlante, la señal de audio se reproduce en forma de voces y sonidos que son réplicas fieles de la información audible enviada por la emisora sintonizada.

Los circuitos del receptor AM CEKIT

En la figura 354 se muestra el diagrama completo de bloques del receptor superheterodino AM CEKIT.

El sistema consta básicamente de las siguientes etapas:

- El circuito de sintonía.
- El amplificador/conversor de RF, constituido por el amplificador de RF, el oscilador local (LO) y el conversor o mezclador propiamente dicho.
- El primer amplificador de FI.
- El segundo amplificador de FI.
- El detector o demodulador de AM.
- El circuito de CAG.
- El preamplificador de audio.
- El amplificador de potencia.
- El filtro de RF.
- La batería o fuente de alimentación.

El circuito de antena se puede sintonizar a cualquier frecuencia en el rango de 530 a 1600 KHz variando el valor del condensador CV1. El conden-

Diagrama de bloques del receptor AM CEKIT

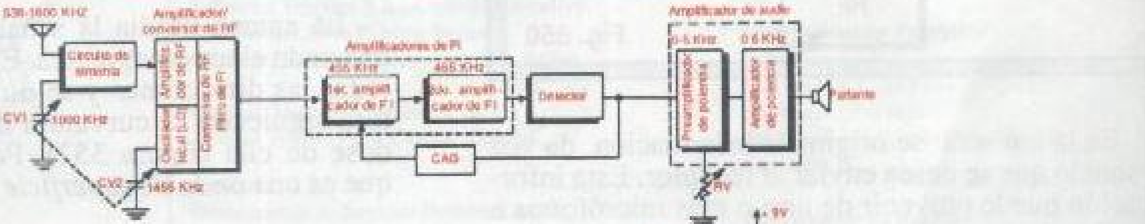
530-1600 KHz
Amplificador/conversor de RF
Circuito de sintonía
Amplificador de RF
Oscilador local (LO)
Conversor de RF
CV1
1600 KHz
CV2
1455 KHz
Amplificadores de FI
455 KHz
1er. amplificador de FI
455 KHz
2do. amplificador de FI
Detector
CAG
Amplificador de audio
0-5 KHz
0-5 KHz
Preamplificador de potencia
Amplificador de potencia
Parlante
RV
9V
7,8 V
Filtro de RF
Fuente de alimentación

Fig. 354

sador CV2 permite variar la frecuencia del oscilador local en el rango de 985 a 2055 KHz. Los dos condensadores están acoplados mecánicamente sobre el mismo eje.

Para simplificar, supongamos que el circuito de antena se sintoniza a 1000 KHz mediante CV1. El oscilador local operará entonces, automáticamente, a 1455 KHz, es decir, 455 KHz por arriba de la frecuencia de resonancia del circuito de sintonía.

A la antena convergen señales provenientes de un gran número de estaciones. El circuito de sintonía selecciona únicamente la correspondiente a una portadora de 1000 KHz, incluyendo sus bandas laterales. Las demás son ignoradas.

La señal de 1000 KHz captada ingresa entonces al amplificador/conversor de RF, donde se amplifica y se mezcla con la señal de 1455 KHz del oscilador local. A la salida de este circuito obtenemos 4 frecuencias diferentes de mezcla: 1000 KHz, 1455 KHz, 2455 KHz y 455 KHz.

Las dos primeras, 1000 KHz y 1455 KHz, corresponden, respectivamente, a las frecuencias originales de las señales de RF y del oscilador local. La señal de 2455 KHz resulta de la suma de ambas frecuencias (2455=1000+1455) y la de 455 KHz de su diferencia (455=1455-1000).

La salida del conversor de RF está sintonizada a 455 KHz, la frecuencia intermedia (FI) del proceso. En consecuencia, sólo esta señal pasa a la etapa siguiente. Los demás productos de la mezcla son ignorados. La señal de 455 KHz contiene exactamente la misma modulación de la portadora de 1000 KHz con una sola excepción: las bandas laterales aparecen invertidas (figura 355).

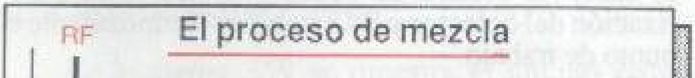

Fig. 355

La señal de salida del conversor se aplica al primer amplificador de FI. Este circuito está sintonizado también a 455 KHz y tiene un ancho de banda suficiente para dejar pasar las bandas laterales.

La señal proveniente del primer amplificador de FI se aplica a un segundo amplificador, también sintonizado a 455 KHz. Este la entrega al detector o demodulador de AM.

El detector rectifica la señal modulada de FI, separa la señal de modulación de la portadora de 455 KHz y la suministra al preamplificador de audio. Al desaparecer la portadora de FI, la señal de salida del detector es la banda base original (figura 356).

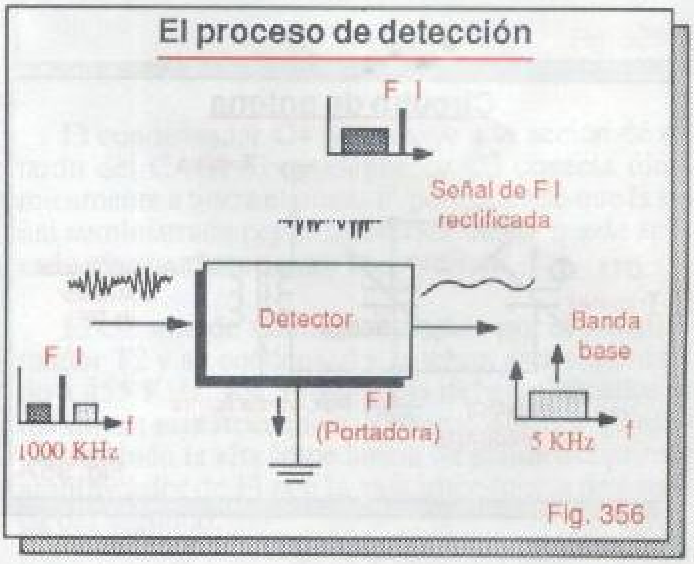

Fig. 356

La señal de salida del detector se aplica también al circuito de control automático de ganancia. El CAG regula la ganancia del primer amplificador de FI en función de la intensidad de la señal de entrada. Las señales fuertes reciben menos amplificación que las débiles. El resultado es un nivel aproximadamente constante de señal de audio a la entrada del preamplificador.

El preamplificador y el amplificador de potencia configuran el amplificador de audio.

El preamplificador aumenta el nivel de voltaje de la señal de audio y el amplificador de potencia le suministra la corriente necesaria para impulsar el parlante. Asociado al preamplificador está el control de volumen (RV).

El receptor obtiene la alimentación de CC necesaria para polarizar adecuadamente todos sus circuitos de una batería o fuente de poder de 9V. La alimentación del amplificador de audio se obtiene directamente de la batería y la de los circuitos de alta frecuencia a través de un filtro de RF.

El filtro de RF evita que las señales de alta frecuencia presentes en las líneas de alimentación interfieran con la operación del amplificador de audio e introduzcan ruido.

El circuito de antena

El circuito de antena capta las señales de radio emitidas por las diferentes estaciones, selecciona una en particular y la suministra al amplificador conversor de RF. La emisora deseada se elige mediante un condensador variable (CV1).

El circuito consiste básicamente de una antena colectora de ondas y un circuito resonante paralelo o tanque. En la figura 357 se muestra su diagrama esquemático.

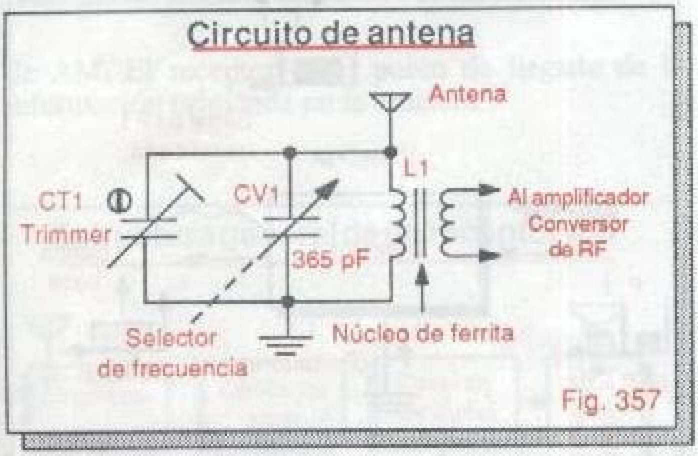
Circuito de antena
Antena
L1
CT1
Trimmer
CV1
365 pF
Al amplificador Conversor de RF
Selector de frecuencia
Núcleo de ferrita

Fig. 357

El circuito tanque lo integran un transformador de núcleo de ferrita (L1), un condensador variable (CV1) y un condensador trimmer o de compensación (CT1). El núcleo de ferrita actúa como antena propiamente dicha.

El condensador CV1 permite variar la frecuencia de resonancia del circuito dentro del rango de 530 a 1600 KHz. El trimmer CT1 permite realizar ajustes finos de frecuencia.

Además de contribuir a la determinación de la frecuencia de resonancia, L1 acopla la alta impedancia del tanque resonante con la baja impedancia de entrada del amplificador/conversor de RF.

El circuito tanque establece la *selectividad de entrada* del receptor, es decir, su habilidad para dejar pasar únicamente la portadora y sus bandas laterales y rechazar las demás señales. A la frecuencia de resonancia, el circuito tanque presenta su máxima respuesta de señal.

El amplificador / conversor de RF

El amplificador/conversor de RF recibe la señal modulada de RF captada en el circuito de antena, la amplifica, la mezcla con la señal del oscilador local y traslada la modulación original de la portadora de RF a una portadora fija de 455 KHz o FI.

La señal de salida se aplica a los amplificadores de FI. La frecuencia del oscilador local se controla mediante un condensador variable (CV2) y es siempre 455 KHz más alta que la frecuencia de resonancia del circuito de antena.

El circuito consiste básicamente de un amplificador sintonizado de RF, un oscilador local y un mezclador de frecuencias. En la figura 358 se muestra el diagrama correspondiente.

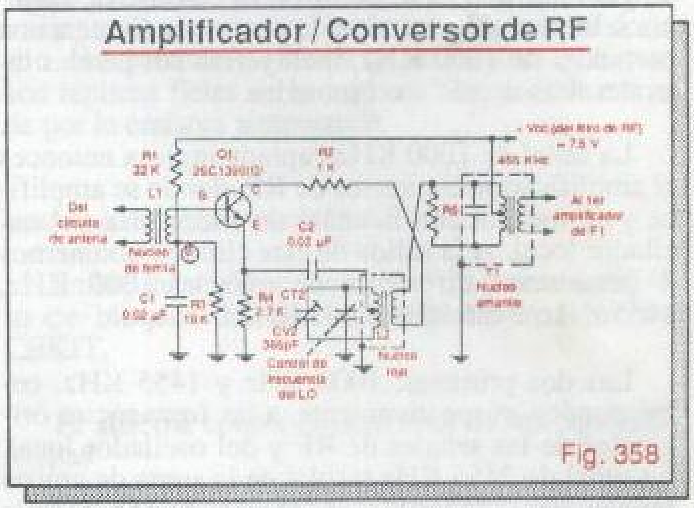
Amplificador / Conversor de RF
- Vcc (del filtro de RF) = 7.5 V
R1 22 K
Q1 2SC1390(O)
R2 1 K
L1
B
C
E
Del circuito de antena
Núcleo de ferrita
C2 0.02 µF
C1 0.02 µF
R3 10 K
R4 2.7 K
CT2
CV2 365pF
Control de frecuencia del LO
L2
Núcleo rojo
455 KHz
R5
Al 1er amplificador de FI
T1
Núcleo amarillo

Fig. 358

El elemento activo de la etapa es el transistor NPN Q1. Este dispositivo, en asocio con los demás componentes pasivos de la etapa (resistencias, condensadores y bobinas) amplifica la señal de RF suministrada por el circuito de antena, genera la señal local de oscilación y la mezcla con la de RF para producir la señal de FI.

Las resistencias R1, R2, R3 y R4 cumplen la función de polarizar adecuadamente el transistor Q1. En particular, R1 y R3 polarizan directamente la base, R2 suministra la tensión inversa de polarización del colector y R4 estabiliza térmicamente el punto de trabajo.

El condensador C2 evita que las variaciones de la señal de entrada aplicada a la base se manifiesten sobre la resistencia de emisor (R4), alterando la polarización de la etapa, e inyecta la señal local de oscilación al emisor, permitiendo que ésta sea la única señal que controla la posición del punto de trabajo durante el proceso de mezcla.

El circuito tanque constituido por el transformador L2, el condensador variable CV2 y el condensador de compensación o trimmer CT2 determinan la frecuencia local de oscilación.

En particular, CV2 proporciona una variación continua de frecuencia entre 985 KHz y 2055 KHz y CT2 permite realizar ajustes finos de frecuencia.

El transformador o bobina osciladora L2 suministra la señal de realimentación positiva necesaria para mantener las oscilaciones.

El condensador C1 conecta a tierra el punto B desde el punto de vista dinámico, es decir, en presencia de señal. De este modo, la señal de RF queda efectivamente aplicada entre base y tierra y la de oscilación local (LO) entre emisor y tierra.

El circuito tanque constituido por el transformador T1, su condensador interno y la resistencia R5 está sintonizado a 455 KHz. Su propósito es seleccionar la señal modulada de FI y transferirla a la etapa siguiente.

El transformador T1 acopla también la alta impedancia de salida del conversor con la baja impedancia de entrada del primer amplificador de FI. La resistencia R5 se utiliza para aumentar el ancho de banda del circuito tanque de salida.

El amplificador/conversor de RF determina la figura de ruido, la sensibilidad y el rango dinámico del receptor. Para que su operación sea eficiente, debe poder manejar señales de entrada fuertes y débiles con la misma facilidad y amplificarlas con el mínimo de ruido posible.

El circuito debe también producir una señal de oscilación local fuerte, espectralmente pura (sin distorsión) y aislarla del circuito de antena para que no se irradie al exterior.

El primer amplificador de FI

Los dos amplificadores de FI reciben la señal modulada de 455 KHz procedente del amplificador conversor de RF, la amplifican a un nivel conveniente y la entregan al detector o demodulador de audio. La ganancia del primer amplificador de FI la regula el circuito de CAG.

En la figura 359 se muestra el circuto esquemático del primer amplificador de FI. Se trata de un amplificador de RF clase A, de ganancia variable, sintonizado a 455 KHz.

El elemento activo de esta etapa es el transistor Q2. Las resistencias R6 y R7, el condensador C5 y la tensión de CAG establecen las condiciones de polarización de la etapa.

En particular, R6 y la tensión de CAG polarizan directamente la base, R7 estabiliza térmicamente el punto de trabajo y C5 impide que las variaciones de la señal de entrada afecten las condiciones de polarización. La tensión de polarización inversa del colector se obtiene del positivo de la fuente a través del primario de T2.

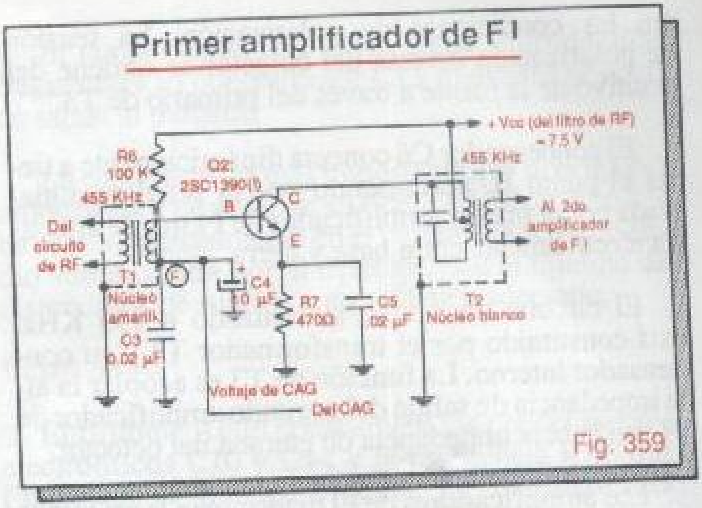
Primer amplificador de F I

Fig. 359

El condensador C4 contribuye a la acción de retardo del CAG. El condensador C3 conecta dinámicamente a tierra el punto F, permitiendo que la señal suministrada por el conversor de RF quede aplicada efectivamente entre base y tierra.

El circuito de salida, constituido por el transformador T2 y su condensador interno, está sintonizado a 455 KHz. Así, la ganancia del amplificador es máxima a esta frecuencia. El transformador T2 acopla también la alta impedancia de salida del primer amplificador de FI con la baja impedancia de entrada del segundo.

El segundo amplificador de FI

En la figura 360 se muestra el circuito esquemático del segundo amplificador de FI. Se trata también de un amplificador de RF clase A, de ganancia fija, sintonizado a 455 KHz.

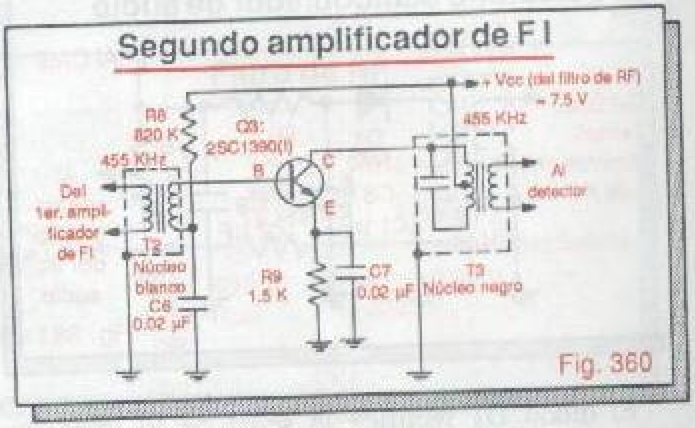
Segundo amplificador de F I

Fig. 360

El elemento activo de esta etapa es el transistor Q3. Las resistencias R8 y R9 y el condensador C7 establecen las condiciones de polarización.

En particular, R8 polariza en directo la base, R9 estabiliza térmicamente el punto de trabajo y C7 impide que las variaciones de la señal de entrada afecten las condiciones de polarización. La tensión de polarización inversa del colector se obtiene del positivo de la fuente a través del primario de T3.

El condensador C6 conecta dinámicamente a tierra el punto H, permitiendo que la señal suministrada por el primer amplificador de FI quede aplicada efectivamente entre base y tierra.

El circuito de salida, sintonizado a 455 KHz, está constituido por el transformador T3 y su condensador interno. La función de T3 es acoplar la alta impedancia de salida del segundo amplificador de FI con la baja impedancia de entrada del detector.

Los amplificadores de FI determinan la selectividad y la sensibilidad finales del receptor. El principal requisito que se les exige es entregar una señal de salida sin distorsión y lo suficientemente fuerte como para ser detectada.

El detector o demodulador

El detector recibe la señal proveniente del segundo amplificador de FI y la demodula, entregando únicamente la componente de audio y eliminando la portadora interna de 455 KHz. El detector provee también la tensión de CAG.

En la figura 361 se muestra el detector en detalle. La etapa consta básicamente de un rectificador de media onda (D1) y de una red filtradora compuesta por C8, R11 y C9. Este filtro pasabajos sigue las variaciones de baja frecuencia de la señal modulada de FI e ignora la portadora de 455 KHz.

Detector o demodulador de audio

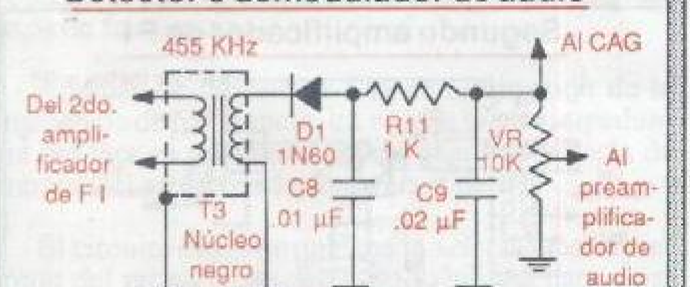

Fig. 361

El diodo D1 rectifica la señal de FI, dejando pasar únicamente la porción negativa y eliminando una de las bandas laterales. Los condensadores C8 y C9 envían a tierra la portadora de FI y transfieren únicamente la señal de audio al potenciómetro VR.

La característica más importante que se exige del detector es su linealidad: el circuito debe reproducir fielmente, sin distorsión, la señal de modulación.

El circuito de CAG

El circuito de CAG utiliza el nivel de CC de la señal de salida del detector para controlar la ganancia del primer amplificador de FI. De este modo garantiza que a la entrada del amplificador de audio se aplique siempre la misma cantidad de señal, sin importar la intensidad de la señal captada.

El circuito de CAG se muestra en la figura 362. El control de la ganancia del primer amplificador de FI se realiza a través de la resistencia R10. Entre más negativo sea el voltaje de CAG, más baja es la ganancia de este amplificador y viceversa.

Control automático de ganancia

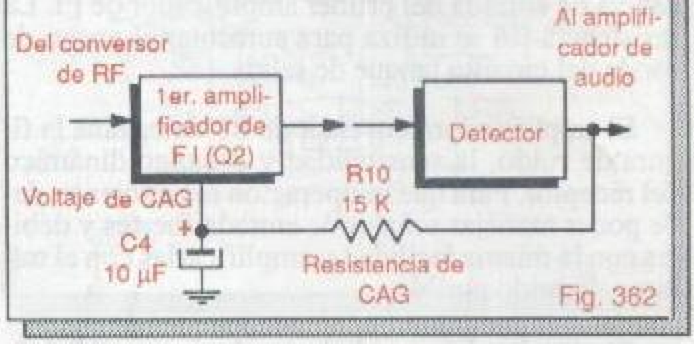

Fig. 362

El amplificador de audio

El amplificador de audio recibe la señal de salida del detector, la amplifica y la reproduce en el parlante. Consta de dos etapas: un preamplificador y un amplificador de salida o "driver".

El preamplificador es un amplificador de voltaje clase A realimentado negativamente y el driver un amplificador de potencia clase AB en configuración "push-pull" o complementaria. Asociado al preamplificador está el control de volumen. La etapa de salida maneja el parlante.

El preamplificador de audio

En la figura 363 se muestra el circuito del preamplificador. El elemento activo de esta etapa es el transistor Q4, conectado en la configuración emisor común. La carga es el amplificador de salida.

El potenciómetro R V actúa como control de volumen, regulando el nivel de la señal de audio suministrada por el detector al amplificador. La señal de audio se inyecta a la base mediante C11.

La resistencia R16 proporciona la realimentación negativa necesaria para minimizar la distorsión de cruce típica del esquema de amplificación "push-pull" utilizado en el amplificador de potencia.

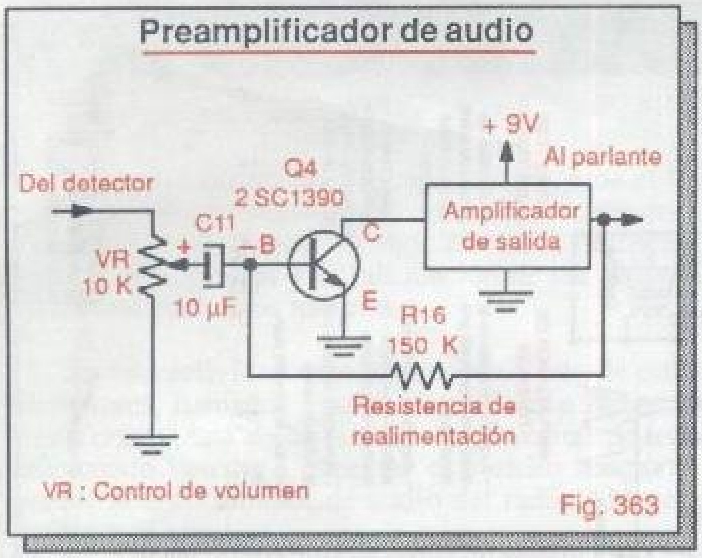
Preamplificador de audio
+ 9V
Al parlante
Q4
2 SC1390
Del detector
C11
Amplificador de salida
VR
10 K
+
– B
C
E
10 µF
R16
150 K
Resistencia de realimentación
VR : Control de volumen
Fig. 363

El amplificador de potencia

En la figura 364 se muestra el circuito del amplificador de potencia. Los elementos activos de esta etapa son los transistores complementarios Q5 y Q6, conectados en la configuración colector común.

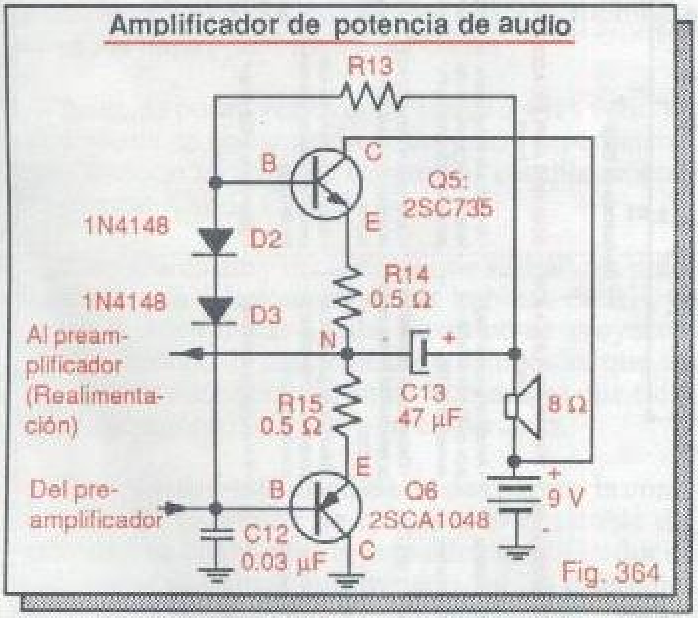
Amplificador de potencia de audio
R13
B
C
Q5:
2SC735
E
1N4148
D2
R14
0.5 Ω
1N4148
D3
Al preamplificador (Realimentación)
N
+
R15
0.5 Ω
C13
47 µF
8 Ω
E
+
Del preamplificador
B
Q6
2SCA1048
9 V
C12
0.03 µF
C
-
Fig. 364

El transistor Q5 suministra impulsos de corriente al parlante durante los ciclos positivos de la señal de entrada y Q6 durante los negativos. Entre los dos reproducen en su totalidad la señal de entrada. A este proceso se le denomina "push-pull".

La polarización directa de la base del transistor Q5 se obtiene del positivo de la batería a través de R13 y el parlante. El transistor Q6 la obtiene de la base de Q5 a través de los diodos D2 y D3. Las resistencias R14 y R15, situadas en los emisores, estabilizan térmicamente ambos transistores.

El condensador C12 actúa como filtro de altas frecuencias. El condensador C13 transfiere la señal de salida al parlante.

La característica más importante que se exige del amplificador de audio es su fidelidad. El circuito debe reproducir fielmente en el parlante la señal de audio suministrada por el detector, con el mínimo de distorsión y de ruido y a un volumen adecuado.

Filtro de RF

El filtro de RF lo constituyen los condensadores electrolíticos C10 y C14 y la resistencia R12. Sus funciones se pueden resumir en los siguientes términos:

- Suministrar la tensión de alimentación de las etapas de alta frecuencia (amplificador/conversor de RF y amplificadores de FI). Estas etapas se alimentan con un voltaje más bajo (≈7.5V) que el del amplificador de audio (9V).

- Aislar las líneas de alimentación de las etapas de RF y FI con respecto a las de la etapa amplificadora de audio. Así, los armónicos de alta frecuencia generados por la operación en clase AB de los transistores Q5 y Q6 no interfieren con la operación de las etapas de RF y de FI y las señales de alta frecuencia manejadas en estas etapas no interfieren con la operación del amplificador de audio.

El circuito del filtro de RF utilizado por el radio AM CEKIT se muestra en la figura 365. La forma como están interconectados los 3 componentes (R12, C10 y C14) se denomina "pi" porque recuerda la forma de esta letra griega (∏).

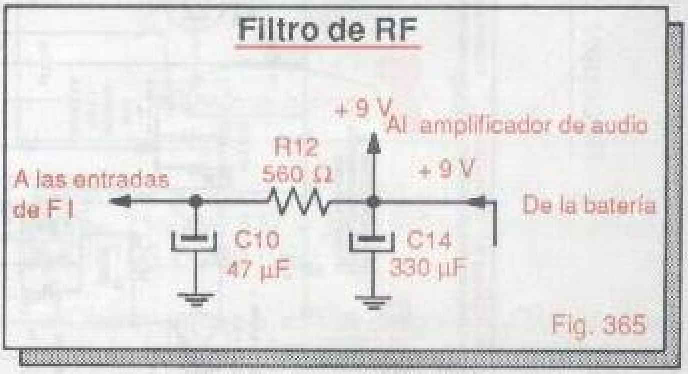
Filtro de RF
+ 9 V
Al amplificador de audio
R12
560 Ω
+ 9 V
A las entradas de F I
De la batería
C10
47 µF
C14
330 µF
Fig. 365

Diagrama esquemático general

A manera de resumen de lo explicado en esta lección, en la figura 366 se muestra el diagrama general del radio AM CEKIT. Se han delimitado todas las etapas que lo conforman, sintetizando su función específica e ilustrando pictóricamente el flujo de las diferentes señales involucradas.

Receptor Superheterodino AM CEKIT

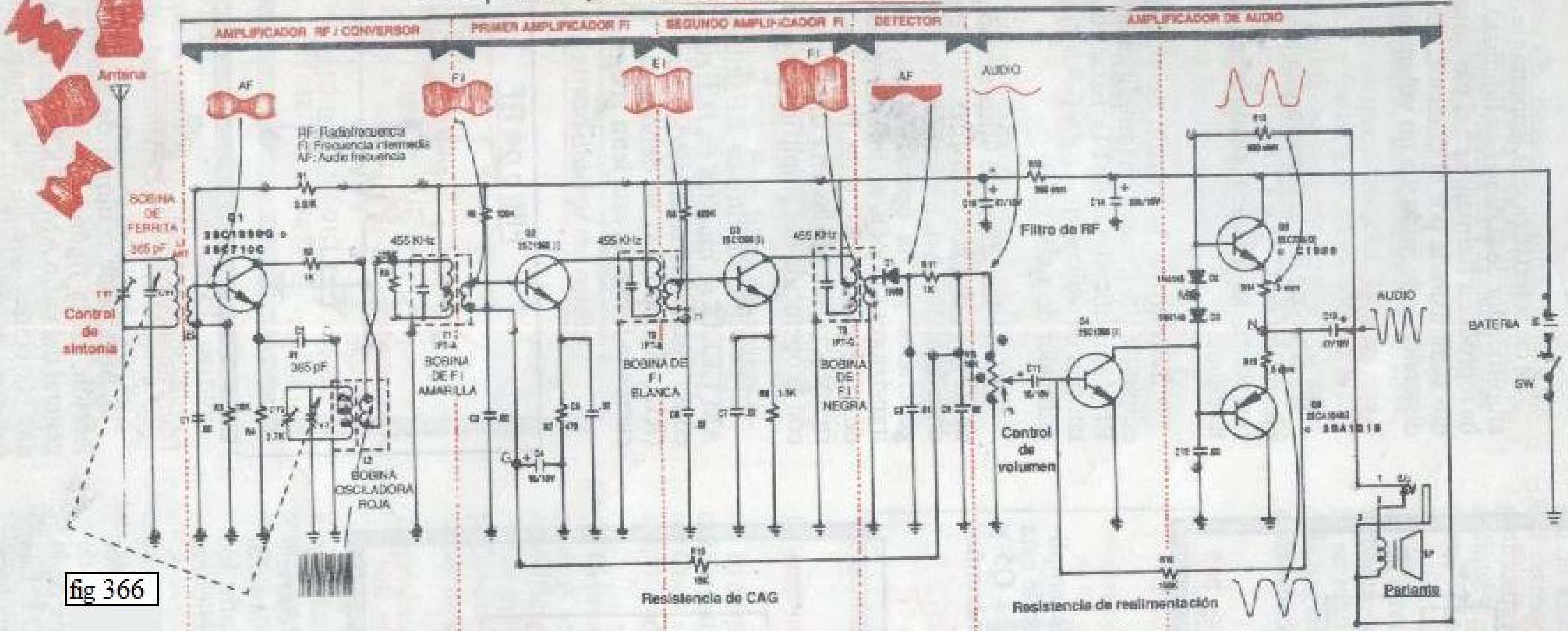
AMPLIFICADOR RF / CONVERSOR
PRIMER AMPLIFICADOR FI
SEGUNDO AMPLIFICADOR FI
DETECTOR
AMPLIFICADOR DE AUDIO
Antena
AF
FI
AF
AUDIO
RF: Radiofrecuencia
FI: Frecuencia intermedia
AF: Audio frecuencia
BOBINA DE FERRITA
365 pF
Control de sintonía
Q1
2SC1390G ó 2SC710C
455 KHz
455 KHz
455 KHz
365 pF
BOBINA DE FI AMARILLA
BOBINA DE FI BLANCA
BOBINA DE FI NEGRA
L2 BOBINA OSCILADORA ROJA
Filtro de RF
Control de volumen
AUDIO
BATERIA
SW
Parlante
Resistencia de CAG
Resistencia de realimentación

fig 366

CIRCUITO DE ANTENA

- Selecciona una emisora en particular.
- Determina la selectividad de receptor.
- Rechaza la frecuencia imagen.
- Sintonizable entre 530 y 1600 KHz.

AMPLIFICADOR / CONVERSOR DE RF

- Amplifica la señal de RF captada.
- Genera señal local de oscilación.
- Mezcla las señales de RF y de LO y produce la señal de FI.
- Invierte las bandas laterales.
- Determina el rango dinámico, la sensibilidad y la figura de ruido de entrada del receptor.
- Aísla la señal del oscilador del circuito de antena.

PRIMER AMPLIFICADOR DE FI

- Amplifica la señal modulada de FI proveniente del conversor de RF.
- Determina la sensibilidad y la selectividad finales del receptor.
- Modifica su ganancia de acuerdo a la intensidad de la señal captada.
- Sintonizado a 455 KHz.

SEGUNDO AMPLIFICADOR DE FI

- Amplifica la señal modulada de FI proveniente del primer amplificador de FI.
- Determina la sensibilidad y selectividad finales del receptor.
- Suministra una señal suficientemente fuerte al detector de audio.
- Sintonizado a 455 KHz.

DETECTOR DE AUDIO

- Separa la señal de audio de la FI.
- Envía a tierra la portadora interna de 455 KHz.
- Entrega la señal de audio al preamplificador.
- Suministra el voltaje de CAG del primer amplificador de FI.

PREAMPLIFICADOR DE AUDIO

- Amplifica en voltaje la señal de audio proveniente del detector.
- Regula la cantidad de señal de audio aplicada al amplificador de potencia.
- Determina la fidelidad del sonido reproducido en el parlante.

AMPLIFICADOR DE POTENCIA

- Amplifica en corriente la señal proveniente del preamplificador.
- Reproduce la señal de audio en el parlante en forma de ondas sonoras.
- Minimiza la distorsión de la señal de salida.
- Utiliza el principio push-pull.

FUENTE DE ALIMENTACION

- Suministra los voltajes de polarización de todas las etapas.
- Determina la máxima potencia de salida del amplificador de audio.

Actividad práctica Nº 15

Armando un radio de galena

En la lección 20 de este curso estudiamos el tema de los receptores de radio en su aspecto general. Además se mostraron algunos circuitos de receptores muy sencillos y la fabricación de las bobinas de antena para estos receptores.

En esta actividad vamos a construir uno de estos receptores, llamados *Radios de Galena o Receptores a cristal*, una de las actividades favoritas de todo aficionado.Vamos a conectar este radio temporalmente al amplificador de audio del radio AM para reforzar su señal.

Una vez que se haya terminado esta actividad, se desconecta otra vez este circuito y continuamos con las otras actividades hasta terminar completamente el radio que se ensambla en este curso.

Es de mucha utilidad para el experimentador o el estudiante ensamblar este tipo de radio, ya que puede verificar y comprobar las lecciones estudiadas hasta el momento, lo mismo que reforzar los conocimientos teóricos adquiridos hasta el momento sobre los temas principales del curso.

Además podrá verificar la utilidad y las ventajas del sistema de transmisión y recepción superheterodino, cuando termine de ensamblar completamente el receptor AM de CEKIT.

También es muy importante que se realicen prácticas sobre la fabricación de las bobinas de RF, ya que se pueden llevar a cabo otro tipo de proyectos más avanzados en transmisión y recepción, que involucran la construcción de las bobinas ya que éstas no se consiguen fácilmente en el mercado.

La actividad está dividida en dos partes: la construcción de la bobina de antena y el ensamble del radio con su conexión a la etapa del amplificador de audio, que ya hemos ensamblado, en las diferentes actividades prácticas que hemos efectuado.

Antes de tratar el tema del ensamble del receptor vamos a estudiar cómo se diseñan y fabrican bobinas de RF y daremos todas las instrucciones necesarias para la construcción de la bobina de antena para esta actividad.

Parte 1

Diseño y construcción de bobinas de RF

Los experimentadores de radio que desean construir aparatos electrónicos especiales que operan en varias gamas de frecuencia, a menudo necesitan diseñar sus propias bobinas. Vamos a tratar dos métodos: un método matemático sencillo y un método gráfico muy rápido.

El sencillo método matemático ayudará al experimentador a construir sus propias bobinas de RF y a determinar la inductancia de bobinas disponibles cuyo valor se desconoce.

El problema se reduce a esto: para la operación a un frecuencia dada, ¿ qué tamaño de forma para bobina, alambre y longitud de devanado se necesitan, y cuántas vueltas ha de tener la bobina ?.

El diseño de una bobina con núcleo de aire de una inductancia dada es relativamente fácil. Y si se conoce la gama de frecuencia que se va a cubrir y el condensador de sintonía que se va a emplear, la determinación de la inductancia requerida es más fácil aún.

Con este método los diseños resultantes no serán de alta precisión, pero sí adecuados para propósitos experimentales. Aunque está orientado hacia el diseño de bobinas, sólo se necesita invertir el procedimiento para determinar las características de las bobinas ya existentes.

Cómo determinar la inductancia

Supongamos que usted desea construir una bobina para la banda de radiodifusión común, y asumamos que emplea un condensador variable o de sintonía de 365 pF y la frecuencia más baja que debe sintonizar es de 540 KHz.

La inductancia L de la bobina en microhenrios (µH) se determina usando la fórmula

$$L = \frac{25400}{f^2\,C}$$

C representa pF y f, MHz

En este problema C = 365 y f = 0.54. Por lo tanto, $L = 25400/0.54^2 \times 365) = 25400/(0.29 \times 365) = 25400/106$, ó 239 µH.

Observe que se usó el extremo de baja frecuencia de la banda en este cálculo. Para determinar el extremo de alta frecuencia de la banda que usted espera cubrir con la bobina de 239 µH, asuma que la capacidad mínima del condensador de sintonización y la capacitancia parásita del circuito es de 30 pF. La fórmula aplicable es:

$$f = \frac{159}{\sqrt{LC}}$$

En este caso, f = 159 / $\sqrt{(239 \times 30)}$ = 1.88 MHz o 1800 Khz. Así esta combinación cubre fácilmente la banda de radiodifusión común y se puede extender el límite de baja frecuencia para asegurarse de que la cubre adecuadamente.

Una fórmula simplificada

Para el diseño de bobinas de RF, con una precisión hasta de 1 ó 2% aproximadamente, la fórmula es:

$$\boxed{n = \frac{\frac{1}{r}}{\sqrt{L\,(9r + 10\,l)}}}$$

En la que L es la inductancia en µH, n es el número de vueltas de la bobina, r es el radio de la bobina en pulgadas, y l es la longitud del devanado en pulgadas. Si se usa un diámetro de 1" (r=1/2") se simplifica la fórmula más aún a:

$$\boxed{n = 2\sqrt{(L\,(4.5 + 10\,l)}}$$

Ahora, aproximemos la inductancia requerida para la banda de radiodifusión común (con el condensador de 365 pF) a 240 µH y asumamos que la forma de la bobina es de 1" de diámetro.

Tenemos que asumir también una longitud de devanado, de forma que probaremos con 1 1/2". El número de vueltas que se necesitan es entonces:

$n = \sqrt{(240(4.5 + 10 \times 1.5))}$ Así,

$n = 2\sqrt{(240 \times 19.5)}$, ó

$n = 2\sqrt{(4680)}$, que son 137 vueltas.

El calibre del alambre que se emplee para devanar la bobina es opcional, siempre que el diámetro sea lo suficientemente pequeño para permitir que queden 137 vueltas en 1.5" de longitud de forma. El devanado es más fácil, por supuesto, si las vueltas se ajustan unas contra otras a lo largo de esta longitud de bobina.

El diámetro del alambre que llenará este requisito es l/n ó 1.5/137, que es .0109".

El alambre Nº 30 tiene .0108" de diámetro y es el más aproximado al diámetro calculado. Por consiguiente, se puede devanar la bobina con 137 vueltas juntas de alambre esmaltado Nº 30.

Método gráfico

Existen gráficos especiales para simplificar este procedimiento. En él hay líneas que tienen marcada la inductancia, el diámetro y el radio (figura A38).

Teniendo el dato de la inductancia requerida, en este caso 240 µH, y escogiendo el diámetro de la bobina según nuestra disponibilidad, hallamos el número de vueltas o espiras que nos da esta inductancia.

Construcción de la bobina

Con este dato, hallado con el método matemático o el método gráfico, procedemos a enrollar sobre un tubo de cartón, de plástico o PVC las vueltas de alambre necesarias. *Esta bobina será entonces una bobina de antena con núcleo de aire.*

En nuestro caso vamos a enrollar en forma apretada 140 vueltas con alambre esmaltado Nº 30 sobre una formaleta o forma de 1 pulgada (1") de diámetro.

Paso 1: Consiga un tubo de plástico o de PVC, de los que se utilizan para las instalaciones eléctricas, de 1" de diámetro y 12 centímetros de longitud. Si no consigue este tubo, puede fabricar la formaleta de cartón delgado enrollándolo sobre algún tubo dándole varias vueltas y pegando las capas entre sí.

Cuando esté firme, retire la base y le quedará el tubo de cartón listo para enrollar la bobina.

Paso 2: Asegure un extremo del alambre en un sitio cercano al borde de la formaleta para la bobina y empiece a enrollar las vueltas o espiras de alambre de tal manera que queden muy juntas entre sí pero sin montarse una sobre otra. Cuando termine las 140 vueltas, asegure el terminal final como se muestra en la figura A39.

Al principio y al final deje unos 15 centímetros como terminales para la conexión de la bobina en el circuito. Como procedimiento adicional, puede aplicar esmalte transparente para uñas en algunas partes de la bobina con el fin de asegurar sus espiras o vueltas sobre la formaleta.

Para terminar la bobina quite o raspe el esmalte de los extremos y estáñelos con el fin de que se pueda soldar fácilmente al circuito.

Parte 2

Ensamble del radio de galena

En la figura A40 tenemos el diagrama esquemático del circuito que vamos a ensamblar. Para hacer-

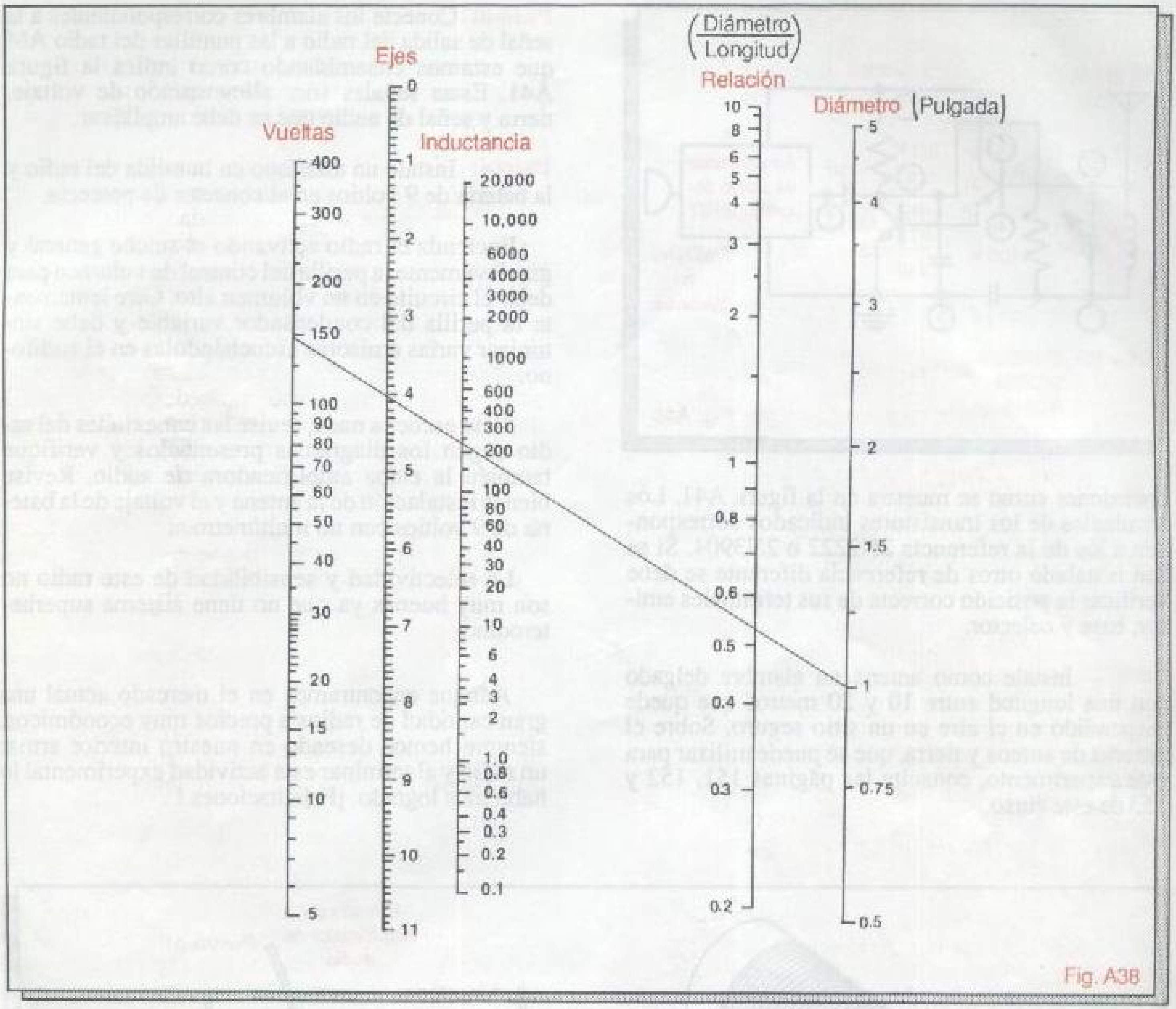

Fig. A38

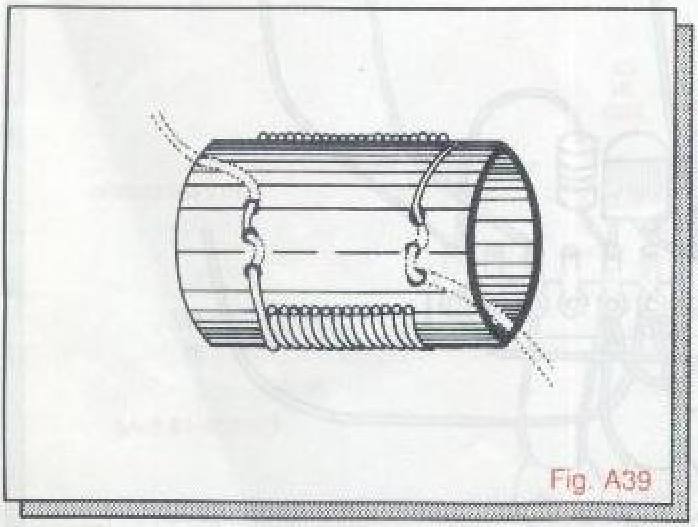

Fig. A39

lo, debemos conseguir los siguientes materiales adicionales a la bobina que hemos construido:

- 2 Transistores NPN Ref. 2N2222, 2N3904, BC549 o cualquier otro similar.
- 1 Condensador variable de 365 µF de una sola banda para AM.
- 2 Condensadores de cerámica de 0.1µF / 25 V
- 1 Condensador de cerámica de 0.01µF / 25 V
- 1 Resistencia de 100 KΩ 1/4 W 5%
- 1 Resistencia de 3.3 KΩ 1/4 W 5%
- 1 Regleta de conexiones de 8 o 10 terminales

Estos materiales son de fácil consecución y bajo costo, por lo que le recomendamos que se anime a realizar esta actividad experimental que le dará una gran satisfacción y le aumentará sus conocimientos.

El condensador variable lo podemos obtener de algún radio de AM que esté dañado.

Paso 1: Tome los componentes y vaya soldando uno por uno de izquierda a derecha en la regleta de

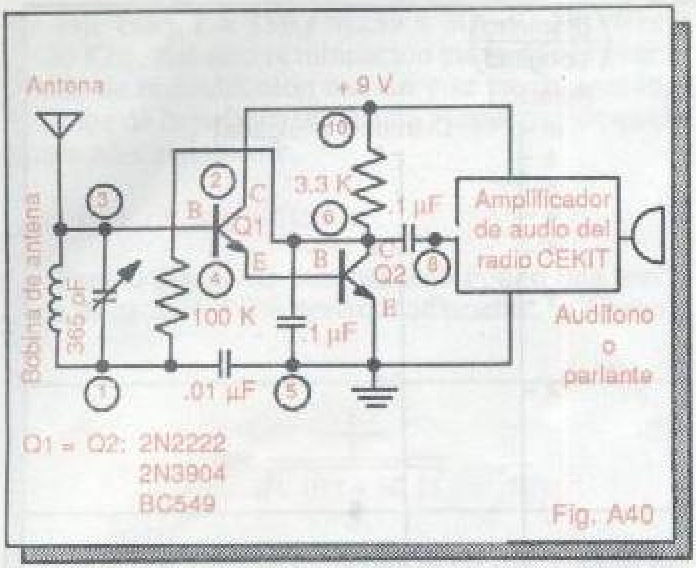
Antena
+9 V
10
2
C
3
B
3.3 K
Q1
6
.1 µF
Amplificador de audio del radio CEKIT
4
E
B
C
Q2
8
E
Bobina de antena
365 pF
100 K
.1 µF
Audífono o parlante
1
.01 µF
5
Q1 = Q2: 2N2222
2N3904
BC549

Fig. A40

conexiones como se muestra en la figura A41. Los terminales de los transistores indicados corresponden a los de la referencia 2N2222 o 2N3904. Si se han instalado otros de referencia diferente se debe verificar la posición correcta de sus terminales emisor, base y colector.

Paso 2: Instale como antena un alambre delgado con una longitud entre 10 y 20 metros que quede suspendido en el aire en un sitio seguro. Sobre el sistema de antena y tierra, que se puede utilizar para este experimento, consulte las páginas 151, 152 y 153 de este curso.

Paso 3: Conecte los alambres correspondientes a la señal de salida del radio a las puntillas del radio AM que estamos ensamblando como indica la figura A41. Estas señales son: alimentación de voltaje, tierra y señal de audio que se debe amplificar.

Paso 4: Instale un audífono en la salida del radio y la batería de 9 voltios en el conector de potencia.

Encienda el radio activando el suiche general y gire levemente la perilla del control de volumen para dejar el circuito en un volumen alto. Gire lentamente la perilla del condensador variable y debe sintonizar varias emisoras escuchándolas en el audífono.

Si no escucha nada, revise las conexiones del radio según los diagramas presentados y verifique también la etapa amplificadora de audio. Revise bien la instalación de la antena y el voltaje de la batería de 9 voltios con un multímetro.

La selectividad y sensibilidad de este radio no son muy buenas ya que no tiene sistema superheterodino.

Aunque encontramos en el mercado actual una gran cantidad de radios a precios muy económicos, siempre hemos deseado en nuestro interior armar un radio y al terminar esta actividad experimental lo habremos logrado. ¡Felicitaciones ! .

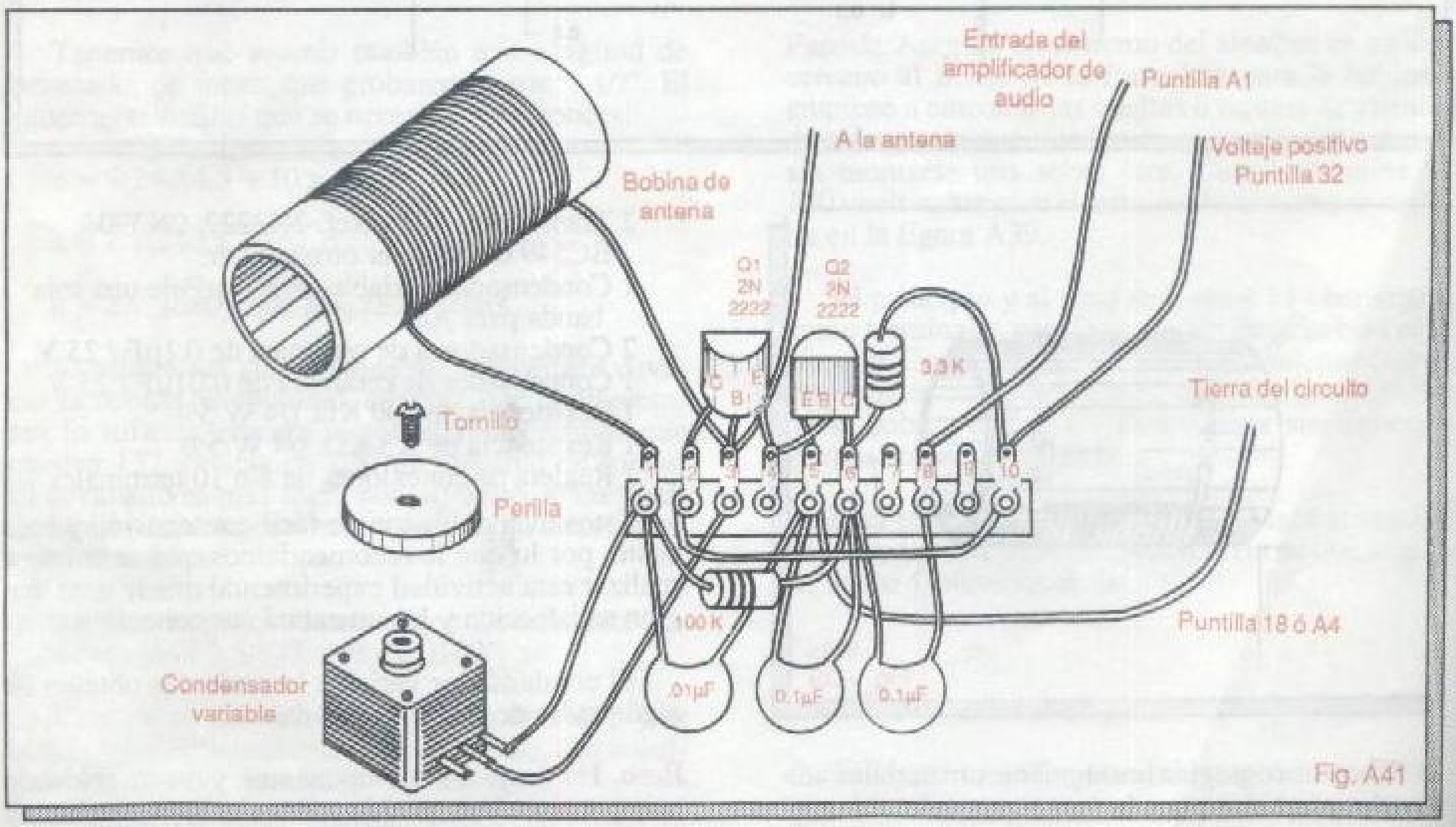
Entrada del amplificador de audio
Puntilla A1
A la antena
Voltaje positivo Puntilla 32
Bobina de antena
Q1 2N 2222
Q2 2N 2222
C E B
E B C
3.3K
Tierra del circuito
Tornillo
Perilla
1
2
3
4
5
6
7
8
9
10
100 K
Puntilla 18 ó A4
Condensador variable
.01µF
0.1µF
0.1µF

Fig. A41

El sistema de FM

Introducción

La modulación de frecuencia o FM, es uno de los métodos de transmisión de información más populares de nuestro tiempo. La principal razón de esta popularidad es la recepción prácticamente libre de ruido que este sistema provee, especialmente si se compara con la modulación de amplitud, su competidor más próximo.

La modulación de frecuencia se utiliza tanto en radioafición como en radiodifusión comercial o "broadcasting" El sonido de TV en VHF y UHF también se transmite en FM.

Las estaciones de radiodifusión comercial en FM transmiten en la banda de 88 a 108 MHz. A cada emisora se le asigna un ancho de banda de 200 KHz.

Para los radioaficionados se han destinado ciertas bandas específicas. La más conocida es la de 2 metros, la cual comprende frecuencias alrededor de 144 MHz.

Otras bandas de radioafición en FM muy populares son las de 52 a 54 MHz, 146 a 148 MHz, 222 a 225 MHz y 440 a 450 MHz, todas en el espectro de VHF.

En esta lección estudiaremos los fundamentos de la modulación de frecuencia, sus orígenes, sus ventajas y desventajas y otros temas de interés.

Conoceremos también cómo están estructurados, en términos generales, un transmisor y un receptor de FM.

Todo esto nos servirá como base para emprender, a partir de la próxima lección, el estudio completo de un receptor experimental de FM que usted mismo podrá construir, analizar y reparar, paso a paso.

Orígenes de la modulación de frecuencia

Los orígenes de la FM moderna se remontan a 1922, cuando Carson comparó matemáticamente los sistemas de AM y FM y demostró, con los métodos de detección conocidos en la época, que la modulación de frecuencia era inferior a la de amplitud, porque requería de más ancho de banda e introducía mayor distorsión.

El trabajo de Carson fue publicado originalmente en la revista "Procedimientos del IRE" de Febrero de 1922 y se titulaba "Notas sobre la Teoría de la Modulación". IRE es la sigla en inglés del Instituto de Ingenieros de Radio, hoy llamado Instituto de Ingenieros Eléctricos y Electrónicos (IEEE).

Carson fue injustamente criticado cuando la detección práctica de FM fue posible, pero sus argumentos prevalecieron por más de una década.

En 1936, el mayor Edwin H. Armstrong, en un intento por reducir los problemas de estática y ruido asociados con las recepciones de AM, desarrolló un método práctico de detección de FM y estableció las bases para lograr que la FM fuera competitiva con la AM.

El trabajo original de Armstrong fue publicado también en la revista "Procedimientos del IRE", pero en mayo de 1936 y se titulaba "Método de Reducir las Distancias en Radio-Señalización mediante un Sistema de Modulación de Frecuencia".

A Armstrong se le considera como el padre del sistema FM, aunque después de él surgieron pioneros como Crosby, Noble, Grammer, Goodman y otros. La modulación de frecuencia no fue un descubrimiento casual: nació primero en el papel y luego se comprobó en la práctica.

En 1939, el profesor Daniel E. Noble, de la Universidad de Connecticut, estableció formalmente los principios de la modulación de frecuencia, en un artículo titulado "Fundamentos de FM", publicado en la revista "QST" de agosto de ese año.

Noble había sido colaborador de Armstrong y entre ambos diseñaron dos estaciones de radiodifusión experimentales de FM de 43 MHz. Noble fue también el pionero de las comunicaciones móviles de dos vías en VHF-FM utilizando repetidoras.

Ventajas y desventajas de la modulación de frecuencia

La principal ventaja de la modulación de frecuencia es su habilidad para producir una alta relación de señal a ruido (SNR) cuando recibe una señal de moderada intensidad.

En otras palabras, la recepción en FM está prácticamente exenta de ruido e interferencia.

Por esta razón, la modulación de frecuencia es el modo elegido por los servicios de comunicaciones móviles y por las emisoras que difunden programas de calidad, en especial emisiones musicales.

En su camino del transmisor al receptor, una señal de RF puede incorporar ruido natural o ruido hecho por el hombre (MMN). El primero proviene de descargas atmosféricas y el segundo de aparatos generadores de chispa, como motores eléctricos, avisos luminosos de neón y sistemas de ignición de automóviles y motocicletas.

Todo ruido, cualquiera que sea su origen, modula en amplitud una onda viajera de RF. Por consiguiente, el ruido afecta la reproducción de la información en transmisiones de AM. La información transmitida puede ser sonido, imagen o de otro tipo.

La anterior es la razón por la cual los receptores de radio AM y de TV son muy susceptibles a los ruidos e interferencias provenientes de electrodomésticos, automóviles, aviones, luces, etc. (figura 367). La información de la imagen en TV se envía modulada en amplitud.

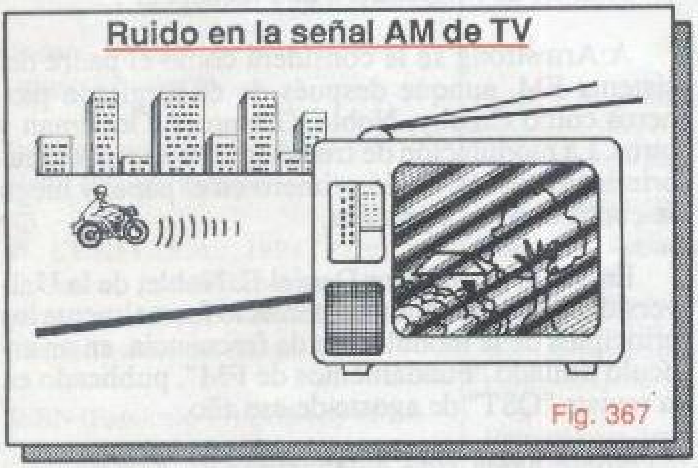
Ruido en la señal AM de TV

Fig. 367

El ruido no afecta la recepción de FM porque los circuitos de detección de información responden solamente a las variaciones de frecuencia e ignoran las variaciones de amplitud. En la figura 368 se muestra el caso de una señal de FM modulada en amplitud por ruido.

El ruido inducido en una señal modulada de RF puede también interferir con la frecuencia de la portadora, desviándola de su valor central.

Este problema se soluciona en los sistemas de FM utilizando, en el transmisor, un circuito de *preénfasis* a la entrada de audio del modulador y, en el receptor, un filtro de *deénfasis* a la salida del detector o demodulador.

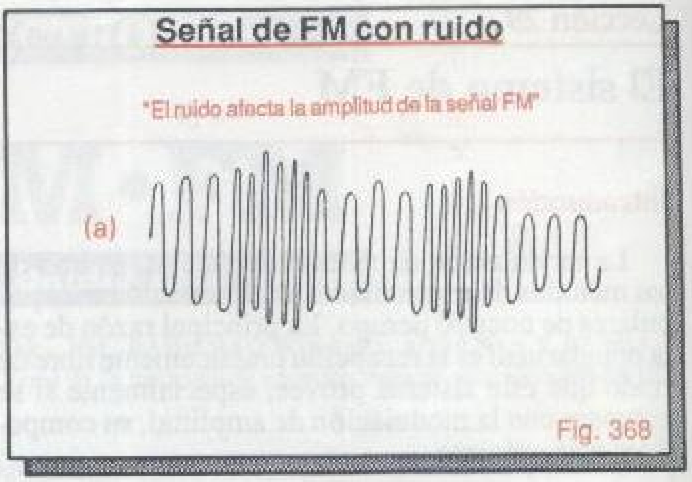
Señal de FM con ruido

Fig. 368

El circuito de preénfasis o preacentuación atenúa las frecuencias más bajas de audio y el de deénfasis las restablece a sus valores originales.

En la figura 369A se muestra un ejemplo de circuito de preénfasis, utilizado en transmisión de FM, y en la figura 369B un ejemplo de circuito de deénfasis, empleado en recepción.

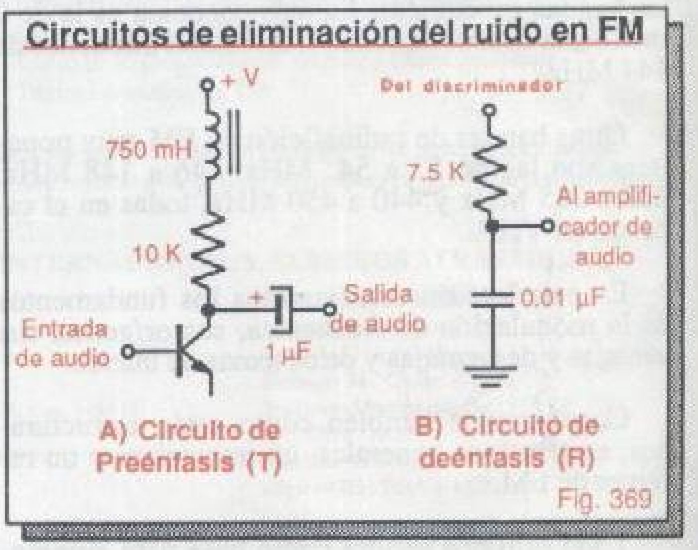
Circuitos de eliminación del ruido en FM

A) Circuito de Preénfasis (T)

B) Circuito de deénfasis (R)

Fig. 369

En contraste con la alta inmunidad al ruido, las señales de FM requieren de un gran ancho de banda y experimentan una severa distorsión cuando se propagan a través de la ionosfera.

En radioafición, esta última circunstancia limita el uso de la modulación de frecuencia a bandas como la de 10 metros y a varios sectores del espectro de VHF, como las bandas de 52 a 54 MHz, 146 a 148 MHz y otras.

El amplio ancho de banda requerido en FM es otro de los factores determinantes de la alta calidad de reproducción de sonido que se consigue por este sistema.

En radiodifusión por FM (88 a 108 MHz), las estaciones transmiten información dentro de un ancho de banda de 150 KHz, con 25 KHz más de seguridad a cada lado de los extremos de la banda. En total, disponen de 200 KHz (figura 370).

Ancho de banda en radiodifusión FM

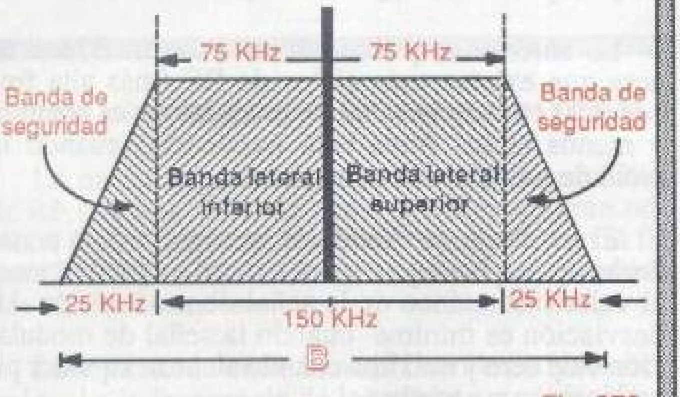
75 KHz
75 KHz
Banda de seguridad
Banda de seguridad
Banda lateral inferior
Banda lateral superior
25 KHz
25 KHz
150 KHz
B

Fig. 370

Este ancho de banda es 20 veces más grande que el utilizado en radiodifusión por AM (530 a 1600 KHz), donde se dispone de solo 10 KHz para enviar una gama muy limitada de tonos.

La utilización de un ancho de banda tan amplio permite introducir en la señal de FM toda la gama audible de sonidos (20 Hz a 20 KHz), incluyendo sus armónicos. Estos últimos son los que le dan el timbre distintivo a cada sonido, imprimiéndole fidelidad y realismo a cada audición.

Métodos de modulación

La propagación del sonido por el espacio tiene un alcance restringido. Por ejemplo, el sonido emitido por la voz humana no puede ir más allá de ciertos límites, incluso si se tiene una voz potente o se utiliza un amplificador de audio (figura 371).

Propagación del sonido

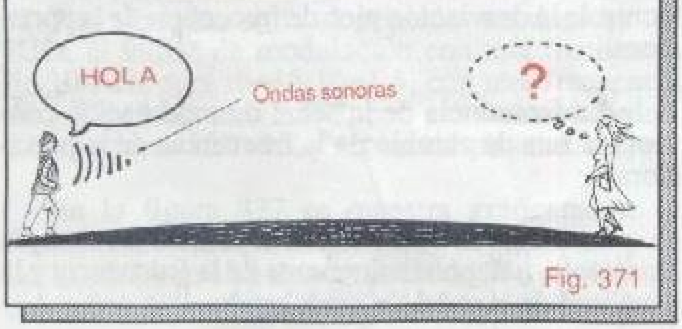
HOLA
Ondas sonoras
?

Fig. 371

Para transmitir sonido y, en general, cualquier información inteligente (voz, música, etc.) a lugares distantes, esta última debe convertirse en una señal eléctrica equivalente y luego enviarse modulada sobre una portadora de alta frecuencia. La señal resultante puede así irradiarse fácilmente al espacio, en forma de una onda electromagnética (figura 372).

Modulación del sonido

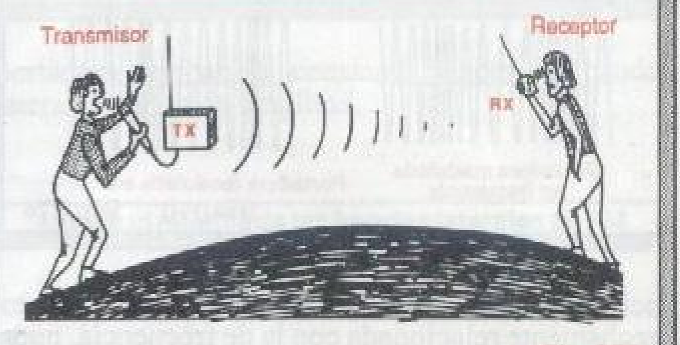
Transmisor
Receptor
TX
RX

Fig. 372

Mediante el proceso de modulación modificamos cualquier característica de la portadora de acuerdo a la información de audio que deseamos transmitir. Las características de una portadora susceptibles de modular son su amplitud, su frecuencia y su fase (figura 373).

Características de una portadora

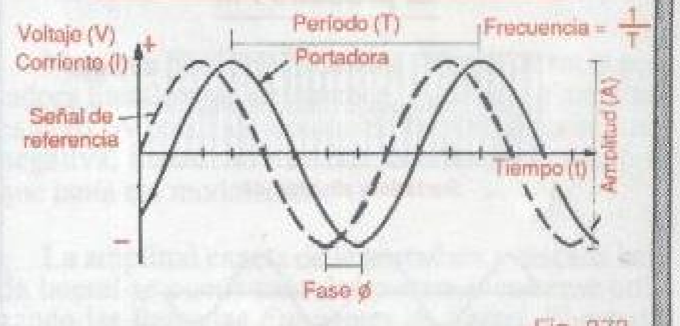
Voltaje (V)
Corriente (I)
+
Período (T)
Frecuencia = 1/T
Portadora
Señal de referencia
Amplitud (A)
Tiempo (t)
-
Fase ø

Fig. 373

Cuando la señal de modulación varía la amplitud de la portadora se obtiene **modulación de amplitud** (AM), cuando varía su frecuencia se obtiene **modulación de frecuencia** (FM) y cuando varía su fase el resultado es **modulación de fase** (PM). En la figura 374 se comparan las señales obtenidas por los tres métodos.

Los fundamentos de la modulación de amplitud se estudian en la Lección 19 de este curso. En esta lección nos ocuparemos exclusivamente de la modu-

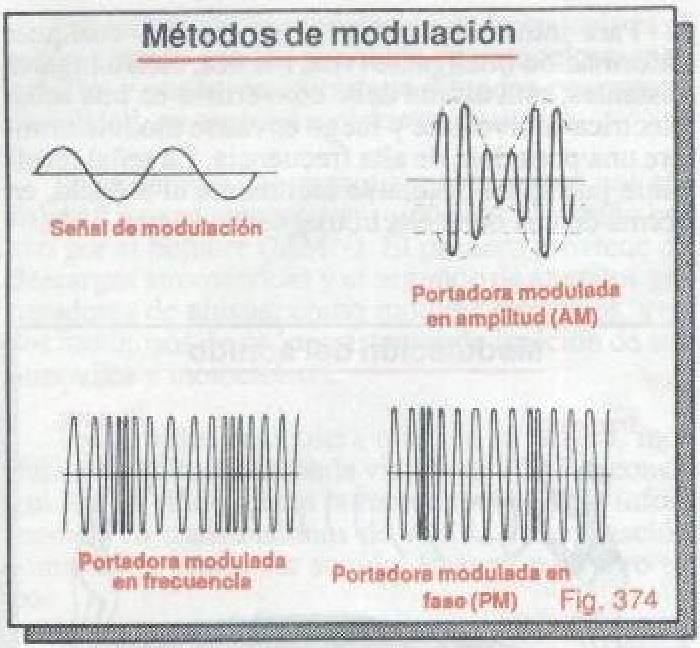
Métodos de modulación

Señal de modulación
Portadora modulada en amplitud (AM)
Portadora modulada en frecuencia
Portadora modulada en fase (PM)

Fig. 374

lación de frecuencia. La modulación de fase está estrechamente relacionada con la de frecuencia, puesto que no se puede variar la frecuencia de una señal sin afectar su fase y viceversa.

La modulación de frecuencia

En la figura 375 se describe gráficamente el proceso de modulación de frecuencia de una portadora de RF mediante una señal de audio. La combinación de estas dos señales la realiza en el transmisor un circuito especializado llamado *modulador*.

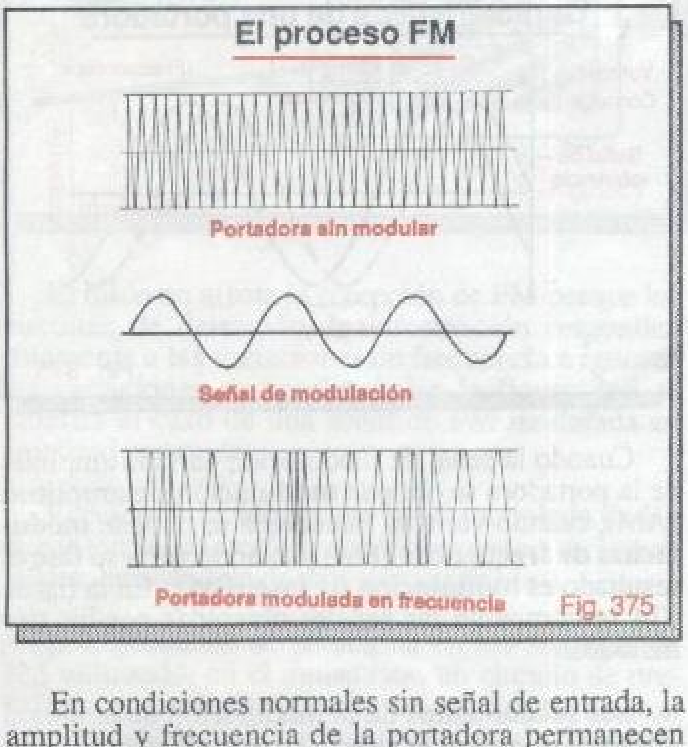
El proceso FM

Fig. 375

En condiciones normales sin señal de entrada, la amplitud y frecuencia de la portadora permanecen constantes. Cuando se aplica la señal de modulación, la frecuencia de la portadora varía por encima y por debajo de su valor central de acuerdo a la amplitud y polaridad de la señal de modulación o de audio aplicada.

Específicamente, la frecuencia de la portadora se incrementa durante los semiciclos positivos de la señal de modulación y se reduce durante los negativos. La frecuencia de la portadora es máxima o mínima cuando la señal de modulación alcanza su valor pico positivo o negativo, respectivamente.

Lo anterior se puede ver en la figura 375 si se nota que existen más ciclos de RF (más alta frecuencia) cuando la señal de modulación es positiva y menos ciclos (más baja frecuencia) cuando la señal de modulación es negativa.

El cambio o *desviación* de frecuencia de la portadora con respecto a su valor central es proporcional al valor instantáneo de la señal de modulación. La desviación es mínima cuando la señal de modulación vale cero y máxima cuando alcanza su valor pico, positivo o negativo.

La desviación máxima o pico se llama Δfc **("delta fc")** y su valor lo define el transmisor. Las estaciones de radiodifusión de FM que transmiten en la banda de 88 a 108 MHz utilizan una desviación máxima de ±75 KHz.

En televisión, Δfc=±25 KHz. En otros sistemas FM menos conocidos, Δfc=±15 KHz.

Para aclarar este concepto daremos un ejemplo. Si se modula una portadora de 100 MHz con una onda seno de 1 KHz y 1 Vp de amplitud en un sistema que tiene una desviación pico de ±15 KHz, la frecuencia de la portadora oscilaría 1000 veces por segundo entre 99985 KHz y 100015 KHz.

Una señal de modulación de 2 Vp y 1 KHz causaría una desviación máxima de ±30 KHz. Una señal de 2Vp y 2 KHz provocaría también una desviación pico de ±30 KHz pero a una rata de cambio de 2000 veces por segundo.

En conclusión, en una señal de FM:

- La amplitud máxima de la señal de modulación controla la desviación pico de frecuencia de la portadora.

- La frecuencia de la señal de modulación controla la rata de cambio de la frecuencia de la portadora.

- La amplitud de la portadora permanece siempre constante, independientemente de la frecuencia y la amplitud de la señal de modulación.

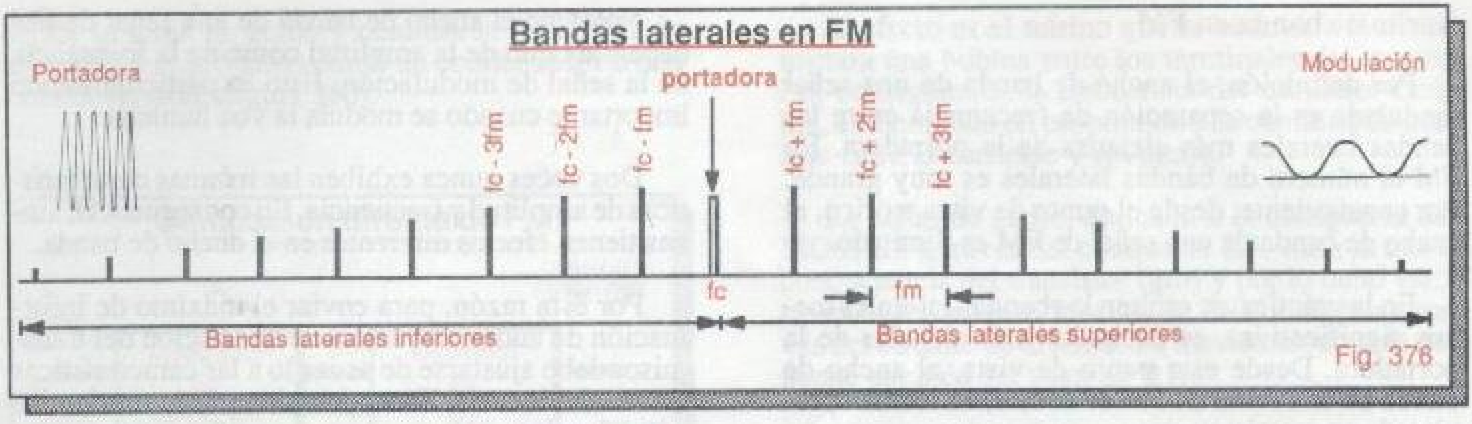
Bandas laterales en FM

Fig. 376

Bandas laterales e índice de modulación en FM

La modulación en frecuencia de una portadora de RF con una señal de audio produce un gran número de bandas laterales, como se muestra en la figura 376. La modulación de amplitud sólo produce dos bandas laterales.

La separación entre bandas laterales adyacentes es igual a la frecuencia de la señal de modulación y su amplitud disminuye progresivamente a medida que se alejan de la portadora. Las bandas laterales más próximas a la portadora son las más fuertes o significativas y las más alejadas, las más débiles.

Por ejemplo, si se modula en frecuencia una portadora de 100 MHz con un tono de 2 KHz, el primer par de bandas laterales estará ubicado en 99.998 MHz y 100.002 MHz, respectivamente, el segundo en 99.994 MHz y 100.004 MHz, el tercero en 99.996 MHz y 100.006 MHz, y así sucesivamente.

La amplitud de la portadora y de las bandas laterales depende del *índice de modulación* del proceso. En FM, el índice de modulación (m) se define como la relación entre la máxima desviación de frecuencia de la portadora (Δfc) y la frecuencia de la señal de modulación (fm). Esto es:

$$\text{Indice de modulación (m)} = \frac{\Delta fc}{fm}$$

Por ejemplo, si la desviación máxima es 15 KHz, el índice de modulación con una frecuencia de 10 KHz será m=15/10=1.5, con una frecuencia de 5 KHz sería m=15/5=3, con una frecuencia de 1 KHz sería m=15/1=15, y así sucesivamente.

En la figura 377 se muestra gráficamente la forma como varía la amplitud de la portadora y de las bandas laterales de una señal de FM en función del índice de modulación. En AM, la amplitud de la portadora permanece constante y sólo las bandas laterales varían de amplitud.

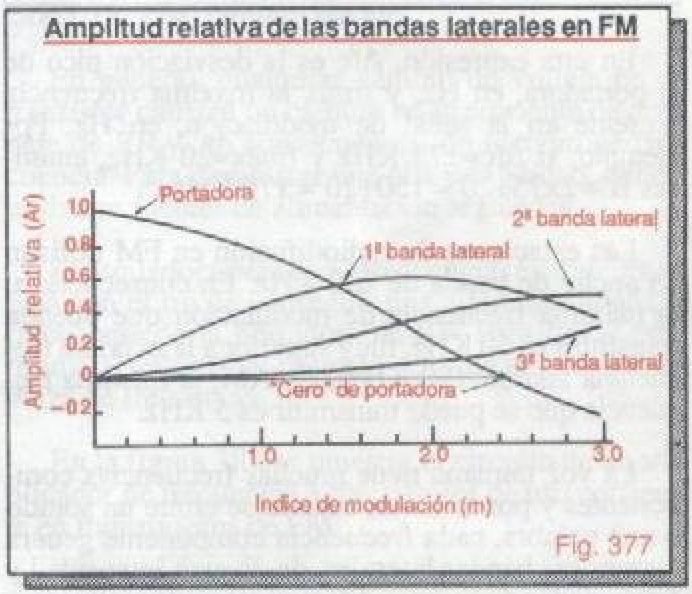
Amplitud relativa de las bandas laterales en FM

Fig. 377

Note que para determinados valores de **m**, la portadora literalmente desaparece, es decir, su amplitud es cero. A partir de entonces, la portadora se hace negativa, invirtiendo su fase con respecto a la fase que tenía sin modulación.

La amplitud exacta de la portadora y de cada banda lateral se puede calcular matemáticamente utilizando las llamadas *Funciones de Bessel*. El estudio de las funciones de Bessel está fuera de los propósitos de este curso.

Los valores de amplitud suministrados por la figura 377 son relativos. El valor real de la portadora o de cualquier par de bandas laterales se obtiene multiplicando el valor relativo obtenido (**Ar**) por la amplitud de la portadora sin modular (**Ac**).

Por ejemplo, si m=2 y Ac=5V, entonces, para el primer par de bandas laterales, Ar=0.9. En consecuencia, la amplitud real de la primera banda lateral es Ar x Ac= 0.9 x 5 = **4.5 V**.

Ancho de banda en FM

Por definición, el ancho de banda de una señal modulada es la separación de frecuencia entre las bandas laterales más alejadas de la portadora. En FM el número de bandas laterales es muy grande. Por consiguiente, desde el punto de vista teórico, el ancho de banda de una señal de FM es ilimitado.

En la práctica, se omiten las bandas laterales menos significativas, es decir, las más alejadas de la portadora. Desde este punto de vista, el ancho de banda de una señal de FM se define, en forma aproximada, como sigue:

$$B = 2\Delta fc + f\,máx$$

En esta expresión, **Δfc** es la desviación pico de la portadora, en Hz, y **fmáx** la máxima frecuencia presente en la señal de modulación, en Hz. Por ejemplo, si Δfc=±75 KHz y fmáx=20 KHz, entonces B = 2x75+20 = 150+20 = 170 KHz.

Las estaciones de radiodifusión en FM utilizan un ancho de banda de 200 KHz. En consecuencia, la máxima frecuencia de modulación que pueden transmitir es 50 KHz, muy superior a la máxima frecuencia audible (20 KHz). En AM, la máxima frecuencia que se puede transmitir es 5 KHz.

La voz humana tiene muchas frecuencias componentes y por lo tanto, cuando se emite un sonido o una palabra, cada frecuencia componente genera numerosas bandas laterales, de diversa intensidad y distanciadas de manera diferente en cada instante.

El número de bandas laterales significativas está también cambiando permanentemente. El resultado es un espectro continuo de bandas laterales a cada lado de la portadora (figura 378).

Espectro de una señal FM compleja

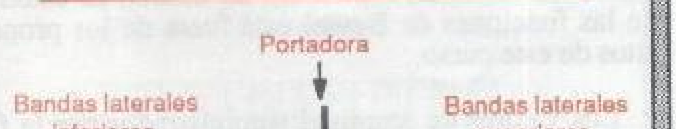

Fig. 378

Note que el ancho de banda de una señal de FM depende tanto de la amplitud como de la frecuencia de la señal de modulación. Esto es particularmente importante cuando se modula la voz humana.

Dos voces nunca exhiben las mismas características de amplitud y frecuencia. En consecuencia, ambas tienen efectos diferentes en el ancho de banda.

Por esta razón, para enviar el máximo de información de audio, el control de desviación del transmisor debe ajustarse de acuerdo a las características de voz del locutor. Esta operación puede resultar incómoda e imprecisa.

Para minimizar este problema, los equipos de FM utilizan un circuito denominado "clipper" o limitador en la sección procesadora de audio del transmisor. En la figura 379 se muestra el diagrama de bloques de la etapa amplificadora de audio de un transmisor típico de FM.

Amplificador de audio con limitador para FM

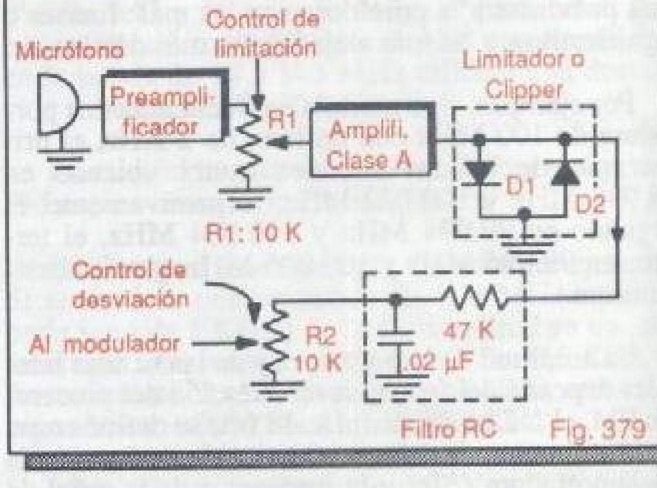

Fig. 379

En este circuito, R1 establece el nivel de limitación o recorte de la amplitud de la señal de audio y R2 la cantidad de desviación deseada. El filtro minimiza la distorsión de la señal de salida. La operación de recorte de los picos de audio la realizan los diodos D1 y D2.

Métodos de modulación de frecuencia

Existen básicamente dos métodos de modulación de frecuencia: el directo y el indirecto. En esta sección se describe el método directo por ser el más sencillo. El método indirecto se estudia en una lección posterior.

El proceso de modulación de frecuencia se realiza en el transmisor en un circuito denominado *modulador de reactancia*. El modulador de reactancia recibe la señal de audio y controla la frecuencia de un oscilador, actuando como un condensador o una bobina variable (figura 380).

Generación directa de FM

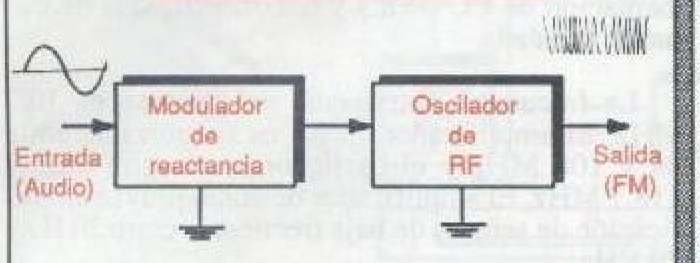
Entrada
(Audio)
Modulador
de
reactancia
Oscilador
de
RF
Salida
(FM)

Fig. 380

En la figura 381 se muestra el circuito de un modulador de reactancia con transistor MOSFET. La compuerta G1 se conecta al circuito tanque L1C1 del oscilador a través de la resistencia R1 y el condensador C2. El condensador C3 representa la capacitancia de entrada del transistor.

Modulador de reactancia para FM con MOSFET

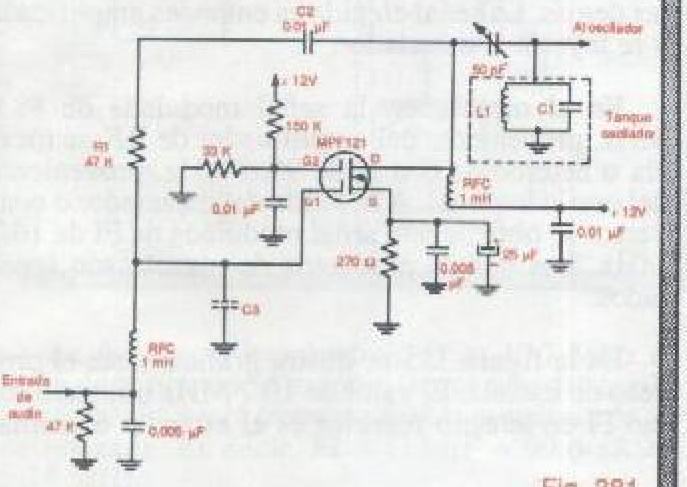
C2
0.01 µF
Al oscilador
50 pF
L1
C1
Tanque
oscilador
+ 12V
150 K
MPF121
R1
47 K
30 K
G2
D
0.01 µF
G1
S
RFC
1 mH
+ 12V
0.01 µF
270 Ω
0.005 µF
25 µF
C3
RFC
1 mH
Entrada
de
audio
47 K
0.005 µF

Fig. 381

La resistencia R1 es relativamente grande comparada con la reactancia de C3 a la frecuencia de operación. Por lo tanto, la corriente de RF a través de R1 y C3 está prácticamente en fase con el voltaje de RF presente en los terminales del circuito tanque.

Sin embargo, el voltaje a través de C3 se retrasará 90° con respecto a la corriente a través suyo. Esto provoca que la corriente de RF del circuito de drenaje se retrase también 90° con respecto al voltaje de RF del circuito tanque.

El efecto es el mismo que se obtendría si se conectara una bobina entre los terminales del tanque. Por consiguiente, la frecuencia del oscilador cambia, aumentando en proporción a la cantidad de desfase entre la corriente y el voltaje.

La señal de audio, aplicada a la compuerta del MOSFET a través del choque de RF, varía la transconductancia del transistor (gm) y por lo tanto varía también la corriente de salida de RF que circula a través del circuito de drenaje. La transconductancia depende también del valor de R1.

La resistencia R1 es, en realidad, el control de desviación de frecuencia del modulador. Para un nivel determinado de audio, la desviación aumenta a medida que aumenta el valor de R1 y viceversa. La desviación también aumenta cuando aumenta la relación L1/C1 del circuito tanque.

En general, cualquier cambio de voltaje en el transistor causará un cambio en la corriente de drenaje de RF y, en consecuencia, un cambio de frecuencia. Para garantizar una alta estabilidad, deben utilizarse fuentes de alimentación reguladas.

El oscilador opera a una frecuencia relativamente baja con el fin de conseguir una alta estabilidad de frecuencia de la portadora de RF. Para llevar esta última al valor final deseado se utilizan multiplicadores de frecuencia.

En la figura 382 se muestra el circuito de un triplicador de frecuencia a transistores de uso corriente en transmisores de FM.

Triplicador de frecuencia para FM

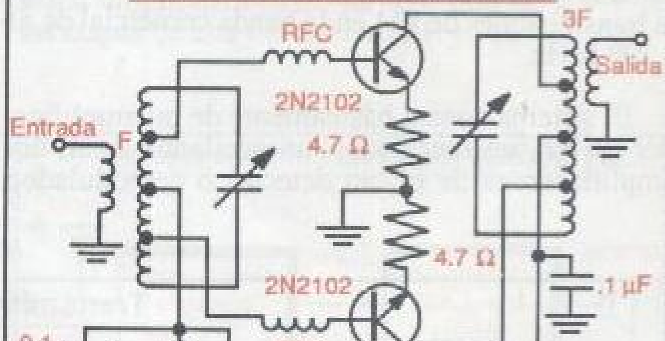
RFC
3F
Salida
2N2102
Entrada
F
4.7 Ω
4.7 Ω
2N2102
.1 µF
0.1
µF
10 Ω
RFC
560 Ω
10 µF
+ V

Fig. 382

Los multiplicadores son, básicamente, amplificadores no lineales y, por lo tanto, generan armónicos, es decir, frecuencias que son múltiplos enteros (2, 3, 4, etc.) de la frecuencia de entrada. El tanque sintonizado de salida selecciona el armónico deseado, en este caso el tercero.

Transmisores de FM

En la figura 383 se muestra el diagrama de bloques de un transmisor de FM típico. El sistema consiste básicamente de una fuente de señal, un procesador de audio, un modulador, un oscilador, varias etapas multiplicadoras, dos amplificadores de RF, un circuito de CAF y la antena.

El primer amplificador de RF se denomina excitador. La función del CAF (control automático de frecuencia) es estabilizar la frecuencia de salida. En el procesador de audio, la señal proveniente del micrófono se somete inicialmente al proceso de preénfasís, luego se amplifica y finalmente se limita.

El procesador de audio incluye el control de limitación de amplitud y el modulador el de desviación de frecuencia. Una vez amplificada y limitada, la señal de audio se aplica al modulador de reactancia, donde se verifica el proceso de modulación de frecuencia anteriormente descrito.

La señal de salida del modulador se aplica a varias etapas multiplicadoras de RF, las cuales desplazan la frecuencia de la portadora hasta obtener la frecuencia final deseada. Una vez obtenida, esta señal se inyecta a los amplificadores de RF y se irradia al exterior a través de la antena.

Receptores de FM

En la figura 384 se muestra el diagrama de bloques de un receptor superheterodino de FM, con la misma estructura del radio FM experimental que se estudiará a la lo largo de este curso. El circuito capta transmisiones de FM en la banda comercial de 88 a 108 MHz.

El sistema consta básicamente de un amplificador de RF, un mezclador, un oscilador local, dos amplificadores de FI, un detector o demodulador, un circuito de control automático de frecuencia (CAF) y un amplificador de audio.

Asociado al amplificador de RF está el circuito de antena. El amplificador de audio maneja el parlante. El mezclador y el oscilador local configuran el conversor de FI. El receptor utiliza tensiones de alimentación de CC de 8.3 y 6.2 V, extraidas de una fuente regulada.

La frecuencia intermedia del proceso es 10.7 MHz. El amplificador de RF es sintonizable entre 88 y 108 MHz y el oscilador local entre 98.7 y 118.7 MHz. El amplificador de audio provee amplificación de señales de baja frecuencia entre 20 Hz y 20 KHz.

Después de esta descripción general, veamos cómo opera el receptor.

Para simplificar la explicación, supongamos que el circuito de sintonía está ajustado para captar la señal enviada por una estación de FM que transmite con una frecuencia de portadora de 88.9 MHz. En consecuencia, la frecuencia del oscilador local es, automáticamente, 88.9+10.7 = 99.6 MHz.

A la antena convergen las señales provenientes de las diferentes estaciones del área. El circuito de sintonía asociado al amplificador de RF se encarga de seleccionar únicamente la de 88.9 MHz e ignora las demás. La señal elegida es entonces amplificada y se inyecta al mezclador.

En el mezclador, la señal modulada de 88.9 MHz, proveniente del amplificador de RF, se mezcla o heterodina con la de 99.6 MHz, proveniente del oscilador local. A la salida del mezclador o conversor se obtiene una señal modulada de FI de 10.7 MHz. Los demás productos de mezcla son ignorados.

En la figura 385 se ilustra gráficamente el proceso de mezcla. El valor de 10.7 MHz utilizado como FI en nuestro receptor es el estándar o norma-

Transmisor de FM típico

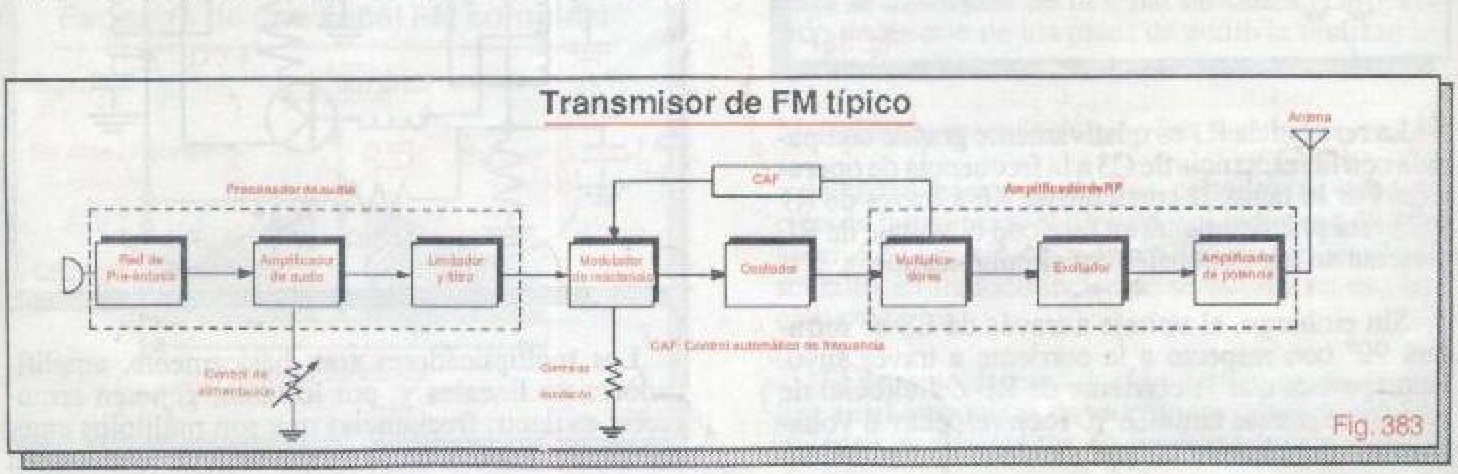

Fig. 383

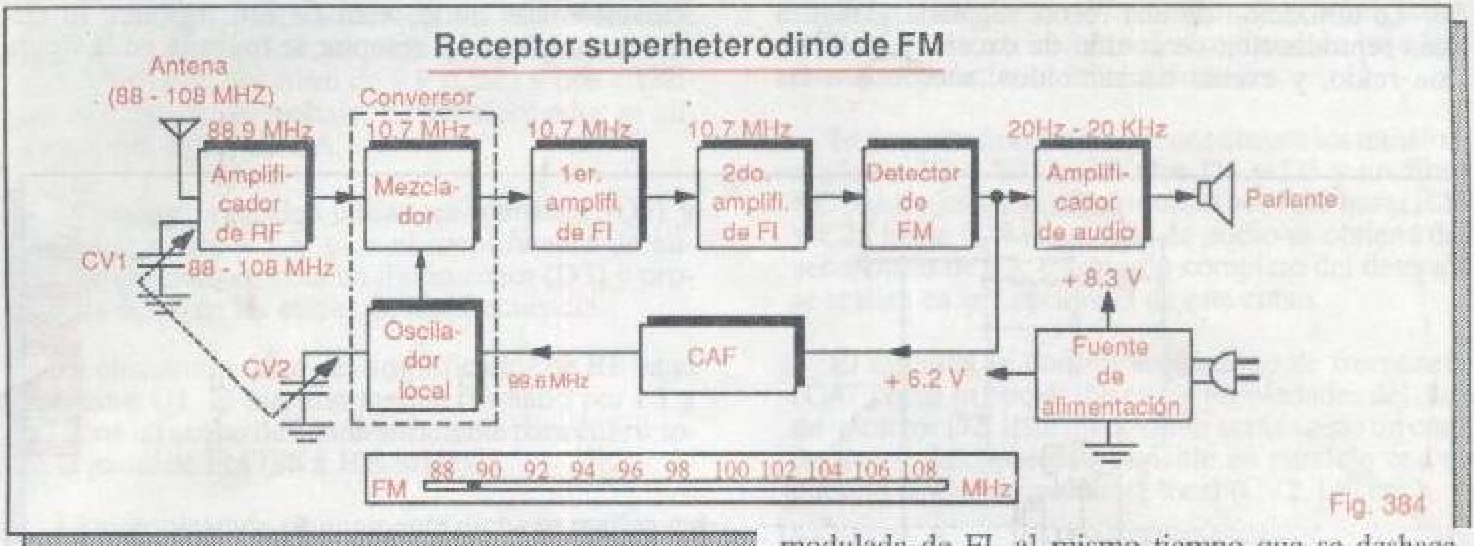
Antena
(88 - 108 MHZ)
88.9 MHz
Amplificador de RF
CV1
88 - 108 MHz
Conversor
10.7 MHz
Mezclador
Oscilador local
CV2
99.6 MHz
10.7 MHz
1er. amplifi. de FI
10.7 MHz
2do. amplifi. de FI
Detector de FM
20Hz - 20 KHz
Amplificador de audio
Parlante
+ 8.3 V
CAF
+ 6.2 V
Fuente de alimentación
FM
88 90 92 94 96 98 100 102 104 106 108
MHz

Receptor superheterodino de FM

Fig. 384

lizado en radiodifusión FM. En radiodifusión AM, la FI es 455 KHz.

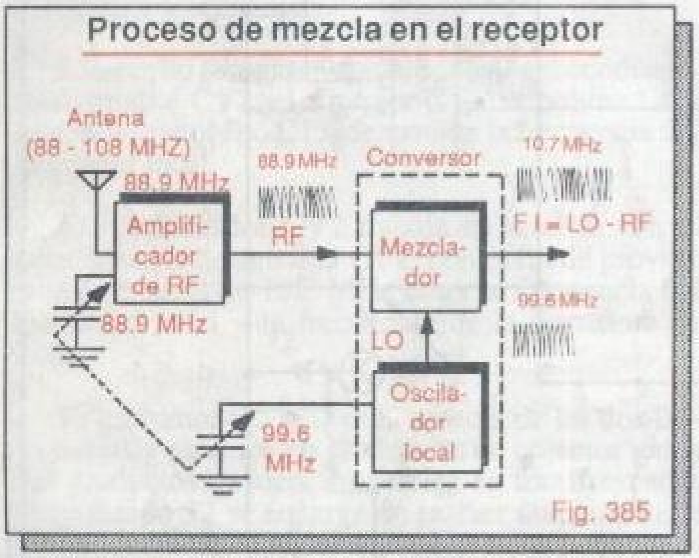
Antena
(88 - 108 MHZ)
88.9 MHz
Amplificador de RF
88.9 MHz
88.9 MHz
RF
Conversor
Mezclador
LO
Oscilador local
99.6 MHz
10.7 MHz
F I = LO - RF
99.6 MHz

Proceso de mezcla en el receptor

Fig. 385

La frecuencia intermedia (FI) de 10.7 MHz resulta de la diferencia entre la frecuencia del oscilador local (LO) y la frecuencia de la portadora de RF sintonizada. Es decir, FI = LO-RF = 99.6-88.9 = 10.7 MHz.

En síntesis, el efecto neto del proceso de mezcla es trasladar la modulación original de la portadora (88.9 MHz) a una portadora más baja (10.7 MHz). La información transportada por ambas señales es la misma.

A continuación, la señal de FI se amplifica en dos etapas sucesivas, alcanzando un nivel suficiente como para ser detectada.

El detector o demodulador cumple la función de extraer la información de audio implícita en la señal modulada de FI, al mismo tiempo que se deshace de la portadora de esta última (10.7 MHz). El proceso de detección utilizado en FM es sustancialmente diferente del empleado en un receptor de AM.

Una vez obtenida la señal de audio a la salida del detector, se envía al amplificador de audio donde alcanza la potencia suficiente como para impulsar el parlante y convertirse en información audible.

El propósito del circuito de control automático de frecuencia (CAF) es mantener constante la frecuencia del oscilador local, evitando su desplazamiento. Este control se logra mediante un nivel de CC proveniente de la salida del detector.

Finalmente, el receptor necesita de una fuente de alimentación regulada de CC para permitir que operen las diferentes etapas. Se requieren dos tensiones de alimentación (figura 386), una de 8.3 V para la etapa amplificadora de audio y otra de 6.2 V para las etapas de alta frecuencia (RF y FI).

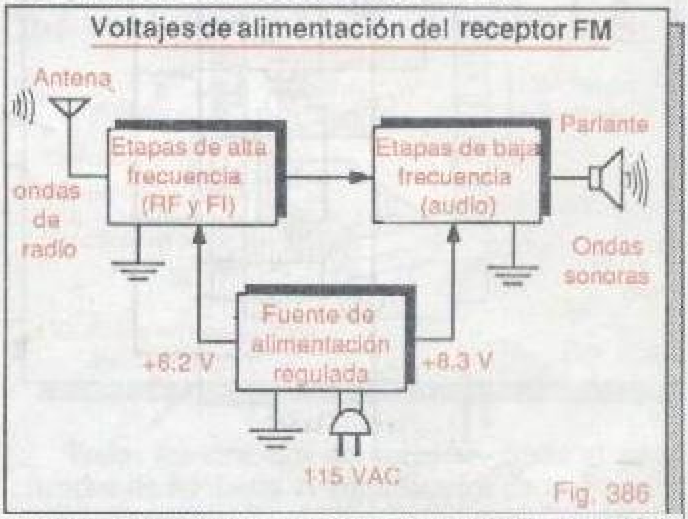
Antena
ondas de radio
Etapas de alta frecuencia (RF y FI)
Etapas de baja frecuencia (audio)
Parlante
Ondas sonoras
Fuente de alimentación regulada
+6.2 V
+8.3 V
115 VAC

Voltajes de alimentación del receptor FM

Fig. 386

La utilización de una fuente regulada garantiza una reproducción de sonido de excelente calidad, sin ruido, y exenta de zumbidos, acorde con las características de la señal de FM recibida. El circuito completo del receptor se muestra en la figura 387.

Diagrama esquemático del receptor FM experimental

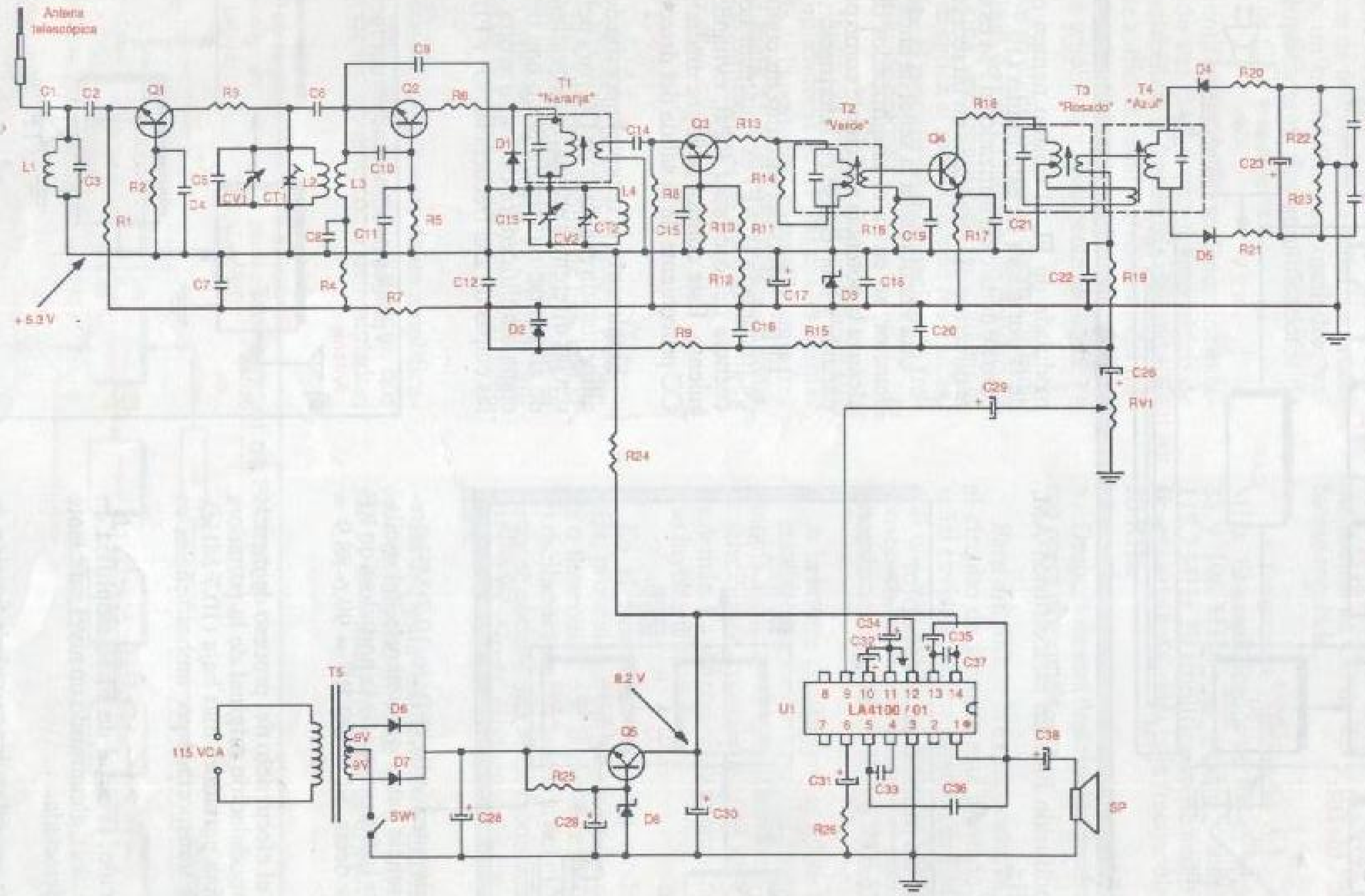
Antena telescópica
C1
C2
Q1
R3
C6
C9
Q2
R6
T1
"Naranja"
C14
Q3
R13
T2
"Verde"
Q4
R18
T3
"Rosado"
T4
"Azul"
D4
R20
C24
C23
R22
R23
C25
L1
C3
R1
R2
C4
C5
CV1
CT1
L2
L3
C8
C10
C11
R5
D1
C13
CV2
CT2
L4
R8
C15
R10
R11
R14
R16
C19
R17
C21
D5
R21
+ 5.3 V
C7
R4
R7
C12
D2
R9
R12
C18
C17
D3
C16
R15
C20
C22
R19
C26
C29
RV1
R24
T5
115 VCA
D6
9V
9V
D7
SW1
C28
R25
C28
Q5
D8
8.2 V
C30
U1
LA4100 / 01
8 9 10 11 12 13 14
7 6 5 4 3 2 1
C34
C32
C35
C37
C31
C33
C36
R26
C38
SP

Fig. 387

La fuente de alimentación consta básicamente de un transformador (T5), un rectificador de onda completa (D6, D7), un filtro de RF (C28) y dos circuitos reguladores de voltaje. El transformador se alimenta con 115 V de CA.

El primer regulador utiliza un transistor (Q5) y proporciona los 8.3 V para el amplificador de audio. El segundo emplea un diodo zener (D3) y provee los 6.2 V de las etapas de alta frecuencia.

El elemento activo del amplificador de RF es el transistor Q1. El circuito tanque formado por L1 y C3 tiene un ancho de banda suficiente para cubrir toda la gama de FM (88 a 108 MHz).

La sintonización propiamente dicha se realiza en el circuito tanque constituido por el condensador variable CV1, el trimmer CT1, el condensador fijo C5 y la bobina L2.

El conversor de FI es el transistor Q2. Este dispositivo realiza simultáneamente las funciones de mezclador y oscilador.

El circuito resonante constituido por el condensador variable CV2, el trimmer CT2, la bobina L4 y el condensador fijo C13, determina la frecuencia del oscilador local.

El condensador CV2 trabaja en tandem con el condensador de sintonía CV1, con el fin de proveer una diferencia de 10.7 MHz entre la frecuencia del oscilador local y la frecuencia de la portadora de FM.

El transistor Q2 realiza la mezcla de las dos frecuencias y produce en su circuito de colector todos los productos de mezcla posibles. El transformador sintonizado T1 se encarga de extraer únicamente la componente de FI, es decir, la que tiene la frecuencia diferencia (10.7 MHz).

La bobina L3 y el condensador C8 constituyen un circuito "trampa" sintonizado a 10.7 MHz. Su función es impedir que la señal de FI se devuelva al emisor y origine oscilaciones parásitas.

El elemento activo del primer amplificador de FI es el transistor Q3, conectado en la configuración base común. El circuito posee un transformador sintonizado a 10.7 MHz en el circuito de colector (T2).

El segundo amplificador de FI es el transistor Q4, conectado en la configuración emisor común. Este circuito posee también un transformador sintonizado a 10.7 MHz en su circuito de salida.

El demodulador de audio utilizado se denomina "detector de relación". Su función consiste en convertir las variaciones de frecuencia de la señal modulada de FI en variaciones de amplitud de baja frecuencia, es decir, en señales de audio.

El detector de relación lo constituyen los transformadores T3 y T4, los diodos D4 y D5 y un filtro RC pasabajos de salida formado por R20 hasta R23 y C23 hasta C25. La salida de audio se obtiene del secundario de T3. El estudio completo del detector se realiza en la Lección 33 de este curso.

El circuito de control automático de frecuencia (CAF) basa su operación en las propiedades del diodo varactor D2. Este dispositivo actúa como un condensador de capacidad variable en paralelo con el circuito tanque de oscilador local (CV2, L4, etc.).

La tensión inversa de polarización de D2 se obtiene de la señal de audio a través de R15, C16, R9 y C20. La operación detallada del CAF se estudia en la Lección 34 de este curso.

El elemento central del amplificador de audio es el circuito integrado LA4100/01. La señal de salida para el parlante se obtiene del pin # 1. La señal de entrada se aplica al pin # 9. El circuito puede proveer hasta 1.5 W de potencia.

La tensión de alimentación del amplificador de audio se aplica a los pines # 14 (+ 8.3V) y # 3 (tierra). Los demás componentes (C32, C33, etc.) cumplen funciones auxiliares.

La operación completa del CI LA 4100/01 se describe en detalle en la Lección 38 de este curso. El dispositivo incorpora internamente 14 transistores y 12 resistencias y tiene el aspecto mostrado en la figura 388.

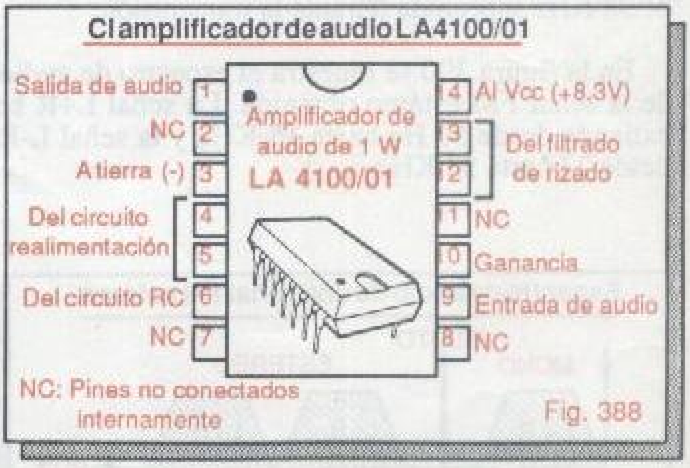

Fig. 388

Todos los circuitos del receptor, desde el amplificador de RF hasta el amplificador de audio se estudian detalladamente a partir de la próxima lección.

El sistema FM estéreo

En la figura 389 se muestra la disposición básica de un transmisor de FM estéreo. Las señales provenientes de los micrófonos derecho (R) e izquierdo (L) se procesan en un mezclador de audio. A la salida del mezclador se obtiene la señal suma L+R y la señal diferencia L-R.

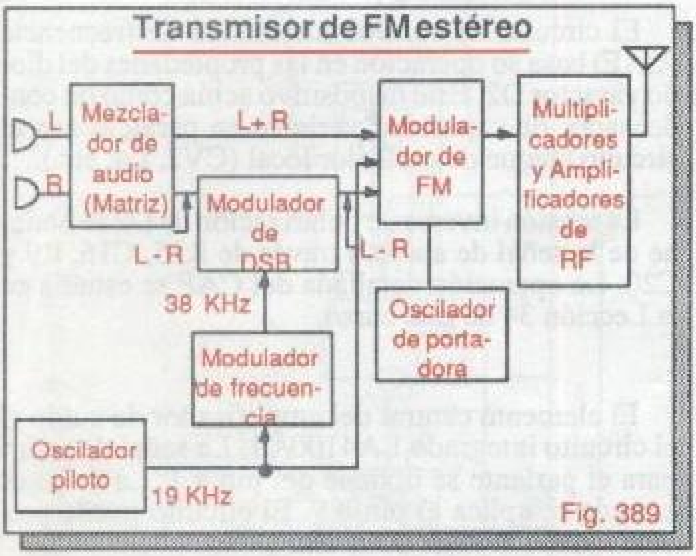
Transmisor de FM estéreo

Fig. 389

La señal L+R modula directamente la portadora de 88 a 108 MHz en el modulador de FM. La señal L-R modula una subportadora de 38 KHz en un modulador de doble banda lateral (DSB). El modulador de DSB suprime la subportadora de 38 KHz y entrega la señal L-R modulada al modulador FM.

La subportadora de 38 KHz proviene de un oscilador de 19 KHz seguido de un multiplicador. La señal de 19 KHz modula también la portadora de FM y se utiliza como señal *piloto*. Su propósito es facilitar en el receptor la regeneración de la subportadora de 38 KHz suprimida durante la transmisión.

En la figura 390 se muestra el espectro de audio de la señal FM estéreo obtenida. La señal L+R se extiende desde 50 Hz hasta 15 KHz y la señal L-R desde 23 hasta 53 KHz.

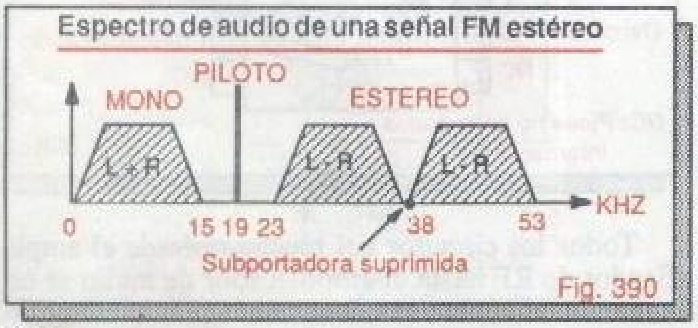
Espectro de audio de una señal FM estéreo

Fig. 390

Actividad práctica Nº 16

Ensamble de la segunda etapa de frecuencia intermedia. (Parte 1)

En esta actividad vamos a instalar en el tablero del radio una parte de los componentes que forman la etapa del segundo amplificador de FI o sea la etapa marcada con color naranja en la lámina del diagrama del radio.

Componentes necesarios

1 Transistor NPN C1390 o su reemplazo. (2N2222 o similar). Q3.

1 Resistencia de 15 KΩ, 1/4 W, 5%, (café, verde, naranja, dorado) R10.

Nota: El transistor Q3, C1390 se puede reemplazar por el 2N2222 que tiene la siguiente configuración en sus terminales (figura A42).

Paso 1: Tome el transistor Q3 y separe ligeramente sus terminales, fijándose en la figura A42. Si el transistor es 2N2222, suelde sus terminales así: la base (B) en la puntilla E4, el emisor (E) en la puntilla E8 y el colector (C) en la puntilla E7.

Paso 2: Instale la resistencia R10 de 15 KΩ entre las puntillas E6 y D18.

Paso 3: Con el alambre de conexión, una las puntillas E10 y 14, dejando unos 3 cms. sobrantes en la puntilla E10 para una conexión posterior. Con el mismo alambre una las puntillas E12 y 40, teniendo en cuenta de no hacer corto con el alambre que va de E7 a E11.

Paso 4: Con el alambre de conexión unas las puntillas E8, E9 y E5.

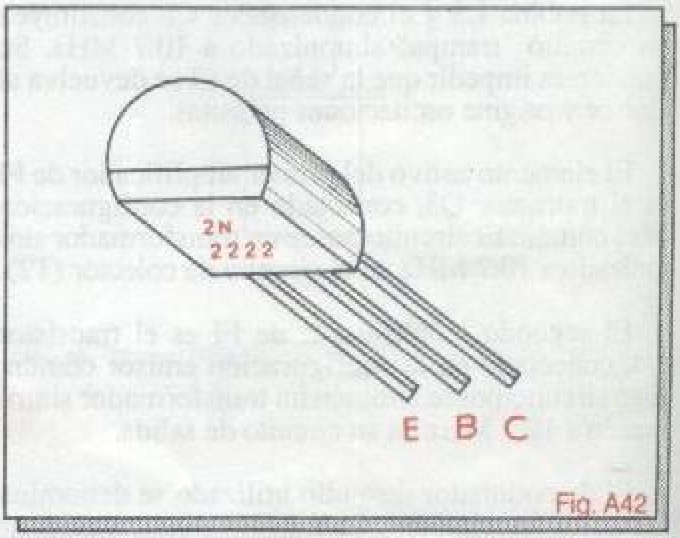

Fig. A42

Continuación actividad práctica Nº 16

Ensamble de la etapa de frecuencia intermedia (Parte 2)

En esta actividad vamos a continuar con la instalación en el tablero del radio de los componentes de la segunda etapa de frecuencia intermedia (FI).

Componentes necesarios

1 Resistencia de 1.5 kΩ, 1/4 W, 5% (café, verde, rojo, dorado). R9.
1 Resistencia de 820 kΩ (gris, rojo, amarillo, dorado). R8.
1 Condensador de cerámica de 0.02 µF/50V. C7.
1 Condensador de cerámica de 0.02 µF/50V. C6.

Paso 1: Tome la resistencia de 1.5 kΩ, marcada en el diagrama como R9 e instálela entre las puntillas E9 y 13.

Paso 2: Tome la resistencia de 820 kΩ, marcada en el diagrama como R8 e instálela entre las puntillas 41 y E3.

Paso 3: Con el alambre de conexión, una las puntillas E1 yE4 cuidando de no cortocircuitar el terminal de la resistencia R8. Para mayor seguridad, doble este alambre hacia arriba en el punto de cruce.

Paso 4: Con el alambre de conexión, una las puntillas E2 y E3.

Paso 5: Tome el condensador C7 de 0.02 µF; éste puede estar marcado como 0.02, .02, 203Z ó 223Z. Instálelo entre las puntillas E5 y 12.

Paso 6: Tome el condensador C6, igual al anterior de 0.02 µF e instálelo entre las puntillas E3 y 11.

Nota : Para terminar el ensamble de esta etapa, sólo falta instalar el transformador de FI identificado como T3 y reconocible por su núcleo negro. Este transformador se colocará en una futura actividad.

R8 polariza directamente la unión base-emisor de Q3, a través del secundario de T2 (segundo transformador de FI, nucleo blanco). R9 conecta el emisor de Q3 a tierra y estabiliza térmicamente el punto de trabajo de este transistor.

Los condensadores C6 y C7 ofrecen un camino de baja resistencia a tierra para las señales de alta frecuencia presentes en el punto de prueba H y en el emisor de Q3, respectivamente. Estos componentes no afectan la polarización de CC de Q3.

Actividad práctica Nº 17

Ensamble de un generador de FI de 455 kHz con modulación

Objetivo

El objetivo de esta actividad es construir un sencillo generador de onda seno de 455 kHz controlado a cristal, con modulación de onda cuadrada incluida. En la figura A43 se muestra la forma de onda de la señal de salida obtenida.

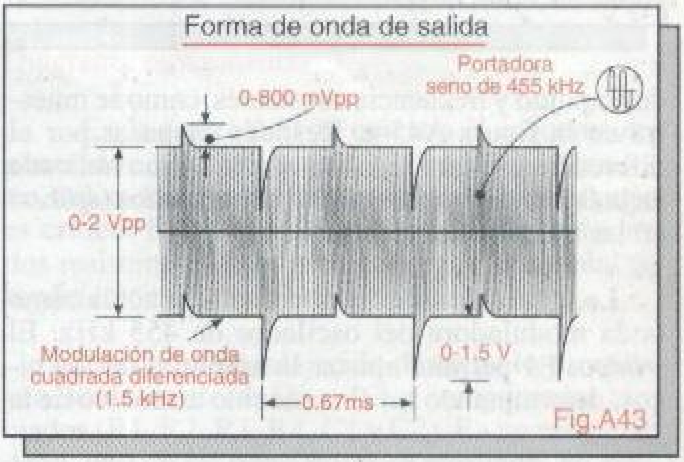

Fig.A43

El circuito resulta muy útil y necesario para la prueba de las etapas de detección y de FI en receptores de AM domésticos y portátiles. La señal de modulación puede también utilizarse para la prueba de amplificadores de audio.

Este generado"r lo utilizaremos posteriormente para alinear y ajustar las etapas de FI del radio AM CEKIT que se está ensamblando a lo largo de este curso. Por su bajo costo y fácil realización, es un reemplazo ideal para los generadores de RF utilizados normalmente en esta operación.

Después de utilizarlo en este curso, este generador será una herramienta muy valiosa para su trabajo tanto profesional como de aficionado.

Teoría de operación

En la figura A44 se muestra el diagrama de bloques del generador de 455 KHz propuesto. El sistéma consta básicamente de un oscilador de relajación de 1.5 kHz, una red RC diferenciadora (filtro pasa altos) y un oscilador a cristal tipo Colpitts de 455 kHz con modulación.

El oscilador de relajación (multivibrador astable) suministra una onda cuadrada simétrica,

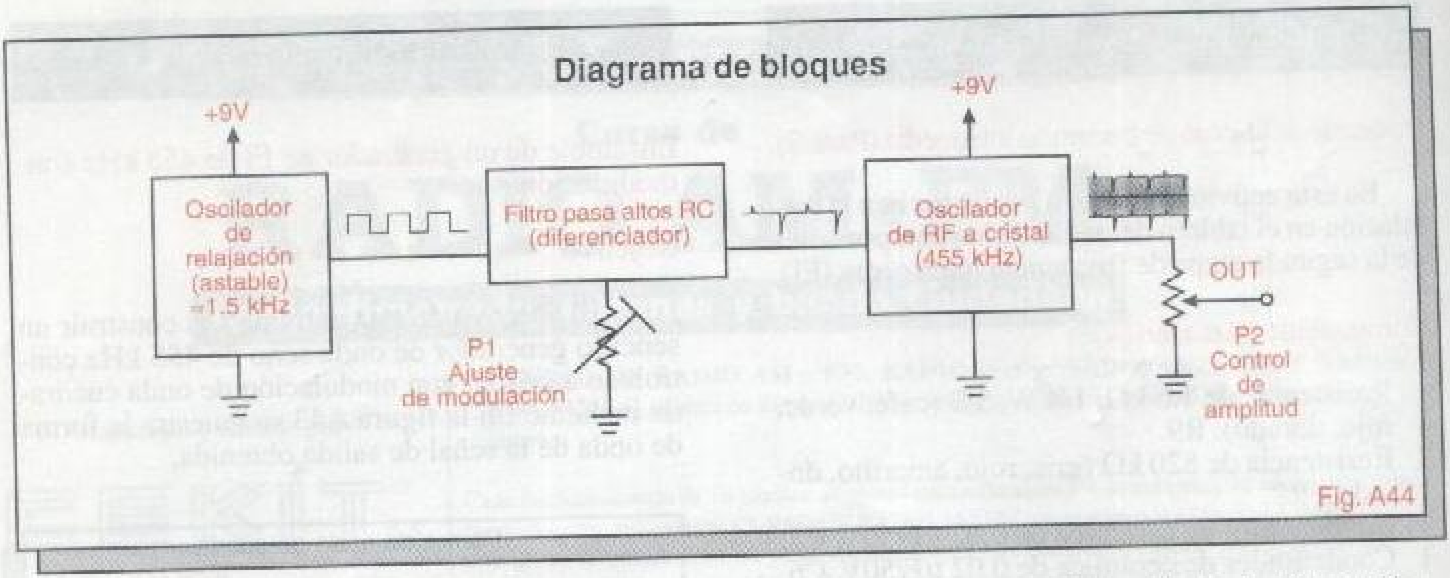
Diagrama de bloques

+9V
Oscilador de relajación (astable) ≈1.5 kHz
Filtro pasa altos RC (diferenciador)
P1 Ajuste de modulación
+9V
Oscilador de RF a cristal (455 kHz)
OUT
P2 Control de amplitud

Fig. A44

de amplitud y frecuencia constantes, como se muestra en la figura A45-a. Después de pasar por el diferenciador, esta señal adopta la forma indicada en la figura A45-b, con picos pronunciados (*spikes*) en las transiciones o flancos.

La señal de salida del diferenciador actúa como onda moduladora del oscilador de 455 kHz. El *trimpot* P1 permite ajustar la amplitud de los picos, determinando así el contenido armónico de la señal de audio y la profundidad de la modulación.

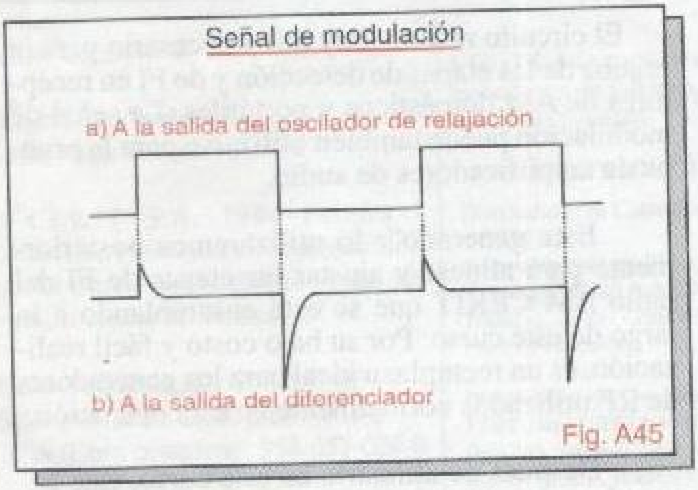
Señal de modulación

a) A la salida del oscilador de relajación

b) A la salida del diferenciador

Fig. A45

Sin señal de entrada (P1 en la posición mínima), el oscilador de FI produce una forma de onda prácticamente sinusoidal, como la que se muestra en la figura A46. La frecuencia de la señal de FI (455 kHz) la determina un cristal cerámico.

Con señal de entrada (P1 al máximo o en una posición intermedia), el oscilador de FI produce una forma de onda como la que se muestra en la figura A47. Se trata, fundamentalmente, de una portadora sinusoidal de 455 kHz modulada en amplitud por la onda diferenciada.

La señal de salida del oscilador de FI se aplica al circuito bajo prueba a través del condensador C7 y el potenciómetro P2. El primero desacopla el nivel de CC de la señal de FI, mientras P2 controla la amplitud de esta señal entre 0 y 6 Vpp.

Es importante destacar que la señal de modulación suministrada por el diferenciador al oscilador de FI produce un gran número de bandas laterales alrededor de la frecuencia fundamental de 455 kHz y sus armónicos (910 kHz, 1365 kHz, etc.). Esta circunstancia se muestra en la figura A48.

Tanto la fundamental como sus armónicos (910 kHz, 1365 kHz, etc.) son ondas sinusoidales puras. Asimismo, cada banda lateral producida por la señal de modulación es también una onda sinusoidal pura.

Si lográramos filtrar la portadora de 455 kHz y su primer par de bandas laterales, obtendríamos una forma de onda como la que se muestra en la figura A49. Esta situación corresponde al clásico ejemplo de modulación de una señal sinusoidal de alta frecuencia (portadora) mediante un tono, también sinusoidal, de baja frecuencia (moduladora).

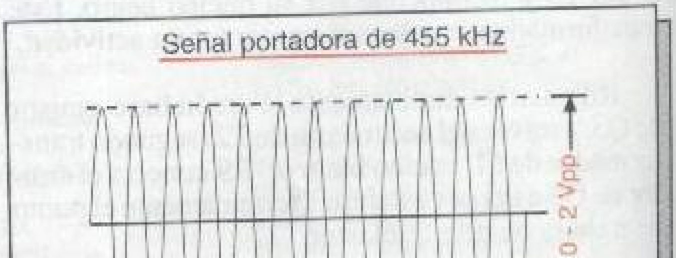
Señal portadora de 455 kHz

Fig. A46

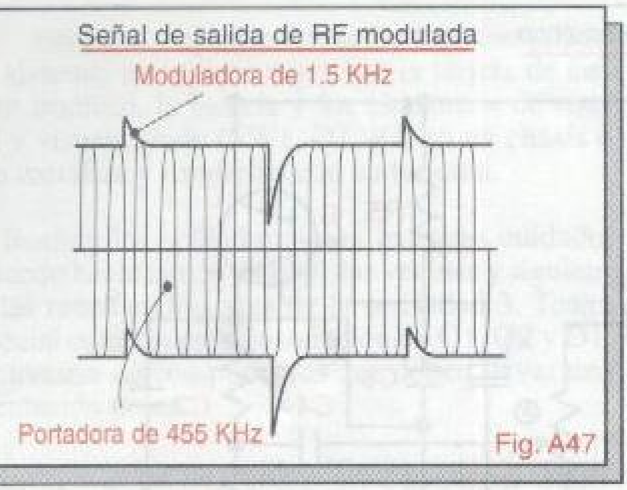

Señal de salida de RF modulada

Fig. A47

Para obtener una señal como la de la figura A49 se necesita disponer de un filtro muy selectivo y, por consiguiente, costoso. Sin embargo, esto no es absolutamente necesario en nuestro caso, puesto que la portadora es una onda sinusoidad y los circuitos de FI del receptor son lo bastante selectivos para lograr un tono casi sinusoidal a la salida del detector.

En realidad, la señal que se observa a la salida del detector es algo más compleja de lo que la explicación precedente sugiere. En la figura A50 se presentan algunos ejemplos de señales obtenidas a la salida del detector de audio del radio AM CEKIT para distintas formas de onda de la señal de prueba. Se supone que está última se aplica al primario del primer transformador de FI (T1).

Una vez familiarizados con los principios de operación del circuito, vamos ahora a proceder a su construcción y calibración. A continuación se suministran el diagrama esquemático, el circuito impreso, la guía de localización de componentes y la lista de materiales. La forma de utilizar este instrumento en la calibración de las etapas de FI del radio AM CEKIT y otras aplicaciones se tratará en futuras actividades.

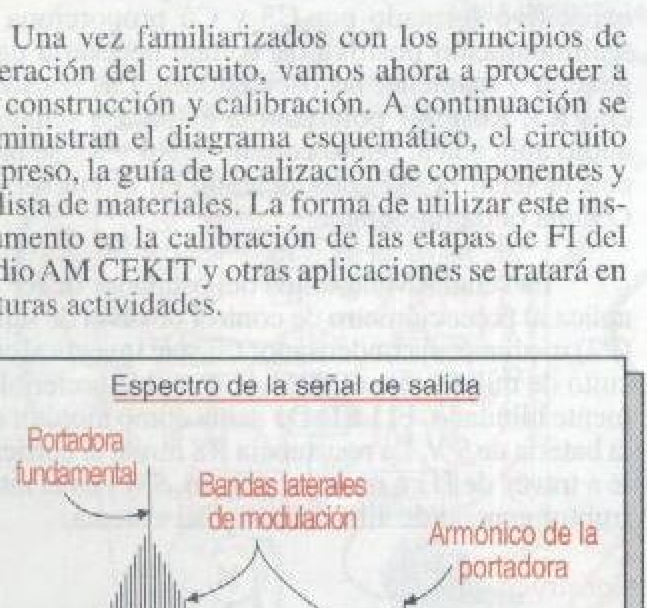

Espectro de la señal de salida

Fig. A48

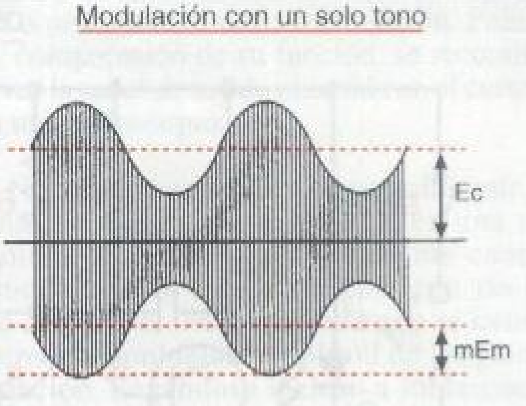

Modulación con un solo tono

Fig. A49

Diagrama esquemático

En la figura A51 se muestra el esquema del generador de FI propuesto. Todos los componentes utilizados son de fácil consecución y ninguno es crítico. El *trimpot* P1 puede ser sustituido por dos resistencias fijas para conseguir una señal de modulación de características tonales fijas.

El oscilador de relajación de audio lo constituyen los transistores Q1 y Q2, y componentes asociados (R1, R2, R3, R4, C1 y C2). En un momento dado, uno de los transistores se encuentra en estado de saturación (conduciendo), mientras el otro permanece en estado de corte (bloqueado).

Asumiendo que Q1 está bloqueado, su voltaje de colector es igual al voltaje de alimentación (+9 V). Puesto que Q2 está conduciendo, su voltaje de colector es muy bajo, del orden de 0.1 V. Esto permite que el condensador C1 se cargue a través de la resistencia R1 y tienda a alcanzar el voltaje de la fuente (9 V).

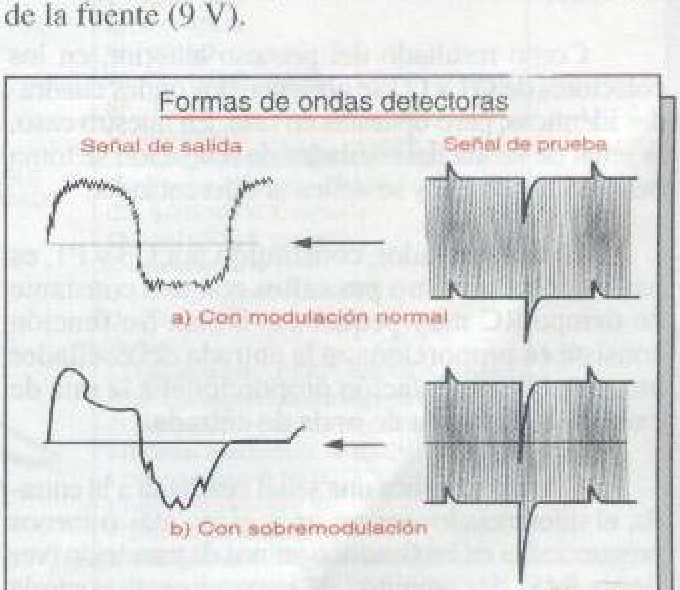

Formas de ondas detectoras

Fig. A50

Diagrama esquemático

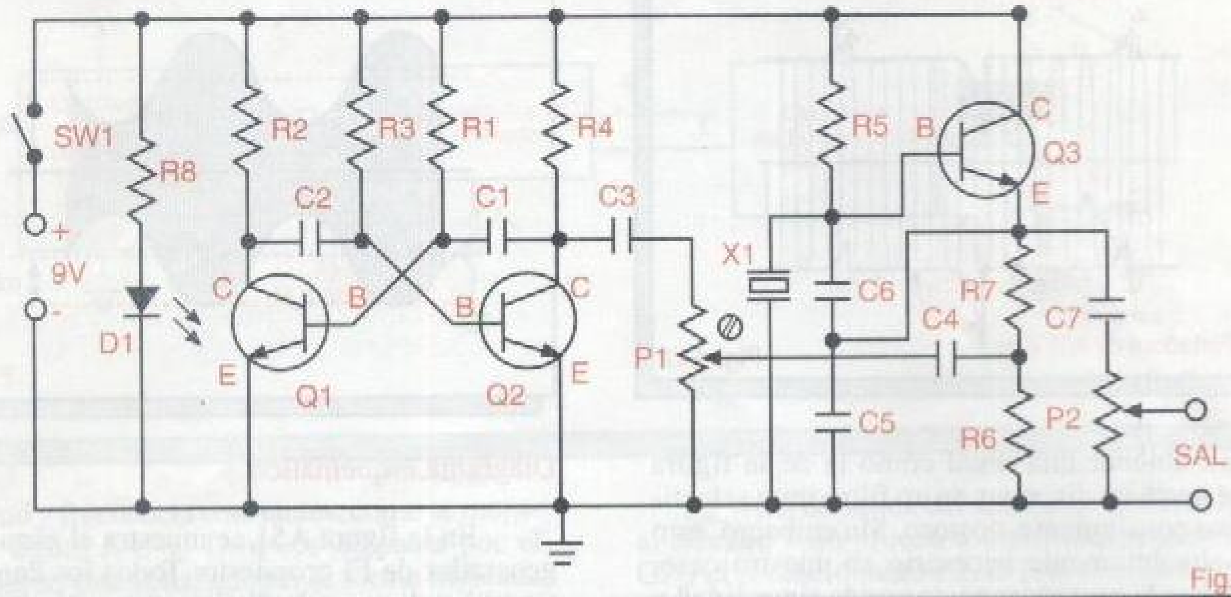

Fig. A51

A medida que C1 se carga, llega un momento en el cual el voltaje en la base de Q1 alcanza un valor suficiente para disparar este transistor y saturarlo. Este voltaje es del orden de 0.6 V.

Al conducir Q1, el voltaje de colector del mismo desciende desde 9 V hasta 0.1 V. Este cambio de voltaje se aplica a la base de Q2 a través de C2, sacando a este transistor del estado de conducción y llevándolo al de no conducción o corte.

Mientras tanto, C2 se carga a través de R3 en dirección del voltaje de alimentación. Cuando el voltaje aplicado por C2 a la base de Q2 llega a 0.6 V, este transistor se dispara y conduce. El ciclo se repite.

Como resultado del proceso anterior, en los colectores de Q1 y Q2 se obtienen dos ondas cuadradas idénticas, pero opuestas en fase. En nuestro caso, la señal de salida del oscilador de relajación se toma del colector de Q2 y se aplica al diferenciador.

El diferenciador, constituido por C3 y P1, es básicamente un filtro pasa altos con una constante de tiempo RC muy pequeña (~2 ms). Su función consiste en proporcionar a la entrada del oscilador una señal de modulación proporcional a la rata de cambio de la forma de onda de entrada.

Cuando se aplica una señal cuadrada a la entrada, el diferenciador genera unos picos más o menos pronunciados en los flancos o puntos de transición (ver figura A45). La amplitud de estos picos se controla mediante el potenciómetro P1. Esta operación afecta la profundidad de la modulación y el contenido armónico de la señal detectada en el receptor.

La señal de salida del diferenciador se aplica al oscilador de 455 kHz a través del condensador C4. Este oscilador está constituido por el transistor Q3, el cristal X1 y componentes asociados (R5, R6, R7, C5 y C6). La teoría general de los osciladores a cristal se discute en la lección 23 de este curso.

El cristal X1 determina la frecuencia de oscilación (455 kHz), mientras que el divisor capacitivo formado por C5 y C6 proporciona la realimentación positiva de voltaje necesaria para generar y mantener las oscilaciones. Las resistencias R5, R6 y R7 polarizan la etapa. La señal de audio aplicada al punto común de R6 y R7 se bate con la señal de RF, produciendo en el emisor de Q3 una onda modulada.

La señal AM de salida del oscilador de RF se aplica al potenciómetro de control de nivel de señal (P2) mediante el condensador C7 y se inyecta al circuito de utilización a través de un cable preferiblemente blindado. El LED D1 actúa como monitor de la batería de 9 V. La resistencia R8 limita la corriente a través de D1 a un valor seguro. SW1 es el interruptor general de alimentación del sistema.

Construcción

En la figura A52 se muestran el diagrama del circuito impreso y la guía de localización de componentes del generador de FI propuesto. La lista de partes se proporciona en el cuadro anexo.

Para una mejor presentación y conservación del sistema, se sugiere organizar la tarjeta de circuito impreso, la batería y los elementos de control y visualización (SW1, D1, P2) en un chasís o caja metálica y rotularlos con su función.

Realice las soldaduras con extremo cuidado, evitando hacer cortos entre pistas vecinas y siguiendo las recomendaciones de la actividad 3. Tenga especial cuidado en la instalación de Q1, Q2 y D1, por tratarse de componentes que deben llevar una orientación única.

Note la forma de identificar el cátodo (C) de D1 y los terminales correspondientes a la base (B), el colector (C) y el emisor (E) de Q1 y Q2. Al soldar X1, tenga la precaución de no calentar sus terminales más allá de lo absolutamente necesario porque puede destruirlo.

Calibración y prueba

El generador de FI prácticamente no requiere de calibración puesto que la frecuencia de la portadora de FI (455 kHz) la determina el cristal X1 y las características de la señal de modulación están predefinidas en el diseño del oscilador de relajación. Eventualmente, el único ajuste requerido es el del reóstato P1.

Este elemento controla la amplitud de la señal de audio aplicada al oscilador y determina el índice o profundidad de la modulación. Para una mejor comprensión de su función, se recomienda observar la señal de salida obtenida en el cursor de P2 en un osciloscopio.

Notará que con P1 en su posición mínima, la señal de salida del generador es una onda sinusoidal de 455 kHz prácticamente continua (sin modulación). Este caso corresponde a un índice cercano al 0%. A medida que se aumenta P1, aumenta también la amplitud de los picos de modulación, llegándose incluso a sobremodular la señal de salida, es decir a tener índices superiores al 100%.

Efectué la prueba del generador en un receptor de AM en buen estado. Aisle el conversor de RF e inyecte la señal modulada de 455 kHz a la entrada de la primera etapa de FI. Como punto de partida, sitúe P1 y P2 en sus posiciones intermedias.

Si la señal de prueba ha sido correctamente aplicada, en el parlante del receptor debe escucharse un sonido característico. Notará que la intensidad y riqueza tonal de este sonido varían con el punto de ajuste de P1 y P2.

Lo anterior indica que las etapas de FI, el detector y el amplificador de audio están trabajando correctamente. Si lo desea, puede realizar el seguimiento de la señal de prueba a través de las distintas etapas mediante un osciloscopio.

Guia de localización de componentes

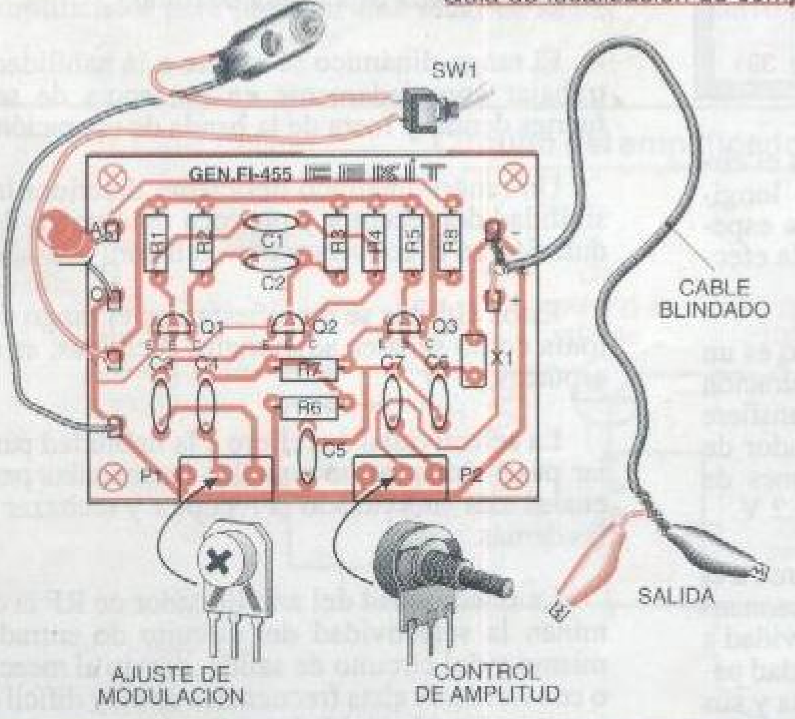

Lista de componentes

Resistencias (1/4 W, 5 %)
R1, R3 - 15 kΩ
R2, R4 - 470 Ω
R5 - 220 kΩ
R6, R7 - 1.2 kΩ
R8 - 1.5 kW
Condensadores
C1, C2, C3 - 0.047 mF/50 V, cerámico
C4, C7 - 0.1 mF/50 V, cerámico
C5 - 0,003 mF/50 V, cerámico
C6 - 680 pF/50 V, cerámico
Semiconductores
Q1, Q2, Q3 - Transistores NPN 2N3904
D1 - Diodo LED rojo estándar (5 mm)
Varios
1 Circuito impreso CEKIT Ref. GEN.FI.455
4 Terminales para circuito impreso (espadines)
1 Cristal cerámico de 455 kHz (X1)
1 Reóstato miniatura de 50 kΩ (P1)
1 Potenciómetro de 50 kΩ (P2)
1 Batería alcalina de 9 V
1 Conector para batería de 9 V
1 Interruptor de codillo miniatura (SW1)
1 Chasis de montaje CEKIT Ref. GEN.FI.455

Fig. A51

El amplificador de RF

Introducción

En esta lección estudiaremos diversos aspectos relacionados con el amplificador de RF del radio FM experimental CEKIT. Este circuito es la etapa de entrada del receptor y la que determina su figura de ruido.

La funcion del amplificador de RF consiste en captar las señales de FM transmitidas en la banda de 88 a 108 MHz, seleccionar una en particular y suministrarla amplificada al conversor de FI con el mínimo de ruido. En la figura 391 se muestra el diagrama de bloques correspondiente.

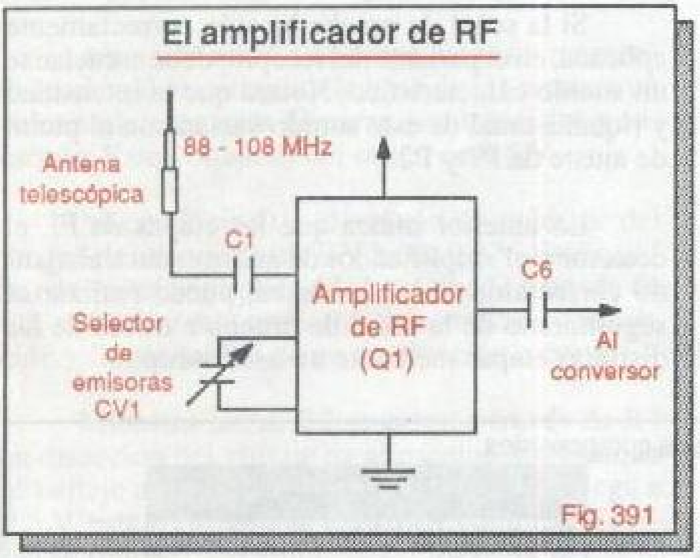

Fig. 391

La captación de señales de FM la realiza el circuito a través de una antena telescópica o de longitud variable. La selección de una frecuencia específica dentro de la banda de 88 a 108 MHz la efectúa un circuito sintonizado LC.

El amplificador de RF propiamente dicho es un transistor NPN (Q1), conectado en la configuración base común. La señal de salida de RF se transfiere al conversor de FI a través de un condensador de paso (C6). El transistor obtiene sus tensiones de polarización de una fuente estabilizada de +6.2 V.

El amplificador de RF está sintonizado tanto a la entrada como a la salida mediante circuitos resonantes LC. Estos circuitos determinan su selectividad a la frecuencia de operación, es decir, su habilidad para amplificar únicamente la portadora elegida y sus bandas laterales.

Antes de comenzar la descripción detallada del amplificador de RF de nuestro receptor es conveniente conocer las características y requisitos de diseño más importantes que debe reunir un amplificador de RF de FM para ser eficiente y no degradar la operación general del sistema.

El conocimiento de estos parámetros nos permitirá analizar y comprender hasta qué punto influye el amplificador de RF en el funcionamiento general del receptor y evaluar posibles fallas.

Después de todo, el amplificador de RF es la etapa por donde entra la señal de FM que queremos escuchar limpia y clara en el parlante.

De nada nos sirve disponer de un amplificador de audio sofisticado o de un demodulador impecable si la señal de entrada ha sido procesada con ruido, está distorsionada o tiene interferencia. Todos estos defectos dependen, en gran parte, del diseño del amplificador de RF.

Consideraciones de diseño de amplificadores de RF para FM

Las principales consideraciones de diseño que debe tener el amplificador de RF de un receptor de FM, tanto de radiodifusión comercial como de radioafición, son un alto rango dinámico, una buena selectividad y una baja figura de ruido.

El rango dinámico se refiere a la habilidad para trabajar apropiadamente en presencia de señales fuertes dentro y fuera de la banda de operación.

Un rango dinámico deficiente deteriora la sensibilidad del receptor y provoca fenómenos de modulación en el conversor o mezclador.

Estos últimos se manifiestan en el rango de sintonía como señales adicionales recibidas, es decir, espúreas.

La selectividad se refiere a la habilidad para dejar pasar únicamente aquellas frecuencias para las cuales está sintonizado el receptor y rechazar todas las demás.

La selectividad del amplificador de RF la determinan la selectividad del circuito de entrada del mismo y del circuito de salida, previo al mezclador o conversor. A altas frecuencias es muy difícil obtener una alta selectividad.

Además de restringir el paso de señales por fuera de la banda de sintonía, los circuitos de entrada y de salida del amplificador de RF deben rechazar también frecuencias imágenes, es decir, señales de interferencia dentro de la misma banda.

La selectividad general del receptor depende no solamente de la selectividad del amplificador de RF sino también de la selectividad del mezclador, de los amplificadores de FI y de las etapas de audio.

La figura de ruido se refiere a la habilidad para captar señales relativamente débiles y producir una señal de salida con un bajo nivel de ruido. La figura de ruido está estrechamente relacionada con la sensibilidad. Entre más baja sea la figura de ruido, más sensible es un receptor

El ruido generado dentro del amplificador de RF debe ser lo suficientemente bajo como para no enmascarar o degradar las señales débiles provenientes de la antena. La figura de ruido se minimiza utilizando dispositivos de bajo ruido como amplificadores de RF.

En general, los transistores de efecto de campo (FETs y MOSFETs) exhiben más bajas figuras de ruido que los transistores bipolares, bajo las mismas condiciones de operación.

Otras consideraciones de diseño importantes de un amplificador de RF para FM son su ganancia y su linealidad.

La ganancia es la cantidad de amplificación que recibe la señal de entrada y la linealidad la habilidad del amplificador para producir una señal de salida sin distorsión. Naturalmente, la señal de salida debe ser, en lo posible, una réplica exacta de la señal de entrada.

Como regla general, un amplificador de RF no debe tener más ganancia de la necesaria para obtener una aceptable figura de ruido. Más ganancia de la necesaria degradaría marcadamente el rango dinámico del receptor. Sin embargo, entre más alta sea la ganancia, más alta será la sensibilidad.

En algunos receptores de FM, el amplificador de RF está sometido a la acción de un circuito de control automático de ganancia (CAG). En algunos casos, esto es necesario pero no es conveniente. En realidad, un receptor bien diseñado no debe tener CAG en el amplificador de RF.

La aplicación de un voltaje de CAG al amplificador de RF degrada su figura de ruido y su rango dinámico. Esto se debe a que el voltaje de CAG altera las características de polarización del amplificador y cambia su punto de trabajo, no permitiendo que opere linealmente, en clase A.

Una vez enterados de estos conceptos, procederemos al análisis del amplificador de RF de nuestro receptor FM. Comenzaremos describiendo su diagrama esquemático, su operación tanto en presencia como en ausencia de señal y las funciones de cada uno de sus componentes.

Operación del amplificador de RF

En la figura 392 se muestra el diagrama esquemático de nuestro amplificador de RF. El dispositivo activo de esta etapa es el transistor Q1, conec-

Circuito del amplificador de RF

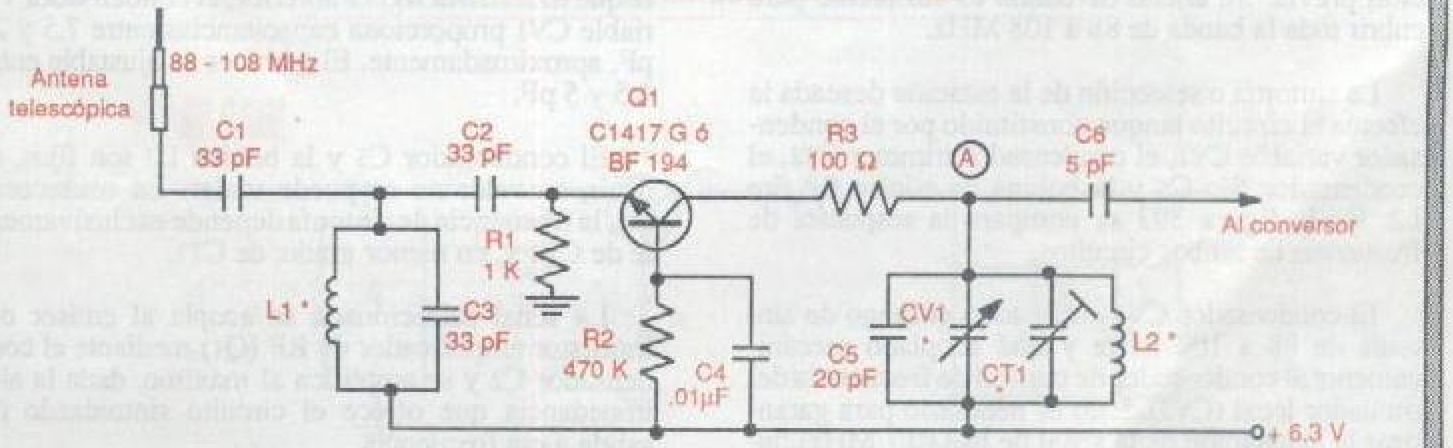

Fig. 392

tado como amplificador de voltaje base común. Esta configuración proporciona una operación con bajo ruido y un buen rango dinámico.

Note que el amplificador de RF está sintonizado tanto a la entrada como a la salida mediante circuitos resonantes LC. Esta doble sintonía le confiere una alta selectividad y un buen rechazo a las frecuencias imagen.

En radiodifusión FM la frecuencia imagen es la frecuencia ubicada 21.4 MHz por encima de la frecuencia sintonizada. Por ejemplo, la frecuencia imagen de 88.5 MHz es 88.5+21.4=109.9 MHz.

Si esta última señal lograra pasar al conversor, a través del amplificador de RF, produciría también una frecuencia intermedia de 10.7 MHz y las dos emisoras (la de 88.5 MHz y la de 109.9 MHz) se escucharían simultáneamente en el parlante.

El concepto de frecuencia imagen se explica en la Lección 26 de este curso para el caso de AM y se ampliará en la Lección 31, cuando estudiemos el conversor de FM.

Después de estas consideraciones generales, veamos cómo opera en detalle el amplificador de RF de nuestro receptor de FM.

A la antena telescópica convergen señales de todo tipo. El circuito resonante conformado por la bobina L1 y el condensador C3 presenta una alta impedancia a las señales de la banda de FM (88 a 108 MHz) y una baja impedancia a las señales por fuera de esta banda.

Las señales de FM se transfieren de la antena al circuito tanque L1C3 a través del condensador de paso C1. Este circuito no es el que realiza la sintonía. El solamente se encarga de realizar una selección previa. Su ancho de banda es suficiente para cubrir toda la banda de 88 a 108 MHz.

La sintonía o selección de la estación deseada la efectúa el circuito tanque constituido por el condensador variable CV1, el condensador trimmer CT1, el condensador fijo C5 y la bobina de núcleo de aire L2. En la figura 393 se compara la respuesta de frecuencia de ambos circuitos.

El condensador CV1 cubre todo el rango de sintonía de 88 a 108 MHz y está acoplado mecánicamente al condensador de control de frecuencia del oscilador local (CV2). Esto es necesario para garantizar la generación de la señal de FI (10.7 MHz) durante la mezcla o conversión.

El condensador CT1, de compensación o trimmer, está incorporado en la misma unidad que contiene a CV1, pero independiente de este último (figura 394). Se utiliza para realizar ajustes finos de sintonía, especialmente en los extremos de la banda.

Curvas de respuesta de frecuencia

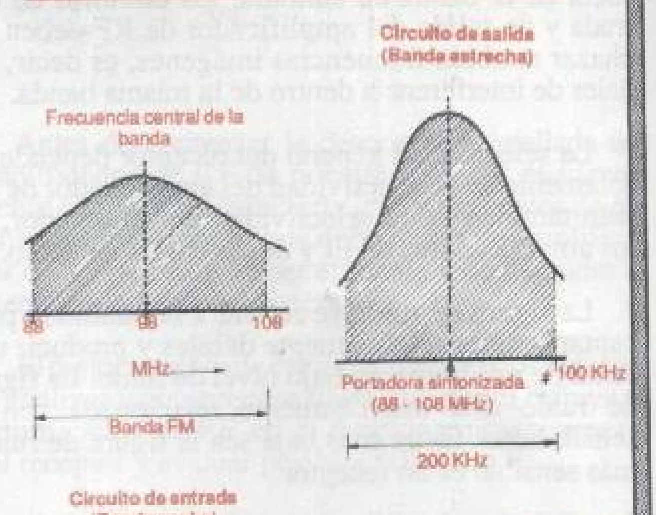

Fig. 393

Condensador de sintonía típico de FM

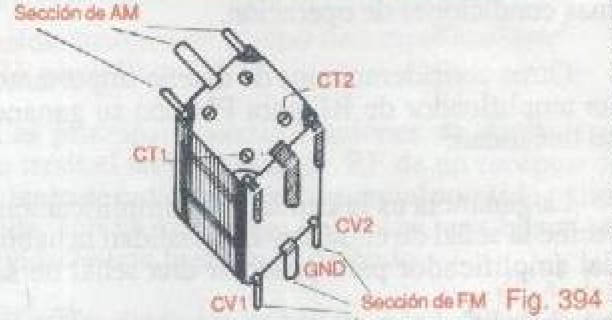

Fig. 394

En una unidad de sintonía típica para FM como la que muestra la figura anterior, el condensador variable CV1 proporciona capacitancias entre 7.5 y 20 pF, aproximadamente. El trimmer es ajustable entre 1.5 y 5 pF.

El condensador C5 y la bobina L2 son fijos, es decir, su valor no se puede variar. En consecuencia, la frecuencia de sintonía depende exclusivamente de CV1 y, en menor grado, de CT1.

La señal seleccionada se acopla al emisor del transistor amplificador de RF (Q1) mediante el condensador C2 y se amplifica al máximo, dada la alta impedancia que ofrece el circuito sintonizado de salida a esa frecuencia.

Lo anterior significa que si, por ejemplo, el circuito tanque de salida está sintonizado a 88.5 MHz, sólo la señal de entrada correspondiente a esta frecuencia recibirá amplificación y pasará a la etapa siguiente. Las demás señales sufrirán atenuación y, en consecuencia, serán ignoradas.

El transistor Q1 utilizado como amplificador de RF (C1417G o BF194) trabaja bien a altas frecuencias y tiene una baja figura de ruido. Esto último es importante para conseguir la característica de operación con bajo ruido que se espera de la etapa.

El transistor Q1 incrementa el nivel de la señal procedente de la antena antes de que alcance el conversor. La cantidad de amplificación la determinan los componentes asociados y puede variar de unos pocos a varios dBs (decibelios). Una amplificación excesiva afecta el rango dinámico.

Una vez que la señal de RF elegida se amplifica, ingresa al conversor, donde se heterodina o mezcla con la señal del oscilador local para establecer la señal de frecuencia intermedia (10.7MHz).

La señal de salida del amplificador de RF se transfiere al conversor a través del condensador C6. Esta señal aparece en los extremos del circuito de sintonía, es decir, entre los puntos "A" y "B" o, lo que es lo mismo, entre el punto "A" y tierra.

Recuerde que la fuente de alimentación es un corto para la señal. En consecuencia, el punto "B", que corresponde a la línea de alimentación de +6.2 V, está conectado dinámicamente a tierra, desde el punto de vista de la señal de RF.

El condensador C4 conecta dinámicamente a tierra la base de Q1 en presencia de la señal de RF. De este modo, la señal de entrada queda aplicada efectivamente entre emisor y tierra.

En la figura 395 se resume el comportamiento del circuito en condiciones dinámicas, es decir, en presencia de señal.

Se ignoran los condensadores C1, C2, C4 y C6 porque su reactancia capacitiva (XC) es despreciable a las frecuencias de operación del circuito, como se comprueba en el ejemplo 1.

Las resistencias R1, R2 y R3 polarizan adecuadamente el transistor Q1, es decir, fijan su punto de trabajo. El amplificador debe operar en clase A, en la región lineal de su característica, con el fin de no distorsionar la señal de entrada y producir una señal de salida fiel.

En general, los amplificadores de RF, al igual que los amplificadores de audio, necesitan estar correctamente polarizados para garantizar una operación confiable. Las tensiones de polarización, en este caso, se obtienen de la fuente de +6.2V.

La resistencia R2 polariza directamente la unión base-emisor de Q1. La corriente de reposo o de polarización de la base (I_{BQ}) la extrae el circuito del positivo de la fuente, a través de R2.

La resistencia R3 polariza inversamente la unión base-colector de Q1. La corriente de polarización o de reposo del colector (I_{CQ}) la extrae el circuito del positivo de la fuente a través de R3 y la resistencia interna de la bobina L2. Para efectos prácticos, esta última es despreciable (0Ω).

La resistencia R1 estabiliza térmicamente el punto de trabajo de Q1, evitando que las corrientes de reposo de la base (I_{BQ}) y del colector (I_{CQ}) varíen con la temperatura.

Circuito equivalente dinámico del amplificador de RF

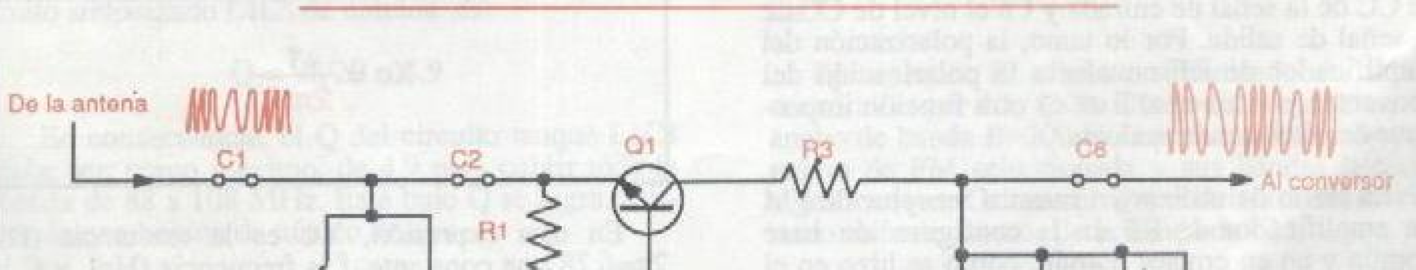

" Los condensadores de desacople C1, C2, C4 y C6 y la fuente de alimentación de CC son cortocircuitos para las fuentes de RF "

Fig. 395

Circuito equivalente estático del amplificador de RF

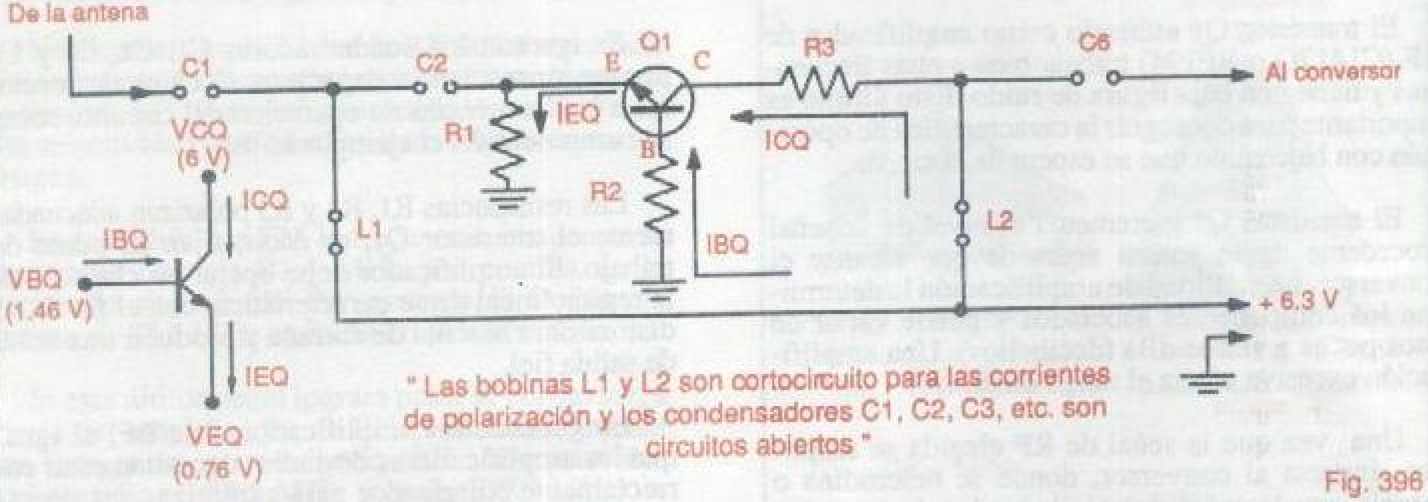

" Las bobinas L1 y L2 son cortocircuito para las corrientes de polarización y los condensadores C1, C2, C3, etc. son circuitos abiertos "

Fig. 396

En la figura 396 se resume el comportamiento del circuito en condiciones de polarización, de reposo o de CC, es decir, en ausencia de señal. También se indican los valores típicos de voltaje obtenidos entre base, emisor y colector, con respecto a tierra.

Observe que en la figura 396 se ignoran los efectos de todos los condensadores y de todas las bobinas presentes en el circuito original. Sólo se incluyen los efectos de las resistencias y de la fuente de alimentación. Las razones de esto se pueden resumir en los siguientes términos:

Los condensadores presentan una alta oposición o resistencia al paso de la CC y, en consecuencia, se comportan como circuitos abiertos. Las bobinas presentan una muy baja resistencia y, por lo tanto, se comportan como cortocircuitos.

Resulta entonces claro que C_2 bloquea el nivel de CC de la señal de entrada y C_6 el nivel de CC de la señal de salida. Por lo tanto, la polarización del amplificador de RF no afecta la polarización del conversor y viceversa. Esta es otra función importante de estos condensadores.

La razón de utilizar en nuestro receptor de FM un amplificador de RF en la configuración base común y no en emisor común, como se hizo en el receptor AM, no es por ganancia sino por estabilidad. Como sabemos, el montaje base común suministra menos ganancia de voltaje que el emisor común.

Sin embargo, el circuito base común está menos sujeto a autooscilaciones. Recordemos que entre más alta es la frecuencia de operación, más alta es la probabilidad de que un amplificador oscile.

En AM trabajabamos con frecuencias entre 530 y 1600 KHz. Ahora estamos trabajando con frecuencias entre 88 y 108 MHz. Por lo tanto, los riesgos de inestabilidad son mayores. Por este motivo, deben prevenirse y la solución más obvia es adoptar una configuración amplificadora estable.

Después de esta descripción cualitativa de la operación del amplificador de RF, vamos a presentar algunos ejemplos numéricos sencillos, los cuales nos permitirán analizar cuantitativamente estos y otros aspectos de su comportamiento.

Ejemplo 1. Calcule la reactancia capacitiva XC de los condensadores C_1, C_2, C_4 y C_6 a una frecuencia de portadora de 100 MHz.

Solución. Tal como vimos en la Lección 7 (página 74), la reactancia capacitiva XC de un condensador está dada por la fórmula:

$$X_C = \frac{1}{2\pi f C}$$

En esta expresión, XC es la reactancia (Ω), 2π=6.28 una constante, **f** la frecuencia (Hz) y C la capacitancia (F).

En nuestro caso, $f=100\ \text{MHz}=1\times10^8$ Hz.

Para $C=C_1=C_2=33\ \text{pF} = 33\times10^{-12}$ F:

$X_{C1,2} = 1/(6.28\times1\times10^8\times33\times10^{-12}) = 48\Omega$

Para $C= C_4=0.01\ \mu\text{F}=1\times10^{-8}$ F:

$X_{C4} = 1/(6.28\text{x}1\text{x}10^8\text{x}1\text{x}10^{-8}) = \mathbf{0.16\ \Omega}$

Para $C=C_6=5$ pF$=5\text{x}10^{-12}$F:

$X_{C5} = 1/(6.28\text{x}1\text{x}10^8\text{x}5\text{x}10^{-12}) = \mathbf{318\ \Omega}$

Estos bajos valores de reactancia indican que los condensadores C1, C2, C4 y C5 prácticamente no presentan oposición al paso de la señal. Por esta razón se ignoran en el circuito equivalente de señal de la figura 395.

Ejemplo 2. El circuito sintonizado L1C3 debe tener un ancho de banda suficiente para cubrir el rango de 88 a 108 MHz. La frecuencia central de esta banda es 98 MHz. Calcule el Q o factor de calidad de este circuito y el valor de L1 para esta frecuencia. Suponga una resistencia interna de 10Ω.

Solución. Como vimos en la Lección 24 (página 180), el ancho de banda de un circuito sintonizado LC está dado por la fórmula:

$$B = \frac{Fo}{Q}$$

En esta expresión, B es el ancho de banda (Hz), Fo es la frecuencia central o de resonancia de la banda (Hz) y Q el factor de calidad del circuito. Por lo tanto:

$$Q = \frac{Fo}{B}$$

En nuestro caso, Fo=98 MHz y B=108-88 = 20 MHz. Por consiguiente, el factor de calidad del circuito sintonizado L1C3 de entrada es:

$$Q = 98/20 = \mathbf{4.9}$$

En consecuencia, el Q del circuito tanque L1C3 debe ser, como máximo, de 4.9 para cubrir toda la banda de 88 a 108 MHz. Este bajo Q se logra utilizando una bobina de núcleo de aire.

En nuestro receptor de FM, L1 se obtiene devanando 5 espiras de alambre esmaltado #20 sobre un núcleo de cartón de 4 mm de diámetro. Esta configuración provee el bajo Q requerido. En la figura 397 se muestra el aspecto de esta bobina.

Como vimos en la Lección 24 (página 180), el Q o factor de calidad de un circuito sintonizado LC está dado por la fórmula:

Especificaciones de la bobina L1

Especificaciones	Valor
Diámetro del núcleo (D)	4 mm
Número de espiras (N)	5
Calibre del alambre	# 20
Diámetro del alambre (ø)	1.2 mm
Longitud devanado (S)	7 mm
Inductancia (L)	.08 μH
Q a 98 MHz (Qo)	5

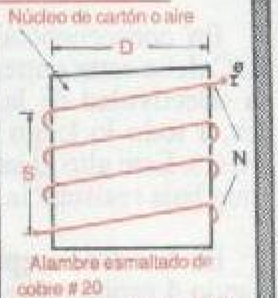

Fig. 397

$$Q = \frac{X_L}{Rs} = \frac{2\pi FoL}{Rs}$$

En esta expresión, 2π=6.28, Q es el factor de calidad, X_L la reactancia inductiva (Ω) de la bobina, Rs su resistencia interna (Ω), L su inductancia (H) y Fo la frecuencia central o de resonancia de la banda (Hz). Por lo tanto:

$$L = \frac{QRs}{2\pi Fo}$$

En nuestro caso, Q=4.9, Rs=10Ω, 2π=6.28 y Fo=98 MHz=$98\text{x}10^6$ Hz. Por consiguiente:

$L = (4.9\text{x}10)/(6.28\text{x}98\text{x}10^6) = \mathbf{0.08\ \mu H}$

Este debe ser el valor de inductancia de L1 necesario para proporcionar una frecuencia central de resonancia de 98 MHz y un Q de 4.9 a esa frecuencia.

Ejemplo 3. El circuito sintonizado de salida, constituido por L2, CV1, CT1 y C5, debe tener un ancho de banda B=200 KHz para amplificar la portadora de FM seleccionada y sus bandas laterales significativas. Calcule el Q que debe tener este circuito en los extremos de la banda de FM.

Solución. El límite inferior de la banda de FM es 88 MHz. A esta frecuencia, el factor de calidad del circuito es mínimo y vale:

$\mathbf{Q(mín)} = Fo(mín)/B = (88\text{x}10^6)/(200\text{x}10^3) = \mathbf{440}$

El límite superior de la banda de FM es 108 MHz. A esta frecuencia, el factor de calidad del circuito es máximo y vale:

$$Q(\text{máx}) = Fo(\text{máx})/B = (108 \times 10^6)/(200 \times 10^3) = \mathbf{540}$$

En consecuencia, el Q del circuito sintonizado L2C debe estar entre 440 y 540 para proveer una alta selectividad de la portadora y sus bandas laterales, a todo lo largo de la banda de FM (88 a 108 MHz). Este alto Q se logra utilizando una bobina de muy baja resistencia interna, es decir, sin pérdidas.

En nuestro receptor de FM, L2 se obtiene devanando 4 espiras de alambre esmaltado #20 sobre un núcleo aislante (aire) de 4 mm de diámetro. Esta configuración provee el alto Q requerido. En la figura 398 se muestra el aspecto de esta bobina.

Especificaciones de la bobina L2

Especificaciones	Valor
Diámetro del núcleo (D)	4 mm
Número de espiras (N)	4 mm
Calibre del alambre	# 20
Diámetro del alambre (σ)	1.2 mm
Longitud devanado (S)	5 mm
Inductancia (L)	.07 µH
Q a 98 MHz (Qo)	540

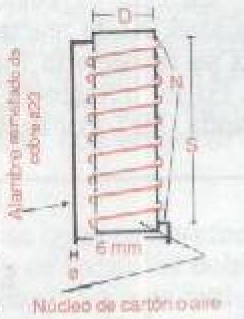
D
N
S
Alambre esmaltado de cobre #20
6 mm
σ
Núcleo de cartón o aire

Fig. 398

Antenas para recepción de FM

La longitud de la antena es una consideración importante y debe estar de acuerdo con la frecuencia de la emisora que se desea sintonizar. Teóricamente, la longitud de la antena debe ser igual a la mitad de la longitud de onda de la portadora.

El concepto de longitud de onda se explica en la Introducción de este curso. Una señal de 100 MHz, por ejemplo, tiene una longitud de onda de 3 metros. Para sintonizar eficientemente una emisora que transmite con esta frecuencia, la antena de recepción deberá tener 1.5 metros, es decir, λ/2.

Para una de 88 MHz, se necesitaría una antena de 1.7 m, para una de 108 MHz se necesitaría una antena de 1.4 m, y así sucesivamente. Por esta razón, para cubrir toda la banda de FM, el receptor utiliza una antena telescópica, es decir, de longitud variable (figura 399A).

La longitud de λ/2 se denomina longitud eléctrica. La longitud física de antena requerida es realmente la mitad de este valor, a sea λ/4. Esto se debe a que el sistema de tierra de la antena crea eléctricamente la antena imagen, es decir, el otro tramo de λ/4 faltante.

Antenas de recepción de FM prácticas

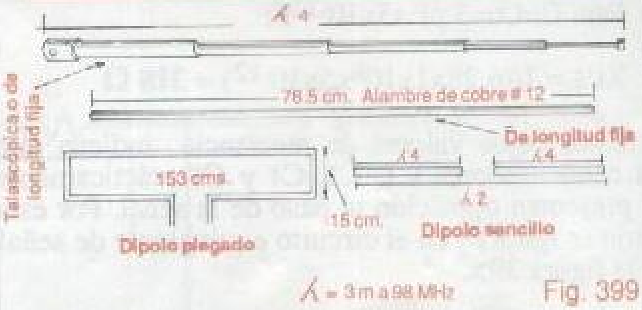
λ/4
Telescópica o de longitud fija
78.5 cm. Alambre de cobre # 12
De longitud fija
153 cms.
15 cm.
Dipolo plegado
λ/4
λ/4
λ/2
Dipolo sencillo
λ = 3 m a 98 MHz

Fig. 399

Este tipo de antena se denomina antena vertical de λ/4 y es muy utilizado. Una antena de este tipo capta igualmente señales en todas las direcciones, excepto a lo largo de su eje (figura 400A). Se dice, entonces, que tiene polarización vertical.

Características de recepción de las antenas de FM

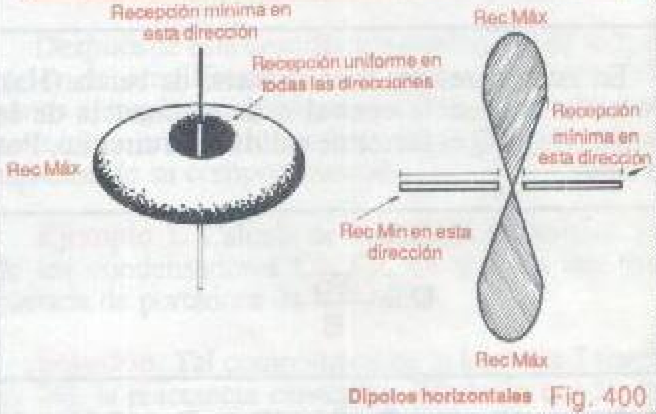
Recepción mínima en esta dirección
Recepción uniforme en todas las direcciones
Rec Máx
Rec Máx
Recepción mínima en esta dirección
Rec Mín en esta dirección
Rec Máx
Dipolos horizontales

Fig. 400

Si no se dispone de una antena telescópica, puede utilizarse un alambre de longitud fija, cortado para 98 MHz, es decir, para la frecuencia central de la banda de FM. En este caso, la longitud de onda es de 3.06 m. Por lo tanto, el alambre de antena deberá ser de 76.5 cm (λ/4) de largo (figura 399B).

Otra alternativa al uso de una antena vertical telescópica o fija es emplear un dipolo de λ/2, simple o plegado (figuras 399C y D). Un dipolo de λ/2 está formado por dos secciones horizontales de alambre de longitud igual a λ/4 cada uno.

Los dipolos tienen un patrón de recepción como el que muestra la figura 400B. La recepción es mínima en la misma dirección del eje de la antena y máxima en la dirección perpendicular al mismo. Por esta razón deben orientarse apropiadamente.

El conversor de FI

Introducción

Continuando con el análisis del receptor experimental de FM CEKIT, en esta lección estudiaremos el conversor de FI o mezclador. Esta etapa es el circuito previo a los amplificadores de FI y el que determina el rango dinámico del receptor.

El conversor de FI lo constituyen dos circuitos: un oscilador local y el mezclador propiamente dicho (figura 401). Su función consiste en mezclar o heterodinar la señal suministrada por el oscilador local con la señal de entrada de RF y producir una señal de frecuencia intermedia (FI) de 10.7 MHz.

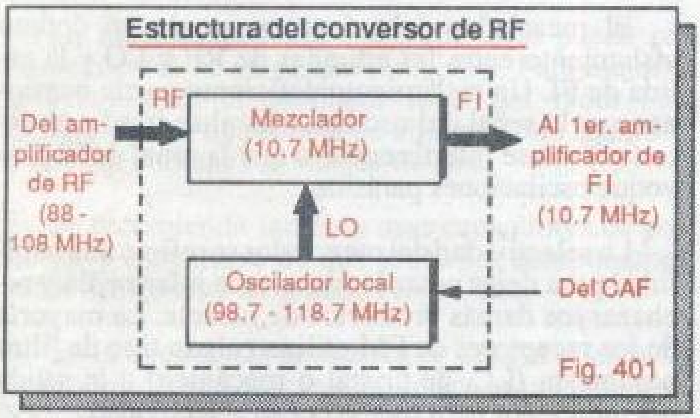
Estructura del conversor de RF
RF
FI
Del amplificador de RF (88 - 108 MHz)
Mezclador (10.7 MHz)
Al 1er. amplificador de FI (10.7 MHz)
LO
Oscilador local (98.7 - 118.7 MHz)
Del CAF
Fig. 401

El oscilador local produce una onda sinusoidal pura, de amplitud fija y frecuencia variable. La frecuencia del oscilador local depende de la frecuencia de la señal suministrada por el amplificador de RF y está siempre 10.7 MHz por encima del valor de portadora de esta última.

Por ejemplo, si se sintoniza una señal de RF de 98 MHz, la frecuencia del oscilador local deberá ser de 98+10.7=108.7 MHz, si se sintoniza una frecuencia de 93 MHz, el oscilador deberá trabajar a 93+10.7=103.7 MHz , y así sucesivamente.

En consecuencia, para cubrir toda la banda de FM (88 a 108 MHz), la frecuencia del oscilador local debe variar continuamente entre 98.7 y 118.7 MHz y estar sincronizada con la frecuencia de la portadora captada.

Para lograr este sincronismo, la frecuencia del oscilador local la controla un condensador variable acoplado mecánicamente al condensador de sintonía del amplificador de RF (figura 402).

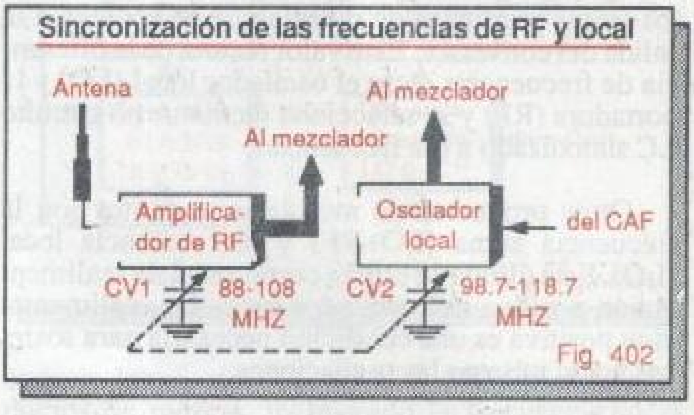
Sincronización de las frecuencias de RF y local

Fig. 402

El condensador CV1 controla la frecuencia de resonancia del circuito de sintonía del amplificador de RF y CV2 la del circuito determinador de frecuencia del oscilador local. De este modo, al variar la primera, lo hace también la segunda, en la misma cantidad y en la misma dirección.

La operación de mezcla la realiza un transistor bipolar (Q1), conectado como un amplificador no lineal en la configuración base común. Este transistor cumple simultáneamente las funciones de mezclador y oscilador. Por esta razón se denomina *conversor*.

La forma como se realiza la mezcla de la señal de RF con la del oscilador local en el receptor de FM se basa en los mismos principios generales expuestos en la Lección 26. Recuerde: dos señales aplicadas a un circuito no lineal siempre se mezclan y generan productos de mezcla.

El efecto neto del proceso de conversión o mezcla en FM es trasladar la modulación de la portadora original (88 a 108 MHz) a una portadora interna más baja (10.7 MHz). Esta última es la frecuencia intermedia (FI) del proceso (figura 403). Recordemos que en el sistema AM, el valor estándar de la FI es 455 KHz.

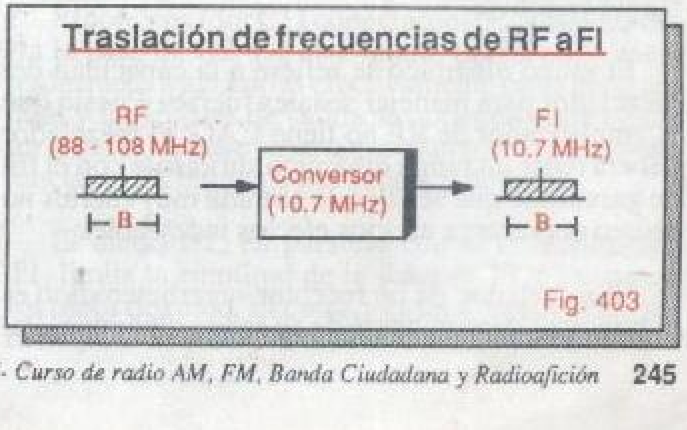
Traslación de frecuencias de RF a FI

Fig. 403

La utilización de una portadora interna más baja que la original permite obtener una alta selectividad y una alta ganancia en las etapas de FI y mantenerlas constantes.

La frecuencia intermedia de 10.7 MHz es uno de los tantos productos de mezcla que se generan a la salida del conversor. Este valor resulta de la diferencia de frecuencias entre el oscilador local (LO) y la portadora (RF) y se selecciona mediante un circuito LC sintonizado a esa frecuencia.

Otros productos de mezcla importantes son la frecuencia suma (LO+RF) y la frecuencia local (LO). Esta última se utiliza como señal de realimentación positiva del oscilador local. La realimentación positiva es una condición necesaria para sostener por sí mismas las oscilaciones.

El oscilador local del conversor está también gobernado por el circuito de control automático de frecuencia o CAF. La función del CAF es mantener constante la frecuencia del oscilador local. Su operación detallada se estudia en la Lección 34.

Una vez trasladada la modulación de la portadora de RF a la portadora interna de FI, esta última señal es amplificada y finalmente detectada. En la próxima lección estudiaremos el proceso de amplificación de la señal de FI. En la Lección 33 se analiza el proceso de detección de la señal FM.

Antes de analizar la operación del conversor de FI del receptor de FM es conveniente conocer las características generales involucradas en el diseño de mezcladores y osciladores locales.

Un mezclador o un oscilador deficientes pueden arruinar el mejor de los receptores, incluso si todas las demás etapas están trabajando óptimamente. Por esto es importante conocer sus requisitos de operación y las formas de hacerlos más eficientes.

Características generales del mezclador

Las principales características que debe reunir el mezclador de un receptor superheterodino de FM son un alto rango dinámico, una baja figura de ruido, una buena capacidad de aislamiento y ser altamente selectivo.

El rango dinámico se refiere a la capacidad del mezclador para manejar señales fuertes. Puesto que el amplificador de RF no tiene CAG, el mezclador deberá tener un rango dinámico suficiente con el fin de garantizar que señales de entrada muy fuertes no causen sobrecarga ni otros efectos indeseables.

El mezclador de un receptor superheterodino es el principal determinante de su rango dinámico. Para una óptima operacion, el mezclador deberá recibir sólo la cantidad de señal necesaria para superar el nivel de ruido propio de la etapa.

Cuando se aplican niveles excesivos de señal, se producen fenómenos como la modulación cruzada y la intermodulación (IMD). Un rango dinámico deficiente afecta también la sensibilidad del receptor y provoca la aparición de espúreas.

Una excesiva ganancia del amplificador de RF causa que el mezclador se sature en presencia de señales moderadas y fuertes. Durante mucho tiempo la solución a este problema fue utilizar el proceso de *doble conversión*, que se estudia en la Lección 32.

La figura de ruido de un receptor de FM la establecen primariamente el amplificador de RF y el mezclador. Para garantizar una baja figura de ruido, se deben utilizar dispositivos activos de bajo ruido en estas etapas, preferiblemente transistores de efecto de campo o FET's.

El mezclador deberá proporcionar un óptimo aislamiento entre las entradas de RF y LO y la salida de FI. Un aislamiento deficiente puede ocasionar que la señal del oscilador local se irradie al exterior y cause interferencia o que la señal de FI provoque oscilaciones parásitas.

La selectividad del mezclador se refiere a su habilidad para dejar pasar la frecuencia intermedia y rechazar los demás productos de mezcla. La mayoría de los receptores de FM utilizan algún tipo de filtro pasabanda (LC, de cristal o mecánico) a la salida del mezclador para establecer su selectividad.

Este filtro sirve también para rechazar productos de mezcla no deseados.

Otra consideración de diseño importante es el tipo de dispositivo activo utilizado como mezclador. Nuestro sencillo receptor de FM utiliza un transistor bipolar para este propósito.

Los receptores más avanzados, especialmente los de uso para estaciones de radioaficionado, utilizan diodos Schottky y transistores de efecto de campo (FETs y MOSFETs) como mezcladores. Estos dispositivos son tan sensibles que el amplificador de RF se puede eliminar en muchos casos.

Características generales del oscilador local

La señal de inyección del mezclador, procedente del oscilador local, debe ser espectralmente pura y altamente estable, estar libre de ruido y espúreas y tener una amplitud adecuada. Así mismo, debe estar convenientemente aislada de la señales de RF y FI.

Esto último es importante porque el oscilador local tiende a sincronizar su frecuencia con la de las señales que le llegan, originando oscilaciones parásitas. Cuando la frecuencia del oscilador sufre una variación debida a factores externos se produce una *deriva* o desplazamiento de frecuencia.

En el oscilador local del radio FM CEKIT se utilizan básicamente dos métodos para minimizar la deriva de frecuencia: el primero emplea una trampa de FI en el mismo conversor y el segundo un circuito de control automático de frecuencia o CAF.

La trampa de FI aísla la señal del oscilador local de la señal de FI. El CAF corrije automáticamente su frecuencia y la restablece a su valor nominal.

Otra característica importante que debe tener el oscilador local es su velocidad de respuesta. Las oscilaciones deben arrancar tan pronto se aplique el voltaje de alimentación y sostenerse indefinidamente sin ningún tipo de ayuda externa.

La inestabilidad del oscilador local puede ser consecuencia de un diseño deficiente. Para mejorar la estabilidad, el oscilador local del radio FM CEKIT utiliza una fuente de alimentación regulada y emplea bobinas con núcleo de aire.

Se recomienda también usar capacitores de poliéster y asegurarse de que todos los componentes estén firmemente asegurados en su sitio.

En la Lección 25 se analiza en detalle el tema de la estabilidad de frecuencia de los osciladores de RF.

El ruido del oscilador se puede mantener a un valor aceptable utilizando circuitos sintonizados de alto Q. Entre más alto sea el Q, más estrecho es el ancho de banda y, por lo tanto, más bajo es el nivel de ruido en el voltaje de salida.

El oscilador local debe también poder suministrar al mezclador una señal suficientemente fuerte para todas las frecuencias de su rango de operación. Esta condición es importante para garantizar la operación no lineal del dispositivo mezclador. Si éste opera linealmente, no genera productos de mezcla.

Operación del conversor de FI

En la figura 404 se muestra el diagrama detallado de bloques del conversor de FI. El circuito recibe dos señales de entrada: una modulada, proveniente del amplificador de RF y otra sin modular o continua, generada localmente (LO).

El mezclador combina o heterodina las dos señales de entrada y produce a su salida la señal de frecuencia intermedia (FI) de 10.7 MHz y otros productos de mezcla, incluyendo la frecuencia local (LO). La señal de FI conserva la misma modulación de la señal de RF.

El conversor de FI en bloques

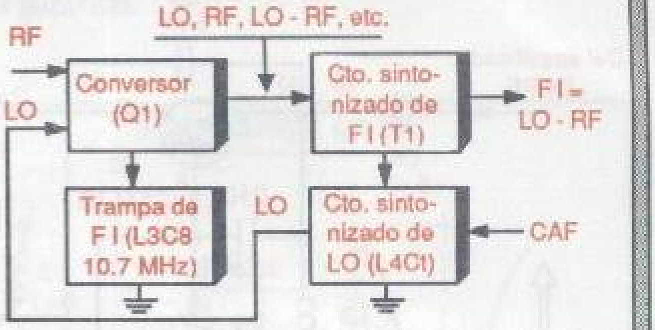
LO, RF, LO - RF, etc.
RF
LO
Conversor (Q1)
Cto. sintonizado de FI (T1)
FI = LO - RF
Trampa de FI (L3C8 10.7 MHz)
LO
Cto. sintonizado de LO (L4Ct)
CAF

Fig. 404

El circuito de sintonía del mezclador selecciona la señal de FI y la transfiere a la etapa siguiente, ignorando las demás. El circuito de sintonía del oscilador local selecciona la señal LO y la realimenta positivamente a la entrada. La trampa de FI envía a tierra cualquier rastro de FI presente a la entrada.

En la figura 405 se muestra el circuito esquemático del conversor de FI. El dispositivo activo de esta etapa es el transistor Q2, el cual actúa simultáneamente como mezclador y como oscilador local. Se utiliza una tensión de alimentación de +6.2 V.

La señal de RF se inyecta al emisor de Q2 a través del condensador C6. La señal local (LO) se aplica al mismo electrodo a través del condensador C9. Estos condensadores sirven únicamente de acople y su reactancia es muy baja a las frecuencias de operación involucradas.

El transformador T1 selecciona la señal de FI y la transfiere a las etapas de FI. El circuito tanque asociado a la bobina L4 selecciona la señal local y la realimenta positivamente a la entrada del conversor. La bobina L3 y el condensador C8 configuran la trampa de FI.

El condensador C11 conecta dinámicamente a tierra la base de Q2. De este modo, las señales de entrada quedan aplicadas efectivamente entre emisor y base y las señales de salida se obtienen entre colector y base. La base sirve como punto común de referencia de los circuitos de entrada y de salida.

El diodo D1, en paralelo con el transformador T1, limita la amplitud de la señal de FI, eliminando los picos de ruido. El condensador C10 se utiliza para mejorar la estabilidad del oscilador.

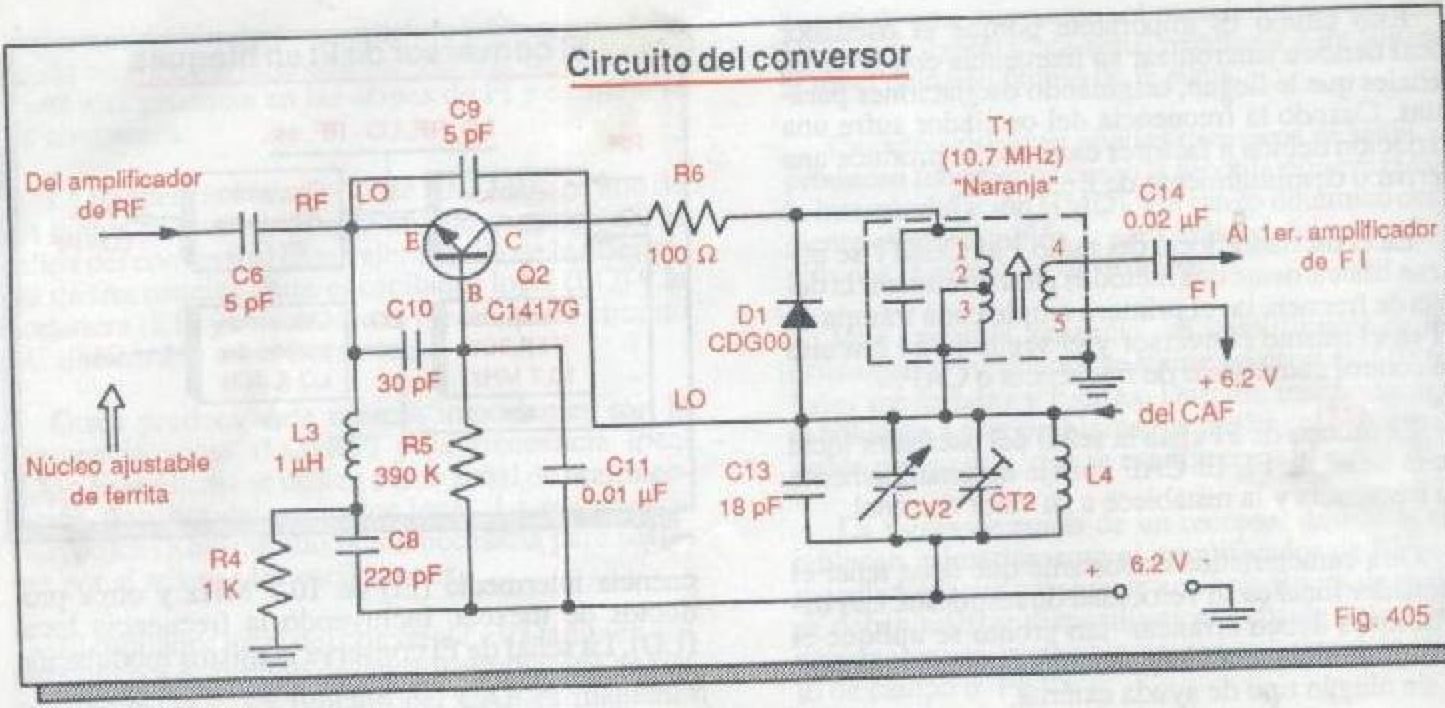
Circuito del conversor
C9
5 pF
Del amplificador
de RF
RF
LO
R6
100 Ω
T1
(10.7 MHz)
"Naranja"
C14
0.02 µF
Al 1er. amplificador
de F I
F I
E
C
Q2
C1417G
B
C6
5 pF
C10
30 pF
D1
CDG00
1
2
3
4
5
+ 6.2 V
del CAF
LO
Núcleo ajustable
de ferrita
L3
1 µH
R5
390 K
C11
0.01 µF
C13
18 pF
CV2
CT2
L4
C8
220 pF
R4
1 K
+ 6.2 V -
Fig. 405

La resistencia R6 reduce el riesgo de oscilaciones parásitas, muy comunes en los osciladores a transistores, y aísla el colector de Q2 de tierra. La resistencia R4 hace lo propio con el emisor y sirve de carga a las señales de RF y local.

Estas resistencias cumplen también la importante función de polarizar adecuadamente el conversor.

El transformador sintonizado de FI

El transformador T1 (figura 406) está sintonizado a 10.7 MHz. Su función consiste en seleccionar la señal de frecuencia intermedia a la salida del mezclador y transferirla al primer amplificador de FI. Proporciona también el necesario acople de impedancias entre ambas etapas.

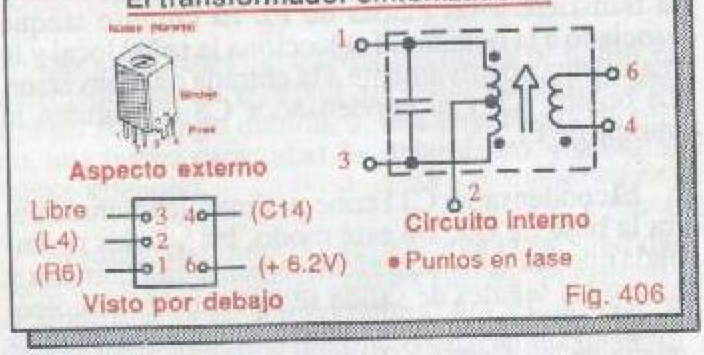
El transformador sintonizado F I
Aspecto externo
1
6
3
4
2
Circuito interno
Libre
(L4)
(R6)
3
4
2
1
6
(C14)
(+ 6.2V)
• Puntos en fase
Visto por debajo
Fig. 406

Este transformador se distingue por poseer un núcleo de ferrita de color naranja. Posee 5 terminales y está blindado magnéticamente con el fin de evitar que irradie energía de FI hacia los circuitos vecinos o reciba señales espúreas.

El oscilador local

El circuito de sintonía del oscilador local lo constituyen la bobina L4, el condensador variable CV2, el condensador "trimmer" CT2 y el condensador fijo C13, en asocio con el CAF (figura 407).

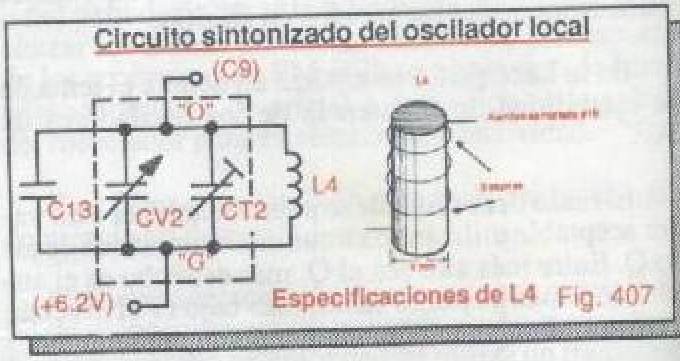
Circuito sintonizado del oscilador local
(C9)
"O"
C13
CV2
CT2
L4
"G"
(+6.2V)
Especificaciones de L4
Fig. 407

El condensador CV2 trabaja en tándem con el condensador CV1 del amplificador de RF. La bobina L4 se obtiene devanando 3 espiras de alambre esmaltado #19 sobre un núcleo de aire de 4 mm. No se utiliza un núcleo magnético por cuestiones de estabilidad de frecuencia.

El circuito tanque de L4 ofrece una impedancia muy alta a la frecuencia propia del oscilador local y una impedancia muy baja a la frecuencia intermedia de 10.7 MHz. El transformador sintonizado T1 se comporta en forma exactamente contraria.

Para los demás productos de mezcla, ambos circuitos presentan una impedancia muy baja. Puesto que los mismos se encuentran en serie con el colector, la señal de FI desarrolla un voltaje máximo

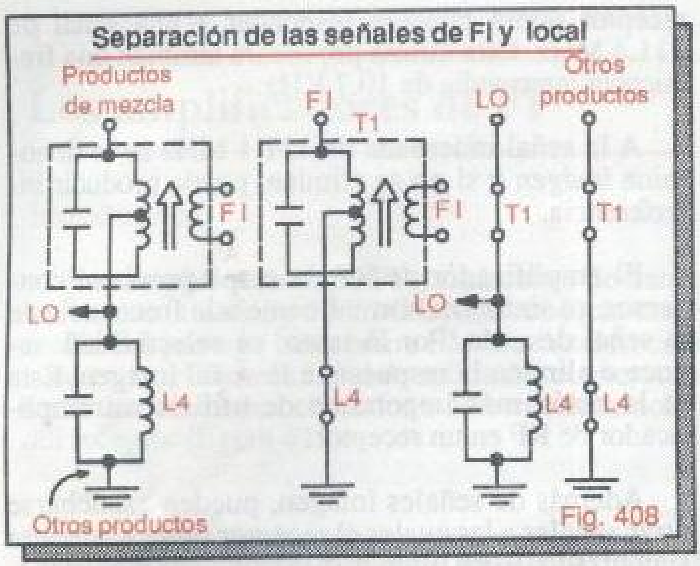
Separación de las señales de FI y local
Productos
de mezcla
FI
T1
LO
Otros
productos
FI
FI
T1
T1
LO
LO
L4
L4
L4
L4
Otros productos

Fig. 408

sobre T1 y la señal local un voltaje máximo sobre L4. En la figura 408 se ilustra este concepto.

El CAF (control automático de frecuencia) es, básicamente, un varactor o diodo de capacitancia variable. Su función es mantener constante la frecuencia del oscilador local. La acción del CAF se estudia en la Lección 34. El CAF se acopla capacitivamente al oscilador a través de C12.

El circuito correspondiente al oscilador local se muestra en la figura 409. Se obtiene del circuito conversor original reemplazando la trampa de FI por un circuito abierto y el transformador sintonizado T1 por un cortocircuito. Así se comportan estos componentes ante la frecuencia local de oscilación.

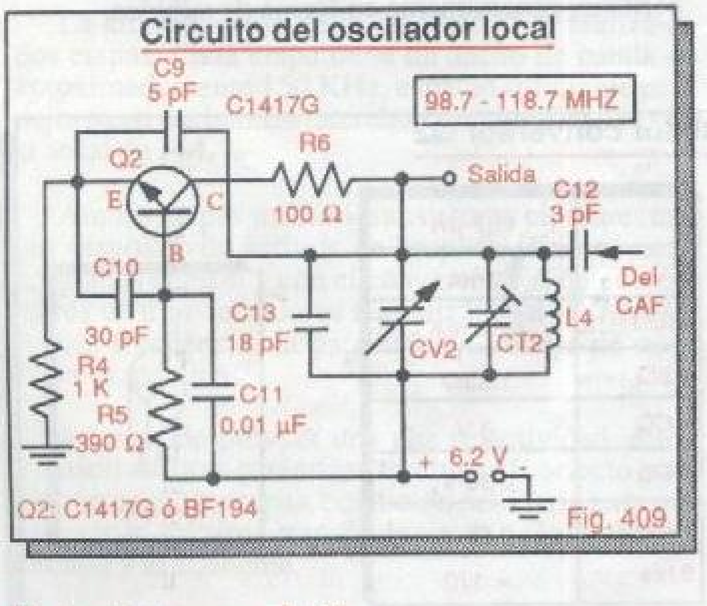
Circuito del oscilador local
C9
5 pF
C1417G
98.7 - 118.7 MHZ
R6
Q2
Salida
E
C
100 Ω
C12
3 pF
B
C10
Del
CAF
C13
18 pF
L4
30 pF
CV2
CT2
R4
1 K
R5
C11
390 Ω
0.01 µF
+ 6.2 V -
Q2: C1417G ó BF194

Fig. 409

El circuito trampa de FI

La bobina L3 y el condensador C8 configuran el circuito trampa de FI, sintonizado a 10.7 MHz (figura 410). Su función consiste en enviar a tierra cualquier rastro de la señal de FI presente en el circuito de entrada, evitando la creación de oscilaciones parásitas.

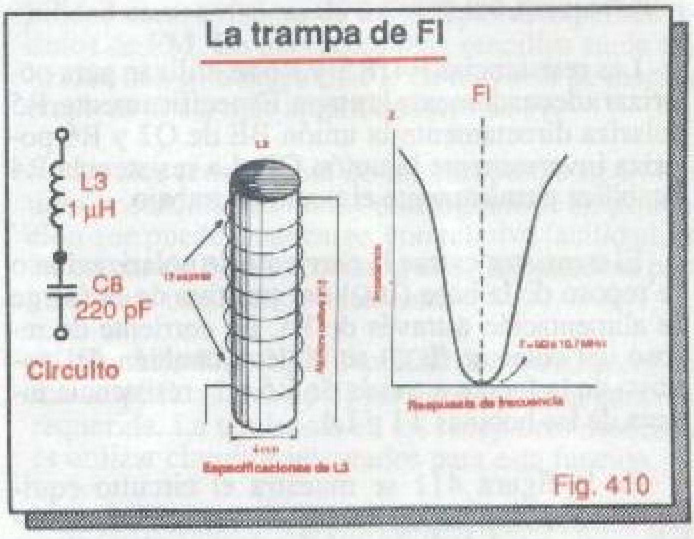
La trampa de FI
L3
1 µH
C8
220 pF
Circuito
FI
Respuesta de frecuencia
Especificaciones de L3

Fig. 410

A la frecuencia de 10.7 MHz (FI), el circuito resonante serie L3C8 presenta su mínima impedancia posible, comportándose como un cortocircuito a esa frecuencia. Bajo esta circunstancia, el emisor de Q2 queda conectado dinámicamente a tierra.

Puesto que las frecuencias de RF y de LO involucradas son relativamente altas comparadas con la FI de 10.7 MHz, la presencia del circuito trampa no las afecta, comportándose ante ellas como un circuito abierto.

Recuerde que el rango de la señal de RF se extiende desde 88 hasta 108 MHz y el de la señal local desde 98.7 hasta 118.7 MHz. Estas frecuencias están muy lejos del ancho de banda útil del circuito trampa L3C8. Por esta razón son ignoradas.

La señal de FI debe mantenerse confinada al circuito de colector de Q2. Cuando esta señal se realimenta a la entrada, interactúa con las señales de RF y LO, provocando efectos de modulación indeseables. La trampa de FI evita que esto suceda.

El uso de trampas es muy común en los circuitos de comunicaciones. La función primaria de una trampa es evitar que ciertas señales con frecuencias específicas pasen a los circuitos subsiguientes.

Algunas trampas dejan pasar una sola señal deseada o un grupo de señales con frecuencias muy próximas a la de ésta, bloqueando todas las demás.

Otras trampas dejan pasar todas las señales excepto aquellas que se encuentran dentro de algún rango determinado de frecuencias.

Por ejemplo, una trampa puede dejar pasar señales cuyas frecuencias sean inferiores a 2 MHz y superiores a 4 MHz. Por consiguiente, serán rechazadas todas las señales cuyas frecuencias estén entre 2 y 4 MHz.

Polarización del conversor

Las resistencias R4, R5 y R6 se utilizan para polarizar adecuadamente la etapa. Específicamente, R5 polariza directamente la unión BE de Q2 y R6 polariza inversamente la unión CB. La resistencia R4 estabiliza térmicamente el punto de trabajo.

El transistor extrae la corriente de polarización o de reposo de la base (IBQ) del positivo de la fuente de alimentación a través de R5. La corriente de reposo del colector (ICQ) se obtiene también del positivo de la fuente a través de R6 y la resistencia interna de las bobinas T1 y L4.

En la figura 411 se muestra el circuito equivalente de polarización del conversor de FI. Para obtener este modelo, sustituímos los condensadores por circuitos abiertos y las bobinas por cortocircuitos. También se indican los valores típicos de voltaje obtenidos entre cada electrodo de Q2 y tierra.

El fenómeno de las frecuencias imágenes y espúreas

Cada frecuencia del oscilador local origina una respuesta de FI para dos valores diferentes de señal de entrada: una mayor y otra menor que la frecuencia del oscilador local.

Si por ejemplo, se ajusta el oscilador local a 100.7 MHz para sintonizar una señal de 90 MHz, el receptor podrá también responder a una señal de 111.4 MHz. Esta última provocará también una frecuencia intermedia de 10.7 KHz.

A la señal indeseada de 111.4 MHz se le denomina *imagen* y si no se elimina, puede producir interferencia.

El amplificador de RF, la etapa previa al conversor, se sintoniza normalmente a la frecuencia de la señal deseada. Por lo tanto, su selectividad reduce o elimina la respuesta a la señal imagen. Esta es la razón más importante de utilizar un amplificador de RF en un receptor.

Además de señales imagen, pueden escucharse otras señales a las cuales el receptor no se encuentra sintonizado. Estas últimas se denominan *espúreas*.

Las armónicas del oscilador local de alta frecuencia pueden mezclarse con señales que se encuentren bastante alejadas de la frecuencia deseada y producir una señal de FI.

Estas respuestas espúreas se pueden reducir utilizando amplificadores de RF muy selectivos y empleando bobinas de FI blindadas. Con el blindaje, se evita la captación de señales parásitas provenientes de otras fuentes distintas de la antena.

Cuando el detector recibe una señal muy potente las armónicas generadas en este circuito pueden inducirse en el amplificador de RF o en el mezclador, procesándose conjuntamente con la señal deseada.

Estas señales indeseables se manifiestan en el parlante, generalmente en forma de silbidos.

Circuito de polarización del transistor conversor Q2

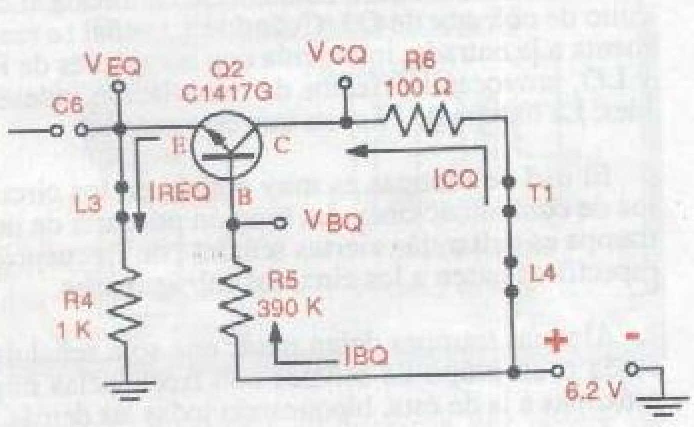

I_{BQ}	11.7 µA
I_{CQ}	2 mA
I_{REQ}	0.98 mA
V_{BQ}	1.63 V
V_{CQ}	6 V
V_{EQ}	0.98 V
V_{BEQ}	0.65 V
β h_{fe}	= 170

Todos los voltajes se miden con respecto a tierra, excepto V_{BEQ}
Nota: IREQ = corriente de reposo a través de la resistencia de emisor (R4)

Fig. 411

Lección 32

Los amplificadores de FI

Introducción

Los amplificadores de FI son circuitos de banda ancha sintonizados a 10.7 MHz, la frecuencia intermedia del proceso heterodino FM. Su función consiste en amplificar la señal de salida del conversor y suministrarla al demodulador o detector de relación del receptor (figura 412).

Etapas amplificadoras de FI

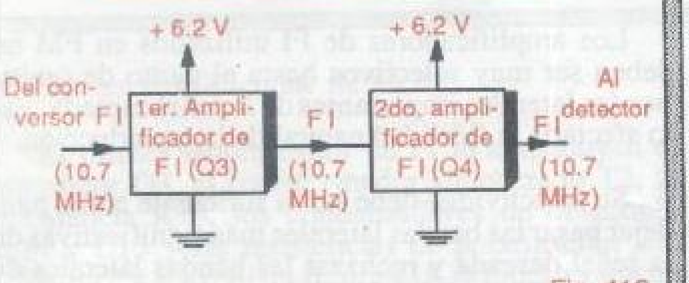

Fig. 412

El detector convierte las variaciones de frecuencia de la señal de FI en una señal de audio correspondiente. La señal de FI suministrada por el conversor tiene una amplitud fija y su frecuencia varía con respecto a la portadora de 10.7 MHz, dependiendo de la modulación original.

La amplificación de la señal de FI se realiza en dos etapas. Cada etapa tiene un ancho de banda de aproximadamente 150 KHz, espacio suficiente para dejar pasar las bandas laterales más significativas de la señal de FM.

Ambas etapas utilizan transistores bipolares como dispositivos activos de amplificación y están acopladas entre sí y con el conversor y el detector, a través de transformdores sintonizados a 10.7 MHz. A esta frecuencia, la ganancia de voltaje de cada etapa es máxima.

Además de proveer una alta selectividad, estos transformadores garantizan también el correcto acoplamiento entre etapas, condición necesaria para garantizar la máxima transferencia de señal entre un circuito y el siguiente.

Características de los amplificadores de FI para FM

Las etapas de frecuencia intermedia son las principales responsables de la selectividad y la sensibilidad características de los receptores superheterodinos de FM. En los radios más sencillos suele utilizarse una sola etapa de FI. La mayoría de diseños emplean dos o más amplificadores de FI.

La alta sensibilidad de los receptores de FM es una consecuencia directa del alto grado de amplificación que puede alcanzarse, con relativa facilidad, en las etapas de FI. Es necesaria una alta ganancia para garantizar una óptima detección.

El radio FM CEKIT utiliza dos transistores como amplificadores de FI para conseguir la ganancia requerida. La tendencia en los receptores modernos es utilizar circuitos integrados para esta función.

Los circuitos integrados amplificadores de FI, tales como el CA3012, el ECG736, el TBA120 y otros, son fáciles de conseguir y de utilizar. Además, resultan relativamente económicos.

En la figura 413 se muestra un ejemplo típico de aplicación del circuito integrado CA3012 como amplificador de FI de 10.7 MHz en un receptor de radiodifusión FM. El CA3012 acepta tensiones de alimentación hasta de +10 V. Este voltaje se aplica entre los pines 10 (positivo) y 8 (tierra).

Receptor de FM con amplificador de F I CA3012

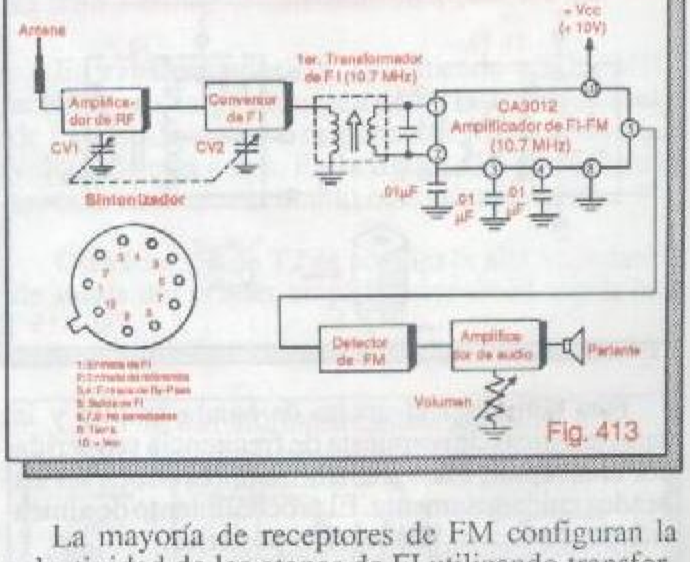

Fig. 413

La mayoría de receptores de FM configuran la selectividad de las etapas de FI utilizando transformadores sintonizados de acoplamiento. En el caso del receptor FM CEKIT se utilizan tres transformadores sintonizados a 10.7 MHz para este propósito (figura 414).

Acoplamiento magnético de las etapas de F I

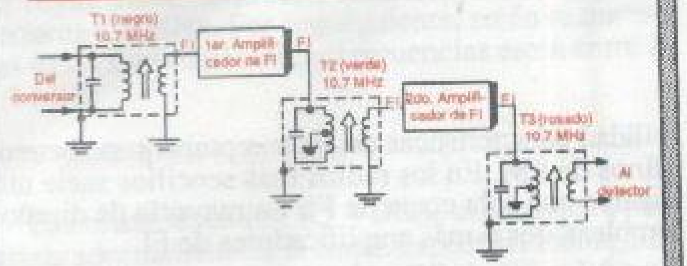

Fig. 414

El primer transformador (T1) acopla la salida del conversor con la entrada de la primera etapa de FI. El segundo (T2) acopla la salida del primer amplificador con la entrada del segundo. El tercero (T3) acopla la salida de la segunda etapa de FI con la entrada del detector.

Estos transformadores son blindados y se devanan sobre núcleos ajustables de ferrita. Vienen provistos de condensadores fijos y se sintonizan variando la posición del núcleo. Los núcleos se distinguen por colores: naranja para T1, verde para T2 y rosado para T3.

En la figura 415 se muestra la estructura típica de los transformadores de FI utilizados en el receptor de FM CEKIT. El blindaje de aluminio se conecta a tierra. Su propósito es confinar la energía de FI dentro de la estructura, evitando que se irradie al exterior. Minimiza también la captación de ruido.

Estructura de los transformadores de F I

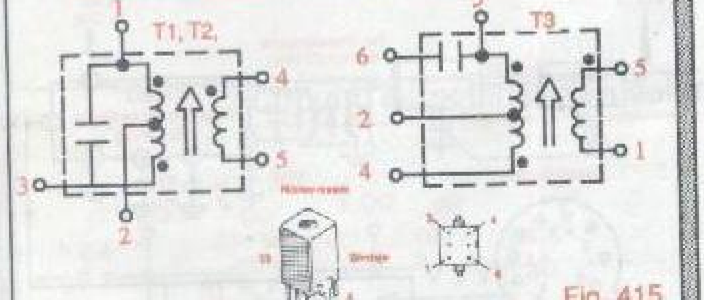

Fig. 415

Para conseguir el ancho de banda amplio y las características de respuesta de frecuencia requeridas por el receptor, estos transformadores deben ser alineados cuidadosamente. El procedimiento de alineación se explica en una lección posterior.

En la figura 416 se muestra una curva típica de respuesta de frecuencia de un amplificador de FI para FM. La frecuencia central de la banda pasante es 10.7 MHz. El circuito debe proporcionar una respuesta plana en un rango de 150 KHz alrededor de la frecuencia central o FI.

Respuesta de frecuencia de los amplificadores de F I

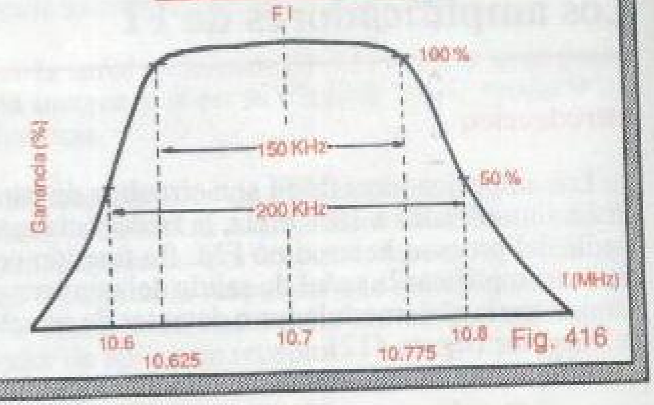

Fig. 416

Los amplificadores de FI utilizados en FM no deben ser muy selectivos hasta el punto de omitir bandas laterales importantes de la señal captada. Esto afectaría la fidelidad natural de la reproducción.

Su selectividad debe ser la suficiente como para dejar pasar las bandas laterales más significativas de la señal deseada y rechazar las bandas laterales de canales adyacentes.

La elección de 10.7 MHz como frecuencia intermedia del proceso heterodino FM es un valor de compromiso entre dos hechos importantes: el rechazo a la frecuencia imagen y la selectividad.

El valor de 10.7 MHz elegido resulta ser lo suficientemente alto como para proporcionar un óptimo rechazo a la frecuencia imagen y lo suficientemente bajo como para garantizar una excelente selectividad y una buena ganancia.

Operación de los amplificadores de FI

En la figura 417 se muestra el diagrama esquemático de las etapas amplificadoras de FI. El elemento activo de la primera etapa es el transistor Q3, conectado en la configuración base común. El elemento activo de la segunda etapa es el transistor Q4, conectado en la configuración emisor común.

La señal de FI suministrada por el conversor se transfiere a la entrada del primer amplificador de FI a través del condensador C14. El acoplamiento entre las dos etapas de FI se realiza a través del transformador T2, sintonizado a 10.7 MHz.

La señal de salida del segundo amplificador de FI se acopla magnéticamente al demodulador mediante el transformador T3, sintonizado también a 10.7 MHZ. Las dos etapas de FI utilizan una tensión de alimentación de 6.2 V.

Circuito de los amplificadores de FI

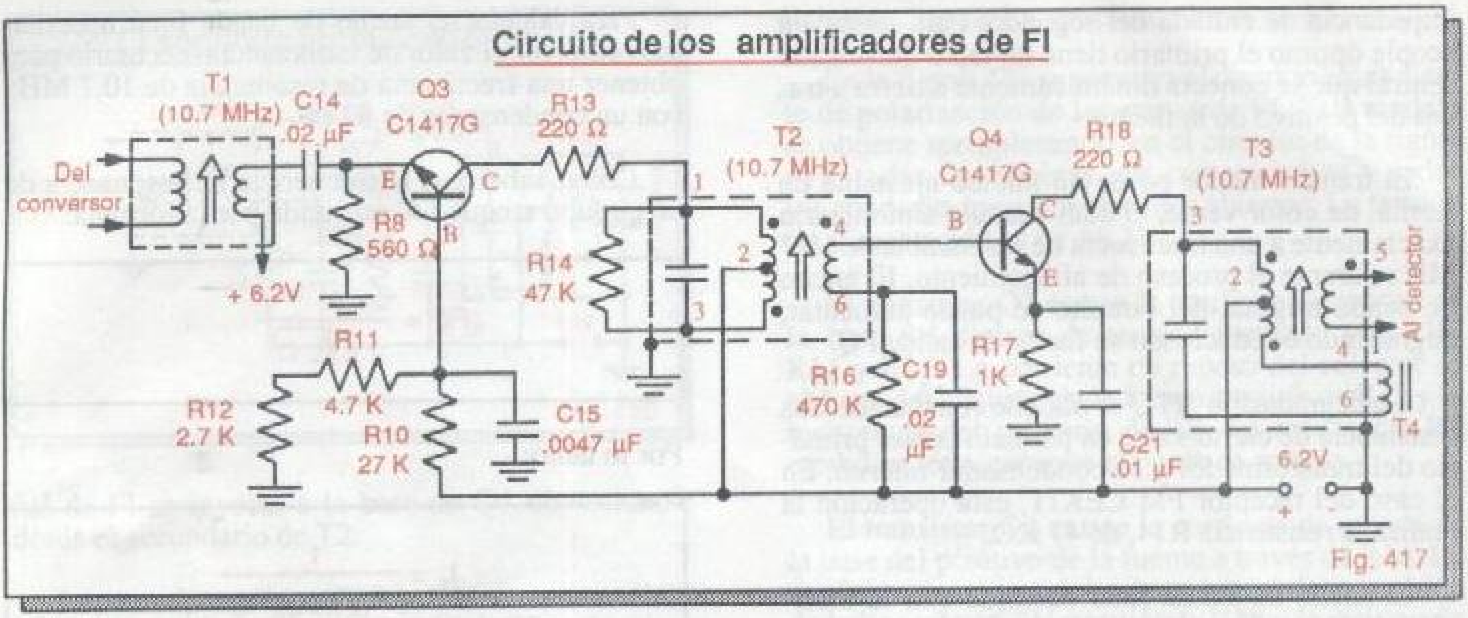
T1
(10.7 MHz)
C14
.02 µF
Q3
C1417G
R13
220 Ω
T2
(10.7 MHz)
Q4
C1417G
R18
220 Ω
T3
(10.7 MHz)
Del
conversor
E
C
B
R8
560 Ω
+ 6.2V
R14
47 K
1
2
3
4
6
R11
4.7 K
R12
2.7 K
R10
27 K
C15
.0047 µF
R16
470 K
C19
.02
µF
R17
1K
C21
.01 µF
6.2V
+
5
4
6
Al detector
T4

Fig. 417

El primer amplificador de FI

El primer amplificador de FI lo constituyen el transistor Q3, el transformador sintonizado T2, la resistencia R13, los condensadores de desacople C14 y C15 y las resistencias de polarización R8, R11, R12 y R13. En la figura 418 se muestra su circuito esquemático.

El primer amplificador de FI

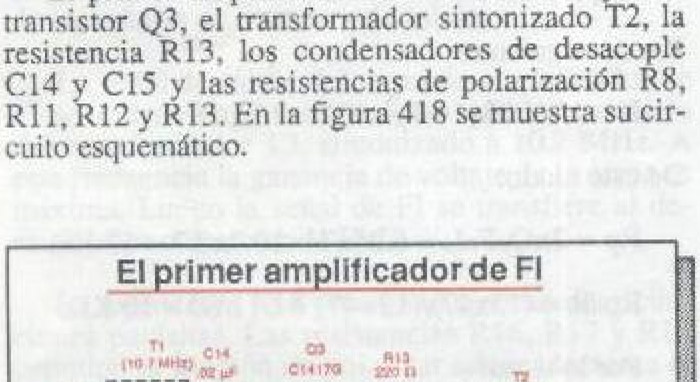
T1
(10.7 MHz)
C14
.02 µF
Q3
C1417G
R13
220 Ω
T2
(10.7 MHz)
Del
conversor
E
C
B
R8
560 Ω
+ 6.2V
R14
47 K
1
2
3
4
6
Al segundo
amplificador
R11
4.7 K
R12
2.7 K
R10
27 K
C15
.0047 µF
6.2V

Fig. 418

La señal de entrada del primer amplificador de FI proviene del conversor. El acoplamiento entre ambas etapas se efectúa a través del transformador T1, sintonizado a 10.7 MHz (FI). Este dispositivo acopla la alta impedancia de salida del conversor con la baja impedancia de entrada del amplificador.

La señal de FI se inyecta al emisor de Q3 a través de C14. Este condensador presenta una muy baja reactancia a 10.7 MHz y aisla el emisor del positivo de la fuente, conservando la polarización de la etapa.

El condensador C15 conecta dinámicamente a tierra la base de Q3. De este modo, la señal de entrada queda aplicada efectivamente entre emisor y base.

La señal de salida del primer amplificador de FI se aplica al primario del transformador T2, sintonizado a 10.7 MHz, y se transfiere a la segunda etapa de FI. A esta frecuencia la ganancia de voltaje de la etapa es máxima.

La resistencia R13 protege al circuito de oscilaciones parásitas. La resistencia R14 se utiliza para disminuir artificialmente el Q del transformador sintonizado T2, aumentando su ancho de banda. Las resistencias R8, R10, R11, R12 y R3 cumplen la función de polarizar adecuadamente el amplificador.

El transformador sintonizado T2

El transformador T2, sintonizado a 10.7 MHz, actúa como carga de salida del primer amplificador de FI y es el principal responsable de la ganancia de voltaje de esta etapa. En la figura 419 se muestra su aspecto típico y su circuito equivalente.

Otra función de T2 es acoplar la alta impedancia de salida del primer amplificador de FI con la baja

El transformador sintonizado T2

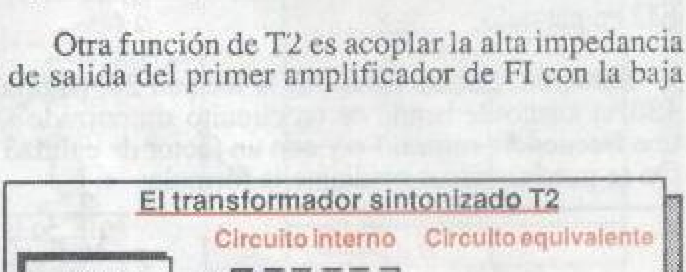
Circuito interno
Circuito equivalente
Núcleo verde
1
4
6
3
2
1
3
Rp
C
L
Aspecto externo

Fig. 419

impedancia de entrada del segundo. Para lograr un acople óptimo el primario tiene un tap o derivación central que se conecta dinámicamente a tierra a través del positivo de la fuente.

El transformador posee un núcleo ajustable de ferrita, de color verde, el cual permite sintonizarlo exactamente a una frecuencia de resonancia de 10.7 MHz durante el proceso de alineamiento. El ancho de banda original del circuito se puede aumentar, degradando o reduciendo su factor de calidad Q.

La degradación del Q se obtiene conectando una resistencia de cierto valor en paralelo con el primario del transformador y su condensador interno. En el caso del receptor FM CEKIT, esta operación la realiza la resistencia R14, de 47 KΩ.

El ancho de banda final obtenido con la inserción de esta resistencia puede evaluarse, en forma aproximada, mediante las siguientes fórmulas:

$$B \approx \frac{2\pi Fo^2 L}{R \parallel Rp}$$

$$Rp = 2\pi Qo Fo L$$

En estas expresiones, $2\pi=6.28$ es una constante, B es el ancho de banda final (Hz), Fo la frecuencia de resonancia (Hz), L la inductancia del primario (H), R la resistencia de degradación (Ω), Rp la resistencia equivalente de pérdidas de T2 (Ω) y Qo el factor de calidad original.

El siguiente ejemplo numérico ilustra el uso de estas fórmulas.

Ejemplo. Considerando que el condensador interno del transformador es de 82 pF y que éste tiene un Qo de 71 (valores típicos), calcule el ancho de banda original del transformador T1 y el ancho de banda obtenido al conectar una resistencia de 47 KΩ en paralelo.

Solución. Según vimos en la lección 24, página 180, el ancho de banda de un circuito sintonizado a una frecuencia central Fo y con un factor de calidad Qo se puede evaluar mediante la fórmula:

$$B = \frac{Fo}{Qo}$$

En nuestro caso, Fo=10.7 MHz y Qo=71. Así:

$B_o = 10.7/71 = 0.15$ MHz = **150 KHz**.

Este es el valor del ancho de banda original del transformador sintonizado T1.

Para calcular el ancho de banda final necesitamos conocer el valor de inductancia necesario para obtener una frecuencia de resonancia de 10.7 MHz con un condensador de 82 pF.

Como sabemos, la frecuencia de resonancia de un circuito tanque LC está dada por la fórmula:

$$Fo = \frac{1}{2\pi\sqrt{LC}}$$

Por lo tanto:

$$L = \frac{1}{(2\pi Fo)^2 C}$$

En nuestro caso, $2\pi=6.28$, Fo=10.7 MHz y C=82 pF.

Por consiguiente:

$L = 1/((6.28 \times 10.7)^2 \times 82) = 2.7$ µH.

De este modo:

$Rp = 2\pi Q_o Fo L = 6.28 \times 71 \times 10.7 \times 2.7 = 13$ KΩ

y

$Rp \parallel R = (13 \times 47)/(13+47) = 611/60 = 10$ KΩ.

Por lo tanto:

$B = (2\pi Fo^2 L)/(R \parallel Rp) = (6.28 \times 10.7^2 \times 2.7)/10$

B = 1940/10 = 194 KHz

Este es el valor final de ancho de banda obtenido al insertar una resistencia de 47 KΩ en paralelo con el primario del transformador sintonizado. La banda de paso se ha extendido en 44 KHz con respecto a su valor original (150 KHz). Esto representa un incremento del 30%, aproximadamente.

El segundo amplificador de FI

El segundo amplificador de FI lo constituyen el transistor Q4, el transformador sintonizado T3, los condensadores de desacople C19 y C21 y las resistencias de polarización R16, R17 y R18. En la figura 420 se muestra su circuito esquemático.

La señal de entrada del segundo amplificador de FI proviene de la primera etapa. El acoplamiento entre ambos circuitos se efectúa a través del transformador T2, sintonizado a la FI de 10.7 MHz. La se-

El segundo amplificador de FI

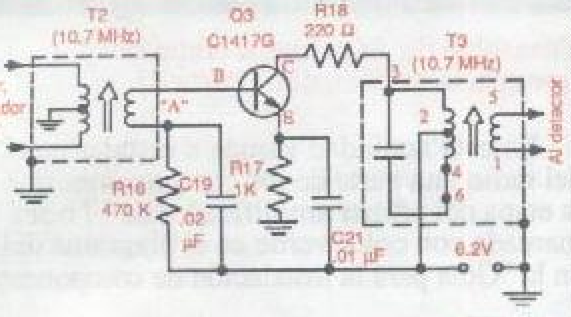

Fig. 420

ñal de FI se inyecta a la base de Q4 directamente desde el secundario de T2.

Los condensadores C19 y C21 conectan dinámicamente a tierra el extremo "A" de T2 y el emisor de Q4, respectivamente. De este modo, la señal de entrada queda aplicada efectivamente entre base y emisor. El emisor es el punto común de referencia de las señales de entrada y de salida.

La señal de salida del segundo amplificador de FI se obtiene en el colector y se aplica al primario del transformador T3, sintonizado a 10.7 MHz. A esta frecuencia la ganancia de voltaje de la etapa es máxima. Luego la señal de FI se transfiere al demodulador o detector de FM.

La resistencia R18 protege al circuito de oscilaciones parásitas. Las resistencias R16, R17 y R18 cumplen la función de polarizar adecuadamente el amplificador.

Polarización de los amplificadores de FI

En la figura 421 se muestra el circuito equivalente de polarización de las etapas de FI. Este modelo se obtiene reemplazando, en el circuito de la figura 417, todas las bobinas por cortocircuitos y todos los condensadores por circuitos abiertos. La tensión de alimentación es de +6.2V.

El transistor Q3 extrae la corriente de reposo de la base del divisor de tensión formado por R10, R11 y R12. La corriente de reposo del colector se obtiene del positivo de la fuente a través de R13 y la resistencia del primario de T2. La resistencia R8 estabiliza térmicamente el punto de trabajo.

El transistor Q4 extrae la corriente de reposo de la base del positivo de la fuente a través de R16. La corriente de reposo del colector se obtiene también de la línea de +6.2V a través de R18 y la resistencia del primario de T3. La resistencia R17 estabiliza térmicamente el punto de trabajo.

Receptores de FM de doble conversión

La mayor parte de los receptores usados en comunicaciones de FM se diseñan para la banda de VHF. En este rango de frecuencias no siempre es posible obtener recepción óptima con un circuito superheterodino convencional debido a las características de la FI.

Como sabemos, una FI alta es conveniente para asegurar una buena selectividad. Sin embargo no garantiza un buen rechazo de las frecuencias imagen.

Redes de polarización de los amplificadores de F I

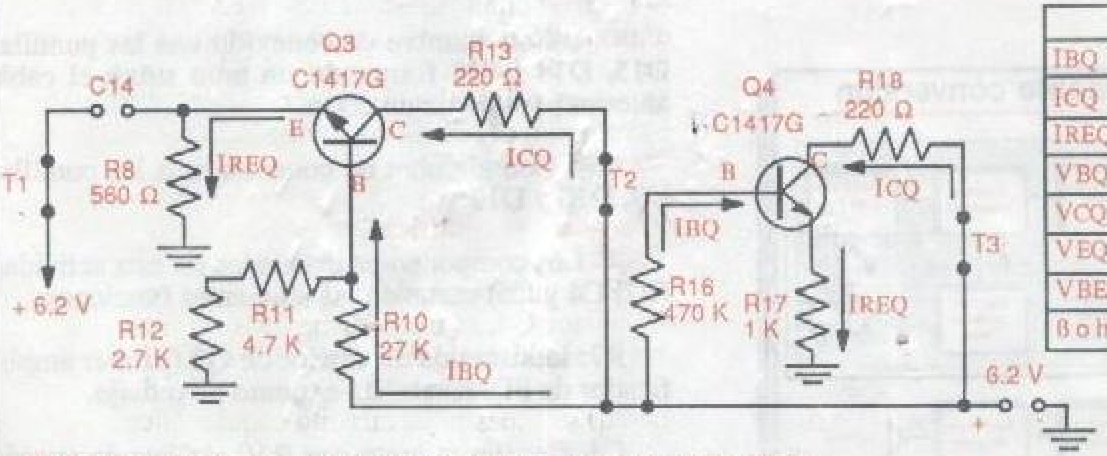

	Q3	Q4
IBQ	0.18 mA	10.8 µA
ICQ	1.36 mA	1.36 mA
IREQ	0.96 mA	0.42 mA
VBQ	1.21 V	1.11 V
VCQ	5.9 V	5.9 V
VEQ	0.54 V	0.42 V
VBEQ	0.67 V	0.69 V
β o hfe	≈ 8	126

Todos los voltajes se miden con respecto a tierra, excepto V_{BEQ}

Nota: I_{REQ} = corriente de reposo a través de las resistencias de emisor (R8, R16)

Fig. 421

Los amplificadores de FI no pueden distinguir entre las señales de FI producidas por la señal deseada y la señal imagen. Una FI baja, aunque deja pasar algunas frecuencias imagen, garantiza que éstas sean menos numerosas y de menor amplitud.

Resulta entonces obvio que una sola FI no puede asegurar al mismo tiempo una buena selectividad y una buena supresión de frecuencias imagen. Para minimizar este inconveniente muchos receptores de FM aplican el proceso de doble conversión.

En la figura 422 se muestra el diagrama de bloques de un receptor de FM de doble conversión.

Por medio de este sistema la señal se convierte primeramente en una FI alta y se amplifica. Así se obtiene la selectividad de frecuencia requerida. A continuación, la FI inicial se convierte en una FI más baja. Para esta operación se necesita otro mezclador y otro oscilador local. Este proceso asegura la supresión requerida de frecuencias imagen.

Los circuitos restantes son los mismos que en un receptor de conversión simple convencional.

El segundo mezclador deberá estar localizado despúes de un filtro altamente selectivo si se desean minimizar las espúreas y las frecuencias imagen.

Los requisitos de diseño del segundo mezclador y del segundo oscilador local no son tan exigentes como para los primeros. Para estas funciones pueden utilizarse transistores bipolares de bajo costo.

Con un esquema de doble conversión es posible eliminar la etapa amplificadora de RF sin degradar seriamente la sensibilidad del receptor, si existe suficiente selectividad antes del primer mezclador. En estos casos resulta muy práctico el uso de FETs y MOSFETS como dispositivos mezcladores.

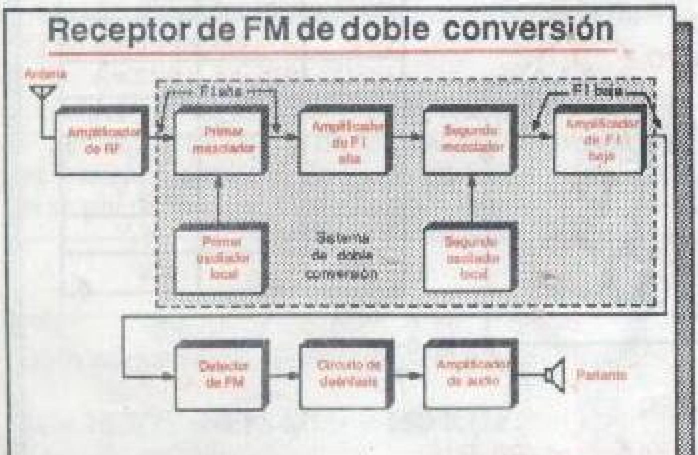
Receptor de FM de doble conversión

Fig. 422

Actividad práctica Nº 18

Ensamble de la primera etapa de frecuencia intermedia. (Parte 1)

En esta actividad vamos a instalar en el tablero del radio una parte de los componentes que forman la etapa del primer amplificador de FI o sea la etapa marcada con color verde en el diagrama del radio y en la "Guía para la instalación de componentes".

Componentes y materiales necesarios

1 Resistencia de 470 Ω, 1/4 W, 5% (amarillo, violeta, café, dorado). R7.
1 Condensador de cerámica de 0.02 µF/50V. C5.
1 Condensador electrolítico de 10 µF/10V ó 16V. C4.
Alambre de conexión.

Paso 1: Instale la resistencia R7 de 470 Ω entre las puntillas D10 y D11. Utilice el terminal interior de la resistencia para unir las puntillas D11 y 8.

Paso 2: Instale el condensador C5 de 0.02 µF entre las puntillas D12 y 9.

Paso 3: Instale el condensador C4 de 10 µF/10V entre las puntillas D6 y D11. Asegúrese de que el terminal negativo quede en la puntilla D11. Recorte los terminales sobrantes.

Paso 4: Con alambre de conexión, una las puntillas 10 y D13 dejando unos tres centímetros libres en el extremo de D13, que se debe conectar después a la carcaza metálica de T2.

Paso 5: Con alambre de conexión una las puntillas D16 y 42, formando ángulo recto como lo muestra la guía para la instalación de componentes.

Paso 6: Con alambre de conexión una las puntillas D15, D14 y D8 formando un arco sobre el cable anterior para evitar un corto.

Paso 7: Con alambre de conexión una las puntillas D9, D10 y D12.

Nota: Los componentes instalados en esta actividad (R7, C4 y C5) cumplen las siguientes funciones:

R7: Resistencia de emisor de Q2 (primer amplificador de FI). Estabiliza el punto de trabajo.

C4: Configura junto con R10 el filtro de retardo del voltaje de CAG de Q2.

C5: Conecta dinámicamente a tierra el emisor de Q2, mejorando la ganancia de CA de la 1º etapa.

Lección 33

El detector de relación o demodulador FM

Introducción

En esta lección se estudia el demodulador o detector del radio FM CEKIT. La función de este circuito es recuperar las variaciones de amplitud de baja frecuencia de la información de audio implícitas en las variaciones de frecuencia de la portadora de FI. En la figura 423 se muestra su bloque funcional.

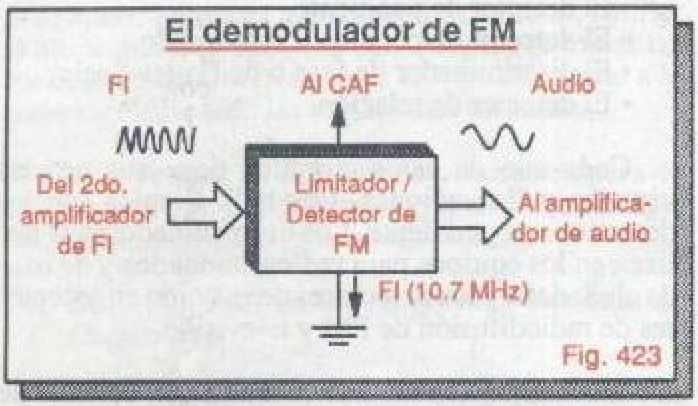

El demodulador de FM

Fig. 423

El detector recibe la señal amplificada de FI suministrada por la segunda etapa, ignora la portadora interna de 10.7 MHz y convierte las variaciones de frecuencia de esta última en variaciones de amplitud de baja frecuencia, es decir, en una señal de audio.

Además de ésta, que es su función básica, el detector actúa como *limitador* dinámico de amplitud de ruido y espúreas y suministra la señal de entrada del CAF o control automático de frecuencia.

En términos generales, el detector convierte la señal de FI modulada en frecuencia en una señal de audio que está modulada tanto en amplitud como en frecuencia. La modulación de la señal de audio se refiere al hecho de que su amplitud y frecuencia instantáneas están cambiando con el tiempo.

La señal de audio extraída por el detector debe ser, en lo posible, una réplica exacta de la información original (modulación) enviada con la portadora de RF desde el transmisor y trasladada a la portadora interna de FI en el receptor.

En este sentido, el detector de un receptor de FM opera en forma exactamente opuesta al modulador del transmisor. Mientras que el modulador transfiere la señal de audio a la portadora de RF para modular su frecuencia, el detector extrae la señal de audio de la portadora de FI.

En otras palabras, el modulador traslada una banda base de baja frecuencia a una portadora de alta frecuencia. El detector retorna la banda base de la portadora a sus condiciones originales. En la figura 424 se ilustra este concepto.

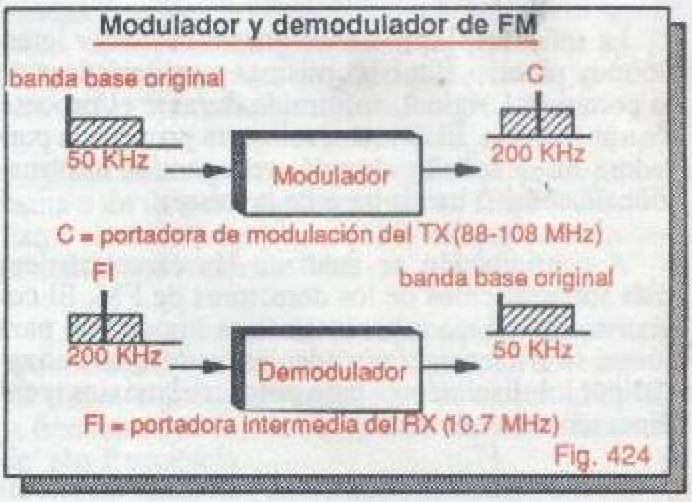
banda base original
C
50 KHz
200 KHz
Modulador
C = portadora de modulación del TX (88-108 MHz)
FI
banda base original
200 KHz
50 KHz
Demodulador
FI = portadora intermedia del RX (10.7 MHz)

Modulador y demodulador de FM

Fig. 424

En general, la demodulación o detección es el proceso de recuperar una señal de audio de una portadora modulada. Cualquier circuito que extraiga una información de una portadora modulada es, por lo tanto, un detector.

Todos lo detectores realizan la misma función básica descrita anteriormente, es decir, recuperar la información original enviada sobre una portadora. La diferencia entre los diferentes tipos de receptores (AM, FM, SSB, etc.) radica, precisamente, en la forma como ellos demodulan la señal recibida.

Las técnicas de detección utilizadas en FM son sustancialmente distintas a las empleadas en AM debido a la forma tan diferente como viene codificada la información en ambos sistemas.

En AM la información se envía codificada en las variaciones de amplitud de la portadora y se mantiene constante su frecuencia. En FM se envía codificada en las variaciones de frecuencia, manteniéndose constante la amplitud de la portadora.

Los circuitos utilizados para detectar señales de AM o FM tienen sólo una entrada: la de la señal de FI. Los utilizados para demodular señales de SSB (banda lateral única) y DSB (banda lateral doble), requieren de dos entradas: una para la señal incidente y otra para una señal local (figura 425).

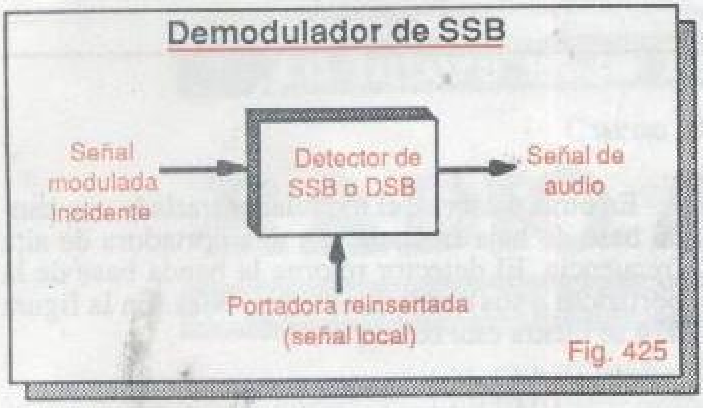

Demodulador de SSB

Fig. 425

La señal local, producida por un oscilador interno muy preciso, tiene las mismas características de la portadora original, suprimida durante el proceso de transmisión. El detector reinserta primero la portadora en la señal y después recupera la información de audio o banda base de la misma.

A continuación se analizan las características más sobresalientes de los detectores de FM. El conocimiento de estos parámetros es importante para juzgar su eficiencia y entender las estrategias utilizadas por los diseñadores para obtener el máximo rendimiento de estos circuitos.

Posteriormente se describen y analizan los circuitos de detección de FM más populares, haciendo especial énfasis en los detectores de fase y de relación que son los más utilizados. Finalmente, se analiza el detector de relación *balanceado* empleado en el radio FM CEKIT.

Características de los detectores de FM

Las principales características de un detector de FM son su linealidad, su sensibilidad, su selectividad y su capacidad de manejo.

La linealidad se refiere a su habilidad para no distorsionar la señal detectada. La señal de audio obtenida debe ser una réplica exacta de las variaciones de frecuencia de la señal de FI.

La sensibilidad se refiere a su capacidad para amplificar la señal al mismo tiempo que la detecta. Los detectores con esta característica utilizan dispositivos activos como tubos, transistores y CIs.

La selectividad se refiere a su habilidad para responder a una banda dada de frecuencias y rechazar frecuencias que estén por fuera de esa banda. El detector del radio FM CEKIT responde a frecuencias dentro de un ancho de banda de 200 KHz.

La capacidad de manejo se refiere a su habilidad para demodular señales de entrada de gran amplitud sin producir distorsión excesiva ni sobrecarga.

Así mismo, un detector de FM deberá ser, en lo posible, insensible a los cambios de amplitud de la señal de entrada y su ajuste y operación no deberán ser factores críticos.

Tipos de detectores de FM

Los circuitos detectores de FM tienen una entrada y una salida. A la entrada se aplica un voltaje (FI) de amplitud constante y frecuencia variable. A la salida se obtiene un voltaje (audio) cuya amplitud instántanea depende de las variaciones de frecuencia de la señal de entrada.

Existen básicamente 4 tipos de detectores de FM:

- El detector de pendiente.
- El detector de pendiente balanceado.
- El discriminador de fase o de Foster-Seeley.
- El detector de relación.

Cada uno de estos circuitos tiene sus propias virtudes y limitaciones. Los más simples son los detectores de pendiente. Los discriminadores se utilizan en los equipos para radioaficionados y de banda ciudadana y los detectores de relación en receptores de radiodifusión de FM y televisión.

El receptor de FM CEKIT utiliza un detector de relación balanceado, que es una de las configuraciones detectoras de FM más eficientes. Por esta razón prestaremos particular atención a los principios de funcionamiento de este último.

El detector de pendiente

En la figura 426 se ilustra el principio de operación de un detector de pendiente. La señal de FI alimenta un circuito sintonizado cuya frecuencia de resonancia (**fr**) está desplazada con respecto a la frecuencia de portadora de la señal de entrada (**fi**) en una cantidad igual a su desviación máxima (**Δf**).

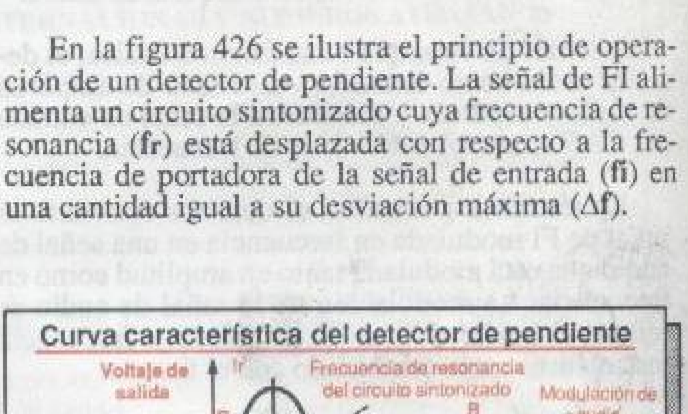

Curva característica del detector de pendiente

Fig. 426

La amplitud instantánea de la señal de salida será siempre directamente proporcional a la desviación instantánea de frecuencia de la portadora de FI. En el punto A, por ejemplo, la desviación de frecuencia es cero. Por lo tanto, la amplitud de la señal de salida es también cero.

En los puntos B y C la desviación de frecuencia es máxima con respecto a la portadora de FI. Bajo estas circunstancias, la amplitud de la señal de salida es también máxima, siendo positiva en el primer caso, porque la desviación es positiva, y negativa en el otro, porque la desviación es negativa.

Algo parecido puede decirse de cualquier punto intermedio. El voltaje de salida del circuito resonante se aplica entonces a un detector de diodo, similar al utilizado en un receptor de AM, para eliminar las variaciones de alta frecuencia (FI) y suministrar a la etapa siguiente sólo las variaciones de audio.

Como vemos, un detector de pendiente opera de una manera extremadamente simple. Sin embargo, es ineficiente por las siguientes razones:

- Sólo es lineal en un rango muy limitado de frecuencias.
- Responde también a cambios de amplitud. Por lo tanto, detecta también cualquier ruido asociado a la señal de entrada.
- Es relativamente complicado de ajustar y su operación es crítica.

El detector de pendiente balanceado

En la figura 427 se muestra la configuración de un detector de pendiente balanceado, llamado también *detector de Travis*, en honor de su inventor. El circuito utiliza dos detectores de pendiente, conectados en oposición de fase a los extremos de un transformador con tap o derivación central.

Detector de pendiente balanceado

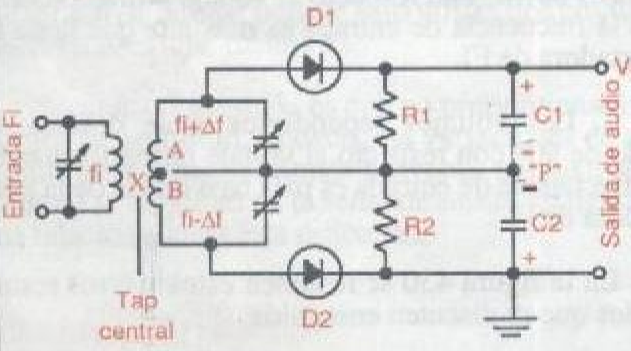

Δf: Desviación máxima (100 KHz)
fi : Frecuencia intermedia (10.7 MHz)

Fig. 427

El primario del transformador se sintoniza a la frecuencia de la portadora de FI (**fi**). El secundario superior se sintoniza a una frecuencia igual a **fi+Δf** y el inferior a una frecuencia igual a **fi-Δf**. Como sabemos, **Δf** es la desviación máxima de frecuencia de la portadora de FI.

En receptores de radiodifusión FM, Δf se toma igual a 100 KHz para cubrir el máximo número de bandas laterales posible. Recuerde que en el transmisor la desviación pico máxima admisible es 75 KHz, lo cual provee un ancho de banda de 200 KHz y una modulación máxima de 50 KHz.

Cada circuito sintonizado del secundario se conecta a un detector de diodo y a un filtro RC de FI. Estos circuitos cumplen aquí la misma función que en un demodulador de AM.

Los diodos rectifican la señal de entrada presente entre el tap central y cada extremo del devanado y los filtros siguen las variaciones de amplitud de baja frecuencia de la misma, ignorando las variaciones de alta frecuencia.

La señal de salida total (**Vo**) es la suma o superposición de las salidas individuales de cada filtro. Veamos entonces cómo opera el detector balancedo.

Cuando la desviación de frecuencia de la señal de entrada es máxima, el voltaje desarrollado en uno de los devanados sintonizados del secundario será el máximo posible, dependiendo de la dirección de la desviación.

Si la desviación máxima es positiva (por encima de la portadora de FI), la frecuencia de la señal de entrada será **fi+Δf** y entrará en resonancia el circuito superior, desarrollándose un voltaje máximo en el devanado "A" y un voltaje mínimo en el devanado "B". El voltaje de salida será máximo positivo.

Si la desviación máxima es negativa (por debajo de la portadora de FI), la frecuencia de la señal de entrada será **fi-Δf** y entrará en resonancia el circuito inferior, desarrollándose un voltaje máximo en el devanado "B" y un voltaje mínimo en el devanado "A". El voltaje de salida será máximo y negativo.

Cuando la desviación instantánea de la portadora de FI está entre estos dos extremos, el voltaje de salida tendrá algún valor intermedio, dependiendo de la magnitud y polaridad de la desviación.

En particular, si la desviación es cero, la frecuencia de la señal de entrada tendrá un valor igual al de la portadora de FI (10.7 MHz). Bajo estas condiciones, ninguno de los circuitos sintonizados entrará en resonancia y el voltaje de salida será prácticamente cero.

Algo similar a esto último sucede cuando la frecuencia de la señal de entrada se sale del rango comprendido entre **f-Δf** y **f+Δf**. El circuito rechaza así señales no deseadas como imágenes y espúreas. En la figura 428 se resume su comportamiento.

Respuesta del detector de pendiente balanceado

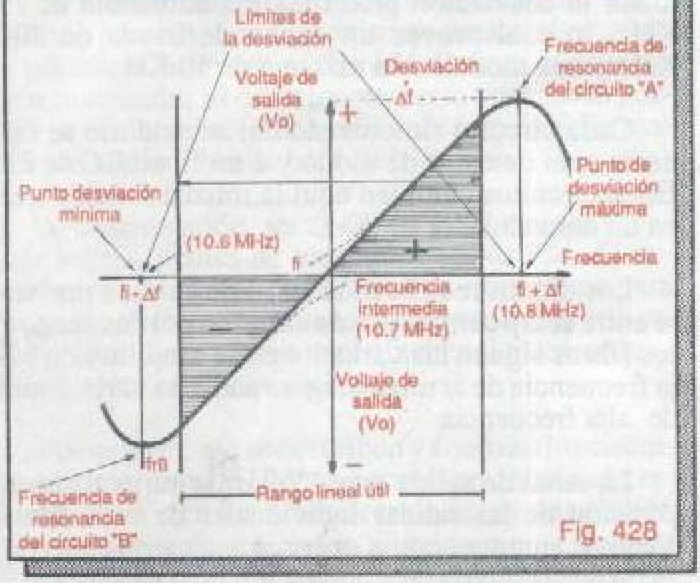

Fig. 428

A pesar de ser considerablemente más eficiente y lineal que el detector de pendiente, la calibración del detector balanceado sigue siendo aún muy crítica. Además, el detector responde también a los cambios de amplitud y su linealidad, aunque es buena, no es óptima.

El detector de fase o discriminador de Foster-Seeley

En la figura 429 se muestra el circuito de un detector de fase, más comúnmente conocido como *discriminador de Foster-Seeley*, en honor de sus inventores. El circuito es similar en su estructura al detector de pendiente balanceado.

Observe que se sigue conservando la misma configuración de diodos y filtros RC a la salida por ser un esquema muy satisfactorio. Sin embargo, los devanados primario y secundario están ambos sintonizados a la portadora de FI (**fi**). Estos circuitos son de alto Q y su inductancia mutua es muy baja.

Discriminador de Foster - Seeley

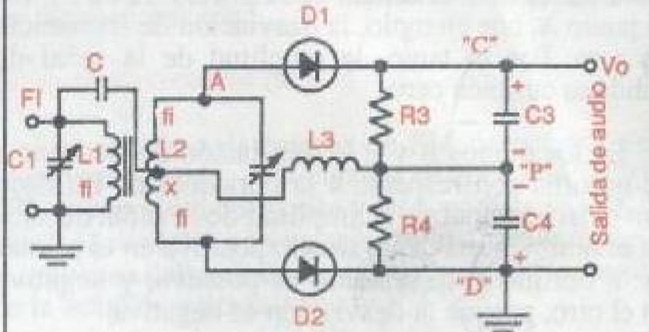

Fig. 429

Además, se ha insertado una bobina (L3) entre el punto común de los filtros (P) y el tap central del secundario y un condensador (C) entre este último y la entrada de señal del primario. Esto asegura que los voltajes aplicados a los diodos varíen linealmente con las desviaciones de frecuencia de la señal FI.

La bobina L3 es un choque de radiofrecuencia (RFC). Su propósito es suministrar una gran reactancia a las frecuencias involucradas. Esta reactancia es muy superior a la de C, C3 y C4.

El voltaje a través de L3 es prácticamente igual al voltaje de entrada (**Vin**). El voltaje aplicado a cada diodo es la suma del voltaje primario y el correspondiente voltaje secundario, medido este último con respecto al tap central (X).

El voltaje de salida (**Vo**), obtenido entre los puntos C y D, es la diferencia de los voltajes sobre C3 y C4. Otras características importantes del circuito, demostrables matemáticamente, son:

1. Los voltajes secundarios están desfasados exactamente 90º con respecto al voltaje primario cuando la frecuencia de entrada es igual a la de la portadora de FI (**fi**).

2. Los voltajes secundarios están desfasados menos de 90º con respecto al voltaje primario cuando la frecuencia de entrada es más alta que la de la portadora de FI.

3. Los voltajes secundarios están desfasados más de 90º con respecto al voltaje primario cuando la frecuencia de entrada es más baja que la de la portadora de FI.

En la figura 430 se resumen estos y otros resultados que se discuten enseguida.

En consecuencia, los voltajes de salida de los diodos D1 y D2, medidos con respecto al punto P, serán iguales sólo cuando la frecuencia de la señal de entrada sea igual a la FI de 10.7 MHz. En este

Condiciones de fase del discriminador Foster - Seeley

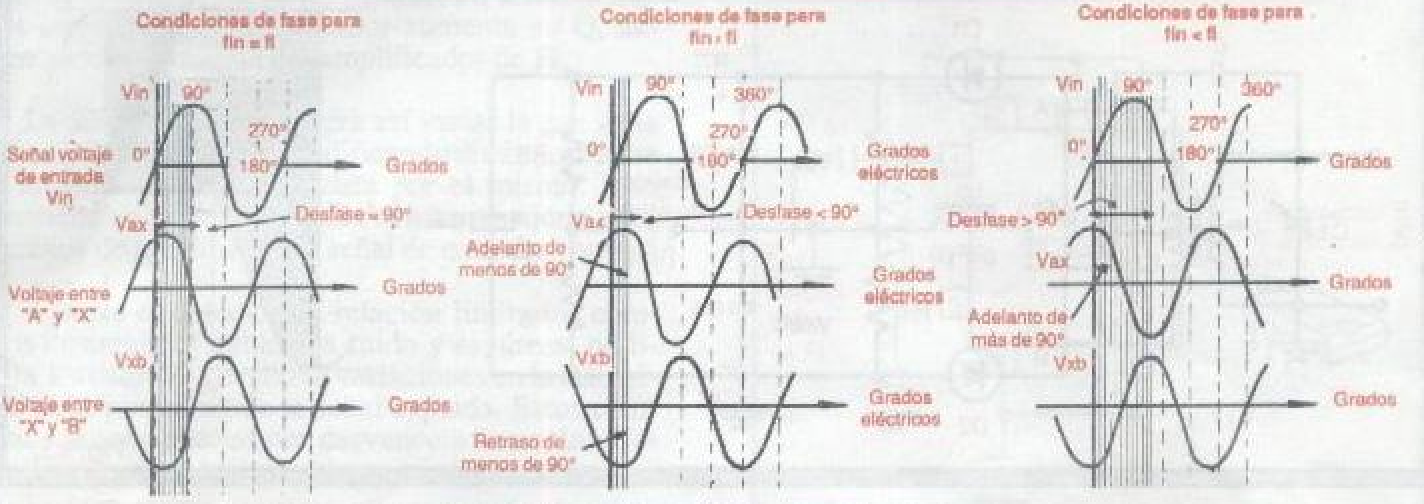
Condiciones de fase para
fin = fi
Vin
90°
270°
0°
180°
Señal voltaje de entrada Vin
Grados
Vax
Desfase = 90°
Voltaje entre "A" y "X"
Grados
Vxb
Voltaje entre "X" y "B"
Grados
Condiciones de fase para
fin › fi
Vin
90°
360°
270°
0°
180°
Grados eléctricos
Vax
Desfase < 90°
Adelanto de menos de 90°
Grados eléctricos
Vxb
Grados eléctricos
Retraso de menos de 90°
Condiciones de fase para
fin < fi
Vin
90°
360°
270°
0°
180°
Grados
Desfase > 90°
Vax
Grados
Adelanto de más de 90°
Vxb
Grados

fin = frecuencia instantánea de la señal de FI   Vax = Voltaje secundario entre los puntos "A" y "X"   Vxb = Voltaje secundario entre los puntos "X" y "B" Fig. 430

Característica del detector de fase

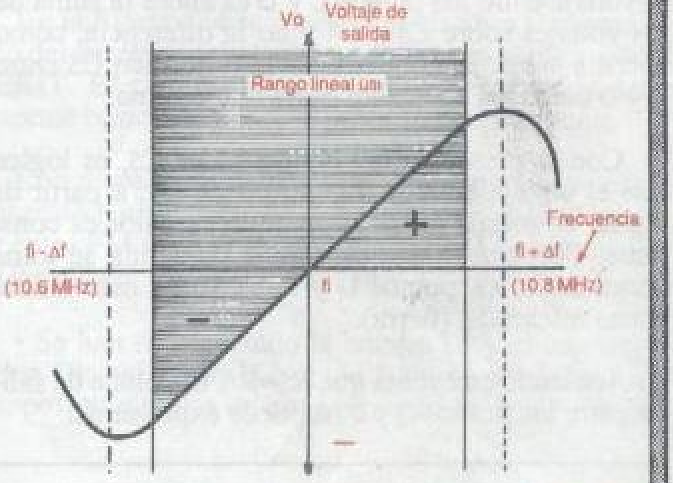

Fig. 431

caso, el voltaje neto de salida (Vo), medido entre los puntos C y D, será cero.

En los demás casos, uno de los voltajes de salida de los diodos D1 o D2 será mayor que el otro, dependiendo de la frecuencia instantánea de la señal de entrada. Bajo estas circunstancias, el voltaje neto de salida será positivo o negativo, según la desviación de frecuencia con respecto a la FI sea positiva o negativa.

De este modo, la magnitud y polaridad del voltaje de salida dependen de la magnitud y polaridad de la desviación de frecuencia de la señal de entrada. Este es precisamente el comportamiento que se espera de un detector de FM.

El detector de Foster-Seeley tiene una curva característica de respuesta de frecuencia como la que se muestra en la figura 431. Puede observarse que es más lineal que la del detector de pendiente balanceado.

Además, el detector de Foster-Seeley es mucho más fácil de alinear porque sólo utiliza dos circuitos sintonizados a la misma FI.

Su única desventaja es que no proporciona limitación de amplitud, es decir, las variaciones de amplitud en la señal de entrada provocan también cambios en la amplitud de la señal de salida. El detector de relación supera esta dificultad.

El detector de relación

En la figura 432 de muestra el circuito y la curva de respuesta de frecuencia de un detector de relación básico. Su característica más notable es que ignora los cambios de amplitud de la señal de entrada y por lo tanto no detecta ruido ni espúreas eventualmente asociadas a la señal de FM.

Con respecto al detector de Foster-Seeley, el detector de relación presenta tres modificaciones fundamentales:

- Se ha invertido la orientación o polaridad de uno de los diodos (el inferior, D2).

- Se ha insertado un condensador electrolítico de gran capacidad (C5) y un divisor de tensión resistivo (R5 y R6) en paralelo con los terminales de salida originales. En la práctica, R5=R6.

- La señal de salida se extrae ahora de otro par de puntos (Q y P).

Detector de relación básico

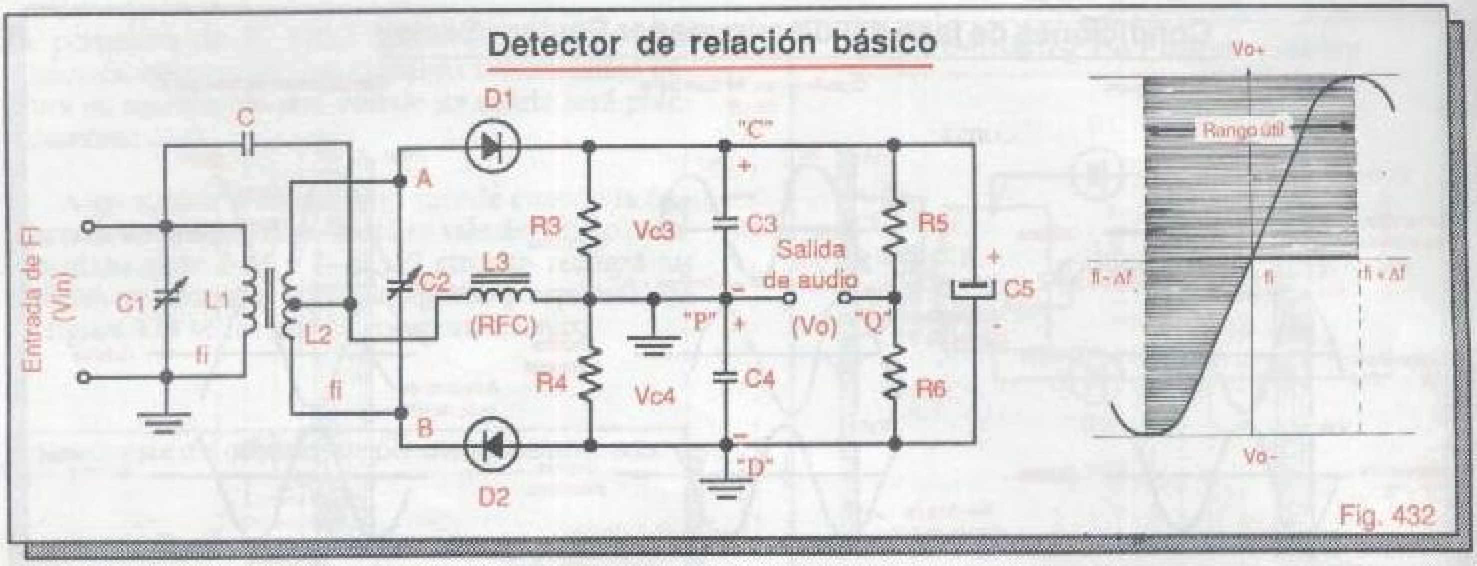

Fig. 432

El circuito opera así:

Con el diodo D2 invertido, el punto P es ahora positivo con respecto al punto D. Por consiguiente, el voltaje entre los puntos C y D es ahora la suma de los voltajes sobre C3 y C4 y no la diferencia, como sucedía antes. El condensador electrolítico C5 entre C y D conserva constante este voltaje suma.

Con la presencia del condensador C5, es lógico que el voltaje entre C y D no puede ser, a partir de ahora, el voltaje de salida porque su valor es constante. Por esta razón, el voltaje de salida se toma ahora entre los puntos Q y P, tomando este último como referencia (tierra).

Teniendo en cuenta que R5=R6, el voltaje de salida entre los puntos Q y P se puede expresar así:

$$Vo = Vqp = \frac{(Vcp - Vdp)}{2} = \frac{(Vc3 - Vc4)}{2}$$

En esta expresión, $V_o = V_{qp}$ es el voltaje de salida, medido entre los puntos Q y P, Vcp=VC3 el voltaje entre los puntos C y P o sea el voltaje sobre C3, medido con respecto a tierra, y Vdp=VC4 el voltaje entre los puntos D y P o sea el voltaje sobre C4, medido también con respecto a tierra.

El detector de relación se comporta exactamente igual ante las desviaciones de frecuencia con respecto a la portadora de la señal de entrada que el discriminador de Foster-Seeley y por eso tienen la misma curva "S" o característica. La diferencia está en cómo se comporta ante los cambios de amplitud.

El detector de relación como limitador de amplitud

Para explicar la forma cómo el detector de amplitud responde a los cambios de amplitud de la señal de entrada, asumamos inicialmente que el voltaje de esta última (Vin) ha permanecido constante durante un tiempo relativamente largo.

Bajo esta condición, C5 ha tenido tiempo de cargarse completamente al voltaje entre los puntos C y D y no fluirá corriente a través suyo.

En otras palabras, el condensador C5 se comportará como un circuito abierto. La resistencia de carga vista por los diodos D1 y D2 será la suma de R3 y R4, las cuales, en la práctica, son muy pequeñas comparadas con R5 y R6.

Si el voltaje de entrada tiende a incrementarse, el condensador C5 se opondrá a cualquier cambio en el voltaje de salida (Vo).

Este fenómeno sucede no sólo porque C5 tiene una gran constante de tiempo RC asociada a él. Ocurre también porque a medida que el voltaje de entrada aumenta, circula más corriente a través de los diodos y la corriente en exceso se inyecta a C5, cargándolo aún más.

En consecuencia, el voltaje entre C y D seguirá permaneciendo constante porque no es posible cambiar instantáneamente el voltaje a través de C5. El efecto neto de esto es que la corriente suministrada a la carga ha aumentado pero el voltaje sobre la carga no ha cambiado.

Lo anterior se interpreta como una disminución en la resistencia de la carga. Al sentirse más cargado el secundario del transformador, éste disminuirá su Q y, por lo tanto, la ganancia del amplificador de FI que maneja el detector disminuirá, compensando así el aumento del voltaje de entrada.

El fenómeno contrario sucede cuando el voltaje de la señal de entrada de FI tiende a disminuir. La corriente a tráves de los diodos también disminuirá pero el voltaje de carga no cambia. Al sentirse menos cargado, el transformador aumenta su Q, aumentando la ganancia del amplificador de FI.

La señal de entrada logra así variar la ganancia del amplificador de FI, modificando dinámicamente la impedancia de carga vista por el mismo. Esto mantiene un voltaje de salida constante, ajeno a los cambios de amplitud de la señal de entrada.

Aunque el detector de relación limita los cambios de amplitud debidos a ruido y espúreas, no limita los cambios debidos a variaciones en la intensidad de la portadora de la señal captada. Estos cambios son ocasionados por desvanecimiento, interferencias, al pasar de una emisora a otra, etc.

Por esta razón los receptores de FM que utilizan detectores de relación emplean algunas veces control automático de ganancia (CAG). El voltaje de CAG lo proporciona el mismo detector. Otros receptores utilizan, además de CAG, un circuito de limitación especializado antes del detector.

En general, el limitador de un receptor de FM es la etapa encargada de recortar el ruido y cualquier modulación de amplitud presente en la señal de entrada. De este modo, únicamente se demodula o detecta la modulación de frecuencia deseada.

Los limitadores utilizan generalmente tubos, transistores o circuitos integrados. En la figura 433 se muestra un ejemplo de un limitador para FM con el CI MC1590 de Motorola, diseñado específicamente para este propósito.

Limitador para FM con el CI MC1590

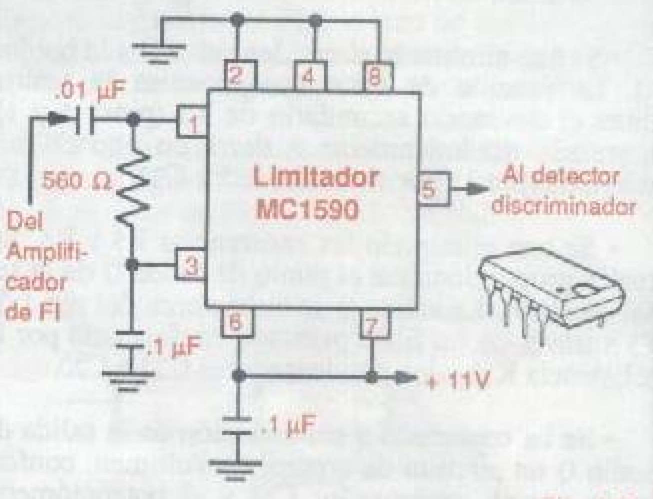
.01 µF
2
4
8
1
560 Ω
Limitador
MC1590
5
Al detector
discriminador
Del
Amplifi-
cador
de FI
3
6
7
.1 µF
+ 11V
.1 µF

Fig. 433

El detector de relación del radio FM CEKIT

En la figura 434 se muestra el circuito esquemático del detector de relación utilizado en el radio FM CEKIT. El circuito es una versión *balanceada* del detector básico de relación descrito anteriormente.

Con respecto al detector de relación básico de la figura 432, en el detector de relación balanceado del receptor FM CEKIT de la figura 434 se han realizado las siguientes modificaciones:

- Se han reemplazado la bobina L1 y el condensador C1, así como la bobina L2 y el condensador C2, por un circuito de acoplamiento de doble sinto-

Detector de relación balanceado del radio FM CEKIT

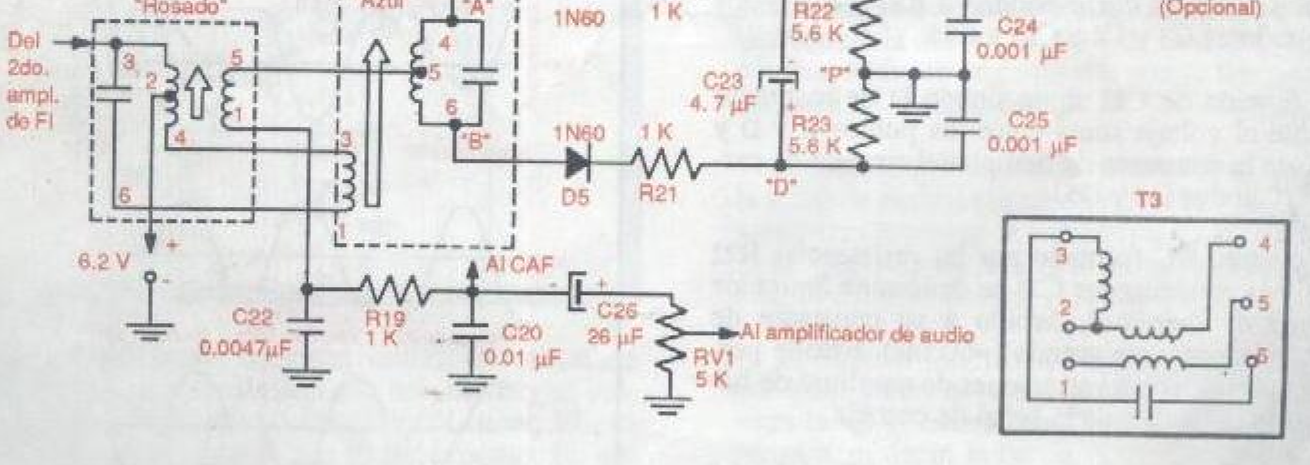
T3
(10.7 MHz)
"Rosado"
T4
(10.7 MHz)
"Azul"
Del
2do.
ampl.
de FI
3
2
5
1
4
6
"A"
4
5
6
"B"
3
1
D4
1N60
R20
1 K
"C"
"N"
Salida de CAG
(Opcional)
R22
5.6 K
C24
0.001 µF
C23
4.7 µF
"P"
R23
5.6 K
C25
0.001 µF
1N60
D5
1 K
R21
"D"
6.2 V
+
-
Al CAF
C22
0.0047µF
R19
1 K
C20
0.01 µF
C26
26 µF
Al amplificador de audio
RV1
5 K
T3
3
4
2
5
6
1

Fig. 434

nía constituido, respectivamente, por los transformadores sintonizados de FI T3 (núcleo rosado) y T4 (núcleo azul).

- Se han eliminado el condensador C y la bobina L3. La función de estos componentes la realiza ahora el devanado secundario de T3 (pines 5 y 1), conectado dinámicamente a tierra en uno de sus extremos (pin 1) por el condensador C22.

- Se han eliminado las resistencias R5 y R6, las cuales proporcionaban el punto de salida Q de la señal de audio. La salida Q se toma ahora del pin 1 de T3 a través de un filtro pasabajo conformado por la resistencia R19 y los condensadores C22 y C20.

- Se ha conectado a continuación de la salida de audio Q un circuito de control de volumen, conformado por el condensador C26 y el potenciómetro RV1.

- Se han conectado las resistencias R20 y R21 en serie con los diodos D4 y D5, respectivamente. Estos diodos cumplen la misma función de D1 y D2 en el detector básico. R20 y R21 amortiguan la corriente de alta frecuencia que circula a través de D4 y D5.

- Se ha invertido la polaridad de los diodos D4 y D5 y la del condensador C23 con respecto a la que tenían D1, D2 y C5, respectivamente, en el detector de relación básico. La razón de esto es permitir que del punto N se pueda tomar una tensión negativa de CAG (opcional) para otras partes del receptor.

En los demás aspectos, el detector de relación básico se ha mantenido sin modificaciones y la operación y características del circuito de la figura 434 son las mismas del circuito de la figura 432.

Con base en la descripción dada para este último usted puede entonces describir en detalle cómo trabaja el detector balanceado del radio FM CEKIT, sólo que ahora el condensador C5 se llama C23, las resistencias R3 y R4 corresponden a R22 y R23 y los condensadores C3 y C4 a C24 y C25.

La función de C23 sigue siendo la de mantener constante el voltaje suma entre los puntos C y D y establecer la constante de tiempo del circuito de carga de los diodos D4 y D5.

El circuito RC formado por las resistencias R22 y R23 y el condensador C23 se denomina *limitador dinámico de amplitud*, debido a su constante de tiempo relativamente grande (≈60 ms). Esto le permite seguir las lentas variaciones de amplitud de baja frecuencia (audio) de la señal de entrada.

Una constante de tiempo más baja dificultaría el alineamiento del circuito porque permitiría que el detector respondiera también a los cambios rápidos de amplitud provocados por el ruido y a los cambios más lentos de la misma ocasionados por señales espúreas.

El circuito RC constituido por los condensadores C22 y C20 y la resistencia R19 es un filtro pasabajos diseñado para eliminar las componentes de alta frecuencia de la señal de audio, en la misma forma como lo hace el filtro de FI del detector de AM (véase Lección 22).

El secundario de T3 se usa también para acoplar la baja impedancia de entrada del secundario de T4 con la alta impedancia del primario de T3, evitando que este último se cargue excesivamente por acción del detector de relación.

La disposición de los transformadores T3 y T4 permite obtener en el secundario de T4 voltajes inducidos cuyas fases dependen del valor instantáneo de la señal de FI aplicada al primario de T3. La amplitud de estos voltajes inducidos es constante dentro de una amplia gama de frecuencias.

Específicamente, al secundario de T4 le llegan dos señales:

a) La inducida por el primario de T3 (pines 3 y 2) en el secundario de T3 (pines 5 y 1) y aplicada por este último al tap central de T4 (pin 5).

b) La inducida por el primario de T4 (pines 3 y 1), el cual está en serie con el primario de T3.

Los voltajes secundarios de T4 aparecen desfasados 180º entre sí y 90º con repecto al voltaje inducido en el secundario de T3 cuando la frecuencia de la señal de entrada es igual a la FI, en menos

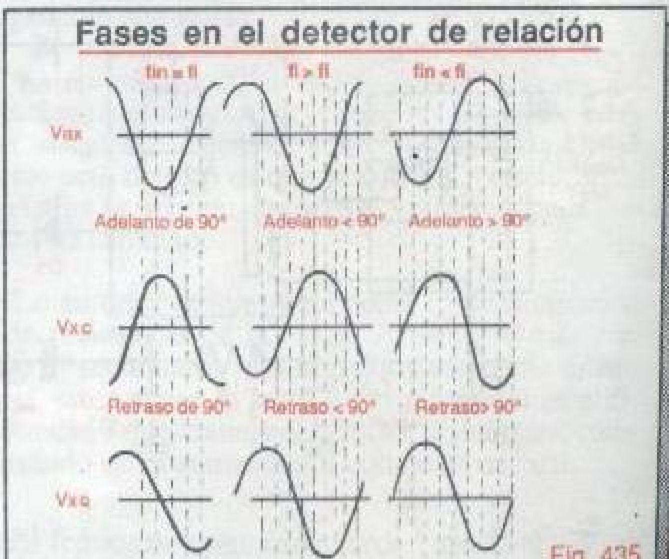
Fases en el detector de relación

Fig. 435

de 90º cuando la frecuencia es mayor y en más de 90º cuando es menor.

En la figura 435 se ilustran estos conceptos. La señal inducida en el secundario de T3 está en fase con la señal de FI suministrada por el segundo amplificador de FI al primario de T3.

Así mismo, el voltaje en el punto de salida será cero cuando la frecuencia de la señal de entrada sea igual a la FI y tendrá otro valor, positivo o negativo, cuando sea diferente, dependiendo de si la desviación de frecuencia con respecto a la FI es positiva o negativa.

De cualquier modo, la señal de salida (audio) será siempre proporcional a la desviación de frecuencia de la señal de entrada. Por consiguiente, el circuito convierte las variaciones de frecuencia de la señal de FI en una señal de audio equivalente.

Una configuración como la utilizada en el detector de relación de la figura 434 puede también utilizarse con el discriminador de fase de la figura 429. Esto no significa que, en la práctica, sea muy fácil cambiar un detector de relación balanceado en un discriminador de fase y viceversa.

Otros diseños de detectores de FM

Las dificultades encontradas en el alineamiento y construcción de los discriminadores de fase y de los detectores de relación han inspirado una serie de diseños de detectores de FM prácticamente libres de ajustes. Uno de tales circuitos es el discriminador a cristal que se muestra en la figura 436.

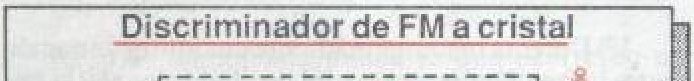

Discriminador de FM a cristal

Fig. 436

El discriminador a cristal utiliza un cristal de cuarzo (XT) en paralelo con una bobina (L2) en sustitución del circuito de doble sintonía usual. El circuito L1C1 es resonante a la FI del proceso. En general, C2=C3. El condensador C4 garantiza que se alimenten iguales cantidades de señal a D1 y D2.

Demodulación de FM con PLL

Desde su aparición en el mercado en forma de circuito integrado, el PLL o lazo de amarre de fase (véase Lección 25) ha revolucionado algunas facetas del diseño de los receptores de radio de FM.

El PLL es, intrínsecamente, un demodulador de FM. En la figura 437 se muestra el circuito práctico de un detector de FM con PLL NE565.

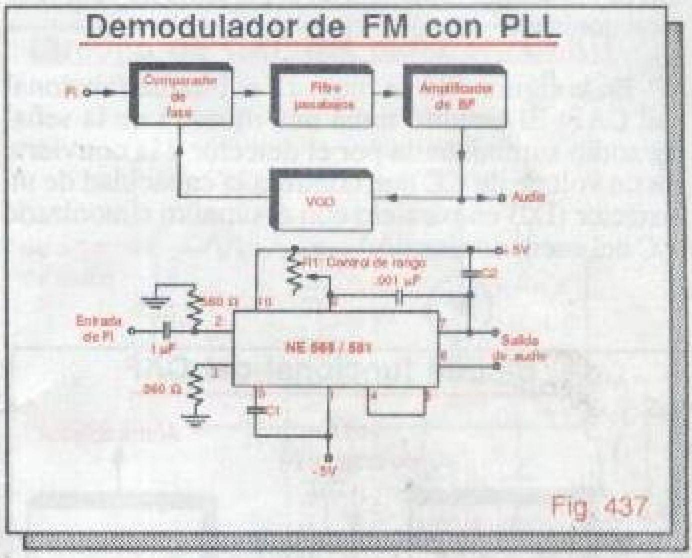

Demodulador de FM con PLL

Fig. 437

Un PLL consiste de un comparador de fase, un filtro pasabajos, un amplificador de baja frecuencia y un oscilador controlado por voltaje o VCO.

El VCO trabaja a una frecuencia muy próxima a la de la señal de entrada y el comparador de fase produce un voltaje de error proporcional a la diferencia de frecuencia entre el VCO y la señal de entrada. Este voltaje de error se aplica al VCO.

Cualquier cambio en la frecuencia de la señal de entrada es detectado por el comparador y convertido en un voltaje de error que reajusta automáticamente la frecuencia del VCO. De este modo, el VCO permanece siempre enganchado con la frecuencia instantánea de la señal de FI.

Debido a que el voltaje de error es una réplica de la señal de audio utilizada originalmente en la emisora para desplazar la frecuencia de la portadora, el PLL funciona directamente como un detector de FM. El filtro determina el ancho de banda.

El circuito R1C1 determina la frecuencia central del VCO y ésta debe ser igual a la FI de la señal de entrada. El condensador C2 determina el *rango de captura*, es decir, la banda de frecuencias alrededor de la FI dentro de la cual el PLL se engancha con la señal de entrada.

Lección 34

Control automático de frecuencia (CAF)

Introducción

La función del circuito de control automático de frecuencia o CAF es mantener constante la frecuencia del oscilador local. La acción del CAF garantiza que la FI tenga siempre su valor correcto de 10.7 MHz y la señal escuchada en el parlante no aparezca desintonizada.

En la figura 438 se muestra el bloque funcional del CAF. El circuito toma una muestra de la señal de audio suministrada por el detector y la convierte en un voltaje de CC que controla la capacidad de un varactor (D2) en paralelo con el circuito sintonizado LC del oscilador local.

Bloque funcional del CAF

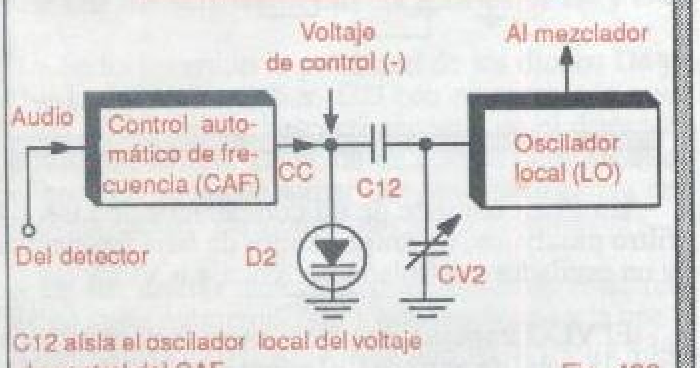

C12 aísla el oscilador local del voltaje de control del CAF

Fig. 438

En general, un circuito de CAF se compone de dos partes: un detector capaz de transformar las variaciones de frecuencia de la señal de FI en un voltaje de CC de control y una reactancia (condensador o bobina) cuyo valor pueda ser variado mediante este voltaje de control.

En nuestro caso utilizamos un varactor como reactancia variable. El varactor altera la frecuencia del oscilador local lo suficiente como para garantizar que al batirse esta frecuencia con la frecuencia de la portadora sintonizada se produzca nuevamente la frecuencia intermedia correcta de 10.7 MHz.

Por normas legales, la frecuencia de portadora de las emisoras de FM que transmiten en la banda de 88 a 108 MHz debe mantenerse constante durante todo el tiempo que duren sus transmisiones.

El anterior es un requisito básico impuesto a las mismas por el Ministerio de Comunicaciones y otros organismos encargados de regular este tipo de comunicaciones en cada país o estado. La máxima desviación admisible de la portadora es ±2 KHz.

Esta alta estabilidad de la frecuencia de transmisión posibilita el proceso heterodino en el receptor, es decir, el traslado de la información transmitida en una portadora de 88 a 108 MHz a una portadora interna de 10.7 MHz o FI.

Como sabemos, el proceso de heterodinación requiere de dos señales para generar la FI. Una es la señal de RF captada. La otra es la señal del oscilador local. Esta última debe ser también de frecuencia constante y estar siempre 10.7 MHz por encima de la frecuencia de la portadora de RF recibida.

En la práctica, por inestabilidad o distintas circunstancias, la frecuencia del oscilador cambia sin que haya cambiado la frecuencia de la señal de entrada. Cuando esto sucede se percibe en el parlante un efecto de distorsión, como si el radio estuviera mal sintonizado.

Esta distorsión ocurre porque el ancho de banda de la señal de FI (200 KHz) es sólo una pequeña fracción de la frecuencia (2%) de la portadora interna de 10.7 MHz. Cualquier error en la frecuencia del oscilador local causa que las etapas de FI.rechacen la señal deseada.

En síntesis, el CAF de un receptor de FM realiza dos funciones:

1º) Detecta instantáneamente cualquier variación entre la frecuencia actual del oscilador local y la frecuencia que debe producir de acuerdo a la señal sintonizada, originando un voltaje de control que es proporcional a la diferencia de ambas frecuencias.

2º) Utiliza el voltaje de control generado para gobernar un varactor y variar así las características de resonancia del circuito LC del oscilador local, retornando su frecuencia al valor adecuado.

Características de los circuitos de CAF

Como hemos visto, la función del CAF consiste en anular o minimizar las desviaciones de frecuencia del oscilador local con respecto al valor correcto necesario para producir una FI de 10.7 MHz.

Para que el circuito opere eficientemente, su acción debe suspenderse automáticamente cuando se alcance el valor deseado. En otras palabras, la tensión de control debe actuar de modo que la sintonía no se salga del ancho de banda del canal cuya frecuencia intermedia se pretende resintonizar.

Si esto no fuera así, una vez sintonizado el receptor a una emisora determinada, mantendría la sintonía para una extensión excesiva de frecuencias próximas, incluyendo las que están fuera del canal. Para escuchar una emisora cercana habría entonces que desconectar el radio y resintonizarlo.

En la figura 439 se ilustra la acción del CAF en un receptor de FM. La recta inclinada refleja el comportamiento entre la frecuencia deseada (eje horizontal) y la frecuencia real (eje vertical) del oscilador local en un receptor sin CAF. Obviamente, las dos frecuencias coinciden.

Acción del CAF en un receptor de FM

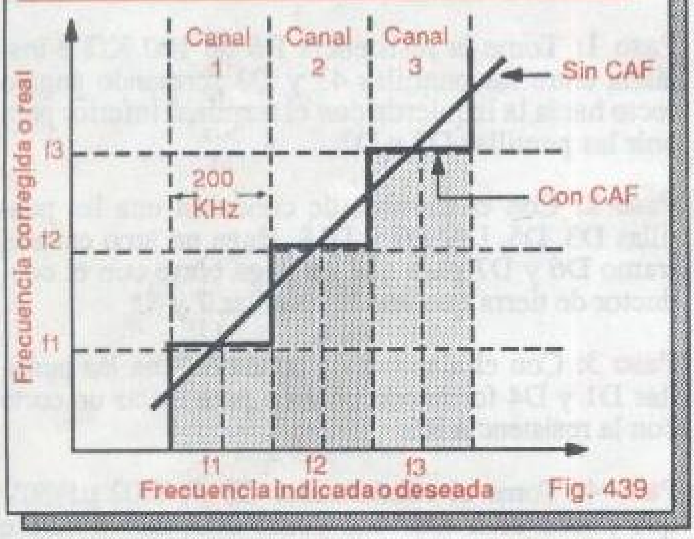

Fig. 439

La zona sombreada indica lo que ocurre en un receptor con CAF. Dentro de cada canal o emisora el CAF mantiene la frecuencia del oscilador local en su valor correcto y no permite que se sintonicen frecuencias de canales adyacentes.

Cuando se sintoniza otra emisora, la frecuencia del oscilador local no cambia hasta que la frecuencia deseada, establecida mediante el dial, se sale del canal de 200 KHz en curso.

La sinionización de la emisora contigua sobreviene entonces repentinamente y la acción del CAF produce instantáneamente la nueva frecuencia de oscilador requerida.

En los receptores de FM se puede incluir un interruptor para abrir el circuito de CAF y anular su efecto. Esto se hace con el fin de poder sintonizar una señal muy débil próxima a una muy fuerte.

Con el CAF conectado, la sintonización de esta señal marginal sería difícil porque el oscilador tendería a sincronizarse con la señal más fuerte, ignorando la más débil.

Operación del circuito de CAF del radio FM CEKIT

En la figura 440 se muestra el circuito de control automático de ganancia del radio experimental FM CEKIT. La frecuencia del oscilador local depende de la capacitancia neta entre los puntos A y B.

Circuito de CAF del radio FM CEKIT

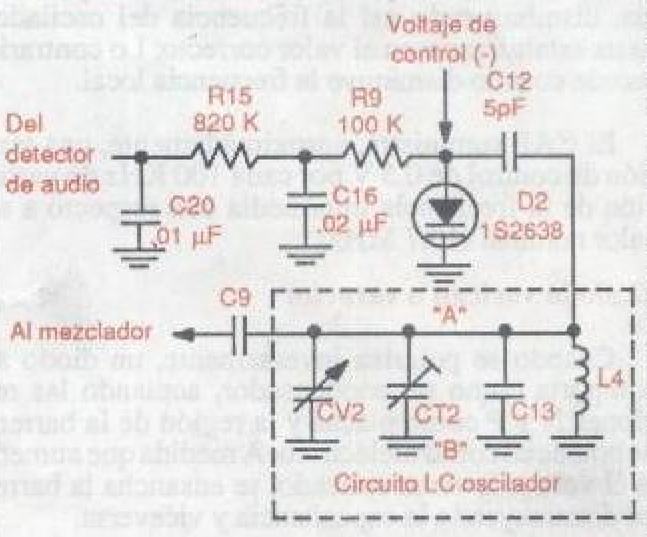

Fig. 440

El funcionamiento del CAF se basa en la acción del diodo varicap o varactor D2 sobre el circuito determinador de frecuencia del oscilador local (L4, CV2, CT2 y C13). Llamaremos a este último *circuito LC oscilador*.

El varactor D2, en serie con el condensador C12 y en paralelo con el circuito LC oscilador, se comporta como un condensador de capacitancia variable controlado por voltaje. El sustituto del varactor en los antiguos receptores a tubos era la llamada *válvula de reactancia*.

Para operar como un condensador variable controlado por voltaje, el diodo varactor D2 debe estar polarizado inversamente. A mayor tensión inversa sobre D2 menor capacidad y viceversa.

La polarización de D2 se obtiene conectando el cátodo a tierra y aplicando a su ánodo una tensión negativa de control proveniente de la salida del detector de audio y filtrada por el circuito pasabajos de las resistencias R15 y R 9 y los condensadores C16 y C20.

Una vez sintonizada una emisora, si la frecuencia del oscilador local se desplaza de su valor nominal, la tensión de control aplicada al ánodo también varía. Este cambio de voltaje sobre D2 ocasiona un aumento o disminución de su capacidad.

Esto a su vez modifica la capacidad total del circuito LC oscilador, el cual corrige automáticamente su frecuencia de resonancia y retorna la frecuencia del oscilador local a su valor nominal. Así se suministra a la salida del mezclador una señal diferencia siempre centrada en 10.7 MHz.

Por ejemplo, si la frecuencia del oscilador local aumenta, disminuye la tensión inversa aplicada al varactor. En consecuencia, aumenta su capacitancia, disminuyendo así la frecuencia del oscilador hasta estabilizarse en el valor correcto. Lo contrario sucede cuando disminuye la frecuencia local.

El CAF suministra, aproximadamente, una tensión de control de 0.5 V por cada 100 KHz de variación de la frecuencia intermedia con respecto a su valor nominal (10.7 MHz).

El diodo varicap o varactor

Cuando se polariza inversamente, un diodo se comporta como un condensador, actuando las regiones N y P como placas y la región de la barrera de potencial como dieléctrico. A medida que aumenta el voltaje inverso aplicado, se ensancha la barrera, disminuyendo la capacitancia y viceversa.

La mayoría de diodos se fabrican para que tengan una capacitancia mínima, excepto los varactores, los cuales se diseñan para proporcionar varios rangos de capacitancias, desde unos pocos pF hasta más de 100 pF.

El valor de la capacitancia se controla variando el ancho de la barrera de potencial mediante un voltaje inverso. En la figura 441 se muestran las características de un diodo varactor típico, similar al utilizado en el CAF del receptor FM CEKIT.

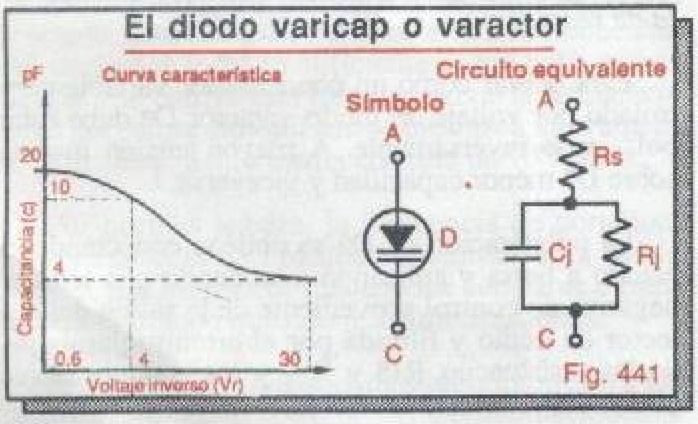

Fig. 441

Continuación Actividad práctica Nº 18

Ensamble de la primera etapa de FI. (Parte 2)

En esta actividad vamos a terminar el armado de la primera etapa de frecuencia intermedia o primer amplificador de FI, distinguido en el plano por el color verde.

Componentes y materiales necesarios

1 Transistor NPN referencia 2SC1390, o C1390 o 2N2222. Q2.

1 Resistencia de 100 KΩ, 1/4W, 5%. (café, negro, amarillo, dorado). R6.

1 Condensador de cerámica de 0.02 μF/50V. C3.

Alambre de conexión

Paso 1: Tome la resistencia R6 de 100 KΩ e instálela entre las puntillas 43 y D3 formando ángulo recto hacia la izquierda con el terminal inferior para unir las puntillas D3 y D2.

Paso 2: Con el alambre de conexión una las puntillas D3, D5, D6, D7 y D18. Haga un arco entre el tramo D6 y D7 para que no haga corto con el conductor de tierra que une las putillas 7 y 8.

Paso 3: Con el alambre de conexión una las puntillas D1 y D4 formando un arco para evitar un corto con la resistencia R6.

Paso 4: Tome el condensador C3 de 0.02 μF/50V que puede estar marcado como 203, 223 ó 0.02 e instálelo entre las puntillas D3 y 7. Si no alcanzan los terminales, se puede agregar alambre de conexión a uno de ellos para completarlo.

Paso 5: Tome el transistor C1390 e instálelo en la posición indicada para Q2 en las puntillas D4, D8 y D9. Fíjese muy bien en la posición de cada uno de los terminales, emisor (E), base (B) y colector (C). Si el transistor es un reemplazo, como el 2N2222, conéctelo como se indica en la figura A53.

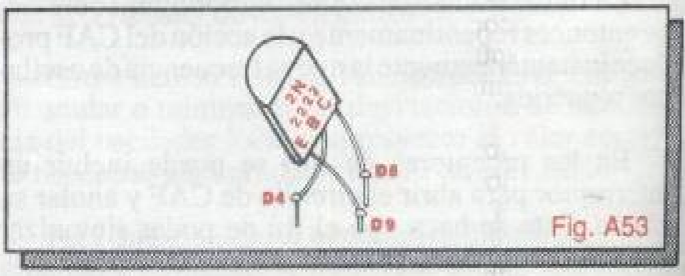

Fig. A53

Operación completa y ensamble del radio FM CEKIT

Introducción

En esta lección se resume la operación completa del receptor experimental de FM CEKIT y se describe el proceso de ensamble por etapas del mismo. Para hacerlo, se suministra el plano general del radio, el diseño de su circuito impreso y toda la información necesaria para su calibración y ajuste.

El ensamble de este receptor de FM le proporcionará a los estudiantes de este curso un gran conocimiento y experiencia sobre la práctica con los radios de FM, aspecto necesario en todo curso de electrónica: sin practicar no se aprende.

El estudio detallado de cada etapa ha sido ya cubierto en las lecciones anteriores. En esta lección se hace particular énfasis en los aspectos prácticos y se supone que el lector está familiarizado con la operación de todas las etapas, à excepción de la fuente de alimentación y el amplificador de audio.

Como equipo mínimo de prueba se necesitará un multímetro. Muchas de las pruebas de ajuste sugieren el uso de un generador de RF-FM y un osciloscopio. Si no posee estos instrumentos no se desanime. Puede utilizar como fuente de señal un receptor de FM que esté trabajando correctamente.

El receptor en bloques

En la figura 442 se muestra el diagrama de bloques del receptor de FM CEKIT. El circuito está formado básicamente por un amplificador de RF,un mezclador, un oscilador local, dos amplificadores de FI, un detector de FM, un circuito de CAF, un amplificador de audio y una fuente de alimentación.

A la antena convergen las señales provenientes de las diferentes estaciones del área. La antena convierte las ondas electromagnéticas interceptadas en señales eléctricas equivalentes y las suministra al amplificador de RF.

El amplificador de RF recibe las señales de FM provenientes de la antena, selecciona una en particular, por ejemplo la de 100 MHz, la amplifica y la suministra al mezclador.

El oscilador local y el mezclador constituyen el conversor de FI. La función del conversor es trasladar la modulación original de cualquier portadora de FM sintonizada en la banda de 88 a 108 MHz a una portadora interna fija de 10.7 MHz llamada *frecuencia intermedia* o FI.

El mezclador heterodina la señal proveniente del amplificador de RF (**rf**) con la onda continua producida por el oscilador local (**lo**) y obtiene la señal de FI de la diferencia de ambas frecuencias (**fi=lo-rf**). La frecuencia local es siempre 10.7 MHz mayor que la de la portadora de rf captada.

Los dos amplificadores de FI amplifican la señal de FI de salida del conversor y la suministran al detector o demodulador de FM.

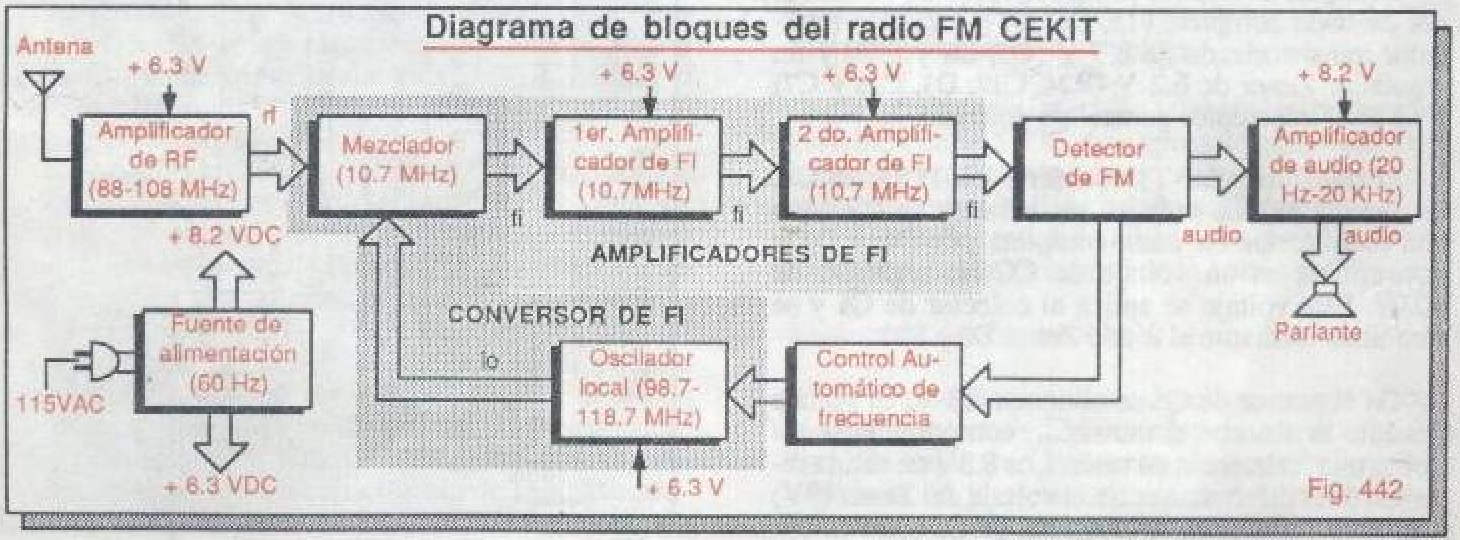
Diagrama de bloques del radio FM CEKIT
Antena
+ 6.3 V
Amplificador de RF (88-108 MHz)
rf
Mezclador (10.7 MHz)
+ 6.3 V
1er. Amplificador de FI (10.7MHz)
fi
+ 6.3 V
2 do. Amplificador de FI (10.7 MHz)
fi
Detector FM
+ 8.2 V
Amplificador de audio (20 Hz-20 KHz)
fi
audio
audio
+ 8.2 VDC
AMPLIFICADORES DE FI
Parlante
Fuente de alimentación (60 Hz)
115VAC
CONVERSOR DE FI
lo
Oscilador local (98.7-118.7 MHz)
Control Automático de frecuencia
+ 6.3 VDC
+ 6.3 V

Fig. 442

El detector recibe la señal de FM proveniente del segundo amplificador de FI, extrae la información de audio y la suministra al amplificador de audio. También limita la amplitud de la señal de FI y proporciona la tensión de control del CAF.

El CAF o control automático de frecuencia convierte la señal de salida del detector en un nivel de CC que se aplica a un diodo varactor y controla la frecuencia del oscilador local, manteniéndola constante y evitando que el receptor se desintonice.

El amplificador de audio recibe la señal de salida del detector, la amplifica y le suministra la potencia suficiente para impulsar el parlante.

El parlante convierte la señal eléctrica de audio en una información sonora audible que es una réplica de la originada en la emisora.

La fuente de alimentación convierte los 115 V de CA de la red pública en un voltaje de CC regulado de 8.3 V para el amplificador de audio y en otro de 6.2Vpara los amplificadores de RF y FI y el conversor. El detector y el CAF son pasivos y no requieren de fuente de alimentación.

El receptor en detalle

En la figura 443 se muestra el circuito esquemático completo del receptor de FM CEKIT. Se observa claramente la delimitación entre las 7 etapas diferentes de que consta el radio. Todos los componentes aparecen con su valores y referencias correspondientes.

También se destacan componentes claves como el circuito integrado amplificador de audio (U1), los dispositivos semiconductores (Q1, Q2, etc.), los transformadores de FI (T1, T2, etc.), el parlante y otros.

La fuente de alimentación consta de un rectificador de onda completa (T5, D6, D7 y C28), un regulador transistorizado de 8.3 V (Q5, D8 y C30) y un regulador Zener de 6.2 V (R24, C17, D3, C18 y C7). SW1 es el interruptor general de potencia.

El transformador T5 convierte los 115V de CA de entrada en dos voltajes secundarios de 9V que son rectificados en onda completa por D6 y D7 y convertidos en un voltaje de CC no regulado de ≈24V. Este voltaje se aplica al colector de Q5 y se estabiliza mediante el diodo Zener D8 y R25.

En el emisor de Q5 se obtienen 8.3 V. El voltaje restante lo absorbe el transistor, comportándose así como una resistencia en serie. Los 8.3V de salida resultan de la diferencia entre el voltaje del Zener (9V) y la caída directa en la unión BE de Q5 (0.7 V).

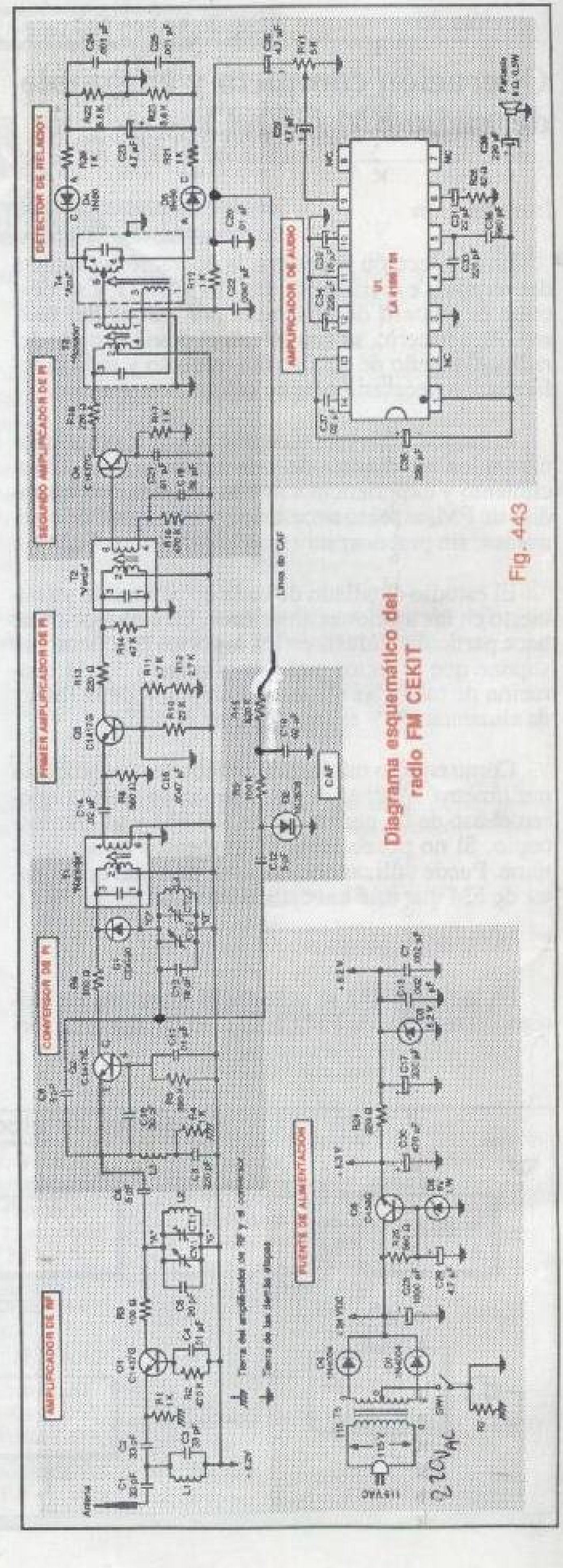

Diagrama esquemático del radio FM CEKIT

Fig. 443

El voltaje de salida del primer regulador (8.3V) alimenta el amplificador de audio y sirve también como voltaje de entrada del segundo regulador. El diodo Zener D3 convierte estos 8.3V en un voltaje estabilizado de 6.2V que alimenta los amplificadores de RF y FI y el conversor. El voltaje restante (2.1V) lo absorbe la resistencia R24.

Los condensadores electrolíticos C28, C30 y C17 eliminan las variaciones de baja frecuencia (ripple o rizado) de los voltajes de CC obtenidos y los condensadores cerámicos C18 y C7 las variaciones de alta frecuencia (RF). La resistencia R7 aísla la tierra de las etapas de RF (amplificador y conversor) de la tierra de las demás etapas.

El tema de las fuentes de alimentación se discute con mayor detalle en la Lección 36.

El amplificador de audio lo constituyen el circuito integrado U1 y componentes asociados. Los condensadores C26 y C27 transfieren la señal de salida del detector a la entrada de audio (pin 9). El potenciómetro RV1 actúa como divisor de señal y es el control de volumen del receptor.

Los condensadores C33 y C36 cumplen la función de realimentar la señal de salida de audio disponible en el pin 1 a los pines 4 y 5 . El parlante se conecta al pin 1 a través del condensador C38. El circuito RC formado por R26 y C31 provee amplificación uniforme para todas las frecuencias.

La tensión de alimentación de +8.2 V se aplica a los pines 14 (Vcc) y 3 (tierra). El condensador C37 es un filtro de RF. C35, C34 y C32 son filtros de rizado. Los pines 2, 7 y 8 no se utilizan.

Los amplificadores de audio con circuitos integrados se tratan con detalle en la Lección 38.

El detector o demodulador de FM lo constituyen los transformadores sintonizados de FI T3 y T4, el filtro pasabajos C22-R19-C20, los diodos detectores D4 y D5, las resistencias amortiguadoras R20 y R21, el circuito limitador dinámico C23-R22-R23 y el divisor capacitivo C24-C25.

La configuración empleada es un detector de relación balanceado. El detector recibe su señal de entrada del segundo amplificador de FI a través del transformador de acoplamiento T3. La señal de salida se aplica al amplificador de audio y al CAF.

El circuito de CAF lo constituyen el circuito pasabajos de dos etapas C20-R15-C16-R9, el diodo de capacitancia variable o varactor D2 y el condensador de desacople C12. La tensión de control proviene del detector. La capacitancia de salida se aplica en paralelo al circuito tanque del oscilador local.

El segundo amplificador de FI es el transistor Q4, conectado en la configuración emisor común y trabajando en clase A. La carga es el transformador sintonizado T3. Las resistencias R16, R17 y R18 son de polarización. Los condensadores C19 y C21 son de desacople de RF.

El primer amplificador de FI es el transistor Q3, conectado en la configuración base común y trabajando en clase A. La carga es el transformador sintonizado T2 y la resistencia R14. Las resistencias R8, R10, R11, R12 y R13 son de polarización. Los condensadores C14 y C15 son de desacople de RF.

El conversor de FI es el transistor Q2, conectado en la configuración base común y trabajando en un modo no lineal. La carga es el transformador sintonizado T1. El circuito tanque L4-CT2-CV2-C13 determina la frecuencia del oscilador local. El circuito L3-C8 es una trampa de FI.

Las resistencias R4, R5 y R6 son de polarización. Los condensadores C6, C9 y C11 son de desacople de RF. El diodo D1 es un limitador de amplitud. El condensador C10 es de estabilización.

El amplificador de RF es el transistor Q1, conectado en la configuración base común y trabajando en clase A. La carga de salida es el circuito sintonizado de banda estrecha L2-CT1-CV1. La carga de entrada es el circuito de banda ancha L1-C3.

Las resistencias R1, R2 y R3 son de polarización. Los condensadores C1, C2 y C4 son de desacople de RF. El condensador CV1, acoplado mecánicamente al condensador CV2 del conversor, es el control de sintonía del receptor. Asociada al amplificador de RF está la antena.

El condensador CV1 permite sintonizar cualquier emisora de la banda de FM (88 a 108 MHz) captada por la antena. El condensador CV2 controla automáticamente la frecuencia del oscilador local (LO) entre 98.7 y 118.7 MHz para que sea siempre 10.7 MHz mayor que la sintonizada.

La antena es una sonda telescópica de longitud variable entre 65 y 85 cms aproximadamente. También se puede utilizar como antena un alambre de cobre de longitud fija cortado a 75 cms. Este valor es 1/4 de la longitud de onda correspondiente a 98 MHz, la frecuencia central de la banda de FM.

Ensamble del receptor de FM

Se describe a continuación el proceso de ensamble del receptor de FM por etapas en una tarjeta de circuito impreso. Comenzaremos por la fuente de alimentación y terminaremos en el amplificador de RF, pasando por los demás circuitos del receptor.

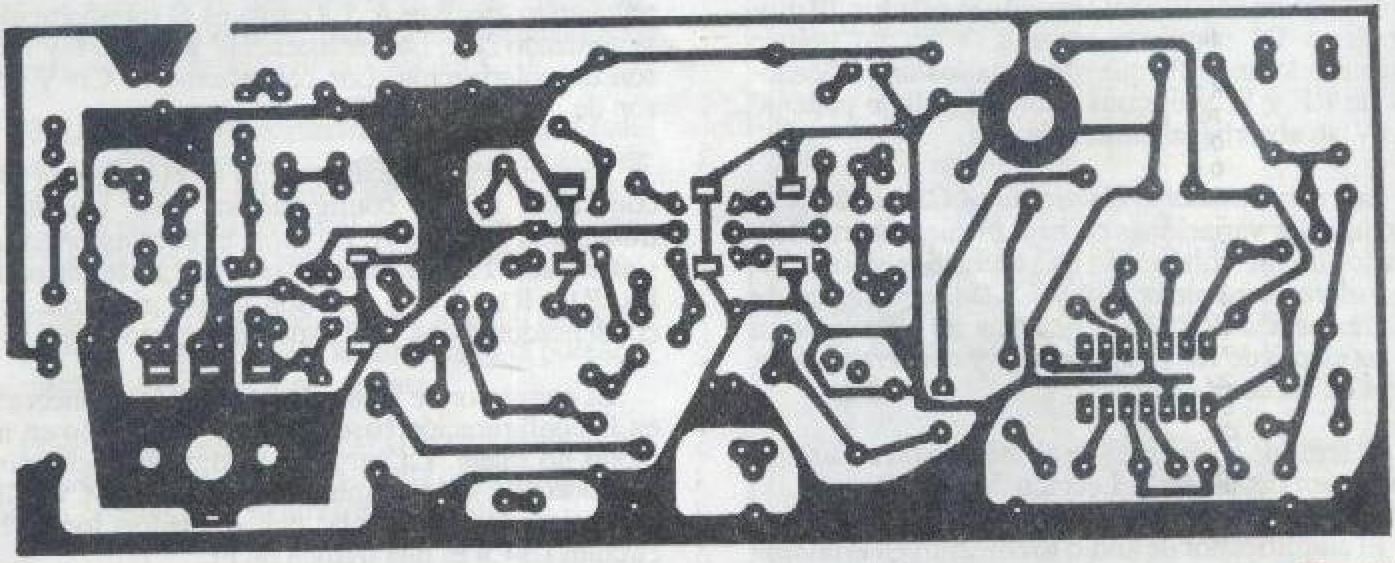

Circuito impreso del radio FMCEKIT. Lado del cobre tamaño real

Fig. 444

En la figura 444 se muestra el circuito impreso completo del radio FM CEKIT visto por el lado del cobre. Todos los componentes, incluyendo los controles de sintonía y de volumen, se alojan en esta tarjeta, a excepción del transformador de potencia, la antena y el parlante que son externos a la misma.

En la figura 445 se muestra el aspecto que presenta la tarjeta de circuito impreso del receptor vista por el lado de los componentes. En la Tabla 1 se relaciona la lista completa de materiales y su referencia en el plano general de la figura 443.

Nota: El circuito impreso del receptor de FM CEKIT lo puede elaborar usted mismo, siguiendo por ejemplo las instrucciones del Kit K550 de CEKIT, o lo puede solicitar directamente a CEKIT. Todos los componentes son de fácil consecución.

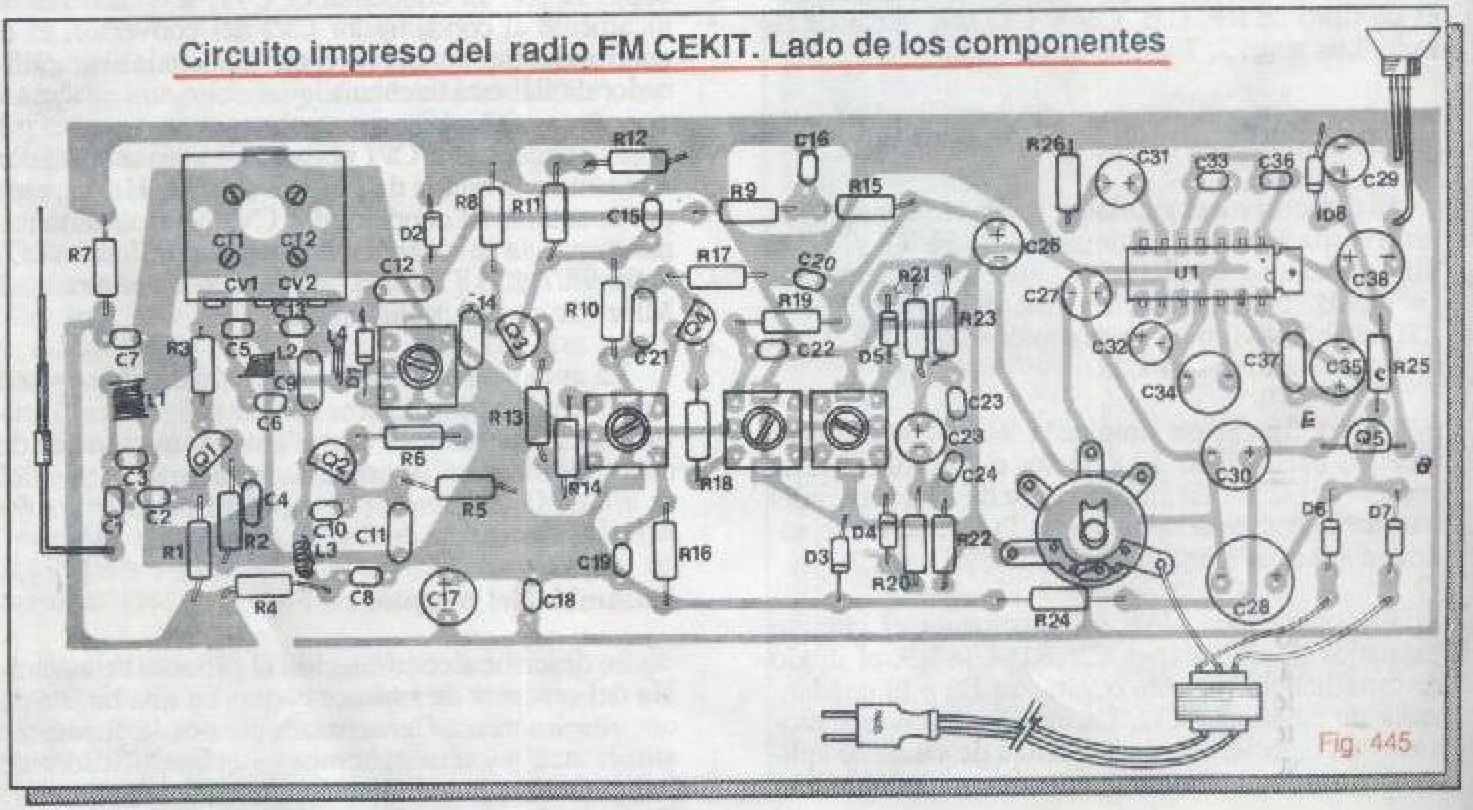
R7
C11
C12
CV1
CV2
C7
L1
R3
C5
L2
C9
C6
Q1
Q2
C3
C1
C2
R1
R2
C4
C10
L3
C11
C8
R4
D2
C12
L4
C14
R6
R5
R8
R11
R12
C15
R9
R15
C16
R17
C20
R10
C21
Q3
Q4
R19
C22
R13
R14
R18
C19
R16
C17
C18
R21
R23
D5
C23
C23
C24
D4
D3
R20
R22
R24
R26
C31
C33
C36
D8
C29
C25
C27
U1
C38
C32
C37
C35
R25
C34
E
Q5
C30
D6
D7
C28

Circuito impreso del radio FM CEKIT. Lado de los componentes

Fig. 445

Tabla 1. Lista general de componentes y materiales del radio experimental FM CEKIT

Resistencias fijas de 1/4W, 5 %

6 Resistencias de 1K (R1, R4, R17, R19, R20, R21).
2 Resistencias de 470 K (R2, R16).
3 Resistencias de 100 Ω (R3, R6, R7).
1 Resistencia de 390 K (R5).
2 Resistencias de 560 Ω (R8, R25).
1 Resistencia de 100 K (R9).
1 Resistencia de 27 K (R10).
1 Resistencia de 4.7 K (R11).
1 Resistencia de 2.7 K (R12).
3 Resistencias de 220 Ω (R13, R18, R24).
1 Resistencia de 47 K (R14).
1 Resistencia de 820 K (R15).
2 Resistencias de 5.6 K (R22, R23).
1 Resistencia de 82 Ω (R26).

Resistencias variables

1 Potenciómetro lineal o logarítmico de 5 K (RV1) con interruptor (SW1) y perilla.

Condensadores fijos de cerámica o poliéster de 50V

3 Condensadores de 33 pF (C1, C2, C3)
5 Condensadores de .01 µF (C4, C11, C19, C20, C21)
1 Condensador de 20 pF (C5)
3 Condensadores de 5 pF (C6, C9, C12)
5 Condensadores de .02 µF (C7, C14, C16, C18, C37)
2 Condensadores de 220 pF (C8, C33)
1 Condensador de 30 pF (C10)
1 Condensador de 18 pF (C13)
2 Condensadores de .0047 µF (C15, C22)
2 Condensadores de .001 µF (C24, C25)
1 Condensador de 560 pF (C36)

Condensadores variables

1 Condensador de sintonía para FM (CV1) con perilla o dial.

Condensadores electrolíticos

4 Condensadores de 220 µF/10V(C17, C34, C35, C38)
4 Condensadores de 4.7µF/10V (C23, C26, C27, C29)
1 Condensador de 1000 µF/16V (C28)
1 Condensador de 470 µF/10V (C30)
1 Condensador de 22 µF/10V (C31)
1 Condensador de 10µF/10V (C32)

Bobinas de núcleo de aire

L1: Alambre #20: 5 espiras. Diámetro del núcleo: 4mm.
L2: Alambre #20: 4 espiras. Diámetro del núcleo: 4 mm.
L3: Alambre #20: 12 espiras. Diámetro del núcleo: 4 mm o
Alambre #23: 20 espiras. Diámetro del núcleo: 3 mm.
L4: Alambre #19: 3 espiras. Diámetro del núcleo: 4 mm.

Transformadores sintonizados de FI para FM

1 Transformador de FI núcleo naranja, 5 pines (T1).
1 Transformador de FI núcleo verde, 5 pines (T2).
1 Transformador de FI núcleo rosado, 6 pines (T3).
1 Transformador de FI núcleo azul, 5 pines (T4).

Transformadores de potencia (220VAC) / CON PUNTO MEDIO (9 + 9) (1A/1A)

T5: Entrada=115V, Salida=18V, con tap central.
Potencia= 2VA. (Omega ODC110 o similar).

Diodos

1 Diodo de conmutación CDG00 o ECG177 (D1).
1 Diodo Varicap o Varactor 1S2638 o ECG611 (D2).
1 Diodo Zener de 6.2V, 1 W (D3)*.
2 Diodos detectores de germanio 1N60 (D4, D5).
2 Diodos rectificadores 1N4004 (D6, D7).
1 Diodo Zener de 9.0 V, 1W (D8)**

* ECG137A o similar
** ECG139A o similar

Transistores NPN

4 Transistores C1417G o BF194 (Q1, Q2, Q3, Q4).
1 Transistor C458G (Q5) o TIP31.

Circuitos integrados

1 CI LA4100 ó LA4101 ó ECG1180 (U1) con base de 14 pines.

Accesorios

- Plaqueta de circuito impreso de 77 x 192 mm^2.
- Parlante de 8Ω, 0.25 a 1 W.
- Antena telescópica de FM o alambre de 75 cm.
- Soldadura de aleación 40/60 (40% de estaño, 60% de plomo).
- Cordón de potencia de 115 V.
- 1 Caja plástica o metálica de proyectos (opcional)
- Cable de conexiones #20.

Herramientas

- 1 Cautín de 25 a 35 W.
- 1 Pinza de corte lateral (cortafríos).
- 1 Pinza larga de punta plana.
- 1 Destornillador plástico de pala para ajuste de transformadores de FI.

Instrumentos

1 Multímetro digital o análogo.
1 Generador de FM o de RF o 1 Radio de FM en servicio.
1 Osciloscopio de 20 MHz (opcional).

Recomendaciones generales para el ensamble

Instale los componentes siguiendo el orden establecido para evitar contratiempos y facilitar la identificación de fallas. Utilice las pinzas planas para doblar los terminales a la longitud adecuada y los cortafríos para eliminar el alambre excedente.

Sea metódico. No continúe con la siguiente etapa hasta no estar seguro del correcto funcionamiento de la etapa que está armando. La mayor parte de las fallas en el armado son causadas por falta de cuidado en la instalación de los componentes y soldaduras deficientes o "frías", entre otros factores.

Antes de instalar los componentes de cada etapa, realice un rápido chequeo de los mismos con el multímetro para asegurarse de que no estén internamente abiertos o en cortocircuito. Si éste es el caso, reemplácelos. Recuerde:

a) Las resistencias presentan un valor de acuerdo a su código. La resistencia de los potenciómetros es fija entre sus extremos y continuamente variable entre el cursor y cualquier extremo, dependiendo de la posición del eje.

b) Los diodos y uniones BE y BC de los transistores presentan una alta resistencia en inverso y una baja resistencia en directo. Si posee o tiene acceso a un probador de transistores y diodos, el chequeo será más exacto.

c) El alambre que constituye las bobinas y los transformadores tiene una muy baja resistencia interna. En un transformador, la resistencia entre cualquier terminal del primario y cualquier terminal del secundario o entre éstos y el blindaje debe ser muy alta.

d) La resistencia interna de un condensador cerámico es muy alta, prácticamente un circuito abierto. Con un capacímetro podrá medir su valor exacto.

e) Los condensadores electrolíticos se cargan rápidamente al voltaje de la pila interna del óhmetro y luego comienzan a descargarse, a una velocidad que depende de su valor. A mayor valor, más lenta es la descarga y viceversa.

f) No intente probar el circuito integrado con el multímetro porque lo puede destruir internamente o causarle averías irreparables.

Para obtener buenas soldaduras, limpie siempre la punta del cautín antes de cada nuevo punto y asegúrese de que las superficies a unir estén limpias. Caliente primero el terminal a soldar y luego acerque la soldadura hasta que se derrita y forme un montículo. En la figura 446 se resume el proceso.

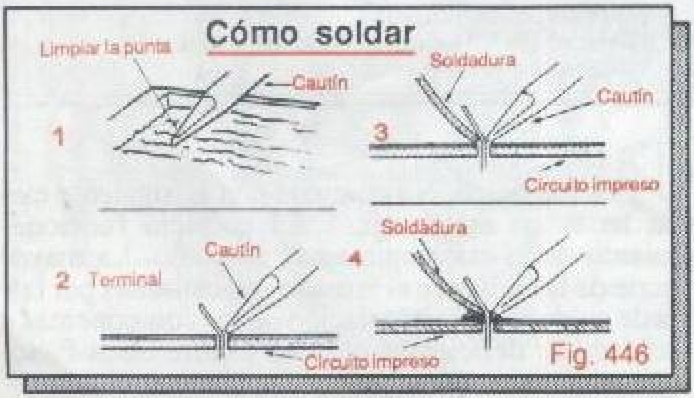

Fig. 446

Antes de comenzar a soldar, estañe la punta del cautín para evitar su oxidación. Límpiela con una esponja húmeda o con una esponjilla y aplíque soldadura hasta formar una capa delgada y uniforme.

Mientras suelda, aplique sólo la cantidad de calor necesaria para permitir que la soldadura fluya uniformemente en la unión, sin levantar el cobre ni calentar excesivamente el componente. Retire entonces el cautín. Asegúrese de no mover el componente hasta que la soldadura se haya enfriado.

Por seguridad, no realice soldaduras con la energía conectada al radio y aísle adecuadamente los terminales del transformador una vez soldados para minimizar el riesgo de un "shock" eléctrico.

Para la instalación del circuito integrado, suelde primero la base de 14 pines al circuito impreso y luego monte sobre ella el integrado. La utilización de una base facilita su montaje y reemplazo.

Si una etapa no le funciona como debe ser, revise cuidadosamente todas las conexiones guiándose por el diagrama esquemático. Efectúe pruebas de continuidad entre los terminales que comparten un punto común del circuito impreso para verificar malos contactos o roturas en el cobre.

Realice con el multímetro todas las medidas que le sean posible e interprételas para localizar la causa de la falla. Si posee un osciloscopio y un generador de señales, el seguimiento de la falla será más sencillo. Verifique que los voltajes de la fuente estén alimentando adecuadamente cada etapa.

En la Lección 41 se trata en detalle el tema de la reparación de radios de FM.

Cuando utilice un radio de FM en servicio como fuente de señal, utilice un condensador en serie de más de 50pF y cable blindado para transferir esa señal al radio bajo prueba. La línea de blindaje debe unir las tierras de ambos receptores.

Una vez finalizado el ensamble y ajuste de su radio, constrúyale algún tipo de contenedor o chasís de plástico, madera o metálico, para protegerlo y darle una apariencia firme de producto terminado.

Fuente de alimentación

En la figura 447 se muestran los detalles constructivos de la fuente de alimentación del radio FM CEKIT. Observe la forma de identificar los terminales del transformador de potencia T5, del transistor Q5 y de otros componentes especiales.

Para probar la operación de la fuente, conecte el cordón de potencia a la red de 115VAC y accione el

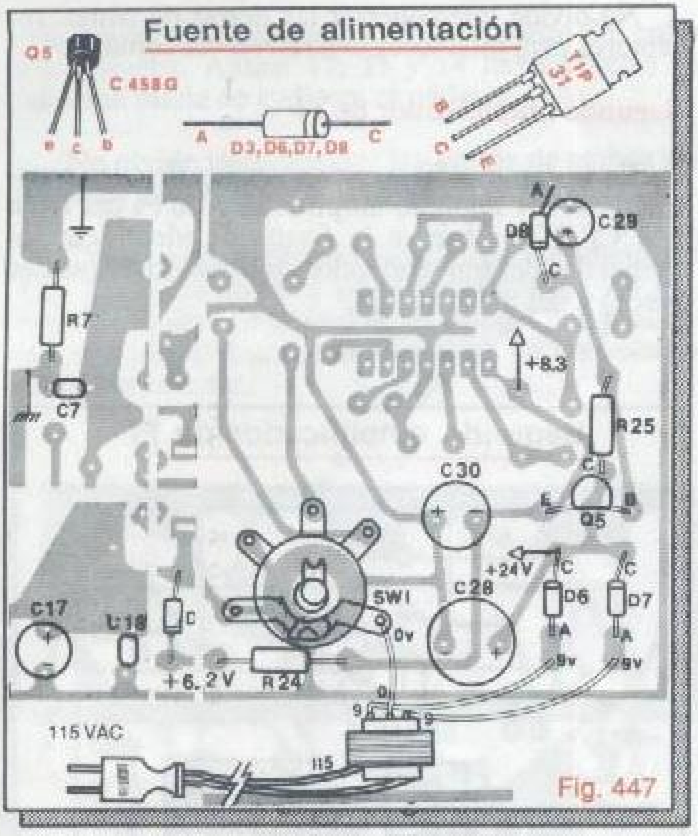
Fuente de alimentación
Q5
C 458G
e c b
A
D3,D6,D7,D8
C
TIP 31
B
C
E
A
D8
C29
C
R7
C7
+8.3
R25
C30
C
E
B
Q5
SW1
0v
+24V
C
C28
D6
A
C
D7
A
C17
C18
C
9v
9v
+6.2V
R24
0
9
9
115 VAC
H5

Fig. 447

interruptor SW1. Ajuste el multímetro como voltímetro de CC en la escala "30VDC".

Conecte la punta negativa a tierra y toque con la punta positiva el colector, la base y el emisor de Q5 y el cátodo de D3. Consigne los valores obtenidos. La tabla 2 puede servirle de guía. Note que el voltaje en el emisor de Q5 es algo mayor de 8.3V debido a que la fuente está sin carga.

VCQ5	VBQ5	VEQ5	V D3
> 18 V	9.1 V	> 8.3 V	6.2 V

Tabla 2

Amplificador de audio

En la figura 448 se muestran los detalles constructivos del amplificador de audio del radio FM CEKIT. Observe la forma de identificar los pines del CI y la manera de conectar el control de volumen. El CI va montado en la base de 14 pines para facilitar su remoción o reemplazo.

Para probar la operación de la etapa, accione el interruptor SW1 y mida nuevamente los voltajes en los terminales del transistor Q5 de la fuente. El amplificador de audio actúa como carga del regulador de 8.3 V. Este debe ser el valor de voltaje medido en el emisor de Q5.

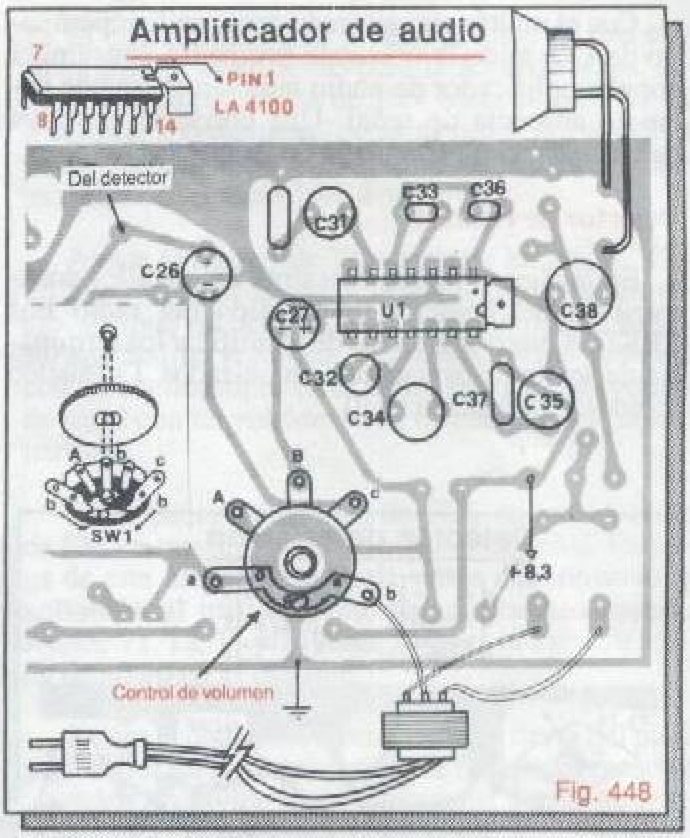
Amplificador de audio
7
PIN 1
LA 4100
8
14
Del detector
C31
C33
C36
C26
C27
U1
C38
C32
C37
C35
C34
A
b
c
b
b
SW1
B
A
c
a
b
+8.3
Control de volumen

Fig. 448

Utilizando un generador de audio o el inyector de señales construido en la actividad práctica Nº 10, inyecte una señal de audio a la entrada del amplificador. Deberá escucharse un tono en el parlante y su intensidad debe aumentar o disminuir dependiendo de la posición del control de volumen.

Con el multímetro ajustado como voltímetro de CC, mida los voltajes en los pines del circuito integrado U1, primero sin señal de entrada (control de volumen al mínimo) y luego con señal (control de volumen en una posición intermedia).

La tabla 3 le servirá de guía para tener una idea del orden de magnitud de los voltajes que debe obtener.

PIN	1	2	3	4	5	6	7	
Volt.	4.1	0.	0.	5.3	0.8	4.1	0	V

PIN	8	9	10	11	12	13	14	
Volt.	3.5	4.1	4.1	0	8.0	8.1	8.2	V

Tabla 3

Con el multímetro ajustado como miliamperímetro de CC, mida la corriente promedio consumida por el amplificador de audio tanto en presencia como en ausencia de señal. Una corriente excesiva puede indicar una condición de cortocircuito.

Detector de relación

En la figura 449 se muestran los detalles constructivos del detector de relación del radio FM CEKIT. Observe la forma de identificar los terminales de los transformadores sintonizados T3 (núcleo rosado) y T4 (núcleo azul).

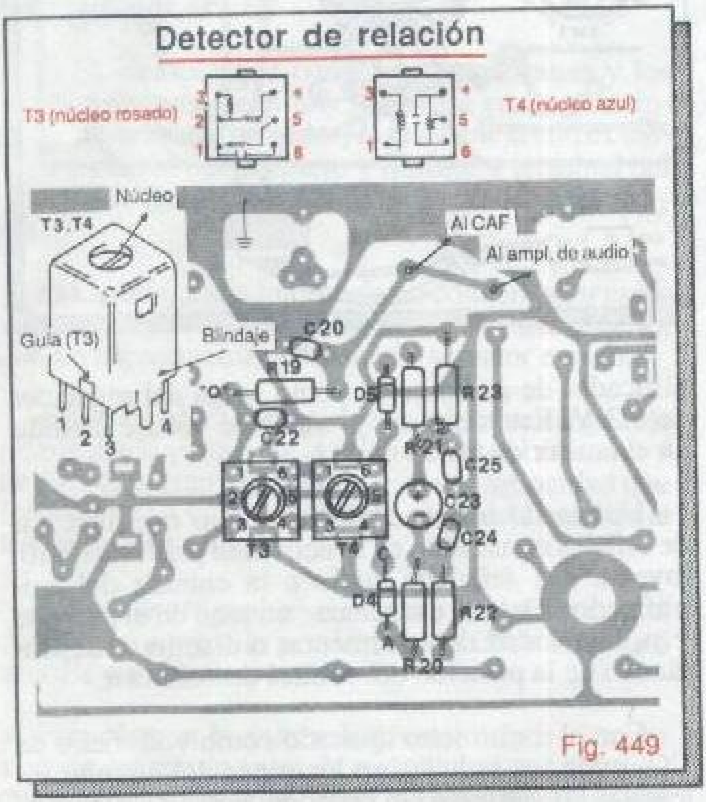
Detector de relación

Fig. 449

Para la prueba de esta etapa, inyecte una señal de FM de 10.7 MHz al terminal de entrada (pin 3 de T3). Deberá escuchar un tono en el parlante. Si suspende la modulación, el tono deberá desaparecer.

Ajuste el núcleo de T3 y T4 hasta obtener la máxima señal de salida de audio. El tono escuchado en el parlante deberá alcanzar su máximo volumen. Visualice la señal de salida con un osciloscopio en el punto Q o mida el voltaje de señal con un voltímetro de CA en paralelo con el parlante.

Si no posee generador de FM, utilice un radio de FM en servicio como generador de señal. Extraiga de este último la señal de salida del segundo amplificador de FI y aplíquelo al mismo punto del radio bajo prueba. Ajuste T3 y T4 hasta obtener la máxima salida de audio en el parlante.

No olvide interconectar las tierras de ambos receptores para establecer una referencia común.

Segundo amplificador de FI

En la figura 450 se muestran los detalles constructivos del segundo amplificador de FI del radio FM CEKIT. Observe la forma de identificar los terminales del transformador sintonizado T2 (núcleo verde) y del transistor Q4.

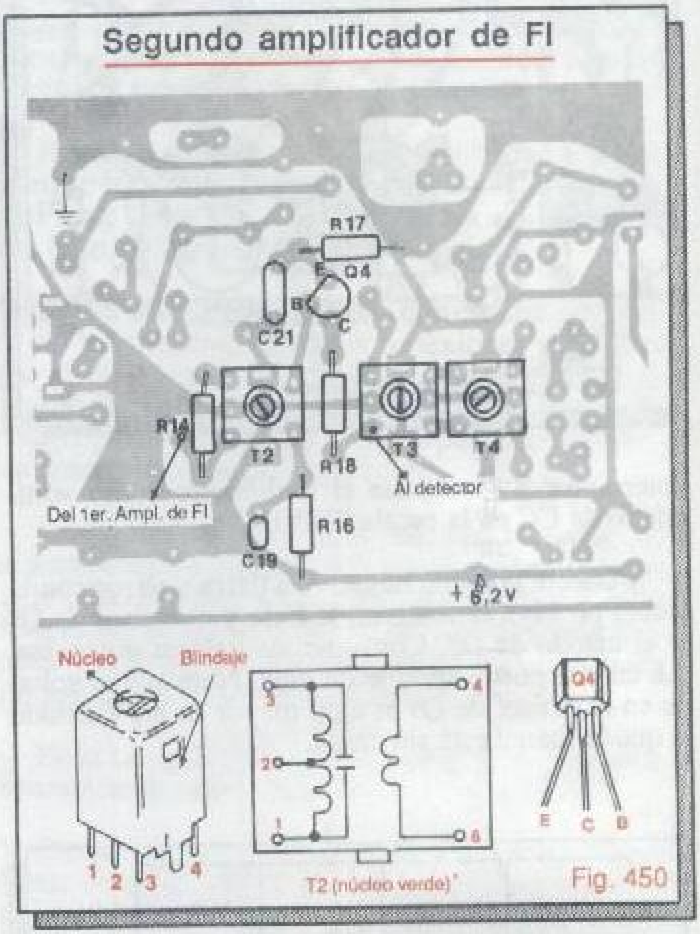
Segundo amplificador de FI

Fig. 450

El procedimiento de prueba y ajuste de esta etapa es similar al utilizado para el detector. Inyecte una señal de 10.7 MHz modulada en frecuencia al terminal de entrada (pin 3 de T2). Deberá escuchar un tono en el parlante. Si suspende la modulación, el tono deberá desaparecer.

Ajuste el núcleo de T2 y reajuste los de T3 y T4 hasta obtener la máxima señal de salida de audio. El tono escuchado en el parlante deberá alcanzar su máximo volumen. Visualice la señal de salida con el osciloscopio o mida el voltaje de señal con el voltímetro de CA en paralelo con el parlante.

Si no posee generador de FM, utilice un radio de FM en servicio como generador de señal. Extraiga de este último la señal de salida del primer amplificador de FI y aplíquelo al mismo punto del radio bajo prueba. Ajuste T2, T3 y T4 hasta obtener la máxima salida de audio en el parlante.

No olvide interconectar las tierras de ambos receptores para establecer una referencia común.

Con el multímetro ajustado como voltímetro de CC, mida los voltajes de polarización en los terminales de Q4. La tabla 4 le servirá de guía para tener una idea del orden de magnitud de los voltajes que debe obtener.

VB	VC	VE	V BE	
1.11	5.9	0.42	0.7	V

Tabla 4

Primer amplificador de FI

En la figura 451 se muestran los detalles constructivos del primer amplificador de FI del radio FM CEKIT. Observe la forma de identificar los terminales del transformador sintonizado T1 (núcleo naranja) y del transistor Q3.

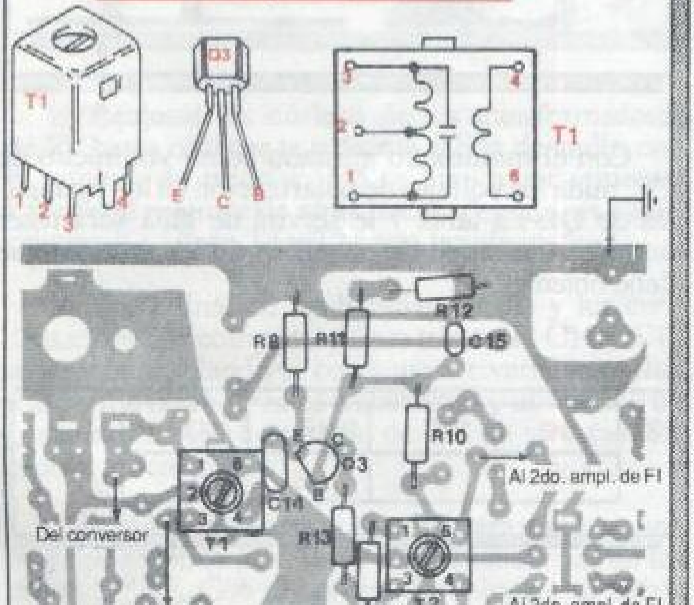
Primer amplificador de FI

Fig. 451

El procedimiento de prueba y ajuste de esta etapa es similar al utilizado para la segunda etapa. Inyecte una señal de FM de 10.7 MHz al pin 1 de T1 y conecte momentáneamente el pin 2 del mismo a tierra. Deberá escuchar un tono en el parlante. Sin modulación, el tono debe desaparecer.

Ajuste el núcleo de T1 y reajuste los de T2, T3 y T4 hasta obtener la máxima señal de salida de audio. El tono escuchado en el parlante deberá alcanzar su máximo volumen. Visualice la señal de salida con un osciloscopio en el punto Q o mida el voltaje de señal con un voltímetro de CA en paralelo con el parlante.

Si no posee generador de FM, utilice un radio de FM en servicio como generador de señal. Extraiga de este último la señal de salida del conversor y aplíquelo al mismo punto del radio bajo prueba. Ajuste T1, T2, T3 y T4 hasta obtener la máxima salida de audio en el parlante.

Retire la conexión momentánea a tierra del pin 2 para evitar un posible corto en la fuente del receptor modelo. No olvide interconectar las tierras de referencia de ambos receptores.

Con el multímetro ajustado como voltímetro de CC, mida los voltajes de polarización en los terminales de Q3. La tabla 5 le servirá de guía para tener una idea del orden de magnitud de los voltajes que debe obtener.

VB	VC	VE	V BE	
1.46	6	0.76	0.7	V

Tabla 5

Conversor de FI y CAF

En la figura 452 se muestran los detalles constructivos del conversor y el CAF del radio FM CEKIT. Observe la forma de identificar los terminales del transistor Q2, el condensador variable CV2 y su trimmer CT2 asociado. Se indica también la forma de fabricar las bobinas L3 y L4.

Para la prueba de esta etapa, inyecte una señal de FM de 10.7 MHz modulada en frecuencia al extremo libre de C6. Ajuste la amplitud de esta señal a un nivel mínimo para no afectar el rango dinámico de las etapas. Deberá escuchar un tono en el parlante. Sin modulación, el tono debe desaparecer.

Reajuste nuevamente los núcleos de T1, T2, T3 y T4 hasta obtener la máxima señal de salida de au-

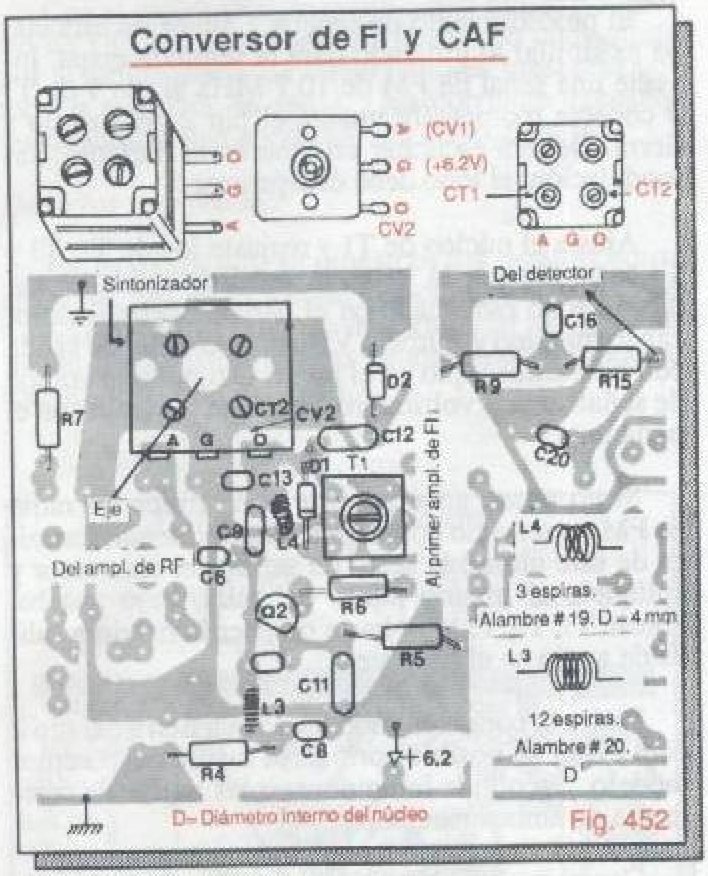
Conversor de FI y CAF
(CV1)
(+6.2V)
CT1
CT2
CV2
A G O
Sintonizador
Del detector
C15
D2
R9
R15
R7
CT2
CV2
A G O
C12
C20
D1
T1
C13
Eje
C9
L4
Al primer ampl. de FI
Del ampl. de RF
C6
Q2
R6
R5
3 espiras.
Alambre # 19. D = 4 mm
L3
C11
12 espiras.
Alambre # 20.
D
C8
R4
+6.2
D= Diámetro interno del núcleo

Fig. 452

dio. El tono escuchado en el parlante deberá alcanzar su máximo volumen. Visualice la señal de salida con el osciloscopio o mida el voltaje de señal con el voltímetro de CA en paralelo con el parlante.

Si no posee generador de FM, utilice un radio de FM en servicio como generador de señal. Extraiga de este último la señal de salida del amplificador de RF y apliquela al mismo punto del radio bajo prueba. Ajuste T1, T2, T3 y T4 hasta obtener la máxima salida de audio en el parlante.

No olvide interconectar las tierras de ambos receptores para establecer una referencia punto.

Con el multímetro ajustado como voltímetro de CC, mida los voltajes de polarización en los terminales de Q2. La tabla 6 le servirá de guía para tener una idea del orden de magnitud de los voltajes que debe obtener.

VB	VC	VE	V BE	
1.63	6	0.98	0.7	V

Tabla 6

Amplificador de RF

En la figura 453 se muestran los detalles constructivos del amplificador de RF del radio FM CEKIT. Observe la forma de identificar los terminales del transistor Q1, el condensador variable CV1 y su trimmer CT1 asociado. Se indica tambien la forma de fabricar las bobinas L1 y L2.

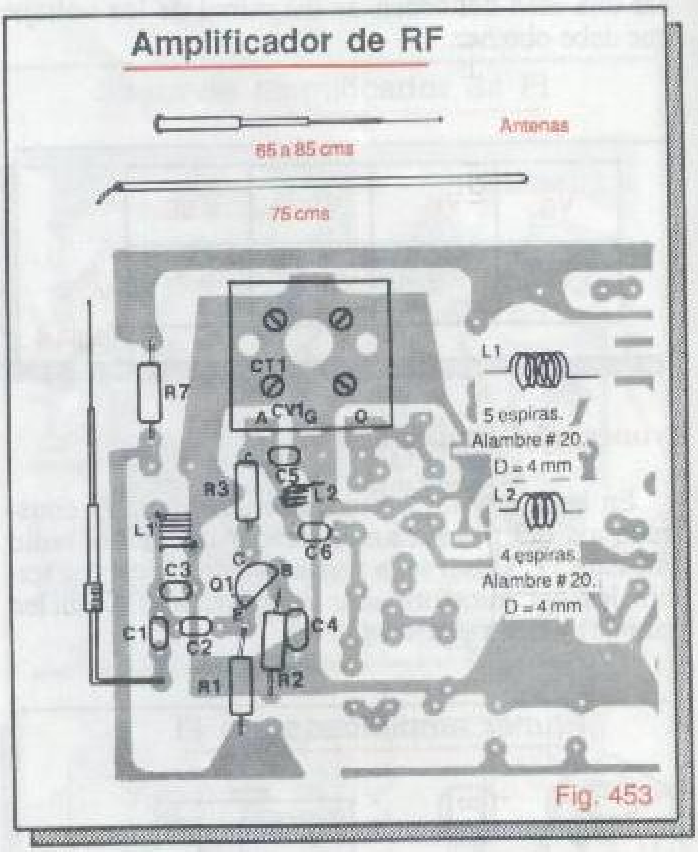
Amplificador de RF
Antenas
65 a 85 cms
75 cms
CT1
CV1
A G O
R7
L1
5 espiras.
Alambre # 20.
D = 4 mm
R3
C5
L2
L2
C6
4 espiras.
Alambre # 20.
D = 4 mm
L1
C3
Q1
C
B
E
C1
C2
C4
R1
R2

Fig. 453

Con el multímetro ajustado como voltímetro de CC, mida los voltajes de polarización en los terminales de Q1. La tabla 7 le servirá de guía para tener una idea del orden de magnitud de los voltajes que debe obtener.

VB	VC	VE	V BE	
1.46	6	0.76	0.7	V

Tabla 7

Una vez terminado el ensamble de esta etapa y con la antena correctamente instalada, el receptor debe poder sintonizar más de una emisora con un nivel aceptable de volumen.

Si éste no es el caso y todas las demás etapas están funcionando correctamente, rectifique las conexiones del amplificador. Busque posibles circuitos abiertos, cortocircuitos, componentes mal instalados, soldaduras "frías", etc. No se desanime.

Si la falta de sintonización persiste y Q1 está trabajando correctamente, realice lo siguiente:

a) Desconecte el cable de antena de su televisor y conecte a uno de sus extremos un condensador de 50 pF en serie con la antena de su radio (figura 454). Sitúe el control de volumen al máximo y gire lentamente la perilla de sintonía (CV1/CV2) hasta sintonizar la emisora más débil.

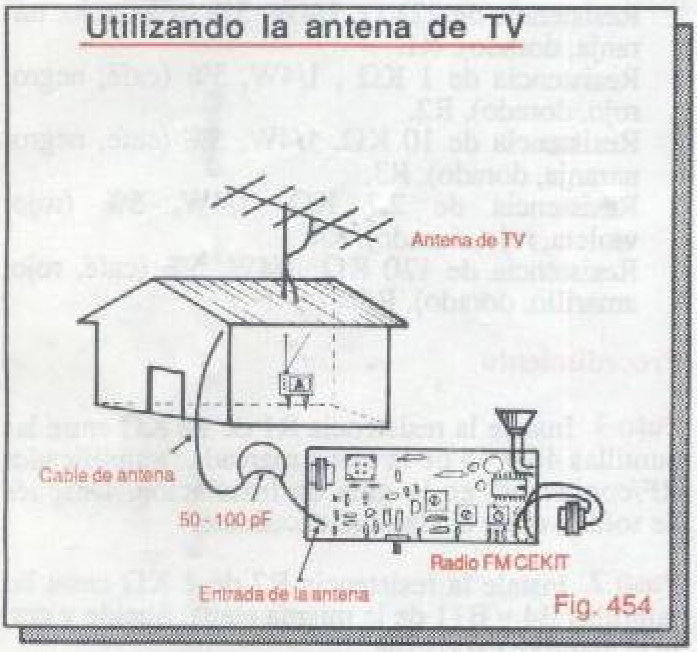
Utilizando la antena de TV

Fig. 454

b) Reajuste los núcleos de los transformadores de FI hasta obtener la máxima salida de audio con el mínimo de ruido, como se hizo anteriormente. Efectúe el reajuste de atrás hacia adelante, es decir, primero T4, luego T3 y T2 hasta terminar con T1.

El ajuste final de las bobinas L2, L4 y los condensadores de compensación o trimmers CT1 y CT2 se realiza acoplando al condensador variable el dial o la perilla indicadora de frecuencias y utilizando un generador de RF o un radio de FM en servicio. Siga este procedimiento:

1º) Calibre el generador de RF para entregar una señal de 88 MHz modulada en frecuencia por un tono de 1 KHz. Con el control de volumen al máximo, sitúe también el dial de su receptor en 88 MHz.

Aproxime o aplique la señal de RF del generador a la antena a través de un condensador de 50 pF o mayor. Abra o cierre los devanados de la bobina osciladora L4 hasta escuchar el tono de modulación de 1 KHz en el parlante.

Si no posee generador de RF-FM, sintonice tanto el radio modelo como el receptor a calibrar en 88 MHz o la frecuencia más proxima. Ajuste L4 hasta escuchar en el parlante de su radio la misma emisora que se está oyendo en el radio modelo, con el máximo volumen y el mínimo de ruido.

2º) Calibre ahora el generador de RF para entregar una señal de FM de 108 MHz. Con el control de volumen al máximo, sitúe también el dial de su receptor en 108 MHz.

Aproxime o aplique la señal de RF del generador a la antena. Con un destornillador no metálico, calibre el trimmer CT2 del oscilador hasta escuchar el tono de modulación en el parlante.

Si utiliza un radio de FM como patrón de frecuencia, sintonice ambos receptores a 108 MHz, o la estación más próxima, y ajuste CT2 hasta escuchar en el parlante la misma emisora, con el máximo volumen y el mínimo de ruido.

3º)Calibre el generador o el radio modelo a 93 MHz y el dial de su receptor a la misma frecuencia. Abra o cierre las espiras de la bobina L2 de sintonía de RF hasta escuchar nítidamente el tono de modulación o la misma emisora en el parlante.

4º) Calibre el generador o el radio patrón a 103 MHz y el dial de su receptor a la misma frecuencia. Ajuste el trimmer CT1 hasta escuchar claramente en el parlante el tono de modulación o la misma emisora sintonizada en el radio de muestra.

Una vez terminada la calibración, recorra el dial para verificar que la frecuencia de la emisora sintonizada coincida con la marcada. Repita los pasos 1º) hasta 4º), si es necesario, hasta obtener una correspondencia óptima.

Una vez ensamblado y ajustado su radio y en pleno funcionamiento, sólo falta conseguirle o crearle un gabinete plástico o metálico adecuado, con acceso a los controles de sintonía y volumen y que sirva de soporte al circuito impreso, la antena, el parlante y el transformador de potencia. Así su radio tendrá una apariencia profesional.

En la figura 455 se muestra un ejemplo de gabinete acrílico abierto que usted puede construir fácilmente para su radio.

En los almacenes especializados se venden cajas plásticas o metálicas terminadas de varias dimensiones, diseñadas específicamente para alojar proyectos. Lo único que tiene que realizar usted son las perforaciones para operar externamente el control de volumen y el control de sintonía y asegurar firmemente el transformador de potencia y el parlante.

Presentación final del radio FM CEKIT

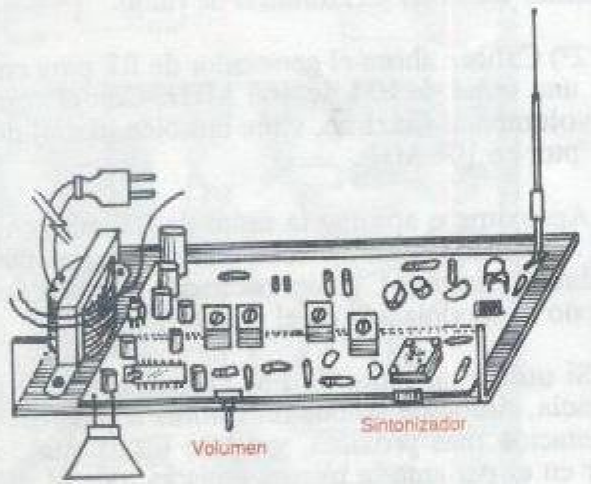

Fig. 455

Con unas pocas herramientas esenciales y un procedimiento adecuado, la construcción de un chasís metálico para su radio es relativamente fácil.

Utilice preferiblemente aluminio, no sólo porque es un material que provee un excelente blindaje si no porque es más fácil de trabajar que otros metales y proporciona buenos contactos a tierra. El aluminio se vende en forma de láminas de diferentes espesores y es muy maleable.

Usted puede ahorrar mucho dinero y energía si dedica suficiente tiempo a la planeación de su trabajo. Cuando todos los detalles han sido previstos, la construcción del chasís se simplifica enormemente.

Cuando perfore agujeros en un metal con un taladro, es importante localizar primero los centros con un centro punto, que es un estilete de acero de punta afilada. De esto modo se evita que la broca se deslice o patine con respecto al centro cuando se va a empezar a perforar.

Los agujeros de más de 6 mm de diámetro deben perforarse en dos pasos: primero con una broca pequeña y finalmente con la broca definitiva. Para realizar círculos muy grandes o extraer superficies arbitrarias como rectángulos, cuadrados, etc. de la lámina, siga estas recomendaciones:

Realice a todo lo largo del perímetro interno de la figura una serie de pequeñas perforaciones muy próximas entre sí. Desprenda la lámina y suavice los bordes con una lima fina.

Actividad práctica Nº 19

Ensamble del amplificador / conversor de RF. Parte 1.

En esta primera parte de la actividad 19 instalaremos en el tablero del radio AM CEKIT una parte de los componentes que forman el amplificador/conversor de RF, la etapa marcada con color amarillo en la lámina correspondiente al diagrama esquemático del receptor.

Componentes necesarios

1 Resistencia de 22KΩ, 1/4W, 5% (rojo, rojo, naranja, dorado). R1.
1 Resistencia de 1 KΩ , 1/4W, 5% (café, negro, rojo, dorado). R2.
1 Resistencia de 10 KΩ, 1/4W, 5% (cafe, negro, naranja, dorado). R3.
1 Resistencia de 2.7 KΩ, 1/4W, 5% (rojo, violeta, rojo, dorado). R4.
1 Resistencia de 120 KΩ, 1/4W, 5% (café, rojo, amarillo, dorado). R5.

Procedimiento

Paso 1. Instale la resistencia R1 de 22 KΩ entre las puntillas 46 y 45 de la etapa marcada "amplificador RF/conversor" en la guía de instalación. Después de soldar, corte el alambre excedente.

Paso 2. Instale la resistencia R2 de 1 KΩ entre las puntillas B4 y B11 de la misma etapa. Suelde y corte el alambre excedente.

Paso 3. Instale la resistencia R3 de 10 KΩ entre las puntillas B3 y 3 de la etapa. Suelde y corte el alambre excedente.

Paso 4. Instale la resistencia R4 de 2.7 KΩ entre las puntillas B6 y 4. Suelde y corte el alambre excedente.

Paso 5. Instale la resistencia R5 de 120KΩ entre las puntillas B14 y B15. Doble hacia la derecha, en ángulo recto, los terminales excedentes de ambas puntillas y conéctelos a las puntillas B17 y B19, respectivamente. Suelde y corte el alambre excedente

R1 y R3 suministran la corriente de polarización de la base del transistor Q1. R2 proporciona la corriente de polarización del colector del mismo y minimiza el riesgo de oscilaciones parásitas. R4 estabiliza térmicamente el punto de trabajo. R5 amplía el ancho de banda del transformador T1.

El transistor Q1 y el transfomador T1 se instalarán en una futura actividad.

Fuentes de poder para radios AM y FM

Introducción

Todos los circuitos electrónicos necesitan algún tipo de energía para poder funcionar; algunas veces se requiere corriente alterna y otras veces continua.

En el caso de los radios de AM y FM, éstos obtienen su energía por medio de dos sistemas: el de pilas o baterías y el de fuentes de poder.

Cuando el radio está fabricado para entregar una baja potencia de sonido, de unos 2 a 5 vatios como máximo, se alimenta con pilas. Si el radio tiene una mayor potencia se alimenta por medio de una fuente de poder.

Las fuentes de poder pueden ser internas o externas. Son internas cuando sus componentes forman parte integral del circuito del radio o de un equipo amplificador de sonido que tiene como uno de sus módulos un radio.

Es externa cuando está fabricada en forma independiente y se conecta al radio por medio de un cable y un conector.

La función básica de una fuente de poder es convertir corriente alterna en corriente continua (figura 456). En la Lección 3 de este curso ya hemos estudiado la naturaleza y el funcionamiento de las pilas y baterías en sus diferentes tipos.

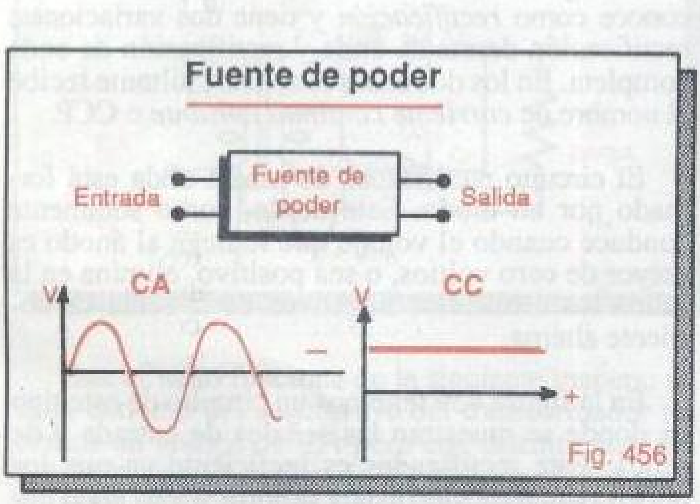

Fig. 456

En esta lección vamos a estudiar el funcionamiento de las fuentes de poder y sus diferentes componentes.

Fuentes de poder no reguladas

Una fuente de poder o convertidor no regulado transforma corriente alterna en corriente continua, con un voltaje de salida que puede variar de acuerdo a las circunstancias de funcionamiento. En algunos circuitos de radio esta variación no afecta su funcionamiento y se puede tolerar.

En la figura 457 tenemos el diagrama de bloques en donde se muestra cómo está formada una fuente de poder no regulada o convertidor de voltaje.

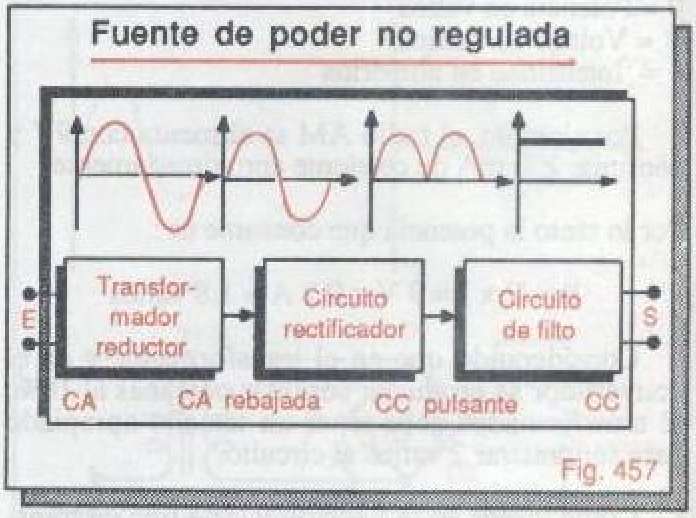

Fig. 457

El primer bloque siempre contiene un transformador que se encarga de rebajar el voltaje de la red de alimentación, ya sea de 110 voltios o de 220 voltios, a un valor menor, generalmente comprendido entre 6 y 12 voltios.

El segundo bloque se llama circuito rectificador y se encarga de eliminar los semiciclos negativos de la señal o voltaje de corriente alterna. Esta operación se puede realizar en diferentes formas que estudiaremos más adelante en esta misma lección.

El tercer bloque del convertidor es el circuito de filtro que se encarga de *pulir* la señal para llevarla a una forma de onda lo más cercana posible a la corriente continua.

El transformador en la fuente de poder

Debido a que los circuitos internos de los radios trabajan con voltajes bajos, el primer paso que se debe realizar es rebajar el voltaje de entrada que es alto, generalmente de 110 voltios.

Este proceso lo realiza el transformador por medio de dos bobinas. La bobina del primario, que está conectada a la entrada y que tiene muchas vueltas y la bobina de salida o secundario, que tiene pocas vueltas.

El tamaño del transformador depende de la potencia que se desea rebajar. Esta potencia se calcula conociendo el voltaje y la corriente del circuito del radio.

La fórmula para calcular la potencia es:

$$\boxed{P = V \times I}$$

Donde:

P = Potencia en vatios
V = Voltaje en voltios
I = Intensidad en amperios

Por ejemplo, el radio AM se alimenta con 9V y consume 200 mA de corriente aproximadamente.

Por lo tanto la potencia que consume es :

$$P = V \times I = 9\ V \times 0.2\ A = 1.8 \text{ vatios}$$

Considerando que en el transformador y en el convertidor se producen pérdidas cercanas al 10%, el transformador debe tener un tamaño apropiado para suministrar 2 vatios al circuito.

Como vimos en la Lección 9, este tipo de transformador recibe el nombre de *transformador reductor* ya que en su salida el voltaje es menor que en la entrada.

Como veremos más adelante, según el tipo de *rectificación* o sistema de convertir la CA en CC, el secundario del transformador puede tener dos o tres terminales. Algunos transformadores utilizados para radios y grabadoras tienen un primario con tres terminales con el fin de que el aparato se pueda conectar a voltajes de entrada de 110 ó 220 voltios.

En la figura 458 tenemos el aspecto físico y el símbolo de estos transformadores que se utilizan para las fuentes de poder de los radios.

Para armar nuestras propias fuentes de poder podemos conseguir los transformadores con los voltajes y corrientes apropiados, en los almacenes especializados en la venta de componentes electrónicos o en las compañías dedicadas a su fabricación.

Transformadores para radios

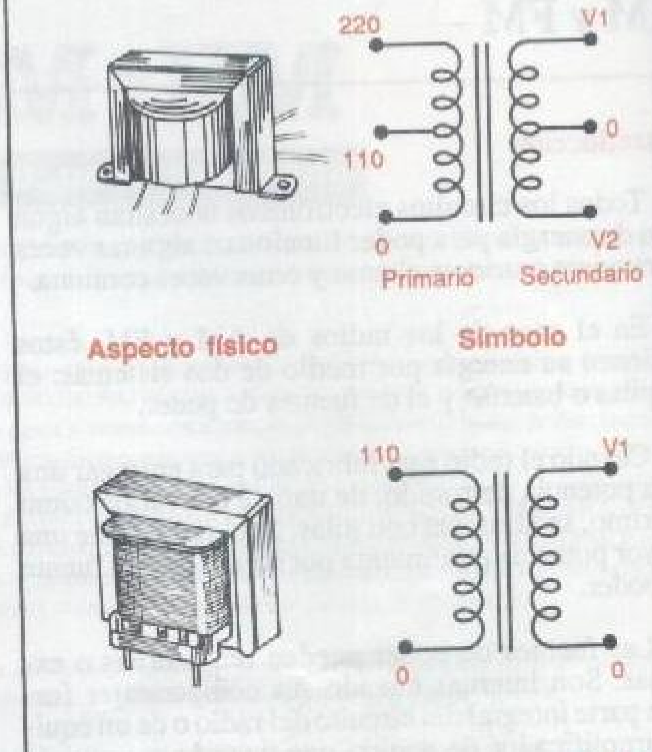

Fig. 458

Los circuitos rectificadores

Según el diagrama de bloques de la figura 457 el segundo circuito de una fuente de poder se llama circuito *rectificador*.

El circuito rectificador recibe del transformador un voltaje de corriente alterna rebajado y elimina la parte negativa de esta señal. A este proceso se le conoce como *rectificación* y tiene dos variaciones; rectificación de media onda y rectificación de onda completa. En los dos casos la señal resultante recibe el nombre de *corriente continua pulsante* o CCP.

El circuito rectificador de media onda está formado por un diodo. Este diodo, como solamente conduce cuando el voltaje que le llega al ánodo es mayor de cero voltios, o sea positivo, elimina en la salida los semiciclos negativos de la señal de corriente alterna.

En la figura 459 tenemos un circuito de este tipo en donde se muestran las señales de entrada y de salida. Este rectificador es ineficiente ya que los pulsos tienen mucha distancia entre sí y es poco utilizado en las fuentes de poder para radios.

Para suplir esta deficiencia se utilizan circuitos *rectificadores de onda completa* que pueden ser de

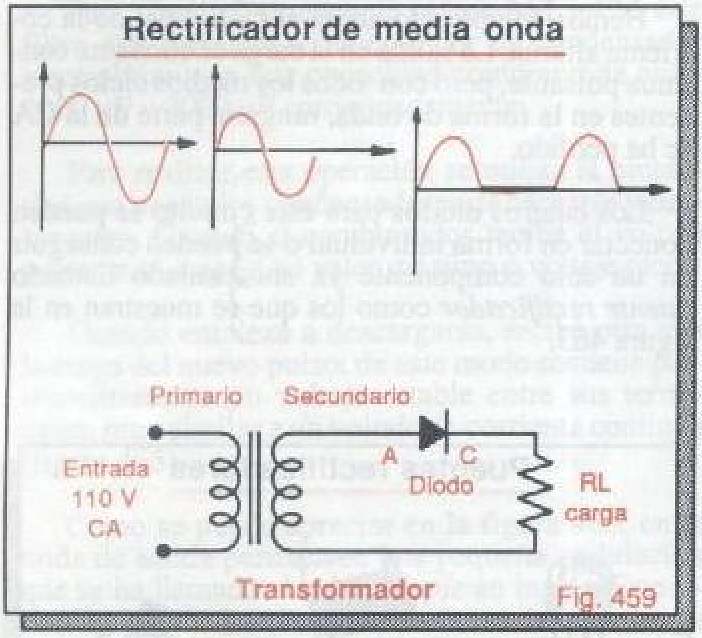
Rectificador de media onda
Primario
Secundario
A
C
Entrada
110 V
CA
Diodo
RL
carga
Transformador

Fig. 459

dos tipos: con dos diodos o con puente rectificador de cuatro diodos.

En la figura 460 tenemos el circuito rectificador de onda completa con dos diodos. En este caso el secundario del transformador tiene tres terminales. El terminal del centro recibe el nombre de *tap central* y por él se obtiene el polo negativo del voltaje rectificado.

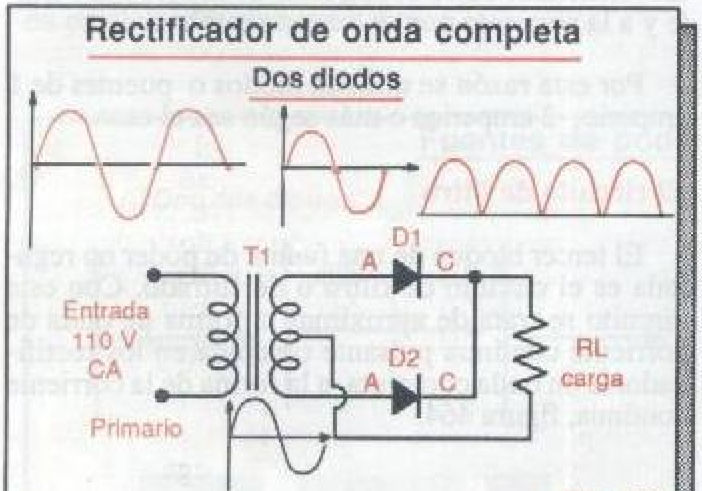
Rectificador de onda completa
Dos diodos
D1
T1
A
C
Entrada
110 V
CA
D2
A
C
RL
carga
Primario

Fig. 460

Este circuito funciona de la siguiente manera: en cada extremo del secundario del transformador tenemos un voltaje de CA pero con distinta fase debido a la configuración de las bobinas. Estos puntos están marcados como A y B en el diagrama.

Si tomamos las señales simultáneamente, el diodo D1 empezará a conducir en el primer semiciclo y D2 estará bloqueado y no dejará pasar la señal.

En el segundo semiciclo, conduce el diodo D2 y se bloquea el diodo D1. Así, cada diodo deja pasar o conduce cada medio ciclo del voltaje de corriente alterna entregando en la salida un voltaje de corriente continua pulsante con todos los ciclos de valor positivo adyacentes.

De esta forma nos vamos acercando a la conversión de corriente alterna en corriente continua.

El otro método de rectificación es el que se muestra en la figura 461 que utiliza cuatro diodos conectados en una configuración especial llamada *puente rectificador*. Con este sistema el secundario del transformador solamente necesita dos terminales y su tamaño puede ser más pequeño que en el caso anterior.

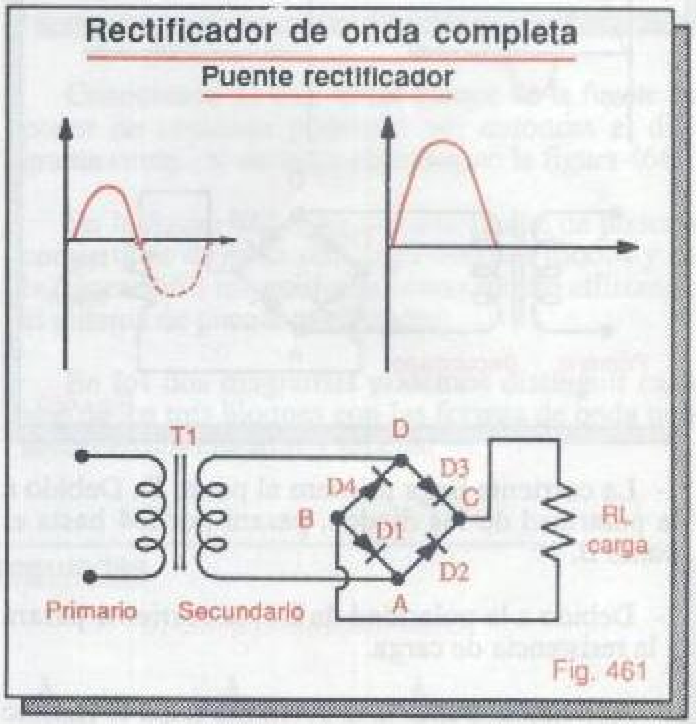
Rectificador de onda completa
Puente rectificador
T1
D
D3
D4
C
B
D1
RL
carga
D2
A
Primario
Secundario

Fig. 461

Este circuito trabaja de la siguiente manera: durante el primer medio ciclo positivo, el terminal superior del transformador es positivo y el terminal inferior es negativo. La secuencia del flujo de corriente será así:

1- La corriente llega al punto A y pasa entonces a D1; ésta no puede pasar por D2, debido a su polarización.

2- Cuando la corriente llega al punto B, ésta pasa a la resistencia de carga. No puede pasar por D4, debido a su polarización.

3- Cuando la corriente llega al punto C, encuentra a D3 y D2, entonces pasa por D3, ya que por D2 no puede pasar debido a su polarización.

4- Cuando la corriente llega al punto D y encuentra el terminal superior del transformador, puede pasar por D4 por su polaridad, entonces pasa por el transformador.

Ahora veamos qué pasa en el medio ciclo negativo: durante el primer medio ciclo negativo, el terminal superior del transformador es negativo y el terminal inferior es positivo. La corriente tendrá entonces la secuencia o camino como la que se muestra la figura 462.

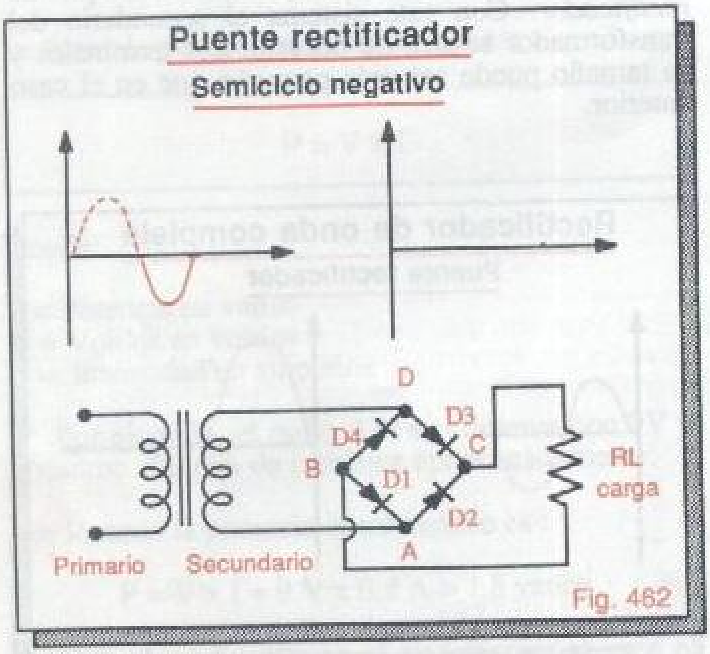
Puente rectificador

Semiciclo negativo

Fig. 462

1- La corriente llega primero al punto D. Debido a la polaridad de los diodos, pasará por D4 hasta el punto B.

2- Debido a la polaridad de D1, la corriente pasará a la resistencia de carga.

3- Al llegar al punto C la corriente toma el camino de D2, ya que su ánodo es positivo.

4- Finalmente en el punto A la corriente encontrará el terminal positivo del transformador.

Lo anterior se puede reducir de la siguiente manera:

En el medio ciclo positivo conducen (dejan pasar la corriente), los diodos D1 y D3. En el medio ciclo negativo conducen los diodos D2 y D4.

En el punto C tenemos deficiencia de cargas negativas. En el punto B, tenemos exceso de cargas negativas.

Entonces siempre tendremos una corriente que va desde B a C, o lo que en otras palabras significa: C está a mayor potencial que B.

Hemos completado ahora el ciclo total de la corriente alterna. La salida en la carga es corriente continua pulsante, pero con todos los medios ciclos presentes en la forma de onda, ninguna parte de la CA se ha perdido.

Los cuatros diodos para este circuito se pueden conectar en forma individual o se pueden conseguir en un sólo componente ya encapsulado llamado *puente rectificador* como los que se muestran en la figura 463.

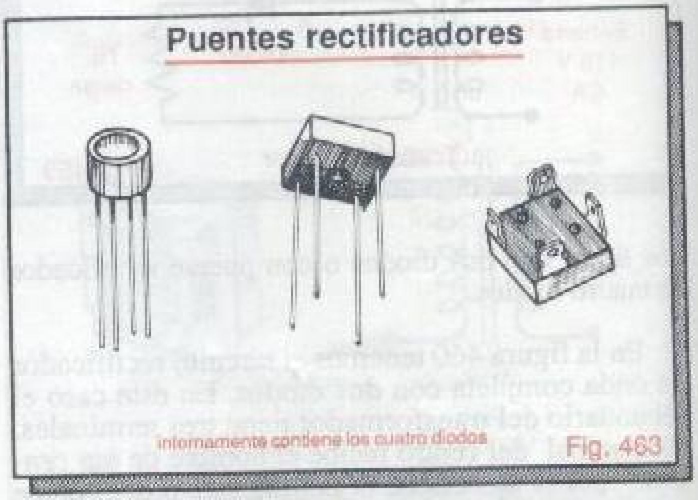
Puentes rectificadores

internamente contiene los cuatro diodos

Fig. 463

Para los casos de rectificadores de onda completa con dos diodos, con cuatro diodos o con puente rectificador éstos deben estar de acuerdo al voltaje y a la corriente que se desean rectificar.

Por esta razón se utilizan diodos o puentes de 1 amperio, 2 amperios o más según sea el caso.

El circuito de filtro

El tercer bloque de una fuente de poder no regulada es el circuito de filtro o de filtrado. Con este circuito se trata de aproximar la forma de onda de corriente continua pulsante obtenida en los rectificadores de onda completa, a la forma de la corriente continua, figura 464.

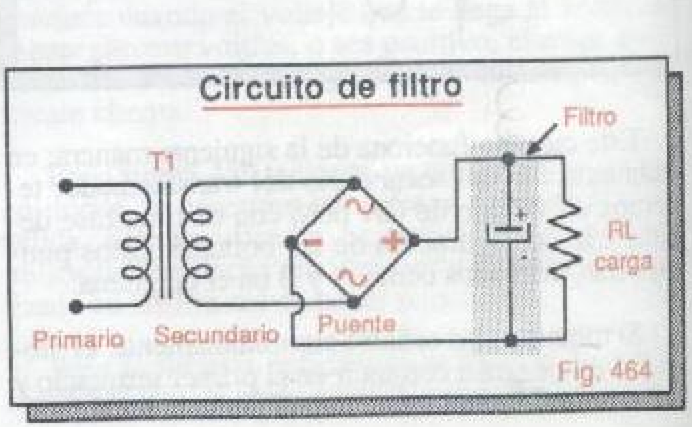
Circuito de filtro

Fig. 464

Para circuitos simples de radios AM o FM este filtro está formado solamente por un condensador electrolítico con una capacidad comprendida entre 1000 μF y 3000 μF aproximadamente.

Para realizar esta operación se utiliza la propiedad que tienen los condensadores de cargarse y descargarse. Cuando el condensador recibe el voltaje pulsante se carga a su valor máximo o voltaje pico.

Cuando empieza a descargarse, recibe otra vez la carga del nuevo pulso; de este modo sostiene permanentemente un voltaje estable entre sus terminales, muy similar a un voltaje de corriente continua (figura 465).

Como se puede apreciar en la figura 465, en la onda de salida permanece una pequeña ondulación que se ha llamado *rizado* o *ripple* en inglés. Este rizado es mayor cuando aumenta la corriente de carga y depende del valor en microfaradios del condensador de filtro, ya que según este factor el condensador se descarga más o menos rápidamente.

Este fenómeno se puede apreciar mejor en la representación de la figura 465 y se puede escuchar en un radio como un ruido de fondo permanente más conocido en el argot técnico como *hum* o *hom*.

Para un determinado circuito se puede rebajar este ruido considerablemente, aumentando el valor en microfaradios del condensador de filtro simplemente conectando otro igual o de más microfaradios en paralelo con el ya existente.

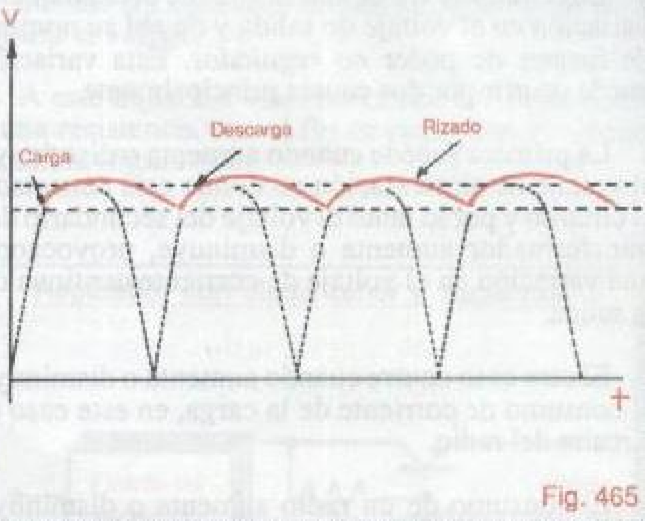
Señal en el circuito de filtro

Fig. 465

Conociendo ya este tercer bloque de la fuente de poder no regulada podemos ver entonces el diagrama completo de estos circuitos en la figura 466.

En la figura 466A tenemos la fuente de poder o convertidor de onda completa con dos diodos y en la figura 466B tenemos el mismo circuito utilizando el sistema de puente rectificador.

En los dos diagramas podemos distinguir cada uno de los tres bloques con las formas de onda presentes en sus entradas y salidas.

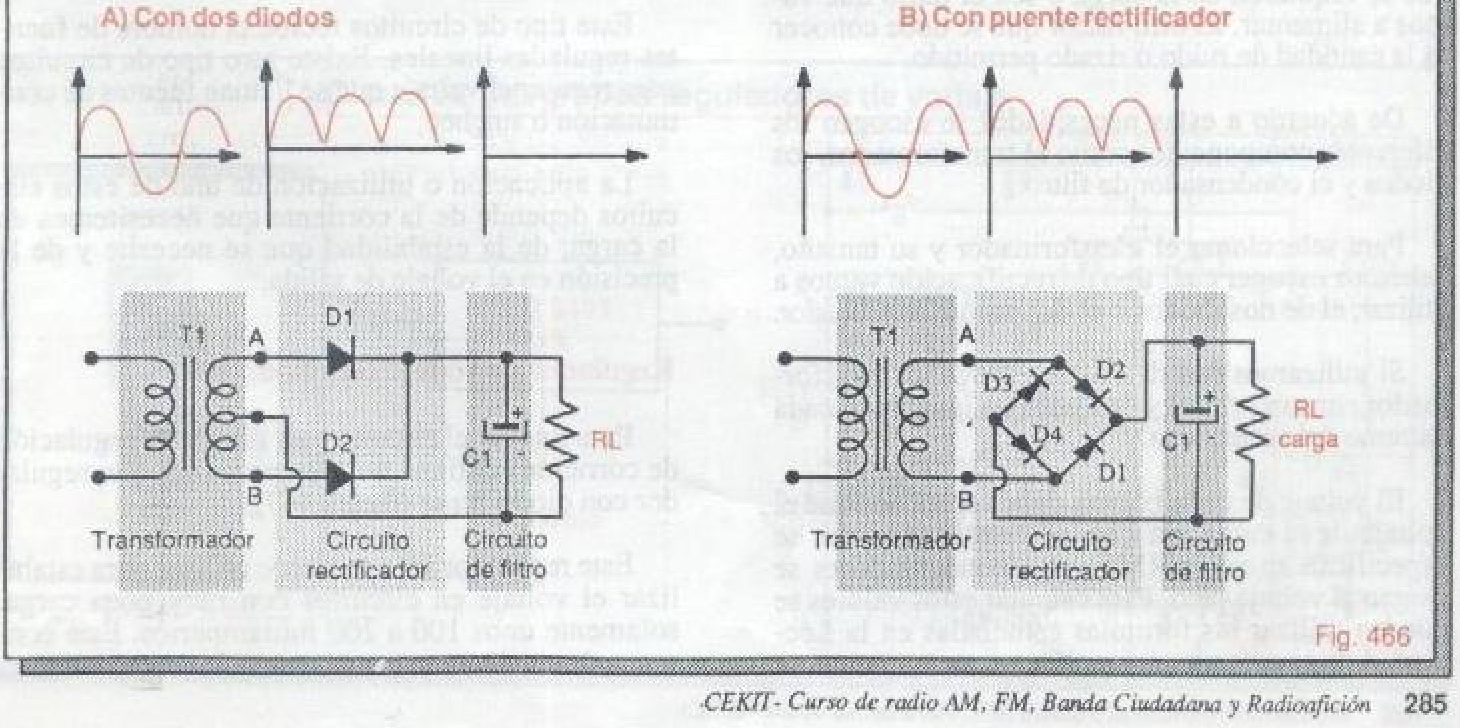
Fuentes de poder no reguladas

A) Con dos diodos

B) Con puente rectificador

Fig. 466

Variación del voltaje de salida en la fuentes de poder no reguladas

Los circuitos vistos anteriormente presentan una variación en el voltaje de salida y de ahí su nombre de fuentes de poder *no reguladas*. Esta variación puede ocurrir por dos causas principalmente.

La primera sucede cuando aumenta o disminuye el voltaje de alimentación en donde está conectado el circuito y por lo tanto el voltaje del secundario del transformador aumenta o disminuye, provocando una variación en el voltaje de corriente continua de la salida.

El otro caso ocurre cuando aumenta o disminuye el consumo de corriente de la carga, en este caso el circuito del radio.

El consumo de un radio aumenta o disminuye proporcionalmente al volumen ajustado en el control de volumen del amplificador de audio.

Cuando hay más volumen, hay más consumo de corriente y viceversa. Esto hace que el voltaje del transformador baje o suba de acuerdo con este consumo de corriente.

En algunos radios esta variación es permisible y en otros no, por lo que se hace necesaria la implementación de circuitos reguladores de voltaje.

Cálculos de los valores de los componentes

Para construir una fuente no regulada debemos tener en cuenta varios factores. Primero que todo debemos conocer cuál es el voltaje y la corriente que se requieren en la carga o sea el radio que vamos a alimentar. El otro factor que se debe conocer es la cantidad de ruido o rizado permitido.

De acuerdo a estas necesidades se escogen los diferentes componentes como el transformador, los diodos y el condensador de filtro.

Para seleccionar el transformador y su tamaño, debemos escoger cuál tipo de rectificación vamos a utilizar, el de dos diodos o el del puente rectificador.

Si utilizamos dos diodos se escoge un transformador con tap central y dos voltajes iguales en cada extremo del secundario.

El voltaje de cada bobina debe ser menor que el voltaje de la carga, ya que los transformadores se especifican en valor RMS y los condensadores se cargan al voltaje pico. Para calcular estos valores se pueden utilizar las fórmulas estudiadas en la Lección 4 de este curso.

Fuentes de poder reguladas

Los circuitos de radio donde se utilizan circuitos integrados y transistores necesitan un voltaje estable y preciso para el cual fueron diseñados. Si el voltaje es menor no trabajan y si es mayor, se pueden quemar fácilmente.

Existen además otros componentes que necesitan voltajes muy estables como son los microprocesadores y las memorias que se están utilizando en todos los equipos de radio modernos.

Otro aspecto que se modifica considerablemente con la variación del voltaje de alimentación es la potencia de salida de la señal de radiofrecuencia en los equipos. Así, encontramos que uno o dos voltios menos se transforman en porcentajes de 20 ó 30% menos de potencia de transmisión, en el caso de los radiotransmisores.

En el caso contrario, si el voltaje de alimentación es muy alto, los circuitos o etapas finales quedan sobrecargados y se pueden quemar fácilmente.

Circuitos reguladores de voltaje

Para evitar las variaciones de voltaje se debe agregar a las fuentes no reguladas un circuito adicional que regule el voltaje de salida de CC, aunque haya variaciones de voltaje en la entrada de CA o varíe la carga conectada a la fuente de poder.

Estos circuitos reguladores pueden ser muy simples, utilizando diodos zener o un poco más complejos: con transistores, diodos zener y resistencias o con circuitos integrados llamados reguladores de voltaje de tres terminales.

Este tipo de circuitos recibe el nombre de fuentes reguladas lineales. Existe otro tipo de circuitos para regular el voltaje que se llaman fuentes de conmutación o suicheo.

La aplicación o utilización de uno de estos circuitos depende de la corriente que necesitemos en la carga, de la estabilidad que se necesite y de la precisión en el voltaje de salida.

Regulador de voltaje con diodo zener

Para ilustrar el circuito más simple de regulación de corriente continua miraremos un circuito regulador con diodo zener (figura 467).

Este regulador sólo se puede utilizar para estabilizar el voltaje en circuitos con muy poca carga, solamente unos 100 a 200 miliamperios. Este consumo de corriente se encuentra en los receptores; en

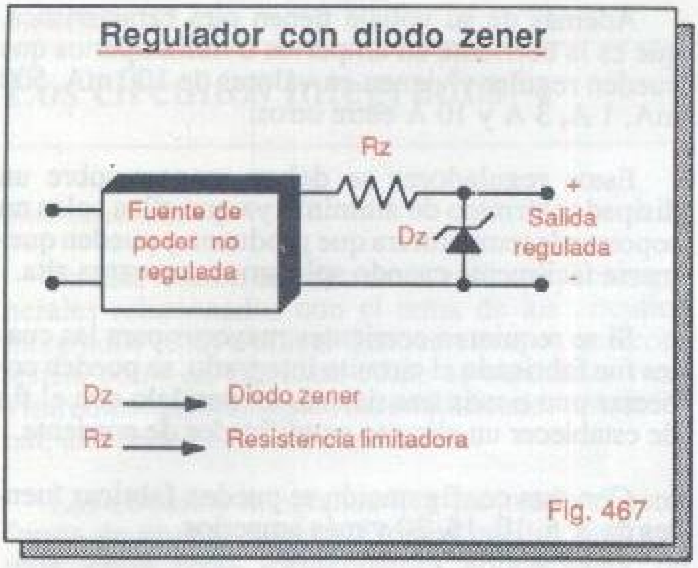

Regulador con diodo zener

Fig. 467

los equipos transmisores de radio de banda ciudadana y radioaficionado tenemos cargas de 3 amperios en adelante.

Note que debe existir siempre una resistencia en serie con el diodo zener. El objetivo de esta resistencia es limitar la corriente que circula por él, con el fin de evitar su destrucción cuando aumenta el voltaje de entrada no regulado.

El valor en ohmios de esta resistencia depende del voltaje de entrada no regulado y del voltaje del diodo zener.Los diodos zener se fabrican para un voltaje determinado, como por ejemplo 5 voltios, 6 voltios, 9 voltios, etc.

Para cargas mayores se recomiendan los circuitos con uno o varios transistores y diodo zener o los circuitos integrados reguladores de voltaje.

Regulador de voltaje con transistor y diodo zener

Cuando la corriente de la carga es mayor, se debe utilizar un transistor en serie con la carga para regular el voltaje.

A este transistor van conectados un diodo zener y una resistencia con el fin de establecer el voltaje de salida (figura 468).

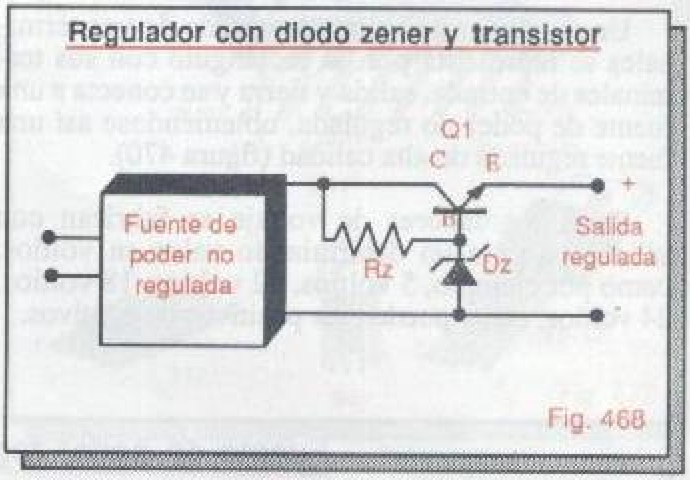

Regulador con diodo zener y transistor

Fig. 468

El transistor se comporta como una resistencia variable. Cuando el voltaje de entrada es mayor, hay mas caída de voltaje entre los terminales emisor y colector del transistor. Igualmente, si varía la carga, varía la caída de voltaje en serie y la salida trata de permanecer estable.

Este tipo de circuito puede manejar corrientes más grandes que el circuito con zener y también soporta mayores variaciones del voltaje de entrada.

Circuitos integrados reguladores de voltaje

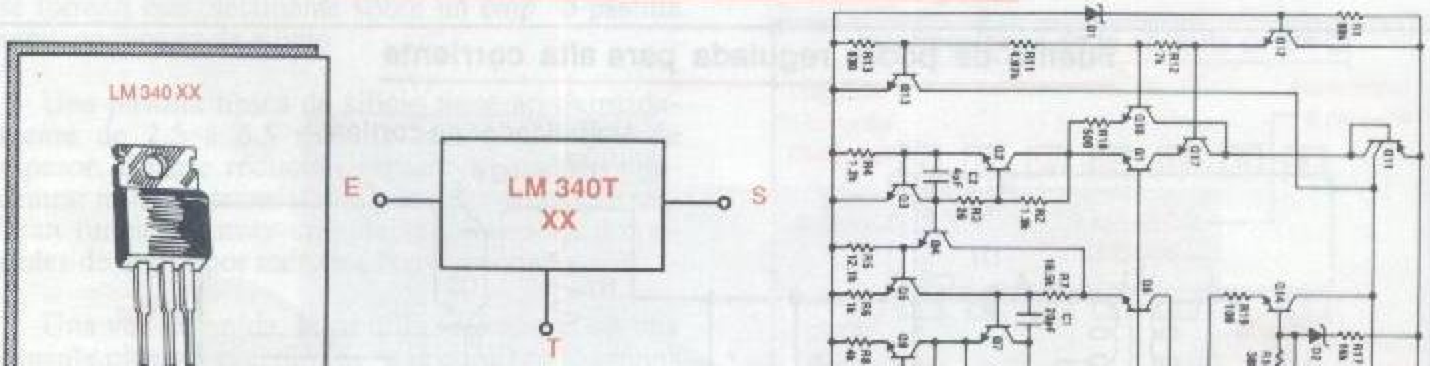

Aspecto físico

Símbolo

Fig. 469

Circuitos integrados reguladores de voltaje de tres terminales

El método más eficiente y fácil de usar actualmente es el de los circuitos integrados reguladores de voltaje de tres terminales.

En estos circuitos integrados se han agrupado una gran cantidad de componentes como transistores, diodos y resistencias para formar un circuito regulador de voltaje muy completo (figura 469).

Un regulador de voltaje integrado de tres terminales se representa por un rectángulo con sus terminales de entrada, salida y tierra y se conecta a una fuente de poder no regulada, obteniéndose así una fuente regulada de alta calidad (figura 470).

Estos reguladores de voltaje se fabrican con salida fija para un determinado valor en voltios, como por ejemplo, 5 voltios, 12 voltios, 18 voltios, 24 voltios, etc, y pueden ser positivos o negativos.

Además de su voltaje tienen otra característica, que es la corriente en amperios o miliamperios que pueden regular y vienen en valores de 100 mA, 500 mA, 1 A, 3 A y 10 A entre otros.

Estos reguladores se deben montar sobre un disipador térmico de aluminio ya que ellos solos no soportan la temperatura que producen y pueden quemarse fácilmente cuando aplicamos una carga alta.

Si se requieren corrientes mayores para las cuales fue fabricado el circuito integrado, se pueden conectar uno o más transistores en paralelo con el fin de establecer un circuito amplificador de corriente.

Con esta configuración se pueden fabricar fuentes de 3, 5, 10, 15, 20 y más amperios.

Este es el caso de las fuentes reguladas que se utilizan para los equipos de radioaficionado (figura 471).

Fuente de poder regulada con circuito integrado

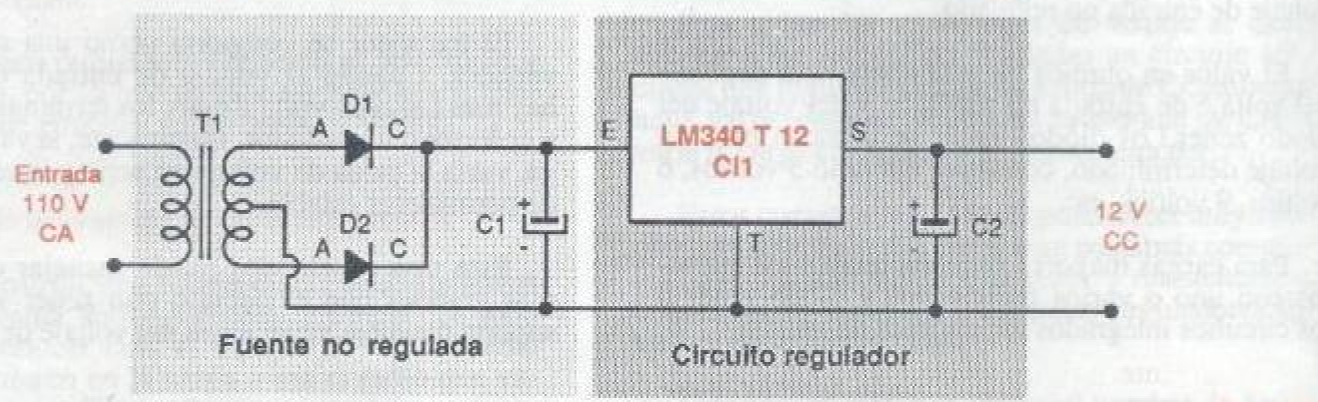
T1
D1
A
C
Entrada
110 V
CA
D2
A
C
C1
+
-
E
LM340 T 12
CI1
S
T
+
C2
12 V
CC
Fuente no regulada
Circuito regulador

Fig. 470

Fuente de poder regulada para alta corriente

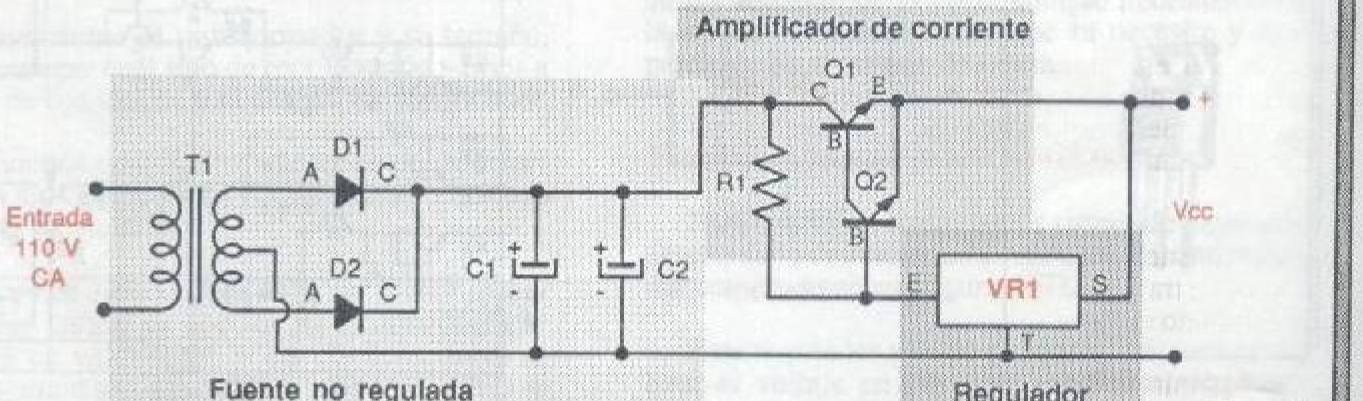
Amplificador de corriente
Q1
C
E
B
R1
Q2
B
E
T1
D1
A
C
Entrada
110 V
CA
D2
A
C
C1
+
-
+
-
C2
E
VR1
S
T
+
Vcc
Fuente no regulada
Regulador

Fig. 471

Lección 37

Los circuitos integrados

Introducción

En esta lección estudiaremos varios aspectos generales relacionados con el tema de los circuitos integrados (CIs) o chips. Conoceremos en qué consisten, cómo se fabrican, cómo se clasifican y revisaremos la teoría básica del amplificador operacional, uno de los CIs más importantes.

Los circuitos integrados han mejorado nuestra forma de vivir y trabajar. Gracias a ellos disfrutamos, entre otras sorprendentes aplicaciones, de computadores personales, relojes digitales, calculadoras de bolsillo, electrodomésticos y equipos médicos, industriales y de oficina altamente sofisticados.

En el campo de las comunicaciones y la radio los circuitos integrados han hecho posible, por ejemplo, la transmisión de información de todo tipo vía satélite, la grabación y reproducción de imágenes, el desarrollo de sofisticados equipos de control y telemetría y otros sucesos.

Por esta razón es imprescindible para toda persona relacionada con el mundo de la electrónica, ya sea a nivel técnico, profesional o aficionado, conocer la existencia de los circuitos integrados y aprender a utilizarlos racionalmente. Es un conocimiento del que no podemos sustraernos.

Qué es un circuito integrado

Básicamente, un circuito integrado es un circuito electrónico completo en el cual todos los componentes, incluyendo transistores, diodos, resistencias, condensadores y alambres de interconexión, se forman completamente sobre un *chip* o pastilla semiconductora de silicio.

Una pastilla típica de silicio tiene aproximadamente de 2.5 a 6.5 mm de lado y 0.5 mm de espesor. En ese reducido espacio se pueden concentrar miles de transistores y componentes que realizan funciones muy complejas y variadas, imposibles de lograr por métodos convencionales.

Una vez obtenida, la pastilla se encierra en una cápsula plástica o cerámica que contiene los pines de acceso a través de los cuales el chip se comunica con el mundo exterior.

En la figura 472 se muestra el aspecto de un circuito integrado con encapsulado DIP o de doble fila y otras presentaciones comunes.

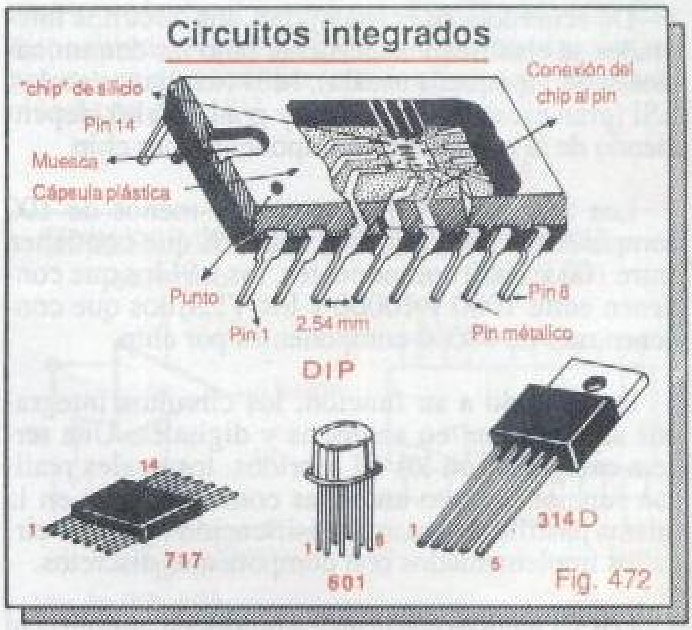
Circuitos integrados
"chip" de silicio
Pin 14
Muesca
Cápsula plástica
Conexión del chip al pin
Punto
Pin 1
2,54 mm
Pin 8
Pin metálico
DIP
14
1
717
1
8
601
1
5
314 D

Fig. 472

En la presentación DIP, la más popular, el pin 1 se identifica mediante una ranura o un punto grabado en la parte superior de la cápsula. El conteo de los demás pines se realiza en sentido contrario al de las manecillas del reloj. Son comunes circuitos integrados DIP de 8, 14, 16, 24, 40 y 64 pines.

La cápsula contiene, además, información respecto al fabricante, la referencia del dispositivo y la fecha de fabricación (figura 473). El logotipo identifica al fabricante del chip (National, Texas, Fairchild, Motorola, etc.). La referencia o número de parte designa específicamente al dispositivo.

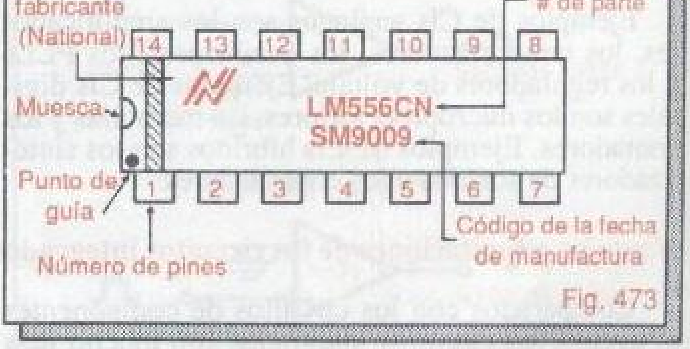
Informaciones de la cápsula de un CI
Logo del fabricante (National)
Referencia o # de parte
14
13
12
11
10
9
8
Muesca
LM556CN
SM9009
Punto de guía
1
2
3
4
5
6
7
Código de la fecha de manufactura
Número de pines

Fig. 473

El código de la fecha informa cuándo fue manufacturado el chip. Las dos primeras cifras indican el año y las dos últimas se refieren al mes o semana de fabricación. Por ejemplo, "8307" significa la séptima semana (mediados de febrero) de 1983.

Clasificación de los circuitos integrados

De acuerdo a su complejidad, los circuitos integrados se clasifican usualmente bajo las denominaciones SSI (pequeña escala), MSI (mediana escala), LSI (gran escala) y VLSI (muy gran escala), dependiendo de la cantidad de componentes por chip.

Los SSI son los que contienen menos de 100 componentes por chip, los MSI los que contienen entre 100 y 1000 componentes, los LSI los que contienen entre 1000 y 10000 y los VLSI los que contienen más de 10000 componentes por chip.

De acuerdo a su función, los circuitos integrados se clasifican en análogos y digitales. Una tercera categoría son los CI híbridos, los cuales realizan funciones tanto análogas como digitales en la misma pastilla. La misma clasificación rige para circuitos implementados con componentes discretos.

Los CI análogos o lineales producen, amplifican o responden a señales variables de voltaje (figura 474A). Los digitales o lógicos producen, procesan o responden a señales que tienen únicamente dos niveles de voltaje, denominados estados **alto** y **bajo** o **1** (uno) y **0** (cero), respectivamente (figura 474B).

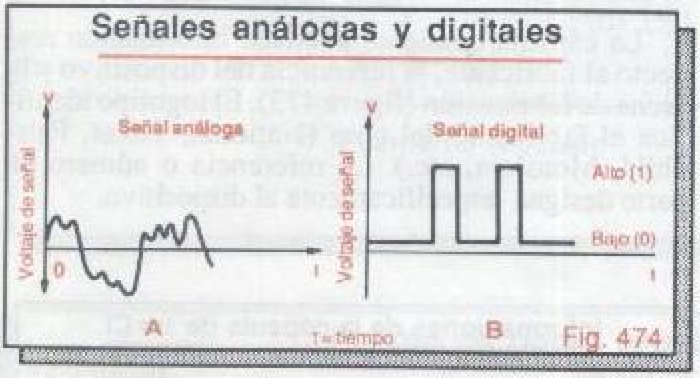

Fig. 474

Ejemplos de CIs análogos son los amplificadores, los temporizadores, los osciladores, los PLLs y los reguladores de voltaje. Ejemplos de CIs digitales son los microprocesadores, las memorias y los contadores. Ejemplos de CIs híbridos son los sintetizadores de sonidos y los controladores.

Ventajas y limitaciones de los circuitos integrados

Comparados con los circuitos de componentes discretos, los circuitos integrados son mucho más pequeños, livianos, económicos y confiables. Sin embargo, su reducido tamaño limita los niveles de voltaje y potencia que pueden manejar.

Además, no se pueden integrar bobinas ni transformadores. Así mismo, en el estado actual de la tecnología, la integración de resistencias y condensadores de gran valor no es eficiente porque estos componentes requieren de mucho espacio.

Cómo se fabrican los circuitos integrados

Prácticamente todos los circuitos integrados (CIs) disponibles en la actualidad se fabrican a partir de pastillas de *silicio* (Si), aunque están emergiendo otras tecnologías como la basada en el *arseniuro de galio* (GaAs).

El procesamiento del silicio para obtener CIs o chips es relativamente complicado pero intentaremos describirlo de una forma sencilla.

Para fabricar un chip, las pastillas de silicio se procesan primero para hacer transistores. Una pastilla de silicio por sí misma es aislante y no conduce corriente. Los transistores se crean agregando impurezas como fósforo o arsénico a determinadas regiones de la pastilla.

Las conexiones se realizan a través de líneas o trazos metálicos muy finos, generalmente de oro.

El proceso de agregado de impurezas se denomina *dopado*. Los transistores y los trazos metálicos de conexión se denominan *rasgos* .

Cada rasgo se forma sobre la pastilla rociando en las regiones seleccionadas un químico protector sensible a la luz llamado *photoresist*. El photoresist forma una película muy delgada sobre la superficie de la pastilla.

La pastilla es entonces bombardeada con luz ultravioleta mediante un proyector deslizante muy preciso llamado *alineador óptico*.

El alineador posee un dispositivo muy pequeño llamado *máscara*, que evita que la luz incida sobre ciertos puntos específicos de la pastilla. Cuando la luz alcanza un área determinada de la pastilla, elimina el photoresist presente en esa zona.

El químico (fósforo, arsénico o metal) se deposita en las regiones descubiertas por la luz e ignora las encubiertas por la máscara. Estas últimas zonas aún permanecen recubiertas de photoresist.

La precisión del alineador óptico determina qué tan fino puede hacerse un rasgo.

A comienzos de los años 70 era difícil hacer transistores de menos de 10 micras de tamaño. Hoy día, los transistores alcanzan tamaños inferiores a una micra. Esto permite una alta densidad y mejora la velocidad de respuesta de los dispositivos. Una micra es la millonésima parte de un metro.

A continuación, la pastilla se calienta a altas temperaturas. Esto origina que el silicio no procesado de la superficie se convierta en óxido de silicio (SiO_2). El SiO_2 se esparce sobre la superficie de la pastilla y forma sobre la misma una delgada película aislante de unas pocas micras de espesor.

De este modo se obtiene el primer nivel de *metalización* del chip. Para obtener una nueva capa de metalización, el SiO_2 se trata nuevamente con photoresist y se expone al alineador óptico, repitiéndose el mismo procedimiento seguido con el silicio del primer nivel.

Las diferentes capas van creciendo una sobre otra formando una estructura parecida a un sandwich, con el SiO_2 como el pan y el metal o el silicio dopado como la salchicha (figura 475). Por cuestiones prácticas, la mayoría de CI's no se hacen con más de tres capas de metalización.

Chip con dos niveles de metalización

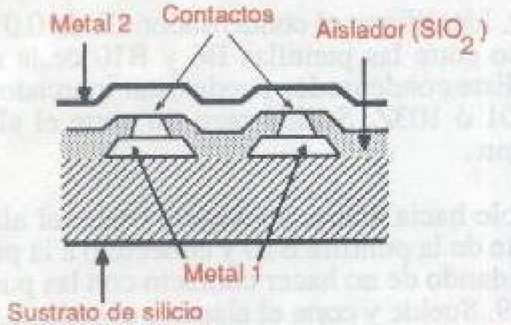

Fig. 475

El amplificador operacional

Un amplificador operacional (AO) es un circuito integrado análogo extremadamente versátil, utilizado para la amplificación de voltajes CA y CC, la generación de formas de onda, la producción de intervalos de tiempo, la comparación de voltajes y una gama muy variada de aplicaciones.

Los primeros amplificadores operacionales fueron desarrollados originalmente para realizar operaciones aritméticas y resolver electrónicamente ecuaciones matemáticas. Originaron los llamados computadores análogos, precursores de los computadores digitales de nuestros días.

Los amplificadores operacionales se caracterizan por su alta ganancia de voltaje, su alta impedancia de entrada y su baja impedancia de salida. La ganancia de voltaje es del orden de 10^5, la impedancia de entrada del orden de 10^6 a 10^{12} Ω y la de salida del orden de 100Ω.

En la figura 476 se muestra el símbolo utilizado para representar un amplificador operacional y la distribución de pines del popular CI LM1458. Este dispositivo posee 2 AOs en una misma cápsula de 8 pines y puede trabajar con tensiones de alimentación hasta de ±18V.

Simbología de amplificadores operacionales

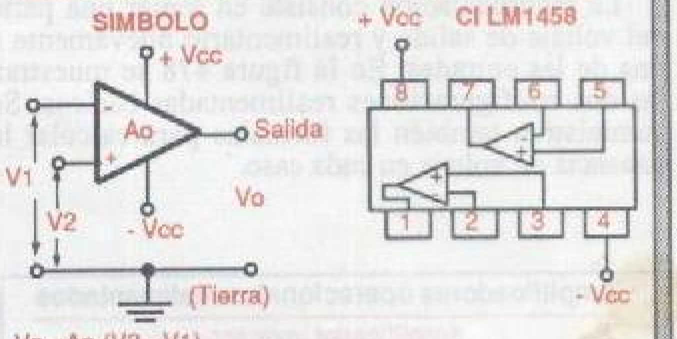

Vo= Ao (V2 - V1)

Ao = Ganancia de lazo abierto

Fig. 476

Un amplificador operacional posee dos entradas y una salida. Las entradas (-) y (+) se denominan inversora y no inversora, respectivamente. El voltaje de salida es proporcional a la **diferencia** de los voltajes aplicados a ambas entradas.

Si se aplica un voltaje a la entrada no inversora, el voltaje de salida tendrá la misma polaridad del voltaje de entrada. Si se aplica un voltaje a la entrada inversora, el voltaje de salida tendrá la polaridad contraria. Estos conceptos se ilustran en la figura 477.

Cómo trabaja el amplificador operacional

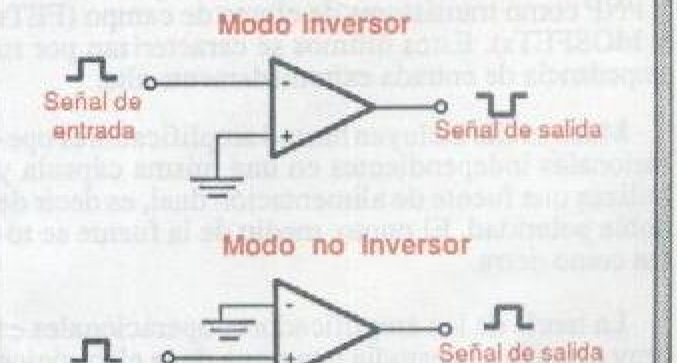

Fig. 477

Los circuitos mostrados en la figura 477 utilizan el amplificador operacional en **lazo abierto**, es decir, sin realimentación. Este modo de operación se emplea principalmente en comparadores de voltaje. Un AO en lazo abierto trabaja con su máxima ganancia de voltaje posible, por ejemplo 80 dB.

La mayoría de circuitos utilizan el amplificador operacional en **lazo cerrado**, es decir realimentado negativamente. El empleo de realimentación negativa permite controlar la ganancia en un amplio rango, proporciona mayor estabilidad, reduce la distorsión e incrementa el ancho de banda.

La realimentación consiste en tomar una parte del voltaje de salida y realimentarlo nuevamente a una de las entradas. En la figura 478 se muestran las dos configuraciones realimentadas básicas. Se suministran también las fórmulas para calcular la ganancia de voltaje en cada caso.

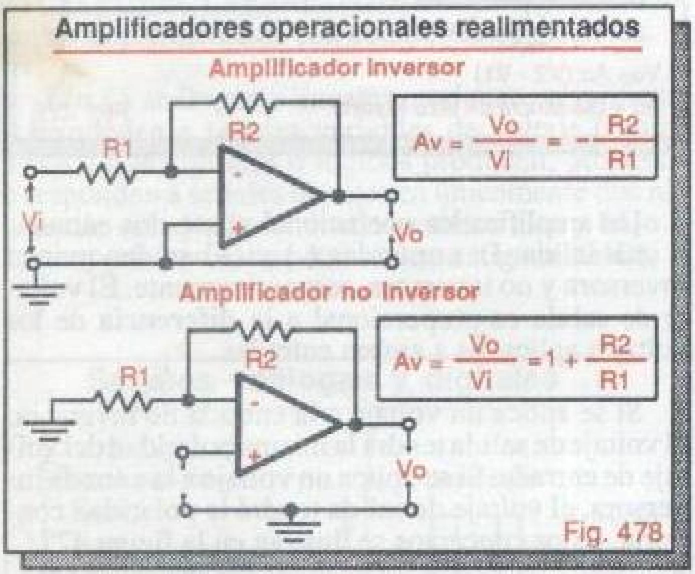
Amplificadores operacionales realimentados
Amplificador inversor
R2
R1
$Av = \frac{Vo}{Vi} = -\frac{R2}{R1}$
Vi
Vo
Amplificador no inversor
R2
R1
$Av = \frac{Vo}{Vi} = 1 + \frac{R2}{R1}$
Vo

Fig. 478

Los amplificadores operacionales integrados se fabrican utilizando tanto transistores bipolares NPN y PNP como transistores de efecto de campo (FETs y MOSFETs). Estos últimos se caracterizan por su impedancia de entrada extremadamente alta.

Muchos CIs incluyen hasta 4 amplificadores operacionales independientes en una misma cápsula y utilizan una fuente de alimentación dual, es decir de doble polaridad. El punto medio de la fuente se toma como tierra.

La teoría de los amplificadores operacionales es muy extensa y la estudia una rama de la electrónica conocida como electrónica lineal o análoga. Por esta razón es imposible cubrirla con detalle en este curso. El lector interesado en conocer más a fondo el tema puede remitirse a la literatura especializada.

Continuación actividad práctica Nº 19

Ensamble del amplificador/conversor de RF. Parte 2.

En esta segunda parte de la Actividad 19 instalaremos en el tablero del radio AM CEKIT otra parte de los componentes que forman el amplificador/conversor de RF, la etapa marcada con color amarillo en la lámina correspondiente al diagrama esquemático del receptor.

Componentes necesarios

2 Condensadores cerámicos de .02µF/50V.C1, C15.
1 Condensador cerámico de .01µF/50V.C2.
1 Transistor NPN C1390 ó 2N2222. Q1
6 centímetros de alambre de conexión.

Procedimiento

Paso 1. Conecte el condensador C1 de 0.02 µF entre las puntillas B1 y 2 y el condensador C15 en paralelo con la resistencia R6 de 100KΩ entre las puntillas 43 y D3 de la etapa amplificadora de RF. Estos condensadores pueden estar marcados como 0.02, .02, 203Z ó 223Z.

Paso 2. Identifique el condensador C2 de 0.01 µF e instálelo entre las puntillas B6 y B10 de la misma etapa. Este condensador puede estar marcado como 0.01, .01 ó 103Z. Suelde pero **no** corte el alambre excedente.

Doble hacia abajo, en ángulo recto, el alambre sobrante de la puntilla B10 y conéctelo a la puntilla B8, cuidando de no hacer contacto con las puntillas B7 y B9. Suelde y corte el alambre excedente.

Doble hacia arriba, en ángulo recto, el alambre sobrante de la puntilla B6 y conéctelo a la puntilla B5. Suelde y corte el alambre excedente.

Paso 3. Identifique la base (B), el colector (C) y el emisor (E) del transistor Q1 C1390 ó 2N2222. Recuerde que el 2N2222 es un reemplazo del C1390 pero tiene una distribución de pines diferente. Para cualquier duda, refiérase a la actividad 16, parte 1.

Separe ligeramente los terminales del transistor Q1 y conéctelos a las puntillas B2, B4 y B5 así: la base (B) a la puntilla B2, el colector (C) a la puntilla B4 y el emisor (E) a la puntilla B5.

Paso 4. Con alambre de conexión conecte las puntillas 47 y B1. Suelde y corte el alambre excedente de la puntilla 47. Doble hacia la derecha, en ángulo recto, el alambre sobrante de la puntilla B1 y conéctelo a la puntilla B3. Suelde y corte el alambre excedente.

Actividad práctica Nº 20

Análisis, prueba y ajuste de la segunda etapa de FI

Introducción

En las actividades prácticas Nº 16 y 18 habíamos ensamblado las etapas de frecuencia intermedia del radio AM. Esta actividad tiene dos partes: primero vamos a conectar el transformador T3 para completar el ensamble de la etapa, y luego vamos a realizar un análisis completo de este circuito.

El estudio comprende un análisis estático, o sea la medida de los voltajes y corrientes de polarización en corriente continua, y un análisis dinámico con el fin de aplicar señales reales a la etapa y lograr la calibración o alineamiento óptimo del transformador T3 para que quede debidamente sintonizado.

Para realizarlo es necesario tener disponible un multímetro análogo o digital que tenga una escala para medir corrientes bajas en miliamperios. Con el segundo análisis ajustamos la etapa para una ganancia o amplificación máxima a la frecuencia intermedia de 455 KHz.

Para esta prueba necesitamos un generador de RF y un osciloscopio. También se puede utilizar el generador de FI de 455 KHz cuya construcción se explicó en la actividad práctica Nº 17.

Si no tiene estos instrumentos, puede conseguirlos prestados en un instituto, colegio, universidad o taller de radio o electrónica. Aprender a utilizarlos es uno de los principales objetivos de este curso.

Si de todas maneras no tiene acceso a ellos, realice solamente el análisis estático con el multímetro y al final daremos las instrucciones para calibrar el radio sin necesidad de estos instrumentos. Este procedimiento es más difícil pero también logra sus resultados.

Instalación del transformador T3

Tome T3 (el segundo transformador de FI), el cual se puede identificar por el núcleo de color negro y suéldelo a las puntillas E11, E12, E13, C1 y C2. Luego conecte la cubierta del transformador T3 al alambre sobrante de la puntilla E10.

Análisis estático No.1 - Consumo de corriente

Este análisis tiene como objetivo medir el consumo total de corriente del amplificador de audio y las dos etapas de FI.

Esta corriente no debe soprepasar el valor de 30 mA. Si esto sucede tendrá que buscar un posible corto en el circuito.

Procedimiento

Paso 1. Coloque un puente de alambre momentáneamente entre las puntillas E3 y E4. Deje conectado este puente durante todo el análisis de la segunda etapa.

Paso 2. Coloque el multímetro en la gama de corriente continua y en una escala igual o un poco mayor a 30 mA.

Paso 3. Conecte la batería al radio intercalando el multímetro en serie entre ésta y el circuito como se muestra en la figura A54.

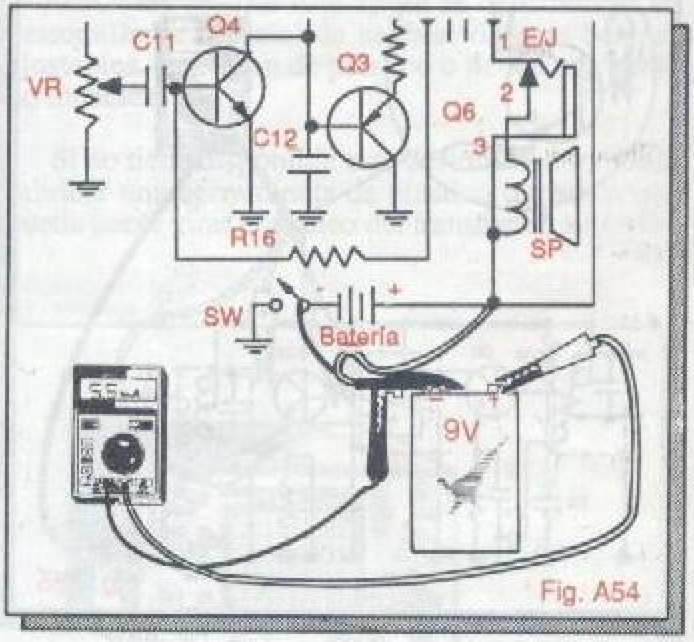

Fig. A54

Paso 4. Encienda el radio y observe la corriente medida por el multímetro; escriba su valor .

IT = ______________ mA (IT = Corriente total)

Esta es la corriente total que consumen las etapas de audio y las de FI.

Análisis estático No. 2

Polarización del transistor Q3

El objetivo de este análisis es medir los voltajes de corriente continua con que están alimentadas las junturas del transistor Q3 y a partir de estas mediciones concluir si están bien polarizadas para que este transistor funcione en su punto óptimo.

Procedimiento

Paso 1. Ajuste el multímetro en una escala un poco mayor a 9 voltios en corriente continua. Conecte el terminal negativo a la tierra del radio. Utilice cualquiera de las puntillas 11,12, 13, 14, 15 o 16 o el alambre que las une.

Paso 2. Una la batería a su conector y encienda el radio.

Paso 3. Vaya tocando con el terminal positivo del multímetro uno a uno los terminales del transistor, el colector (E7), la base (E4) y el emisor (E8). Anote cada medida de voltaje en la tabla 1. Figura A55.

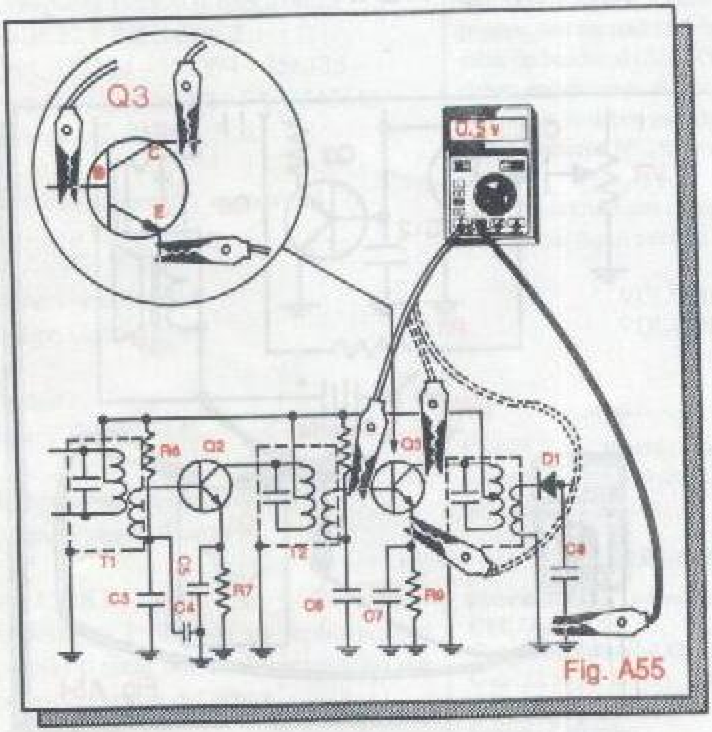

Fig. A55

Transistor	Tipo	VC	VE	VB	VBE= VB-VE	VCB= VC-VB
Q3	NPN					

Tabla 1

Como ya se había dicho anteriormente, un transistor conduce sólo cuando su juntura E-B está directamente polarizada y su juntura C-B está inversamente polarizada.

Como Q3 es un transistor NPN, la juntura B-E está directamente polarizada cuando la base es positiva con respecto al emisor (VBE = positivo) y la juntura C-B estará polarizada cuando el colector es positivo con respecto a la base (VCB = positivo).

Análsis estático No. 3

Midiendo la corriente del colector de Q3

En este análisis mediremos la corriente de colector del transistor Q3 colocando el multímetro en el *modo de corriente y sus terminales en serie con* el colector de Q3.

De esta forma obtenemos un dato real de la corriente circulante por este transistor y podemos concluir si Q3 está trabajando correctamente.

Procedimiento

Paso 1. Coloque el multímetro en una escala para medir corriente en corriente continua un poco mayor a 5 mA.

Paso 2. Mida la corriente de colector de Q3 desconectando el terminal del colector de la puntilla E7 como se muestra en la figura A56, teniendo en cuenta la polaridad de los terminales. El valor medido es

Ic de Q3 = ____________ mA

Si la aguja del multímetro análogo se desplaza hacia la izquierda o el multímetro digital muestra un valor negativo, invierta sus terminales.

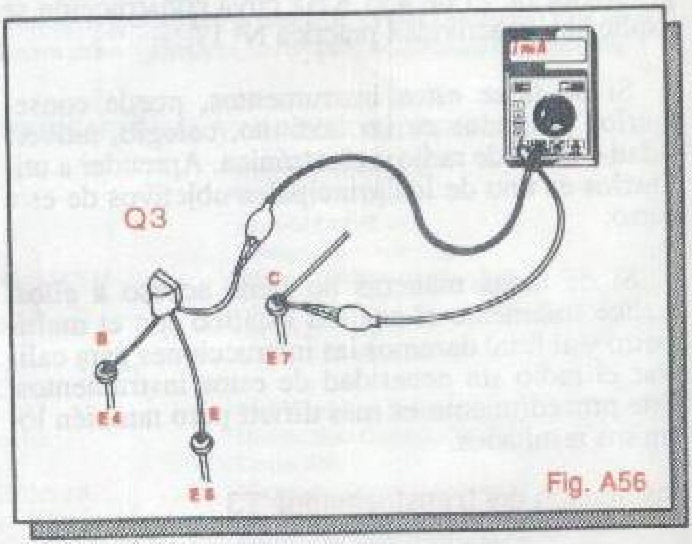

Fig. A56

Si el circuito está correcto, la corriente debe tener un valor entre 3 y 10 mA. Si no hay corriente o está muy alta, se debe revisar el circuito.

Análisis dinámico No. 1

Alineamiento o calibración del transformador T3 y cálculo de la ganancia de voltaje del transistor Q3

Este análisis tiene como objetivo colocar al transformador T3 en su punto óptimo de trabajo, es decir, que produzca una ganancia máxima de esta etapa para una señal de 455 KHz.

Calcularemos también la amplificación de voltaje de esta segunda etapa de FI para dicha señal.

Para realizar este ajuste vamos a inyectar por medio de un generador de RF o con el generador de 455 KHz, una señal de FI sin modular y con una amplitud aproximada de 40 mVpp (milivoltios pico a pico) en la base del transistor Q3.

Procedimiento

Paso 1. Encienda el generador de RF y el osciloscopio.

Paso 2. Conecte el caimán negro (tierra) del generador de RF y del osciloscopio a la tierra del radio (puntillas 1 a 21).

Paso 3. Conecte el terminal del osciloscopio a la base de Q3 (puntilla E4) como se muestra en la figura A57.

Paso 4. Coloque la perilla TIME/DIV del osciloscopio en 2 µs y la perilla VOLT/DIV en 1 mV preferiblemente.

Si su osciloscopio no posee estas escalas, elija otra similar o cercana a estos valores.

Paso 5. Ajuste el generador de RF para tener en su salida una señal sin modulación de 455 KHz y con una amplitud de 40 mVpp.

Cuando esté seguro que tiene la señal de 455 KHz a 40 mVpp en la base de Q3 puede continuar con el paso siguiente.

Paso 6. Conecte la batería en su borne y encienda el radio.

Paso 7. Conecte el otro terminal del osciloscopio en el colector de Q3 (figura A57) y tenga en cuenta que este canal debe estar en su escala mayor, por ejemplo VOLT/DIV en 0.2 VOL/DIV, ya que hay que tener en cuenta la amplificación del transistor.

Paso 8. Para realizar este ajuste se debe utilizar un destornillador especial de calibración para bobinas ajustables. Estos son de plástico o tienen una punta no metálica.

Si no tiene disponible este destornillador, puede fabricar una herramienta de plástico o madera que pueda hacer girar el núcleo del transformador.

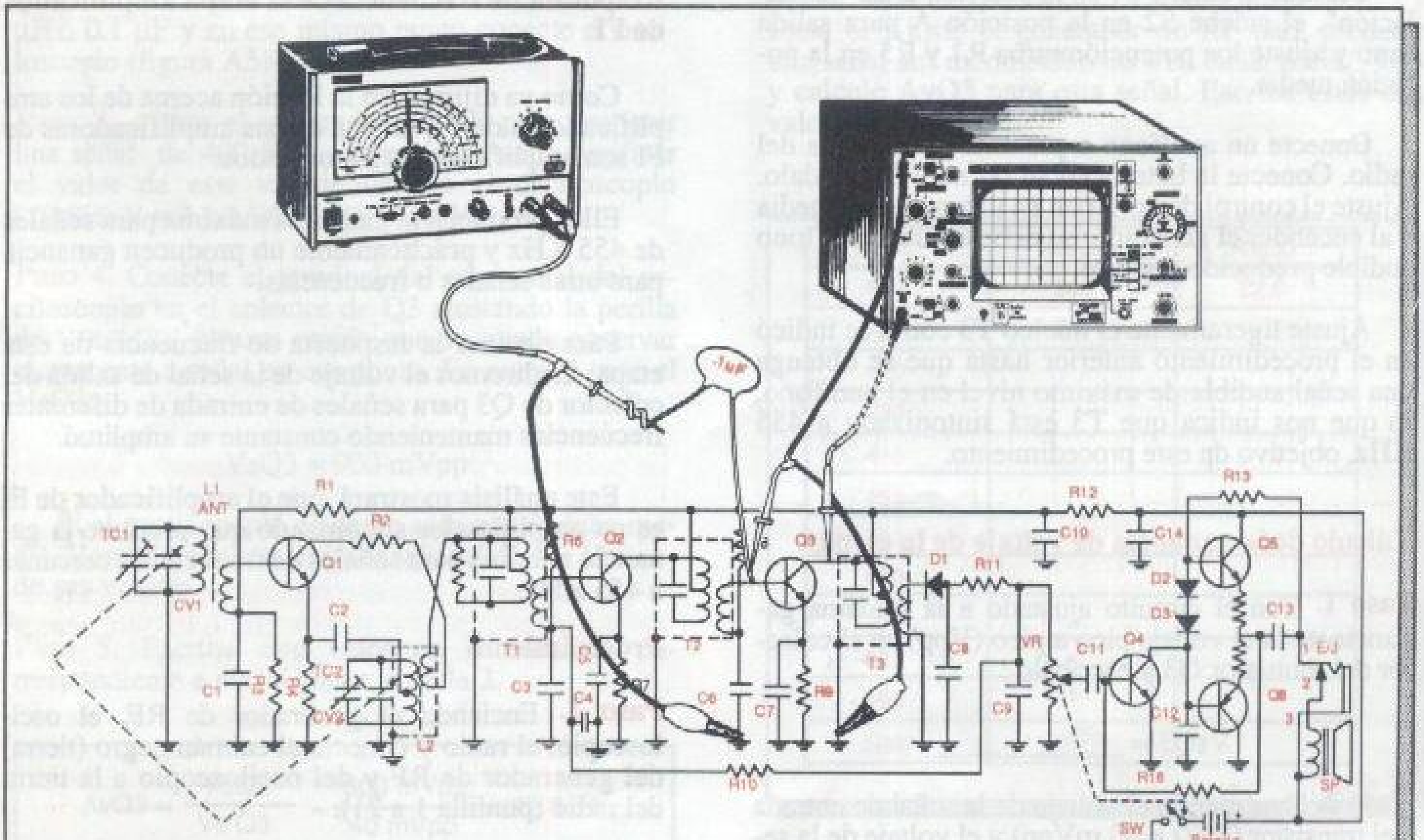

Fig. A57

Con esta herramienta ajuste el núcleo del transformador T3 hasta obtener la máxima amplitud de la señal en la pantalla del osciloscopio.

Gire lentamente el tornillo del núcleo hacia la izquierda o derecha observando en cuál dirección va tomando la señal una mayor amplitud.

Trate con mucho cuidado estos núcleos ya que son muy delicados.

Cuando se obtiene la máxima señal, el osciloscopio nos indica que el transformador está sintonizado en la máxima resonancia a la frecuencia de FI de 455 KHz.

Si usted no tiene acceso a un osciloscopio ni a un generador de RF, puede utilizar el generador de 455 KHz con señal de modulación para ajustar el transformador de FI, T3. Antes de utilizar el generador, éste debe estar calibrado para que entregue una señal con la frecuencia exacta.

Para hacerlo, utilice la salida de onda seno del generador por el borne J4. Conecte la tierra del generador a la tierra del radio y el otro terminal a la base de Q3 por medio de un condensador de cerámica de 0.1 µF en serie.

Coloque el suiche S1 en la posición B de "modulación", el suiche S2 en la posición A para salida seno y ajuste los potenciómetros R1 y R5 en la posición media.

Conecte un audífono o parlante en la salida del radio. Conecte la batería en su borne y enciéndalo. Ajuste el control de volumen en una posición media y al encender el generador se debe escuchar un tono audible producido por éste.

Ajuste ligeramente el núcleo T3 como se indicó en el procedimiento anterior hasta que se obtenga una señal audible de máximo nivel en el audífono, lo que nos indica que T3 está sintonizado a 455 KHz, objetivo de este procedimiento.

Cálculo de la ganancia de voltaje de la etapa

Paso 1. Con el circuito ajustado a la máxima ganancia mida el voltaje pico a pico (Vpp) en el colector del transistor Q3 y escríbalo :

VsQ3 = ____________

Paso 2. Conociendo el voltaje de la señal de entrada del transistor (V_{EQ} = 40 mVpp) y el voltaje de la señal de salida (VsQ3), calcule la ganancia de voltaje realizada por Q3.

$$AvQ3 = \frac{VsQ3}{VE\ Q3} = \frac{\qquad}{40\ mVpp}$$

AvQ3 = Ganancia de voltaje de Q3

Nota: Para calcular AvQ3 y VsQ3 debe utilizarse el mismo submúltiplo de voltaje.

Por ejemplo, ambos voltajes deben estar en milivoltios (1V = 1000 mV). La ganancia de voltaje puede expresarse también en decibeles o decibelios de la siguiente forma:

$$AvdB = 20\ Log\ \frac{VsQ3}{VE\ Q3}$$

Calcule la amplificación de Q3 en dB.

AvQ3 (dB) = 20 Log (   )/(   )

Análisis dinámico No. 2

Respuesta de frecuencia de la etapa amplificadora de FI

Como ya dijimos en la lección acerca de los amplificadores de FI, ambas etapas amplificadoras de FI son amplificadores sintonizados.

Ellos producen la ganancia máxima para señales de 455 KHz y prácticamente no producen ganancia para otras señales o frecuencias.

Para obtener la respuesta de frecuencia de esta etapa, mediremos el voltaje de la señal de salida del colector de Q3 para señales de entrada de diferentes frecuencias manteniendo constante su amplitud.

Este análisis mostrará que el amplificador de FI es un amplificador sintonizado que produce la ganancia máxima para señales de frecuencias cercanas a 455 KHz.

Procedimiento

Paso 1. Encienda el generador de RF, el osciloscopio, el radio y conecte el caimán negro (tierra) del generador de RF y del osciloscopio a la tierra del radio (puntilla 1 a 21).

Paso 2. Conecte el terminal del generador de RF a la base de Q3 por medio de un condensador de 0.01

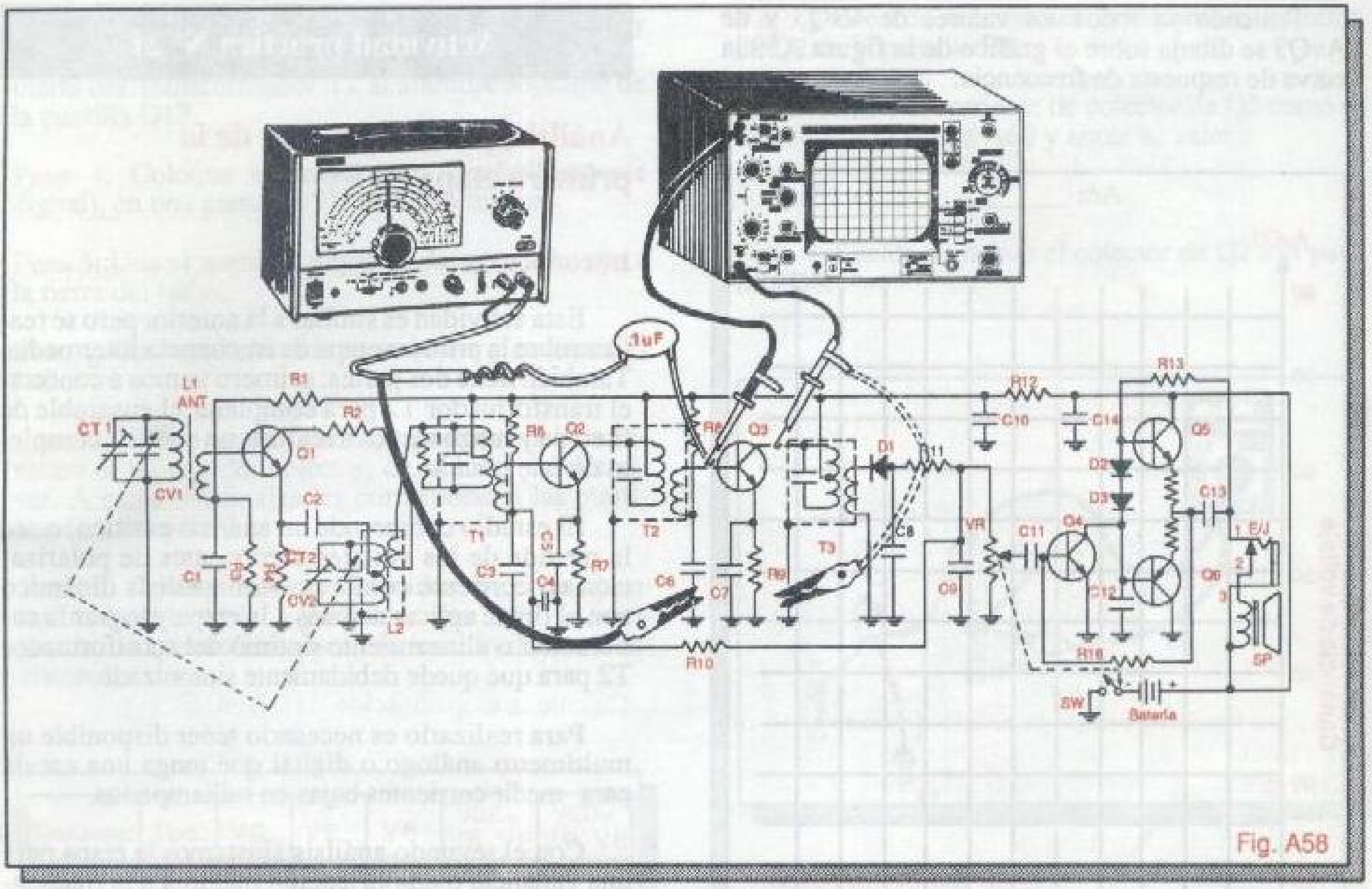
.1uF
L1
ANT
CT1
CV1
R1
R2
Q1
C2
C1
R3
CT2
CV2
L2
T1
R5
Q2
C3
C4
C5
R7
T2
R6
Q3
C6
C7
R8
R10
T3
D1
R11
C8
VR
C9
C11
R12
C10
C14
R13
Q5
D2
D3
C13
Q4
Q6
C12
R16
E/J
1
2
3
SP
SW
Batería

Fig. A58

µF ó 0.1 µF y en ese mismo punto conecte el osciloscopio (figura A58).

Paso 3. Ajuste el generador de RF para producir una señal de 400 KHz, 40 mVpp. Puede verificar el valor de este voltaje usando el osciloscopio (TIME/DIV = 2 µs, VOLT/DIV = 1mV).

Paso 4. Conecte el terminal del otro canal del osciloscopio en el colector de Q3 ajustando la perilla de VOLT/DIV en una escala que se pueda observar claramente a señal en ese punto. En nuestro caso el VsQ3:

$$VsQ3 = 900\ mVpp$$

Si el osciloscopio es de un solo canal, debe medir alternadamente la entrada y la salida e ir anotando sus valores.

Paso 5. Escriba este valor en la columna correspondiente a 400 KHz en la tabla 2.

$$AvQ3 = \frac{VsQ3}{VE\ Q3} = \frac{900\ mVpp}{40\ mVpp} = 22.5$$

AvQ3 = Ganancia de voltaje de Q3

Paso 6. Ajuste el generador de RF para producir una señal sin modulación de 410 KHz, mida VsQ3 y calcule AvQ3 para esta señal. Escriba estos dos valores en la tabla 2.

KHz	mVpp	AvQ3
400	900	22.5
410		
420		
430		
440		
450		
455		
460		
470		
480		
490		
500		

Tabla 2

Paso 7. Repita el paso 6 para las frecuencias mayores a 410 KHz y anote en la tabla 2 hasta llenarla.

Teniendo ya todos los valores de VsQ3 y de AvQ3 se dibuja sobre el gráfico de la figura A59 la curva de respuesta de frecuencia.

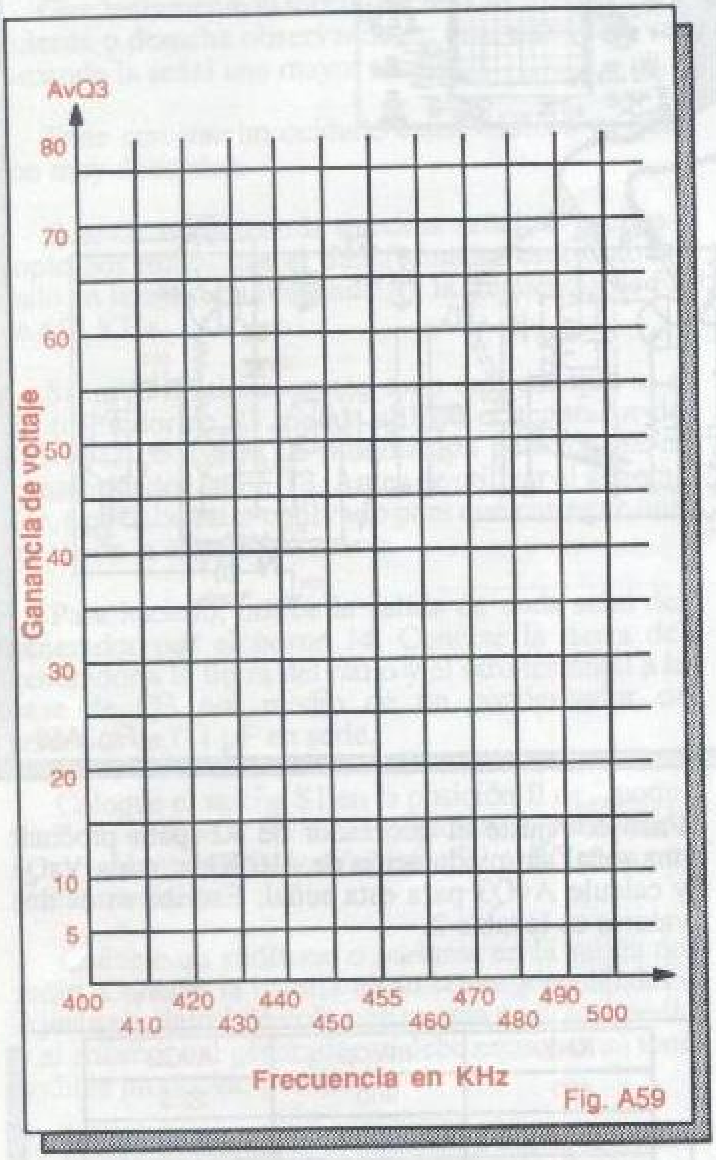

Fig. A59

En el eje horizontal están las respectivas frecuencias que corresponden a las ganancias calculadas y que están situadas en el eje vertical.

Por ejemplo, si para 400 KHz nos dio una ganancia de 22.5, situándose sobre este valor subimos hasta encontrar la ganancia de 22.5 en el eje vertical.

Marque con el lápiz el punto de cada frecuencia en la figura. Después de haber dibujado todos los puntos sobre la gráfica, únalos para obtener la curva de respuesta de frecuencia.

Esta curva debe tener su amplitud máxima para el valor de frecuencia de 455 KHz tal como se había estudiado en las lecciones teóricas sobre este tema.

Actividad práctica Nº 21

Análisis, prueba y ajuste de la primera etapa de FI

Introducción

Esta actividad es similar a la anterior pero se realiza sobre la primera etapa de frecuencia intermedia. También tiene dos partes: primero vamos a conectar el transformador T2 para completar el ensamble de la etapa y luego vamos a realizar un análisis completo de este circuito.

El estudio comprende un análisis estático, o sea la medida de los voltajes y corrientes de polarización en corriente continua, y un análisis dinámico con el fin de aplicar señales a la etapa y lograr la calibración o alineamiento óptimo del transformador T2 para que quede debidamente sintonizado.

Para realizarlo es necesario tener disponible un multímetro análogo o digital que tenga una escala para medir corrientes bajas en miliamperios.

Con el segundo análisis ajustamos la etapa para una ganancia o amplificación máxima a la frecuencia intermedia de 455 KHz.

Para esta prueba necesitamos un generador de RF y un osciloscopio. También se puede utilizar el generador de FI de 455 KHz cuya construcción se explicó en la actividad práctica Nº 17.

Análisis estático No. 1

Medición del voltaje de polarización de Q2

Procedimiento

El objetivo de este análisis es medir los voltajes de polarización de corriente continua del transistor Q2 y concluir del estudio de estas medidas si el transistor está correctamente polarizado o no.

Paso 1. Haga un puente temporal entre las puntillas D4 y D3. Este puente permanecerá durante todo el análisis de esta actividad.

Paso 2. Tome T2 (el primer transformador de FI), el cual se puede identificar por el núcleo de color blanco y suéldelo a las puntillas D15, D16, D17, E1 y E2.

Compruebe que los alambres ya soldados en las puntillas no se hayan desconectado.

Paso 3. Asegúrese que las puntillas D15, D16 y D17 no se toquen entre sí. Luego conecte la cubierta del transformador T2 al alambre sobrante de la puntilla D13.

Paso 4. Coloque el multímetro (preferiblemente digital), en una gama de 9V en CC o mayor.

Paso 5. Una el terminal negativo de su multímetro a la tierra del radio.

Paso 6. Conecte la batería en su borne y encienda el radio.

Paso 7. Mida con el terminal positivo del multímetro el voltaje del colector, de la base y del emisor. A estos terminales les corresponden las puntillas D8, D4 y D9.

Paso 8. Con los valores obtenidos en el paso anterior llene la tabla 3 y observe que aquí se comprueba lo que definimos en la lección de los transistores.

Transistor	Tipo	VC	VE	VB	VBE = VB - VE	VCB= VC - VB
Q2	NPN					

Tabla 3

Estos conducirán cuando la juntura B-E esté directamente polarizada y la juntura C-B esté inversamente polarizada.

Siendo Q2 un transistor NPN, la juntura B-E está directamente polarizada cuando la base es positiva con respecto al emisor (VCE = positivo).

La juntura C-B está inversamente polarizada cuando el colector es positivo con referencia a la base (VCB = positivo).

Análisis estático No . 2

Midiendo la corriente de colector del transistor Q2

En el análisis estático No. 2 debemos verificar que el transistor Q2 tenga el valor correcto de corriente de colector.

Procedimiento

Paso 1. Coloque el multímetro en la gama de corriente continua y en una escala mayor o igual a 5 mA.

Paso 2. Desconecte momentáneamente el colector de Q2 de la puntilla D8.

Paso 3. Mida la corriente de colector de Q2 como se muestra en la figura A60 y anote su valor :

IcQ2 = ______________ mA

Paso 4. Suelde de nuevo el colector de Q2 a la puntilla D8.

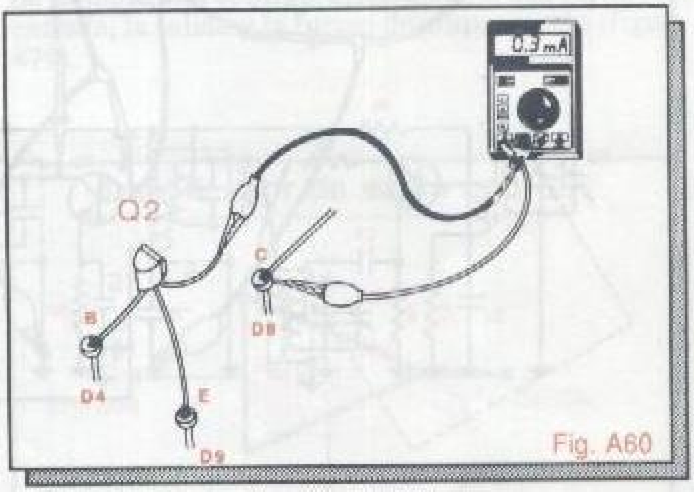

Fig. A60

Análisis dinámico No. 3

Calibración del transformador T2 y cálculo de la ganancia de voltaje del transistor Q2

El objetivo de este análisis es alinear o calibrar el transformador T2 para producir una ganancia máxima de la etapa para señales de 455 KHz. Calcularemos también la amplificación o ganancia de voltaje de la etapa para esta misma señal.

Procedimiento

Paso 1. Encienda el generador de RF (o su generador CEKIT), y el osciloscopio.

Paso 2. Conecte el terminal negativo del generador de RF y la tierra del osciloscopio al negativo del radio (puntillas 1 a 21).

Paso 3. Conecte el terminal de señal del generador de RF a la base de Q2 (puntilla D40) intercalando en serie un condensador de 0.01 µF o de 0.1µF.

Paso 4. Conecte un canal del osciloscopio también a la base de Q2. Observe la figura A61.

Paso 5. Coloque la perilla TIME/DIV del osciloscopio en 2 µS y la perilla VOLT/DIV en 1 mV. Ajuste el generador de RF para producir una señal sin modulación de 455 KHz.

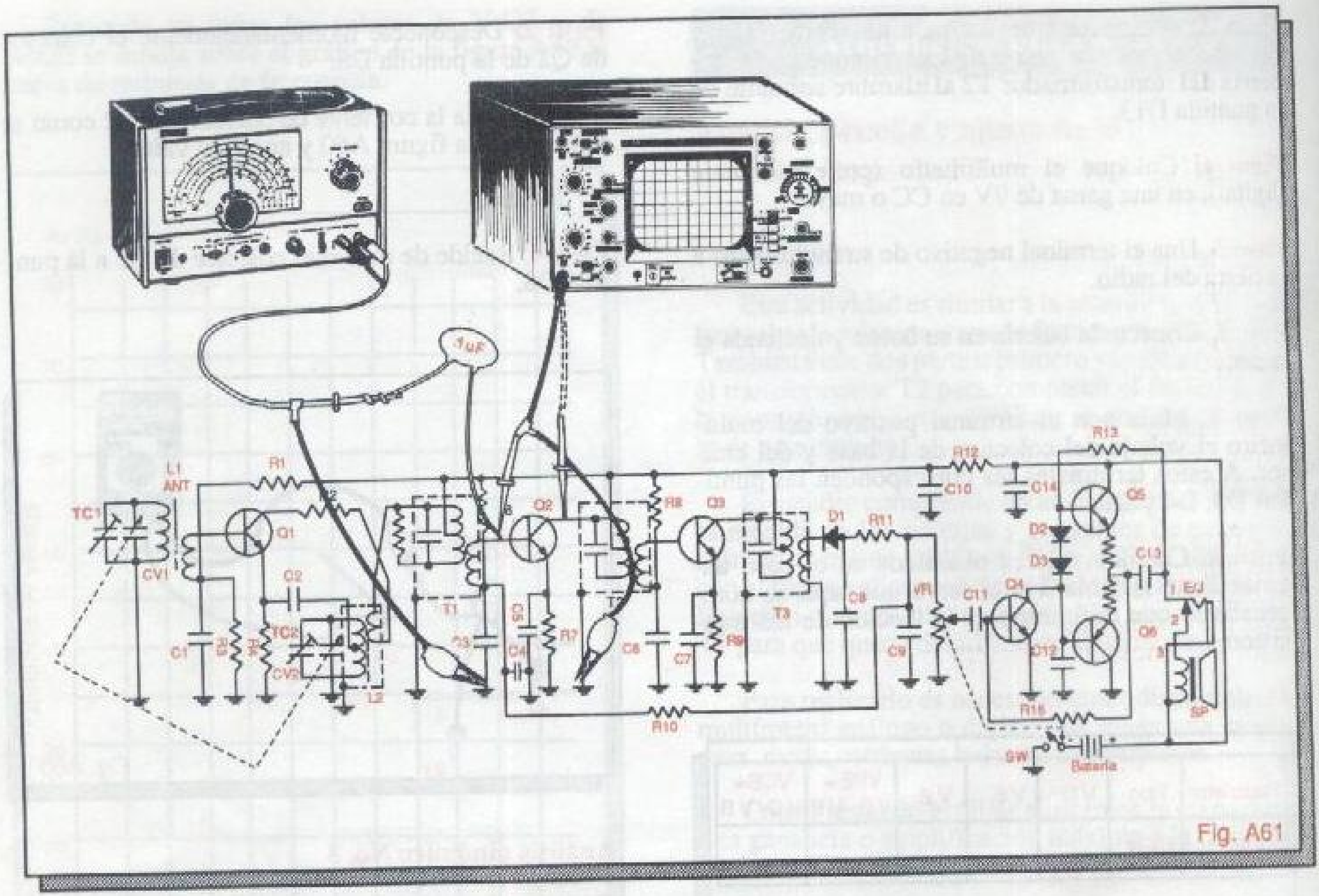

Fig. A61

Ajuste la perilla de nivel de salida del generador de RF para producir una señal de 40 mVpp. Usted medirá el voltaje de esta señal usando un osciloscopio; por lo tanto, recuerde la atenuación de los terminales.

Paso 6. Una la batería a su conector de batería y encienda el radio.

Desenganche el terminal del osciloscopio de la base de Q2 y conéctelo al colector de Q2.

Cambie la posición de la perilla VOLT/DIV del osciloscopio para ver la señal en el colector. (VOLT/DIV = 0.2 voltios).

Paso 7. De la misma forma que se explicó en la actividad anterior, ajuste el tornillo sintonizador de T2 hasta obtener la máxima amplitud de la señal en la pantalla del osciloscopio.

Calcule el voltaje pico a pico (VsQ2) de la señal y escríbala.

VsQ2 = ________ mVpp

Paso 8. Apague el radio, el generador de RF y el osciloscopio y desenganche los terminales.

Paso 9. Conociendo el voltaje de la entrada (VEQ2 = 40 mVpp) y la señal de salida del transistor Q2, calcule la ganacia de voltaje.

$$AvQ2 = \frac{VsQ2}{VE\,Q2} = \frac{\qquad}{40\text{ mVpp}}$$

Ahora calcule la ganancia de voltaje en dB.

AvQ2 = 20 Log AvQ2 =

Nota: Probablemente usted notará que la ganancia de voltaje de esta etapa es más pequeña que la de la segunda etapa amplificadora de FI.

Esto es así porque estamos inyectando una señal relativamente fuerte en la base de Q2. Por eso, el CAG causa que la amplificación de esta etapa sea pequeña.

Paso 10. Desconecte el puente entre las puntillas D4 y D3.

Si ha realizado este ajuste correctamente, la primera etapa de FI estará alineada en este momento.

Circuitos amplificadores de audio para radios AM y FM con circuitos integrados

En la Lección 18 estudiamos el tema general de los amplificadores de audio y como aplicación se analizó el amplificador del radio AM CEKIT que ha servido como actividad de práctica y experimentación en este curso.

En esta lección se analizaron circuitos con transistores bipolares y con transistores de efecto de campo o FET's.

Actualmente se ha desarrollado con gran rapidez la tecnología de los circuitos integrados. Dentro de los diferentes tipos de éstos, se encuentra una familia muy numerosa y de gran aplicación que es la de los amplificadores de audio integrados.

Debido a su pequeño tamaño, gran eficiencia y fácil utilización se ha popularizado su uso en etapas de audio de radio AM, FM, aparatos de TV, equipos de sonido, pasacintas para automóviles, radioteléfonos, intercomunicadores y muchos otros.

Un circuito integrado amplificador de audio contiene en su interior todos o casi todos los componentes que forman una etapa como las que se estudiaron en la Lección 18.

Existe actualmente una gran cantidad de modelos o referencias con potencias desde 0.5W hasta 100W o más, elaborados por varios fabricantes.

Debido a que este tema es de gran actualidad e importancia en cualquier estudio de radio vamos a presentar varios circuitos de este tipo con algunos comentarios sobre su diagrama, conexión y funcionamiento.

Para circuitos de radio vamos a conocer circuitos con potencias hasta de 10 vatios, suficiente para irradiar el sonido con muy buen volumen en un recinto cerrado.

Como ya lo hemos mencionado, una de las características más importantes de estos circuitos es la facilidad para su conexión.

Con unos pocos componentes adicionales, cualquier persona con conocimientos básicos de electrónica puede armar un amplificador de audio ya sea para aumentar el volumen de un radio, una grabadora o un tocadisco.

Configuración general

Un circuito integrado amplificador de audio tiene siempre una configuración general que incluye la entrada, la salida y la fuente de alimentación (figura 479).

Amplificador de audio con CI

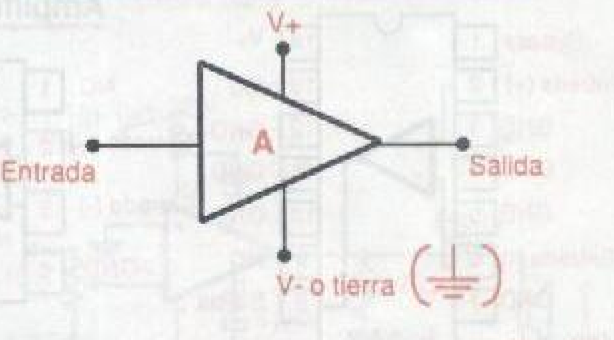

Fig. 479

Además de estos pines tiene unos terminales auxiliares para la conexión de algunos componentes externos que se utilizan para determinar las características de la etapa. El número de estos terminales varía según el modelo o referencia del circuito integrado y depende de su complejidad interna.

En la figura 480 tenemos el diagrama, el circuito interno y el aspecto físico de uno de estos circuitos integrados, el LM 386 fabricado por National Semiconductor.

Como se puede observar, en su interior están conectados varios transistores, diodos y resistencias de tal manera que cumplen la función de amplificador de audio. Más adelante mostraremos la forma de conectarlo como amplificador de señal.

A continuación vamos a conocer algunos de los modelos más comunes de este tipo de pines y sus circuitos de aplicación.

Conoceremos la familia de circuitos de National Semiconductor, que es una de las fábricas que más referencias tiene, especialmente en circuitos de baja potencia.

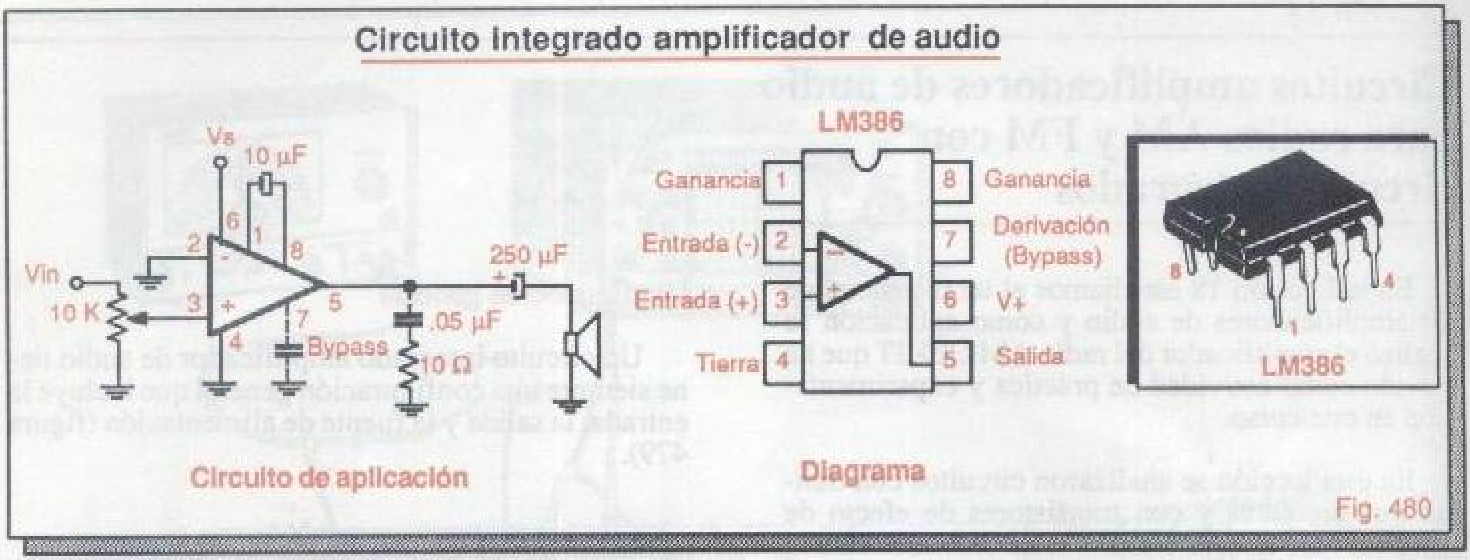
Circuito integrado amplificador de audio

Vs
10 µF
6
1
2
8
Vin
10 K
3
+
5
7
4
Bypass
250 µF
.05 µF
10 Ω
Circuito de aplicación
LM386
Ganancia 1
Entrada (-) 2
Entrada (+) 3
Tierra 4
8 Ganancia
7 Derivación (Bypass)
6 V+
5 Salida
Diagrama
8
4
1
LM386

Fig. 480

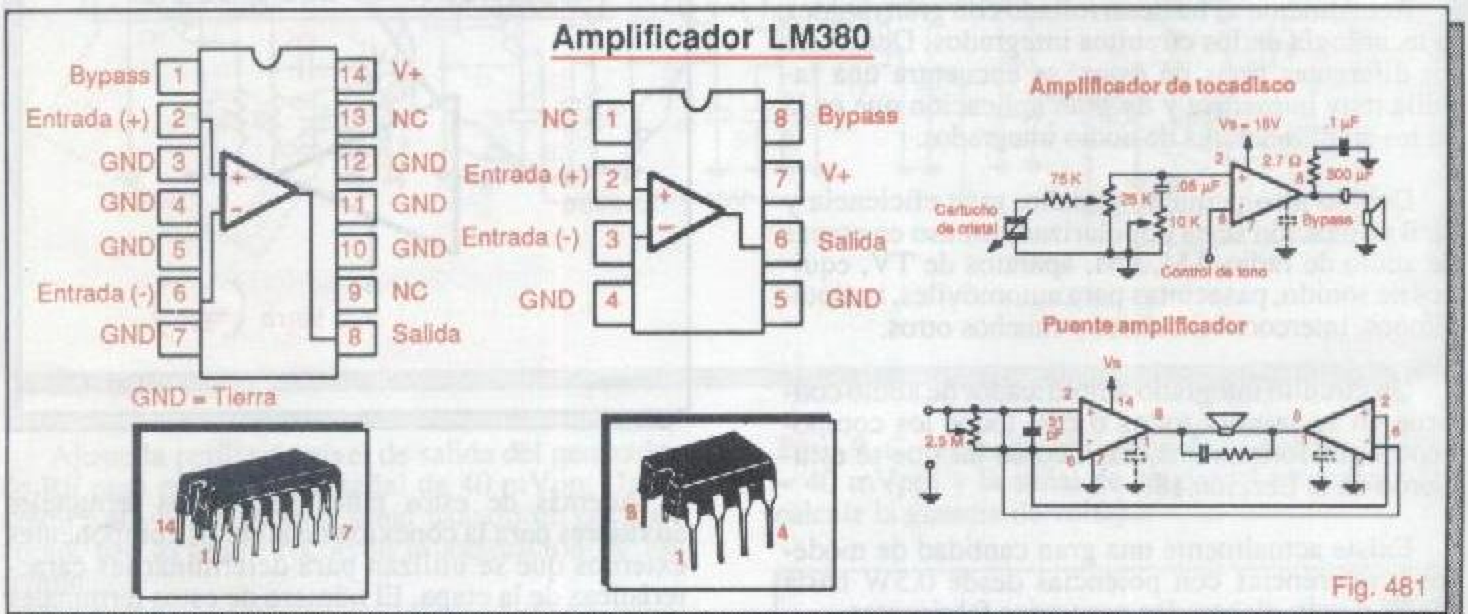
Amplificador LM380

Bypass 1
Entrada (+) 2
GND 3
GND 4
GND 5
Entrada (-) 6
GND 7
14 V+
13 NC
12 GND
11 GND
10 GND
9 NC
8 Salida
GND = Tierra
14
1
7
NC 1
Entrada (+) 2
Entrada (-) 3
GND 4
8 Bypass
7 V+
6 Salida
5 GND
8
1
4
Amplificador de tocadisco
Vs = 18V
.1 µF
2
2.7 Ω
8
300 µF
75 K
25 K
.05 µF
10 K
Cartucho de cristal
6
7
Bypass
Control de tono
Puente amplificador
Vs
2
14
8
2.5 M
6
7
1
2
6

Fig. 481

Circuito LM380

Este fue uno de los primeros circuitos de este tipo y aún se sigue utilizando ampliamente. Se fabrica en dos versiones: de 14 pines y de 8 pines.

En la figura 481 tenemos sus diagramas de pines, características eléctricas y dos circuitos de aplicación. Note que en el segundo circuito se utiliza una configuración en puente con la que se obtiene mayor potencia conectando dos circuitos iguales.

Este tipo de configuración es muy común en estos circuitos como lo veremos más adelante.

Para aplicar un circuito integrado amplificador de audio a un radio se toma la señal del detector y se lleva a la entrada del circuito. La fuente de alimentación y la tierra deben ser comunes. El parlante va a la salida del circuito integrado (figura 482).

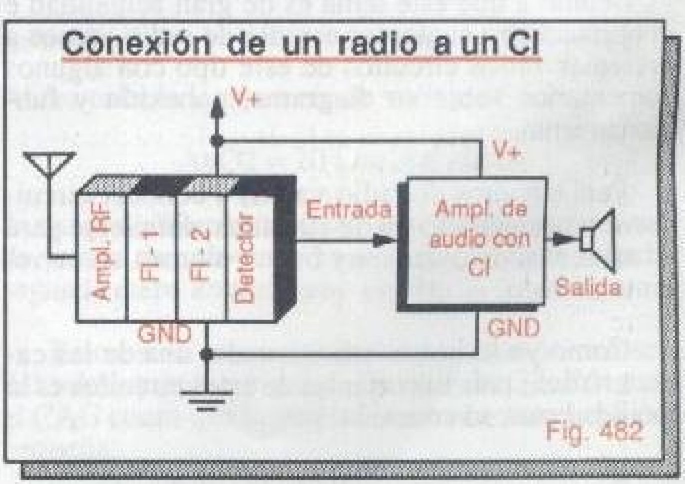
Conexión de un radio a un CI

V+
V+
Ampl. RF
FI - 1
FI - 2
Detector
Entrada
Ampl. de audio con CI
Salida
GND
GND

Fig. 482

Circuito LM386

Este circuito está diseñado para trabajar a bajos voltajes y es muy práctico para ser utilizado en ra-

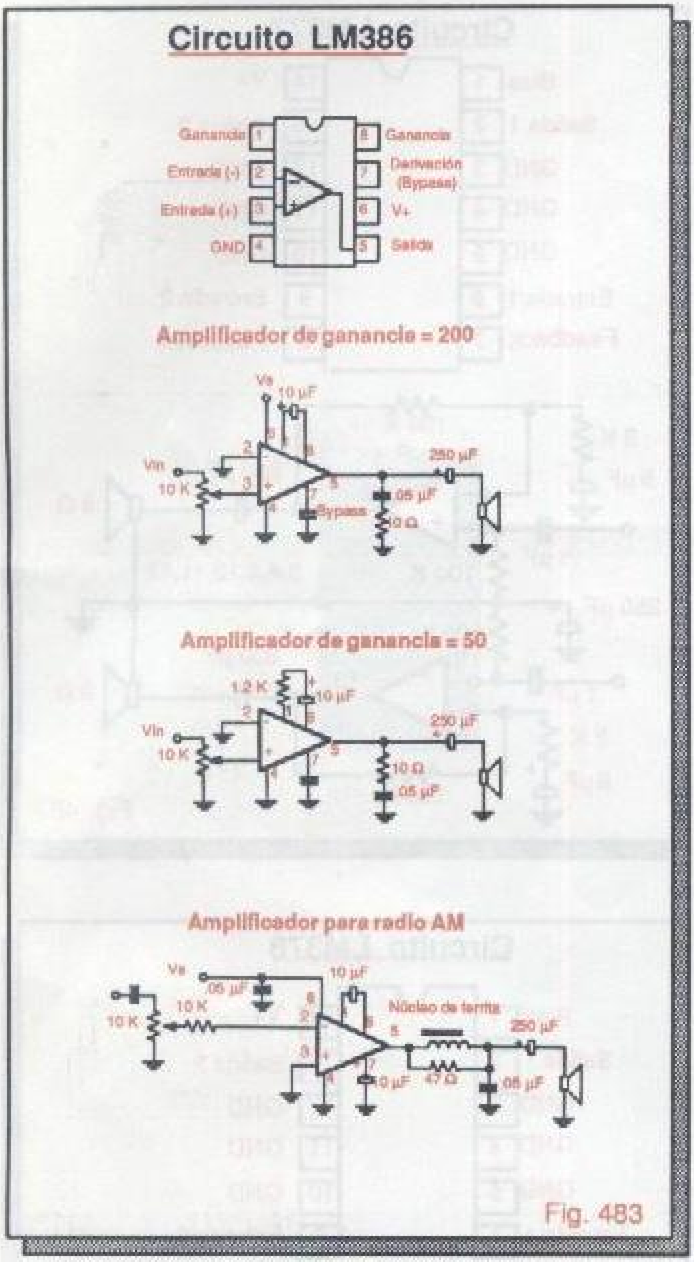
Ganancia 1
Entrada (-) 2
Entrada (+) 3
GND 4
8 Ganancia
7 Derivación (Bypass)
6 V+
5 Salida
Vs
10 µF
Vin
10 K
250 µF
.05 µF
10 Ω
Bypass
1.2 K
10 µF
250 µF
10 K
10 Ω
.05 µF
Vs
.05 µF
10 µF
10 K
10 K
Núcleo de ferrita
250 µF
47 Ω
10 µF
.05 µF

Circuito LM386

Amplificador de ganancia = 200

Amplificador de ganancia = 50

Amplificador para radio AM

Fig. 483

dios AM y FM pequeños. La figura 483 muestra su diagrama de pines y algunos circuitos de aplicación.

Una de sus funciones es la de poder variar su ganancia por medio de componentes externos conectados entre los pines 1 y 8. Si no hay nada conectado, la ganancia es de 20; con un condensador la ganancia es de 200 y con una resistencia en serie con un condensador se puede variar la ganancia entre 20 y 200.

Los componentes que están conectados con líneas punteadas son opcionales y mejoran la calidad del sonido.

Circuito LM384

EL LM384 es un circuito integrado que suministra una potencia hasta de 5 vatios. Por su configuración interna, requiere muy pocos componentes externos, lo que hace muy fácil su conexión. En la figura 484 tenemos el diagrama de pines y un circuito de aplicación.

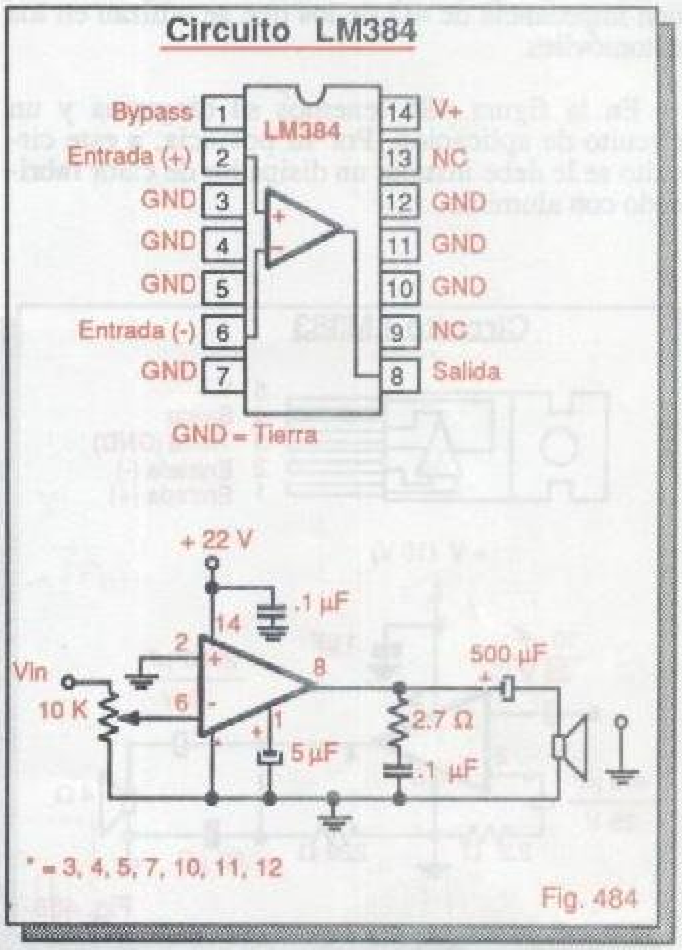
Bypass 1
Entrada (+) 2
GND 3
GND 4
GND 5
Entrada (-) 6
GND 7
LM384
14 V+
13 NC
12 GND
11 GND
10 GND
9 NC
8 Salida
GND = Tierra
+ 22 V
.1 µF
14
2
8
Vin
10 K
6
1
5 µF
500 µF
2.7 Ω
.1 µF
* = 3, 4, 5, 7, 10, 11, 12

Circuito LM384

Fig. 484

La salida tiene protección contra cortocircuito por medio de limitación térmica. Debido a su potencia, este circuito alcanza una temperatura alta y se le debe instalar un disipador de calor como el que se muestra en la figura 485.

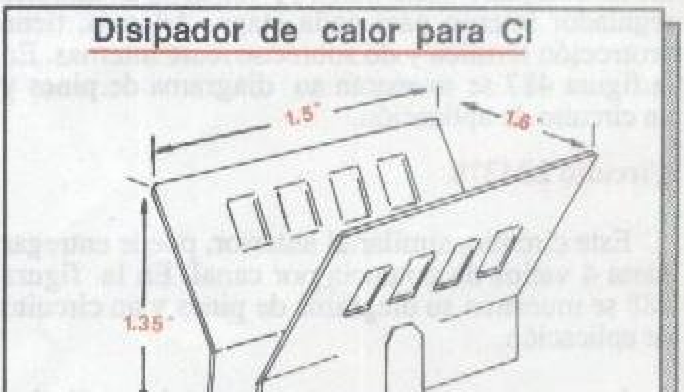

Disipador de calor para CI

Fig. 485

Circuito LM383

El LM383 es un circuito integrado que puede entregar una potencia hasta de 7 vatios. Por sus características de manejo de corrientes altas (hasta de 3.5 Amps) se puede utilizar para manejar parlantes con impedancia de 4Ω de los que se utilizan en los automóviles.

En la figura 486 tenemos su diagrama y un circuito de aplicación. Por su potencia, a este circuito se le debe instalar un disipador de calor fabricado con aluminio.

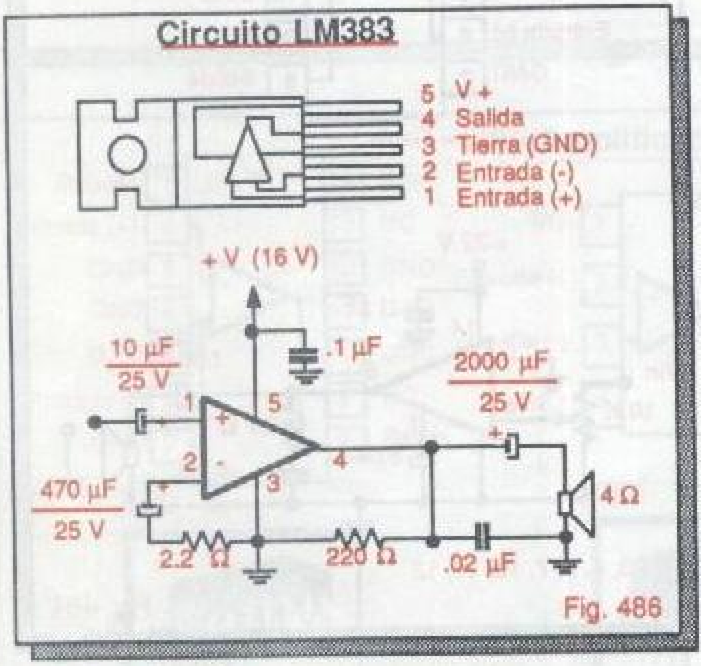
Circuito LM383
5 V +
4 Salida
3 Tierra (GND)
2 Entrada (-)
1 Entrada (+)
+V (16 V)
.1 µF
10 µF
25 V
2000 µF
25 V
1
5
2
4
3
470 µF
25 V
4 Ω
2.2 Ω
220 Ω
.02 µF

Fig. 486

Circuito LM377

El LM377 es un amplificador integrado doble de potencia, de 2 vatios por canal, que se puede utilizar para tocadiscos, grabadoras, radios FM, etc.

Este circuito está diseñado para utilizar el mínimo de componentes externos y contiene un circuito regulador interno para cada etapa. Además, tiene protección térmica y de sobrecorriente internas. En la figura 487 se muestran su diagrama de pines y un circuito de aplicación.

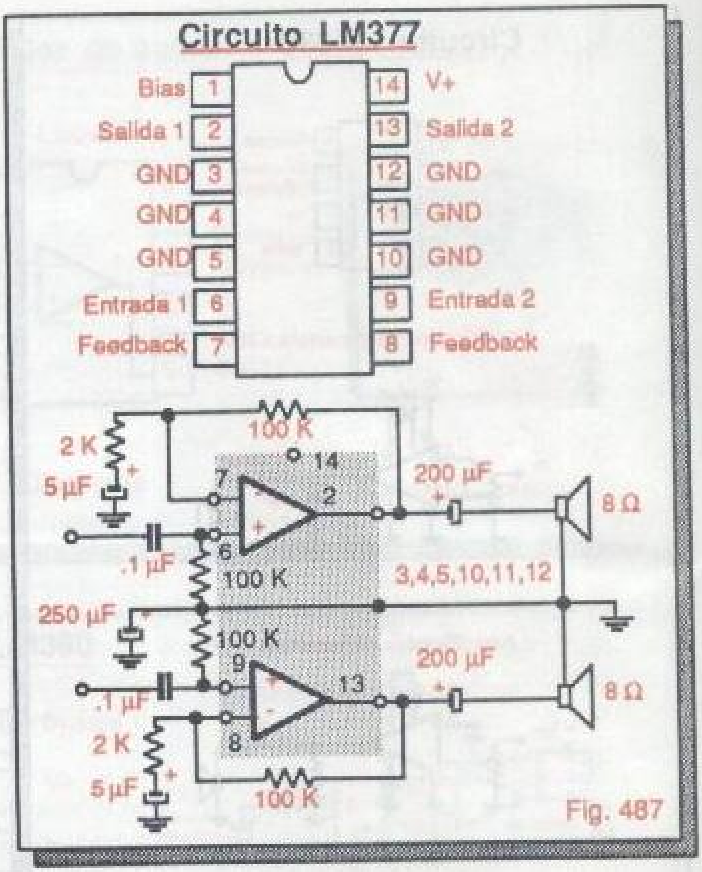
Circuito LM377
Bias 1
Salida 1 2
GND 3
GND 4
GND 5
Entrada 1 6
Feedback 7
14 V+
13 Salida 2
12 GND
11 GND
10 GND
9 Entrada 2
8 Feedback
100 K
2 K
5 µF
14
7
2
200 µF
8 Ω
.1 µF
6
100 K
3,4,5,10,11,12
250 µF
100 K
9
13
200 µF
8 Ω
.1 µF
8
2 K
5 µF
100 K

Fig. 487

Circuito LM378

Este circuito, similar al anterior, puede entregar hasta 4 vatios de potencia por canal. En la figura 488 se muestran su diagrama de pines y un circuito de aplicación.

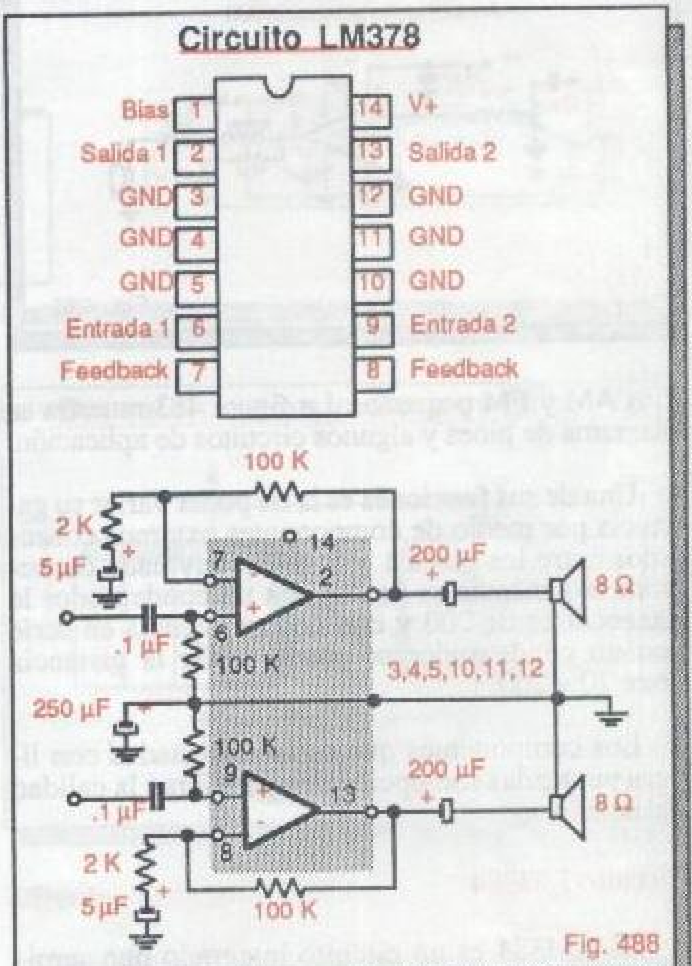
Circuito LM378
Bias 1
Salida 1 2
GND 3
GND 4
GND 5
Entrada 1 6
Feedback 7
14 V+
13 Salida 2
12 GND
11 GND
10 GND
9 Entrada 2
8 Feedback
100 K
2 K
5 µF
14
7
2
200 µF
8 Ω
.1 µF
6
100 K
3,4,5,10,11,12
250 µF
100 K
9
13
200 µF
8 Ω
.1 µF
8
2 K
5 µF
100 K

Fig. 488

Los componentes extenos son mínimos. En los dos circuitos anteriores se debe utilizar un disipador de calor similar al usado por el LM384.

Otros circuitos integrados para AM y FM

Introducción

En esta lección se estudia una serie de circuitos integrados especializados en el procesamiento de señales AM, FM y FM estéreo. Todos estos dispositivos requieren de muy pocos componentes externos para configurar receptores muy compactos y ofrecen unas características de calidad sorprendentes.

La mayoría trae incorporado un amplificador de audio. Otros necesitan un amplificador de audio externo, transistorizado o integrado. Además, son muy flexibles en cuanto al rango de voltajes de alimentación con que pueden operar.

En la primera parte se estudian los sistemas integrados de recepción de AM adecuados para la banda doméstica de 530 a 1600 KHz. Se analiza en detalle el CI ZN414Z de Ferranti y se describen otros, destacándose los chips LM1863 y LM3820 de National.

En la segunda parte se estudian los sistemas integrados de recepción de FM adecuados para la banda estándar de 88 a 108 MHz. Se analiza en detalle el CI TDA7000 de SGS-ATES y se describen brevemente otros como el LM1865 y el LM 3089.

En la tercera parte se estudian los sistemas integrados de recepción simultánea de AM y FM. Se analiza en detalle el CI LM1868 de National y se describen brevemente otros como los de la serie TDA1220 de SGS.

En la parte final se estudian los sistemas integrados de demodulación de FM estéreo. Se amplían los conceptos de FM estéreo introducidos en la Lección 29 y se analiza la operación del CI CA3089 de RCA. También se describen otros como el LM1800 y el LM1870.

En todos los casos se suministra información práctica muy útil sobre la operación y utilización de los chips más representativos, y las funciones de cada pin. También se proporcionan circuitos de aplicación típicos, fáciles de implementar con componentes comunes y se explica su funcionamiento.

Sistema de recepción de AM ZN414Z

El CI ZN414Z de Ferranti es un sistema completo de recepción de AM que contiene, en una misma cápsula T0-8 de 3 pines, un amplificador de RF de 4 etapas con CAG y un detector a transistor. El circuito consiste internamente de 10 transistores, 15 resistencias y 4 condensadores.

En la figura 489 se muestra el aspecto físico, el símbolo y el diagrama interno de bloques del CI ZN414Z. La entrada de RF (pin 2) es de muy alta impedancia (4 MΩ). La salida de audio (pin 1) puede manejar directamente un audífono de 400 a 600 Ω sin necesidad de amplificador de audio.

El circuito integrado ZN414Z

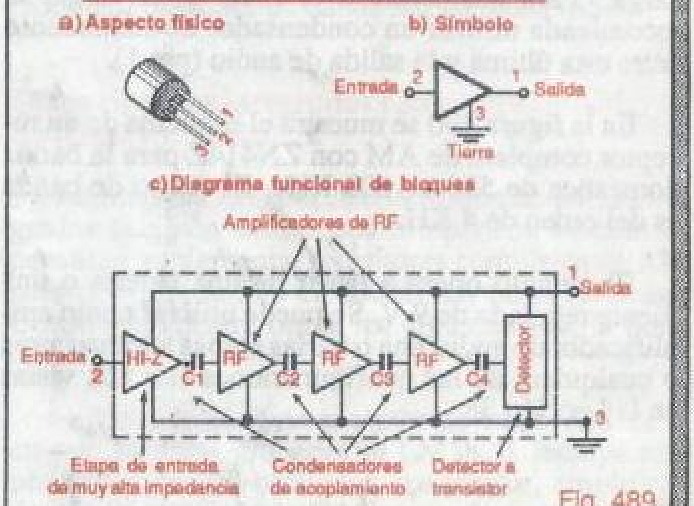

Fig. 489

El ZN414Z puede operar con tensiones de alimentación desde 1.2 hasta 1.6 V. El fabricante recomienda utilizar una fuente de 1.3 V. Bajo estas condiciones, el consumo de corriente es típicamente de 0.3 a 0.5 mA.

El ZN414Z se especifica para trabajar en el rango de frecuencias comprendido entre 150 KHz a 3 MHz. Su detector ha sido diseñado para responder a señales de AM dentro de esta gama. Sin embargo, ha sido utilizado con éxito a frecuencias superiores a 7 MHz.

Esta característica lo hace apropiado para ser utilizado como receptor de AM en las bandas de onda larga (150 a 300 KHz), onda media o AM estándar (530 a 1600 KHz) y de señalización de tiempo o WWV (2.5, 5, 10 y 15 MHz) de la NBS (Buró Nacional de Estándares) de los Estados Unidos.

Las emisiones de WWV las origina la NBS desde Boulder (Colorado, EE-UU). Consisten en tonos alternados de 440 y 680 Hz que modulan en amplitud la portadora principal y se utilizan para calibrar la frecuencia de equipos de comunicaciones tanto profesionales como de radioafición.

La ganancia de potencia típica del ZN414Z es de 72 dB pero se han reportado ganancias de potencia hasta de ¡110 dB! La salida de audio es del orden de 30 mV (valor rms). Con una fuente de 1.3 V, el umbral de sensibilidad (mínima señal de entrada discernible sin ruido) es de 50 µV.

El ZN414Z puede manejar directamente un audífono de 400 a 600 Ω. Para impulsar un parlante de baja impedancia, la salida de audio (pin 1) debe acoplarse capacitivamente a la entrada de un amplificador de audio apropiado.

La acción del control automático de ganancia (CAG) del ZN414Z depende de la impedancia de la carga. Para independizar el CAG de la carga, se recomienda utilizar un condensador de aislamiento entre esta última y la salida de audio (pin 1).

En la figura 490 se muestra el esquema de un receptor completo de AM con ZN414Z para la banda doméstica de 530 a 1600 KHz. El ancho de banda es del orden de 4 KHz.

El circuito opera a partir de una batería o una fuente regulada de 9 V. Se puede utilizar como amplificador de audio una o varias etapas a transistores o cualquiera de las configuraciones con CIs vistas en la Lección 38.

El regulador de voltaje constituido por el transistor Q1, las resistencias R2, R3 y R4, el potenciómetro R5 y el condensador C2 permite que el receptor opere confiablemente a partir de la fuente de 9V. El condensador C3 aisla el circuito de CAG del ZN414Z del amplificador de audio.

El potenciómetro R5 se utiliza para ajustar el voltaje entre los pines 1 y 3 (puntos A y B) del ZN414Z al nivel de 1.3 V recomendado por el fabricante. Una vez fijado este voltaje, R5 puede ser sustituido por una resistencia de valor apropiado.

La bobina L1 y el condensador variable CV1 son unidades estándar de 30 µH y 365 pF, respectivamente, del mismo tipo de las utilizadas en el receptor de AM CEKIT. Estos componentes deben mantenerse aislados del resto del circuito para minimizar efectos de realimentación no deseables.

En la figura 491 se muestra una versión simplificada del circuito de la figura 490 utilizando un audífono de 400 a 600 Ω y una pila de 1.5 V. Este receptor "mínimo" capta estaciones locales de AM con sorprendente claridad y a un buen volumen.

El circuito de la figura 491 puede convertirse fácilmente en un receptor de WWV de 2.5 MHz con unas pocas modificaciones. En la figura 492 se muestra el diagrama correspondiente. Para proveer una óptima recepción de la señal WWV, utilice como antena un alambre de 4 a 5 m de longitud.

Las bobinas L2 y L4 son unidades estándar de 30 µH con núcleos ajustables de ferrita. La derivación de L2 debe realizarse a 10 espiras del extremo

Receptor completo de AM con ZN414Z

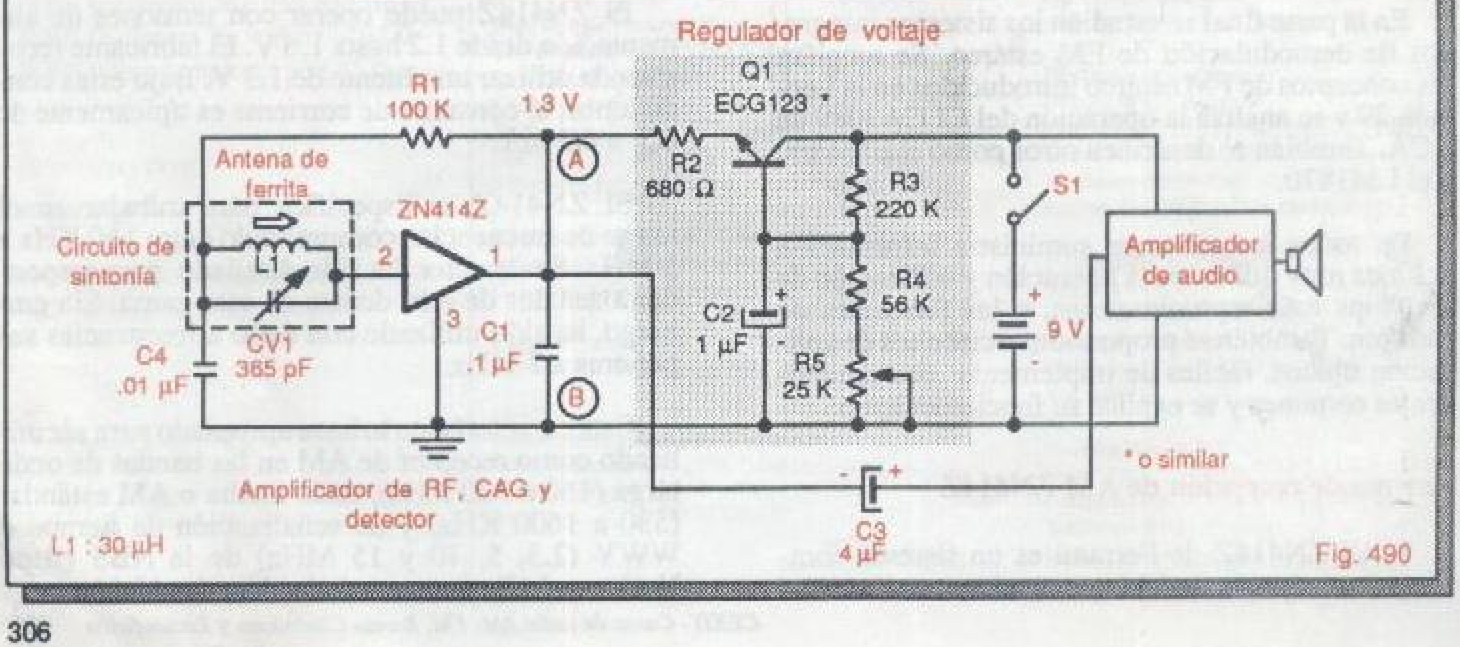

Fig. 490

Receptor mínimo de AM con ZN414Z

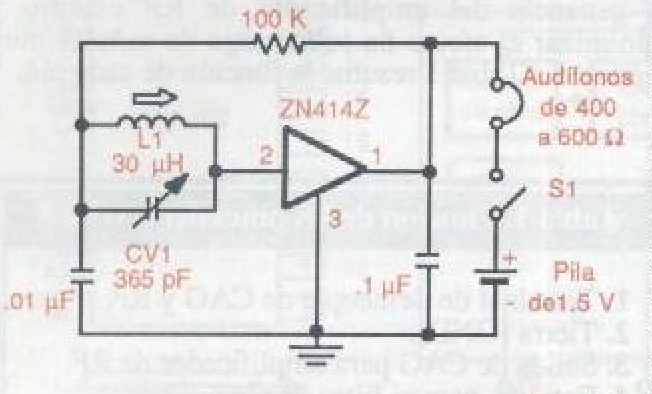

Fig. 491

inferior para proporcionar un buen punto de acoplamiento con L3.

Receptor de WWV con ZN414

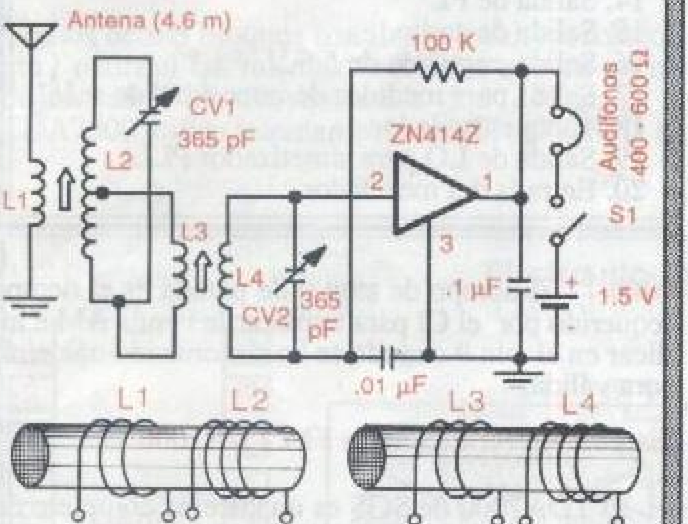

Fig. 492

Las bobinas L1 y L3 consisten de 10 espiras de alambre aislado o Litz y se acoplan magnéticamente a L2 y L4 respectivamente, como se indica. El acoplamiento magnético puede también efectuarse devanando L1 alrededor de L2 y L3 alrededor de L4.

Cuando sintonice el receptor de la figura 492, utilice los condensadores variables CV1 y CV2 para la sintonía propiamente dicha y desplace los núcleos de L2 y L4 para conseguir ajustes finos.

Manipule estos componentes hasta conseguir la recepción más fuerte y clara posible de la señal WWV de 2.5 MHz. Una vez conseguido el punto óptimo de calibración, inmovilice las bobinas y condensadores con cemento plástico o similar.

Recuerde que la acción del CAG del ZN414Z depende fuertemente de la impedancia de los audífonos. Por esta razón, utilice preferiblemente audífonos de 400 a 600 Ω. Tenga también presente que la señal WWV sólo se emite a determinados momentos del día.

Además de receptores de AM y WWV, el ZN414Z puede también utilizarse en una gran variedad de aplicaciones. Es sin duda uno de los chips de experimentación con radio más versátiles y económicos de que se dispone actualmente.

Por ejemplo, si se sintoniza a una frecuencia fija de 1.75 MHz (FI) y se continúa con un convertidor de 2 a 54 MHz, pueden captarse transmisiones de onda corta. Si se introduce realimentación positiva de RF y se hace oscilar, el ZN414Z permite demodular emisiones de banda lateral única (SSB).

Otros circuitos integrados receptores de AM. LM1863 y LM3820.

Además del ZN414Z, existen otros circuitos integrados que, con unos pocos componentes externos, permiten implementar receptores completos de AM para la banda de 530 a 1600 KHz y otros servicios. Destacamos, entre otros, los siguientes de la línea de National:

•LM3820. Sistema receptor de radio AM de 3 etapas. 14 pines. Superior al LM1820. Incluye amplificador de RF, oscilador, mezclador, amplificador de FI, detector de CAG y regulador de voltaje. Tensión de alimentación máxima: 16 V.

•LM1863. Sistema receptor de AM para radios de sintonía electrónica. 20 pines. Se utiliza en receptores de AM con búsqueda automática de emisoras. Puede manejar directamente un sintetizador PLL. Compatible con AM estéreo. Voltaje de alimentación máximo: 16 V.

En la figura 493 se muestra la configuración de pines y el diagrama funcional de bloques del CI LM3820. El sistema incorpora, en una cápsula DIP de 14 pines, un amplificador de RF, un oscilador local, un mezclador, un amplificador de FI, un circuito de CAG y un regulador Zener .

En la figura 494 se muestra un circuito típico de aplicación del LM3820. El tanque doblemente sintonizado T1-T2, que acopla la salida del mezclador (pin 14) con la entrada del amplificador de FI (pin 7), provee una alta selectividad.

El circuito integrado LM3820

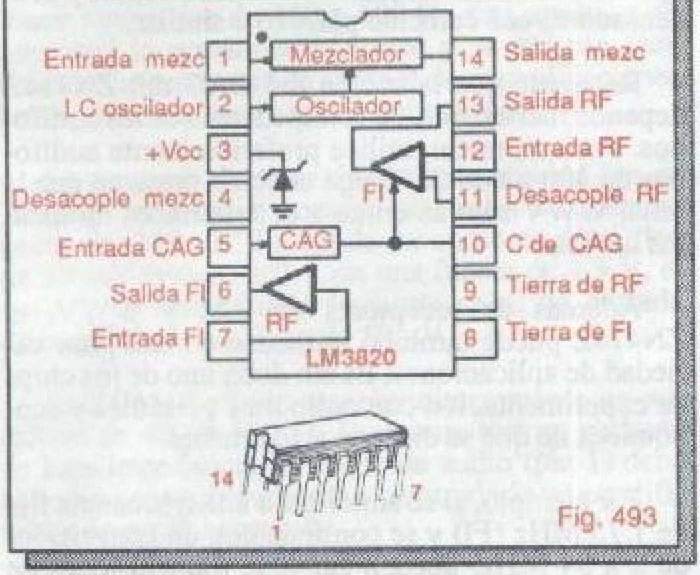

Entrada mezc 1
LC oscilador 2
+Vcc 3
Desacople mezc 4
Entrada CAG 5
Salida FI 6
Entrada FI 7
14 Salida mezc
13 Salida RF
12 Entrada RF
11 Desacople RF
10 C de CAG
9 Tierra de RF
8 Tierra de FI
Mezclador
Oscilador
FI
CAG
RF
LM3820
14
7
1

Fig. 493

Las bobinas de antena (L1) y osciladora (L2), así como los transformadores de FI (T1, T2, T3) y el condensador variable de dos secciones (CA, CB), son estándares, del mismo tipo de las utilizadas en el receptor de AM CEKIT.

Receptor de AM con CI LM3820

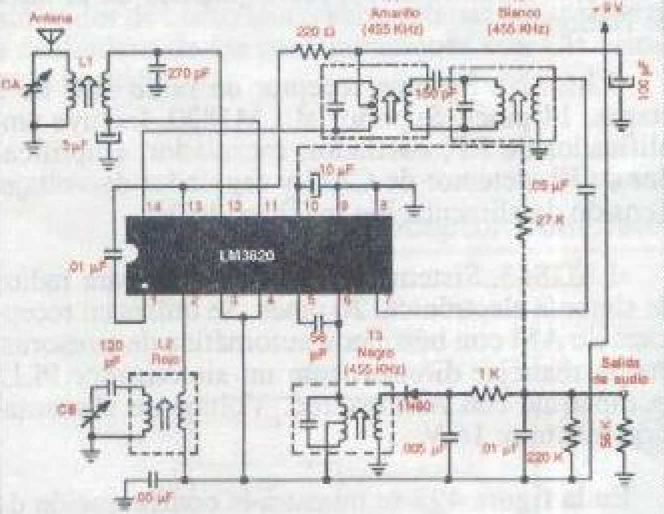

Antena
CA
L1
270 pF
5 pF
220 Ω
T1
Amarillo
(455 KHz)
T2
Blanco
(455 KHz)
150 pF
+9 V
100 µF
10 µF
.05 µF
27 K
14
13
12
11
10
9
8
LM3820
.01 µF
1
2
3
4
5
6
7
56 pF
120 pF
L2
Rojo
CB
T3
Negro
(455 KHz)
1N60
1 K
.005 µF
.01 µF
220 K
Salida de audio
.05 µF

Fig. 494

En la figura 495 se muestra la configuración de pines y el diagrama interno de bloques del CI LM1863. El sistema incorpora, en una cápsula SO o de montaje superficial de 20 pines, un oscilador local, un mezclador, un amplificador de FI de dos etapas, un detector y un regulador de voltaje.

El oscilador local puede manejar directamente un sintetizador o rastreador de frecuencias externo. El chip utiliza un detector de CAG de RF para reducir la ganancia del amplificador de RF externo y minimizar el efecto de sobrecarga de señales muy fuertes. La Tabla 1 resume la función de cada pin.

Tabla 1. Función de los pines del LM1863

1. Terminal de desacople de CAG y RF.
2. Tierra (GND).
3. Salida de CAG para amplificador de RF.
4. Entrada para el filtro de CAG.
5. Terminal de ajuste del tiempo de stop.*
6. Entrada de la fuente de alimentación (Vcc).
7. Entrada a cristal del detector de emisoras.
8. No se utiliza.
9. Salida digital de control del sintetizador.
10. Salida del mezclador.
11. Entrada de FI.
12. Desacople de FI.
13. No se utiliza.
14. Salida de FI.
15. Salida de audio.
16. Salida regulada de 5.6 V.
17. Salida para medidor de intensidad de señal.
18. Tanque oscilador.
19. Salida de LO para sintetizador PLL.
20. Entrada del mezclador.

Nota *: El tiempo de stop o de parada es el tiempo requerido por el CI para explorar la banda AM e indicar en el pin 9 cuándo se ha sintonizado una emisora válida.

Sistema de recepción de FM TDA7000

El TDA7000 de SGS es un sistema completo de recepción de FM que incluye, en una misma cápsula DIP de 18 pines, un oscilador local, un mezclador, un filtro de FI de dos etapas, un amplificador limitador de FI, un demodulador de FM y un circuito de *mute* o de silenciamiento de ruido.

El filtro de FI de dos etapas determina la selectividad del receptor. Se caracteriza por ser *activo*. La frecuencia central, así como las frecuencias inferior y superior de la banda de paso, se establecen mediante condensadores externos.

Los filtros activos realizan las mismas funciones de los filtros pasivos LC estudiados en la Lección 11 pero no utilizan bobinas. Operan a base de condensadores, resistencias y amplificadores operacionales. En el TDA7000, estos dos últimos tipos de componentes están incorporados en el chip.

El circuito integrado LM1863

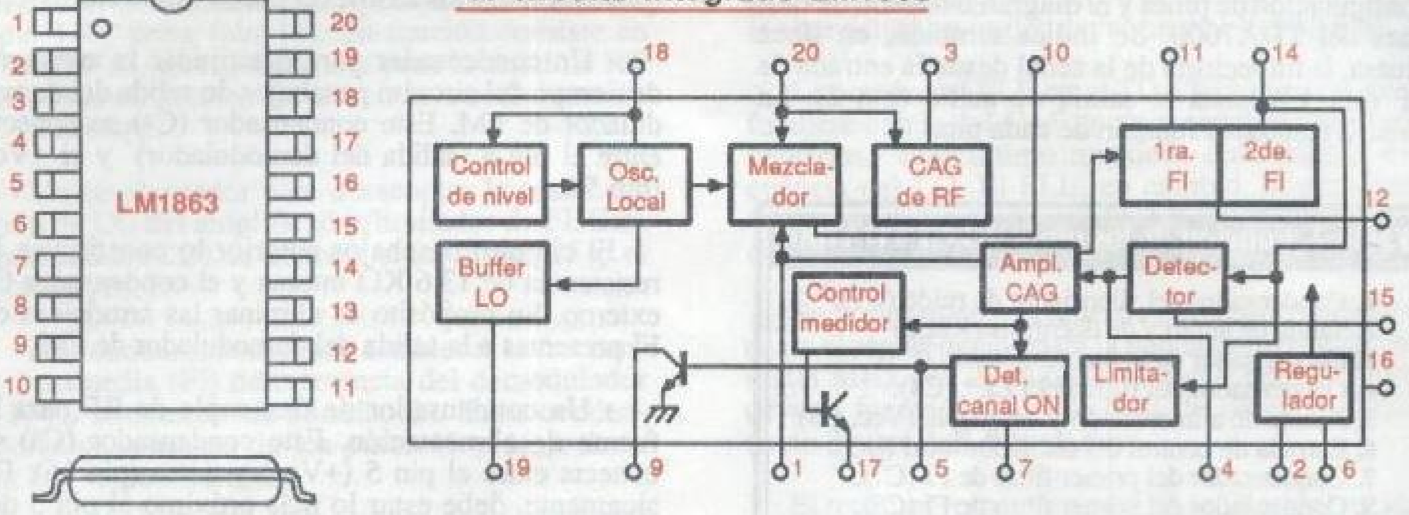
1
2
3
4
5
6
7
8
9
10
20
19
18
17
16
15
14
13
12
11
LM1863
18
20
3
10
11
14
Control de nivel
Osc. Local
Mezclador
CAG de RF
1ra. FI
2de. FI
12
Buffer LO
Ampl. CAG
Detector
15
Control medidor
Det. canal ON
Limitador
Regulador
16
19
9
1
17
5
7
4
2
6

Fig. 495

El circuito silenciador de ruido impide la recepción de espúreas y elimina pulsos de ruido superpuestos a la señal de audio. Lo controla internamente un *detector de correlación* asociado al demodulador de FM.

Los demás bloques (oscilador local, mezclador, etc.) realizan las mismas funciones que en un receptor de FM convencional. Sin embargo, el TDA7000 utiliza internamente una FI de 70 KHz en lugar de la FI estándar de 10.7 MHz. Las razones de esto último son puramente técnicas y se explican más adelante.

El TDA7000 resulta adecuado para radios de FM portátiles por la mínima cantidad de componentes externos o periféricos que requiere. También se puede utilizar como receptor en equipos de banda ciudadana, teléfonos inalámbricos, sección de sonido de televisores y otros sistemas FM.

El circuito integrado 7DA7000

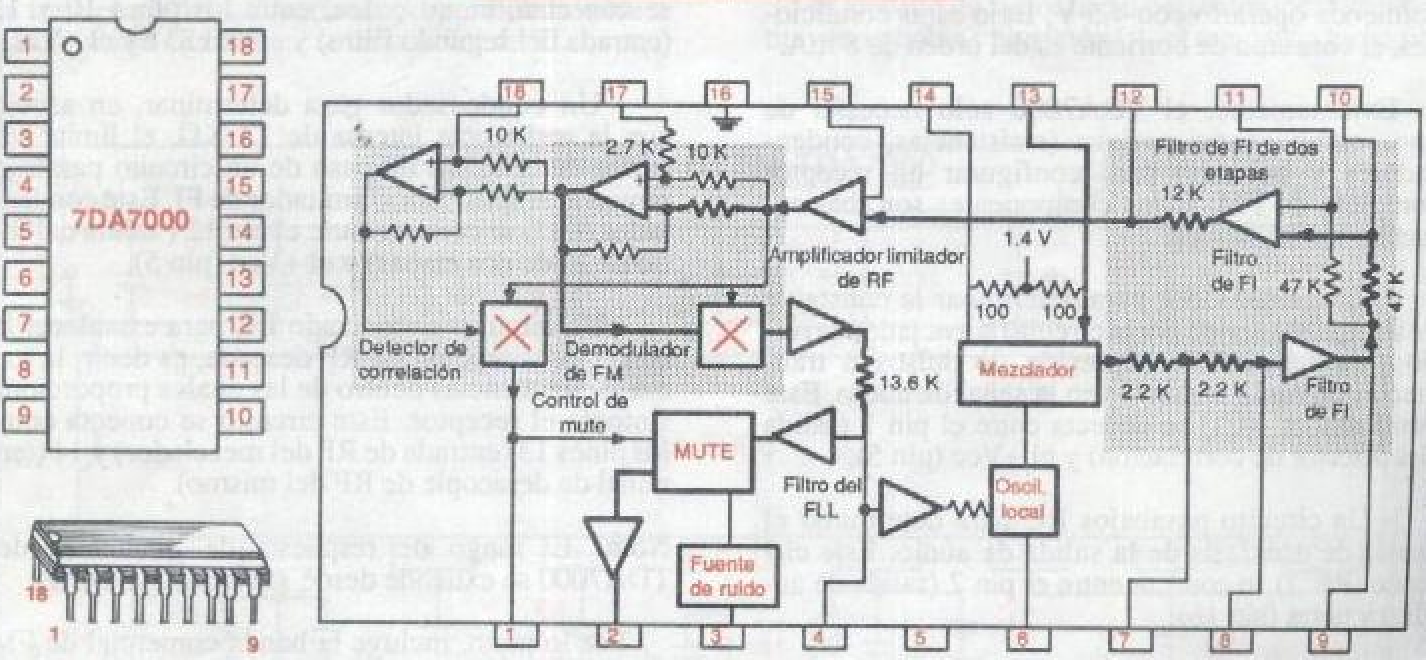
1
2
3
4
5
6
7
8
9
18
17
16
15
14
13
12
11
10
7DA7000
18
1
9
10 K
2.7 K
10 K
Filtro de FI de dos etapas
12 K
Filtro de FI
47 K
47 K
1.4 V
100
100
Amplificador limitador de RF
Detector de correlación
Demodulador de FM
13.6 K
Control de mute
MUTE
Mezclador
2.2 K
2.2 K
Filtro de FI
Filtro del FLL
Oscil. local
Fuente de ruido

Fig. 496

En la figura 496 se muestra el aspecto físico, la configuración de pines y el diagrama interno de bloques del TDA7000. Se indica también, en línea gruesa, la trayectoria de la señal desde la entrada de RF (pin 13) hasta la salida de audio (pin 2). La Tabla 2 resume la función de cada pin.

Tabla 2. Función de los pines del CI TDA7000

1. Condensador del silenciador de ruido (C1).
2. Salida de audio y de deénfasis (R2C2)
3. Condensador de la fuente de ruido (C3).
4. Condensador del filtro del FLL* (C4).
5. Fuente de alimentación y desacople (Vcc, C5).
6. Entrada de control del oscilador local (LO).
7. Condensador del primer filtro de FI (C7).
8. Condensador del primer filtro de FI (C8).
9. Condensador del primer filtro de FI (C7).
10. Condensador del segundo filtro de FI (C10).
11. Condensador del segundo filtro de FI (C11).
12. Condensador de salida del filtro de FI (C12).
13. Entrada de RF del mezclador (RF)
14. Desacople del mezclador (C14)
15. Desacople del amplificador de FI (C15)
16. Terminal de referencia (GND o tierra).
17. Condensador del demodulador (C17).
18. Condensador del detector de correlación (C18).

* Nota : FLL es la abreviatura de frequency-locked loop (lazo de amarre de frecuencia). Se trata de un circuito interno de control del oscilador local. Su operación se explica más adelante.

El TDA7000 puede trabajar con tensiones de alimentación desde 3V hasta 12 V. El fabricante recomienda operarlo con 4.5 V. Bajo estas condiciones, el consumo de corriente es del orden de 8 mA.

Externamente, el TDA7000 sólo necesita de unos componentes pasivos (resistencias, condensadores y bobinas) para configurar un receptor completo de FM. Estos componentes son, básicamente, los siguientes:

• Un condensador para determinar la constante de tiempo requerida por el circuito silenciador de ruido para asegurar la supresión de pulsos o transientes de ruido presentes en la señal de audio. Este condensador (C1) se conecta entre el pin 1 (salida del detector de correlación) y el +Vcc (pin 5).

• Un circuito pasabajos RC para determinar el grado de deénfasis de la salida de audio. Este circuito (R2C2) se conecta entre el pin 2 (salida de audio) y tierra (pin 16).

• Un condensador para determinar el nivel de salida del generador de ruido asociado al circuito silenciador. Este condensador (C3) se conecta entre el pin 3 (entrada del generador de ruido) y el +Vcc (pin 5). En ciertos casos, C3 puede omitirse.

• Un condensador para determinar la constante de tiempo del circuito pasabajos de salida del demodulador de FM. Este condensador (C4) se conecta entre el pin 4 (salida del demodulador) y el +Vcc (pin 5).

El circuito pasabajos anterior lo constituyen la resistencia de 13.6 KΩ interna y el condensador C4 externo. Su propósito es eliminar las armónicas de FI presentes a la salida del demodulador de FM.

• Un condensador de desacople de RF para la fuente de alimentación. Este condensador (C5) se conecta entre el pin 5 (+Vcc) y tierra (pin 16). Físicamente, debe estar lo más próximo al pin 5 del TDA7000 como sea posible.

• Un circuito sintonizado LC para establecer la banda de operación del oscilador local. Este circuito se conecta entre el pin 6 (entrada de control del LO) y el +Vcc (pin 5).

• Dos condensadores para establecer, en asocio con las resistencias internas de 2.2 KΩ, la frecuencia central o de corte del filtro de FI de dos etapas. Estos condensadores (C7 y C8) se conectan, en su orden, entre los pines 7 y 9 (salida del primer filtro de FI) y entre el 8 y el +Vcc (pin 5).

• Dos condensadores para establecer, en asocio con las resistencias internas de 4.7 KΩ, los límites inferior y superior de la banda de paso del filtro de FI de dos etapas. Estos condensadores (C10 y C11) se conectan, en su orden, entre los pines 10 y 11 (entrada del segundo filtro) y entre el 11 y el +Vcc.

• Un condensador para determinar, en asocio con la resistencia interna de 12 KΩ, el límite superior de la banda de paso de un circuito pasabajo previo al amplificador/limitador de FI. Este condensador (C12) se conecta entre el pin 12 (salida del filtro de FI de dos etapas) y el +Vcc (pin 5).

• Un circuito sintonizado LC para establecer la banda de recepción de RF deseada, es decir, la gama de frecuencias dentro de las cuales proporciona sintonía el receptor. Este circuito se conecta entre los pines 13 (entrada de RF del mezclador) y 14 (terminal de desacople de RF del mismo).

Nota: El rango de respuesta de frecuencia del TDA7000 se extiende desde 1.5 hasta 110 MHz.

Por lo tanto, incluye la banda comercial de FM estándar de América y Europa (88 a 108 MHz), la banda de FM estándar del Japón (76 a 91 MHz) y otras bandas populares de FM.

- Un condensador de desacople de RF para el mezclador. Este condensador (C14) se conecta entre el pin 14 y tierra (pin 16). Su función consiste en establecer una tierra dinámica para la señal de RF aplicada al mezclador sin afectar las condiciones de CC de la etapa.

- Un condensador para desacoplar la realimentación de CC del amplificador/limitador de FI. Este condensador (C15) se conecta entre el pin 15 y el +Vcc (pin 5).

- Dos condensadores para establecer la frecuencia intermedia (FI) de referencia del demodulador de FM y el detector de correlación. Estos condensadores se conectan de los pines 17 y 18 a tierra (pin 16), respectivamente.

Recuerde que el TDA7000 utiliza una FI interna de 70 KHz y no la estándar de 10.7 MHz.

El empleo de una FI tan baja permite lograr una integración total, proporciona un adecuado rechazo de frecuencias imagen y espúreas y garantiza una buena relación señal a ruido. Sin embargo, puede introducir distorsión.

La distorsión es particularmente crítica en el caso de señales de entrada fuertemente moduladas, es decir, con desviación máxima. Para obviar este inconveniente, el TDA7000 comprime la desviación pico de la señal de FI (±75 KHz) a ±15KHz.

Esta compresión del margen de desviación se logra desplazando la frecuencia del oscilador local mediante la señal de salida del demodulador de FM. La frecuencia del LO varía en proporción inversa a la desviación de frecuencia de la señal de FI.

El circuito de control así constituido se denomina **FLL** o bucle de enganche de frecuencia. FLL es la abreviatura en inglés de frequency-locked loop.

La acción del FLL en el TDA7000 no debe confundirse con la del CAF en un receptor de FM convencional. Este último mantiene constante la frecuencia del LO. El FLL, en cambio, la varía permanentemente, dependiendo del valor instantáneo de desviación de la señal de FI.

En la figura 497 se muestra el circuito completo de un receptor de FM para la banda doméstica de 88 a 108 MHz con TDA7000. El radio opera a partir de una fuente de 4.5 V, fácil de obtener conectando 3 pilas de 1.5 V en serie.

El receptor responde a niveles señales de entrada desde 1.5 µV hasta 200 mV, sin introducir distorsión ni ruido. El primer valor (1.5 µV) es su *umbral de sensibilidad*, es decir, la señal más débil que puede discernir sin ruido. La salida de audio es del orden de 70 mV (valor rms).

Se puede utilizar como amplificador de audio una o varias etapas a transistores o cualquiera de las configuraciones con circuitos integrados vistas en la Lección 38.

Las funciones de los condensadores C1 hasta C18 y del filtro pasabajos R2C2 son las mismas explicadas anteriormente. Con los valores de componentes suministrados, la FI es de 70 KHz y el ancho de banda es del orden de 92 KHz. La constante de tiempo R2C2 es de ≈40 µs.

El circuito de sintonía del oscilador local lo constituyen la bobina ajustable L1, el condensador varia-

Receptor completo de FM TDA7000

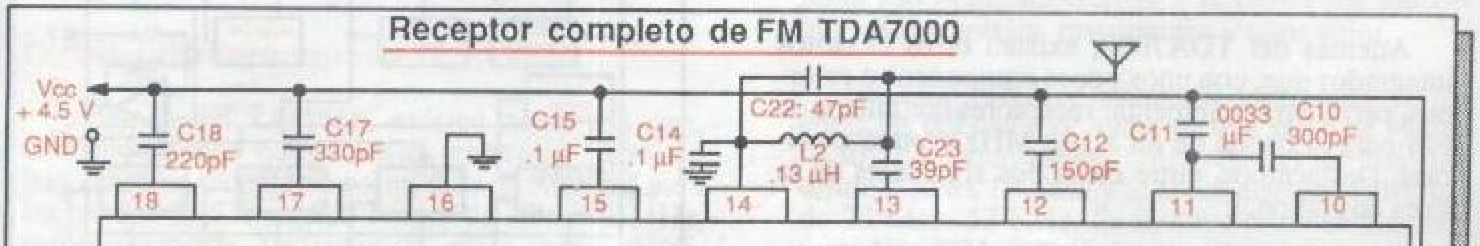

Fig. 497

ble C20 y los condensadores fijos C19 y C21. El circuito de sintonía de RF lo forman la bobina L2 y los condensadores C22 y C23. El único componente que requiere ajuste es la bobina L1.

Para la recepción de otras bandas de FM distintas a la estándar (88 a 108 MHz), deben modificarse los valores de L1 y L2. Por ejemplo, para la banda japonesa de 76 a 91 MHz, L1 debe ser de 0.078 µH y L2 de 0.15 µH.

En la figura 498 se muestra la forma de modificar el receptor de la figura 497 utilizando un diodo varactor como sintonizador. El varactor actúa como un condensador en paralelo con la bobina L1. Su capacitancia se controla mediante el voltaje inverso suministrado por el potenciómetro de 100 KΩ.

Sintonía por varactor del receptor FM

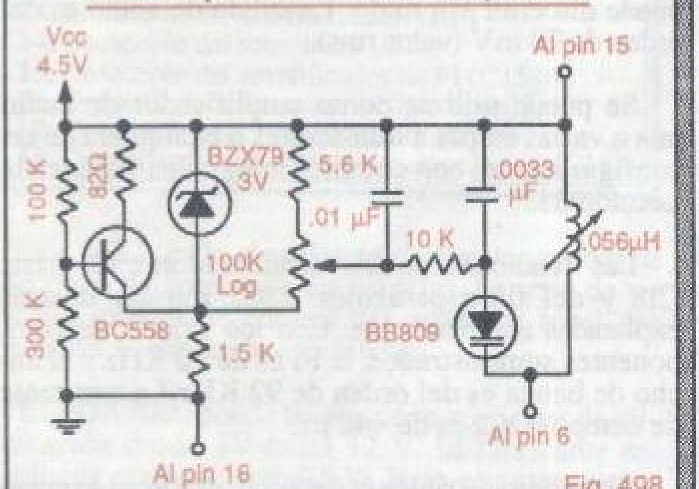

Fig. 498

Otros receptores integrados de FM

Además del TDA7000, existen otros circuitos integrados que, con unos pocos componentes externos, permiten implementar receptores completos de FM para la banda de 88 a 108 MHz y otros servicios. Destacamos, entre otros, los siguientes de la línea de National:

•**LM1865, 1965 y 2065.** Sistemas avanzados de FI para FM. Todos son de 20 pines. El LM1865 y el LM 2065 han sido diseñados para receptores de FM con sintonía electrónica (búsqueda automática de emisoras) y el LM1965 para receptores de sintonía manual. Voltaje de alimentación: 16V, máximo.

•**LM3089.** Sistema de FI para FM. 16 pines. Incluye amplificador de audio. Apropiado para autoradios, equipos de alta fidelidad y de comunicaciones y otros sistemas FM. Reemplaza el CA3089E de RCA. Voltaje de alimentación máximo: 16V.

•**LM3189:** Sistema de FI para FM superior al LM3089. 16 pines. Incluye amplificador de audio, amplificador/limitador de FI de 3 etapas, detector de FM doblemente balanceado, regulador de voltaje y otras funciones. Opera con voltajes entre 8.5 y 16 V. Apropiado para equipos de alta fidelidad.

•**LM3361A.** Sistema de FI para FM de banda estrecha. Incluye oscilador, mezclador, amplificador limitador de FI, demodulador y otras funciones. Similar al MC3361 de Motorola. Trabaja con menos de 2V. Voltaje de alimentación máximo: 12V.

•**TBA 120S.** Amplificador de FI y detector de FM. 14 pines. Adecuado para la detección de audio en receptores de TV y FM. Incluye amplificador de FI de 8 etapas con limitador, detector balanceado, preamplificador de audio y regulador de voltaje. Opera con voltajes entre 6 y 18 V.

Sistema receptor de AM y FM LM1868

El LM1868 es un sistema integrado de recepción de AM y FM que proporciona toda la circuitería necesaria para implementar un radio de dos bandas AM/FM de 0.5 W. En la figura 499 se muestra su distribución de pines y el diagrama interno de bloques.

El circuito integrado LM1868

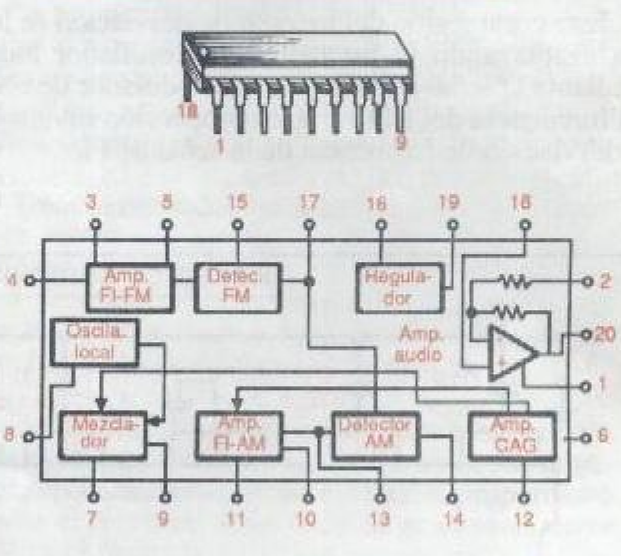

Fig. 499

El LM1868 puede operar a partir de una batería o una fuente de 9 V. Como puede observarse, incorpora un amplificador de audio y un regulador de voltaje. Estos bloques son comunes a los circuitos de AM y FM. El amplificador de audio puede manejar directamente un parlante de 8Ω.

La sección de FM consta de un amplificador de FI y un detector. La sección de AM consta de un oscilador local, un mezclador, un amplificador de FI y un amplificador de CAG. El oscilador y el mezclador configuran un conversor de AM.

El amplificador de CAG interno de AM gobierna la operación tanto del mezclador como del amplificador de FI. Las salidas de los detectores de AM y FM se superponen en el pin 17, que es la salida común de audio. La entrada del amplificador de audio es el pin 18.

En el modo AM, la señal de antena se acopla capacitivamente al pin **7** (entrada del mezclador). La frecuencia del oscilador local se programa mediante un circuito sintonizado LC conectado al pin **8** (entrada del LO).

La FI se selecciona mediante un circuito sintonizado conectado al pin **9** y se acopla inductivamente al pin **11** (entrada del amplificador de FI). Después de amplificada, la señal de FI se aplica al detector y se controla mediante una trampa LC conectada al pin **13** (entrada del detector).

Una vez detectada, la señal de audio se obtiene en el pin **17**. Esta señal se puede acoplar a un amplificador de audio externo o a la entrada del amplificador interno (pin **18**). Una vez amplificada, la señal de audio impulsa, desde el pin **20**, el parlante. Los demás pines AM (10, 12 y 14) son de control.

En el modo FM, la señal de FI proveniente de un mezclador o conversor de FM externo se aplica a la entrada del amplificador de FI y se detecta. La FI de referencia del detector de FM se establece mediante un circuito LC conectado al pin 15. Una vez detectada, la señal de audio se suministra al pin 17.

Otros receptores integrados de AM/FM.

Además del LM1868, existen otros CIs que, con unos pocos componentes externos, permiten implementar receptores completos de AM/FM para las bandas de 530 a 1600 KHz y de 88 a 108 MHz, respectivamente. Destacamos, entre otros, los siguientes de la línea de SGS-ATES:

·TDA1220A. Radio AM-FM. 16 pines. Diseñado para su uso en radios portátiles y domésticos de AM y FM. Trabaja con tensiones de alimentación desde 3 hasta 16 V.

La sección AM incluye oscilador local, mezclador con preamplificador, amplificador de FI con CAG, detector balanceado y preamplificador de audio. La sección de FM incluye amplificador de FI, detector de cuadratura y preamplificador de audio.

·TDA 1220B. Radio AM-FM de alta calidad. 16 pines. Versión mejorada del TDA1220A. Adecuado para producción en serie. Utiliza un mezclador doblemente balanceado en la sección AM. Cubre bandas de AM hasta 30 MHz. Compatible con FM estéreo. Opera con 3 a 12 V.

·TDA1220L. Radio AM-FM de bajo voltaje. 16 pines. Diseñado para su uso en radios portátiles que operan con baterías de 4.5 o 6V. El máximo voltaje de alimentación admisible es 12 V. La misma configuración interna del TDA1220B.

Fundamentos de la recepción de FM estéreo

Como vimos al final de la Lección 29 (página 232), en una emisora de FM estéreo la información de audio proveniente de los micrófonos izquierdo (L) y derecho (R) se procesa en un mezclador de audio. Este entrega una señal de audio suma (L+R) o monofónica y una señal diferencia (L-R) o estéreo.

La señal L+R modula directamente la portadora de 88 a 108 MHz como en el sistema FM convencional o monofónico. La señal L-R modula previamente en amplitud una subportadora de 38 KHz que se suprime durante el proceso de transmisión.

Las dos señales modulan la portadora principal de FM conjuntamente con un tono piloto de 19 KHz que facilita, en el receptor, la regeneración de la subportadora suprimida de 38 KHz. Esto permite demodular la señal estéreo L-R y recuperar las señales L y R originales.

Además de señales estéreo, las emisoras de FM pueden también transmitir sobre subportadoras de 53 a 75 KHz música de fondo, reportes de tiempo y otros programas privados. Este servicio se denomina FM *multiplex* o SCA y se presta por suscripción ($$) a hoteles, restaurantes y otros sitios.

En la figura 500 se muestra la distribución de frecuencias del canal de audio de una señal de FM estéreo. La señal monofónica (L+R) se extiende desde 0 hasta 15 KHz y la señal estéreo (L-R) desde 23 hasta 53 KHz. Se incluye también una subportadora SCA o multiplex de 67 KHz.

Las señales mono y estéreo modulan la portadora principal del canal en un 80% o 90%, dejando un 10% para el tono piloto de 19 KHz y un 10% para la señal multiplex o SCA, cuando se utiliza.

Un receptor de FM estéreo detecta la señal L-R e individualiza la información original de los canales izquierdo (L) y derecho (R) en dos parlantes. Los receptores monofónicos de FM, como el radio FM CEKIT, detectan la señal compuesta L+R pero no separan sus componentes L y R individuales.

Espectro de audio de la Señal FM estéreo

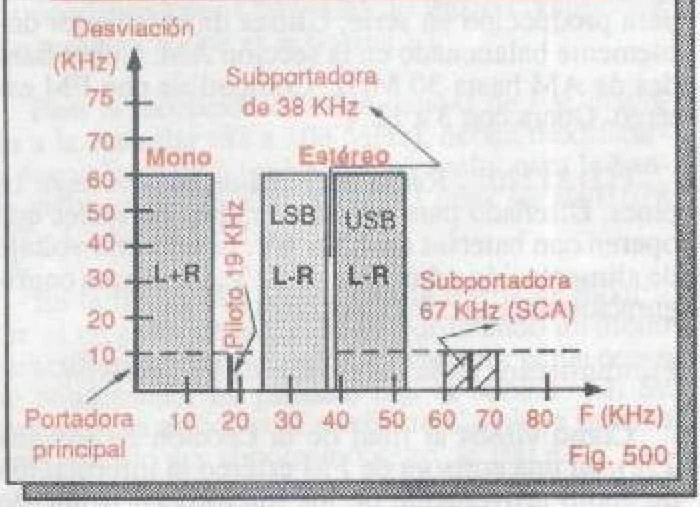

Fig. 500

Existen varias maneras de recuperar, en forma independiente, la información de los canales L y R de una señal de FM estéreo. Uno de los métodos más eficientes es utilizar un lazo de amarre de fase o PLL. Además, el PLL se presta a la integración.

Demodulador integrado de FM estéreo CA3090

La primera compañía en incorporar el principio PLL en un decodificador integrado de FM estéreo fue la RCA. En la figura 501 se muestra el diagrama interno de bloques del decodificador PLL de FM estéreo CA3090, precursor de toda una generación de dispositivos como el popular LM1800.

La regeneración de la subportadora de 38 KHz la realiza un circuito PLL constituido por un VCO (oscilador controlado por voltaje) de 76 KHz, una cadena de divisores de frecuencia que genera las señales requeridas por el proceso y un detector de fase de 19 KHz.

El voltaje de CC de salida de este último es proporcional a la diferencia de fase entre la salida del divisor de frecuencia de 19 KHz y la señal piloto incidente de 19 KHz. Para operar, el VCO requiere de una bobina externa entre los pines 15 y 16.

Se utiliza una frecuencia de VCO de 76 KHz para asegurar que la portadora reinsertada de 38 KHz sea perfectamente simétrica y la separación entre los canales de audio (L y R) sea nítida.

La salida del detector de fase de 19 KHz es cero cuando la frecuencia del VCO es correcta (76 KHz) o cuando no hay señal piloto de 19 KHz. Este último es el caso de emisiones de FM monofónicas. Para señalizar la presencia de una emisión de FM estéreo se utiliza un detector de 19 KHz adicional.

Los componentes externos conectados a los pines 6, 7 y 8 establecen el umbral de sensibilidad y la constante de tiempo de este detector de presencia de la señal piloto de 19 KHz.

La salida del divisor de frecuencia está en fase con la señal piloto y proporciona la señal de control de un conmutador electrónico de mono a estéreo del tipo *Schmitt trigger*.

Este interruptor Schmitt-trigger dispara el detector de la señal L-R, habilitando la recepción de programas estereofónicos, y excita la lámpara indicadora de estéreo.

La señal de salida del detector L-R se procesa en un circuito sumador o matriz con la señal estéreo de FM recibida y aplicada al pin 1. A la salida de la matriz se obtienen individualmente las señales de los canales izquierdo (L) y derecho (R). La señal estéreo de entrada es previamente amplificada.

Las señales L y R se amplifican por separado y se suministran a los pines 9 y 10 respectivamente. De estos puntos se envían a los amplificadores de potencia externos de cada canal. La separación entre canales es del orden de 40 dB y el rechazo a las señales SCA de 55 dB.

El circuito integrado CA3090

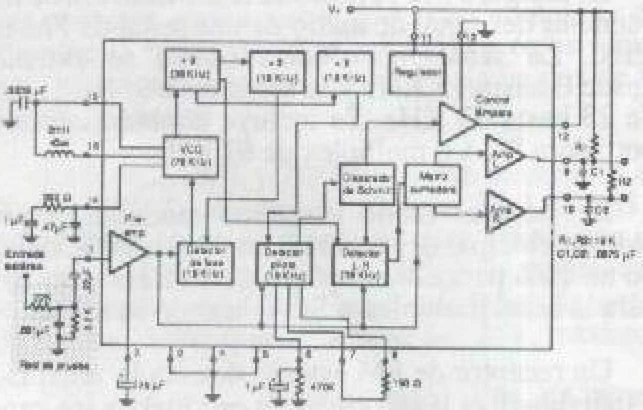

Fig. 501

Otros demoduladores de FM estéreo integrados

Además del CA3090, existen otros circuitos integrados que, con unos pocos componentes externos, permiten implementar receptores completos de FM estéreo para la banda de 88 a 108 MHz y otros servicios. Destacamos, entre ellos, los siguientes:

•**LM1800.** Demodulador de FM estéreo con PLL. Utiliza técnicas de PLL (lazo de amarre de fase) para regenerar la subportadora de 38 KHz. 16 pines. Voltaje de alimentación máximo: 18 V.

•**LM1870.** Demodulador de FM estéreo con reducción de ruido.Utiliza técnicas de PLL para regenerar la subportadora de 38 KHz y control por CC para reducir el ruido durante la recepción de señales débiles. Opera con voltajes de alimentación entre 7 y 15 V.

•**LM4500A.** Demodulador de FM estéreo de alta fidelidad con reducción de ruido. 16 pines. Utiliza una novedosa técnica de demodulación de FM que minimiza la interferencia entre emisoras adyacentes debida a los armónicos de la subportadora de 38 KHz. Máximo voltaje de alimentación: 16 V.

•**TEA1330.** Decodificador de FM estéreo. 16 pines. Utiliza técnicas de PLL. No requiere bobinas externas. Sólo necesita ajuste del VCO. Puede trabajar con voltajes entre 3 y 14 V.

•**MC1310P.** Decodificador de FM estéreo. 14 pines. No requiere bobinas. Utiliza técnicas de PLL. Sólo necesita ajuste del VCO. Separación entre canales de 40 dB. Rechazo de SCA de 80 dB. Trabaja con voltajes entre 8 y 16 V.

En la figura 502 se muestra el diagrama de bloques y un circuito típico de aplicación del CI LM 1800. El chip trabaja con tensiones hasta de 18 V y puede manejar directamente una lámpara indicadora de estéreo de 100 mA conectada al pin 7.

El LM18000 incorpora, en una cápsula de 16 pines, un regulador de voltaje, un preamplificador de audio, dos detectores de fase, un amplificador de baja frecuencia, un VCO de 76 KHz, una cadena de divisores de frecuencia, un conmutador electrónico de mono a estéreo y un decodificador de estéreo.

El amplificador de baja frecuencia, el VCO, el divisor de frecuencia y uno de los detectores de fase configuran un circuito PLL que produce la subportadora de 78 KHz. El otro detector de fase detecta la presencia del tono piloto de 19 KHz y controla la operación del conmutador electrónico.

La señal compuesta de audio, proveniente del detector de FM del receptor, se acopla mediante C6 a la entrada del preamplificador de audio (pin 1). Este último proporciona una señal de salida (V) para el decodificador y otra (W) para los detectores de fase del PLL y el detector de piloto.

La señal de audio V es accesible externamente desde el pin 2. La entrada de audio de los detectores de fase es el pin 12. El acople entre ambos puntos se realiza a través del condensador C7.

En el circuito PLL, la frecuencia de salida de 76 KHz del VCO se divide sucesivamente por 2 para obtener una señal de 38 y dos de 19 KHz. Una de las señales de 19 KHz se aplica a la otra entrada del detector de fase.

Esta señal de 19 KHz se puede monitorear en el pin 11 mediante un frecuencímetro (medidor de frecuencia) para ajustar exactamente a 76 KHz la fre-

El circuito integrado LM1800

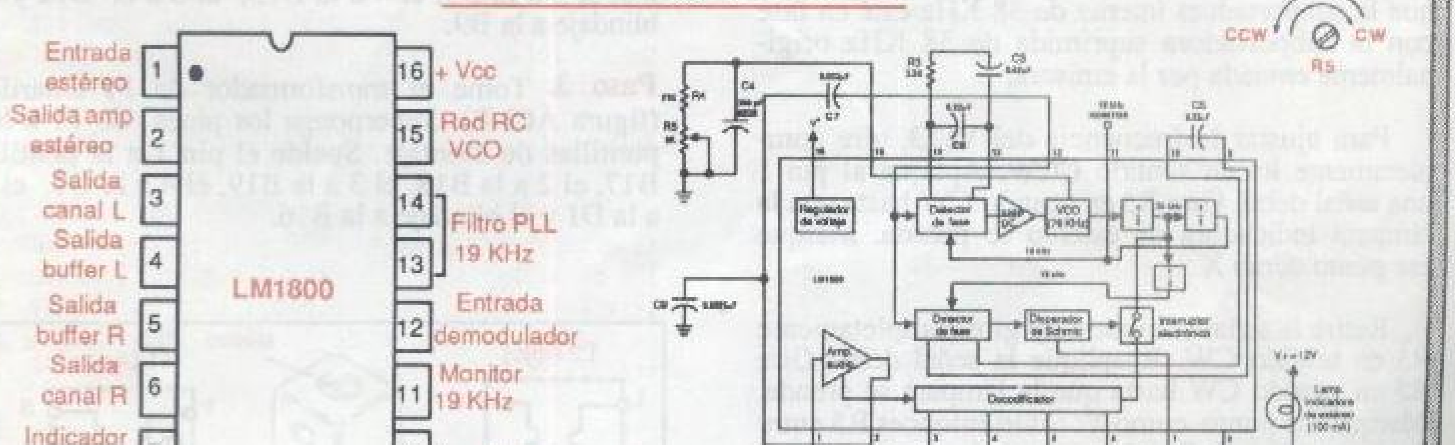
Entrada estéreo
Salida amp estéreo
Salida canal L
Salida buffer L
Salida buffer R
Salida canal R
Indicador estéreo
GND
1
2
3
4
5
6
7
8
16
15
14
13
12
11
10
9
+ Vcc
Red RC VCO
Filtro PLL 19 KHz
Entrada demodulador
Monitor 19 KHz
Filtro interruptor estéreo
LM1800
CCW
CW
R5
C4
R4
R5
C7
R3
C3
C5
C8
C9
Regulador de voltaje
Detector de fase
VCO (76 KHz)
LM1800
Detector de fase
Disparador de Schmit
Interruptor electrónico
Amp audio
Decodificador
V+ = 12V
Lamp. indicadora de estéreo (100 mA)
C6
Entrada estéreo
C1
R1
R2
C2
Salidas

Fig. 502

cuencia del VCO mediante R5. La frecuencia del VCO se puede también calibrar manualmente utilizando R5 y la lámpara indicadora de estéreo.

La salida del detector de fase del PLL mantiene la frecuencia del VCO constante en 76 KHz. La salida de 38 KHz del primer divisor de frecuencia se toma como subportadora regenerada para demodular la información L-R en el decodificador.

La señal amplificada de audio W y la otra señal de 19 KHz generada en el PLL manejan el detector de piloto. La salida de voltaje de este detector es proporcional a la amplitud del tono piloto de 19 KHz. Cuando este tono es suficientemente fuerte, el conmutador electrónico S se dispara.

Al dispararse, el conmutador prende la lámpara indicadora de estéreo y conecta la subportadora de 38 KHz al decodificador para permitir la demodulación de la señal L-R.

En el modo monofónico (S abierto) la salida del decodificador es la misma señal de audio de entrada aplicada al pin 1 y amplificada internamente. Los circuitos R1C1 y R2C2 son redes de deénfasis.

En el modo estéreo (S cerrado), el decodificador demodula la información L-R, la suma o *matriza* con la información L+R y produce como resultado dos señales separadas L y R que se suministran a los pines de salida 4 y 5.

El circuito R3-C3-C8 establece el rango de captura del PLL. C8 contribuye también a mantener la separación adecuada entre canales, garantizando que la información del uno no se mezcle con la del otro.

El condensador C9, conectado al pin 2, asegura que la subportadora interna de 38 KHz esté en fase con la subportadora suprimida de 38 KHz originalmente enviada por la emisora.

Para ajustar la frecuencia del VCO, gire completamente R5 en sentido CCW. Aplique al pin 1 una señal débil. Gire R5 en sentido CW hasta que la lámpara indicadora de estéreo se prenda. Marque ese punto como X.

Retire la señal compuesta y gire completamente R5 en sentido CW. Reaplique la señal débil. Gire R5 en sentido CW hasta que la lámpara se prenda. Marque ese punto como Y. Sitúe entonces R5 entre las marcas X e Y. Ese es su punto de ajuste.

El circuito anterior se puede adaptar fácilmente al radio FM CEKIT para convertirlo en un receptor de FM estéreo. Simplemente, inyecte la señal de audio al pin 1 del LM1800 y utilice dos amplificadores de audio separados para las señales L y R.

Actividad práctica Nº 22

Ensamble de la etapa osciladora

En esta actividad instalaremos en el tablero del radio AM CEKIT los transformadores L2 y T1 del amplificador/conversor de RF, la etapa marcada con color amarillo en la lámina correspondiente al diagrama esquemático del receptor.

Componentes necesarios

1 Bobina osciladora de núcleo rojo. L2
1 Transformador de FI de núcleo amarillo. T1.

Procedimiento

Paso 1. Con alambre realice las siguientes conexiones punto a punto y corte el excedente:

- Una las puntillas B11 y B13. Asegúrese de que este alambre no haga contacto la puntilla B12.

- Una las puntillas B12 y B14 y el extremo superior de R5. Asegúrese de que no haga contacto con el alambre que conecta las puntillas B13 y B11.

- Una las puntillas B18 y 44. Cerciórese de que este alambre no haga contacto con el alambre que conecta las puntillas B14 y B17. Realice una elevación en el punto de cruce para mayor seguridad.

-Una la puntilla B9 a la 5 y la B16 a la 6

Paso 2. Tome la bobina osciladora roja L2 (figura A62-A). Superponga los pines de L2 a las puntillas de montaje. Suelde el pin 1 a la puntilla B7, el 2 a la B8, el 3 a la B9, el 4 a la B13, el 6 a la B12 y el blindaje a la B9.

Paso 3. Tome el transformador de FI amarillo (figura A62-B). Superponga los pines de T1 a las puntillas de montaje. Suelde el pin 1 a la puntilla B17, el 2 a la B18, el 3 a la B19, el 4 a la D2, el 6 a la D1 y el blindaje a la B16.

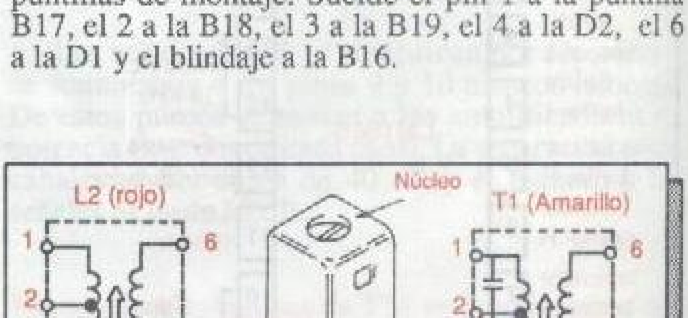

Fig. A62

Actividad práctica Nº 23

Ensamble de la etapa de antena

En esta actividad instalaremos en el tablero del receptor AM CEKIT la bobina de antena y el condensador de sintonía. Estos componentes completan el amplificador/conversor de RF, la etapa marcada con color amarillo en el esquema del radio.

Componentes y materiales necesarios

1 Bobina de antena de núcleo de ferrita con soportes plásticos. L1.
1 Condensador variable con perilla (dial), tornillo y cinta de montaje. CV1/CV2.

Procedimiento

Paso 1. Introduzca la bobina de antena en sus soportes y asegure el conjunto a la tabla con pegante sintético, como se indica (figura A63A). Recorte las prolongaciones de los soportes si no dispone de taladro o herramienta de perforación.

Paso 2. Tome el condensador variable y adhiéralo a la tabla con cinta de montaje, pegante sintético o lámina de auluminio como se indica (figura A63B). Instale la perilla y asegúrela mediante el tornillo. La perilla debe girar libremente, sin rozar la tabla.

Paso 3. Identifique los terminales de la bobina L1 y conéctelos así: el negro a la puntilla 1, el rojo a la B2, el azul a la B1 y el rosado al terminal "A" del condensador variable. Prolongue con cable aislado.

Paso 4. Identifique los terminales del condensador variable y conéctelos así: el oscilador ("O") a la puntilla B7, el común ("G") a la puntilla 1 y el de antena ("A") al rosado de L1. Utilice cable aislado.

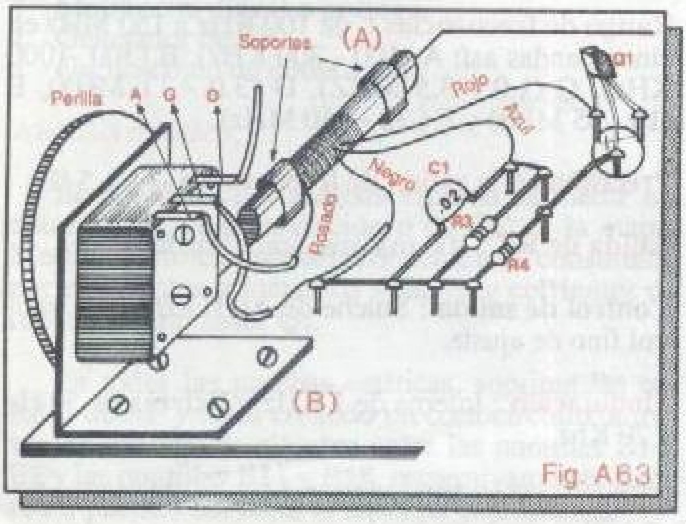

Fig. A63

Actividad práctica Nº 24

Funcionamiento y manejo de un Generador de RF

En la Lección 15 de este curso habíamos mencionado la importancia de algunos instrumentos de medida para la correcta calibración y ajuste de los diferentes aparatos electrónicos.

En el trabajo con circuitos de radio y comunicaciones en general, uno de los instrumentos más utilizados para su alineación y prueba es el generador de radiofrecuencia o RF.

Un generador de RF es un instrumento que entrega una señal de alta frecuencia con forma de onda sinusoidal.

Las principales características que debe tener son: frecuencia variable en un amplio rango, amplitud de la señal ajustable en la salida y posibilidad de modular, con audio, la señal de RF.

En la figura A64 tenemos un modelo común de generador de RF. A continuación vamos a estudiar su manejo y funcionamiento, que se puede asimilar a otros modelos con pequeñas variaciones.

Los números encerrados en los círculos de la figura indican los diferentes controles, que tienen las siguientes funciones:

1. Control de frecuencia. Con esta perilla debemos seleccionar la frecuencia de RF deseada teniendo en cuenta la posición del selector de bandas 2.

2. Selector de bandas. En cada posición selecciona una banda de frecuencias. Ejemplo: si está en la posición C el generador puede entregar entre 1 y 3.5 MHz, dependiendo de la posición del dial.

3. Borne para cristal. Permite conectar allí un cristal de cuarzo para obtener una señal muy precisa.

4. Suiche general. Se utiliza para encender y apagar el equipo.

5. Lámpara piloto. Se enciende cuando se oprime el suiche general.

6. Selector de modo. Permite seleccionar modulación externa, modulación interna de 1KHz o activar el generador con el cristal si está instalado en sus bornes.

7. Bornes para modulación externa. Cuando se selecciona el modo de modulación externa, se conecta en ellos el generador de audio.

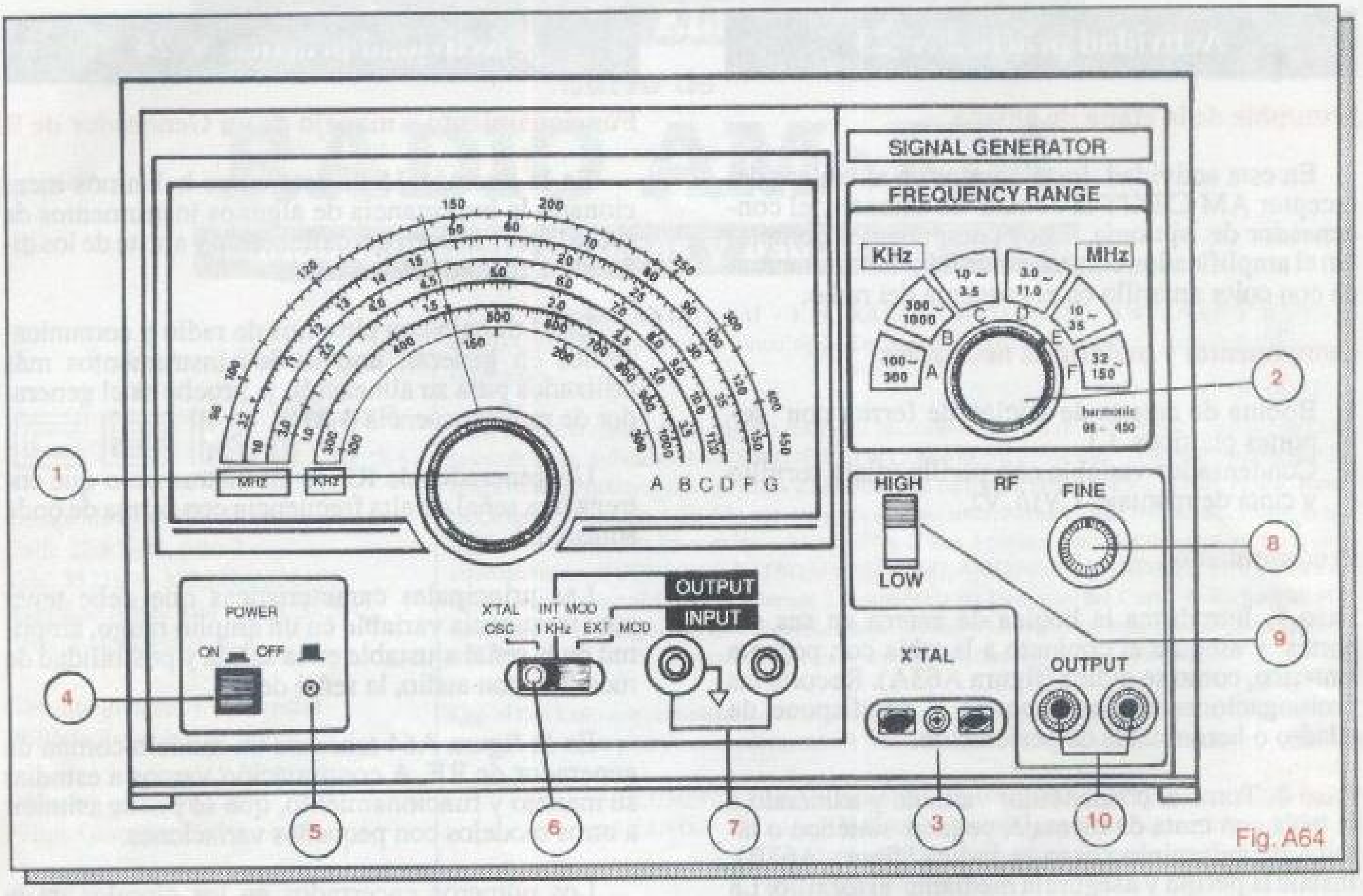
SIGNAL GENERATOR
FREQUENCY RANGE
KHz
MHz
100 ~ 300
300 ~ 1000
1.0 ~ 3.5
3.0 ~ 11.0
10 ~ 35
32 ~ 150
A
B
C
D
E
F
harmonic 96 ~ 450
MHz
KHz
A B C D F G
POWER
ON
OFF
X'TAL OSC
INT MOD 1 KHz
EXT MOD
OUTPUT
INPUT
HIGH
LOW
RF
FINE
X'TAL
OUTPUT
1
2
3
4
5
6
7
8
9
10

Fig. A64

8. Control de ajuste. Permite regular el nivel de salida de la señal.

9. Selector de rango de salida. Permite establecer dos niveles, alto (HIGH) y bajo (LOW), en la salida.

10. Terminales de salida. Por estos bornes sale la señal de RF del generador.

Operación

Para lograr una determinada señal de RF con este generador debemos proceder de la siguiente forma:

Paso 1. Antes de conectar el generador al circuito se debe ajustar su frecuencia y su nivel de salida. Para hacerlo, coloque el selector de bandas en el rango de frecuencias deseado, ya sea en KHz o en MHz.

Paso 2. Con el control de frecuencia obtenga la frecuencia buscada moviendo el dial hasta que la marque en la escala.

Paso 3. Con el control de ajuste y el selector de rango de salida lleve el nivel de la señal de salida al valor deseado. Este valor se puede medir con un multímetro digital o con un osciloscopio.

Se debe tener en cuenta si el valor medido es RMS, pico o pico a pico.

Paso 4. Con el selector de modo seleccione si la señal es modulada internamente, externamente o sin modulación. Con estos pasos ya está listo el generador para su utilización.

Para cambiar de señal, en otras pruebas, se deben repetir los pasos anteriores.

Características generales de este generador de RF

Rango de frecuencias : de 100 KHz a 150 MHz en cinco bandas así: A (100 - 300 KHz), B (300 -1000 KHz), C (1.0 - 3.5 MHZ), D (3.0 - 11 MHz), E (10 - 35 MHz) y F (32 - 150 MHz).

Precisión en la salida de frecuencia : + - 1.5 %

Salida de RF : 100 mVrms, aproximadamente.

Control de salida : Suiche de ALTO-BAJO y control fino de ajuste.

Modulación : Interna de 1 KHz y Externa de 50 Hz a 20 Khz.

Alimentación : 110 o 220 Voltios de CA.

Actividad práctica Nº 25

Análisis y prueba del conversor de RF

Objetivo

En esta actividad alinearemos el circuito amplificador/conversor de RF del receptor de AM CEKIT y analizaremos su comportamiento, tanto en condiciones estáticas como dinámicas. Al final de esta actividad su radio quedará plenamente calibrado y trabajando en condiciones óptimas.

Las pruebas que se realizaran son las siguientes:

Estáticas o de CC (sin señal)

- Corriente total consumida por el radio.
- Voltajes de polarización de Q1.
- Corriente de reposo de colector de Q1.

Dinámicas o de CA (con señal)

- Alineamiento del transformador T1.
- Visualización de la señal del oscilador local.
- Alineamiento del LO (TC2, L2).
- Alineamiento del circuito de antena (TC1, L1).

Para la realización de estas pruebas necesitará de un multímetro, un osciloscopio y un generador de RF como equipo mínimo de prueba. Se supone que usted ya está familiarizado con el uso de estos instrumentos y con las técnicas de medición de corriente, voltaje, resistencia y frecuencia.

Si usted no posee osciloscopio ni generador de RF, en la próxima actividad le explicaremos cómo calibrar su radio sin instrumentos, utilizando otro radio de AM en servicio como modelo.

Equipo de prueba necesario

1 Multímetro digital o análogo.
1 Generador de RF.
1 Osciloscopio de doble trazo.

Análisis estático

El propósito del análisis estático es medir las condiciones de polarización o de CC de la etapa. Mediremos inicialmente la corriente total consumida por el receptor y luego los voltajes y corrientes de reposo del transistor Q1.

En todas las pruebas estáticas, suprima las señales de RF y local creando un cortocircuito, a través de puentes de alambre entre las puntillas B1 y B2 y las puntillas B11 y B14, respectivamente. Retire los puentes del corto al terminar este análisis.

Medición de la corriente total consumida

Programe el multímetro como miliamperímetro de CC en la escala de 30 mA. Conéctelo en serie a la batería con la polaridad apropiada. Encienda el interruptor y registre la corriente medida. Retire el medidor y restablezca la conexión original.

I_T = ______________ mA

Una corriente excesiva puede indicar la presencia de un cortocircuito en la etapa conversora. Si este es el caso, desconecte la batería y revise todas las conexiones para detectar posibles cortos. Realice pruebas de continuidad con el óhmetro para detectar la causa de la falla.

Medición de las condiciones de polarización

En esta prueba mediremos los voltajes de polarización del colector, la base y el emisor de Q1 así como la corriente de reposo del colector. Utilizaremos un multímetro, digital o análogo, como instrumento de prueba.

Paso 1. Programe el multímetro como voltímetro de CC en la escala de 12V. Envíe el terminal negativo a tierra y toque sucesivamente con el terminal positivo el colector (C), el emisor (E) y la base (B) de Q1.

Registre los valores obtenidos en la Tabla 1. Retire el voltímetro.

V_C	V_E	V_B	V_{BE}	
				V

Tabla 1

Paso 2. Programe el multímetro como miliamperímetro de CC en la escala de 3 mA. Conéctelo en serie a colector de Q1 con la polaridad apropiada. Registre la corriente medida. Retire el medidor y restablezca la conexión original.

I_{CQ} = __________ mA

La corriente de colector se puede también evaluar de un modo "indirecto" midiendo la caída de voltaje sobre la resistencia R2 (1KΩ) y aplicando la Ley de Ohm. Es decir:

$I_{CQ} = V_{R2}/R_2 = V_{R2}/1K\Omega =$ ________ mA

Análisis dinámico

Durante el análisis dinámico alinearemos el primer transformador de FI (T1) y observaremos la señal del oscilador local. Veremos como CV2, CT2 y L2 afectan la frecuencia de esta última, sin modificar sustancialmente su amplitud.

Posteriormente, alinearemos CV2, CT2 y L2 para establecer la banda de operación del oscilador local entre 985 y 2055 KHz, es decir, 455 KHz por encima de la banda de recepción de AM (530 a 1600 KHz).

Finalmente, alinearemos los componentes del circuito de antena (CV1, CT1 y L1) para establecer la banda de sintonía del receptor entre 530 y 1600 KHz.

Las pruebas de analisis dinámico descritas requieren el uso de un osciloscopio y un generador de RF. En la próxima actividad veremos cómo puede alinearse el convertidor de RF prescindiendo de estos instrumentos y utilizando un receptor de AM en servicio.

Alineamiento del transformador T1

En esta prueba alinearemos el transformador T1 operando el transitor Q1 como amplificador de RF. Inhibiremos el efecto del oscilador local e inyectaremos una señal de 455 KHz, proveniente de un generador de RF, al circuito de antena.

Ajustaremos entonces el núcleo de T1 hasta conseguir la máxima ganancia de la etapa conversora. Posteriormente, realinearemos T2 y T3 hasta lograr la máxima ganancia de señal desde el conversor hasta el detector.

Paso 1. Inhabilite el oscilador local desoldando o separando el condensador C2 de la puntilla B6. Esta operación impide que la señal de realimentación local se inyecte al emisor de Q1 durante la prueba. Puede utilizar un cortafríos.

Paso 2. Inhabilite el circuito de antena desoldando el alambre rosado de la bobina L1 del terminal "A" del condensador variable. Esta operación impide la recepción de señales externas durante la prueba.

Paso 3. Programe el generador de RF para producir una señal modulada de 455 KHz y amplitud máxima, por ejemplo 500 mVpp.

Paso 4. Acople inductivamente la señal del generador de RF a la entrada del conversor conectando el caimán negro a cualquier puntilla de tierra del radio y el rojo al alambre rosado (libre) de L1, como se muestra en la figura A65.

Con la batería conectada y el control de volumen al máximo, se deberá escuchar el tono de modulación en el parlante o el audífono, aunque sea levemente. Después de verificar esto, elimine la modulación interna y deje sólo la onda continua de 455 KHz.

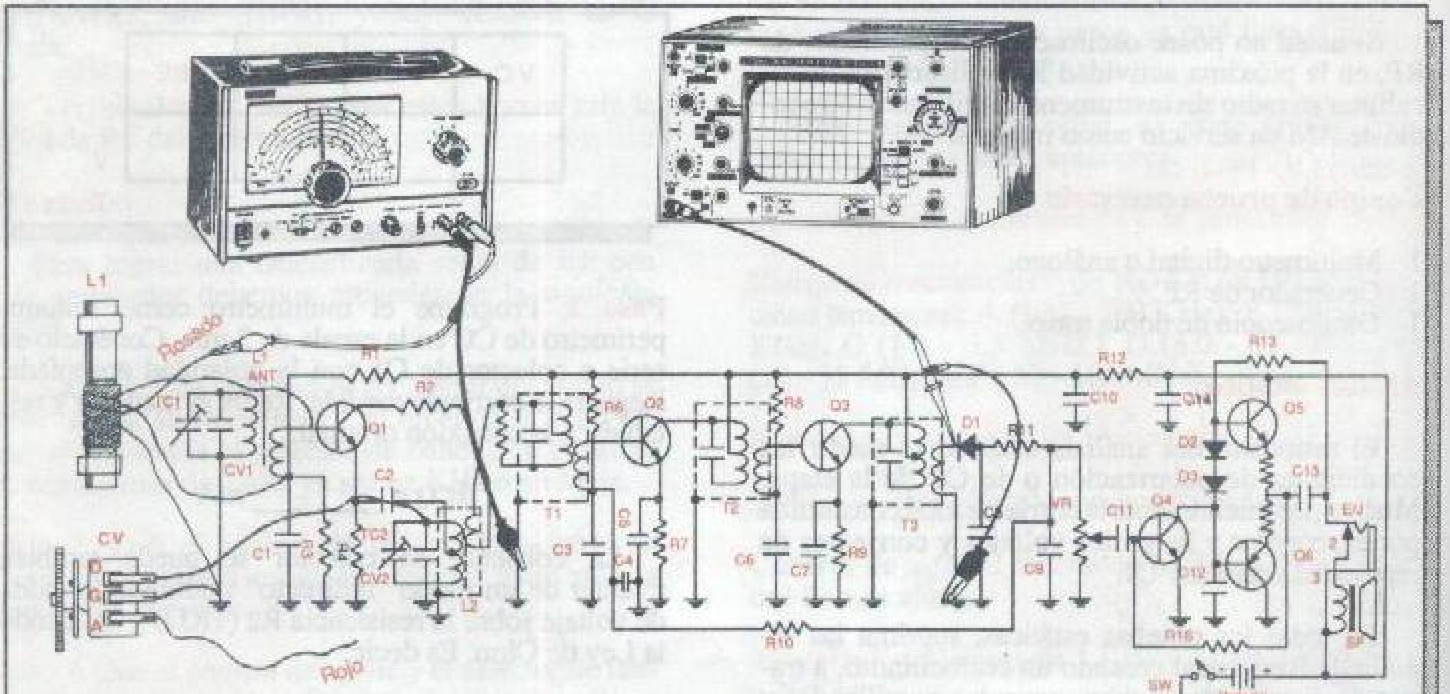

Fig. A65

Paso 5. Sitúe el control de atenuación vertical del osciloscopio en 0.01 V/div y la base de tiempo en cualquier rango de 1 a 5 µs. Conecte la sonda al cátodo del diodo detector D1 (puntilla C1) y el caimán negro a cualquier puntilla de tierra (figura A65).

Paso 6. Con el generador de RF sin modulación interna, ajuste lentamente el tornillo del transformador T1 (amarillo) hasta obtener la máxima señal en el osciloscopio. Utilice un destornillador plástico.

Paso 7. Reajuste ligeramente, en su orden, los transformadores T2, T3 y T1 hasta obtener, con cada uno, la máxima señal en el osciloscopio.

Paso 8. Repita el paso **7** varias veces, hasta convencerse de que la señal visualizada en el osciloscopio es, en definitiva, la máxima posible.

Una vez terminada esta prueba, los tres transformadores de FI (T1, T2 y T3) estarán perfectamente alineados a 455 KHz. La ganancia del conversor y de las etapas de FI será, bajo estas condiciones, máxima.

Paso 9. Conecte nuevamente la modulación interna del generador de RF. Notará que el tono en el parlante o el audífono se escucha ahora con mayor intensidad. Si lo desea, inmovilice los núcleos con resina epóxica, parafina derretida o similar para prevenir desajustes.

Paso 10. Apague el radio. Reconecte el condensador C2 a la puntilla B6 y el alambre rosado de la bobina L1 al terminal "A" del condensador variable. Desconecte los instrumentos del radio.

Encienda el radio y sitúe el control de volumen al máximo o en una posición intermedia. Mueva lentamente el dial del condensador variable. Notará que se sintonizan varias emisoras de su localidad.

Visualización de la señal del oscilador local

En esta prueba observaremos en el osciloscopio la señal del oscilador local. Comprobaremos cómo L2, CV2 y CT2 influyen en la frecuencia de la misma. Realizaremos también algunas medidas previas de voltaje y frecuencia en los extremos de la escala. La alineación definitiva del circuito se realiza en la prueba siguiente.

Paso 1. Encienda su radio, sitúe el control de volumen en una posición intermedia y gire completamente la perilla de sintonía (CV2) en sentido CCW (contrario a las manecillas del reloj).

Paso 2. Programe el osciloscopio para una atenuación vertical de 0.1 V/div y ajuste la base de tiempo en 0.5 ó 1 µs/div. Conecte la sonda a la puntilla B6 y el caimán negro a cualquier puntilla de tierra.

Observe la señal del oscilador local en el osciloscopio. Deberá ser una onda prácticamente sinusoidal y sin distorsión, de amplitud uniforme y con una determinada frecuencia. Registre los valores de amplitud y frecuencia obtenidos.

Aos = ________ mVpp

Fos (mín) = ________ KHz

Paso 3. Gire lentamente la perilla de sintonía en dirección CW (el mismo de las manecillas del reloj) y observe las formas como cambia (aumenta) la frecuencia de la señal observada a medida que avanza el dial. A intervalos regulares escuchará algunas emisoras de AM de su localidad.

Notará también que la amplitud de la señal local se mantiene prácticamente constante. Registre los valores de amplitud y frecuencia obtenidos cuando el dial llega a su tope máximo.

Aos = ________ mVpp

Fos (mín) = ________ KHz

Paso 4. Sitúe el dial en una posición intermedia y observe el efecto del condensador trimmer CT2 en la frecuencia del oscilador local.

Gire el tornillo de ajuste de este condensador (figura A66) mientras observa en el osciloscopio la señal del LO. Utilice un destornillador plástico. Derive sus propias conclusiones.

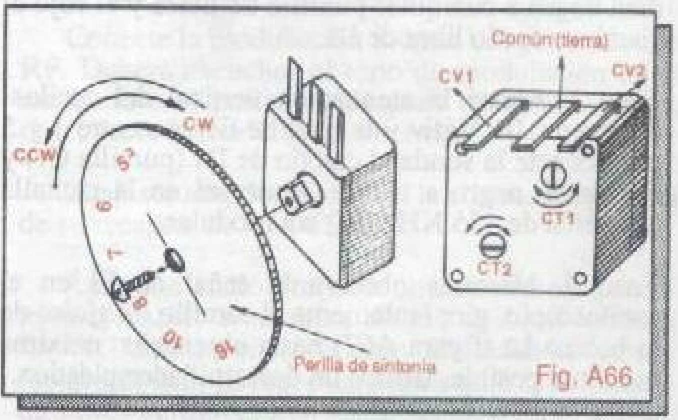

Fig. A66

Paso 5. Con el dial y el tornillo de ajuste de CT2 en posiciones intermedias, observe el efecto de la bobina L2 en la frecuencia del oscilador local.

Gire el tornillo de ajuste de esta bobina (figura A67) en una u otra dirección mientras observa en

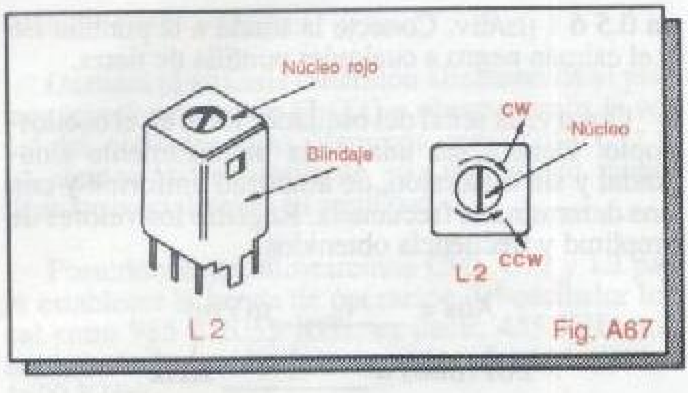

Fig. A67

el osciloscopio la señal del LO. Utilice un destornillador plástico. Derive sus propias conclusiones.

Alineamiento del oscilador local. Método 1

En esta prueba alinearemos el oscilador local mediante el generador de RF. Acoplaremos la señal del generador a la base de Q1 por inducción, inyectándola al primario de L1. Utilizaremos una señal de RF relativamente fuerte para "enmascarar" las señales de AM que convergen a la antena.

Estableceremos el límite inferior de frecuencia del LO calibrando el núcleo ajustable de la bobina L2 y el límite superior calibrando el tornillo de ajuste del trimmer CT2.

Paso 1. Desconecte el alambre rosado de la bobina L1 del terminal "A" del condensador variable. Gire completamente la perilla de sintonía en sentido CCW (antihorario). Prenda su radio y sitúe el control de volumen en una posición intermedia.

Paso 2. Programe el generador de RF para suministrar una señal sin modulación de 530 KHz y amplitud máxima, digamos 500 mVpp. Conecte el caimán negro a cualquier puntilla de tierra y el rojo al alambre rosado libre de L1.

Paso 3. Ajuste la atenuación vertical del osciloscopio en 0.01 V/div y la base de tiempo entre 1 y 5 µs. Conecte la sonda al cátodo de D1 (puntilla C1) y el caimán negro a tierra. Observará en la pantalla una señal de 455 KHz (FI) sin modular.

Paso 4. Mientras observa la señal de FI en el osciloscopio, gire lentamente el tornillo de ajuste de la bobina L2 (figura A67) hasta obtener la máxima amplitud posible. Utilice un destornillador plástico.

Conecte la modulación interna del generador de RF. Deberá escuchar el tono de modulación en el audífono o el parlante. Después de verificar esto, desconecte la modulación interna.

Paso 5. Gire completamente la perilla de sintonía en sentido CW (horario).Programe el generador de RF para suministrar una señal sin modulación de 1600 KHz. Mantenga conectado el osciloscopio a la puntilla C1.

Paso 6. Mientras observa la señal de FI en el osciloscopio, gire lentamente el tornillo de ajuste del trimmer CT2 (figura A66) hasta obtener la máxima amplitud posible. Utilice un destornillador plástico.

Conecte la modulación interna del generador de RF. Deberá escuchar el tono de modulación en el audífono o el parlante. Después de verificar esto, desconecte la modulación interna. El ajuste anterior completa la alineación del oscilador local.

Apague el radio y retire los instrumentos. Conecte nuevamente el alambre rosado de L1 al terminal "A" del condensador variable. Encienda el radio y recorra el dial. Note cómo ha mejorado la sintonía.

Alineación del oscilador local. Método 2

En esta prueba alinearemos el oscilador local mediante el generador de RF. Acoplaremos la señal del generador a la base de Q1 capacitivamente, es decir mediante un condensador.

Desconectaremos el circuito de antena para minimizar la influencia de señales externas. Estableceremos los límites inferior y superior de frecuencia del LO de la misma forma anterior.

Paso 1. Desconecte el alambre rosado de la bobina L1 del terminal "A" del condensador variable y el alambre rojo de la base del transistor Q1. Gire completamente la perilla de sintonía en sentido CCW (antihorario).

Prenda su radio y sitúe el control de volumen en una posición intermedia.

Paso 2. Programe el generador de RF para suministrar una señal sin modulación de 530 KHz y el osciloscopio con una atenuación de 0.005 V/div y una base de tiempo entre 1 y 5 µs.

Conecte el caimán negro del generador de RF a tierra y el rojo a la base de Q1 a través de un condensador de desacople de 0.01 µF. Con el mismo terminal de este condensador, realice un puente temporal entre las puntillas B2 y B1 (figura A68).

Conecte la sonda del osciloscopio a la base de Q1 (puntilla B2) y el caimán negro a tierra. Ajuste el control de amplitud del generador de RF hasta visualizar en el osciloscopio una señal de 10 mVpp.

Después de verificar lo anterior, desconecte la sonda de la base de Q1. Los demás pasos, a partir del 3 son los mismos del método 1.

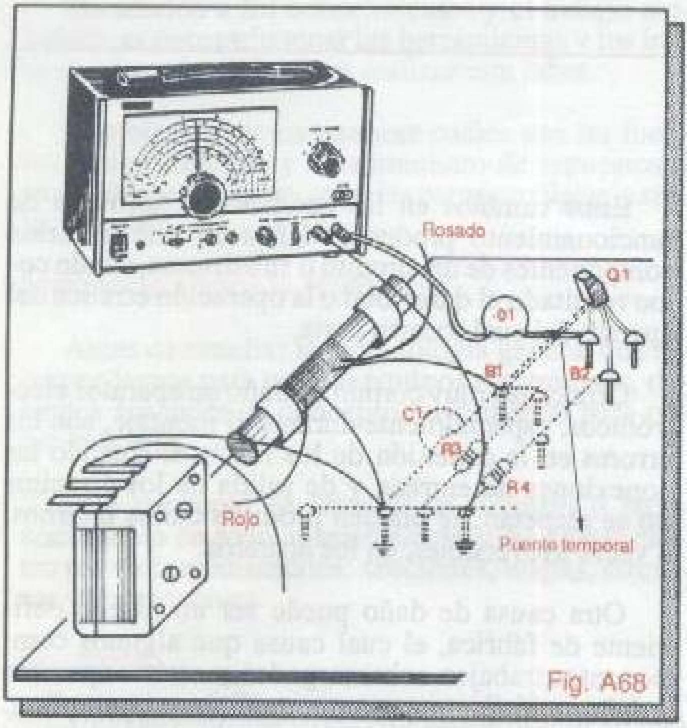

Fig. A68

Al concluir el alineamiento, apague el radio y retire los instrumentos. Elimine el puente temporal entre las puntillas B1 y B2. Conecte nuevamente el alambre rosado de L1 al terminal "A" del condensador variable y el rojo a la base de Q1.

Alineamiento del circuito de antena

En esta prueba alinearemos el circuito de antena mediante el generador de RF. Acoplaremos la señal del generador a la base de Q1 por inducción, inyectándola al primario de L1. Utilizaremos una señal de RF relativamente fuerte para "enmascarar" las señales de AM que convergen a la antena.

Estableceremos el límite inferior de frecuencia de la banda de sintonía (530 KHz) encontrando la posición óptima de L1 en su núcleo y el límite superior calibrando el tornillo de ajuste del trimmer CT1.

Paso 1. Desconecte el alambre rosado de la bobina L1 del terminal "A" del condensador variable. Gire completamente la perilla de sintonía en sentido CCW (antihorario). Prenda su radio y sitúe el control de volumen en una posición intermedia.

Paso 2. Programe el generador de RF para suministrar una señal sin modulación de 530 KHz y amplitud máxima, digamos 500 mVpp. Conecte el caimán negro a cualquier puntilla de tierra y el rojo al alambre rosado libre de L1.

Paso 3. Ajuste la atenuación vertical del osciloscopio en 0.01 V/div y la base de tiempo entre 1 y 5 µs. Conecte la sonda al cátodo de D1 (puntilla C1) y el caimán negro a tierra. Observará en la pantalla una señal de 455 KHz (FI) sin modular.

Paso 4. Mientras observa la señal de FI en el osciloscopio, desplace lentamente la bobina L1 a lo largo del núcleo de ferrita (figura A69) hasta obtener la máxima amplitud posible. Preserve el punto de ajuste inmovilizando la bobina L1 al núcleo con pegante epóxico o parafina derretida.

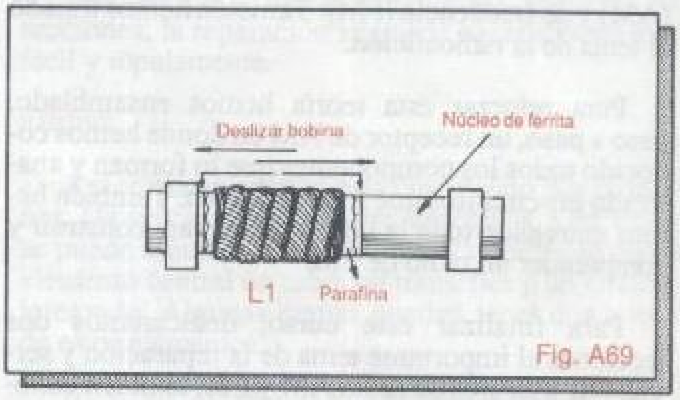

Fig. A69

Conecte la modulación interna del generador de RF. Deberá escuchar el tono de modulación en el audífono o el parlante. Después de verificar esto, desconecte la modulación interna.

Paso 5. Gire completamente la perilla de sintonía en sentido CW (horario).Programe el generador de RF para suministrar una señal sin modulación de 1600 KHz. Conecte el osciloscopio a la puntilla C1.

Paso 6. Mientras observa la señal de FI en el osciloscopio, gire lentamente el tornillo de ajuste del trimmer CT1 (figura A66) hasta obtener la máxima amplitud posible. Utilice un destornillador plástico.

Conecte la modulación interna del generador de RF. Deberá escuchar el tono de modulación en el audífono o el parlante. Después de verificar esto, desconecte la modulación interna.

El ajuste anterior completa la alineación general de su receptor de AM CEKIT .

Apague el radio y retire los instrumentos. Conecte nuevamente el alambre rosado de L1 al terminal "A" del condensador variable. Encienda el radio y recorra el dial. El receptor debe captar las emisoras de AM que transmiten en la banda de 530 a 1600 KHz de su localidad.

En la próxima actividad realizaremos el alineamiento definitivo de esta etapa utilizando un receptor de AM en servicio. Si el radio no le funciona apropiadamente, la Lección 40 le enseñará un método para localizar el origen de la falla.

Reparación de radios AM

Introducción

En las lecciones anteriores hemos estudiado la teoría básica de las comunicaciones por radio a través de los sistemas de modulación de amplitud (AM) y de frecuencia (FM). También hemos tratado el tema de la radioafición.

Para reforzar esta teoría hemos ensamblado, paso a paso, un receptor de AM en donde hemos conocido todos los componentes que lo forman y analizado los circuitos que lo constituyen. También hemos entregado toda la información para construir y comprender un radio de FM.

Para finalizar este curso, dedicaremos dos lecciones al importante tema de la reparación y servicio de aparatos de radio AM y FM, también conocido como *troubleshooting* en el argot técnico.

Partiendo de la base de que usted ya posee un bagaje importante de conocimientos teóricos y prácticos, adquiridos a lo largo de las distintas lecciones y actividades de este curso, en las Lecciones 40 y 41 lo orientaremos sobre una metodología a seguir en la repararación éxitosa de aparatos de radio.

Esta metodología es de aplicación general y por lo tanto se puede derivar hacia la reparación de otros aparatos electrónicos.

En la Lección 40 trataremos inicialmente el tema en forma general y luego aplicaremos la metodología de reparación al caso específico del radio AM CEKIT que hemos ensamblado en este curso. En la Lección 41 haremos lo propio con el radio FM CEKIT.

Reparación de equipos electrónicos

La reparación de equipos electrónicos es una técnica que ha estado ligada al desarrollo de la electrónica misma desde sus comienzos. La razón de su existencia se debe al hecho de que los componentes electrónicos no son completamente inmunes a condiciones anormales de funcionamiento y se pueden dañar en cualquier momento.

Aunque la gran mayoría de los circuitos electrónicos se diseñan para que puedan soportar cambios bruscos en sus parámetros eléctricos, todo aparato está expuesto a sobrevoltajes, sobrecorrientes o cortocircuitos, así como a variaciones bruscas de temperatura, golpes e impactos mecánicos.

Estos cambios en las condiciones normales de funcionamiento producen daños en uno o varios componentes de un circuito o subsistema, dando como resultado el daño total o la operación errática del aparato del cual forman parte.

Otra causa muy común de daño en aparatos electrónicos, especialmente durante el montaje, son los errores en la conexión de los mismos. Cuando las conexiones de entrada y de salida de los circuitos no se respetan, se pueden producir daños internos, a veces irreversibles, en los aparatos.

Otra causa de daño puede ser un diseño deficiente de fábrica, el cual causa que algunos componentes trabajen sobrecargados y estén expuestos a dañarse fácilmente a corto, mediano o largo plazo.

Además, ciertos componentes electrónicos tienen una vida útil determinada, al cabo de la cual ya no funcionan correctamente o dejan de funcionar totalmente, causando daños en el aparato en que están conectados.

Esta es sólo una parte de las múltiples causas por las que se pueden dañar los aparatos electrónicos y con las cuales usted se irá familiarizando a medida que tenga experiencia en la reparación de ellos.

La reparación de aparatos electrónicos es una profesión muy difundida en todo el mundo debido a la gran cantidad de éstos que hay instalados. Cada vez más, el hombre utilizará un mayor número de aparatos electrónicos, lo que incrementa la necesidad de técnicos especializados en su reparación.

La profesión de técnico se puede convertir en una actividad muy lucrativa. Existen, incluso, empresas completas dedicadas exclusivamente al mantenimiento y servicio de equipos electrónicos.

Las cualidades más importantes que debe reunir un técnico en electrónica son: ser estudioso, estando siempre al día en las tecnologías del momento, ser muy metódico, realizando su trabajo con conocimiento y ser muy organizado, planeando muy bien sus actividades.

En la reparación de equipos electrónicos, además de tener buenos conocimientos teóricos y prácticos, es necesario adoptar un método eficiente y seguro que nos garantice efectuar con éxito la gran mayoría de los trabajos que emprendamos.

En adición a los conocimientos y el trabajo metódico, es necesario tener las herramientas y los instrumentos adecuados para realizar esta labor.

También debemos conocer cuáles son las fuentes de información y de suministro de repuestos o partes de reemplazo necesarias para completar las reparaciones.

Metodología para la reparación

Antes de estudiar la metodología general que recomendamos para reparar equipos electrónicos, debemos aprender un concepto fundamental para poder aplicarla.

Este concepto se trata de la división o reconocimiento en todo aparato electrónico de sus cuatro partes fundamentales: secciones, etapas, circuitos y componentes.

Secciones

Todo aparato electrónico se puede dividir en secciones. Una sección es la parte de un aparato que realiza una función específica diferente a las demás de acuerdo a las señales presentes en ella. Esta es la mínima división que podemos hacer y la debemos tener claramente definida antes de iniciar la labor de reparación.

En la figura 503 tenemos como ejemplo la división de un radio AM en sus correspondientes secciones.

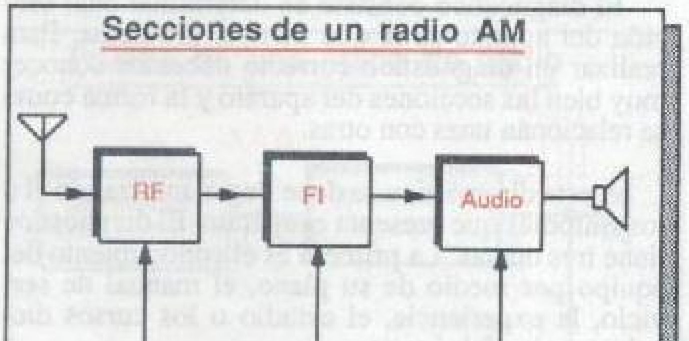

Fig. 503

Para conocer estas secciones lo primero que debemos tener o conseguir es el manual de servicio o el diagrama esquemático o plano del aparato que vamos a reparar. En ellos encontraremos la información necesaria para establecer esta división.

Si no tenemos esta información debemos elaborarla dibujando el diagrama del aparato a partir del circuito real, lo que es un poco difícil pero no imposible. El tiempo empleado en esta labor nos ahorra el tiempo gastado en reparar un equipo sin información y que puede ser mucho mayor.

Una vez que se han comprendido claramente las secciones, la reparación se puede hacer mucho más fácil y rápidamente.

Etapas

Así como un aparato está compuesto por secciones, las secciones se dividen en etapas. Una etapa se puede identificar claramente, ya que tiene como elemento central un tubo, un transistor o un circuito integrado. Algunas etapas pueden tener dos o más de estos elementos.

En la figura 504 podemos observar la división de las secciones del radio AM en sus etapas correspondientes. Todas las etapas de una sección trabajan con el mismo tipo de señal.

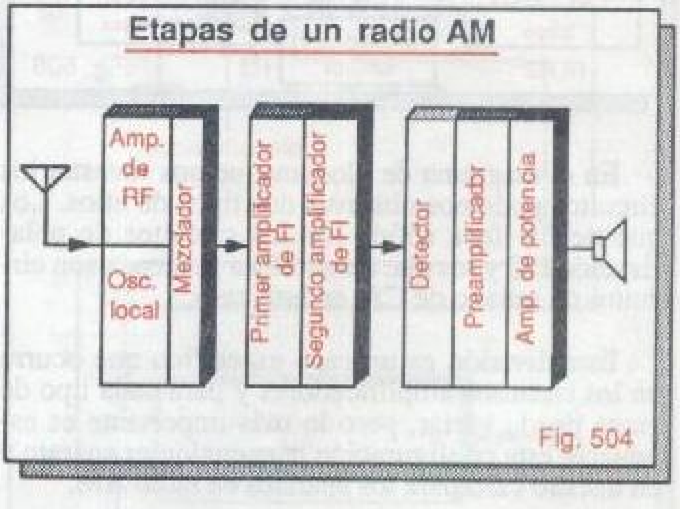

Fig. 504

Hay un concepto muy importante que podemos observar en la etapa mezcladora de la sección de RF y en la etapa detectora de la seción del amplificador de audio.

Estas etapas se utilizan para unir dos secciones y generalmente convierten un tipo de señal en otro. Estas etapas reciben el nombre de *etapas acopladoras* o de interfase.

Estas etapas forman parte de dos secciones, su entrada es parte de la sección anterior y su salida es parte de la sección siguiente.

Circuitos

La siguiente división que debemos hacer es la de los *circuitos*. Comúnmente se ha llamado a una etapa un circuito pero esto es equivocado. Los circuitos son las diferentes partes de una etapa.

Para entender mejor este concepto vamos a analizar un ejemplo real utilizando una etapa con transistor como la que se muestra en la figura 505.

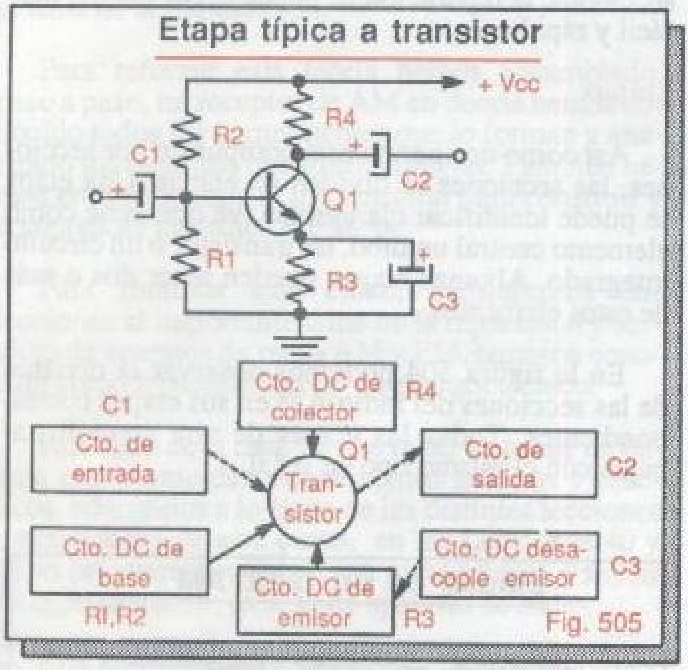

Fig. 505

En el diagrama de bloques que nos muestra los circuitos podemos observar dos tipos de ellos. Los que tienen línea sólida son los circuitos de polarización DC y los que tienen línea punteada son circuitos de señal o de CA, en este caso.

Esta división es un caso específico que ocurre en los circuitos amplificadores y para cada tipo de etapa puede variar, pero lo más importante es establecer esta configuración para cualquier aparato y en nuestro caso para los aparatos de radio AM.

En la figura anterior podemos observar que hay tres circuitos de polarización claramente definidos: el circuito de colector, el circuito de emisor y el circuito de base. Además de éstos hay dos circuitos de señal uno de entrada y otro de salida.

Esta identificación de los circuitos de una etapa también es muy útil para efectuar reparaciones fácilmente.

Conociendo los circuitos podemos realizar medidas de voltaje y de corriente y evaluar la calidad de los componentes.

Componentes o partes

La última división de un aparato es cuando establecemos los componentes o partes que forman un circuito. Estos componentes ya son elementos únicos como transistores, resistencias, condensadores, etc., y se denominan generalmente por su nombre y la función que desempeñan en el circuito.

Así tenemos por ejemplo, la resistencia de emisor, el condensador de acoplamiento, el transistor conversor, etc. El comportamiento de cada componente se debe diferenciar si se hace un análisis de polarización en CC o se analiza el paso de una señal de CA.

Una vez conocida esta división de un aparato en componentes vamos a estudiar el método recomendado a seguir en las reparaciones.

Método para las reparaciones

El método para las reparaciones está directamente relacionado con la división establecida en las páginas anteriores.

Este método también está compuesto por cuatro pasos o partes que son:

1. Diagnóstico.
2. Localización.
3. Aislamiento.
4. Sustitución.

1. Diagnóstico

El diagnóstico consiste en determinar cuál sección del aparato es la que tiene el problema. Para realizar un diagnóstico correcto debemos conocer muy bien las secciones del aparato y la forma como se relacionan unas con otras.

A este diagnóstico se debe llegar analizando el o los síntomas que presenta el aparato. El diagnóstico tiene tres etapas. La primera es el conocimiento del equipo por medio de su plano, el manual de servicio, la experiencia, el estudio o los cursos dictados por los fabricantes.

Con base en este conocimiento se debe hacer una inspección rápida, que puede ser visual, auditiva, olfativa o táctil, observando, escuchando, oliendo o tocando los componentes de esa sección.

Por último analizamos los síntomas para empezar a tomar decisiones. Entre los síntomas pueden estar: el equipo no trabaja nada, o trabaja parcialmente con algunas deficiencias en su funcionamiento, o sus controles no trabajan, o trabaja sólo un momento, o trabaja en forma intermitente.

Establecer el síntoma correcto es el primer paso para lograr la reparación. Con la experiencia se logra una gran habilidad para asociar inmediatamente el síntoma con la sección que lo produce. Para llegar a estas conclusiones se necesita tener razonamiento lógico.

No debemos apresurarnos a seguir siempre el método de los cuatro pasos en forma mecánica. También la intuición juega un papel importante en cualquier reparación.

2. Localización

Una vez que estamos seguros de la sección que tiene el problema, debemos hallar la etapa en esa sección que causa el problema. El objetivo de este paso no es encontrar todavía la parte o componente defectuoso: se trata de acercarnos a él.

En este paso debemos empezar a utilizar los instrumentos. Con las pruebas y la observación de los síntomas podemos determinar rápidamente cuál es la etapa defectuosa.

La mejor manera de determinar si una etapa de un radio está trabajando bien es inyectar una señal similar a la que utiliza esa etapa en su normal funcionamiento.

Para hacerlo debemos tener disponible generadores de señal como los que hemos estudiado en este curso: un generador de RF, un generador de audio, un inyector de señales o similares que se utilizan según la etapa a probar.

Prueba de una etapa de audio

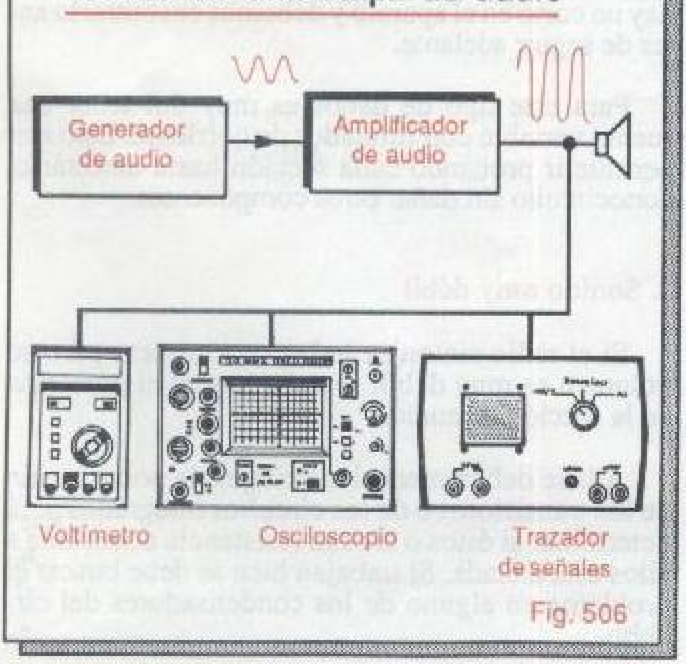

Fig. 506

Una vez aplicada la señal debemos tener la forma de comprobar que ésta es bien procesada por la etapa que estamos analizando.

En las etapas amplificadoras de audio, por ejemplo, podemos aplicar una señal débil audible en la entrada la cual debe resultar amplificada en la salida de esa etapa. Para comprobar este proceso necesitamos un voltímetro, un osciloscopio o un trazador de señales. Figura 506.

Un trazador de señales es un amplificador de audio que se utiliza para detectar si hay sonido en una etapa. Este es uno de los instrumentos más versátiles para la reparación de radios y lo podemos construir fácilmente.

Para ello se puede emplear uno de los circuitos amplificadores de audio con circuito integrado estudiados en la Lección 38 de este curso. Para hacer el circuito aplicable a señales de RF se le conecta en la entrada un circuito detector de RF (figura 507).

Trazador de señales

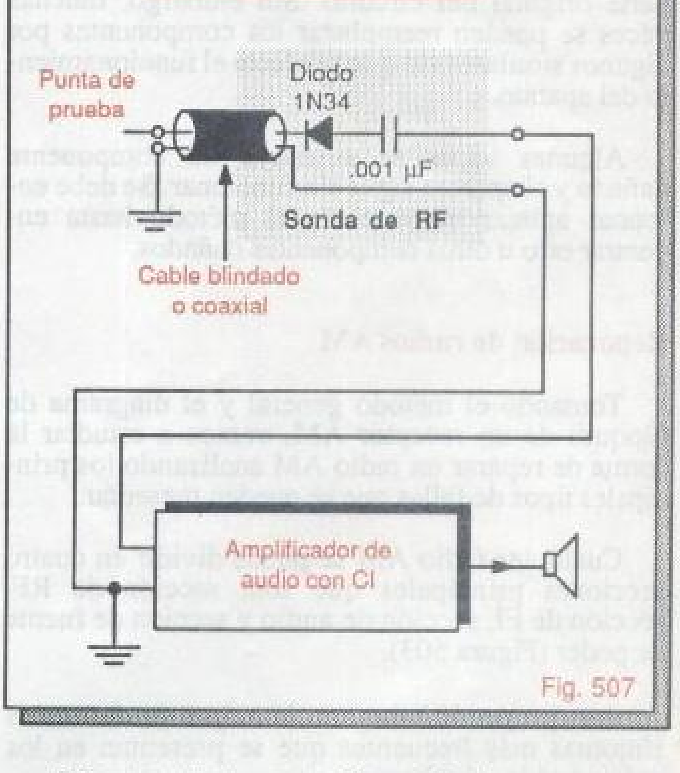

Fig. 507

Algunas etapas, como los osciladores, producen su propia señal y solamente debemos observar si entrega la señal correcta de acuerdo a lo estudiado en el curso.

Una vez establecido que una etapa nos entrega la señal que debe suministrar, debemos proceder a medir voltajes y corrientes en CC hasta determinar el componente o la conexión defectuosa.

3. Aislamiento

En este paso del método nos vamos acercando más al objetivo de la reparación y debemos *aislar* el circuito defectuoso. Este circuito está formado por pocos componentes y ya podemos verificar su funcionamiento uno por uno sin tomarnos mucho tiempo.

Midiendo en un transistor sus voltajes de colector, emisor o base podemos determinar si el transistor está dañado o si alguno de sus componentes relacionados, especialmente resistencias, está defectuoso.

Si hay dudas sobre un componente podemos desconectarlo y medirlo por fuera del circuito. Esto elimina la posibilidad de interacción con los otros componentes del circuito.

4. Sustitución

El paso final de toda reparación es la sustitución del o de los componentes dañados. Esta sustitución debe hacerse en lo posible por un componente o parte original del circuito. Sin embargo, muchas veces se pueden reemplazar los componentes por algunos similares sin que se afecte el funcionamiento del aparato.

Algunas veces se sustituye un componente dañado y el aparato sigue sin funcionar. Se debe entonces aplicar nuevamente el método hasta encontrar otro u otros componentes dañados.

Reparación de radios AM

Tomando el método general y el diagrama de bloques de un receptor AM, vamos a estudiar la forma de reparar un radio AM analizando los principales tipos de fallas que se pueden presentar.

Cualquier radio AM se puede dividir en cuatro secciones principales que son: sección de RF, sección de FI, sección de audio y sección de fuente de poder (Figura 503).

Aunque puede haber muchos tipos de daños, los síntomas más frecuentes que se presentan en los radios son los siguientes:

1. El radio no funciona totalmente.
2. Sonido muy débil.
3. Recepción intermitente y ruidosa.
4. Ruido de fondo muy alto.
5. Distorsión en el sonido.
6. Ruidos de alta frecuencia.
7. Sonido demasiado agudo.

Como se mencionó en el método debemos aprender a buscar la sección del radio que se relaciona con los síntomas presentados. Vamos ahora a tratar brevemente cada uno de los síntomas y el procedimiento a seguir.

1. El radio no funciona

Este primer caso, en donde un radio no produce ningún sonido, es quizás el más fácil de reparar. Probablemente, el daño se puede encontrar en la sección de audio o en la de la fuente de poder.

Lo primero que debemos hacer es verificar con un multímetro el estado de las pilas o baterías con el radio encendido. Si estos están bien, se debe verificar el funcionamiento del interruptor general para ver si está pasando el voltaje al aparato.

Si ese no es el problema, se debe probar el parlante. Luego podemos inyectar señal a la entrada de la sección de audio y se debe escuchar algo en el parlante. Si no ocurre esto debemos dividir la sección de audio e inyectar señal a cada una de sus etapas hasta encontrar cuál está defectuosa.

Una vez encontrada la etapa debemos medir sus voltajes de polarización hasta detectar el o los componentes defectuosos.

Si el radio tiene fuente de poder se verifica su funcionamiento. Muchas veces se daña el transformador de entrada que alimenta el circuito. Otro daño muy común en esta sección es un fusible de entrada.

Este se debe reemplazar siempre por uno de igual valor y si se sigue quemando nos indica que hay un corto en el aparato y debemos encontrarlo antes de seguir adelante.

Para este tipo de daños es muy útil tener una fuente variable con limitador de corriente. Esto nos permite ir probando cada sección hasta detectar el cortocircuito sin dañar otros componentes.

2. Sonido muy débil

Si el radio sintoniza todas las emisoras pero su volumen es muy débil, se debe buscar el problema en la sección de audio.

Allí se deben medir los voltajes de polarización de los transistores o de los circuitos integrados para determinar si éstos o alguna resistencia conectada a ellos está dañada. Si trabajan bien se debe buscar el problema en alguno de los condensadores del circuito.

Un condensador puede estar en cortocircuito o en circuito abierto. Si este último es el caso, no deja pasar la señal.

3. Recepción intermitente y ruidosa

En este caso, el radio sólo funciona cuando se coloca en determinadas posiciones o se golpea. Incluso se escucha un chisporroteo de fondo. Una forma práctica de encontrar la etapa a partir de la cual se manifiesta la perturbación es utilizando un analizador dinámico o trazador de señales.

La causa más frecuente de esta falla es la rotura de una de las pistas de cobre del circuito impreso de la etapa. Para identificar la pista defectuosa, provoque una avería total, presionando levemente la plaqueta hasta que la recepción se interrumpa por completo. Así es más fácil arreglar el problema.

Realice a continuación pruebas de continuidad con el óhmetro en todas las pistas de cobre de la etapa, especialmente las que llevan la corriente de alimentación. Recuerde que el óhmetro debe utilizarse con la energía desconectada.

Una vez identificada la pista defectuosa, reconstruya la conexión original con soldadura o alambre y pruebe el radio. Si la falla persiste, localice con el trazador de señales la siguiente etapa a partir de donde comienza la falla y efectúe las pruebas de continuidad.

Si después de medir y corregir la continuidad de las pistas el problema persiste, la causa debe buscarse en soldaduras "frías", terminales quebrados o componentes defectuosos.

4. Ruido de fondo muy alto

La mayoría de receptores presentan un ruido de fondo más o menos alto cuando se gira el control de volumen al máximo. Para calificar este ruido como excesivo, debe compararse con el de otro receptor bajo las mismas condiciones.

Un ruido excesivo puede originarse por un ajuste defectuoso de la RF o la FI, el cual obliga a utilizar al máximo la amplificación de audio. También puede ser provocado por un potenciómetro de control de volumen ruidoso o localizarse en el preamplificador de audio.

5. Distorsión en el sonido

Una forma de distorsión muy común en los receptores de radio es la que se manifiesta únicamente durante la recepción de una emisora local. La causa más probable es un desajuste en el circuito de control automático de ganancia (CAG).

Un segundo tipo de distorsión es la que es perceptible en la recepción de todas las emisoras, especialmente en las que transmiten con baja potencia. La causa más común es una polarización insuficiente de la etapa de salida del amplificador de audio.

Algunos receptores traen un potenciómetro para ajustar la corriente de polarización de los transistores de salida. Este potenciómetro debe ajustarse sólo lo necesario para que desaparezca la distorsión, sin provocar una corriente excesiva. Se recomienda utilizar un osciloscopio.

Otro tipo de distorsión implica también la pérdida de potencia. La recepción es normal a bajo volumen pero cuando se aumenta éste, la señal se distorsiona. Sucede generalmente porque los transistores de salida no son simétricos o uno de ellos ha dejado de trabajar por completo.

6. Ruidos de alta frecuencia

Los silbidos de alta frecuencia o "enganches" provienen, por regla general, de los amplificadores de FI o del conversor de RF. No se producen en todas las frecuencias.

Los silbidos en las etapas de FI se originan al variar la capacitancia interna entre colector y base de uno de los transistores. La forma de compensar esta variación es ajustar el condensador de neutralización, si lo trae, o reemplazar el transistor.

En otros casos, los ruidos de alta frecuencia son provocados por variaciones espontáneas en las características y valores de los condensadores de neutralización y de desacople y/o acoplamiento de FI.

Los silbidos de alta frecuencia originados en la etapa de conversión son particularmente estridentes y se deben a una reacción demasiado intensa del oscilador local. Una forma de corregir esta falla es reducir el voltaje de alimentación del conversor, aumentando el valor de la resistencia limitadora.

Otro recurso para minimizar la reacción del oscilador local es insertar una resistencia de 10 a 50 Ω en serie con el emisor del transistor conversor. Naturalmente, debe abrirse el circuito impreso en ese punto.

7. Sonido demasiado agudo

Cuando la reproducción de las frecuencias bajas es deficiente se presenta el fenómeno de la sonoridad aguda, caracterizado por la proliferación de altas frecuencias de audio. En estos casos debe examinarse el parlante para detectar algún bloqueo accidental en la bobina móvil o en la membrana.

El fenómeno de la sonoridad aguda va acompañado generalmente de distorsión y pérdida de potencia y sensibilidad. Si el parlante está en buen estado, la causa debe buscarse en un condensador de acoplamiento o de desacople.

Si la deficiencia en la reproducción de tonos graves se manifiesta cuando se conecta el trazador de señales al control de volumen, lo más probable es que esté fallando el condensador de desacople del CAG.

Guía de reparación de fallas del radio AM CEKIT

Si después de ensamblar todos los componentes en el tablero y seguir los procedimientos de alineación explicados en las actividades 25 y 26 su radio no le funciona como debe ser, identifique la categoría de la falla y realice las acciones correctivas que se recomiendan a continuación.

Localice la etapa defectuosa por el método explicado anteriormente. Compare su circuito con la guía a color. Busque componentes mal colocados, invertidos o en cortocircuito, así como soldaduras frías o puentes de alambre o de soldadura entre puntillas que no deben estar conectadas.

En cada etapa que revise, mida los voltajes de polarización de los transistores. En la tabla 1 se resumen los valores de voltaje de polarización típicos que deben obtenerse en diferentes puntos del receptor.

Transistor	Base (B)	Colector (C)	Emisor (E)
Q1	2.3 V	6.8 V	1.9 V
Q2	0.58 V	7.7 V	0.13
Q3	2.2 V	7.8 V	1.6 V
Q4	0.73 V	5.4 V	0.0 V
Q5	6.7 V	9.0 V	6.1 V
Q6	5.4 V	0.0 V	6.1 V

Vcc : Batería alcalina de 9V

Tabla 1

Si después de realizar todas las pruebas y estar seguro de que no hay componentes defectuosos ni errores de montaje, su radio continúa sin trabajar, deberá realinearlo paso a paso.

Recuerde que su receptor consta de fuente de alimentación, amplificador de audio, detector, segundo amplificador de FI, primer amplificador de FI y amplificador/conversor de FI. De este último forman parte el oscilador local y el mezclador.

Mida la corriente consumida por el radio para detectar la presencia de un posible cortocircuito. En condiciones normales, el radio consume de 3 a 15 mA. Si la corriente es excesiva, apague el radio, desconecte la resistencia y mídala nuevamente.

Si sigue siendo excesiva, el problema está en el amplificador de audio. Si no es así, localícelo en las secciones de FI o RF. Compare su circuito con la guía a color para encontrar la causa del corto. En especial, busque componentes mal colocados, invertidos o en cortocircuito interna o externamente.

Si el problema no es por cortocircuito, mida los voltajes de suministro del amplificador de audio (puntilla 38) y de las demás secciones (puntilla 39). En la puntilla 38 deberá medir 9V. En la puntilla 39 deberá medir 7 V, aproximadamente. Si no es así, pruebe la resistencia R12.

Pruebe el amplificador de audio aplicando a la entrada (puntilla A1) una señal de audio proveniente de un inyector o un generador. Deberá escuchar el tono en el parlante. Si no es así, revise el amplificador de audio para encontrar el componente defectuoso.

Pruebe el detector inyectando una señal modulada de 455 KHz a la puntilla C1. Puede usar el generador de FI construido en la Actividad 17. Deberá escuchar el tono de modulación en el parlante. Si no es así, revise el detector para encontrar el componente defectuoso.

Pruebe la segunda etapa de FI inyectando una señal modulada de 455 KHz a la puntilla E4 a través de un condensador de 0.01 μF. Deberá escuchar el tono de modulación en el parlante. Si no es así, revise la etapa para encontrar el componente defectuoso.

Pruebe la primera etapa de FI inyectando una señal modulada de 455 KHz a la puntilla D4 a través de un condensador de 0.01 μF. Deberá escuchar el tono de modulación en el parlante. Si no es así, revise la etapa para encontrar el componente defectuoso.

Pruebe el amplificador/conversor de RF inyectando una señal modulada de 455 KHz a la puntilla B2 a través de un condensador de 0.01 μF. Deberá escuchar el tono de modulación en el parlante. Si no es así, revise la etapa para encontrar el componente defectuoso.

Lección 41

Reparación de radios de FM

Introducción

En esta lección enfocaremos el tema de la reparación de receptores de FM desde un punto de vista general y lo particularizaremos al radio FM CEKIT como aplicación práctica. La metodología seguida refuerza los principios básicos expuestos en la lección anterior.

Se asume que usted está familiarizado con los diversos aspectos teóricos y técnicos involucrados en la operación de los receptores de FM. Si no es el caso, le sugerimos repasar las lecciones relacionadas con el tema. De este modo, adquirirá una mayor seguridad en sus conocimientos.

La reparación de un aparato electrónico es una aplicación práctica y lógica de conceptos técnicos. Si usted maneja estos conceptos, dispondrá de elementos de juicio razonables para enfocar el problema y localizar el área donde se origina la falla.

Es absurdo pretender reparar un radio de FM si no sabemos cómo funciona ni podemos identificar, tanto en el plano como en el receptor mismo, sus secciones, etapas y circuitos o desconocemos la función de componentes claves como CIs, semiconductores, circuitos sintonizados, etc.

Para lograr esta solidificación de conceptos es importante refrescar, actualizar y renovar conocimientos a través del estudio y la investigación, practicados de manera asidua y disciplinada. Sin embargo, un buen reparador de radio no se forma sólo a través de la lectura.

Un reparador electrónico deja de ser un empírico y se convierte en un experto cuando consigue desarrollar, a través de la experiencia, un modo de proceder lógico y eficiente sugerido por su propio razonamiento y adquiere habilidad en el manejo de herramientas, instrumentos e información técnica.

No pretendemos que con estas dos últimas lecciones usted se convierta en un experto en la reparación de radios AM y FM. Sólo queremos orientarlo hacia la búsqueda de una metodología personal donde sus conocimientos, habilidades e interés juegan un papel muy importante.

Pruebas básicas

Antes de comenzar a buscar y analizar sistemáticamente las causas de una falla, revise primero las cosas fáciles de reparar. De este modo, si el daño es sencillo, podrá ahorrarse tiempo y procedimientos innecesarios.

Por simple rutina, mida siempre los voltajes de CA y de CC asociados a la fuente de alimentación. Revise las tensiones de entrada, de salida y de control del transformador, del rectificador y de los reguladores de tensión. Mida también la corriente total consumida para detectar posibles cortos.

En la figura 508 se muestra el diagrama esquemático de la fuente de alimentación del radio FM CEKIT y se indican los valores típicos de voltajes AC y DC presentes en diferentes puntos del circuito en condiciones normales de funcionamiento.

Voltajes de la fuente de alimentación

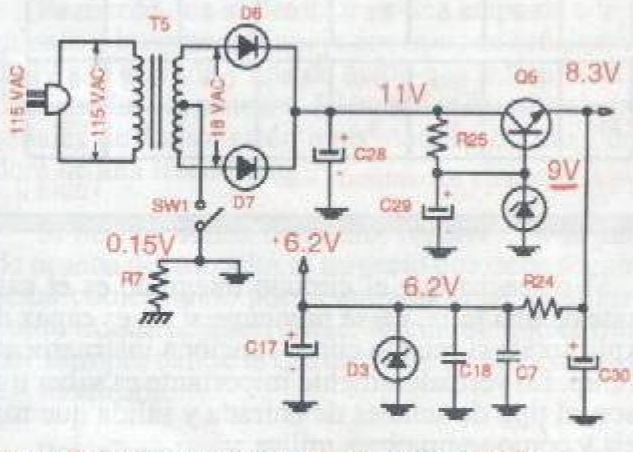

Nota: Todos los voltajes de CA son valores RMS.
Todos los voltajes de CC están medidos con respecto a tierra ⏚

Fig. 508

De la misma forma mida los voltajes de CC asociados a todos los transitores. De este modo podrá saber si alguno de ellos está abierto o en cortocircuito o no está siendo correctamente polarizado por los circuitos externos. Un transitor mal polarizado no puede trabajar apropiadamente.

En la Tabla 2 se relacionan los voltajes de polarización típicos de los transistores de RF y FI del radio FM CEKIT. VC1-VC4 son los voltajes de reposo de colector de Q1-Q4, VB1-VB4 los voltajes de base y VE1-VE4 los de emisor. Todos estos voltajes están medidos con respecto a tierra.

	Q1	Q2	Q3	Q4
VC	7.59 V	7.59 V	7.34 V	7.16 V
VE	0.68 V	1.20 V	0.84 V	1.13 V
VB	1.34 V	1.68 V	1.53 V	1.84 V

Nota: Voltaje de alimentación = 7.75 V

Tabla 2

Mida también los voltajes de CC presentes en cada uno de los pines de los circuitos integrados utilizados en el receptor. Así podrá detectar posibles cortocircuitos internos o entre pines y evaluar si están siendo o no correctamente alimentados.

En la Tabla 3 se relacionan los valores típicos de voltaje de CC obtenidos en los pines del CI amplificador de audio LA4101 del radio FM CEKIT. Todas estas tensiones están medidas con respecto a tierra y pueden variar de acuerdo a la tolerancia de los componentes.

Pin	14	13	12	11	10	9	8
V	8.7	8.5	8.5	0.0	4.4	4.3	3.9
V	4.3	0.0	0.0	5.1	0.4	4.3	0.0
Pin	1	2	3	4	5	6	7

Nota: Voltaje de alimentación = 8.7 V

Tabla 3

Si sospecha que el circuito integrado es el causante de una falla, no se preocupe si no es capaz de explicarse a sí mismo cómo funciona internamente el chip. Lo verdaderamente importante es saber qué hace, el tipo de señales de entrada y salida que maneja y cómo se prueba y utiliza.

Toda esta información se la suministra el fabricante del CI en las hojas de datos del producto o en los manuales técnicos. Si el dispositivo no trabaja de acuerdo a estas especificaciones, es porque está defectuoso y debe reemplazarse.

Busque visualmente cables sueltos, conexiones abiertas o resistencias quemadas. Muchas veces una mala recepción se debe sólo al hecho de que el cable que va a la antena se ha desoldado, hace un contacto intermitente o está abierto.

Con el control de volumen al máximo, mueva cuidadosamente con un destornillador o un estilete no metálico resistencias, condensadores, transistores, bobinas y tanques de FI para detectar posibles soldaduras frías o roturas de los terminales.

Mueva también el condensador variable de sintonía. Si el radio ha recibido un impacto o se ha dejado caer al suelo, es posible que una de las patillas se haya desoldado del circuito impreso o esté haciendo un contacto intermitente.

Flexione suavemente la tarjeta de circuito impreso, especialmente a lo largo de los costados. Es probable que durante esta operación, el radio recupere repentinamente su funcionamiento normal, indicando que existe una conexión abierta o suelta que ha sido temporalmente restablecida.

Si cuando prende el radio escucha un "click" en el parlante, será una indicación clara de que la bobina del parlante y el circuito de salida del amplificador de audio están trabajando bien. Si no lo escucha, algo anormal está pasando en esa etapa.

Toque con un dedo el terminal central o "cursor" del potenciómetro que sirve como control de volumen con este último al máximo. Si se produce un ruido en el parlante (hum), el amplificador de audio está trabajando correctamente.

Cuando estas pruebas no son suficientes para localizar la falla, debemos acudir necesariamente al uso de instrumentos como el generador de RF y el trazador de señales. El procedimiento que sigue se basa en razonamientos lógicos del tipo "si esto sucede, entonces proceda a esto" y es aplicable a la reparación de cualquier aparato electrónico.

Puebas con generadores y trazadores de señal.

En la figura 509 se muestra la forma de utilizar un generador de RF/FI, un trazador de señales y un generador de audio para revisar el funcionamiento de un receptor. El método es aplicable tanto a radios de AM como de FM. Recuerde que en AM la FI es de 455 KHz y en FM es de 10.7 MHz.

Rastreo de señales en un radio

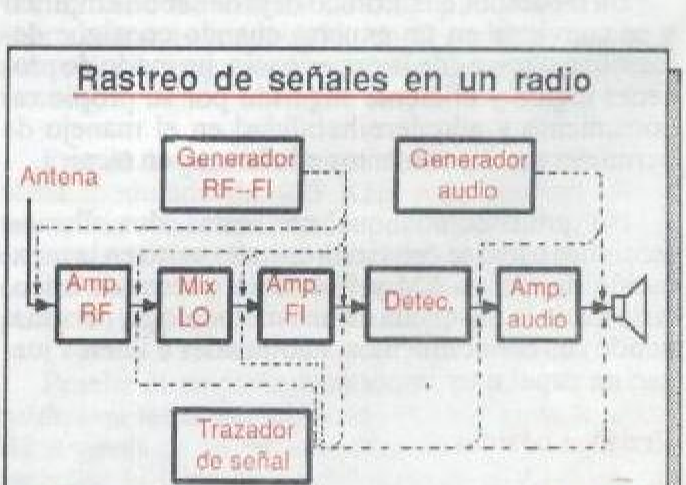

Fig. 509

Si no posee generador de audio, puede utilizar en su lugar el inyector de señales construido en la Actividad 11. El generador de RF debe poder entregar señales de FM tanto de 10.7 MHz como de 88 a 108 MHz. El trazador de señales es un detector seguido de un amplificador de audio.

Pruebe el amplificador de audio inyectando la señal del generador de AF inicialmente al parlante y luego a la entrada del amplificador. En ambos casos deberá escuchar un tono en el parlante. Si no es así, el parlante está defectuoso o el amplificador de audio tiene una falla.

En el radio FM CEKIT, puede tomarse como entrada del amplificador de audio el terminal (-) de C26 y como salida el terminal (-) de C38. El camino recorrido por la señal de audio desde la entrada hasta la salida se ilustra con línea gruesa en el circuito de la figura 510.

Flujo de señal en el amplificador de audio

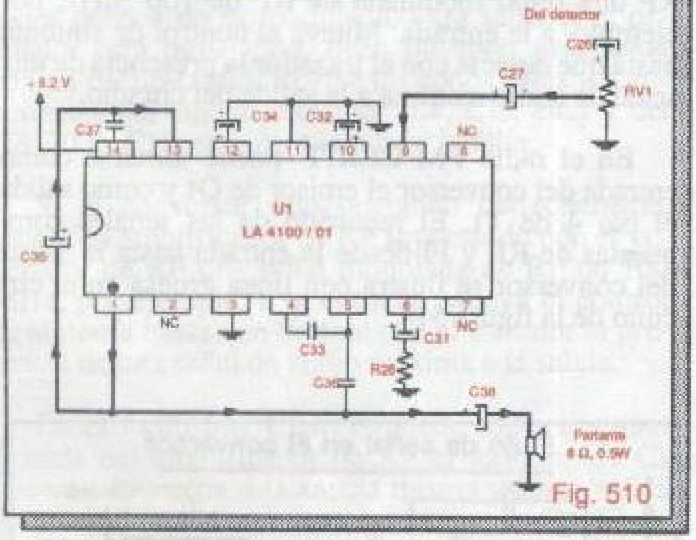

Fig. 510

Si no hay salida de audio, recorra con la punta de prueba del trazador el trayecto que debe seguir la señal comenzando por la entrada hasta localizar el punto donde se interrumpe. Por ejemplo, si hay audio en el pin 9 de U1 pero no en el pin 1, lo más probable es que el CI esté defectuoso.

Si la señal de audio se escucha en el parlante con distorsión, revise los condensadores de realimentación C33 y C36. Si la reproducción de altas y bajas frecuencias no es uniforme, revise C31 y R26. Si hay ruido de fondo (hum), revise C32 y C34.

Si el amplificador de audio está operando correctamente, pruebe el demodulador inyectando con el generador de RF una señal modulada de FI de 10.7 MHz a la entrada. Detecte con el trazador la presencia de señal de audio a la salida del detector.

En el radio FM CEKIT, la entrada de FI del detector de relación es el pin 3 de T3. La salida se puede tomar del potenciómetro RV1. El recorrido de la señal compuesta de FI desde la entrada hasta la salida del detector se ilustra con línea gruesa en el circuito de la figura 511.

Flujo de señal en el detector de FM

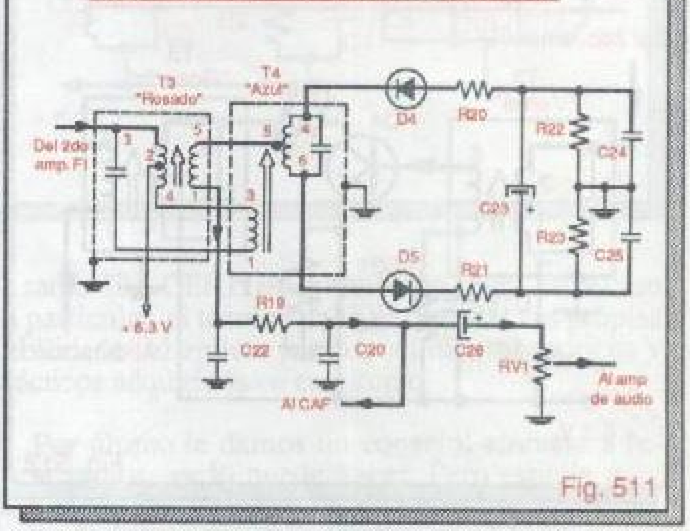

Fig. 511

Recuerde que el detector es una etapa de acoplamiento o interface. Maneja dos tipos de señales: una de FI a la entrada y una de audio a la salida. Esta última modula la primera. El trazador detecta siempre señales de audio, estén o no modulando una portadora de alta frecuencia.

Si no hay salida de audio, recorra con la punta de prueba del trazador el trayecto que debe seguir la señal comenzando por la entrada hasta localizar el punto donde se interrumpe. Recuerde que el trazador siempre extrae la componente de audio de la señal modulada.

Por ejemplo, si hay audio en C22 pero no en C20, lo más probable es que R19 esté abierta o se ha interrumpido una de las pistas de circuito impreso que conectan este componente a C22 o C20.

Revise también los componentes del filtro de FI (C22, R19 y C20) cuando se presente el fenómeno de sonoridad aguda (ruidos de alta frecuencia) en la señal de audio escuchada en el parlante y los del limitador (C23, R22 y R23) cuando el volumen de esta señal no sea constante.

Si el detector de relación está operando correctamente, pruebe el segundo amplificador de FI inyectando con el generador de RF una señal modulada de FI de 10.7 MHz a la entrada. Detecte con el trazador la presencia de señal de audio a la salida del circuito.

En el radio FM CEKIT puede tomarse como entrada del segundo amplificador de FI la base de Q4 y como salida el pin 3 de T3. El recorrido de la señal compuesta de FI desde la entrada hasta la salida del amplificador se ilustra con línea gruesa en el circuito de la figura 512.

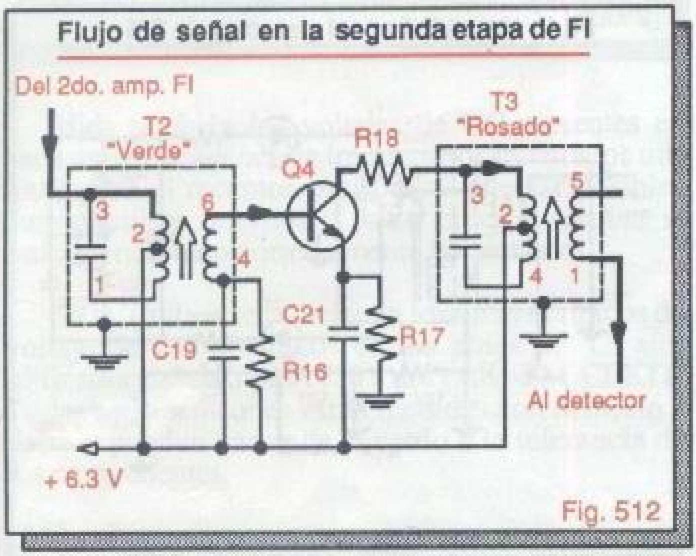
Del 2do. amp. FI
T2
"Verde"
3
2
6
4
1
Q4
R18
T3
"Rosado"
3
2
5
4
1
C19
R16
C21
R17
Al detector
+ 6.3 V

Flujo de señal en la segunda etapa de FI

Fig. 512

Si no hay salida de audio, recorra con la punta de prueba del trazador el trayecto que debe seguir la señal comenzando por la entrada hasta localizar el punto donde se interrumpe.

Por ejemplo, si el trazador detecta audio en la base de Q4 pero no en el colector, lo más probable es que este transistor esté defectuoso o desconectado. Puede suceder también que el circuito esté oscilando y por lo tanto ignora la señal aplicada a la entrada. En este caso, revise también R18.

Si la segunda etapa de FI está operando correctamente, pruebe la primera etapa inyectando con el generador de RF una señal modulada de FI de 10.7 MHz a la entrada. Detecte con el trazador la presencia de señal de audio a la salida del circuito.

En el radio FM CEKIT puede tomarse como entrada del primer amplificador de FI el pin 4 de T1 y como salida el pin 6 de T2. El recorrido de la señal compuesta de FI desde la entrada hasta la salida del amplificador se ilustra con línea gruesa en el circuito de la figura 513.

Si no hay señal de audio a la salida, recorra con la punta de prueba del trazador el trayecto que debe seguir la señal comenzando por la entrada hasta localizar el punto donde se interrumpe.

Por ejemplo, si el trazador detecta audio en el pin 3 de T2 pero no en el pin del mismo, puede estar sucediendo que este transformador está abierto internamente o su pin 4 está desconectado y no tiene retorno a tierra a través de C19 o R16.

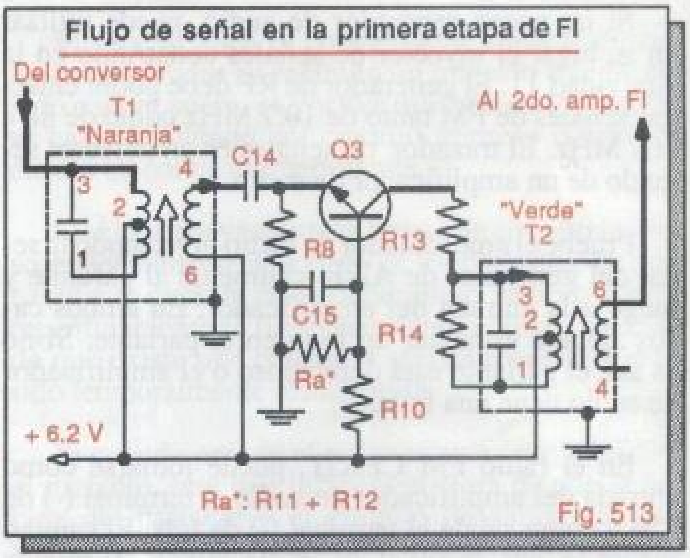
Del conversor
T1
"Naranja"
3
2
4
6
1
C14
Q3
Al 2do. amp. FI
R8
R13
"Verde"
T2
3
2
6
1
4
C15
R14
Ra*
R10
+ 6.2 V
Ra*: R11 + R12

Flujo de señal en la primera etapa de FI

Fig. 513

Si la primera etapa de FI está operando bien, pruebe el conversor inyectando con el generador de RF una señal modulada de RF de 100 MHz, por ejemplo, a la entrada. Mueva el control de sintonía hasta que detecte con el trazador la presencia de una señal de audio máxima a la salida del circuito.

En el radio FM CEKIT puede tomarse como entrada del conversor el emisor de Q1 y como salida el pin 4 de T1. El recorrido de las señales compuestas de RF y FI desde la entrada hasta la salida del conversor se ilustra con línea gruesa en el circuito de la figura 514.

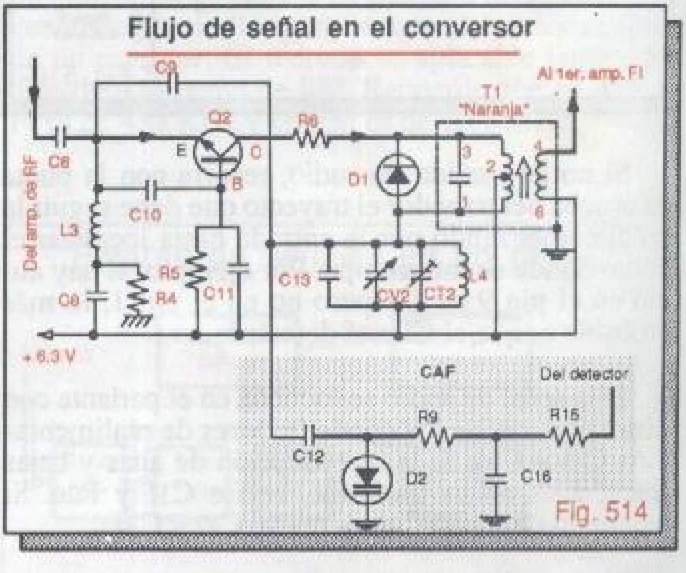
Del amp. de RF
C9
C8
Q2
E
C
B
R6
T1
"Naranja"
Al 1er. amp. FI
3
2
4
1
6
D1
C10
L3
R5
C11
C6
R4
C13
CV2
CT2
L4
+ 6.3 V
CAF
Del detector
R9
R15
C12
D2
C16

Flujo de señal en el conversor

Fig. 514

Recuerde que el conversor es una etapa acopladora o de interface. Maneja tres tipos de señales: la de RF de entrada, la de FI de salida y la de LO (oscilador local) de control. Las dos primeras son moduladas y por lo tanto llevan señal de audio. La señal LO no debe contener modulación.

Si no hay señal de audio a la salida, recorra con la punta de prueba del trazador el trayecto que debe seguir la señal comenzando por la entrada hasta localizar el punto donde se interrumpe. No olvide mover el control de sintonía hasta localizar el punto donde se produce la FI de 10.7 MHz.

Por ejemplo, si el trazador detecta audio en el emisor de Q2 pero no en el colector, puede ser porque este transistor está defectuoso o se ha interrumpido la conexión con R6 o está abierto el primario de T1, entre otras posibilidades.

Si el trazador detecta audio a la salida del conversor pero la señal no se escucha en el parlante a pesar de que todas las demás etapas subsiguientes están operando correctamente, lo más probable es que el oscilador local está trabajando erráticamente o el transformador T1 está desintonizado.

La señal de LO no la puede detectar con el trazador porque no es modulada. Para saber si el oscilador está trabajando, debe observar esta señal con un osciloscopio. Revise las conexiones y componentes del tanque oscilador (L4, C13, etc.) y del CAF (D2, C12, etc.) para localizar la falla.

Si el conversor está operando correctamente, pruebe el amplificador de RF inyectando con el generador de RF una señal modulada de RF de 100 MHz, por ejemplo, a la entrada. Mueva el control de sintonía hasta que detecte con el trazador la presencia de una señal de audio máxima a la salida.

En el radio FM CEKIT puede tomarse como entrada del amplificador de RF el terminal de C1 que va a la antena o la antena misma y como salida el terminal de C6 conectado al emisor de Q2. El recorrido de la señal compuesta de RF desde la entrada hasta la salida del amplificador de RF se ilustra con línea gruesa en el circuito de la figura 515.

Si no hay audio en la señal de RF de salida, recorra con la punta de prueba del trazador el trayecto que debe seguir la señal comenzando por la entrada hasta localizar el punto donde se interrumpe. No olvide mover el control de sintonía hasta encontrar el punto de máxima ganancia de RF.

Por ejemplo, si el trazador detecta audio en el punto de unión de C1 y C2 pero no en el emisor de Q1, es probable que C2 esté abierto o desconectado o que R1 esté en cortocircuito.

Con estas instrucciones creemos que hemos dado una guía suficiente para emprender nuestras primeras reparaciones. Hemos utilizado como modelo el radio FM CEKIT. Sin embargo, para cada caso en particular, el técnico debe desarrollar sus propias habilidades al aplicar los conocimientos teóricos y prácticos adquiridos en este curso.

Flujo de señal en el amplificador de RF

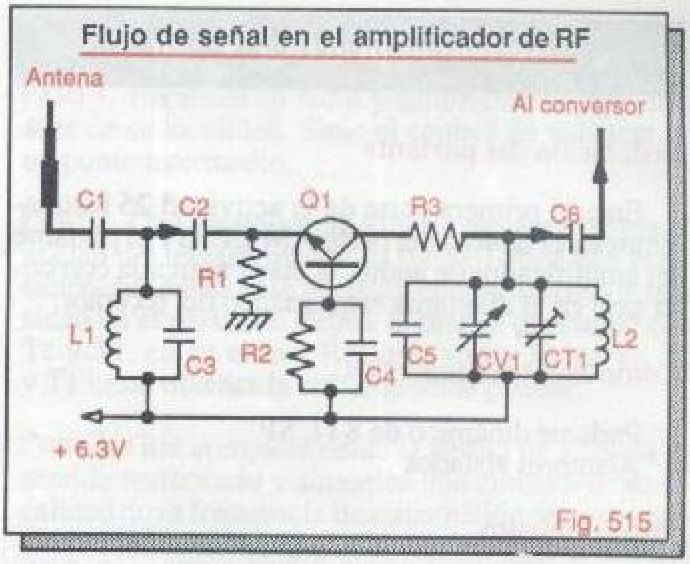

Fig. 515

Por último le damos un consejo: atrévase a reparar radios, ya lo puede hacer. Pero estudie muy bien el aparato y realice los procesos técnicamente. Antes de actuar, planee mentalmente cada paso.

Si después de intentar una reparación por un buen tiempo tiene la sensación de que no puede hacerla o no se consigue un determinado componente o repuesto, devuelva el aparato a su dueño y sea sincero con él: dígale que usted no está en capacidad de repararlo. No prometa que lo hará mañana, la próxima semana o el próximo mes. Tal vez nunca pueda hacerlo.

Una regla de oro que se debe seguir es que si no logra reparar algo en un tiempo prudencial, es mejor renunciar a ese trabajo e iniciar rápidamente otro en el cual pueda lograr una reparación exitosa.

Es necesario resaltar la importancia de los instrumentos de prueba y de medida que debe usted tener en su laboratorio. A lo largo de este curso hemos aprendido a manejar el multímetro, el osciloscopio, el generador de audio, el generador de RF y otros aparatos y sabemos de su utilidad.

También debe disponer de herramientas básicas de trabajo como pinzas, cortafríos, cautín y destornilladores tanto metálicos como plásticos. Estos últimos son imprescindibles para ajustar los núcleos de ferrita de las bobinas de alta frecuencia utilizadas en los receptores de radio.

Debe también poseer un buen stock de componentes de emergencia como resistencias y condensadores de varios valores.

Actividad práctica Nº 26

Instalación del parlante

En esta primera parte de la actividad 26 instalaremos en el tablero del radio AM CEKIT el parlante del amplificador de audio, la etapa marcada con color azul en el diagrama esquemático del receptor.

Componentes necesarios

1 Parlante dinámico de 8 Ω. SP
2 Alambres aislados

Procedimiento

Paso 1. Tome el primer alambre aislado. Suelde un extremo a la puntilla A30 y el otro extremo a uno de los bornes del parlante.

Paso 2. Tome el segundo alambre aislado. Suelde un extremo a la puntilla A31 y el otro extremo al borne restante del parlante.

Paso 3. Asegure el parlante a la tabla de montaje como se muestra en la figura A70-A. Puede utilizar cinta adhesiva de doble faz o pegante sintético. La calidad del sonido mejora notablemente si monta el parlante en un baffle o caja acústica similar a la mostrada en la figura A70-B.

Si se decide por utilizar el parlante externamente, no necesita conectarlo a las puntillas A30 y A31. Puede utilizar el jack del audífono, conectando el parlante a un plug monofónico como se muestra en la figura 70-C.

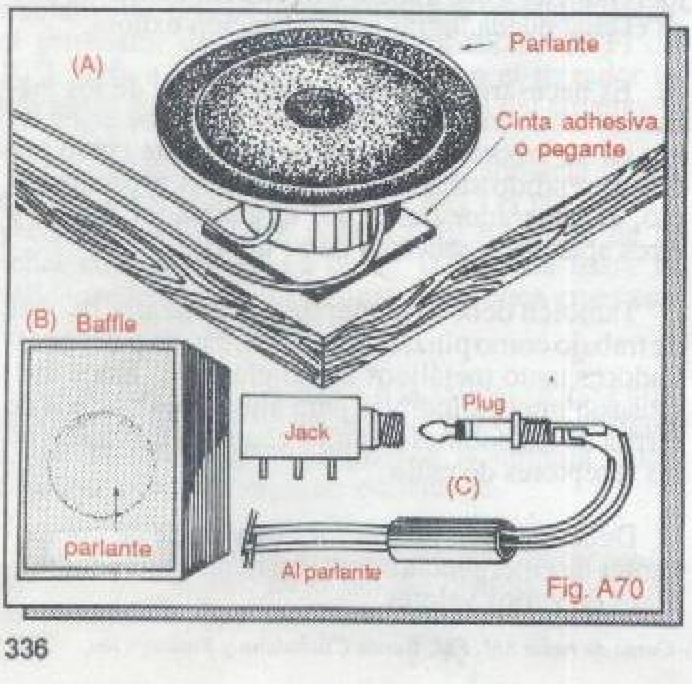

Fig. A70

Alineamiento del radio con generador de RF

Se describe a continuación la forma de calibrar el radio AM CEKIT utilizando únicamente un generador de RF. Alinearemos los transformadores de FI (T3, T2 y T1), las bobina osciladora (L2), la bobina de antena (L1) y el bloque de sintonía (CV1, CV2, CT1, CT2).

Nota: Si realizó la Actividad 25 hasta el final y su radio ya está calibrado, no necesita efectuar estas pruebas. Sólo reajuste ligeramente, en su orden, los tornillos de CT1 y CT2 y los núcleos de L1 y L2 para que los extremos de 1600 y 530 KHz de su dial coincidan con los de un radio de AM en servicio.

Si no posee amplificador de RF, puede utilizar el generador de FI construido en la actividad 17 como fuente de señal modulada de 455 KHz.

Paso 1. Conecte la batería al radio y enciéndalo. Sitúe el control de volumen en su punto medio.

Paso 2. Programe el generador de RF para producir una señal AM de 455 KHz. Conecte el caimán negro a tierra e inyecte la señal de RF a la base de Q3 (puntilla E4) con un condensador de 0.01 μF.

Paso 3. Ajuste el núcleo del transformador T3 (negro) con un destornillador plástico hasta que la intensidad del tono de modulación escuchado en el parlante sea máxima. Desconecte el generador de RF de la base de Q3.

Paso 4. Inyecte la señal de RF de 455 KHz a la base de Q2 a través del condensador de 0.01 μF. Ajuste el núcleo del transformador T2 (blanco) hasta que el tono escuchado en el parlante sea máximo. Desconecte la sonda del generador de RF de la base de Q2 y el caimán negro de la puntilla de tierra.

Paso 5. Inhabilite el oscilador local desconectando el condensador C2 de la puntilla B6. Desconecte también el alambre rosado de la bobina de antena del terminal "A" del condensador variable.

Paso 6. Conecte el caimán negro del generador de RF a la sonda de prueba del mismo formando un bucle. Sitúe este bucle cerca de la bobina de antena (L1) para acoplar por inducción la señal modulada de RF de 455 KHz al radio.

Paso 7. Sitúe el control de amplitud del generador de RF al máximo. Deberá escuchar el tono de modulación en el parlante. Ajuste el núcleo del transformador T1 (amarillo) hasta que el tono escuchado en el parlante sea máximo.

Paso 8. Reajuste los transformadores T3, T2 y T1, en ese orden, hasta lograr, con cada uno, el máximo volumen posible. Una vez obtenido el punto óptimo de ajuste, inmovilice los núcleos con parafina derretida o similar.

Paso 9. Conecte nuevamente el alambre rosado de la bobina de antena al terminal "A" del condensador variable y el condensador C2 a la puntilla B6.

Paso 10. Gire completamente la perilla de sintonía del receptor en sentido antihorario (CCW). Programe el generador de RF para producir una señal AM de 530 KHz. Acerque el bucle de RF a la bobina de antena. Deberá escuchar el tono de modulación en el parlante.

Paso 11. Ajuste el núcleo de la bobina L2 (roja) hasta que el tono escuchado en el parlante sea de amplitud máxima. Una vez obtenido el punto de ajuste óptimo, inmovilice el núcleo de L2 con parafina derretida.

Paso 12. Ajuste la posición de la bobina de antena (L1) en su núcleo desplazándola lentamente hacia adelante y hacia atrás hasta que el tono escuchado sea máximo. Una vez obtenido el punto de ajuste, solidarice la bobina al núcleo con parafina derretida.

Paso 13. Gire completamente la perilla de sintonía del receptor en sentido horario (CW). Programe el generador de RF para producir una señal AM de 1600 KHz. Acerque el bucle de RF a la bobina de antena. Deberá escuchar el tono de modulación en el parlante.

Paso 13. Ajuste el tornillo del condensador trimmer CT2 hasta que el tono escuchado en el parlante alcance su máxima intensidad. Una vez obtenido el punto óptimo de ajuste de CT2, realice la misma operación con el trimmer CT1 hasta que el tono sea máximo.

Alineamiento del radio sin equipo de prueba

Paso 1. Encienda su radio y sintonícelo en una emisora de su localidad. Sitúe el control de volumen en un punto intermedio.

Paso 2. Con un destornillador no metálico, ajuste el núcleo del transformador T3 (negro) hasta que la estación sintonizada se escuche con la máxima intensidad en el parlante. Repita la misma operación con T2 y T1, en su orden. Reajuste varias veces T3, T2 y T1 hasta obtener la máxima señal posible.

Paso 3. Gire completamente la perilla de sintonía en sentido antihorario y sintonice una emisora de su localidad cuya frecuencia de transmisión sea muy próxima a 530 KHz. Ajuste el núcleo de las bobinas osciladora (L2, roja) y de antena (L1) hasta que la señal captada sea máxima.

Paso 4. Gire completamente la perilla de sintonía en sentido antihorario y sintonice una emisora de su localidad cuya frecuencia de transmisión sea muy próxima a 1600 KHz. Ajuste los tornillos de CT1 y CT2 hasta que la señal captada sea máxima. De este modo habrá completado la calibración de su radio.

Voltajes de polarización del radio AM CEKIT

Una forma práctica y segura para diagnosticar si uno de los transistores del radio es el causante de una falla determinada es medir sus voltajes de polarización de la base, el colector y el emisor. Un transistor mal polarizado no puede realizar adecuadamente su función amplificadora o de otro tipo.

En la figura A71 se muestran los valores típicos de voltajes de CC presentes en diferentes puntos del radio AM CEKIT. Todas estas tensiones están medidas con respecto a tierra.

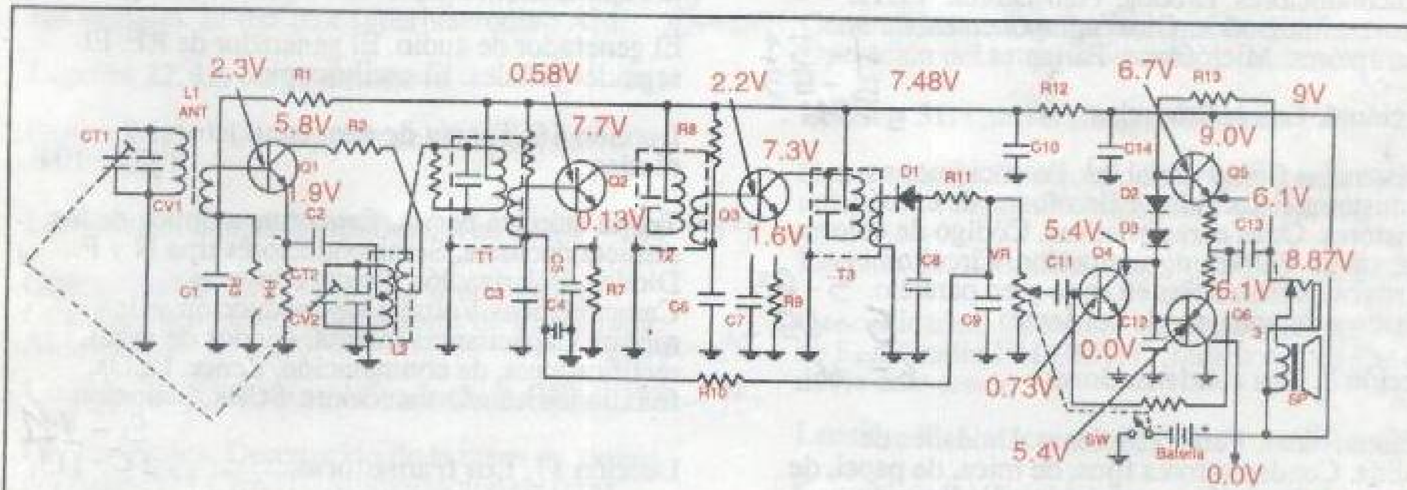

Fig. A71

INDICE GENERAL DEL CURSO DE RADIO AM Y FM

INDICE DE LECCIONES

Prólogo T 6

Introducción. Teoría básica de Radio 1- 9-16

Radio. Radiocomunicación. Ondas. Historia de la radio.Ondas de radio. Tipos de ondas. Transmisión y recepción de ondas de radio. Componentes utilizados en los aparatos de radio.

Lección 1. La materia y los electrones 2- 17-28

Materia. Elementos. Moléculas. Atomos. Cargas. Teoría Atómica. Electricidad. Conductores, aislantes y semiconductores.

Lección 2. Circuitos y Ley de Ohm 2- 22

Corriente, voltaje, resistencia y potencia. Circuitos. Aplicaciones de la Ley de Ohm. Formas de la corriente eléctrica.

Lección 3. Corriente Continua (CC ó DC) 2-23-28

Generalidades. Usos. Métodos de producción. -3-29 Fuentes de poder. Pilas y baterías. Pilas secas, alcalinas, de mercurio, de níquel - cadmio y solares. Acumuladores. Conexiones en serie y paralelo de pilas.

Lección 4. Corriente alterna (CA o AC) 3- 35-40

Características.Formas de onda. Ruido. Valores pico, pico a pico, RMS y promedio de un voltaje alterno. Concepto de Amplitud. Ciclo. Frecuencia. Período. Relación de fase. Armónicos 3- 44

Lección 5. Los componentes electrónicos 4- 45

Componentes pasivos. Resistencias. Condensadores. Bobinas. Transformadores. Semiconductores. Diodos. Transistores. FETs. Circuitos integrados. Otros componentes. Interruptores. Micrófonos. Parlantes. 4-51 5-53

Lección 6. Las resistencias 5- 54

Resistencias fijas y variables. Potenciómetros. Termistores.Fotoceldas. Micrófonos de carbón. Varistores. Otras características. Código de colores. Tolerancia. Valores normalizados. Circuitos resistivos. Resistencias en serie y en paralelo. 5-63 Divisores de voltaje y de corriente. 6-65

Lección 7. Los condensadores 6- 66

Características. Funcionamiento. Unidades de medida. Condensadores fijos, de mica, de papel, de cerámica, electrolíticos, variables. Códigos de identificación. Condensadores en serie y en paralelo. Aspectos prácticos. Reactancia.

Lección 8. Las bobinas 6- 75

Inductancia. Magnetismo. Imanes. Materiales magnéticos y diamagnéticos. Electromagnetismo. Bobinas. Núcleos. Producción de electricidad por medios magnéticos. Inducción magnética. Circuitos inductivos. Comportamiento de una bobina con CC y CA. Relaciones de fase.Bobinas de núcleo de hierro, aire, ferrita y variables. Bobinas en serie y en paralelo. Construcción de bobinas. 6-76 7-81

Lección 9. Los transformadores 7- 82

Elementos. Funcionamiento. Transformadores de RF, de audio y de potencia.

Lección 10. La impedancia 7- 84

Generalidades. Impedancia de circuitos RL, RC, RLC serie y RLC paralelo.

Lección 11. Los filtros 7- 86

Generalidades. Filtros pasabajos, pasaaltos, pasabanda y de rechazo de banda.

Lección 12. Resonancia 7- 88

Resonancia serie y paralelo. Cálculo. 8-89

Lección 13. Los instrumentos de medida 8-90

Generalidades. Instrumentos de bobina móvil.

Lección 14. El multímetro electrónico 8-93

Generalidades. Cómo medir corriente, voltaje y resistencia. Manejo. Instrumentos digitales.

Lección 15. Instrumentos para el ajuste y la reparación de aparatos de radio 9- 101

El generador de audio. El generador de RF. El seguidor señales. El osciloscopio.

Lección 16. Teoría de semiconductores y diodos 9- 104

Teoría atómica básica. Estructura atómica de los semiconductores. Semiconductores tipo N y P. Diodos. Polarización directa e inversa. Características.Voltajes de conducción y de ruptura.Capacitancia interna. Diodos de señal, rectificadores, de conmutación, Zener, LEDs, fotodiodos, túnel, varactores, SCRs, Schottky. 9-111

Lección 17. Los transistores 10-113

Generalidades. Clases. Encapsulados. Códigos de identificación. Transistores bipolares.

Características. Transistores NPN y PNP. El transistor como amplificador. Circuitos base común, emisor común y colector común. Parámetros alfa y beta. Prueba de transistores bipolares. Transistores FET. Transconductancia. Transistores MOSFET. Protección contra ESD. Modos de agotamiento y realce. Los FET y MOSFET como amplificadores. Circuitos de compuerta común, fuente común y drenador común.

10-123

Lección 18. Amplificadores de audio 11-125

Generalidades. Características. Ganancia. Impedancia de entrada y de salida. Respuesta de frecuencia y ancho de banda.Amplificadores de voltaje, de corriente y de potencia. Amplificadores de audio y de RF. Amplificadores de señal pequeña y de señal grande. Rendimiento de un amplificador de potencia. Amplificadores de potencia push-pull. Clases de operación A, B, AB y C.Amplificadores realimentados. Importancia de la realimentación negativa. Descripción del amplificador de audio del radio AM CEKIT.

11-135

Lección 19. Cómo se produce la señal de una emisora de radio 12-137

Estructura general de una emisora. Fuentes de señales de audio. Micrófonos, tocadiscos y grabadoras.Fuentes de señales de RF. Osciladores Hartley, Colpitts y a cristal. Modulación de amplitud (AM) y de frecuencia (FM). Amplificación de la señal modulada de RF.

12-147
13-149

Lección 20. Receptores de radio 13-150

Generalidades. Receptores sencillos. Sistemas de antena y de tierra. Construcción de bobinas de antena. Características generales. Selectividad. Sensibilidad.Figura de ruido. Estabilidad. Fidelidad. Rango dinámico.

Lección 21. El receptor superheterodino 155 - 13

Generalidades. Heterodinación y FI. Frecuencias intermedias en AM y FM. La superheterodinación y sus ventajas. El receptor superheterodino AM.

13-157

Lección 22. La etapa detectora 14-161

Generalidades. Descripción del detector del radio AM CEKIT.

Lección 23. Etapas amplificadoras de FI 164 - 14

Generalidades.Amplificadores sintonizados. Descripción de los amplificadores de FI del radio AM CEKIT.

167 - 15

Lección 24. Etapa amplificadora de RF 178 - 15

Generalidades. Descripción de la etapa de antena del radio AM CEKIT. Sintonización de otras bandas de AM. Descripción del amplificador de RF del radio AM CEKIT. Amplificadores de RF con FET. Amplificadores de RF en cascada. Sintonización sincrónica y escalonada de amplificadores de RF. Neutralización de amplificadores de RF.

15-184

Lección 25. Circuitos osciladores 16-185

Osciladores Colpitts, Clapp y Hartley. Descripción del oscilador local del radio AM CEKIT. Importancia de la estabilidad de frecuencia en los osciladores de RF. Osciladores a cristal. VXOs y VFOs. Sintetizadores heterodinos de frecuencia. VCOs y PLLs.

196-16

Lección 26. Mezcladores de RF 17-197

Generalidades. Mezcladores de transmisión y recepción. Descripción del conversor de RF del radio AM-CEKIT. Características. Mezcladores con transistores y FETs.

Lección 27. Control Automático de Ganancia (CAG) 17-205

Generalidades.Descripción del CAG del radio AM-CEKIT. Filtrado de la tensión de CAG.

17-207

Lección 28. Descripción general del receptor de AM CEKIT 18-209

Transmisión y recepción. Los circuitos del receptor AM CEKIT. Sintonizador. Amplificador/Conversor de RF. Primer y segundo amplificador de FI. Detector. CAG. Amplificador de audio. Filtro de RF. Diagrama esquemático general.

18-216

Lección 29. El sistema de FM 19-221

Generalidades.Orígenes.Ventajas y desventajas. Fundamentos. Bandas laterales, índice de modulación y ancho de banda. Métodos de modulación de frecuencia. Transmisores y receptores de FM. El sistema FM estéreo.

19-232

Lección 30. Amplificadores de RF 20-238

Generalidades. Consideraciones de diseño. Operación del amplificador de RF del radio FM.

20-244

Lección 31. Conversores de FI para FM 245 -31

Generalidades.Operación del conversor de FI del radio FM. Frecuencias imágenes y espúreas.

Lección 32. Amplificadores de FI 21-251

Generalidades. Operación de los amplificadores de FI del radio FM CEKIT. Receptores de FM de doble conversión.

21-256

Lección 33. Detectores de FM 22-257

Generalidades.Tipos de detectores de FM. Detectores de pendiente, balanceados, de Foster-Seeley y de relación. El detector de relación como limitador de amplitud. Descripción del detector de relación del radio FM-CEKIT. Otros diseños. Demoduladores de FM con PLL 22-265

Lección 34. Control Automático de Frecuencia (CAF) 22-266

Generalidades. Operación del circuito de CAF del radio FM-CEKIT. 22-268

Lección 35. Operación completa y ensamble del Radio FM CEKIT 23-269

Generalidades. El receptor en bloques y en detalle. Ensamble. Fuente de alimentación. Amplificador de audio. Detector de relación. Amplificadores de FI. Conversor de FI y CAF. Amplificador de RF. 23-280

Lección 36. Fuentes de poder para radios AM y FM 24-281

Generalidades. Fuentes de poder no reguladas. Transformador de potencia. Rectificadores. Filtros. Variación del voltaje de salida en fuentes no reguladas. Fuentes reguladas. Reguladores con diodos zener y transistores. Circuitos integrados reguladores de voltaje.

Lección 37. Los circuitos integrados 24-289

Generalidades. Clasificación. Ventajas y limitaciones. Fabricación. El amplificador operacional.

Lección 38. Circuitos amplificadores de audio para radios AM y FM con CI 301-25

Generalidades. Descripción de los circuitos integrados LM380, LM386, LM384, LM383, LM377 y LM378. Aplicaciones típicas. 304-25

Lección 39. Otros CI para AM y FM 305-26

Generalidades. Receptor de AM ZN414Z y otros. Receptor de FM TDA7000 y otros. Receptor de AM/FM LM1868 y otros. Fundamentos de la recepción de FM estéreo. Demodulador de FM estéreo CA3090 y otros.

Lección 40. Reparación de radios AM 324-27

Generalidades. Metodología y procedimientos de reparación. Identificación de secciones, etapas, circuitos y componentes. Diagnóstico, localización, aislamiento y sustitución de fallas. Reparación de radios AM. Fallas típicas y correcciones.

Lección 41. Reparación de radios FM. 28-331

Generalidades. Pruebas básicas. Pruebas con generadores y trazadores de señal.

Indice general del Curso 338

INDICE DE ACTIVIDADES PRACTICAS

T

Nº 1. Fijación del plano de montaje. 16-1

Nº 2. Instalación de las puntillas de soporte. 28-2

Nº 3. Aprendiendo a soldar. 40-3

Nº 4. Alambrando el circuito de potencia. 52-4

Nº 5. Instalación de R12, R13, R14 y R15. 64-5

Nº 6. Instalación de C11, C12, C13 y C14. 76-6

Nº 7. Instalación de D2 y D3. 100-8

Nº 8. Completando las conexiones del amplificador de audio. 112-9

Nº 9. Instalación de Q4, Q5 y Q6. 124-10

Nº 10. Ensamble de un inyector de señales. 136-11

Nº 11. Manejo del generador de audio. 148-12

Nº 12. Manejo del osciloscopio. 158-13
160-13

Nº 13. Ensamble final y análisis general del amplificador de audio. 173-15- 167-14
172-14

Nº 14. Instalación de C8, C9, R11 y D1. 196-16
Prueba del detector. Instalación de C10. 208-17

Nº 15. Armando un radio de galena. 217-18

Nº 16. Instalación de Q3 y R10. 232-19
Instalación de R8, R9, C6 y C7 233-20

Nº 17. Ensamble de un generador de FI. 233-20

Nº 18. Instalación de R7, C4 y C5. 256-21
Instalación de Q2, R6 y C3 268-22

Nº 19. Instalación de R1, R2, R3, R4 y R5 280-23
Instalación de C1, C2 y Q1 292-24

Nº 20. Análisis, prueba y ajuste de la segunda etapa de FI. Instalación de T3. 293-25

Nº 21. Análisis, prueba y ajuste de la primera etapa de FI. Instalación de T2. 298-25

Nº 22. Instalación de L2 y T1. 316-26

Nº 23. Instalación de L1, CV1, CV2, CT1 y CT2. 317-27

Nº 24. Manejo del generador de RF. 317-27

Nº 25. Análisis y prueba del conversor de RF. 27-319-2

Nº 26. Instalación del parlante. Alineamiento final del radio. 336-28

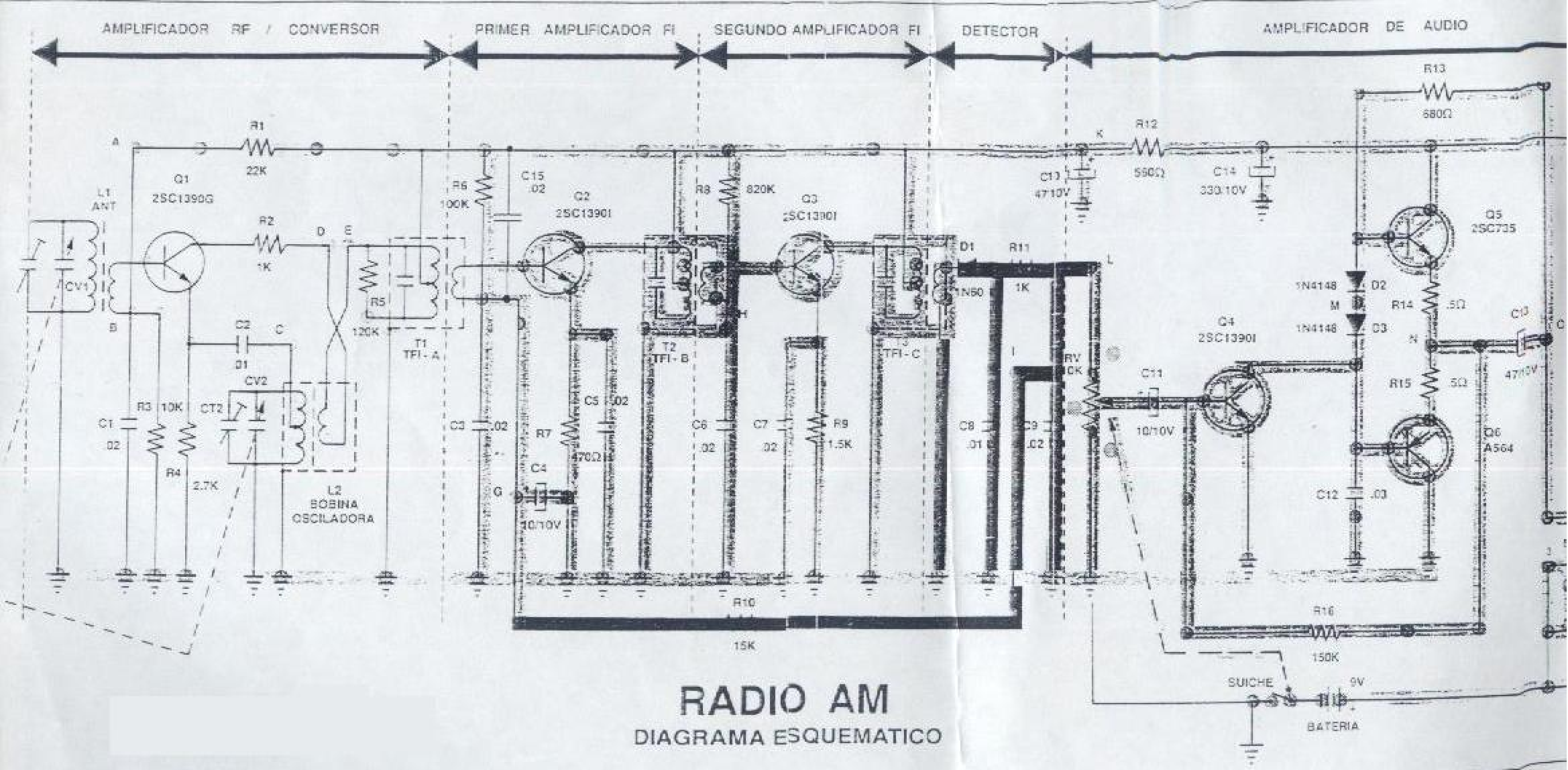
AMPLIFICADOR RF / CONVERSOR
PRIMER AMPLIFICADOR FI
SEGUNDO AMPLIFICADOR FI
DETECTOR
AMPLIFICADOR DE AUDIO
A
R1
22K
L1
ANT
CV1
B
Q1
2SC1390G
R2
1K
D
E
C2
D1
C
CV2
CT2
C1
.02
R3
10K
R4
2.7K
L2
BOBINA
OSCILADORA
R5
120K
T1
TFI - A
R6
100K
C15
.02
Q2
2SC1390I
C3
.02
R7
470Ω
C5
.02
C4
G
10/10V
R8
820K
T2
TFI - B
H
Q3
2SC1390I
C6
.02
C7
.02
R9
1.5K
R10
15K
T3
TFI - C
D1
1N60
R11
1K
L
I
C8
.01
C9
.02
RV
10K
K
C13
47/10V
R12
560Ω
C14
330/10V
R13
680Ω
Q5
2SC735
1N4148
D2
M
1N4148
D3
R14
.5Ω
N
Q4
2SC1390I
C11
10/10V
R15
.5Ω
Q6
A564
C12
.03
C10
47/10V
R16
150K
SUICHE
9V
BATERIA

RADIO AM
DIAGRAMA ESQUEMATICO

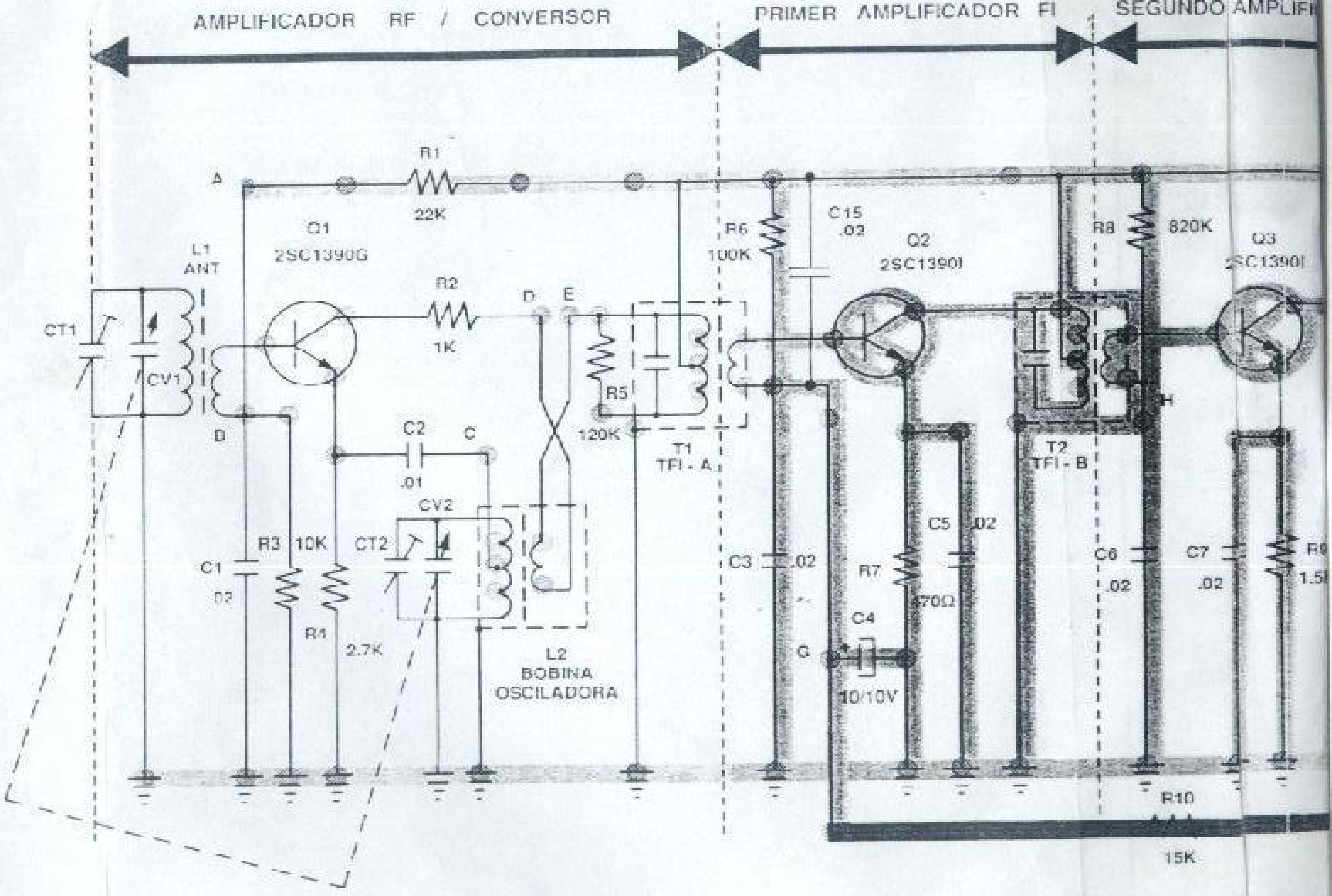
AMPLIFICADOR RF / CONVERSOR
PRIMER AMPLIFICADOR FI
SEGUNDO AMPLIF
A
R1
22K
Q1
2SC1390G
L1
ANT
CT1
CV1
D
R2
1K
D
E
C2
.01
C
CV2
CT2
C1
.02
R3 10K
R1
2.7K
L2
BOBINA
OSCILADORA
R5
120K
T1
TFI - A
R6
100K
C15
.02
Q2
2SC1390I
C3
.02
G
C4
10/10V
R7
470Ω
C5 .02
R8
820K
T2
TFI - B
H
Q3
2SC1390I
C6
.02
C7
.02
R9
1.5
R10
15K

RADIO A

DIAGRAMA ESQU

CEKIT S.A.

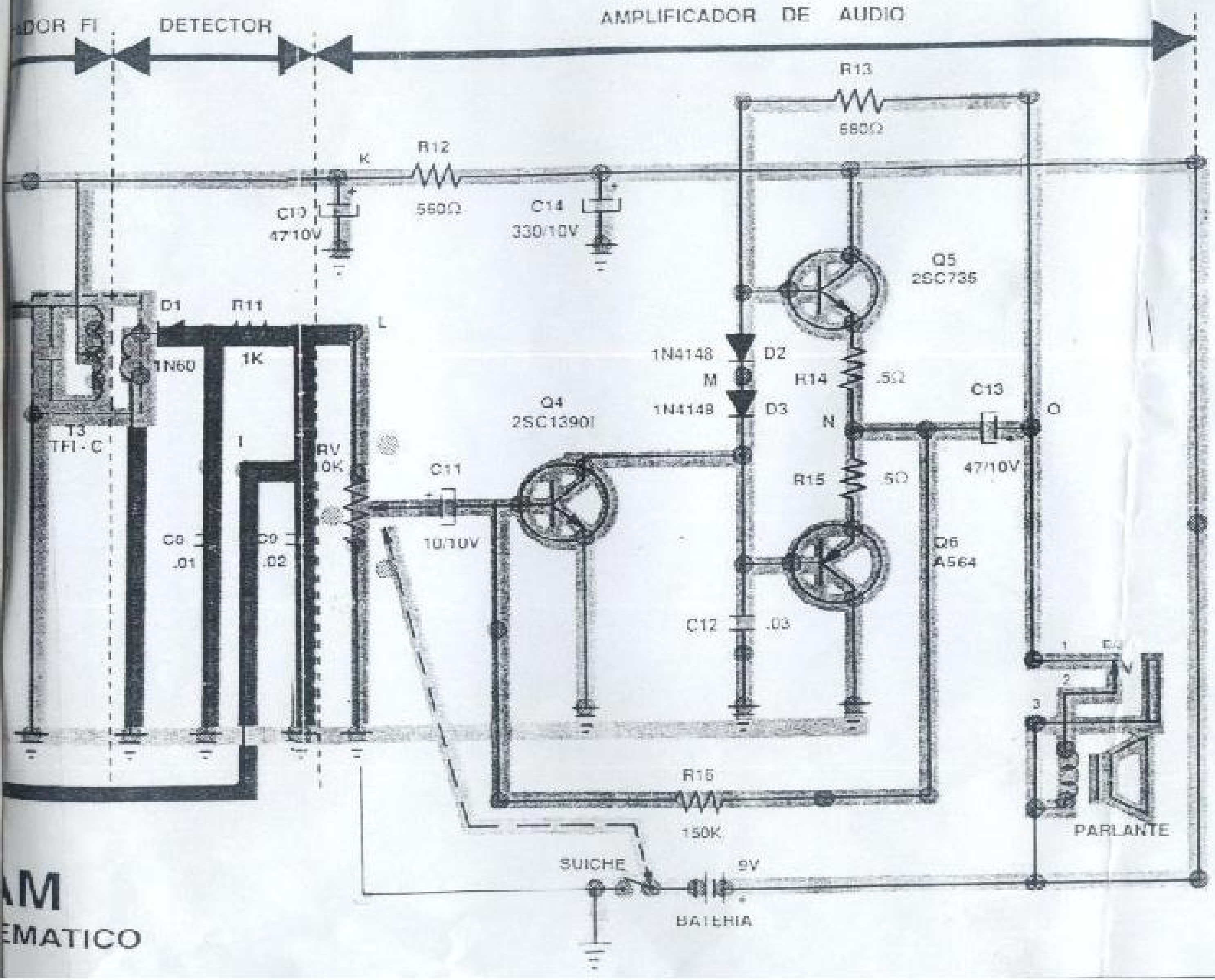
ADOR FI
DETECTOR
AMPLIFICADOR DE AUDIO
R13
680Ω
K
R12
560Ω
C10
47/10V
C14
330/10V
Q5
2SC735
D1
R11
1N60
1K
L
T3
TFI - C
1N4148
D2
M
1N4148
D3
R14
.5Ω
C13
O
N
47/10V
Q4
2SC1390I
RV
10K
C11
10/10V
C8
.01
C9
.02
R15
.5Ω
Q6
A564
C12
.03
1
2
3
R15
150K
PARLANTE
SUICHE
9V
BATERIA

AM

EMATICO

An FM Crystal Set

Updated 2010-11-28:

Since this page was first posted in June 2010, I have made a number of improvements to the receiver described below. I've also received a few requests for additional information. In reviewing this page, I realized that it was long overdue for an update. The original information is unchanged. The updated information appears at the bottom of the page.

Introduction

Crystal sets–totally passive receivers without any amplification–have been popular for many years. The vast majority of them are designed and built for the Medium Wave band, 530 to 1700 kHz. However, there have also been some built for the shortwave bands, and occasionally some for the FM band (88 to 108 MHz).

Having several local powerful FM stations, I'd occasionally noticed that I was picking up FM on a medium wave crystal set. Considering that there was a strong enough signal to receive FM when it wasn't wanted, it should be possible to build a receiver to pick up the signal intentionally.

Most FM crystal sets use slope detection. That is, they are tuned slightly off the centre frequency of the station, and as the frequency modulated carrier moves in and out of the receiver's resonant frequency, an audio signal approximating the original modulation is produced. True FM detection requires a frequency discriminator circuit, and the complexity generally makes it unsuitable for a passive receiver. I'm only aware of one such project. It was designed and built by Edward Richley, and was written up in the Xtal Set Society Newsletter, January and March 1996. In order to achieve the high Q required, the circuit used a coaxial resonator constructed from 2 inch copper pipe. It was ingenious in that, while most FM discriminators have two tuned circuits, this one only required a single tuned circuit. This was an important feature, because trying to adjust two tuned circuits when tuning to different stations, would be extremely difficult, if not impossible.

I've redrawn the Richley circuit here:

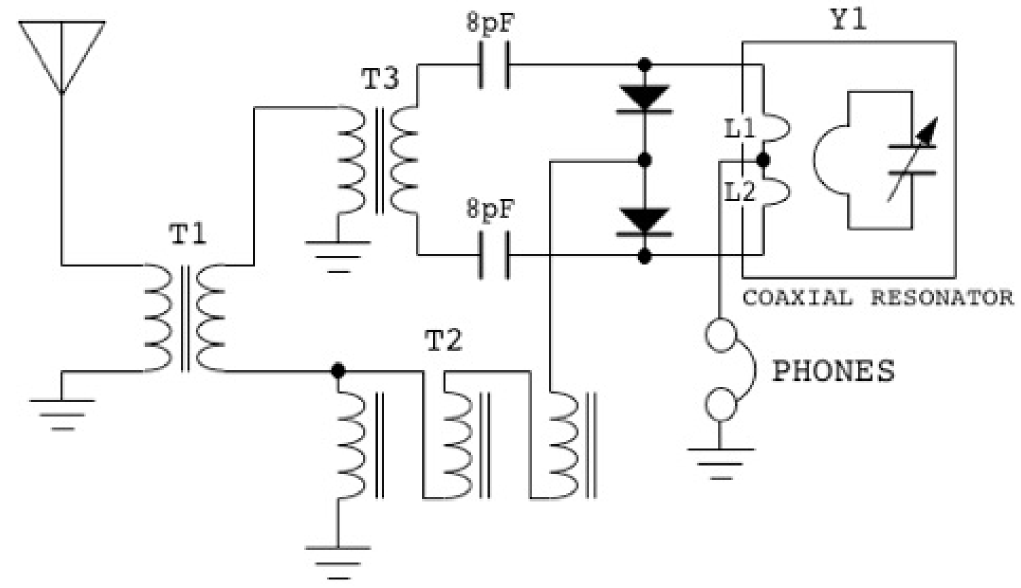

COAXIAL RESONATOR Y1:
CENTRE CONDUCTOR: COPPER TUBING 0.5" DIAMETER, 25.5" LONG;
SHIELD: COPPER PIPE 2" DIAMETER, 29.5" LONG;
TUNING BY MEANS OF MOVABLE SLUG IN TOP OF RESONATOR;
L1 & L2: SINGLE TURN WIRE LOOPS 2.75" LONG X 0.5" WIDE MOUNTED AT BASE END OF RESONATOR ON OPPOSITE SIDES OF CENTRE CONDUCTOR.

A very brief description of the circuit follows. For a more thorough discussion, please refer to the original newsletter articles.

As described in the diagram notes, the resonator is constructed from a two foot piece of 2 inch diameter copper pipe, with a centre conductor of 1/2 inch copper pipe. The resonator is designed for an unloaded Q of 2000, and a loaded Q of about 500, which is required for acceptable selectivity. FM channel bandwidth is 200 kHz, or 0.2 MHz. Therefore, at a frequency of 100 MHz, required Q will be 100/0.2=500.

Transformers T1, T2, and T3 are made from standard 300:75 ohm baluns which are included in the kit of parts that come with virtually every new television or radio. T1 is used unmodified. T2 is rewound as 1:3 autotransformer. T3 has a two turn primary and a two turn secondary. Transformer T1 converts the incoming unbalanced signal to a balanced signal which then feeds T2 and T3. The secondary of T3 is loosely coupled via the 8 pF capacitors to the resonator coupling loops, which develop a 90° phase shifted signal due to the normal behaviour of a coupled circuit at resonance. The secondary of T2 is connected to the common point between the diodes. This signal is 180° out of phase with the incoming RF. Hence, there are quadrature signals applied across the diodes which act as a synchronous detector. As the frequency modulated radio carrier moves above and below the resonant frequency, the phase shift varies above and below 90°. The detector converts this phase shift into an audio frequency signal which matches the original audio source.

Initially, I planned to build a copy of this receiver, but after doing a bit of research, I concluded that a helical resonator would be more practical at these frequencies. My new plan was to duplicate this circuit, except that I would substitute a helical resonator for the coaxial resonator.

The first step then was to design a helical resonator with an unloaded Q of 2000 at a resonant frequency of 110 MHz, the top of the FM band. Using standard helical resonator design formula, the resulting resonator has an inner coil constructed from 4.3 turns of 9.5mm (3/8") copper tubing. Coil outside diameter is 65mm (2.6"), and coil length is 84mm (3.3").

The shield is constructed from copper sheet, and is 84mm (3.3") square by 200mm (8") long. It is housed inside a wood enclosure for rigidity. The shield is about 2" longer than required to allow room for coupling loops at the bottom, and for a tuning assembly at the top. The photo to the right shows the shield parts prior to assembly. The tuning mechanism is a small plate attached to a length of brass tubing which runs through the end cap. The rod is moved in and out by means of a screw. This changes the fringing capacitance at the open end of the coil. The closer the plate is to the coil, the lower the resonant frequency. The next photo shows

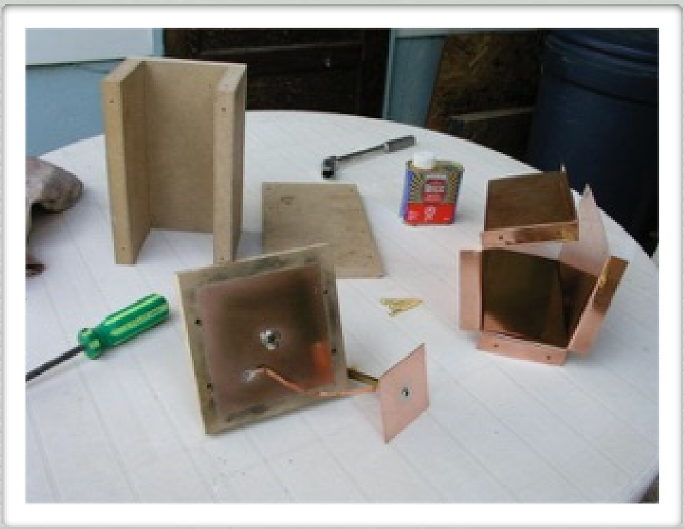

the top cap with tuning mechanism assembled. The tubing has a square cross section to prevent the assembly from turning. It slides inside a slightly larger piece of square tubing. This type of tubing is available at hobby shops and comes in a range of sizes suitable for close sliding fits. The coil and the inside of the shield were polished to remove oxidization, and hence maximize Q. The partially assembled resonator is shown in the next two photos. The photo on the left is of the top end, and the photo on the right is of the bottom end.

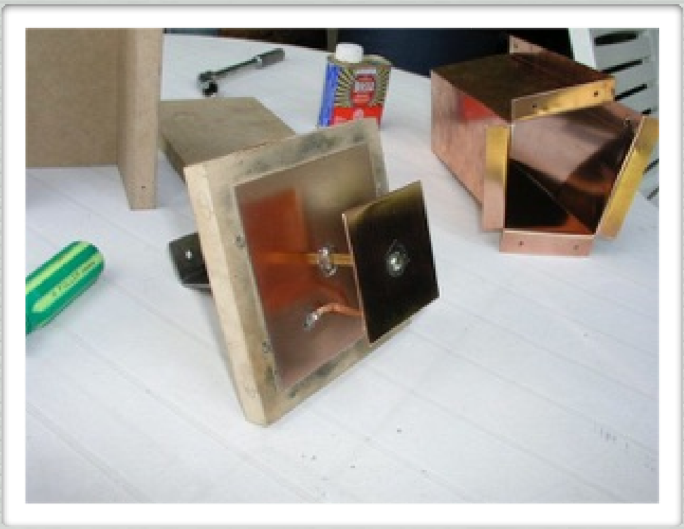

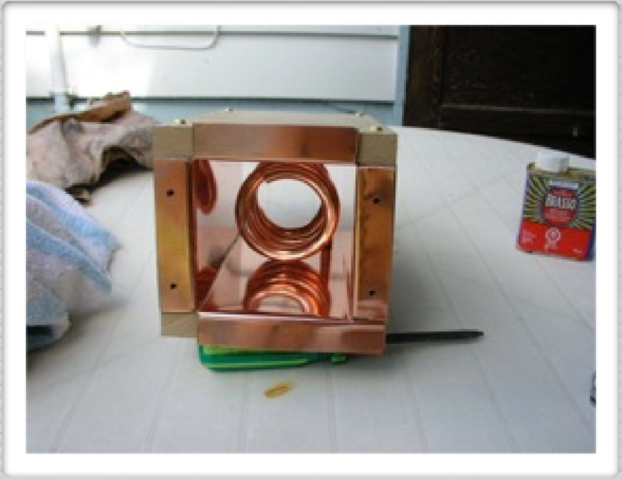

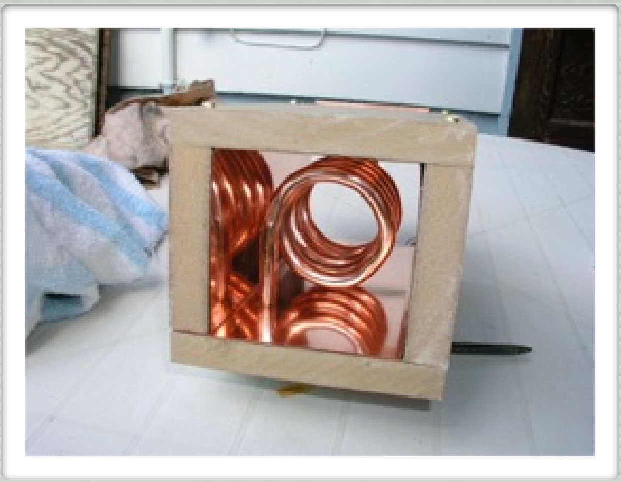

The next photo shows the top cap installed and the exterior part of the tuning mechanism:

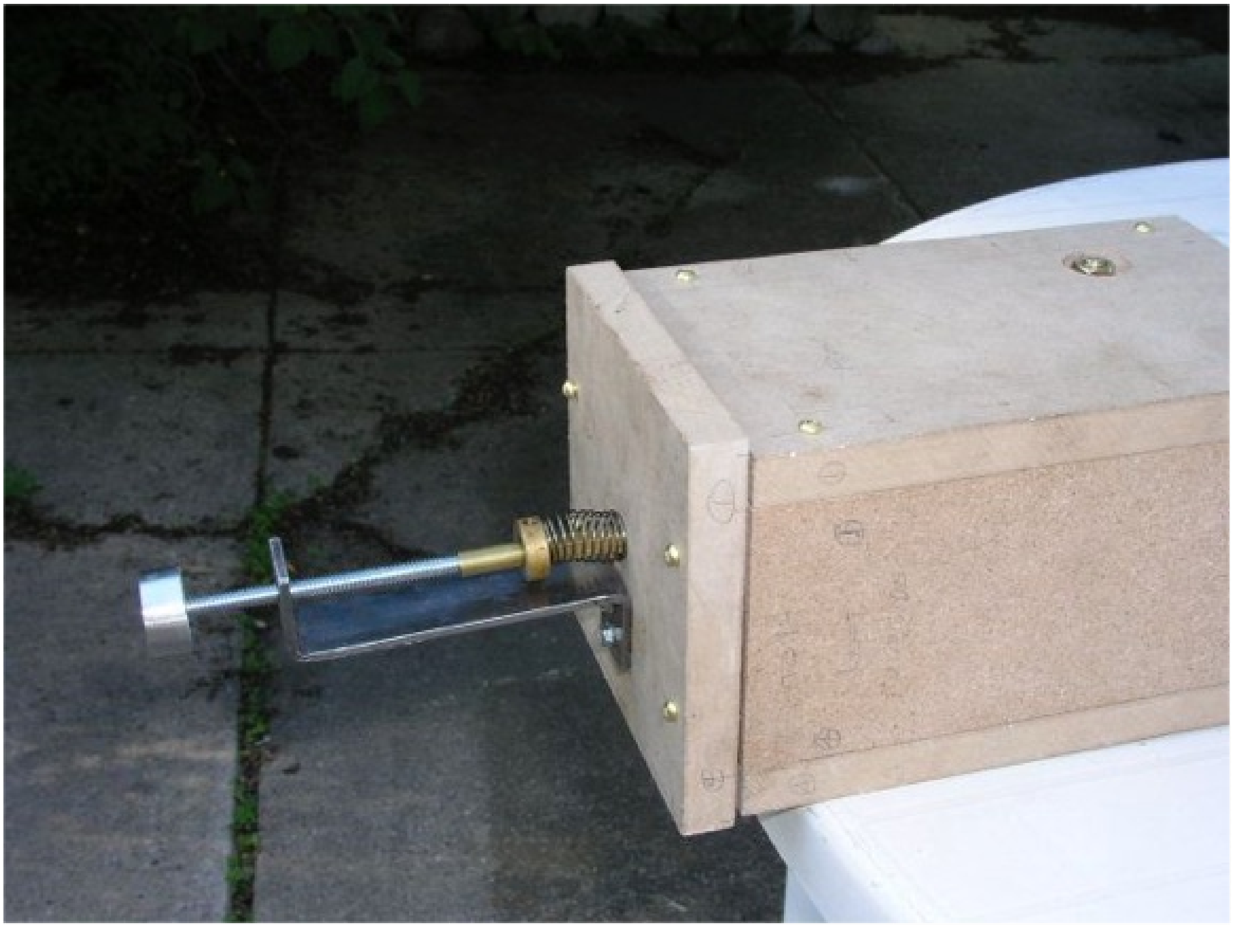

I decided that for initial testing, it would be best to keep things simple and use a slope detector circuit. This is shown in the following schematic:

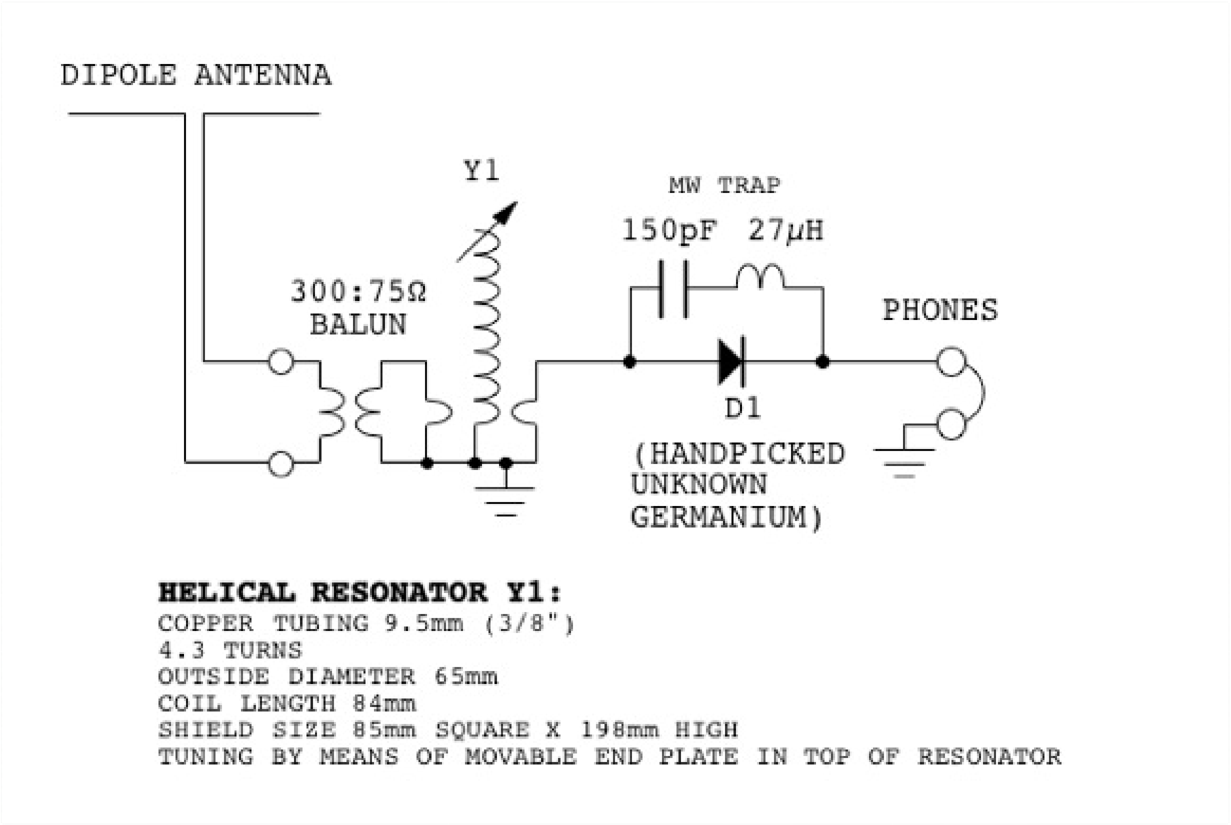
DIPOLE ANTENNA
Y1
MW TRAP
150pF 27μH
300:75Ω
BALUN
PHONES
D1
(HANDPICKED
UNKNOWN
GERMANIUM)
HELICAL RESONATOR Y1:
COPPER TUBING 9.5mm (3/8")
4.3 TURNS
OUTSIDE DIAMETER 65mm
COIL LENGTH 84mm
SHIELD SIZE 85mm SQUARE X 198mm HIGH
TUNING BY MEANS OF MOVABLE END PLATE IN TOP OF RESONATOR

The 27uH choke and 150pF cap are a low pass filter to trap out a local high powered AM station that was causing some interference problems. For a diode, I used a 1N34A initially, but found another unknown type in my parts box which worked slightly better. It may in fact also be a 1N34A from a different manufacturer. The detector and coupling loops were built on a small circuit board and mounted on a block of wood so that the assembly could be repositioned easily to obtain the best performance. The assembly is shown below.

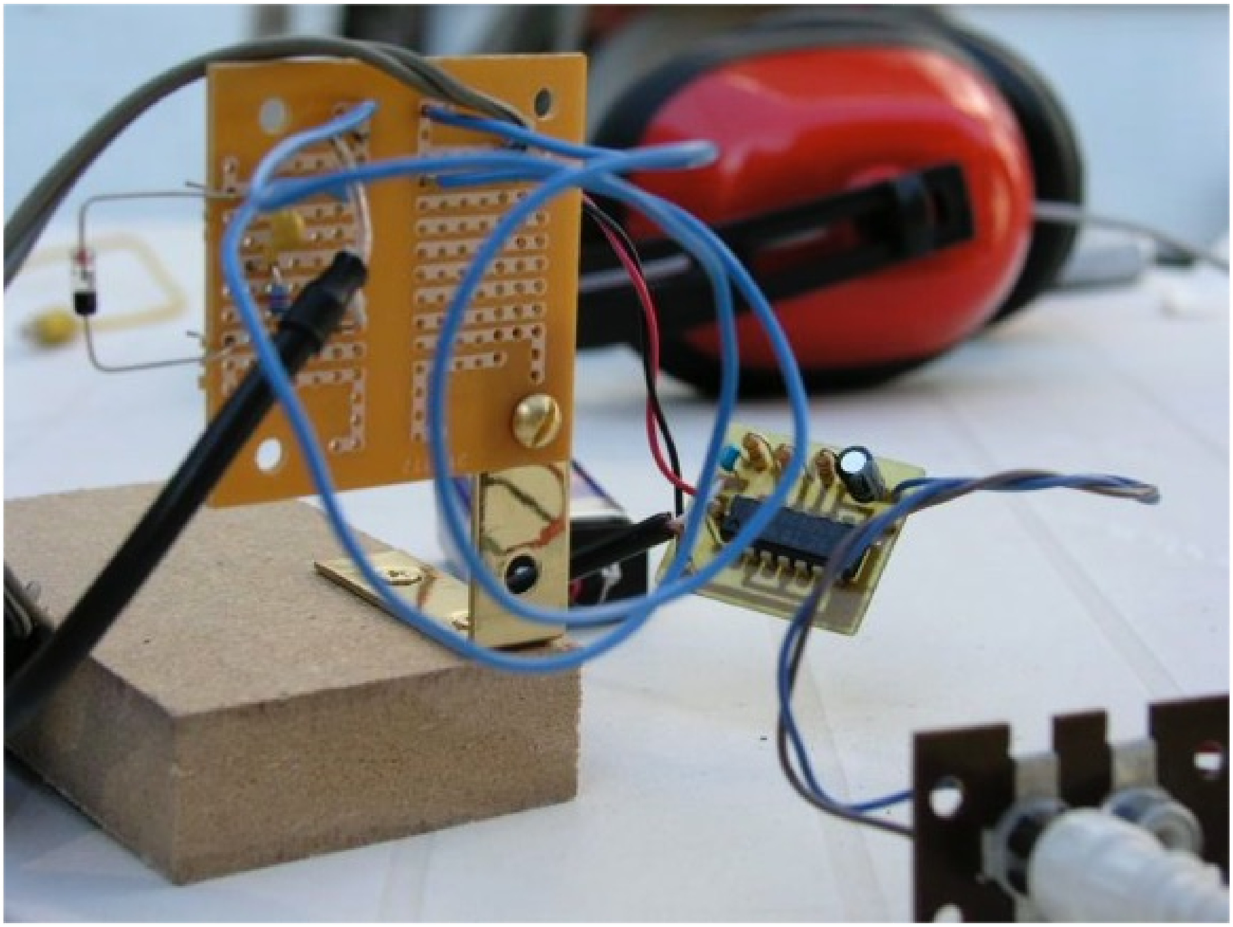

The RF and detector coupling loops are blue hook-up wire. The little postage stamp sized circuit board is a low power headphone amplifier which was used in conjunction with my relatively insensitive headphones to make testing a bit more painless. The initial working set-up is shown below:

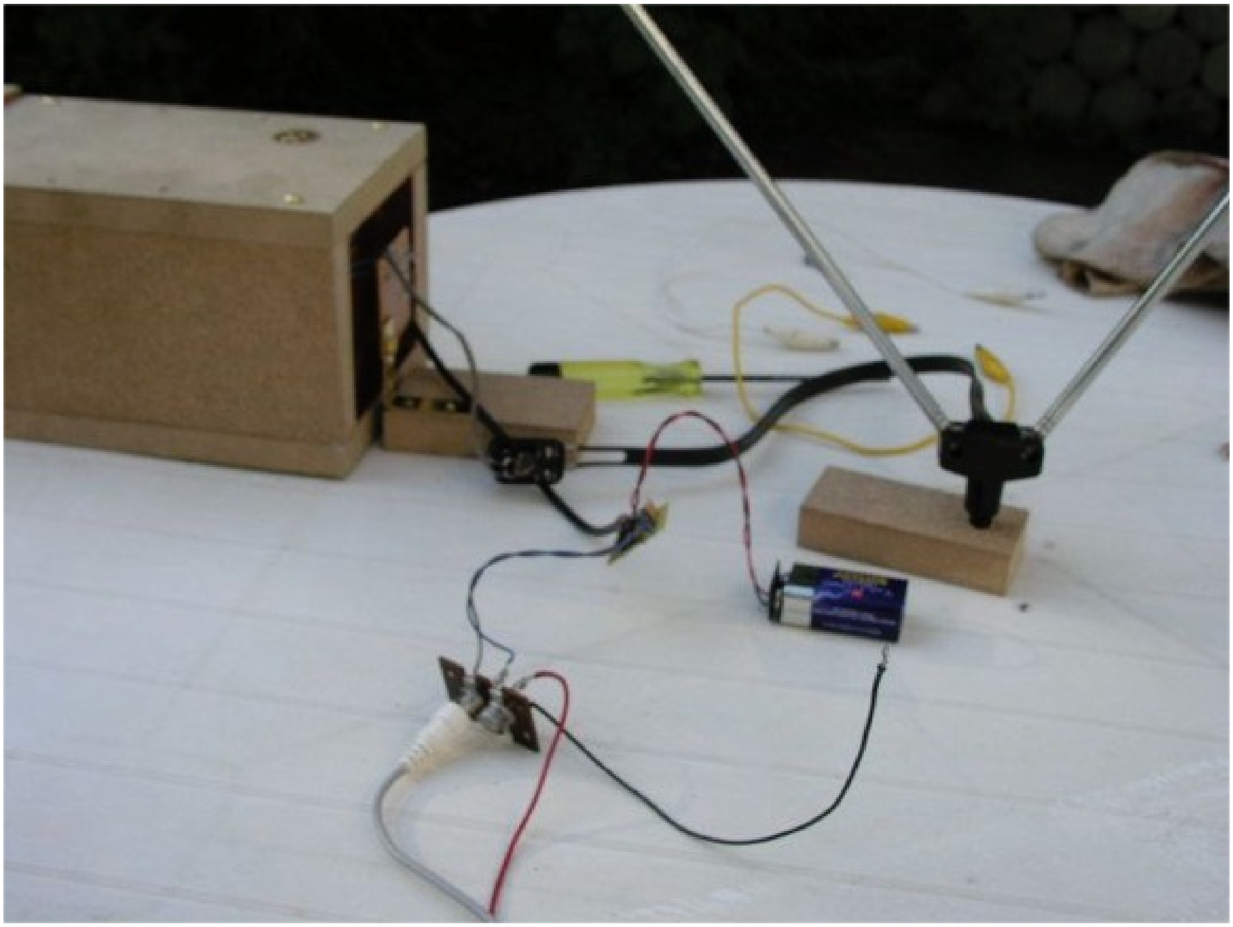

Reception was fairly good. The signal was weak using my piezo phones without any amplification. So, I connected in the headphone amplifier, and did all the remaining testing with it. Selectivity was very good. However, I don't have any easy means to measure the Q. I found that I could easily pick either side of a station's centre frequency for detection, and if I tuned dead centre, the signal became very distorted. That leads me to believe that the Q is very high. I couldn't get that kind of precision with earlier experimental sets which used a coil/capacitor tuned circuit.

So with that bit of success, I started thinking about making a true discriminator. As mentioned above, I had intended to duplicate Ed Richley's circuit except that the coaxial resonator would be replaced with the helical one. However, after reviewing the original circuit, I concluded that two of the transformers could be eliminated. The first transformer is a balun to convert from an unbalanced RF signal from the antenna to a balanced one. Since I intended to use a dipole antenna, this was unnecessary.

Rather than excite the resonator via coupling loops through the small capacitors, I decided to drive the resonator directly at a tap near the ground end of the resonator coil. This eliminated the need for the balanced secondary winding on T3.

I decided that I could combine T2 and T3 into a single transformer. I had disassembled several baluns and found some to have two-hole cores and some were simple toroids. I used the toroid from one of these and wound a 4-turn primary, and 4-turn secondary. The secondary has a tap after the first turn which is grounded. The schematic follows:

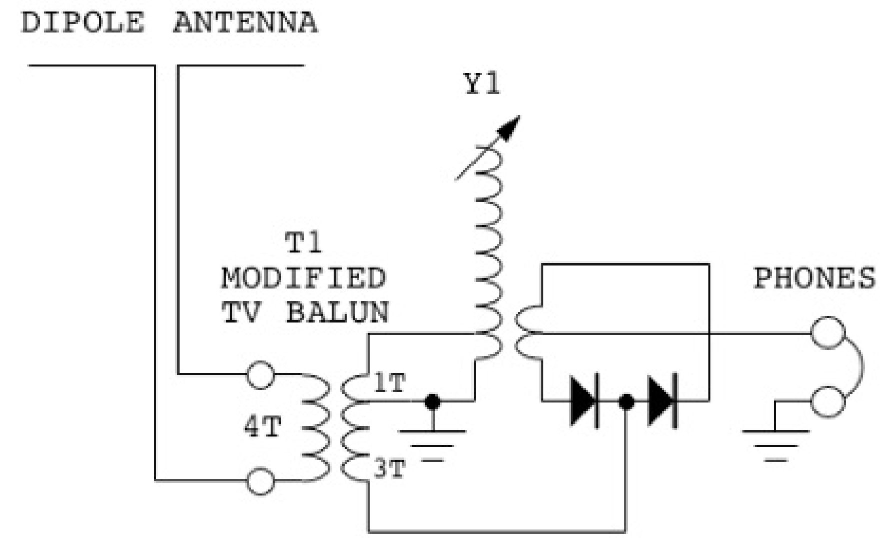

HELICAL RESONATOR Y1:
COPPER TUBING 9.5mm (3/8")
4.3 TURNS
OUTSIDE DIAMETER 65mm
COIL LENGTH 84mm
SHIELD SIZE 85mm SQUARE X 198mm HIGH
TUNING BY MEANS OF MOVABLE END PLATE IN TOP OF RESONATOR

The secondary arrangement still gives the same 3:1 power ratio between the diodes and the resonator excitation, but the circuit is much simpler. The revised coupler/detector board is shown below:

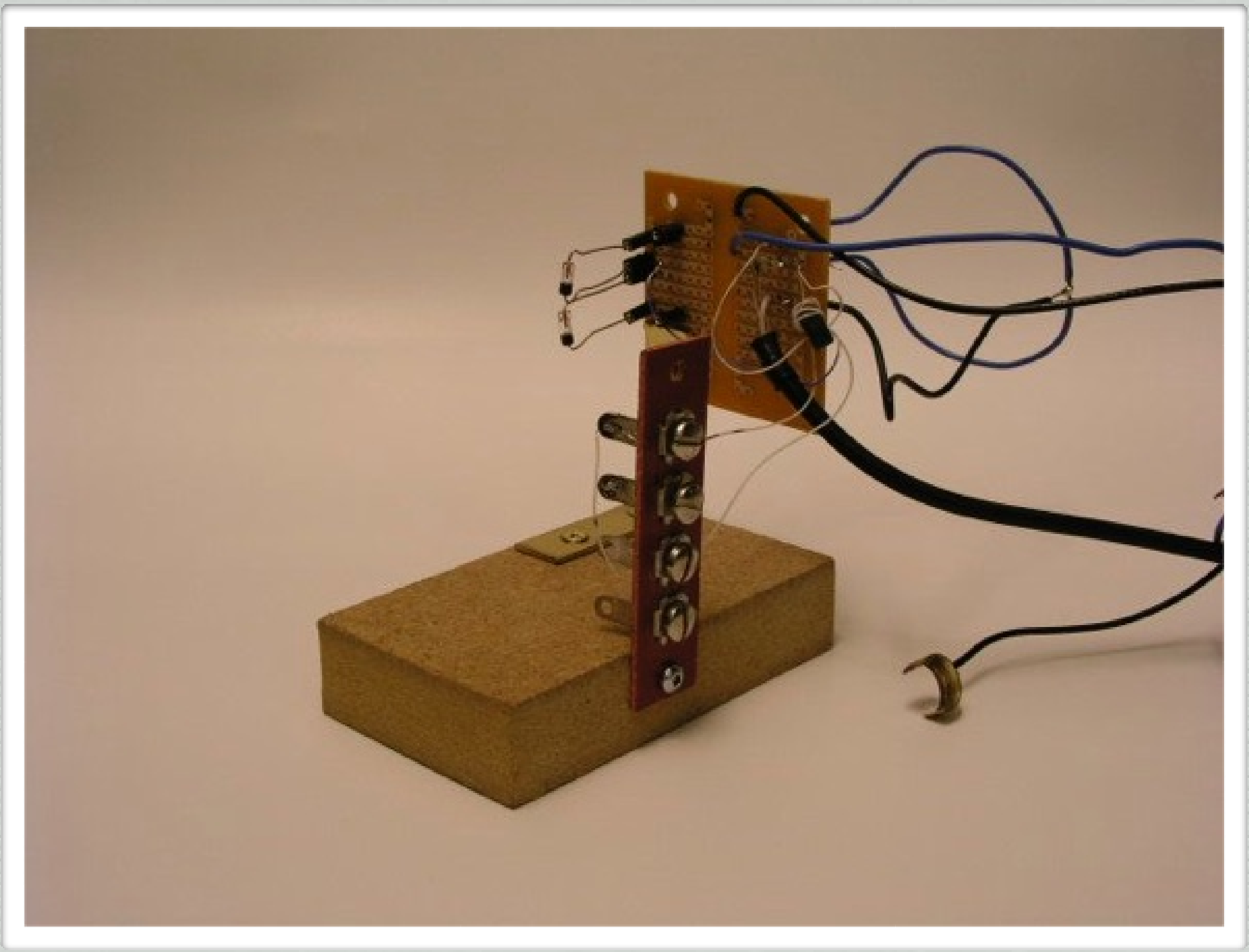

The black wire with the brass clip is for connecting to a tap on the resonator coil. I made it this way so that I could quickly try different positions. Optimum turns out to be about 40mm (1.5") from ground.

The pickup loop is the single loop of blue wire with the audio output tap at the halfway position (black wire).

Diodes are socket mounted so that different ones may be tried without fuss. I didn't put in an AM trap this time, because the circuit arrangement is not conducive to AM pickup, since the diodes are essentially shorted at MW frequencies by the pickup loop.

Here's a brief description of how the circuit works:

Input transformer T1 splits the incoming RF (untuned) into two signals 180 degrees out of phase with reference to ground. The low level side goes to the resonator and the high level side goes to the common point of the diodes. The voltage induced by the resonator into the pickup loop will be 90 degrees out of phase with the the input signal. This is the nature of a loosely coupled tuned circuit at resonance. The phase of the signal on the pickup coil will vary on either side of 90 degrees as the carrier is frequency modulated. The "90 degree" signal is applied to the outside ends of the diodes while the 180 degree signal is applied to the diode common point which then act as a quadrature detector. Hence it is essentially the same mode of operation as Richley's original circuit.

The following photo shows the discriminator assembly in position:

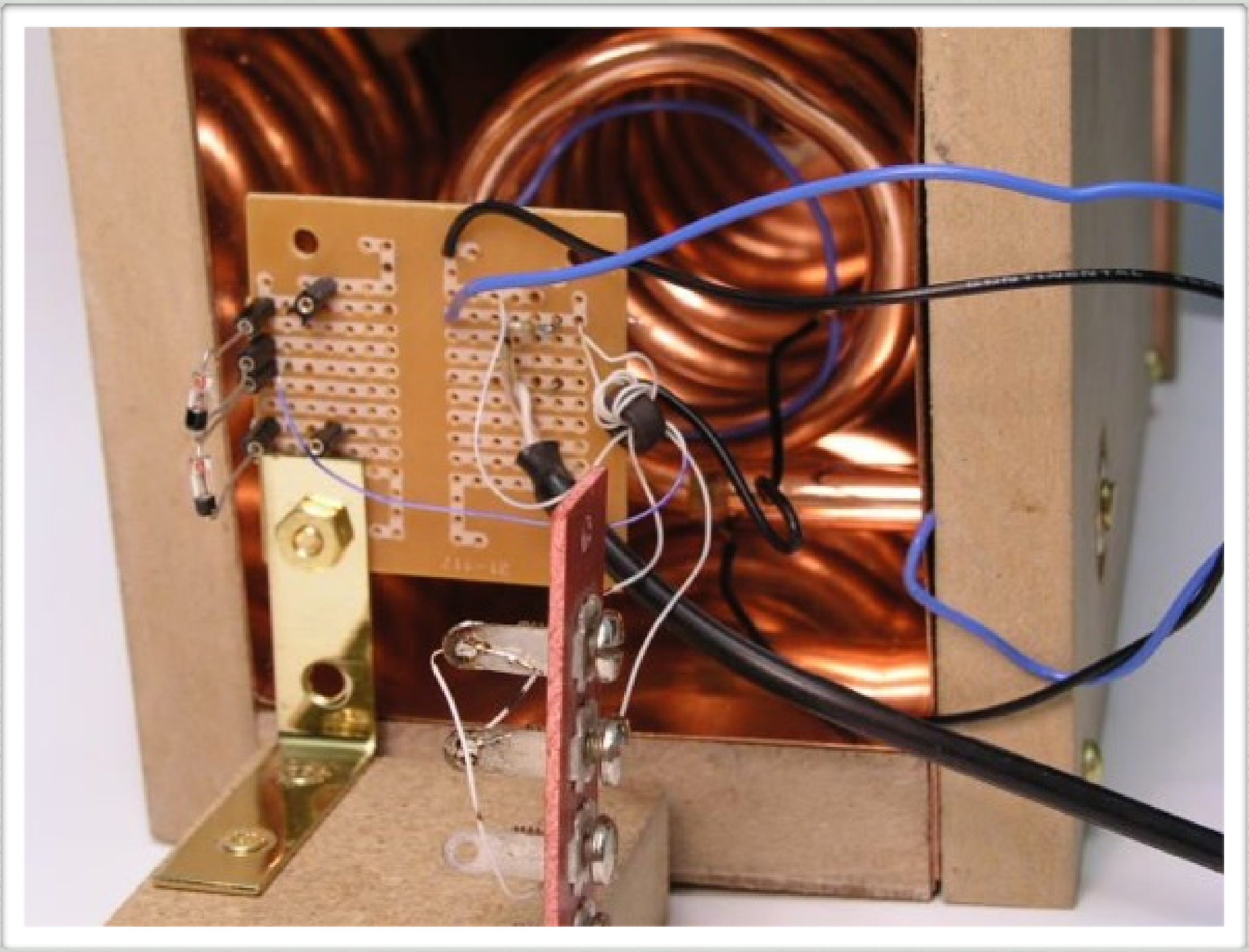

Now, with everything connected up, and after finding the best position for the pickup coil I found that the discriminator appears to be working correctly. Tuning is sharper, and there are no double slopes to select as I could previously. In fact tuning this receiver has very much the same feel as tuning a commercially made FM receiver. Audio level has increased significantly too. This is something I had been wondering about. I was concerned that a true discriminator might be less sensitive than a slope detector, but it appears not to be the case.

2010-11-28 Update:

Since the original material first appeared on this page, I've continued to tinker with this circuit, resulting in several incremental improvements. Although the improvements came in many small increments, the cumulative result is that this is now a very good performing receiver. Improvements listed in order of importance (most important first) are:

- Changing the detector diodes to Avago HSMS-2850;
- Changing the turns ratio on the RF input transformer;
- Getting a good set of sound powered headphones;
- Permanently mounting the detector loop;
- Finding a good antenna location;
- Shortening the resonator helix.

Diodes

The Avago (formerly Agilent) HSMS-2850 diodes are rather odd ones. I had purchased these along with a number of other different types of detector diodes, more or less at random, to test their effectiveness. These particular diodes were described as "**zero bias**" which made them too intriguing to pass up. When they arrived, I set them aside and didn't think about them again until I ran across this page on Dick Kleijer's site. He had also built an FM crystal set and had tested a number of different diodes. Among them were the HSMS-2850's which turned out to be his best performers. I dug through my diode assortment and confirmed that they were the same ones I had. These are surface mount, and so I had to construct a small printed circuit adapter board. Otherwise, I would have tried them much sooner. When I tried them, there was a considerable improvement in performance. I mentioned that these diodes are odd ones. They have a very low R_0 (zero crossing resistance) of about 5000 ohms (many thanks to Mike Tuggle who measured them for me). Coincidentally, this is about the same R_0 as a galena detector. The low R_0 would normally make them a bad choice as a crystal set detector (galena aficionados notwithstanding). However, it appears that this particular resonator circuit has a very low resonant RF output resistance, which turns out to be a good match to the diodes. For more info on this, see Article #4 on Ben Tongue's site. This low RF resistance has another consequence which is discussed below.

In addition to the above mentioned diodes, I also recently tested Sanyo 1SS351 diodes. These are a surface mount type, with two diodes in a single package, and while not quite as sensitive as the HSMS-2850's they are significantly better than any of the germaniums or other Schottkys that I've tested. I also tried paralleling two of these, and got a further improvement in performance. In a listening test (which was completely subjective) I found it difficult to tell much difference between the parallel 1SS351's and the singleHSMS-2850's.

RF Input Transformer

Getting the best turns ratio for the RF input transformer was an iterative procedure which also affected the optimum input tap position on the resonator. The schematic appearing earlier on this page shows a primary with 4 turns, and a secondary with 4 turns, with a grounded tap at 1 turn. After much experimentation (and wire breakage), I currently have a primary with 6 turns, and a secondary with 8 turns, with the grounded tap again at 1 turn. This is the current schematic:

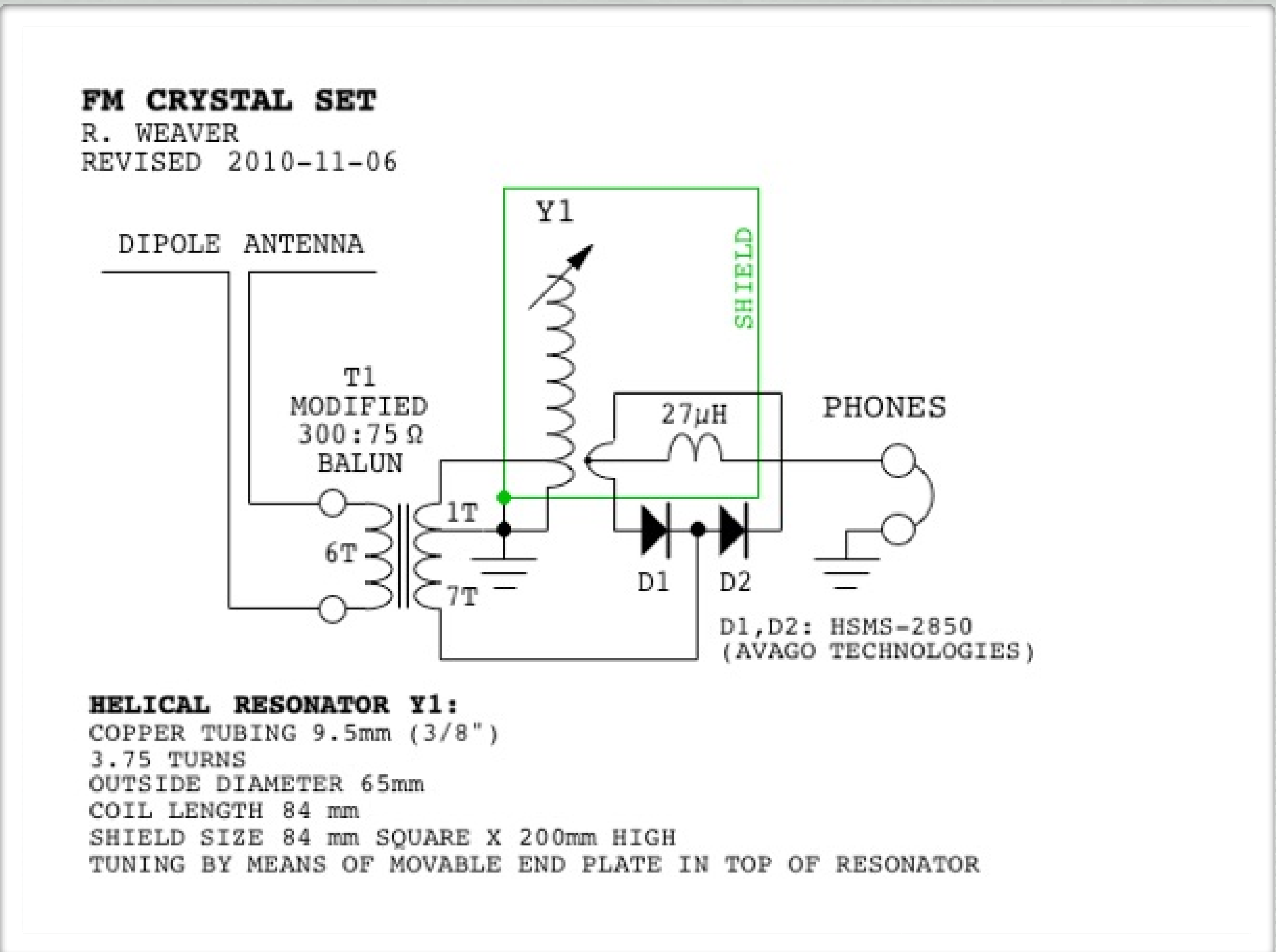
FM CRYSTAL SET
R. WEAVER
REVISED 2010-11-06
Y1
SHIELD
DIPOLE ANTENNA
T1
MODIFIED
300:75 Ω
BALUN
27μH
PHONES
1T
6T
7T
D1
D2
D1,D2: HSMS-2850
(AVAGO TECHNOLOGIES)
HELICAL RESONATOR Y1:
COPPER TUBING 9.5mm (3/8")
3.75 TURNS
OUTSIDE DIAMETER 65mm
COIL LENGTH 84 mm
SHIELD SIZE 84 mm SQUARE X 200mm HIGH
TUNING BY MEANS OF MOVABLE END PLATE IN TOP OF RESONATOR

I might have arrived at this turns ratio a bit sooner if I hadn't made a bit of a blunder, confusing a dipole antenna with a folded dipole antenna, and thinking my antenna impedance was 300 ohms when in fact it is 75 ohms. (You may pick up on this goof in the earlier text of this page, which I haven't corrected yet.) I'm not yet at the point where I will claim this transformer configuration to be optimum, but it is much improved over what I had before.

Another addition is the 27µH choke connected between the centre tap on the detector loop and the phones. This seems to help isolate the audio wiring from the RF part the circuit, and vice versa. This improvement is subjective at this point (it seems to increase volume), but regardless, it doesn't appear to be a bad idea.

There were a few other changes made to the circuit as well, but in the end, except for the diodes, the revised RF transformer windings, and the addition of the 27µH choke, the other changes resulted in negligible improvements. Hence, they were dropped.

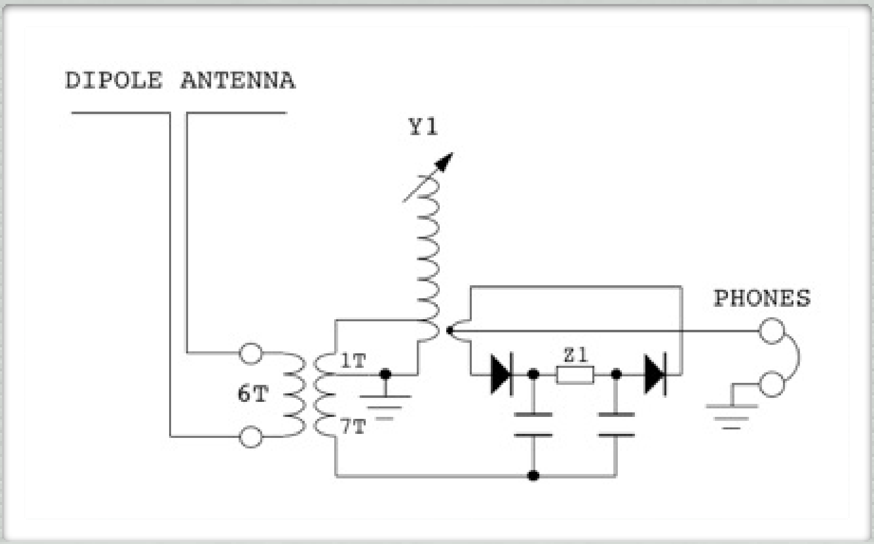

While on the subject of the schematic diagram, alert readers will note the apparent short circuit at audio frequencies if the diodes happen to be conducting. The consequence of this is that when receiving a very strong signal, there is an upper limit to how much audio the detector will produce. My earlier circuit simulation verified this, but I haven't yet encountered a real life signal strong enough to hit this limit. However, I have allowed for circuit modifications should this ever become a problem. It would involve the addition of a resistive or reactive component (Z1) between the diodes as shown in the diagram here. The circuit model used in the simulation does not yet have sufficiently accurate parameters for the diodes or other components, to give a good idea what values to use for Z1 or the coupling capacitors. My circuit board has several strategically located jumper plugs to allow for this future addition without the need for any soldering or de-soldering.

Sound Powered Phones

There's not much I can say about sound powered phones other than: If you're serious about crystal radios, eventually you'll have to get a pair. There is nothing else that compares in sensitivity. Piezo phones are not bad if you have nothing else available, but simply don't compare to sound powered phones . I was lucky enough to get a pair of Western Electric SP phones (thanks to Darryl Boyd), and with the other improvements that were made in the receiver, there is no longer any need for an amplifier, not even for testing purposes.

The Western Electric phones (model D173014) have an impedance of about 600 ohms per element, and I have them wired in series for a total impedance of 1200 ohms. This turns out to be such a good match for the low impedance detector, that I have not found any advantage to using an audio matching transformer. Originally, I tested these with a Bogen T725 transformer wired as an autotransformer. The best transformer configurations were the 2:1 and 4:1 ratios. However, I couldn't honestly detect any improvement over the direct connection using no transformer at all. I did notice a slight difference in frequency response (better bass with the transformer), but no net improvement in volume using the transformer. I'm guessing that the optimum impedance would be somewhere around 2400 to 3000 ohms, but the transformer's insertion loss probably negated any benefits from the better impedance match.

Detector Loop Mounting

Originally, as shown in the early photos above, I mounted the diodes and detector loop on a separate moveable base. Once I'd determined the optimum position for the loop, I mounted it permanently through the side of the enclosure. This eliminated quite a bit of excess wire which is never good to have when dealing with VHF. The RF transformer and diodes are now mounted on the side of the enclosure where the leads come through. Here is a detail of how the detector mounts:

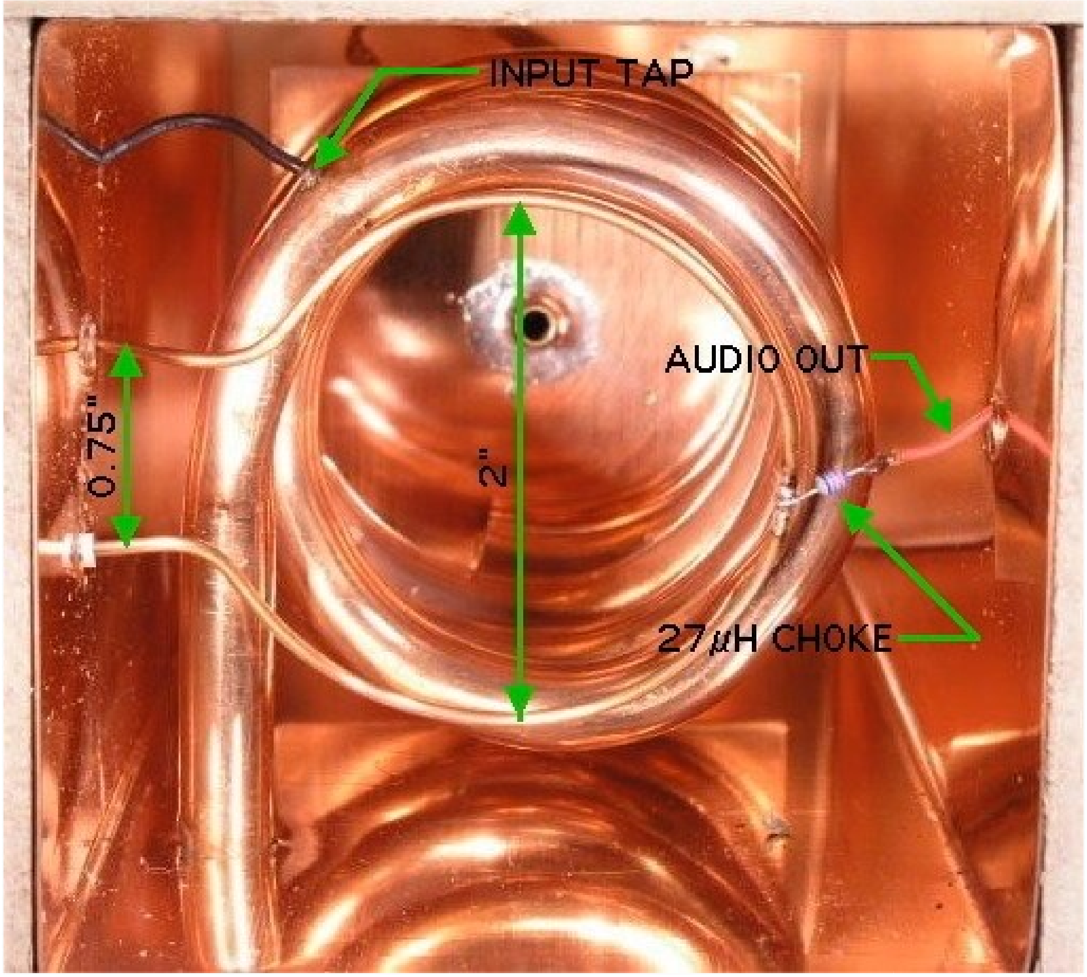

The detector loop is 50 mm (2") diameter #14 AWG (1.6 mm) copper wire, which is stiff enough to maintain optimum shape and position, once those have been determined. It is a friction fit through the side of the enclosure. Both leads are insulated where they pass through the shield.

The detector loop could mount through any of the four sides of the shield. However I chose the arrangement shown here, where both ends of the loop cross the ground end of the helix at 90°. It's my belief that this would minimize any possible unbalance in the loop due to ground capacitance effects.

What isn't shown in the photo, is the separation between the detector loop and the bottom of resonator helix. The distance is approximately 16 mm (5/8"). The loop has a bit of a twist in it, to follow the pitch of the helix, so that separation is a fairly constant 16 mm at all points, except where it bends around the ground end of the helix. Audio is taken off at the middle of the loop through the 27µH choke. The input tap (black wire) is connected to the resonator helix about 75 mm (3") from the grounded end.

Antenna Location

Unfortunately, antenna location is much more critical when dealing with VHF than when dealing with medium wave. A change in position of less than a metre can drastically affect reception. Initially, I found that to receive some stations, the antenna would have to be in one position, and to receive other stations, the antenna would have to be moved somewhere else. In all cases the antenna would have to be rotated to get the strongest signal.

My antenna is a simple set of rabbit ears (dipole) that was supplied with some long forgotten television receiver. I have it mounted on a camera tripod with a pan head, making it easy to rotate.

Currently, it sits in one corner of my house where I find that it gives good reception for all of my local FM stations (except one; more on that below). I have no explanation for why this particular location is best. No doubt, there are many complicated factors. However, I do note that this is in a far corner of my house which likely has the least amount of electrical wiring in the walls and ceiling.

Resonator Helix Length

Originally the helix was 4.3 turns of tubing. According to my design calculations this should have given a resonant frequency of about 110 MHz, just above the top of the FM band. In practice, it gave a resonant frequency of just a bit over 103 MHz. I expect this is mainly because I wasn't able to make the helix diameter exactly as designed.

The resonant frequency is also affected by fringing capacitance between the ungrounded end of the helix and the outer shield. If the helix is not perfectly centred in the shield (particularly the ungrounded end) the fringing capacitance will be higher and the resonant frequency will be lower. The resonant frequency can often be increased slightly by adjusting the position of the helix (ie., making sure it's exactly centred). However I wasn't able to increase the resonance enough, using this method. So, I trimmed pieces off the end of the helix until the resonance increased to an acceptable value.

After trimming, the helix was stretched out lengthwise to return the overall length to the design value of 84mm (3.3").

Currently, the the helix is 3.75 turns, and it tunes up to 106.3 MHz which covers all of the FM stations in my area, but not quite the whole FM band.

Performance

Shortly after building the phase discriminator version of the detector circuit, I was asked–quite legitimately–whether I was certain that the circuit was behaving as a proper phase detector, and not just an oddly configured slope detector. So, it was quite important that I verify its operation. Previously, I'd run a circuit simulation which indicated it would work. Then subsequently, I put a digital voltmeter on the audio output and verified that as an unmodulated carrier was slowly and manually swept across the resonant frequency of the receiver, the output went negative, crossed through zero, became positive and then dropped to zero again. But that's a far cry from actually seeing a decent response curve on a scope.

Unfortunately, I don't have a VHF signal generator which can be suitably frequency modulated. However, I do have a cheap FM modulator (intended for playing portable music players through a car's FM radio). It uses a Rohm BA1404 FM modulator IC. The RF output is tied back into the 12V power cable which plugs into the car's cigar lighter. This was not suitable for providing a reliable and consistent signal to the receiver. So, I had to perform some surgery. Thanks to an online data sheet for the BA1404, I was able to locate the RF output (pin 7, for anyone interested), and connect a separate RF output lead.

Once I managed to separate the RF output, I got a nice clean signal. My audio sweep source is a sine wave from an audio generator. The only remaining complication was that the FM modulation was slightly out of phase relative to the input audio signal. So, I had to build an adjustable phase shift circuit to go between the audio generator and the scope X-input. Once it was adjusted properly, the waveform cleaned up very nicely, and I was quite thrilled to see a fairly good S-shaped discriminator response curve which is shown here:

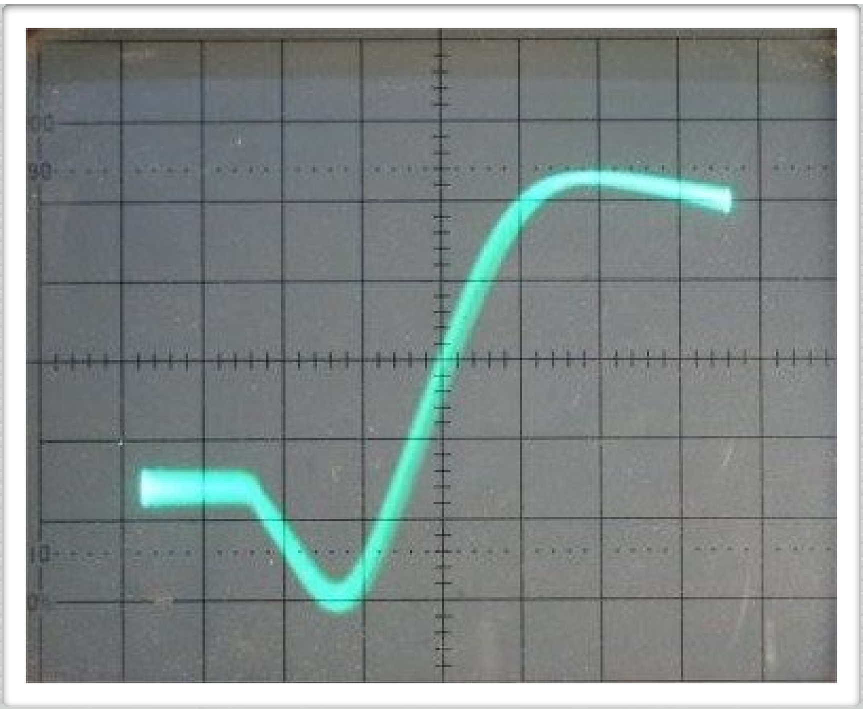

This is the actual DC response of the detector with an input FM signal of about 100 MHz. Mid position on the graticule is 0 Volts DC. Some asymmetry can be seen in the response curve; the skirt on the right tapers off much more gradually than that on the left. At some point I will probably investigate why this is happening, though I'm not too concerned about it. The flat section on the left is an artifact of the FM modulator. All things considered, I'm quite happy with the results.

Unfortunately, because the modulator is uncalibrated, I'm unable to say just what the actual bandwidth is. I hope to address this in the future.

As a side note, it can be seen that the trace is somewhat fuzzy. This is because there is a superimposed 19 kHz stereo pilot tone which is produced by the FM modulator. This had me baffled for a while until it dawned on me what it was. By adjusting the audio generator frequency, the 19 kHz signal can be made almost stationary, but still moves too fast for the camera to catch it.

The test set-up, used to generate the above response curve, is shown here:

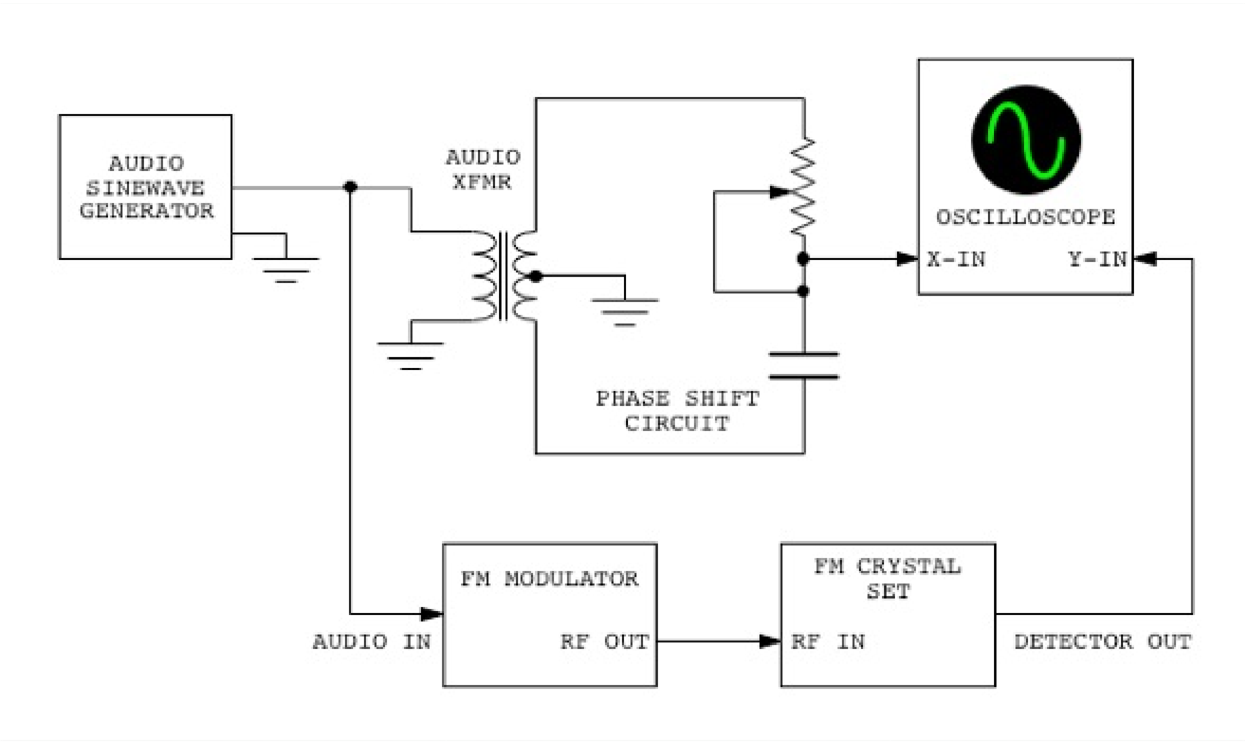

Listening Test

I have twelve local FM stations (within approx. 30 km). With this receiver I can clearly hear all except for two of them. Of those two, the first (91.7 MHz) seems to be temporarily off the air (although I can pick up its unmodulated carrier on other receivers). The second (100.3 MHz) is so weak that I can barely pick it up on a regular FM receiver, and even then, it is buried in noise and is barely intelligible.
So, I wasn't too upset about not picking it up on the crystal set.

The ten local FM stations, which I can hear clearly, range in power from 1300 Watts to 100,000 Watts. All of them can be heard at pleasant volume in a reasonably quiet room. The 1300 Watt station is one of the loudest. However, its transmitter is probably closer to my listening location than the others. The receiver is selective enough that there is absolutely no overlap in stations. Audio quality is very good with the sound powered phones.

Four of the stations are strong enough that they can be heard fairly well with Sony MDR-W08 Sport Walkman phones when properly matched with a Bogen T725 transformer. Although the volume was considerably lower than with the sound powered phones, it was still quite intelligible.

Here is an audio clip tuning through the ten local stations starting at

88.7 MHz and ending at 105.5 MHz, which gives a general idea of the audio quality. File size is 1.6 MB. It was created by connecting the crystal set audio output through a simple RC de-emphasis filter, into a microphone preamp and then into the audio input of my computer running Audacity 1.25 software. The audio is exactly as recorded with no editing or other processing. You will notice that the sound quality changes rather erratically during tuning. This is due to backlash in the tuning mechanism (particularly troublesome at the bottom of the band), and simultaneous adjustment of antenna direction.

More to Come

While this project is close to being complete, a few things remain to be done.

Firstly, the tuning mechanism is very non linear. Tuning at the low end of the band is very touchy, and backlash in the adjusting screw makes tuning these stations a bit frustrating. I'm currently looking at an alternative tuning method which should result in more linear tuning.

Secondly, I would like to test the receiver with a good quality outdoor FM antenna.

Finally, I need to do some final clean-up of the construction, and make

Be a Radio Technician

As a civilian or member of the Armed Forces, Radio offers you opportunities for better pay and valuable, intelligent service to your Country. Start learning Radio quickly; right in your own home, in spare time—through the time-tested, well known N.R.I. Course.

Shortage of Trained Radio Technicians Already Serious

Since new Radios are no longer being made, the service and repair of existing receivers is vitally necessary work. But many of the leading Radio Servicemen in every community are now in the Army, Navy or Marines. This means beginners must get ready fast to take over the great volume of important, highly paid Radio work which must be done.

Radio factories, working to produce over TWO BILLION dollars worth of Military Radio equipment, need more trained men at good pay and overtime. Civilian Radio jobs are also open, to men who can meet Civil Service requirements, in the Army Signal Corps and the Army Air Force as civilian maintenance men and instructors.

All these great civilian demands for Radio Technicians are ADDED to urgent calls by the Army and Navy for MORE Radio Technicians to accept attractive assignments of duty with good chances for advancement in rank and pay.

Learn Radio by the Tested, Practical N.R.I. Method

When you learn Radio with N.R.I., you follow a plan based on over 25 successful years of teaching Radio. You study a Course prepared by Radio experts who are also teachers.

You learn Radio by the method which has put hundreds of men on the road to success. From my students and graduates, working in every field of Radio, I get letters telling me of money they are making—honors they are winning—service they are rendering to their country because of N.R.I. Training.

N.R.I. Will Help You Get a Quick Start

From your first N.R.I. Lesson, you will find practical, money making information. Within a few months, you can be ready to do many kinds of Radio work. It's up to YOU. N.R.I. will cooperate, help you make rapid progress. These days, some students complete four or five lessons a week. We will help you do as well or better.

As an example of how quickly N.R.I. helps you get ready, I want to point out that by the time you finish 12 lessons you should be familiar with Power Supplies, Radio Frequency Amplifiers, Detectors, Oscillators, and important circuits in a Radio receiver and transmitter. There is no wasted time in N.R.I. training.

Mail the Coupon for Sample Lesson and Free Book

Act Today. Mail coupon now for Sample Lesson (Getting Acquainted with Receiver Servicing), and my 64-page book "Win Rich Rewards in Radio." Both books are FREE. They give you an idea of what my Course is like, point out Radio's opportunities, show you letters from men I trained, telling what they are doing and earning. Find out what Radio offers **you**. MAIL COUPON in an envelope or paste on a postcard—NOW.

J. E. SMITH, President
National Radio Institute
Dept. 3EC-5 Washington, D. C.

J. E. Smith, President, Dept. 3EC-5
National Radio Institute, Washington, D. C.

Send me FREE, without obligation, Sample Lesson and 64-page book "Win Rich Rewards in Radio" which tells about Radio's spare time and full time opportunities and how I can train at home for them. (Write plainly.)

Name.. Age............

Address ....

City.. State.................. 4FR-2

ERNSBACK'S EDUCATIONAL LIBRARY No 8

RADIO for BEGINNERS

10 CENTS

By HUGO GERNSBACK

RECEIVER
SENDER

ADIO PUBLICATIONS . 25 WEST B'WAY - NEW YORK, N.Y.

Suggest A Good Book Title

WIN TEN DOLLARS!

SINCE 1908, the name Gernsback has been synonymous with scientific and radio literature. Since that time, many millions of Gernsback magazines and books have been distributed and read by people all over the globe.

Recently, Hugo Gernsback decided to place upon the a new, popular priced series of books under the name of GE BACK'S EDUCATIONAL LIBRARY. It is the intent of the publisher to continuously add books to the library, from time to time, under various titles. These books will cover a great many subjects, which will not necessarily be restricted to radio or science, but which will encompass many other arts.

It is Mr. Gernsback's plan to have all the books of uniform size. Their price being extremely low, everyone will ultimately be able to possess the entire library, especially since they are purchased piece-meal as the books are published.

Don't forget either, that Gernsback books and magazines always have given you your full money's worth. The present series of books again proves this. Never in the history of technical publishing have you been able to buy so much for so little money. The answer to this is in turning out a greater number of books and in large quantities.

In bringing out new books, we want your help. Perhaps you have an idea for a certain book title (subject for a new book) that you think we should publish in one of these 10c volumes. If you have such an idea, **communicate it to us; we will pay $10.00 for each new book title which we accept. If you submit more than one title, write each one on a separate sheet of paper.** In case of a tie, identical prizes will be paid to the contestants.

Address all letters to

RADIO TITLE EDITOR

RADIO PUBLICATIONS
25 West B'way NEW YORK City.

PRIZE WINNERS

ALFRED CETTIE
4115 Midvale Avenue
Seattle, Wash.
Title: ALL ABOUT AERIALS, Book 4

RUDY RUZICKA
3441 California Avenue
St. Louis, Mo.
Title: BEGINNERS' RADIO DICTIONARY, Book 5

KENNETH WARNER
R. F. D. No. 2
Burbank, Ohio
Title: HOW TO READ RADIO DIAGRAMS, Book 7

CLYDE SORRELL
504 North 10th Street
Rockyford, Colo.
Title: RADIO FOR BEGINNERS, Book 8

(Watch the List Grow)

Copyright 1938, by H. Gernsb
Printed in U.S.

Radio for the Beginner

By HUGO GERNSBACK

Chapter I

HISTORIC

THE true art of radio was unquestionably discovered by Heinrich Hertz, a German professor, living at Frankfort. His first technical papers on his epoch-making invention were published in 1887. Hertz was the first to send electric waves through space by means of an electric spark. His apparatus was simple; he had an electric spark coil that made intermittent sparks, and by proper arrangement of this station, he could receive sparks at a distance by the simple arrangement of cutting a single wire hoop and leaving a small gap. Between the two free ends, small sparks jumped whenever sparks were made to jump on his spark coil a few yards away.

Guglielmo Marconi, an Italian youth, read of these experiments and being gifted along these lines, he duplicated Hertz's experiments. Soon his mind conceived the idea of using the invention for transmitting intelligence over a distance. He endeavored to send a message without wires over miles where Hertz used yards. Instead of the wire hoop, Marconi devised and used a more sensitive apparatus. He found that an instrument called the coherer was enormously sensitive to the new electric waves, and soon was transmitting signals for many hundreds of yards on the estate of his father in Italy. By diligent labor he increased this distance and shortly was telegraphing without wires across the English channel. Not many years later, he transmitted the letter "S" in telegraphic code across the Atlantic by means of wireless.

Radio telephony, contrary to popular opinion is not a new invention either. It was first invented by Valdemar Poulsen, the "Danish Edison." Instead of using a crashing spark at his sending station, he used a silent electric arc, with certain adjuncts.

Chapter II
WAVE ANALOGIES

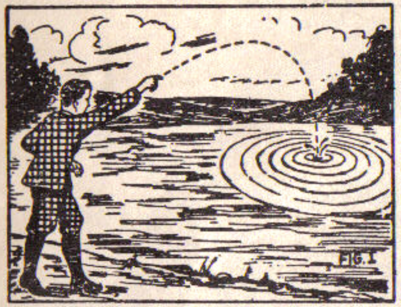

What is a radio wave? It is not any different physically than a sound wave or a wave in the ocean. If we throw a heavy stone in a still lake, it makes a splash. This wave rapidly extends in the form of circles, as shown in Fig. 1.

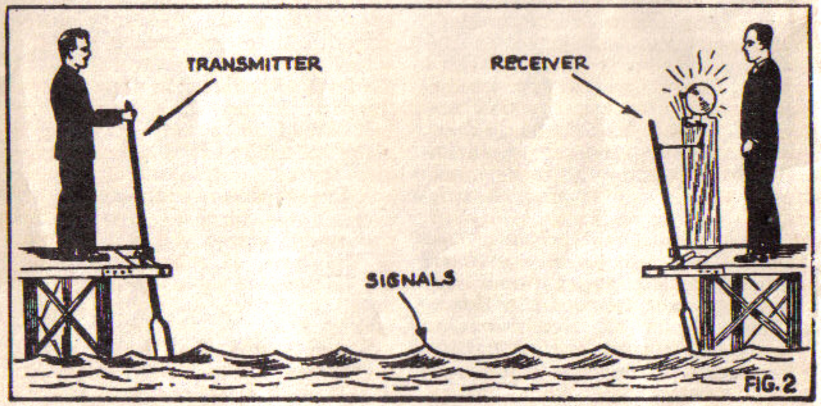

The analogy between water waves and radio waves can be carried a step further. We have seen how waves may be produced; now let us study how they may be received.

Take, for instance, a tank of water 20 or 30 feet in length. At the two opposite ends, platforms have been built, as illustrated in Fig. 2. On one of these platforms a large paddle has been arranged, so that a person may operate its handle. Now, if the paddle is moved back and forth, a series of waves extending in all directions from this source of creation will be formed. The waves spread further and further away from the paddle in concentric rings until their strength is completely expended. In this instance the tank is small, and the waves are sufficiently powerful to reach the opposite end, where the other platform is built.

On the other platform, located at the other end of the tank, we have a smaller paddle, on the handle of which a hammer has been arranged to strike a gong. It is obvious that the waves moving the paddle will cause the gong to ring, informing the operator on that platform that the operator on the other platform is moving the paddle and creating waves on the surface of the water. By skillful manipulation of the larger paddle, it is possible to cause the smaller paddle to ring the bell periodically, as desired; and if a series of signals has been prearranged, the operator with the larger paddle may communicate certain information by properly operating its handle. This represents both the transmitting and receiving stations of the wireless telegraph, the larger paddle being the transmitter, and the conducting medium being the water, while the smaller paddle is the receiver. It is exactly so in radio. If, by means of certain electrical apparatus connected to an aerial, we excite this aerial electrically, waves are set up in the space exactly as water waves are set-up on the lake. Radio waves, just as do the water waves, branch out in all directions. With the water waves this is not so true. A true water wave, as we know, is carried along only upon the surface of the water. A few feet below the water and immediately above the water, no water waves are found. A better analogy would be sound waves. Take, for instance, a church bell. By giving it a blow with a hammer, we excite this bell. Sound waves are set up in the air in all directions from the bell. Whether you are on the street level, 100 feet below, or in a building on the same level as the bell—in all these positions you will clearly hear the ringing of the bell (Fig. 3). The sound waves are propagated in every direction in the form of waves, invisible to the eye, but "visible" to the ear. These waves are exactly of the same shape as are the ocean waves or water waves, with the difference that the sound waves go out in the air in the form of spheres. In other words, the first sound wave leaving the bell would be a sort of invisible globe all around it. The wave rapidly spreads out, becoming larger and larger, always remaining, however, in the form of a sphere. If the sound waves did not go out in the form of spheres, it would not be possible for us to hear them in all directions as we have seen in Fig. 3.

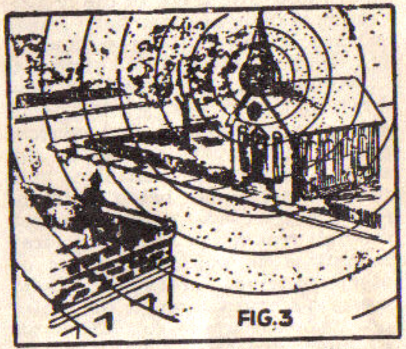

Sound waves that leave a bell branch out, above, below, sideways; in fact in all directions. It is exactly so in radio. The aerial of the broadcasting station, or other radio transmitting station radiates as does a bell. Both are transmitters of waves.

We can hear a bell even if windows are closed. The invisible sound waves pass through the window panes. Radio waves do exactly the same thing, except that they pass through solids far better than do sound waves. Radio waves even pass through mountains, providing these mountains do not contain ores or other metallic substances. Radio waves also pass through the water just as sound waves do. But, as we all know, the farther away we go from a ringing bell, the more difficult it is to hear it. The reason is that the original wave, as we increase the distance between ourselves and the bell, becomes larger and larger and soon covers a tremendous area. Finally there comes a point where we no longer can hear the bell.

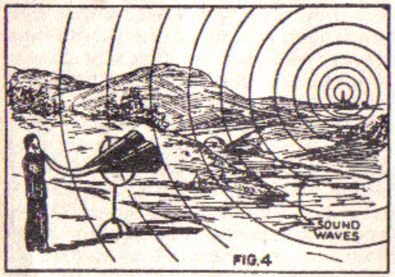

FIG.4

If, now we were to take two horns and point them in the direction of the bell, as shown in Fig. 4, and apply the ear pieces to our ears, we would be able to hear the bell again, although without these appliances, we would not be able to hear it at all. The reason is that the vibrations that reach our ears normally are too weak to be intercepted by our small ears. By enlarging our ears, as shown in Fig. 4, we intercept many more weak sound waves, and these waves, all being collected into our ears—bunched together, so to speak—are sufficient to impress the diaphragm in the ear. This holds true in radio as well. If we have a transmitting station, or a broadcasting station, we can hear it only up to a certain distance with a given apparatus. If we take a small aerial which we can liken to a normal ear, we can use it only for a given distance, let us say 25 miles. If we move this aerial 30 miles away from the radio broadcasting station we can no longer hear it. The case here is exactly as with the sound waves. The radio waves now have to cover much larger areas, and there are not enough waves, so to speak, to leave any impression upon our small aerial. If, however, we were to double or triple the size of the aerial, we would do physically the same thing as we were doing when we attached the two horns to our ears. By having a larger or increased aerial with more wires, we would, by means of this, intercept more waves than we could with a small normal aerial; consequently with such an aerial we could hear the broadcasting station again, even though we were removed 35 miles from it. The analogy between the sound wave and the radio wave holds pretty true, all the way through. Of course, in radio we have other means to bring in the signals even if we are removed still greater distances. It would not always be practical to make the aerial tremendously large in order to hear greater distances, also we would not expect to hear our bell 20 miles away by means of even large horns. We would have to devise some other more sensitive means to hear the bell, and there are such means at hand today in super-sensitive electrical microphones which magnify the very weakest sounds. So too in radio it is not necessary to build a larger and larger aerial, the more we remove ourselves from the broadcasting or transmitting station. Instead, we use more sensitive apparatus which will magnify the sounds in an electrical manner, so that we can hear the station even though we are removed thousands of miles from it.

Let us now return to our stone which we dropped into the water. If we were to place our eye on a level with the water, and someone was to throw a stone into a quiet surface of water, what would we see? Fig. 5 shows this. Any water wave is composed of two distinct parts, the crest and the trough. In other words, the water first comes up, then dips below the original surface, then up again above the original surface etc. In our illustration, we have shown in dotted lines the original surface of the water. The disturbance of the stone has caused the water to expand into waves. Now then, the wave length is that portion which extends fron crest to crest. In Fig. 5 we see what a wave length consists of. It starts at the top of the crest, covers the trough and again up to the crest. This is exactly one wave length, because it embraces the total make-up of one complete wave.

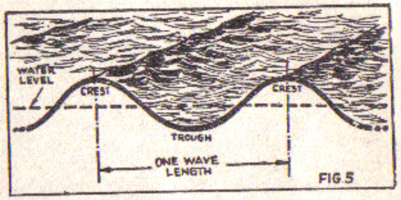

FIG.5

In radio we have the same sort of waves, and these waves go out into space in all directions, as we have learned before. In radio we can make a wave length from a fraction of one meter up to several thousand meters and over. This all depends upon the apparatus we use.

RADIO TELEGRAPH AND RADIO TELEPHONE WAVES

In radio telegraphy we simply hear the plain wave in our telephone receivers. If the operator in the transmitting station presses his key, groups of waves are sent out into space as long as the key is depressed. At the receiving side we hear the waves making a buzzing sound for the length of time that the key is depressed at the sending station. If the key is pressed down for a second, we hear a buzz for a second. If the key is depressed for two seconds we hear the buzz for two seconds, and by means of this buzzing sound the telegraphic signals are reproduced. Usually a code such as the Morse or the Continental is used. For instance, a short buzz will be the letter "E" while "SOS" would stand for the following - - - — — — - - - (a short dash being a short buzz, a long dash being a long buzz).

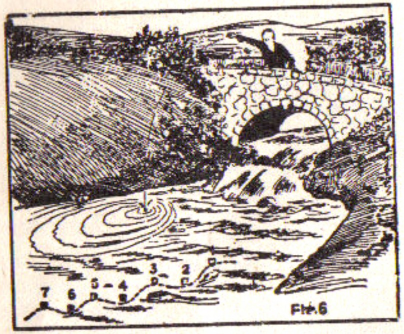

In radio telephony (broadcasting) however, we have a different and more complicated action. In the first place, we hear sounds, words, and music exactly as they are produced at the broadcasting or transmitting station. Two distinct things happen. The aerial is made to send out a radio wave that is continuous. This wave cannot be heard by the human ear with ordinary receiving apparatus. It is what is technically called C.W. or Continuous Wave. It is also used to carry along the human speech. At this point we must resort again to our water wave. Suppose we throw a stone into a river. At the same time that the stone is thrown we also throw a cork into the water, at the same spot. What happens? The cork is carried along by the current as shown in Fig. 6. First we see the cork in position 1. A little later we see it in position 2. Still later in position 4 as shown on the dotted lines. The cork, therefore, is carried along by the wave as well as by the current. As the waves progress, the cork progresses also. Exactly the same thing happens when the human speech is impressed upon the radio carrier wave. By certain means too technical to go into here, the vibrations made by the voice are carried along upon the carrier wave, exactly as the cork is carried upon the water wave. At the receiving side we only hear the words or music, for the reason that the carrier wave is inaudible. Hence, nothing but the words or speech are heard by us in our loud speakers.

Radio waves travel with the speed of light, namely, the enormous speed of 186,000 miles per second. We, therefore, can understand that if a message is sent out anywhere on our globe, it will be received at any place almost instantaneously; the greatest distance that a radio wave or a message could travel would be 12,000 miles, for the reason that the circumference of the earth is 24,000 miles. You will see, therefore, that a radio wave would travel around the earth at the rate of almost eight times in one second.

POPULAR MISCONCEPTION AS TO RADIO WAVES

Many people have an idea that radio waves broadcast by a transmitting or broadcasting station, change their form as they are sent out into space. The length of the wave never changes between the transmitting and the receiving stations.

It stands to reason that if all stations were to send at exactly the same wave length, we would get nothing but a jumble.

Suppose you have six pianos in one room, which are all tuned alike; if we have six players sitting down at the pianos and each hits the same key, we will only hear that one note, let us say A. You could not possibly detect it if five were striking the key A, because all of the players are transmitting on the same sound wave length which transmits only the note A. Suppose, however, that one operator is striking the key A while another strikes the key E. We can immediately eliminate one or the other, and by a little concentration of our ear, we can hear either A or E. In other words the two pianos are now transmitting at different sound wave lengths, the wave length of E being different from the wave length of A, and vice versa. It is exactly so in radio, only we have better means in radio because we can tune out entirely one station or another by means of tuning appliances so that we can hear either one at will. That is the reason why different transmitting stations send on different wave lengths.

●

Chapter III

TRANSMITTING

General

There are several ways of transmitting by radio; the oldest and historical method is shown in Fig. 7. Here we have an ordinary spark coil such as is used in automobiles, a few dry cells, a key and the so-called spark gap which may consist of wire nails or better two zinc balls. Every time we press the key a spark jumps across the open space in the spark gap. By connecting one end of an aerial to the ground, which may be a water pipe, or a steam radiator, radio waves are sent out into space. We might compare this to a string held between two nails and plucked with the finger, when we would hear some sort of noise.

Just exactly as a manufacturer of a piano knows what the sound wave length of the longest string of his piano is, so the radio engineer will know on what wave length a given aerial will send.

Roughly speaking, an aerial 100 feet long will give a wave length of about 140 meters, while an aerial 200 feet long will give a wave length of exactly twice the length of the shorter one or 280 meters.

Suppose, with our little outfit shown in Fig. 7, we wish to send out a wave length of a thousand meters; we could do this by making an aerial 833 feet in length. That however, would not be practicable because not in all instances could we find that much room for the aerial.

We, therefore, resort to another means, and we build an aerial in-

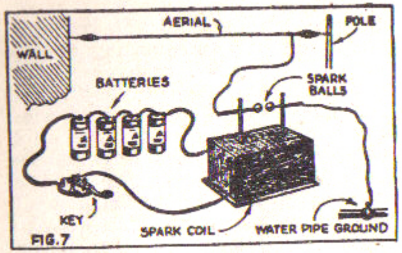

FIG. 7

doors which we attach to the original aerial—a sort of sending tuning coil, which we show in Fig. 8. This tuning coil is the same wire which we use for the aerial, wrapped around a frame or tube, as shown in Fig. 8. This coil means simply the additional wire which is necessary to lengthen our aerial in order to make it long enough to give us our thousand meters. By means of the slider, which runs up and down the wire convolutions, we now have the means of changing the wave length merely by adding more or less wire. If this is not entirely plain, take a violin as an example. When the violinist wishes to transmit a certain sound wave he plucks his string first without touching his hand to it. As he presses down on the string, he automatically makes it shorter and shorter, and the further down his fingers slide, the higher and higher the note becomes. He does here exactly the same thing as the slider does on our sending tuning coil, that is, he changes his sound wave length. Fig. 9.

In other words, if he wants a long sound wave he slides his finger down the small end of the violin, and if he wants a short sound wave, he slides his finger towards his chin. This changes the sound wave length in exactly the same way as our sending coil changes the radio wave length. Both are fundamentally the same.

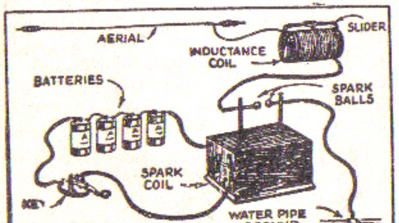

FIG. 8

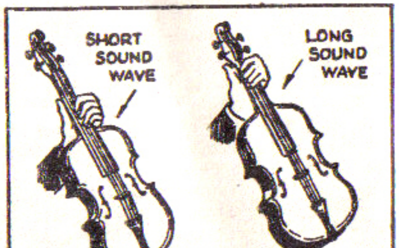

FIG. 9

When Marconi first rigged up his little sending station, as shown in Fig. 7, he naturally could only send out radio telegraphic signals. Every time he pressed the key, radio waves were sent out. When he pressed the key for a second, a buzzing noise was heard for a second in the distant receiving telephone receivers. If he pressed the key for two seconds, a buzzing sound for two seconds was heard. By this means the telegraphic code is made up. At the present time, the Continental code is used almost exclusively, and today, as in Marconi's time, when

the operator at his transmitter presses down his key for a short duration, this is interpreted as a dot at the receiving side, and when he presses his key down for a longer period this becomes a dash on the other end. By means of dots and dashes, the telegraphic code is made up.

In Fig. 7, we showed a simple sender. Of course, it goes without saying that soon after Marconi started his experiments, more complicated sending apparatus was designed.

There are now many different transmitters in use, as for instance, the vacuum tube which may be used for transmitting. This has the advantage of giving rise to what is called continuous waves; this is made clear by the diagram shown in Fig. 10A. When we press the key of the old Marconi outfit, we send out into space radio waves which have somewhat the form shown in Fig. 10A. These waves start with a high pitch, as we might say, and die out rapidly. This happens a great many thousand times each second, but these waves are not continuous. They are small wavelets, as we might term them, which are disrupted and do not form a continuous line. Look at Fig. 10B; this is what we might term a continuous wave and is a wave which is sent out by a vacuum tube transmitter such as is now used universally at broadcast stations. As long as the sending set is transmitting, a continuous wave is sent out into space which does not vary. It does not take a technical mind to know that the waves sent out, as shown in Fig. 10B, must be better and clearer than the interrupted waves sent out in Fig. 10A. This is the day of the Continuous Wave commonly called C. W. It is the Continuous Wave that makes broadcasting possible.

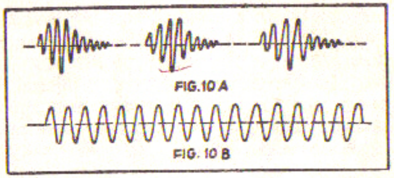

Let us make a comparison again, which can be easily understood, and which may serve to make the interrupted wave and the Continuous Wave clear in our minds. Take a number of pipes as shown in Fig. 11A; one person stands at one end and another at the other end. One talks into this interrupted pipe, which may be 100 feet long, and as will be readily seen the person at the other end will have a great disadvantage in hearing the speaker because the pipe, being interrupted so many times, breaks up the speech. This is the analogy for spark waves. Now turn to Fig. 11B; here we have a long pipe the same as we use in our speaking tubes,

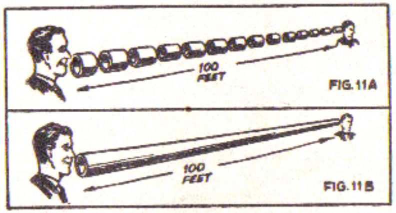

which is free from interruptions, and is continuous all the way through. You can readily understand why the person at the other end will have no trouble in hearing what the speaker says, for the reason that the pipe is continuous all the way through. This continuous long pipe stands for a continuous wave. This of course, is not a strict analogy, but may serve to implant in the reader's mind the difference between an interrupted spark wave and a continuous wave.

●

Chapter IV

RECEIVING

General

No matter what receiving station you have, it can receive either radio telegraphy or radio telephony. The receiving station has the exact counterpart in your ear, which receives any and all sound and noises that are floating about in the air. So it is with the radio receiving station; with it you can hear, if properly adjusted, any and all disturbances that are flung out into space by the various broadcast stations. Receiving instruments are becoming more and more sensitive for which reason we can hear the sending station further and further away. If we have a broadcast station which is sending out a band concert, and if we were to use Marconi's first instrument, the coherer, for receiving purposes, it would not be possible for us to receive this concert at all because Marconi's coherer is totally unsuited to receive broadcast radio music. After Marconi's coherer came the auto-coherer a somewhat more sensitive instrument. With such an instrument a broadcasting station could possibly be heard five to ten miles, but no further. Next came the crystal detectors; with a good one we may hear the broadcast station at a distance of 25 miles or more. Still later came the audion or vacuum tube. This instrument, being enormously more sensitive than a crystal detector, at once increased the range up to a thousand miles and over.

How do we get more sensitive radio receivers? A single vacuum tube is only able to detect radio signals for a given distance. By adding more vacuum tubes, more "stages" as we call them in radio parlance, we step up the faint signals until finally a radio signal that could not be heard at all with a pair of telephone receivers, and a crystal detector, will roar out of the loud-speaker with ear-splitting strength. If conditions are not right, for instance, if our insulation is bad, or if the adjustments of the apparatus are not correct, we will hear the signals faintly, and often not at all. Receiving radio waves is not any different from receiving light waves. If you go to the opera you would not think of using the opera glass unless it was properly adjusted — tuned — to your particular eye. You also would not have the lenses covered with finger marks. You know in advance that you would not see much of the opera if you were to do that. The same thing holds true of your receiving set. We must have perfect insulation; all metal parts that carry the current must make good contact—all parts must be perfectly adjusted. Only in this case will the receiving be 100 per cent., or rather approaching it, because we have not as yet reached the stage where we can receive 100 per cent.

We have mentioned before that radio waves pass as readily through a stone wall as through the air. It, therefore, does not surprise us that we can have a modern radio receiver in our library without an outdoor aerial at all and the waves will be received just as well as if the radio set was stationed on top of the roof or out in the yard. This is true only if the detecting apparatus of the set is very sensitive, otherwise we will not be able to detect the waves, although they are there.

As a general thing, it has been found in receiving that the higher up our receiving aerial is, the better we can receive. It also has been found that one can receive further with a given receiver over water than over land. Roughly speaking, one can hear twice as well over water as over land.

Presence of steel buildings also tend to cut down the receiving range. Thus, for instance, if a crystal of moderate sensitivity is located in the heart of the New York downtown district, we will hear practically nothing from the neighboring broadcast stations, unless of course the aerial extends above the buildings. All these facts should be borne in mind when erecting a good receiving station.

Another point to be remembered is that reception during the night time as a rule is better than during the day. The reason for this is that during the day time the sun's rays ionize the air, which means that the sunlight makes the air partly conductive. That cuts down the receiving range as well. It is not a rare occurrence that distant stations are heard twice as far during the night time as during the day.

STATIC

A few words as to this greatest nuisance that the radio man has to contend with. Static disturbance is nothing but atmospheric electricity. We are not bothered much with static in the winter time, but during the months of May, June, July, August and September, there is plenty of it, particularly, if we have an aerial extending up into the air. Static makes itself heard in our receivers in a sort of irregular noise that cannot be controlled today. Very often we hear sharp clicks in our receivers which vary up to a loud roar particularly when a thunder storm is approaching. Sometimes the air, even on a perfectly clear day, is so highly charged with electricity that if we bring the lead-in from our aerial, close to the ground wire, small sparks will jump from the aerial to the ground, proving conclusively that static electricity is collecting upon the aerial. These static noises so far have not been corrected, as no way has been found to weed out or entirely tune out these disturbances.

Chapter V

RECEIVING INSTRUMENTS

The earliest instrument for detecting radio waves was the coherer. This was a rather complicated little instrument, and one that was difficult to keep adjusted. Furthermore, it was not at all sensitive, compared with the detecting instruments of today.

One of the earlier detectors, still in use today, is the crystal detector. Dunwoody was perhaps the first man to use such a crystal, viz., carborundum. This particular crystal is seldom used today.

Another of the early detectors, which has almost entirely vanished from the radio picture is the silicon detector. Silicon is a manufactured substance, which is a by-product of the electric oven in the manufacture of abrasives; it is a cousin to carborundum. Silicon is a hard rock-like substance of a dark silver-gray color. A small piece of silicon broken from a larger piece by means of a hammer or in a vice, about ¼ inch by ¼ inch, is first imbedded into a soft solder. The idea of this pellet is that contact is made on five sides with the metal, which is simply cast around the silicon, and the crystal part of this round pellet is afterwards placed in contact with the contact member as shown in Fig. 13.

The galena and other common crystals of today are similarly mounted. The contact member to the silicon detector is nothing but a piece of brass, which is not very sharp at the end, but rather blunt. The amount of pressure upon the pellet is varied by a spring. In detectors of this kind, not every point of the silicon is equally sensitive. Some points are very sensitive while others are not. Some of the sensitive points require more pressure than others. All this is found out by experiments. The silicon detector is quite sensitive; it was the first detector invented that required no battery whatsoever to detect radio signals, and for that reason was a favorite instrument with early experimenters. The silicon detector had also the great advantage in that it was not easily "knocked out," as most other detectors are, nor did it burn out easily. When connected, as in Fig. 12, a set of receivers of at least 1,000 ohms should be used

for best results. A 75-ohm receiver, such as is used in house phones, should never be used in connection with radio waves. The results are very poor.

Soon after the silicon detector was invented, Greenleaf W. Pickard, the inventor of the silicon detector, invented a host of other detectors all of which use a native mineral crystal, such as, for instance iron pyrite, copper pyrite, bornite, etc. All of these detectors are used similarly to the silicon detector, the crystal being cast into a soft metal in pellet form. This pellet is used in the same way as the silicon detector; sometimes a sharp brass contact point is used with some minerals and at other times a fine wire is used, which latter is termed a **Catwhisker.** Such a detector is shown in Fig. 13. This detector uses as a sensitive member the mineral or crystal known as the iron pyrite. As with all other crystal detectors, no battery is used in connection with it. In this detector no sharp point is used, but rather a fine wire catwhisker. A catwhisker is a piece of fine wire about No. 26 or No. 28 B & S gauge phosphor bronze. This is usually attached to some sort of handle or other adjusting means so that the pressure of the wire upon the surface of the mineral may be varied.

One of the most sensitive and most widely used detectors is made of Galena, a lead ore of which there are different grades. It is known under many trade names as well. A good piece of galena is probably as sensitive as any crystal yet discovered, but it is not stable. A catwhisker, as explained above, is used with the galena crystal and the amount of pressure has to be found by experiment. Ordinary galena is not sensitive on every spot, but there are certain grades which are equally sensitive over the entire surface; this is known as Argentiferous Galena, which means that it is silver bearing. On the other hand not all argentiferous galena is equally sensitive, and there is no hard and fast rule about it. It must be found by experiment. A good connection for the crystal is shown in Fig. 26. No battery is used with galena, and as a matter of fact a battery will destroy the usefulness of it by burning out the sensitive points. With galena, a fine brass wire No. 24 or No. 26 B & S gauge is used; a stiff gold wire of the same dimension may also be used, as it is non-oxidizing.

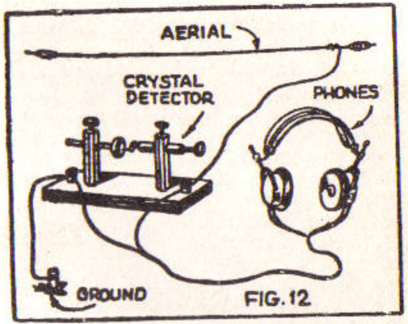

FIG. 12

All crystals are sensitive only if absolutely clean, and their usefulness becomes destroyed immediately upon being handled with bare fingers.

Although we have stated a little further back that most crystal detectors have the sensitive mineral embedded in a metal pellet, the amateur or experimenter does not always require this, and Fig. 13A shows several simple home-made detectors. The illustrations are so self-explanatory, that no further details need be given. Base boards may be of wood, hard rubber or any good insulator. As will be seen in these illustrations, the detector mineral is clamped by simple holding devices; anything that will hold the crystal down so that it will not move, and at the same time make good contact with it, may be used. The catwhisker wire is best, a No. 24 or No. 26 B & S brass gauge, or phosphor bronze wire. It may be straight, or coiled in pig-tail fashion, either will work equally well. The clever experimenter can change the design to suit his own individual tastes, and the chances are that the device will work well. The trouble with most mineral detectors is that their adjustment does not keep for any length of time. Jars, or static surges in the aerial will cause the detector to become inoperative, after which it must again be adjusted.

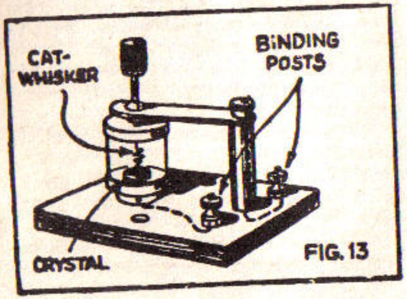

FIG. 13

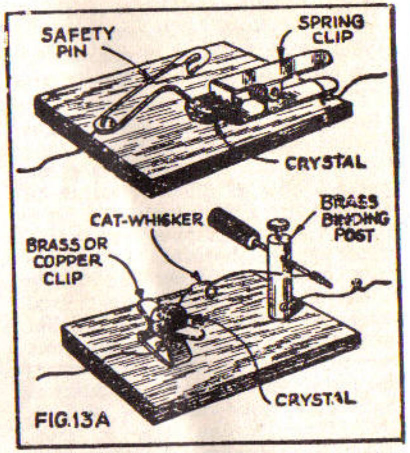

FIG. 13A

We now come to a vastly different sort of detector, namely the Audion, or as it is commonly called, the "Vacuum Tube." This detector works upon an entirely different principle from any of the former ones described, and is in general use today for reasons which we shall learn presently. Fig. 14 shows a standard vacuum tube where we have the filament, which is the same as that used in an incandescent lamp; this is heated by means of a battery of from two to six volts or by means of current from the house lighting lines. We next have the grid which may be in the form of a gridiron or a spiral, it making little difference which. Opposite the filament and with the grid in the middle, we find the plate, usually a small piece of nickel or other metal. The connection of the simplest audion is shown in Fig. 15. If we make the plate positive with respect to the filament, we find that highly charged electrical particles called "electrons" travel constantly from the filament to the cold plate.

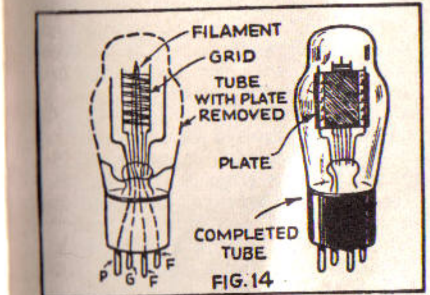

FIG. 14

It was soon found that the vacuum tube acted as a sort of valve for the electrical current, allowing the high frequency currents as they came over the aerial to travel in one direction in a vacuum tube but not in the other. In this respect the vacuum tube is the same as a crystal detector, which also acts as a valve, permitting currents to pass one way only.

The vacuum tube was first invented by Dr. Fleming, to whom belongs the honor of using it first as a detector for radio. He was using only a two-element tube, viz., an exhausted bulb containing a filament and a plate. Dr. Lee De Forest conceived the idea of introducing a third electrode into the tube, as explained above. The purpose of this electrode which he called the grid serves only to control the flow of the electrons attracted by the cold plate. It is the grid that makes the vacuum tube the exceedingly sensitive apparatus that it is. Making the grid alternately positive and negative varies the amount of current that flows from the hot filament to the plate, decreasing, and even stopping it entirely. The grid simply acts as a gate valve which controls the plate current. The curious thing about the grid is that it uses no great amount of power. A modern vacuum tube is exhausted to a very high degree, because it was found that unless the vacuum was perfect, the sensitivity of the tube was very poor. It is not necessary here to go into a very technical discussion of the vacuum tube, as we are merely interested in its functioning. The study of the vacuum tube, however, is a science in itself today, and for that reason it can only be treated generally here. We must, however, add that the vacuum tube is far more sensitive than other detectors, particularly when used in connection with other vacuum tubes. It was found, for instance, that this was the case when several tubes

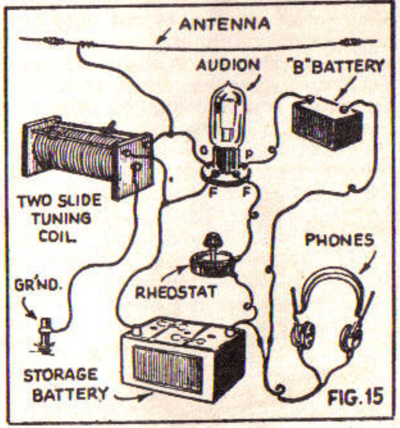

FIG. 15

were coupled together; this gives us the so-called two-step or three-step amplifier, which will be discussed later on. The idea of these amplifiers is for each to step up the exceedingly weak current received from the first tube. By means of such a stepping up process, it is possible to bring in signals over tremendous ranges, a thing impossible to do with any other detector known at this time.

Tuning Devices

We have seen in previous chapters that each radio wave-length is dependent upon the length of wire of each aerial. If it were possible to make all aerials of exactly the same length and capacity, and if all stations were transmitting at exactly the same wave-length, we would not need any tuning devices. Unfortunately this is not the case. When we install an aerial, we cannot always make it of the length or capacity which we desire, but are hampered by physical and geographical limitations. In other words, our aerial is usually a compromise. On the other hand, the various transmitting stations all send on different wave lengths and, for that reason, many different tuning devices are used. One of these, and the oldest, is the Tuning Coil, shown in Fig. 16. This is nothing but an aerial wound upon a cardboard tube or other circular or square piece of insulating material. The tuning coil is simply an extension of the aerial. Even though we have an aerial which is only 100 feet long, by attaching more wire to it in the form of the tuning coil, we thereby lengthen the aerial. The tuning coil, as shown in the illustration, is simply an insulated wire wrapped upon a cardboard tube, its size is immaterial. Tuning coils may be made in almost any size, from the smallest one wound upon a pencil, to the largest, as big as a barrel. The more wire we use, the more wave-length our tuning coil will be able to absorb, so to speak. Of course, in practice tuning coils are built for a certain capacity, all depending upon what it is to be used for. If, for instance, we have but a little aerial and wish to receive from stations having a wave length of say 650 meters, a small coil about 6 inches long and 2 inches in diameter and wound with No. 24 B & S gauge wire will do nicely. The purpose of the slider is simply to add more or less wire to the aerial; it is but an adjustment. It goes without saying that the slider of the tuning coil must touch the wire, as otherwise no connection would be made. In Fig. 16, we show the simplest connection for a tuning coil. This, as will be seen, duplicates the connection of the crystal detector. We have here merely added the tuning coil in order to tune the circuit. By means of this tuning coil, it now becomes possible to tune out unwanted stations merely by moving the slider back and forth and so connecting more or less wire to the aerial. For instance, if two stations are sending at the same time, by moving the slider backward and forward it becomes possible to tune in or out the unwanted station, and listen only to the one we desire to hear. In Fig. 17, we show the same tuning coil, but with two sliders. The two sliders are somewhat of an improvement, for the reason that better tuning is accomplished with them.

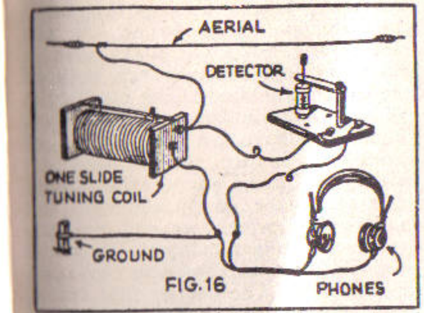

FIG. 16

Loose Couplers

The Loose Coupler, Fig. 18, is another form of tuning coil and this instrument, which was formerly used a great deal, is really one of the best tuning devices known. Instead of using just one coil, as for instance the tuning coil just described, a loose coupler uses two coils—one sliding into the other without touching. The loose coupler is an electrical transformer, as it has been found that if a radio current traversed one coil, another tuning coil standing close by would be affected, although no wire touched the first coil. This is termed an "inductive effect." In other words, the energy is radiated from one coil to another, the same as a stove radiates heat to objects that are close to it.

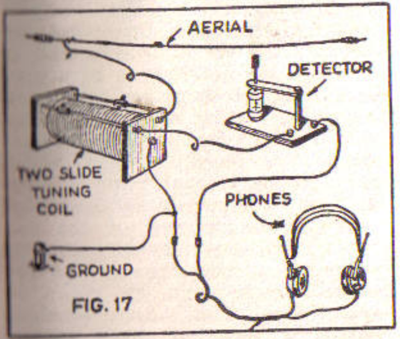

FIG. 17

As we just mentioned, the loose coupler is a transformer. The current that comes in over the aerial in the form of radio waves is a high frequency current. By that we mean that waves swing back and forth very rapidly. It is the purpose of the loose coupler to change this energy into a more suitable form. We again take recourse to an analogy. In Fig, 19 we show, by means of a lever action, the principle of the transformer. We are all familiar with the lever action whereby a man who weighs only 150 lbs. can raise a weight of 1000 lbs. by means of the lever. Is he getting something for nothing in this case? Certainly not! You cannot get free energy, but the experiment in Fig. 19 simply shows that force plus time may be transformed into something else. In this case the man

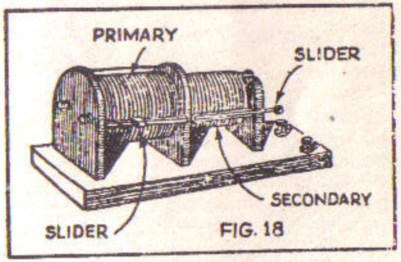

FIG. 18

who weighs 150 lbs. is the force and the time is the interval that it takes him to reach from point A to point B. The two added together are sufficient to raise the 1000 lb. weight, the distance from C to D. The longer the lever arm L, the more weight we can raise. Archimedes told us that "give him a long enough lever he could raise the earth from its hinges." Always providing that he has a sufficiently long lever and a fulcrum, or point of rest which is shown at F in Fig. 19. Summing up, we understand now that by means of a small weight we are able to lift a much heavier one.

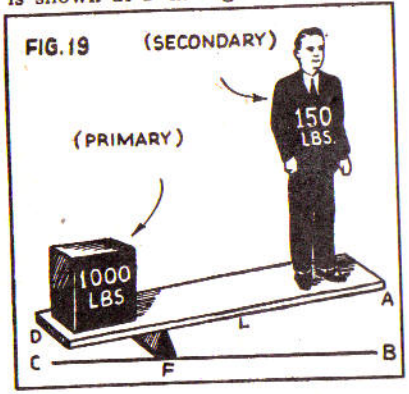
FIG. 19

This analogy holds with our loose coupler, which we have shown in Fig. 18. The loose coupler has two coils; the primary, which is usually the outer tube, is always wound with a coarser wire, while the inner tube is wound with a finer wire. As in the tuning coil we have a slider upon the primary, or if we do not use a slider we may have taps (connections) taken at every few turns of wire, if we so desire. On the secondary also, we may have a slider or taps brought out, both of which are the same. The inner tube is made to slide back and forth upon sliding rods so that the degree of coupling, as we call it, may be changed. If the inner tube, called the secondary, is pushed into the outer tube, which we call the primary, we have a complete electrical lever system, as shown in our analogy Fig. 19. The energy that comes into the primary is now raised exactly as the weight is raised by means of a lever, and we get a marked effect from the secondary. The more we pull out the secondary tube, the less our lever action becomes. It is as if the man in Fig. 19 were to move down to the point F where it would become impossible for him to raise the weight at all, or even to budge it. By using the loose coupler, we do not get something for nothing any more than if we raise a stone we perform no work.

Reverting back to Fig. 18, the tubes of the tuning coil may be of cardboard, hard rubber or composition, or any good insulating material. No steel or iron should be used in the construction of a good loose coupler. Its size is immaterial providing the proportions are right, this being determined by experiment. The important part is that the secondary must come as close as possible to the primary. In other words, the diameter of the two tubes must be so that the secondary tube, when moved inside the primary will take up the entire air space without, however, touching the outer tube. The closer the two windings come together, the better it is. In Fig. 20, we show the simplest connection for a loose coupler, crystal detector and phones. Very good results are had with this circuit, and the loose coupler is particularly efficient for tuning out interference to a certain degree. It gives what is called sharp tuning, because if two stations operate at a close wave length, let us say one at 360 meters and another at 320 meters, the loose coupler will give very good results by reason of its sharp tuning.

The Vario-coupler shown in Fig. 21 is a development of the loose coupler. In this instrument we have an outer tube wound with a heavy wire, while the inner tube which rotates upon its axis is wound with a finer wire.

In a radio circuit, in order to do fine tuning, we must often take recourse to the condenser which instrument is used to do just what its name implies, viz., condensing the electric current. This is perhaps not exactly accurate, for there is no condensing done in radio work, but rather storing of energy.

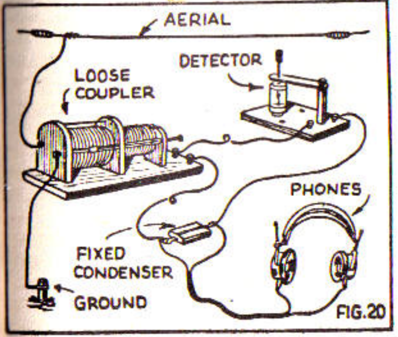
FIG. 20

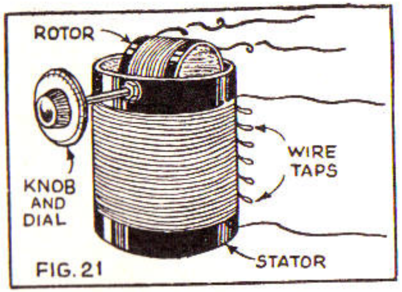
FIG. 21

Consider Fig. 22. Here we have a spring which we compress by means of a weight. As soon as we take the weight away, the spring returns to its original position. What have we done? We have simply stored energy into the spring. The electrical condenser is used in exactly the same way, viz., to store electrical energy. However, that is not its only purpose. Just as the spring may be used for other purposes besides that of storing mechanical energy, so the electrical condenser may be used for other purposes also.

A condenser is a capacity or a vessel in which electrical energy is stored. The simplest form of electrical condenser is shown in Fig. 23, where we have a metal plate A, a glass plate B, and another metal plate C. By means of this arrangement, we may store electrical energy upon the surfaces, of glass plate B. The larger we make the metal plates, the more electrical energy may be stored.

The form shown in Fig. 23, is

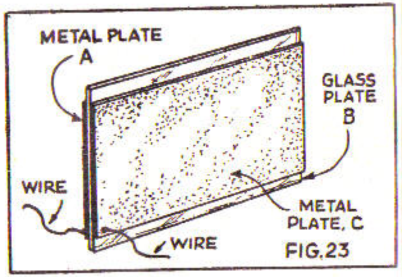
METAL PLATE
A
GLASS PLATE
B
WIRE
METAL PLATE, C
WIRE

FIG.23

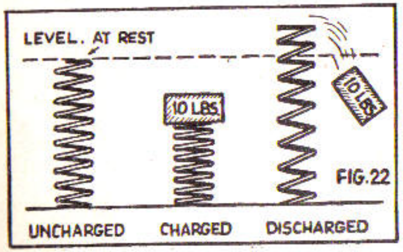
LEVEL. AT REST
10 LBS
10 LBS
UNCHARGED
CHARGED
DISCHARGED

FIG.22

used in many condensers today. The metal plates A and C may be any form of metal, such as, for instance, tin or metal foil, while the glass plate B may be replaced by a piece of paraffin paper. In other words, any good metallic conductors may be used if coupled with a good insulator. The better the insulator, the better the condenser will be and the greater its electrical capacity. In the commercial condensers, paraffin paper, varnished silk, sulphur, sealing compounds, or mica, is usually used. In Fig. 24, is shown a simple condenser; this is also shown opened up. It is made by rolling together two strips of tinfoil between several strips of paraffin paper. The whole, when rolled together and assembled, becomes the finished condenser. By rolling it together, it takes up less room. In radio work, where fine regulation is required, we make use of still another condenser, as shown in Fig. 25. This condenser, instead of being fixed, is variable. As will be seen there are a number of plates which are usually made of brass, zinc or aluminum, which mesh into each other to a more or less degree. The more plates we have and the closer they come to each other, the higher will be the capacity of that condenser. For certain purposes we need only a small condenser of a few plates, while in others we need a larger one of a great many plates. It is just like having a small spring and still another very large one. Both have their uses, and both are very necessary, all depending upon what work they are required to do. In Fig. 26, we show the simplest elementary connection, where we have a crystal detector, a tuning coil without a slider, a pair of phones and a telephone condenser. This is a peculiar connection because in it we wish to show that we can tune by means of the condenser. As will be seen in this tuning coil, we do not use any slider by which the length of the aerial may be changed, which would thus change the wave length. This is performed entirely by the variable condenser. When we adjust the latter, we also change the relation of the tuning coil, and in fact are changing the wave length until a point is found where the signals come in best. This is a finely balanced circuit, and the amount of wire on a little tuning coil should be in direct relation to the condenser. In other words, if there is too much wire on the tuning coil and the capacity of the condenser is small, we cannot do much tuning. For the best results, as for instance for receiving broadcast music on a wave length of 360 meters, we could

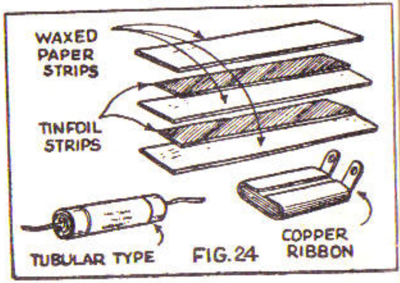
WAXED PAPER STRIPS
TINFOIL STRIPS
TUBULAR TYPE
COPPER RIBBON

FIG.24

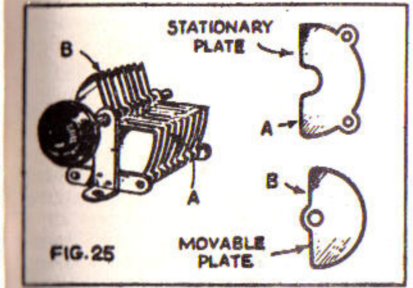
STATIONARY PLATE
B
A
A
B
MOVABLE PLATE

FIG.25

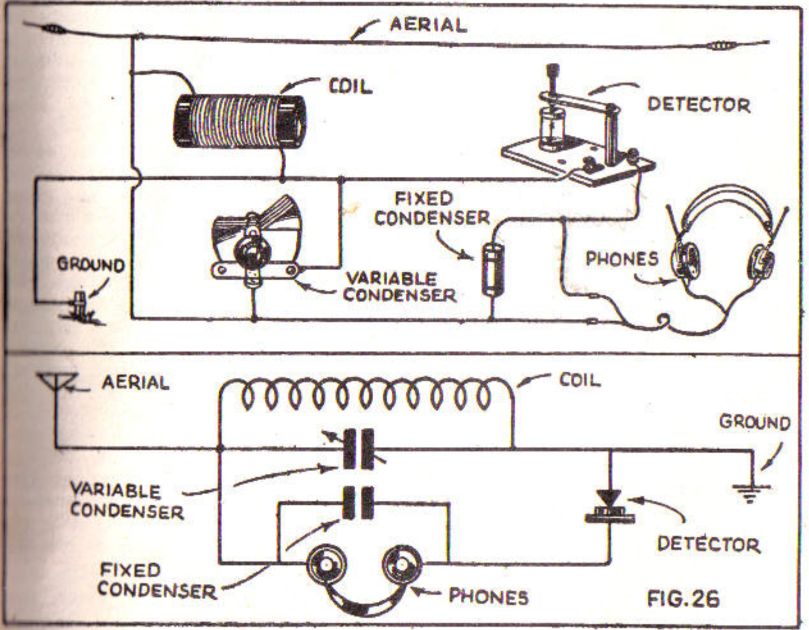
AERIAL
COIL
DETECTOR
FIXED CONDENSER
GROUND
PHONES
VARIABLE CONDENSER
AERIAL
COIL
GROUND
VARIABLE CONDENSER
DETECTOR
FIXED CONDENSER
PHONES

FIG.26

use a small coil, one inch in diameter, wound with about 70 or 80 turns of No. 18 enamel wire, while the condenser should be of the commercial variety known as a 23-plate condenser. Then, all we have to do is simply adjust the condenser until the signals are heard best.

In this illustration, we also see where the phone condenser is located. This phone condenser stores the energy of the circuit, and discharges it into the telephone receivers which enables us to hear the signal more loudly.

The two forms of condensers shown here are of course not the only ones, as many more types of either fixed or variable condensers are made.

VACUUM TUBE ACCESSORIES

We have learned something about the vacuum tube which was described previously in this article, and in Fig. 15 we have shown the simplest connection of an audion detector. There are, however, a number of other auxiliary instruments used in vacuum tube systems which give certain refinements.

The vacuum tube, when it is used singly, acts as a detector and detects the signals the same as a crystal detector. Also, we might state here, that the crystal detector is a better rectifier "valve" than the vacuum tube.

With a crystal detector, or in the ordinary single vacuum tube circuit, the incoming signals act upon the phones and we hear the signals with a certain strength. Let us now consider the vacuum tube and the incoming signal. We may indeed, by certain means boost up the very weakest of signals and amplify or magnify it a hundred or a thousand or a million times its original strength. It is just as if you take a piece of film such as is used in a moving picture theatre and examine it with your eye. The figures are so small that you can hardly distinguish them. The regular film which is about the size of a postage stamp here stands in our analogy as a single vacuum tube. We can, however, take that film, and by using a powerful light enlarge the little picture (no larger than a postage stamp,) by projecting it upon the screen. We thereby amplify or magnify the original picture several thousand times. We can amplify or enlarge it a million times if necessary all depending upon the amount of light we put behind a film and the distance from the screen. This is graphically shown in Fig. 27.

We may do precisely the same thing with a vacuum tube, but we must use additional energy, the same as in our film where we use energy (the electric current which produces the light) to project the film upon the screen. In other words, we can take the detector tube and enlarge the original small and weak signal, and boost it up until the sound comes out loud from a loudspeaker, which in our analogy stands for the moving picture screen. The electrical connections for a vacuum tube amplifier are shown in Fig. 28. This is what is called technically a two-step amplifier. We show this connection simply because without it, it is almost impossible to bring home the meaning of the vacuum tube auxiliary instruments with which the reader is as yet unfamiliar. In this circuit we have, as before, the aerial, the ground, the variable condenser, the blocking or phone condenser and several other instruments as well. We find for instance, that several transformers are used, viz., what is technically called the Audio Frequency Transformer.

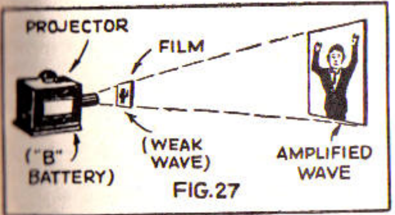

FIG. 27

AUDIO FREQUENCY TRANSFORMER

This transformer (E in Fig. 28), consists of just an iron core upon which is first wound a coarse wire termed the primary, and on top of this a finer wire termed the secondary. The ratio of these transformers is usually such that, electrically speaking, the value of the secondary is from three to ten times as much as that of the primary. The audio frequency transformer is in principle the same as the loose coupler, which we studied before, and the purpose of the audio frequency transformer is to transform the energy from a low level to a high one. The purpose of this transformer, as shown in Fig. 28, is to boost up the weakest signals, transforming them into stronger ones. The transformer by itself could never accomplish this, and in order

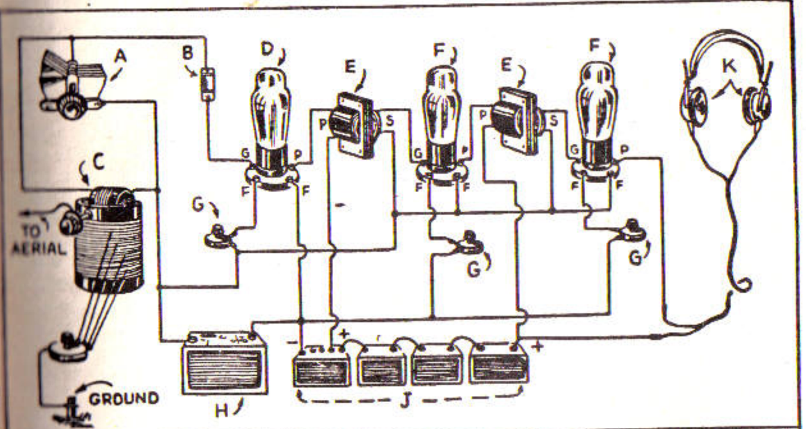

A- VARIABLE CONDENSER ; B-GRID LEAK AND CONDENSER ; C-VARIO-COUPLER ; D-DETECTOR TUBE ; E-AMPLIFIER TRANSFORMER , F-AMPLIFIER TUBE , G-RHEOSTAT ; H-STORAGE BATTERY J-"B" BATTERIES . K- EARPHONES OR MAGNETIC LOUDSPEAKER FIG. 28

to make the lever action work perfectly, we take recourse to a battery which is connected to the transformer and with the vacuum tube, as shown. By means of this additional electrical energy we are now in a position to boost up and relay the weak signal. In this connection, we have shown first a detector tube, while the other tubes are amplifier tubes. By this we mean that the first tube receives the signal, while the other tubes are merely used as pumps to boost up the electrical energy until the signal finally comes from the phones so loudly that if we connect the phones with a loud speaker, loud signals or music will issue from it. The battery used in this case is a so-called "B" battery, or high voltage battery which has been found necessary to aid in boosting up the weak signals. As a rule batteries anywhere from 24 to 300 volts are used, all depending upon the circuits.

It should be understood that the audio frequency transformer is used only to boost up the weak signal as it leaves the first detector tube. It is not in the province of this transformer to do anything save amplify the signal which is detected by the detector tube.

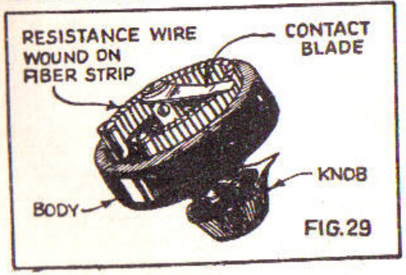

FIG. 29

RHEOSTATS

In Fig 28, we also find another new instrument, the Rheostat, shown in detail in Fig. 29. This is simply an electrical resistance and is used solely to increase or decrease the glowing of the vacuum tube filament. When signals are received, it has been found that the filament must glow at a certain intensity. Some signals come in best when the filament is burning very brightly, while with other tubes the signals come in best when the filament is only a cherry red.

GRID LEAK

In Fig. 28 we have another newcomer, which is termed the Grid Leak, and its condenser. It has been found that when the grid condenser is used, as shown in the illustration, the signals will come in about twice as well as if none was used. However, this condenser alone would not be sufficient, for the reason that the accumulation of electrons, which are highly charged electrical particles on the grid of the vacuum tube, would interfere with the normal working of the tube. We must provide a means to let the surplus electrons leak out without, however, letting them out too quickly. It is as if we had a boiler under which a constant fire was maintained. In order to provide a remedy, we install a safety valve. This valve is used for the purpose of giving off the surplus steam and so keep the boiler free from harm. It is the same with the vacuum tube. While of course, the vacuum tube would not burst, even if we did not use the grid leak, electrically speaking, the tube would not function properly. Hence, the grid leak, which is a sort of safety valve to let the surplus accumulation of electrons run off. The grid leak is nothing but a very high resistance, sometimes millions of ohms high. It may consist of only pencil lines drawn across a piece of stiff Bristol board; these pencil lines are but slight electrical conductors, but the resistance is enormous. It suffices, however, to allow the surplus electrons to leak off. There are various ways and means to make grid leaks, and a popular form is shown in Fig. 30. Here we have a piece of cardboard or fibre upon which is traced a fine line in India ink. This line acts the same as a pencil line. The whole is enclosed in a tube to prevent moisture or dust from settling upon the grid leak. Connections are made on the ends by metal clips. Fig. 30 shows a grid leak and condenser combined as two instruments, which are usually used in conjunction. The grid leak condenser is small and is similar to a telephone blocking condenser, and the grid leak is traced by means of China ink upon a piece of fibre; the whole is enclosed in waxed paper.

RADIO FREQUENCY TRANSFORMER

In Fig. 28 we have learned about the audio frequency transformer. We know that this transformer amplifies static and also other disturbances, as well as the signals. For that reason it is not possible to use many such transformers, or, technically termed, many steps of audio amplification. If we use more than three such transformers and their respective vacuum tubes, additional noises are all amplified, and the amount of noise which we get in the phones is tremendous. For that reason, we take recourse to what is termed a Radio Frequency Transformer. The radio frequency transformer may consist of only two windings, one adjacent to the other on a cardboard tube. The simplest form is shown in Fig. 31. The wire used on this is usually exceedingly fine, No. 24 B & S wire, or even thinner. The two windings act upon each other by induction, and do not make connections physically. A radio frequency stage amplifies the weak signals as they are coming in over an aerial. The radio frequency transformer steps up these weak signals, amplifying them and passing them on to the detector tube. We now get the net result, with the detector tube in a position to detect already fairly strong signals which may then be amplified in the audio frequency amplifiers, and boosted up further by a second or a third transformer, if so desired.

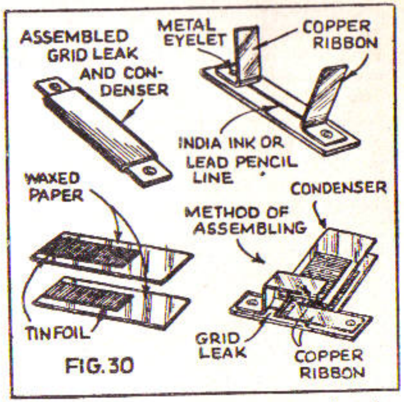

FIG. 30

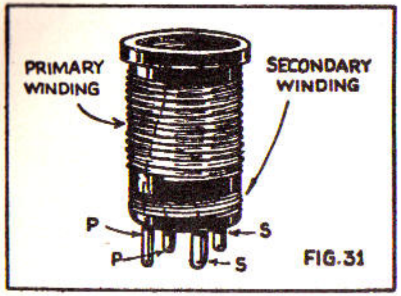

FIG. 31

To resume and in a few words, we may say that the radio frequency transformer boosts up the very weak radio frequency currents so that the detector gives maximum results, whereas the audio frequency boosts up the audible signals. The radio frequency transformers, in other words, amplify signals that would be lost otherwise, while the audio frequency transformers give volume to signals which are already audible.

TELEPHONE RECEIVERS

In order to receive signals or broadcast entertainment by ear, we use a telephone receiver, of which two simple types are shown in Fig. 32; this consists of the following: First we have a powerful magnet which attracts to it a thin iron diaphragm. This diaphragm is clamped tight like a drum head along its outer edge. Upon the magnet are mounted two pole pieces around which are wound many thousand turns of exceedingly fine wire, almost as fine as the human hair.

Ordinarily when no current is sent into the telephone receiver, the diaphragm is pulled down somewhat to the pole pieces, although it must never touch them. If it does, no sound will be received. If, however, a weak electrical current passes through these spools the diaphragm will either be pulled down more if the current is in the right direction, or if the current is in the wrong direction, it will weaken the magnetism on the pole pieces. In this case, the diaphragm is not attracted. These little variations make the diaphragm vibrate more or less. These vibrations are passed on to the air, and the air vibrating in unison with the diaphragm is changed into sound waves, which are sent on to our ear, where we hear them.

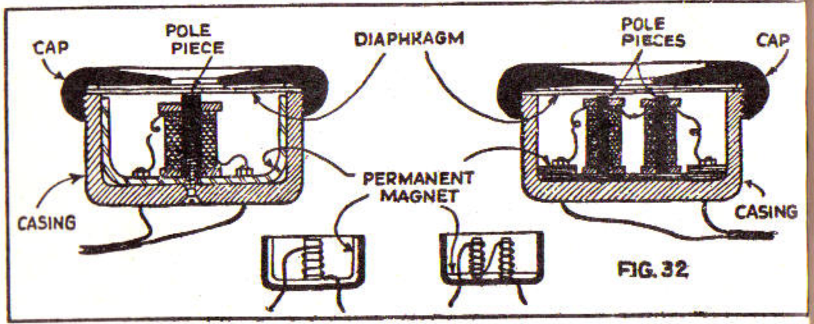

FIG. 32

LOUD SPEAKERS

The dynamic loud speaker, now so widely used, works upon a principle where a small coil, through which the received current flows, is influenced by a powerful electromagnet. It is another case of boosting up the sound which is received from the last amplifier tube. Such tone amplifiers can throw the voice or music over distances of one-half mile and more, and if a person stands in front of one of these giant horns, the amount of sound that issues from it is simply terrific. Of course, not all tone amplifiers work so loudly. Those made for home or parlor purposes do not use so much current, and therefore do not give so much power. There are a number of types of tone amplifiers, but most of them work along the same electro-magnetic lines, and if they do not use the outside electric current in order to create a strong electromagnetic field, they either use strong magnets to accomplish the same result, or necessitate the use of a high tension current in the amplifier. Such tone amplifiers are nothing but transformers or relays, transforming or relaying a weak sound into a loud one.

●

Chapter VI

TUNING

In former chapters we learned something about tuning; this is nothing but resonance. We all know the experiment of standing near the piano and singing a certain note into it; when we reach the correct or fundamental note, the piano begins to sound that particular note in sympathy. We may then say that we are in tune with that particular string which sounds in our ears. Likewise in radio, we make use of a similar system, except that we use electrical tuning instead of acoustical tuning. Tuning consists as a rule in merely attuning our aerial electrically to the same length as the aerial that is transmitting to us. In other words, if a broadcasting station is transmitting on a wave of 360 meters, we must attune our aerial to the same wave, namely, 360 meters. If we have an aerial which is 260 meters long, electrically speaking, it stands to reason that we must add 100 meters to this aer-

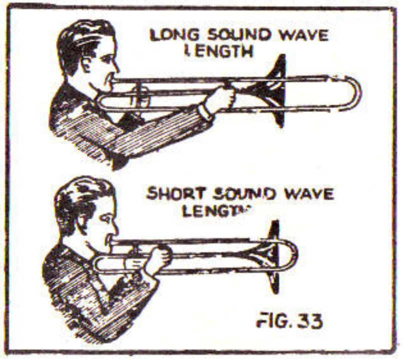

FIG. 33

ial in order to receive the wave at all. We have learned in other chapters how this may be accomplished. If we have a receiving outfit, all we have to do is move the slider of our tuning coil backward and forward until the signals come in at maximum strength. When that point is reached, we know that our aerial, electrically speaking, must be 360 meters long. We have also seen in Fig. 26 that we need not have sliders on the tuning coil in order to tune. We may use a condenser for tuning purposes because its electrical equivalent is the same as a tuning coil slider. By adding more or less capacity to the condenser and therefore to the tuning coil, we changed the electrical value of the tuning coil, and also its wave length. This is not literally true, technically speaking, but we must use this language to bring home the meaning.

We therefore learn that we may tune either by lengthening the aerial with additional wire, or by using a capacity or condenser in connecton with a wire coil. Both, if correctly apportioned, give the same results. Before we can receive signals, or amplify them, it is of the greatest importance that we tune in to the right wave length. An aerial must be in electrical sympathy with the sending station before we hope to receive signals. In Figs. 17 and 26, we have shown the elemental methods of tuning. Of course, there are many other ways of tuning, all of which, however, are along the same principles as those just enumerated.

Perhaps an analogy in tuning will not be amiss here, and we have a particular analogy that covers tuning nicely. Take the musical instrument, the trombone shown in Fig. 33. You all have seen this instrument, as nearly every orchestra boasts of one or more. It is known by all of us that while the musician blows into the mouth piece, he varies the length of the trombone by moving the sliding member back and forth. If he wants to get a deep note, he pulls the sliding member almost all the way out, and this gives him a long sound wave. If he wishes a high note, he must have a short sound wave. This means that he must push the sliding member all the way in. It is literally, as well as scientifically, true that the lengthened trombone gives a long wave length, while the shortened trombone gives a short wave length. These are, of course, sound waves with which we have to do here. In radio we do exactly the same thing in tuning. When we wish a long wave length, we must add more wire or its equivalent to the aerial. If we want a short wave length we must either have a short aerial or subtract some wire from the aerial. The reader should remember that in order to decrease the wave length of an aerial, all that is necessary is to put a condenser in series with the aerial, which actually decreases the wave length; it does not increase it as some people seem to think. The variable condenser, therefore, gives us the best practical means to decrease the wave length; this point is quite important to remember. Suppose you have a long aerial, say 200 feet, in connection with a small tuning coil, or suppose you have a short aerial and live on the tenth floor of an apart-

ment house. The only available ground would be the water pipe. This water pipe, however, would be so long that it would add extra meters to your wave length, and something must be done to decrease it, if you wish to receive signals sent out from a broadcasting station operating on a short wave length of 360 meters. The only way you could then tune in would be in the former case of the long aerial, to put a variable condenser in the aerial circuit, or in the other case where you have a long ground to interpose the condenser in the ground lead.

●

Chapter VII

AERIALS, LOOP AERIALS, GROUNDS

An aerial is used to intercept radio waves; that is its sole function in the receiving set. It does not amplify or make the signals come in clearer by itself. Hundreds of different aerials have been invented, and there is hardly anything in this field that has not been tried out. An aerial, properly speaking, is an elevated wire that is well insulated, and is usually placed outside of the building or house.

An aerial can be made of most any metallic wire, but the best material is copper wire. A still better wire to use is a stranded wire, which is composed of several copper or phosphor bronze wires twisted together. As a rule, we may say that the larger the wire, the better it is for radio purposes. Very thick wires, as a rule, cost much and are very heavy, and therefore are not very practicable. A No. 14 B & S gauge wire is a standard as used today and gives excellent results. For radio broadcast reception it has been found that a single wire 50 feet long gives excellent results. Illustrations 34 and 35 show such a type of aerial.

Unless you wish to go to a great deal of inconvenience, make your lead-in of the same wire as the main aerial. This may be done very simply with a single wire aerial, for the reason that no soldered connections are necessary. This is also shown in Fig. 34 and Fig. 35. The next things to consider are the insulators, which are quite important. The insulator serves to insulate the aerial, and unless we use good ones, a great deal of energy will be dissipated uselessly. We show in Fig. 36 various types of insulators that may be used. One of the simplest is the ordinary porcelain cleat, but when this type is chosen, an unglazed cleat should be avoided. Insist upon getting a glazed cleat which is a better insulator. When using cleats, put them in tandem, two or three strung in a row, as shown. The more insulators we add, the better the insulation.

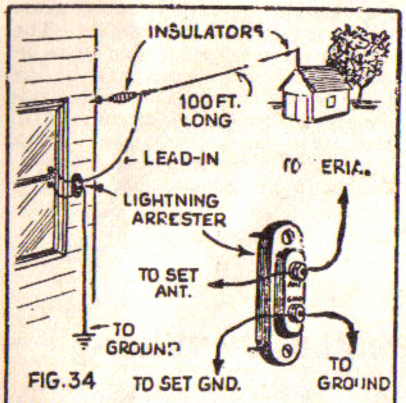

FIG. 34

It is, however, hardly necessary to use more than three in a row. We next have the small spool insulators, which are very good and may also be strung in pairs, or sets of three. Various other types are shown.

When putting up an aerial, it should be remembered always that the aerial proper must be at least a foot away from all buildings, barns, trees and the like. In other words, it should be away from all objects.

The height of the aerial is often important. It should always be placed at least 20 to 30 feet above the ground.

As a rule, an aerial in the country may be stretched from the attic window to a flagpole, or if such is not at hand, a barn, garage, or even a tree could be made use of. If a tree is used, some means must be had to compensate for the swaying of the tree. Such a method is to have a pulley attached to a tree by means of a rope; the end of the aerial is then run over this pulley and a fairly heavy weight secured to the open end. As the tree sways back and forth, more or less aerial rope is paid out or taken in, and a good compensation is thus had. The weight may be 50 to 100 pounds.

When an aerial is erected in the city, let us say on an apartment house, it should be at least 10 feet above the roof, particularly if the apartment has steel construction.

LEAD - IN

The lead-in is that part of the aerial that goes into the building or house to establish connection with the instruments. In a single-wire aerial, the lead-in is simply the aerial wire itself leading into the house and thence to the receiver. The lead-in wire should be of the same size as the aerial. In other words, about No. 14 B & S wire. It should be insulated at the point where it nears the building, or if this is not possible in the case of a single-wire aerial, the lead-in is strung on insulators, the wire being always at least 2 inches away from buildings,

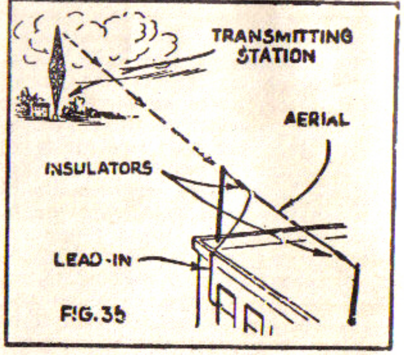

FIG. 35

walls, etc., until it reaches the points where actual entrance is made into the building.

We now come to an aerial which is entirely different from those of which we have spoken before. We refer to the loop aerial, which is shown in Fig. 37. It should be understood that a loop aerial is hardly, if ever, used in connection with a crystal set. It is used almost exclusively with a vacuum tube set, where it serves several purposes. In the first place, it does away with the ground connection. Secondly, the loop aerial may be made in any size from a few inches square up to 20 feet square. The loop aerial is highly directive; by that we mean that it will only receive with maximum intensity if the loop is turned in the direction of the coming signals. This is shown clearly in Fig. 38. Here we see how an ordinary loop aerial is placed in a building and we also see how the waves are propagated from a distant sending station. It will be found that the signals are strongest when the loop points exactly in the direction from which the waves are coming.

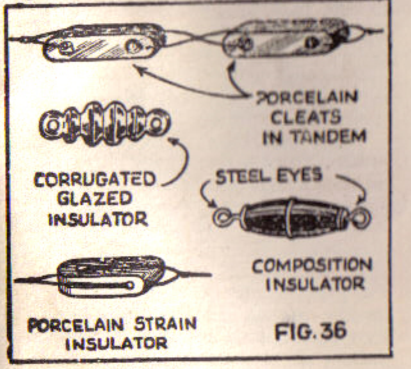
PORCELAIN CLEATS IN TANDEM
CORRUGATED GLAZED INSULATOR
STEEL EYES
COMPOSITION INSULATOR
PORCELAIN STRAIN INSULATOR

FIG. 36

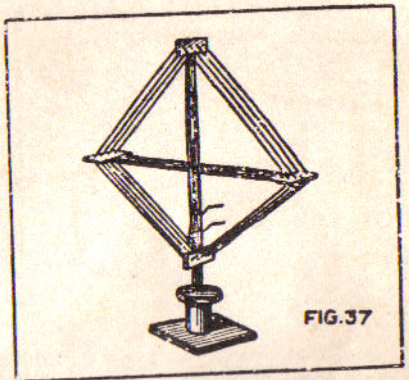

FIG. 37

GROUNDS

In radio, in connection with the usual aerial, it becomes necessary to use a ground, which as its name implies, is a connection made with the earth. Fig. 39 shows the simplest and perhaps the best. It is simply a wire fastened to the cold water pipe, which is found in almost every house and apartment. In order to make a good connection, we use a ground clamp, as shown in Fig. 39. By means of some clamping arrangement, which differs for every ground clamp, a strong mechanical connection is made. The ground wire is then fastened to the screw or binding post attached to the ground clamp. The ground wire need not be insulated. An ordinary bare No. 14 B & S wire will do nicely; in other words, the same wire which we use on an aerial may be used. It is not necessary to run the ground wire on insulators, as is done with the aerial

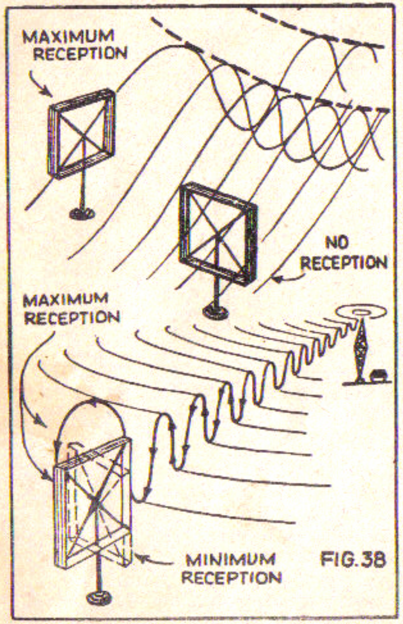

FIG. 38

lead-in, but it may be attached to the wall by means of nails which serve the purpose equally well. Of course, the ground wire should not be longer than is absolutely necessary. If it is not possible to find a cold water pipe, a radiator pipe may be used, although the results may not be as good as from the cold water pipe. It is against the law to connect a ground to a gas pipe, and it should therefore never be done.

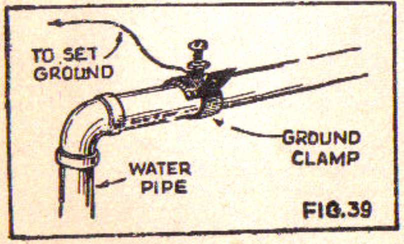

FIG. 39

When we are out in the country, for instance, when camping, it is not always possible to have a water pipe, and in that case we have to establish contact with Mother Earth direct. This is usually accomplished by driving a metal rod into moist earth, as is shown in Fig. 40.

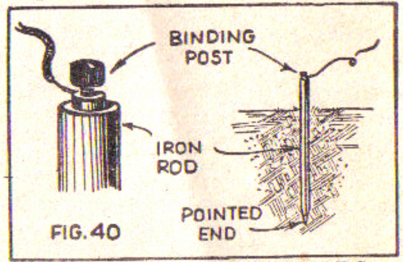

FIG. 40

LIGHTNING ARRESTERS

The properly installed aerial, when used with a lightning arrester, is the best protection a building or house could have against lightning. The aerial is a lightning conductor itself, and will actually protect the house, and will never endanger it if properly installed.

The lightning arrester itself is nothing but a small spark gap either in a vacuum or in the atmosphere, which gap breaks down when a current of a few hundred volts strikes the aerial. Instead of going through the instruments which have a high resistance, the current travels directly to the ground, which has a low resistance. Secondly, the instruments are not damaged.

Here Are Nine Additional Titles in the Famous
GERNSBACK'S EDUCATIONAL LIBRARY:

No. 1 How To Make Four Doerle Short Wave Sets
No. 2 How To Make The Most Popular All-Wave 1- and 2-Tube Receivers
No. 3 Alternating Current for Beginners
No. 4 All About Aerials
No. 9 Simple Electrical Experiments
No. 5 Beginners' Radio Dictionary
No. 6 How To Have Fun With Radio
No. 7 How To Read Radio Diagrams
No. 10 Television

10c. EACH

BOOKS ARE ALL UNIFORM
Every book in the GERNSBACK'S EDUCATIONAL LIBRARY has 32 pages—with illustrations varying from 30 to 66 in number. Each title volume contains over 15,000 words. POSITIVELY THE GREATEST 10c VALUE IN RADIO BOOKS EVER OFFERED TO THE PUBLIC.

Order books by number — remit by check, money order, cash or unused U. S. Postage Stamps. Books are mailed POSTPAID.

RADIO PUBLICATIONS . 25 WEST B'WAY - NEW YORK, N.Y.

April 18, 1933. **C. J. AANENSON** 1,904,773

RADIO GROUND

Filed March 10, 1931 2 Sheets-Sheet 1

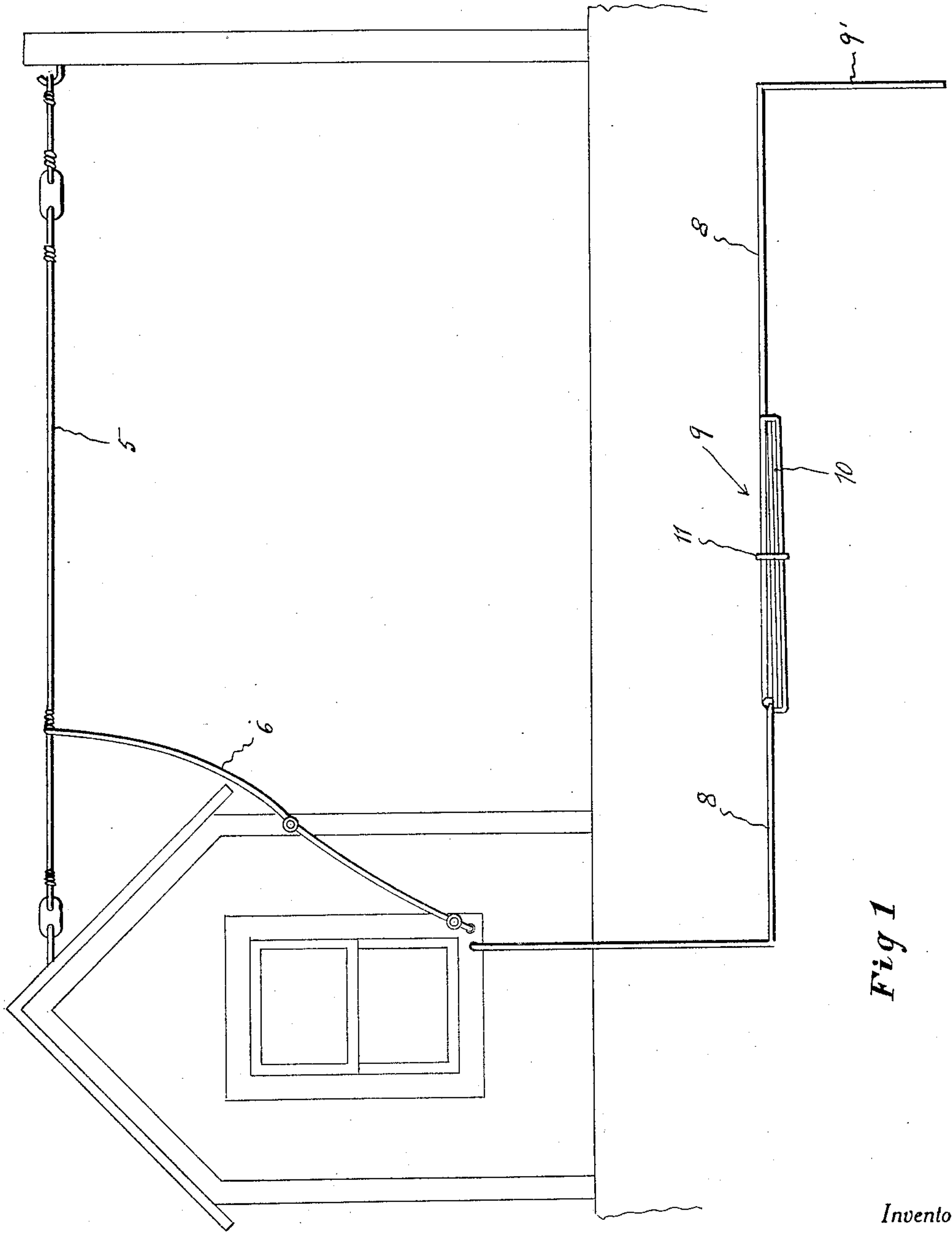

Fig 1

Inventor

Christ J. Aanenson

By Clarence A. O'Brien
Attorney

April 18, 1933. **C. J. AANENSON** 1,904,773

RADIO GROUND

Filed March 10, 1931 2 Sheets-Sheet 2

Fig 2

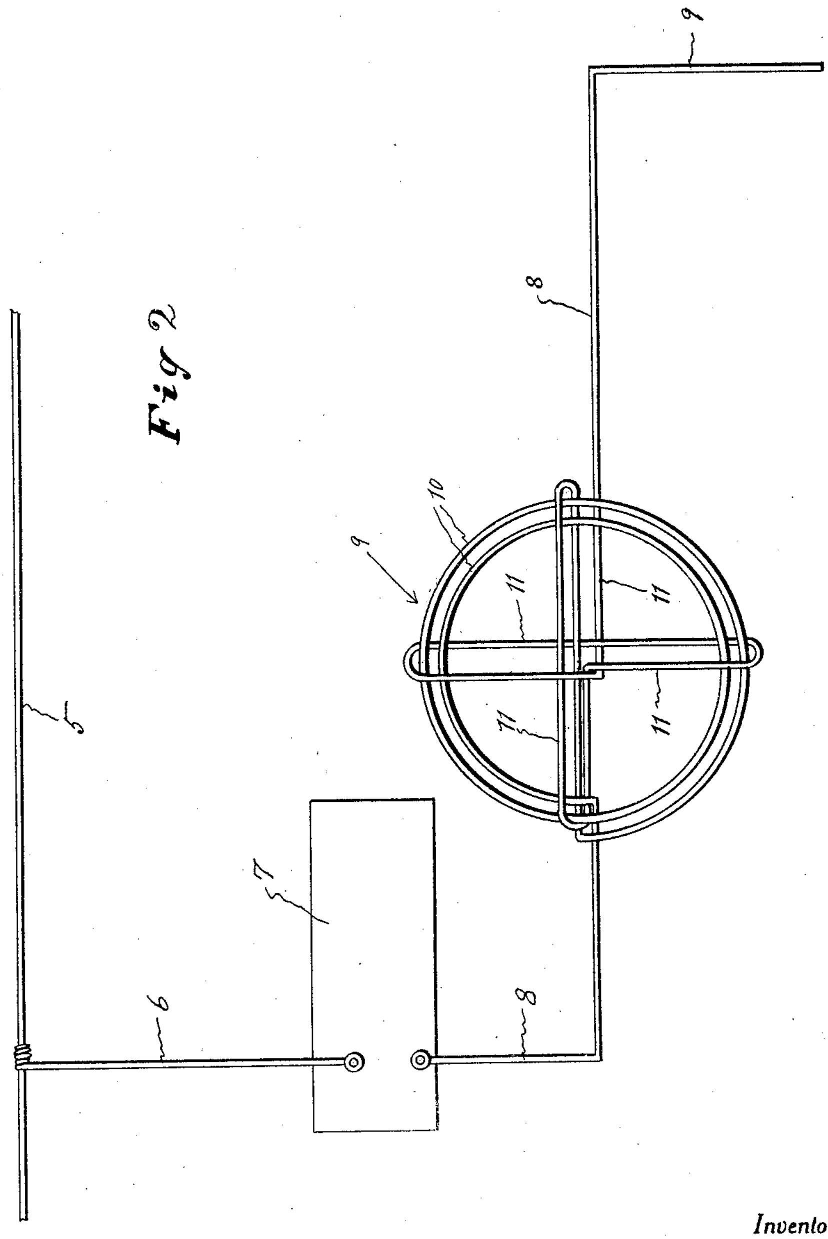
9'
8
10
9
11
11
11
11
5
7
6
8

Inventor

Christ J. Aanenson

By Clarence A. O'Brien

Attorney

Patented Apr. 18, 1933 **1,904,773**

UNITED STATES PATENT OFFICE

CHRIST J. AANENSON, OF MAGNOLIA, MINNESOTA

RADIO GROUND

Application filed March 10, 1931. Serial No. 521,539.

This invention appertains to improvements in ground devices for radio apparatus.

The principal object of this invention resides in the provision of a novel radio ground device for the elimination of fading of volume, increase of range and the promotion of selectivity.

Other important objects and advantages of the invention will become apparent to the reader of the following specification and claim.

In the drawings:—

Figure 1 represents a diagrammatic view disclosing the ground and the manner in which it is associated with the usual aerial.

Fig. 2 represents a diagrammatic view disclosing in plan the formation of the annular construction of the ground.

Referring to the drawings wherein like numerals designate like parts, it can be seen that numeral 5 represents the usual antenna from which extends the lead-in wire 6 connected to the usual antenna post on the receiver 7 as shown in Fig. 2.

Extending from the ground post of the receiver 7 is the wire 8 of suitable constructive material, which leads to the annular formation generally referred to by numeral 9. This formation along the horizontal portion of the wire 8 is approximately ten inches under the surface of the earth. Extending from the formation 9 is the same wire 8, of which the formation 9 is constructed and this leads to the downwardly disposed portion 9′ which may be in the form of a peg of copper or some other suitable conductive material.

It will be observed in Fig. 2, that the annular formation 9 simply consists in the wire 8 into several convolutions 10 and training the same diametrically across the convolutions to provide right angularly disposed spoke portions 11. This annular formation is disposed horizontally under the earth in the manner shown in Fig. 1 and it is submitted that this annular formation 9 materially increases the efficiency of the ground device.

While the foregoing specification sets forth the invention in specific terms, it is to be understood that numerous changes in the shape, size and materials may be resorted to without departing from the spirit and scope of the invention as claimed hereinafter.

Having thus described my invention, what I claim as new is:—

In a ground terminal for radios, a single length of wire having an intermediate portion formed with a plurality of convolutions and diametrical elongated loops, said loops being at right angles to each other and having the intermediate portions thereof located approximately at the axis of said convolutions and being interlocked, said convolutions being extended through and being restrained by the outer portions of said elongated loops.

In testimony whereof I affix my signature.

CHRIST J. AANENSON.

TUTTA LA Radio

Supplemento al N. 5 di Tecnica Pratica
Spedizione in abbonamento postale - gruppo III

L. 500

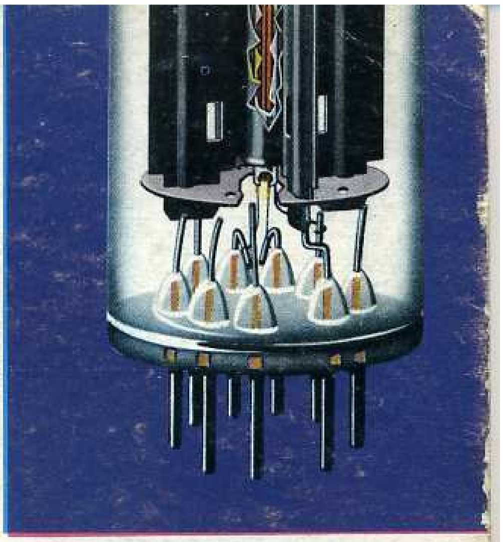

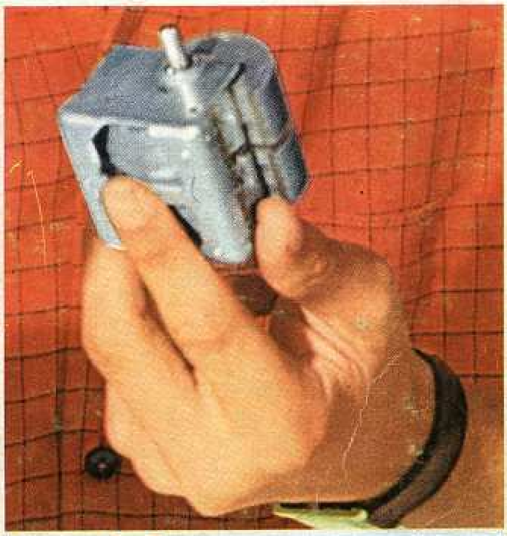

in 36 ore

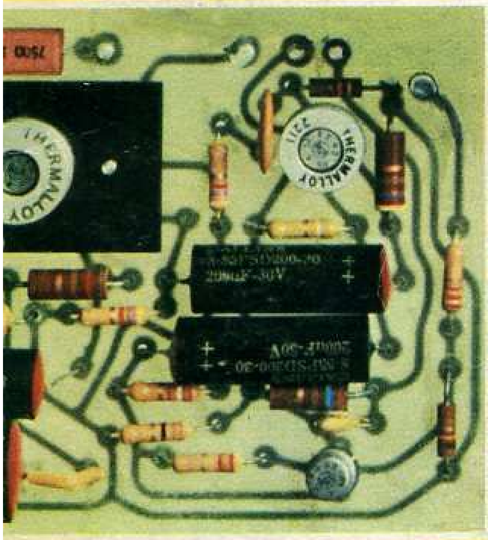

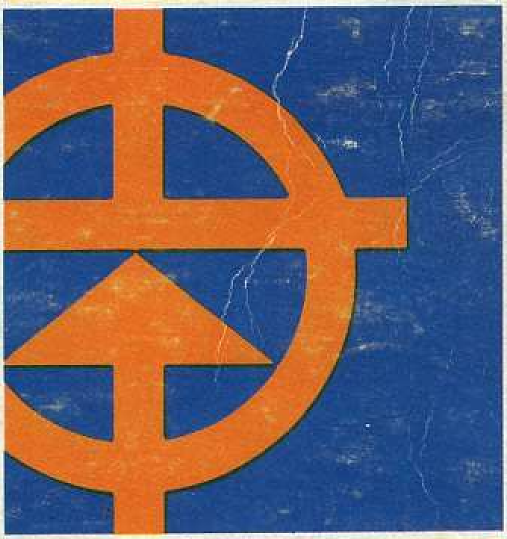

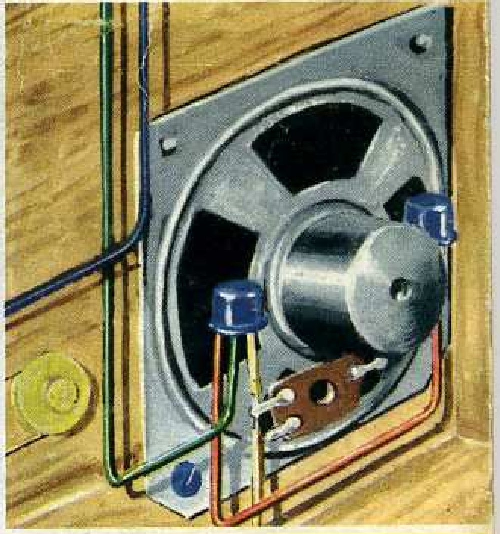

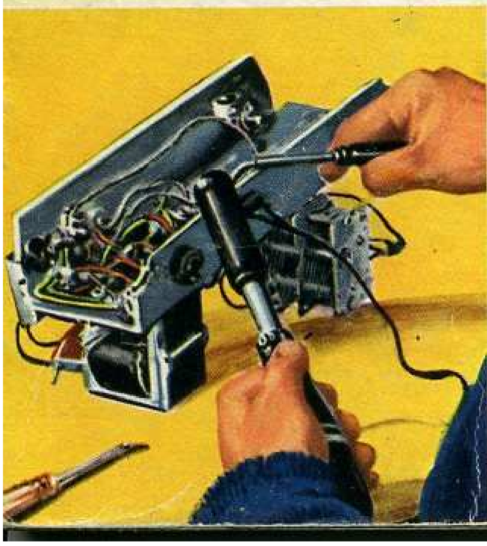

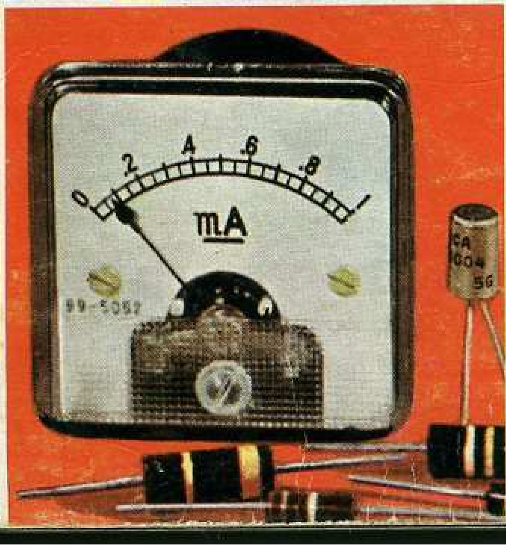

tecnica
pratica
L. 200
COSTRUITEVI QUESTO RICETRASMETTITORE SULLA GAMMA DEI 20 mt

tecnica
pratica
L. 200
IL PIÙ BEL REGALO DI NATALE!
MONTATELA VOI STESSI

ESPERIENZE DI RADIO ELETTRONICA
L. 200
tecnica
TV - FOTOGRAFIA
COSTRUZIONI
pratica
MONTATELO
VOI STESSI!

I FASCICOLI ARRETRATI

di tecnica pratica

SONO UNA MINIERA D'IDEE E DI PROGETTI

Fate richiesta di uno o più fascicoli arretrati inviando la somma di L. 300 (comprese spese di spedizione) anticipatamente a mezzo vaglia o C.C.P. 3/49018 intestato a « TECNICA PRATICA », Via Gluck 59, Milano. Ricordiamo però che i fascicoli arretrati dallo aprile 1962 al gennaio 1963 sono TUTTI ESAURITI.

SONO DISPONIBILI SOLO DAL FEBBRAIO '63 IN AVANTI

ESPERIENZE DI RADIO ELETTRONICA
L. 250
tecnica
TV - FOTOGRAFIA
COSTRUZIONI
pratica
provate l'emozione d'inviare la vostra voce nelle vie dell'etere
UN BEL TRASMETTITORE CHE TUTTI POSSONO COSTRUIRE
CON SOLE 4500 LIRE

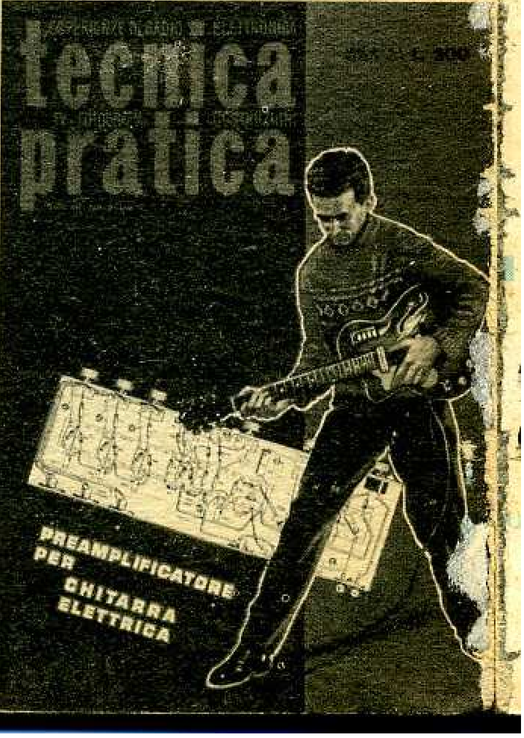
ESPERIENZE DI RADIO ELETTRONICA
tecnica
TV - FOTOGRAFIA
COSTRUZIONI
pratica
L. 300
PREAMPLIFICATORE PER CHITARRA ELETTRICA

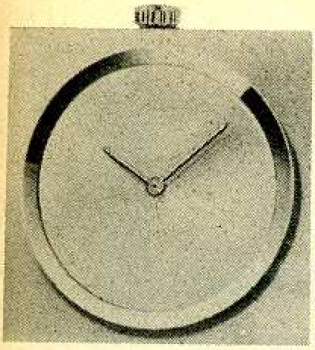

TUTTA LA RADIO IN 36 ORE!

sommario

1ª ora	Un regno di fili e di colori - Dalla emittente alla ricevente - La radio in simboli - Il circuito radio.	pag. 4
2ª ora	Il più semplice radioricevitore.	pag. 10
3ª ora	Il più semplice radiomontaggio - L'astuccio contenitore - Procedimento meccanico.	pag. 14
4ª ora	Cablaggió - La saldatura - I terminali del transistore - Il calore guasta il transistore - Le polarità del diodo.	pag. 17
5ª ora	L'interruttore - L'ascolto - Costruzione dell'antenna - Sintonizzazione.	pag. 22
6ª ora	La corrente elettrica - Misura della corrente elettrica.	pag. 25
7ª ora	Le resistenze - Misure delle resistenze - Codice a colori delle resistenze.	pag. 27
8ª ora	Il circuito a reazione - I condensatori fissi.	pag. 30
9ª ora	Montaggio del ricevitore a reazione - Collegamento di pile in serie.	pag. 33
10ª ora	La tensione elettrica - Modulazione e rivelazione.	pag. 36
11ª ora	Il circuito reflex.	pag. 38
12ª ora	Montaggio del ricevitore reflex.	pag. 41
13ª ora	Ricevitore reflex con amplificazione B.F. - Il trasformatore d'uscita - L'altoparlante.	pag. 44
14ª ora	Montaggio del ricevitore con amplificatore B.F.	pag. 43
15ª ora	Potenza elettrica delle resistenze.	pag. 50

Numero unico. Supplemento al N. 5 del maggio 1966 di Tecnica Pratica. Direttore responsabile A. D'Alessio. Autorizzaz. del tribunale di Milano numero 6156 del 21-1-63. Redazione: Via Gluck, 59 - Milano - Distribuzione: Messaggerie Italiane - Via Carcano, 32 - Milano. Tipolitografia « La Veltro » Via Brunelleschi, 26 - tel. 9.12.13.26 - Cologno Monzese.

Prezzo di una copia L. 500

EDIZIONI CERVINIA

16ª ora Potenziometri a grafite e a filo - Volt-lavoro dei condensatori. pag. 51

17ª ora Le valvole elettroniche. pag. 53

18ª ora Funzionamento delle valvole Diodo. pag. 55

19ª ora Diodo raddrizzatore e rivelatore - Triodo - Tetrodo e pentodo. pag. 59

20ª ora Ricevitore a una valvola - Alimentatore - Trasformatore di alimentazione - Il cambiotensione. pag. 61

21ª ora Montaggio del ricevitore a una valvola - Il raddrizzatore al selenio - Il condensatore elettrolitico a vitone - Cablaggio del trasformatore di alimentazione. pag. 65

22ª ora Ricevitore a due valvole. pag. 68

23ª ora Ricevitore reflex monovalvolare. pag. 71

24ª ora Montaggio del ricevitore reflex monovalvolare. pag. 74

25ª ora Zoccoli portavalvole. pag. 75

26ª ora Circuito supereterodina. pag. 77

27ª ora Conversione di frequenza. pag. 79

28ª ora Il primo stadio del ricevitore. pag. 80

29ª ora Frequenza dell'oscillatore - Stadio amplificatore di media frequenza. pag. 82

30ª ora Controllo automatico di volume. pag. 83

31ª ora Stadio preamplificatore di B.F. - Stadio amplificatore finale - L'altoparlante. pag. 84

32ª ora Stadio alimentatore - Procedimento meccanico di montaggio. pag. 86

33ª ora Cablaggio del circuito supereterodina. pag. 88

34ª ora Consigli e suggerimenti. pag. 90

35ª ora Interpretazione simbolica e piano di cablaggio. pag. 91

36ª ora Messa a punto e taratura. pag. 94

Entusiasmo ben riposto

Se avete acquistato il presente manuale, ciò significa che avete aderito alla nostra proposta, apparentemente un po' azzardata, a prima vista, forse, troppo lusinghiera. Ma vuol dire anche che intimamente sentivate la necessità di questo nuovo mezzo di cultura popolare. Comunque, per confermarvi la bontà della vostra scelta e per mantenere acceso il vostro ben riposto entusiasmo, è doveroso da parte nostra una breve spiegazione di premessa.

Tutta la radio in 36 ore! Ma è dunque possibile? Possibilissimo, rispondiamo noi! I corsi delle scuole elementari, del liceo, ad esempio, si compiono in 5 anni, in un tempo pari a 43.000 ore. Ecco pertanto che le nostre 36 ore equivalgono a 36 giorni e, per i più fortunati che hanno maggior tempo a disposizione, a 18 o 20 giorni. E in un'ora con la nostra guida si può far molto.

Si possono assimilare piacevolmente gli elementi fondamentali di un settore della materia; ciò perchè il concetto si rivela spontaneo attraverso una sequenza di elementari esercitazioni pratiche e di avvincenti applicazioni.

Con questa moderna meccanica di insegnamento giungerete, ora per ora, a capire tutta la radio. Proprio tutta? Sì, per poter seguire pubblicazioni specializzate. Sì, per poter interpretare progetti elettronici, ma soprattutto per poter realizzare da soli, con soddisfazione, radioapparati più o meno complessi, che altri hanno potuto affrontare dopo lungo e pesante studio.

Da questo momento schiacciate dunque il pulsante del vostro cronometro e partite di buona lena!

EDIZIONI CERVINIA

UN REGNO DI FILI E DI COLORI

Con la radio si comincia sempre allo steso modo. Capita sottomano un manuale, una rivista, un libro di radiotecnica, si dà, per caso, una sbirciatina a quel piccolo regno fatto di arnesi, di fili e di colori che è il radiolaboratorio, così misterioso da una parte e così pieno di fascino dall'altra, e si rimane... contagiati.

Il neofita della radio ha bisogno sempre di una guida amica, paziente e generosa.

È accaduto a noi e accade ogni giorno e in ogni luogo ad altri: ai giovanissimi e agli uomini già maturi. La passione per la tecnica della radio non conosce limiti di età, di classi sociali, di luoghi o di tempo. Esplode improvvisa, impetuosa, all'insegna dell'entusiasmo e un po' del sacrificio, creando dappertutto, e di continuo, schiere nuove ed agguerrite.

Ma il neofita della radio è un po' come il bambino, che ha bisogno della mamma per imparare a muovere i primi passi, per crescere. Da solo, completamente da solo, non può davvero farcela. Una guida amica, paziente e generosa, pronta a consigliare e ad incoraggiare quando ci si trovi a contatto con le prime difficoltà o coi primi insuccessi, è assolutamente necessaria. Sì, perchè quando ci si trova da soli, troppo soli, in un mondo tanto affascinante ma anche tanto vasto, come quello della radiotecnica, è facile smarrirsi e abbandonare presto il campo solo per non aver trovata e seguita la via maestra.

La passione per la radio si traduce spesso in una gran fretta di costruire, di toccare con mano fili conduttori e componenti radio elettrici.

Esistono oggi molte pubblicazioni, specializzate in materia di radio, che tengono il passo con il continuo progredire dell'elettronica, presentando circuiti e progetti nuovi, talora con dovizia di particolari e assai spesso con linguaggio accessibile soltanto ai più preparati. Ma chi non sa nulla di radio ed è preso soltanto da interesse e passione per la radiotecnica, ama e preferisce il linguaggio tecnico più semplice, quello più elementare, di continuo confortato dalla possibilità di toccare con mano ciò che è detto in teoria e di tradurre in pratica almeno i concetti fondamentali. E l'amore per la pratica si traduce, quasi sempre, in una gran fretta di costruire, in un bisogno immediato di raggiungere la meta più ambita: quella di ascoltare voci e suoni, provenienti dall'etere, per mezzo di un piccolo apparecchio, che sia veramente la nostra creatura, composta e realizzata tutta da noi. Dunque, vi insegneremo presto a costruire il vostro primo radioricevitore. Intendiamoci, però! Non si tratterà di un apparato capace di farvi ascoltare molte emittenti e neppure troppo lontane da voi. Con esso riceverete bene le emittenti locali e un po' meno bene le altre; ma, quel che più importa, avrete acceso le polveri del vostro entusiasmo e, badate bene, con una minima spesa e senza aver logorato il cervello in uno studio faticoso, complicato da astruse formule o da lunghi calcoli.

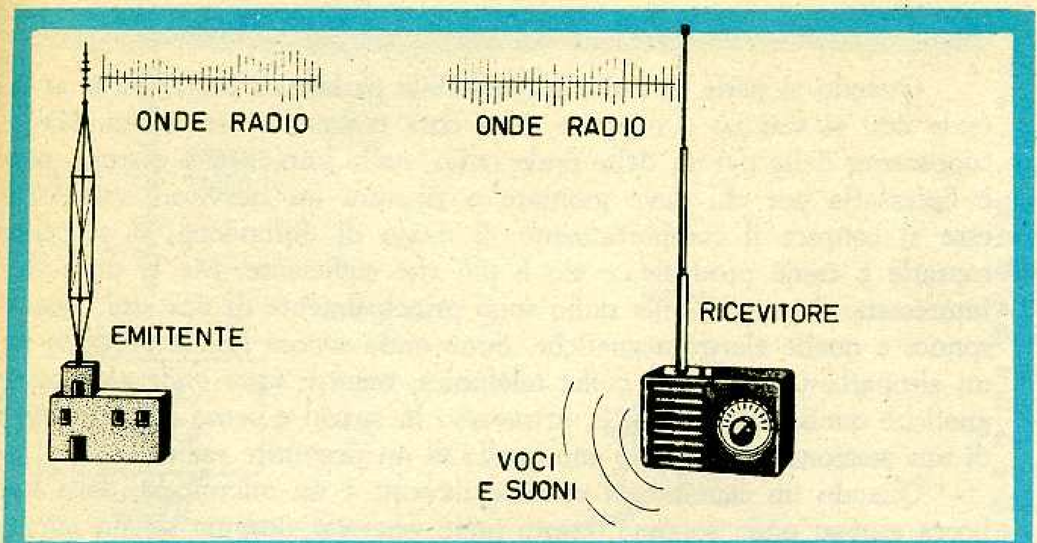
ONDE RADIO
ONDE RADIO
EMITTENTE
RICEVITORE
VOCI
E SUONI

Le onde radio sono invisibili e presenti in ogni dove. Anche le onde sonore, sono invisibili. Le prime, per compiere qualunque viaggio, non abbisognano di alcun mezzo di « sostegno »: esse si propagano anche attraverso il vuoto. Le seconde, per propagarsi, necessitano di un mezzo gassoso, liquido o solido, e il più naturale fra questi è rappresentato dall'aria.

Il microfono capta le onde sonore e le trasforma in corrente elettrica; questa corrente viene rinforzata (amplificata) da un apparecchio, denominato amplificatore. In ogni stazione trasmittente viene generata corrente oscillante, che viene mescolata con quella proveniente dall'amplificatore. La corrente risultante da tale mescolanza viene inviata all'antenna trasmittente dalla quale si diparte nello spazio in forma di onde radio.

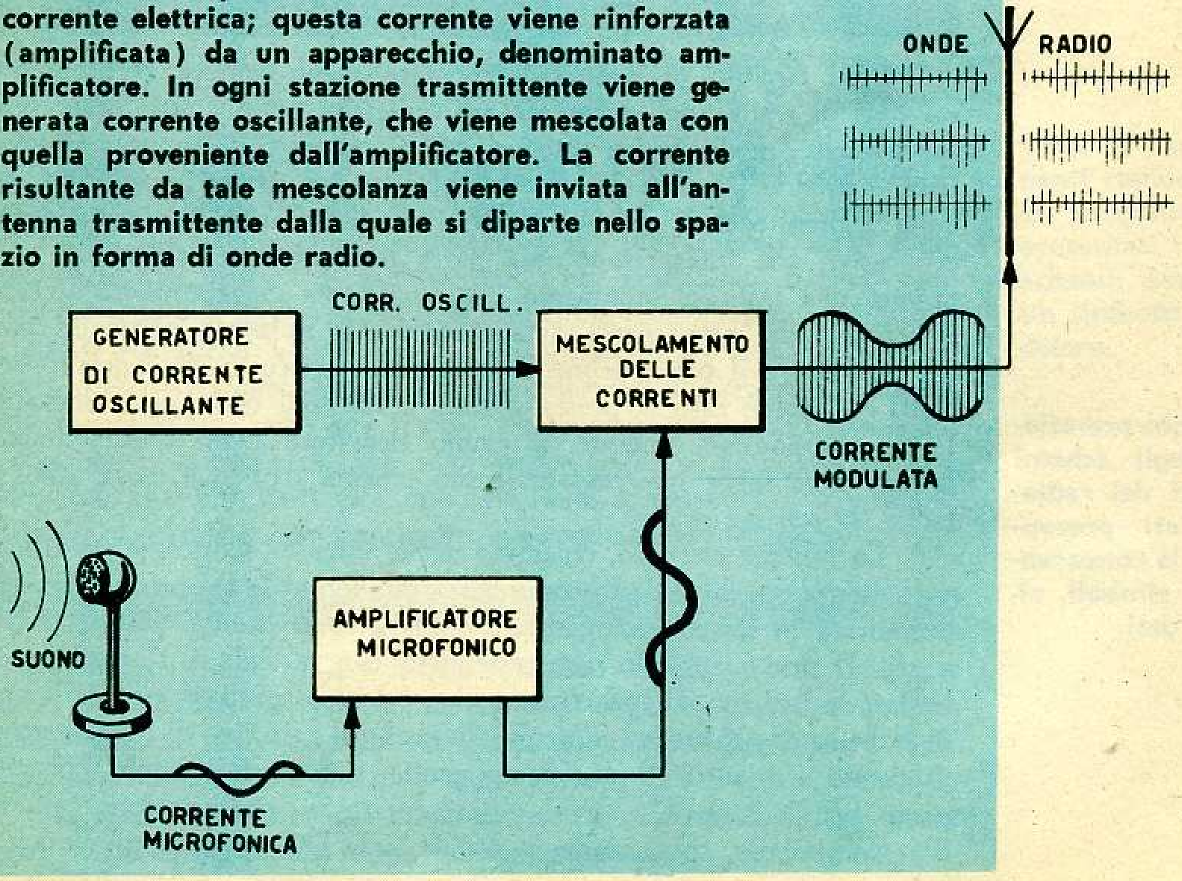
ONDE
RADIO
CORR. OSCILL.
GENERATORE DI CORRENTE OSCILLANTE
MESCOLAMENTO DELLE CORRENTI
CORRENTE MODULATA
SUONO
AMPLIFICATORE MICROFONICO
CORRENTE MICROFONICA

DALLA EMITTENTE ALLA RICEVENTE

Quando si parla di radio è inevitabile parlare di onde, anche se le onde non si vedono e neppure si sa cosa realmente esse siano. Ma la conoscenza della natura delle onde radio, nella loro intima essenza, non è necessaria per chi deve montare o riparare un ricevitore radio. Di esse si conosce il comportamento, il modo di diffondersi, si sa come captarle e come produrle, e ciò è più che sufficiente. Ma le onde che interessano il mondo della radio sono principalmente di due tipi: quelle sonore e quelle elettromagnetiche. Sono onde sonore quelle prodotte da un altoparlante o da una cuffia telefonica, mentre sono onde elettromagnetiche quelle che collegano, attraverso lo spazio e senza fili, l'antenna di una stazione trasmittente con quella di un ricevitore radio.

Due tipi fondamentali di onde: onde sonore e onde elettromagnetiche.

Quando un cantante si esibisce davanti a un microfono, dalla sua bocca escono onde sonore; queste onde vengono ricevute da un microfono e da questo trasformate in corrente elettrica; questa corrente elettrica subisce uno speciale trattamento, attraverso apparati più o meno complessi, e viene inviata, trasformata in onde elettromagnetiche, alla antenna trasmittente. La corrente elettrica che esce dal microfono prende il nome di **corrente microfonica** o **corrente di bassa frequenza**; quella inviata all'antenna trasmittente prende il nome di **corrente ad alta frequenza**.

Ma lasciamo da parte il processo di trasmissione per entrare subito in quello, più interessante per chi legge questo volumetto, della ricezione dei segnali radio.

LA RADIO IN SIMBOLI

Ogni circuito di ogni radioapparato viene sempre rappresentato, sui libri di testo, sugli schemari, sui bollettini pubblicitari, sulle riviste specializzate, mediante un disegno, tutto composto di simboli radioelettrici, e prende il nome di **schema elettrico** o **circuito teorico**. Talvolta il circuito elettrico è confortato dalla presenza di uno o più disegni esplicativi, che interpretano una parte meccanica del ricevitore radio, oppure una fase del montaggio. Ma lo schema maggiormente accettato e ricercato dai principianti rimane sempre lo **schema pratico**, quello che riproduce, il più fedelmente possibile, il circuito reale dell'apparecchio, così come esso si presenta nella realtà agli occhi dell'osservatore. Prima di montare, dunque, il primo ricevitore radio sperimentale, il lettore deve saper interpretare il suo schema teorico e, cosa assai più facile, lo schema pratico.

La interpretazione degli schemi teorici dei radioapparati presuppone la conoscenza di simboli, sigle e dati.

Lo schema elettrico, chiamato anche circuito teorico, è un disegno tutto composto di simboli, come detto più sopra. E nel primo ricevitore, presentato in questo volumetto, i simboli sono ridotti, principalmente, a sei. Il primo simbolo, contrassegnato con la sigla L 1, vuol rappresentare la **bobina di induttanza** che, nel caso specifico, prende il nome di **bobina di sintonia** (vedremo più avanti il principio di funzionamento di questo e di tutti gli altri componenti); il secondo simbolo, contrassegnato con la sigla C 1, vuol rappresentare un condensatore variabile; il terzo simbolo, contrassegnato con la sigla DG, vuol rappresentare un

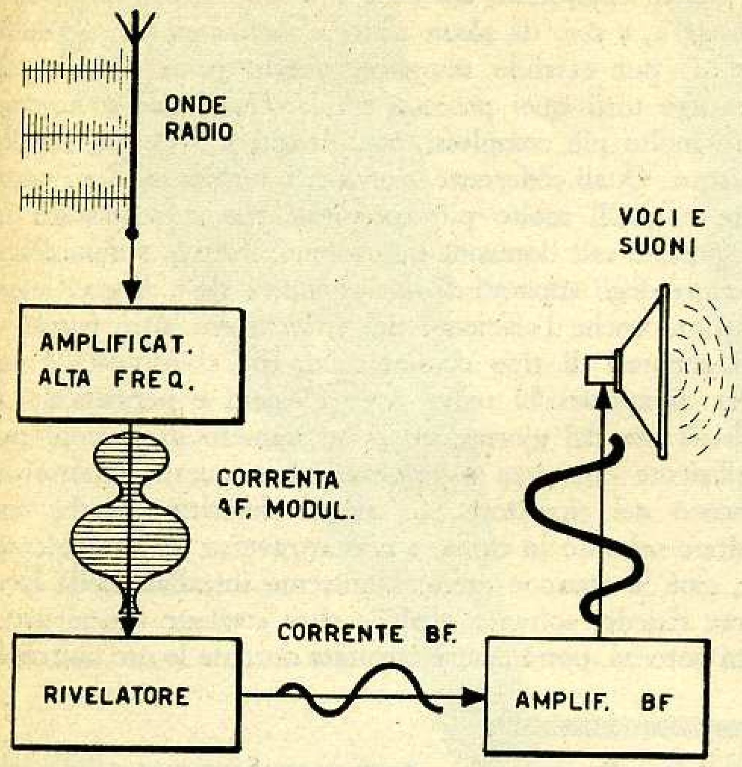

Le onde radio, quando arrivano all'antenna ricevente, sono molto deboli a causa del lungo viaggio da esse affrontato. Il ricevitore radio provvede subito a rinforzarle (amplificarle). Successivamente viene operata la separazione fra la corrente oscillante e quella microfonica; soltanto quest'ultima viene inviata, dopo un ulteriore processo di rinforzo, all'altoparlante, che provvede a trasformarla in voci e suoni.

diodo al germanio; il quarto simbolo rappresenta un componente radioelettrico il cui nome è oggi sulla bocca di tutti: il transistore; il quinto simbolo rappresenta la cuffia, cioè quell'apparato elettroacustico, che viene tenuto contro le orecchie e sostenuto con il capo; questo componente permette di ascoltare il suono; il sesto e ultimo simbolo vuol rappresentare una pila, e la dicitura « 9 V. » vuol significare che la tensione della pila è appunto di 9 volt.

Taluni dei componenti radioelettrici fin qui elencati sono conosciuti anche dai profani di radiotecnica. Chi, infatti, non ha mai visto una cuffia, una pila da 9 volt, di quelle montate nei ricevitori a transistor portatili? Ma anche la conoscenza più intima degli altri componenti diverrà presto familiare, perchè basterà osservare il disegno del componente rappresentato nella sua realtà, confrontarlo con quello acquistato in commercio e ritenerne ben fisso nella mente il simbolo, per non dimenticare mai più, nella vita, che cosa siano una bobina di induttanza, un diodo al germanio, un transistore, un condensatore variabile.

Tutti i componenti radioelettrici vengono rappresentati, negli schemi, mediante un simbolo particolare.

IL CIRCUITO RADIO

Il circuito radio, di qualunque tipo esso sia, è un po' come una strada, un po' tortuosa, talvolta un tantino lunga, caratterizzata principalmente da un ingresso e da un'uscita. All'ingresso entrano i segnali radio, invisibili e sempre presenti nello spazio che ci circonda; all'uscita, che può essere costituita da un'altoparlante o da una cuffia, gli stessi segnali radio escono sotto forma di voci e di suoni.

Ma nel tipo di ricevitore più semplice e più avanti descritto, questa « strada » non è affatto complessa, tortuosa e lunga; al contrario, essa è semplice e « lineare », e non dà alcun adito a confusioni o a difficoltà di interpretazione. E pur essendo semplice, questo primo circuito di ricevitore radio svolge tutti quei processi fondamentali che si attuano nei ricevitori radio molto più complessi, con circuiti a 5 e più valvole o a 7 e più transistori. Quali differenze, dunque, intercorrono fra questo semplice ricevitore e quelli molto più complessi che si acquistano in commercio? La risposta a tale domanda può essere intuitiva e immediata.

La veste esteriore degli apparati di tipo commerciale è di gran lunga migliore ed è migliore anche l'efficienza del circuito. In altre parole si vuol dire che gli apparati di tipo commerciale, che si acquistano nei negozi di rivendita di apparecchi radio, sono eleganti e permettono di ascoltare, in qualsiasi ora del giorno, un buon numero di stazioni trasmittenti, con sufficiente chiarezza e potenza; tutte queste caratteristiche, invece, mancano nel ricevitore più avanti descritto, perchè con esso si potrà ascoltare soltanto in cuffia, e non attraverso un altoparlante, l'emittente locale, cioè la stazione radiotrasmittente installata nella località in cui il lettore risiede; soltanto qualche altra stazione trasmittente, dotata di una certa potenza, potrà essere ascoltata durante le ore notturne.

Ogni circuito radio può essere paragonato ad una strada, più o meno lunga, più o meno tortuosa, sempre caratterizzata dalla presenza di una « entrata » e di una « uscita ».

Lo schema elettrico di ogni apparecchio radio prende anche il nome di circuito teorico. Esso è tutto composto di simboli, che permettono di semplificare il disegno e di rendere immediata la conoscenza e l'analisi del circuito. Nello schema qui riportato ogni simbolo risulta accoppiato, a parte, con il disegno reale del componente cui esso si riferisce. Per soli motivi di semplificazione didattica è stato sostituito, nello schema, il vero simbolo della pila con due cerchietti contrassegnati con i segni + e —, che si riferiscono al morsetto positivo e a quello negativo della pila.

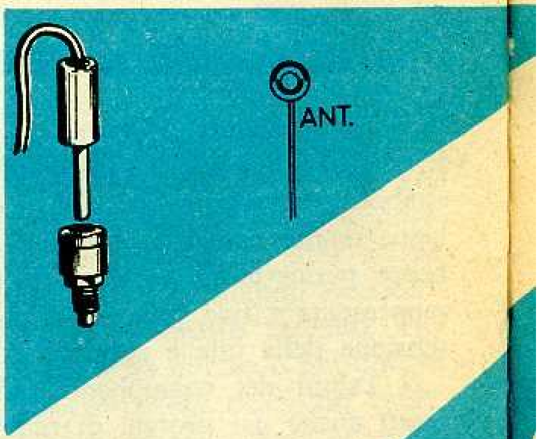

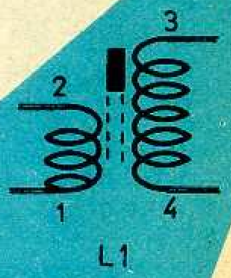

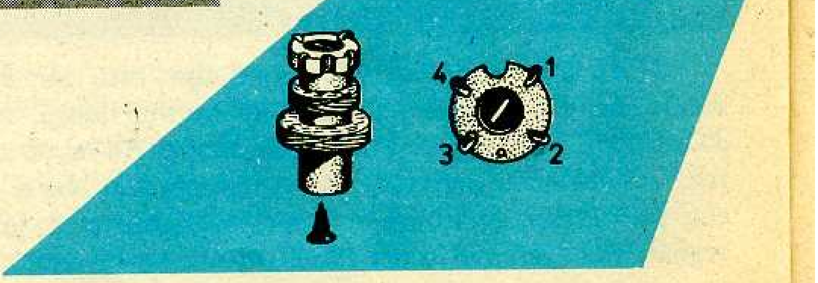

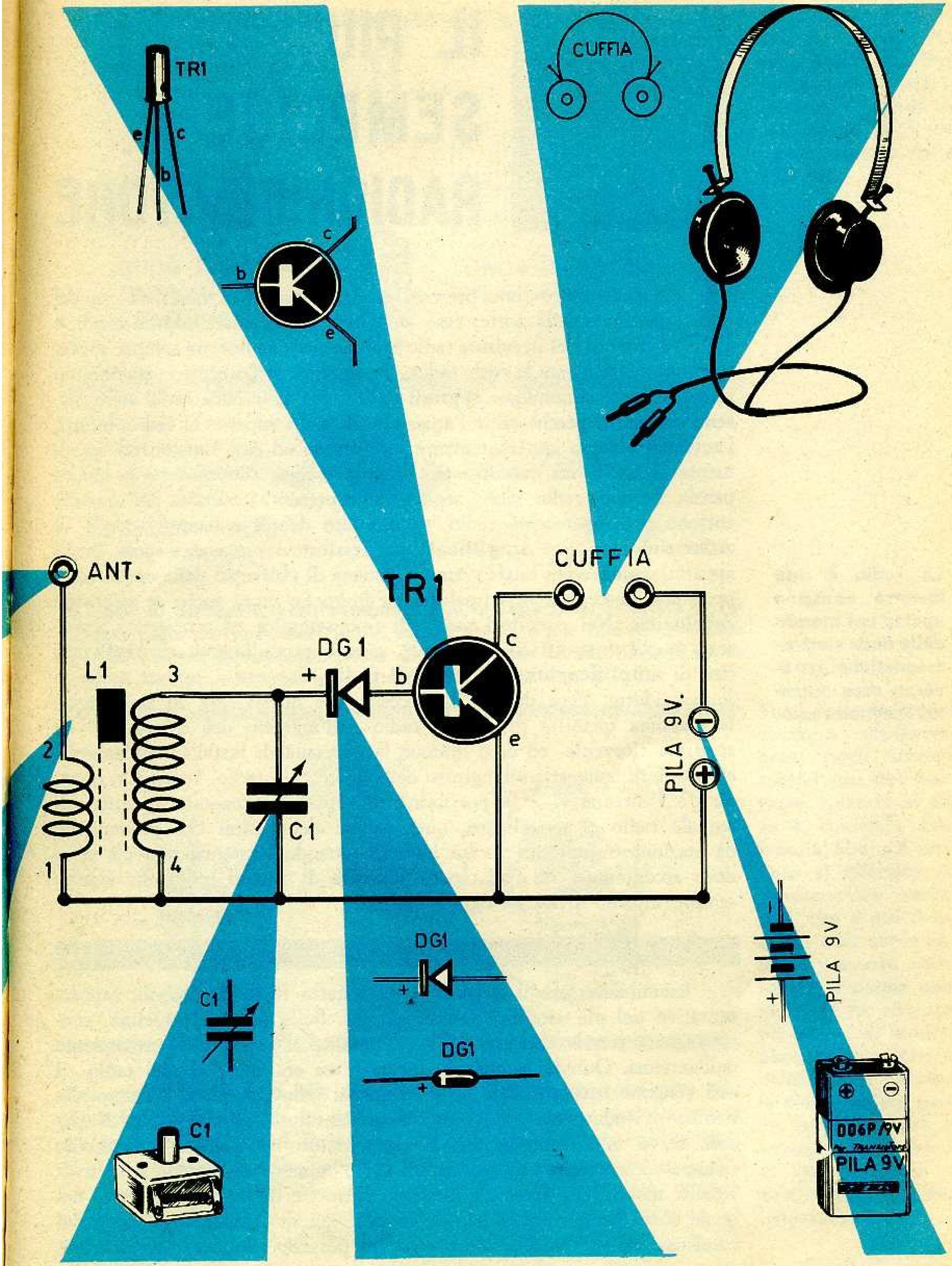
TR1
e
c
b
b
c
e
CUFFIA
ANT.
L1
2
1
3
4
C1
DG1
+
b
TR1
c
e
CUFFIA
PILA 9V.
–
+
DG1
+
DG1
+
C1
C1
–
+
PILA 9V
DD6P/9V
PILA 9V

IL PIU' SEMPLICE RADIORICEVITORE

Le onde radio sono presenti in ogni dove e in qualsiasi ora del giorno e della notte; esse sono nelle nostre case, intorno a noi, e l'antenna del ricevitore radio rappresenta una finestra sempre aperta e pronta a far entrare le onde radio, che in gergo radiotecnico vengono più semplicemente denominate **segnali radio**. Ma se le onde radio sono presenti dovunque perchè serve l'antenna? In molti apparecchi radio, infatti, l'antenna, almeno apparentemente, non esiste ed essi funzionano ugualmente bene. È un concetto questo che si può interpretare in poche parole. Le onde radio, cioè i segnali radio presenti intorno a noi, quando entrano nell'apparecchio radio, sono molto deboli e hanno bisogno di essere rinforzati, cioè **amplificati**, per trasformarsi in voci e suoni. Negli apparati commerciali esiste tutto un sistema di **rinforzo** delle onde radio, per cui si riesce sempre a trasformarle in voci e suoni anche se esse sono debolissime. Nei ricevitori radio, di tipo semplice ed economico, come sono quelli costruiti dai dilettanti, questo procedimento di **rinforzo**, cioè di **amplificazione dei segnali radio**, non esiste, oppure esiste in forma ridotta. Ecco dunque la necessità, in questi casi, di far entrare la massima quantità di segnali radio nell'ingresso del circuito di un apparato ricevente, ed ecco dunque la necessità di installare un'antenna esterna e di collegarla all'ingresso dell'apparecchio radio. Insomma, quando c'è l'antenna vi è la possibilità di captare la massima quantità di segnale radio e, soprattutto, quei segnali debolissimi che provengono da stazioni trasmittenti molto lontane; quando l'antenna non c'è ci si deve accontentare di quella poca quantità di segnali radio che stanno sempre intorno a noi, in ogni luogo.

La radio è una finestra sempre aperta nel mondo delle onde elettromagnetiche: attraverso essa entrano i segnali radio.

ESAME DEL CIRCUITO TEORICO

Esaminiamo ora il circuito teorico, fatto tutto di simboli, rappresentativo del più semplice radioricevitore. Sulla presa di antenna, contrassegnata con la dicitura ANT., si applica il conduttore proveniente dall'antenna. Quindi, attraverso questa presa entrano i segnali radio ed essi vengono **intrappolati** nella bobina di induttanza L 1. Accanto alla bobina di induttanza è presente il condensatore variabile C 1. A che cosa serve tale componente? Il condensatore variabile è un piccolo sistema meccanico composto da un certo numero di lamelle fisse e di lamelle mobili; un perno di comando permette di far ruotare l'asse nel quale sono incorporate le lamelle mobili; quindi azionando il perno del condensatore variabile, le lamelle mobili possono assumere una determi-

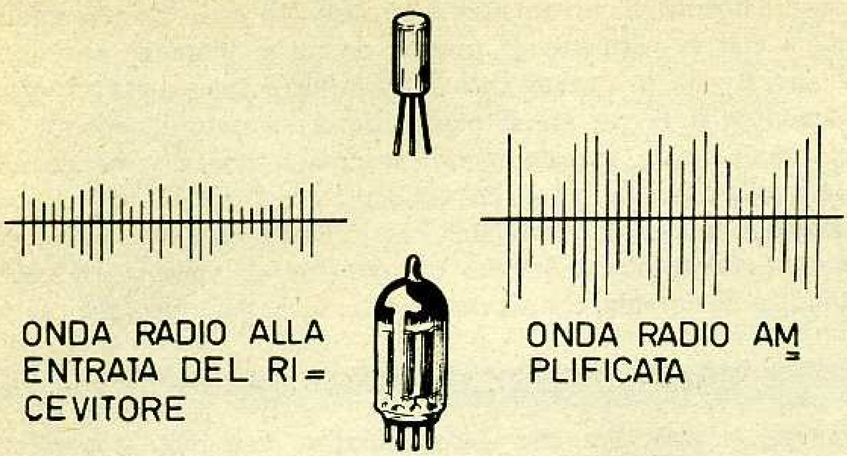

Il transistore e la valvola elettronica rappresentano due componenti radioelettrici in grado di rinforzare le onde medie captate dalla antenna.

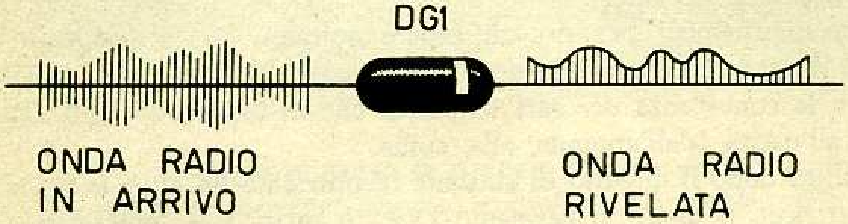

Il diodo al germanio, contrassegnato con la sigla DG1, si lascia attraversare dalla corrente elettrica in un sol senso, e per tale motivo esso prende il nome di semiconduttore. Esso provvede ad eliminare le semionde di uno stesso nome dei segnali radio in arrivo.

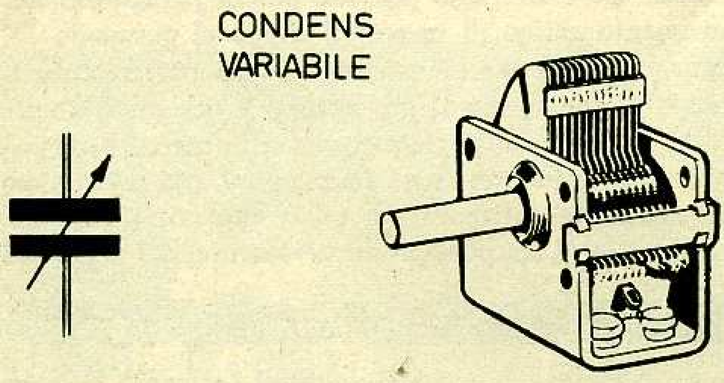

Tutti i condensatori vengono contrassegnati negli schemi teorici con due barrette; quando in esse è riportata una freccia, allora il simbolo si riferisce al condensatore variabile, che è qui disegnato così come esso si presenta nella realtà. Il perno permette di far ruotare un blocco di lamine parallele fra loro e della stessa forma; la rotazione del perno fa allontanare o avvicinare il blocco di lamine mobili a quello composto con lamine fisse.

nata posizione rispetto alle lamelle fisse. Questa posizione crea una condizione radioelettrica del condensatore variabile per la quale non tutti i segnali radio presenti intorno a noi possono circuitare nel sistema elettrico composto dalla bobina L1 e dal condensatore variabile C1. Le condizioni radioelettriche, create dal condensatore variabile, permettono, principalmente, la presenza di un solo segnale radio all'ingresso del ricevitore. L'insieme della bobina L1 e del condensatore variabile C1 prende il nome di **circuito di sintonia**. Vi siete mai chiesti che cosa avviene internamente al vostro ricevitore radio di casa quando azionate la manopola che vi permette di passare da un programma radiofonico ad un altro? Anche in questo caso la risposta è immediata. Ruotando quella manopola si fa ruotare il perno del condensatore variabile, cioè si fa in modo che le lamelle mobili assumano una diversa posizione rispetto a quelle fisse. Al perno del condensatore variabile, nei ricevitori di tipo commerciale, è applicata una particolare meccanica che trascina, contemporaneamente al movimento di rotazione del perno del condensatore, l'indice di sintonia, che scorre lungo la scala parlante del ricevitore.

SELEZIONE DEI SEGNALI RADIO

Dunque, si può dire che il condensatore variabile rappresenti la chiave in grado di aprire molte porte, e in grado di far entrare nel ricevitore radio il segnale preferito, quello della stazione trasmittente che si desidera ricevere.

Accontentiamoci per ora di sapere soltanto ciò a proposito del circuito di sintonia e procediamo con l'esame del circuito del ricevitore, cioè con la conoscenza dei vari fenomeni che in esso si attuano, dall'ingresso all'uscita, dall'antenna alla cuffia.

Il condensatore variabile rappresenta, in pratica, il componente radioelettrico che permette di selezionare tra i molti segnali radio presenti quello che si vuol ascoltare.

Subito dopo il circuito di sintonia incontriamo un altro componente radioelettrico: il diodo al germanio DG; in realtà tale componente può essere costruito con aspetti esteriori diversi, ma la funzione è sempre la stessa: esso impedisce il passaggio a quella parte delle onde radio che serve soltanto a trasportare i segnali rappresentativi delle voci e dei suoni, lasciando passare solo questi ultimi. Sì, perchè le onde radio sono costituite da una mescolanza di segnali: quelli che fungono soltanto da elementi trasportatori e quelli che, come è stato detto, rappresentano le voci e i suoni. Le onde radio possono quindi paragonarsi ad un autocarro in movimento carico di merce; il diodo al germanio DG rappresenta un segnale di stop per l'autocarro e lascia passare invece la merce in esso contenuta. Impareremo, più avanti, a designare l'autocarro con il termine di « segnali di alta frequenza » e la merce in esso contenuta con il termine di « segnali di bassa frequenza ». Ma per ora, anche senza conoscere il significato intrinseco di tali espressioni, possiamo far uso di esse per appropriarci, sia pure progressivamente, del gergo radiotecnico.

GLI STADI

E intanto possiamo dire che il diodo al germanio DG separa il circuito in esame in due parti importanti, che prendono il nome di **stadi**. Lo stadio che precede il diodo al germanio prende il nome di « stadio di alta frequenza », lo stadio che succede al diodo (e comprende lo stesso diodo) viene designato col nome di « stadio di bassa frequenza ». Ma i segnali di bassa frequenza, che risultano presenti al di là del diodo, sono troppo deboli per essere trasformati in suono: essi devono essere **rinforzati**, cioè **amplificati.** A tale processo di rinforzo dei segnali provvede un altro componente radioelettrico, quello contrassegnato con la sigla TR 1, che è rappresentativo del moderno transistore. Internamente

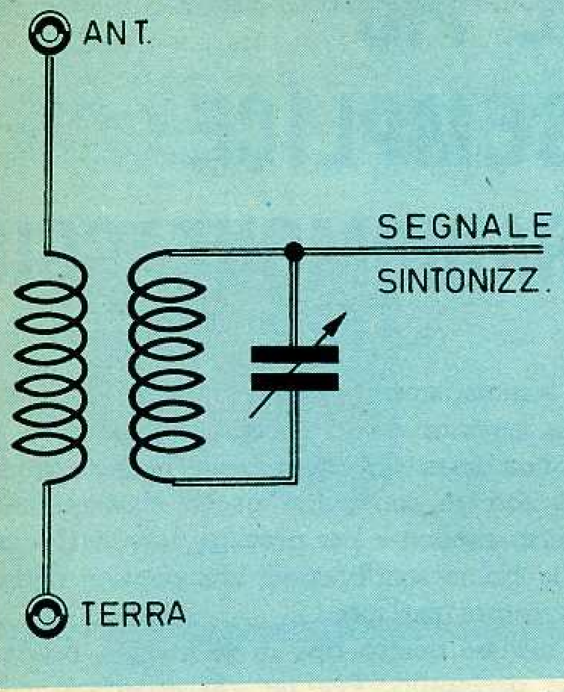

Il circuito di entrata del ricevitore radio è rappresentato da un avvolgimento di fili, che prende il nome di bobina di induttanza. Ad essa sono collegati i circuiti di antenna e di terra. La bobina di induttanza è accoppiata ad una seconda bobina che, unitamente al condensatore variabile, compone il circuito di sintonia del ricevitore.

al transistore si svolge il processo di rinforzo dei segnali radio, che vengono successivamente prelevati da esso e applicati alla cuffia, nella quale avviene la vera e propria trasformazione dei segnali radioelettrici in segnali acustici, cioè in voci e suoni. Il lettore si chiederà a questo punto per quale motivo nel circuito è inserita una pila da 9 volt, e ciò è presto detto: il transistore TR 1, per poter adempiere al compito di rinforzo dei segnali radio, deve essere percorso da una piccola quantità di corrente erogata da una pila, ma anche su tale argomento avremo modo di intrattenerci più avanti.

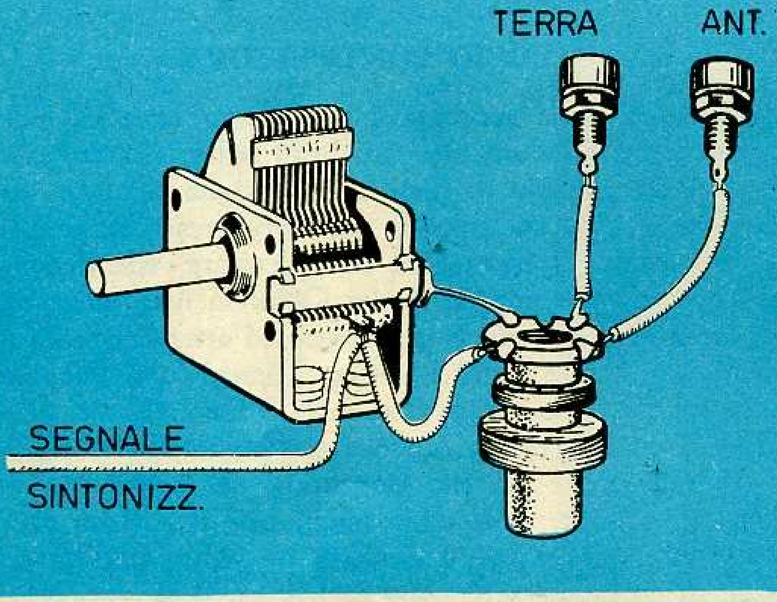

Così si presenta nella realtà il circuito antenna-terra del ricevitore e quello di sintonia. I due avvolgimenti di fili risultano effettuati in un unico supporto di forma cilindrica; i quattro terminali dei due avvolgimenti risultano legati a quattro ancoraggi riportati nella parte superiore del supporto: su di essi si effettuano le saldature a stagno dei fili conduttori collegati alle boccole, al terminale utile del condensatore variabile e alla sua carcassa (collegamento di massa).

IL PIU' SEMPLICE RADIOMONTAGGIO

Dopo aver descritto sommariamente il ricevitore e dopo aver interpretato in forma generica i vari processi radioelettrici che in esso si svolgono, non resta ora che passare immediatamente sul piano della pratica, per toccare con mano quegli elementi finora rappresentati soltanto attraverso simboli e per provare, cosa assai importante, le prime emozioni e le prime soddisfazioni che possono derivare dalla costruzione di una... creatura parlante.

Il materiale necessario per la costruzione.

È ovvio che per poter iniziare questo tipo di montaggio, occorrerà prima munirsi del materiale necessario. Per far ciò occorrerà recarsi presso un negozio specializzato nella rivendita di materiali radioelettrici al dettaglio e chiedere il seguente materiale: una bobina tipo Corbetta CS3/BE, un condensatore variabile, isolato a mica o a aria, del valore capacitivo compreso fra i 300 e i 500 picofarad, un diodo al germanio di qualuque tipo, un transistore di tipo OC 75 (o corrispondente), una pila da 9 volt, una cuffia da 500 o 1000 ohm di impedenza, due boccole, due spinotti, una manopola adatta per il perno del condensatore variabile che si è acquistato. Occorrono ancora: un saldatore, un po' di stagno in tubetto e un po' di filo ricoperto in plastica, adatto per collegamenti radio; anche questa seconda parte del materiale può essere acquistata presso lo stesso negoziante di materiali radioelettrici. All'atto dell'acquisto del saldatore, è consigliabile orientarsi sui tipi dotati di punta di media grandezza e della potenza elettrica aggirantesi intorno ai 50 watt.

L'ASTUCCIO CONTENITORE

Il primo elemento da approntarsi per questo semplice montaggio è rappresentato dal contenitore del ricevitore radio; per contenitore si intende un qualsiasi astuccio di plastica, di cartone rigido o di legno, di dimensioni tascabili e sul tipo di quello rappresentato nel disegno raffigurativo del piano di montaggio del ricevitore, che denominiamo schema pratico. Quel che importa è che l'astuccio non venga costruito in metallo, perchè le onde radio non riescono ad attraversare gli ostacoli metallici. E qui il lettore potrà obiettare: perchè non può essere di metallo il contenitore, dato che in esso entra il filo collegato all'antenna e per mezzo del quale vengono convogliati i segnali radio nel circuito di sintonia? L'obiezione è più che giustificata. Ma occorre tener conto che il sistema di rinforzo delle onde radio, cioè il sistema di amplificazione, è alquanto rudimentale e modesto in questo ricevitore, e non ci si deve quindi accontentare dei soli segnali radio provenienti dall'an-

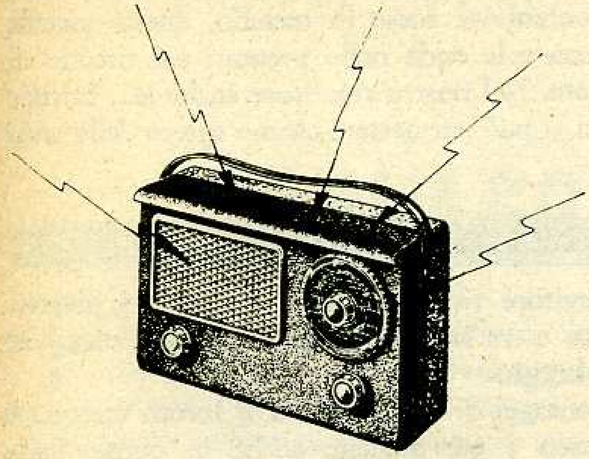

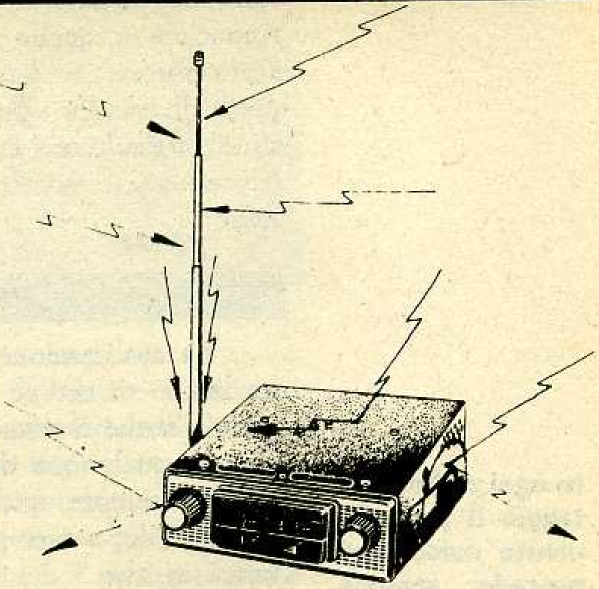

Le onde radio sono presenti dovunque, anche attorno e dentro il ricevitore. Nei moderni apparati, con circuito transistorizzato, la ricezione può avvenire senza l'impiego di un'antenna vera e propria esterna al ricevitore (disegno in alto a sinistra); nei ricevitori muniti di custodia metallica l'uso dell'antenna è necessario (disegno in alto a destra).

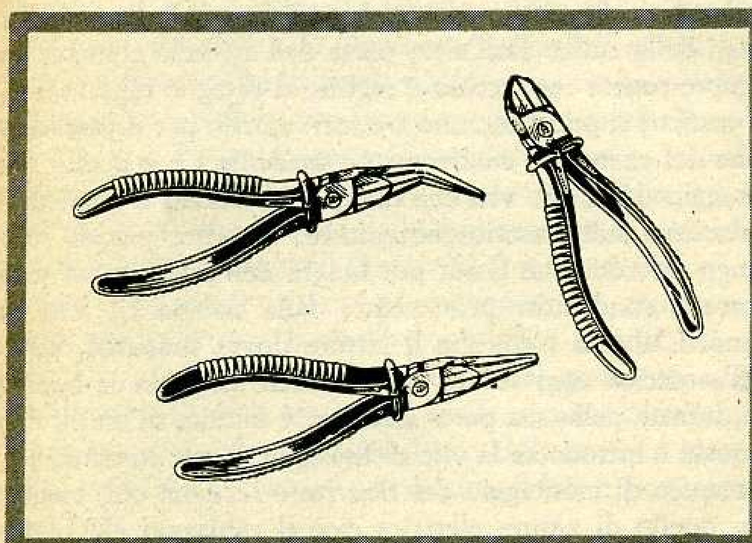

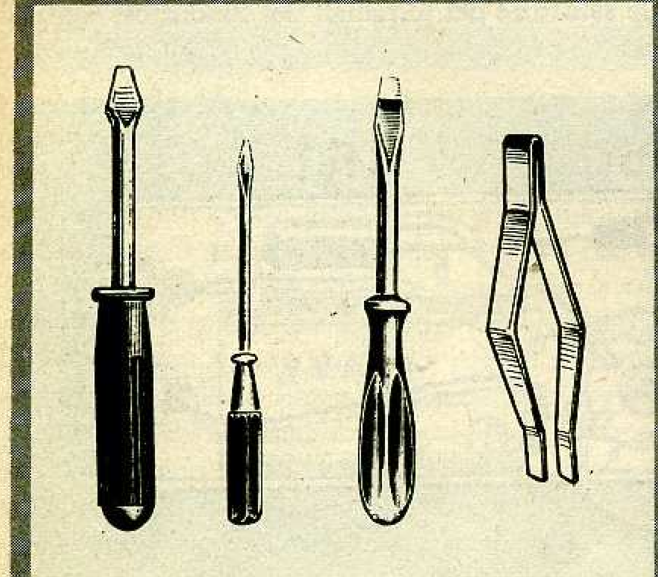

Gli attrezzi necessari per il montaggio di un semplice ricevitore radio possono essere ridotti a pochi elementi: due pinze, un tronchesino, alcuni cacciaviti, un pinzettino e un trapano a mano.

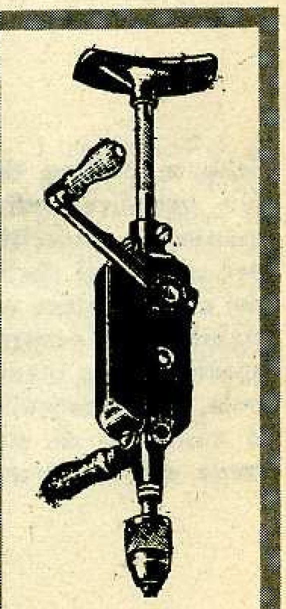

tenna, ma occorre anche utilizzare quei segnali che si trovano inevitabilmente intorno al circuito di sintonia; in altre parole non è possibile rinunciare a quella piccola parte di energia presente nell'aria intorno al ricevitore. Se l'astuccio contenitore fosse in metallo, questa piccola parte di energia sarebbe esclusa e le onde radio presenti nel circuito di sintonia risulterebbero attenuate. Nel nostro ricevitore anche le... briciole devono essere raccolte e non ci si può permettere alcuno spreco delle onde radio!

PROCEDIMENTO MECCANICO

La realizzazione del ricevitore richiede due tipi di lavori diversi: un lavoro di ordine meccanico e un lavoro di natura radioelettrica, che prende anche il nome di **cablaggio.**

In ogni radiomontaggio il procedimento meccanico precede sempre quello elettrico.

In qualunque tipo di montaggi di radioapparati il lavoro meccanico precede sempre quello elettrico e ciò avviene anche in questo caso. Dunque, dopo aver preparato l'astuccio contenitore, si provvederà a praticare in esso i due fori in cui verranno avvitate le due boccole per la presa di cuffia, le due boccole cioè in cui verranno innestati i due spinotti collegati ai terminali della cuffia. Dall'altra parte dell'astuccio contenitore (il lettore deve sempre tenere sott'occhio e seguire il disegno rappresentativo dello schema pratico) si praticheranno tre fori: quello per il passaggio e la libera rotazione del perno del condensatore variabile C1 e quelli, più piccoli, per il passaggio delle due viti che matengono fermo il condensatore stesso e lo bloccano sull'astuccio contenitore. Un altro piccolo foro verrà praticato lungo uno dei due bordi più lunghi dell'astuccio, ed esso servirà a far passare il conduttore proveniente dalla bobina L1, che va collegato all'antenna. L'ultimo foro, che il lettore dovrà praticare, apparirà sul fondo dell'astuccio: esso servirà a mantenere bloccata la bobina L1; questa bobina, infatti, nella sua parte inferiore è munita di un piccolo foro filettato, nel quale è introdotta la vite di fissaggio di tale componente.

Il lavoro meccanico di montaggio del ricevitore termina qui, mentre può essere iniziato quello di natura elettrica cioè il cablaggio che, come è stato detto, consiste nella saldatura dei terminali dei conduttori elettrici e di quelli dei componenti.

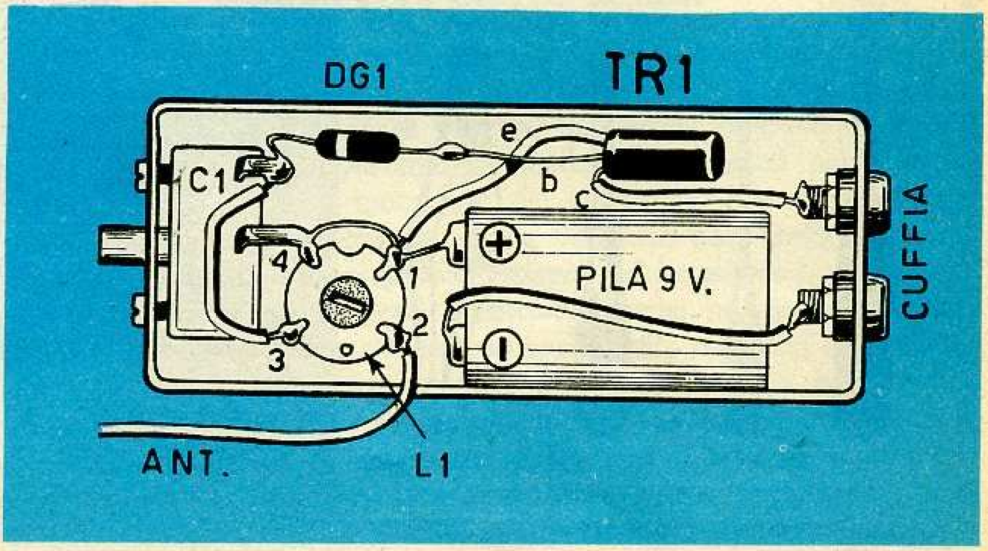

Schema pratico del più semplice radiomontaggio descritto nel testo. Sul perno del condensatore variabile C1 occorre applicare una manopola, che costituirà il comando di sintonia del ricevitore.

CABLAGGIO

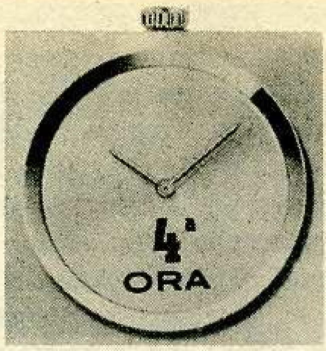

Prima di por mano al saldatore, cioè prima di iniziare a saldare i componenti, il lettore dovrà preparare i conduttori, nelle misure adatte. Per esempio si comincerà a ritagliare uno spezzone di filo di lunghezza sufficiente a congiungere una delle due boccole di cuffia con il morsetto negativo della pila da 9 volt; nelle opportune misure si prepareranno anche gli spezzoni di filo che congiungono i terminali della bobina L1 con il morsetto positivo della pila e con il terminale del condensatore variabile C1. Dopo aver preparati questi spezzoni di filo (conviene utilizzare filo di rame rigido ricoperto in plastica), il lettore dovrà provvedere a pulirne accuratamente le estremità. Tutti i conduttori metallici, infatti, vanno soggetti ad ossidazioni ed è quindi necessario provvedere alla loro pulizia. Questo lavoro di pulizia dei terminali si esegue servendosi di una lametta da barba; con essa si raschia accuratamente ogni estremità dei conduttori, fino a che il rame appare nel suo brillante colore naturale. E questo stesso procedimento va applicato anche ai morsetti della pila, alle due linguelle (terminali) uscenti dal condensatore variabile C1, ai due terminali del diodo al germanio e ai tre terminali del transistore. Questo lavoro di raschiamento, comunque, va eseguito soltanto sull'estremità del conduttore. Soltanto ora tutti i componenti possono considerarsi pronti per essere saldati e non resta che por mano al saldatore.

Il termine cablaggio deriva dal francese « cable » e sta ad indicare le operazioni di collegamento dei fili conduttori e dei componenti.

LA SALDATURA

La saldatura rappresenta una « giuntura » tra le estremità di due o più conduttori metallici; essa deve garantire la continuità dei conduttori e deve essere eseguita osservando alcune norme fondamentali. La saldatura può essere esatta e perfetta, ma può essere anche errata. Nel primo caso il gergo radiotecnico si serve dell'espressione « saldatura calda », nel secondo caso si usa l'espressione « saldatura fredda ». La saldatura calda è quella che garantisce una perfetta continuità di conduzione elettrica: essa appare lucida e a superficie curva. La saldatura fredda può presentare queste stesse caratteristiche ma, sottoposta ad un leggero sforzo di trazione esercitato con le dita, provoca il distacco dei conduttori e dello stagno. Per i primi tipi di saldature, dunque, il dilettante farà sempre bene ad accertarsi di avere eseguito una saldatura calda, semplicemente esercitando una trazione sui conduttori con le dita della mano; in altre parole si tirano i conduttori in verso opposto a quello in cui è stata eseguita la saldatura.

I radiotecnici distinguono le saldature buone da quelle cattive con le espressioni di: « saldatura calda » e « saldatura fredda ».

Anche il saldatore va soggetto ad ossidazione, cioè la sua punta di rame saldante si ricopre di un velo scuro, che è cattivo conduttore del calore, e questa è una delle cause che concorrono assai spesso alla esecuzione di saldature fredde. La punta di rame del saldatore, quindi, deve essere pulita di quando in quando, servendosi di uno spazzolino metallico e non della lama di un temperino o, peggio ancora, della lima, come purtroppo fanno alcuni radiotecnici; con questi sistemi, infatti, si riduce sempre più la massa del rame e la punta saldante si assottiglia.

Fatte queste premesse, non resta che prendere il saldatore in mano e inziare con la prima saldatura, per esempio quella del conduttore che unisce il morsetto negativo della pila con una delle due boccole della presa di cuffia. Il terminale del conduttore va tenuto stretto fra i becchi di una pinza e appoggiato sull'estremità filettata della boccola. Con la mano destra si impugna il saldatore e si appoggia la punta sopra il terminale del conduttore, sciogliendo una certa quantità di stagno; lo stagno si scioglie soltanto se posto in contatto con la punta del saldatore. Occorrerà sempre ricordarsi di non accontentarsi della liquefazione dello stagno e non togliere la punta del saldatore dalla saldatura; è bene che la punta del saldatore rimanga ferma sul punto in cui si effettua la saldatura per alcuni secondi, di modo che il calore possa distribuirsi uniformemente e nella massima quantità sulle parti metalliche. Quando si è convinti di aver eseguito la saldatura, si toglie il saldatore e si attende per qualche istante per dar tempo allo stagno di rapprendersi; quindi si esercita una leggera trazione sul conduttore, muovendolo un po' da una parte e un po' dall'altra, in modo da accertarsi del suo completo fissaggio sulla boccola e della buona qualità della saldatura ottenuta. In ogni caso la pratica sarà sempre la migliore maestra per tutti e le operazioni di saldatura diverranno sempre più semplici e istintive col passare del tempo; la regola fondamentale rimarrà sempre la stessa: pulizia perfetta dei terminali da saldare ed esecuzione relativamente lenta della saldatura.

I compiti del saldatore sono due: diffondere il calore sulle parti da saldare e sciogliere lo stagno.

I TERMINALI DEL TRANSISTORE

Dentro il transistore vi sono tre pezzettini di cristallo di germanio; ad ognuno di questi tre pezzettini di cristallo è collegato un conduttore sottile, lungo, che fuoriesce dall'involucro di questo componente per costituire i terminali del transistore. Ma la natura dei cristalli di germanio, contenuti internamente al transistore, non è la stessa per tutti e tre i cristalli; esistono delle differenze nella natura intrinseca del cristallo stesso, ed ognuno di essi prende un nome preciso. Ma lasciamo da parte quel che è contenuto dentro il transistore ed occupiamoci soltanto dei suoi terminali, che rappresentano gli elementi reali con cui il radiotecnico ha a che fare quando lavora. È ovvio che il nome assunto dai tre cristalli di germanio contenuti dentro il transistore si estende anche ai terminali del componente. Questi nomi sono: EMITTORE-BASE-COLLETTORE.

I tre conduttori, che costituiscono i terminali del transistore, sono perfettamente uguali, pur tuttavia l'allievo radiotecnico deve imparare a riconoscerli e a distiguerli con la più assoluta precisione. Purtroppo

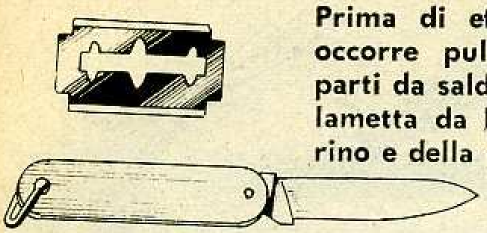

Prima di effettuare la saldatura occorre pulire accuratamente le parti da saldare, servendosi di una lametta da barba o di un temperino e della pasta-salda.

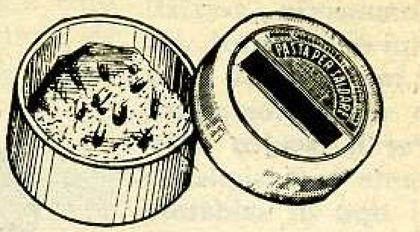

Lo stagno in tubetto, contenente internamente la pasta salda, viene venduto in commercio in rocchetti (disegno in basso a sinistra). Prima di effettuare la saldatura occorre raschiare energicamente il terminale del conduttore fino a mettere in evidenza la lucentezza del metallo.

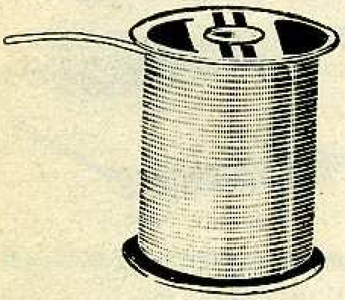

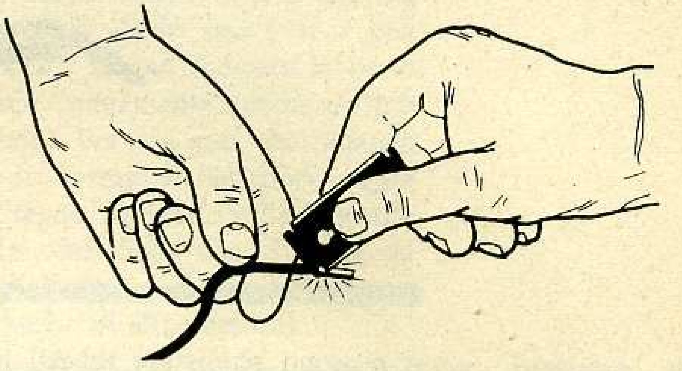

non tutti i transistori sono uguali esteriormente e il metodo di riconoscimento dei terminali varia spesso fra un tipo di transistore e un altro.

Il transistore montato nel nostro semplice ricevitore appartiene ad una grande categoria di transistori, per i quali il metodo di « lettura » dei terminali è sempre lo stesso. Sull'involucro esterno di questo tipo di transistori, in prossimità della base dell'involucro cilindrico, è impresso un puntino colorato, che può essere rosso, verde, blu, ecc. Ebbene, il terminale di COLLETTORE è sempre quello che si trova da quella parte in cui è presente il puntino colorato; il terminale di BASE si trova in posizione centrale, mentre quello di EMITTORE si trova alla estremità opposta. Questi tre terminali possono essere distribuiti lungo una stessa linea oppure lungo una SEMICIRCONFERENZA. Ma tale distribuzione dei terminali del transistore non disturba affatto il metodo di riconoscimento, purché risulti ben visibile il puntino colorato che sta sempre ad indicare il terminale di COLLETTORE.

Nello schema pratico del ricevitore questi tre terminali sono stati indicati con le lettere minuscole **e** (emittore), **b** (base), **c** (collettore). Il lettore pignolo potrà osservare che nello schema pratico del ricevitore il terminale di base (b) è « nudo », cioè così come esso si presenta all'atto dell'acquisto del transistore stesso; i terminali di emittore (e) e di collettore (c) sono invece disegnati con linee più grosse. Qual'è il motivo di tale differenza? I conduttori di emittore e di collettore sono stati rivestiti con due pezzetti di tubetto di plastica flessibile, allo scopo di isolarli elettricamente e cioè in modo che essi non si possano toccare tra loro in alcun modo. Presso i negozianti di materiali radioelettrici

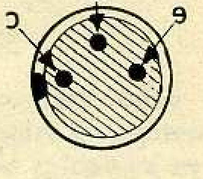

Vista in pianta della parte inferiore di un transistore. Le tre lettere minuscole si riferiscono ai conduttori di emittore (e), base (b), collettore (c).

Esistono oggi in commercio svariati tipi di saldatori, più o meno costosi e più o meno complessi. Per il tecnico dilettante è consigliabile il tipo di saldatore più semplice, quello rappresentato nel disegno in basso.

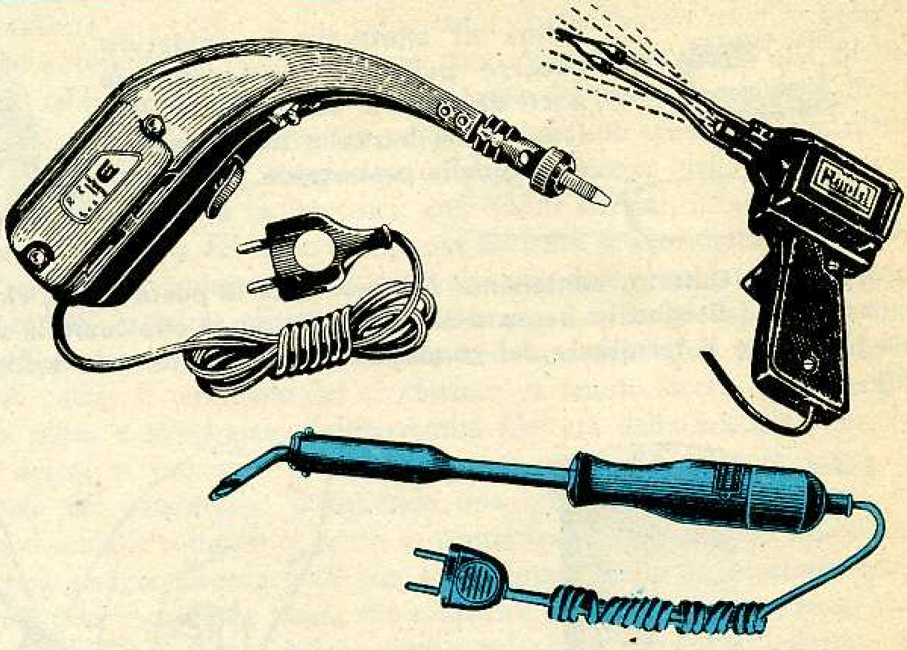

si possono acquistare tubetti isolanti di qualunque tipo e sezione e di colori diversi. Il colore del tubetto isolante costituisce un motivo di grande importanza nei montaggi radioelettrici, perchè con il colore si riesce ad individuare subito un tipo di circuito o un insieme di collegamenti che hanno funzioni analoghe. Impareremo, più avanti, che nei circuiti radio vi è un insieme di conduttori destinati ad essere percorsi dalle correnti ad alta tensione; tutti questi conduttori possono essere rivestiti, ad esempio, con tubetto dello stesso colore, preferibilmente il rosso; i conduttori che non sono percorsi da correnti a tensioni elevate possono essere ricoperti con tubetto di color bianco; in pratica questa differenziazione di colori serve anche a tutelare l'incolumità del radiotecnico, il quale sa che toccando i fili rossi potrebbe prendere una forte scossa elettrica, talvolta pericolosa, mentre toccando i fili bianchi non c'è alcun pericolo.

Il calore guasta il transistore.

Parlando della saldatura si è detto che questa non deve essere mai eseguita affrettatamente, e tale consiglio è valido per qualunque tipo di saldatura dei terminali dei conduttori e dei componenti radioelettrici. Non è valido, invece, per le operazioni di saldatura dei terminali dei transistori. I transistori sono componenti di natura solida e sopportano facilmente gli urti; il transistore, quindi, non si danneggia cadendo per terra o subendo un colpo inavvertitamente impresso con le pinze e con il cacciavite. Esso si danneggia invece quando è sottoposto a una grande quantità di calore e ciò può verificarsi appunto nel momento in cui si eseguono le saldature dei suoi terminali. Dunque, bisogna evitare che il calore eccessivo entri, attraverso il terminale, nell'interno del transistore. E tale accorgimento è rispettato soltanto quando si eseguono saldature rapide per mezzo di un saldatore munito di punta sottile e ben calda. Tale avvertimento, peraltro, non vuole infondere paura nell'allievo radiotecnico ma vuol soltanto ricordare che la tecnica di saldatura dei terminali del transistore si differenzia da quella adottata per ogni altro tipo di saldatura. Del resto, i terminali dei transistori sono costituiti da fili conduttori molto sottili e per ottenere in essi una saldatura « calda » non occorre indugiare a lungo con il saldatore.

LE POLARITA' DEL DIODO

In gergo radiotecnico si usa dire che il diodo al germanio, contrassegnato con la sigla DG 1 negli schemi elettrico e pratico del nostro semplice ricevitore, è un componente polarizzato. Ciò vuol significare, praticamente, che questo componente non può essere inserito nel circuito in un modo qualunque, ma in un preciso verso. Avete notato che nel disegno rappresentativo dello schema pratico il diodo DG 1 reca una fascetta bianca? Ebbene, quella fascetta bianca sta ad indicare la parte in cui si trova il terminale positivo di tale componente (vedremo più avanti il significato preciso di tale espressione). Per ora quel che importa è che l'allievo nell'effettuare le saldature dei terminali del diodo segua scrupolosamente l'indicazione rilevata dal disegno, saldando sulla linguella del condensatore variabile C 1 il terminale del diodo che si trova da quella parte in cui l'involucro esterno del diodo è contrassegnato con la fascetta bianca (tale fascetta può essere anche di altro colore).

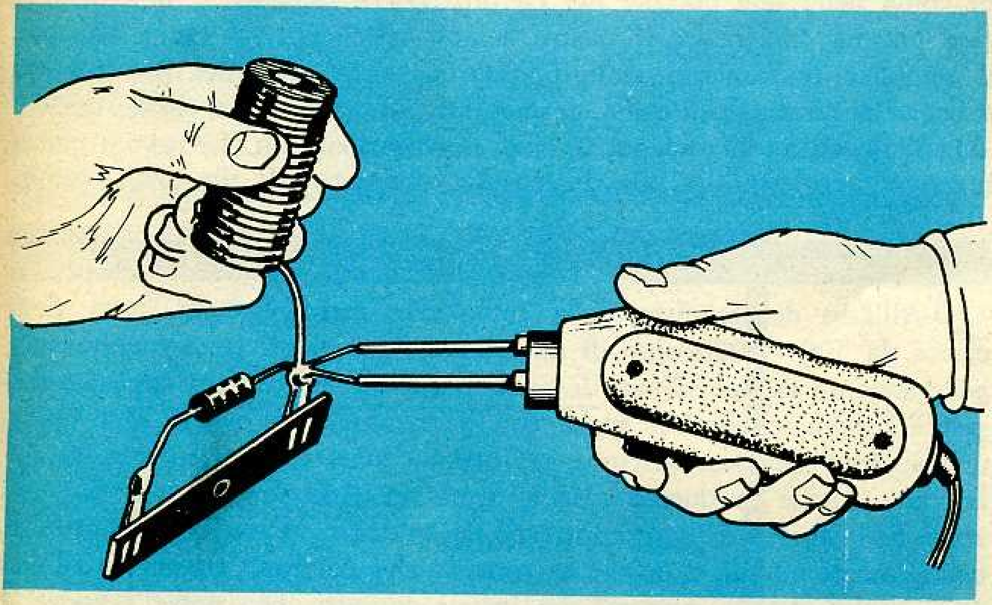

Per ottenere una buona saldatura occorre indugiare qualche istante con la punta del saldatore aderente alle parti da saldare, in modo da permettere al calore di diffondersi uniformemente e in maggiore quantità.

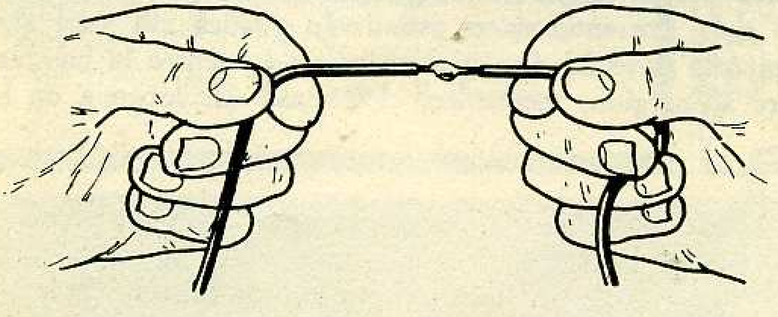

È sempre bene effettuare un controllo della saldatura esercitando una certa forza di trazione con le mani fra i conduttori saldati tra loro.

L' INTERRUTTORE

Ogni apparecchio radio è munito di un interruttore che serve ad accendere e spegnere l'apparecchio stesso. Nell'apparecchio radio di casa nostra esiste una manopola che, con un semplice movimento di rotazione, permette di accendere il ricevitore; molto spesso su questa stessa manopola è applicato anche il comando di volume del ricevitore, cioè il comando che permette di ascoltare la radio più o meno forte.

Nel nostro semplice ricevitore manca un interruttore vero e proprio, però anche in questo caso è stato previsto il sistema di accensione e spegnimento del circuito. Infatti, basta infilare gli spinotti dei terminali di cuffia nelle due boccole del circuito per « accendere » il ricevitore, cioè per far funzionare il nostro semplice apparecchio radio. Dunque, il semplice inserimento nelle boccole del ricevitore dei due spinotti collegati ai terminali di cuffia, permette di accendere il ricevitore; togliendoli il ricevitore si « spegne », cioè non funziona più. In pratica questa manovra permette alla pila di erogare corrente o di rimanere inattiva.

CONTATTI OTTIMI

All'atto dell'acquisto della cuffia che, come è stato detto, deve essere da 500 - 1000 ohm di impedenza, i terminali sono rappresentati generalmente da due fili sottili rigidi. In questi fili occorre applicare due spinotti (sono conosciuti anche con il termine di « banane ») e questi due spinotti devono esercitare un ottimo contatto elettrico con le due boccole. Anche in questo caso si tratta di collegamenti, ma essi non sono fissi come quelli ottenuti con la saldatura a stagno; si tratta di collegamenti mobili, e poiché le correnti che fluiscono attraverso ad essi sono molto deboli, occorre che questi contatti, per poter favorire il passaggio delle correnti, siano ottimi. In pratica ciò vuol significare che lo spinotto deve aderire in maniera rigida dentro la boccola, senza permettere alcun gioco meccanico, cioè senza dar luogo a un contatto lasco.

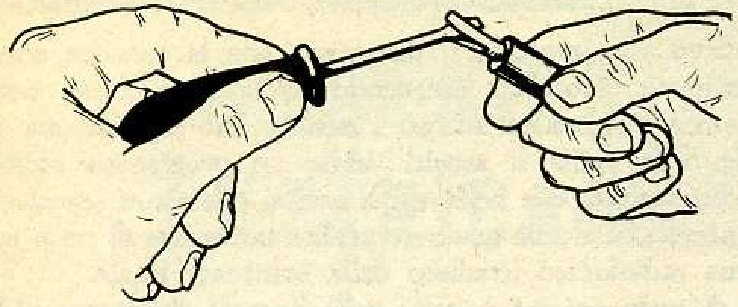

Gli spinotti, chiamati anche « banane », devono sempre esercitare un perfetto contatto elettrico con le boccole nelle quali vengono inseriti. Per ottenere ciò basta introdurre la lama di un cacciavite nel taglio dello spinotto distanziando tra loro leggermente le due parti.

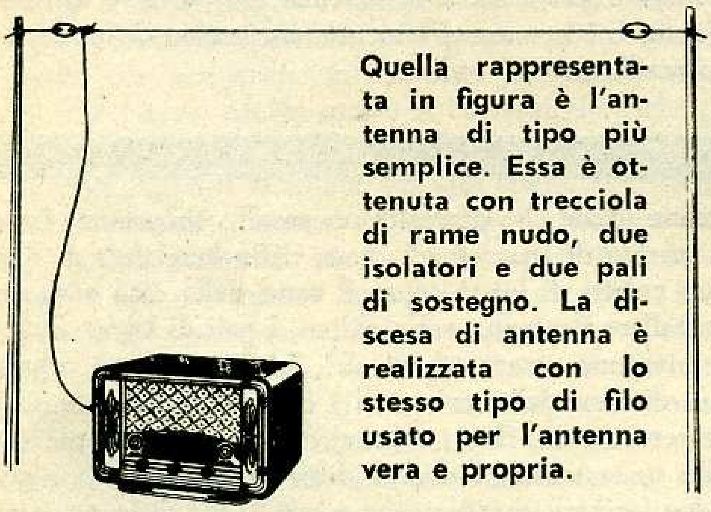

Quella rappresentata in figura è l'antenna di tipo più semplice. Essa è ottenuta con trecciola di rame nudo, due isolatori e due pali di sostegno. La discesa di antenna è realizzata con lo stesso tipo di filo usato per l'antenna vera e propria.

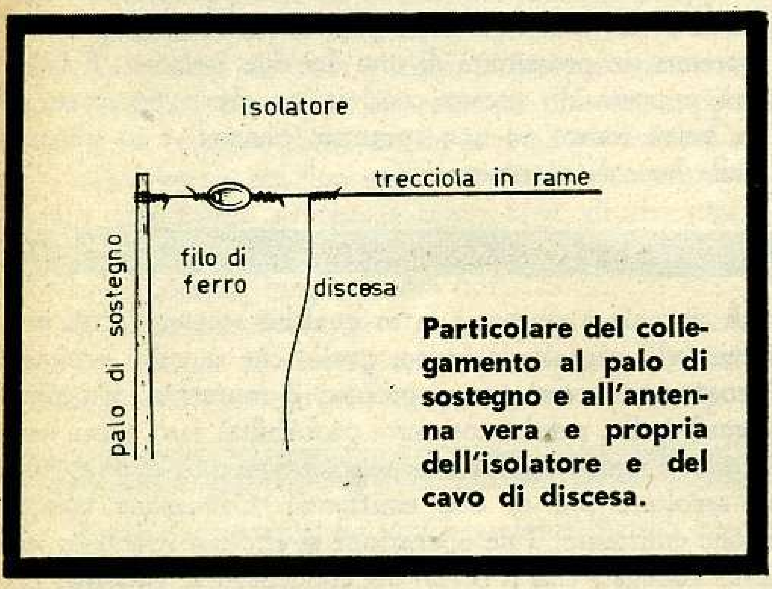

Particolare del collegamento al palo di sostegno e all'antenna vera e propria dell'isolatore e del cavo di discesa.

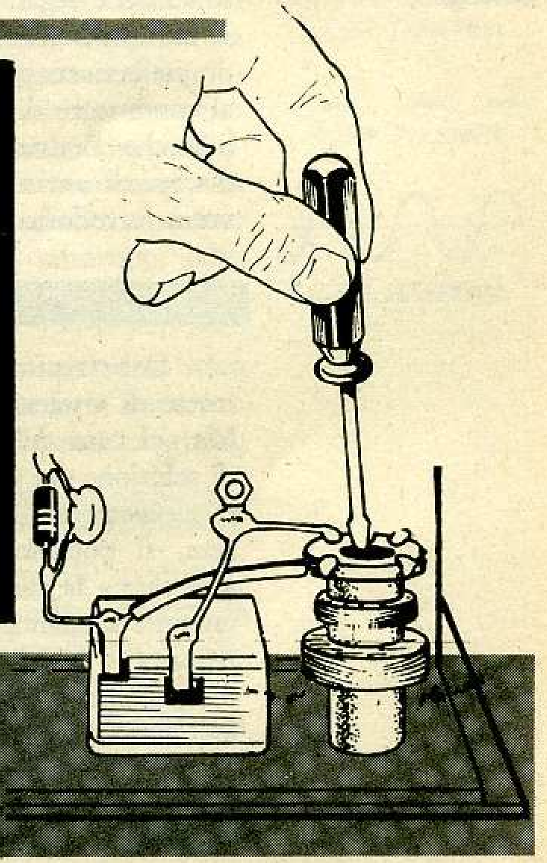

[illegible] bobina di sintonia è avvitato [illegible] parte che, in fase di messa a [illegible] dove essere avvitato o [illegible] per ottenere l'ascolto più forte

L'ASCOLTO

Se l'allievo avrà eseguito il montaggio con la massima attenzione, senza commettere errori ed eseguendo saldature perfette, potrà ora mettere la cuffia in testa ed iniziare l'ascolto. Può capitare, ma ciò sarà difficile, che nella cuffia si ascolti subito un programma radiofonico; ma può anche capitare che nella cuffia non si oda alcun segnale. Infatti si rendono necessarie alcune manovre preliminari prima di poter ascoltare il programma radiofonico irradiato dalla emittente locale.

Prima di tutto occorre inserire nella boccola di antenna (ANT.) lo spinotto collegato alla discesa di antenna, poi si deve agire sul condensatore variabile C 1 e, successivamente, sul nucleo di ferrite della bobina L 1. Ma procediamo con ordine.

COSTRUZIONE DELL'ANTENNA

Trecciola di rame. Isolatori. Pali di sostegno.

L'antenna ideale che permette un ascolto abbastanza forte è rappresentata da un filo di trecciola di rame, della lunghezza di 4 metri circa, tesa fra due paletti di legno sopra il tetto della casa o su un poggiolo. Volendo installare l'antenna nel giardino, i pali di legno, di sostegno dell'antenna, dovranno essere molto alti. L'allievo dovrà acquistare, oltre alla trecciola di rame della sezione di 1 o 2 millimetri, due isolatori. I terminali di antenna vanno fissati a questi due isolatori, semplicemente avvolgendo il filo su se stesso. Sull'altro foro dell'isolatore (gli isolatori sono muniti di due fori) va introdotto un pezzo di filo di ferro, attorcigliato su se stesso e legato, all'estremità opposta, al palo di sostegno dell'antenna. Il filo di discesa è rappresentato da una trecciola di rame dello stesso tipo di quella usata per l'antenna; esso va avvolto per una decina di giri attorno al conduttore di antenna, in prossimità di uno dei due isolatori. È intuibile che l'estremità opposta di questo conduttore, che rappresenta la discesa di antenna, verrà fissato ad uno spinotto (banana) e lo spinotto verrà introdotto nella boccola di antenna.

SINTONIZZAZIONE

Del circuito di sintonia è stato già fatto qualche accenno. Esso permette di « selezionare » il segnale preferito, quello che si vuol ascoltare. Ma nel caso del nostro ricevitore non è proprio il momento di parlare di selezione dei segnali radio, perchè con tutta probabilità esso permetterà di ricevere uno o due segnali radio al massimo (in casi più fortunati, di sera, si potranno ascoltare più di due emittenti). Comunque occorre « cercare » la stazione emittente. Tale operazione si effettua ruotando lentamente la manopola collegata con il perno del condensatore variabile C1. Appena si sentirà qualcosa in cuffia ci si fermerà e, mediante il cacciavite, si provvederà ad avvitare o a svitare quel cilindretto nero, avvitato nell'interno della bobina L1, che prende il nome di « nucleo di ferrite ». Il lettore si accorgerà, dopo una serie di tentativi di avvitamento e di svitamento del nucleo, che esisterà un punto in cui i segnali radio assumeranno la massima intensità sonora.

LA CORRENTE ELETTRICA

Attraverso i conduttori e i componenti di ogni ricevitore radio passano le correnti elettriche. Nel gergo radiotecnico, le correnti elettriche che percorrono un circuito prendono svariati nomi, a seconda della loro natura. Molto spesso le correnti elettriche prendono il nome di « segnali »; ma i segnali sono sempre rappresentati da correnti elettriche. Esistono correnti alternate, correnti continue, correnti pulsanti, correnti ad alta frequenza, correnti a bassa frequenza, correnti di grande intensità e di piccola intensità. Ma in ogni caso la corrente elettrica è rappresentata da un movimento invisibile di particelle atomiche che predono il nome di « elettroni ». Dunque, attraverso i conduttori e i componenti di ogni circuito radio scorre un flusso di elettroni, cioè una corrente elettrica. Se gli elettroni si muovono sempre in un unico verso, allora si dice che la corrente elettrica, cui essi danno vita, è una **corrente elettrica continua;** se gli elettroni si muovono in avanti e all'indietro, con movimento oscillatorio, allora si dice che la corrente da essi determinata è una **corrente alternata.** Se il movimento in avanti e all'indietro degli elettroni è rapidissimo, allora si dice che la corrente da essi prodotta è di tipo alternato ma ad **alta frequenza;** se il movimento in avanti e all'indietro degli elettroni è lento, la corrente generata è di tipo alternato ma a **bassa frequenza.** La corrente pulsante rappresenta una particolarità della corrente continua: in essa gli elettroni si muovono sempre in uno stesso verso, ma non uniformemente, cioè con una velocità costante; il loro movimento avviene a scatti, cioè gli elettroni si muovono, subiscono un arresto per riprendere poi nuovamente il loro viaggio, sempre nello stesso verso, lungo i conduttori.

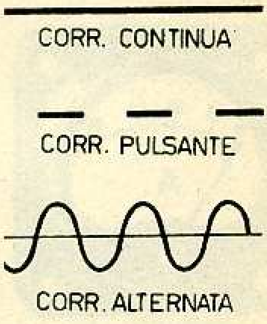

Quando gli elettroni, che compongono la corrente elettrica, sono molto numerosi, si dice di essere in presenza di una corrente intensa o forte; quando gli elettroni sono pochi, la corrente è poco intensa o debole.

CORRENTI DI ALTA FREQUENZA

Tutte le correnti fin qui esaminate hanno bisogno di conduttori metallici per potersi trasferire da un punto ad un altro; fa eccezione la sola corrente ad alta frequenza, che tende a sfuggire dai conduttori metallici e a « viaggiare » anche attraverso l'aria e lo spazio vuoto. Queste correnti, non sono più costituite da elettroni, quando si trovano fuori dai conduttori metallici, ma sono rappresentate da campi elettromagnetici.

Tutti voi certamente avrete avuto in mano una calamita, e avrete avuto modo di constatare che questa attira i corpi metallici. Vi siete mai

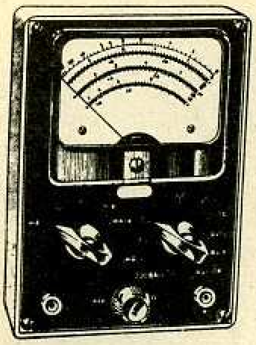

In radiotecnica le correnti elettriche si misurano con un particolare tipo di strumento, che prende il nome di « tester »; questo strumento permette di effettuare molti altri tipi di misure di varie grandezze elettriche.

chiesti il perchè di tale fenomeno? Ebbene, ve lo ricordiamo. La calamita dà origine ad un campo magnetico, invisibile, fatto tutto di forze magnetiche che si articolano nello spazio circostante i poli della calamita stessa.

Si suol dire che questo tipo di campo magnetico è un campo magnetico statico, mentre quello delle onde radio è un campo elettromagnetico dinamico, cioè in oscillazione continua. Il lettore non si preoccupi della astrusità di tali concetti che, ascoltati per la prima volta, possono risultare incomprensibili. Quel che importa è ritenere a mente che le correnti di alta frequenza, quando fluiscono attraverso un conduttore, sviluppano intorno ad esso un insieme di forze invisibili, cioè un campo elettromagnetico oscillante, che è in grado di originare una corrente oscillante vera e propria su un altro conduttore posto a una certa distanza.

Alle antenne trasmittenti vengono inviate correnti alternate di alta frequenza; queste correnti danno luogo, intorno all'antenna stessa, ad un campo elettromagnetico oscillante, la cui intensità diminuisce a mano a mano che ci si allontana dall'antenna. Dunque, le onde radio altro non sono che campi elettromagnetici oscillanti che, appena investono un conduttore metallico danno vita in esso ad un movimento di elettroni, cioè a una corrente elettrica ad alta frequenza. Quando si dice, ad esempio, che attraverso l'antenna del ricevitore radio entrano i segnali radio, si vuol dire che quell'antenna è investita dal campo elettromagnetico oscillante generato dall'antenna trasmittente, che genera nell'antenna del ricevitore radio un movimento di elettroni in avanti e all'indietro, molto rapido, cioè una corrente elettrica alternata ad alta frequenza.

MISURA DELLA CORRENTE ELETTRICA

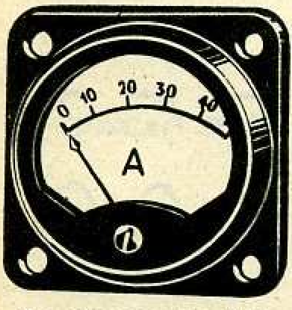

Lo strumento elettrico atto a misurare la sola intensità di corrente prende il nome di « amperometro ». Esso viene usato dagli elettrotecnici.

La corrente elettrica, intesa come movimento di elettroni, è una entità fisica, che può essere misurata. Di essa si può misurare la frequenza, se si tratta di corrente alternata, e l'intensità nel caso di corrente alternata, continua o pulsante. Lasciamo per ora da parte la misura della frequenza ed occupiamoci della misura di intensità di corrente.

L'intensità di corrente, che esprime la quantità di elettroni che attraversano nel tempo la sezione di un conduttore, ha un'unità di misura, che prende il nome di « ampere » e che abbreviatamente si indica con la sigla « A ». Anche per questa unità di misura, così come avviene per le unità di misura di lunghezza, di peso, di volume, ecc. sono stati stabiliti dei valori che sono sottomultipli dell'ampere. Essi sono:

Milliampere = un millesimo di ampere (simbolo mA)
Microampere = un milionesimo di ampere (simbolo μA)

Per misurare i pesi si usa la bilancia, per misurare le lunghezze si usa il metro; per misurare le correnti elettriche, cioè l'intensità delle correnti elettriche, si usa uno strumento che prende il nome di « amperometro », « milliamperometro » o « microamperometro », a seconda dell'entità delle correnti che si vuol misurare. In radiotecnica le correnti elettriche si misurano con un particolare strumento, che prende il nome di « tester » e che permette di effettuare anche molti altri tipi di misure delle varie grandezze elettriche.

LE RESISTENZE

I conduttori, presenti nei circuiti radio, danno via libera al passaggio degli elettroni, cioè delle correnti elettriche, che possono essere erogate da una pila, da un accumulatore o da una presa della rete-luce di casa. In molti punti di un circuito radioelettrico, tuttavia, è necessario dosare la corrente elettrica, cioè limitarne l'entità. Occorrono, insomma, lungo i percorsi della corrente elettrica delle « porte di sbarramento », che permettano di dosare a piacimento l'intensità della corrente elettrica, che può essere quella generata dalla pila o quella generata dalle onde radio. Queste particolari « porte di sbarramento » prendono il nome di RESISTENZE ELETTRICHE. Le resistenze, che prendono anche il nome di resistori, possono essere di diversi tipi e dimensioni. Ma una prima suddivisione viene fatta fra i due tipi fondamentali di resistenze: quelle fisse e quelle variabili. Le resistenze fisse rappresentano un impedimento costante al flusso di elettroni, le resistenze variabili rappresentano un impedimento che può essere variato a piacere in qualsiasi momento.

RESISTENZE CHIMICHE E A GRAFITE

I tipi di resistenze più note, in radiotecnica, sono le resistenze a filo, le resistenze chimiche, le resistenze a grafite. Le più usate sono senza dubbio le resistenze chimiche e quelle variabili a grafite.

Un esempio molto noto di resistenza variabile a grafite è rappresentato dal cosiddetto « potenziometro ». Il potenziometro è un componente presente in tutti i ricevitori radio, di qualsiasi tipo e grandezza; esso serve per regolare il volume sonoro dell'apparecchio ed è munito di un perno al quale viene applicata, appunto, la manopola di comando di volume del ricevitore radio. Il potenziometro di volume, essendo una resistenza variabile, permette di regolare la corrente rappresentativa dei segnali radio che vengono rinforzati, cioè amplificati dai circuiti dell'apparecchio radio.

Una resistenza variabile molto importante: il « potenziometro ».

MISURA DELLE RESISTENZE

Anche le resistenze hanno una unità di misura che prende il nome di « ohm »; ma molto spesso, in radiotecnica, si impiegano resistenze di valori relativamente elevati, per cui si fa uso dei seguenti multipli dell'ohm:

Chiloohm = mille ohm (KΩ)
Megaohm = un milione di ohm (MΩ)

VALORI DELLE RESISTENZE

Le resistenze usate un tempo nei circuiti radio portavano impresso, sul loro involucro esterno, l'indicazione del loro valore. Tale sistema di indicazione del valore della resistenza in cifre viene usato ancor oggi in alcuni tipi di resistenze. Ma il sistema più usato, adottato quasi universalmente dai costruttori di resistenze, è quello dell'indicazione del valore ohmmico mediante uno speciale codice a colori; sull'involucro esterno della resistenza vengono impresse alcune fascette variamente colorate; dal colore di tali fascette e dall'ordine con cui esse si succedono, si deduce l'esatto valore della resistenza. Non riuscendo a determinare il valore di una resistenza, perchè i colori sono scomparsi dal suo involucro oppure perchè essi non sono più chiaramente distinguibili, si effettua la misura della resistenza mediante uno strumento elettrico che prende il nome di OHMMETRO. I radiotecnici non fanno uso dell'ohmmetro vero e proprio, ma ricorrono all'impiego del TESTER, che è uno strumento universale che permette di effettuare una vasta gamma di misure, compresa quella della resistenza.

CODICE A COLORI DELLE RESISTENZE

Colore	1° anello	2° anello	3° anello
Nero	—	0	—
Marrone	1	1	0
Rosso	2	2	00
Arancione	3	3	000
Giallo	4	4	0000
Verde	5	5	00000
Blu	6	6	000000
Viola	7	7	—
Grigio	8	8	—
Bianco	9	9	—

4° anello	
Oro:	tolleranza ± 5 %
Argento:	tolleranza ± 10 %

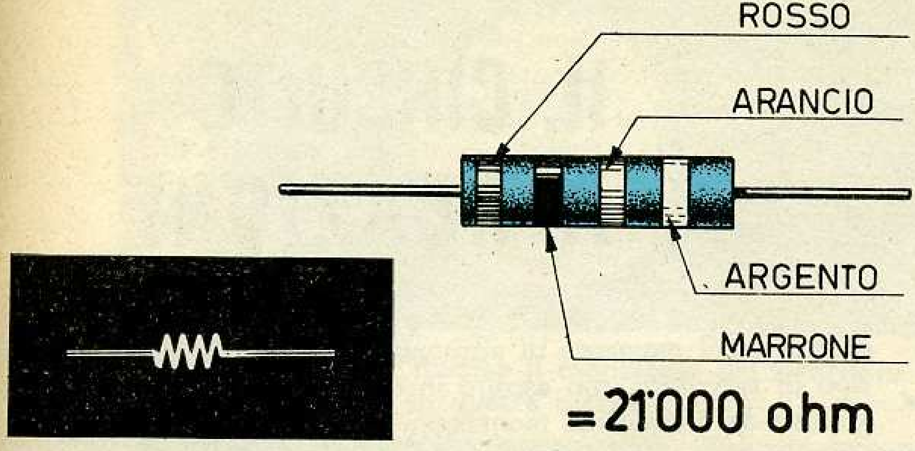

Il valome ohmmico di questa resistenza è di 21.000 ohm; esso viene facilmente dedotto dal codice a colori delle resistenze. Il disegno a sinistra rappresenta il simbolo elettrico della resistenza.

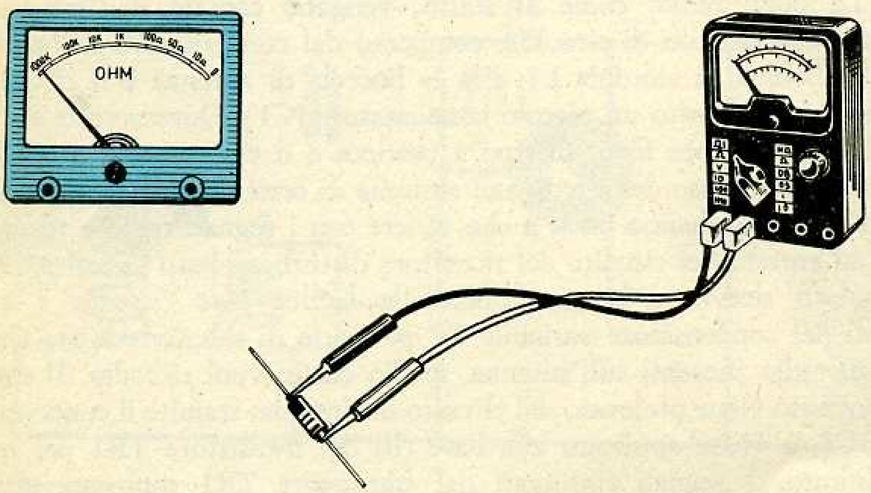

La misura della resistenza elettrica si ottiene mediante lo impiego di uno strumento che prende il nome di « ohmmetro. Anche per questa misura i radiotecnici fanno uso del tester commutato nella misura ohmmica.

Un esempio servirà ottimamente a mettere il lettore in condizioni di saper usare il codice a colori delle resistenze. Si supponga di avere in mano una resistenza in cui il primo anello sia di color rosso (il primo anello è sempre quello che si trova all'estremiatà opposta rispetto all'anello di color argento o oro), il secondo anello sia di color marrone, il terzo anello sia di color arancione, il quarto di color argento. Dal codice si rileva che per il primo anello di color rosso fa corrispondenza il numero 2; per il secondo anello, al color marrone corrisponde il numero 1, per il terzo anello, al color arancione corrispondono tre zeri; mettendo in fila uno dopo l'altro questi numeri si ottiene il valore di quella resistenza, che è di 21.000 ohm, mentre il quarto anello, di color argento, sta a significare che la tolleranza di quella resistenza è di ± 10 %. La tolleranza sta ad indicare la percentuale di discordanza, in più o in meno, tra il valore effettivo della resistenza e il valore indicato. Nel caso della resistenza presa come esempio, essa può avere i valori limiti di 23.100 ohm e di 18.900 ohm, senza che nessuna colpa possa attribuirsi alla casa costruttrice.

IL CIRCUITO A REAZIONE

È giunto ora il momento di affrontare il montaggio di un secondo tipo di ricevitore, con ascolto in cuffia, e un pochino più impegnativo del precedente montaggio. Esaminiamo prima di tutto lo schema elettrico di questo ricevitore, che come al solito è fatto tutto di simboli e di sigle.

Un ricevitore dotato di grande sensibilità.

Le onde radio, come al solito, vengono captate dall'antenna ed entrano nel circuito di sintonia, composto dal condensatore variabile C2 e dalla bobina di sintonia L1. Fra la boccola di antenna e il circuito di sintonia è interposto un piccolo condensatore (C1). Questa volta si tratta di un condensatore fisso, di tipo a pasticca e il cui compito è quello di opporsi all'ingresso del circuito di sintonia di certi disturbi, cioè di taluni segnali che non hanno nulla a che vedere con i segnali radio e se riuscissero ad entrare nel circuito del ricevitore disturberebbero l'ascolto. Anche in questo ricevitore la posizione delle lamine fisse rispetto a quelle mobili del condensatore variabile C2 permette di selezionare, tra i molti segnali radio presenti sull'antenna, quello che si vuol ricevere. Il segnale sintonizzato viene prelevato dal circuito di sintonia, tramite il condensatore fisso C3, e viene applicato alla base (b) del transistore TR1 per essere amplificato. I segnali rinforzati dal transistore TR1 sono presenti sul collettore (c) di TR1 e vengono applicati alla cuffia per essere trasformati in voci e suoni.

Rinforzo, teoricamente infinito, delle onde radio.

E vediamo ora in che cosa consiste il potenziamento di questo circuito rispetto a quello precedentemente esaminato e realizzato. Sull'emittore (e) di TR1 è presente una piccola parte del segnale amplificato; esso viene applicato alla resistenza variabile R1, che prende il nome di potenziometro; questo segnale, che determina una debolissima corrente elettrica, è presente anche nell'avvolgimento secondario della bobina L1. La bobina L1, infatti, è composta da due avvolgimenti vicini tra loro, ma isolati elettricamente; l'avvolgimento contrassegnato con i numeri 3-4 prende il nome di « avvolgimento primario della bobina »; l'avvolgimento contrassegnato con i numeri 1-2 prende il nome di « avvolgimento secondario della bobina ». Trattandosi di una corrente di alta frequenza, questa si trasferisce, attraverso l'aria, nell'avvolgimento primario della bobina L1. Avevamo già detto che le correnti di alta frequenza tendono a sfuggire dai conduttori e a trasferirsi altrove sotto forma di campi elettromagnetici.

Con tale strattagemma, dunque, si riesce a riportare nel circuito di sintonia, costituito dal condensatore variabile C2 e dall'avvolgimento

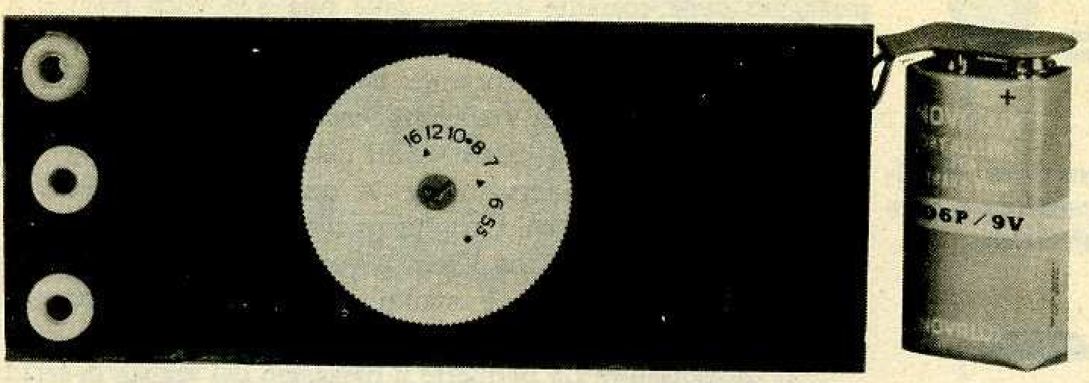

La foto qui riportata illustra il montaggio del ricevitore a reazione analizzato in questa ottava ora.

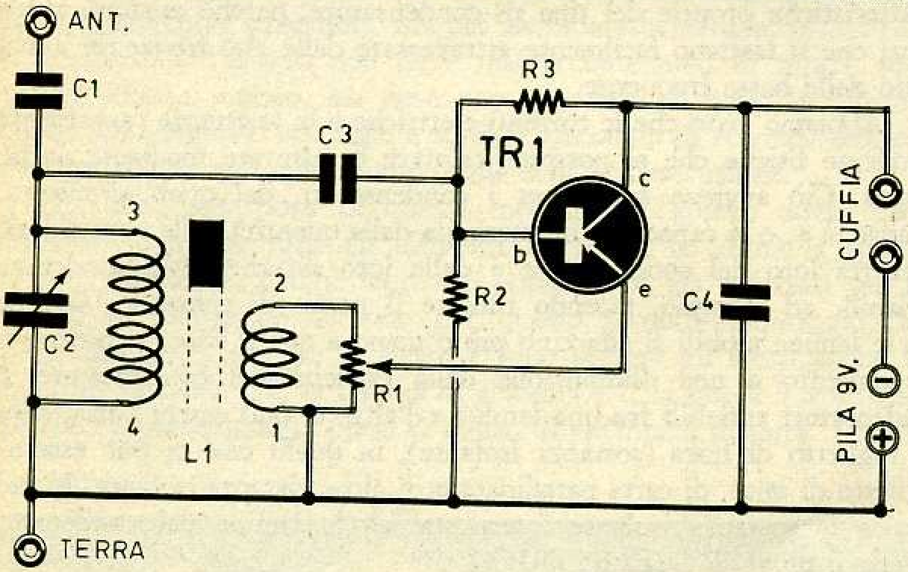

Disegno rappresentativo del circuito teorico del ricevitore a reazione.

primario della bobina L1, una parte del segnale di alta frequenza già amplificato; esso ritorna poi al transistore TR1 dove subisce un ulteriore processo di amplificazione. E questo ritorno del segnale si protrarrebbe per una infinità di volte; ma in questo caso si avrebbe un'amplificazione talmente grande da produrre nella cuffia un fischio acutissimo; occorre dunque regolare questo ciclo di successive amplificazioni del segnale di alta frequenza, e ciò si ottiene intervenendo sulla resistenza variabile R1. Questo strattagemma prende il nome di REAZIONE e permette di ottenere una grande amplificazione dei segnali radio.

I CONDENSATORI FISSI

Tutti i condensatori, di tipo variabile o fisso, sono componenti che permettono il passaggio, attraverso di essi, delle sole correnti alternate, siano esse di bassa o di alta frequenza; attraverso i condensatori, invece, non passano le correnti continue; essi rappresentano, dunque, una porta chiusa alle correnti erogate dalle pile e dagli accumulatori. Nel caso del

Condensatore in polistirolo e simbolo elettrico.

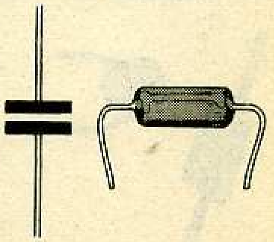

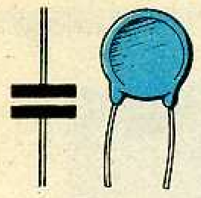

Condensatore di tipo a pasticca e simbolo elettrico.

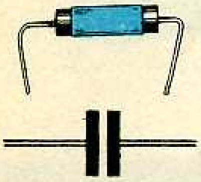

Condensatore a carta e simbolo.

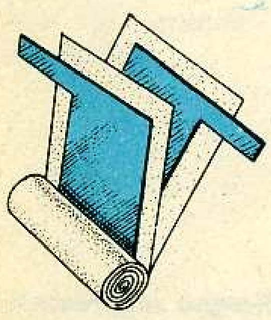

Il disegno illustra la composizione interna di un condensatore di tipo a carta.

Esempi di condensatori a mica.

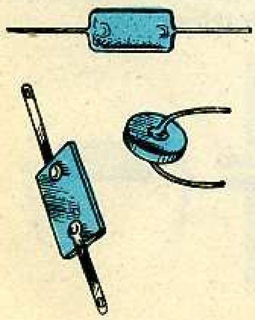

circuito elettrico in esame, i condensatori C3 e C4 non permettono il passaggio della corrente erogata dalla pila a 9 volt, mentre si lasciano attraversare dai segnali radio. Il condensatore C4, ad esempio, non può essere attraversato dalla corrente della pila, mentre lascia passare gli eventuali segnali di alta frequenza ancora presenti in prossimità della cuffia, che darebbero luogo a disturbi nella ricezione.

Il lettore avrà osservato che in questo secondo tipo di ricevitore manca il diodo al germanio, che provvede a sbarrare il passo ai segnali di alta frequenza lasciando via libera a quelli di bassa frequenza. Tale compito, in questo circuito, è svolto dallo stesso transistore TR1, che amplifica i segnali radio secondo il principio della reazione.

Il condensatore C3 permette il passaggio dei segnali di alta frequenza, provenienti dall'antenna, sino alla base (b) del transistore TR1, ma non permette che, attraverso ad esso, possano passare i segnali di bassa frequenza, presenti sul collettore (c) di TR1. E ciò avviene in virtù delle caratteristiche proprie del tipo di condensatore, perchè esistono condensatori che si lasciano facilmente attraversare dalle alte frequenze e molto meno dalle basse frequenze.

Abbiamo visto che le correnti elettriche e le resistenze rappresentano grandezze fisiche che si possono valutare e misurare mediante unità di misura. Ciò avviene anche per i condensatori, dei quali si misura la « capacità »; e la capacità è determinata dalla quantità delle superfici affacciate tra loro del condensatore e dalla loro distanza. Nei condensatori variabili, ad esempio, facendo ruotare il perno di comando, si ottiene che le lamine mobili si affaccino più o meno a quelle fisse; si ottiene così un aumento o una diminuzione della capacità del condensatore. Nei condensatori variabili fra una lamina e l'altra ci può essere l'aria, oppure un foglietto di mica (sostanza isolante), in quelli fissi ci può essere un foglietto di mica, di carta paraffinata o di altra sostanza isolante. In radiotecnica la sostanza isolante interposta fra le lamine del condensatore, prende il nome di DIELETTRICO.

L'unità di misura della capacità è il « farad » (abbrev. F.). Poichè il farad rappresenta un'unità di misura molto grande, in radiotecnica si fa impiego dei sottomultipli del farad. Essi sono:

il microfarad µF = un milionesimo di farad
il picofarad pF = un milionesimo di milionesimo di farad

Il picofarad viene usato, generalmente, per valori fino a 100.000 pF, mentre per le capacità più grandi si usa il microfarad. I condensatori di capacità elevata prendono il nome di « elettrolitici ». In questo tipo di condensatori il dielettrico è costituito da uno strato di ossido che si forma sulle superfici affacciate di due nastri di alluminio, separate da un elettrolita, quando esse sono sottoposte alle forze elettriche.

Sui terminali dei condensatori elettrolitici viene sempre indicata la polarità; in essi esiste sempre, quindi, un terminale positivo e un terminale negativo; ciò significa che i condensatori elettrolitici, contrariamente per quanto avviene per gli altri tipi di condensatori, non possono essere inseriti nei circuiti in un verso qualsiasi, perchè l'inversione delle polarità danneggerebbe irreparabilmente il condensatore.

MONTAGGIO DEL RICEVITORE A REAZIONE

Tutti i montaggi radioelettrici vengono eseguiti in due tempi: prima si esegue il lavoro di ordine meccanico, quello che richiede l'uso delle pinze, del cacciavite, del trapano, ecc., poi si esegue il cablaggio, cioè quel lavoro che si effettua con il saldatore in mano; così è stato fatto per il primo tipo di ricevitore e così occorre fare per questo ricevitore con circuito a reazione. Ma per assimilare bene ogni concetto di natura costruttiva, l'allievo non deve mai accontentarsi di tenere sott'occhio lo schema pratico, ma deve seguire prima di tutto lo schema teorico, quello composto con sigle e simboli, cercando di rendersi sempre conto del perchè di ogni operazione, prima della sua esecuzione.

Anche per l'esecuzione di questo montaggio, l'allievo dovrà preoccuparsi di acquistare tutti gli elementi necessari; in verità la spesa, questa volta è minore, perchè molti dei componenti già utilizzati per il primo tipo di montaggio servono per questo secondo montaggio; la cuffia, ad esempio, è sempre la stessa e sono sempre gli stessi anche il condensatore variabile C2, la bobina di sintonia L1, la pila a 9 volt e le boccole che rappresentano le prese di cuffia, di antenna e di terra. I com-

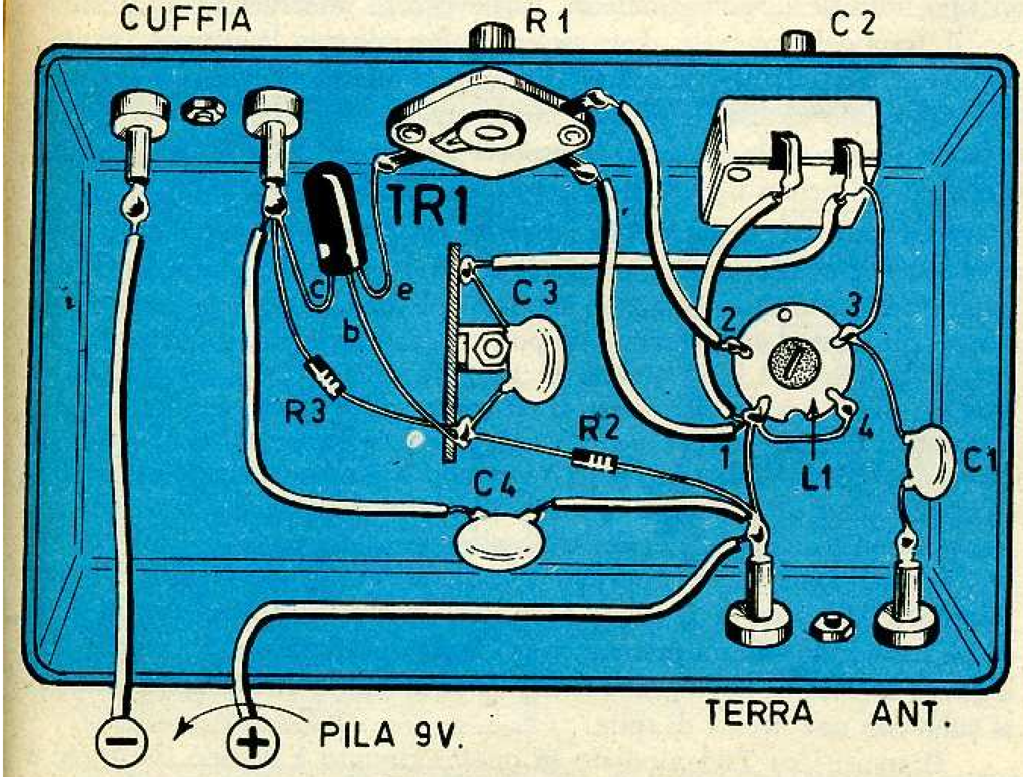

Piano di montaggio (schema pratico) del ricevitore a reazione con ascolto in cuffia e alimentazione con pila a 9 volt.

COMPONENTI DEL RICEVITORE A REAZIONE

C1 = 47 pF (condensatore a pasticca)
C2 = 500 pF (condensatore variabile)
C3 = 250 pF (condensatore a pasticca)
C4 = 10.000 pF (condensatore a pasticca)
R1 = 1.000 ohm (potenziometro a variazione logaritmica)
R2 = 150.000 ohm (marrone-verde-giallo)
R3 = 1 megaohm (marrone-nero-verde)
TR1 = OC44 (transistore tipo pnp)
L1 = bobina tipo Corbetta CS3/BE

ponenti che si dovranno acquistare si riducono, quindi, a tre condensatori fissi, a due resistenze fisse e a una resistenza variabile (potenziometro); occorre acquistare inoltre un altro tipo di transistore. Comunque nell'elenco componenti sono riportati tutti i dati necessari per poter acquistare con sicurezza e precisione il materiale occorrente. Il condensatore C3, come si nota nello schema pratico del ricevitore, è fissato sui terminali di una basetta di bachelite che, a sua volta, è fissata sulla scatola-contenitore per mezzo di una squadretta metallica, di una vite e di un dado.

Basette e ancoraggi.

Si fa ricorso molto spesso, nei montaggi radioelettrici, all'impiego di basette a due o più terminali, allo scopo di conferire al montaggio una certa rigidità e compattezza. Queste morsettiere vengono vendute in commercio in misure di un metro circa e spetta all'allievo il compito di ritagliare la basetta nella misura utile. Il potenziometro R1 disegnato nello schema pratico è di tipo miniatura, ma l'allievo potrà acquistare anche un tipo normale di potenziometro, sprovvisto di interruttore, ricordando che il terminale centrale deve essere collegato con l'emittore (e) del transistore TR1. Le operazioni di cablaggio di questo ricevitore a reazione non richiedono alcuna particolare interpretazione, purchè l'allievo ricordi e faccia tesoro delle spiegazioni esposte nel corso della descrizione del cablaggio del primo tipo di ricevitore.

Se in cuffia si ode un fischio, occorrerà intervenire sul perno del potenziometro R1, al quale sarà stata applicata una manopola, ruotando fino al punto in cui il fischio scompare. Agendo sul condensatore variabile C2 si cerca l'emittente locale e ruotando poi, con un cacciavite, il nucleo di ferrite della bobina L1, si ottiene la massima intensità sonora. La stessa antenna usata per il primo tipo di ricevitore deve essere applicata anche a questo ricevitore a reazione. Occorre peraltro anche un conduttore di terra, applicando sulla boccola di terra uno spinotto collegato ad un filo di rame legato, all'estremità opposta, al rubinetto dell'acqua o a una conduttura dell'acqua potabile, del termosifone o del gas.

Questa legatura deve essere effettuata avvolgendo il filo di rame attorno al rubinetto o al tubo, dopo aver raschiato le parti con la lama di un temperino, sino a mettere in evidenza la lucentezza dei metalli. La legatura si effettua avvolgendo il filo di rame molto strettamente attorno al tubo con una decina di spire.

Il transistore TR1 montato in questo circuito è di tipo OC44 ed è prodotto dalla Philips come quello montato nel primo tipo di ricevitore. L'allievo, tuttavia, non trovando in commercio un transistore di tipo OC44, potrà far ricorso a qualsiasi altro tipo di transistore, purchè esso sia adatto per l'alta frequenza, sia di tipo pnp (vedremo più avanti il significato di tale espressione) e non sia un transistore di potenza; i risultati, pur operando la sostituzione del transistore con uno di tipo diverso, saranno pressochè identici.

Anche per questo tipo di ricevitore, l'interruttore, che permette di accendere e spegnere il circuito, è rappresentato dall'innesto o dalla estrazione dei due spinotti della cuffia dalle relative boccole. L'allievo tenga presente che facendo rimanere inserita la cuffia nel circuito, la pila si consuma rapidamente.

COLLEGAMENTO DI PILE IN SERIE

La pila da 9 volt, usata per questo tipo di ricevitore, è di tipo normale, di quelle normalmente montate sui ricevitori a transistori di tipo tascabile. Questa pila, volendo mantenere acceso il ricevitore per molto tempo, può scaricarsi presto. Ma volendo avere una lunga autonomia di funzionamento del ricevitore, senza ricorrere al ricambio della pila, si può ricorrere all'impiego di due pile da 4,5 volt ciascuna, di quelle usate per le lampade tascabili. Queste due pile devono essere collegate in « serie » tra loro.

Cosa significa tale espressione? Ve lo diciamo subito. Collegare in serie due o più pile significa collegare il morsetto positivo di una con quello negativo dell'altra; il morsetto positivo della seconda con quello negativo della terza e così via; i due morsetti che rimangono liberi rappresentano i morsetti utili della batteria di pile. Nel nostro caso, necessitando una pila da 9 volt, si dovranno utilizzare due pile da 4,5 volt, collegando tra loro il morsetto positivo di una con quello negativo dell'altra; i due morsetti rimasti liberi, quello positivo e quello negativo vanno collegati al circuito del ricevitore.

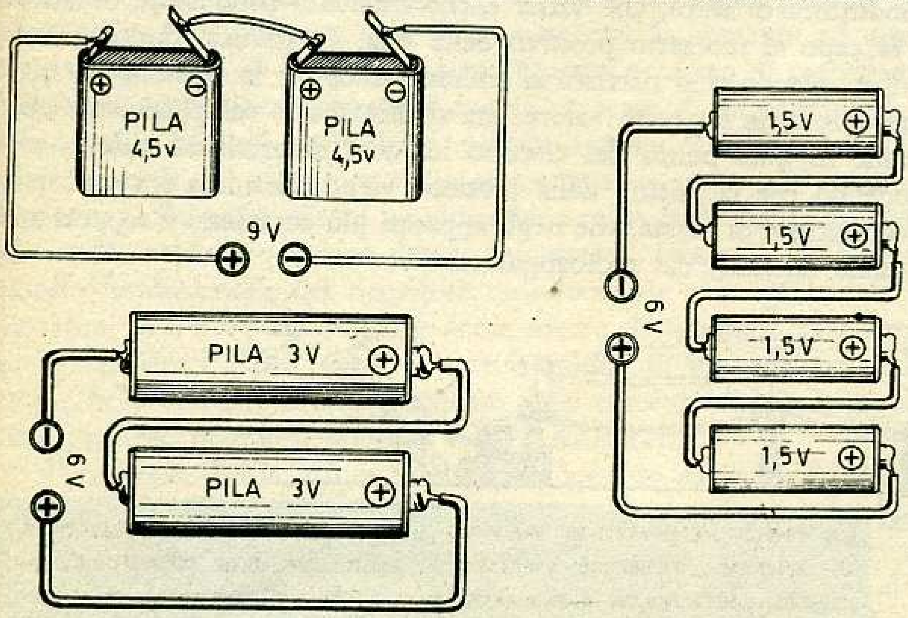

Esempi di collegamenti di pile collegate in serie tra di loro. Quando i valori delle tensioni delle pile inserite nel collegamento sono uguali, allora il valore della tensione risultante è dato dal prodotto del valore della tensione di una singola pila per il numero delle pile.

LA TENSIONE ELETTRICA

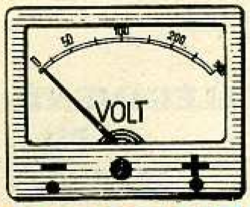

Per la misura della tensione elettrica si usano valori multipli e sottomultipli del volt; i più comuni, in radiotecnica, sono:
Chilovolt = mille volt (simbolo kV)
Millivolt = un millesimo di volt (simbolo mV)
Microvolt = un milionesimo di volt (simbolo µV).

La tensione elettrica rappresenta una delle grandezze fisiche fondamentali dell'elettrotecnica in generale e della radiotecnica in particolare. Si tratta di un dato rilevabile in tutti i punti di un circuito radio con valori diversi. Anche la tensione elettrica, come accade per la corrente e per la resistenza, è un dato che si rileva mediante uno strumento che prende il nome di VOLTMETRO. I radiotecnici non fanno uso di un voltmetro vero e proprio, ma impiegano il tester, che è uno strumento di misure elettriche molteplice e nel quale è compreso anche il voltmetro.

La tensione elettrica è una grandezza che si misura sempre fra due punti e non in un punto solo. Essa rappresenta la forza elettrica in grado di generare una corrente elettrica, cioè di mettere in movimento attraverso un conduttore gli elettroni. Sui morsetti delle pile, ad esempio, si misura una tensione, che può essere di 1,5 volt, 3 volt, 4,5 volt, 9 volt; questa tensione vuol significare che tra i morsetti della pila esiste una forza elettrica, che a sua volta trae origine dalle forze di natura chimica presenti all'interno della pila stessa, e che è in grado di produrre corrente elettrica. Sul collettore del transistore TR1 del ricevitore a reazione esiste una tensione; ma tale affermazione ha significato preciso quando si sottointende che tale tensione viene misurata fra il collettore del transistore e il conduttore di terra, che viene anche chiamato conduttore di massa e che fa capo al morsetto positivo della pila. In pratica, dunque, non si usa dire, quando ci si riferisce ai circuiti radio, che la tensione fra questo e quel punto ha un certo valore, ma si dice molto semplicemente che la tensione in quel punto del circuito ha quel determinato valore, sottointendendo che la misura della tensione viene effettuata fra quel punto e il conduttore di massa, che negli apparati più complessi è rappresentato dal telaio (chassis) del radioapparato.

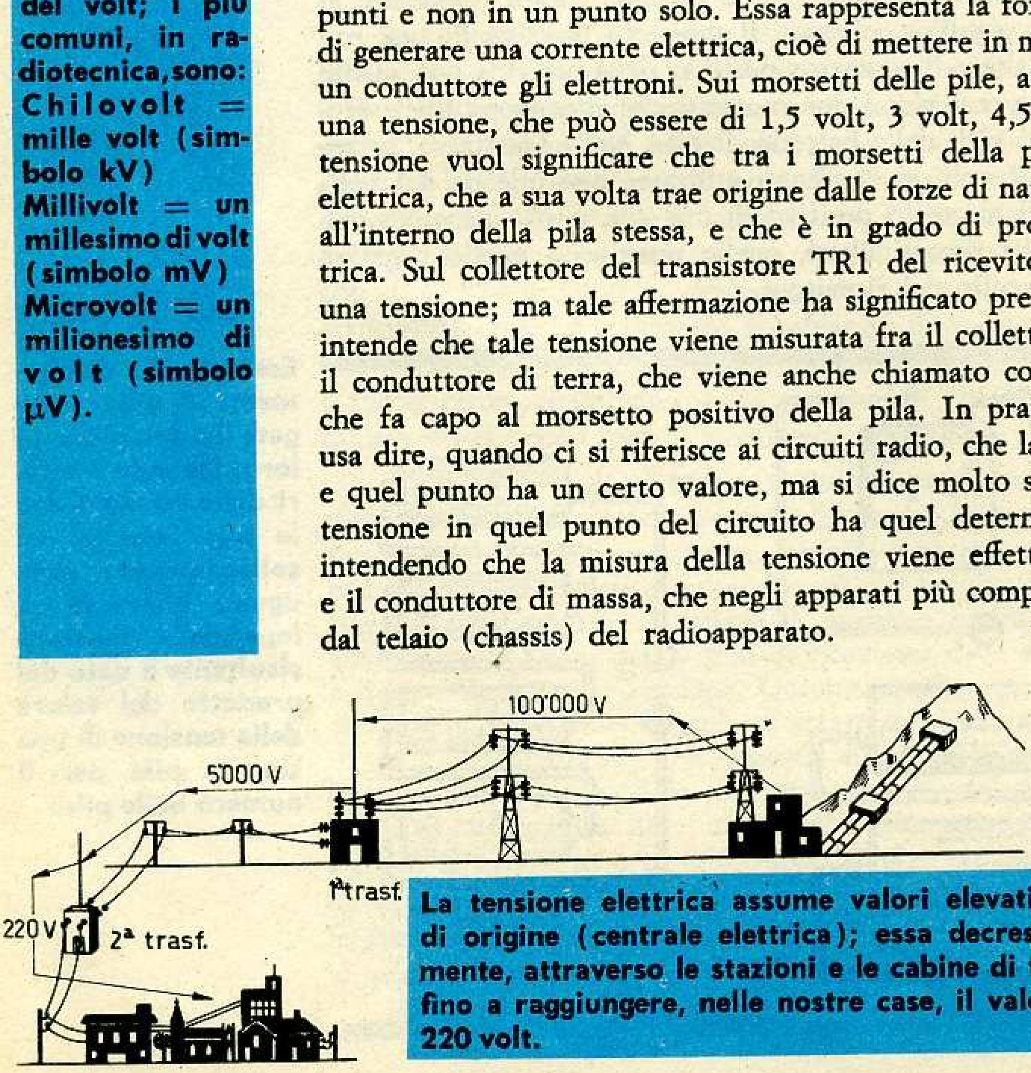

La tensione elettrica assume valori elevati nella stazione di origine (centrale elettrica); essa decresce progressivamente, attraverso le stazioni e le cabine di trasformazione, fino a raggiungere, nelle nostre case, il valore massimo di 220 volt.

L'unità di misura della tensione elettrica è il « volt » (abbrev. V.). Si dice che quando sui terminali di una resistenza di 1 ohm viene applicata la tensione di 1 volt, quella resistenza è percorsa dalla corrente di 1 ampere. Ma per avere una sensazione più reale del concetto di tensione elettrica è bene ricordare taluni valori delle tensioni elttriche, che sono fondamentali e con i quali tutti noi, più o meno, abbiamo a che fare ogni giorno. Delle pile si è già detto: la loro tensione oscilla fra 1,5 volt e 9 volt; la tensione presente nelle nostre case può essere di 130 volt o di 220 volt. Negli apparecchi radio normali si hanno, in molti punti del circuito, tensioni comprese fra i 200 e i 300 volt. Sulle linee elettriche di distribuzione di energia, fra le centrali e le località di distribuzione, la tensione è a volte superiore ai 200.000 volt.

MODULAZIONE E RIVELAZIONE

Pur avendo affrontato finora, direttamente, i circuiti di due ricevitori radio, e pur avendone interpretato, nelle loro linee essenziali, i concetti fondamentali della radiotrasmissione e radioricezione, occorre, ora, prima di procedere oltre, assimilare due concetti fondamentali della radiotecnica: quello della « modulazione » e quello della « rivelazione ». È stato detto che le correnti elettriche ad alta frequenza (correnti oscillanti) presentano la caratteristica di « sfuggire » dai conduttori elettrici e di diffondersi nell'aria sotto forma di onde elettromagnetiche come sono, appunto, le onde radio. Ma se alle antenne trasmittenti venissero convogliate soltanto correnti elettriche ad alta frequenza, negli apparecchi radio non si ascolterebbero voci e suoni. L'intero processo di trasmissione e ricezione radio si svolge così: le onde sonore, quelle generate dalla voce umana o da uno strumento musicale, vengono captate dal microfono e trasformate, da tale componente, in corrente elettrica di bassa frequenza, così come avviene quando si parla al telefono; questa corrente elettrica di bassa frequenza, viene fatta mescolare, nelle stazioni radiotrasmittenti, con la corrente oscillante di alta frequenza e il tutto viene inviato all'antenna trasmittente che diffonde nell'etere le onde radio. Dunque, la corrente oscillante funge da veicolo per la corrente di bassa frequenza: le onde radio portano con sè le voci e i suoni proprio perchè nella stazione trasmittente si è fatta mescolare la corrente ad alta frequenza con quella di bassa frequenza. Questo processo di mescolamento della corrente di bassa frequenza con la corrente di alta frequenza, che avviene nelle stazioni trasmittenti, prende il nome di MODULAZIONE.

Smistamento delle frequenze.

L'antenna del ricevitore radio capta le onde radio e il circuito del ricevitore radio provvede ad amplificarle, perchè così come esse entrano nel circuito sarebbero troppo deboli per essere... lavorate; ma il secondo compito fondamentale del ricevitore radio consiste nel... gettar via l'alta frequenza, che è servita soltanto come mezzo di trasporto attraverso lo spazio delle voci e dei suoni. Si tratta, quindi, di un processo di separazione della corrente di alta frequenza dalla corrente di bassa frequenza, e tale processo prende il nome di RIVELAZIONE.

Il processo di rivelazione può essere compiuto da un diodo al germanio (primo tipo di ricevitore descritto) oppure da un transistore (secondo tipo di ricevitore descritto) o, come vedremo, da una valvola elettronica. La corrente rappresentativa del segnale di bassa frequenza, uscente da tali componenti, prende il nome di corrente di rivelazione e i segnali radio prendono il nome di « segnali rivelati ».

IL CIRCUITO REFLEX

Dopo il ricevitore a circuito semplice, quello con circuito a reazione, la terza tappa della radiotecnica pratica è rappresentata dal circuito riflesso, denominato « reflex ». Che cosa significa circuito riflesso?

È presto detto. Significa sottoporre lo stadio amplificatore di alta frequenza anche ad un processo di amplificazione dei segnali di bassa frequenza; in altre parole, si fa... riflettere sul primo stadio amplificatore il segnale più amplificato e rivelato.

Ma passiamo senz'altro all'esame del circuito elettrico del ricevitore reflex, per poter comprendere più intimamente il principio di funzionamento. I segnali radio captati dall'antenna sono applicati, tramite il condensatore C1 al circuito di sintonia, rappresentato dal condensatore variabile C2 e dalla bobina L1. La presenza del condensatore C1 e il principio di funzionamento del circuito di sintonia sono sempre gli stessi, quindi è inutile ripeterci. Il passaggio del segnale sintonizzato dall'avvolgimento L1 all'avvolgimento L2 avviene per « induzione »; tale espressione vuole interpretare un fondamentale concetto di elettrotecnica: quello del trasferimento di energia elettrica da un circuito a un altro sotto

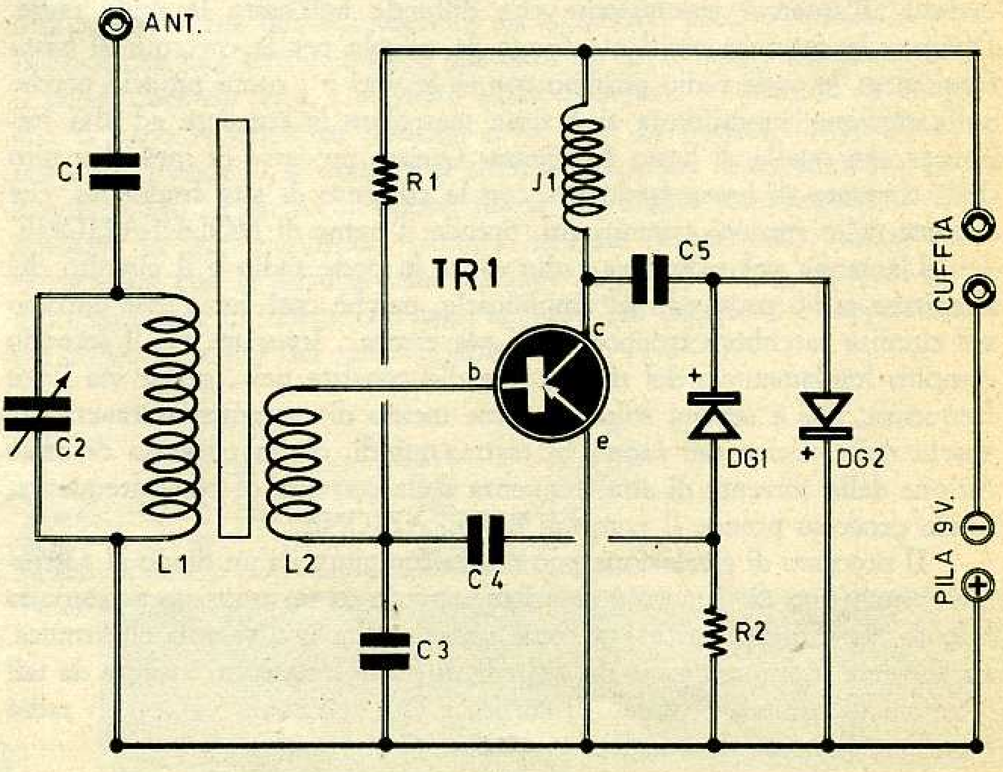

Schema elettrico del ricevitore con ricezione in cuffia e funzionante in circuito reflex.

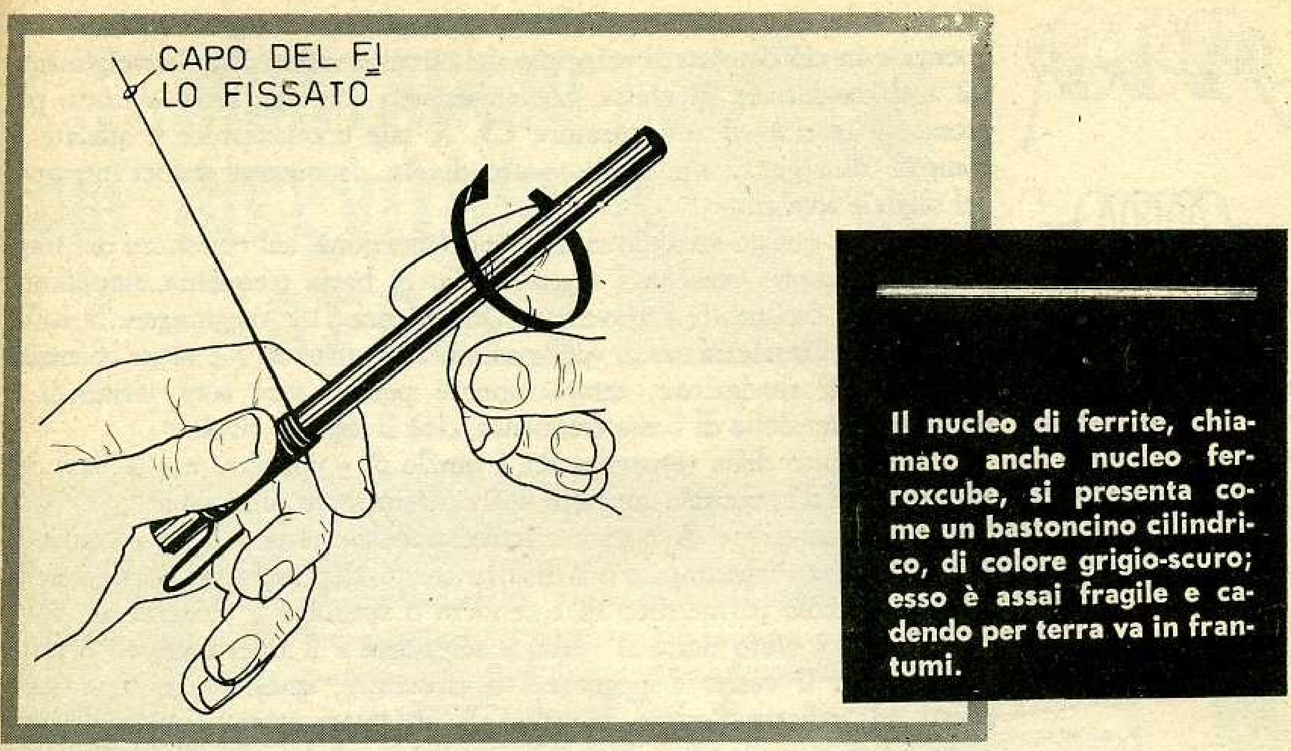

Il nucleo di ferrite, chiamato anche nucleo ferroxcube, si presenta come un bastoncino cilindrico, di colore grigio-scuro; esso è assai fragile e cadendo per terra va in frantumi.

Gli avvolgimenti delle bobine L 1 ed L2 vengono effettuati su uno stesso nucleo di ferrite, nella misura standard 8 x 140 mm. Le due bobine, che devono essere autocostruite, vengono realizzate praticamente con l'aiuto delle due mani; con una mano si mantiene teso il filo e lo si guida durante l'avvolgimento; con l'altra mano si imprime un movimento di rotazione lento e continuo al nucleo cilindrico.

forma di energia elettromagnetica. Spieghiamoci meglio; le correnti di alta frequenza, presenti nell'avvolgimento L1, generano dei campi elettromagnetici, che avvolgono l'avvolgimento L2 e provocano in esso una corrente, più intensa, di alta frequenza dello stesso tipo di quella presente nell'avvolgimento L1. Il nucleo di ferrite, disegnato con un rettangolo allungato, sul quale vengono effettuati praticamente i due avvolgimenti L1 ed L2, favorisce il trasferimento di energia dall'avvolgimento L1 all'avvolgimento L2 e funge anche da antenna captatrice dei segnali radio, rinforzando gli stessi segnali captati dall'antenna. Sull'avvolgimento L2 è presente il solo segnale radio di alta frequenza selezionato dal circuito di sintonia; esso viene applicato alla base (b) del transistore TR1, che provvede ad amplificarlo. All'uscita del transistore TR1, cioè sul suo collettore (c) sono presenti i segnali radio di alta frequenza amplificati, che incontrano due ipotetiche vie di transito: quella del condensatore C5 e quella dell'avvolgimento contrassegnato con la sigla J1, che vuol essere un'impedenza di alta frequenza. L'impedenza di alta frequenza è pur essa una bobina, il cui compito è quello di costituire uno sbarramento ai segnali di alta frequenza e di lasciar via libera a un passaggio dei segnali di bassa frequenza. Dunque, i segnali di alta frequenza amplificati, uscenti dal collettore del transistore TR1, non possono far altro che prendere la via del condensatore C5. Subito dopo questo condensatore sono presenti due diodi al germanio: DG1 e DG2; questi due diodi provvedono al processo di rivelazione dei segnali radio, cioè trasformano i segnali radio di alta frequenza in segnali radio di bassa frequenza. Perchè sono stati impiegati due diodi anzichè uno solo? Semplicemente per poter raddoppiare la tensione rivelata e rendere più potente il ricevitore. La tensione rivelata viene prelevata tramite il condensatore C4 e viene riportata nell'avvolgimento L2 e quindi nella base del transistore TR1.

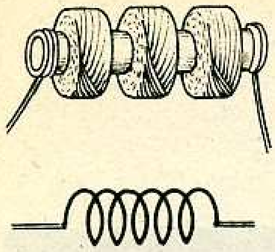

Disegno reale (sopra) e simbolo (sotto) della impedenza di alta frequenza J 1 montata nel circuito reflex.

Questa volta al transistore si chiede di amplificare i segnali di bassa frequenza e in ciò consiste il principio del circuito reflex. Fra l'avvolgimento L2 e il conduttore di massa, contrassegnato con una linea a tratto più grosso, è inserito il condensatore C3. A tale condensatore è affidato il compito di... gettar via quella parte di alta frequenza ancora presente nel segnale rivelato.

E dopo questo secondo ciclo di amplificazione, sul collettore del transistore TR1 sono presenti i segnali radio di bassa frequenza amplificati, che possono facilmente attraversare l'impedenza J1 e raggiungere la cuffia nella quale si trasformano in voci e suoni. La resistenza R2 viene chiamata resistenza di rivelazione, semplicemente perchè suoi suoi terminali è presente la tensione di bassa frequenza, cioè il segnale rivelato.

Il compito della resistenza R1 è quello di « polarizzare » la base del transistore TR1, ma tale concetto verrà interpretato più avanti.

L'alimentazione di questo circuito è ottenuta con la solita pila da 9 volt, mentre l'inserimento o il disinserimento degli spinotti di cuffia nelle relative boccole permettono di accendere o spegnere il ricevitore.

Finora è stato usato il verbo « accendere » il ricevitore, ed è stato usato anche il verbo « spegnere » il ricevitore; questi verbi non sono adatti ad indicare l'esatta funzione di un interruttore. Invece di dire « accendere o spegnere » il ricevitore, si deve dire « chiudere o aprire » il circuito del ricevitore; infatti, l'inserimento degli spinotti di cuffia, nelle due boccole, « chiude » il circuito di alimentazione elettrica del ricevitore, mentre l'operazione di disinserimento « apre » il circuito.

Queste stesse operazioni siamo soliti eseguirle ogni giorno nelle nostre case, quando accendiamo e spegnamo la luce; infatti quando ruotiamo la levetta di un interruttore o ne premiamo il pulsante, per accendere una lampadina o un lampadario, non facciamo altro che « chiudere » il circuito di accensione elettrica dell'impianto di casa, unendo tra loro i terminali di due conduttori elettrici mediante lo stesso interruttore; quando si spegne la luce, invece, si « apre » il circuito, staccando tra loro mediante l'interruttore, i terminali di due conduttori elettrici. Questi due verbi « chiudere » e « aprire » il circuito devono entrare nella parlata comune del futuro radiotecnico, perchè essi appartengono appunto al gergo radiotecnico.

Negli impianti elettrici domestici si ritrovano gli esempi più elementari che interpretano i concetti « circuito chiuso » e « circuito aperto ». Quando la lampadina è accesa (schema riportato a sinistra), allora si dice che il circuito elettrico è « chiuso »; quando la lampadina è spenta (schema rappresentato a destra), allora si dice che il circuito è « aperto ».

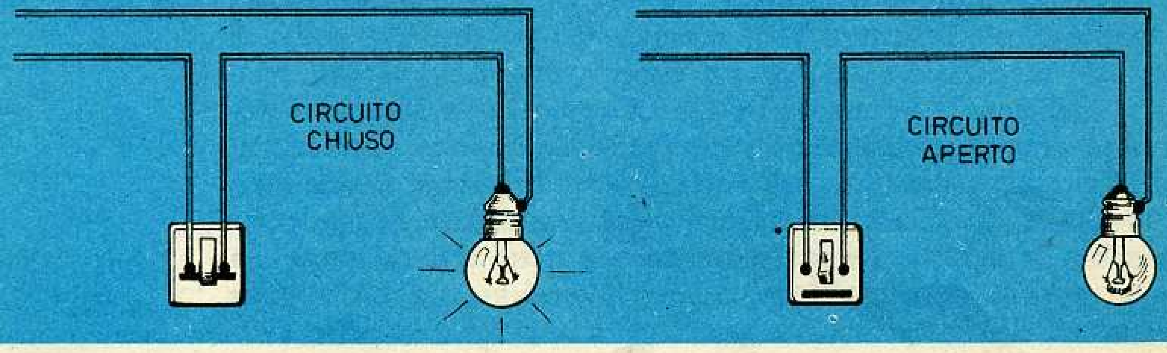

MONTAGGIO del RICEVITORE REFLEX

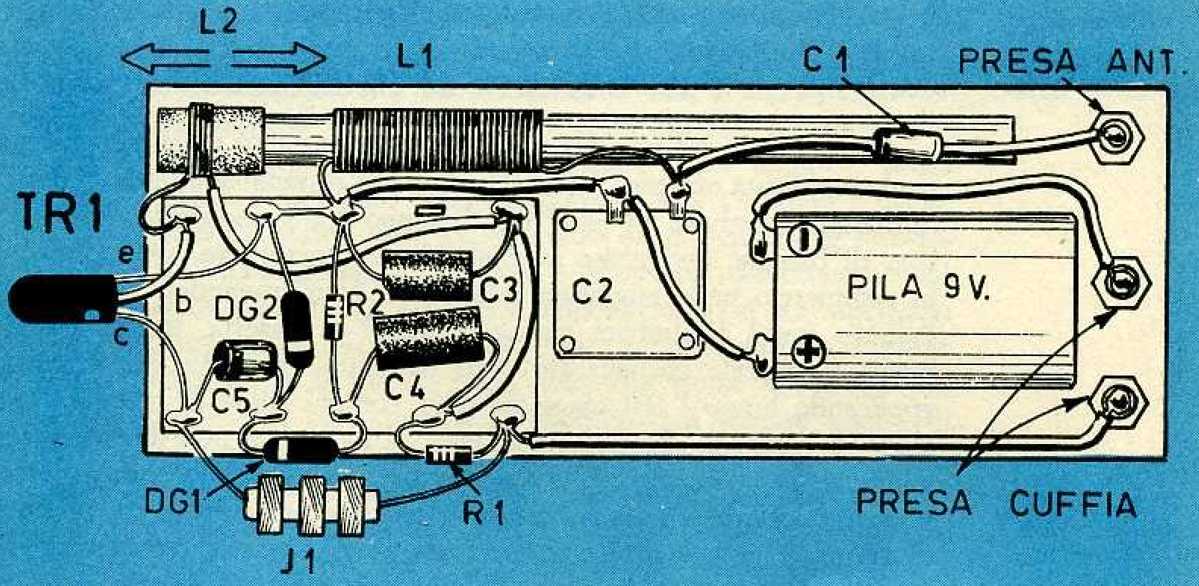

Schema pratico del ricevitore funzionante in circuito reflex. All'estrema sinistra dello schema è visibile la piccola basetta di forma rettangolare, munita di cinque terminali, distribuiti in parti uguali lungo i lati maggiori: essa permette di semplificare le operazioni di cablaggio e di realizzare un montaggio in modo razionale e compatto.

Il montaggio del ricevitore con circuito reflex deve essere fatto nel modo indicato nello schema pratico. Anche per questo tipo di montaggio si comincerà col preparare prima tutti gli elementi necessari ed eseguendo poi ogni lavoro meccanico. Nel disegno rappresentativo dello schema pratico il montaggio del ricevitore è fatto su una basetta di forma rettangolare, che può essere di legno oppure di bachelite. Questa stessa basetta, a lavoro ultimato, potrà essere introdotta in una scatolina con funzione di mobile. Trattandosi, tuttavia, di un montaggio sperimentale, come lo sono i precedenti, ci si potrà accontentare della sola basetta rettangolare, perchè quel che importa di più non è tanto la realizzazione di un ricevitore radio dotato di un piacevole aspetto esteriore, quanto l'esercizio pratico dell'allievo radiotecnico.

COMPONENTI DEL RICEVITORE REFLEX

C1 = 47 pF (condensatore in polistirolo)
C2 = 500 pF (condensatore variabile)
C3 = 50.000 pF (condensatore a carta)
C4 = 100.000 pF (condensatore a carta)
C5 = 1000 pF (condensatore in polistirolo)
R1 = 1 megaohm (marrone-nero-verde)
R2 = 38.000 ohm (arancione-grigio-arancione)
DG1-DG2 = diodi al germanio
L1-L2 = bobine di sintonia
J1 = impedenza di alta frequenza tipo Geloso 558
TR1 = OC44 (transistore tipo pnp)
pila = 9 volt

LA BASETTA DI BACHELITE

All'estrema sinistra del disegno rappresentativo dello schema pratico risulta applicata una piccola basetta di forma rettangolare, munita di cinque terminali, distribuiti in parti uguali lungo i lati maggiori. Anche questo tipo di basette, così come avviene per le morsettiere, vengono vendute in commercio nella lunghezza desiderata. Non trovandole in commercio, l'allievo potrà facilmente realizzare da sè la basetta, munendosi di un pezzo di cartone bachelizzato o di bachelite, di forma rettangolare, ed applicando lungo i lati maggiori dieci comuni ribattini metallici in funzione di terminali. Basterà ricordarsi che i ribattini non devono essere di alluminio per realizzare un ottimo lavoro; sull'alluminio, infatti, lo stagno non... attacca, cioè non si possono effettuare saldature a stagno su parti di alluminio. Ma la realizzazione della piccola basetta rettangolare non è strettamente necessaria; essa è soltanto utile perchè semplifica le operazioni di cablaggio, cioè permette di realizzare il montaggio di una buona parte dei componenti in modo razionale e compatto. Anche in questo caso il successo dipende dalla precisione dei collegamenti, cioè da un cablaggio privo di errori, effettuato con saldature « calde ». L'applicazione dei due diodi al germànio DG1 e DG2 deve essere fatta tenendo conto, come al solito, della loro polarità; nello schema pratico, tuttavia, tale indicazione è chiaramente evidenziata dalla posizione della fascetta colorata di questi due componenti. Anche il riconoscimento dei terminali del transistore TR1 si effettua sempre con lo stesso metodo, tenendo conto della presenza del puntino colorato in corrispondenza del terminale di collettore (c). Il terminale di base (b) è ricoperto con un tubetto di plastica, allo scopo di isolare elettricamente il conduttore e in modo che esso non possa creare danni pur toccando il conduttore di emittore o quello di collettore, perchè se tali fili dovessero toccarsi tra loro il ricevitore non funzionerebbe e la... salute del transistore TR1 verrebbe compromessa.

Tenete conto delle polarità dei diodi.

COSTRUZIONE DELLA BOBINA

Nei precedenti montaggi la bobina di sintonia era di tipo commerciale, cioè la si poteva acquistare già bell'e pronta in commercio. Per questo tipo di ricevitore reflex è stato adottato un sistema diverso, con lo scopo di mettere l'allievo nelle condizioni di realizzare, per la prima volta, con le proprie mani, una bobina di sintonia. A tale scopo occorrerà acquistare in commercio quel bastoncino cilindrico visibile nello schema pratico. In termini radiotecnici, quel bastoncino cilindrico prende i nomi di « nucleo di ferrite » o « nucleo ferroxcube ». In commercio se ne vendono di diversi tipi e forme, ma quello necessario per il presente montaggio deve avere una lunghezza di 140 millimetri e un diametro di 8 millimetri; si tratta di misure standard di facile reperibilità in commercio. Il nucleo ferroxcube è di colore metallico scuro ed è molto fragile; se esso cade per terra si spezza e va in frantumi; ciò significa che l'allievo dovrà manipolare tale componente con una certa attenzione, in modo da non lasciarselo mai sfuggire dalle mani e farlo cadere.

Arrestate i terminali degli avvolgimenti con nastro adesivo.

Il primo avvolgimento da effettuarsi è quello indicato con la sigla L 1. Per ottenerlo occorre acquistare non meno di 2 metri di filo di rame smaltato di diametro 0,2 mm. A partire da una distanza di 25 - 30 mm. circa da una delle due estremità del nucleo di ferrite, si inizierà l'avvolgimento L 1, avvolgendo con le mani 70 spire di filo compatte; ciò significa che le spire dovranno essere aderenti l'una all'altra, senza alcuno spazio vuoto. I terminali dell'avvolgimento L 1 verranno bloccati con due pezzettini di nastro adesivo, in modo da impedire che l'avvolgimento si sfili. L'avvolgimento L 2 non è effettuato direttamente sul nucleo. Su questo cilindretto, di facile realizzazione, si avvolgeranno quattro spire compatte di filo dello stesso tipo. Se l'allievo non riuscisse a trovare in commercio il filo di rame smaltato, potrà acquistare del filo di rame coperto di seta, purché il diametro sia sempre quello di 0,2 mm. (un diametro leggermente inferiore o superiore ai 0,2 mm. non compromette l'esito del lavoro). Nel disegno rappresentativo dello schema pratico è stata riportata una freccia in corrispondenza dell'avvolgimento L 2; questa freccia sta a significare che l'avvolgimento L 2 deve poter scorrere nei due sensi sul nucleo di ferrite; ciò serve per la messa a punto del ricevitore; infatti, a lavoro ultimato, dopo aver sintonizzato il ricevitore sulla emittente locale, si proverà a spostare in avanti e all'indietro l'avvolgimento L 2 e lo si bloccherà, con un pezzetto di nastro adesivo, in quel punto in cui la ricezione è più forte.

Si può evitare di collegare in maniera definitiva, per mezzo di saldature a stagno, i morsetti della pila al circuito, facendo uso delle apposite prese polarizzate, nelle quali il terminale positivo e quello negativo si distinguono tra loro per la diversa colorazione dei conduttori.

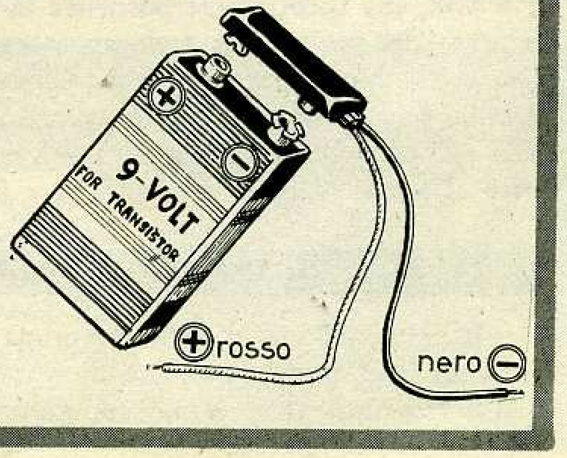

RICEVITORE REFLEX con AMPLIFIC. B.F.

I progetti dei ricevitori fin qui presentati sono tutti muniti di un solo stadio amplificatore. In altre parole i segnali radio sono sempre stati sottoposti ad un processo di rinforzo mediante un solo transistore. È pur vero che nei secondi due progetti si è fatto « lavorare » il transistore con il massimo sfruttamento; nel ricevitore a reazione il transistore amplifica più volte lo stesso segnale di alta frequenza; nel circuito reflex il transistore amplifica i segnali di alta frequenza e quelli di bassa frequenza; ma non è possibile con un solo transistore far funzionare un altoparlante, perchè l'ascolto in altoparlante richiede una certa potenza di segnali.

Con l'aggiunta di un secondo transistore (TR 2) che « pilota » uno stadio di amplificazione di bassa frequenza è possibile far aumentare di molto la corrente di bassa frequenza che... contiene le voci e i suoni.

Tutta la parte iniziale del ricevitore (l'allievo faccia riferimento allo schema elettrico) è uguale a quella del precedente progetto. Sul conduttore che nel precedente progetto arrivava alla cuffia è stato applicato il potenziometro R 3 (resistenza variabile). La pila, quando si agisce sull'interruttore S 1 per chiudere il circuito, invia la sua corrente a tutta la parte iniziale del ricevitore attraverso il potenziometro R 3. Quindi sul potenziometro R 3 è presente la tensione rivelata di bassa frequenza; essa può essere prelevata nella quantità desiderata per mezzo del cursore del potenziometro stesso. Tale tensione viene applicata, tramite il condensatore C 6, alla base (b) del transistore amplificatore di bassa frequenza TR 2. Ma per far funzionare questo transistore occorre applicare sul suo collettore (c) la tensione di 9 volt erogata dalla pila. Non è possibile, tuttavia, applicare direttamente sul collettore di TR 2 la tensione di 9 volt, ma occorre interporre fra la pila e il collettore di TR 2 un elemento dal quale poter prelevare i segnali di bassa frequenza. Questo elemento è rappresentato dall'avvolgimento primario del trasformatore d'uscita T 1. Attraverso questo avvolgimento fluisce la corrente erogata dalla pila e quella rappresentativa del segnale di bassa frequenza, amplificato al punto di poter far funzionare un altoparlante.

Simbolo elettrico del trasformatore d'uscita e (sotto) disegno del componente reale.

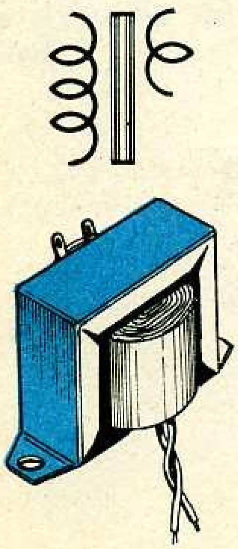

IL TRASFORMATORE D'USCITA

Il trasformatore d'uscita è, prima di tutto, un trasformatore di corrente; esso prende il nome di « trasformatore d'uscita » perchè in tutti i ricevitori radio viene applicato all'uscita del circuito, cioè fra

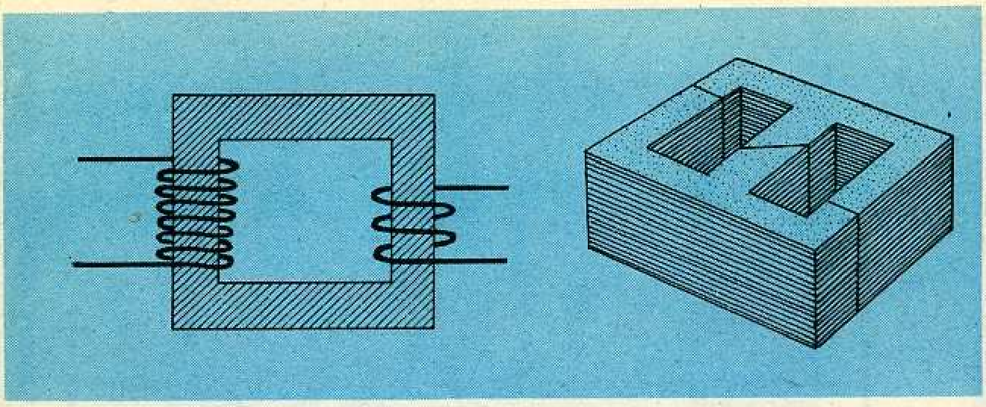

Il disegno a sinistra interpreta il concetto di trasformatore; i due avvolgimenti, che prendono il nome di « avvolgimento primario » e « avvolgimento secondario » risultano elettricamente isolati tra di loro. A destra è rappresentato il pacco lamellare con il quale si costruisce il trasformatore di uscita.

il transistore amplificatore finale e l'altoparlante; negli apparati a valvole esso viene collegato fra la valvola amplificatrice finale e l'altoparlante. Il trasformatore d'uscita è costituito da un pacco di lamierini di ferro sovrapposti, che prende il nome di NUCLEO; sul nucleo vengono avvolti due avvolgimenti di filo di rame; uno di questi due avvolgimenti prende il nome di « avvolgimento primario » del trasformatore: esso è costituito da un elevato numero di spire (qualche migliaio) di filo di rame sottile. L'altro avvolgimento, che prende il nome di « avvolgimento secondario », è costituito da un centinaio di spire di filo di diametro maggiore ed esso va collegato all'altoparlante, più precisamente alla bobina mobile di questo.

L'ALTOPARLANTE

L'altoparlante, come la cuffia, serve a trasformare le correnti di bassa frequenza in voci e suoni. Esso costituisce un dispositivo radioelettrico che prende il nome di « trasduttore elettroacustico ». L'altoparlante si compone, principalmente, di un cono di carta speciale soggetto a continue vibrazioni; esso prende anche il nome di « cono diffusore ». Le vibrazioni meccaniche del cono diffusore producono delle vibrazioni nelle masse d'aria che si trovano intorno al cono stesso, e le vibrazioni dell'aria altro non sono che suoni. Ma come fa a vibrare il cono di un altoparlante? Sul vertice del cono di carta vi è un cilindretto di cartone rigidamente fissato ad esso; su questo cilindretto è presente un avvolgimento di un solo strato di filo di rame smaltato sottile; questo cilindretto, con il suo avvolgimento, prende il nome di BOBINA MOBILE. I terminali della bobina fuoriescono dalla parte posteriore del cono e sono

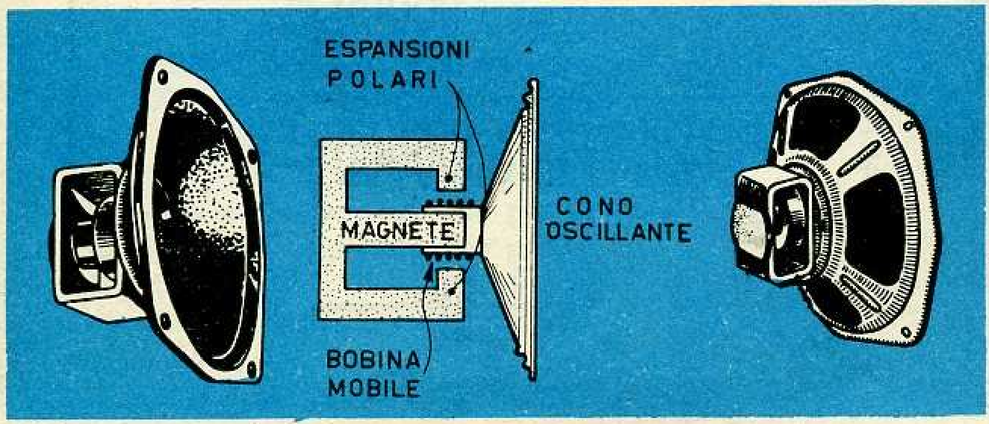

La bobina mobile dell'altoparlante (disegno centrale) è libera di muoversi longitudinalmente sul corpo centrale del magnete permanente.

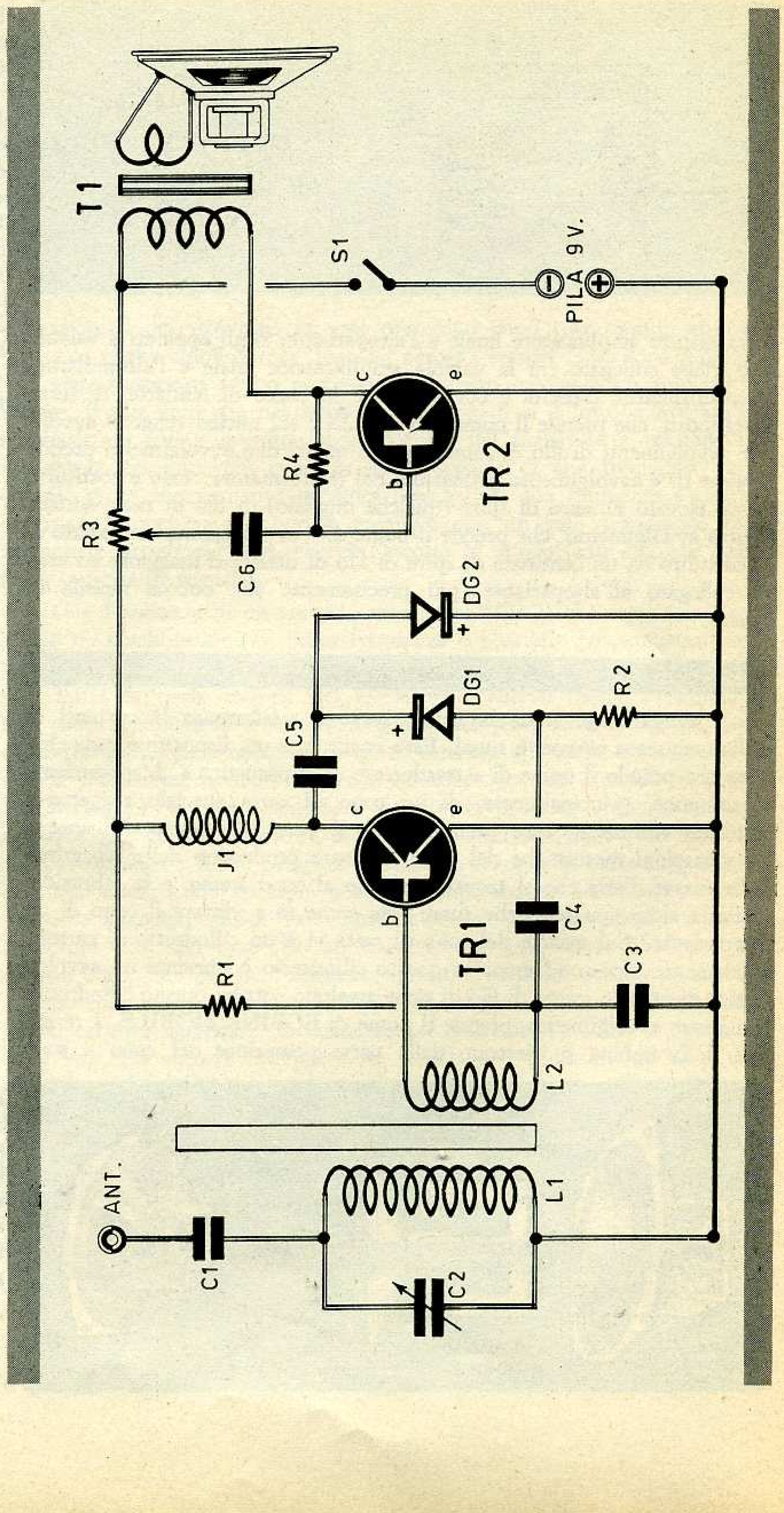
T1
S1
PILA 9 V.
c
e
b
TR 2
R4
R3
C6
DG 2
+
DG1
+
R 2
C5
c
e
b
J1
TR1
C4
C3
R1
L2
ANT.
L1
C1
C2

Schema elettrico del ricevitore reflex munito di stadio amplificatore di bassa frequenza.

Il ricevitore reflex può essere montato in un mobiletto di legno o di plastica per assumere il tipico aspetto di un ricevitore commerciale.

fissati ad una piastrina di bachelite o a due terminali metallici isolati. Su questi due terminali si applica la corrente di bassa frequenza presente all'uscita del ricevitore radio; più precisamente, su questi due terminali vengono saldati i due terminali dell'avvolgimento secondario del trasformatore d'uscita.

La bobina mobile è immersa fra le espansioni polari di un magnete permanente (calamita). Quando la bobina mobile dell'altoparlante è attraversata dalla corrente di bassa frequenza, essa produce all'intorno un campo elettromagnetico variabile, che contrasta con il campo magnetico generato dal magnete dell'altoparlante; da tale contrasto scaturisce la forza che imprime un movimento in avanti e all'indietro alla bobina mobile e quindi al cono di carta dell'altoparlante.

COMPONENTI DEL RICEVITORE REFLEX CON AMPLIFICATORE BF

C1 = 47 pF (condensatore in polistirolo)
C2 = 500 pF (condensatore variabile)
C3 = 50.000 pF (condensatore a carta)
C4 = 100.000 pF (condensatore a carta)
C5 = 1.000 pF (condensatore in polistirolo)
C6 = 100.000 pF (condensatore a carta)
R1 = 1 megaohm (marrone-nero-verde)
R2 = 38.000 ohm (arancione-grigio-arancione)
R3 = 5.000 ohm (potenziometro con interruttore S1 incorporato)
R4 = 100.000 ohm (marrone-nero-giallo)
DG1-DG2 = diodi al germanio
L1-L2 = bobine di sintonia
J1 = impedenza di alta frequenza tipo Geloso 558
TR1 = OC44 (transistore tipo pnp)
TR2 = OC75 (transistore tipo pnp)
T1 = trasformatore d'uscita (vedi testo)
S1 = interruttore incorporato con R3
pila = 9 volt

MONTAGGIO DEL RICEVITORE con AMPLIFIC. B.F.

La maggior parte dei componenti utilizzati nel precedente montaggio servono ora per la realizzazione pratica del ricevitore reflex dotato di stadio amplificatore finale. L'allievo dovrà acquistare soltanto il potenziometro R 3, la resistenza R 4, il condensatore C 6, il transistore TR 2, il trasformatore d'uscita T 1, l'altoparlante e alcune viti e dadi.

All'atto dell'acquisto di questo materiale si dovrà chiedere, al negoziante, un potenziometro da 5.000 ohm, a variazione logaritmica, munito di interruttore (S 1), un transistore di tipo OC 75, un condensatore a carta da 100.000 pF, una resistenza da 100.000 ohm, un trasformatore d'uscita per push-pull di OC 72 e un altoparlante di diametro 70 - 80 mm. di impedenza identica a quella dell'avvolgimento secondario del trasformatore d'uscita T 1. Il trasformatore d'uscita per push-pull di OC 72 è dotato di un avvolgimento primario e di un avvolgimento secondario; l'avvolgimento primario è munito di tre conduttori uscenti, ma l'allievo dovrà utilizzare soltanto due di questi conduttori, lasciando inutilizzato quello centrale; uno di questi due conduttori va collegato al collettore del transistore TR 2, l'altro va collegato all'interruttore S 1, come si vede nello schema pratico. L'interruttore S 1 è incorporato nel potenziometro R 3. Il montaggio del ricevitore si effettua su una tavoletta di legno di forma rettangolare, oppure su una basetta di bachelite. E poiché questo montaggio ripete, nella massima parte, il precedente montaggio, è inutile ripetere tutte le operazioni di cablaggio già ampiamente descritte. Il trasformatore di uscita T 1 viene saldato direttamente sul castello metallico dell'altoparlante, mentre lo stesso altoparlante è fissato alla tavoletta di legno o alla basetta di bachelite mediante alcune linguelle metalliche fissate sulla basetta con vite e dado.

Montatelo su una tavoletta di legno o su una basetta di bachelite di forma rettangolare.

OPERAZIONI DI TARATURA

Le operazioni di taratura sono le stesse già descritte per il precedente montaggio. Intervenendo sul perno del condensatore variabile C 2 si sintonizza il ricevitore sulla emittente voluta; spostando leggermente, verso destra o verso sinistra, lungo il nucleo di ferrite, l'avvolgimento L 2, si riesce a tarare il ricevitore in modo da raggiungere la massima sensibilità (massima potenza sonora nell'altoparlante). Intervenendo sul perno del potenziometro R 3, al quale si applicherà una manopola di comando, si riesce ad accendere o spegnere il ricevitore, cioè a chiudere o ad aprire il circuito; ma con il potenziometro R 3 si riesce a effettuare un'altra importante operazione, quella di aumento o di abbassamento del volume sonoro del ricevitore. Durante le fasi di messa a punto e taratura del ricevitore, sarà bene mantenere il comando di volume al valore massimo.

Questo circuito non richiede il conduttore di terra e si può fare a meno anche dell'antenna se il montaggio viene effettuato dentro una scatolina di legno, di plastica o altro materiale isolante. Utilizzando un mobiletto metallico, sarà necessario servirsi ancora dell'antenna.

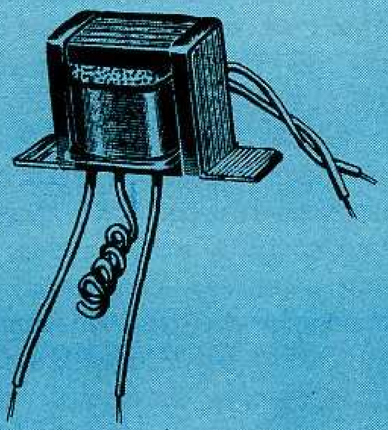

Dal trasformatore d'uscita fuoriescono cinque conduttori; due di questi sono ottenuti con filo di rame smaltato di maggior sezione e rappresentano i terminali dell'avvolgimento secondario. L'avvolgimento primario è munito di tre conduttori; di questi si dovrà lasciare inutilizzato il conduttore centrale.

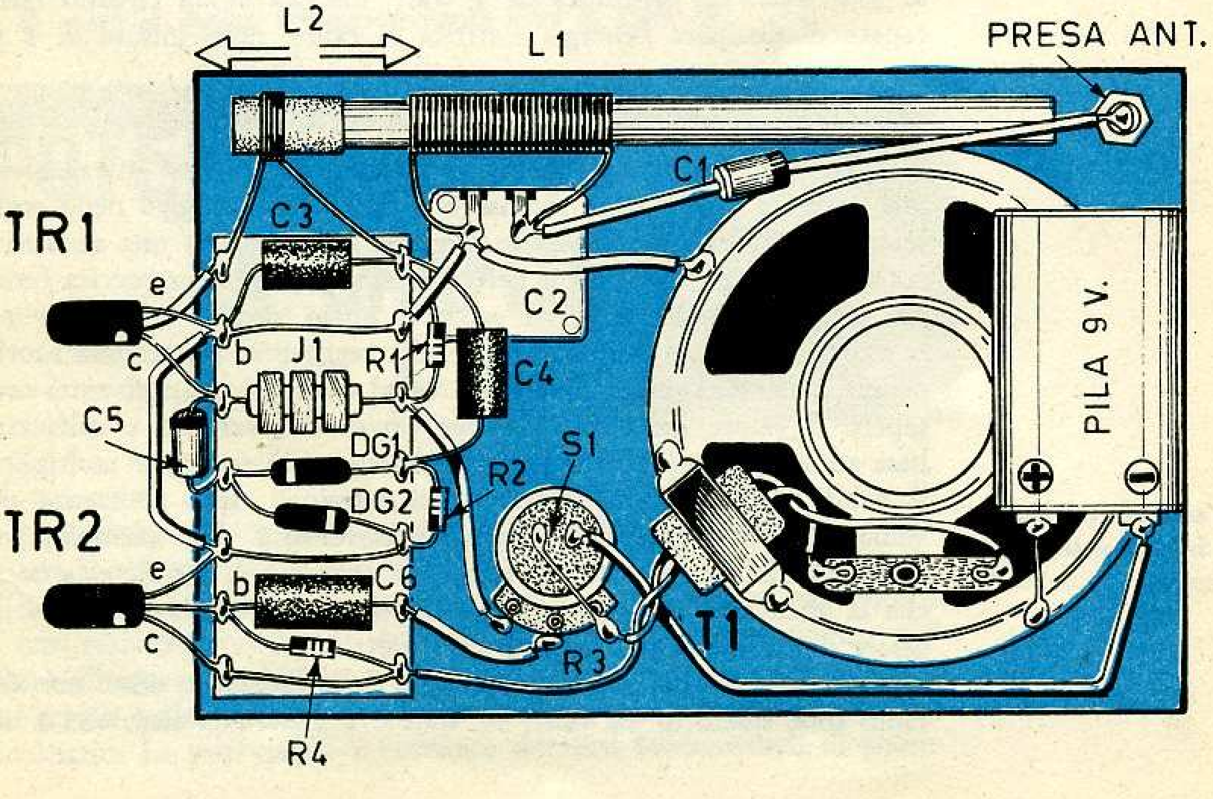

Schema pratico del ricevitore con circuito reflex dotato di stadio amplificatore finale.

POTENZA ELETTRICA delle RESISTENZE

Quando l'allievo si recherà presso il negoziante di prodotti radioelettrici per acquistare una resistenza da 1 megaohm, da 38.000 ohm, da 100.000 ohm, ecc., potrà sentirsi chiedere: « Da quanti watt »? È questa una domanda naturale di tutti i negozianti, quando il cliente si limita soltanto a dichiarare il valore ohmmico della resistenza richiesta, omettendo di specificare il wattaggio. Ma le espressioni watt e wattaggio sono nuove e devono essere accuratamente interpretate.

Il watt è l'unità di misura della potenza elettrica, così come lo è il volt per la tensione elettrica, l'ampere per la corrente e l'ohm per la resistenza. Ma che cos'è questa potenza elettrica? Anche qui si tratta di un concetto fondamentale, che l'allievo radiotecnico deve necessariamente assimilare. Essa sta ad indicare l'energia elettrica che un qualsiasi componente elettrico consuma trasformandola in altra forma di energia.

Watt (W.): unità di misura della potenza elettrica.

Quando si acquista una lampadina da 50 watt, ciò sta a significare che la potenza elettrica di quella lampadina è di 50 watt e che essa trasforma l'energia elettrica in energia luminosa e termica nella misura di 50 watt. Una resistenza elettrica da 350 watt, inserita in un ferro da stiro, trasforma l'energia elettrica in energia termica nella misura di 350 watt. La resistenza da 1 watt, inserita in un circuito radio, è capace di dissipare l'energia elettrica in calore nella misura di 1 watt.

LE RESISTENZE SI RISCALDANO

Tutte le resistenze inserite nei circuiti radio sono attraversate da una corrente più o meno intensa e a causa del passaggio della corrente esse si riscaldano; non si può tollerare, tuttavia, che una resistenza si scaldi troppo, prima di tutto perchè questa non è la sua precisa funzione nei circuiti radioelettrici, in secondo luogo perchè l'eccessivo calore potrebbe causare la « bruciatura » della resistenza, mettendola fuori uso definitivamente. Dunque, quando si va ad applicare una resistenza occorre sapere il valore della corrente che dovrà attraversarla e calcolare, in base ad essa, il valore della potenza elettrica. Ciò si ottiene moltiplicando il valore della tensione presente sui terminali della resistenza per il valore della corrente destinata ad attraversarla. Per esempio, se sui terminali di una resistenza si misura la tensione di 2 volt e la corrente che la attraversa è di 0,5 ampere, la potenza di quella resistenza dovrà essere di: 2 volt x 0,5 ampere = 1 watt.

L'eccessivo calore « brucia » le resistenze.

I valori più comuni del wattaggio delle resistenze usate nei circuiti radio sono quelli di ¼ watt, ½ watt e 1 watt. Nei ricevitori a transi-

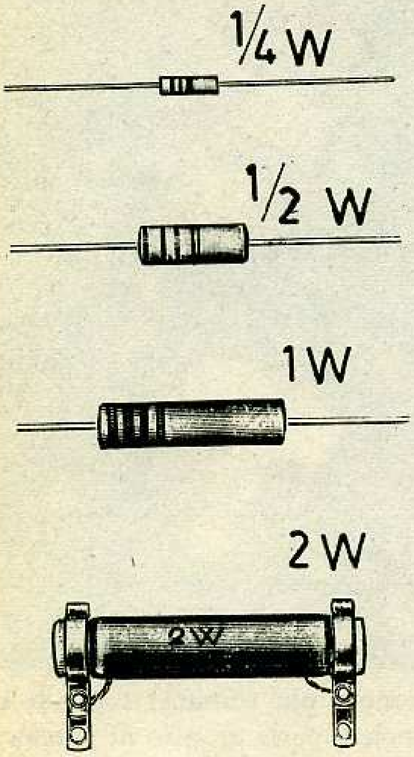

La potenza elettrica delle resistenze può essere approssimativamente valutata ad occhio nudo; le resistenze da 1/4 di watt sono molto piccole e vengono montate nei ricevitori a transistori e, più generalmente, nei circuiti miniaturizzati, nei quali le correnti elettriche in gioco sono molto deboli. Le resistenze di qualche unità di watt sono generalmente di tipo a filo. Le resistenze dei ferri da stiro (disegno riportato qui sotto) si aggirano intorno al centinaio di ohm e la loro potenza è, normalmente, di 350 watt.

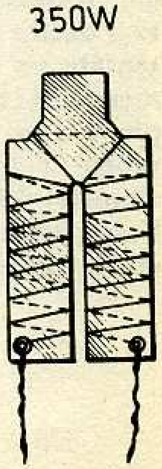

stori si usano resistenze da ¼ watt e da ½ watt; nei ricevitori a valvole si usano resistenze da ½ watt e da 1 watt e, talvolta, una o più resistenze di valore superiore a 1 watt; le resistenze di valore superiore a 1 watt non sono più di tipo chimico, bensì a filo.

Il watt si indica abbreviatamente con la sigla W.

POTENZIOMETRI A GRAFITE E A FILO

I potenziometri sono componenti radioelettrici che servono a regolare il volume sonoro e la tonalità della voce nell'apparecchio radio. Ma nella sua espressione radioelettrica il potenziometro altro non è che una resistenza variabile; dunque anche per il potenziometro può capitare di dover considerare la potenza elettrica, cioè il wattaggio, anche se ciò capita raramente. Ma per risolvere tale problema, l'industria radioelettrica ha provveduto a costruire due tipi fondamentali di poten-

I potenziometri usati in radiotecnica sono normalmente di due tipi: a strato di grafite (disegni in alto) e a filo (disegni in basso). In alto a destra è rappresentato un potenziometro di tipo doppio e con comando unico. I tre potenziometri disegnati in basso sono sprovvisti di interruttore.

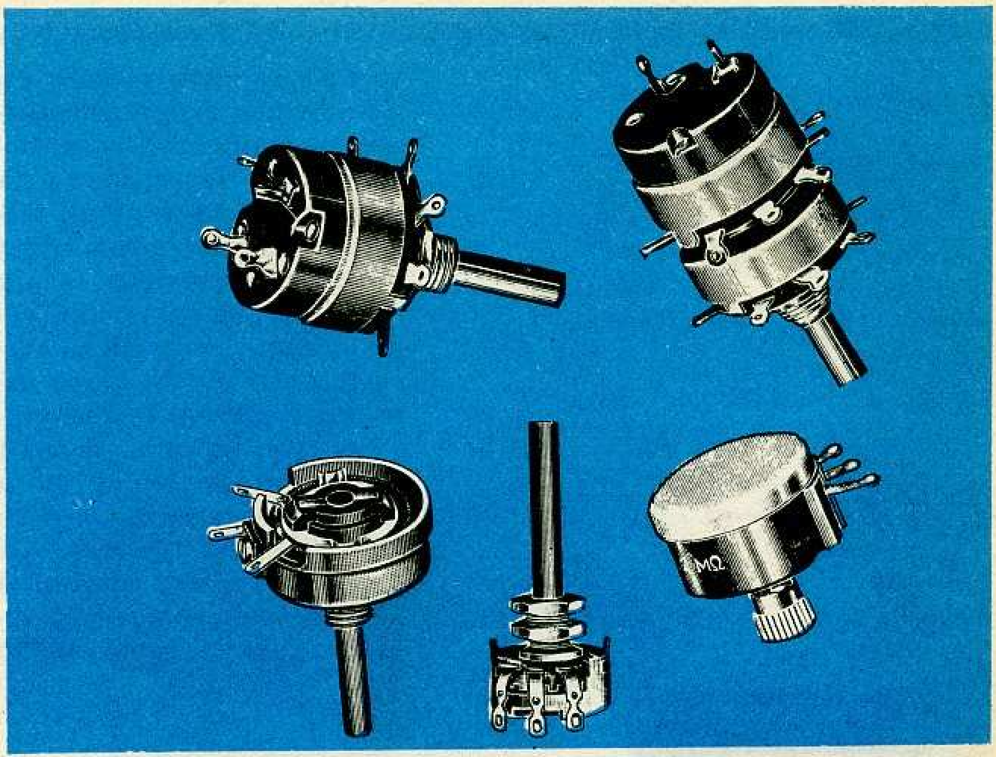

ziometri: quelli a strato di grafite (che sono i più comuni) e quelli a filo. Per i potenziometri a strato di grafite non si parla proprio di potenza elettrica o di wattaggio, perchè si suppone sempre che la corrente che li attraversa sia molto bassa; la resistenza di questi potenziometri è rappresentata da uno strato di grafite, sul quale scorre il cursore del potenziometro che fa capo al terminale di centro; i due terminali estremi fanno capo, invece, alle due estremità dello strato di grafite che è distribuito, internamente alla carcassa metallica del potenziometro, lungo una circonferenza. Quando il potenziometro è destinato ad essere attraversato da una corrente di una certa intensità, esso deve essere necessariamente di tipo a filo; in questo tipo di potenziometri esiste un avvolgimento di filo da resistenza distribuito lungo la circonferenza interna del potenziometro; il cursore scorre lungo questo avvolgimento e fa capo, come nel caso precedente, al terminale centrale del potenziometro; i due terminali estremi fanno capo ai due terminali estremi dell'avvolgimento di filo da resistenza.

VOLT - LAVORO DEI CONDENSATORI

I condensatori sono componenti radioelettrici che non si lasciano attraversare dalle correnti continue, mentre si lasciano attraversare dalle

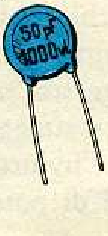

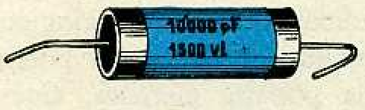

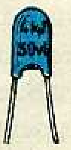

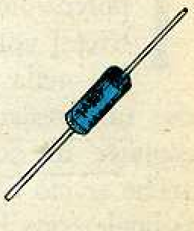

correnti alternate; ma le correnti alternate, come le correnti continue, possono essere più o meno intense e quindi il condensatore deve essere in grado di sopportare la corrente elettrica alternata senza rimaner danneggiato da questa. Si è detto che le resistenze elettriche devono avere un wattaggio proporzionato alla quantità di corrente che le attraversa per non rimanere « bruciate »; la stessa cosa avviene press'a poco nei condensatori, nei quali può accadere che si perfori il dielettrico, cioè lo strato di sostanza isolante interposto fra le lamine del condensatore stesso. Capita difficilmente, tuttavia, che il negoziante, all'atto dell'acquisto di un condensatore, chieda i volt-lavoro del condensatore, perchè le case costruttrici dei condensatori hanno provveduto a realizzare tali componenti per tensioni molto elevate, in modo da poter essere usati in molti tipi di circuiti senza alcuna preoccupazione per la tensione elettrica.

I condensatori a disco, ad esempio, vengono costruiti per una tensione di lavoro di 1000 volt, che è valore molto elevato e che si adatta alla maggior parte dei circuiti radio. La tensione di lavoro, invece, è necessaria per i condensatori elettrolitici, ma di ciò sarà detto più avanti.

LE VALVOLE ELETTRONICHE

Le valvole montate nei circuiti radio prendono il nome di valvole elettroniche proprio perchè, internamente ad esse, fluisce una corrente di elettroni. E chi non ha mai visto una valvola elettronica? Tutti voi certamente, se proprio non avete ancora avuto in mano una

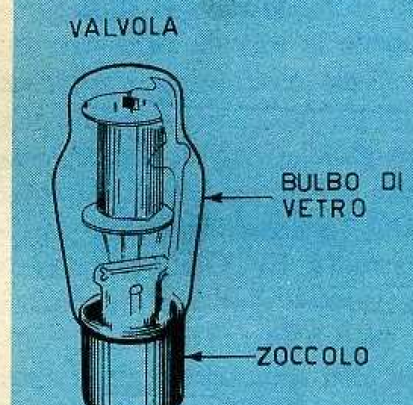

Il disegno a sinistra riproduce una valvola elettronica di vecchio tipo; sono visibili in essa lo zoccolo e i piedini, che risultano collegati agli elettrodi contenuti nel bulbo di vetro. Il disegno a destra riproduce, in spaccato, un moderno tipo di valvola elettronica.

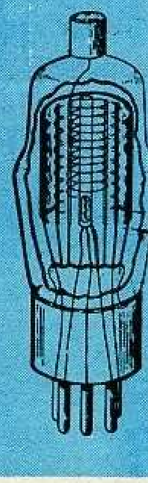

Circuiti elettrici... custoditi sotto vetro.

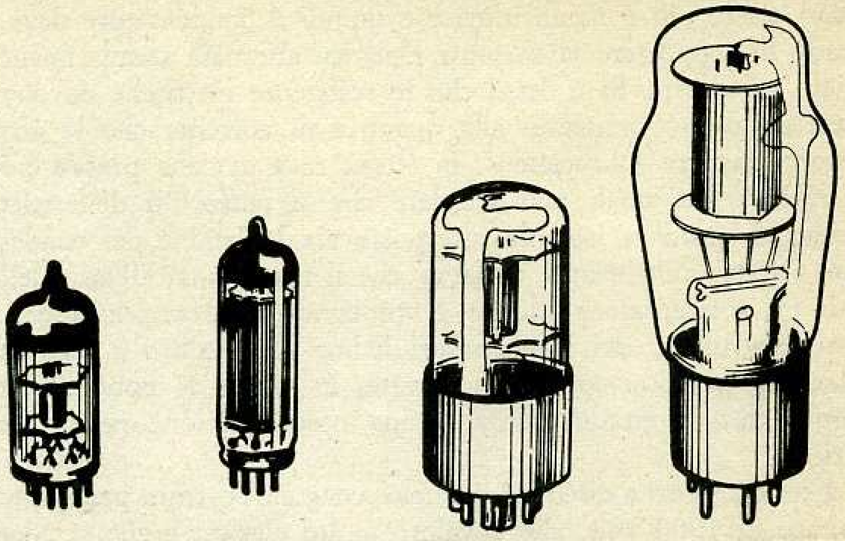

Il progresso della tecnica elettronica permette di costruire, oggi, valvole tuttovetro di piccolissime dimensioni (primi due disegni a sinistra). Le due valvole rappresentate a destra del disegno si riferiscono a tipi di vecchia costruzione.

valvola elettronica, l'avrete più di qualche volta adocchiata, curiosando dietro il vostro ricevitore radio. Essa si presenta come un cilindretto di vetro, dentro il quale, quasi sempre, si vede una lucetta rossa. Ma le valvole elettroniche vengono costruite oggi in molte maniere. Vi sono valvole di medie e di grandi dimensioni, come quelle impiegate nelle stazioni radiotrasmittenti. In ogni caso, la valvola elettronica racchiude con il suo involucro, che è quasi sempre di vetro, ma che può essere anche metallico, un certo numero di elementi che prendono il nome di ELETTRODI. Si tratta di elementi metallici che si presentano come fili o come lamierini e che sono sostenuti e isolati tra loro da foglietti di mica (materiale isolante e insensibile al calore). Più grande è il numero di elettrodi, contenuti nella valvola, e più complessa è la valvola stessa. Le valvole di costruzione moderna sono sempre più o meno complesse, e quasi sempre sono comprensive di due o più valvole semplici. Insomma, dentro uno stesso bulbo di vetro l'industria attuale è riuscita a racchiudere due e più valvole di tipo semplice, che un tempo rappresentavano una valvola sola. Ogni elettrodo, contenuto dentro la valvola, fa capo, all'esterno, ad alcuni bastoncini metallici che prendono il nome di PIEDINI. I piedini possono essere disposti in modo diverso sulla parte inferiore del bulbo e tutti assieme compongono lo zoccolo della valvola. Il numero di elettrodi, contenuto nella valvola elettronica, attribuisce alla valvola stessa il nome di battesimo. Se gli elettrodi sono due soltanto, la valvola prende il nome di DIODO; se gli elettrodi sono tre, essa prende il nome di TRIODO. Coll'aumentare del numero degli elettrodi le valvole prendono i nomi di: TETRODO (quattro elettrodi); PENTODO (cinque elettrodi); ESODO (sei elettrodi); EPTODO (sette elettrodi); OPTODO (otto elettrodi); ecc. Le valvole multiple prendono i nomi di: DOPPIO DIODO - DOPPIO TRIODO - TRIODO PENTODO ecc. Anche le valvole, come tutti gli altri componenti elettronici, vengono rappresentate negli schemi elettrici mediante simboli che, in pratica, sono dei cerchietti contenenti all'interno un certo numero di trattini, e ciascun trattino sta a simboleggiare un elettrodo della valvola.

Il numero degli elettrodi contenuti nella valvola elettronica attribuisce alla valvola stessa il nome di battesimo.

FUNZIONAMENTO DELLE VALVOLE

Qualunque corpo metallico riscaldato, come ad esempio un ferro arroventato, risulta avvolto da una « nube » di elettroni. Si tratta di un fenomeno ben noto in elettrotecnica, che prende il nome di fenomeno termoelettrico. Questi elettroni escono dalla superficie del corpo riscaldato, vagano ad una certa distanza da esso e in esso ricadono per liberarsi poi nuovamente. La nube di elettroni, quindi, non è statica, ma è formata da un continuo movimento di elettroni, che entrano ed escono dal corpo riscaldato. Nelle valvole vi è un particolare elettrodo, che prende il nome di CATODO, che viene riscaldato e che provvede all'emissione di elettroni. Per ottenere il riscaldamento del catodo si fa passare una corrente elettrica relativamente intensa attraverso un filo, del tipo di quelli usati nei fornelli e nelle stufe elettriche; questo filo, al passaggio della corrente diviene rosso; esso prende il nome di FILAMENTO. Quando si guarda nella parte posteriore di un apparecchio radio a valvole e si vedono quelle luci rosse internamente alle valvole, esse sono proprio le luci emesse dal filamento acceso. Il filamento, quindi, non svolge alcun compito elettrico che interessi il funzionamento vero e proprio della valvola; ad esso è affidato il solo compito di fungere da generatore di calore. Quando il catodo viene riscaldato dal filamento, il catodo emette una certa quantità di elettroni, che rimarrebbero attorno ad esso se non intervenissero delle forze elettriche atte a metterli in

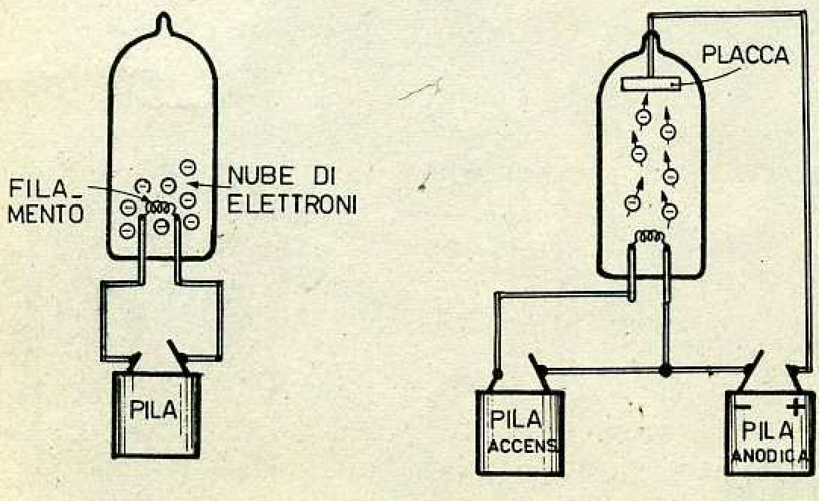

Il disegno a sinistra illustra il concetto di emissione di elettroni da parte del filamento. La pila eroga corrente e il filamento si riscalda al passaggio di essa avvolgendosi di una nube di elettroni. L'applicazione di una seconda pila (disegno a destra), con il morsetto positivo collegato alla placca, costringe gli elettroni a muoversi verso la placca stessa, dando origine alla corrente elettronica internamente alla valvola.

A sinistra è rappresentato il simbolo elettrico della valvola DIODO nella sua espressione più semplice.

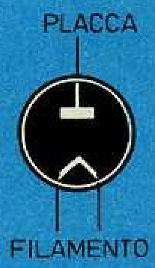

A destra il simbolo di un diodo munito di catodo. Nel primo caso il filamento funge anche da catodo, mentre nel secondo caso esso funge soltanto da elemento riscaldante.

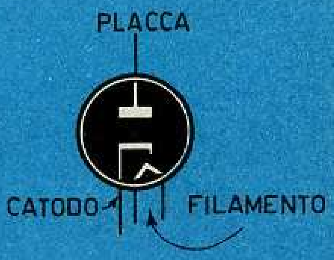

movimento. Gli elettroni sono cariche elettriche negative e nella valvola esiste un elettrodo, fondamentale, che prende il nome di PLACCA, al quale viene applicata una tensione positiva, che attrae a sé gli elettroni emessi dal catodo, determinando una corrente elettronica attraverso il vuoto; sì, perchè dentro il bulbo della valvola tutta l'aria è stata eliminata ed esiste un vuoto spinto.

DIODO

La valvola di tipo più semplice è costituita soltanto da un filamento e da una placca, e prende il nome di diodo. In questo caso il filamento funge anche da catodo; ma esistono anche diodi che contengono tutti e tre gli elettrodi essenziali: filamento, catodo e placca. Le valvole sprovviste di catodo sono chiamate « valvole a riscaldamento diretto ». Le valvole provviste di catodo vengono chiamate « valvole a riscaldamento indiretto ». Ma i diodi provvisti di catodo, pur essendo muniti di tre elettrodi, conservano il nome di diodi; infatti, gli elettrodi fondamentali, quelli che interessano l'impiego della valvola sono due: il catodo e la placca.

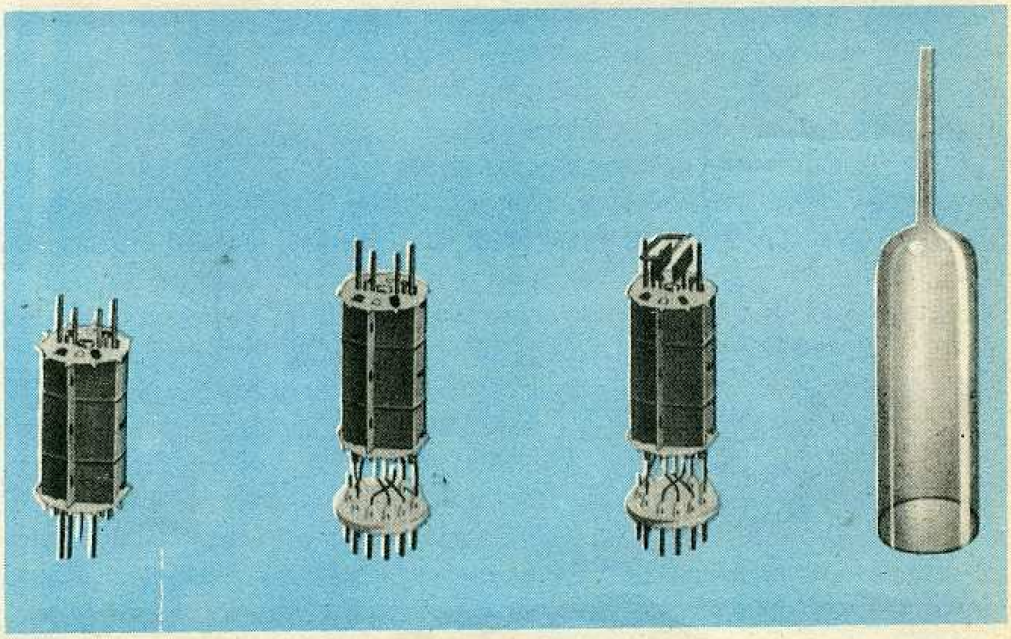

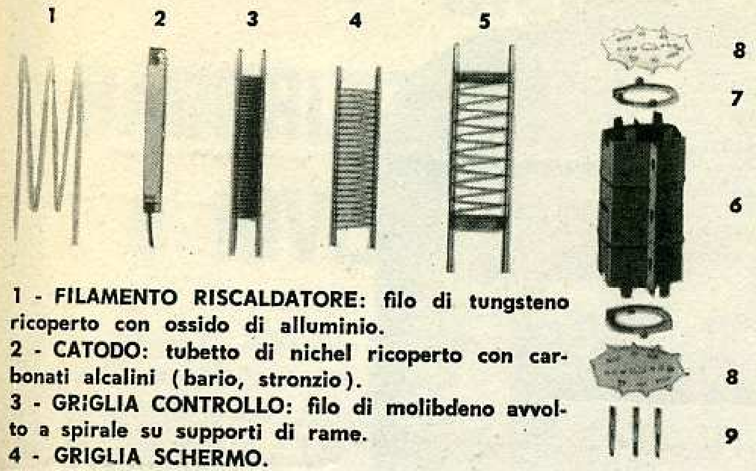

1 - FILAMENTO RISCALDATORE: filo di tungsteno ricoperto con ossido di alluminio.
2 - CATODO: tubetto di nichel ricoperto con carbonati alcalini (bario, stronzio).
3 - GRIGLIA CONTROLLO: filo di molibdeno avvolto a spirale su supporti di rame.
4 - GRIGLIA SCHERMO.
5 - GRIGLIA SOPPRESSORE.
6 - ANODO: ferro placcato con alluminio e successivamente annerito per migliorare la irradiazione del calore.
7 - SCHERMI PER I PONTI DI MICA: nichel.
8 - PONTI DI MICA: ricoperti con ossido di magnesio.
9 - NASTRINI DI COLLEGAMENTO: rame nichelato.

Elettrodo (in tre pezzi)

Elettrodo: nichel.

Passante: lega ferro-nichelato con rivestimento in rame.

Piedino: nichel.

Fondello

Radiatori (ferro ricoperto con alluminio annerito)

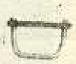

Getter

Il sistema elettrodico viene disposto sul « fondello » formato da un dischetto di vetro dove si trovano i « piedini »; indi si collegano, mediante saldatura, le parti terminali degli elettrodi con i « piedini » stessi; si applica su di un supporto del metallo (bario) ed infine si inserisce tutto il sistema elettrodico completato in un tubo di vetro (bulbo o palloncino) nel quale si crea il vuoto.

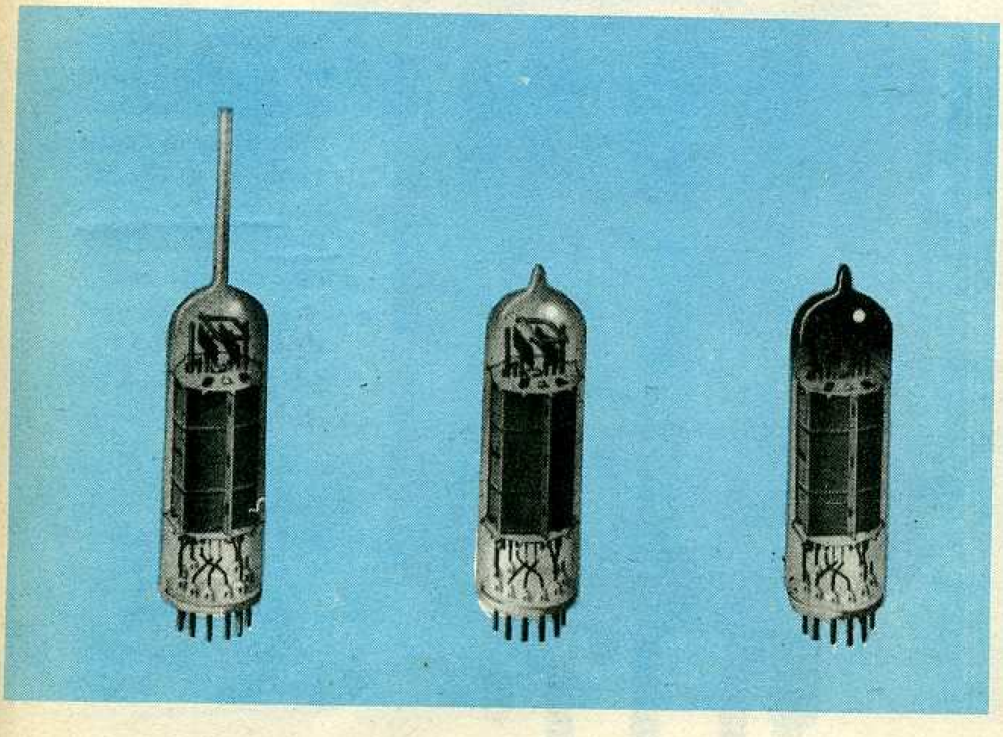

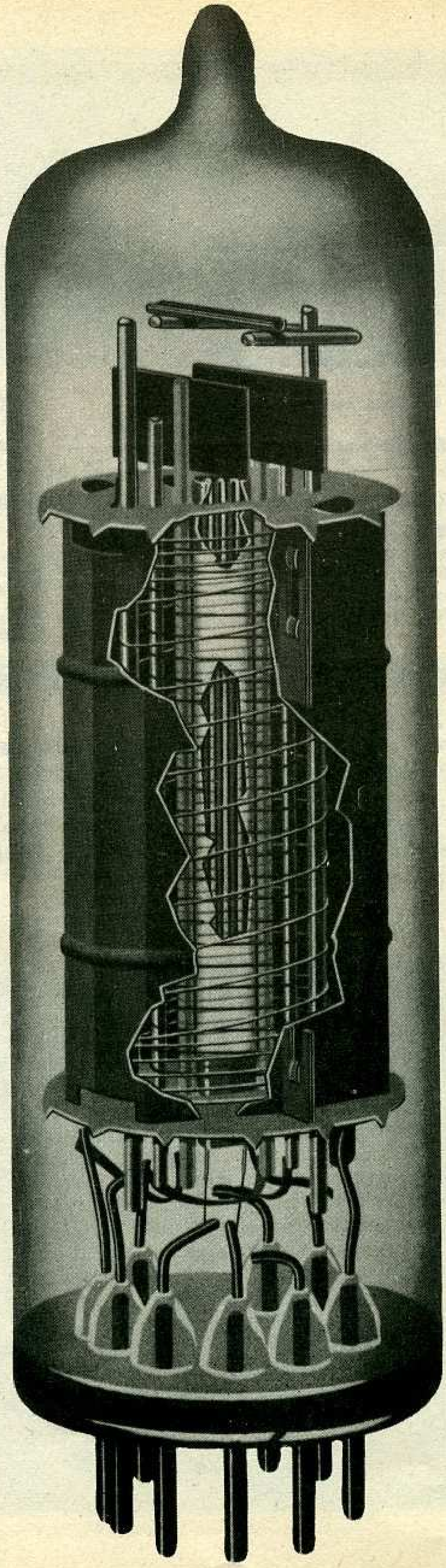

La figura illustra un moderno tipo di valvola elettronica vista in spaccato. Si tratta di un moderno pentodo impiegato nei circuiti radio in qualità di valvola amplificatrice dei segnali di bassa frequenza.

DIODO RADDRIZZATORE e RIVELATORE

La valvola di tipo più semplice è costituita soltanto da un filamento e da una placca: essa prende il nome di DIODO. Nei circuiti radio le valvole a due elettrodi, cioè diodi, vengono impiegate per svolgere due compiti diversi e fondamentali. Il primo compito è quello di « raddrizzare » le correnti alternate, cioè trasformare le correnti alternate in correnti pulsanti, e in questo caso la valvola prende il nome di DIODO RADDRIZZATORE. Il secondo compito, altrettanto importante, cui sono adibiti i diodi, è quello di rivelare i segnali radio, in questo caso il diodo prende il nome di DIODO RIVELATORE (esso sostituisce il diodo al germanio).

a)
b)

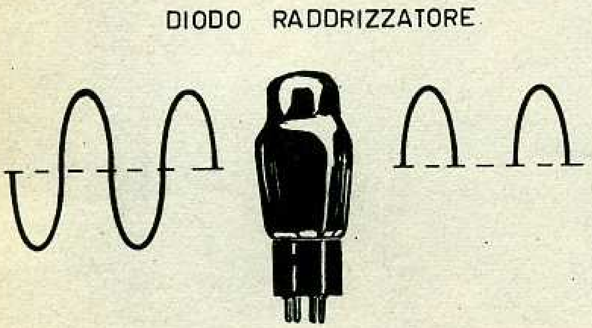

La valvola raddrizzatrice può essere un diodo semplice o un doppio-diodo; attraverso essa la corrente alternata si trasforma in corrente pulsante; praticamente questa valvola elimina le semionde negative della corrente alternata.

TRIODO

Nella valvola denominata triodo vi sono tre elettrodi: il catodo, la placca e la griglia; questo terzo elettrodo viene anche chiamato GRIGLIA CONTROLLO o GRIGLIA PILOTA; a tale elettrodo è affidato il compito di regolare il flusso elettronico internamente alla valvola, cioè la quantità di elettroni che, partendo dal catodo, raggiungono la placca; la griglia controllo si comporta, dunque, come un rubinetto che permette di regolare il flusso dell'acqua. Questa regolazione si ottiene applicando alla griglia una debole tensione negativa, perchè la tensione negativa, a seconda del suo valore, si oppone più o meno energicamente al movimento degli elettroni fra il catodo e la placca. Il compito principale, cui sono stati adibiti i triodi nei circuiti radio, è quello di amplificare i segnali radio, così come fanno i transistori.

TETRODO E PENTODO

Il tetrodo è una valvola dotata di quattro elettrodi; ai tre elettrodi già noti si aggiunge un quarto elettrodo che prende il nome di GRIGLIA SCHERMO. La griglia schermo rappresenta un elemento che permette alla valvola tetrodo di amplificare maggiormente i segnali radio. Nella valvola pentodo il quinto elemento prende il nome di GRIGLIA SOPPRESSORE. A tale quinto elettrodo si applica, normalmente, la tensione zero oppure una tensione leggermente negativa. Il compito della griglia soppressore è quello di ricacciare verso la placca quegli elettroni che, rimbalzando su di questa, non fossero stati attratti dalla placca stessa. I tetrodi e i pentodi rappresentano valvole largamente impiegate nei circuiti radio in qualità di valvole amplificatrici.

In ogni tipo di valvola il catodo è posto in posizione centrale; attorno ad esso risultano sistemate le griglie e, per ultima, la placca.

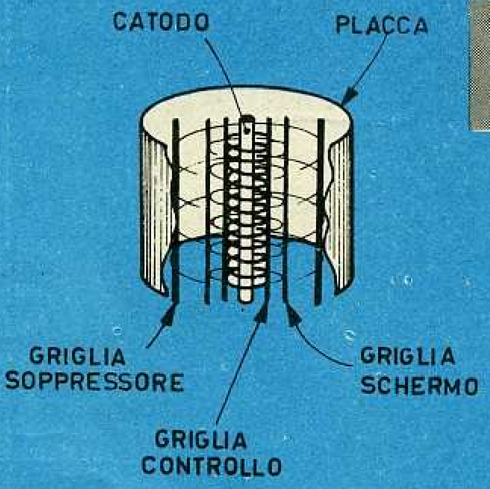
CATODO
PLACCA
GRIGLIA SOPPRESSORE
GRIGLIA SCHERMO
GRIGLIA CONTROLLO

Corrispondenza tra gli elementi simbolici di un triodo e i suoi componenti reali.

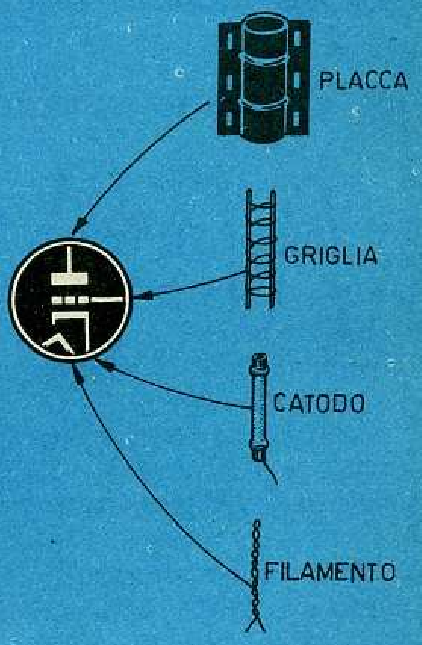
PLACCA
GRIGLIA
CATODO
FILAMENTO

Il triodo, impiegato come valvola amplificatrice, rinforza i segnali radio applicati alla griglia controllo; i segnali amplificati sono presenti all'uscita della valvola, cioè sulla placca.

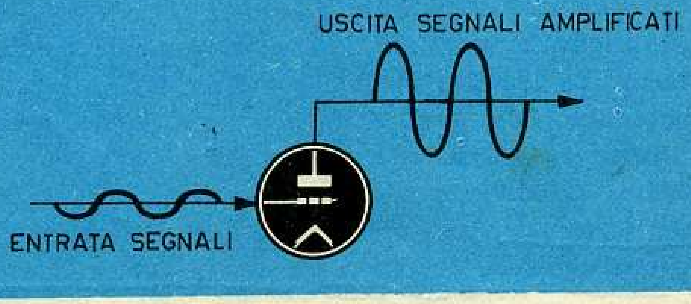
USCITA SEGNALI AMPLIFICATI
ENTRATA SEGNALI

RICEVITORE A UNA VALVOLA

Per poter costruire un ricevitore radio con una sola valvola bisogna pretendere dalla valvola stessa il massimo lavoro possibile; occorre cioè far svolgere alla valvola diversi compiti in una volta sola. La valvola V 1 rappresentata nello schema elettrico del ricevitore, pur essendo stata costruita per funzionare come pentodo amplificatore, viene sfruttata in questo caso come valvola rivelatrice e amplificatrice dei segnali di bassa frequenza. Ma passiamo all'analisi del circuito teorico, per poter comprendere meglio il funzionamento del ricevitore. Il circuito di sintonia è sempre lo stesso, quello già analizzato e montato nei circuiti a transistori, pur essendo stata impiegata, questa volta, una bobina di sintonia di tipo diverso, più precisamente la bobina Corbetta CS 2. Il segnale sintonizzato viene applicato, tramite il condensatore C 3 e la resistenza R 1, alla griglia controllo della valvola V 1 (piedino 1). Fra la griglia controllo e il catodo (piedino 7) fluisce la corrente di alta frequenza e questa esce dal catodo della valvola già rivelata; dunque, fra la griglia controllo e il catodo della valvola avviene il processo di rivelazione. La tensione rivelata è presente sui terminali della resistenza R 1, che prende anche il nome di resistenza di rivelazione. Il segnale rivelato viene poi amplificato, internamente alla valvola e viene prelevato dalla placca (piedino 5) ed applicato alla cuffia. Si dice che la cuffia svolge, in questo caso, il compito di « carico anodico » della valvola, perchè sui suoi terminali estremi è presente la tensione amplificata di bassa frequenza. Il potenziometro R 2 permette di regolare la tensione di griglia schermo (piedino 6) e in questo modo si riesce contemporaneamente a regolare il volume sonoro del ricevitore. Il condensatore C 5, collegato fra griglia schermo e massa, permette di livellare la tensione di griglia schermo. Il condensatore C 4 permette di fugare a massa quella parte di alta frequenza ancora presente nei segnali amplificati, impedendo che questa possa raggiungere la cuffia e manifestarsi sotto forma di fischio.

La cuffia funge da carico anodico.

ALIMENTATORE

I ricevitori radio fin qui costruiti erano tutti alimentati dalla pila. Ciò è stato possibile in quanto i transistori consumano poco; facendo impiego di una valvola il consumo della pila potrebbe essere eccessivo e la valvola stessa non potrebbe funzionare perchè essa richiede sulla placca una tensione superiore ai 100 volt. Meglio dunque ricorrere alla tensione della rete-luce, adattandola, per mezzo di un alimentatore, ai valori di tensione richiesti dal circuito radio. L'alimentatore si compone di un tra-

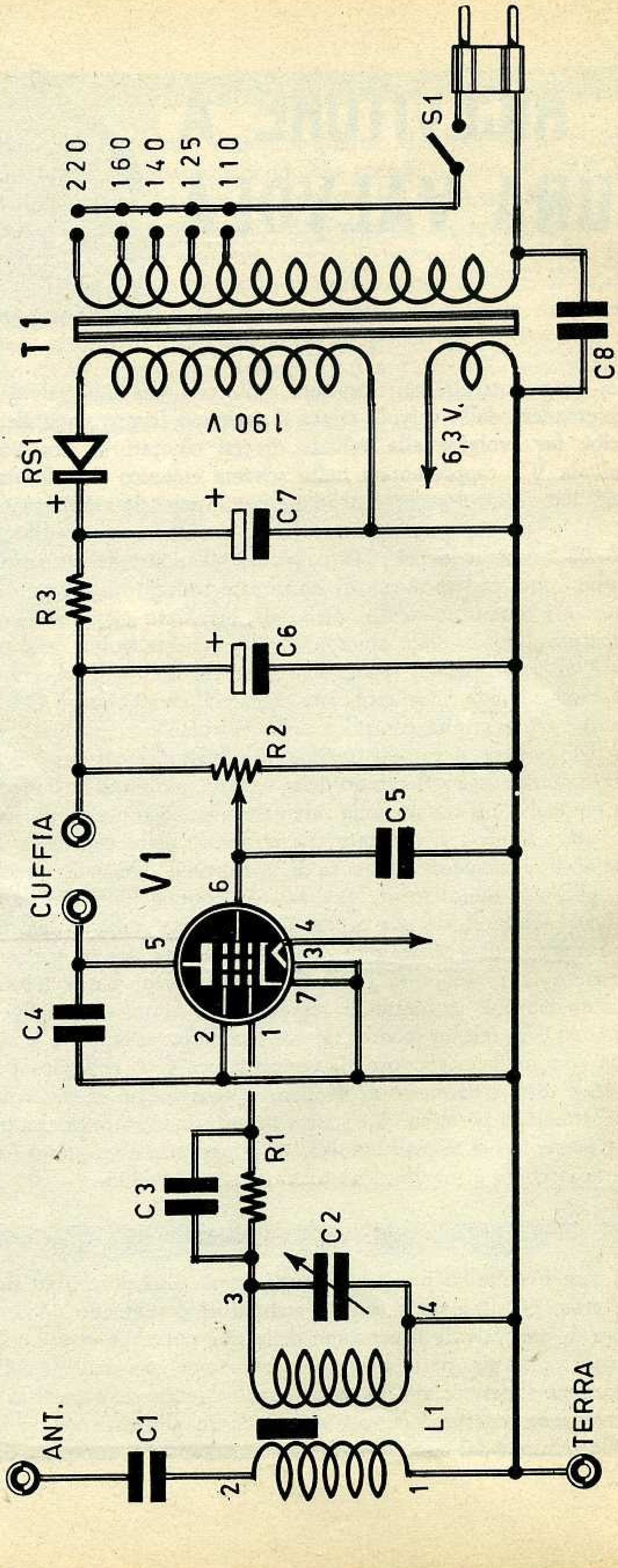

ANT.
C1
2
L1
1
TERRA
C3
R1
C2
3
4
C4
2
1
7
3
4
5
6
V1
CUFFIA
C5
R2
+
C6
R3
+
C7
+
RS1
190 V
6,3 V.
T1
C8
220
160
140
125
110
S1

Schema elettrico del ricevitore ad una valvola con ricezione in cuffia. L'alimentazione è ottenuta dalla rete-luce mediante un circuito alimentatore.

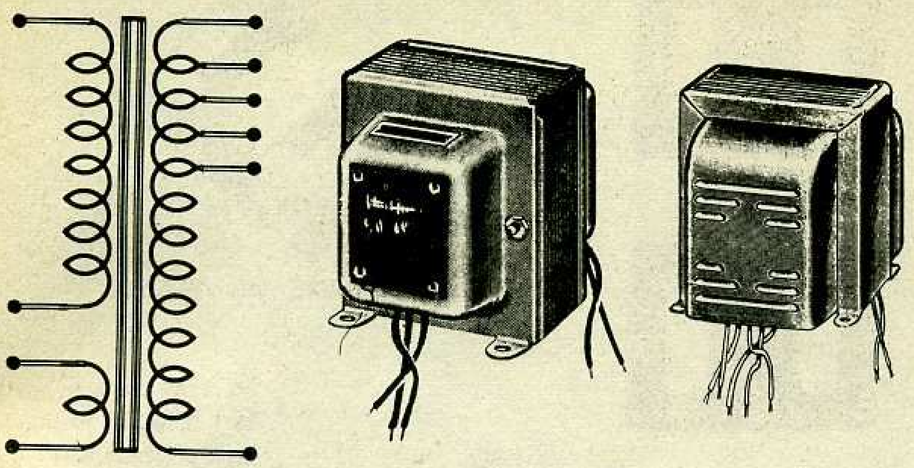

A sinistra è rappresentato lo schema elettrico di un normale trasformatore di alimentazione per ricevitori radio. I due disegni a destra riproducono due tipi di trasformatori; quello all'estrema destra prende il nome di trasformatore carrozzato, perchè completamente racchiuso in custodia metallica.

sformatore (T 1), di un raddrizzatore al selenio (RS 1), di un condensatore elettrolitico doppio a vitone (C 6-C 7), di una resistenza (R 3), di un condensatore (C 8) e di un interruttore (S 1), che in questo caso è incorporato nel potenziometro di controllo del volume sonoro del ricevitore (R 2).

TRASFORMATORE DI ALIMENTAZIONE

Il trasformatore di alimentazione, come il trasformatore d'uscita, è dotato di un avvolgimento primario, ma di due o più avvolgimenti secondari. L'avvolgimento primario risulta elettricamente isolato dagli avvolgimenti secondari. E poichè l'avvolgimento primario risulta direttamente collegato alla presa-luce, gli avvolgimenti secondari, che alimentano il ricevitore radio, determinano un isolamento completo del ricevitore dalla

COMPONENTI DEL RICEVITORE A1 VALVOLA

C1 = 100 pF (condensatore ceramico)
C2 = 500 pF (condensatore variabile)
C3 = 75 pF (condensatore ceramico)
C4 = 2.200 pF (condensatore ceramico)
C5 = 50.000 pF (condensatore a carta)
C6-C7 = 32 + 32 microfarad - 250-300 Vl. (condensatore elettrolitico doppio a vitone)
C8 = 10.000 pF - 1500 Vl. (condensatore a carta)
R1 = 2 megaohm (rosso-nero-verde)
R2 = 500.000 ohm (potenziometro a variazione logaritmica con interruttore S1 incorporato)
R3 = 2.200 ohm - 1 watt (rosso-rosso-rosso)
V1 = valvola tipo 6BA6
L1 = bobina di sintonia tipo Corbetta CS2
RS1 = raddrizzatore al selenio 250 volt - 50 mA
T1 = trasformatore di alimentazione tipo GBC H/188

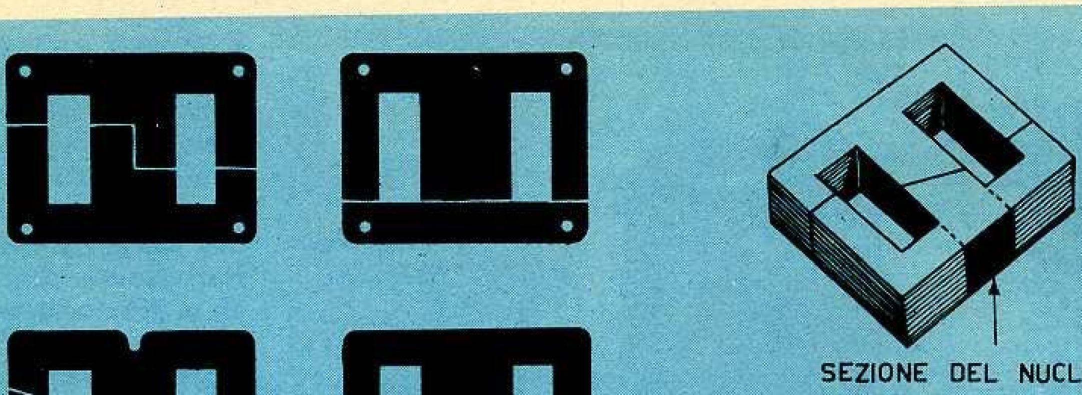

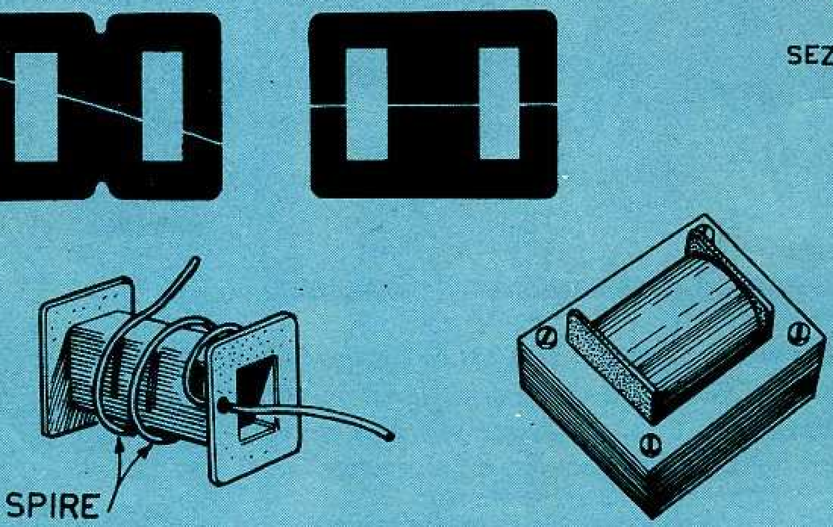

In alto sono riportati i disegni di alcuni tipi di lamierini per trasformatore di alimentazione; in basso, all'estrema sinistra, è disegnato il cartoccio, sul quale si effettuano gli avvolgimenti del trasformatore.

rete-luce. L'avvolgimento primario è dotato, generalmente, di sei terminali. Due di questi rappresentano gli estremi dell'avvolgimento; gli altri quattro costituiscono le prese intermedie, corrispondenti ai diversi valori delle tensioni di rete. In pratica, ad uno dei due conduttori che compongono il cordone di alimentazione viene collegato il terminale estremo dell'avvolgimento primario del trasformatore, mentre l'altro conduttore viene collegato alla presa centrale di un componente che prende il nome di CAMBIOTENSIONE, che permette di collegare il secondo conduttore del cordone di alimentazione sul terminale intermedio desiderato dell'avvolgimento primario. Nel trasformatore di alimentazione T 1 vi sono due avvolgimenti secondari; da uno di essi si preleva la tensione a 190 volt necessaria per alimentare il circuito anodico del ricevitore, cioè la placca e la griglia schermo della valvola; il secondo avvolgimento secondario, a 6,3 volt, serve soltanto per accendere il filamento della valvola V 1.

IL CAMBIOTENSIONE

Il cambiotensione costituisce in pratica un « ponte » di collegamento tra uno dei conduttori del cordone di alimentazione e una delle prese intermedie dell'avvolgimento primario del trasformatore. Con tale componente è assai semplice e rapido adattare l'avvolgimento primario del trasformatore di alimentazione al valore della tensione della rete-luce. In commercio esistono diversi tipi di cambiotensione, ma il principio di funzionamento è sempre lo stesso.

MONTAGGIO del RICEVITORE a UNA VALVOLA

Lo stesso ordine progressivo di montaggio dei ricevitori a transistori vale anche per i ricevitori a valvole. Quindi, anche per il montaggio di questo ricevitore monovalvolare l'allievo comincerà con l'eseguire tutte quelle operazioni che richiedono un lavoro di ordine meccanico. Si comincerà dapprima col preparare un telaio metallico (chassis) come quello indicato nel disegno dello schema pratico. Sul telaio si applicano il trasformatore di alimentazione T 1 nella parte superiore, il condensatore elettrolitico doppio a vitone C 6-C 7, il potenziometro R 2, il cambiotensione, il raddrizzatore al selenio RS 1, lo zoccolo portavalvola, le quattro

Il montaggio del ricevitore a una valvola è realizzato su telaio metallico, denominato anche con il termine francese « chassis ».

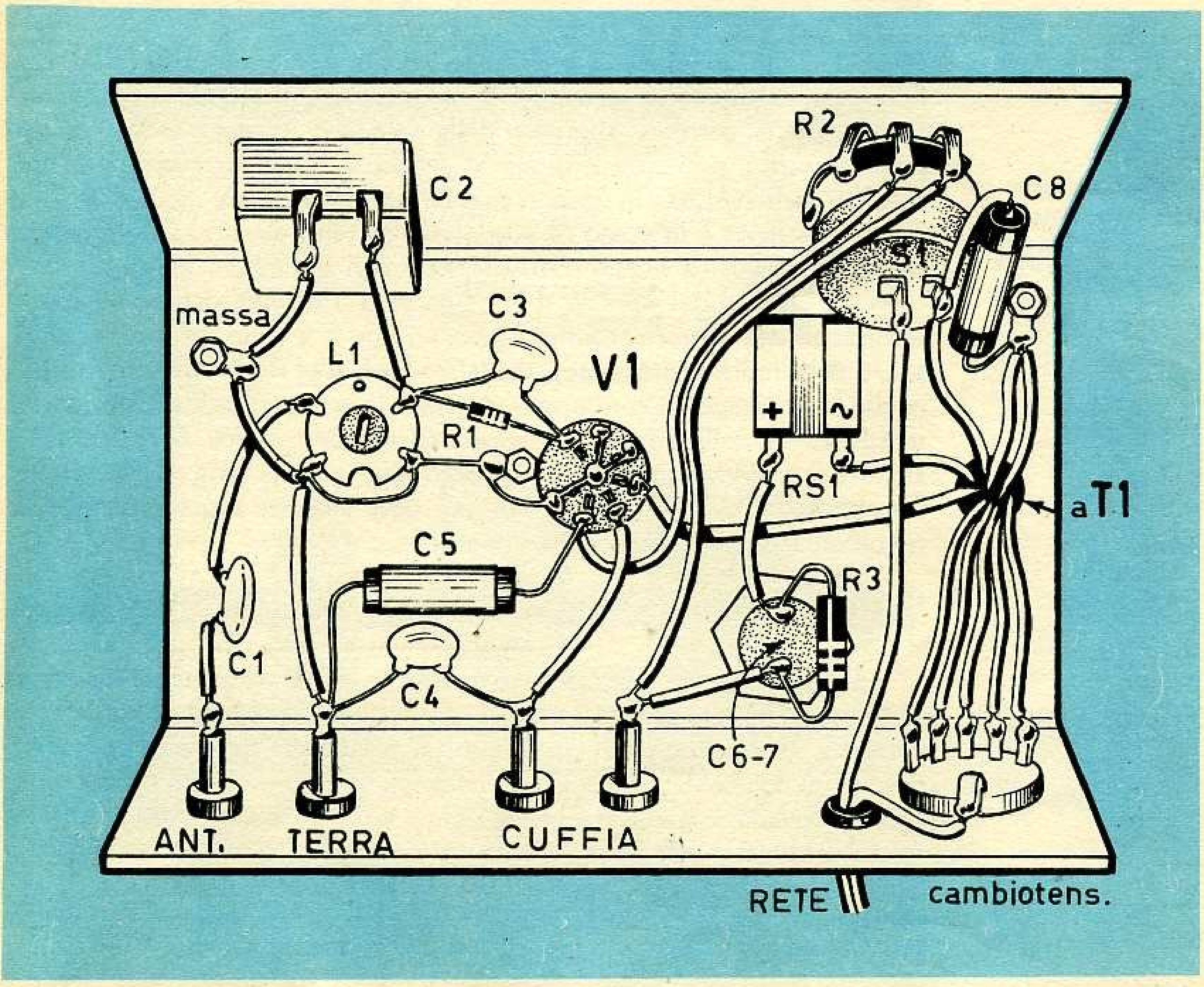

Il telaio funge da conduttore di massa.

boccole, la bobina L 1 e il condensatore C 2. Il telaio funge da conduttore di massa e ciò significa che tutti i telai dei ricevitori radio a valvole devono essere di metallo. Poichè risulta poco agevole effettuare le saldature a stagno direttamente sul telaio, conviene applicare in esso alcune viti con dado, interponendo un capocorda in funzione di ancoraggio di massa: su di esso si effettueranno le saldature dei terminali e dei componenti che devono essere collegati a massa; naturalmente la vite deve essere ben stretta, in modo da stabilire un perfetto contatto elettrico con il telaio. La possibilità di riconoscere fra i molti fili uscenti dal trasformatore di alimentazione quelli relativi all'avvolgimento primario e all'avvolgimento secondario, è concessa da uno speciale cartellino che si trova sempre inserito nel trasformatore stesso all'atto del suo acquisto. In questo cartellino è rappresentato lo schema elettrico del trasformatore e i valori delle tensioni, riportati sui terminali, trovano corrispondenza, nella realtà pratica, con i colori dei conduttori. Il conduttore a 220 volt, ad esempio, avrà un colore diverso da quelli a 160 volt, 140 volt, ecc. Anche i conduttori dei due avvolgimenti secondari hanno colori diversi; i due terminali a 190 volt hanno colore uguale e così accade anche per i due terminali dell'avvolgimento a 6,3 volt.

IL RADDRIZZATORE AL SELENIO

Trasformazione della corrente alternata in corrente pulsante.

Il raddrizzatore al selenio è un componente che si comporta un po' come il diodo al germanio e come la valvola a due elettrodi (diodo). A differenza del diodo al germanio, il raddrizzatore al selenio è destinato a raddrizzare la corrente alternata della rete-luce, trasformandola in corrente pulsante, e mentre il diodo al germanio viene attraversato da una corrente debolissima (in caso contrario andrebbe fuori uso), il raddrizzatore al selenio è in grado di sopportare correnti elevate. Esso è munito di due terminali: a quello contrassegnato con il simbolo indicatore della corrente alternata si applica il conduttore a 190 volt dall'avvolgimento secondario del trasformatore di alimentazione T 1 (uno dei due terminali a 190 volt, indifferentemente); dall'altro terminale del raddrizzatore RS 1, quello contrassegnato con la crocetta, esce la tensione raddrizzata; questo terminale va collegato con il condensatore elettrolitico C 7. L'applicazione del raddrizzatore RS 1 al telaio può essere fatta mediante una fascetta metallica che, ovviamente, non deve stabilire alcun contatto elettrico con i due terminali del componente.

IL CONDENSATORE ELETTROLITICO A VITONE

Il condensatore elettrolitico a vitone è un componente di forma cilindrica; l'intera carcassa metallica del condensatore, cioè il contenitore metallico costituisce, in ogni punto, il terminale di massa (negativo) comune ai due condensatori C 6-C 7 contenuti nel medesimo componente; quindi, quando si applica il condensatore elettrolitico, il collegamento di massa risulta stabilito automaticamente mediante il fissaggio del dado esagonale che blocca il componente sul telaio; i due terminali uscenti, rappresentano, invece, i due terminali positivi dei due condensatori elettrolitici da 32 μF ciascuno.

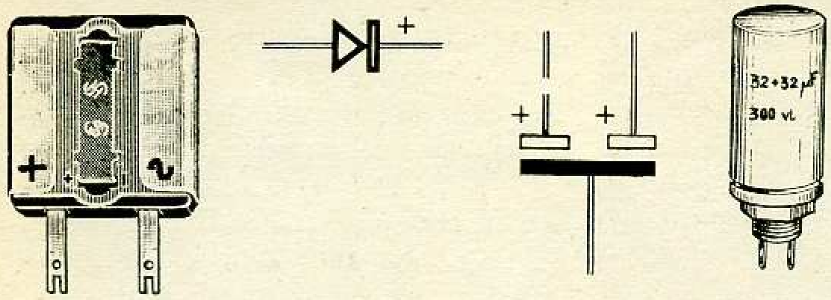

I due disegni a sinistra riproducono il raddrizzatore al selenio e il suo simbolo elettrico. Sui terminali del raddrizzatore sono riportati i simboli della tensione positiva uscente e di quella alternata entrante. I due disegni a destra riproducono il condensatore elettrolitico doppio a vitone e il suo simbolo elettrico.

CABLAGGIO DEL TRASFORMATORE DI ALIMENTAZIONE

Il cablaggio del ricevitore va fatto nel modo indicato nello schema pratico, e va iniziato con le saldature dei terminali dell'avvolgimento primario del trasformatore di alimentazione al cambiotensione. Per l'avvolgimento secondario si collegheranno a massa uno dei due terminali dell'avvolgimento a 6,3 volt e uno dei due terminali dell'avvolgimento a 190 volt, così come è indicato nello schema elettrico del ricevitore. Il terminale 0 volt dell'avvolgimento primario va collegato ad uno dei due morsetti dell'interruttore S1. Il secondo conduttore a 6,3 volt, dell'avvolgimento secondario di T1, va collegato al piedino n. 5 dello zoccolo portavalvola.

Per gli altri collegamenti non v'è alcun bisogno di interpretazione, perchè essi sono ben evidenziati nello schema pratico del ricevitore. La lettura dei piedini dello zoccolo portavalvola va fatta in senso orario, cioè secondo il verso delle lancette dell'orologio; i piedini n. 2-3-7 sono collegati con il cilindretto metallico centrale dello zoccolo e con la massa (il terminale n. 3 rappresenta il secondo terminale del filamento ed è collegato a massa così come è collegato a massa l'altro conduttore a 6,3 volt dell'avvolgimento secondario di T1).

La foto qui sotto riportata illustra il montaggio pratico del ricevitore a una valvola.

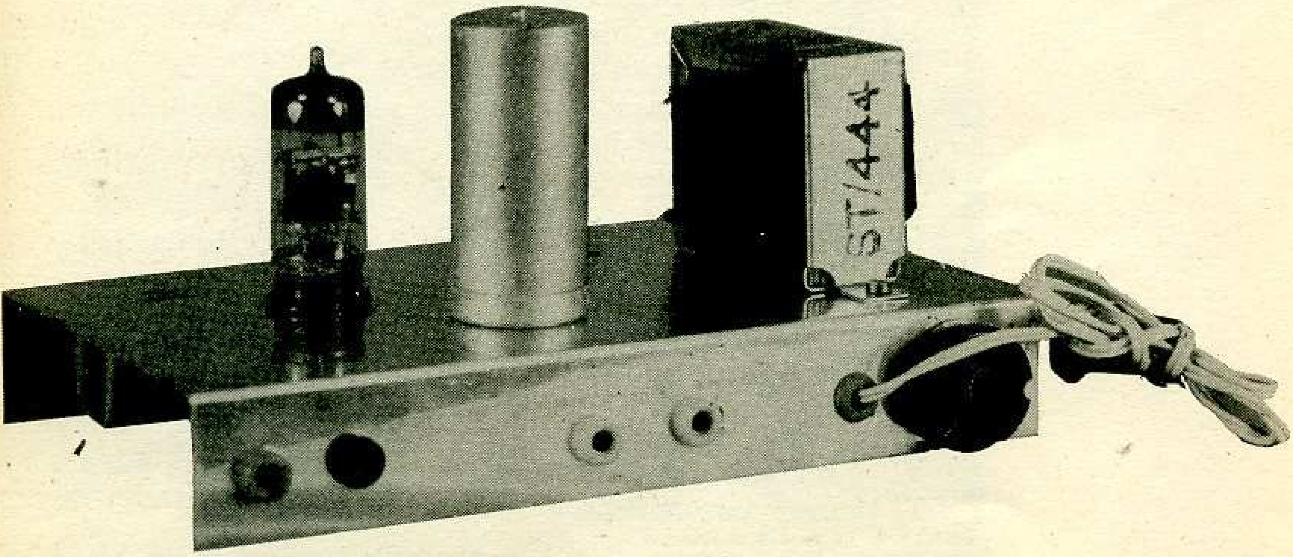

RICEVITORE A 2 VALVOLE

Questo ricevitore a due valvole vuol essere una estensione del precedente ricevitore ad una sola valvola. Con l'aggiunta di una valvola amplificatrice di bassa frequenza è possibile, infatti, trasformare il precedente ricevitore in un ricevitore adatto per l'ascolto in altoparlante. Lo stadio di alta frequenza del ricevitore a due valvole è uguale a quello del ricevitore ad una sola valvola. Il segnale rivelato di bassa frequenza, anzichè essere applicato alla cuffia, viene inviato alla griglia controllo di una valvola tipo 6AQ5 (V2), che è un tetrodo, cioè una valvola a quattro elettrodi (placca-griglia schermo-griglia controllo-catodo).

COMPONENTI DEL RICEVITORE A DUE VALVOLE

C1 = 1.000 pF (condensatore ceramico) - 500 Vl.
C2 = 500 pF (condensatore variabile)
C3 = 75 pF (condensatore ceramico) - 500 Vl.
C4 = 2.200 pF (condensatore ceramico) - 500 Vl.
C5 = 20.000 pF (condensatore a carta) - 500 Vl.
C6 = 50.000 pF (condensatore a carta) - 500 Vl.
C7 = 2.000 pF (condensatore ceramico) - 1500 Vl.
C8 = 50 microfarad - 25 Vl. (condensatore elettrolitico)
C9-C10 = 32 + 32 microfarad - 300 Vl. (condensatore elettrolitico doppio a vitone)
C11 = 10.000 pF (condensatore ceramico) - 1500 Vl.
R1 = 2 megaohm (rosso-nero-verde)
R2 = 220.000 ohm (rosso-rosso-giallo)
R3 = 1 megaohm (marrone-nero-verde)
R4 = 500.000 ohm (potenziometro a variazione logaritmica con interruttore S1 incorporato)
R5 = 220 ohm (rosso-rosso-marrone) - 1 watt
R6 = 1.200 ohm (marrone-rosso-rosso) - 1 watt
Le resistenze per le quali non è indicato il wattaggio si intendono tutte da 1/2 watt.
RS1 = raddrizzatore al selenio: 250 volt - 85 mA
L1 = bobina di sintonia tipo Corbetta CS2
V1 = valvola tipo 6BA6
V2 = valvola tipo 6AQ5
T1 = trasformatore d'alimentazione tipo GBC H/188
T2 = trasformatore d'uscita - 5000 ohm di impedenza

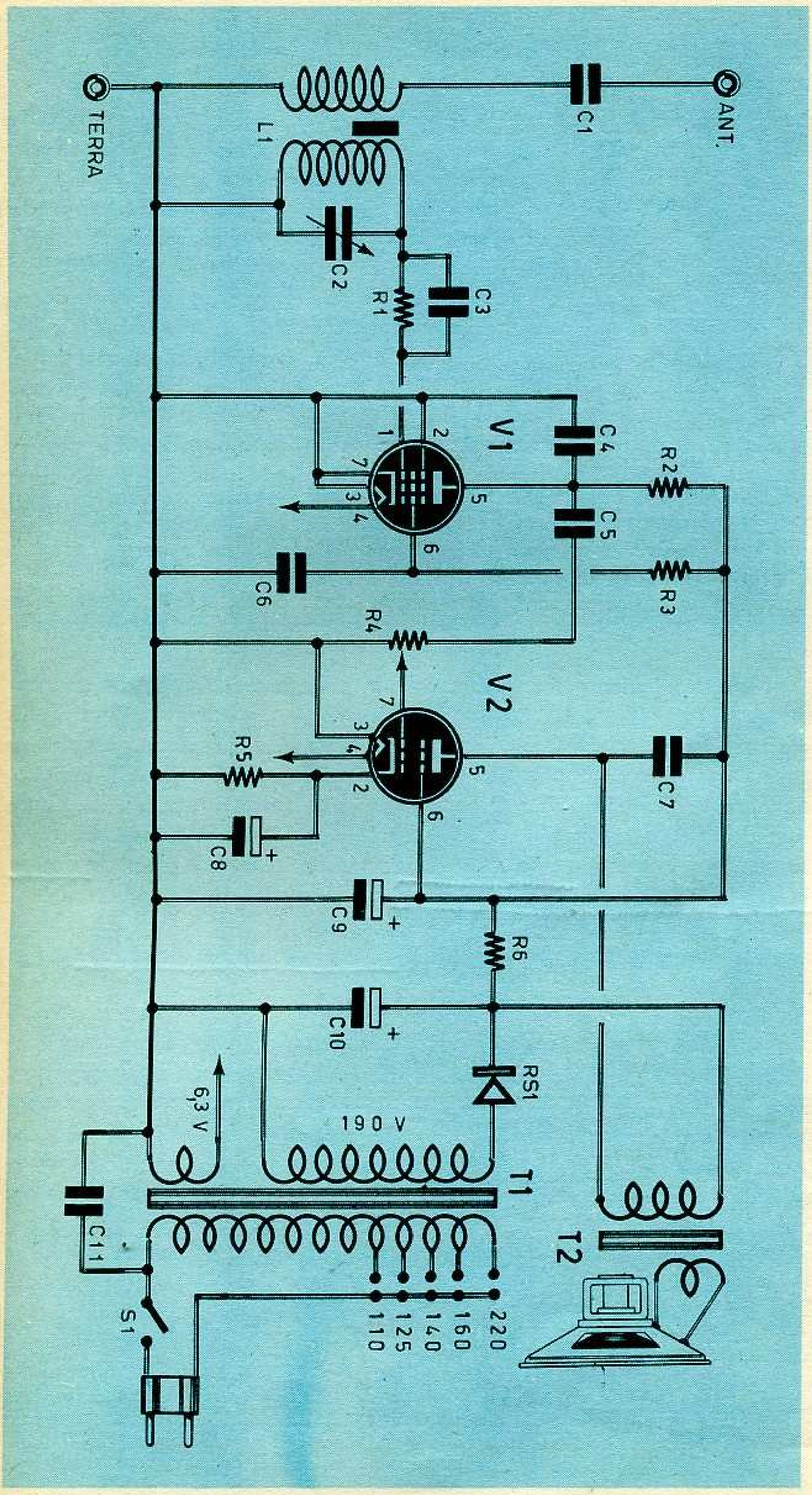
TERRA
ANT.
L1
C1
C2
C3
R1
V1
C4
C5
R2
R3
C6
R4
V2
R5
C7
C8
C9
R6
C10
RS1
6,3 V
190 V
T1
T2
C11
S1
220
160
140
125
110

Schema elettrico del ricevitore a due valvole.

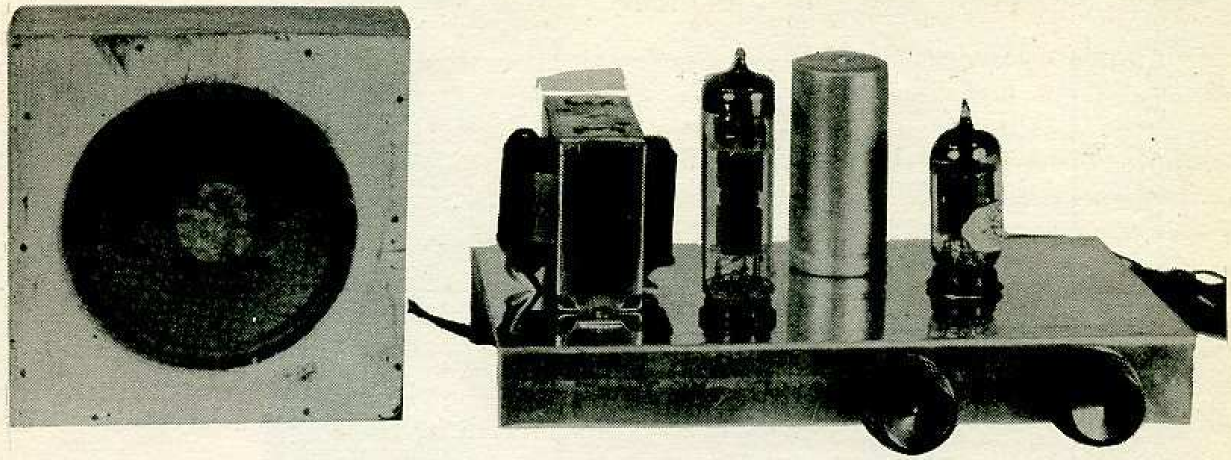

La foto qui sopra riproduce il montaggio sperimentale del ricevitore a due valvole; allo scopo di raggiungere una buona riproduzione sonora, l'altoparlante è stato racchiuso in una cassettina di legno che funge da mobiletto acustico. Il disegno a piè di pagina raffigura lo schema pratico del ricevitore.

L'accoppiamento fra le due valvole avviene per mezzo del condensatore a carta C5, che impedisce alla tensione anodica, applicata alla placca della valvola V1 per mezzo della resistenza R2, di raggiungere la griglia controllo della valvola V2; attraverso il condensatore C5 transitano soltanto i segnali radio di bassa frequenza, sotto forma di una debole corrente alternata; questa debole corrente stabilisce una tensione fra i terminali del potenziometro R4; per mezzo del cursore è possibile dosare la quantità di segnale, cioè la tensione, che si applica nella griglia controllo della valvola V2 (piedino 7 dello zoccolo). È ovvio che dosando il segnale non si fa altro che regolare il volume sonoro del ricevitore.

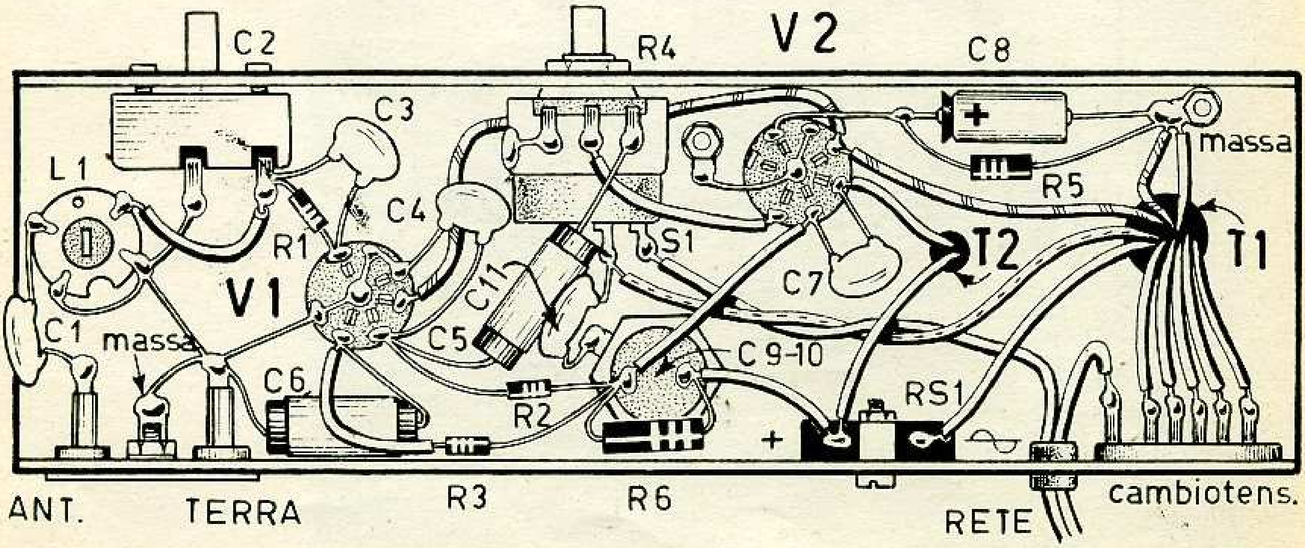

STADIO FINALE

Il segnale amplificato di bassa frequenza, uscente dalla placca della valvola V2 (piedino 5) è presente sull'avvolgimento primario del trasformatore d'uscita. L'alimentatore è identico a quello del precedente ricevitore; il solo valore della resistenza R6 è cambiato. Anche il montaggio di questo ricevitore non differisce di molto da quello precedente. Esso va effettuato su telaio metallico, seguendo i consigli precedentemente esposti. L'altoparlante deve avere un diametro di 70-120 millimetri e la impedenza della sua bobina mobile deve essere uguale a quella dell'avvolgimento secondario del trasformatore d'uscita T2. All'atto dell'acquisto del trasformatore T2 si dovrà chiedere un trasformatore da 5000 ohm di impedenza. Sul catodo della valvola V2, in parallelo alla resistenza R5, risulta collegato il condensatore elettrolitico C8, che prende il nome di condensatore catodico. A questo condensatore è affidato il compito di stabilizzare la tensione di polarizzazione della valvola, cioè la tensione negativa di piccolo valore esistente fra la griglia controllo della valvola e il catodo, in virtù della presenza della resistenza R5; la tensione di polarizzazione è stabile soltanto quando non vi è presenza di segnale; in presenza di segnale la tensione di polarizzazione varia col variare del segnale applicato all'entrata della valvola.

Il condensatore catodico.

RICEVITORE REFLEX MONOVALVOLARE

Il principio del circuito reflex è già stato analizzato in sede di presentazione di un precedente ricevitore a transistor. Questo tipico circuito di ricevitore radio, che tutti i dilettanti realizzano in qualunque corso di radiotecnica, può essere realizzato anche con una valvola, come è il caso del ricevitore qui presentato. La valvola V1, infatti, funziona da amplificatrice dei segnali di alta frequenza, rivelatrice e amplificatrice dei segnali di bassa frequenza. Il circuito di sintonia è sempre lo stesso e i segnali di alta frequenza vengono applicati alla griglia controllo (piedino 6) della valvola V1; essi escono, amplificati dalla placca (piedino 2) e sono presenti nell'avvolgimento primario della bobina L2, che funge da impedenza di alta frequenza, cioè non permette il passaggio dei segnali di alta frequenza. Per induzione questi segnali si trasferiscono nell'avvolgimento secondario della bobina L2, che costituisce un secondo circuito di sintonia e che permette, manovrando il perno del condensatore variabile C8, di raggiungere un ottimo grado di selettività. Da questo

Una sola valvola per tre lavori diversi.

Schema elettrico del ricevitore monovalvolare con circuito reflex.

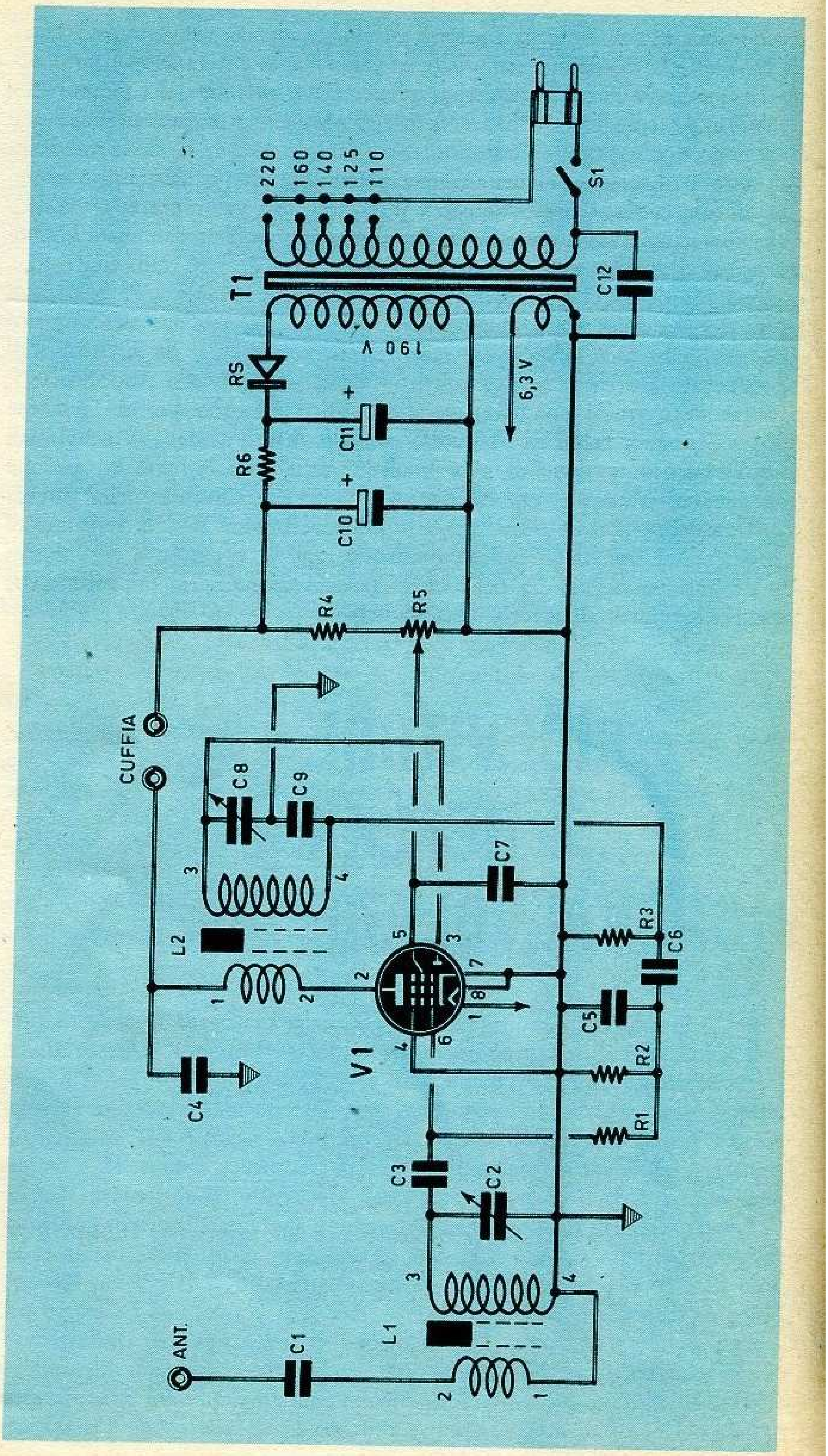

ANT.
C1
L1
1
2
3
4
C2
C3
R1
R2
C4
V1
1
2
3
4
5
6
7
8
C5
C6
R3
C7
L2
1
2
3
4
C8
C9
CUFFIA
R4
R5
C10
+
C11
+
R6
RS
190 V
6,3 V
T1
C12
220
160
140
125
110
S1

secondo circuito di sintonia i segnali di alta frequenza vengono prelevati ed applicati alla placchetta della valvola (piedino 3), che assieme al catodo rappresenta la sezione diodo della valvola V1 e permette il processo di rivelazione dei segnali radio. Di conseguenza, anche i segnali radio di bassa frequenza vengono amplificati ed escono dalla placca della valvola V1; essi attraversano l'avvolgimento primario della bobina L2 e raggiungono la cuffia. Il potenziometro R5 permette di regolare la tensione di griglia schermo e quindi anche il volume sonoro del ricevitore. L'alimentatore è sempre dello stesso tipo. Poichè attraverso la cuffia fluisce la corrente di placca, cioè la corrente anodica della valvola V1, la cuffia stessa funge da carico anodico della valvola V1; la sua impedenza deve essere di 2000-4000 ohm.

Se questo ricevitore viene usato in prossimità della emittente locale, si potrà sostituire la cuffia con un altoparlante, interponendo un trasformatore di uscita come nel caso del precedente ricevitore. Per l'ascolto in cuffia occorre un'antenna della lunghezza di almeno 2 metri.

Le due bobine dei due circuiti di sintonia sono perfettamente identiche e sono di tipo CS3/BE; anche i due condensatori variabili C2 e C8 sono uguali. L'uso di questo ricevitore è semplice: prima si sintonizza il segnale per mezzo del condensatore variabile C2, successivamente, per mezzo del condensatore variabile C8 si raggiunge un più elevato grado di sintonizzazione del segnale, che si manifesta attraverso un aumento del volume e una eventuale eliminazione di disturbi accompagnati ai segnali dell'emittente.

COMPONENTI DEL RICEVITORE REFLEX MONOVAIVOLARE

C1 = 100 pF - 500 Vl. (condensatore ceramico)
C2-C8 = 500 pF (condensatori variabili)
C3 = 100 pF - 500 Vl. (condensatore ceramico)
C4 = 2.200 pF - 500 Vl. (condensatore ceramico)
C5 = 200 pF - 500 Vl. (condensatore ceramico)
C6 = 10.000 pF - 500 Vl. (condensatore ceramico)
C7 = 50.000 pF - 500 Vl. (condensatore a carta)
C8 = Vedi C2
C9 = 1.000 pF - 500 Vl. (condensatore ceramico)
C10-C11 = 32 + 32 microfarad - 300 Vl. (condensatore elettrolitico doppio a vitone)
C12 = 10.000 pF - 1500 Vl. (condensatore a carta)
R1 = 470.000 ohm (giallo-viola-giallo)
R2 = 470.000 ohm (giallo-viola-giallo)
R3 = 470.000 ohm (giallo-viola-giallo)
R4 = 100.000 ohm (marrone-nero-giallo)
R5 = 500.000 ohm (potenziometro con interruttore S1 incorp.)
R6 = 2.200 ohm - 1 watt (rosso-rosso-rosso)
Tutte le resistenze per le quali non viene indicato il wattaggio si intendono da 1/2 watt.
L1-L2 = bobina tipo Corbetta CS3/BE
V1 = valvola tipo EAF42
T1 = trasformatore d'alimentazione tipo GBC H/188

MONTAGGIO RICEVITORE REFLEX MONOVALVOLARE

La realizzazione del ricevitore reflex ad una sola valvola va fatta seguendo lo schema pratico del ricevitore. Il montaggio è ottenuto su telaio metallico. Nella parte anteriore del telaio affiorano i tre perni relativi ai due condensatori variabili e al potenziometro di volume, che ha incorporato l'interruttore S1. Nella parte posteriore del telaio sono presenti il cambiotensione, le due boccole che rappresentano le prese di cuffia e la boccola di presa antenna. A quest'ultima boccola verrà collegato lo spinotto che fa capo al conduttore di discesa di antenna. Ma per ricevere in cuffia non occorre un'antenna esterna molto lunga, perchè basterà installare, nello stesso locale dove funziona il ricevitore, un'antenna di soli due metri. Coloro invece che vorranno far impiego di un altoparlante, collegando in corrispondenza delle boccole di cuffia i due terminali dell'avvolgimento primario del trasformatore di uscita, dovranno provvedere all'installazione di una buona antenna installata nella parte più alta del tetto. Occorre, infatti, aumentare la sensibilità del ricevitore, cioè occorre ricevere i segnali radio con la massima intensità possibile.

Il cablaggio di questo ricevitore deve essere iniziato non appena ulti-

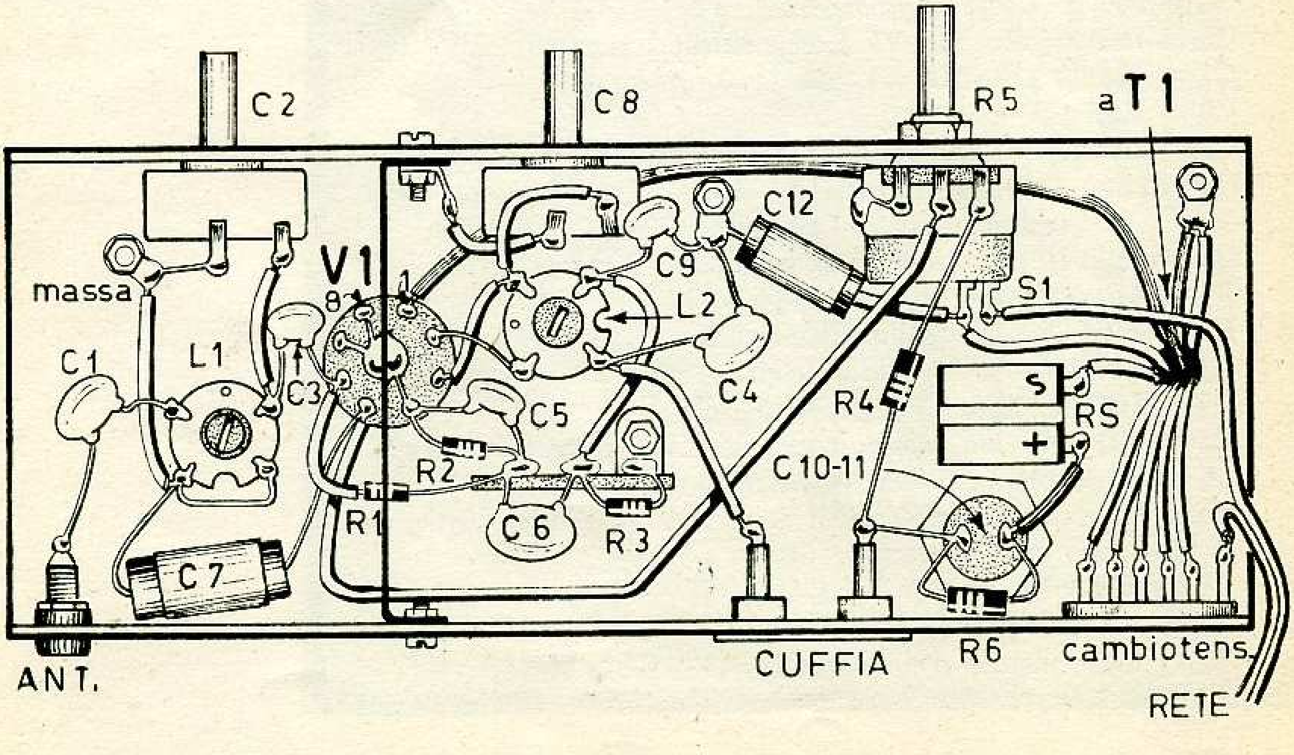

Schema pratico del ricevitore reflex monovalvolare. L'ascolto è ottenuto in cuffia ma è anche possibile collegare un altoparlante munito di trasformatore di uscita nelle boccole contrassegnate con la dicitura: CUFFIA.

mate tutte le operazioni di ordine meccanico; le prime saldature da farsi sono quelle dei terminali uscenti dal trasformatore di alimentazione T1, così come si è fatto per il precedente ricevitore. In pratica questo ricevitore si distingue dal primo tipo di ricevitore ad una sola valvola con ascolto in cuffia per l'impiego di due bobine di sintonia e di due condensatori variabili, cioè di due circuiti accordati (circuiti di sintonia). Ma in fase di montaggio di questo ricevitore occorre realizzare un accorgimento meccanico, cioè una schermatura particolare sullo zoccolo portavalvola della valvola V1. Si tratta di far attraversare il telaio, proprio in corrispondenza del cilindretto metallico dello zoccolo portavalvola, con una sbarretta di ottone o di ferro (non di alluminio perchè non potrebbe essere saldata). Questa sbarretta è ripiegata alle estremità e fissata al telaio, da ambo le parti, mediante due viti e due dadi. In corrispondenza del cilindretto metallico dello zoccolo, questa sbarretta deve essere accuratamente saldata a stagno. In questa stessa sbarretta occorrerà praticare, mediante il trapano, due fori: uno serve per il passaggio della resistenza R1 (il diametro di questo foro deve essere tale da non permettere che la resistenza R1 tocchi il metallo), l'altro serve per il passaggio del conduttore che porta la tensione alla griglia schermo della valvola (piedino 5 dello zoccolo). Lo zoccolo della valvola V1 si differenzia da quelli delle precedenti valvole, che erano a 7 piedini, mentre questo è a 8 piedini.

Occorre schermare lo zoccolo della valvola.

ZOCCOLI PORTAVALVOLE

La disposizione dei piedini delle valvole, che fanno capo agli elettrodi contenuti nel bulbo di vetro, prende il nome di « zoccolo della valvola ». La sua lettura si effettua sempre capovolgendo la valvola con la testa all'ingiù e contando la progressione numerica dei piedini, a partire dal piedino 1, secondo il verso delle lancette dell'orologio (verso orario).

Normalmente i piedini della valvola sono disposti lungo una circonferenza e sono equidistanti tra di loro. Tale distanza, invece, è più grande tra il primo e l'ultimo piedino dello zoccolo, ed è proprio tale distanza che permette di individuare il primo e l'ultimo piedino dello zoccolo. Ciò accade nelle valvole a 7 piedini, a 8 piedini e a 9 piedini (valvole noval).

Vi sono delle valvole denominate RIMLOCK nelle quali tutti i piedini dello zoccolo, che sono in numero di 8, sono equidistanti tra di loro; ma in queste valvole, sul bulbo di vetro, in prossimità dello zoccolo

Zoccolo per valvola Rimlock.

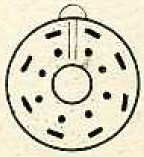

è ricavata una tasca-guida, che permette di individuare il piedino n. 1 e quello n. 8, perchè proprio questa tacca si trova in corrispondenza della immaginaria linea di divisione del primo e dell'ultimo piedino.

Tutte le valvole, per poter essere correttamente impiegate, devono essere infilate nell'apposito zoccolo portavalvola; e in pratica l'allievo radiotecnico, quando deve « leggere » i piedini della valvola, non fa mai riferimento allo zoccolo della valvola, ma soltanto e sempre allo zoccolo portavalvola, « leggendolo » nella sua parte posteriore, cioè in quella opposta alla faccia in cui verrà infilata la valvola. Dunque, quel che importa è saper leggere lo zoccolo portavalvola; in esso non vi sono piedini, ma soltanto terminali adatti ad essere saldati ed occorre necessariamente conoscere l'ordine numerico progressivo, perchè all'atto pratico il radiotecnico abbina l'ordine numerico degli elettrodi della valvola rilevati nel simbolo elettrico della valvola con l'ordine numerico dei terminali dello zoccolo portavalvola che, in pratica, vengono pur essi chiamati piedini.

Come si fa a riconoscere il piedino n. 1 nello zoccolo portavalvola di una valvola di tipo RIMLOCK come quella montata sul ricevitore reflex?

È cosa assai semplice. Fra il piedino 1 e il piedino 8 è impresso, nello zoccolo portavalvola, un trattino; guardando lo zoccolo con il trattino rivolto verso l'osservatore, il terminale immediatamente a sinistra del trattino rappresenta il terminale 1, mentre quello immediatamente a destra è il terminale n. 8. Nello schema pratico del ricevitore reflex ad 1 valvola questo trattino non è visibile, perchè è ricoperto dal tratto indicativo in lamierino di ottone che funge da schermo elettromagnetico e che serve ad evitare l'insorgere di fischi o di altri disturbi a causa della notevole amplificazione operata dalla valvola V1.

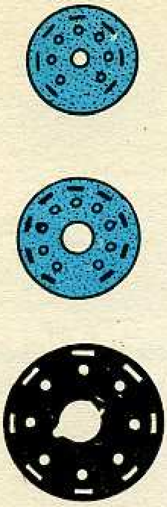

Dall'alto al basso: zoccolo per valvola miniatura a 7 piedini; zoccolo noval per valvola a 9 piedini; zoccolo octal per valvola a 8 piedini.

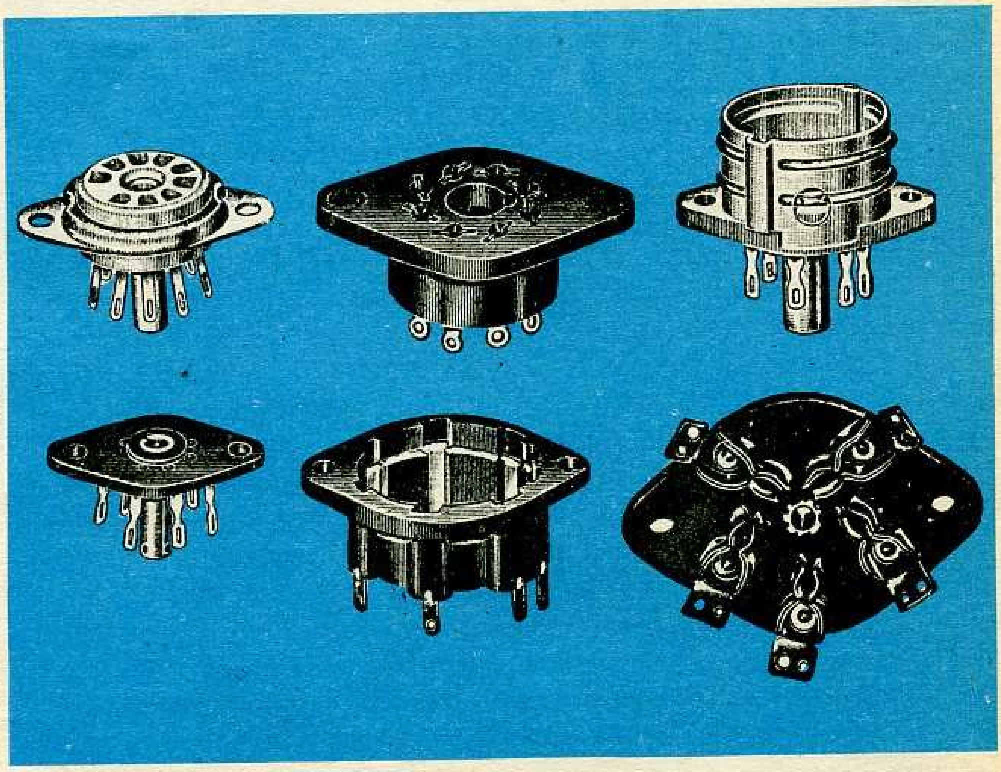

Principali tipi di zoccoli portavalvole usati in apparati radio moderni e di vecchio tipo.

CIRCUITO SUPER ETERODINA

Può sembrare inverosimile, eppure esistono oggi molti tecnici, dilettanti e professionisti, che hanno mosso i loro primi passi, lungo il cammino della radiotecnica, cominciando proprio con la supereterodina.

Generalmente ciò non avviene. E chi, per la prima volta, è preso dalla passione per la radio, volge il suo interessamento ai montaggi più semplici, ai piccoli ricevitori a cuffia, che traggono dall'antenna la sola energia capace di farli funzionare. I circuiti più complessi, quelli ad una e più valvole, alimentati dalle pile, prima, e dalla tensione della rete-luce poi, vengono affrontati soltanto in un secondo tempo, quando si è divenuti più esperti e più... amici delle resistenze, dei condensatori, dei trasformatori, dei circuiti teorici e di quelli pratici.

Al circuito supereterodina ci si arriva dopo aver effettuato una lunga serie di montaggi, via via più complessi e più impegnativi, che arricchiscono, progressivamente, il bagaglio di cognizioni acquisite ed affinano la tecnica.

È il circuito di tutti gli apparati di tipo commerciale.

Ma le eccezioni ci sono sempre e proprio in questo caso possiamo dire che esse sono molto frequenti. Sono frequenti, specialmente, tra coloro che hanno già delle nozioni generali di elettrotecnica ed hanno la fortuna di avere un parente od un amico radiotecnico, disposto ad offrire la propria preparazione ed esperienza nell'insegnare ed assistere. Sono frequenti, ancora, fra i giovani dotati di spirito di intraprendenza e d'avventura, che non temono l'insuccesso e vogliono riuscire ad ogni costo, chiedendo consigli ed insegnamenti a chi ne sa di più fra amici e conoscenti e rivolgendosi, persino, agli stessi tecnici professionisti dei laboratori radio.

Del resto, il montaggio di un ricevitore a circuito supereterodina, non è poi un'impresa tanto ardua e difficile, quando v'è passione, v'è modo di essere guidati, consigliati, assistiti. Con ciò, tuttavia, intendiamoci bene, non si vuol qui esortare l'allievo ad improvvisarsi radiotecnico di punto in bianco, per invitarlo a costruire subito il ricevitore supereterodina, eliminando il montaggio di tutti gli apparati fin qui descritti. Assolutamente no! Perchè si insegnerebbe al lettore di buttar via tempo e danaro per nulla! Ma coloro che sono appassionati di radio ed hanno potuto già acquisire in precedenza una certa famigliarità con i circuiti radio, potranno sorvolare sulla realizzazione dei sette apparati fin qui presentati per realizzare il ricevitore a circuito supereterodina, che costituisce un capitolo a sè, piuttosto ampio, in ogni corso di radiotecnica.

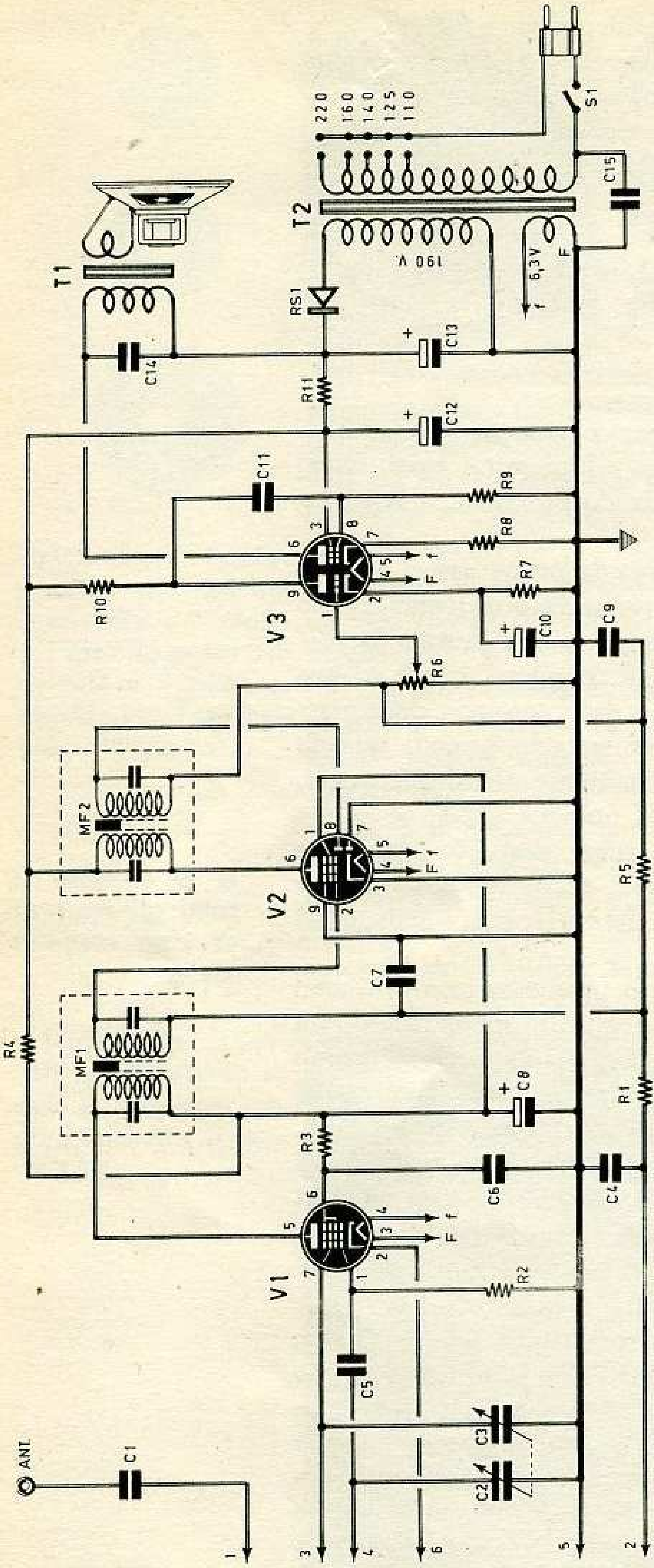

220
160
140
125
110
S1
C15
T2
190 V
6,3 V
F
T1
RS1
C14
C13
R11
C12
C11
R9
R8
R7
R10
V3
C10
C9
R6
MF2
V2
R5
C7
MF1
R4
C8
R1
R3
C6
C4
V1
R2
C5
C3
C2
ANT.
C1
1
3
4
6
5
2

Schema elettrico del ricevitore a circuito supereterodina descritto in queste pagine. Il numero delle valvole, per motivi di semplicità e di economia, è stato ridotto al minimo.

COMPONENTI

Condensatori

C1 = 2.000 pF - 1500 Vl. (a carta)
C2 - C3 = 450 + 450 pF (variabile doppio)
C4 = 50.000 pF - 500 Vl. (a carta)
C5 = 100 pF - 500 Vl. (polistirolo)
C6 = 50.000 pF - 1500 Vl. (a carta)
C7 = 50.000 pF - 1500 Vl. (a carta)
C8 = 8 microfarad - 350 Vl. (elettrolitico)
C9 = 250 pF - 500 Vl. (polistirolo)
C10 = 50 microfarad - 25 Vl. (elettrolitico)
C11 = 10.000 pF - 500 Vl. (ceramico)
C12 - C13 = 32 + 32 microfarad - 350 Vl. (elettrolitico a doppio vitone)
C14 = 2.000 pF - 1000 Vl. (a carta)
C15 = 2.000 pF - 1000 Vl. (a carta)

Resistenze

R1 = 500.000 ohm
R2 = 22.000 ohm
R3 = 4.000 ohm
R4 = 1.000 ohm
R5 = 1 megaohm
R6 = 500.000 ohm (potenziometro con interruttore S1 incorp.)
R7 = 1.500 ohm
R8 = 220 ohm - 1 watt
R9 = 470.000 ohm
R10 = 220.000 ohm
R11 = 1.200 ohm - 1 watt

Le resistenze per le quali non è stato indicato il wattaggio si intendono tutte da 1/2 watt.

Varie

V1 = 6BE6
V2 = EBF89
V3 = ECL86
MF1 = primo trasformatore di media frequenza
MF2 = secondo trasformatore di media frequenza
T1 = trasformatore d'uscita - 5.000 ohm di impedenza
T2 = trasformatore d'alimentazione tipo GBC H/188
RS1 = raddrizzatore al selenio 250 V - 85 mA
Gruppo A.F. - tipo Corbetta CS/23 R.

CONVERSIONE DI FREQUENZA

Cercheremo di contenere nello spazio di poche pagine tutto ciò che è necessario dire, col proposito di esprimerci con l'allievo con la massima chiarezza, senza lesinare sui consigli e suggerimenti. La teoria è necessaria, sì, ma quel che più importa all'allievo, lo sappiamo per esperienza, è la spiegazione ordinata, completa, delle successive fasi di montaggio, che conducono con sicurezza alla realizzazione pratica di un ricevitore perfettamente funzionante e privo di difetti. Ed è proprio in questa seconda fase della nostra esposizione, che il lettore incontrerà una vera assistenza tecnica.

Che cosa significa SUPERETERODINA? Significa semplicemente: circuito radio a CONVERSIONE DI FREQUENZA. Per dirla in altre parole, ciò significa che nei ricevitori a circuito supereterodina le frequenze dei segnali in arrivo, di qualunque valore esse siano, vengono sempre trasformate in un'altra frequenza che è sempre la stessa per ogni tipo di ricevitore.

Questa nuova frequenza può essere, come spesso avviene, di 470 Kc/s (chilocicli al secondo). Si suol dire allora che la MEDIA FREQUENZA dell'apparecchio è di 470 Kc/s.

Significato della parola « supereterodina ».

Nel ricevitore che presentiamo il valore della media frequenza è di 467 Kc/s. Pertanto, qualunque sia la frequenza del segnale radio in arrivo, essa viene sempre cambiata in quella di 467 Kc/s. Se l'apparecchio ad esempio, è accordato su una stazione ad onde medie, di 1000 Kc/s, tale frequenza viene cambiata in quella di 467 Kc/s. Se la frequenza della stazione è di 800 Kc/s, anche essa viene cambiata in quella di 467 Kc/s; se l'apparecchio è accordato su una stazione ad onde corte, ad esempio di 10.000 Kc/s, anche questa frequenza di 10.000 Kc/s viene cambiata in quella di 467 chilocicli.

Convertire la frequenza del segnale in arrivo in un'altra frequenza qualsiasi, è cosa facile. A tale scopo provvede la prima valvola del circuito, quella contrassegnata con V1 nello schema elettrico. Ed è proprio per questo motivo che la prima valvola di un circuito supereterodina viene chiamata CONVERTITRICE.

La valvola V1, che è di tipo 6BA6, è un pentodo che provvede ad amplificare i segnali di alta frequenza in arrivo, a generare una tensione oscillante e a sovrapporre le due tensioni. Le due tensioni, quella in arrivo dovuta alla captazione delle onde radio, e quella generata dalla sezione oscillatrice della valvola, si sovrappongono, e da tale sovrapposizione risulta il cambiamento di frequenza del segnale in arrivo.

IL PRIMO STADIO del RICEVITORE

Cominciamo ora con l'esaminare il primo stadio del ricevitore supereterodina, seguendo attentamente lo schema elettrico.

Per conferire una maggiore semplicità allo schema, è stato volutamente omesso il disegno delle varie bobine di antenna e di oscillatore, relative alle onde medie e alle onde corte e dei pochi condensatori, di piccola capacità, ad esse collegate, sostituendo tutto ciò direttamente con il cosiddetto GRUPPO ALTA FREQUENZA proprio perchè, in pratica, si ha a che fare soltanto con il gruppo A.F. sul quale si interverrà, a lavoro ultimato, per le poche e semplici operazioni di taratura.

IL GRUPPO A.F.

Il gruppo A.F. è un apparato radioelettrico e meccanico insieme. In esso sono contenute, principalmente, due coppie di bobine: una coppia serve per l'ascolto delle onde medie e una coppia serve per l'ascolto delle onde corte. Ciascuna coppia è costituita dalla bobina d'aereo e da quella d'oscillatore. Il perno, di cui è dotato il gruppo A.F., serve per inserire, manualmente, ora l'una ora l'altra delle due coppie di bobine nel circuito del ricevitore.

Analizziamo ora il funzionamento del primo stadio del ricevitore quando nel suo circuito è inserita la coppia di bobine relativa alle onde medie (il principio di funzionamento è lo stesso quando risulta inserita la coppia di bobine relativa alle onde corte).

La bobina d'aereo è costituita da un avvolgimento primario e da uno secondario. Ai capi dell'avvolgimento primario vengono collegate l'antenna e la terra (numeri 1 e 5 del gruppo A.F.). I due terminali, invece, dell'avvolgimento secondario risultano così collegati: uno (terminale 4 del

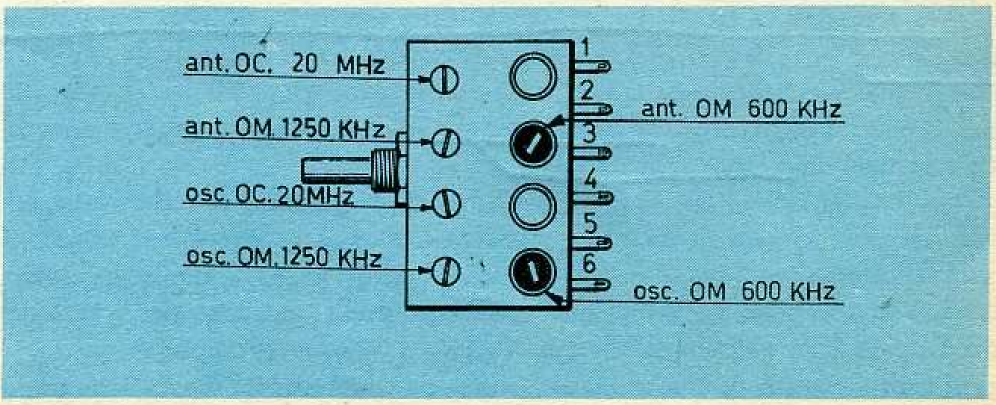

In fase di messa a punto del ricevitore, l'allievo dovrà intervenire con un cacciavite sulle quattro viti allineate lungo un bordo della piastrina superiore del gruppo e sui due nuclei allineati sul bordo opposto.

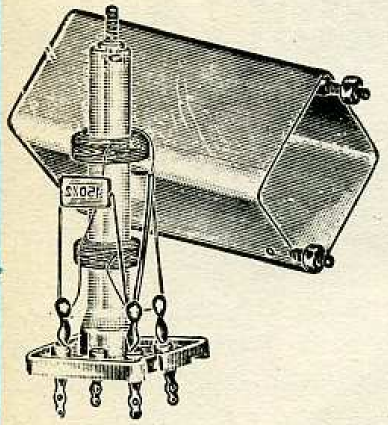

La media frequenza è in pratica un trasformatore con due avvolgimenti (primario e secondario). Tutto il circuito viene racchiuso in una custodia metallica che ha funzioni di schermo elettromagnetico.

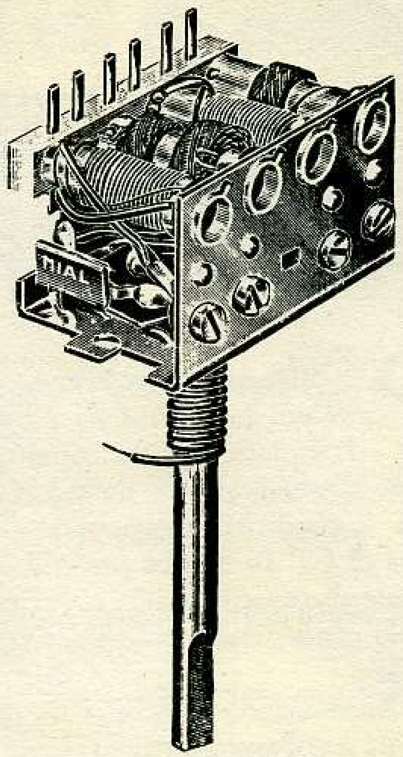

Nel gruppo A.F. vi sono quattro bobine, alcuni condensatori e un commutatore multiplo, che permette, mediante il perno di comando, di commutare il ricevitore sulla gamma delle onde medie e su quella delle onde corte.

gruppo) è collegato al condensatore variabile C2 e, tramite il condensatore C5 alla griglia controllo (piedino 1 dello zoccolo) della valvola convertitrice V1. L'altro terminale dell'avvolgimento secondario della bobina d'aereo è collegato (terminale 2 del gruppo) ad un circuito chiamato C.A.V. e di cui vedremo più avanti il significato.

Il gruppo A.F. permette di commutare il ricevitore anche nella posizione «fono».

Il secondario della bobina d'aereo ed il condensatore variabile C2 costituiscono il primo circuito accordato d'entrata del ricevitore.

Sempre alla stessa valvola V1 risulta pure collegato il secondo circuito accordato, costituito dalla bobina d'oscillatore (contenuta nel gruppo) e dal condensatore variabile C3.

La bobina d'oscillatore è costituita da due avvolgimenti i quali hanno entrambi un terminale collegato a massa. Gli altri due terminali sono collegati al condensatore variabile C3 (terminale 3 del gruppo) e al catodo (terminale 6 del gruppo). Dunque la bobina d'oscillatore è dotata, complessivamente, di tre terminali: uno è collegato a massa l'altro alla griglia mescolatrice (piedino 7 dello zoccolo), e il terzo è collegato al catodo.

FREQUENZA dell' OSCILLATORE

Per ottenere la conversione di frequenza è necessario che il circuito
periore a quella della tensione A.F. in arrivo. Per tale motivo la
d'oscillatore risulti costantemente accordato ad una frequenza su-
bobina d'oscillatore ha un'induttanza minore della bobina d'entrata ed anche il condensatore variabile C3 deve avere una capacità inferiore a quella di C2. In pratica però C2 e C3 hanno la stessa capacità e la diminuzione di capacità di C3 si ottiene collegando in serie ad esso un altro piccolo condensatore contenuto nel gruppo, detto « correttore » o « padding ».

Stadio amplificatore di media frequenza.

Riassumendo i concetti fin qui esposti, diciamo che alla prima valvola V1 è collegato il primo circuito accordato, quello d'aereo e il secondo circuito accordato, quello d'oscillatore. La conversione di frequenza dei segnali radio in arrivo avviene nell'interno della valvola V1. Più precisamente la tensione oscillante viene prelevata dal catodo (piedino 2 dello zoccolo) e introdotta nella valvola attraverso la griglia mescolatrice (piedino 7 dello zoccolo).

All'uscita, quindi, della prima valvola, e cioè sulla placca (piedino 5), si ritrova il segnale radio, che il condensatore variabile C2 ha selezionato tra i tanti segnali in arrivo sull'antenna, amplificato e con una frequenza diversa, esattamente quella di 476 Kc/s. Questa nuova frequenza, come abbiamo detto, viene anche chiamata MEDIA FREQUENZA.

I trasformatori di media frequenza sono costituiti da due circuiti accordati alla frequenza di 467 Kc/s, dei quali uno è detto « primario », l'altro viene chiamato « secondario ».

Facendo sempre riferimento allo schema elettrico del ricevitore si osserva che tra la prima valvola V1 e la seconda V2 è interposto un trasformatore, contrassegnato in figura con il simbolo MF1. È questo il primo di due trasformatori (il secondo viene subito dopo la valvola V2 ed è contrassegnato in figura con MF2) detti TRASFORMATORI DI MEDIA FREQUENZA o, anche, più semplicemente, MEDIE FREQUENZE. Ciascuno di questi due trasformatori assolve il compito principale di accoppiare l'uscita della valvola precedente con l'entrata di quella seguente, oltre quello di operare una ulteriore selezione del segnale.

Nel caso di MF1 il primario è collegato alla placca di V1 mentre il secondario è collegato alla griglia controllo di V2 (piedino 2) che costituisce la valvola amplificatrice di media frequenza.

Questo sistema di collegare tra loro due stadi amplificatori, conferisce al ricevitore un elevato grado di selettività, appunto perchè essendo le due medie frequenze accordate alla stessa frequenza fissa di 467 Kc/s esse vietano il passaggio a tutti quei segnali radio di frequenza diversa che eventualmente si fossero introdotti nel circuito.

CONTROLLO AUTOMATICO del VOLUME

Le emittenti locali determinano all'entrata dell'apparecchio tensioni A.F. fortissime, che non è necessario amplificare molto; in tal caso la valvola V2 amplifica poco.

Le emittenti molto lontane determinano all'entrata dell'apparecchio tensioni A.F. debolissime, che è necessario amplificare al massimo, come effettivamente avviene.

A ciò provvede un semplicissimo dispositivo detto « Controllo Automatico di Volume » o, abbreviatamente, C.A.V.

Ma per arrivare alla descrizione di questo dispositivo, seguiamo ancora per un momento il percorso del segnale radio.

L'amplificazione della seconda valvola non è costante per tutte le tensioni A.F. in arrivo, ma varia automaticamente.

Il segnale di media frequenza amplificato dalla valvola V2 viene applicato al primario della seconda media frequenza. Da questo, per induzione, passa al secondario che è collegato da un lato al diodo rivelatore, più precisamente alla placchetta del diodo rivelatore (piedino 8 di V2), mentre dall'altro lato è collegato alla resistenza variabile R6 che costituisce il potenziometro di volume, con il quale si regola manualmente il volume del ricevitore.

Il secondario della seconda media frequenza, la placchetta del diodo (piedino 8), il catodo (piedino 3), il potenziometro R6 costituiscono gli elementi del circuito di RIVELAZIONE del ricevitore.

Ritorniamo ora al circuito C.A.V. e consideriamo ancora il potenziometro R6. Ai suoi capi, evidentemente, è presente la tensione del segnale rivelato.

Il lato della resistenza variabile R6 collegato a massa, in pratica al telaio del ricevitore, è positivo. L'altro lato è negativo, e ad esso è collegata l'uscita del secondario del primo trasformatore di media frequenza, tramite la resistenza R5. Maggiore è il segnale, maggiore è tale tensione negativa e minore è l'amplificazione della valvola V2. Il condensatore C7 da 50.000 pF, serve a livellare la tensione, in modo che essa sia continua. È questo il circuito C.A.V.

La tensione CAV viene prelevata dal potenziometro di volume.

Dopo la resistenza R5 risulta collegata la resistenza R1, attraverso la quale viene convogliata la tensione C.A.V. sul secondario della bobina d'aereo per cui anche l'ampiezza del segnale radio in arrivo viene automaticamente controllata e di conseguenza viene anche controllata l'amplificazione della prima valvola V1. Il terminale del secondario della bobina d'aereo cui viene applicata la tensione negativa C.A.V. è contrassegnato, sul gruppo A.F., con il numero 2.

STADIO PREAMPLIFICATORE DI B.F.

La valvola V2 amplifica i segnali di media frequenza. La valvola V3 amplifica i segnali di bassa frequenza. Questa valvola è di tipo doppio, cioè essa racchiude nel suo bulbo due valvole contemporaneamente: un triodo e un pentodo. Il triodo funge da valvola preamplificatrice di bassa frequenza, mentre il pentodo funziona da valvola amplificatrice di potenza. Sul potenziometro R6 è presente la tensione del segnale rivelato, cioè di bassa frequenza, privo di ogni componente di alta frequenza, che risulta fugata a massa tramite il condensatore C9. Tale tensione viene applicata, nella dose preferita, alla griglia controllo (piedino 1) della sezione triodica di V3. Dosando la tensione rivelata da applicarsi alla griglia controllo, si regola il volume sonoro nell'altoparlante. Dalla placca, (piedino 9) il segnale di bassa frequenza amplificato viene inviato allo stadio amplificatore finale di potenza, cioè alla sezione pentodo della valvola V3, tramite il condensatore di accoppiamento C11. Il segnale viene introdotto nella griglia controllo (piedino 8).

STADIO AMPLIFICATORE FINALE

Una valvola di potenza per pilotare l'altoparlante.

Lo stadio amplificatore finale è rappresentato dalla sezione pentodo della valvola V3. I segnali preamplificati vengono applicati alla griglia controllo (piedino 8) e risultano sufficientemente amplificati sulla placca, piedino 6. Dopo questo ulteriore processo di amplificazione, i segnali radio sono in grado di pilotare l'altoparlante. La placca (piedino 6) della valvola V3 è collegata con l'avvolgimento primario del trasformatore d'uscita T1; l'altro terminale di questo avvolgimento è collegato con la tensione anodica raddrizzata dal raddrizzatore al selenio RS1.

L'ALTOPARLANTE

L'altoparlante è munito, nella parte posteriore, di due ancoraggi (terminali); questi terminali vanno collegati all'avvolgimento secondario del trasformatore di uscita T1, che è quello dotato di un minor numero di spire ed è realizzato con filo di rame smaltato di spessore molto più elevato di quello usato per l'avvolgimento primario e ciò lo rende facilmente riconoscibile. L'altoparlante deve essere di tipo magnetico e di diametro superiore ai 10 centimetri, per poter sopportare la potenza di uscita del ricevitore. In molti tipi di ricevitori radio il trasformatore di uscita T1 anzichè essere applicato al telaio, è fissato direttamente sul cestello dell'altoparlante.

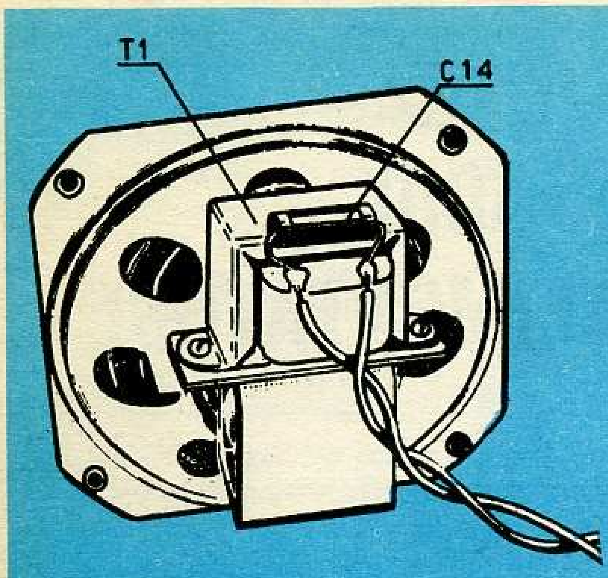

Il condensatore, collegato in parallelo all'avvolgimento primario del trasformatore d'uscita, ha il compito di evitare che il trasformatore oscilli, disturbando la ricezione. Con valori capacitivi bassi la riproduzione sonora è « più chiara », con valori capacitivi elevati la voce diviene « più scura ».

L'applicazione diretta del trasformatore d'uscita sul cestello dell'altoparlante costituisce una grande comodità di montaggio, perchè permette di eliminare un componente dal telaio del ricevitore. Si tratta quindi di economia di spazio, che permette di ridurre le dimensioni dei piccoli ricevitori a circuito supereterodina a valvole.

Un accorgimento tecnico che semplifica il montaggio meccanico.

L'applicazione del trasformatore d'uscita sul cestello dell'altoparlante è un'operazione assai semplice. Occorre preparare, a parte, una piastrina di alluminio, di forma rettangolare, la cui lunghezza deve essere uguale a quella che si misura tra le estremità delle squadrette di fissaggio del trasformatore d'uscita, quelle nelle quali sono riportati i due fori per il passaggio delle viti di fissaggio. Sulle estremità di questa piastrina si dovranno ricavare due fori in corrispondenza di quelli praticati sulle estremità dello schermo del trasformatore d'uscita. La piastrina viene fatta passare internamente al cestello, al di sopra del cilindretto che rappresenta il magnete permanente dell'altoparlante. Una piastrina di bachelite, della stessa forma di quella metallica, deve essere intercalata fra il trasformatore d'uscita e la parte superiore metallica del cestello. Le due viti dovranno essere bloccate saldamente mediante dadi e interposizione di quattro rondelle.

I terminali dei trasformatori d'uscita vengono attuati con due diversi sistemi: quello dei fili uscenti dal trasformatore e quello degli ancoraggi rigidi. Il secondo sistema è più comodo per il fissaggio del condensatore a carta collegato in parallelo all'avvolgimento primario del trasformatore d'uscita. In entrambi i casi è facile riconoscere l'avvolgimento primario da quello secondario mediante l'uso dell'ohmmetro, oppure semplicemente a vista. L'avvolgimento primario ha sempre una resistenza di qualche migliaio di ohm, mentre l'avvolgimento secondario ha una resistenza di pochi ohm. In ogni caso le estremità dell'avvolgimento secondario sono rappresentate da due fili di rame di sezione relativamente elevata, mentre i terminali dell'avvolgimento primario sono rappresentati da conduttori sottilissimi. Non riuscendo ad individuare, a vista, i due avvolgimenti, è sempre consigliabile l'uso dell'ohmmetro.

STADIO ALIMENTATORE

All'estrema destra dello schema elettrico del ricevitore supereterodina è rappresentato lo stadio alimentatore. Esso provvede a trasformare la tensione di rete in due valori diversi di tensione: quella a 190 volt per l'alimentazione anodica del circuito e quella a 6,3 volt per l'alimentazione dei filamenti delle tre valvole ed eventualmente anche di una o due lampadine per l'illuminazione della scala parlante. Il trasformatore di alimentazione T2 è dotato di un avvolgimento primario, a presa universale, cioè adatto a tutte le tensioni di rete. Subito dopo la spina-luce è indicato il simbolo dell'interruttore, contrassegnato con la sigla S1; questo interruttore si trova incorporato nel potenziometro di volume R6; mediante la rotazione del perno di questo potenziometro, dapprima si fa scattare l'interruttore e poi si regola il volume del ricevitore. In uno dei due conduttori provenienti dalla spina-luce è collegato un condensatore antidisturbo, contrassegnato con la sigla C15, che ha il compito di eliminare eventuali disturbi presenti nell'impianto elettrico di casa. L'altro terminale va collegato al cambiotensione. Uno dei due terminali dell'avvolgimento secondario a 190 volt è collegato al raddrizzatore al selenio RS1; questo componente provvede a trasformare la corrente alternata in corrente unidirezionale pulsante. A valle del raddrizzatore al selenio vi è la cellula di filtro, composta dalla resistenza R11 e dai due condensatori elettrolitici C12 e C13; questa cellula provvede a trasformare la corrente unidirezionale pulsante, uscente dal raddrizzatore, in corrente continua.

La cellula di filtro trasforma la corrente pulsante in corrente continua.

PROCEDIMENTO MECCANICO DI MONTAGGIO

Prima cosa da fare è quella di applicare al telaio i tre zoccoli sui quali, a lavoro ultimato, verranno infilate le tre valvole. Successivamente si fissa, mediante viti e dadi, il trasformatore di alimentazione T2 e si passa quindi all'applicazione del cambiotensione, delle varie prese di massa, del gruppo A.F., del condensatore elettrolitico doppio a vitone di C12 - C13, del potenziometro R6 e dei due trasformatori di media frequenza MF1 ed MF2. A proposito di questi due trasformatori ricordiamo che all'atto dell'acquisto essi risultano accompagnati da un foglietto indicatore dal quale è possibile dedurre quale è il primo e quale il secondo e a quali avvolgimenti corrispondono i quattro terminali. Molte volte queste indicazioni sono direttamente riportate sulla scatola di cartone che li contiene.

Quando i perni dei potenziometri sono troppo lunghi occorre segarli. Il potenziometro deve essere fissato sulla morsa nel modo indicato nel disegno, in modo che lo sforzo meccanico esercitato con la sega non venga risentito dal potenziometro vero e proprio.

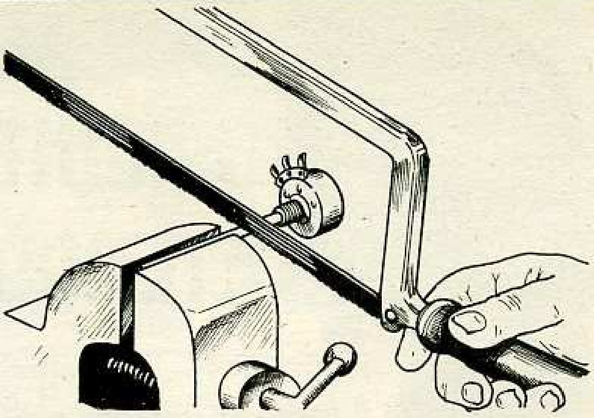

Ciò risulta molto importante perchè nel fissare i due trasformatori al telaio si rende necessario disporli in modo che i terminali risultino vicini ai piedini delle valvole a cui verranno collegati in fase di cablaggio: soltanto rispettando questa disposizione si possono ottenere dei collegamenti corti come è necessario.

Il fissaggio delle prese di massa costituisce un'operazione molto importante, perchè proprio da essa può dipendere, in buona parte, il funzionamento del ricevitore. Quello che importa è che gli ancoraggi di massa formino un perfetto contatto elettrico con il telaio. È dunque necessario provvedere alla pulizia delle parti prima del loro fissaggio; ciò significa che il telaio dovrà essere ben raschiato nel punto in cui si deve fissare l'ancoraggio di massa; la raschiatura, che ha lo scopo di eliminare l'eventuale patina di ossido ed eventuali impurità formatesi sul telaio stesso, può essere effettuata mediante un cacciavite o un temperino. Questa stessa operazione va fatta, ovviamente, anche sull'ancoraggio di massa. Vite e dado dovranno essere ben stretti, mediante interposizione di rondelle. Negli apparati di classe e, in particolar modo, negli amplificatori ad alta fedeltà e in quelli stereofonici, si preferisce a questo sistema di ritorni di massa un conduttore unico; il conduttore unico di massa si ottiene collegando tra loro tutti gli ancoraggi di massa con filo di rame nudo di sezione elevata. I ritorni di massa vengono effettuati, in tal caso, direttamente sul conduttore unico.

Ancoraggi di massa in intimo contatto elettromeccanico con il telaio.

I trasformatori di media frequenza, come abbiamo detto, sono rinchiusi in contenitori metallici (quasi sempre di alluminio), che fungono da schermi elettromagnetici; tali schermi devono essere ben collegati a massa se si vuole garantire ad essi la loro precisa funzione radioelettrica: quella di impedire che eventuali frequenze presenti nelle vicinanze dei trasformatori stessi possano giungere direttamente ad essi e quella di impedire che la media frequenza, presente nei trasformatori, possa interferire con altri segnali presenti nel ricevitore radio. È dunque importante che lo schermo di ogni trasformatore di media frequenza risulti ben collegato a massa, se si vogliono evitare fischi ed inneschi. Anche la carcassa metallica del condensatore variabile deve risultare ben collegata a massa; in tutti i condensatori variabili è presente un ancoraggio collegato meccanicamente con le lamine mobili, che deve essere collegato a massa.

CABLAGGIO del CIRCUITO SUPERETERODINA

Terminato il lavoro di cacciavite e di pinze, passiamo ora a quello di saldatore e cioè al cablaggio. I primi collegamenti da effettuarsi sono quelli relativi al trasformatore di alimentazione. In questo caso è necessario servirsi dell'apposito cartellino indicatore di cui tutti i trasformatori di alimentazione sono corredati all'atto dell'acquisto. In esso è possibile rilevare a che cosa corrisponde ciascun filo uscente, perchè vi è rappresentato lo schema elettrico del trasformatore con le indicazioni di tutte le tensioni in corrispondenza dei vari colori dei conduttori. È difficile, dunque, sbagliarsi con l'aiuto di questo cartellino. Cominciate, allora, a preparare il cordone di alimentazione con la spina e fissatene un capo al terminale distanziato del cambiotensione e l'altro ad un terminale dell'interruttore incorporato nel potenziometro di volume R6.

Saldate quindi i conduttori relativi all'avvolgimento primario di T2 agli altri terminali del cambiotensione, ricordando che il terminale corrispondente allo zero del trasformatore va collegato all'altro terminale dell'interruttore S1.

Il cambiotensione.

Si può ora procedere alla saldatura dei conduttori relativi all'avvolgimento secondario.

Dei due terminali dell'avvolgimento secondario a 190 volt uno risulta collegato al terminale del raddrizzatore al selenio RS1 contrassegnato con il simbolo della corrente alternata, l'altro terminale è collegato a massa. Per il collegamento del secondario a 6,3 volt le cose sono più semplici. Basta collegare a massa un terminale e connettere l'altro ai piedini 5 delle valvole V2 e V3 e al piedino 4 della valvola V1. In un secondo tempo si provvederà a collegare a massa i piedini 4 delle valvole V2 e V3 e il piedino 3 della valvola V1. Il circuito di accensione delle tre valvole può considerarsi ora ultimato. Completato il collegamento del trasformatore di alimentazione, si procederà nel cablaggio con la saldatura del terminale positivo del raddrizzatore con quello del condensatore C13, e si procederà poi alla saldatura di tutti gli altri elementi seguendo attentamente lo schema pratico. È consigliabile segnare con una crocetta, sullo schema elettrico, tutti i componenti e i vari conduttori che via via vengono saldati durante la fase di cablaggio; solo così si potrà essere certi di non dimenticare nulla e di non commettere errori. Il continuo controllo della corrispondenza fra lo schema elettrico, disegnato tutto in simboli, e quello pratico, che si sviluppa a mano a mano che si procede con il lavoro di cablaggio, rappresenta un esercizio utilissimo per l'allievo, perchè abitua a ragionare e a formare una mentalità radiotecnica.

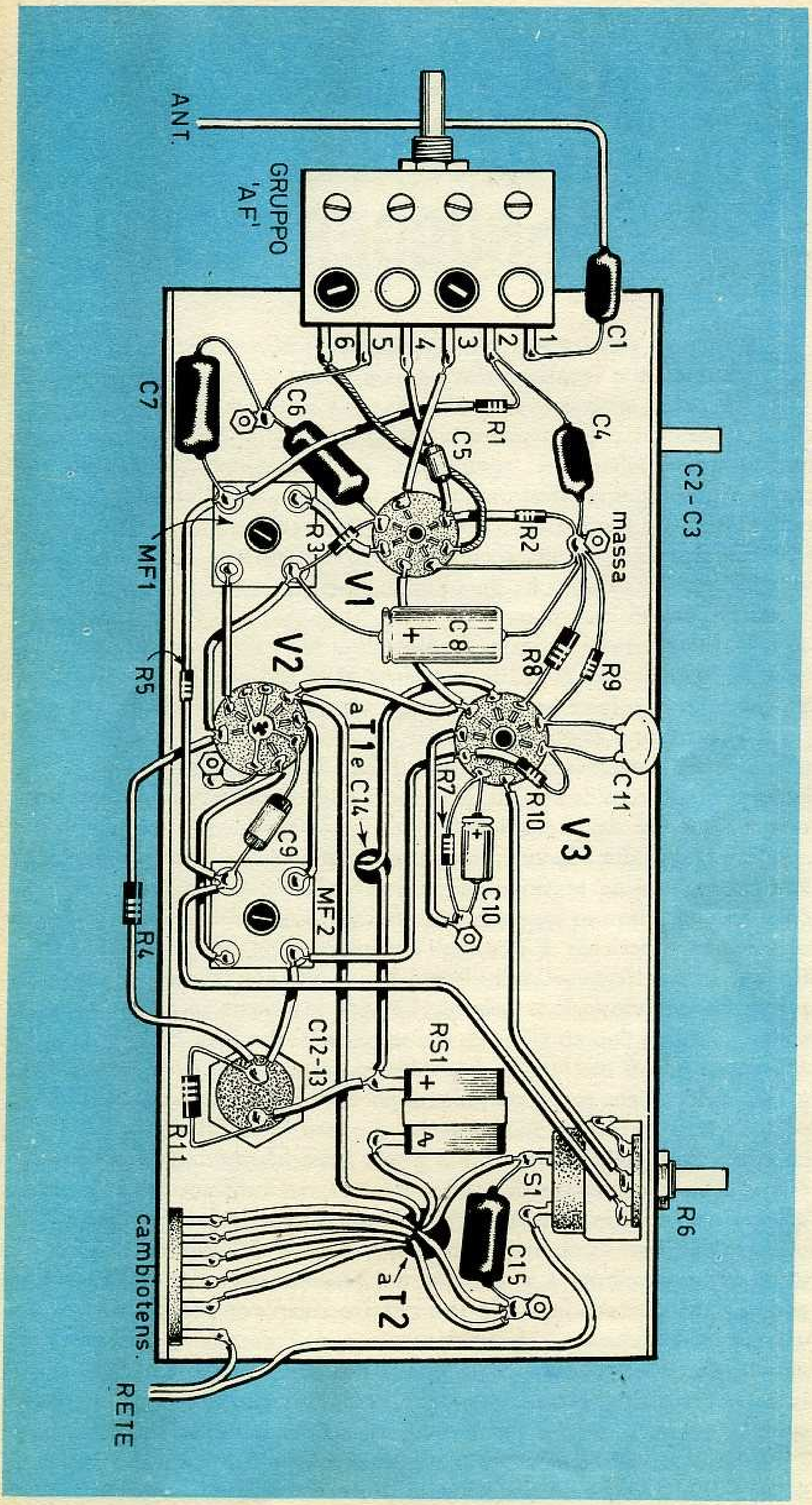
ANT.
GRUPPO 'A F'
C1
1
2
3
4
5
6
C2-C3
C4
C5
C6
C7
R1
R2
R3
massa
MF1
V1
C8
+
R8
R9
V2
R5
a T1 e C14
C11
R7
R10
V3
C9
C10
+
MF2
R4
C12-13
RS1
+
R11
S1
R6
C15
cambiotens.
a T2
RETE

Schema pratico del ricevitore con circuito supereterodina a tre valvole. Il condensatore variabile è applicato nella parte superiore del telaio.

CONSIGLI E SUGGERIMENTI

Prima di passare alla fase finale della realizzazione della supereterodina, che consiste nella messa a punto e taratura del ricevitore, si rende necessario elencare una serie di avvertimenti e di consigli che risulteranno della massima utilità per ogni allievo radiomontatore. Nel fissare al telaio quei componenti che richiedono l'impiego di viti come, ad esempio, il trasformatore di alimentazione, il trasformatore di uscita, le medie frequenze, il condensatore variabile, il condensatore elettrolitico doppio a vitone, occorrerà sempre ricordarsi di stringere bene le viti, perchè esse devono garantire un perfetto contatto tra il telaio e le varie parti ad esso fissate.

Collegamenti corti e schermature perfette.

Il collegamento tra la media frequenza MF1 e il piedino 2 della valvola V2, e quello tra la media frequenza MF2 e il piedino 8 della valvola V2 dovranno risultare corti il più possibile. Anche i collegamenti al gruppo A.F. dovranno essere molto corti. Devono pure risultare molto corti i terminali del condensatore di fuga sul C.A.V. C4. Tutte le saldature debbono risultare « calde » e prima di ogni saldatura occorre raschiare i terminali in modo da liberarli dall'ossido e da eventuali altre impurità. I conduttori dell'alta tensione, provenienti dalla cellula di filtro, cioè quelli che convogliano la tensione anodica alle placche e alle griglie schermo delle valvole, devono essere mantenuti distanziati dal telaio in modo da scongiurare il pericolo di eventuali cortocircuiti. Il condensatore elettrolitico catodico C10 e quello di filtro dell'alta tensione C8 devono essere applicati rispettando le loro polarità. Quando si prende in mano il condensatore variabile doppio C2 - C3, per applicarlo al telaio, bisogna far bene attenzione a non mettere le dita della mano sulle lamine e, soprattutto, a non farlo cadere per terra. Il gruppo A.F. va trattato con delicatezza, perchè potrebbe essere facile, anche durante le saldature, interrompere un conduttore delle bobine con la punta del saldatore. Sulla carcassa metallica del potenziometro di volume R6 è ricavato un terminale, che nello schema pratico non appare, e che va collegato a massa. Sui due fori di ingresso del conduttore di antenna e del cordone di alimentazione occorre mettere due gommini a protezione dei conduttori stessi, realizzando un nodo nei conduttori, in modo che essi incontrino resistenza nel caso dovessero essere soggetti a forze di trazione dall'esterno. Quando si è terminato anche il lavoro di cablaggio, si può dire di essere arrivati ad un buon punto della costruzione, ma non certamente alla fine. Occorrerà, infatti, accertarsi ancora una volta dell'esattezza del lavoro, ripassando punto per punto tutto lo schema teorico e quello pratico, in un lavoro di confronto con quello realizzato.

INTERPRETAZIONE SIMBOLICA e PIANO di CABLAGGIO

È assai importante che l'allievo radiotecnico mandi bene a memoria tutti i simboli caratteristici dei vari componenti radioelettrici. A tale scopo sono stati raggruppati, nelle pagine seguenti, i simboli fondamentali dei componenti del ricevitore radio con circuito supereterodina. Il solo gruppo di alta frequenza non possiede un particolare simbolo, perchè esso rappresenta un circuito, talvolta anche complesso, in cui sono contenuti condensatori, compensatori, bobine, contatti meccanici; certamente fra tutti questi elementi i più importanti sono quelli rappresentativi delle bobine di aereo e di oscillatore. Per ciascuna gamma d'onda esiste sempre una coppia di queste bobine e se il ricevitore è dotato delle onde medie e delle onde corte le bobine risulteranno complessivamente in numero di quattro.

Per quel che riguarda la preparazione del piano di cablaggio ricordiamo che i componenti devono sempre essere disposti sul telaio secondo un citerio logico. I componenti che partecipano alla composizione dello stadio di alta frequenza dovranno trovarsi ad una estremità del telaio, mentre quelli che partecipano alla formazione dello stadio amplificatore finale dovranno trovarsi all'estremità opposta; al centro del telaio risulteranno installati lo stadio di media frequenza, lo stadio rivelatore e gli eventuali stadi preamplificatori di bassa frequenza. In ogni caso, prima di iniziare un montaggio, l'allievo dovrà sempre dare un'occhiata all'elenco dei componenti, per vedere se già possiede qualche componente e recarsi subito dopo dall'abituale fornitore di materiali radioelettrici per acquistare quanto occorre. Il pezzo più importante (e proprio questo non appare quasi mai negli elenchi dei componenti!) è il telaio. Chi vuole risparmiare nella spesa, può costruirlo da sè, ricavandolo da una lamiera di alluminio. Ma chi vuole evitare questo fastidio lo può trovare già pronto in negozio. In questo caso si potrà acquistare anche la scala parlante, già adatta per il tipo di telaio scelto. Il telaio già pronto ha il vantaggio d'avere tutti i fori necessari per la applicazione dei vari componenti. In ogni caso, decidendo di introdurre l'apparecchio radio in un mobile, occorrerà sempre, prima di iniziare il montaggio, fare acquisto del mobile, perchè solo con il mobile sotto mano si possono stabilire le lunghezze esatte dei perni di comando dei potenziometri, del condensatore variabile, del gruppo di alta frequenza. La grandezza del mobile condiziona inoltre il diametro dell'altoparlante che si può applicare in esso. Ricordiamo che bisogna sempre evitare di segare i perni dei componenti quando questi sono già montati sul telaio, perchè il seghetto da ferro è un attrezzo pesante che finisce sempre per rovinare i componenti.

Il telaio rappresenta sempre il primo componente da acquistare o da costruire per qualsiasi tipo di montaggio.

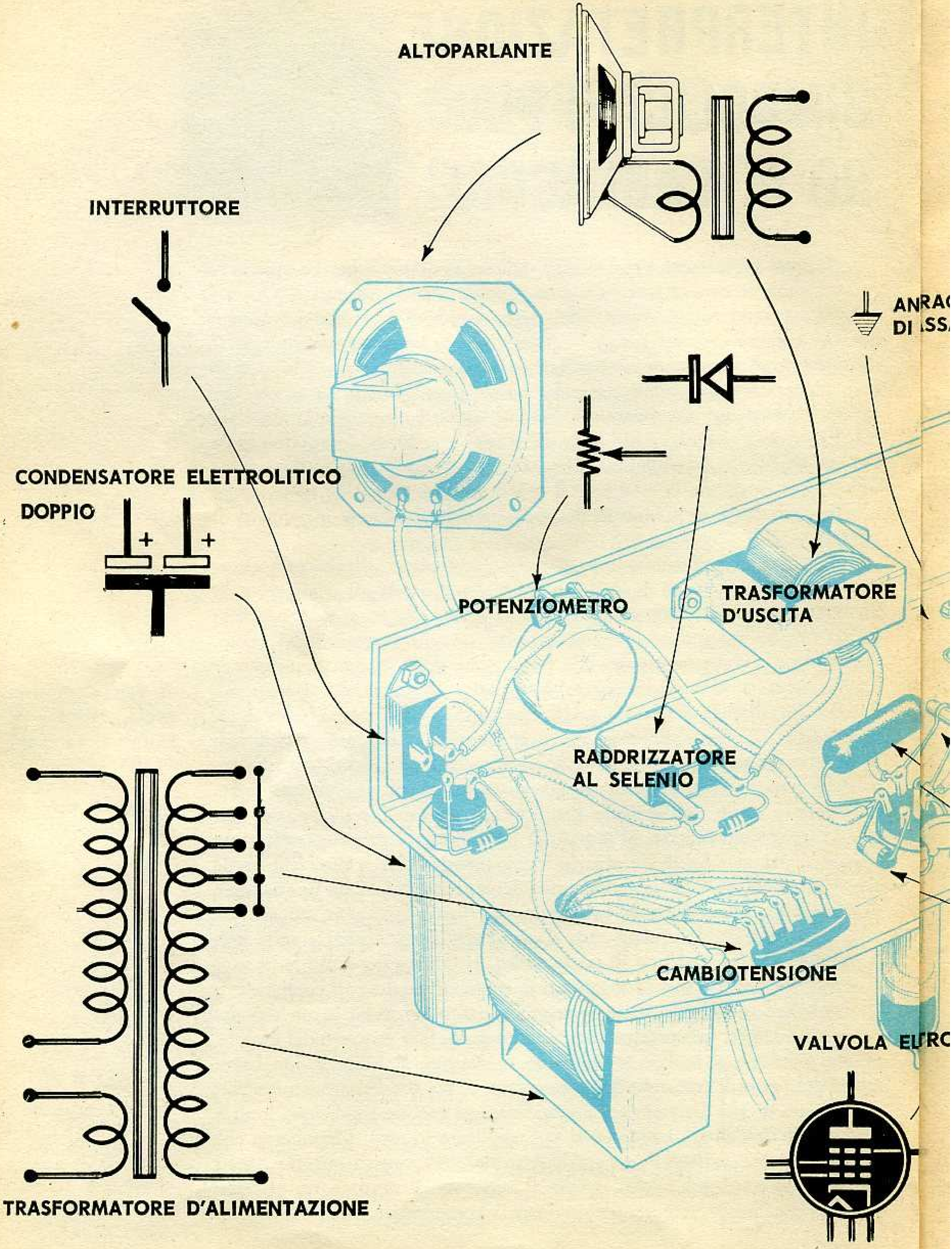
ALTOPARLANTE
INTERRUTTORE
CONDENSATORE ELETTROLITICO
DOPPIO
+
+
ANRAC
DI SSA
POTENZIOMETRO
TRASFORMATORE
D'USCITA
RADDRIZZATORE
AL SELENIO
CAMBIOTENSIONE
VALVOLA ELTRO
TRASFORMATORE D'ALIMENTAZIONE

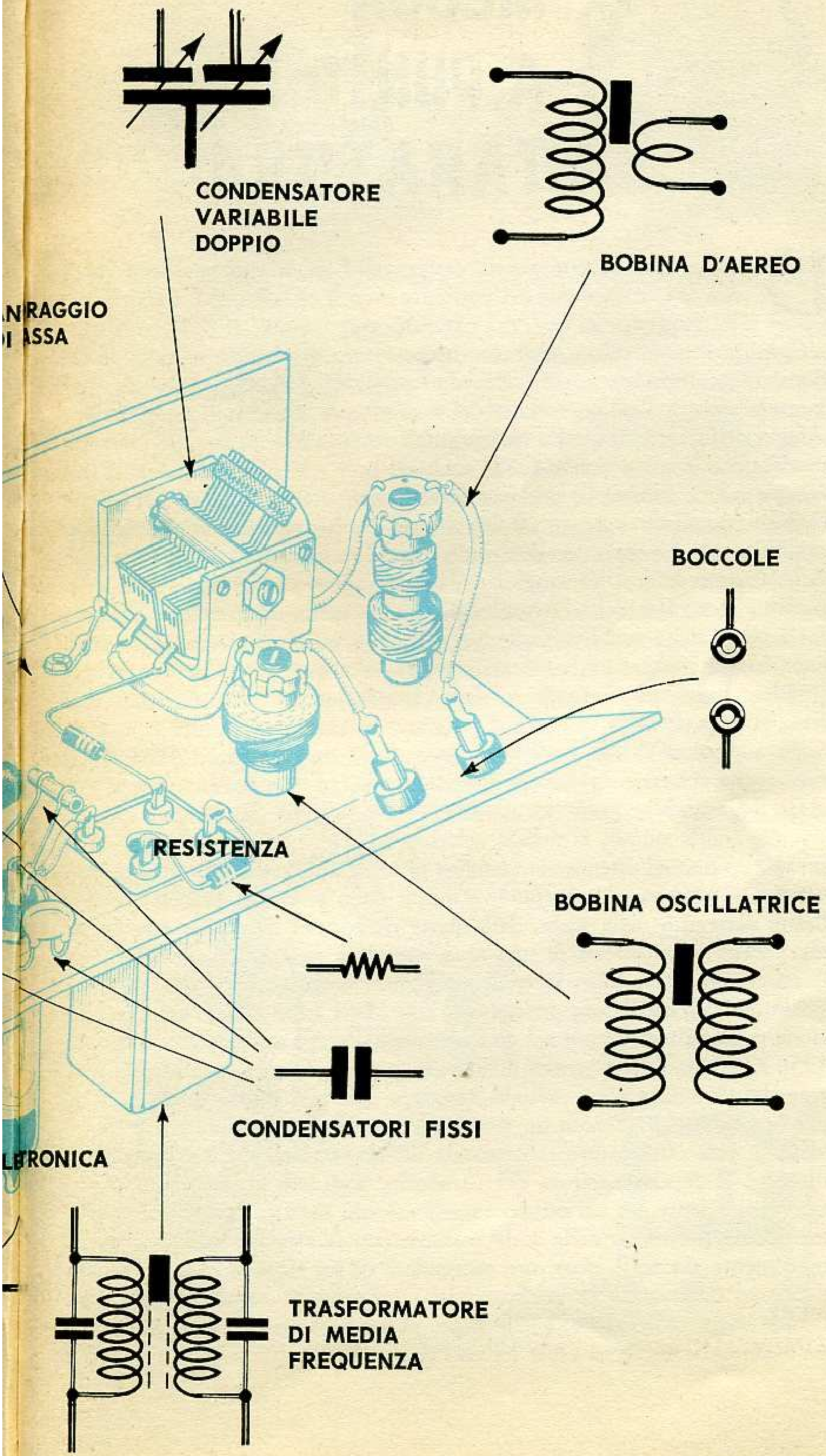
CONDENSATORE VARIABILE DOPPIO
BOBINA D'AEREO
RAGGIO
ASSA
BOCCOLE
RESISTENZA
BOBINA OSCILLATRICE
CONDENSATORI FISSI
RONICA
TRASFORMATORE DI MEDIA FREQUENZA

MESSA A PUNTO E TARATURA

Dopo essersi accertato dell'esattezza del montaggio, l'allievo potrà infilare le tre valvole nei rispettivi zoccoli ed accendere il ricevitore. Nel migliore dei casi, ma ciò non capita spesso, si potrà verificare un pronto funzionamento del ricevitore. In caso contrario occorre procedere alla taratura dei circuiti accordati prima di decidere se si sono commessi errori, oppure se vi sono dei componenti che non funzionano per difetto di fabbricazione. La taratura del ricevitore si esegue nel seguente modo: si sintonizza una stazione molto debole e si regolano i nuclei (inferiore e superiore) delle due medie frequenze, cominciando dalla seconda (MF2), fino ad ottenere un ascolto che sia il più potente possibile. Questa operazione va ripetuta per due o tre volte. Per quanto riguarda il gruppo A.F. si procede così. Si applica sulla parte anteriore del telaio una scala parlante, completa della meccanica di trascinamento di un indice. Si porta la lancetta della scala parlante verso quella estremità della scala dove sono indicate le stazioni che trasmettono sulla lunghezza d'onda più alta (minima frequenza) facendo corrispondere la lancetta con una emittente italiana nota di cui si conosce il programma trasmesso in quel momento. Si regola quindi il nucleo dell'oscillatore onde medie (osc. OM 600 KHz) fino a che si arriva a ricevere l'emittente su cui è stata fermata la lancetta della scala.

Si regola quindi il nucleo relativo all'aereo onde medie (ant. OM 600 KHz) fino ad ottenere la massima potenza d'uscita. Questa stessa operazione si esegue, poi, portando l'indice della scala del ricevitore verso l'altra estremità della scala, dalla parte delle onde più corte (frequenze alte) sopra l'indicazione di una nota emittente italiana di cui si conosce il programma trasmesso in quel momento. Si agisce dapprima sul compensatore dell'oscillatore onde medie (osc. OM 1250 KHz) fino alla ricezione dell'emittente e poi sul compensatore d'aereo onde medie (ant. OM 1250 KHz) fino ad ottenere la massima uscita. Questo metodo di taratura del ricevitore è un metodo empirico, che può essere seguito da chi per la prima volta monta un ricevitore supereterodina e da chi è sprovvisto di una attrezzatura strumentale di laboratorio, che ha lo scopo di ottenere la precisa taratura del ricevitore. Tuttavia questo metodo può essere sufficiente per un ottimo funzionamento del ricevitore. Chi volesse evitare l'applicazione della scala parlante con la relativa meccanica potrà ugualmente tarare il ricevitore servendosi di un ricevitore radio funzionante ed osservando le posizioni del condensatore variabile, cioè delle lamine mobili rispetto a quelle fisse, in corrispondenza delle due emittenti sintonizzate verso i due estremi di gamma.

I trasformatori di media frequenza vengono tarati introducendo negli appositi fori un cacciavite corto, munito di manico isolante, e avvitando o svitando i nuclei di ferrite.

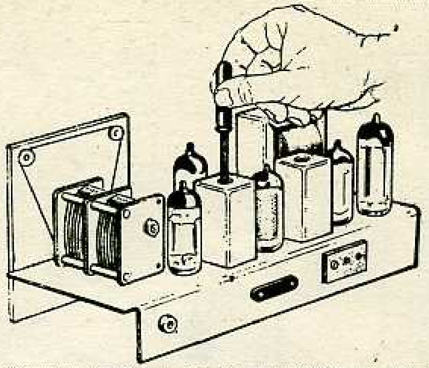

La taratura corretta e precisa del ricevitore supereterodina si esegue mediante l'impiego di un particolare strumento: l'oscillatore modulato.

Prima cosa da fare è quella di dissaldare il collegamento sulla griglia controllo di V2 (piedino 2 dello zoccolo). Su questo stesso piedino di V2 va collegato il cavo proveniente dallo strumento che sarà stato, in precedenza, tarato sulla frequenza di 467 Kc/s. Quindi si regolano i nuclei di MF2 fino ad ottenere, nell'altoparlante, la massima uscita (massima intensità del segnale prodotto dallo strumento). Chi disponesse di un voltmetro a corrente alternata, potrà collegare questo secondo strumento in parallelo alla bobina mobile dell'altoparlante. Ciò consentirà un controllo visivo più esatto di quello auditivo. Tarata così la seconda media frequenza, si stacca dal piedino della valvola V2 il cavo proveniente dall'oscillatore e si rifà il collegamento prima dissaldato. Questa stessa operazione di taratura va ripetuta ora per la prima media frequenza MF1. Si dissalda il collegamento alla griglia controllo di V1 (piedino 1) e si collega allo stesso piedino lo strumento, ricordandosi di collegare a massa, provvisoriamente, la griglia oscillatrice di V1 (piedino 1). Si regolano quindi i nuclei di MF1 fino ad ottenere la massima uscita nell'altoparlante.

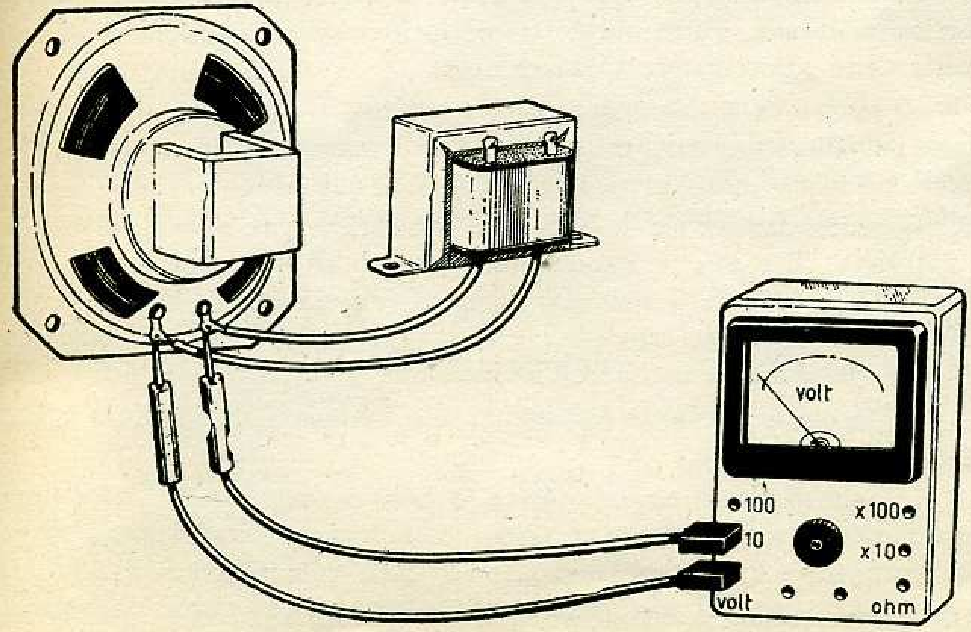

Durante la taratura delle medie frequenze si può seguire il procedimento di messa a punto osservando l'indice di un tester commutato nella posizione voltmetro. Il punto ottimo di taratura è raggiunto quando l'indice raggiunge la massima deviazione.

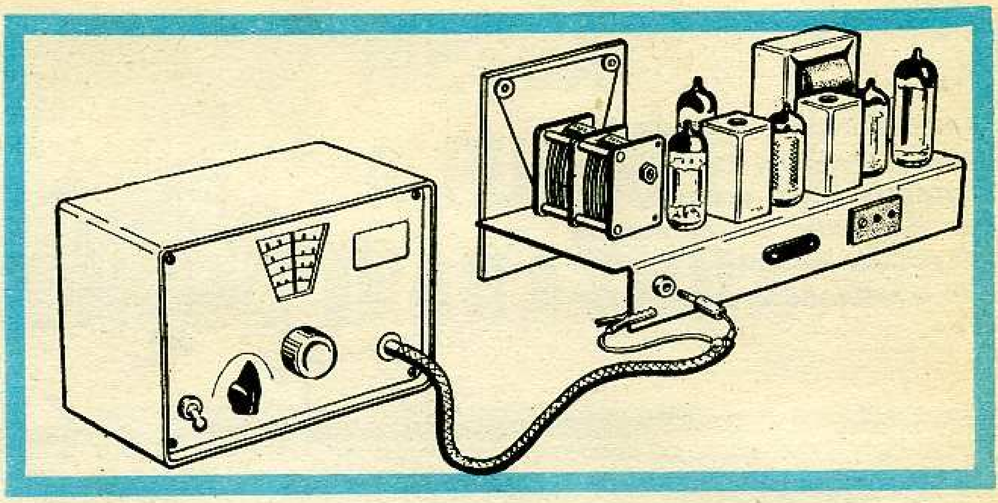

L'allineamento di un radioricevitore viene fatto mediante l'impiego dell'oscillatore modulato, il cui cavo schermato va inserito nella boccola di antenna del ricevitore; la calza metallica del cavo deve essere collegata al telaio dell'apparecchio.

A questo punto le due operazioni, ora eseguite, vanno ripetute per una seconda volta, ricominciando con MF2 e poi con MF1.

La taratura delle medie frequenze è così conclusa. Si tratta ora di accordare l'oscillatore del ricevitore.

Prima, però, di descrivere questa operazione, ricordiamo al lettore che anche il gruppo A.F. così come le medie frequenze e il trasformatore di alimentazione, è accompagnato, quando lo si acquista, da un foglietto con le indicazioni di riferimento ai nuclei e alle viti regolabili, che risultano collegate, internamente al gruppo, ai perni di piccoli condensatori variabili, che prendono il nome di compensatori.

Chiusa questa parentesi informativa procediamo con la taratura.

Per tarare l'oscillatore del ricevitore sulla frequenza di 600 Kc/s si inserisce lo strumento, tarato sulla frequenza di 600 Kc/s, nella presa d'antenna del ricevitore, interponendo un condensatore da 200 pF tra la presa d'antenna e il cavo proveniente dallo strumento. Si regola la sintonia del ricevitore in modo che la lancetta della scala parlante coincida con la frequenza di 600 Kc/s. Si tenga presente che la calza metallica del conduttore proveniente dallo strumento va collegata al telaio dell'apparecchio ricevitore. Fatto questo si regola il nucleo del gruppo A.F. che corrisponde all'oscillatore OM 600 Kc/s.

Il nucleo va regolato in modo da ottenere la massima uscita del segnale nell'altoparlante del ricevitore. Successivamente si agisce nello stesso modo sul nucleo che corrisponde all'ant. OM 600 KHz.

Si sintonizza poi sia lo strumento che il ricevitore sulla frequenza di 1.250 Kc/s. Si regola il compensatore relativo all'oscillatore onde medie (osc. OM 1.250 KHz) e quindi il compensatore d'aereo onde medie (ant. OM 1.250 KHz). Anche in questi casi la regolazione va fatta in modo da ottenere la massima uscita nell'altoparlante.

Si ripete nuovamente la taratura alla frequenza di 600 Kc/s indi a quella di 1.250 Kc/s.

A questo punto occorre tarare la gamma onde corte. L'oscillatore modulato risulterà collegato al ricevitore sempre nel modo prima citato. Lo strumento e il ricevitore verranno accordati sulla frequenza di 20 MHz e si regolerà prima il compensatore d'oscillatore e poi quello di antenna.

SONO disponibili annate ARRETRATE

di *Il* "SISTEMA "A"

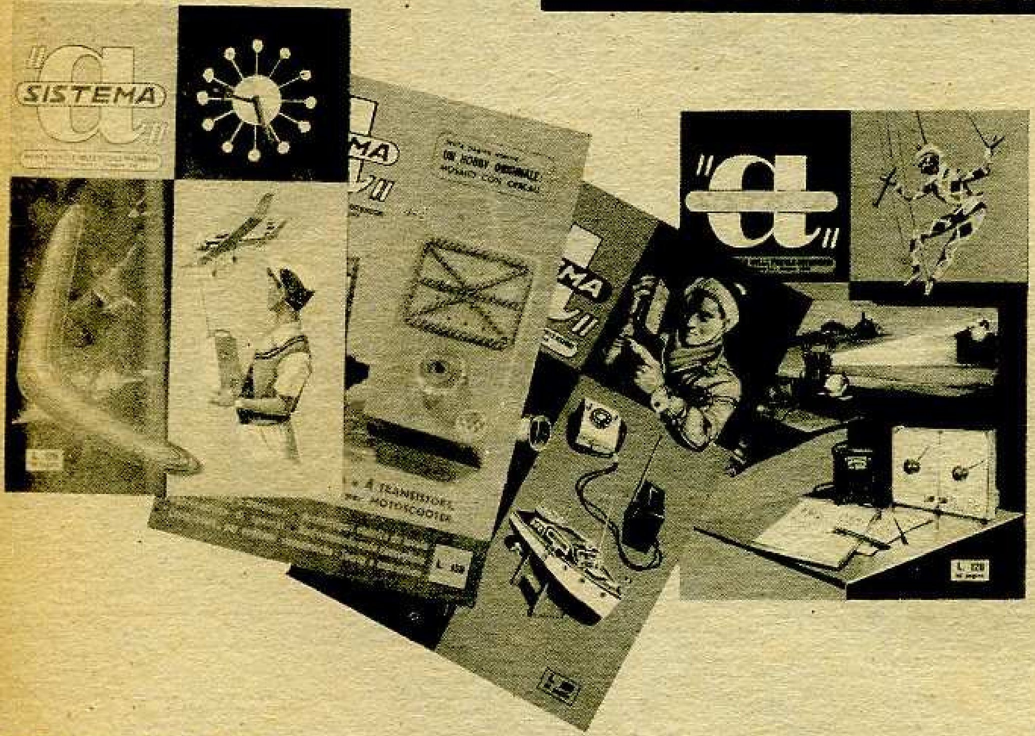

SE VI MANCA ***un'annata per completare la raccolta di questa interessante "PICCOLA ENCICLOPEDIA" per arrangisti, è il momento per approfittarne***

POSSIAMO INVIARVI dietro semplice richiesta, con pagamento anticipato o in contrassegno le seguenti annate:

1955 . . .	**L. 2000**	**1959 . . .**	**L. 2000**
1956 . . .	**L. 2000**	**1960 . . .**	**L. 2000**
1957 . . .	**L. 2000**	**1961 . . .**	**L. 2000**
1958 . . .	**L. 2000**	**1962 . . .**	**L. 2000**

indirizzate le vostre richieste a :

"SISTEMA A" Via Gluck, 59 - Milano

rimettendo l'importo sul conto corrente postale n. 3/49018

RADIOMANUALE
10 MANUALI IN 1
1 - Utensili, attrezzi, strumenti del radiolaboratorio
2 - Come si ripara il ricevitore a valvole
3 - Come si ripara il ricevitore a transistori
4 - Calcolo dei
nenti radio -
Tabelle - D.
Dati utili
5 - Tabelle di sostituzione dei transistori
6 - Progetti pratici di ricevitori a valvole e a transistori
7 - Progetti pratici di trasmettitori a valvole e a transistori
8 - Progetti pratici di amplificatori a valvole e a transistori
9 - Prontuario delle valvole europee
10 - Prontuario delle valvole americane
EDIZIONI CERVINIA - MILANO

TUTTO TRANSISTOR
EDIZIONI CERVINIA - MILANO

10 Manuali in 1: un libro che per l'appassionato di radiotecnica è più prezioso dell'esperienza stessa; 340 pagine, L. 3.000.

Novità 1966. Un autentico ferro di mestiere per il laboratorio, di agevole consultazione e di utile conforto per tutti; 300 pagine, L. 3.000.

I NOSTRI LIBRI DI SUCCESSO

È un'opera di consultazione immediata che risolve, completa, chiarisce ogni dubbio tecnico; 2200 voci, 600 pagine, 300 illustrazioni, L. 3.500.

Ogni progetto è corredato da fotografie, da schemi elettrici e pratici oltre ad una chiara descrizione delle fasi di montaggio. L. 500.

ENCICLOPEDIA
di
tecnica pratica
EDIZIONI CERVINIA - MILANO

20 novità PROGETTI
20 REALIZZAZIONI
20 SUCCESSI
a TRANSISTOR e a VALVOLE
L. 500

Per entrare in possesso di queste pubblicazioni basta farne richiesta direttamente alle EDIZIONI CERVINIA Via Gluck, 59 Milano, inviando anticipatamente l'importo a mezzo vaglia postale, o c.c.p. n° 3/49018.

[11] Patent Number: 4,685,047
[45] Date of Patent: Aug. 4, 1987

[54] APPARATUS FOR CONVERTING RADIO FREQUENCY ENERGY TO DIRECT CURRENT

[76] Inventor: **Raymond P. Phillips, Sr.**, 1200 S. Frances, Terrell, Tex. 75160

[21] Appl. No.: **886,138**
[22] Filed: **Jul. 16, 1986**
[51] **Int. Cl.**⁴ **H02M 7/06**
[52] **U.S. Cl.** **363/126**; 307/151
[58] **Field of Search** 363/125, 126; 307/151

[56] References Cited

U.S. PATENT DOCUMENTS

649,621	11/1901	Tesla	455/39
685,957	5/1902	Tesla	455/619
3,434,678	3/1969	Brown et al.	307/151
3,445,767	5/1969	Beard	324/78
3,678,365	8/1972	Fremiot	307/151
3,723,891	3/1973	Whitely	307/246
3,760,257	9/1973	Fletcher et al.	307/151
3,784,845	1/1974	Haas	307/519
3,792,290	2/1974	Brocker	307/522
3,944,935	3/1976	Plant	307/519
4,079,268	3/1978	Fletcher et al.	307/151
4,079,325	3/1978	Mawhinney et al.	307/261
4,085,377	4/1978	Turski et al.	307/519
4,195,237	3/1980	Meslener	307/261
4,297,557	10/1981	Tyler et al.	219/10.55 B
4,360,741	11/1982	Fitzsimmons et al.	307/151
4,360,865	11/1982	Yasumura et al.	363/126
4,393,441	8/1983	Enge	363/126

Primary Examiner—Patrick R. Salce
Assistant Examiner—Jeffrey Sterrett
Attorney, Agent, or Firm—Millen & White

[57] ABSTRACT

Apparatus for and methods of converting radio frequency energy into direct current for generating electric power includes a dipolar antenna for receiving radio frequency energy and a circuit connected thereto for converting the radio frequency energy to direct current. The circuit has a positive output line connected to one pole of the antenna and a negative output line connected to the other pole of the antenna. A positive transmitting diode is in the positive output line and a negative transmitting diode is in the negative output line. First and second bus lines and a pair of tuned circuits of opposite polarity couple the positive output line and negative line to the bus line with one of the bus lines being connected to ground. Each tuned circuit includes a first bridging line connecting the positive output line to the first and second ground lines and a second bridging line connecting the negative output line to the first and second ground lines. Each bridging line has a diode therein oriented at a polarity which is reverse with respect to the input diode. The bridging lines of each tuned circuit are connected to one another by an inductance and have capacitors disposed between the diode and the bus lines. A direct current device is connected to the positive line of the circuit.

11 Claims, 1 Drawing Figure

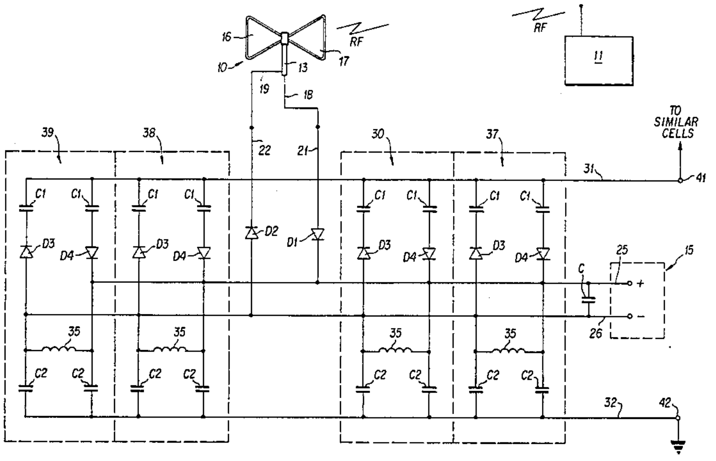

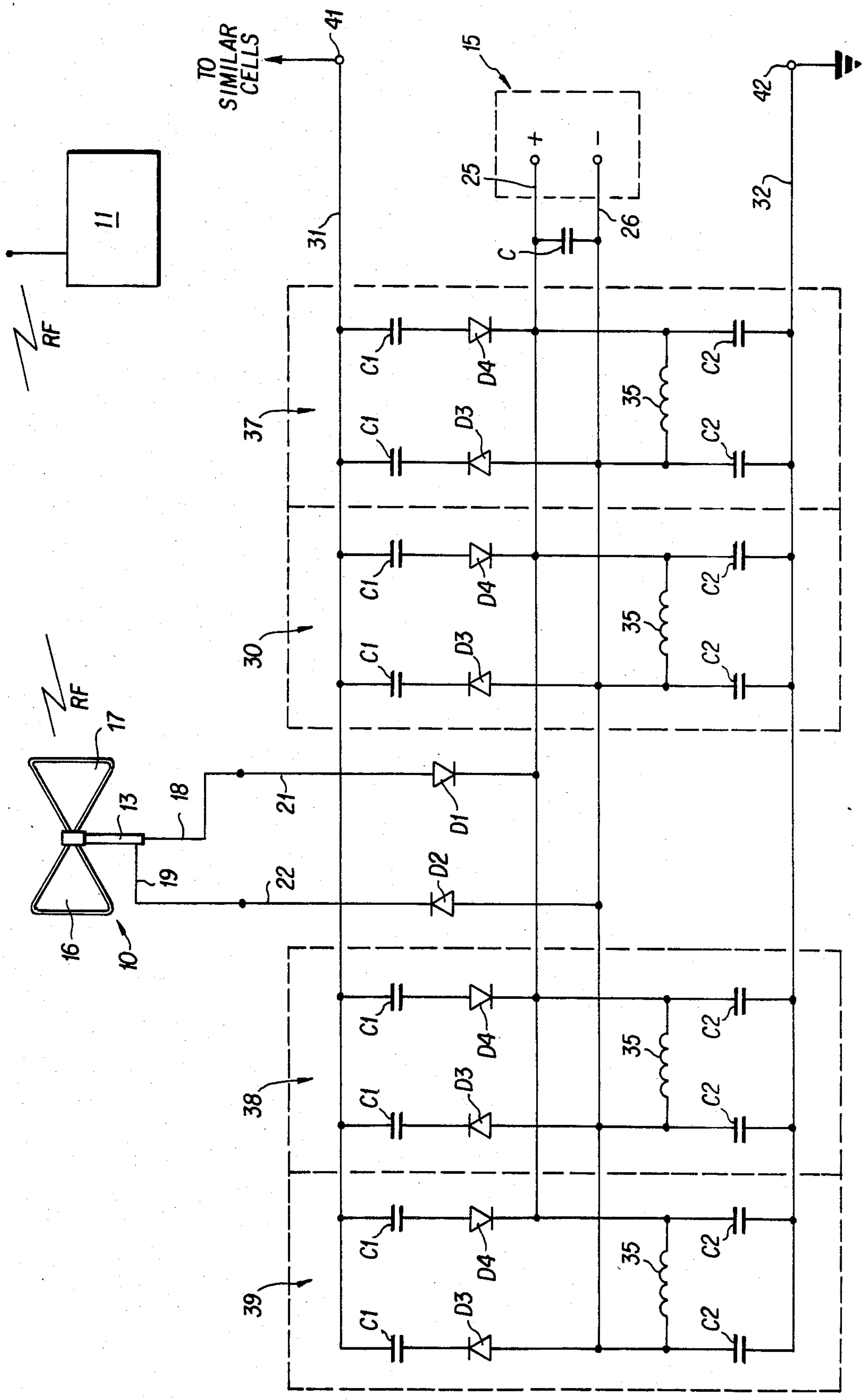
TO SIMILAR CELLS
+
−
C
RF
RF
C1
C2
D3
D4
D1
D2

APPARATUS FOR CONVERTING RADIO FREQUENCY ENERGY TO DIRECT CURRENT

BACKGROUND OF THE INVENTION

1. Field of the Invention

The instant invention relates to methods of an apparatus for converting radio frequency energy to direct current; more particularly, the instant invention relates to methods and apparatus for converting radio frequency energy to direct current wherein the direct current is of sufficient magnitude to power devices such as battery charges and electric motors without the use of amplification.

2. Technical Considerations in Prior Art

There has long been interest in technology directed to transmitting electrical energy over a distance without utilizing conductors, such as wire conductors. Development of such a technology has enormous potential. This was first recognized by Nikola Tesla who in 1899 constructed a 200 foot Tesla coil rated at 300 kilowatts and 150 kilocycles. Tesla hoped to set up standing waves of electrical energy around the whole surface of the earth, so that receiving antennas set at optimum points could tap the power when needed. Tesla was able to light hundreds of lamps at a distance of about 40 kilometers with his device without utilizing electrical conductors. The scheme has generally remained a scientific curiosity but has provided the initial groundwork for current developments wherein attempts are being made to transmit power by microwaves. However, power transmitted by microwaves is envisioned in the form of a beam of very high intensity which is focused from a microwave generator to a receiving antenna. This technology is envisioned as being used for many types of purposes such as transmitting microwave energy collected from gigantic solar power satellites and "star wars"-type weapons systems. However, the focused microwave beam is not suitable for many applications in that the beam must be directed toward a receiving antenna and cannot be transmitted through most objects, including living objects, without destroying the objects.

The instant invention relies on converting energy from standing waves which are emitted from radio frequency antennas in the RF range rather than the microwave range. Of particular interest are very low frequencies which are not used in communications and are available for transmitting power. Also of interest with respect to the instant invention are the low frequency waves emitted by the earth due to pulsation thereof caused by its magnetic field. These low frequency standing "earth" waves can be picked up by receivers tuned thereto.

SUMMARY OF THE INVENTION

It is an object of the invention to provide new and improved methods of an apparatus for converting radio frequency currents to direct current for practical uses other than communications, wherein the direct current energy converted from the radio frequency input energy does not require amplification.

Upon further study of the specification and appended claims, further objects and advantages of this invention will become apparent to those skilled in the art.

The instant invention contemplates an RF antenna for receiving radio waves. The RF antenna connected to a circuit configured in accordance with the principles

of the instant invention to convert the RF signals to direct current. The radio frequency signals received by the antenna are transmitted to first and second leads, with one lead being rectified to transmit positive voltage and the other lead being rectified to transmit negative voltage. The positive voltage lead being connected directly to a positive output line and the negative voltage lead being connected directly to a negative output line. The positive output line is connected to a pair of bus lines through a first pair of capacitors, while the negative output line is connected to the pair of bus lines by a second pair of capacitors. Disposed between the first bus line and the positive output line is a reverse diode of negative polarity, while disposed between the negative output line and first bus line is a reverse diode of positive polarity. The positive and negative output lines are connected to one another through an inductance which is in parallel with the capacitors of the first and second pair connected between the second bus line and the positive and negative output lines.

In accordance with one embodiment of the invention the afore-described circuit is duplicated for each positive and negative output line. In accordance with another embodiment of the invention, the afore-described circuitry is coupled to additional circuits identically configured in order to increase the direct current output of the arrangement.

In accordance with a further configuration of the invention, the antenna utilized is a dipolar antenna of aluminum wire arranged in a "butterfly" configuration.

The instant invention further contemplates the method of utilizing the afore-described elements so as to generate direct current having sufficient power to perform tasks such as charging batteries, lighting lamps and powering direct current electric motors without the use of amplifiers.

BRIEF DESCRIPTION OF THE DRAWING

Various other objects, features and attendant advantages of the present invention will be more fully appreciated as the same becomes better understood when considered in connection with the accompanying drawing, in which like reference characters designate the same or similar parts throughout the several views, and wherein:

There is shown a diagram of a circuit in accordance with the instant invention in combination with a driven device and a dipolar antenna which receives radio frequency waves which are converted to DC current for powering the driven device.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring now to the drawing, there is shown a dipolar antenna, designated generally by the numeral **10**, which receives radio frequency waves from an RF transmitter **11**. The radio frequency waves are transmitted to a radio frequency to direct current converting circuit, designated generally by the numeral **12**, through a coaxial cable **13** and DC current from the circuit **12** is used to power an output device, designated generally by the numeral **15**, which may for example by a battery charger DC motor or lighting device. The circuit **12** has no other power inputs other than the radio frequency energy transmitted thereto by the antenna **10** and therefore includes no amplifiers for amplifying the RF energy.

The source of radio frequencies convertible to direct current by the circuit shown may include sources of high frequency, low frequency (LF), very low frequency (VLF) and extremely low frequency (ELF) radio waves as well as seismic vibration of the earth's magnetic fields.

Preferably, the dipolar antenna **10** is formed of two triangular loops of aluminum wire **16** and **17**, one of which is connected to the annular conductor **18** of the coaxial cable **13** and the other of which is connected to the center conductor **19** of the coaxial cable. The size of the bipolar antenna **10** is dependent on the particular application to which it is put. In accordance with one embodiment of the invention, the antenna **10** is approximately 12 inches in width and 18 inches in length. Such an antenna is used to receive five watt energy, such as that generated by a walkie-talkie or citizen-band radio.

The annular conductor **18** of the coaxial cable **13** is connected to a positive lead **21** of the circuit **12**, while the center conductor **19** of the coaxial cable is connected to a negative lead **22** of the circuit. A positive transmitting diode D1 is disposed between the lead **21** and the remainder of the circuit **12** while a negative transmitting diode D2 is disposed between the lead **21** to a positive output line **25** while the negative diode D2 is connected to a negative output line **26**. Accordingly, the positive voltages with respect to ground are produced on output line **25** and negative voltages with respect to ground are produced on output line **26**.

In order to provide a DC output of sufficient power, a plurality of inductance-capacitance, RF, tuned circuits **30**, each forming a positive cell, or a negative cell, are utilized for connecting the positive output line **25** and negative output line **26** to first and second bus lines **31** and **32**, respectively. Bus line **32** is connected to ground while bus line **32** can be connected to circuits similar to circuit **12**. The positive output line **25** is connected by a first bridging line **33** to the first and second bus lines **31** and **32** while the negative output line **26** is connected by a second bridging line **34** to the first and second bus lines. The bridging line **33** has capacitors C1 and C2 disposed between the positive output line **25** and the first and second bus lines **31** and **32**, while the bridging line **34** also has capacitors C1 and C2 dispensed between the negative output line **26** and the first and second bus lines **31** and **32**. Connected between the bridging lines **33** and **34**, is an inductor **35** which serves as an RF choke, while disposed between the positive output line **25** and the capacitor C1 there is a negative polarity diode D4 referred to herein as a bridging diode and disposed between the negative output line **26** and capacitor C1 in line **34** there is a positive polarity diode D3 referred to herein as a bringing diode. As is seen of the drawing, the RF tuned circuit cell **30** is repeated a plurality of times. In the specific example shown, the circuit **12** has separate cells **30**, **37**, **38** and **39**. The cells **30** and **38** are of opposite polarity and balance one another while the cells **37** and **39** of opposite polarity and also balance one another. In order for the system to function, a pair of opposite polarized cells must be utilized. The particular number of cells **30** and the value of the components thereof are determined by the configuration of the dipolar antenna **10** and the power and frequency of the RF transmitter **11**.

The radio frequency to direct current conversion circuit **12** may itself be connected to a duplicate circuit via pins **41** so as to provide additional direct current output on lines similar to positive output line **25** and negative output line **26** the output lines may be connected together in order to boost the total output of the system.

An operative embodiment of the invention utilizes the following elements:

Diodes D1, D2, D3 and D4—Germanium Diodes, Archer **1** N34A,Catalog #1123.
Inductor **35**-47 Milli henry R. F. Choke
Capacitors C1 and C2-0.47 Pico Farads at 200 volts
Coaxial Cable **13**—50 ohms
Dipolar Antenna **10**—aluminum wire triangular loops approximately 12 inches by 18 inches.

From the foregoing description, one skilled in the art can easily ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

What is claimed is:

1. Apparatus for converting radio frequency energy into direct current for generating electric power, the apparatus comprising:

a dipolar input for receiving radio frequency energy;

a positive output line connected to one pole of the dipolar input and a negative output line connected to the other pole of the dipolar input;

a positive transmitting input diode means in the positive output line and a negative transmitting input diode means in the negative output line;

first and second bus lines and a pair of tuned circuits of opposite polarity coupling the positive output line and negative output line to the bus lines, one of the bus lines being connected to ground;

each tuned circuit including a first bridging line connecting the positive output line to the first and second bus lines and a second bridging line connecting the negative output line to the first and second bus lines, each bridging line having a bridging diode means therein oriented at a polarity which is reverse with respect to the transmitting input diode means of the respective output line; the bridging lines of each tuned circuit being connected to one another by an inductance and having capacitors disposed between the bridging diode means thereon and the bus lines, and

a direct current device connected to the bridging lines of the tuned circuit.

2. The apparatus of claim **1**, wherein there are a plurality of similarly configured tuned circuits connected between the output lines and the bus lines.

3. The apparatus of claim **2**, wherein there are a plurality of radio frequency-to-direct current conversion circuits connected to one another to provide a direct current power array tuned to a specific radio frequency such as a high frequency source, low frequency source, very low frequency source, extremely low frequency source, or source created by seismic vibrations of the earth's magnetic field.

4. The apparatus of claim **3** wherein the dipolar input is a dipolar antenna.

5. The apparatus of claim **1**, wherein the device connected to the output lines is a direct current motor.

6. The apparatus of claim **1**, wherein the device connected to the output line is an illuminating device.

7. The apparatus of claim **1**, wherein the device connected to the output lines is a battery charger.

8. The apparatus of claim **1**, wherein the device connected to the output lines is a DC-to-AC inverter.

9. The apparatus of claim **1**, wherein there are a plurality of radio frequency to direct current converter circuits connected to one another to form an array of circuits tuned to a particular radio frequency.

10. The apparatus of claim **1** wherein the dipolar input is a dipolar antenna.

11. The apparatus of claim **10**, wherein the dipolar antenna utilizes aluminum wire arranged in pair of triangular loops.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 4,685,047

DATED : August 4, 1987

INVENTOR(S) : Raymond P. Phillips, Sr.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

On the Title Page "11 Claims" should read -- 12 Claims --.

Column 4, Claim 1, Line 21:

Reads: "into direct current for generating electric power, the"

Should Read: --into direct current, the--

Column 4, Claim 5, Line 62:

Reads: "nected to the output lines is a direct current motor."

Should Read: --nected to the output lines is a direct-current motor.--

Column 4, Claim 8, Line 68:

Reads: "nected to the output lines is a DC-to-AC inverter. "

Should Read: --nected to the output lines is a DC-to-AC converter.--

Please add the following new claim:

--12. The apparatus of claim 1 wherein the dipolar input is a coazial cable.--

Signed and Sealed this

Third Day of November, 1987

Attest:

DONALD J. QUIGG

Attesting Officer

Commissioner of Patents and Trademarks

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 4,685,047 Page 1 of 4

DATED : August 4, 1987

INVENTOR(S) : Raymond P. Phillips, Sr.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

On the Title Page, Item [57],
IN THE ABSTRACT:

Line 1: Reads: "Apparatus for and methods of converting radio fre-"
Should Read: --Apparatus for converting radio fre- --

Line 9: "transmitting diode is in the positive output line and a"
--transmitting diode is located in the positive output line and a"

Line 10: "negative transmitting diode is in the negative output"
--negative transmitting diode is located in the negative output--

Line 24: "to the positive line of the circuit"
--to the output lines of the circuit--

COL. 1
Line 9: "the instant invention relates to methods of an appara-"
--the instant invention relates to appara- --

COL. 1
Line 8: "to methods and apparatus for converting radio fre-"
--to apparatus for converting radio fre- --

COL. 1
Line 14: "as battery charges and electric motors without the use"
--as battery chargers and electric motors without the use--

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 4,685,047

DATED : August 4, 1987

INVENTOR(S) : Raymond P. Phillips, Sr.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

COL. 1
Line 29: "tific curiosity but has provided the initial groundwork"
--tific curiosity, but has provided the initial groundwork--

COL. 1
Line 57: "improved methods of an apparatus for converting radio"
--improved apparatus for converting radio--

COL. 2
Line 6: "tive voltage. The positive voltage lead being connected"
--tive voltage. The positive voltage lead is connected--

COL. 2
Line 8: "age lead being connected directly to a negative output"
--age lead is connected directly to a negative output--

COL. 2
Line 58: "ted to a radio frequency to direct current converting"
--ted to a radio frequency-to-direct current converting--

COL. 2
Line 62: "by the numeral 15, which may for example by a battery"
--by the numeral 15, which may for example be a battery--

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 4,685,047 Page 3 of 4

DATED : August 4, 1987

INVENTOR(S) : Raymond P. Phillips, Sr.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

COL. 3
Line 23: "and the remainde of the circuit 12 while a negative"
--and the remainder of the circuit 12, while a negative--

COL. 3
Line 25: "to a positive output line 25 while the negative diode D2"
--a positive output line 25, while the negative transmitting diode D2--

COL. 3
Line 31: "a plurality of inductance-capacitance, RF, tuned cir-"
--a plurality of inductance-capacitance, RF-tuned cir- --

COL. 3
Line 32: "cuits 30, each forming a positive cell, or a negative cell,"
--cuits 30, each forming either a positive cell or a negative cell,--

COL. 3
Line 50: "polarity diode D4 referred to herein as a bridging diode"
--polarity diode D4, referred to herein as a bridging diode--

COL. 3
Line 53: "D3 referred to herein as a bringing diode. As is seen of"
--D3, referred to herein as a bridging diode. As is seen of--

COL. 3
Line 58: "other while the cells 37 and 39 of opposite polarity and"
--other, while the cells 37 and 39 of opposite polarity--

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 4,685,047 Page 4 of 4

DATED : August 4, 1987

INVENTOR(S) : Raymond P. Phillips, Sr.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

COL. 3
Line 67: "via pins 41 so as to provide additional direct current"
--via pin 41 so as to provide additional direct current--

COL. 4
Line 1: "negative output line 26 the output lines may be con-"
--negative output line 26. The output lines may be con- --

Claim 12, line 2: Change "coazial" to --coaxial--.

Signed and Sealed this

Eleventh Day of July, 1989

Attest:

DONALD J. QUIGG

Attesting Officer

Commissioner of Patents and Trademarks

United States Patent [19]
Berley

[11] **4,410,805**
[45] **Oct. 18, 1983**

[54] RADIOMETER GENERATOR

[76] Inventor: **Lawrence F. Berley**, 19 Chipmunk La., Media, Pa. 19063

[21] Appl. No.: **918,772**

[22] Filed: **Jun. 26, 1978**

[51] **Int. Cl.**^{3} **F03G 7/00**
[52] **U.S. Cl.** .... **290/1 R**; 356/216
[58] **Field of Search** .... 356/213, 216; 290/1 R; 60/641; 310/306; 73/355 R; 350/1.1; 250/215

[56] **References Cited**

U.S. PATENT DOCUMENTS

1,000,831	8/1911	Martin	356/216
2,004,392	6/1935	Rah	356/216 X
3,348,374	10/1967	Schalkowsky	310/306
4,146,407	3/1979	Litsenko et al.	136/89 PC

FOREIGN PATENT DOCUMENTS

Ad.10769	of 1890	United Kingdom	356/216
1395137	5/1975	United Kingdom	250/215

OTHER PUBLICATIONS

Series of papers published in the Philosophical Transactions of the Royal Society of London (1873–1881), par. 452.

"Radiant Matter", A Resume of the Principal Lectures and Papers by Prof. William Crookes on the Fourth State of Matter, published by James W. Queen & Co., No. 924, Chestnut Street, Philadelphia, Pa., pp. 6-9.

Primary Examiner—J. V. Truhe
Assistant Examiner—W. E. Duncanson, Jr.
Attorney, Agent, or Firm—John F. A. Earley

[57] **ABSTRACT**

A self-contained system embodying the principles of the radiometer, the otheoscope, and the dynamoelectric machine is provided to derive an output of electrical energy from an input of radiant energy, and in particular infra-red energy, from the surrounding environment, and for amplifying that output. The system is suitable for airborne and space as well as terrestrial applications.

27 Claims, 34 Drawing Figures

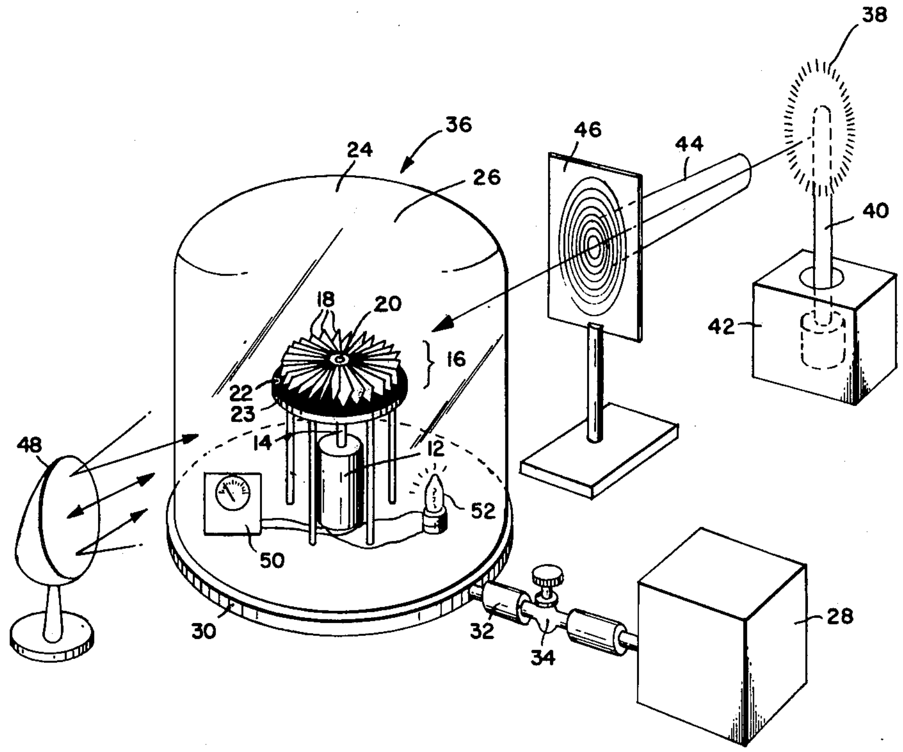

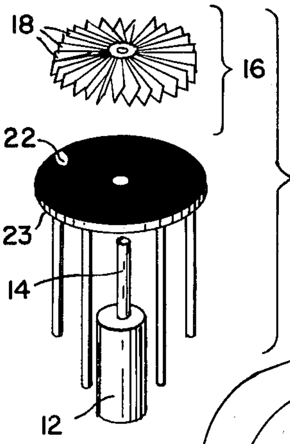
18
16
22
23
14
12

FIG. 3.

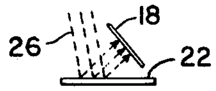

FIG. 4.

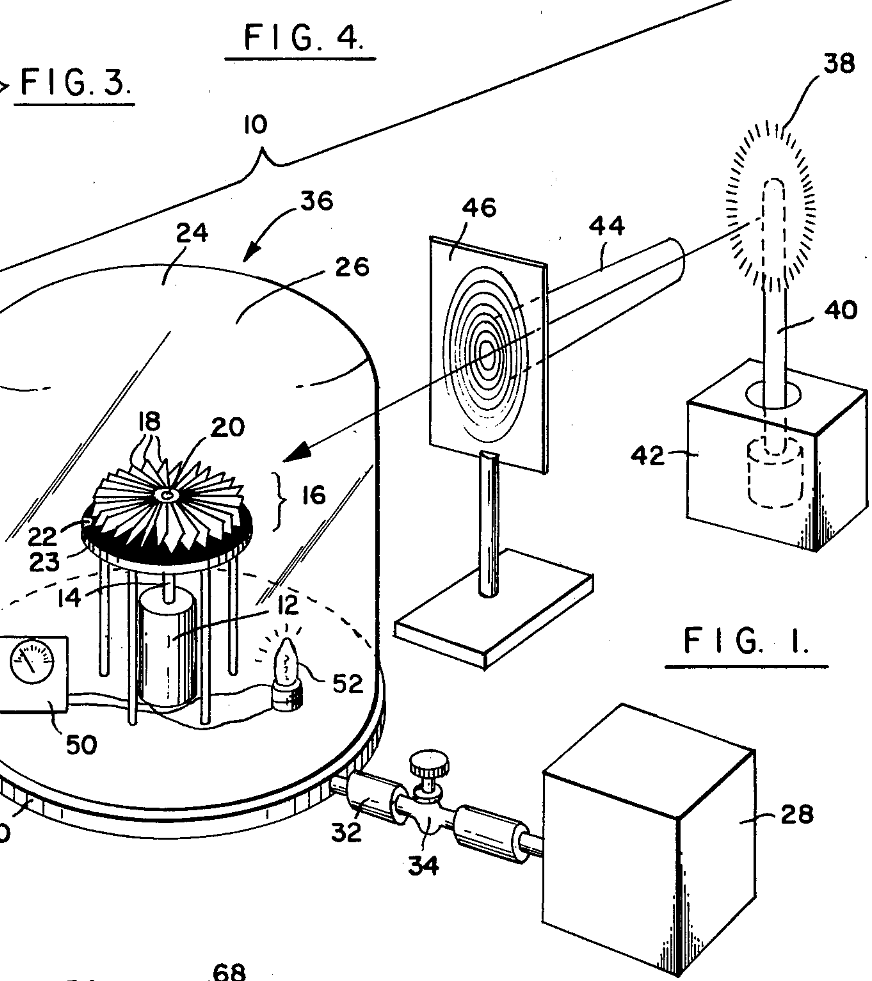
10
38
36
24
26
46
44
40
42
18
20
16
22
23
14
12
48
52
50
30
32
34
28

FIG. 1.

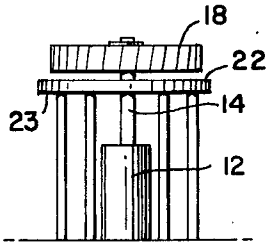
18
22
23
14
12

FIG. 2.

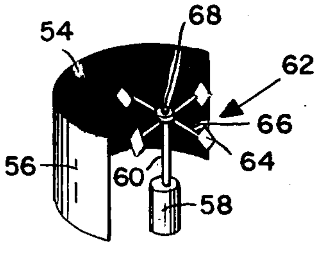
54
68
62
66
64
56
60
58

FIG. 5.

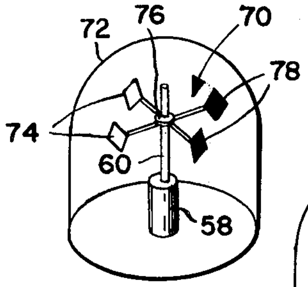
72
76
70
78
74
60
58

FIG. 6.

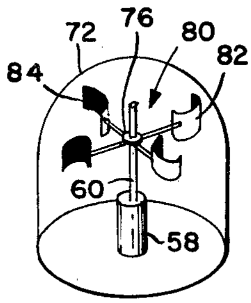
72
76
80
82
84
60
58

FIG.7.

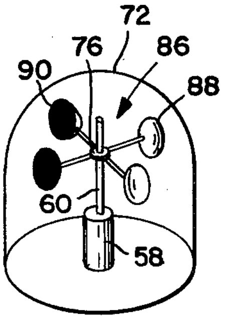
72
76
86
90
88
60
58

FIG. 8.

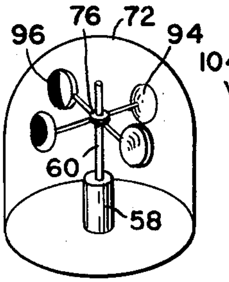
96
76
72
94
60
58

FIG. 9.

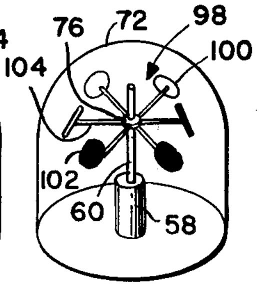
76
72
98
100
104
102
60
58

FIG. 10.

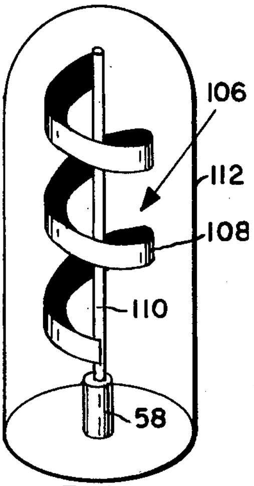
106
112
108
110
58

FIG. 11.

FIG. 12.

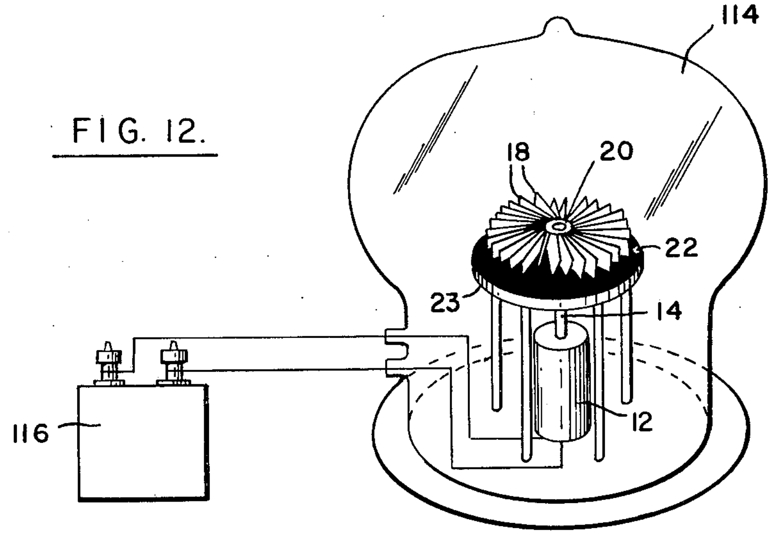
114
18
20
22
23
14
12
116

FIG. 13.

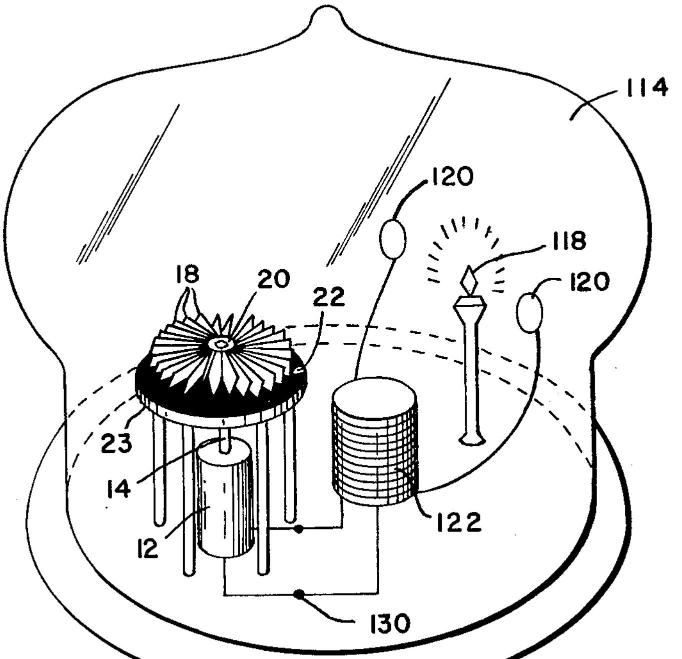
114
120
118
120
18
20
22
23
14
12
122
130

FIG. 14.

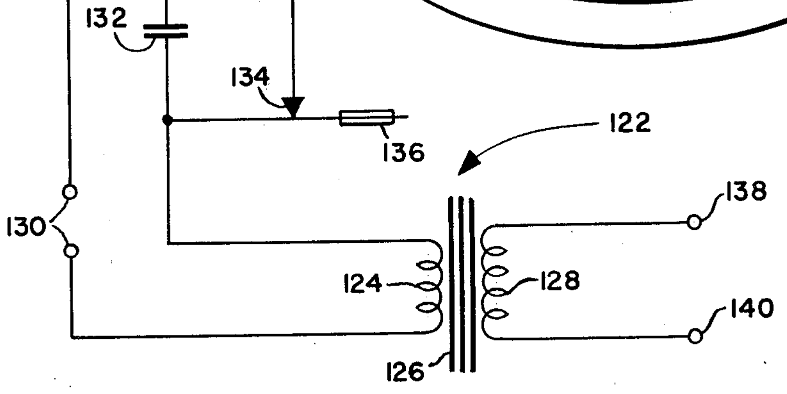
132
134
136
122
138
130
124
128
140
126

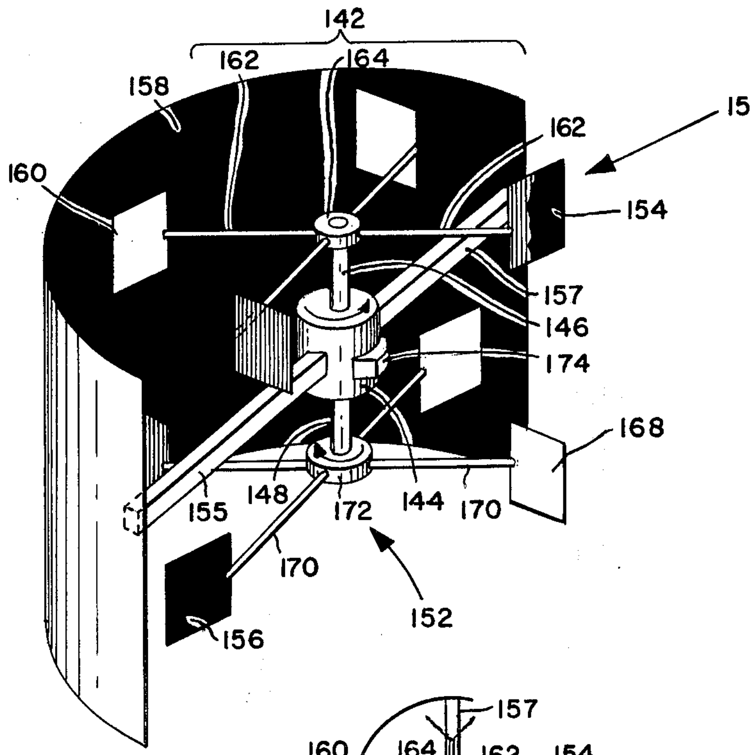
142
162
164
158
150
162
160
154
157
146
174
168
155
148
172
144
170
170
152
156

FIG. 15.

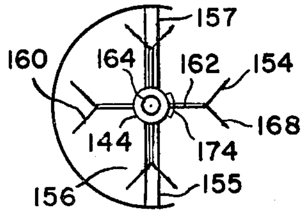
157
160
164
162
154
168
144
174
156
155

FIG. 16.

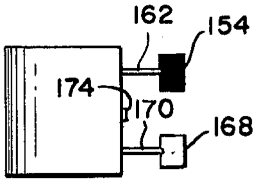
162
154
174
170
168

FIG. 17.

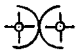

FIG. 18

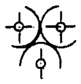

FIG. 19.

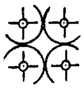

FIG. 20.

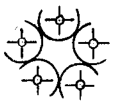

FIG. 21.

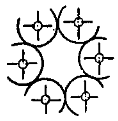

FIG. 22.

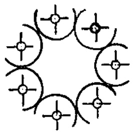

FIG. 23.

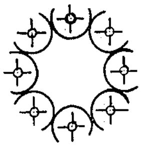

FIG. 24.

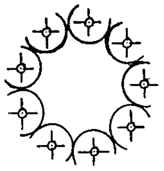

FIG. 25.

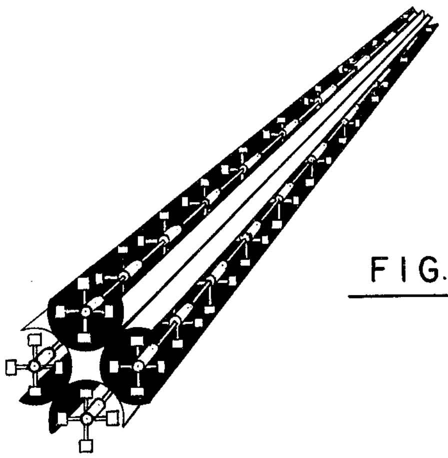

FIG. 26.

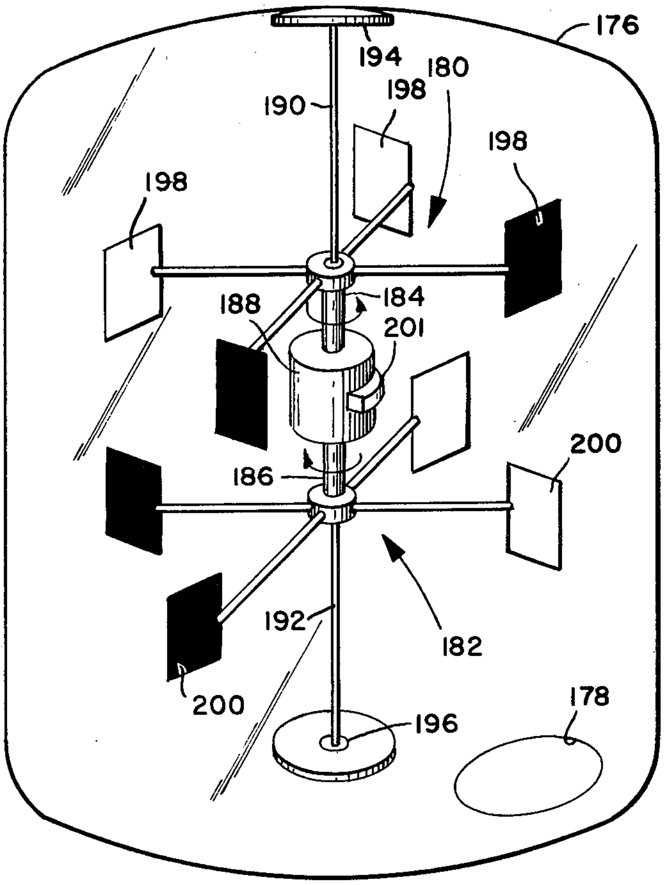
176
194
180
190
198
198
198
198
184
188
201
200
186
192
182
200
196
178

FIG. 27.

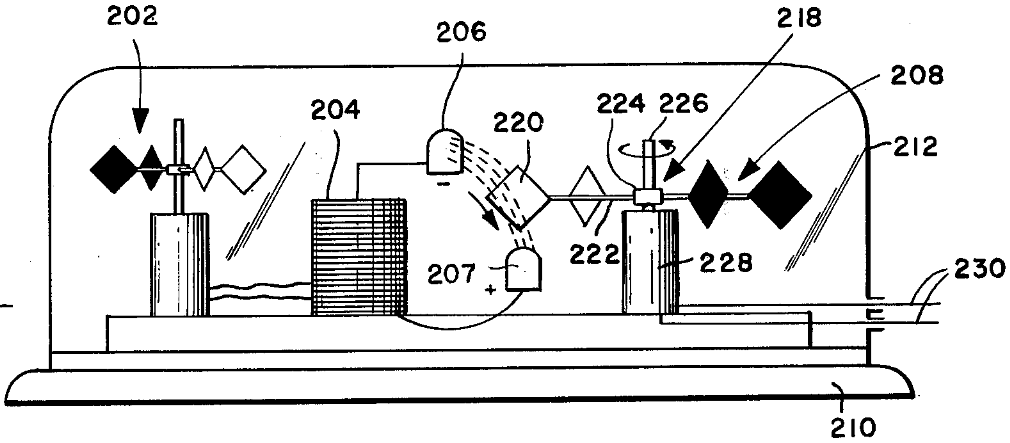
202
206
218
208
204
220
224
226
212
222
228
230
207
+
−
210

FIG. 28.

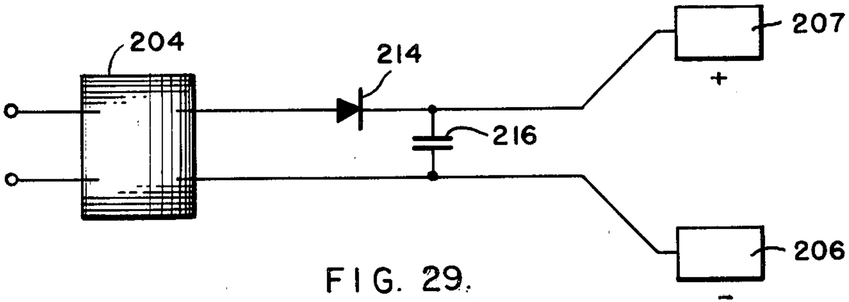
204
214
216
207
+
206
−

FIG. 29.

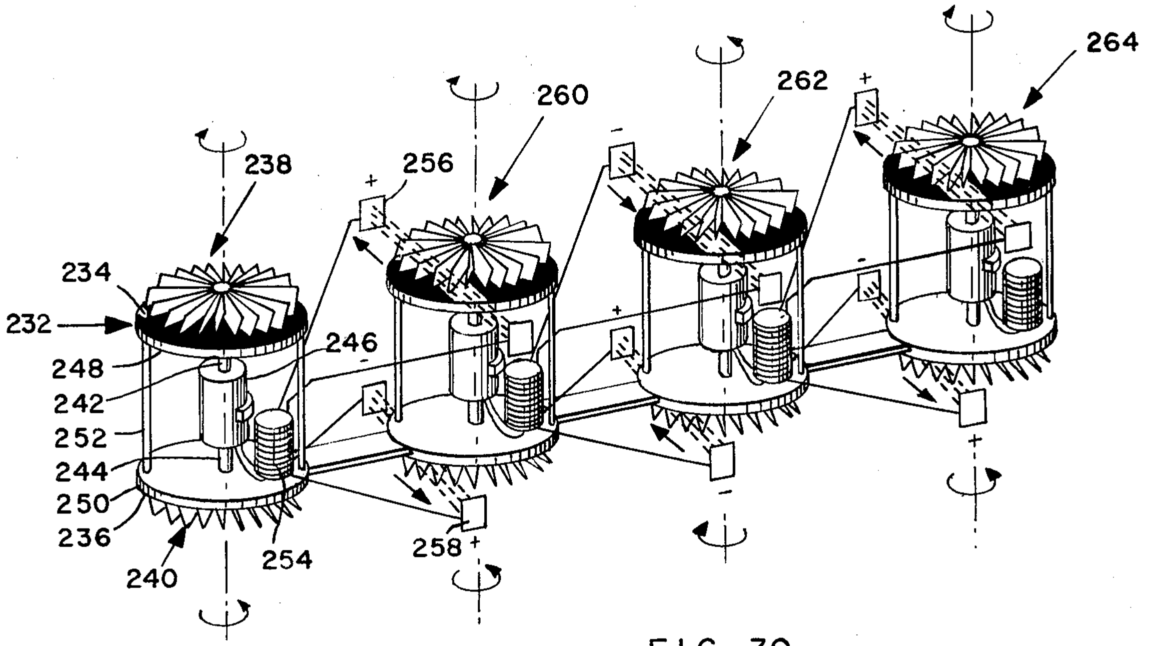

FIG. 30.

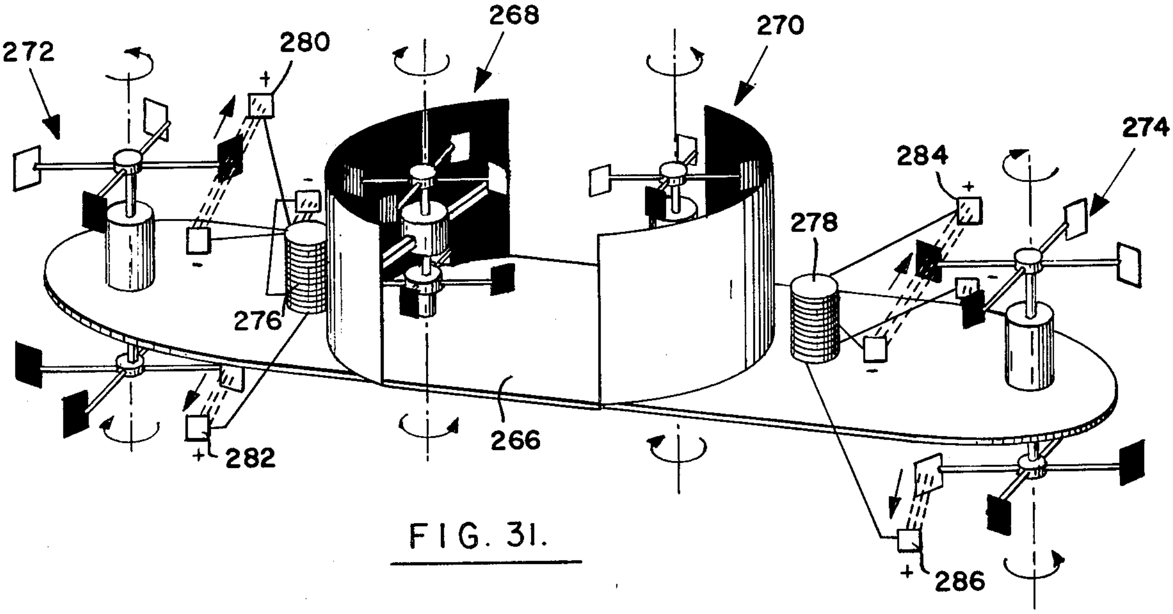

FIG. 31.

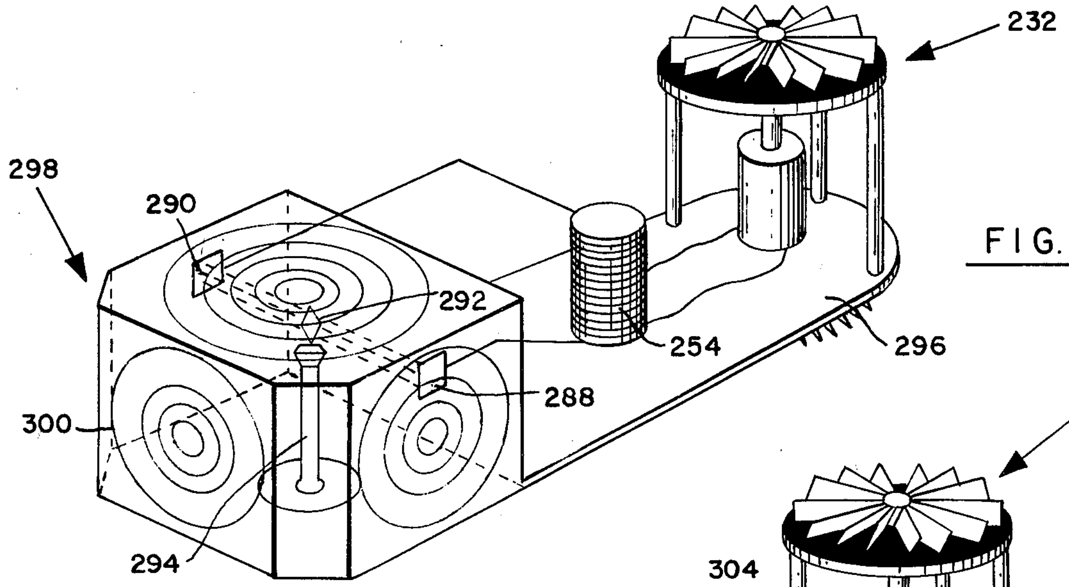
232
298
290
292
254
296
288
300
294

FIG. 32.

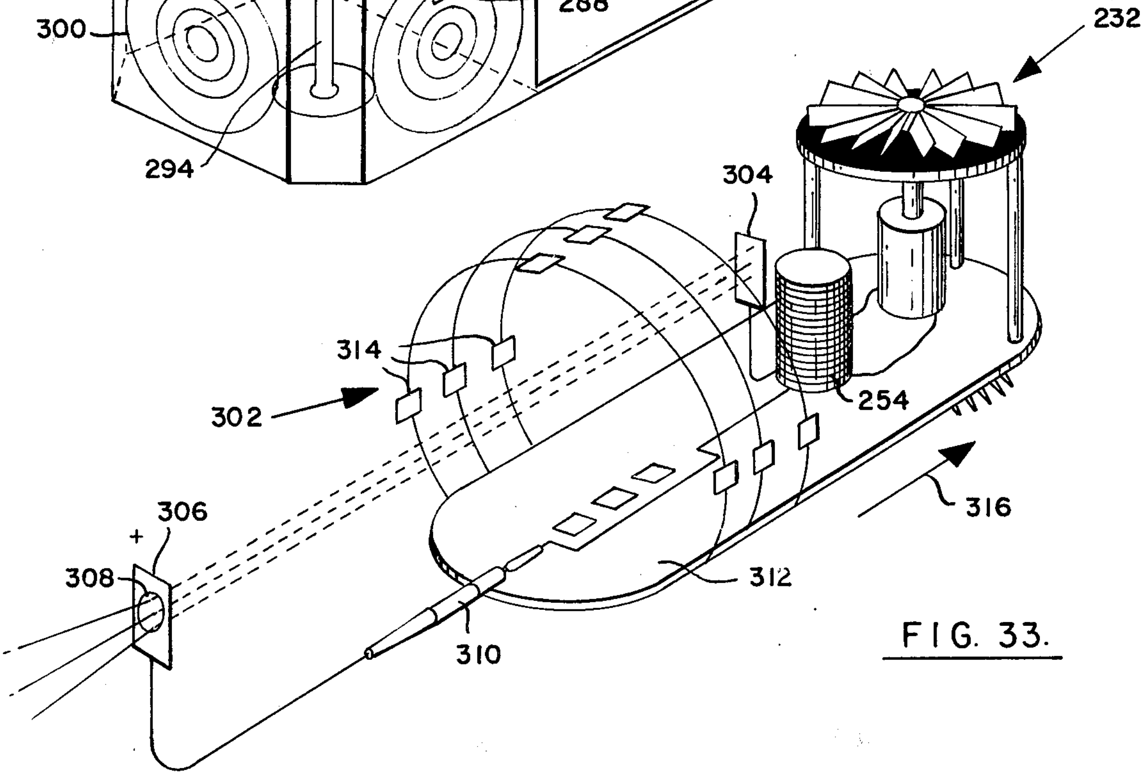
232
304
314
302
254
306
+
308
316
312
310

FIG. 33.

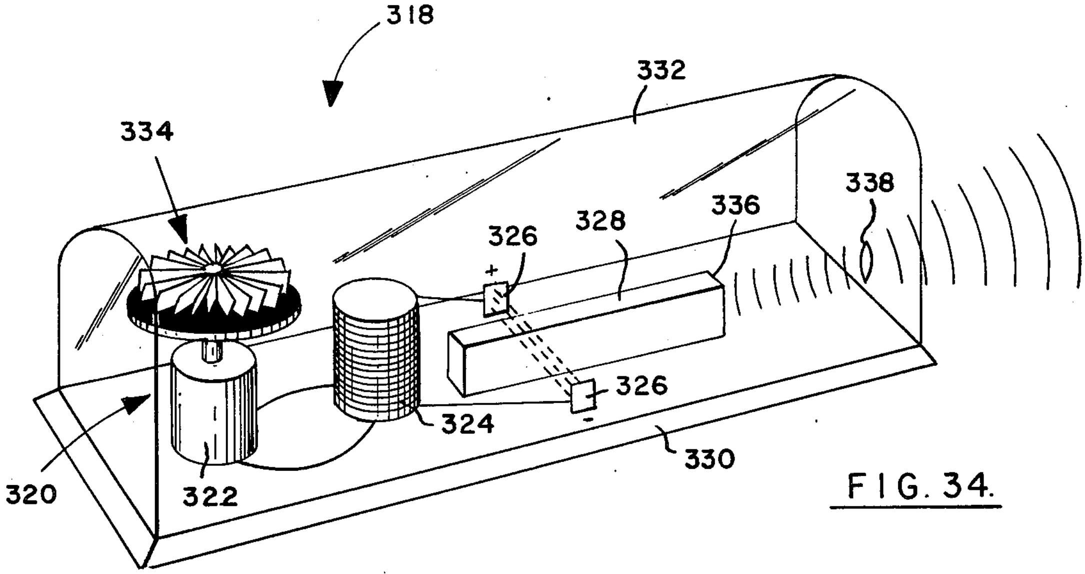
318
332
334
338
326
328
336
+
326
-
324
330
320
322

FIG. 34.

RADIOMETER GENERATOR

BACKGROUND OF THE INVENTION

1. Field of the Invention

The invention relates to a self-contained system for utilizing environmentally available radiant energy, specifically infra-red energy, to provide a new source of electrical energy.

2. Description of the Prior Art

Many proposals have been made in the prior art for the conversion of environmentally available radiant energy directly into electrical energy. Specifically, it has been proposed to use the thermoelectric, piezoelectric, and photovoltaic effects for this purpose. Typical U.S. patents showing such generators are U.S. Pat. Nos. 3,757,146 granted Sept. 4, 1973, 3,198,969 granted Aug. 3, 1965, and 3,296,469 granted Jan. 3, 1967. While the underlying principles of such generators have been known for many years, attempts at the utilization of these principles have been beset with problems that have impeded the development of such generators as practical sources of electrical power.

Other proposals for the conversion of infra-red energy into electrical energy have utilized the principle of the decrease in magnetism of bodies, particularly iron and certain other metals, alloys and their compositions as their temperature is increased. Generators embodying this principle are disclosed in U.S. Pat. Nos. 476,093, granted June 14, 1892 and 2,016,100 granted Oct. 1, 1935. Here, again, notwithstanding the apparent potential of such generators, the art has been unable to develop the technology as a practical electrical energy source. Among the problems involved with such generators is the complex and consequently expensive magnetic field structure that is required.

a phenomenon that has long been known as providing an output of kinetic energy from an input of radiant energy is the radiometer. The radiometer is an instrument that was devised by Sir William Crookes to measure the intensity of incident radiation. It operates on the principle that a body delicately suspended in a vacuum is apparently repelled by radiation. The instrument, in a commonly known form, includes a system of four spaced light-weight vanes of poor heat conductivity which are blackened on one side, the other side being bright, and so suspended in a highly exhausted bulb that they are free to rotate about a common vertical axis with the black surfaces all facing in the same direction. Incident radiation from the sun or an adjacent warm body is absorbed and emitted by the black surfaces of the vanes more than the bright surfaces, and as a result, gas molecules near the black surfaces are energized and press more on the black than the bright surfaces causing the vanes to rotate about their axis. The more intense the incident radiation, the more rapid is the rate of rotation. Actually, an outside source of lower temperature such as ice, dry ice, evaporating ether, or other cold material will draw radiant energy from the molecules causing them to move in the direction of the cold material, also driving the vanes. The container enclosing the vanes is necessary to maintain the high vacuum and to provide a barrier to rebound activated molecules back to the vanes, driving them further.

The radiometer is described in a series of papers that were published in the Philosophical Transactions of the Royal Society of London (1873–1881). Reference is made particularly to Paragraphs 145–148. The publication entitled "'RADIANT MATTER', A Resume of the Principal Lectures and Papers by Prof. William Crookes on the Fourth State of Matter", published by James W. Queen & Co., No. 924 Chestnut Street, Philadelphia Pennsylvania also is concerned with the phenomenon.

In another form of radiation responsive instrument which he called the otheoscope and that is described also in the aforementioned Philosophical Transactions, particularly in Paragraphs 450–454, Crookes provided a larger radiant energy absorbing and emitting surface consisting of a large stationary plate of good heat conductivity having a blackened surface. Thus, in the otheoscope the surface that generates the molecular disturbance is not the one that rotates. The blackened surface of the plate, however, is placed close to the movable vanes. Because of the larger surface area and closer proximity, the blackened stationary plate activates a greater number of molecules and directs them more accurately to the movable vanes resulting in a greater speed or rotation than is achievable by the radiometer with the same intensity of incident radiation. Another advantage of the otheoscope is that an enclosure is not necessary when a vacuum such as outer space is already present. The small space between the stationary blackened surface and movable vanes provide a rebounding space within which molecular rebound can occur and rebound off a container wall is not necessary to provide the moving force of the vanes. Another form of the otheoscope involved a stationary blackened disc above which light-weight blades radiating from a common center were positioned giving the appearance of a turbine.

Although the inventions of the radiometer and the otheoscope were made more than a hundred years ago, the only use proposed heretofore for the instruments and their inherent property of converting radiant energy into kinetic energy has been the measurement of the intensity of radiation falling upon them. That this available kinetic energy might be utilized to provide a useful electrical energy source has gone unnoticed. This notwithstanding the need for new energy sources and the intensive search for such new energy sources that has long been underway. Significant in this connection is the highly developed technology that has been available for the effective utilization of such kinetic energy, specifically the production of an output of electrical energy by the mechanical relative rotative motion of the magnetic field and the electrical conductors of the dynamoelectric machine.

SUMMARY OF THE INVENTION

An object of the invention is to bring together in a self-contained system apparatus embodying the principles of the radiometer, the otheoscope, and the dynamoelectric machine thereby to derive, as a new source of energy, an output of electrical energy from the dynamoelectric machine directly from an input of radiant energy to the system from the surrounding environment.

Another object of the invention is to provide such a self-contained system that is characterized by light-weight, thereby rendering it suitable for airborne and space as well as terrestrial applications.

A further object of the invention is to include in combination with such an electrical generating system a voltage amplifying device such as an induction coil.

A further object of the invention is to provide a series of such electrical generating systems each including a voltage amplifying device the output of which is utilized to provide a driving force for the succeeding stage.

Another object of the invention is to include in combination with such an electrical generating system and amplifying device self-contained photon-emitting devices, for example, solid materials such as gemstones and specially enclosed liquids and gases thereby to provide a self-contained photon-emitting device.

Still another object of the invention is to provide in a self-contained system, combination and cluster arrangements of such electrical generating systems and amplifying devices thereby to increase the electrical power output.

In accomplishing the foregoing objects and other objectives of the present invention, there are provided selfcontained radiometer and otheoscope generating systems for generating electrical energy from impinging or incident external environmental sources of radiant energy such as heat from the sun or any source of infrared radiant energy such as a heat pipe attached to a solar energy heat collector, an atomic pile, an airplane engine, or in general from heat derived in way from space, the air, land or water. The system will function as well if it is giving up radiant energy to a source much colder than it such as ice, dry ice, liquid nitrogen or evaporating fluids.

The system, in its specific aspects, includes an energy source such as a heat pipe, means for concentrating and directing the energy from the heat pipe including a collimator for directing and making parallel the rays of radiant energy to a fresnel lens and thereby to a radiant energy absorbing/emitting surface. There is also provided a parabolic, hyperbolic, or elliptical mirror for the further reflection of rays of radiant energy and then focusing them on the absorbing/emitting surface. The system further includes a dynamoelectric machine and torque converter which cooperates with the absorbing/emitting surface and nearby molecules to drive the dynamoelectric machine. Thus the incident radiant energy on the absorbing/emitting surface is converted to kinetic molecular energy and then to kinetic mechanical energy, specifically a moment to produce the relative rotative motion of the magnetic field and electrical conductors of the dynamoelectric machine. The dynamoelectric machine may comprise a machine of either the direct or alternating current generating type. The rotative moment at the output of the torque converter provides in cooperation with the dynamoelectric machine a capability for high electrical energy output in the form of high voltage or high electrical current. The system of the present invention is distinguished in this important respect from the prior art.

In accordance with one form of the invention, the radiant energy absorbing/emitting surface, torque converter and dynamoelectric machine are suitably mounted in a highly evacuated container. The collimator, fresnel lens and mirror may be located external of the container. The electrical output produced by the system may be employed to energize electrical components that may be located internally or externally of the container, as desired, or if desired, stored in batteries for future use. When utilized in outer space, the container may be omitted since no special provision is then required to provide the evacuated internal environment of the system. In this latter use, structures employing the principles of the otheoscope would be most effective, the close disposition of the stationary blackened surface and movable vanes providing the necessary space within which molecular rebound can occur. That is to say, a container wall is not necessary to provide the moving force of the vanes, as in the radiometer.

In another aspect of the invention advantage is taken of a special property of the highly evacuated internal environment of the system that solid materials such as gemstones, and specially enclosed liquids and gases, in such an environment may be caused to emit photons by virtue of the interaction of high-energy molecules and a spanning high voltage electrical field, thus providing a self-contained photon-emitting device. Such a device emits light rays useful as a beacon or for lighting living areas as a long term lamp that need not be turned off.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a perspective view of an embodiment of the system of the present invention utilizing an otheoscopic generator;

FIG. **2** is a side elevation illustrating the radiation absorber/emitter, the torque converter, and dynamoelectric machine of FIG. **1**;

FIG. **3** is an exploded perspective view of the apparatus of FIG. **2**;

FIG. **4** is a side elevation illustrating the action of heated molecules emitted from the radiation absorber/emitter on a vane of the torque converter after absorption of radiant energy by the absorber/emitter;

FIG. **5** is a perspective view illustrating another form of otheoscopic generator that may be employed in the system of FIG. **1**;

FIGS. **6** through **11** are perspective views illustrating different forms of radiometer generators that may be employed in the system of FIG. **1**;

FIG. **12** is a perspective view illustrating a battery charging system according to the present invention;

FIG. **13** is a perspective view of a self-contained photon-emitting system according to the present invention;

FIG. **14** is a wiring diagram illustrating a typical form of induction coil that may be used in the system of FIG. **13**;

FIG. **15** is a perspective view illustrating another embodiment of the system of the present invention utilizing a dual otheoscopic generator and having particular application in outer space.

FIGS. **16** and **17** show top plan and side elevational views, respectively, of the system of FIG. **15**;

FIGS. **18** through **25** illustrate top plan views of combination and cluster arrangements of the system of FIG. **15**;

FIG. **26** illustrates a perspective view of a cluster arrangement of the basic FIG. **15** system;

FIG. **27** illustrates an embodiment of a dual radiometer generator for outerspace use that is enclosed in a glass or other heat conducting container;

FIG. **28** illustrates an embodiment of the radiometer generator and induction coil as a power amplifier;

FIG. **29** shows a diode rectifier circuit for use with the apparatus embodiment of FIG. **28**;

FIG. **30** illustrates an embodiment linking dual otheoscopic generator power amplifiers into a chain thereby providing increasingly more powerful generators;

FIG. **31** illustrates another embodiment of a power amplification system using dual otheoscopic and also radiometric forms of generators in outer space;

FIG. 32 illustrates an embodiment of the invention employed as a lighthouse in outer space;

FIG. 33 illustrates an embodiment of the invention employed as a propulsive mechanism in outer space; and

FIG. 34 illustrates an embodiment of the invention for providing a source of coherent light.

DESCRIPTION OF THE PREFERRED EMBODIMENT

THE EMBODIMENTS OF FIGS. 1-11

Referring now to the drawings, and in particular to FIG. **1**, there is shown an otheoscopic generator comprising a self-contained system indicated at **10**. This system requires no external power for its operation, other than environmental incident radiation, for converting such incident radiation into useful electrical energy and other desired forms of radiant energy. The system includes a dynamoelectric machine or generator **12** having a drive shaft **14**. The output electrical current from the generator **12** can be alternating or direct, as desired. In FIG. **1**, the generator **12** is shown as a miniature direct current generator having minimal inertial turning resistance of the shaft, minimum armature resistance, and an output voltage of 1.8 volts per 1000 RPM. A permanent magnet not shown provides an internal magnetic field. Where subsidiary current to maintain an internal magnetic field is required, a suitable additional otheoscopic generator of the type herein illustrated in FIG. 1 would be added to provide such current.

The system of FIG. **1** further includes a torque converter indicated at **16** comprising a light-weight circular aluminum fan having a bank or plurality of light-weight vanes **18** each approximately 1/32" thick radiating from a center disc **20** attached perpendicularly to the vertically positioned drive shaft **14** of generator **12**, as for example, with epoxy glue.

A radiant energy absorber/emitter surface indicated generally by the reference numeral **22** may consist of lampblack painted on the upper side or surface of a plate **23** facing the torque converter **16**. The plate **23** preferably is made of a good heat conductor such as silver, gold, copper or aluminum. A desired property of the blackened surface is high absorbancy and emission of infra-red radiant energy. Material other than lampblack may be employed to provide the desired high absorbant black surface **22** for the plate **23**, as is known in the art.

The generator **12**, torque converter **16** and plate **23** with its absorber/emitter surface **22** are positioned within a Bell jar **24**, being suitably mounted therein. The internal environment **26** of the Bell jar **24** consists of molecules representative, for example, of normal sea level air or other gas mixtures by design evacuated to a high vacuum of 15 to 100 microns of mercury pressure. (Note that at 1 micron of mercury pressure there are still 400 billion molecules per cubic centimeter of vacuum space.) To this end, in FIG. **1**, there is provided a vacuum pump **28** that is attached to the Bell jar **24** by way of a suitable platform or base plate **30** on which the Bell jar **24** is mounted, and also by way of a vacuum hose **32** and high vacuum stop cock **34**.

The environment external of the Bell jar **24** is the normal sea level air for the region in which the Bell jar **24** is located under the influence of various climatological and geophysical factors such as temperature, humidity, pressure, gravitational, electrical, and magnetic field gradients. This environment contains many natural sources of radiant energy indicated generally at **38**, including the sun. The energy source **38** chosen for purposes of illustration in FIG. **1** is a 12" heat pipe **40** that is arranged to transfer heat energy from heated water stored in an insulated container indicated at **42**. The environment shown also includes a short collimator **44** that is provided to direct and make parallel the rays of radiant energy from the close source of radiant energy, the heat pipe **40**. The collimator **44** directs the radiant energy to a fresnel lens **46** having plane dimensions of approximately 12" on a side and ⅛" thick. The fresnel lens **46** concentrates the radiant energy as it is directed toward the absorber/emitter surface **22** and focuses the energy thereon. A parabolic mirror **48** provided on the other side of the Bell jar **24** also reflects radiant energy impinging upon it from the source **40** and focuses it directly on the absorber/emitter surface **22**. Hyperbolic or elliptical mirrors may also be employed.

In FIG. **1** a suitable electrical meter **50** connected to the output of the generator **12** provides a measure of the output of the generator **12**. This output is in the form of a voltage or amperage, as desired. Additionally, a penlight lamp **52**, rated at 5 volts and 60 microamperes, is shown as being energized by the output from the generator **12**.

As those skilled in the art will understand, the absorber/emitter surface **22** and torque converter **16** provide in combination with the generator **12** a capability, in an installation in outer space, for substantial electrical energy output in the form of voltage or electrical current depending upon the electrical design parameters of the generator **12**. Thus, while in FIGS. **1-3** the generator **12** is described as a miniature direct current generator, this is because of physical limitations imposed by housing the structure of the absorber/emitter surface **22** and torque converter **16** within the Bell jar **24**. If housed in a larger container and thus freed of its physical constraints, or located in outer space wherein no housing or container would be required to provide the evacuated space, the torque converter **16**, vanes **18** and absorber/emitter surface **22** may be made of such size as to provide a substantially larger turning moment on the center disc **20**. In this case the absorber/emitter surface **22** would be separate but close to vanes **18** and center disc **20**. Accordingly, the drive shaft **14** and generator **12** may then be increased proportionately in size to provide a correspondingly enhanced output of electrical energy.

The otheoscopic torque converter of FIGS. **1** through **4** is characterized in that the vanes **18** may be transparent so as to allow a substantially unimpeded impingement of radiation on the surface **22** of plate **23** and is further characterized in that the surface **22** that generates the molecular activity in response to incident radiation is stationary, and therefore, is not restricted as to material, weight, shape, size and area of surface. Accordingly, the most suitable heat conducting material and the most efficient size, shape, weight and means of support for the plate **23** can be employed to the end of utilizing the incident radiation to best advantage. The vanes **18** of the torque converter **16**, moreover, can be located very close to the surface **22**, preferably within a millimeter or two, and the angular position of the vanes **18** with respect to the surface so chosen as to take the most efficient mechanical advantage of the radiation produced "molecular wind", as it was termed by Crookes, and as illustrated by way of example in FIG. **4**.

In FIGS. 1 through 3 the stationary plate 23 on which surface 22 is provided is shown as supported on vertical ports that rest on the base plate 30, and the vanes 18 of the torque converter 16 are disposed in a fan shaped configuration close to the surface 22, each vane presenting itself to that surface as illustrated in FIG. 4.

Another form of otheoscopic torque converter that may be employed in the system of FIG. 1 is illustrated in FIG. 5 wherein the surface that generates the molecular activity in response to incident radiation, indicated at 54, is provided on a metal sheet or plate 56 that is curved to provide a semi-circular cylindrical portion having as its axis the shaft of a dynamoelectric machine or generator 58 having a vertically positioned drive shaft 60. Generator 58 and shaft 60 may be identical to the generator 12 and shaft 14, respectively, of FIGS. 1 through 3.

The arrangement of FIG. 5 further includes a torque converter indicated at 62 comprising four light-weight, transparent diamond-shaped vanes 64 each approximately 1/32" thick, that are attached to the end of a respectively associated rod 66. The rods 66 radiate from a center disc 68 that is attached perpendicularly to the end of drive shaft 60 in any suitable manner. Each vane 64, as shown, is vertically positioned on the end of its associated rod 66, being disposed at an angle displaced from normal with respect thereto, and in each revolution about the axis of shaft 60, passes close, within a millimeter or two, to the molecular generating surface 54. The angle of each vane 64 with respect to surface 54 is similar to that between vane 18 and surface 22, as illustrated in FIG. 4, and thus molecular activity at the surface 54 produces a moment that causes the vane 64 to move with respect to the surface 54, thus producing rotation of shaft 60. The direction of rotation of the shaft 60 depends upon the direction from normal of the angle of the vane 64 with respect to its supporting radial rod 66, the angle preferably being the same for each of the vanes 64.

If employed in an environment such as outer space where vacuum conditions exist naturally, it is not necessary to enclose the otheoscopic torque converter 62 within an evacuated container such as Bell jar 24 of FIG. 1. When not so employed, the converter 62 would be positioned in an environment such as that provided by Bell jar 24 with the generator 58 and the curved plate 56 suitably supported on the base plate 30.

FIGS. 6 through 11 illustrate alternative radiometric forms of torque converters that may be employed in place of the otheoscopic torque converter in the system of FIG. 1. The radiometer generators of FIGS. 6 through 11 differ significantly from the otheoscopic torque converter in that the absorber/emitter surface is embodied on the moving members or vanes of the generator instead of on the stationary member, and differs further, as will become apparent, in the requirement for an enclosing wall of the evacuated space to be close to the path of the moving vanes. That is to say, where the absorber/emitter surface is on the moving vanes of the torque converter, an enclosure or wall of the evacuated space must be placed close to the vanes for the latter to be driven when exposed to radiant energy. As noted in connection with the otheoscopic torque converter of FIG. 1, when the absorber/emitter is provided on a surface separate from the movable vanes, the vanes need not have any lampblack coating.

Each of the apparatus embodiments of FIGS. 6 through 11 includes a glass container within which the moving vane system and dynamoelectric machine are placed, which container, in turn, would be placed inside a Bell jar such as the Bell jar 24 of FIG. 1 and evacuated, also, together with the bell jar.

As illustrated, the torque converter 70 of FIG. 6 is enclosed in a glass container 72 and includes four lightweight vanes 74 each approximately 1/32" thick radiating from a center disc 76 attached perpendicularly to a vertically positioned drive shaft 60 of a generator 58 that may be identical to the correspondingly numbered components of FIG. 5. One side 78 of each of the vanes 74 is blackened, as with lampblack, the black surfaces all facing in the same direction around the torque converter. The other sides of the vanes are bright. Preferably the material of which the vanes 74 of FIG. 6 are made, and the vanes of the torque converters described in connection with FIGS. 7 through 11 are made, is a poor heat conductor to the end that the high absorbance of infra-red radiant energy on the black side has little or no effect on the temperature of the bright side.

Radiant energy, and in particular, infra-red energy is differentially absorbed and emitted by the blackened surfaces of the torque converter vanes imparting kinetic energy to adjacent gas molecules. The lack of a similar energy exchange between the obverse bright surface and its adjacent gas molecules results in a pressure differential whereby the more energetic gas molecules close to the blackened surface of the vanes give rise to a resultant driving force turning or rotating the center disc 76.

FIG. 7 illustrates a torque converter 80 that is similar to that of FIG. 6, but includes vanes that consist of semicircular metal foils, as indicated at 82. The foils 82 are each blackened on one side indicated at 84, the blackened sides all facing the same direction as in torque converter 70 of FIG. 6 and also in the torque converters illustrated in FIGS. 8 through 10.

The vanes of the torque converter 86 of FIG. 8 consists of hemispheric cups 88 each of which is blackened on the inside indicated at 90, all facing the same direction.

The vanes of the torque converter 92 of FIG. 9 comprise hemispheric cups indicated at 94. Each cup 94 is provided with a blackened screen 96 that covers the outer half of each cup 94.

The FIG. 10 form of torque converter 98 is a "turbine" form and includes elliptically shaped vanes or plates 100 that present a substantially circular outline in the primary direction of the radiation source. The plates 100 are blackened on the outer faces indicated at 102, and are oriented at 45° from the vertical, and also from the rods 104 on the ends of which plates 100 are mounted. The apparatus embodiment of FIG. 10 allows for continuous exposure of the absorber/emitter surfaces 102 of the vanes 100 to incident light and does not require the spacing of the vanes illustrated in connection with the apparatus embodiments of FIGS. 6 through 9.

The torque converter 106 of FIG. 11 includes a single elongated vane 108 of spiral form that is exposed to incident radiant energy emanating from almost any direction. In FIG. 11, generator 58 is provided with an elongated shaft 110 to the remote end of which one end of of vane 108 is attached, the other end of the spiral of vane 108 being connected to the shaft 110 adjacent the generator 58. As illustrated, the glass container, indicated at 112, for torque converter 106, is elongated and shaped to conform to the general outline of the spiral vane **108**.

THE EMBODIMENTS OF FIGS. 12 THROUGH 14

In FIG. **12** there is illustrated another embodiment of the invention in which a glass container **114** that may be similar in size to an ordinary household light bulb is employed, and which embodies the otheoscopic form of torque converter illustrated in FIG. **1**. The glass container **114** is hermetically sealed after placement of the generator system therein and evacuation of the internal environment to 1 to 100 microns of mercury pressure (preferably one micron). The output terminals of generator **12**, in FIG. **12**, are shown connected to an electrical storage device or battery **116** for charging the latter, the electrical charge stored in battery **116** being available for later use.

FIG. **13** illustrates a further embodiment of the invention employing a container **114** of FIG. **12** and the otheoscopic form of torque converter of FIG. **1** wherein a gemstone **118** is positioned between two spaced plate electrodes **120**. The electrodes **120** are connected to the high voltage output of an induction coil indicated at **122** and actuated by the output of the generator **12**.

A typical electrical circuit for the induction coil **122** of FIG. **13** is illustrated in FIG. **14**. As shown, the induction coil **122** includes a primary winding **124**, consisting of only a few turns, wound on an iron core **126**. A secondary winding **128**, wound over the primary winding, consists of a large number of turns. The primary winding **124** is connected to the output terminals **130** of the generator **12** in series with an electrical capacitor **132**. An electrical switch **134** actuated by an armature **136** is connected across the capacitor **132**. The armature **136** is positioned in cooperative relation with the iron core **126**, being normally biased in upward direction to close the switch but is attracted by the iron core **126**, upon magnetization of the latter, to open the switch **134** and thereby break the circuit to primary coil **124**. As a result the armature **136** is returned to its initial position by a spring, not shown, and closes switch **134**, again restoring the circuit to the primary coil. The cycle is rapidly repeated.

In the operation of the induction coil, the flow of electrical current in the primary winding **124** produces a magnetic field in the iron core **126**. Opening of switch **134** results in a collapse of the magnetic field that induces a high voltage in the secondary winding **128**. The capacitor **132** is placed across the switch **134** to reduce the voltage across the switch as it is opened to minimize sparking.

The output terminals **138** and **140** of the secondary winding are each connected electrically to one of the electrodes **120**. Hence, there is established between the electrodes **120** an electrical field in accordance with the A.C. high voltage output of the induction coil **122**. The gemstone **118**, although a solid material, has the property of emitting photons in the environment described and thus provides a self-contained photon-emitting device.

THE EMBODIMENT OF FIGS. 15 THROUGH 26

When positioned in outer space, it is noted that unless the generator **12** is mounted on a stable platform **30** that is able to resist the turning or rotative moment resulting from radiation impinging on the absorber/emitter surface **22**, there is a tendency for the rotative moment to be applied to the generator **12** itself as well as to the drive shaft **14** thereof. Consequently, after a time the absorber/emitter surface **22**, torque converter **16**, drive shaft **14** and generator **12** will all effectively be rotating as a unit. As a result, there will first be a depletion and then, after a time, no generation at all of electrical power. This undesirable situation may be corrected, as indicated, by mounting the generator **12** on a stable platform that effectively resists such turning or rotative moments, and hence, holds the body of generator **12** stationary with respect to the relatively rotating shaft **14**, torque converter **16** and absorber/emitter surface **22**. In outer space, when the blackened absorber/emitter surface remains part of the system of movable vanes **18**, an enclosure or shroud must be provided to return high energy molecules to further drive the vanes **18** in a rotary motion. Such an enclosure is not required with the otheoscopic form of generator in which the absorber/emitter surface is part of the stationary system.

The embodiment of the invention illustrated in the perspective view of FIG. **15** and in top and side views, respectively, of FIGS. **16** and **17**, shows another way in which to arrange a radiation generator system so as to avoid the need for a platform to hold the body of the generator stable with respect to the associated drive shaft torque converter and absorber/emitter. The radiation generator system **142** of FIG. **15** includes a dynamoelectric machine or generator **144** having a pair of drive shafts **146** and **148**, a pair of torque converters **150** and **152**, and a pair of absorber/emitter surfaces **154** and **156**.

The generator **144** can provide an output of alternating or direct current as desired. Desirably, in the apparatus embodiment of FIG. **15**, the generator **144** comprises two separate generators, each having its own rotatable armature, the magnetic field sections thereof being fastened together and arranged within and fastened to the same external, common casing. The drive shaft **146** is attached to the rotatable armature of one generator. The drive shaft **148** is attached to the rotatable armature of the other generator, being axially in alighnment with the drive shaft **146**. The common casing of the generators is supported by brace members indicated at **155** and **157**, between the ends of a curved absorber/emitter surface **158**, at the center of curvature thereof.

The torque converter, indicated at **150**, comprises a plurality of vanes **160** extending from the ends of rods **162** radiating from a center disc **164** that is attached perpendicularly to the end of drive shaft **146** in any suitable manner. The vanes **160**, as shown, are positioned parallel to the shaft **146** and are so arranged as to have an edge pass close, within a millimeter or two, in each revolution about shaft **146**, to the curved absorber/emitter surface **158** that may be similar to the surface **54** of FIG. **5**. The angle of each vane **160** with surface **158** in such that molecular disturbance at the surface **158** produces a turning on the drive shaft **146**, in the counter-clockwise direction, for example. Additionally, one side of each vane **160** is blacked to provide absorber/emitter surface **154**. This provides a supplemental turning movement in the counterclockwise direction on the drive shaft **146**.

Similarly, the torque converter, indicated at **152**, comprises a plurality of vanes **168** extending parallel to drive shaft **148** from the ends of rods **170**. The rods **170** radiate from a center disc **172** that is attached, in any suitable manner, perpendicularly to the end of drive shaft **148**. The vanes **168** are positioned parallel to drive shaft **148** and arranged to have an edge pass close, within a millimeter or two, to the curved absorber/emitter surface **158**. The angle presented by each of the vanes **168** to the surface **158**, however, is opposite to that of the vanes **160** of the upper torque converter **150**, whereby incident radiation tends to produce clockwise rotation of drive shaft **148**. Additionally, if desired, one side of each of the vanes **168** is blackened to supplement the turning moment of drive shaft **146**, the blackened sides being oppositely disposed with respect to blackened sides of vanes **160** of torque converter **150**.

With this construction, it is evident that any tendency of the rotative moment introduced by one of the torque converters **150** and **152** to rotate the common casing of dual generator **144** is offset or counterbalanced by the rotative moment on the generator casing by the other torque converter thereby providing a substantially balanced system.

The arrangement of FIG. **15** is further advantageous in that the generator **144** effectively comprises two separate generators the outputs of which may be combined whereby a correspondingly larger electrical output is obtained. The output of the generator **144** may be utilized, for example, to energize a device **174** suitably mounted on the body of the generator **144**, externally thereof, as indicated. The device **174** may comprise a lamp providing a beacon, for example, a microwave generator transmitting generated power to receiving antenna or earth, or a radio transmitter transmitting intelligence concerning space environmental factors sensed by detecting devices not shown but supported and electrically powered by the otheoscopic generator **142**. Alternatively, or in addition thereto, the electrical output of the generator **144** may be taken via suitable slip rings provided on the appropriate shaft **146** or **148**.

Another way in which the radiation generating system of FIG. **15** may be arranged so as to be balanced, thereby avoiding the depletion, after a time, of generation of electrical power is to employ but a single generator **144** having two drive shafts **146** and **148** that are axially aligned but extend in opposite directions. With this structure the brace members **155** and **157** for supporting the generator **144** would be carried by a band (not shown) encircling the casing of the generator and disposed, for example, in a groove to preclude vertical displacement but to allow rotation of the casing of the generator relatively to the band and the brace members. One drive shaft **146** of the generator **144** would serve, for example, as the drive shaft **146** of upper torque converter **150** and would be attached to the rotatable armature of the generator. The other drive shaft **148** would be attached to the lower torque converter **152** and also to the body or casing of the generator **144**. It is evident that since the turning moments on the two shafts would be in opposite directions that the system would be balanced and would continue to provide an electrical output irrespective of whether one turning moment were greater than the other. The electrical outputs of the two generators could be connected in parallel or series, as desired.

In accordance with the invention, it is contemplated that otheoscopic generating systems of the kind illustrated in FIGS. **15** through **17** may be arranged in groups or clusters to increase the electrical power output. Thus, FIGS. **18** through **25** illustrate top plan views of combinations of whole system units, from two to nine, each of which may be identical to the structure of FIG. **15**, except for the inclusion on each system unit of a device **174**. FIG. **26** shows a perspective view of constructions linking four whole system units into chains. Other numbers of whole system units may be provided, the chains may be of any length for outer space use. If desired, such chains may be deployed in a star configuration thereby providing desired structural strength as well as contributing to a further increase in electrical power output.

As illustrated in FIG. **26**, it is contemplated that the curved absorber/emitter surfaces would extend, suitably supported, the entire length of each chain. Additionally, more than one set of vanes **160** may be associated with each drive shaft **146** and more than one set of vanes **168** may be associated with the drive shaft **148**. With this arrangement, there is made available a significant increase in the energy for driving each of the generators making up the dual generator **144**. It will further be understood that the kinetic energy outputs of the entire four and if desired, five or more whole system chains that could be provided, may be utilized in parallel, by the provision of suitable gearing, to provide an even more powerful total output.

FIG. **27** illustrates an embodiment of the radiometer generator system of the present invention for outer space use. This system is enclosed in a container **176** that is transparent to infra-red radiation, which container may be glass and may be provided with an opening **178** in the wall to facilitate evacuation to the level of outer space. The system includes two torque converters **180** and **182** that are arranged to rotate oppositely disposed drive shafts **184** and **186** of a generator **188**. The body of generator **188** is directly connected to drive shaft **184**, while the armature of the generator is connected to the drive shaft **186**.

The radiometer generator system of FIG. **27** is rotatively suspended in the container **176** by upper and lower suspension members **190** and **192**, respectively. Thus, as shown, upper suspension member **190** extends between the upper end of drive shaft **184** and a disc **194** that is attached to the wall of container **176**. Lower suspension member **192** extends between the lower end of drive shaft **186** and a disc **196** that is attached to the opposite wall of container **176**.

The torque converters **180** and **182** are shown as of the type illustrated in FIG. **6**, and include four lightweight vanes indicated at **198** and **200**, respectively. One side of each of the vanes is blackened and the other is bright, the black sides of vanes **198** all facing in the same direction, the clockwise direction, as shown, and the bright sides of vanes **200** all facing in the counterclockwise direction. As a result, radiant energy differentially absorbed and emitted by the blackened surfaces of the vanes **198** and **200** impart kinetic energy to adjacent molecules to provide a turning moment on the drive shafts **184** and **186** of generator **188** in the counterclockwise and clockwise directions, respectively. A device **201** that may be similar to the device **174** of FIG. **15** may be provided for utilizing the electrical output of generator **188**.

It will be understood that combinations of whole systems units, each identical to the structure of FIG. **27**, may be provided. Such combinations may be similar to those illustrated in FIGS. **18** through **25**. Additionally, similarly to the constructions of FIG. **26**, the structure of FIG. **27** may be embodied in constructions linking two and more whole system units into elongated chains.

FIG. 28 illustrates a power amplifier embodiment of the present invention in which the output voltage of a radiometer generator system 202, which may be similar to that illustrated in FIG. 6 is amplified by an arrangement consisting of an induction coil 204, a pair of spaced electrodes 206 and 207, and a second radiometer generator system 208, the entire system being mounted on a base plate 210 and enclosed in a glass container 212 that has been appropriately evacuated, as previously described.

The output terminals of radiometer generator system 202 are connected to the input terminals of the induction coil 204 and the high voltage output terminals of the latter are connected through a suitable diode rectifier 214 to the terminals of a high voltage capacitor 216 as shown in FIG. 29, the terminals of capacitor 216 being connected to the electrodes 206 and 207. While shown as of the solid state variety, the rectifier 214 may be of the vacuum type. The induction coil 204 may be similar in construction to the induction coil 122 that has been described by reference to FIGS. 13 and 14. The high voltage alternating current output of induction coil 204, however, is converted to direct voltage by rectifier 214 and capacitor 216.

The radiometer generator system 208 may be similar in construction to the generator system of FIG. 6, but preferably includes a torque converter 218 in which vanes 220 are somewhat larger, and moreover, are supported on the ends of rods 222 at a greater radial distance from a center plate 224 that is attached perpendicularly to the end of a drive shaft 226 of a dynamoelectric machine 228 having a greater output than generator of system 202.

As illustrated the electrodes 206 and 207 are so spaced and so positioned with respect to the circular path transversed by the vanes 220 as to intercept a substantial portion of that path without, however, providing any physical obstruction to the movement of the vanes 220. The polarity of the high voltage applied to electrodes 206 and 207 is such as to make electrode 207 positive with respect to electrode 206 whereby a high voltage molecular stream is established in the direction indicated by the arrow in FIG. 28. This molecular stream or flow of electrons between the electrodes impinges upon each of the vanes 220 as they successively traverse an arcuate path between the electrodes, thereby to produce a turning moment on the drive shaft 226 in the counterclockwise direction, as illustrated. This turning moment supplements the turning moment produced on the drive shaft 226 by the effect of radiation impingement upon the blackened surfaces of the vanes 220, one side of each vane 220 being blackened and the other being bright and with the blackened surfaces all facing in the same direction, that required to produce a counterclockwise turning moment. Hence, the high voltage molecular stream enables the turning of the generator 228 to provide an electrical output greater than that obtainable from the generator of radiometer generator system 202. The amplified power output from generator 228 is available from terminals 230 thereof.

FIG. 30 illustrates an embodiment of the otheoscopic form of radiation generator for use in outer space wherein dual othescopic generators are utilized with associated induction coils to provide power amplifiers, and wherein the power amplifiers are linked into a chain or brigade to produce increasingly higher electrical fields to drive increasingly more powerful generators. Specifically, there is provided in FIG. 30 a first dual otheoscopic generator indicated at 232 consisting of respective upper and lower radiant energy absorber/emitter surfaces 234 and 236, respective upper and lower torque converters 238 and 240, respective upper and lower drive shafts 242 and 244, and a common dynamoelectric machine 246. Each of the otheoscopic generators may be similar to that illustrated and described by reference to FIG. 1. Thus, the absorber/emitter surfaces 234 and 236 may each consist of a blackened surface provided on a respectively associated plate 248 and 250 which are disposed in parallel relation to each other, rigidly supported by suitable rods 252. The surfaces 234 and 236 are respectively associated with torque converters 238 and 240, each of the latter comprising a fan shaped bank or plurality of light weight vanes radiating from a center disc which, in turn, is attached perpendicularly to a respectively associated drive shaft 242 and 244. The upper drive shaft 242, similarly to the arrangement of FIG. 15, is attached to the armature of a first generator, while the lower drive shaft is attached to the armature of a second generator, the magnetic field sections of the generators being connected together and attached to the outer case. As described in connection with FIG. 15, the torque converters 238 and 240 are so arranged with respect to their associated absorber/emitter surfaces as to drive their associated shafts 242 and 244 in opposite directions. As a result, the otheoscopic generator 232 is dynamically balanced.

The output terminals of generator 246 are connected to the input terminals of an induction coil 254 that may be similar to that illustrated in FIG. 14. Additionally, as in the apparatus embodiment of FIG. 28, a diode rectifier-capacitor circuit desirably is provided to convert the high voltage alternating current output of the induction coil to a high voltage direct current potential. A circuit similar to that shown in FIG. 29 is suitable for this purpose.

The high voltage direct current potential so obtained from the dual otheoscopic generator is applied, in parallel, to a first or upper air of spaced electrodes 256 and to a second or lower pair of spaced electrodes 258, as seen in FIG. 30. The spaced electrodes 256 are associated with a fan shaped bank of vanes of an upper torque converter of a second dual otheoscopic generator 260, being so polarized as to supplement the turning moment on the drive shaft connected to the torque converter that is provided by the associated absorber/emitter surface. The spaced electrodes 258 are similarly associated with the vanes of the lower torque converter of the otheoscopic generator 260, the electrodes 258 being polarized oppositely to that of the electrodes 256 since the turning moment produced by lower torque converter is opposite to that of the upper torque converter. The otheoscopic generator 260 may be similar to the otheoscopic generator 232, the output thereof being applied to an induction coil similar to the induction coil 254 and similarly rectified to provide a high voltage unidirectional potential output. Each successive otheoscopic generator may be somewhat larger than the preceding one, four such generators being shown in FIG. 30, thereby to provide a brigade creating increasingly higher voltage fields to drive increasingly more powerful dual otheoscopic generators, as indicated at 262 and 264.

FIG. 31 illustrates another embodiment of a power amplification system, according to the invention, that utilizes a combination of a plurality of otheoscopic generators and a plurality of radiometric generators. Thus, there is provided a platform **266** on which are mounted two otheoscopic generators **268** and **270**, each of which may be similar to the otheoscopic generator illustrated in FIG. **15**, and two radiometric generators **272** and **274**, each of which may be similar to the generator illustrated in FIG. **27**. Desirably, each of the radiometric generators **272** and **274** is enclosed in a glass or heat conducting container that is not shown to avoid undue complication of the drawing.

As illustrated, induction coils **276** and **278** are provided, being respectively associated with the otheoscopic generators and each including a suitable rectifier and capacitor as described in connection with FIGS. **28** and **29**. As in the arrangement of FIG. **30**, two pairs of electrodes **280** and **282**, and **284**,**286** are provided in respective association with the outputs of each of the induction coils **276** and **278**, and the circular path of the vanes of the radiometric generators **272** and **274**. The polarities of the electrodes are so arranged that the molecular streams between electrodes supplements the turning moments provided by the absorber/emitter surfaces of the vanes of the respective radiometric generators, thereby providing increased power output therefrom.

FIG. **12** illustrates another modification of the invention comprisiing an otheoscopic form of radiation generator for use in outer space linked to an induction coil and spaced electrodes to produce photon emission from a gem stone. The light from the gem stone may be transmitted in substantially all coordinate directions through fresnel lenses. Thus, there is provided a self-contained arrangement that aptly may be termed a lighthouse in outer space.

The otheoscopic generator and induction coil of FIG. **12** may be similar to the dual otheoscopic generator **232** and the induction coil **254** of FIG. **30**, and accordingly, have been designated by the same reference numbers. The output terminals of the induction coil **254** are connected through a diode rectifier and capacitor (not shown) to a pair of spaced electrodes **288** and **290**, the latter having a positive potential. A gem stone **292** is positioned between the electrodes **288** and **290**, being mounted on a pedestal **294**. The pedestal **294**, induction coil **254** and generator **232** are all supported on a base plate **296**. The gem stone **292** is positioned centrally of a box-like container **298** that is also supported on the base plate **296**. Each wall of container **298** includes a fresnel lens **300** in that side of container **298**.

FIG. **33** illustrates an otheoscopic form of radiation generator linked to an induction coil for producing a molecular stream of high velocity in a given direction thereby to provide means of propulsion in outer space, specifically a form of rocket engine, for a space craft indicated by the reference numeral **302**. Since it is known that the shape and direction of a high-voltage, high velocity stream can be influenced by a magnetic field, advantage is taken of this in the arrangement of FIG. **33** to direct the electrical field by changing the magnetic field strength.

The otheoscopic radiation generator and induction coil of FIG. **33** may be similar to the dual otheoscopic generator **232** and induction coil **254** of FIG. **30**, and in FIG. **33**, have been designated by those numerals. In practice, it is contemplated that several such otheoscopic generator and induction coil arrangements as illustrated in FIG. **30**, would be utilized. The output terminals of induction coil **254** are connected through a suitable diode rectifier and capacitor, not shown, to a pair of spaced electrodes, indicated at **304** and **306**, the latter being the positive electrode. Electrode **306** has the form of a plate with a centrally located opening indicated at **308** and is positioned on the end of a rudder **310**. Rudder **310** is carried at the rear of a platform **312**, being mounted on a suitable universal joint. Otheoscopic generator **232** and induction coil **254** are shown mounted on the platform **312**, forwardly thereof. Also mounted on platform **312**, by means not shown, are a plurality of spaced magnetic coils, indicated at **314**, that are positioned about and along a portion of the length of the path of the high velocity electrical stream between the negative electrode and **304** and the positive electrode **306**. As those skilled in the art understand, reaction between the magnetic fields established by the magnetic coils **314** and the magnetic field established by the high velocity electrical stream about itself is effective to deflect the stream laterally and vertically in accordance with the relative directions and intensities of the magnetic fields produced by the spaced coils **314**. It will be understood that electrical current for energizing the coils **314** may be obtained from another or even several otheoscopic generators of a type similar to the generator **232**.

In FIG. **33** it is contemplated that a portion of the high velocity molecular stream directed toward the positive electrode or plte **306** escapes through the opening **308** providing a force on the negative electrode **304** that is unopposed, and accordingly, that is useful to propel the space craft **302** in the opposite direction, that indicated by the arrow designated by the numeral **316**. Since the shape and direction of the high-voltage, high velocity stream can be influenced by a magnetic field, advantage is taken of this to direct the stream by changing the relative magnetic field strengths, longitudinally of the stream, in each of the four quadrants about the stream. Thus, for example, deflection of the stream effected to the left rear or port side, as seen in FIG. **33**, will result in the craft turning its bow to starboard. It is contemplated that the strength and spatial intensity of the magnetic fields established by coils **314** would be activated by navigational and computer equipment of conventional or known type, not shown.

Without a magnetic field as produced by magnetic coils **314** present, it will be understood that control of the direction of the craft could be obtained by mechanical movement of the position of the positive electrode or plate **306**, as by mechanical deflection of the rudder **310** about its universal mounting, thereby enabling manual adjustment of the direction in which space craft **302** is propelled.

FIG. **34** illustrates an embodiment of the invention that provides a self contained source of coherent light. Thus, there is provided in FIG. **34** a self-contained laser system indicated at **318** comprising an otheoscopic generator **320**, a dynamoelectric machine or electric generator **322**, an induction coil **324**, spaced electrodes **326**, a gem stone **328**, a base **330** and a glass enclosure **332**.

The otheoscopic generator **320** may be of the type illustrated in FIG. **1** and includes a fan shaped torque converter **334** that is connected to the drive shaft of generator **322**. The induction coil **324**, which may be of the type illustrated in FIG. **14**, has its input connected to the output terminals of generator **322**. The output terminals of induction coil **324** has its output terminals connected through a suitable rectifier and capacitor, not shown, for converting the high-voltage alternating current output to a high voltage unidirectional voltage output, as described in connection with FIG. **29**. The negative output terminal of induction coil **324** is applied to one of the spaced electrodes **326**, and the positive terminal is applied to the other one of the electrodes **326**. Gem stone **328**, desirably a ruby rod, having a half silvered mirrored surface at one end, is placed between the spaced electrodes **326** for bombardment by the electrical field established by the electrodes in a direction perpendicular to the long axis of the ruby rod. The various otheoscopic generator laser components are suitably mounted on the base **330** and located within glass container **332** which is suitably evacuated.

In the operation of the laser system, the high-voltage electrical field established between the electrodes **326** stimulates the ruby rod **328** to the emission of light. Stimulated emission of light occurs, as is known in the art, as radiant energy travelling back and forth between the mirrored surfaces at the ends of the long axis of the ruby rod **328** builds up to a level of magnitude enabling it to escape through the escape half-silvered morror surface indicated at **336** as a spaced frequency distribution of coherent light. In FIG. **34** a lens **338** is provided in the wall of container **332** for diverging or converging the radiation emitted by the ruby rod **328**, as needs vary. Other kinds of lenses may be employed, if desired.

There has thus been provided, according to the present invention self-contained radiometric and otheoscopic generator system and several variations, combinations and cluster arrangements thereof for generating electrical energy from incident radiation emanating from external environmental sources including the sun or any source of infra-red energy whether derived from space, the air, land or water.

Where the development of an increased electrical output, voltage or current, from the available environmental radiation is desired or required, it has been shown that the electrical outputs of a plurality of otheoscopic and/or radiometric generators may be connected in series or parallel, the series connection providing an increased voltage output and the parallel connection an increased current output.

Additionally, for the development of increased kinetic energy from the incident radiation and correspondingly increased electrical energy from each otheoscopic or radiometric generator, it has been shown that several torque converters each having its own bank or plurality of vanes may be associated with the drive shaft of the generator, arranged either in tandem or in parallel, appropriate gearing being provided for the parallel arrangement, and that further increase in the electrical energy output may be provided by arranging the generator combinations in clusters, particularly, for use in outer space.

Embodiments of the invention involving dual otheoscopic generators linking power amplifiers in a chain or brigade thereby to provide more powerful generators have also been provided. The use of the invention for space ship propulsion also is illustrated.

Embodiments of the invention have also been provided for self-contained photon-emitting devices, providing for example, beacons or lighthouses in space and for providing coherent source of light.

I claim:

1. An electric generator comprising a dynamoelectric machine having an input drive shaft and output electrical terminals, a torque converter having an output member connected to rotate said drive shaft and energy receiving means for driving said drive shaft, and means including a radiation absorber/emitter surface operatively associated with said torque converter for absorbing radiation and transmitting it to nearby gas molecules and imparting energy to them to cause them to travel and impart energy to said energy receiving means, said absorber/emitter surface and nearby gas molecules imparting kinetic energy to said torque converter responsively to incident radiation.

2. A generator as specified in claim **1** wherein said torque converter energy receiving means includes a plurality of movable vanes that radiate from said output member about an axis coincident with said drive shaft.

3. A generator as specified in claim **2** wherein said absorber/emitter surface means includes a black surface formed on a plate that has good heat conductivity and is positioned adjacent the path of movement of said movable vanes.

4. A generator as specified in claim **3** wherein said plate is flat and said movable vanes comprise the vanes of a circular fan.

5. A generator as specified in claim **3** wherein said output member includes a center disc that is attached to the end of said drive shaft, and wherein said plate is curved to provide a semicircular cylindrical portion, and wherein each of said movable vanes is attached to the end of an associated rod that is attached to and radiates from said center disc, the angle subtended by each vane with respect to said curved plate being such that molecular disturbance emanating from the absorber/emitter surface of said curved plate produces a turning moment on said vane.

6. A generator as specified in claim **2** including an evacuated container, said torque converter and said absorber/emitter surface being positioned within said container, and wherein said absorber/emitter surface comprises a blackened radiant energy absorbant surface that is provided on one face of each of said vanes, the other face being bright, with the blackened surfaces all facing in the same direction around said output member, and with each of said vanes so positioned as to have at least a component of the blackened surface thereof parallel to the shaft of said dynamoelectric machine.

7. A generator as specified in claim **2** wherein a surface of each said vanes is so positioned as to be substantially parallel to the drive shaft of said dynamoelectric machine.

8. A generator as specified in claim **2** including an evacuated container and a base plate for said container, and wherein said dynamoelectric machine, torque converter and absorber/emitter surface are supported on said base plate within said container.

9. A generator as specified in claim **8** including a fresnel lens and a parabolic lens, said lenses being positioned externally of said container, one on one side thereof and one on the other, and each directing environmental radiation toward said absorber/emitter surface.

10. A generator as specified in claim **9** including a heat pipe comprising a radiant energy source and a collimator positioned between said heat pipe and said fresnel lens for directing rays of radiant energy to said fresnel lens and for making them parallel whereby said rays of radiant energy are concentrated by said fresnel lens on said absorber/emitter surface.

11. A generator as specified in claim **10** including at least one actuatable electrical device supported on said base plate within said container, said electrical device being connected to the output electrical terminals of said dynamoelectric machine.

12. A generator as specified in claim **11** wherein said actuatable electrical device comprises an electric lamp, and further including a meter within said container connected to the output electrical terminals of said dynamoelectric machine.

13. A generator as specified in claim **8** wherein said dynamoelectric machine produces direct electrical current and wherein said evacuated container includes a pair of terminals positioned in a wall thereof, said pair of terminals being connected to the output terminals of said dynamoelectric machine.

14. A generator as specified in claim **8** including an induction coil having a pair of input terminals and a pair of output terminals, a pair of electrodes, and a gem stone positioned within said container between said electrodes, the input terminals of said induction coil being connected to the output terminals of said dynamoelectric machine, the output terminals of said induction coil being connected to said electrodes, and said gem stone being positioned between said electrodes so that it is activated by the emissions from the electrodes.

15. A generator as specified in claim **1** wherein said dynamoelectric machine includes a second input drive shaft disposed in axial alignment with said first mentioned input drive shaft, and further including a second torque converter having an output member connected to rotate said second drive shaft, said second torque converter being so arranged with respect to said radiation absorber/emitter surface that the latter and nearby gas molecules impart kinetic energy to said second torque converter, responsively to incident radiation, in a sense opposite to the kinetic energy imparted to said first mentioned torque converter.

16. A generator as specified in claim **15** wherein said first and second input drive shafts are positioned at the opposite ends of said dynamoelectric machine, and said absorber/emitter surface is provided on a plate that is curved to provide a semi-circular cylindrical portion, and wherein each of said torque converters includes a center disc attached to the end of each of said input drive shafts, a plurality of rods radiating from each of said centers discs, and a vane carried on the end of each of said rods for movement in a path closely adjacent said semi-circular cylindrical portion of said plate.

17. A generator as specified in claim **15** wherein each of said torque converters includes a plurality of movable vanes, wherein the vanes associated with each of said torque converters comprise the vanes of an individual circular fan, and wherein said radiation/emitter surface is formed on one surface of each of a pair of oppositely disposed parallel plates, said one surface of each of said plates being positioned adjacent the path of movement of the vanes of an associated one of said circular fans.

18. A generator as specified in claim **15** wherein a second radiation absorber/emitter surface is associated with said second torque converter, said second absorber/emitter surface and nearby gas molecules imparting kinetic energy to said second torque converter responsively to incident radiation, said first and second torque converters each including a plurality of movable vanes that radiate respectively from the associated output member of said first and second torque converters about an axis coincident with said aligned first and second input drive shafts of said dynamoelectric machine, all of said vanes being so positioned as to be substantially parallel to said aligned drive shafts, and wherein said first and second radiation absorber/emitter surfaces respectively associated with said first and second torque converters each comprise a blackened radiant energy absorbant and emitting surface provided on one face of each the vanes of the associated torque converter, the other face being bright, with the blackened surfaces of the vanes of said first torque converter all facing in one direction around the shaft of said dynamoelectric machine and the blackened surfaces of the vanes of said second torque converter facing in the opposite direction.

19. A generator as specified in claim **1** including at least one additional torque converter having an output member connected to said drive shaft of said dynamoelectric machine, and a radiation absorber/emitter surface associated with each such additional torque converter, each such additional radiation absorber/emitter surface and nearby gas molecules imparting kinetic energy to the associated additional torque converter in the same sense as the kinetic energy imparted to said first mentioned torque converter by said first mentioned absorber/emitter surface and nearby gas molecules responsively to incident radiant energy.

20. A generator as specified in claim **19** wherein said torque converters are connected in tandem to said drive shaft of said dynamoelectric machine.

21. A generator as specified in claim **19** wherein said torque converters are connected in parallel to said drive shaft of said dynamoelectric machine.

22. A generator as specified in claim **2** wherein said dynamoelectric machine provides direct electrical current at its output electric terminals and including at least another torque converter having movable vanes and associated absorber/emitter surface that are similar to the first mentioned torque converter and absorber/emitter surface, an induction coil having a pair of input electrical terminals and a pair of output electrical terminals, electrical rectifying means, and a pair of spaced electrodes, said input terminals of said induction coil being connected to said output electrical terminals of said dynamoelectric machine, said output terminals of said induction coil being connected through said rectifying means to said spaced electrodes whereby an alternating electrical current output of said induction coil is converted to a direct current potential output that is applied to said spaced electrodes, said spaced electrodes being positioned to intercept a portion of the path of the movable vanes of said another torque converter whereby a molecular stream established between said electrodes by said direct current potential output impinge upon said last mentioned movable vanes to supplement the kinetic energy output of said second torque converter.

23. An electric generator comprising

a dynamoelectric machine having an input drive shaft and output electrical terminals,

a torque converter connected to said shaft,

said torque converter including a plurality of lightweight vanes radially connected to said shaft,

radiation absorber/emitter means for driving said vanes and rotating said shaft,

said absorber/emitter means including surface means for absorbing radiant energy and heating up and emitting heat to agitate nearby gas molecules and thereby create a molecular wind that rotates said vanes to rotate said input drive shaft of said dynamoelectric machine and generate electricity.

24. A power amplifier comprising a plurality of dual otheoscopic generators and individually associated dynamoelectric machines linked into a chain to produce increasingly more powerful generators, said otheoscopic generators each including a plurality of movable vanes, said amplifier further including a plurality of pairs of spaced electrodes, a plurality of rectifying means, and a plurality of induction coils each having an input associated with the output and an individual one of said dynamoelectric machines and an output associated with one of said rectifying means, the output of each of said rectifying means being applied between the spaced electrodes of an associated one of said pairs of spaced electrodes, successive ones of said pairs of spaced electrodes being positioned to intercept a portion of the path of the movable vanes of the successive otheoscopic generator in the chain whereby a molecular stream established between pairs of said electrodes impinge upon said movable vanes to supplement the kinetic energy output of said successive generator.

25. A power amplifier as specified in claim **24** wherein each of said dynamoelectric machines includes a pair of oppositely disposed rotary drive shafts, and an otheoscopic generator of each of said dual otheoscopic generators being associated with an individual one of each of said shafts, the direction of rotation of each of said shafts being in opposite directions whereby each of said dynamoelectric machines are dynamically balanced.

26. A power amplifier as specified in claim **25** wherein the electrodes of each of the pairs of spaced electrodes associated with the otheoscopic generators of each dual otheoscopic generator are oppositely polarized whereby the individual molecular streams established between the pairs of electrodes and impinging upon the movable vanes of the associated otheoscopic generator of each dual otheoscopic generator are oppositely directed and have a minimal effect on the dynamic balance of the associated dynamoelectric machine.

27. An electric generator as specified in claim **1** further including means providing an evacuated space, said dynamoelectric machine, torque converter and absorber/emitter surface being wholly contained within said evacuated space.

* * * * *

United States Patent [19]

Minovitch

[11] **Patent Number: 5,052,638**

[45] **Date of Patent: Oct. 1, 1991**

[54] **ELECTROMAGNETIC RAMJET**

[76] Inventor: **Michael A. Minovitch**, 2832 St. George St. #6, Los Angeles, Calif. 90027

[21] Appl. No.: **330,632**

[22] Filed: **Mar. 30, 1989**

[51] **Int. Cl.**^{5} **B64C 39/00**; B64D 27/02

[52] **U.S. Cl.** .... **244/53 R**; 60/202; 244/62; 244/172

[58] **Field of Search** .... 60/202; 244/53 R, 62, 244/166, 172

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,114,517	12/1963	Brown	244/62
3,174,705	3/1965	Schiff et al.	244/62
3,322,374	5/1967	King, Jr.	244/62
3,436,918	4/1969	Brogan et al.	60/202
3,542,316	11/1970	Hart	244/53 R
3,818,700	6/1974	Kantrowitz et al.	244/53 R
3,825,211	7/1974	Minovitch	244/62
3,891,160	6/1975	Minovitch	244/62
4,006,597	2/1977	Dick	60/721

FOREIGN PATENT DOCUMENTS

2212773	8/1989	United Kingdom	244/53 R

OTHER PUBLICATIONS

Astronautics & Aeronautics, Kantrowitz, Propulsion to Orbit by Ground-Based Lasers, May 1972, pp. 74–76.

Primary Examiner—Charles T. Jordan
Assistant Examiner—Richard W. Wendtland
Attorney, Agent, or Firm—Christie, Parker & Hale

[57] **ABSTRACT**

An electromagnetic ramjet propulsion system is provided for accelerating high mass payloads through the upper atomsphere at orbital velocities. The propulsion system comprises a plurality of coplaner self-supporting superconducting dipole coils several hundred meters in diameter that is initially accelerated to high altitude and supersonic speed by magnetic repulsive forces generated by a plurality of superconducting field coils several kilometers in diameter embeded beneath the earth's surface. The ramjet is accelerated to orbital velocities by a multigigawatt microwave beam that is transmitted from the earth's surface. A reflecting grid of conducting wires is mounted inside the inner dipole which shock ionizes the low dentisy atmospheric gas passing through it. The frequency of the microwave beam is adjusted to produce electron cyclotron resonance with the free electrons passing through the magnetic field of the dipoles thereby accelerating them away from the dipoles by magnetic repulsive forces, taking the positively charged ions with them. The system becomes a giant hypervelocity ionospheric ramjet propelled by magnetic repulsive forces without any velocity limitations. By utilizing a plurality of microwave transmitters located along the ground track of the ramjet, it is possible to accelerate payloads of several thousand tons to orbit.

133 Claims, 14 Drawing Sheets

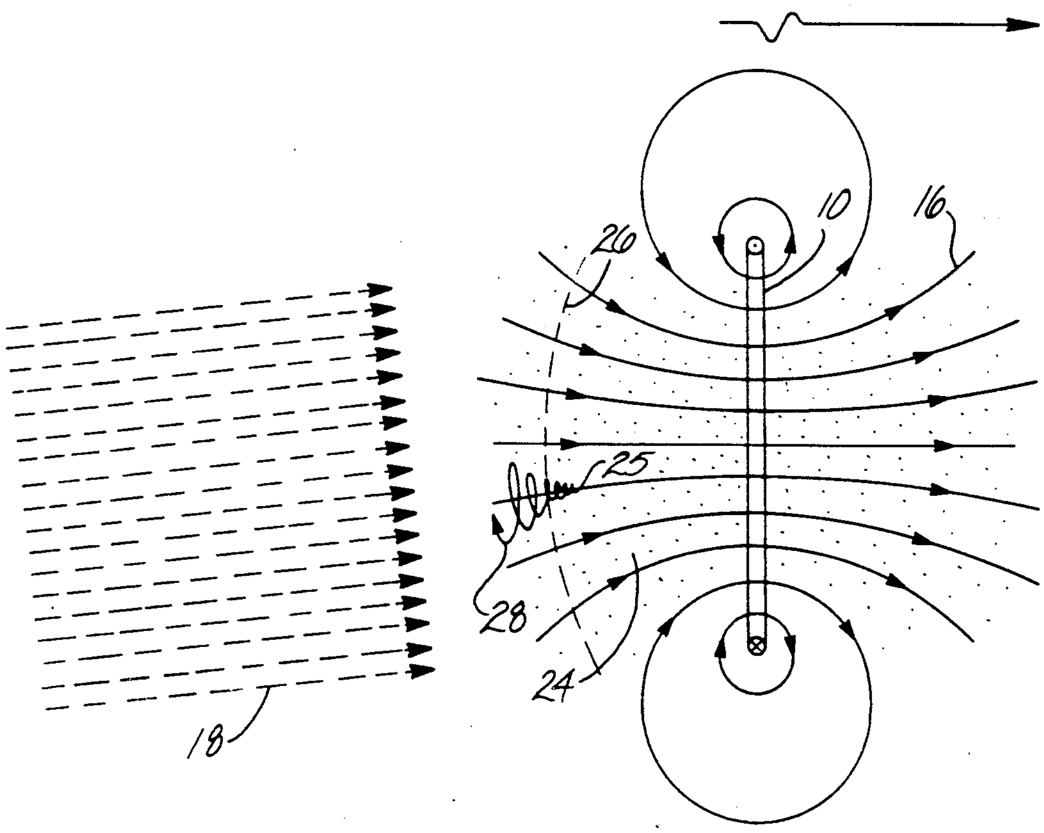

Fig. 1

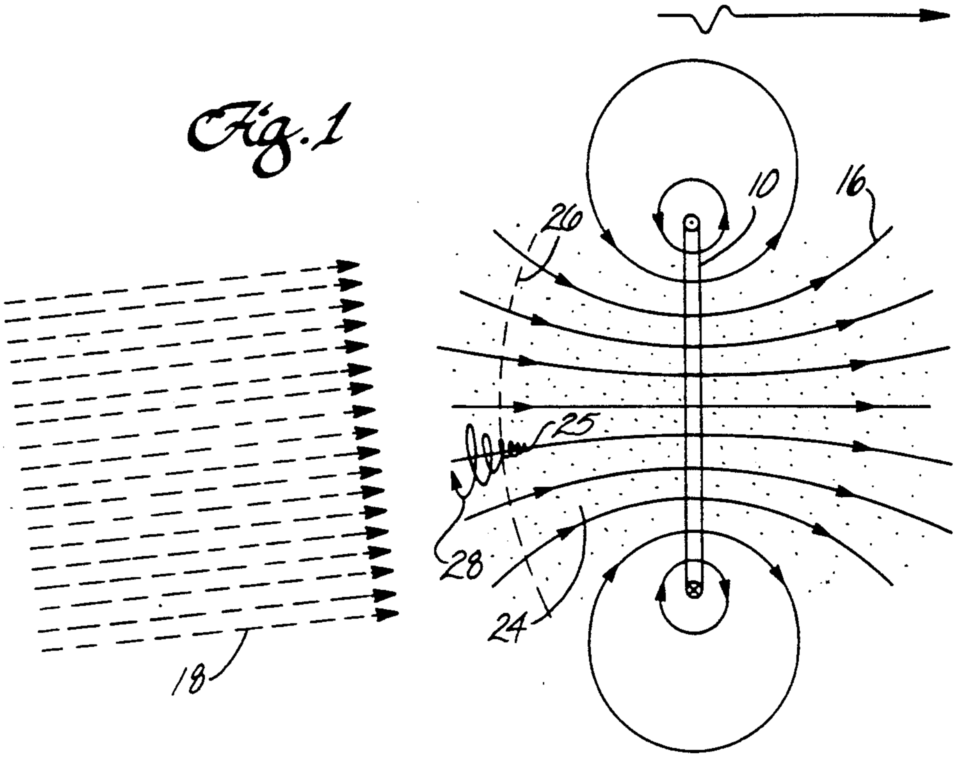
10
16
26
25
28
24
18

Fig. 2

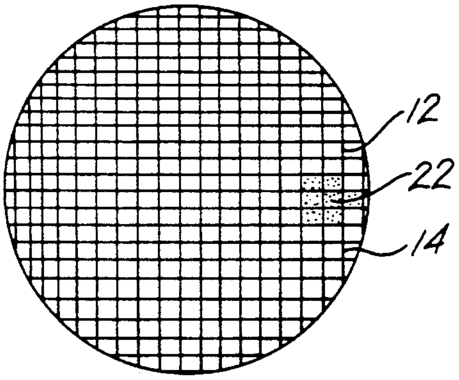

Fig. 3

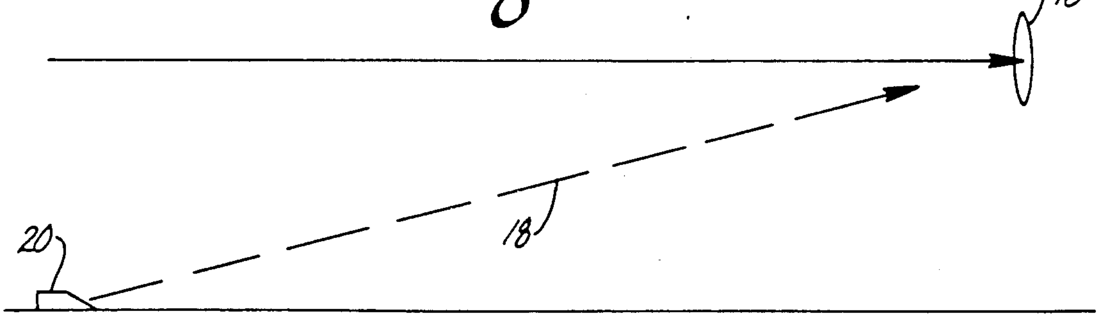

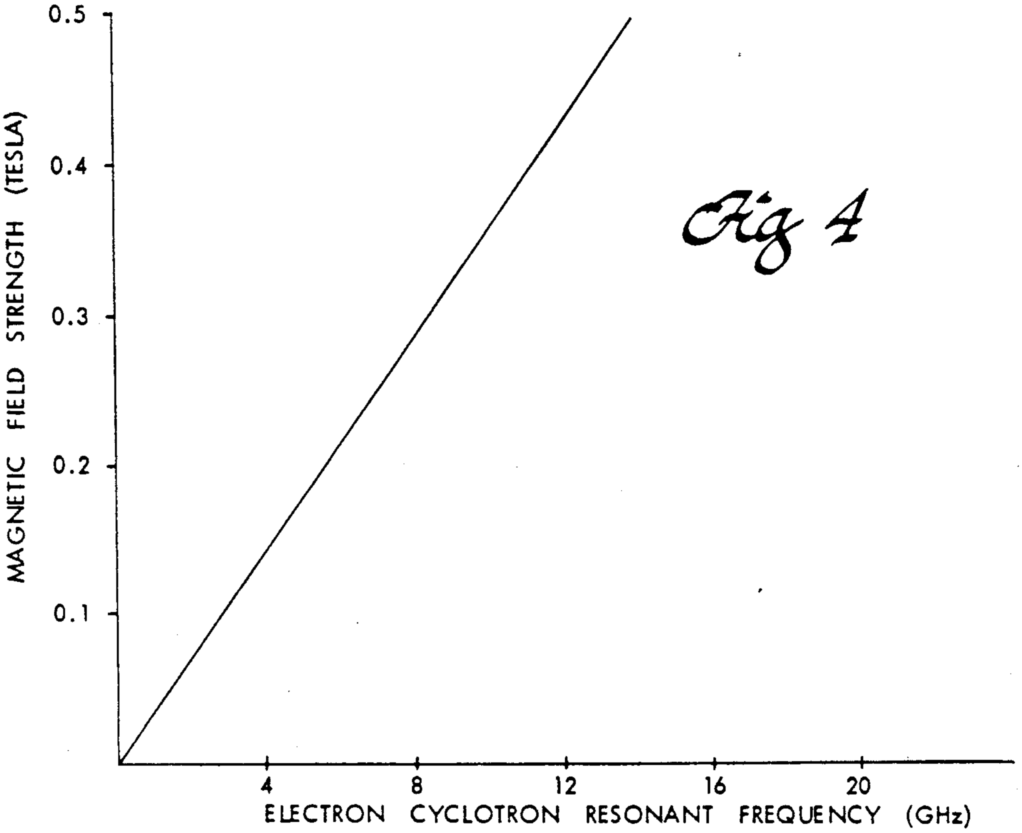
MAGNETIC FIELD STRENGTH (TESLA)
0.5
0.4
0.3
0.2
0.1
4
8
12
16
20
ELECTRON CYCLOTRON RESONANT FREQUENCY (GHz)

Fig 4

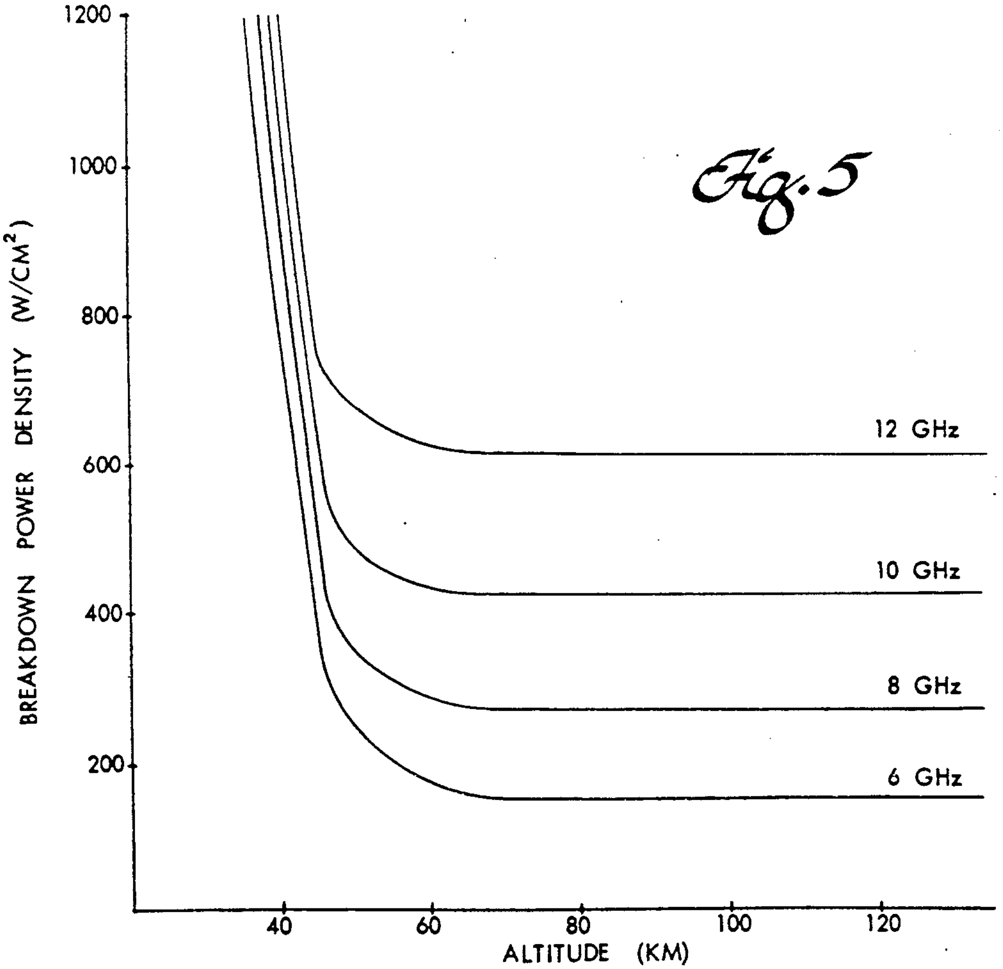
BREAKDOWN POWER DENSITY (W/CM^2)
1200
1000
800
600
400
200
12 GHz
10 GHz
8 GHz
6 GHz
40
60
80
100
120
ALTITUDE (KM)

Fig. 5

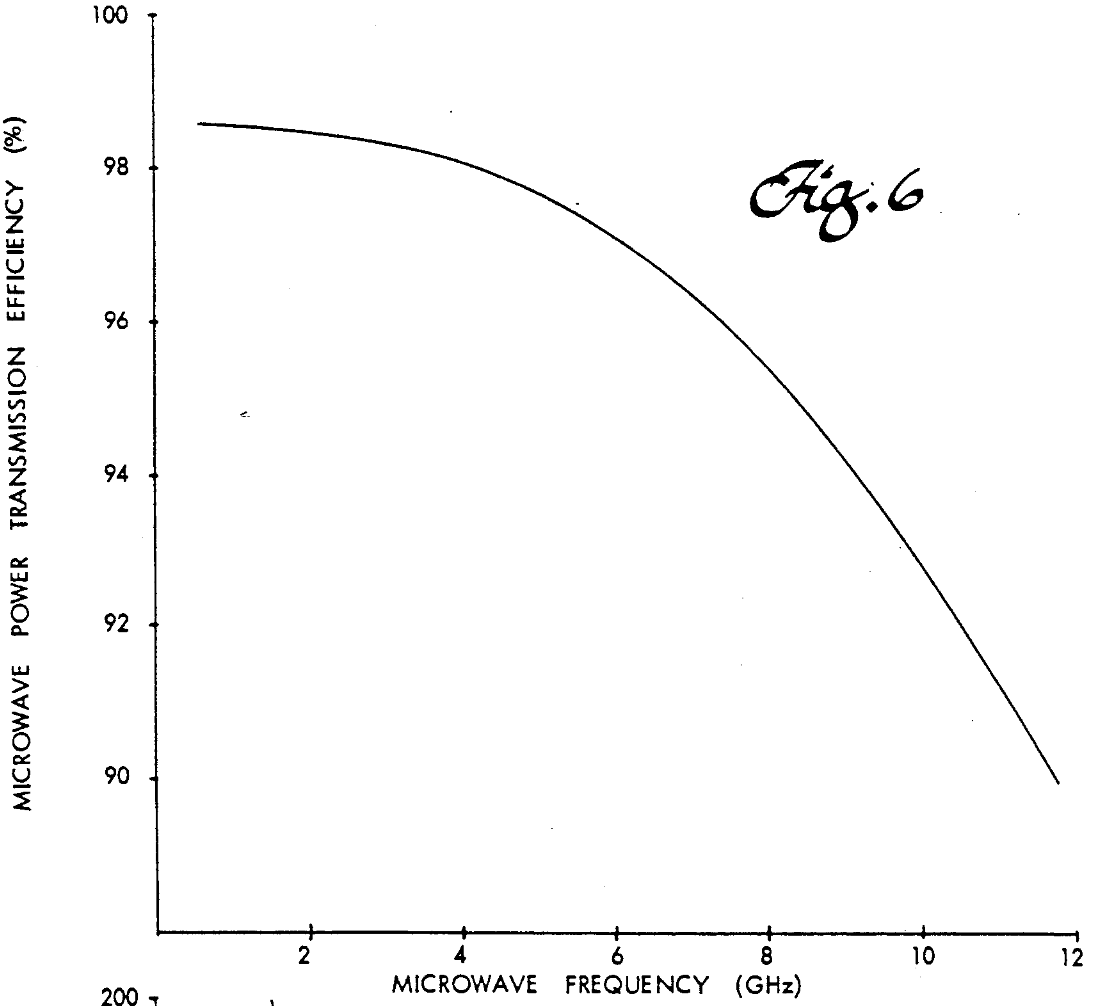
100
98
96
94
92
90
MICROWAVE POWER TRANSMISSION EFFICIENCY (%)
2
4
6
8
10
12
MICROWAVE FREQUENCY (GHz)

Fig. 6

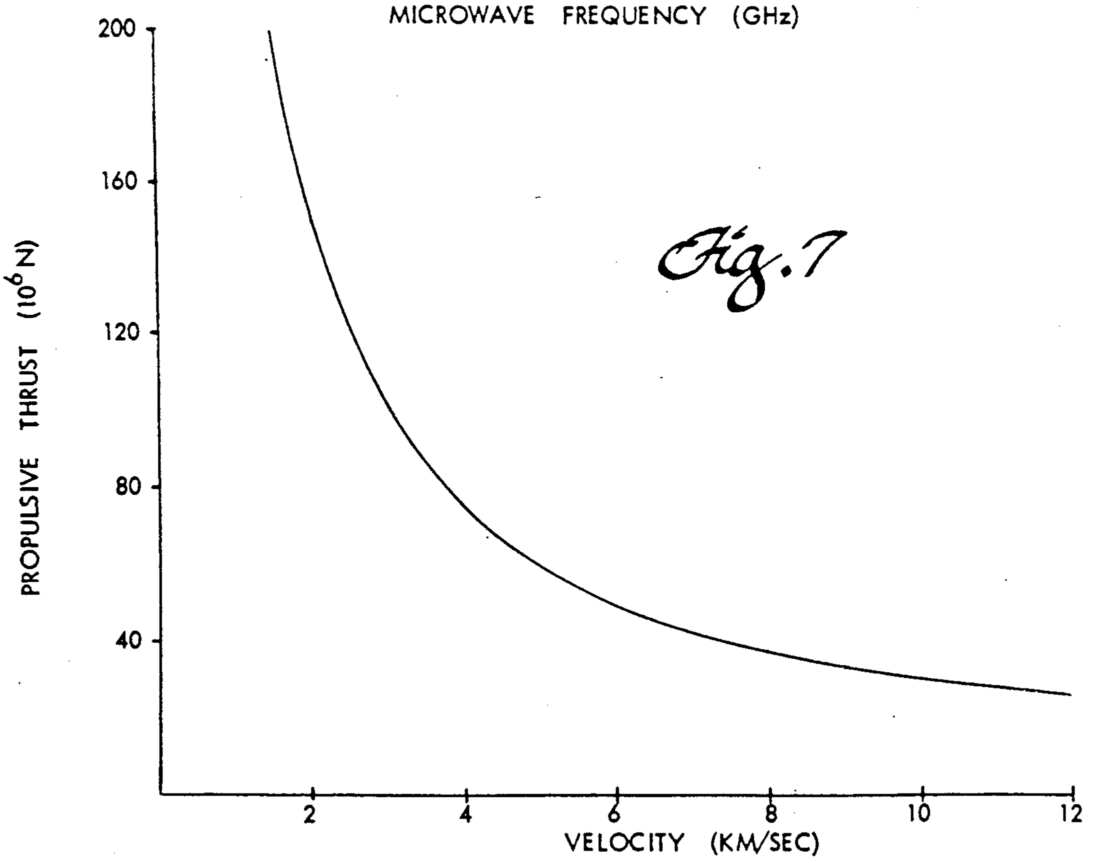
200
160
120
80
40
PROPULSIVE THRUST (10^6N)
2
4
6
8
10
12
VELOCITY (KM/SEC)

Fig. 7

Fig. 8

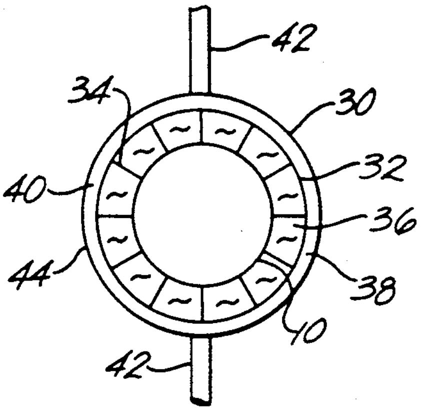
42
34
30
32
40
36
44
38
40
42

Fig. 9

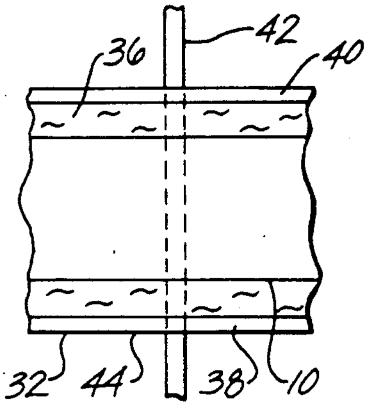
42
36
40
32
44
38
10

Fig. 10

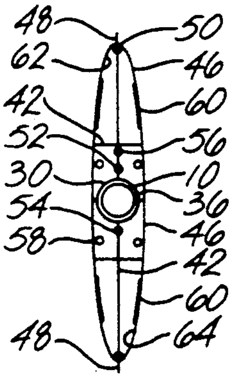
48
50
62
46
42
60
52
56
30
10
54
36
58
46
42
60
48
64

Fig. 11

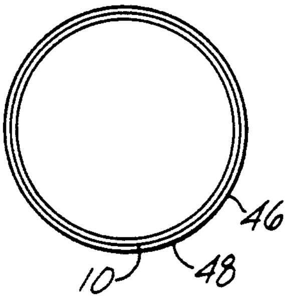

Fig. 12

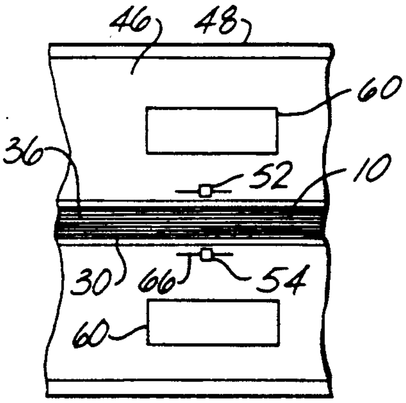
46
48
60
36
52
10
30
66
54
60

Fig. 16

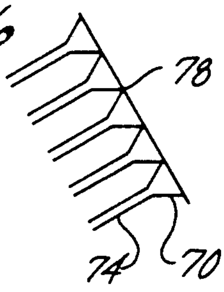

Fig. 13

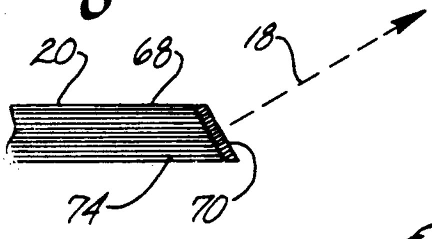
20
68
18
74
70

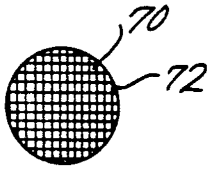

Fig. 14

Fig. 15

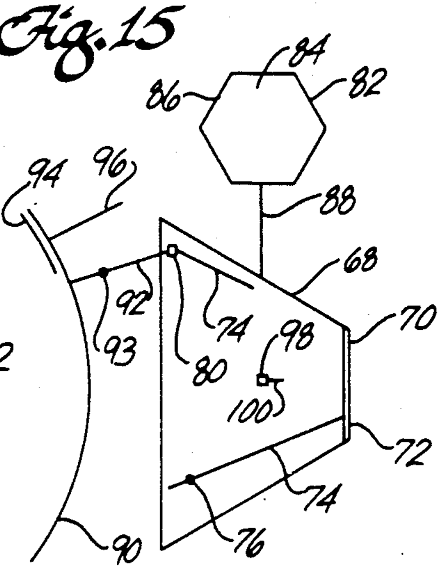
84
86
82
94
96
88
68
92
93
74
80
98
70
100
72
74
76
90

Fig. 17

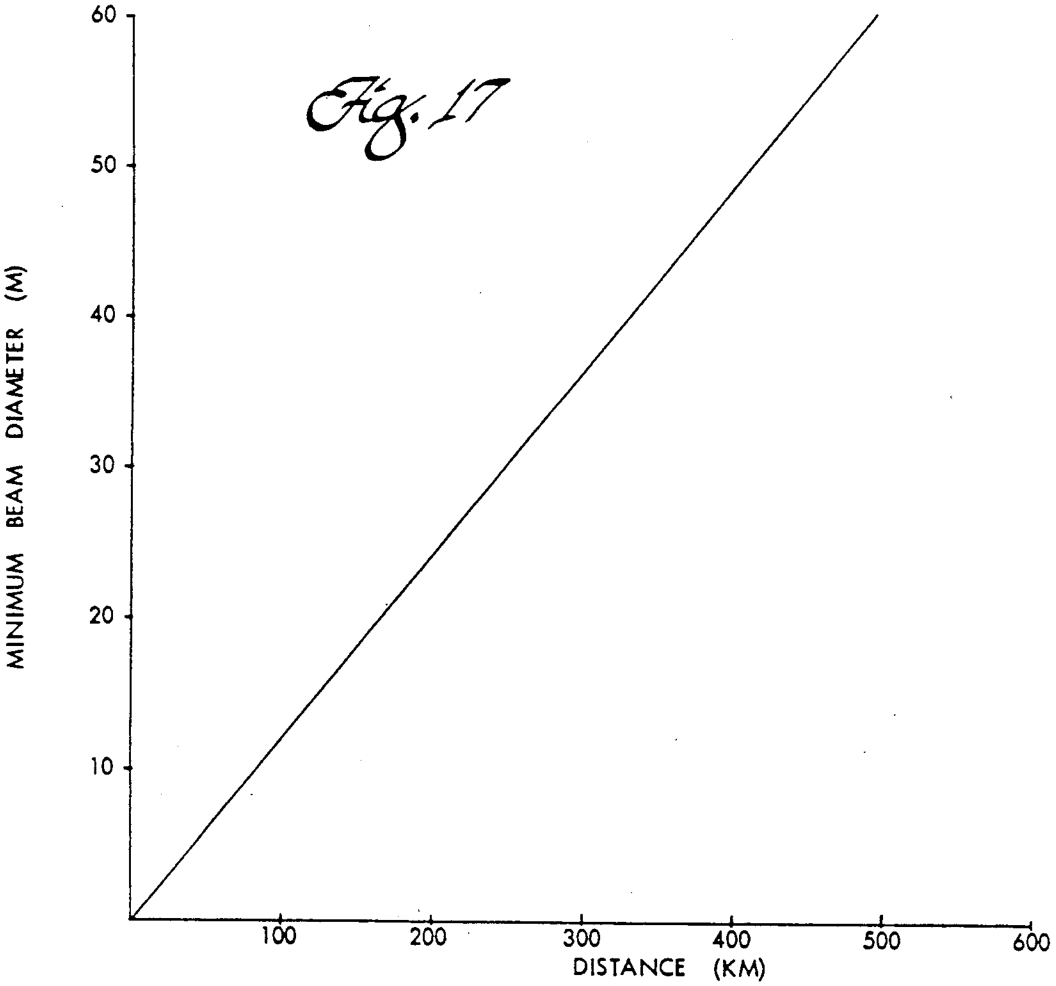
60
50
40
30
20
10
MINIMUM BEAM DIAMETER (M)
100
200
300
400
500
600
DISTANCE (KM)

Fig. 18

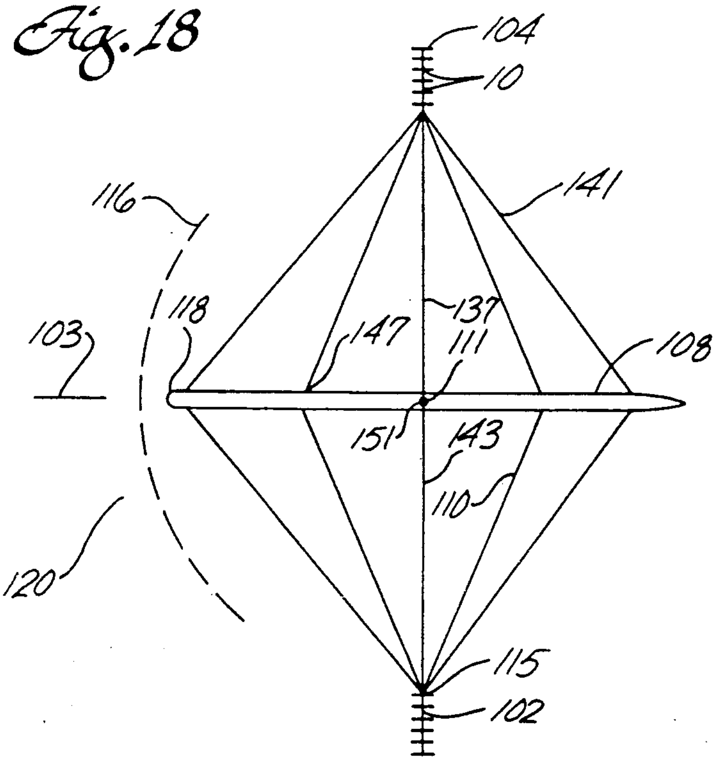
104
10
141
116
137
118
147
111
108
103
151
143
110
120
115
102

Fig. 19

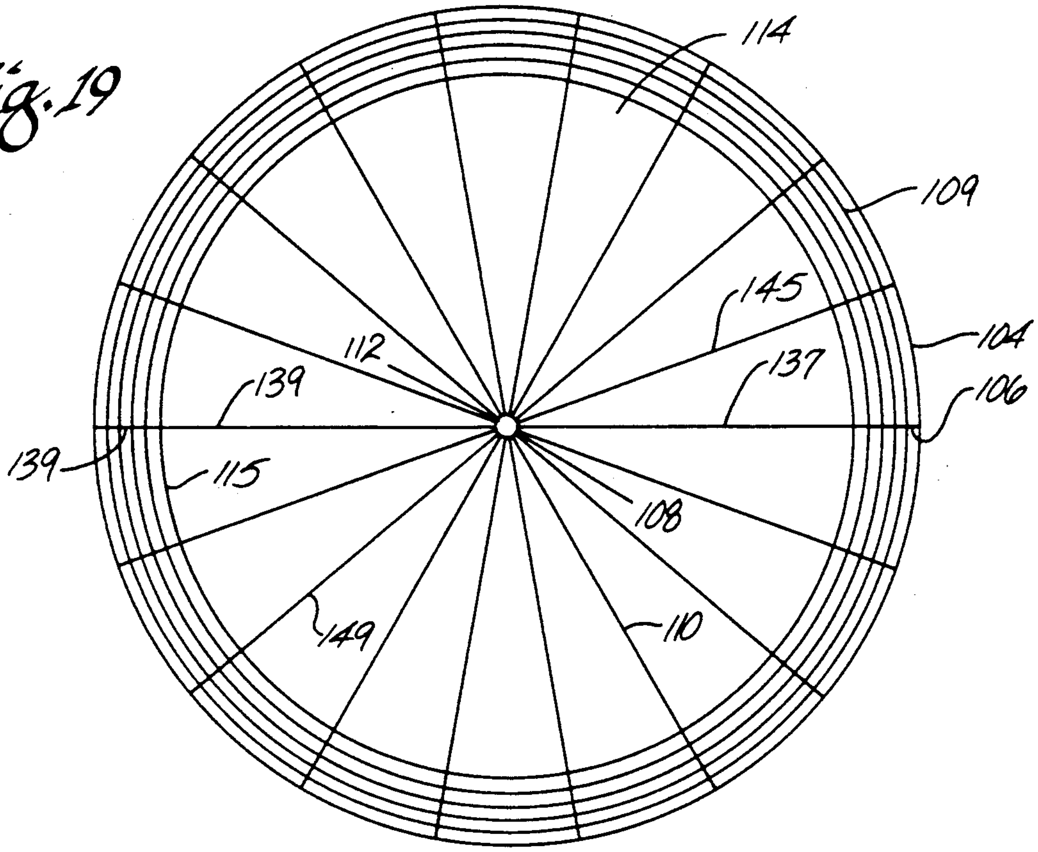
114
109
145
104
112
137
139
106
139
115
108
149
110

Fig. 20

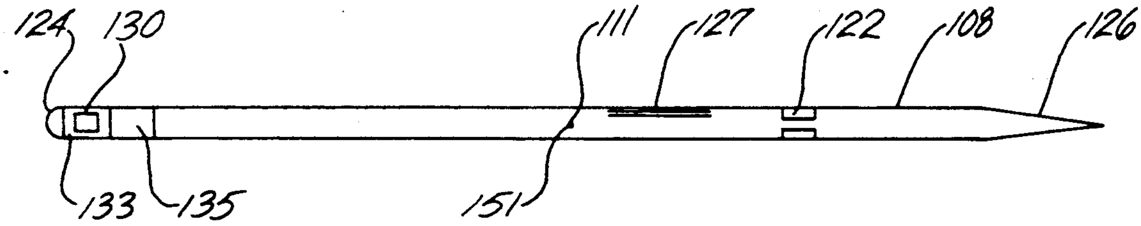
124
130
111
127
122
108
126
133
135
151

Fig. 21

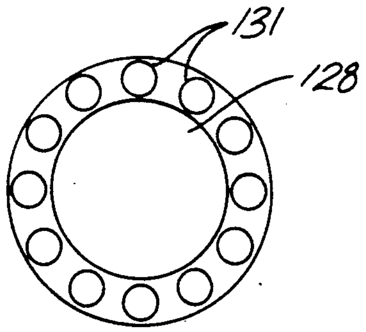

Fig. 22

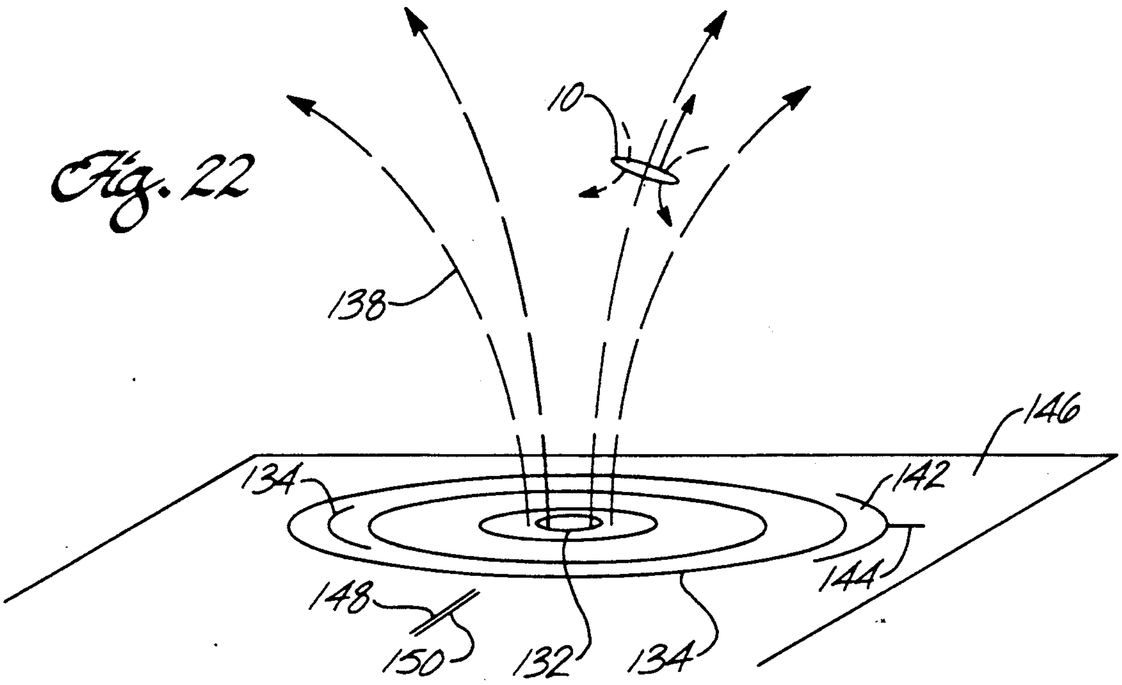
10
138
146
134
142
144
148
150
132
134

Fig. 23

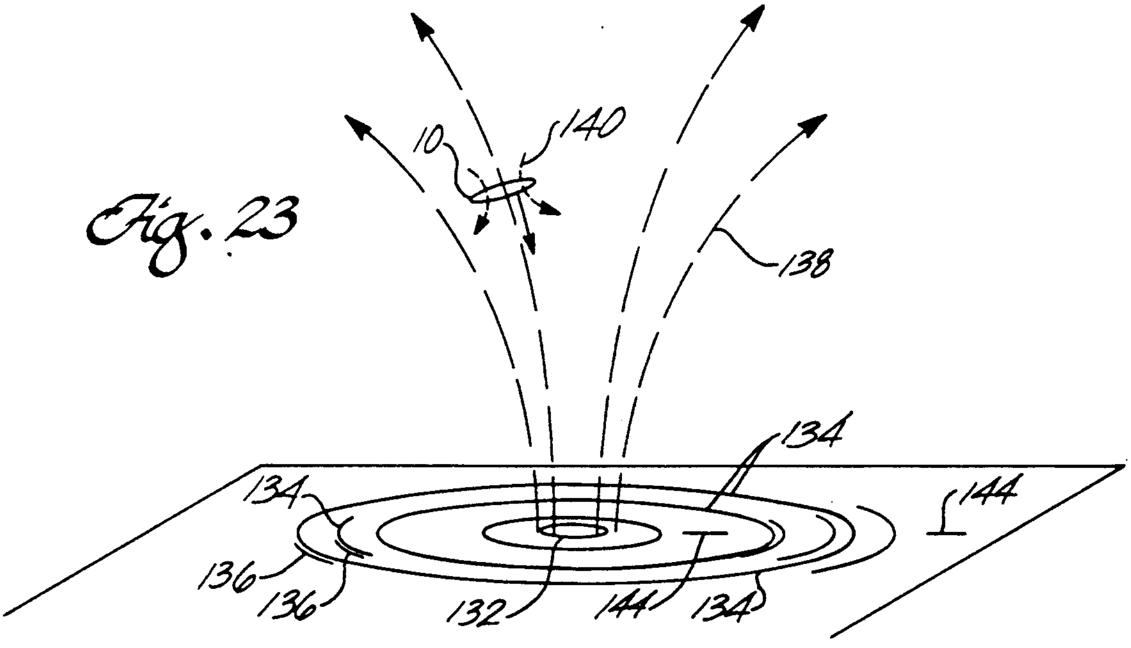
10
140
138
134
134
144
136
136
132
144
134

Fig. 24

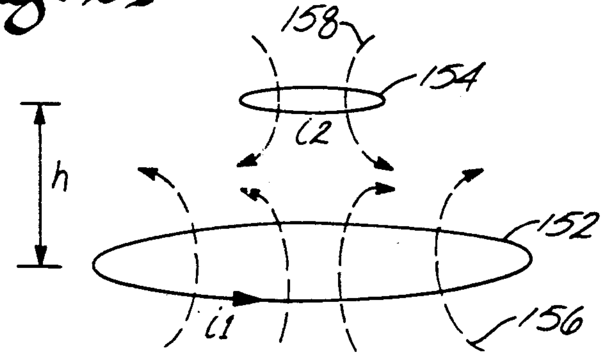
158
154
i2
h
152
i1
156

Fig. 25

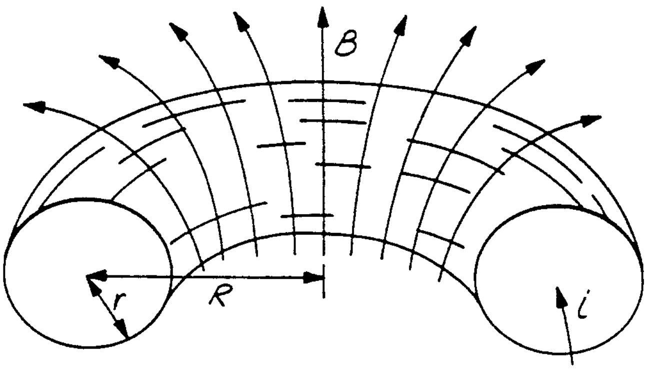

Fig. 26

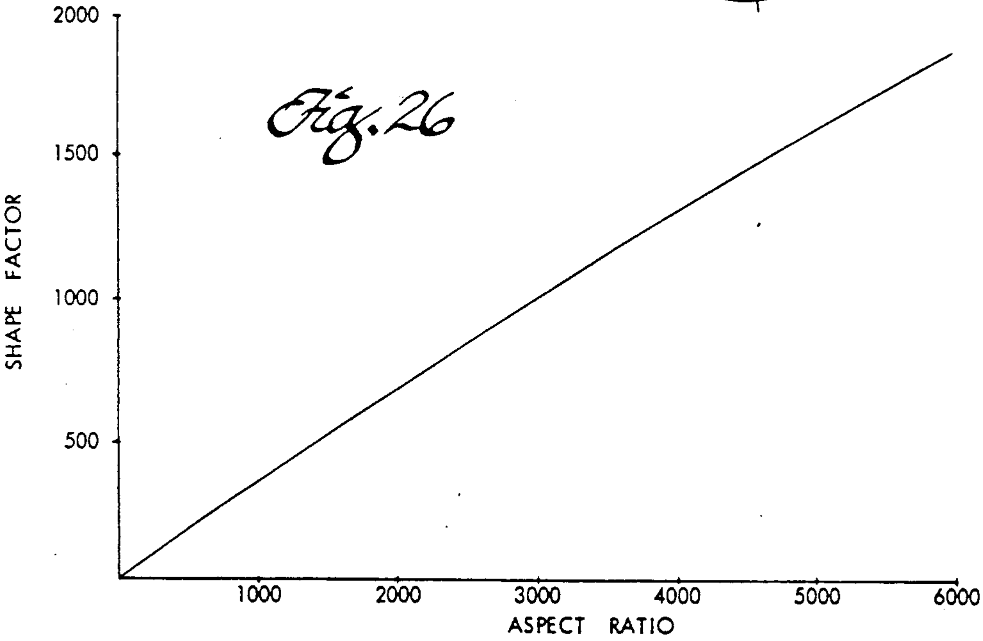

Fig. 27

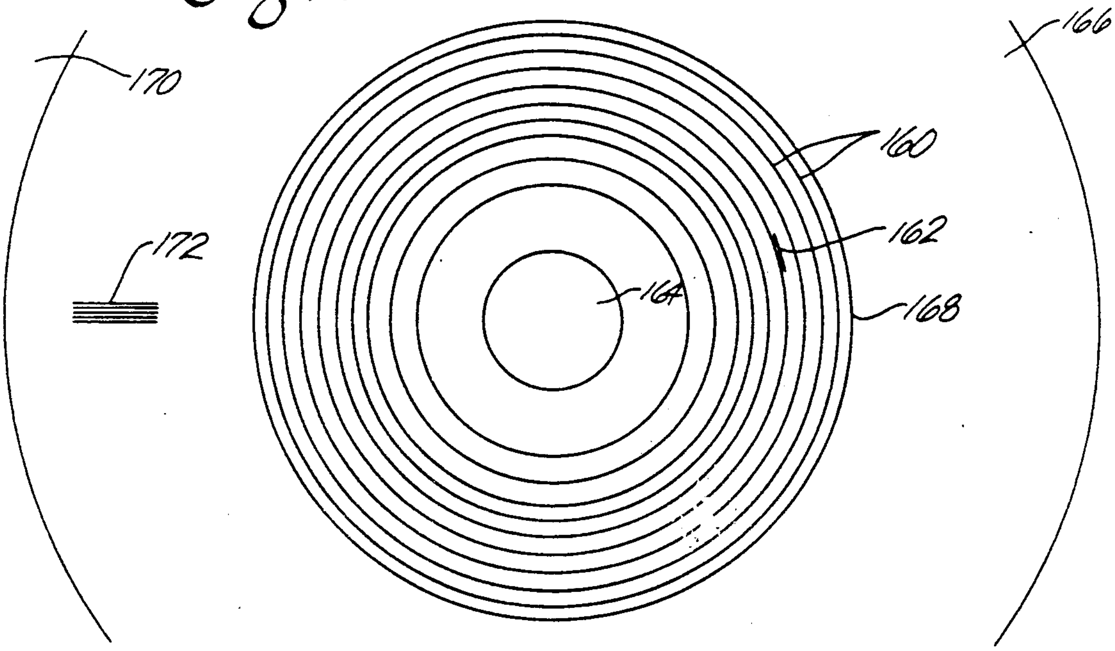
166
170
160
162
172
164
168

Fig. 28

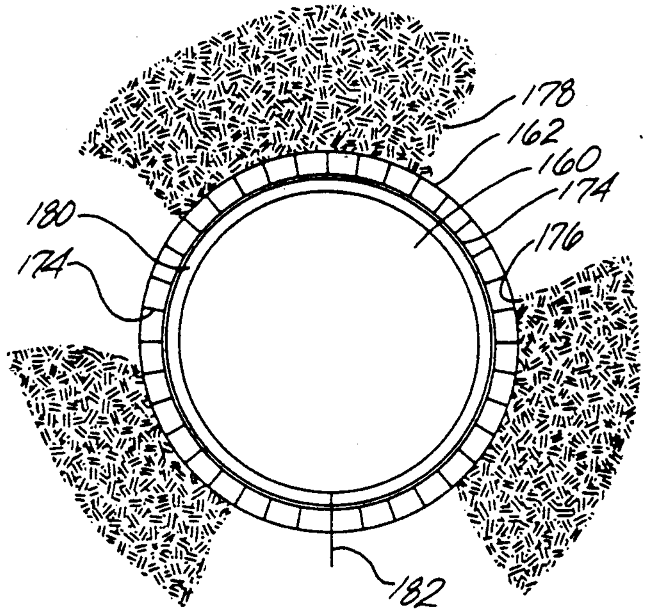
178
162
160
174
180
174
176
182

Fig. 29

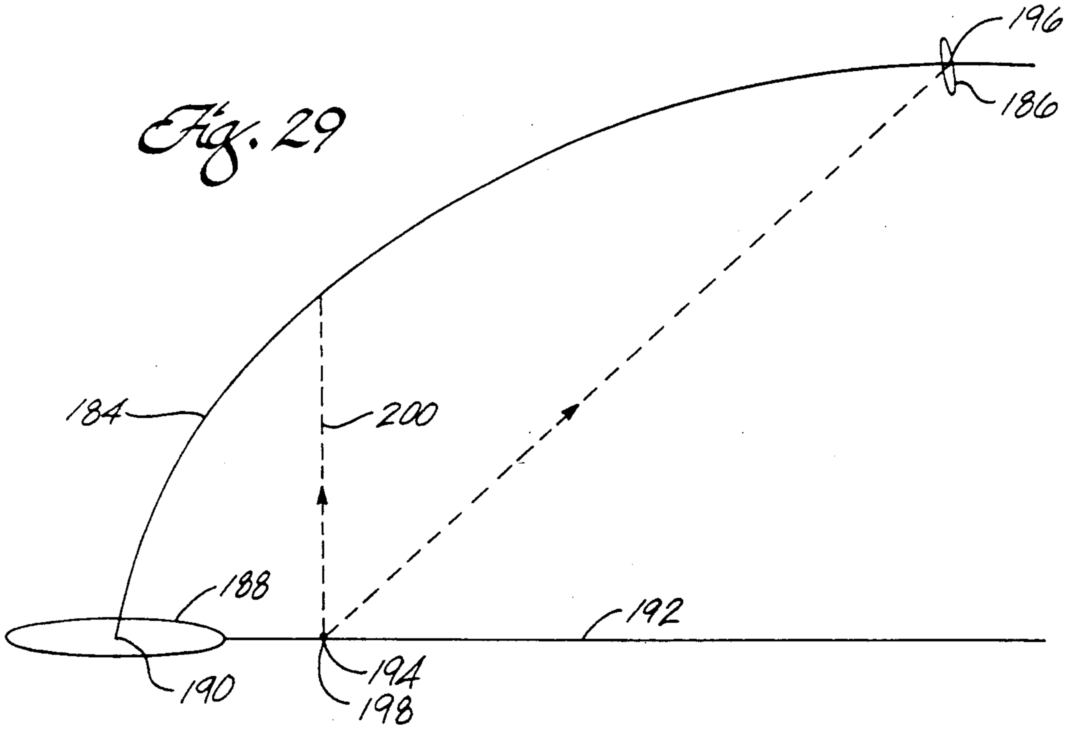
196
186
184
200
188
192
190
194
198

Fig. 30

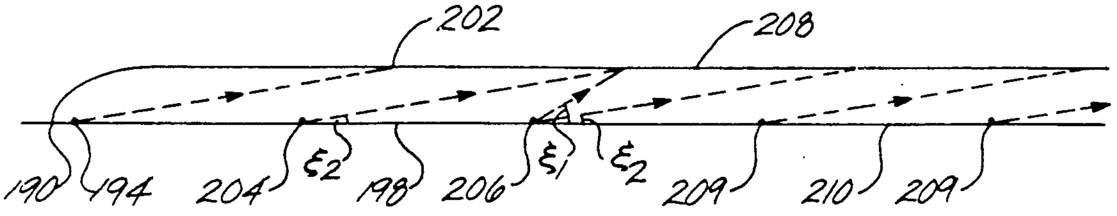
202
208
190
194
204
ε2
198
206
ξ1
ξ2
209
210
209

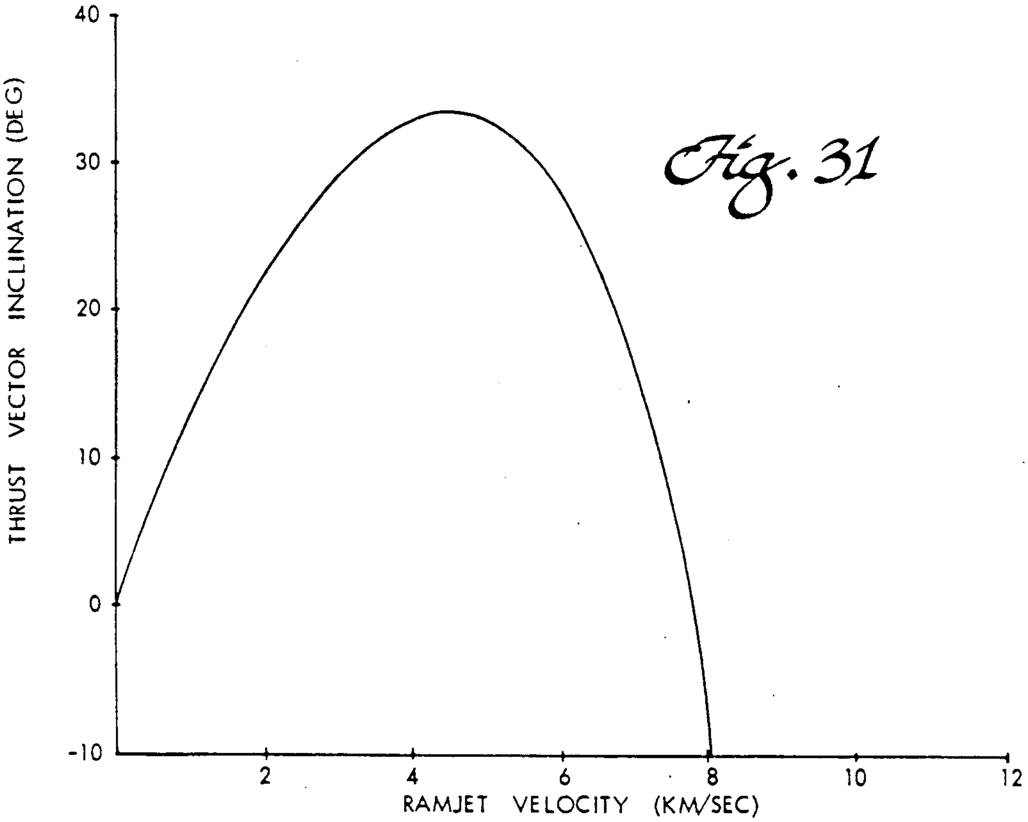
40
30
20
10
0
-10
THRUST VECTOR INCLINATION (DEG)
2
4
6
8
10
12
RAMJET VELOCITY (KM/SEC)

Fig. 31

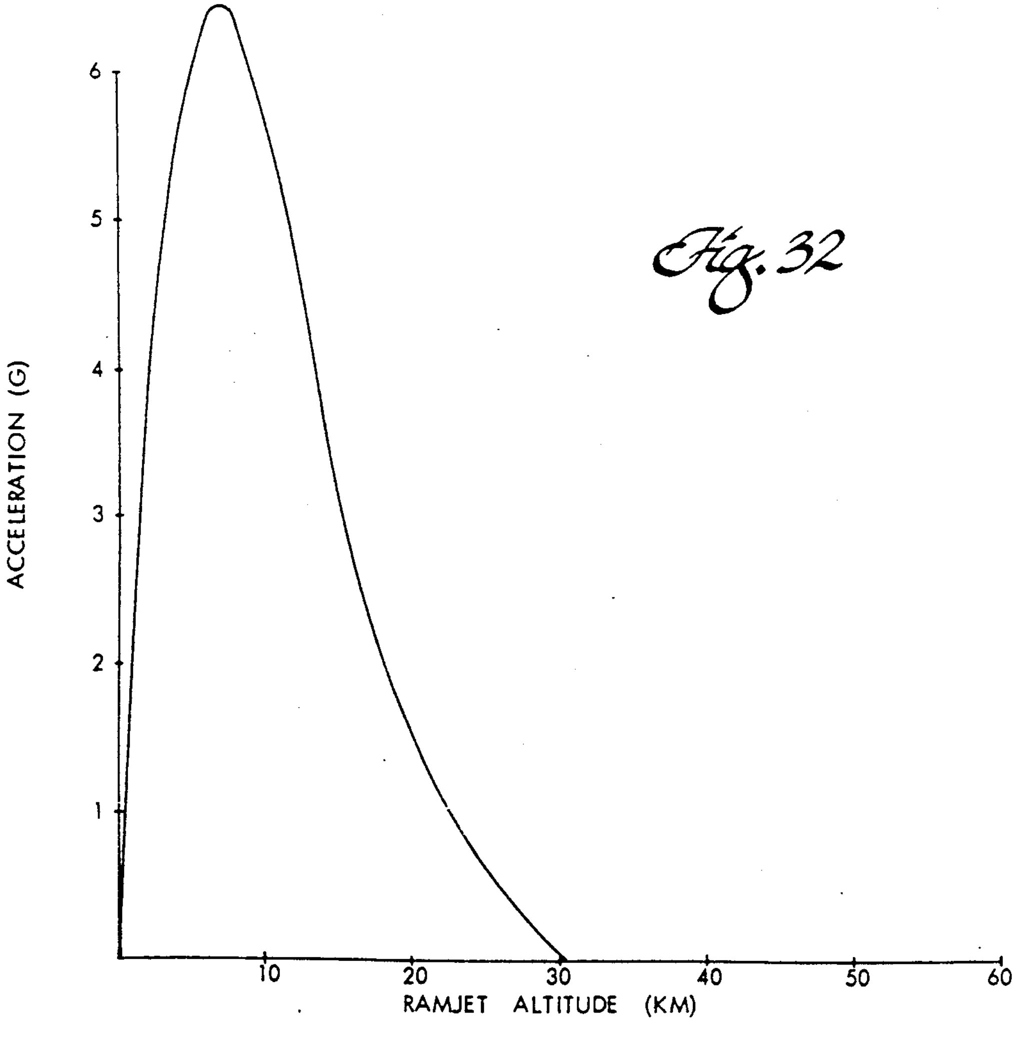
6
5
4
3
2
1
ACCELERATION (G)
10
20
30
40
50
60
RAMJET ALTITUDE (KM)

Fig. 32

Fig. 33

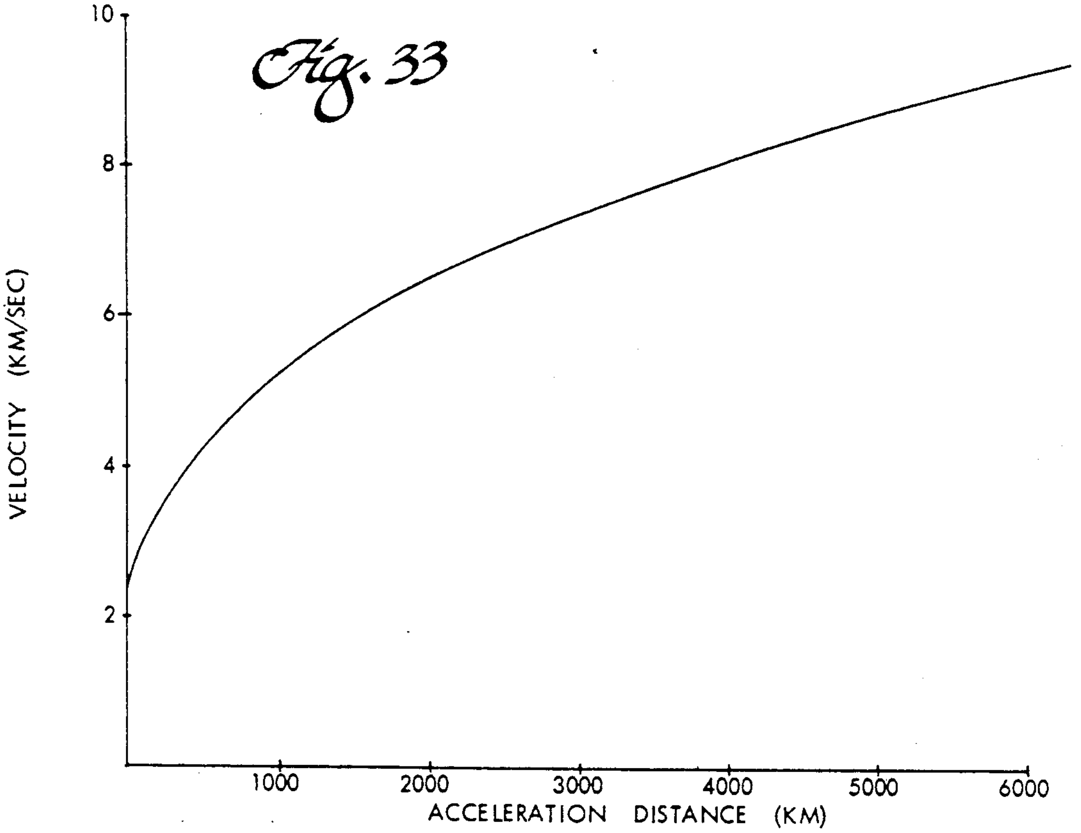
10
8
6
4
2
VELOCITY (KM/SEC)
1000
2000
3000
4000
5000
6000
ACCELERATION DISTANCE (KM)

Fig. 34

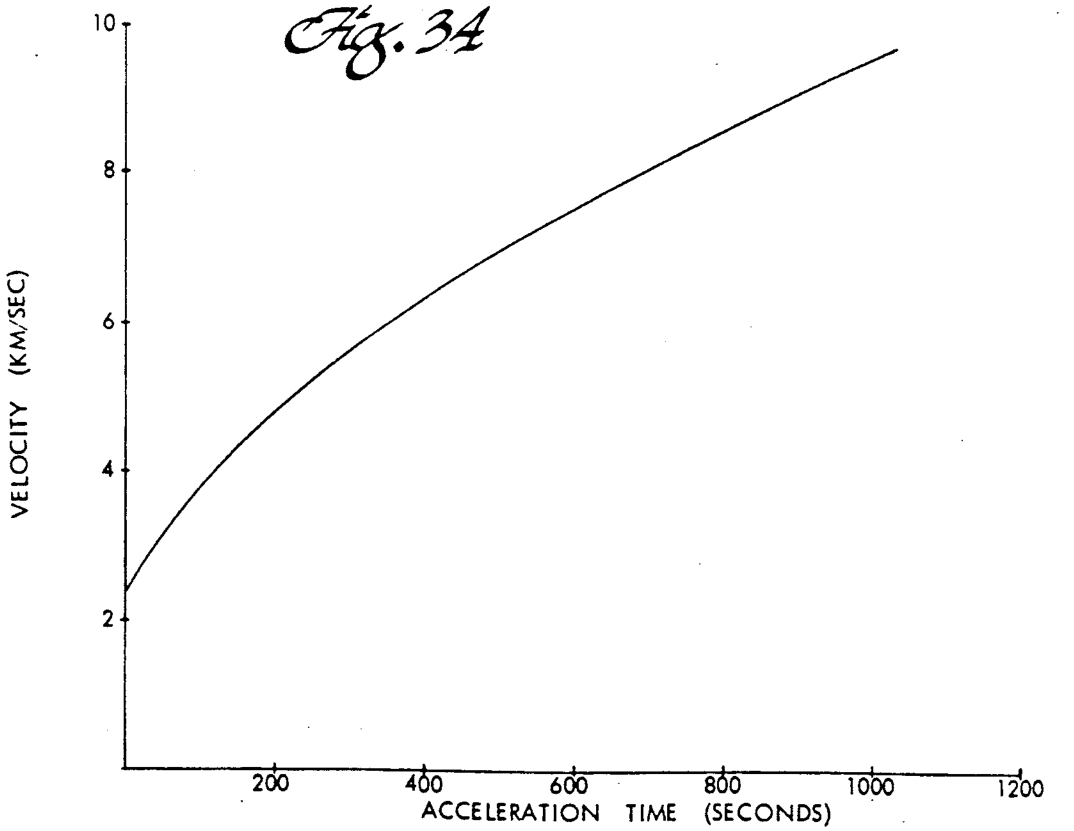
10
8
6
4
2
VELOCITY (KM/SEC)
200
400
600
800
1000
1200
ACCELERATION TIME (SECONDS)

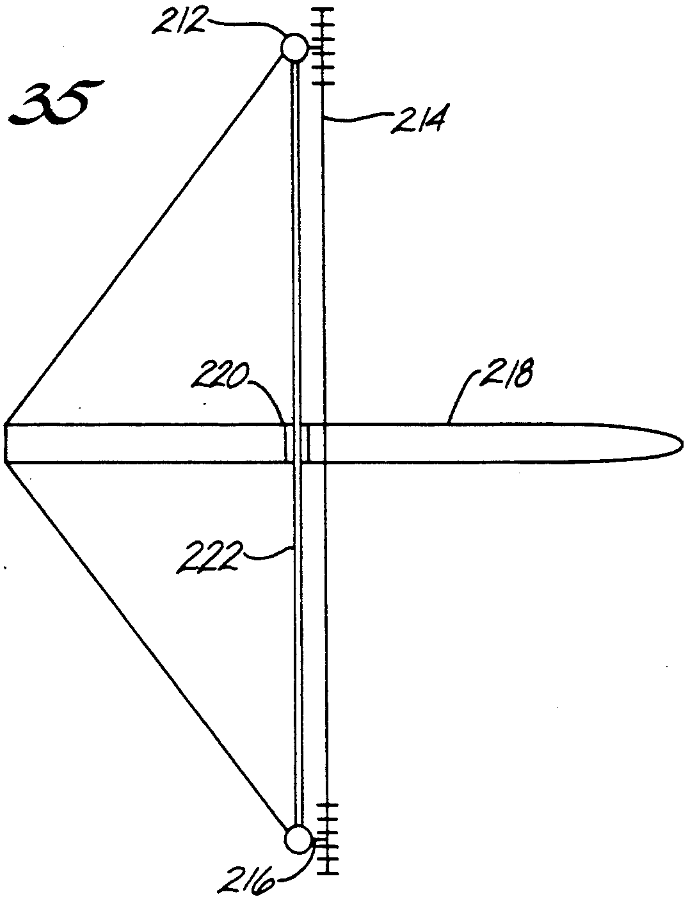
212
214
220
218
222
216

Fig. 35

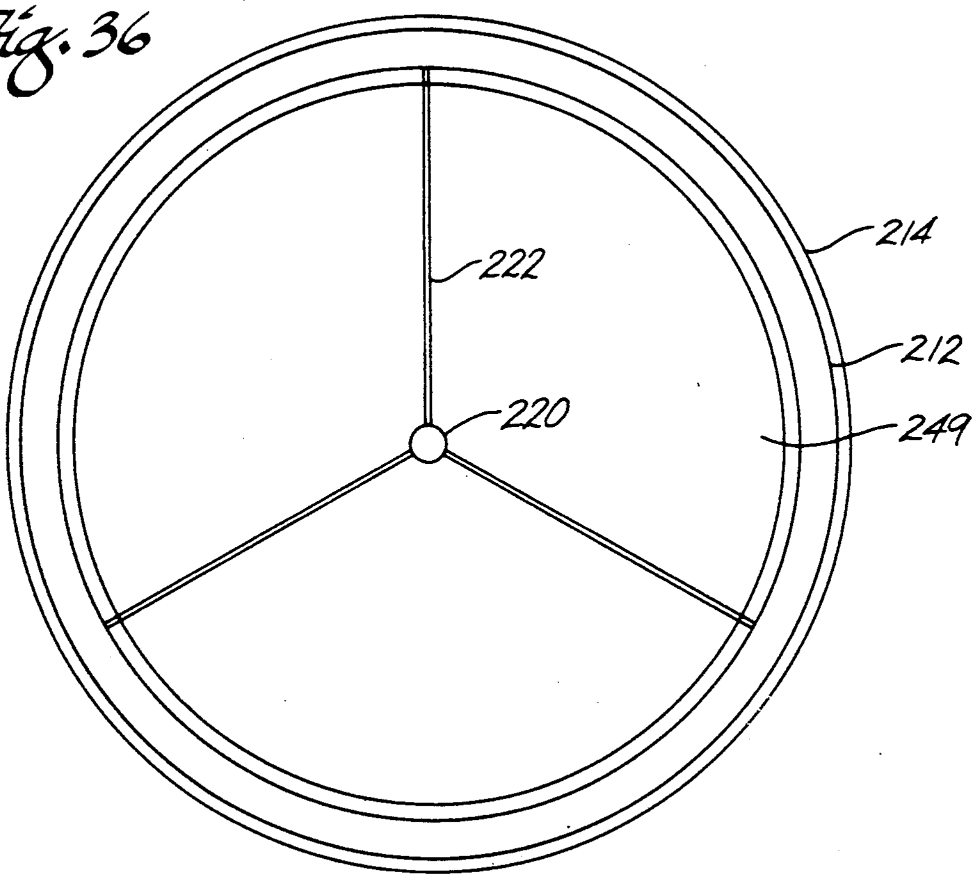
222
214
212
220
249

Fig. 36

Fig. 37

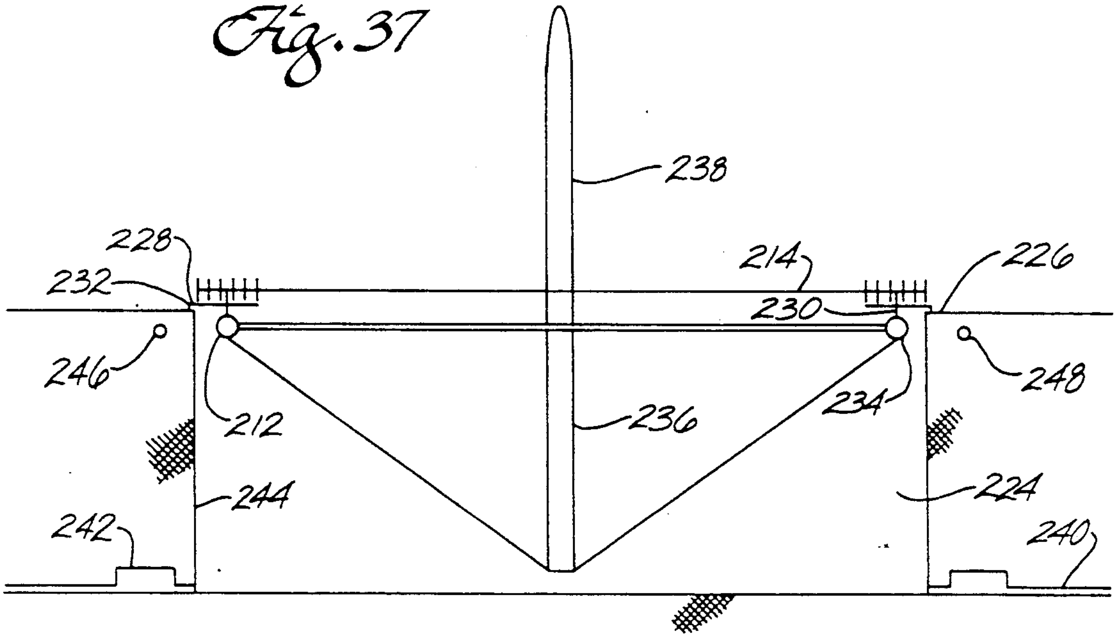
238
228
232
214
226
230
246
248
212
236
234
244
224
242
240

Fig. 38

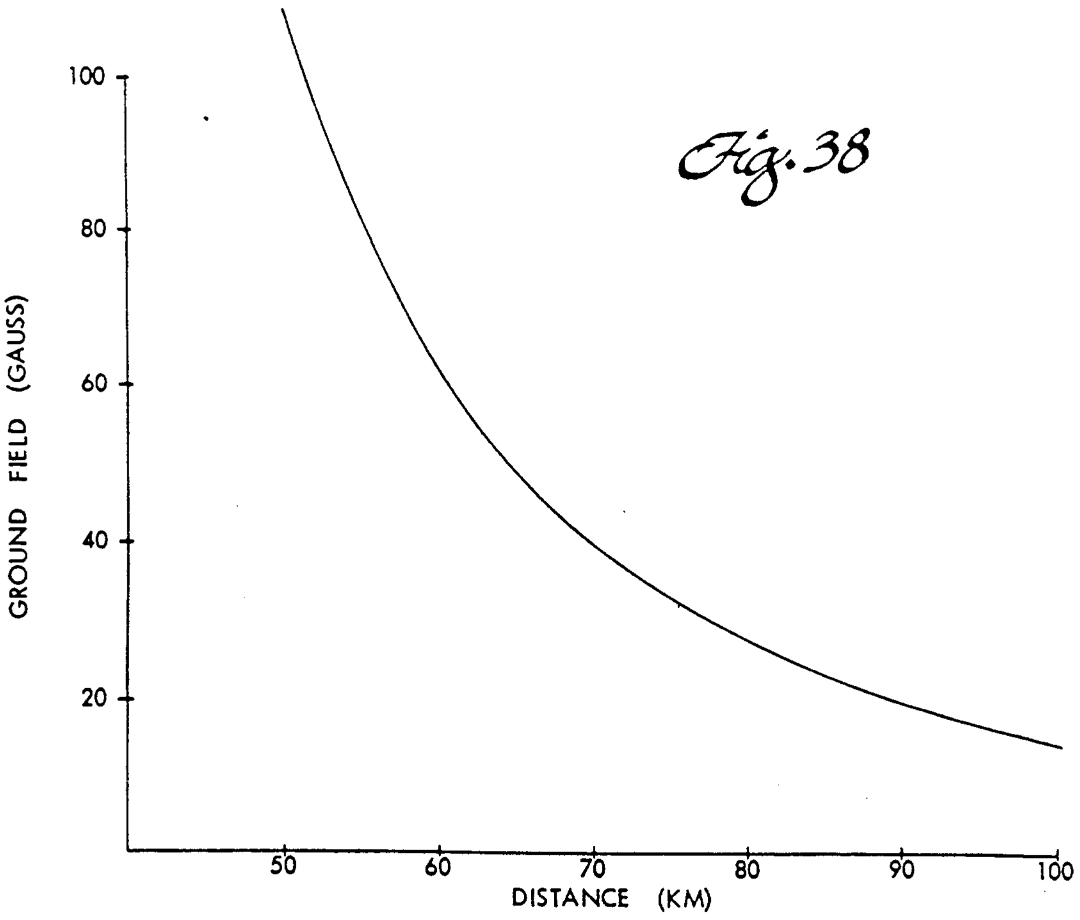
100
80
60
40
20
GROUND FIELD (GAUSS)
50
60
70
80
90
100
DISTANCE (KM)

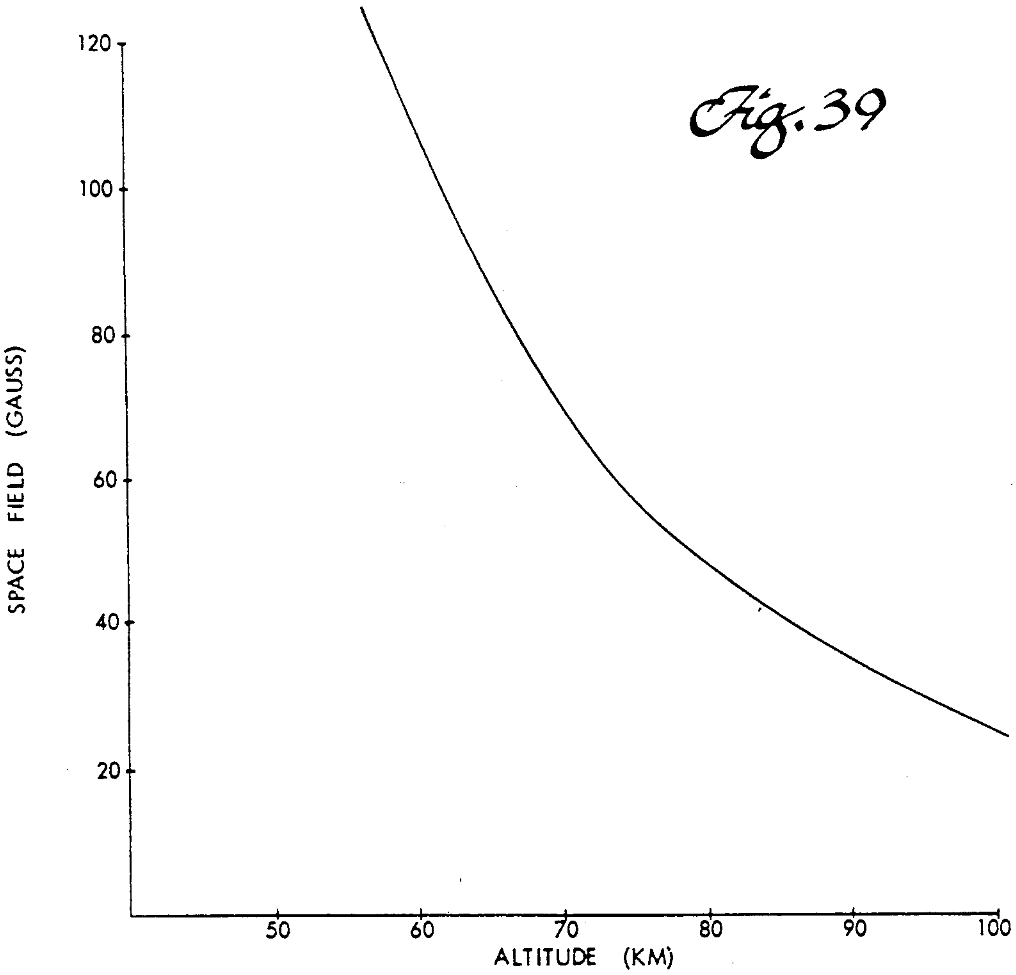
120
100
80
60
40
20
SPACE FIELD (GAUSS)
50
60
70
80
90
100
ALTITUDE (KM)

Fig. 39

Fig. 40

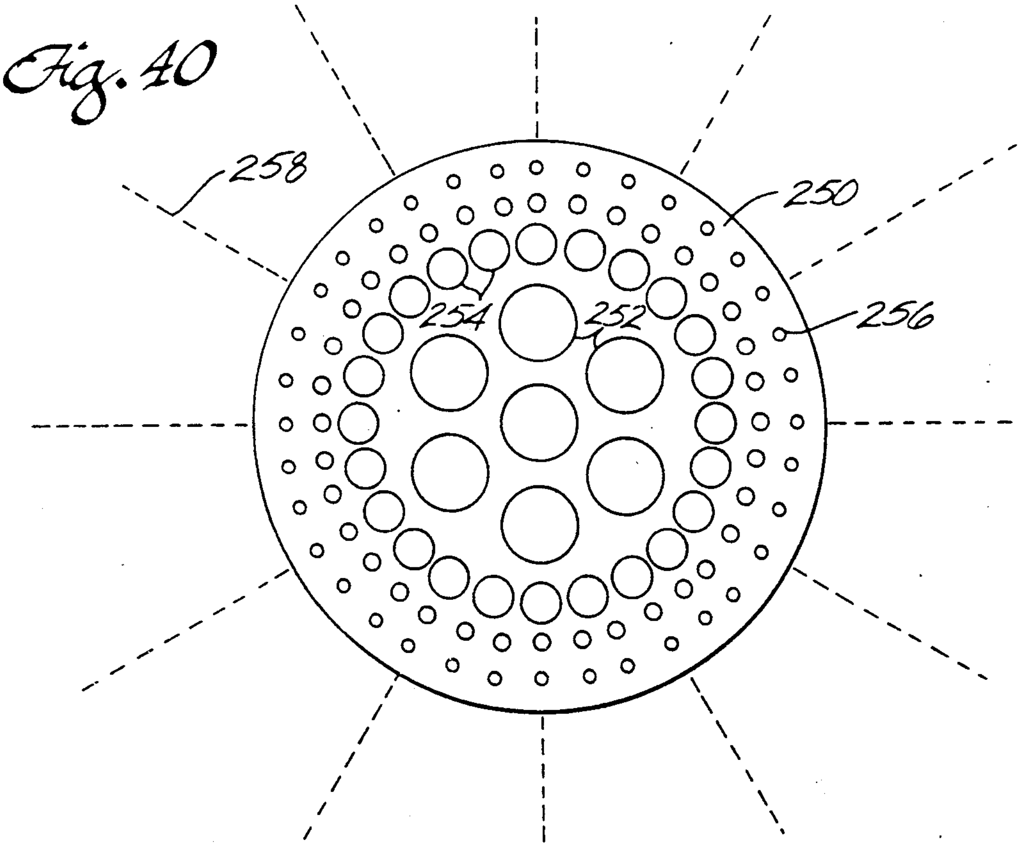
258
250
254
252
256

ELECTROMAGNETIC RAMJET

BACKGROUND

The problem of launching payloads into orbit from the earth's surface involves the development of a thrust generating mechanism powerful enough to lift a payload off the earth's surface and accelerate it to orbital velocities. In the prior art, this has been achieved with the development of high thrust rocket engines using the reaction principle. Unfortunately, rocket engines consume huge quantities of propellant that must be stored onboard the vehicle. Hence, in order to launch even relatively small payloads into orbit, the size of a rocket propelled launch vehicle must be enormous in order to contain all of the propellant. (The required propellant mass is over 100 times greater than the payload mass.)

The "scramjet" designed for the proposed ground-to-orbit aerospace plane, represents another type of propulsion system for accelerating payloads to high velocities. Basically, this engine is a ramjet in that it ingests air at some ingestion velocity (which is essentially equal to the engine's forward velocity through the atmosphere), sprays it with fuel which ignites forcing the expanding combustion gases to leave the engine with a velocity greater than the ingestion velocity. Since the ingestion velocity is very high, the ignition time must be very short. Only hydrogen fuel can be used in these scramjet engines because of the very short ignition time. Unfortunately, liquefied hydrogen has a very low density which requires fuel tanks five times larger than conventional fuel tanks (which must be thermally insulated since liquefied hydrogen is a very low temperature cryogenic fluid). This results in a high inert structural mass with a corresponding decrease in payload mass. See "Will the Aerospace Plane Work," *Technology Review*, January 1987, pp. 42–51 by S. W. Korthals-Altres. This problem is compounded by the fact that since the kinetic energy which the vehicle must develop in order to reach orbital velocity is so high (4×10^9 Joules/kg) the amount of hydrogen fuel that must be carried by the vehicle to achieve orbital velocity is about 56% of the total initial vehicle mass, even assuming optimal combustion efficiency. See "From Earth To Orbit In A Single Stage," *Aerospace America*, August 1987, pp. 32–34, by R. A. Jones and C. D. Donaldson. Thus, there are inherent fundamental engineering problems with this propulsion concept that cannot be circumvented even if the scramjet propulsion system can be made to operate as envisioned at orbital velocity (which may be a physical impossibility). Moreover, another problem will be introduced involving a severe limitation on payload size because nearly all of the interior volume of the vehicle will have to be filled with liquefied hydrogen.

Since the scramjet propelled reusable aerospace plane is currently believed to represent the cheapest method for achieving orbit (with an estimated cost of $200/lb compared to $2,000/lb for the Space Shuttle) the prospects for commercial space travel by private individuals in the 21st century appear to be very remote. (For example, it would cost a 200 lb passenger without luggage $40,000 to be transported to orbit on a one-way flight with the scramjet propelled vehicle.)

By utilizing the concept of beamed energy transmission where a vehicle's propulsive energy is generated off the vehicle at some remote power generating plant and transmitted to it via microwave or laser beams, it is possible to construct a ramjet which does not require the vehicle to carry any fuel. A microwave propelled ramjet was proposed as early as 1959 by William Brown. (See his U.S. Pat. No. 3,083,528 entitled "Microwave Engines" filed May 12, 1959.) This microwave propelled ramjet concept was later developed by Schad and Moriarty to propel payloads from the earth's surface directly into orbit. (See "Microwave Rocket Concept," *International Astronautical Congress*, Vol. 16, Athens, 1965, pp. 175–199.) Unfortunately, the problem of designing a sufficiently large receiving antenna for capturing the beamed microwave power and focusing it into the ramjet engine required for long range power transmission rendered the concept impractical. Since the wavelength of optical radiation is much smaller than micro wave radiation, it is possible to achieve significantly greater power transmission distances using a laser beam as the energy transmitting medium. A laser propelled ramjet was invented in 1972 by Kantrowitz and Rose. (See U.S. Pat. No. 3,818,700 entitled "Ram Jet Powered By A Laser Beam".) However, all of these prior art beamed energy ramjet propulsion concepts generate propulsive thrust by using the incoming beam energy to heat the ingested air to high temperatures so that it can be expelled through a conventional exhaust system with a velocity greater than the ingestion velocity. Thus, the incoming beam energy is not converted directly into propulsive thrust. These systems require the intermediate step of heating. Since all thermal rocket engines require a relatively small heating region, there is a practical upper limit on the amount of beam power that can be converted into propulsive thrust in thermal rocket engines that is set by power density limitations and by the thermal limitations of the surrounding structure. This intermediate heating step also results in a considerable reduction of efficiency. (More than 50% of the beam energy is wasted by radiative heat losses.) Since laser generators are not very efficient to begin with, prior art beamed power propulsion systems using laser beams will consume large amounts of input energy in the conversion to vehicle kinetic energy.

In theory, the most efficient method for propelling payloads into orbit is by means of an electromagnetic accelerator because the cost essentially reduces to the cost of generating an amount of electrical energy equal to the kinetic and potential energy of the total mass that is accelerated to orbit. For example, if the cost of generating electrical energy is 10¢/KW-hr, this cost is 90¢/kg or 41¢/lb for a 200 km high circular orbit. This is 5,000 times cheaper than the U.S. Space Shuttle and about 500 times cheaper than the proposed aerospace plane. Although there are several different types of prior art electromagnetic accelerators (which are also called mass drivers) that have been designed to accelerate bodies to high velocities (i.e., orbital velocities) from the earth's surface, they all have one common and very undesirable characteristic: they all require an evacuated launch tube through which the payload is accelerated. Therefore, unless an evacuated tube of several hundred kilometers is provided, the acceleration of prior art electromagnetic ground to orbit launchers are inherently high, and the mass and physical dimensions of the payload are too small to be of any significant practical value. See "Electromagnetic Launchers," *IEEE Transactions On Magnetics*, Vol. MAG-16, No. 5, Sept. 1980, pp. 719–721 by H. Kolm et al.

Large objects, such as completely assembled space based interplanetary transfer vehicles (ITVs) with diameters exceeding 25 m and lengths exceeding 100 m would be completely impossible to accelerate to orbit from the earth's surface by any prior art electromagnetic accelerator. In fact, completely assembled payloads with these dimensions could not be accelerated into orbit by any prior art ground to orbit transportation system (or any such system proposed for the future) because they would simply be too large to fit inside any launch vehicle. It is assumed without question and taken for granted that large objects designed for operating in earth orbit will have to be transported there piece by piece, in relatively small sections, and assembled in orbit.

The high cost and payload limitations of prior art launch vehicles is a result of the basic thrust generating principles used for their propulsion. For all practical purposes, these prior art propulsion principles are believed to be essentially unchangeable because they involve basic laws of physics. However, the discovery of a fundamentally new physical phenomenon or principle be applied to develop a completely new thrust generating propulsion concept. The electromagnetic ramjet propulsion concept disclosed herein is based upon one such discovery—superconducting materials with high critical temperatures. It will be shown that this propulsion concept will enable payloads to be orbited with mass and physical dimensions far beyond that which were previously believed to be possible. Moreover, since this propulsion concept is basically electromagnetic, it also enables payloads to be transported to orbit with minimum cost. The vacuum environment usually required for electromagnetic accelerators is provided by operating the ramjet in the ionosphere at very high altitude—instead of on the earth's surface inside a vacuum tube.

BRIEF SUMMARY OF THE INVENTION

With the foregoing in mind, the present invention provides an ultra high thrust electromagnetic ramjet propulsion system for accelerating high mass payloads with very large dimensions from the earth's surface directly into orbit. The propulsion system comprises a plurality of coplanar self-supporting superconducting dipole coils several hundred meters in diameter with very high aspect ratios that is initially accelerated to supersonic speed by magnetic repulsive forces generated by a plurality of giant superconducting field coils several kilometers in diameter mounted coaxially in circular underground tunnels located in a remote region. All of the superconducting coils are constructed with superconducting material having a critical temperature above liquid nitrogen thereby eliminating the need for expensive liquefied helium cryogenic refrigeration systems. The electromagnetic ramjet, along with the attached payload, is launched vertically and initially accelerated along a line of magnetic induction generated by the field coils.

The propulsion coils of the ramjet are mounted inside circular hypersonic wing-like airfoils equipped with movable aerodynamic control surfaces for guidance. The surfaces are moved by a plurality of electric servo motors controlled by a high accuracy inertial guidance system. The propulsion system can also be controlled from the ground, or from an orbiting satellite via radio signals. The individual superconducting dipole coils and their surrounding circular airfoils are maintained in a rigid concentric spaced-apart relationship by a plurality of radial airfoils that are mounted between adjacent circular airfoils. This system of circular airfoils containing the superconducting dipole propulsion coils and the radial airfoils mounted between them comprise the electromagnetic ramjet propulsion system. It is a self-contained automated reusable launch vehicle which is able to traverse through the atmosphere in controlled flight and land back at the launch site.

The payload is mounted inside a long cylindrical central body (fuselage) extending along the perpendicular longitudinal axis of the dipoles that is connected to the ramjet by a plurality of detachable cables. A reflecting grid of conducting wires is mounted across the inner dipole such that after the system is initially accelerated to supersonic speed at high altitude by the field coils, the low density atmospheric gas passing through the inner dipole is shock ionized via collision processes.

After the electromagnetic ramjet is launched from the earth's surface and accelerated to a relatively high initial altitude by the field coils, it is accelerated to orbital velocity by means of an ultra high power microwave beam. The frequency of the microwave beam is adjusted to produce electron cyclotron resonance with free electrons of the shock ionized gas moving away from the magnetic field of the dipoles. When the free electrons move into this resonance region they are rapidly accelerated in circular transverse paths that are perpendicular to the lines of magnetic induction of the dipoles. The electrons are thus pumped into very high energy cyclotron orbits by the microwave beam. A strong coupling between the electrons and the microwave beam is created resulting in the absorption of essentially all of the beam energy by the electrons. But this absorption is not thermalization. Almost all of this energy is pumped into extremely high energy electron cyclotron orbits. The high energy orbiting electrons become strong magnetic dipoles by virtue of their circular motion. Their magnetic moments are essentially aligned antiparallel to the coil's magnetic field. The magnetic field of the dipole coils decreases with increasing distance from its center giving rise to a magnetic gradient. This gradient is felt by the electron dipoles as a strong magnetic repulsive force which rapidly accelerates them away from the dipole coils. Although the remaining positively charged ions are not affected by the microwave beam, or by the magnetic field, the resulting charge separation accelerates them after the receding electrons. This generates a retarding force on the electrons. However, the electrons keep accelerating ahead of the ions propelled by magnetic repulsive forces. A strong magnetic repulsive recoil force is then exerted on the dipole coils propelling them forward. Basically, the orbiting electrons, in their aggregate, create a huge magnetic field behind the ramjet propelling it forward by magnetic repulsive forces. The process is purely electromagnetic so that in principal, the microwave beam can be converted into propulsive thrust with an efficiency approaching 100%. The dipole coils become a giant electromagnetic ionospheric ramjet several hundred meters in diameter that is propelled by enormous magnetic repulsive forces.

By constructing the microwave transmitter on the earth's surface as a close-packed, electronically steered, phased array with several million active elements each radiating microwave power on the order of one megawatt, it is possible to obtain CW beam powers of thousands of gigawatts for accelerating extremely large payloads of several thousand tons. The DC electric current used to operate the microwave transmitter can be obtained from the field coils which are utilized as a giant inductive superconducting energy storage system. By employing several microwave transmitters located along a prescribed ground track, the acceleration process can be extended over distances of several thousand kilometers thereby reducing the peak acceleration loads to relatively low levels that can be easily tolerated by human passengers. The propulsion system is so powerful that it will be possible to transfer huge payloads of thousands of tons directly from the earth's surface to lunar orbit.

Since the magnetic field strength directly over the field coils at orbital altitudes is inversely proportional to the cube of the altitude, the field at orbital altitudes directly over the coils is negligibly small (smaller than the earth's natural magnetic field) and decreases very rapidly with increasing altitude. Likewise, the magnetic field generated by the coils on the earth's surface in a region surrounding the outer perimeter of the coils (in the plane of the coils) also decreases very rapidly with distance. Thus, the magnetic fields generated by the field coils resulting from fundamental principles of electromagnetism are ideally suited for the propulsion concept proposed herein because these fields are sufficiently strong locally to generate enormous propulsive thrust, but decrease very rapidly with increasing distance so as to not affect the surrounding region at any significant distance.

A system of annular arrays of photovoltaic solar cells are mounted around the outer periphery of the field coils for generating electrical energy. This electrical energy is used for initially charging up the superconducting field coils and maintaining them with a desired inductive energy. The amount of inductive energy in the fully charged field coils is several orders of magnitude greater than the amount of energy used to launch a payload. Thus, the system will enable multiple launches to take place nearly simultaneously with massive payloads without any significant decrease in the magnetic field. By constructing large solar arrays with diameters of many kilometers, the system will generate significantly more electrical energy than is used for launching payloads. Thus, the cost of the electrical energy used in launching payloads is zero and the excess electrical energy generated by the arrays can be sold to utility companies for generating income revenue that is significantly greater than the total cost of operating the system.

Since the solar arrays can only generate electric power during daylight hours, the field coils are also utilized for accumulating a portion of the electrical energy generated during daylight hours so that it can be released into a power distribution system to various utility companies during the non-daylight hours. Thus, the field coils also serve as a giant load-leveling system that allows electric power to be fed info the distribution system at a nearly uniform rate 24 hours per day.

A large fleet of automated electromagnetic ramjet accelerators are provided operating from a large central launching/landing region using the same field coils and microwave transmitters ranging in size from a few dozen meters in diameter for accelerating small payloads of a few thousand kilograms to a diameter exceeding one kilometer for launching payloads exceeding 5×10^8 kg. The ground-to-orbit transportation system is envisioned as a vast circular complex with an outer diameter exceeding 80 km and an inner launching/landing region 10 km in diameter for providing very low cost transportation to orbit. It is also envisioned as a giant electric power generating plant.

DRAWINGS

These and other advantages and features of the invention will be apparent from the disclosure, which includes the specification with the foregoing and ongoing description, the claims, and the accompanying drawings wherein:

FIG. 1 is a schematic longitudinal cross section of an electromagnetic ramjet having a single superconducting dipole propulsion coil illustrating its basic operating principles;

FIG. 2 is a schematic transverse cross section of a single coil electromagnetic ramjet illustrating its basic design features;

FIG. 3 is a schematic longitudinal perspective view illustrating a microwave transmitter on the earth's surface accelerating an electromagnetic ramjet by a high power microwave beam;

FIG. 4 is a graph of magnetic field strength B versus the electron cyclotron resonant frequency of a microwave beam propelling an electromagnetic ramjet;

FIG. 5 is a parametric system of graphs illustrating the atmospheric breakdown power density for a microwave beam versus ambient atmospheric pressure and corresponding altitude for various frequencies;

FIG. 6 is a graph illustrating the percentage of microwave power that is transmitted through the atmosphere corresponding to a beam elevation angle of 30° as a function of frequency on a rainy day with a precipitation rate of 2 mm/hr.

FIG. 7 is a graph of propulsive thrust F versus velocity V generated by an electromagnetic ramjet powered by a 300 GW microwave beam and operating with an efficiency of 100%;

FIG. 8 is a schematic transverse cross section of a superconducting propulsion dipole mounted inside a tubular cryogenic Dewar system that is mounted inside a ramjet's circular airfoil illustrating the design and construction;

FIG. 9 is a schematic longitudinal cross section further illustrating the design and construction of a cryogenic Dewar system for a superconducting dipole coil that is mounted inside a circular airfoil of an electromagnetic ramjet;

FIG. 10 is an enlarged transverse cross section through the minor axis a propulsion coil illustrating the design and construction of a circular hypersonic winglike airfoil mounted around the propulsion coil with movable aerodynamic control surfaces;

FIG. 11 is a perpendicular cross section of **FIG. 10**;

FIG. 12 is a longitudinal cross section of a portion of the airfoil shown in **FIG. 10** further illustrating its design and construction;

FIG. 13 is a schematic longitudinal cross section of a giant multigigawatt phased array microwave transmitter;

FIG. 14 is a schematic transverse cross section of **FIG. 13** further illustrating the design and construction of a phased array microwave transmitter

FIG. 15 is a schematic plan view of a giant multigigawatt phased array microwave transmitter further illustrating its design and construction;

FIG. 16 is an enlarged longitudinal cross section of a portion of the phased array microwave transmitter illustrating the design and construction of the individual beam launching horns and the connecting waveguides;

FIG. **17** is a graph of the minimum beam diameter d_s versus transmission distance D corresponding to a phased array microwave transmitter with an effective transmitting aperture diameter of 600 m operating at a frequency of 10 GHz;

FIG. **18** is a schematic longitudinal cross section illustrating the design and construction of a multiple-coil electromagnetic ramjet with a cylindrical central body containing the payload;

FIG. **19** is a transverse view of the multiple-coil electromagnetic ramjet shown in FIG. **18**;

FIG. **20** is an enlarged schematic longitudinal cross section of the electromagnetic ramjet's rotatable central body which contains the payload and which is accelerated to orbital velocity by the ramjet;

FIG. **21** is an enlarged schematic transverse cross section through the rear of a central body illustrating its high thrust propulsion system;

FIG. **22** is a schematic perspective view showing a superconducting dipole coil of an electromagnetic ramjet accelerating away from a plurality of coaxial field coils buried beneath the earth's surface by magnetic repulsive forces;

FIG. **23** is a schematic perspective view showing a superconducting dipole coil of an electromagnetic ramjet being decelerated during its return to the launching site via magnetic repulsive forces generated by the field coils;

FIG. **24** is a schematic perspective view illustrating the magnetic repulsive force between two parallel coaxial current carrying coils;

FIG. **25** is a perspective view of half of a dipole coil showing its relative geometrical relationships;

FIG. **26** is a graph of a propulsion coil's shape factor versus its aspect ratio;

FIG. **27** is a schematic plan view illustrating the field coils in the preferred embodiment and a plurality of annular arrays of photovoltaic solar cells for generating electrical energy;

FIG. **28** is a schematic transverse cross section illustrating a superconducting field coil mounted inside an underground tunnel;

FIG. **29** describes the initial ascent trajectory of an electromagnetic ramjet after a vertical launch from the center of the field coils;

FIG. **30** shows the electromagnetic ramjet being sequentially accelerated along a 100 km high circular trajectory by a series of microwave transmitters positioned at 400 km intervals along the trajectory's ground track;

FIG. **31** is a graph of an electromagnetic ramjet's angle of inclination relative to its flight path versus velocity while accelerating a mass of 28,000,000 kg with a 1,200 GW microwave beam;

FIG. **32** is a graph of acceleration versus altitude of an electromagnetic ramjet's initial ascent trajectory to an altitude of 60 km propelled by magnetic repulsive forces generated by the field coils;

FIG. **33** is a graph of velocity versus acceleration distance for an electromagnetic ramjet accelerating a mass of 28,000,000 kg with a 1,200 GW microwave beam;

FIG. **34** is a graph of velocity versus time for an electromagnetic ramjet accelerating a mass of 28,000,000 kg with a 1,200 GW microwave beam;

FIG. **35** is a schematic longitudinal cross section of a fully assembled and manned 660 m diameter 25,000 ton toroidal space station attached to a 700 diameter electromagnetic ramjet that lifts the space station off the earth's surface and accelerates it to orbital velocity;

FIG. **36** is a transverse cross section of FIG. **35**;

FIG. **37** is a schematic longitudinal cross section illustrating how the space station shown in FIG. **35** is launched from the earth's surface by the electromagnetic ramjet and initially accelerated to an altitude of 60 km by magnetic repulsive forces generated by the surrounding field coils;

FIG. **38** is a graph showing the magnetic field strength on the earth's surface generated by the field coils at various distances from the center;

FIG. **39** is a graph showing the magnetic field strength in space directly above the center of the field coils at various altitudes; and

FIG. **40** is a schematic plan view of the central launching/landing complex inside the inner field coil with a plurality of individual circular launching/landing terminals for accommodating various electromagnetic ramjets with different diameters.

DESCRIPTION OF THE PREFERRED EMBODIMENT

The ramjet propulsion system disclosed herein is radically different from prior art ramjet propulsion systems in that it does not involve any thermodynamic processes and operates at very high altitudes. It uses electromagnetic energy to generate propulsive thrust instead of chemical energy obtained by burning fuel. Moreover, since this electromagnetic energy is converted directly into propulsive thrust without any intermediate thermalization stage, the input electromagnetic energy is converted into kinetic energy with nearly 100% efficiency. Since the ramjet operates in the ionosphere, aerodynamic drag is essentially zero. The electromagnetic energy is generated at a remote power generating plant and transmitted to the vehicle by a microwave beam. Consequently, it does not have any energy limitations. Since it does not have to accelerate any dead inertial mass such as an energy generating source or fuel load, it is capable of very high performance. The thrust generating mechanism involves converting the microwave power directly into magnetic repulsive forces generated by very low density atmospheric gas passing through the ramjet at high velocity. Consequently, unlike all prior art ramjet propulsion systems designed for operation within the earth's atmosphere, this electromagnetic ramjet propulsion system has no inherent velocity limitations. It is basically a giant electromagnetic accelerator that does not require a vacuum tube. Although it does require a partial vacuum environment, this environment is not provided artificially within a tube. Rather, it is obtained by operating the system at very high altitudes. Therefore, this accelerator has no size or payload mass limitations.

The underlying propulsion principles upon which the invention is based are made possible by the recent discoveries of superconducting material with critical temperatures, critical fields, and current densities significantly greater than previously believed possible. This fundamentally important technological breakthrough enables current carrying superconductors to be operated at liquid nitrogen temperature instead of much colder liquid helium temperature. Since liquid hydrogen is colder than liquid nitrogen and can be easily stored and maintained aboard space vehicles, it is possible to maintain extremely large superconducting coils in orbit that are capable of generating very large and intense magnetic fields indefinitely.

FIGS. **1** and **2** are schematic longitudinal and transverse cross sections of an electromagnetic ramjet with a single dipole coil illustrating its basic design and operating principles. As is illustrated in these figures, the ramjet is extremely simple and comprises a superconducting self-supporting dipole coil 10 several hundred meters in diameter. The dipole **10** moves through the rarified upper atmosphere at supersonic speed such that the plane of the dipole is nearly perpendicular to its velocity vector. A grid **12** of conducting wires **14** is mounted as a screen across the inside periphery of the dipole. This screen has two functions. It provides a means for ionizing the low density atmospheric gas passing through it at supersonic speed by generating small shock waves via collision processes; and it also serves as a microwave reflector which prevents the microwave beam from passing through the dipole. The wires **14** are mounted inside small stress bearing tubular filaments of fused silica glass (or some other suitable high temperature material) as thermal protection for the wire screen.

As is shown in FIG. **1**, the superconducting dipole coil **10** is charged with electric current which generates a very large magnetic field **16**. An ultra high power microwave beam **18**, is transmitted to the dipole **10** into the magnetic field **16** from a ground based transmitter **20** located several hundred kilometers directly behind the dipole and on its ground track, FIG. **3**. (In some embodiments, the microwave transmitter may be located in orbit.) The microwave beam **18** is plane polarized such that its electric field vector is perpendicular to the dipole's central magnetic field. When the low density atmospheric gas **22** passes through the screen **12** inside the dipole coil **10**, a plurality of small shock waves are generated behind the screen **12**. This system of multiple shock waves intersect a short distance behind the screen **12** and forms a low density cloud of ionized gas **24** moving away from the dipole **10** with a velocity approximately equal to the dipole's forward velocity through the atmosphere (i.e., the ramjet's intake velocity) In order for the ramjet to generate forward propulsive thrust, the ionized gas cloud **24** must be accelerated away from the dipole with a relative velocity significantly greater than the intake velocity. This acceleration will be accomplished by the principle of electron cyclotron resonance. The electromagnetic acceleration mechanics of this principle can be understood by referring to FIG. **1**.

The frequency of the microwave beam **18** is adjusted to produce electron cyclotron resonance with the free electrons **25** passing into a region **26** of the magnetic field of the dipole **10** located several meters behind the shock screen **12**. This resonance frequence is given by the equation

$$f=(B/2\pi)(e/m) \qquad (1)$$

where the term e/m denotes the charge to mass ratio of the electron which is 1.7592×10^{11} coloumb per kg and where B denotes the field strength. When the free electrons **25** of the ionized gas **24** move into this region **26** they are rapidly accelerated in circular transverse paths **28** that are perpendicular to the magnetic field lines of induction **16**. The free electrons are thus pumped into very high energy cyclotron orbits by the transverse electric field vector of the microwave beam that is oscillating at a frequency f. A strong coupling between the low density plasma cloud and the microwave beam is created resulting in the absorption of essentially all of the beam energy by the plasma cloud **24**. But this absorption is not thermalization. Most of the energy is pumped into extremely high energy electron cyclotron orbits. Any gas passing through the dipole screen **12** that remains unionized, is ionized when it enters the resonance region **26**.

The high energy orbiting electrons become strong magnetic dipoles by virtue of their circular motion. However, because of "Lenz's Law", their magnetic moments are aligned anti-parallel with the magnetic field. The magnetic field strength decreases with increasing distance from the dipole **10** giving rise to a strong magnetic gradient ∇B. This gradient is felt by the electron dipoles j as a magnetic repulsive force which rapidly accelerates them away from the dipole **10**. Although the remaining positively charged ions are not affected by the magnetic gradient ∇B, the resulting charge separation accelerates them after the accelerating electrons. A magnetic repulsive recoil effect is then exerted on the dipole coil **10** thereby creating a propulsive reaction force. Eventually, the orbital energies of the electrons decay to relatively low levels thus converting their kinetic energy from transverse circular motion to directed linear motion. Basically, the orbiting electrons, in their aggregate, create a huge magnetic field behind the dipole coil **10** propelling the coil forward by magnetic repulsive forces. The inertia of the positively charged ions enables the repulsive forces generated on the dipole coil **10** to be very strong. With this underlying accelerating principle, it is possible to convert essentially 100% of the microwave beam power directly into magnetic propulsive thrust that is exerted on the superconducting dipole coil **10** by the accelerating gas cloud **24**.

FIG. **4** is a graph of magnetic field strength B versus the electron cyclotron resonant frequency f of a microwave beam propelling an electromagnetic ramjet given by equation (1).

Since the superconducting dipole coil which represents the ramjet is several hundred meters in diameter, the total mass flow rate of low density atmospheric gas passing through the coil is fairly high. Since the diameter of the microwave beam is equal to the diameter of the dipole coil, the beam power can reach levels of thousands of gigawatts. Consequently, this high altitude electromagnetic ramjet concept can generate propulsive thrust far beyond any prior art ramjet, rocket engine or any other propulsion system.

The technique of accelerating a low density plasma to high velocities by the principle of electron cyclotron resonance using microwave power as the energy source is usually applied in high energy physics laboratories concerned with the investigation of nuclear processes. The principle has also been studied as a means for accelerating and expelling propellant stored onboard a space vehicle for generating small amounts of propulsive thrust the microwave power is only on the order of a few kilowatts and generated onboard the vehicle. (See the article "Cyclotron Resonance Thruster Design Techniques," *AIAA Journal,* Vol. 4, No. 5, May 1966, pp. 835–840.) However, the principle of electron cyclotron resonance has never been considered as a means for accelerating atmospheric gas passing outside a vehicle in the form of a ramjet using microwave power transmitted to the vehicle from a remote power generating station.

The propulsive thrust F generated by the electromagnetic ramjet can be expressed as

$$F = \frac{P}{V} \qquad (2)$$

where P is equal to the beam power and V is equal to the dipole velocity. Assuming that the diameter of the microwave beam is equal to the diameter of the propulsion dipole **10**, the beam power P can be expressed as

$$P=\pi R^2\rho \qquad (3)$$

where R is equal to the dipole radius and ρ is equal to the power density. The maximum power density ρ_{max} of the microwave beam that can be transmitted to the dipole is limited by the breakdown limitations of the atmosphere. This maximum power density can be expressed as

$$\rho_{max}=Cp^2[1+(f/5.3\times 10^9 p)^2 \qquad (4)$$

where p is equal to the ambient atmospheric pressure in Torr, and C is a constant equal to 119.38 W/(cm^{2}Torr^2). FIG. **5** is a parametric system of graphs of ρ_{max} versus atmospheric pressure p corresponding to several microwave frequencies f. (The altitude corresponding to various values of atmospheric pressure is also shown in FIG. **5**.) For a more detailed technical description of microwave breakdown in air at various pressures see "Breakdown Limitations on the Transmission of Microwave Power Through the Atmosphere," *IEEE Transactions On Antennas And Propagation,* Vol. 12, Nov. 1964, pp. 709-717, by W. E. Scharfman et al. FIG. **5** indicates that in order to obtain high power densities, the frequency f of the microwave beam should be as high as possible.

If the frequency of the microwave beam is too high, atmospheric attenuation begins to reduce the transmission efficiency of microwave power through the atmosphere. This is due to the absorption of microwave power by water vapor, oxygen, and various scattering processes. FIG. **6** is a graph showing the percentage of microwave power that is transmitted through the atmosphere with a beam elevation angle of 30° as a function of frequency during a cloud covered rainy day with a precipitation rate of 2 mm/hr. Thus, for example, at a frequency of 10 GHz, the microwave power transmission through the atmosphere with a beam elevation angle of 30° exceeds 95% even when it is raining with a precipitation rate of 2 mm/hr and total cloud cover. A laser beam operating at optical frequencies would be completely absorbed by the atmosphere under these conditions. A detailed technical description of atmospheric attenuation of microwave power can be found in the article, Atmospheric Attenuation of Microwave Power," *The Journal of Microwave Power,* Vol. 5, No. 4, Dec. 1970, pp. 269-278 by V. Falcone.

FIGS. **5** and **6** are important for determining the optimum operating frequency for the microwave beam. In order to be able to obtain a high power density ρ, the frequency should be as high as possible. However, if the frequency is too high, atmospheric attenuation will reduce the power transmission efficiency. Since it will be shown herein that the transmitter will be able to provide almost unlimited beam power, power losses due to atmospheric attenuation will be of secondary importance. Thus, the frequency selected for the preferred embodiment will be relatively high in order to obtain a high power density. A frequency f=10 GHz will provide a power density of 425 W/cm^2 and is selected for the preferred embodiment. It should be noted however, that this frequency will still provide very high transmission efficiencies which, on a clear day, will approach 98% for beam elevation angles on the order of 30°.

In order to comprehend the very high performance capabilities of the electromagnetic ramjet suppose R=150 m and f=10 GHz (f=10^{10}). The maximum power density ρ_{max} of the microwave beam operating at this frequency is 425 watts/cm^2. Hence, the maximum effective beam power that can be used to propel the ramjet will be 300 GW. This power is many orders of magnitude greater than any prior art beamed power propulsion concept—and this power is converted directly into propulsive thrust by electromagnetic processes with an efficiency of nearly 100%. FIG. **7** is a graph of propulsive thrust F versus velocity V that could be generated by the ramjet using 300 GW of propulsive power.

The reason why it is possible to operate the ramjet at these very high power levels is because of its huge dimensions—and these dimensions are possible because the superconducting propulsion coil can create and sustain enormous magnetic fields with a relatively small amount of material.

The most efficient operating altitudes of the electromagnetic ramjet (where the beam power to propulsive thrust conversion efficiency approaches 100%) is between 100 km and 120 km. At an altitude of 100 km the particle density of the atmosphere is about 10^{13} molecules/cm^3 and the ramjet operation will be very nearly collisionless. Thus, there will be essentially zero thermalization and the ramjet operation will involve purely electromagnetic processes. Since altitudes on the order of 100 km are well within the ionosphere the ramjet can be viewed as "ionospheric ramjet."

It should be pointed out that the shock screen **12** used to ionize the molecules passing through the dipole coil **10** is not necessary at 100 km altitudes. The ionosphere at 100 km altitudes contains some fraction of the molecules that are already ionized by natural solar radiation. When the superconducting dipole coil **10** sweeps through this region with its large magnetic field, it acts as a giant ion scoop several kilometers in diameter which collects the ionized molecules and free electrons from all directions. Thus, the fraction of ionized molecules passing through the dipole is several orders of magnitude greater than the natural fraction at this altitude. When the free electrons pass into the electron cyclotron resonance region **26** (FIG. **1**) behind the moving dipole coil and rapidly accelerated into expanding circular cyclotron orbits they will, in themselves, form an ionizing region, which will ionize most of the remaining unionized gas molecules passing through it in an avalanche effect. Although this process will reduce the overall beam power to propulsive thrust conversion efficiency of the ramjet, this reduction will only on the order of 2 or 3%.

At low altitudes below the ionosphere the fraction of ionized gas molecules is not large enough to start an ionizing process. The shock screen **12** therefore provides a means for generating a plasma source when the ramjet is moving at relatively low altitudes.

It should also be pointed out that the microwave generators could easily be designed to change frequency. This capability would allow the electromagnetic ramjet to be "fine tuned" from the ground while it is accelerating a payload to optimize its propulsive efficiency. It follows from equation (1) that by varying the frequency f of the microwave beam, it will be possible to move the resonance region of the orbiting electrons to give optimum performance. Consequently, in view of the large dimensions of the resonance region and the nearly perfect beam quality, it will be possible to operate the electromagnetic ramjet close to its theoretical performance limits.

One of the most important technical aspects of the invention is the fact that the superconducting dipole coil **10** is constructed with superconducting material having a relatively high critical temperature such that it can be maintained at liquid hydrogen temperature (18° K) instead of very low liquid helium temperature (4° K). FIGS. **8** and **9** illustrate the design and construction of a cryogenic Dewar system **30** that is designed to maintain the coil **10** at liquid hydrogen temperature. The system **30** comprises a double walled tube **32** that completely encloses the coil **10**. A plurality of stand-off mounting struts **34** keep coil centered inside the tube **32**. The tube **32** is filled with liquid hydrogen **36** such that the coil **10** is completely immersed in this cryogenic fluid at 18° K. The tube **32** and mounting struts **34** are constructed with material having very low thermal conductivity. The space **38** between the double walls of the tube **32** is filled with multiple layers of cryogenic thermal insulation **40**. A plurality of larger mounting struts **42** are attached to the outer tube wall **44** and used for supporting an outer circular wing-like, high-speed airfoil **46** that completely encloses the Dewar **30**.

FIGS. **10**, **11**, and **12** describe the detail design and construction of a high-speed airfoil **46** that is mounted around the coil **10** and its Dewar **30**. As described above, the airfoil **46** is attached to the Dewar **30** by a plurality of supporting struts **42**. Some portions of the airfoil **46** are equipped with movable aerodymamic control surfaces **48** that enable the ramjet propulsion system to be steered while traversing through the lower atmosphere at high velocity. The control surfaces **48** are moved by electrically driven servo motors **50** that are controlled by a guidance computer **52** and an inertial guidance system **54**. A back-up radio control system **56** is provided to enable the control surfaces **48** to be controlled via radio signals transmitted from the launch site or from an orbiting satellite. A plurality of small orbiting maneuvering thrusters **58** (which could be high pressure vessels of compressed gas) are also provided for generating small amounts of propulsive thrust while traversing through the ionosphere (or while in orbit above the atmosphere). These thrusters **58** are controlled by the guidance system **54** and can also be used for providing lateral guidance when the propulsion system is moving at low velocities near the earth's surface. The airfoil **46** is also equipped with a plurality of high speed dive brakes **60** for providing additional decelerating propulsive thrust while traversing through the lower portions of the atmosphere. The dive brakes **60** are also controlled by the guidance system **54**.

The airfoil **46** and the movable control surfaces **48** are constructed with high strength, high temperature composite material such as graphite fiber with negligible magnetic susceptibility. (See, "New World for Aerospace Composites," *Aerospace America*, Oct. 1985, pp. 36–42, by W. F. De Mario.) This protective high-speed airfoil **46** gives the ramjet propulsion system a fairly high lift to drag ratio which enables it to glide and maneuver a considerable distance through the lower atmosphere without any propulsive forces.

As is shown in FIGS. **10**, **11**, and **12**, the airfoil **46** is designed with a small transverse cross section with a chord to thickness ratio of about 8 to 1. This design enables the total structural mass of the airfoil **46**, including all of its internal components except the dipole coil **10**, to have a total mass approximately equal to 25% of the coil mass. Thick blankets of evacuated multilayer cryogenic thermal insulation **62** are mounted along the inside walls **64** of the airfoil **46**. Essentially all of the structure surrounding the coil **10** is constructed with material having very low magnetic susceptibility. Small superconducting magnetic shielding coils **66** are provided to shield various electron command and control systems from the magnetic field of the dipole **10**.

The circular airfoil **46**, all of its internal structure and operating systems, and the superconducting propulsion dipole **10** (along with its cryogenic Dewar system) comprise a single-coil electromagnetic ramjet. It is a self-contained, automated reusable launch vehicle (i.e., accelerator) which is able to maneuver in space and traverse through the atmosphere in controlled flight and land vertically back at the launch site. For simplicity, FIGS. **1**, **2**, **3**, **10**, **11**, and **12** illustrating the design and construction of the electromagnetic ramjet do not show the payload in order to keep the discussion focused on its operating principles.

Although the construction of a CW beamed power transmitter capable of generating a 300 GW laser beam at optical frequencies is well beyond the state of the art, it can be easily achieved at microwave frequencies. FIGS. **13** and **14** illustrate the design and construction of a giant phased array microwave transmitter **20**. As is shown in FIG. **13**, the transmitter **20** is rigidly mounted in a housing **68** with multiple radiating elements **70** having a broadside beam elevation angle of 30°. The beamed power electromagnetic ramjet propulsion system envisioned herein will require a plurality of microwave power transmitters. The first transmitter, located closest to the launch point, will have a broadside beam elevation angle of **45°** . All other transmitters will be designed with a broadside beam elevation angle of 30°. In the preferred embodiment, the array **72** is circular (FIG. **14**) with a diameter of 600 m. Each radiating element **70** (which comprises a beam launching horn) is fed CW microwave power via a waveguide **74**. As is shown in FIG. **13** and in FIG. **15** (which is a plan vie of the transmitter) the waveguides **74** extend outward from the back of the array **72** in diverging directions for several hundred meters so as to provide some space between adjacent waveguide **74** that is large enough for access by technicians. These are where the phase shifters **76** are mounted. The microwave beam **18** is steered electronically by these phase shifters **76** which control the phase of the microwave power that is fed to the beam launching horns **70**. FIG. **16** shows an enlarged longitudinal cross section of some of the beam launching horns **70** and the waveguides **74** that are connected to them. It is possible to design this microwave feed system with low loss waveguides such that over 99.9% of the microwave power fed into the system is transmitted.

In the preferred embodiment, the transmitting horns **70** and waveguides **74** have square transverse cross sections. The transmitting end of the horns **70** have aperture dimensions 4λ by 4λ where λ denotes the wavelength of the microwave beam. For the preferred frequency of 10 GHz, $\lambda=3$ cm. Hence, the end of each transmitting horn **70** has inside dimensions of 12 cm by 12 cm. The mounting of the array **72** is such that the spacing **78** between the inner walls of adjacent horns **70** is $\lambda/2=1.5$ cm so as to form a close-packed array with negligible sidelobes. This close-packing design of the array will enable essentially 100% of the radiated power to be contained in the main lobe which forms the microwave beam. With these dimensions, the array **72** will be composed of 15,514,038 separate beam launching horns **70** and a like number of waveguides **74** and phase shifters **76**. Since the technology of microwave power transmission enables relatively small waveguides to convey very high levels of microwave power (because the power density can be as high as 1.2 MW/cm^2) over long distances with virtually zero loss, it is possible to achieve a total beam power of essentially any desired value by simply feeding a sufficient amount of microwave power into each individual waveguide. This is an important operating feature of the invention as it enables the electromagnetic ramjet to be designed to operate with essentially any desired level of microwave power. In the preferred embodiment of the invention, the microwave transmitters will be designed to transmit a maximum beam power of 10,000 ^{GW} (10^{13} watts). In order to obtain a maximum beam power of 10,000 GW, it is only necessary to feed 645 KW of microwave power into each individual waveguide **74**. Since 645 KW microwave generators operating at 10 GHz are readily available as commercial "off the shelf items", the possibility of achieving a beam power of 10,000 GW (and much greater power) is very easy. It is just a matter of numbers. For example, since 1 MW amplitron microwave generators operating at 10 GHz (with an overall electric-to-microwave efficiency of 95%) are readily available and could be used to feed each waveguide **74**, a beam power of over 15,500 GW could be achieved. (See, "High Power Microwave Generators of the Crossed-Field Type," *Journal of Microwave Power,* Vol. 5, No. 4, December 1970, pp. 245-259, by William Brown.) The reason why the waveguides **74** diverge and extend in back of the array for several hundred meters is to separate them to allow individual access and to enable the heat that is generated by the very small inefficiency losses to be carried off by various cooling systems.

The particular method adopted for generating the microwave power and feeding it into each waveguide will depend upon economics which is beyond the intended scope of this disclosure. However, it is instructive to identify a few of the possibilities. The first method involves connecting each waveguide **74** to its own separate microwave generator as described above. However, since the number of waveguides is so enormous, this method may be too expensive. Another method would be to generate several gigawatts of microwave power in single unit microwave generators and use the output to feed several thousand waveguides. In particular, a single microwave generator operating with an efficiency of 90% in the form of a 1.22 m long, 2.75 m diameter cylinder with a mass of only 5,400 kg would be capable of generating 10 GW of microwave power at 10 GHz. (See pages 176–180 of the paper, "Microwave Rocket Concept,"*International Astronautical Congress* Vol. 16, Athens, 1965 by J. Schad and J. Moriarity.) It would only require 1,000 of these units (each of which feeding 15,504 waveguides) to obtain the desired maximum 10,000 GW beam power. For definiteness, this is the method adopted herein. Thus, as shown in FIGS. **13** and **15**, the 10 GW microwave generating units **80** are positioned at various locations, each of which are connected to 15,504 waveguides **74**.

In order to dissipate the waste heat, a large cooling system **82** is provided. This cooling system **82** takes large quantities of water **84** from a nearby reservoir **86** and circulates it via conduits **88** around the microwave generators **80**, phase shifters **76**, and large portions of the waveguides **74** and beam launching horns **70**. Since the transmitter will have to operate for only few minutes, the heat build up will not overwhelm the cooling system.

The electrical energy used to power the 1,000 microwave generators **80** is obtained from a large underground superconducting energy storage system **90**. Such energy storage systems will be easily capable of supplying over 10,000 GW of DC electric current for the microwave generators **80** since superconducting inductor coils can be discharged at very high power levels. The electric current is fed to the generators **80** via superconducting power transmission lines **92** that are connected to the inductor **90** via superconducting on/off control switches **93**. A large cryogenic cooling system **94** is provided for the inductor **90** and for the various superconducting transmission lines **92** and switches **93**. Since an electromagnetic ramjet receiving propulsive power from a particular transmitter will usually be within range of the transmitter for less than 2 minutes, the total amount of inductive energy stored in the superconducting energy storage system could be easily achieved. It would require a storage capacity of about 350 GW-hr to provide 10,000 GW for 2 minutes. Such systems are well within engineering feasibility. (See for example, "Superconductive Energy Storage For Diurnal Use By Electric Utilities," *IEEE Transactions On Magnetics,* Vol. MAG-17, No. 1, January 1981, pp. 340–343 by R. Boom.) After the superconducting energy storage system is discharged by feeding current to the microwave generators **80** at very high power levels, it is recharged with current via other superconducting power transmission lines **96** at a much lower power level, but over a much longer time period. These recharging transmission lines could be connected to the field coils that are used to launch the ramjet vertically off the earth's surface. However, since some of the transmitters will be located several thousand kilometers from the launch point, it may be more practical to recharge the superconducting energy storage systems **90** by specially designed "stand-alone" power plants. A large superconducting MHD power plant would be ideal for this purpose because this type of power plant is very efficient and could be easily designed to operate on an intermittent basis with the capability of generating very large amounts of power. (See *Investigation of MHD Power Generation, Vol. II Integration With Cross-Field Microwave Devices,* RADC-TDR-62-464, Contract AF 30 (602)-2487, by W. C. Brown.) For example, such a power plant could be designed to generate 100 GW of power for 3½ hours which would be sufficient to recharge a completely discharged superconducting energy storage system **90**. However, since the first transmitter will be located near the field coils, the electric current it uses to power its microwave generators will be taken directly from the superconducting field coils.

In this case, the inductive energy storage system for the transmitter would be the field coils themselves and thus there would be no need for a separate system **90**.

There is another method for operating the microwave power generation and transmission system that may be the most economical. This method is based upon taking maximum advantage of the enormous microwave power that can be efficiently transmitted through circular large-diameter pressurized gas high frequency waveguides. For example, if d (meters) denotes the diameter of such a waveguide, the microwave power P (gigawatts) that could be transmitted through it is given approximately by

$$P=(5d)^2\ GW$$

(See, "Economic Feasibility of Microwave Power Transmission in Circular Waveguide," Ch. 3.6, pp. 256–269, D. Dunn and W. Loewenstern, Jr., *Microwave Power Engineering*, Vol. 1, Academic Press, 1968, E. Okress, ed.) Consequently, a waveguide of this type having a diameter of only 14.14 m (46.39 ft) would be capable of transmitting 5,000 GW. Two such waveguides would be capable of transmitting the maximum desired microwave power. Thus, instead of having to generate 10,000 GW of microwave power at each transmitter, it could be generated at the first transmitter (via the generators **80**) and fed to the other transmitters sequentially by a pair of these giant waveguides mounted in underground tunnels that are connected to all of the transmitters located within the continental United States. The process of turning off the microwave power at one transmitter and turning it on at the next transmitter would simply involve switching the power flowing through the waveguides from one transmitter to the next transmitter. Consequently, with this method, there would be no need for any microwave generators **80** or superconducting energy storage systems **90** at these transmitters. The microwave power would be fed directly into the small waveguides **74** that feed it past the phase shifters **76** to the beam launching horns **70** at each phased array transmitter.

Although the phased array microwave transmitter may appear to be a very complex structure and operating system, it is, in reality, a simple bee-hive type housing with millions of waveguides. Since there are no moving parts, there is nothing to ware out. It would not require very much maintenance and it would be essentially 100% reliable. The economics of automated mass production techniques could be employed to reduce the total cost of the array.

As pointed out above, the microwave beam **18** is steered electronically by controlling the phase of each radiating element **70** in the array by the phase shifters **76**. Each of these phase shifters **76** are controlled by a beam steering computer **98** via electric conduits **100**. This computer **98** generates the steering commands such that the beam is automatically pointed at the electromagnetic ramjet as it traverses through the ionosphere several hundred kilometers away. The phase shifters **76** also enable the beam to be focused on the dipole such that the beam diameter is always maintained exactly equal to the dipole diameter as it moves away from the transmitter. An excellent technical exposition on the design and operating principals of electronically steered phased array microwave transmitters is given in the book: *Theory And Analysis Of Phased Array Antennas*, John Wiley & Sons, Inc., 1972 by N. Amitay et al.

The focused minimum possible diameter d_s of the microwave beam at a distance D is given by the formula

$$d_s=2.44\ D\ (\lambda/d_t) \tag{5}$$

where d_t is equal to the diameter of the transmitting antenna, which is, in the preferred embodiment, equal to 600 m. Since $\lambda=0.03$ m, the minimum possible beam diameter d_s at a distance $D=2{,}459$ km (2.5×10^6 m) is 300 m. Consequently, the effective operational range of this transmitter when beaming propulsive power to an electromagnetic ramjet with a diameter of 300 m is 2,459 km. FIG. is a graph of d_s versus D corresponding to this transmitter where $\lambda=0.03$ m, an $dt=600$ m. Since the maximum required distance will only be about 500 km (which corresponds to a ramjet altitude of 100 km, and a beam elevation angle of 10°) this 2,459 km range is much more than is actually required. The minimum beam diameter d_s for this transmitter at the maximum 500 km range is 61 m. Consequently, in order to operate this transmitter at maximum power transmission efficiency (where the beam diameter is always equal to the smallest dipole diameter) the dipole diameter must be equal to or greater than 61 m. In order not to exceed the maximum power density of 425 W/cm^2 (so as to avoid atmospheric breakdown) the maximum beam power that could be transmitted to an electromagnetic ramjet with this diameter is 12.42 GW (which is only 0.12% of the transmitter's maximum power generating capability of 10,000 GW).

As pointed above, the actual design of the electromagnetic ramjet will comprise not one but several superconducting dipole coils. This enables the effective magnetic field of the central region (i.e., the electron cyclotron resonance region) to be increased by the superposition of the magnetic fields generated by all the dipoles, thereby significantly increasing the propulsive thrust. The diameter of the microwave beam is continuously focused so as to be equal to that of the inner most dipole coil. FIGS. **18** and **19** are schematic longitudinal and transverse cross sections of a typical electromagnetic ramjet **102** with six concentric coplanar superconducting dipole coils **10** mounted inside circular concentric airfoils **104**. (These airfoils **104** are essentially identical to the airfoil **46** described in FIGS. **10**, **11** and **12**.) The concentric circular dipole airfoils **104** are maintained in rigid, spaced apart positions relative to each other by a secondary system of straight radial airfoils **106** that are mounted between adjacent dipole airfoils **104**.

A cylindrical central body **108**, (i.e., fuselage), which represents the ramjet's payload, is mounted along the central longitudinal axis of the dipole via a plurality of cables **110** made of very high strength, high temperature material (such as fused silica glass fibers). The central body **108** has a length less than the diameter of the inner airfoil **115** and contains several large in dependent cargo bays for the payload. It is mounted such that its center of mass **111** coincides with the center of mass of the ramjet **102** (which is at the geometrical center of the dipole coils). As is shown in FIG. **19**, the transverse cross sectional area **112** of the central body **108** is a small fraction of the area **114** of the inner airfoil **115**. Therefore, its presence inside the center of the electromagnetic ramjet does not restrict the air flowing through it and does not reduce its thrust generating capabilities. The electron cyclotron resonance region **116** of the ramjet is designed to be behind the end **118** of the central body **108** such that the central body **108** is outside the accelerating region **120** of the ionized atmospheric gas passing through the ramjet and does not interfere with it. Unlike prior art ramjet propulsion systems that are mounted inside a fuselage or attached to wings (such as the "scramjets" proposed for the ground-to-orbit aerospace plane), the electromagnetic ramjet proposed herein is so huge that the entire fuselage of the vehicle is mounted inside the ramjet and the "wings" become the external structure of the ramjet itself.

FIG. **20** is an enlarged schematic longitudinal view of the central body **108** illustrating a plurality of large aerodynamic dive brakes **122** that are used to assist in decelerating the ramjet as it traverses back through the atmosphere while returning to the launch site on the earth's surface. The rear of the central body **108** is fitted with large hemispherical doors **124** made of high temperature material that provides thermal protection during the high-speed descent through the atmosphere back to the launch site. The front **126** of the central body **108** has a streamlined shape to reduce aerodynamic drag while accelerating to orbit through the atmosphere. Large portions of the inside walls of the central body **108** are fitted with superconducting shielding coils **127** to prevent the magnetic field generated by the ramjet's dipole coils from entering the central body.

FIG. **21** is a schematic transverse cross section of the rear of the central body directly behind the hemispherical thermal doors **124** illustrating the exhaust duct **128** of a very large multigigawatt electron cyclotron resonance (ECR) accelerator **130**, surrounded by the exhaust nozzles **131** of twelve high thrust chemical rocket engines **133** using propellant **135** stored inside the central body **108**. These engines **130**, **133** provide the vehicle with a considerable amount of orbiting maneuvering propulsion that is used after the primary ramjet propulsion system accelerates the vehicle to a desired launch velocity. For example, this auxiliary propulsion system can be used for plane changing maneuvers and for circularizing the vehicle in a desired circular orbit. Since the electromagnetic ramjet propulsion system is capable of accelerating the vehicle through the ionosphere to lunar transfer velocities, this self-contained propulsion system enables the vehicle to decelerate into lunar orbit, unload cargo, and accelerate back to earth. By constructing a plurality of superconducting field coils on the moon's surface, the ramjet (and its payload) will be able to land on its surface and be catapulted back to earth via magnetic repulsive forces, thereby providing direct surface-to-surface transportation between the earth and the moon for passengers and/or cargo. But before this magnetic propulsion system could be constructed on the moon, it will be necessary to perform lunar propulsive maneuvers using the conventional reaction principle of rocket propulsion.

The electrical energy used to operate the ECR engine will be obtained from the ramjet's superconducting dipole coils **109** which are utilized as a large inductive energy storage system. Since these dipole coils **109** will have to be nearly completely discharged before the ramjet can be returned back to the launch site in order to be properly decelerated by the magnetic repulsive forces of the field coils, the utilization of this stored inductive energy for maneuvering propulsion is very convenient. (Since the total mass of the returning vehicle will be significantly lower than the launch mass, it is necessary to reduce the magnetic field of the returning ramjet so as to reduce the magnetic repulsive forces generated by the field coils and allow the ramjet to land.) For a detailed technical description of the ECR engine see my paper "Generalized Theory of Rocket Propulsion for Future Space Travel," AIAA *Journal of Propulsion and Power*, Vol. 3, No. 4, July-August 1987, pp. 320–328.

When the electromagnetic ramjet is being accelerated by the microwave beam, the intensity of the repulsive magnetic field generated by the free electrons moving in their resonant cyclotron orbits behind the ramjet will be several orders of magnitude greater than the intensity of the earth's natural magnetic field. Consequently, during this time, the earth's natural magnetic field will have no effect on the ramjet. However, when the microwave propulsion is terminated, the earth's natural magnetic field will exert a magnetic torque on the dipole coils that will force the dipole plane of the ramjet to remain oriented perpendicular to the earth's magnetic field. Therefore, the central body **108** is mounted on the ramjet **102** such that after the microwave propulsion is terminated and the vehicle is in orbit, the central body **108** will be free to rotate 360° around a rotation axis **137** that passes through its center, perpendicular to its longitudinal axis **103** which extends along the diameter of the circular airfoils **104**. This central rotation axis **137** is a tubular beam **137** mounted inside co-linear radial airfoils **139** that are attached to the circular airfoils **104**. A plurality of co-planer supporting cables **141** (FIG. **18**) extending from the central body **108** are attached to the rotation axis **137** thereby providing additional support for the central body **108** while the central body **108** rotates around the rotation axis **137**. A plurality of superconducting cables **143** are mounted inside the central beam **137** and enable electric current taken from the dipole coils **109** to pass into the central body **108** and fed into the ECR engine **130**.

When the ramjet is being accelerated by the field coils, and then by the microwave beams, the thrust forces generated by the dipole coils **109** will be very great. In order to transmit these thrust forces to the central body **10** without causing any significant coil deformations, a large number of supporting cables **145** are mounted at close intervals around each circular airfoil **104** (FIG. **19**) that are attached to various bulkheads **147** (FIG. **18**) on the central body **108**. After the microwave propulsion is terminated, those supporting cables **149** that are not attached to the rotation axis **137** are detached from the airfoils **104** and withdrawn into the central body **108** thereby allowing the central body to rotate freely around the rotation axis **137**. The reflector grid **12** is also disconnected from the inner airfoil **115** and withdrawn into the central body **108**. Although the dipole plane must always remain perpendicular to the earth's magnetic field, the longitudinal axis **103** of the central body **108** can be pointed in any direction by rolling the vehicle around the longitudinal axis **103** and by rotating the central body **108** around the rotation axis **137**. Consequently, the propulsion system of the central body **108** can generate propulsive thrust in any desired direction while the coil plane remains perpendicular to the earth's magnetic field and such that the thrust vector always passes through the vehicle's center of mass. A large attitude control moment gyro system **151** is mounted at the center of mass **111** inside the central body **108** (FIG. **20**) in order to carry out these rolling and rotating movements.

When the vehicle is preparing to return to the earth's surface prior to reentering the atmosphere, the central body **108** is rotated around the rotation axis **137** such that the longitudinal central axis **103** of the central body **108** is aligned with the longitudinal central axis of the ramjet **102**. The alignment is such that the hemispherical doors **124** on the rear of the central body **108** are heading into the direction of motion. Since the dipole coils **109** are nearly discharged (which will be required when returning to earth) the magnetic torque exerted by the earth's natural magnetic field will be relatively weak and could be cancelled by the attitude control system. However, when the ramjet approaches to within a few hundred kilometers of the launch site, the magnetic field generated by the field coils will become stronger than the earth's natural magnetic field. This will create a new and stronger magnetic torque that will keep the dipole plane perpendicular to the magnet field of the field coils and this is precisely what is required during the final descent to the landing area.

As described above, the electromagnetic ramjet, along with its payload, is launched vertically from the earth's surface and accelerated to supersonic speed at high altitude by magnetic repulsive forces generated by a plurality of coaxial superconducting field coils several kilometers in diameter embedded beneath the earth's surface. Since the superconducting dipole coils of the electromagnetic ramjet generate such a huge magnetic field (when fully charged), this field is utilized to initially lift the vehicle off the earth's surface and to accelerate it to an altitude where the ramjet can begin operating by the microwave beam. Thus, after the vehicle reaches a sufficiently high altitude, the electromagnetic ramjet propulsion system is started using microwave power transmitted from the earth's surface which accelerates the vehicle to orbital velocity. After the payload is removed from the central body, the vehicle is decelerated back to the launch site by continuous aerodynamic braking while gliding through the upper atmosphere at high altitudes over extended distances gradually descending to lower altitudes and landing vertically at the launch site with final decelerating thrust provided by magnetic repulsive forces generated by the same field coils used to launch the vehicle. These launching and landing methods are illustrated in FIGS. **22** and **23** respectively. For simplicity, the ramjet appears as one dipole propulsion coil in these figures and the central body is omitted.

As is illustrated in FIG. **22**, the dipole coil **10** of an electromagnetic ramjet accelerated vertically from an annular launching platform **132** by magnetic repulsive forces generated by a plurality of superconducting field coils **134** (which are also dipoles) mounted inside circular coaxial underground tunnels **136** in a remote region. These field coils **134** generate a resultant magnetic field **138** (represented by the vector summation or "superposition" of the individual magnetic fields) which opposes the magnetic field **140** generated by the propulsion dipole **10** which has opposite polarity. This results in a magnetic repulsive force exerted on the propulsion coil **10**, and magnetic repulsive forces exerted on the field coils **134**. Multiple arrays **142** of photovoltaic solar cells are mounted around the outer perimeter of the field coils for generating electrical energy. This electrical energy is used to charge up the superconducting field coils. As described above, the field coils are also utilized as a giant superconducting inductive energy storage system for supplying current to the microwave generators at the first transmitter. A plurality of underground superconducting power transmission lines **144** are provided for transferring excess electrical energy not needed to operate the propulsion system to the national intertie power grid for distribution across the United States.

FIG. **23** is a schematic perspective view showing how a superconducting propulsion coil **10** is decelerated back to the launch site by the stationary field coils **134** embedded beneath the earth's surface **146**. The inductive coupling between the field coils **134** and the magnetically decelerated propulsion coil **10** enable a portion of the potential and kinetic energy of the coil **10** to be reconverted into electrical energy when the coil **10** is brought to rest back on the launching platform **132**. This inductive coupling appears as an increasing repulsive magnetic field generated by the dipole coil **10** when approaching the field coils that generate strong magnetic repulsive forces decelerating the coil and enabling it to lose all of its velocity at the moment of contact with the earth's surface in the center of the field coils. This is a beautiful situation as it allows the propulsion dipole **10** to be automatically partially recharged with current that is used to launch the next payload. However, a great deal of additional current must be fed into the dipole coil because the launch mass will be many times greater than the return mass. But the cost of this additional inductive energy that must be added to the propulsion coil will be zero since the solar arrays are conducted to be an integral part of the entire system.

Payloads are brought to the launch site via an underground access tunnel **148**. The surrounding walls of this tunnel **148** are fitted with superconducting shielding coils **150** to prevent the magnetic field **138** generated by the field coils **134** from entering the access tunnel **148**.

In order to demonstrate the basic engineering feasibility of lifting the electromagnetic ramjet off the earth's surface (along with its payload) and accelerating it to very high altitude via magnetic repulsive forces generated by a plurality of superconducting field coils embedded beneath the earth's surface, it is necessary to conduct a mathematical investigation. FIG. **24** illustrates two parallel coaxial circular coils **152**, **154** with radii R_1 and R_2 separated by a distance h and carrying currents i_1 and i_2 in opposite directions respectively. Hence, the corresponding magnetic fields **156**, **158** generated by the currents i_1 and i_2 have opposite directions, and the coils are therefore repelled by a magnetic repulsive force F. The exact mathematical expression for this repulsive force is given by

$$F = \frac{\mu_o i_1 i_2 h}{\sqrt{h^2 + (R_1 + R_2)^2}} \left[\left(\frac{{R_1}^2 + {R_2}^2 + h^2}{(R_1 - R_2)^2 + h^2} \right) E_2 - E_1 \right] \tag{6}$$

where E_1 and E_2 and the complete elliptic integrals of the first and second kinds, respectively given by

$$E_1 = \int_o^{\pi/2} \frac{dx}{\sqrt{1 - k^2 \sin^2 x}}$$

and

$$E_2 = \int_o^{\pi/2} \sqrt{1 - k^2 \sin^2 x}\, dx$$

where

-continued

$$k^2 = \frac{4R_1R_2}{h^2 + (R_1 + R_2)^2}$$

If $R_1 >> R_2$, equation (6) can be expressed to a good approximation by the equation

$$F = \frac{3\pi\mu_o R_1^2 R_2^2 i_1 i_2 h}{2(R_1^2 + h^2)^{2.5}} \tag{7}$$

where $\mu_o = 4\pi \times 10^{-7}$ henry/m (MKS units are used). See, *Introduction to Electromagnetic Fields And Waves*, Addison-Wesley Publishing Company, 1968, pp. 219–220, by E. V. Bohn. It is important to point out the fact that equation (7) demonstrates that the propulsive force F acting on the second coil **154** increases with the square of its radius R_2.

Suppose that the first coil **152** (field coil) is at rest on the earth's surface and the second coil **154** (propulsion coil) is being repelled above it by the magnetic repulsive force F. The downward force of gravity W acting on the second coil **154** is equal to m_2g where m_2 is equal to its mass and $g = 9.81$ m/sec^2 is equal to the average gravitational acceleration at the earth's surface. If the second coil **154** has a mass density ρ and a current density J, the thrust to weight ratio F/W for the magnetically repelled coil **154** can be expressed as

$$\frac{F}{W} = \frac{3\mu_o R_1^2 i_1 R_2 J h}{4g\rho(R_1^2 + h^2)^{2.5}} \tag{8}$$

It is apparent from this equation that the thrust to weight ratio of the magnetically repelled coil **154** can be increased by increasing its radius R_2 and its current density J. If the coils **152**, **154** are superconducting, the values of i_1 and J can be made very large. Moreover, by utilizing superconducting coils, all internal electrical resistance vanishes, and the currents i_1 and i_2 do not have to be sustained by any external power source to make up for resistive losses.

There exists the practical problem of charging the second coil **154** with current i_2 without it being catapulted away from the first coil **152** (which is assumed to be fully charged with current i_1) before the second coil **154** can be fully charged. However, in view of equation (8), there is no upward thrust exerted on the second coil when $h=0$. In order for the coil **154** to be magnetically repelled away from the field coil **152**, it must have some relatively small initial altitude h_o where $F/W=1$. Thus, the second coil **154** can be charged with its current i_2 while resting on the earth's surface (at $h=0$) in the center of the larger field coil **152**.

Let B_1 denote the magnetic field strength generated by the first coil **152** at its center. Hence

$$B_1 = \frac{\mu_o i_1}{2R_1} \tag{9}$$

If A_2 denotes the area of the second coil **154**, its magnetic dipole moment $M_2 = i_2A_2$. Therefore, since $A_2 = \pi R_2^2$, the magnetic dipole moment M_2 of the second coil **154** can be expressed as

$$M_2 = i_2\pi R_2^2 \tag{10}$$

If the second coil **154** is charged with current i_2 when it is on the earth's surface in the center of the field coil **152**, its magnetic potential energy E_m, which is equal to B_1M_2, can be expressed as

$$E_m = \frac{\mu_o\pi i_1 i_2 R_2^2}{2R_1} \tag{11}$$

This magnetic potential energy E_m is converted into kinetic and potential energy by the second coil when it accelerates away from the first coil (which remains stationary). The mathematical expression of this energy represented by equation (11) clearly demonstrates that an enormous amount of magnetic potential energy E_m can be fed into the second coil (by charging it with current i_2) if the coil is constructed with a very large radius R_2. This method of launching the ramjet by magnetic repulsive forces involves the recognition (through equations 7, 8 and 11) that enormous amounts to stored propulsive energy E_m and propulsive thrust F can be achieved with the principle of magnetic repulsion by utilizing superconducting coils and scaling up their physical dimensions many orders of magnitude—far beyond anything previously contemplated. By constructing a superconducting field coil **152** with a radius of several kilometers and charging it with an enormous current i_1, it is possible to construct the superconducting propulsion coil **154** with a radius of hundreds of meters and charge it with current i_2 to obtain an initial magnetic launch energy E_m far greater than the initial chemical launch energy contained in the propellant of any prior art chemically propelled launch vehicle. In fact, equation (11) shows that, for all practical purposes, there is virtually no limit on the amount of launch energy E_m than can be obtained for the magnetically propelled coil—and this launch energy E_m will be converted into kinetic and potential energy with an efficiency of nearly 100% compared to only about 2% for chemically propelled launch vehicles. (Almost all of the initial chemical launch energy of conventional rocket propelled launch vehicles is wasted by having to lift and accelerate the enormous propelled load.)

In order to better understand this magnetic propulsion concept, it may be helpful to consider a simple numerical example. Suppose the field coil **152** has a current $i_1 = 1.5 \times 10^9$ amp and a radius $R_1 = 10{,}000$ m. Suppose the current density of the second coil **154** is $J = 5 \times 10^8$ amp/m^2 and has a radius $R_2 = 1{,}000$ m with a mass density $\rho = 3{,}000$ kg/m^3. At an altitude $h = 10{,}000$ m (32,800 ft) the thrust to weight ratio for the magnetically repelled coil **154** would be equal to 42.46. Thus, at this altitude, the coil **154** would be accelerating upward by the magnetic repulsive force with an acceleration of 41.46 g and would be moving with a velocity of 3.156 km/sec (7,059 MPH or Mach 10,2).

The corresponding velocity V at any altitude h can be calculated from the equation

$$V = \{C[(R_1^2 + h_o^2)^{-1.5} - (R_1^2 + h^2)^{-1.5}] - 2G[(r_e + h_o)^{-1} - (R_e + h)^{-1}\}^{\frac{1}{2}} \tag{12}$$

where

$$C = \frac{\mu_o R_1^2 i_1 R_2 J}{2\rho}$$

and G=3.981×10^{14} m^3/sec^2 (earth's gravitational constant) R_e=6,371,315 m (earth's mean radius) and h_o is equal to the launch altitude (which is equal to the initial height where F/W=1). For the above example h_o=41.16 m and C=1.571×10^{19} m^5/sec^2. The coil would be accelerated to a maximum altitude of 916 km. The altitude where the magnetic repulsive force is equal to the gravitational force (F/W=1) is 37.73 km (123,800 ft). This altitude could be sustained indefinitely by the propulsion coil without expending any energy.

This simple example demonstrates the very powerful propulsive forces than can be generated by magnetic repulsion when large superconducting coils are used. These propulsive forces are significantly greater than that which can be generated by conventional rocket engines. The most important fact however, that has profound importance in the field of space travel, is that these powerful propulsive forces are generated without using any propellant or any power generating system on the propulsion coil.

Before proceeding, it should be pointed out that although the construction of a 20 km diameter superconducting field coil would be expensive, it would cost significantly less than the $3 billion cost of the Superconducting Super Collider underground particle accelerator, with a diameter of 60 km that is being designed and constructed for high energy physics research. (See "The SSC: A Machine For The Nineties," *Physics Today,* March 1985, pp. 28-37, by S. Gladshow and L. M. Lederman). Using mass production techniques for manufacturing the superconducter, the cost of installing a 20 km diameter field coil could probably be kept to about $10 million per kilometer or $625 million. Thus, the concept of magnetic propulsion introduced herein is well within economic feasibility.

It follows from equation (8) that the thrust to weight ratio F/W of the propulsion dipole **154** can be increased by maximizing the product R_2J. But the radius R_2 and current density J are not independent variables. They are related by the mechanical stress limitations generated by magnetic forces. Any current carrying conductor generates a magnetic field. This magnetic field generates $\vec{J}\times\vec{B}$ Lorentz forces on the conductor which, in turn, generates various mechanical stresses that must be contained by some supporting mass. In the case of superconducting propulsion coils described herein, the conductor itself is designed to contain these forces. Thus, in order to minimize the mass of a propulsion coil it will be designed to be a self supporting superconducting dipole. The stresses are contained by the superconducting cable used to construct the dipole.

The inductive energy E of a superconducting coil carrying a current i with self inductance L is given by the equation

$$E=\tfrac{1}{2}Li^2 \tag{13}$$

The self inductance of a dipole coil with major radiums R and minor radius r (FIG. 25) is given by

$$L=\mu_oR[\log(8R/r)-1.75] \tag{14}$$

Consequently, the total radial outward force F acting on the dipole is

$$F=\frac{\partial E}{\partial R}=\frac{1}{2}i^2\frac{\partial L}{\partial R}=\frac{1}{2}i^2\mu_o[\log(8R/r)-.75]$$

where i denotes the total current flowing through the coil's total cross sectional area A_t. If F_t denotes the total tension in the coil, its stress $\sigma=F_t/A_t=F_t/\pi r^2$. Consequently, since $F_t=F/2\pi$, it follows that

$$F_t=\frac{F}{2\pi}=\frac{i^2}{4\pi}\mu_o[\log(8R/r)-.75]=\pi r^2\sigma \tag{15}$$

The magnetic filed B_o at the center of the dipole is given by

$$B_o=\frac{\mu_oi}{2R} \tag{16}$$

and the maximum magnetic filed B_m which occurs on the surface of the dipole closest to the center, is given by

$$B_m=\frac{\mu_oi}{2\pi r} \tag{17}$$

Consequently, since the aspect ratio γ=R/r, equation (15) can be expressed as

$$B_o{}^2\gamma^2[\log(8\gamma)-0.75]=\sigma\pi^2\mu_2 \tag{18}$$

Since the current i is related to the current density J by $i=JA_t=J\pi r^2$, it follows from equation (16) that

$$JR=\frac{2B_o\gamma^2}{\pi\mu_o} \tag{19}$$

Hence, in view of equation (18) it follows that

$$JR=2(\sigma/\mu_o)^{\frac{1}{2}}[\gamma/(\log(8\gamma)-0.75)^{\frac{1}{2}}] \tag{20}$$

This equation is of fundamental importance because it relates the product JR to the tensile strength σ of the superconducting cable and the aspect ratio γ of the propulsion dipole.

The function F(γ) in the brackets of equation (20) defined by

$$F(\gamma)=\gamma/[\log(8\gamma)-0.75]^{\frac{1}{2}} \tag{21}$$

is dependent only on γ and hence the shape of the dipole (i.e., only on the relative thickness of its cross section). It will therefore be called the "shape factor". A graph of this function is shown in FIG. **26**. It is very nearly equal to a straight line with a slope of 0.3. Hence, the product JR can be expressed to a good approximation by the equation

$$JR=.600\,\gamma\sqrt{\sigma/\mu_o} \tag{22}$$

This equation clearly demonstrates that the value of JR can be increased to large values by increasing the aspect ratio γ of the dipole and by constructing the superconducting cable of the dipole with a very high tensile strength σ.

The aspect ratio γ of a dipole coil can also be determined to a good approximation by the equation

$$\gamma = \frac{.3\pi\sqrt{\sigma\mu_o}}{B_o} \tag{23}$$

which follows from equation (18) using the approximately $F(\gamma)=0.3\gamma$. However, it should be emphasized that the exact value of γ corresponding to a central field B_o is obtained by solving equation (18).

In view of equations (8) and (22) the thrust to weight ratio F/W of the dipole can be maximized by constructing the superconducting cable with a high strength to density ratio σ/ρ. The fabrication of such a cable is disclosed in my U.S. Pat. No. 4,078,747 filed June 2, 1975 entitled "Orbiting Solar Power Station." Basically this fabrication method involves reinforcing high strength, low density, superconducting cable previously fabricated by vapor depositing a thin layer of superconducting material onto high strength carbon fibers, with a tensile strength of 0.28×10^{10} N/m^2, with super high strength fused silica glass fibers, with a tensil strength of 1.4×10^{10} N/m^2. However, by using new high temperature ceramic superconducting material, the superconducting layer can be vapor deposited directly onto the fused silica glass fibers thereby eliminating the need for any carbon fibers. The technique of vapor depositing superconducting material onto high strength fibers is discussion in the article "Superconducting Properties of Thin Film Niobium Carbonitrides on Carbon Fibers," *IEEE Transactions on Magnetics,* Vol. Mag-11, No. 2, March 1975, pp. 185–188 by G. E. Pilce et al. Since experiments involving high temperature ceramic superconducting material indicate that such material is intrinsically stable, there is no need for any stabilizer material such as copper or aluminum. The resulting tension bearing superconducting cable fabricated by this technique will have a tensil strength $\sigma=1.4\times10^{10}$ N/m^2, and a mass density $\rho=2{,}160$ kg/m^3. This is the cable that will be used to construct the superconducting propulsion dipole.

In view of equation (1), the magnetic field B_o at the center of the dipole coil **10** is determined by the electron cyclotron resonance frequency f of the microwave beam. Hence

$$B_o=2\pi f(m/e) \tag{24}$$

Therefore, since $f=10^{10}$ (10 GHz) it follows this equation that $B_o=0.3572$ T. Consequently, when this value is substituted into equation (18), with $\sigma=1.4\times10^{10}$ N/m^2, the resulting aspect ratio $\gamma=429{,}094$.

If the major radius R_2 of the propulsion dipole were 300 m, the minor radius r_2(see FIG. **25**) would be $R_2/\gamma=0.6991$ m. Since the density ρ of the coil is 2,160 kg/m^3, the total coil mass would be equal to 6,252,000 kg. This can be significantly reduced by employing multiple dipole coils. Thus, in the preferred embodiment of the electromagnetic ramjet, the magnetic field is generated by six dipole coils instead of one, and each coil generates an equal field strength at the center. Therefore, the magnetic field generated at the center of each coil will be equal to $B_o/6=0.0595$ T. When this value is substituted into equation (18) the resulting aspect ratio γ is 2,324.1, which is a factor of 5.416 times greater than that of a single coil. Therefore, in view of equations (8) and (22) the thrust to weight ratio F/W of each of the six dipole coils will be increased by a factor of 5.416 over that of the single dipole. If the radii of these six dipole coils were 300 m, 310 m, 320 m, 330 m, 340 m and 350 m, the total mass would be 1,629,300 kg which represents a reduction by a factor of 3.8 over the single dipole design. The minor radius r_2 of the inner dipole would be 300 m/2,324.1=0.129 m (12.9 cm or 5.1 in) and the minor radius of the outer coil would be 350 m/2,324.1=0.151 m.

The general formula for calculating the total coil mass m_c of an electromagnetic ramjet with n dipole coils with equal aspect ratio is

$$m_c = \frac{2\pi^2\rho}{\gamma^2}\sum_{j=1}^{n} R_{2}{}^{3}{}_{j} \tag{25}$$

The magnetic launch energy E_m of the dipoles of an electromagnetic ramjet can be arbitrarily increased to any desired value without increasing its size or current by simply increasing the current i_1 in the field coils and/or by simply increasing the number of field coils. Since the field coils remain stationary on earth, their mass is of no consequence. Moreover, they do not have to be self-supporting as in the case of the propulsion dipoles. Since they will be embedded in underground tunnels the stresses could be supported by the surrounding earth itself.

In the preferred embodiment of the invention, there will be many (coaxial) field coils instead of only one. In view of the principle of field superposition where the effective propulsive field is equal to the vector sum of the individual fields, the propulsive thrust and energy can be increased to achieve unlimited values by simply increasing the number of field coils—which can be easily accomplished since the field coils remain fixed in underground tunnels. In this case, the thrust to weight ratio given by equation (8) becomes

$$\frac{F}{W} = \left(\frac{3\mu_o R_2 J h}{4g\rho}\right)\sum_{j=1}^{n}\frac{i_{1j}R_{1j}^2}{(R_{1j}^2+h^2)^{2.5}} \tag{26}$$

where R_{1j} is equal to the radius of the j'th field coil and there are n field coils altogether. The current in each field coil is equal to i_{ij}. In view of equation (20) the product R_2J will be equal for each individual dipole coil of a ramjet since the aspect ratio γ is constant. Therefore, the thrust to weight ratio for each dipole will be equal regardless of their radii.

The thrust to weight ratio of a multiple coil electromagnetic ramjet with coil mass m_c propelled by multiple field coils is given by

$$\frac{F}{W} = \left(\frac{3\mu_o h M^2}{2gm_c}\right)\sum_{j=1}^{n}\frac{i_{1j}R_{1j}^2}{(R_{1j}^2+h^2)2.5} \tag{27}$$

where M_2 is equal to the total magnetic dipole moment of the ramjet.

The total magnetic dipole moment M_2 of a multiple coil electromagnetic ramjet is equal to the summation of the individual dipole moments M_{2j} of each coil. Hence

$$M_2 = \sum_{j=1}^{n} M_{2j} = \sum_{j=1}^{n} i_{2j}\pi R_{2j}^2$$

The current i_{2j} flowing through the j'th coil is equal to its density J_j multiplied by its transverse cross sectional area A_{tj} which is equal to πr_{2j}^2 where r_{2j} is equal to the coil's minor radius. Since $r_{2j}=R_{2j}/\gamma$, the current i_j can be expressed as

$$i_j = \frac{\pi(j_jR_{2j})R_{2j}}{\gamma^2}$$

However, in view of equation (20), the value of J_jR_{2j} is the same for each coil. Consequently, the total magnetic dipole moment of the ramjet can be expressed as

$$M_2 = \frac{2\pi^2}{\gamma}\left[\frac{\sigma}{\mu_o[\log(8\gamma) - .75]}\right]^{\frac{1}{2}} \sum_{j=1}^{n} R_{2j}^3 \tag{28}$$

The magnetic dipole moment M_2 of the ramjet with six propulsion coils given in the example above with radii R_{2j} equal to 300 m, 310 m, 320 m, 340 m, 350 m will be equal to 6.18179×10^{13} amp m^2. The magnetic dipole moment corresponding to the ramjet with a single coil with radius 300 m would be 4.82220×10^{13} amp m^2.

The velocity V of the ramjet at any altitude h can be calculated by

$$V = \left\{ C_o \sum_{j=1}^{n} i_{1j}R_{1j}^2 \left[(R_{1j}^2 + h_o^2)^{-1.5} - (R_{1j}^2 + h^2)^{-1.5}\right] - 2G[(R_e - h_o)^{-1} - (R_e + h)^{-1}] \right\}^{\frac{1}{2}}$$

where

$$C_o = \frac{\gamma}{\rho}\left[\frac{\sigma\mu_o}{\log(8\gamma) - .75}\right]^{\frac{1}{2}} \approx \frac{.3\gamma\,(\mu_o\sigma)^{\frac{1}{2}}}{\rho}$$

When using the formula for determining the ascent velocity of an electromagnetic ramjet carrying a payload, the density ρ has to be multiplied by the mass ratio m_tm_c where m_t the total vehicle mass and m_c is the total coil mass.

The total magnetic potential energy E_m of the propulsion dipole is equal to

$$M_2 \sum_{j=1}^{n} B_{1j}$$

where B_{1j} is equal to the magnetic field strength at the center of the j'th coil generated by the j'th coil, and where M_2 is equal to the total magnetic dipole moment of the ramjet.

Let B_1 denote the central magnetic field strength generated by the field coils at the launch altitude h_o where the thrust to weight ratio F/w=1. This altitude h_o represents the initial height above the plane of the field coils. Since the launch altitude above the earth's surface is assumed to be o, the field coils are installed in underground tunnels at a depth equal to h_o. Therefore, the strength of the magnetic field due to the field coils at the earth's surface is given by

$$B_1 = \tfrac{1}{2}\mu_o \sum_{j=1}^{n} \frac{i_{1j}R_{1j}^2}{(R_{1j}^2 + h_o^2)^{1.5}} \tag{29}$$

Consequently, the initial magnetic launch energy E_m of the electromagnetic ramjet (which is equal to M_2B_1) can be expressed as

$$E_m = \tfrac{1}{2}\mu_o M_2 S_1 \tag{30}$$

where

$$S_1 = \sum_{j=1}^{n} \frac{i_{1j}R_{1j}^2}{(R_{1j}^2 + h_o^2)^{1.5}} \tag{31}$$

The magnetic inductive energy E_c stored in a multiple coil electromagnetic ramjet is given by

$$E_c = \tfrac{1}{2} \sum_{j=1}^{n} L_j\, i_{2j}^2$$

where L_j denotes the self-inductance of the j'th dipole coil given by equation (14). Since $i_{2j}=\pi R_{2j}(JR_{2j})/\gamma^2$, it follows from equations (14) and (18) that the ramjet's inductive energy E_c can be expressed by

$$E_c = \frac{2\pi^2\sigma}{\gamma^2}\left[\frac{\log(8\gamma) - 1.75}{\log(8\gamma) - 0.75}\right] \sum_{j=1}^{n} R_{2j}^3 \tag{32}$$

As described above, this stored inductive energy will be used for maneuvering the vehicle outside the earth's atmosphere by the onboard ECR engine after the ramjet propulsion system accelerates the vehicle to the required launch velocity.

Since the superconducting field coils are stationary and embedded in underground tunnels, the mechanical stress generated by their magnetic fields can be contained by the surrounding structure and supported by the earth itself. Thus, by eliminating the stress limitations on the field coils, the field coils can be designed to carry significantly more current than the propulsion coils (i.e., the current density J of the field coils can be designed to be much greater than that of the propulsion coils). Although the actual current densities J_o of the current carrying superconductor will be about the same for propulsion coils and field coils, the conductor to non-conductor cross sectional area ratio of the field coils will be much greater than that of the propulsion coils.

In order to take advantage of the very favorable scaling laws that are inherent in the magnetic launching system disclosed herein, the preferred embodiment will be very large. In the preferred embodiment there will be ten superconducting field coils **160** (FIG. **27**) mounted coaxially in underground tunnels **162** with increasing radii R_{ij} (j=1,2, . . . , 10).

In order to obtain a more uniform acceleration for the initial ascent trajectory, these field coils **160** will be concentrated toward the outer coil with a gradually decreasing spacing between adjacent coils determined by the function $2\times10^7/R$. Consequently, the radii R_{ij} of the ten field coils will be equal to 10,000 m, 12,000 m, 13,667 m, 15,130 m, 16,452 m, 17,668 m, 18,800 m, 19,864 m, 20,871 m, and 21,829 m. These field coils will all be designed to carry a current i_{ij} equal to 1.5×10^9 amp. Since the field coils will cover a circular region 44 km (27 miles) in diameter, it will be constructed in some remote, and relatively flat area within the western part of continental United States. (One remote desert region in the north west corner of the state of Nevada would be ideal.) A relatively flat circular central region **164** is provided with a diameter of 10 km, that serves as the launching and landing region The strength of the magnetic field in this region will be 0.6 T (6,000 Gauss).

By employing large scale production facilities, the cost of manufacturing the field coils will be relatively low. For example, studies by John Gilman, at the Lawrence Berkeley Laboratory in Berkeley, Calif. have shown that new ceramic oxide high temperature superconducting material could be mass produced at a cost of only $20/ton if the rate of production were 500,000 ton per year. This cost could be reduced to $2/ton by producing 50,000,000 tons/year. See the article, "The Commercial Potential Of High-T_c Superconductors," *Physics Today*, March 1988, by J. J. Gilman. The cost of producing fused silica glass (which would represent most of the mass of the propulsion coils) could be manufactured at a much lower cost. The underground tunnels would have a maximum depth of about 500 m and could be rapidly excavated by employing ten high speed tunnel boring machines operating simultaneously. The required tunnel diameter would be about **12** or **13** m. The construction of the coils could proceed by automated winding machines that continuously travel around each tunnel with spools of superconducting cable that build up each coil to the required thickness.

Since the local magnetic field on the earth's surface in an annual region **166** surrounding the outer field coil **168** will be many times greater than the earth's natural magnetic field, it will be kept isolated from the more distant surrounding region **170** where the field is not as great. In order to make good use of this isolated region **166**, it will be used for generating electric energy via photovoltaic solar arrays. The presence of a magnetic field does not affect the operation of photovoltaic solar cells. Thus, in the preferred embodiment of the invention, a huge annular array **172** of solar cells, 18 km wide with an inner radius of 22 km, and an outer radius of 40 km, is constructed around the outer field coil **168** for generating bulk electric power. This array **172** will cover a total land area of about 3,500 km^2.

The solar array **172** can be constructed from long continuous sheets of thin film amorphous silicon cells and mounted in rigid, 10 m long by 3 m wide modules produced in huge automated factories to reduce unit cost. The raw materials needed for the production, namely silicon, is the second-most abundant element on earth. Hence, by increasing the scale of the production facilities, it should be possible to construct the array with a unit cost of about $30/m^2. (See "Photovoltaic Power," *Scientific American* April 1987, pp. 87-92, by Y. Hamakawa.) Thus, the total cost of the array will be about $100 billion with an average operating efficiency of 15% the total amount of electric power that could be generated by the system is 526 GW. Since the solar array can only generate electric power during the daylight hours, the average daily power output of the system will be about 100 GW. This represents about 30% of the entire electric generating capacity of the United States and is much more than required to operate the transportation system. Thus, the complete ground to orbit transportation system envisioned herein also represents a giant solar-electric power generating plant. Since this electric generating system does not burn any combustible or nuclear fuel, and has no moving parts, it would require very little maintenance and it could remain operating indefinitely. The operating cost would be nearly zero. A large portion of the excess power could be fed into the existing U.S. electric power intertie grid for distribution throughout the United States and Canada. Large portions could also be fed into Mexico and other countries via superconducting power transmission lines. This electric power generating capability of the system would therefore enable many thermal electric power plants to be dismantled thereby eliminating atmospheric and thermal pollution caused by these plants. It would also have a substantial beneficial effect on the total U.S, economy since it would not only reduce the national balance of payments, but reverse it on a colossal scale.

The initial construction cost of $100 billion for the array could be recovered by the commercial sale of the electric power generated therefrom. For example, if the average output of the array is 50 GW per year, and if the average commercial sale of the electrical energy is 10¢/KW-hr, the amount of income revenue that could be generated by the array over a one year time period would $44 billion. At this rate the entire cost of the array could be paid in less than 3 years time.

Since the solar array **172** can only generate electric power during the daylight hours, the field coils **160** will be utilized as a giant underground superconducting inductive energy storage system for accumulating part of the electrical energy generated during the daylight hours so that it can be released into the power distribution system during the non-daylight hours. Thus, the field coils serves as a giant inductive energy storage system and load leveling system that allows electric power to be fed into the distribution system at any desired rate, 24 hours per day. Since the total amount of inductive energy stored in the field coils is so great, the amount of additional energy put into the system (or withdrawn from the system) represented by the output of the array over five or six days operation is relatively small and will not cause any significant change in the magnetic field. A portion of the inductive energy of the field coils could also be used in national defense. In particular, it could be used for instantly energizing various directed energy weapon systems in a national emergency. This capability would be extremely valuable to our country's critically important SDI program since the engineering feasibility of such beam weapon systems depends largely upon the availability of a sufficiently large power source

FIG. **28** is a schematic transverse cross section of a superconducting field coil **160** mounted inside an underground tunnel **162** and supported by a surrounding structure **174**. This structure **174** is itself supported by the tunnel walls **176** and the surrounding earth **178**. A liquefied nitrogen cryogenic Dewar system **180** is provided for maintaining the coil **160** at liquefied nitrogen temperatures. The coil's minor radius r is designed to be 5.0 m. Hence, it follows from equation (17) that the maximum magnetic field B_m on the surface of the coil will be 60 T. The coil's current density $J=i_1/(\pi r^2)=1.9099\times10^7$ amp/m^2. (Since the actual current carrying superconducting material will represent about 25% of the superconductor will be about 7.6×10^7 amp/m^2). Current is fed into and withdrawn drawn from the field coils **160** by a plurality of superconducting power transmission mission lines **182** that are connected to the solar array **172**; to the microwave power generators of the first transmitter; and to the national intertie power grid.

After the solar array is constructed, the electrical energy generated therefrom would first be used to charge up the ten field coils. If L_j denotes the inductance of the j'th field coil which carries a current $i_1 = 1.5 \times 10^9$ amp, the total electrical energy E_c that would have to be fed into the field coils to charge up the system is given by

$$E_c = \tfrac{1}{2} i_1{}^2 \sum_{j=1}^{n} L_j$$

where

$$L_j = \mu_o R_j [\log(8R_j/r) - 1.75]$$

The radius r of each coil is equal to 5 m. Consequently, $E_c = 1.98982 \times 10^{18}$ Joules. Assuming that the solar array generates an average output of 50 GW per day, it would require a total of 460.61 days (1.26 years) to charge up the field coils. Although this charging time is fairly long, it is reasonable. It would represent an enormous energy reservoir. After the charging up process is completed, almost all of the power generated by the array could be sold to commercial utility companies to generate vast amounts of revenue that could continue indefinitely with very little operating expense.

The total specific energy $\hat{E}_o$ of the electromagnetic ramjet at any velocity V and altitude h above the plane of the field coils is equal to the sum of its kinetic and gravitational potential energy which is given by

$$\hat{E}_o = \tfrac{1}{2} V^2 + G\left[\frac{1}{R_e} - \frac{1}{R_e + (h - h_o)}\right] \tag{33}$$

The field coils **160** are designed to lift and magnetically accelerate the ramjet to high altitude at supersonic speed where the atmospheric pressure is sufficiently low to allow the electromagnetic ramjet propulsion system to begin operating with the microwave beam. This altitude will be about 60 km (200,000 ft) where the pressure is 0.2 Torr. Since sonic velocity (Mach 1.0) is about 320 m/sec at this altitude, a velocity of 1,200 m/sec (Mach 3.75) will be sufficient to start the ramjet. Consequently, the specific energy $\hat{E}_o$ corresponding to this end point of the boost phase of the launch trajectory is 1.302928×10^6 Joules/kg (which is obtained from equation (33) with h=60,500 m and h_o=500 m.) The altitude above the earth's surface is equal to $h = h_o = 60{,}000$ m.

The total initial specific magnetic launch energy E_m that the ramjet must have in order to lift the vehicle from the earth's surface and accelerate it to the initial point for microwave propulsion is

$$\hat{E}_m = \hat{E}_o + \hat{\Delta}_m \tag{34}$$

where $\hat{\Delta}_m$ is equal to the specific magnetic potential energy at an altitude of 60 km. If this altitude is significantly greater than the diameter of the outer field coil, then $\hat{\Delta}_m \approx 0$. However, since this is not the case for the preferred embodiment, $\hat{\Delta}_m$ cannot be ignored. If the total launch mass is m_t, then $\hat{\Delta}_m$ can be expressed as

$$\Delta_m = \frac{\mu_o M_2 S_2}{2m_t} \tag{35}$$

-continued

where

$$S_2 = \sum_{j=1}^{n} \frac{i_{1j} R_{1j}^2}{(R_{1j}^2 + h^2)^{1.5}} \tag{36}$$

The required magnetic launch energy $E_m = m_t \hat{E}_m$. Consequently, in view of equation (30), the total launch mass mt can be expressed as

$$m_t = \frac{\mu_o M_2 (S_1 - S_2)}{2E_o} \tag{37}$$

For the field coils described in the preferred embodiment h_o=500 m and $S_1 = 9.54098 \times 10^5$ amp/m and $S_2 = 1.41324 \times 10^4$ amp/m. The magnetic dipole moment M_2 corresponding to the six coil electromagnetic ramjet described above is 6.18179×10^{13} amp m². When these values are substituted into equation (37) the total launch mass m_t is equal to 27,928,651 kg (61,582,675 lbs or 30,791 tons).

Since the total mass of the six dipole coils comprising the electromagnetic ramjet is only 1,639,300 kg, this total launch mass is 17.036 times greater than the total coil mass. Thus, the weight lifting capability of the ramjet's superconducting dipole coils is truly enormous. This launch mass is many times greater than the largest mass ever launched from the earth's surface.

In order to understand how great this launch mass really is, it is helpful to make a comparison—it is actually greater than an average World War II Battleship or an Essex class aircraft carrier. Moreover, all of this launch mass will be accelerated into orbit (or to the moon) by the electromagnetic processes system. Since this will be accomplished by electromagnetic processes by converting electric power into propulsive thrust (with an efficiency approaching 100%) using electrical energy generated by the solar array, the cost will be zero. (Since the electromagnetic ramjet operates in the ionosphere at very high altitudes, the drag forces are essentially zero.)

The total amount of inductive energy that is taken out of the field coils to launch the vehicle will be equal to $m_t \hat{E}_o = 3.6389 \times 10^{13}$ Joules. Since this amount of inductive energy represents less than 0.002% of the total inductive energy stored in the field coils, the launch will have virtually no effect on its magnetic field. Moreover, if the solar array is used to replace this energy with a power of 100 GW, it would only require 364 seconds (6.06 minutes).

By combining equations (25), (28), and (37), the "mass ratio" m_t/m_c of the electromagnetic ramjet that is launched by the field coils can be expressed as

$$\frac{m_t}{m_c} = \frac{\gamma(S_1 - S_2)}{2\hat{E}_o \rho}\left[\frac{\mu_o \sigma}{\log(8\gamma) - .75}\right]^{\int} 8 \tag{38}$$

which, in view of equation (21) can be expressed to a good approximation by

$$\frac{m_t}{m_c} = \frac{.15\gamma(S_1 - S_2)(\mu_o \sigma)^{\frac{1}{2}}}{\rho \hat{E}_o} \tag{39}$$

Since the parameters σ, ρ, and E_o are fixed constants, the mass ratio of the ramjet depends only upon the aspect ratio γ of its dipole coils. Thus, the mass ratio (weight lifting capability of the ramjet) can be arbitrarily increased to any desired value by increasing the aspect ratio. (For the magnetic propulsion concept introduced herein, the efficiency of the propulsion system increases with increasing mass ratio instead of decreasing with mass ratio as in prior art rocket propulsion systems.)

The velocity V along the initial ascent trajectory at any altitude high above the plane of the field coils can be expressed by the equation

$$V = \left\{ \frac{\mu_o M_2(S_1 - S_2)}{m_t} - 2G\left[\frac{1}{R_e} - \frac{1}{R_e + (h - h_o)}\right]\right\}^{\frac{1}{2}} \tag{40}$$

The initial ascent trajectory **184** (FIG. **29**) of the ramjet **186** into the ionosphere will be along a curve of magnetic induction generated by the field coils **188**. This trajectory **184** is vertical at the launch point **190** and becomes parallel to the earth's surface **192** at an altitude of 100 km. However, when the ramjet **186** reaches an altitude of 60 km, the microwave beam from the first transmitter **194** is turned on and the vehicle is propelled by the combined thrust of the field coils **188** and the microwave beam. As the vehicle is accelerated further away from the field coils **188**, their propulsive effect gradually decreases. When the vehicle is ascending from the launch point **190** along the ascent trajectory **184** propelled by the magnetic repulsive force of the field coils **188**, the longitudinal central axis of the ramjet is maintained tangent to the ascent trajectory (i.e., tangent to the line of magnetic induction) so as to generate maximum repulsive thrust from the field coils.

Omitting the mathematical details it can be shown that a line of magnetic induction generated by the field coils passing through its center will follow a curve that can be represented in polar coordinates by the equation $r = C \sin^2\theta$ 8 where C is a constant. The vertical line passing through the center **190** of the field coils **188** represents the polar axis where $\theta = 0$. The condition that this curve be horizontal to the earth's surface **192** at an altitude of 100 km determines the value of C to be equal to 259.808 km. At this point $\theta = 54.736°$ and $r = 173.207$ km. The curve defined by the equation $r = 259.808 \sin^2\theta$ therefore defines the ascent trajectory from the launch point **190** to the point **196** where it reaches an altitude of 100 km. After the vehicle reaches 100 km, its trajectory remains circular at this altitude until the microwave propulsion is terminated.

When the altitude of the vehicle reaches 60 km, $\theta = 31.330°$, $r = 70.242$ km and the ground track **198** is 36.352 km east of the launch point **190**. This is the location of the first microwave transmitter **194**. Thus, as is shown in FIG. **29**, initially the microwave beam **200** will be pointing vertically upward into the inner dipole coil. The tangent line to the ascent trajectory **184** when the vehicle is directly over the first transmitter makes an angle $\phi = 48.258°$ with the verticle. In order to cancel out the gravitational force acting on the vehicle, the ramjet's thrust vector will have to be inclined upward at an angle $\alpha = 14.968°$ to this tangent line. Hence, the angle between the dipole axis and the incident microwave beam will be $\phi - \alpha = 33.290°$. If the radius of the ramjet's inner dipole coil is R_{21}, the initial beam diameter d_s will be equal to $2R_{21} \cos(\phi - \alpha) = 2R_{21} \times 0.8359$ which is nearly equal to the dipole's diameter $2R_{21}$. However, it should be noted that since the electric field vector of the microwave beam **200** is polarized to oscillate in a plane perpendicular to the central magnetic field lines of the ramjet, a strong coupling will result between the microwave beam and the free electrons passing through the resonance region which will pump them into very high energy electron cyclotron resonance orbits without requiring the beam to have a zero angle of incidence. Since the vehicle will be moving at a relatively slow initial velocity (1,200 m/sec) when the microwave propulsion begins, equation (2) shows that the initial beam power P does not have to be very high in order to achieve high propulsive thrust F.

As the vehicle passes over and begins to recede from the first transmitter **194**, (FIG. **29**) the angle of incidence between the microwave beam and dipole axis begins to decrease thereby enabling the beam diameter to be larger for greater propulsive power. When the vehicle reaches an altitude of 100 km, the vehicle will be 144.845 km from the first transmitter **194**, and the beam elevation angle relative to the local horizon will be 43.192°. At this point **196**, the dipole axis is inclined to its circular flight path by an angle $\alpha \approx 25°$ so as to provide an upward component for the thrust vector. Thus, the beam diameter d_s at this point **196** will be equal to $2R_{21} \cos(43{,}192 - \alpha) = 2R_{21} \times 0.95$ which is essentially equal to the inner dipole diameter. Since the vehicle will be moving at suborbital velocities, this upward thrust component will be necessary to maintain the vehicle's altitude. This vehicle flight attitude will also enable the microwave beam to have a nearly zero angle of incidence with the dipole axis thereby allowing a maximum beam diameter and hence maximum power transmission.

The first transmitter **194** continues to transmit prepulsive power until the elevation angle ϵ of the microwave beam **200** decreases to 10° relative to the local horizon. At this time, the first transmitter **194** is 477.398 km from the vehicle **202** (FIG. **30**) and 463.288 km from its ground track **198**. At this instant, the first transmitter **194** is turned off and the second transmitter **204** is turned on such that there is no interruption of propulsive power transmitted to the ramjet. In the preferred embodiment, the second transmitter **204** is located 400 km from the first transmitter **194** such that the initial beam elevation angle ϵ_1 of the second transmitter is 57.184° and the range is 118.609 km. As in the case of the first transmitter **194**, the second transmitter **204** continues to transmit microwave power until its beam elevation angle ϵ_2 decreases to 10° and the range is 477.398 km. At this instant, the second transmitter **204** is turned off and the third transmitter **206**, located 400 km further along the vehicle's ground track, is turned on. The vehicle is accelerated a distance of 406.278 km along its circular flight path **208** while being accelerated by the second microwave transmitter **204**. All succeeding transmitters operate with these same characteristics (i.e., maximum power transmission, and initial and terminal beam elevation angles ϵ_1, ϵ_2 of 57.184° and 10° respectively; beginning and ending beam transmission distances D_1, D_2 of 118.609 km and 477.398 km respectively; and a vehicle acceleration distance s_{12} of 406.278 km).

When the second transmitter **204** stops transmitting its propulsive power, the third transmitter **206** takes over in a precisely synchronized power relay operation such that the propulsive thrust generated by the electromagnetic jet is a smooth and continuous process that may involve many transmitters over a very long acceleration flight path **208**, This process of transmitting the microwave power by utilizing a series of transmitters **209** positioned at 400 km intervals along the vehicle's ground track **210** continues until the vehicle reaches the required orbital velocity. The ascent, magnetically propelled trajectory, and the microwave propelled trajectory are always assumed to have the same azimuth so that the ground track **210** for each launch remains unchanged. (However, the altitudes may vary to obtain optimal propulsion efficiency.)

In the case of very massive payloads as in the above numerical example, the electromagnet ramjet will require a long acceleration time that extends over several thousand kilometers in order to reach orbital velocities. The flight path **208** (FIG. **30**) will take it directly over a large number of microwave power transmitters that are located at 400 km intervals along the ground track. Assuming that the launch point is located in the north east corner of Nevada, the ground track will follow a southeasterly great circle arc across the United States and across the state of Florida. Since the system will be capable of generating extremely high launch velocities if the acceleration path is extended over long distances, the preferred embodiment will be designed to take maximum advantage of this operating feature. Consequently, in the preferred embodiment there will be 17 microwave transmitters constructed along a 6,800 km long ground track that begins in the north east corner of Nevada and extends southeasterly across the United States and West Indies, with the last transmitter located on Barbados Island. Six of the transmitters will be located on a chain of small islands in the West Indies. Since the unit cost of each transmitter system will decrease because of the economics of mass construction, the total cost of these transmitters should not exceed $30 billion.

At this point it is necessary to digress a moment to demonstrate some important operating characteristics of the electromagnetic ramjet. As pointed out above, in order for the vehicle to maintain an altitude of 100 km while it being accelerated by the microwave beam, the thrust vector $\vec{F}_p$ generated by the ramjet must be inclined at some angle α relative to the vehicle's circular flight path in order to cancel out the downward force of gravity. In particular, the vertical upward component F_{up} of the vehicle's thrust vector $\vec{F}_p$ together with the upward centrifugal force F_c must be equal to the downward force of gravity F_g. These forces can be expressed as

$$F_{up} = F_p \sin \alpha$$

$$F_c = \frac{m_t V^2}{R}$$

$$F_g = mg$$

where V is equal to the vehicle's velocity along the circular flight path of radius $R = R_e + 100{,}000\ m = 6{,}471{,}315\ m$ and where the gravitational constant at 100 km altitude is equal to 9.5062 m/sec^2.

When the vehicle is moving along the 100 km high circular trajectory **208** propelled by the microwave beam **200**, the propulsive power P will always be assumed to be maximum P_{max}. As described above $P_{max} = \pi R_{21}{}^2 P_{max}$ where P_{max} is equal to the maximum

power density limit (425 watts/cm^2) of the microwave beam set by atmospheric breakdown and where R_{21} is equal to the radius of the ramjet's inner dipole coil. In view of equation (2) the ramjet's propulsive force F_p along the thrust vector can be expressed as

$$F_p = \frac{P_{max}}{V_p}$$

where V_p is equal to the vehicle's velocity component along the thrust vector given by

$$V_p = V \cos \alpha$$

Therefore, the condition $F_{up} + F_c = F_g$ results in the equation

$$\tan \alpha = \frac{m_t}{P_{max}} \left(g - \frac{V^2}{R} \right) V \tag{41}$$

which can be used to determine the angle of inclination α. FIG. **31** is a graph of α versus vehicle velocity V corresponding to the above numerical example where $m_t/P_{max} = 2.32417 \times 10^{-5}\ kg/watt$. This inclination is important because it enables the microwave beam that is transmitted from the ground to intercept the coil plane of the ramjet at a nearly perpendicular angle (i.e., nearly broadside) for maximum power reception. Since circular orbital velocity at an altitude of 100 km is 7,843.32 m/sec, α becomes zero when V is equal to this value.

Although the ramjet's propulsive thrust vector $\vec{F}_p$ is inclined to the vehicle's circular trajectory in order to cancel out the downward force of gravity so that it can maintain its altitude, the tangental component F_t of the thrust vector along the trajectory which accelerates the vehicle along the trajectory is not diminished by this fact and is independent of α. This can be demonstrated as follows:

The component F_t of the ramjet's thrust vector along the flight path is equal to F_p cos α. However, $F_p = P_{max}/V_p$ where $V_p = V$ cos α. Consequently, $F_t = [P_{max}/(V \cos \alpha)] \cos \alpha = P_{max}/V$. Therefore, the propulsive force generated by the ramjet along the trajectory is independent of α and is equal to that generated when there is no gravitational force and the thrust vector can be directed along the flight path with magnitude given by equation (2).

If a_c denotes the vehicle's acceleration along the circular trajectory, the tangental propulsive force $F_t = m_t a_c = P_{max}/V$. Since $a_c = dv/dt$, this equation gives rise to a simple differential equation which can be solved to determine the time interval T required to achieve a velocity V while accelerating along the circular trajectory under maximum propulsive power. This time is given by the equation

$$T = \frac{3(V^2 - V_o^2)}{2K} \tag{42}$$

where V_o is equal to the initial vehicle velocity when it begins moving along 100 km high circular trajectory at point **196** (when T=0) and where K is a constant given by

$$K = \frac{3\,P_{max}}{m_t} \tag{43}$$

The distance s_c traversed by the vehicle accelerating along the circular trajectory can be expressed as

$$s_c = \frac{V^3 - V_o^3}{K} \tag{44}$$

Consequently, the vehicle velocity can be expressed as a function of s_c by the expression

$$V=[Ks_c+V_o^3]^{\frac{1}{3}} \tag{45}$$

The values of V_o and K will depend upon the particular situation. This will be illustrated by continuing the numerical example given above.

Table 1 is a detailed computer simulation of the magnetically propelled ascent trajectory from launch to an altitude of 60 km corresponding to the above example where the total launch mass m_t=27,928,651 kg. The various flight paramaters described in the table are:

H=altitude above earth's surface (km)
T=elapsed time from lift off (sec)
ϕ=flight path angle (from vertical, deg)
s=distance traveled along ascent trajectory (km)
F=propulsive thrust (10^6 N)
P=propulsive power (GW)
a=acceleration (g)
V=velocity (m/sec)
M=Mach number

TABLE 1

Flight Parameters of A 700 m Diameter Electromagnetic Ramjet Accelerating a Launch Mass of 27,928,651 kg To An Altitude of 60 km at Mach 3.75 By Magnetic Repulsive Forces Generated By 10 Underground Coaxial Superconducting Field Coils

h	T	ϕ	s	F	P	a	V	M
0	0	0	0	0	0	0	0	0
2	14.84	7.570	2.009	1,248	338	3.563	270.67	0.794
4	20.02	10.748	4.035	1,812	926	5.631	511.23	1.530
6	23.37	13.216	6.080	1,994	1,419	6.305	711.39	2.174
8	25.98	15.322	8.144	1,906	1,661	5.994	871.04	2.735
10	28.21	17.202	10.228	1,695	1,688	5.233	995.31	3.216
12	30.28	18.923	12.332	1,448	1,579	4.338	1090.67	3.632
14	32.17	20.526	14.457	1,206	1,402	3.464	1163.12	3.998
16	33.97	22.039	16.603	989	1,204	2.683	1217.68	4.262
18	35.72	23.479	18.772	805	1,013	2.020	1258.30	4.359
20	37.44	24.861	20.965	656	845	1.488	1288.06	4.416
22	39.15	26.194	23.181	534	700	1.053	1309.35	4.445
24	40.85	27.486	25.423	433	574	0.694	1323.98	4.453
26	42.55	28.745	27.690	354	473	0.417	1333.35	4.446
28	44.27	29.975	29.985	292	391	0.201	1338.54	4.423
30	46.00	31.181	32.309	241	323	0.023	1340.38	4.392
32	47.76	32.366	34.661	197	263	−0.127	1339.52	4.352
34	49.54	33.535	37.045	163	218	−0.239	1336.45	4.300
36	51.35	34.689	39.460	140	186	−0.312	1331.56	4.242
38	53.19	35.832	41.910	117	155	−0.383	1325.17	4.182
40	55.07	36.966	44.395	95	126	−0.451	1317.53	4.119
42	56.99	38.093	46.917	83	109	−0.483	1308.81	4.055
44	58.95	39.215	49.478	72	94	−0.511	1299.18	3.988
46	60.96	40.335	52.080	60	78	−0.542	1288.76	3.921
48	63.02	41.454	54.726	52	67	−0.558	1277.65	3.863
50	65.14	42.574	57.418	46	58	−0.570	1265.93	3.828
52	67.32	43.697	60.159	39	48	−0.582	1253.66	3.798
54	69.56	44.825	62.952	34	42	−0.585	1240.89	3.787
56	71.87	45.960	65.800	29	35	−0.591	1227.68	3.774
58	74.25	47.103	68.707	25	30	−0.590	1214.04	3.761
60	76.71	48.258	71.678	24	29	−0.578	1200.00	3.750

This table clearly demonstrates that the magnetic field generated by the field coils will exert a tremendous magnetic repulsive force on the electromagnetic ramjet such that it will lift the ramjet (along with the enormous payload mass) off the earth's surface and accelerate it to supersonic speed at a high altitude very rapidly. For example, 28.21 seconds after launch, the vehicle will be at an altitude of 10 km (32,808 ft) and moving nearly vertically upward with a velocity of 995 m/sec (Mach 3.22). At 37.44 seconds after launch, the vehicle will be at 20 km (65,618 ft) and moving at 1,288 m/sec (Mach 4.42). The vehicle essentially leaves the sensible atmosphere 46.00 seconds after launch with a velocity of 1,340 m/sec (Mach 4.39) at an altitude of 30 km (98,425 ft). Relatively little initial magnetic launch energy is lost by aerodynamic drag while traversing through the atmosphere. The system is therefore able to convert initial electrical energy into kinetic and potential energy with an efficieny of nearly 100%. FIG. **32** is a graph of acceleration (g versus altitude H (km) corresponding to this example.

Perhaps the most spectacular parameters in the table are propulsive thru F and propulsive power P. At any altitude of 6 km, the propulsive force is 1,994,000,000^N which is more than 60 times greater than the maximum lift off thrust generated by the giant Saturn V Apollo launch vehicle. The corresponding propulsive power is 1,419 GW. This enormous propulsive power is being generated by magnetic repulsive forces acting on the ramjet's superconducting dipole coil by the field coils.

It is important to point out and emphasize that the initial launch trajectory from the earth's surface to an altitude of 60 km described in Table 1, where the ramjet is propelled by the magnetic repulsive forces generated by the field coils, will be the same for each vehicle regardless of the ramjet's size or total launch mass. This is because the velocity of a vehicle along this initial portion of the trajectory which is given by equation (40) is determined only by the ratio of the ramjet's magnetic dipole moment divided by the total launch mass M_2/m_t. But in view of equation (37) this ratio is always the same and given by

$$\frac{M_2}{m_t} = \frac{2\hat{E}_o}{\mu_o(S_1 - S_2)} \tag{46}$$

Consequently, the acceleration along the initial launch trajectory is always the same regardless of launch mass and is given by Table 1 and described graphically in FIG. **32**.

When the vehicle reaches an altitude of 60 km and is moving at 1,200 m/s (Mach 3.75) the electromagnetic ramjet is started by transmitting microwave power from the first transmitter **194** (FIG. **29**). The amount of gravitational potential energy represented by the vehicle climbing from 60 km to 100 km (which is equal to 40,000 g) while being propelled by the ramjet will appear as a velocity loss. A small fraction of this loss will be canceled by the remaining magnetic potential energy $\hat{\Delta}_m$ generated by the field coils given by equation (35) which is equal to 19,654.41 m^2/sec^2. This residual magnetic potential energy is equivalent to a gravitational potential energy difference of 2,003.51 m (assuming $g=9.81/m/sec^2$). Consequently, the gravity losses will be equal to that resulting from a climb of 37,996.40 m, which will be equal to a velocity loss of 863.42 m/sec. Therefore, by assuming that the vehicle's initial velocity at the altitude of 60 km is (1,200.00−863.42) m/sec=336.58 m/sec instead of 1,200 m/sec, the combined effects of gravity and the residual magnetic field along the remaining portion of the trajectory can be ignored.

For simplicity, the microwave power will begin at a relatively low level and is gradually increased to give the vehicle a constant acceleration a_o along the remaining portion of the ascent trajectory (from 60 km to 100 km altitude) such that when the vehicle reaches an altitude of 100 km, the microwave power is at its maximum possible level $P_{max}=1.20166\times10^{12}$ watts. Consequently, if Δs is equal to the distance the vehicle travels along the remaining portion of its ascent trajectory (from 60 km altitude to 100 km altitude) which is equal to 114,732.50 m, the vehicle's velocity V_o after reaching 100 km altitude will be given by the equation

$$V_o = 336.58 \text{ m/sec} + \sqrt{2a_o\,\Delta s}$$

The vehicle's propulsive force F_o at the beginning of the circular 100 km high trajectory is given by

$$F_o = \frac{P_{max}}{V_o}$$

However, at this point the vehicle's acceleration is equal to a_o. Consequently since $F_o=m_t a_o$ it follows that

$$V_o = \frac{P_{max}}{m_t a_o} \tag{47}$$

By combining these two equations for V_o, the required acceleration a_o can be determined by solving the equation

$$\frac{P_{max}}{m_t a_o} = 336.58 \text{ m/sec} + 2a_o\Delta s \tag{48}$$

The result is $a_o=18.114$ m/sec^2. Consequently, $V_o=2{,}375.32$ m/sec.

With the value V_o determined, it is now possible (via equation **44**) to determine the required accelerating distance s_c the vehicle must travel along the circular trajectory **208** in order to reach a desired launch velocity V. For example, if the required launch velocity is 8,000 m/sec, the required distance $s_c=3{,}862.76$ km. In this case, the vehicle would be accelerated by only 11 of the 17 transmitters. As soon as the required launch velocity is reached, the microwave power transmission is terminated. The time T required to carry out this acceleration process can be determined by equation (42). This time is 678.17 sec (11.303 minutes). The peak acceleration is equal to a_o (which is 1.85 g). After reaching the required launch velocity, the vehicle moves out of the atmosphere along a free-fall trajectory. In order to remain out of the atmosphere during the next orbit revolution, the vehicle has to execute a small propulsive maneuver using onboard propulsion of a few meters per second when passing through its apogee.

FIG. **33** is a graph of velocity V versus acceleration distance s_c given by equation (45) corresponding to this numerical example where $V_o=2{,}375.32$ m/sec. FIG. **34** is a graph of V versus time T given by equation (42) corresponding to this numerical example.

In view of Table 1, the maximum overall acceleration will occur during the initial portion of the ascent trajectory when the vehicle is propelled by the repulsive magnetic field generated by the field coils. This maximum will be about 6.5 g and will occur when the vehicle is at an altitude of about 6,300 m (20,670 ft) and moving at Mach 2.3. Since this acceleration decreases very rapidly, it could be easily tolerated by human passengers. (The passengers would be lying in a prone position on special acceleration couches designed to enable them to withstand very high acceleration.)

In some situations it may not be possible to propel the vehicle to the required launch velocity during one pass over the transmitters. For example, suppose that the required launch velocity is not 8,000 m/sec, but 11,035.72 m/sec which would be required for a direct flight to the moon. In order to achieve this launch velocity, an acceleration distance $s_c=10{,}309$ km would be required. However, the system would still be able to give the vehicle the required launch velocity by carrying out two propulsive passes over the microwave power transmitters instead of only one pass.

In order to enable the vehicle to make another propulsive pass over the transmitters, the initial orbit must be such that it will take the vehicle back over the initial ground track over the transmitters at some future time. However, since the earth's surface is moving relative to an earth centered inertial frame by virtue of its rotation, the initial ground track will not be under the vehicle after the vehicle makes one complete revolution in its orbit. The orbital mechanics of the situation are such that in order for the vehicle to pass over the initial ground track at some future time, it must have an orbital period that is a rational fraction of one sidereal day (i.e., its orbital period P_v must be expressible as $(n_1/n_2)P_s$ where P_s is one sidereal day and n_1 and n_2 are integers). These orbits are called "synchronous orbits". If $n_1=1$, the vehicle will pass over the launch site exactly one sidereal day after it is launched after making n_2 complete orbit revolutions. This is the shortest possible time interval between successive pass overs. Thus, the initial orbit of the vehicle will be designed such that its orbital period $P_v=P_s/n_2$ where $p_s=86,164.099$ seconds and n_2 will be some integer $\leqq 16$. Since the period P_v of an orbit is related to its semi-major axis a by the equation

$$P_v = 2\pi\sqrt{\frac{a^3}{G}}$$

each integer n_2 corresponds to a unique value of a synchronous orbit's semi-major axis a_n that is determined by the equation

$$P_v = 2\pi n_2\sqrt{\frac{a_n^3}{G}}$$

Each of these values corresponds to a unique launch velocity V_n given by

$$V_n = \sqrt{G\left(\frac{2}{R_o} - \frac{1}{a_n}\right)}$$

where $R_o=6,471,315$ m, and each of these launch velocities corresponds to a required acceleration distance s given by equation (44).

For example, if $n_2=16$, then $a_n=6,637,660.85$ m, $V_n=7,940.990$ m/sec and $s_c=3,775.63$ km. If $n_2=15$, then $a_n=6,929,483.65$ m, $V_n=8,098.463$ m/s and $s_c=4,011.03$ km. If $n_2=2$, then $a_n=26,550,643.40$ m, $V_n=10,394.29$ m/sec and $s_c=8,596.42$ km. Since this distance exceeds the maximum acceleration distance of the electromagnetic ramjet, $n_2=2$ is too small. If $n_2=3$, then $a_n=20,261,933.10$ m, $V_n=10,167.97$ m/sec and $s_c=8,040.40$ km. Hence this value of n_2 is also too small. Values of 4 and 5 are also too small for n_2. However, if $n_2=6$, then $a_n=12,764,218.01$ m, $V_n=9,583.66$ m/sec and $s_c=6,715,49$ km which is within the maximum 7,000 km long acceleration distance. The best design strategy is to choose the initial synchronous orbit having the smallest eccentricity so that the ramjet will stay within the atmosphere over a long distance during its second pass over the transmitters. Therefore, the integer $n_2=16$ is selected. Since the earth's rotation will given the vehicle an additional 368.76 m/sec relative to an earth centered inertial frame, the actual launch velocity needed to achieve this synchronous orbit will be (7,940.99−368.76) m/sec=7,572.23 m/sec. The required acceleration distance $s_c=3,259.88$ km.

After the vehicle is launched onto this $n_2=16$ synchronous orbit (with an eccentricity $e=0.02506$) it will reach an apogee altitude (one-half revolution later) of 432.69 km. It will then make another half revolution and pass through its perigee point at an altitude of 100 km. Atmospheric drag will be very small and probably could be disregarded. However, if experience shows that this drag cannot be neglected, its effect can be canceled by increasing the initial launch velocity a small amount such that after 16 complete revolutions, the vehicle passes directly over the initial ground track and over the microwave power transmitters.

The initial velocity V_o corresponding to the second ramjet propulsive maneuver will be equal to the initial launch velocity, which is 7,572.33 m/sec. Taking into consideration the effect of the earth's rotation, the required launch velocity for a direct trajectory to the moon is equal to (11,035.72−368.76) m/sec=10,666.96 m/sec. Consequently, the required acceleration distance needed to reach this velocity corresponding to the second pass over the microwave power transmitters is 6,039.35 km. It will be completed in 655.94 seconds (10.93 minutes). The vehicle acceleration at the beginning and end of this second propulsive maneuver is 5.682 m/sec^2 (0.579 g) and 4.045 m/sec^2 (0.412 g) respectively. The trip to the moon would take 2.1 days.

It is important to point out the fact that since the acceleration path over the microwave transmitters takes place in a fixed plane relative to the earth's surface, the launch trajectory to a particular target and intercept point will have to satisfy this condition. In particular, the latitude ϕ_o and azimuth angle β_o of the launch point will depend upon the launch velocity. For example, low launch velocities will probably have a launch point over the United States, while high launch velocities may require a long accelerating distance that ends near Barbados Island with a latitude of only 11°.

If γ_{01} represents the free-fall transfer trajectory's sweep angle between the launch point and the target intercept point relative to an earth centered inertial frame, the declination angle δ_1 at the intercept point is determined by the equation

$$\sin\delta_1=\sin\phi_o\cos\gamma_{01}+\cos\phi_o\sin\gamma_{01}\cos\beta_o \qquad (49)$$

In the case of the lunar transfer trajectory given in the above example, $\phi_o=12°$ $\beta_o=124°$ and $\gamma_{01}=169.2°$. Consequently, the declination angle of the intercept point will be equal to −17.862°. Since the moon's declination angle is constrained by its orbital motion to lie within the range ±28.5°, the above described transfer trajectory could be easily accomplished with the electromagnetic ramjet accelerating system. (A detailed mathematical discussion of earth to moon free fall trajectories can be found in the book *Fundamentals of Astrodynamics,* Dover Publications, Inc. 1971, by Roger Bate.)

The vehicle would approach the moon with an asymptotic approach velocity of 1,198 m/sec. It could go into orbit around the moon by executing a retro propulsive maneuver using its onboard chemical rocket engines. For example, if the desired orbital altitude is 100 km, the vehicle would approach the moon along a hyperbolic trajectory to within 100 km and execute a ΔV retro maneuver of 968.39 m/sec. Assuming that the onboard chemical rocket propulsion system has a specific impulse of 490 sec (with an exhaust velocity of 4,806.9 m/sec) the amount of rocket propellant consumed during this retro propulsive maneuver would be 5,095,914 kg. Although this represents a very large amount of propellant, it is a relatively small fraction of the vehicle's actual payload. This payload (which would be about 20,000 tons) could be transferred down to the moon's surface by a fleet of relatively small lunar excursion molecules. (The propellant used to operate these vehicles could be manufactured from lunar material.)

After the payload is removed from the vehicle, the vehicle could accelerate out of lunar orbit and return to earth by means of its ECR propulsion system using the inductive energy stored in the ramjet's superconducting dipole coils. Since the ECR propulsion system will be capable of generating a much higher specific impulse than the chemical propulsion system, a significant saving in propellant mass can be achieved by using the ECR engine. Moreover, since it will be desirable to discharge the ramjet's dipole coils before reentering the earth's atmosphere in order to eliminate magnetic torque generated by the earth's magnetic field, it is very convenient to discharge the dipole coils by using the inductive energy to power the ECR engine.

Let P_e denote the propulsive power generated by the ECR system. If u_e denotes the exhaust velocity, the rate of propellant mass flow $\dot{m}_e$ is given b

$$m_e = \frac{2P_e}{u_e^2} \quad (50)$$

For any given propulsive power P_e, the rate of propellant mass flow m_e can be increased or decreased by decreasing or increasing the exhaust velocity u_e respectively. The amount of propellant m_p used to carry out the propulsive maneuver is given by

$$m_p = \dot{m}_e T_e \quad (51)$$

where T_e is equal to the time interval required to carry out the propulsive maneuver. Consequently, the required velocity increase ΔV can be expressed as

$$\Delta V = u_e \log\left(\frac{M_1}{M_1 - m_e T_e}\right) \quad (52)$$

where M_1 is equal to the total vehicle mass before the propulsive maneuver. In view of equation (50) this equation can be expressed as

$$\Delta V = u_e \log\left(\frac{M_1}{M_1 - (2P_e T_e/u_e^2)}\right) \quad (53)$$

The product P_eT_e in equation (53) is equal to the amount of propulsive energy used to carry out the maneuver. The exhaust velocity u_e of the ECR propulsion system will be designed such that the ramjet's dipole coils are completely discharged when the required $\Delta V = \Delta V_r$ is achieved. (For the lunar mission $\Delta V_r = 968.39$ m/sec.) Consequently, if η_e denotes the overall efficiency of the ECR engine, the required exhaust velocity u_e can be determined by solving the equation

$$\Delta V_r = u_e \log\left(\frac{M_1}{M_1 - (2\eta_e E_c/u_e^2)}\right) \quad (54)$$

where E_c is equal to the total inductive energy of the dipole coils given by equation (32). Suppose $M_1 = 3{,}000{,}000$ kg and $\eta_e = 0.90$. Therefore, since $E_c = 9.455 \times 10^{12}$ Joules, the exhaust velocity u_e required to return the vehicle back to earth with $\Delta V_r = 968.39$ m/sec using all of the inductive energy stored in the dipole coils would be equal to 6,318.57 m/sec. (This corresponds to a specific impulse of 644 sec.) The expelled propellant mass $m_p = 426{,}281$ kg. If the propulsive power of the ECR engine were 3 GW, the corresponding rate of propellant mass flow $\dot{m}_e = 150.28$ kg/sec, and the time T_e required to carry out the maneuver would be equal to 2,836.50 sec (47.28 minutes). If chemical rocket propulsion were used to carry out the ΔV_r maneuver, the required propellant mass would be 547,387 kg. Consequently, by using the ECR propulsion system instead of the chemical propulsion system, a propellant mass saving of 121,106 kg will achieved.

The vehicle returns to earth along an eccentric trajectory having a perigee altitude that dips into the earth's upper atmosphere. This altitude is designed to be sufficiently low to enable the vehicle to be decelerated to suborbital velocity by atmospheric breaking. Since the vehicle's mass will be relatively low, this initial deceleration will be relatively easy. After the vehicle is decelerated to suborbital velocity, the deceleration process can be very gradual, extending over several thousand kilometers so as to avoid excessive aerodynamic heating. Guidance and control of the vehicle are accomplished automatically via computerized navigation and flight control systems. After the vehicle is decelerated via aerodynamic breaking to a relatively low velocity, it executes a controlled glide to the landing area (i.e., launch site) and makes a final descent along a line of magnetic induction generated by the field coils. The resulting magnetic repulsive forces allow the vehicle to be decelerated such that its vertical velocity is exactly zero at the instant it makes contact with the landing-/launching platform. The vehicle's aerodynamic control surfaces, dive brakes, attitude control gyros and reaction thrusters provide a high level of vehicle control to enable it to land automatically with very high precision.

Although the dipole coils of the electromagnetic ramjet are discharged when it reenters the atmosphere, they are partially recharged by the inductive coupling with the field coils. This creates magnetic repulsive forces between the ramjet and the field coils that allows the vehicle to be magnetically decelerated as it approaches the launch site, and descends for a vertical landing.

Since the space transportation system made possible by the electromagnet ramjet propulsion concept disclosed herein allows vast quantities of payload to be transported from the earth's surface directly to the moon, it will be possible to construct a system of superconducting field coils, similar to that described for the earth, on the moon so that vehicles could be decelerated by magnetic repulsive forces and land directly on the moon's surface. It is instructive to describe the required magnetic field that would have to be generated by the superconducting coils.

If the vehicle on the above described lunar transfer trajectory were allowed to free-fall onto the moon's surface without any retro propulsion, the impact velocity $V_m = 2{,}658.63$ m/sec. Consequently, if lunar field coils are used to decelerate the vehicle by magnetic repulsive forces such that its velocity is zero immediately prior to making contact, the total amount of specific kinetic and potential energy E_o that must be converted into specific magnetic potential energy is given by

$$\hat{E}_m = \hat{E}_o = \tfrac{1}{2} V_m^2 = 3.5341469 \times 10^6 \text{ Joules/kg}$$

The total magnetic potential energy $E_m = m_t \hat{E}_m$. If B_m represents the magnetic field strength at the center of the field coils, the magnetic potential energy

$E_m=M_2B_m$ where M_2 represents the magnetic dipole moment of the electromagnetic ramjet. Consequently, the central magnetic field strength of the field coils can be determined by

$$B_m=\frac{m_tE_o}{M_2}$$

Assuming that the magnetic dipole moment of the electromagnetic ramjet remains unchanged with a value of 6.18179×10^{13} amp m^2, the required magnetic field strength of the lunar field coils will be 1.59669 T (15,966.9 Gauss). Thus, the superconducting coil system would have to be larger than the one described for the earth. However, if such a system were constructed and put into operation it would represent the ultimate in transportation efficiency. Since there is no atmosphere on the moon, nearly 100% of the total kinetic energy of a vehicle being decelerated to a gentle landing on moon's surface will be converted into electrical energy and stored inductively in the coils by the inductive coupling between the magnetically decelerated dipole coils of the ramjet and the field coils. This energy can therefore be used to launch the vehicle back to earth. However, since the mass of the vehicle returning to earth will be significantly lower than its arrival mass, the system will generate vast amounts of electrical energy that could be used to launch many magnetically propelled vehicles with payloads taken from the moon for other destinations. Such a system would be very versatile because unlike prior art electromagnetic accelerators designed for the moon's surface that require a long accelerating track with a fixed launch azimuth, this electromagnetic accelerating system does not require any accelerating track and will provide any launch azimuth desired. It will also be much safer since the vehicle being accelerated does not have to move near any stationary body (such as a guide rail or drive coils) where any accidental contact could result in a catastrophic disintegration of the entire system.

In order to determine the actual payload mass m_{pay} that could be delivered to the moon in the above example, the total mass of the ramjet propulsion system and connecting cables must be determined. Since the structural mass of the airfoil surround the propulsion coils, including all internal components except the coils themselves, will have a mass equal to about 25% of the total coil mass, this mass will be about 410,000 kg. The airfoils will have a chord of about 4.3 m and a thickness of about 60 cm (20 in). The total mass of the radial airfoils (FIG. **18**) will be approximately 25,000 kg. Thus, the total mass of the electromagnetic ramjet m_r will be equal to about 2,074,300 kg. Therefore, the total mass m_o that will be accelerated by the ramjet will be about 25,854,400 kg. Since the maximum acceleration a_m (including 1.0 g from gravity) will be 7.6 g, the corresponding maximum accelerating force $F_m=m_oa_m=1.902\times10^9$ N. Assuming that the total length of the central fuselage **108** is 400 m (1,312 ft) and is attached to the ramjet in a central position such that the fuselage extends 200 m on each side of the ramjet's coil plane, the supporting cables **110** attached to the inner airfoil and the rear of the fuselage **108** will make an angle of 56.3° with the central longitudinal axis of the ramjet. Consequently, the total tension F_t acting on these cables **110** when the vehicle is under maximum acceleration a_m will be equal to $F_m/\cos 56.3°=3.429\times10^9$ N. If A_t denotes the total cross sectional area of these cables and σ denotes the tensil strength, then $\sigma A_t=F_t$. Assuming that the supporting cables are constructed with high strength fused glass fibers where $\sigma=1.4\times10^{10}$ N/m^2, the required cross-sectional area $A_t=3.429\times10^9\ N/1.4\times10^{10}\ N/m^2=0.2449\ m^2$. The length of the cables will be $[300^2+200^2]^{\frac{1}{2}}=500\ m$. Therefore, since the density $\rho=2{,}160$ kg/m^3, the total mass of these cables will be 500 $A_t\rho=264{,}514$ kg. Thus, the total mass m_{cab} of all the supporting cables attached to the fuselage will be about 300,000 kg.

If the fuselage **108** is cylindrical with a maximum diameter of 30 m (98 ft) and is constructed with high strength composite material with a mass density of 3,000 kg/ms, the total empty mass m_f of the fuselage **108** will be about 700,000 kg. Consequently, if the total propellant mass $m_p=5{,}600{,}000$ kg, the total payload mass m_{pay} that could be delivered to the moon in the above numerical example would be $m_{pay}=m_t-m_r-m_{cab}-m_f-m_p=19{,}254{,}400\ kg$ (42,456,000 lbs or 21,300 tons).

It is estimated that in the future, the lowest cost of transporting payloads to the moon from the earth's surface will be about $25,000/lb using the most efficient prior art propulsion systems. Hence, it would cost over one trillion dollars (\$$10^{12}$) to deliver the same amount of payload mass m_p as described in the example. Using the electromagnetic ramjet propulsion system disclosed herein, this cost would be essentially zero. The electromagnetic ramjet propulsion system would also be much more reliable than any prior art propulsion system because the engine has no moving parts and burns no combustible fuel.

There is another economic aspect which will have a profound impact on the SDI Program. The current planning for SDI payloads calls for 50 million pounds to be orbited at a total launch cost of $130 billion. The electromagnetic propulsion system envisioned herein will be easily capable of delivering all of this payload in a single launch at zero cost. Moreover, the propulsion system will also be capable of transporting extremely large and bulky SDI payloads that could be completely assembled and tested on the ground before being transported to orbit as a single unit. This would eliminate the very tedious, costly, and time consuming process of having to assemble large SDI systems in orbit by transporting small components piece by piece.

The electromagnetic ramjet propulsion system could easily launch a huge fully assembled space station that is even larger and more elaborate than the giant toroidal design illustrated in Arthur C. Clarke's famous science fiction novel "2001 Space Odyssey"—fully manned with a crew of several hundred. In these ultra large payloads, an outer ablative protective jacket could be mounted around the hull that is designed to protect it while traversing through the atmosphere at high speed. (It could be enclosed in a detachable structure while being transported to orbit.) FIGS. **35** and **36** are schematic longitudinal and transverse verse cross sections illustrating how a fully assembled 25,000 ton toroidal space station **212** with a major and minor radius of 325 m and 20 m respectively, is attached to an electromagnetic ramjet **214** via detachable cables **216**. This toroidal space station **212** is equipped with two, 250 m long, 30 m diameter, central column cylinders **218** that are connected to the central hub **220** of the space station via swivel joints so as to remain stationary while the toroidal hull is rotating to provide an artificial earth-like gravitational field for the inhabitants by centrifugal force. Three relatively small 4 m diameter spoke cylinders **222** connect the hub **220** with the toroidal hull. When the ramjet transports the space station to a certain altitude, the calbes **216** are detached from the ramjet, and the ramjet returns to the launch site without any central structure. The space station is equipped with a plurality of ECR propulsion systems that maneuver the station to a desired final orbit. The design of the space station could be similar to the one disclosed in my U.S. Pat. No. 4,730,797, entitled "Inflatable Core Orbital Construction Method and Space station," filed Aug. 12, 1985.

FIG. **37** is a schematic longitudinal cross section of the electromagnetic ramjet **214** attached to a fully assembled toroidal space station **212** prior to launch. The space station **212** is assembled inside a large subsurface cylindric cavity **224** that is 220 m deep with a diameter of 700 m. The electromagnetic ramjet **214** is positioned around the rim **226** of the cavity **224** on a plurality of movable launching/landing platforms **228**. The main connecting cables **230** are positioned between the platforms **228**. When the space station is launched vertically upward, the platforms **228** simultaneously swing away from the rim **226** of the cavity **224** so that the ramjet can lift the space station upward by magnetic repulsive forces acting between its superconducting dipole coils and the surrounding superconducting field coils. The platforms **228** are fitted with superconducting current carrying cables **232** that charge up the ramjet coils prior to launch and discharge the coils after a landing. The toroidal hull **234** of the space station **212** can be constructed inside the cavity **224** using the method described in my U.S. Pat. No. 4,730,797. The two column cylinders **236,238** and all of their internal structures and systems can be constructed outside the cavity **224** and transported into it via large diameter underground access tunnels **240**. A plurality of large underground subassembly workrooms **242** are provided around the cavity **224**. All of the walls of the tunnels, workrooms, and cavity are fitted with superconducting shielding coils **244** which prevent the magnetic field of the surrounding field coils from entering the work areas.

It should be noted that although magnetic shielding is provided to prevent the surrounding magnetic field generated by the field coils from entering the underground facilities of the launch complex, there is no medical evidence indicating that exposure to 6,000 Gauss magnetic fields is biologically harmful. See "Human Tolerance to Magnetic Fields," *Astronautics*, March 1962 by D. E. Beischer. However, in order to minimize the risk of such exposure, some personnel working inside the underground launch complex can be provided with flexible suits that are designed to keep out a magnetic field.

A superconducting dipole coil **246** is mounted inside a tunnel **248** near the rim **226** of the cavity. This coil **246** is designed to generate a relatively weak local magnetic field with variable strength so as to assist in guiding the ramjet **214** to a precise landing on the landing/launching platforms **228**.

Large toroidal space stations represent ideal payloads for the electromagnetic ramjet because the area inside the circular periphery of a toroidal space station provides a large, relatively unobstructed central region **249** (FIG. **36**) for low density atmospheric gas to pass through the ramjet. Other payloads such as the 400 m long cylindrical central body **108** (FIGS. **18,19,20**) can also be introduced into the cavity **224** and attached to the electromagnetic ramjet **214** via detachable connecting cables. In this case, the central body would be brought into the cavity in two completely assembled sections, (loaded with payloads) and joined together inside the cavity **224** in a vertical position. This central body may contain two completely assembled self-refueling interplanetary transfer vehicles for carrying out ultra high speed interplanetary voyages around the entire solar system. Since these vehicles generate their own propellant while undergoing decelerating propulsive maneuvers, they never have to be refueled with propellant from external sources. The ability to launch fully assembled self-refueling interplanetary transfer vehicles would open up the entire solar system for manned exploration and eventual colonization. (See "A Generalized Theory of Classical Rocket Propulsion For Future Space Travel," *Journal of the British Interplanetary Society*, Vol. 40, No. 8, August 1987, pp. 341-352.)

In the practical application of the proposed ground to orbit space transportation system using the electromagnetic ramjet, it is important to consider the strength of the magnetic field generated by the field coils at various distances on the earth's surface. The field strength inside the central region **164** (FIG. **27**) will be nearly constant and (for the preferred embodiment) equal to 0.6 T (6,000 Gauss). For distances D from the center that are outside the outer field coil **168**, the magnetic field strength is given by

$$B = \frac{\mu_o}{4D^3} \sum_{j=1}^{n} i_{1j} R_{1j}^2$$

For the particular field coils in the preferred embodiment

$$B = \frac{1367.8\ (T\,\mathrm{km}^3)}{D^3}$$

where D is the distance in kilometers. Thus, the field strength in the region surrounding the field coils decreases very rapidly, inversely with the cube of the distance from the center. For example, at a distance D=30 km (8.17 km from the outer field coil **168**) the ground field is 0.0507 T (507 Gauss). At a distance D=50 km (28.17 km from the outer field coil) the ground field 0.0103 T (103 Gauss). At a distance of 100 km from the outer field coil, the ground field would only be 8 Gauss. These correspond to very weak magnetic fields and would be essentially imperceptible. At distances greater than about 400 km, the ground field would be less than the earth's natural magnetic field. The annular region close to the outer field coil where the magnetic field strength is high is occupied by the solar array **172** which is unaffected by the field. FIG. **38** is a graph of the magnetic field strength (Gauss) generated by the field coils on the earth's surface versus distance from the center D (km) for the preferred embodiment.

The strength of the magnetic field at various altitudes h above the center of the field coils is given by

$$B = \frac{\mu_o}{2} \sum_{j=1}^{n} \frac{i_{1j} R_{1j}^2}{[R_{1j}^2 + h^2]^{1.5}}$$

In the preferred embodiment of the field coils, the magnetic field strength at an altitude of 100 km directly over the center of the field coils would be 26 Gauss. At an altitude of 200 km, the field would only be 3 Gauss, and at an altitude of 500 km, it would only be 0.2 Gauss (which is approximately equal to the earth's natural magnetic field at this altitude). Thus, the region in space directly above the field coils **134** would be very weak at orbital altitudes. Consequently the field coils will not affect any artificial satellites moving in orbit above them. This is a beautiful situation as it allows the field to be relatively strong in a region where it is needed to lift the electromagnetic ramjet off the earth's surface and accelerate it to supersonic speed at a sufficiently high altitude to enable it to be accelerated by the microwave beam, but it decreases very rapidly beyond this region where it is no longer needed. FIG. **39** is a graph of the magnetic field strength in space directly above the field coils versus altitude.

Since the microwave power generating and transmitting stations are designed to transmit a maximum beam power of 10,000 GW (10^{13} watts), it is possible to construct a much larger electromagnetic ramjet to take full advantage of this enormous power capability. It will be instructive to carry out this calculation as it will provide additional teaching as to how the various equations developed herein are applied in the actual design of an electromagnetic ramjet. The radius R_{21} of the inner dipole coil of a multicoil electromagnetic ramjet designed to operate with the maximum 10^{13} watt microwave beam is determined by the equation

$$\pi R^2_{21}\,\rho_{max}=10^{13}\ W$$

where ρ_{max} is equal to the beam's power density at the onset of atmospheric max breakdown, which is 425 w/cm^2. Hence, $R_{21}=865.4$ m. Suppose that the ramjet has ten coaxial superconducting dipole coils with **10** m spacing between adjacent dipoles and that the magnetic field strength at the center generated by each coil is equal. Consequently, since the total strength of the magnetic field inside the inner dipole coil must be equal to 0.3572 T in order to generate electron cyclotron resonance with the 10 GHz microwave beam (as determined by equation 1) the central field of each dipole coil is equal to 0.03572 T. Upon substituting this value for B_o into equation (18), the aspect ratio γ of each coil in the ramjet will be equal to 3,772.0. The total magnetic dipole moment M_2 of the electromagnetic ramjet can now be calculated by equation (28). The result is $M_2=1.35167\times10^{15}$. The total launch mass m_t can now be calculated from equation (37). The result is $m_t=611{,}023{,}607$ kg. Since the total mass m_c of the dipole coils is 22,679,413 kg, this launch mass is a factor of 26.9 times greater than the coil mass. In the previous example, this "mass ratio" was 17.0. It is greater in this example because the aspect ratio γ of the dipole coils is greater than in the previous example (3,772.0 instead of 2,324.1). As is demonstrated by equation (38), the greater the aspect ratio, the greater the mass ratio (i.e., weight lifting efficiency).

Assuming that this mass is boosted to the 100 km high circular accelerating orbit along the same ascent trajectory as in the previous example, the initial velocity V_o, determined from equation (47) and (48) will be 1,786.46 m/sec The initial acceleration a_o along the circular accelerating orbit will be equal to 9.1611 m/sec^2(0.934 g). In order to reach orbital velocity (7,843 m/sec), the ramjet would have to accelerate the vehicle a distance of 8,419 km, which is obtained from equation (44). Since the maximum acceleration distance in the preferred embodiment is 7,000 km, it will not be possible to accelerate the launch mass to orbital velocity. The maximum mass that could be accelerated to orbit velocity with the preferred embodiment can be calculated by determining the value of K in equation (44) where $s_c=7\times10^6$ m, $V=(7{,}843-370)$ m/sec$=7{,}473$ m/sec and $V_o=1{,}786$ m/sec. The result is $K=58.805$ w/kg. Consequently, it follows from equation (43) that the maximum possible vehicle mass $\overline{m}_t=3\ P_{max}/K=510{,}156{,}788$ kg. Since the total mass of the ramjet's 10 dipole coils is 22,679,413 kg, and since the total mass of the ramjet's airfoils and all internal systems is assumed to be equal to 25% of the total coil mass, the total mass of the ramjet would be 28,349,266 kg. Hence, the total mass that could be transported to low earth orbit with this system would be equal to 481,807,522 kg (1,062,385,586 lbs or 531,193 tons). Since the ground to orbit transportation cost for the Space Shuttle is $2,000/lb, it would cost $2.125 trillion to orbit this amount of payload in that prior art system—and it would be in the form of 16,000 small pieces! Moreover, at a rate of one Shuttle launch every week, it would require over 300 years to deliver it. (By using the proposed scramjet propelled aerospace plane the cost could be reduced by one order of magnitude, but it would still require over a hundred years to deliver it—piece by piece.)

Although the capability of orbiting huge payloads with a mass of this magnitude may appear to be beyond any conceivable practical application, I do not believe this is the case. A ground-to-orbit transportation system with this capability would enable giant sized, completely assembled, toroidal habitats to be transported into orbit along with all of its inhabitants that may number in the tens of thousands.

The possibility of developing a ground-to-orbit transportation system having this capability is so remote in the prior art of astronautics that it is not even discussed in science fiction novels. However, the present disclosure introduces such a system that is well within present engineering feasibility. It is made possible by the electromagnetic ramjet propulsion concept and the recent developments in high temperature superconducting material which makes it possible. (Superconducting cable constructed with new ceramic high temperature superconducting material with critical temperature above liquid nitrogen is already being manufactured on a commercial basis. See "Superconducting Wire," *Popular Science,* March 1989, p. 17.)

The above example describes the performance capabilities of the most powerful electromagnetic ramjet that could be used in the preferred embodiment of the invention. It is therefore instructive to consider the performance capability of a ramjet having the smallest inner dipole diameter. This diameter is 61 m and is set by the minimum beam diameter at an assumed maximum range of 500 km. In this case $R_{21}=30.5$ m. Consequently, the ramjet's maximum propulsive power $P_{max}=\pi R_{21}{}^2\rho_{max}=1.242\times10^{10}$ w (12.42 GW). It will be assumed that this ramjet has 3 dipole coils separated by 2 m intervals with equal aspect ratio γ. Hence, $R_{22}=32.5$ m and $R_{23}=34.5$ m. In order to generate the 0.3572 T central magnetic field that is required for electron cyclotron resonance with 10 GHz microwave beam, each of the three coils must generate a central magnetic field $B_o=0.11907$ T. Hence, the aspect ratio of each coil will be 1,205.7 which is obtained by equation (18). The ramjet's total magnetic dipole moment $M_2=6.17777\times10^{10}$ amp m^2 (which is obtained by equation 28). The total launch mass $m_t=28{,}003$ kg (which is obtained from equation 37). Since the total coil mass $m_c=3{,}043$ kg, the mass ratio $m_t/m_c=9.20$. This mass ratio can also be obtained directly from equation (38). The total ramjet mass will be about $m_c+0.25\ m_c=3{,}804$ kg. Hence, the non-ramjet payload mass will be 24,199 kg.

Assuming, as in the previous examples, that the vehicle is accelerated to the 100 km high circular accelerating trajectory along the same ascent trajectory, the initial velocity $V_o=4{,}895.90$ m/sec and the initial acceleration $a_o=90.59$ m/sec^2 (9.23 g) which is obtained from equations (47) and (48). In order to reach an orbital velocity of 7,843 m/sec (which would require a launch velocity of 7,493 m/sec because of the earth's rotation) the ramjet would only have to accelerate the vehicle a distance of 228 km along the circular flight path, which is obtained from equation (44). Hence, only the first microwave transmitter would be used in this case. (More than half of the required orbital velocity would be obtained while climbing along the initial ascent trajectory). Since the initial launch trajectory between lift-off and 60 km altitude is the same for every vehicle regardless of its mass (because the ratio M_2/m_t which determines the launch trajectory given by equation (46) is unchanged) the vehicle's maximum acceleration will be 9.23 g. Although this acceleration is rather high, it is still within the limits of human toleration. (It could be easily lowered by reducing the propulsive power.) Consequently, since this smallest ramjet propelled vehicle would have the highest acceleration, every ramjet propelled vehicle in the preferred embodiment of the invention would be capable of transporting human passengers. However, these very small ramjet propelled vehicles would be ideal as automated unmanned ground to orbit supply vehicles.

Since any number of electromagnetic ramjets could be used to accelerate payloads into orbit (one at a time) by the same field coils and microwave transmitters, the ground-to-orbit transportation system envisioned herein is very large and includes many individual ramjets of various sizes. FIG. **40** is an enlarged schematic plan view of the central 10 km diameter launching/landing region **250**. The center of this region contains 7 launching/landing terminals **252** designed for the largest ramjets with inner and outer dipole diameters of 1,731 m and 1,911 m respectively. These terminals are surrounded with other launching/landing terminals **254** using smaller ramjets. The smallest ramjets have an inner dipole diameter of **61** m and use the smallest outer launching/land terminals **256**. A network of underground access tunnels **258** connect the various terminals. Passengers could be transported to and from various launching/landing terminals via high speed trains operating in some of the underground tunnels **258**.

The field coils **160** (FIG. **27**) may also have many other embodiments. For example, if it is possible to initiate the microwave propelled ramjet propulsion system at relatively low altitudes instead of at an altitude of 60 km, only one field coil would have to be constructed. In fact, in some embodiments of the invention, a vehicle propelled by an electromagnetic ramjet propulsion system could be boosted to some initial altitude and speed by a conventional chemical rocket system without requiring any field coils. After reaching this initial altitude, the rocket booster could be separated from the vehicle and the vehicle could be propelled by the electromagnetic ramjet receiving propulsive power from a microwave beam. However, as described herein, the preferred embodiment of the invention uses a plurality of large superconducting field coils to initially lift the vehicle off the earth's surface and to propel it to some initial altitude because this will enable the system to launch much greater payloads into orbit.

It should also be pointed out that the magnetic field generated by the field coils used in the preferred embodiment of the invention, may be increased by the presence of underground deposits of iron ore or other natural ferromagnetic substances in the earth near the field coils. This would be very desirable as it would reduce the number of field coils required to establish a desired magnetic field in the vicinity of the launch site.

In another embodiment of the invention, an orbiting microwave transmitter could be used to power the electromagnetic ramjet instead of a plurality of ground based transmitters. This would enable the accelerating distance s_c of the ramjet to be essentially unlimited. It would also enable the maximum power density ρ_{max} of the microwave beam to be increased because atmospheric breakdown at very high altitudes requires much higher power densities.

In still other embodiments, the frequency of the microwave beam may be other than 10 GHz. This is a design detail that would be determined after a more detailed study of the invention involving optimization analysis.

It should also be pointed out and emphasized that the construction of the self-supporting dipole propulsion coils described above, where thin layers of ceramic superconducting material is vapor deposited directly onto fused silica glass fibers, is not a necessary construction feature in the practice of this invention. The superconducting material can be deposited on some intermediate fiber substrate having high tensil strength σ and low density ρ such as carbon fibers. The superconducting tension bearing cable could then be fabricated by embedding the carbon filaments in a matrix of fused silica glass fibers as described in my previously mentioned U.S. Pat. No. 4,078,747.

Although constructing the ramjet's superconducting dipole coils with new high temperature superconducting material is an important construction feature of the invention, this feature is not a necessary construction feature in the practice of this invention. The dipole coils could be constructed with ordinary superconducting material that is vapor deposited onto the stress bearing fibers as disclosed in my U.S. Pat. No. 4,078,747. However, in this case the liquid hydrogen cryogenic refrigerant **36** (FIGS. **8**,**9**) used for maintaining the ramjet's superconducting dipole coils **10** at cryogenic temperature would have to be replaced with liquid helium.

It should be emphasized that the ground to orbit electromagnetic propulsion concept and operating system disclosed herein is radically different from prior art electromagnetic propulsion concepts that also use magnetic forces generated by current carrying coils to generate propulsive thrust. In prior art electromagnetic propulsion systems, (which are called "Mass drivers" or "coaxial electromagnetic accelerators") the moving propulsion coil (which is mounted around the payload) is very small, usually only a few centimeters in diameter, and moves through an evacuated tube surrounded by a plurality of identical coaxial drive coils mounted at various intervals along the tube. The drive coils are exited (i.e., pulsed with a current surge) in a sequence timed with the moving coil to generate a travelling magnetic field. Thus, the moving coil is accelerated by series of relatively small incremental steps as it passes through each drive coil. These prior art systems are very complicated electronically (because of the required timing circuits) and can only accelerate very small payloads that are restricted in size by the diameter of the evacuated accelerating tube. Moreover, if the moving coil makes accidental contact with the inside walls of the tube, it could cause the system to disintegrate because of the very high speeds involved Since the tube must be several kilometers long in order to reduce acceleration loads, its construction will be very expensive and prone to catastrophic failure. These are fundamental problems inherent in prior art electromagnetic propulsion systems that cannot be circumvented. Thus, the practical utility of prior art ground-to-orbit launch systems using electromagnetic forces is severely limited. A good technical description of this prior art can be found in the following papers: "Basic Principles Of Coaxial Launch Technology," *IEEE Transactions On Magnetics,* Vol. MAG-20, No. 2, March 1984, pp. 227-230, by H. Kolm and P. Mongeau; and "An Alternative Launching Medium," IEEE Spectrum, April 1982, pp. 30-36, by H. Kolm and P. Mongeau.

The electromagnetic propulsion concept disclosed herein generates its magnetic propulsive force by creating a moving repulsive magnetic field behind superconducting dipole propulsion coils by utilizing low density ionized gas in the earth's upper atmosphere. This is achieved by directing a microwave beam at the propulsion dipoles having a frequency designed to produce electron cyclotron resonance with free electrons moving through the coils. When free electrons move into this resonance region they are rapidly accelerated into circular cyclotron orbits perpendicular to the lines of magnetic induction of the coils. The electrons are thus pumped into very high energy cyclotron orbits by the microwave beam. A strong coupling between the electrons and the microwave wave beam is created resulting in the absorption of essentially all of the beam energy by the electrons. The electrons become strong magnetic dipoles which, in their aggregate, generate a strong repulsive magnetic field behind the propulsion coils that accelerates the coils forward by magnetic repulsive forces. The unique relationship between the magnetic field of the propulsion coils and the frequency of the microwave beam required to produce electron cyclotron resonance is such that very large coil diameters are possible so that thousands of gigawatts of beam power could be used for accelerating giant size payloads with dimensions and mass many orders of magnitude greater than could be accelerated by any prior art propulsion concept. This is at the heart of the invention.

Unlike prior art beamed power propulsion concepts, where a beam is used to heat a working fluid in a relatively small heating region, so that it can be expelled at high velocity to generate propulsive thrust, the beam in the present invention is not absorbed by thermalization processes. Almost all of the beam energy is used to pump free electrons into high energy cyclotron orbits in a very large region. Thus, the microwave beam is converted directly into propulsive thrust with an efficiency that could, in principle, approach 100%. Consequently, the cost of transporting payloads into orbit with this system is essentially equal to the cost of generating an amount of electrical energy equal to the orbital energy. Since the system generates its own electrical energy, this cost is essentially zero.

Although the propulsion system is basically a high altitude ramjet, it is fundamentally different from all prior art ramjets in that it does not require any fuel, and does not involve any heating or combustion processes. Consequently, it has no inherent velocity limitations and, since it generates its propulsive power from a microwave beam, it has no energy limitations. Since the electromagnetic ramjet propulsion system does not have any moving parts and operates by electrical processes, it would be extremely reliable.

Perhaps the most striking operating feature of the electromagnetic ramjet propulsion system disclosed herein is the fact that it can be operated as a reusable self-contained payload acceleration system whereby pre-assembled payloads with very large dimensions and virtually unlimited mass can be lifted off the earth's surface and accelerated into orbit at essentially zero cost. This is a fundamental physical impossibility within the prior art of space travel and will have truly revolutionary implications. For example, it will enable huge completely assembled and fully equipped space stations, manned with hundreds (or thousands) of occupants with a mass of tens of thousands of tons to be transported into essentially any orbit around the earth or the moon at zero cost. Hundreds (of even thousands) of such stations, in the form of orbiting self-contained cities, could be transported into space on such a vast scale that a significant fraction of the total world population could be living in them before the end of the next century. It could also be used to transport huge, completely assembled self-refueling space based interorbital/interplanetary vehicles operating under the generalized theory of rocket propulsion to transport hundreds of passengers at very low cost on high speed interplanetary trajectories throughout the entire solar system. This could lead to the colonization of the solar system.

From the foregoing description, it will thus be evident that the present invention provides a vastly improved method for propelling payloads from the earth's surface into orbit for achieving economical space transportation. As various changes and modifications can be made in the above system and operating method without departing from the spirit or scope of the invention, it is intended that all matter contained in the above description or shown in the accompanying drawings, should be interpreted as illustrative and not in a limiting sense.

What is claimed is:

1. A method for generating propulsive thrust comprising the steps of:

charging a superconducting coil with electric current to generate a magnetic field;

providing a source of ionized gas moving through a portion of said magnetic field; and

directing a beam of electromagnetic radiation into said magnetic field with a frequency designed to produce electron cyclotron resonance with free electrons of said ionized gas moving through said magnetic field thereby accelerating said ionized gas away from said coil and generating propulsive thrust.

2. A method as set forth in claim **1** wherein said gas is atmospheric air that is moving through said coil.

3. A method as set forth in claim **1** further comprising the step of mounting said superconducting coil inside a circular wing-like airfoil having an open duct-like central region.

4. A method as set forth in claim **3** further comprising the step of maintaining said coil at cryogenic temperature by cryogenic cooling means mounted inside said airfoil.

5. A method as set forth in claim **3** wherein said ionized gas is atmospheric air and further comprising the step of moving said airfoil through the atmosphere so as to provide a continuous source of gas moving through said coil

6. A method as set forth in claim **5** wherein said beam of electromagnetic radiation is transmitted from a transmitter located on the earth's surface.

7. A method as set forth in claim **5** further comprising the step of mounting movable aerodynamic control surfaces on said airfoil adapted for guiding said airfoil through the atmosphere.

8. A method as set forth in claim **5** wherein a portion of said ionized air is generated by the step of moving said airfoil at supersonic speed thereby generating supersonic shock waves and creating ionized gas.

9. A method as set forth in claim **5** wherein said airfoil establishes a central longitudinal axis passing through the center of said coil and perpendicular to the plane of said coil and furhter comprising the step of mounting a payload carrying central body in the center of said airfoil extending along said longitudinal axis.

10. A method as set forth in claim **9** further comprising the step of rotating said central body about an axis perpendicular to said longitudinal axis when said airfoil is moving outside the earth's atmosphere.

11. A method as set forth in claim **10** further comprising the step of mounting secondary thrust generating means inside said airfoil and inside said central body for controlling the attitude of said airfoil and for providing maneuvering thrust when said airfoil is moving outside said atmosphere.

12. A method as set forth in claim **5** further comprising the steps of:

creating a static magnetic field on the earth's surface that extends upward into the atmosphere;

launching said airfoil from the earth's surface to some initial altitude at some initial velocity by magnetic repulsive forces exerted on said coil by said static magnetic field; and

accelerating said airfoil to higher velocity by directing said beam or electromagnetic radiation into said magnetic field generated by said superconducting coil when said airfoil reaches said initial altitude,

13. A method as set forth in claim **5** further comprising the steps of:

mounting aerodynamic breaking means on said airfoil; and

decelerating said airfoil by said aerodynamic breaking means so that said airfoil means can be returned to the earth's surface.

14. A method as set forth in claim **3** wherein said step of providing a source of ionized gas comprises the step of moving said airfoil through the earth's ionosphere and utilizing said ionosphere as a source for said ionized gas.

15. A method as set forth in claim **1** wherein said superconducting coil is self-supporting dipole with an aspect ratio exceeding 100.

16. A method for generating propulsive thrust comprising the steps of:

moving a current carrying superconducting coil means generating a magnetic field through the atmosphere such that atmospheric air flows through said coil means; and

creating a secondary repulsive magnetic field behind said coil means by free electrons in said atmospheric air moving in circular electron cyclotron orbits in the magnetic field of said coil means energized by a microwave beam tuned to the electron cyclotron resonant frequency of said orbiting electrons.

17. A method as set forth in claim **16** wherein said microwave beam is transmitted from a transmitter located on the earth's surface.

18. A method as set forth in claim **16** furhter comprising the step of mounting said superconducting coil means inside circular wing-like airfoil means having an open duct-like central region.

19. A method as set forth in claim **16** wherein said superconducting coil means comprises a plurality of superconducting coils further comprising the steps of

mounting said coils inside a like plurality of circular wing-like airfoils having an open duct-like central region; and

maintaining said airfoils in a co-axial spaced-apart relationship with a plurality of radial supporting airfoils mounted between adjacent circular airfoils.

20. A method as set forth in claim **18** further comprising the step of maintaining said coil means at cryogenic temperature by cryogenic cooling means mounted inside said airfoil means.

21. A method as set forth in claim **18** further comprising the step of mounting movable aerodynamic control surfaces on said airfoil means adapted for guiding said airfoil means through the atmosphere.

22. A method as set forth in claim **18** further comprising the step of generating free electrons in the atmosphere by moving said airfoil means at supersonic speed thereby shock ionizing a portion of said air flowing through said airfoil means.

23. A method as set forth in claim **18** further comprising the step of moving said airfoil means through the earth's ionosphere thereby utilizing natural ionized gas present in the ionosphere as a source for free electrons.

24. A method as set forth in claim **18** wherein said airfoil means establishes a longitudinal central axis passing through the center of said airfoil means and further comprising the step of mounting a payload carrying central body in the center of said airfoil means extending along said longitudinal axis.

25. A method as set forth in claim **24** further comprising the step of rotating said central body about an axis perpendicular to said longitudinal axis when said airfoil is moving outside the earth's atmosphere thereby allowing said central body to be aligned along any desired direction while said coil means is constrained by magnetic torque exerted by the earth's natural magnetic field.

26. A method as set forth in claim **25** further comprising the step of mounting secondary thrust generating means inside said airfoil means and inside said central body adapted for controlling the attitude of said airfoil means and for providing maneuvering thrust when said airfoil means is moving outside said atmosphere.

27. A method as set forth in claim **18** further comprising the steps of:

creating a static magnetic field on the earth's surface that extends upward into the atmosphere;

launching said airfoil means from the earth's surface to some initial altitude at some initial velocity by magnetic repulsive forces exerted on said coil means by said static magnetic field; and

accelerating said airfoil means to higher velocity when said airfoil mean reaches said initial altitude by directing said microwave beam into said magnetic field generated by said coil means thereby creating said secondary repulsive magnetic field behind said coil means.

28. A method as set forth in claim **18** further comprising the steps of:

mounting aerodynamic breaking means on said airfoil means; and

decelerating said airfoil means by said aerodynamic breaking so that airfoil means can be returned to the earth's surface.

29. A method as set forth in claim **16** wherein said superconducting coil means comprises a plurality of self-supporting dipoles having aspect ratios exceeding 100.

30. A method for generating propulsive thrust comprising the steps of:

charging a superconducting coil means with electric current thereby generating a magnetic field; and

accelerating ionized gas away from said coil means by a microwave beam having a frequency tuned to produce electron cyclotron resonance with free electrons of said ionized gas moving in said magnetic field.

31. A method as set forth in claim **30** wherein said ionized gas is atmospheric air moving through said coil means.

32. A method as set forth in claim **30** further comprising the steps of:

mounting said superconducting coil means inside circular wing-like air-foil means having an open duct-like central region; and

moving said airfoil means through the atmosphere such that said gas is atmospheric air moving through said duct-like central region.

33. A method as set forth in claim **32** further comprising the step of ionizing a portion of said atmospheric air moving through said airfoil means.

34. A method as set forth in claim **32** further comprising the step of maintaining said coil means at cryogenic temperature by cryogenic cooling means mounted inside said airfoil means.

35. A method as set forth in claim **32** wherein said microwave beam is transmitted from a transmitter located on the earth's surface.

36. A method as set forth in claim **32** further comprising the step of mounting movable aerodynamic control surfaces on said airfoil means for guiding said airfoil means through the atmosphere in controlled flight.

37. A method as set forth in claim **32** wherein said circular airfoil means establishes a central longitudinal axis passing through the center of said airfoil means further comprising the step of mounting a payload carrying central body in the center of said airfoil means extending along said longitudinal axis.

38. A method as set forth in claim **37** further comprising the step of rotating said central body about an axis perpendicular to said longitudinal axis when said airfoil means is moving outside the earth's atmosphere thereby allowing said central body to be aligned along any desired direction while said coil means is constrained by magnetic torque exerted by the earth's natural magnetic field.

39. A method as set forth in claim **38** further comprising the step of mounting secondary thrust generating means inside said airfoil means and inside said central body for controlling the attitude of said airfoil means and for providing maneuvering thrust when said airfoil means is moving outside said atmosphere.

40. A method as set forth in claim **32** further comprising the step of moving said airfoil means through the atmosphere at supersonic speed thereby generating shock waves and ionizing a portion of said atmospheric air moving through said airfoil means.

41. A method as set forth in claim **32** further comprising the steps of:

creating a static magnetic field on the earth's surface that extends upward into the atmosphere;

launching said airfoil means from the earth's surface to some initial altitude at some initial velocity by magnetic repulsive forces exerted on said coil means by said static magnetic field; and

propelling said airfoil means to higher velocity when said airfoil means reaches said initial altitude by accelerating ionized gas away from said airfoil means by transmitting said microwave beam tuned to said electron cyclotron resonant frequency.

42. A method as set forth in claim **32** further comprising the steps of:

mounting aerodynamic breaking means on said airfoil means; and

decelerating said airfoil means by said aerodynamic breaking means so that said airfoil means can be returned to the earth's surface.

43. A method as set forth in claim **32** further comprising the steps of:

moving said airfoil means through the earth's ionosphere while generating said propulsive thrust; and

utilizing said ionosphere as a source for ionized gas.

44. A method as set forth in claim **30** wherein said superconducting coil means comprises a plurality of circular superconducting coils further comprising the steps of:

mounting said coils inside a like plurality of circular wing-like airfoils with an open duct-like central region; and

maintaining said airfoils in a co-axial, spaced-apart relationship with a plurality of radial supporting airfoils mounted between adjacent circular airfoils.

45. A method as set forth in claim **30** wherein said superconducting coil means comprises a plurality of self-supporting dipoles having aspect ratios exceeding 100.

46. A method for accelerating a body to high velocity comprising the steps of

charging a superconducting coil means with electric current so as to create a magnetic field;

mounting said superconducting coil means inside an airfoil means adapted for flight through the atmosphere;

attaching said body to said airfoil means; and

creating a travelling repulsive magnetic field behind said airfoil means by inducing free electrons in the atmosphere to move in circular electron cyclotron resonance orbits by a microwave beam thereby accelerating said airfoil means by magnetic repulsive forces.

47. A method as set forth in claim **46** wherein said microwave beam is transmitted from a transmitter means located on the earth's surface.

48. A method as set forth in claim **46** further comprising the step of maintaining said superconducting coil means at cryogenic temperature by cryogenic cooling means mounted inside said airfoil means.

49. A method as set forth in claim **46** further comprising the step of ionizing a portion of atmospheric air for generating free electrons.

50. A method as set forth in claim **46** wherein said airfoil means has a circular wing-like shape with an open circular duct-like central region that is surrounded by said coil means and further comprising the step of moving said airfoil means so that atmospheric air passes through said open duct-like region.

51. A method as set forth in claim **50** further comprising the step of mounting movable aerodynamic control surfaces on said airfoil means for guiding said airfoil means through the atmosphere in controlled flight.

52. A method as set forth in claim **50** further comprising the step of mounting secondary thrust generating means inside said airfoil means for controlling the attitude of said airfoil means and for providing maneuvering thrust when said airfoil means is moving outside the earth's atmosphere.

53. A method as set forth in claim **50** further comprising the step of moving said airfoil means through the atmosphere at supersonic speed thereby generating shock waves and ionizing a portion of said atmospheric air moving through said open duck-like airfoil means.

54. A method as set forth in claim **50** wherein said body has a length less than the diameter of said open duct-like region further comprising the step of mounting said body inside said duct-like region on a beam extending along a diameter of said duct-like region and such that said body can be rotated around said beam.

55. A method as set forth in claim **54** wherein said body defines a longitudinal central axis that is perpendicular to said beam further comprising the steps of:

rolling said body about said longitudinal axis; and

rotating said body around said beam so that said longitudinal axis can be pointed in any direction independent of the orientation of said airfoil means.

56. A method as set forth in claim **47** further comprising the steps of:

creating a static magnetic field on the earth's surface that extends upward into the atmosphere;

launching said airfoil means from the earth's surface to some initial altitude at some initial speed by magnetic repulsive forces exerted on said coil means by said static magnetic field; and

accelerating said airfoil means to higher velocity when said airfoil means reaches said initial altitude by creating said travelling repulsive magnetic field by said microwave beam transmitted from said transmitter means.

57. A method as set forth in claim **50** further comprising the steps of:

mounting aerodynamic breaking means on said airfoil means; and

decelerating said airfoil means by said aerodynamic breaking means so that said airfoil means can be returned to the earth's surface.

58. A method as set forth in claim **46** further comprising the steps of:

moving said airfoil means through the earth's ionosphere; and

utilizing the ionosphere as a source for ionized gas for creating said travelling repulsive magnetic field.

59. A method as set forth in claim **46** wherein said superconducting coil means comprises a plurality of circular superconducting coils and further comprising the steps of:

mounting said coils inside a like plurality of circular wing-like airfoils with an open duct-like central region; and

maintaining said airfoils in a co-axial, spaced-apart relationship with a plurality of radial supporting airfoils mounted between adjacent circular airfoils.

60. A method as set forth in claim **46** wherein said superconducting coil means comprises a plurality of self-supporting dipoles having aspect ratios exceeding 100.

61. An apparatus for generating propulsive thrust comprising:

superconducting coil means for generating a magnetic field;

means for passing ionized gas through a portion of said magnetic field; and

means for transmitting a microwave beam into said magnetic field with a frequency designed to produce electron cyclotron resonance with free electrons of said ionized gas moving through said magnetic field thereby accelerating said ionized gas away from said coil means and generating propulsive thrust.

62. An apparatus as set forth in claim **61** wherein said ionized gas is atmospheric air.

63. An apparatus as set forth in claim **61** further comprising:

circular wing-like airfoil means having a duct-like central region; and

means for mounting said superconducting coil means inside said airfoil means.

64. An apparatus as set forth in claim **63** further comprising cryogenic cooling means mounted inside said airfoil means for maintaining said superconducting coil means at cryogenic temperature.

65. An apparatus as set forth in claim **63** wherein said circular airfoil means is moved through the atmosphere such that atmospheric air flows through said duct-like central region and through said magnetic field.

66. An apparatus as set forth in claim **65** wherein said means for transmitting said microwave beam comprises a phased array microwave transmitter located on the earth's surface.

67. An apparatus as set forth in claim **65** further comprising:

movable aerodynamic control surfaces mounted on said wing-like airfoil means;

actuator means for moving said movable control surfaces mounted inside said airfoil means;

guidance and navigation computer means mounted inside said airfoil means; and

electrical conduit means connecting said computer means with said actuator means enabling said airfoil means to be guided automatically through the atmosphere along a pre-calculated flight path.

68. An apparatus as set forth in claim **65** further comprising means for generating supersonic shock waves mounted inside said duct-like central region of said airfoil means such that when said airfoil means is moving at supersonic speed, a portion of said atmospheric air flowing through said duct-like central region is ionized by said shock waves.

69. An apparatus as set forth in claim **65** wherein said duct-like region establishes a longitudinal central axis further comprising means for mounting a payload carrying central body in the center of said airfoil means extending along said longitudinal axis wherein the transverse cross sectional area of said central body is significantly less than the transverse cross sectional area of said open duct-like central region so as to not restrict the atmospheric air flowing through said central region.

70. An apparatus as set forth in claim **69** wherein said central body has a length less than the diameter of said airfoil means and further comprising means for rotating said central body about an axis perpendicular to said longitudinal axis when said airfoil means is moving outside the earth's atmosphere thereby allowing said central body to be aligned along any desired direction while said coil means is constrained by magnetic torque exerted by the earth's natural magnetic field.

71. An apparatus as set forth in claim **70** further comprising attitude control and maneuvering propulsion means mounted inside said airfoil means and inside said central body for providing attitude control and maneuvering propulsion when said airfoil means is moving outside the earth's atmosphere.

72. An apparatus as set forth in claim **65** further comprising:

means for creating a static magnetic field on the earth's surface that extends upward into the atmosphere; and

means for launching said airfoil means from the earth's surface to some initial altitude at some initial velocity by magnetic repulsive forces exerted on said coil means by said static magnetic field such that when said airfoil means reaches said initial altitude, said microwave beam is turned on from a microwave transmitter located on the earth's surface thereby accelerating said airfoil means to higher velocity.

73. An apparatus as set forth in claim **72** further comprising a plurality of spaced-apart microwave transmitters located on the earth's surface such that when said airfoil means moves out of range of one transmitter, the next transmitter is turned on in a precisely synthronized manner so as to enable said airfoil means to be continuously accelerated sequentially through the atmosphere by a series of transmitters until said airfoil means reaches a desired terminal velocity.

74. An apparatus as set forth in claim **65** further comprising aerodynamic breaking means mounted on said airfoil means for decelerating said airfoil means when said airfoil means is traversing through the atmosphere so that said airfoil means can be returned to the earth's surface.

75. An apparatus as set forth in claim **61** wherein said superconducting coil means is a self-supporting dipole with an aspect ratio exceeding 100.

76. An apparatus for generating propulsive thrust comprising:

superconducting current carrying coil means generating a magnetic field;

means for moving said coil means through the atmosphere such that atmospheric air flows through said coil means; and

means for creating a secondary repulsive magnetic field behind said moving coil means by free electrons in said atmospheric air moving in circular electron cyclotron orbits in the magnetic field of said coil means energized by a microwave beam tuned to the electron cyclotron resonant frequency of said orbiting electrons.

77. An apparatus as set forth in claim **76** wherein said microwave beam is transmitted from a phased array transmitter located on the earth's surface.

78. An apparatus as set forth in claim **76** further comprising:

circular wing-like airfoil means having a duct-like central region; and

means for mounting said superconducting coil means inside said airfoil means.

79. An apparatus as set forth in claim **78** wherein said circular airfoil means is moved through the atmosphere such that atmospheric air flows through said duct-like central region and through said magnetic field generated by said coil means.

80. An apparatus as set forth in claim **76** wherein said superconducting coil means comprises a plurality of superconducting coils further comprising:

a like plurality of circular wing-like airfoils having a duct-like central region;

means for mounting said coils inside said airfoils; and

means for maintaining said circular airfoils in a coaxial, spaced-apart relationship by a plurality of radial supporting airfoils mounted between adjacent circular airfoils.

81. An apparatus as set forth in claim **79** further comprising:

cryogenic cooling means; and

means for mounting said cryogenic cooling means inside said airfoil means for maintaining said superconducting coil means at cryogenic temperature.

82. An apparatus as set forth in claim **79** further comprising:

movable aerodynamic control surfaces mounted on said circular airfoil means;

actuator means for moving said movable control surfaces mounted inside said airfoil means;

guidance and navigation computer means mounted inside said airfoil means; and

electrical conduit means connecting said computer means and said actuator means enabling said airfoil means to be guided automatically through the atmosphere along a pre-calculated flight path.

83. An apparatus as set forth in claim **79** further comprising means for generating supersonic shock waves mounted inside said duct-like central region of said airfoil means such that when said airfoil means is moving at supersonic speed, a portion of said atmospheric air flowing through said duct-like central region is ionized by said shock waves.

84. An apparatus as set forth in claim **79** wherein said duct-like central region establishes a longitudinal central axis further comprising means for mounting a payload carrying central body in the center of said airfoil means extending along said longitudinal axis wherein the transverse cross sectional area of said central body is significantly less than the transverse cross sectional area of said airfoil duct so as to not restrict the atmospheric air flowing through said duct.

85. An apparatus as set forth in claim **84** wherein said central body has a length less than the diameter of said airfoil means and further comprising means for rotating said central body about an axis perpendicular to said longitudinal axis when said airfoil means is moving outside the earth's atmosphere thereby allowing said central body to be aligned along any desired direction while said coil means is constrained by magnetic torque exerted by the earth's natural magnetic field.

86. An apparatus as set forth in claim **85** further comprising attitude control and maneuvering propulsion means mounted inside said airfoil means and said central body for providing attitude control and maneuvering propulsion when said airfoil means is moving outside the earth's atmosphere.

87. An apparatus as set forth in claim **79** further comprising:

means for creating a static magnetic field on the earth's surface that extends upward into the atmosphere; and

means for launching said airfoil means from the earth's surface to some initial altitude at some initial velocity by magnetic repulsive forces exerted on said coil means mounted inside said airfoil means by said static magnetic field such that when said airfoil means reaches said initial altitude, said microwave beam is turned on from a microwave transmitter located on the earth's surface thereby accelerating said airfoil means to higher velocity.

88. An apparatus as set forth in claim **87** furhter comprising a plurality of spaced-apart microwave transmitters located on the earth's surface such that when said airfoil means moves out of range of one transmitter, the next transmitter is turned on in a precisely synchronized manner so as to enable said airfoil means to be continuously accelerated sequentially through the upper atmosphere by a series of transmitters until said airfoil means reaches a desired terminal velocity.

89. An apparatus as set forth in claim **84** further comprising aerodynamic breaking means mounted on said airfoil means and on said central body for decelerating said airfoil means when said airfoil means is traversing through the atmosphere so that said airfoil means can be returned to the earth's surface.

90. An apparatus as set forth in claim **76** wherein said superconducting coil means comprises a plurality of self-supporting dipoles having aspect ratios exceeding 100.

91. An apparatus for generating propulsive thrust comprising:

superconducting coil means for generating a magnetic field;

means for passing ionized gas through said magnetic field;

a microwave transmitter located on the earth's surface;

means for transmitting a microwave beam from said microwave transmitter into said magnetic field with a frequency tuned to produce electron cyclotron resonance with free electrons of said ionized gas moving in said magnetic field thereby accelerating said ionized gas away from said coil means and generating propulsive thrust.

92. An apparatus as set forth in claim **91** wherein said ionized gas is atmospheric air.

93. An apparatus as set forth in claim **91** further comprising:

circular wing-like airfoil means having a duct-like central region; and

means for mounting said superconducting coil means inside said airfoil means.

94. An apparatus as set forth in claim **93** further comprising:

cryogenic cooling means; and

means for mounting said cryogenic cooling means inside said airfoil means for maintaining said superconducting coil means at cryogenic temperature.

95. An apparatus as set forth in claim **93** wherein said circular airfoil means is moved through the atmosphere such that atmospheric air flows through said duct-like central region and through said magnetic field.

96. An apparatus as set forth in claim **95** wherein said superconducting coil means comprises a plurality of superconducting coils further comprising:

A like plurality of circular wing-like airfoils having a duct-like central region;

means for mounting said coils inside said airfoils; and

means for maintaining said circular airfoils in a coaxial spaced-apart relationship by a plurality of radial airfoils mounted between adjacent circular airfoils.

97. An apparatus as set forth in claim **95** further comprising means for ionizing a portion of said atmospheric air flowing through said duct-like central region.

98. An apparatus as set forth in claim **95** wherein said microwave transmitter is an electrically steered phased array.

99. An apparatus as set forth in claim **95** further comprising:

movable aerodynamic control surfaces mounted on said wing-like airfoil means;

actuator means for moving said movable control surfaces mounted inside said airfoil means;

guidance and navigation computer means mounted inside said airfoil means; and

electrical conduit means connecting said computer means with said actuator means enabling said airfoil means to be guided automatically through the atmosphere along a pre-calculated flight path.

100. An apparatus as set forth in claim **95** wherein said duct-like region establishes a longitudinal central axis further comprising means for mounting a payload carrying central body in the center of said airfoil means extending along said longitudinal axis wherein the transverse cross sectional area of said central body is signficantly less than the transverse cross sectional area of said airfoil duct so as to not restrict the atmospheric air flowing through said duct.

101. An apparatus as set forth in claim **100** wherein said central body has a length less than the diameter of said airfoil means and further comprises means for rotating said central body about an axis perpendicular to said longitudinal axis when said airfoil means is moving outside the earth's atmosphere thereby allowing said central body to be aligned along any desired direction while said coil means is constrained by magnetic torque exerted by the earth's natural magnetic field.

102. An apparatus as set forth in claim **101** further comprising attitude control and maneuvering propulsion means mounted inside said airfoil means and inside said central body for providing attitude control and maneuvering propulsion when said airfoil means is moving outside the earth's atmosphere.

103. An apparatus as set forth in claim **95** further comprising means for generating supersonic shock waves mounted inside said duct-like central region of said airfoil means such that when said airfoil means is moving at supersonic speed, a portion of said atmospheric air flowing through said duct-like central region is ionized by said shock waves.

104. An apparatus as set forth in claim **95** further comprising:

means for creating a static magnetic field on the earth's surface that extends upward into the atmosphere; and

means for launching said airfoil means from the earth's surface to some initial altitude at some initial velocity by magnetic repulsive forces exerted on said coil means mounted inside said airfoil means by said static magnetic field such that when said airfoil means reaches said initial altitude said microwave beam is turned on from a microwave transmitter located on the earth's surface thereby accelerating said airfoil means to higher velocity.

105. An apparatus as set forth in claim **104** further comprising a plurality of spaced-apart microwave transmitters located on the earth's surface such that when said airfoil means moves out of range of one transmitter, the next transmitter is turned on in a precisely synchronized manner so as to enable said airfoil means to be continuously accelerated sequentially through the upper atmosphere by a series of microwave transmitters until said airfoil means reaches a desired terminal velocity.

106. An apparatus as set forth in claim **105** wherein said central duct-like region of said airfoil means has a radius exceeding 100 m and wherein said microwave beams transmitted by said transmitters have beam powers exceeding 100 GW and further comprising cable means attached to said airfoil means adapted for accelerating a toroidal space station having a maor radius greater than said duct radius, from the earth's surface into orbit.

107. An apparatus as set forth in claim **95** further comprising aerodynamic breaking means mounted on said airfoil means for decelerating said airfoil means when said airfoil means is traversing through the atmosphere.

108. An apparatus as set forth in claim **91** wherein said superconducting coil means comprises a plurality of self-supporting dipoles having aspect ratios exceeding 100.

109. A high altitude ramjet propulsion system comprising:

superconducting coil means for generating a primary magnetic field;

circular duct-like airfoil means;

means for mounting said superconducting coil means inside said duct-like airfoil means;

a microwave transmitter located on the earth's surface; and

means for transmitting a microwave beam from said transmitter into said primary magnetic field with a frequency designed to produce electron cyclotron resonance with free electrons moving in said primary magnetic field thereby creating a secondary repulsive magnetic field behind said primary magnetic field and generating propulsive thrust.

110. An electromagnetic accelerator for accelerating a body to high velocity comprising:

movable superconducting coil means for generating a primary magnetic field;

means for attaching said body to said movable superconducting coil means;

microwave transmitter means mounted on the earth's surface; and

means for creating a travelling repulsive magnetic field behind said superconducting coil means within the earth's atmosphere by a microwave beam transmitted from said transmitter means thereby accelerating said coil means by magnetic repulsive forces.

111. An electromagnetic accelerator as set forth in claim **110** wherein said means for creating said travelling repulsive magnetic field behind said movable superconducting coil means comprises transmitting said microwave beam into said primary magnetic field with an electron cyclotron resonant frequency, such that free electrons from the atmosphere moving through said primary magnetic field are pumped into high energy electron cyclotron resonant orbits which, by virtue of their circular paths, become magnetic dipoles creating said travelling repulsive magnetic field.

112. An electromagnetic accelerator set forth in claim **111** further comprising:

circular wing-like airfoil means having an open circular central duct-like region; and

means for mounting said movable superconducting coil means inside said airfoil means such that said coil means surrounds said open central duct-like region and such that said primary magnetic field passes through said duct-like region.

113. An electromagnetic accelerator as set forth in claim **112** wherein said circular airfoil means is moved through the atmosphere such that atmospheric air flows through said central duct-like region and through said primary magnetic field.

114. An electromagnetic accelerator as set forth in claim **113** further comprising cryogenic cooling means mounted inside said airfoil means for maintaining said superconducting coil means at cryogenic temperature.

115. An electromagnetic accelerator as set forth in claim **113** further comprising:

movable aerodynamic control surfaces mounted on said airfoil means;

actuator means for moving said movable control surfaces mounted inside said airfoil means;

guidance and navigation computer means mounted inside said airfoil means; and

electrical conduit means connecting said computer means with said actuator means enabling said airfoil means to be guided automatically through the atmosphere along a pre-calculated flight path.

116. An electromagnetic accelerator as set forth in claim **113** wherein said open circular central duct-like region defines a central longitudinal axis and wherein said body has a length less than the diameter of said duct-like central region and a transverse cross sectional area less than 10% of the transverse cross sectional area of said duct-like region and wherein said mounting means comprises:

a beam extending along a diameter of said open central duct-like region perpendicular to said longitudinal axis connected to said airfoil means; and

means for mounting said body on said beam such that the center of mass of said body passes through said beam.

117. An electromagnetic accelerator as set forth in claim **116** wherein said body defines a longitudinal central axis that is perpendicular to said beam further comprising:

means for rolling said body around its longitudinal central axis; and

means for rotating said body around said beam such that said longitudinal axis of said body can be pointed in any direction independent of the orientation of said airfoil means.

118. An electromagnetic accelerator as set forth in claim **117** further comprising attitude control and maneuvering propulsion means mounted inside said airfoil means and inside said central body for providing attitude control and maneuvering propulsion when said airfoil means is moving outside the earth's atmosphere.

119. An electromagnetic accelerator as set forth in claim **113** further comprising means for generating supersonic shock waves mounted inside said duct-like central region of said airfoil means such that when said airfoil means is moving at supersonic speed, a portion of said atmospheric air flowing through said duct-like region is ionized by said shock waves.

120. An electromagnetic accelerator as set forth in claim **113** further comprising:

means for creating a static magnetic field on the earth's surface that extends upward into the atmosphere; and

means for launching said airfoil means from the earth's surface to some initial altitude at some initial velocity by magnetic repulsive forces exerted on said coil means mounted inside said airfoil means by said static magnetic field such that when said airfoil means reaches said initial altitude, said microwave beam is turned on thereby creating said travelling repulsive magnetic field behind said airfoil means accelerating said airfoil means to higher velocity.

121. An electromagnetic accelerator as set forth in claim **120** further comprising a plurality of spaced-apart microwave transmitters mounted on the earth surface such that when said airfoil means moves out of range of one transmitter, the next transmitter is turned on in a precisely synchronized manner so as to enable said airfoil means to be accelerated sequentially through the upper atmosphere by a series of microwave transmitters until said airfoil means reaches a desired terminal velocity.

122. An electromagnetic accelerator as set forth in claim **113** further comprising aerodynamic breaking means mounted on said airfoil means for decelerating said airfoil means when said airfoil means is traversing through the atmosphere.

123. An electromagnetic accelerator as set forth in claim **113** wherein said superconducting coil means comprises a plurality of superconducting coils further comprising:

a like plurality of circular wing-like airfoils having a circular duct-like central region;

means for mounting said coils inside said circular airfoils such that said coils surround said open central duct-like region; and

means for maintaining said circular airfoils in a coaxial spaced-apart relationship by a plurality of radial airfoils mounted between adjacent circular airfoils.

124. An electromagnetic accelerator as set forth in claim **110** wherein said superconducting coil means comprises a plurality of self-supporting dipoles having aspect ratios exceeding 100.

125. An electromagnetic accelerator as set forth in claim **118** further comprising:

secondary electromagnetic propulsion means mounted inside said body energized by electric current; and

means for utilizing electric current stored inside said superconducting coil means as an energy source to power said secondary propulsion means.

126. An electromagnetic accelerator as set forth in claim **120** further comprising control and guidance means for landing said airfoil means back on the earth's surface while decelerating said airfoil means by magnetic repulsive forces generated by said static magnetic field.

127. An electromagnetic accelerator for accelerating a body to high velocity above the earth's surface comprising:

movable superconducting coil means for generating a primary magnetic field;

means for attaching said body to said movable superconducting coil means;

microwave transmitter means mounted on the earth's surface; and

means for generatig a travelling secondary repulsive magnetic field behind said superconducting coil means by a microwave beam transmitted from said transmitter means.

128. An electromatnetic accelerator as set forth in claim **127** wherein said means for generating said travelling repulsive magnetic field comprises accelerating free electrons from ionized gas moving in said primary magnetic field into circular cyclotron orbits by said microwave beam operating with an electron cyclotron resonant frequency.

129. An electromagnetic accelerator as set forth in claim **128** wherein said ionized gas is ionized atmospheric air.

130. An electromagnetic accelerator as set forth in claim **129** wherein said microwave transmitter means comprises a series of spaced-apart microwave transmitters mounted on the earth's surface further comprising switching means such that when said coil means moves out of range of one transmitter, said energizing microwave beam is transmitted from the next transmitter in the series so that said coil means can be accelerated over a long accelerating distance.

131. An apparatus for generating propulsive thrust comprising:

superconducting coil means for generating a primary magnetic field; and

means for generating a secondary repulsive magnetic field behind said superconducting coil means by a microwave beam.

132. An apparatus as set forth in claim **131** wherein said means for generating said secondary repulsive magnetic field comprises accelerating free electrons from ionized gas moving in said primary magnetic field into circular cyclotron orbits by said microwave beam operating with an electron cyclotron resonant frequency.

133. An apparatus as set forth in claim **132** wherein said ionized gas is ionized atmospheric air.

* * * * *

United States Patent [19]

Logan et al.

[11] **Patent Number: 5,043,739**

[45] **Date of Patent: Aug. 27, 1991**

[54] **HIGH FREQUENCY RECTENNA**

[75] Inventors: **B. Grant Logan**, Danville; **William J. Orvis**, Livermore, both of Calif.

[73] Assignee: **The United States of America as represented by the United States Department of Energy**, Washington, D.C.

[21] Appl. No.: **472,548**

[22] Filed: **Jan. 30, 1990**

[51] **Int. Cl.**^{5} **H01Q 1/26**

[52] **U.S. Cl.** **343/701**; 343/893

[58] **Field of Search** .... 343/701, 753, 700 MS, 343/846, 893

[56] **References Cited**

U.S. PATENT DOCUMENTS

2,855,543	4/1953	White	315/34
3,114,517	12/1963	Brown	244/1
3,434,678	5/1965	Brown	244/1
3,535,543	10/1970	Dailey	307/149
3,544,998	12/1970	Vandenplas et al.	343/701
3,711,848	1/1973	Martens	343/701
3,789,471	2/1974	Spindt et al.	357/6
3,852,755	12/1974	Works et al.	343/701
4,008,477	2/1977	Babij et al.	343/701
4,079,268	3/1978	Fletcher et al.	343/700 MS
4,155,155	5/1979	Bourdon et al.	29/590
4,516,037	5/1985	Shackle	307/252 A
4,587,545	5/1986	Berthold et al.	357/38
4,721,885	1/1988	Brodie	313/576
4,853,703	8/1989	Murakami et al.	343/701
4,853,705	8/1989	Landt	343/701
4,855,636	8/1989	Busta et al.	313/306
4,955,562	9/1990	Martin et al.	244/62

OTHER PUBLICATIONS

"The History of Power Transmission by Radio Waves", by William C. Brown, IEEE Transactions on Microwave Theory and Techniques, vol. MIT-32, No. 9, 9/1984.

"Rectenna Technology Program: Ultra Light 2.45 GHz Rectenna and 20 GHz Rectenna", by William C. Brown, Prepared for National Aeronautics and Space Administration, NASA Lewis Research Center, Contract NAS3-22764.

"Initiative for the 21st Century: Advanced Space Power and Propulsion Based on Lasers", by B. G. Logan, Lawrence Livermore National Laboratory Report No. UCRL-98520.

Primary Examiner—Michael C. Wimer
Assistant Examiner—Hoanganh Le
Attorney, Agent, or Firm—H. P. Sartorio; L. E. Carnahan; William R. Moser

[57] **ABSTRACT**

The invention provides an inexpensive array of rectifying antennas which employ field emission diodes for rectifying electromagnetic waves of microwave frequencies and higher frequencies.

15 Claims, 5 Drawing Sheets

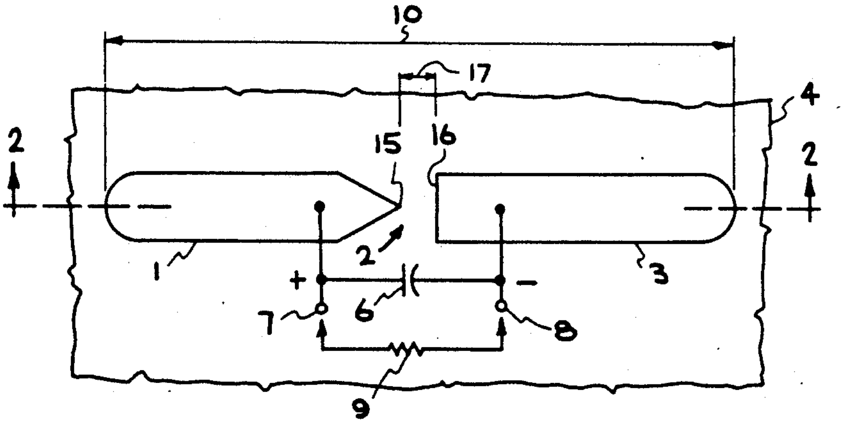

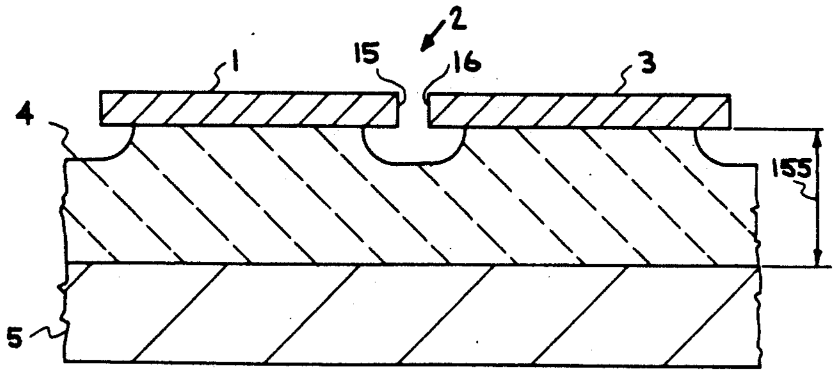

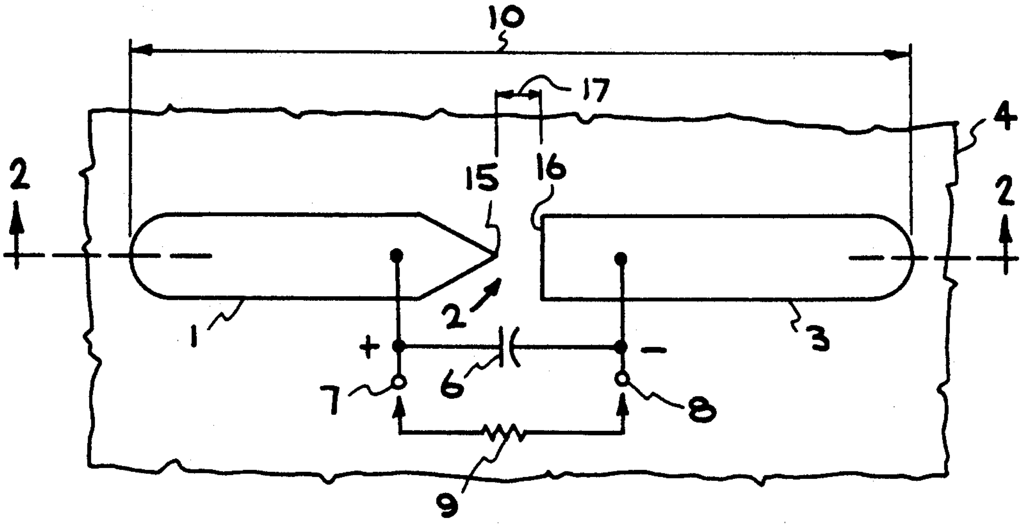
10
17
4
2
15
16
2
2
1
3
+
−
6
7
8
9

FIG. 1

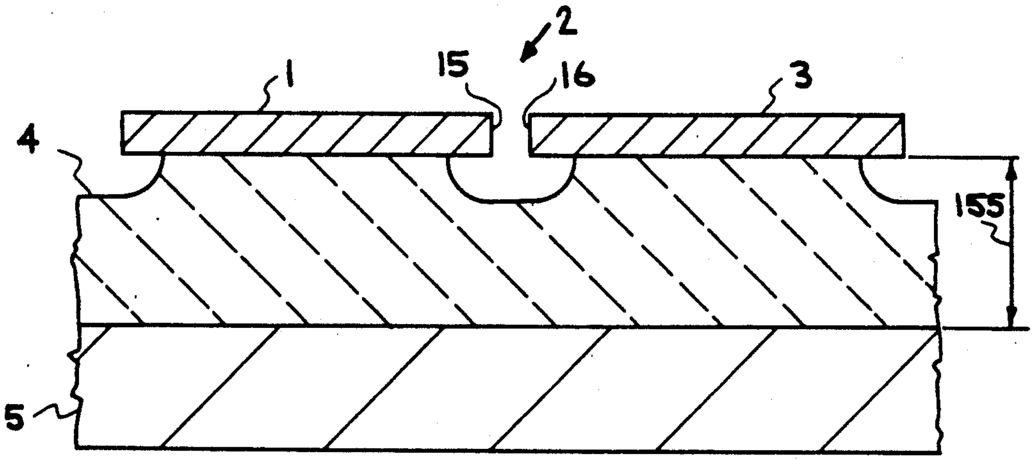
2
15
16
1
3
4
155
5

FIG. 2

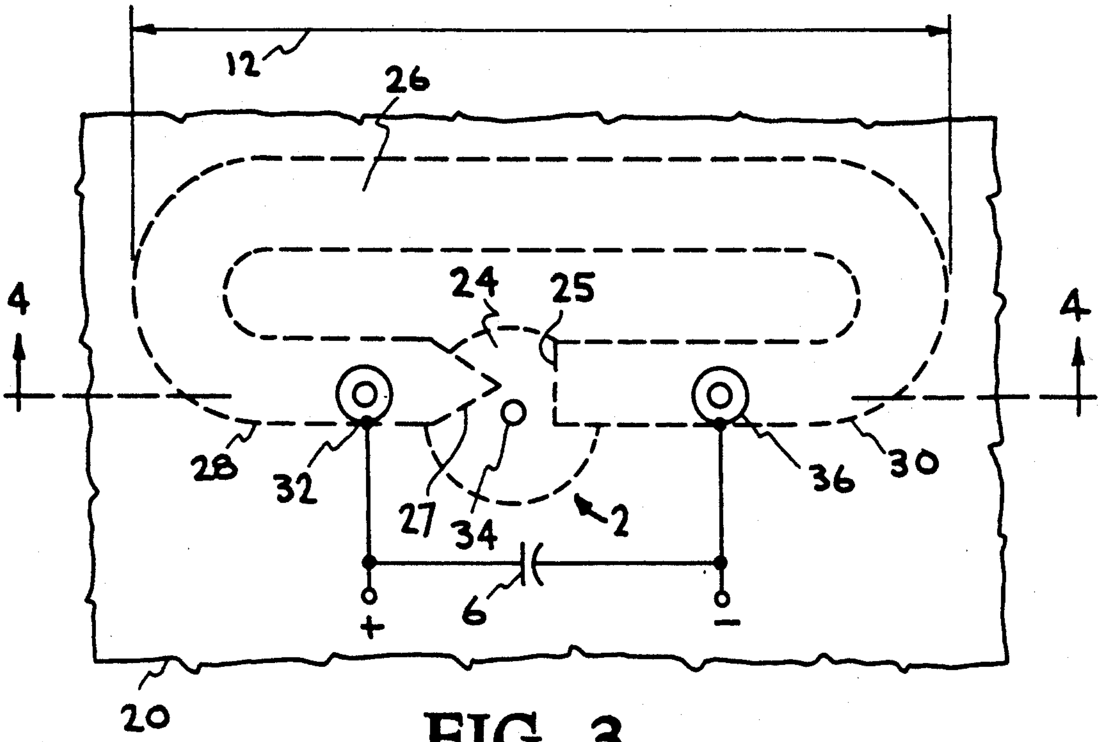
+
−

FIG. 3

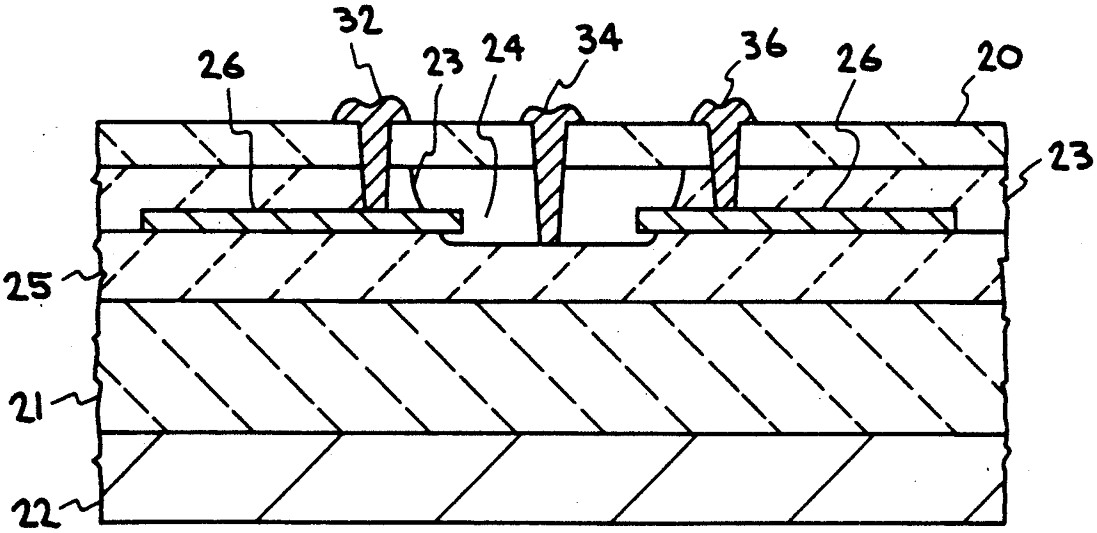

FIG. 4

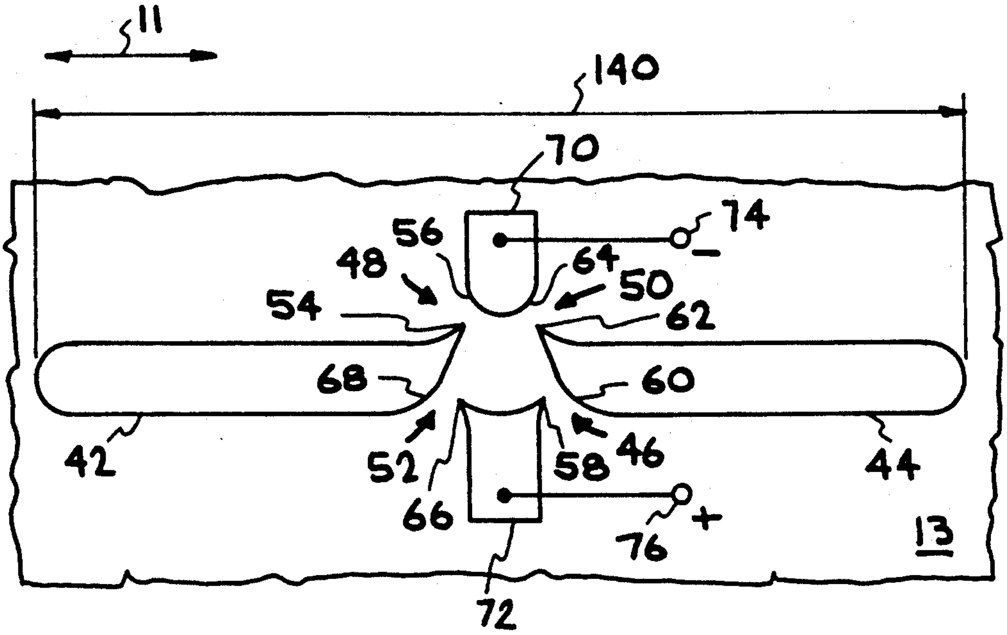
–
+

FIG. 5

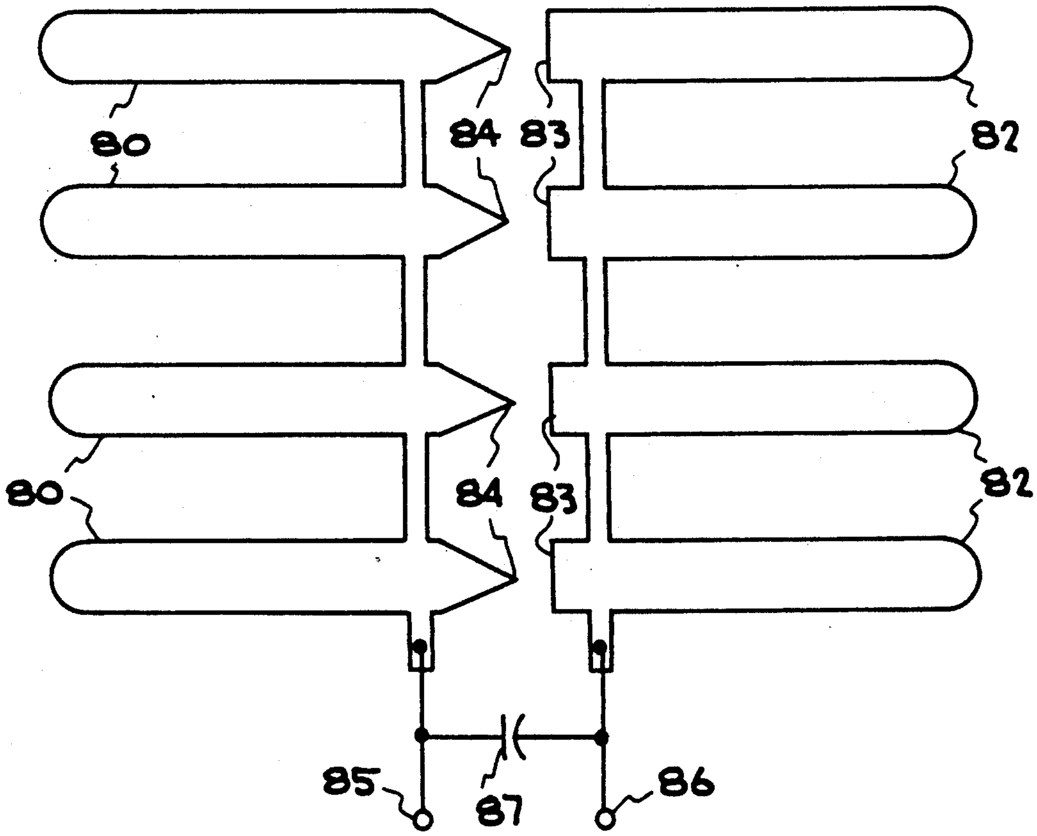

FIG. 6

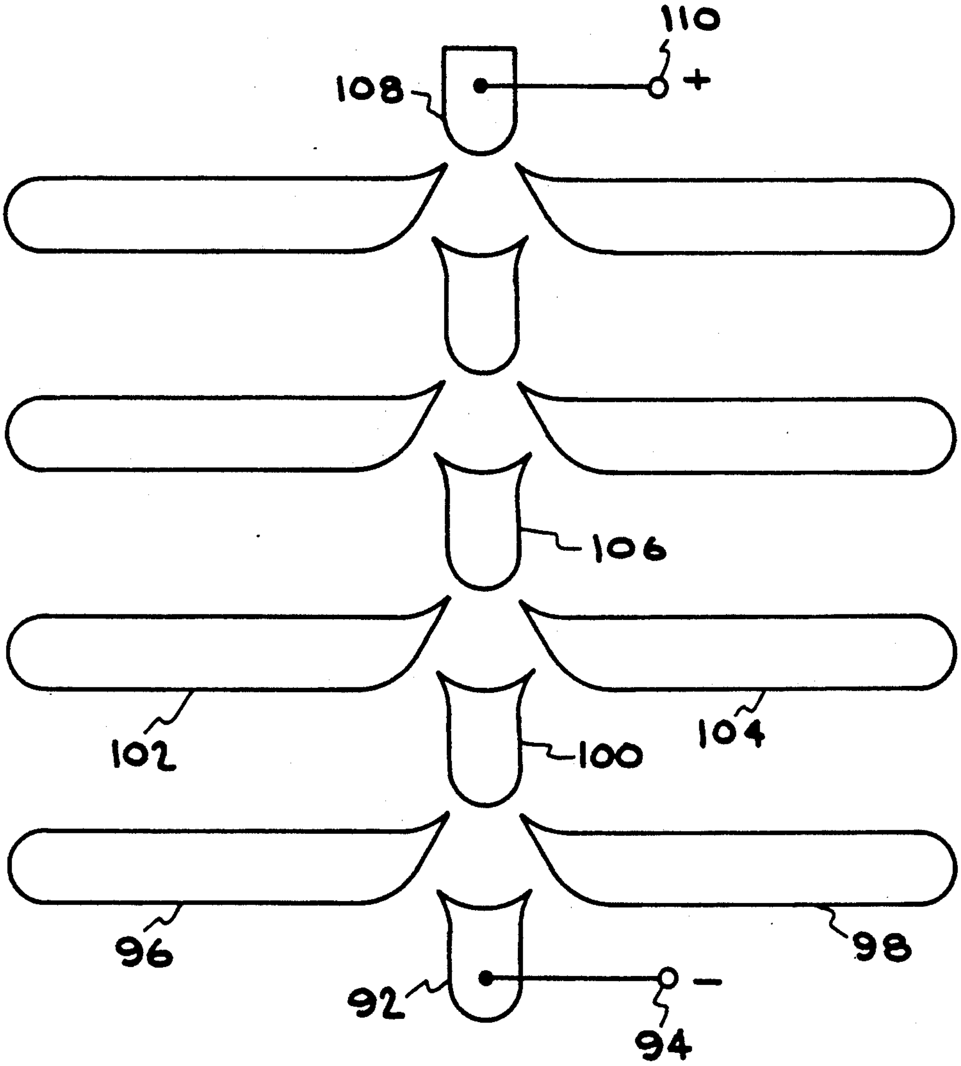
110
108
+
106
102
104
100
96
98
92
−
94

FIG. 7

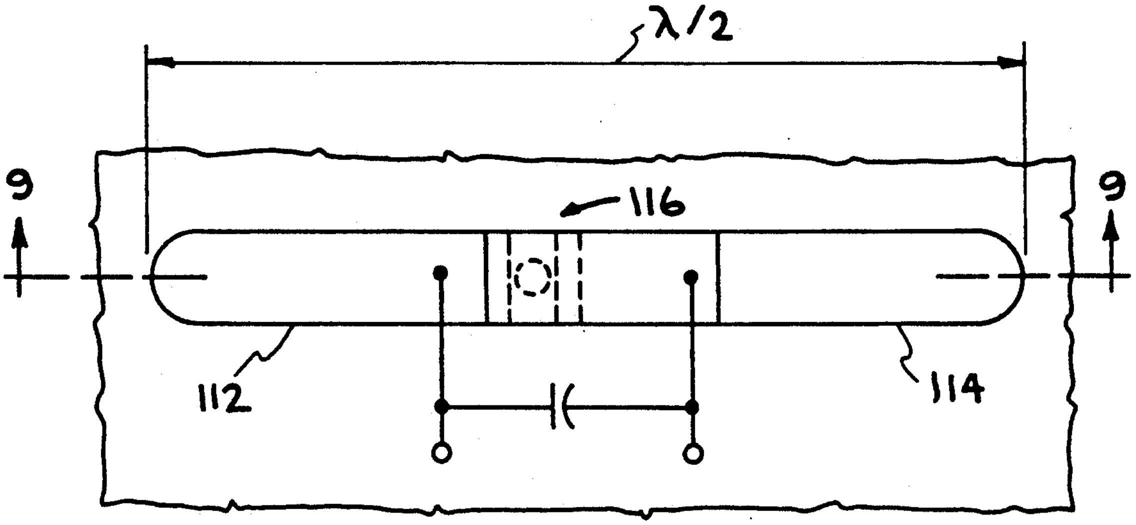
λ/2
9
9
116
112
114

FIG. 8

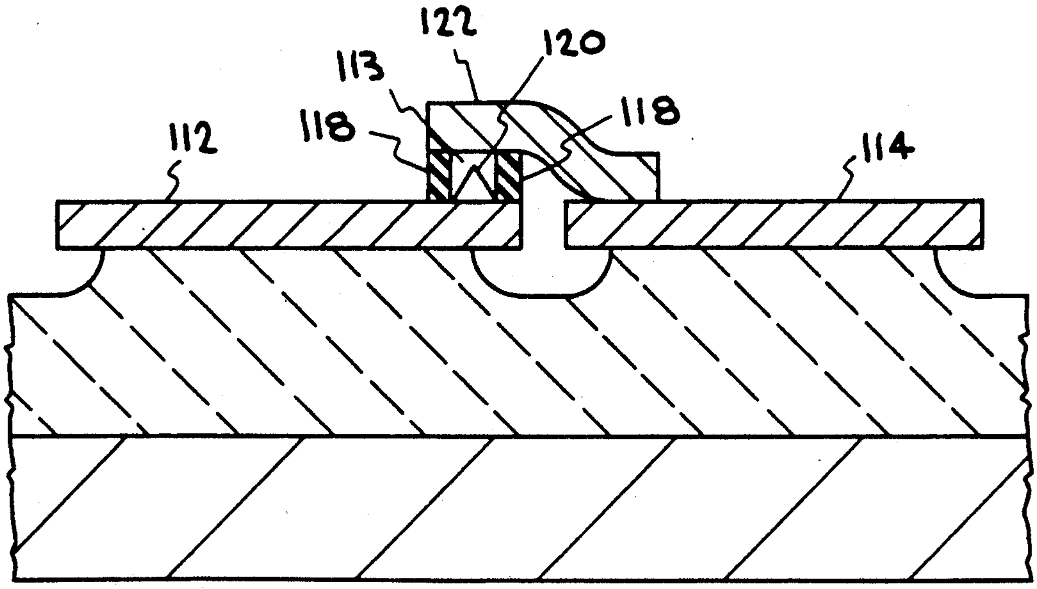
122
120
113
118
118
112
114

FIG. 9

HIGH FREQUENCY RECTENNA

BACKGROUND OF THE INVENTION

The United States Government has rights in this invention pursuant to Contract No. W-7405-ENG-48 between the U.S. Department of Energy and the University of California, for the operation of Lawrence Livermore National Laboratory.

A rectenna, or rectifying antenna, comprises an antenna and a diode rectifier. In operation, the antenna gathers electromagnetic energy at a frequency consistent with its size and shape, and the diode rectifies that energy into D.C. current.

Presently, rectennas utilize a semiconductor diode or large vacuum diode to rectify electromagnetic waves, with frequencies ranging up to the radio frequency spectrum, to direct current. Rectennas utilizing a semiconductor diode have an upper frequency limit and an upper intensity limit due to the inherent limitations of semiconductor diodes. These limitations include the device capacitance and the fragility of high speed semiconductor devices. Rectennas using large vacuum tubes also have upper frequency limits and are expensive to build. Large arrays of rectennas using conventional vacuum tubes would be prohibitively expensive to build, and could not be produced in a compact form.

William C. Brown in "The History of Power Transmission of the Rectenna," in the IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-32, No. 9, September 1984, describes the development of rectennas for rectifying electromagnetic signals with a frequency of 2.45 Giga Hertz (GHz). The article describes research into an apparatus that would be able to allow power to be sent from the earth to a device above the earth or to allow power to be sent from above the earth to the earth via a 2.45 GHz radio signal. The author states that the best diodes that he found for the rectennas are either silicon or gallium arsenide (GaAs) Schottky-barrier diodes.

William C. Brown in "Rectenna Technology Program: Ultra Light 2.45 GHz Rectenna and 20 GHz Rectenna," published by NASA as PT-6902, CR 179558, Mar. 11, 1987 describes the development of a 2.45 GHz rectenna and proposes a 20 GHz rectenna. The article first describes the development of a 2.45 GHz rectenna. The GaAs Schottky barrier diodes for the 2.45 GHz rectenna have specifications including a zero bias capacitance of 3 pf and a reverse breakdown voltage of 60–70 volts. These specifications and other qualities of the Schottky barrier diodes limit the efficiency and the maximum power capabilities of these rectennas. The article goes on to propose a 20 GHz and higher frequency rectenna. The article states that the losses in the diode will be considerably greater at these higher frequencies and the losses in the scaled circuit structure will become greater by the square root of the frequency scaling factor because of skin losses. Because semiconductor diodes break down at high temperatures, semiconductor diodes must be cooled. In the design proposed in the article for a 20 GHz rectenna, reasonable cooling is provided if the rectenna is limited to a DC power output of 0.4 watts per diode element with an output voltage of 8.9 volts and a reverse voltage breakdown of 20 volts. At higher intensities, the resistance in the rectenna will create more heat elevating the semiconductor temperature and causing the semiconductor to breakdown. These parameters help illustrate the inefficiency and fragileness of semiconductor diodes in a rectenna.

Because of the inherent limitations of semiconductor rectennas and large vacuum tube rectennas, there is a need for rectennas that can rectify high frequency and high intensity electromagnetic waves.

The invention provides an array of rectennas, which are able to rectify high frequency microwave or high frequency and high intensity electromagnetic radiation. These arrays may be useful in transportation of power through electromagnetic waves and conversion of intense high frequency electromagnetic radiation into direct current. Applications of the invention include sending power to satellites or high altitude devices from the earth by electromagnetic waves, transmission of power from satellites which collect solar power and convert it to electromagnetic waves which are transmitted to the earth, and direct conversion of microwaves generated by fusion reactions into a direct current.

SUMMARY OF THE INVENTION

It is an object of the invention to provide a rectenna which can rectify electromagnetic radiation with wavelengths equal to or shorter than the wavelengths of microwaves.

It is another object of the invention to provide a rectenna which can rectify electromagnetic waves at high intensities.

It is another object of the invention to provide a rectenna which can rectify electromagnetic radiation with wavelengths equal to or shorter than the wavelengths of microwaves and with high intensities, and with an efficiency greater than 40%.

It is another object of the invention to provide a rectenna which is resilient to high energy radiation and high temperatures.

It is another object of the invention to provide an array of rectennas which can be manufactured in large numbers at a low cost.

It is another object of the invention to provide a compact array of rectennas.

Additional objectives, advantages and novel features of the invention will be set forth in part in the description which follows, and in part will become apparent to those skilled in the art upon examination of the following or may be learned by practice of the invention. The objects and advantages of the invention may be realized and attained by means of the instrumentalities and combinations particularly pointed out in the appended claims.

The invention comprises an antenna, comprising a first half and a second half, and at least one field emission vacuum diode electrically connected between the first half and the second half of the antenna to provide a rectenna. The diode is made an integral part of the antenna so that arrays of rectennas can be used for rectifying electromagnetic radiation with a frequency higher than or equal to microwave frequencies. If the diode was not integral to the antenna, and if an array of antennas was connected together, then phase differences from one side of the array to the other would cause the combined outputs to cancel, significantly decreasing the output. By making the diode an integral part of the antenna, the current is rectified at the source, and phase differences will have little effect on the combined currents.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated and form a part of the specification, illustrate several embodiments of the invention and, together with the description, serve to explain the principles of the invention.

FIG. **1** is a top view of a preferred embodiment of the invention. It illustrates a half wave rectenna using a dipole antenna and a single lateral field emission diode.

FIG. **2** is a cross-sectional view of a half wave rectenna illustrated in FIG. **1** along lines **2**.

FIG. **3** is a top view of another embodiment, a half wave rectenna using a folded dipole antenna and a single lateral field emission diode.

FIG. **4** is a cross-sectional view of a half wave rectenna illustrated in FIG. **3** along lines **4**.

FIG. **5** is a top view of another embodiment, a full wave rectenna using a dipole antenna and a full wave bridge rectifier using four lateral field emission diodes.

FIG. **6** is a top view of an array of half wave rectennas in parallel.

FIG. **7** is a top view of an array of full wave rectennas in series.

FIG. **8** is a top view of another embodiment, a half wave rectenna using a dipole antenna and a single vertical field emission diode rectifier.

FIG. **9** is a cross-sectional view of a half wave rectenna illustrated in FIG. **8** along lines **9**.

DESCRIPTIONS OF THE PREFERRED EMBODIMENTS

FIGS. **1** and **2** show a top and side view of a preferred embodiment of the invention. The invention receives and rectifies an electromagnetic wave with wavelength L. It includes a dipole antenna (**1** and **3**), coupled with an integral lateral field emission diode **2** at the center, and an external capacitor **6**. In the preferred embodiment the total length **10** of the dipole antenna is approximately L/2.

The embodiment may be constructed using conventional semiconductor manufacturing techniques on an insulating substrate **4** attached to a conducting ground plane **5** (FIG. **2**). The ground plane makes the antenna directional, with the vertical direction being the preferred direction. The separation **155** between the antenna and the ground plane is preferably approximately L/4 to provide a maximum reflection of electromagnetic radiation which passes through the antenna **1** and **3** to the ground plane. The actual distance would depend on the material used as the insulating substrate.

Using conventional semiconductor techniques, a layer of conductive material (tungsten, or doped polysilicon for example) is deposited on the insulator and then shaped by photolithography and wet or dry etching. The insulator under the layer of conductive material is then etched, undercutting the conductor. This undercutting frees the edge of the conductor to prevent surface leakage currents and surface breakdown along the insulator between the two sides of the antenna.

The method of making a field emission vacuum triode is disclosed in U.S. Pat. No. 4,855,636 by Busta et al, which is incorporated by reference. Using the method described in Busta et al. a field emission diode in a vacuum could be created. By using a differently shaped mask the field emission diode can be integrated into a dipole antenna. Even though Busta discloses the creation of a triode, the field emission diode could be created in a similar manner by either sealing the etching holes described in Busta with an insulating material or by not applying a charge to the gate of the triode. For cases where the mean free path of an electron in a gas at some pressure is comparable to the separation of the anode and cathode, the diode will operate in that gas at that pressure as if it were in a vacuum. In addition, a low pressure or near vacuum can also be established around the field emission diode by sealing the entire rectenna in a near vacuum container.

Electromagnetic radiation of wavelength L is absorbed by the dipole antenna **1** and **3**. According to well known principles of antenna design, the electromagnetic radiation will set up an oscillating potential difference between the first half **1**, and the second half **3** of the dipole antenna. The potential difference will oscillate with the same frequency as the absorbed electromagnetic radiation. This oscillating potential difference will manifest itself across the lateral field emission diode **2** whose cathode, manifested by cathode tip **15** which is a pointed edge of the first half **1** of the antenna, is integral with the first half **1** of the antenna, and whose anode, manifested by anode surface **16** which is a nonpointed edge of the second half **3** of the antenna, is integral with the second half **3** of the antenna. The pointed edge making the cathode tip **15** is located on a part of the first half **1** of the antenna that is located closest to the second half **3** of the antenna as shown. The nonpointed edge making the anode surface **16** is located on a part of the second half **3** of the antenna that is located closest to the first half **1** of the antenna.

During the positive half cycle of the electromagnetic radiation, the second half **3** of the antenna is positive with respect to the first half **1** of the antenna, and field emission of electrons occurs off of the tip of the cathode **15**. The field emitted electrons will be absorbed at the anode surface **16**, causing a surplus of electrons in the second half **3** of the antenna. During the negative half cycle, the second half **3** of the antenna is negative with respect to the first half **1** of the antenna. Since the anode surface **16** is flat, it does not emit an appreciable amount of electrons and thus no field emission current flows in the negative half of the cycle. Thus, a voltage is applied to the capacitor **6**, and any load **9** attached to terminals **7** and **8**. This allows the surplus of electrons in the second half **3** of the antenna to flow through the attached load **9** back to the first half **1** of the antenna. The capacitor **6** smooths the voltage available at the terminals by charging from the antenna during the negative half cycle and discharging through the load **9** during the positive half cycle. This illustrates how the cathode tip **15** forming the cathode is integrally and electrically connected to the first half **1** of the antenna and how the anode is integrally and electrically connected to the second half **3** of the antenna and how a field emission diode **2** comprising the anode **16** and cathode **15** are electrically connected between the first half **1** and the second half **3** of the antenna.

In the preferred embodiment shown in FIG. **1**, the rectenna is designed to rectify an electromagnetic wave having a predetermined frequency between 1,000 GHz and 10 GHz. This would require that the separation **17** between the anode and cathode be less than 10 microns, and preferably less than 3 microns.

The rectenna is not limited to a dipole antenna design, nor to the half wave rectifier described above. FIG. **3** shows a similar structure, but using a folded dipole antenna **26** instead of the simple dipole antenna. The folded dipole antenna **26** typically has a higher impedance than the simple dipole, which may make a better coupling between the antenna and the diode. The folded dipole antenna **26** could be manufactured in the same way as the embodiment shown in FIG. **1**, having the same cross-sectional view. For illustrative purposes, the folded dipole antenna shown in FIG. **3** is made using the method described in U.S. Pat. No. 4,855,636 by Busta et al. FIG. **4** is a cross-sectional view of FIG. **3** along lines **4**. In this embodiment a silicon wafer **21** is attached to a conducting layer **22**. Preferably the silicon wafer **21** has a thickness equal to approximately ¼ the wavelength of the electromagnetic radiation to be rectified. As described in U.S. Pat. No. 4,855,636 approximately 1 micron of thermal silicon dioxide is grown on the silicon wafer **21** forming a first insulating layer **25**. A conducting layer is deposited on the first insulating layer **25** to form the folded dipole antenna **26** with a first half **28** and second half **30** of the antenna and an integrated anode **25** and cathode **27**. In the preferred embodiment the length **12** of the folded dipole antenna **26** is approximately equal to L/2, where L is the wavelength of the electromagnetic radiation to be rectified. The entire structure is then covered by a layer of undoped, highly resistive polycrystalline silicon **23**. The upper surface of the polycrystalline silicon **23** is thermally oxidized to form a top insulating layer **20**. Windows are made through the top insulating layer **20** to allow the etching of a chamber **24** and holes for contacts **32** and **36**. A metallic layer is then applied over the top insulating layer **20** and then etched back to create contacts **32** and **36** and for creating a window plug **34**.

The folded dipole rectenna shown in FIGS. **3** and **4** rectifies electromagnetic waves in the same manner as the dipole rectenna described above. Using the described method of building the folded dipole rectenna, the integrated diode can be made so that the separation between the cathode tip **27** and the anode surface **25** would be in the above described range for rectifying signals with frequencies higher than 10 GHz. At an operating voltage of 500 volts it is estimated that the leakage current due to parasitic current through the undoped polycrystalline silicon layer **23** would be 500 pico Amps. This leakage could be further reduced by replacing the polycrystalline silicon side walls with silicon dioxide side walls. The integration of field emission diodes with an antenna in this embodiment provides a very efficient rectenna that can rectify signals with frequencies equal to or greater than microwaves. In addition, the provided diodes can easily handle 500 volts, wherein the Schottky barrier diodes used in the prior art had a reverse breakdown voltage of 60–70 volts.

FIG. **5** shows a rectenna with a dipole antenna, comprising a first strip of conductive material on a substrate **13** with a length along a first direction **11** less than L/2, where L is the wavelength of the electromagnetic radiation to be rectified, forming a first half **42** of the antenna and a second conductive strip on a substrate **13** with a length along the first direction **11** less than L/2 forming a second half **44** of the antenna, and a full wave bridge rectifier. The second conductive strip **44** is located along the first direction **11** with respect to the first conductive strip **42**. The rectifier utilizes four field emission diodes **46**, **48**, **50** and **52** arranged in a full wave bridge configuration. The configuration illustrated in FIG. **5** is manufactured in the same manner as the embodiment illustrated in FIG. **1**. It would also be possible for this embodiment to be manufactured in the same manner as the embodiment illustrated in FIG. **3**.

Electromagnetic radiation of wavelength L is absorbed by the dipole antenna **42** and **44**. According to well known principles of antenna design, incident electromagnetic radiation will set up an oscillating potential difference between the first half **42**, and the second half **44** of the dipole antenna. The potential difference will oscillate at the same frequency as the absorbed electromagnetic radiation. This oscillating potential difference will manifest itself across the lateral field emission diodes.

During the positive half cycle of the electromagnetic radiation, the second half **44** of the antenna is positive with respect to the first half **42** of the antenna, and field emission of electrons occurs off of the tip of a cathode tip **54** of a first diode **48**, wherein the cathode tip **54** is an integral part of the first half **42** of the antenna formed by a pointed edge of the first strip of conductive material which forms the first half **42** of the antenna. The field emitted electrons will be absorbed at the anode surface **56** of the first diode **48**, which is formed by a nonpointed edge of a third strip of conductive material forming an anode terminal **70**, causing a surplus of electrons in the anode terminal **70**. Also during the positive half cycle of the electromagnetic radiation, field emission of electrons occurs off of the tip of cathode **58** of a second diode **46**, wherein the cathode tip **58** is formed by a pointed edge of a fourth strip of conductive material on the substrate **13**, wherein the fourth strip of conductive material forms a cathode terminal **72**. The field emitted electrons will be absorbed at the anode surface **60** of the second diode **46**, wherein the anode surface **60** is an integral part of the second half **44** of the antenna, causing a surplus of electrons in the second half **44** of the antenna. During the negative half cycle, the second half **44** of the antenna is negative with respect to the first half **42** of the antenna, and field emission of electrons occurs off of the tip of the cathode tip **62** of a third diode **50**, wherein the cathode tip **62** is an integral part of the second half **44** of the antenna. The field emitted electrons will be absorbed at the anode surface **64** of the second diode **50**, which is part of the anode terminal **70**, causing a surplus of electrons in the anode terminal **70**. Also during the negative half cycle of the electromagnetic radiation, field emission of **5** electrons occurs off of the tip of cathode **66** of a fourth diode **52**, wherein the cathode tip **66** is part of the cathode terminal **72**. The field emitted electrons will be absorbed at the anode surface **68** of the fourth diode **52**, wherein the anode surface **68** is an integral part of the first half **42** of the antenna, causing a surplus of electrons in the first half **42** of the antenna.

Thus, during the full wavelength electrons flow from the cathode terminal **72** to either the first **42** or second **44** half of the antenna and then to the anode terminal **70**. The surplus of electrons gathered at the anode terminal **70** can flow through the negative load terminal **74** electrically connected to the anode terminal and then through a load to a positive load terminal **76** electrically connected to the cathode terminal **72**. The anode terminal **70** acts as an anode for both the first part **42** of the antenna and the second part **44** of the antenna, and the cathode terminal **72** acts as a cathode for both the first part **42** of the antenna and the second part **44** of the antenna. Although a capacitor is not needed, it may be used to smooth the current flow between terminals **74**, **76**.

In the preferred embodiment of the invention, the total length **140** of the first half **42** and the second half **44** of the antenna along the first direction **11** is approximately equal to L/2. As shown in FIG. **5**, the four rectifying field emission diodes are electrically connected between the first half **42** and the second half **44** of the antenna making two parallel sets of two diodes in series.

An advantage of the invention is that a large number of rectennas could be inexpensively combined in an array using conventional semiconductor manufacturing techniques. FIGS. **6** and **7** illustrate different arrays of rectennas. FIG. **6** shows a parallel array of the half wave rectennas shown in FIG. **1**. For clearer illustration of the embodiment, only the pattern of conductive strips is shown. Each half wave rectenna in the array has a first half **80** and a second half **82** of an antenna. Each first half **80** of an antenna has an integrally connected cathode tip **84**. Each second half **82** of an antenna has an integrally connected anode surface **86**. Each first half **80** of an antenna is electrically connected to a first load terminal **85** and a first half of a capacitor **87**. Each second half **82** of an antenna is electrically connected to a second load terminal **86** and a second half of the capacitor **87**. Each individual rectenna would function like the rectenna embodiment illustrated in FIG. **1**. Attaching them in a parallel configuration as illustrated would increase the current available at the terminals.

FIG. **7** shows a series array of the full wave rectennas shown in FIG. **5**. For clear illustration only the pattern of conductive strips is shown. The array comprises a first terminal **92** electrically connected to a first load terminal **94**. The electromagnetic signal to be rectified will cause electrons to flow from the first terminal **92** to the first half **96** of a first antenna and the second half **98** of the first antenna. The electromagnetic signal will also cause electrons to flow from the first half **96** of the first antenna and the second half **98** of the first antenna to a second terminal **100**. The second terminal **100** acts as an anode terminal to the first antenna. The electromagnetic signal will cause electrons to flow from the second terminal **100** to the first half **102** of a second antenna and the second half **104** of the second antenna. The electromagnetic signal will also cause electrons to flow from the first half **102** of the second antenna and the second half **104** of the second antenna to a third terminal **106**. The second terminal **100** acts as a cathode terminal to the second antenna, and the third terminal **106** acts as an anode terminal to the second antenna. The electrons continue to flow through the subsequent antenna elements and terminals until they reach a final terminal **108**, which is electrically connected to a second load terminal **110**. This series configuration will increase the voltage across the first **94** and second **110** load terminals.

In either the parallel or the series configurations, an external capacitor could be connected to the load terminals.

In addition to the lateral diode shown so far, a vertical field emission diode can also be used as the rectifier in a rectenna. FIGS. **8** and **9** show a half wave rectenna using a half wave vertical field emission diode **116** and a dipole antenna with a first half **112** and a second half **114**. The vertical field emission diode comprises an insulator **118** placed on the first half **112** of the antenna. In a hole **113**, formed in the insulator, a conductive cone or pyramid **120** is formed in electrical contact with the first **112** half of the antenna. A conducting cap **122** is placed over the hole **113**, that also makes electrical contact with the second half **114** of the antenna. The hole **113** can be evacuated and sealed, forming an integral vacuum diode. In operation this embodiment works in the same manner as the embodiment illustrated in FIGS. **1** and **2**, except that the vertical diode causes the electrons to be emitted vertically off of the tip of the cone or pyramid **120** instead of horizontally off of the point. The manufacturing of vertical field emission diodes is described in U.S. Pat. No. 4,721,885 by Brodie, and U.S. Pat. Nos. 3,755,704 and 3,789,471 by Spindt incorporated by reference.

The foregoing description of preferred embodiments of the invention have been presented for purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed, and obviously many modifications and variations are possible in light of the above teaching. The embodiments were chosen and described in order to best explain the principles of the invention and its practical application to thereby enable others skilled in the art to best utilize the invention in various embodiments and with various modifications as are suited to the particular use contemplated. It is intended that the scope of the invention be defined by the claims appended hereto.

We claim:

1. A rectifying antenna for rectifying electromagnetic radiation with a wavelength shorter than 3.0 centimeters, comprising:

an antenna having a first half and a second half; and

a first field emission diode for rectifying a current flow, wherein the first field emission diode has a first anode and first cathode electrically connected between the first half of the antenna and the second half of the antenna, and wherein the first cathode is separated from the first anode by a distance of less than 10 microns and wherein the antenna has a length which is approximately equal to half of the wavelength of the electromagnetic radiation to be rectified.

2. A rectifying antenna as claimed in claim **1**, further comprising:

a first load terminal electrically connected to the first half of the antenna; and

a second load terminal electrically connected to the second half of the antenna, and wherein the first cathode is electrically connected to the first half of the antenna, and the first anode is electrically connected to the second half of the antenna.

3. A rectifying antenna as claimed in claim **2**, wherein the first cathode is integral with the first part of the first antenna, and wherein said first cathode comprises a pointed edge of the first part of the antenna, and wherein the first anode is integral with the second part of the antenna, and wherein the first anode comprises a nonpointed edge of the second part of the antenna.

4. A rectifying antenna as claimed in claim **2**, wherein the first field emission diode is a linear field emission diode.

5. A rectifying antenna as claimed in claim **2**, wherein the first field emission diode is a vertical field emission diode.

6. A rectifying antenna as claimed in claim **1**, further comprising:

- a first terminal, wherein the first field emission diode is electrically connected between the first terminal and the second half of the antenna with the first anode being electrically connected to the first terminal and the first cathode being electrically connected to the second half of the antenna;
- a second terminal;
- a second field emission diode with a second cathode and second anode electrically connected between the first terminal and the first half of the antenna with the second anode being electrically connected to the first terminal and the second cathode being electrically connected to the first half of the antenna;
- a third field emission diode with a third cathode and a third anode electrically connected between the second terminal and the first half of the antenna with the third anode being electrically connected to the first half of the antenna and the third cathode being electrically connected to the second terminal; and
- a fourth field emission diode with a fourth cathode and a fourth anode electrically connected between the second terminal and the second half of the antenna with the fourth anode being electrically connected to the second half of the antenna and the fourth cathode being electrically connected to the second terminal.

7. A rectifying antenna as claimed in claim 6, wherein the first cathode is formed by a pointed edge of the second half of the antenna, and wherein the second cathode is formed by a pointed edge of the first half of the antenna, and wherein the first and second anodes are formed by nonpointed edges of the first terminal, and wherein the third anode is formed by a nonpointed edge of the first half of the antenna, and wherein the fourth anode is formed by a nonpointed edge of the second half of the antenna, and wherein the third and fourth cathodes are formed by pointed edges of the second terminal.

8. A rectifying antenna as claimed in claim 6, wherein the field emission diode is a lateral field emission diode.

9. A rectifying antenna for rectifying electromagnetic radiation with a wavelength shorter than 3.0 centimeters, comprising:

- a substrate comprising a layer of insulating material;
- a first cathode;
- a first anode;
- a first conductive strip on the substrate, with a length in a first direction, wherein the first conductive strip is integrally and electrically connected to the first cathode;
- a second conductive strip on the substrate, with a length in the first direction, wherein the second conductive strip is integrally and electrically connected to the first anode, and wherein the second conductive strip is located along the first direction with respect to the first conductive strip, and wherein the length of the first conductive strip in the first direction is less than half of the wavelength and wherein the length of the second conductive strip in the first direction is less than half of the wavelength, and wherein the length of the first conductive strip in the first direction summed with the length of the second conductive strip in the first direction make a sum which is approximately equal to half of the wavelength;
- a first load terminal electrically connected to the first conducting strip; and
- a second load terminal electrically connected to the second conducting strip, wherein the first cathode is located on the part of the first conductive strip closest to the second conductive strip, and wherein the first anode is located on the part of the second conductive strip closest to the first cathode, and wherein the first anode and the first cathode are separated by a distance less than 10 microns forming a first field emission diode.

10. A rectifying antenna as claimed in claim 9, wherein the cathode is formed by a pointed edge of the first conductive strip, and wherein the anode is formed by a nonpointed edge of the second conductive strip.

11. A rectifying antenna as claimed in claim 9, wherein the first field emission diode is a vertical field emission diode.

12. A rectifying antenna for rectifying electromagnetic radiation with a wavelength shorter than 3.0 centimeters, comprising:

- a substrate comprising a layer of insulating material with a first surface;
- a first conductive strip on a first surface of the substrate with a length in a first direction wherein a pointed edge of the first conductive strip forms a cathode of a linear field emission diode; and
- a second conductive strip on the first surface of the substrate with a length in the first direction wherein the second conductive strip forms an anode of a linear field emission diode, and wherein the second conductive strip is located along the first direction with respect to the first conductive strip, and wherein the length of the first conductive strip in the first direction is less than half the wavelength, and wherein the length of the second conductive strip in the first direction is less than half the wavelength and wherein the length of the first conductive strip in the first direction summed with the length of the second conductive strip in the first direction make a total length which is approximately equal to half of the wavelength.

13. A rectifying antenna as claimed in claim 12, further comprising a conductive layer on a second surface of the substrate, wherein the conductive layer is separated from the first conductive strip by a distance which is approximately equal to one fourth of the wavelength.

14. A rectifying antenna as claimed in claim 13, further comprising:

- a first load terminal electrically connected to the first conductive strip; and
- a second load terminal electrically connected to the second conducting strip, wherein the cathode is located on the part of the first conductive strip closest to the second conductive strip, and wherein the anode is located on the part of the second conductive strip closest to the cathode, and wherein the first anode and the first cathode are separated by a distance less than 10 microns forming a field emission diode.

15. A rectifying antenna as claimed in claim 13, further comprising:

- a third conductive strip on the first surface of the substrate with first and second nonpointed edges that form first and second anodes, wherein a pointed edge of the second conductive strip forms a cathode, and the first anode of the third conductive strip is located less than 10 microns away from the cathode on the first conductive strip forming a first field emission diode and the second anode of the third conductive strip is located less than 10 microns away from the cathode on the second conductive strip forming a second field emission diode; and

a fourth conductive strip on the first surface of the substrate with first and second pointed edges that form first and second cathodes, wherein a non-pointed edge of the first conductive strip forms an anode, and the first cathode of the fourth conductive strip is located less than 10 microns away from the anode on the first conductive strip forming a third field emission diode and the second cathode of the fourth conductive strip is located less than 10 microns away from the anode on the second conductive strip forming a fourth field emission diode.

* * * * *

1 | Monitoring for cloud apps AI powered, full stack, automated — Free trial! dynatrace.com

2 | Voted #1 VPN Service Full Access to Censored Apps & Websites. 100% Secure & Anonymous. Fastest Speeds. ipvanish.com

Homebrew 5.8 GHz patch FPV antenna

Posted on September 26, 2014 (http://vk4ghz.com/homebrew-5-8-ghz-patch-fpv-antenna/) Posted in FPV (http://vk4ghz.com/category/multirotors/fpv/), Multirotors (http://vk4ghz.com/category/multirotors/)

Like Share 35 people like this. Be the first of your friends.

These 5.8 GHz PCB patch antennas come from Kent Electronics (http://www.wa5vjb.com/) in Texas, USA.

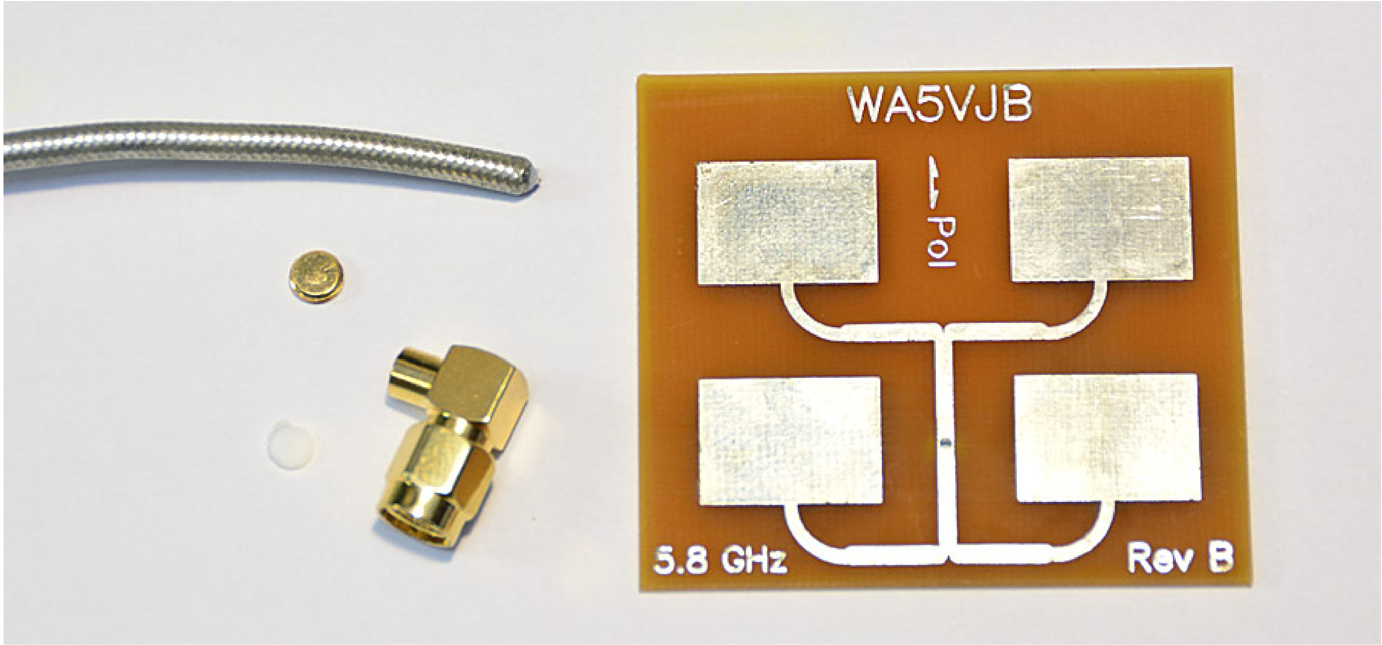

You can download the product PDF from here (http://www.wa5vjb.com/pcb-pdfs/5800MHzPatchArrays.pdf).

The quad patch array (pictured above) offers 11 – 12dBi gain, or 8.8 – 9.8 dB gain over a (unity gain) dipole.

To put that into a simplistic terms, is almost 10 times better than your dipole.

Along with some UT141 semi-rigid hardline coaxial cable, and a right-angle SMA (m) connector, both from RF Supplier (http://www.rfsupplier.com/) in Shenzhen, China, a low-cost patch antenna can be made for multirotor FPV use.

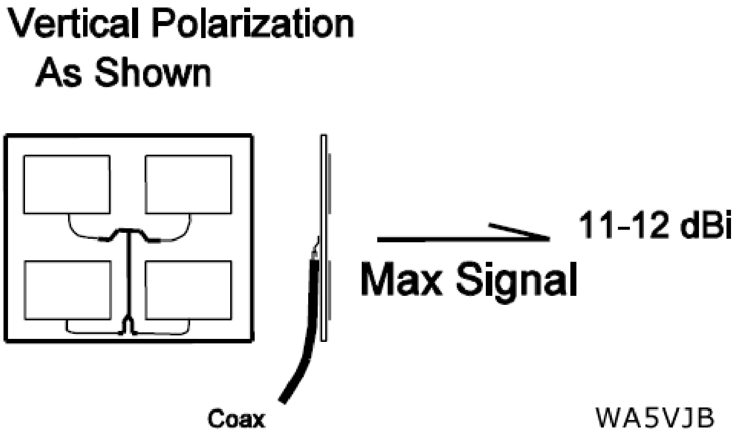

Another handy store for connectors and coax in Brisbane is The RF Shop (http://rfshop.com.au/), located in Oxley.
RF Shop is more expensive than a Chinese online store, (and most of their product is better quality, from Taiwan), and operates out of a home, but give them a call first, as they don't mind if you drop in and pay & pick up.

First step, prepare one end of the 141 to be soldered to the SMA (m) connector.

Working with 141 and SMAs is almost idiot proof.
141 has a teflon dielectric, so unlike RG58 type cable for CB or ham use, won't deform/melt with abuse from a soldering iron!

Start by preparing one end to suit soldering to the SMA (m) R/A connector.,

Score the outer sheath with a blade, as you rotate the coax (being careful not to cut into the dielectric):

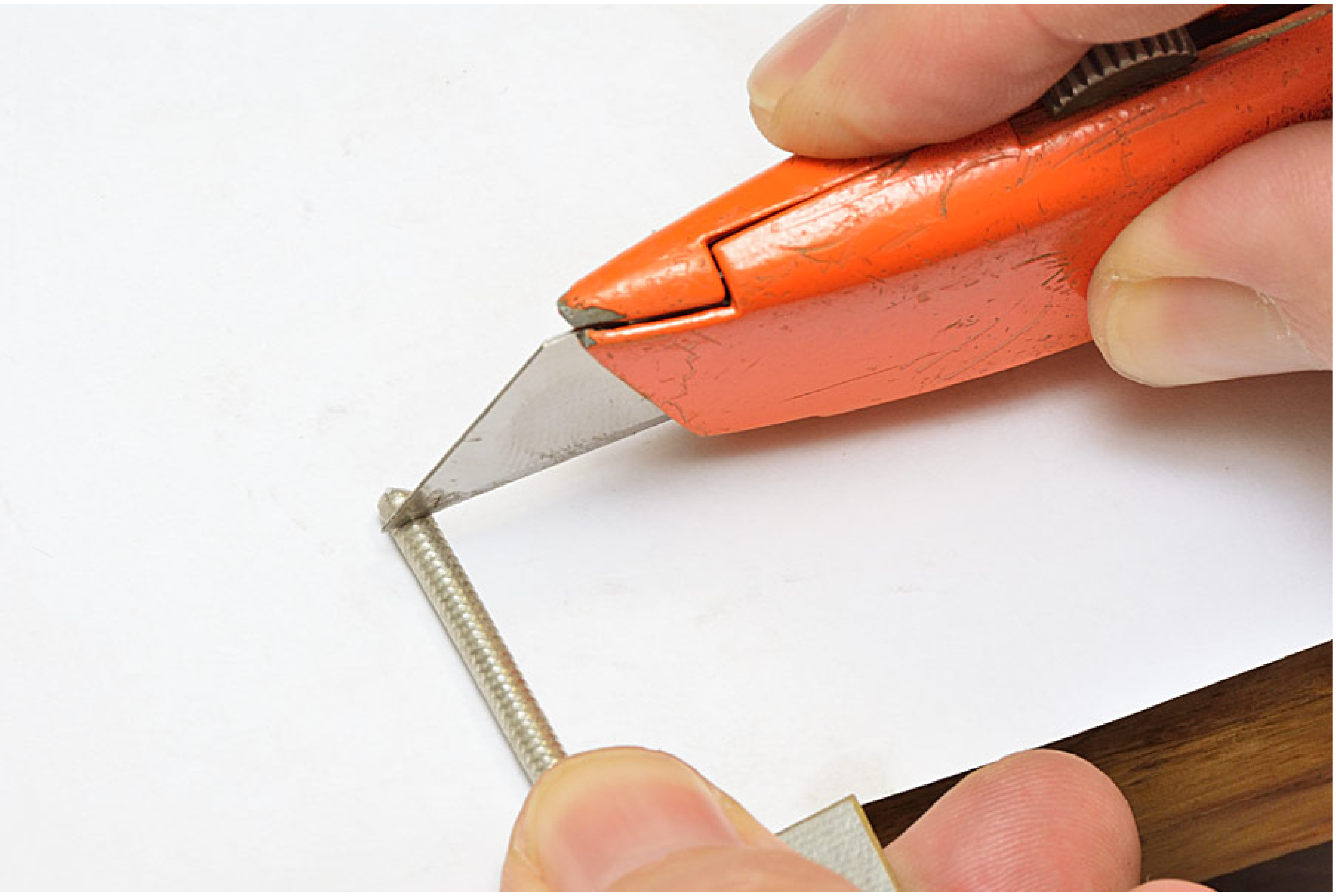

Using pliers, twist and pull off the outer sheath, like so:
(I'm cheating, using a picture from further down)

Coax end prepared, ready to solder to R/A SMA (m) connector:

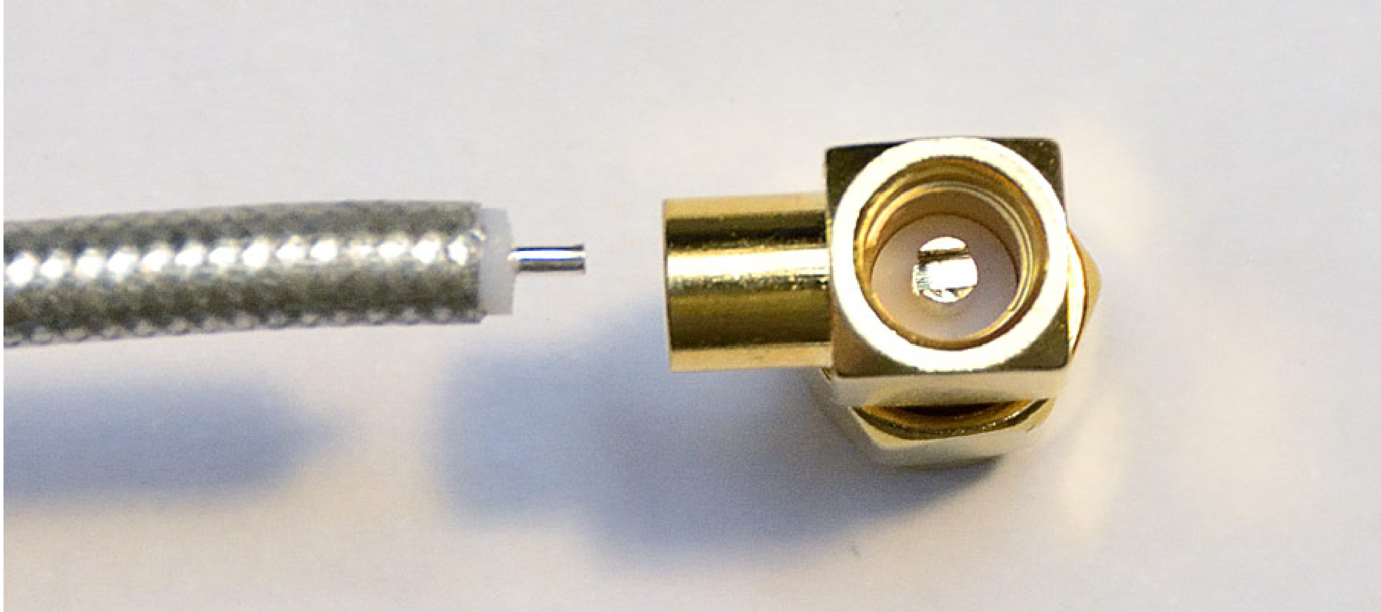

Coax inner soldered to centre pin:

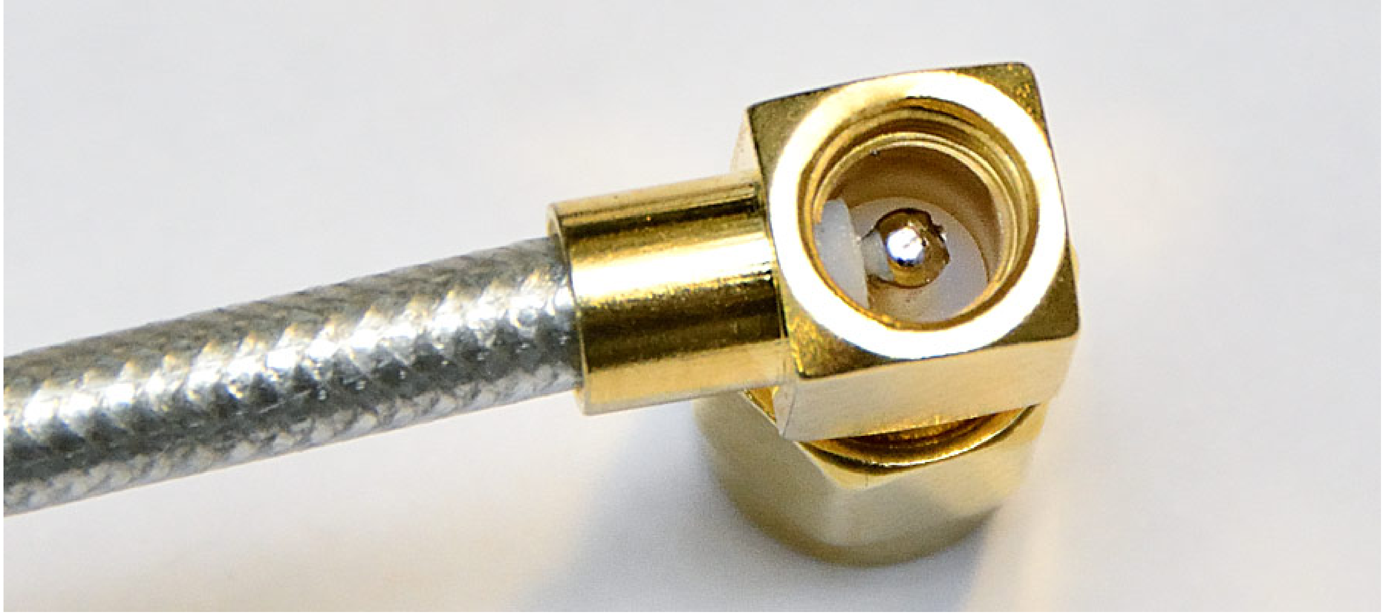

Add fillet of solder around outer sheath/connector body:

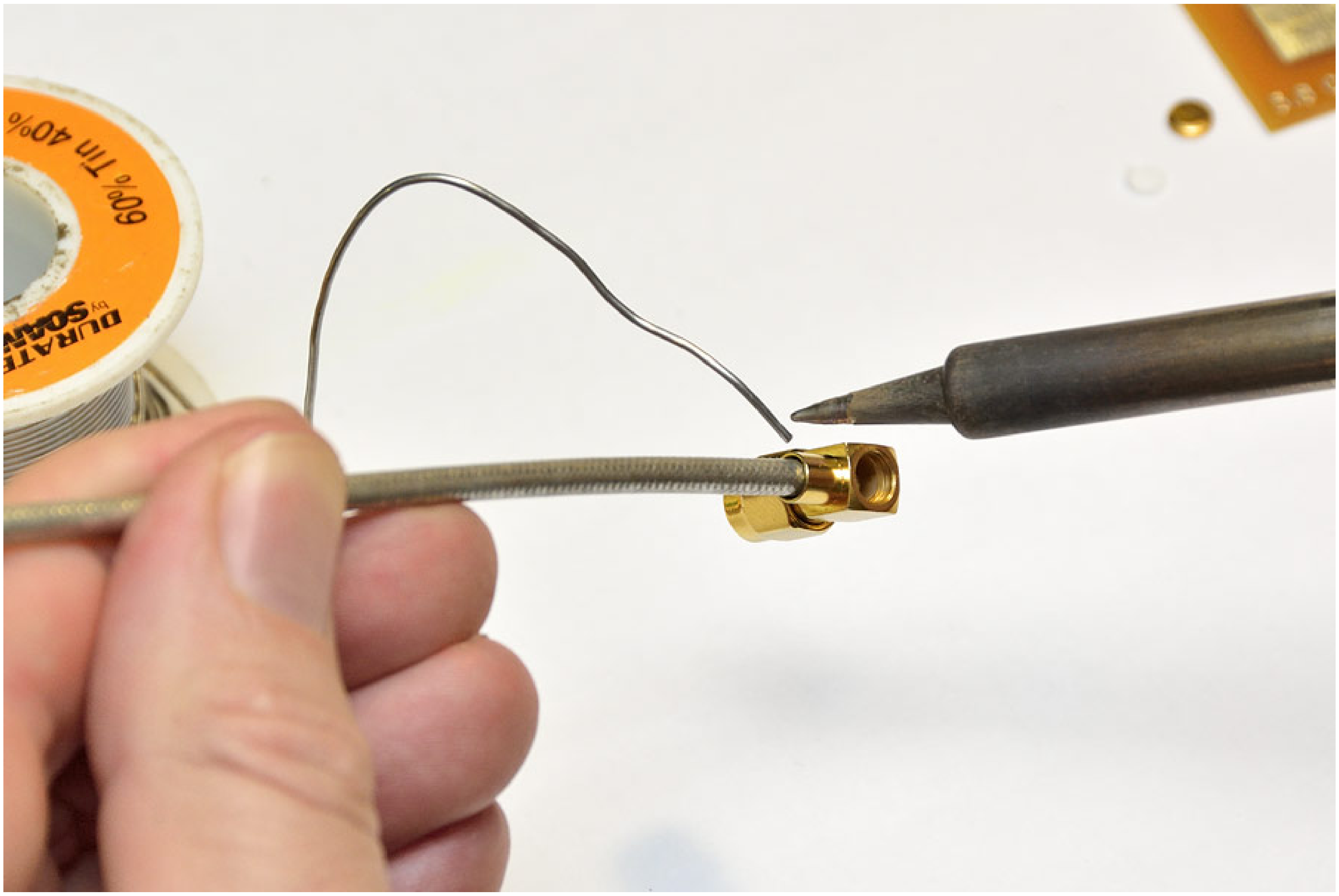
60% Tin 40%
DURATECH by SOLDER

End cap screwed on:
Terminated!

A nominal length of coax is chosen:

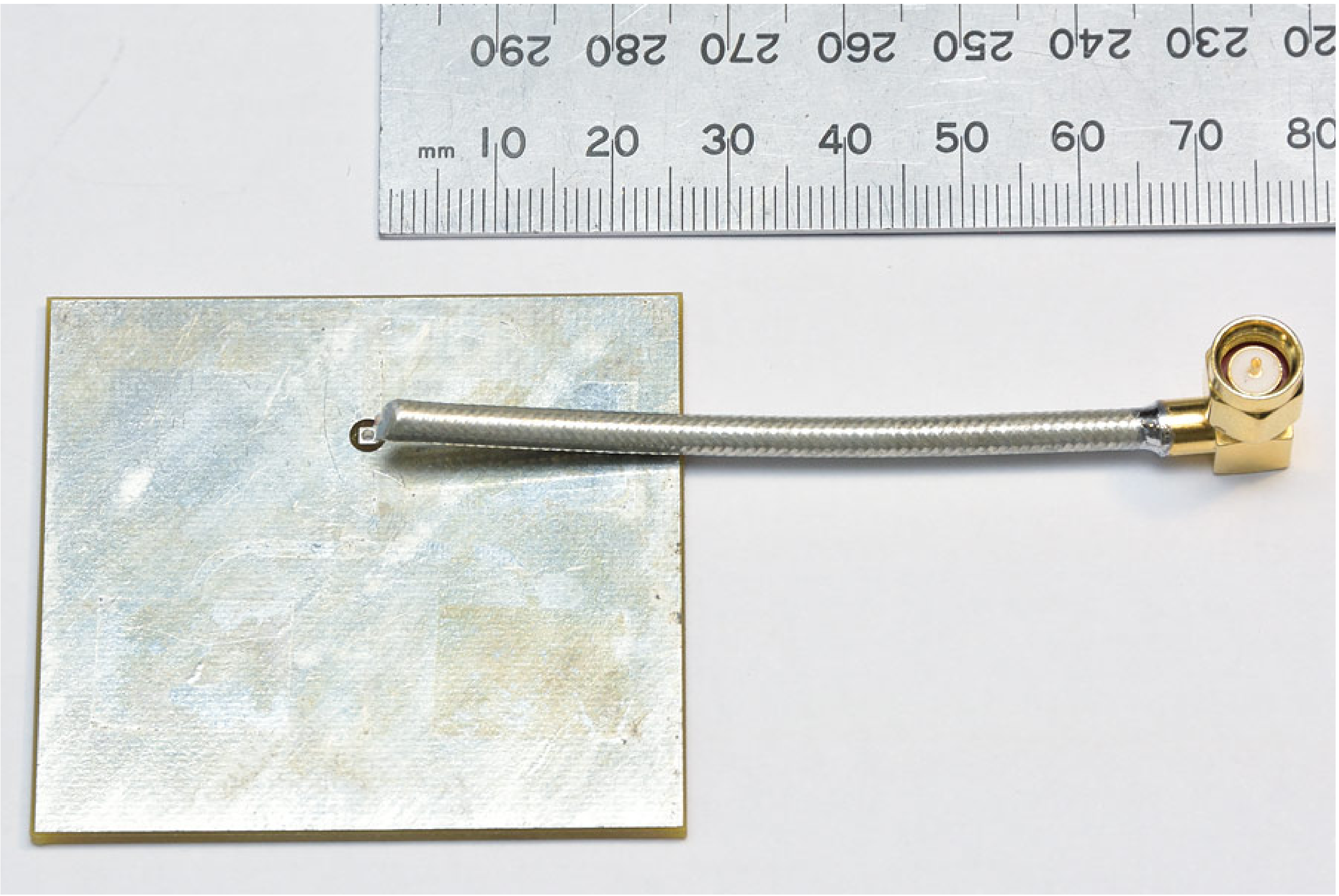
mm
10
20
30
40
50
60
70
290
280
270
260
250
240
230

Sorry Americans, in millimetres… about time you adopted the metric system, like everywhere else! 😉

By the way, nothing fancy in the soldering department… just a bog standard Weller:

AUSTRALIAN MADE
Weller®
WTCPT
Output For TC 201
Or WTL 24 Irons Only
Use Only Original
Weller Tips

Coax end prepared, ready to solder to PCB:

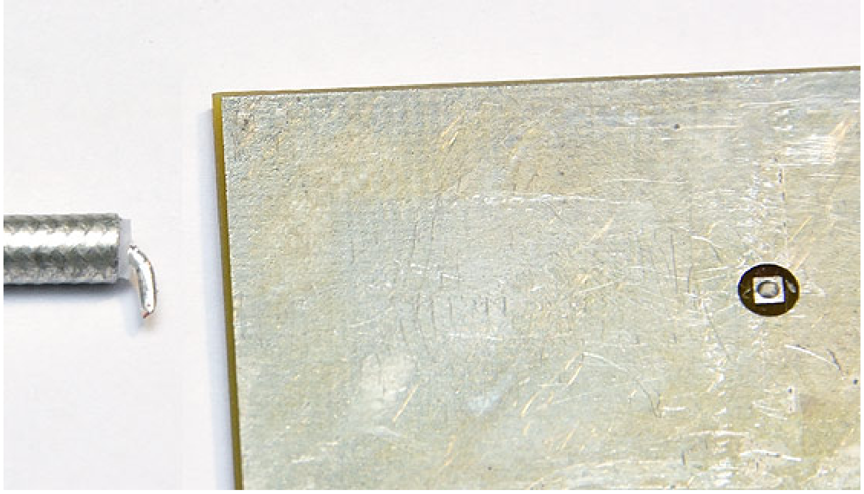

Coax soldered to the PCB antenna:

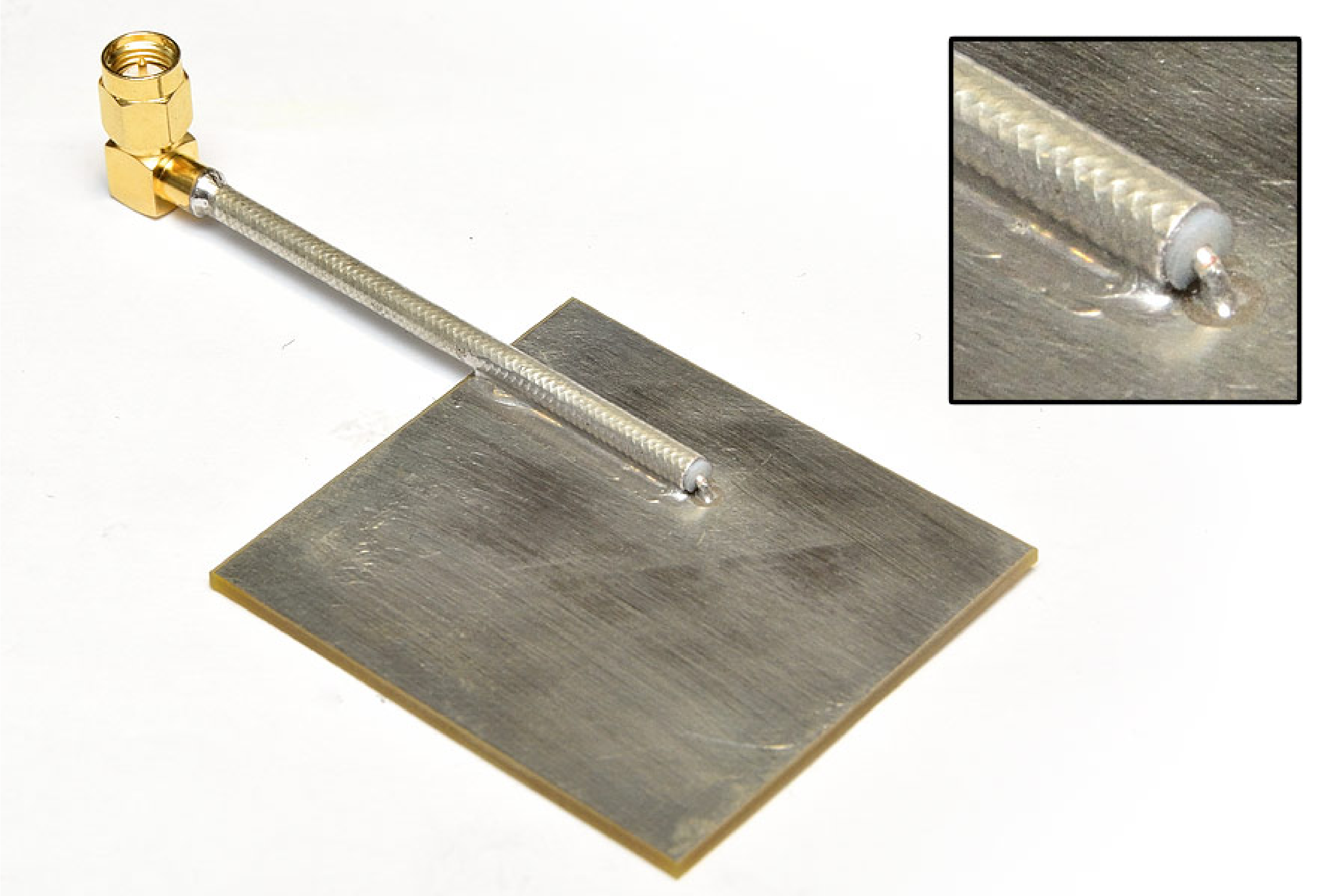

Note the fillet of solder that runs along the coax (on both sides), fastening the outer sheath to the ground plane of the PCB.

Installed on the rear of the 5.8 GHz diversity receiver/monitor, set up for horizontal polarisation:

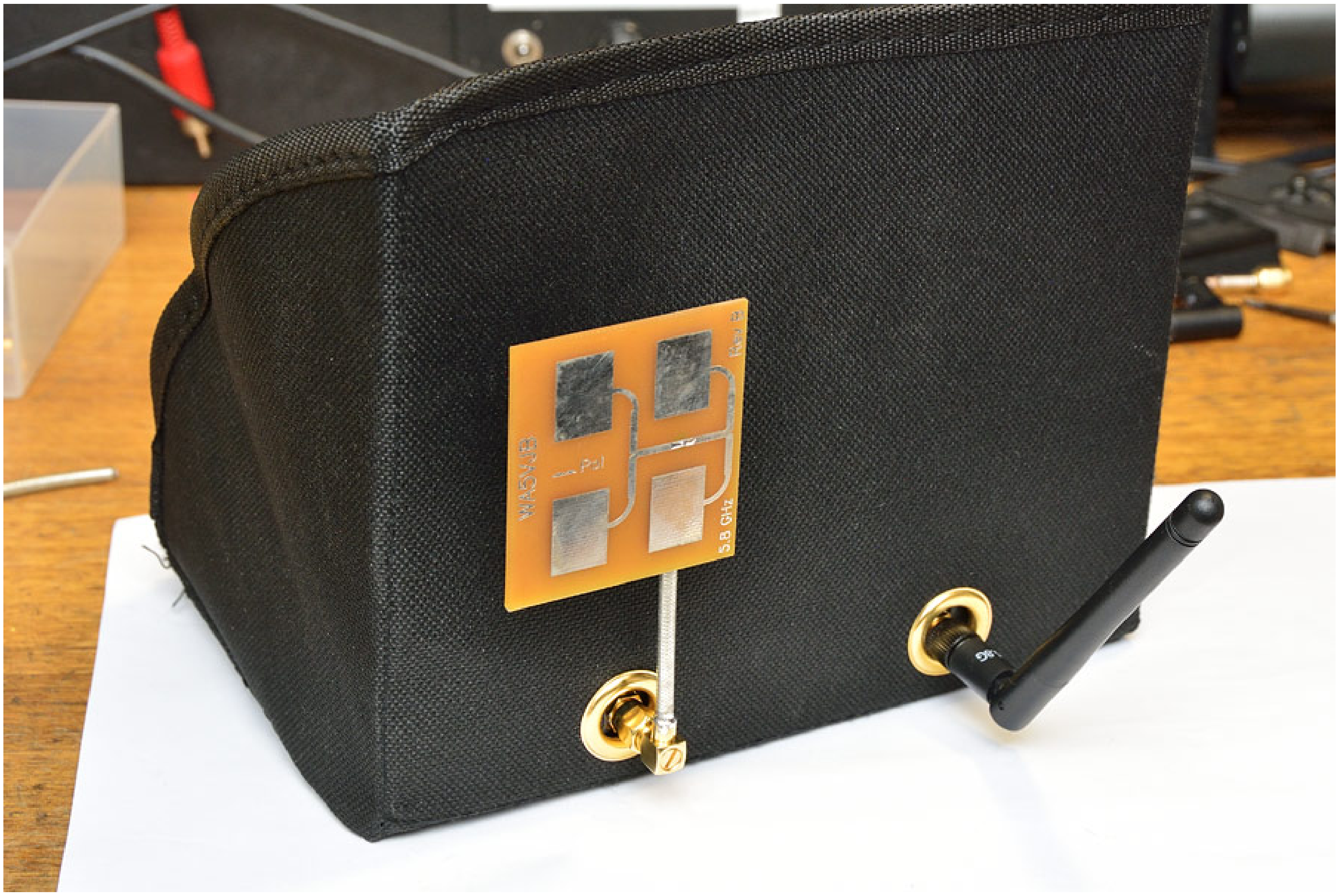

A dual patch antenna was constructed as well.
Same procedure:

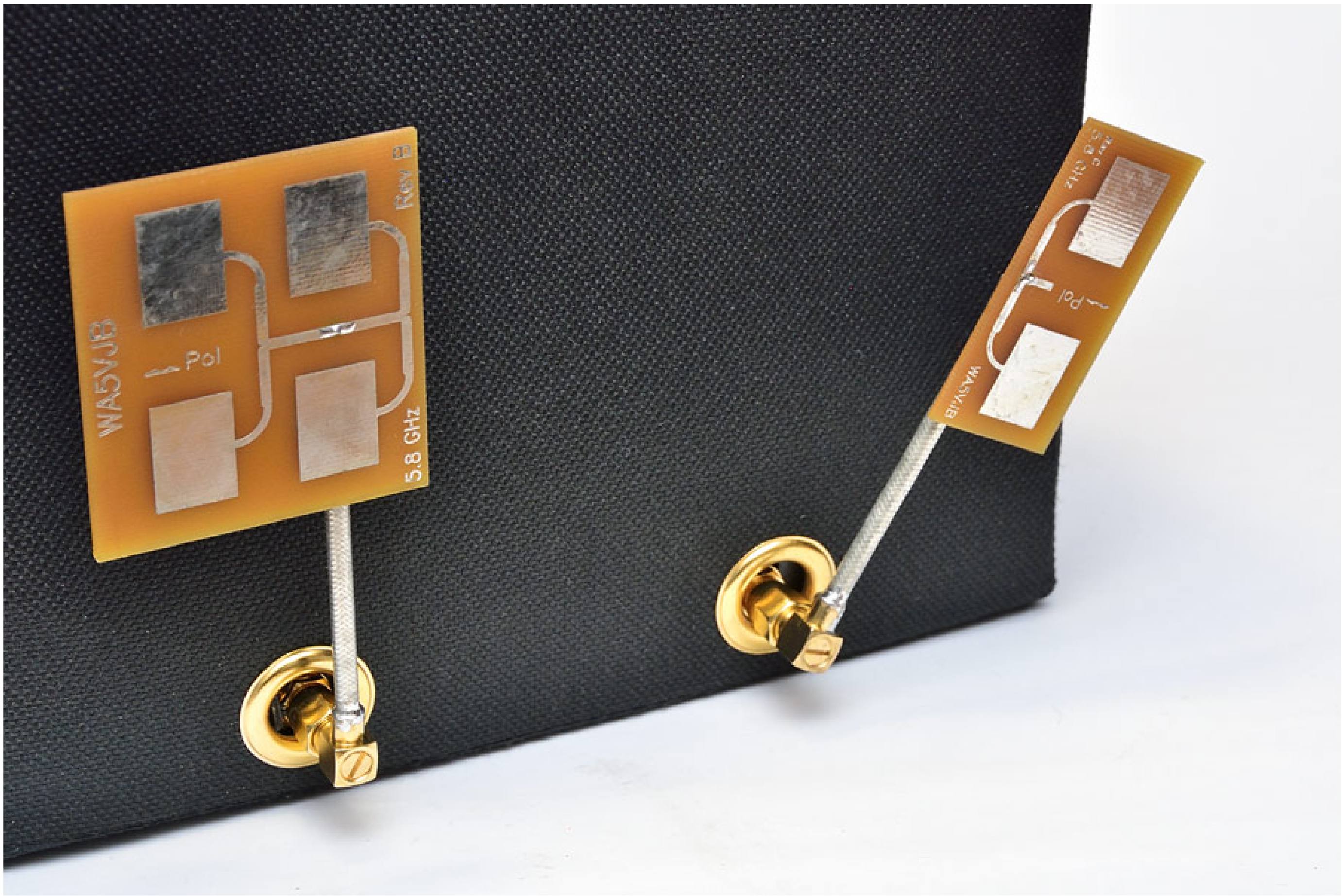
WA5VJB
Pol
5.8 GHz
Rev B
WA5VJB
Pol
5.8 GHz

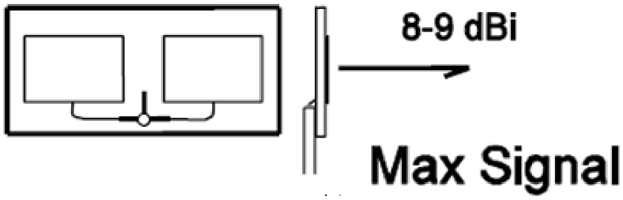
8-9 dBi
Max Signal

WA5VJB

Also on hand is a short in-line flexible “adapter”, with a male SMA at one end, a 30mm length of 141, and a female SMA, should a small angle be required:

As all hams operating on the microwave bands know, when dealing with SMAs, a 5/16″ spanner is always a handy tool to have:

Super Cheap Auto often have a bargain $1 bin on various size spanners.
Have a look for a 5/16″ next time you’re in SCA.

Super Value
SPANNER SINGLE
VARIOUS SIZES
$ 1.00
PLU - 7905
PA 16109
everything auto & much more

« Hobby King T-Style CF Propellers review (http://vk4ghz.com/hobby-king-t-style-cf-propellers-review/)

5.8 GHz patch FPV antenna gain (http://vk4ghz.com/fpv-quad-patch-antenna-gain/) »

HPE 3PAR Storage
Chat Live with HPE & Learn About 3PAR StoreServ Storage hpe.com

© Adam Maurer. All rights reveserved.

Translate

SCIENCE ALERT
An Open Access Publisher

Subscribe

- Abstract
- Full Text (PDF)
- References
- XML

Research Article

A Feed Network for a Novel E-H Shaped Microstrip Patch Antenna Array

N. Misran, **M. Tariqul Islam** and **K.J. Ng**

CrossMark
click for updates

ABSTRACT

This research presents a beamforming feed network for a novel E-H shaped microstrip patch antenna array. The broadband microstrip antenna array was developed based on the novel L-probe fed inverted hybrid E-H (LIEH) shaped element design arranged into 4x1 uniform linear array antenna, fed by a feed network. The corporate 4-ways Wilkinson power divider which was fabricated in-house covering the frequency range from 1.83-2.29 GHz (460 MHZ) were integrated to form the beamforming feed network. The beamforming feed network comprises of commercial variable attenuators (KAT1D04SA002), variable phase shifters (KPH35OSC000) and the corporate 4-ways Wilkinson power divider which was developed in-house. The developed 4-ways Wilkinson power divider achieves bandwidth of 22.33% at VSWR 1.25. The measured isolation and insertion loss are 30.72 and 0.32 dB correspondingly. The feed network developed and fabricated in-house has been shown to function well within its design ability.

Services
- Related Articles in ASCI
- Similar Articles in this Journal
- Search in Google Scholar
- View Citation
- Report Citation

How to cite this article:

N. Misran, M. Tariqul Islam and K.J. Ng, 2008. A Feed Network for a Novel E-H Shaped Microstrip Patch Antenna Array. *Journal of Applied Sciences, 8: 1982-1986.*

DOI: 10.3923/jas.2008.1982.1986

URL: http://scialert.net/abstract/?doi=jas.2008.1982.1986

INTRODUCTION

The evolution of **wireless communication**s has a deep impact on the field of antenna design. One essential component of the antenna array is its feeding technique. In an antenna array multiple radiating elements are assembled in an electrical and geometrical configuration. The antenna array provides beamforming capability to extract the desired signal and simultaneously filtering out the unwanted interference signal and environmental noise.

In designing array antenna, dipole, horn, waveguide slot or microstrip antenna can be used as radiating elements. However such radiating elements are unable to meet the requirements to be aesthetic and light weighted if implemented on the cell sites. To be aesthetic, the radiating element of the antenna array is desired to be small in size, low profile, conformal and compatible with integrated circuit and etc. Therefore, the radiating elements are preferred to be fabricated by microstrip technology that offers its non-electrical characteristic advantages (Pozar and Schaubert, 1995). Microstrip antennas are low profile and light weight. They can also be made conformable and well suited to be integrated with microwave integrated circuit (Lau *et al*., 2006; Zhang and Wang, 2006). In terms of fabrication, such system offers simplicity, which allows mass production and cost-effective manufacturing as well as high performance.

A combination of X- and C-band slots and patches were investigated and it was concluded that C-band patch/X-band slot concept had the greatest merit. In that design, the C-band feed network was on the same plane with the patches while the X-band feed was behind the ground plane. A combination of stacked perforated L-band patch overlaid over an array of C-band patches to achieve shared-aperture L/Cb was reported in Shafai *et al*. (2000). Employing similar feeding principles, the C-band feed network behind the ground plane was aperture-coupled to the patches. The L-band feed network was on the same plane with the lower L-band perforated patch. This idea was further developed and applied to L- and X-band array in Pozar and Targonski (2001).

In their studies, Targonski and Pozar (1998) noted that it was impossible to accommodate two linear polarizations for two separate frequencies on a single feed layer. Maci and Gentili (1997), adding that single feed network for two frequencies was the most critical problem to solve for a dual-band array. Current dual-band shared-aperture technology suggests employing the higher- frequency feed network behind the ground plane (aperture-couple) and the lower-frequency feed network above the ground plane (coplanar feed). This may be because higher frequency feed produces higher spurious radiation, due to shorter wavelength and longer layout length. In a small array, feed layout above the ground plane may not produce excessive spurious radiation. However, in a larger array, feed line radiation may not be ignored.

In this research, the beamforming feed network was constructed using a four-way Wilkinson power divider of corporate structure type and a phase shifter-attenuator networks, connected in series to the ports to examine the scanning capability of the array. Due to the high cost and complexity of the design for planar and high resolution array, the design focuses on the development of a four-element (4x1) array antenna.

BEAMFORMING FEED NETWORK

A feed network is integrated with the antenna array. The radiating element of the antenna array is based on the LIEH shaped Microstrip Patch Antenna (MPA) that is arranged in a 4x1 linear array configuration which is termed as Broadband Microstrip Patch Antenna Array (BMPAA). The antenna array is constructed using two dielectric layer arrangement where a thick air-filled substrate was sandwiched between top-loaded dielectric substrate or superstrate with inverting radiating patch and an aluminium ground plane. The antenna array is used contemporary design techniques namely, the L-probe feeding, inverted patch and slotted patch techniques to meet the design requirement.

The difference in amplitude and phase of the array is fed to each element by feed network. This causes the radiating field of each element combines constructively in the intended direction and cancel out each other at other undesired directions in the far field (Balanis, 1997). Thus higher power gain and more flexibility in controlling the shape of the beamwidth and sidelobe levels can be achieved from this method. Moreover, the feed network allows the radiation pattern to be varied for beam scanning without any physical antenna motion.

Arrays can be fed in many ways, but the two general types of feed arrangement for microstrip patch arrays are series feed and corporate feed. Microstrip patch array in a series feed network is fed by a single transmission line, whereas patch array in a corporate feed network is fed by multiple transmission lines. Series feed network is much easier to design and fabricate compared to the corporate feed network. However this method will result in progressive phase delay between elements thus making unsuitable for phase scanning. Another problem with series feed arrangement is the high VSWR that is caused by the additive mismatches at the various junctions between the elements (Zurcher and Gardiol, 1995).

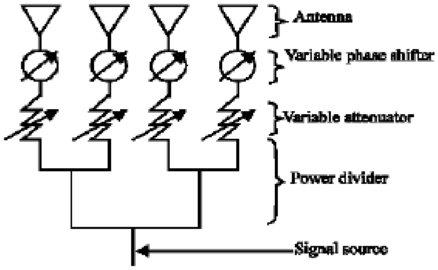

Fig. 1: Schematic diagram of a beamforming feed network

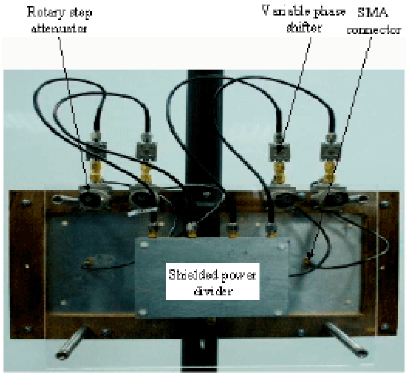

Fig. 2: Rear view of the 4x1 LIEH shaped BMPAA with a beamforming network

Corporate feed network is more flexible and offers better phase control over the performance of each array element. It is best suited for scanning phased array, multi beam arrays or shaped beam arrays. However, the use of corporate feed network is more preferable over s work as it is less affected by frequency scan. A corporate feed network supplies excitation individually to each array element and use equal line lengths and power dividers for each element.

The beamforming feed network for the BMPAA consists of a 4-ways corporate structure Wilkinson power divider and a network of phase shifter-attenuator as shown in Fig. 1.

The phase shifter-attenuator network comprises of a commercial off-the-shelf (COTS) variable phase shifter KPH350SC00 and step-rotary attenuator KATID04SA002, both are from KMW Inc. The 4-way Wilkinson power divider was developed and fabricated in-house by using Taconic RF-35 with the dielectric permittivity, ε_{r1} of 3.5, thickness h = 0.76 mm and with the dimension 25x20 mm. Table 1 gives the summary of the beamforming feed network design and its electrical parameters.

Figure 2 shows the fabricated beamforming feed network connected to the antenna array. As shown in the Fig. 2, the power divider is covered by an aluminium shield and the antenna array is behind the power divider.

Table 1: Summary of the beamforming feed network

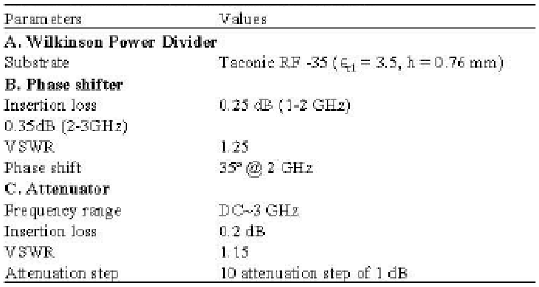

Parameters	Values
A. Wilkinson Power Divider	
Substrate	Taconic RF -35 (ε_{r1} = 3.5, h = 0.76 mm)
B. Phase shifter	
Insertion loss	0.25 dB (1-2 GHz)
0.35dB (2-3GHz)	
VSWR	1.25
Phase shift	35° @ 2 GHz
C. Attenuator	
Frequency range	DC~3 GHz
Insertion loss	0.2 dB
VSWR	1.15
Attenuation step	10 attenuation step of 1 dB

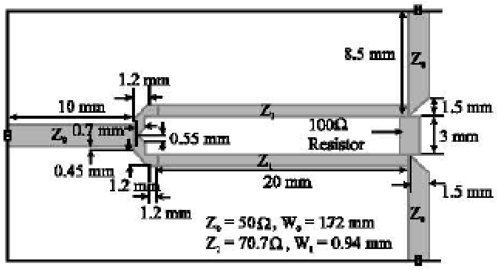

Fig. 3: The layout of 2-ways Wilkinson power divider on substrate Taconic RF 35 (ε_{r1} of 3.5, h = 0.76 mm and δ = 0.0018)

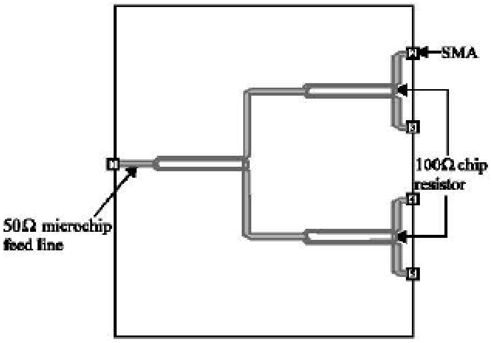

Fig. 4: The layout of 4-ways Wilkinson power divider

A 4-ways power divider is designed as a corporate feed network to feed the radiating elements with equal amplitude and phase. This power divider is based on Wilkinson power divider which offers good matching in all ports with good isolation between the output ports. The Wilkinson power divider has a useful property of being lossless when the output ports are matched, that is, only reflected power is dissipated (Pozar, 1990). The design schematic for the 2-ways Wilkinson power divider layout is shown in Fig. 3 and the layout of a 4-ways Wilkinson (3 dB) power divider is shown in Fig. 4.

Translate

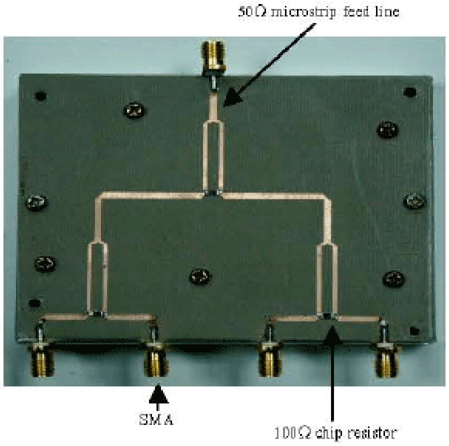

Fig. 5: Fabricated 4-ways Wilkinson power divider

Figure 5 shows the fabricated 4-ways Wilkinson power divider which is fabricated on Taconic RF-35.

RESULTS AND DISCUSSION

The antenna array and the power divider are simulated by *Sonnet® Suite em* simulator. The fabricated antennas and power divider are measured using the Agilent PNA E8358A network analyzer, Agilent ESG-DP Series E4436B signal generator, Advantest R3131A spectrum analyzer and the standard gain LPDA-0803 log periodic dipole antenna. SMA connector calibration kit, SMA male to male cable and automatic controlled rotator are also required for the measurement. Measurement is conducted in the UKM microwave lab and the UKM open field. The return loss measurement of 4-ways Wilkinson power divider is taken by using the Agilent PNA E8358A network analyzer. Return loss of every terminal is measured by terminating other element with 50 Ω load. Figure 6 shows the return loss at 1st terminal in 4-ways Wilkinson power divider. The measurement result shows higher value than the simulation results. This is due to the effect from SMA connector of every terminal. The aluminium ground plane is used to support SMA connector that is not been grounded properly. This installation problem causes the higher value of measurement; besides the used SMA connector, probe pin is not tapered. Simulated and measured return loss results are in good agreement in all terminals. The operating range for power divider is from 1.83-2.29 GHz (460 MHz) with bandwidth 22.33% at VSWR 1.25 considering all terminals.

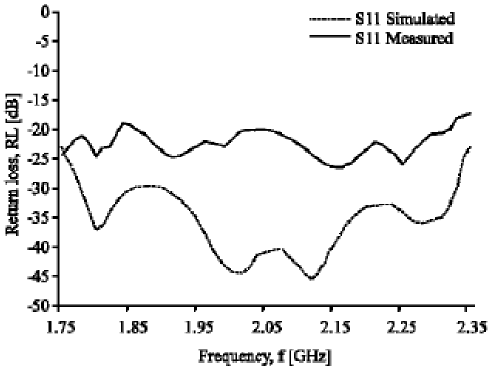

Fig. 6: Measured and simulated return loss of port 1 of the 4-ways Wilkinson power divider

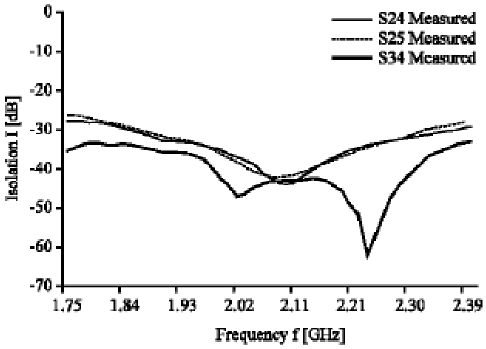

Fig. 7: Measured isolation of the 4-ways Wilkinson power divider

Table 2: Summarized result of power divider

Translate

Properties	Measurement results
Bandwidth	460 MHz (1.83-2.29 GHz at VSWR = 1.25)
Isolation	30.72 dB
Insertion loss	0.32 dB
Amplitude balance	0.70 dB
Phase balance	5.40°

Isolation of 4-ways Wilkinson power divider is shown in Fig. 7. The maximum value of isolation in the operating frequency range is at 30.72 dB. The measurement results of power divider are shown in Table 2.

Figure 8 and 9 show the E-plane and H-plane radiation patterns of the antenna array at 1.91 and 2.14 GHz (resonance frequencies) by using feed network. Chebyshev amplitude distribution is used to have narrow beamwidth and low equal side lobe level (-20 dB). Chebyshev distributions value is shown in Table 3 as normalized in dB.

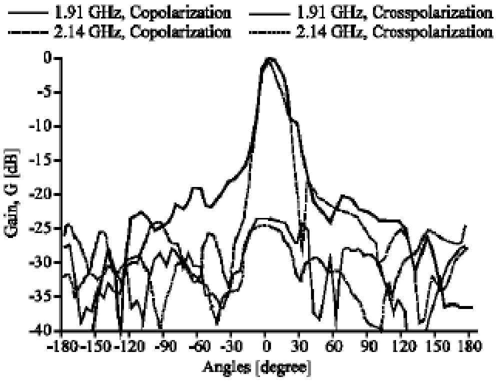

Fig. 8: Measured E-plane radiation patterns for the LIEH shaped BMPAA at 1.91 and 2.14 GHz

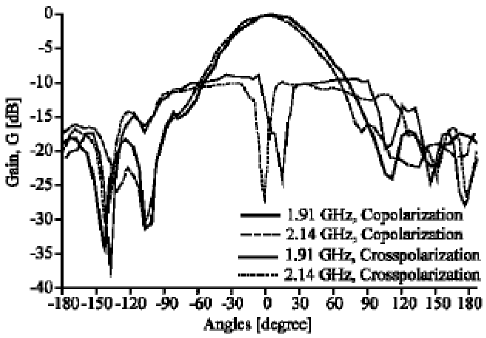

Fig. 9: Measured H-plane radiation patterns for the LIEH shaped BMPAA at 1.91 and 2.14 GHz

Table 3: Chebyshev distributions for each element of 4x1 BMPAA

Element-1	Element-2	Element-3	Element-4
0.48312 (3.1594 dB)	1(0 dB)	1(0 dB)	0.48312 (3.1594 dB)

Table 4: Amplitude and phase value for measurements

0° Scanning Element-1	Element-2	Element-3	Element-4
-10.06 dB	-6.5 dB	-7.5 dB	-9.65 dB
100.0°	-50.85°	149.3°	2.55°

However, the exact amplitude and phase distribution were unable to obtain due to the limitation of the available variable attenuator and phase shifter. Therefore, the antenna radiation pattern result is expected to be distorted from the desired performance.

Scanning pattern is measured for 0°.Amplitude and phase values are chosen by considering the following initial phase angle and by employing progressive phase shift equations. Amplitude and phase values for measurement are shown in Table 4.

Initial phase angle: Port 1: 100°
Port 2: 105°
Port 3: 101°
Port 4: 105°

The phase in each element of the array is progressively shifted by α = -βdcosθ, where β is the phase shift factor and d is the inter-element spacing.

Translate

T atterns show some fluctuations due to the reflection from some obstacles in the field, however, they have good beam patterns and crosspolarization level, which is shown in Fig. 8 and 9. The 3 dB beamwidth is closed to 25° for the E-plane while the 3 dB beamwidth is about 65° for the H-plane. The H-plane radiation pattern is virtually symmetrical, while the E-plane radiation pattern exhibits some asymmetries, which is similar in the report by Huynh and Lee (1995) using a thick substrate. As shown in Fig. 8, the sidelobe levels are unequally distributed. The first side lobe levels at 1.91 and 2.14 GHz are -18.9 dB (at -60°) and -20.47 dB (at 45°), respectively. This result is due to the amplitude/phase unbalances in the beamforming feed network. The maximum crosspolarization of the array is in the order of -23.47 and -10 dB for E-plane and H-plane, respectively.

CONCLUSION

The beamforming feed network was constructed using a four-way Wilkinson power divider of corporate structure type and off the shelf variable phase shifter (KPH350SC00@ KMW Inc) attenuator (KATID04SA002@ KMW Inc) networks to measure the beamforming capability of the array antenna. Due to the high cost and complexity of the design for planar and high resolution array, the design focused on the development of a uniform four-element (4x1) array antenna. The array designed by employing novel hybrid E-H shaped design, inverted patch, slotted patch and L-probe. The developed 4-ways Wilkinson power divider enjoys the bandwidth of 22.33% at VSWR 1.25. The measured isolation and insertion loss are 30.72 and 0.32 dB correspondingly.

ACKNOWLEDGMENTS

The authors would like to thank the IRPA Secretariat, Ministry of Science, Technology and Environmental of Malaysia, for sponsoring this research. This research is funded by IPRA 04-02-02-0029.

REFERENCES

Balanis, C.A., 1997. Antenna Theory: Analysis and Design. 1st Edn., John Wiley and Sons Inc., New York.

Huynh, T. and K.F. Lee, 1995. Single-layer single-patch wideband microstrip antenna. Electron. Lett., 31: 1310-1312.
Direct Link |

Lau, K.L., K.M. Luk and K.F. Lee, 2006. Design of a circularly-polarized vertical patch antenna. IEEE Trans. Antennas Propagat., 54: 1332-1335.
Direct Link |

Maci, S. and G.B. Gentili, 1997. Dual-frequency patch antennas. IEEE Trans. Antennas Propagat. Maga., 39: 13-20.
CrossRef | Direct Link |

Pozar, D.M. and D.H. Schaubert, 1995. Microstrip Antennas, the Analysis and Design of Microstrip Antennas and Arrays. IEEE Press, New York.

Pozar, D.M. and S.D. Targonski, 2001. , A shared-aperture dual-band dualpolarized microstrip array. IEEE Trans. Antennas Propagat., 49: 150-157.
Direct Link |

Pozar, D.M., 1990. Microwave Engineering. Addison Wesley, New York.

Shafai, L.L., W.A. Chamma, M. Barakat, P.C. Strickland and G. Seguin, 2000. 2000 Dual-band dual-polarized perforated microstrip antennas for SAR applications. IEEE Trans. Antennas Propagat., 48: 58-66.
Direct Link |

Targonski, S.D. and D.M. Pozar, 1998. A dual-band dual-polarized printed antenna element. Electron. Lett., 34: 2193-2194.
Direct Link |

Zhang, Y.P. and J.J. Wang, 2006. Theory and analysis of differentially-driven microstrip antennas. IEEE Trans. Antennas Propagat., 54: 1092-1099.
Direct Link |

Zurcher, J.F. and F.E. Gardiol, 1995. Broad band patch antenna. Artech House Inc., Boston, New York.

© 2017 Science Alert. All Rights Reserved

Bachelor thesis
Electrical engineering with emphasis
on telecommunication
May 2015

Mobile Phone Antenna Design

By

Nazem Alsmadi & Khalid Saif

(nazemalsmadi@hotmail.com) (khaled_s_23@yahoo.com)

25th of May 5, 2015

Supervisor

Dr. Benny Lövström

Examiner

Dr. Sven Johansson

Department of applied signal processing
Blekinge Institute of Technology
37179 Karlskrona, Sweden

BLEKINGE TEKNISKA HÖGSKOLA
BTH

Abstract

This thesis focuses on mobile phones antenna design with brief description about the historical development, basic parameters and the types of antennas which are used in mobile phones. Mobile phones antenna design section consists of two proposed PIFA antennas. The first design concerns a single band antenna with resonant frequency at GPS frequency (1.575GHz). The first model is designed with main consideration that is to have the lower possible PIFA single band dimensions with reasonable return loss (S11) and **the** ***efficiencies. Second design concerns in a wideband PIFA antenna which cover the range from 1800MHz to 2600MHz. This range covers certain important bands: GSM (1800MHz & 1900MHz), UMTS (2100MHz), Bluetooth & Wi-Fi (2.4GHz) and LTE system (2.3GHz, 2.5GHz, and 2.6GHz). The wideband PIFA design is achieved by using slotted ground plane technique. The simulations for both models are performed in COMSOL Multiphysics.***

The last two parts of the thesis present the problems of mobile phones antenna. Starting with Specific absorption rate (SAR) problem, efficiency of Mobile phones antenna, and hand-held environment.

Acknowledgements

The work of the thesis has been carried out of Blekinge Institute of Technology (BTH), Karlskrona, Sweden under the supervision Dr. Benny Lövström.

Firstly, thanks to Allah who give us the ability and the strength to be managed to complete our thesis work and we really thank the people who support us and guide as to success this thesis work.

We would like to gratitude our thesis supervisor Dr. Benny Lövström for his encourages and his comment he was behind this success and we extend our thanks for Prof. Ansel Berghuvud.

We would like to thank Dr. Sven Johansson.

Contents

Abstract I

Acknowledgment II

Contents III

Acronyms V

Chapter 1

1.1 Introduction 1
1.2 Thesis Motivation 2
1.3 The Development Of Mobile Phone Antennas 2

Chapter 2

2. Basic Mobile Phone Antennas 8
2.1 Parameters Of Mobile phone Antennas 8

Radiation pattern 8

Directivity 10

Polarization 12

Impedance 14

Antenna gain 14

Antenna efficiency 15

Antenna Beamwidth 15

Effective area 16

Band width 16

VSWR 17

2.2 Antenna Types 18

Dipole antenna 18

Monopole antenna 19

PIFA antenna 20

Fractal antenna 22

Chapter 3

3 Mobile Phone Antennas Design 23

3.1 Mobile Phone GPS Antenna Design 23

Model descriptions 25

Calculations 26

Simulations 27

Result and Analysis 29

3.2 Wideband PIFA Antenna Design 34

Model description 34

Calculation 35

Result and analysis 39

3.3 Wi-Fi And Bluetooth 41

Chapter 4

4 Specification Absorption Rate (SAR) 42

4.1 How To Measure The SAR 43

4.2 The Technique Of Reducing SAR 44

4.3 The Limit Of SAR In Different Countries 46

4.4 SAR With Different Types Of Antenna 47

Chapter 5

5 Problem Of Mobile Phone Antenna Design 50

5.1 The Radiation Efficiency Of The Mobile Phone Antenna 50

5.2 Bandwidth 52

5.3 Mutual Coupling Antenna To Antenna Loss 54

5.4 The Hand-Held Environment Problem 55

Summary And Conclusion 59

Reference 60

Acronyms

LAN	Local area network.
GSM	Global system for mobile communication.
PIFA	Planar inverted – F Antenna.
UMTS	Universal mobile telecommunication system.
DECT	Digital enhanced cordless tele communication.
PDC	Personal digital cellular.
RFID	Radio frequency identification.
HPBW	Half – power beam width.
FNBW	First null beam width.
IFS	Iterative function system.
SAR	Specific absorption rate.
RF	Radio frequency.
PMA	planar monopole antenna.
MOM	Method of moments.
FDTD	Finite difference time – domain.
ESA	Electrically small antenna.
Wi-Fi	Wireless fidelity
GPS	Global positioning system
RHCP	Right Hand circularly polarized
IFA	Inverted F-antenna
PTFE	Polytetrafluoroethylene
PEC	Perfect electrically conductor
LTE	Long term evaluation

Chapter1

1.1 Introduction

The huge development of the mobile phones have grown up rapidly in the last years, frequency bands have come up and the market is asking for smaller mobile phones with more services which give the user the ability to use the mobile phone with good signal performance and helps to use the mobile phone around the all world. On top of that it is important to reduce the risks affecting in the human body because of the antenna radiation.

In the past the mobile phone was so heavy, big and had external antenna on the top of the phone which effect badly on the human head and most of the signal radiate is reflected and absorbed by the human head which lead to bad efficiency.

Nowadays the internal antenna has been using instead of the external antenna the main reason of that is the internal antenna has a good relation with SAR rate, on the other hand the size of the phone became smaller.

Recently there are many types of the internal antennas for example PIFA antenna (Planar Inverted- F Antenna), fractal antenna and monopole antenna. Those kinds of antennas can cover a single band, dual band, wideband and multiband based on the design of the antenna.

PIFA antennas are used widely in mobile phone antennas design due to its advantages such as SAR rate and less interaction with hand-held environment, but one of the significant problems of PIFA antennas that’s PIFA antennas have a narrow bandwidth. Wideband and multiband PIFA antenna can be a solution of that problem. By designing multiband antenna or wideband antenna it’s possible to have one antenna can cover more frequency such as using one antenna that can cover the very important bands which are in use in most countries GSM, UMTS, Wi-Fi and LTE.

1.2 Thesis Motivation

The work in this thesis is motivated by the necessity to investigate antenna structures which can be integrated in today is mobile phones which have a small available space for antennas inside. Nowadays the demand on low profile antennas increases rapidly that’s in response to the huge developments of mobile phone devices either in functions or in size.

The other motivation is to investigate and design a wideband PIFA antenna. In general PIFA antennas have a disadvantage that’s PIFA antennas have a narrow bandwidth, but on other hand PIFA antennas have many advantages such as best SAR rate, low profile and easy to be integrated inside mobile phones which make PIFA antennas the best candidate for mobile phones antennas. There are many papers published which study and discuss the narrow bandwidth of PIFA antennas, but most of those papers give the solution of the problem in dual band PIFA or in multiband design which in both cases do not solve the problem totally. Since there are some bands remain without covering, consequently the need of several antennas appears for covering low and high frequency bands.

So this thesis presents a study of a proposed wideband PIFA antenna with bandwidth more than 30% and the bandwidth cover an important range which could cover GSM, UMTS, and some bands of LTE system.

1.3 The Development of Mobile Phone Antennas

In the first generation 1G of mobile system, the system was operating at 800 MHz and as known it was analog. The first antenna handset for one quarter of wavelength has length about 9.4 cm, and it was one antenna only. The first mobile phone was Motorola DynaTAC8000X[1] with type antenna is sleeve dipole as shown in the figure 1.3.1 which is not use any more in modern design of mobile phones antennas, however it still is used in different wireless LAN access points.

Sleeve dipole antenna has an efficient performance; the length of it is about half the wave length at its frequency. So at 850 MHz the antenna should be 176mm.

With the improvement on mobile phone and due to the dramatic minimizing size of the mobile phone, there were no needs for a sleeve dipole to be proportional to the cellular phone.

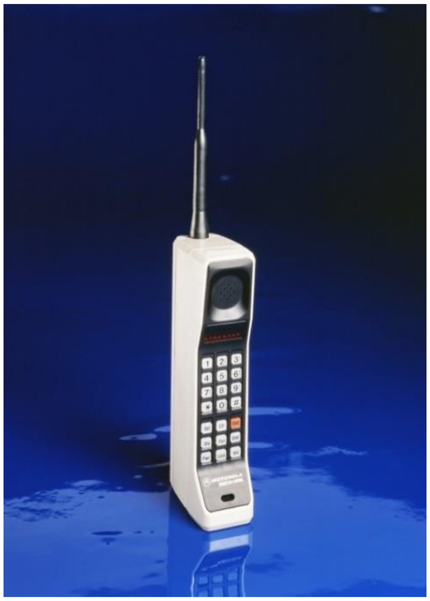

Photo was loaded from http://mashable.com/2014/03/13/first-cellphone-on-sale/

Figure 1.3.1 Motorola DynaTAC8000X with Sleeve dipole antenna.

In 1990s 2G launched which offered new services such as text message, and it is operated at GSM 900 MHz, where later on at 1800 MHz unlikely 1G, the second generation handset antenna has two antennas monopole and helix with only single band supporting as shown below in figure 1.3.2 and 1.3.3.

http://tech-kid.com/nokia-phone.html *http://www.northstandchat.com/showthread.php?289468-Your-first-ever- Mobile -Phone-and-what-make-and-model-was-it-!/page3*

Figure 1.3.2 Nokia 1011 supports only GSM900 single band band. *Figure1.3.3 Motorola m300 support only GSM1800 single*

In 1997, Motorola produced a mobile device Motorola mr601 which was the first dual band GSM phone, it supported GSM900 and GSM1800 dual band and its antenna consists of two antennas helix antenna which has travelling wave in the shape of corkscrew with circularly polarized and whip antenna which can be consider as dipole antenna and it is Omni-directional radiation pattern. That's model phone offered the ability to access network in over 70 countries by the end 1997.

In the first dual band PIFA operating at GSM900 and GSM1800, invented by Prof. Peter Hall in 1996 U.K, the first dual band PIFA a slot with a certain geometric dimensions, lead to support two different bands.

The development timeline of mobile phone has several significant steps, one of them was in 1999, when Nokia launched Nokia3210 as shown in figure 1.3.4 the first mobile with fully internal antenna, support both GSM900 and GSM1800 dual band, and was one of the most popular handset with over 160 million being sold [2].

In 1999, Prof. Peter Hall comes back with the first Triple Band PIFA which operates at GSM800/1800/1900. He designed Triple Band PIFA with two slots in the ground plane of the PIFA antenna, that's let the antenna to transmit and receive different band of frequencies.

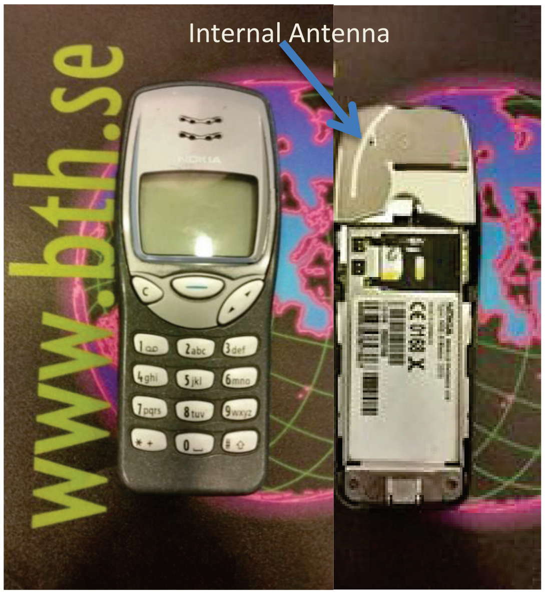
Internal Antenna
www.bth.se

Figure 1.3.4 Nokia 3210 and its internal antenna which is installed to the top left of the back view of the mobile phone.

The first commercial 3G network was launched by Hutchison Telecommunications branded as 3G in June 2003, and in 2004 Nokia released Nokia6630,as showing in figure1.5 which was the first mobile phones allows truly global roaming, and first mobile with dual mode tri-band, GSM900/1800/1900 and UMTS 2100.

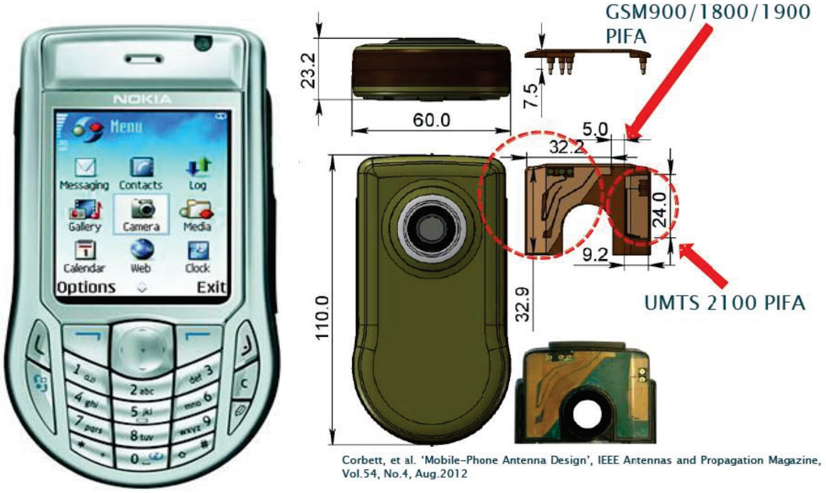
GSM900/1800/1900
PIFA
23.2
7.5
60.0
5.0
32.2
24.0
9.2
32.9
110.0
UMTS 2100 PIFA

Corbett, et al. 'Mobile-Phone Antenna Design', IEEE Antennas and Propagation Magazine, Vol.54, No.4, Aug.2012

The photo to the left was loaded from http://www.mondomobileblog.com/2010/03/31/codici-segreti-nokia-6630/

Figure 1.3.5 Nokia6630 mobile and its internal antennas.

Table1.1 The timeline of development antenna use in mobile phones [1].

Year	**1900**	**1950**	**1970**	**1990**	**2000**	**2010**
Frequency	< 10 kHz	< 30 MHz	< 800 MHz	< 1.9 GHz	< 2 GHz	3 GHz 5 GHz
System	Telegraph / telephone for train, ship, police cars	Vehicle and portable systems for business, Pager	Mobile phones (analog), cordless phones, pagers, GPS	Mobile phones (digital),aircraft (voice and data) personal phone	Mobile phones (multimedia), Wireless access, Bluetooth	Mobile phones (high data rate), Ultra wideband,
Antenna	Monopole dipole, whip,	Blades, ferrite coil, helical antenna	Corner reflector, PIFA, helix	Meander line, normal mode helix, ceramic chip, adaptive array	Wideband, multiband built-in antennas, adaptive array, MIMO	Small compact functional antennas, implant antennas

Table 1.3.1 shows the time line of development in both sides' frequencies and antenna which are used for mobile phones systems. The frequencies are raised from 10 kHz till 3 or 6 GHz which is considered as a 4G system.

Cordless phones navigation systems and multichannel access systems have contributed dramatically to the development of antennas design, while the personalization of mobile terminals raised the needs of small antennas. Other needs for small antennas are hybrid systems, the handset of multiband GSM combined with DECT (Digital Enhanced Cordless Telecommunications) for example.

The specifications of the handset small antennas are low profile, compact and lightweight, which has led to modify and develop the antennas. One of the antennas is a modified PIFA, which has changed that it is difficult to recognize it as an original PIFA. The modified PIFA was first used in GSM handsets and has become more common in several different mobile terminals. Other modified antennas such as monopole were used in PDC (Personal Digital Cellular) with a normal mode helical antenna.

In modern mobile phone, we can conclude a summary of the considerations in design process as following:

- The requirement of multiband and built–in antenna.
- The requirement of small size antenna due to small size mobile terminals, for example RFID units (Radio-frequency identification).
- The requirement of function antennas such as adaptive antennas and MIMO for transmission high data rate.

Printed antenna has been used in mobile terminals; also Printed antennas have the smallest dimensions comparing with other installed small antennas. Chip antennas or a printed antenna consists of dielectric material and a ground plane which is a part of the antenna. Chip antennas offer a high gain, appropriate radiation pattern, and high efficiency.

Chapter 2

Basic Mobile Phone Antennas Principles

2.1 Parameters of Mobile Phone Antennas

Radiation pattern

Radiation pattern can be describe by power radiated through the antenna, which depends on the direction, and it can be represented as graphical or mathematical function of the radiation properties.

There are two common types of antenna pattern , the first one is a power pattern and the second is field pattern , the first type is considering the power radiated and the position of the spherical coordinate and the second type which depend on the field if we have electrical field or magnetic field with the position of spherical coordinate.

The field region of the antenna pattern can be dived to three regions. The first region is Reactive Near Field Region which is the closest region to the antenna so it gives possibility to control the energy and standing waves. The second region which is locate between the Reactive near field region and the far field region where control the radiation field, The last region is the far field region which has a far distance of the antenna , as showing in figure 2.1.1.

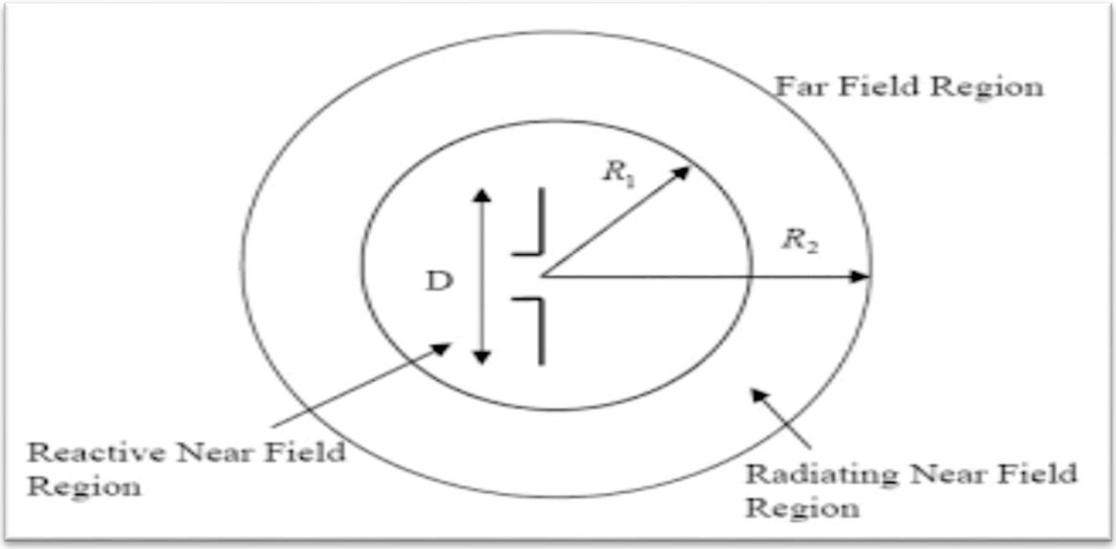

http://antenna-practices.blogspot.se/2008/11/antenna.html

Figure 2.1.1 The radiation pattern and the field regions.

There are many types of antenna pattern

1- Isotropic pattern the radiation of the isotropic is symmetric in all directions which is the ideal antenna but it is not found until these days.
2- Directional pattern which radiate more focusing with one direction.
3- Plane pattern which consist of two principles the electrical field plane (E-plane) and the magnetic field (H-plane).

What is (E-plane) and (H-plane)?

E-plane consists of electrical field vector combined with the direction of maximum radiation.

H-plane consists of magnetic field vector combined with the direction of maximum radiation.

Parameters of antenna pattern

1- Main lobe is the radiation lobe going in the same direction of the maximum radiation.
2- Side Lobe is the radiation lobe of any directional lobe except the intended direction.
3- Back lope which has the opposite direction of the main lobe.
4- Half-Power Beam width (HPBW) the angular between two directions the half power point and the beam width which equal to -3db.
5- First Null Beam width (FNBW) - angular width between two directions the first zero and other side in the side lobe.

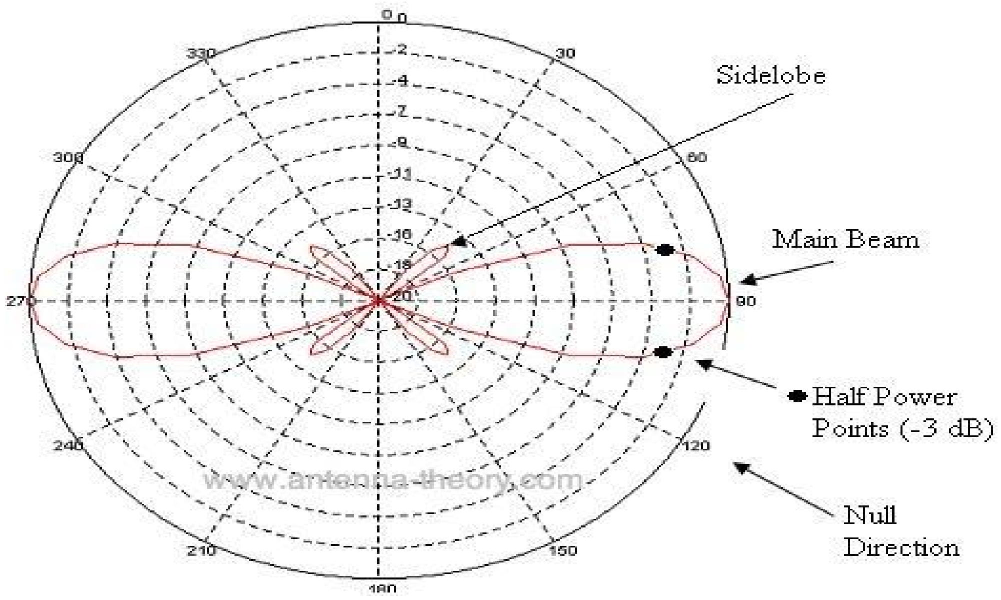

http://www.antenna-theory.com/basics/whyantennasradiate.php

Figure 2.1.2 A polar figure of an antenna with performance Parameters.

Directivity

Directivity is one of the antenna parameters that guide us to measure the direction of the radiation pattern, the directivity help us to determine the efficiency of the antenna, directivity of isotropic antenna equal to 1 (0 dB).

The directivity can be calculated by determining the maximum value of the magnitude of the radiation pattern which is defined by F(θ, ϕ) and the average power of all the direction which is equal to 1 so, the formula of calculating the directivity it can be written as:

$$\boldsymbol{D = \frac{1}{\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}|F(\theta,\phi)|^2 \sin\theta d\theta d\phi}} \quad \textbf{(2.1.1)}$$

$\theta =$ The angle between the z-axis and the vector from the origin to the point

$\phi =$ The angle between the x-axis and the projection onto the x-y plane.

Mobile phone antenna should have a low directivity, because is not useful to use an antenna with high directivity, that is the received signal it can be from different directions. On the other hand the base station needs a high directivity.

Table 2.1.1 Directivity of different types of antennas [4].

Antenna Type	Typical Directivity	Typical Directivity (dB)
Short Dipole Antenna	1.5	1.76
Half-Wave Dipole Antenna	1.64	2.15
Patch (Microstrip) Antenna	3.2-6.3	5-8
Horn Antenna	10-100	10-20
Dish Antenna	10-10,000	10-40

As showing in the table 2.1.1 above the directivity of different types of antenna so, the properties of the antenna have to be chosen regarding to what we need, as we have mentioned if you want to choose antenna for mobile station as cell phone, car or computer we will choose antenna with low directivity.

Polarization

One of the characteristics for any types of antennas is the direction of the electrical field vector of the electromagnetic wave. The polarizations of electromagnetic wave depend on the electrical field, which is perpendicular with the magnetic field.

The polarization of antennas can be classified in three different types based on electrical field

1- **Linear polarization:** as seen in the figure 2.1.3 the electrical field travels with the same direction of the plane wave. There are two types of linear polarization:
   - **Vertical polarization:** the electrical field travelling orthogonally to the y-axis and the magnetic field is travelling horizontally to x-axis.
   - **Horizontal polarization:** the electrical field is travelling horizontally to the x-axis and the magnetic field is travelling orthogonally to the y-axis.

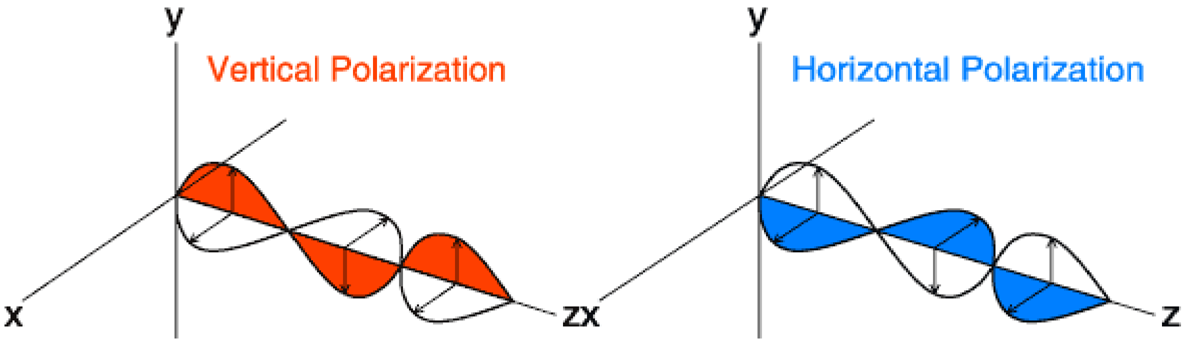

http://www.srh.noaa.gov/hun/?n=dualpol

Figure 2.1.3 Vertical polarization and horizontal polarization.

2- **Circular polarization:** the electrical field travels circularly as seen in figure 2.1.4. It can be divided to two types regarding to direction of the circle:
   - Right Circular polarization: the direction of the circle is clockwise.
   - Left circular polarization: the direction of the circle is anti-clockwise.

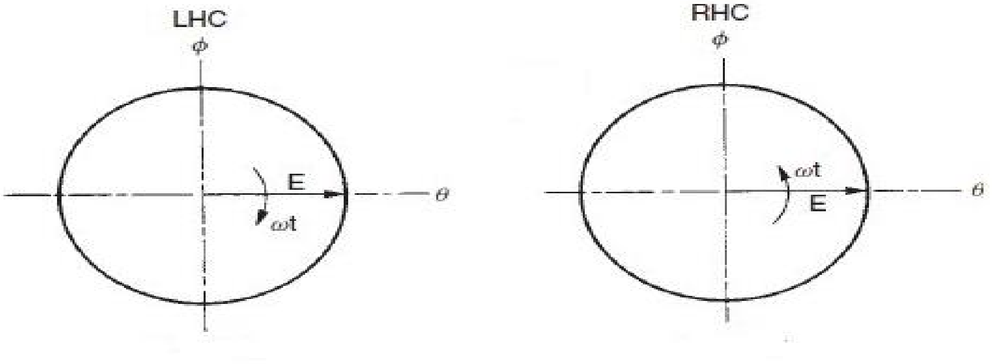
LHC
ϕ
E
ωt
θ
RHC
ϕ
ωt
E
θ

http://www.antetec.com/news/2010/01/27-371.html

Figure 2.1.4 Right circular polarization and left circular polarization.

3- **Elliptical polarization:** the electrical field travels with an elliptical shape based on the direction of the electrical field figure 2.1.5 it can be divided to two types:

- Right elliptical polarization which is has direction with clockwise.
- Left elliptical polarization which has direction anti-clockwise.

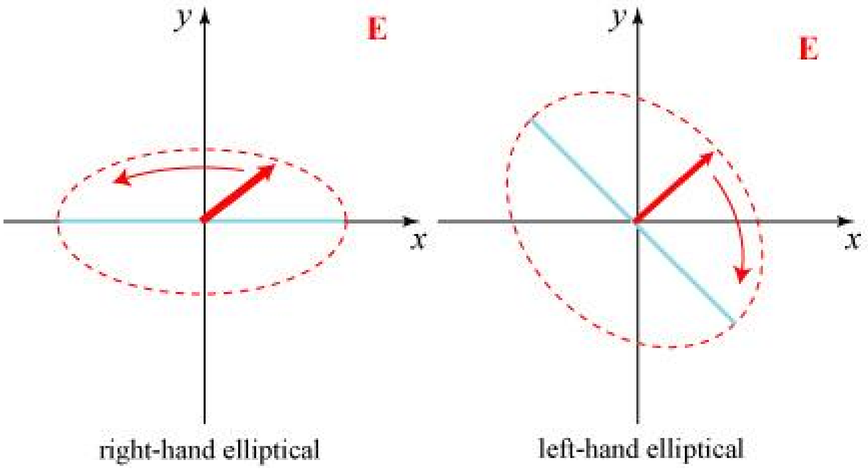
y
E
x
right-hand elliptical
y
E
x
left-hand elliptical

http://skullsinthestars.com/2012/09/23/hairy-balls-in-optics/

Figure 2.1.5 Right elliptical polarization and left elliptical polarization.

Impedance

The impedance of the antenna mainly depends on the input voltage and current for the power losses equal to the power transmission of the antenna. The benefits of knowing the input impedance it is lead you that to choose the suitable transmission line.

The impedance of an antenna ZA must be matched with its impedance transmission line ZS (feeder). Otherwise there will be amount of a reflection wave back to the signal source. To have maximum power at the antenna from source as shown in figure 2.1.6 ZS should be equal to ZA

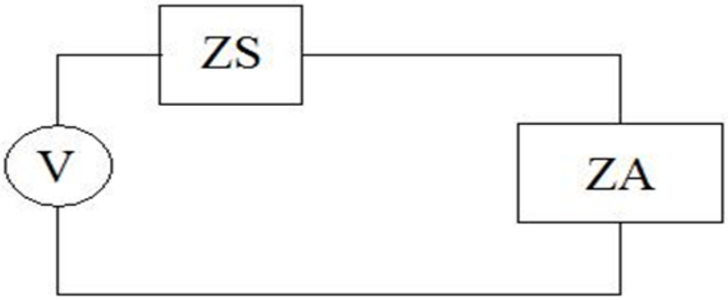

http://www.antenna-theory.com/basics/whyantennasradiate.php

Figure 2.1.6 The equivalent circuit of an antenna and the transmission line.

Antenna gain

One of the antenna properties to describe the value of the power transmission to the peak radiation direction. The efficiency of the antenna always relate with the antenna gain so if the efficiency is perfect 100% that’s mean the gain and the directivity are equal to each other, it can be describe by the following formula:

$$G = \varepsilon_R D \qquad (2.1.2)$$

Antenna efficiency

The relation between the input power to the antenna and the dissipated power which is the losses power within antenna, regarding to the relation between the power radiate and the power input the efficiency of the antenna can be written by the following formula:

$$\varepsilon_R = \frac{P_r}{P_{in}} \qquad \textbf{(2.1.3)}$$

Antenna Beamwidth (θ)

The angle between the two points of the half power of the radiation pattern, the beamwidh could be clarify as the half power, the directivity become high if the beamwitdh very small. The beamwidth is found by determining the two points with 3db as showing in figure 2.1.7.

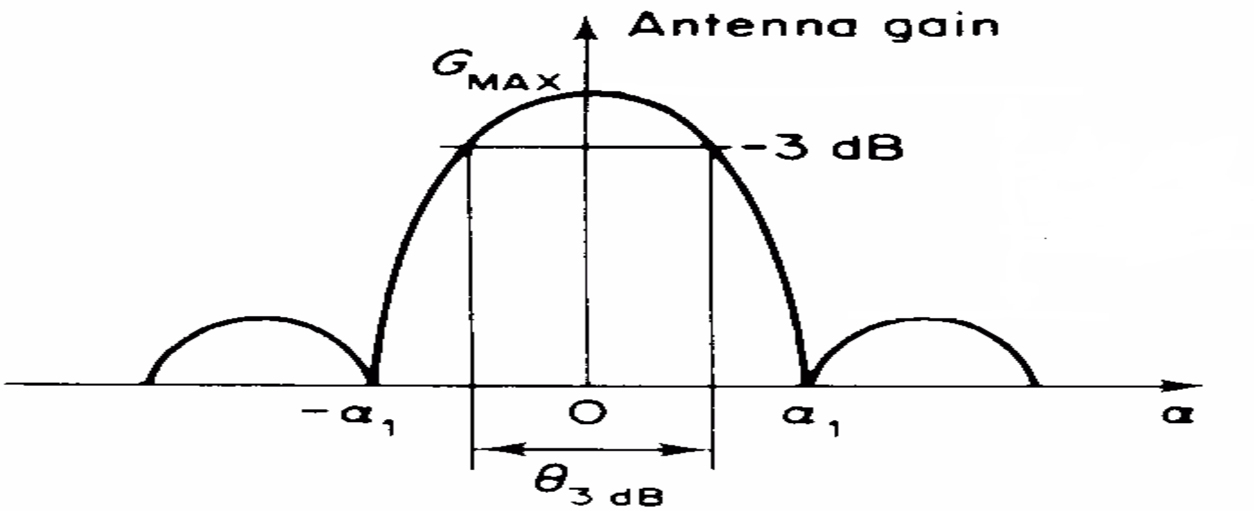

http://happy.emu.id.au/lab/rep/rep/9510/txtspace/9510_010.htm

Figure 2.1.7 Half power Beamwidth.

Effective area

To calculate the transmission power (P_t) and determine how much power has been lost due to the plane wave, regarding to the relation between the power density (P) and the effective area (A_e) the following formula can be written as:

$$P_t = PA_e \qquad \textbf{(2.1.4)}$$

To calculate the effective area we should consider the relation between the effective area and the antenna gain the formula can be written as:

$$A_e = \frac{\lambda^2}{4\pi} G \qquad \textbf{(2.1.5)}$$

Bandwidth

One of the other important parameters is bandwidth which determines the range of the frequencies where the antenna is operating, table 2.1.2 shows different common types of antennas with frequency range of each antenna. The percentage of the bandwidth can be calculated as following:

$$Bandwidth\% = \frac{f_{max} - f_{min}}{f_{Center}} \times 100 \qquad \textbf{(2.1.6)}$$

Table 2.1.2 different common types of antennas with frequency range of each antenna [4].

Antenna	Center frequency	Frequency range	Percentage bandwidth
Patch	1000MHz	985-1015MHz	3%
Dipole	1000MHz	960-1040MHz	8%
Horn	1000MHz	154-1848MHz	169.40%
Spiral	1000MHz	95-1900MHz	180.50%

VSWR

Voltage standing wave ration or S_{11} which describe the reflection power coefficient or the return losses of the antenna, the VSWR can be defined by the following formula:

$$\boldsymbol{VSWR = \frac{1+|\Gamma|}{1-|\Gamma|}} \qquad \textbf{(2.1.7)}$$

Γ : the reflection coefficient

If the voltage standing wave has a small value that means the value of the power transmission will deliver to the antenna better but does not necessarily mean the power delivered sometimes also radiated. The value of the ideal VSWR is 1 and that's will happen if the voltage of the transmission wave matched in this situation we will not have a return losses on the other hand if the voltage in the transmission waves haven't matched with the antenna that's mean there is some losses.

VSWR is a measure of how much power is delivered by a voltage source to an antenna. Which means that's an antenna dose not radiate all the received power; there is always amount of reflection voltage due to mismatching network.

Based on the VSWR concept we find that an antenna with VSWR less than 2 that's good enough to have a good impedance matching, on the other hand the transmission of more power to the radio can be lead to damage the radio.

2.2 Antenna types

Dipole Antenna

One of the most common antennas today in radio communications is a dipole antenna due to its simplicity. Dipole antenna has two rods or wires (poles) which are identical conductive elements, and fed by a signal source or feed energy. As shown in figure 2.2.1 below the two elements, current, and the voltage causes electromagnetic wave.

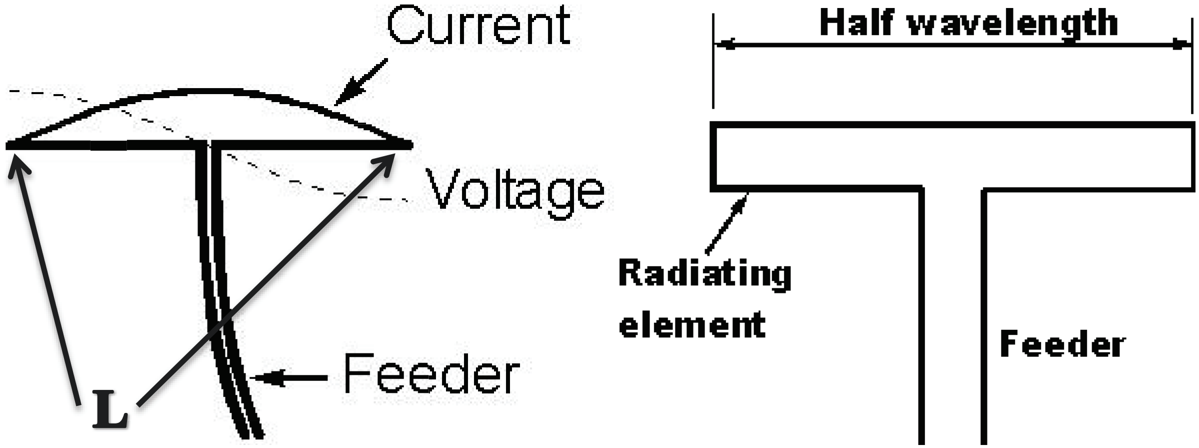

http://www.radio-electronics.com/info/antennas/dipole/feed//

http://www.radio-electronics.com/info/antennas/dipole/folded_dipole.php

Figure 2.2.1 The half wave dipole antenna with center feed point, and to the right Half-wave folded dipole antenna.

Dipole antenna types

Half wave dipole antenna is the most popular type. But still there several types which could be used.

- **Short dipole**
  A dipole antenna is called a short dipole antenna if its rods or wires length (L) shorter than half of a wavelength (λ), for example if L< λ/10.
- **Half wave dipole antenna**
  As shown in figure # above a half wave dipole antenna consists of two wires each of them quarter of the wavelength, and the feeder is connecting to both wires.
- **Multiple half wave dipole antenna**
  A dipole antenna with odd multiple of half wavelength long.

- **Folded dipole antenna**
  A folded dipole antenna is half wavelength dipole but its wires are folded back as shown in figure 2.2.1.

Monopole antenna

Unlikely dipole antenna Monopole antenna is an antenna which consists of a one straight rod conductor, and is installed over a ground plane (conductive surface) as seen in figure 2.2.2. Monopole antennas are fed at lower end, near to ground plane which works as a reflector. In case the ground plane conductive and valid in size, the efficiency of the ground plane is as well as a vertically installed dipole.

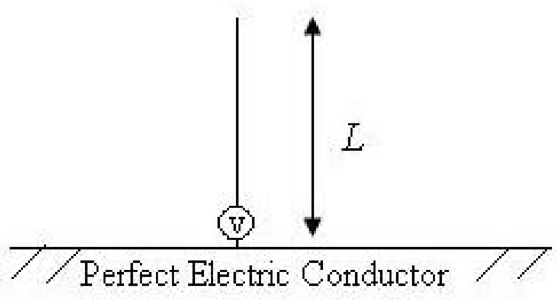

http://www.antenna-theory.com/antennas/monopole.php

Figure 2.2.2 Monopole antenna with length L.

There are other common types of monopole antenna such as Whip, helical, T-antenna, and mast radiator. Monopole antenna like dipole antenna both have omnidirectional radiation pattern, for that the monopole it could be used in cell phones and other applications, for example indoor applications such as airplane, or shopping center.

Monopole antennas are widely used in mobile communications, specially the quarter wave monopole -the length of the monopole L is approximately quarter of wavelength- which is a suitable for MIMO wireless communications systems, due to its low angle radiation and minimized ground losses [3].

For better efficiency of monopole antennas, many methods are applied, one of them type of the ground plane; The ground plane can be infinite or finite with different geometric shape such as spherical, cylindrical, or rectangular sheet. That's affect characteristics monopole specially radiation pattern, and less effects in monopole's impedance [4].

Based on fact that, monopole antennas are half the size (half length) of their dipole. So we can describe the directivity of monopole in term of its dipole, for example if the directivity of a dipole of length 2L has a directivity D1 [decibels], then the directivity of monopole with length L will be D1+3 [decibels], this because there is no radiation happens below the ground plane [4].

PIFA antenna

PIFA is the short name of Planar inverted-F antennas which is the most common type of cellular phone antenna nowadays, it has the same structure of microstrip antenna but it shorted with the ground because of this short pin the PIFA produce a quarter wave length. The main reason of name PIFA because has change of using wire to use ground plane. The PIFA structure consist of a rectangular planar which is installed on the ground plane, it is always parallel to reduce the height. Figure 2.2.3 is showing the geometric dimensions of the PIFA antenna the length which is L, The width which symbolized by W1, The short pin is W2, for height is h and the distance between the short pin and the feed point is D.

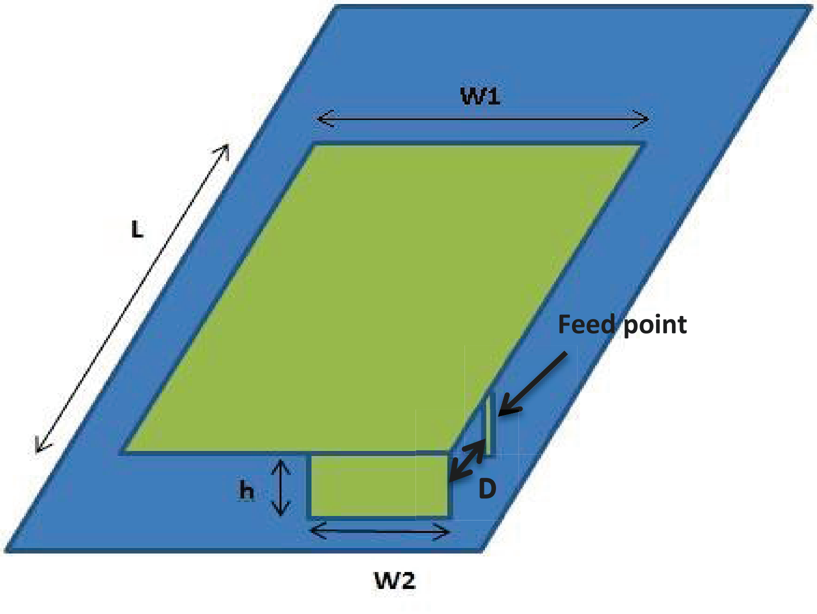

http://www.raymaps.com/index.php/planar-inverted-f-antenna-pifa/

Figure 2.2.3 PIFA antenna and its geometric dimensions.

The structure of the PIFA can deal with different frequency band

- Single band
- Multiband
- Reconfigurable

The main advantages of using PIFA antenna

1- Due to the small size of PIFA antenna it gives the availability to insert the antenna inside the cellular phone.

2- PIFA has availability to get high gain for both polarization states vertically and horizontally there for it can receive the reflection wave easily from different directions.

3- The backward radiation of the PIFA has be reduced that’s mean the electromagnetic waves power has reduced which lead us to get a good deal to less damage of the human health.

4- The design and the material of the PIFA are not costly on top of that it has a very high efficiency also It is easy to fabricate.

As we know the development of communication always come up with a new evaluation but until now there are no antennas can work without disadvantages, So there is a problem in PIFA antenna could not be solved until now which is the sensitivity of the bandwidth but in general PIFA prove that is the best cellar phone antenna have been used regarding to the comparing between the advantage and the disadvantage of the antenna.

There are a few di-electrical material can be used in PIFA, and when the air is the one which is used, the gain will be better comparing with other material.

Fractal antenna

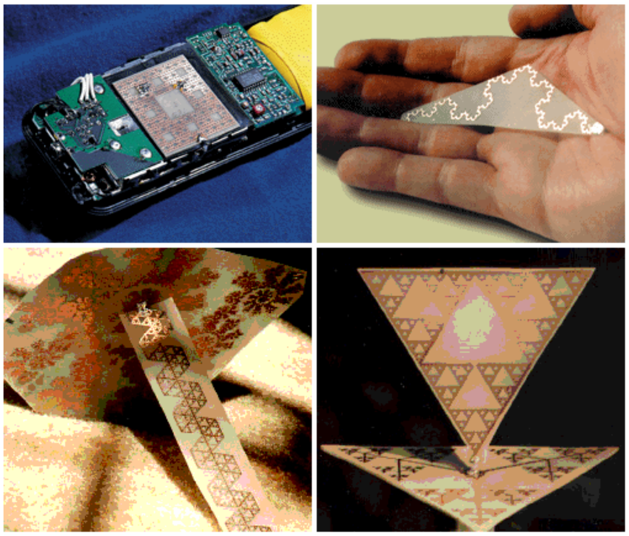

Photo was loaded from http://www.antenna-theory.com/antennas/fractal.php

Figure 2.2.4 Triangle Fractal Geometry. Upper Left: Patch Antenna Fractal. Lower Left: Monopole Fractal. Lower Right: Bowtie Monopole Fractal Antenna.

One of the problems of popular designs of antenna, that's the antenna which is smaller than a quarter of the wavelength, is very sensitive to a narrow band of frequencies. This problem is a significant issue for mobile phone, where the requirements are always to minimize the size of the mobile phone which leads to the problem of sensitive narrow band frequencies.

As shown in figure 2.2.4 the fractal antenna has a complex shape which offers many electric current modes to exist, that's causes, the increasing in radiation, consequently very wide bandwidth.

Fractal antenna designs can solve sensitive narrow band frequencies problem. Experiments have shown that antennas built with only a small number of iterations of a fractal process can exhibit sensitivity at several frequencies, and it is worthy to mention that's fractal antenna can work efficiently more at one quarter of the wavelength than other antennas types [5].

Geometry of fractals

The shape of a fractal can be formed by iterative mathematical process called Iterative Function Systems (IFS) [6] and by using affine transformations as following:

$$W(x, y) = (ax+by+e,\ cx+dy+f)$$

a,b,c,d control rotation and scaling e , f control linear translation by repeating the upper calculations, we get the IFS sequence that converges to the final image.

Chapter 3

Mobile phones antenna design

Mobile phones consist of several antennas for several purposes such as Wi-Fi antenna, GPS antenna, low and high frequencies antennas. Each type of antennas is designed according to certain considerations, of course, the small sizes the main issue.

As it is mentioned in the previous sections of this thesis. Dipole antenna is a useful type of antennas; easy to be constructed, and integrated in cell phones

.3.1 Mobile phone GPS antenna design

GPS antenna is an omnidirectional small antenna with only receive mode that can be one of main design's considerations. The GPS frequency is 1.575 GHz, with practically no bandwidth, and the GPS satellites are RHCP (table 3.1.1) to avoid Faraday rotation problems, and there is one other benefit, that's circular polarization does not need rotational alignment of a circularly polarized antenna at mobile phone.

For a maximum received power the user terminal's antenna must be RHCP. User terminal's antenna which is linearly polarized will have a loss 3dB in received power due to polarization issue. By looking at the power density of the GPS received wave. We find that it is extremely low level which equals to -160dB. Therefore, we need an efficient antenna at the receiver.

Table 3.1.1 Cell phone GPS antenna specifications and goals [4].

Transmit and receive frequency	**Receive only 1.575 GHz**
Efficiency	-3 dB to -9 dB
Polarization	RHCP
Quarter wavelength(L)	About 4.75cm or 1.87 inches

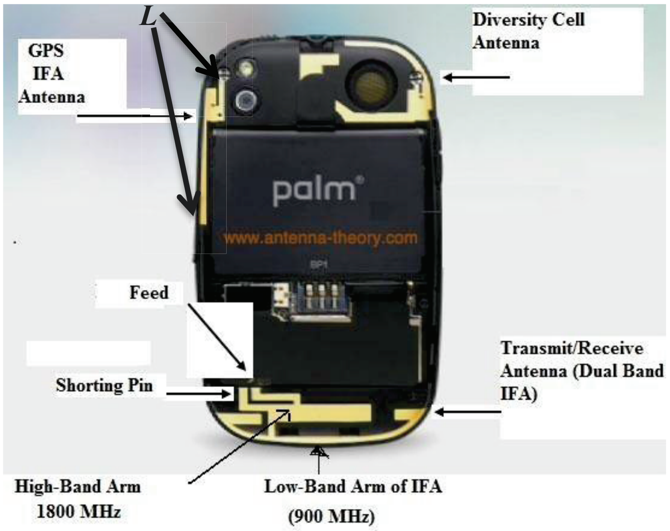
L
GPS IFA Antenna
Diversity Cell Antenna
palm
www.antenna-theory.com
Feed
Shorting Pin
Transmit/Receive Antenna (Dual Band IFA)
High-Band Arm 1800 MHz
Low-Band Arm of IFA (900 MHz)

http://www.antenna-theory.com/antennas/aperture/ifa.php

Figure 3.1.1 Palm Pre Antennas viewable by Removing Back Cover.

As shown in figure 3.1.1 in this design, quarter wavelength IFA antenna is used. The ground plane (the chassis) of the mobile phone is considered as one arm of dipole antenna and the feeder is placed on opposite side at a point which helps for matching impedance. The GPS antenna is placed on the top of the mobile phone to guarantee best performance; that's because the users hold the mobile phone vertically when they want check current location for example, and in that case their hands will be on lower part of mobile phone, far from GPS antenna.

As obvious in figure 3.1.1 IFA antenna is designed with its geometric dimensions to get the better efficiency, and of course to take into account the available low profile for the antenna in mobile phone.

To obtain circularly polarized dipole antenna, we use two crossed dipoles to provide the two orthogonal field components, and if the two dipoles are fed with a 90° time-phase difference the polarization will be circular along zenith [28].

Single band PIFA Antenna design

GPS antennas are received mode antennas. So we do not have to worry about SAR measurements. In this section we will design a single band antenna for GPS use and simulate the model by COMSOL.

Model descriptions

The desired bandwidth is a single frequency 1.575 GHz (table 3.1.1). Thus we try to make the resonant frequency of the PIFA antenna close to 1.575 GHz to provide high receiving efficiency as much as possible. The gain of the antenna should be between -3dB to 0dB.

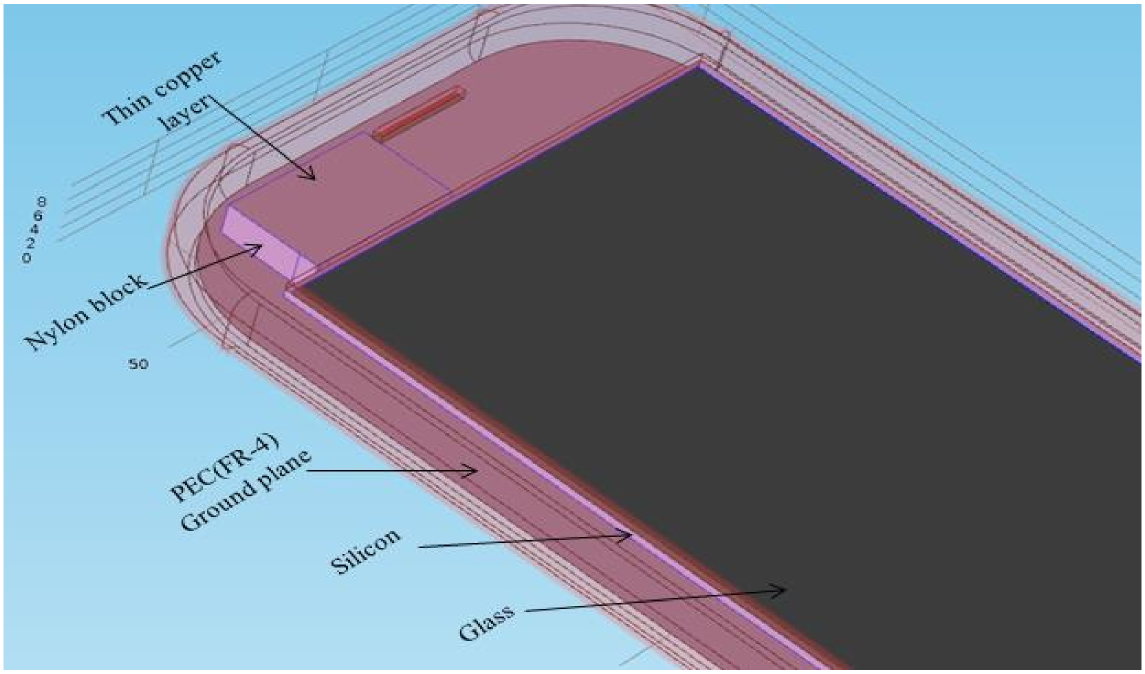

Figure 3.1.2 Zoom view of PIFA antenna block. It consists of Nylon block and the radiating part and the materials which are used in design.

Calculations

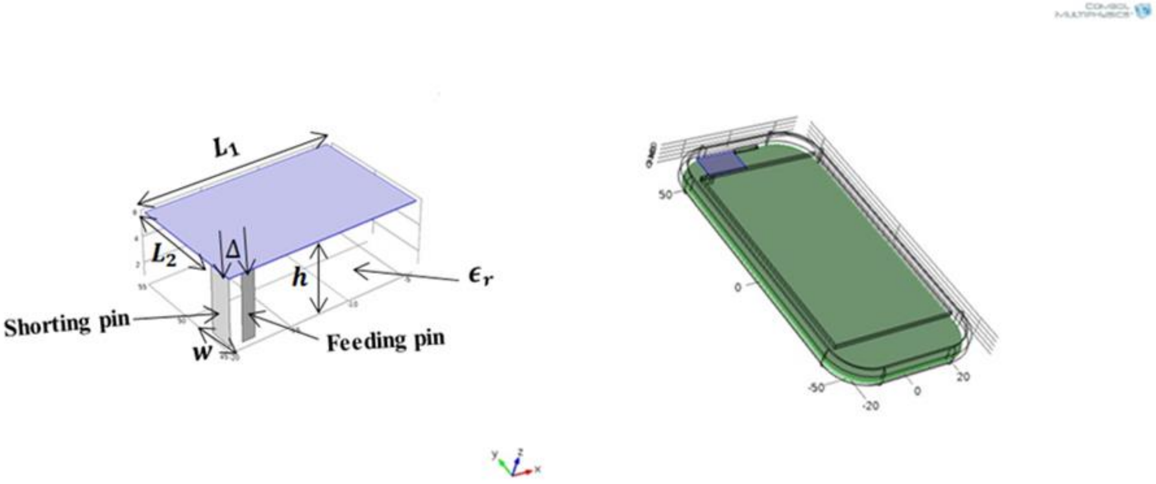

Figure 3.1.3 The antenna to the left and its dimensions $L_1 = 20mm, L_2 = 10mm, w = 2mm$, $h = 4mm$ $\epsilon_r = 3.8$, $\Delta = 0.2mm$ *to the right the casing's dimensions have length =119mm, width =60mm.*

The resonant wavelength of a PIFA antenna can be calculated as following:

$$\boldsymbol{L_1 + L_2 - W = \frac{\lambda_0}{4}} \qquad \textbf{(3.1.1)}$$

And the relation between the resonant wavelength and the resonant frequency can be determined by the equation:

$$\boldsymbol{\lambda_0 = \frac{c_0}{f_0\sqrt{\epsilon_r}}} \qquad \textbf{(3.1.2)}$$

λ_0: Resonant wavelength of PIFA.

c_0: The speed of light in space.

f_0: Resonant frequency.

ϵ_r: Relative permittivity.

Mathematically, PIFA antenna's size can be reduced by using factor $\sqrt{\epsilon_r}$ (equation 3.1.2)

Therefore if we use a material with higher ϵ_r ,that reduces the size of the PIFA which is a useful way to get a low profile antenna but that could affect the gain of the antenna as well. It is a big challenge to keep the efficiency without any effects when the small size is required, so the PIFA dimensions which are used in this design were chosen after many simulations, and based on the best efficiency, lowest value in return loss S11, and the smallest size. We determined the dimensions.

By applying equation 3.1.1 we find that:

$20 + 10 - 2 = \frac{\lambda_0}{4}$, λ_0 = 112 mm

Hence,

$$f_0 = \frac{3 \times 10^8}{112 \times 10^{-3}\sqrt{3.8}} = 1.374 \text{ GHz}$$

The calculated resonant frequency is not very close too much to the desired frequency 1.575 GHz. But by increasing and decreasing the gap impedance change, the resonant frequency will shift to upper or lower than the calculated resonant frequency. Increasing and decreasing of the impedance gap affects the input impedance, consequently effects at VSWR for the resonant frequency.

Simulations

In simulation, first we installed the antenna to left of the casing figure 3.1.3 and inserted the materials as shown in figure 3.1.2 and with dielectric constant as table 3.1.2 in addition the outer shell of the casing is simulated with Polytetrafluoroethylene (PTFE). We use whole chassis as a ground plane for better efficiency.

A 50 Ω lumped port is used to excite the antenna and determine the input impedance. The lumped port is mounted between two metallic boundaries: the vertical feeding and the ground plane of FR4 board as seen in figure 3.1.2. The distance Δ, the impedance matching gap affects significantly the matching impedance. So another strip shorted to the ground plane is added. The dielectrically material which will be used for the antenna block is Nylon.

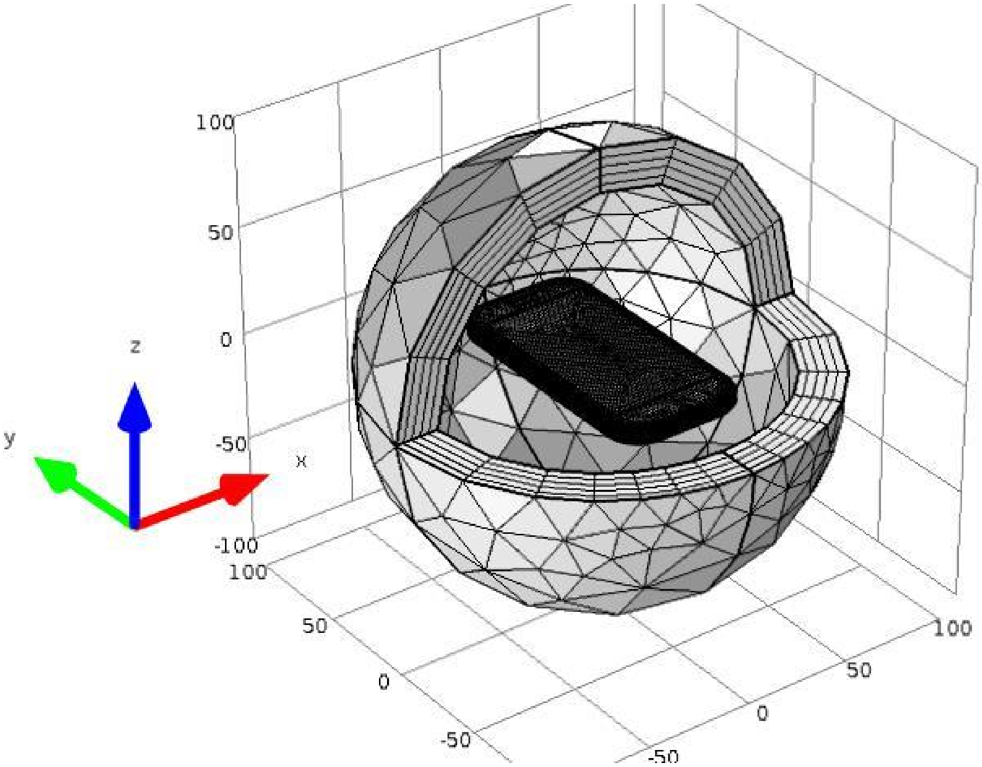

Figure 3.1.4 Mesh model for the whole simulated model: PIFA antenna, casing of the mobile phone and the surrounded perfect matched layers.

The model consists of several domains; every domain has a certain material figure 3.1.2. In addition there is one more domain encloses the casing. This domain is simulated as a sphere with radius 100mm, material air, and has five perfect matched layers in order for the radiation to be able to travel anywhere as shown in figure 3.1.2. The metal part of the antenna element at frequency 1.575 GHz can be modeled using perfect electric conductor boundaries.

Table 3.1.2 the materials of the simulated model and its Dielectric constant according COMSOL's materials libraries except Nylon.

Material	**Dielectric constant (ϵ_r)**
Air	1
FR-4	4.5
Nylon	3.8*
Glass (quartz)	4.2
Silicon	11.7
PTEF	2.1

* The Dielectric constant of Nylon accordinghttp://www.professionalplastics.com/professionalplastics/ElectricalPropertiesofPlastics.pdf

Results and Analysis

E-field norm on the xy-plane slice is shown below in figure 3.1.5. The plot shows that the electric field is strong at one of the top metallic surface shell far from the feeding point. This looks alike the E-field distribution of a quarter wavelength monopole antenna, which the PIFA derived from.

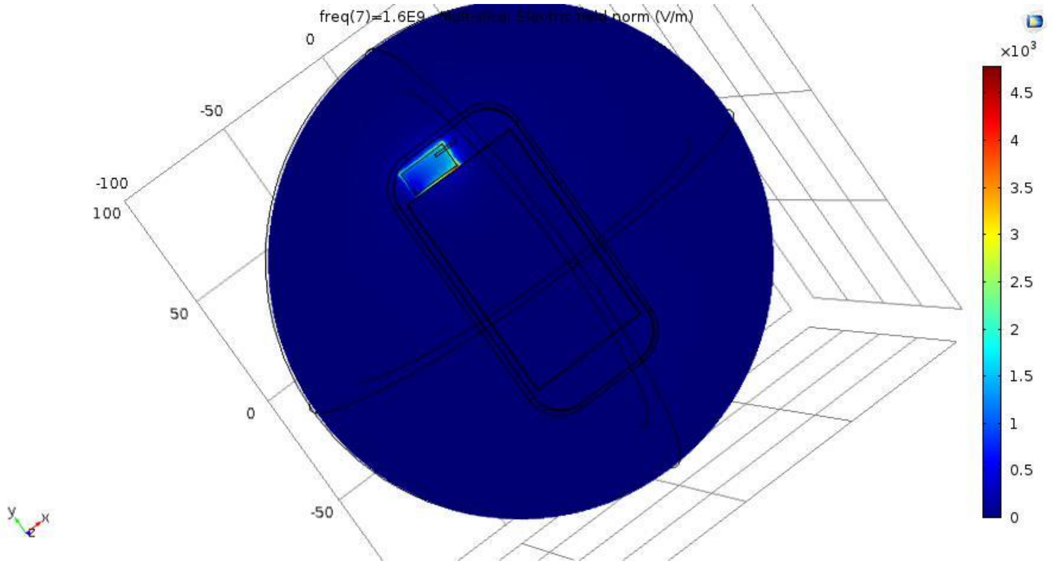

Figure 3.1.5 E-field norm distribution on the top of the PIFA.

The polar plot of the far field radiation pattern of the antenna is shown in figure 3.1.6 As obvious the antenna gain on xy-plane varies from about -6dBi to 1.5dBi. The azimuthal radiation pattern is not Omni-directional any more, since the antenna is mounted on the ground plane and miniaturized.

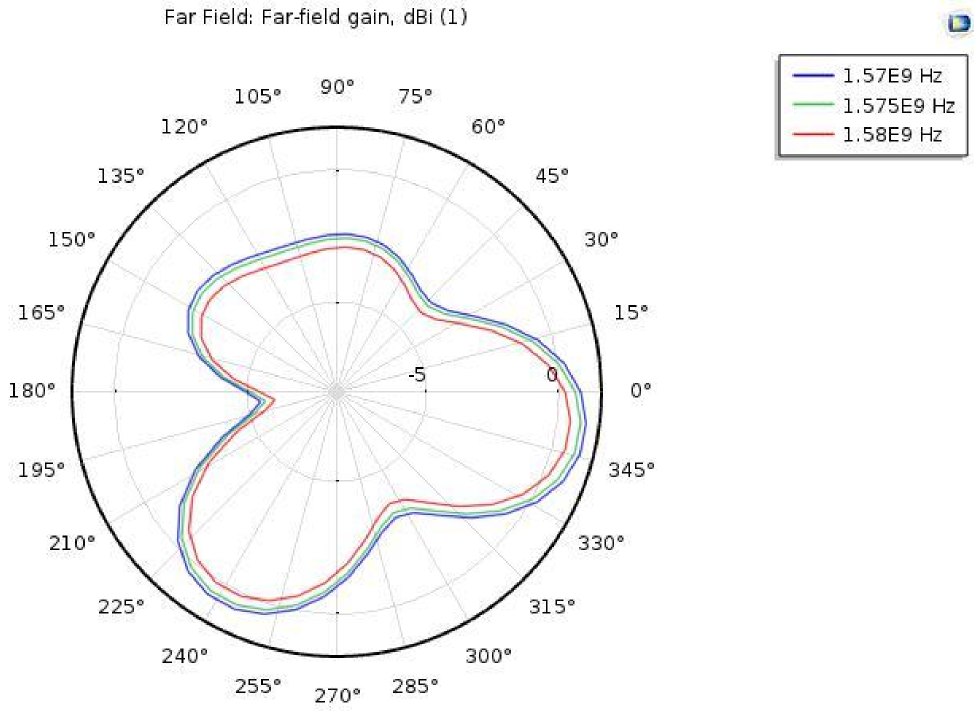

Figure 3.1.6 Antenna gain pattern on the xy-plane.

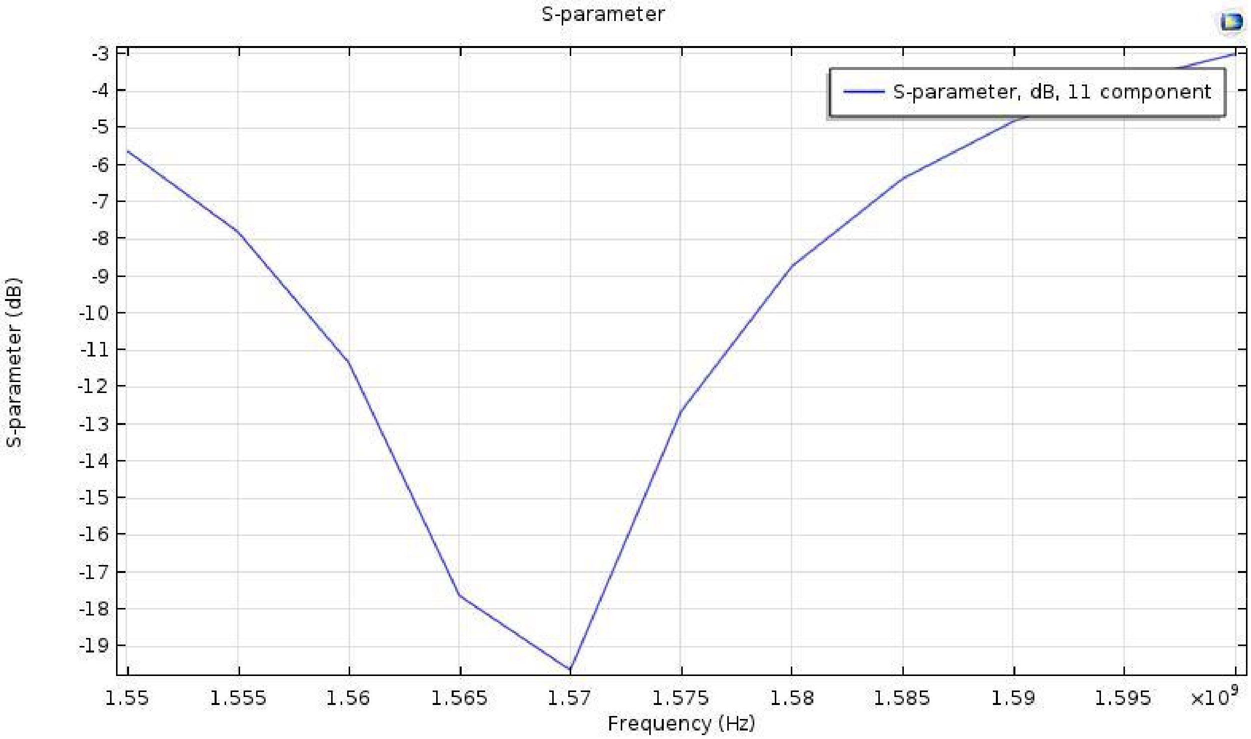

Figure 3.1.7 Return loss S-parameters (S11) for the antenna by simulation.

S-parameters (S11) measurements indicate that at 1.575GHz is -13dB which means that the reflected power is 5%. This describes how well the antenna input impedance is matched to the 50 Ω reference impedance.

Antenna's bandwidth regarding figure 3.1.7 is a narrow bandwidth:

$$Bandwidth \approx \frac{1.577 - 1.557}{1.570} \times 100 = 1.2\%$$

The wide bandwidth for GPS antenna is not required. So the bandwidth above is sufficient.

Keeping the feeding point near to the shorting pin as much as possible reduces the antenna size but that causes a narrow bandwidth. The electrical characteristics of PIFA are affected by the size of ground plane. By varying the size of ground plane, the efficiency and the bandwidth are changed; to increase the bandwidth we need the whole possible free space in the chassis. In simulation we use the whole chassis as ground plane in order to get a better efficiency.

The effective of the PIFA antenna geometric dimensions on the resonant frequency can be summarized by the following table 3.1.3

Table 3.1.3 Effect of geometric dimensions on resonant frequency.

	L_1	L_2	W	H
f_0↑	↓	↓	↑	↓

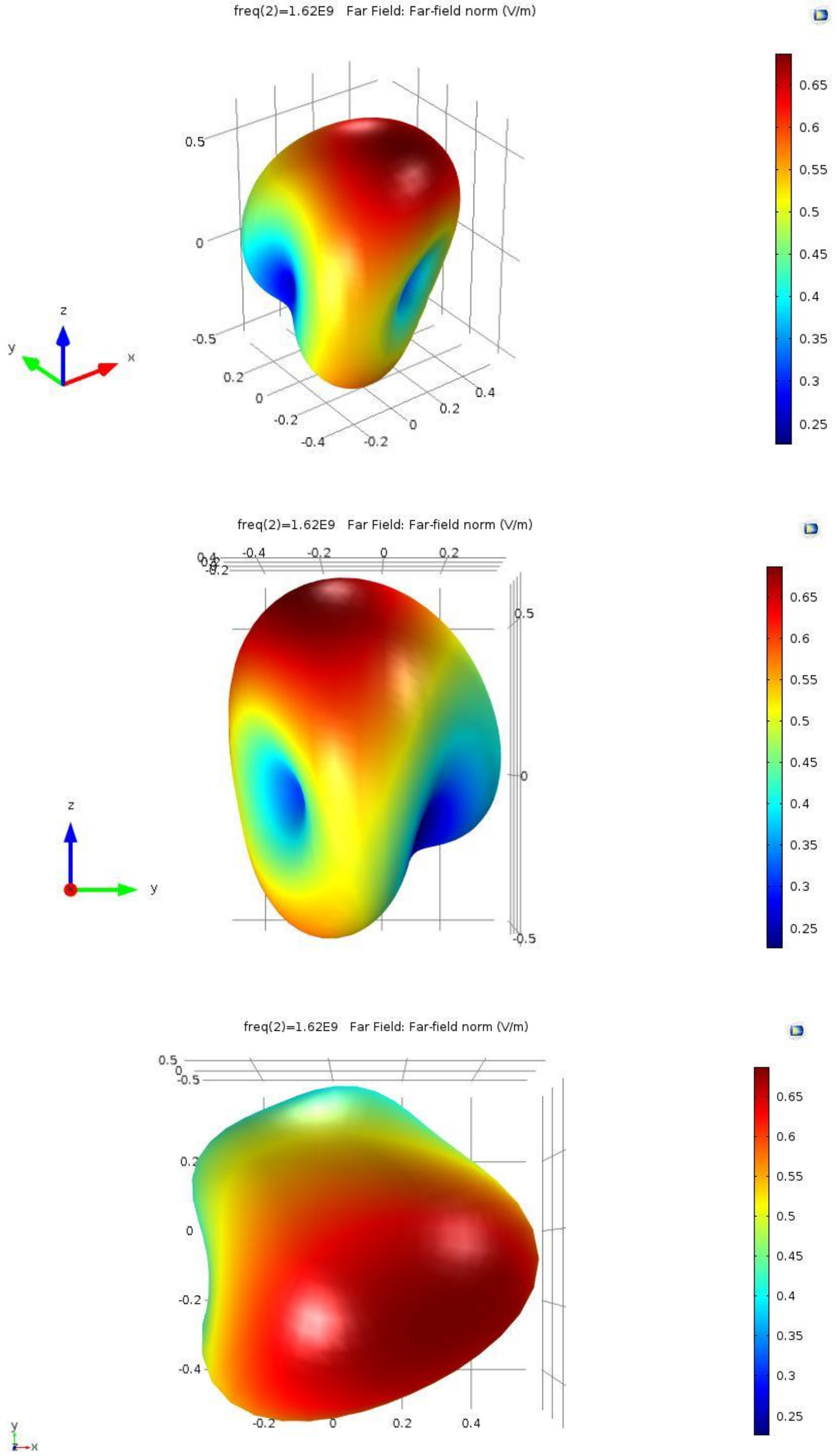

Figure 3.1.8 3D far-field radiation pattern shown from three different angles.

Table 3.1.4 several mobile phones models and its internal type antenna [29].

Model	Antenna	Length	Width	Height	Effective Volume (cc)	Form Factor	Frequency Bands
Apple iPhone 2G	Planar Monopole	55.6	22.8	10.4	6.9	Candybar	GSM850/900/1800/1900
Apple iPhone 3G	Planar Monopole	55	20	8.5	9.0	Candybar	GSM850/900/1800/1900+3G
Apple iPhone 4	Planar Monopole	58	14	6.1	4.8	Candybar	GSM850/900/1800/1900+3G
ASUS M307	Planar Monopole	35	24	0.2	1.6	Clamshell	GSM900/1800/1900
ASUS P525	PIFA	34	20	7.6	5.1	Candybar	GSM850/900/1800/1900
BenQ Siemens EF-71	Planar Monopole	38	10.1	5.3	2.0	Clamshell	GSM900/1800/1900
Blackberry 8100	Planar Monopole	42	9	6.7	7.7	Candybar	GSM850/900/1800/1900
Blu 233/Sendo M570	Planar Monopole	35	16.4	1.5	0.9	Clamshell	GSM900/1800
Geo GC688	PIFA	36.5	10.7	5.8	2.3	Slider	GSM850/900/1800/1900
Hagenuk	Slot	72.4	49.6	6.0	12.6	Candybar	GSM900
Motorola E398	PIFA	37.2	20.1	8.7	5.5	Candybar	GSM900/1800/1900
Motorola KRZR K1	Planar Monopole	37.5	7.6	7.0	2.4	Clamshell	GSM850/900/1800/1900
Motorola L2000/P7389	Helix	44.2	5	5.0	2.1	Candybar	GSM900/1800/1900
Motorola L6	Planar Monopole	36.1	7.5	7.0	1.9	Candybar	GSM900/1800/1900
Motorola T193	PIFA	27	17.8	8.1	2.9	Candybar	GSM1900
Motorola T720i	Helix	29	10	10	1.6	Clamshell	GSM900/1800
Motorola V690	PIFA	38	16.9	9.5	6.1	Clamshell	GSM900/1800/1900
Motorola W208	PIFA	35.6	24	6.5	2.1	Candybar	GSM900/1800
Nokia 2626	PIFA	38	22.3	7.7	3.0	Candybar	GSM900/1800
Nokia 2652	PIFA	41.4	18.6	6.6	5.1	Clamshell	GSM900/1800
Nokia 5210	PIFA	34	18	6.6	4.4	Candybar	GSM900/1800
Nokia 5300	Planar Monopole	36.5	8.4	9.3	2.8	Slider	GSM900/1800/1900
Nokia 5320	PIFA	42.6	28.5	8	6.4	Candybar	GSM850/900/1800/1900+3G
Nokia 5500	PIFA	34.4	38.8	8.5	10.5	Candybar	GSM900/1800/1900
Nokia 6030	PIFA	38	22.3	9.1	6.6	Candybar	GSM900/1800
Nokia 6100	PIFA	37.9	28.6	6.3	6.1	Candybar	GSM900/1800/1900
Nokia 6108	PIFA	37.9	28.6	5.6	5.5	Candybar/ Flip	GSM900/1800/1900
Nokia 6111	Planar Monopole	41.4	9.1	6.8	2.5	Slider	GSM900/1800/1900
Nokia 6210	PIFA	42.6	19.8	9.9	8.4	Candybar	GSM900/1800

3.2 Wideband PIFA Antenna Design

PIFA antenna is used widely in mobile phones today as shown in the table 3.1.3, that's due to many advantages of PIFA which are mentioned in chapter 2 Antennas types. Hence in this section PIFA antenna will be used to design a wideband antenna and simulate the model in COMSOL.

Model descriptions

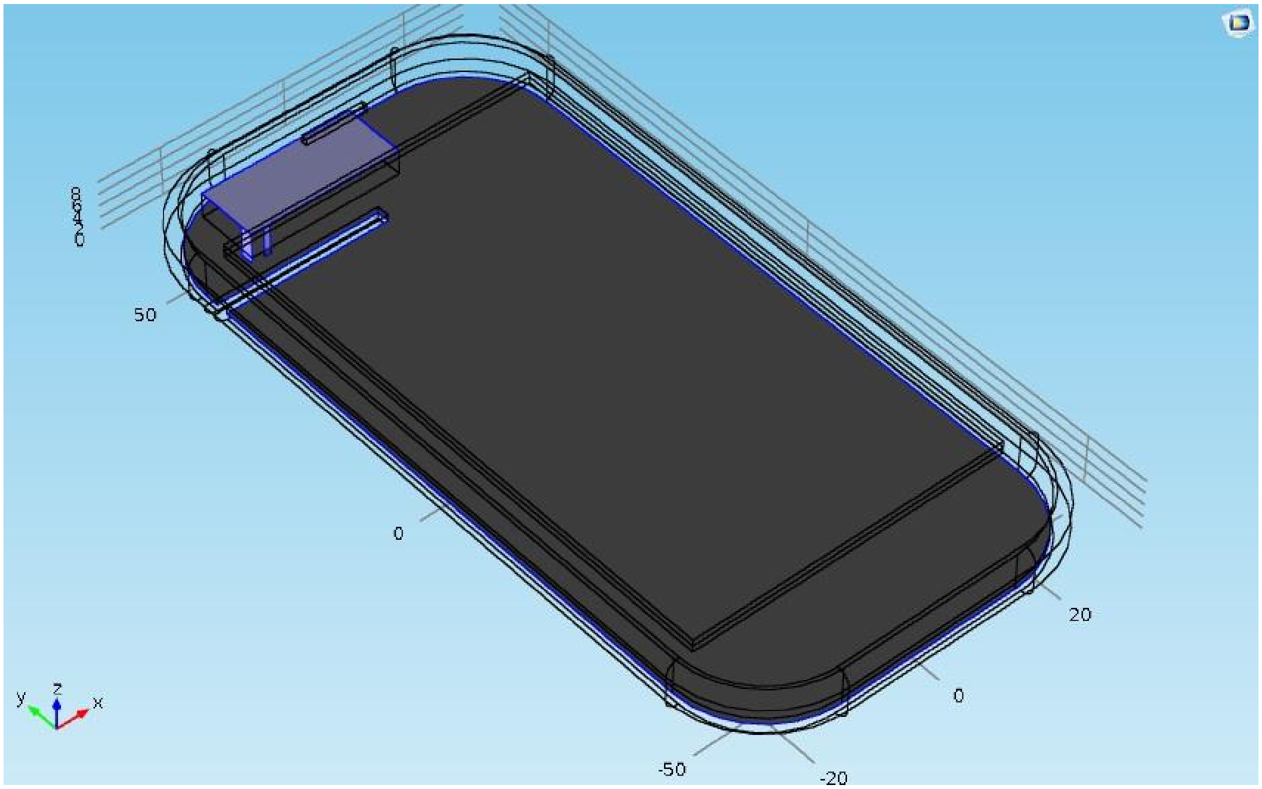

Figure 3.2.1 the wideband PIFA antenna is mounted to left top of the slotted ground plane.

PIFA antenna design using slot technique, with desired range of frequency from 1800MHz to 2600MHz. This important range cover GSM (1800MHz & 1900MHz), UMTS (2100MHz), Bluetooth and Wi-Fi (2.4GHz), and LTE system (2.3GHz, 2.5GHz, and 2.6GHz).

The wideband PIFA antenna design using slot technique model consists of the same materials which are used in previous design table 3.1.2, except the dielectric material between the PIFA and the ground plane which is air with dielectric constant (ϵ_r) equals to 1 to have a better efficiency and a wider bandwidth.

Calculations

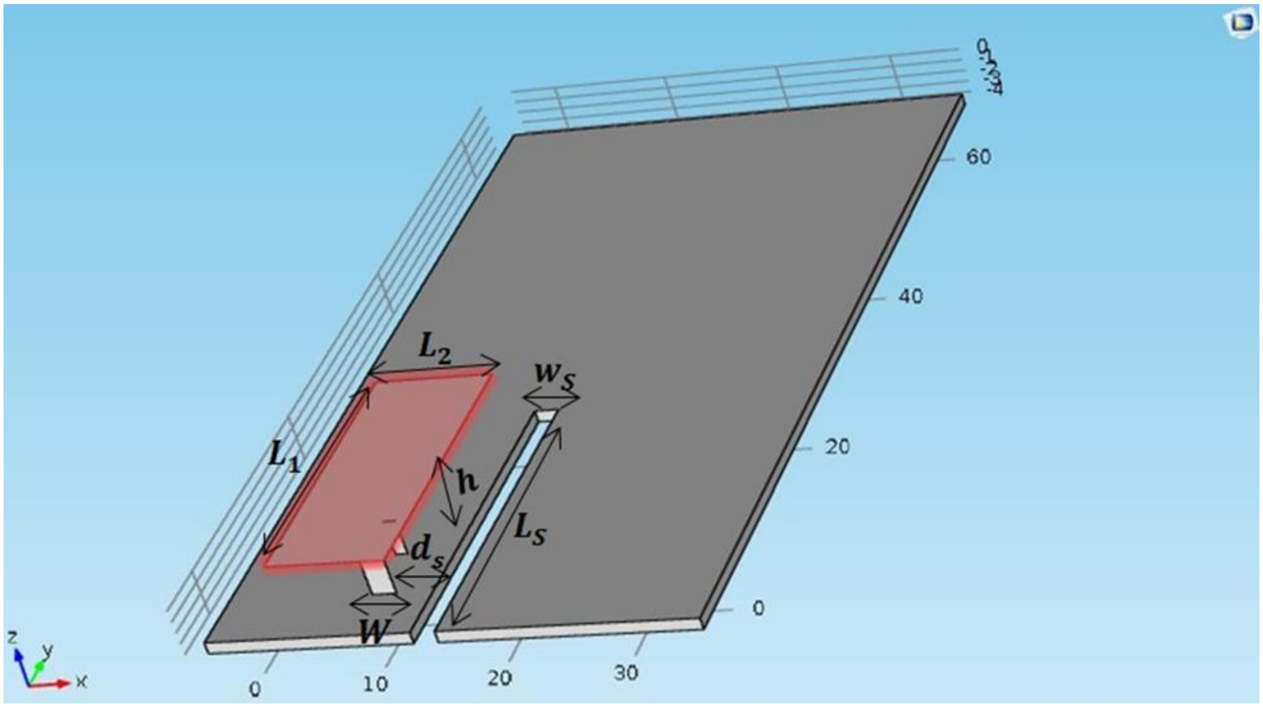

Figure 3.2.2 The illustration dimensions of the PIFA $L_1 = 24mm$, $L_2 = 10mm$, $h = 4mm$, $W = 2mm$ *and the ground plane with its slot dimension* $d_s = 5m, Ls = 28mm$, $W_s = 2mm$.

The resonant frequency of the PIFA antenna can be calculated using 3.1.1 and 3.1.2 formulas respectively.

$24 + 10 - 2 = \frac{\lambda_0}{4}$, $\lambda_0 = 128\text{mm}$

$f_0 = \frac{3\times10^8}{128\times10^{-3}\times\sqrt{1}} = 2.343GHz$

The desired bandwidth percentage:

$f_{center} = \frac{2600+1800}{2} = 2200MHz$

$Bandwidth\% = \frac{2600-1800}{2200} \times 100 = 36.36\%$

If the bandwidth percentage higher than 20%, that's bandwidth is considered as a wide bandwidth.

Simulations

In this design we use a slotted ground plane in order to improve PIFA bandwidth. A slot in a ground plane creates different paths of current flow which leads to kind of diversity in electric field, consequently a wider bandwidth. The length of the slot should be proportional with the desired resonant frequency. The simulations are performed for two cases, one without the slot and other with slot, for different positions of the slot on the ground plane and for different dimensions of PIFA except the width of the planar which is fixed at 10mm and width of the short pin is 2mm.

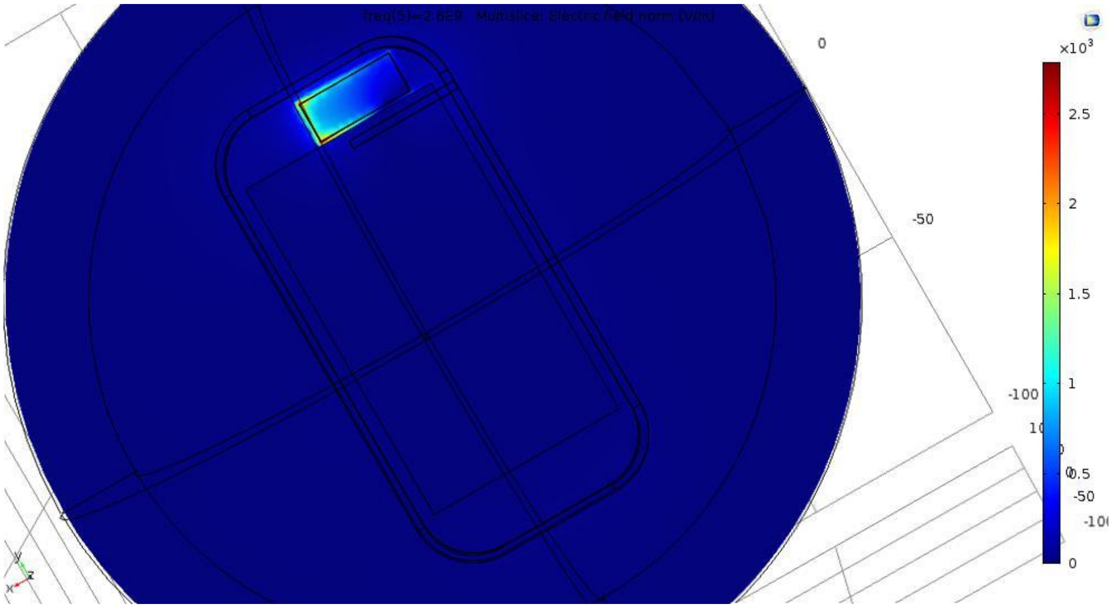

Figure 3.2.3 Back view of E-field norm distribution on the top of the wideband PIFA.

In figure 3.2.3 the plot of the E-field shows that the electric field is strong at one of the top metallic surface shell far from the feeding point on both corners of the planar which similar to E-field plot in previous design.

Figure 3.2.4 shows the far-field gain in dBi for every frequency center of the desired bands and in figure 3.2.5 the 3D far-field radiation pattern of wideband PIFA shown from three different angles.

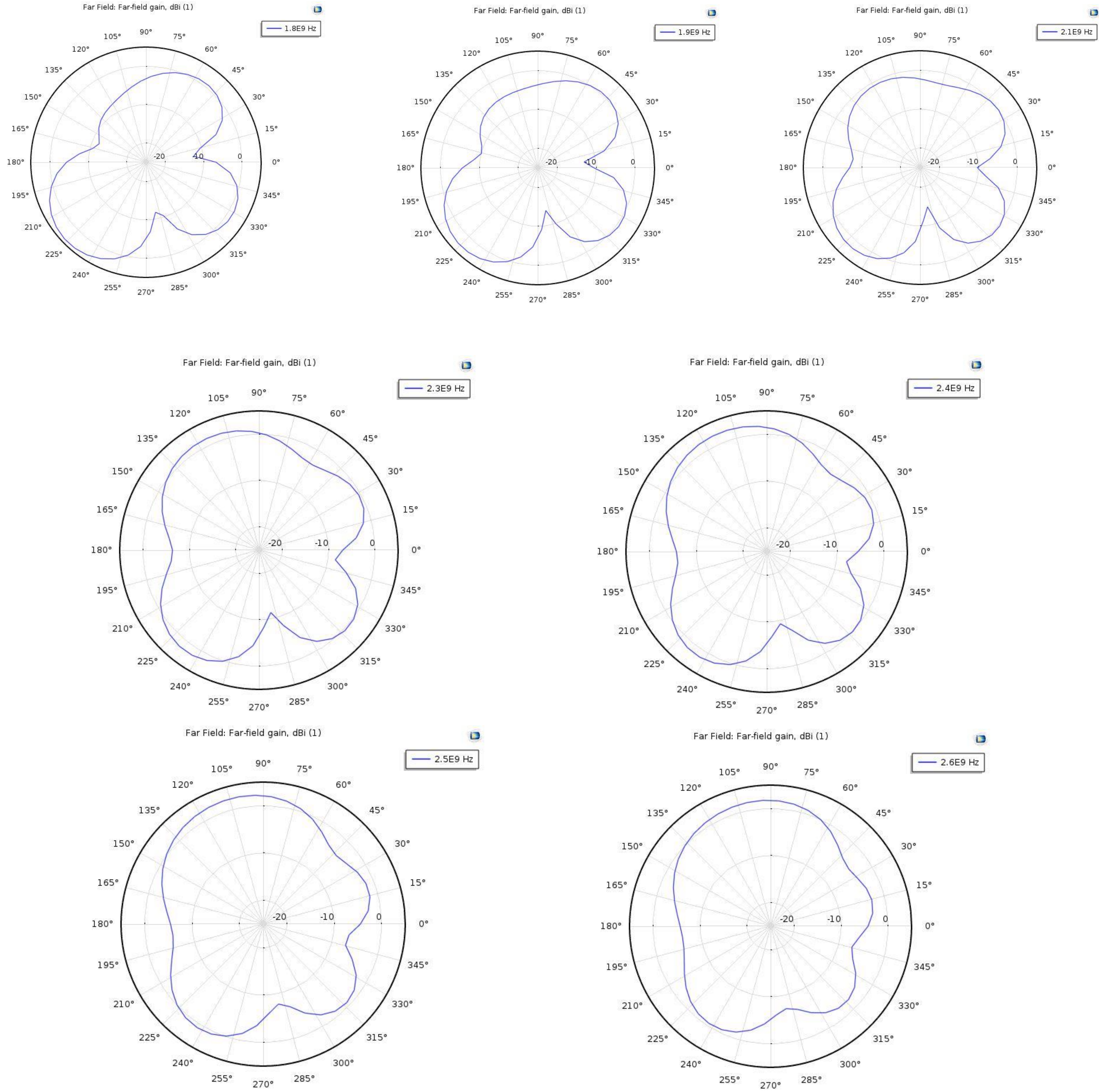

Figure 3.2.4 Antenna gain pattern on the xy-plane for each frequency band.

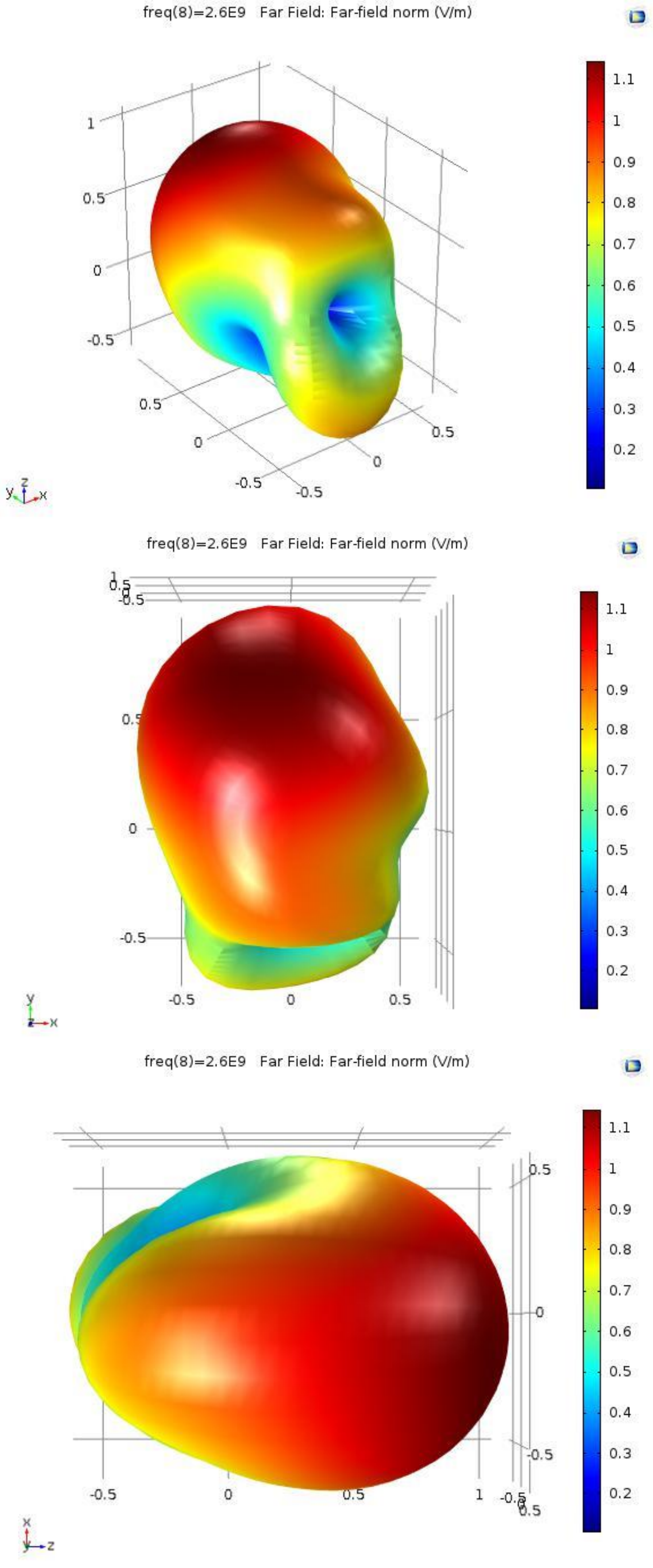

Figure 3.2.5 3D far-field radiation pattern of wideband PIFA shown from three different angles.

Results and Analysis

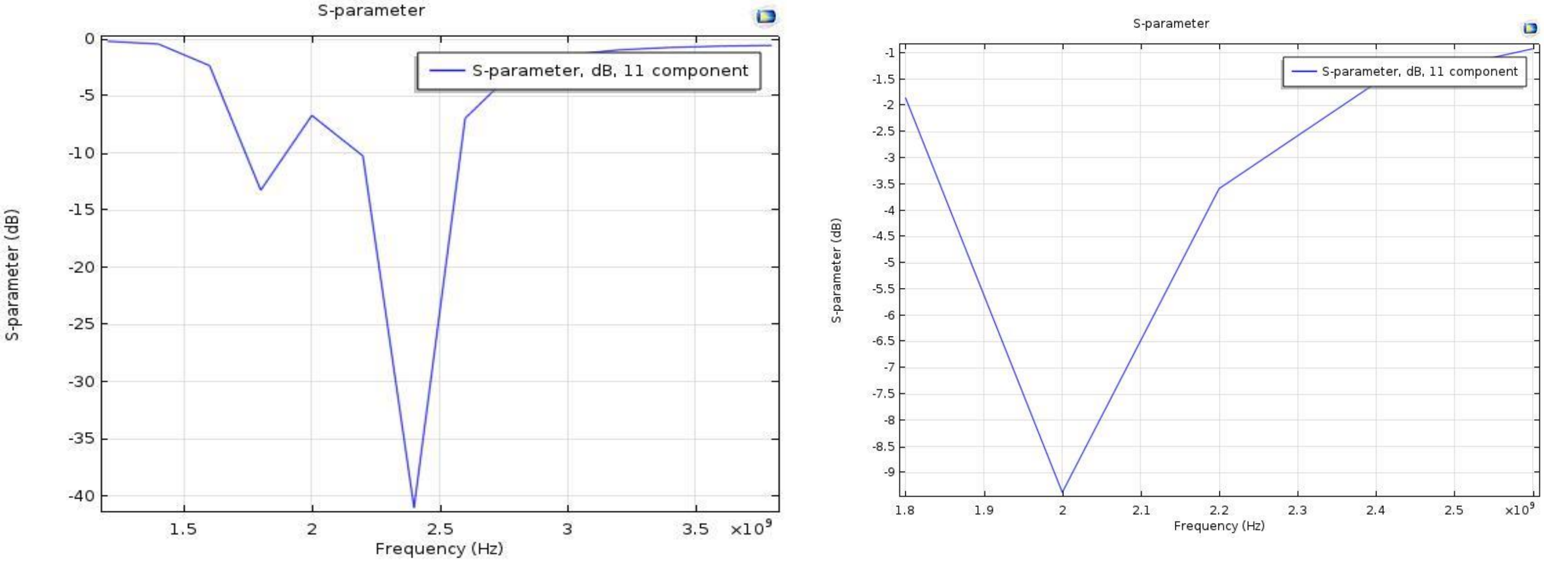

Figure 3.2.5 (a) Return loss (S11) for the wideband PIFA using slotted ground plane by simulation. *(b) Return loss (S11) for PIFA without slot.*

As shown in figure 3.2.5 (a) the resonant frequency is 2.4GHz with -40dB and return loss of the frequencies 1800MHz,1900MHz,2300MHz,2400MHz and 2500MHz is $\leq -10dB$. At 2600MHz is -8dB and at 2100MHz is -9dB which it's still reasonable. With considering 2600MHz has a sufficient low return loss, the bandwidth will be 36.36%.

Figure 3.2.5 (b) shows the PIFA antenna without using a slot on ground plane with resonant frequency at 2GHz, return loss -10dB and the bandwidth can be calculated as following:

$$Bandwidth\% \approx \frac{2.01-1.99}{2} = 1\%$$

Based on what is mentioned above, we can conclude that's the slot has a big significant effect on either bandwidth or on better matching impedance network. In bandwidth there is a huge difference; without using slot the bandwidth is 1% and with slot is 36.36%. Even if we exclude the band 2600MHz the bandwidth 32.36% which is still much wider rather, comparison is nonessential. On other hand the effect of the slot on return loss (S11) is clear as well. Where PIFA with slotted ground plane resonate at 2.4GHz with -40dB, and without slot resonate at 2GHz with poor return loss -10dB.

As seen in figure 3.2.4 it's obvious that all frequencies bands have an Omni-directional behavior with reasonable gain table 3.2.1 summary the maximum and minimum gain.

Table 3.2.1 Simulated gain of the frequencies which shown in figure 3.2.4.

Frequency	Maximum gain	Minimum gain
1800MHz	4.95dB	-12.48dB
1900MHz	3.97dB	-13.86dB
2.1GHz	2.89dB	-14.72dB
2.3GHz	2.54dB	-11.31dB
2.4GHz	2.78dB	-9.33dB
2.5GHz	2.89dB	-7.79dB
2.6GHz	2.25dB	-7.1dB

In this design several methods are followed in order to improve the Bandwidth and the gain, where the relation between the Bandwidth and the gain is not linear. Using dielectric material with high dielectric constant causes a degraded gain; on other hand the effects on Bandwidth could be neglected.

Techniques which are used to increase the Bandwidth for proposed PIFA:

- Bandwidth depends very much on the size of the ground plane. So for better performance, the whole available size of the chassis should be used as a ground plane. Reducing the ground plane can effectively limit the bandwidth of the PIFA antenna.
- Using slotted ground plane: using a slot with proper length to get other resonant frequencies.
- Using Air as a dielectric material between the PIFA element and the ground plane this technique improves the Bandwidth and enhances the gain.

A wideband PIFA antenna has been designed and presented. The proposed PIFA antenna occupies a compact envelope dimension of $24 \times 10 \times 4mm^3$while covering the required wide band with a sufficient impedance matching (S11 ≤ -10 dB) covering GSM (1800MHz &1900MHz), Bluetooth & Wi-Fi (2.4GHz), and LTE system (2.3GHz, 2.5GHz) except UMTS (2.1GHz) and LTE (2.6GHz) which has acceptable return loses.

3.3 Wi-Fi and Bluetooth Antenna

Wi-Fi and Bluetooth antennas work by transmitting and receiving the electromagnetic waves from the antenna to the receiver. They are one of the most useful services in the new smart phones, Wi-Fi antenna is the smallest antenna could be finding in the mobile phones because it works with the highest frequencies.

One of the best advantages of the Wi-Fi antenna that can support dual-band, the first band is from 2400 MHz to 2484MHz and also can support 5150-5850MHz; therefore the Wi-Fi antenna can cover the band of the Bluetooth too.

Nowadays, the most useful Wi-Fi antenna in mobile phone that's is connect to small chip which work for the Wi-Fi and the Bluetooth antennas in the same time, also we can have two antennas one for the Wi-Fi antenna which cover 5GHz and the other one for the Bluetooth which cover 2.4GHz but it is not common to use two types of antennas it will need more space and it will be costly than use one antenna.

As we mentioned Wi-Fi antenna is the smallest antenna in the mobile phones, the antenna half wavelength which covering the Bluetooth 2.4 GHz supposed to be 6.25 cm and at 5GHz which is mostly dipole antenna the half wavelength for sure will be smaller, it is 3 cm. also the quarter wavelength could be used which mean that the size of the antenna will be much smaller which is usually PIFA or PMA.

The quality of the Wi-Fi and Bluetooth antenna connection depend mainly on the gain of the antenna as we have mentioned in the second chapter, by determining the antenna power gain we can identify the efficiency and the directivity of the antenna, the order of the Wi-Fi mobile phone antenna efficiency from -6 dB to -2dB.

In case the primary antenna of the mobile phone does not cover the Bluetooth and Wi-Fi frequencies the Wi-Fi and Bluetooth antenna integrated on the top of the phone close to the GPS antenna due to the hand held of the user.

Chapter 4

Specification Absorption Rate (SAR)

As the world care about the development of wireless communication and with the really huge goals they have achieved on the other hand we should consider the human health and the risk which is effect negatively in the human health.

The meaning of SAR is a short name of specification absorption rate which is the measurement of the energy has absorbed by the human body during transmit the radio frequency electromagnetic field. The human body absorbed the energy that's mean we will lose some energy the second problem that's mean will affect the human body badly.

SAR can be calculated by integrating or averaging over 1 gram or 10 gram:

$$\boldsymbol{SAR = \int \frac{\sigma(r)|E(r)|^2}{\rho(r)} dr} \qquad \textbf{(4.1)}$$

Regarding to the equation 4.1 it explains that SAR is a function of the induced electrical field which is from the radiated energy E it can be measured by $volts\ /\ meter$, the electrical conductivity σ which can measured in $Siemens\ /\ meter$ and the mass density ρ can be measured in $g\ /\ (cubic\ meter)$, finally the unit of the SAR is $W\ /\ kg$.

The SAR have different values regarding to the design of the mobile phone and the location of the antenna in the mobile phone, therefore always, the mobile phone antenna located on the bottom of the phone to keep the radiate as far as much of the user. Low SAR means that it is safer than the high SAR while all the mobile phones has radio frequencies. On the other hand we don't forget to mention that the SAR will effect on the quality of the power by reducing the power level because of the absorption.

4.1 How To Measure The SAR?

Measuring the SAR by the DASY measurement system as shown in figure 4.1.1

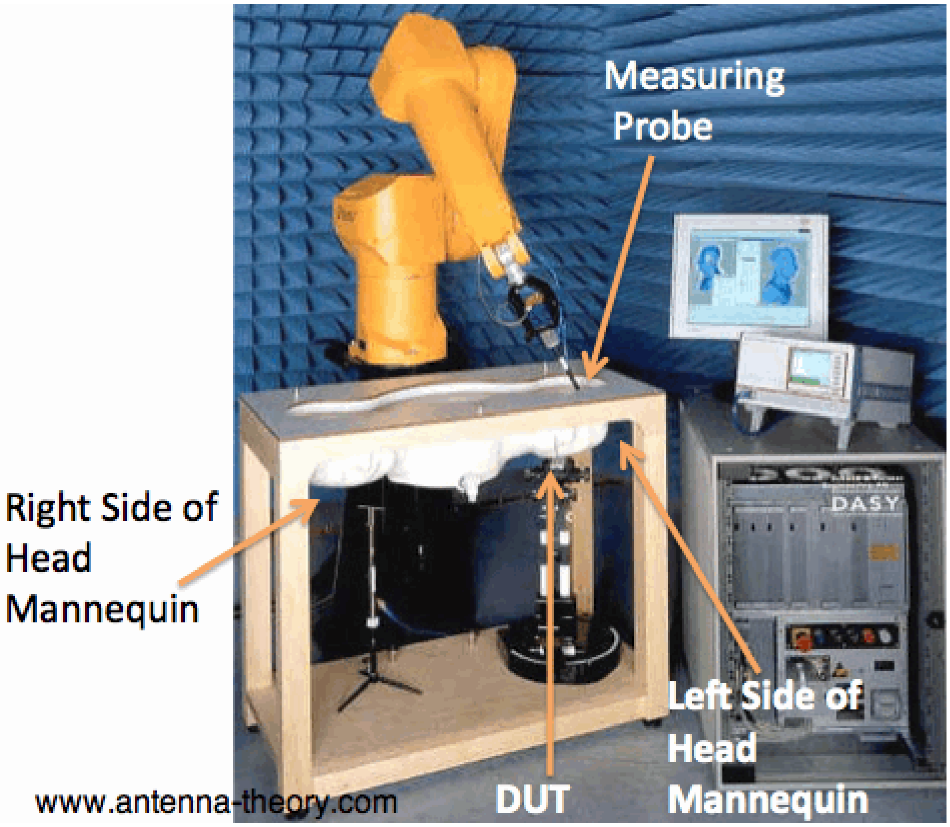
Measuring
Probe
Right Side of
Head
Mannequin
DASY
Left Side of
Head
Mannequin
DUT
www.antenna-theory.com

http://www.antenna-theory.com/definitions/sar.php

Figure 4.1.1 The DASY SAR measurement system.

If we look at the figure 4.1.1 it shows different equipment which gives us availability for picturing the same situation of using the mobile phone in the real life.

The hollow tub in the figure 4.1.1 has the same shape of the right side and the left side of the human head, also there is a yellow robot with long arm which has in the end of the arm a measuring probe, the measuring probe moves around the all the region of the right side of human head and the left side too to lead us the measuring the SAR. But to get the correct measuring of the SAR we should have the correct conductivity and correct density therefore the tub must be filled with special fluid which has almost the same properties of the human tissue.

It is really worth to mention that the fluid has to change the standardized of the human tissue regarding to the frequency we are dealing with.

The DUT which has a long arm, it is located under the hollow tub and close to the head shape due to our need to place the mobile phone in the edge of the head as it must be in the reality.

Finally, the measurement shouldn't be in one side of the human head and we should consider that the measurement of the SAR very sensitive so we can get a very big different result because of the complex nature of the near field.

4.2 The Techniques Of Reducing SAR

Reducing the electrical magnetic field of the antenna will effect to reduce the value of SAR in human head that can be made by using the RF shield to the front side of the phone. The RF shield which has been made of ferromagnetic material so it can help to reducing the value of the SAR because of the suppression of the surface current on the front of the mobile phone box as shown in figure 4.2.1.

During the travelling of the electrical magnetic wave through the free space the wave will be divided to three parts, the first part of the wave going to transmitted, other parts will reflect and absorbed, the function of the electrical magnetic wave material absorbed the energy of the wave as magnetic loss then convert it to energy which it can be feel as heat.

The main materials that are used for making electromagnetic shield are sheet metal, metal foam and material, the holes in the shield should be smaller than the wave length to keep it away.

The value of the reduction depends mainly on the material that have been used, the hickness of the shield, the frequency of the field, the shape, the directivity of the aperture to indicate the electromagnetic field and the size of the shield.

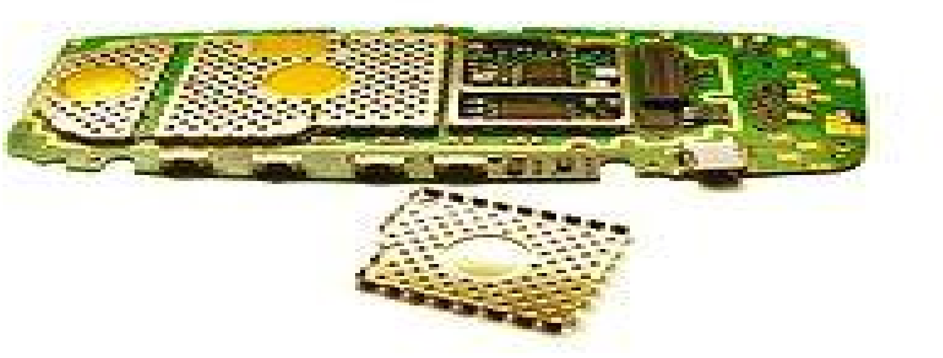

http://en.wikipedia.org/wiki/Electromagnetic_shielding

Figure 4.2.1 The RF shield in mobile phone.

The identification of the RF shield can be determined by determining the value of the SAR reducing factor (SRF) so if the value of SRF is very large that's mean its need to greater RF shield.

SRF can be calculating by the following formula:

$$\boldsymbol{SRF_{Total}(\%) = \frac{P_{abs} - P_s}{P_{abs}} \times 100} \quad (4.2.1)$$

$$\boldsymbol{SRF_{1g}(\%) = \frac{SAR_{1g} - SAR_{1g,s}}{SAR_{1g}} \times 100} \quad (4.2.2)$$

$$\boldsymbol{SRF_{10g}(\%) = \frac{SAR_{10g} - SAR_{10g,s}}{SAR_{10g}} \times 100} \quad (4.2.3)$$

- (SRF) is the total SRF
- (P_{abs}) is the power absorbed in the head model, (P_s) is the power in the RF shield.

- (SRF_{1g}, SRF_{10g}) is the SRF for 1g and 10g peak SAR
- (SAR_{1g}, SAR_{10g}) is 1g and 10g peak SAR (without RF shielding)
- $(SAR_{1g,s}, SAR_{10g,s})$ is 1g and 10g peak SAR (with RF shielding)

Headset vs SAR

Based on the study of UK consumers association reported that the headset can reduce the value of SAR, because of the wire that's connect the headset to mobile phone therefore the wire is working as channeled unshielded radiation and antenna to the human head. The SARtest discovered that can reduce the absorbed radiation until 70%.

4.3 The Limit of SAR in Different Countries

TABLE 4.3.1 The limit of the SAR in different countries [13].

Country/Region	SAR measurement protocol	Reference to SAR limit	Limit
Europe	European Specification ES 59005 (1998)	ICNIRP Guidelines 1998 (ICNIRP 1998)	2.0 W / kg in 10g of tissue
Australia	Australian Communication Authority (ACA) standard (ACA RS 1999)	Australian standard AS/NZS 2772.1	1.6 W / kg in 1g of tissue
US	Federal Communications commission(FCC) Guidelines(FCC1987)	American Standard ANSI C95.1 (ANSI 1992)	1.6 W / kg in 1g of tissue

Regarding to the table 4.3.1 the average limit of European countries is 2.0 $W\ /\ kg$ in 10g of tissue is absorbing the most signal, Australia and US has the same limitation 1.6 $W\ /\ kg$ in 1g of tissue.

In India they were following the same limitation as Europe until 2012 then they have switched to follow the same limitation as US, and the test of India in SAR is done by Telecommunication engineering center. There are many countries have been following the same limitation of US such as Korea, Bolivia and Canada 1.6 $W\ /\ kg$ in 1g of tissue.

4.4 SAR With Different Types Of Antenna

There are some types of antenna that have a really good deal with SAR, the table 4.1.3 shows the dealing between the level of SAR with different types of antenna.

Table 4.4.1 The level of SAR in different types of antenna.

Type of antenna	Level of SAR
Helix	High SAR
Slot	LOW SAR
PIFA	LOW SAR
PMA	High SAR

As showing in figure 4.4.1 first shape show the human head with Helix external antenna which used since 1995 to 1999 with thickness 20 mm and the value of SAR 1.68 mW/g , we notice in the second shape it had used the same type of antenna but the thickness have changed from 20mm to 10mm because of this change of the thickness the value has increased to 2.26 mW/g, the used of PIFA from 1997 to 2009 which has a really good deal with SAR as we mentioned in table 4.4.1 and PMA which used above 2009 with SAR value 2.0 mW/g.

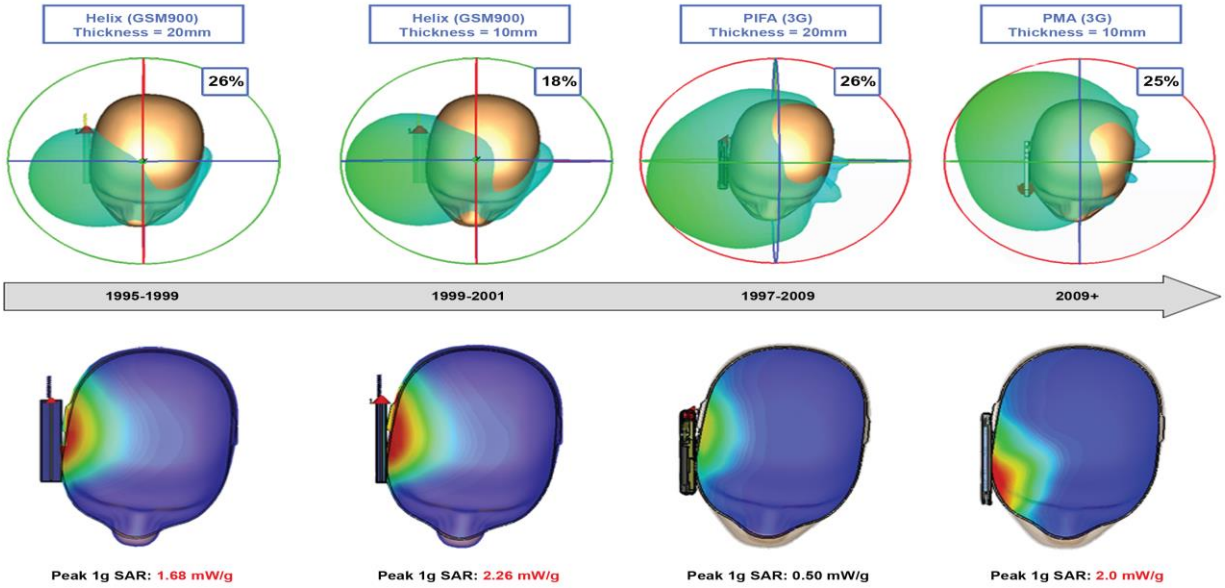
Helix (GSM900)
Thickness = 20mm
Helix (GSM900)
Thickness = 10mm
PIFA (3G)
Thickness = 20mm
PMA (3G)
Thickness = 10mm
26%
18%
26%
25%
1995-1999
1999-2001
1997-2009
2009+
Peak 1g SAR: 1.68 mW/g
Peak 1g SAR: 2.26 mW/g
Peak 1g SAR: 0.50 mW/g
Peak 1g SAR: 2.0 mW/g

http://www.scribd.com/doc/254155865/Mobile-Antenna-Design#scribd

Figure 4.4.1 The evaluation and the performance of GSM antenna types from 1995 to 2011.

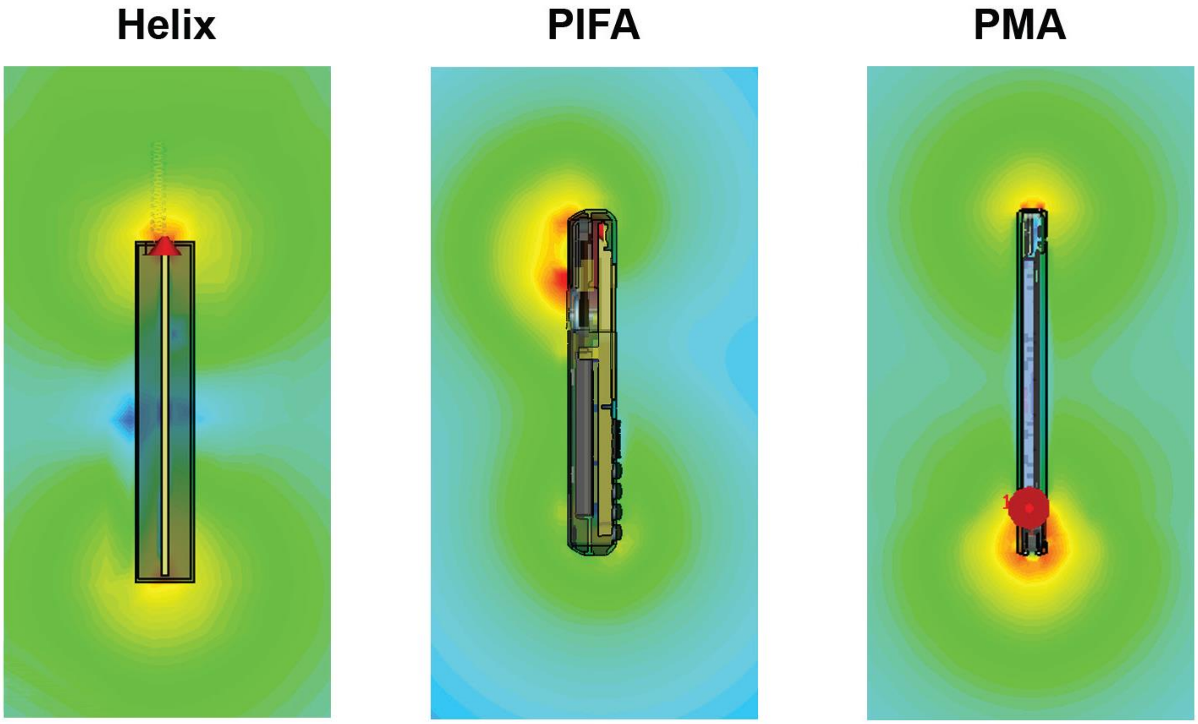
Helix
PIFA
PMA

http://www.scribd.com/doc/254155865/Mobile-Antenna-Design#scribd

Figure 4.4.2 Comparing between the radiation of Helix, PIFA and PMA.

We can notice from the previous figure 4.4.2 that he radiation of Helix and PMA are like a dipole antenna but PIFA has different characteristics of radiation which guide us to notice that it is a type of microstrip antenna. So we can mention that the dipole antenna hasn’t a good relation with SAR.

Table 4.4.1 The highest SAR values in different well known mobile phone manufacture companies [12].

Manufacture company	Model	SAR in W / kg
Apple	IPhone 5	1.25
Nokia	Lumia 630	1.52
Sony Ericsson	Z1010	1.41
Samsung	P400	1.18
Blackberry	Curve 9320	1.56

Table 4.4.2 The lowest SAR values in different well known mobile phone manufacture companies [12].

Manufacture company	Model	SAR in W / kg
Apple	IPhone 4,5c,4s,6	1.18
Nokia	9300	0.07
Sony Ericsson	Tz600	0.16
Samsung	X830	1.18
Blackberry	Curve 8900	1.01

Chapter 5

Problems of Mobile Phone Antennas Design

Mobile phones antenna design is a complicated process. There are a lot critical considerations, the place where the antennas will be installed in mobile phone chassis for example, and regulatory requirements, what are the possible bands of frequencies could be used. In addition, nowadays, mobile phones have multiple antenna for different functions.

5.1 The Radiation Efficiency of the Mobile Phone Antennas

The performance of an antenna depends on the antenna element itself and the ground plane of the mobile phone. That leads to the small size antenna problem, which means that there are limitations for how much the antenna can be minimized, in term of available space in mobile phone or in the expectable good performance. Mobile phone antennas have low efficiencies, and low input resistance and high input reactance, that's create a difficult for matching impedance between the antenna element and the transmission line.

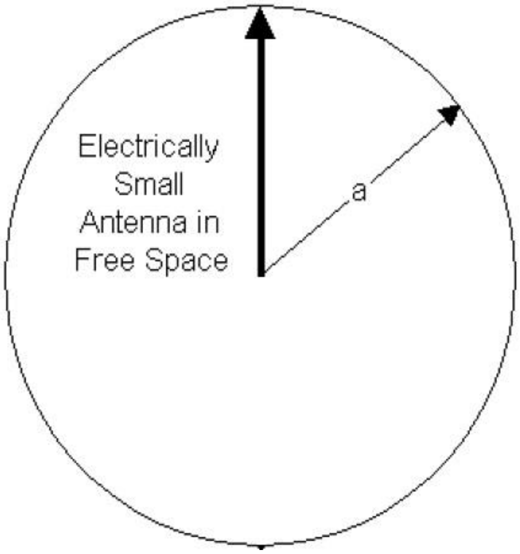

Figure 5.1.1 Imaginary Sphere encloses an antenna in free space with radius a.

At the beginning Wheeler introduced a method to define the maximum volume radiation for electrically small antenna by two measurements, first one in free space and the other within an imaginary closed sphere with radius $a = \frac{\lambda}{2\pi}$ as shown in figure 5.1.1 [18]. Wheeler considers that the antenna resonance frequency can be described as two series resistances; radiation resistance R_{rad} and loss resistance R_L (of used material) [18]. The radiation efficiency of a small antenna can be determined by the following formula:

$$\boldsymbol{\eta_{rad} = \frac{R_{rad}}{R_L + R_{rad}}} \qquad \textbf{(5.1.1)}$$

The current distribution in the small electrically antennas has small space to flow as it’s obvious, that’s make the antenna to behave as a capacitor, and consequently the current will be zero, that means there is no radiation power. In that case we must satisfy a proper matching impedance network in term to get radiation power by the small antenna. Radiation resistance for small antennas becomes very small due to weak radiation [26].

Efficiency of a system- which consists of antenna and matching network- can be expressed with help of the radiation efficiency which described in equation 5.1.2 by:

$$\boldsymbol{\eta_s = \eta_{rad}\eta_m} \qquad \textbf{(5.1.2)}$$

$\boldsymbol{\eta_s}$: Efficiency of the system.

$\boldsymbol{\eta_m}$: Efficiency of the matching network.

Based on the equations 5.1.1 and 5.1.2 we can conclude that in order to get higher efficiency of a system, we have to take in account several parameters:

1. Maximum Radiation resistance.
2. The lowest value of loss resistance.
3. The lower mismatch network as possible (lower |Γ|).

In addition, the mobile phone case's material type, and the surrounding environment (such as hand-held, which it will be discussed later) have an additional effect on the radiation efficiency. Antennas with bigger size have better efficiencies than small antennas. On other hand the space which is available for antenna in chassis of mobile phones is very restricted. Thus there is always tradeoff between size and performance.

5.2 Bandwidth

We can determine the bandwidth of an antenna by determining the impedance for all frequencies in that desired range. But in small antennas case the bandwidth depends on others factors. Small antennas such as PIFA have less real input resistance (it approaches to zero) and highly reactive input impedance, which leads to mismatching network problem. To match the impedance of a small antenna with its feeder's impedance is a quite critical issue. The bandwidth depends on the reflection coefficient and matching networks.

The quality factor Q is a limit of the accessible impedance bandwidth at certain efficiency. Sometimes the small antenna is defined by Q.

As it's mentioned in the previous chapters, we can define the bandwidth by the following formula:

$$\boldsymbol{Bw = \frac{f_{max} - f_{min}}{f_0} = \frac{1}{Q}} \qquad \textit{Valid for } Q >> 1 \qquad \textbf{(5.2.1)}$$

For obtaining wide bandwidth we need to have the most possible minimum value of Q that's a hard challenge to solve. In 1947 Wheeler started to define radiation power factor, and the work on that issue is continued until nowadays. The most famous paper was published by Mclean 1996 [90] which describe the Q factor for a linearly polarized in free space based on Wheeler's concept as following:

$$\boldsymbol{Q_{LP} = \frac{1}{k^3a^3} + \frac{1}{ka}} \qquad \textbf{(5.2.2)}$$

Where, $k = \frac{2\pi}{\lambda}$

For circularly polarized antenna Q has slightly different requirement:

$$\boldsymbol{Q_{CP} = \frac{1}{2}\left[\frac{1}{k^3a^3} + \frac{1}{ka}\right]} \qquad \textbf{(5.2.3)}$$

The small antennas with higher values of Q have narrow bandwidth and that due to the low radiation resistance and high reactance. The Q factor for a small antenna is defined as following:

$$\boldsymbol{Q_{minimum} = \frac{2\omega max(W_e, W_m)}{P_{rad}}} \qquad \textbf{(5.2.4)}$$

ω: The angular frequency.

W_e: The stored electric energy.

W_m: The stored magnetic energy.

P_{rad}: Radiation power.

With assuming that's there is no stored energy (W_e, W_m) inside the sphere (figure 5.1.1) which means reactive fielded equal to zero.

Based on equation 5.2.4 we note that’s, mathematically to get the lowest value of Q the power radiation should be maximized, that’s lead us to the efficiency problem. The mobile phones antenna design’s problems are related with each other. Consequently more challenges are added to design process.

Multiband antenna can be used in order to solve the bandwidth limitations problem partially, since there is no small multiband antennas can cover all desired bands. As obvious from the discussion above, the wide bandwidth is a hard challenge to achieve by an ESA due to low profile (small size). In modern mobile phones, there are several antennas for different functions as mentioned in mobile phones antennas design’s chapter. Each antenna operates with certain band; for example one antenna for higher band frequency, and one for GPS etc. Then all desired bands can be covered by that method. On other hand we have to take care on matching network, since we have several antennas to deal with, mutual coupling antenna to antenna loss, and the available space in mobile phone’s chassis.

5.3 Mutual Coupling Antenna to Antenna Loss

The mutual coupling is the amount of the absorbed energy by an antenna when another nearby antenna is operating [4]. This interaction between near antennas is unwanted because that’s amount of the absorbed energy should be radiated instead. Thus the mutual coupling reduces the efficiency of the antenna on both receiving and transmitting mode.

Mobile phones consist of several antennas, and with mobile phones geometrical dimensions, we find that the antennas near to each other in such way make the mutual coupling loss inevitably exists. So a good isolation is required to keep the antennas in mobile phone’s chassis far from each other as possible. For that the antennas are generally distributed on top and bottom of the chassis which reduce the coupling loss. The coupling loss can be 1-2dB for antenna efficiency. Isolation values for smartphones which have same ground plane at the low band are about 10 dB, and can be 20 dB for the high band [28]. The coupling loss is not constant, it’s varying; the coupling loss at low frequencies and at high frequencies are different.

The isolation can be increased by:

- Minimizing the correlation coefficient between the antennas.
- Increasing the separation space between antennas as much as possible.
- Using different polarizations for the antennas.

- Using proper filters with the antennas to decrease the effects of opposite antennas frequencies.
- Using dielectric walls between the antennas.

The isolation between two antennas is measured by connecting both of them to a Vector Network Analyzer and measuring of S12.

5.4 The Hand-Held Environment Problem

Table 5.4.1 Dipole performance at different locations of simulated distance from human head [24].

Distance from Head	cm	Free space	5	4	3	2	1
Input Impedance	Ω	75+j1.3	64+j27	60+j28	62+j35	52+j17	60+j23
Radiation Efficiency	**%**	100	72.3	63.4	50.5	42.7	29.1
Max. Directive Gain	**dB**	2.15	4	2.15	0	-1	-2
Min. Directive Gain	**dB**	2.15	-7	-7.5	-8	-9	-12

As shown in table 5.4.1, when the antenna is closer to the human head, the input impedance decreases, consequently the resonant frequency of the antenna decreases also. The absorption which occurs by human head reduces the radiation efficiency to 29% when the distance 1cm as shown above.

Other effect of handheld is in the directivity of the antenna. The difference between two positions at free space and 5cm far from the head, shows that the directivity increases from 2.15 dB to 4 dB, that's because the sum of radiation fields and fields scattered by the head which somehow compensate the loss in radiation of efficiency[24].

In general we can conclude that when the mobile phone is closer to the head the gain will fall quickly and the directivity is lost due to polarized fields.

Another model of a dipole antenna with a simulated distance is 0.6 cm from human head, and by applying the method of moments (***MoM***). The input impedance is 65 + j13 Ω, directional properties are same as 1cm and 15% radiation efficiency [25].

In figure 5.4.1 we can note that all three of the patterns radiation show -10dB nulls directly behind the user because absorption of energy by the modeled tissue, the attenuation increases as distance from the body decreases, and is worst for operation near the human torso [24].

Table 5.4.2 The effects of iPhone4 cases on antenna efficiency and absorption in the user's hand [17].

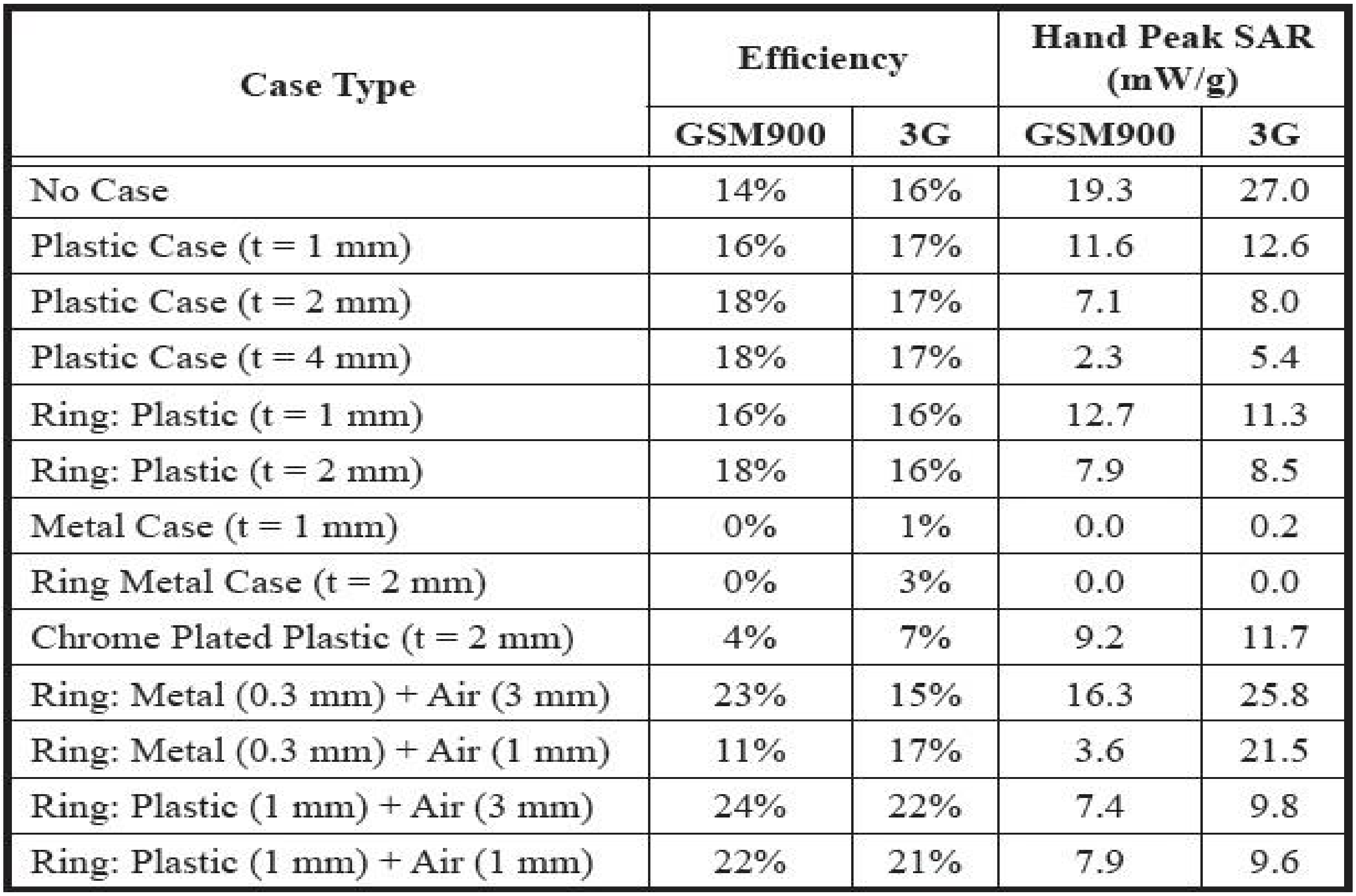

Case Type	Efficiency		Hand Peak SAR (mW/g)	
	GSM900	3G	GSM900	3G
No Case	14%	16%	19.3	27.0
Plastic Case (t = 1 mm)	16%	17%	11.6	12.6
Plastic Case (t = 2 mm)	18%	17%	7.1	8.0
Plastic Case (t = 4 mm)	18%	17%	2.3	5.4
Ring: Plastic (t = 1 mm)	16%	16%	12.7	11.3
Ring: Plastic (t = 2 mm)	18%	16%	7.9	8.5
Metal Case (t = 1 mm)	0%	1%	0.0	0.2
Ring Metal Case (t = 2 mm)	0%	3%	0.0	0.0
Chrome Plated Plastic (t = 2 mm)	4%	7%	9.2	11.7
Ring: Metal (0.3 mm) + Air (3 mm)	23%	15%	16.3	25.8
Ring: Metal (0.3 mm) + Air (1 mm)	11%	17%	3.6	21.5
Ring: Plastic (1 mm) + Air (3 mm)	24%	22%	7.4	9.8
Ring: Plastic (1 mm) + Air (1 mm)	22%	21%	7.9	9.6

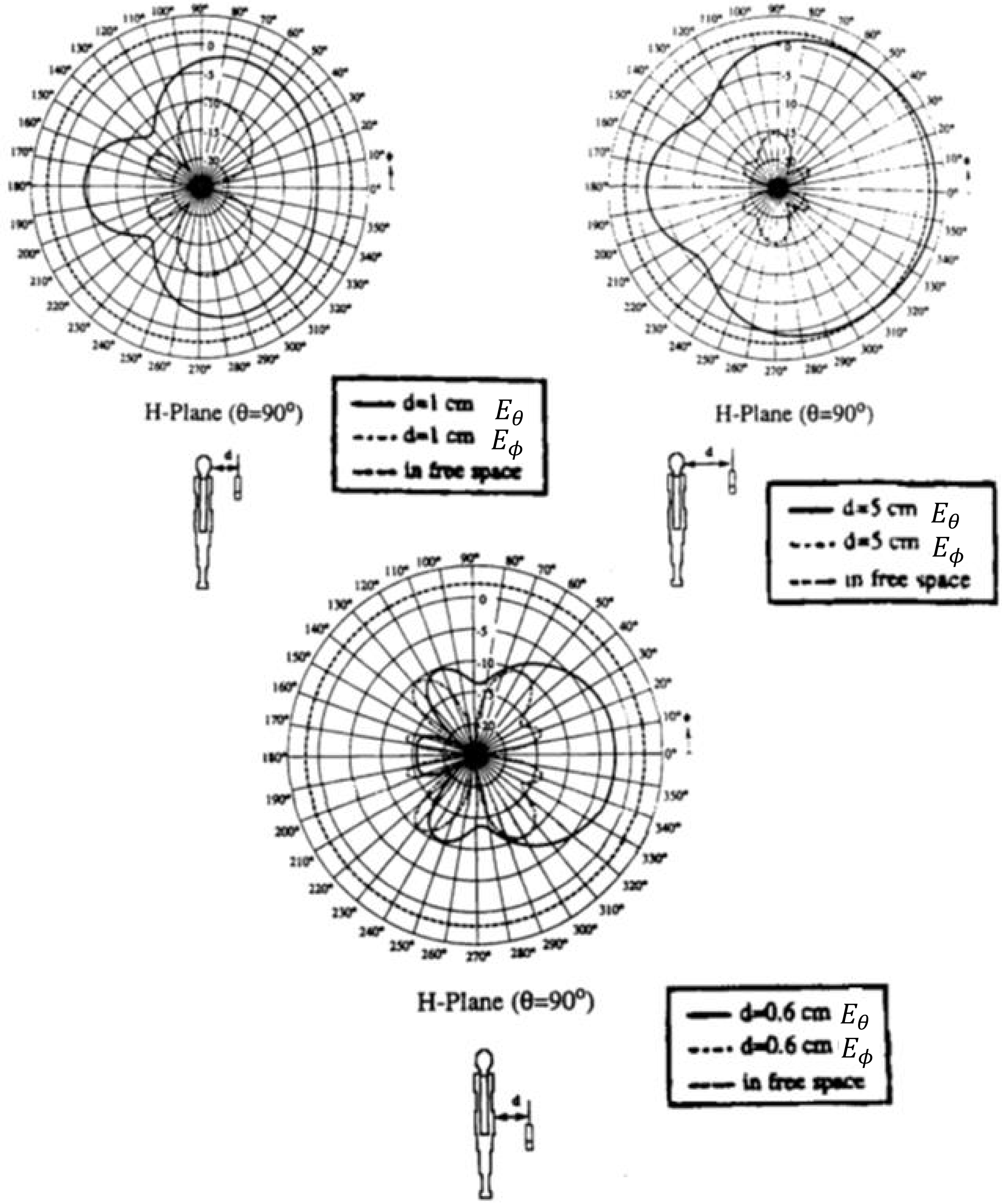

Chuang, H. R. “Human Operator Coupling Effects on Radiation Characteristics of a Portable Communication Dipole Antenna”., IEEE Transactions on Antennas and Propagation, v. 42, n.4, April 1994, pp. 556-560.

Figure 5.4.1 .Free space dipole radiation pattern compared to the calculated (using MoM) radiation patterns for operation of a dipole antenna near the simulated human head and torso.

Coupling of energy to the hand of the user is another problem caused by hand-held environment. Figure 5.4.1 shows the experimental and measured results of the reflected power radio at different distance from the human hand for internal PIFA antenna by using an FDTD model for human hand. As seen below in figure 5.4.2 when the distance d becomes less more and more (the hand of the user gets closer to the antenna) the resonance frequency is shifted, and the reflected power radio becomes more, which means that's the antenna's efficiency is obviously degraded.

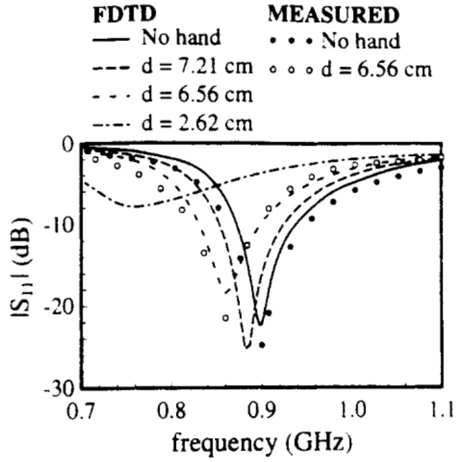

Photo was loaded from: Jensen, M., Rahmat-Samii, Y. "EM Interaction of Handset Antennas and a Human in Personal Communications".

Proceedings of the IEEE, v. 83, n. 1, Jan. 1995, pp. 7-17.

Figure 5.4.2 The experimental and measured results of the reflected power radio at different distance from the human hand.

The hand-held environment problems add more challenges in design process. Since it's so important to take in account the different scenarios. Some of that's issues can be solved by design and installed the antenna in the lower part of the mobile phone to keep the antenna far as possible from the hand of user. Another critical factor can play big role in this term is a good isolation for antenna.

IPhone 4 suffers from isolation antenna problem. Many users reported that signal strength of the phone was reduced when touching the lower left of the phone, which causes dropped calls in some areas [26]. On other hand Apple recommended the consumers to do not grip the phone at that corner.

On July 16, 2010, Steve Jobs advertised that Apple would offer a case for the consumers to help solve the antenna issue [26]. Table 5.4.2 shows an iPhone4 and the effects of different material of casing on efficiency and SAR. Metal cases with thickness equal to 1mm, and 2mm, keep the efficiency without any effects and SAR values at almost zero.

Summary and Conclusion

Mobile phone antennas have several critical parameters such as the geometric dimensions, dialectical materials which are used in design, and the ground plane of the antenna. All of that should be taken into account for optimizations process. The fast development of mobile phones devices add more demand on optimizations of the antennas.

In this thesis we investigate PIFA antenna type by design two models of it. The first model handles one single band with resonant frequency 1.575GHz which could be suitable at that frequency for GPS signal with reasonable gain from -6dBi to 1.5dBi and with return loss (S11) -13dB. The bandwidth 1% which is sufficient since there is no bandwidth practically.

The second proposed PIFA is a wideband antenna with dimensions $24 \times 10 \times 4mm^3$cover range from 1800MHz to 2600MHz. This important range covers GSM (1800MHz & 1900MHz), UMTS (2100MHz), Bluetooth & Wi-Fi (2.4GHz), and LTE system (2.3GHz, 2.5GHz, and 2.6GHz). The Bandwidth is 36.36% which is considered as a wide bandwidth. This wideband PIFA antenna is designed using slotted ground plane and air as a dielectric material between the planar element and the ground plane.

Mobile phone antennas design is a big challenge. Mobile phone antennas designs have lot problems. Low efficiencies can be the most critical problem which leads to tradeoff between size and performance.

Reference

[1] Mobile Antenna Systems Handbook page 17-19

[2] http://www.businessinsider.com/complete-visual-history-of-cell-phones-2011-5?op=1&IR=T

[3] C. Chiau, "Study of diversity antenna array for MIMO wireless communication systems," Ph.D. dissertation, Queen Mary University of London, UK, April 2006.

[4] http://www.antenna-theory.com/

[5] http://classes.yale.edu/fractals/panorama/ManuFractals/FractalAntennas/FractalAntennas.html

[6] Mandelbrot, B. B., the Fractal Geometry of Nature, W.H. Freeman and Company, New York, 1983.

[7] http://www.iject.org/vol4/spl3/c0113.pdf

[8] http://www.ijstr.org/final-print/may2014/Response-Of-Planar-Inverted-F-Antenna-Over-Different-Dielectric-Substrates.pdf

[9] http://www.slideshare.net/NaveenKumar11/thesis-viva-presentation?related=1

[10] http://www.raymaps.com/index.php/planar-inverted-f-antenna-pifa/

[11] http://www.ece.msstate.edu/~donohoe/ece4990notes2.pdf

[12] http://www.s21.com/sar.htm

[13] http://sarvalues.com/what-is-sar-and-what-is-all-the-fuss-about/

[14] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.206.2791&rep=rep1&type=pdf

[15] http://www.ijrte.academypublisher.com/vol02/no05/ijrte02055862.pdf

[16] http://telecom.hellodirect.com/docs/Tutorials/HeadsetBenefits.1.110200.asp

[17] Mobile-phone antenna design, Author: ROWELL, CR; Lam, EYM, Issued Date 2012 URL: http://hdl.handle.net/10722/185908 Rights: IEEEE Antennas and Propagation Magazine. Copyright © IEEE.

[18] H. A. Wheeler, "The radiansphere around a small antenna", Proceedings of the IRE, pp. 1325-1331, August1959.

[19] T. Taga and K. Tsunekawa, "Performance analysis of a built-in planar inverted F antenna for 800 MHz band portable radio units," IEEE J.Select. Areas Commun, vol. SAC-5, pp. 921–929, June 1987.

[20] K. Sato, K. Matsumoto, K. Fujimoto, and K. Hirasawa, "Characteristics of a planar inverted-F antenna on a rectangular conducting body," Electron. Commun. Japan, pt. 1, vol. 72, pp. 43–51, 1989.

[21] T. Taga, "Analysis of planar inverted-F antennas and antenna design for portable radio equipment," in Analysis, Design, and Measurement of Small and Low-Profile Antennas, K. Hirasawa and M. Haneishi,Eds. Norwood, MA: Artech House, 1992, pp. 161–180.

[22] P. Vainikainen, J. Ollikainen, O. Kivekäs, and I. Kelander, “Resonator- based analysis of the combination of mobile handset antenna and chassis,” IEEE Trans. Antennas Propagat., vol. 50, pp. 1433–1444, Oct. 2002.

[23] http://scholar.lib.vt.edu/theses/available/etd-7697-21043/unrestricted/CH1_2.PDF

[24]Chuang, H. R. “Human Operator Coupling Effects on Radiation Characteristics of a Portable

Communication Dipole Antenna”., IEEE Transactions on Antennas and Propagation, v. 42, n. 4,

April 1994, pp. 556-560.

[25] http://en.wikipedia.org/wiki/IPhone_4

[26] The Annual Workshop and Feder Award Ceremony 2010. Speaker: Prof. Raphael Kastner, Tel Aviv University.

[27] Research Article: Novel Wideband MIMO Antennas That Can Cover the Whole LTE Spectrum in Handsets and Portable Computers, Mohamed Sanad1 and Noha Hassan2

[28] Balanis, Constantine A. "Antenna Theory - Analysis and Design", 2005, 3rd Edition, John Wiley & Sons ///page 80.

[29] IEEE Antennas and Propagation Magazine, Vol. 54, No. 4, August 2012.

[30] http://ecee.colorado.edu/~bart/ecen6355/app-04.pdf

High-Performance Indoor VHF-UHF Antennas:

Technology Update Report

15 May 2010

(Revised 16 August, 2010)

M. W. Cross, P.E. (Principal Investigator)

Emanuel Merulla, M.S.E.E.

Richard Formato, Ph.D.

Prepared for:

National Association of Broadcasters

Science and Technology Department

1771 N Street NW

Washington, DC 20036

Mr. Kelly Williams, Senior Director

Prepared by:

MegaWave Corporation

100 Jackson Road

Devens, MA 01434

Contents:

Section	Title	Page
1.	Introduction and Summary of Findings	3
2.	Specific Design Methods and Technologies Investigated	7
2.1	Advanced Computational Methods	7
2.2	Fragmented Antennas	22
2.3	Non-Foster Impedance Matching	26
2.4	Active RF Noise Cancelling	35
2.5	Automatic Antenna Matching Systems	37
2.6	Physically Reconfigurable Antenna Elements	58
2.7	Use of Metamaterials in Antenna Systems	75
2.8	Electronic Band-Gap and High Impedance Surfaces	98
2.9	Fractal and Self-Similar Antennas	104
2.10	Retrodirective Arrays	112
3.	Conclusions and Design Recommendations	128

1.0 Introduction and Summary of Findings

In 1995 MegaWave Corporation, under an NAB sponsored project, developed a broadband VHF/UHF set-top antenna using the continuously resistively loaded printed thin-film bow-tie shown in Figure 1-1. It featured a low VSWR (< 3:1) and a constant dipole-like azimuthal pattern across both the VHF and UHF television bands.

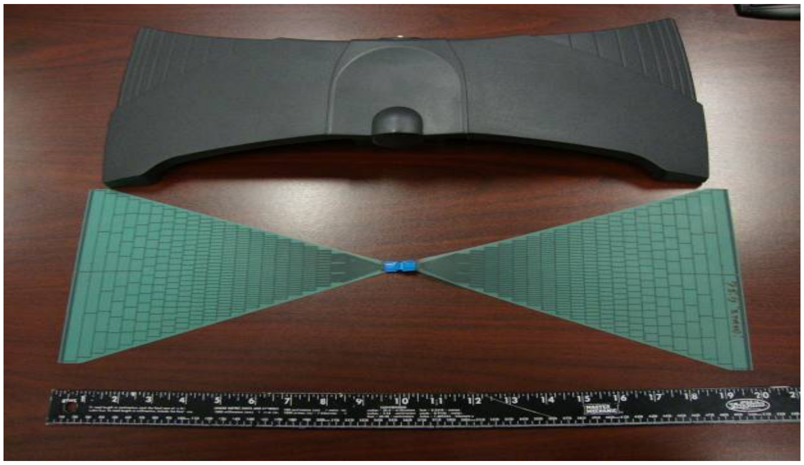

Figure 1-1: MegaWave 54-806 MHz Set Top TV Antenna, 1995

In the 15 years since then much technical progress has been made in the area of broadband and low-profile antenna design methods and actual designs. These improvements have been published in: technical textbooks, peer-reviewed articles, patents, government research and development reports, and seminar proceedings. As a developer of advanced antenna systems, primarily for the U.S. government, MegaWave constantly reviews these sources and acquires the latest computer based EM simulation tools in order preserve its competitive advantage. In this project, this knowledge was used to identify ten candidate design methods and technologies that have the potential to materially improve the performance of indoor VHF-UHF TV antennas. This report describes each candidate and its potential to improve indoor ”set-top” reception of DTV signals between 54 and 698 MHz.

Of course, it must be kept in mind that, while advanced design methods and actual physical designs exist, so do the laws of electromagnetics. Maxwell’s equations have resulted both in practical as well as, what Dr. R. C. Hansen humorously calls, “Pathological Antennas”. These pathological designs are described in his most recent textbook [1], especially in the area of electrically-small and broadband designs. It is instructive to apply these fundamental limitations to the problem at hand, the set-top TV antenna.

[1] Hansen, R.C., “Electrically Small, Superdirective, and Superconducting Antennas,” Wiley, 2006

Consider that a half-wavelength in the low VHF TV band varies between 9.2 and 5.6 feet; between 34 and 27 inches in the high VHF band and between 12.6 and 8.5 inches in the UHF (470-698 MHz) band. A dipole antenna whose physical length is less than its wavelength divided by pi (λ/π) is considered to be an electrically "small" antenna (ESA). ESAs unfortunately are characterized by narrow bandwidths and low gains. Assuming 2 to 3 feet as a maximum acceptable length for an indoor or set-top antenna, it definitely falls into the ESA category in the low VHF band. But, in addition to size constraints and the resulting difficulty in obtaining acceptable performance from a single antenna over the 54 to 698 MHz spectrum, there are other concerns. Indoor and set-top antennas are fundamentally disadvantaged due to building penetration losses and by proximity to sources of manmade radio noise. The former effect is more pronounced at UHF and the latter at low VHF channels. Both can have a significant deleterious effect on antenna performance. This brief discussion highlights the difficult problems inherent in designing efficient, high performance antennas for the indoor/set-op TV environment. Fortunately, emerging technologies may effectively address these concerns.

This report is organized as follows. Sections 2.1 through 2.10 contain summaries of each advanced method and hardware technology identified as a potential candidate for high-performance indoor VHF-UHF DTV antennas. Each section includes a list of references and, in many cases, photographs and performance data for multiple implementations of the technology that is described. Section 3 includes conclusions and a conceptual design for a practical indoor/set top VHF-UHF antenna system.

The authors evaluated each technology and arrived at the conclusions and design concept after sorting the nine hardware candidates into three categories as follows:

- Mature technologies that **do not** require any CE-909-A channel designator or signal quality information from the DTV receiver:
  - Fragmented Antennas (Section 2.2)
  - Non-Foster Impedance Matching (Section 2.3)
- Mature technologies that **do** require channel and quality data from the receiver:
  - Active RF Noise Cancelling (Section 2.4)
  - Automatic Antenna Matching Systems (Section 2.5)
  - Physically Reconfigurable Antenna Elements (Section 2.6)
- Emerging technologies that show promise, but are not sufficiently mature or practical at this time:
  - Metamaterials (Section 2.7)
  - Electromagnetic Band Gap (EBG) Materials (Section 2.8)
  - Fractal/Self Similar Antennas (Section 2.9)
  - Retrodirective Arrays (Section 2.10)

A common thread connects each of these technology areas: ***advanced computational methods***. Whether a particular technology is mature and immediately applicable or emerging and highly speculative, various schemes for antenna design optimization are universally applicable and described in Section 2.1. These methodologies apply to all of the candidate technologies discussed in Sections 2.2 through 2.10, and accordingly was placed at the beginning of Section 2. If even one of the optimization algorithms described had been available during the development of MegaWave's 1995 broadband set top antenna, it is likely that markedly better gain performance would have resulted, especially in the low and high VHF bands. Another attractive and potentially very significant capability offered by optimization algorithms is the possibility of discovering entirely new antenna geometries, rather than simply optimizing a pre-existing geometry.

Table 1-1 ***subjectively*** ranks the nine identified candidate hardware technologies (2.2 - 2.10). A score of 10 represents perfection. By maturity we mean how close to off-the-shelf a particular technology's hardware is and how well it basic principle of operation has been vetted in the literature. The term SWAP refers to size/weight and power.

	Method/ Technology	**Active/ Passive**	**Requires CE-909-A Interface**	**Maturity**	**Vetted**	**Risk**	**Design Complexity**	**SWAP**	**Comments**
2.1	Adv. Comp. Methods	N/A	N/A	9	9	Very Low	N/A	N/A	Applies to all Technologies
2.2	Fragmented	Passive	No	7	7	Low	Low	9	Planar
2.3	Non-Foster	Active	No	6	7	Low	Moderate	7	Limited Bandwidth
2.4	Active Noise Cancelling	Active	Yes	7	6	Low	High	3	Requires I&Q
2.5	Automated Antenna Matching	Active	Yes	7	7	Moderate	High	7	Requires complex TV interface
2.6	Reconfigurable Antennas	Active	Yes	6	7	Moderate	High	6	Control of MEMS w/DC
2.7	Metamaterials	Passive	No	3	4	High	High	8	Emerging/ Availability an issue
2.8	EBG	Passive	No	5	6	High	High	5	Inherently Narrow Band Maybe useful for shielding
2.9	Fractal/Self Similar	Passive	No	6	4	Moderate	Moderate	8	Controversial Performance Gain
2.10	Retrodirective	Active	Yes	4	5	Very High	Very High	2	Narrow - Band, Large

Table 1-1: Candidate Technologies Considered and Their Ranking

As an example of how advanced computational methods could be combined with an advanced hardware technique, that does not require a CE-909-A interface, is described at the end of Section 3 and summarized here.

Using the genetic algorithm described in Section 2.1.5 a fragmented antenna was designed and combined with a non-Foster-matching circuit to provide a planar 54-698 MHz dipole approximately 13 by 13 inches with significantly better gain, especially in the 54-88 and 174-216 MHz bands, than the 1995 MegaWave/NAB set top antenna. Figure 1 shows the broadband fragmented planar element's design obtained after approximately 24 hours of computational time on a PC. Details of the specific method used are in Section 2.2 of this report. It is well matched across the UHF DTV band, but requires some passive matching in the high VHF band (which would also serve as the band combiner) and the more robust matching capability of the active Non-Foster-Matching technique, described in Section 2.3, for the low-VHF band.

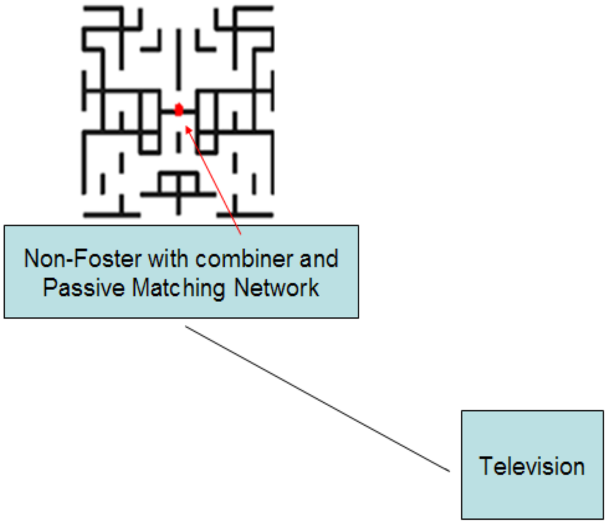

Figure 1. 13 x 13 Inch Planar Fragmented Non-Foster Matched VHF-UHF Antenna

An omni-directional version could also be designed. It should be stressed that the above is included here only to illustrate the notion of combining advanced computational broadband antenna element designs with emerging electronic antenna matching capabilities and that other antenna element geometries are also possible, depending on the starting conditions, trade space dimensions and performance goals provided to the optimizer.

The authors want to make clear that 90 percent of the techniques and ideas contained in this study are the work of others, as published in the open literature and referenced herein.

2.0 Specific Design Methods and Hardware Technologies Investigated

2.1 Advanced Computational Methods

2.1.1 Summary

Optimization methodologies abound, and they are extensively used in every aspect of engineering design, in particular antenna design. Optimization algorithms are useful in two ways. They can be used to optimize the design parameters for a user-specified antenna geometry (for example, element spacing, length and diameter in a Yagi-Uda array). They also can *generate designs* that are impossible to achieve otherwise. In both cases, optimization involves meeting specific performance objectives (typically, VSWR, gain, bandwidth, and so on).

Optimization algorithms have become progressively more important as the limitations of classic analytical techniques have become progressively more apparent. While the equations underlying electromagnetic theory are well understood and accurately describe all electromagnetic phenomena, in most practical cases they cannot be solved analytically or, oftentimes, even numerically. Designing better antennas requires improved methodologies, and state-of-the-art optimization algorithms have proven very effective. There is no question that these techniques are applicable to the set-top antenna design problem, and that they should receive considerable attention in future design activities.

There are many different optimization methodologies that fall into two broad categories: analytical methods and heuristic methods. Analytical methods are based on precise mathematical formulations of the optimization problem. Even though they may be fundamentally numerical in nature, they involve standard mathematical operations such as computing derivatives or evaluating integrals. Heuristic methods may involve equations, but the equations are not the result of an analysis. Instead, they are offered without "proof" based on the fact that they "work."

Many optimization heuristics are Nature inspired. The steps an algorithm performs to optimize an *antenna* are based, for example, on how *bacteria* forage for food. As disparate as these entities may seem, there is a connection, at least in the sense that bacteria finding a good food source is similar to finding an antenna with a good gain-bandwidth product. Optimization algorithms of this type are usually referred to as "metaheuristics," a term intended to emphasize that the method is both *empirical* and *conceptual* in nature. Thus, an effective bacteria foraging algorithm can be implemented in many different ways because the bacteria foraging metaheuristic simply suggests an analogy to Nature that is implemented in a computer algorithm working on an antenna problem. The metaheuristic thus is an *algorithmic framework* instead of a list of steps or instructions.

Several Nature inspired metaheuristics are described. A brief summary of each algorithm is provided, and several example antenna problems solved by a variety of algorithms are discussed. The algorithms include Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Simulated Annealing (SA), Central Force Optimization (CFO), Invasive Weed Optimization (IWO), Intelligent Water Drop (IWD) algorithm, and Bacteria Foraging Optimization (BFO). There are many other optimization algorithms [for example, Space Gravitation Optimization (SGO), Integrated Radiation Optimization (IRO)], but they have not been applied to antennas or antenna related problems.

Each of these algorithms, except one, is inherently *stochastic* because its Nature inspired algorithmic model relies on randomness in its functioning. The underlying equations contain true *random variables* whose values are computed from a probability distribution and consequently cannot be known in advance. As a result, every time a stochastic optimizer run is made, its results are different than the previous run even when exactly the same run setup parameters are used. The performance of stochastic optimizers is necessarily characterized statistically (for example, average values, standard deviations). This may be a limitation in the utility of optimization algorithms if they are used in a set-top antenna on a real time basis. For example, a self-structuring antenna (SSA) must reconfigure itself in real time in response, for example, to a changing environment.

The one algorithm that is not inherently stochastic is Central Force Optimization (CFO) whose Nature inspiring metaphor is gravitational kinematics, the branch of physics that deals with the motion of masses moving under the influence of gravity. The underlying equations are Newton's equations of motion, which are completely *deterministic*. CFO analogizes these equations in "CFO space" by flying "probes" that are similar to small satellites to search a decision space "landscape" for the maximum (optimal) values of a function (for example, antenna gain as a function of element length and polar angle). CFO has been applied to antenna design and network synthesis, and tested against many recognized benchmark functions used to evaluate optimization algorithms. It therefore may be especially useful for the set-top antenna problem.

2.1.2 Introduction

This section describes developments in antenna design optimization over the past fifteen years or so that have been driven largely by the availability of progressively more powerful computers. A plethora of new optimization algorithms have been introduced and tested and are now in widespread use. The new antenna designs often are non-intuitive, occasionally even counter-intuitive, but all share the common feature of not being accessible in any other way. State-of-the-art optimization algorithms can effectively solve intractable problems that have no analytical solutions or are too complex to apply traditional analytical techniques. These approaches are useful right now in designing set-top television antennas, and they will continue to be useful whatever form future set-stop systems take. Some of the more important and interesting optimization algorithms are described here.

Optimization Methodologies. The problem of locating the maximum values of a function is generally referred to as "multidimensional search and optimization." As pointed out above, any problem involving three or more design parameters ("decision variables") is a multidimensional problem, and simple methods such as plotting the function to be maximized cannot be used. Methods for solving these problems fall into two broad categories: analytical methods and heuristic methods. Analytical methods, which involve computing derivatives and gradients, are of limited use, especially in the complex landscapes associated with antenna design. Stringent performance requirements in terms of bandwidth, radiation pattern, and standing wave ratio (SWR) make antenna optimization problems particularly difficult because the landscape is usually extremely multimodal with narrow resonances and often high sensitivity to slight parameter variations. Heuristic optimization methodologies, which are inherently numerical in nature, are effective in dealing with these issues, and consequently they are considered here while analytical approaches are not.

An entire class of heuristic optimization algorithms are "Nature inspired", and these appear to be the most effective. A Nature inspired algorithm is a computer search and optimization program whose function mimics some natural process. These programs are described as being "metaphorical" because they analogize some natural process without precisely modeling it. For example, "Ant Colony Optimization" (ACO) is an algorithm that simulates (to some degree) the behavior of ants seeking food. Thus, ACO is *inspired* by the *metaphor* of ant foraging. All such algorithms *evolve* a solution to the optimization problem over a series of steps or iterations, and almost all such algorithms are *stochastic population*-based methodologies. An initial population (of ants, for example) randomly (stochastically) moves through the decision space (landscape) step-by-step (iterating) in such a way that it converges on the largest food supply (maximum function value). The ants' progress is controlled by a set of equations that mimic real ant behavior in Nature. There are many Nature inspired algorithms, ACO being one of the earliest ones. The more important algorithms are discussed below with examples of their application to antenna optimization.

2.1.3 Ant Colony Optimization

Figure 1 illustrates the basic idea behind Ant Colony Optimization (ACO) [1]. The irregular objects represent the ants' nest (bottom) and a desirable source of food (top). It has been observed that ants seeking food eventually traverse the shortest path between the nest and food by marking that trail with a chemical pheromone that each ant can sense (probably by smell). If the path is unobstructed [(a) in the figure], then the ants simply walk a more-or-less straight line between home and the food supply. But, if an obstruction is imposed [(b) and (c) in the figure], then more ants eventually end up on the shorter trail between the food and the nest, which in turn results in a greater pheromone concentration along that "optimal" trail. By depositing progressively more pheromone on the shortest path, almost all of the ants eventually end up on that path, and the "best" solution has been found. The red lines in the bottom part of the diagram illustrate the path evolution with the eventual result that the shortest path is identified.

The ACO algorithm mimics the ants' behavior using equations that represent the *random* motion of individual ants subject to their pheromone environment. Instead of searching for food, the metaphorical ACO ants search the landscape of a decision space for the maximum value of the function to be maximized. But the process they follow is a simplified model of ant behavior as observed in Nature. And, just as real ants eventually discover the best food source, ACO's "ants" eventually converge on the function's global maximum value.

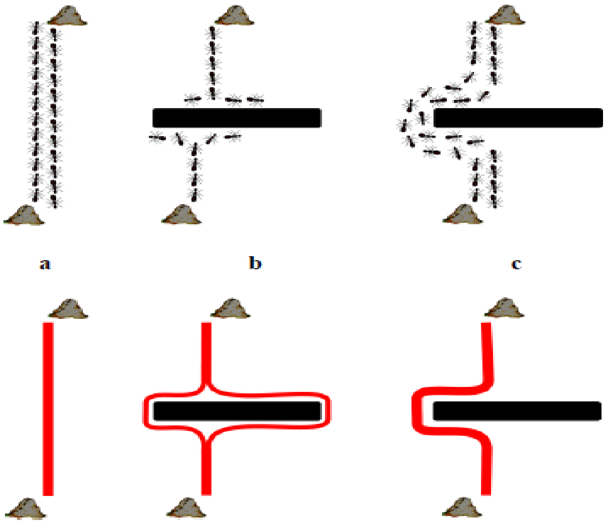

Fig. 1 Ant Colony Optimization: ants searching the shortest route using a pheromone trail.

Figure 1. Ant Colony Optimization Metaheuristic (reproduced from [1]).

2.1.4 Particle Swarm Optimization

Particle Swarm Optimization (PSO) [2] is another stochastic population-based Nature inspired evolutionary algorithm. PSO analogies the swarming behavior of fish or bees seeking food. Unlike ACO in which each "ant" creates a pheromone trail for other ants to follow, PSO's population of "agents" collectively communicate two pieces of information: each individual agent's "best" solution (greatest food concentration) and the population's overall (global) best solution. Equations that mimic bee and fish swarming then control each agent's subsequent motion in the decision space based on the competing tendencies of moving toward the global best and randomly exploring the vicinity of its best solution. As shown in Figure 2 for bees swarming around a flower concentration, after many steps PSO agents converge on the global best solution (highest flower concentration) because the local search fails to reveal any better solutions.

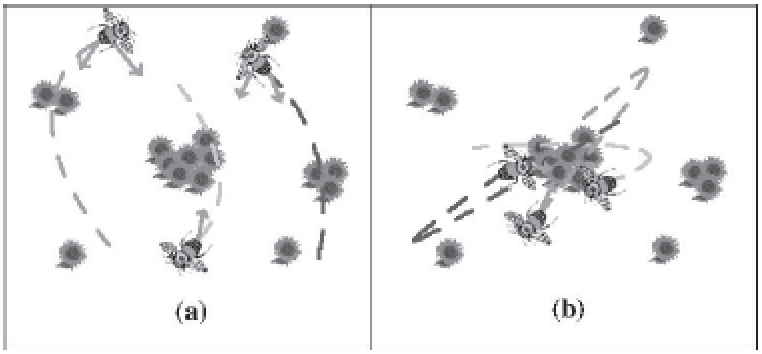

Figure 2. Particle Swarm Optimization metaheuristic (reproduced from [2]).

2.1.5 Genetic Algorithms

A Genetic Algorithm (GA) [3] analogizes the process of natural evolution or "survival of the fittest." When biological parents "mate," they exchange DNA to create a new individual ("child") whose characteristics are drawn from both parents by combining the parents' DNA. A GA creates successive generations of children who then serve as parents for the next generation whose children, in turn, will exhibit better "fitness" than the previous generation. In the context of search and optimization, the fitness is the value of the function to be maximized, so that the "best" fitness corresponds to the function's global maximum. As the GA progresses generation after generation, the best discovered fitness improves and eventually converges on the function's global maximum.

Figure 3 shows a typical GSA flowchart for an antenna optimization algorithm. It starts with a definition of the decision space (parameters to be optimized) and the “fitness function” to be maximized (for example, antenna directivity, or some specified combination of performance parameters such as gain, bandwidth, and so on). An initial population of “individuals” is randomly created, and each individual is defined by a *chromosome* that may be a binary sequence or a real number. Each chromosome comprises a set of *genes*, and each gene is one of the design parameters. For example, if the three design parameters were element length, inter-element spacing, and element diameter in a four element Yagi-Uda array, then there is a total of eleven design parameters, and each one is a gene. Thus, the optimization problem is defined on an 11-dimensional decision space, and the objective is to determine each of the eleven parameters so as to maximize some specific *fitness* function, say, the array’s gain-bandwidth product. A separate computer program is used compute the fitness at each step for each chromosome (the “evaluate fitness” box in Figure 3).

After the initial population’s fitnesses are evaluated, the “selection” process chooses two parent chromosomes that will mate (“crossover”) to produce two children chromosomes in the next generation. The selection and crossover processes take many varied forms. For example, the selection of parents may be random, or “best mates worst,” or best pairs pair wise through the population, and so on. The crossover operation likewise can take many forms. For example, the parents’ chromosomes may be split at the midpoint with first and second parts being swapped, or a random break point might be used, or some other combinatorial approach taken. Finally, the children thus created are subject to some level of *mutation*, a random perturbation of the chromosome structure just as real chromosomes are mutated in Nature. The steps described thus far are essentially common to all Gas, but the next step in the flowchart (“elitist model”) is not. In this GA, the worst individual in the new generation is replaced by the best individual from the previous generation, thus preserving the best solution from generation to generation as the algorithm progresses. As a final step, the best fitness is tested for convergence, and the process repeated until convergence is achieved.

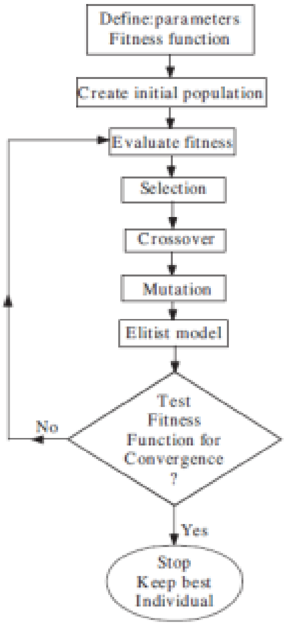
Define:parameters
Fitness function
Create initial population
Evaluate fitness
Selection
Crossover
Mutation
Elitist model
Test
Fitness
Function for
Convergence
?
No
Yes
Stop
Keep best
Individual

Figure 2. Flow chart diagram of real-coded GA.

Figure 3. GA flow chart (reproduced from [3]).

2.1.6 Simulated Annealing

Simulated Annealing (SA) [4] is a stochastic algorithm based on a metaphor drawn from physics instead of biology, as ACO, PSO, and GA are. SA analogizes the statistical mechanics of physical systems in thermal equilibrium with many degrees of freedom. In particular, the physical processes involved in annealing a solid as it cools forms the basis of the SA optimization algorithm, which has proven effective in optimizing problems with large numbers of decision variables. Because of SA's complexity, the algorithm is not described in detail. Instead its performance against a classic test problem is discussed.

The Traveling Salesman Problem (TSP) is a recognized example of combinatorial optimization that SA was used to solve because it constitutes a good test of an algorithm's effectiveness. The salesman must visit *N* different cities once each and return to his starting point. The problem is to determine the least costly route using a "cost" or "objective" function that is specified beforehand. Minimizing the cost is the same as maximizing its negative (note that minimization and maximization problems are exactly the same except for multiplying the objective function by -1). The TSP is a multidimensional search and optimization problem in the same vein as an antenna optimization problem, so that an algorithm suitable for one very likely is applicable to the other.

For the SA test, the TSP cost function is simply the total distance travelled by the salesman (to be minimized). Two different distance metric can be used, the standard Euclidean distance ("square root of the sum of the squares"), or the "Manhattan" metric (sum of the separations along the two coordinate axes), the latter being used in this case because it is simpler (less computationally intensive). Evolved solutions for TSP appear in Figure 6 and show a clear tendency towards removing redundancy in the travelled route, with the final solution (d) being close to optimal as discussed in [4].

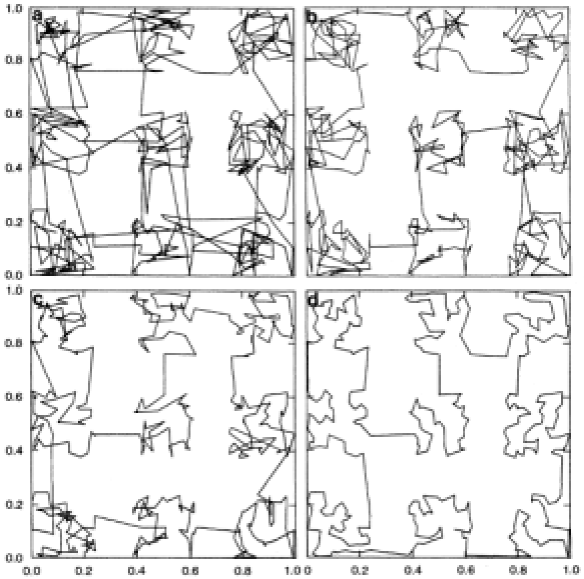

Figure 4. Evolution of SA solutions to TSP (reproduced from [4]).

2.1.7 Central Force Optimization

Central Force Optimization (CFO) [5] is a new algorithm that departs significantly from all other Nature inspired metaheuristics. ACO, PSO, SA, and the other algorithms described below are all inherently *stochastic*. Every run with the same setup parameters in general produces a different set of solutions. No two runs yield the same results because these algorithms rely on true randomness in their functioning. The values of certain key variables in the algorithm are, by definition, random variables that are computed from a probability distribution. The values of these variables must vary from one calculation to the next, and their values are completely unknown and unknowable until the probabilistic calculation is performed.

CFO is quite different. It is based on an analogy drawn from gravitational kinematics, which in turn is based on Newton's laws of gravity and motion. Newton's laws are mathematically precise (completely *deterministic*) and, as a result, so too is CFO. CFO searches the decision space by "flying" probes through it whose trajectories are computed by deterministic equations that analogize Newton's laws of motion.

Figure 5 shows how CFO's probes move through a 3D decision space at each time step sampling the decision space by computing the fitness of the function to be maximized (shown by the darkened circles). CFO thus provides some major advantages over stochastic algorithms, *viz*, every run with the same setup returns exactly the same answers, and because of that characteristic only one run is necessary (stochastic algorithms usually are run many times and the results averaged). CFO has been effectively used for antenna optimization, and it holds considerable promise for use in set-top antenna design.

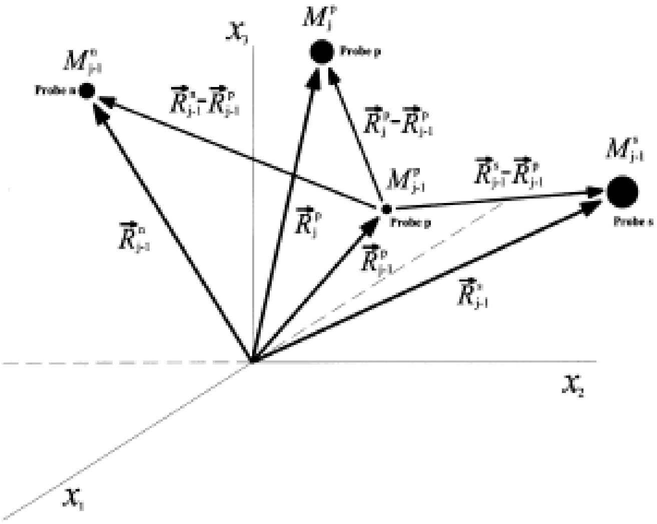

Figure 5. Central Force Optimization metaheuristic (reproduced from [5]).

2.1.8 Invasive Weed Optimization

Invasive Weed Optimization (IWO) [6] draws its inspiration from the colonization characteristics of invasive flora as understood from weed biology and ecology. Like ACO, PSO, GA, and SA, IWO is a population-based stochastic algorithm. Weeds exhibit a very strong tendency to opportunistically occupy (colonize) the interstitial spaces is a cropping field. Spaces not occupied by crops, which usually do not spread, become weed-filled, and the weed then grows and propagates by consuming unutilized resources in the field. The weed that uses these resources most effectively becomes the dominant (fittest) weed. When a weed flowers, it produces seeds that are randomly dispersed throughout the field until all interstitial space is occupied and all resources utilized.

Figure 6 shows a flow chart the IWO implementation used to solve electromagnetic problems in [6]. This flowchart starts out much the same as the GA flowchart with a randomly generated population whose fitness is evaluated in the initial step. Each weed then produces a number of seeds (reproduction) based on its fitness, with weeds having better fitnesses being allowed to produce more seeds. The seeds are then randomly dispersed through the decision space using a normal (Gaussian) distribution of random numbers with mean value equal to the weed's location. After the new seeds have been dispersed, they are allowed to grow into new flowering weeds, and the process is repeated until a convergent solution is generated. Because the number of weeds grows constantly, a maximum weed population serves as a ceiling on weed count. Whenever it is exceeded, the bottom worst plants are "weeded out" by being discarded.

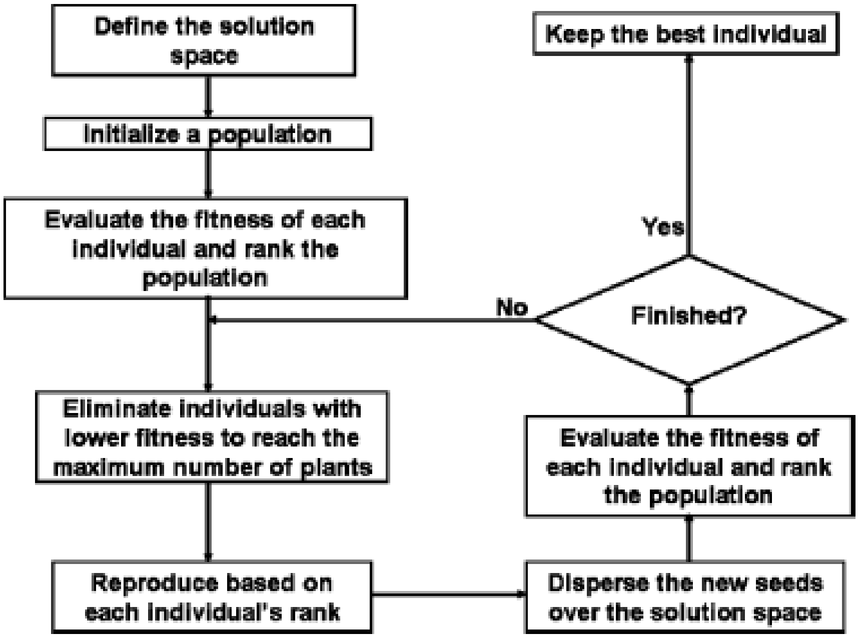
Define the solution space
Initialize a population
Evaluate the fitness of each individual and rank the population
Eliminate individuals with lower fitness to reach the maximum number of plants
Reproduce based on each individual's rank
Disperse the new seeds over the solution space
Evaluate the fitness of each individual and rank the population
Finished?
No
Yes
Keep the best individual

Figure 6. Invasive Weed Optimization flow chart (reproduced from [6]).

2.1.9 Intelligent Water Drop Algorithm

The Intelligent Water Drop algorithm (IWD) [7], like SA and CFO, analogizes a physical process. But, like SA and unlike CFO, it is stochastic in nature instead of deterministic. IWD is inspired by the notion that the seemingly random meanders in a river or stream bed are, in fact, based on mechanisms that can be applied to effectively solve optimization problems. Two principal factors are considered in IWD: water velocity and soil characteristics. Each IWD flows from a source to a destination, initially with non-zero velocity and zero soil. Along its route, velocity and soil both may be gained and lost. Soil inhibits drop velocity, so that between points the IWD's velocity increases inversely with soil (in a non-linear manner). Figure 7 shows typical IWD results for the Travelling Salesman Problem (see also discussion above on SA and TSP).

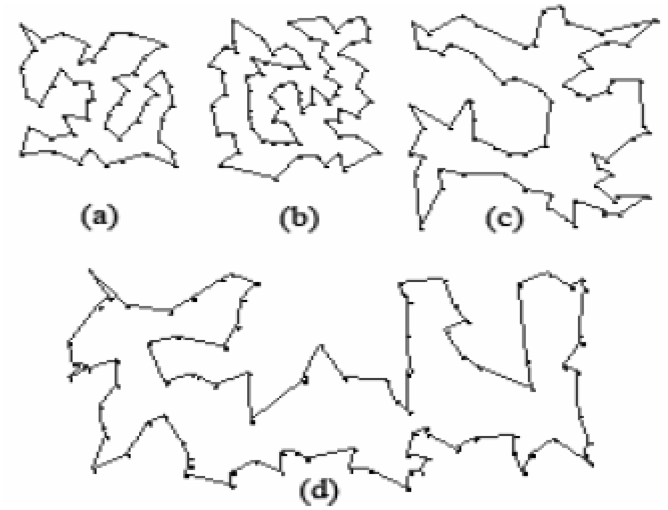

Figure 7. IWD results for the Traveling Salesman Problem (reproduced from [7]).

2.1.10 Bacteria Foraging Optimization

Bacteria Foraging Optimization (BFO) [8] mimics the natural behavior of bacterial seeking food. The motion of individual bacteria is driven by avoiding noxious elements in the environment while "swimming" upward along the food concentration gradient hopefully locating a denser source of food. *Chemotaxis* is the process by which a bacterium tumbles to orient itself, swims a fixed distance, and samples the food concentration at its new location. If the concentration is greater than at the previous location, then the bacterium continues in the same direction for another step. But if the concentration is lower, then the bacterium tumbles into a new direction and explores it instead. Each bacterium has a finite lifetime that limits the number of steps it can take. At the end their lifetime bacteria that are in the highest food concentration regions are allowed to "reproduce" by splitting into two new bacteria, while those in other regions die. Because the bacteria are dispersed, after the reproduction step, only the best half are retained, while the others die, thus preserving the total number of bacteria. The user specifies the maximum number of chemotactic and reproductive steps, maximum swim length, and maximum swims in a given direction.

2.1.11 An Antenna Optimization Example

This section presents a specific example of an optimized antenna design, a Self-structuring antenna (SSA) that can alter its electrical shape in near-real time in response to factors such as a changing environment or a change in mission. Typically these structures comprise wire segments that can be interconnected using control signals from a micro-controller device. An example of a SSA that was optimized using three different optimization algorithms appears in Figure 8 [9]. The antenna in this case is an asymmetric wire array containing 39 switches resulting in 549 *billion* possible antenna configurations.

The specific problem addressed in [9] was whether or not the optimization algorithms would provide better performance than a simple random search for a configuration that met specific performance criteria.

The wire structure was modeled with the Numerical Electromagnetics Code (NEC) following all modeling guidelines with respect to wire segmentation and segment length to diameter and wavelength ratios. The objective was to obtain low VSWR (< 2) values at frequencies of 40, 100 and 400 MHz. The antenna grid measured 0.6 meter square (0.08λ on a side at 40 MHz), which electrically is quite small. Figures 8, 9 and 10 compare the results of a random search to those produced by the three optimization algorithms. Random search performed very poorly at the lowest frequency, while each of the optimization algorithms performed relatively well. The GA produced the “tightest” results at 40 MHz (minimum standard deviation), and achieved the design objective in 60% of the trials. This example shows that real-time-optimized SSAs are within reach for the set-top antenna application at this time, and that fairly economical designs may be possible.

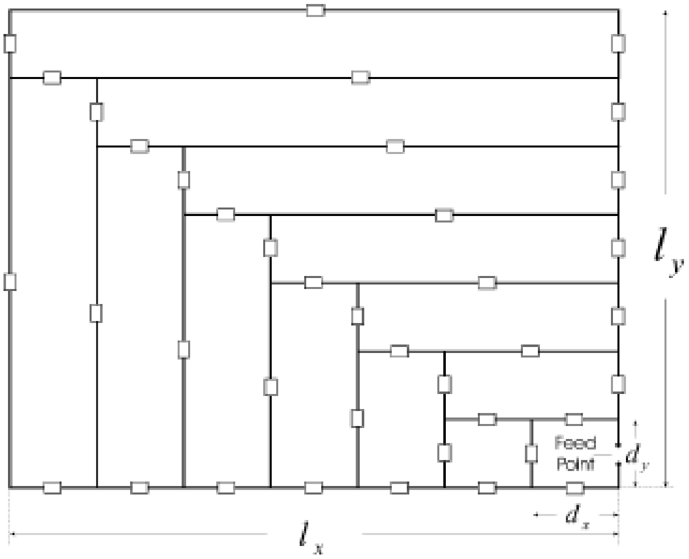

Figure 8. SSA geometry (reproduced from [9]).

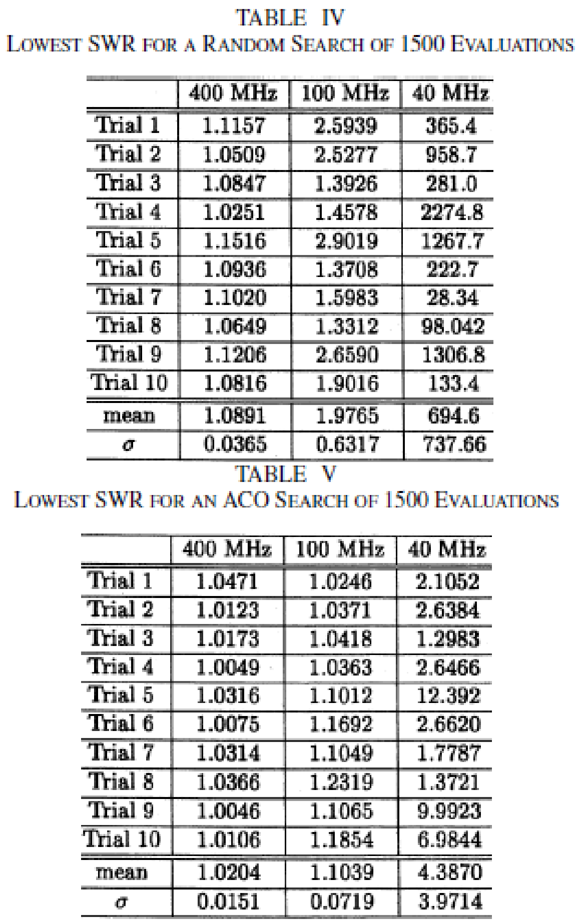

TABLE IV

Lowest SWR for a Random Search of 1500 Evaluations

	400 MHz	100 MHz	40 MHz
Trial 1	1.1157	2.5939	365.4
Trial 2	1.0509	2.5277	958.7
Trial 3	1.0847	1.3926	281.0
Trial 4	1.0251	1.4578	2274.8
Trial 5	1.1516	2.9019	1267.7
Trial 6	1.0936	1.3708	222.7
Trial 7	1.1020	1.5983	28.34
Trial 8	1.0649	1.3312	98.042
Trial 9	1.1206	2.6590	1306.8
Trial 10	1.0816	1.9016	133.4
mean	1.0891	1.9765	694.6
σ	0.0365	0.6317	737.66

TABLE V

Lowest SWR for an ACO Search of 1500 Evaluations

	400 MHz	100 MHz	40 MHz
Trial 1	1.0471	1.0246	2.1052
Trial 2	1.0123	1.0371	2.6384
Trial 3	1.0173	1.0418	1.2983
Trial 4	1.0049	1.0363	2.6466
Trial 5	1.0316	1.1012	12.392
Trial 6	1.0075	1.1692	2.6620
Trial 7	1.0314	1.1049	1.7787
Trial 8	1.0366	1.2319	1.3721
Trial 9	1.0046	1.1065	9.9923
Trial 10	1.0106	1.1854	6.9844
mean	1.0204	1.1039	4.3870
σ	0.0151	0.0719	3.9714

Figure 9. SSA VSWR, random search [left], ACO [right] (reproduced from [9]).

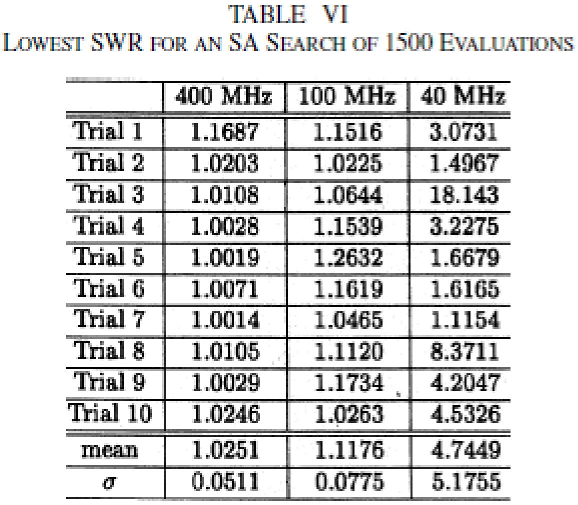

TABLE VI
LOWEST SWR FOR AN SA SEARCH OF 1500 EVALUATIONS

	400 MHz	100 MHz	40 MHz
Trial 1	1.1687	1.1516	3.0731
Trial 2	1.0203	1.0225	1.4967
Trial 3	1.0108	1.0644	18.143
Trial 4	1.0028	1.1539	3.2275
Trial 5	1.0019	1.2632	1.6679
Trial 6	1.0071	1.1619	1.6165
Trial 7	1.0014	1.0465	1.1154
Trial 8	1.0105	1.1120	8.3711
Trial 9	1.0029	1.1734	4.2047
Trial 10	1.0246	1.0263	4.5326
mean	1.0251	1.1176	4.7449
σ	0.0511	0.0775	5.1755

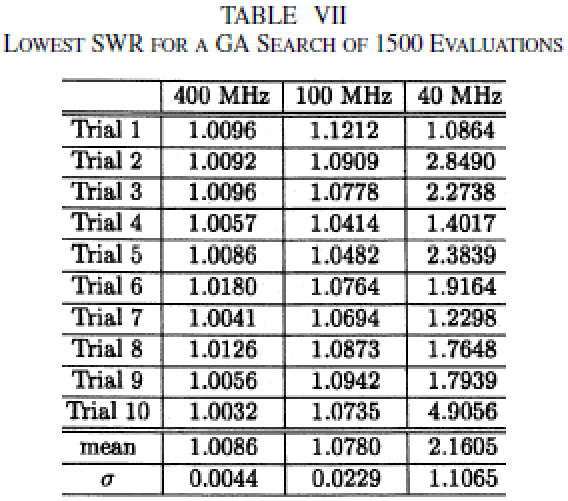

TABLE VII
LOWEST SWR FOR A GA SEARCH OF 1500 EVALUATIONS

	400 MHz	100 MHz	40 MHz
Trial 1	1.0096	1.1212	1.0864
Trial 2	1.0092	1.0909	2.8490
Trial 3	1.0096	1.0778	2.2738
Trial 4	1.0057	1.0414	1.4017
Trial 5	1.0086	1.0482	2.3839
Trial 6	1.0180	1.0764	1.9164
Trial 7	1.0041	1.0694	1.2298
Trial 8	1.0126	1.0873	1.7648
Trial 9	1.0056	1.0942	1.7939
Trial 10	1.0032	1.0735	4.9056
mean	1.0086	1.0780	2.1605
σ	0.0044	0.0229	1.1065

Figure 10. SSA VSWR, SA [left], GA [right] (reproduced from [9]).

2.1.12 References

[1] Pechac, P., "Electromagnetic Wave Propagation Modeling Using the Ant Colony Optimization Algorithm," *Radioengineering*, **11**, No. 3, September 2002, p. 1.

[2] Robinson, J., and Rahmat-Samii, Y., "Particle Swarm Optimization in Electromagnetics," *IEEE Trans. Antennas and Propagation*, **52**, no. 2, Feb. 2004, p. 397.

[3] Mahanti, G. K., Pathak, N., and Mahanti, P., "Synthesis of Thinned Linear Antenna Arrays with Fixed Sidelobe Level Using Real-Coded Genetic Algorithm," *Prog. in Electromagnetics Research*, **PIER 75**, 319-329, 2007.

[4] Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P., "Optimization by Simulated Annealing," *Science*, **220**, no. 4598, 13 May 1983, p. 671.

[5] Formato, R. A., "Central Force Optimization: A New Metaheuristic with Applications in Applied Electromagnetics," *Prog. in Electromagnetics Research*, **PIER 77**, 425-491, 2007.

[6] Karimkashi, S., and Kishk, A. A., Invasive Weed Optimization and its Features iElectromagnetics," *IEEE Trans. Antennas and Propagation*, **58**, no. 4., April 2010, p. 1269.

[7] Shah-Hosseini, H., "The Intelligent Water Drops Algorithm: A Nature-Inspired Swarm- Based Optimization Algorithm," *Int. J. Bio-Inspired Computation*, **1**, nos.1/2, p. 71, 2009.

[8] Datta, T., and Misra, I. S., "Adaptive antenna Array Processing: The Bacteria-Foraging Algorithm and Particle-Swarm Optimization," *IEEE Antennas and Propagation Magazine*, **51**, no. 6, December 2009, p. 69.

[9] Coleman, C. M., Rothwell, E. J., and Ross, J. e., "Investigation of Simulated Annealing, Ant-Colony Optimization, and Genetic Algorithms for Self-Structuring Antennas, *IEEE Trans. Antennas and Propagation*, **52**, 4, p. 1007, April 2004.

2.2 Fragmented Antennas

As an additional illustration of the power of an electro-magnetic (EM) optimization algorithm, this section describes the computer program ***FAopt*** that begins with a grid of wires as shown in Figures 1 and 2 and optimizes the placement of these wires using a Binary Genetic Algorithm (BGA) and the NEC-4 [1] Method of Moments (MoM) FORTRAN code. Each bit in the BGA chromosome is either a 1 or a 0, representing the presence or absence of a wire, respectively. Quadrant 1 and 3 are a mirror image of each other as well as quadrant 2 and 4, so that the antenna is symmetrical about its diagonal. The resulting antenna is comprised of a series of plated conductors, some of which are connected to one another. Some are capacitively/inductively coupled and act as parasitic elements to increase the element impedance and radiation pattern bandwidth. These types of radiating elements, when implemented as pixels rather than wires, (see Maloney et al of GTRI [2]) are generally called "fragmented" antennas [3]. ***FAopt*** allows the user to choose the polarization as either vertical or horizontal, and the antenna can be either directional or omni-directional. The user also can choose the dimensions for the desired antenna, as well as the frequency range. This program was used to design a set-top antenna to operate in the VHF and UHF bands. (See Section 3.0). During optimization a window is shown that updates the progress of the optimizer with the best "figure of merit" (FoM) displayed along with the standing wave ratio.

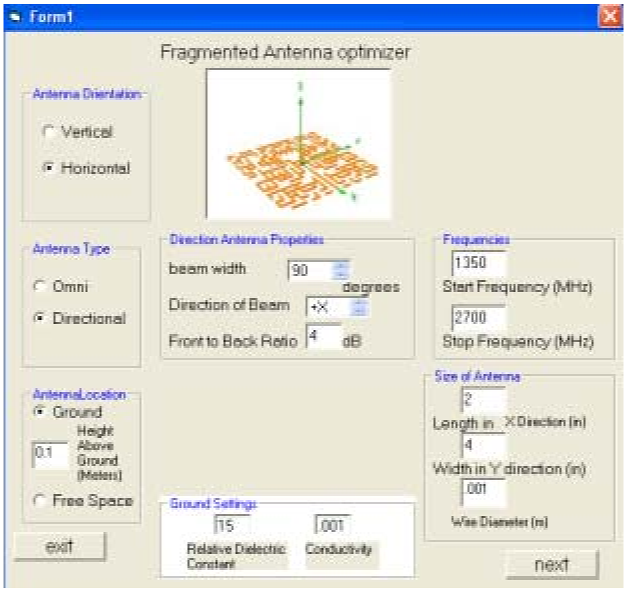

Antenna Orientation
○ Vertical
● Horizontal

Antenna Type
○ Omni
● Directional

AntennaLocation
● Ground
0.1 Height Above Ground (Meters)
○ Free Space

Direction Antenna Properties
beam width 90 degrees
Direction of Beam +X
Front to Back Ratio 4 dB

Frequencies
1350 Start Frequency (MHz)
2700 Stop Frequency (MHz)

Size of Antenna
2 Length in X Direction (m)
4 Width in Y direction (m)
.001 Wire Diameter (m)

Ground Settings
15 Relative Dielectric Constant
.001 Conductivity

Figure 1: FAopt Code's GUI

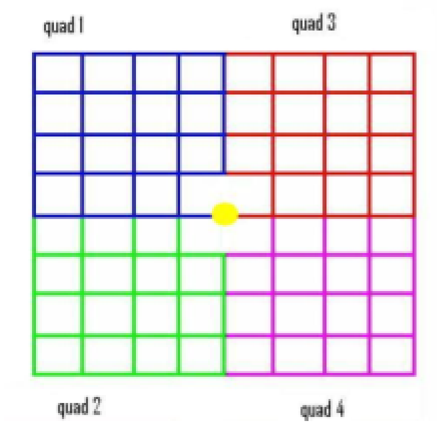

Figure 2: Wire Grid

As an example, a thin planar directional antenna was optimized using ***FAopt*** as a horizontally polarized antenna from 2400-2480MHz with 8 blocks per row/column per quadrant. The orientation appears in Figure 1, and the antenna model is shown in Figure 3. The dimensions for this antenna are 2” by 2”. A prototype was built as shown in Figure 4.

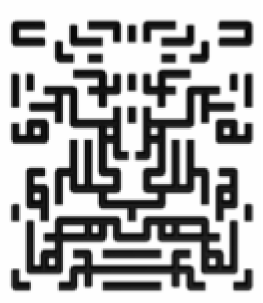

Figure 3. Fragmented Directional WiFi Antenna

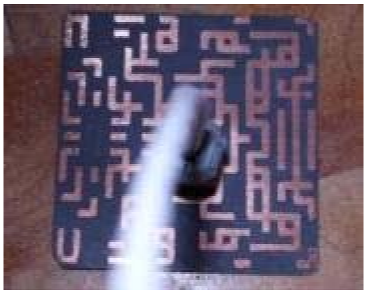

Figure 4. Fragmented Directional WiFi Antenna

The simulation results, which were confirmed by measurements, shows the average gain of this antenna to be better than 1.9 dBi. This antenna is a directional WiFi antenna optimized to have a F/B ratio of about 6 dBi. Its radiation pattern is very consistent across the band as shown in Figure 6, and its measured VSWR is better than 2:1 across the band. Further work, such as changing the grid template or the optimization routine, could be undertaken to make this approach more efficient for designing set-top VHF/UHF antennas.

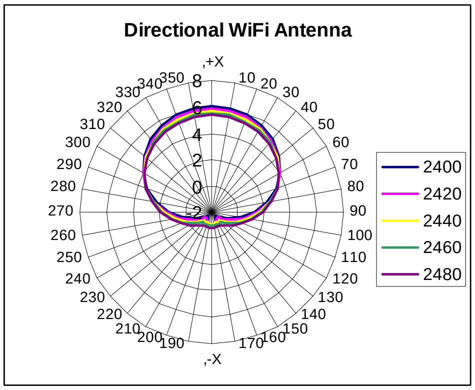

Figure 5. Pattern of Directional WiFi Antenna

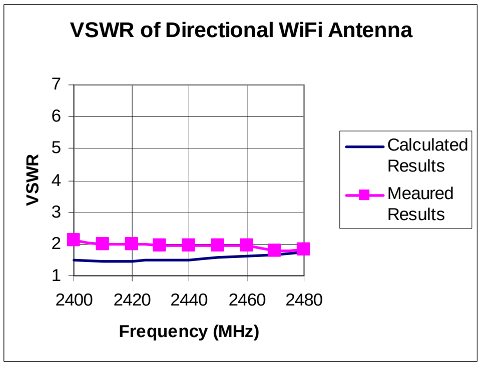

Figure 6. VSWR of Directional WiFi Antenna

The F/B ratio of simulated and measured are in fair agreement as shown in Figure 7.

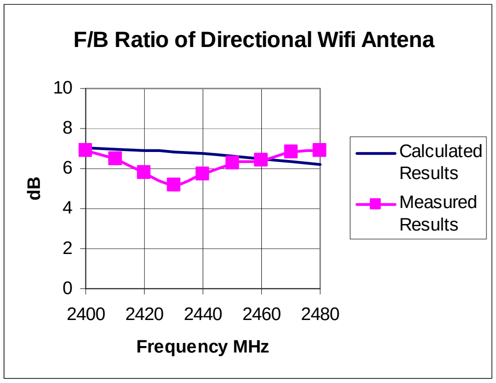

Figure 7. F/B Ratio of Directional WiFi Antenna

References

[1] G. J. Burke, “Numerical Electromagnetics Code-NEC-4 Method of Moments”, Lawrence Livermore National Laboratory, 1996

[2] J. G. Maloney, M. P. Kesler, P. H. Harms, G. S. Smith, United States Patent No. US 6323809 B1, Nov 27,2001

[3] B. Thors, H. Steyskal and H. Holter, “Broad-Band Fragmented Aperture Phased Array element Design Using Genetic Algorithms”, IEEE Trans. *Antennas Propag.,* Vol 53, no. 10, pp. 3280-3287

2.3 Non-Foster Impedance Matching

Foster's reactance theorem, which dates to 1924, states that a lossless reactance must have a positive slope with frequency. Any lossless matching network presumably satisfies the theorem, but a great deal of research that has been done on "non-Foster matching" in the past 15 years suggests otherwise. This section examines how non-Foster networks may be useful in set-top antennas to match an electrically small element (characterized by a low real and high negative imaginary complex impedance) within, for example, the low-VHF band.

Until recently no one has created a practical antenna system using this method that showed much of an improvement in performance. This was due to limitations in the semiconductor technology at the time that permitted only narrow band solutions. Semiconductor technology has improved considerably in recent years resulting in increased bandwidth, lowered noise, and decreased losses in the devices. Electrically small antennas present high Q impedances characterized by large reactances and small radiation resistances. For such antennas, the effectiveness of passive matching circuits is severely limited by gain bandwidth theory, which predicts narrow bandwidths and/or poor gain. The result inevitably is a poor signal to noise ratio compared with a larger antenna.

Non-Foster matching uses negative impedance converters (NIC) to create "negative" capacitors and "negative" inductors [1-2]. It is possible, for example, to use negative capacitance to cancel out the reactance in a short dipole or monopole. This leaves only a small real impedance which could then be matched to, say, 75 Ω using a transformer. One disadvantage of this approach is that transformers add to the antenna size which could be undesirable. An alternative to matching could be achieved by placing the non-Foster matching circuit away from the feed to control both the real impedance as well as the reactance of the antenna [3]. The potential improvement in antenna performance is very significant. Theoretical bandwidth improvements have been shown for a small loop (increasing bandwidth from 50 MHz to 300MHz at a center frequency of 700MHz), as well as a small broadband dipole (increasing bandwidth from 250MHz to 1000MHz, lowering the lower operating frequency from 250 MHz to 190 MHz).

It should be noted that the Figures and Tables in this section were taken from references 2 through 5, listed at the end of this section.

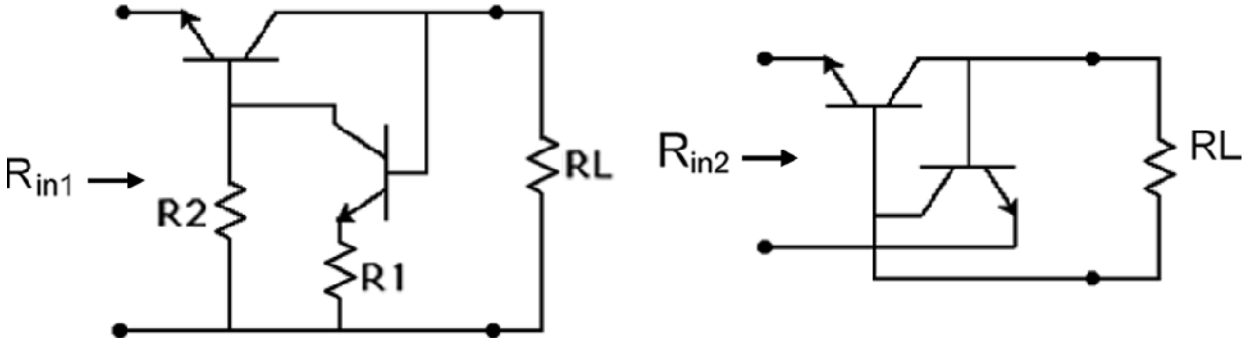

Figure 1. Linvill’s ideal OCS NICs terminated in a resistance.

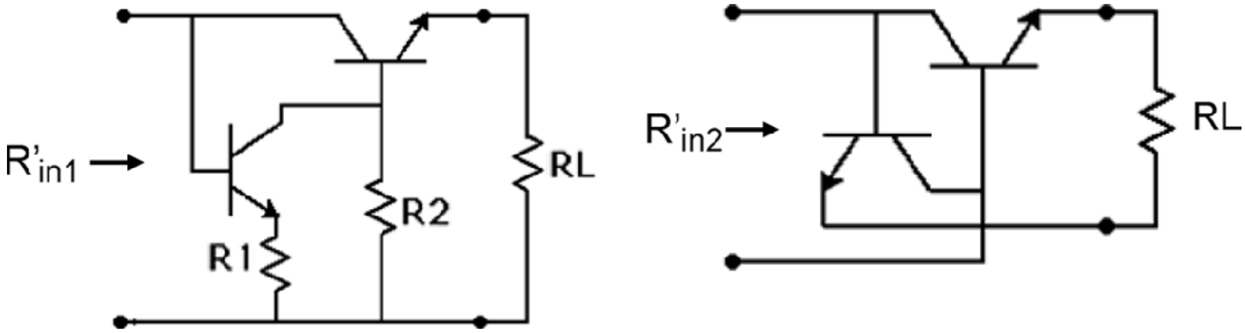

Fiugre 2. Linvill’s ideal SCS NICs terminated in a resistance.

With an ideal transistor, a pure negative resistance is achievable. In Linvill’s actual realizations (Figures 1 and 2), a substantial reactive component of the input impedance Z accompanies the negative resistance, resulting in a low Q, where the “resistive” Q is defined as: Q=Re(Z)/Im(Z). The circuits shown are “open-circuit stable” (OCS), meaning, practically, that if a very large resistance terminates the negative-resistance one-ports on the left, then the overall network will be stable. The networks of Figure 1 can be turned around as shown in Figure 2, where negative resistances of the “short-circuit stable” (SCS) variety are obtained. Again, practically speaking, this means these networks will be stable if a very large conductance is placed across the input. An initial step in improving the power efficiency of non-Foster matching is simply to use a single-transistor version of the NIC [4].

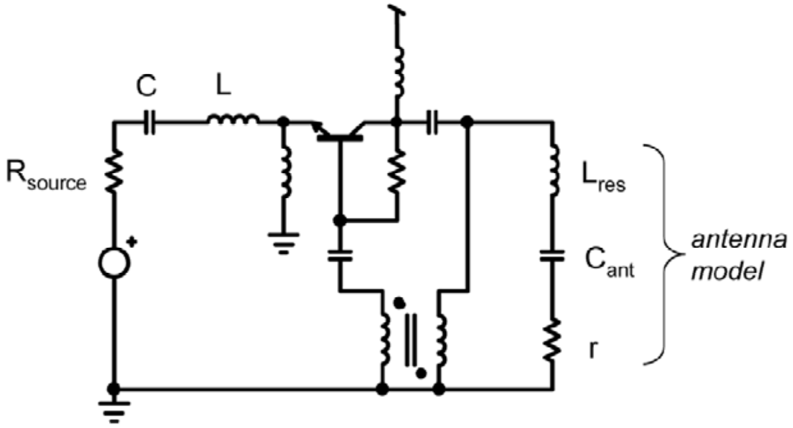

Figure 3. Non Foster Matching

An example of a non-Foster matching circuit is shown in Figure 3. Although not shown in the schematic, a transformer would be used at the input to make the effective value of R_{source}, as seen by the matching circuit, to be 18.75Ω (4:1 impedance ratio).The circuit negates antenna's Capacitance C_{ant} and the voltage-managing resonating inductor L_{res}, as well as the radiation resistance r. The idea is to cancel the negated L_{res} and C_{ant} with C and L on the left, respectively. The negated r is absorbed by the much larger source resistance and does not affect operation of the circuit. The source resistance would be 75 ohms in this case.

An example of non-Foster matching for a 12 inch dipole is shown in Figure 4. It is shown in Figure 5 that this non-Foster matching network creates a negative capacitance that is used to cancel out the reactance of the 12" dipole. This is used along with a transformer to help match the radiation resistance of the antenna to the source impedance of 50 Ω. The non-Foster matching network is compared to a similar dipole with a lossy passive network for improvements in signal to noise as shown in Figures 6 and 7. The improvement in SNR is quite remarkable, as much as 25 dB around 44 MHz.

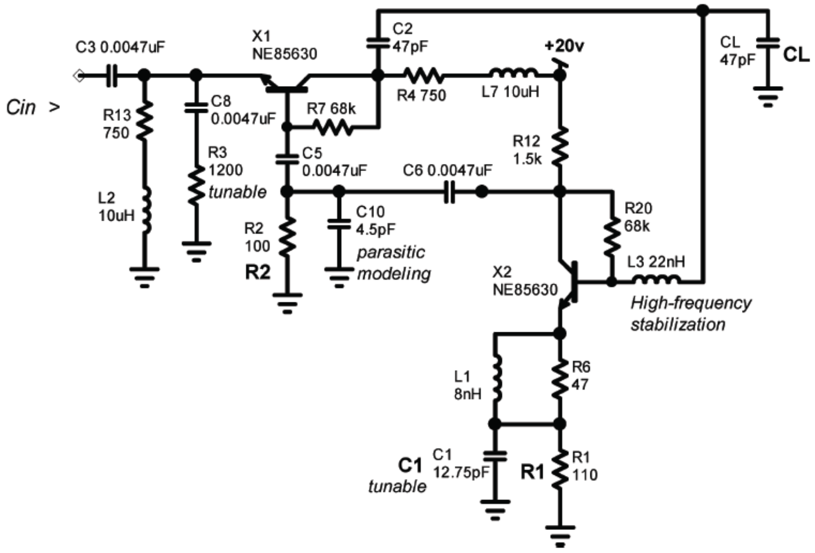
C3 0.0047uF
Cin >
R13 750
L2 10uH
C8 0.0047uF
R3 1200 tunable
X1 NE85630
C2 47pF
R7 68k
C5 0.0047uF
R2 100
R2
C10 4.5pF
parasitic modeling
R4 750
L7 10uH
+20v
R12 1.5k
C6 0.0047uF
R20 68k
L3 22nH
High-frequency stabilization
X2 NE85630
L1 8nH
R6 47
C1 12.75pF
C1 tunable
R1
R1 110
CL 47pF
CL

Figure 4. Non-Foster Matching Network for a 12" Dipole.

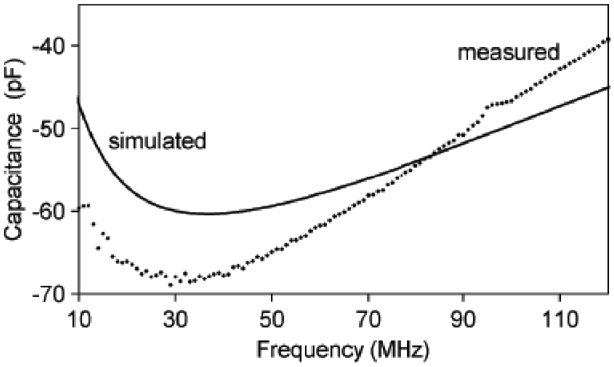

Figure 5. Measured and simulated results for capacitance of non-Foster matching network.

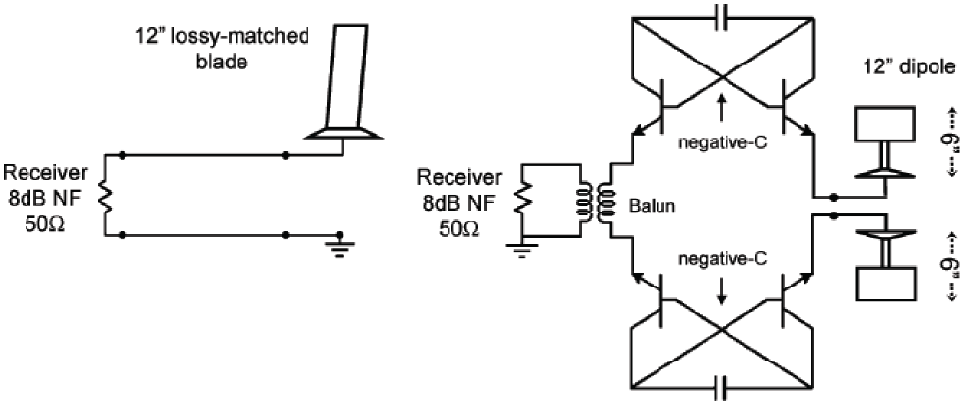

Figure 6. Setup for Signal to Noise improvement measurement.

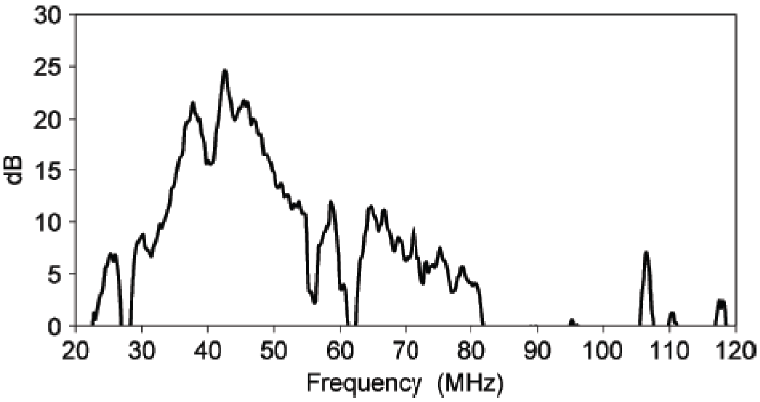

Figure 7. Measured Improvement of Signal to Noise

Non-Foster matching may be particularly useful for the set-top application because it works best with electrically small antennas. Therefore, this type of match could be useful when designing a small VHF/UHF DTV set-top antenna. Improvements in signal to noise measurements have been shown from 20MHz to 400 MHz [5]. In particular, the use of non-Foster techniques to impedance match a lossy electrically-small dipole antenna has been quite effective. On an antenna range, there was a measured 30 dB gain improvement over 60 – 200 MHz, with several dB of improvement as high as 400 MHz. Again, the comparison was to the same antenna with no matching at all. Although no S/N measurements were made, the circuits used were based upon the same low-noise designs developed earlier for lower-frequency circuits. Because of the lossiness of the antenna, passive matching can do a little better than no matching at all, and these results are illustrated in simulation. This computer simulation designed a number of "best-effort" passive matching networks and calculated the transducer gain (S21) between a 50 ohm source and the complex impedance of a 6-inch monopole. S21 for no matching and S21 for a single, ideal negative capacitance whose value (-5.54pF) exactly cancels the antenna reactance at 30 MHz, were calculated. Plots of the real and imaginary parts of the antenna impedance are shown in Figure 8. The various matching networks are illustrated in Figure 9; each of these would appear in place of the NIC in Figure 10b.

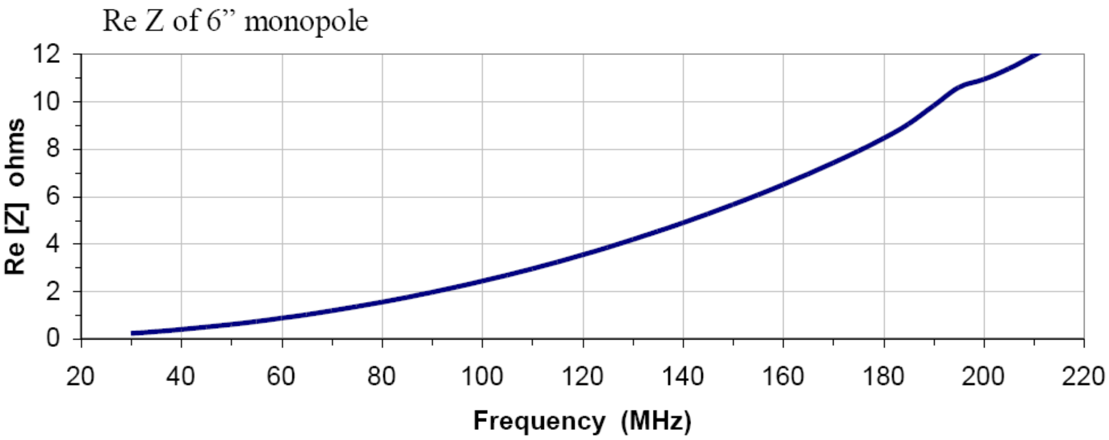

Figure 9-A. Real Part of Antenna Impedance for a six inch monopole

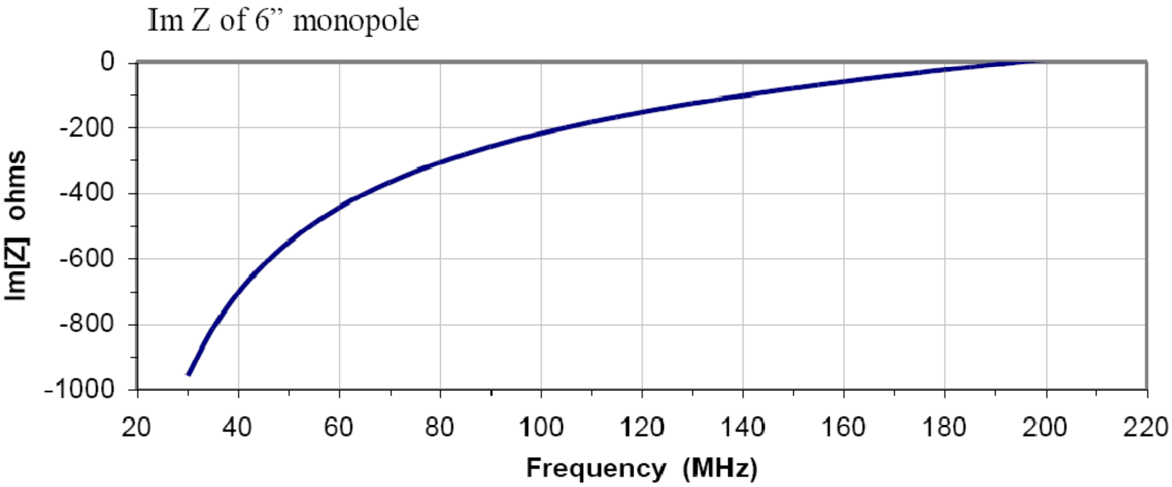

Figure 9-B. Imaginary Part of Antenna Impedance for a six inch monopole

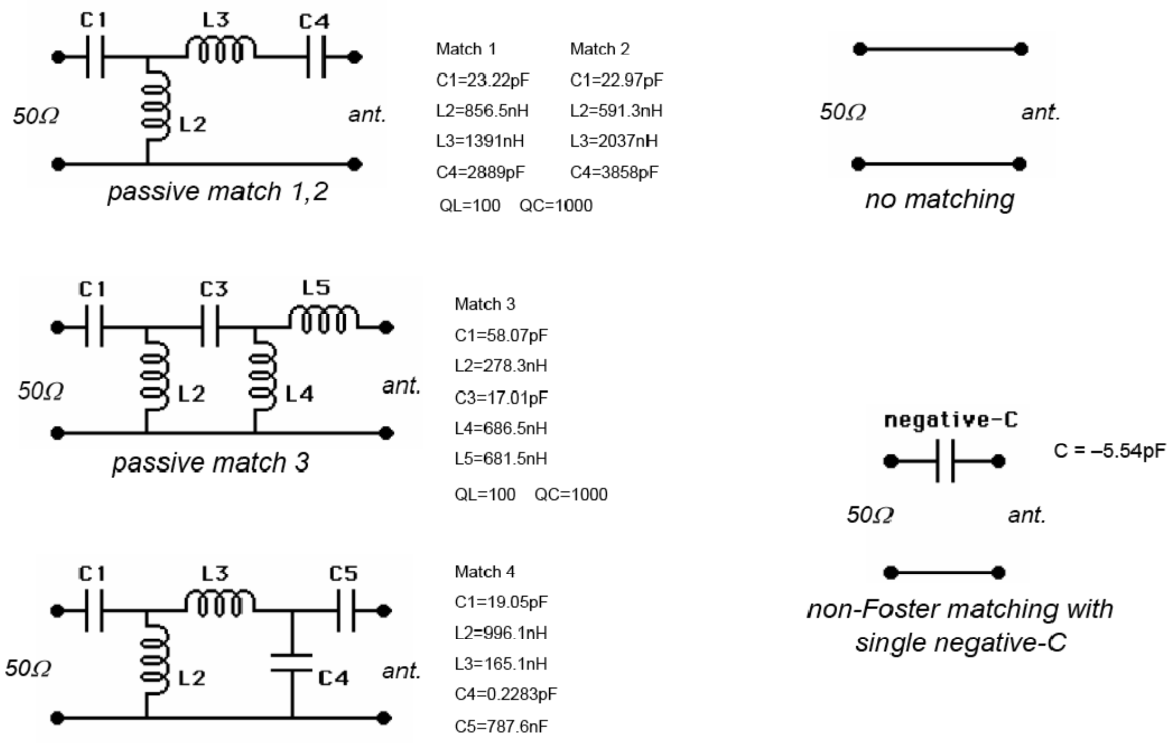
C1
L3
C4
50Ω
L2
ant.
passive match 1,2
Match 1
C1=23.22pF
L2=856.5nH
L3=1391nH
C4=2889pF
Match 2
C1=22.97pF
L2=591.3nH
L3=2037nH
C4=3858pF
QL=100 QC=1000
50Ω
ant.
no matching
C1
C3
L5
50Ω
L2
L4
ant.
passive match 3
Match 3
C1=58.07pF
L2=278.3nH
C3=17.01pF
L4=686.5nH
L5=681.5nH
QL=100 QC=1000
negative-C
C = –5.54pF
50Ω
ant.
non-Foster matching with
single negative-C
C1
L3
C5
50Ω
L2
C4
ant.
Match 4
C1=19.05pF
L2=996.1nH
L3=165.1nH
C4=0.2283pF
C5=787.6nF

Figure 10. Various Matching Techniques for six inch monopole

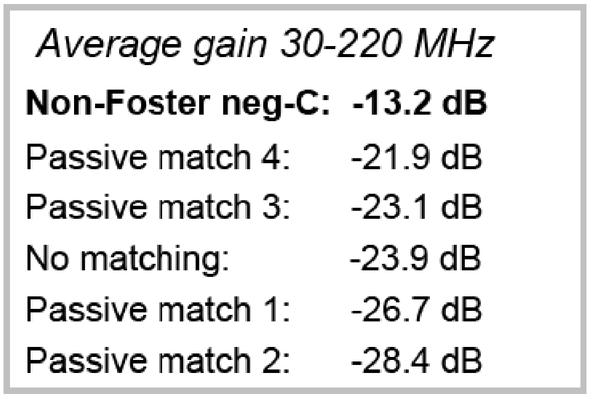

Average gain 30-220 MHz

Non-Foster neg-C:	**-13.2 dB**
Passive match 4:	-21.9 dB
Passive match 3:	-23.1 dB
No matching:	-23.9 dB
Passive match 1:	-26.7 dB
Passive match 2:	-28.4 dB

Table 1. Computed Average Gain of all matching techniques.

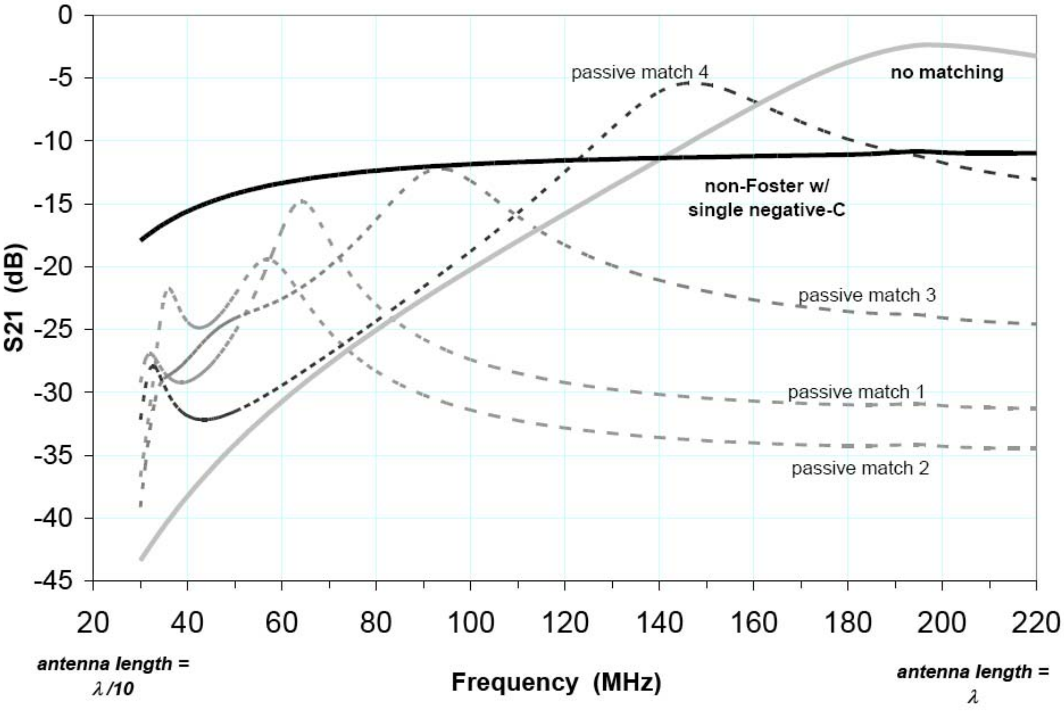

Figure 11. Plots of Computed Gains of all matching techniques for a six inch monopole

The six inch monopole was then built using both the non-Foster matching network and an unmatched antenna. They were both measured with the results shown in Figure 12. In the design, NE68019 BJTs were used in the NICs because these devices provided good overall performance over the entire frequency band. Again, the results are quite remarkable, with the non-Foster antenna showing average gain improvements between about 2 and 28 dB. It is noteworthy that the greatest improvements are at lower frequencies, between about 60 and 210 MHz, approximately the low and high VHF bands.

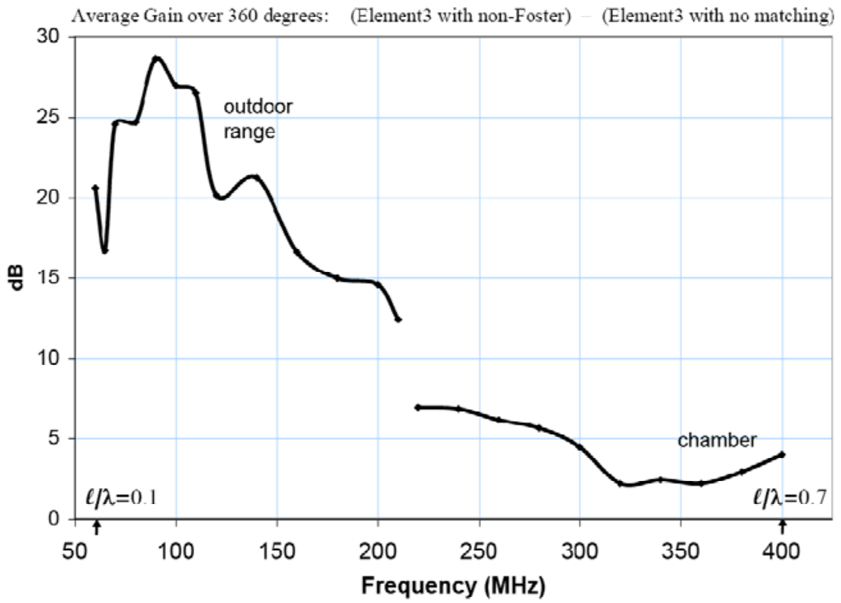

Figure 12. Gain of a non-Foster matching network over the same antenna with no matching.

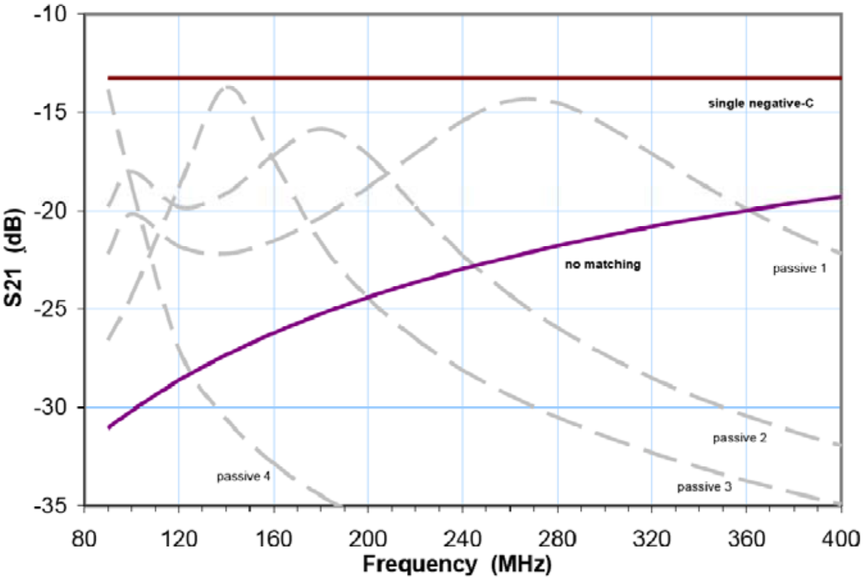

Figure 13. Plots of Computed Gains of all matching techniques for a six inch dipole

Results for an electrically small lossy dipole are shown in Figure 13 (this antenna was simulated, not measured). The actual gain may be less due to noise of the device if this were to be built and tested. Thus, realization of non-Foster technology is still limited by the analog circuitry performance. With respect to the set-top application, it is very reasonable to expect that as better silicon devices are developed covering the entire television frequency range will be possible with substantially better antenna performance in terms of gain and signal-to-noise ratio.

References

[1] R.C Hansen, “Electrically Small Super directive and Super conducting Antennas”, John Wiley & Sons Inc. 2006

[2] J. T Aberle, R. Leopsinger-Romak, C. A Balanis, “Antennas with Non-Foster Matching Networks”, Morgan & Claypool Publishers, 2007

[3] S. Koulouridis and J.L. Volakis, “Non-Foster circuits for small broadband antennas”,

[4] S. E. Sussman-Fort and R. M. Rudish “Non-Foster Impedance Matching of Electrically-Small Antennas”,IEEE Transactions on Antennas and Prop., Vol 57,NO. 8, August 2009

[5] S. E. Sussman-Fort and R. M. Rudish, “Non-Foster impedance matching of a lossy, electrically-small antenna over an extended frequency range” presented at the Antenna Applicat. Symp., Allerton Park, IL, Sep. 2007

2.4 Active RF Noise Cancelling

Considerable work has been done to mitigate noise from devices near receiving antennas, as is particularly useful in set-top antennas. An antenna placed on top of a television is especially vulnerable because it may pick up noise from internal circuitry. But this noise can be reduced, which in turn reduces the noise floor, or, equivalently, increases the signal-to-noise ratio (SNR). Intersil [1] has created the QHx220 chip that accomplishes this e.g. at UHF frequencies, or more precisely from 300 MHz up to >3GHz. MegaWave has evaluated the Intersil noise cancelation chip QHx220 shown in Figure 1. The chip was tested for signal to noise improvements at 535MHz as shown in Figure 2. Measured results showed an improvement of about 12dB in SNR as shown in Figures 3 and 4. This technology has been extended down to FM covering the VHF III band [2], which is necessary for this approach to be viable in the set-top environment.

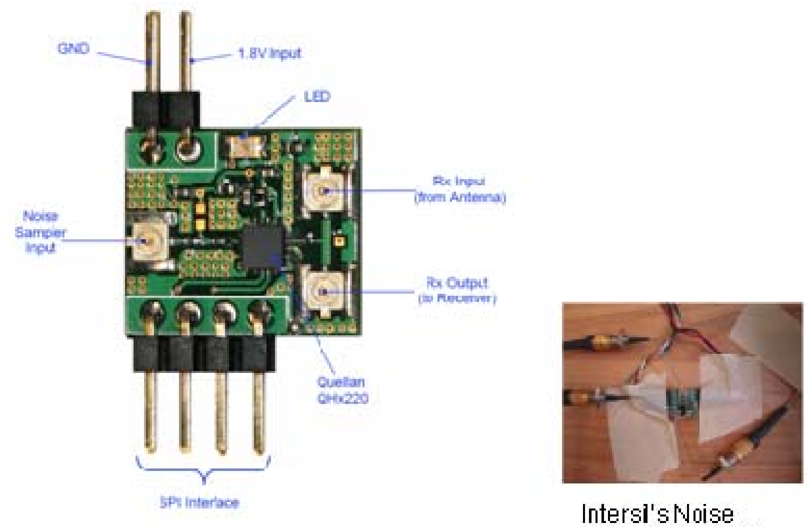

Intersil's Noise Cancellation Chip QHX-220

Figure 1: Intersil's Noise Cancellation Chip QHx220

This system works on a principle similar to Bose and Sony noise cancelling headsets, but at a ~1,000,000 times higher frequency. It requires an I and Q setting obtained from the DTV receiver's channel setting and the system's link quality parameter (BER, SNR, RSSI, etc.). This is done on either a micro-processor or in the baseband processor, which run a set of small algorithms. As a result the noise minimum is centered in the desired TV channel and signals - formerly buried in the noise floor - are being restored (Figure 4, e.g. at 549MHz).

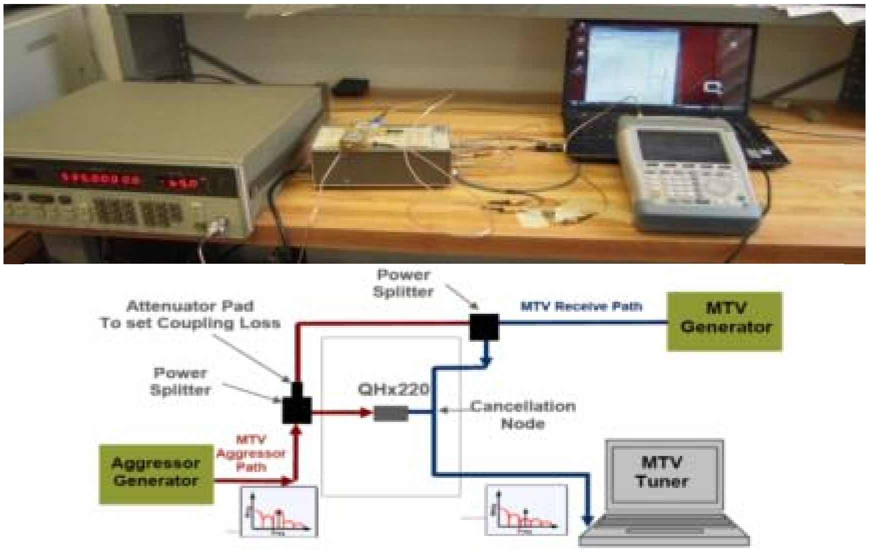
Power Splitter
Attenuator Pad
To set Coupling Loss
MTV Receive Path
MTV Generator
Power Splitter
QHx220
Cancellation Node
MTV Aggressor Path
Aggressor Generator
MTV Tuner

Figure 2 Noise Cancellation Test and Evaluation Setup

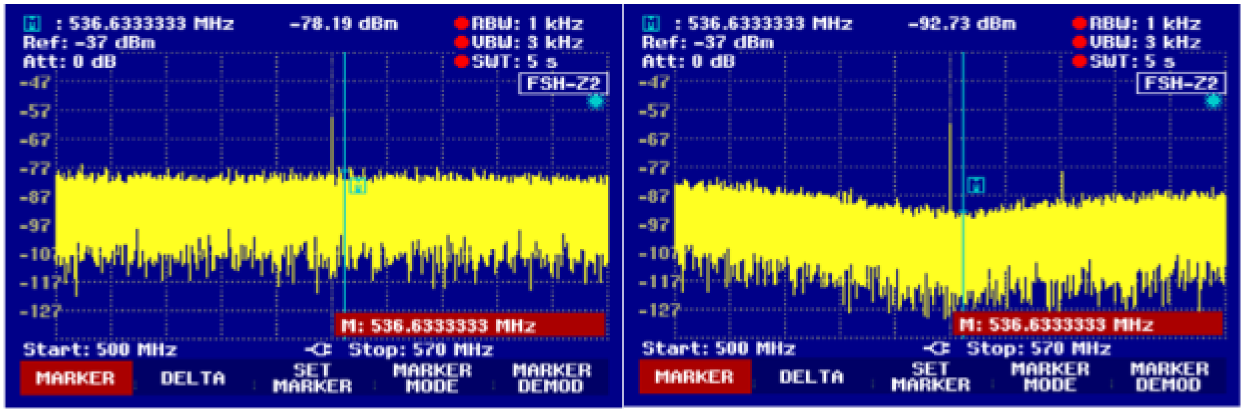

Figures 3-4: Measured results before and after noise cancelation

References

[1] Intersil Corporation, 1001 Murphy Ranch Rd, Milpitas, Ca 95035, www.Intersil.com

[2] QHx120 (development) works from 75MHz up to 225MHz.

2.5 Automatic Antenna Matching Systems

2.5.1 Summary

From an impedance matching point of view, the ideal antenna has a perfectly flat, purely resistive input impedance across its entire operating frequency range. The value of that impedance should be 70 Ω because this is the nominal industry standard characteristic impedances of the coaxial cable used in television receivers. Of course, no antenna is perfect. Quite to the contrary, most antennas' impedance variations with frequency are usually quite dramatic. This is particularly true for electrically small antennas, which tend to exhibit low radiation resistances and very high reactances in a narrow frequency range. Set-top DTV television antennas are not necessarily electrically small in the high VHF and UHF bands, but are at low VHF. Due to their required bandwidth (54-698 MHz) they invariably exhibit wide impedance fluctuations.

An impedance mismatch leads to losses and reflected energy that is not transferred to the receiver. These inefficiencies can be mitigated to a great degree by precise impedance matching that is provided by an "antenna tuner" (AT) connected between the antenna output terminals and the receiver input port. The AT may comprise a network of discrete components, or it may be a distributed, continuously varying structure (a tapered transmission line, for example), or some combination of both.

Lumped element matching networks go back to the earliest days of radio, and they often require manual adjustment of the matching elements (usually variable capacitors or roller inductors). In the early 1960s the military developed "ALE" (automatic link establishment) systems that employed remotely controlled automatic antenna tuners. These networks are usually located at the antenna feed and remotely tuned automatically based on VSWR (voltage standing wave ratio) measurements made at the transmitter. Essentially the same approach is taken with the built-in automatic ATs that are common in modern transmitters (most amateur radios contain built-in auto ATs, for example).

Modern low-power ATs use relays to switch lumped elements (inductors, capacitors) in and out of the matching network until the desired match is achieved. This technology is highly developed and readily available for use in set-top TV receive antennas. It therefore is not reviewed in this report. Instead, the emphasis here is on new, developing technologies that may be useful for antenna impedance matching. Four broad categories have been identified.

(a) *Voltage controlled reactive elements* are capacitors and inductors whose values are controlled by an applied voltage. Instead of switching discrete components in and out of a matching network using relays, the same operation is accomplished by varying the voltage applied to the reactive element. This type of non-mechanical system provides faster response time and more continuous control in a less noisy and probably smaller space than a relay-based AT.

(b) *Integrated ATs* are "antenna tuners on a chip," a complete device that is fabricated on a single chip or in a very small module with only a few external components. Integrated ATs have received considerable attention for cellular applications, and they work quite well. Chip-level devices successfully match up to 10:1 VSWRs in the highly fluctuating cellular antenna environment with rapid response times and high efficiency. These devices are constantly improving, and it is likely that existing technology can meet set-top AT requirements.

(c) *Tunable tracking filters* are already used across the VHF/UHF TV bands for wideband DTV (digital television) tuners, and they work quite well, providing very flat passbands, steep skirts, and harmonic suppression from the tuner's local oscillator. Extremely small devices have been designed and computer-simulated, and some measured data are available confirming achievement of design objectives. This is an evolving technology that may be transferrable to set-top AT requirements.

(d) *Software defined radio (SDR)* is a new concept in wireless communications that seeks to move system properties from the hardware layer to a high-level software layer by developing hardware modules that are fully software-controllable and configurable. SDR radios are gaining wide acceptance in many applications, and it is a developing technology. SDR "smart antenna" technology, which may be applicable to the set-top environment, consequently is a candidate technology for a television AT.

2.5.1.1 **Voltage Controlled Reactive Elements**

(a) *BZN (pryrochlore bismuth zinc niobate)* is a non-ferroelectric dielectric material that shows great promise for a new class of voltage tunable thin film capacitors [1]. When it was introduced in 2006, BZN exhibited the lowest loss of any room temperature complex oxide film with a very high dielectric constant and high tunability of (loss tangent ~0.0005, dielectric constant ~180, ~55% tuning range @ 1 MHz). Losses increase in the microwave range, but are controllable by modifying the electrode structure.

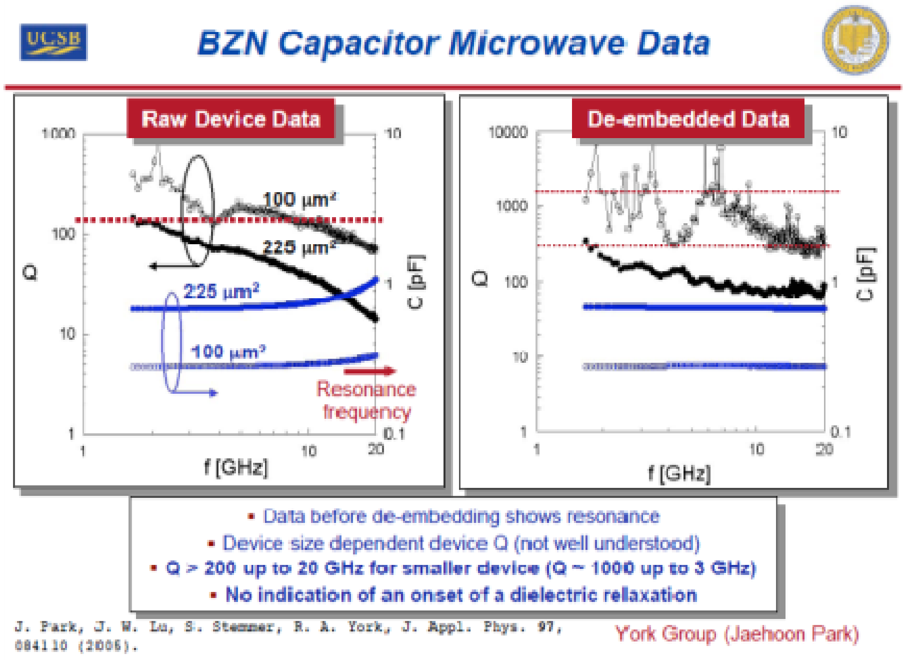

BZN Capacitor Microwave Data

- Data before de-embedding shows resonance
- Device size dependent device Q (not well understood)
- **Q > 200 up to 20 GHz for smaller device (Q ~ 1000 up to 3 GHz)**
- **No indication of an onset of a dielectric relaxation**

J. Park, J. W. Lu, S. Stemmer, R. A. York, J. Appl. Phys. 97, 084110 (2005).

York Group (Jaehoon Park)

Figure 1. BZN capacitor high-frequency measurements (reproduced from [1])

Measured data for different size BZN thin film capacitors at higher frequencies appear in Figure 1. For small devices (100 µm^2) the *Q*-factor exceeds 200 up to 20 GHz (de-embedded data, right plot). Through about 3 GHz *Q* ~ 1000. In addition, self-resonance in the BZN thin film structure is well above 20 GHz. By comparison, the best state-of-the-art thin-film BST (barium strontium titanate) devices in 2006 exhibited *Q*s that decreased monotonically from about 100 at 1 GHz to around 20 at 20 GHz. BZN is a far superior tunable dielectric. BZN and similar materials (see below) hold considerable promise for the development of completely solid-state ATs that can be used to tune set-top TV broadcast band antennas.

(b) *PZN (lead zinc niobate)* thin film interdigital capacitors (IDC) have been fabricated to increase tunability and reduce bias voltage compared to BZN IDC implementations. Cubic pyrochlore PZN thin film dielectrics provide superior performance through the microwave range. Figure 2 shows a typical configuration. The IDC is fabricated on a silicon substrate in the usual coplanar waveguide (CPW) configuration. Figure 3 plots the measured low-frequency dielectric constant (blue) which increases from just over 200 to about 230 at 10 MHz. The tenability as a function of applied bias voltage decreases from unity to about 0.75 with an applied electric field of 2 mV/cm (bias voltage of 5.5 volts). At 1 GHz the device *Q* falls to about 10 at all bias voltages, but the tunability remains at 26% (compared to 25% at 1 MHz) for 5.5 volts applied bias. The PZN tunable capacitor thus is a strong candidate for voltage-tunable set-top television antenna ATs, and its performance is comparable or better than that of BZN devices.

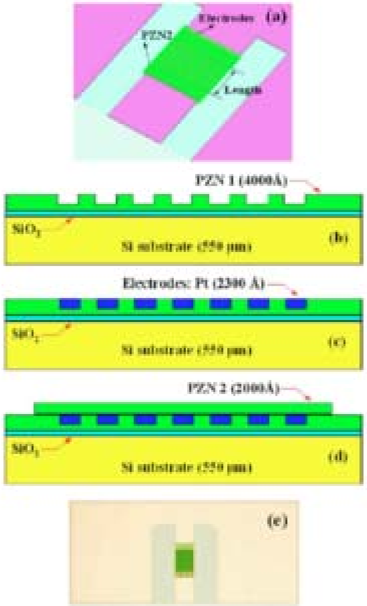

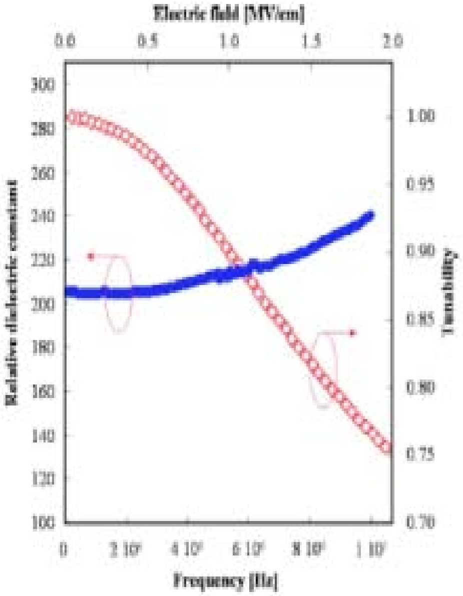

Figure 2. PZN IDC structure (reproduced from [2]). Figure 3. PZN IDC performance (reproduced from [2]).

(c) *DuNE™ capacitors* are a new patent-pending technology developed by Peregrine Semiconductor [3] for DTCs (Digitally Tunable Capacitors). The manufacturing process is based on Peregrine's proprietary UltraCMOS™ technology supplemented by its HaRP™ design methodology. UltraCMOS™ is a patented low-power SOI (silicon on insulator) variant chip architecture that reduces transistor capacitance resulting in increased switching speed. HaRP™ is a design technology that produces significant improvements in harmonic results, linearity and overall RF performance.

The company claims that its DTCs provide digitally controllable capacitances in the range 0.5 to 10 pF with either 3:1 or 6:1 tuning ratios and 32 addressable states (5 bits). *Q*s range from 40 to 80 between 1 and 2 GHz with device switching times below 5 μsec. Power handling capability is quite good (+38 dBm @ 50 Ω), while power consumption is low (100 μamps @ 2.4-3.5 volts DC). The manufacturer claims its DTCs provide better performance than currently available MEMS (*microelectromechanical system*) or BST devices, and they are readily available as off-the-shelf components.

The digitally tuned capacitor schematic diagram and its "flip-chip" package appear in Figure 4. The DTC acts as a series capacitor with two RF terminals. It is powered by a single DC line, and controlled by a three-wire serial interface. The control word is five bits long. The device is extremely small (1.26 mm x 0.81 mm) and thus well-suited for use in TV set-top antennas. Measured performance is shown in Figure 5. Capacitance variation is very linear with control state (32 increments), and the variation ratio is 3:1, from 1.15 pF to 3.4 pF. The DTC *Q*-factor at 900 MHz, a representative cellular frequency, varies from about 53 to about 72, which is quite good.

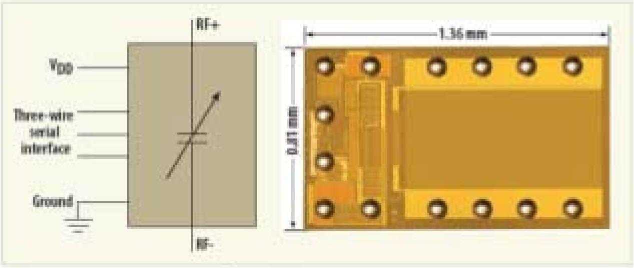

5. The schematic diagram shows the DuNE DTC (left) while the photograph (right) shows a flip-chip DuNE DTC prototype that measures just 1.36 x 0.81 mm.

Figure 4. Digitally Tuned Capacitor (reproduced from [4]).

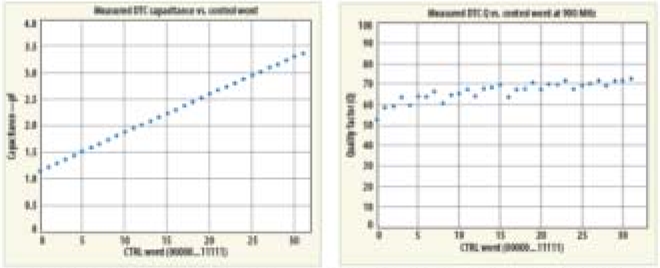

6. The measured performance of a 5-b DuNE DTC prototype shows linear tuning characteristic and a 3:1 capacitance range from 1.15 to 3.4 pF.

7. These measurements show the DuNE DTC prototype with quality factor (Q) designed to be in the range of 60 to 70 at 900 MHz.

Figure 5. DTC measured performance (reproduced from [4]).

DTCs have been effectively applied in cellular phone ATs. A typical configuration is shown in Figure 6. The transceiver, in this case a cellular handset, is connect to its antenna through an AT comprising four main elements: (1) serial interface; (2) digital mismatch sensor; (3) tuning algorithm; and (4) DTC core. This structure is representative of ATs based on other tuning technologies besides DTCs. For example, this same architecture would be used with MEMS-based or tunable thin film capacitor ATs. Note that the diagram does not show the RF path between the antenna and transceiver, only the AT control data path. The serial interface exchanges digital data with the transceiver on the AT status. The mismatch sensor responds to reflected power resulting from any impedance mismatch between the antenna and the system's characteristic impedance. Digital mismatch level data are processed by the tuning algorithm that determines which DTCs should be activated and at what capacitance level. Control words are sent to the DTCs in the matching network (DTC core) to effect the impedance match between the antenna and transceiver.

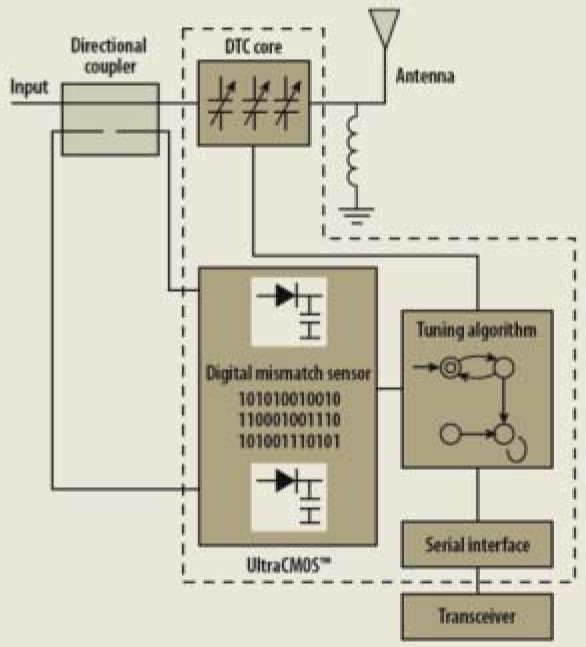
Directional coupler
Input
DTC core
Antenna
Digital mismatch sensor
101010010010
110001001110
101001110101
Tuning algorithm
UltraCMOS™
Serial interface
Transceiver

Figure 6. DTC-based AT (reproduced from [4]).

(d) *Dual-gap Tunable MEMS Capacitors* are made using a novel fabrication process that creates two gap device that has an extremely wide tuning range (as much as a factor of 15) with high *Q* [5]. Figure 1 shows the dual-gap structure schematically and as implemented using a "two hump" sacrificial layer which, when removed, creates two MEMS gaps instead of one. In the schematic, the two moveable electrodes are indicated by the double downward arrows. As these suspended elements move closer to the substrate electrodes the capacitance increases. The structure is a MEMS device designed for a linear capacitance variation with applied control voltage.

In the fabricated prototype, the central region (E_c) is 250µm x 80µm (LxW) with a 1.5µm fixed gap. The two "beams" are 800µm long by 80µm wide creating a variable gap because they are movable. The "actuation area" patches (E_a) are 200µm long and 80 um wide with a gap of 4.5µm with the beams not deflected. This structure produces a minimum capacitance of 0.12 pF that is voltage controllable to a maximum capacitance of 1.77 pF. The resulting tuning range is $\frac{C_{\max} - C_0}{C_0} = 1375\%$, which is quite remarkable. The required bias ("pull in") voltage is less than 12 volts. This device is not commercially available, but it represents a class of tunable devices that should be very useful in the set-top AT when (and if) they become COTS (commercial off-the-shelf) items.

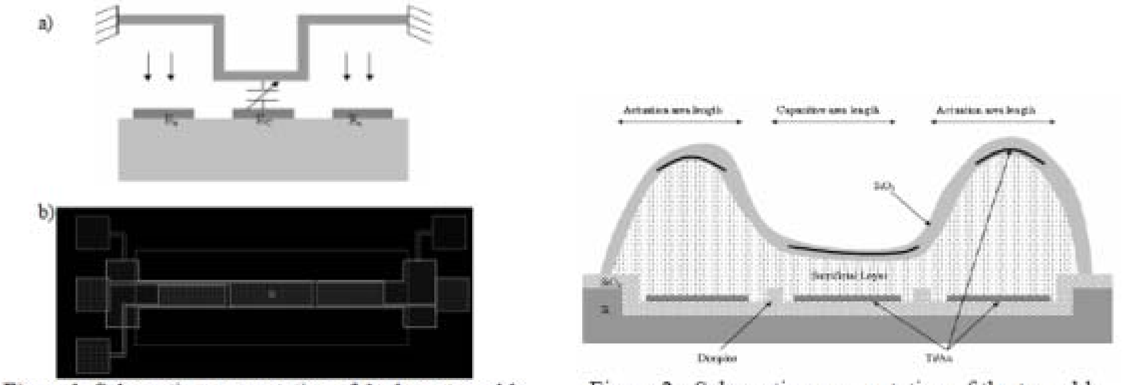
a)
b)
Actuation area length
Capacitive area length
Actuation area length
Sacrificial Layer

Figure 1: Schematic representation of dual gap tuneable capacitor - a) Cross section - b) Layout top view.

Figure 2 : Schematic representation of the tuneable capacitor based on dual pick profile of the sacrificial layer

Figure 7. Dual-gap tunable MEMS capacitor (reproduced from [5]).

(e) *MEMS floating dielectric capacitors* (MFDCs) are a recent promising development [6]. The movable dielectric is a new actuation principle in which a floating movable dielectric is electrostatically maneuvered to vary the capacitance. Figure 8 shows the new concept schematically. A mechanical spring returns the movable dielectric to its undisturbed position when no force is applied by the electrostatic comb drive. When the dive is activated, the dielectric moves closer to the top and bottom capacitor plates thereby increasing the capacitance. Preferentially the RF path is through the plates, not through the dielectric and spring, which increases losses.

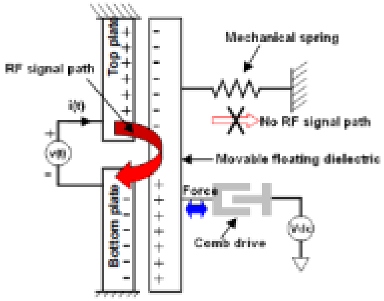
Top plate
Bottom plate
RF signal path
i(t)
v(t)
Mechanical spring
No RF signal path
Movable floating dielectric
Force
Comb drive
Vdc

Fig. 1: Conceptual schematic view of the proposed tunable capacitor with an electrically floating dielectric.

Figure 8. MFDC concept (reproduced from [6]).

Figure 9 shows schematic RF signal path superimposed on the MFDC's actual structure as fabricated. In the left pane the signal flows through the spring, which increases the capacitance at the expense of increased losses. In the right pane the signal flows between the top and bottom plates through the movable dielectric. This configuration results in lower losses, but lower capacitance as well. Of course, either configuration can be used, depending upon the specific application.

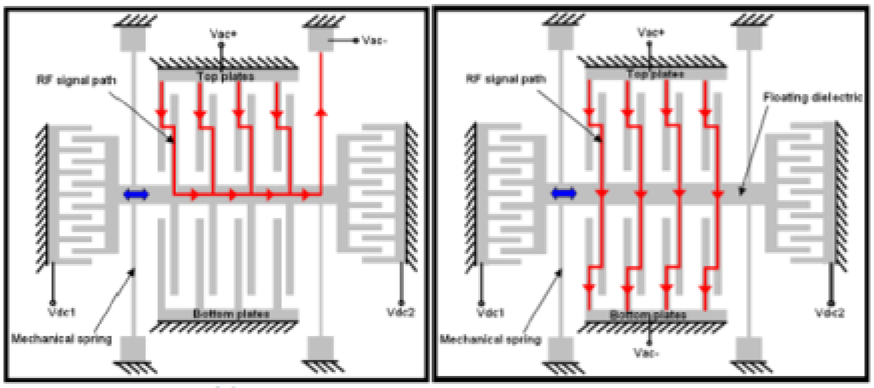
Vac+
Vac-
RF signal path
Top plates
Bottom plates
Vdc1
Vdc2
Mechanical spring
Floating dielectric

(a) (b)

Fig. 2: Schematic views of tunable capacitor design with RF signal through spring (a) and dielectric (b).

Figure 9. MFDC as fabricated and measured data (reproduced from [6]).

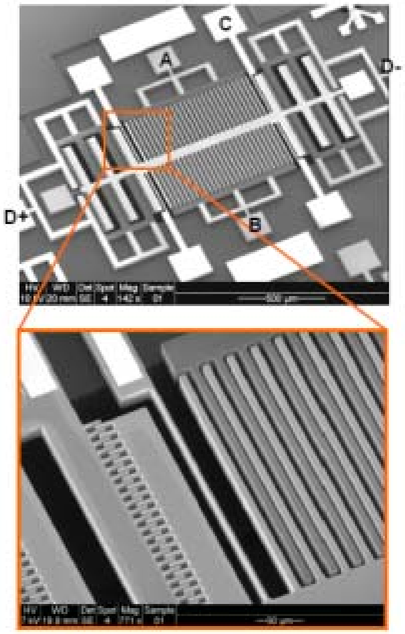

Fig. 4: SEM image of the micromachined tunable capacitor.

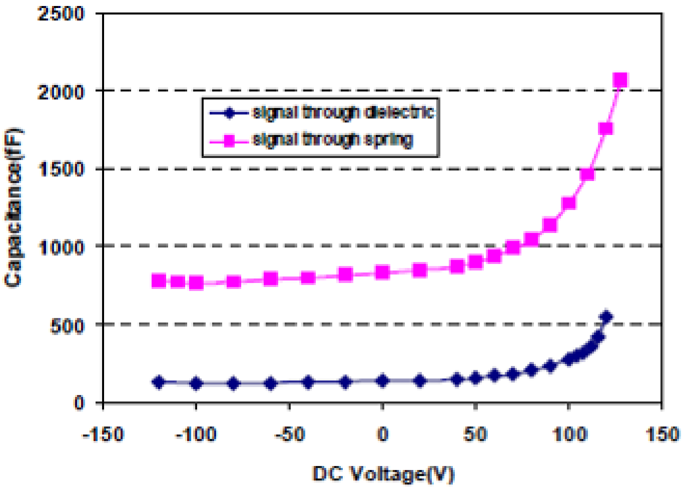

Fig. 5: Measured capacitor changes vs. applied dc bias

Figure 10. MFDC as fabricated and measured data (reproduced from [6]).

Figure 10 provides a scanning electron (SEM) microscope image of a fabricated MFDC (the area outlined in orange, lower pane) and measured performance data. Comparing Figs. 9 and 10, the features in the schematic are readily identifiable in the SEM photo. The actuator pads are labeled D+ and D- in the photo, while the spring is attached to pad C. Pads A and B are the connections for their associated comb plates.

The applied bias voltage ranged from -120 VDC to +120VDC, resulting in capacitance ranges of about 760 fF (femtoFarad) to 2100 fF for RF signals passing through the spring element (pink curve). The initial capacitance (no bias voltage) was 830 fF, leading to a tuning range of approximately 170%. The 1 GHz *Q*-factor was 0.35, which is low by comparison to other technologies. When the RF path is through the movable dielectric (blue curve), capacitance ranged from just over 100 fF to about 550 fF yielding a tuning range of approximately 370% with an initial capacitance of 135 fF. As expected, the reduced loss resulting from routing the signal away from the spring results is a much better *Q*-factor of 56 at 1 GHz. Overall, this performance is not as good as that provided by other technologies, but MFDCs are a new concept that requires further development. The MFDC approach certainly merits consideration as a potentially useful future technology for the set-top AT.

(f) The v*oltage controlled semiconductor inductor* (VCSI) is another reactive component that should be useful in set-top ATs. Inductor values are usually fixed, so that obtaining specific value of inductance in an AT is usually accomplished by mechanically switching small inductors in and out of the matching circuit using relays or MEMS RF switches. Recently disclosed VCSI devices [7] should be particularly useful at broadcast TV frequencies. The device addresses the problem of limited tuning range provided by voltage tunable capacitors by varying the inductance instead.

Generally is comprises regular coil turns of wire interconnected by semiconductor diodes that can connect individual turn to create a VCSI. Including a resistor and capacitor creates a complete tunable RLC circuit. Figure 11 illustrates this patented technology. The left pane shows a perspective view of the device, which includes conductive loops (205) connected to a semiconductor bar comprising P- and N-type regions at each end (208 and 210, respectively) connected by a depletion region (212). Applying a voltage across terminals L_A and L_B varies the length of the depletion region which acts as an insulator. Because its size is proportional to the applied bias voltage, individual coil turns are either connected or disconnected in proportion to the voltage, thereby creating a voltage-variable inductor.

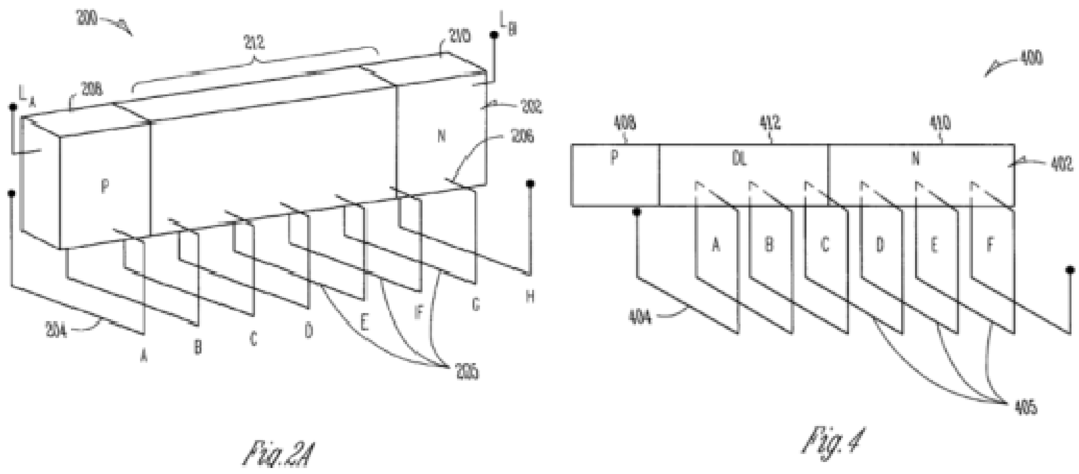

Figure 11. Voltage controlled semiconductor inductor (reproduced from [7]).

A helical-turn implementation of the VCSI is shown in the right pane of Figure 11. The same semiconductor diode structure comprising P- and N-type end regions connected to a central depletion layer (DL) whose length is voltage-variable. The figure provides a schematic representation of how the three regions' lengths vary with bias voltage compared to the "off" state shown in the left pane. In the diagram, coil turns C through F are electrically short-circuited because they are outside the insulating DL. In this case, only turns A and B are active elements in the inductor.

The VCSI may become an important element in set-top ATs because it provides a complete RLC tuning in a single chip-level device. A bank of VCSIs, for example, could comprise the switchable reactive elements in a set-top AT that are completely voltage-controllable, thereby eliminating the need for mechanical switching relays. VCSI therefore is an attractive emerging technology that merits watching.

2.5.1.2 Fully Integrated ATs

In addition to the emerging component-level technologies described in §1, fully integrated ATs have been developed that also merit consideration for set-top tuners. This section examines developments in that area.

(a) A *reconfigurable RF-MEMS-based matching network* is described in [8]. The chip-level device's circuit diagram is shown in Figure 12. It comprises two stages, the first of which is a Pi-match section with four shunt capacitor-series inductor (CL) sections. A total of eight RF-MEMS switches are employed yielding 2^8 impedances. The shunt capacitors are formed from bi-valued MEMS varactors. The variable capacitor in series with each fixed inductor, also a MEMS varactor, has the effect of adjusting the series inductance. The second section is a phase-shifter comprising a 3-dB 90-degree coupler connected to a reflective load. MEMS varactors adjust the load reflection coefficient to control the overall phase shift. A total of 2^3 phase rotations are possible in the phase shifter, resulting in as much as 340 degrees of total phase shift that can be applied to the impedances at the output of the Pi-match. The network is designed to work at a 50 Ω impedance level.

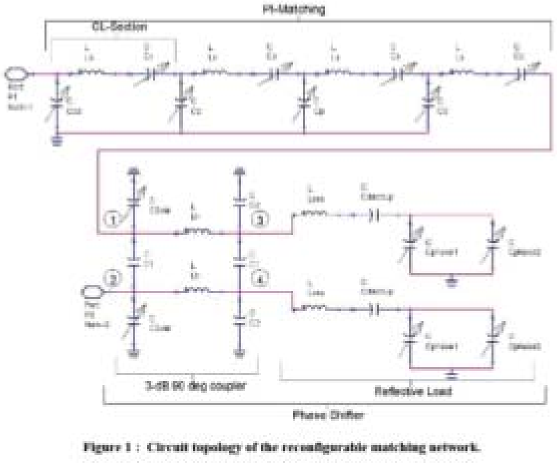

Figure 12. Reconfigurable RF-MEMS-based matching network circuit (reproduced from [8]).

The chip layout and fabricated device are shown in Figure 13 with the major sections being labeled on the layout diagram. The chip area is extremely small, only 40 mm^2 (slightly larger than 3x12 mm). The published report indicates that the chip has been fabricated and was undergoing testing, but no actual measured data were reported. Instead simulated performance at a single frequency (620 MHz) was calculated at each of the possible 2,048 impedance combinations and plotted on a Smith chart as shown in Figure 14. The circles are impedance values computed for the first stage alone, while the dots represent the show the performance of the complete reconfigurable matching circuit. The fairly uniform distribution of dots throughout the Smith chart suggest that the matching circuit will effectively match an extremely wide range of impedances to 50 Ω. This technology is directly applicable to the set-top AT and appears to be on the verge of realization. Future published results for reconfigurable RF-MEMS-based matching networks clearly bear watching.

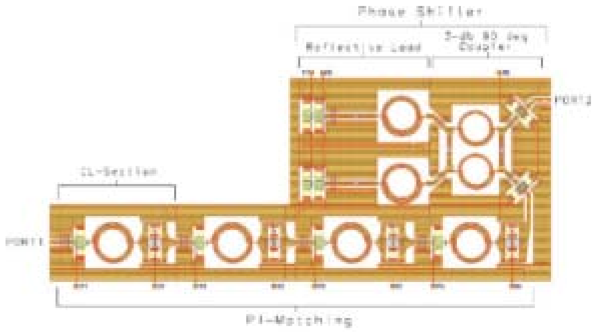

Figure 3 : Layout of the circuit as implemented in ADS

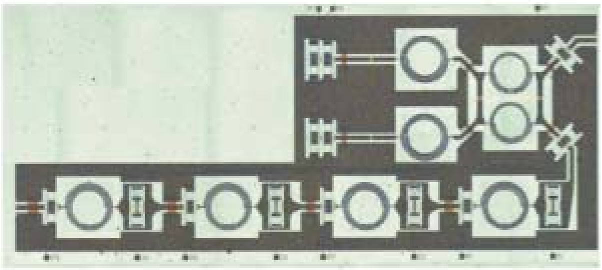

Figure 4 : Photograph of the fabricated circuit

Figure 13. Reconfigurable RF-MEMS-based matching network layout & fabrication (reproduced from [8]).

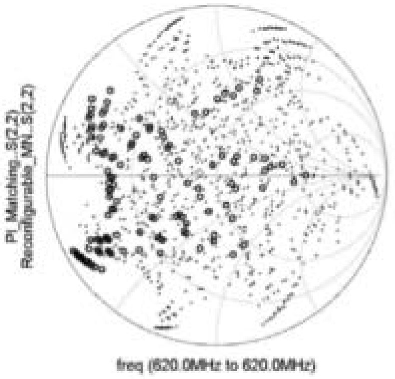

Figure 2 : Simulated impedance values (circles: first stage; dots: full reconfigurable matching network).

Figure 14. Reconfigurable RF-MEMS-based matching network performance (reproduced from [8]).

(b) The *AT using only RF signal amplitudes* described in [9] may be useful for the set-top application because it specifically addresses the issue of highly variable, unpredictable and uncontrollable environments. This new technology was developed for mobile applications such as cellular transceivers because their fluctuating environment often causes VSWR spikes approaching 10:1 in nominal 50 Ω systems. These fluctuations are frequently transient on time scales in the millisecond range. The new AT design accommodates this environment, and it may be useful in the highly variable TV set-top antenna environment as well.

The architecture appears in Figure 15, which includes schematic Smith chart representations of the impedance at various points. ATs generally comprise an impedance sensor, a tunable matching network, and control circuitry that changes the network parameters in order to achieve an acceptable VSWR (usually relative to 50 Ω with VSWR < 2-3:1). The general approach shown in Figure 15 therefore is applicable to any matching network. The new matching concept involves a two-step process: (1) the reactance is essentially tuned out using a series or shunt reactance; and (2) a tunable "transformer" changes the remaining resistive component to the desired value of impedance.

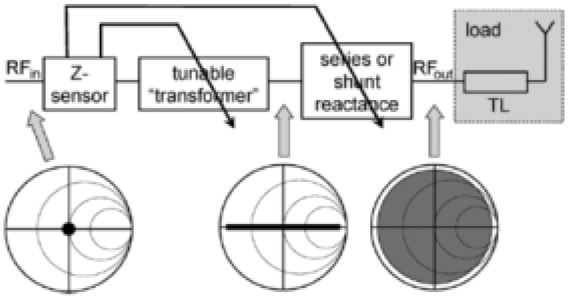

Fig. 2. Basic concept of the automatic antenna tuning system: tune out the imaginary part and tune the real part separately.

Figure 15. New AT architecture for fluctuating environment (reproduced from [9]).

In the new AT, the antenna impedance sensor makes a quasi-DC log-peak measurement of the RF signal amplitude at three points in the signal chain (V1, V2, V3 in Figure 16). Details of the log-peak detector circuit are shown in Figure 17. These data combined with the known transfer functions of the reactive elements (jX_{ext}) permit a calculation of the impedance's imaginary part. The real part of the impedance is determined from the signal level $V_{\Delta R}$ and an offset voltage that brings this level to a target value corresponding to 1:1 VSWR. The tunable transformer comprises a T-network lumped transmission line made of two adjustable inductors and an adjustable capacitor. Any appropriate device can be used to implement these elements.

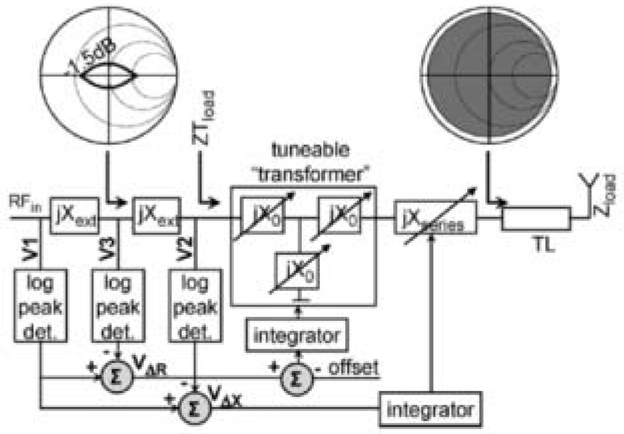
-1.5dB
ZT_load
tuneable
"transformer"
RF_in
jX_ext
jX_ext
jX_0
jX_0
jX_series
TL
Z_load
V1
V3
V2
jX_0
log peak det.
log peak det.
log peak det.
integrator
V_ΔR
- offset
V_ΔX
integrator

Fig. 3. Automatic antenna tuning system with low-frequency impedance detection; the target impedance (− 1.5-dB contour) is used in Section IV-A.

Figure 16. AT functional block diagram (reproduced from [9]).

Two demonstration versions of this new AT design was built and tested at 900 MHz. Measurements confirmed that they could reliably and quickly automatically tune antennas with $Z_{in} = [5-500] + j[15.8-158]\,\Omega$ (10:1 VSWR). Although the sizes were not reported, these demonstration units were intended for use in cellular handsets, so that the size is certainly consistent with use in a set-top television AT.

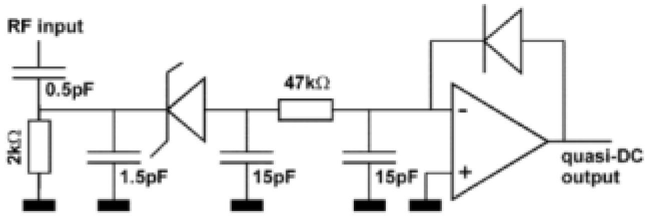
RF input
0.5pF
2kΩ
1.5pF
47kΩ
15pF
15pF
-
+
quasi-DC
output

Fig. 4. Schematic diagram of the log-peak detector implementation.

Figure 17. Details of log-peak detector circuit in Figure 16 (reproduced from [9]).

2.5.1.3 On-Chip Tracking Filters

While tracking filters are not AT's, their technology likely is applicable to antenna tuners ATs, and consequently should be monitored for application in set-top devices. This section describes three different types of tracking filter.

A device developed specifically for DTV tuners is described in [10]. A new architecture is proposed that includes two separate filters, a harmonic rejection tracking filter (HRTF) and an RF tracking filter (RFTF) that are controlled by complementary digital switches. The new architecture is shown in Figure 18. The entire device is fabricated as a single integrated circuit (IC) chip. The filter stage is shown in blue located between the low noise amplifier (LNA) connected to the antenna and the DTV tuner.

Unlike narrowband RF receivers, wideband DTV tuners are prone to interference from local oscillator odd harmonics mixing with signals at the lower television channels (48-287 MHz). This problem is addressed by the HRTF that provides a high-order band-pass tracking filter with 3^{rd}-order harmonic rejection greater than 60 dB. On the UHF television channels the RFTF tunes 287-860 MHz with narrow band response (20 MHz at -3 dB). This stage is implemented using a cascade of tunable 2^{nd}-order bandpass tracking filters.

Computer simulation of the new architecture's performance predict 48-860 MHz operation with tunable adjustable bandwidth of 8-20 MHz, 5-15 dB N+2 channel rejection (16 MHz offset), and 3^{rd}-order harmonic rejection of 60 dB from the HRTF. The RFTF stage provides 4.2 dB of N+2 channel rejection above 287 MHz. This device can be fabricated on-chip using 0.13 µm CMOS technology with 1.2 VDC supply voltage and 19.8 mA current draw (total of 24 mW power consumption).

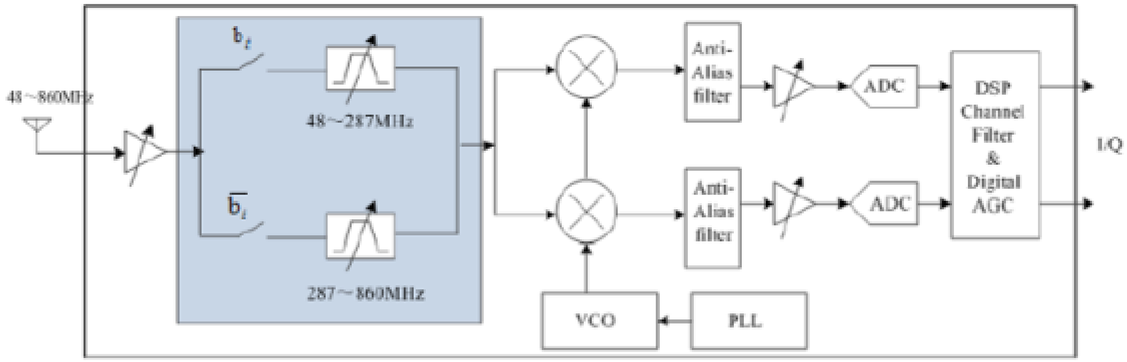

Figure 1. Direct Conversion TV tuner architecture

Figure 18. New DTV tracking filter architecture (reproduced from [10]).

Another potentially important emerging filter technology is the use of low-temperature co-fired ceramic (LTCC) technology to fabricate fully integrated multi-layer tunable filters for RF and microwave use. A typical device is shown in Figure 18. Passive

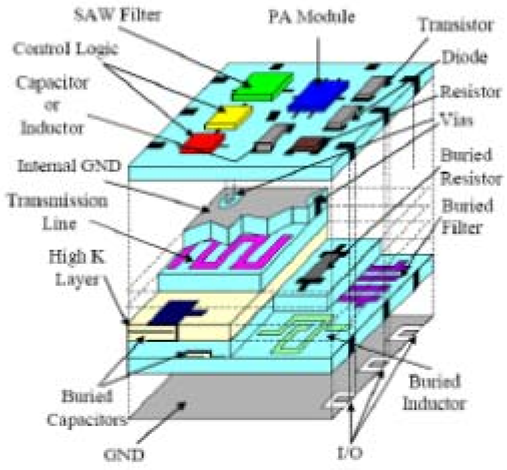
SAW Filter
PA Module
Transistor
Control Logic
Diode
Capacitor or Inductor
Resistor
Vias
Internal GND
Buried Resistor
Transmission Line
Buried Filter
High K Layer
Buried Capacitors
Buried Inductor
GND
I/O

Figure 1. A typical LTCC structure

Figure 18. LTCC filter structure (reproduced from [11]).

Elements (resistors, inductors, capacitors) are integrated on the surface or embedded in a multilayer substrate. Various configurations have been demonstrated, including bandpass filters (380 MHz-2.4 GHz), 3-stage Butterworth bandpass filters (1.2 GHz), and an electronically tunable microstrip combline filter. These prototype devices point to LTCC technology's utility for RF and microwave applications. At this point the new multilayer architecture proposed in [11] that involves switching between layers for tuning is being computer-modeled, but working devices based on that approach have not been fabricated. LTCC technology may be very attractive in the television frequency range because of its potential for very high levels of integration resulting from the multilayer design.

A third example of an on-chip tracking filter is provided by [12]. A complete tunable structure was fabricated and tested. This chip occupied an area of only 2.8 mm^2 (fabricated with 0.18 µm CMOS) requiring 34-120 mA at 1.8 VDC. A photomicrograph of the chip appears in Figure 19, and its architecture in Figure 20. The device comprises cascaded RLC sections as shown in Figure 20, each containing a digitally programmable on-chip capacitor and resistor, and an off-chip fixed inductor. The resistor is adjusted with 8-bit resolution, while the capacitor uses a 10-bit control signal.

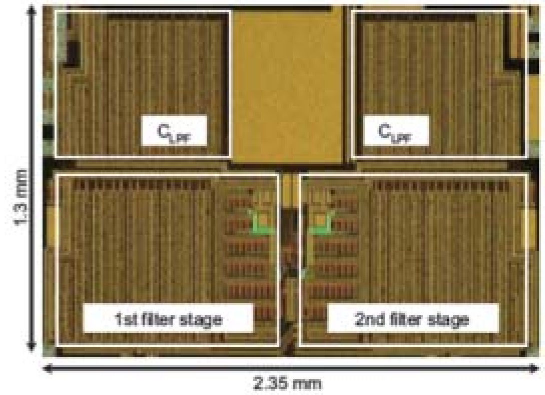
C_LPF
C_LPF
1.3 mm
1st filter stage
2nd filter stage
2.35 mm

The chip area excluding bonding pads is 2.8 mm².

Fig. 7. Chip micrograph

Figure 19. Tunable LC-Tracking filter as fabricated (reproduced from [12]).

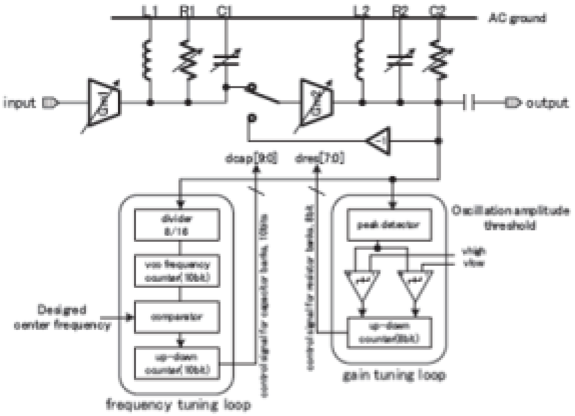
L1
R1
C1
L2
R2
C2
AC ground
input
output
dcap[9:0]
dres[7:0]
divider 8/16
vco frequency counter(10bit)
comparator
up-down counter(10bit)
Desired center frequency
frequency tuning loop
peak detector
Oscillation amplitude threshold
vhigh
vlow
up-down counter(8bit)
gain tuning loop

Fig. 1. RF tracking filter architecture

Figure 20. Tunable LC-Tracking filter architecture (reproduced from [12]).

Figure 21 shows the tracking filter's measured performance data from 125 MHz to 1.06 GHz. Its response in the 5.6 MHz passband is very flat with a ripple less than 0.2 dB. The noise figure in this device appears to be somewhat high (16.8-19.5 dB), but the third-order intercept points are good (~128/167 dBμV, in/out of band). Frequency selectivity is good at > 36 dB, and the power consumption quite low (<~0.2 W maximum). This example shows that very effective single-chip RF tracking filters with minimal off-chip components (in this case two inductors) can be designed and fabricated for set-top use using currently available technologies.

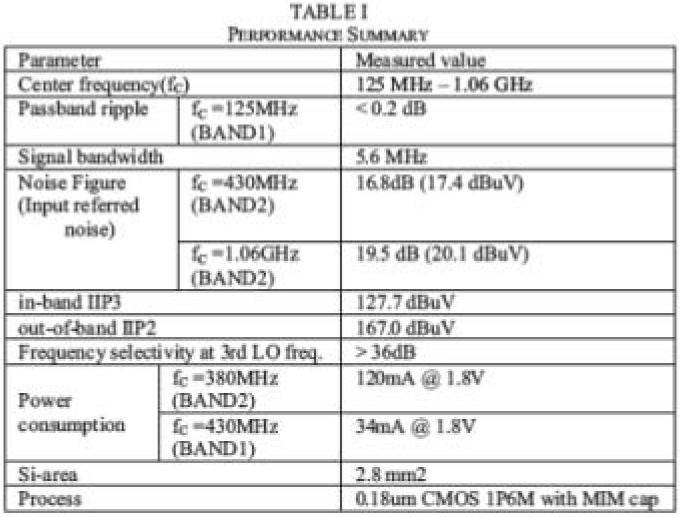

TABLE I
PERFORMANCE SUMMARY

Parameter		Measured value
Center frequency(f_C)		125 MHz – 1.06 GHz
Passband ripple	f_C =125MHz (BAND1)	< 0.2 dB
Signal bandwidth		5.6 MHz
Noise Figure (Input referred noise)	f_C =430MHz (BAND2)	16.8dB (17.4 dBuV)
	f_C =1.06GHz (BAND2)	19.5 dB (20.1 dBuV)
in-band IIP3		127.7 dBuV
out-of-band IIP2		167.0 dBuV
Frequency selectivity at 3rd LO freq.		> 36dB
Power consumption	f_C =380MHz (BAND2)	120mA @ 1.8V
	f_C =430MHz (BAND1)	34mA @ 1.8V
Si-area		2.8 mm2
Process		0.18um CMOS 1P6M with MIM cap

Figure 21. Measured performance of tunable LC-Tracking filter (reproduced from [12]).

2.5.1.4 Software Defined Radios

A software defined radio (SDR) is an element of a wireless communication network whose operational modes and parameters can be changed or augmented post-manufacture via software. The essential idea is that a flexible hardware layer exists whose function can be controlled and modified entirely by a computer program, as opposed to requiring hardware modifications of any kind. The SDR concept spans many radio network technologies including cellular systems, personal communications services (PCS), 3rd and 4th generation wireless (3G and 4G), mobile data, emergency services, paging, messaging, and military/government communications, and any future modifications to these existing services or entirely new ones. The FCC (Federal Communications Commission) definition is more restrictive in that it applies only to the transmitter side of an SDR. But, as a practical matter, the SDR concept applies to any wireless device whose characteristics are software-controllable, whether it be the transmitter, receiver, both, or some other element such as a modem.

SDR technology is relevant to the television set-top AT because there is a developing standard that specifically addresses the issue of “smart antennas” (SAs) in the context of SDR. This type of antenna and its associated AT may be useful for the set-top application and consequently should be monitored as an emerging technology. The high-level SDR smart antenna architecture appears in Figure 22. The hardware layer comprises M transmit antennas and N receive antennas because SDR in general supports two-way communication (in the step-top TV receive application, of course, there are no transmit antennas). Each antenna is has a separate RF/IF processing chain with the smart antenna signal processing (“waveform application”) being applied to the baseband signal. Device drivers in the middleware layer control various programmable hardware devices, such as ASSPs (application-specific standard processors), FPGAs (field-programmable gate arrays), DSPs (digital signal processors), and GPPs (general purpose processors).

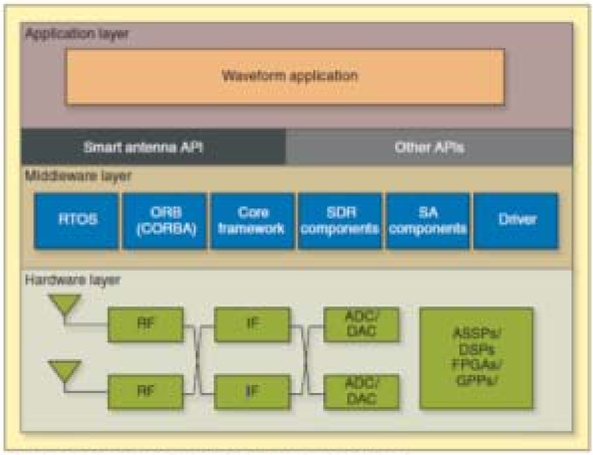
Application layer
Waveform application
Smart antenna API
Other APIs
Middleware layer
RTOS
ORB (CORBA)
Core framework
SDR components
SA components
Driver
Hardware layer
RF
IF
ADC/ DAC
RF
IF
ADC/ DAC
ASSPs/ DSPs FPGAs/ GPPs/

Figure 1. Example of an SDR-based smart antenna open architecture.

Figure 22. SDR smart antenna high-level architecture (reproduced from [13]).

Figure 23 illustrates a typical deployment of an SA API (application program interface). The SA control device (for example, an antenna tuner, tracking filter, MEMS-based controller, fluidic element controller, and so on) is operated by a GPP controlled by a CORBA interface (common object request broker architecture). The high-level SA algorithm controls appropriate CORBA-compatible drivers for DSP baseband processing, antenna control, and other functions such as synchronization and other SDR devices that may be controlled by the SA algorithm. The basic concept in this structure is that any hardware module involved in controlling the SA or processing its signal is controllable by the API with variable parameters.

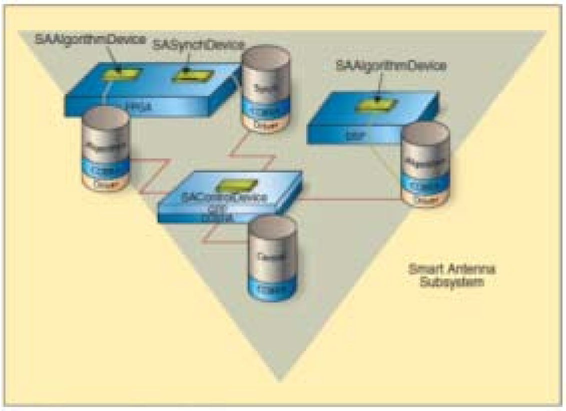
SAAlgorithmDevice
SASynchDevice
FPGA
Synch
SAAlgorithmDevice
DSP
SAControlDevice
Control
Smart Antenna Subsystem

Figure 6. Deployment diagram of SA API.

Figure 23. SDR Smart Antenna API deployment (reproduced from [13]).

A typical SA hardware implementation for the SDR baseband signal processing module is shown in Figure 24. The two DSP integrated circuits (ICs), the GPP, and the FPGA ICs are labeled. This type of module is simply inserted into a backplane containing the other SDR modules to create a complete SDR. All chip-level functions are fully controllable by the SA API. As modifications are required, a simple download of the updated API is all that is necessary. No hardware modifications or swaps are necessary. The SDR approach thus provides exceptional flexibility in customizing radio performance and it well may be a very useful approach to developing effective set-top television ATs.

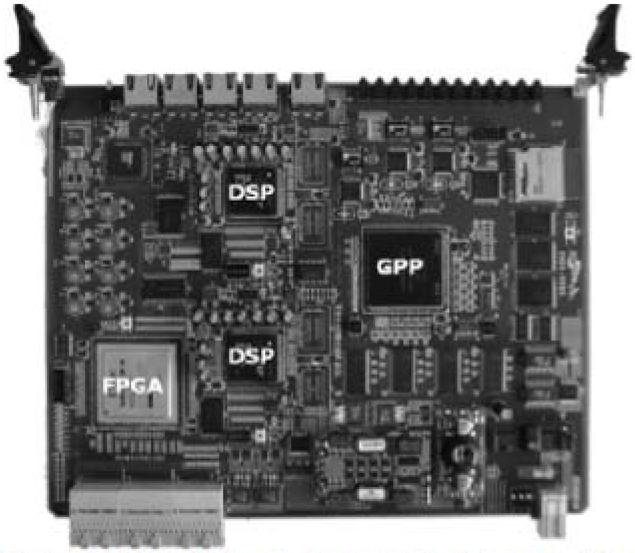

Figure 6: Photograph of implemented Baseband Signal Processing part of SDR smart antenna system.

Figure 24. Typical SDR smart antenna hardware implementation (reproduced from [14]).

2.5.2 References

[1] Stemmer, S., "Structure-Property Relationships of Tunable Thin Film Dielectrics for Microwave Applications," *Ferroelectrics UK Conference*, Birmingham, UK, May 23-26, 2006.

[2] Lee, Y. C., Hong, Y P., and Ko, K. H., "Low-voltage and High-tunability Interdigital Capacitors Employing Lead Zinc Niobate Thin Films," *Applied Physics Letters*, **90**, 182908 (2007).

[3] Peregrine Semiconductor Corp., 9380 Carroll Park Drive, San Diego, CA 92121 USA (www.psemi.com).

[4] Ranta, T., and Novak, R., "Antenna Tuning Approach Aids Cellular Handsets," *Microwaves & RF*, p. 82, November, 2008.

[5] Soulimane, S., Casset, F., Chapuis, F., Charvet, P. L., and Aid, M., "Tuneable Capacitor based on Dual Picks Profile of the Sacrificial Layer," *DTIP of MEMS & MOEMS*, Stresa, Italy, 25-27 April 2007 (EDA Publishing/DTIP 2007, ISBN: 978-2-35500-000-3).

[6] Zhu, Y., Yuce, M. R., Moheimani, S. O. R., "A Low-Loss MEMS Tunable Capacitor with Movable Dielectric,**"** *IEEE Conference on Sensors* (*IEEE SENSORS 2009),* pp. 651-654, October, 2009.

[7] Subramanian, K. M., "Voltage-Controlled Semiconductor Inductor and Method," U.S. Patent No. US 7,511,356 B2, Mar. 31, 2009.

[8] Bedani, M., Carozza, F., Gaddi, R., Gnudi, A., Margesin, B., Giacomozzi, F., "A Reconfigurable Impedance Matching Network Employing RF-MEMS Switches," *DTIP of MEMS & MOEMS*, Stresa, Italy, 25-27 April 2007 (EDA Publishing/DTIP 2007, ISBN: 978-2-35500-000-3).

[9] Firrao, E. L., Annema, A-J., Nauta, B., "An Automatic Antenna Tuning System Using only RF Signal Amplitudes," *IEEE Trans. Circuits and Systems – II: Express Briefs*, 55, No. 9, pp. 833-837, Sept. 2008.

[10] Sun, Y., Lee, J-s., Lee, S-g., "On-chip Active RF Tracking Filter with 60dB 3rd-Order Harmonic Rejection for Digital TV Tuners," *SoC Design Conference, 2008. ISOCC '08*, 24-25 Nov. 2008, Busan, vol. 01, pp I_406-I_409.
DOI: 10.1109/SOCDC.2008.4815658

[11] Suma, M. N., and Suhas, K., "Performance Analysis and Process Parameters of Novel LTCC Filters," *International Journal of Recent Trends in Engineering*, **1**, No. 3, p. 346, May 2009.

[12] Kanmazawa, Y., Fujimoto, Y., Izuka, K., "A 130M to 1GHz Digitally Tunable RF LC-Tracking Filter for CMOD RF Receivers," IEEE Asian Solid-State Circuits Conference, Fukuoka, Japan, pp., 469-472, November 3-5, 2008 (978-1-4244-2605-8/08/$25.00 ©2008 IEEE)

[13] Hyun, S., Kim, J., Choi, S., Pucker, L., Fette, B., "Standardizing Smart Antenna API for SDR Networks," *RF Design*, p. 34, Sept. 2007.

[14] Hyeon, S., kim, J., Choi, S., "Smart Antenna APIS: From Concept to Practice," *Proc. SDR 07 Technical Conference and Product Exposition*, SDR Forum, http://www.wirelessinnovation.org

2.6 Physically Reconfigurable Antenna Elements

2.6.1 Summary

A *reconfigurable antenna* (RA) is an antenna whose physical and/or electrical properties can be changed in real time in order to achieve certain performance characteristics. The panoply of possible RA implementations makes RAs especially attractive for the television set-top antenna application. This section discusses some of the "typical" RA implementations, as well as some rather unusual technologies that may point the way to future developments that could be useful for set-top devices.

The essential concept underlying RAs is that some antenna property is changed "on the fly" to accomplish some objective, say tuning. An example is the tunable antenna in which the electrical length of a radiating element is changed by switching in and out reactive elements (capacitors/inductors) or remotely moving the tap on a roller-type tuning inductor. This type of reconfigurable antenna has been used for at least fifty years (see, for example, [1] p. 20.46 *et seq*.).

While an RA may be any electrical size, typically they are a substantial fraction of a wavelength, so that they are not (necessarily) electrically small antennas (ESAs). The problems associated with ESAs, narrow bandwidth, low radiation efficiency, highly reactive input impedance, are mitigated in RAs that are a substantial fraction of a wavelength in size. Nevertheless, the RA concept is applicable to ESAs as well, and it may be useful for ESA candidates for the set-top application. As a general proposition, RA technology is applicable across the entire range of antenna electrical sizes, which makes it particularly attractive for set-top devices.

Most recent examples of effective RAs are at frequencies well above the broadcast television channels, typically in the microwave region. However, because these RAs often are comparable to the operating wavelength in size, they may be readily used at lower frequencies, especially at UHF TV channel frequencies. On the lower VHF channels, RA approaches may be effective when applied to standard antennas such as bowties or spirals whose size is consistent wit set-top requirements.

Five candidate RA technologies hold promise for short or long term application to television set-top antennas:

(a) *Microelectromechanical System* (MEMS) RAs are antennas based on MEMS RF switches. These devices, which have been fabricated for well over a decade, have come to the foreground as the preferred approach for RAs. They can be used in all frequency ranges, and, in fact, perform better at lower frequencies, although current applications emphasize the high UHF and microwave ranges. Other similar switching devices such as PIN diodes or GaAs (gallium arsenide) solid-state switches also are used to build RAs, but these technologies are older and perform less well. The latest developments in MEMS, nano-MEMS and the use of carbon nano-tubes (CNT), reduce sizes and improve performance to the point where there seems essentially no question that MEMS-based set-top RAs will be achievable in the near future.

(b) *Fluidic* RAs are a very recent development that promises simple, cost-effective, rugged reconfigurable wire antenna structures in almost any shape imaginable. Fluidic elements are fabricated from a flexible eslastomer enclosing human-hair-thin channels filled with eutectic gallium/indium (EGaln), a liquid metal. The resulting "wire" provides electrical performance similar to copper and can be stretched, bent, rolled, or twisted into almost any shape simply by applying mechanical stress. The wire returns to its original linear structure upon relieving the stress. These elements have been used to fabricate a simple center-fed dipole that is tuned by stretching. A similar approach using a multiplicity of wires or other geometries (for example, bow ties or spirals) may well be the basis of a simple, effective set-top antenna. This technology has only recently been reported, and it certainly bears watching as a candidate for set-top RAs.

(c) *Pixelated piston* RAs comprise small individually controllable and addressable "pistons" in a two-dimensional matrix. Each piston is controlled by an actuator which, in turn, is operated by an electronic controller. A conductive patch is mounted on the end of each piston, which in their quiescent position form a continuous, flat ground plane. Each piston moves back and forth above and below the ground plane to form reconfigurable antenna radiating elements or transmission lines. This technology, which is patented, has been demonstrated to provide high-gain, beam steerable, antennas operating from 500 MHz to 18 GHz with reconfiguration times less than one millisecond. Its utility for the set-top application is self-evident, but at this point in its development the cost likely is prohibitive.

(d) *Liquid crystal* (LC) RAs are fabricated using LC substrates sandwiched between two electrodes that apply a DC bias voltage to control the LC's dielectric constant by deforming its molecules. Very small unit cells are fabricated that can be formed or applied onto surfaces of various shapes. At this point in its development, LC RAs have been demonstrated only on planar surfaces, and only at extremely high frequencies (high microwave through millimeter wave ranges). An effective LC RA reflectarray antenna has been demonstrated at 77 GHz. It provided complete electronic beam steering over an angular range of 35 degrees with good gain and sidelobe performance. Whether or not this technology can be ported to low VHF television set-top channels is not clear, but the technology bears watching because it is demonstrably effective for designing and building reconfigurable antennas.

(e) *Plasma* RAs are similar to the pixilated piston antennas, but instead of using conductive patches, they utilize ionized gas (plasma) as the conductive element. Each plasma element is turned “on” (ionized) or “off” (non-conducting) by a control voltage, so that the entire antenna aperture is electronically reconfigurable. Unlike pixilated piston RAs, the plasma elements lie in a single plane and cannot be positioned above or below the plane. This limitation, however, is minor in the context of set-top receive antennas because it is very likely that good receiving structures can be formed on a planar surface, which also simplifies the RA. The plasma RA has another potentially significant advantage over other candidate technologies. Existing plasma display technology, which is very highly developed and sophisticated, may be directly applicable to the design and manufacture of plasma RAs for the set-top application. Leveraging this existing technology may speed time to market and reduce costs substantially.

2.6.2 MEMS-based RAs

Microelectromechanical system (MEMS) radio-frequency switches have gained widespread acceptance as a standard element for implementing in RA structures. RF MEMS are routinely used to change both antenna feed networks to accomplish impedance matching and radiating element topology to control radiation pattern and efficiency. Most applications are in the microwave range, but MEMS use is not frequency-limited.

While there are other switching elements besides MEMS, notably sold-state PIN diodes and GaAs switches, MEMS offer several advantages. They are inexpensive, which is a major advantage for set-top applications, and they exhibit low insertion loss and low power consumption [2]. In recent years their availability as a commercial item has grown considerably, and they are now readily available as off-the-shelf components [3,4,5]. Demonstrated lifetimes exceed 100 billion cycles, which is consistent with consumer levels of usage over periods of several years.

Figure 1 shows a typical bi-layer curled MEMS structure. It is extremely small in size (300x1500µm). The two-layer switch element comprises materials with different thermal coefficients of expansion that pull it up away from the contact after annealing as shown in (a). The switch therefore is normally open (NO) when no control voltage is applied. Applying a DC voltage between the switch post and bottom electrode (1.5 µm thick polysilicon) creates an electrostatic field that pulls the switch element onto the contact as shown in (b). The microphotograph below the switch diagram shows its structure as actually fabricated.

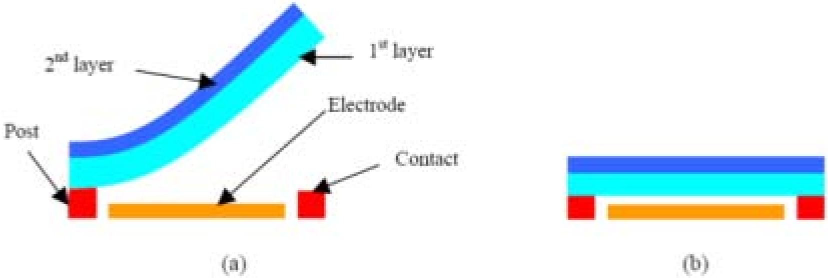

Figure 11. Operation of a bi-layer curled MEMS switch. (a) OFF state and (b) ON state.

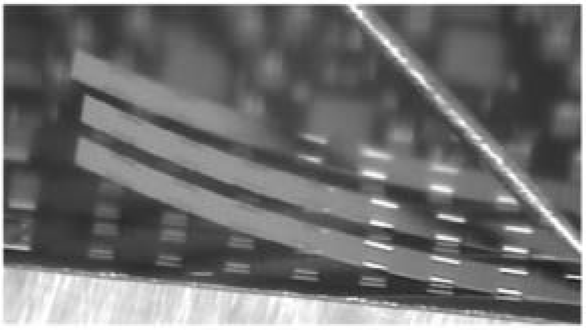

Figure 12. A photograph of the fabricated MEMS switches.

Figure 1. Typical RF MEMS switch (reproduced from [2]).

MEMS technology is improving at a rapid pace, and the latest generation of these switches are nano-scale [6]. *Nanoelectromechanical system* (NEMS) are laterally ~10-100 times smaller than a typical MEMS device, and they operate with actuation voltages below 10 volts, compared to 30-80 volts for MEMS switches. NEMS also provide faster switching times (<~1 µsec) compared to MEMS (~10-20 µsec). Figure 2 shows the new NEMS double-arm cantilevered switch schematically (a) and as fabricated (b) imaged by a scanning electron microscope.

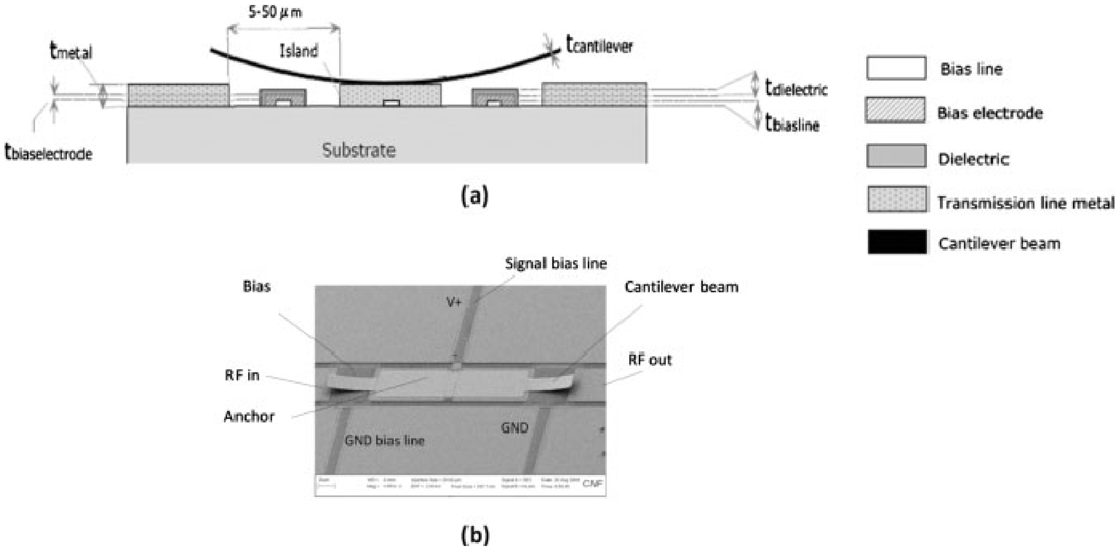
5-50 μm
Island
t_{metal}
$t_{cantilever}$
$t_{dielectric}$
$t_{biasline}$
$t_{biaselectrode}$
Substrate
(a)
Bias line
Bias electrode
Dielectric
Transmission line metal
Cantilever beam
Signal bias line
Bias
Cantilever beam
V+
RF in
RF out
Anchor
GND
GND bias line
CNF
(b)

Figure 2. DC-contact NEMS switch (reproduced from [6]).

Another important recent development is the use of carbon nanotubes (CNT) to fabricate NEM switches [7]. These devices promise extreme longevity and reliability because of CNT's unusual physical and electrical properties. Carbon nanotubes are extremely strong with high electrical conductivity. In addition, their flexibility and extremely low mass suggest that switching speeds in the GHz range eventually may be attainable. Figure 3 shows two types of CNT-based NEM switches. Measured switching times were 26 ns for these devices, which while far short of the GHz range, are nonetheless faster than MEMS devices.

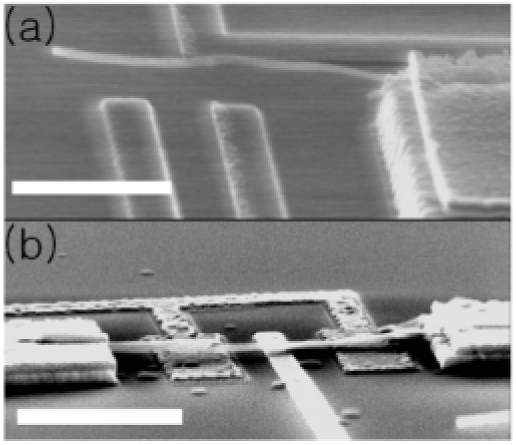

Figure 3. Two types of CNT-based NEMS switches (reproduced from [7]).

A typical MEMS-based reconfigurable array structure appears in Figure 4. This is an example of a complex RECAP (Reconfigurable Aperture) structure whose objective is wide bandwidth. Because a MEMS-based antenna can be reconfigured essentially in real-time, it can operate in multiple frequency bands. Aperture reconfiguration is achieved by switching in and out receiving and/or transmitting elements (radiating a signal, of course, is not an issue with set-top TV antennas). The structure shown in Figure 4 typically would be implemented using solid-state switches, but MEMS technology has leap-frogged that approach, so that MEMS provides better performance. In the array application, only the MEMS network departs from an otherwise standard design. The receiving/radiating element structure and its feed network are independent of the MEMS switch layer, as are the bias, FSS (frequency selective surface) and PBG (photonic band gap) layers. These last two elements may well not be required in an application like the set-top antenna because the operating frequencies are well below the microwave range where FSS and PBG structures are common.

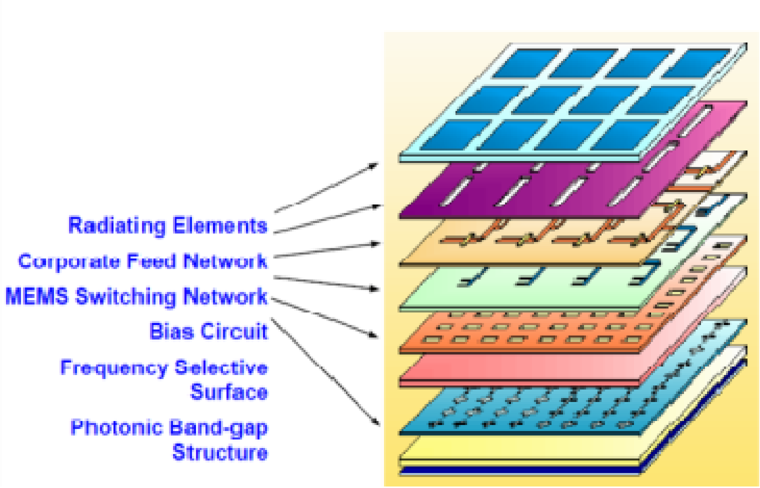

Figure 4. Reconfigurable MEMS-based antenna array (reproduced from [8]).

Another example of a state-of-the-art multiband MEMS-based antenna is reported in [9]. The design emphasizes a symmetric, repeatable topology which lends itself well to scalability, so that additional frequency bands are readily added. Scalability could be an important attribute in applying MEMS technology to the set-top environment, and this antenna demonstrates that scalability is achievable with MEMS. As configured the antenna covers four bands: 800-900 MHz; 1.7-2.5 GHz; 3.3-3.6 GHz; and 5.1-5.9 GHz. The lowest band lies just above the highest UHF television range, suggesting that this design likely could be easily modified to cover some portion of the upper TV channels. Table 2 reproduced below from [9] shows that this antenna indeed provides excellent performance in the four design bands. Its structure appears in Figure 5.

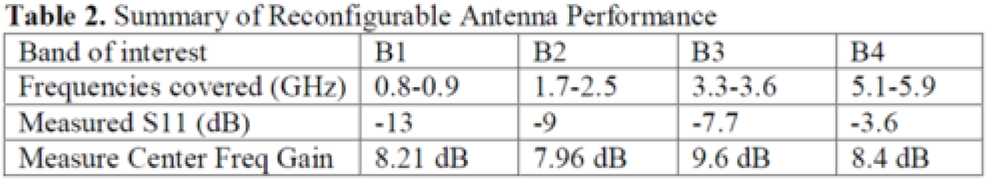

Table 2. Summary of Reconfigurable Antenna Performance

Band of interest	B1	B2	B3	B4
Frequencies covered (GHz)	0.8-0.9	1.7-2.5	3.3-3.6	5.1-5.9
Measured S11 (dB)	-13	-9	-7.7	-3.6
Measure Center Freq Gain	8.21 dB	7.96 dB	9.6 dB	8.4 dB

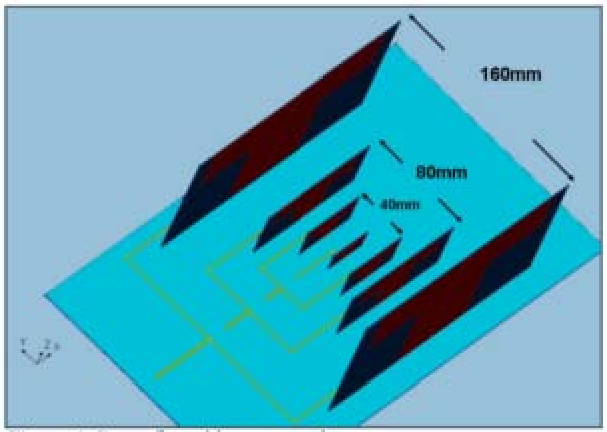

Figure 1. Reconfigurable antenna element.

Figure 5. Four band MEMS-based antenna (reproduced from [9]).

2.6.3 Fluidic RAs

An interesting new RA technology is the *fluidic* antenna. A fluid metallic alloy, eutectic gallium/indium (EGaln), comprising the metals gallium and indium, that remains liquid room temperature is injected into a microfluidic channel made from the silicone elastomer PDMS (polydimethylsiloxane). The channels are very small, about the width of a human hair, open at each end, and of any desired shape. After the channel is filled with EGaln, the alloy's surface oxidizes, creating a "skin" that holds the alloy in place while allowing it to retain its liquid properties. The alloy's mechanical properties are determined by the elastomer casing because it remains liquid.

The "wires" created by this process can be bent, stretched, rolled and twisted by applying a stress. Figure 6 shows a photograph of a twisted fluidic wire. Upon relieving the stress, the liquid conductor/elastomer reversibly returns to its original shape without hysteresis. An antenna fabricated of fluidic elements consequently is fully reconfigurable. The structure is self-healing to small cuts, and it is highly flexible and durable. Life expectancy therefore is very high, and fabrication costs reasonable because the elastomer's channels are formed using a technique known as "soft lithography" which avoids milling or etching in fabricating the fluidic antenna elements. However, the EGaln liquid metal is expensive, and at this time may be prohibitive for consumer applications.

A simple center-fed dipole was constructed from the fluidic element and tested. The antenna is shown in Figure 7. Its radiation efficiency was measured at approximately 90% over a frequency range of 1910-1990 MHz tuned by stretching the antenna. The efficiency is comparable to a copper element in this frequency range, so that the fluidic antenna is expected to perform as well as any metallic structure. In the television set-top application, this type of RA may be very attractive, especially if the price of EGaln alloy can be reduced because of volume. Of course, there are unanswered questions, for example, to what degree the elastomer casing can be stretched. The important conclusion that can be drawn from the reported research is that an entirely new fluidic technology may be close enough to maturity that it soon will be viable for reconfigurable set-top receive antennas.

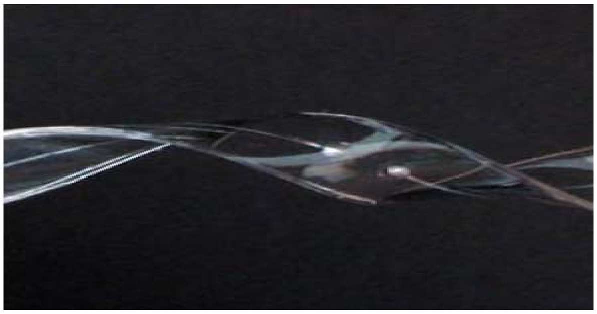

Figure 6. Twisted fluidic wire element.

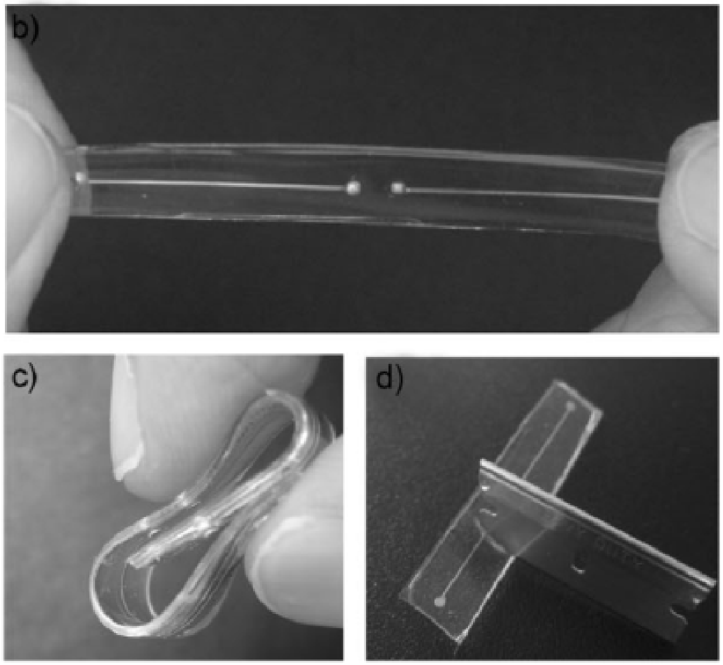

Figure 7 (a. b, c) Photographs of a prototype antenna being stretched and rolled. There is no hysteresis in the spectral properties of the antenna as it is returned to the ‘‘relaxed’’ state. d) The antenna self-heals in response to sharp cuts, such as those inflicted by a razor blade. Fluidic dipole antenna (reproduced from [10]).

2.6.4 Pixel Piston RAs

An interesting and potentially set-top feasible reconfigurable antenna comprises addressable conductive segments ("pixels") attached to the top of a movable piston in a two-dimensional array of pistons [11]. This RA appears in Figure 8. The piston includes a conductive patch (1010) sitting on top of a dielectric section (1020) that is attached to a conductive segment of the same triangular shape (1030), although other shapes are possible as well. The lower circular shaft (1040) is a dielectric material. This antenna is designed to operate at very high frequencies, 20 GHz being representative, and its dimensions are consequently commensurate. Typically the conductive pixel is about $\lambda/20$ across (0.7 mm at 221 GHz), while the height of the dielectric segment is $\lambda/10$ (1.4 mm at 21 GHz) with an overall height of about 11 mm.

The reconfigurable antenna is created by raising and lowering the pistons to form a transmission line and radiating elements in any desirable allowable pattern. If a pixel is in its lower position, it comprises an element in the array's ground plane. Otherwise, it is part of a transmission line or radiating element as shown in Figure 8 on the right. The entire structure is addressable on a pixel-by-pixel basis by a controller that operates a two-dimensional actuator for each pixel. Typical reconfiguration times are less than one millisecond, which is consistent with consumer requirements for the set-top application.

While the objective of this patented technology is electronically reconfigurable microwave patch antenna arrays, it may also be applicable to the set-top TV receive antenna if alternative element configurations are considered. For example, the pixel piston structure could be useful in creating folded element antennas, spirals, or bow-tie type elements. Each of these could be tuned to the required channel by sizing it appropriately.

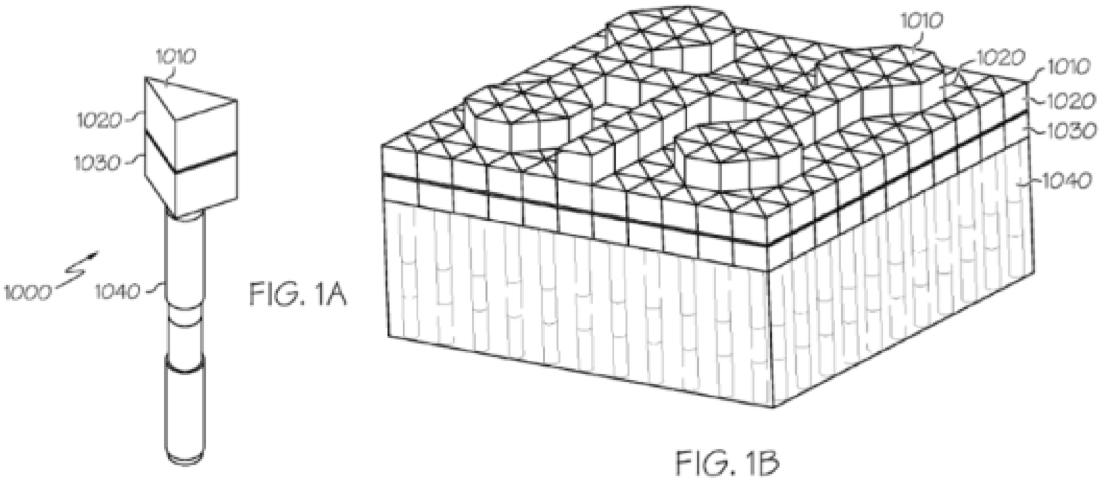

FIG. 1A

FIG. 1B

Figure 8. Pixel piston reconfigurable array (reproduced from [11]).

This technology has been implemented in commercially available antennas [12]. PARCA (Pixel-Addressable Reconfigurable Conformal Antenna) arrays are being developed for military applications requiring UWB (ultra wide-band) transmit antennas that adaptively reconfigure operating frequency, gain (beam width), and polarization while handling high power (~2 KW). A typical implementation operating from 1 to 18 GHz using approximately 100,000 ~1.65 mm diameter pixels would be 0.5 meter (20.5 in) square in size with a thickness of about 20 mm (0.8 in). Approximate weight is 12 lbs. This PARCA array would provide gain of 15 dBi and 40 dBi at 1 GHz and 18 GHz, respectively. The antenna is shown schematically in Figure 9.

Testing of microstrip transmission line and patch pixelated antennas has been done with good performance being demonstrated from 500 MHz to 18 GHz [13]. Whether or not this technology can be extended an order or magnitude lower in frequency for the set-top application with reasonable overall size is an open question. But it is clear from the current state-of-the-art that pixilated RAs constitute a technology to watch for the set-top receive antenna.

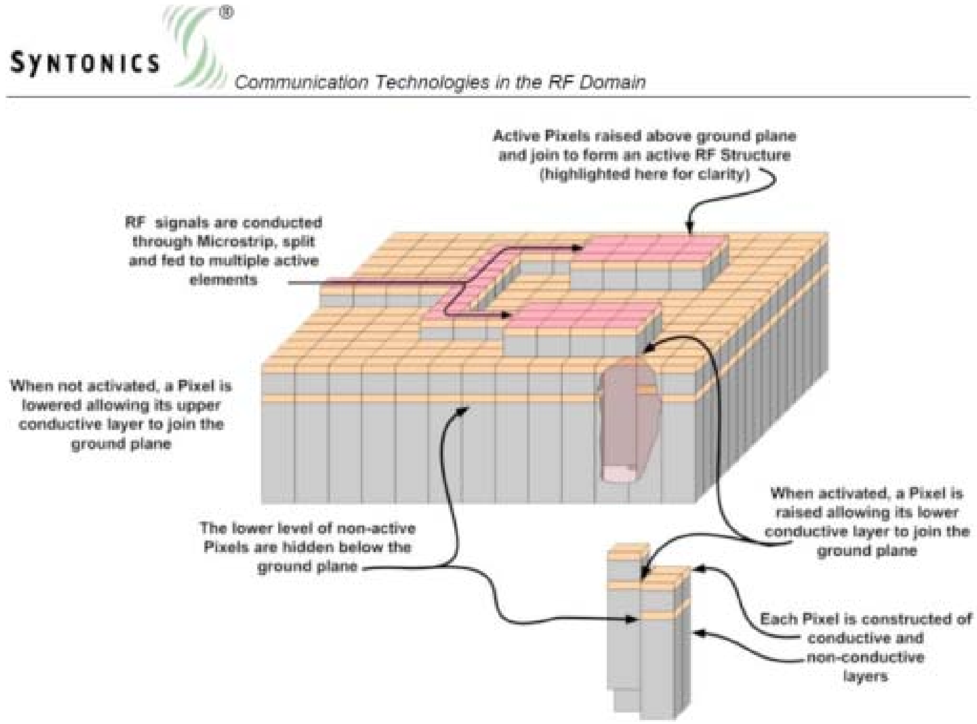
SYNTONICS
Communication Technologies in the RF Domain
Active Pixels raised above ground plane and join to form an active RF Structure (highlighted here for clarity)
RF signals are conducted through Microstrip, split and fed to multiple active elements
When not activated, a Pixel is lowered allowing its upper conductive layer to join the ground plane
When activated, a Pixel is raised allowing its lower conductive layer to join the ground plane
The lower level of non-active Pixels are hidden below the ground plane
Each Pixel is constructed of conductive and non-conductive layers

Figure 9. UWB PARCA array (reproduced from [13]).

2.6.5 Liquid Crystal Reconfigurable Antennas

Liquid crystal (LC) technology has recently been applied to RF front end components and array antennas for satellite applications, typically at quite high frequencies (> 20 GHz). Porting this technology to the low VHF range for a television set-top receive antenna simply may not be possible. Nevertheless, LC technology is worth monitoring because it may migrate to substantially lower frequencies, or it may inspire development of similar crystal-based technologies that are applicable in the VHF range.

Liquid crystals are inherently anisotropic and are so named because they simultaneously exhibit properties of liquid and crystalline materials. The class of *nematic* LCs provide the best dielectric characteristics from the microwave through millimeter wave bands, and therefore are the material of choice for very short wavelength RAs. A nematic LC molecule has a more or less rod-like shape a shown in Figure 10. Anisotropy is reflected in different perpendicular and parallel (relative to an incident electric field) dielectric constants, typical values being ~2.5 and ~2.8-3.5, respectively. Application of a bias voltage to an LC causes a deformation of the molecular alignment as shown schematically in Figure 10, the effect of which is to change the LC's dielectric properties. These changes, in turn, constitute the basis of a reconfigurable antenna element.

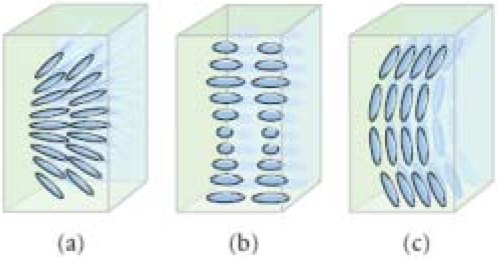

FIGURE 3: (a) The splay, (b) twist, and (c) bend deformation.

Figure 10. LC deformation modes (reproduced from [14]).

A reflectarray RA was fabricated using a dielectric substrate whose properties could be tuned using a bias voltage less than about 20 volts. The RA is a waveguide-fed planar reflecting surface capable of accurately steering the radiated beam into a wide range of directions. A unit cell comprising a single microstrip patch of dielectric sandwiched between two parallel plate electrodes is shown in cross-section in Figure 7 (left). The upper electrode patch is printed onto the dielectric substrate as is the lower electrode which acts as the antenna's ground plane. Initial LC molecule alignment is achieved using a thin (~300 nm) polyimide film on the electrode patch and ground plane. The applied external electric field created by the bias voltage reorients the LC molecules, thereby changing its dielectric properties. The figure shows the variation of dielectric constant as the bias voltage increases from zero to a value slightly exceeding the LC's threshold voltage, V_{th}, and then to a value that is much greater. As the effective dielectric constant changes, so does the capacitance per unit length of the microstrip patch, which allows the cell to be voltage-tunable as a reconfigurable antenna element.

The reflectarray antenna comprises 16x16 unit cells spaced 0.55λ. Each cell is 2.2 mm across with an LC cavity height of 50 μm. The complete antenna is shown in Figure 11, in which the individual cells are clearly visible, as are the control lines (bias lines). The antenna is fed using a right-angle bent WR-10 open-ended waveguide as shown in the figure. Figure 12 shows measured pattern data for different bias voltage levels. The main lobe is clearly steered into the desired directions of zero (blue), +25 (black), and -10 (red) degrees with good relative main lobe gain and sidelobe suppression.

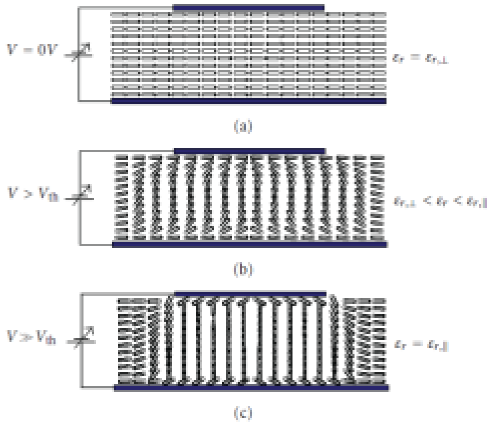
V = 0V
$\varepsilon_r = \varepsilon_{r,\perp}$
(a)
V > V_th
$\varepsilon_{r,\perp} < \varepsilon_r < \varepsilon_{r,\parallel}$
(b)
V ≫ V_th
$\varepsilon_r = \varepsilon_{r,\parallel}$
(c)

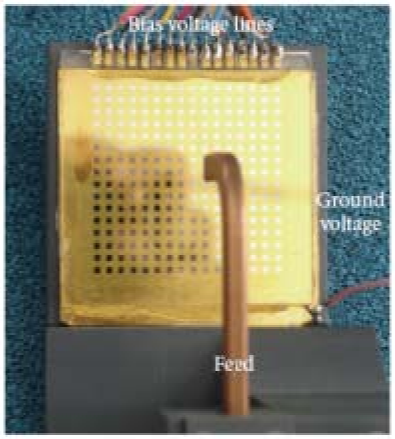
Bias voltage lines
Ground voltage
Feed

FIGURE 13: Fully assembled 77 GHz LC reflectarray with feed.

Figure 11. 77 GHz LC reconfigurable reflectarray (reproduced from [14]).

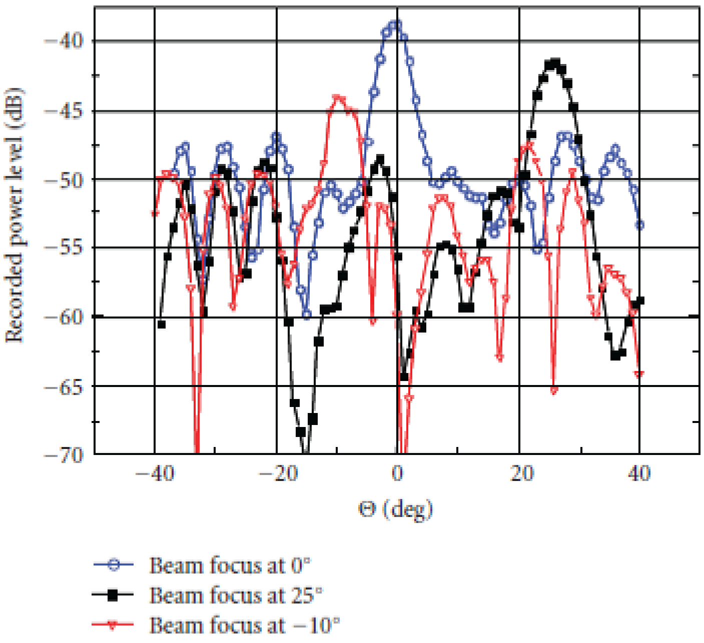

FIGURE 14: Three recorded E-plane patterns for three different voltage configurations, with the main beam pointing at different angles: −10°, 0°, and +25° (at 77.2 GHz).

Figure 12. LC reflectarray measured patterns (reproduced from [14]).

2.6.6 Plasma Reconfigurable Antennas

The plasma RA [15] is conceptually similar to the PARCA antenna in §4. It also uses electrically conductive/non-conductive surfaces to create radiating elements and associated components. But instead of conductive metallic patches, the plasma RA utilizes gas enclosures in which the gas can be made electrically conductive (plasma state) or non-conductive. The enclosures are interconnected and can be turned on (conducting) and off (non-conducting) by applying a control voltage. Figure 13 shows the basic plasma RA structure. It comprises square cells that contain four control electrodes, one on each edge. In the diagram, a cell that is turned on is shown by wavy lines between the electrodes. For example, cell 110 is "on" as shown by the wavy lines which indicate that the gas within the cell has been ionized into a plasma state. By contrast, cell 202 is "off" as indicated by the absence of wavy lines; its gaseous contents have not been ionized. Turning individual cells on and off in the manner forms the basis for a reconfigurable antenna system.

FIG. 2

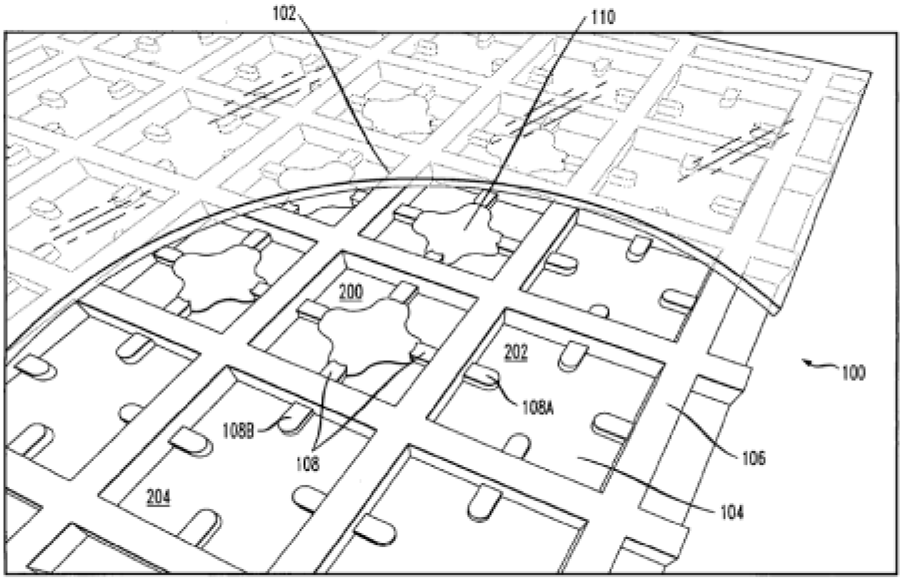

Figure 13. Plasma RA structure (reproduced from [15]).

While the reconfigurable aperture in Figure 13 can be used on a stand-alone basis, it frequently is coupled to backplane as shown in Figure 14. The backplane comprises control elements (304) that are arrayed on an insulating substrate (302) with conductive traces (312) that form the control lines. When mated with the plasma structure, the control elements turn individual gas enclosures on and off. A typical complete plasma RA is shown in Figure 15, which adds a ground plane (404) printed on a dielectric substrate (402). The sandwich comprising the individually controllable plasma enclosures, the groundplane, and the backplane form an electronically reconfigurable antenna system.

The plasma RA may be particularly useful as a set-top antenna. One advantage is has over other approaches is that it almost certainly can benefit from the highly developed plasma screen technology used in consumer televisions. A plasma RA should be much simpler in structure and fabrication than a large plasma screen display. Applying existing display technology to manufacturing plasma RAs may be only an incremental step, thereby reducing time to market and costs.

FIG. 3

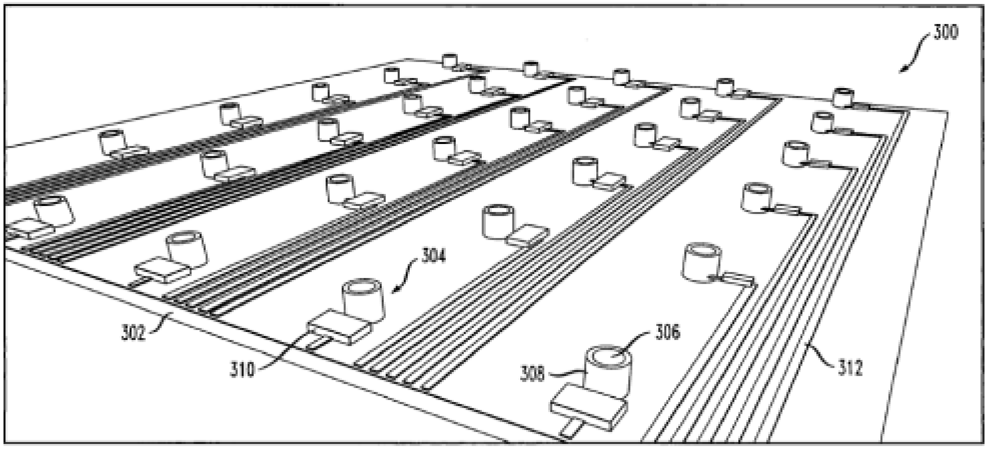

Figure 14. Plasma RA backplane (reproduced from [15]).

FIG. 4

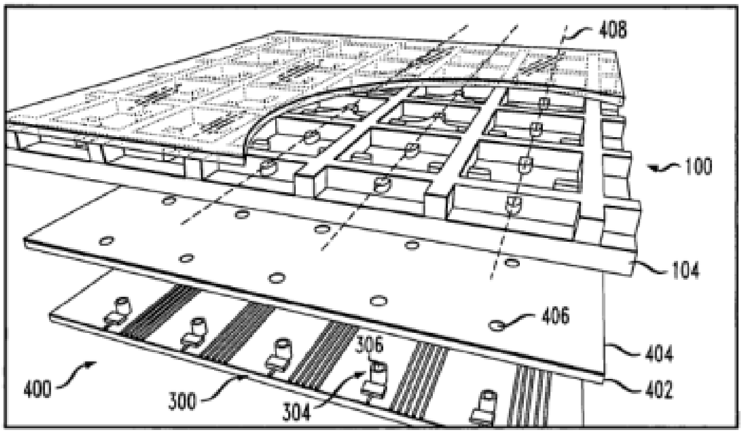

Figure 15. Complete plasma RA (reproduced from [15]).

2.6.7 References

[1] Reed, D. (ed.), **The ARRL Handbook (2004)**, American Radio Relay League, Newington, CT 06111 USA.

[2] Ativanichayaphong, T., Cai, Y., Wang, J., Chiao, M., and Chiao, J.-C., "Design Considerations of Reconfigurable Antennas using MEMS Switches," SPOIE 2005 Microelectronics, MEMS, and Nanotechnology Symposium, Brisbane, Australia, Dec. 11-15, 2005.

[3] RadantMEMS, Inc., 255 Hudson Road, Stow, MA 01775 USA.

[4] DowKey Microwave Corp., 4822 McGrath Street, Ventura, CA 93003 USA.

[5] MEMScomm, LLC, 419r East Main Street, Morehead, KY 40351 USA.

[6] Cetiner, B., Biyikli, N., Yildirim, B. S., and Damgaci, "Nanoelectromechanical Switches for Reconfigurable Antennas," *Microwave and Optical Technology Letters*, **52**, No. 1, January 2010.

[7] Lee, S., Eriksson, A., Abdelrahim, A. S., Campbell, E. E. B., "Carbon-nanotube-based Nano Electromechanical Switches," *Jour. Korean Physical Society*, 55, No. 3, pp. 957-961, Sept. 2009.

[8] Sabel, K. F., Katehi, L. P. B., Sarabandi, K., "Modeling and Design of MEMS-Based Reconfigurable Antenna Arrays," *IEEE Aerospace Conference, 2003. Proceedings*, p. 2-1135. DOI: 10.1109/AERO.2003.1235528

[9] Wu, T., Li, R. L., Eom, S. Y., Lim, K., Jeon, S. I., Laskar, J., and Tentzeris, M. M., "A Multiband/Scalable Reconfigurable Antenna for Cognitive Radio Base Stations," IEEE *Antennas and Propagation Society International Symposium, 2008. AP-S 2008,* 5-11 July 2008, pp. 1-4. DOI: 10.1109/APS.2008.4619490.

[10] So, J-H., Thelen, J., Qusba, A., Hayes, G. J., and Dickey, M. D., **"Reversibly Deformable and Mechanically Tunable Fluidic Antennas, "** ***Advanced Functional Materials*****, 19, pp. 3632-3637 (2009). DOI: 10.1002/adfm.200900604.**

[11] Walton, *et al*., "Reconfigurable Antenna Using Addressable Pixel Pistons," U.S. Patent No. 7,561,109 B2, July 14, 2009.

[12] Syntonics, LLC, 9160 Red Branch Road, Columbia, MD 21045 USA.

[13] Syntonics, http://www.syntonicscorp.com/

[14] Gaebler, A., Moessinger, A., Goelden, F., Manabe, A., Goebel, M., Follmann, R., Koether, D., Modes, C., Kipka, A,. Deckelmann, M., Rabe, T., Schulz, B., Kuchenbecker, P., Lapanik, A., Mueller, S., Haase, W., and Jakoby. R., "Liquid Crystal-Reconfigurable Antenna Concepts for Space Applications at Microwave

and MillimeterWaves," *International Journal of Antennas and Propagation*, Volume 2009, Article ID 876989, 7 pp. DOI:10.1155/2009/876989.

[15] Metz, C., "Reconfigurable Plasma Antenna with Interconnected Gas Enclosures," U.S. Patent Application Publication No. US 2006/0220980 A1, Oct. 5, 2006.

2.7 Use of Metamaterials in Antenna Systems

2.7.1 Summary

Traditional antennas are fabricated from electrical conductors such as copper or aluminum and insulators such as plastic or rubber. Their performance is limited by their electrical size, that is, the largest dimension measured in wavelengths. Electrically small antennas are narrow-band, inefficient, and difficult to impedance-match. The demands on a television set-top antenna are for the most part antithetical. It must be small, but it requires a very wide bandwidth to cover all VHF/UHF broadcast channels with good gain and efficiency. Traditional antennas therefore have fallen short of providing good set-top performance.

A very promising emerging technology that may soon well address this limitation is the use of *metamaterials* (MTM) in the set-top antenna. This section examines whether or not MTM can sufficiently improve the performance of electrically small antennas to make MTM-based designs attractive for set-top use. While MTM have been available since the mid-1990s and the subject of much literature, they have not gained widespread commercial acceptance until recently. A recent survey paper examines this question and suggests some explanations for the MTM's slow acceptance, much of which is not based on the technology itself.

More recent developments show that well-designed MTM-based antennas indeed can provide significantly improved performance over traditional designs, but the reported designs are in frequency bands well above the TV VHF band. Wireless communication antennas have been designed, fabricated, and measured that improve on the "Chu limit" for small antenna performance by orders of magnitude using MTM. Substantially improved bandwidth, high efficiency, and good gain are available from antennas with maximum dimensions less than one-tenth wavelength. The question is whether or not this technology can be ported to the low end of the television broadcast band; and, while the answer is not guaranteed, it appears to be in the affirmative.

The MTM-based antenna is an attractive candidate for the set-top application, and it should be vigorously studied as a candidate architecture. Because MTM typically are formed from simple wire structures, they likely can be optimized for the set-top application, which may lead to designs that are different than those used at higher frequencies that still meet set-top performance objectives.

2.7.2 What is a *metamaterial*?

The term *metamaterial* (MTM) applied to antennas does not connote a material in its commonly used sense. The materials used to fabricate an antenna broadly fall into the categories of conductors (metallic elements), insulators (non-conductors), and, occasionally, semiconductors. Materials like copper, aluminum, silver, various plastics, rubber, paper, and so on, are what come to mind as typical antenna materials; and all of these share some common features. Whether naturally occurring or man-made, they are homogeneous with electrical properties determined by their chemical structure. Like all other naturally occurring substances (at least all that are known at this time), antenna materials exhibit positive relative permittivity (dielectric constant), $\varepsilon_r \geq 1$, and positive relative permeability, $\mu_r \geq 1$. Their electrical performance in the antenna is determined by these fundamental physical parameters.

In sharp contrast, a *metamaterial* is a "composite" whose electrical properties are not determined by chemical structure, but instead by the shapes and patterns of inclusions that often are imbedded in some other material or simply arrayed in a pattern on a substrate [1]. A common inclusion is the split-ring resonator (SRR), and often combinations of SRRs and small wires are arranged to create a "unit cell" from which the metamaterial is built. Arrays of cells on some surface or within some volume constitute the actual "metamaterial." As a concrete example, Figure 1 shows a unit cell comprising a double SRR on an insulating substrate and a parallel wire. The SRR and wire are normally conducting materials such as copper or aluminum. An array of unit cells deployed on or embedded in a surface of any shape or throughout some volume becomes the *metamaterial*, which evidently is quite different from a homogeneous piece of plastic, for example, whatever its shape.

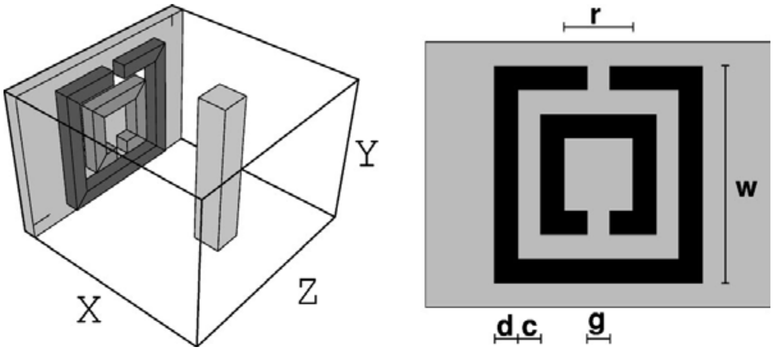

Figure 1. SRR/wire unit cell (reproduced from [1]).

What makes metamaterials attractive as antenna components is that they are *left-handed*, in distinction from all naturally occurring materials, which are *right-handed*. In right-handed materials (RHM), the electric field, $\vec{E}$, the magnetic intensity, $\vec{H}$, and propagation (wave) vector, $\vec{k}$, are related by the "right-hand rule," *viz*, $\vec{E} \text{ x } \vec{H} \approx \vec{k}$. The phase and group velocities in RHM are in the same direction, parallel to the wave vector. A *left-handed material* (LHM) obeys a "left-handed rule" in which the phase and group velocities propagate in *opposite* directions. Physically this results in energy flow (Poynting vector) antiparallel to the direction of wave propagation. For LHM, both the effective permittivity and permeability are negative, $\varepsilon_r \leq 0$ and $\mu_r \leq 0$, which gives rise to LHM's unusual electromagnetic (EM) properties.

While the notion that this type of behavior can arise from as simple a unit cell as an SRR close to a wire perhaps seems counter-intuitive, it is not. There are examples of very simple wire structures exhibiting similar behavior. For example, the three-dimensional mesh of orthogonal crossed wires in Figure 2 exhibits a negative dielectric constant for frequencies $\omega < \omega_p$ where $\omega = 2\pi f$, $\omega_p = (c/\mathrm{d})[2\pi/\ln(\mathrm{d}/\mathrm{w})]^{1/2}$, and c is the velocity of light [2]. Thus, properly arrayed unit cell structures comprising conducting and insulating elements simply exhibit a similar behavior that is determined by the specific details of how the unit cell is structured.

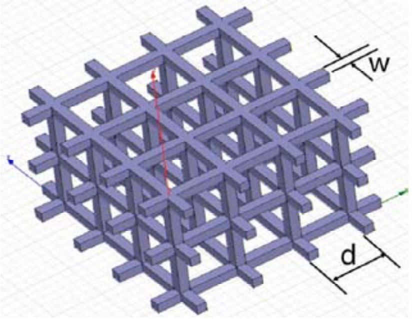

Figure 2. Crossed-wires metamaterial (reproduced from [2]).

Without doubt, the most important consequence of having $\varepsilon_r \leq 0$ and $\mu_r \leq 0$ is that the metamaterial's refractive index is negative, $n < 0$, which leads to some very non-intuitive EM effects. Besides energy flow opposite to the propagation direction, $n < 0$ results in *negative refraction* as shown in Figure 3. In the figure, air lies above the interface with either LHM or RHM. For both types of lower materials, a reflection takes place at the interface that follows the usual law of reflection (incidence and reflection angles being equal, rays 1 and 2). If the lower medium is RHM (normal material), the propagating ray (#4) is bent towards the normal. This is the usual refraction that takes place with all normal materials, and, for example, bends the image of a pencil in water towards the vertical. But, if the lower medium is LHM, just the opposite occurs. The propagating ray (#3) is bent *away* from the vertical, not towards it.

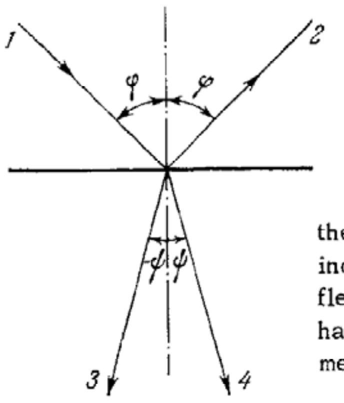

FIG. 3. Passage of a ray through the boundary between two media. 1 – incident ray; 2 – reflected ray; 3 – reflected ray if the second medium is left-handed; 4 – refracted ray if the second medium is right-handed.

Figure 3. LHM and RHM rays at air-metamaterial interface (reproduced from [3]).

2.7.3 Slow Acceptance of LHM Cellular Antennas

While there is no question that metamaterials offer intriguing possibilities for novel antennas, their use has not become widespread. Some possible explanations for this situation are offered in [4], which specifically addresses application to mobile-phone antennas. Because the cellular band (825-894 MHz, $\lambda = 0.36 - 0.335m$) lies just above the broadcast television band (54-806 MHz, $\lambda = 5.55 - 0.37m$), the issues raised in [4] are largely applicable to TV antennas.

An important consideration in whether or not LHM are useful is the antenna's *electrical size*, that is, its size measured in wavelength, λ. The wavelength at the highest TV channel 69 is the same order as the cell phone wavelengths, whereas TV channel 2 has a wavelength about 15 times larger than the longest cellular wavelength. This observation raises the question of whether or not general statements or conclusions made about antennas for cellular frequencies hold up for low-band TV antennas. The answer is a clear "maybe."

A typical cell phone antenna with a largest dimension of 0.1m has an electrical size of 0.28λ at 825 MHz. A channel 2 TV antenna has the same electrical size if its largest dimension is 1.55m, which is too large for a set-top antenna. A more reasonable set-top dimension might be 0.5m, which is 0.09λ at 54 MHz. So the question is whether or not observations about a 0.28λ antenna can be applied to one that is about three times smaller. There is no definitive answer, but the electrical sizes are close enough that some attributes of metamaterials applied to cellular antennas can be ascribed to TV antennas as well. Thus, some of the reasons in [4] why LHM are not widely used in cellular antennas may apply to TV antennas as well. Among those limitations are:

1. *Bandwidth.* Metamaterials tend to exhibit strong resonances accompanied by substantial dispersion and losses. The result is a high-*Q*, narrow-band structure that makes it very difficult to cover the entire range of VHF and UHF television frequencies. An antenna covering all cellular frequencies has a fractional bandwidth (BW) of 8% (relative to band center). If a set-top TV antenna contains two antennas, one low-band (channels 2-13, 54-216 MHz) and the other high-band (channels 14-69, 470-806 MHz), they must have fractional bandwidths of 120% and 53%, respectively, to cover all channels. If achieving 8% BW in a LHM cellular antenna is difficult, then meeting the requirements of a set-top antenna is even more difficult.

2. *Other Factors.* Whether or not an antenna is "good" for a specific application depends on many factors, many of which are not electrical in nature. Antennas that perform well electrically (gain, bandwidth, efficiency) may be too costly, too large, too heavy, too difficult to manufacture, and so on. Reports of successful LHM-based antennas usually do not address these factors, and as a result it may be difficult to analyze the benefits and risks associated with that type of antenna. This consideration has slowed the adoption of LHM antennas for mobile phones and likely applies to TV set-top antennas as well.

3. *Marketing*. Like any consumer product or component thereof, LHM antennas must be properly marketed to be effectively used. Many types of man-made materials are used in antennas, especially in the microwave range, and the resulting designs are considered "traditional." LHM-based designs often resemble traditional ones, so that they share many common features. Yet LHM antennas are frequently described as being fundamentally "different," which leads to some measure of reluctance in adopting them. A better approach would be to characterize LHM antennas as *alternatives* to traditional designs rather than a fundamental departure.

2.7.4 Examples of LHM Antennas

Even though the speculation in [4] provides possible reasons why MTM-based antennas are not more widely used, recent developments show rather convincingly that technical objections are not valid reasons. Quite to the contrary, state-of-the-art MTM antenna designs support the view that they well may be able to meet set-top TV performance requirements. This section consequently examines some typical designs in the context of porting them to the low VHF television channels.

(a) *Compact Broadband Monopole.* An example of the state-of-the-art in small broadband LHM antennas is shown in Figure 4(a), which is reproduced from [5]. The antenna comprises a PCB (printed circuit board) monopole loaded with a left-handed negative-refractive-index transmission-line (NRI-TL) metamaterial. The loading reduces the electrical size of the antenna while providing large bandwidth.

This particular antenna was designed for dual-band use in the WiMax and WiFi bands [3.3-3.8 GHz, ($\lambda = 0.09\text{m} - 0.08\text{m}$) and 5.15-5.85 GHz ($\lambda = 0.06\text{m} - 0.05\text{m}$), respectively]. Its physical size of 0.03m [W_g in Figure 4(a)] corresponds to 0.3λ at the lowest frequency, which is comparable to the electrical size of a cellular antenna and about three times larger than a typical 0.5m set-top antenna at channel 2. Adding NRI-TL loading increased BW over the unloaded structure by approximately 1.5 GHz without increasing its size. The LHM monopole's BW (return loss below -10 dB) is 3.84 GHz from 3.15-6.99 GHz (fractional BW of 76%). The computed radiation efficiency is close to 90% across both the WiMax and WiFi bands. Input impedance is plotted in Figure 4(b), and it shows moderate values across a very wide frequency range, from about 5GHz to 10 GHz, so that good impedance matching should be readily achievable without requiring a balun.

The compact LHM monopole provides a good example of enhancements directly attributable to NRI-TL loading of a simple structure: excellent impedance bandwidth and high efficiency in an electrically small structure. Although this particular antenna would not meet a set-top TV antenna's requirement of 120% BW (see §3), it certainly suggests that with further improvements that objective might be reached.

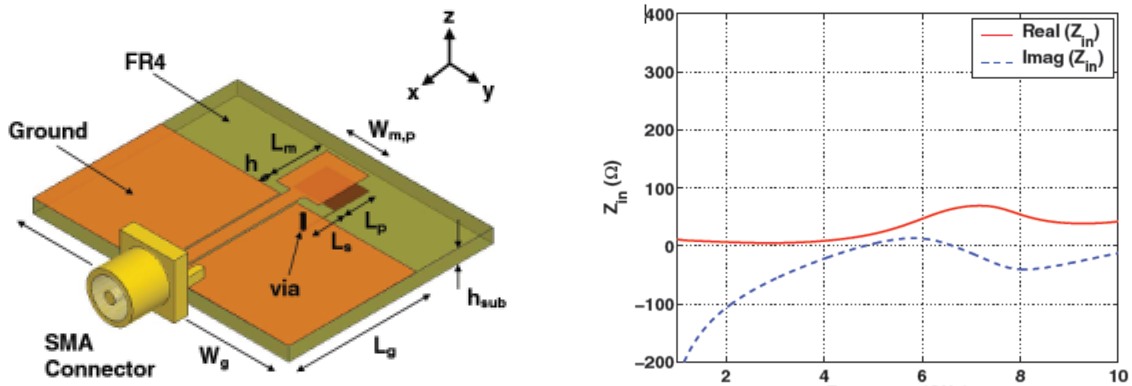

Fig. 1. NRI-TL metamaterial monopole antenna. All dimensions are in mm: $L_m = 6$, $W_m = 5$, $L_g = 15$, $W_g = 30$, h = 1, $L_p = 4$, $W_p = 5$, $L_s = 3.45$, $W_s = 0.1$, $h_{sub} = 1.59$, $S_{cpw} = 0.2$, $W_{cpw} = 1.55$, via diam. = 0.5.

Fig. 4. Simulated input impedance characteristics of the unloaded monopole antenna from Ansoft HFSS.

(a) (b)

Figure 4. Broadband LHM monopole: Structure (a); Input impedance (b) (reproduced from [5]).

(b) *Efficient Electrically Small Antenna (EESA).* An MTM-based Efficient Electrical Small Antenna (ESSA) [6] is shown in Figure 5. It provides another example of how LHM antennas can achieve performance that is not possible with traditional designs, and suggests that porting MTM designs to TV set-top antennas is a viable approach to achieving improved performance.

The ESSA comprises a monopole surrounded by an SNG (single negative) MTM structure mounted on a plane substrate. It achieves a voltage standing wave (VSWR) bandwidth ≤ 2 of 470 MHz (1.75-2.22 GHz) [fractional VSWR bandwidth of 24%] in a package with an electrical size of only 0.07λ. The addition of LHM in this case resulted in a much broader bandwidth than could achieved otherwise because the effect MTM's effect is to lower the system *Q*-factor from the theoretical "Chu limit" of 16.72 for traditional materials to 2.23. Significantly, the increased bandwidth is not accomplished at the expense of radiation efficiency or gain, which often must be sacrificed in traditional designs. The ESSA's measured maximum gain and efficiency were 2.44 dBi and 83.61%, respectively.

The radiation patterns generally exhibited broad lobes with fairly shallow nulls ($<\approx -25\text{dB}$, Figure 6), which lend themselves well to omnidirectional reception as required in a set-top antenna. Because the ESSA's electrical size is comparable to that required for a set-top antenna (see §3), it may be possible to "stack" several ESSA's covering different band segments in a fairly small set-top package. This particular ESSA design suggests that MTM-based antennas may well meet TV set-top reception requirements.

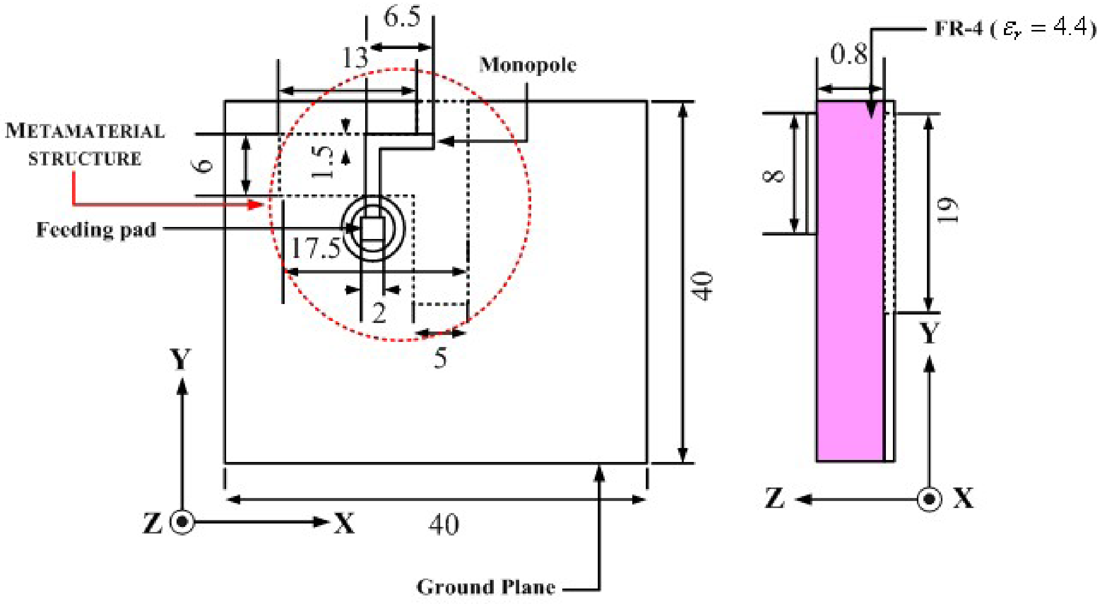

Figure 5. MTM ESSA geometry (reproduced from [6])

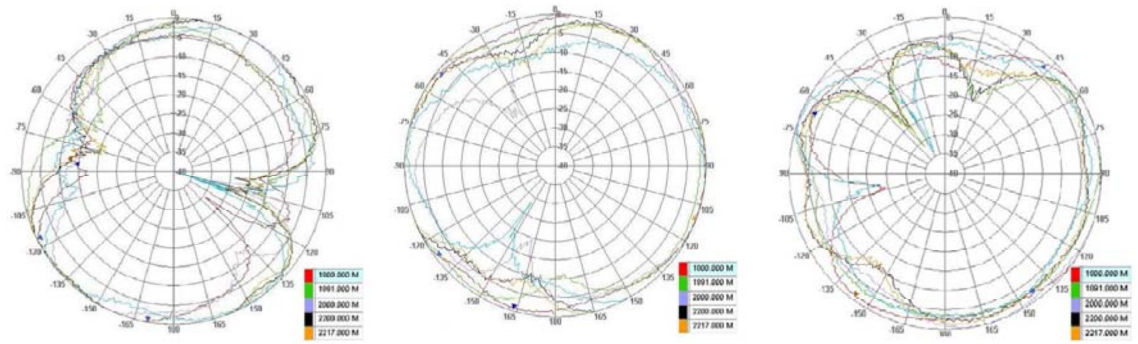

Figure 6. ESSA measured radiation patterns (reproduced from [6]).

(c) *LHM Topologies.* There are some general approaches to MTM-based antennas that appear to be useful regardless of the specific type of radiating element [7], and these likely should be the focus of future efforts to apply LHM to television set-top antennas. Various types of materials are available as shown in Figure 7. “Normal” materials have positive permittivity and permeability, and are referred to as Double Positive (DPS). “Epsilon negative” (ENG) materials can occur naturally in unusual states of matter (for example, plasmas), while *homogeneous* “mu negative” (MNG) materials can be man-made (for example, ferrite). LHM materials, on the other hand, do not occur in Nature and are “Double Negative” (DNG) because both permittivity and permeability are negative. While composite MTMs usually are DNG, ENG and MNG types also can be made and sometimes are desirable in antenna applications.

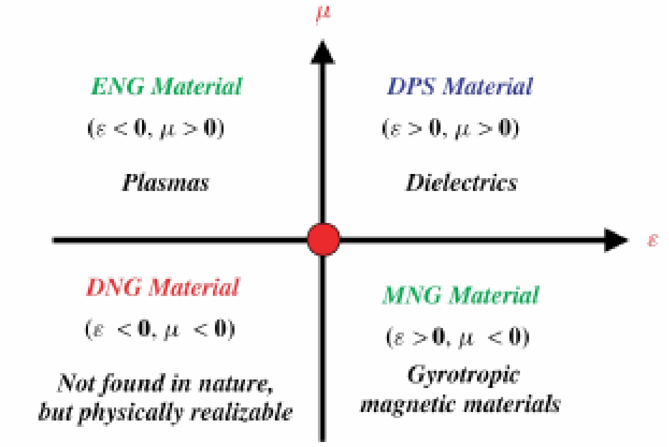
μ
ENG Material
$(\varepsilon < 0, \mu > 0)$
Plasmas
DPS Material
$(\varepsilon > 0, \mu > 0)$
Dielectrics
ε
DNG Material
$(\varepsilon < 0, \mu < 0)$
Not found in nature, but physically realizable
MNG Material
$(\varepsilon > 0, \mu < 0)$
Gyrotropic magnetic materials

Figure 7. Material classification (reproduced from [7]).

Figure 8 shows two LHM configurations that should be useful in a wide range of applications and, in particular, TV set-top antennas. In Figure 8(a) a PCB antenna is placed on an MTM substrate in order to decrease radiating element size while in creasing radiation efficiency and bandwidth. In this example, the substrate is an MNG array of square SRRs imbedded in a homogeneous dielectric medium. Figure 8(b) shows an ENG shell surrounding a radiator fed against a PEC (perfectly electrically conducting) ground plane. The shell can be very thin ($<\approx 0.01\lambda$) so that it has little effect on attenuating signal level, yet it substantially reduces antenna input reactance over a fairly wide range of frequencies. While the "Chu *Q*-factor" is the recognized theoretical limit for antennas made from normal materials, the ENG shell can lower *Q* by a factor of 1,583 while providing a good match to 50 Ω and efficiencies approaching 98%, at least over some range of frequencies. These examples illustrate LHM's potential benefits for TV set-top antennas, and they should be vigorously pursued for the next generation of set-top antennas.

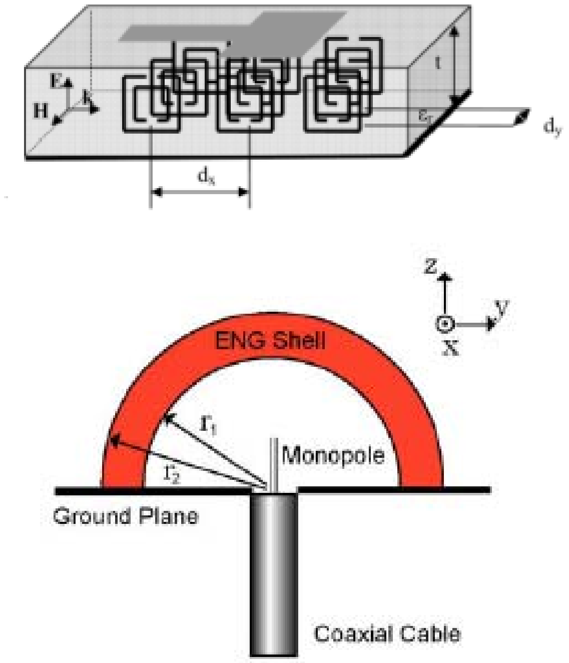

(a) (b)

Figure 8. Generally useful LHM configurations (reproduced from [7]).

2.7.5 LHM Antenna Patents

An online search at Patent Storm (http://www.patentstorm.us/) using the keyword 'metamaterial' returns 521 U.S. issued patents and applications, many of which are related to antennas or optics. The application of MTM to antennas is a very active technology area, both in the scientific and engineering literature and in the realm of patented proprietary technology. While a thorough review of the patent literature is beyond the scope of this report, it is instructive to examine a specific development as representative of the state-of-the-art.

Figure 9 shows a typical implementation disclosed in U.S. patent #7,592,957 B2, which issued recently (Sept. 22, 2009) and consequently representative of the state-of-the-art. The figure illustrates one implementation of many described in the patent (forty-nine drawing sheets, many with multiple figures). This patent discloses composite right- and left-handed (CRLH) MTM structures as the basis of electrically small antennas and antenna arrays for use in wideband multi-band communications. One design, for example, provides 5.6 dBi gain with a 2.35-4.42 GHz BW (S_{11}<-10 dB). This level of performance is representative of what can be achieved today for wireless communication applications, and it is quite good compared to the performance of traditional antennas. It bodes well that LHM-based antenna designs likely can be useful for the TV set-top application.

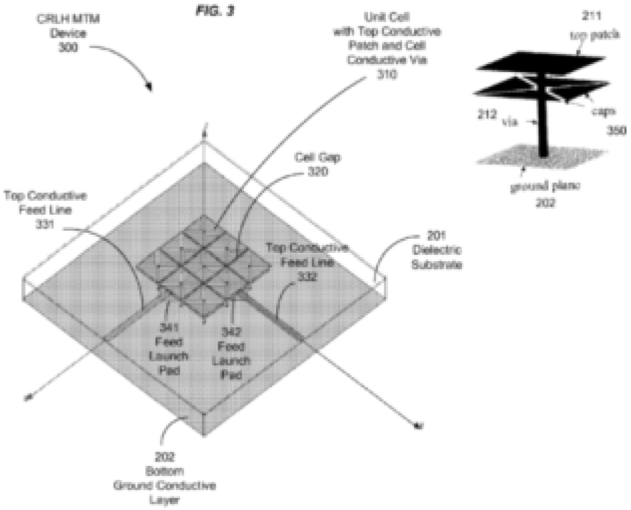
FIG. 3
CRLH MTM Device 300
Unit Cell with Top Conductive Patch and Cell Conductive Via 310
211 top patch
212 via
caps 350
ground plane 202
Cell Gap 320
Top Conductive Feed Line 331
Top Conductive Feed Line 332
201 Dielectric Substrate
341 Feed Launch Pad
342 Feed Launch Pad
202 Bottom Ground Conductive Layer

Figure 9. Typical CRLH MTM 2-D array (reproduced from U.S. patent #7592957)

2.7.6 Theory of LHM

A *metamaterial* (MTM) or *left-handed material* (LHM) has simultaneously negative values of relative permittivity (ε_r) and permeability (μ_r). There are no known naturally occurring substances with this attribute, but there are artificial media exhibiting left-handedness. MTM are useful antenna components because unusual electromagnetic effects occur when ε_r and μ_r are both negative. For example, the direction of power flow, wave refraction, Doppler shift, and the direction of Cherenkov radiation are all dramatically different than they are in “normal” media. These and perhaps other unusual electromagnetic properties of MTM may be useful in antenna system design, particularly in the microwave region.

Artificial MTM can be manufactured from arrays of conducting elements, such as small wire segments, split ring resonators, or other similar structures. These discrete elements may mimic continuous media, because, surprisingly, their left-handedness can extend to frequencies where the wavelength is much larger than either the element size or spacing. MTM slabs may be useful as coatings on metallic antenna parts. If the metamaterial can be "worked" appropriately, then major antenna components might be fabricated from it directly. The utility of these materials is an open question, because they are not well understood either theoretically or experimentally. In fact, serious investigation of these materials has occurred only in about the past decade. What follows is a discussion of how metamaterials behave and some of the issues in using them as antenna components.

Maxwell's Equations

All electromagnetic (EM) phenomena, including the EM behavior of metamaterials, are described by Maxwell's equations. Maxwell's equations are completely general, and apply to EM phenomena in all media at all frequencies, including electrostatics ("zero" frequency). MTM behavior must be consistent with Maxwell's equations in every limiting case for which theoretical results are available, whether electrostatic or time-varying. Limiting cases may be useful in understanding the counter-intuitive behavior of metamaterials.

Maxwell's equations are:

$$\nabla \cdot \mathbf{E} = \frac{\rho_t}{\varepsilon_r \varepsilon_0} \tag{1}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{2}$$

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \tag{3}$$

$$\nabla \times \mathbf{B} - \varepsilon_r \varepsilon_0 \mu_r \mu_0 \frac{\partial \mathbf{E}}{\partial t} = \mu_r \mu_0 \mathbf{J}_m \tag{4}$$

in which: $\mathbf{E}$ is the electric field (volts/meter).

$\rho_t = \rho_f + \rho_b$ is the total charge density (coulombs/cubic meter).

ρ_f is the free charge density, and ρ_b is the bound charge density.

$\vec{B}$ is the magnetic induction (teslas or webers/square meter).

$\mathbf{J}_m = \mathbf{J}_f + \partial \mathbf{P} / \partial t + \nabla \times \mathbf{M}$ is the total current density (amperes/square meter), consisting of $\mathbf{J}_f$, the current density of free charges; $\partial\mathbf{P}/\partial t$, the polarization current density; and $\nabla\times\mathbf{M}$, the magnetization current density.

$\mathbf{M}$ is the magnetization (amperes/meter).

ε_0 is the permittivity of free space (farads/meter).

ε_r is the relative permittivity, or dielectric constant.

μ_0 is the permeability of free space, which is arbitrarily set equal to $4\pi\times10^{-7}$ teslas meter/ampere (henrys/meter).

μ_r is the relative permeability.

In materials where Ohm's law applies, the free current density is related to the electric field by $\mathbf{J}_f = \sigma\mathbf{E}$, where σ is the conductivity (Siemens/meter).

Physically, the two fundamental vector fields are $\mathbf{E}$ and $\mathbf{B}$, because it is these fields that produce forces on electric charge according to the Lorentz force law $\mathbf{F} = q(\mathbf{E}+\mathbf{v}\times\mathbf{B})$. It is customary, however, to write Maxwell's equations in terms of the fields $\mathbf{E}$ and $\mathbf{H}$, where $\mathbf{H}$, the "magnetic field" or "magnetic intensity" (amperes/meter), is proportional to $\mathbf{B}$. The primary reason for doing this is that the quantity $\mathbf{E}\times\mathbf{H}$ is the energy flux density, which is of paramount importance is determining a system's electromagnetic behavior.

Media Parameters

It is necessary to specify the electrical and magnetic properties of the medium supporting the electromagnetic field in order to solve the field equations, subject to appropriate boundary conditions. Without this information, Maxwell's equations do not form a determinate system.

The EM behavior of any medium is determined by its material properties, the electric permittivity and magnetic permeability. There are other material parameters, but they are secondary in nature. For example, in linear media the electric susceptibility is the proportionality factor between the applied electric field and induced polarization. Similarly, the magnetic susceptibility relates magnetization to the applied magnetic field. Each of these parameters is included indirectly in the relative permittivity and permeability, respectively, and therefore they have no independent significance.

Various classes of media exit. The simplest is *free space*, in which the relative permittivity and permeability are both unity. The EM properties of *isotropic* media are independent of the direction of the EM fields. Media in which the electrical properties are the same everywhere are *spatially homogeneous*. In some media, the EM properties vary in time, and such media are described as *temporally inhomogeneous*. The most complex media are *anisotropic*. The EM properties are direction-dependent, and may also vary spatially or temporally. The permittivity and permeability of anisotropic media are described by tensors instead of real or, in the case of lossy media, complex scalars.

Wave Velocity and Refractive Index

Time-varying fields in the form of propagating electromagnetic waves follow directly from Maxwell's equations. The equations are combined to obtain a "wave equation" whose solutions are coupled electric and magnetic field that move through the medium carrying energy as they do.

In a lossless medium, ε_r and μ_r are real numbers. A monochromatic EM wave's phase velocity, the velocity at which the wavefront (surface of constant phase) propagates, is completely determined by these two material parameters. The propagation velocity is

$$v = \frac{1}{\sqrt{\varepsilon_r \varepsilon_0 \mu_r \mu_0}} = \frac{c}{\sqrt{\varepsilon_r \mu_r}}, \qquad (5)$$

where c is the velocity of light in free space, or simply "speed of light" (very nearly $3\text{x}10^8$ meters/second).

In lossless "normal" media, ε_r and μ_r are positive real numbers greater than 1 (this is not true in a metamaterial). Waves in normal media always propagate slower than the speed of light. The velocity ratio is the medium's *refractive index*, given by

$$n = \frac{c}{v} = \sqrt{\varepsilon_r \mu_r} \; . \qquad (6)$$

When n is real, EM waves can propagate in the medium. If n is purely imaginary at some frequency, then no wave mode (field configuration) can propagate at that frequency.

In lossless anisotropic or inhomogeneous media, ε_r and μ_r vary with propagation direction or position, and are represented mathematically as matrices of positive real numbers. In lossy materials ε_r and μ_r are complex numbers with non-zero imaginary parts that result in energy dissipation as the wave propagates through the medium. In such media the refractive index varies with direction or position or both.

n^2 may be negative (n is imaginary) in some naturally occurring media. At frequencies below its plasma frequency, for example, a gaseous plasma has a negative dielectric constant but positive permeability, resulting in $n^2 < 0$. This situation occurs in the earth's ionosphere, and provides an example of unusual wave propagation resulting from the negative permittivity. Instead of being refracted, as it would be if n^2 were positive, below the plasma frequency an incident radio wave is instead reflected. This is the mechanism that supports long distance high-frequency (HF) radio communication.

In normal media, both ε_r and μ_r are frequency-dependent, often strongly so. For example, the dielectric constant of water decreases from 81 at DC and very low frequencies, to 78 at 1 Mhz, 34 at 10 Ghz, and about 2 at optical frequencies. The refractive index varies correspondingly from 9 at DC to 1.4 in the optical region. In magnetic materials (water is not one), the frequency variation of relative permeability can be several orders of magnitude.

The effective ε_r and μ_r of metamaterials are also are frequency-dependent (in fact they must be), and they change sign just as the dielectric constant does in a plasma. As a result, in certain frequency ranges the metamaterial's novel EM properties disappear, and there is a strong analogy between the behavior of a metamaterial and naturally occurring plasma. The variation of permittivity and permeability with frequency may be very important in practical antenna applications, particularly in view of the wide spectrum used by modern communication systems. Metamaterials may be quite useful in the microwave range, but less so at lower frequencies.

Right and Left Handed Media and Power Flow

Another important parameter describing the propagation of an EM wave is its *wave vector,* $\mathbf{k}$, whose magnitude is the *wavenumber*. $\mathbf{k}$ is in the direction of propagation and perpendicular to the wavefront. In normal media, not in a metamaterial, it also coincides with the direction of energy flow.

The wavefront consists of electric and magnetic fields which are orthogonal to each other, and to the direction of propagation. The triad of vectors $\mathbf{E}$, $\mathbf{H}$, and $\mathbf{k}$ constitute a mutually orthogonal set related by the "right hand rule". When the vector

$\mathbf{E}$ is rotated into the direction of $\mathbf{H}$ through the included angle ϕ in the plane containing $\mathbf{E}$ and $\mathbf{H}$, curling one's fingers in the direction of rotation with the thumb extended perpendicular to the fingers determines the direction of the wave vector $\mathbf{k}$ (hence, right "hand" rule). The *handedness* of a medium is determined by the vector triplet $\mathbf{E}$, $\mathbf{H}$, and $\mathbf{k}$. The right hand rule relationship holds in normal media, which are consequently described as *right handed*. A metamaterial, by contrast, is *left handed* as discussed below.

The direction of energy flow in the wavefront is $\mathbf{E} \times \mathbf{H}$, which for a plane wave in normal media corresponds to the direction of the wave vector $\mathbf{k}$. The vector $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ is known as the *Poynting vector*, in units of watts/square meter. Integrating $\mathbf{S}$ over a closed surface gives the amount of energy per unit time (joules/second) flowing out of that surface.

Theoretical Basis of Metamaterials

Veselago's Model: In a seminal paper published in the late 1960s, the Russian physicist V. G. Veselago theoretically investigated the effects on EM fields of negative values of relative permittivity and permeability [3]. Because the refractive index is equal to the square root of their product, changing the signs of *both* ε_r and μ_r produces no change in the refractive index. Veselago suggested three possible interpretations of what simultaneously negative values of ε_r and μ_r might mean:

(1) The electromagnetic properties of any material are independent of whether or not ε_r and μ_r are both positive or both negative.

(2) Having simultaneously negative values for ε_r and μ_r is inconsistent with the laws of physics, and no such material can ever exist.

(3) Materials with negative permittivity and permeability may exist and possess unusual electromagnetic properties.

Veselago acknowledged that, at the time of his paper, there were no materials, naturally occurring or artificially made, with $\varepsilon_r \langle 0$ and $\mu_r \langle 0$. He nevertheless proceeded to analyze the properties of such media in the event that they might be discovered or man-made.

In the case of a monochromatic plane wave described by the propagation factor $e^{-i(kz-\omega t)}$, Maxwell's equations lead to the following relationships:

$$\mathbf{k} \times \mathbf{E} = \frac{\omega}{c} \mu_r \mu_o \mathbf{H} \tag{7}$$

$$\mathbf{k} \times \mathbf{H} = -\frac{\omega}{c} \varepsilon_r \varepsilon_o \mathbf{E} \tag{8}$$

If $\varepsilon_r > 0$ and $\mu_r > 0$, then the triad $\mathbf{E}$, $\mathbf{H}$, and $\mathbf{k}$ are related by the right hand rule. The medium is therefore *normal* or *right-handed* (RH). If, instead, $\varepsilon_r < 0$ and $\mu_r < 0$, the sign of the right hand side of equations (7) and (8) changes. As a result, the vector triplet is now left-handed (curling the fingers of the *left* hand). The direction of the wave vector $\mathbf{k}$ therefore is reversed when the material parameters are both negative. A metamaterial with negative permittivity and permeability is consequently called a *left- handed* (LH) material (LHM).

Power Flow in LHM: Compared to its orientation in RHM, the wave vector $\mathbf{k}$ in LHM points in the opposite direction relative to $\mathbf{E}$ and $\mathbf{H}$. An immediate consequence is that the Poynting vector and wave vector are in opposite directions. Because the Poynting vector $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ always forms a right-handed triad with $\mathbf{E}$ and $\mathbf{H}$, regardless of the sign of the permittivity and permeability, allowing $\varepsilon_r < 0$ and $\mu_r < 0$ changes the direction of $\mathbf{k}$ but not of $\mathbf{S}$.

The wave propagates in one direction, but the power flow is in the opposite direction. Veselago points out that this situation corresponds to a negative group velocity ($V_{group} = \frac{\partial \omega}{\partial k}$, computed at the average value of the wavenumber). While perhaps somewhat surprising, there are examples of just this effect. Some naturally occurring anisotropic media and certain media in which spatial dispersion exists exhibit negative group velocity.

Doppler Effect and Cherenkov Radiation in LHM: Other unusual effects occur in LHM. The Doppler effect is reversed. The measured frequency of an approaching emitter will be lower than the stationary frequency, not higher as in RHM. Cherenkov radiation, electromagnetic radiation emitted by a charged particle moving at high speed in a LHM, will be directed backward relative to the direction of motion, which is the opposite of what happens in RHM.

Reflection and Refraction at a Plane Interface: EM plane waves incident upon a plane material surface are *refracted* as they enter the medium. For waves propagating between normal media, the wave bends towards the normal upon entering a denser medium (higher refractive index), and away upon entering a less dense medium. Figure 10 shows this effect in which the upper medium is less dense than the lower one. Φ is the incidence angle, and Ψ the refraction angle. If the wave propagates from a RHM into a denser LHM, the refracted wave is bent towards the normal, but on the opposite side of the normal as shown in Figure 2. Interestingly, the direction of the reflected wave is the same (reflection angle equals the incidence angle) regardless of the nature of the lower medium.

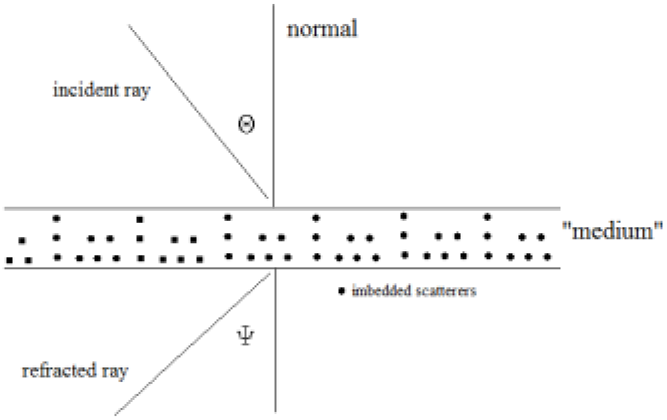

Figure 10. Refraction at MTM interface

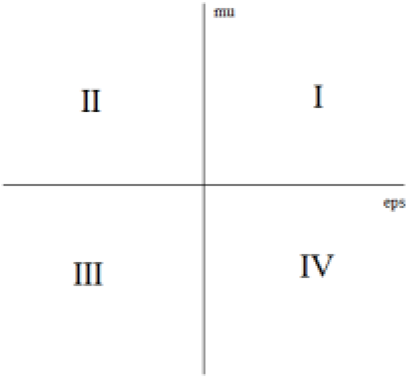

Figure 11. Media classification

Media Classification: In order to classify media based on permittivity and permeability, Veselago introduced the $\varepsilon - \mu$ diagram shown in Figure 11. The plane is divided into the four standard quadrants, labeled I through IV proceeding counter-clockwise from the upper right-most quadrant. Considering only isotropic materials, normal dielectrics with $\varepsilon_r > 0$ and $\mu_r > 0$ are contained in quadrant I. Quadrant II contains media with $\varepsilon_r < 0$ and $\mu_r > 0$, which is characteristic of gaseous and solid-state plasmas. The dielectric constant of a gaseous plasma is given by $\varepsilon = 1 - \sum \frac{\omega_o^2}{\omega^2}$, where ω_0 is the plasma frequency, and the summation is over all species in the plasma. At sufficiently low frequencies, the second term in this expression becomes greater than 1, which leads to a negative permittivity. Quadrants II and IV, however, are unoccupied, at least by isotropic materials, because there is no known naturally occurring material with negative permeability.

The $\varepsilon - \mu$ diagram can be used only for scalar values of permittivity and permeability. Certain anisotropic materials with tensor permittivity and permeability fall into quadrants III and IV for propagation in specific directions within the medium, because in those directions ε_r and μ_r are scalars. An example is propagation along the magnetic field direction in a gyrotropic medium, such as a plasma in a magnetic field. In the propagation direction ε_r and μ_r are scalars, while generally in the medium they are second-rank tensors (3x3 matrices).

Requirement of Dispersion: A non-dispersive LHM cannot exist. This result follows from the fact that the energy density of an electromagnetic field is given by $W = \varepsilon_r \varepsilon_o \mathbf{E} \cdot \mathbf{E} + \mu_r \mu_o \mathbf{H} \cdot \mathbf{H}$. If both ε_r and μ_r are negative, then the total energy is negative, which is not possible.

In order to assure positive total energy, the medium must be *dispersive*. In a dispersive material, the expression for energy density is different, and includes the variation of permittivity and permeability with frequency:

$$W = \varepsilon_o \frac{\partial(\varepsilon_r \omega)}{\partial \omega} E^2 + \mu_o \frac{\partial(\mu_r \omega)}{\partial \omega} H^2 . \tag{9}$$

W is always positive if $\frac{\partial(\varepsilon_r \omega)}{\partial \omega} > 0$ and $\frac{\partial(\mu_r \omega)}{\partial \omega} > 0$. In general, therefore, the permittivity and permeability must be frequency-dependent in LHM. Media in which both ε_r and μ_r are negative and constant with respect to frequency cannot exist. It does appear possible to have only one parameter depend on frequency, as long as its derivative is large enough to avoid a negative total energy; but Veselago did not consider this possibility.

The same conclusion about the necessity of dispersion can also be reached by considering the limiting case of a parallel plate capacitor at DC. Neglecting fringing, the capacitance is $C = \frac{\varepsilon_r \varepsilon_o A}{d}$, where A is the plate area, and d is the plate separation in consistent units. Physically, capacitance is the ratio of free charge on the plates, Q, to the potential difference (voltage) between them, V, taken as positive quantities. Thus, $C = \frac{Q}{V}$, must be positive at DC. If a metamaterial were used as a dielectric between the capacitor plates, its permittivity therefore must be positive in the limit of zero frequency (DC). This is necessary to avoid the physically impossible case of a negative capacitance. But, by definition, ε_r in a metamaterial must be negative at some frequency in order for the medium to be a metamaterial. Consequently, the metamaterial's relative permittivity must change sign with frequency, from a positive value at DC to a negative value at some non-zero frequency. This is dispersion. Note that this brief discussion does not address how the capacitor behaves as a circuit element at frequencies where $\varepsilon_r < 0$. Doing so should be quite interesting.

Realization of Metamaterials

Because metamaterials do not occur naturally, the only way to obtain a metamaterial is to manufacture it. The very nomenclature, *metamaterial*, conjures up the image of a "substance", generally a solid material with some specific chemical composition. Existing metamaterials, however, could not be further from that image.

Smith [8] recently described a LHM that was modeled theoretically and measured experimentally. Building upon the split ring resonator (SRR) medium created by Pendry [9], short, small-diameter wires were added to a an array of SRRs to create the left-handed metamaterial. The medium is actually an array of discrete electromagnetic scatterers arranged so that the *effective* ε_r and μ_r are negative at some frequencies. It is not at all a "substance" in the usual sense.

Figure 10 represents this concept schematically. Imagine two parallel planes in space, between which an array of scattering elements is deployed. The scatterers, represented by dots in the figure, might be short wire segments, other geometrical shapes, SRRs, other resonant structures, or any number of EM elements fabricated from conductors and/or dielectrics.

In the realm of every possible configuration, it seems intuitively plausible that some arrangements will result in anomalous propagation such as seen in metamaterials. In Figure 10, for example, it is not difficult to imagine that there is some scatterer arrangement that will "refract" a wavefront incident at the angle Φ in the direction opposite to "normal" refraction as shown. This type of propagation is a characteristic of a left-handed material. If the volume between the planes is thought of as a "medium", it would be properly described as one with negative effective permittivity and permeability, that is, a metamaterial.

Metamaterials that are "substances" may eventually be developed or discovered. But current research focuses on creating LHMs using imbedded scatterers. This is not a limitation, however, and indeed may actually provide more flexibility than attempting to create purely chemical substances. Designing the scattering matrix may permit engineering metamaterials whose properties could not exist in chemical substances. Creating solid materials consisting of a matrix containing small conductive scatterers, possibly even carbon nanotube scatterers, might permit the designer to actually tailor the LHM properties to the EM application.

The idea of using a matrix of scatterers to create a "medium" with a negative effective permittivity is not new. In the 1950s, Bracewell [10] and Rotman [11] experimentally investigated ionospheric propagation using a three-dimensional array of intersecting thin, straight wires. Dispersion in this “medium” is similar to that in a neutral plasma, such as the ionosphere.

Just as the real ionosphere does, the artificial one has $\varepsilon_r < 0$ for frequencies below the “plasma frequency”. However, unlike the ionosphere’s plasma frequency, which depends on the charged particle density, and particle charge and mass, the artificial medium’s “plasma frequency” depends on the arrangement and dimensions of the wires. Because LHM must exhibit $\varepsilon_r < 0$ at some frequencies, and because the dispersion relation has the same form as a gaseous plasma’s, metamaterials made from scatterers are often described in terms of a *plasma frequency*. It is important to remember that the term used in the manner is a purely formal construct, not to be confused with the parameter of the same name describing a gaseous plasma.

The effective permittivity of Smith’s wire medium is given by

$$\varepsilon_{eff}(\omega) = 1 - \frac{\omega_p^2}{\omega^2} \tag{10}$$

where ω_p is that medium’s plasma frequency. ω_p depends in a complicated way on the wire scatterers’ electrical properties and geometrical arrangement. At frequencies below ω_p, ε_{eff} is negative, which is one of the requirements for LHM. The other requirement is that μ_r simultaneously be negative. The SRR medium has an effective permeability given by

$$\mu_{eff}(\omega) = 1 - \frac{F\omega_o^2}{\omega^2 - \omega_o^2 - i\omega\Gamma} \tag{11}$$

where F is the fractional area of the unit cell occupied by the SRR’s interior, and Γ is the SRR dissipation factor. ω_o is the SRR medium’s plasma frequency. Equations (10) and (11) assume that the array of wires and the array of SRRs do not interact appreciably, which was true in Smith’s medium.

When the scatterer arrays are uncoupled, the wire array produces the medium’s ε_{eff}, and the SRR array produces its μ_{eff}. Under these conditions, the effective refractive index of the composite medium is $n_{eff}(\omega) = \sqrt{\varepsilon_{eff}(\omega)\mu_{eff}(\omega)}$. If there is an overlapping band of frequencies in which the effective permittivity and permeability are both negative, then the effective index of refraction is real, and propagating modes exist. The medium is LH, however, because of the negative material parameters, so that the unusual propagation effects associated with LHM occur.

The SRR, wire segment, wire-SRR media were modeled using a commercial EM software package to generate dispersion curves (the problem is far too complex for an analytical solution). The SRR array alone showed a pronounced band gap from 4.2-4.6 Ghz. No modes would propagate in this band. The wire array alone exhibited an even wider band gap, from just above DC to 13 GHz, the array's plasma frequency. No modes would propagate in this range. The reason propagation cannot occur is that μ_{eff} and ε_{eff}, respectively, are separately negative in these bands, which results in a purely imaginary refractive index.

A metamaterial is created by combining the wire and SRR arrays. The wire-SRR array has overlapping bands from 4.2-4.6 GHz where the simultaneously negative material parameters combine to create a metamaterial. Computer modeling of the dispersion characteristics of the combined arrays clearly showed propagation in the otherwise forbidden band. Experimental measurements further confirmed the LH nature of the medium by observing propagation between 4.2 and 4.6 GHz which was not possible in either array alone.

Analysis of Metamaterials

In calculating a metamaterial's performance, it is unlikely that analytical results are attainable for all but the simplest scatterer geometries. Analytically solving Maxwell's equations is almost certainly impractical. Computer modeling is likely to be the best approach, and possibly the only truly useful one. Metamaterials made of imbedded arrays of scattering elements can, in fact, be analyzed with existing modeling software, at least to some degree. Commercially available packages, and government-developed programs such as NEC-4, provide a starting point for modeling metamaterials. But many questions arise.

Most EM software approximates the modeled structure using some form of discretization. In NEC-4, for example, continuous wires are approximated by short segments. Continuous surfaces are approximated by patches of surface area. Other programs use the same or similar approaches.

One characteristic of the discretized models is that very substantial computer resources (memory and speed) are often required to obtain good results. In NEC-4, for example, the model size (a measure of required memory) is proportional to the square of the number of segments, and the computation time (a measure of required processor speed) scales similarly. Other types of models exhibit similar characteristics. Disproportionately greater resources are required with increasing complexity, modeling accuracy, and resolution in the frequency domain. It may be expected that these issues will be exacerbated for metamaterials, because LHMs fabricated from large numbers of complex individual scatterers are indeed very complex structures. This level of complexity is likely to push existing modeling capabilities to the limit.

Metamaterials should be studied experimentally, not only to characterize available materials, but also to develop the experimental techniques and fixtures needed for such studies. Because of the anomalous propagation, it is likely that new measurement techniques will have to be developed to fully explore metamaterial properties. For example, direct measurement of the negative values of permittivity and permeability, key parameters in characterizing a metamaterial, is required. But it is not clear that suitable instrumentation or techniques for doing so exist.

Future of MTM

MTM hold great promise as components in antenna design, particularly at microwave frequencies. Whether or not they can be used effectively a low VHF frequencies is an open question because there simply is not enough data to reach a conclusion. At this time not enough is known about MTM's electromagnetic behavior. There are no "design handbooks" and essentially no low frequency data. Nevertheless, the performance of MTM-based antennas at microwave frequencies strongly suggests hat MTM antennas will be useful at VHF as well.

2.7.7 References

[1] Markos, P., and Soukoulis, C. M., "Numerical studies of left-handed materials and arrays of split ring resonators," *Physical Review E*, **65**, p. 036622-1 (2002). DOI: 10.1103/PhysRevE.65.036622.

[2] Shapiro, M. A., Shvets, G., Sirigiri, J. R., and Temkin, R. J., "Spatial dispersion in metamaterials with negative dielectric permittivity and its effect on surface waves," *Optics Letters*, **31**, 13, p. 2051 (2006).

[3] Veselago, V. G., "The electrodynamics of substances with simultaneously negative values of ε and μ," *Soviet Physics Uspekhi*, **10**, 4, Jan.-Feb. 1968

[4] Ikonen, P., "Electrically small metamaterial-based antennas – have we seen any real practical benefits?" arXiv:0902.1820v1 [cond-mat.mtrl-sci] (2009).

[5] Antoniades, M. A., and Eleftheriades, G., V., "A compact broadband NRI-TL metamaterial monopole antenna," *13th Annual International Symposium on Antenna Technology and Applied Electromagnetics and the Canadian Sciences Meeting* (2009).

[6] Kim, J-H., Kim, H-Y., Lee, H-M., “Design of efficient electrically small dual band monopole antenna using negative permittivity metamaterial structures,” *IEEE Antennas and Propagation Society International Symposium, 2008,* San Diego, CA, pp.1-4. DOI: 10.1109/APS.2008.4619487

[7] Slyusar, V. I., “Metamaterials on antenna solutions,” *International Conference on Antenna Theory and Techniques*, Lviv, Ukraine, pp. 19-24, 6-9 October 2009.

[8] Smith, D. R., Padilla, W. J., Vier, D. C., Nemat-Nasser, S. C., Schultz, S., “Composite medium with simultaneously negative permeability and permittivity,” *Phys. Rev. Lett.*, **84**, 18, 1 May 2000

[9] Pendry, J. B., Holden, A. J., Robbins, D. J., Stewart, J., “Extremely low frequency plasmons in metallic microstructures,” *Phys. Rev. Lett.*, **76**, 4773 (1996).

[10] Bracewell, R. N., “Analogues of an ionized medium, applications to the ionosphere,” *Wireless Engineer*, **31**, p. 320 (1954).

[11] Rotman, W., “Plasma simulation by artificial dielectrics and parallel-plate media,” *IRE Trans. Ant. Prop*., **AP10**, p. 82 (1962).

2.8 Electronic Band-Gap and High Impedance Surfaces

Introduction

Electromagnetic band-gap (EBG) structures recently have attracted a great deal of interest among researchers because EBG structures can significantly reduce surface waves, thereby improving patch antenna performance [1]. Reflection phase characteristics also can have a pronounced negative effect on antenna performance [2,3]. It is well known that when a dipole antenna is operated near a metallic ground plane as shown in Figure 1, the radiated waves are reflected from the ground plane, so that interference may occur. Antenna bandwidth and gain are degraded, which is readily demonstrated by applying image theory.

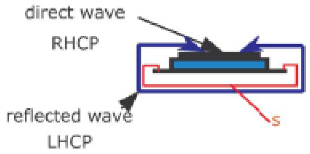

Figure 1. Antenna Near Ground Plane

To overcome these difficulties, dipoles are operated at a height of 0.2% or higher above the ground plane. In [1] the concept and feasibility of developing wideband, ultra-thin (substrate thickness < O.1) directional dipoles and dipole arrays using EBG structures are addressed. The reflection phase angle of an EBG surface varies from -180 to 180 degrees with frequency. The goal is to make the antenna operating frequency band coincide with the EBG stop band, and then to flatten the reflection phase curve within the stop-band region, which then will ensure wide-band operation of the antenna. This technology could be useful in an integral television antenna as well as set-top antennas that lay flat on the television.

The parameters generally adopted to characterize the performance of EBG structures include the surface wave attenuation and reflectivity. A double benefit can be derived from the use of artificial magnetic conductor type of EBGs because not only are surface waves suppressed, but also in-phase image currents are induced. The so-called reflection phase—frequency diagram gives information on how the structure reacts to a wave impinging on it. Normally, the configuration behaves similar to a metal plate (perfect electric conductor, “PEC”). However, another characteristic feature of EBGs is the existence of a certain frequency range over which an incident electromagnetic wave does not experience *any* phase reversal upon reflection (a rather remarkable phenomenon). In this range the structure behaves as a perfect magnetic conductor (“PMC”). The frequency range where the phase reversal is within from the PMC point (phase reversal) is conventionally used to define the bandwidth and is called the “bandgap.” The concept here is to introduce a ground plane for shielding purposes (reducing the contributions of LHCP signals), and cover this plate with EBG cells (cancelling out the surface currents)

Types of EGB-Cells

Various EBG structures from literature have been investigated. A small sample of different geometries is displayed in Figure 2. These EBG cells can be with or without vias to a ground plane on the opposite side of the substrate. A Mushroom EBG substrate is shown in figure 3. The EBG cells in this figure are shown with vias (each cell is shorted to the ground plane).

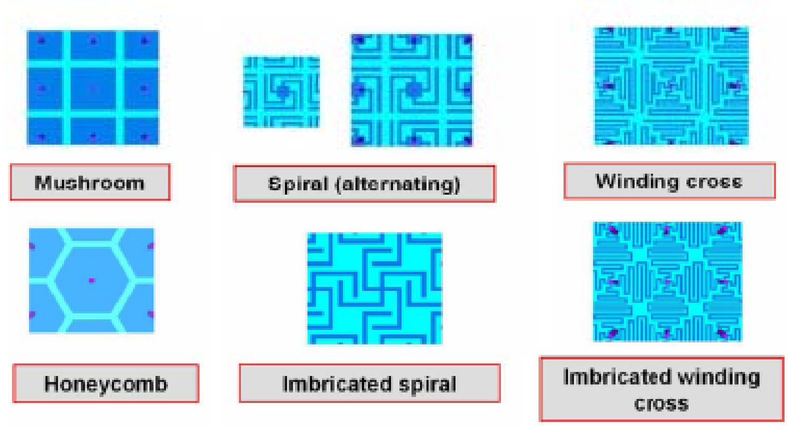

Figure 2. Different Types of EBG Structures.

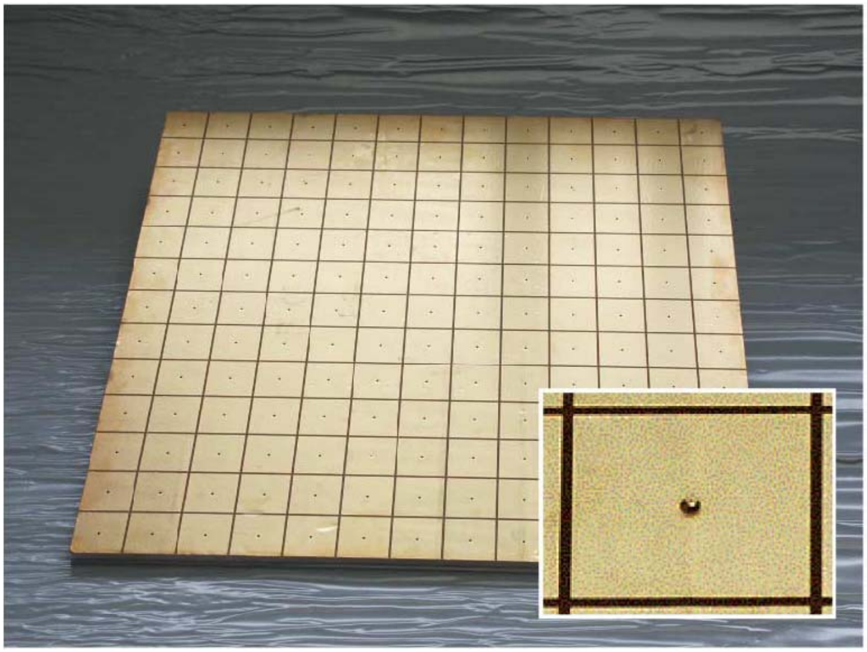

Figure 3 Mushroom EBG substrate (inset: Zoom on a single cell and a golden nail)

Performance of a Dipole Antenna on a Constant Reflection Phase EBG Surface

For a dipole antenna operating against a ground plane, the driving-point impedance is the algebraic sum of the self and mutual impedances when image theory is applied. For a PEC ground plane the driving point impedance is the self minus the mutual impedance since the reflection phase is 180 degrees. The driving point impedance of a dipole on an EBG structure with any other constant reflection phase characteristics can also be calculated. For example, for a thin wire dipole the self and mutual impedances can be computed using the induced EMF theory. Of course, it may not be possible to achieve a constant reflection phase over a band of frequencies using an EBG structure, but that is the goal. As an example, the dipole under consideration is infinitely thin and resonant at 2.9 GHz in free-space with a 2:1 VSWR bandwidth of *5%.* The driving point resistance is negative (at low frequencies) for phase angles of -100 degrees to -160 degrees because the mutual resistance is larger than the self resistance and adds destructively with it. But for phase angles of 40 degrees to 110 degrees the mutual resistance adds constructively with the self-resistance. Thus, the driving point reactance for different phase angles is capacitive at low frequencies and inductive at high frequencies.

An EBG surface can be characterized by its reflection phase angle given by

$$\phi(f) = \mathrm{Im}\left[\ln\left(\frac{af^2 + bf - c}{af^2 + bf + c}\right)\right]$$

where a, b and c are the EBG design parameters with $a=4\pi^2\eta LC$, $b=2\pi L$ and $c=\eta$ [4]**.** Here ***L C*** and η represent the inductance, capacitance and intrinsic impedance of free-space, respectively. The slope of the reflection phase curve can be changed by varying the values of a and b*.* The bandwidth of this dipole antenna has been enhanced by flattening out the reflection phase curve at the useful phase angle regions, which corresponds to the EBG surface of having a stop band center frequency at 2.82 GHz.

EBG Dipole

Work on an EBG Dipole was completed in the past by MegaWave Corp. Figure 4 shows an EBG structure modeled in XFDTD. It consists of a double layer, mushroom-like structure embedded in FR-4 substrate with the bottom layer consisting of an array of 6 by 12 square patches and the top layer of an array of 5 by 10 square patches. The overall dimensions are 2”x4” with a thickness of 0.18”. The computed reflected phase of the EBG is plotted in Figure 5. The reflected phase was generated by comparing the phase of a plane wave scattered from the EBG surface with that from a PEC surface. The bandwidth of the EBG material was estimated to be from 1.66 to 2.35GHz as determined by the reflection phase being in the range of 90°±45°[2].

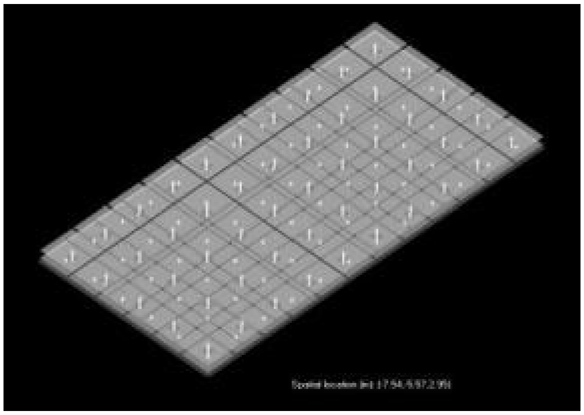

Figure 4: XFDTD Model of Dual Layer EBG Material

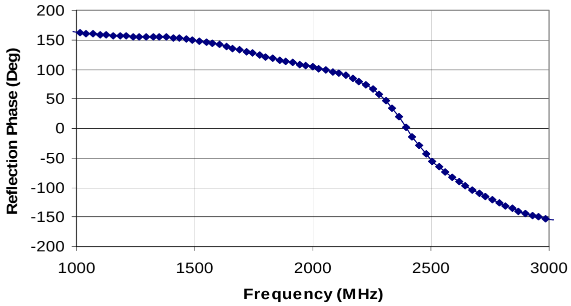

Figure 5: Computed Reflection Phase of EBG Material

A prototype of the MegaWave-designed EBG structure was fabricated as shown in Figure 6, and Figure 7 shows a wideband dipole placed 0.25” away from the EBG structure. The measured VSWR was less than 2.6:1 across a very wide operating frequency band from 1.35 to 2.7GHz as shown in Figure 8.

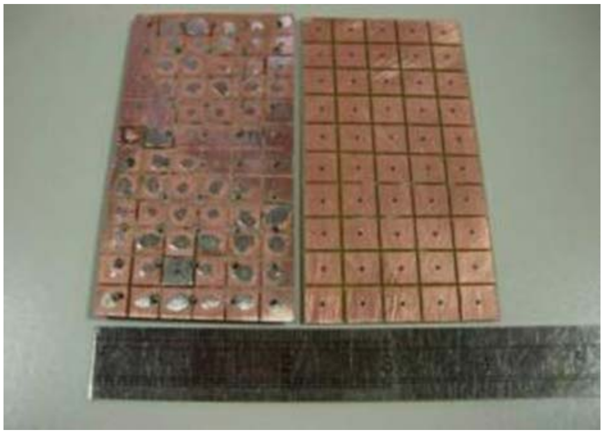

Figure 6: Lab Prototype of EBG Structure (Top and Bottom Layers Shown)

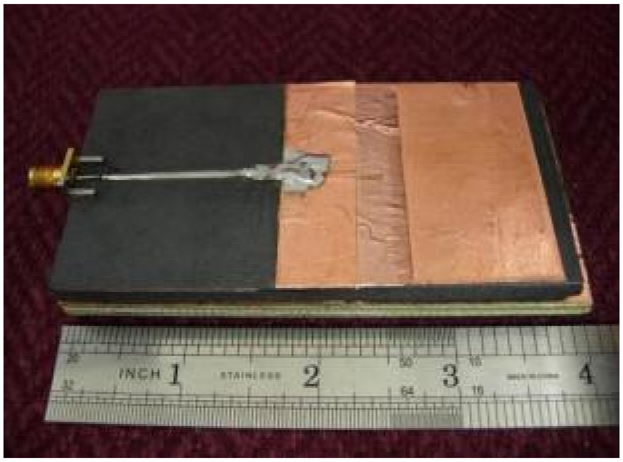
INCH 1
STAINLESS
2
3
4

Figure 7: Lab Prototype of Wideband Dipole on EBG Material

The free space radiation patterns of both the wideband dipole on the EBG structure and the Phase I directional dipole antenna were measured and compared in MegaWave's laboratory. Figures 9 and 10 show the measured free-space patterns. This EBG technology could be implemented in the television UHF and VHF bands with an expectation of greatly improved set-top antenna performance.

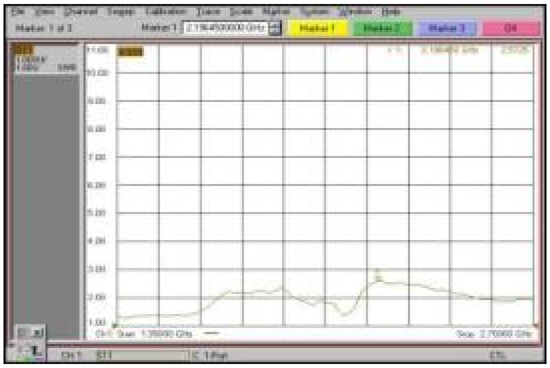
Marker 1
Marker 2
Marker 3
11.00
10.00
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
Start 1.35000 GHz
Stop 2.75000 GHz

Figure 8: VSWR of Wideband Dipole on EBG Material

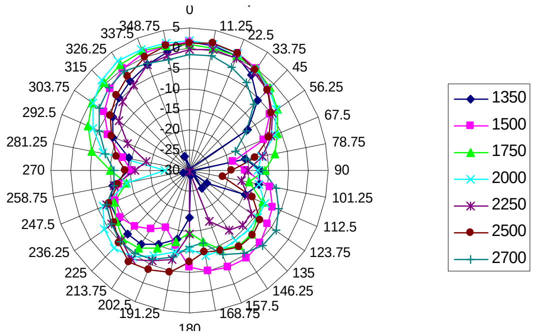

Figure 9: Measured Radiation Patterns of Antenna

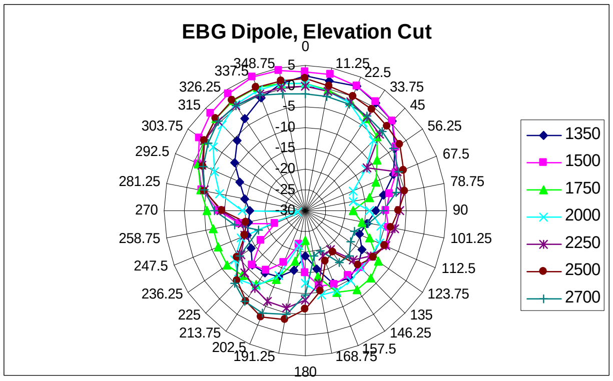

Figure 10: Measured Radiation Patterns of Wideband Dipole on EBG

References

[1] R. Baggen, M. Martinez-Vazquez, J. Leiss, S. Holzwarth, L. S. Drioli and P. D.Maagt, “Low Profile GALILEO Antenna Using EBG Technology”,IEEE, Antennas and Prop.,Vol.56,Num. 3, March 2008, pp 667-673

[2] Fan Yang and Yahya Rahmat-Samii, “Reflection Phase Characterization of the EBG Ground Plane for Low Profile Wire Antenna Applications”, IEEE Transactions on Antennas and Propagation, Vol.51, No.10, October 2003.

2.9 Fractal and Self-Similar Antennas

There is no question that, among all the antennas existing today, one of the most interesting and visually appealing is the Fractal antenna. However it not clear exactly what benefits in bandwidth and gain they provide as compared to non-fractical elements of the same shape and area, but without the fine detailed geometries. The reader is directed to the section on pathological antennas (1.5) and its subsection 1.5.8 on fractals in Dr. R.C. Hansen's recent textbook, "Electrically Small, Superdirective, and Superconducting Antennas," Wiley Interscience, 2006.

First publications about electrodynamic characteristics of fractal structures were published in the 1980's, but first practical approach appeared 10 years later. Dr. Nathan Cohen, a professor at Boston University, designed, engineered and patented many fractal antenna designs, some of which are shown in Figure 1.

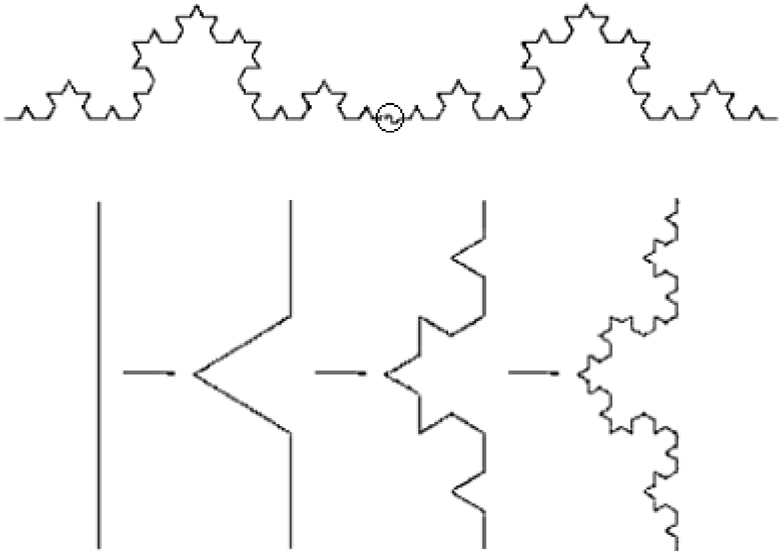

Figure 1. Fractal Designs

This fractal figure is built as follows: first line (length is z) is cut in 3 equal pieces z/3. In the middle the triangle is formed with same side lengths z/3 and same angles. This creates a single element template. This process is repeated with other segments in which sizes diminish 3 times (z/9), then again 3 times (z/27), and so on. A fractal antenna is created using fractal geometry, a self-similar pattern built from the repetition of a simple shape.

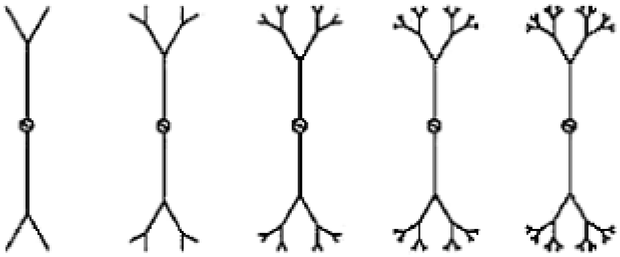

Figure 2. Fractal Dipole

The Fractal dipole as shown in Figure 2, is a recursive tree. Each of the new iterations multiplies the branches by two, in the process lowering the resonance frequency. By using many iterations, it is possible to use an antenna at lower frequencies without increasing the overall antenna size. Dipole antennas usually have narrow bandwidth, about 2.4% of the center frequency. If the 5th fractal iteration is used, then this parameter increases to 3.1%. If a 3D tree is used (when there are 4 branches used instead of 2) then this parameter increases 12.7%. Besides Dipole antennas there are resonant loop antennas used. Some typical loop fractal antennas built using Koch figures as shown in Figure 3 [7,8].

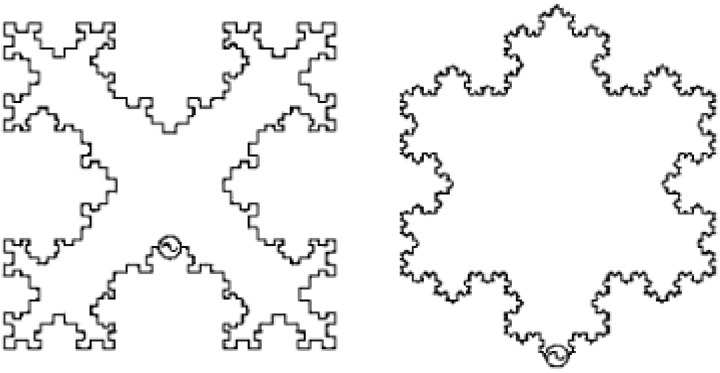

Figure 3. Fractal Loop Examples

RF MEMS could also be used in conjunction with fractal antenna structure as the basis of a new reconfigurable array antenna approach. The RF MEMS switches permit the connectivity of sections of the antenna's conductive parts, and therefore enhance the coupling between the fractal elements allowing multiple frequency operation with one antenna. Using this scheme a set-top antenna could be designed to change shape and/or size based upon which frequency (channel) the television was currently tuned to. The use of fractal shapes provides the multiband characteristic through the property of self similarity at equal or different physical scales. The electromagnetic performance of the RF MEMS switches is ideal for this application, and their placement is accomplished by small physical connections of the antenna's adjacent conducting pans [1].

Most fractal antennas, including the Sierpinski antenna, have been studied extensively over the last few years [3,4]. A modified Sierpinski antenna is shown in Figure 4. The centerline of the CPW [1] feed provides the RF input and dc voltage for MEMS switch actuation. Switches S1 and S1’ actuate at a low voltage, switches S2 and S2’ at a medium voltage, and switches S3 and S3’ at a high voltage. Figure 5 shows the different states of the antenna. An alternate implementation could be for a bowtie antenna as shown in Figure 7.

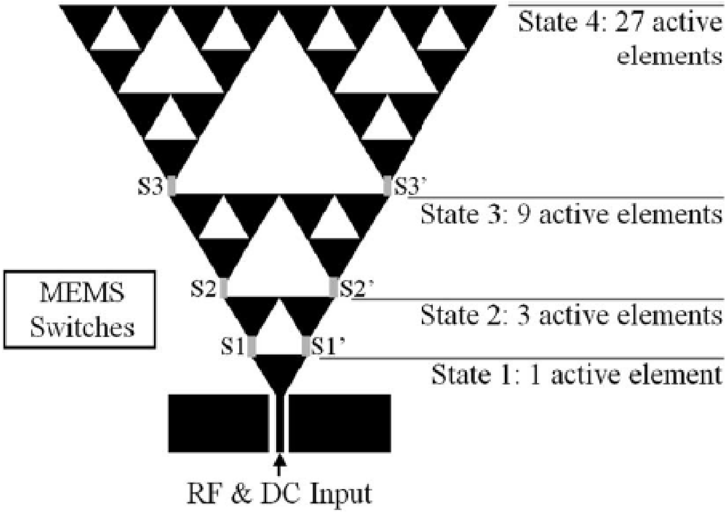
State 4: 27 active elements
S3
S3’
State 3: 9 active elements
MEMS Switches
S2
S2’
State 2: 3 active elements
S1
S1’
State 1: 1 active element
RF & DC Input

Figure 4. Illustration of a MEMS reconfigurable Sierpinski antenna.

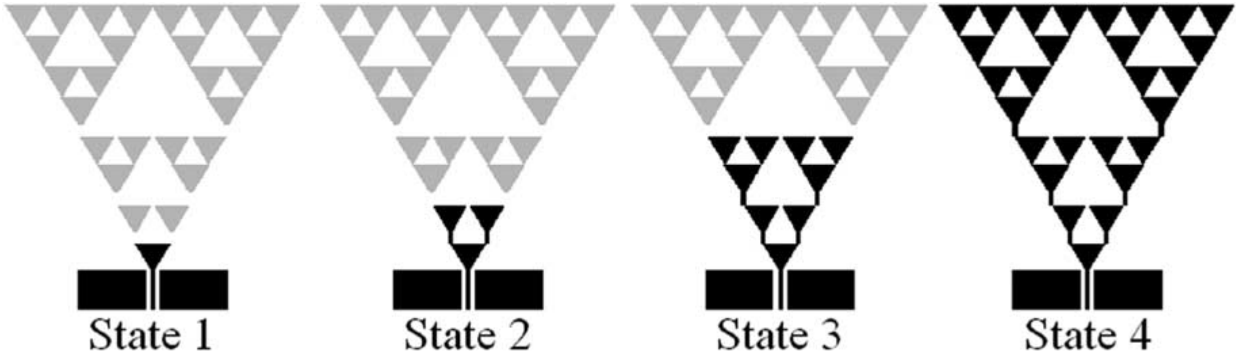
State 1
State 2
State 3
State 4

Figure 5. States of Reconfigurable Sierpinski Antenna

In most of the literature, integration has been accomplished on rigid and nonflexible semiconducting or organic polymer substrates such as silicon and FR-4. The idea of integrating RF MEMS switches into a multiband self-similar antenna was first implemented in [2], where the entire system, including the RF MEMS, the planar self-similar antenna, and the CPW–coplanar strip transition, was fabricated on silicon. The integration is achieved on a very thin and flexible liquid crystal polymer (LCP) substrate. Since the switches share a common dc feed, this technique provides reconfigurability without the need for additional bias lines. This is advantageous since dc bias lines take up space, add loss, and reduce the bandwidth of a device. This technology is particularly useful for antennas where bias lines can have a pronounced effect on radiation patterns, and may be especially useful in the set-top application.

The implementation of a sequentially activated antenna is shown in Figure 5. All of the MEMS switches used are single supported (cantilever type) and ohmic. Regardless of the applied voltage, the triangular element that is closest to the RF/dc input is always active. When no dc voltage is applied, the antenna radiates at its highest frequency. When a low dc voltage is applied to the signal line, the first set of MEMS switches (S1 and S1’) actuate, and this activates the second level of triangular elements. The antenna now radiates at a lower frequency. Since all of the switches are ohmic, the low voltage is also present at the membrane of the next set of switches (S2 and S2’). However, these switches are designed to actuate at a higher voltage so they are unaffected by the voltage present. When a higher dc voltage is applied, the first set of MEMS switches (S1 and S1’) remains in the “ON” position while the second set of switches (S2 and S2’) actuates. This activates the next iteration, consisting of six additional radiating elements. Again, this higher voltage is present at the next set of switch membranes (S3 and S3’), but the electrostatic force created is not sufficient for actuation. Finally, when the voltage is increased to its highest value, the first two sets of switches (S1 and S1’ and S2 and S2’) remain in the “ON” position while the remaining set of the switches (S3 and S3’) actuates .

Thus, the voltage “cascades” from one state to the next like a sequence of overflowing buckets. This technique could not be used if the switches were capacitive since they do not pass dc voltage. The four different states are illustrated in Figure 8, where all of the activated regions for different voltages are dark in color. This biasing technique allows for direct actuation of the electrostatic MEMS switches through the RF feed structure. Since only the RF feed is dc grounded, the switches actuate with the use of a floating ground. That is, the signal pin of the CPW feed is connected to the dc cathode, and the ground pins are corrected to the anode of a bias tee. The electrostatic charges that are created during switch actuation can dissipate through the substrate and be removed by the dc-grounded RF feed when in the “OFF” state. This method has been successfully documented in [5]. The reduction or elimination of bias lines is highly advantageous because they can significantly distort the radiation patterns and they can introduce additional unwanted resonances. To change the actuation voltage of a MEMS switch, there are four parameters that can be changed.

1) *Membrane material:* Switch membranes are almost always made of metal. This is due to their pliable nature. Stiffer metals (that is, those with a high Young’s Modulus*, E*) will have a higher actuation voltage than those with a lower *E*.
2) *Bridge thickness:* The thicker the bridge, the stiffer the membrane, resulting in a higher actuation voltage.

3) *Bridge height:* The higher the bridge, the larger the gap between the metal layers, resulting in decreased the electrostatic force and increased actuation voltage.
4) *Membrane geometry:* Springs can be designed into the shape of the membrane to lower the actuation voltage.

Of these parameters, only the fourth does not require any fabrication changes. Making changes to a fabrication process can be a costly endeavor and may create additional variables of concern.

LCP was chosen as the substrate for its many advantages. LCP is a thin (100 μm), flexible, low-loss (tan $\delta \approx 0.004$), low-moisture absorbing material with low permittivity ($\varepsilon r \approx 3$) [9]. Since the material is a polymer, there are additional packaging and cost advantages. All of these characteristics make it an ideal substrate for antennas, particularly at high frequencies. With respect to the geometry, the antenna elements have a 60◦ flare angle and maintain the resonant structure's self similarity with a log-periodicity of $\delta = 2$. The antenna is fed through a 6-mm-long CPW transmission line with a 50-μm gap, a 1.3-mm signal conductor width, and a 1.5-μm-thick aluminum layer. The overall size of the antenna, including the feed, is 20 × 25 mm. The CPW feed was chosen to facilitate the measurement setup. This reconfigurable antenna operates at four different principle frequencies. For each frequency, the antenna maintains its multiband performance. It was verified that the antenna has a different first resonant frequency for each of the four states. Since the antenna is self-similar with a log-periodicity of 2, each time the antenna transitions to the next state, the frequency should be halved. That is, the resonant frequency of state 2 should be half that of state 1. This is shown in Figure 6.

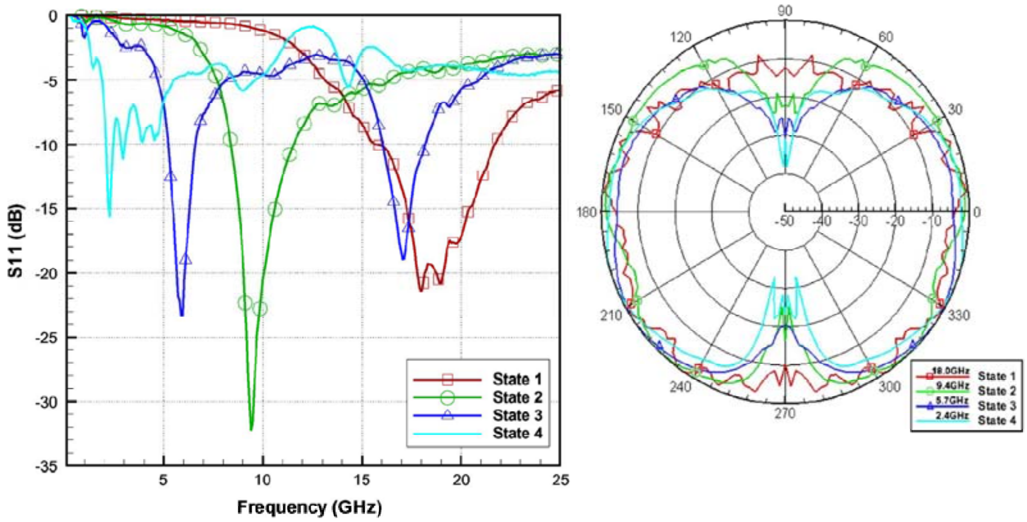

Figure 6. Measured Results of Built antenna for S11 and pattern

Fractal Antenna with RFID-MEM's

The general problem with using MEMS in a fractal antenna is how to power them. This was accomplished by putting DC on the RF of the feed of the antenna with a floating ground in the last section. If the goal is a dipole antenna as shown in Figure 7, then the floating ground would become a problem. One possible solution for this difficulty is known as "RFID-MEM's." The RFID could power the MEM to switch a section on and off. It would be controlled with a reader placed close to the antenna. If the television sends a digital signal back on the coaxial line that has information on what channel is currently tuned, then this information could be used by the reader to activate the correct switches to optimize the antenna for the best reception on the tuned channel. The RFID would need to operate either below the television spectrum, such as 13.56MHz, or above it, such as an UWB RFID. This system should have minimal effects on the antenna pattern of the bowtie antenna and could cover the entire VHF and UHF bands.

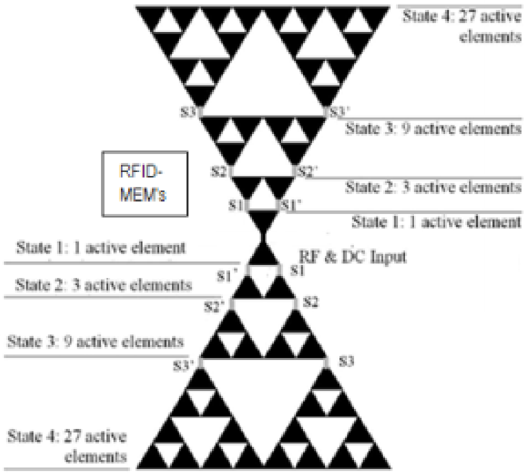
State 4: 27 active elements
S3
S3'
State 3: 9 active elements
RFID-MEM's
S2
S2'
State 2: 3 active elements
S1
S1'
State 1: 1 active element
State 1: 1 active element
RF & DC Input
State 2: 3 active elements
S1'
S1
S2'
S2
State 3: 9 active elements
S3'
S3
State 4: 27 active elements

Figure 7. Bowtie Antenna

Fractennas

There have been some commercially available fractal/self similar antennas made by Fractal Antenna systems Inc. [10]. Fractal Antenna Systems has fabricated a wideband dipole as shown in Figure 8. The antenna has a self similar geometry present at multiple resolutions. Along two of the axes there is a substantial mirror image. Along the center there is a thin dipole that runs all the way through the antenna. The long thin dipole geometry creates a loading effect rather than predominately providing the function of radiating electromagnetic energy, which instead is provided by the more distinct V shaped dipole as shown in Figure 9. This antenna is claimed to radiate energy between 400 MHz and 6000MHz and are fabricated using deterministic fractal geometry.

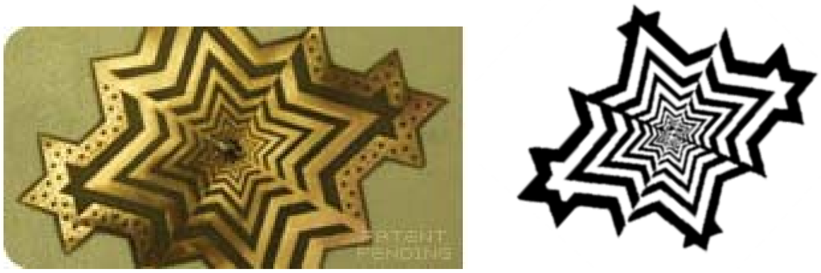

Figure 8 Wideband dipole by Fractenna

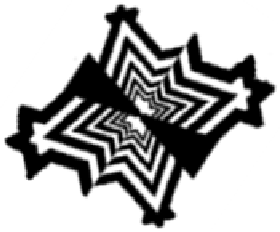

Figure 9 Wideband Dipole (Bowtie) by Fractenna

REFERENCES

[1] N. Kingsley, D.E. Anagnostou, M. Tentzeris and J. Papapolymerou, "RF MEMS Sequentially Reconfigurable Sierpinski Antenna on a Flexible Organic Substrate With Novel DC-Biasing Technique", J. OF MICROELECTROMECHANICAL SYSTEMS, VOL. 16, NO. 5, OCTOBER 2007, pp. 1185-1192

[2] D. E. Anagnostou *et al.*, "Design, fabrication, and measurements of an RF-MEMS-based self-similar reconfigurable antenna," *IEEE Trans. Antennas Propag.*, vol. 54, no. 2, pt. 1, pp. 422–432, Feb. 2006.

[3] B. Cetiner, J. Qian, H. Chang, M. Bachman, G. Li, and F. DeFlaviis, "Monolithic integration of RF MEMS switches with a diversity antenna on PCB substrate," *IEEE Trans. Microw. Theory Tech.*, vol. 51, no. 1, pp. 332–334, Jan. 2003.

[4] R. Jackson and R. Ramadoss, "A MEMS-based electrostatically tunable circular microstrip patch antenna," *J. Micromech. Microeng.*, vol. 17, no. 1, pp. 1–8, Jan. 2007.

[5] N. Kingsley and J. Papapolymerou, "Organic 'wafer-scale' packaged miniature four-bit RF MEMS phase shifter," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 3, pp. 1229–1236, Mar. 2006.

[6] C. P. Baliarda, J. Romeu, and A. Cardama, "The Koch monopole: A small fractal antenna," *IEEE Trans. Antennas Propag.*, vol. 48, no. 11, pp. 1773–1781, Nov. 2000.

[7] N. Cohen and R. Hohlfeld, "Fractal loops and the small loop approximation," *Commun. Q.*, pp. 77–81, Winter 1996.

[8] J. Anguera, C. Puente, C. Borja, and J. Soler, "Fractal-shaped antennas: A review," in *Wiley Encyclopedia of RF and Microwave Engineering*, vol. 2, K. Chang, Ed. New York: Wiley, 2005, pp. 1620–1635.

[9] D. Thompson, O. Tantot, H. Jallageas, G. Ponchak, M. Tentzeris, and J. Papapolymerou, “Characterization of liquid crystal polymer (LCP) material and transmission lines on LCP substrates from 30 to 110 GHz,” *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 4, pp. 1343–1352, Apr. 2004.

[10] Fractal Antenna Systems, Inc. 130 Third Avenue Waltham MA 02451, www.fractenna.com

2.10 Retrodirective Arrays

2.10.1 Summary

Retrodirective antennas are self-adapting systems that transmit a signal or direct their beam in the direction of an incoming signal without any prior knowledge of the signal's angle of arrival or intensive signal processing to determine the direction. Two types of retrodirective antennas have been developed, the Van Atta (VA) array and the Phase Conjugating (PC) array. The VA array is a fundamentally a passive, planar structure (although active VA arrays exist as well). The PC array is an active antenna that can be fabricated on a non-planar surface. Thus, the VA array generally is simpler and, consequently, easier to design and fabricate. Retrodirective systems are attractive candidates for the set-top television antenna.

While existing retrodirective systems typically operate at frequencies well above the television band, the technology is advancing rapidly and much of it at this time or in the short-term future likely will be applicable to set-top antennas. This section describes several state-of-the-art implementations, mostly for VA systems, but some for PC systems. It also describes an important development in conformal PC array design using genetic algorithm optimization that allows a retrodirective array to be designed for a specific surface shape. Finally, some new work on near field scatterer effects is discussed that shows how a retrodirective antenna can mitigate near field phase distortions in environments such as the interior of a building (a typical set-top environment).

(a) Retrodirective Concept

A retrodirective antenna is conceptually similar to an optical corner reflector, three plane reflecting surfaces intersecting at right angles. A light beam incident onto the corner reflector, also known as a retroreflector or cataphote, is automatically reflected back along the direction from which it came. The basic retrodirective array concept is shown schematically in Figure 1. There are two basic structures: (1) the Van Atta array, and (2) the Phase Conjugating array. The VA array is simpler than the PC array because it does not require active devices, whereas the PC array does. VA and PC arrays operate differently, but both systems achieve the same goal of directing the antenna's beam in the direction of an incoming signal. Without any knowledge of the signal's direction of arrival, a retrodirective antenna is able to transmit a signal back toward the transmitter along the direction of that incoming signal without using phase shifters.

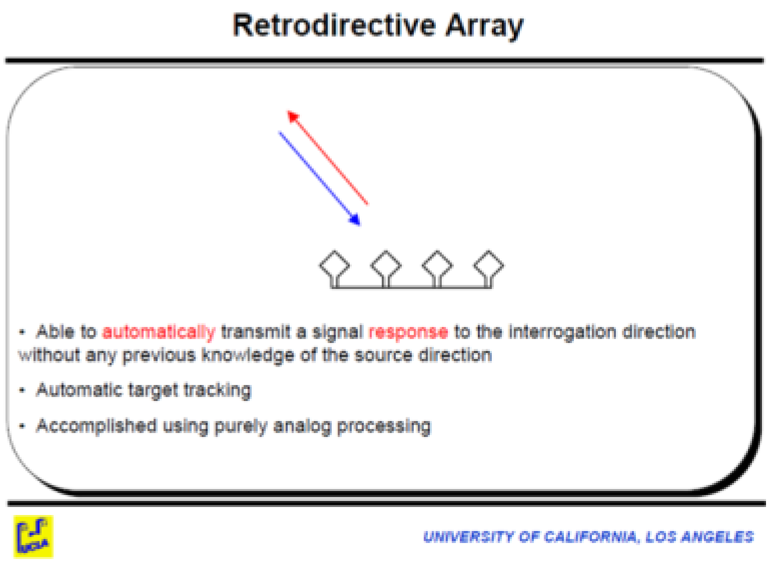
Retrodirective Array

- Able to automatically transmit a signal response to the interrogation direction without any previous knowledge of the source direction
- Automatic target tracking
- Accomplished using purely analog processing

UNIVERSITY OF CALIFORNIA, LOS ANGELES

Figure 1. Basic retrodirective concept (reproduced from [1].

(b) Van Atta Array

The VA concept [2] is shown in Figure 2. Pairs of antenna elements, which can be any type of radiator (in this case horns), are equally spaced about the array's center point and interconnected with waveguides (or transmission lines) that have equal length or are multiples of a wavelength. Each antenna element acts as both a receiving antenna and a transmitting antenna. The incident field received by each element is reradiated by its paired counterpart after propagating along the interconnecting transmission line or waveguide. The effect of the resulting propagation delay is to reverse the phase of the reradiated field distribution relative to the received field distribution, causing the reradiated field to be directed back along the line of sight of the incident field. While Van Atta's original antenna was strictly a passive retroreflector, active systems have been developed that utilize the basic Van Atta structure [3,4,5]. Figure 3 provides a schematic illustration of the Van Atta concept for passive and active implementations.

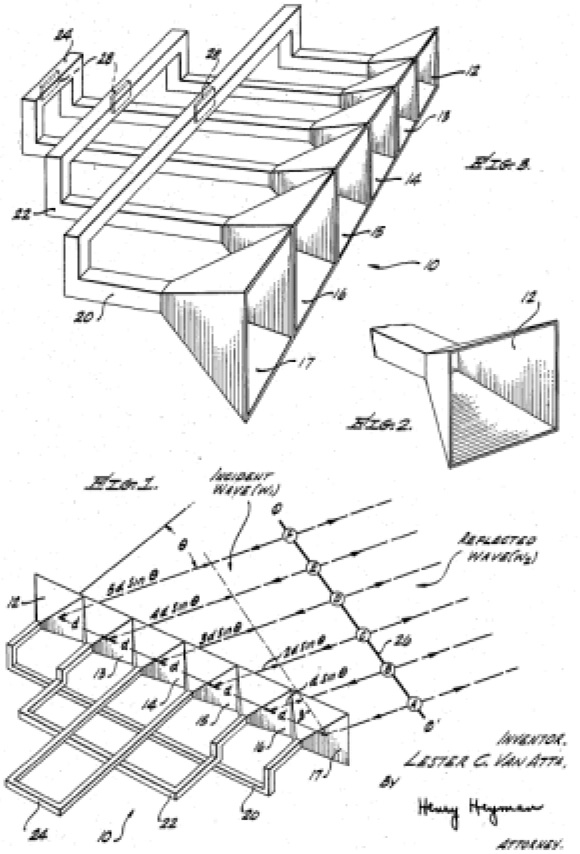
24
28
28
12
13
14
15
16
17
10
20
22
FIG. 3.
FIG. 2.
FIG. 1.
INCIDENT WAVE (W1)
REFLECTED WAVE (W2)
θ
θ'
5d sin θ
4d sin θ
3d sin θ
2d sin θ
d sin θ
d
26
INVENTOR,
LESTER C. VAN ATTA,
BY
Henry Heyman
ATTORNEY.

Figure 2. Van Atta retrodirective array (reproduced from [2]).

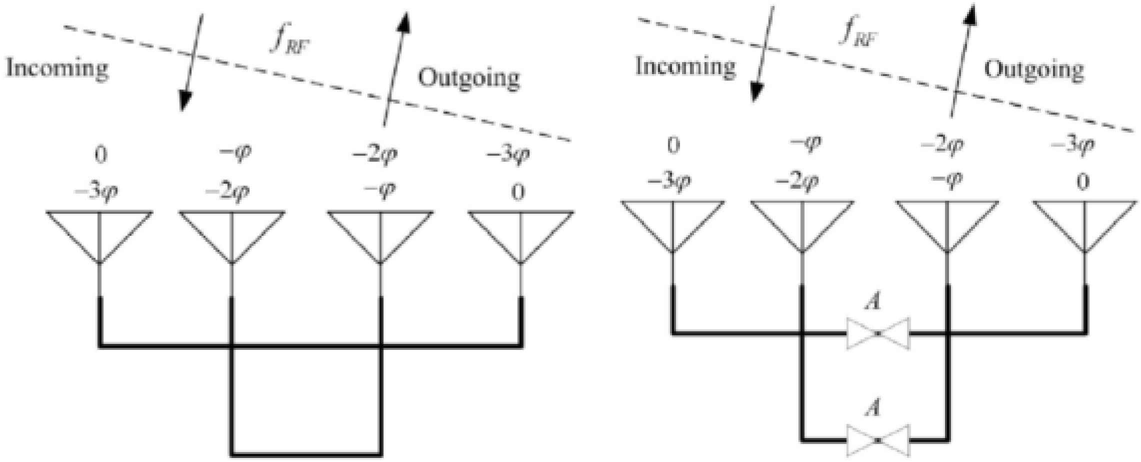
Incoming
f_{RF}
Outgoing
0
-3φ
$-\varphi$
-2φ
-2φ
$-\varphi$
-3φ
0
Incoming
f_{RF}
Outgoing
0
-3φ
$-\varphi$
-2φ
-2φ
$-\varphi$
-3φ
0
A
A

Figure 3. Passive (a) and active (b) Van Atta array phase front (reproduced from [3]).

(c) Phase-Conjugating Array

Unlike the Van Atta array, the PC retrodirective array requires active elements because its operation relies on actively changing the relative phase of its elements. The direction of an emitted signal, or, by reciprocity, the maximum directivity for received signals, is controlled by the relative phase and amplitude of the individual radiating sources or receiving elements in a multi-element array as shown in Figure 4. By operating a local oscillator (LO) at twice the received signal's radio frequency (RF), mixing the LO and RF signals results in a lower sideband product whose frequency is the same as the incident signal's but whose phase is conjugated (reflected about the real axis as shown in Figure 5).

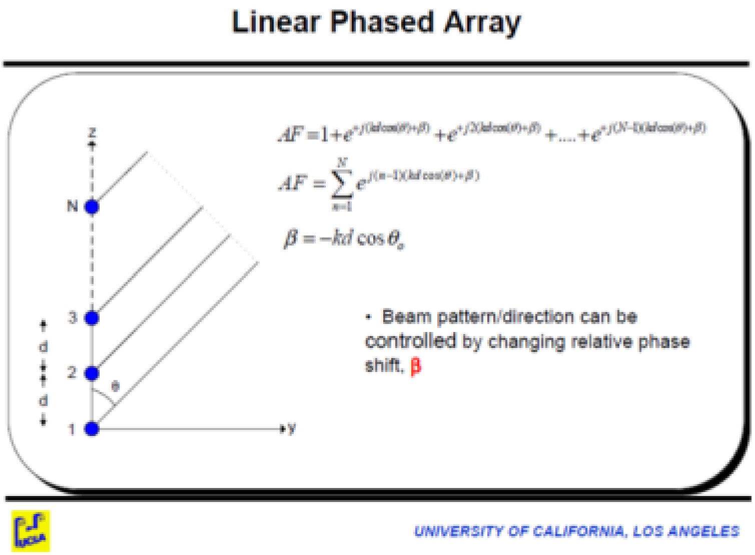
Linear Phased Array

$$AF = 1 + e^{+j(kd\cos(\theta)+\beta)} + e^{+j2(kd\cos(\theta)+\beta)} + \dots + e^{+j(N-1)(kd\cos(\theta)+\beta)}$$

$$AF = \sum_{n=1}^{N} e^{j(n-1)(kd\cos(\theta)+\beta)}$$

$$\beta = -kd\cos\theta_o$$

- Beam pattern/direction can be controlled by changing relative phase shift, β

UNIVERSITY OF CALIFORNIA, LOS ANGELES

Figure 4. Beam direction control (reproduced from [1]).

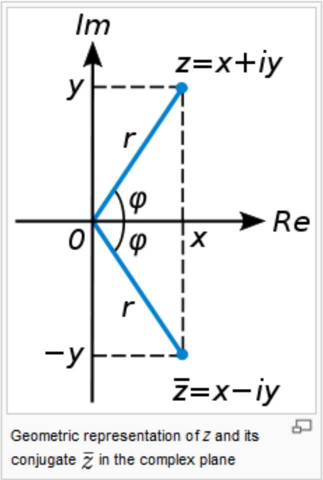

Geometric representation of z and its conjugate $\bar{z}$ in the complex plane

Figure 5. Phase conjugation.

Conjugating the phase for each antenna element steers the antenna beam back along the direction of the incoming signal as shown in Figure 6. If the antenna is used for transmitting, then the radiated signal is transmitted along back along the direction of the received signal. If the antenna is used for receiving, then the radiation pattern is steered with is maximum directivity in the direction of the received signal.

A conjugating retrodirective antenna, whether VA or PC, could be very useful as a set-top television antenna because it provides optimal gain in the direction of the received signal regardless of its incidence direction. A retro receive antenna would automatically adapt itself to the signal environment allowing a fixed antenna to provide maximum response to signals from any direction. The following sections describe some state-of-the-art implementations that may be applicable in the set-top television environment.

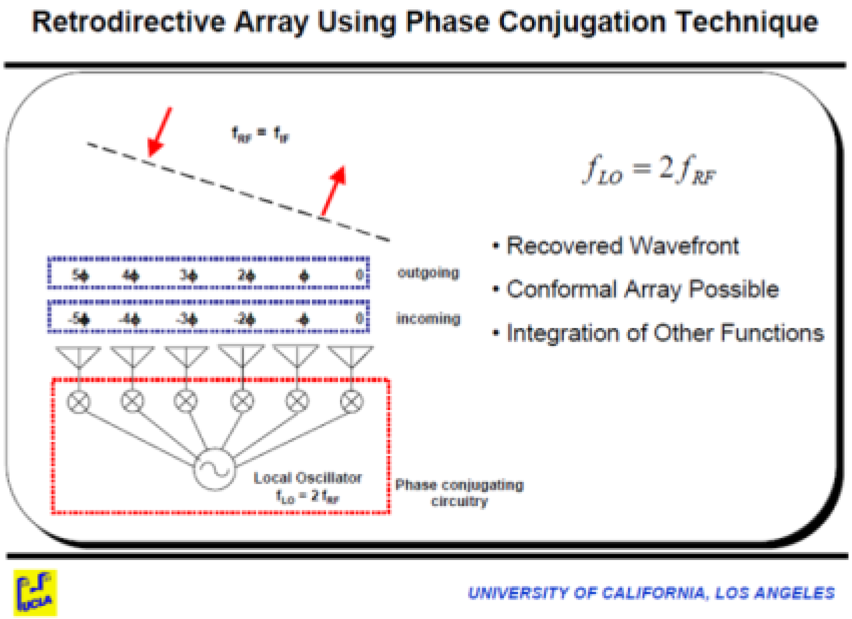

Retrodirective Array Using Phase Conjugation Technique

$f_{LO} = 2 f_{RF}$

- Recovered Wavefront
- Conformal Array Possible
- Integration of Other Functions

$f_{RF} = f_{IF}$
5ϕ 4ϕ 3ϕ 2ϕ ϕ 0
outgoing
-5ϕ -4ϕ -3ϕ -2ϕ ϕ 0
incoming
Local Oscillator
$f_{LO} = 2 f_{RF}$
Phase conjugating circuitry
UCLA
UNIVERSITY OF CALIFORNIA, LOS ANGELES

Figure 6. Retrodirective phase conjugating architecture (reproduced from [1]).

(d) Retrodirective Array Implementations

Broadband Van Atta Array. Figure 7 shows the schematic diagram for a frequency-offset broadband Van Atta array [6]. While the VA architecture is inherently more broadbanded than PC arrays because bandwidth-limiting LO and phase conjugation mixers are not required, there nevertheless are active-element implementations that improve performance while maintaining substantial bandwidth as shown in the figure. In this particular application, the antenna is used for two-way communication, which is why optional amplifiers that increase effective radiated power are included and dual-polarized antennas are used (orthogonal polarizations for transmission/reception to increase isolation).

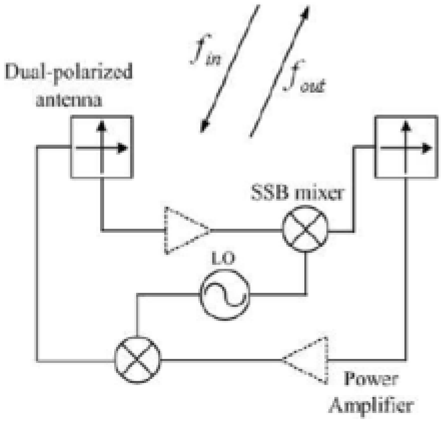
Dual-polarized antenna
f_{in}
f_{out}
SSB mixer
LO
Power Amplifier

Figure 7. Diagram of broadband two-way Van Atta array (reproduced from [6]).

Figure 8 shows the structure of the prototype broadband VA array antenna. A patch radiator is fed by two excitation slots at right angles to provide orthogonal polarization. Microstrip feed lines excite the apertures and are deposited as shown the opposite side of the aperture substrate. The reflecting plane is included to improve front-to-back ratio. Two such antennas comprise the prototype array which operates from 2.8 to 3.4 GHz. The antenna as fabricated is shown in Figure 9. The patch antennas are evident on the surface of the front circuit board substrate, while the separate SSB mixer circuit boards are seen in the background. This antenna achieved beam steering over approximately ±30º from the broadside direction (perpendicular to the plane of the antenna patches). This active VA array shows that substantial retrodirectivity can be achieved even in a simple two-element array. It perhaps points the way towards retrodirective set-top antennas based on the same or similar architecture that utilize different antenna elements, such as dipoles or bow-ties, that are better suited to the lower television frequencies.

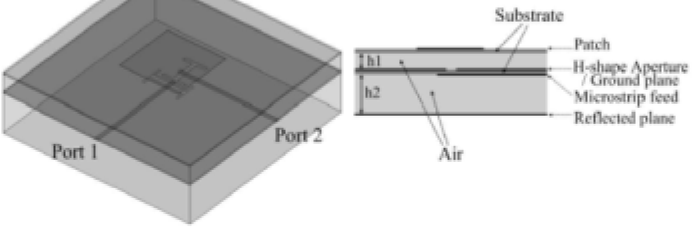
Port 1
Port 2
Substrate
Patch
h1
h2
H-shape Aperture / Ground plane
Microstrip feed
Reflected plane
Air

Figure 8. Broadband two-way Van Atta array antenna element (reproduced from [6]).

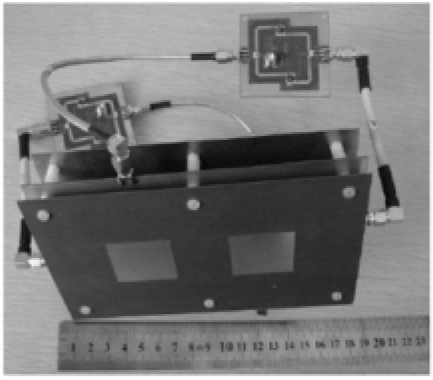

Figure 9. Broadband two-way Van Atta array as fabricated (reproduced from [6]).

Passive Array with Hybrids. An interesting example of a passive VA array implementation involves using 90º hybrid couplers as the phasing elements [7]. This design relies on a clever use of hybrid couplers whose incident and reflected signals' phases are reversed when the coupler is terminated with loads having the same reflection coefficient. Figure 10 shows how a phase difference $\Delta\phi$ between signals into ports 1 and 2 (input/isolated ports, respectively) is transformed into a phase difference of $-\Delta\phi$ between the reflected signals at those ports as long as ports 3 and 4 (direct/coupled, respectively) are terminated with the same reflection coefficient. In practice, these ports simply float unterminated because Γ then is the same for each (~1).

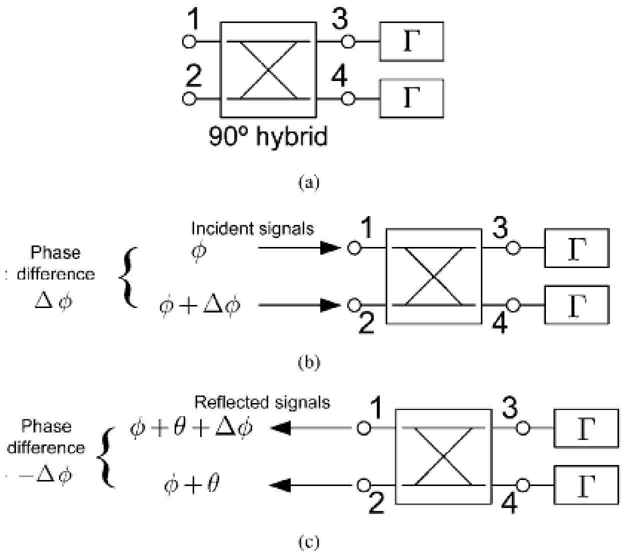
1
3
2
4
Γ
Γ
90° hybrid
(a)
Incident signals
Phase difference Δφ
φ
φ+Δφ
1
3
2
4
Γ
Γ
(b)
Reflected signals
Phase difference −Δφ
φ+θ+Δφ
φ+θ
1
3
2
4
Γ
Γ
(c)

Fig. 2. (a) A terminated 90° hybrid with (b) and (c) to illustrate the phase difference reversal characteristics of the input signals and the reflected signals.

Figure 10. Passive VA array using hybrid couplers (reproduced from [7]).

A particularly attractive feature of this design is ease of implementation. Figure 11 shows 3-, 4- and 6-element retrodirective array circuits fabricated on 1.6 mm thick FR4 substrate using planar 90º microstrip branch line couplers. It is significant, too, that this retrodirective design does not require an even number of array elements as does the basic Van Atta design. The prototype array was fabricated with 2.9 GHz patch antenna elements on the same type substrate spaced 0.48λ. Measurements showed effective beam steering over a range of ±40º. This type of retrodirective antenna may be particularly attractive for the set-top application because of its simplicity.

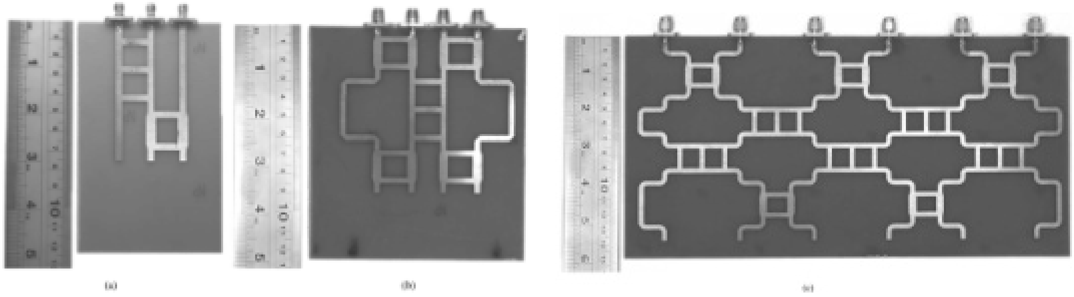

Figure 11. Passive 3/4/6-element VA array circuits as fabricated on PCB (reproduced from [7]).

Feed Systems for Van Atta Arrays. Two interesting and potentially useful VA array designs are described in [8]. Two planar VA arrays were designed and simulated using commercially available electromagnetic modeling software (ANSOFT HFSS). The first antenna (Figure 12, (a)/(b) left) used a microstrip line edge-fed patch at 7.58 GHz as the antenna element (overall dimensions ~13 x 30 mm). This element was arrayed in a two-dimensional 4x4 grid and evaluated for bandwidth and radiation characteristics. The second antenna comprised aperture-coupled microstrip patch antenna elements operating at 9 GHz also arrayed in a 4x4 2-D grid (Figure 12, (a)/(b) right). The second design was found to be superior because it could be more easily fabricated using both sides of the substrate and it provided better beam steering (±80º @ -10 dB). This simulation shows that very effective designs with small footprints are possible at very short wavelengths. This technology may be applicable directly or with modification or extension to set-top antenna requirements.

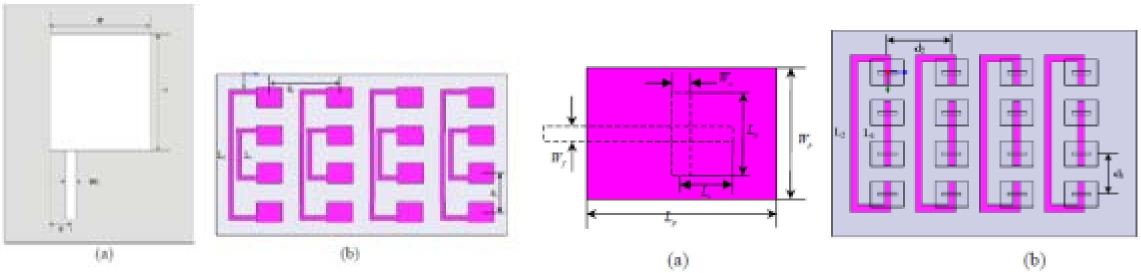

Figure 12. Planar edge-fed VA arrays (reproduced from [8]).

A Fractal VA Antenna Element. Dual-polarized Van Atta array antenna elements provide advantages over singly polarized elements, in particular increased transmit/receive (T/R) isolation on two-way links, multipath mitigation, and polarization diversity. While T/R isolation is not a consideration in the set-top television antenna, the other factors are. Additionally, size is a concern, and newer antenna element designs emphasize small size. An interesting fractal-based element is reported in [9] that may be useful for set-top VA arrays. Figure 13 shows the overall antenna in perspective and plan views. There are two square (40 mm) 1-mm thick material dielectric substrates, the top one being PTFE material ($\varepsilon_r = 2.65$), and the bottom FR3 ($\varepsilon_r = 4.4$). The First Minkowski fractal patch is shown on the right with the vertical and horizontal polarization feed ports marked V and H, respectively.

Details of the patch, which is designed to operate from ~4.4 to 5 GHz, appear in Figure 14. Its overall size is quite small at 21 x 16.8 mm, and it provides similar electrical performance to a square patch that is about 20% larger. Figure 15 shows details for the H-shaped slots that function as aperture coupling structures to excite the fractal patch. It is evident that, once dimensions are specified, the components are readily fabricated and easily assembled, which are very desirable characteristics for the set-top television antenna.

Computed data show that the antenna element exhibits H-polarization VSWR < 2 from ~4.82-5.17 GHz with 16.8% impedance bandwidth. The corresponding V-polarization figures are ~4.27-5.2 GHz, bandwidth 19.8%. Measured results confirm the modeling with bandwidth figures of 19.7% and 20.5%, respectively, which is good agreement. Directive gain is ~8 dB with >20 dB front-to-back ratio. These results clearly show that the microstrip Minkowski fractal antenna element is useful in VA arrays at microwave frequencies, and that it well may provide excellent performance for the higher television bands as well.

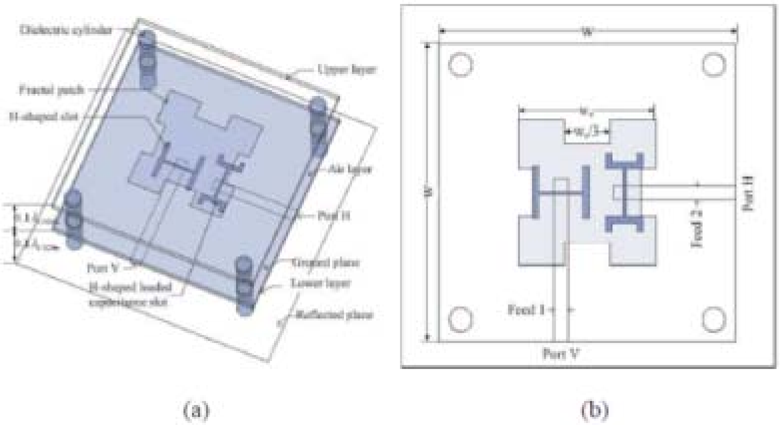
Dielectric cylinder
Upper layer
Fractal patch
H-shaped slot
Air layer
Port H
Ground plane
Lower layer
Reflected plane
Port V
H-shaped loaded cophasing slot
(a)
W
w_p
$w_p/3$
W
Port H
Feed 2
Feed 1
Port V
(b)

Figure 13. Proposed microstrip fractal antenna element (reproduced from [9]).

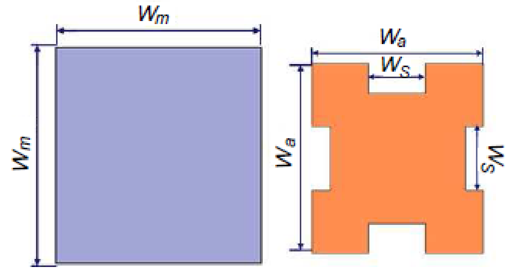

Figure 2. Structure of the conventional square patch and the Minkowski fractal patch ($w_m = 21\,\mathrm{mm}$, $w_a = 16.8\,\mathrm{mm}$).

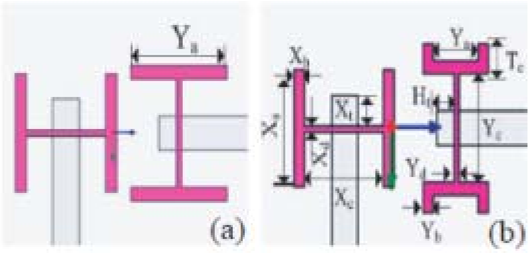

Figure 3. The aperture-coupled structures. (a) Double H-shaped slots, (b) H-shaped and H-shaped loaded capacitance slots.

Figure 14. Fractal patch (reproduced from [9]).

Figure 15. Aperture details (reproduced from [9]).

Dual-Ring Antenna Elements. The dual-ring antenna element proposed in [10] is another novel device that can be used in retrodirective Van Atta arrays. Inner and outer concentric rings form an inherently two resonant-frequency structure. Each of the rings is parasitic with the other ring excited, and by properly choosing dimensions fairly widely separated resonance points can mark the upper and lower limits of a fairly flat response curve. The structure of the dual-ring element is shown in Figure 16. It is implemented in a two-layer microstrip structure with the rings deposited on the outer surface of an upper substrate underlain by a transmission line for feeding and a ground plane as shown. Element dimensions were optimized using HFSS modeling software, and the fabricated element tested. The measured bandwidth (-10 dB) was just under 34% (31-42.8 GHz), which agreed quite well with the predicted value of 34.3% (30-42 GHz). Between 30 and 40 GHz the average measured gain was ~4.5 dBi.

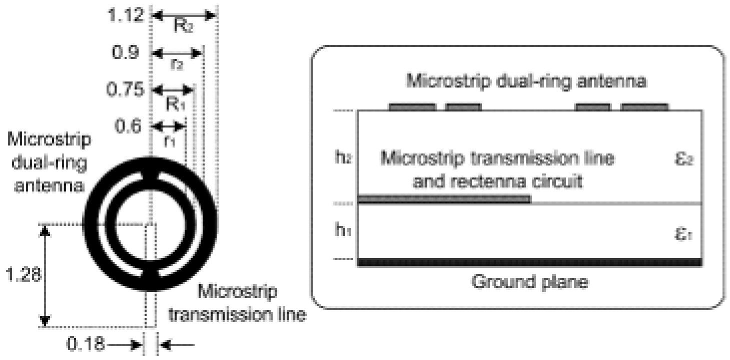
1.12
R2
0.9
r2
0.75
R1
0.6
r1
Microstrip dual-ring antenna
1.28
0.18
Microstrip transmission line
Microstrip dual-ring antenna
h2
Microstrip transmission line and rectenna circuit
ε2
h1
ε1
Ground plane

Fig. 3. Geometry of the broadband proximity-coupled dual-ring antenna and the two-layer dielectric structure. All dimensions are in mm.

Figure 16. Dual ring antenna element (reproduced from [10]).

The Van Atta array implemented using dual-ring elements was fabricated using 8 sub-arrays comprising 16 dual-ring elements in a 4 x 4 2-D grid. Figs. 17 and 18 show the sub-array and complete array structures, respectively. Measurements confirm the new array's predicted performance, which shows excellent beam steerability out to about ±45º. The large array exhibits a very narrow main half-power beamwidth on the order of ~2º that is very stable between 32 and 49 GHz. Of course, this prototype antenna operates far beyond the highest television operating frequency, but its technology may be applicable to the TV set-top antenna with appropriate modifications.

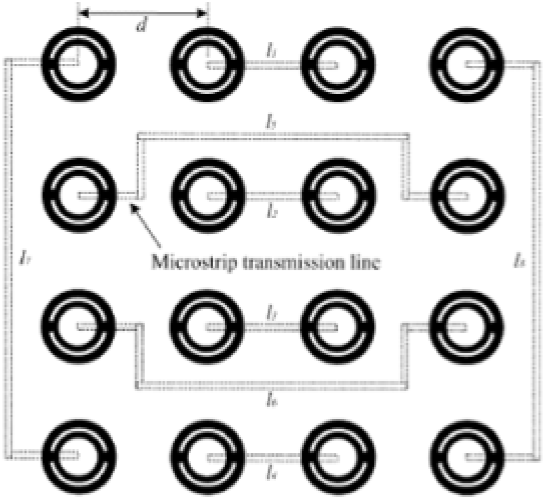
d
l1
l5
l2
Microstrip transmission line
l7
l8
l3
l6
l4

Fig. 5. Geometry of the 4x4 retrodirective sub-array where the dash-lines are the microstrip transmission line networks.

Figure 17. Dual ring element sub-array (reproduced from [10]).

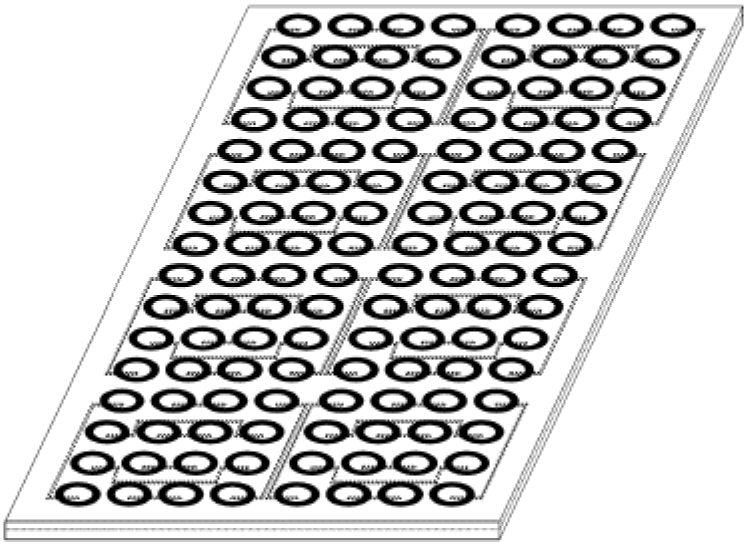

Fig. 6. Geometry of the 8x16 retrodirective array that consists of eight 4x4 sub-arrays.

Figure 18. Retrodirective array of dual ring elements (reproduced from [10]).

Conformal Retrodirective Arrays. In some cases it may be desirable to deploy a retrodirective antenna on a non-planar surface (the systems described above all utilize planar structures). Unfortunately the simpler Van Atta design cannot be used in such cases, so that Phase Conjugating systems are required instead. An interesting example of a genetically optimized, thinned PC array is described in [11]. The array's schematic representation appears in Figure 19, and its final physical configuration in Figure 20.

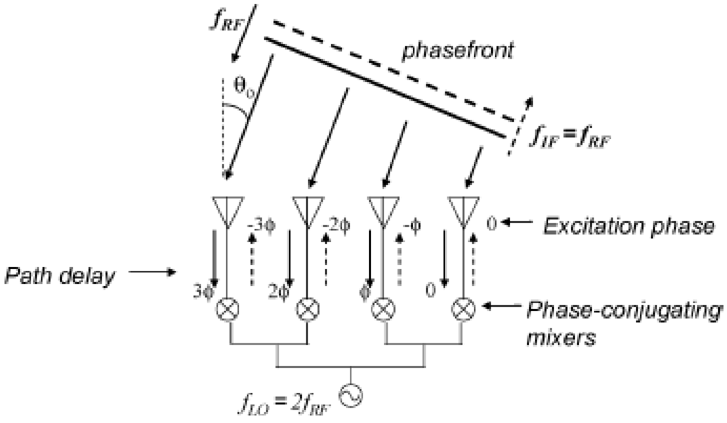
f_{RF}
phasefront
θ_0
$f_{IF} = f_{RF}$
-3ϕ
-2ϕ
$-\phi$
0
Excitation phase
Path delay
3ϕ
2ϕ
ϕ
0
Phase-conjugating mixers
$f_{LO} = 2f_{RF}$

Fig. 1. A retrodirective array using heterodyne technique [5].

Figure 19. Conformal PC retrodirective array schematic diagram (reproduced from [11]).

In this example, the array was thinned to a total of 16 elements whose positions were optimized to achieve reasonable sidelobe levels over a scanning range of ~±120º. The shape and size of the optimized array aperture cannot be determined analytically, but a suitable optimization algorithm permits design of conformal PC arrays. This type of structure could be important in the set-top environment as a space-saving measure, but it would incur the additional expense and complexity associated with active PC arrays compared to the simpler passive Van Atta design.

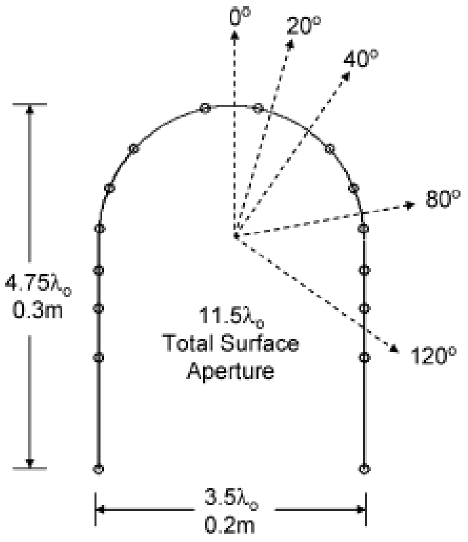

Fig. 2. The shape and the physical size of the aperture [4].

Figure 20. Conformal PC array structure (reproduced from [11]).

Near-Field Effects. Retrodirective antennas typically operate primarily in free space or near-free space environments, so that near field obstructions are not a factor in antenna performance. This typical environment is quite different from the set-top environment in which nearby objects and structures will definitely be in the antenna's near field and consequently a potential source of interference. Near field effects were studied in [12] in order to investigate these effects. Figure 21 shows the experimental antenna in an anechoic chamber and typical measured data. Three different near field scatterers were studied: (a) straight wire grid; (b) partially obscuring low loss MDF (medium density fiberboard); and (c) partially or totally obscuring concrete blocks.

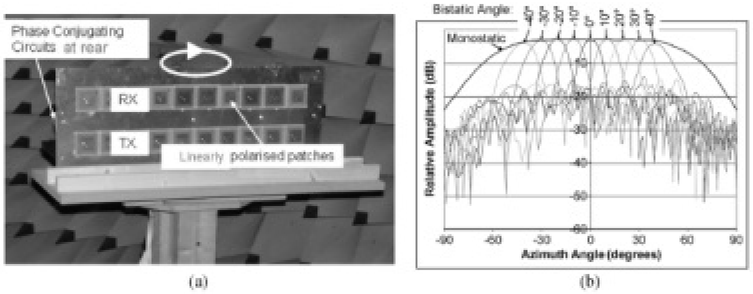

Figure 21. Measurement system for near field effects (reproduced from [12]).

Measurements were made at 2.4 GHz using a 10-element retrodirective array scanning ±40º with a half-power beamwidth of ~12º. The retrodirective antenna performed quite well in the presence of obstructions. For example, using gapped concrete block scatterers, it was able to fully compensate for phase distortions that reduced a basic reference antenna's gain by 11 dB. This study was the first to demonstrate that retrodirective antennas offer demonstrable performance advantages over other antennas in difficult propagation environments such as the interior of a building. This finding is an important reason for exploring retrodirective technology as it might be used in a television set-top antenna.

2.10.2 References

[1] Leong, K., and Itoh, T., "Active Retrodirective Arrays," Electrical Engineering Dept., Univ. of California Los Angeles, available online: http://www.rish.kyoto-u.ac.jp/jusps/P15.pdf.

[2] Van Atta, L. C., "Electromagnetic Reflector," U.S. Patent No. 2,908,002, Oct. 6, 1959.

[3] Chung, S-J., Chen, S-M., Lee, Y-C., "A Novel Bi-Directional Amplifier with Applications in Active Van Atta Retrodirective Arrays," *IEEE Trans. Microwave Theory and Techniques*, **51**, No. 2, Feb. 2003, p. 542.

[4] Guo, Y. C., Shi, X. W., Chen, L., "Retrodirective Array Technology," *Prog. In Electromagnetics Research B* (PIER B), **5**, pp. 153-167 (2008).

[5] Goshi, D. S., Leong, K. M. K. H., and Itoh, T., "Recent Advances in Retrodirective System Technology," *IEEE Radio and Wireless Symposium*, pp. 459-462, 2006.

[6] Chen, L., Wang, X.-H., Shi, X.-W., Zhang, T.-L., Ting, J.-Z., "Design of a Broadband Frequency Offset Van Atta Array," *Prog. In Electromagnetics Research Letters*, **13**, pp. 161-171 (2010).

[7] Hsieh, S-N., and Chu, T-H., “Linear Retro-Directive Antenna Array Using 90º Hybrids,” *IEEE Trans. Antennas and Propagation*, **56**, No. 6, June 2008.

[8] Yao, H-Y., and Wang, C-F., “Investigation of Electromagnetics Properties of Two Types of Passive Retrodirective Antenna Arrays,” IEEE Antennas & Propagation Society International Symposium 2008 (*AP-S 2008*). DOI: 10.1109/APS.2008.4619052.

[9] Cheng, H. R., Chen, X. Q., Chen, L., and Shi, X. W., “Design of a Fractal Dual-Polarized Aperture Coupled Microstrip Antenna,” *Progress in Electromagnetics Research Letters*, **9**, pp. 175-181 (2009).

[10] Ren, Y-J., and Chang, K., “A New Millimeter-Wave Broadband Retrodirective Antenna Array,” *Microwave Symposium, 2007. IEEE/MTT-S International.* DOI: 10.1109/MWSYM.2007.380038.

[11] Sun, J. S., Goshi, D. S., Itoh, T., “Optimization and Modeling of Sparse Conformal Retrodirective Array,” *IEEE Trans. Antennas and Propagation*, **58**, No. 3, March 2010.

[12] Fusco, V. F., and Buchanan, N., “Retrodirective Array Performance in the Presence of Near Field Obstructions,” *IEEE Trans. Antennas and Propagation*, **58**, No. 3, March 2010.

3.0 Conclusions and Design Recommendations

Upon consideration of Table 1-1 at the end of Section 1 of this report we have concluded the following:

- There have been significant and potentially useful antenna design methods and actual hardware developed over the past 15 years that will improve the performance of low-profile, compact indoor/set-top DTV antennas.
- In the area of electro-magnetic computational and design optimization methods Genetic Algorithms (GA) and the Central Force Optimizer (CFO) have proven themselves as powerful tools in the design of broadband, compact antennas.
- Using the above, we have included an example of an advanced antenna technology, specifically the Fragmented Antenna as part of this project. We submit that it would be nearly impossible for a human to replicate its design and performance.
- The Non-Foster active broadband matching technology and its required semiconductors have matured to the point where they could be incorporated into a commercial indoor/set-top design to provide a matched antenna where it is considered as being electrically small. In the example shown below it would be used to provide acceptable performance with the Fragmented dipole within the 54-88 MHz band. Other element geometries are also possible.
- Many of the other advancements listed in Table 1-1, specifically 2.4 through 2.6, require a CE-909-A type interface with the DTV receiver, and while potentially useful if and when manufacturers start adding this feature to their products they offer no practical improvement for the near term.
- The technologies designated as 2.7 through 2.10 while interesting are either too embryonic (2.7 & 2.8), too complicated and too large (2.10) or not well vetted (2.9) to be considered in the near term.

Considering the above it is recommended that the NAB give consideration to further developing a computer generated/optimized broadband dipole antenna element, such as a Fragmented Planar Dipole with/without a Non-Foster Matching circuit in order to demonstrate that it is feasible to design a VHF/UHF antenna system that is both simple, have positive gain, be of reasonable size and weight and not require a CE-909-A interface with the DTV receiver.

As an example, a fragmented planar antenna was optimized for 54-698 MHz to cover all three DTV television bands. The antenna has dimension of 13” x 13” across the VHF and UHF bands, a block diagram of the proposed antenna system is shown in Figure 1. The computed directivity, after optimization, along the horizon is good at all DTV frequencies as shown in Figure 2. The antenna shows a good match at UHF frequencies, but as shown in Figures 4 and 5 some matching, using a simple L-C passive network of SMT components and Non-Foster Matching circuitry at the feed-point would be necessary for good performance in the high and low VHF bands respectively.

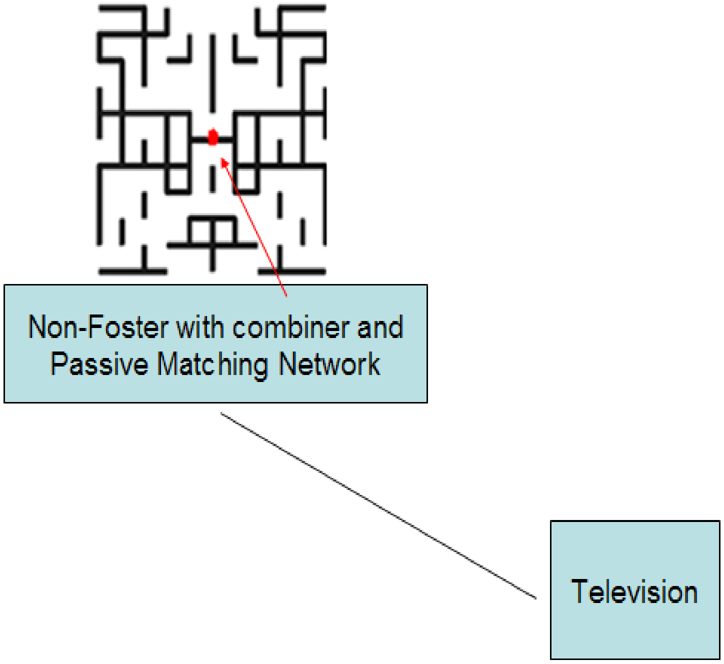
Non-Foster with combiner and Passive Matching Network
Television

Figure 1. Fragmented Antenna Example, 13”x13”

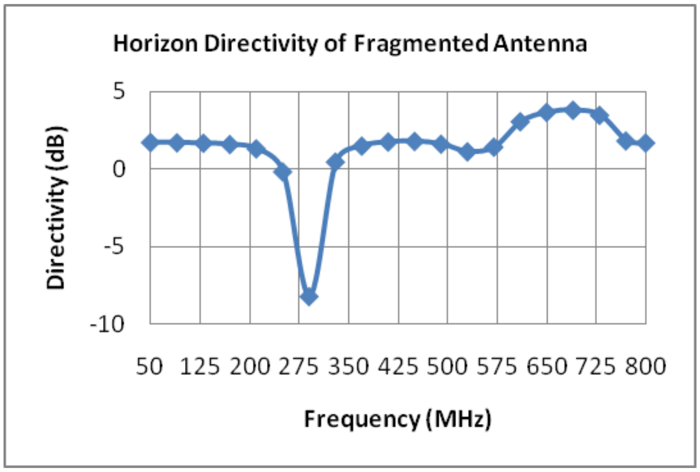

Figure 2. Directivity Along Horizon for Fragmented Antenna

The pattern of the antenna as currently designed is a “figure eight” dipole pattern at all frequencies, as shown in Figure 3. Figure 6 indicates that without any matching, a very low VSWR across the UHF 470-698MHz band.

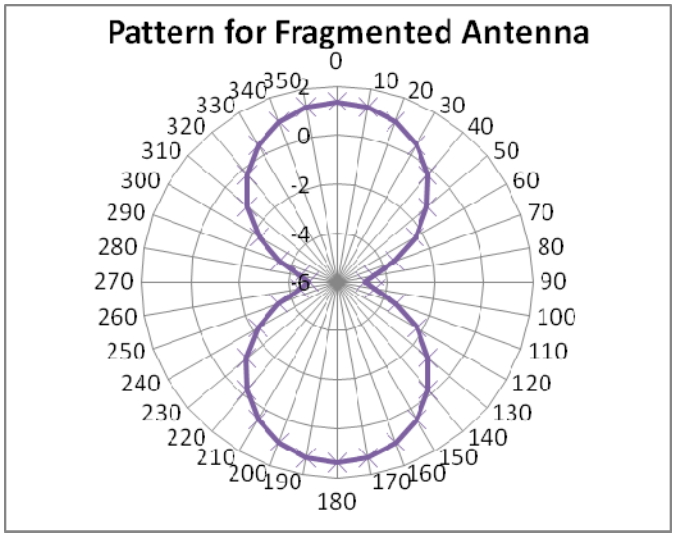

Figure 3. Azimuthal Pattern of Fragmented Antenna

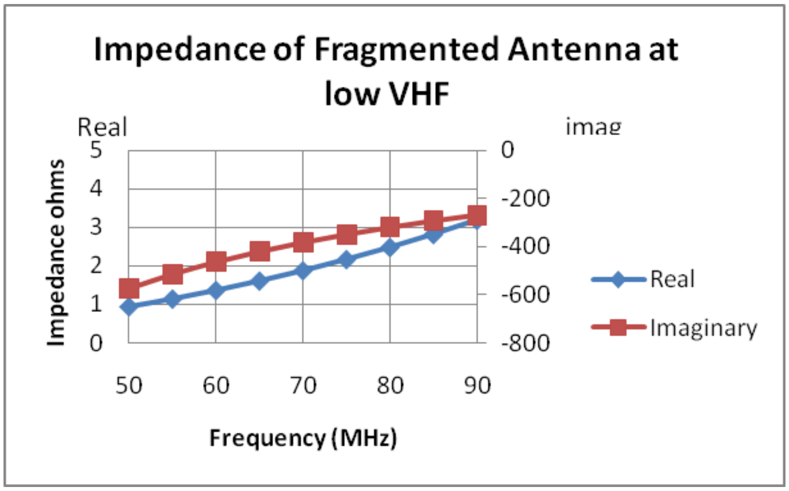

Figure 4. Impedance of Fragmented Antenna at low VHF

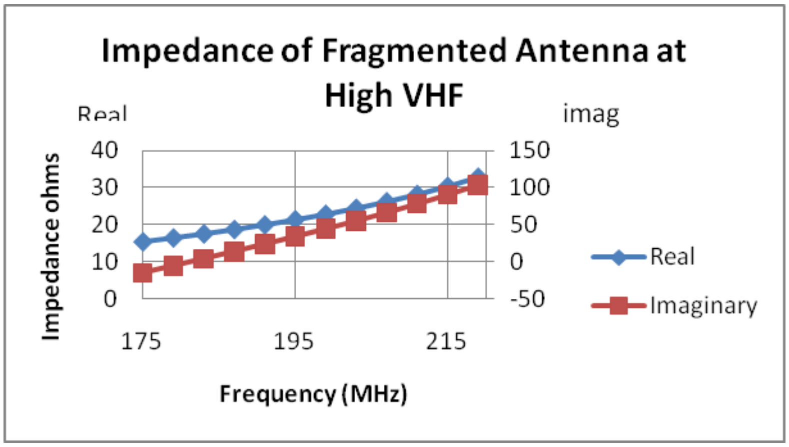

Figure 5. Impedance of Fragmented Antenna at high VHF

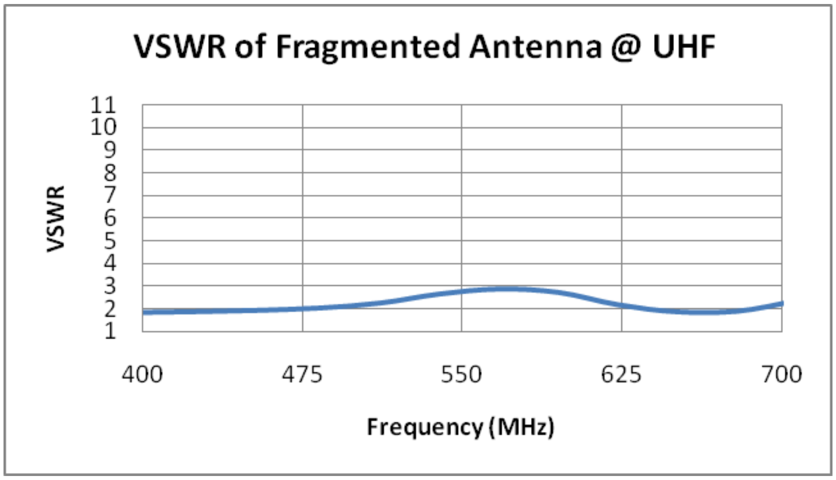

Figure 6. VSWR of Fragmented Antenna at UHF

A Microstrip Patch Antenna Manufactured with Flexible Graphene-Based Conducting Material

Sayeed Z. Sajal and Benjamin D. Braaten
Electrical and Computer Engineering Department
North Dakota State University
Fargo, ND 58108-6050, USA
Email: sayeed.sajal@ndsu.edu, benbraaten@ieee.org

Val R. Marinov
Industrial and Manufacturing Engineering Department
North Dakota State University
Fargo, ND 58108-6050, USA
Email: val.marinov@ndsu.edu

***Abstract*—In this paper, a unique process for fabricating microstrip patch antennas with flexible graphene-based conductors is presented. In particular, this manufacturing process uses a commercially available micro-cutter to cut the outline of the patch antenna from the flexible graphene-based conductors, and then this piece is attached to a grounded FR4 substrate using adhesive to create a unique printed antenna. The design was modeled using a commercial simulator, and a prototype was fabricated and measured. Overall, it was shown that the S-parameter simulations agreed fairly well with measurements, and that this manufacturing process has the potential to develop more complicated designs, such as meander-line dipoles for example, that are difficult to cut-out manually.**

***Index Terms*—Patch antenna and graphene-based conductors.**

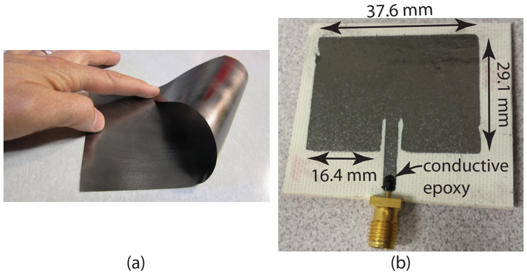

Fig. 1. Photograph of the commercially available flexible graphene-based conducting material and (b) a photograph of the prototype (with dimensions) of the fabricated graphene-based patch antenna.

I. INTRODUCTION

Modern wireless communication systems are being required to operate in evermore complicated environments. This is because these systems are being applied to problems that (1) involve surfaces that change shape with time, (2) include wearable networks and/or (3) are subjected extreme environmental temperature/pressure changes [1]. Because of the wireless nature of these systems, the antenna is a major part of the design and conformal antennas have the potential to overcome some of the aforementioned difficulties. However, one drawback of using a traditional conformal antenna can be the copper conductors. This is because of the weight and the potential for copper failure due to repetitive bending deformations, which was noticed in the work reported in [1]. To mitigate some of these issues, a flexible graphene-based conducting material [2] is being explored as a possibility of replacing the etched copper in a conformal antenna. A picture of the material is shown in Fig. 1(a). However, an efficient manufacturing process for fabricating complicated antenna designs based on this material is required. Thus, the objective of this paper is to present a suitable manufacturing process and introduce the graphene-based microstrip patch antenna shown in Fig. 1(b). This provided a standard design that has been well understood and was marginally complicated.

An initial investigation on graphene patch antennas at microwave frequencies was reported in [3] and later discussed in [4]. The authors of these papers summarized the benefits of graphene as a mono-atomic structure, whereas the presented work builds on these ideas and explores the potentials of using graphene-based conductors that are commercially available and much thicker than a single layer for printed antenna purposes.

II. THE MANUFACTURING PROCESS

Several steps are required to prepare the graphene-based conductor material in Fig. 1(a) for manufacturing. Initially, the temporary spray adhesive (Sulky KK 2000 [6]) shown in Fig. 3 is applied to the top surface of a 100 μm thick sheet of paper and bonded to the bottom surface of the graphene-based conducting material (Fig. 2). Next, the same spray adhesive is applied to the bottom surface of a 100 μm thick transparency film and bonded to the top surface of the graphene-based conducting material (also shown in Fig. 2). As a result, the three layers are bonded together for manufacturing, giving the 3-layer structure shown in Fig. 2. The 3-layer structure is then placed on the adhesive cutting mat shown in Fig. 3. The graphene-based conducting layer is now ready for cutting out the shape of the microstrip patch. It should be noted that the three layers were pressed together by hand and the temporary adhesive was cured at room temperature for 2 - 3 minutes before cutting. This allowed for the separation of the three layer after the cutting process was completed. Finally, the layout of the patch is defined in the software included with the micro-cutter, which is discussed in the next section.

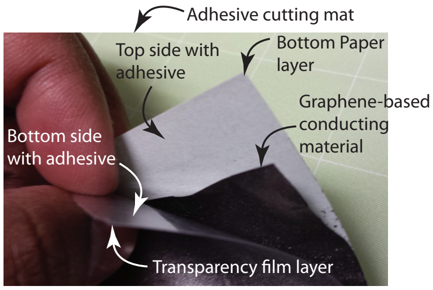

Fig. 2. Photograph of the flexible graphene-based conductors placed in between different layers to fabricate the antenna.

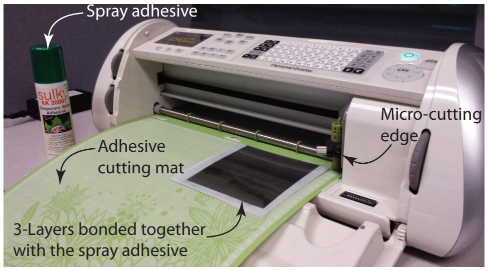

Fig. 3. Photograph of the commercial micro-cutter [5] and the temporary adhesive spray [6].

III. DESIGN AND EVALUATION OF THE PROTOTYPE MICROSTRIP PATCH ANTENNA

To develop the patch antenna in Fig. 1, the conductive properties of the graphene-based material [2] were determined by measuring the S-parameters of a known printed microstrip transmission line that used the graphene-based conductors instead of copper. The conductivity was determined to be σ = 1.94×10^5 S/m for a thickness of 25 μm. Next, this thickness and conductivity were used in the commercial simulation software ADS [7] to determine the geometry of the patch antenna shown in Fig. 1(b). The simulated S-parameters are shown in Fig. 5. The resonant frequency was predicted to be 2.4 GHz and the grounded FR4 substrate was 1.5 mm thick with ε_r = 4.5 (the bottom layer of the FR4 was 0.5 oz copper). Then, this geometry was drawn in the software included with the micro-cutter and used to cut out the graphene-based conducting layers shown in Fig. 4. As a final step, the graphene-based conducting patch was removed from the 3-layer structure and attached to the FR4 substrate. The result was the prototype shown in Fig. 1(b). The center conductor of the SMA connector was attached to the patch with conductive epoxy. Next, the $|S_{11}|$ values of the prototype were measured in an anechoic chamber and are shown in Fig. 5. The resonant frequency was 2.6 GHz.

Several comments can be made about the results in Figs. 1(b) and 5. Closer observation of the edges in Fig. 1(b) shows that the micro-cutter blade tends to wobble slightly during cutting, which may be the cause of the differences in the S-parameters in Fig. 5. This behaviour could be due to the blade depth setting and further exploration is planned in the near future. On the other hand, the results in Fig. 5 show that a new patch antenna can be developed using this flexible graphene-based material, which may be difficult to otherwise manufacture with the existing PCB techniques.

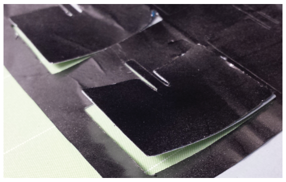

Fig. 4. Fabrication step to separate the flexible graphene-based conducting layer of the patch from the protected layers.

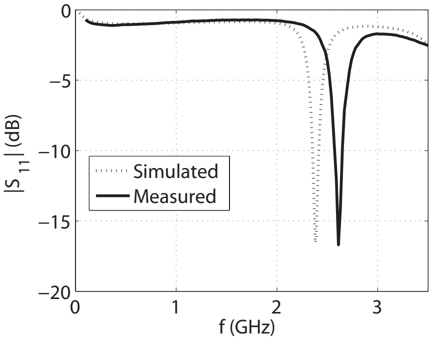

Fig. 5. Measured and simulated $|S_{11}|$ values of the prototype graphene-based antenna.

IV. CONCLUSION

An initial investigation of a new patch antenna that uses flexible graphene-based conductive material was introduced in this paper. Additionally, a unique process of manufacturing these patch antennas with the graphene-based material was presented. The design was simulated in ADS using extracted material properties, and a prototype was fabricated and tested. Overall, fair agreement between the $|S_{11}|$ values was observed and it was demonstrated that moderately complex geometries could be manufactured using a simple and cost-efficient commercially available micro-cutter.

REFERENCES

[1] B. D. Braaten, M. A. Aziz, S. Roy, S. Nariyal, I. Irfanullah, N. F. Chamberlain, M. T. Reich and D. E. Anagnostou, "A Self-Adapting Flexible (SELFLEX) Antenna Array for Changing Conformal Surface Applications," *IEEE Trans. Antennas Propag.,* vol. 61, no. 2, Feb., 2013, pp. 655 - 665.
[2] Graphene Laboratories Inc., [online] www.graphene-supermarket.com
[3] J. S. Gomez-Diaz and J. Perruisseau-Carrier, "Microwave to THz Properties of Graphene and Potential Antenna Applications," *Proc. of ISAP2012,* OCT. 29 - Nov. 2, 2012, Nagoya, Japan, pp. 239 - 241.
[4] J. Perruisseau-Carrier, "Graphene for Antenna Applications: Opportunities and Challenges from Microwaves to THz," *2012 Loughborough Antennas and Propag. Conf,* Nov. 12 - 13, Loughborough, UK.
[5] Cricut, Inc., [online], www.cricut.com.
[6] Sulky, Inc., [online], www.sulky.com.
[7] Advanced Design System (ADS) by Agilent Technologies, [online] www.agilent.com.

United States Patent [19]

Murakami et al.

[11] **Patent Number: 4,853,703**

[45] **Date of Patent: Aug. 1, 1989**

[54] **MICROSTRIP ANTENNA WITH STRIPLINE AND AMPLIFIER**

[75] Inventors: **Yuichi Murakami; Kiyokazu Ieda,** both of Tokyo, Japan

[73] Assignee: **Aisin Seiki Kabushikikaisha,** Aichi, Japan

[21] Appl. No.: **26,705**

[22] Filed: **Mar. 17, 1987**

[30] **Foreign Application Priority Data**

Mar. 17, 1986 [JP] Japan 61-58692

[51] **Int. Cl.**^{4} **H01Q 1/38**

[52] **U.S. Cl.** .... **343/700 MS;** 343/701

[58] **Field of Search** .... 343/700 MS, 701, 846; 333/247; 455/327, 333, 325, 291, 282; 330/277, 1 R, 286

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,508,269	4/1970	Snyder	343/701
4,054,874	10/1977	Oltman	343/700 MS
4,203,116	5/1980	Lewin	343/700 MS
4,276,558	6/1981	Ho et al.	333/247
4,383,260	5/1983	Ryan	343/701
4,491,809	1/1985	Noda et al.	330/277
4,612,512	9/1986	Powell et al.	330/286
4,724,443	2/1988	Nysen	343/700 MS
4,761,654	8/1988	Zaghloul	343/700 MS

FOREIGN PATENT DOCUMENTS

77306	5/1983	Japan	343/700 MS
90810	5/1983	Japan	330/277
137314	8/1983	Japan	330/277
59-52708	4/1984	Japan .	
77701	5/1984	Japan	333/247
59-91016	6/1984	Japan .	
112701	6/1984	Japan	333/247
171206	9/1984	Japan	333/247

Primary Examiner—Rolf Hille
Assistant Examiner—Michael C. Wimer
Attorney, Agent, or Firm—Sughrue, Mion, Zinn, Macpeak & Seas

[57] **ABSTRACT**

A high frequency amplifier is mounted in the immediate vicinity of a feeder of a microstrip antenna. A microstrip antenna includes a dielectric member and a grounded conductor member which is applied thereto. A second dielectric member is disposed on the side of the grounded conductor member which faces away from the dielectric member, and the high frequency amplifier is mounted on the surface of the second dielectric member which faces away from the grounded conductor member, and is connected to the feeder of the microstrip antenna. A transmission distance between the feeder and the high frequency amplifier, which represents an initial amplifier stage, is greatly reduced, thus drastically reducing disturbances caused by external noises and improving the noise figure of a receiving equipment.

8 Claims, 4 Drawing Sheets

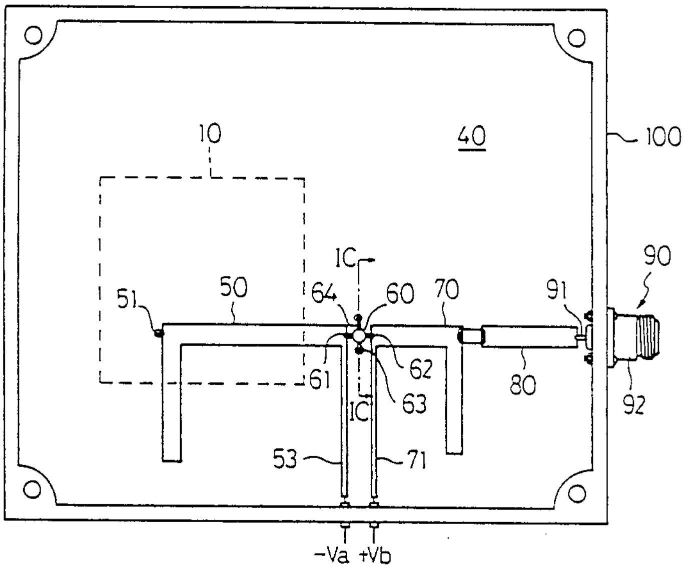

Fig.1a

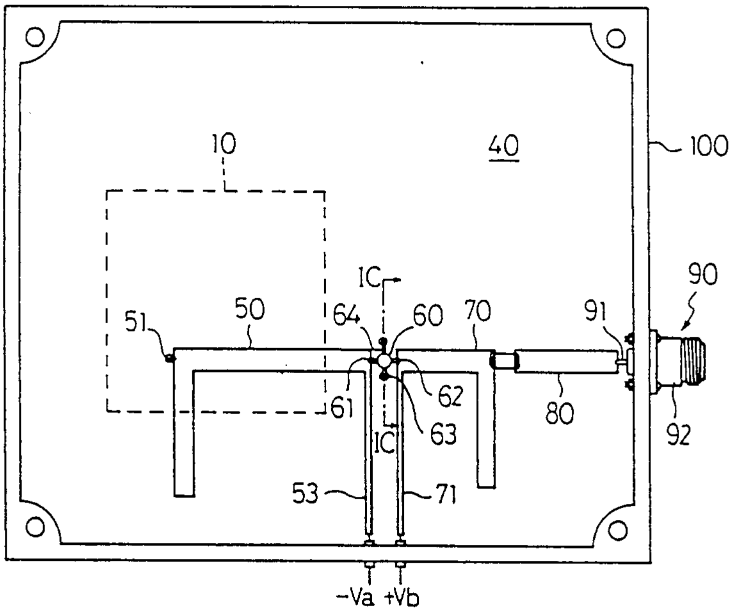
10
40
100
IC
90
51
50
64
60
70
91
61
62
63
80
92
IC
53
71
-Va
+Vb

Fig.1b

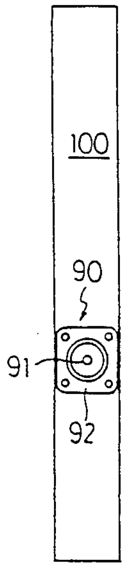

Fig.1c

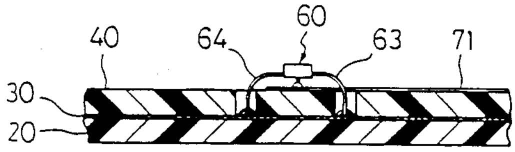
60
40
64
63
71
30
20

Fig. 1d

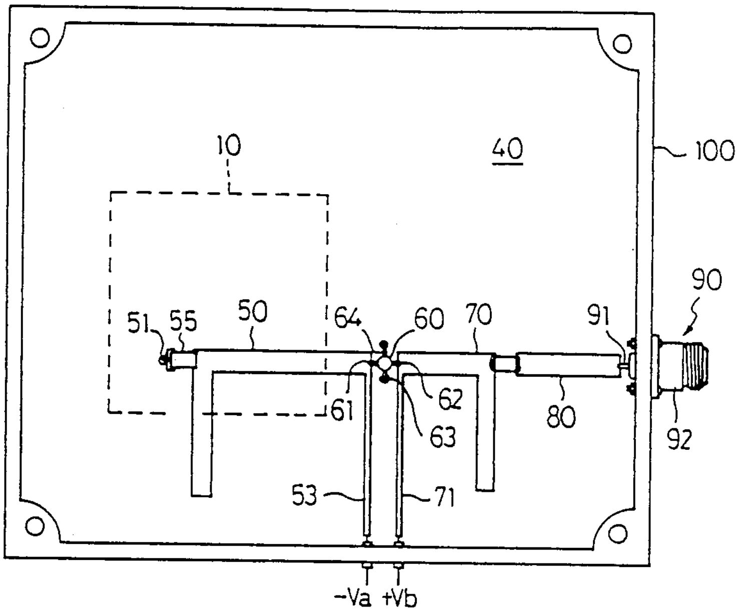
10
40
100
90
51
55
50
64
60
70
91
61
62
63
80
92
53
71
-Va
+Vb

Fig.2a

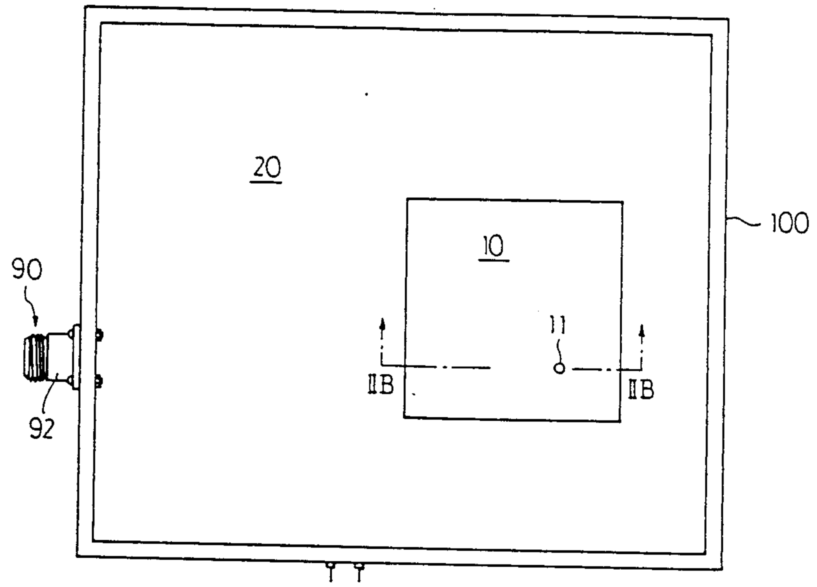
20
10
100
90
11
IIB
IIB
92

Fig.2b

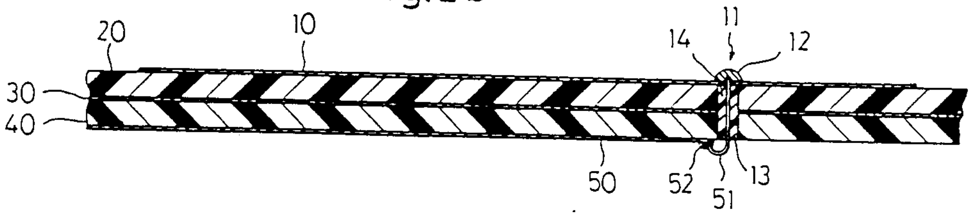
11
20
10
14
12
30
40
50
52
51
13

Fig. 3
PRIOR ART

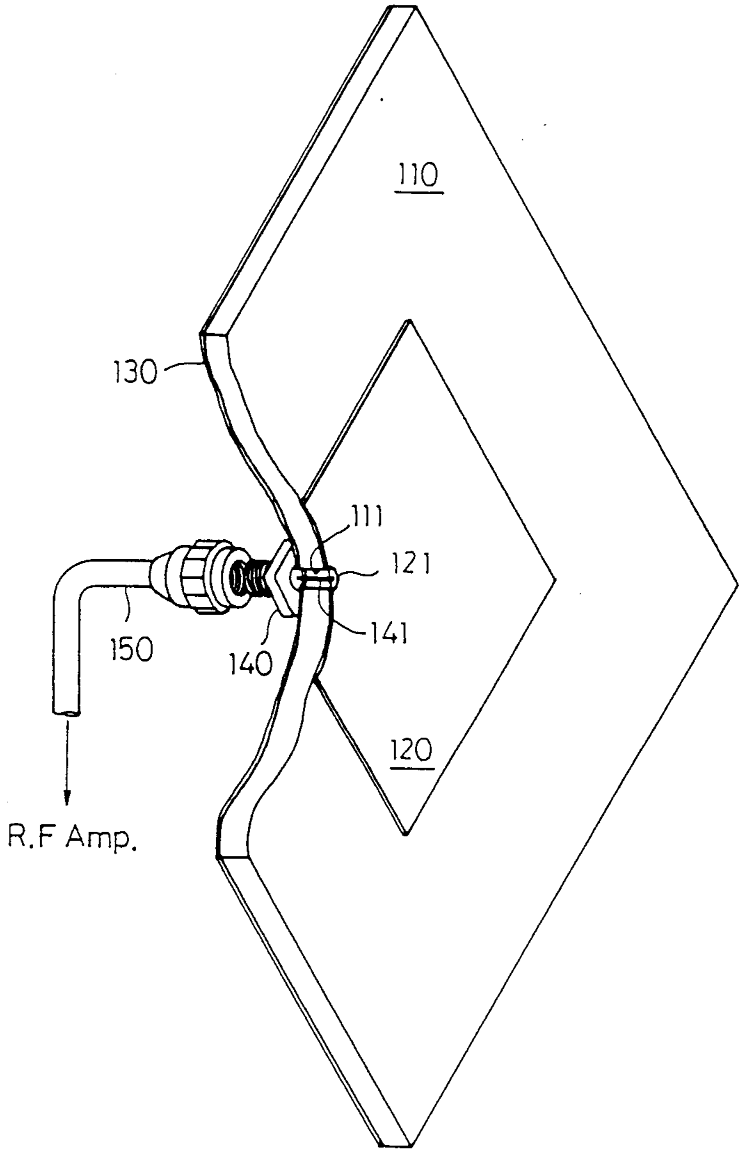
110
130
111
121
141
150
140
120
R.F Amp.

MICROSTRIP ANTENNA WITH STRIPLINE AND AMPLIFIER

BACKGROUND OF THE INVENTION

The invention relates to an antenna device including a microstrip antenna.

A microstrip antenna comprises a dielectric member, a conductor member mounted on the dielectric member, and a ground conductor member mounted on the opposite surface of the dielectric member from the conductor member, and represents an antenna which utilizes a radiation loss of an open plane resonance circuit. Attention is now being directed to such microstrip antenna due to its low profile, reduced weight, compactness and ease of manufacture.

FIG. **3** illustrates one form of conventional microstrip antenna device. As shown, the device comprises a dielectric plate **110**, one surface of which is applied with a radiating conductor sheet **120** formed by a copper foil while the opposite surface is applied with a ground conductor sheet **130** again formed by a copper foil. The device includes a feeder **121** in which a small hole **111** is formed extending through the dielectric plate **110**, conductor sheet **120** and ground sheet **130**. A connector **140**, or more precisely, an external conductor associated therewith, is soldered to the ground sheet **130**, and an internal conductor or core of the connector **140** is connected to a gold plated wire **141** which is soldered to the feeder portion of the radiating conductor sheet **120**. The hole **111** is filled with an insulating material, not shown, thereby insulating the wire **141** from the ground sheet **130**. The connector **140** is connected with a coaxial cable **150** which is in turn connected to a high frequency amplifier of a receiver unit.

In a receiver unit which is herein understood to be a circuit portion extending from the antenna feeder to an output device such as a loudspeaker, a cathode ray tube or the like, disturbances caused by external noises which are applied in a region between the antenna feeder and a first stage amplifier or the high frequency amplifier of the receiver unit have a great influence upon the noise figure of the receiving unit because they are amplified by the high frequency amplifier and every amplifier in an electrical path subsequent thereto. Obviously, the greater the length between the feeder and the initial amplifier, the greater the influence of these disturbances.

By way of example, when an antenna device such as shown in FIG. **3** is assembled into an outside panel of a vehicle with a coaxial cable **150** being used to provide a connection between the connector **140** and an input terminal of a receiver which is disposed within an instrument panel of the vehicle, a substantial length is required of the cable **150** and a received signal from the feeder **121** experiences an increased attenuation during its transmission through the connector **140** and through the cable **150** due to attenuation coefficients of the connector **140** and the coaxial cable **150**. On the other hand, the coaxial cable is subject to disturbances of external noises over its full region including the connector **140**, and this considerably degrades the noise figure of the receiving unit, i.e. between the antenna feeder **121** and an output source such as a loudspeaker or cathode ray tube unit.

SUMMARY OF THE INVENTION

It is an object of the invention to reduce disturbances caused by external noises which occur between a feeder of a microstrip antenna and a high frequency amplifier as much as possible.

The above object is accomplished in accordance with the invention by providing a second dielectric member on the opposite side of the ground conductor member from and in opposing relationship with the first mentioned dielectric member of the strip antenna, and mounting a high frequency amplifier which is connected to the feeder of the microstrip antenna on the opposite surface of the second dielectric member from the ground conductor member.

With this arrangement, the distance of transmission between the feeder of the microstrip antenna and the initial amplifier or the high frequency amplifier can be greatly reduced, thus substantially reducing any effect of disturbances caused by external noises.

Other objects and features of the invention will become apparent from the following description of an embodiment thereof with reference to the drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1***a* is a rear view of an antenna device according to an embodiment of the invention;

FIG. **1***b* is a right-hand side elevation of the device shown in FIG. **1***a*;

FIG. **1***c* is a cross section taken along the line IC—IC shown in FIG. **1***a*;

FIG. **1***d* is a rear view of a modification of the antenna device;

FIG. **2***a* is a front view of the antenna device shown in FIG. **1***a*;

FIG. **2***b* is a cross section taken along the line IIB—IIB shown in FIG. **2***a*; and

FIG. **3** is a perspective view, partly broken away, of a conventional antenna device.

DETAILED DESCRIPTION OF EMBODIMENTS

Initially referring to FIG. **2***a*, an antenna device according to the invention includes a radiating conductor sheet **10** which is applied to a first dielectric plate **20**. In this embodiment, the sheet **10** comprises a copper foil having a thickness of 35 μm applied to the first dielectric plate **20** which is formed by PTFE (polytetrafluoroethylene), or commonly referred to as "Teflon", glass substrate having a thickness of 1.588 mm.

FIG. **2***b* shows a cross section taken along the line IIB—IIB shown in FIG. **2***a*. It will be noted that a ground conductor sheet **30** is applied to the entire surface of the first dielectric plate **20** which is located on the opposite side from the radiating conductor sheet **10**. In this manner, the combination of the radiating conductor sheet **10**, the first dielectric plate **20** and the ground conductor sheet **30** forms a microstrip antenna for use in a frequency range from 1.5 to 2 GHz. A second dielectric plate **40** is applied to the outer surface of the ground sheet so as to hold the latter between the both dielectric plates, thus forming a laminar structure. In the embodiment, the ground sheet **30** comprises a copper foil having a thickness of 35 μm and the second dielectric plate **40** comprises PTFE glass substrate in the same manner as the first dielectric plate **20**.

A small hole **13** extends through the radiating conductor sheet **10**, the first dielectric plate **20**, the ground conductor sheet **30** and the second dielectric plate **40**, which form together a laminar structure, and a metal golded wire **51** is supported therein as insulated from the ground sheet **30** by an insulating material **14**. The end of the wire **51** located nearer the radiating conductor sheet is soldered to a feeder **11** thereof as shown at **12**. The other end of the wire **51** is soldered to a strip conductor **50**, to be mentioned later, as shown at **52**.

FIG. 1*a* shows a rear view of the antenna device. As will be noted, the strip conductor **50** which forms an input matching circuit, a field effect transistor **60**, which may be 2SK571, a strip conductor **70** which forms an output matching circuit and a 50 Ω line strip conductor **80** are applied to the surface of the second dielectric plate **40** which is opposite from the ground conductor sheet **30**. Each of the strip conductors **50**, **70** and **80** comprises a copper foil having a thickness of 35 μm. A combination of the strip conductor **50**, the second dielectric plate **40** and the ground conductor sheet **30** forms a strip line, which forms an input matching circuit which matches the feeder **11** of the microstrip antenna to a gate electrode **61** of the transistor **60**. A combination of the strip conductor **70**, the second dielectric plate **40** and the ground conductor sheet **30** forms another strip line, which forms an output matching line which matches a drain electrode **62** of the transistor **60** to a line having a characteristic impedance of 50 μ.

A first bias line **53** which is integral with the strip conductor **50** applies a bias voltage of −Va to the gate electrode **61** while a second bias line **71** which is integral with the strip conductor **70** applies a bias voltage of +Vb to the drain electrode **62** of the transistor **60**. The transistor **60** has two source electrodes **63** and **64** which are both connected to the ground conductor sheet **30** as shown in FIG. 1*c* which indicates a cross section taken along the line IC—IC shown in FIG. 1*a*. In this manner, the transistor **60** forms a high frequency amplifier together with the input and the output matching circuit. In the present embodiment, the amplifier has a gain of approximately 15 dB. A combination of the strip conductor **80**, the second dielectric plate **40** and the ground conductor sheet **30** forms a strip line having a characteristic impedance of 50 Ω.

The antenna device constructed in the manner mentioned above is fixedly mounted on a flange, not shown, of a frame **100**, which comprises an electrically conductive material and which is connected to the ground conductor sheet **30**. It will be seen that a connector **90** having a characteristic impedance of 50 Ω is secured to the right-hand side of the frame **100**, as viewed in FIG. 1*a*, and includes an inner conductor **91** which is connected to the strip conductor **80** while its outer conductor **92** is threadably engaged with the frame **100**. In other words, the outer conductor **92** is connected to the ground conductor sheet **30** through the frame **100**.

A conductive cover member, not shown, is mounted on the frame **100** in opposing relationship with a surface of the second dielectric sheet **40** on which the strip conductors **50**, **70**, **80** and FET **60** are applied. The combination of the cover member and the frame which is formed of a conductive material constitutes together a shielded casing, preventing the high frequency amplifier and 50 Ω strip line from being subject to disturbances caused by external noises. A cover member, not shown, of a dielectric material is mounted on the frame **100** in opposing relationship with the radiating conductor sheet **10**, functioning to protect the latter from dusts.

In the present embodiment, a gate bias −Va for FET **60** is applied to the first bias line **53** which is integral with the strip conductor **50** while a grounded conductor sheet **30** is connected to the frame **100**. Accordingly, FET **60** may be destroyed in the event of occurrence of an electrical short-circuit between the radiating conductor sheet **10** and the frame **100** for some abnormality. As mentioned previously, a cover, not shown, is mounted on the front side of the frame **100** in order to avoid such interferences, but such interference may occur when the cover is not yet mounted as during a testing or mounting of the antenna device. Accordingly, in a modification of the invention shown in FIG. 1*d*, a chip capacitor **55** is interposed between the gold plated wire **51** and the strip conductor **50** and acts as a d.c. blocking capacitor which protects FET **60** by blocking a flow of a d.c. current which may be developed in the event of a short-circuit. To provide a sufficiently low impedance in a frequency range from 1.5 to 2 GHz in which the antenna device of the embodiment is used, the capacitor **55** has a capacitance on the order of 1,000 to 2,000 pF.

In the described embodiments, the radiating conductor sheet **10**, the first dielectric sheet **20** and the grounded conductor sheet **30** form together a single microstrip antenna, but it will be evident that the invention can be equally applied to a microstrip antenna array having a plurality of radiating conductor sheets, with an increased number of corresponding high frequency amplifiers associated therewith.

As described, according to the invention, a high frequency amplifier is located in the immediate vicinity of a feeder of a microstrip antenna, thus greatly reducing a transmission distance between the feeder and the amplifier and thus drastically reducing disturbances caused thereon by external noises.

What we claimed is:

1. An antenna device comprising:

a microstrip antenna formed on a first dielectric member, said microstrip antenna including feeder means and further including a radiating conductor member disposed on a first major surface of said first dielectric member, and a grounded conductor member disposed on an opposing major surface of said first dielectric member;

a second dielectric member disposed on an opposite side of said grounded conductor member from said first dielectric member, such that said grounded conductor member is sandwiched between first and second dielectric members; and

a high frequency circuit mounted on a surface of said second dielectric member which faces away from said grounded conductor member, said high frequency circuit comprising a high frequency amplifier and a strip conductor member constituting a strip line with said second dielectric member and said grounded conductor member for electrically matching said high frequency amplifier with said feeder means of said microstrip antenna.

2. An antenna device according to claim **1**, in which said high frequency amplifier includes input matching means and output matching means.

3. An antenna device according to claim **2**, in which said input matching means is connected to the feeder means of said microstrip antenna.

4. An antenna device according to claim **3**, in which said input matching means comprises a strip line including a first conductor line member, said second dielectric member and said grounded conductor member.

5. An antenna device according to claim **4**, in which said input matching means comprises a d.c. blocking means and is connected to the feeder means of said microstrip antenna through an electrical conduction path provided by said blocking means.

6. An antenna device according to claim **5**, in which said blocking means comprises a capacitor.

7. An antenna device according to claim **2**, in which said output matching means comprises a strip line including a second conductor line member, said second dielectric member and said grounded conductor member.

8. An antenna device comprising:

a microstrip antenna formed on a first dielectric member, said microstrip antenna including a radiating conductor member disposed on a first major surface of said first dielectric member, and a grounded conductor member disposed on an opposing major surface of said first dielectric member;

a second dielectric member disposed on an opposite side of the grounded conductor member form said first dielectric member, such that said grounded conductor member is sandwiched between said first and second dielectric members; and

a high frequency circuit mounted on a surface of said second dielectric member which faces away from said grounded conductor member, said high frequency circuit being electrically connected to a feeder of said microstrip antenna, such that an electrical conduction path is provided between said high frequency circuit and said radiating conductor member, wherein said high frequency circuit comprises a high frequency amplifier and a strip conductor member constituting a strip line with said second dielectric member of electrically matching said high frequency amplifier with said feeder of said microstrip antenna.

* * * * *

United States Patent [19]

Phillips, Sr.

[11] **Patent Number:** **4,685,047**

[45] **Date of Patent:** **Aug. 4, 1987**

[54] **APPARATUS FOR CONVERTING RADIO FREQUENCY ENERGY TO DIRECT CURRENT**

[76] Inventor: **Raymond P. Phillips, Sr.,** 1200 S. Frances, Terrell, Tex. 75160

[21] Appl. No.: **886,138**

[22] Filed: **Jul. 16, 1986**

[51] **Int. Cl.⁴** **H02M 7/06**

[52] **U.S. Cl.** **363/126**; 307/151

[58] **Field of Search** 363/125, 126; 307/151

[56] **References Cited**

U.S. PATENT DOCUMENTS

649,621	11/1901	Tesla	455/39
685,957	5/1902	Tesla	455/619
3,434,678	3/1969	Brown et al.	307/151
3,445,767	5/1969	Beard	324/78
3,678,365	8/1972	Fremiot	307/151
3,723,891	3/1973	Whitely	307/246
3,760,257	9/1973	Fletcher et al.	307/151
3,784,845	1/1974	Haas	307/519
3,792,290	2/1974	Brocker	307/522
3,944,935	3/1976	Plant	307/519
4,079,268	3/1978	Fletcher et al.	307/151
4,079,325	3/1978	Mawhinney et al.	307/261
4,085,377	4/1978	Turski et al.	307/519
4,195,237	3/1980	Meslener	307/261
4,297,557	10/1981	Tyler et al.	219/10.55 B
4,360,741	11/1982	Fitzsimmons et al.	307/151
4,360,865	11/1982	Yasumura et al.	363/126
4,393,441	8/1983	Enge	363/126

Primary Examiner—Patrick R. Salce
Assistant Examiner—Jeffrey Sterrett
Attorney, Agent, or Firm—Millen & White

[57] **ABSTRACT**

Apparatus for and methods of converting radio frequency energy into direct current for generating electric power includes a dipolar antenna for receiving radio frequency energy and a circuit connected thereto for converting the radio frequency energy to direct current. The circuit has a positive output line connected to one pole of the antenna and a negative output line connected to the other pole of the antenna. A positive transmitting diode is in the positive output line and a negative transmitting diode is in the negative output line. First and second bus lines and a pair of tuned circuits of opposite polarity couple the positive output line and negative line to the bus line with one of the bus lines being connected to ground. Each tuned circuit includes a first bridging line connecting the positive output line to the first and second ground lines and a second bridging line connecting the negative output line to the first and second ground lines. Each bridging line has a diode therein oriented at a polarity which is reverse with respect to the input diode. The bridging lines of each tuned circuit are connected to one another by an inductance and have capacitors disposed between the diode and the bus lines. A direct current device is connected to the positive line of the circuit.

11 Claims, 1 Drawing Figure

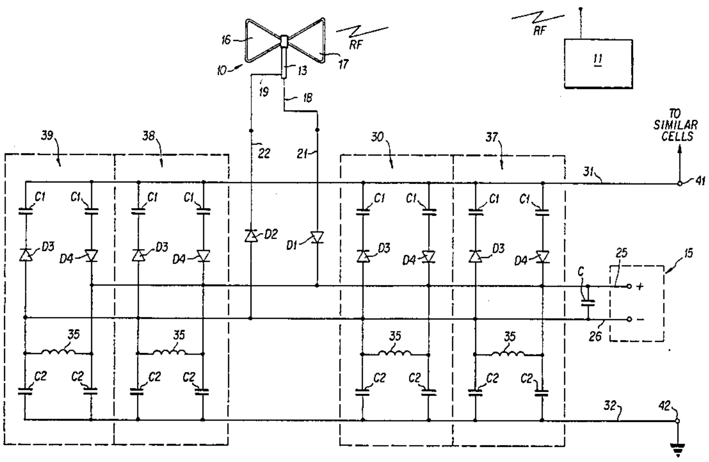

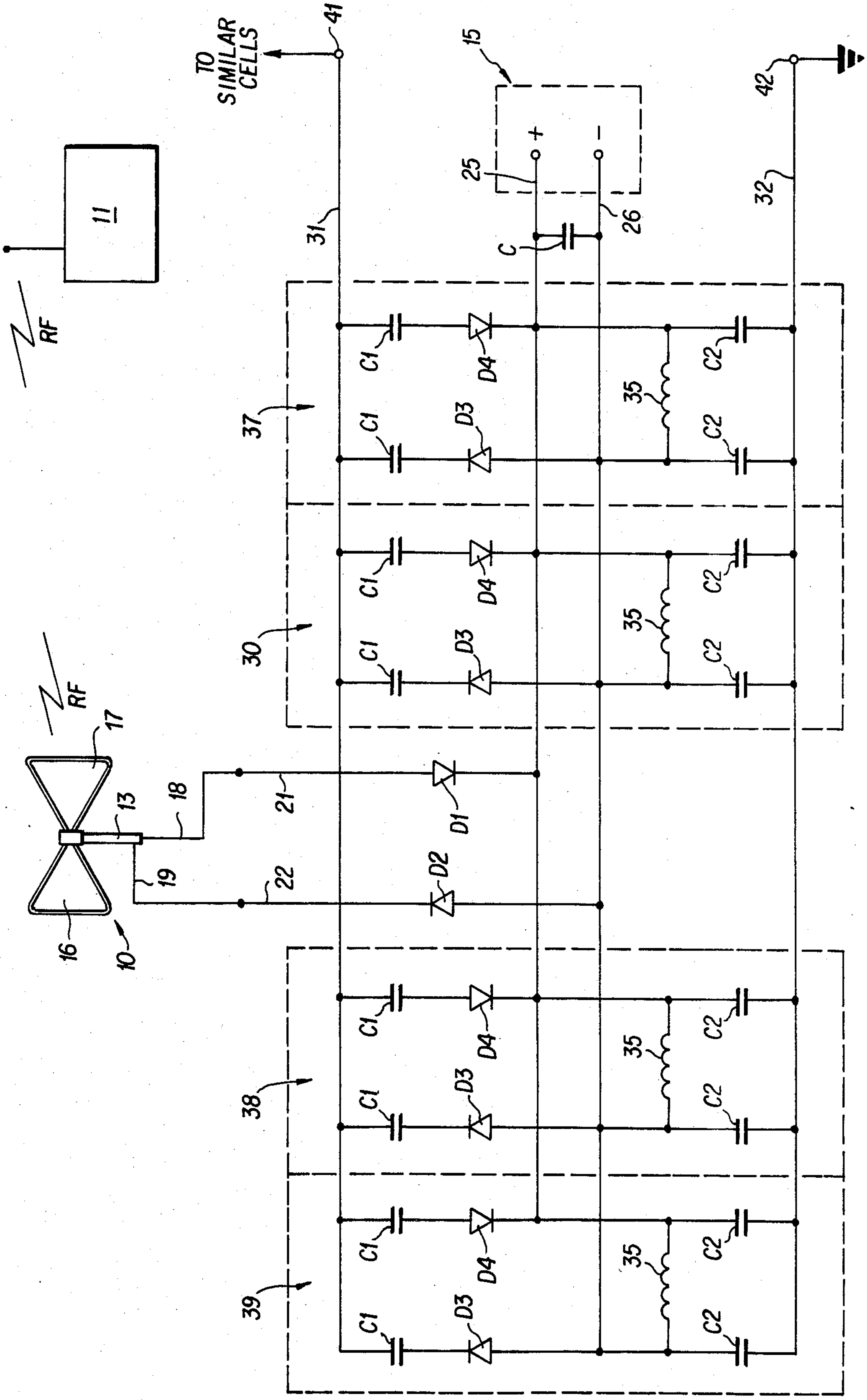
TO SIMILAR CELLS
+
−
C
RF
C1
C2
D3
D4
RF
D1
D2

APPARATUS FOR CONVERTING RADIO FREQUENCY ENERGY TO DIRECT CURRENT

BACKGROUND OF THE INVENTION

1. Field of the Invention

The instant invention relates to methods of an apparatus for converting radio frequency energy to direct current; more particularly, the instant invention relates to methods and apparatus for converting radio frequency energy to direct current wherein the direct current is of sufficient magnitude to power devices such as battery charges and electric motors without the use of amplification.

2. Technical Considerations in Prior Art

There has long been interest in technology directed to transmitting electrical energy over a distance without utilizing conductors, such as wire conductors. Development of such a technology has enormous potential. This was first recognized by Nikola Tesla who in 1899 constructed a 200 foot Tesla coil rated at 300 kilowatts and 150 kilocycles. Tesla hoped to set up standing waves of electrical energy around the whole surface of the earth, so that receiving antennas set at optimum points could tap the power when needed. Tesla was able to light hundreds of lamps at a distance of about 40 kilometers with his device without utilizing electrical conductors. The scheme has generally remained a scientific curiosity but has provided the initial groundwork for current developments wherein attempts are being made to transmit power by microwaves. However, power transmitted by microwaves is envisioned in the form of a beam of very high intensity which is focused from a microwave generator to a receiving antenna. This technology is envisioned as being used for many types of purposes such as transmitting microwave energy collected from gigantic solar power satellites and "star wars"-type weapons systems. However, the focused microwave beam is not suitable for many applications in that the beam must be directed toward a receiving antenna and cannot be transmitted through most objects, including living objects, without destroying the objects.

The instant invention relies on converting energy from standing waves which are emitted from radio frequency antennas in the RF range rather than the microwave range. Of particular interest are very low frequencies which are not used in communications and are available for transmitting power. Also of interest with respect to the instant invention are the low frequency waves emitted by the earth due to pulsation thereof caused by its magnetic field. These low frequency standing "earth" waves can be picked up by receivers tuned thereto.

SUMMARY OF THE INVENTION

It is an object of the invention to provide new and improved methods of an apparatus for converting radio frequency currents to direct current for practical uses other than communications, wherein the direct current energy converted from the radio frequency input energy does not require amplification.

Upon further study of the specification and appended claims, further objects and advantages of this invention will become apparent to those skilled in the art.

The instant invention contemplates an RF antenna for receiving radio waves. The RF antenna connected to a circuit configured in accordance with the principles

of the instant invention to convert the RF signals to direct current. The radio frequency signals received by the antenna are transmitted to first and second leads, with one lead being rectified to transmit positive voltage and the other lead being rectified to transmit negative voltage. The positive voltage lead being connected directly to a positive output line and the negative voltage lead being connected directly to a negative output line. The positive output line is connected to a pair of bus lines through a first pair of capacitors, while the negative output line is connected to the pair of bus lines by a second pair of capacitors. Disposed between the first bus line and the positive output line is a reverse diode of negative polarity, while disposed between the negative output line and first bus line is a reverse diode of positive polarity. The positive and negative output lines are connected to one another through an inductance which is in parallel with the capacitors of the first and second pair connected between the second bus line and the positive and negative output lines.

In accordance with one embodiment of the invention the afore-described circuit is duplicated for each positive and negative output line. In accordance with another embodiment of the invention, the afore-described circuitry is coupled to additional circuits identically configured in order to increase the direct current output of the arrangement.

In accordance with a further configuration of the invention, the antenna utilized is a dipolar antenna of aluminum wire arranged in a "butterfly" configuration.

The instant invention further contemplates the method of utilizing the afore-described elements so as to generate direct current having sufficient power to perform tasks such as charging batteries, lighting lamps and powering direct current electric motors without the use of amplifiers.

BRIEF DESCRIPTION OF THE DRAWING

Various other objects, features and attendant advantages of the present invention will be more fully appreciated as the same becomes better understood when considered in connection with the accompanying drawing, in which like reference characters designate the same or similar parts throughout the several views, and wherein:

There is shown a diagram of a circuit in accordance with the instant invention in combination with a driven device and a dipolar antenna which receives radio frequency waves which are converted to DC current for powering the driven device.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring now to the drawing, there is shown a dipolar antenna, designated generally by the numeral **10**, which receives radio frequency waves from an RF transmitter **11**. The radio frequency waves are transmitted to a radio frequency to direct current converting circuit, designated generally by the numeral **12**, through a coaxial cable **13** and DC current from the circuit **12** is used to power an output device, designated generally by the numeral **15**, which may for example by a battery charger DC motor or lighting device. The circuit **12** has no other power inputs other than the radio frequency energy transmitted thereto by the antenna **10** and therefore includes no amplifiers for amplifying the RF energy.

The source of radio frequencies convertible to direct current by the circuit shown may include sources of high frequency, low frequency (LF), very low frequency (VLF) and extremely low frequency (ELF) radio waves as well as seismic vibration of the earth's magnetic fields.

Preferably, the dipolar antenna **10** is formed of two triangular loops of aluminum wire **16** and **17**, one of which is connected to the annular conductor **18** of the coaxial cable **13** and the other of which is connected to the center conductor **19** of the coaxial cable. The size of the bipolar antenna **10** is dependent on the particular application to which it is put. In accordance with one embodiment of the invention, the antenna **10** is approximately 12 inches in width and 18 inches in length. Such an antenna is used to receive five watt energy, such as that generated by a walkie-talkie or citizen-band radio.

The annular conductor **18** of the coaxial cable **13** is connected to a positive lead **21** of the circuit **12**, while the center conductor **19** of the coaxial cable is connected to a negative lead **22** of the circuit. A positive transmitting diode D1 is disposed between the lead **21** and the remainder of the circuit **12** while a negative transmitting diode D2 is disposed between the lead **21** to a positive output line **25** while the negative diode D2 is connected to a negative output line **26**. Accordingly, the positive voltages with respect to ground are produced on output line **25** and negative voltages with respect to ground are produced on output line **26**.

In order to provide a DC output of sufficient power, a plurality of inductance-capacitance, RF, tuned circuits **30**, each forming a positive cell, or a negative cell, are utilized for connecting the positive output line **25** and negative output line **26** to first and second bus lines **31** and **32**, respectively. Bus line **32** is connected to ground while bus line **32** can be connected to circuits similar to circuit **12**. The positive output line **25** is connected by a first bridging line **33** to the first and second bus lines **31** and **32** while the negative output line **26** is connected by a second bridging line **34** to the first and second bus lines. The bridging line **33** has capacitors C1 and C2 disposed between the positive output line **25** and the first and second bus lines **31** and **32**, while the bridging line **34** also has capacitors C1 and C2 dispensed between the negative output line **26** and the first and second bus lines **31** and **32**. Connected between the bridging lines **33** and **34**, is an inductor **35** which serves as an RF choke, while disposed between the positive output line **25** and the capacitor C1 there is a negative polarity diode D4 referred to herein as a bridging diode and disposed between the negative output line **26** and capacitor C1 in line **34** there is a positive polarity diode D3 referred to herein as a bringing diode. As is seen of the drawing, the RF tuned circuit cell **30** is repeated a plurality of times. In the specific example shown, the circuit **12** has separate cells **30**, **37**, **38** and **39**. The cells **30** and **38** are of opposite polarity and balance one another while the cells **37** and **39** of opposite polarity and also balance one another. In order for the system to function, a pair of opposite polarized cells must be utilized. The particular number of cells **30** and the value of the components thereof are determined by the configuration of the dipolar antenna **10** and the power and frequency of the RF transmitter **11**.

The radio frequency to direct current conversion circuit **12** may itself be connected to a duplicate circuit via pins **41** so as to provide additional direct current output on lines similar to positive output line **25** and negative output line **26** the output lines may be connected together in order to boost the total output of the system.

An operative embodiment of the invention utilizes the following elements:

Diodes D1, D2, D3 and D4—Germanium Diodes, Archer **1** N34A,Catalog #1123.

Inductor **35**-47 Milli henry R. F. Choke

Capacitors C1 and C2-0.47 Pico Farads at 200 volts

Coaxial Cable **13**—50 ohms

Dipolar Antenna **10**—aluminum wire triangular loops approximately 12 inches by 18 inches.

From the foregoing description, one skilled in the art can easily ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

What is claimed is:

1. Apparatus for converting radio frequency energy into direct current for generating electric power, the apparatus comprising:

a dipolar input for receiving radio frequency energy;

a positive output line connected to one pole of the dipolar input and a negative output line connected to the other pole of the dipolar input;

a positive transmitting input diode means in the positive output line and a negative transmitting input diode means in the negative output line;

first and second bus lines and a pair of tuned circuits of opposite polarity coupling the positive output line and negative output line to the bus lines, one of the bus lines being connected to ground;

each tuned circuit including a first bridging line connecting the positive output line to the first and second bus lines and a second bridging line connecting the negative output line to the first and second bus lines, each bridging line having a bridging diode means therein oriented at a polarity which is reverse with respect to the transmitting input diode means of the respective output line; the bridging lines of each tuned circuit being connected to one another by an inductance and having capacitors disposed between the bridging diode means thereon and the bus lines, and

a direct current device connected to the bridging lines of the tuned circuit.

2. The apparatus of claim **1**, wherein there are a plurality of similarly configured tuned circuits connected between the output lines and the bus lines.

3. The apparatus of claim **2**, wherein there are a plurality of radio frequency-to-direct current conversion circuits connected to one another to provide a direct current power array tuned to a specific radio frequency such as a high frequency source, low frequency source, very low frequency source, extremely low frequency source, or source created by seismic vibrations of the earth's magnetic field.

4. The apparatus of claim **3** wherein the dipolar input is a dipolar antenna.

5. The apparatus of claim **1**, wherein the device connected to the output lines is a direct current motor.

6. The apparatus of claim **1**, wherein the device connected to the output line is an illuminating device.

7. The apparatus of claim **1**, wherein the device connected to the output lines is a battery charger.

8. The apparatus of claim **1**, wherein the device connected to the output lines is a DC-to-AC inverter.

9. The apparatus of claim **1**, wherein there are a plurality of radio frequency to direct current converter circuits connected to one another to form an array of circuits tuned to a particular radio frequency.

10. The apparatus of claim **1** wherein the dipolar input is a dipolar antenna.

11. The apparatus of claim **10**, wherein the dipolar antenna utilizes aluminum wire arranged in pair of triangular loops.

* * * * *

US008253639B2

(12) **United States Patent**
Cohen

(10) **Patent No.: US 8,253,639 B2**
(45) **Date of Patent: Aug. 28, 2012**

(54) **WIDEBAND ELECTROMAGNETIC CLOAKING SYSTEMS**

(76) Inventor: **Nathan Cohen**, Belmont, MA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 498 days.

(21) Appl. No.: **12/547,104**

(22) Filed: **Aug. 25, 2009**

(65) **Prior Publication Data**

US 2010/0156556 A1 Jun. 24, 2010

Related U.S. Application Data

(60) Provisional application No. 61/189,966, filed on Aug. 25, 2008.

(51) **Int. Cl.**
H01Q 15/02 (2006.01)
H01Q 19/06 (2006.01)

(52) **U.S. Cl.** **343/753**; 343/909

(58) **Field of Classification Search** 343/753, 343/909, 910; 333/135
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,471,224 A * 11/1995 Barkeshli 343/909
5,714,961 A * 2/1998 Kot et al. 343/769
6,473,048 B1 * 10/2002 Diaz 343/753

* cited by examiner

Primary Examiner — Michael C Wimer
(74) *Attorney, Agent, or Firm* — McDermott Will & Emery LLP

(57) **ABSTRACT**

Arrangement of resonators in an aperiodic configurations are described, which can be used for electromagnetic cloaking of objects. The overall assembly of resonators, as structures, do not all repeat periodically and at least some of the resonators are spaced such that their phase centers are separated by more than a wavelength. The arrangements can include resonators of several different sizes and/or geometries arranged so that each size or geometry corresponds to a moderate or high "Q" response that resonates within a specific frequency range, and that arrangement within that specific grouping of akin elements is periodic in the overall structure. The relative spacing and arrangement of groupings can be defined by self similarity and origin symmetry.

14 Claims, 4 Drawing Sheets

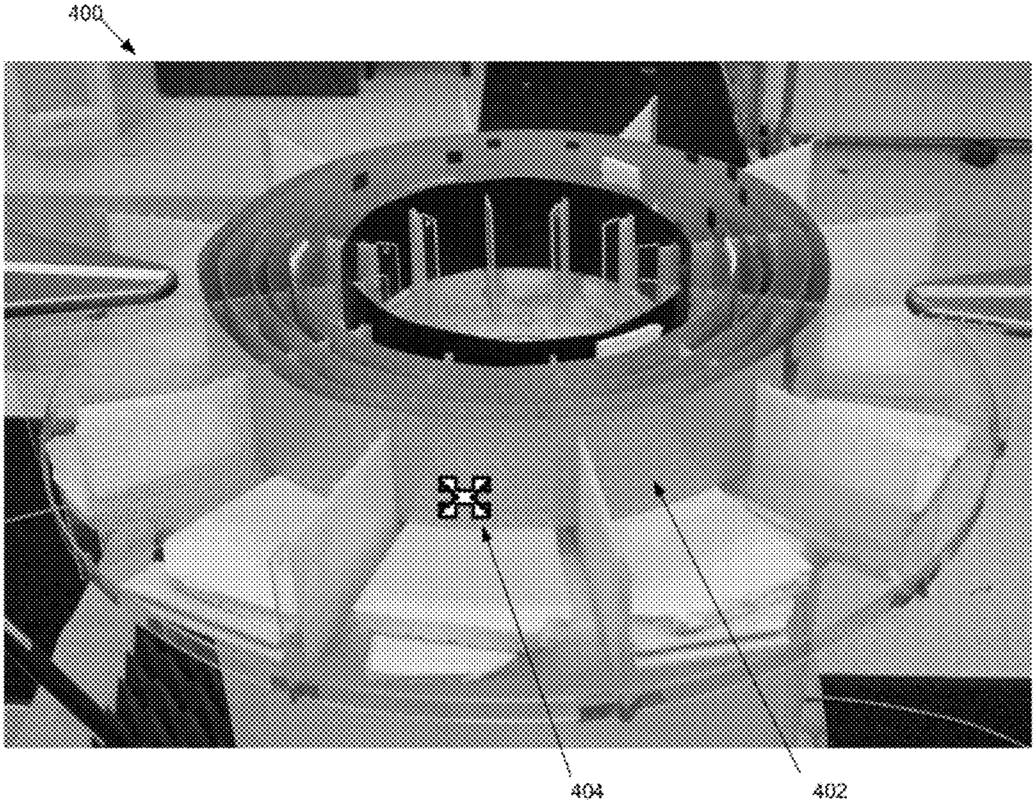

FIG. 1

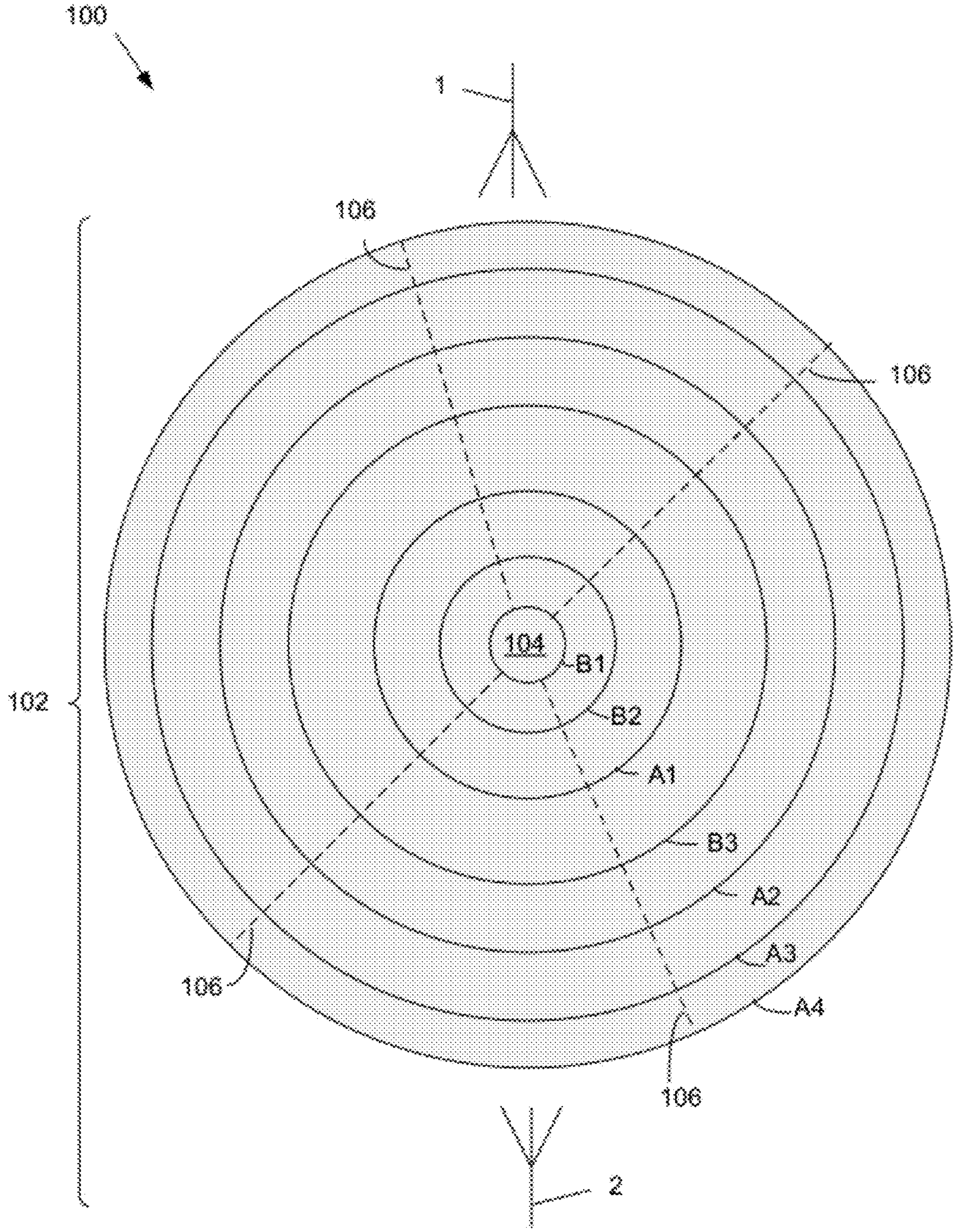
100
1
106
106
102
104
B1
B2
A1
B3
A2
A3
A4
106
106
2

FIG. 2

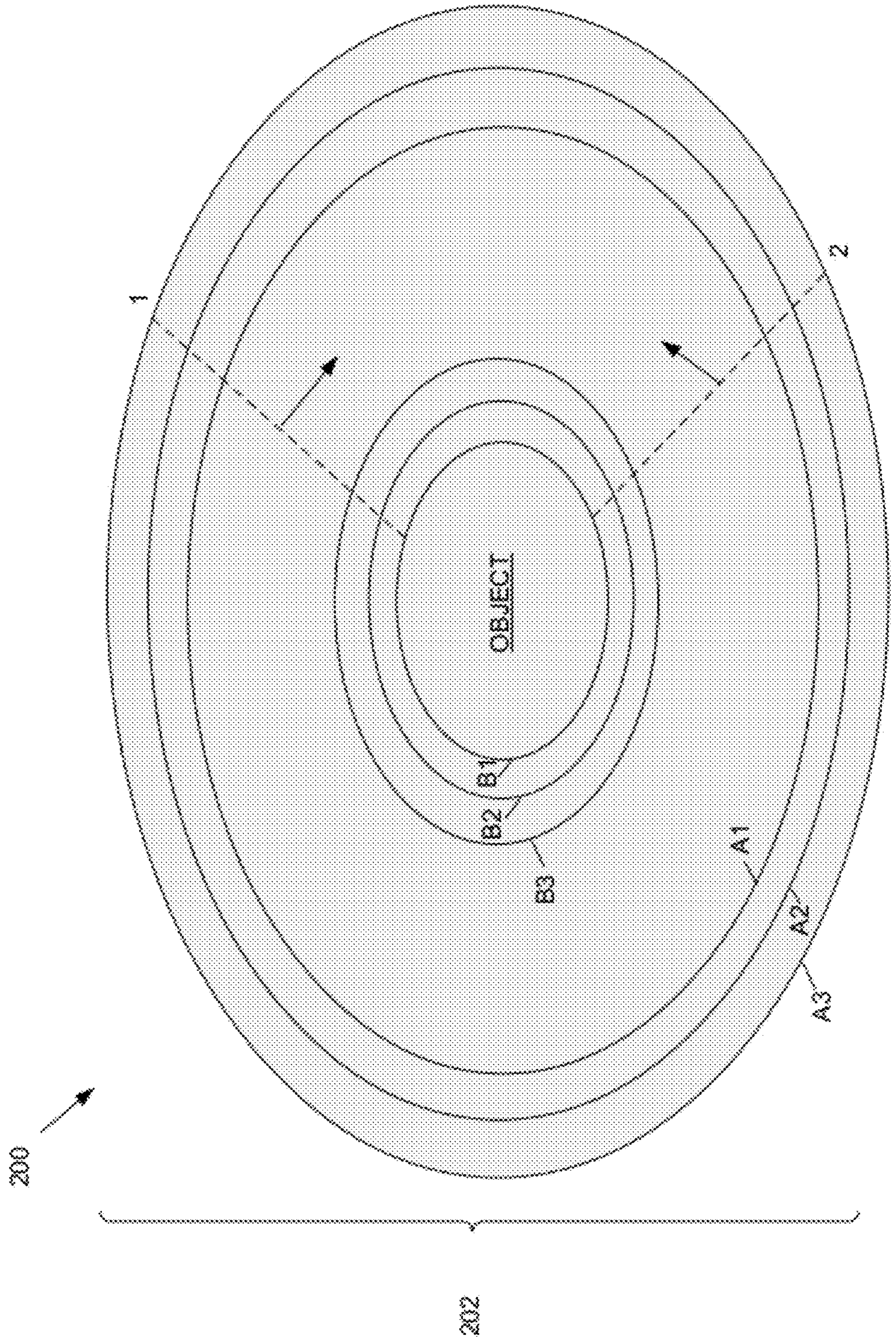
200
202
1
2
OBJECT
B1
B2
B3
A1
A2
A3

FIG. 3

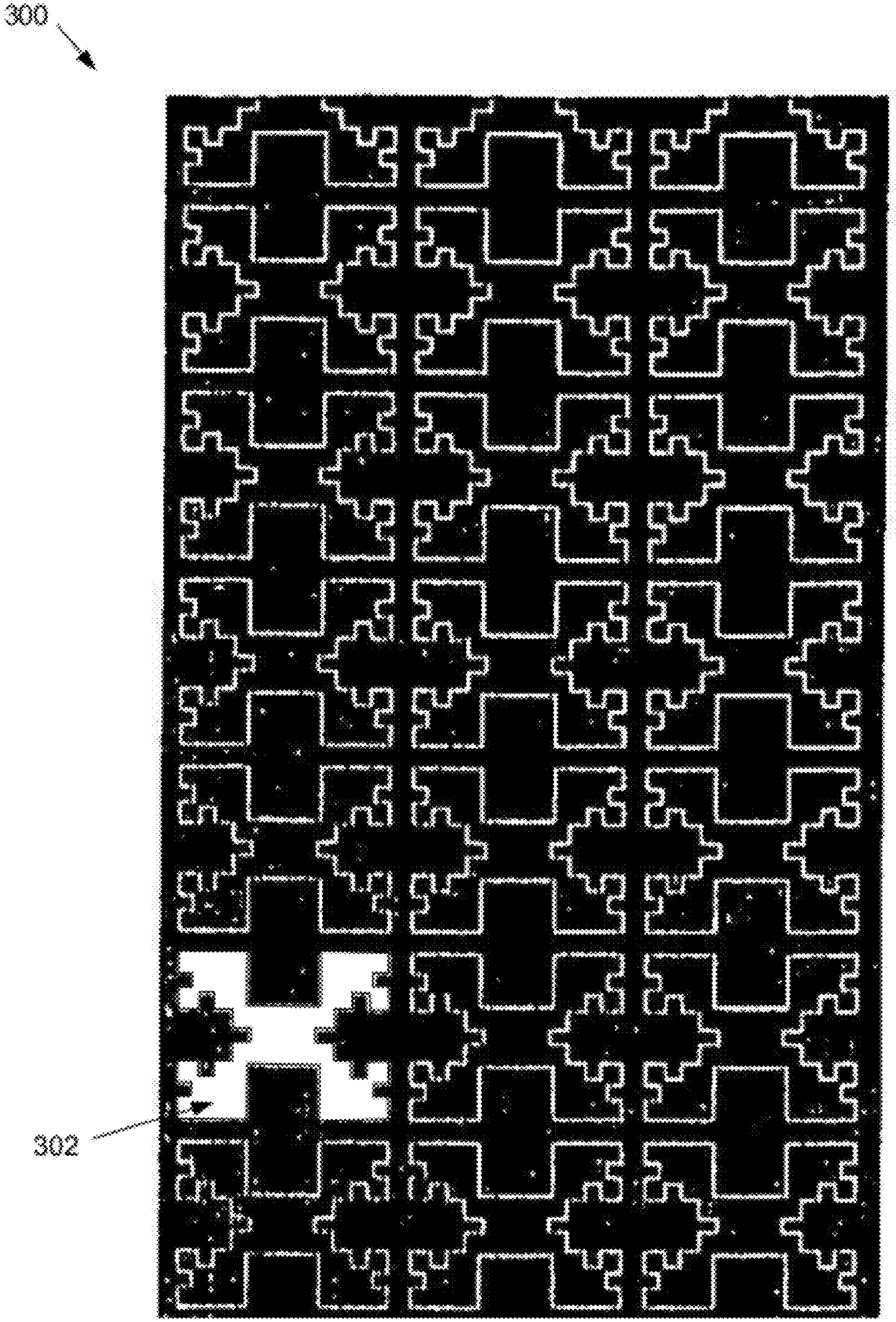

FIG. 4

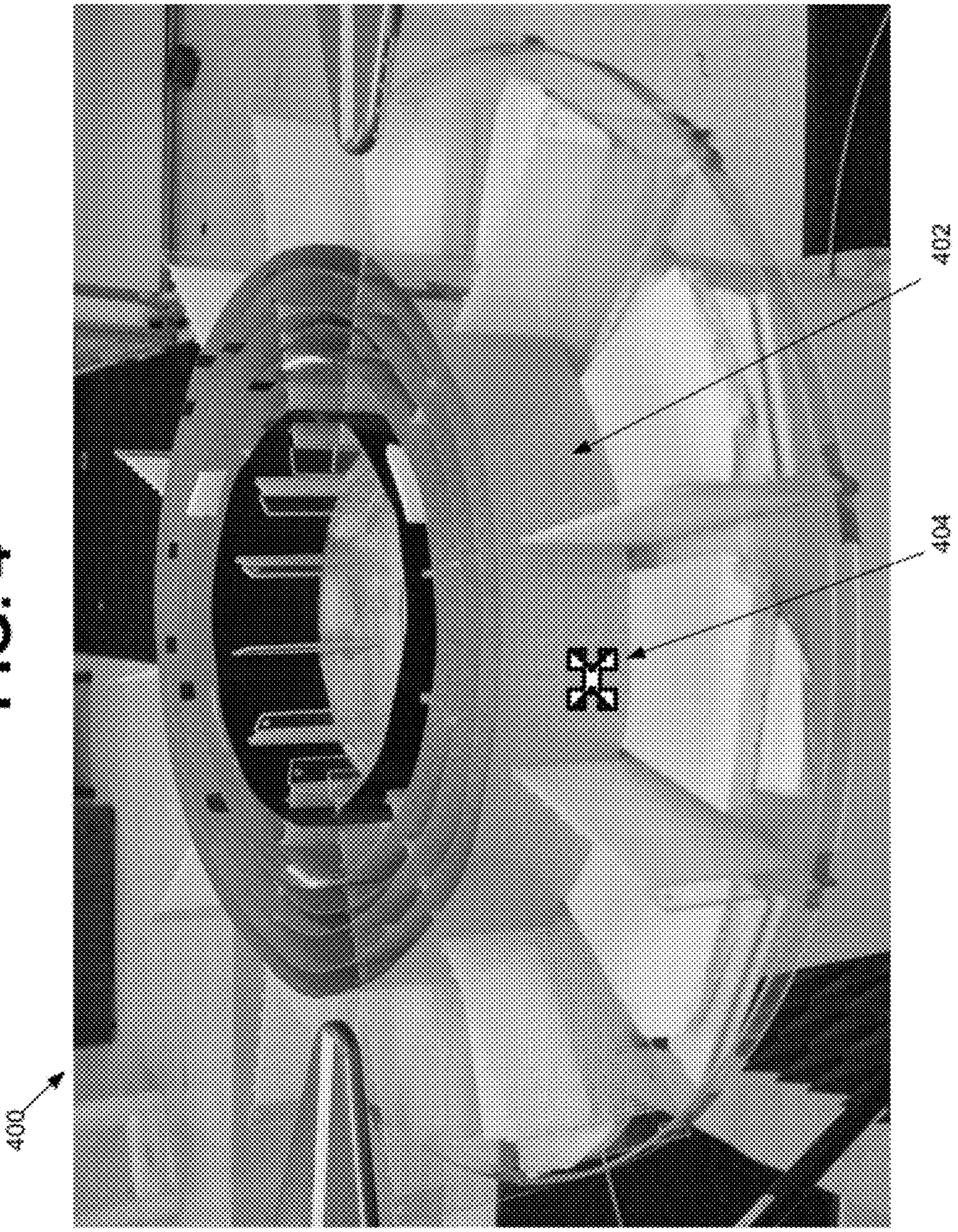

WIDEBAND ELECTROMAGNETIC CLOAKING SYSTEMS

RELATED APPLICATION

This application claims priority to U.S. Provisional Patent Application No. 61/189,966, filed 25 Aug. 2008 and entitled "Method and Apparatus for Wideband Electromagnetic Cloaking, Negative Refractive Index Lensing and Metamaterial Applications," the entire contents of which are incorporated herein by reference.

BACKGROUND

Much time and effort has been devoted to the quest for so-called invisibility machines. Beyond science fiction, however, there has been little if any real progress toward this goal.

Materials with negative permittivity and permeability leading to negative index of refraction were theorized by Russian noted physicist Victor Veselago in his seminal paper in *Soviet Physics* USPEKHI, 10, 509 (1968). Since that time, metamaterials have been developed that produce negative index of refraction, subject to various constraints. Such materials are artificially engineered micro/nanostructures that, at given frequencies, show negative permeability and permittivity. Metamaterials have been shown to produce narrow band, e.g., typically less than 5%, response such as bent-back lensing. Such metamaterials produce such a negative-index effect by utilizing a closely-spaced periodic lattice of resonators, such as split-ring resonators, that all resonate. Previous metamaterials provide a negative index of refraction when a sub-wavelength spacing is used for the resonators.

In the microwave regime, certain techniques have been developed to utilize radiation-absorbing materials or coatings to reduce the radar cross section of airborne missiles and vehicles. While such absorbing materials can provide an effective reduction in radar cross section, these results are largely limited to small ranges of electromagnetic radiation.

SUMMARY

Embodiments of the present disclosure can provide techniques, including systems and/or methods, for cloaking objects at certain wavelengths/frequencies or over certain wavelength/frequency ranges (bands). Such techniques can provide an effective electromagnetic lens and/or lensing effect for certain wavelengths/frequencies or over certain wavelength/frequency ranges (bands).

The effects produced by such techniques can include cloaking or so-called invisibility of the object(s) at the noted wavelengths or bands. Representative frequencies of operation can include, but are not limited to, those over a range of 500 MHz to 1.3 GHz, though others may of course be realized. Operation at other frequencies, including for example those of visible light, infrared, ultraviolet, and as well as microwave EM radiation, e.g., K, Ka, X-bands, etc. may be realized, e.g., by appropriate scaling of dimensions and selection of shape of the resonator elements.

Exemplary embodiments of the present disclosure can include a novel arrangement of resonators in an aperiodic configuration or lattice. The overall assembly of resonators, as structures, do not all repeat periodically and at least some of the resonators are spaced such that their phase centers are separated by more than a wavelength. The arrangements can include resonators of several different sizes and/or geometries arranged so that each size or geometry ("grouping") corresponds to a moderate or high "Q" (that is moderate or low bandwidth) response that resonates within a specific frequency range, and that arrangement within that specific grouping of akin elements is periodic in the overall structure—even though the structure as a whole is not an entirely periodic arrangement of resonators. The relative spacing and arrangement of groupings (at least one for each specific frequency range) can be defined by self similarity and origin symmetry, where the "origin" arises at the center of a structure (or part of the structure) individually designed to have the wideband metamaterial property.

For exemplary embodiments, fractal resonators can be used for the resonators in such structures because of their control of passbands, and smaller sizes compared to non-fractal based resonators. Their benefit arises from a size standpoint because they can be used to shrink the resonator (s), while control of passbands can reduce or eliminates issues of harmonic passbands that would resonate at frequencies not desired.

It should be understood that other embodiments of wideband electromagnetic resonator or cloaking systems and methods according to the present disclosure will become readily apparent to those skilled in the art from the following detailed description, wherein exemplary embodiments are shown and described by way of illustration. The systems and methods of the present disclosure are capable of other and different embodiments, and details of such are capable of modification in various other respects. Accordingly, the drawings and detailed description are to be regarded as illustrative in nature and not as restrictive.

BRIEF DESCRIPTION OF THE DRAWINGS

Aspects of the disclosure may be more fully understood from the following description when read together with the accompanying drawings, which are to be regarded as illustrative in nature, and not as limiting. The drawings are not necessarily to scale, emphasis instead being placed on the principles of the disclosure. In the drawings:

FIG. **1** depicts a diagrammatic plan view of a resonator cloaking system utilizing a number of cylindrical shells, in accordance with exemplary embodiments of the present disclosure;

FIG. **2** depicts a diagrammatic plan view of a resonator cloaking system utilizing a number of shells having an elliptical cross-section, in accordance with an alternate embodiment of the present disclosure;

FIG. **3** depicts an exemplary embodiment of a portion of shell that includes repeated conductive traces that are configured in a fractal-like shape; and

FIG. **4** depicts a perspective view (photograph) of an exemplary embodiment of the present disclosure.

While certain embodiments depicted in the drawings, one skilled in the art will appreciate that the embodiments depicted are illustrative and that variations of those shown, as well as other embodiments described herein, may be envisioned and practiced within the scope of the present disclosure.

DETAILED DESCRIPTION

The present disclosure is directed to novel arrangements of resonators useful for obscuring or hiding objects at given bands of electromagnetic radiation. Embodiments of the present disclosure can provide techniques, including systems and/or methods, for hiding or obscuring objects at certain wavelengths/frequencies or over certain wavelength/frequency ranges or bands. Such techniques can provide an

effective electromagnetic lens and/or lensing effect for certain wavelengths/frequencies or over certain wavelength/frequency ranges or bands. The effects produced by such techniques can include cloaking or so-called invisibility of the object(s) at the noted wavelengths or bands.

Representative frequencies of operation can include, but are not limited to, those over a range of about 500 MHz to about 1.3 GHz, though others may of course be realized. Operation at other frequencies, including for example those of visible light, infrared, ultraviolet, and as well as microwave EM radiation, e.g., K, Ka, X-bands, etc. may be realized, e.g., by appropriate scaling of dimensions and selection of shape of the resonator elements.

Embodiments of the present disclosure include arrangement of resonators or resonant structures in aperiodic configurations or lattices. The overall assembly of resonator structures can include nested or concentric shells, that each include repeated patterns of resonant structures. The resonant structures can be configured as a close-packed arrangement of electrically conductive material. The resonant structures can be located on the surface of a circuit board.

The overall assemblies, as structures, do not all repeat periodically and at least some of the resonators are spaced such that their phase centers are separated by more than a wavelength. The arrangements can include resonators of several different sizes and/or geometries arranged so that each size or geometry ("grouping") corresponds to a moderate or high quality-factor "Q" response (that is, one allowing for a moderate or low bandwidth) that resonates within a specific frequency range, and that arrangement within that specific grouping of like elements is periodic in the overall structure—even though the structure as a whole is not an entirely periodic arrangement of resonators. The relative spacing and arrangement of groupings (at least one for each specific frequency range) can be defined by self similarity and origin symmetry, where the "origin" arises at the center of a structure (or part of the structure) individually designed to have the wideband metamaterial property.

For exemplary embodiments, fractal resonators can be used for the resonators because of their control of passbands, and smaller sizes. A main benefit of such resonators arises from a size standpoint because they can be used to shrink the resonator(s), while control of passbands can reduce/mitigate or eliminate issues of harmonic passbands that would resonate at frequencies not desired.

Exemplary embodiments of a resonator system for use at microwave (or nearby) frequencies can be built from belts of circuit boards festooned with resonators. These belts can function to slip the microwaves around an object located within the belts, so the object is effectively invisible and "see thru" at the microwave frequencies. Belts, or shells, having similar closed-packed arrangements for operation at a first passband can be positioned within a wavelength of one another, e.g., 1/10λ, 1/8λ, 1/4λ, 1/2λ, etc.

An observer can observe an original image or signal, without it being blocked by the cloaked object. Using no power, the fractal cloak can replicates the original signal (that is, the signal before blocking) with great fidelity. Exemplary embodiments can function over a bandwidth from about 500 MHz to approximately 1500 MHz (1.5 GHz), providing 3:1 bandwidth; operation within or near such can frequencies can provide other bandwidths as well, such as 1:1 up to 2:1 and up to about 3:1.

FIG. **1** depicts a diagrammatic plan view of a cloaking system **100** and RF testing set up in accordance with exemplary embodiments of the present disclosure. As shown in FIG. **1**, a number of concentric shells (or bands) **102** are placed on a platform (parallel to the plane of the drawing). The shells include a flexible substrate (e.g., polyimide with or without composite reinforcement) with conductive traces (e.g., copper, silver, etc.) in fractal shapes or outlines. The shells **102** surround an object to be cloaked (shown as **104** in FIG. **1**). A transmitting antenna **1** and a receiving antenna **2** are configured at different sides of the system **100**, for verifying efficacy of the cloaking system **100** and recording results. The shells **102** can be held in place by radial supports **106** (while only four are shown, **12** were used in the exemplary embodiment indicated).

The shells indicated in FIG. **1** are of two types, one set (A**1**-A**4**) configured for optimal operation over a first wavelength/frequency range, and another set (B**1**-B**3**) configured for optimal operation over a second wavelength/frequency range. (The numbering of the shells is of course arbitrary and can be reordered, e.g., reversed.)

For an exemplary embodiment of system **100**, the outer set of shells (A**1**-A**4**, with A**1** being the innermost and A**4** the outmost) had a height of about 3 to 4 inches (e.g., 3.5 inches) and the inner set of shells had a height of about 1 inch less (e.g., about 2.5 to 3 inches). The spacing between the shells with a larger fractal shape (A**1**-A**4**) was about 2.4 cm while the spacing between shells of smaller fractal generator shapes (B**1**-B**3**) was about 2.15 cm (along a radial direction). In a preferred embodiment, shell A**4** was placed between shell B**2** and B**3** as shown. The resonators formed on each shell by the fractal shapes can be configured so as to be closely coupled (e.g., by capacitive coupling) and can serve to propagate a plasmonic wave.

It will be appreciated that while, two types of shells and a given number of shells per set are indicated in FIG. **1**, the number of shell types and number of shells for each set can be selected as desired, and may be optimized for different applications, e.g., wavelength/frequency bands.

FIG. **2** depicts a diagrammatic plan view of a cloaking system (or electrical resonator system) according to an alternate embodiment in which the individual shells have an elliptical cross section. As shown in FIG. **2**, a system **200** for cloaking can include a number of concentric shells (or bands) **202**. These shells can be held in place with respect to one another by suitable fixing means, e.g., they can be placed on a platform (parallel to the plane of the drawing) and/or held with a frame. The shells **202** can include a flexible substrate (e.g., polyimide with or without composite reinforcement) with a close-packed arrangement of electrically conductive material formed on the first surface. As stated previously for FIG. **1**, the closed-packed arrangement can include a number of self-similar electrical resonator shapes. The resonator shapes can be made from conductive traces (e.g., copper, silver, gold, silver-based ink, etc.) having a desired shape, e.g., fractal shape, split-ring shape, and the like. The shells **202** can surround an object to be cloaked, as indicated in FIG. **2**.

As indicated in FIG. **2** (by dashed lines **1** and **2** and arrows), the various shells themselves do not have to form closed surfaces. Rather, one or more shells can form open surfaces. This can allow for preferential cloaking of the object in one direction or over a given angle (solid angle). Moreover, while dashed lines **1** and **2** are shown intersecting shells B**1**-B**3** and A**1**-A**3** of system **200**, one or more shells of each group of shells (B**1**-B**3** and A**1**-A**3**) can be closed while others are open.

With further regard to FIGS. **1**-**2**, it should be appreciated that the cross-sections shown for each shell can represent closed geometric shapes, e.g., spherical and ellipsoidal shells.

As indicated previously, each shell of a cloaking system can include multiple resonators. The resonators can be repeated patterns of conductive traces. These conductive traces can be closed geometric shapes, e.g., rings, loops, closed fractals, etc. The resonator(s) can being self similar to at least second iteration. The resonators can include split-ring shapes, for some embodiments. The resonant structures are not required to be closed shapes, however, and open shapes can be used for such.

In exemplary embodiments, the closed loops can be configured as a fractals or fractal-based shapes, e.g., as depicted by **302** in FIG. **3** for an exemplary embodiment of a shell **300**, or **402** in FIG. **4**. The dimensions and type of fractal shape can be the same for each shell type but can vary between shell types. This variation (e.g., scaling of the same fractal shape) can afford increased bandwidth for the cloaking characteristics of the system (e.g., system **100** of FIG. **1**) This can lead to periodicity of the fractal shapes of common shell types but aperiodicity between the fractal shapes of different shell types.

Examples of suitable fractal shapes (for use for shells and/or a scatting object) can include, but are not limited to, fractal shapes described in one or more of the following patents, owned by the assignee of the present disclosure, the entire contents of all of which are incorporated herein by reference: U.S. Pat. Nos. 6,452,553; 6,104,349; 6,140,975; 7,145,513; 7,256,751; 6,127,977; 6,476,766; 7,019,695; 7,215,290; 6,445,352; 7,126,537; 7,190,318; 6,985,122; 7,345,642; and, U.S. Pat. No. 7,456,799.

Other suitable fractal shape for the resonant structures can include any of the following: a Koch fractal, a Minkowski fractal, a Cantor fractal, a torn square fractal, a Mandelbrot, a Caley tree fractal, a monkey's swing fractal, a Sierpinski gasket, and a Julia fractal, a contour set fractal, a Sierpinski triangle fractal, a Menger sponge fractal, a dragon curve fractal, a space-filling curve fractal, a Koch curve fractal, an Iypanov fractal, and a Kleinian group fractal.

FIG. **3** depicts an exemplary embodiment of a shell **300** (only a portion is shown) that includes repeated conductive traces that are configured in a fractal shape **302** (the individual closed traces). For the exemplary embodiment shown, each resonator shape **302** is about 1 cm on a side. Such resonator could, e.g., be used for the fractal shapes of shells B**1**-B**3** of FIG. **1**, in which case similar fractal shapes of larger size (e.g., about 1.5 cm on a side) could be used for shells A**1**-A**4**. The conductive trace is preferably made of copper. While exemplary fractal shapes are shown in FIG. **3**, the present disclosure is not limited to such and any other suitable fractal shapes (including generator motifs) may be used in accordance with the present disclosure.

It will be appreciated that the resonant structures of the shells may be formed or made by any suitable techniques and with any suitable materials. For example, semiconductors with desired doping levels and dopants may be used as conductive materials. Suitable metals or metal containing compounds may be used. Suitable techniques may be used to place conductors on/in a shell, including, but no limited to, printing techniques, photolithography techniques, etching techniques, and the like.

It will also be appreciated that the shells may be made of any suitable material(s). Printed circuit board materials may be used. Flexible circuit board materials are preferred. Other material may, however, be used for the shells and the shells themselves can be made of non-continuous elements, e.g., a frame or framework. For example, various plastics may be used.

FIG. **4** depicts a perspective view (photograph) of an exemplary embodiment of a cloak system **400** according to the present disclosure. As shown, the system includes a number of resonator shells **402** each having a close-packed arrangement of electrically conductive material (self-similar resonators **404**) formed on one surface. Two different shell configurations are shown, four larger shells and two smaller shells. The smaller shells included close-packed arrangements of resonator structures in which each resonator shape (as shown by **302** in FIG. **3**) was about 1 cm on a side. Similar fractal shapes of larger size (e.g., about 1.5 cm on a side) were used for the larger shells.

In FIG. **4**, a transmitting (source) antenna and a receiving antenna are shown as triangular shapes on the left and right, respectively (though functionally of each could of course be interchanged for the other). Twelve radially arrayed spacers are shown in FIG. **4**. The system **400** is shown supported on a Nalgene tank and Delrin platform and Delrin supports (radial supports) RF absorbers were placed in the immediate vicinity of the set up; further RF tripods (e.g., available from ETS) were used; all such materials were substantially transparent at the RF frequencies investigated/used. The cloak system **400** consists of six belts of fractal metamaterial (i.e., fractal-resonant structures shown in FIG. **3**) on flexible Taconic EF35 (low loss) circuit board. The belts are shown surround a scattering ring (object). The arrangement is supported by RF transparent plastics in a comb support. The entire system **400** was shown to be easily built up and broken down within a minute or two. The scale in FIG. **4** is about 0.7 meters across. The height of each shell can of course be selected as desired depending on the situation/application.

While embodiments are shown and described herein as having shells in the shape of concentric rings (circular cylinders), shells can take other shapes in other embodiments. For example, one or more shells could have a generally spherical shape (with minor deviations for structural support). In an exemplary embodiment, the shells could form a nested arrangement of such spherical shapes, around an object to be shielded (at the targeted/selected frequencies/wavelengths). Shell cross-sections of angular shapes, e.g., triangular, hexagonal, while not preferred, may be used.

One skilled in the art will appreciate that embodiments and/or portions of embodiments of the present disclosure can be implemented in/with computer-readable storage media (e.g., hardware, software, firmware, or any combinations of such), and can be distributed and/or practiced over one or more networks. Steps or operations (or portions of such) as described herein, including processing functions to derive, learn, or calculate formula and/or mathematical models utilized and/or produced by the embodiments of the present disclosure, can be processed by one or more suitable processors, e.g., central processing units ("CPUs") implementing suitable code/instructions in any suitable language (machine dependent on machine independent).

While certain embodiments and/or aspects have been described herein, it will be understood by one skilled in the art that the methods, systems, and apparatus of the present disclosure may be embodied in other specific forms without departing from the spirit thereof.

For example, while certain wavelengths/frequencies of operation have been described, these are merely representative and other wavelength/frequencies may be utilized or achieved within the scope of the present disclosure.

Furthermore, while certain preferred fractal generator shapes have been described others may be used within the scope of the present disclosure. Accordingly, the embodi-

ments described herein are to be considered in all respects as illustrative of the present disclosure and not restrictive.

What is claimed is:

1. An electrical resonator system, comprising:

a plurality of concentric electrical resonator shells, each shell including a substrate having first and second surfaces and a close-packed arrangement of electrically conductive material formed on the first surface, wherein the closed-packed arrangement comprises a plurality of self-similar electrical resonator shapes and is configured to operate at a desired passband of electromagnetic radiation;

wherein the close-packed arrangements of at least two of the electrical resonator shells are different in size and/or shape; and

wherein a resonator in the close-packed arrangement comprises a second order or higher fractal.

2. The system of claim **1**, wherein said passband is about 2:1.

3. The system of claim **2**, wherein said passband is about 3:1.

4. The system of claim **1**, wherein the electrical system is configured and arranged so that radiation incident on the system from a given direction has an intensity on a point-by-point basis such at each respective antipodal point, relative to an object placed at the center of the system, the radiation has the same or similar intensity.

5. The system of claim **1**, wherein the system is configured and arranged so that radiation incident on the system from a direction in cylindrical coordinates has the same or similar intensity at the antipodal point after having traversed the system.

6. The system of claim **1**, wherein the plurality of shells comprises a first pair of shells having similar closed-packed arrangements for operation at a first passband, wherein the two shells are positioned within ⅛λ of one another.

7. The system of claim **6**, wherein the plurality of shells comprises a second pair of shells having similar closed-packed arrangements for operation at a second frequency band, wherein the two shells are positioned within ⅛λ of one another.

8. The system of claim **1**, wherein the plurality of shells are hemispherical.

9. The system of claim **1**, wherein the plurality of shells are cylindrical.

10. The system of claim **1**, wherein the plurality of shells are spherical.

11. The system of claim **1**, wherein said fractal is selected from the group consisting of a Koch fractal, a Minkowski fractal, a Cantor fractal, a torn square fractal, a Mandelbrot, a Caley tree fractal, a monkey's swing fractal, a Sierpinski gasket, and a Julia fractal.

12. The system of claim **1**, wherein the fractal is selected from the group consisting of a contour set fractal, a Sierpinski triangle fractal, a Menger sponge fractal, a dragon curve fractal, a space-filling curve fractal, a Koch curve fractal, a Lypanov fractal, and a Kleinian group fractal.

13. The system of claim **1**, wherein the plurality of concentric electrical resonator shells are configured and arranged for operation at K band, Ka band, or X-band.

14. The system of claim **1**, wherein the system is operational over a bandwidth from about 500 MHz to about 1500 MHz.

* * * * *

US005590031A

United States Patent [19]

Mead, Jr. et al.

[11] **Patent Number: 5,590,031**

[45] **Date of Patent: Dec. 31, 1996**

[54] **SYSTEM FOR CONVERTING ELECTROMAGNETIC RADIATION ENERGY TO ELECTRICAL ENERGY**

[76] Inventors: **Franklin B. Mead, Jr.**, 44536 Avenida Del Sol, Lancaster, Calif. 93535; **Jack Nachamkin**, 12314 Teri Dr., Poway, Calif. 92064

[21] Appl. No.: **281,271**

[22] Filed: **Jul. 27, 1994**

[51] **Int. Cl.^{6}** **H02M 1/00**

[52] **U.S. Cl.** **363/8**; 363/178; 342/6

[58] **Field of Search** 363/8, 13, 178; 342/6, 61, 73, 173, 175

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,882,503	5/1975	Gamara	343/100 R
4,725,847	2/1988	Poirier	343/840
5,008,677	4/1991	Trigon et al.	342/17

Primary Examiner—Peter S. Wong
Assistant Examiner—Adolf Berhane
Attorney, Agent, or Firm—Chris Papageorge

[57] **ABSTRACT**

A system is disclosed for converting high frequency zero point electromagnetic radiation energy to electrical energy. The system includes a pair of dielectric structures which are positioned proximal to each other and which receive incident zero point electromagnetic radiation. The volumetric sizes of the structures are selected so that they resonate at a frequency of the incident radiation. The volumetric sizes of the structures are also slightly different so that the secondary radiation emitted therefrom at resonance interfere with each other producing a beat frequency radiation which is at a much lower frequency than that of the incident radiation and which is amenable to conversion to electrical energy. An antenna receives the beat frequency radiation. The beat frequency radiation from the antenna is transmitted to a converter via a conductor or waveguide and converted to electrical energy having a desired voltage and waveform.

14 Claims, 8 Drawing Sheets

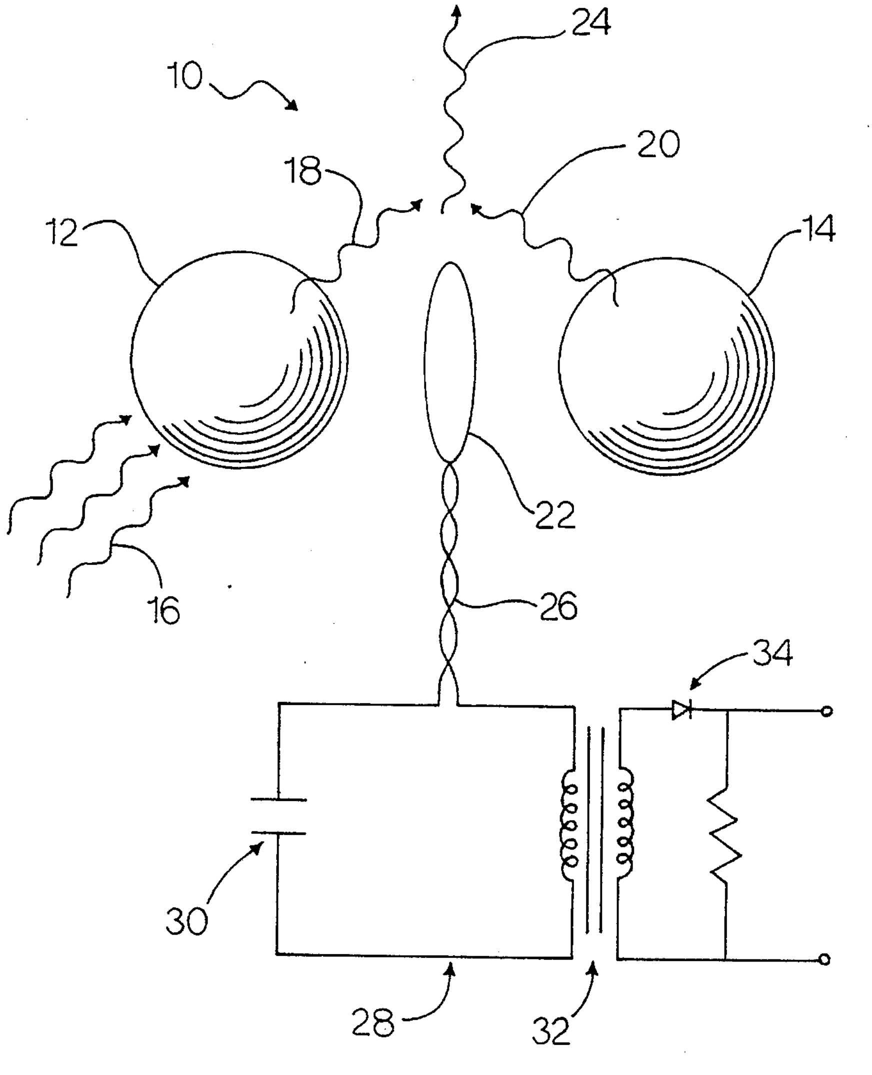

FIG. 1

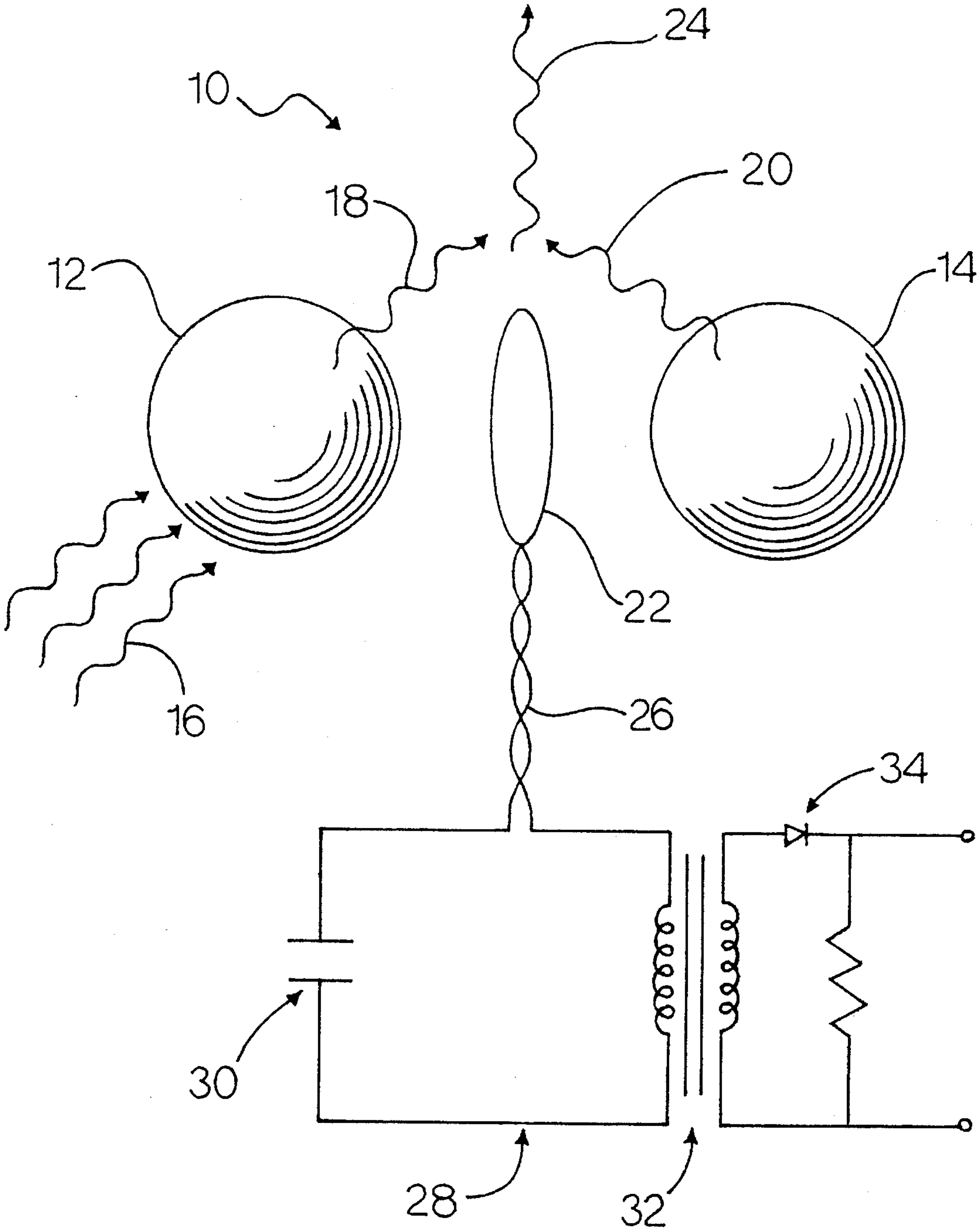
24
10
20
18
12
14
22
16
26
34
30
28
32

FIG. 2

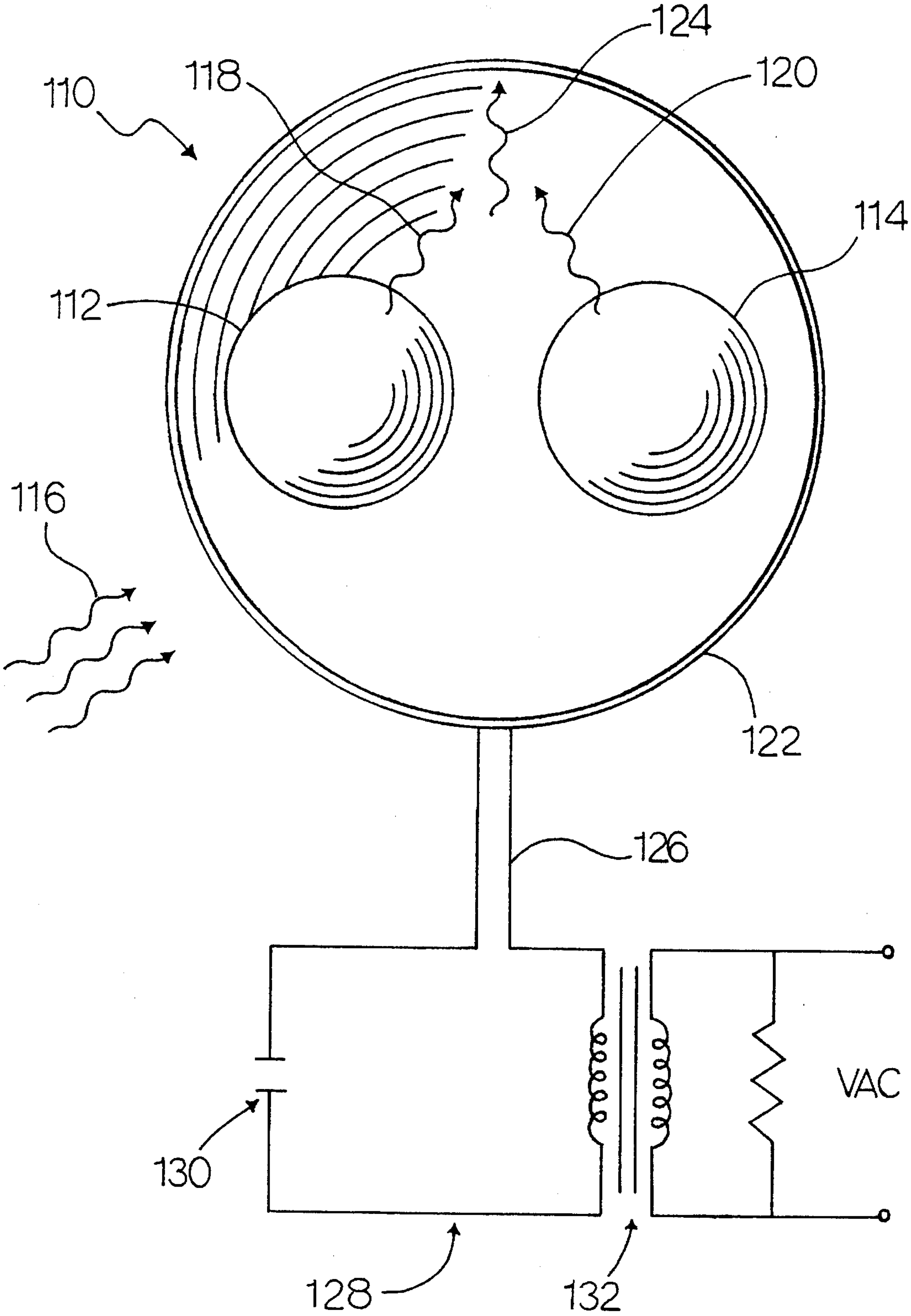
124
110
118
120
114
112
116
122
126
VAC
130
128
132

FIG. 3

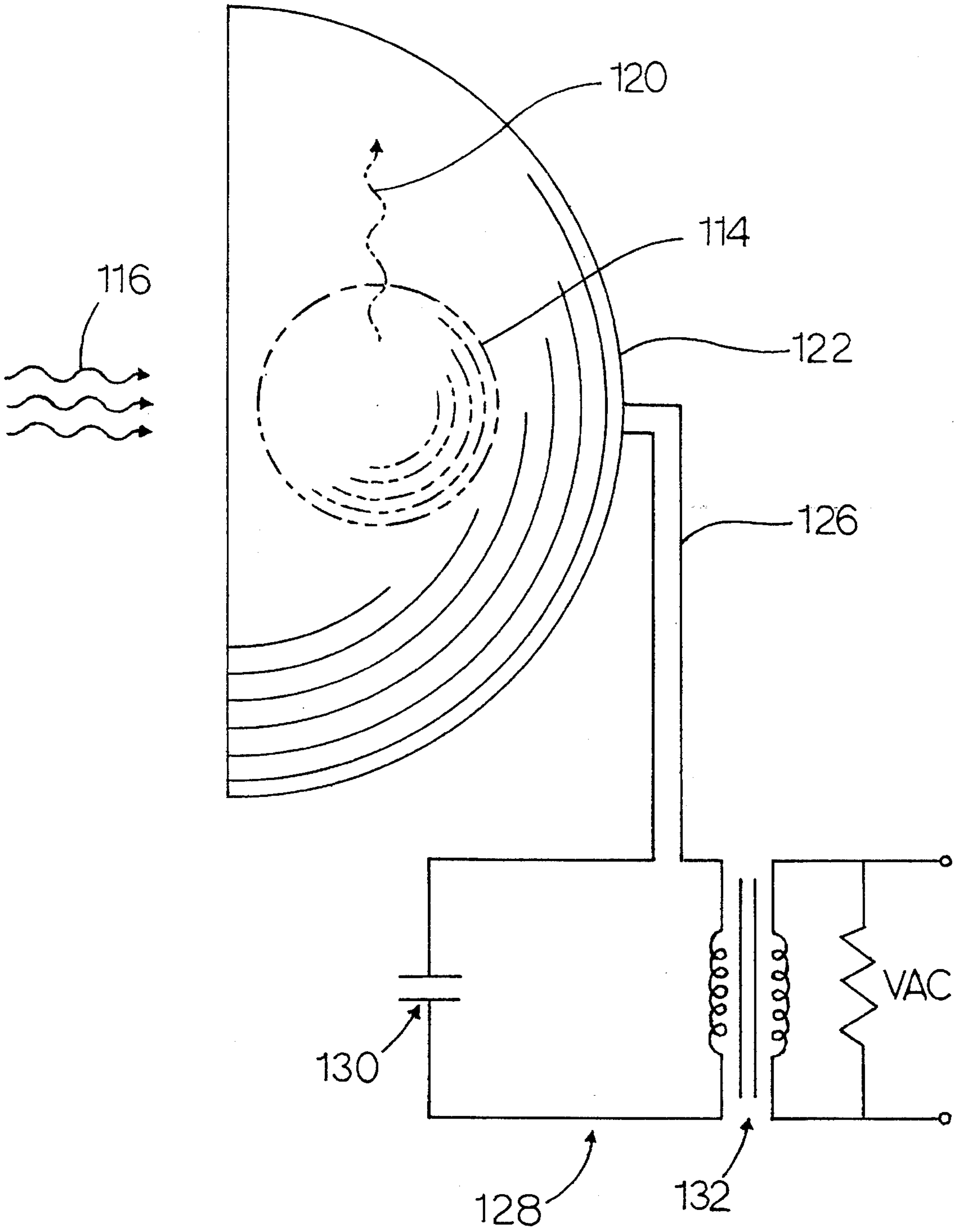
120
114
116
122
126
VAC
130
128
132

FIG. 4

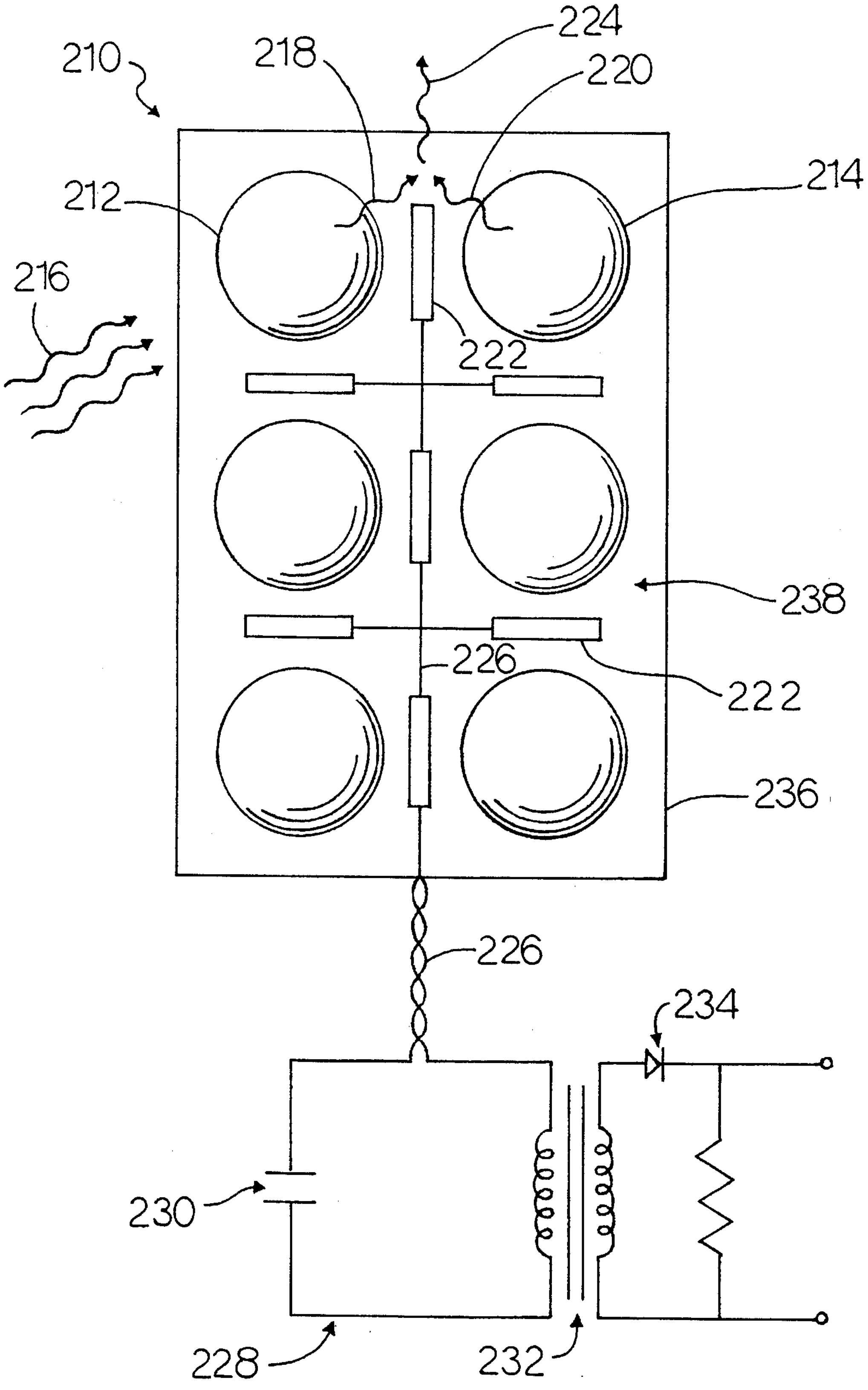

FIG. 5

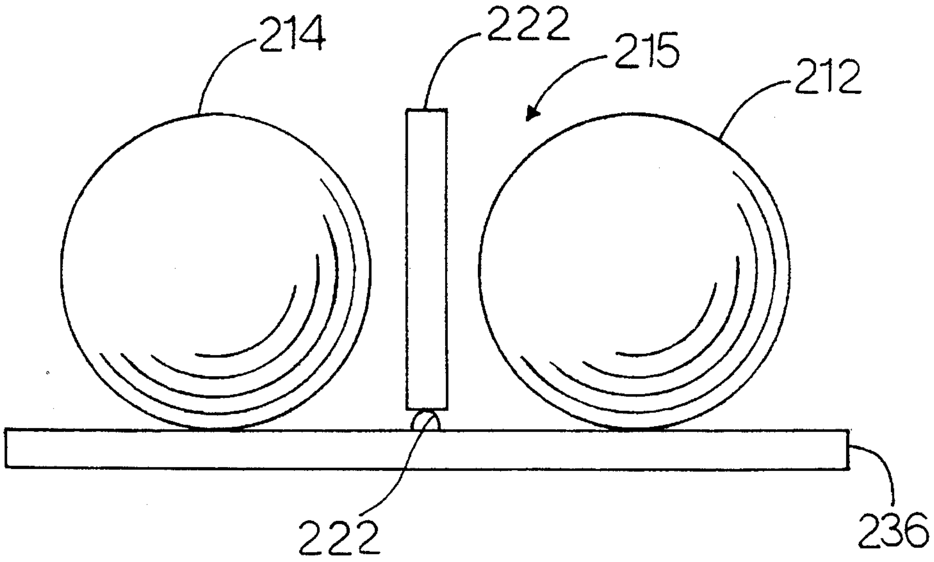
214
222
215
212
222
236

FIG. 6

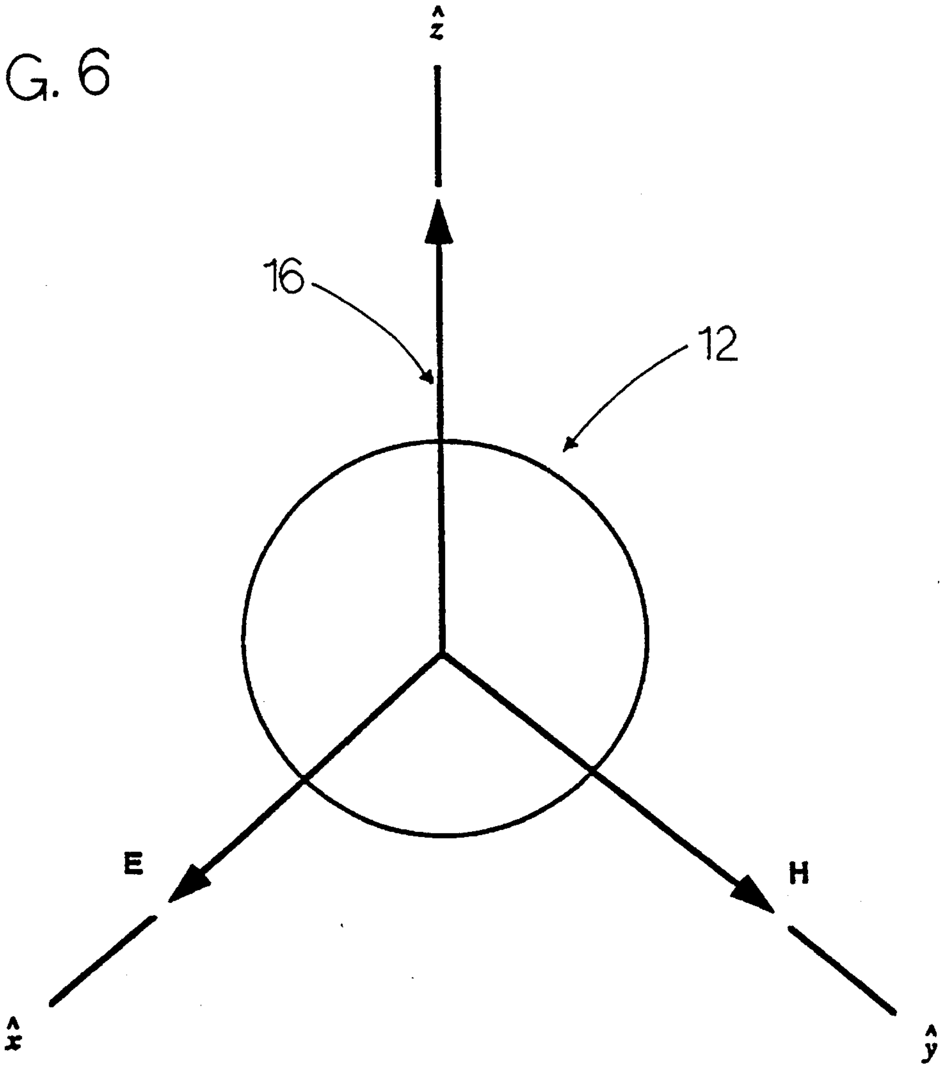
$\hat{z}$
16
12
E
H
$\hat{x}$
$\hat{y}$

FIG.7

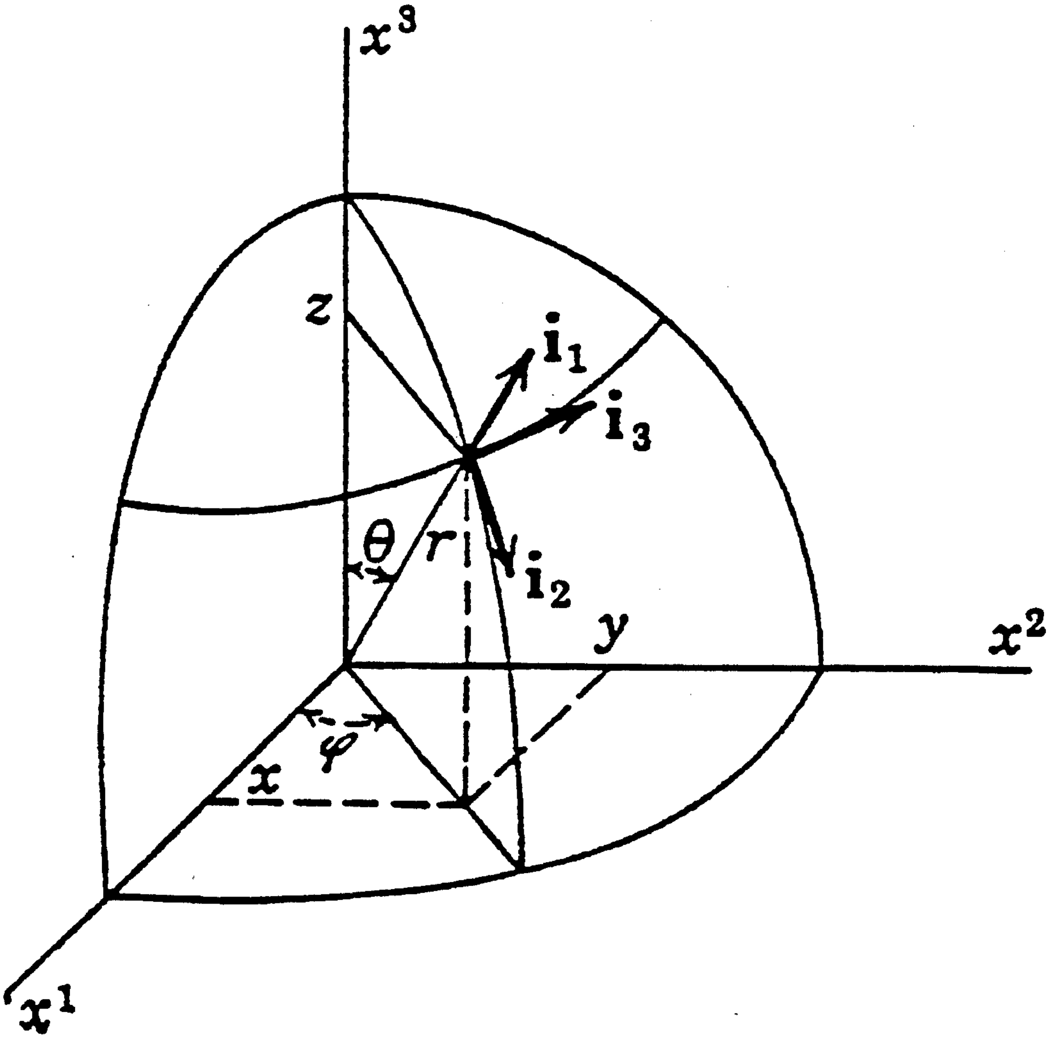
x^3
z
i_1
i_3
θ
r
i_2
y
x^2
φ
x
x^1

FIG. 8

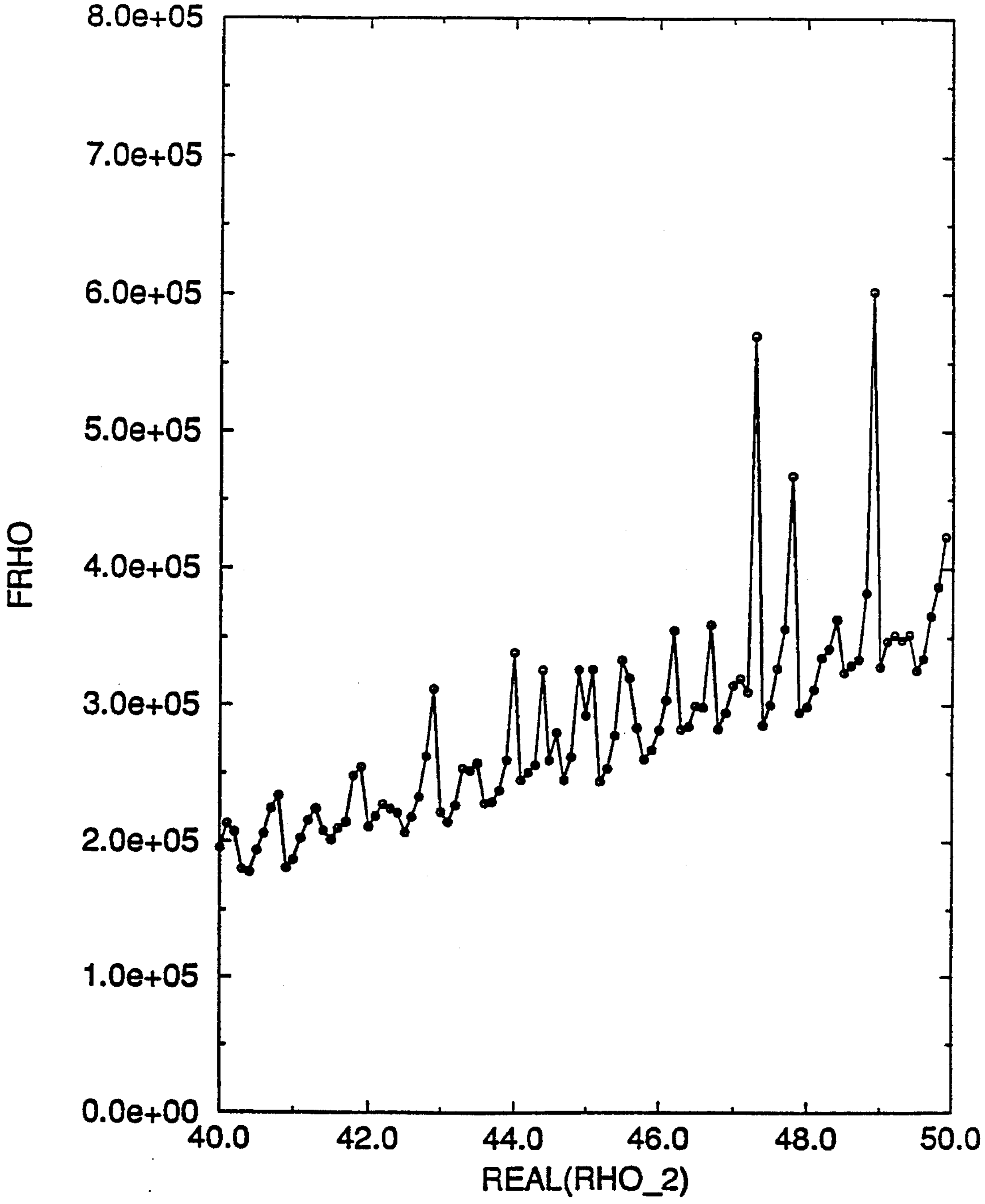
8.0e+05
7.0e+05
6.0e+05
5.0e+05
4.0e+05
3.0e+05
2.0e+05
1.0e+05
0.0e+00
FRHO
40.0
42.0
44.0
46.0
48.0
50.0
REAL(RHO_2)

FIG.9

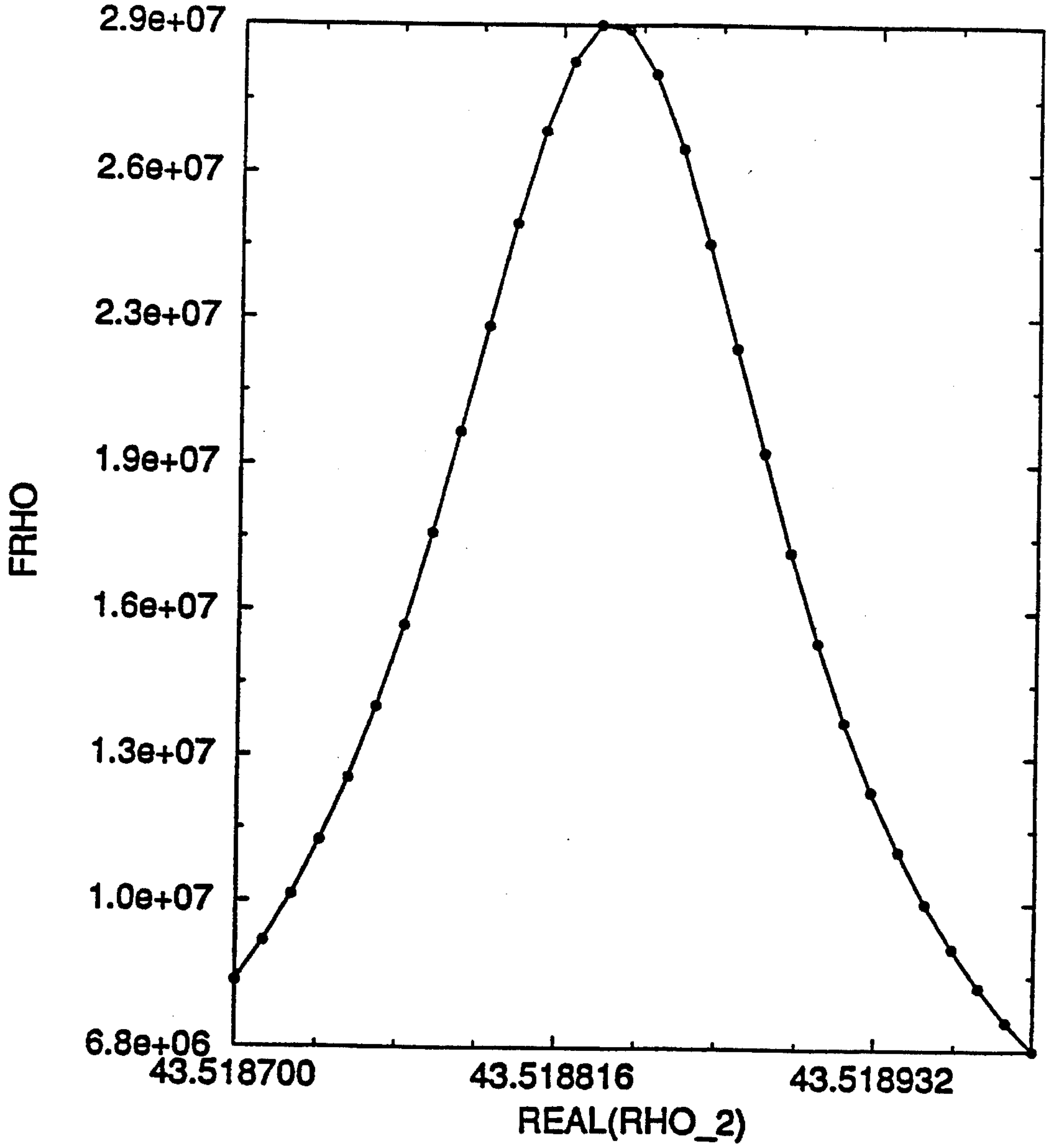
2.9e+07
2.6e+07
2.3e+07
1.9e+07
1.6e+07
1.3e+07
1.0e+07
6.8e+06
FRHO
43.518700
43.518816
43.518932
REAL(RHO_2)

SYSTEM FOR CONVERTING ELECTROMAGNETIC RADIATION ENERGY TO ELECTRICAL ENERGY

BACKGROUND OF THE INVENTION

The invention relates generally to conversion of electromagnetic radiation energy to electrical energy, and, more particularly, to conversion of high frequency bandwidths of the spectrum of a type of radiation known as zero point electromagnetic radiation to electrical energy.

The existence of zero point electromagnetic radiation was discovered in 1958 by the Dutch physicist M. J. Sparnaay. Mr. Sparnaay continued the experiments carried out by Hendrik B. G. Casimir in 1948 which showed the existence of a force between two uncharged parallel plates which arose from electromagnetic radiation surrounding the plates in a vacuum. Mr. Sparnaay discovered that the forces acting on the plates arose from not only thermal radiation but also from another type of radiation now known as classical electromagnetic zero point radiation. Mr. Sparnaay determined that not only did the zero point electromagnetic radiation exist in a vacuum but also that it persisted even at a temperature of absolute zero. Because it exists in a vacuum, zero point radiation is homogeneous and isotropic as well as ubiquitous. In addition, since zero point radiation is also invariant with respect to Lorentz transformation, the zero point radiation spectrum has the characteristic that the intensity of the radiation at any frequency is proportional to the cube of that frequency. Consequently, the intensity of the radiation increases without limit as the frequency increases resulting in an infinite energy density for the radiation spectrum. With the introduction of the zero point radiation into the classical electron theory, a vacuum at a temperature of absolute zero is no longer considered empty of all electromagnetic fields. Instead, the vacuum is now considered as filled with randomly fluctuating fields having the zero point radiation spectrum. The special characteristics of the zero point radiation which are that it has a virtually infinite energy density and that it is ubiquitous (even present in outer space) make it very desirable as an energy source. However, because high energy densities exist at very high radiation frequencies and because conventional methods are only able to convert or extract energy effectively or efficiently only at lower frequencies at which zero point radiation has relatively low energy densities, effectively tapping this energy source has been believed to be unavailable using conventional techniques for converting electromagnetic energy to electrical or other forms of easily useable energy. Consequently, zero point electromagnetic radiation energy which may potentially be used to power interplanetary craft as well as provide for society's other needs has remained unharnessed.

There are many types of prior art systems which use a plurality of antennas to receive electromagnetic radiation and provide an electrical output therefrom. An example of such a prior art system is disclosed in U.S. Pat. No. 3,882,503 to Gamara. The Gamara system has two antenna structures which work in tandem and which oscillate by means of a motor operatively attached thereto in order to modulate the radiation reflected from the antenna surfaces. The reflecting surfaces of the antennas are also separated by a distance equal to a quarter wavelength of the incident radiation. However, the Gamara system does not convert the incident radiation to electrical current for the purpose of converting the incident electromagnetic radiation to another form of readily useable energy. In addition, the relatively large size of the Gamara system components make it unable to resonate at and modulate very high frequency radiation.

What is therefore needed is a system which is capable of converting high frequency electromagnetic radiation energy into another form of energy which can be more readily used to provide power for transportation, heating, cooling as well as various other needs of society. What is also needed is such a system which may be used to provide energy from any location on earth or in space.

SUMMARY OF THE INVENTION

It is a principle object of the present invention to provide a system for converting electromagnetic radiation energy to electrical energy.

It is another object of the present invention to provide a system for converting electromagnetic radiation energy having a high frequency to electrical energy.

It is another object of the present invention to provide a system for converting zero point electromagnetic radiation energy to electrical energy.

It is another object of the present invention to provide a system for converting electromagnetic radiation energy to electrical energy which may used to provide such energy from any desired location on earth or in space.

It is another object of the present invention to provide a system for converting electromagnetic radiation energy to electrical energy having a desired waveform and voltage.

It is an object of the present invention to provide a miniaturized system for converting electromagnetic radiation energy to electrical energy in order to enhance effective utilization of high energy densities of the electromagnetic radiation.

It is an object of the present invention to provide a system for converting electromagnetic radiation energy to electrical energy which is simple in construction for cost effectiveness and reliability of operation.

Essentially, the system of the present invention utilizes a pair of structures for receiving incident electromagnetic radiation which may be propagating through a vacuum or any other medium in which the receiving structures may be suitably located. The system of the present invention is specifically designed to convert the energy of zero point electromagnetic radiation; however, it may also be used to convert the energy of other types of electromagnetic radiation. The receiving structures are preferably composed of dielectric material in order to diffract and scatter the incident electromagnetic radiation. In addition, the receiving structures are of a volumetric size selected to enable the structures to resonate at a high frequency of the incident electromagnetic radiation based on the parameters of frequency of the incident radiation and propagation characteristics of the medium and of the receiving structures. Since zero point radiation has the characteristic that its energy density increases as its frequency increases, greater amounts of electromagnetic energy are available at higher frequencies. Consequently, the size of the structures are preferably miniaturized in order to produce greater amounts of energy from a system located within a space or area of a given size. In this regard, the smaller the size of the receiving structures, the greater the amount of energy that can be produced by the system of the present invention.

At resonance, electromagnetically induced material deformations of the receiving structures produce secondary fields of electromagnetic energy therefrom which may have eva-

nescent energy densities several times that of the incident radiation. The structures are of different sizes so that the secondary fields arising therefrom are of different frequencies. The difference in volumetric size is very small so that interference between the two emitted radiation fields, and the receiving structures at the two different frequencies produces a beat frequency radiation which has a much lower frequency than the incident radiation. The beat frequency radiation preferably is at a frequency which is sufficiently low that it may be relatively easily converted to useable electrical energy. In contrast, the incident zero point radiation has its desirable high energy densities at frequencies which are so high that conventional systems for converting the radiation to electrical energy either cannot effectively or efficiently so convert the radiation energy or simply cannot be used to convert the radiation energy for other reasons.

The system of the present invention also includes an antenna which receives the beat frequency radiation. The antenna may be a conventional metallic antenna such as a loop or dipole type of antenna or a rf cavity structure which partially encloses the receiving structures. The antenna feeds the radiation energy to an electrical conductor (in the case of a conventional dipole or comparable type of antenna) or to a waveguide (in the case of a rf cavity structure). The conductor or waveguide feeds the electrical current (in the case of the electrical conductor) or the electromagnetic radiation (in the case of the waveguide) to a converter which converts the received energy to useful electrical energy. The converter preferably includes a tuning circuit or comparable device so that it can effectively receive the beat frequency radiation. The converter may include a transformer to convert the energy to electrical current having a desired voltage. In addition, the converter may also include a rectifier to convert the energy to electrical current having a desired waveform.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a plan view of the receiving structures and antenna of a first embodiment of the system of the present invention with a schematic view of the conductor and converter thereof and also showing the incident primary and emitted secondary electromagnetic radiation.

FIG. **2** is a front view of the receiving structures, antenna and waveguide of a second embodiment of the system of the present invention with a schematic view of the converter thereof and also showing the incident primary and emitted secondary electromagnetic radiation.

FIG. **3** is a perspective view of the receiving structures, antenna and waveguide of the second embodiment shown in FIG. **2** with a schematic view of the converter thereof and also showing the incident primary and emitted secondary electromagnetic radiation.

FIG. **4** is a front view of the substrate and a plurality of pairs of the receiving structures and a plurality of antennas of a third embodiment of the system of the present invention with a schematic view of the conductor and converter thereof and also showing the incident primary and emitted secondary electromagnetic radiation.

FIG. **5** is a top view of some of the components of the third embodiment of the system of the present invention showing two of the plurality of pairs of receiving structures and two of the plurality of antennas mounted on the substrate.

FIG. **6** is a diagram of a receiving structure of the system of the present invention showing an incident electromagnetic plane wave impinging on the receiving structure and illustrating the directions of the electric and magnetic field vectors thereof.

FIG. **7** is a diagram of a spherical coordinate system as used in the formulas utilized in the system of the present invention.

FIG. **8** is a graph showing an imaginary ρ parameter plotted against a real ρ parameter illustrating the values thereof at resonance as well as values thereof at other than resonance.

FIG. **9** is a graph showing a portion of the graphical representation shown in FIG. **8** illustrating the real and imaginary ρ values at or near a single resonance.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring to the drawings, a first embodiment of the present invention is generally designated by the numeral **10**. The system **10** includes a first and second means for receiving **12** and **14** incident electromagnetic radiation **16**. The means for receiving **12** and **14** are preferably a pair of spherical structures **12** and **14** which are preferably composed of a dielectric material. Alternatively, the spheres **12** and **14** may be cubical structures or any other suitable shape. The spheres **12** and **14** may be mounted on a suitable foundation by any suitable mounting means (not shown), or spheres **12** and **14** may be suspended from a suitable foundation by any suitable suspension means (not shown). The spheres **12** and **14** are preferably composed of a dielectric material. The dielectric spheres **12** and **14** scatter and concentrate electromagnetic waves. At very sharpely defined frequencies, the spheres **12** and **14** will have resonances wherein the internal energy densities can be five orders of mangitude larger than the energy density of the incident electromagnetic field driving the spheres **12** and **14**. At resonance, the electromagnetic stresses, equivalent to pressures proportional to the energy density, can cause material deformation of the spheres **12** and **14** which produce a secondary electromagnetic field. The spheres **12** and **14** are preferably positioned proximal to each other, as shown in FIG. **1**. Although the proximity of the spheres to each other will adversely affect the resonances, the very high "Q"s of the isolated-sphere resonances results in such adverse affect being relatively small. However, the proximity of the spheres **12** and **14** allows the spheres to interact electromechanically which increases the magnitude of the secondary radiation emitted therefrom.

The electromagnetic radiation incident upon the spheres **12** and **14** which drives the spheres to resonance is preferably zero point radiation **16**. However, other types of electromagnetic radiation may also be used to drive the spheres **12** and **14**, if desired.

The effect of a dielectric sphere such as **12** or **14** on an incident electromagnetic radiation such as a plane wave thereof is shown in FIG. **6**. The plane wave propagates in the z axis direction and is diffracted by the sphere **12** resulting in scattering thereof. This scattering is commonly known as Mie scattering. The incident radiation wave has an electric vector component which is linearly polarized in the x axis direction and a magnetic vector component which is linearly polarized in the y axis direction.

An electromagnetic wave incident upon a structure produces a forced oscillation of free and bound charges in synch with the primary electromagnetic field of the incident electromagnetic wave. The movements of the charges produce a

secondary electromagnetic field both inside and outside the structure. The secondary electromagnetic radiation comprising this secondary electromagnetic field is shown in FIG. **1** and designated by the numerals **18** and **20**. An antenna which is shown simply as a loop antenna but may also be a dipole or any other suitable type of antenna is also shown in FIG. **1** and designated by the numeral **22**. The nonlinear mutual interactions of the spheres produces interference between the secondary electromagnetic radiation **18** and **20** produces a beat frequency radiation **24** which is preferably at a much lower frequency than the primary radiation **16**. It is this beat frequency radiation **24** which is desired for conversion into electrical energy because it preferably is within the frequency range of rf radiation which may be converted into electrical energy by generally conventional systems. Thus, the radiation **24** received by the antenna **22** is fed via an electrical conductor **26** to a means for converting the beat frequency radiation **24** to electrical energy. This means for converting is designated by the numeral **28** and preferably includes a tuning capacitor **30** and a transformer **32** and a rectifier (preferably a diode) **34**. Instead of including the capacitor **30**, transformer **32** and rectifier **34**, the converter **28** may alternatively include an rf receiver of any suitable type.

The resultant field at any point is the vector sum of the primary and secondary fields. For the equations that follow, the structure receiving the incident plane wave is a sphere of radius a having a propagation constant k_1 positioned in an infinite, homogeneous medium having a propagation constant k_2. The incident plane wave propagates in the z axis direction and is as shown in FIG. **6**. The spherical coordinate system used for the vector spherical wave functions is shown in FIG. **7**. Expansion of the incident field provides:

$$E_i = E_o e^{-i\omega t} \sum_{n=1}^{\infty} i^n \frac{2n+1}{n(n+1)} \left(m^{(1)}_{01n} - i n^{(1)}_{eln}\right)$$

$$H_i = -\frac{k_2}{\omega\mu_2} E_o e^{-i\omega t} \sum_{n=1}^{\infty} i^n \frac{2n+1}{n(n+1)} \left(m^{(1)}_{eln} + i n^{(1)}_{01n}\right)$$

where E is the electric field and H is the magnetic field; and

$$\overset{(1)}{m}_{\substack{01n\\ e1n}} = \pm \frac{1}{\sin\theta} j_n(k_2R) P_n{}^1(\cos\theta) \begin{matrix}\cos\\ \sin\end{matrix} \phi i_2 - j_n(k_2R) \frac{\partial p_n{}^1}{\partial\theta} \begin{matrix}\sin\\ \cos\end{matrix} \phi i_3$$

$$\overset{(1)}{n}_{\substack{0ln\\ e1n}} = \frac{n(n+1)}{k_2R} j_n(k_2R) P_n{}^1(\cos\theta) \begin{matrix}\sin\\ \cos\end{matrix} \phi i_1 + \frac{1}{k_2R} [k_2Rj_n(k_2R)]' \times$$

$$\frac{\partial P_n{}^1}{\partial\theta} \begin{matrix}\sin\\ \cos\end{matrix} \phi i_2 \pm \frac{1}{k_2R\sin\theta} [k_2Rj_n(k_2R)]' P_n{}^1(\cos\theta) \begin{matrix}\cos\\ \sin\end{matrix} \phi i_3.$$

The electric and magnetic fields of the incident wave transmitted into the sphere i.e., R<a, can be similarly expanded:

$$E_t = E_0 e^{-i\omega t} \sum_{n=1}^{\infty} i^n \frac{2n+1}{n(n+1)} \left(a_n^{tm}\,{}^{(1)}_{01\mathrm{n}} - i b_n^{tn}\,{}^{(1)}_{\mathrm{eln}} \right)$$

$$H_t = \frac{k_2}{\phi\mu_1} E_0 e^{-i\omega t} \sum_{n=1}^{\infty} i^m \frac{2n+1}{n(n+1)} \left(b_n^{tm}\,{}^{(1)}_{\mathrm{eln}} - i a_n^{tn}\,{}^{(1)}_{01\mathrm{n}} \right)$$

If $j_n(k_2R)$ is replaced by $h_n{}^{(1)}(k_2R)$ in the previous equations, the functions $m^{(1)}$ and $n^{(1)}$ become $m^{(3)}$ and $n^{(3)}$. The outgoing fields i.e., R>a, are represented by:

$$E_R = E_0 e^{-i\omega t} \sum_{n=1}^{\infty} i^n \frac{2n+1}{n(n+1)} \left(a_n^{rm}\,{}^{(3)}_{01\mathrm{n}} - i b_n^{rn}\,{}^{(3)}_{\mathrm{eln}} \right)$$

$$H_r = \frac{k_2}{\phi\mu_I} E_0 e^{-i\omega t} \sum_{n=1}^{\infty} i^m \frac{2n+1}{n(n+1)} \left(b_n^{rm}\,{}^{(3)}_{\mathrm{eln}} - i a_n^{rn}\,{}^{(3)}_{01\mathrm{n}} \right)$$

where H_r represents the resultant wave in the medium surrounding the sphere. At resonance, the values of ρ at resonance require that the $a_n{}^t$ and $b_n{}^t$ coefficients be infinite. In order to determine these values of $a_n{}^t$ and $b_n{}^t$, the boundary conditions at the sphere radius are needed. Since there must be continuity of the E and H values at the surface, the following equations are used:

$$i_1\times(E_i+E_r)=i_1\times E_t \text{ and}$$

$$i_1\times(H_i+H_r)=i_1\times H_t$$

which lead to two pairs of inhomogeneous equations:

$$a_n{}^tj_n(N\rho)-a_n{}^rh_n{}^{(1)}(\rho)=j_n(\rho)$$

$$\mu_2a_n{}^t[N\rho j_n(N\rho)]'-\mu_1a_n{}^r[\rho h_n{}^{(1)}(\rho)]'=\mu_1[\rho j_n(\rho)]' \text{ and}$$

$$\mu_2Nb_n{}^tj_n(N\rho)-\mu_1b_n{}^rh_n{}^{(1)}(\rho)=\mu_1j_n(\rho)$$

$$b_n{}^t[N\rho j_n(N\rho)]'-Nb_n{}^r[\rho h_n{}^{(1)}(\rho)]'=N[\rho j_n(\rho)]'$$

where $k_1=Nk_2$, $\rho=k_2a$, $k_1a=N\rho$. Spherical Bessel functions of the first kind are denoted by j_n, while those of the third kind are denoted by $h_n{}^{(1)}$. The resulting equations are:

$$a_n{}^t = \frac{\mu_1 j_n(\rho)[\rho h_n^{(1)}(\rho)]' - \mu_1 h_n^{(1)}\rho)[\rho j_n(\rho)]'}{\mu_1 j_n(N\rho)[\rho h_n^{(1)}(\rho)]' - \mu_2 h_n^{(1)}(\rho)[N\rho j_n(N\rho)]'}$$

and

$$b_n{}^t = \frac{\mu_1 N j_n(\rho)[\rho h_n^{(1)}(\rho)]' - \mu_1 N h_n^{(1)}\rho)[\rho j_n(\rho)]'}{\mu_2 N^2 j_n(N\rho)[\rho h_n^{(1)}(\rho)]' - \mu_1 h_n^{(1)}(\rho)[N\rho j_n(N\rho)]'}$$

At a resonance, the denominator of either $a_n{}^t$ or $b_n{}^t$ will be zero. Thus, ρ values are found using the above equations that correspond to a resonant combination of angular frequency (ω) and radius (a) for a given sphere material and given surrounding medium. In determining such values of ρ, the following equations are also specifically used:

$$\rho = ak_2 = a\omega\sqrt{\epsilon_2\mu_2} \text{ and}$$

$$\rho_1 = (k_l/k_2)\rho$$

where ρ_1 corresponds to the sphere material. An iterative method is preferably used to find the desired values of ρ at resonance. In calculating ρ utilizing the above equations for purposes of example, it was assumed that $\mu_1=\mu_2=\mu_0=4\pi\times10^{-7}$ and $\epsilon_2=\epsilon_0=8.85419\times10^{-12}$.

One major root of ρ which was found has a value of:

Real (ρ)=+66.39752607619131

Imaginary (ρ)=−0.6347867071968998.

These particular values are not shown in FIG. **8**. However, other values of ρ found using the equations set forth herein are shown in FIG. **8**. The peaks in FIG. **8** are the resonances. One of these resonances shown in FIG. **8** is shown in detail in FIG. **9**. These resonance values are shown for purposes of example. Other resonances also exist which have not been determined; thus, not all possible resonance values are shown in FIGS. **8** and **9**.

Calculation of these values also allows the determination of a possible am combination which would have these root values. For ρ, ε (epsilon)=ϵ_0 and $\mu=\mu_0$, and

$$\rho = a\omega\sqrt{\epsilon_0\mu_0} = a\omega/c.$$

Expressed in SI units, the speed of light $c=2.99792458\times10^{14}$ m/s. If an a value of 10^{-6} m is assumed for the examples shown herein, then:

$$\omega=\rho c/a\approx1.9919\times10^{16}-i1.9044\times10^{14}\text{ radians/s.}$$

This is an example of the angular frequency required within the impingent EM radiation in order to create a resonant situation. Examples of other resonances were indicated, and these are shown in FIG. **8**. No complex-frequency plane waves exist. Therefore, the calculations were made by considering only the real portion of the above root and setting the imaginary portion equal to zero. However, upon doing this, the iterative calculation procedure becomes insensitive to any root in the vicinity of the root's real portion. In the iterative calculation procedure, initially a range of ρ values is input into the equations. These ρ values are in the neighborhood of the prospective root. A range of ρ values is subsequently studied to find any imaginary ρ i.e., fρ (a function of ρ), peaks in that range. Next, once a peak has been chosen, the function order n giving the dominant fρ is determined. This also gives a clue as to whether the peak is due to a magnetic resonance (a_n approaches infinity) or an electrical resonance (b_n approaches infinity). A large number of Newton-Raphson iterations is preferably performed in order to converge upon a root ρ value.

FIGS. **2** and **3** show a second embodiment of the present invention generally designated by the numeral **110**. Embodiment **110** is essentially the same as embodiment **10** except that the antenna is a rf cavity structure **122** which feeds the received beat frequency radiation **124** to a waveguide **126**. Embodiment **110** also preferably includes two spheres **112** and **114** which receive the primary incident electromagnetic radiation **116** and emit the secondary electromagnetic radiation **118** and **120**. As with the spheres **18** and **20** of embodiment **10**, spheres **118** and **120** are preferably composed of a dielectric material. Embodiment **110** also includes converter **128**, capacitor **130**, transformer **132** and rectifier **134** which are essentially identical to the correspondingly numbered elements of embodiment **10**. Therefore, a description of these components of embodiment **110** will not be repeated in order to promote brevity. In addition, the same equations and method of calculation set forth above with regard to embodiment **10** also apply to embodiment. Therefore, their description will not be repeated in order to promote brevity.

FIGS. **4** and **5** show a third embodiment of the present invention generally designated by numeral **210**. Embodiment **210** is essentially identical to the first embodiment **10** except that the embodiment **210** includes a plurality of pairs **215** of receiving means (spheres) **212** and **214** mounted on a substrate **236**. The spheres **212** and **214** are thus in the form of an array **238**. The pairs **215** of the array **238** are preferably positioned proximal to each other in order to maximize the amount of energy extracted from a particular area or space of a given size. Since, as set forth hereinabove, the energy density of the zero point radiation increases as the frequency of the radiation increases, it is desirable that the spheres resonate at as high a bandwidth of frequencies as possible. Because the spheres **212** and **214** must be small in direct proportion to the wavelength of the high frequencies of the incident electromagnetic radiation **216** at which resonance is desirably obtained, the spheres **212** and **214** are preferably microscopic in size. Current lithographic techniques are capable of manufacturing such microscopically small spheres mounted on a suitable substrate thereby providing a suitably miniaturized system **210**. A miniaturized system enhances the energy output capability of the system by enabling it to resonate at higher frequencies at which there are correspondingly higher energy densities. Consequently, utilization of array **238** in the system **210** enhances the maximum amount of electrical energy provided by the system **210**.

Lithographic techniques may be more amenable to manufacturing microscopically small receiving structures **212** and **214** which may be disc shaped, semispherical or have another shape other than as shown in FIGS. **4** and **5**. Consequently, the receiving means **212** and **214** may accordingly have such alternative shapes rather than the spherical shape shown in FIGS. **4** and **5**. In addition, a large number of small spheres may be manufactured by bulk chemical reactions. Packing a volume with such spheres in close proximity could enhance the output of energy.

Embodiment **210** also includes a plurality of antennas **222** positioned preferably between the spheres **212** and **214** which receive the beat frequency radiation **224** produced by the interference between the secondary radiation **218** and **220**. The antennas **222** are shown as loop antennas **222** but may be any other suitable type of antennas as well.

Embodiment **210** has a plurality of electrical conductors **226** which preferably include traces mounted on the substrate **236** which occupies a finite volume. The electrical conductors **226** feed the electrical output from the antennas **222** to a suitable converter **228** which preferably includes tuning capacitor **230**, transformer **232** and rectifier **234**, as with embodiments **10** and **110**. Except as set forth above, the components of embodiment **210** are identical to embodiment **10** so the detailed description of these components will not be repeated in order to promote brevity. In addition, the same equations and method of calculation set forth above for embodiment **10** also apply to embodiment **210**. Therefore, the description of these equations and method of calculation will not be repeated in order to promote brevity.

Accordingly, there has been provided, in accordance with the invention, a system which converts high frequency zero point electromagnetic radiation into electrical energy effectively and efficiently and thus fully satisfies the objectives set forth above. It is to be understood that all terms used herein are descriptive rather than limiting. Although the invention has been specifically described with regard to the specific embodiments set forth herein, many alternative embodiments, modifications and variations will be apparent to those skilled in the art in light of the disclosure set forth herein. Accordingly, it is intended to include all such alternatives, embodiments, modifications and variations that fall within the spirit and scope of the invention as set forth in the claims hereinbelow.

What is claimed is:

1. A system for converting incident electromagnetic radiation energy to electrical energy, comprising:

a first means for receiving incident primary electromagnetic radiation, said means for receiving producing emitted secondary electromagnetic radiation at a first frequency, said first means for receiving having a first volumetric size selected to resonate at a frequency within the frequency spectrum of the incident primary electromagnetic radiation in order to produce the secondary electromagnetic radiation at the first frequency at an enhanced energy density;

- a second means for receiving the incident primary electromagnetic radiation, said means for receiving producing emitted secondary electromagnetic radiation at a second frequency, the secondary radiation at the first frequency and the secondary radiation at the second frequency interfering to produce secondary radiation at a lower frequency than that of the incident primary radiation, said second means for receiving having a second volumetric size selected to resonate at a frequency within the frequency spectrum of the incident primary electromagnetic radiation in order to produce the emitted secondary electromagnetic radiation at the second frequency at an enhanced energy density;
- an antenna for receiving the emitted secondary electromagnetic radiation at the lower frequency, said antenna providing an electrical output responsive to the secondary electromagnetic radiation received;
- a converter electrically connected to said antenna for receiving electrical current output from said antenna and converting the electrical current output to electrical current having a desired voltage and waveform.

2. The system of claim **1** wherein:

- said first means for receiving is composed of a dielectric material; and
- said second means for receiving is composed of a dielectric material.

3. The system of claim **1** wherein:

- said first means for receiving is spherical; and
- said second means for receiving is spherical.

4. A system for for converting incident zero point electromagnetic radiation energy to electrical energy, comprising:

- a first means for receiving incident primary zero point electromagnetic radiation, said means for receiving producing emitted secondary electromagnetic radiation at a first frequency;
- a second means for receiving the incident primary zero point electromagnetic radiation, said means for receiving producing emitted secondary electromagnetic radiation at a second frequency, the secondary radiation at the first frequency and the secondary radiation at the second frequency interfering to produce secondary radiation at a beat frequency which is lower than that of the incident primary radiation;
- an antenna for receiving the emitted secondary electromagnetic radiation at the lower frequency, said antenna providing an electrical output responsive to the secondary electromagnetic radiation received;
- means for transmitting the emitted secondary electromagnetic radiation at the beat frequency from said antenna, said means for transmitting connected to said antenna;
- a converter connected to said means for transmitting for receiving the emitted secondary electromagnetic radiation at the beat frequency from said antenna and converting the same to electrical current having a desired voltage and waveform.

5. The system of claim **4** wherein:

- said first means for receiving has a first volumetric spherical size selected to resonate in response to the incident primary electromagnetic radiation in order to produce the secondary electromagnetic radiation at the first frequency at an enhanced energy density; and
- said second means for receiving has a second volumetric spherical size selected to resonate in response to the incident primary electromagnetic radiation in order to produce the emitted secondary electromagnetic radiation at the second frequency at an enhanced energy density, said first and second volumetric sizes selected based on parameters of propagation constant of said first and second means for receiving, propagation constant of medium in which said first and second means for receiving are located and frequency of the incident primary electromagnetic radiation.

6. The system of claim **5** wherein the first and second volumetric sizes are selected by utilizing the formulas:

$$a_n{}^t=\frac{\mu_1 j_n(\rho)[\rho h_n^{(1)}(\rho)]'-\mu_1 h_n^{(1)}\rho)[\rho j_n(\rho)]'}{\mu_1 j_n(N\rho)[\rho h_n^{(1)}(\rho)]'-\mu_2 h_n^{(1)}(\rho)[N\rho j_n(N\rho)]'}$$

$$b_n{}^t=\frac{\mu_1 N\mathrm{j}_n(\rho)[\rho h_n^{(1)}(\rho)]'-\mu_1 \mathrm{N} h_n^{(1)}\rho)[\rho j_n(\rho)]'}{\mu_2 N^2 j_n(N\rho)[\rho h_n^{(1)}(\rho)]'-\mu_1 h_n^{(1)}(\rho)[N\rho j_n(N\rho)]'}$$

$$\rho=a\omega\sqrt{\epsilon_2\mu_2}$$

wherein at a resonance, the denominator of either equation for $a_n{}'$ or $b_n{}'$ will be approximately zero and wherein k_1=propagation constant of the means for receiving, k_2=propagation constant of medium through which the incident electromagnetic radiation propagates, a is the radius of either means for receiving, $N=k_1/k_2$, $\rho=k_2a$, $k_1a=N\rho$, $a_n{}'$=magnitude of oscillations of the electric field of the nth order, $b_n{}'$=magnitude of oscillations of the magnetic field of the nth order, ω=angular frequency of the incident electromagnetic radiation, ϵ is the permittivity of the material or medium and μ is the permeability of the material or medium.

7. The system of claim **6** wherein the radius of the first means for receiving is different from the radius of the second means for receiving, difference between the radius of said first means for receiving and the radius of said second means for receiving selected so that the beat frequency resulting from the difference is a frequency which facilitates conversion of the beat frequency electromagnetic radiation to electrical energy.

8. The system of claim **4** wherein:

- said first means for receiving is composed of a dielectric material; and
- said second means for receiving is composed of a dielectric material.

9. The system of claim **4** wherein:

- said first means for receiving is spherical; and
- said second means for receiving is spherical.

10. The system of claim **4** wherein said antenna is positioned generally between said first and second means for receiving.

11. The system of claim **4** wherein said antenna is a loop antenna.

12. The system of claim **4** wherein said antenna is a generally concave shell partially enclosing said first and second means for receiving.

13. The system of claim **4** wherein said means for transmitting is a waveguide.

14. A system for for converting incident zero point electromagnetic radiation energy to electrical energy, comprising:

- a substrate;
- a plurality of pairs of first means for receiving incident primary zero point electromagnetic radiation and second means for receiving incident primary zero point electromagnetic radiation, said plurality of pairs of means for receiving mounted on said substrate, said first means for receiving producing emitted secondary

electromagnetic radiation at a first frequency, said second means for receiving the incident primary zero point electromagnetic radiation producing emitted secondary electromagnetic radiation at a second frequency, the secondary radiation at the first frequency and the secondary radiation at the second frequency interfering to produce secondary radiation at a beat frequency which is lower than that of the incident primary radiation, said first means for receiving having a first volumetric size selected to resonate in response to the incident primary electromagnetic radiation in order to produce the secondary electromagnetic radiation at the first frequency at an enhanced energy density, and said second means for receiving having a second volumetric size selected to resonate in response to the incident primary electromagnetic radiation in order to produce the emitted secondary electromagnetic radiation at the second frequency at an enhanced energy density, said first and second volumetric sizes selected based on parameters of propagation constant of said first and second means for receiving, propagation constant of medium in which said first and second means for receiving are located and frequency of the incident primary electromagnetic radiation, said first and second volumetric sizes being different from each other;

a plurality of antennas for receiving the emitted secondary electromagnetic radiation at the lower frequency, said antenna providing an output responsive to the secondary electromagnetic radiation received, said plurality of antennas mounted on said substrate, each of said plurality of antennas receiving the emitted secondary electromagnetic radiation of one of said pairs of first and second means for receiving;

means for transmitting the emitted secondary electromagnetic radiation at the beat frequency from said antenna, said means for transmitting connected to said plurality of antennas;

a converter connected to said means for transmitting for receiving the emitted secondary electromagnetic radiation at the beat frequency from said antenna and converting the same to electrical current having a desired voltage and waveform.

* * * * *

US007057514B2

(12) **United States Patent**
Mickle et al.

(10) **Patent No.: US 7,057,514 B2**
(45) **Date of Patent: Jun. 6, 2006**

---

(54) **ANTENNA ON A WIRELESS UNTETHERED DEVICE SUCH AS A CHIP OR PRINTED CIRCUIT BOARD FOR HARVESTING ENERGY FROM SPACE**

(75) Inventors: **Marlin H. Mickle**, Pittsburgh, PA (US); **Christopher C. Capelli**, Pittsburgh, PA (US); **Harold Swift**, Gibsonia, PA (US)

(73) Assignee: **University of Pittsburgh - Of the Commonwealth System oif Higher Education**, Pittsburgh, PA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **10/857,527**

(22) Filed: **Oct. 5, 2004**

(65) **Prior Publication Data**

US 2005/0030181 A1 Feb. 10, 2005

Related U.S. Application Data

(60) Provisional application No. 60/475,151, filed on Jun. 2, 2003.

(51) **Int. Cl.**
G08B 13/14 (2006.01)

(52) **U.S. Cl.** **340/572.7**; 340/572.4; 343/701; 343/703

(58) **Field of Classification Search** .... 340/572.2, 340/572.6, 572.4, 572.7; 343/895, 879, 795, 343/793, 700 MS, 701, 703; 257/491, 678, 257/679, 778
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,471,344	A *	9/1984	Williams	340/572.2
4,724,427	A	2/1988	Carroll	
4,857,893	A	8/1989	Carroll	
5,296,866	A	3/1994	Sutton	
6,054,925	A *	4/2000	Proctor et al.	340/572.7
6,100,804	A *	8/2000	Brady et al.	340/572.7
6,177,872	B1 *	1/2001	Kodukula et al.	340/572.7
6,289,237	B1	9/2001	Mickle et al.	
6,335,686	B1 *	1/2002	Goff et al.	340/572.4
6,373,447	B1 *	4/2002	Rostoker et al.	343/895
6,615,074	B1	9/2003	Mickle et al.	
6,856,291	B1	2/2005	Mickle et al.	

* cited by examiner

Primary Examiner—Anh V. La
(74) *Attorney, Agent, or Firm*—Arnold B. Silverman; Philip E. Levy; Eckert Seamans Cherin & Mellott, LLC

(57) **ABSTRACT**

A wireless untethered device which may be an integrated circuit chip or printed circuit board having at least one integrated antenna and circuitry operatively associated with the antenna for receipt of energy harvested from space by the antenna. The antenna has an effective area greater than its physical area. A system which receives power from the harvested energy after the energy is processed by the circuitry may also be provided with mismatched impedance between the antenna and a circuit resistance.

12 Claims, 40 Drawing Sheets

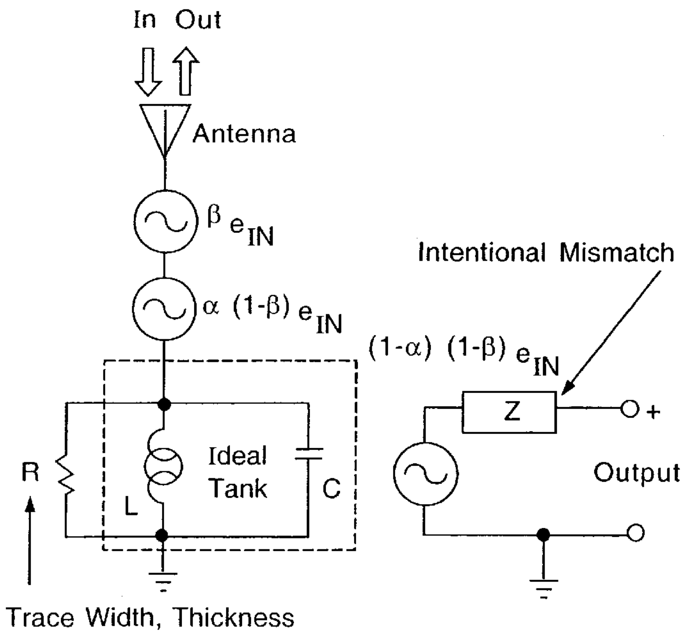

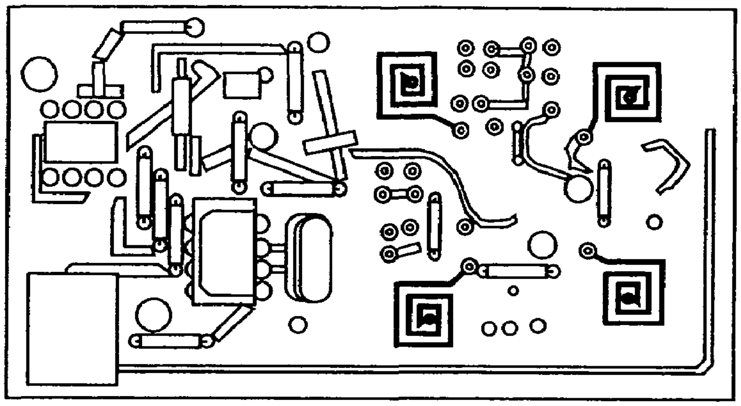

FIG. 1

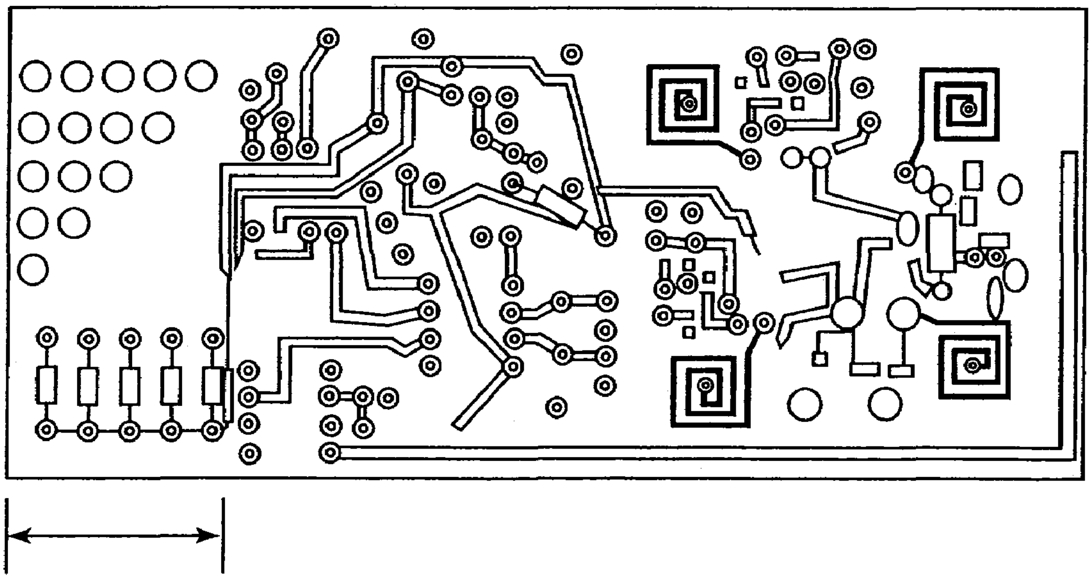

Power and voltage measurement circuitry

FIG. 2

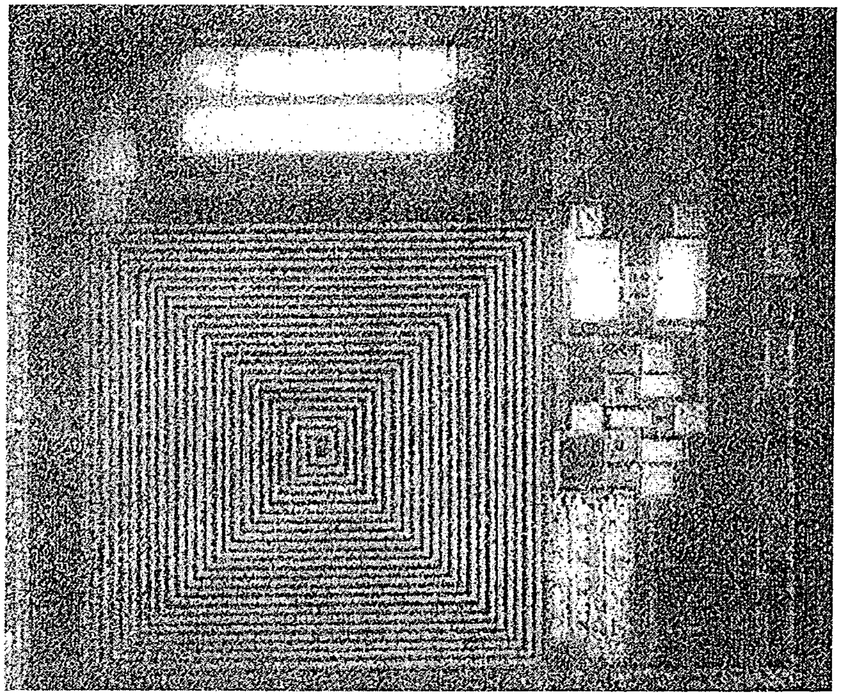

FIG. 3B

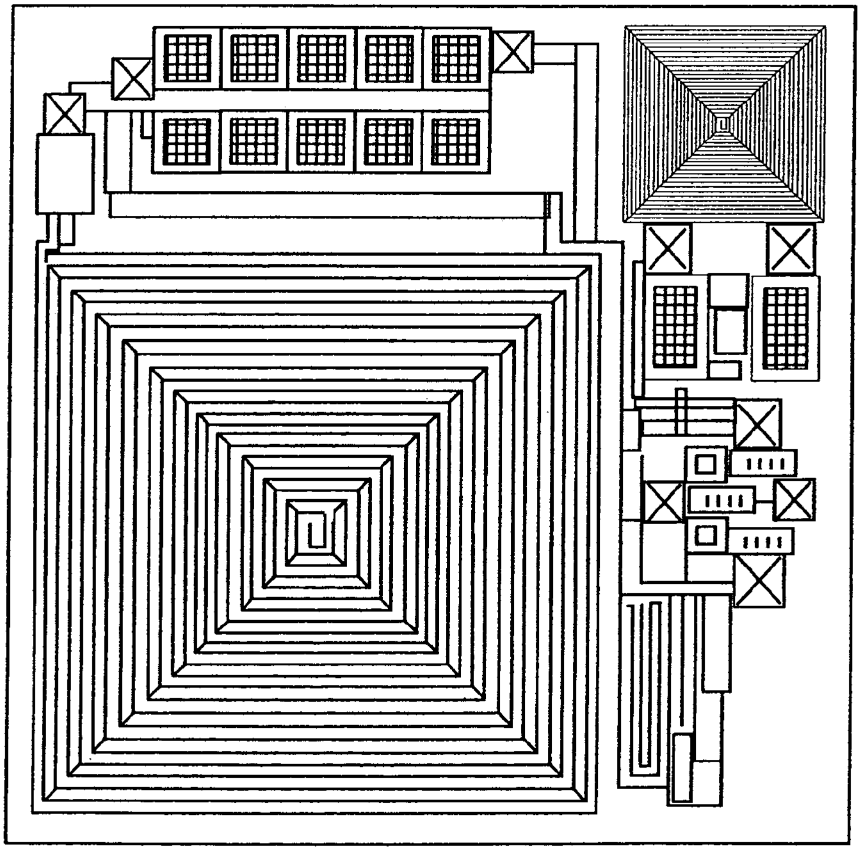

FIG. 3A

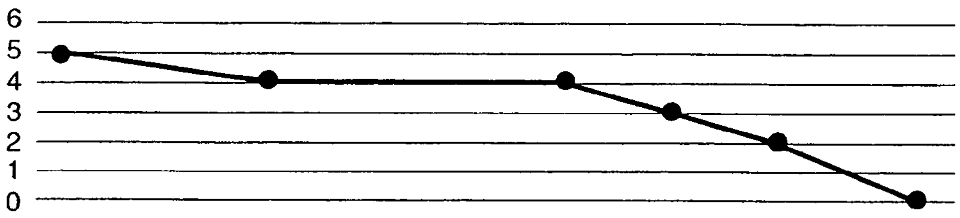
6
5
4
3
2
1
0

FIG. 4a

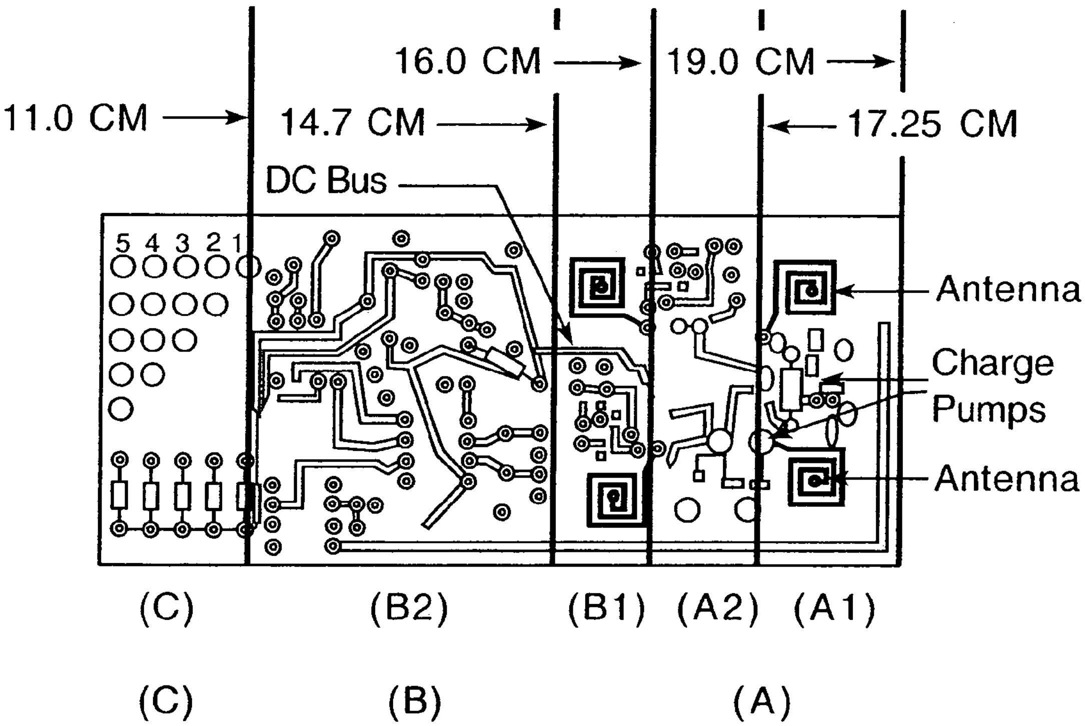
16.0 CM
19.0 CM
11.0 CM
14.7 CM
17.25 CM
DC Bus
5 4 3 2 1
Antenna
Charge Pumps
Antenna
(C)
(B2)
(B1)
(A2)
(A1)
(C)
(B)
(A)

FIG. 4b

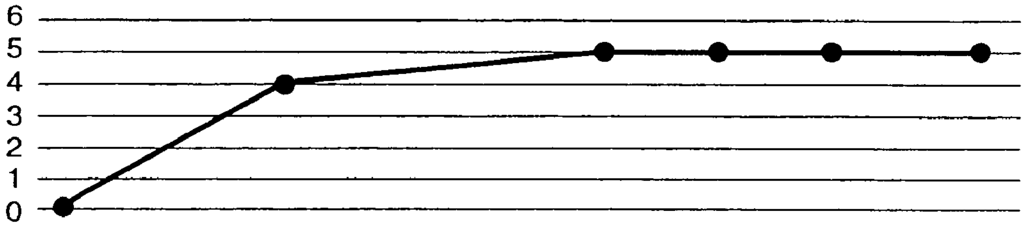
6
5
4
3
2
1
0

FIG. 5a

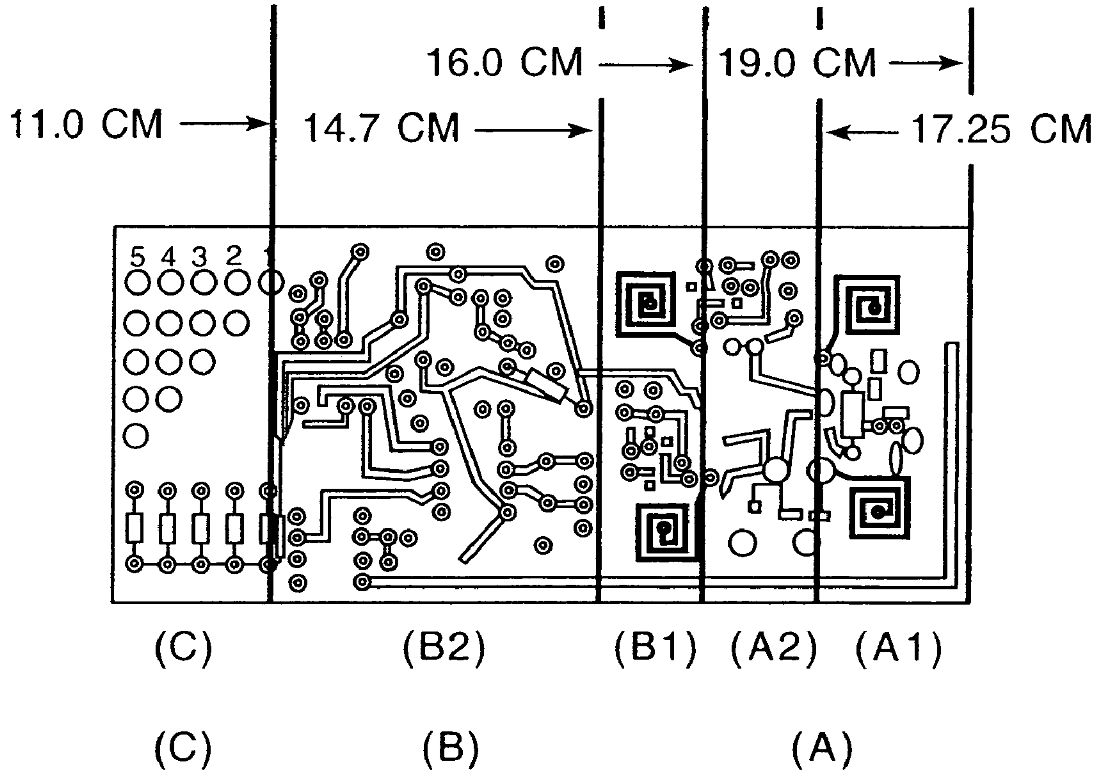
11.0 CM
14.7 CM
16.0 CM
19.0 CM
17.25 CM
5 4 3 2 1
(C)
(B2)
(B1)
(A2)
(A1)
(C)
(B)
(A)

FIG. 5b

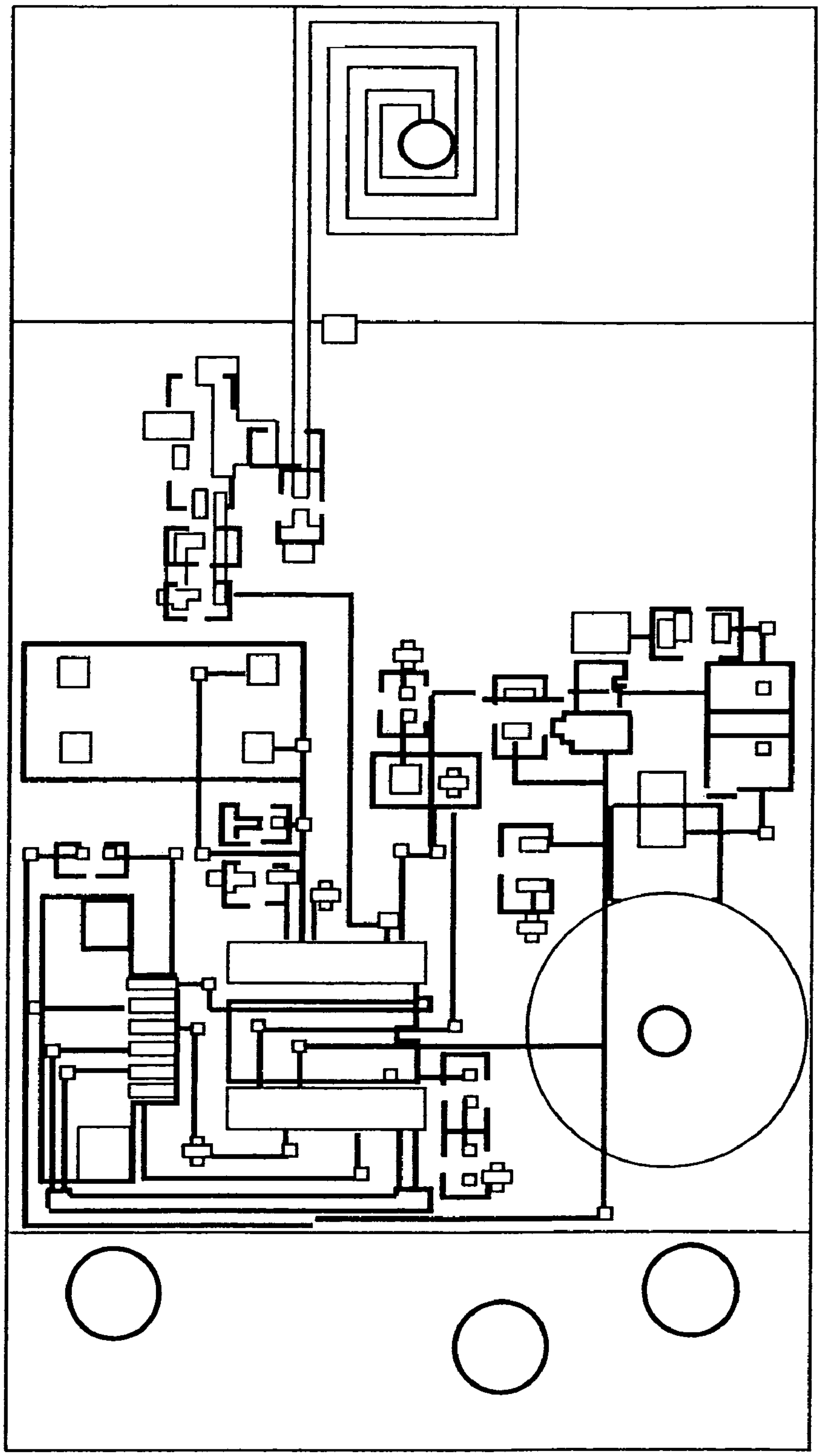

FIG. 6a

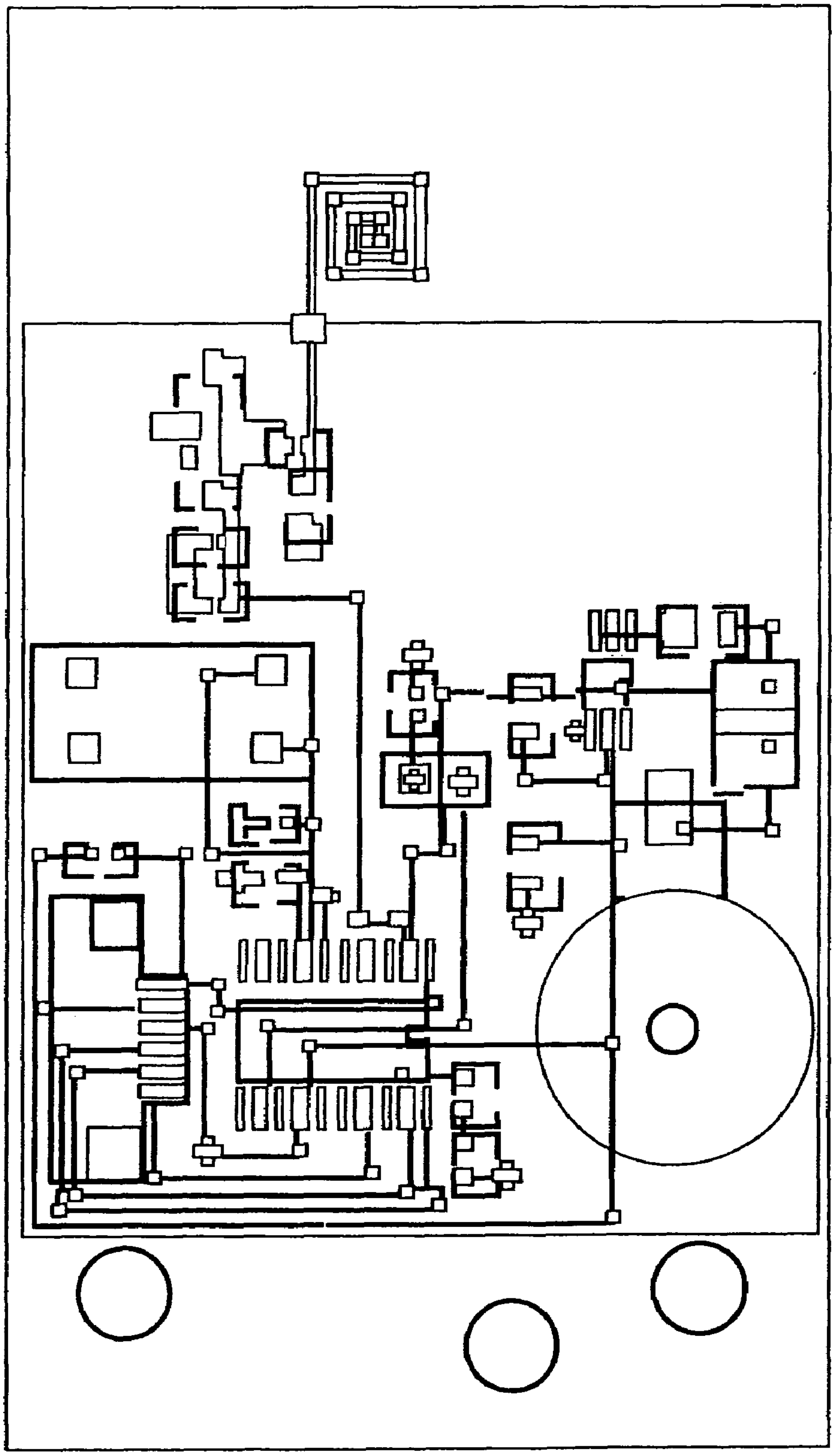

FIG. 6b

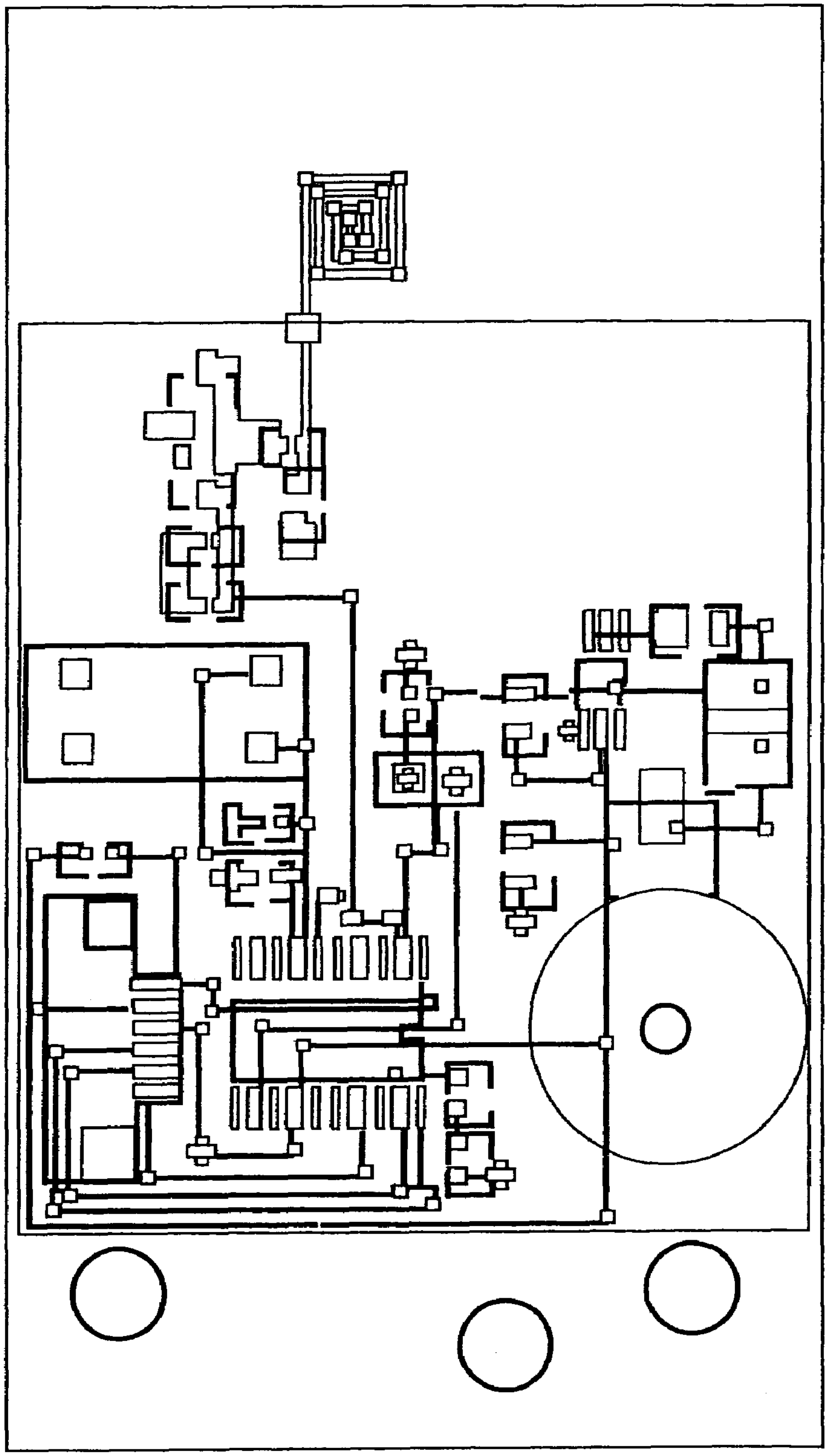

FIG. 6c

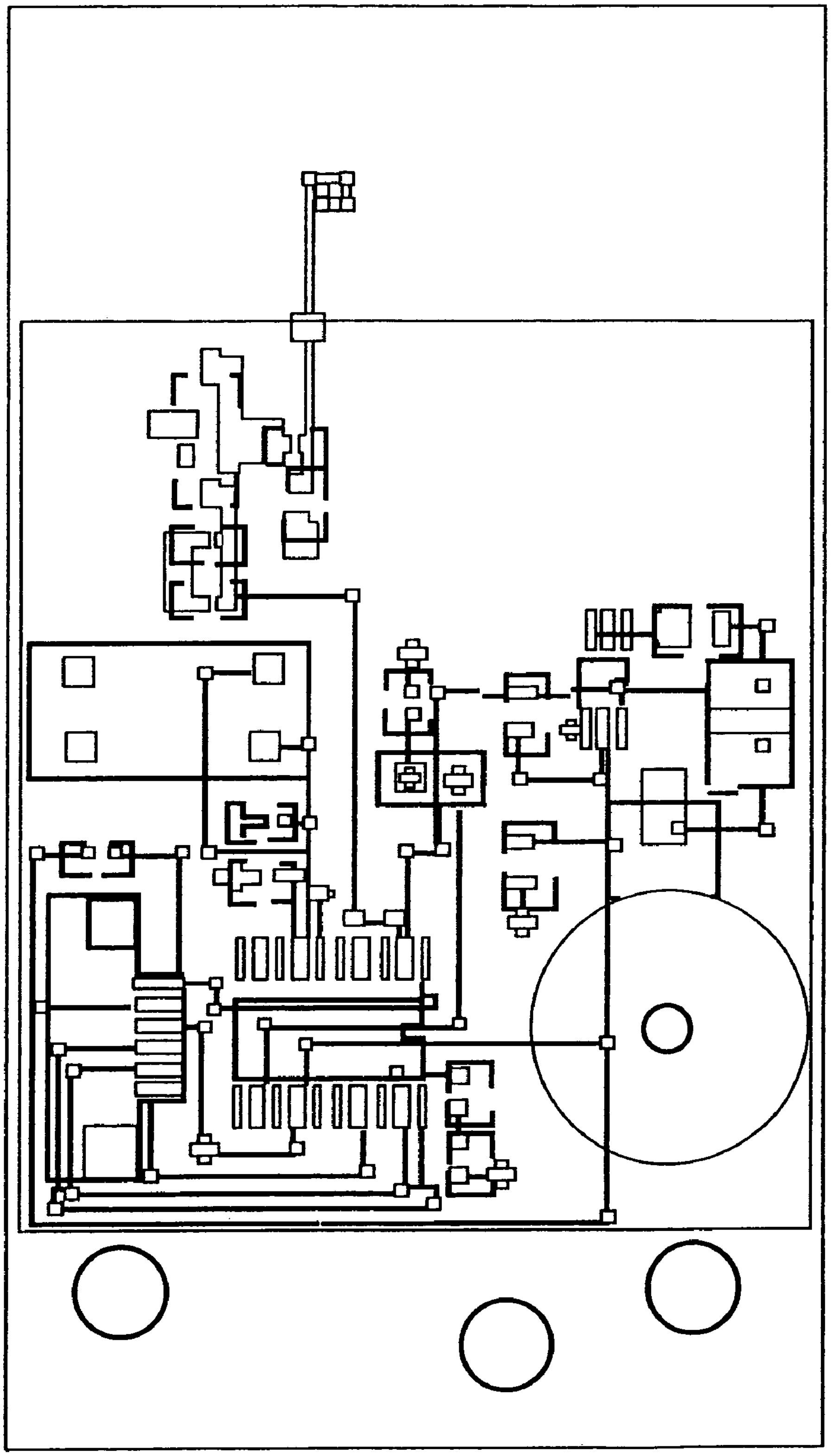

FIG. 6d

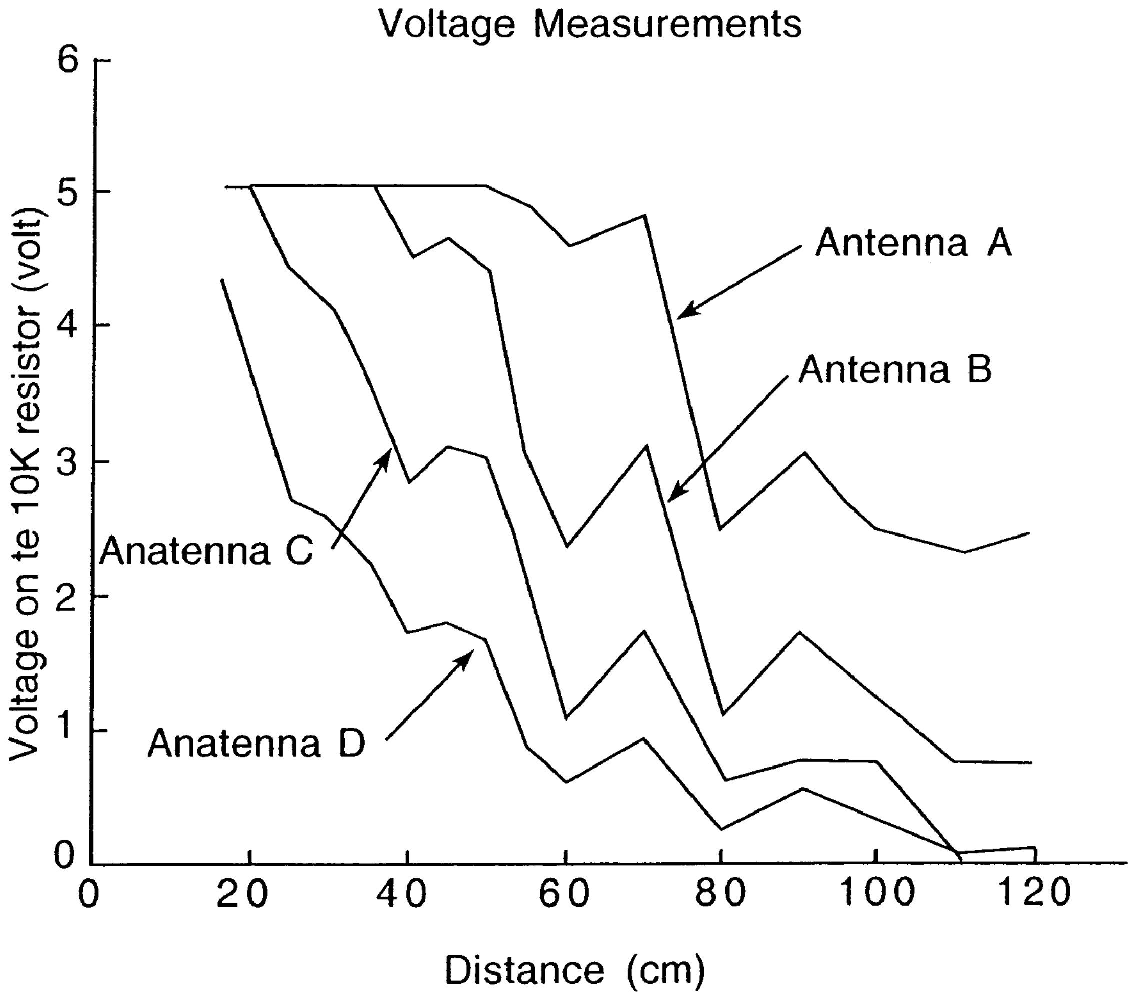
Voltage Measurements
Voltage on te 10K resistor (volt)
0
1
2
3
4
5
6
0
20
40
60
80
100
120
Distance (cm)
Antenna A
Antenna B
Anatenna C
Anatenna D

FIG. 7

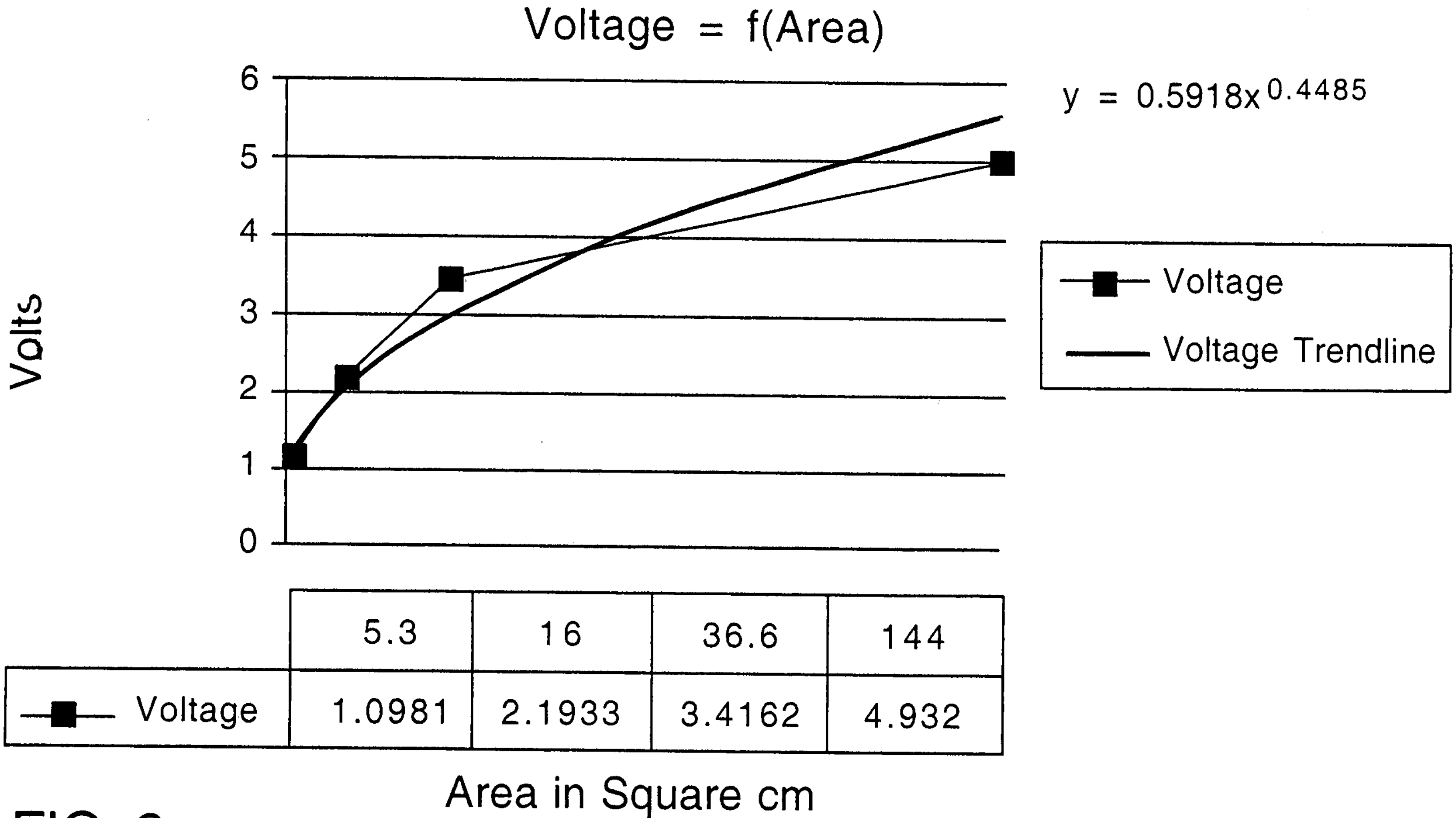
Voltage = f(Area)
y = 0.5918x^0.4485
Volts
6
5
4
3
2
1
0
Voltage
Voltage Trendline
5.3
16
36.6
144
Voltage
1.0981
2.1933
3.4162
4.932
Area in Square cm

FIG. 8

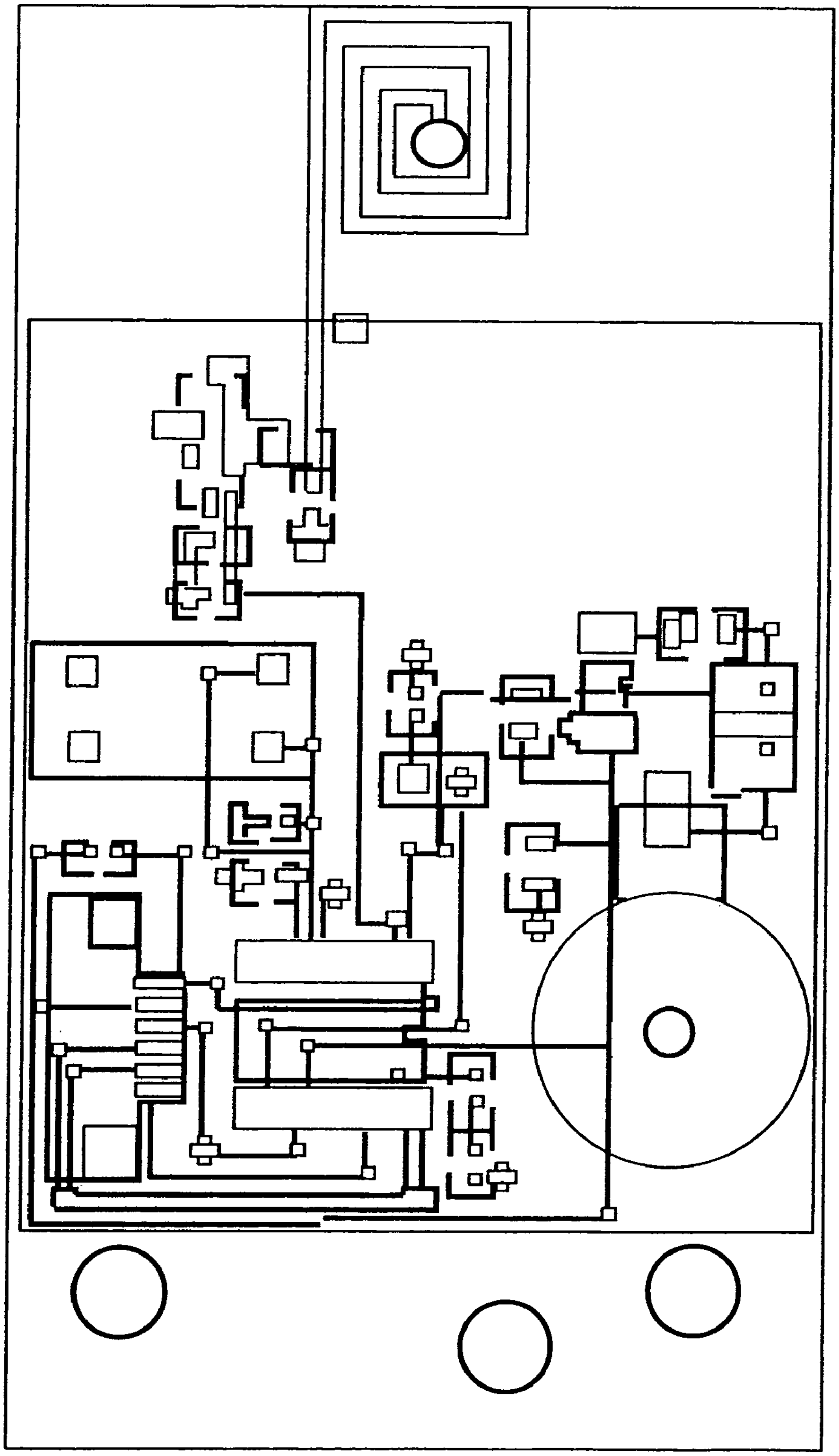

FIG. 9a

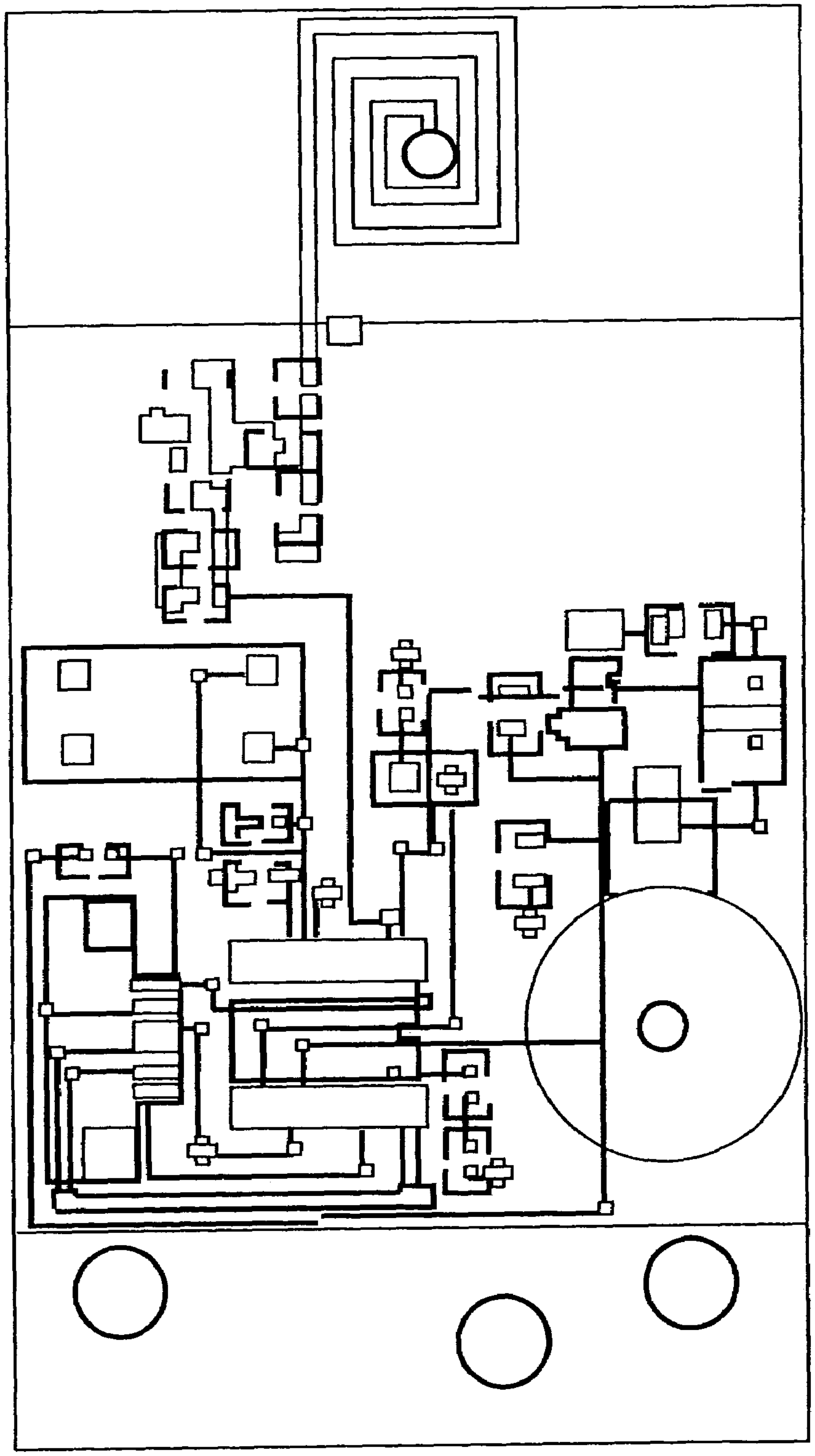

FIG. 9b

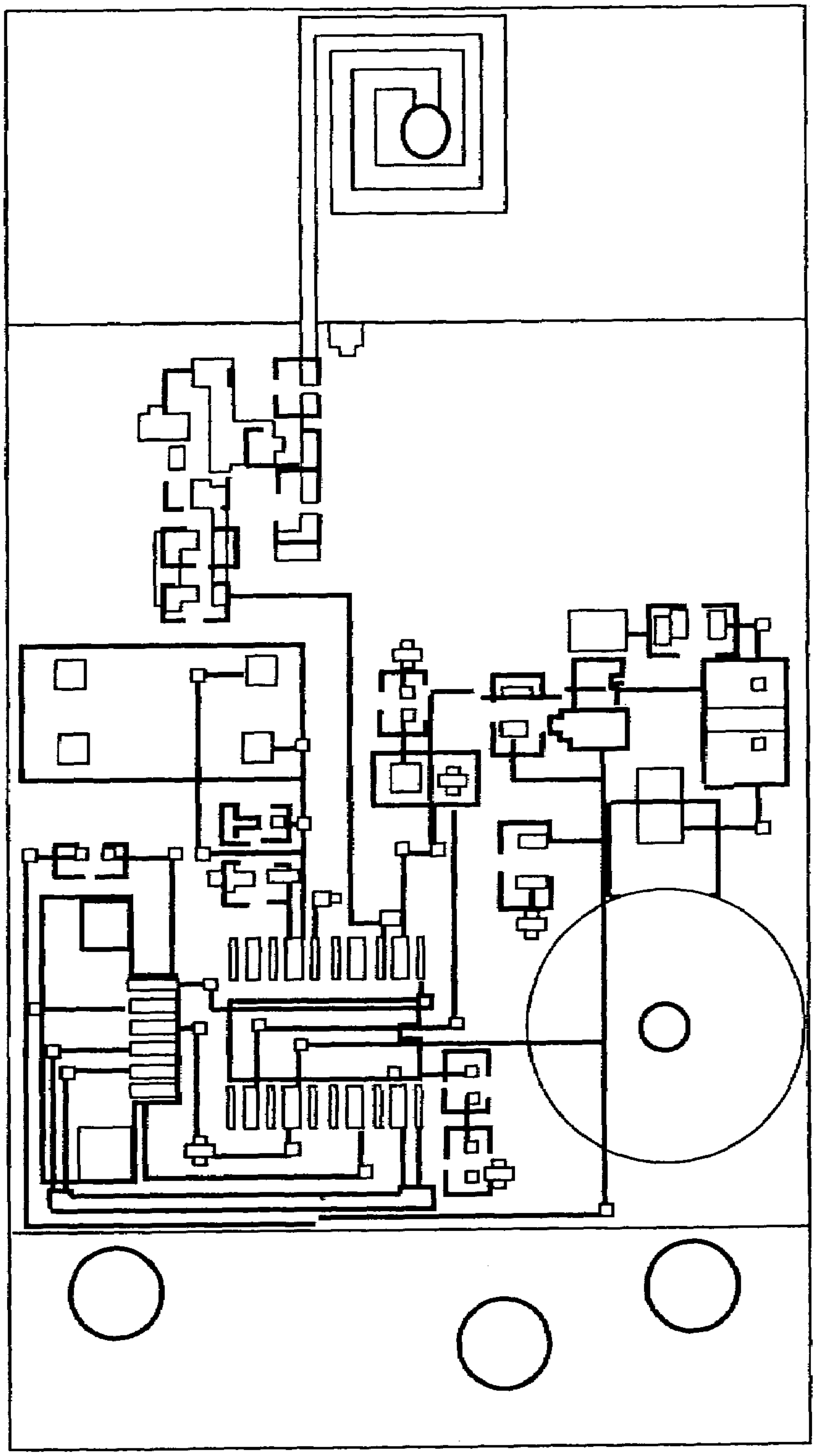

FIG. 9c

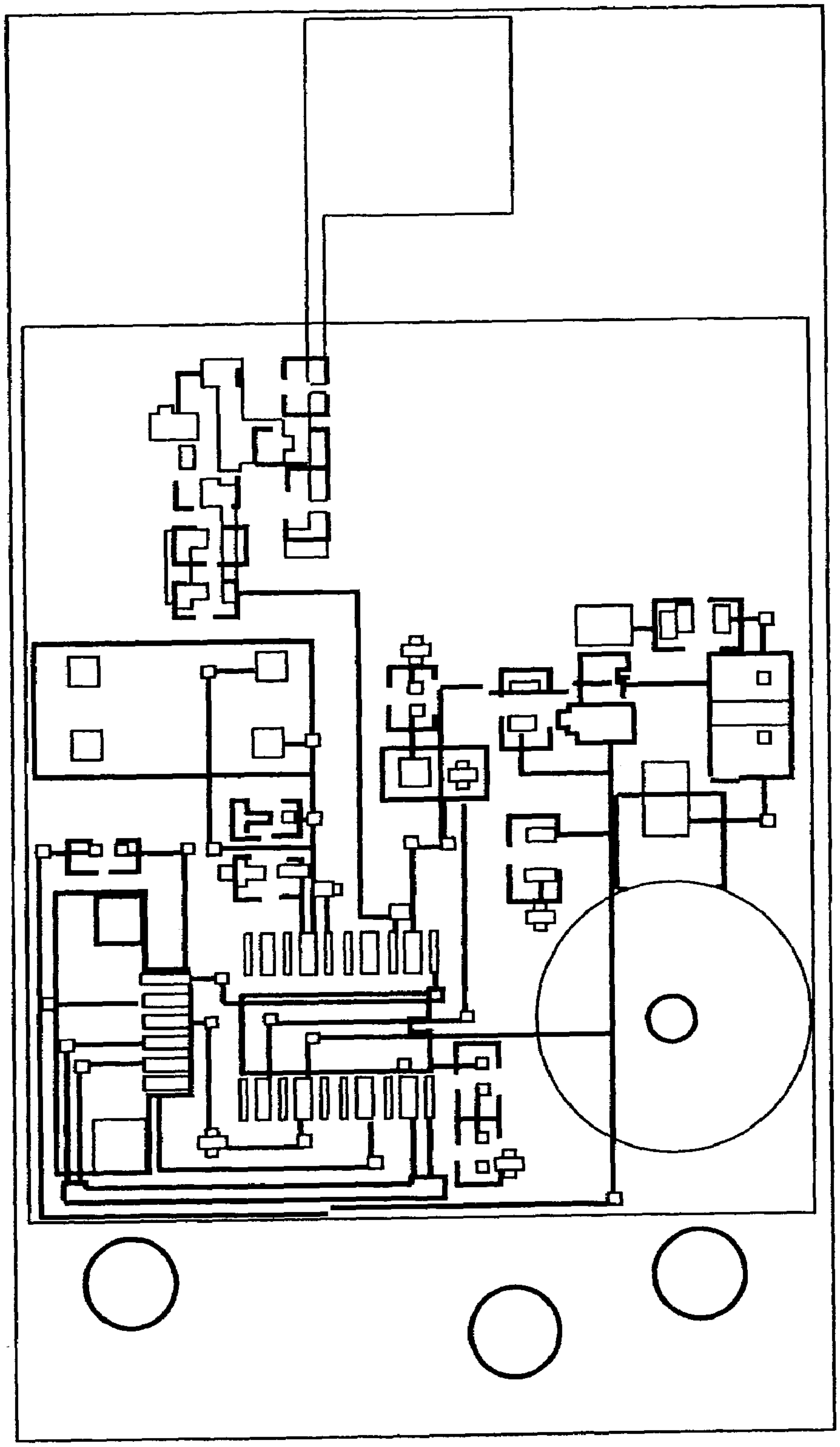

FIG. 9d

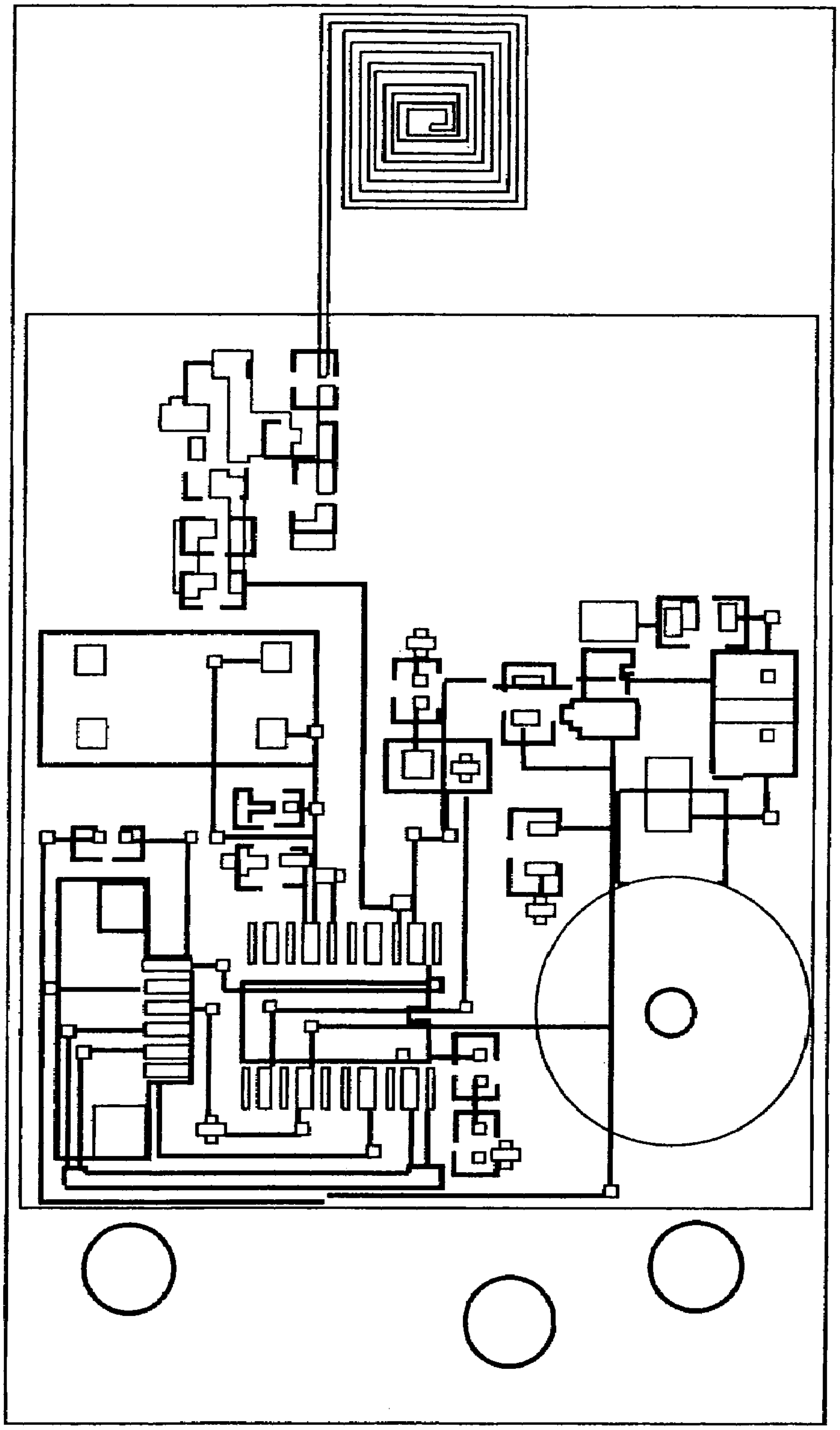

FIG. 9e

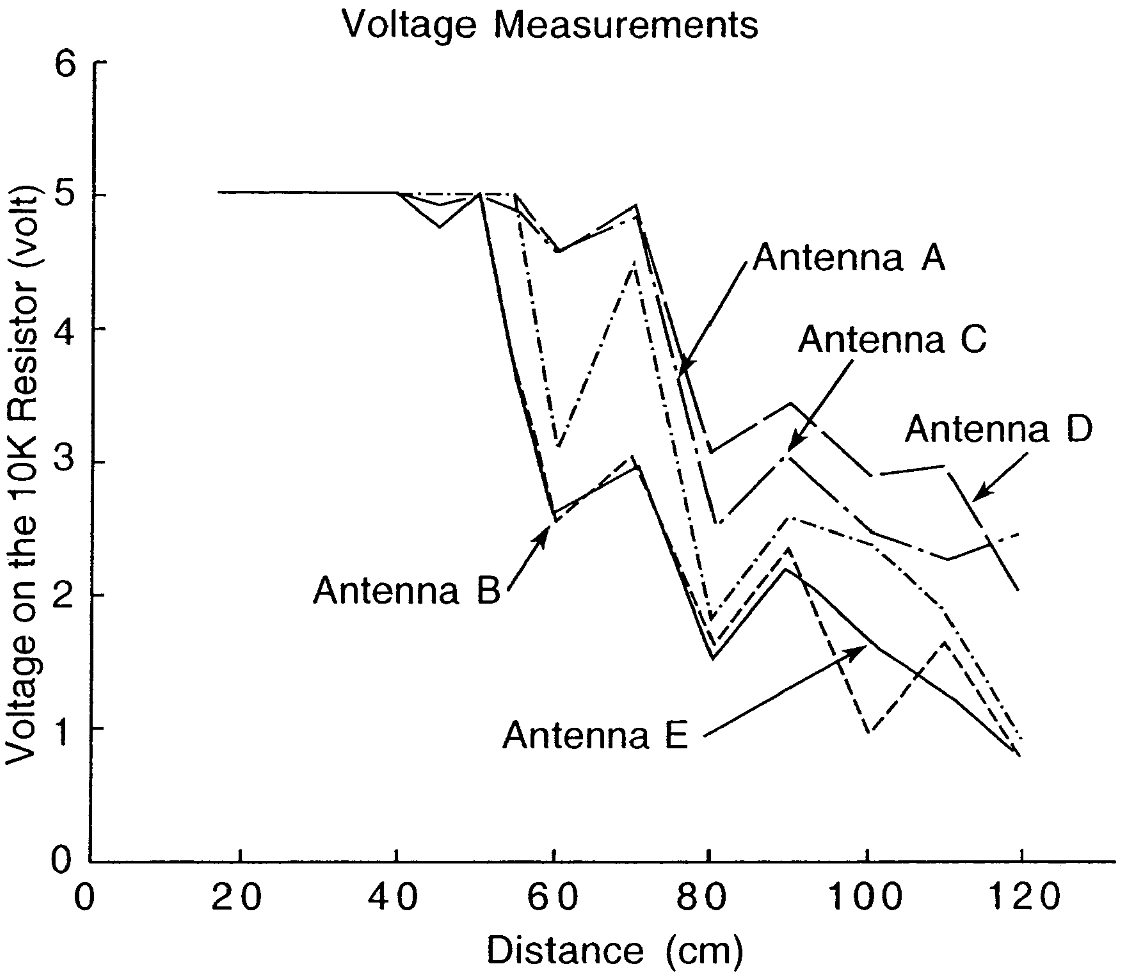
Voltage Measurements
Voltage on the 10K Resistor (volt)
6
5
4
3
2
1
0
0
20
40
60
80
100
120
Distance (cm)
Antenna A
Antenna C
Antenna D
Antenna B
Antenna E

FIG. 10

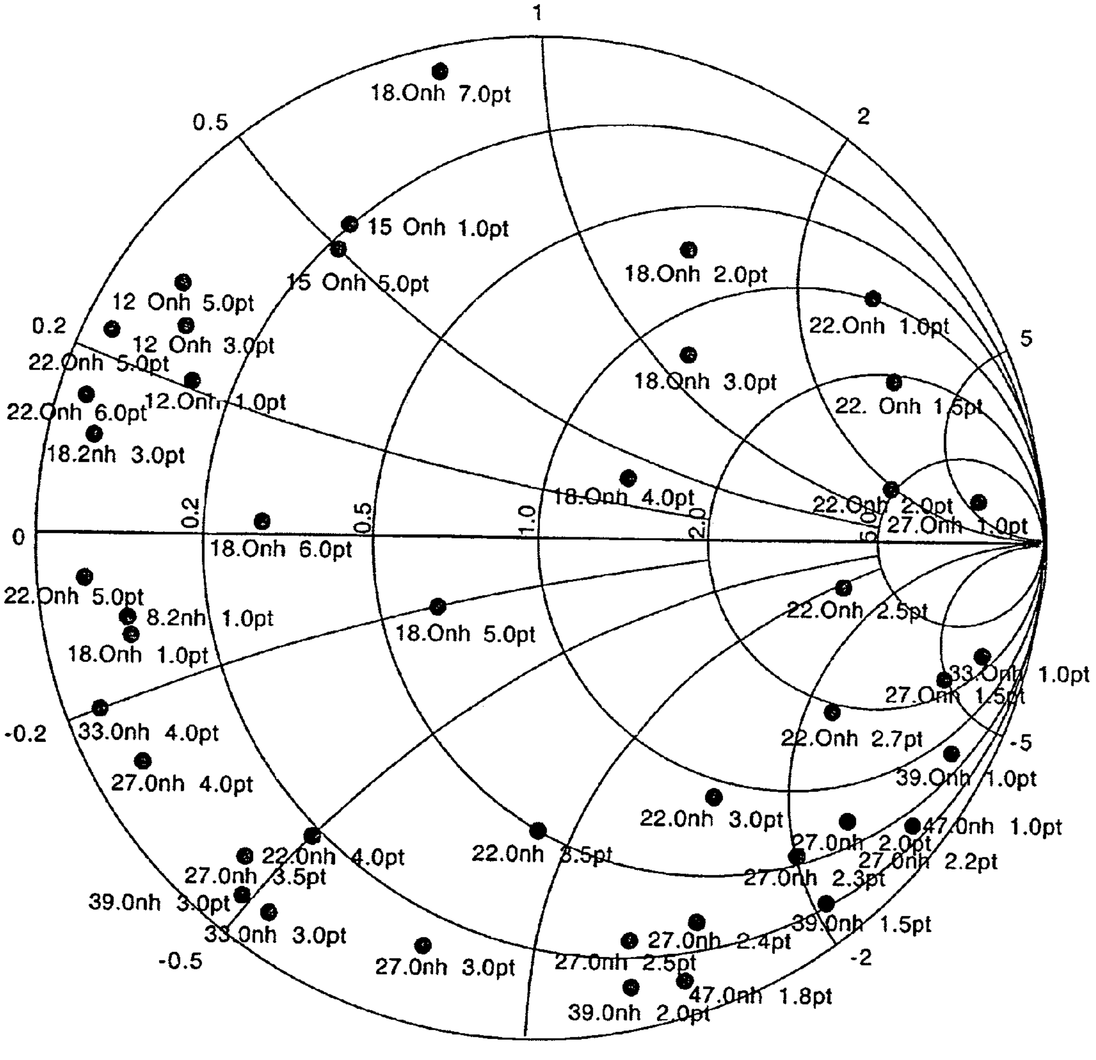
1
0.5
2
0.2
5
0
-0.2
-5
-0.5
-2
-1
0.2
0.5
1.0
2.0
5.0
18.0nh 7.0pt
15 0nh 1.0pt
15 0nh 5.0pt
12 0nh 5.0pt
12 0nh 3.0pt
22.0nh 5.0pt
12.0nh 1.0pt
22.0nh 6.0pt
18.2nh 3.0pt
18.0nh 2.0pt
22.0nh 1.0pt
18.0nh 3.0pt
22. 0nh 1.5pt
18.0nh 4.0pt
22.0nh 2.0pt
27.0nh 1.0pt
18.0nh 6.0pt
22.0nh 5.0pt
8.2nh 1.0pt
18.0nh 1.0pt
18.0nh 5.0pt
22.0nh 2.5pt
33.0nh 1.0pt
27.0nh 1.5pt
33.0nh 4.0pt
22.0nh 2.7pt
27.0nh 4.0pt
39.0nh 1.0pt
22.0nh 3.0pt
47.0nh 1.0pt
27.0nh 2.0pt
27.0nh 2.2pt
22.0nh 4.0pt
22.0nh 3.5pt
27.0nh 2.3pt
27.0nh 3.5pt
39.0nh 3.0pt
39.0nh 1.5pt
33.0nh 3.0pt
27.0nh 2.4pt
27.0nh 3.0pt
27.0nh 2.5pt
47.0nh 1.8pt
39.0nh 2.0pt

FIG. 11

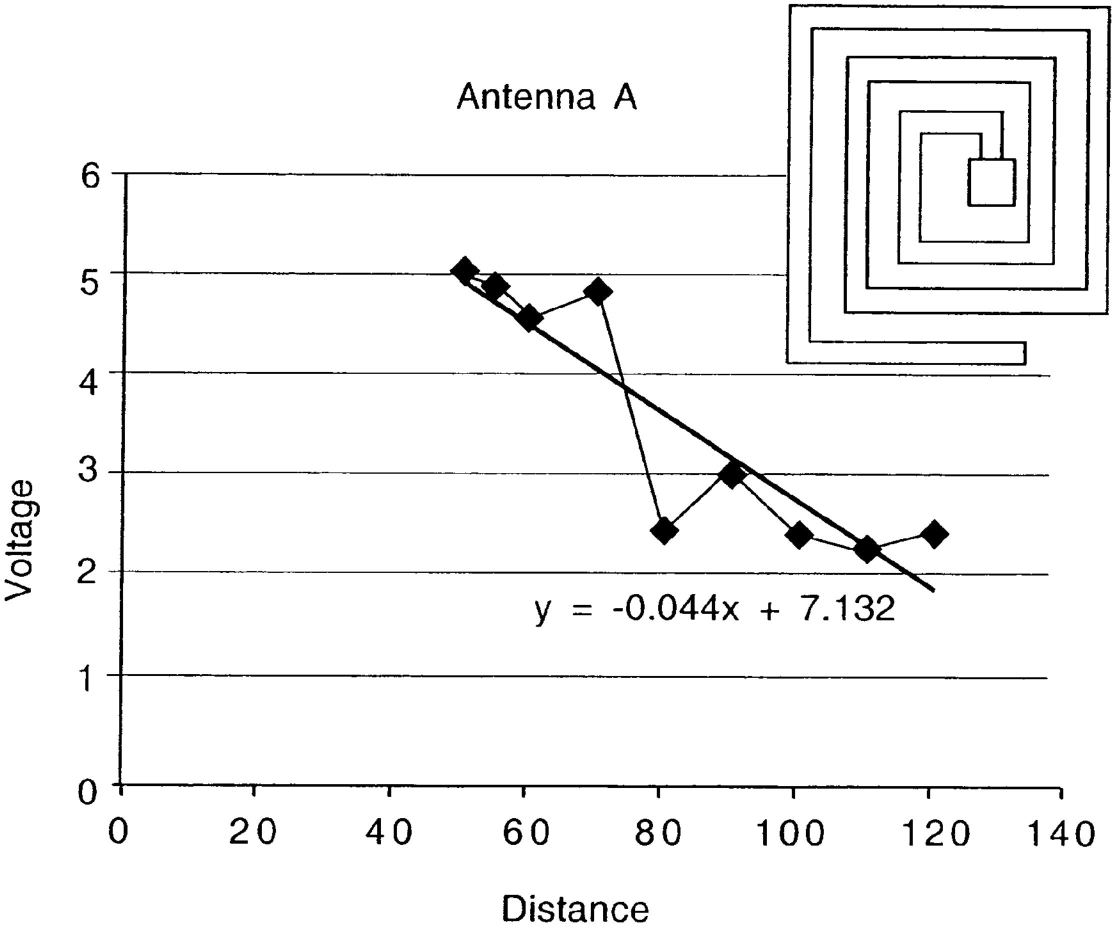
Antenna A
Voltage
6
5
4
3
2
1
0
y = -0.044x + 7.132
0
20
40
60
80
100
120
140
Distance

FIG. 12a

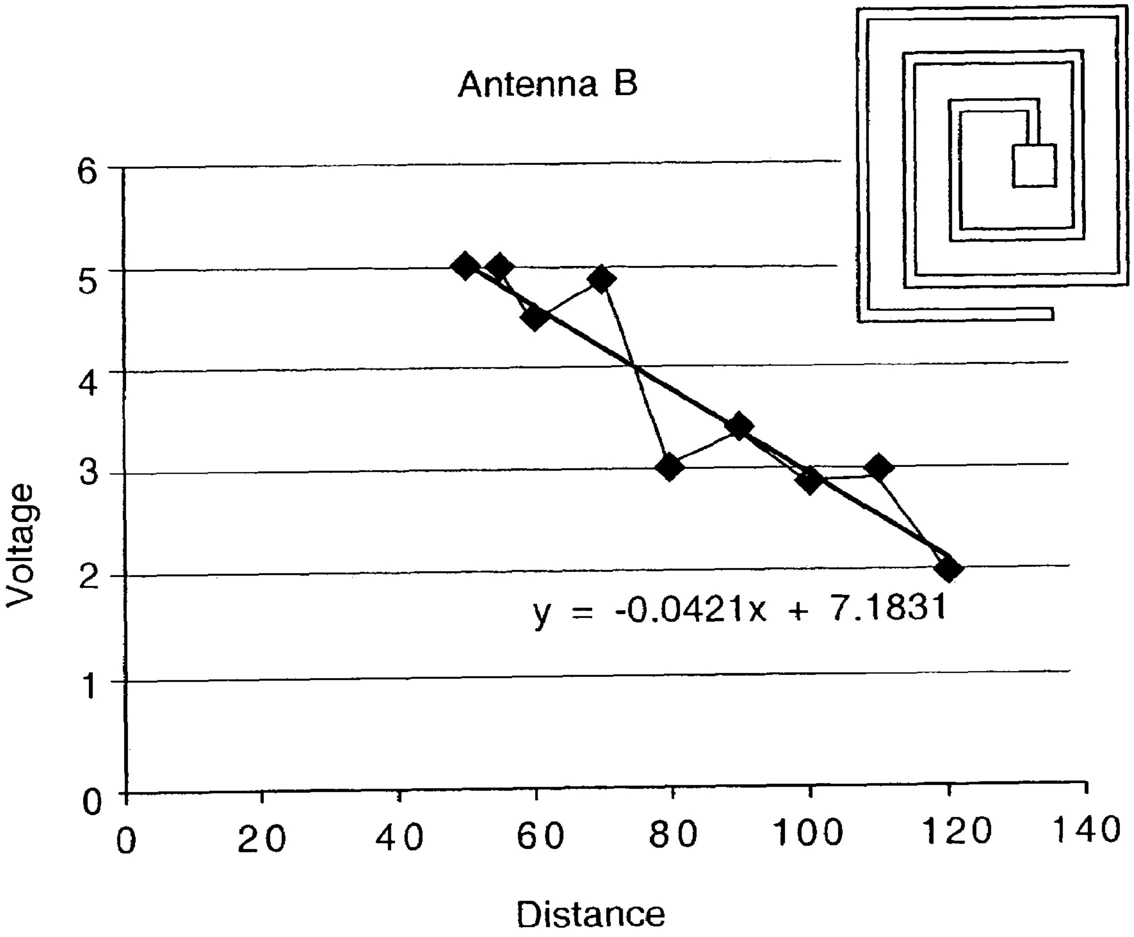
Antenna B
Voltage
6
5
4
3
2
1
0
y = -0.0421x + 7.1831
0
20
40
60
80
100
120
140
Distance

FIG. 12b

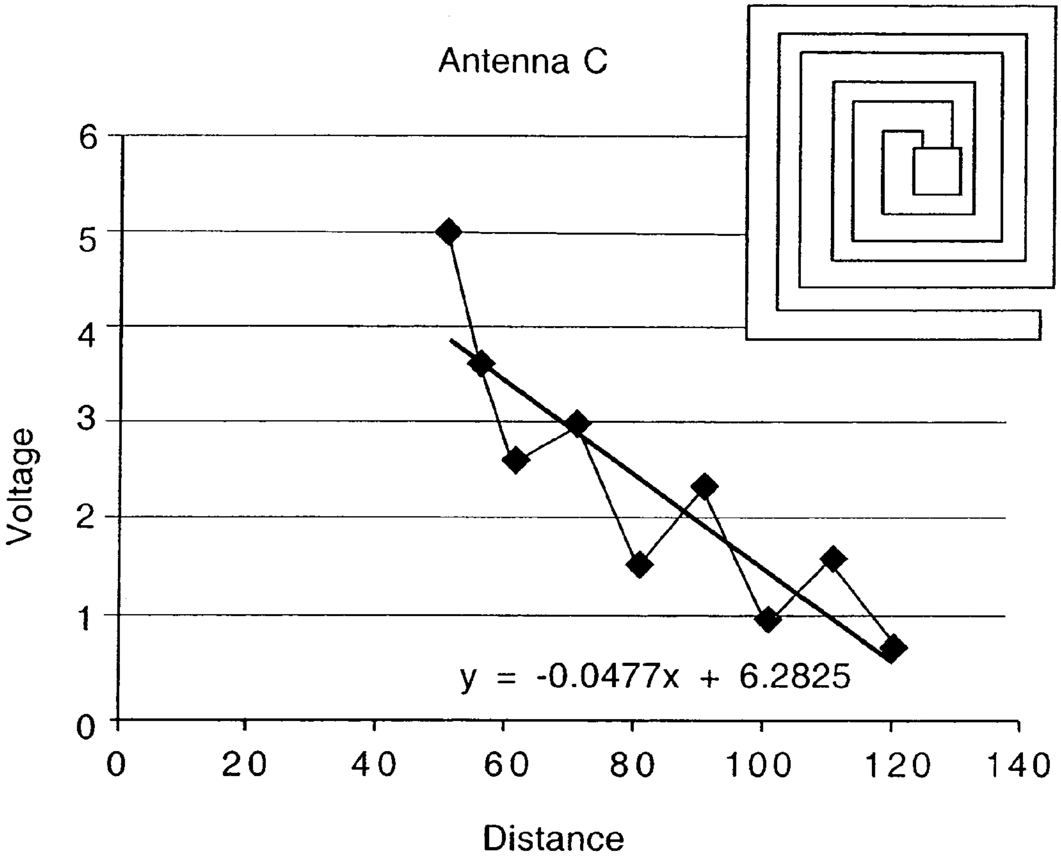
Antenna C
6
5
4
3
2
1
0
Voltage
0
20
40
60
80
100
120
140
Distance
y = -0.0477x + 6.2825

FIG. 12c

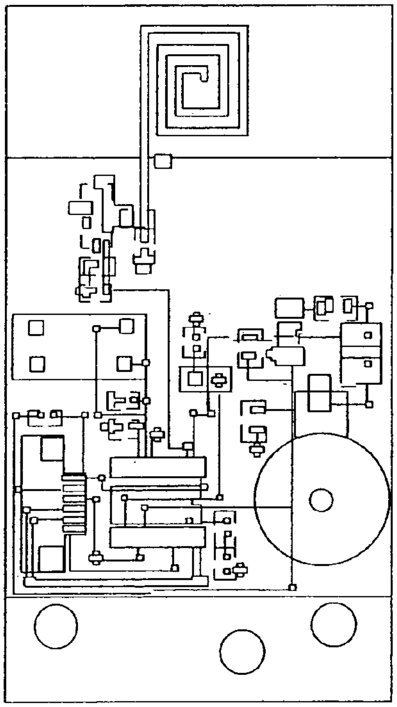

Number of Measurement	Thickness
1	55.1 um
2	223 um
3	234um

FIG. 13a

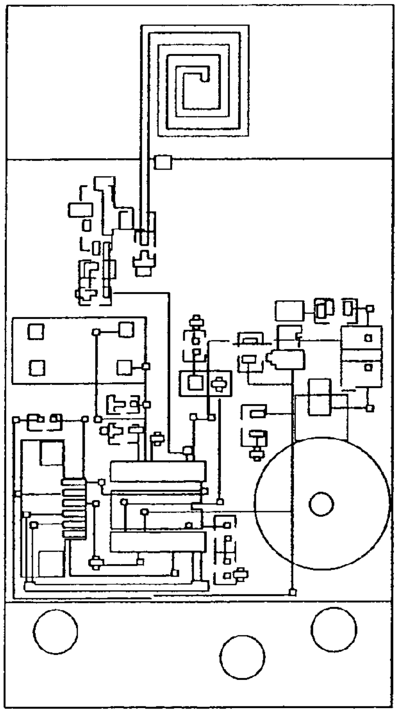

Number of Measurement	Thickness
1	55.1um
2	82.3um
3	273um

FIG. 13b

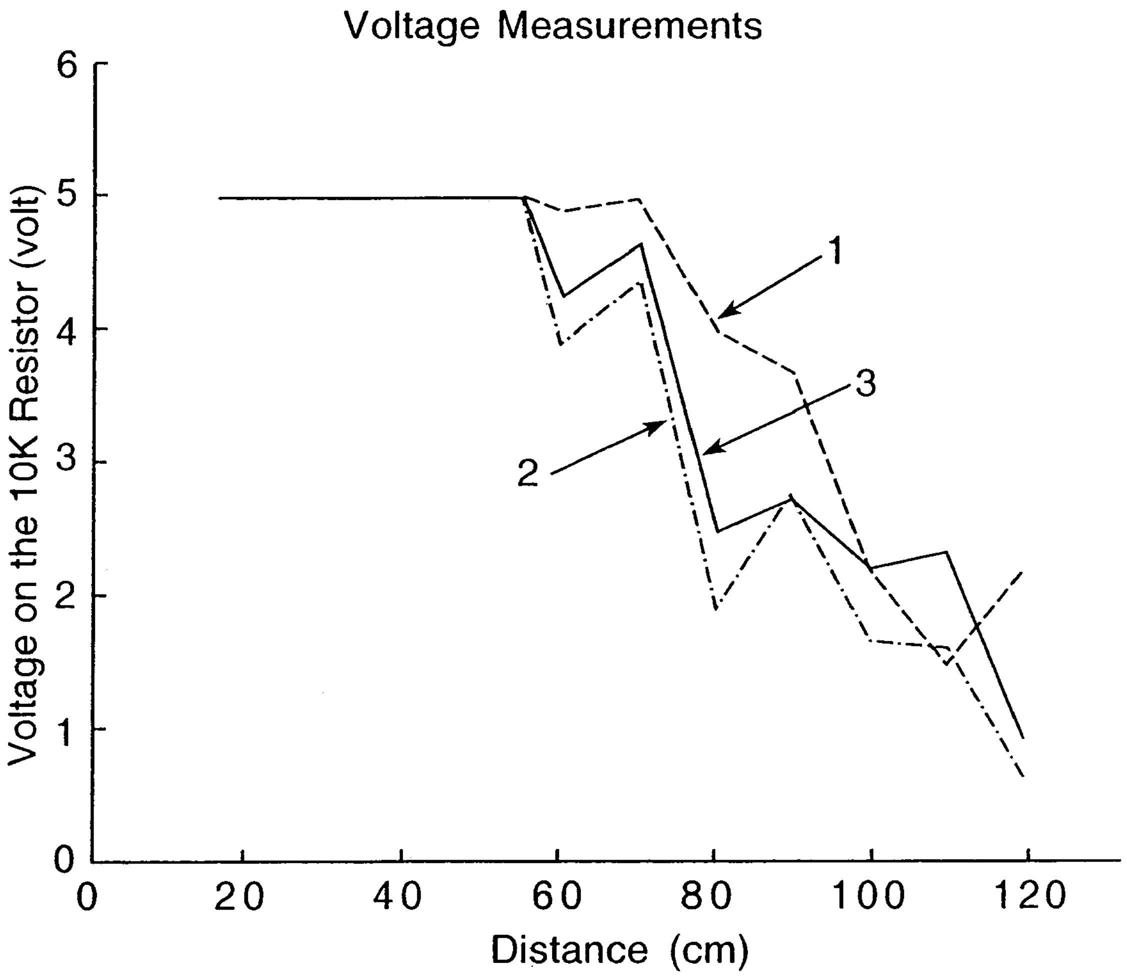
Voltage Measurements
Voltage on the 10K Resistor (volt)
0
1
2
3
4
5
6
0
20
40
60
80
100
120
Distance (cm)
1
2
3

FIG. 14a

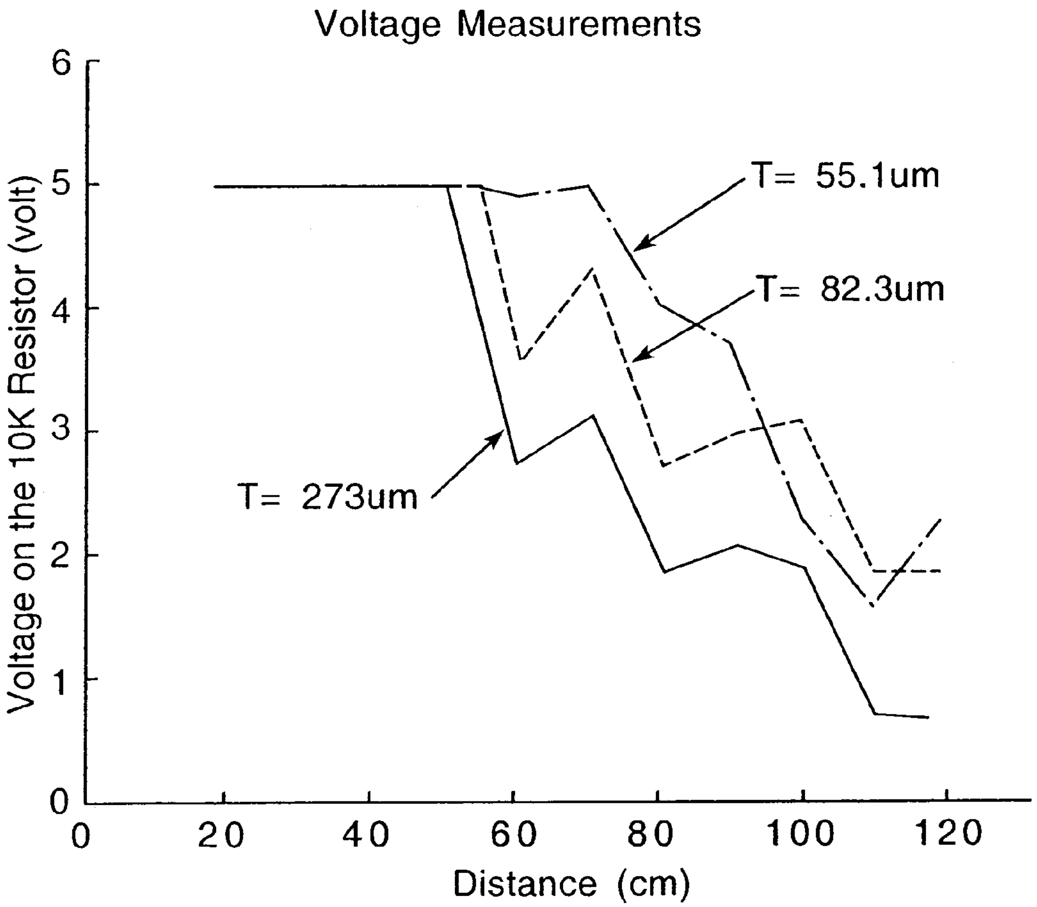
Voltage Measurements
Voltage on the 10K Resistor (volt)
0
1
2
3
4
5
6
0
20
40
60
80
100
120
Distance (cm)
T= 55.1um
T= 82.3um
T= 273um

FIG. 14b

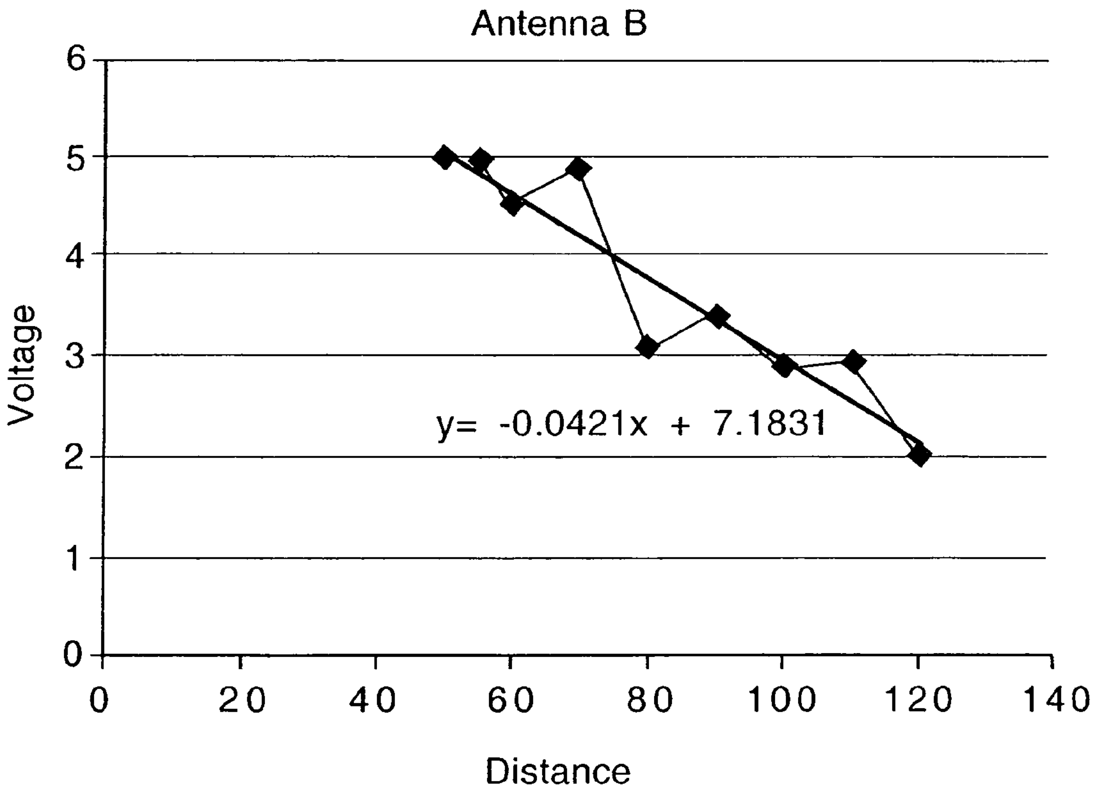
Antenna B
Voltage
0
1
2
3
4
5
6
Distance
0
20
40
60
80
100
120
140
y= -0.0421x + 7.1831

FIG. 15a

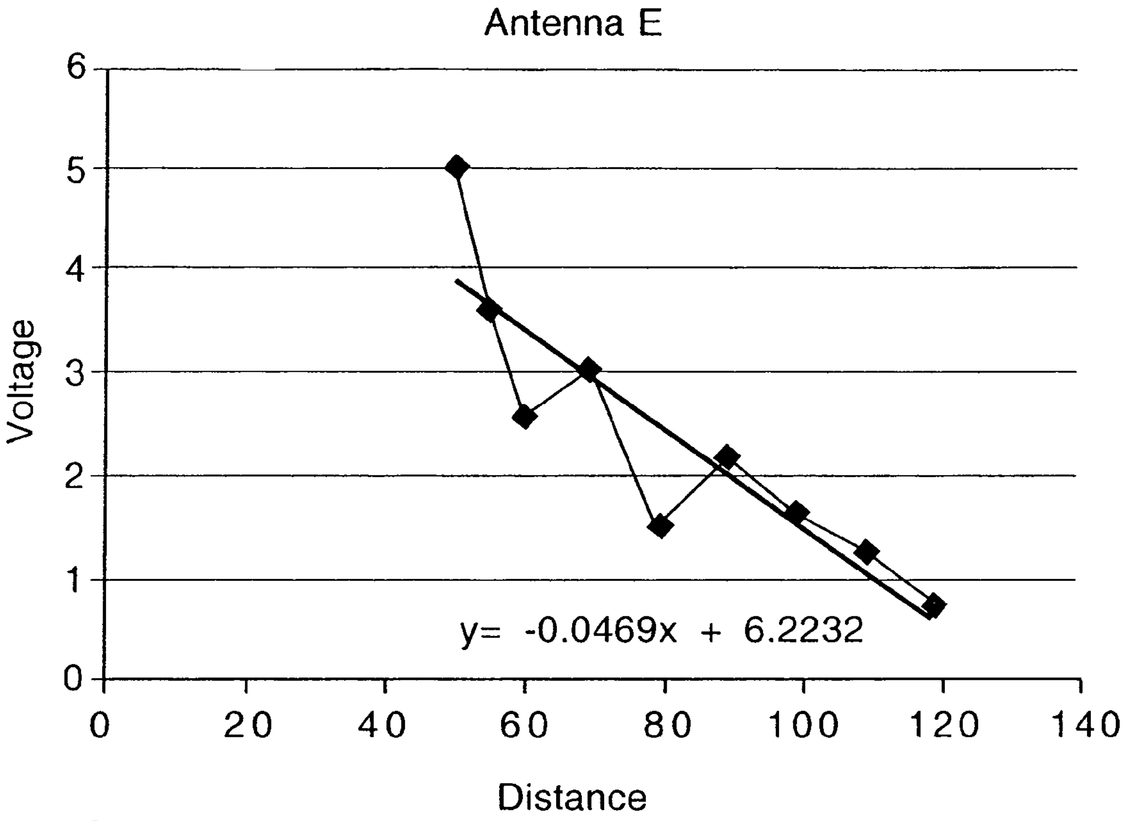

Antenna E
Voltage
6
5
4
3
2
1
0
y= -0.0469x + 6.2232
0
20
40
60
80
100
120
140
Distance

FIG. 15b

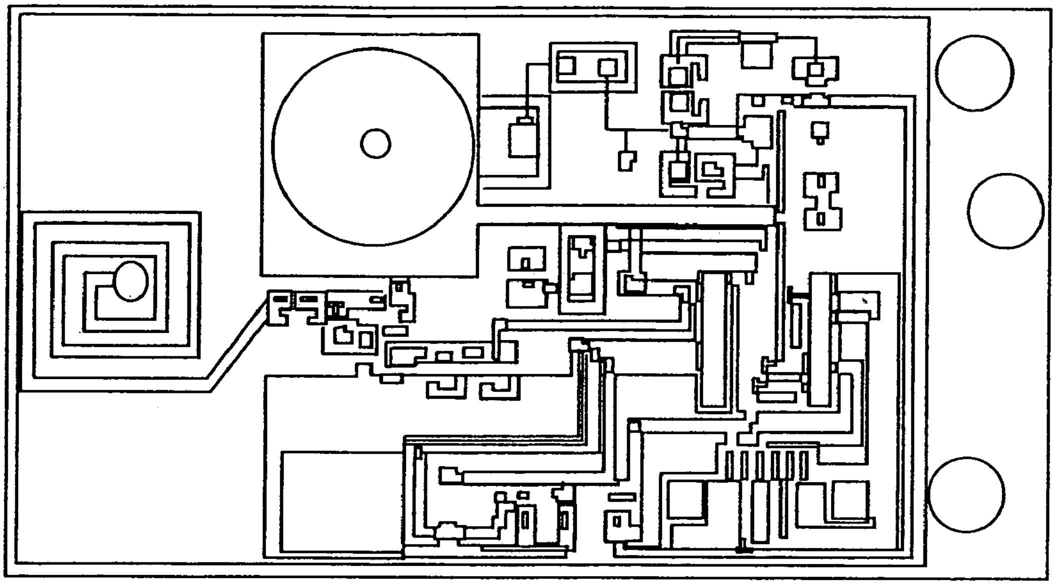

FIG. 16a

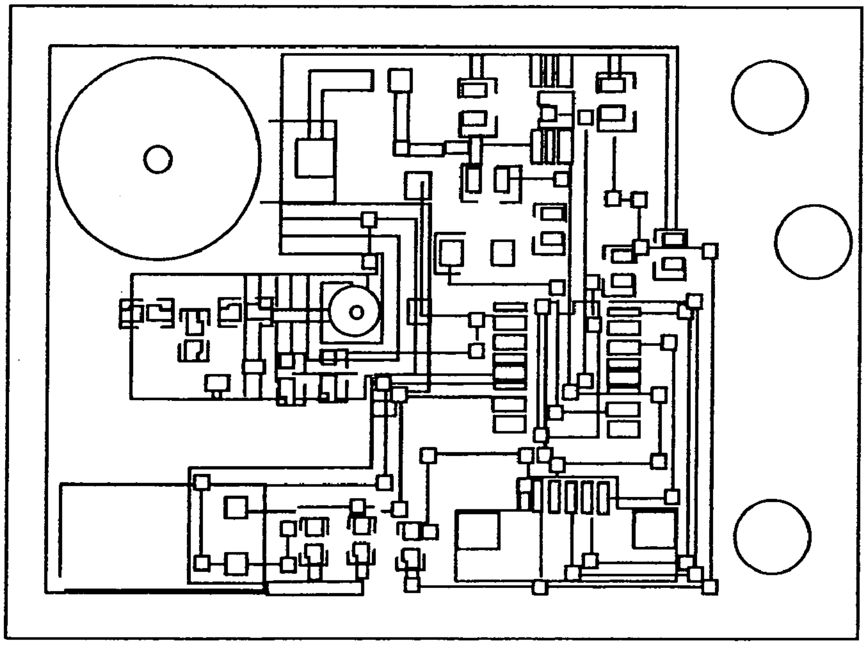

FIG. 16b

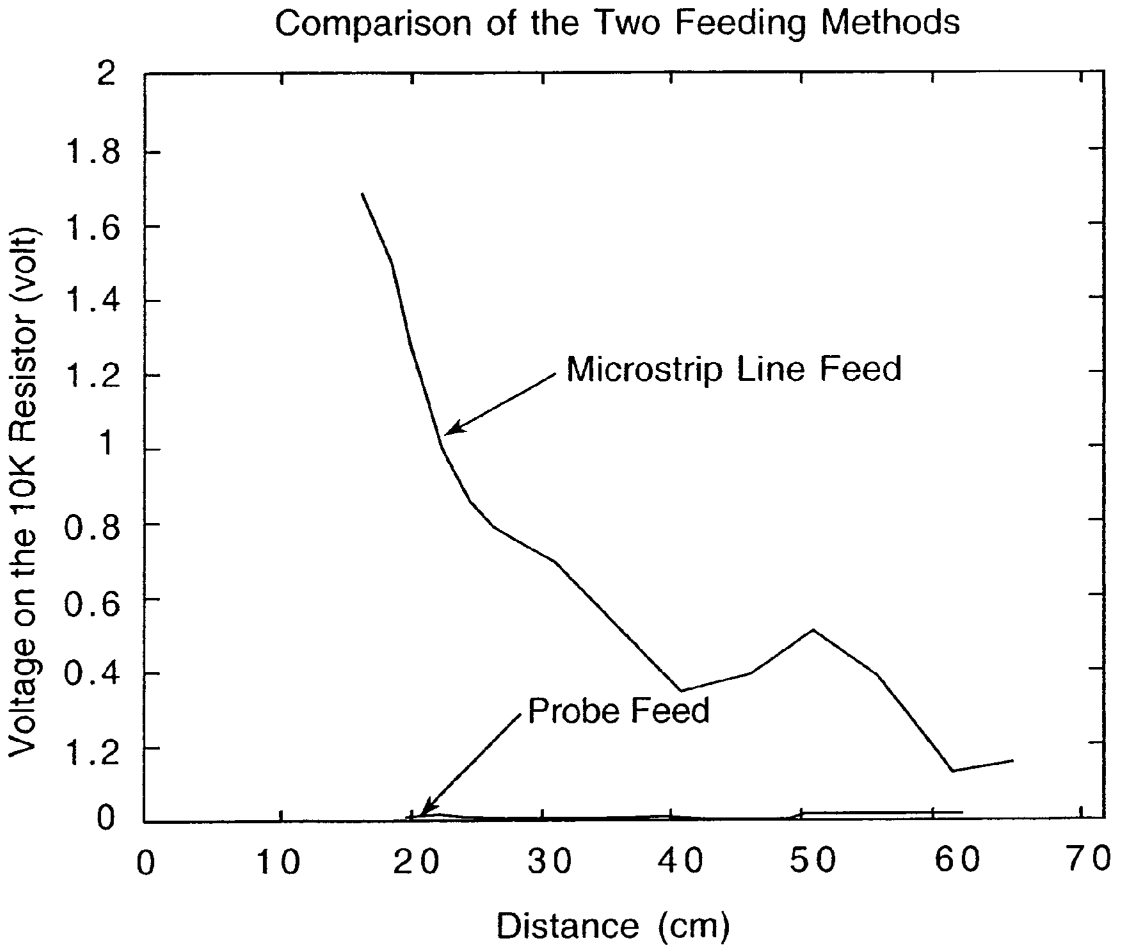
Comparison of the Two Feeding Methods
Voltage on the 10K Resistor (volt)
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
1.2
0
0
10
20
30
40
50
60
70
Distance (cm)
Microstrip Line Feed
Probe Feed

FIG. 17

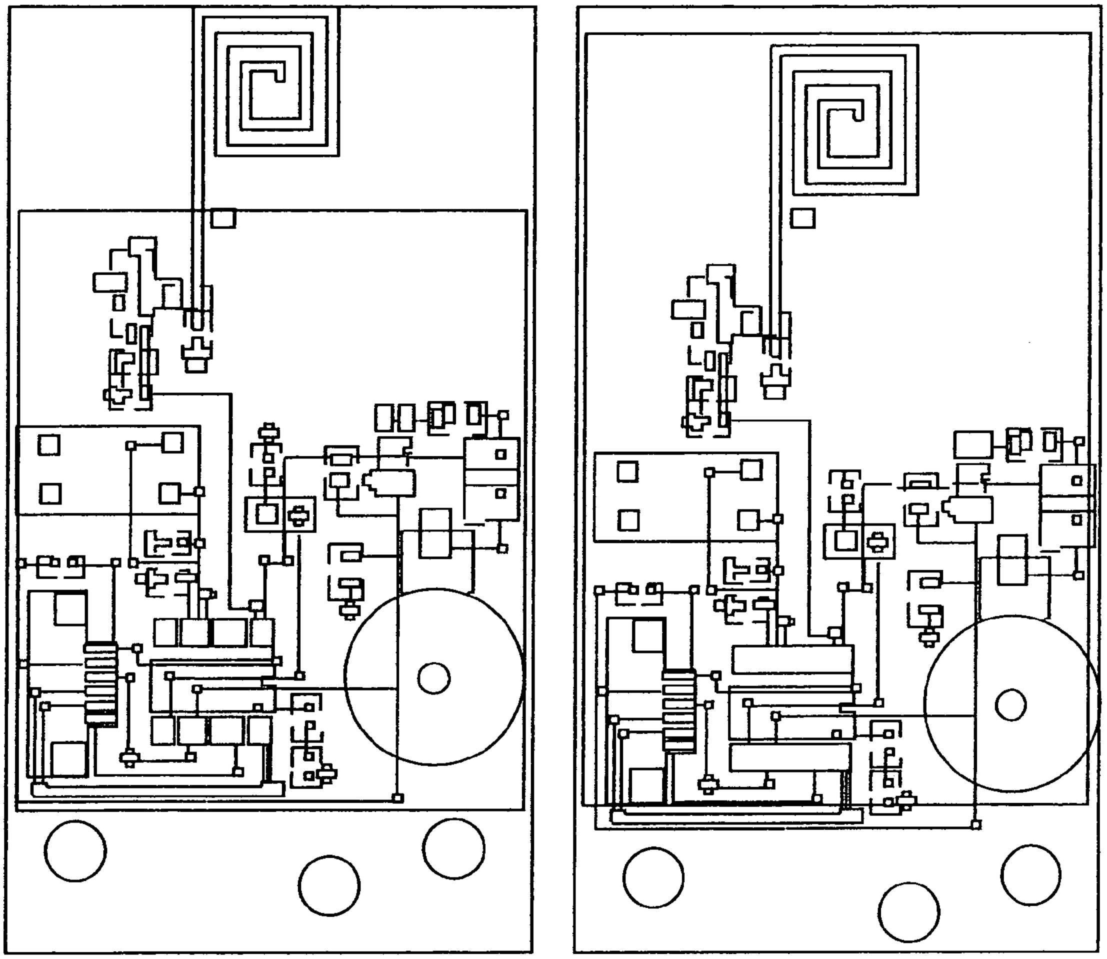

Antenna A
(No Ground Plane)

Antenna B
(With Ground Plane)

FIG. 18a

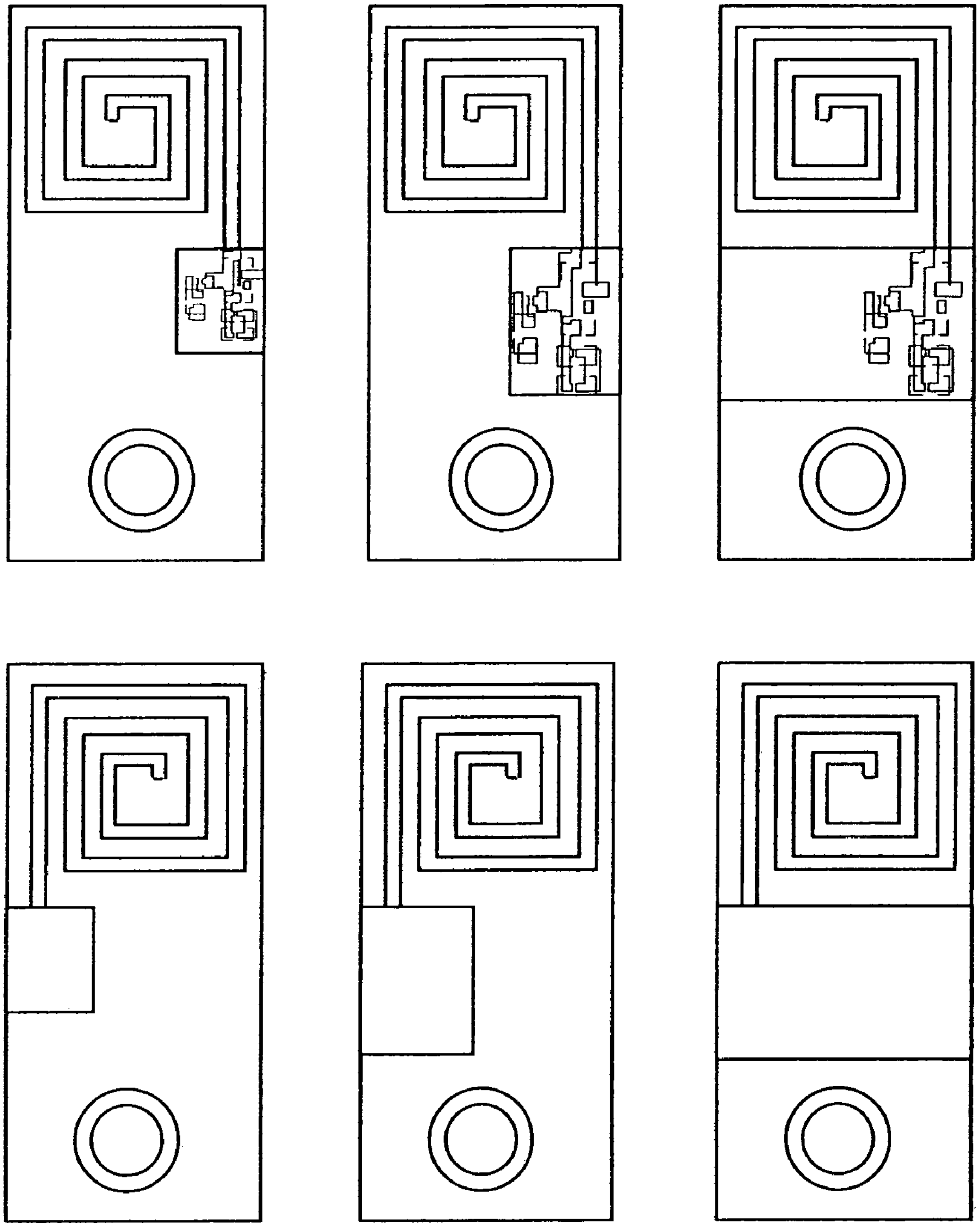

FIG. 18b

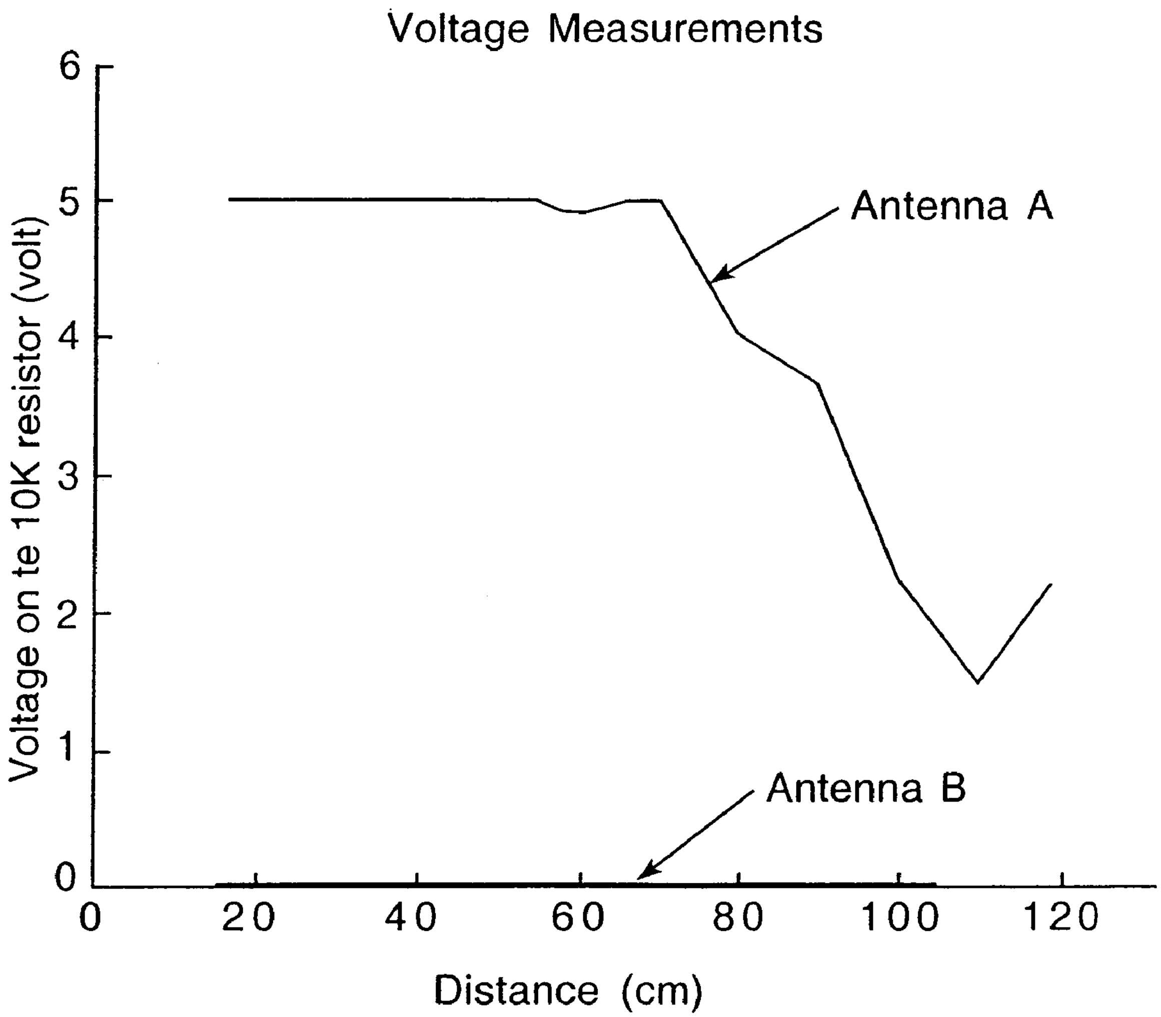
Voltage Measurements
Voltage on te 10K resistor (volt)
6
5
4
3
2
1
0
Antenna A
Antenna B
0
20
40
60
80
100
120
Distance (cm)

FIG. 19

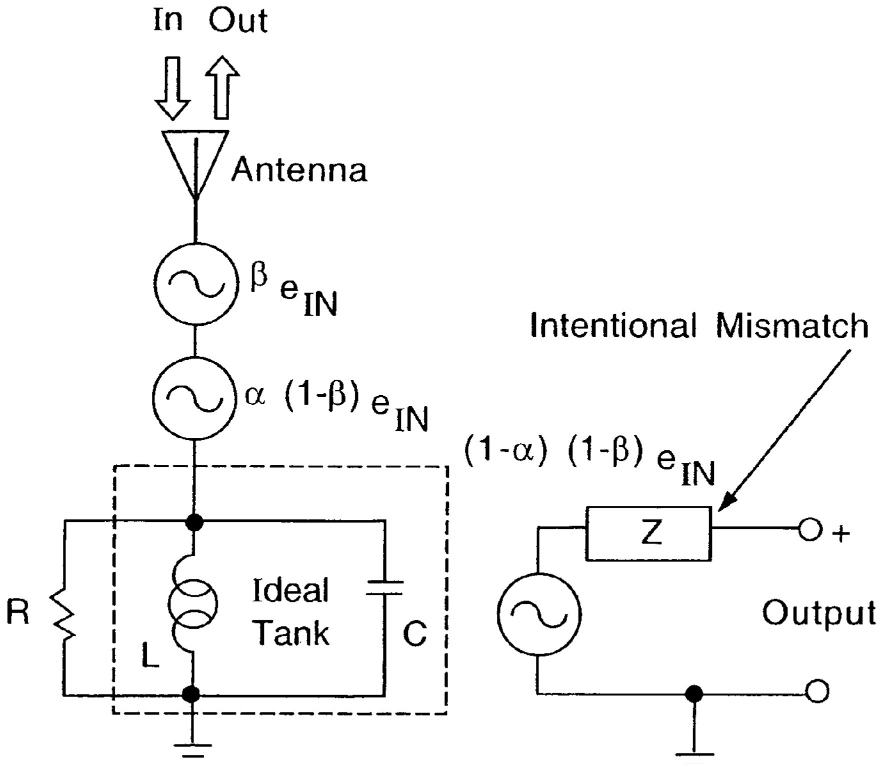
In Out
Antenna
$\beta\ e_{IN}$
$\alpha\ (1-\beta)\ e_{IN}$
Intentional Mismatch
$(1-\alpha)\ (1-\beta)\ e_{IN}$
Z
+
Output
R
L
Ideal
Tank
C

FIG. 20

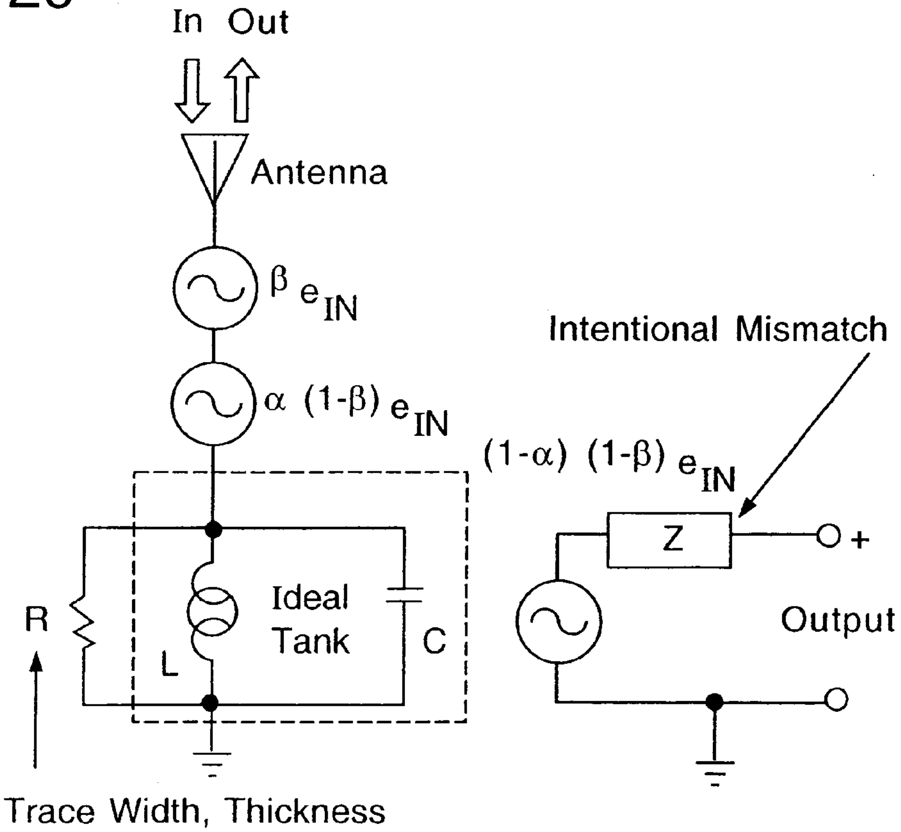
In Out
Antenna
$\beta\ e_{IN}$
$\alpha\ (1-\beta)\ e_{IN}$
Intentional Mismatch
$(1-\alpha)\ (1-\beta)\ e_{IN}$
Z
+
Output
R
L
Ideal
Tank
C
Trace Width, Thickness

FIG. 21

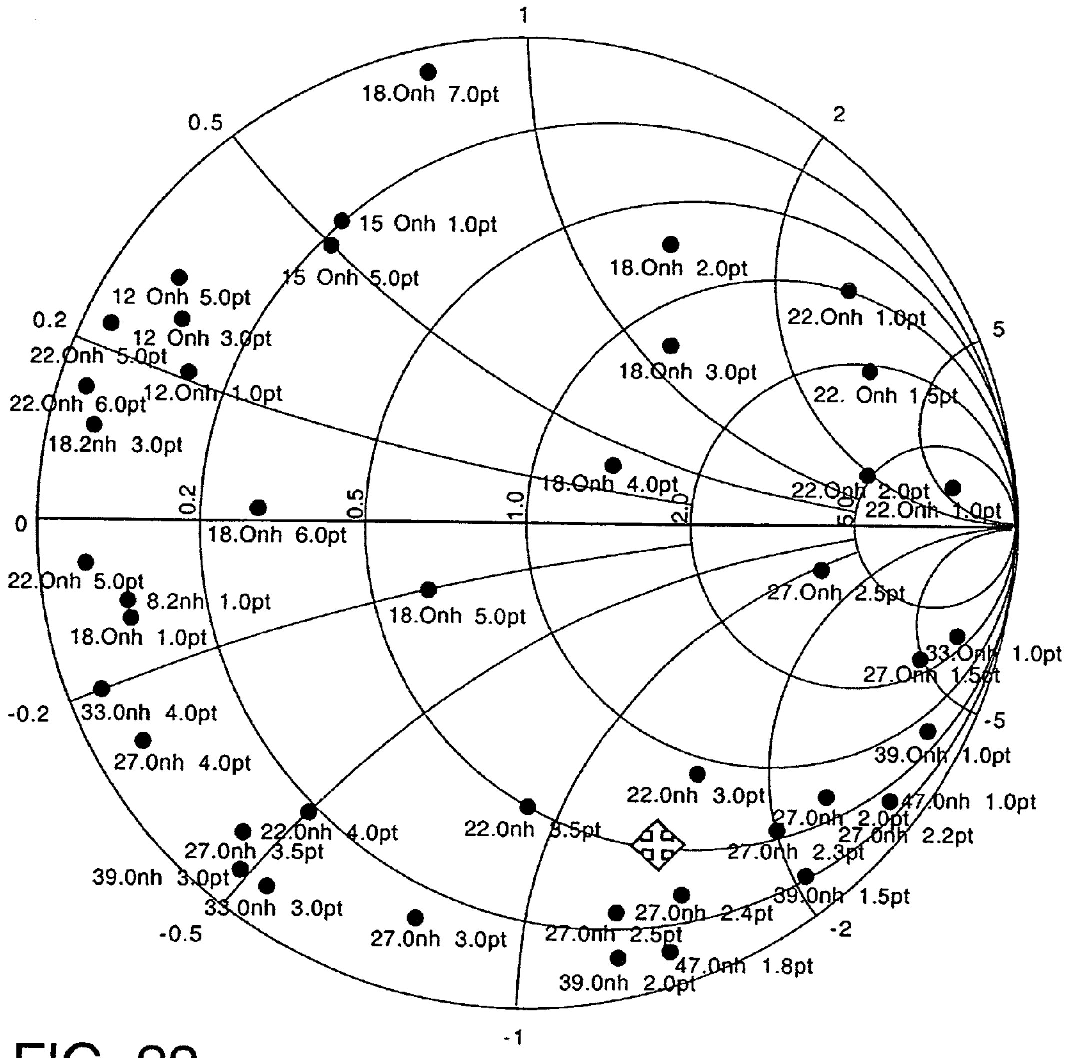
1
18.Onh 7.0pt
0.5
2
15 Onh 1.0pt
15 Onh 5.0pt
18.Onh 2.0pt
12 Onh 5.0pt
22.Onh 1.0pt
0.2
12 Onh 3.0pt
5
22.Onh 5.0pt
18.Onh 3.0pt
12.Onh 1.0pt
22. Onh 1.5pt
22.Onh 6.0pt
18.2nh 3.0pt
18.Onh 4.0pt
22.Onh 2.0pt
22.Onh 1.0pt
0
0.2
0.5
1.0
2.0
5.0
18.Onh 6.0pt
22.Onh 5.0pt
8.2nh 1.0pt
18.Onh 5.0pt
27.Onh 2.5pt
18.Onh 1.0pt
33.Onh 1.0pt
27.Onh 1.5pt
-0.2
33.0nh 4.0pt
-5
27.0nh 4.0pt
39.Onh 1.0pt
22.0nh 3.0pt
47.0nh 1.0pt
27.0nh 2.0pt
22.0nh 4.0pt
22.0nh 3.5pt
27.0nh 2.2pt
27.0nh 3.5pt
27.0nh 2.3pt
39.0nh 3.0pt
39.0nh 1.5pt
33.0nh 3.0pt
27.0nh 2.4pt
-0.5
27.0nh 3.0pt
27.0nh 2.5pt
-2
47.0nh 1.8pt
39.0nh 2.0pt
-1

FIG. 22

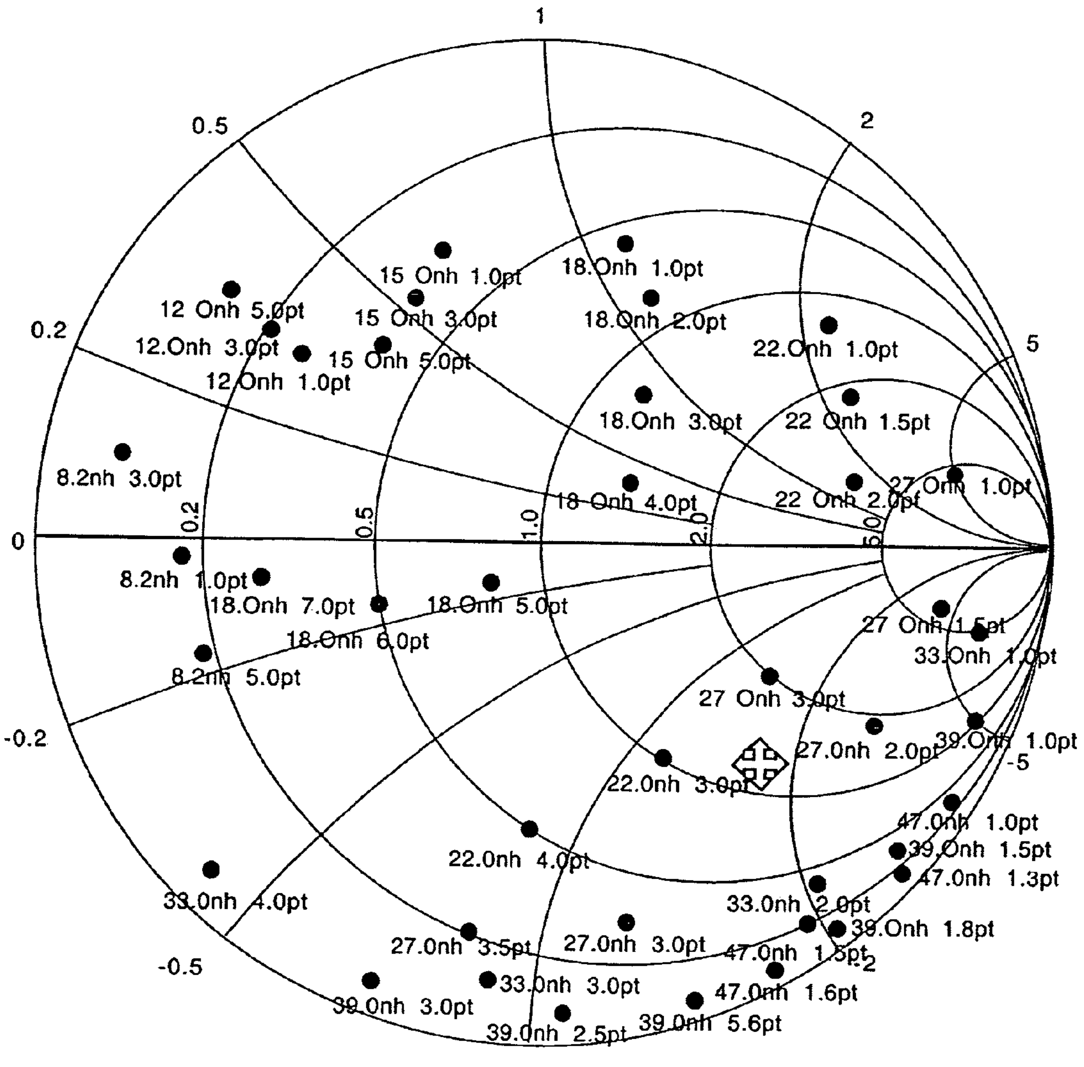
1
0.5
2
0.2
5
0
0.2
0.5
1.0
2.0
5.0
-0.2
-5
-0.5
-2
-1
15 0nh 1.0pt
18.0nh 1.0pt
12 0nh 5.0pt
15 0nh 3.0pt
18.0nh 2.0pt
22.0nh 1.0pt
12.0nh 3.0pt
15 0nh 5.0pt
12.0nh 1.0pt
18.0nh 3.0pt
22 0nh 1.5pt
8.2nh 3.0pt
27.0nh 1.0pt
18.0nh 4.0pt
22 0nh 2.0pt
8.2nh 1.0pt
18.0nh 7.0pt
18.0nh 5.0pt
18.0nh 6.0pt
27 0nh 1.5pt
33.0nh 1.0pt
8.2nh 5.0pt
27 0nh 3.0pt
27.0nh 2.0pt
39.0nh 1.0pt
22.0nh 3.0pt
22.0nh 4.0pt
47.0nh 1.0pt
39.0nh 1.5pt
47.0nh 1.3pt
33.0nh 4.0pt
33.0nh 2.0pt
39.0nh 1.8pt
27.0nh 3.5pt
27.0nh 3.0pt
47.0nh 1.5pt
33.0nh 3.0pt
47.0nh 1.6pt
39.0nh 3.0pt
39.0nh 2.5pt
39.0nh 5.6pt

FIG. 23

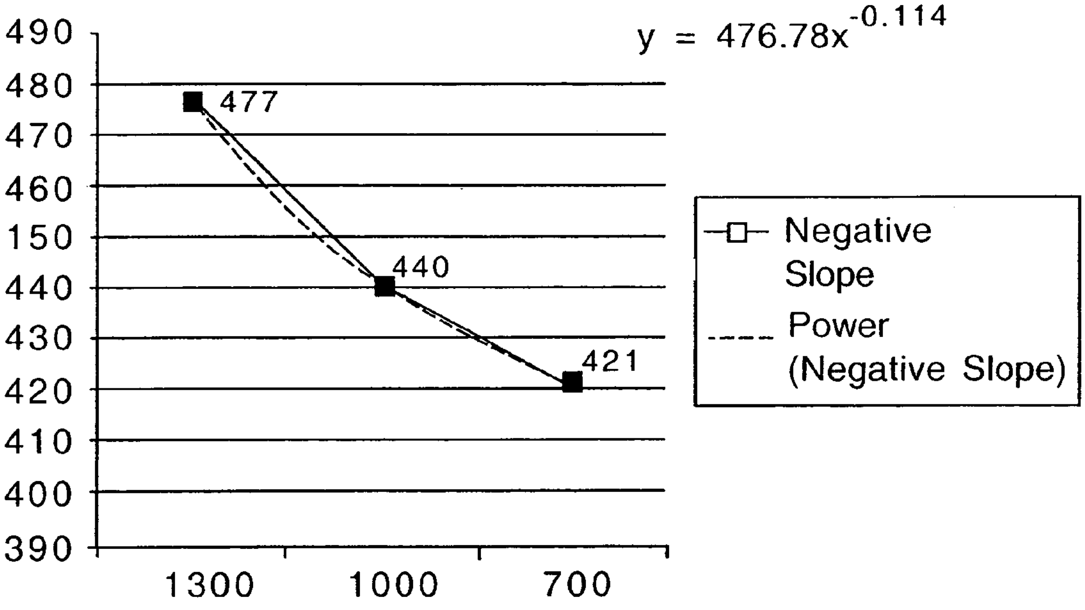

FIG. 24a

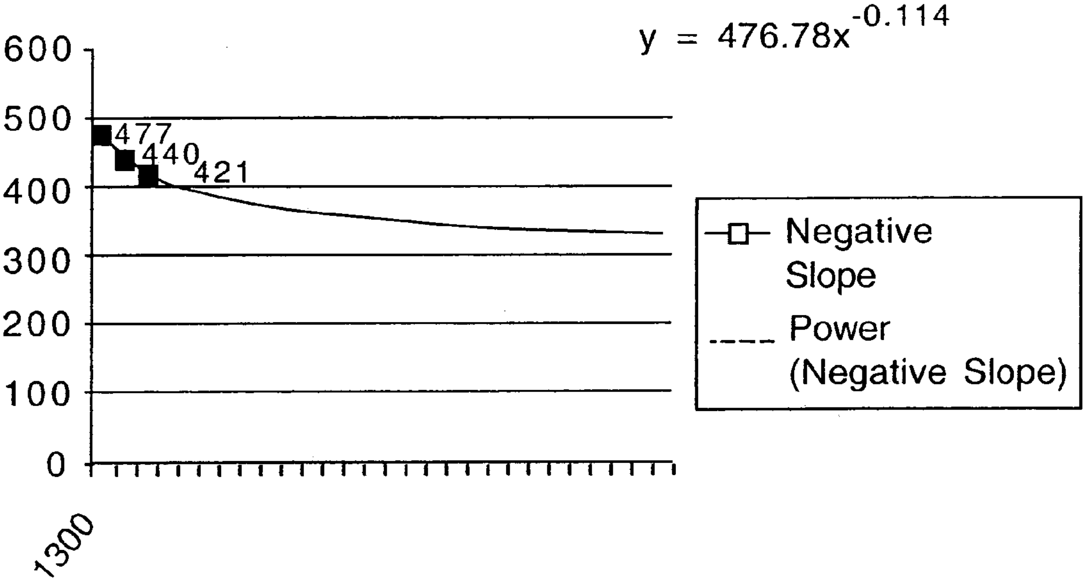

FIG. 24b

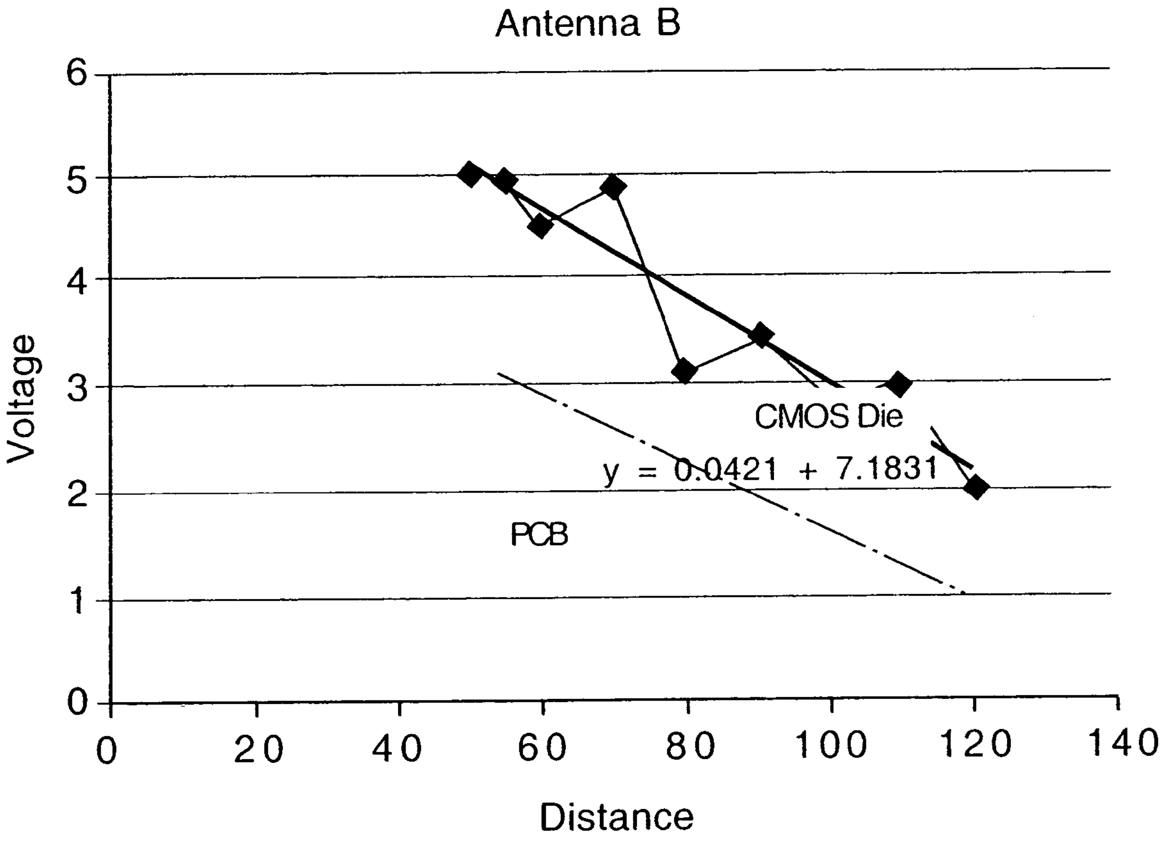
Antenna B
6
5
4
3
2
1
0
Voltage
CMOS Die
y = 0.0421 + 7.1831
PCB
0
20
40
60
80
100
120
140
Distance

FIG. 25

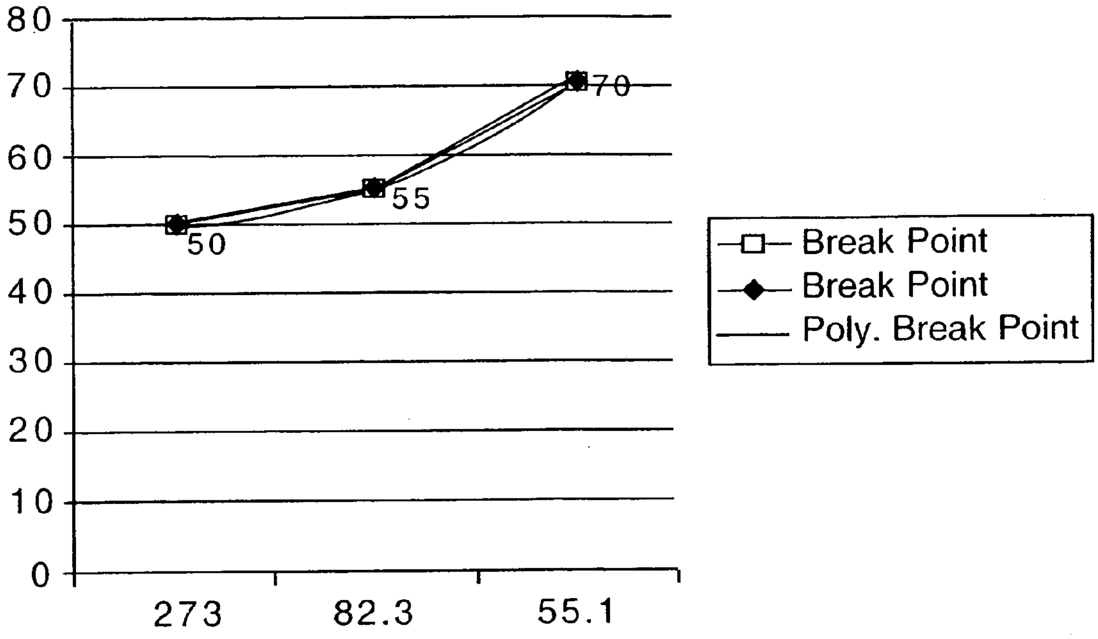

FIG. 26a

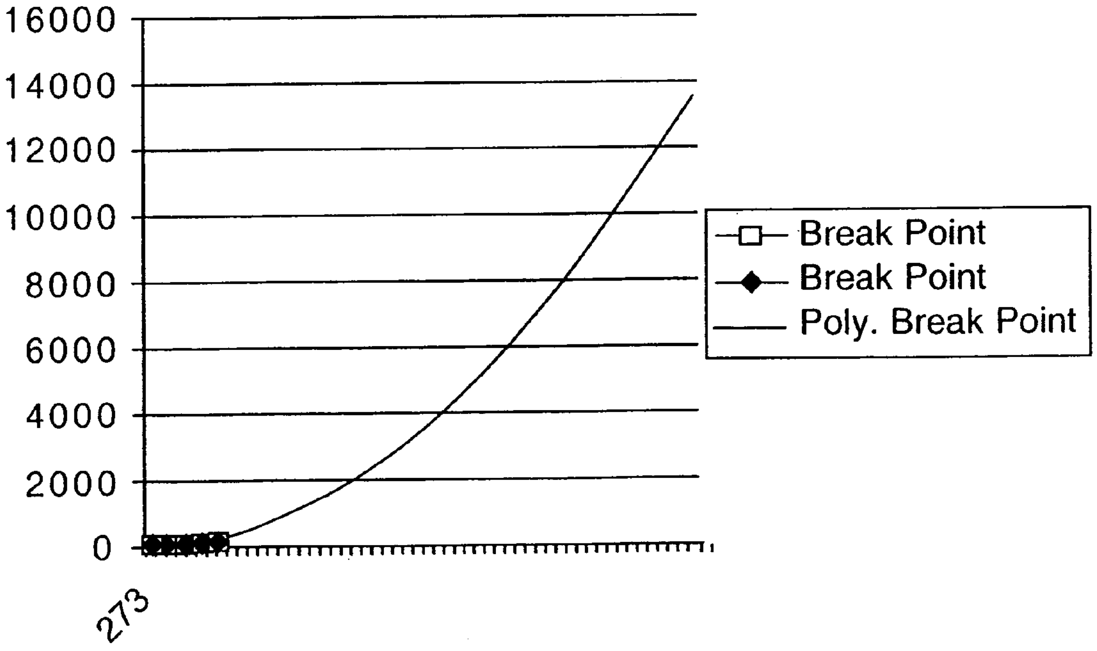

FIG. 26b

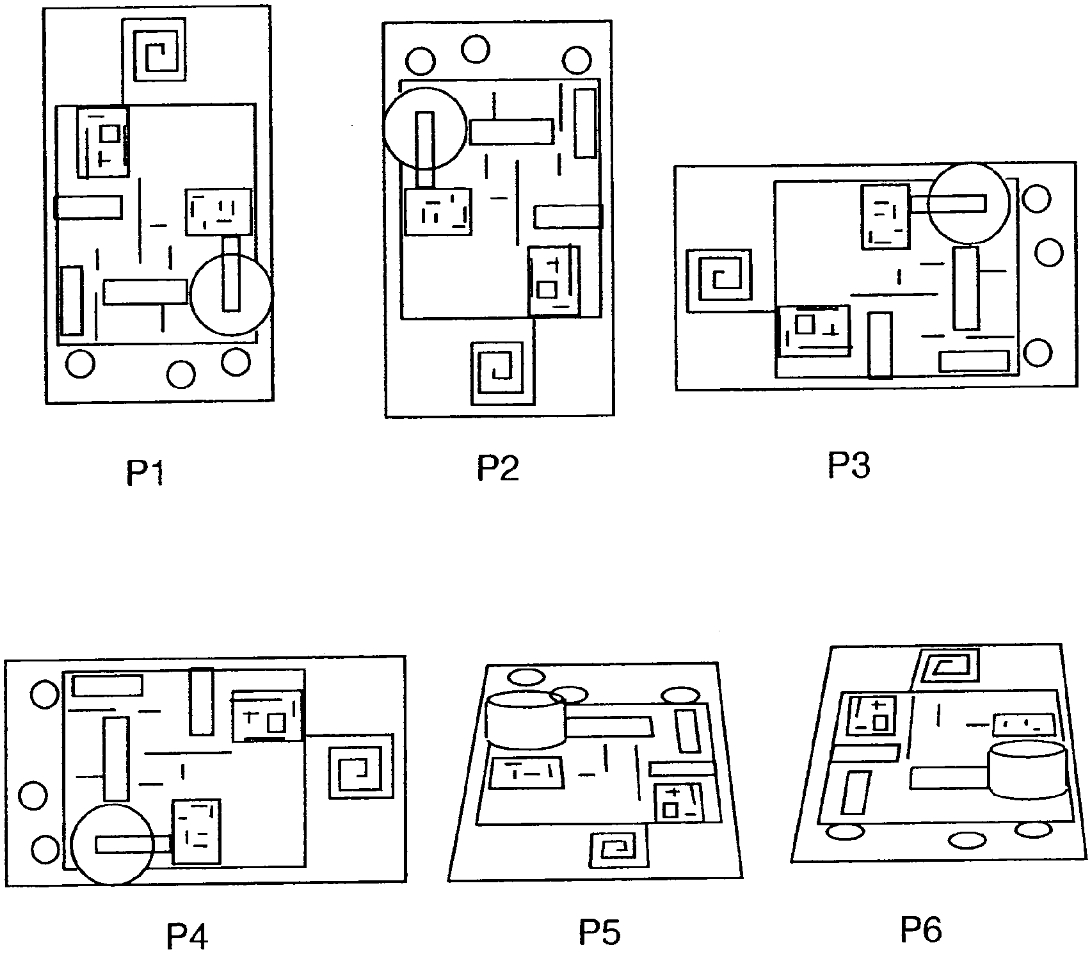
P1 P2 P3

P4 P5 P6

FIG. 27

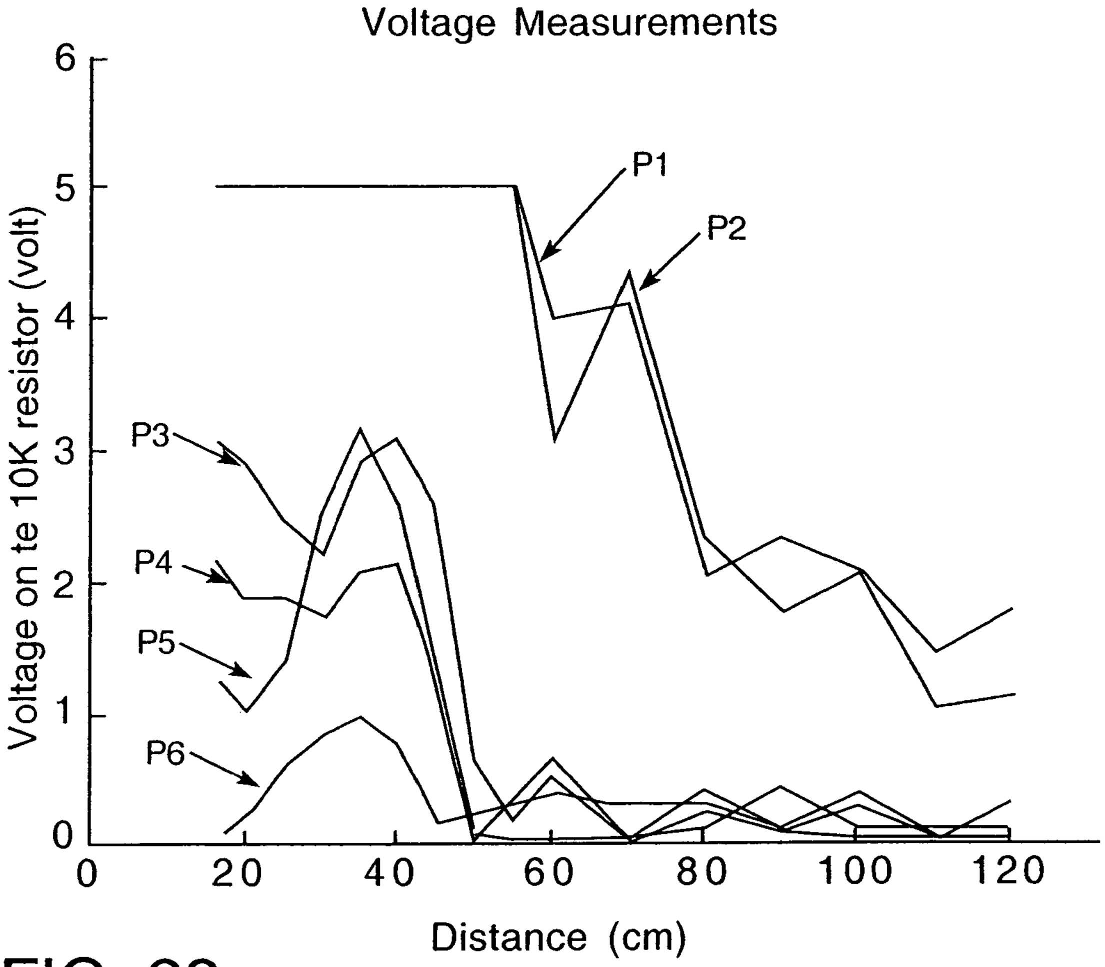
Voltage Measurements
Voltage on te 10K resistor (volt)
Distance (cm)
0
1
2
3
4
5
6
0
20
40
60
80
100
120
P1
P2
P3
P4
P5
P6

FIG. 28

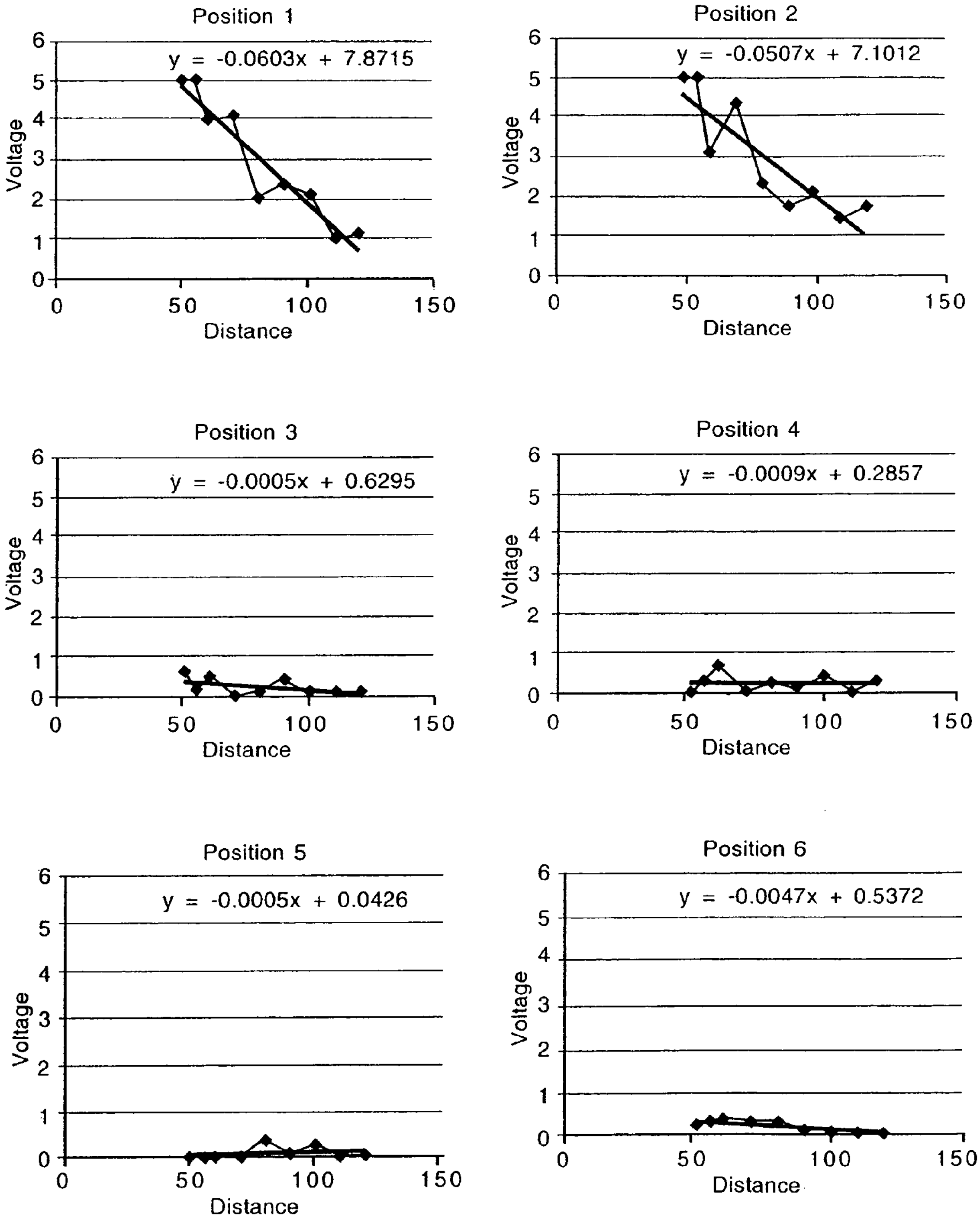
Position 1
y = -0.0603x + 7.8715
Voltage
Distance
0
1
2
3
4
5
6
0
50
100
150
Position 2
y = -0.0507x + 7.1012
Voltage
Distance
Position 3
y = -0.0005x + 0.6295
Voltage
Distance
Position 4
y = -0.0009x + 0.2857
Voltage
Distance
Position 5
y = -0.0005x + 0.0426
Voltage
Distance
Position 6
y = -0.0047x + 0.5372
Voltage
Distance

FIG. 29

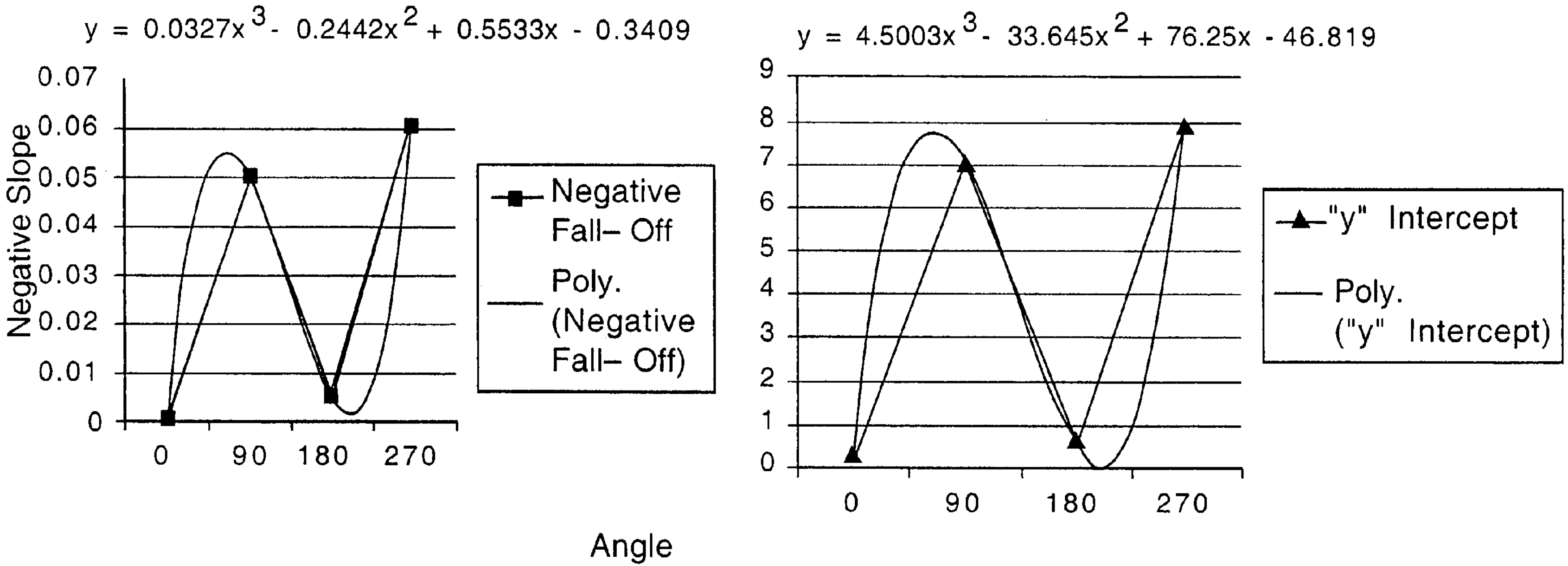
y = 0.0327x 3 - 0.2442x 2 + 0.5533x - 0.3409
Negative Slope
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
90
180
270
Negative Fall- Off
Poly. (Negative Fall- Off)
y = 4.5003x 3 - 33.645x 2 + 76.25x - 46.819
9
8
7
6
5
4
3
2
1
0
0
90
180
270
"y" Intercept
Poly. ("y" Intercept)
Angle

FIG. 30

ANTENNA ON A WIRELESS UNTETHERED DEVICE SUCH AS A CHIP OR PRINTED CIRCUIT BOARD FOR HARVESTING ENERGY FROM SPACE

CROSS-REFERENCE

This application claims the benefit of U.S. Provisional Application Ser. No. 60/475,151 entitled "ANTENNA ON A WIRELESS UNTETHERED DEVICE SUCH AS A CHIP OR PRINTED CIRCUIT BOARD FOR HARVESTING ENERGY FROM SPACE" filed Jun. 2, 2003.

BACKGROUND OF THE INVENTION

1. Field of the Invention

A device is disclosed comprising a semiconductor integrated circuit (IC) chip having disposed on it antenna and the circuitry for the IC chip, wherein antenna and the circuitry of the IC chip are integrated so as to provide the IC chip with an improved antenna effective area. The design of the antenna placement on the chip is independent of the location of the antenna on the chip. As a result, antenna-on-a-chips of this invention are simpler to design and manufacture than the designs disclosed in the prior art. With increasing dependence on local and wide area wireless networks, particularly those with low power (range) requirements and local sensing or object identification, there is a need to have an antenna structure integrated onto single semiconductor devices. One important lack has been an on-chip antenna structure that overcomes simple deficiencies of size.

The length and material of antennas normally determine the frequency and intensity of signals that may be received or sent from the antenna. However with smaller and smaller local area wireless networks being contemplated, the concept of a room sized network area or building area with antennas mounted in walls and ceilings (whether independent separate antennas or multi-use antennas, such as using electrical wiring or telephone wiring as an antenna structure), the feasibility of using very low power antenna structures to transfer information from a local network to a wireless IC device or system containing such wireless IC device becomes practicable.

2. Description of the Prior Art

It has been known to provide antennae and circuitry at specific locations on chips. U.S. Pat. Nos. 4,857,893 and 4,724,427 disclose the use of planar antennas that are included in circuitry of a transponder on a chip. As disclosed therein, the antenna coil is etched around the periphery of the chip substrate. In the center of the antenna coil are found a custom logic circuit, a programmable memory array, and memory control logic. The planar antenna of the transponder was said to employ a magnetic film inductor on the chip in order to allow for a reduction in the number of turns and thereby simplify fabrication of the inductors. It disclosed an antenna having a multi-turned spiral coil and having a 1 cm×1 cm outer diameter. When a high frequency current was passed in the coil, the magnetic films were said to be driven in a hard direction, and the two magnetic films around each conductor served as a magnetic core enclosing a one turn coil. The magnetic films were said to increase the inductance of the coil, in addition to its free-space inductance.

One of the problems with the approach as taught in U.S. Pat. Nos. 4,857,893 and 4,724,427 is the need to fabricate small, air core inductors of sufficiently high inductance and Q for integrated circuit applications. The small air core inductors were said to be made by depositing a permalloy magnetic film or other suitable material having a large magnetic permeability and electric insulating properties in order to increase the inductance L of the coil. Such an approach increases the complexity and cost of the antenna on a chip and also limits the ability to reduce the size of the antenna because of the need for the magnetic film layers between the antenna coils.

Another problem with the approach as taught in U.S. Pat. Nos. 4,857,893 and 4,724,427 is the attendant integrating of an antenna on an integrated circuit (IC) chip. In the case of an antenna disposed about the periphery of the chip, as described by the patents discussed hereinabove, the location of the antenna interferes with conventional bond pad layout about the periphery of the chip. Also, the electromagnetic fields within the central area of an antenna laid out about the periphery of a chip can interfere with the operation of circuits located within the antenna.

U.S. Pat. Nos. 6,373,447 discloses the use of one or more antennas that are formed on an integrated circuit chip connected to other circuitry on the chip. The antenna configurations include loop, multi-turned loop, square spiral, long wired and dipole. The antenna could have two or more segments which could selectively be connected to one another to alter effective length of the antenna. Also, the two antennas are said to be capable of being formed in two different metallization layers separated by an insulating layer.

U.S. Pat. No. 6,373,447 attempts to overcome the problems of U.S. Pat. Nos. 4,857,893 and 4,724,427 by having the antenna formed on other than a peripheral area of an integrated circuit (IC) chip. U.S. Pat. No. 6,373,447 describes devices comprising an integrated circuit (IC) chip having a surface wherein the surface is divided into a peripheral area and a central area. The central area comprises a first portion and a second portion wherein the circuitry of the IC chip is disposed within the first portion of the central area; and an antenna disposed within the second portion of the central area. This patent teaches that one or more antennas can be formed on an integrated circuit chip connected to other circuitry on the chip. The antenna configurations include loop, multi-turned loop, square spiral, long wired and dipole. The antenna could have two or more segments which could selectively be connected to one another to alter the effective length of the antenna.

U.S. Pat. No. 6,373,447 attempts to avoid the problems of the antenna interfering with the conventional bond pad layout about the periphery of the chip by placing the antenna in the central area of the chip. However, in all cases disclosed by the inventor the antenna is located in one portion of the central area, and IC circuitry is disposed on the other portion of the central area. The inventors do state in the body of the patent that the antennas can also be located on "one or two or three (but not four) edges of the chip, so long as there are no bond pads in that area".

A problem with the prior art as discussed above is that disposition of the antenna on the chip is limited to specific different parts of the chip. In the case of U.S. Pat. Nos. 4,857,893 and 4,724,427 the antenna is on the periphery of the chip and in the case of U.S. Pat. No. 6,373,447 the antenna is in the central area of the chip.

A further problem with the antenna on chip designs as disclosed by prior art is that the antenna and IC circuitry are disposed as discrete components in separate portions of the chip to minimize the interaction of the antenna and the circuitry of the IC chip. In the case of U S. Pat. No. 6,373,447, the antenna is one portion of the central area and the IC circuitry is the other portion of the central area. This again limits the design of placing an antenna on a IC chip. Furthermore, it makes for devices that are not truly integrated.

SUMMARY OF THE INVENTION

The present invention provides an integrated circuit chip which has at least one integrated and preferably integrally formed antenna with circuitry operatively associated with the antenna for receipt of energy harvested by the antenna from space. In another embodiment, a printed circuit board or other wireless untethered device is employed. The antenna has an effective area greater than its physical area. A system which receives power from the circuitry generally at DC power resulting from rectification by the circuitry of the energy harvested may also be provided on the chip.

The antenna and circuitry may be positioned at any location on the chip without adversely effecting performance.

The antenna may be a spiral antenna such as a square spiral antenna, with the thickness and width of the conductive portion being adjusted to maximize performance. Impedance may be adjusted to provide for an enhanced effective antenna area.

It is an object of the present invention to provide an integrated antenna on a chip or printed circuit board with associating circuitry to facilitate efficient receipt of energy from space by the antenna and conversion of the same by the circuitry to DC power.

It is a further object of the invention to provide such a semiconductor integrated chip or printed circuit board which will deliver DC power from the circuitry to an electronic system.

It is another object of the present invention to provide such a system which does not require the use of a power storage element or hard wiring.

These and other objects of the invention will be made fully understood from the following description on reference to the appended drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** illustrates a circuit which is an operational device but has no battery or wires for powering.

FIG. **2** illustrates a circuit of the type shown in FIG. **1** with an associated display to measure the energy harvested.

FIG. **3***a* shows a chip having an on-chip antenna for harvesting data and an on-chip on chip antenna for transmitting data.

FIG. **3***b* is a photograph of the chip represented in FIG. **3***a*.

FIGS. **4***a* and **4***b* show, respectively, (i) a plot of experimental results and (ii) a circuit and related power measuring device along with a plot of variations in result due to use of a conductive ground plane.

FIGS. **5***a* and **5***b* show, respectively, a plot and measurement in a representation of the circuit and power representation of ground plane influence.

FIGS. **6***a*–**6***d* show **4** square antenna profiles on a printed circuit board.

FIG. **7** is a plot of distance versus voltage for the four antennas shown in FIG. **6**.

FIG. **8** is a plot of voltage harvested at load for a harvesting distance of 50 cm.

FIGS. **9***a*–**9***e* show five antennas with four being spiral and one being a square patch.

FIG. **10** is a plot of voltage versus distance for the antennae of FIGS. **9***a*–**9***e*.

FIG. **11** is a representation of impedances for various parameters.

FIGS. **12***a*–**12***c* illustrate three variations in trace width for an antenna with plots of voltage versus distance.

FIGS. **13***a*–**13***b* illustrate data for antennas of two different materials and three different thicknesses.

FIGS. **14***a*–**14***b* show plots of voltage versus distance for the antennas of FIGS. **13***a*–**13***b*.

FIGS. **15***a*–**15***b* show plots of voltage versus distance for different lengths.

FIGS. **16***a*–**16***b* illustrate two different feed arrangements for antennas.

FIG. **17** is a plot of voltage versus distance for the two feed lines of FIGS. **16***a*–**16***b*.

FIGS. **18***a*–**18***b* illustrate several antennas with respect to ground plane variations.

FIG. **19** illustrates a plot of distance versus voltage for some of the antennas of FIGS. **18***a*–**18***b*.

FIG. **20** illustrates a circuit in terms of impedance matching elements.

FIG. **21** illustrates the circuit of FIG. **20** with impedance related markings.

FIGS. **22** and **23** illustrate, respectively, experimental and simulated impedance matching plots.

FIGS. **24***a*–**24***b* illustrate, respectively, plots of PCB data points and a projection of the same to about 25 mm.

FIG. **25** shows plots of voltage versus distance for PCB and CMOS traces.

FIGS. **26***a*–**26***b* show, respectively, plots of original data for copper and extrapolation of the same in respect of thickness.

FIG. **27** illustrates six antenna orientations.

FIG. **28** is a plot of voltage versus distance for the orientations of FIG. **27**.

FIG. **29** shows trend lines for the six plots of voltage versus distance as related to FIG. **27**.

FIG. **30** shows the negative slope and y intercept of the trend lines.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

As employed herein, the tern "in space" means that energy or signals are being transmitted through the air or similar medium regardless of whether the transmission is within or partially within an enclosure, as contrasted with transmission of electrical energy by a hard wired or printed circuit boards.

As employed herein the term "wireless untethered device" means a device which is remotely powered by RF energy and shall include, but not be limited to, IC chips, printed circuit boards and other suitable substrates having discrete or integrated circuit components.

What is further needed is an antenna on a chip wherein the antenna is not a discrete component to the IC circuit. In that respect there are two issues; (1) the interference of the received RF energy on the proximate IC circuitry, and (2) the interference of the proximate IC circuitry on the RF energy. While the first issue is important to any functional IC, the separation of the energy harvesting function from the payload IC electronic circuitry, U.S. Pat. No. 6,289,237, discloses isolating the negative effect of the RF from the payload electronics with respect to RF signals as opposed to RF energy. Our experiments have shown that while the payload electronics are in the proximate area of the energy harvesting antenna and associated circuitry, the payload will still function satisfactorily.

The second issue is in fact a benefit to the energy harvesting antenna and adds to the effective area of such an antenna(s). The functional devices for the experiments of the previous paragraph are pictured in FIGS. **1** and **2**. In both cases, the proximate circuitry has been demonstrated to operate satisfactorily in an applied RF energy field while simultaneously enhancing the harvested energy through providing additional area to the effective area to the antenna. The ability to obtain a tractable closed form mathematical formula to describe the indicated situation is extremely difficult at best due to the multitude of effects and interactions by the interconnection circuitry and the proximate components themselves.

FIG. **1** illustrates an operational device with all discrete components and conducting traces in place with no battery or wires for powering. The energy is supplied by beaming radio frequency (RF) energy to the device, rectifying the RF and converting it to direct current (DC), and powering electronics to read a temperature sensor and transmit the result to a receiver. The device requires five milliwatts of power to function. This device is functional as illustrated.

FIG. **2** is a device to measure power where the energy harvesting circuitry is the same as the device in FIG. **1**, but with five sets of light emitting diodes (LED's) connected in parallel to measure voltage without the need to connect wires to the device energy harvesting circuitry. Relative power can also be estimated by observing the brightness of the LED's. The conductors to interconnect the payload components of FIG. **1** are included in FIG. **2** while the proximate electronic components themselves have been removed.

FIG. **3** shows a single chip RFID design using an on-chip antenna for energy harvesting and an on-chip antenna for transmitting data. FIG. **3**.*a*. is the "flatten" picture of the layout from the Computer Aided Design (CAD) tool, and FIG. **3**.*b*. is a photograph of the actual fabricated chip (die).

There is an uncertainty in any in situ energy or power measurements due to the measuring circuitry interfering with the device under test. FIGS. **4***a*–**4***b* include the power measuring device of FIG. **2** along with a graph showing the number of LED's lighted when the space between the energy source and the device is interrupted by a conducting plane inserted perpendicular to a line drawn between the source patch antenna and the device, both of which are parallel. The top of the device was 19.0 cm above the laboratory table top. Conducting planes of 11 cm, 14.7 cm, 16.0 cm, 17.25 cm and 19.0 cm were placed between the antenna and the device. This resulted in lighted LED's numbering 5 (no conducting plane), 4, 4, 3, 2, and 0 respectively for the conducting planes. FIG. **5** is the same set-up as FIG. **4**, where the conducting planes were inserted in the order 17.25 to 19.0 cm, 16.0 to 19.0 cm, 14.7 to 10.0 cm, 11.0 to 19.0 cm, and 0.0 to 19 cm. In FIGS. **5***a*–**5***b*, the LED's that were lighted were 5, 5, 5, 5, 4, and 0.

In the first embodiment of this invention, a device is disclosed comprising an integrated circuit (IC) chip having disposed on it antenna and the circuitry for the IC chip. The antenna or antennas and the circuitry of the IC chip are integrated so as to provide the IC chip with an improved antenna effective area.

In other embodiments, a wireless untethered device such as a printed circuit board may be provided with the desired antenna and circuitry.

"Effective area" of the antenna refers to the functional area of an antenna as compared with the physical area. A tuned antenna may have an effective area that is larger than its geometric area. The phenomenon was disclosed by Reinhold Rudenberg in 1908 [Rudenberg, Reinhold, "Der Empfang Elektrischer Wellen in der Drahtlosen Telegraphie" ("The Receipt of Electric Waves in the Wireless Telegraphy") Annalen der Physik IV, 25, 1908, p. 446–466.] and the description of the phenomena has been expanded upon over the years by many other writers.

U.S. Pat. No. 5,296,866 states that "Rudenberg teaches that the antenna interacts with the incoming field, which may be approximately a plane wave, causing a current to flow in the antenna by induction. The [antenna] current, in turn, produces a field in the vicinity of the antenna, which field, in turn, interacts with the incoming field in such a way that the incoming field lines are bent. The field lines are bent in such a way that the energy is caused to flow from a relatively large portion of the incoming wave front, having the effect of absorbing energy from the wave front into the antenna from a wave front which is much larger than the geometrical area of the antenna."

While the concept of effective area may be known, implementing it in antenna design and construction is not easy or obvious. U.S. Pat. No. 5,296,866 teaches making active antennas that had effective areas greater than the geometric areas through use of special circuitry. The antennas detailed in this disclosure were non-planar coil antennas.

Co-pending U.S. patent application Ser. No. 09/951,031, the disclosure of which is expressly incorporated herein by reference, discloses an antenna on a chip having an effective area 300–400 times greater than its physical area. The effective area is enlarged through the use of an LC tank circuit formed through the distributed inductance and capacitance of a spiral conductor. This is accomplished through the use on the antenna of inter-electrode capacitance and inductance to form the LC tank circuit. This, without requiring the addition of discrete circuitry, provides the antenna with an effective area greater than its physical area. It also eliminates the need to employ magnetic film. As a result, the production of the antenna on an integrated circuit chip is facilitated, as is the design of ultra-small antennas on such chips. See also U.S. Pat. No. 6,289,237, the disclosure of which is expressly incorporated herein by reference.

The invention of this disclosure integrates the antenna and the circuitry of a wireless untethered device such as an integrated circuit (IC) chip or printed circuit board (PCB) to achieve an effective area that is greater than what would be expected from the antenna disposed on the chip alone. By "integrate" or "integration" it is meant that the antenna is disposed on the chip using the circuitry of the IC chip or PCB as part of the antenna function specifically the energy harvesting effective area. Integrating the antenna and proximate circuitry of the IC chip or PCB simplifies the design of the chip because placing the discrete antenna and proximate IC chip or PCB circuit components in discrete areas of the chip is no longer necessary.

In general, a preferred approach is to have the antenna designed as an integral element which is connected to the circuitry with proper impedance matching. If desired, the circuitry can be designed and then the antenna can be designed for maximum performance efficiency in the presence of the circuitry with particular emphasis of use in small, wireless untethered devices.

One advantage of integrating the antenna and the proximate circuitry of the IC chip or PCB is that the physical size of the antenna component is no longer the sole driving force in obtaining greater effective area. Using the embodiment of this invention, the effective area of the antenna is a combination of the physical size of the antenna and proximate IC circuitry on the IC chip or PCB. In essence, the antenna on the IC chip or PCB for all practical purposes is the whole IC chip or PCB and not a specific portion of IC chip or PCB that has the physical antenna disposed on it. It will be appreciated that in the IC chip, PCB or other embodiments the effective area of an antenna may be greater than the physical area of the IC chip, the PCB or other substrate. In fact, isolating the physical antenna into separate portions of the IC chip or PCB (i.e., peripheral or a distinct area of the central portion) can be detrimental to making IC chips or PCBs with effective areas greater than the effective area from the isolated physical antenna.

The design of an antenna on chip or PCB involves a number of sets of information; (1) frequency of operation (center frequency), (2) bandwidth at the half power points, (3) Q—ratio of peak performance to that of the half power points, (4) the electrical and magnetic properties along with thickness of each process layer—typically termed the technology files.

Given the appropriate design information, an architecture, e.g., whip, dipole, square spiral, round spiral, etc. is selected. Next, a range of parameters such as conductor length equal to wavelength (center frequency) divided by 4, resonance of the antenna at the center frequency, etc.

Using the resonance strategy, the antenna parameters and configuration are then adjusted to achieve resonance using the parasitic and inter-electrode inductance and capacitance including the electronic and magnetic properties as given in the technology files.

The resonance is normally checked through computer aided design software including simulation.

FIGS. **4***a* and **4***b* illustrate the test apparatus including; (A) two spiral energy harvesting antennas with circuitry for rectification and voltage “doubling” (charge pumps), (B) interconnection circuitry, and (C) the power, voltage measurement circuitry (power meter).

The measurement of power and/or voltage on a remote autonomous device is complicated by the inability to connect non-interfering wires to the device, and the fact that as the size of the device decreases, there is an inherent uncertainty concerning the effect/contribution of the measurement circuitry itself on the measurement. The effective area measurements in the Co-pending U.S. patent application Ser. No. 09/951,031, involved connecting wires in order to effect the power measurements.

A series of experiments were run with the device of FIG. **4***b* in the field of an RF source antenna at 915 MHz. In these experiments, a flat conducting surface (plane) was placed between the source antenna providing the RF energy and the device under test at an approximate distance of 0.25 inches from the device. Five conducting planes were used where the planes shielded areas of the device at the 5 heights shown in FIGS. **4***a* and **4***b*.

The height of 19.0 cm represents the height of the top of the device above the laboratory bench. The 5 heights of the conducting plane in the 5 experimental positions were 19.0 cm, 17.25 cm, 16.0 cm, 14.7 cm and 11.0 cm respectively. The points on the graph above the device in FIG. **4***a,* represents the number of LEDs lighted on the power meter section (C) of the device. The number of LEDs lighted is proportional to the voltage across the LEDs and represents a lower bound on the voltage dropped across the 1 K ohm resistor that can be seen in area (B**2**) of FIG. **4***b*.

As can be noted from the FIG. **4***a* graph, while exposing the strict antenna area (A**1**) to the RF field does in fact harvest sufficient energy to light **2** LEDs, exposing additional device area facilitates additional energy harvesting, including the power meter area (C).

It is noted that for the proximate electronics in (B**2**) and (C), the function of the circuitry is strictly DC including the DC bus noted in area (B**1**).

A functioning power meter (C) when totally removed from the device parts identified as (A) and (B) shows no lighted LEDs. Thus, as an energy harvesting device, the power meter by itself will not harvest sufficient energy to light a single LED. However, when connected to the device, the area and circuitry contribute to the energy harvesting by an amount of 74.9 milliwatts in terms of measured power and sufficient voltage to light an LED.

In FIGS. **4***a* and **4***b,* the conducting planes were introduced in the sequence 19.0 cm, 17.25 cm, 16.0 cm 14.7 cm and 11.0 cm. The results of a similar experiment with the sequence of conducting planes were introduced in the sequence 17.25 to 19.0 cm, 16.0 to 19.0 cm, 14.7 to 19.0 cm 11.0 to 19.0 cm and 0 to 19.0 cm, respectively.

While no mathematical model is proposed for the results of FIGS. **4** and **5**, it is obvious that the proximate interconnections (B) and circuitry (A**2**) contribute to the total amount of energy harvested by increasing the effective area of the harvesting antennas.

It is clear that more than one factor contributes to the effectiveness of the antenna/device combination in harvesting energy. We have chosen to note the effective area do to the results of Rudenberg as Rudenberg effective area, and the remainder of the effective area due to components and interconnections as proximate effective area.

The present invention is concerned with the design and analysis of small antennas for the purpose of harvesting energy on a printed circuit board (PCB) or a complementary metal oxide semiconductor (CMOS) die in order to provide power to on board circuitry without the need for batteries or external connections as a source of power.

Earlier design of such devices has led to the desire to use a square spiral coil as the profile for the antenna. Thus, this disclosure will be mostly concerned with the square spiral antenna although a patch design will be included from time to time.

Given the square spiral profile, there are still many parameters of the design that must be considered and chosen. This disclosure is concerned with a series of experiments that were run in an effort to evaluate these alternatives using PCB implementations. However, the ultimate goal is to implement the square spiral on a CMOS die thus opening up the question of what differences in performance can be expected with the CMOS implementation.

The approach to the analysis is strictly empirical although a few references may be made to classical approaches either to antennas or electronic circuits. The sole purpose is given a square spiral implementation, what are the parameters values that provide the greatest amount of power or energy at the payload circuitry in this case a 10 K ohm resistor.

Printed circuit test boards each containing a target antenna have been constructed where the voltage across the target load is measured by an analog to digital converter, frames and transmitted by an infra red transmitter to a receiver connected to a data gathering personal computer. The test/measurement circuitry is powered by a small battery. All test/measurement circuitry and the battery are located behind a ground plane for the circuitry so as to shield the circuitry from an energizing radio frequency (RF) source whereby a contribution to the power/energy can be attributed to this circuitry.

Other research has demonstrated that the proximate circuitry can be used to advantage in harvesting energy.

Previous attempts to measure the energy harvested by small antennas have been complicated due to the measurement equipment that is connected to the device under test (DUT). Thus, the attempt here is to include the test circuitry directly on the DUT. As indicated previously, the circuitry included with the DUT can have an effect on the results thus leading to the shielding of the test circuitry behind a conducting plane as shown in FIGS. **6***a*–**6***d*.

Over many decades it has been difficult if not impossible to prove that the test circuitry does not have an effect on the results of the experiment, the results reported in this paper suggest that the effect is minimal through some of the profiles of the DUT, e.g., antenna area as shown in FIGS. **6***a*–**6***d* for antennas A through D.

The design of antennas in a limited space such as on a printed circuit board or a Complementary Metal Oxide Semiconductor (CMOS) chip or die requires the consideration of a number of physical factors of the antenna to obtain an optimum performance. These factors include:

(1) The area allocated to the antenna including the actual metallic conductor forming the antenna and any space required to form or shape the metallic conductor. The metallic conductor has certain characteristics, e.g., shape (profile), conductor width, conductor thickness, and conductor length.

(2) One of the profiles considered in this research is a square spiral. The topological nature of the square spiral is that connection can be made in the plane of the antenna from the last outside trace of the spiral. Such a connection will be noted as a micro-strip line feed. The inner end of the spiral must be connect through a via to another layer of the PCB or CMOS die. Such a connection will be noted as a probe feed. Two general profiles are considered in the research reported herein, a square spiral and a square patch.

(3) The antenna environment includes any form of conductor, insulator, etc., that exist between the source of the radio frequency (RF) energy and the antenna or behind the antenna. One form of element can be a ground plane formed as a part of the PCB or CMOS die.

(4) The antenna is connected to a circuit (energy harvesting in this case) where both the antenna and the circuit (load) have an impedance. For optimum performance, it has long been held that the two impedances should be matched as complex conjugates, i.e., impedance matching, with equal real (resistance) components.

(5) There are two areas involved in the antenna design. The first is the physical area required to implement the antenna. The second is the effective area resulting from dividing the power received at the load by the power density in the field in which the antenna is place, i.e., Power Harvested=Power density*Effective Area.

In summary, the factors involved in the design of an antenna in a limited space, e.g., a PCB or a CMOS die are:

(A) Physical Area
(B) Profile
(C) Antenna Conductor Width
(D) Antenna Conductor Thickness
(E) Antenna Conductor Length
(F) Feed Type
(G) Ground Plane Existence
(H) Impedance Matching

(A) Physical Area

Four antennas of a square profile with four different areas as illustrated in FIGS. **6***a*–**6***d*, have been constructed. The voltage response measured at the load of the four antennas in shown in FIG. **2**. The antennas are compared at a harvesting distance of 50 centimeters by plotting the voltage harvested at the load as shown in FIG. **7**. Based on the results of FIG. **8**, it is clear, as expected, that the amount of energy harvested increases with increased area. This area is termed the physical area.

(B) Profile

Two profiles are considered in this study, a square spiral and a square patch. FIGS. **9***a*–**9***e* illustrate five antennas, four of which are square spirals and one is a square patch (Antenna D, Board H**1**).

The voltage results for the profile test cases are illustrated in FIG. **10**. It can be argued from FIG. **10** that the best results are obtained with the square patch. The differences are not great, but there are differences. However, the square patch has a disadvantage in that the impedance for impedance matching requires an inductor as the patch has a capacitive impedance. In addition, the impedance of the antenna is fixed by the physical area. In the case of the square spiral antennas, the impedance of the antenna may be varied while maintaining a fixed physical area. In the case of the square patch, the matching requires an inductor that is difficult to fabricate on a CMOS die and not easy to precisely match with an on die inductor or a discrete inductor for the PCB implementation. The inductor is required to form the complex conjugate as the square patch is capacitive.

The square spiral can be modified by trace width, thickness and relative spacing to have essentially a continuum of impedances many of which are illustrated in FIG. **11**. This variation is particularly useful in coordinating the matching impedance between the load and the antenna. Thus, from a design standpoint, the square spiral is the profile of choice.

Antenna E in FIG. **9***e*, while a square spiral, has length greater that the ¼× length of Antennas A, B, and C. The length of antenna E is greater than ¾×. Thus, the length of the antenna conductor is a factor in designing the most efficient PCB or die (chip) antenna. While the ¼× length is better for design purposes as the added length does not gain any appreciable advantage, this does not constitute a proof for the optimum length. The major contribution here is that the antenna conductor length is a factor that must be considered in the antenna design.

(C) Antenna Conductor Width

The antennas A, B and C of FIGS. **9***a*–**9***c* vary in trace width. Based on the results of FIG. **10**, the trend lines are shown in FIG. **12**.

From FIG. **12**, the slope, i.e., rate of voltage decrease is the least for antenna B. In addition, the trend line for antenna B has the largest y intercept. Both of these results support antenna B with the narrowest trace width as being the optimum for this set of experiments.

(D) Antenna Conductor Thickness

Two independent experiments were run for antenna thickness. The first began with one of the type A antennas to which additional conducting solder was first added and then some of the added solder was removed. This gave an original thickness, a much thicker trace and thirdly a trace of intermediate thickness. These steps produced three different thicknesses of the antenna conductor thicknesses indicated in FIG. **13***a*.

Using another set of independent boards and steps, all solder was removed from the antenna exposing the bare copper giving the thinnest trace thickness. The copper was then electroplated onto the antenna to give different thicknesses of copper and thus three different values of antenna conductor thicknesses as indicated in FIG. **13***b*. The measured results for both lead and copper additions are illustrated in FIGS. **13***a*–**13***b*.

Based on the results shown in FIGS. **14***a*–**14***b*, it is quite clear that the best energy harvesting, highest voltage levels, is accomplished antenna conductor thickness is the thinnest. The results of the two thicker traces in FIG. **14***a* are very close and no conclusion should be drawn solely with respect to these two antennas. The removal of the solder from the thickest trace caused a minor amount of melting of the solder having a minimal effect on the antenna profile. However, these two thicknesses relative to the thinnest trace thickness can be considered significant.

(E) Antenna Conductor Length

The antenna conductor length results can be observed considering FIGS. **9***a*–**9***e* and **10**. While there does appear to be some equivalence in the results of antenna B and antenna E in FIG. **19**, the results at the 50 centimeter distance was chosen as the point of comparison show that nothing significant is gained by the additional length. In fact, the length is definitely less useful when compared to antennas A and C. Thus it is an initial conclusion that the primary criterion for the antenna to minimize area and maximize the amount of energy harvested that the length by held at or near ¼ wavelength.

FIG. **15***a*. is the trend line for Antenna B, and FIG. **15***b* is the trend line for Antenna E. In comparing B and E, the rate of drop off for E is greater and E also has a lower y intercept

There is perhaps another significant point in comparing the two antennas B and E. From a -traditional antenna design standpoint, it could be argued that the two antennas should have performances with more significant differences than are shown in these results.

The explanation offered here is that the most significant feature of the small antennas is the fact that there exists a resonant tank disturbing the field which modified the effective area of the antenna. Thus, these two antennas of significantly different lengths produce similar results supporting the notion that the resonant tank circuit is the most significant aspect of the small antenna.

(F) Feed Type

FIGS. **16***a*–**16***b* provide an illustration of the two type of feed from the antenna, i.e., (a) the micro-strip line feed and (b) the probe feed.

The voltage results of the two type of line feeds are shown in FIG. **17** where it is quite clear that the preferred type of connection to the antenna is the micro-strip line feed.

(G) Ground Plane Existence

In many applications of different antenna profiles such as a patch, it is desirable to provide a ground plane that is separated from the antenna by a distance that can be calculated based on the dielectric constant of the medium between the ground plane and the antenna. However, it the case of a PCB or CMOS die implementation, the distance between the antenna and the potential location of the ground plane and the dielectric constant of the intervening medium are both fixed by the process independent of the antenna design. Thus, as a rule, the resulting ground plane is not located at the proper position for the fixed dielectric, or the dielectric is incorrect for the position of the ground plane.

The current PCB experiments included provisions for testing the ground plane existence and the net effect of some form of a ground plane in the proximity of the antenna. Two types of ground plane experiments are illustrated in FIGS. **18***a*–**18***b*. The first is a conventional ground plane directly behind the antenna, and the second is a ground plane in the proximity of the antenna on the opposite side of the PCB. The corresponding voltage measurements for the case of FIG. **18***a* are shown in FIG. **19**. The results for the three cases illustrated in FIG. **18***b* were indistinguishable.

Based on FIGS. **18***a*–**18***b*, it is clear that the ground plane is not a desirable feature for the PCB implementation. As a curious note, it would be possible to find an antenna/frequency combination for which the ground plane could lead to an advantage. However, such a result would not be practical in view of the restrictions on frequency both from regulatory and power/energy considerations.

As noted previously, the results for the three different sized ground planes illustrated in FIG. **18***b* were indistinguishable. However, the optimum voltage present in all three cases was the result of different physical impedance matching of the three antennas. Thus, while the proximate ground planes of FIG. **18***b* do not affect antenna performance, they do provide an affect on the antenna and feed line impedances.

The conclusion here is that in general, it is not desirable to include a ground plane for a PCB or CMOS die type of antenna. In additional a physical ground effect object in the proximate area may be useful in achieving a certain impedance or impedance match.

(H) Impedance Matching

As indicated previously, it is in general desirable to match the impedance of the antenna to the impedance of the load as complex conjugates. However, in the case of the use of an antenna for the sole purpose of energy harvesting, there is reason to believe that another form of matching or designed mismatching may not be of an advantage. Consider the model of an energy harvesting antenna as shown in FIG. **20**. As reported by Rudenberg, the antenna forms a non-ideal tank that has a nonzero resistance. This resistance results in a retransmission of some of the incoming energy that in turn interferes with the transmitted field resulting in an increased effective area of the antenna.

As a result of this research, it is conjectured that an additional amount of retransmitted energy may be of an advantage in disturbing the field to provide an even larger effective area. It is recognized that this increased effective area is not without limit.

The Rudenberg retransmitted energy is a function of the resistance R shown in the tank circuit of FIG. **20**. We conjecture that in addition, the intentional mismatch of the impedance Z of FIG. **20** can result in increased effective area.

The effective area increase by the resistor, R, of the tank in FIG. **20**, will be referred to as the Rudenberg effective area. The increased effective area caused by the mismatch of Z and the resulting reflected energy from the load is termed the proximate effective area.

The net effect of these two effective area contributions is modeled as two voltage sources just below the ideal antenna symbol of FIG. **20**.

These two features contributing to the effective area are highlighted in FIG. **21**.

The basis of the conjecture on the mismatch of the impedance Z resulting in reflected energy from the load is arrived at from the repeated attempts to correlate the simulated results for impedance matching with the experimental results found when actually choosing an impedance to increase the harvested energy/voltage experimentally. The results of the simulation and experimental impedance matching are shown in FIGS. **22** and **23**.

FIG. **22** represents the compiled voltage results with many different values of impedance used to obtain the most effective match through an organized searching procedure. The N/S/E/W cross of FIG. **22** represents the centroid of the four best matches that could be achieved by physical matching using electronic elements.

FIG. **23** represents the compiled values of matched impedance along with color coding for strength. The cross of FIG. **23** represents the centroid of the four best matches for the simulations. It is important to note that the simulation may obviously in general not represent the actual circuit due to minute differences that are extremely hard to model. However, numerous attempts were made to calibrate the simulation based on the model and the corresponding test results. However, none of these was successful.

Thus, at this time, the difference in the simulated matching of FIG. **23** based on the complex conjugate impedances and the actual best match in the physical tests is attributed to the mismatching of the impedance z in FIG. **21**. There would be a limit on the effectiveness of this mismatch.

As pointed out previously, the two voltage sources shown in FIGS. **22** and **23** immediately below the ideal antenna symbol are attributed to the retransmission of the energy as a result of the R in the non-ideal tank and the reflected energy from the load due to the mismatch of the impedance Z.

CONCLUSIONS

One of the important aspects of this research is the ability to empirically evaluate the feasibility of certain results that require much additional research other than the antenna. For example, consider a system on a chip (SOC) implementation that requires other circuitry to be designed that will in turn be powered by the energy harvesting technology being discussed here. A considerable amount of work may be required to develop the payload circuitry, and it is desired to have some idea of what level of power can be achieved.

FIG. **8** gives the trend line for the harvested voltage across a 10K ohm load. Therefore, with an antenna that is 1.58 mm on a side, the power that can be expected to be harvested with the thickness of antenna A is:

$$\text{Power}=\{[O.5918*(1.582)J0{\sim}{\sim}85]2\}/104=80\ j.t\ \text{watts}$$

However, based on the narrower and thinner trace, the voltage can be expected to be at least doubled. This would result in:

$$\text{Power}=\{[2*0.591\ 8*(1.582)]04485]2\}/104=178\ j.i\ \text{watts}$$

The results appear to show that the highest voltage is achieved with a narrower and thinner conductor trace. This is not unexpected from the energy harvesting scenario. At the frequencies involved, the current tends to flow on the outside surface with possible eddy current flowing in the interior that will contribute losses.

Consider reducing only the width of the traces. The PCB traces are on the order of 700 to 1300 microns. A plot of the data for the slope in the linear fall off($y=-Mx+b$, M is the linear fall off or slope of the trend line) as shown in FIG. **7** for antennas A, B and C is shown in FIG. **24***a*.

The point of this discussion is solely to see the fall off rate for the harvested energy as the antenna trace pattern is reduced in width. The trend lines of FIG. **10**, for example, have both a slope and a y intercept, M and b in the above equation for y. The discussion here concerns solely the slope, M, of the trend line. The y intercept appears to be primarily a function of the antenna area and is outside this current discussion.

Based on the projection shown in FIG. **24***b,* the fall off rate in energy is, −0.0328, compare with, −0.0421, for Antenna B as the antenna trace width is reduced to 25 microns.

Thus, based solely on the projection in FIG. **24***b,* the rate improvement of the fall off in going to a CMOS die from a PCB implementation is about 22%. Therefore, in general, it can be expected that the deterioration in performance for a CMOS die will be much better (less per unit distance) than a PCB implementation as the distance between the source and device increases.

An example of the effect of this difference can be seen in FIG. **25**, assuming the intercept point that would make the harvested voltages at 50 centimeters to be the same.

Another interesting look at the projections in going from PCB antennas to those of a CMOS die can be obtained by considering the thickness results from FIG. **13***b*. The results of the decrease in thickness are shown below in FIG. **26***a*. Based on the trend line fitted to these data points, an extrapolation to approximately 1 micron is shown in FIG. **26***b*. f it can be assumed that both the PCB and the CMOS die have the same value at say 50 centimeters, it can be conjectured that the break point as shown in FIG. **14***b,* can be extended to 13,600 centimeters or 446 feet.

Obviously the simple extension to 446 feet from 70 centimeters is difficult to defend. The point here is that there are certain of the parameters in the design of the antenna that must be taken into account that provide direction into the experimental development that will eventually produce a working result although the theory as to why may be delayed by a considerable time.

There is no suggestion, other than the extrapolation, that the 446 feet can be achieved. This result is for copper, and most current fabrications are done with aluminum.

One major conclusion of these results is that the transition from a PCB implementation to a CMOS die need not be considered to be an unfavorable transition.

The primary reduction in the PCB to CMOS die transition is the attendant desire to reduce area because of the relative cost of a PCB compared to a CMOS die per unit (equivalent) area.

All of the previous results were obtained with the antennas in the same orientation with respect to the source antenna. In order to evaluate the orientation issue, FIG. **27** shows six orientations that were tested.

The voltage as a function of distance curves for the six configurations are shown in FIG. **28**. Obviously, the most favorable positions are positions **1** and **2**. Position **1** is the position that was used for all of the previous results. The two orientations, **5** and **6**, are included for the sake of completeness. However, these two are not at all favorable. One reason for their inclusion was that these two were quite favorable in the early tests on the Delta **1** CMOS antenna. At this time there has been no attempt to discover why the PCB and CMOS results for the horizontal antenna appear to be different.

The trend lines for the curves of FIG. **28** are given in FIG. **29**. From the trend lines, position **1** is the best in terms of the y intercept. However, the fall-off rate for position **2** is the best for the two positions. If we consider the longest distance for a typical PIC microcontroller to function based on 2.5 volts, Position **2** yields 90.8 centimeters and Position **1** yields 89.1 centimeters. These two results are extremely close based on the trend lines.

The negative slope and the y intercept of the trend lines is shown below in FIG. **30**. The angle shown along the abscissa is the feed line of the micro-strip type feed where the angle is from the feed line in the direction of the load circuitry. In FIG. **27**, P**4** is 0 degrees, P**2** is 90 degrees, P**4** is 180 degrees, and P**1** is 270 degrees.

Based on the results reported here, the conclusions are summarized as follows:

1. Impedance matching is essential but does not guarantee optimum performance for a fixed area of a small antenna
2. A ¼ wavelength antenna appears to be the optimum antenna for a fixed area
3. The area appears to be the most important factor—given 1 and 2
4. The spiral antenna profile is preferred because it is much easier to match impedances
5. The narrowest trace (width) appears to be optimum
6. The thinnest trace (height) appears to be optimum
7. Increased conductor length does not appear to offer any benefit for energy harvesting
8. The micro-strip line feed is superior to the probe feed
9. There should be no ground plane under the antenna
10. The square spiral is preferred to the square patch based on impedance matching considerations
11. Both tank resistance and reflected load power due to mismatched impedances appear to increase the effective area of the small antenna

While for convenience of disclosure primary attention has been directed toward use of the preferred embodiments of the invention which employ IC chips or printed circuit boards, the invention may also be employed advantageously in connection with other substrates in wireless untethered devices.

Whereas particular embodiments of the invention have been disclosed herein for purposes of illustration, it will be evident to those skilled in the art that numerous variations of the details may be made without departing from the invention as set forth in the appended claims.

The invention claimed is:

1. A wireless untethered device comprising:

at least one antenna;

circuitry operatively associated with said antenna for receipt of energy harvested by said antenna from space, said circuitry being structured to convert said energy to direct current; and

a system structured to be energized by the direct current from said circuitry, said system and said circuitry having an impedance mismatch with respect to said antenna wherein said antenna, said circuitry and said system together form an integrated antenna and wherein said antenna has an effective area greater than its physical area.

2. The wireless untethered device of claim **1** including

said wireless untethered device being selected from the group consisting of an integrated circuit chip and a printed circuit board.

3. The wireless untethered device of claim **1** including

said wireless untethered device being an integrated circuit chip.

4. The integrated circuit chip of claim **3** including

said antenna being integrally formed on said chip.

5. The integrated circuit chip of claim **4** including

said antenna structured to receive RF energy and deliver it to said circuitry.

6. The integrated circuit chip of claim **4** including

a second integral antenna for transmitting data.

7. The integrated circuit chip of claim **4** including

said antenna being a spiral antenna.

8. The integrated circuit chip of claim **7** including

said spiral antenna being a square spiral.

9. The integrated circuit chip of claim **3** including

said circuitry having a tank circuit for increasing the effective area of said antenna.

10. The integrated circuit chip of claim **5** including

said antenna having a length of about ¼ wavelength of said RF energy.

11. The integrated circuit chip of claim **3** including

a micro-strip feed line connected to said antenna.

12. The integrated circuit chip of claim **3** including

said circuitry having a non zero resistance which results in retransmission of some of said energy to thereby interfere with transmitted energy and increase the effective area of said antenna.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 7,057,514 B2 Page 1 of 2
APPLICATION NO. : 10/857527
DATED : October 5, 2004
INVENTOR(S) : Marlin H. Mickle, Christopher C. Capelli and Harold Swift

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 4, line 46, “tern” should be changed to --term--.

Column 8, line 30, “do” should be changed to --due--.

Column 8, line 57, “what” should be changed to --that--.

Column 9, line 56, “place” should be changed to --placed--.

Column 10, line 5, “in” should be changed to --are--.

Column 10, line 39, “that” should be changed to --than--.

Column 11, line 11, the word “where” should be inserted after the word --accomplished--.

Column 11, line 60, “it” should be changed to --in--.

Column 11, line 48, “type” should be changed to --types--.

Column 11, line 51, “type” should be changed to --types--.

Column 12, line 24, “affect” should be changed to --effect--.

Column 12, line 28, “additional” should be changed to --addition,--.

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 7,057,514 B2 Page 2 of 2
APPLICATION NO. : 10/857527
DATED : October 5, 2004
INVENTOR(S) : Marlin H. Mickle and Christopher C. Capelli and Harold Swift

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 12, line 32, “in general” should be changed to --generally--.

Column 14, line 29, “f” should be changed to --If--.

IN THE FIGURES

FIGS. 7, 19 and 28, “te” should be changed to --the--.

Signed and Sealed this

Fourteenth Day of November, 2006

Jon W. Dudas

JON W. DUDAS
Director of the United States Patent and Trademark Office

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 7,057,514 B2 Page 1 of 2
APPLICATION NO. : 10/857527
DATED : June 6, 2006
INVENTOR(S) : Marlin H. Mickle, Christopher C. Capelli and Harold Swift

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 4, line 46, “tern” should be changed to --term--.

Column 8, line 30, “do” should be changed to --due--.

Column 8, line 57, “what” should be changed to --that--.

Column 9, line 56, “place” should be changed to --placed--.

Column 10, line 5, “in” should be changed to --are--.

Column 10, line 39, “that” should be changed to --than--.

Column 11, line 11, the word “where” should be inserted after the word --accomplished--.

Column 11, line 60, “it” should be changed to --in--.

Column 11, line 48, “type” should be changed to --types--.

Column 11, line 51, “type” should be changed to --types--.

Column 12, line 24, “affect” should be changed to --effect--.

Column 12, line 28, “additional” should be changed to --addition,--.

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 7,057,514 B2 Page 2 of 2
APPLICATION NO. : 10/857527
DATED : June 6, 2006
INVENTOR(S) : Marlin H. Mickle and Christopher C. Capelli and Harold Swift

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 12, line 32, “in general” should be changed to --generally--.

Column 14, line 29, “f” should be changed to --If--.

IN THE FIGURES

FIGS. 7, 19 and 28, “te” should be changed to --the--.

This certificate supersedes Certificate of Correction issued November 14, 2006.

Signed and Sealed this

Second Day of January, 2007

Jon W. Dudas

JON W. DUDAS
Director of the United States Patent and Trademark Office

International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 13 Issue 1 –MARCH 2015.

A PATCH ANTENNA TO HARVEST AMBIENT ENERGY FROM MULTIBAND RF SIGNALS FOR LOW POWER DEVICES

K.Yuvaraj[#1]and A.Ashwin Samuel[*2]
[#] *B.E, ECE,Srinivasa Institute of Engineering & Technology, Chennai, India*
[*] *B.E, ECE,Srinivasa Institute of Engineering & Technology, Chennai, India*

Abstract—In **this digitized world, every device is battery dependent. But battery is cannot to supply the require energy to devices for longtime it should be recharge after regular interval. It is not only impossible but also improbable to carry a power source or a battery along with us everywhere. We have ambient signals flying around us such as Mobile signals, FM signals, and Wi-Fi signals. So, why do not we use these signals to power up our devices? This paper would give an insight about how the above could be achieved.**

Electromagnetic energy harvesting holds a promising future for energizing low power electronic devices in wireless communication circuits. This article presents an RF energy harvesting system that can harvest energy from the ambient surroundings at the downlink radio frequency ISM (International Standard for Mobile communication) band range of GSM-900. The harvesting system is aimed to provide an alternative source of energy for energizing low power devices.

Index Terms**—Rectangular microstrip patch antenna (RMPA), Metamaterial (MTM), Voltage multiplier Directivity, Impedance, Bandwidth, Return loss, Gain.**

I. INTRODUCTION

There is an active research area investigating a number of alternative ways to extract energy from the environment and convert it into electrical energy for energizing low power electronic circuits directly or store it for later use. One such energy is from radio frequency. RF energy harvesting from the ambient will have an important role in the future microelectronic circuits. This work is being carried out by many researchers for the following reasons:

- The energy is freely available in space.
- Complementing the low power sources used for energizing the low power electronic devices, as an application to green technology.

RF energy harvesting from ambient sources have great potential to impact on the cellular phones and portable electronic devices. This concept needs an efficient antenna along with a circuit capable of converting RF signals into DC voltage.

In [1] the rectifier circuit with 50mV input for frequency range 902{928MHz achieved a maximum efficiency of 60%}.

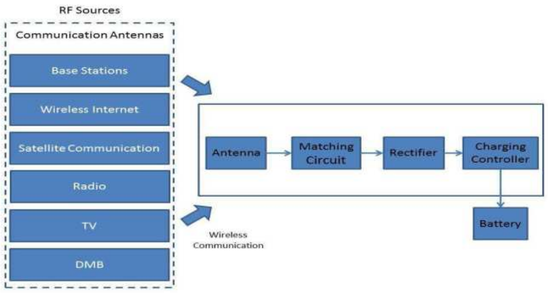

Fig 1. Block Diagram of RF Harvesting System

In [2] the authors proposed a system tocharge a small battery or to run a low power sensor at frequency range of 950 MHz{960 MHz. Due to frequency hopping of signals suggested to use a DSP processor, in which main challenge was powering up the DSP processor. The authors in [3] designed a square microstrip antenna with 50- impedance and used a 6-stage Schottky diode voltage doubler to form an RF energy harvesting system at 900MHz band. This work achieved an output voltage of 2.78V and 0.87V at a distance of 10m and of 50m respectively for a received power levels of 4dBm and 5dBm from a 900MHz cell tower.

In this article a novel RF energy harvesting system is designed with partial ground plane for energizing low power devices. Types of ambient resources and overview of this paper is shown in block diagram at Fig 1. The special features of this designed system are in the impedance of the antenna, matching network, ground plane technique and frequency of operation (downlink radio frequency range of GSM-950 band) which all differs from the previous works described.

The antenna and Pi matching network have been optimized through electromagneticsimulation software, CST Microwave studio2014.

The uniqueness of the system lies in the partial ground plane and the alignment of induced electric field for maximum current flow in the antenna that maximizes the captured RF energy. The design and simulation of the voltage multiplier circuit were performed using Multisim software. All the three modules were integrated and fabricated on a double sided FR-4 printed circuit board.

II.MICROSTRIP PATCH ANTENNA

A Microstrip Patch antenna consists of a radiating patch on one side of a dielectric substrate which has a ground plane on the other side as shown in Fig 2.The patch is normally made of conducting material such as copper or gold and can take any possible shape. The radiating patch and the feed lines are usually photo etched on the dielectric substrate. In order to simplify analysis and performance estimation, generally square, rectangular, circular, triangular, and elliptical or some other common shape patches are used for designing a microstrip antenna.

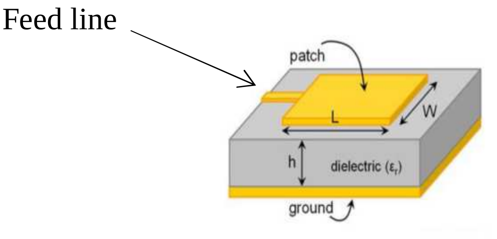

Fig 2. Microstrip Patch Antenna

For a rectangular patch, the length L of the patch is usually $0.3333\lambda_o < L < 0.5\lambda_o$, where λ_o is the free-space wavelength. The patch is selected to be very thin such that $t \ll \lambda_o$ (where "t" is the patch thickness). The height "h" of the dielectric substrate is usually $0.003\lambda < h < 0.05\lambda_o$. The dielectric constant (ε_r) is typically in the range $2.2 \leq \varepsilon_r \leq 12$.Microstrip patch antennas radiate primarily because of the fringing fields between the patch edge and the ground plane.

For good performance of antenna, a thick dielectric substrate having a low dielectric constant is necessary since it provides larger bandwidth, better radiation and better efficiency. However, such a typical configuration leads to a larger antenna size. In order to reduce the size of the Microstrip patch antenna, substrates with higher dielectric constants must be used which are less efficient and result in narrow bandwidth. Hence a trade-off must be realized between the antenna performance and antenna dimensions.

III.ANTENNA DESIGN

The proposed antenna was an E-shaped single patch from the conventional wide band microstrip antennas. The topology of the antenna was designed on a FR-4 substrate. It is a grade designation assigned to glass-reinforced epoxy laminate sheets and printed circuit boards (PCBs). This is one of the popular industry-wide standard substrate material formats for electronic circuit boards. The key property of FR-4 relates to the flame retardant qualities of the material. The important specifications chosen in simulation for this design are the thickness of substrate 1.6 mm, the thickness of copper 0.035 mm, the relative permittivity 3.9 and the loss tangent 0.01. The antenna size is characterized by its length, width and height (*L*, *W*, *h*) and is fed by Pi matching network, feed line and is followed by a partial ground plane. The antenna is designed and optimized to capture the energy from the ambient at downlink radio frequency range of GSM- 950 band.

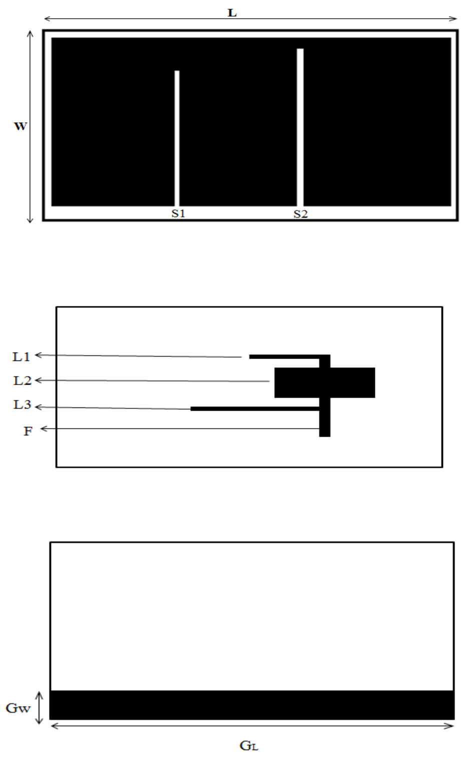

Fig 3. Design Consideration

In order to expand its bandwidth, two parallel slots are incorporated into this patch. The pi matching network is designed and optimized to provide an impedance matching for antenna 377 ohm impedance to the complex load 63-*j*117 ohm impedance (input of the RF- DC convertor). The feed line is appropriately positioned at the upper leg of the E-shaped patch antenna as shown in Fig 3. The slot length, width, and position are important parameters in controlling the achievable bandwidth. Traditionally, the property of the patch antenna is suitable for narrow bandwidth applications.The antenna design required to look into the permittivity or dielectric constant of the substrate, width, length of the patch antenna and the ground plane. The permittivity of the substrate plays a major role in the overall performance of the antenna. It affects the width, the characteristic impedance, the length and therefore the resonant frequency that resulting to reduce the transmission efficiency.

IV. DESIGN PROCEDURE OF MICROSTRIP PATCH ANTENNA

In the typical design procedure of Microstrip patch antenna, three essential parameters are,

A. *Frequency of operation (Fr)*

The resonant frequency of the antenna must be selected appropriately. The frequency ranges and the bandwidth requirement for various wireless applications are shown in the table. The resonant frequencies selected for my design is GSM 950MHz for ISM band mobile communication.

B. *Dielectric constant of the substrate* (εr)

The dielectric constant of substrate material plays an important role in the patch antenna design. A substrate with a high dielectric constant reduces the dimensions of the antenna but it also affects the antenna performance. So, there is a trade-off between size and performance of patch antenna. The dielectric constant of the substrate (εr) is typically in the range $2.2 \leq \varepsilon r \leq 12$.

C. *Height of dielectric substrate (h)*

For the Microstrip patch antenna to be used in wireless communication systems, it is essential that the antenna is not bulky. Hence, the height of the dielectric substrate should be less. In our design the height is 1.6mm.

After the proper selection of above three parameters, the next step is to calculate the radiating patch and ground plane width and length.

The width of the antenna is calculated by equation,

$$W = \frac{1}{2fr\sqrt{\mu o \varepsilon o}\sqrt{\frac{\varepsilon r+1}{2}}}$$

$$W = \frac{Co}{2fr\sqrt{\frac{\varepsilon r+1}{2}}} \quad (1)$$

The effective dielectric is calculated by equation,

$$\varepsilon reff = \frac{\varepsilon r+1}{2} + \frac{\varepsilon r-1}{2}\left[1 + 12\left(\frac{h}{W}\right)\right]^{1/2} \quad (2)$$

The effective length is calculated using equation,

$$\text{Lreff}=\frac{c}{2fr\sqrt{\varepsilon reff}} (3)$$

The length extension is calculated using equation,

For, $\frac{W}{h} > 1$,

$$\Delta L = 0.412h \times \frac{(\varepsilon reff \pm 0.3)\left(\frac{W}{h}+0.264\right)}{(\varepsilon reff-0.258)\left(\frac{W}{h}+0.8\right)} (4)$$

The actual length is obtained from the equation,

$$\text{L}=\frac{lambda}{2} - \Delta L$$
$$\text{L} = \text{L}reff - 2\Delta\text{L} \qquad (5)$$

The ground plane dimensions are calculated as,

Length

$$(G_L) = 6h+L \qquad (6)$$

Width

$$(G_W) = 6h+W \qquad (7)$$

TABLE I. Dimensions of Antenna Geometry

Basic Configuration	variable			Dimensions (mm)
Patch antenna	Patch		W	85
			L	106
	Slot	S1	W	79
			L	2
		S2	W	81
			L	3
Pi Matching network	L1		W	2
			L	10
	L2		W	13
			L	8
	L3		W	2
			L	20
Feed Line (F)	W			8
	L			3
	G_W			8

Ground plane		
	G_L	3

The two slots on the antenna are denoted by $S1$ and $S2$. The three components of the pi matching network are represented by $L1$, $L2$ and $L3$ respectively. The feed line is denoted by F. The structure of the E shaped patch antenna and the pi matching network are printed on one side of the FR-4 substrate with the partial ground plane on the other side. The partial ground plane is denoted by $G_L$and $G_W$as shown in Fig3. The design parameters such as dimension (length and width) of patch, slots, pi matching network, the feed line and the partial ground plane are optimized to obtain the best return loss and wide impedance bandwidth. The dimensions of the proposed antenna, pi matching network, feed line and the partial ground plane are as shown in Table 1.

V.MATCHING NETWORK

The matching network used to transform the complex load 63-j117 ohm impedance at the input of the voltage doubler (RF-DC conversion circuit) to the 377 ohm impedance of the antenna. The transmission line calculator is used to obtain microstrip components in the matching network. The line calculator uses Equation (for the characteristic impedance Z_0).

$$\varepsilon eff=\frac{\varepsilon r+1}{2}+\frac{\varepsilon r-1}{2}\left[\frac{1}{\sqrt{1+12\left(\frac{H}{W}\right)}}+0.4\left(1-\frac{W}{H}\right)^2\right] \qquad (8)$$

For, $\frac{W}{H}<1$

$$Z0=\frac{60}{\sqrt{\varepsilon eff}}\ln\left(\frac{8H}{W}+\frac{W}{4H}\right) ohm \qquad (9)$$

VI.ENERGY CONVERSION MODULE (RF-DC)

The energy conversion module is a voltage multiplier circuit used to convert the harvested energy from ambient radio frequency into DC voltage. The design contains stages of Villard voltage multiplier circuit. The function of the energy conversion module is to convert the (RF) signals into direct-current (DC) voltage at the given frequency band to power the low power devices (or) circuits. The design is based on the Villard voltage doubler circuit. A 7-stage Schottky diode voltage doubler circuit is designed, modeled, simulated.

Multisim 2011 was used for the modeling and simulation. For an equivalent incident signal 40dBm, the circuit produced 3.137V across a 100 k ohm. The voltage multiplier circuit in this design uses zero bias Schottky diode 10BQ015. The attractive features of this diode are low forward voltage, low substrate leakage, very fast switching and uses the non-symmetric properties of a diode that allows unidirectional flow of current under ideal condition but leads to a fabrication overhead.

Each independent stage with its dedicated voltage doubler circuit can be seen as a single battery with open circuit output voltage $V0$, internal resistance $R0$. With load resistance RL, the output voltage, $Vout$is expressed as in.

$$V_{out}=\frac{Vo}{RO+RL}RL \qquad (10)$$

When n number of these circuits are put in series and connected to a load of RL in Equation the output voltage $Vout$obtained is given by

$$V_{out}=\frac{nVo}{nRO+RL}RL \qquad (11)$$

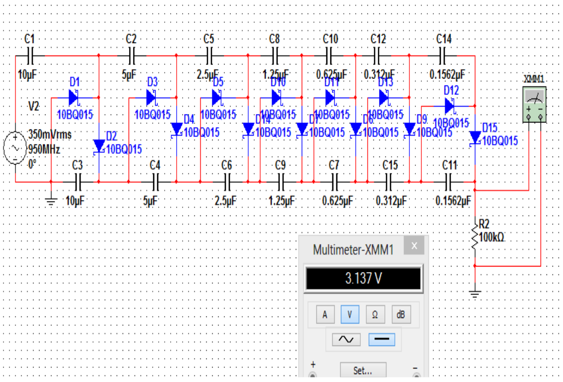

Fig 4.Seven stage Voltage Multiplier

In the above Fig 4, capacitance value was varied in such a way that, from one stage to the next, it was halved. For example, if the first stage was 3.3 nF, the second stage was 1.65 nF, third stage was 825 pF, fourth stage was 415 pF and so on. But keeping in view of testing, the capacitance values were chosen to have a close match with the standard available values in the market.

The number of stages in the system has the greatest effect on the output voltage. The capacitors, both in the stages and at the final stage of the circuit, affect the speed of the transient response and the stability of the output signal. The number of stages is essentially directly proportional to the amount of voltage obtained at the output of the system but the current rating is decreased directly proportional to the number of stages.

VII.RESULT ANALYSIS

The properties and performance of the proposed antenna have been predicted and optimized through electromagnetic simulation software in CST microwave studio 2014.

The results of return loss are shown in Fig 5. It is observed that the structure is resonating at 950MHz with return loss of 34.2dB giving an impedance bandwidth of 215MHz.We suggest that there is good matching at the frequency points below the -10 dB region. The maximum achievable gain over the entire frequency band of and for 950MHz is 3.3dB.

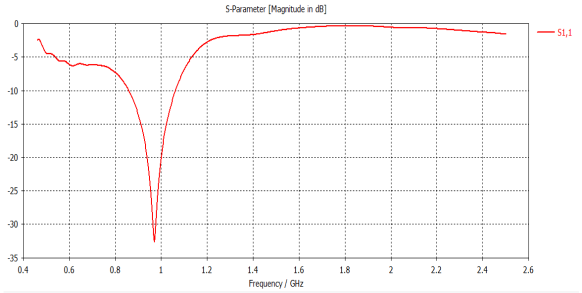

Fig 5. Graph shows the Return Loss of the antenna

The achieved antenna impedance is 50 ohm which is exact to the required impedance of 50 ohm. The VSWR ratio at 950MHz frequency is 1.127 is shown in Fid 6. It should lie in between 1 and 2. The optimum feed positions have been determined for good impedance matching because the input impedance is controlled by the position of the feed to patch connection point. The used microstrip feed line is designed for microstrip patch antenna made of impedance 50 ohm.

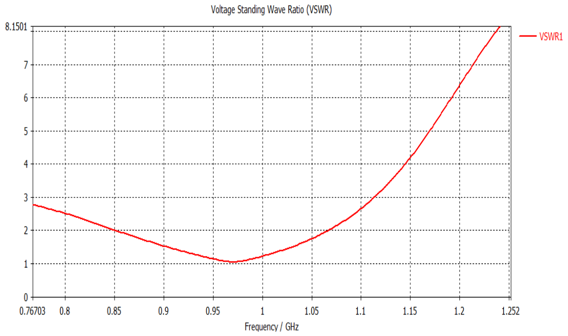

Fig 6. Graph displays the VSWR value 1.127

The fig.7 shows distribution of current through the antenna. The maximum value of this design is 32.78dB.Both electric and magnetic field can be analyzed with the help of current distribution. Depending upon the distribution of current inside the antenna, the radiation efficiency will increase.

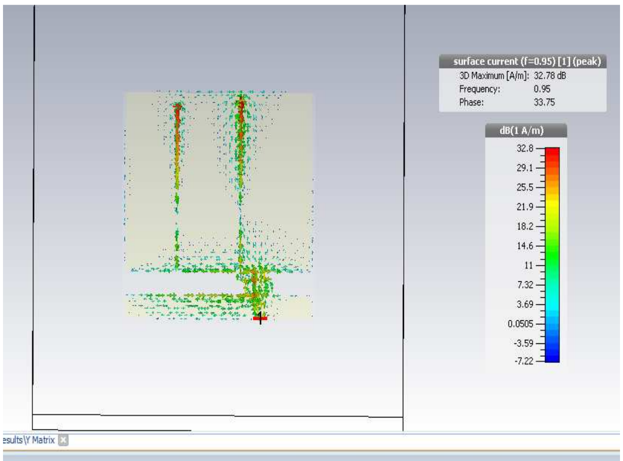
surface current (f=0.95) [1] (peak)
3D Maximum [A/m]: 32.78 dB
Frequency: 0.95
Phase: 33.75

Fig 7.Current Distribution

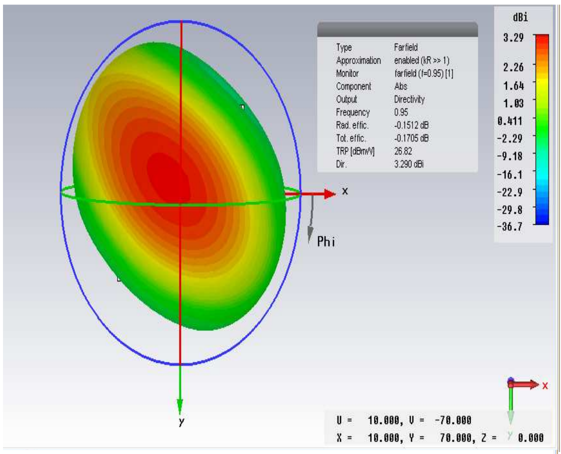
Type Farfield
Approximation enabled (kR >> 1)
Monitor farfield (f=0.95) [1]
Component Abs
Output Directivity
Frequency 0.95
Rad. effic. -0.1512 dB
Tot. effic. -0.1705 dB
TRP [dBmW] 26.82
Dir. 3.290 dBi

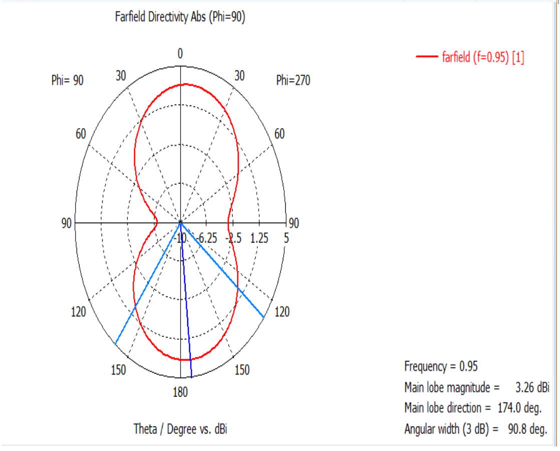

Frequency = 0.95
Main lobe magnitude = 3.26 dBi
Main lobe direction = 174.0 deg.
Angular width (3 dB) = 90.8 deg.

Fig 7.Farfield directivity

The farfield 3D graph shows the directivity of the antenna, the main lobe magnitude is 3.26dBi with frequency 950 MHz.Maximum value of signal is observed from the angle 180 degree and angular width 90.8 deg. It is more enough to receive the ambient RF signals.

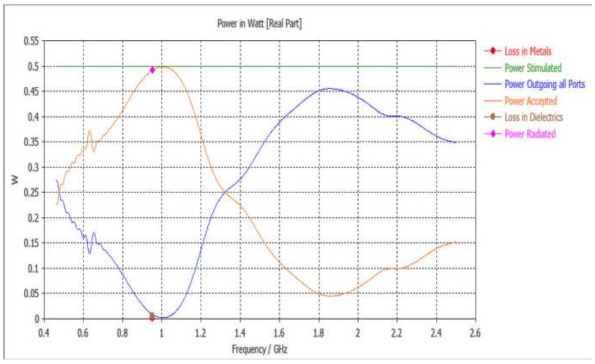

Fig 8. Power Excitation

The above graph shows total simulated power(0.5 W), outgoing power from the ports, level of power accepted to and radiated from the antenna

VII.CONCLUSION

A novel 950MHz RF energy harvesting system for powering low powersensors has been analyzed, discussed, designed, and stimulated. The novelty lies in the partial ground plane in the antenna structure which resulted in maximizing the energy captured and generating a higherDC output voltage that can power low power devices. Firstly, a 377 ohm E-shaped patch antenna with partial ground plane withcompacted size was designed, simulated.

Compared to the established performance characteristics of wideband micro strip 50 ohm patch antennas, the antenna with the partial groundplane was successes in achieving a wide impedance bandwidth formatching to a free space impedance 377ohm subsequently the pi matching network located in betweenantenna and the RF-DC conversion module is designed simulated andimplemented to provide a good match from the load to the source.Lastly the energy conversion module that comprises of 7-stages voltage multiplier circuit with zero bias Schottky diodes was successfullydesigned, implemented, tested and found to be efficient in convertingthe RF signals captured by the antenna to the required DC output voltage harvested.

We have used pure copper (annealed)as it is a low budget project. The output obtained using copper is discussed in our paper, if we have used either Gold or Metamaterial, the output receiving power capacity would be improved and the design become efficient.

ACKNOWLEDGMENT

We would like to acknowledge and thank our college faculties members Mrs.N.padmaavathiM.Tech, Mr. R.RameshBabu M.E,Mrs.M.RathiM.Tech, they have guided us throughout the project and enhance our knowledge by motivating us.

REFERENCES

1. Le, T. T., “Efficient power conversion interface circuits for energy harvesting applications,” PhD Thesis, Oregon State University, USA, Jun. 2008.

2. Asefi, M., S. H. Nasab, L. Albasha, and N. Qaddoumi, “Energizing low power circuits by using an RF signal harvester," *16th Telecommunications Forum TELFOR*, Nov. 2008.

3. Arrawatia, M., M. S. Baghini, and G. Kumar, “RF energy harvesting system from cell towers in 900MHz band,” National Conference on Communication (NCC) 2011
.

4. Devi, K. K. A., S. Sadasivam, N. M. Din, C. K. Chakrabarthy, and S. K. Rajib, “Design of a wideband 377 ohm E-shaped patch antenna for RF energy harvesting,” *Microwave and Optical Technolog Letters*, Vol. 54, No. 3, 569, Mar. 2012.

Yuvarajperusing B.E. graduation in Electronics and Communication Engineering at Srinivasa Institute of Engineering and Technology (SIET), Poonamallee. I am fond of electronics and I would like to achieve something great in this field.

Ashwin Samuel perusing B.E. graduation in Electronics and Communication Engineering at Srinivasa Institute of Engineering and Technology(SIET), Poonamallee. I am interested towards new innovation.

March 25, 1969 W. C. BROWN ET AL 3,434,678

MICROWAVE TO DC CONVERTER

Filed May 5, 1965

Sheet 1 of 3

FIG. 1

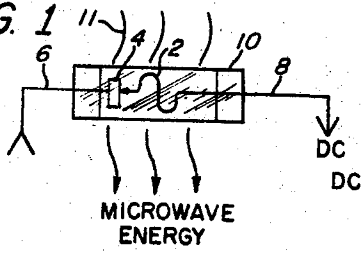
11
4
2
10
6
8
DC
MICROWAVE
ENERGY

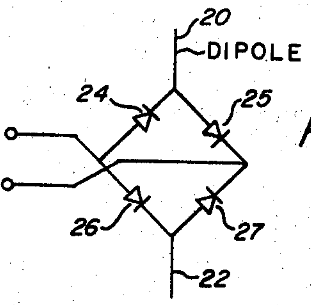
20
DIPOLE
24
25
DC LOAD
26
27
22

FIG. 2

FIG. 3

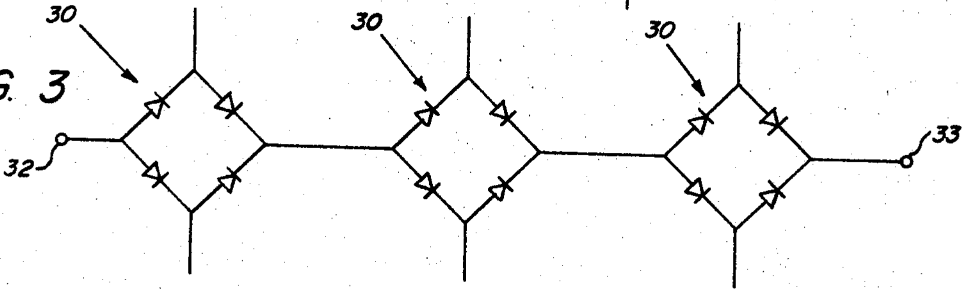
30
30
30
32
33

FIG. 4

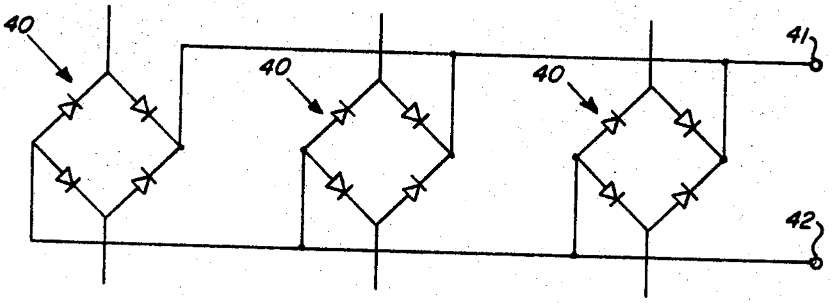
40
40
40
41
42

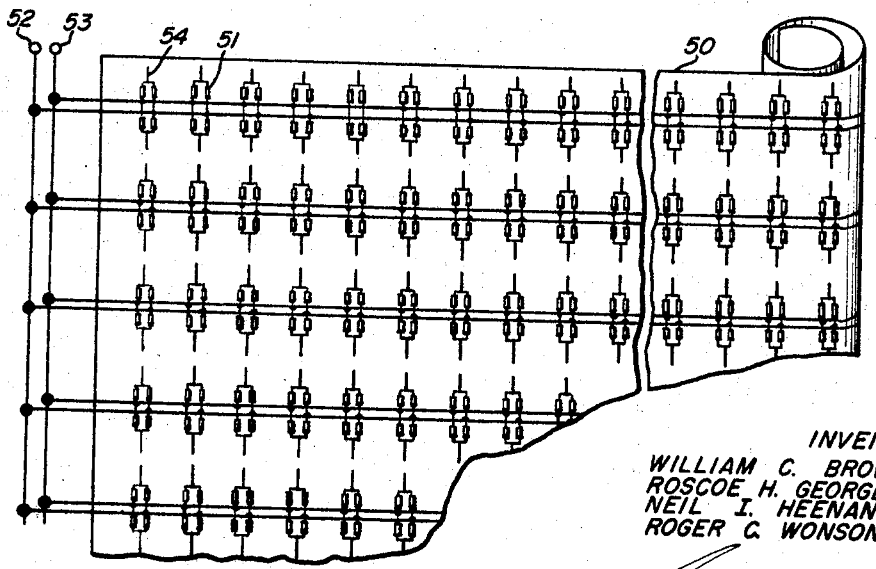
52
53
54
51
50

FIG. 5

INVENTORS
WILLIAM C. BROWN
ROSCOE H. GEORGE
NEIL I. HEENAN
ROGER C. WONSON

BY Edgar O. Rost

ATTORNEY

March 25, 1969 W. C. BROWN ET AL 3,434,678

MICROWAVE TO DC CONVERTER

Filed May 5, 1965 Sheet 2 of 3

FIG. 7

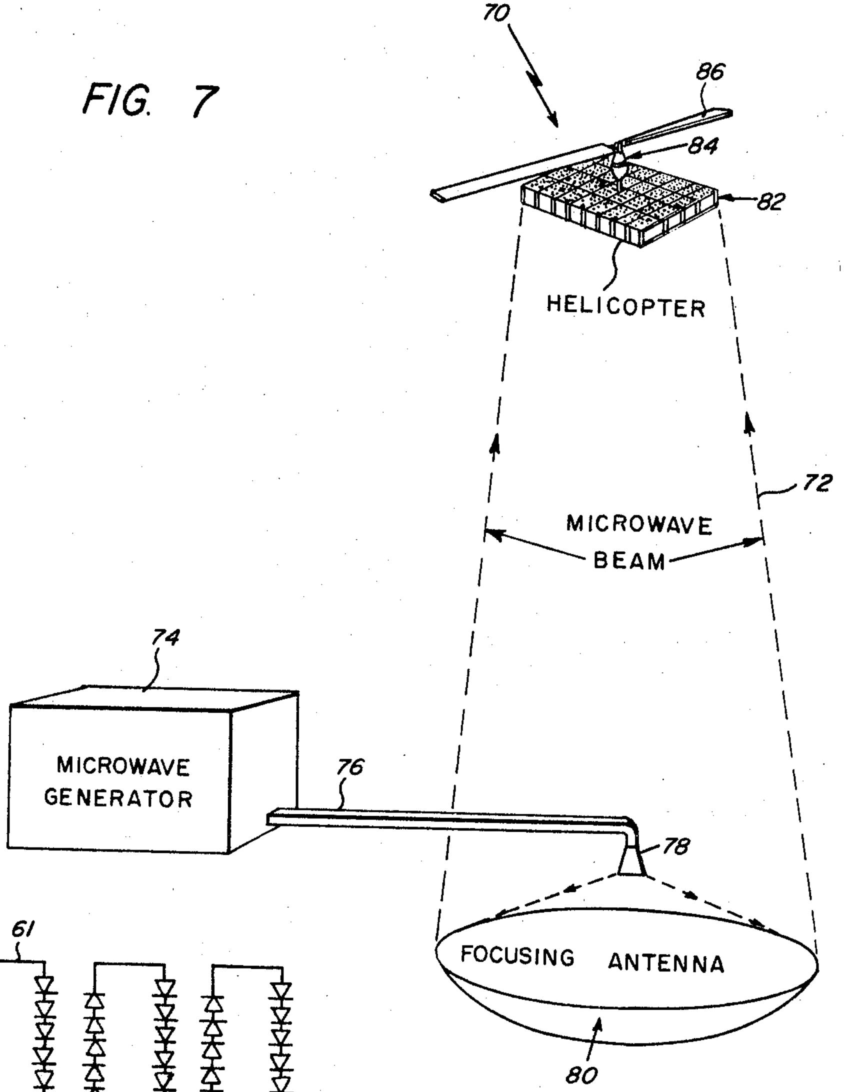
70
86
84
82
HELICOPTER
72
MICROWAVE
BEAM
74
MICROWAVE
GENERATOR
76
78
FOCUSING ANTENNA
80

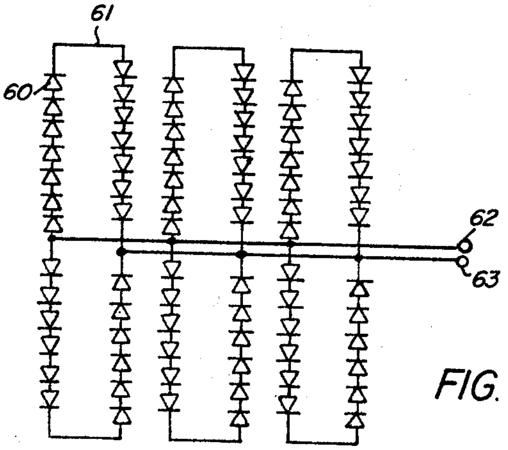

FIG. 6

INVENTORS
WILLIAM C. BROWN
ROSCOE H. GEORGE
NEIL I. HEENAN
ROGER C. WONSON

BY Edgar O. Rost

ATTORNEY

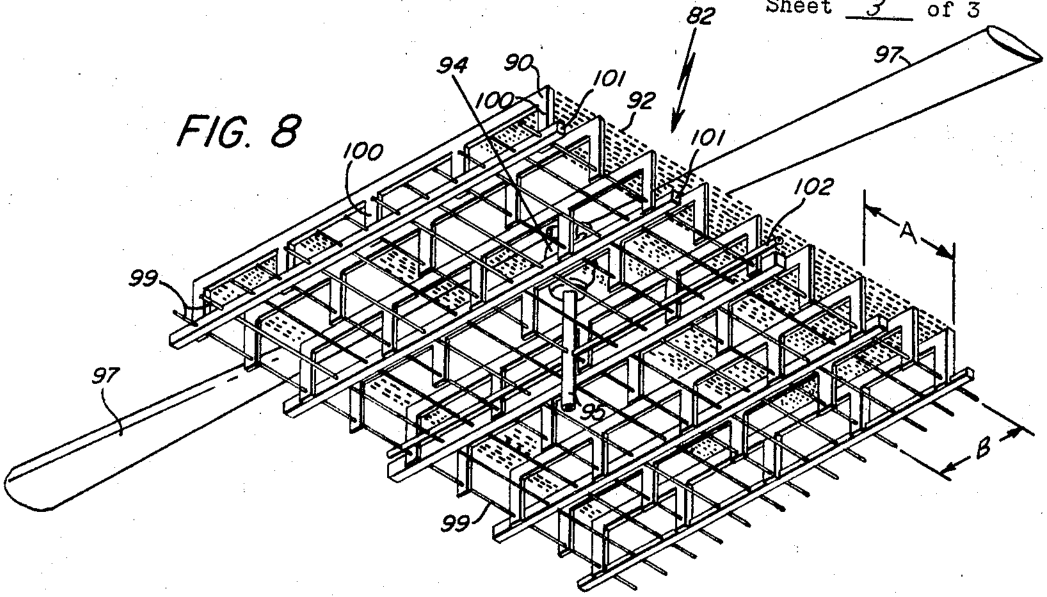
82
97
94
90
101
100
92
101
100
102
A
99
97
95
B
99

FIG. 8

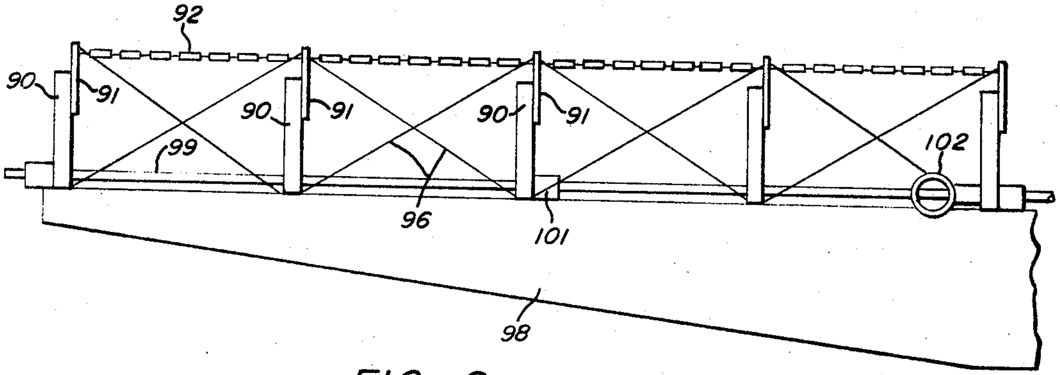
92
90
91
90
91
90
91
99
102
96
101
98

FIG. 9

INVENTORS
WILLIAM C. BROWN
ROSCOE H. GEORGE
NEIL I. HEENAN
ROGER C. WONSON

BY Edgar O. Rost
ATTORNEY

United States Patent Office
3,434,678
Patented Mar. 25, 1969

3,434,678

MICROWAVE TO DC CONVERTER

William C. Brown, Weston, Mass., Roscoe H. George, West Lafayette, Ind., and Neil I. Heenan, Needham, and Roger C. Wonson, Beverly, Mass., assignors to Raytheon Company, Lexington, Mass., a corporation of Delaware

Filed May 5, 1965, Ser. No. 453,415
Int. Cl. B64g *1/00*
U.S. Cl. 244—1
3 Claims

ABSTRACT OF THE DISCLOSURE

A combined antenna and conversion mechanism for reception of beamed high frequency electromagnetic energy in space including a large array of unidirectional current semiconductor rectifier devices. A self-supporting space vehicle utilizing the rectified DC electrical energy for propulsion is disclosed in an illustrative embodiment.

The present invention relates in general to the transfer of energy by means of an electromagnetic wave beam and more particularly to interception and rectification of such energy into low frequency electrical DC energy with a high degree of efficiency.

Improved technology in the field of microwave energy generation at superpower levels has resulted in the realization of electrical energy transmission over considerable distances for remote energization of devices or vehicles without the aid of wires. The transmission of microwave electromagnetic energy into space has been commonly employed in the radar pulse echo systems for the detection and orientation of desired objects within a predetermined scanning range of a transmitting antenna. Beams of a similar nature may now be employed for other useful purposes and the advantages attendant the utilization of electromagnetic energy in the microwave region in contrast with other wavelengths may now be enumerated.

Microwaves have been generally defined as high frequency radio waves whose wavelength is less than 30 centimeters, with a lower wavelength limit on the order of 1 millimeter sometimes being applied to what is commonly referred to as the "microwave region." The superiority of high frequency microwaves is due in part to the fact that it is generally desirable to focus the transmitted energy so as to achieve a high power density at a remote point or area with respect to a given power source. In accordance with the laws of optics, the sharpness of the microwave beam produced by a transmitting antenna varies as the ratio of antenna dimensions to the wavelength of the transmitted energy. Therefore, for a given or desired power density or beam sharpness, a decrease in the wavelength of the transmitted energy permits a corresponding decrease in the dimensions of the antenna. From the standpoint of mechanical considerations, it is desirable to employ small antennas and other components, and it is therefore advantageous to employ high frequency energy of very short wavelength. In addition, the difficulties encountered in long wave transmission as a result of natural and man-made interference or noise do not occur with any appreciable significance at microwave frequencies. Further, in aerospace applications with considerable distances separating the transmitter at an earth or mother planet location and the employment of shorter wavelength beamed energy is preferred since longer wave signals will generallly be reflected at certain altitudes by reflecting layers in the atmosphere.

In view of certain losses due to absorption which may occur in the atmosphere, microwaves in the region having the approximate bounds of 2 and 30 centimeters are

readily adaptable to the convenient radiation of power to remote points without the utilization of wires. The preferred wavelengths are of the order of 5 or 10 centimeters to provide efficient focusing with existing transmitting antenna systems which may be maintained at a reasonable size. An illustrative device of the superpower high frequency microwave generators operative in the desired band is the so-called Amplitron which is an amplifier having a broad bandwidth and excellent performance characteristics for the focusing of the beam. Such devices are capable of producing 15 or 20 kilowatts of average continuous wave power in the neighborhood of 10 centimeters in wavelength with capabilities expected in the region of 500 kilowatts or more average power with 50 megawatts peak power. A complete description of such devices may be had by referring to Patent No. 2,933,723 issued Apr. 19, 1960 to William C. Brown and assigned to the assignee of the present invention.

With microwave energy capable of being generated and directed over longer distances conversion of such high frequency electromagnetic energy is of paramount concern. One conversion mechanism in the prior art involves direct conversion of such energy into heat which may then be utilized directly or indirectly for propulsion or generation of flight-producing forces. Examples of such devices for heat energy exchange as well as space vehicles utilizing such energy may be noted in Patent No. 3,174,-705, issued Mar. 23, 1965, to D. Schiff et al., as well as U.S. Letters Patent No. 3,083,528, issued Apr. 2, 1963 and No. 3,114,517, issued Dec. 17, 1963, to William C. Brown. The heat exchanger method of conversion of electromagnetic energy into useful power is limited by the overall efficiencies of approximately 25 percent in the conversion of heat into mechanical or electrical work. Desirable, therefore, would be the direct rectification of the high frequency electromagnetic energy into low frequency electrical energy for the operation of many useful aerospace devices as well as systems.

The present invention has for its primary object the conversion of high frequency electromagnetic energy in the microwave region directly into low frequency electrical energy.

A further object of the present invention is the provision of a combined nondirectional receiving antenna and microwave electromagnetic energy to low frequency electrical energy conversion means in a unitary structure.

A still further object of the present invention is a provision of a new and novel combined nondirectional receiving antenna and microwave to DC energy converter for aerospace applications.

Another object of the present invention is the provision of a new and novel nondirectional receiving antenna and microwave to DC energy converter having a high degree of efficiency.

Still another object of the present invention is the provision of a new and novel aerospace vehicle with nondirectional receiving antenna and microwave to DC energy converter means with said vehicle being capable of being supported by its own energy generation means at a distance spaced apart from the power generation means.

In accordance with the teachings of the present invention, the above and other objects are achieved by the employment of efficient unidirectional microwave power rectifiers and dipole antenna means. Such rectifying devices, while being individually limited in power-handling capabilities, normally in the order of fractions of watts, have been found to be highly efficient means for the rectification of microwave power when assembled in large numbers in various arrays. It is interesting to note that the observed collective efficiency was on the order of 40 to 70 percent. In an illustrative embodiment, point-

contact semiconductor diodes were arranged in four arm bridge connected networks with the networks interconnected in various configurations such as series, parallel and series-parallel.

In discussing aerospace applications, an additional problem is encountered in the beaming of microwave energy to a remote point and the interception and utilization of such electrical energy. In such applications the advantages of a vehicle which may be maintained in space for indeterminate periods of time without employing a local fuel source are readily apparent. Such devices could readily provide communication networks, surveillance functions using radar techniques along with numerous other functions. The acapture of the beamed high frequency electromagnetic energy raises the need for an efficient antenna means capable of intersecting the beam at high altitudes. Conventional techniques employed in microwave radar usage such as receiving antenna horns are capable of intersecting only a small portion of the beam energy and add considerable weight in applications involving heavier-than-air vehicles. In an exemplary embodiment of the invention a space vehicle, namely a helicopter, is disclosed for either moving flight or a stationary location with self-supporting electrically operative propulsion means. The semi-conductor diode rectifier arrays have been demonstrated to fulfill the receiving antenna functions as well as the electrical energy rectification means in a highly efficient manner. Such combined antenna and rectifier means has also assisted in reduction of the weight problem in airborne devices. Further, it has provided a nondirectional means for the interception of the microwave energy to thereby reduce the problems of focusing inherent in prior art directional horn type receiving antennas.

With the above features, advantages and objects in mind the invention will now be described by reference to the following detailed description together with the accompanying drawings in which:

FIG. 1 is a perspective view of an illustrative diode rectifier;

FIG. 2 is a schematic circuit diagram of a bridge connected diode network with dipole antenna means;

FIG. 3 is a schematic circuit diagram of a plurality of bridge connected networks arranged in series;

FIG. 4 is a schematic circuit diagram of a parallel bridge connected network array;

FIG. 5 is a perspective view of an illustrative embodiment of a combined antenna and rectifier array in a folded or rolled up configuration;

FIG. 6 is a schematic circuit diagram illustrating the bridge connected diode array incorporated in the aerospace vehicle shown in FIG. 7;

FIG. 7 is a schematic representation in elevation illustrative of a heavier-than-air aerospace vehicle incorparating the structure of the present invention;

FIG. 8 is a perspective view of the aerospace vehicle embodiment as viewed from the under portion thereof; and

FIG. 9 is an enlarged partial view in elevation of a portion of the illustrative embodiment shown in FIG. 8.

FIG. 1 illustrates a point-contact semiconductor diode rectifier of the type employed in radar microwave receiver apparatus to rectify returned radar pulses. Any of the high burnout semiconductor diodes having high rectification characteristics are preferred and are commercially available, such as the 1N82 or 1N830. The rectifying junction is formed by whisker element **2** contacting the semiconductor element **4** respectively connected to leads **6** and **8**. Silicon is preferred over germanium for element **4** because of its ability to operate at higher temperatures and thereby handle higher powers. Envelope **10** houses the rectifying elements and may be of a hermetically sealed dielectric material or combination metal and ceramic composition. The inherent characteristic of such diode rectifiers is that the microwave energy is intercepted and rectified in a unidirectional manner and the line **11** indicate pictorially the rays of the beamed electromagnetic microwave energy in a plane normal to the envelope. In FIG. 2 a full-wave bridge connected diode network is illustrated with the forward direction of the rectified DC electric current indicated by the direction of the arrow symbols. The network shown consists of half-wave dipoles **20** and **22** each terminated with a diode rectifier element **24** to **27** in an arm of the bridge connected network. The dipole elements **20** and **22** are of the half-wave configuration and may be spaced apart from each other a one-half wavelength at the frequency of the beamed electromagnetic energy.

Referring now to FIG. 3, an array of bridge connected diode networks each with the half-wave dipoles are shown connected in series. Each network is referred to by the numeral **30** and is similar in the bridge connections to the single element network shown in FIG. 2. The DC output of the collective rectified energy is coupled by means of terminals **32** and **33**. In FIG. 4, a similar number of individual bridge connected diode-dipole networks are shown connected in a parallel array. Each network is indicated by the numeral **40**, and the output terminals are indicated as **41** and **42**.

Any number of diode-dipole networks may be provided and in FIG. 5 such a multi-element array is illustrated by mounting on a flexible material **50** which may be rolled or folded into any desired package or enclosed within a capsule to be launched and released at a predetermined point in space. Any flexible material which is pervious to electromagnetic energy is preferred. The total power desired would be the determining factor in a number of individual diode-dipole elements required. In this embodiment, the bridge connected networks **51** are connected in parallel to the output load indicated by terminals **52** and **53**, and representative measurements of electrical characteristics have shown that approximately five watts of DC electrical energy is realizable for each square foot of area of the combined antenna-rectifier. While the dipole elements **54** have been indicated in a particular array, it is within the scope of the invention to stagger the placement of such dipoles to increase the over-all efficiency of the antenna-rectifier.

To further increase the DC powder output, the full-wave bridge connected networks are preferably arranged with a plurality of diodes in series in each arm of the bridge. An illustrative schematic circuit diagram of such a configuration is shown in FIG. 6 wherein seven diodes **60** are shown in each arm of the bridge circuit and are connected in series for a total of twenty-eight diodes in each bridge network. The dipole members will then be the substantially U-shaped end portions **61** at the ends of each brace of seven diodes. In the illustration three such twenty-eight diode bridge networks are shown connected in parallel to terminals **62** and **63**. This closer spacing and compact arrangement has been shown to be a source of improved power output and is capable of a high degree of reliability through the redundant nature of the parallel series connections within each bridge network. If one of the diode rectifiers fails to function the over-all voltage drop across this element would be divided among the six remaining diode rectifiers. If any of the connecting wires between the diode elements should break, the adjacent arms of the other bridge assemblies would take the additional load due to the close proximity of the respective arms to each other. In addition, it is possible to have a number of open connections or inoperative diodes dispersed throughout the array without any serious impairment in performance.

In relation to the array concept to be hereinafter described it may be stated that within a six inch square area ten such individual bridge networks each containing twenty-eight diode rectifiers for a total of 280 diode rectifiers may be deployed in such a manner as to provide

maximum exposed area for each diode as well as the connecting leads. Such an arrangement will be hereinafter referred to as a "module" and a DC output in excess of fourteen watts has been measured for such a module. Any number of such modules could be connected provided for a desired power yield and this module concept readily lends itself to use in certain aerospace applications now to be described.

In FIG. 7 a propelled type of space vehicle **70** is shown wholly supported by means of the transfer and rectification of continuous wave electromagnetic energy via a microwave beam **72**. The source of the microwave energy which may be of the Amplitron type device as described in the aforementioned issued Patent No. 2,933,723 is indicated as **74**. This energy is fed by waveguide means **76** to a transmitting horn **78** to illuminate an ellipsoidal beam forming focusing antenna **80** for the transmission of the microwave beam **72**. It will be appreciated by those skilled in the art that the representations of the microwave generation and transmitting antenna means are pictorial representations to illustrate the usage of the invention in diagrammatic form and the present invention is not limited to any particular source of microwave energy or transmitting antenna assembly. It may be stated the reflector of the antenna assembly is considerably larger than most of the reflectors of the prior art in order to focus a large amount of the microwave power at high altitudes for use in the transfer of energy to space vehicles. Such antenna assemblies may be partially supported in a large hollowed area on the earth's surface or other convenient means of support.

The space vehicle or helicopter **70** can be described as a main body member supporting antenna-rectifier means **82** including a large number of the so-called modules connected together and rigidly supported in a planar parallel array. A motor **84** is supported by the combined body member and the receiving antenna-rectifier means and actuates the rotor **86** of conventional design employed in such self-propelled hovering vehicles. The disclosed vehicle provides for the illumination of the planar array of the semiconductor diode dipole elements by the microwave beam and the direct conversion of the microwave power transmitted by the beam into usable electrical energy for the self-propulsion of the device without any local fuel supply being required.

FIGS. 8 and 9 illustrate a space vehicle **82** comprising a plurality of the combined receiving antenna-rectifier module means for interception and rectification of the electromagnetic microwave energy beam emanating from an earth or mother planet source. A planar array of the antenna-rectifier modules is mechanically supported by means of structural members **90** of any lightweight wood or metal. Insulators **91** positioned coextensive with the members **90** support the diode rectifier array and avoid interference with the receiving and electrical performance characteristics by the structural support members. Carrying forward the module concept of 280 diode rectifiers to provide an approximate power output of 14 watts, it was noted that any number of such modules may be coupled together since the individual module outputs are relatively insensitive to a wide range of load resistances connected to the common output terminals. To achieve the desired electrical output of approximately 120 volts and 250 watts of power, subgroups of four modules each were assembled and parallel connected with an approximate 30 volts available for each subgroup. Four such subgroups were series-connected to result in a total of 4,480 diode rectifiers or 16 modules assembled in a two foot square self-supporting planar parallel array structure. The individual diode rectifiers connected in each arm of the bridge network are indicated by the numeral **92**. An exemplary module configuration would extend within the area delineated by the dotted lines and reference letter A on one side and similar dotted lines and reference letter B on the other side.

A motor **94** is connected to the DC side of the overall array and may be additionally supported by tubular member **95**. A shaft and propulsion means consisting of rotor blades **97** provide for the upward lift of the overall vehicle for the self-supporting of same in space applications. Additional structural support such as interlaced rigging **96** of a high tensile strength material such as nylon or steel wire, as well as bracing member **98**, may be employed for strengthening of the body means to withstand the vibrational forces and downwash from the propulsion means.

In accordance with the well known technology of microwave transmission the combined array of diode rectifiers and propulsion means presents a specific load impedance which must be suitably matched to the transmitted microwave energy beam to result in maximum efficiency. In aerospace applications a mismatch of approximately ten to one may be evident. Matching of the load impedance to a value of approximately 377 ohms as the free space value will be provided by a plurality of coplanar parallel metallic rod members **99** disposed in a grating array in front of the diode rectifiers a predetermined distance. Rod members **99** are linearly disposed and extend in a similar direction as the assembled diode rectifiers. A selected frontal spacing of one-quarter of the wavelength of the microwave frequency being transmitted has been experimentally determined to be suitable for impedance matching purposes. An approximate spacing of two inches between the respective members was preferred for a selected microwave frequency of 2,450 megacycles. Each of the members **90** are provided with lateral sections **100** to support the elongated bar members **101** which in turn maintain the rod members **99** in the desired position. A tubular member **102** of a lightweight metal may also be provided to combine with the motor support member **95** for structural support.

The combined antenna-rectifier array provides a source of electrical energy to render any space vehicle self-supporting. The diode rectifier elements when assembled in the antenna array have been found to be nondirectional with respect to interception of the beamed microwave energy. This represents a large step forward in the utilization of high power microwave energy over the prior art horn-type receiving antennas which must be accurately focused and pointed in a particular direction for the reception of any energy. The connections between the respective members of the diode rectifier array and deployment in the parallel configuration serves to provide maximum exposed area. Such connections and in particular the end loop portions adjacent the terminus of each arm of the bridge networks serves as an efficient dipole for the interception of the microwave energy.

Although it is not intended as a full explanation of the high degree of efficiency attained with the disclosed antenna-rectifier array, it is believed that the whisker elements within the semiconductor diodes themselves are a contributing factor and may function as additional dipole elements. The disclosed embodiment functioned efficiently when illuminated by microwave energy generating a vertically polarized beam. Hence, an efficient and light weight energy conversion apparatus is disclosed which may be self-supporting without the requirement of a large local fuel supply payload.

It may be within the purview of the invention to use the available rectified electrical energy for performing many functions in addition to the actuation of the propulsion means. Hence, communications' payloads may be maintained at predetermined positions in space in a hovering attitude utilizing a portion of the electrical energy available. Relay signals to other such vehicles or return signals to ground stations would then be within the realm of possibility. Such available energy may also be employed for servomechanisms, stabilizing and countertorque systems for the navigation of such vehicles.

The electrical efficiencies realized with the combined

receiving antenna and rectifier means have also provided certain weight advantages over other energy converters in aerospace applications. Examples of such converters would be heat exchangers or solar cells. In comparison to the present invention where five to eight pounds per kilowatt of energy realized is a normal characteristic, other energy conversion means weigh in the vicinity of 150 pounds per kilowatt of realizable energy. The inherent advantages of the present invention are therefore apparent. While the technology in the diode rectifier art is being continually advanced, new diode power rectifiers as well as integrated circuit techniques are readily available to future configurations of the present invention. The so-called Schottky barrier diodes could be employed to produce combined antenna-rectifier means weighing even less than two pounds per kilowatt of available energy.

Although the foregoing detailed description has referred to DC power rectification it will be evident that with suitable circuit components low frequency AC energy may also be made available. In addition, other propulsion means may be readily substituted using electrical energy. The embodiments disclosed herein are illustrative only and other modifications or alterations will be apparent to those skilled in the art which do not depart from the scope of the broadest aspects of the present invention as defined in the appended claims.

What is claimed is:

1. A space vehicle comprising:
body means;
said body means including spaced structural support members;
combined antenna and DC electrical energy rectification means for the interception and rectification of incident high frequency electromagnetic microwave energy carried by said support members in a planar parallel array;
said rectification means comprising a plurality of four arm full-wave bridge connected rectifier circuit networks each having a plurality of unidirectional semiconductors in each arm;
said networks being electrically interconnected to common output terminals;
electrically operable propulsion means comprising a motor and rotor members carried by said body means and connected to said terminals for the utilization of said rectified DC energy; and
means for matching the load impedance of said combined antenna and electrical energy rectification means to the incident microwave energy.

2. A space vehicle according to claim 1 wherein said load impedance matching means are arranged in a coplanar array coextensive with said antenna and energy rectification means array, and spaced therefrom a distance of approximately one-quarter of a wavelength at the frequency of the microwave energy.

3. A space vehicle according to claim 2 wherein said load impedance matching means comprise a plurality of parallel disposed elongated metallic members.

References Cited

UNITED STATES PATENTS

1,217,149	2/1917	Caldwell ----------	321—27 X
2,165,055	7/1939	Kafka --------------	321—27
2,444,458	7/1948	Master ---------------	321—8
2,927,321	3/1960	Harris -------------	343—6.8 X
3,098,971	7/1963	Richardson --------	325—592 X
3,174,705	3/1965	Schiff et al. ----------	244—1

RODNEY D. BENNETT, *Primary Examiner.*

MALCOMB F. HUBLER, *Assistant Examiner.*

U.S. Cl. X.R.

307—151; 318—16; 321—27; 325—494; 343—100

United States Patent [19]
Phillips, Sr.

[11] **Patent Number: 4,685,047**
[45] **Date of Patent: Aug. 4, 1987**

[54] APPARATUS FOR CONVERTING RADIO FREQUENCY ENERGY TO DIRECT CURRENT

[76] Inventor: **Raymond P. Phillips, Sr.,** 1200 S. Frances, Terrell, Tex. 75160

[21] Appl. No.: **886,138**

[22] Filed: **Jul. 16, 1986**

[51] **Int. Cl.**⁴ **H02M 7/06**
[52] **U.S. Cl.** **363/126**; 307/151
[58] **Field of Search** 363/125, 126; 307/151

[56] References Cited

U.S. PATENT DOCUMENTS

649,621	11/1901	Tesla	455/39
685,957	5/1902	Tesla	455/619
3,434,678	3/1969	Brown et al.	307/151
3,445,767	5/1969	Beard	324/78
3,678,365	8/1972	Fremiot	307/151
3,723,891	3/1973	Whitely	307/246
3,760,257	9/1973	Fletcher et al.	307/151
3,784,845	1/1974	Haas	307/519
3,792,290	2/1974	Brocker	307/522
3,944,935	3/1976	Plant	307/519
4,079,268	3/1978	Fletcher et al.	307/151
4,079,325	3/1978	Mawhinney et al.	307/261
4,085,377	4/1978	Turski et al.	307/519
4,195,237	3/1980	Meslener	307/261
4,297,557	10/1981	Tyler et al.	219/10.55 B
4,360,741	11/1982	Fitzsimmons et al.	307/151
4,360,865	11/1982	Yasumura et al.	363/126
4,393,441	8/1983	Enge	363/126

Primary Examiner—Patrick R. Salce
Assistant Examiner—Jeffrey Sterrett
Attorney, Agent, or Firm—Millen & White

[57] ABSTRACT

Apparatus for and methods of converting radio frequency energy into direct current for generating electric power includes a dipolar antenna for receiving radio frequency energy and a circuit connected thereto for converting the radio frequency energy to direct current. The circuit has a positive output line connected to one pole of the antenna and a negative output line connected to the other pole of the antenna. A positive transmitting diode is in the positive output line and a negative transmitting diode is in the negative output line. First and second bus lines and a pair of tuned circuits of opposite polarity couple the positive output line and negative line to the bus line with one of the bus lines being connected to ground. Each tuned circuit includes a first bridging line connecting the positive output line to the first and second ground lines and a second bridging line connecting the negative output line to the first and second ground lines. Each bridging line has a diode therein oriented at a polarity which is reverse with respect to the input diode. The bridging lines of each tuned circuit are connected to one another by an inductance and have capacitors disposed between the diode and the bus lines. A direct current device is connected to the positive line of the circuit.

11 Claims, 1 Drawing Figure

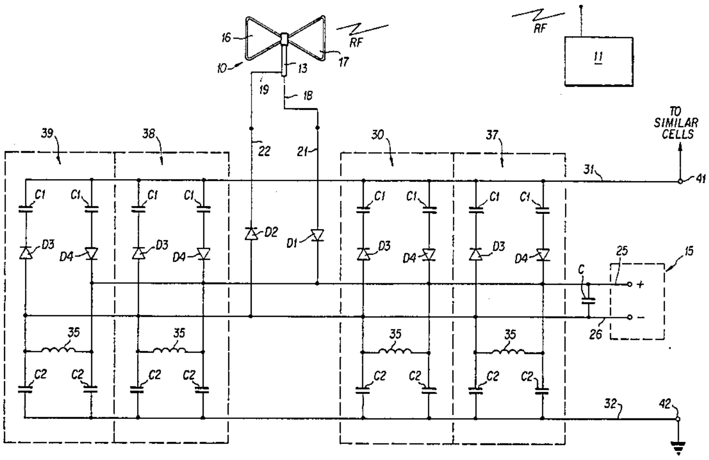

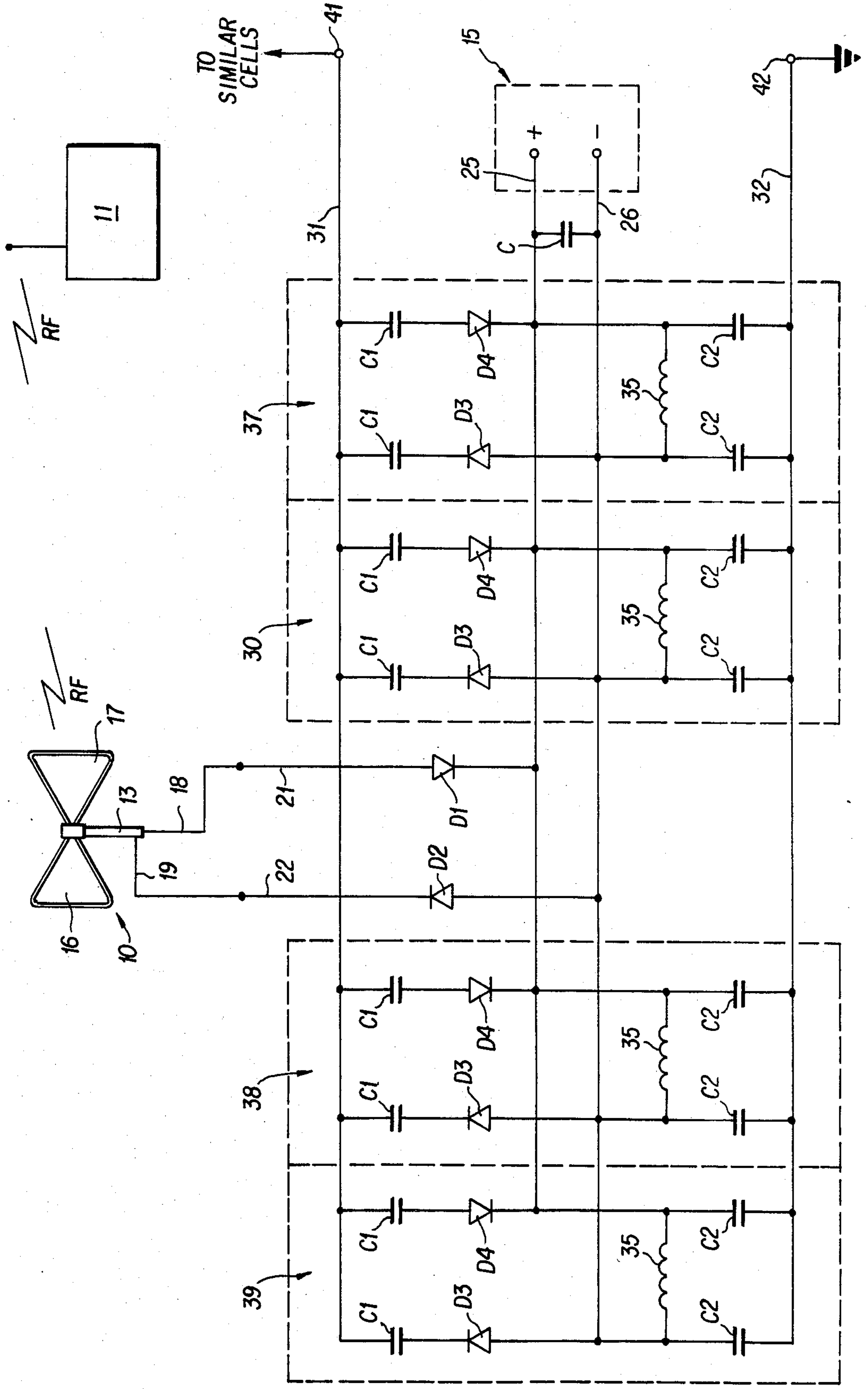
11
RF
TO SIMILAR CELLS
41
15
+
-
25
26
C
31
32
42
C1
D4
C2
C1
D3
C2
35
37
C1
D4
C2
35
30
C1
D3
C2
RF
17
13
18
21
D1
19
22
D2
16
10
C1
D4
C2
35
38
C1
D3
C2
C1
D4
C2
35
39
C1
D3
C2

APPARATUS FOR CONVERTING RADIO FREQUENCY ENERGY TO DIRECT CURRENT

BACKGROUND OF THE INVENTION

1. Field of the Invention

The instant invention relates to methods of an apparatus for converting radio frequency energy to direct current; more particularly, the instant invention relates to methods and apparatus for converting radio frequency energy to direct current wherein the direct current is of sufficient magnitude to power devices such as battery charges and electric motors without the use of amplification.

2. Technical Considerations in Prior Art

There has long been interest in technology directed to transmitting electrical energy over a distance without utilizing conductors, such as wire conductors. Development of such a technology has enormous potential. This was first recognized by Nikola Tesla who in 1899 constructed a 200 foot Tesla coil rated at 300 kilowatts and 150 kilocycles. Tesla hoped to set up standing waves of electrical energy around the whole surface of the earth, so that receiving antennas set at optimum points could tap the power when needed. Tesla was able to light hundreds of lamps at a distance of about 40 kilometers with his device without utilizing electrical conductors. The scheme has generally remained a scientific curiosity but has provided the initial groundwork for current developments wherein attempts are being made to transmit power by microwaves. However, power transmitted by microwaves is envisioned in the form of a beam of very high intensity which is focused from a microwave generator to a receiving antenna. This technology is envisioned as being used for many types of purposes such as transmitting microwave energy collected from gigantic solar power satellites and "star wars"-type weapons systems. However, the focused microwave beam is not suitable for many applications in that the beam must be directed toward a receiving antenna and cannot be transmitted through most objects, including living objects, without destroying the objects.

The instant invention relies on converting energy from standing waves which are emitted from radio frequency antennas in the RF range rather than the microwave range. Of particular interest are very low frequencies which are not used in communications and are available for transmitting power. Also of interest with respect to the instant invention are the low frequency waves emitted by the earth due to pulsation thereof caused by its magnetic field. These low frequency standing "earth" waves can be picked up by receivers tuned thereto.

SUMMARY OF THE INVENTION

It is an object of the invention to provide new and improved methods of an apparatus for converting radio frequency currents to direct current for practical uses other than communications, wherein the direct current energy converted from the radio frequency input energy does not require amplification.

Upon further study of the specification and appended claims, further objects and advantages of this invention will become apparent to those skilled in the art.

The instant invention contemplates an RF antenna for receiving radio waves. The RF antenna connected to a circuit configured in accordance with the principles of the instant invention to convert the RF signals to direct current. The radio frequency signals received by the antenna are transmitted to first and second leads, with one lead being rectified to transmit positive voltage and the other lead being rectified to transmit negative voltage. The positive voltage lead being connected directly to a positive output line and the negative voltage lead being connected directly to a negative output line. The positive output line is connected to a pair of bus lines through a first pair of capacitors, while the negative output line is connected to the pair of bus lines by a second pair of capacitors. Disposed between the first bus line and the positive output line is a reverse diode of negative polarity, while disposed between the negative output line and first bus line is a reverse diode of positive polarity. The positive and negative output lines are connected to one another through an inductance which is in parallel with the capacitors of the first and second pair connected between the second bus line and the positive and negative output lines.

In accordance with one embodiment of the invention the afore-described circuit is duplicated for each positive and negative output line. In accordance with another embodiment of the invention, the afore-described circuitry is coupled to additional circuits identically configured in order to increase the direct current output of the arrangement.

In accordance with a further configuration of the invention, the antenna utilized is a dipolar antenna of aluminum wire arranged in a "butterfly" configuration.

The instant invention further contemplates the method of utilizing the afore-described elements so as to generate direct current having sufficient power to perform tasks such as charging batteries, lighting lamps and powering direct current electric motors without the use of amplifiers.

BRIEF DESCRIPTION OF THE DRAWING

Various other objects, features and attendant advantages of the present invention will be more fully appreciated as the same becomes better understood when considered in connection with the accompanying drawing, in which like reference characters designate the same or similar parts throughout the several views, and wherein:

There is shown a diagram of a circuit in accordance with the instant invention in combination with a driven device and a dipolar antenna which receives radio frequency waves which are converted to DC current for powering the driven device.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring now to the drawing, there is shown a dipolar antenna, designated generally by the numeral **10**, which receives radio frequency waves from an RF transmitter **11**. The radio frequency waves are transmitted to a radio frequency to direct current converting circuit, designated generally by the numeral **12**, through a coaxial cable **13** and DC current from the circuit **12** is used to power an output device, designated generally by the numeral **15**, which may for example by a battery charger DC motor or lighting device. The circuit **12** has no other power inputs other than the radio frequency energy transmitted thereto by the antenna **10** and therefore includes no amplifiers for amplifying the RF energy.

The source of radio frequencies convertible to direct current by the circuit shown may include sources of high frequency, low frequency (LF), very low frequency (VLF) and extremely low frequency (ELF) radio waves as well as seismic vibration of the earth's magnetic fields.

Preferably, the dipolar antenna **10** is formed of two triangular loops of aluminum wire **16** and **17**, one of which is connected to the annular conductor **18** of the coaxial cable **13** and the other of which is connected to the center conductor **19** of the coaxial cable. The size of the bipolar antenna **10** is dependent on the particular application to which it is put. In accordance with one embodiment of the invention, the antenna **10** is approximately 12 inches in width and 18 inches in length. Such an antenna is used to receive five watt energy, such as that generated by a walkie-talkie or citizen-band radio.

The annular conductor **18** of the coaxial cable **13** is connected to a positive lead **21** of the circuit **12**, while the center conductor **19** of the coaxial cable is connected to a negative lead **22** of the circuit. A positive transmitting diode **D1** is disposed between the lead **21** and the remainder of the circuit **12** while a negative transmitting diode **D2** is disposed between the lead **21** to a positive output line **25** while the negative diode **D2** is connected to a negative output line **26**. Accordingly, the positive voltages with respect to ground are produced on output line **25** and negative voltages with respect to ground are produced on output line **26**.

In order to provide a DC output of sufficient power, a plurality of inductance-capacitance, RF, tuned circuits **30**, each forming a positive cell, or a negative cell, are utilized for connecting the positive output line **25** and negative output line **26** to first and second bus lines **31** and **32**, respectively. Bus line **32** is connected to ground while bus line **32** can be connected to circuits similar to circuit **12**. The positive output line **25** is connected by a first bridging line **33** to the first and second bus lines **31** and **32** while the negative output line **26** is connected by a second bridging line **34** to the first and second bus lines. The bridging line **33** has capacitors **C1** and **C2** disposed between the positive output line **25** and the first and second bus lines **31** and **32**, while the bridging line **34** also has capacitors **C1** and **C2** dispensed between the negative output line **26** and the first and second bus lines **31** and **32**. Connected between the bridging lines **33** and **34**, is an inductor **35** which serves as an RF choke, while disposed between the positive output line **25** and the capacitor **C1** there is a negative polarity diode **D4** referred to herein as a bridging diode and disposed between the negative output line **26** and capacitor **C1** in line **34** there is a positive polarity diode **D3** referred to herein as a bringing diode. As is seen of the drawing, the RF tuned circuit cell **30** is repeated a plurality of times. In the specific example shown, the circuit **12** has separate cells **30**, **37**, **38** and **39**. The cells **30** and **38** are of opposite polarity and balance one another while the cells **37** and **39** of opposite polarity and also balance one another. In order for the system to function, a pair of opposite polarized cells must be utilized. The particular number of cells **30** and the value of the components thereof are determined by the configuration of the dipolar antenna **10** and the power and frequency of the RF transmitter **11**.

The radio frequency to direct current conversion circuit **12** may itself be connected to a duplicate circuit via pins **41** so as to provide additional direct current output on lines similar to positive output line **25** and negative output line **26** the output lines may be connected together in order to boost the total output of the system.

An operative embodiment of the invention utilizes the following elements:

Diodes D1, D2, D3 and D4—Germanium Diodes, Archer **1** N34A,Catalog #1123.
Inductor 35-47 Milli henry R. F. Choke
Capacitors C1 and C2–0.47 Pico Farads at 200 volts
Coaxial Cable 13—50 ohms
Dipolar Antenna 10—aluminum wire triangular loops approximately 12 inches by 18 inches.

From the foregoing description, one skilled in the art can easily ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

What is claimed is:

1. Apparatus for converting radio frequency energy into direct current for generating electric power, the apparatus comprising:

a dipolar input for receiving radio frequency energy;

a positive output line connected to one pole of the dipolar input and a negative output line connected to the other pole of the dipolar input;

a positive transmitting input diode means in the positive output line and a negative transmitting input diode means in the negative output line;

first and second bus lines and a pair of tuned circuits of opposite polarity coupling the positive output line and negative output line to the bus lines, one of the bus lines being connected to ground;

each tuned circuit including a first bridging line connecting the positive output line to the first and second bus lines and a second bridging line connecting the negative output line to the first and second bus lines, each bridging line having a bridging diode means therein oriented at a polarity which is reverse with respect to the transmitting input diode means of the respective output line; the bridging lines of each tuned circuit being connected to one another by an inductance and having capacitors disposed between the bridging diode means thereon and the bus lines, and

a direct current device connected to the bridging lines of the tuned circuit.

2. The apparatus of claim **1**, wherein there are a plurality of similarly configured tuned circuits connected between the output lines and the bus lines.

3. The apparatus of claim **2**, wherein there are a plurality of radio frequency-to-direct current conversion circuits connected to one another to provide a direct current power array tuned to a specific radio frequency such as a high frequency source, low frequency source, very low frequency source, extremely low frequency source, or source created by seismic vibrations of the earth's magnetic field.

4. The apparatus of claim **3** wherein the dipolar input is a dipolar antenna.

5. The apparatus of claim **1**, wherein the device connected to the output lines is a direct current motor.

6. The apparatus of claim **1**, wherein the device connected to the output line is an illuminating device.

7. The apparatus of claim **1**, wherein the device connected to the output lines is a battery charger.

8. The apparatus of claim **1**, wherein the device connected to the output lines is a DC-to-AC inverter.

9. The apparatus of claim **1**, wherein there are a plurality of radio frequency to direct current converter circuits connected to one another to form an array of circuits tuned to a particular radio frequency.

10. The apparatus of claim **1** wherein the dipolar input is a dipolar antenna.

11. The apparatus of claim **10**, wherein the dipolar antenna utilizes aluminum wire arranged in pair of triangular loops.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 4,685,047
DATED : August 4, 1987
INVENTOR(S) : Raymond P. Phillips, Sr.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

On the Title Page "11 Claims" should read -- 12 Claims --.

Column 4, Claim 1, Line 21:

Reads: "into direct current for generating electric power, the"
Should Read: --into direct current, the--

Column 4, Claim 5, Line 62:

Reads: "nected to the output lines is a direct current motor."
Should Read: --nected to the output lines is a direct-current motor.--

Column 4, Claim 8, Line 68:

Reads: "nected to the output lines is a DC-to-AC inverter. "
Should Read: --nected to the output lines is a DC-to-AC converter.--

Please add the following new claim:

--12. The apparatus of claim 1 wherein the dipolar input is a coazial cable.--

Signed and Sealed this

Third Day of November, 1987

Attest:

DONALD J. QUIGG

Attesting Officer

Commissioner of Patents and Trademarks

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 4,685,047
DATED : August 4, 1987
INVENTOR(S) : Raymond P. Phillips, Sr.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

On the Title Page, Item [57],
IN THE ABSTRACT:

Line 1: Reads: "Apparatus for and methods of converting radio fre-"
Should Read: --Apparatus for converting radio fre- --

Line 9: "transmitting diode is in the positive output line and a"
--transmitting diode is located in the positive output line and a"

Line 10: "negative transmitting diode is in the negative output"
--negative transmitting diode is located in the negative output--

Line 24: "to the positive line of the circuit"
--to the output lines of the circuit--

COL. 1
Line 9: "the instant invention relates to methods of an appara-"
--the instant invention relates to appara- --

COL. 1
Line 8: "to methods and apparatus for converting radio fre-"
--to apparatus for converting radio fre- --

COL. 1
Line 14: "as battery charges and electric motors without the use"
--as battery chargers and electric motors without the use--

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 4,685,047 Page 2 of 4

DATED : August 4, 1987

INVENTOR(S) : Raymond P. Phillips, Sr.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

COL. 1
Line 29: "tific curiosity but has provided the initial groundwork"
--tific curiosity, but has provided the initial groundwork--

COL. 1
Line 57: "improved methods of an apparatus for converting radio"
--improved apparatus for converting radio--

COL. 2
Line 6: "tive voltage. The positive voltage lead being connected"
--tive voltage. The positive voltage lead is connected--

COL. 2
Line 8: "age lead being connected directly to a negative output"
--age lead is connected directly to a negative output--

COL. 2
Line 58: "ted to a radio frequency to direct current converting"
--ted to a radio frequency-to-direct current converting--

COL. 2
Line 62: "by the numeral 15, which may for example by a battery"
--by the numeral 15, which may for example be a battery--

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 4,685,047 Page 3 of 4

DATED : August 4, 1987

INVENTOR(S) : Raymond P. Phillips, Sr.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

COL. 3
Line 23: "and the remainde of the circuit 12 while a negative"
--and the remainder of the circuit 12, while a negative--

COL. 3
Line 25: "to a positive output line 25 while the negative diode D2"
--a positive output line 25, while the negative transmitting diode D2--

COL. 3
Line 31: "a plurality of inductance-capacitance, RF, tuned cir-"
--a plurality of inductance-capacitance, RF-tuned cir- --

COL. 3
Line 32: "cuits 30, each forming a positive cell, or a negative cell,"
--cuits 30, each forming either a positive cell or a negative cell,--

COL. 3
Line 50: "polarity diode D4 referred to herein as a bridging diode"
--polarity diode D4, referred to herein as a bridging diode--

COL. 3
Line 53: "D3 referred to herein as a bringing diode. As is seen of"
--D3, referred to herein as a bridging diode. As is seen of--

COL. 3
Line 58: "other while the cells 37 and 39 of opposite polarity and"
--other, while the cells 37 and 39 of opposite polarity--

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 4,685,047

Page 4 of 4

DATED : August 4, 1987

INVENTOR(S) : Raymond P. Phillips, Sr.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

COL. 3
Line 67: "via pins 41 so as to provide additional direct current"
--via pin 41 so as to provide additional direct current--

COL. 4
Line 1: "negative output line 26 the output lines may be con-"
--negative output line 26. The output lines may be con- --

Claim 12, line 2: Change "coazial" to --coaxial--.

Signed and Sealed this

Eleventh Day of July, 1989

Attest:

DONALD J. QUIGG

Attesting Officer

Commissioner of Patents and Trademarks

Oct. 20, 1970 C. C. DAILEY 3,535,543

MICROWAVE POWER RECEIVING ANTENNA

Filed May 1, 1969

3 Sheets-Sheet 1

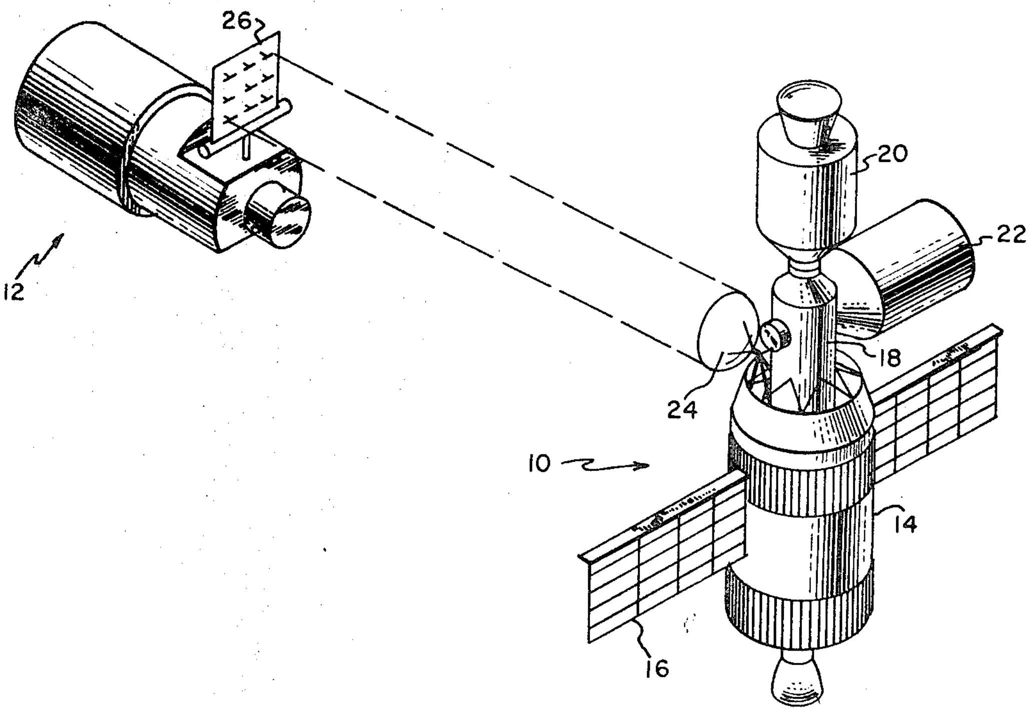
26
20
22
12
18
24
10
14
16

FIG. 1

INVENTOR
CARROLL C. DAILEY

BY George J. Porter
ATTORNEYS

Oct. 20, 1970 C. C. DAILEY 3,535,543

MICROWAVE POWER RECEIVING ANTENNA

Filed May 1, 1969

3 Sheets-Sheet 2

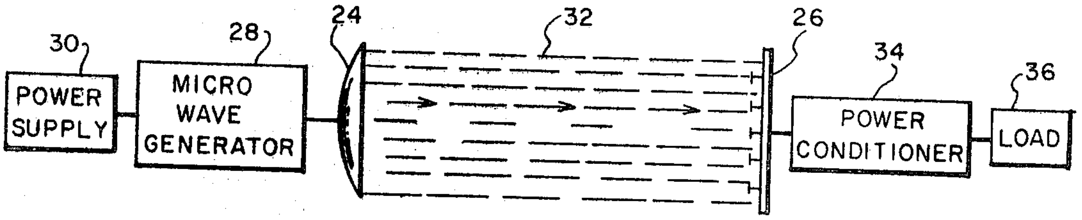
30
POWER SUPPLY
28
MICRO WAVE GENERATOR
24
32
26
34
POWER CONDITIONER
36
LOAD

FIG. 2

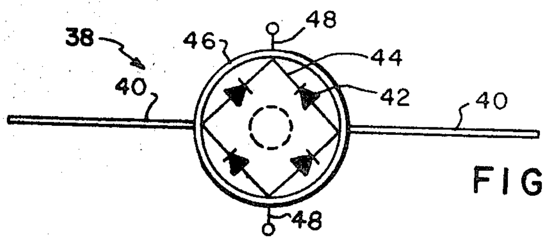
38
46
48
44
40
42
40
48

FIG. 3

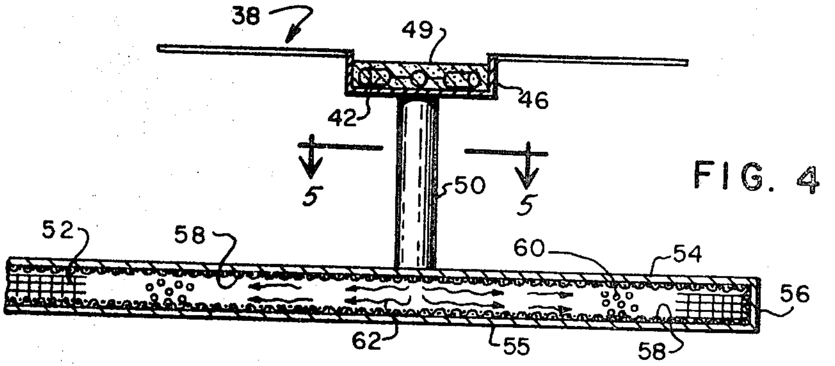
38
49
46
42
5
50
5
52
58
60
54
56
62
55
58

FIG. 4

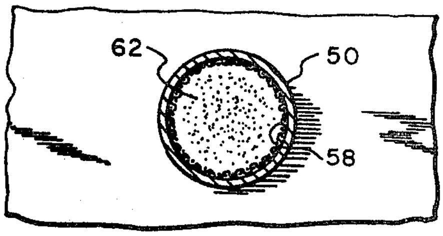

FIG. 5

INVENTOR
CARROLL C. DAILEY

BY George J. Porter
ATTORNEYS

Oct. 20, 1970 C. C. DAILEY 3,535,543

MICROWAVE POWER RECEIVING ANTENNA

Filed May 1, 1969

3 Sheets-Sheet 3

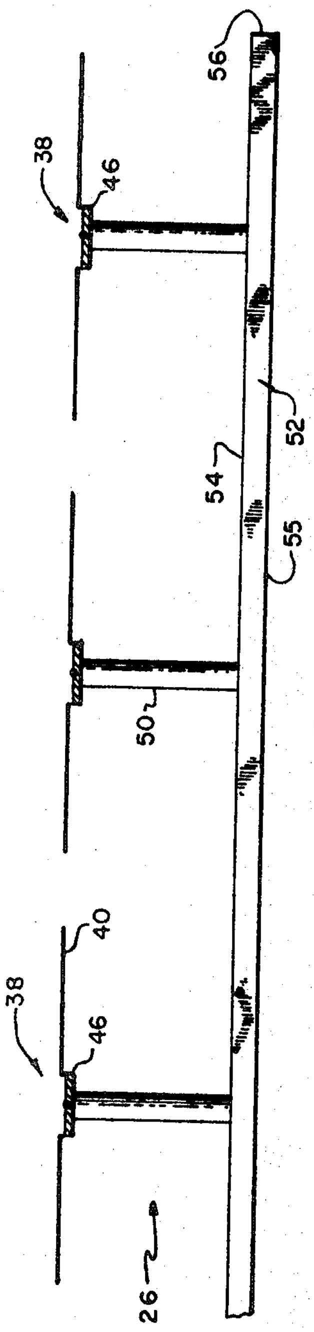
56
38
46
52
54
55
50
40
38
46
26

FIG. 6

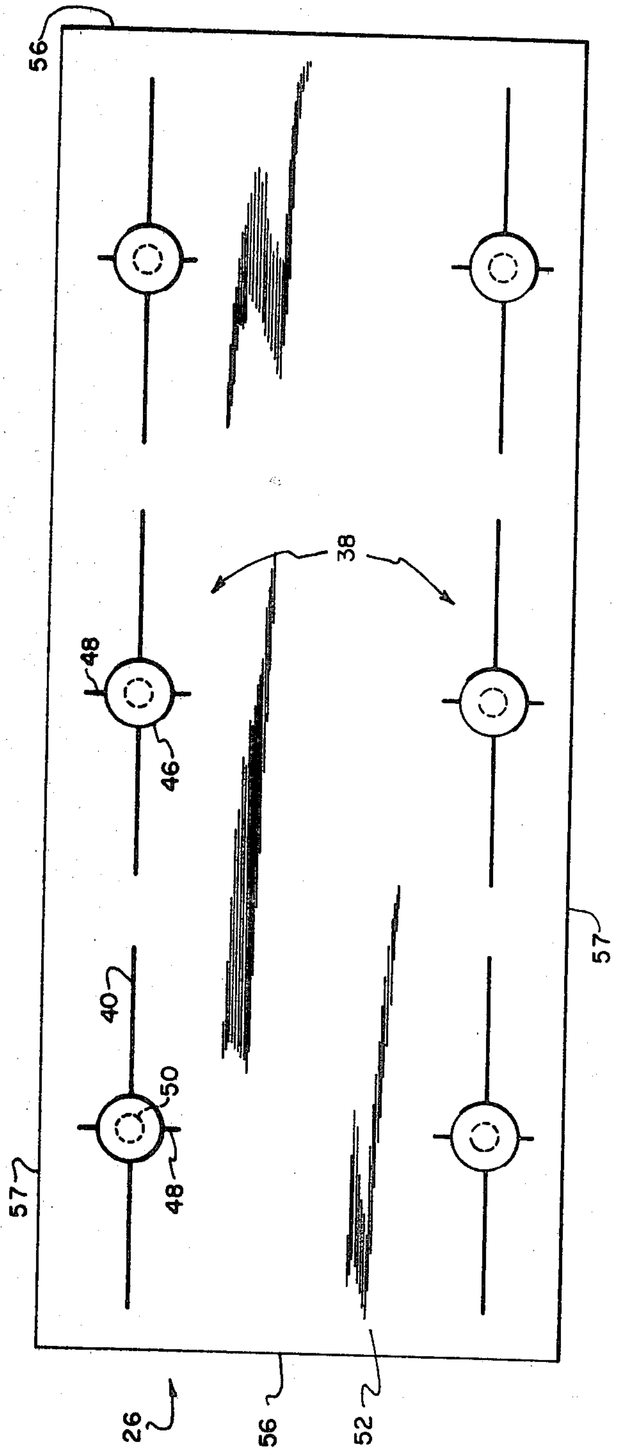
56
38
48
46
40
50
48
57
57
26
56
52

FIG. 7

INVENTOR
CARROLL C. DAILEY

BY

George J. Porter
ATTORNEYS

3,535,543
MICROWAVE POWER RECEIVING ANTENNA
Carroll C. Dailey, Huntsville, Ala., assignor to the United States of America as represented by the Administrator of the National Aeronautics and Space Administration
Filed May 1, 1969, Ser. No. 820,965
Int. Cl. H02j *1/00*
U.S. Cl. 307—149 6 Claims

ABSTRACT OF THE DISCLOSURE

A microwave power receiving antenna array having a solid-state rectifier circuit at the center of each of a plurality of dipole antennas for conversion of the high-frequency energy to direct current. The device effectively and efficiently solves the problem of heat dissipation from the diode rectifier enclosure by construction of the dipole supporting posts, the antenna reflecter and the dipole elements as heat pipe devices. Each supporting post and the antenna reflector may either communicate for greater efficiency in dissipating heat or be physically separated to simplify fabrication.

ORIGIN OF THE INVENTION

The invention described herein was made by an employee of the United States Government and may be manufactured or used by or for the Government for governmental purposes without the payment of any royalties thereon or therefor.

BACKGROUND OF THE INVENTION

Field of the invention

This invention relates to power receiving antennas and more particularly to a rectifying dipole antenna array having a highly efficient structure for dissipation of heat from the rectifying circuit.

Description of the prior art

One of the comparatively recent modes of electrical power transmission is power transfer by microwaves using frequencies as high as about 30 gigahertz. One attractive use of this type of power transmission is for powering helicopters or other types of aircraft from a remote location. Another use which appears attractive is transmission of power from a central manner space station, which can service a number of small independent subsatellites or experiment modules located distances as much as several kilometers away. Batteries in a subsatellite can be recharged with energy generated in the main space station, thus precluding the need for solar arrays or extra batteries on the subsatellite, increasing its versatility, and prolonging its useful life.

In the power receiving antenna for a microwave power transmission system, a rectifying circuit may be used to convert the high frequency energy to direct current. The diodes which transform the high frequency energy to DC are usually quite efficient (approximately 70 to 90 percent). However, there is some loss of energy which appears as heat. This heat raises the temperature of the diodes. Also, the diodes and the antenna or antenna array are subject to the heating of the sun's rays. These heating effects have the inherent disadvantage of limiting the amount of electrical power that can be handled by the array because of the temperature limits of the diodes. For diodes of interest in space applications, this limit typically occurs at about 120° centigrade, although specific applications may show limits somewhat higher or lower.

In order to keep the diode temperature down, it is necessary to radiate heat to space. The area of the radiating surface available and the coatings applied to the surface determine how much heat can be radiated at a specific temperature. In the case of antenna dipoles the radiating area is small, since the conductors are narrow and the dipoles are also very small.

SUMMARY OF THE INVENTION

Accordingly, it is an object of this invention to provide an improved microwave receiving antenna.

Still another object of this invention is to provide a microwave receiving antenna having the capability of efficient dissipation of heat from its rectifier circuits.

Yet another object of this invention is to provide a more efficient microwave receiving antenna by modifying conventional elements of antenna structure so as to better radiate the heat from the antenna's rectifier circuit.

These and other objects are accomplished in the present invention which provides at least one pair of dipole antenna elements supported by an enclosure containing a rectifying circuit comprising a plurality of diodes. Each enclosure and its corresponding pair of dipoles is mounted on a heat conducting support post, which in turn is mounted on a large antenna reflector.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be more fully understood by the following detailed description when taken together with the accompanying drawings in which:

FIG. 1 is an isometric view of an orbiting space laboratory transmitting electrical power to a subsatellite.

FIG. 2 is a block diagram of a microwave power system capable of utilizing the present invention.

FIG. 3 is a plan view of a dipole having a bridge rectifier at its center, illustrating part of one embodiment of the present invention.

FIG. 4 is a side view of part of a microwave receiving antenna, which illustrates one embodiment of the present invention.

FIG. 5 is a horizontal sectional view of the support post taken along line 5—5 of FIG. 4 and showing the inner construction of the heat pipe support post.

FIG. 6 is a side view of the microwave power receiving antenna array showing one row of dipoles.

FIG. 7 is a plan view of one embodiment of the microwave power receiving antenna array, showing two rows of three dipoles mounted on a reflector.

DESCRIPTION OF THE PREFERRED EMBODIMENT

With continued reference to the accompanying figures wherein like numerals designate similar parts throughout the various views and with initial attention directed to FIG. 1 there is illustrated an orbital space laboratory, designated generally by numeral **10**, transmitting electrical power by microwaves to a subsatellite, designated generally by numeral **12**. Laboratory **10** includes a spent rocket stage **14** having solar panels **16**, docking adapter **18** and command modules **20** and **22**. Solar panels **16** may be replaced by a nuclear power source in other versions of laboratory **10**.

Referring now to FIG. 2, a microwave power system which includes the present invention is shown in block diagram form. Electrical power in the form of microwaves is furnished by microwave power generator **28** having power supply **30**. Power transmitting antenna **24** is aimed at power receiving antenna **26** and transmits electrical power in the form of microwave beam **32** to receiving antenna **26**. Subsatellite **12** uses this power received from beam **32** to charge its batteries or fuel cells or for powering electric thrusters. The energy in beam **32** is converted from high-frequency energy (approximately 2 megahertz to 30 gigahertz) to a direct current which is fed into power conditioning equipment **34** to transform it to the actual voltages required for the load **36**.

FIG. 3 shows a self-supporting, single, half-wave, Hertz antenna, designated generally by numeral **38**, having a pair of dipole antenna elements **40**. The conversion from high-frequency energy to direct current is accomplished by a plurality of diodes **42** which are connected in a bridge rectifier circuit **44** (shown in diagrammatic form). Rectifier circuit **44** is contained in dipole center enclosure **46** and is connected between the dipole elements **40** and a pair of output terminals **48**. Dipole elements **40** are securely fastened to opposite sides of enclosure **46**.

FIG. 4 shows a single, half-wave dipole antenna **38** mounted on enclosure **46**. Enclosure **46** is shown in vertical section so that diodes **42** may be seen, contained in potting compound **49**. Enclosure **46** is mounted on support post **50** which is in turn mounted on an antenna reflector **52**. Support post **50** is a hollow pipe which is closed at its upper end where it is fastened to enclosure **46** and open at its lower end where it is fastened to antenna reflector **52**. Reflector **52** comprises a pair of metal walls **54** and **55**, a pair of ends **56**, and a pair of sides **57** (see FIG. 7) all of which make reflector **52** a closed container. However, the cavity of reflector **52** communicates with the lower end of support post **50**, as described above.

Both support post **50** and antenna reflector **52** utilize known principles of operation of a heat pipe. As may be seen in both FIG. 4 and the horizontal sectional view of the support post shown in FIG. 5, all the interior surfaces of both support post **50** and reflector **52** are covered with a wicking material **58**, which may be screen wire or a similar material. The communicating space enclosed by poth the support post **50** and the antenna reflector **52** together contains a heat transfer fluid **60**, which may be water, lithium or a number of other substances which are easily vaporized.

FIGS. 6 and 7 are side and plan views, respectively, of a microwave power receiving antenna **26** having a plurality of half-wave dipole antennas **38**. Each dipole **38** is supported by a corresponding dipole center enclosure **46** and support post **50**. All dipoles **38** are mounted on one antenna reflector **52**, so as to provide a combined broadside and collinear antenna array. Each support post **50** communicates with the antenna reflector **52** in a manner already described for FIG. 4 above. The internal construction of each support post **50** and the antenna reflector **52**, including wicking material **58** and the presence of heat transfer fluid **60**, is also as described for FIG. 4.

One cycle of operation of the microwave power receiving antenna **26** follows: Microwave power beam **32** is received on the dipole elements **40** of half-wave antenna **38**. Bridge rectifier circuit **44** comprising diodes **42** rectifies the incoming high-frequency energy to convert it to direct current. Power conditioner **34** changes the form of the direct current as desired and transmits it to the load **36**. Heat developed within enclosure **46** is absorbed by the top end of support post **50**. Heat transfer fluid **60** absorbs heat from the end of support post **50** and is vaporized. The vapor **62** then moves under vapor pressure down the support post **50** and passes into the inner portion of antenna reflector **52**. When the vapor **60** reaches a comparatively "cool spot" on the surface of antenna reflector **52** (which would theoretically be midway between the "hot spots" caused by the heat input from the support post **50** or as far as possible from a support post **50**), the vapor **62** condenses back to fluid **60** and forms deposits on the inner sides of antenna reflector walls **54** and **55**, antenna reflector ends **56**, and antenna reflector sides **57**. The fluid **60** (condensate) then moves by means of capillary flow through the wicking material **58** back to the top of support post **50**, where it absorbs more heat and starts the cycle over. In condensing, vapor **62** gives up heat which is absorbed by walls **54** and **55**, ends **56**, and sides **57**. Thus, heat is distributed evenly to all parts of the surface of antenna reflector **52** so that reflector **52** is able to effectively dissipate this heat by radiation into space.

In an alternative arrangement of the invention the interior cavity of each support post **50** does not communicate with the interior cavity of the antenna reflector **52**. Instead, the upper sidewall **54** has no openings. In this embodiment, each support post **50** as well as the antenna reflector **52** contains its own supply of heat transfer fluid **60**, which cycles within the cavity available. Thus, heat is transferred down the support post **50** and passes by conduction through wall **54** of the antenna reflector **52**. The heat is then transferred evenly to the walls of the antenna reflector **52** where it is dissipated in the manner already described above.

In another alternative arrangement, the invention may be made with a conventional flat or concave antenna reflector **52**, with a corresponding sacrifice in the heat dissipation capability of the antenna array.

Any of the above-described arrangements of the invention may be made with still another variation in its construction. Each of the antenna dipole elements **40** is made in the form of a heat pipe having its interior walls and ends lined with a wicking material **58**. Also, each dipole element **40** is completely enclosed and contains its own supply of heat transfer fluid **60**. Fluid **60** cycles within the cavity of each dipole element **40** in a manner already described above for support post **50**, taking on heat from enclosure **46** and removing it to the outside end of each dipole **40**, where it is radiated to space.

From the foregoing it may be seen that applicant has invented a novel type of microwave power receiving antenna capable of more efficient dissipation of heat than antennas previously known. The outputs from the individual dipole antennas may be connected either in series, in parallel or in series-parallel, as desired. Also, diodes **42** may be mounted directly in the cavity of support post **50** to improve heat transfer efficiency, provided they are properly insulated electrically and the appropriate electrical connections are made through the upper end of support post **50**. This approach, although more difficult from the standpoint of fabrication, is desirable at high power levels.

Of the various embodiments described in detail above, the embodiment which has support posts which communicate with the antenna reflector and which also has heat pipe conducting dipole elements is, of course, the most efficient. The simpler embodiments, although they sacrifice efficiency which is highly desirable in the invention, do have the advantage of being cheaper and easier to manufacture.

What is claimed is:

1. A microwave power receiving antenna comprising:
   (a) at least one pair of dipole antenna elements,
   (b) at least one dipole center enclosure, the inside ends of each pair of said antenna elements being mounted on one said enclosure,
   (c) a rectifying circuit in each said enclosure, said rectifying circuit comprising a plurality of diodes,
   (d) at least one support post, each said enclosure being supported by one said support post,
   (e) an antenna reflector, each said support post being mounted on said antenna reflector.

2. The microwave power receiving antenna of claim 1 wherein each said support post comprises:
   (a) a cylindrical section of pipe having two closed ends,
   (b) a fluid contained inside said pipe, for absorbing heat at the end of said pipe attached to said center enclosure, and discharging heat to said antenna reflector at the end of said pipe attached to said antenna reflector,
   (c) a wick positioned along the inner surfaces of said cylindrical section of said pipe and said pipe ends, for returning said fluid from said end of said pipe at-

tached to said antenna reflector to said end of said pipe attached to said diode enclosure.

3. The microwave power receiving antenna of claim **2** wherein said antenna reflector is a completely enclosed, flat, double-wall enclosure, said enclosure containing:

(a) a fluid for absorbing heat at the point of attachment of each said support post to said antenna reflector and discharging the absorbed heat to points on the surface of said antenna reflector remote from poinst of attachment of each said support post, so as to distribute the absorbed heat to the whole surface area of said antenna reflector,

(b) a wick positioned on substantially all of the inner surface area of said antenna reflector, for returning said fluid from said remote points to said point of attachment of said support post.

4. The microwave power receiving antenna of claim **3** wherein each said dipole antenna element comprises:

(a) a cylindrical section of tubing having two closed ends,

(b) a fluid contained inside said tubing, for absorbing heat at the end of said tubing attached to said center enclosure and discharging heat at the opposite end of said closed section of tubing,

(c) a wick positioned along the inner surface of said tubing for returning said fluid from said opposite end of said tubing to said end of said tubing mounted on said center enclosure.

5. The microwave power receiving antenna of claim **1** wherein:

(a) each said support post comprises an enclosed, cylindrical section of pipe,

(b) said antenna reflector comprises an enclosed, flat, double-wall enclosure, the enclosed area of each said support post communicating with the enclosed area of said antenna reflector, to form a total enclosed area including the enclosed areas of each said support post and said antenna reflector,

(c) said total enclosed area containing:

  (1) a fluid for absorbing heat from said center enclosure and discharging the absorbed heat to points on the surface of said antenna reflector remote from points of attachment of each said support post, so as to distribute the absorbed heat to the whole surface area of said antenna reflector,

  (2) a wick positioned on substantially all of the inner surface areas of said antenna reflector and each said support post, for returning said fluid from said remote points to said end of each said support post attached to said center enclosure.

6. The microwave power receiving antenna of claim **5** wherein each said dipole antenna element comprises:

(a) a cylindrical section of tubing having two closed ends,

(b) a fluid contained inside said tubing, for absorbing heat at the end of said tubing attached to said center enclosure and discharging heat at the opposite end of said closed section of tubing,

(c) a wick positioned along the inner surface of said tubing for returning said fluid from said opposite end of said tubing to said end of said tubing mounted on said center enclosure.

References Cited

UNITED STATES PATENTS

3,432,690	3/1969	Blume.

ROBERT K. SCHAEFER, Primary Examiner

H. J. HOHAUSER, Assistant Examiner

U.S. Cl. X.R.

174—15; 317—100; 313—105; 321—8; 310—4

United States Patent [19] [11] **3,760,257**
Fletcher et al. [45] **Sept. 18, 1973**

[54] **ELECTROMAGNETIC WAVE ENERGY CONVERTER**

[76] Inventors: **James C. Fletcher,** Administrator of the National Aeronautics and Space Administration with respect to an invention of; **Robert L. Bailey,** Gainesville, Fla.

[22] Filed: **Sept. 27, 1972**

[21] Appl. No.: **292,698**

[52] **U.S. Cl.** **321/1.5,** 136/89, 250/212
[51] **Int. Cl.** **H02m**
[58] **Field of Search** 136/89; 250/212; 333/21; 321/1.5

[56] **References Cited**

UNITED STATES PATENTS

3,026,439	3/1962	Geer	250/212 X
3,121,648	2/1964	Jensen	136/89
3,232,795	2/1966	Gillette et al.	250/212 X
3,350,234	10/1967	Ule	250/212 X
3,427,200	2/1969	Lapin et al.	136/89
3,448,273	6/1969	Webb	250/212 X
3,615,853	10/1971	Paine	136/89

Primary Examiner—William M. Shoop, Jr.
Attorney—R. F. Kempf et al.

[57] **ABSTRACT**

Electromagnetic wave energy is converted into electric power with an array of mutually insulated electromagnetic wave absorber elements each responsive to an electric field component of the wave as it impinges thereon. Each element includes a portion tapered in the direction of wave propagation to provide a relatively wideband response spectrum. Each element includes an output for deriving a voltage replica of the electric field variations intercepted by it. Adjacent elements are positioned relative to each other so that an electric field subsists between adjacent elements in response to the impinging wave. The electric field results in a voltage difference between adjacent elements that is fed to a rectifier to derive d.c. output power. The element pairs may be arranged in a two-dimensional array to provide power conversion of randomly polarized electromagnetic waves, such as sunlight.

19 Claims, 3 Drawing Figures

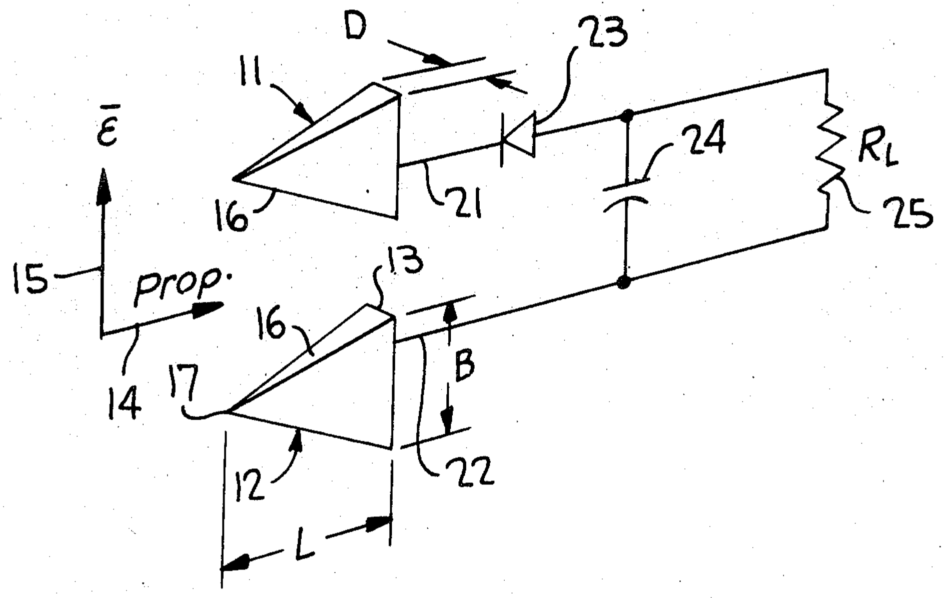

FIG. 1

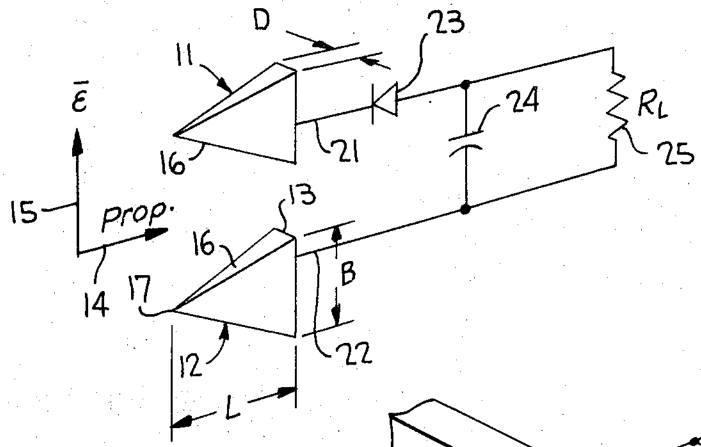
D
23
11
$\bar{\varepsilon}$
24
R_L
16
21
25
15
prop.
13
16
B
14
17
12
22
L

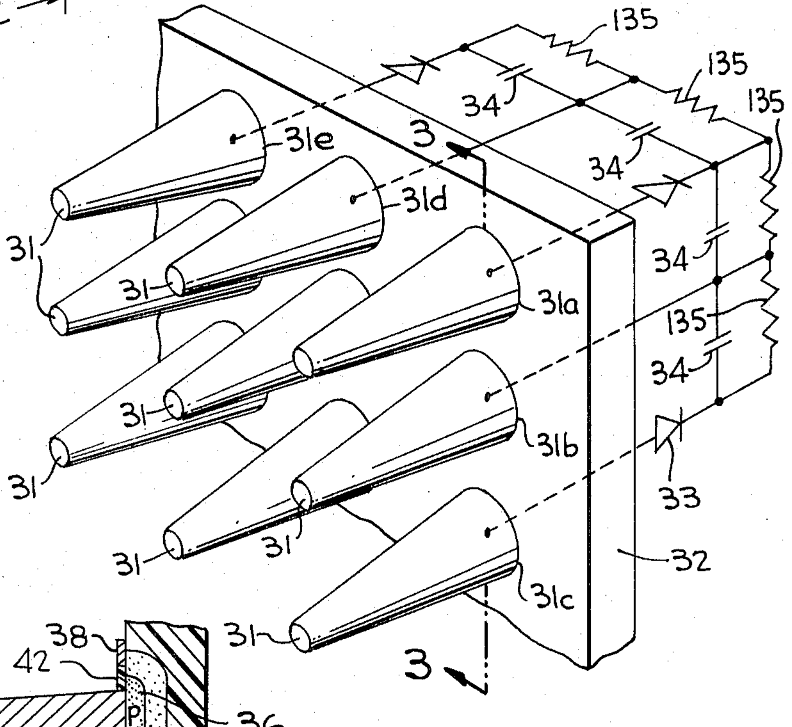
135
135
135
34
3
31e
34
31d
34
31
31a
31
135
34
31
31
31b
33
31
31
31
32
31c
31
3

FIG. 2

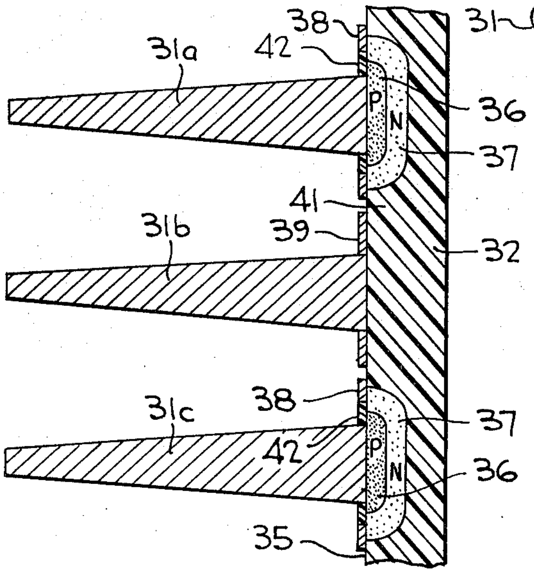
38
31a
42
P
N
36
37
41
31b
39
32
38
37
31c
42
P
N
36
35

FIG. 3

ELECTROMAGNETIC WAVE ENERGY CONVERTER

ORIGIN OF THE INVENTION

The invention described herein was made in the performance of work under a NASA contract and is subject to the provisions of Section 305 of the National Aeronautics and Space Act of 1958, Public Law 85-568 (72 Stat. 435; 42 U.S.C. 2457).

FIELD OF INVENTION

The present invention relates generally to devices for converting electromagnetic wave energy into electric power and, more particularly, to a device including a number of relatively closely spaced electromagnetic wave absorber elements having tapered portions responsive to wide band electromagnetic wave radiation.

BACKGROUND OF THE INVENTION

Devices for converting radiant energy into usable power have been considered extensively in the past. Presently, the most commonly utilized devices for converting radiant energy into usable power are solar cells which are adapted to convert solar energy directly into electricity for power generating purposes. Solar cells generally include oppositely doped semiconductor junctions which generate current in response to solar energy impinging thereon. Solar cells depend for their operation on quantum properties of light energy to provide charge separation and hence current flow in the junction region. To date, the maximum efficiency of solar cells in converting solar energy into electric energy is approximately 13 percent. In addition to the relatively low conversion efficiency of solar energy into electric power, solar cells are very expensive, as well as fragile, and usually must be mounted on a rigid, preferably flat substrate. The requirement for a rigid, flat substrate frequently substantially increases the combined cost of the solar cells and their supporting structure; if the supporting structure is a spacecraft panel, problems of folding the panel and maintaining it in a rigid condition in use are encountered.

BRIEF DESCRIPTION OF THE INVENTION

In accordance with the present invention, the wave properties of electromagnetic energy are utilized to convert wide band electromagnetic wave energy into usable electric power. The wide band energy impinges on an array of tapered absorber elements mutually insulated from each other. The length, base area, and separation of the elements are such that relatively wide band energy can be absorbed. To achieve maximum conversion of wide band energy, each element has a length, in the direction of wave propagation, equal to several wavelengths of the longest wavelength energy to be converted. Adjacent elements have spaced, sloping converging sides to enable maximum conversion of energy wavelengths in the spectrum to be converted. A voltage replica of the electric field variations absorbed by each element is derived at an output of each element. The difference between the voltage variations derived from adjacent elements is derived and supplied to a rectifier to provide the power conversion. To maximize the voltage difference, bases of adjacent absorber elements are spaced from each other by less than one wave length and preferably less than ½ wavelength, of the received energy. The close spacing, relatively long length and tapered sides also enable the absorber elements to function effectively as absorbers and prevent substantial re-radiation and reflection of the energy impinging thereon because the wave radiation is trapped in the converging interstices between the abosrbing elements.

In a preferred embodiment, the absorbers are formed of metallic elements, such as copper or other suitable material; however, absorbers may consist of dielectric elements which function to direct the electromagnetic wave energy onto an electromagnetic wave energy-to-electric voltage converter that derives a voltage replica of the wave energy.

The invention may be utilized in conjunction with converting wave energy from the microwave region through the visible light spectrum. If the invention is utilized for wave lengths in the solar spectrum, it is ideally suited for use as a solar energy-to-electric power converter. In such a configuration, the absorber elements and electric components utilized for converting the voltage wave replicas into electric power might be formed utilizing integrated circuit type manufacturing processes.

If the device is utilized for absorbing plane polarized electromagnetic waves, adjacent elements in the direction of the wave electric field are aligned with each other in pairs such that maximum electric field variations are derived between them but adjacent elements at right angles to the electric field are packed as close as possible to each other without electric contact, to maximize the active absorber area. If the electromagnetic wave radiation is circularly or randomly polarized, as in the case of solar energy, a two-dimensional array of elements is provided. The elements of the two-dimensional array are positioned in mutually orthogonal directions having aligned columns and rows, and the elements preferably have symmetrical bases, such as a square or circle. Voltage differences are derived between adjacent pairs in both orthogonal directions to provide maximum conversion efficiency of the circularly or randomly polarized wave energy.

Because the conversion process is in response to the wave properties of the impinging electromagnetic wave energy, rather than the quantum properties of such waves, and because of the small number of loss mechanisms individually and collectively optimal, it is believed possible to achieve conversion efficiencies considerably greater than existing solar cell type devices. A major advantage of the present invention is the separation of the wave absorption means and the conversion means permitting each to be individually optimized for the incident wave electromagnetic power spectrum. Another major advantage is that by suitable choice of geometry for the absorbing elements the device can be made to match the incident radiation spectrum. There is no known means of achieving this desirable result with present art solar cells. Another advantage of the present invention is that it does not utilize temperature-sensitive active semiconductor elements, for the basic absorption process; the only semiconductor components are passive diodes. A further advantage of the present invention is that the substrate on which the elements are mounted can be mechanically flexible, to substantially eliminate many of the problems inherent with prior art fragile solar cells mounted on rigid substrates.

It is, accordingly, an object of the present invention to provide a new and improved device for converting electromagnetic wave energy into electric power.

Another object of the invention is to provide a new and improved device for converting wide band electromagnetic wave energy into electric power with relatively high efficiency, utilizing the wave properties of the energy.

An additional object of the invention is to provide a new and improved device for converting circularly or randomly polarized electromagnetic wave energy into electric power in response to electric field gradients.

A further objet of the invention is to provide a relatively inexpensive device for converting electromagnetic wave energy into electric power, which device is relatively temperature-insensitive and can be mounted on a flexible substrate.

An additional object is to provide a new and improved integrated circuit type device for converting solar energy into electric power.

An additional object is to provide a new and improved means of matching the response of the device to the incident electromagnetic wave spectrum.

The above and still further objects, features and advantages of the present invention will become apparent upon consideration of the following detailed description of several specific embodiments thereof, especially when taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWING

FIG. 1 is a perspective view schematically illustrating the principles of the present invention;

FIG. 2 is a perspective diagram illustrating an embodiment of the invention particularly adapted for converting solar energy into electric power; and

FIG. 3 is a side sectional view of the embodiment illustrated in FIG. 2, along the lines 3—3.

DETAILED DESCRIPTION OF THE DRAWING

Reference is now made to FIG. 1 of the drawing wherein is illustrated a pair of mutually insulated, substantially identical, pyramidal or conical, aligned and relatively closely packed, preferably metallic, as copper or other suitable material, electromagnetic wave absorber elements **11** and **12** which are responsive to plane polarized wide band electromagnetic wave energy that is propagating in a direction indicated by arrow **14** and has an electric field with a gradient in a direction indicated by arrow **15.** Elements **11** and **12** are constructed so that the shapes of every cross section parallel to their bases **13** are similar; the cross-sectional areas decrease for increasing distances from the bases **13** of the elements. Each of elements **11** and **12** includes a co-planar base **13**, which may have any desired configuration, such as a rectangle, square or circle, that lies in a plane at right angles to the direction of wave propagation and is parallel to the electric field gradient **15.** Elements **11** and **12** have a substantial height, L, or length in the direction of wave energy propagation, between their base **13** and apex **17.** The height is generally several (up to approximately seventy five) wave lengths of the longest wavelength of the incident electromagnetic energy. Elements **11** and **12** include adjacent, tapered side surfaces **16** that converge toward base **13** and are positioned relative to each other to intercept the electric field **15** and absorb energy over a wide band of the electromagnetic wave. The apex of absorber elements **11** and **12** may be truncated in which case L is the length of the truncated absorber elements.

Voltage variations derived in elements **11** and **12** in response to the electric field variations of the electromagnetic wave energy impinging on surfaces **16** are transduced or converted into an electric voltage at terminals **21** and **22** of the elements. The voltages at terminals **21** and **22** are replicas of the electric field variations impinging on absorber elements **11** and **12,** whereby the voltage difference between terminals **21** and **22** is proportional to the electric field gradient established by the plane polarized wave between facing surfaces **16** of elements **11** and **12.**

The centerline spacing between adjacent elements **11** and **12** is such that for the incident electromagnetic wave spectrum, maximum voltage between **21** and **22** occurs. It is of the order of a wavelength or less of the incident electromagnetic energy. The present invention responds to this voltage difference as derived over the desired spectrum of the incident wave energy, to convert the wave energy into electric power.

The voltage difference between terminals **21** and **22** is converted into useful d.c. electric power by a rectifier **23** and a filter capacitor **24,** the voltage across which is supplied to a suitable external load, such as resistor **25.** The cathode is connected between the anode of the rectifier diode and terminal **22.** It is to be understood that a full wave rectifier can be employed and that the d.c. voltage developed across external load **25** can be supplied to any suitable device, such as a d.c.-to-a.c. power frequency device.

The breadth,B, and depth,D, of base **13** as well as the length, L, and spacing between the elements **11** and **12** enable the band width to which the elements **11** and **12** are responsive to be adjusted to provide optimum matching to the spectrum of the electromagnetic wave energy impinging on absorber elements **11** and **12.** Close spacing between adjacent elements **11** and **12,** in addition to enabling the electric field gradient to be coupled with the greatest voltage difference to terminals **21** and **22,** enables the electric field to be effectively trapped in the converging interstices between facing, adjacent elements **11** and **12.** Trapping of the electric field variations occurs because elements **11** and **12** are absorptive to the electric electromagnetic wave energy, whereby the energy is not reflected and re-radiated from the elements.

Because the device of FIG. **1** is responsive to a plane polarized electromagnetic wave, it is preferable for the depth, D, of each of the elements **11** and **12** to be relatively narrow, less than ¼ wave length, so that adjacent elements (not shown) can be packed as closely as possible to either side of elements **11** and **12** in the horizontal direction. Close packing of the elements enables maximum conversion of the electromagnetic wave energy into electric power over the entire area of an array which may be fabricated out of a multiplicity of element pairs as illustrated in FIG. **1.** It is to be noted that the device of FIG. **1** is responsive most efficiently to electromagnetic waves polarized so that the E field is in a plane parallel or coplanar with the long dimension of bases **13.** If the electromagnetic wave polarization direction were rotated 90°, there would be no substantial electric field variation between the adjacent elements **11** and **12** and there would be substantially zero

voltage developed between terminals 21 and 22 with a resulting zero conversion of electromagnetic wave energy into electric power.

The power of the electromagnetic wave incident on absorber elements 11 and 12 is the incident power density times the effective area of the two elements and can thereby be stated approximately as: $2(\bar{E}\times\bar{H})BD$, where $\bar{E}$ and $\bar{H}$ are respectively the electric and magnetic fields of the wave energy. The power indicated by the equation is available so that it can be converted into electric power supplied to a load, except for losses in elements 11 and 12, rectification, and stray losses. The losses of elements 11 and 12 are principally due to skin effect or di-electric losses, while the rectifier losses are the series resistance of the diode and stray losses are of the capacitor, and supporting substrate for elements 11 and 12. Stray losses can be minimized by locating elements 11 and 12 on a low loss dielectric substrate to minimize loss currents between the abosrber terminals.

The maximum wavelength restriction (about 75λ of the longest wavelength of the received spectrum) is provided to enable the shortest wavelength of the desired spectrum to be trapped between adjacent elements and to provide a relatively large surface area for extracting power from it. If the maximum wavelength restriction, which corresponds approximately with the length of a cone in an eyeball of a mammal (about 75λ of the longest wavelength of the received spectrum), is exceeded by the incident electromagnetic energy spectrum, then the voltage output between terminals 21–22 of the device decreases.

If the device of FIG. 1 is utilized to convert plane polarized microwave electromagnetic wave energy into electric power, the rectifier circuit can be connected to the terminals by a coaxial cable having a center conductor connected directly to terminal 21 and a shielded outer conductor, possibly with a conventional balun, connected to terminal 22. In the alternative, the rectifier circuit can be connected to elements 11 and 12 by a wave guide excited by the voltage difference between terminals 21 and 22 so that propagation in the wave guide is in one of the transverse electric modes, such as TE_{01}. In such a configuration, upper and lower conducting surfaces of a rectangular wave guide, between which the electric field is developed, are respectively connected to terminals 21 and 22.

One microwave absorber actually constructed and tested in accordance with the present invention includes a pair of sheet copper pyramidal elements 11 and 12 having a length (L) of 13.3 centimeters, a breadth (B) of 6.3 centimeters, and a depth (D) of 2.0 centimeters. This absorber pair had a center frequency of approximately 475 Mhz and a pass band between approximately 200 and 700 Mhz. Increasing the length of the abosrber elements, without any changes in the base dimensions of the pyramid, lowered the pass band center frequency, as well as the upper and lower cut-off frequencies.

The basic converter structure of FIG. 1 can be modified, as illustrated in FIG. 2, to respond to circularly polarized electromagnetic waves or randomly polarized electromagnetic waves, as subsist in solar energy. In FIG. 2, a multiplicity of absorber elements 31 are arranged in a two-dimensional array of aligned orthogonally directed columns and rows. Elements 31 are mutually insulated from each other and mounted on a dielectric substrate 32, which is preferably flexible but may be rigid. Each of elements 31 preferably has a symmetrical base and cross section parallel to the base, which may be either circular or square, to enable equal voltages to be derived between adjacent ones of elements 31 in the two orthogonal directions in response to orthogonal electric field components at the same wavelength. Thereby, the total array power output of FIG. 2 is insensitive to the polarization direction of the energy impinging thereon and the array can respond to circularly polarized, as well as randomly polarized electromagnetic wave energy.

To convert the electromagnetic wave energy impinging on elements 31 into electric power, adjacent elements are interconnected with each other by a half wave rectifying network similar to that illustrated in FIG. 1. In particular, aligned vertical elements 31*a*, 31*b*, and 31*c* are interconnected so that the anodes of rectifier diodes 33, connected to elements 31*a* and 31*c* (which are separated from each other by element 31*b*) are connected directly to the elements, while the cathodes of the rectifier diodes are connected to one electrode of different capacitors 34, the other electrodes which have a common connection to element 31*b*. Similarly, horizontally aligned elements 31*a*, 31*d*, and 31*e* are connected to rectifier circuits such that the center element of the triad has a common connection to a pair of capacitors 34 which are connected to the cathodes of diodes 33, having anodes connected to be responsive to the voltage replicas respectively derived by absorber element pairs 31*a*–31*d* and 31*e*–31*d*. D. C. voltages developed across capacitors 34 are supplied to a matrix of load resistors 135, one of which is connected across each of the capacitors.

Elements 31 of FIG. 2 are illustrated as frustoconical structures having upper bases with considerably smaller areas than the bases of the elements that are secured to substrate 32. The frusto-conical configuration is preferred in certain instances because sharp points, as illustrated in the pyramidal elements 11 and 12, FIG. 1, may have a tendency to fracture.

To enable the structure illustrated in FIG. 2 to function as a device for converting solar electromagnetic wave energy, which has the majority of its power in the spectrum from 0.3 to 1.1 microns, into d.c. power, the element dimensions and inter-element spacing must be on the order of a micron, whereby integrated circuit techniques are preferably employed in fabrication. To this end, a solar energy converter in accordance with the present invention may take the form illustrated by the cross-sectional view illustrated in FIG. 3 but not limited to this embodiment. In FIG. 3, electrically conductive cones 31*a*, 31*b*, and 31*c* project from surface 35 of substrate 32. Alternate ones of cones 31, such as cones 31*a* and 31*c*, are ohmically connected to a P-doped regions 36 of substrate 32, which are formed on surface 35 of substrate 32 in super-position with the bases of cones 31*a* and 31*c*. A junction is formed between P-doped regions 36 and N-doped regions 37 which are formed in substrate 32 beneath P-doped regions 36. Regions 36 and 37 respectively correspond with the anode and cathode of diodes 33 connected to cones 31*a* and 31*c*, FIG. 2. Annular, dielectric oxide films 42 are formed on face 35 over the intersections between the otherwise exposed portions of the junctions between regions 36 and 37. Oxide films 42 cover the peripheries of P-doped regions 36 and the inner circumference of N-doped regions 37. On the portion of

P-doped regions **36** not covered by oxide films **42** and the inner radial portion of the films, metal elements **31***a* and **31***c* are placed. The bases of elements **31***a* and **31***c* are ohmically connected to P-doped regions **36.** Annular, metal films **38** are formed on face **35** in superposition with the otherwise exposed regions **37** to form ohmic contacts with the N-doped regions. Metal films **38** are electrically insulated from elements **31***a* and **31***c* by dielectric films **42** which are disposed between metal films and elements. Metal film **39** is formed on face **35** at the base of element **31***b* and is electrically connected to the base. The areas **41** of substrate **32** between adjacent metal films **38** and **39** form dielectrics for capacitors **34,** having electrodes comprised of the metal film portions **38** and **39.** A d.c. load is connected between metal film portions **38** and **39** so that the film portions form load terminals.

The structure of FIG. 3 is formed by diffusing P and N regions **36** and **37** onto surface **35** of substrate **32,** utilizing conventional integrated circuit techniques. Regions **36** and **37** are diffused only onto the areas beneath and slightly to the sides of where cones **31***a* and **31***d* are to be located. Thereafter, annular oxide regions **42** are formed over the exterior portion of P-doped region **36** and over a slight segment of the interior of N-doped region **37** on surface **35.** After oxide regions **42** have been formed, thin metal film contacts **38** are vacuum vapor deposited on the exposed portions of N regions **37.** Simultaneously with the formation of contacts **38** thin, metal film, annular contact **39** is vacuum vapor deposited on surface **35** in the region outside of cone **31***b*. After metal films **38** and **39** have been deposited, cones **31***a*, **31***b*, and **31***d* are suitably attached.

While metal is believed preferable in most instances for the absorber elements, it is to be understood that dielectric elements can also be utilized as elements for directing electromagnetic wave radiation to an optical to d.c. converter diode. The dielectric elements would preferably have the same configuration and dimensions as indicated supra while the optical to d.c. converter diode may be of a type described by Javin in the IEEE Spectrum, October, 1971, page 91.

While there have been described and illustrated several specific embodiments of the invention, it will be clear that variations in the details of the embodiments specifically illustrated and described may be made without departing from the true spirit and scope of the invention as defined in the appended claims.

I claim:

1. A device for converting electromagnetic wave energy having an electric field into electric power comprising a plurality of mutually insulated electromagnetic wave absorber elements each responsive to the electric field, adjacent ones of said elements converging toward each other in the direction of the wave propagation, each element including means for deriving a voltage replica of variations of the electric field intercepted by the element, adjacent ones of said elements being positioned relative to each other so that an electric field subsists between said adjacent elements, and means responsive to the voltage difference between the derived voltage replicas between two adjacent elements,

2. The device of claim **1** wherein each of the absorber elements is metallic.

3. The device of claim **1** where each of the absorber elements comprises a metallic surface and has a length in the direction of wave propagation several wavelengths of the impinging electromagnetic wave.

4. The device of claim **3** wherein the adjacent elements have substantially coplanar bases, and the spacing between the adjacent elements at the bases thereof is of the order of no more than a wavelength of the electromagnetic energy.

5. The device of claim **1** wherein the adjacent elements have substantially coplanar bases, and the spacing between the adjacent elements at the bases thereof is of the order of no more than a wavelength of the electromagnetic energy.

6. The device of claim **1** wherein the conversion means responsive to the voltage difference includes means for rectifying the voltage difference to derive a d.c. voltage.

7. A device for converting circularly polarized or randomly polarized electromagnetic wave energy having an electric field into electric power comprising a two-dimensional array of mutually insulated electromagnetic wave absorber elements each responsive to a component of the electric field, each elements converging toward each other in the direction of propagation of the wave and including means for deriving a voltage replica of variations of the electric field intercepted by the element, adjacent ones of said elements being positioned relative to each other so that an electric field subsists between said adjacent elements, and means responsive to the voltage difference between the voltage replicas derived between two adjacent elements, said pairs being spaced at right angles to each other in the array.

8. The device of claim **7** wherein each of the absorber elements is metallic.

9. The device of claim **7** where each of the absorber elements comprises a metallic surface and has a length in the direction of wave propagation several wavelengths of the electromagnetic wave.

10. The device of claim **9** wherein the adjacent elements have substantially co-planar bases and the spacing between the adjacent elements at the bases thereof of the order of one wavelength or less of the electromagnetic energy.

11. The device of claim **7** wherein the adjacent elements have substantially coplanar bases and the spacing between the adjacent elements at the bases thereof is of the order of one wavelength or less of the electromagnetic energy.

12. A device for converting solar, randomly polarized electromagnetic wave energy in the wave length band from approximately 0.3 to 1.1 microns into electric power comprising a dielectric substrate having a face substantially at right angles to the direction of propagation of the solar electromagnetic wave energy, a plurality of mutually insulated electromagnetic wave, metallic absorber elements mounted on said face, each of said elements including tapered portion extending away from the face in the direction of propagation of the wave energy, each of said elements having a decreasing cross-sectional area as the distance from the face increases and a length in the direction of propagation of the wave several wavelengths of the shortest wavelength of the spectrum, an integrated circuit diode in said substrate connected to be responsive to wave energy intercepted by alternately spaced ones of said

elements, a metallic film on said substrate ohmically connected to the remaining metallic elements, whereby a capacitor is formed in the substrate dielectric between adjacent ones of said metallic films.

13. The device of claim **12** wherein each of the elements includes a base with a symmetrical geometry in contact with said face, each element having a cross section parallel to the base similar to the base.

14. The device of claim **13** wherein said face has a square cross section.

15. The device of claim **13** wherein said face has a circular cross section.

16. The device of claim **13** wherein each of said elements is a frustum.

17. The device of claim **12** wherein the spacing between adjacent ones of said elements is of the order of one wavelength or less of the shortest wavelength of the spectrum.

18. The device of claim **12** wherein the several wavelengths are between 1.25 and 75.

19. A device for converting solar, randomly polarized electromagnetic wave energy in the wave length band from approximately 0.3 to 1.1 microns into electric power comprising a dielectric substrate having a face substantially at right angles to the direction of propagation of the solar electromagnetic wave energy, a plurality of mutually insulated electromagnetic wave, metallic absorber elements mounted on said face, each of said elements including a tapered portion extending away from the face in the direction of propagation of the wave energy, each of said elements having a decreasing cross-sectional area as the distance from the face increases, adjacent ones of said elements being spaced from each other by a distance on the order of a wavelength or less of said energy, an integrated circuit diode in said substrate connected to be responsive to wave energy intercepted by alternately spaced ones of said elements, a first metallic film on said substrate ohmically connected to one electrode of each of said diodes, a second metallic film on said substrate ohmically connected to the remaining metallic elements, said first and second films including load terminals whereby a capacitor is formed in the substrate dielectric between adjacent ones of said metallic films, and said load terminals.

* * * * *

343-700 MS AU 256 EX

3/14/78 XR 4,079,268

United States Patent [19] [11] 4,079,268

Fletcher et al. [45] Mar. 14, 1978

[54] THIN CONFORMAL ANTENNA ARRAY FOR MICROWAVE POWER CONVERSION

[76] Inventors: **James C. Fletcher**, Administrator of the National Aeronautics and Space Administration with respect to an invention of; **Richard M. Dickinson**, La Crescenta, Calif.

[21] Appl. No.: **730,045**

[22] Filed: **Oct. 6, 1976**

[51] **Int. Cl.**^{2} **H01Q 1/38**; H02M 1/00

[52] **U.S. Cl.** .... **307/151**; 361/395; 343/700 MS

[58] **Field of Search** .... 343/700 MS, 708, DIG. 2, 343/846; 307/151; 361/395

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,778,717	12/1973	Okoshi et al.	331/108 D
3,887,925	6/1975	Ranghelli et al.	343/814
3,947,850	3/1976	Kaloi	343/846
3,987,455	10/1976	Olyphant	343/700 MS

Primary Examiner—Eli Lieberman

Attorney, Agent, or Firm—Monte F. Mott; John R. Manning; Paul F. McCaul

[57] ABSTRACT

A structure of a circularly polarized, thin conformal, antenna array which may be mounted integrally with the skin of an aircraft employs microstrip elliptical elements and interconnecting feed lines spaced from a circuit ground plane by a thin dielectric layer. The feed lines are impedance matched to the elliptical antenna elements by selecting a proper feedpoint inside the periphery of the elliptical antenna elements. Diodes connected between the feed lines and the ground plane rectify the microwave power, and microstrip filters (low pass) connected in series with the feed lines provide DC current to a microstrip bus. Low impedance matching strips are included between the elliptical elements and the rectifying and filtering elements.

7 Claims, 6 Drawing Figures

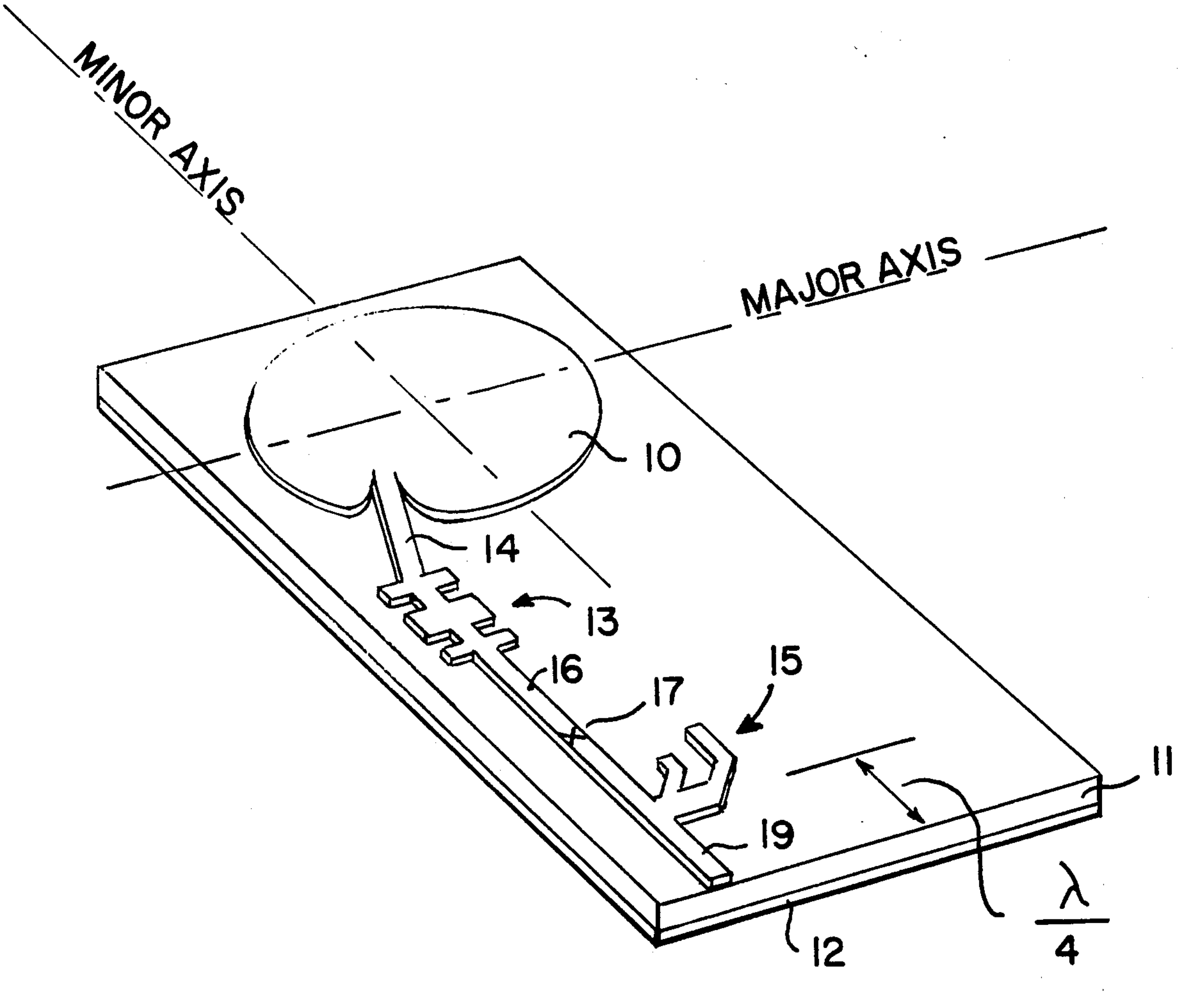

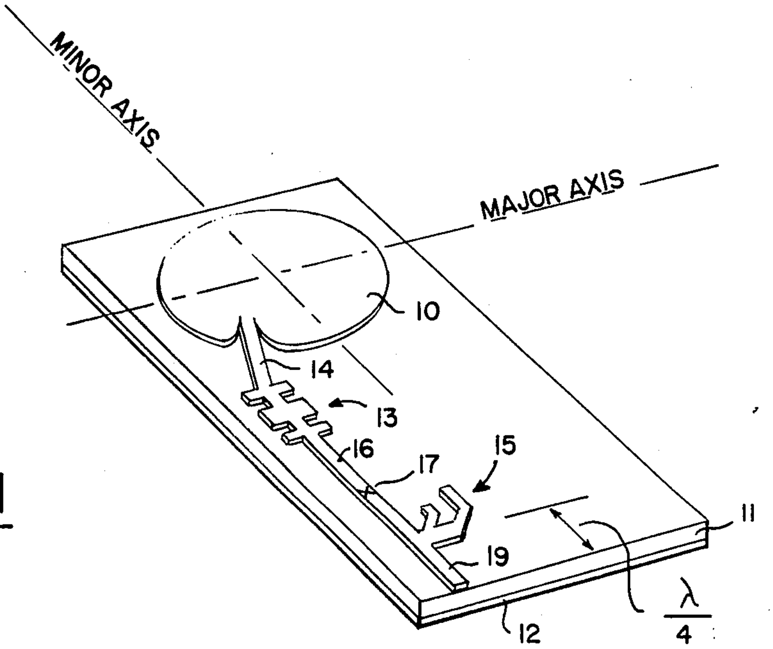
MINOR AXIS
MAJOR AXIS
10
14
13
16
17
15
11
19
12
λ/4

FIG. 1

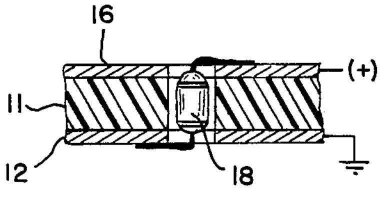

FIG. 2a

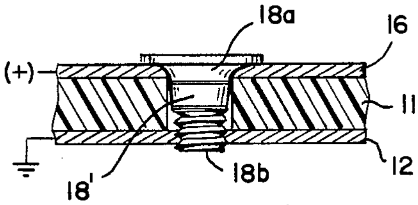
18a
16
(+)
11
12
18b
18'

FIG. 2b

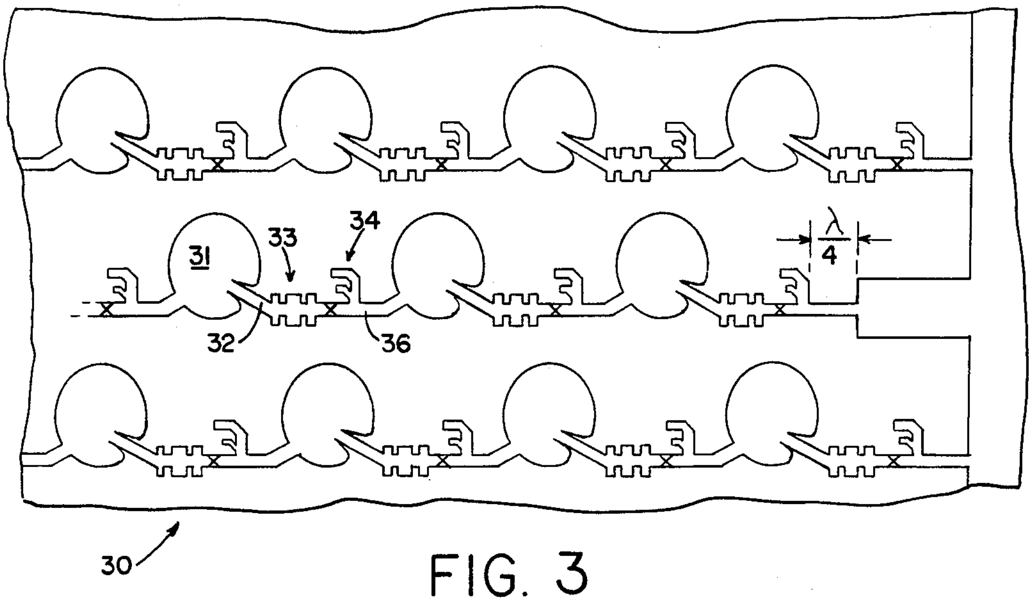
34
33
31
λ/4
32
36
30

FIG. 3

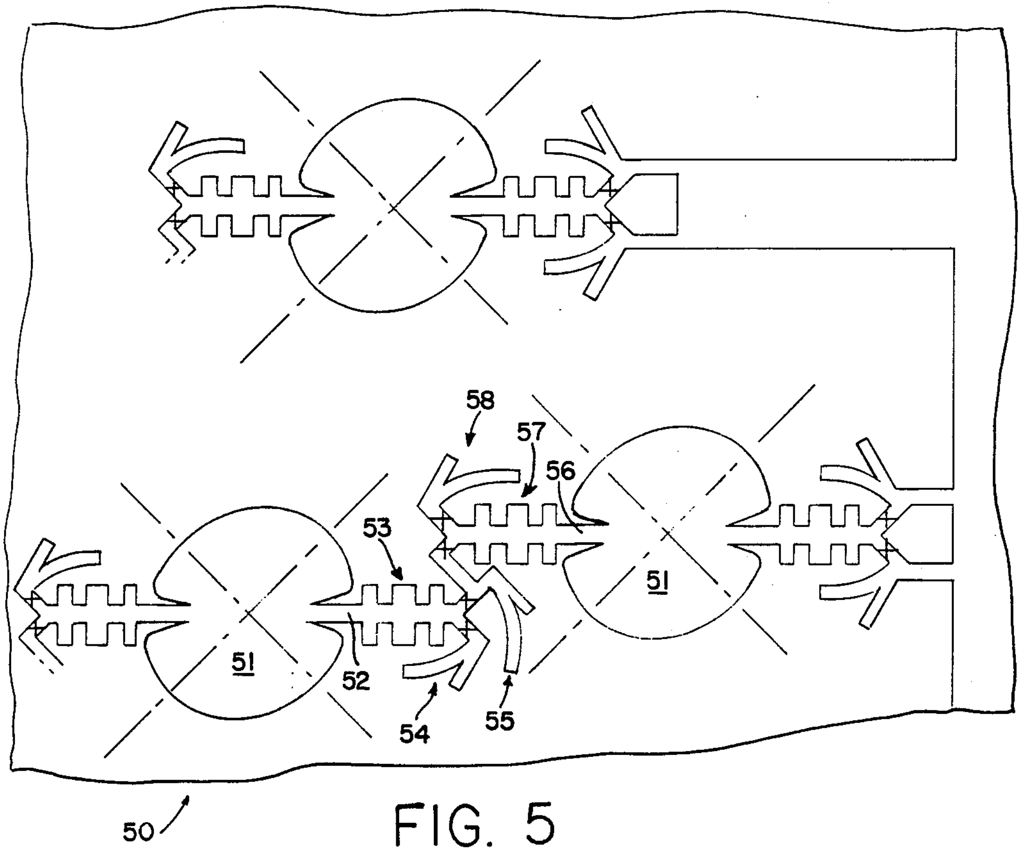
58
57
56
53
51
51
52
54
55
50

FIG. 5

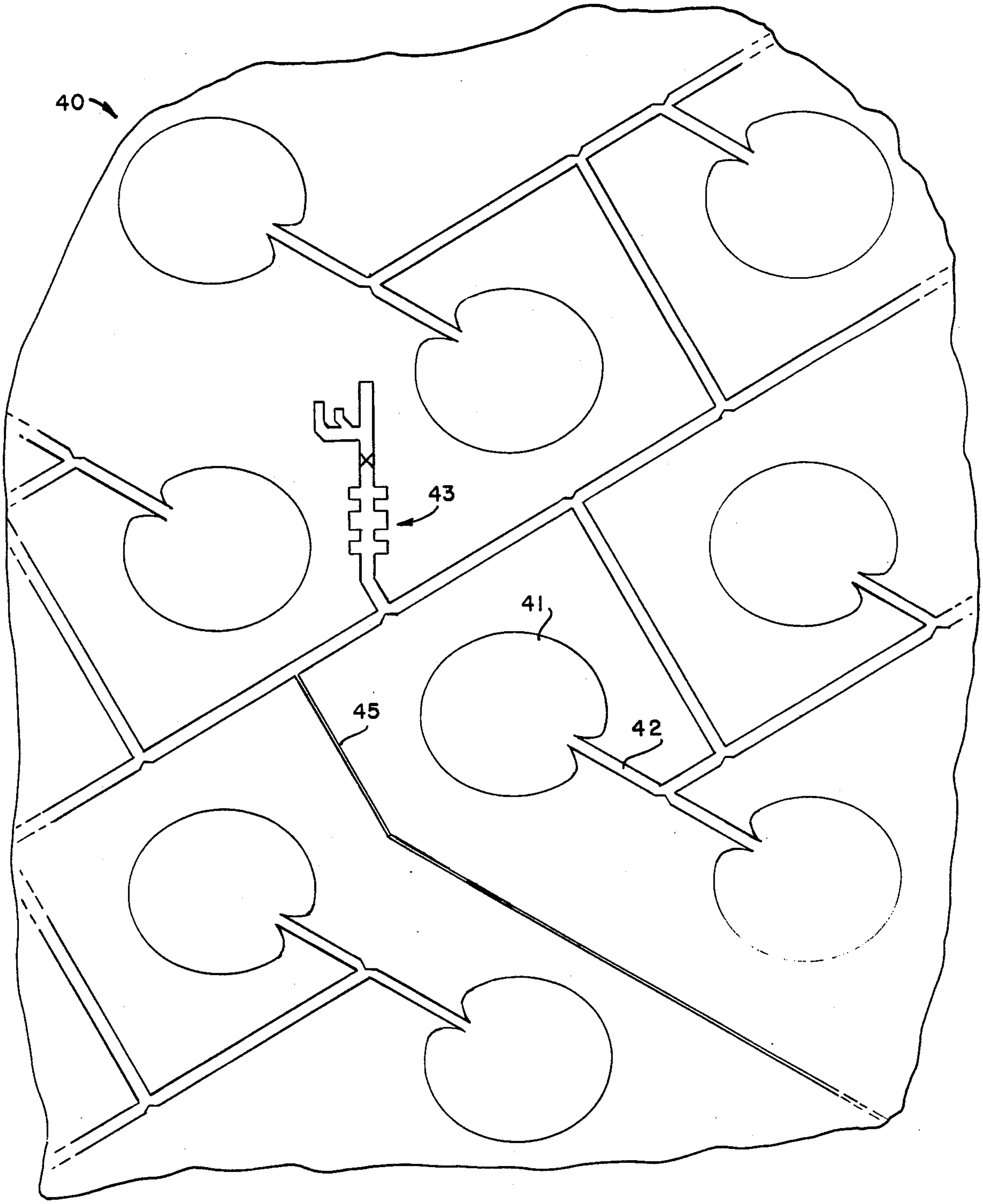
40
43
41
45
42

FIG. 4

THIN CONFORMAL ANTENNA ARRAY FOR MICROWAVE POWER CONVERSION

ORIGIN OF THE INVENTION

The invention described herein was made in the performance of work under a NASA contract and is subject to the provisions of Section 305 of the National Aeronautics and Space Act of 1958, Public Law 85-568 (72 Stat. 435; 42 USC 2457).

BACKGROUND OF THE INVENTION

This invention relates to improvements in a microstrip circularly-polarized antenna array and more particularly to such an array for receiving and converting microwave power to DC current.

There is a long-standing need for low cost, low profile conformed antenna arrays which can be mass produced. That need has been more fully satisfied by microstrip antenna techniques than by any other prior art techniques reviewed in U.S. Pat. No. 3,803,623. According to that review of the technique, circular, elliptical, or even rectangular (but preferably elliptical) antenna elements are made by etching copper clad low-loss printed circuit board material, and a feedpoint is selected to match a low transmission line characteristic impedance, typically 50 ohms. A problem with laying out an array of antenna elements is providing the proper feedpoint for the desired input impedance in elements of a given geometry. It has been discovered that this problem is greatly alleviated by setting in the feedpoint sufficiently to reduce the input impedance to the value required for whatever feedpoint is selected.

When an array of dipole elements is employed to convert microwave energy for power, as for propulsion of a vehicle, it is desirable for the energy received to be rectified at the antenna array. Such an arrangement is commonly referred to as a "rectenna" because it combines the function of rectification with that of receiving microwave energy in the antenna array by placing rectifying diodes between dipole elements to break up the large aperture of the array into the smaller apertures of the dipoles as disclosed in U.S. Pat. No. 3,434,678 in which each dipole is effectively terminated by a diode whose DC output feeds into a common DC load either in series or in parallel with dipole elements. A problem with such a rectenna is that optimum performance of the dipoles requires a fixed orientation with respect to the polarization angle of the transmitted beam. Other examples of rectennas are to be found in U.S. Pat. Nos. 3,535,543 and 3,852,755. An example of how rectennas are to be used may be found in U.S. Pat. No. 3,114,517. It is desirable to combine with the technology of rectennas the technology of microstrip antennas of which U.S. Pat. Nos. 3,803,623 and 3,921,177 are exemplars, but in a thin conformal array particularly suited to mounting integrally with thin structures, such as aircraft wings, requiring minimal space.

SUMMARY OF THE INVENTION

In accordance with the invention, a plurality of circularly polarized (disk) antenna elements are fabricated on a metal sheet clading a layer of dielectric material by etching the metal sheet in a pattern which includes conductive strips connecting the elements to a common load terminal, each antenna element having at least one microstrip low-pass filter connected to low impedance matching stubs, and including at least one rectifying diode lodged in the dielectric material with one terminal connected to the conductive strip between the filter and low impedance matching stubs, and the other terminal connected to a metal backing sheet. For low power, all disk elements are connected in parallel to the load terminal with a single microstrip filter connected to matching stubs and a rectifying diode. For moderate power levels, disk elements are connected in series-parallel with a number of disk elements in series, each feeding into a microstrip filter and matching stubs with a diode connected between the microstrip filter and the matching stubs. For high power levels, disk elements of an array are connected in series-parallel, each with a plurality of filters matching stubs and rectifying diode assemblies. For impedance matching the feedpoint of a disk element to a microstrip filter, a feed line (conductive strip) is inset to a point of proper impedance within the periphery of the disk element by etching away a notch on the disk element on each side of the strip to the depth of the feedpoint selected. Each disk element is symmetrical about orthogonal major and minor axes, and preferably elliptical with a major axis not much greater than the minor axis. The feedpoint is selected to be between the major and minor axes for maximum efficiency of response to circularly polarized radiation incident upon the array; the precise impedance of the feedpoint is then selected to match the input impedance of the filter by insetting the connecting strip to a point inside the periphery of the array that is of proper impedance.

The novel features that are considered characteristic of this invention are set forth with particularity in the appended claims. This invention will best be understood from the following description when read in connection with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is an isometric view of a single disk element having a strip line filter, a discrete diode rectifier and a stripline impedance matching stub assembly connected to a feedpoint within the periphery of the disk element.

FIGS. *2a* and *2b* are sectional views of FIG. **1** along a line **2—2** showing two alternate ways in which a diode may be lodged in a dielectric layer and connected between a metal back plate and the microstrip circuit between a low-pass filter and a matching stub.

FIG. **3** is a plan view of a rectenna array for moderate power applications.

FIG. **4** is a plan view of a rectenna array for low power applications.

FIG. **5** is a plan view of a rectenna array for high power applications.

DESCRIPTION OF PREFERRED EMBODIMENTS

Referring now to the drawings, FIG. **1** shows a circularly polarized microstrip antenna element **10** of copper foil on a layer **11** of dielectric material, such as teflon-impregnated fiberglass, backed by a sheet **12** of conductive material, such as copper or aluminum, which serves as a ground plane. A microstrip low-pass filter **13** is coupled to the antenna element by a microstrip feed line **14**. Microstrip impedance matching stubs **15** are connected to the filter element by a microstrip conductor **16** in order to match the impedance of the filter to a low impedance load for the first and all even and odd harmonics of the frequency of the radiant energy being received. A diode D_1 is connected between the strip conductor **16** and the ground plane **12** at a point **17** of the strip conductor that is spaced a distance from the matching stubs and the low-pass filter to optimize coupling into the low impedance diode. The strip conductor **16** between the filter and the diode is selected to be of a width and length that will match the impedance of the diode.

FIGS. **2***a* and **2***b* illustrate alternate ways in which a diode, is inserted through the dielectric layer **11** and connected between the conductive strip **16** and the ground plane **12**. In FIG. **2***a*, the diode is encapsulated in dielectric material **18**, such as glass or ceramic, so that inserting and connecting the diode is a simple matter of drilling a hole through the microstrip structure at the appropriate location, bending the diode leads and soldering or otherwise fusing the leads to the conductive strip **16** and the ground plane **12**. In FIG. **2***b*, the diode is encapsulated in a ceramic body **18'** between metal ends **18***a* and **18***b*. The end **18***b* is threaded to screw into tapping threads in the ground plane **12**. In either case, the diode is inserted and connected with the polarity desired for the voltage to be derived from the microwave beamed power. For the polarity indicated in FIGS. **2***a* and **2***b*, the ground plane is the negative DC output terminal. The microstrip etched strip **16** of conductive material forms the positive DC output terminal. Since the sheet of conductive material from which the strip **16** is etched is everywhere at the same DC potential, it is possible to connect a lead anywhere, but it is preferable to connect the current collection lead in the form of a conductive strip **35** (FIG. **3**) connecting to the quarter wavelength interconnecting line **19** of microwave power conversion units similar to the one unit shown in FIG. **1**.

Included in the matching stubs **15** is an outer stub of a length selected to be λ/4 for the first harmonic, i.e., for the fundamental frequency of the received microwave power beam, and an inner stub of a length selected to be λ/4 for the second harmonic. The length of the conductive strip **19** from the matching stubs to the load (other antenna element or a power bus) is selected to be a quarter wavelength. In the case of connecting disk antenna elements in cascade, each with its own filter, rectifier and matching stubs, the length of the conductive strip **19** to the next antenna element **10** is selected to be λ/4.

The feed line **14** is connected to the antenna element **10** at a location midway between the major and minor axes such as to yield maximum response to circular polarized radiation. The impedance drops to a minimum as the connecting point is shifted toward the center of the disk (or the major axis as described in the aforesaid U.S. Pat. No. 3,803,623). Improved matching is achieved in accordance with one feature of this invention by setting in the feed element inside the periphery of the antenna element to a point where the impedance matches the characteristic impedance of the microstrip filter. For example, if the impedance of the antenna element is found to be 75 ohms on the periphery where the feed element is selected to be connected for laying out an array pattern, and the characteristic impedance of the microstrip filter is 50 ohms, the feed element is simply set in, as shown in FIG. **1**, to a point where the impedance of the antenna element is found to be 50 ohms. The more the feed element is set in, the lower the impedance. This facilitates impedance matching without significantly affecting the efficiency of the antenna element.

A rectenna assembly using microstrip technology and a circularly polarized (disk) antenna is useful for electrically powered vehicles, such as aircraft, because it virtually eliminates losses due to polarization angle misalignment. The microstrip implementation using an etched copper foil on a thin dielectric substrate provides sufficient flexibility to conform with vehicle structures such as aircraft wings or fuselage. However, for such applications as electrically powered vehicles, considerably greater power is required than can be received by one or two antenna elements. FIGS. **3**, **4** and **5** illustrate arrays for moderate, low and high power applications, respectively. These large arrays can be conformed to the configuration of the metal surface of the aircraft. The ground plane can be mounted integrally with that metal surface, but in most cases would be an aluminum or copper sheet or foil backing on the dielectric connected electrically to the metal surface of the vehicle. An RF transparent coating can be applied over the microstrip antenna array for aerodynamic flow and protection of etched circuit elements.

Referring now to FIG. **3**, which utilizes this novel combination of microstrip and rectenna technologies with circularly polarized antenna elements, a fragmentary section **30** of an array of rectenna units comprised of antenna disk elements **31**, each having an inset feed line **32**, a conductive strip filter **33**, matching stubs **34** and a rectifying diode indicated by an "X", all as described with reference to FIG. **1**. The rectenna units formed as shown and described in FIG. **1**, are DC connected in cascade. A bus **35** along one edge of the array is used to connect all cascaded units in parallel.

The DC connection from one unit to the next in cascade is made by a conductive strip **36** of a length selected to be λ/4, as noted hereinbefore, which has as its function the provision of an RF isolated DC connection from the DC output of one unit to the disk element of the next. The conductive strip from the last unit connected to the bus bar **35** is likewise an RF isolated DC connection the length of which is selected to be a quarter wavelength long. The disk elements are spaced with their centers approximately 6/10 λ apart in every direction, where λ is measured in free space. The major axis of each disk element is selected to be approximately half a wavelength of the received RF energy as measured in the dielectric layer.

FIG. **4** illustrates a fragmentary section of a rectenna for low power applications in which the disk antenna elements **41** are spaced as before with their centers spaced approximately 6/10 λ and with feed lines **42** set in as before. A microstrip conductive pattern shown connects the feed lines to a conductive strip filter **43**. The output of the filter is coupled to matching stubs **44**, and a diode indicated by an X rectifies the output of the diode as in the single element rectenna of FIG. **1**. The feed lines **42** are set in as necessary for RF impedance matching with the input of the filter. The DC at the output of the filter is connected to a load through a thin, high impedance conductive strip **45** at any convenient point in the microstrip pattern, such as the geometric center of the pattern. All points of the microstrip pattern are at the same DC potential with respect to the ground plane, so the main concern other than RF impedance matching, in connecting together all of the feed elements **42** at the input of the filter, and in connecting the rectified and filtered output to a load, is the resistance of the microstrip lines to the total DC current received. That total will be limited for low power applications by the capacity of the single rectifying diode, which may range up to 50 watts.

For high power applications, a rectenna constructed in accordance with the present invention would be as shown for a fragmentary section **50** of an array shown in FIG. **5**. Antenna elements **51** are arranged as before. A feed line **52** is inset as before and connected to a microstrip filter **53**. The output conductive strip of the filter is bifurcated, and the branches are connected to matching stubs **54** and **55**. In each branch there is a rectifying diode indicated by an X. The rectified and filtered DC output of one rectenna unit is DC connected to a second feed line **56** inset as the first feed line **52**, and connected to a second conductive strip filter **57**. The output of the second conductive strip filter **57** is connected to a bifurcated conductive strip, each branch having a rectifying diode. It should be noted that one branch shares the matching stubs **55** while the other has its own matching stubs **58**. The result is a symmetrical arrangement of two feed lines connected to separate filters with two rectifying diodes at the output of each filter for a total of four rectifying diodes per disk element. The power capacity is thus four times that of the moderate-power arrangement of FIG. **3**.

Limitations on the high flux density that may be received by the arrangement of FIG. **5** also include thermal dissipation as well as breakdown of the diodes, while the limitation on the low flux density by the arrangement of FIG. **4** is the directivity of the angular region within which RF energy is received due to any increase in the number of antenna disk elements arrayed. In other words, for the low flux density variant of FIG. **4**, more elements would be desired, but the greater directivity that results limits the use of that variant because the array must then be pointed more accurately toward the low flux RF source in order to receive maximum power. Therefore it would be desirable to collect the incident RF power at a number of separate rectenna arrays of FIG. **4**, each with a reasonably small nunmber of disk elements, and to sum the outputs of the separate rectenna arrays. Thus, for low power a number of disk elements may be arrayed and summed via proper RF lines of proper lengths and characteristic impedances before presenting the summed RF energy to a conductive strip filter and diode. This allows the diode to operate at its most efficient conversion level of RF into DC output power. For the other extreme of high flux density, the arrangement whereby multiple diodes are fed RF power from a single antenna element allows the high level RF power to be divided among several diodes thereby preventing excess RF input to an individual diode thus promoting long life and high conversion efficiency of high power RF to high power level DC output. By paralleling diodes or subdividing the RF output of the antenna disk element before applying it to a rectenna converter circuit, the flexible microstrip array can efficiently operate at high level incident RF flux densities. An operating limit due to waste heat temperature rise in the microstrip ground plane obviously exists, therefore maximum RF flux density limits do exist which depend upon the conversion efficiency and array cooling.

Although particular embodiments of the invention have been described and illustrated herein, it is recognized that other embodiments may readily occur to those skilled in the art as modifications of those illustrated, or equivalent to those illustrated. For example, in the high-power arrangement of FIG. **5**, the number of diodes connected in parallel to an antenna disk element is only limited by the skill with which the microstrip layout may be arranged to add other feed lines, filters, diodes and matching stubs without individual rectifying and filtering circuits touching, or coming into too close proximity. Thus as many as two, four, six, or even eight diodes per disk element are possible although only four diodes per disk element are shown in FIG. **5**. Consequently, it is intended that the claims be interpreted to cover such modifications and equivalents.

What is claimed is:

1. An antenna array for microwave power conversion comprising a metal sheet cladding a layer of dielectric material, said layer of dielectric material being disposed over a metal backing sheet, and a plurality of circularly polarized antenna elements fabricated on said metal cladding sheet by etching said metal cladding sheet in a pattern which includes conductive strips connecting said antenna elements to a common load terminal and between said antenna elements and said load terminal, at least one microstrip low-pass filter connected to low impedance matching stubs and lodging a rectifying diode in said dielectric material with one terminal connected between said filter and matching stubs, and the other terminal connected to said metal backing sheet.

2. An antenna array for microwave power conversion as defined in claim **1** wherein each conductive strip connected to an antenna element as a feed line to a filter is inset to a point inside the periphery of the antenna element of proper impedance to match the impedance of the filter.

3. An antenna array as defined in claim **2** for low microwave power conversion wherein all antenna elements are connected in parallel to said load through a single microstrip low-pass filter connected to matching stubs, and a single rectifying diode lodged in said dielectric material with one terminal connected to said conductive strip between said filter and said low impedance matching stubs, and the other terminal connected to a metal backing sheet.

4. An antenna array as defined in claim **2** wherein all of said antenna elements are connected in series-parallel with a plurality of groups of said antenna elements connected in cascade with each feeding into a separate microstrip filter connected to separate matching stubs, and with a separate diode connected between the microstrip filter and the matching stubs, and wherein the last of each group of antenna elements so cascaded is connected to a common load through a last microstrip filter connected to matching stubs, and with a separate diode connected between said last microstrip filter and connected matching stubs.

5. An antenna array as defined in claim **4** for high power conversion wherein each of said antenna elements is connected to a plurality of microstrip filters, matching stubs and rectifying diodes.

6. An antenna array for microwave power conversion comprising a plurality of circularly polarized antenna elements fabricated on a metal front sheet cladding a layer of dielectric material over a metal backing sheet by etching said metal front sheet in a pattern which includes conductive strips connecting the said antenna elements to a common load terminal, each antenna element having at least one microstrip low-pass filter connected to low impedance matching stubs, and at least one rectifying diode lodged in said dielectric material with one terminal connected to said conductive strip between said filter and low impedance matching stubs, and the other terminal connected to said metal backing sheet, each conductive strip feeding from an antenna element into a filter being inset within the periphery of the antenna element to a point of proper impedance to match the input impedance of the filter.

7. An antenna array as defined in claim 6 wherein each antenna element is symmetrical about orthogonal major and minor axes, and the place for the connecting conductive strip is selected to be between the major and minor axes for maximum response to circularly polarized radiation incident upon the array and the precise impedance of the selected feedpoint is then selected to match the input impedance of the filter by insetting the connecting strip to a point inside the periphery of the array to a point that is of proper impedance.

* * * * *

US007456803B1

(12) **United States Patent**
Sievenpiper

(10) **Patent No.: US 7,456,803 B1**
(45) **Date of Patent: Nov. 25, 2008**

(54) **LARGE APERTURE RECTENNA BASED ON PLANAR LENS STRUCTURES**

(75) Inventor: **Daniel F. Sievenpiper**, Santa Monica, CA (US)

(73) Assignee: **HRL Laboratories, LLC**, Malibu, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 101 days.

(21) Appl. No.: **11/594,350**

(22) Filed: **Nov. 7, 2006**

Related U.S. Application Data

(62) Division of application No. 10/944,032, filed on Sep. 17, 2004, now Pat. No. 7,154,451.

(60) Provisional application No. 60/470,028, filed on May 12, 2003, provisional application No. 60/470,027, filed on May 12, 2003.

(51) **Int. Cl.**
H01Q 15/02 (2006.01)

(52) **U.S. Cl.** .... **343/909**; 343/700 MS; 343/705; 343/753

(58) **Field of Classification Search** .... 343/705, 343/708, 700 MS, 753, 754, 756, 909
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,267,480	A	8/1966	Lerner	343/911
3,560,978	A	2/1971	Himmel et al.	343/106
3,810,183	A	5/1974	Krutsinger et al.	343/708
3,961,333	A	6/1976	Purinton	343/872
4,045,800	A	8/1977	Tang et al.	343/854
4,051,477	A	9/1977	Murphy et al.	343/700 MS
4,119,972	A	10/1978	Fletcher et al.	343/844
4,123,759	A	10/1978	Hines et al.	343/854
4,124,852	A	11/1978	Steudel	343/854
4,127,586	A	11/1978	Rody et al.	260/308 B
4,150,382	A	4/1979	King	343/754
4,173,759	A	11/1979	Bakhru	343/100
4,189,733	A	2/1980	Malm	343/100 SA
4,217,587	A	8/1980	Jacomini	343/100 SA
4,220,954	A	9/1980	Marchand	343/113 R
4,236,158	A	11/1980	Daniel	343/100 LE
4,242,685	A	12/1980	Sanford	343/770
4,266,203	A	5/1981	Saudreau et al.	333/21 A
4,308,541	A	12/1981	Seidel et al.	343/786
4,367,475	A	1/1983	Schiavone	343/767

(Continued)

FOREIGN PATENT DOCUMENTS

DE	196 00 609 A1	4/1997

(Continued)

OTHER PUBLICATIONS

Balanis, C., "Aperture Antennas," *Antenna Theory, Analysis and Design*, 2nd Edition, Ch. 12, pp. 575-597 (1997).

(Continued)

Primary Examiner—Tan Ho
(74) *Attorney, Agent, or Firm*—Ladas & Parry

(57) **ABSTRACT**

A rectenna structure comprising a flexible, dielectric sheet of material; a plurality of metallic lenslets disposed on the sheet of material; and a plurality of diodes disposed on the sheet of material, each diode in said plurality of diodes being arranged at a focus of a corresponding one of said plurality of metallic lenslets.

13 Claims, 7 Drawing Sheets

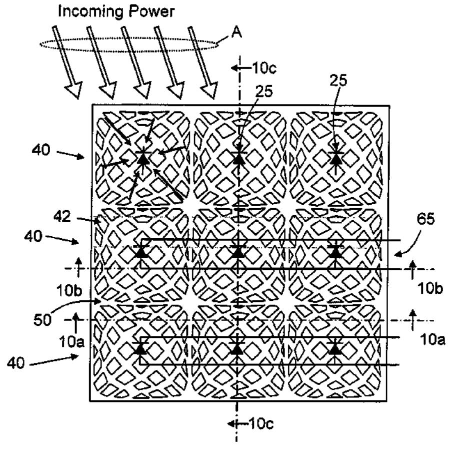

U.S. PATENT DOCUMENTS

4,370,659	A		1/1983	Chu et al.	343/772
4,387,377	A		6/1983	Kandler	343/756
4,395,713	A		7/1983	Nelson et al.	343/713
4,443,802	A		4/1984	Mayes	343/729
4,590,478	A		5/1986	Powers et al.	343/700 MS
4,594,595	A		6/1986	Struckman	343/770
4,672,386	A		6/1987	Wood	343/770
4,684,953	A		8/1987	Hall	343/725
4,700,197	A		10/1987	Milne	343/837
4,737,795	A		4/1988	Nagy et al.	343/712
4,749,966	A		6/1988	Tresselt	343/700 MS
4,760,402	A		7/1988	Mizuno et al.	343/713
4,782,346	A		11/1988	Sharma	343/795
4,803,494	A		2/1989	Norris et al.	343/770
4,821,040	A		4/1989	Johnson et al.	343/700 MS
4,835,541	A		5/1989	Johnson et al.	343/713
4,843,400	A		6/1989	Tsao et al.	343/700 MS
4,843,403	A		6/1989	Lalezari et al.	343/767
4,853,704	A		8/1989	Diaz et al.	343/767
4,903,033	A		2/1990	Tsao et al.	343/700 MS
4,905,014	A		2/1990	Gonzalez et al.	343/909
4,916,457	A		4/1990	Foy et al.	343/770
4,922,263	A		5/1990	Dubost et al.	343/797
4,958,165	A		9/1990	Axford et al.	343/770
4,975,712	A		12/1990	Chen	
5,021,795	A		6/1991	Masiulis	343/700 MS
5,023,623	A		6/1991	Kreinheder et al.	343/725
5,070,340	A		12/1991	Diaz	343/767
5,081,466	A		1/1992	Bitter, Jr.	343/767
5,115,217	A		5/1992	McGrath et al.	333/246
5,146,235	A		9/1992	Frese	343/895
5,148,182	A		9/1992	Gautier et al.	
5,158,611	A		10/1992	Ura et al.	106/499
5,208,603	A		5/1993	Yee	343/909
5,218,374	A		6/1993	Koert et al.	
5,235,343	A		8/1993	Audren et al.	343/816
5,268,696	A		12/1993	Buck et al.	342/372
5,268,701	A		12/1993	Smith	343/767
5,278,562	A		1/1994	Martin et al.	342/1
5,287,116	A		2/1994	Iwasaki et al.	343/700 MS
5,287,118	A		2/1994	Budd	343/909
5,402,134	A		3/1995	Miller et al.	343/742
5,406,292	A		4/1995	Schnetzer et al.	343/700 MS
5,519,408	A		5/1996	Schnetzer	343/767
5,525,954	A		6/1996	Komazaki et al.	333/219
5,531,018	A		7/1996	Saia et al.	29/622
5,532,709	A		7/1996	Talty	343/819
5,534,877	A		7/1996	Sorbello et al.	343/700 MS
5,541,614	A		7/1996	Lam et al.	343/792.5
5,557,291	A		9/1996	Chu et al.	343/725
5,581,266	A		12/1996	Peng et al.	343/770
5,589,845	A		12/1996	Yandrofski et al.	343/909
5,598,172	A		1/1997	Chekroun	
5,600,325	A		2/1997	Whelan et al.	342/13
5,611,940	A		3/1997	Zettler	73/514.16
5,619,365	A		4/1997	Rhoads et al.	359/248
5,619,366	A		4/1997	Rhoads et al.	359/248
5,621,571	A		4/1997	Bantli et al.	
5,638,946	A		6/1997	Zavracky	200/181
5,644,319	A		7/1997	Chen et al.	343/702
5,694,134	A		12/1997	Barnes	343/700
5,721,194	A		2/1998	Yandrofski et al.	505/210
5,767,807	A		6/1998	Pritchett	342/374
5,808,527	A		9/1998	De Los Santos	333/205
5,828,344	A	*	10/1998	Alder et al.	343/755
5,874,915	A		2/1999	Lee et al.	342/375
5,892,485	A		4/1999	Glabe et al.	343/789
5,894,288	A		4/1999	Lee et al.	343/770
5,905,465	A		5/1999	Olson et al.	343/700 MS
5,923,303	A		7/1999	Schwengler et al.	343/853
5,926,139	A		7/1999	Korisch	343/702
5,929,819	A		7/1999	Grinberg	343/754
5,943,016	A		8/1999	Snyder, Jr. et al.	343/700 MS
5,945,951	A		8/1999	Monte et al.	343/700 MS
5,949,382	A		9/1999	Quan	343/767
5,966,096	A		10/1999	Brachat	343/700 MS
5,966,101	A		10/1999	Haub et al.	343/767
5,991,474	A	*	11/1999	Baldi	385/15
6,005,519	A		12/1999	Burns	343/700 MS
6,005,521	A		12/1999	Suguro et al.	343/700 MS
6,008,770	A		12/1999	Sugawara	343/767
6,016,125	A		1/2000	Johansson	343/702
6,028,561	A		2/2000	Takei	343/767
6,028,692	A		2/2000	Rhoads et al.	359/245
6,034,644	A		3/2000	Okabe et al.	343/767
6,034,655	A		3/2000	You	345/60
6,037,905	A		3/2000	Koscica et al.	343/701
6,040,803	A		3/2000	Spall	343/700 MS
6,046,655	A		4/2000	Cipolla	333/137
6,046,659	A		4/2000	Loo et al.	333/362
6,054,659	A		4/2000	Lee et al.	200/181
6,061,025	A		5/2000	Jackson et al.	343/700 MS
6,075,485	A		6/2000	Lilly et al.	343/700 MS
6,081,235	A		6/2000	Romanofsky et al.	343/700 MS
6,081,239	A		6/2000	Sabet et al.	343/753
6,097,263	A		8/2000	Mueller et al.	33/17.1
6,097,343	A		8/2000	Goetz et al.	343/708
6,118,406	A		9/2000	Josypenko	343/700 MS
6,118,410	A		9/2000	Nagy	343/713
6,127,908	A		10/2000	Bozler et al.	333/246
6,150,989	A		11/2000	Aubry	343/767
6,154,176	A		11/2000	Fathy et al.	343/700 MS
6,166,705	A		12/2000	Mast et al.	343/853
6,175,337	B1		1/2001	Jasper, Jr. et al.	343/770
6,175,723	B1		1/2001	Rothwell, III	455/63
6,188,369	B1		2/2001	Okabe et al.	343/767
6,191,724	B1		2/2001	McEwan	342/21
6,198,438	B1		3/2001	Herd et al.	343/700 MS
6,198,441	B1		3/2001	Okabe et al.	343/702
6,204,819	B1		3/2001	Hayes et al.	343/702
6,218,912	B1		4/2001	Mayer	333/106
6,218,997	B1		4/2001	Lindenmeier et al.	343/725
6,246,377	B1		6/2001	Aiello et al.	343/700
6,252,473	B1		6/2001	Ando	333/105
6,285,325	B1		9/2001	Nalbandian et al.	343/700 MS
6,307,519	B1		10/2001	Livingston et al.	343/767
6,317,095	B1		11/2001	Teshirogi et al.	343/785
6,323,826	B1		11/2001	Sievenpiper et al.	343/909
6,331,257	B1		12/2001	Loo et al.	216/13
6,337,668	B1		1/2002	Ito et al.	343/833
6,366,254	B1		4/2002	Sievenpiper et al.	343/700
6,373,349	B2		4/2002	Gilbert	333/126
6,380,895	B1		4/2002	Moren et al.	343/700 MS
6,388,631	B1		5/2002	Livingston et al.	343/767
6,392,610	B1		5/2002	Braun et al.	343/876
6,404,390	B2		6/2002	Sheen	343/700 MS
6,404,401	B2		6/2002	Gilbert et al.	343/780
6,407,719	B1		6/2002	Ohira et al.	343/893
6,417,807	B1		7/2002	Hsu et al.	343/700 MS
6,424,319	B2		7/2002	Ebling et al.	343/911 L
6,426,722	B1		7/2002	Sievenpiper et al.	343/700 MS
6,440,767	B1		8/2002	Loo et al.	438/52
6,473,362	B1		10/2002	Gabbay	367/119
6,483,480	B1		11/2002	Sievenpiper et al.	343/909
6,496,155	B1		12/2002	Sievenpiper et al.	343/770
6,515,635	B2		2/2003	Chiang et al.	343/834
6,518,931	B1		2/2003	Sievenpiper	343/700
6,525,695	B2		2/2003	McKinzie, III	343/756
6,538,621	B1		3/2003	Sievenpiper et al.	343/909
6,552,696	B1		4/2003	Sievenpiper et al.	343/909
6,624,720	B1		9/2003	Allison et al.	333/105
6,642,889	B1		11/2003	McGrath	343/700 MS
6,657,525	B1		12/2003	Dickens et al.	335/78
6,741,207	B1		5/2004	Allison et al.	

6,822,622	B2	11/2004	Crawford et al.	
6,864,848	B2	3/2005	Sievenpiper	343/767
6,897,810	B2	5/2005	Dai et al.	
7,026,993	B2	4/2006	Zhang et al.	
7,154,451	B1 *	12/2006	Sievenpiper	343/909
2001/0035801	A1	11/2001	Gilbert	333/126
2002/0036586	A1	3/2002	Gothard et al.	342/374
2003/0112186	A1	6/2003	Sanchez et al.	
2003/0122721	A1	7/2003	Sievenpiper	
2003/0193446	A1	10/2003	Chen	343/893
2003/0222738	A1	12/2003	Brown et al.	
2003/0227351	A1	12/2003	Sievenpiper	333/105
2004/0113713	A1	6/2004	Zipper et al.	
2004/0135649	A1	7/2004	Sievenpiper	333/105
2004/0227583	A1	11/2004	Shaffner et al.	333/32
2004/0227667	A1	11/2004	Sievenpiper	343/700 MS
2004/0227678	A1	11/2004	Sievenpiper	343/702
2004/0228668	A1	11/2004	Sievenpiper	343/700 MS
2004/0263408	A1	12/2004	Sievenpiper et al.	343/757
2005/0012667	A1	1/2005	Noujem	
2006/0044199	A1	3/2006	Furuhi et al.	

FOREIGN PATENT DOCUMENTS

EP	0 539 297		4/1993
EP	1 158 605	A1	11/2001
FR	2 785 476		5/2000
GB	1145208		3/1969
GB	2 281 662		3/1995
GB	2 328 748		3/1999
JP	61-260702		11/1986
WO	94/00891		1/1994
WO	96/29621		9/1996
WO	98/21734		5/1998
WO	99/50929		10/1999
WO	00/44012		7/2000
WO	01/73891	A1	10/2001
WO	01/73893	A1	10/2001
WO	03/009501		1/2003
WO	03/098732	A1	11/2003

OTHER PUBLICATIONS

Balanis, C., "Microstrip Antennas," *Antenna Theory, Analysis and Design*, 2nd Edition, Ch. 14, pp. 722-736 (1997).

Brown, W.C., "The History of Power Transmission by Radio Waves," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-32, No. 9, pp. 1230-1242 (Sep. 1984).

Chen, P.W., et al., "Planar Double-Layer Leaky-Wave Microstrip Antenna," *IEEE Transactions on Antennas and Propagation*, vol. 50, pp. 832-835 (2002).

Chen, Q., et al., "FDTD diakoptic design of a slot-loop antenna excited by a coplanar waveguide," *Proceedings of the 25th European Microwave Conference 1995*, vol. 2, Conf. 25, pp. 815-819 (Sep. 4, 1995).

Fay, P., "High-Performance Antimonide-Based Heterostructure Backward Diodes for Millimeter-Wave Detection," *IEEE Electron Device Letters*, vol. 23, No. 10, pp. 585-587 (Oct. 2002).

Gianvittorio, J.P., et al., "Reconfigurable MEMS-enabled Frequency Selective Surfaces," *Electronic Letters*, vol. 38, No. 25, pp. 1627-1628 (Dec. 5, 2002).

Gold, S.H.,et al., "Review of High-Power Microwave Source Research," *Rev. Sci. Instrum.*, vol. 68, No. 11, pp. 3945-3974 (Nov. 1997).

Jensen, M.A., et al., "Performance Analysis of Antennas for Hand-Held Transceivers Using FDTD," *IEEE Transactions on Antennas and Propagation*, vol. 42, No. 8, pp. 1106-1113 (Aug. 1994).

Koert, P., et al., "Millimeter Wave Technology for Space Power Beaming," *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, No. 6, pp. 1251-1258 (Jun. 1992).

Lee, J.W., et al., "TM-Wave Reduction From Grooves In A Dielectric-Covered Ground Plane," *IEEE Transactions on Antennas and Propagation*, vol. 49, No. 1, pp. 104-105 (Jan. 2001).

Lezec, H.J., et al., "Beaming Light from a Subwavelength Aperture," *Science*, vol. 297, pp. 820-821 (Aug. 2, 2002).

Linardou, I., et al., "Twin Vivaldi Antenna Fed By Coplanar Waveguide," *Electronics Letters*, vol. 33, No. 22, pp. 1835-1837 (1997).

Malherbe, A., et al., "The Compenasation of Step Discontinues in TEM-Mode Transmission Lines," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-26, No. 11, pp. 883-885 (Nov. 1978).

Maruhashi, K., et al., "Design and Performance of a Ka-Band Monolithic Phase Shifter Utilizing Nonresonant FET Switches," *IEEE Transactions on Microwave Theory and Techniques*, vol. 48, No. 8, pp. 1313-1317 (Aug. 2000).

McSpadden, J.O.,et al., "Design and Experiments of a High-Conversion-Efficiency 5.8-GHz Rectenna," *IEEE Transactions on Microwave Theory and Techniques*, vol. 46, No. 12, pp. 2053-2060 (Dec. 1998).

Oak, A.C., et al. "A Varactor Tuned 16 Element MESFET Grid Oscillator," *Antennas and Propagation Society International Symposium*. pp. 1296-1299 (1995).

Perini, P., et al., "Angle and Space Diversity Comparisons in Different Mobile Radio Environments," *IEEE Transactions on Antennas and Propagation*, vol. 46, No. 6, pp. 764-775 (Jun. 1998).

Ramo, S., et al., *Fields and Waves in Communication Electronics*, 3rd Edition, Sections 9.8-9.11, pp. 476-487 (1994).

Rebeiz, G.M., et al., "RF MEMS Switches and Switch Circuits," *IEEE Microwave Magazine*, pp. 59-71 (Dec. 2001).

Schaffner, J., et al., "Reconfigurable Aperture Antennas Using RF MEMS Switches for Multi-Octave Tunability and Beam Steering," *IEEE Antennas and Propagation Society International Symposium, 2000 Digest*, vol. 1, of 4, pp. 321-324 (Jul. 16, 2000).

Schulman, J.N., et al., "Sb-Heterostructure Interband Backward Diodes," *IEEE Electron Device Letters*, vol. 21, No. 7, pp. 353-355 (Jul. 2000).

Semouchkina, E., et al., "Numerical Modeling and Experimental Study of A Novel Leaky Wave Antenna," *Antennas and Propagation Society*, IEEE International Symposium, vol. 4, pp. 234-237 (2001).

Sievenpiper, D., et al., "Beam Steering Microwave Reflector Based On Electrically Tunable Impedance Surface," *Electronics Letters*, vol. 38, No. 21, pp. 1237-1238 (Oct. 1, 2002).

Sievenpiper, D., et al., "Eliminating Surface Currents With Metallodielectric Photonic Crystals," *1998 MTT-S International Microwave Symposium Digest*, vol. 2, pp. 663-666 (Jun. 7, 1998).

Sievenpiper, D., et al., "Low-Profile, Four-Sector Diversity Antenna On High-Impedance Ground Plane," *Electronics Letters*, vol. 36, No. 16, pp. 1343-1345 (Aug. 3, 2000).

Sievenpiper. D.F., et al., "Two-Dimensional Beam Steering Using an Electrically Tunable Impedance Surface," *IEEE Transactions on Antennas and Propagation*, vol. 51, No. 10, pp. 2713-2722 (Oct. 2003).

Sor, J., et al., "A Reconfigurable Leaky-Wave/Patch Microstrip Aperture For Phased-Array Applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, No. 8, pp. 1877-1884 (Aug. 2002).

Strassner, B., et al., "5.8-GHz Circularly Polarized Rectifying Antenna for Wireless Microwave Power Transmission," *IEEE Transactions on Microwave Theory and Techniques*,vol. 50, No. 8, pp. 1870-1876 (Aug. 2002).

Swartz, N., "Ready for CDMA 2000 1xEV-Do?," *Wireless Review*, 2 pages total (Oct. 29, 2001).

Vaughan, Mark J., et al., "InP-Based 28 Gh$_z$ Integrated Antennas for Point-to-Multipoint Distribution," *Proceedings of the IEEE/Cornell Conference on Advanced Concepts in High Speed Semiconductor Devices and Circuits*, pp. 75-84 (1995).

Vaughan, R., "Spaced Directive Antennas for Mobile Communications by the Fourier Transform Method," *IEEE Transactions on Antennas and Propagation*, vol. 48, No. 7, pp. 1025-1032 (Jul. 2000).

Wang, C.J., et al., "Two-Dimensional Scanning Leaky-Wave Antenna by Utilizing the Phased Array," *IEEE Microwave and Wireless Components Letters*, vol. 12, No. 8, pp. 311-313, (Aug. 2002).

Wu, S.T., et al., "High Birefringence and Wide Nematic Range Bis-Tolane Liquid Crystals," *Appl. Phys. Lett.*, vol. 74, No. 5, pp. 344-346 (Jan. 18, 1999).

Yang, F.R., et al., "A Uniplanar Compact Photonic-Bandgap(UC-PBG) Structure and Its Applications for Microwave Circuits," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, No. 8, pp. 1509-1514 (Aug. 1999).

Yang, Hung-Yu David, et al., "Theory of Line-Source Radiation From A Metal-Strip Grating Dielectric-Slab Structure," *IEEE Transactions on Antennas and Propagation*, vol. 48, No. 4, pp. 556-564 (2000).

Bialkowski, M.E., et al., "Electronically steered antenna system for the Australian Mobilesat," IEEE Proc. Microw. Antennas Proag., vol. 143, No. 4, (Aug. 1996).

Bradley, T.W., et al., Development Of Voltage-Variable Dielectric (VVD), Electronics Scan Antenna, Radar 97, (Oct. 14-16, 1997)

Cognard, J., "Alignment of Nematic Liquid Crystals and their Mixtures", Molecular Crystals Supplement Series (Table of Contents), date is not available.

Doane, J.W., et al., "Field Control light scattering from nematic microdroplets," Appl. Phys. Lett. 38 (4) (Jan. 27, 1986).

Grbic, A., et al., "Experimental verification of backward-wave radiation from a negative refractive index metamaterial," Journal of Applied Physics, vol. 92, No. 10, (Nov. 15, 2002).

Hu, Cheng-Nan, et al., "Analysis and design of large leaky-mode array employing the coupled-mode approach," IEEE Transactions on Microwave Theory and Techniques, vol. 49, No. 4, (Apr. 2001).

Jablonski, W., et al., "Microwave Schottky Diode With Beam-Lead Contacts, "Institute of Electron Technology, date is not available.

Jensen, M.A., et al., "EM Interaction of Handset Antennas and a Human in Personal Communications," Proceedings of the IEEE, vol. 83, No. 1, (Jan. 1995).

Sievenpiper, D., "High-Impedance Electromagnetic Surfaces," dissertation (1999).

Sievenpiper, D., "High-Impedance Electromagnetic Surfaces With a Forbidden Frequency Band," IEEE Transactions on Microwave Theory and Techniques, vol. 47, No. 11, (Nov. 1999).

Yashchyshyn, Y., et al., "The Leaky-Wave Antenna With Ferroelectric Substrate," Institute eof Radioelectronics, date is not available.

* cited by examiner

FIG. 1

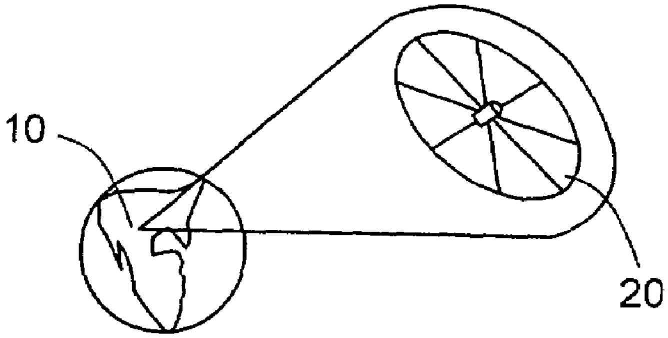

FIG. 2

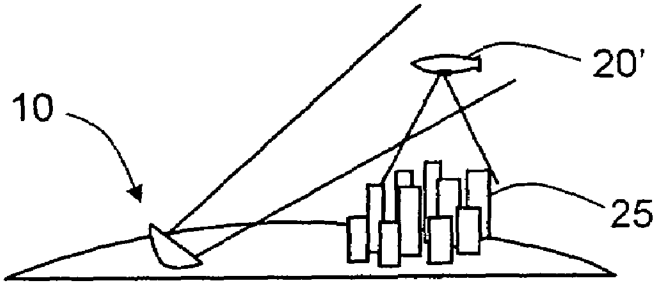

FIG. 3

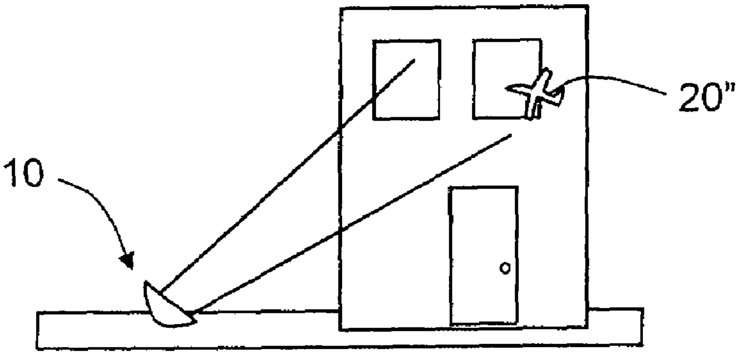

FIG. 4

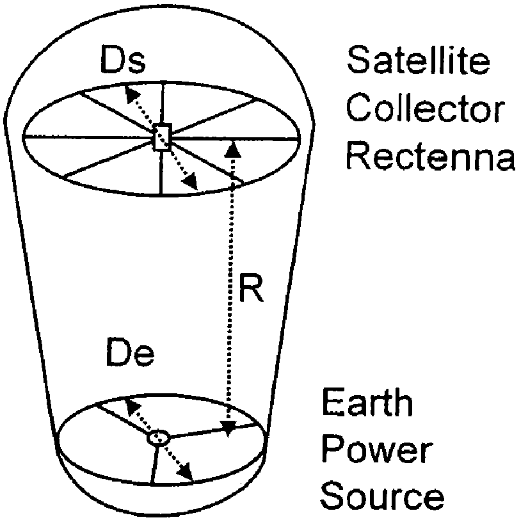
Ds
Satellite
Collector
Rectenna
R
De
Earth
Power
Source

FIG. 5

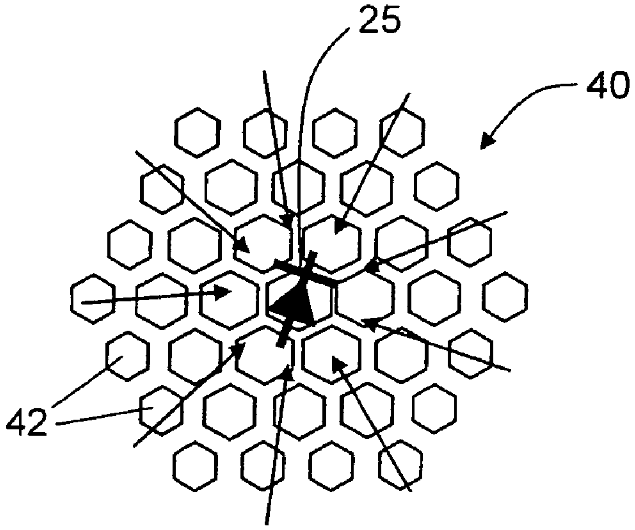

FIG. 6

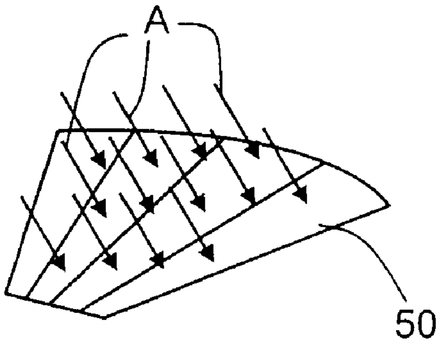

FIG. 7

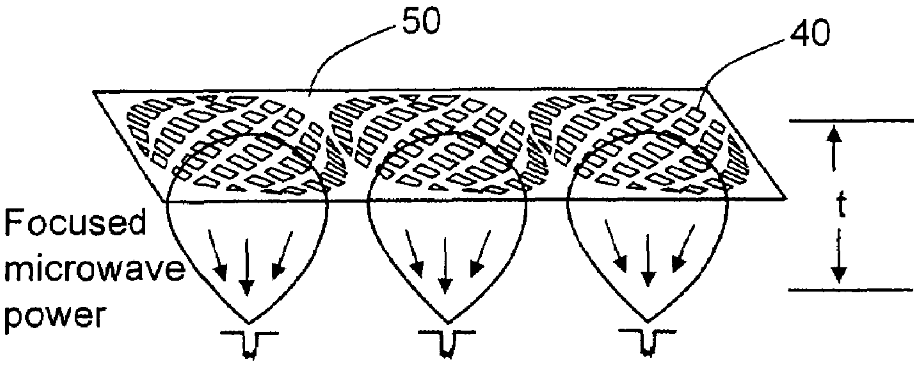
50
40
Focused
microwave
power
t

FIG. 8a

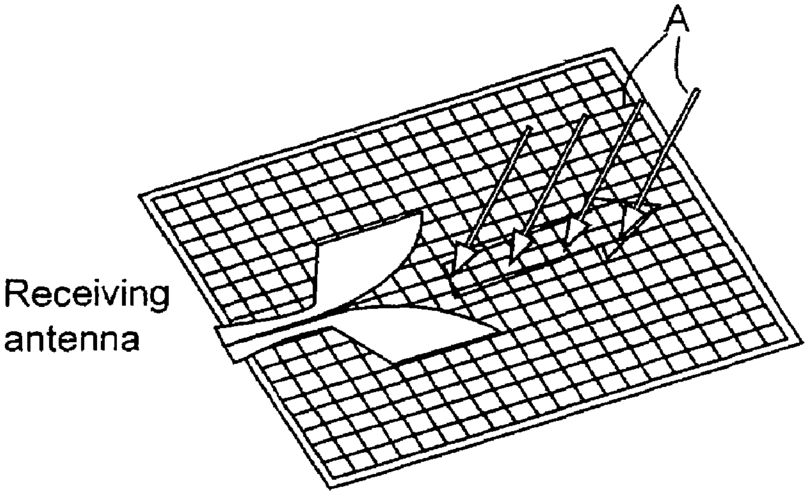
A
Receiving
antenna

FIG. 8b

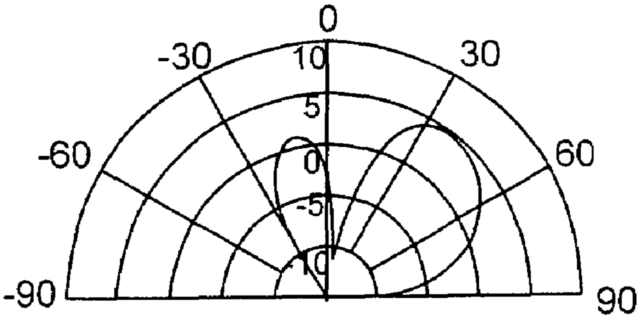

FIG. 8c

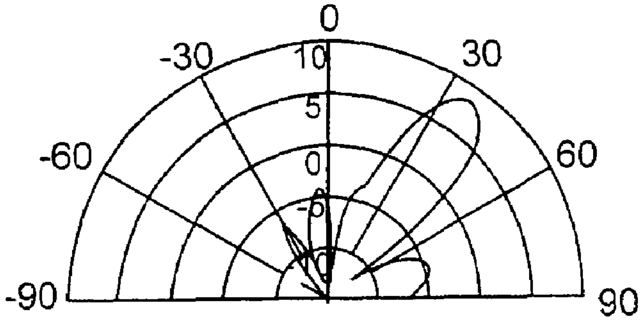

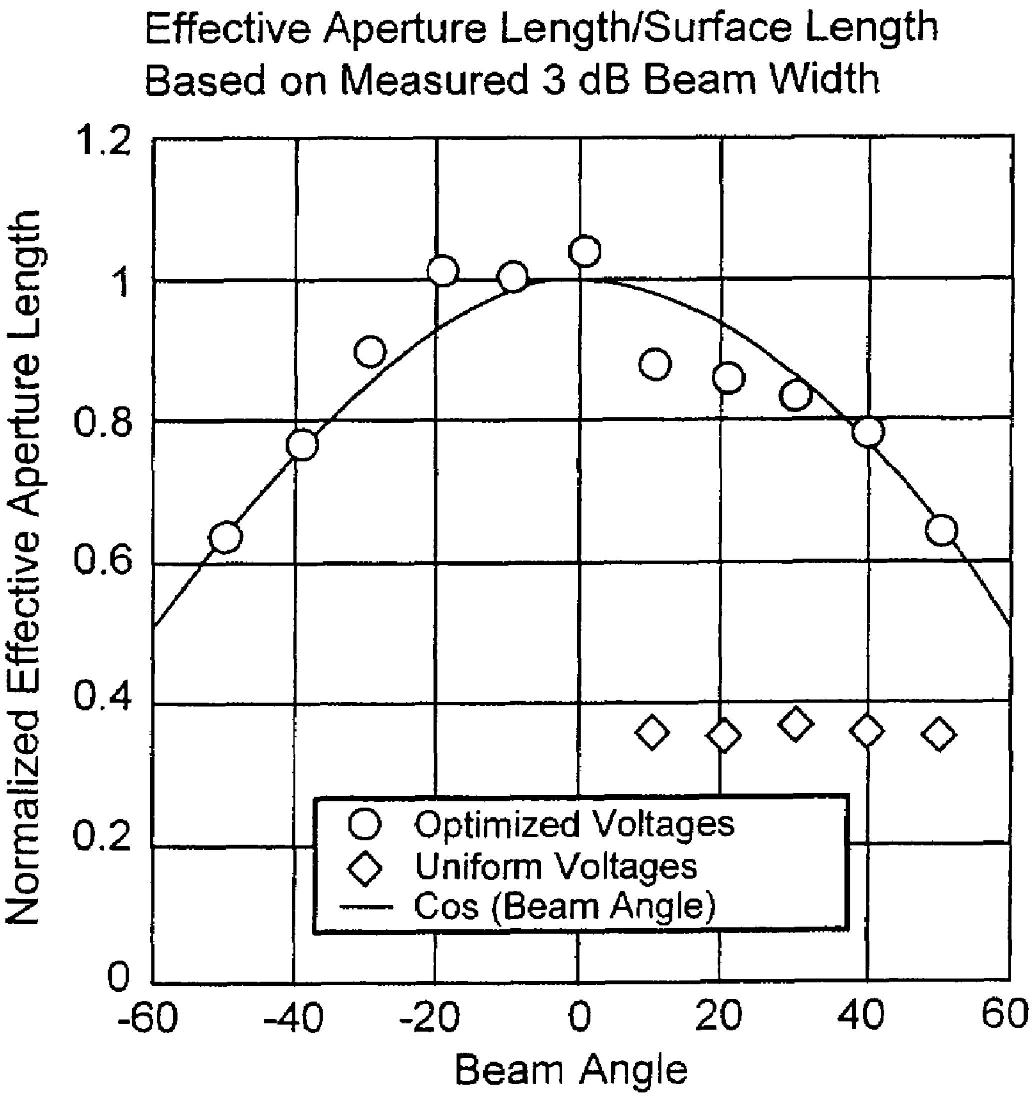
Effective Aperture Length/Surface Length
Based on Measured 3 dB Beam Width
Normalized Effective Aperture Length
1.2
1
0.8
0.6
0.4
0.2
0
-60
-40
-20
0
20
40
60
Beam Angle
Optimized Voltages
Uniform Voltages
Cos (Beam Angle)

FIG. 9

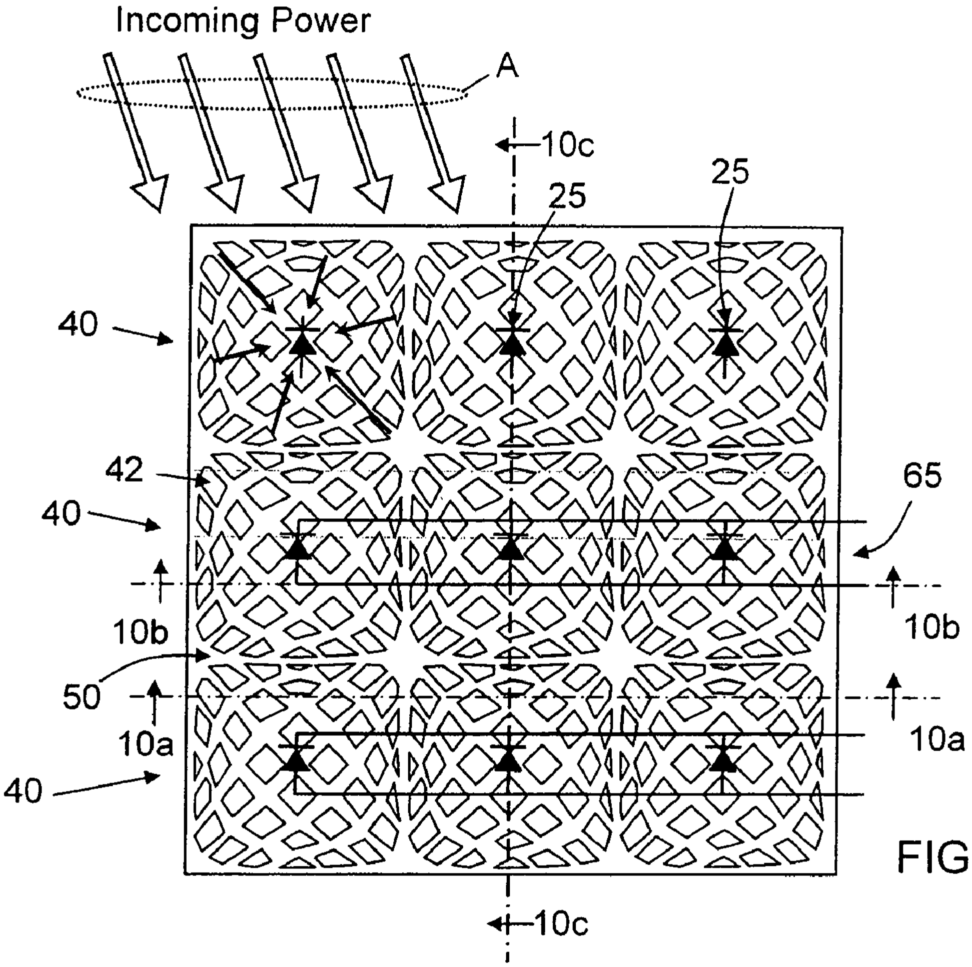
Incoming Power
A
10c
25
25
40
42
65
40
10b
10b
50
10a
10a
40
10c

FIG. 10

FIG. 10a

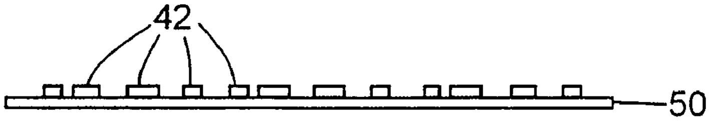

FIG. 10b

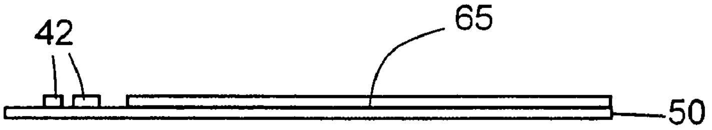

FIG. 10c

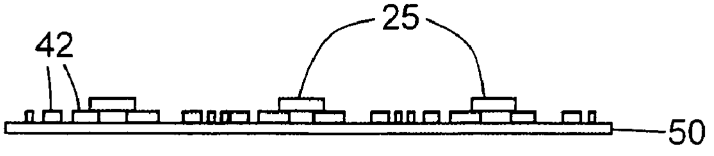

FIG. 11

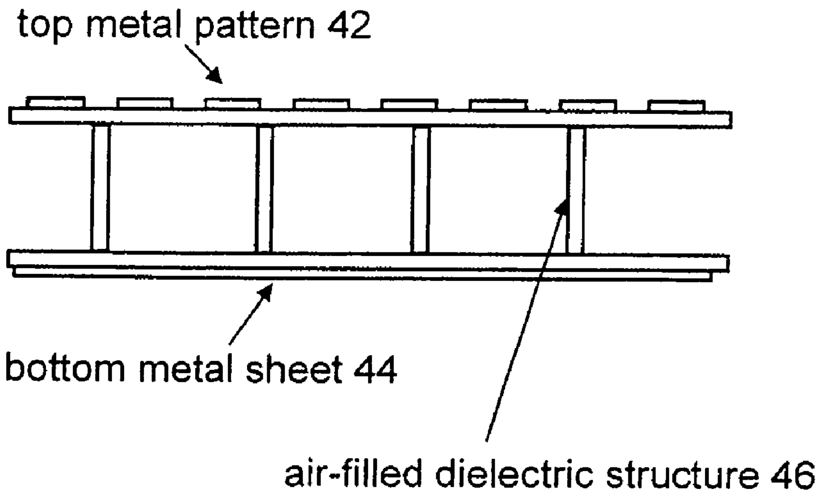
top metal pattern 42
bottom metal sheet 44
air-filled dielectric structure 46

FIG. 12

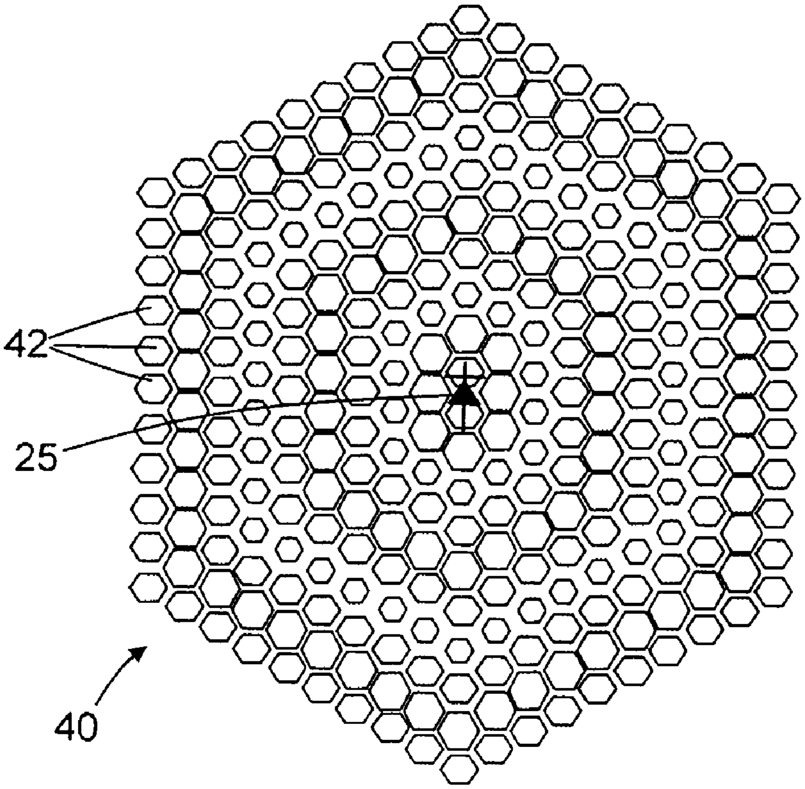

FIG. 13

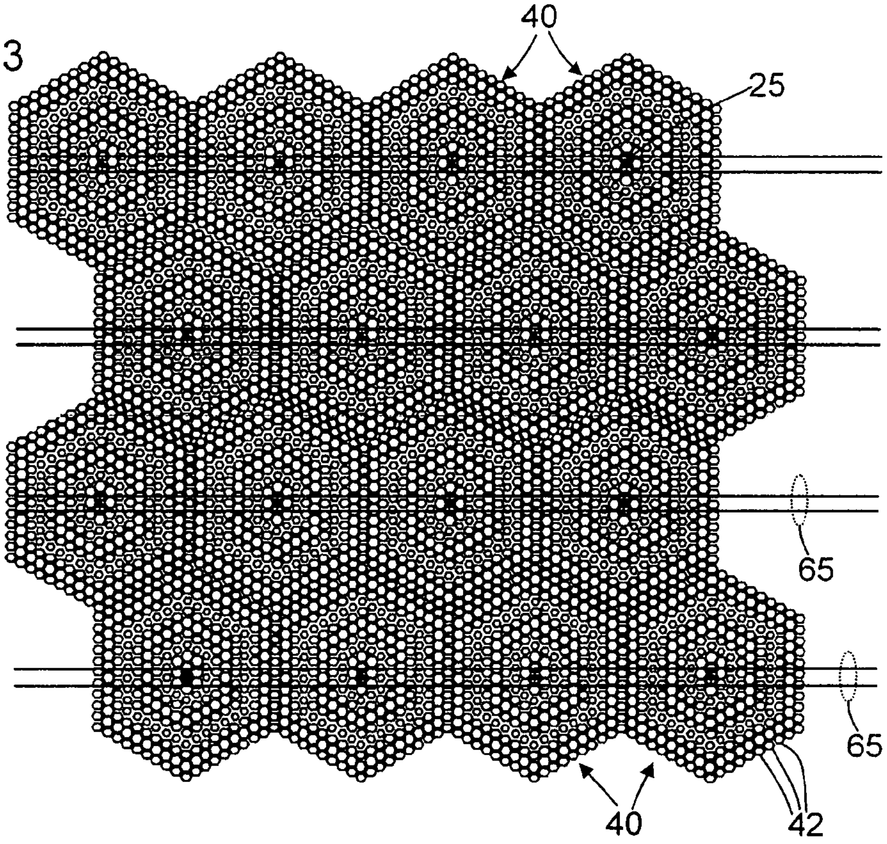
40
25
65
65
40
42

FIG. 14

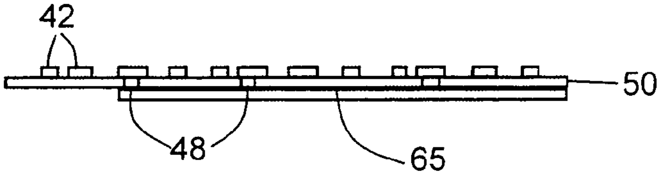

LARGE APERTURE RECTENNA BASED ON PLANAR LENS STRUCTURES

CROSS-REFERENCE TO RELATED APPLICATIONS

This patent application is a divisional application of U.S. application Ser. No. 10/994,032, filed Sep. 17, 2004 now U.S. Pat. No. 7,154,451 which claims the benefit of U.S. Provisional Application Ser. No. 60/470,027 entitled “Meta-element Antenna and Array” filed May 12, 2003 and to U.S. Patent Application Ser. No. 60/470,028 entitled “Steerable Leaky Wave Antenna Capable for Both Forward and Backward Radiation” filed May 12, 2003. The disclosures of these applications are hereby incorporated herein by reference. This disclosure is also related to two non-provisional applications that were filed claiming the benefit of the aforementioned applications. The two non-provisional applications have Ser. Nos. 10/792,411 (now U.S. Pat. No. 7,068,234 issued Jun. 27, 2006) and 10/792,412 and were both filed on Mar. 2, 2004. The disclosures of these two non-provisional applications are also incorporated herein by reference.

TECHNICAL FIELD

The technology disclosed herein relates to a lightweight, high-efficiency rectenna and to a method or architecture for making same. Rectennas can be useful for a variety of applications in the field of beaming RF power, which can be useful for satellites, zeppelins, and UAVs.

BACKGROUND OF THE INVENTION

Rectennas are antenna structures that intentionally incorporate rectifying elements in their designs.

Satellites are an integral part of modern communication systems, and their importance can be expected to grow in the coming years. As future generations of satellites with greater capabilities become possible, it is expected that they can take an even more active role in future military conflicts.

The design of present-day satellites often involves tradeoffs among such aspects as weight, power, and electronic capabilities. Each new electronic system adds weight, and must compete for power with other required systems such as station keeping. The limits of these tradeoffs are eased only gradually from one generation to the next, by the evolution of electronics, batteries, propulsion systems, and so on. Thus, developing new technologies that significantly expand the available design space is crucial to the enablement of satellites with radically improved capabilities over the present generation.

Power supply or generation is one area where revolutionary changes could significantly expand satellite capabilities. Presently, power sources are limited to solar panels or on-board power supplies. Solar panels require continuous exposure to the sun, or the use of batteries to supply power during periods of darkness. Any on-board power system such as a battery adds weight, which reduces the number of electronic systems that can be flown. Furthermore, a system of solar panels and/or on-board sources is best suited to continuous power at moderate levels, and cannot easily supply high-energy bursts without significant additional weight in order to collect and store, and then release the energy.

One way of providing a more flexible power source is to beam the power from a ground station **10** to a satellite **20**, as illustrated in FIG. **1**. This concept has been explored in the past, but in the opposite direction: beaming power to earth

(which seemed attractive during the energy crisis). Sending power in the space to earth direction faces certain fundamental limits that make it impractical, but these limits are eased in the earth to space direction, leading to a system that is within the realm of possibility.

In addition to satellites, there are many other applications where beaming power could be important. For example, it is possible to replace hundreds of civilian cellphone base stations with a single zeppelin **20'**, shown in FIG. **2**, which could service a large metropolitan area **25** with mobile telephony, as well as such other services as “satellite” television. This would provide a low-cost alternative to satellites for many commercial wireless applications.

Furthermore, other applications include small UAVs (Unmanned Aerial Vehicles) that could be powered by beamed energy. See FIG. **3**. As the size of a UAV is reduced, the amount of weight that it can carry limits its lifespan significantly. For example, 100-gram airplanes have been built, but their lifetime is limited to six minutes with currently available batteries. By beaming power to a micro-UAV 20", it could stay aloft much longer. This would be useful for such applications as law enforcement, surveillance, hazardous site investigation, etc., in addition to the obvious military applications.

The embodiments of FIGS. **1-3** assume that the source of power is from a ground station **10**. However, the source of power need not necessarily be terrestrial. The source of power could be airborne or even in space.

Any beamed power system must confront the fundamental limits summarized by the Friis transmission equation, which relates the total power transmitted to the gain, G, of the transmitting and receiving antennas, the distance between them, R, and the wavelength λ of the radiation used.

$$P_{Rx} = G_{Tx} G_{Rx} \left(\frac{\lambda}{4\pi R}\right)^2 P_{Tx} \qquad [1]$$

Assuming for simplicity that both antennas are circular, the gain of each is related to its diameter, D.

$$G = \left(\frac{\pi D}{\lambda}\right)^2 \qquad [2]$$

If one assumes for the moment that very little power will be lost to spillover (this requirement can be relaxed) these equations can be combined to yield an expression for the required sizes of the transmitting and receiving antennas, as a function of their separation, and the wavelength of the radiation used. See FIG. **4**, which depicts the geometry involved in equation 3, to determine the required diameters for the transmitting and receiving antennas.

$$D_{Tx} D_{Rx} = \frac{4}{\pi} R\lambda \qquad [3]$$

For a given separation, reducing the wavelength reduces the size requirements of the transmitter and/or receiver. One tempting solution is to use optical wavelengths, and beam power to space with a large earth-based laser. This has several drawbacks, including scattering by atmospheric turbulence and airborne particles, the typically low wall-plug efficiencies of lasers compared to microwave sources, and the losses in conversion back to DC by photovoltaic cells. Lasers may be viable alternatives for stationary, near-earth applications such as zeppelins, but not for moving applications, such as micro-UAVs. Their utility for satellites is questionable.

The next candidate wavelength range after optical (skipping terahertz frequencies, which are currently not feasible) is millimeter waves. In the 90-100 GHz range, the attenuation for a one-way trip through the atmosphere can be as little as 1 dB (See Koert, 1992, infra). Furthermore, efficient high-power sources are available, such as the gyrotron, which can produce as much as 200 kW of continuous power at millimeter wave frequencies, at an efficiency of 50% (See Gold, 1997, infra). For higher power applications, arrays of klystrons have been proposed that could produce tens of megawatts of power. These existing high-power sources suggest that it could be possible to temporarily supply a satellite with much higher power from the ground than can currently be produced in orbit. For comparison, the most powerful commercial satellite that is available, the Boeing 702, operates at 25 kW from on-board solar panels. These power sources would be more than adequate for airship applications, and the power required for micro-UAVs would only be on the order of watts.

The most significant engineering challenge for efficient earth to space power transmission is the design of the transmitting and receiving antennas. Fortunately, the receiver design is greatly simplified by the development of the rectenna, (See Brown, 1984, infra) which consists of an array having a rectifier diode at each element. Converting to DC directly at each antenna eliminates the requirement for a perfectly flat phase front, and permits the receiving aperture to take any shape. The transmitter must still produce a coherent beam, so a parabolic dish or other method of phase control is necessary. This is one reason why space to earth transmission is impractical. To illustrate the possibility of high-efficiency earth to space transmission, consider the following example.

Assume that 100 GHz radiation is to be used. The maximum transmitter gain is determined by the ability to accurately build a large dish with the necessary smoothness. The Arecibo dish, which operates at 10 GHz, is 300 meters in diameter. First, assume that a 100 GHz dish could be similarly built with a diameter of 30 meters.

Next, assume that a low-earth-orbit (LEO) satellite is utilized, at an altitude of 500 km. Using equation 3, the required receiver diameter for high transmission efficiency is about 60 meters. This can be compared to the Boeing 702 solar panel wingspan of 47 meters. Thus, structures of the required sizes can be built, both on earth and in space.

However, existing rectenna designs are not practical for space power applications because they require an enormous number of diodes to cover such a large area. For the example just described, one diode per half-wavelength at 100 GHz equates to 6 billion diodes. Using 12-inch wafers, and assuming an area of 1 mm square per diode, this represents the yield of 20,000 wafers; the weight and cost of the diodes alone would be prohibitive.

Another problem with space power applications using traditional rectenna designs is that the power density is too low to achieve significant efficiency. The efficiency, h, of a rectenna is related to the voltage across the diodes, V_D, and the built-in diode voltage, V_{bi} (See McSpadden, 1998, infra).

$$\eta \propto \frac{1}{1 + V_{bi} / V_D} \qquad [4]$$

Designs with efficiencies as high as 90% have been demonstrated, [Strassner, 2002] but the power densities involved were much higher than one could expect to encounter in space. For the LEO example given above, the power density would be 6 mW/cm^2, which corresponds to only 0.2 volts generated across each diode—on the order of the typical built-in voltage for a Schottky diode. The practical limitations of a space power system are thus the large number of diodes needed, and the low voltage generated across each diode. The efficiency could also be improved by placing each diode inside a high Q resonant structure, or by using diodes with lower built-in voltage. However, either of these solutions alone would not solve the problem of the large number of required diodes.

As such there is a need for lens-like structures that will allow the number of diodes to be reduced.

In terms of the prior art and a better understanding of the background to the present invention, the reader is directed to the following articles:

W. Brown, "The History of Power Transmission by Radio Waves", *IEEE Transactions on Microwave Theory and Techniques*, vol. 32, no. 9, pp. 1230-1242, September 1984.

P. Fay, J. N. Schulman, S. Thomas III, D. H. Chow, Y. K. Boegeman, and K. S. Holabird, "High-Performance Antimonide-Based Heterostructure Backward Diodes for Millimeter-wave Detection", *IEEE Electron Device Lett.* 23, 585-587 (2002).

S. Gold, G. Nusinovitch, "Review of High Power Microwave Source Research", *Review of Scientific Instruments*, vol. 68, no. 11, pp. 3945-3974, November 1997.

P. Koert, J. Cha, "Millimeter Wave Technology for Space Power Beaming", *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, no. 6, pp. 1251-1258, June 1992.

H. J. Lezec, A. Degiron, E. Devaux, R. A. Linke, L. Martin-Moreno, F. J. Garcia-Vidal, and T. W. Ebbesen, "Beaming Light from a Subwavelength Aperture", *Science*, vol. 297, pp. 820-822, Aug. 2, 2002.

J. McSpadden, L. Fan, K. Chang, "Design and Experiments of a High-Conversion-Efficiency 5.8 GHz Rectenna", *IEEE Transactions on Microwave Theory and Techniques*, vol. 46, no. 12, pp. 2053-2060, September 1984.

J. N. Schulman and D. H. Chow, "Sb-Heterostructure Interband Backward Diodes," *IEEE Electron Device Lett.,* 21, 353-355 (2000).

D. Sievenpiper, J. Schaffner, H. Song, R. Loo, G. Tangonan, "Two-Dimensional Beam Steering Using an Electrically Tunable Impedance Surface", *IEEE Transactions on Antennas and Propagation*, special issue on metamaterials, October 2003.

B. Strassner, K. Chang, "5.8 GHz Circularly Polarized Rectifying Antenna for Wireless Microwave Power Transmission", *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 8, pp. 1870-1876, August 2002.

F. Yang, Y. Qian, T. Itoh, "A Uniplanar Compact Photonic Bandgap (UCPBG) Structure and its Applications for Microwave Circuits", *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 8, pp. 1509-1514, August 1999.

BRIEF DESCRIPTION OF THE INVENTION

Briefly and in general terms, the disclosed technology, in one aspect comprises a rectenna structure comprising: a flexible, dielectric sheet of material; a plurality of metallic lenslets disposed on the sheet of material; and a plurality of diodes disposed on the sheet of material, each diode in said plurality of diodes being arranged at a focus of a corresponding one of said plurality of metallic lenslets.

In another aspect, the disclosed technology relates to a method of generating electrical power for use aboard an aircraft or a satellite, the method comprising: deploying a sheet of dielectric material in an orientation, the sheet of dielectric material being associated with, coupled to and/or forming a part of said aircraft or satellite, the sheet of dielectric material having a plurality of metallic lenslets disposed on the sheet of dielectric material and a plurality of diodes disposed on or adjacent the sheet of dielectric material, each diode in said plurality of diodes being arranged at a focus of a corresponding one of said plurality of metallic lenslets, the diodes being coupled together for supplying electrical power for use by systems aboard said aircraft or a satellite, and directing the orientation of the sheet of dielectric material to receive incident radiation from a source of electromagnetic radiation.

BRIEF DESCRIPTION OF THE FIGURES

FIG. **1** depicts beaming microwave power to an orbiting satellite as an alternative to the use of traditional solar panels, batteries, and other power sources on the satellite.

FIG. **2** depicts another application for beamed RF power which includes airships that would supply cities with wireless services, for example.

FIG. **3** depicts an application where beamed RF power may also be used to power "slow-flight" micro-UAVs, which could be used for law-enforcement or surveillance, or investigations made at a hazardous site, for example.

FIG. **4** shows the geometry involved in equation [3], which equation can be used to determine the required diameters for the transmitting and receiving antennas.

FIG. **5** depicts radiation from a relatively larger being concentrated onto each diode, using, for example, a lightweight, planar resonant structure that may be printed on a thin, flexible plastic film.

FIG. **6** depicts a large plastic film, printed with metallic lens structures, and populated with rectifier diodes, (not shown) that would serve as a lightweight collector for microwave power. Such a structure could be built to cover tens of meters with minimal weight. The use of printed metallic lenses would reduce the number of diodes, and would increase the voltage across each diode for improved efficiency.

FIG. **7** depicts an embodiment where power is focused onto a sparse array of diodes using a lightweight plastic-film that is patterned with metallic lenses.

FIG. **8***a* depicts a coplanar antenna on a diode-tuned surface and experiments using such tunable textured surfaces has led the inventor named herein to believe that planar lens structures can focus power onto a coplanar antenna, yielding a completely flat structure.

FIG. **8***b* is a graph of the gain for uniform surface impedance.

FIG. **8***c* is a graph of the gain for an optimized, non-uniform surface.

FIG. **9** is a graph depicting the effective aperture for an antenna mounted on an optimized impedance surface can be nearly equal to the entire surface area. The effective aperture size assumes the expected cosine function when optimized for different elevation angles. The effective aperture for a uniform surface is shown for comparison. By optimizing a textured surface, one can make a large-area, planar collector for microwave radiation.

FIG. **10** is a plan view of a lightweight, high-efficiency rectenna system based on printed metallic lenslets, and a sparse array of rectifier diodes. The lenslets collect power over many square wavelengths and route it to the diodes. This provides greater power per diode (thus improving efficiency) and also reduces the diode count to a practical number.

FIGS. **10***a*, **10***b* and **10***c* are side elevation views taken through the structure depicted in FIG. **10**.

FIG. **11** is a side elevation view of an embodiment with a ground plane spaced from the lenslets using a honeycomb-like structure.

FIG. **12** depicts an array of metal plates with a period of one-quarter wavelength, and features that vary on a length scale of one wavelength, with radial symmetry.

FIG. **13** depicts an array of lenses depicted by FIG. **12**, each lens having a diode at its center, with those diodes being connected by DC lines.

FIG. **14** is similar to FIG. **10***b*, but with a DC line being shown on a reverse side of the dielectric sheet.

DETAILED DESCRIPTION

A problem in trying to develop a practical earth to space power transmission system is that the voltage across diodes used in a rectenna has not been sufficient in a prior art rectenna to be of practical use to such an application.

However, the voltage across each diode **25** can be increased while reducing the number of diodes by using a lens-like structure or lenslet **40**, shown in FIG. **5**, to concentrate power from a large area over a small number of diodes **25** in an array of lenslets **40**. For example, if one wanted to generate 20 volts across each diode **25**, then the incident power from 40 square wavelengths, or a 2-cm diameter area, needs to be collected. This would not only boost the voltage across each diode—and also the diode's efficiency—it would also reduce the number of diodes to about 3 million for the example described above, which equates to about 100 wafers' worth of diodes. Further reductions in the number of diodes required could be achieved with an even larger collection area per diode. Each lenslet **40** comprises a geometric array of electrically conductive patches **42** disposed on a supporting surface. The conductive patches **42** are preferably formed by thin, individual metallic patches formed on a supporting surface, such as, for example, a thin sheet of a plastic material.

Of course, a traditional dielectric lens would be impractical, but a metallic lens imprinted on a lightweight plastic film **50**, which may be unfolded over a large area and could be utilized in a space environment, is practical. This concept for building a practical microwave space power system is illustrated in FIG. **6** where ground-based radiation is represented by arrows A. The thin plastic film **50** would be patterned with sub-wavelength resonant metallic regions or lenslets **40**, which would focus the incoming power to a sparse array of diodes **25**. Such a film **50** could be made by the tens of meters, but nevertheless would have minimal weight. The metallic lenses would serve the dual purposes of minimizing the number of diodes **25** required, while also improving the efficiency by increasing the voltage across each diode **25**.

In accordance with the presently disclosed technology, a structure having a thin plastic film **50** that is covered with a plurality of thin metal patterns, each pattern comprising a plurality of small electrically conductive patches **42** forming a lenslet **40**, is disclosed. This technology may be used in applications such as the earth to space power transmission system discussed above. Each metal pattern or lenslet **40** is made such that it behaves as a planar lens, with a focal length of zero. That is, it focuses the incoming power in such a way that a relatively high energy field is created at one point on the surface of the lens **40**. The high-energy field has a higher energy than the average energy density of the electromagnetic waves impinging the plastic film **50**. The creation of the high-energy fields allows a rectifier diode **25** to be placed at the focus or center of the high-field location, so that all of the power impinging on the lens **40** is rectified by that diode **25**. This results in two improvements over existing rectenna designs: (1) It requires far fewer diodes, and (2) it allows the voltage per diode to be higher, which results in more efficient operation. As will be seen, an embodiment of the present invention includes the combination of a planar lens and a sparse array of rectifier diodes to create a lightweight, efficient rectenna.

The design of the planar lens can be summarized as follows: (1) assume that the plastic film **50** is preferably planar and is patterned with metallic or other electrically conductive patches **42** that can be considered as resonators, with a certain resonance frequency. (2) Characterize the patches **42** in terms of scattered field (magnitude and phase) for various frequencies with respect to the resonance frequency. (3) Choose the condition that the fields from all of the metal patches **42** should add up in phase at a single point at the focus of a lens **40**, or alternatively choose some other point on the lens. (4) Build a scattering matrix that describes the field at the chosen point on the lens, as a function of the incoming field. This must include the interaction among the various metallic patches. (5) Optimize the resonance frequencies of the metal patches **42** so that the field at the chosen point is a maximum. Of course, diodes **25** would be placed at the focal points of the lenses **40**.

Concentrating microwave power from a large area (several tens of square wavelengths) onto a single device, using a thin, patterned metal film can be done in several ways, including by using a non-uniform frequency selective surface (FSS). These structures have been studied for many years for filtering radomes, and other applications. A non-uniform FSS could be designed to have lens-like behavior, and focus incoming waves from a large area onto a single receiving antenna. This is similar to the Fresnel zone plate that is known in optics, but it can have high efficiency because the metal patterns can be designed to provide only a phase shift, with minimal absorption. A series of microwave lenslets **54** could be patterned over a large area of thin plastic film **50**, as shown in FIG. **6**, to focus the low-density microwave power onto a sparse array of diodes **25**.

One drawback of the traditional FSS approach, shown in FIG. **7**, is that it is not uniplanar, because of the need to focus over a distance “t” which would be roughly equal to the diameter of the lenslets **40**. It would be difficult to unfold such a structure over an area of many square meters while maintaining the required spacing “t” within suitable tolerances for efficient energy collection.

An alternative is to consider structures where the receiving antennas and the diodes are arranged in a coplanar alignment with the metallic lens structures. This concept has already been demonstrated at HRL Laboratories of Malibu, Calif., through work with tunable, textured electromagnetic surfaces. See, for example, the patent applications mentioned above. A metallic surface texture can be made (through proper optimization) to focus power from many square wavelengths, onto an antenna that is coplanar with the textured surface, as illustrated in FIG. **8**. The metallic surface can be quite thin and for most applications, the thinner the metal on the metallic surface the better (due to weight considerations). The experiments at HRL Laboratories involved a diode-tuned impedance surface consisting of many small metallic patches linked by varactor diodes. See the paper by Sievenpiper, 2003, supra, and the patent applications referred to above. Experiments using tunable textured surfaces suggest that planar lens structures can focus power onto a coplanar antenna, yielding a completely (or essentially) flat structure as shown in FIG. **8***a*. The structure of FIG. **8***a* can be made so flat that the antenna and tunable textured surface is nearly imperceptible to one’s fingers. Of course, the structures can be more pronounced in some embodiments (so that they would not generally be called flat), but, generally speaking, flat or nearly flat structures would be preferred in most applications, particularly where the plastic film **50** is to be unfolded someplace, such as in space, where human intervention (due to snags and the like), may well not be possible, convenient or desirable.

FIG. **8***b* is a graph of the gain for uniform surface impedance while FIG. **8***c* is a graph of the gain for an optimized, non-uniform surface. Using this surface with a coplanar antenna, it was determined that the pattern of capacitance between the metallic patches could be optimized to increase the gain of the antenna mounted on the surface. In this way, the effective aperture size could be extended to cover the entire area of the surface, as plotted in FIG. **9**, which indicates that by proper design of a non-uniform impedance surface, power can be collected over a relatively large area (for example a relatively large area may be as large as perhaps hundreds or thousands of square of wavelengths of the impinging radiation), and routed to a single diode on the non-uniform impedance surface for rectification. For RF space power applications, the surface texture would consist of a lattice of fixed capacitors, built into metallized plastic. The capacitors are formed edge to confronting edge of the plates making up each lens. The afore-mentioned capacitors come from the fact that there are small metallic plates that are very close together. These are edge-to-edge capacitors, rather than conventional parallel plate capacitors. Any two conductors that are brought near each other will have some amount of capacitance. A ground plane is not needed here, but it could be used to provide improved efficiency, at the expense of greater weight. The values of the capacitors and the shape of the metal particles would be determined by electromagnetic simulations, and an optimization algorithm.

The results described above with reference to FIGS. **8***a*-**8***c* are for a two-layer structure containing vertical metallic vias—a high-impedance surface—that was built using printed circuit board technology as described in my issued U.S. patents and published U.S. patent applications. Lighter weight structures are needed in order to make this general concept practical and sufficiently lightweight for convenient use in space or even for use on an airship such as the airship shown in FIG. **2**. For example, the planar lenslets **40**, the collection antennas **60**, and the rectifier diodes **25** should preferably be built on a single surface that would preferably be printed on a single-layer plastic, dielectric film. Of significant importance would be the elimination of the vertical vias, and preferably the ground plane itself, leading to an entirely uniplanar structure. Simple uniplanar structures have been studied (See Yang, 1999, supra), but techniques for optimizing complex non-uniform surfaces need to be developed.

Furthermore, if the ground plane is eliminated, methods for minimizing transmission through the structure also would need to be considered. The structure could be analyzed as a complex parasitic array, where the individual patches in the patterned metallic surface could be considered as parasitic antennas. Their shape would be optimized so that the scattered power from each of them would be maximized at one point, where the rectifier diodes would be placed.

A microwave structure embodiment is depicted by FIG. **10**. FIGS. **10***a*, **10***b* and **10***c* provide section views through the embodiment depicted by FIG. **10**. In this embodiment a lattice of printed metallic lenslets **40**, each formed by arrays of thin metal patches **42**, focus power onto a sparse array of rectifier diodes **25**, which could be coplanar with the lenslets or mounted on the adjacent patches **42** as shown by FIG. **10***c*. DC power lines **65** (which are preferably incorporated into the structure) could then carry power from diodes **25** to the satellite **10**, for example, for distribution to the onboard electronic systems. The entire rectenna could be printed on a thin, lightweight, plastic film **50**, which could be unfolded to cover an area comprising many square meters. Like all rectennas, it would not need to assume a particular shape, because rectification is done right at each antenna element. However, since each lenslet would provide some directivity, the surface would need to be roughly pointed toward (i.e. be orthogonal to) the source of energy, such as the ground station **10** source. The required pointing accuracy, among other things, would govern the size of the lenslets **40**.

A ground plane may be helpful in some embodiment. It could increase the efficiency, by not allowing any energy to pass through the structure. The metallic pattern on the top of film **50** would be qualitatively similar to that without the ground plane, but in detail it would probably be a different pattern to compensate for the presence of the ground plane. The ground plane would have to be separated from the top metal patterns by some distance, typically 1/100 to 1/10 wavelength, depending on the tolerances allowed in the manufacturing of the metallic patterns. (This is not due to the tolerance of the film thickness. It is due to the fact that the overall thickness will affect the bandwidth. If the bandwidth is very narrow, then the metallic patterns will have to be defined very accurately to get the capacitance right.) In order to allow some spacing, but not to have a very heavy structure, an embodiment with a ground plane **44** may be ribbed, air-filled structure **46**, such as that seen in FIG. **11**. This might be similar to flexible “bubble wrap”, or a rigid honeycomb-like dielectric that is commonly used in airframes and other such things.

In summary, the rectenna consists of a rectifying diode **25** and a generally planar lens structure **40**. The lens structure comprises a thin dielectric (such as plastic) sheet **50** that is patterned with metallic regions **42**. The metallic regions **42** scatter electromagnetic energy, and they are arranged so that the collective scattered energy from all of them is focused into the diode **25**. Each rectifying diode **25** is attached between two adjacent ones of the metal regions **42**. The diodes **25** are also attached to long conductive paths **46** (wires) that traverse the entire width of the structure, or are otherwise routed so that they supply current to a common location (such as an edge) where it may be collected and used to supply electrical power to a satellite or other device. The wires **46** are preferably coplanar with the metal patches **42** that make up the lens **40**, and they are preferably oriented transverse to the expected polarization of the energizing RF field, so that they have a minimum scattering effect. The metal pattern of the lens **40** can also be optimized to account for the scattering of the wires **46**. The lens **40** and indeed the thin dielectric sheet **50** preferably have a planar configuration and indeed the rectenna, when designed, will very likely be assumed to have a planar configuration in order to simplify its design (see the foregoing discussion). But those skilled in the art should appreciate the fact that the sheet **50** may well assume a non-parallel configuration in use, either by design or by accident. If designed for a planar configuration, the extent by which the in-use sheet **50** deviates from a planar configuration will adversely affect its effectiveness. But if the in-use design is close to being planar, the loss in efficiency is likely to be very small. Of course, the rectenna can be designed initially with a non-planar configuration in mind, but a non-planar configuration will doubtlessly complicate finding a desirable arrangement of the patches **42** for the various lenslets **40**. Making an assumption that the sheet **50** and the lenslets **40** will all be planar should simplify the design of the rectenna significantly.

The lenses (or lenslets) **40** are ideally designed and optimized using a computer. A random collection of scatterers is simulated, and the collected power is calculated using an electromagnetic solver. The sizes, shapes, and locations of the scatterers are varied according to an optimization method. Such methods are known to those skilled in the art, and include the method of steepest descent, genetic algorithms, and many others. The geometry that provides the greatest power to the diode **25** is then apt to be chosen as the ideal structure.

Such methods are good for determining the best geometry when nothing is known about that geometry beforehand. However, in the case of the present invention, much is known about the required geometry, and one can design a simple structure by hand. The preferred design method is then to start with a known good structure using the calculations described below, and then to optimize it using a computer as described above.

It can be shown that a wave having wave vector k_0, propagating on a periodic structure with effective refractive index n_{eff} will be scattered by the periodicity of that structure k_p to an angle θ given by:

$$\theta = \mathrm{Sin}^{-1}\left(\frac{k_0 n_{eff} - k_p}{k_p}\right) \qquad [4]$$

The planar lens structure should be designed so that energy scatters from the normal direction (θ=0°) into the plane of the surface where the diode is located. Assuming that the dielectric layer is thin, we have $n_{eff}=1$, so we are left with $k_p=k_0$. Therefore, the periodicity of the structure should be roughly one free-space wavelength.

In order to have independent control over the magnitude and phase of the radiation from the feed point, (or conversely in the present case, the collected energy at the diode **25**) it is necessary to have the periodicity be much greater. For independent control over two parameters, the array should be oversampled by a factor of at least two, which means that the individual metal patches **42** should be spaced at most one-quarter wavelength apart, with their properties varying periodically on a length scale of one wavelength. The structure should have close to radial symmetry, so that energy is scattered inward toward a central point. However, the symmetry can vary from perfect radial symmetry to account for polarization effects (leading to a slight deviation which has mirror symmetry) or for practical reasons due to the discrete nature of the individual patches **42**. An example of such a structure is shown in FIG. **12**.

FIG. **12** depicts an array of metal plates **42** located on centers spaced with a period of one-quarter wavelength, and the features thereof (size in this embodiment) vary on a length scale of one wavelength, with radial symmetry. The scattered energy from the metal plates **42** combines coherently at the diode **25** located at the center of the geometric pattern formed by plates **42**.

This single planar lens **40** consists of metal patches **42** having a periodicity of one-quarter wavelength, and having properties (the patch size in this embodiment) varying with a period of one wavelength. The planar lens **40** shown has a diameter of about four wavelengths. It collects power over its entire surface, and directs it toward the diode **25** at the center of the pattern, which diode is preferably connected between a pair of the closest patches **42**. This lens **40** forms a single element of a larger array **65**, shown in FIG. **13**, in which the diodes **25** are also connected in parallel by rows.

FIG. **13** depicts an array of planar lenslets or lenses **30**, each having a diode **25** at the center thereof, with those diodes **25** being connected by DC lines **46**. The lines are preferably oriented transverse to the electric field of the incoming radiation, so that they do not interfere significantly with the scattered waves. They can be printed on the same side of the sheet **50** (see FIG. **10***b*) as the metallic patches **42**, in which case the metal pattern of the lines **46** would simply be combined with that of the patches **42**, or they can be printed on the reverse side of sheet **50**, and attached to the diodes **25** by small metal plated via holes **48** in the plastic sheet **50**, as shown by FIG. **14**.

This design requires far fewer diodes than do conventional rectennas, because the diodes **25** are spaced every four wavelengths, rather than every half-wavelength. The result is a factor of close to 64 times reduction in the number of required diodes, and a corresponding factor of 64 times increase in the voltage generated per diode. This is particularly useful in cases where the incoming power density is low (such as space applications), where it would otherwise be difficult to get the induced voltage above the diode threshold voltage. Thus, this design also has higher efficiency due to the greater induced voltage at lower power levels.

Having described this technology in connection with certain embodiments thereof, modification will now doubtlessly suggest itself to those skilled in the art. As such, the protection afforded hereby is not to be limited to the disclosed embodiments except as is specifically required by the appended claims.

What is claimed is:

1. A method of generating electrical power for use aboard an aircraft or a satellite, the method comprising:

a. deploying a sheet of dielectric material in an orientation, the sheet of dielectric material being associated with, coupled to and/or forming a part of said aircraft or satellite, the sheet of dielectric material having a plurality of metallic lenslets disposed on the sheet of dielectric material and a plurality of diodes disposed on or adjacent the sheet of dielectric material, each diode in said plurality of diodes being arranged at a focus of a corresponding one of said plurality of metallic lenslets, the diodes being coupled together for supplying electrical power for use by systems aboard said aircraft or satellite, and

b. directing the orientation of the sheet of dielectric material to receive incident radiation from a source of electromagnetic radiation.

2. The method of claim **1** wherein the source of electromagnetic radiation is a terrestrial source of electromagnetic radiation.

3. The method of claim **1** wherein the sheet of dielectric material is deployed in said orientation such that the sheet of dielectric material assumes a planar configuration and is oriented so that the planar configuration is orthogonal to a direction pointing to said source of electromagnetic radiation.

4. The method of claim **3** wherein the source of electromagnetic radiation is a transmitting antenna disposed on or immediately adjacent a surface of the planet earth.

5. The method of claim **4** wherein the metallic lenslets each comprise a geometric arrangement of metallic patches.

6. The method of claim **5** wherein the focus of each lenslet corresponds to a center of the geometric arrangement of metallic patches comprising the lenslet.

7. The method of claim **5** wherein the plurality of lenslets and diodes define a rectenna and wherein the rectenna is designed to be responsive to incident radiation for converting the incident radiation to electrical energy and wherein metallic patches in each of said geometric arrangements have centers which are spaced from centers of neighboring metallic patches by a distance equal to one-quarter wavelength of said incident radiation.

8. The method of claim **7** wherein the patches in each of said lenslets have a property that varies along a radial direction from the focus of the lenslet with a period equal to one wavelength of said incident radiation.

9. The method of claim **8** wherein the property that varies along a radial direction from the focus of the lenslet is the geometric size of the individual patches.

10. The method of claim **9** wherein the geometric arrangement is a hexagonal arrangement.

11. The method of claim **10** wherein the individual patches each have a hexagonal shape when viewed in a plan view.

12. The method of claim **9** wherein the geometric arrangement is a square arrangement.

13. A method of designing an electromagnetic energy receiving surface having a dielectric film with conductive patches formed thereon and arranged as a lens, the receiving surface being responsive to an incoming electromagnetic field, the method comprising:

(1) assume that the dielectric film is preferably planar and is patterned with the conductive patches disposed in a geometric arrangement;

(2) treat the conductive patches as resonators having a resonance frequency;

(3) characterize the metal patches in terms of scattered field (magnitude and phase) for various frequencies with respect to the resonance frequency;

(4) select a shape and a position of the conductive patches such that fields from all of the metal patches add up in phase at a focal point or focus of the lens;

(5) build a scattering matrix that describes the field at a chosen point on the lens as a function of the incoming electromagnetic field, the scattering matrix including interaction among the conductive patches forming the lens; and

(6) optimize the resonance frequencies of the conductive patches so that the field at the focal point or focus of the lens is a maximum.

* * * * *

US009935370B2

(12) United States Patent
Casse et al.

(10) **Patent No.: US 9,935,370 B2**
(45) **Date of Patent: Apr. 3, 2018**

(54) MULTIBAND RADIO FREQUENCY (RF) ENERGY HARVESTING WITH SCALABLE ANTENNA

(71) Applicant: **Palo Alto Research Center Incorporated**, Palo Alto, CA (US)

(72) Inventors: **Bernard D. Casse**, Saratoga, CA (US); **George Daniel**, Palo Alto, CA (US); **Armin R. Volkel**, Mountain View, CA (US); **Victor Liu**, Mountain View, CA (US)

(73) Assignee: **Palo Alto Research Center Incorporated**, Palo Alto, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 325 days.

(21) Appl. No.: **14/582,033**

(22) Filed: **Dec. 23, 2014**

(65) **Prior Publication Data**

US 2016/0181868 A1 Jun. 23, 2016

(51) **Int. Cl.**
H01F 27/42 (2006.01)
H01Q 5/25 (2015.01)
(Continued)

(52) **U.S. Cl.**
CPC ***H01Q 5/25*** (2015.01); ***H01Q 1/248*** (2013.01); ***H01Q 9/26*** (2013.01); ***H02J 50/27*** (2016.02)

(58) **Field of Classification Search**
None
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,929,594	A	12/1975	Fromson
4,065,364	A	12/1977	Fromson

(Continued)

FOREIGN PATENT DOCUMENTS

CN	102778144		11/2012
CN	103312042	A	9/2013

(Continued)

OTHER PUBLICATIONS

Aoyama, Takahiko et al. "Energy response of a full-energy-absorption neutron spectrometer using boron-loaded liquid scintillator BC-523", Nuclear Instruments and Methods in Physics Research A 333 (1993) 492-501, 10 pages.

(Continued)

Primary Examiner — Cassandra Cox

(74) *Attorney, Agent, or Firm* — Bever, Hoffman & Harms, LLP

(57) ABSTRACT

A radio frequency (RF) energy harvesting device including a scalable metamaterial resonator antenna and a rectifying circuit formed on a flexible plastic substrate. The metamaterial resonator antenna includes a metal (e.g., silver) structure that is conformally fixedly disposed (i.e., either printed or deposited/etched) on the flexible substrate and configured to resonate at RF frequencies using primary and secondary antenna segments connected by linking segments such that captured RF signals are generated at two antenna end points that are 180° out-of-phase with each other. The rectifying circuit including additional metal structures that are also printed or otherwise formed on the flexible substrate, and one or more circuit elements that are configured to pass positive voltage pulses from the captured RF signals to an output node. Various metamaterial resonator antenna configurations are disclosed.

22 Claims, 10 Drawing Sheets

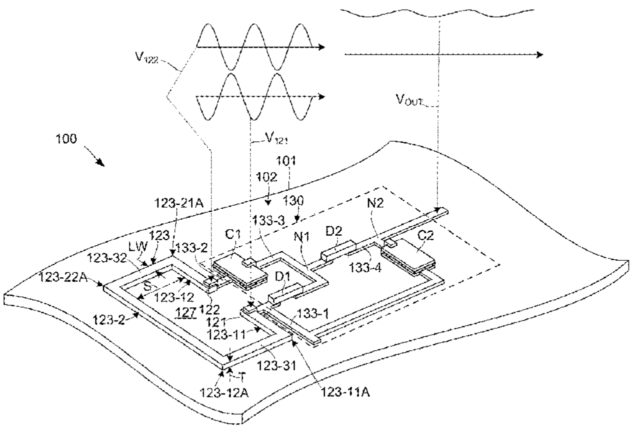

(51) **Int. Cl.**

H01Q 1/24	(2006.01)
H01Q 9/26	(2006.01)
H02J 50/27	(2016.01)

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,315,873	A	2/1982	Smith et al.	
4,779,000	A	10/1988	Ing	
5,043,739	A	8/1991	Logan et al.	
5,712,166	A	1/1998	Broan	
5,905,263	A	5/1999	Nishizawa	
7,002,517	B2	2/2006	Noujeim	
8,621,245	B2	12/2013	Shearer et al.	
8,680,945	B1	3/2014	Wang	
2002/0180639	A1	12/2002	Rickett	
2004/0207486	A1	10/2004	York	
2004/0238751	A1	12/2004	Penn	
2006/0234505	A1	10/2006	Asano et al.	
2008/0049228	A1	2/2008	Chan	
2008/0143192	A1	6/2008	Sample	
2009/0067208	A1	3/2009	Martin	
2009/0152954	A1	6/2009	Le	
2009/0207000	A1	8/2009	Mickle et al.	
2009/0220802	A1	9/2009	Faber et al.	
2009/0284351	A1	11/2009	Rossman et al.	
2011/0242863	A1	10/2011	Park	
2011/0303850	A1	12/2011	Barillon	
2011/0309686	A1*	12/2011	Scherbenski	H02J 17/00 307/104
2012/0133547	A1	5/2012	MacDonald et al.	
2012/0314541	A1	12/2012	Matsuura	
2013/0076570	A1*	3/2013	Lee	H01Q 9/045 343/700 MS
2013/0187830	A1*	7/2013	Warnick	H01Q 5/35 343/893
2014/0131023	A1	5/2014	Raman et al.	
2014/0266946	A1	9/2014	Bily et al.	
2014/0300520	A1	10/2014	Nguyen et al.	
2014/0355381	A1	12/2014	Lai et al.	
2015/0214927	A1	7/2015	Greene et al.	
2015/0229028	A1	8/2015	Bily et al.	
2015/0236551	A1	8/2015	Shearer et al.	
2015/0276489	A1	10/2015	Cumming	
2015/0318618	A1	11/2015	Chen et al.	
2015/0372389	A1	12/2015	Chen et al.	
2015/0380973	A1	12/2015	Scheb	
2016/0145214	A1	5/2016	Douce	
2016/0181867	A1*	6/2016	Daniel	H01Q 9/16 307/104
2016/0254844	A1*	9/2016	Hull	H04B 5/0062 340/6.1
2016/0336198	A1*	11/2016	Singleton	H01L 21/561
2016/0359378	A1*	12/2016	Kuhn	H02M 1/126

FOREIGN PATENT DOCUMENTS

WO	200143228 A	6/2001
WO	2007015281	2/2007
WO	2013039926	3/2013
WO	2015038203	3/2015

OTHER PUBLICATIONS

Flaska, Marek et al., "Digital pulse shape analysis for the capture-gated liquid scintillator BC-523A", Nuclear Instruments and Methods in Physics Research A 599 (2009) 221-225, 5 pages.

Vanier, Peter E., et al., "Directional detection of fission-spectrum neutrons", 1-4244-1302-8/07, 2007 IEEE, 5 pages.

Vanier, Peter E., et al., "Calibration and Testing of a Large-Area Fast-Neutron Directional Detector", Brookhaven National Laboratory, BNL-79632-2007-CP, 8 pages.

Mascarenhas, Nicholas, et al., "Directional Neutron Detectors for Use with 14 MeV Neutrons", Sandia Report, SAND2005-6255, printed Oct. 2005, 32 pages.

Mirenda, Martin, et al., "Ionic liquids as solvents for liquid scintillation technology, Cerenkov counting with 1-Butyl-3-Methylimidazolium Chloride", Radiation Physics and Chemistry 98 (2014) 98-102, 5 pages.

Swiderski, L., et al., "Further Study of Boron-10 Loaded Liquid Scintillators for Detection of Fast and Thermal Neutrons", IEEE Transactions on Nuclear Science, vol. 57, No. 1, Feb. 2010, 6 pages.

Pratap et al., "Plasmonic Properties of Gold-Coated Nanoporous Anodic Alumina With Linearly Organized Pores," Pramana—J. Phys. (Dec. 2014), vol. 83, No. 6, pp. 1025-1033.

Noh et al., Highly Self-Assembled Nanotubular Aluminum Oxide by Hard Anodization, (Jan. 29, 2011), J. Mater. Res., vol. 26, Issue 2, pp. 186-193.

Juan Li et al., "Facile Method for Modulating the Profiles and Periods of Self-Ordered Three-Dimensional Alumina Taper-Nanopores," ACS Appl. Mater. Interfaces 2012, 4, 5678-5683.

Juan Li et al., "Tailoring Hexagonally Packed Metal Hollow-Nanocones and Taper-Nanotubes by Template-Induced Preferential Electrodeposition," ACS Appl. Mater. Interfaces 2013, 5, 10376-10380.

* cited by examiner

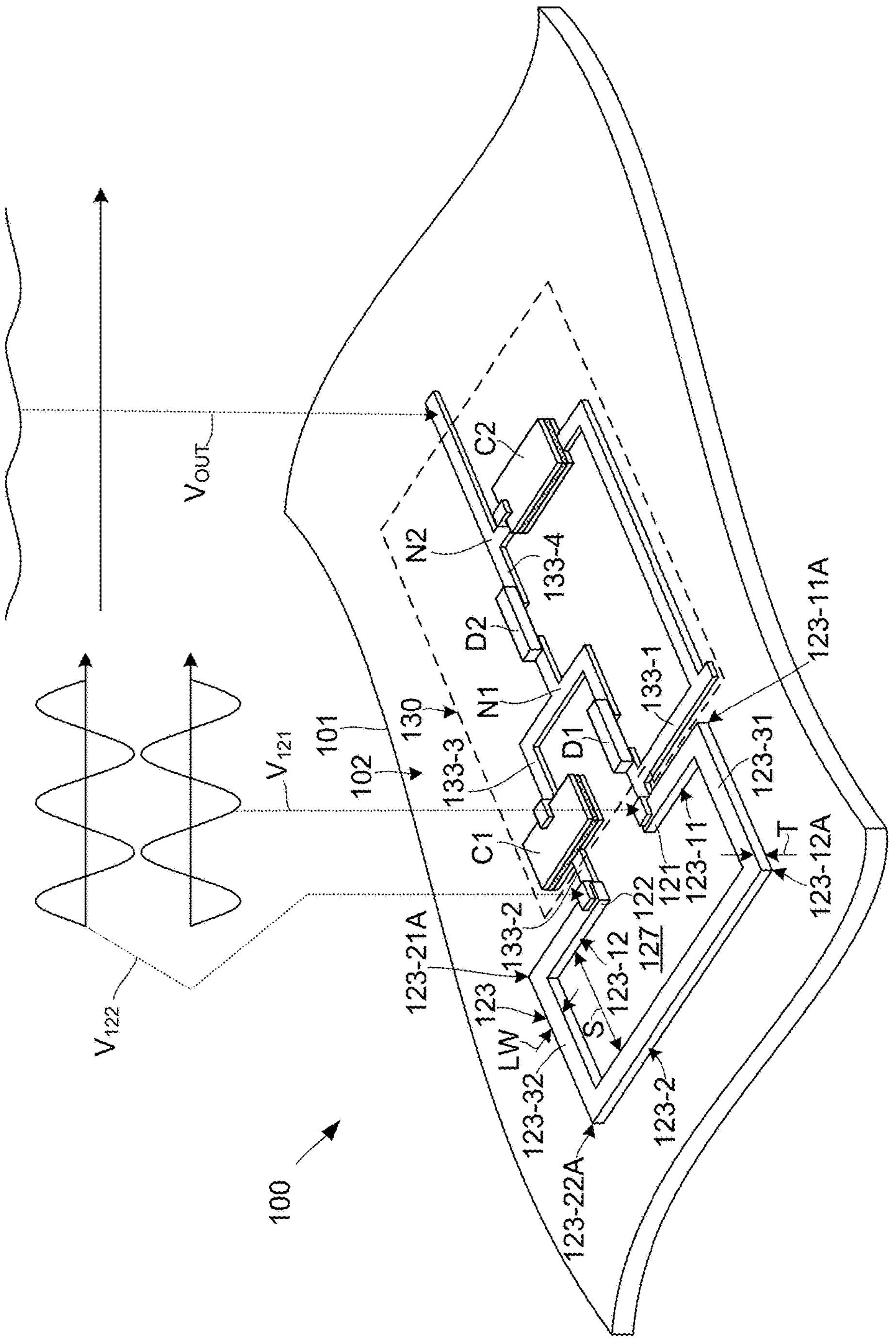
100
V_{122}
V_{121}
V_{OUT}
101
102
130
N2
133-4
D2
N1
133-3
D1
133-1
C2
C1
133-2
123-21A
123
LW
123-32
123-22A
123-2
S
123-12
127
122
121
123-11
123-31
123-12A
T
123-11A

FIG. 1

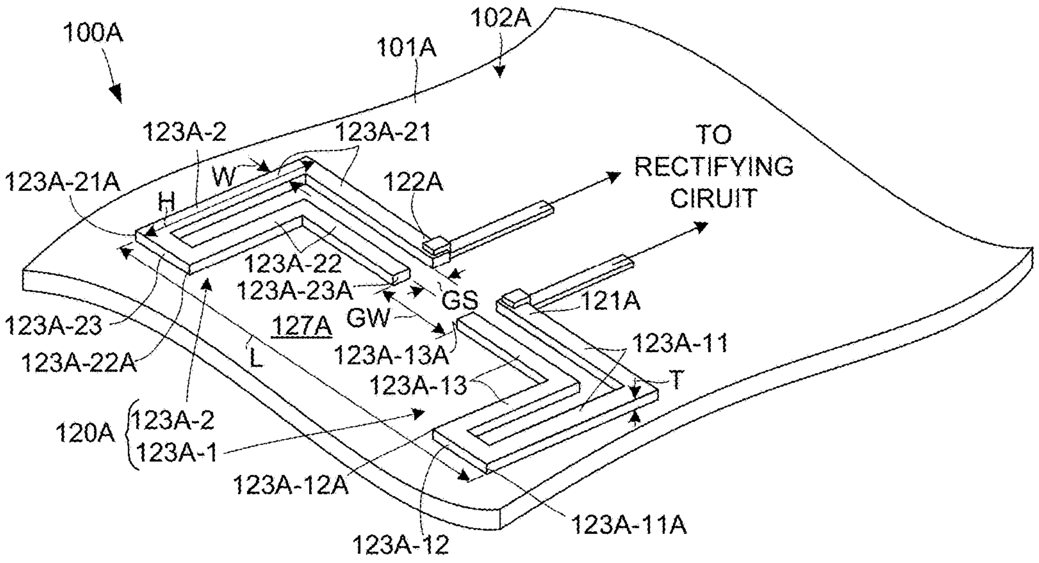
100A
102A
101A
123A-2
123A-21
TO
RECTIFYING
CIRUIT
123A-21A
W
H
122A
123A-22
123A-23A
GS
127A
GW
121A
123A-23
123A-13A
123A-11
123A-22A
L
123A-13
T
120A
123A-2
123A-1
123A-12A
123A-11A
123A-12

FIG. 2

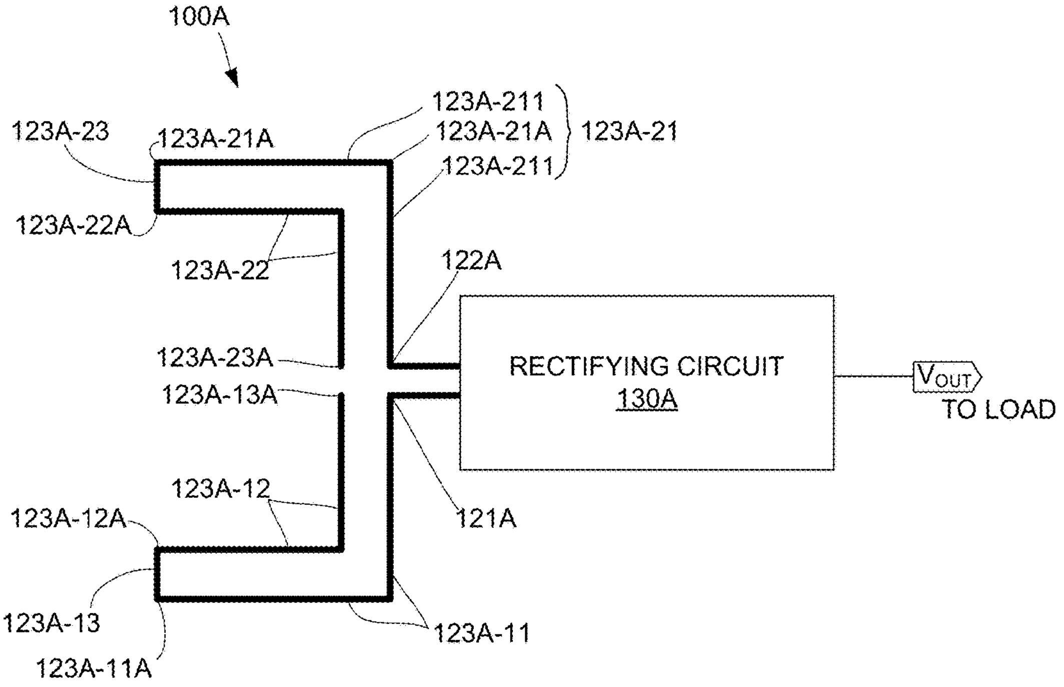
100A
123A-211
123A-23
123A-21A
123A-21A
123A-21
123A-211
123A-22A
123A-22
122A
123A-23A
RECTIFYING CIRCUIT
130A
V_{OUT}
TO LOAD
123A-13A
123A-12
123A-12A
121A
123A-13
123A-11
123A-11A

FIG. 3A

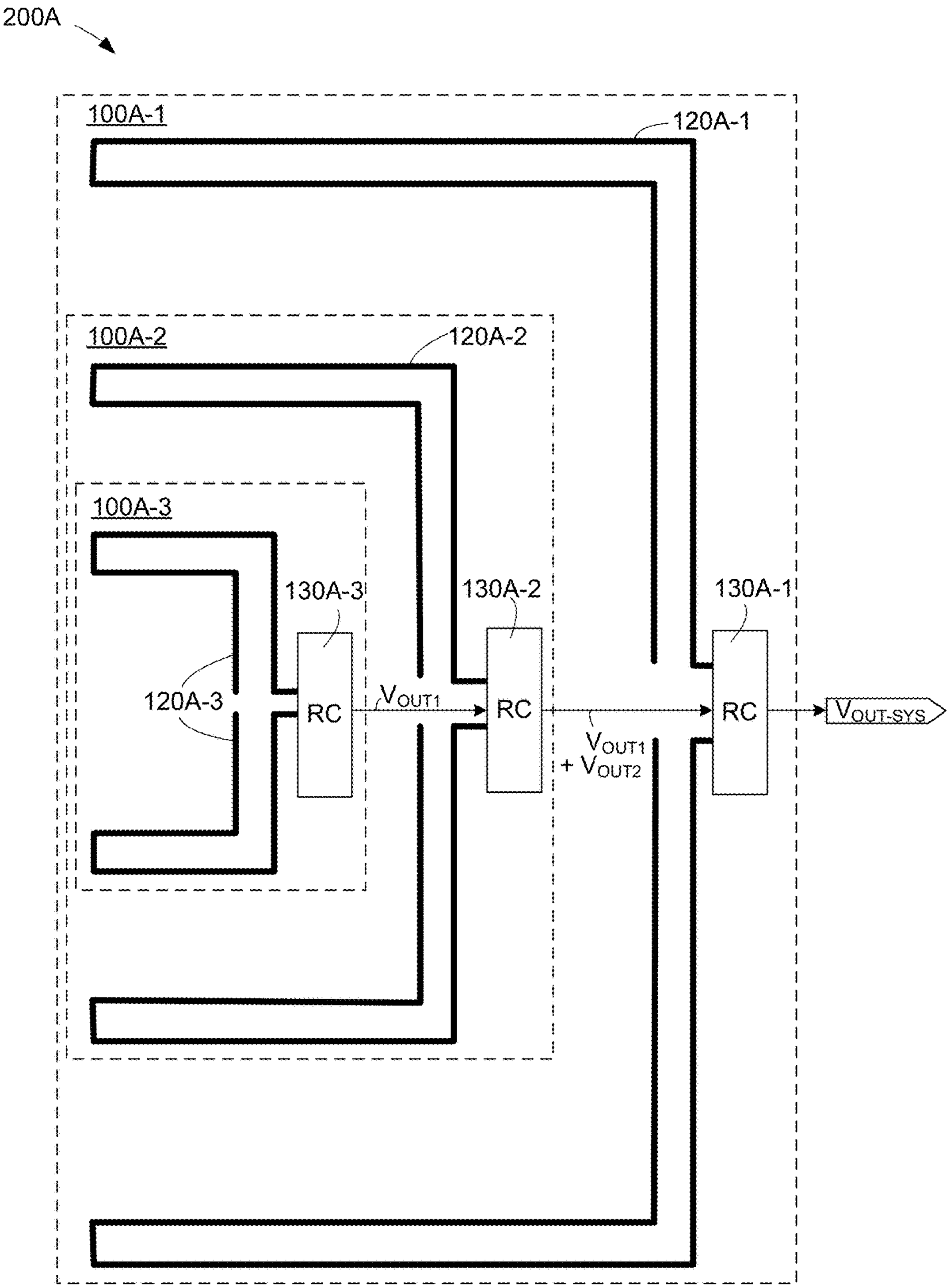
200A
100A-1
120A-1
100A-2
120A-2
100A-3
130A-3
130A-2
130A-1
120A-3
RC
RC
RC
V_{OUT1}
V_{OUT1} + V_{OUT2}
$V_{OUT-SYS}$

FIG. 3B

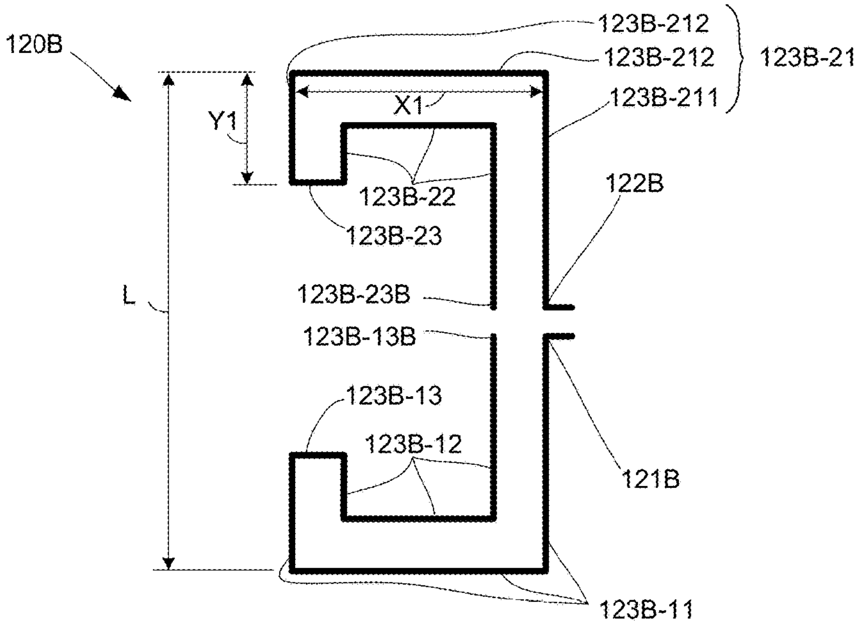
120B
123B-212
123B-212
123B-21
123B-211
X1
Y1
123B-22
123B-23
122B
L
123B-23B
123B-13B
123B-13
123B-12
121B
123B-11

FIG. 4A

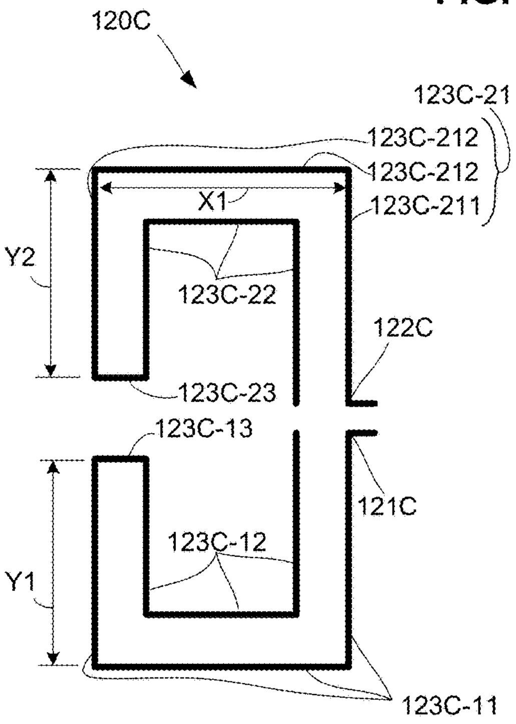
120C
123C-21
123C-212
123C-212
123C-211
X1
Y2
123C-22
122C
123C-23
123C-13
121C
123C-12
Y1
123C-11

FIG. 4B

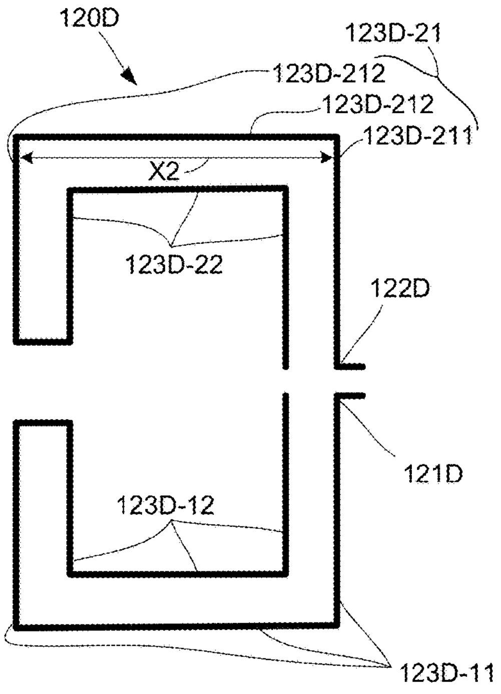
120D
123D-21
123D-212
123D-212
123D-211
X2
123D-22
122D
121D
123D-12
123D-11

FIG. 4C

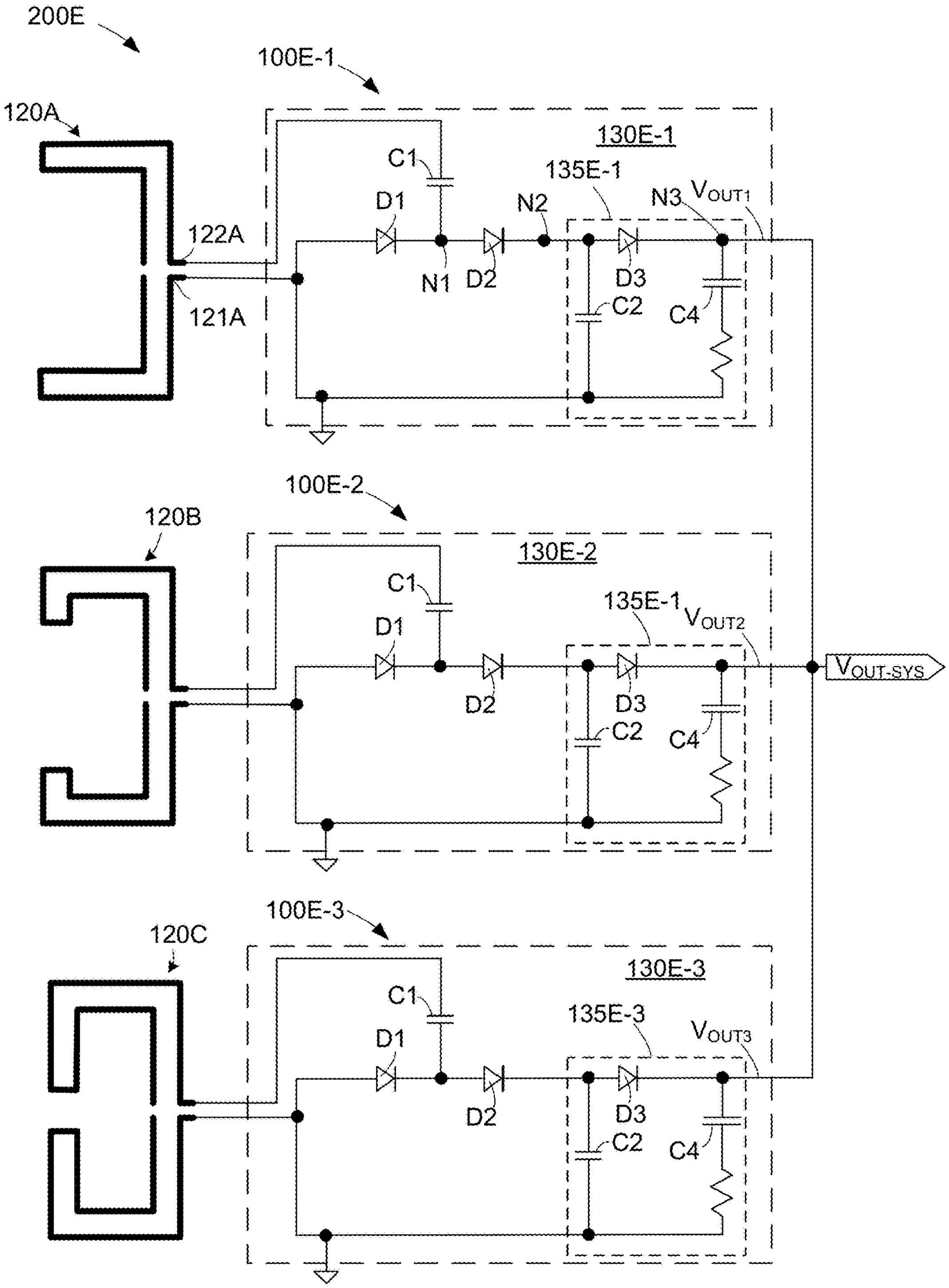
200E
100E-1
120A
130E-1
C1
135E-1
D1
N2
N3
VOUT1
122A
121A
N1
D2
D3
C2
C4
100E-2
120B
130E-2
C1
135E-1
VOUT2
D1
D2
D3
VOUT-SYS
C2
C4
100E-3
120C
130E-3
C1
135E-3
VOUT3
D1
D2
D3
C2
C4

FIG. 5

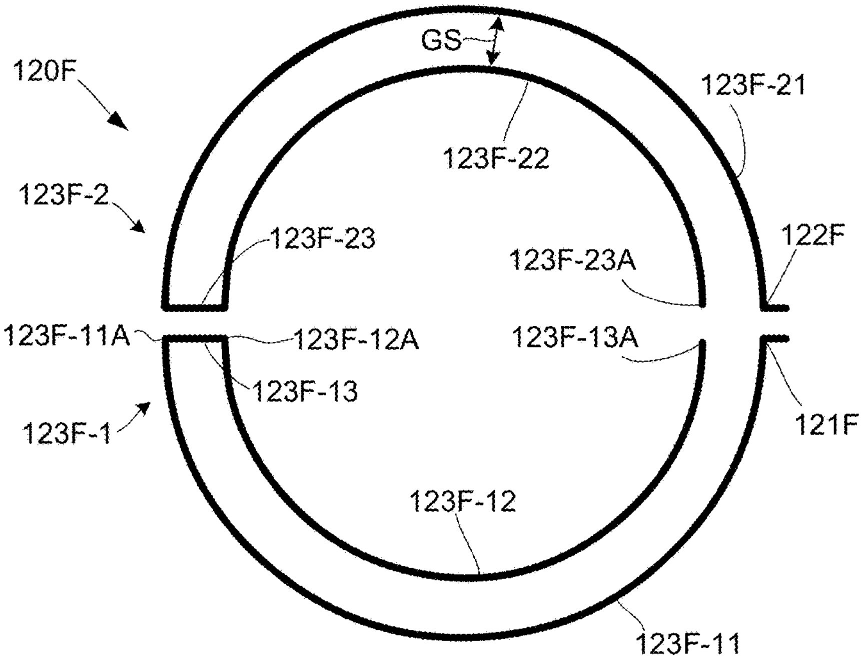
GS
120F
123F-21
123F-22
123F-2
123F-23
122F
123F-23A
123F-11A
123F-12A
123F-13A
123F-13
123F-1
121F
123F-12
123F-11

FIG. 6A

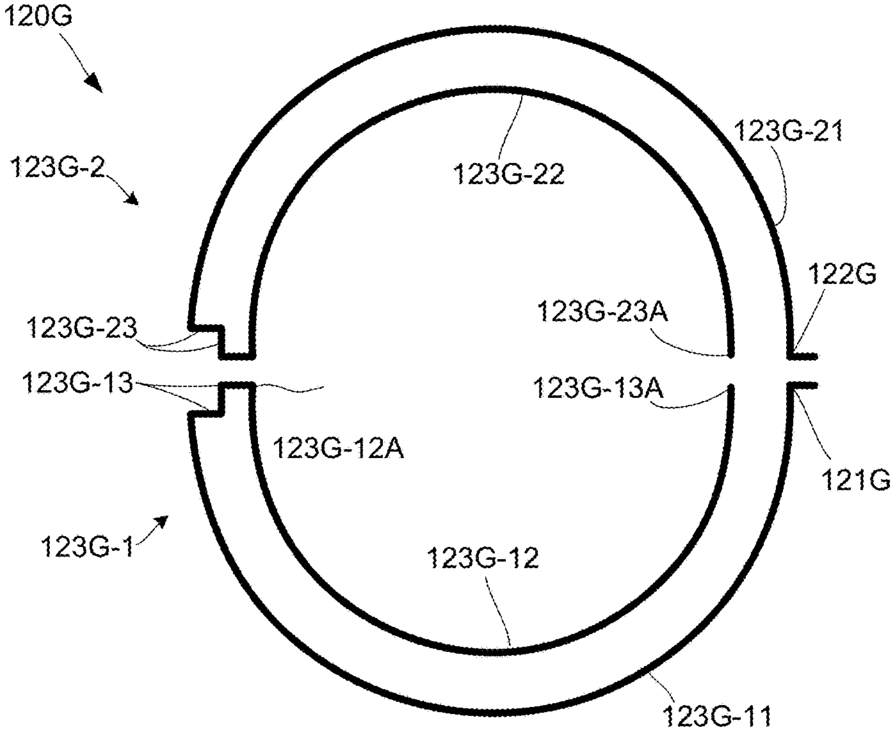
120G
123G-21
123G-2
123G-22
122G
123G-23A
123G-23
123G-13
123G-13A
123G-12A
121G
123G-1
123G-12
123G-11

FIG. 6B

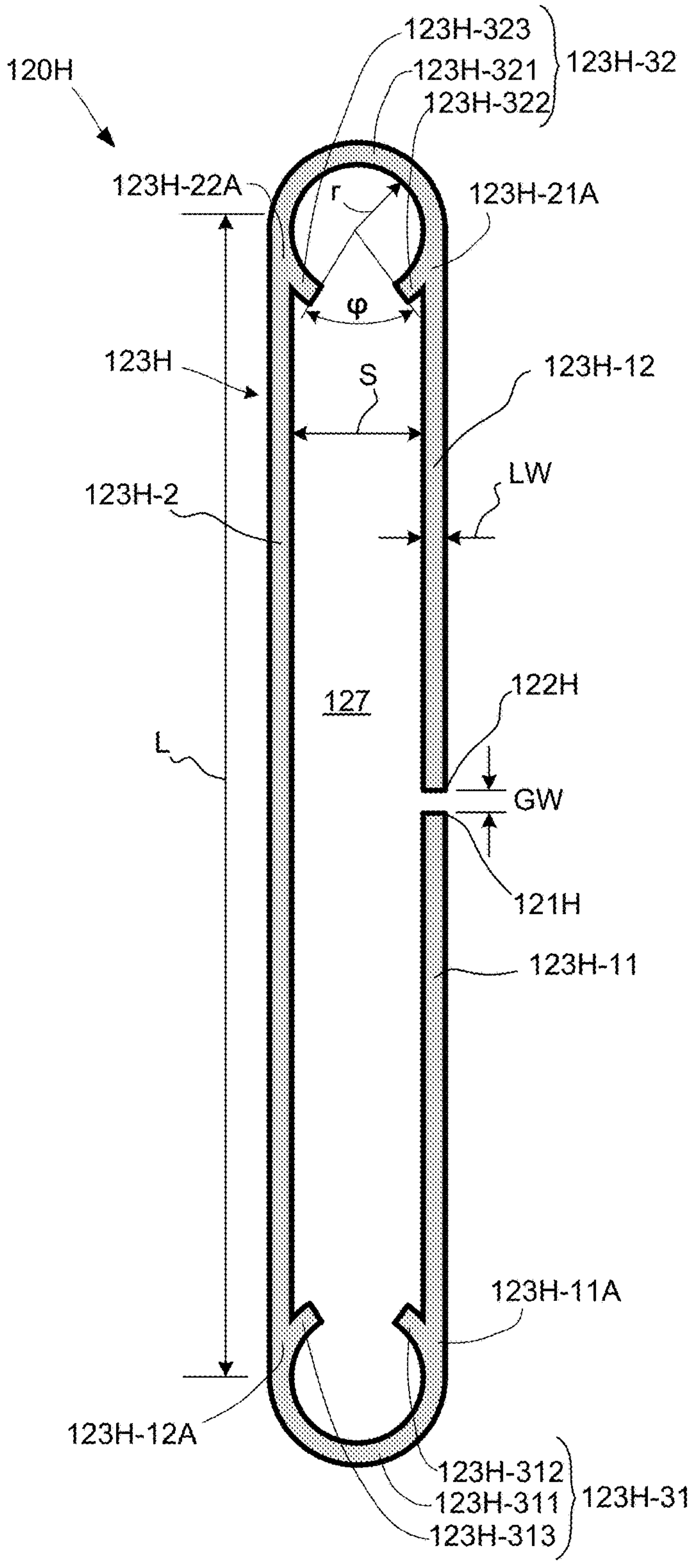
123H-323
123H-321
123H-322
123H-32
120H
123H-22A
r
123H-21A
φ
123H
S
123H-12
LW
123H-2
127
122H
L
GW
121H
123H-11
123H-11A
123H-12A
123H-312
123H-311
123H-31
123H-313

FIG. 7

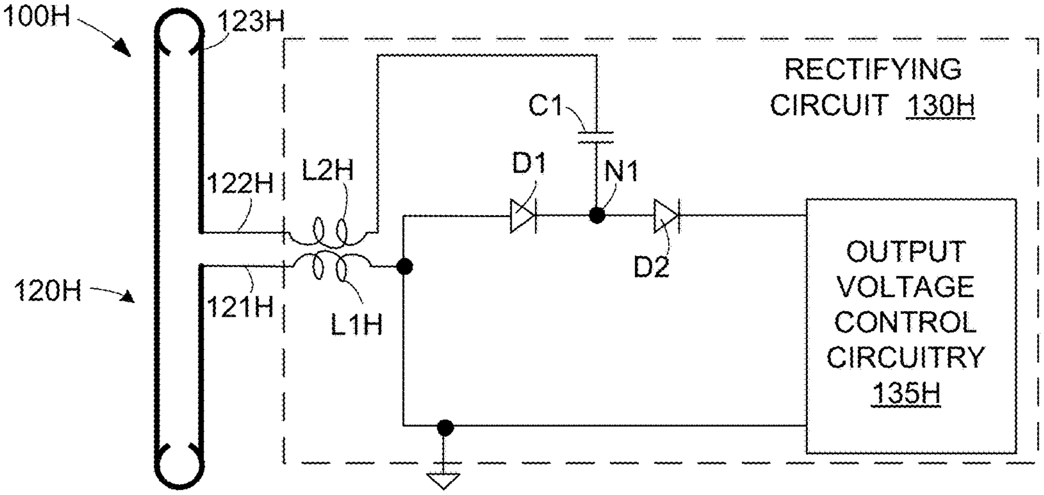
100H
123H
RECTIFYING
CIRCUIT 130H
C1
D1
N1
122H
L2H
120H
121H
L1H
D2
OUTPUT
VOLTAGE
CONTROL
CIRCUITRY
135H

FIG. 8

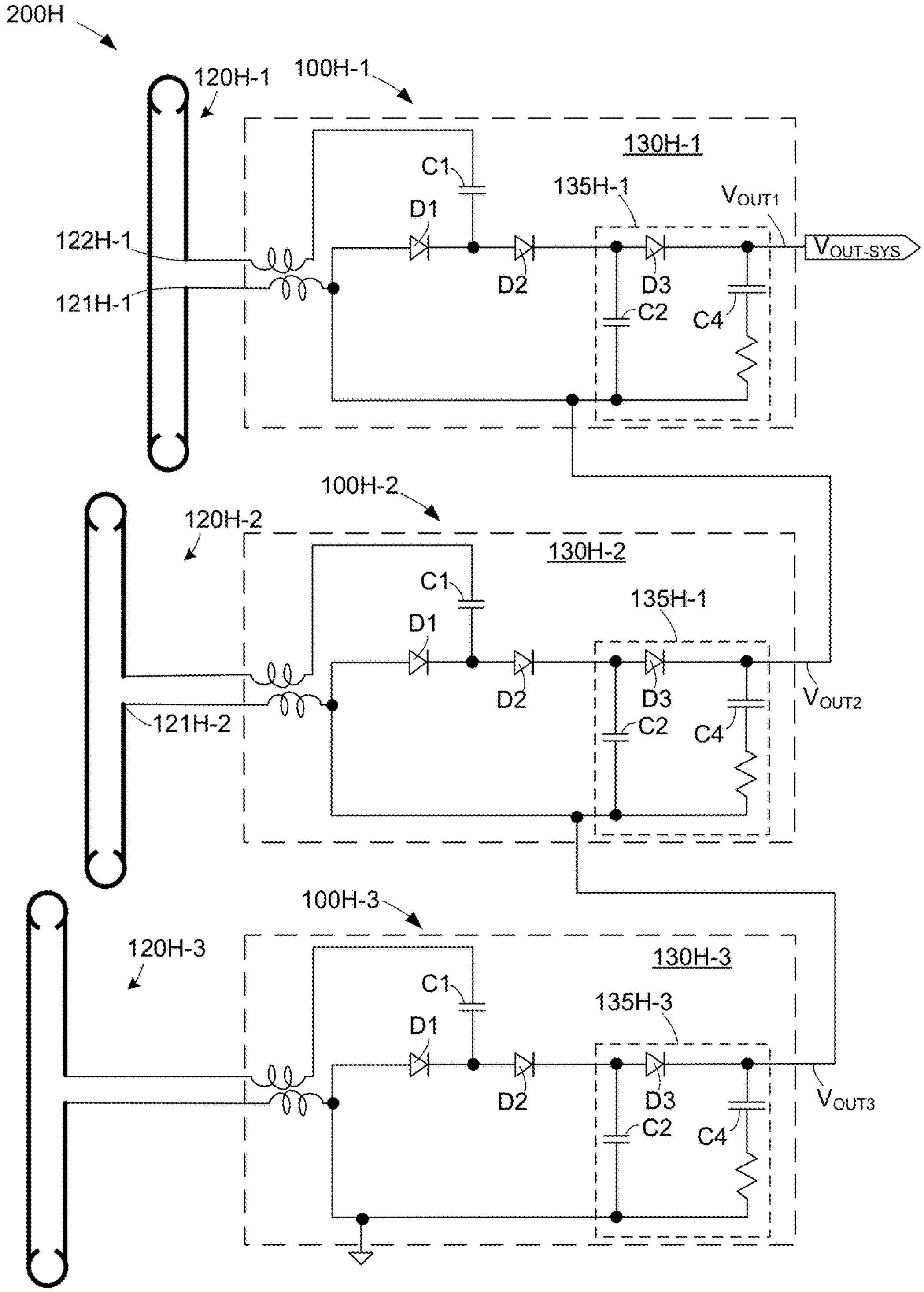
200H
120H-1
100H-1
130H-1
C1
135H-1
V_{OUT1}
D1
122H-1
$V_{OUT-SYS}$
D2
D3
121H-1
C2
C4
100H-2
120H-2
130H-2
C1
135H-1
D1
D2
D3
V_{OUT2}
121H-2
C2
C4
100H-3
120H-3
130H-3
C1
135H-3
D1
D2
D3
V_{OUT3}
C2
C4

FIG. 9

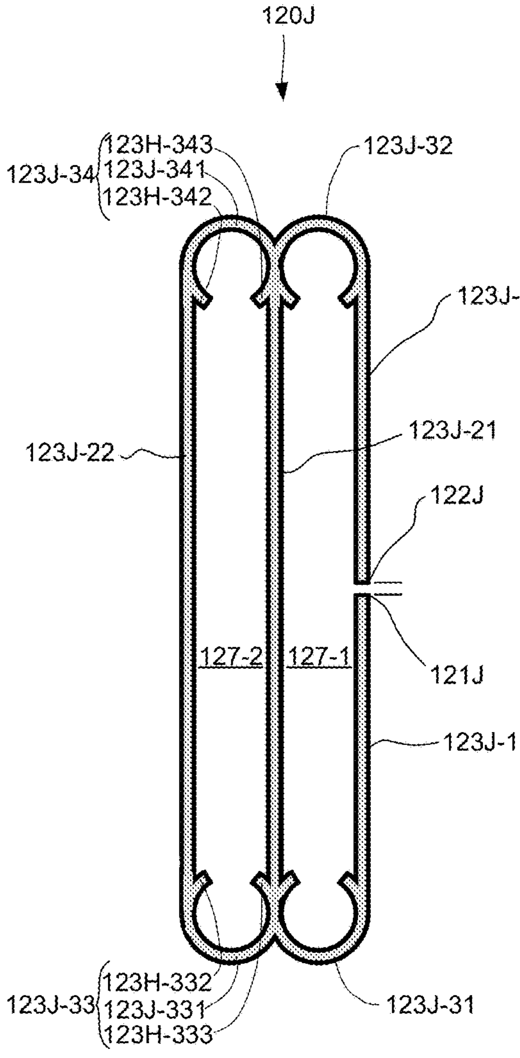
120J
123H-343
123J-34
123J-341
123H-342
123J-32
123J-2
123J-22
123J-21
122J
127-2
127-1
121J
123J-1
123H-332
123J-33
123J-331
123H-333
123J-31

FIG. 10A

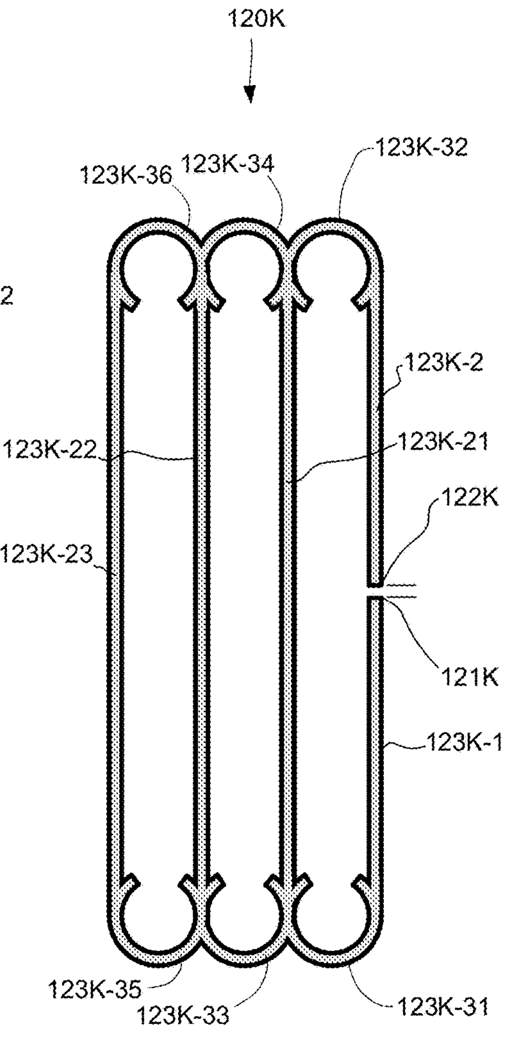
120K
123K-36
123K-34
123K-32
123K-2
123K-22
123K-21
122K
123K-23
121K
123K-1
123K-35
123K-33
123K-31

FIG. 10B

MULTIBAND RADIO FREQUENCY (RF) ENERGY HARVESTING WITH SCALABLE ANTENNA

FIELD OF THE INVENTION

This invention relates to multiband radio frequency (RF) energy harvesting, and more particularly to a low-cost, scalable RF energy harvesting device.

BACKGROUND OF THE INVENTION

A rectenna (rectifying antenna) is a special type of antenna that is used to convert radio wave energy into direct current electricity. Rectennas are used in wireless power transmission systems that transmit power by radio waves. A typical rectenna element consists of a dipole antenna with a diode connected across the dipole elements. The diode rectifies the alternating current (AC) induced in the antenna by the radio waves to produce direct current (DC) power, which is then used to power a load connected across the diode. Schottky diodes are usually used because they have the lowest voltage drop and highest speed and therefore have the lowest power losses due to conduction and switching. Large rectennas consist of an array of many such dipole elements.

There exists a significant motivation to develop rectennas that efficiently capture RF energy. Due to the increasing use of radio frequency (RF) signals to transmit information, the presence of RF energy is almost ubiquitous. Moreover, the unused portions of transmitted RF signals (i.e., portions that are not captured and converted by a receiver) are essentially "wasted" in that the energy transmitted in unused RF signals is absorbed in ambient matter and thus lost. As such, in addition to providing power sources for remote/portable devices, a suitable rectenna capable of efficiently re-converting unused ambient RF energy would reduce overall energy demand by replacing batteries and other remote power sources.

Conventional rectennas capable of converting RF energy to DC power are unable to harvest most ambient RF energy. Existing RF rectennas can only produce usable DC voltages in the presence of high RF power density levels (typically above 1 W/m^2), and are therefore typically used in wireless power transmission systems where dedicated high-power transmitters generate RF energy at sufficiently high energies. Ambient RF energy is typically very low (from mW down to μW), and thus the use of conventional RF rectennas for energy harvesting has proven to be insufficient in remote locations. Also, conventional rectenna devices utilize antennas are single band (i.e., harvest narrow-band RF signals) and rectifying circuits are bulky, and therefore not scalable.

What is needed is a low-cost, high-conversion-efficiency and electrically-small RF rectenna that is capable of reliably harvesting ambient RF energy. In particular, what is needed is an RF energy harvesting device capable of harvesting multiband RF signal energy having low energy levels (i.e., tens of mW and below) including an antenna that can be scaled to accommodate different RF frequencies, thereby facilitating energy generation for remote devices (e.g., sensors) using a wide range of RF signal sources.

SUMMARY OF THE INVENTION

The present invention is directed to space-efficient RF energy harvesting devices (rectennas) that utilize novel metamaterial resonator antennas that are electrically-small and are capable of near-perfect RF energy absorption. The metamaterial resonator antenna includes one or more metal structures that are conformally fixedly disposed on a substrate (e.g., formed by printing a conductive (e.g., silver) ink or depositing/etching a metal layer) in a manner that facilitates low-cost fabrication using printed electronics techniques. The at least one metal structure forming the metamaterial resonator antenna includes two primary antenna segments and at least one secondary antenna segment that are disposed in a spaced-apart (parallel) folded arrangement and connected by two linking segments such that the overall antenna structure resonates at RF frequencies that are within an associated bandwidth range, and are configured such that captured RF signals generated at the two antenna end points are 180° out-of-phase from each other. Metamaterial resonator antennas such as those described herein can be tailored to have high impedance and exhibit near-perfect absorption. The rectifying circuit is coupled to the antenna end points, and includes circuit elements (one or more diodes, capacitors, inductors and/or resistors) that are configured to combine positive voltage pulses from the two captured RF signals to generate a usable output voltage. In some embodiments, the metamaterial resonator antenna arrangement facilitates nesting a smaller antenna within open spaces of a larger antenna, thereby facilitating capturing a wide range of RF frequencies without significantly changing the device's footprint (size), and thereby providing a highly scalable platform that facilitates remotely powering widely dispersed IC devices (e.g., sensors, processors, etc.) using either pre-existing RF signals (e.g., from ambient Wi-Fi signals) or dedicated charging RF sources. Moreover, because the metamaterial resonator antenna arrangement is formed on a substrate and can be configured for different RF frequencies, the present invention facilitates the low-cost production of RF energy harvesting device using inexpensive flexible plastic substrates and cost-effective printed electronics techniques, thereby minimizing overall manufacturing costs while maximizing the RF frequency band that can be converted into usable energy.

According to an embodiment of the invention, the rectifying circuit includes conductive structures that are conformally disposed on the substrate surface (i.e., generated during the same ink-jet printing run) with the antenna in order to further minimize manufacturing costs. In one embodiment, the rectifying circuit utilizes a conventional voltage doubler circuit (e.g., a Greinacher circuit) to combine the two captured RF signals in order to generate a usable output voltage. In a presently preferred embodiment, the rectifying circuit utilizes two dissimilar Zero Bias Schottky diodes and a capacitor that are configured to combine the two captured RF signals in a way that generates an output voltage with minimal energy loss. When the metamaterial resonator antennas of the present invention are combined with these novel rectifying circuits, the resulting RF energy harvesting devices exhibit high energy conversion efficiencies.

According to various embodiments of the present invention, the metamaterial resonator antenna is implemented using a novel two-part dipole antenna arrangement made up of two bilaterally symmetrical metal elements. In this case, each metal element includes one of the primary antenna segments, one of the linking segment, and a secondary antenna segment portion that extending from the linking segment to a distal end point located adjacent to one of the antenna end point. The primary antenna segment and the secondary antenna segment portion are concentrically arranged (i.e., such that they are separated by a uniform gap spacing), which produces the desired captured RF signal characteristics. In one series of specific embodiments, the primary and secondary segment portions are formed by straight (linear) wire segments that are connected at one or more 90° angles (corners), wherein "tuning" the antenna to resonate at a targeted RF frequency involves increasing the length of one or more of the straight segments, or utilizing two or three interconnected segments. With this arrangement, the antenna can be "tuned" to capture a wide range of RF frequencies without substantially increasing the antenna footprint. In other embodiments the primary and secondary segment portions are formed by concentric curved (e.g., semi-circular) wire segments that are connected by a linking segment. In one embodiment the linking segment is implemented using a step-shaped wire that allows the primary and secondary segments to have identical lengths.

According to various other embodiments of the present invention, the metamaterial resonator antenna is implemented using a novel "racetrack-type" configuration formed by a single contiguous metal structure having a shape similar to an oblong rectangle with rounded ends, where portions of the rounded end "curves" extends into the interior "infield" regions. In this case, the primary antenna segments are collinear and extend in opposite directions (e.g., downward and upward) from the antenna end points, a single secondary antenna segment is disposed parallel to and spaced from the primary segments, and a pair of "open circle" linking segments **123**H-**31**/**32** are used to connect the primary and secondary segments, thereby forming a racetrack-shaped structure including curved end segments extending from each of the "open circle" linking segments into the "infield" portion of the racetrack. This arrangement was found to capture RF signals over a substantially higher range than the two-part antenna design (described above), and is further "tunable" by way of changing the gap separating adjacent curved end points, or by adding one or more additional "racetracks" to the antenna structure.

The present invention is further directed to systems in which multiple RF energy harvesting devices including metamaterial resonator antennas and associated rectifying circuits are interconnected to generate a combined output voltage.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other features, aspects and advantages of the present invention will become better understood with regard to the following description, appended claims, and accompanying drawings, where:

FIG. **1** is a top side perspective view showing an RF energy harvesting device according to an embodiment of the present invention;

FIG. **2** is a top side perspective view showing an RF energy harvesting device including first two-part antenna configuration according to a specific embodiment of the present invention;

FIG. **3**A is a simplified diagram showing an RF energy harvesting device including the two-part antenna of FIG. **2**;

FIG. **3**B is a simplified diagram showing an RF energy harvesting system including three RF energy harvesting devices of FIG. **3**A in a nested arrangement;

FIGS. **4**A, **4**B and **4**C are simplified diagrams showing two-part antenna configurations according to alternative specific embodiments of the present invention;

FIG. **5** is a simplified diagram showing an RF energy harvesting system utilizing two-part antennas according to another specific embodiment of the present invention;

FIGS. **6**A and **6**B are simplified diagrams showing curved two-part antenna configurations according to alternative specific embodiments of the present invention;

FIG. **7** is a top plan view showing a "racetrack-type" metamaterial resonator antenna according to another specific embodiment of the present invention;

FIG. **8** is a simplified diagram showing an RF energy harvesting device including the "racetrack-type" antenna of FIG. **7**;

FIG. **9** is a simplified diagram showing an RF energy harvesting system utilizing "racetrack-type" antennas according to another specific embodiment of the present invention; and

FIGS. **10**A and **10**B are top plan views showing two-loop and three-loop "racetrack" antennas, respectively, according to alternative embodiments of the present invention.

DETAILED DESCRIPTION OF THE DRAWINGS

The present invention relates to an improvement in RF energy harvesting. The following description is presented to enable one of ordinary skill in the art to make and use the invention as provided in the context of a particular application and its requirements. The terms "coupled" and "connected", which are utilized herein, are defined as follows. The term "connected" is used to describe a direct connection between two circuit elements, for example, by way of a metal line formed in accordance with normal integrated circuit fabrication techniques. In contrast, the term "coupled" is used to describe either a direct connection or an indirect connection between two circuit elements. For example, two coupled elements may be directly connected by way of a metal line, or indirectly connected by way of an intervening circuit element (e.g., a capacitor, resistor, inductor, or by way of the source/drain terminals of a transistor). Various modifications to the preferred embodiment will be apparent to those with skill in the art, and the general principles defined herein may be applied to other embodiments. Therefore, the present invention is not intended to be limited to the particular embodiments shown and described, but is to be accorded the widest scope consistent with the principles and novel features herein disclosed.

FIG. **1** is a top side perspective view showing an RF energy harvesting device **100** according to a generalized embodiment of the present invention. RF energy harvesting device **100** utilizes a novel metamaterial resonator antenna **120** and a rectifying circuit **130**.

Metamaterial resonator antenna **120** includes at least one metal structure **123** that is conformally fixedly disposed on a substrate **101** (e.g., formed by printing a conductive (e.g., silver) ink or depositing/etching a metal layer) in a manner that facilitates low-cost fabrication using printed electronics techniques. Metal structure **123** includes two primary antenna segments **123-1** and **123-2** and at least one secondary antenna segment **123-2** that are disposed in a spaced-apart (parallel) folded arrangement and connected by two linking segments **123-31** and **123-32**. Specifically, first primary antenna segment **123-11** extends from a first antenna end point **121** to a first midpoint **123-11**A, and second primary antenna segment **123-12** extends from a second antenna end point **122** to a second midpoint **123-21**A. Secondary antenna segment **123-2** is disposed parallel to and spaced from the first and second primary antenna segments by a uniform gap distance S, with secondary antenna segment **123-2** having a first end disposed at a third midpoint **123-12**A and an opposing second end disposed at a fourth midpoint **123-22**A. First linking segment **123-31** is connected between first and third midpoints **123-11**A and **123-12**A, and second linking segment **123-32** connected between second and fourth midpoints **123-21**A and **123-22**A, with these linking segments being substantially shorter than primary and secondary segments. Each of these segments consists of an elongated metal line structure having a thickness T (e.g., 8 microns) and a line width LW (e.g., 2.5 mm). By adjusting the overall length of primary antenna segments **123-11** and **123-12** and secondary antenna segment **123-2**, the overall antenna structure is tailored to resonate at RF frequencies that are within an associated target bandwidth range, and are configured such that captured RF signals V_{121} and V_{122} generated at antenna end points **121** and **122** are 180° out-of-phase from each other. Metamaterial resonator antenna **120** is electrically-small and, when combined with certain rectifying circuits, facilitates near-perfect RF energy absorption. An advantage of metamaterial resonator antennas formed in accordance with the present invention, such as those described below, can be tailored to have high impedance and exhibit near-perfect absorption. The rectifying circuit **130** is coupled to the antenna end points **121**/**122**, and includes circuit elements (one or more diodes, capacitors, inductors and/or resistors) that are configured to combine positive voltage pulses from the two captured RF signals V_{121}&V_{122} to generate a usable output voltage V_{OUT}. In some embodiments, the metamaterial resonator antenna arrangement facilitates nesting a smaller antenna within open spaces of a larger antenna, thereby facilitating capturing a wide range of RF frequencies without significantly changing the device's footprint (size), and thereby providing a highly scalable platform that facilitates remotely powering widely dispersed IC devices (e.g., sensors, processors, etc.) using either pre-existing RF signals (e.g., from ambient Wi-Fi signals) or dedicated charging RF sources. Moreover, because the metamaterial resonator antenna arrangement is formed on a substrate and can be configured for different RF frequencies, the present invention facilitates the low-cost production of RF energy harvesting device using inexpensive flexible plastic substrates and cost-effective printed electronics techniques, thereby minimizing overall manufacturing costs while maximizing the RF frequency band that can be converted into usable energy.

Referring again to FIG. **1**, rectifying circuit **130** includes conductive structures **133-1** to **133-4** that are conformally disposed on the substrate surface **102** (i.e., generated during the same ink-jet printing run as that used to form antenna **120**) in order to further minimize manufacturing costs. In the illustrated and presently preferred embodiment, rectifying circuit **130** utilizes two dissimilar Zero Bias Schottky diodes D**1** and D**2** and a capacitor C**1** that are configured to combine the two captured RF signals V_{121} and V_{122} in a way that generates an output voltage V_{OUT} with minimal energy loss. Specifically, first diode D**1** is coupled by way of conductive structure **133-1** between first antenna end point **121** and a first node N**1** (which in this case is implemented by conductive structure **133-3**. A first terminal of capacitor C**1** is coupled to second antenna end point **122** by way of conductive structure **133-2**, and a second terminal is connected to conductive structure **133-3** (node N**1**). Second diode D**2** is connected between first node N**1** and a second node N**2**, which is implemented by conductive structure **133-4**. A second capacitor C**2** is coupled between first antenna end point **121** and node N**2**, and in one embodiment conductive structure **133-1** is coupled to a ground or other voltage source. In an exemplary embodiment, diode D**1** is implemented using a (first) Zero Bias Schottky diode having a very low forward voltage (e.g., 120 mV or less), a peak inverse voltage of 1.0 V or less, and a reverse current resistance of 5 kΩ. Diode D**2** is implemented using a (second) Zero Bias Schottky diode having a forward voltage of 150 mV or less (i.e., equal to or higher than that of diode D**1**), and a peak inverse voltage of at least 2.0 V (i.e., higher than that of diode D**1**), and a reverse current resistance of at least 10 kΩ (i.e., higher than that of diode D**1**). Capacitors C**1** and C**2** are either discrete (e.g., surface mounted) capacitor structures, or a fabricated capacitor structures (e.g., using printed electronics), and have a value determined in accordance with the characteristics (current and voltage) of DC output voltage V_{OUT}, and in a practical embodiment have a common capacitance value of 10 pF. With this arrangement, a first intermediate voltage generated at first node N**1** comprises a sum of positive voltage pulses of first RF signal V_{121}, which are passed through first diode D**1**, and second AC signal V_{122}, which is passed through first capacitor C**1**. With this arrangement, a second intermediate voltage is generated at second node N**2** in accordance with positive voltage pulses that are passed from the first node N**1** through the second diode D**2**. When the metamaterial resonator antennas of the present invention are combined with this novel rectifying circuit, the resulting RF energy harvesting devices exhibit high energy conversion efficiencies than when conventional rectifying circuits are utilized. However, in other embodiments (not shown), RF energy harvesting devices including the metamaterial resonator antennas of the present invention and conventional rectifying circuits (e.g., using a voltage doubler circuit, such as a Greinacher circuit, to combine the two captured RF signals) may be beneficially utilized to generate a usable output voltage.

FIG. **2** is a top side perspective view showing an RF energy harvesting device **100**A including an open C-shaped two-part antenna **120**A formed by metal structures that are conformally disposed on upper surface **102**A of flexible plastic substrate **101**A. Antenna **120**A represents a first example of a two-part dipole antenna arrangement made up of two bilaterally symmetrical metal elements **123**A**-1** and **123**A**-2**. Metal element **123**A**-1** includes a first primary antenna segment **123**A**-11** extending from first antenna end point **121**A, a first linking segment **123**A**-13** attached to the end of primary antenna segment **123**A**-11**, and a first secondary antenna segment **123**A**-12** extending from linking segment **123**A**-13** to a first distal end point **123**A**-13**A disposed adjacent to first antenna end point **121**A. Note that first primary antenna segment **123**A**-11** and first secondary antenna segment **123**A**-12** are concentrically arranged such that a uniform gap spacing GS separates first primary antenna segment **123**A**-11** and first secondary antenna segment **123**A**-12** along their entire length. Similarly, second metal element **123**A**-2** includes a second primary antenna segment **123**A**-21** extending from second antenna end point **122**A, a second linking segment **123**A**-23**, and a second secondary antenna segment **123**A**-22** extending from second linking segment **123**A**-23** to a second distal end point **123**A**-23**A disposed adjacent to second antenna end point **122**A, where second primary antenna segment **123**A**-21** and second secondary antenna segment **123**A**-22** are also concentrically arranged and separated by gap spacing GS along their entire length. Metal elements **123**A**-1** and **123**A**-2** collectively form an open C-shaped pattern surrounding central region **127** on three sides.

FIG. **3**A is a simplified diagram showing RF energy harvesting device **100**A of FIG. **2**, where device **100**A shows open C-shaped two-part antenna **120**A using simple line structures in order to simplify the description of other embodiments set forth below. That is, FIG. **3**A shows in simplified form antenna **120**A coupled to a generalized rectifying circuit **130**A, where primary antenna segments **123**A-**11** and **123**A-**21**, secondary segments **123**A-**12** and **123**A-**22**, and linking segments **123**A-**13** and **123**A-**23** are indicated by solid black lines. As further illustrated in FIG. **3**A, each of primary antenna segments **123**A-**11** and **123**A-**21** and secondary segments **123**A-**12** and **123**A-**22** comprises at least two linear segments connected by a corner portion. For example, referring to the upper portion of FIG. **3**A, second primary antenna segment **123**A-**21** includes two linear segments **123**A-**211** and **123**A-**212** connected by a corner (90° angle) portion **123**A-**21**A.

FIG. **3**B is a simplified diagram showing an RF energy harvesting system **200**A including three RF energy harvesting devices **100**A-**1**, **100**A-**2** and **100**A-**3** configured in a nested arrangement that illustrates an advantage of the open C-shaped configuration of antenna **120**A (described above). Specifically, devices **100**A-**1**, **100**A-**2** and **100**A-**3** respectively include three open C-shaped antennas **120**A-**1**, **120**A-**2** and **120**A-**3**, each coupled to an associated rectifying circuit **130**A-**1**, **130**A-**2** and **130**A-**3** facilitate generate an combined system output voltage $V_{OUT-SYS}$ utilizing, for example, a parallel configuration similar to that described below with reference to FIG. **5**. As illustrated in FIG. **3**B, an advantage to forming metamaterial resonant antennas using the open C-shaped configuration described above is that multiple antennas of different sized can be concentrically "nested", which facilitates capturing RF energy over a wide range of RF frequencies without increasing the overall device/system footprint. That is, because antennas **120**A-**1**, **120**A-**2** and **120**A-**3** have different sizes, each has a different inductance and hence resonates at a different RF frequency. By nesting antennas **120**A-**2** and **120**A-**3** inside the footprint of antenna **120**A-**1**, system **200**A facilitates capturing energy over a wider range of RF frequencies than would be possible using antenna **120**A-**1** alone, without increasing the substrate area required for antenna **120**A.

FIGS. **4**A, **4**B and **4**C are simplified diagrams showing three additional two-part dipole antenna configurations that illustrate other approaches (i.e., in addition to increasing antenna size, as illustrated in FIG. **3**B) for tailoring antenna inductance according to alternative specific embodiments of the present invention. In each case, the primary antenna segments and secondary antenna segments include three interconnected linear sections that collectively form a C-shaped pattern, where dimensions of the three linear sections are changed to illustrate how metamaterial resonator antennas are tailored to achieve difference inductance values, whereby the antennas are altered to resonate at (and thereby collect energy from) different RF frequencies. For example, referring to FIG. **4**A, antenna **120**B includes primary segments **123**B-**11** and **123**B-**21** and secondary segments **123**B-**12** and **123**B-**22** that are connected by linking sections **123**B-**13** and **123**B-**23** in a manner similar to that described above with reference to antenna **120**A (FIG. **2**). Assuming the overall length L and width X1 of antenna **120**B is the same as that of antenna **120**A (FIG. **2**), FIG. **4**A illustrates that the inductance of antenna **120**B can be tailored to achieve different inductances with little or minimal change to the overall antenna footprint by adding a third (vertical) section to each of the primary and secondary antenna segments. For example, second primary antenna segment **123**B-**21** includes three interconnected linear sections **123**B-**211**, **123**B-**212** and **123**B-**213** that form a C-shaped pattern, where vertical antenna section **123**B-**213** increases the length of primary antenna segment **123**B-**21** by length Y**1** without increasing the overall footprint of antenna **120**B. FIG. **4**B indicates that further tailoring is achieved by further lengthening these third (vertical) sections. Antenna **120**C includes three-section primary segments **123**C-**11** and **123**C-**21** and secondary segments **123**C-**12** and **123**C-**22** that are connected by linking sections **123**C-**13** and **123**C-**23** in a manner similar to that of antenna **120**B (FIG. **4**A), however the inductance of antenna **120**C is altered by way of lengthening the additional section. Specifically, second primary antenna segment **123**C-**21** includes three interconnected linear sections **123**C-**211**, **123**C-**212** and **123**C-**213**, where vertical antenna section **123**C-**213** has a length Y**2** that increases the inductance of antenna **120**C over that of antenna **120**B without increasing the overall footprint of antenna **120**C. FIG. **4**C illustrates antenna **120**D, which achieves even higher inductance by way of lengthening the horizontal sections of primary segments **123**D-**11** and **123**D-**21** and secondary segments **123**D-**12** and **123**D-**22** to width X**2**. For example, second primary antenna segment **123**D-**21** includes three interconnected linear sections **123**D-**211**, **123**D-**212** and **123**D-**213**, where horizontal antenna section **123**D-**212** has a width X**2** that increases the inductance of antenna **120**D over that of antenna **120**C without significantly increasing the overall footprint of antenna **120**C.

FIG. **5** is a simplified circuit diagram showing a RF energy harvesting system **200**E including RF energy harvesting devices **100**E-**1**, **100**E-**2** and **100**E-**3** that are connected in parallel between common ground sources and common system output nodes to maximize the current level of the system output voltage $V_{OUT-SYS}$. To facilitate the generation of combined system output voltage $V_{OUT-SYS}$, devices **100**E-**1**, **100**E-**2** and **100**E-**3** respectively include similar rectifying circuits **130**E-**1**, **130**E-**2** and **130**E-**3** that are optimized to provide their respective output voltages V_{OUT1} to V_{OUT3} in a manner that facilitates combination. Each rectifying circuit **130**E-**1**, **130**E-**2** and **130**E-**3** includes diodes D**1** and D**2** and capacitors C**1** configured as described above with reference to FIG. **1**. In addition, rectifying circuits **130**E-**1**, **130**E-**2** and **130**E-**3** are similar in that they respectively include output control circuits **135**E-**1**, **135**E-**2** and **135**E-**3** configured to convert the intermediate voltage generated at node N**2** into DC output voltages V_{OUT1} to V_{OUT3}. To facilitate combining these output voltages, output control circuits **135**E-**1**, **135**E-**2** and **135**E-**3** include a (second) capacitor C**2** that has essentially the same capacitance as capacitors C**1**, a third diode D**3** and a storage capacitor C**4** that are configured to store DC output voltages V_{OUT1} to V_{OUT3} for combination at system output node $V_{OUT-SYS}$. For example, referring to output control circuit **135**E-**1**, diode D**3** is connected between nodes N**2** and N**3** (which serves as an output node for rectifying circuit **130**E-**1**), a first terminal of storage capacitor C**4** is connected to node N**3**, and a resistor R**2** connected between a second terminal of storage capacitor C**4** and first antenna end point **121**A (e.g., to the applied ground potential). In a practical embodiment, diode D**3** is implemented using the same diode element as that utilized to implement diode D**2** (e.g., a HSMS2850 Surface Mount Zero Bias Schottky Detector Diode manufactured by Agilent Technologies of Santa Clara, Calif., USA). Capacitor C**3** is implemented using a capacitive element sized in accordance with the characteristics (current and voltage) of DC output voltage V_{OUT}, and the size and characteristics of the system voltage present on system output node $V_{OUT-SYS}$. In an exemplary practical embodiment, utilizing the diode and capacitor component values mentioned above, capacitor C**3** has a capacitance of 100 pF, and resistor R**2** has a nominal resistive value of 1Ω. Output control circuit **135**E-**2** and **135**E-**3** are configured identically to output control circuit **135**E-**1**.

According to an aspect of the embodiment shown in FIG. **5**, each device **100**E-**1**, **100**E-**2** and **100**E-**3** includes a different two-part dipole metamaterial resonator antenna **120**A, **120**B and **120**C, whereby devices **100**E-**1**, **100**E-**2** and **100**E-**3** are configured to convert RF energy having different frequencies into output voltages V_{OUT1}, V_{OUT2} and V_{OUT3}. Two-part dipole metamaterial resonator antenna **120**A, **120**B and **120**C are described in detail above. Because antennas **120**A, **120**B and **120**C have different configurations (sizes), each generates an output voltage dependent on the availability of sufficient RF energy at the frequencies to which each antenna resonates. System **200**E is therefore configured to generate output current over a much broader RF frequency range than would be possible using a single antenna by utilizing three different antenna arrangements.

FIGS. **6**A and **6**B are simplified diagrams showing two curved two-part metamaterial resonator antennas **120**F and **120**G according to additional alternative specific embodiments of the present invention. Similar to the straight-section embodiments described above, each antenna **120**F and **120**G includes two bilaterally symmetrical metal elements **123**F-**1** and **123**F-**2** and **123**G-**1** and **123**G-**2**, each including concentric, continuously-curved (e.g., substantially semi-circular) primary and secondary segments wire segments that are connected by a short linking segment.

Referring to FIG. **6**A, metal element **123**F-**1** of antenna **120**F includes a curved primary segment **123**F-**11** extending from antenna end point **121**F to a (first) midpoint **123**A-**11**A, a straight linking segment **123**F-**13** extending from midpoint **123**F-**11**A to midpoint **123**F-**12**A, and a curved secondary antenna segment **123**F-**12** extending from the second midpoint **123**F-**12**A to a distal end point **124**F-**13**A. Curved primary segment **123**F-**11** and curved secondary antenna segment **123**F-**12** are concentrically arranged such that they are separated along their entire length by a uniform gap spacing GS. Similarly, metal element **123**F-**2** includes a curved primary segment **123**F-**21** extending from antenna end point **121**F to straight linking segment **123**F-**23**, and a curved secondary antenna segment **123**F-**22** extending from linking segment **123**F-**23** to a distal end point **124**F-**23**A, where curved primary segment **123**F-**21** and curved secondary antenna segment **123**F-**22** are concentrically arranged.

Referring to FIG. **6**B, metal element **123**G-**1** of antenna **120**G also includes a curved primary segment **123**G-**11** extending from antenna end point **121**G to a midpoint **123**A-**11**A, a straight linking segment **123**G-**13** extending from midpoint **123**G-**11**A to midpoint **123**G-**12**A, and a curved secondary antenna segment **123**G-**12** extending from the second midpoint **123**G-**12**A to a distal end point **124**G-**13**A, with curved primary segment **123**G-**11** and curved secondary antenna segment **123**G-**12** being concentrically arranged, and metal element **123**G-**2** includes a curved primary segment **123**G-**21** extending from antenna end point **122**G to straight linking segment **123**G-**23**, and a curved secondary antenna segment **123**G-**22** extending from linking segment **123**G-**23** to a distal end point **124**G-**23**A, where curved primary segment **123**G-**21** and curved secondary antenna segment **123**G-**22** are concentrically arranged. Antenna **120**G differs from antenna **120**F in that, where linking segments **123**F-**13** and **123**F-**23** of antenna **120**F are implemented by straight (linear shaped) wire structures, linking segments **123**G-**13** and **123**G-**23** of antenna **120**G comprise two-segment step-shaped wire structures arranged such that the length primary antenna segments **123**G-**11** and **123**G-**21** is equal to the lengths of first secondary antenna segments **123**G-**12** and **123**G-**22**, which facilitates improved antenna performance.

FIG. **7** is a top plan view showing a "racetrack-type" metamaterial resonator antenna **120**H according to another specific embodiment of the present invention. Unlike the two-part dipole embodiments described above, antenna **120**H has a single-piece (integral, contiguous) metal element **123**H including (first and second) primary antenna segments **123**H-**11** and **123**H-**12** disposed on a first side of the "racetrack", a secondary antenna segment **123**H-**2** disposed on the second side of the "racetrack", and two open-circle linking segments **12**H-**31** and **123**H-**32** forming the two "curves" of the "racetrack". Primary antenna segments **123**H-**11** and **123**H-**12** are linear metal structures that are collinearly arranged and separated by a gap GW defined between the first and second antenna end points **121**H and **122**H. That is, referring to FIG. **7**, primary antenna segment **123**H-**11** extends straight downward from end point **121**H to a first midpoint **123**H-**11**A, and primary antenna segment **123**H-**21** extends upward from end point **122**H to a second midpoint **123**H-**21**A. Secondary antenna segment **123**H-**2** is a continuous metal structure disposed parallel to and spaced at a distance S from primary antenna segments **123**H-**11** and **123**-**12**, and extends substantially the entire length L between third and fourth midpoints **123**H-**12**A and **123**H-**22**A. Linking segment **123**H-**31** includes a semi-circular section **123**H-**311** connected between a lower end of primary antenna segment **123**H-**11** and a lower end of secondary antenna segment **123**H-**2** (i.e., between midpoints **123**H-**11**A and **123**H-**12**A), and (first and second) curved end sections **123**H-**312** and **123**H-**313** extending from semi-circular section **123**H-**311** into space **127** (i.e., into a region between primary antenna segment **123**H-**11** and secondary antenna segment **123**H-**2**). Linking segment **123**H-**32** includes a semi-circular section **123**H-**321** connected between upper ends of second primary antenna segment **123**H-**12** and secondary antenna segment **123**H-**2** (i.e., between midpoints **123**H-**21**A and **123**H-**22**A), and (third and fourth) curved end sections **123**H-**322** and **123**H-**323** extending from semi-circular section **123**H-**321** into space **127**. The resulting "racetrack-type" antenna configuration resonates at an associated bandwidth range such that captured RF signals generated on antenna end points **121**H and **122**H are 180° out-of-phase from each other. An advantage provided by antenna **120**H is that this "racetrack-type" antenna configuration was found to capture RF signals over a substantially higher frequency range than the two-part dipole antenna configurations described above. Similar to the two-part configurations, the impedance of antenna **120**H is tunable by way of increasing or decreasing the antenna footprint (size). In addition, the "racetrack-type" antenna configuration facilitates further "tunability" without changing the overall footprint by way of changing the gaps separating the curved end sections (e.g., the distance separating end sections **123**H-**322** and **123**H-**323**), or by adding one or more additional "racetrack loops" to the antenna structure, as described in additional detail below. In a practical embodiment, a "racetrack-type" antenna was formed by the present inventors using silver ink (10μΩ/cm) disposed on a PET substrate (250 μm thick) with segments having a thickness of 8 μm and line widths LW of 2.5 mm, where racetrack structure included a straight-section length L of 120 mm and linking segments having an inner radius r of 6.45 mm (i.e., having an overall antenna length of approximately 5.8 inches), with the curved end segments formed using an opening angle φ of 60°, and with a gap width GW of 2.5 mm. The resulting structure was found to resonate with peak efficiency at 914 MHz, and exhibited less than 10% loss in accepted power in the range of approximately 825 MHz to 1050 MHz.

FIG. **8** is a simplified diagram showing an RF energy harvesting device **100**H including "racetrack-type" metamaterial resonator antenna **120**H of FIG. **7**. In this case, rectifying circuit **130**H includes two discrete inductors L1H and L2H that are needed to facilitate proper transmission of captured RF signals to the other elements of rectifying circuit **130**H. Specifically, inductor L1H is coupled between first antenna end point **121**H and diode D**1**, and inductor L2H is coupled between antenna end point **122**H and capacitor C**1**. In one embodiment, inductors L1H and L2H have the same (common) inductance value, and form a balanced circuit with antenna **120**H and the elements of rectifying circuit **130**H (i.e., capacitor C**1**, diodes D**1** and D**2**, and output control circuit **135**H, which may be implemented using any of the configurations described in copending application Ser. No. 14/582,002 entitled "RECTIFYING CIRCUIT FOR MULTIBAND RADIO FREQUENCY (RF) ENERGY HARVESTING", which is incorporated herein by reference in its entirety.

FIG. **9** is a simplified diagram showing an RF energy harvesting system **200**H made up of three devices **100**H-**1**, **100**H-**2**, and **100**H-**3** respectively including "racetrack-type" metamaterial resonator antennas **120**H-**1** to **120**H-**3**, where the output terminals of devices **100**H-**1**, **100**H-**2**, and **100**H-**3** are connected in series such that the second antenna end point of at least one RF energy harvesting device is connected to the output node of an adjacent RF energy harvesting device. According to an aspect of the disclosed embodiment, each energy harvesting device **100**H-**1** to **100**H-**3** includes a substantially identical "racetrack" antenna configuration. For example, device **100**H-**1** includes "racetrack" antenna **120**H-**1** configured as described above to resonate at RF frequencies within an associated bandwidth range. Device **100**H-**2** includes antenna **120**H-**2** and device **100**H-**3** includes antenna **120**H-**3**, where antennas **120**H-**2** and **120**H-**3** are substantially identical in configuration with antenna **120**H-**1** (i.e., such that all three antennas resonate at the same RF frequency, and such that all three antennas generate substantially the same amount of energy). Each energy harvesting device **100**H-**1** to **100**H-**3** also includes a substantially identical rectifying circuit. For example, device **100**H-**1** includes rectifying circuit **130**H-**1** including a first diode D**1** coupled between antenna end point **121**H-**1** and a first node N**1**, a first capacitor C**1** coupled between end point **122**H-**1** and node N**1**, and a second diode D**2** connected between node N**1** and a second node N**2**. Rectifying circuit **130**H-**1** also includes an output control circuit **135**H-**1** coupled to node N**2** that is configured (i.e., by way of third diode D**3**, second capacitor C**2** that has the same capacitance as capacitor C**1**, and storage capacitor C**4**) to convert the intermediate voltage generated on node N**2** into DC output voltage V_{OUT1}. Similarly, energy harvesting devices **100**H-**2** and **100**H-**3** respectively include rectifying circuits **130**H-**2** and **130**H-**3** having diodes D**1** and D**2** and capacitor C**1**, and output control circuits **135**H-**2** and **135**H-**3** including diodes D**3** and capacitors C**2** and C**4**) that generate respective DC output voltages V_{OUT2} and V_{OUT3}.

According to the series-connected arrangement implemented by system **200**H, RF energy harvesting devices **100**H-**1** to **100**H-**3** are connected such that DC output voltages V_{OUT1}, V_{OUT2} and V_{OUT3} are combined such that system output voltage $V_{OUT\text{-}SYS}$ is maximized. Specifically, first antenna end point **121**H-**2** of RF energy harvesting device **100**H-**2** is connected to the output node of RF energy harvesting device **100**H-**3** (i.e., such that the ground plane of device **100**H-**2** is DC output voltage V_{OUT3}), whereby a voltage level of DC output voltage V_{OUT2} generated at the output node of device **100**H-**2** is "boosted" by DC output voltage V_{OUT3}. Similarly, antenna end point **121**H-**1** of RF energy harvesting device **100**H-**1** is connected to the output node of RF energy harvesting device **100**H-**3**, whereby a voltage level of DC output voltage V_{OUT3} generated at the output node of device **100**H-**1** is "boosted" by DC output voltages V_{OUT2} and V_{OUT3}. Hence, the voltage level of system output voltage $V_{OUT\text{-}SYS}$ is maximized by the series arrangement.

Although not illustrated, it is understood that the "racetrack-type" antenna configuration may also be used in a nested-device arrangement similar to that shown and described above with reference to FIG. **3**B.

FIGS. **10**A and **10**B are top plan views respectively showing a two-loop "racetrack-type" metamaterial resonator antenna **120**J and a three-loop "racetrack" "racetrack-type" metamaterial resonator antenna according to alternative embodiments of the present invention. In addition to tailoring the impedance of a racetrack antenna by altering the length and radius diameters of the "racetrack" shape, and by changing the gap distance between the curved end portions (as described above), it is also possible to tailor the antenna's impedance by way of adding one or more additional "loop" structures (i.e., such that the single integral metal antenna structure includes two or more "loops" as described below).

Referring to FIG. **10**A, similar to the embodiment described above with reference to FIG. **7**, antenna **120**J includes a first "loop" structure surrounding region **127**-**1** formed by primary antenna segments **123**J-**11** and **123**J-**12** respectively extending from end points **121**J and **122**J, a (first) secondary antenna segment **123**J-**21**, and (first and second) circular linking segments **123**J-**31** and **123**J-**32**, which are configured in the manner described above with reference to FIG. **7**. In addition, antenna **120**J includes a second "loop" structure surrounding region **127**-**2** formed by first secondary antenna segment **123**J-**21**, a second secondary antenna segment **123**J-**22** disposed parallel to primary antenna segments **123**J-**11** and **123**J-**12** and to first secondary antenna segment **123**J-**21** (i.e., such that first secondary antenna segment **123**J-**21** is between primary antenna segments **123**J-**11** and **123**J-**12** and second secondary antenna segment **123**J-**22**), a (third) linking segment **123**J-**33** including a third semi-circular section **123**J-**331** connected between the first lower end of the first secondary antenna segment **123**J-**21** and a first lower end of the second secondary antenna segment **123**J-**22**, and (fifth and sixth) curved end sections **123**J-**332** and **123**J-**333** extending from third semi-circular section **123**J-**331** into region **127**-**2**, and a (fourth) linking segment **123**J-**34** including a fourth semi-circular section **123**J-**341** connected between an upper end of secondary antenna segment **123**J-**22** and an upper end of the secondary antenna segment **123**J-**21**, and (seventh and eight) curved end sections **123**J-**342** and **123**J-**343** extending from semi-circular section **123**J-**341** into space **127**-**2**. The addition of the second "loop" structure increases the overall inductance of antenna **120**J over that of antenna **120**H (FIG. **7**), whereby antenna **120**J is tailored to resonate at a different RF frequency than antenna **120**H.

Referring to FIG. **10**B, similar to the embodiment described with reference to FIG. **10**A, antenna **120**K includes a first "loop" structure formed by primary antenna segments **123**J-**11** and **123**J-**12** extending from end points **121**K and **122**K, respectively, a (first) secondary antenna segment **123**J-**21**, and (first and second) circular linking segments **123**J-**31** and **123**J-**32**, and a second "loop" structure formed by (first) secondary antenna segment **123**J-**21**, a (second) secondary antenna segment **123**J-**22**, and (third and fourth) circular linking segments **123**J-**33** and **123**J-**34**. In addition, antenna **120**J includes a third "loop" structure formed by (second) secondary antenna segment **123**J-**22**, a (third) secondary antenna segment **123**J-**23**, a (fifth) linking segment **123**J-**35** connecting lower ends of second antenna segments **123**J-**22** and **123**J-**23**, and a (sixth) linking segment **123**J-**36** connecting upper ends of second antenna segments **123**J-**22** and **123**J-**23**. The addition of the third "loop" structure further increases the overall inductance of antenna **120**K over that of antenna **120**J (FIG. **10**A).

Although fabrication costs associated with producing the RF energy harvesting devices/systems of the present invention are minimized using the printing techniques described above, many of the benefits described herein may be achieved using other fabrication techniques. For example, the antenna and rectifying circuit may be fabricated using normal printed circuit board methods on either rigid or flexible substrate. Alternatively, the devices systems may be formed by machining or sintering.

Although the present invention has been described with respect to certain specific embodiments, it will be clear to those skilled in the art that the inventive features of the present invention are applicable to other embodiments as well, all of which are intended to fall within the scope of the present invention.

The invention claimed is:

1. A radio frequency (RF) energy harvesting device comprising:

a substrate having a substantially planar surface;

a metamaterial resonator antenna including at least one metal element that is conformally fixedly disposed on the surface of the substrate, said at least one metal element including:

a first primary antenna segment extending from a first antenna end point to a first midpoint,

a second primary antenna segment extending from a second antenna end point to a second midpoint,

at least one secondary antenna segment disposed parallel to and spaced from the first and second primary antenna segments by a uniform gap distance, said at least one secondary antenna segment having a first end disposed at a third midpoint and an opposing second end disposed at a fourth midpoint,

a first linking segment connected between the first and third midpoints, and

a second linking segment connected between the second and fourth midpoints,

wherein the first and second primary antenna segments and said at least one secondary antenna segment are configured to resonate at RF frequencies that are within an associated bandwidth range such that a first captured RF signal is generated on the first antenna end point and a second captured RF signal is generated on the second antenna end point, said first capture RF signal being 180° out-of-phase from the second captured RF signal; and

a rectifying circuit coupled to the first and second antenna end points, said rectifying circuit including circuit elements configured to combine positive voltage pulses from said first and second captured RF signals to an output node.

2. The RF energy harvesting device of claim **1**, wherein said at least one elongated metal structure forming said metamaterial resonator antenna comprise a silver ink material.

3. The RF energy harvesting device of claim **1**, wherein said rectifying circuit further comprises a plurality of conductive structures conformally fixedly disposed on the surface of the substrate.

4. The RF energy harvesting device of claim **3**, wherein said rectifying circuit further comprises:

a first diode coupled between the first antenna end point and a first node;

a first capacitor coupled between the second antenna end point and the first node; and

a second diode coupled between the first node and a second node,

wherein the first diode has a lower forward voltage than the second diode, and the first diode has a lower reverse voltage than the second diode, whereby a first intermediate voltage generated at the first node comprises a sum of positive voltage pulses of the first RF signal that are passed through the first diode and the second AC signal passed through the first capacitor, and whereby a second intermediate voltage is generated at the second node in accordance with positive voltage pulses that are passed from the first node through the second diode.

5. The RF energy harvesting device of claim **3**, wherein said rectifying circuit comprises a Greinacher circuit.

6. The RF energy harvesting device of claim **1**, wherein said antenna comprises a two-part dipole antenna including first and second bilaterally symmetrical metal elements:

wherein said first metal element includes said first primary antenna segment, said first linking segment, and a first secondary antenna segment extending from the first linking segment to a first distal end point disposed adjacent to the first antenna end point, said first primary antenna segment and said first secondary antenna segment being concentrically arranged such that a uniform gap spacing separates the first primary antenna segment and said first secondary antenna segment along their entire length; and

wherein said second metal element includes said second primary antenna segment, said second linking segment, and a second secondary antenna segment extending from the second linking segment to a second distal end point disposed adjacent to the second antenna end point, said second primary antenna segment and said second secondary antenna segment being concentrically arranged such that said uniform gap spacing separates the second primary antenna segment and said second secondary antenna segment along their entire length.

7. The RF energy harvesting device of claim **6**, wherein each of said first and second primary antenna segments and said first and second secondary wire segments comprises at least two linear segments connected by a corner portion.

8. The RF energy harvesting device of claim **7**, wherein each of said first and second primary antenna segments and said first and second secondary wires comprises an L-shaped pattern.

9. The RF energy harvesting device of claim **7**, wherein each of said first and second primary antenna segments and said first and second secondary antenna segments comprises three interconnected linear sections forming a C-shaped pattern.

10. The RF energy harvesting device of claim **6**, wherein each of said first and second primary antenna segments and said first and second secondary antenna segments comprises continuously curved wire structures.

11. The RF energy harvesting device of claim **10**, wherein each of said first and second linking segments comprises a linear shaped wire structure.

12. The RF energy harvesting device of claim **10**, wherein each of said first and second linking segments comprises a step-shaped wire structure configured such that a length of said first primary antenna segment is equal to a length of said first secondary antenna segment, and such that a length of said second primary antenna segment is equal to a length of said second secondary antenna segment.

13. The RF energy harvesting device of claim **1**, wherein said antenna comprises a single integral metal element comprising:

said first and second primary antenna segments comprising linear structures that are collinearly arranged and separated by a gap defined between the first and second antenna end points;

said secondary antenna segment disposed parallel to the first and second primary antenna segments;

said first linking segment comprising a first open-circle pattern including a first semi-circular section connected between the first primary antenna segment and a first end of the secondary antenna segment, and first and second curved end sections extending from said first semi-circular section into a space between said first primary antenna segment and the secondary antenna segment, and

said second linking segment comprising a second open-circle pattern including a second semi-circular section connected between the second primary antenna segment and a second end of the secondary antenna segment, and third and fourth curved end sections extending from said second semi-circular section into a space between said second primary antenna segment and the secondary antenna segment.

14. The RF energy harvesting device of claim **13**, wherein said single integral metal element further comprises:

a second secondary antenna segment disposed parallel to the first and second primary antenna segments and to the first secondary antenna segment such that the first secondary antenna segment is between first and second primary antenna segments and the second secondary antenna segment;

a third linking segment comprising a third open-circle pattern including a third semi-circular section connected between the first end of the first secondary antenna segment and a first end of the second secondary antenna segment, and fifth and sixth curved end sections extending from said third semi-circular section into a space between said first and second secondary segments, and

fourth linking segment comprising a fourth open-circle pattern including a fourth semi-circular section connected between the second end of the first secondary antenna segment and a second end of the second secondary antenna segment, and seventh and eight curved end sections extending from said fourth semi-circular section into the space between said first and second secondary segments.

15. A radio frequency (RF) energy harvesting system including a plurality of RF energy harvesting devices, each RF energy harvesting device comprising:

a metamaterial resonator antenna including at least one metal element that is conformally fixedly disposed on the surface of a substrate, said at least one metal element including:

a first primary antenna segment extending from a first antenna end point to a first midpoint,

a second primary antenna segment extending from a second first antenna end point to a second midpoint,

at least one secondary antenna segment disposed parallel to and spaced from the first and second primary antenna segments by a uniform gap distance, said at least one secondary antenna segment having a first end disposed at a third midpoint and an opposing second end disposed at a fourth midpoint,

a first linking segment connected between the first and third midpoints, and

a second linking segment connected between the second and fourth midpoints,

wherein the first and second primary antenna segments and said at least one secondary antenna segment are configured to resonate at RF frequencies that are within an associated bandwidth range such that a first captured RF signal is generated on the first antenna end point and a second captured RF signal is generated on the second antenna end point, said first capture RF signal being 180° out-of-phase from the second captured RF signal; and

a rectifying circuit coupled to the first and second antenna end points, said rectifying circuit including circuit elements configured to pass positive voltage pulses from said first and second captured RF signals to an output node,

wherein the plurality of RF energy harvesting devices are coupled together to generate a combined system output voltage.

16. The RF energy harvesting system of claim **15**, wherein the output control circuit of each of the rectifying circuits comprises:

a first diode coupled between the first antenna end point and a first node;

a first capacitor coupled between the second end point and the first node;

a second diode coupled between the first node and a second node, wherein the first diode has a lower forward voltage than the second diode, and the first diode has a lower reverse voltage than the second diode; and

an output control circuit coupled between the second node and an output node, said output control circuit being configured to convert a second intermediate voltage generated on the second node into a direct current (DC) output voltage, and

wherein the plurality of RF energy harvesting devices are coupled together to generate a combined system output voltage.

17. The RF energy harvesting system of claim **16**, wherein the output nodes of all of the plurality of RF energy harvesting devices are connected to a common system output node.

18. The RF energy harvesting system of claim **16**, wherein the antenna of each of the RF energy harvesting devices comprises a two-part dipole antenna including first and second bilaterally symmetrical metal elements:

wherein said first metal element includes said first primary antenna segment, said first linking segment, and a first secondary antenna segment extending from the first linking segment to a first distal end point disposed adjacent to the first antenna end point, said first primary antenna segment and said first secondary antenna segment being concentrically arranged such that a uniform gap spacing separates the first primary antenna segment and said first secondary antenna segment along their entire length; and

wherein said second metal element includes said second primary antenna segment, said second linking segment, and a second secondary antenna segment extending from the second linking segment to a second distal end point disposed adjacent to the second antenna end point, said second primary antenna segment and said second secondary antenna segment being concentrically arranged such that said uniform gap spacing separates the second primary antenna segment and said second secondary antenna segment along their entire length.

19. The RF energy harvesting system of claim **16**, wherein the plurality of RF energy harvesting devices are connected in series such that the second antenna end point of a first said RF energy harvesting device is connected to the output node of a second said RF energy harvesting device.

20. The RF energy harvesting system of claim **19**, wherein the antenna of each of the RF energy harvesting devices comprises single integral metal element comprising:

said first and second primary antenna segments comprising linear structures that are collinearly arranged and separated by a gap defined between the first and second antenna end points;

said secondary antenna segment disposed parallel to the first and second primary antenna segments;

said first linking segment comprising a first open-circle pattern including a first semi-circular section connected between the first primary antenna segment and a first end of the secondary antenna segment, and first and second curved end sections extending from said first semi-circular section into a space between said first primary antenna segment and the secondary antenna segment, and

said second linking segment comprising a second open-circle pattern including a second semi-circular section connected between the second primary antenna segment and a second end of the secondary antenna segment, and third and fourth curved end sections extending from said second semi-circular section into a space between said second primary antenna segment and the secondary antenna segment.

21. A metamaterial resonator antenna comprising at least one metal element including:

a first primary antenna segment extending from a first antenna end point to a first midpoint,

a second primary antenna segment extending from a second antenna end point to a second midpoint,

at least one secondary antenna segment disposed parallel to and spaced from the first and second primary antenna segments by a uniform gap distance, said at least one secondary antenna segment having a first end disposed at a third midpoint and an opposing second end disposed at a fourth midpoint,

a first linking segment connected between the first and third midpoints, and

a second linking segment connected between the second and fourth midpoints,

wherein the first and second primary antenna segments and said at least one secondary antenna segment are configured to resonate at frequencies that are within an associated bandwidth range such that a first captured signal is generated on the first antenna end point and a second captured signal is generated on the second antenna end point, said first capture signal being 180° out-of-phase from the second captured signal,

wherein said at least one elongated metal element forming said metamaterial resonator antenna comprises a silver ink material.

22. The metamaterial resonator antenna of claim **21**, wherein said at least one metal element comprises a single integral metal element comprising:

said first and second primary antenna segments comprising linear structures that are collinearly arranged and separated by a gap defined between the first and second antenna end points;

said secondary antenna segment disposed parallel to the first and second primary antenna segments;

said first linking segment comprising a first open-circle pattern including a first semi-circular section connected between the first primary antenna segment and a first end of the secondary antenna segment, and first and second curved end sections extending from said first semi-circular section into a space between said first primary antenna segment and the secondary antenna segment, and

said second linking segment comprising a second open-circle pattern including a second semi-circular section connected between the second primary antenna segment and a second end of the secondary antenna segment, and third and fourth curved end sections extending from said second semi-circular section into a space between said second primary antenna segment and the secondary antenna segment.

* * * * *

Crystal Radio

KV3540/AK9823/AK23/K23

Advise:

Make sure there is correct number of components in your package. Use 30W iron, 60% solder. Take extra caution against overheating during soldering process. Transistors and ICs can be easily damaged by high temperature so do not leave the iron on the board for too long. Too many or too little solders may cause a defected circuit. Please weld the IC socket first then plug in the IC chip.

Circuit Principal:

The place we live on earth is filled with electric magnetic waves, which are produced by human. Radio station transmit all the usable information at high frequency, through the antenna we receive these signals, then after some processing it is possible to obtain the original information. AM broadcast is using high frequency carrier with low frequency signal. It transmits the original signal by amplitude modulation of the high frequency carrier wave. To receive the signal from the air, we need an antenna. Antenna for AM channel is not very strict, usually larger area and less obstruction will have better reception. Signal received by the antenna is very weak and covers a wide band, so we need to separate the AM frequency we want and filter out the other frequency bands. LC resonance is a good filtering circuit. We get maximum impedance when the frequency is equal to the formula. Using an inductor and one variable capacitor we can produce different frequencies base on different LC values. We can select different radio channels in this way. However the channel received is still in high frequency signal range and can't be heard by human ear, we need to remove the high frequency carrier wave. Diode conducts in positive direction only so used with RC filter we can demodulate the original signal. With a sensitive earphone you can receive clear AM channel. No electricity is needed! This is the first generation semi-conductor radio invented in World War II.

If living in city area, place the antenna stick out of window, and put the earphone on, rotate the variable capacitor to receive radio channels. Adjust the antenna and it's direction can improve reception, if still not satisfy, get a outdoor antenna. Caution, never connect to ground, equal potential and phase will cancel out the signal received. Formula: $f = \frac{1}{2\pi\sqrt{LC}}$

Circuit Explanation:

Electric wave received by antenna pass through the coupling capacitor C1 to LC circuit (consist of inductor L1 and variable capacitor VC1). Use an adjusting stick to slowly adjust capacitor to the required frequency. (Fig. 1 is AM amplitude modulated wave). The frequency signal selected pass through the diode then through the filtering circuit R2 and C2. Finally the sound signal is send to the earphone through the coupling capacitor C3.

Caution:

1. The diode has polarity, antenna need to be straighten up.
2. If sound signal is very weak, point the antenna to a direction with less obstruction. Make close contact to large metal plate to increase the receiving sensitivity.
3. The earphone must be tight with ear without any gap.

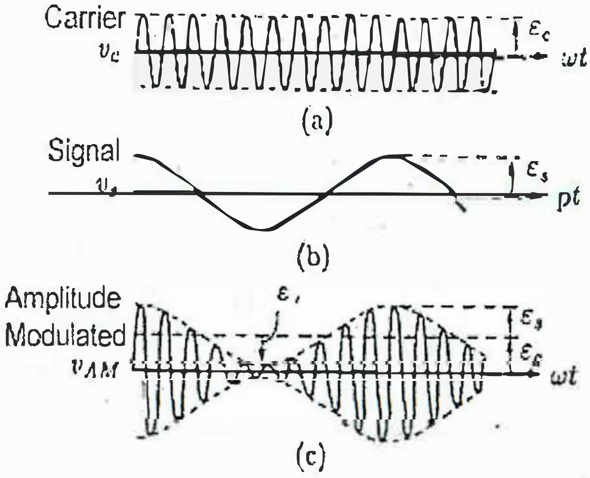

Fig. 1

Parts list:

Serial No.	Component	Qty.	Remark
R1	5.6MΩ	1	GnBlGnGo
R2	2MΩ	1	ReBkGnGo.
C1	332pF	1	3300pF
C2	33pF	1	
C3	10μF	1	
D1	Diode	1	
VC1	Variable capacitor	1	
L1	Inductor (AM)	1	
Earphone	Earphone	1	
Antenna		1	
K23	PCB	1	

Code: Bk:Black, Re:Red, Gn:Green, Bl:Blue, Go:Gold.

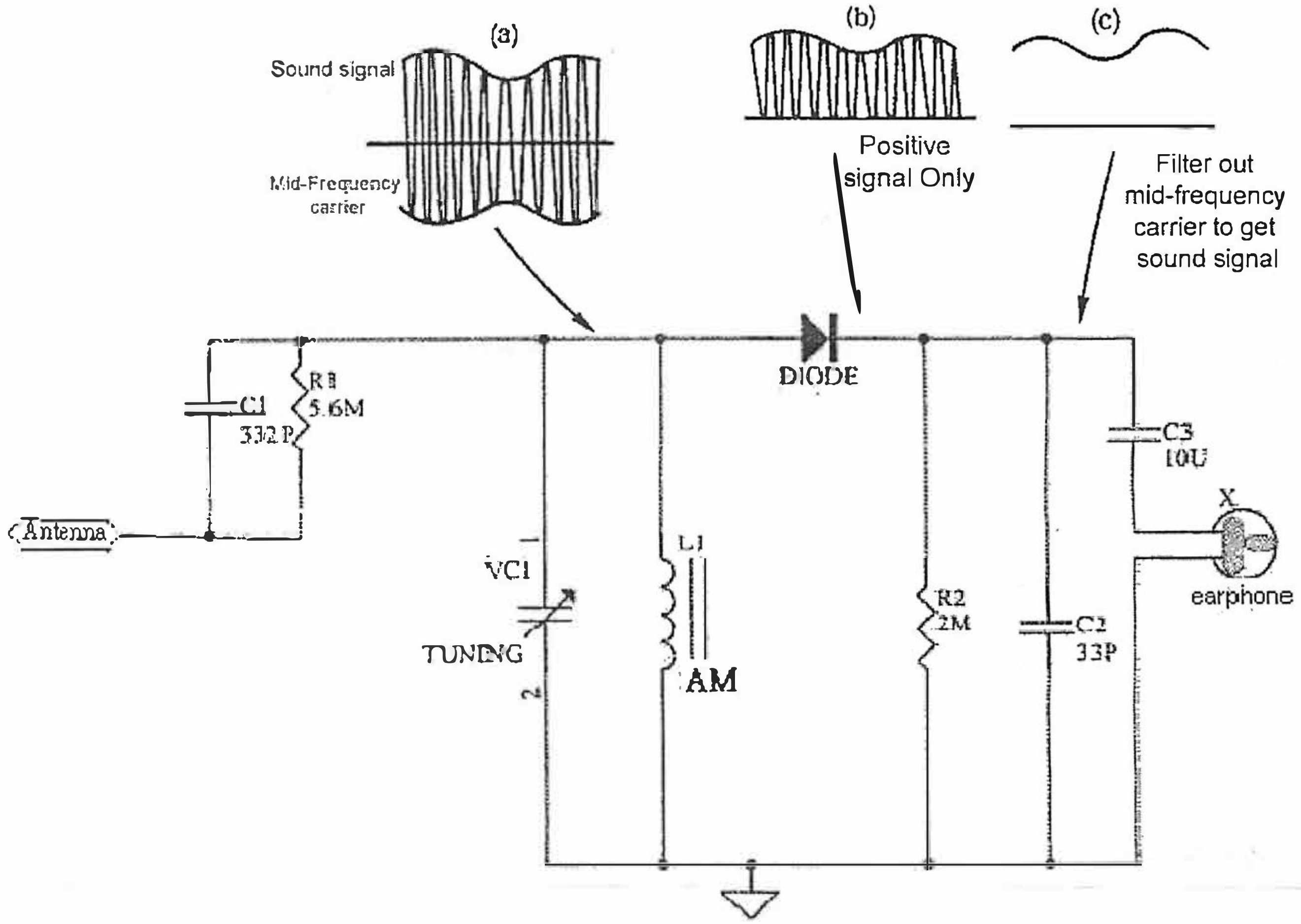
(a)
Sound signal
Mid-Frequency carrier
(b)
(c)
Positive signal Only
Filter out mid-frequency carrier to get sound signal
C1 332P
R1 5.6M
Antenna
VC1
TUNING
L1
AM
DIODE
R2 2M
C2 33P
C3 10U
X
earphone

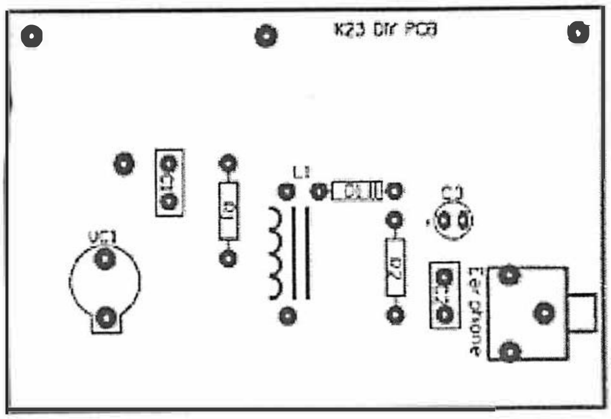
K23 DIY PCB
VC1
C1
R1
L1
D1
C2
R2
C3
Earphone

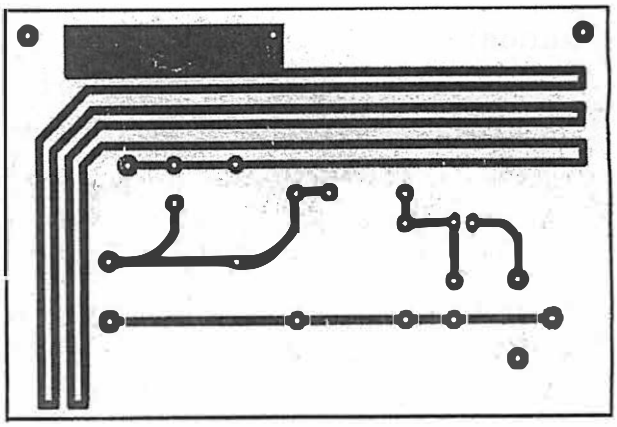

The More Standard Parts Crystal Set - V2

Update September 2011; If you're looking for a kit of modern parts that will reproduce the design below quite closely I recommend the Useful Components Choccy Block Crystal Set Radio Kit.

In another page on this site I describe how to make a pretty good loud crystal set which I'm calling the version 1. This uses fairly high impedance, fairly sensitive magnetic headphones, an OA81 diode, a transistor radio ferrite rod and some other parts that were a bit awkward to find.

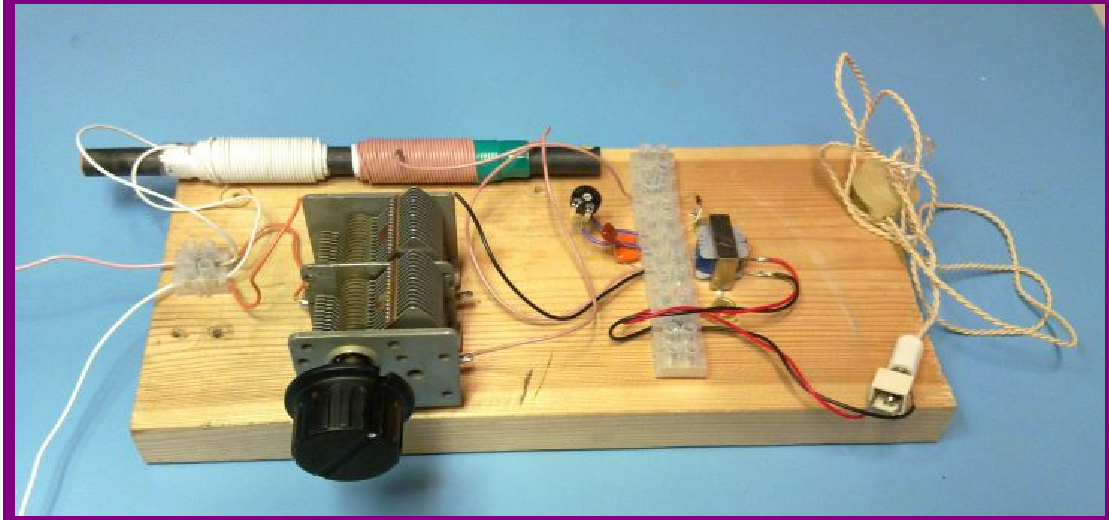

Parts That I Used

1 Standard 120mm ferrite rod with pre-wound MW/LW coils
1 aerial coupling coil that you make (or, see text)
1 output coil that you make (or, see text)
1 1N60 or BAT48 low Vf diode
1 AM variable capacitor from an old transistor radio
1 standard ceramic high impedance earpiece
1 1nF capacitor
1 100nF capacitor
1 100k Ohm variable resistor
1 LT44 interstage audio transformer.
1 20 metres or so of aerial wire
1 2 metres earth connection wire
1 choccy block terminal block connector
1 lump of offcut pine

Some more background

As a kid I played with crystal radios using the little brown crystal earpieces and various circuit configurations with some limited and variable success, but always with quite poor aerials. I couldn't reproduce this limited success using my V1 set and a wire aerial sloping down from an upstairs room to the end of the garden, a mere few metres. This suggested to me that I might be able to get something even louder going using those crystal earpieces, even if they are a bit uncomfortable. I'm calling them crystal earpieces, and so they used to be. When I took one apart in the 1970s they consisted of a light foil diaphragm with two stiff wires leading from the wire terminals to what I assume was a piezo crystal lump, bonded to the middle of the diaphragm with glue at the apex. The new versions contain something which is more like a standard piezo sounder element so might be more accurately described as ceramic earpieces. I suspect that they will have more capacitance than the old ones, which I can't confirm without finding a vintage part. They do seem to work just as well as the old ones though, and are probably more consistant.

Here's the circuit:

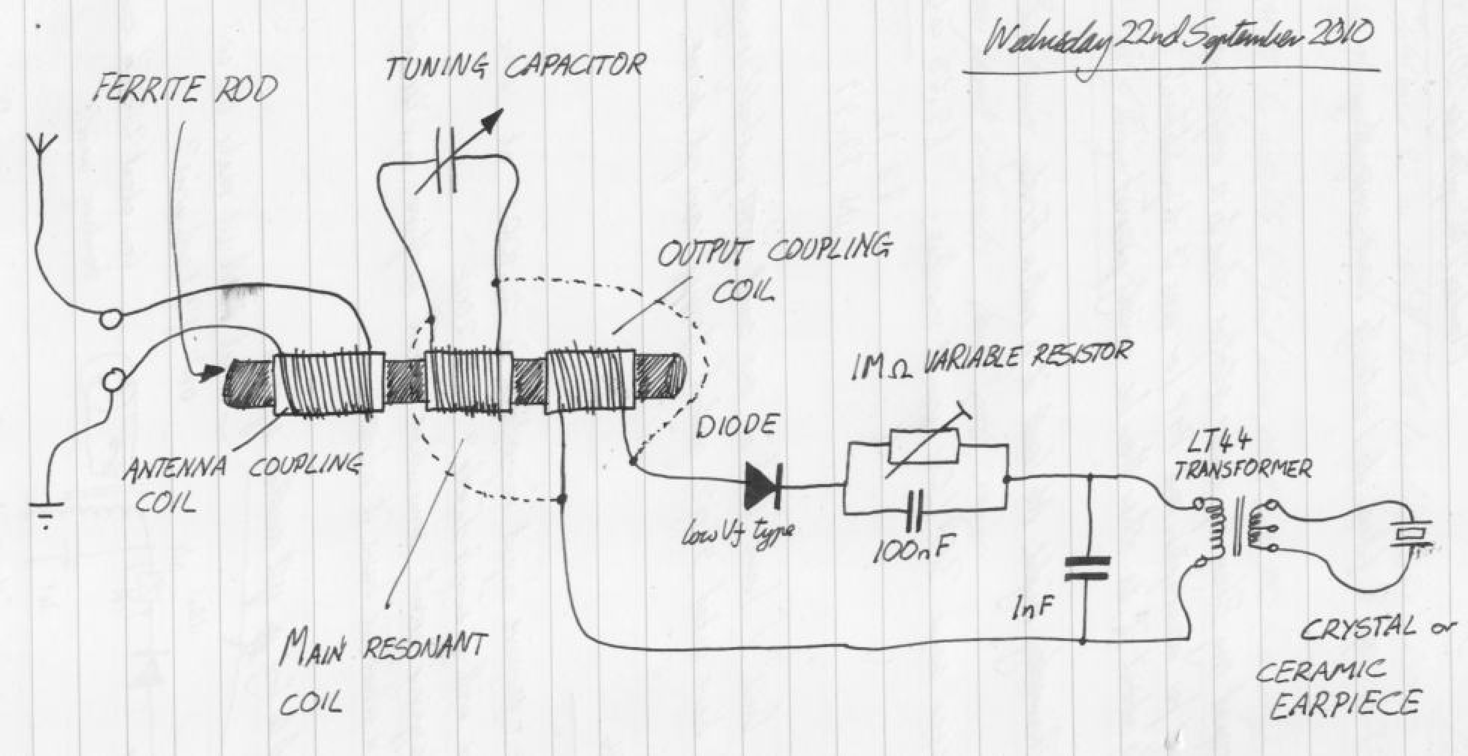

Like the V1 circuit this consists of three coils fitted onto the ferrite rod in such a way that they can slide up and down on it. The reasons for this and plenty more discussion besides are dealt with on the V1 page, so let's deal with the differences here. First, we're using a ceramic earpiece because they are very sensitive and you can buy them easily. I measured a few of these and they are about 27nF capacitance with very high resistance at d.c. At 1kHz, the audio mid-band this is about 6K Ohms impedance. That's quite high, though not high enough to go connecting it directly across the tuning tank. Normally you'd need a resistor connected across it to bleed away the d.c. offset generated by the diode and that makes it a bit less efficient. Here I tried using an LT44 audio transformer instead. This provides a d.c. path to ground for the diode, multiplies the impedance of the earpiece up still further and by coincidence, provides just about the right inductance to resonate with the capacitance of the earpiece at 1kHz. This is a fortunate coincidence indeed, if you're looking for loudness and aren't too bothered about the audio being a bit tinny sounding. When measuring high impedance side of the transformer at 1kHz it actually increases when you connect the earpiece load up to the low side.

Across the high impedance side of the transformer there is a 1nF capacitor. I wasn't expecting to need this, but at low signal levels it increases the volume significantly. A bit of a mystery that. Feeding into the top end of the transformer there is another difference to the V1, that being the inclusion of a d.c. resistor shorted across with a capacitor. These two components together are colloquially known as a "Benny" which I believe is after their inventor and nothing to do with Crossroads. The idea is that the diode works better when the d.c. current is equal to the a.c. current, so you adjust the variable resistor for maximum volume after tuning in your station. It makes no difference at low signal levels, but when you get a decent aerial erected it really does work. I had been very puzzled by this effect when I'd come across it when playing with the V1 set.

The diode is conventional and you can insist on using a germanium point cantact type if you wish, but I've found that silicon low Vf BAT48 schottkys and 1N60s are just as good, even at low signal levels.

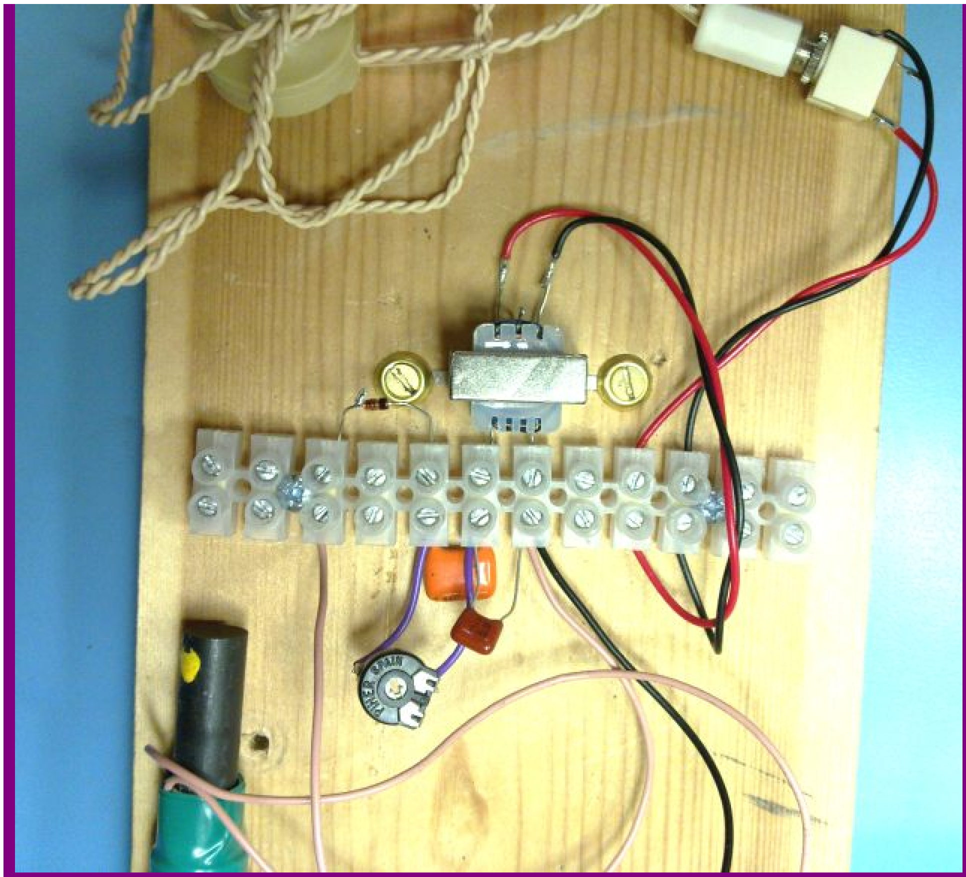

The picture shows small parts on the circuit mounted in a handy 2A terminal strip "choccy block" block. You can guarantee that the insulation between the terminals on these are low loss, which is not necessarily the case if you are trapping the wires under screwcups screwed into variable wood with unknown properties.

Why bother with the audio transformer?

Why bother with the audio transformer when the earpiece is 6KOhms at the frequency of interest and moving your output coupling coil up and down can get the matching correct as in the V1 set? The reasons are that we've changed to a crystal earpiece, so we need that efficient d.c. path to ground instead of a bleed resistor. It also reduces losses due to diode Vf by keeping the signal voltage in the demodulator higher overall. Also, the output coil position should be less critical as we've brought the demodulator stage up in impedance, closer to what would be required for connecting directly across the tank. I've actually used this quite successfully at low signal levels connecting directly across the main tuning coil, but the tuning does become a bit wide. A good compromise might well be to connect the diode about halfway up the main tuning coil. Removing the transformer and using a bleed resistor gives better audio quality particularly at the lower end.

For an aerial coupling coil I have used another LW coil when operating on LW for main tuning and another MW coil when using MW for main tuning. I think there is an element of double resonance going on there with the position of the aerial coil affecting the tuning a little and vice-versa. If you don't want to do that I suggest winding plain insulated wire on a cardboard former until you have enough turns. You have enough when putting more turns on doesn't make it any louder and begins to make it quieter when the aerial and main coils are close together. A similar rule applies for the output coil.

These are all preliminary results based on quite low signal levels, where my Benny has no real effect and maximum output coupling is required to get some fairly meagre sound. I'll be trying to improve the aerial situation at home after which I will be able to be more specific about some issues. For the moment, I can say that this V2 set with the crystal earpiece works better with a small aerial than the V1 so I'd hope that it is also louder in better operating conditions.

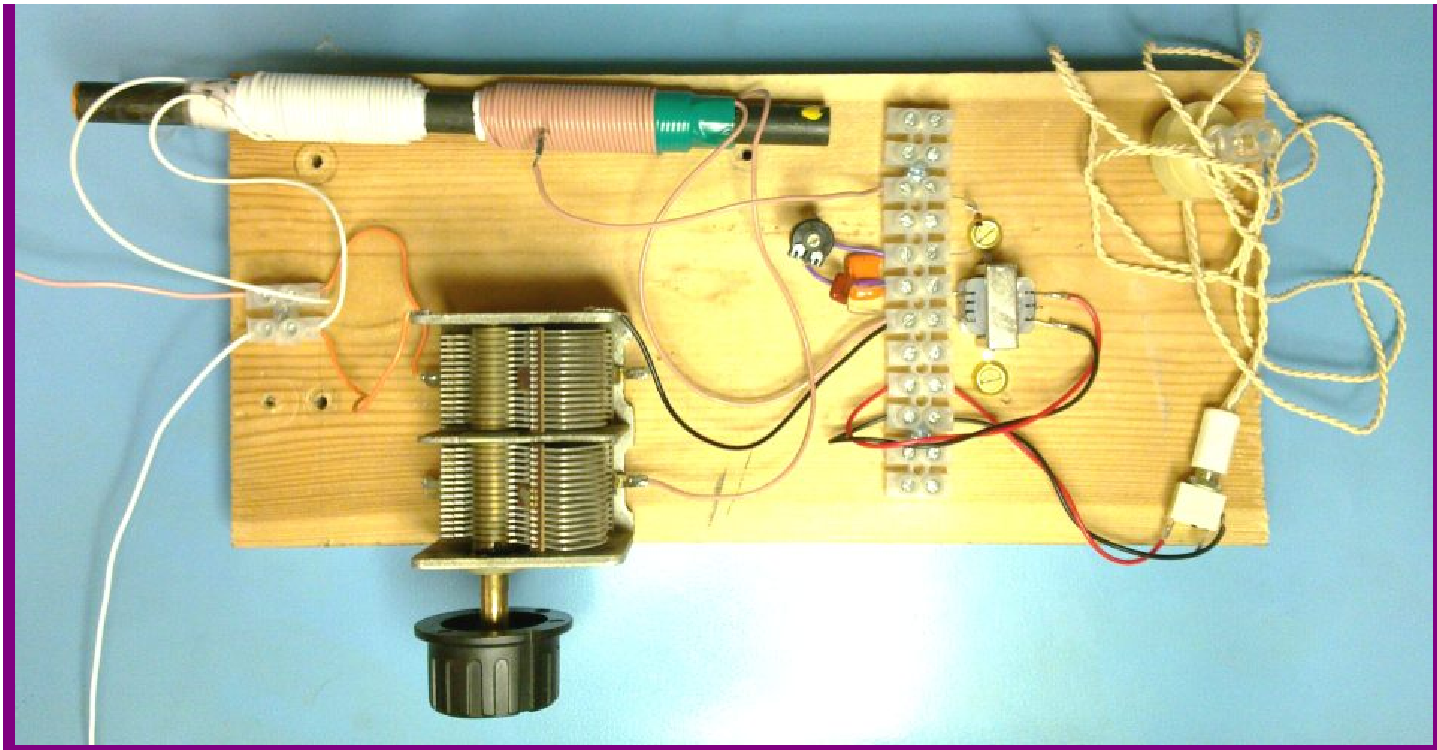

In the picture at the top of the page and the one directly above, you can see that I ended up taking the output coupling off as a tap from the main tuning coil rather than having a separate coil, for this set with the output transformer and ceramic earpiece. In the end it didn't make much difference and the final arrangement with the tap was just more convenient.

Navigate Up

BIRMINGHAM, ALABAMA

CRYSTAL RADIO GROUP

crystalradio.us

- Home
- E-Mail
- B-Ham Sets
    - 80M Receiver
    - Brian's
    - Dee's Radio
    - Gil's Superhet
    - 80M Homebrew
    - Hop's Radio
    - Jack's 1AD
    - BC-221 Regen
    - JB's XtalSet
    - Lem's Regen
    - Lem's Rook Set
    - Jack's Bandpass
    - MosFet Set
    - Progressive
    - Dave's Regen
    - Steve's 221
    - Tom's 1 Tube
- Crystal
    - 2004-1
    - 2004-2
    - 2004-3
    - 2005-1
    - 2005-2
    - 2006-1
    - 2006-2
    - 2006-3
    - 2007-1
    - 2007-2
    - 2008-1
    - 2008-2
    - 2008-3
    - 2008-4
    - 2009-1
    - 2009-2
    - 2009-3
    - 2009-4
    - 2009-5
    - 2009-6
    - 2010-1
    - 2010-2
    - 2010-3
    - 2010-4
- 1AD Radios
    - 2005-1
    - 2006-1
    - 2006-2
    - 2007-1
    - 2007-2
    - 2008-1
    - 2009-1
    - 2009-2
    - 2009-3
- Contest Results

- 2004
- 2005
- 2006
- 2007
- 2008
- 2009
- 2010
- 2006 Sprint
- 2008 Sprint
- 2009 Sprint

The Progressive Crystal Set

This set is called a "progressive" set since you start with the basic unit and can get progressively more sophisticated by adding a trap and antenna matching network.

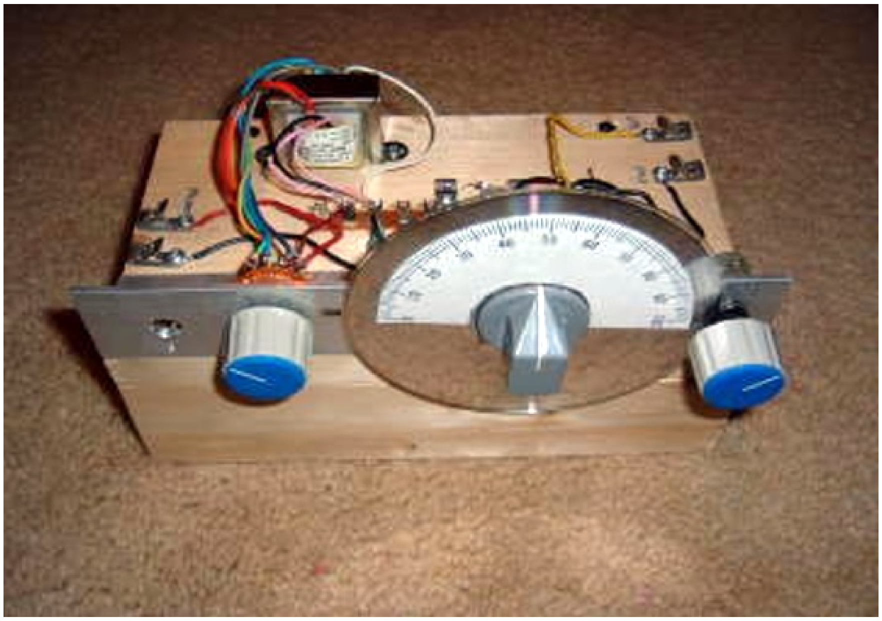

The prototype version uses a homemade vernier made from a CD and rubber grommet. The tapped transformer will match many different headset or speaker impedances. of plexiglass and poly-coated wood.

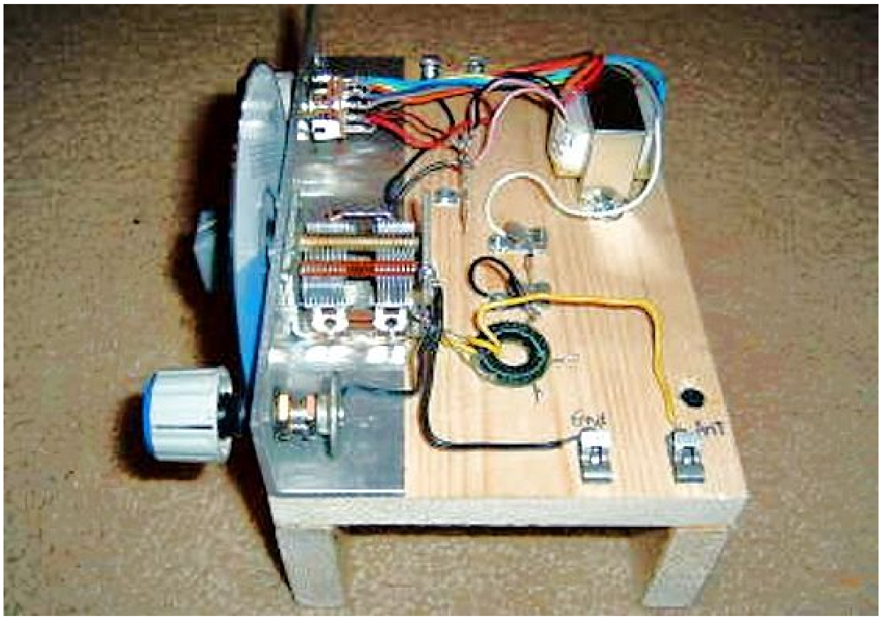

The tuning knob shaft is in a panel mount piece made for a 1/4" shaft. The threaded part of a an old potentiometer could be used, along with its 1/4 inch shaft.

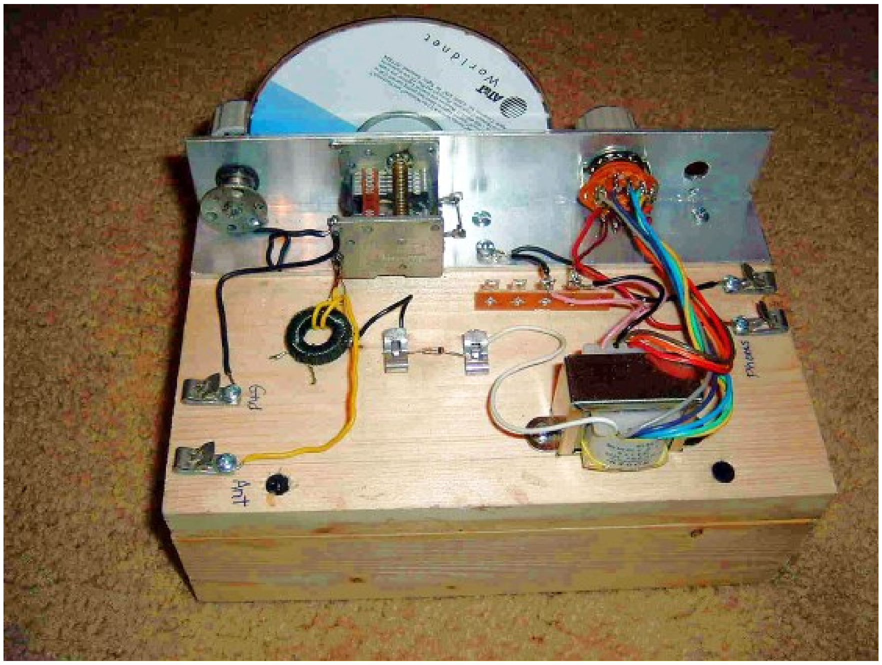
Gnd
Ant
Phones

Here is a rear view of the set. Notice the Bogen model T725 transformer.

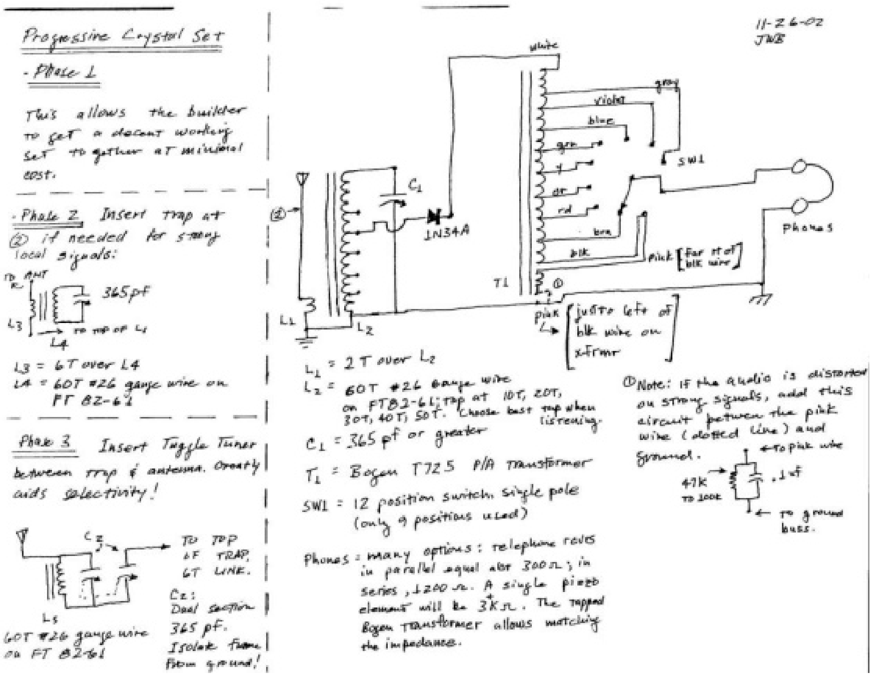
11-26-02
JWB
white
gray
violet
blue
grn
y
or
rd
brn
blk
SW1
C1
1N34A
②
T1
①
L1
L2
Phones
pink [far rt of blk wire]
pink [just to left of blk wire on xfrmr]
TO ANT
365pf
L3
L4
TO TOP OF L1
C2
TO TOP OF TRAP, 6T LINE.
L5
47K TO 100K
.1uf
← to pink wire
← to ground buss.

Progressive Crystal Set

- Phase 1

This allows the builder to get a decent working set together at minimal cost.

- Phase 2 Insert trap at ② if needed for strong local signals:

L3 = 6T over L4
L4 = 60T #26 gauge wire on FT 82-61

Phase 3 Insert Toggle Tuner between trap & antenna. Greatly aids selectivity!

60T #26 gauge wire on FT 82-61

C2: Dual section 365 pF. Isolate frame from ground!

L1 = 2T over L2
L2 = 60T #26 gauge wire on FT82-61; tap at 10T, 20T, 30T, 40T, 50T. Choose best tap when listening.
C1 = 365 pf or greater
T1 = Bogen T725 P/A transformer
SW1 = 12 position switch single pole (only 9 positions used)
Phones = many options: telephone recvs in parallel equal abt 300Ω; in series, 1200 Ω. A single piezo element will be 3K+Ω. The tapped Bogen transformer allows matching the impedance.

① Note: If the audio is distorted on strong signals, add this circuit between the pink wire (dotted line) and ground.

Schematic of set.
[Click for large view.]

Back to the [Birmingham Group Radios].

RF Spectrum Analyzer - Made in Germany / up to 20GHz
Handheld, USB, Remote & 19" Analyzers LF & RF / 1Hz to 20GHz / -170dBm DANL aaronia.com/rf-analyzers

Select Menu

- [Home]

- About Us
- Contest Results
- 1AD Contests
- 1AD Entries
- Antenna Matching
- Antennas
- BHAM Radios
- Crystal Contests
- Crystal Radio Entries
- Sprint Contest
- Headphones
- Projects, Mods
- Favorite Links
- Contact Jack

Radios

- B-Ham Index
- Brian Wingard
- Dee Haynes
- Hop Hays
- Lem's Set
- Jack Bryant
- Mosfet
- Progressive
- BC-221 crystal set
- Jack's 1AD superhet
- JB McCreath crystal set
- Gil's 1AD superhet
- BC-221 regen
- Lem's regen
- Dave's regens
- SW One Tuber
- 80m HB Receiver
- 80m HB Station

RF Spectrum Analyzer - Made in Germany / up to 20GHz

Handheld, USB, Remote & 19" Analyzers LF & RF / 1Hz to 20GHz / -170dBm DANL

aaronia.com

SureLock
Digital TV Signal Finder

SOLID SIGNAL

© Copyright 2000 – 2012, Jack Bryant, [crystalradio.us](). All rights reserved.

Website by [Dave Schmarder]()

rimstar.org
... an addiction to science, renewable energy and building stuff

How to make a piezoelectric crystal speaker

Related links:
[[Make Rochelle crystal](#)], [[Crystal radios](#)]

- Home
- Renewable Energy
- Science/Electronics
  - Projects
    - Build Van de Graaff
    - Wimshurst machine
    - Kelvin water dropper
    - Electroscope
    - Corona motors
    - Ball bearing motor
    - Crystal radios
    - AM Radio Transmitter
    - Fresnel lens
    - Electrets
    - Piezoelectric
      - Make Rochelle crystal
      - Make bigger crystal
      - Test Rochelle crystal
      - Piezo from BBQ lighter
      - Piezo from lighter
      - Piezo disk from card
      - Piezo power from rain
      - **Piezoelectric speaker**
    - Electrolytic capacitor
    - Joule thief
    - Peltier/Seebeck effect
    - Franklins bell
    - Can Stirling
    - Big Stirling
    - 555 timer music
    - TEA laser
    - Laser diode fun
    - How Ion Prop. Works
    - Lifter/ionocraft
    - Star Trek Enterprise
    - Smoke precipitator
    - Moving with sound
    - Straw music instruments
    - Gravity light - DIY
    - Nitinol wire
    - Pyramid hologram
    - Arduino speech/Talkie
    - Arduino controlled Skull
    - Pinhole camera
    - Wireless electricity
    - Fly swatter powered CFL
    - Chlorophyll/fluorescence
    - Small SG Tesla coil
    - Photophone
    - Laser communicator
    - Hero's steam engine
    - Alcohol stove
    - Sunflower heat engine
    - BB-8 Droid
    - Ball cyclotron
    - Atmos. electricity
    - Photoelectric effect
    - Bottle rocket
    - Neural networks
    - Ornithopter
    - Curie temp. experiment
    - Lightning in bottle
    - Gyroscope - vinyl records
    - Raspberry Pi
    - Pepper's ghost phone
    - Potato chip circuit
    - Ion wind rotor
  - Resources
  - Power supplies
  - HV Probes/measuring
  - Miscellaneous
- Efficiency & Conservation
- Composting
- Food
- Non-conv. Energy
- Non-conv. Propulsion
- Space
- Stories

Donate

Plating Rectifier
Udylite by Process Electronics Rectifier Sales, Parts & Service pecrectifier.com

When a voltage is applied across a piezoelectric crystal, the crystal expands and then contracts again. By connecting the crystal to the sound output of a device where a normal speaker would usually go, a radio's output for example (see diagram below), then the device will supply a varying voltage and the crystal's expansion and contraction will produce sounds waves. You now have a speaker!

How the piezoelectric speaker works.

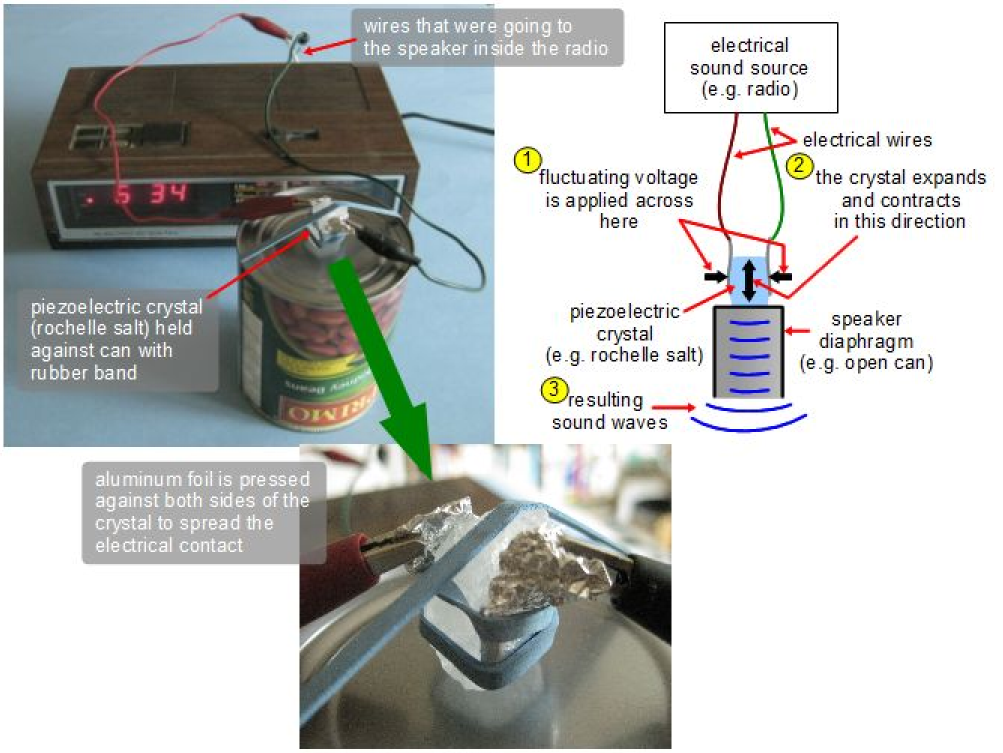
wires that were going to the speaker inside the radio
electrical sound source (e.g. radio)
electrical wires
① fluctuating voltage is applied across here
② the crystal expands and contracts in this direction
piezoelectric crystal (rochelle salt) held against can with rubber band
piezoelectric crystal (e.g. rochelle salt)
speaker diaphragm (e.g. open can)
③ resulting sound waves
aluminum foil is pressed against both sides of the crystal to spread the electrical contact

Stepping up the voltage to the piezo crystal with a transformer

If the voltage coming from the source device isn't high enough then you can step it up with a transformer. Unless you design a transformer specifically for radio frequencies then this won't be ideal, but it does work.

Stepping up the voltage with a doorbell transformer

I first tried a doorbell transformer (see photos below), ones normally used to convert 120V to 12V (for example) for powering a doorbell from household power. Simply wire it backwards, the 12V side to the device's output and the 120V side to the piezoelectric crystal. In my case, my transformer had two 6V options so I used one of those instead.

Of course the signal from the radio isn't 6V and the output from the transformer isn't going to be 120V. I'm just using the transformer for it's ability to step up voltages, in this case the radio's output voltage.

Doorbell transformer.

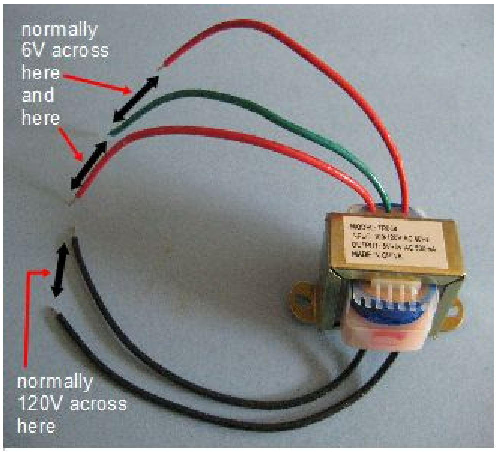
normally 6V across here and here
normally 120V across here

AdChoices
DIY Speaker
Speaker Repair
Plating Rectifier
Udylite by Process Electronics Rectifier Sales, Parts & Service
pecrectifier.com

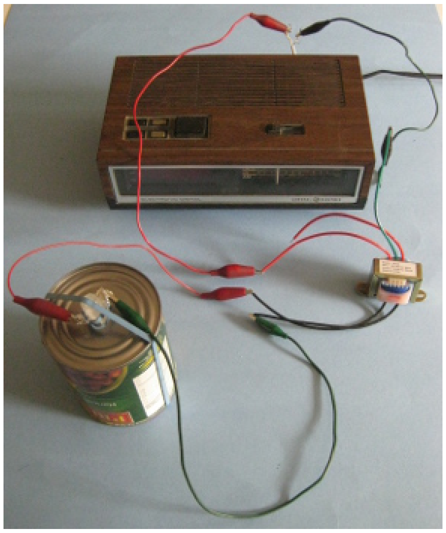

Using the doorbell transformer made a significant improvement in the speaker's volume over no transformer at all.

Doorbell transformers are available wherever household electrical supplies are sold.

Stepping up the voltage with a microwave oven transformer

If the sound from the crystal is still too quiet then you can try increasing it with a microwave oven transformer instead (see photos and diagrams below.) This gave me a big improvement. These transformers can easily be scavanged from old microwave ovens.

The tricky part was that the output coil has only one wire lead coming off it. I'm assuming that the other end of the coil is connected to the transformer case. So for the other output I connected one of the inputs from the radio to the case and took the other output from the case, treating that side as if it were ground, even though the radio has no ground.

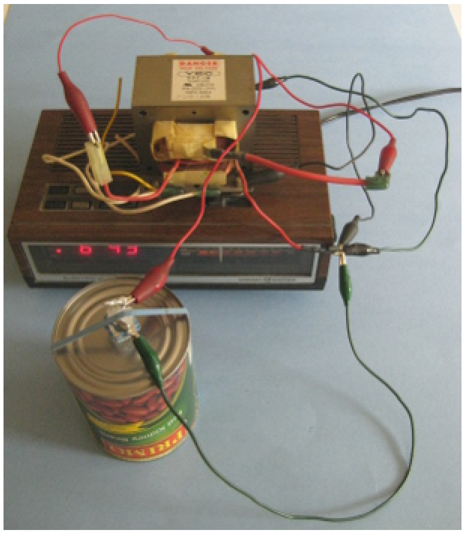

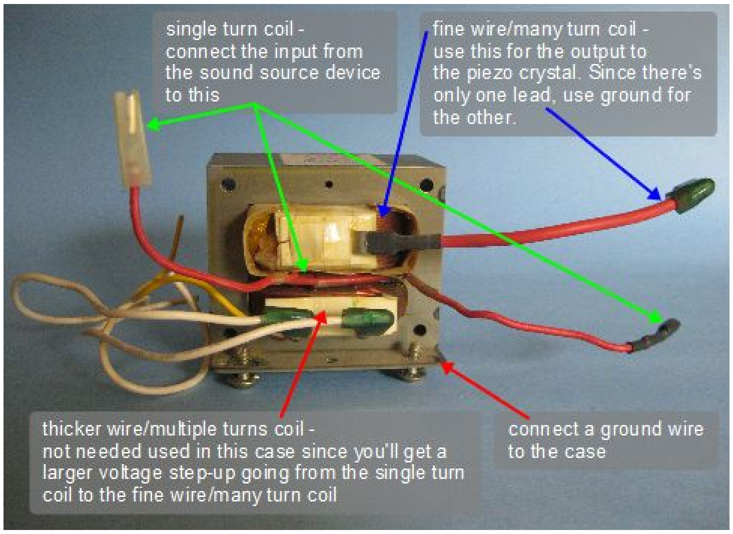

single turn coil - connect the input from the sound source device to this

fine wire/many turn coil - use this for the output to the piezo crystal. Since there's only one lead, use ground for the other.

thicker wire/multiple turns coil - not needed used in this case since you'll get a larger voltage step-up going from the single turn coil to the fine wire/many turn coil

connect a ground wire to the case

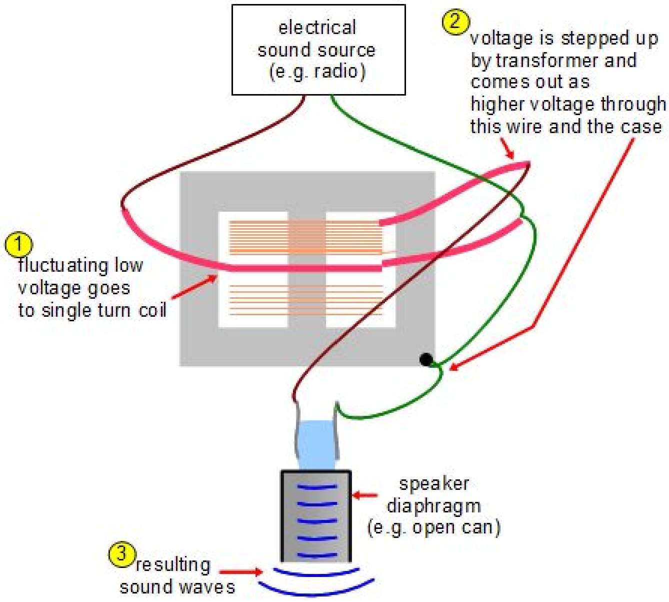

electrical sound source (e.g. radio)
① fluctuating low voltage goes to single turn coil
② voltage is stepped up by transformer and comes out as higher voltage through this wire and the case
speaker diaphragm (e.g. open can)
③ resulting sound waves

Where to find it in a microwave oven.

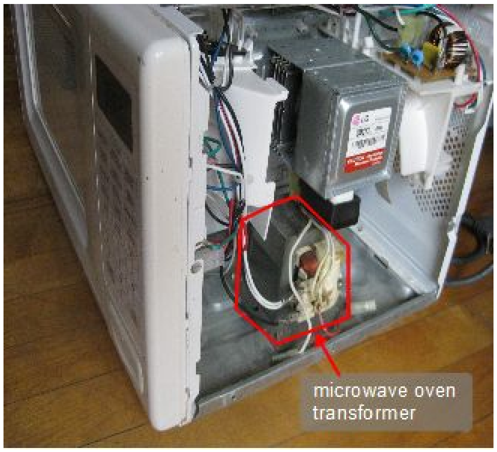

microwave oven transformer

Using the microwave oven transformer made a significant improvement in volume over the doorbell transformer.

WARNING: Be very careful taking apart a microwave oven. There is a high voltage capacitor in it that may still be charged. Do some homework first.

Paper cone piezo speakers with various diaphragm materials

I also experimented with making speakers using paper cones, with the aim of making a good earpiece that would channel the sound into your ear. I tried with an aluminum foil diaphragm but it was to hard to keep it stiff. A paper diaphragm stayed stiff but an aluminum flashing diaphragm gave louder sound. I found that the sound from these was much clearer than the above speaker made from a can. The volume was less though.

Three paper cones for speakers.

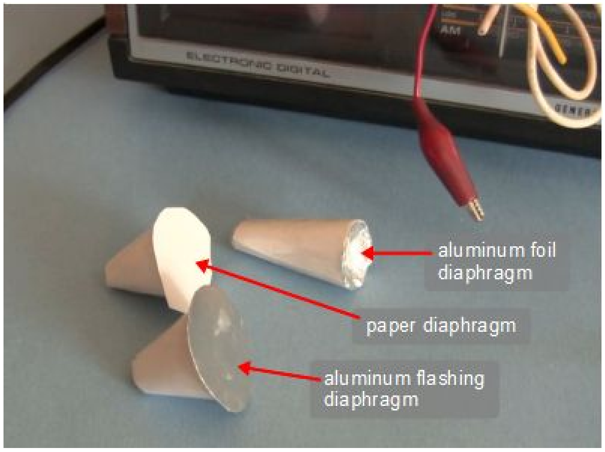

Paper cone speaker with aluminum flashing diaphragm.

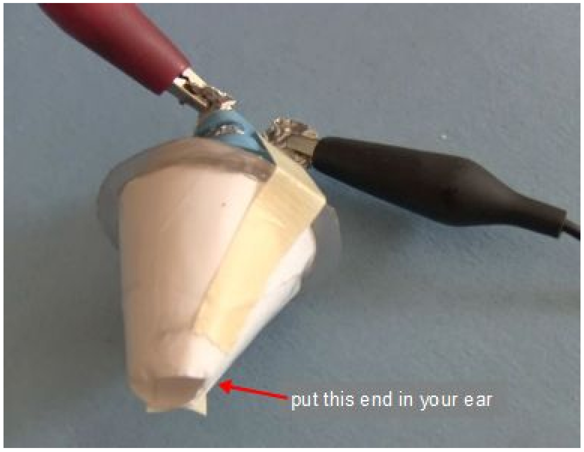

Videos for piezoelectric speakers

This video shows my experimenting starting with no transformer, then the doorbell transformer and finally, the microwave oven transformer. It also shows various piezoelectric crystals I tried, as well as some different speaker configurations.

This video is of me showing that a small rochelle salt crystal works too and how a window can be used as the speaker diaphragm with loud results.

voltage adapter regulator
dual voltage automatic regulators converters, inverters, transformers
welectronics.com

Liked this? Share it with: Digg del.icio.us StumbleUpon Google

Related links:

Donate

[Make Rochelle crystal], [Crystal radios]

rimstar.org - Share your project on rimstar.org - About - Privacy policy - © 2011 Steven Dufresne -

Contact: stevend@rimstar.org

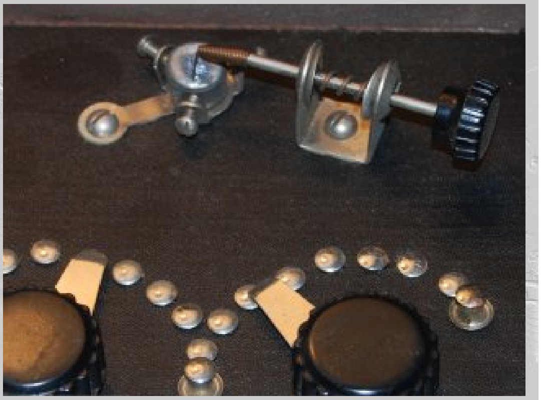

American Radio Stores, Inc.

Back to Radio..

An interesting vintage crystal set recently came up for bidding on Ebay, an American Radio Stores, Inc radio in good working condition. This set interested me because, from the photos provided, it looked to have a fairly classic 1920's circuit with tuning accomplished via two switches to taps on the coil. I presumed that they are units and 10's taps as outlined on the classic Bureau of standards Circuar #121 set. A modest bid brought the radio to my doorstep a couple weeks later.

Looking at the radio when it arrived showed me that this was, even for the time, a fairly inexpensive set lightly built with cheap but workable parts. As expected, the radio has fair sensitivity and abysmal selectivity. (Recall, I bought it for the classic design not its DX potential). Somewhat to my consternation, the front panel was attached to the box by four small brads that could not be removed. No chance to view the inductor in the interior without compromising the vintage nature of the radio. I agonized over the question of whether to tear into the box or preserve the radio. After deliberating some 30 - 40 nanoseconds I was scheming for the most efficient and least destructive procedure to access the interior. My handy Dremmel neatly sliced the small brads one by one leaving no marks on the box or panel.

Inside I expected to find a typical tapped selenoid, the box is certainly large enough for one. What I found suprised and rather shocked me. It still does! The coil is bundle-wound, a mass of 26-30 dcc wire about 2.25 inches diameter. I was unaware that such a design would work. It is impossible to check how the tap wires are attached, they seem to just get wound into the coil with no apparent connection. It appears to be continuous wires from one switch set wound around and around and then exiting to the second switch set. That sounds like eight independent coils wound together. Another odd circuit characteristic is the fact that the tap points between the two tap switches are wired to each other (see photo and schematic). Perhaps connecting in such a manner combines them into a single longer coil, hard to speculate. How this coil works is different from anything I have experienced in crystal radio. Completing the schematic, rectification takes place via an ordinary cat's whisker and small potted galena crystal, whew! This is certainly an interesting design if not otherwise efficient or selective. It is a welcome addition to my small menagerie of vintage sets.

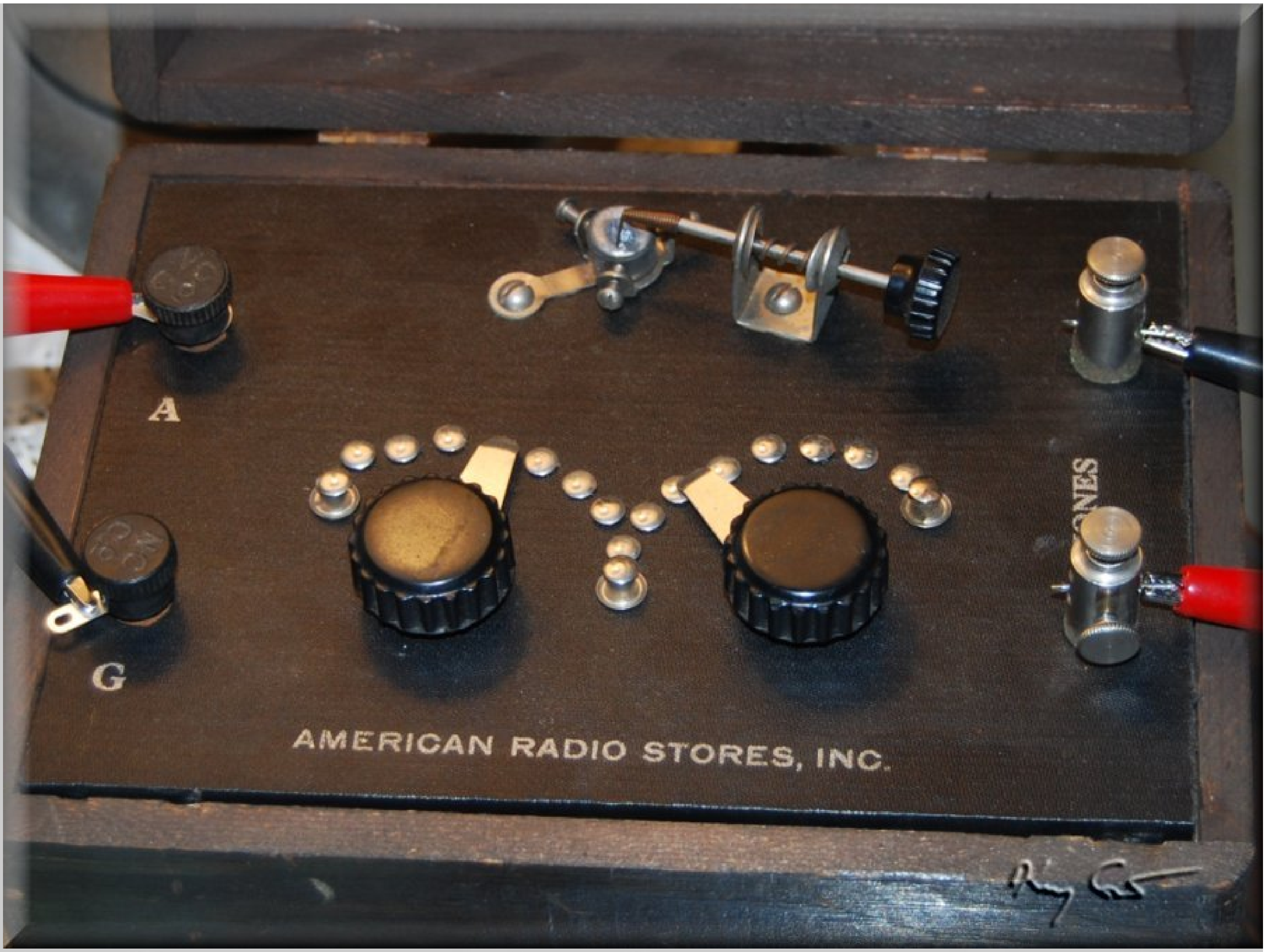
A
G
AMERICAN RADIO STORES, INC.
ONES

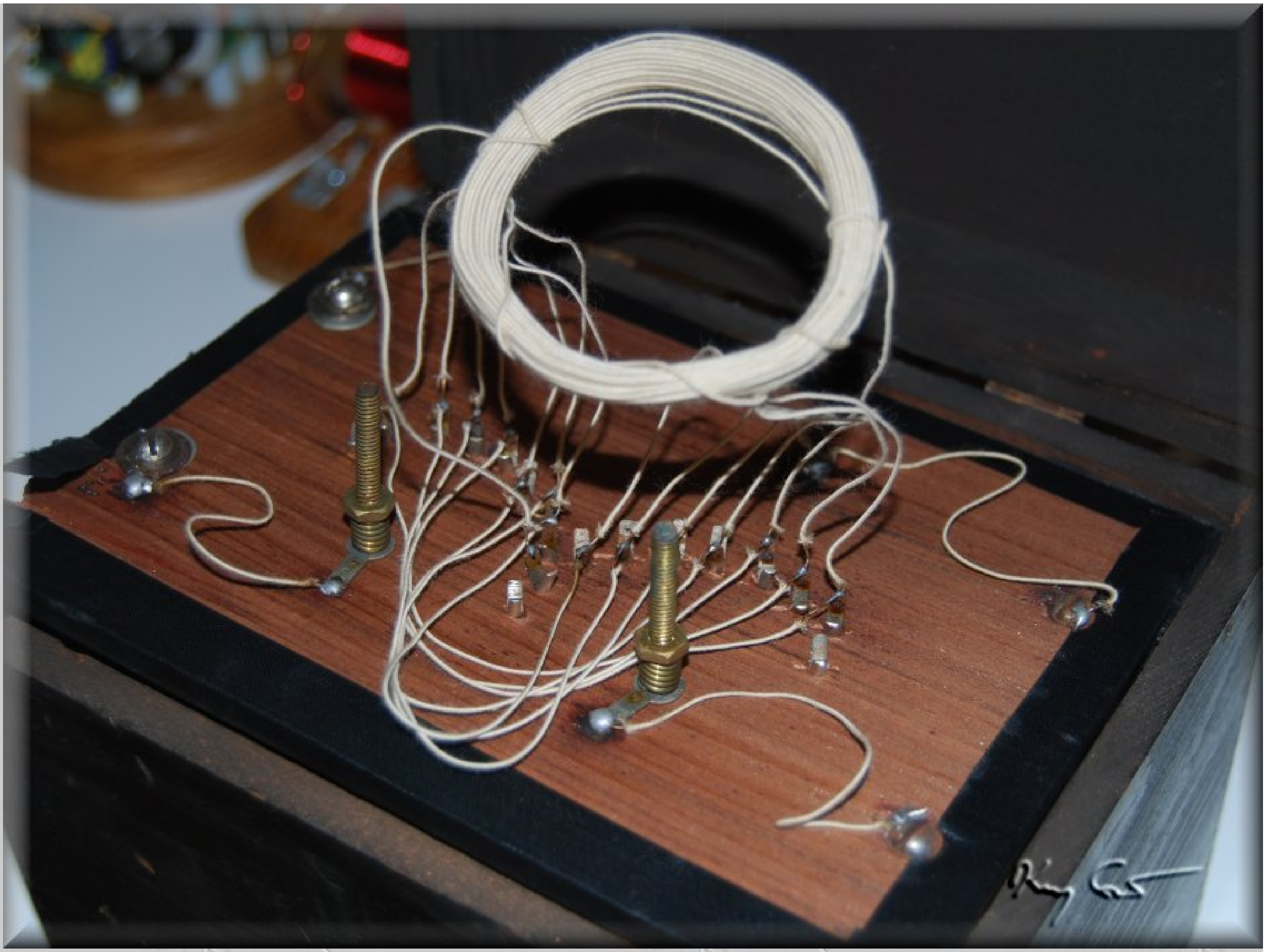

Circuit schematic for the set.

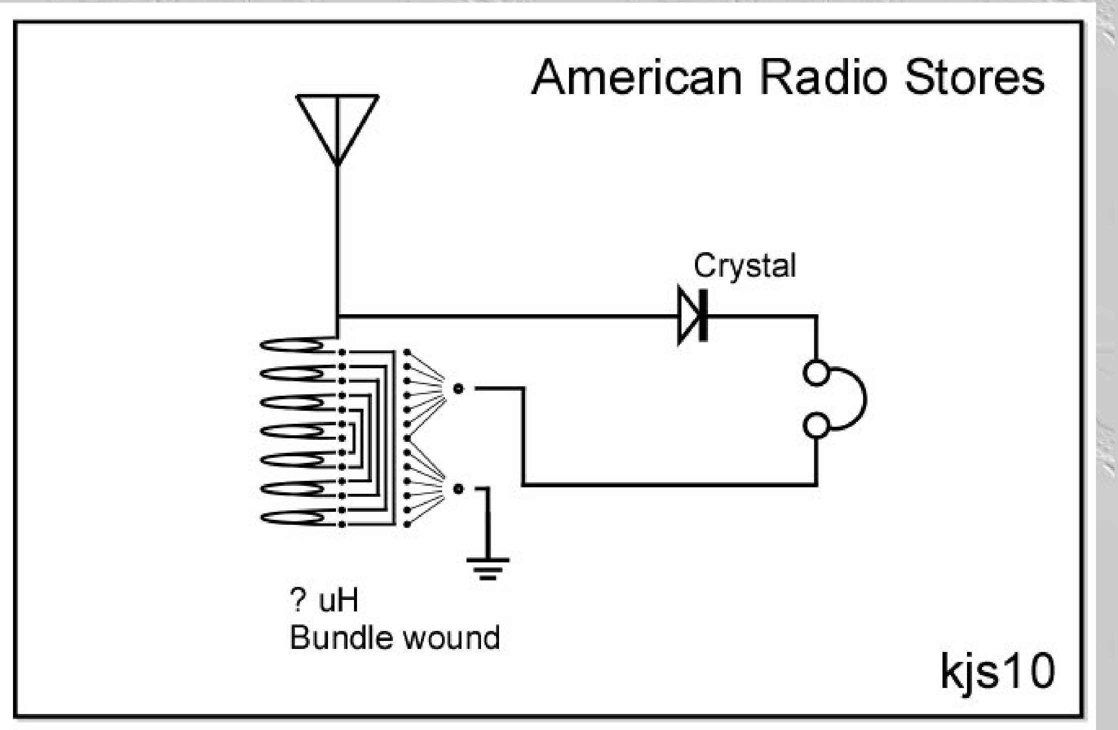
American Radio Stores
Crystal
? uH
Bundle wound
kjs10

Portrait, set and box.

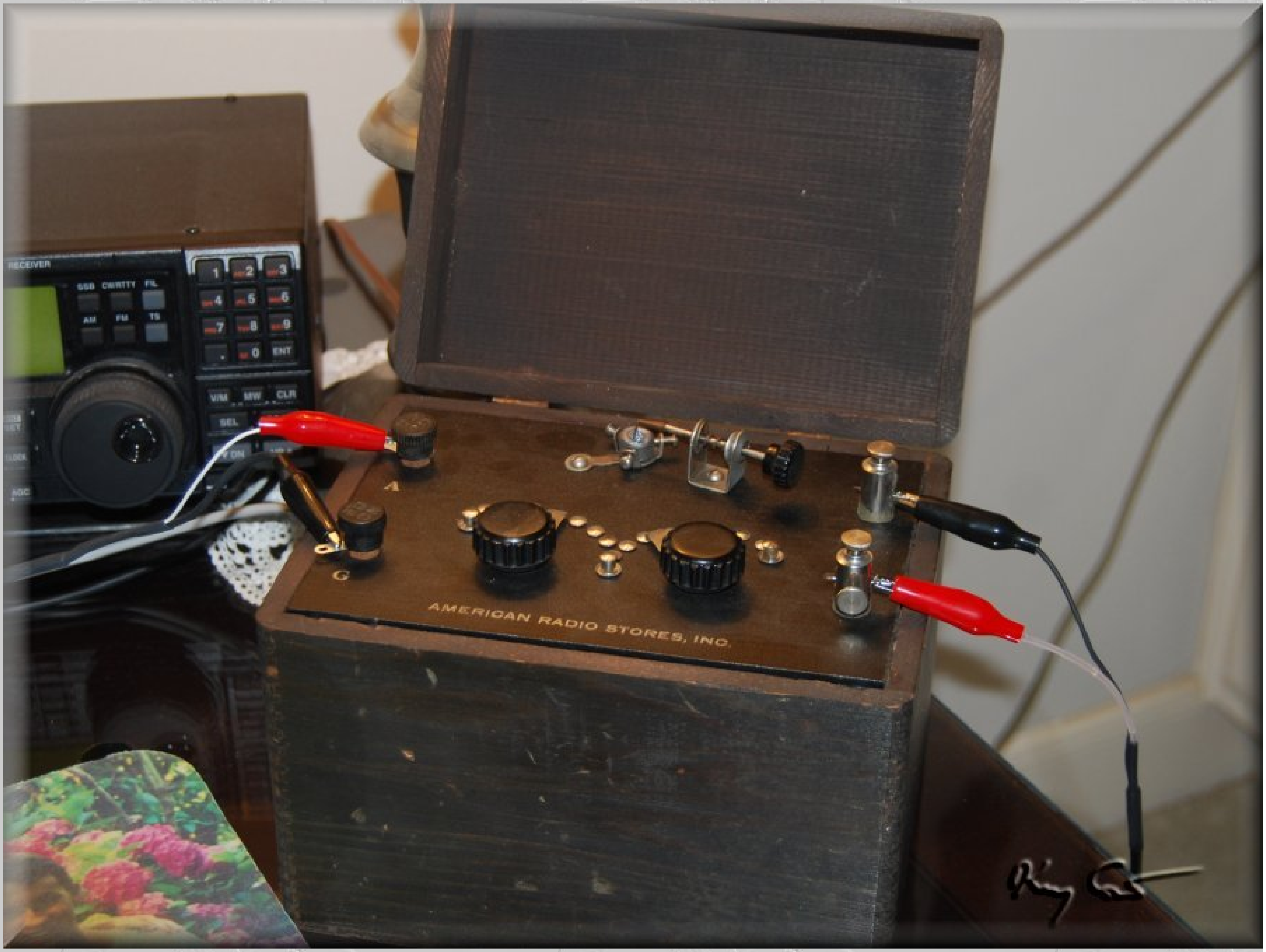

[Return..](#)

[Back to Main....](#)

Friday, 27-Oct-2017 23:44:44 CDT

Comments and/or Correspondance to: kj-smith@sbcglobal.net

.::. Alsace .::. Life List .::. Ionosphere .::. Crystal Radio .::. Snapshots .::. Thoughts .::.
.::. Sitemap .::. Beer Labels .::. International Pix .::. Western US Pix .::. Pinhole Pix .::. Petrology .::.

A Spark-Gap Transmitter

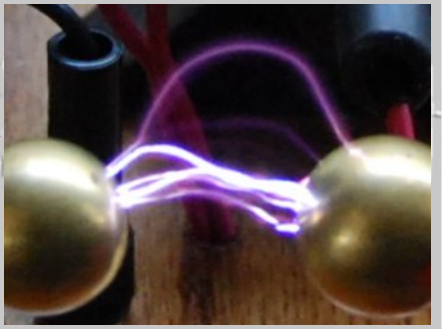

Back to Radio..

Why a spark-gap transmitter? And why here on a crystal radio page? Well, the second question I'll answer first as it is easiest. The transmitter is a vintage technology device that is no longer current in the modern world and so fits thematically with the purpose of my crystal radio page. The first Question is more interesting, why build it? In late 2009 my elder brother, founder and owner of Northwest Research Engineering, LLC., in Seattle, challenged me to construct a spark-gap transmitter. He enjoyed seeing my vintage-style crystal sets and we enjoyed discussing the early history of radio together. So, it seemed a logical step to progress in this direction. Additionally, as his work is in cutting edge ionospheric research, he has little opportunity to study such retro technologies up close and personal.

> DISCLAIMER:
> I am not a licensed radio-amateur and in any case a spark-gap transmitter is not legal to use in the United States. As such, this project was taken on as a technology-demonstration exercise only and is not intended for actual use. Such use is not permitted. Nor is this page intended to promote the construction and use of such transmitters by others. I DO give detail on the construction and theory of the transmitter in order to better understand and appreciate the technologies used at the dawn of the modern wireless era.

The concept behind the project is to build an operating spark-gap transmitter using technology as close as possible to that used at the turn of the 20th century. Technology that would be recognizable to Marconi or Fleming. This meant no chips, no transistors, no diodes, (OK, a galena and cat's whisker is technically a diode, but you know what I mean...) no solid-state. It does not mean using vintage components as these are for the most part unobtainable and/or prohibitively expensive. With each component new or homebrew the question is: would Fleming recognize this? Sometimes I needed to stretch things somewhat, especially concerning the relays, but still I believe I have remained true to this original goal.

The transmitter was conceived as two modules, a power module and an oscillation module. Size constraints forced me to break the oscillation module further into two breadboards, one for the variable HV capacitance and the other for the oscillation transformer. The power module is the most complex piece having on it the induction coil, interrupter, capacitor, and the spark gap itself. To keep this page length manageable and allow enough space to give a good description of each component, its theory and function, I have broken the descriptions into separate links below.

INDUCTION COIL..

COIL INTERRUPTOR..

SPARK GAP..

OSCILLATION CIRCUITS..

PERFORMANCE TESTING..

The following circuit diagram and photo shows the complete transmitter and how the modules integrate together:

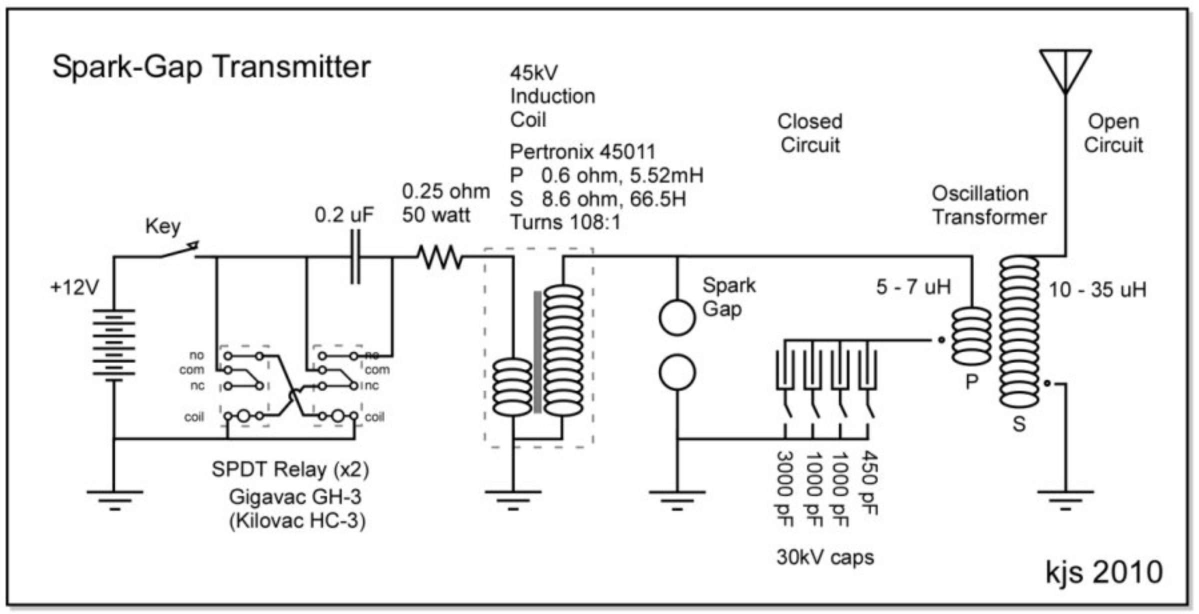

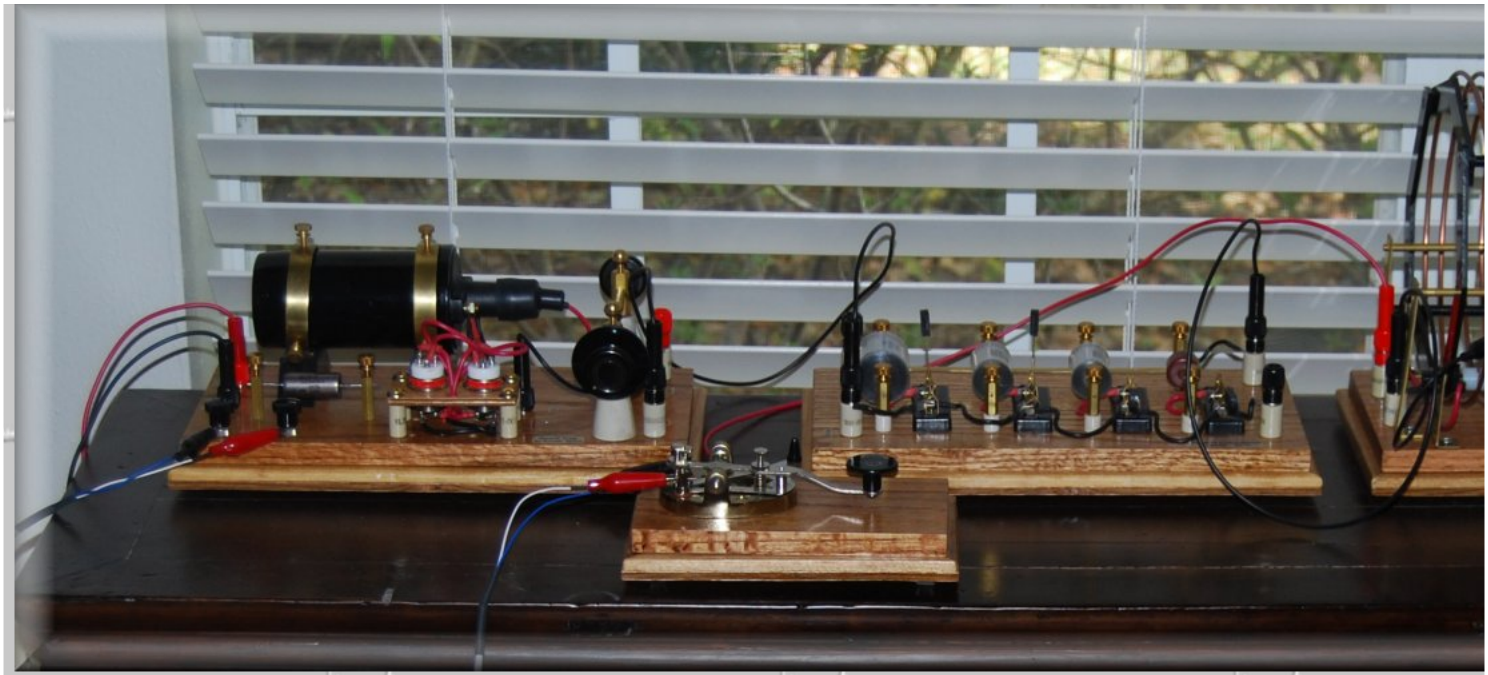

Return..

Back to Main....

Friday, 27-Oct-2017 23:41:47 CDT

Comments and/or Correspondance to: kj-smith@sbcglobal.net

.::. Alsace .::. Life List .::. Ionosphere .::. Crystal Radio .::. Snapshots .::. Thoughts .::.
.::. Sitemap .::. Beer Labels .::. International Pix .::. Western US Pix .::. Pinhole Pix .::. Petrology .::.

THE PROJECT CRYSTAL RADIO

(updated 4 Feb 01)

The Project Crystal Radio involves the construction, testing and operation of a series of crystal radios, starting with very basic set and ending up with a medium performance dx set, complete with antenna tuner. While it uses a smallish gage enameled wire (#24) and the cores of toilet paper or towel rolls, it works rather well throughout, and the coils are easy to wind. In addition to the suggested parts used here, you will also need a decent antenna and ground, preferably with alligator clips attached so you can fix them to your sets wherever needed. Throughout, I selected parts based on availability and cost. Using standard catalog parts from Mouser and Radio shack, you should be able to complete the project for about $25, less if you shop around, more if you decide to go with more expensive components. Here are the electronic parts I used:

Rf tuner capacitors- 2 - ElectronixExpress p/n 14VCRF10-280-P @ $1.35 ea.
Screws for tuner caps - 2 - Mouser p/n 48ss003 @ $1.20/100
Capacitors with knobs and mounting screws are also available at kitsrus
Capacitor with knob also available at Ocean State Electronics
Ceramic earphone- 1 - Mouser p/n 25CR060 @ 1.85
Magnet wire #24- 1 lb - Mouser p/n 501MW24H-1LB @ 10.42 (this gets you about 750 feet, you will only need about 200)
1N34A diode -Xtal set society, Radio shack, and others

47 kohm resistor- Radio Shack - @ .49 for 5 - either 1/4 or 1/2 watt will do
Parts from Mouser can be gotten from their web site; search using the part numbers I have given here.

Other stuff I used:

Fahnestock clips - 3 - Mouser p/n 524-11-4034C @ .21 ea.
Alligator clips - 8 - Radio Shack - p/n 270-356 or something similar -about 2.29.
Double sided sticky tape - couple of inches will do - rtv or hot glue will do as well, but is more permanent. Use for mounting the capacitors
Antenna, ground and hookup wire - a 90 foot roll of #22 stranded hookup wire from Radio Shack at 3.99 will do.
1 foot of 1 x 4 or 1 x 6 pine
1 foot of 1/4 inch lath (see photo)
1/2 inch wood screw - 3 - to hold fahnestock clips to base.
1 inch wood screw - 2 - to hold uprights onto base
Two plastic bottle caps for tuning knobs
Two small washers from hardware store (Ace), about 1/3 inch diameter, to fit under screws for capacitors and hold "knobs" in place.
1/8 to 1/4 inch dowel, 13 inches
Toilet paper cores - 3. Paper towel rolls cores also are ok. Diameter should be about 1 3/4 inch.

The two larger coils we will use are about 400 uH each, and the capacitors will go from about 20 to 220 pF -enough to cover the AM broadcast band. If you want to play around with different size coils, wire, capacitors, etc., go ahead if you know how to adjust. If not, I recommend you keep to the directions.

Construction: First, let's look at a picture of the completed base of the set and the two main coils you'll wind.

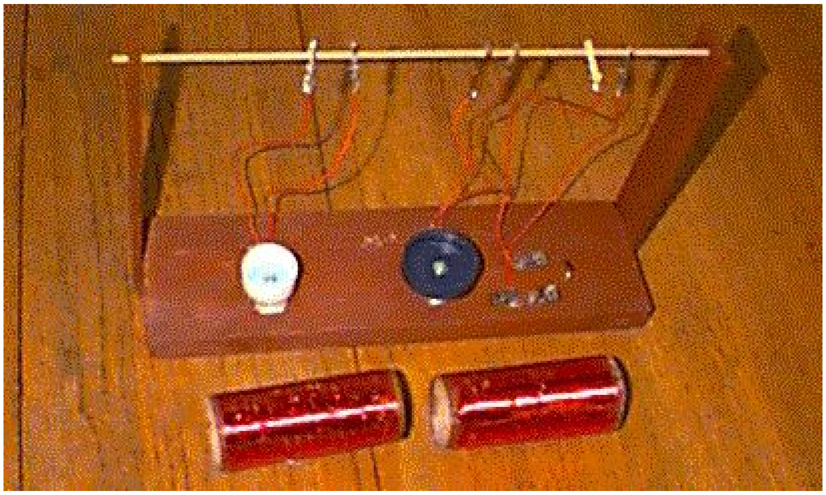

The coil on the left is coil#1, and has 17 taps. The other coil is coil #3. The vertical strips are the lath material, screwed to the 1 foot long base. Holes drilled in the ends of the lath hold the dowel rod; it can be removed to change coils. The bottle caps are the knobs of the two polyfilm rf tuner capacitors, which are stuck to the base with sticky tape. On the right of the base are screwed 3 fahnestock clips. The detector is connected between the rear clip and the right front clip. Leads with alligator clips are connected to the right rear and front left fahnestock clip, and leads and alligator clips are also connected to each capacitor. The two front fahnestock clips are for phones. Antenna and ground leads should be connected to the coils using alligator clips. You can use the picture to figure out how to put the base and parts together. I will show a "ready to listen" picture later.

Before you wind coils, I recommend you put your base together as shown above. You can be much less "elegant" than this, such as putting the whole mess on a shoe box top. I used the wooden base for stability, however, and the elevated coils to give better access to the coil connections and to get the coils a bit away from surrounding metal, which can adversely affect coil Q.

Preparation of the capacitors:
The capacitors I used have three connections. The center one is the base, and the outside ones are for the rotors of the two sections. You need to connect the two outside ones, and connect them to one lead, about 8 inches long, and connect the center connection to another lead. If you look on the back of the capacitor, you will notice two small trimmer capacitors with a screwdriver adjustment. Adjust both these to minimum capacitance (you will see a full circle instead of a half-moon shape); this will give you full BC band coverage when used with the coils you will wind next. Cut holes in the center of the two plastic bottle caps ,and shape them with a knife to fit snugly over the tuning shaft of the capacitors, and then hold them in place with a small washer and screw. Use the double sided sticky tape to mount the capacitors on the base, and fasten alligator clips to the ends of the leads. To avoid working the delicate capacitor leads, I used an ordinary stapler to staple the leads to the base close to the caps.

Construction of coil #1: This takes about 72 feet of #24 wire. Using a pin, punch two holes about 1/4 inch apart close to and parallel to one end of your coil form. Thread about an inch of wire into one hole and out the other to anchor your coil. Now wind about 152 turns of wire in a single, closely wound layer along the coil; at every 10 turns (yes 10), make a small loop that sticks above the coil, each loop being about 1/4 inch in diameter - you will get the hang of this quickly. To avoid crowding these loops, which you will use as taps, you might stagger them so every other tap is offset on the coil by about 1 inch; you will end up with two rows of taps. After the last turn, make a final tap and, use the pin to make two more small holes as you did in the beginning, and thread about an inch of wire in and out to anchor the other end of the coil. You can use either cement or white glue to hold your finished turns onto the coil form. Tape is ok too; just don't cover the taps. Now comes the really tedious part; getting the enamel off the taps. I usually take a small flame, as from a lighter or candle, and burn the enamel off the tap, then lightly sandpaper them clean. If you like, do as I do and tin the taps with solder. This coil will eventually be used for your antenna tuner, hence all the taps. I don't draw well, so here is a link to what a tap should look like:

http://crystalradio.net/beginners2/index.shtml

After you have completed coil #1, assembled the base, and put up your antenna and ground, you are ready to build the first couple or 3 radios in the project.

Construction of coil #2: For this coil I recommend 30 turns on the same diameter core, with taps for connections only at either end. Takes about 14 feet of wire. If you want to, you can actually skip making this coil, and use coil 1 or 3 in its place, using the clips to select a different number of "active" turns.

Construction of coil #3: This coil can be identical to coil #1, but, since I didn't think I needed all the taps, I just put them at 0, 15, 25 50, 75, and 152.

Let's make a radio: The first radio is a simple conductively coupled set, using only the antenna/ground, coil #1 and the detector and phones (earplug). It is commonly built as the "oatmeal box" set. Here is the schematic for it:

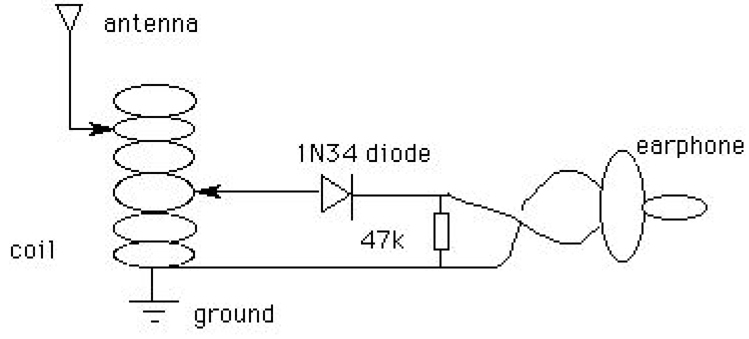

things to try:

The set is "tuned" by using the inductance of the coil and the capacitance between the coil turns by attaching the antenna to various taps. Put the detector tap on different taps for the best signal. You will notice that this set tunes rather broadly, but may receive some shortwave as well as broadcast band stations. You can try disconnecting the ground connection from the coil and connecting it and the detector tap to each other, putting the detector and earphone in series with the coil. Try reversing the diode, and try it without the 47 kohm resistor.

Radios number two and three. These use the same components as above, but with the addition of one of the variable capacitors. Leads from the capacitor go to both ends of the coil for radio number two, creating a parallel tank circuit. For starters, tap the detector about midway on the coil, and connect the antenna to a tap close to the ground end. Adjust both taps for best reception. At this point, you should be wondering "what should I be able to hear?" An educated guess can be obtained by using the scan function on your car radio. You should expect to hear with your xtal set the stations that the radio stops on every time it tunes through the AM band. Strong stations may dominate parts of the band on your xtal set, but we'll work on that later. For some reason, most renditions of this set put the detector and the antenna right at the top of L1/C1; this never works very well for me, but putting the detector between 30 and 50 percent up the coil and the antenna tap somewhere near the bottom seems about right. For a simple radio, just connect the antenna to the top of the tank circuit, and forget the ground; this is the basic kid radio, and you can go around touching the antenna connection to various metal objects around the house to see what you can get. I get a couple of locals this way. Here is a schematic of this radio.

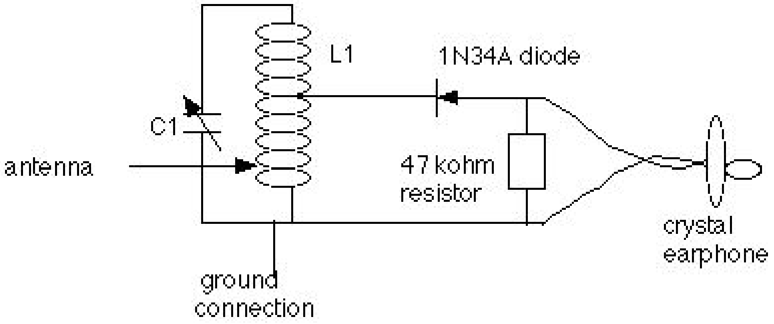

An alternative to this arrangement is to put the coil and capacitor in series instead of in parallel, thereby making radio number three. Your choice as to which is closest to the antenna. Give them both a try. Tap the detector and phones at several points and see which is best. By this time you should be getting a feel for the best taps for selectivity and sensitivity (read loudness) for your detector and phone taps. Feel free to reverse them, to reverse the diode, and to try different diodes or add diodes in series or parallel with each other to see the results. Oh yeah, by tapping the antenna way down the coil, you may pick up some shortwave stations, particularly at night. The series xtal set is shown here:

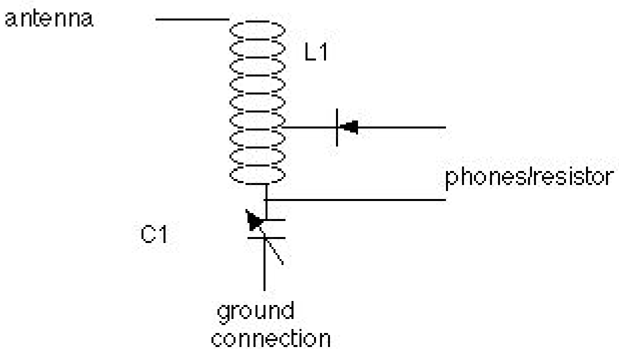

Radio number four. By now you may be wanting a little more selectivity, particularly if you have two strong locals close together on the dial. Here is one way to do it. First, return your radio to the parallel tank circuit of radio number 2. Put coil #2 on the rod near the ground end of coil #1 (the end connected to the earphone lead). Now connect the antenna and ground to coil #2. By separating coils 1 and 2 you vary the coupling between the antenna system and the set, and can pick up some selectivity. This should allow you to keep the detector on the best all round tap, and not lose sensitivity by having to go to a lower tap to gain selectivity. Feel free to experiment. Here is the arrangement for this set:

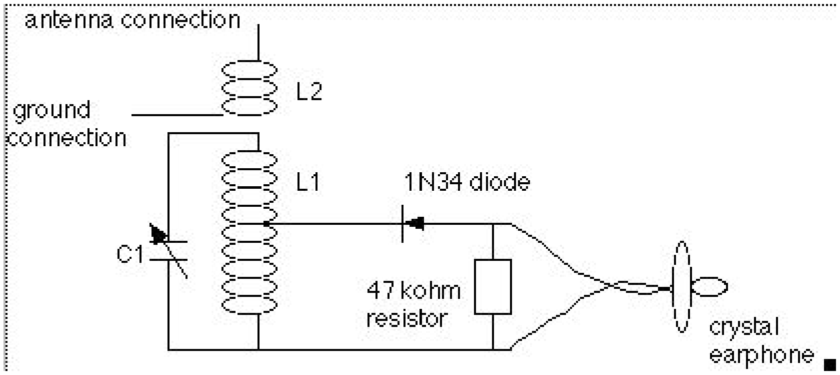

Radio number five. This one uses coil #3, which has been waiting patiently in the wings; if not, then wind it now. For this set, you will take coil #2 off the dowel, and put on coil #3 in its place. Now connect the detector/phone leads to coil #3. This gives you a separate antenna tuner using coil#1, inductively coupled to the set's tank circuit which now uses coil #3. Adjust for the best taps. Use coil #1 in the series or parallel tank circuit configuration of radios two and three, or just use as an antenna loading coil, without the capacitor. A schematic of this arrangement withcoil #1 as a series tuned antenna tuner is here: (note: as built, the antenna tuner configured as a series tuner shown below didn't tune belowabout 650 kHz; to go lower I had to shift to a parallel aarrangement, and connect the antenna to a tap on the coil)

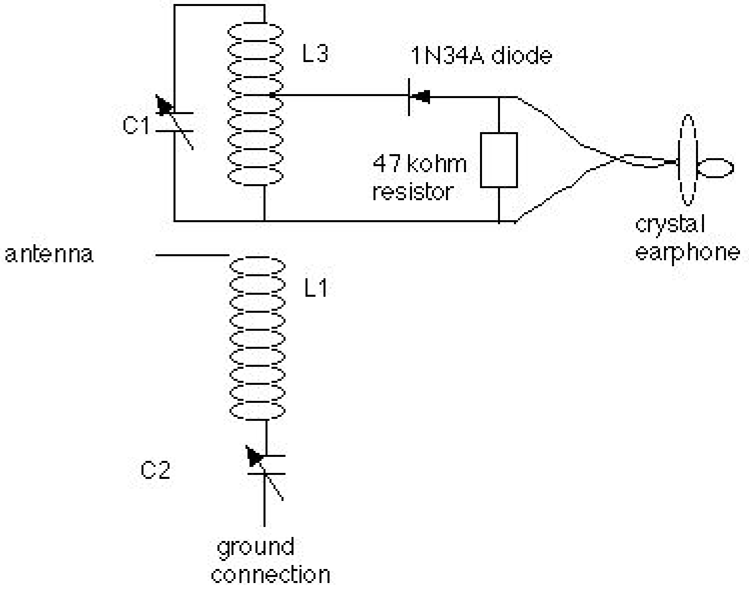

From here on out, we will use the configuration of radio number five, and try out different ways to tune the antenna. With the set you now have, you can try out any of the configurations on my antenna page, linked below, with the single exception of the one with two capacitors on the antenna tuning coil. For that, you will need to get either a third capacitor or a ganged dual capacitor of sufficient capacity - I recommend the latter. From here on out you're on your own, but you now have a set that will pick up ssome good dx, and give you some experience in preparing to build your own ultimate rig Have fun.

Oh yeah, here's a picture of the set in a "ready to listen configuration" - just add antenna and ground.

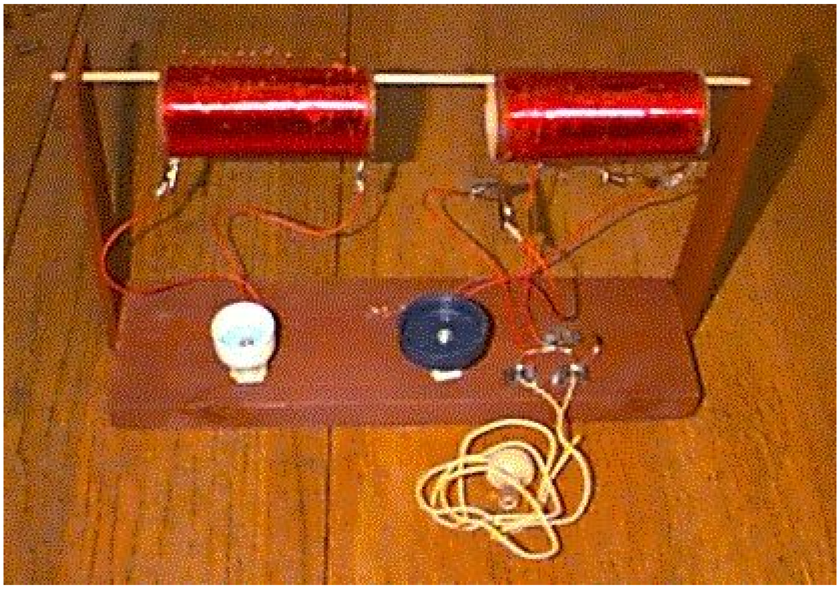

One final after note: This set has a number of compromises which were made on the basis of parts availability and cost. In fact, I scaled it up to incorporate some more advanced (and costly) features for my own use. The basic design is sound however, and it is my new "gold standard". It also has the flexibility to allow me to try out new ideas with little difficulty. The only claim I make is that you cannot buy a better set, either ready-made or in a kit. During the latest (2000) dx contest, I used this set as shown, but with the earphone under a pair of sound isolation muffs, and with one of my in-line wave traps. Got 52 stations just listening in my spare time, which wasn't bad at all. The rig I sent in my logs for was a scaled up version of this with some detector and headphone mods, and logged 98.

For a nicer looking version of this set, see what Dennis Foster, KK5PY sent me:

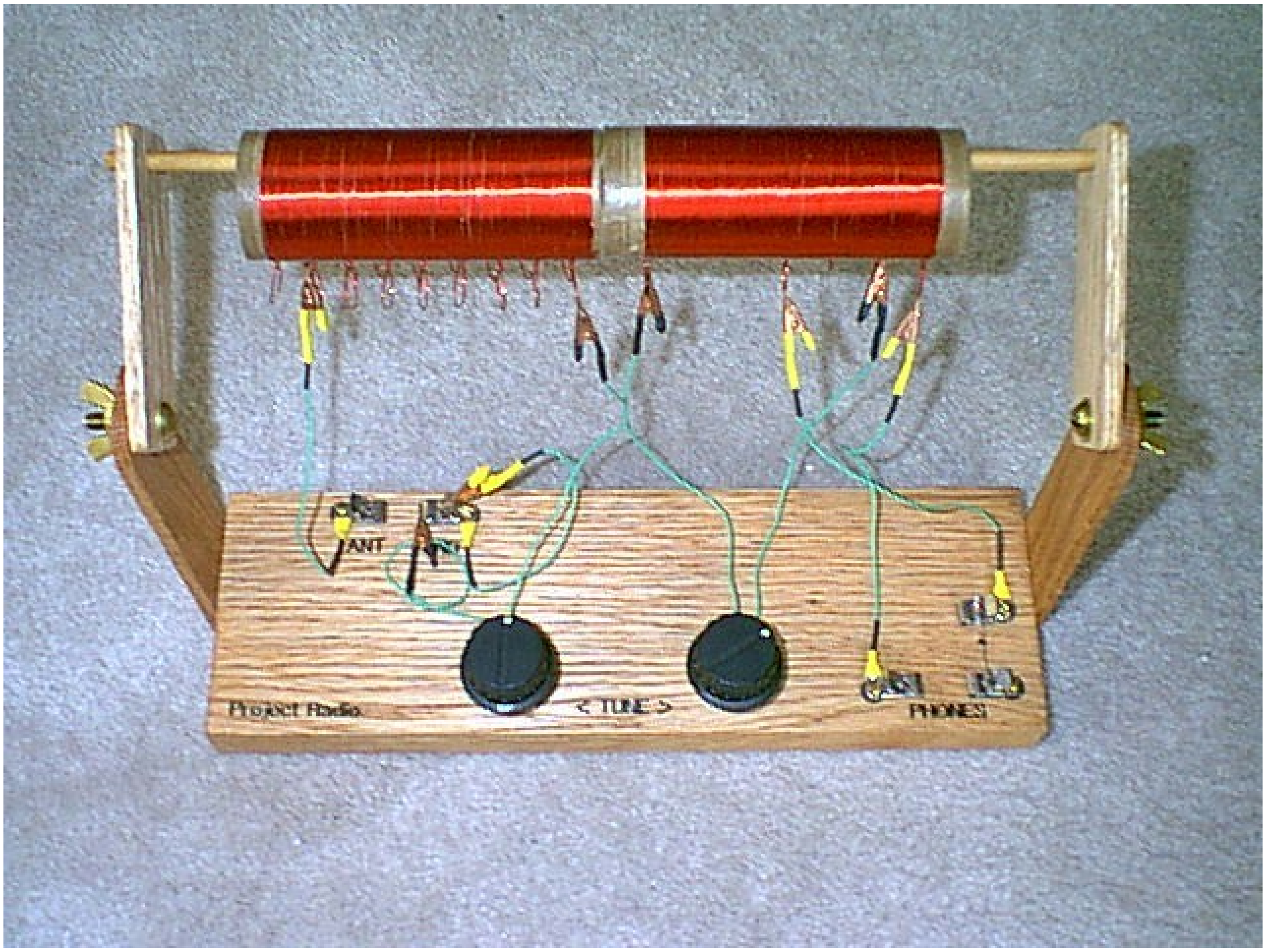

HOME LINKS
CONSTRUCTION TIPS COMMENTS
ANTENNAS
TESTING
REVIEW

US 20110175461A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2011/0175461 A1**

Tinaphong (43) **Pub. Date: Jul. 21, 2011**

(54) **METHOD AND APPARATUS FOR HARVESTING ENERGY**

(75) Inventor: **Prapan Paul Tinaphong**, Westfield, IN (US)

(73) Assignee: **AUDIOVOX CORPORATION**, Hauppauge, NY (US)

(21) Appl. No.: **12/930,421**

(22) Filed: **Jan. 6, 2011**

Related U.S. Application Data

(60) Provisional application No. 61/335,459, filed on Jan. 7, 2010, provisional application No. 61/402,481, filed on Aug. 31, 2010, provisional application No. 61/404,290, filed on Sep. 30, 2010.

Publication Classification

(51) **Int. Cl.**
H02J 17/00 (2006.01)

(52) **U.S. Cl.** **307/149**

(57) **ABSTRACT**

An energy harvesting circuit includes one or more broadband or narrow band antennas to detect WIFI (wireless fidelity) or other RF (radio frequency) signals. The signals are rectified and voltage multiplied, and the resultant DC voltage is provided to a power management circuit. The output of the power management circuit charges a lithium battery or other storage device within the energy harvesting circuit. The energy stored in the battery or storage device is provided through a DC/DC converter circuit to a USB output connector to power or recharge the batteries of an external electronic device connected to the USB output connector of the energy harvesting circuit.

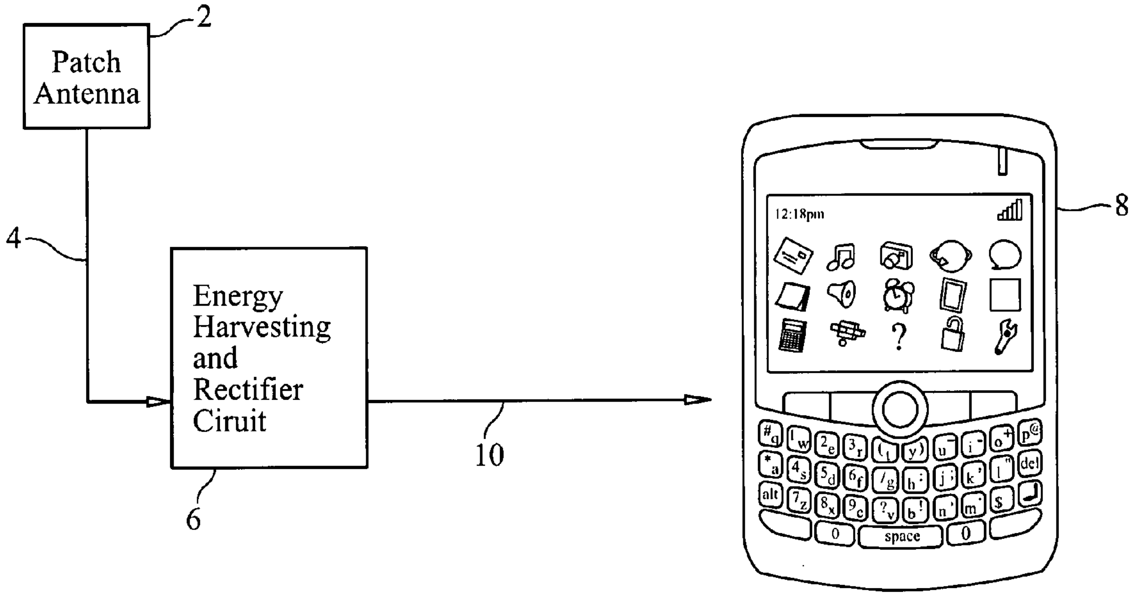

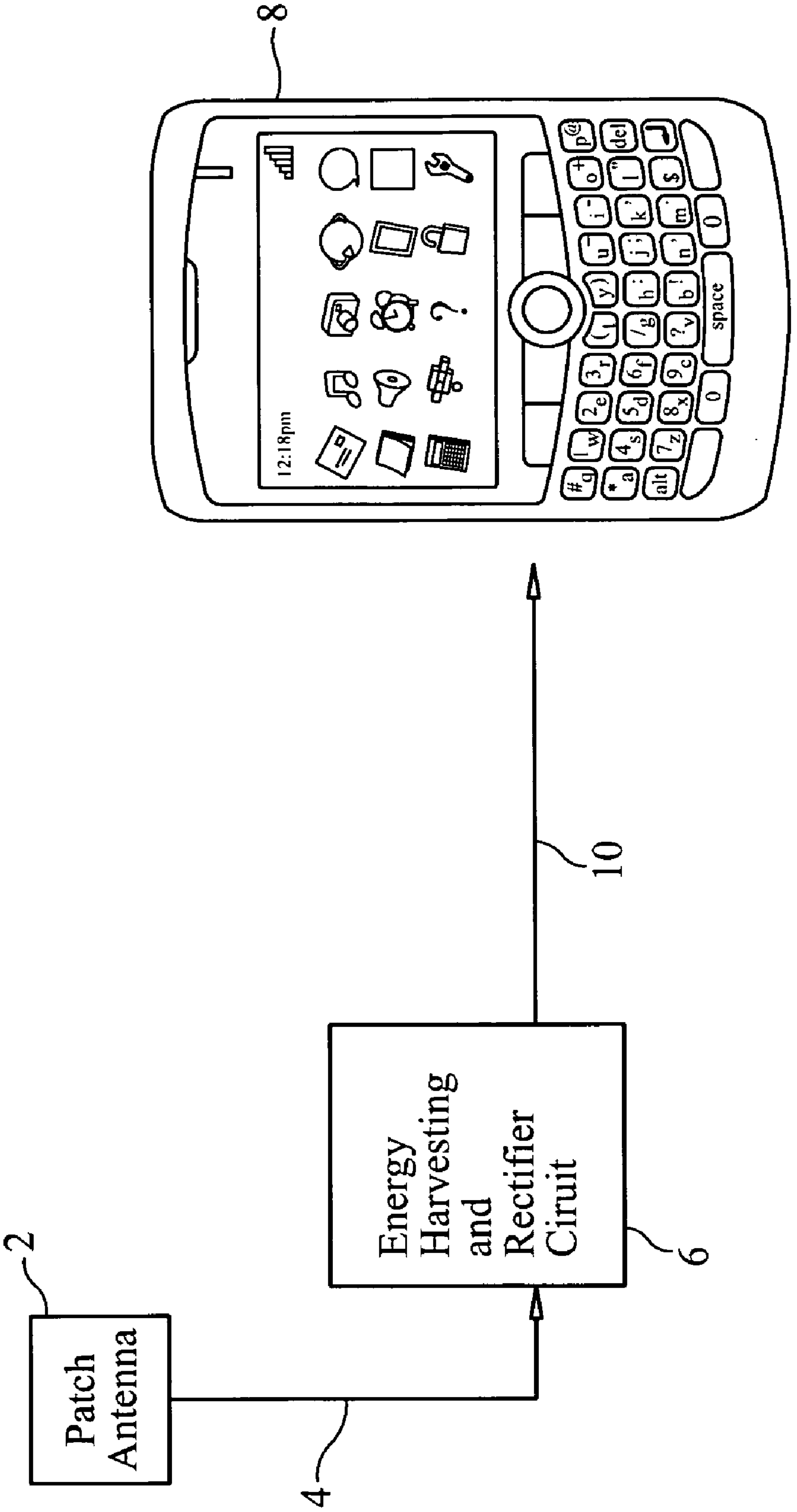
Patent Antenna
2
4
Energy Harvesting and Rectifier Ciruit
6
10
8
12:18pm
space

FIG. 1

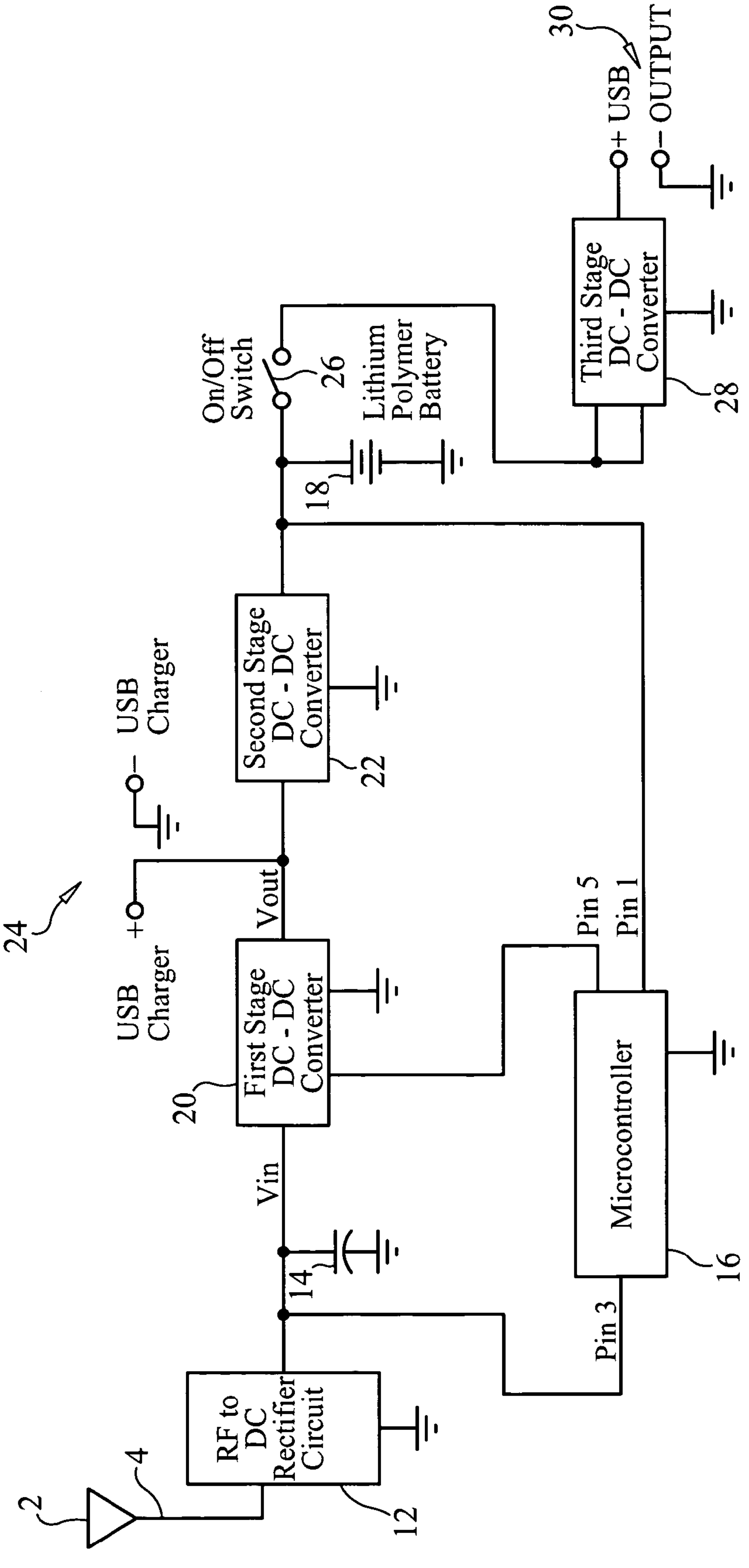
2
4
RF to DC Rectifier Circuit
12
14
Pin 3
Vin
First Stage DC - DC Converter
20
USB Charger
24
+
USB Charger
−
Vout
Microcontroller
16
Pin 5
Pin 1
Second Stage DC - DC Converter
22
18
Lithium Polymer Battery
On/Off Switch
26
Third Stage DC - DC Converter
28
30
+ USB
− OUTPUT

FIG. 2

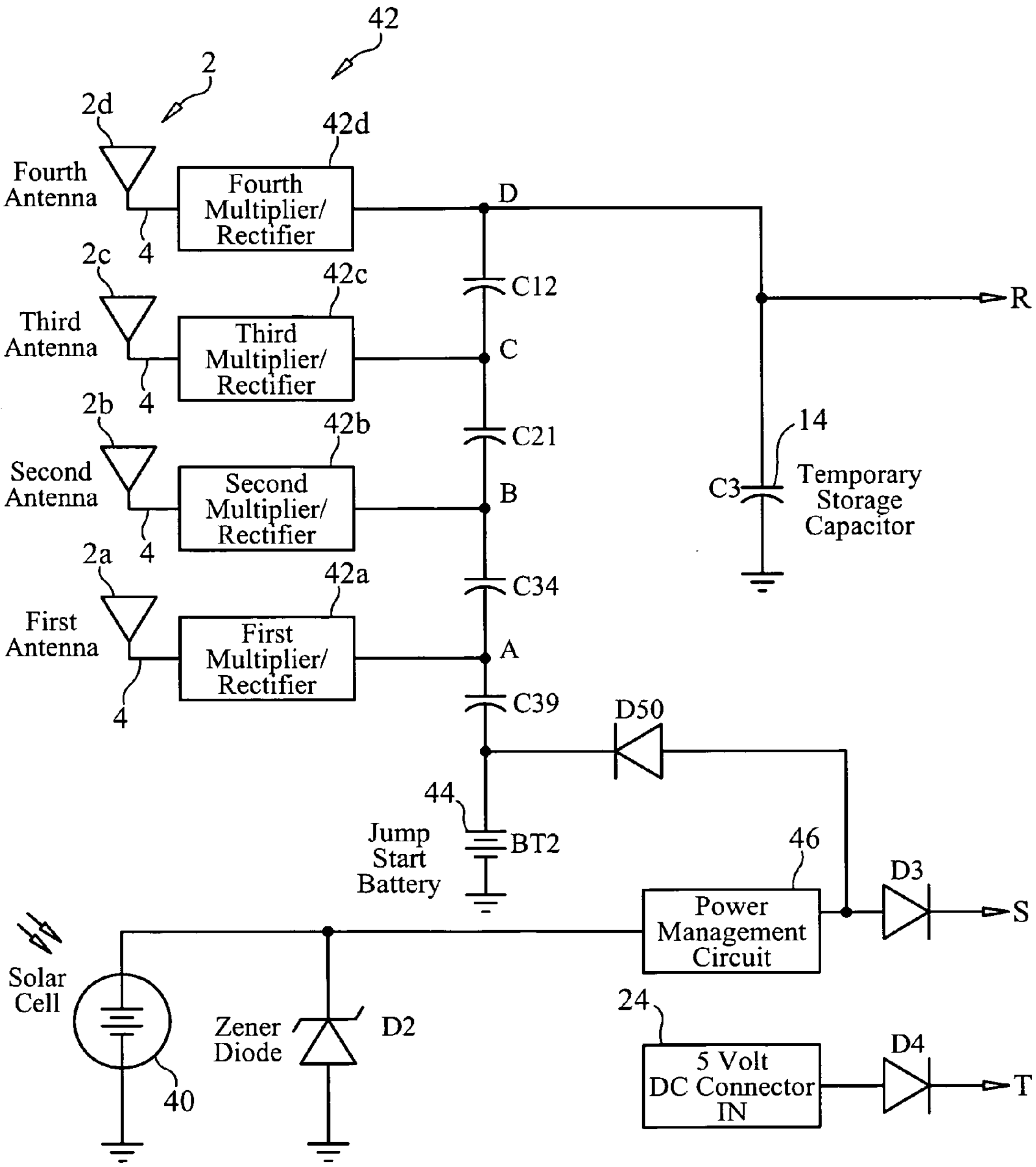
42
2
2d
42d
Fourth Antenna
Fourth Multiplier/ Rectifier
D
4
2c
42c
C12
R
Third Antenna
Third Multiplier/ Rectifier
C
2b
4
42b
C21
14
Second Antenna
Second Multiplier/ Rectifier
B
C3
Temporary Storage Capacitor
2a
4
42a
C34
First Antenna
First Multiplier/ Rectifier
A
4
C39
D50
44
Jump Start Battery
BT2
46
D3
Power Management Circuit
S
Solar Cell
40
Zener Diode
D2
24
5 Volt DC Connector IN
D4
T

FIG. 3A

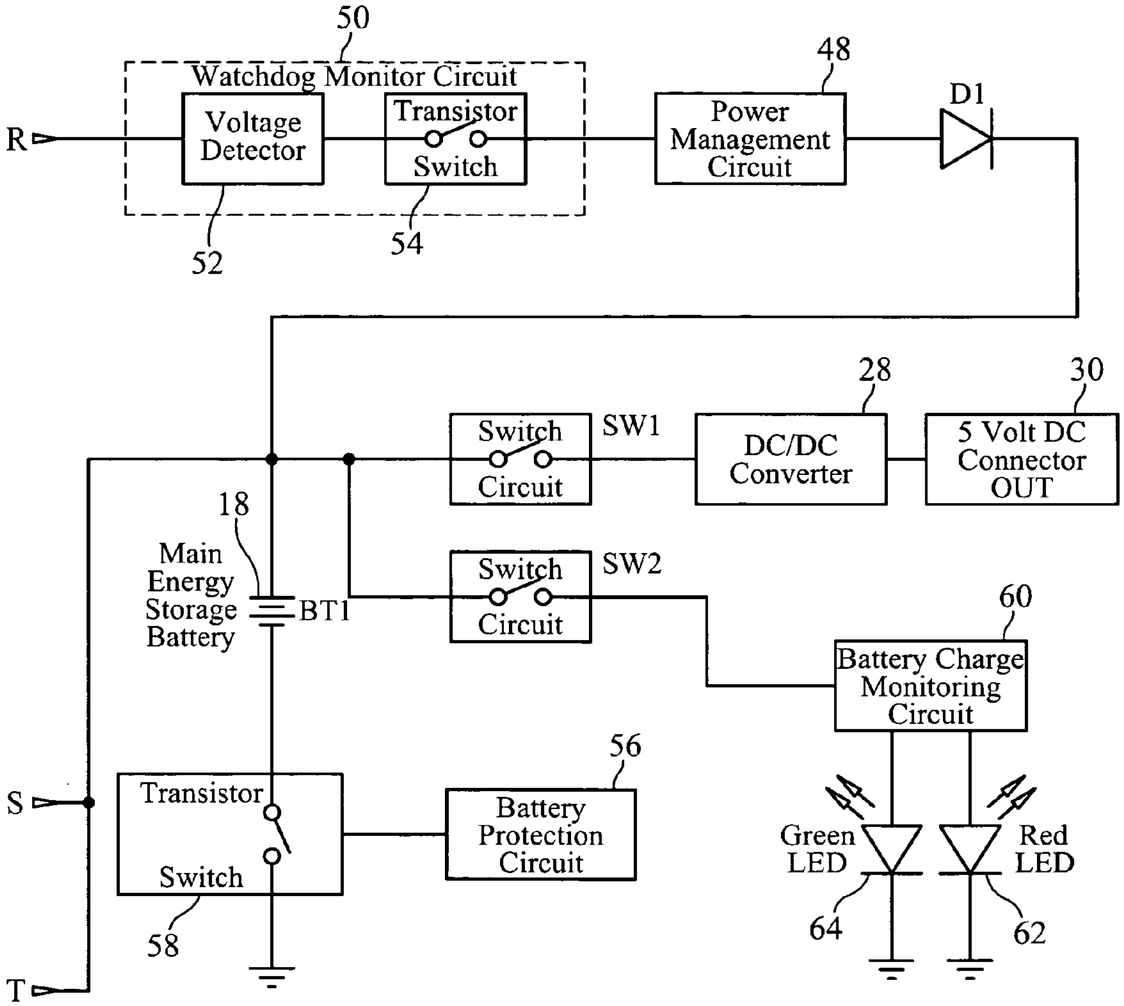
50
Watchdog Monitor Circuit
R
Voltage Detector
Transistor Switch
52
54
48
Power Management Circuit
D1
28
30
Switch Circuit
SW1
DC/DC Converter
5 Volt DC Connector OUT
18
Main Energy Storage Battery
BT1
Switch Circuit
SW2
60
Battery Charge Monitoring Circuit
56
S
Transistor Switch
Battery Protection Circuit
Green LED
Red LED
58
64
62
T

FIG. 3B

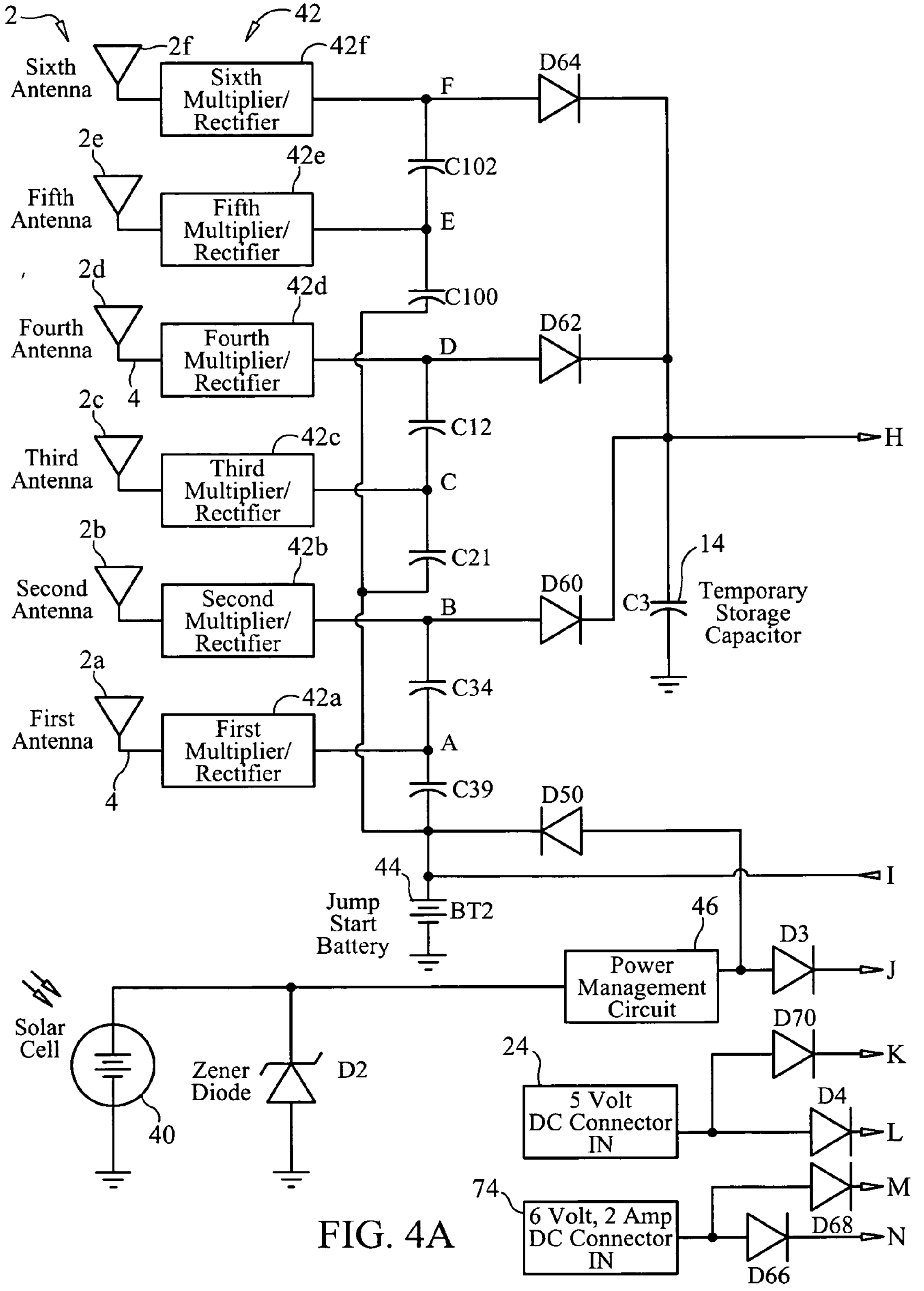
2
42
Sixth Antenna
2f
Sixth Multiplier/ Rectifier
42f
F
D64
C102
2e
Fifth Antenna
42e
Fifth Multiplier/ Rectifier
E
2d
Fourth Antenna
42d
Fourth Multiplier/ Rectifier
C100
D
D62
4
2c
Third Antenna
42c
Third Multiplier/ Rectifier
C12
C
H
2b
Second Antenna
42b
Second Multiplier/ Rectifier
C21
B
D60
C3
14
Temporary Storage Capacitor
2a
First Antenna
42a
First Multiplier/ Rectifier
C34
A
C39
D50
4
44
I
Jump Start Battery
BT2
46
Power Management Circuit
D3
J
Solar Cell
40
Zener Diode
D2
24
5 Volt DC Connector IN
D70
K
D4
L
74
6 Volt, 2 Amp DC Connector IN
M
D68
N
D66

FIG. 4A

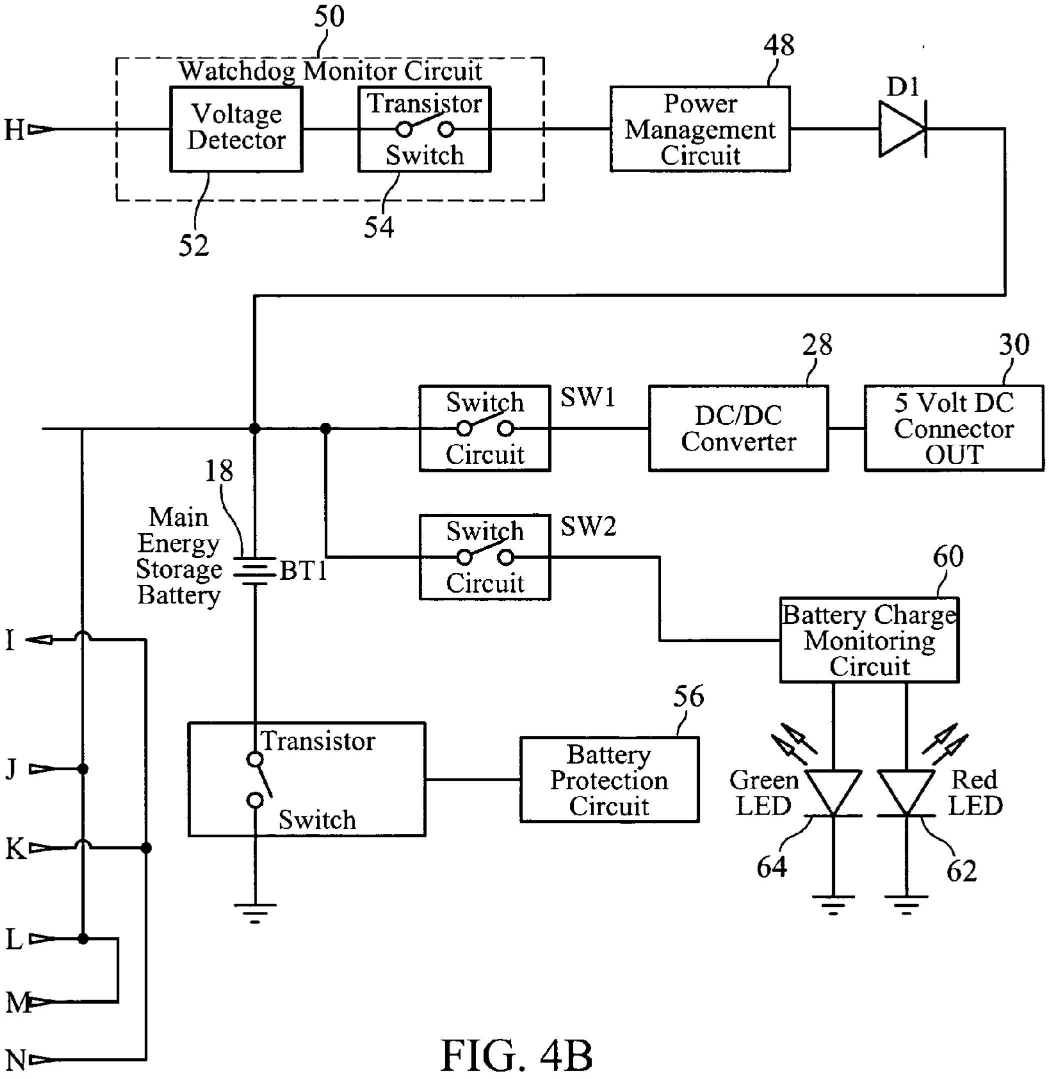
50
Watchdog Monitor Circuit
Voltage Detector
Transistor Switch
52
54
H
48
Power Management Circuit
D1
28
30
Switch Circuit
SW1
DC/DC Converter
5 Volt DC Connector OUT
18
Main Energy Storage Battery
BT1
Switch Circuit
SW2
60
Battery Charge Monitoring Circuit
I
J
K
L
M
N
Transistor Switch
56
Battery Protection Circuit
Green LED
Red LED
64
62

FIG. 4B

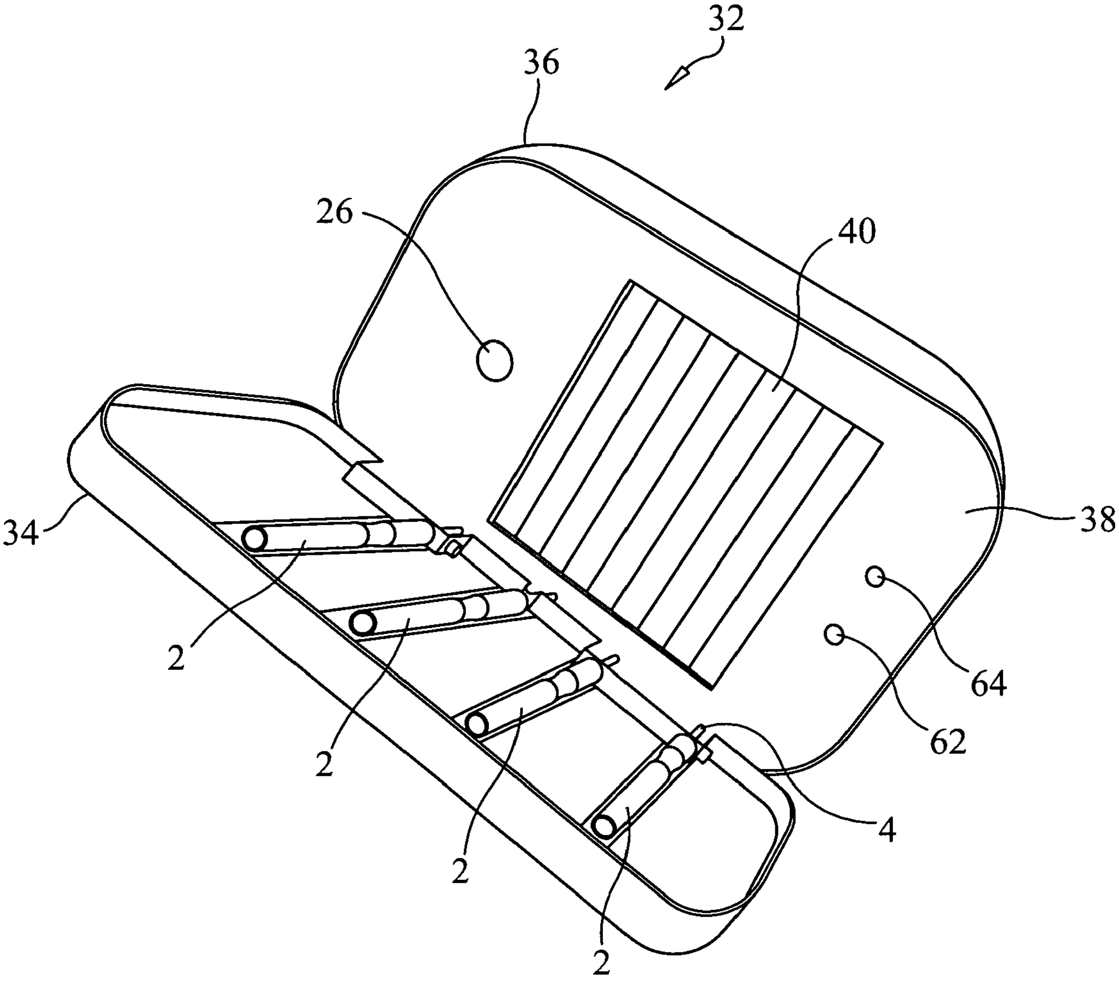
32
36
26
40
34
38
2
64
2
62
2
4
2
2

FIG. 5

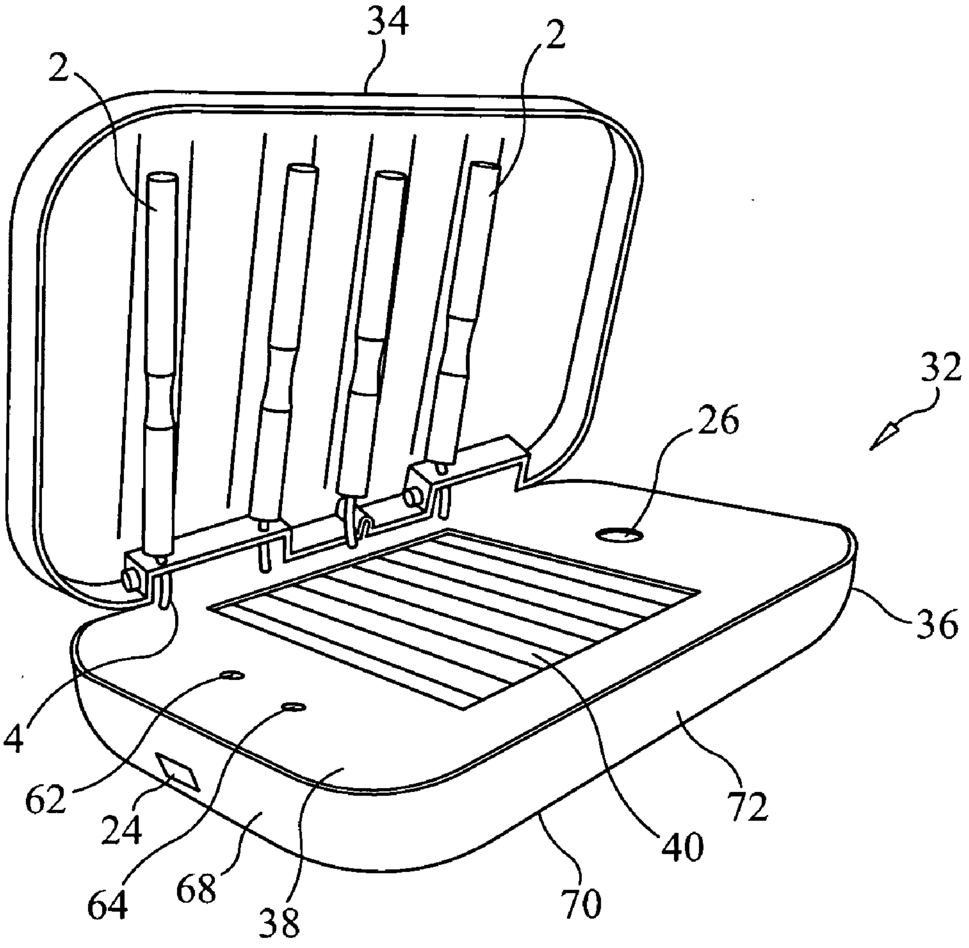
2
34
2
32
26
36
4
62
24
64
68
38
70
40
72

FIG. 6

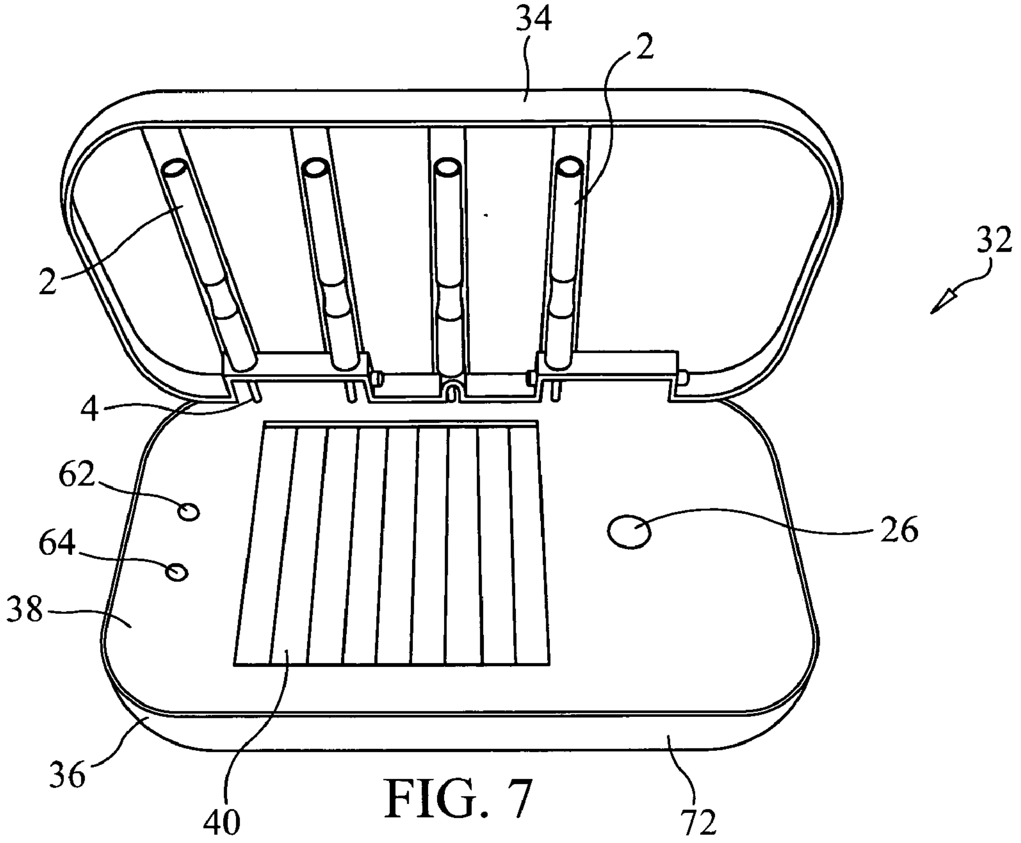
34
2
2
32
4
62
64
38
26
36
40
72

FIG. 7

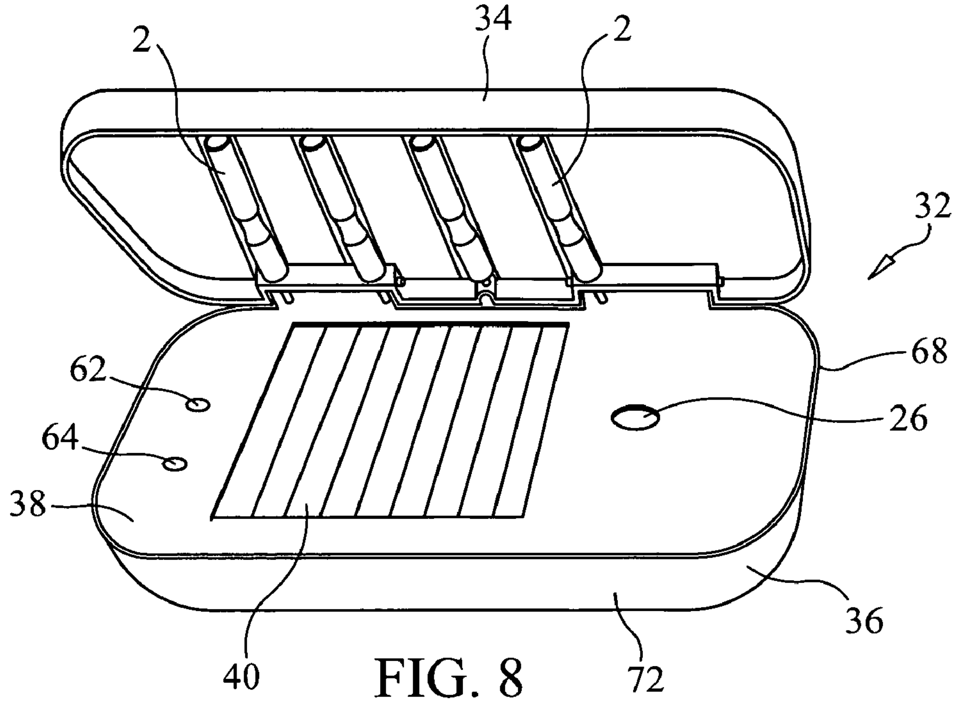
2
34
2
32
68
62
26
64
38
36
40
72

FIG. 8

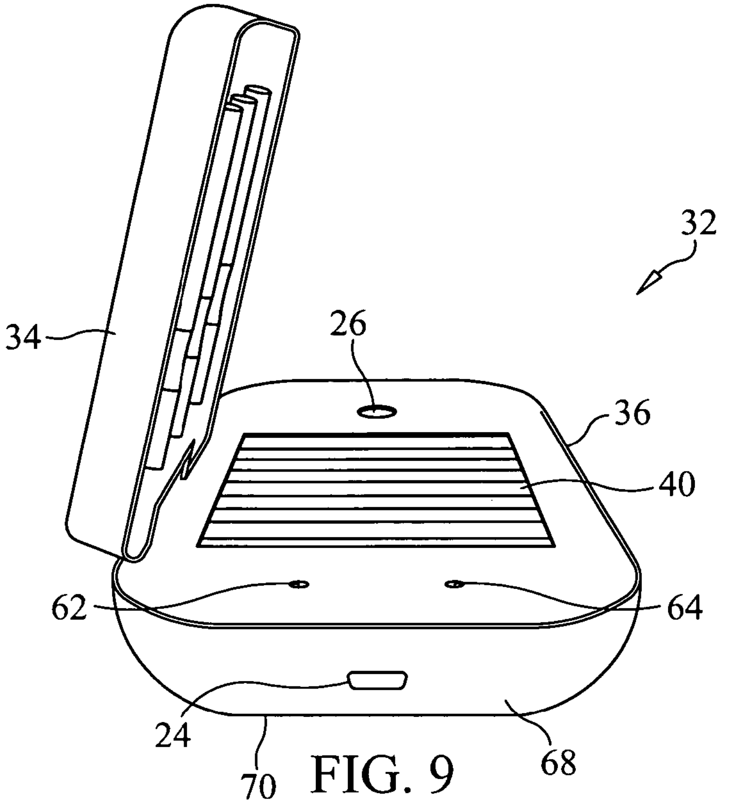
32
34
26
36
40
62
64
24
70
68

FIG. 9

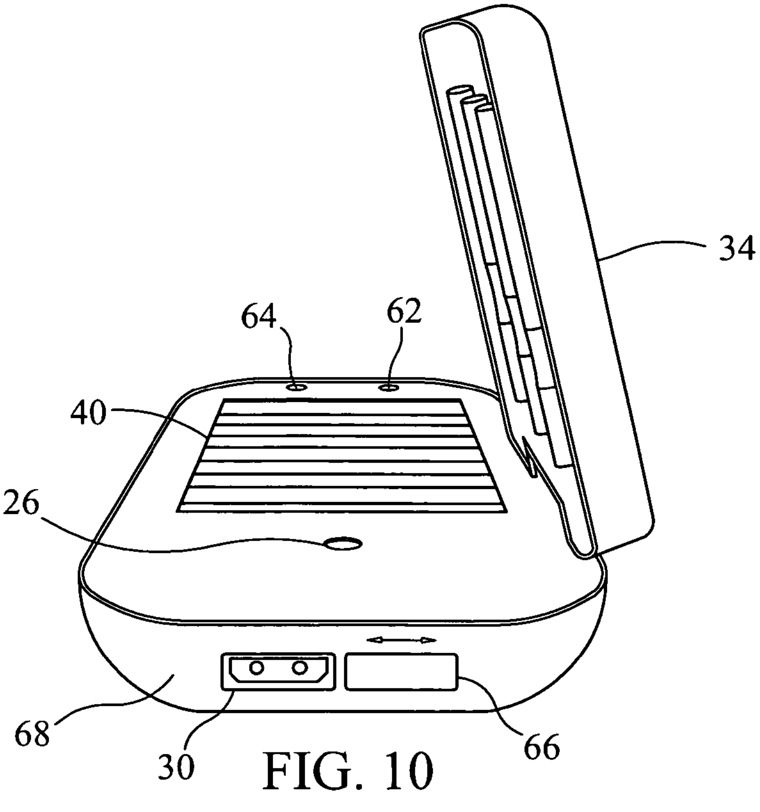
34
64
62
40
26
68
30
66

FIG. 10

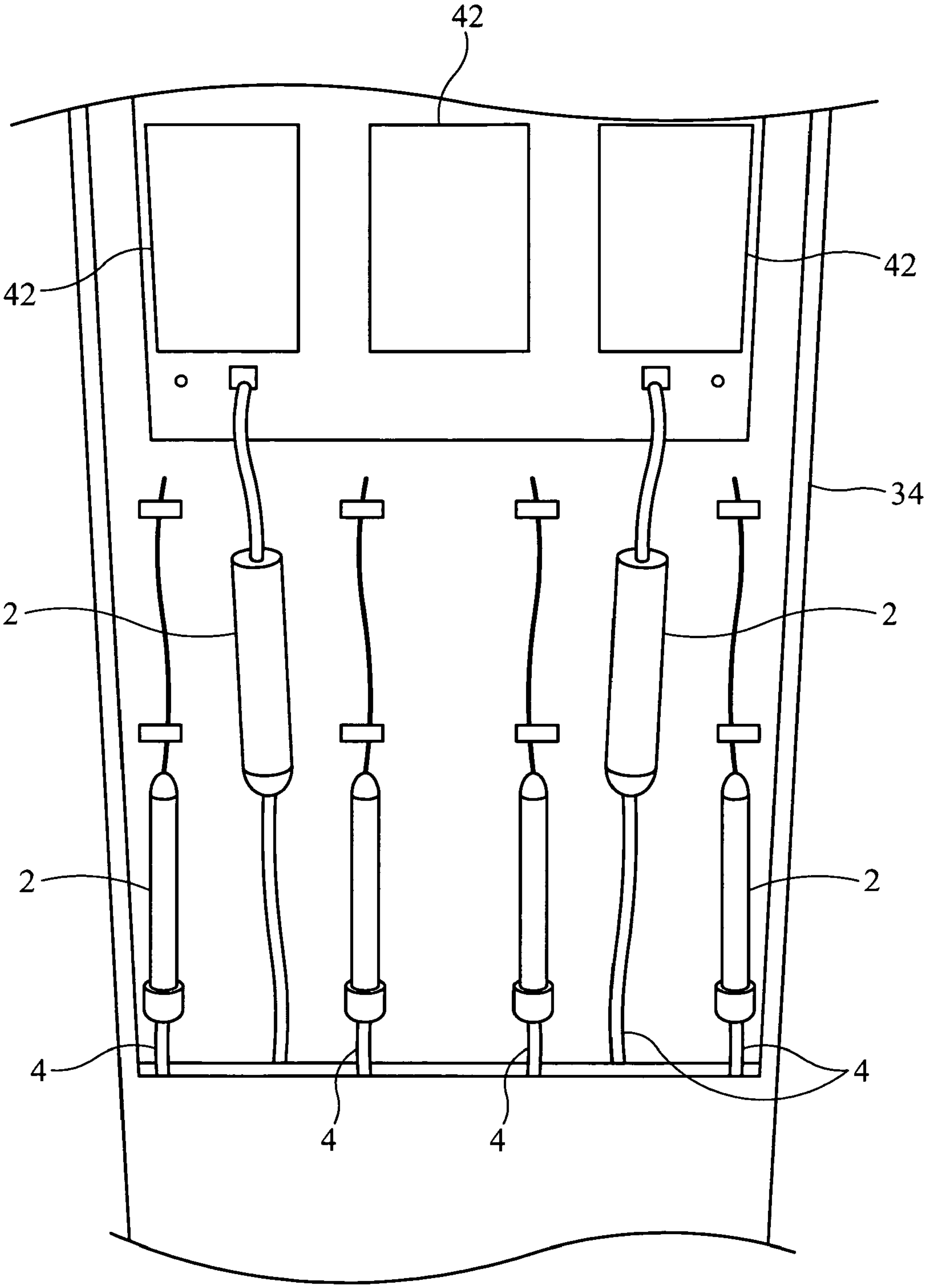
42
42
42
34
2
2
2
2
4
4
4
4

FIG. 11

METHOD AND APPARATUS FOR HARVESTING ENERGY

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is related to U.S. Provisional Application Ser. No. 61/335,459, filed on Jan. 7, 2010, and entitled “Method and Apparatus for Harvesting Energy from WIFI Signals”, U.S. Provisional Application Ser. No. 61/402,481, filed on Aug. 31, 2010, and entitled “Method and Apparatus for Harvesting Energy”, and U.S. Provisional Application Ser. No. 61/404,290, filed on Sep. 30, 2010, and entitled “Method and Apparatus for Harvesting Energy”, the disclosures of which are incorporated herein by reference and on which by priority is hereby claimed.

BACKGROUND OF THE INVENTION

[0002] 1. Field of the Invention

[0003] The present invention relates to energy conservation methods and devices, and more particularly relates to devices which can harvest energy.

[0004] 2. Description of the Prior Art

[0005] Many devices have been proposed or implemented which harvest energy from different types of naturally-occurring energy sources. These energy sources include solar, wind, tide, current and wave power, the energy of which is converted by such devices into electrical power to supplement or replace electrical power provided by a power company. Except for solar cells powering hand held calculators and other small electronic devices, most energy harvesting apparatus are relatively large, expensive and inefficient. To the knowledge of the inventor, there are no RF (radio frequency) commercially available devices which harvest energy from WIFI (wireless fidelity), WiMax, Satellite Radio, Cellular Phone (850 MHz to 900 MHz and 1.8 GHz-1.9 GHz), or any 2.4 GHz and UHF TV signal sources to power or recharge the batteries of small electronic devices, such as cellular phones, personal digital assistants (PDAs), MP3 players or the like.

OBJECTS AND SUMMARY OF THE INVENTION

[0006] It is an object of the present invention to provide a device which harvests energy from WIFI, WiMax, Satellite Radio, Cellular Phone (850 MHz to 900 MHz and 1.8 GHz-1.9 GHz), and any 2.4 GHz and UHF TV signals or any signal from outdoor antenna or CATV feed through any “F” type connectors at consumers’ homes.

[0007] It is another object of the present invention to provide a method and device for harvesting energy from two or more alternative sources of energy such as solar and vibration from a microphone for powering or recharging the batteries of an electronic device.

[0008] It is yet another object of the present invention to provide an energy harvesting device which stores energy for later use to power an electronic device connected thereto.

[0009] It is a further object of the present invention to provide a device for harvesting energy from both WIFI or other RF signals and solar power to power an external electronic device connected thereto.

[0010] It is yet a further object of the present invention to provide apparatus for harvesting energy from WIFI or other RF signals which converts WIFI or RF signals received thereby into a DC (direct current) voltage, the energy from which is stored in a storage device within the apparatus, and which further has the capability of being connectable to a source of electrical power for storing energy therefrom in the storage device for later consumption by a user to power an external electronic device connected to the energy harvesting apparatus.

[0011] In accordance with one form of the present invention, a device for harvesting energy from WIFI or other RF signals includes one or more broadband antennas which are respectively connected to one or more voltage multiplier/rectifier circuits. Each voltage multiplier/rectifier circuit converts the WIFI or RF signals received by the respective antenna coupled thereto into a DC voltage, having an amplitude multiplied a number of times, such as by a factor of 4, 6 or 8. The DC voltage output signals from the voltage multiplier/rectifier circuits are added in series and in the same polarity (i.e., the voltages are “stacked”) or in parallel to provide a combined, detected DC voltage signal, which is provided to a temporary storage device, such as a capacitor.

[0012] When the voltage on the capacitor reaches a predetermined level, a watchdog monitor circuit effectively connects the temporary storage capacitor to the input of a power management circuit. The power management circuit includes a charge pump circuit which increases the voltage on the input of the power management circuit to a higher, more useable voltage on its output for charging a storage device, such as a lithium battery (for example, lithium polymer, lithium-ion or any lithium chemical type). Accordingly, the output voltage of the power management circuit is at a predetermined voltage level to charge the lithium battery or other storage device.

[0013] In a preferred form of the present invention, the device further includes a solar cell which converts solar energy from light impinging thereon to an electrical voltage, which is also filtered and provided to a second power management circuit. The output of this second power management circuit also provides a voltage at a predetermined level to the lithium battery or other storage device in response to the voltage from the solar cell resulting from the light impinging thereon.

[0014] The output of the lithium battery or storage device is connected to the input of a DC/DC boost converter circuit, which provides preferably a 5 volt DC voltage on a USB output connector to which is connectable the external electronic device to be powered thereby or whose battery is to be charged.

[0015] In a preferred form of the invention, the energy harvesting device of the present invention may include an input connector for receiving a charging voltage from an external power source, such as another electronic device, or a power transformer for such device which is connected to a cigarette lighter of an automobile, or to an AC/DC adapter for connection to a 110 volt AC wall power receptacle, for example. The power signal from this external electrical power source is provided to the lithium battery or other storage device internal to the energy harvesting circuit so that the lithium battery or storage device may be charged thereby.

[0016] These and other objects, features and advantages of the present invention will be apparent from the following detailed description of illustrative embodiments thereof, which is to be read in connection with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] FIG. **1** is a block diagram of a circuit for harvesting energy transmitted from an RF (radio frequency) source to charge and/or power an electronic device, such as a cellular telephone or a PDA or a Blackberry™ device, formed in accordance with the present invention.

[0018] FIG. **2** is a schematic diagram of a first preferred form of an electrical circuit constructed in accordance with the present invention for harvesting energy from WIFI signals.

[0019] FIG. **3** (i.e., FIGS. **3**A and **3**B) is a block diagram of a second preferred form of the electrical circuit used in the energy harvesting device of the present invention.

[0020] FIG. **4** (i.e., FIGS. **4**A and **4**B) is a block diagram of third preferred form of the electrical circuit used in the energy harvesting device of the present invention.

[0021] FIG. **5** is a top perspective view of one form of a housing of the energy harvesting device of the present invention, illustrating the cover thereof in an open position.

[0022] FIG. **6** is a top perspective view of the housing of the energy harvesting device of the present invention shown in FIG. **5** taken from a different angle and illustrating the cover thereof in an open position.

[0023] FIG. **7** is a front perspective view of the housing of the energy harvesting device of the present invention shown in FIGS. **5** and **6**, illustrating the cover thereof in an open position.

[0024] FIG. **8** is a front perspective view of the housing of the energy harvesting device of the present invention shown in FIGS. **5-7** taken from a different angle and illustrating the cover thereof in an open position.

[0025] FIG. **9** is a first side perspective view of the housing of the energy harvesting device of the present invention shown in FIGS. **5-8** and illustrating the cover thereof in an open position.

[0026] FIG. **10** is a second side perspective view of the housing of the energy harvesting device of the present invention shown in FIGS. **5-9** and illustrating the cover thereof in an open position.

[0027] FIG. **11** is a perspective view of a portion of the cover of another form of a housing of the energy harvesting device of the present invention, illustrating the cover thereof in an open position and an alternative arrangement of antennas thereon.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0028] The present invention is designed to receive radio frequency signals from any WIFI 2.4 to 2.5 GHz source, WiMax, Satellite Radio, Cellular Phone (850 MHz to 900 MHz and 1.8 GHz-1.9 GHz), and UHF TV signals or CATV signals and convert the received radio frequency signal into another form of energy, such as a DC voltage. Even more preferably, the energy harvesting device of the present invention is capable of receiving RF signals from about 500 MHz to about 2.5 GHz, and convert the received signals into electrical energy. The same circuit configuration can also be designed and extended to convert RF signals in VHF, FM and UHF frequencies (such as signals from CATV) into DC electrical energy. This energy is utilized to recharge a battery inside any cell phone through a USB connector. The invention is designed to use WIFI signals, in particular, for energy harvesting because many WIFI signal transmitters are provided and are available in public places, such as airport terminals, coffee shops, fast food restaurants, etc., where an AC power outlet may not be easily accessible for a user to charge or power an electronic device by connecting it to the outlet.

[0029] There are many types of energy harvesting devices or systems that are known, such as solar cell, windmill, water pressure to provide movement in power generators, FM radio signals to charge small toys or flying devices (for example, spy monitor devices) inductive chargers, and the like. However, no one heretofore has taken advantage of the many WIFI signals, cell phone signals, Satellite Radio or UHF TV signals that are mostly available in public places, and the energy from those transmitted signals is wasted and not utilized. Such WIFI signals, Satellite Radio, UHF TV signals or cell phone signals are broadcast 24 hours a day, every day.

[0030] Most people think that the energy from WIFI or other high frequency radio signals is very minute and cannot be used, but the inventor has found otherwise. Generally, the present invention is directed to harvesting the WIFI, UHF TV, Satellite Radio or cell phone signals and converting the electromagnetic energy in the signals to a DC voltage, and providing a "top charge" to charge the batteries of an external electronic device connected thereto, or to continuously power an internal battery within the energy harvesting device until its storage capacity is full, and that battery energy may be reused when needed to power an electronic device, such as a cellular phone, PDA, MP3 player, camera or other electronic device connected thereto.

[0031] The present invention utilizes WIFI or cell phone signals that are available and easy to find in any public place or in a consumer's home almost anywhere in the world. There are many high powered WIFI broadcasting stations whose transmitted signals create a WIFI hotspot, and also WIMAX signals to be implemented at some future time are projected to be rather strong signals that have a 32 mile radial range, and the energy from these signals may also be harvested.

[0032] Provided above is a general description of the purpose and implementation of the energy harvesting method and apparatus of the present invention. Now, a detailed description of preferred forms of the present invention is provided.

[0033] A block diagram of a circuit for harvesting energy transmitted from an RF (radio frequency) source is shown in FIG. **1** of the drawings. An antenna **2**, such as patch antenna, receives an RF signal, such as a WIFI signal transmitted in public places, such as airport terminals, coffee shops, fast food restaurants and the like, and converts the RF signal to an electrical signal. This electrical signal is provided via a transmission line or signal cable **4** to a circuit **6** which rectifies and harvests the energy of the transmitted RF signal. This circuit **6** converts the RF electrical signal from the antenna **2** to a DC signal of sufficient magnitude to charge and/or power an electronic device **8**, such as a cellular telephone, a PDA or a Blackberry™ device connected to the energy harvesting and rectifier circuit **6** by a USB cable **10** or the like.

[0034] FIG. **2** is a schematic diagram of a first preferred form of an electrical circuit constructed in accordance with present invention for harvesting energy from WIFI or other RF signals. The energy harvesting circuit, in accordance with this first embodiment, includes an antenna **2** which receives RF signals, such as WIFI signals, and converts the RF signals to an electrical signal which is provided by a transmission line or cable **4** to the input of an RF-to-DC rectifier circuit **12**. The rectifier circuit **12** converts the RF signal to a DC (direct current) signal on its output. This DC signal is provided to a first storage device **14**, which is preferably a 2F (farad), 2.5 volt supercapacitor. The first storage device **14** stores and accumulates charge thereon from the rectified (DC) signal outputted by the rectifier circuit **12**.

[0035] The DC signal outputted by the rectifier circuit **12** is also provided to the input (Pin **3**) of a flash-based CMOS microcontroller **16** preferably having Part No. 12F683-0847 manufactured by Microchip Technology Inc. of Chandler, Ariz. The microcontroller **16** is used as a charge circuit control to prevent a lithium battery used as a second storage device **18** in the energy harvesting circuit from overcharging.

[0036] The voltage on the supercapacitor **14** is also provided to the input of a first stage DC-to-DC converter circuit **20**, which acts as a charge pump, or stated another way, a step up voltage regulator and battery management circuit. Preferably, the first stage DC-to-DC converter circuit **20** is Part No. MAX1672 manufactured by Maxim Integrated Products of Sunnyvale, Calif.

[0037] When the voltage on the supercapacitor **14** reaches a certain level, such as about 1.5 volts, which is provided to the input of the microcontroller **16**, the microcontroller **16** generates a pulsed signal on its output (Pin **5**), which is provided to the enable input of the first stage DC-to-DC converter circuit **20** to cause the converter circuit **20**, acting as a charge pump, to increase the voltage on its input from about 1.5 volts to about 4.5 volts on its output. If the voltage on the supercapacitor **14** is below a predetermined level, such as 1.5 volts, the microcontroller **16** does not provide a pulsed signal to the enable input of the first stage DC-to-DC converter circuit **20** to effectively cause the converter circuit **20** from operating, so as not to deplete charge on the supercapacitor **14** if the supercapacitor is insufficiently charged.

[0038] The output of the first stage DC-to-DC converter circuit **20** is provided to the input of a second stage DC-to-DC converter circuit **22**, preferably being the same integrated circuit as the first stage DC-to-DC converter circuit **20** and having Part No. MAX1672. Again, the preferred voltage on the output of the first stage DC-to-DC converter circuit **20** is about 4.5 volts.

[0039] As can be seen from FIG. **2**, the energy harvesting circuit also includes a USB charger port or connector **24** for providing a connection to an external source of voltage (preferably, about 5 volts DC), to charge the internal second storage device **18**, which as will be explained is preferably a lithium polymer battery, using a lap top computer or other device connectable to the USB charger connector **24**. The USB charger connector **24** has its +5 volt input pin connected to the connection between the output of the first stage DC-to-DC converter circuit **20** and the input of the second stage DC-to-DC converter circuit **22**.

[0040] The second stage DC-to-DC converter circuit **22** acts as a low drop out, step down voltage converter/regulator to convert the 4.5 volts from the first stage DC-to-DC converter circuit **20** (or the 5 volt DC signal provided on the USB charger connector **24** from an external source) to a 3.7 volt DC signal on the output of the second DC-to-DC converter circuit **22**. This output voltage of the second stage DC-to-DC converter circuit **22** is provided to the positive side of a lithium polymer battery **18** forming part of the energy harvesting circuit, the negative side of the battery being preferably grounded. Preferably, the lithium polymer battery is a 3.7 volt, 1.5 amp battery which is used as a second storage device **18** which is charged by the rectified and DC-converted voltage generated by the energy harvesting circuit from the RF signal received by the antenna **2**, or by an external DC power source connected to the USB charger connector **24**.

[0041] The voltage on the lithium polymer battery **18** is monitored so as not to exceed approximately 3.7 volts. The positive side of the lithium polymer battery **18** is connected to a feedback input (Pin **1**) on the microcontroller **16** in order to limit the voltage generated by the energy harvesting circuit provided to the lithium polymer battery **18** and to an external electronic device **8** connected to the energy harvesting circuit. When the voltage on the feedback input of the microcontroller **16** reaches a predetermined threshold voltage, such as 3.7 volts, the microcontroller **16** ceases providing a pulsed signal on its output (Pin **5**) connected to the enable input on the first stage DC-to-DC converter circuit **20** so as to prevent the first stage DC-to-DC converter circuit **20** from operating as a charge pump beyond a predetermined voltage conversion limit. When the voltage on the lithium polymer battery **18** falls to below the predetermined threshold voltage, which is provided to the feedback input of the microcontroller **16**, the microcontroller **16** again provides a pulsed signal on its output to the enable input of the first stage DC-to-DC converter circuit **20** so that the converter circuit **20** resumes operation as a charge pump to increase the input voltage provided thereon to approximately 4.5 volts. However, no pulsed signal will be provided to the enable input of the first stage DC-to-DC converter circuit **20** from the microcontroller **16** if the voltage on the supercapacitor **14** is below a predetermined threshold voltage, such as 1.5 volts.

[0042] The positive side of the lithium polymer battery **18** is connected through a user controlled single pole, single throw switch **26** to the input of a third stage DC-to-DC converter circuit **28**, also acting as a charge pump. The third stage DC-to-DC converter circuit **28** is preferably realized using an integrated circuit having Part No. TPS61200DRCT, manufactured by Texas Instruments of Dallas, Tex. The third stage DC-to-DC converter circuit **28** increases the voltage provided by the lithium polymer battery **18** (3.7 volts) to the input of the third stage DC-to-DC converter circuit **28** to approximately 5 volts DC on the output thereof. The output signal of the third stage DC-to-DC converter circuit **28** is provided to a particular pin (+5 volt DC pin) on a USB output connector **30**, having another certain pin grounded. An external electronic or electrical device **8**, such as a cellular phone, PDA, MP3 player, camera or the like, may be connected to the USB output connector **30** of the energy harvesting circuit using a standard USB cable **10** so that the energy harvesting circuit of the present invention may charge the batteries of the external electronic device **8** or power the external electronic device **8** connected to the energy harvesting circuit by the user.

[0043] Referring now to FIG. **3** of the drawings, a block diagram of a second preferred form of the electrical circuit of the energy harvesting device of the present invention is shown. Preferably, a plurality of broadband antennas **2**, each preferably the same, is used. In one form of the present invention, four broadband antennas **2***a*, **2***b*, **2***c*, **2***d* are used, and each broadband antenna **2***a*-**2***d* is particularly situated within the housing **32** for the energy harvesting circuit so that each antenna **2***a*-**2***d* will be capable of receiving WIFI or other RF signals which are transmitted no matter what disposition the housing **32** is in. Preferably, each broadband antenna **2***a*-**2***d* is capable of receiving signals from about 500 MHz to about 2.5 GHz.

[0044] Even more preferably, and as shown in FIGS. **5**-**10** of the drawings, there are four monopole broadband antennas **2***a*-**2***d* which are mounted on the inside surface of a transparent or translucent cover **34** of the housing **32**. The cover **34** is pivotally joined at one edge thereof to a corresponding lateral or top edge of the main body or base **36** of the housing **32**, so that it may be raised and lowered respectively away from and in proximity to the top surface of the main body or base **36**, thus covering and uncovering a solar cell **40** which is exposed through an opening formed in the top surface **38** of the main body or base **36**. Although the four monopole antennas **2***a*-**2***d* will receive RF energy when the cover **34** is disposed in its lower position, covering the solar cell **40** of the housing **32**, the four monopole antennas will more effectively receive RF energy transmitted from an RF source or WIFI source when the cover **34** is disposed in a raised position, such as shown in FIG. **5** of the drawings. The features of the housing **32** of the energy harvesting device of the present invention will be described in greater detail.

[0045] Each antenna **2***a*-**2***d* of the plurality of antennas is coupled to a respective voltage multiplier/full wave rectifier circuit **42**, as shown in FIG. **3**. More specifically, each voltage multiplier/rectifier circuit **42** includes a voltage multiplier so that the WIFI or RF signals received by the respective antenna **2***a*-**2***d* coupled thereto is converted into a DC voltage having an amplitude multiplied a number of times, such as by a factor of 4, 6 or 8, for example.

[0046] More specifically, the first antenna **2***a* is coupled to the first multiplier/rectifier circuit **42***a*, which converts the WIFI or RF signals the first antenna **2***a* receives into a DC voltage multiplied by a factor of 4, 6, or 8, for example. The DC output voltage of the first multiplier/rectifier circuit **42***a*, as a result of the detected signal of the first antenna **2***a* coupled thereto, is provided to node A and a first leg of a capacitor C**39** (see FIG. **3**).

[0047] The second multiplier/rectifier circuit **42***b* is very similar to the first multiplier/rectifier circuit **42***a*. The second multiplier/rectifier circuit **42***b* converts to a DC voltage the WIFI or RF signals received by the second antenna **2***b* coupled to it, multiplied by a desired factor (e.g., 4, 6 or 8), and this voltage is provided to node B and a first leg of a capacitor C**34**, whose second leg is connected to node A and the first leg of capacitor C**39**. Thus, the output voltage of the second multiplier/rectifier circuit **42***b* is effectively stacked on top of the voltage outputted by the first multiplier/rectifier circuit **42***a*.

[0048] The third multiplier/rectifier circuit **42***c* is similar to the second multiplier/rectifier circuit **42***b*, and it is coupled to the third antenna **2***c* in the same manner that the first and second multiplier/rectifier circuits **42***a*, **42***b* are respectively coupled to the first and second antennas **2***a*, **2***b*. Again, the DC voltage outputted by the third multiplier/rectifier circuit **42***c*, which is provided to node C and the first leg of a capacitor C**21**, whose second leg is connected to node B and the first leg of capacitor C**34**, is stacked on top of the DC voltages provided by the first and second multiplier/rectifier circuits **42***a*, **42***b*.

[0049] The fourth multiplier/rectifier circuit **42***d* is similar to the third and second multiplier/rectifier circuits **42***c*, **42***b* described previously, and is coupled to the fourth antenna **2***d*. The DC voltage outputted by the fourth multiplier/rectifier circuit **42***d* is provided to node D and the first leg of a capacitor C**12**, whose second leg is connected to node C and the first leg of capacitor C**21**. As a result of RF signals received by the fourth antenna **2***d* coupled to the fourth multiplier/rectifier circuit **42***d*, the DC output voltage of the fourth multiplier/rectifier circuit **42***d* is stacked on top of the DC voltages outputted by the third, second and first multiplier/rectifier circuits **42***c*, **42***b*, **42***a*. Thus, the potential at the first leg of capacitor C**12**, measured with respect to the second leg of capacitor C**39**, could be as much as, or possibly more than, 25 volts DC as a result of the stacked arrangement of output voltages from the first, second, third and fourth multiplier/rectifier circuits **42***a*-**42***d*. As will be explained in greater detail, the second leg of capacitor C**39** is coupled to a "jump start" battery BT**2** (also referred to as a jump start voltage device **44**, and the stacked output voltages from the four multiplier/rectifier circuits **42***a*-**42***d* are raised by the DC voltage provided by battery BT**2**.

[0050] The combined stacked voltage generated by the interconnected multiplier/rectifier circuits **42***a*-**42***d* is provided to a capacitor C**3**. Capacitor C**3** acts as a temporary storage device **14** for the voltages detected and rectified by the four multiplier/rectifier circuits **42***a*-**42***d* as a result of RF energy received by the four receiving antennas **2***a*-**2***d*. More specifically, the first leg of capacitor C**3** is connected to the first leg of capacitor C**12** and the second leg of capacitor C**3** is grounded. Preferably, capacitor C**3** is a low leakage tantalum capacitor, although other storage devices and types of capacitors may be used.

[0051] As mentioned previously, capacitor C**3** is provided for the temporary storage of the stacked DC voltage outputted by the four multiplier/rectifier circuits **42***a*-**42***d*, as a result of RF or WIFI energy received by the four antennas **2***a*-**2***d* connected thereto, as well as the DC voltage of the jump start battery BT**2** which raises the stacked voltage at node D.

[0052] The circuit of the present invention also preferably harvests light energy by using a solar cell **40** and circuit. In this regard, reference should continue to be had to FIG. **3** of the drawings. A solar cell **40** is provided in the energy harvesting device of the present invention and is exposed through an opening in the top surface **38** of the main body or base **36** of the housing **32** of the energy harvesting device. The solar cell **40** can generate between about 3.1 volts with indoor lighting and about 5.5 volts in direct sunlight. The negative terminal of the solar cell **40** is preferably grounded, and the positive terminal of the solar cell **40** is connected to the cathode of a zener diode D**2**, whose anode is connected to ground. Preferably, zener diode D**2** is a 5.6 volt zener diode to limit the output voltage provided by the solar cell **40** so as not damage any circuitry connected thereto.

[0053] The output of the solar cell **40** is connected to the VDD input of a power management integrated circuit **46**, such as Part No. CBC3112 manufactured by Cymbet Corporation of Elk River, Minn., or any similar power management circuit made by a number of manufacturers. The output voltage of the power management circuit **46**, which is regulated and selected to be about 4.2 volts, is provided through an isolation diode D**3** to the positive side of the main energy storage battery BT**1** (i.e., second or main storage device **18**) to help maintain the charge on battery BT**1** when no or negligible WIFI or RF signals are present or can be detected and when ambient light conditions permit the charging of battery BT**1** by the solar cell **40**.

[0054] To help "jump start" or boost the charge on the temporary storage capacitor C**3**, a 3.7 volt, 60 milliamperehour or greater, lithium polymer battery BT**2** is used (see FIG. **3**). The negative side of battery BT**2** is grounded, while the positive side is coupled to the second leg of capacitor C**39** in order to raise the stacked voltage generated by the four interconnected multiplier/rectifier circuits **42***a*-**42***d* by 3.7 volts, i.e., the voltage of battery BT**2**. Accordingly, the voltage across capacitor C**3** is the voltage resulting from the combination of the voltage from battery BT**2** and the RF or WIFI energy harvested by the circuit of the present invention. It should be noted that the 25 volt DC signal resulting from the stacked arrangement of the four multiplier/rectifier circuits **42***a*-**42***d* is an unloaded voltage, and when used under load to charge capacitor C**3**, the voltage drops significantly.

[0055] The voltage from battery BT**2** is used to boost the voltage temporarily stored on **42***d* capacitor C**3** and to supplement the voltage outputted by the four multiplier/rectifier circuits **42***a*-**42***d* resulting from the WIFI and RF energy which is harvested. Furthermore, the "boost" voltage from battery BT**2** helps insure that a second power management circuit **48** (see FIG. **3**) has a sufficient voltage on its VDD input in order to operate.

[0056] It is estimated that a 60 milliampere-hour battery used as battery BT**2** will last for approximately 1500 hours, or about 62 days, when the energy harvesting device of the present invention is stored in a dark room, thus with no voltage being generated by the solar cell **40**, as the circuit draws only about 25 microamperes to about 40 microamperes of current from battery BT**2**.

[0057] To maintain the charge on battery BT**2**, the output voltage generated by the power management circuit **46** is provided to the anode of an isolation diode D**50**, whose cathode is connected to the positive side of battery BT**2**. In this way, the power management circuit **46** acts as a buffer between the solar cell and battery BT**2** to prevent battery BT**2** from overcharging, which may result in the battery overheating.

[0058] Alternatively, to charge battery BT**2**, the positive lead of the solar cell **40** may be coupled to the positive side of battery BT**2** through a diode (not shown), the anode of the diode being connected to the positive lead of the solar cell **40**, and the cathode of the diode being connected to the positive side of battery BT**2**. Optionally, a PTC (positive temperature coefficient) thermistor (not shown) may be connected to battery BT**2** between the solar cell **40** and the battery BT**2** to provide some protection to the circuit in the remote chance that battery BT**2**, which is preferably a lithium polymer battery, overheats due to overcharging or short circuiting.

[0059] Referring again to FIG. **3** of the drawings, it will be seen that a second power management circuit **48** is provided, and is indirectly coupled to the temporary storage capacitor C**3**. Preferably, a watchdog monitor circuit **50** is disposed between capacitor C**3** and the VDD input of the power management circuit **48**, as will be explained in greater detail. Power management circuit **48** may be Part No. CBC3112 manufactured by Cymbet Corporation of Elk River, Minn., or any similar power management circuit made by a number of manufacturers. The VBAT output of the power management circuit **48** is connected to the anode of diode D**1**, as will be explained in greater detail.

[0060] Since the power management circuit **48**, like the power management circuit **46**, requires about 2.6 volts on its VDD input in order to operate, the energy harvesting circuit of the present invention preferably includes a watchdog monitor circuit **50** situated between the temporary storage capacitor C**3** and the power management circuit **48** in order to minimize the unnecessary discharge of energy stored in capacitor C**3** when there is insufficient voltage on capacitor C**3** for the power management circuit **48** to operate. The watchdog monitor circuit **50** may include a voltage detector circuit **52** and a transistor switch circuit **54** connected thereto and controlled thereby. The voltage detector circuit **52** may be a constant voltage power monitor with hysteresis and having Part No. S-80825 manufactured by Seiko Instruments, Inc. of Chiba, Japan.

[0061] The watchdog monitor circuit **50** comprising the voltage detector circuit **52** and the transistor switch circuit **54** is used to minimize any leakage from the temporary storage capacitor C**3** by preventing a direct connection between capacitor C**3** and the VDD input of the second power management circuit **48** when the voltage on the temporary storage capacitor C**3** is below a predetermined voltage. More specifically, when the voltage is below about 2.6 volts on capacitor C**3**, the voltage detector circuit **52** of the watchdog monitor circuit **50** provides an output signal to the transistor switch circuit **54** to open the transistor switch circuit **54** and prevent the voltage on capacitor C**3** from being connected therethrough to the VDD input of the second power management circuit **48** so that circuit **48** does not drain the energy from the WIFI or RF energy harvesting circuit stored on capacitor C**3**. However, if the voltage on capacitor C**3** rises above about 2.6 volts, the voltage detector circuit **52** of the watchdog monitor circuit **50** outputs a signal to close the transistor switch circuit **54** so that the voltage on capacitor C**3** is provided to the VDD input of the second power management circuit **48**. Typically, the voltage on the temporary storage capacitor C**3** will range between about 2.5 volts and about 2.65 volts, and only when this voltage is equal to or above about 2.6 volts will capacitor C**3** be effectively connected to the input of the second power management circuit **48**. A general form of the watchdog monitor circuit **50** described above may be found in the application notes accompanying the data sheet of the S-808xxC series of voltage detectors disseminated by Seiko Instruments Inc. (see FIGS. **14** and **20** thereof).

[0062] The power management circuit **48** has a charge pump which converts the 2.6-2.65 voltage (or, more specifically, any voltage above 2.63 volts) on its input to between about 4.1 volts DC and about 4.2 volts DC on its VBAT output. The VBAT output of the second power management circuit **48** is connected to the anode of a low voltage drop (about 0.1 volts) isolation diode D**1**. The cathode of diode D**1** is connected to the positive side of a 3.7 volt, 1000 milliampere-hour, lithium polymer battery BT**1**, which acts as the main energy storage battery or device **18**.

[0063] Preferably, a battery protection circuit **56** is connected to the battery BT**1** to provide overcharge protection and over-discharge protection. The battery protection circuit **56** is preferably Part No. DW01 manufactured by Fortune Semiconductor Corporation of Taipei, Taiwan, which monitors the voltage on the positive side of battery BT**1**, and drives a transistor switch circuit **58** situated between the negative side of the battery and ground to selectively connect and disconnect the negative side of battery BT**1** to and from ground.

[0064] When the voltage on the lithium battery BT**1** exceeds a certain predetermined overcharge protection voltage, charging of battery BT**1** is inhibited by the battery protection circuit **56** opening the transistor switch circuit **58** to disconnect the battery BT**1** from ground. On the other hand, when the voltage of the lithium battery BT**1** falls below a predetermined over-discharge protection voltage, discharging of battery BT**1** is inhibited by the battery protection circuit **56** opening the transistor switch circuit **58** to disconnect the battery BT**1** from ground. Essentially, the negative side of battery BT**1** is disconnected from ground when either an overcharge condition or an over-discharge condition exists. When neither condition exists, the battery protection circuit **56** closes the transistor switch circuit **58** to connect the negative side of the battery BT**1** to ground.

[0065] The energy harvesting circuit of the present invention also preferably includes a battery charge monitoring circuit **60** to provide an indication to the user of the energy harvesting device whether the lithium battery BT**1** is sufficiently charged to power an external electronic device **8** connected to the energy harvesting circuit or to charge the battery or batteries of the external device. Preferably, the battery charge monitoring circuit **60** includes a voltage divider network (having interconnected resistors) (not shown) operatively coupled to the positive side of the battery BT**1** and used to bias a transistor (not shown) on or off, depending on the voltage of the battery. A single pole, single throw (or equivalent) momentary push button switch SW**2** (or switch circuit) is connected between the positive side of the lithium battery BT**1** and the battery charge monitoring circuit **60**, and is preferably mounted on the top surface **38** of the housing **32** so as to be accessible to the user of the energy harvesting device. Accordingly, so as not to drain the charge from battery BT**1**, the battery charge monitoring circuit **60** will only work when the user presses on momentary push button switch SW**2** to provide power to the battery charge monitoring circuit **60**.

[0066] A dual color (red/green) light emitting diode (LED) or separate red and green LEDs **62**, **64** are operatively coupled to and driven by the battery charge monitoring circuit **60**, and are preferably mounted on the top surface **38** of the housing **32**, or elsewhere on the housing, to be viewable by the user of the energy harvesting device.

[0067] When the voltage on battery BT**1** is 3.3 volts or greater, the battery charge monitoring circuit **60** will cause the green LED **64** to light when the user presses on momentary push button switch SW**2**. If the voltage on battery BT**1** is 3.0 volts or lower, the battery charge monitoring circuit **60** will cause the red LED **62** to illuminate when the user presses on momentary push button switch SW**2**. The 3.3 volt threshold and the 3.0 volt threshold may be adjusted by changing the values of resistors from which the voltage divider network is formed to turn the transistor on or off at different voltage levels.

[0068] As shown in FIG. **3**, a single pole, single throw switch or switch circuit SW**1** has one side connected to the positive side of the lithium polymer battery BT**1** or other storage device. The switch or switch circuit SW**1** is provided to ensure that stored energy is not drawn from the battery until desired by the user. As will be explained in greater detail, the switch SW**1** is coupled to a spring biased, sliding cover **66** on the housing **32** for the energy harvesting circuit (see FIG. **10**) that selectively covers an output USB type A connector **30**, and will be automatically activated to provide electrical conduction therethrough when the user slides the USB connector cover **66** in one direction on the main body or base **36** of the housing **32**.

[0069] The other side of the switch or switch circuit SW**1** is connected to the input of a DC/DC boost converter circuit **28**, which may be an integrated circuit having Part No. TPS61200DRCT, manufactured by Texas Instruments of Dallas, Tex. This DC/DC converter circuit **28** is configured with external components to convert the 3.7 volts of the lithium polymer battery BT**1** to a 5 volt DC output.

[0070] A 5 volt DC regulated voltage is preferably provided by the DC/DC boost converter circuit **28** to the USB connector **30** of the energy harvesting circuit to which an external electronic device **8** is connectable. Preferably, pin **1** (VCC) of the output connector **30** provides a 5 volt DC output signal to the external electronic device **8**. As shown in FIG. **10** of the drawings, USB output connector or port **30** is preferably positioned on one of the narrower sides **68** of the housing **32** of the energy harvesting device.

[0071] Thus, the circuit configuration described above will allow the output of this USB type A connector **28** to be able to provide the correct charge voltage and current for iPhone™, iPod™ and iPad™ devices, which are manufactured by Apple, Inc. Also, this circuit configuration will be compatible and provide correct charge voltage and current to most if not all smart phones, PDAs, MP3 players and Blackberry™ devices.

[0072] As mentioned previously, the energy harvesting device of the present invention may include an input connector **24** for receiving a charging voltage from an external power source, such as another electronic device, or a power transformer which may be connected to a cigarette lighter of an automobile, or may be an AC/DC adapter which is connectable to a standard house wall outlet, for example. The power signal from this external electrical power source is provided to the anode of an isolation diode D**4**, whose cathode is in electrical communication with the positive side of the lithium battery BT**1** or other storage device internal to the energy harvesting circuit so that the lithium battery or storage device may be charged thereby.

[0073] More specifically, and as shown in FIGS. **3** and **9** of the drawings, a mini USB connector **24** is provided on one side (preferably on one of the narrower sides **68** not occupied by output USB connector **30**) of the housing **32** of the energy harvesting device. Input connector **24** is provided so that the user may connect to connector **24** an AC/DC wall outlet adapter, or another adapter or transformer which is connectable to an auxiliary power outlet or cigarette lighter in an automobile, or to an output port providing a 5 volt DC output signal of an electronic device, such as a laptop computer. Preferably, pin **1** (VCC) of connector **24** provides a 5 volt DC input signal to the circuit.

[0074] The 5 volt DC input signal on connector **24** is provided to the anode of isolation diode D**4**. The cathode of diode D**4** is electrically connected directly or through a resistor (not show) to the positive side of the lithium battery BT**1** so that the lithium battery of the energy harvesting circuit may be charged externally by the AC/DC adapter, power transformer, automobile power outlet, laptop computer or the like connected to the mini USB connector **24**.

[0075] Also, with an external power source connected to the input USB connector **24**, and with switch SW**1** in a closed position, the external power source will provide power to the output USB connector **30** and to an external electronic device **8** coupled thereto.

[0076] Referring now to FIGS. **5**-**10** of the drawings, it will be seen that a preferred form of the housing **32** for the energy harvesting circuit includes a main body or base **36** which is rectangular in shape overall, and a transparent or translucent cover **34** that is pivotally joined to a longer edge (or, alternatively, a narrower edge) of the top surface **38** of the main body **36**. The solar cell **40** is mounted on the top surface **38** of the main body **36** or through a window formed through the thickness of the main body so as to be exposed to ambient light. The transparent cover **34** is pivotable on the top surface **38** between a closed position, where it covers the top surface **38**

or at least the solar cell **40** mounted on the top surface, and an open position, where it is raised from the top surface **38** of the main body **36** to a position which is substantially perpendicular to the plane in which the top surface **38** of the main body resides. Thus, if the bottom surface **70**, which is opposite the top surface **38**, of the main body **36** rests on a horizontal supporting surface, such as a table or desk, then the transparent cover **34** may be pivoted to a substantially vertical position.

[0077] As mentioned previously, there are preferably four antennas **2***a*-**2***d* which are spaced apart from each other a predetermined distance, which are mounted to the transparent cover **34** on the inside surface thereof. For best WIFI or RF signal reception, it is preferred that these broadband antennas **2***a*-**2***d*, preferably being monopole antennas, are disposed in a vertical position. As such, it is preferred if the user of the energy harvesting device of the present invention raises the transparent cover **34** to a vertical position when he chooses to charge the internal battery BT**1** of the energy harvesting circuit with WIFI or RF signals present in the location where the energy harvesting device is situated.

[0078] Preferably, the cover **34** is transparent or translucent, as mentioned previously, so that the solar cell **40** of the energy harvesting circuit will receive light energy through the cover **34** to charge the internal lithium battery BT**1**, even if the cover **34** is in a closed position.

[0079] Also, as mentioned previously, so as not to discharge the internal lithium battery BT**1** unnecessarily when no external electronic device **8** is being powered by the energy harvesting circuit, a sliding cover **66** is provided over the output USB connector **30**, and the switch SW**1** is operatively coupled to, and activated by, movement of the USB connector sliding cover **66**.

[0080] More specifically, the USB output connector **30** is preferably situated on a smaller side **68** of the main body **36** of the housing **32** opposite that of the input USB connector **24** (although the connectors **30** and **24** could be mounted on opposite wider sides **72** of the housing **32**). There is a small cover **66** for the output USB connector **30** so that a user may slide this cover **66** and connect the output USB connector **30** to an external electronic device **8** in order to power or recharge the batteries of the external electronic device. When the sliding cover **66** is positioned such that it uncovers the output USB connector **30**, it automatically engages and activates the single pole, single throw switch or switch circuit SW**1** so that the battery BT**1** or other storage device is electrically connected to the input of the DC/DC boost converter **28** to provide a 5 volt output voltage on the output USB connector **30**.

[0081] When the sliding cover **66** is returned to its closed position, covering the output USB connector **30**, it deactivates the switch or switch circuit SW**1**, breaking the connection between the lithium battery BT**1** or other storage device within the energy harvesting circuit and the DC/DC converter circuit **28**, so that no power is provided to the output USB connector **30**. The sliding cover **66**, in this closed position, covers the output USB connector **30** to prevent access thereto. In a preferred form of the present invention, the sliding cover **30** may be biased by a spring (not shown) so that it automatically returns to a closed position, covering the output connector **30** and deactivating switch SW**1**, when no external electronic device **8** is connected to output connect or connector **30** of the energy harvesting device.

[0082] The purpose of having the sliding cover **66**, covering and uncovering the output USB connector **30**, is to minimize any energy drain from the lithium battery BT**1** when there is no need to provide power to an external electronic device **8**. It should be realized that the user controlled switch **26** mounted on the top surface **38** of the base **36** may be omitted if the sliding cover **66** coupled to switch SW**1** to activate switch SW**1** is used. Alternatively, the switch labeled with reference number **26** shown in FIGS. **5**-**10** may instead be used as switch SW**2**, which may be a momentary push button switch, which the user presses to momentarily activate the battery charge monitoring circuit **60** to illuminate the red LED **62** or the green LED **64** so as to minimize any drain on the lithium battery BT**1**.

[0083] In one form of the present invention, the four antennas **2***a*-**2***d* are identical to one another and have the same or similar broadband bandwidth characteristics. In an alternative form of the present invention, the energy harvesting circuit may include several narrow band antennas **2**, such as one UHF antenna to receive signals in the range from between about 500 MHz and about 700 MHz, one cell phone and 900 MHz antenna to receive signals in the range from between about 850 MHz and about 928 MHz, one GSM band antenna to receive signals in the range from between about 1.8 GHz and about 1.9 GHz, and one WiFi/WiMAx/Satellite radio antenna to receive signals in the range from between about 2.3 GHz and about 2.5 GHz, each antenna **2** being connected to a respective multiplier/rectifier circuit **42** through a coaxial cable **4** (see FIG. **9**).

[0084] Reference should now be had to FIGS. **4** and **11** of the drawings, which show another (third) preferred embodiment of the energy harvesting device of the present invention. The circuit shown in the schematic diagram of FIG. **4** is similar in many respects to the circuit shown in FIG. **3**. However, in this preferred form, there are six (6) antennas **2***a*-**2***f* provided (also as shown in FIG. **10**) and six (6) multiplier/rectifier circuits **42***a*-**42***f*. Therefore, there are nodes A-F associated with the outputs of the first multiplier/rectifier circuit **42***a* through the sixth multiplier/rectifier circuit **42***f*, respectively.

[0085] In this alternative embodiment of the circuit shown in FIG. **4**, the output voltages from the first multiplier/rectifier circuit **42***a* and the second multiplier/rectifier circuit **42***b* are stacked together in series on top of the voltage of the jump start battery BT**2**, and the stacked voltage of the first multiplier/rectifier circuit **42***a* and the second multiplier/rectifier circuit **42***b* is provided at node B and to the temporary storage capacitor C**3** through a diode D**60**, whose anode is connected to node B and whose cathode is connected to temporary storage capacitor C**3**.

[0086] Similarly, the output voltages of the third multiplier/rectifier circuit **42***c* and the fourth multiplier/rectifier circuit **42***d* are stacked in series on the voltage of the jump start battery BT**2**. The stacked voltage of the fourth multiplier/rectifier circuit **42***d* and the third multiplier/rectifier circuit **42***c* is provided at node D and to the temporary storage capacitor C**3** through diode D**62**, whose anode is connected to node D and whose cathode is connected to the temporary storage capacitor C**3**.

[0087] In the same manner, the output voltage of the fifth multiplier/rectifier circuit **42***e* and the output voltage of the sixth multiplier/rectifier circuit **42***f*, whose inputs are respectively connected to the fifth antenna **2***e* and the sixth antenna **2***f*, are stacked in series on the voltage of the jump start battery BT**2**. The stacked voltage of the sixth multiplier/rectifier circuit **42***f* and the fifth multiplier/rectifier circuit **42***e* is provided at node F (node E is the output voltage at the fifth multiplier/rectifier circuit **42***e*) and is provided to the temporary storage capacitor C**3** through diode D**64**, whose anode is connected to node F and whose cathode is connected to the temporary storage capacitor C**3**.

[0088] More specifically, the positive side of the jump start battery BT**2** is connected to a first leg of each of capacitors C**39**, C**21** and C**100**. The second leg of each of capacitors C**39**, C**21** and C**100** is respectively connected to node A and capacitor C**34** (and the output of the first multiplier/rectifier circuit **42***a*); node C and capacitor C**12** (and the output of the third multiplier/rectifier circuit **42***c*); and node E and the first leg of capacitor C**102** (and the output of the fifth multiplier/rectifier circuit **42***e*). The second leg of capacitor C**102** is coupled to node F, the anode of diode D**64** and the output of the sixth multiplier/rectifier circuit **42***f*.

[0089] In other words, the outputs of the first and second multiplier/rectifier circuits **42***a*, **42***b* are connected in series, the outputs of the third and fourth multiplier/rectifier circuits **42***c*, **42***d* are connected in series, and the outputs of the fifth and sixth multiplier/rectifier circuits **42***e*, **42***f* are connected in series, to form three groups, and the output voltages of the series-connected groups are effectively connected in parallel with each other and provided to the temporary storage capacitor C**3**. This provides approximately 8-12 volts to charge the temporary storage capacitor C**3**, but at greater current capacity (because of the parallel arrangement of multiplier/rectifier groups) than having the outputs of the individual multiplier/rectifier circuits **42***a*-**42***f* all connected in series in a stack.

[0090] It should be realized that a greater or fewer number of parallel-connected groups of series-connected multiplier/rectifier circuits **42***a*-**42***f* may be used to generate the voltage to charge the temporary storage capacitor C**3**. For example, the third group consisting of the fifth and sixth multiplier/rectifier circuits **42***e*, **42***f*, and the corresponding fifth and sixth antennas **2***e*, **2***f* to which the fifth and sixth multiplier/rectifier circuits **42***e*, **42***f* are respectively connected, may be omitted to satisfy space constraints in the arrangement of the antennas **2** on the housing **32** of the energy harvesting device and to provide sufficient separation between adjacent antennas **2** in order to minimize any cross-coupling of the antennas.

[0091] This alternative circuit of the energy harvesting device of the present invention also includes a 6 volt DC, 2 amp (or the like) input connector **74**, which allows the main energy storage battery BT**1** to be charged up to approximately 75 percent of its capacity in about 30 minutes. As can be seen from FIG. **4** of the drawings, the output voltage on the quick charge connector **74** is provided to the main energy storage battery BT**1** through diode D**68**, whose anode is operatively connected to the output of connector **74** and whose cathode is operatively connected to the main energy storage battery BT**1**. Also, the output voltage on quick charge connector **74** is provided to the jump start battery BT**2** to charge that battery as well, through diode D**66**, whose anode is operatively connected to the output of the connector **74** and whose cathode is operatively connected to the positive side of jump start battery BT**2**.

[0092] Furthermore, the 5 volt DC connector **24** provides a charging voltage to the main energy storage battery BT**1** through diode D**4**, as mentioned previously with the circuit shown in FIG. **3**, but also provides a charging voltage to the jump start battery BT**2** through diode D**70**, whose anode is operatively connected to the output of connector **24** and whose cathode is operatively connected to the positive side of jump start battery BT**2**.

[0093] Similarly, the output voltage of the power management circuit **46**, which is derived from the solar cell voltage, is provided through diode D**3** to the main energy storage battery BT**1** and through diode D**50** to the positive side of jump start battery BT**2**, in the same manner as shown in the schematic diagram for the circuit shown in FIG. **3**.

[0094] With reference to FIG. **11** of the drawings, it can be seen that the six (6) antennas **2***a*-**2***f* are arranged side-by-side on the cover **34** of the housing **32**, which may be raised from the base **36** of the housing vertically so as to optimize the reception capabilities of the antennas. Although in FIG. **11** the six antennas **2***a*-**2***f* are shown mounted on the cover **34** transversely along the narrower side of the cover, which is hingedly joined to the base **36** of the housing **32**, it is envisioned to be within the scope of the present invention to arrange the six antennas **2***a*-**2***f* along the longer side of the cover **34** and having the longer side of the cover hingedly mounted to the longer side **72** of the base **36** of the housing **32** to provide a greater distance between adjacent antennas **2***a*-**2***f* and to minimize or eliminate any possible coupling between adjacent antennas.

[0095] The energy harvesting circuit of the present invention, in its preferred form, detects WIFI signals, WiMax, Satellite Radio, Cellular Phone (850 MHz to 900 MHz and 1.8 GHz-1.9 GHz), any 2.4 GHz and UHF TV signals and light to harvest energy therefrom to charge an internal battery BT**1** or other storage device for use on demand to power or recharge the batteries of an external electronic device **8** connected thereto. When not in use, the energy which is detected and harvested is stored in the battery BT**1** or other storage device so that the battery or other storage device is fully charged and ready to power an external electronic device **8** when connected thereto by the user.

[0096] It should be further realized that the energy harvesting device or at least portions of the electrical circuit thereof may be incorporated directly into a cellular phone, PDA, camera, Blackberry™ device or other electronic device **8** to power the device or charge the batteries thereof, without using a stand alone energy harvesting device within a separate housing **32**, such as shown in FIGS. **5**-**11**. The antenna or antennas **2**, and the solar cell **40**, of the energy harvesting device may be mounted on or within the housing of the electronic device to receive transmitted RF signals and light energy, respectively, in order to power the electronic device **8** or charge the batteries thereof. It is also envisioned to be within the scope of the present invention to incorporate the electrical circuit of the energy harvesting device directly on or in the battery used to power the electronic device **8**.

[0097] The various embodiments of the energy harvesting device of the present invention will now further be described in detail.

[0098] In one form of the present invention, an energy harvesting device preferably includes an antenna **2** for receiving a transmitted radio frequency (RF) signal, a rectifier circuit **6** operatively coupled to the antenna **2**, the rectifier circuit **6** converting the RF signal received by the antenna **2** to a direct current (DC) voltage, and an output connector **30** operatively coupled to the rectifier circuit **6** and providing the DC voltage thereon, an electronic device **8** being connectable to the output connector **30** for receiving the DC voltage provided on the output connector **30**.

[0099] In another form of the present invention, an energy harvesting device preferably includes an antenna **2** for receiving a transmitted radio frequency (RF) signal, the antenna **2** converting the received RF signal to an RF electrical signal, an RF-to-DC rectifier circuit **12**, the RF-to-DC rectifier circuit **12** being operatively connected to the antenna **2** and converting the RF electrical signal to a direct current (DC) voltage, and a first storage device **14**, the first storage device **14** being operatively connected to the RF-to-DC rectifier circuit **12** and being provided with the DC voltage and being charged thereby. The energy harvesting device preferably further includes a microcontroller **16**, the microcontroller **16** having a first signal input operatively connected to the first storage device **14** to monitor the DC voltage on the first storage device **14**, the microcontroller **16** generating a pulsed output signal if the DC voltage on the first storage device **14** is above a first pre-determined voltage level, a first stage DC-to-DC voltage converter circuit **20**, the first stage DC-to-DC voltage converter circuit **20** being operatively connected to the first storage device **14** and generating an output DC voltage which is greater than the DC voltage on the first storage device **14**, and a second stage DC-to-DC voltage converter circuit **22**, the second stage DC-to-DC voltage converter circuit **22** being operatively connected to the first stage DC-to-DC voltage converter circuit **20** and generating an output DC voltage which is less than the output DC voltage generated by the first stage DC-to-DC voltage converter circuit **20**. The energy harvesting device preferably further includes a second storage device **18**, the second storage device **18** being operatively connected to the second stage DC-to-DC voltage converter circuit **22** and being provided with the output DC voltage generated by the second stage DC-to-DC voltage converter circuit **22** and being charged thereby, a third stage DC-to-DC voltage converter circuit **28**, the third stage DC-to-DC voltage converter circuit **28** being operatively connected to the second storage device **18** and generating an output DC voltage which is greater than the DC voltage on the second storage device **18**, and an output connector **30**, the output connector **30** being operatively connected to the third stage DC-to-DC voltage converter circuit **28** and providing the output DC voltage generated by the third stage DC-to-DC voltage converter circuit **28** thereon. An electronic device **8** is connectable to the output connector **30** for receiving thereon the output DC voltage provided to the output connector **30**.

[0100] The energy harvesting device described above may further include an input connector **24**. The input connector **24** is operatively connected to the second stage DC-to-DC voltage converter circuit **22**. An external source of DC voltage is connectable to the input connector **24** for providing the DC voltage from the external source to the second stage DC-to-DC voltage converter circuit **22**. The second stage DC-to-DC voltage converter circuit **22** generates the output DC voltage of the second stage DC-to-DC voltage converter circuit **22** in response to the DC voltage from the external source provided on the input connector **24**.

[0101] In a preferred form of the invention, the first storage device **14** of the energy harvesting device includes a capacitor. In a preferred form of the invention, the second storage device **18** of the energy harvesting device includes a battery.

[0102] Preferably, the microcontroller **16** includes a second signal input operatively connected to the second storage device **18** to monitor the DC voltage on the second storage device **18**. The microcontroller **16** does not generate the pulsed output signal if the DC voltage on the second storage device **18** is above a second predetermined voltage level.

[0103] Another preferred form of the energy harvesting device is described as follows. The energy harvesting device includes a plurality of antennas **2** for receiving a transmitted radio frequency (RF) signal. The antennas **2** generate RF electrical signals in response to the transmitted RF signals received thereby. A plurality of multiplier and rectifier circuits **42** is also included. Each multiplier and rectifier circuit **42** is responsive to a respective RF electrical signal generated by an antenna **2** of the plurality of antennas and generates a direct current (DC) voltage in response thereto. At least some of the DC voltages are stacked in series to provide a combined DC voltage, the combined DC voltage being greater in magnitude than each individual DC voltage generated by a respective multiplier and rectifier circuit **42**.

[0104] The energy harvesting device of this preferred embodiment further includes a temporary storage device **14**. The temporary storage device **14** is responsive to the combined DC voltage and provides a temporarily stored DC voltage in response thereto. The temporary storage device **14** may be or include a capacitor. A first power management circuit **46** is also preferably included, the first power management circuit **46** being responsive to the temporarily stored DC voltage and generating a DC charging voltage in response thereto.

[0105] The energy harvesting device further preferably includes a main energy storage device **18**. The main energy storage device **18** is responsive to the DC charging voltage generated by the first power management circuit **46** and provides a main energy storage DC voltage in response thereto. The main energy storage device **18** may be or include a battery, such as a lithium polymer battery. An output connector **30** is also preferably included in the device, the output connector **30** providing an output connector DC voltage thereon in response to the main energy storage DC voltage. An electronic device **8** is connectable to the output connector **30** for receiving thereon the output connector DC voltage provided on the output connector **30**.

[0106] In another preferred form of the present invention, the energy harvesting device may further include a jump start DC voltage device **44**. The jump start DC voltage device **44** provides a jump start DC voltage. The combined DC voltage includes the at least some of the DC voltages generated by the plurality of multiplier and rectifier circuits **42** stacked in series and the jump start DC voltage, if such a jump start DC voltage device **44** is included. The jump start DC voltage device **44** may be or include a battery.

[0107] An input connector **24** may be included in the energy harvesting device described above. The input connector **24** is connectable to a source of DC voltage external to the energy harvesting device, the DC voltage of the external source being provided to at least one of the main energy storage device **18**, the jump start DC voltage device **44** and the output connector **30**.

[0108] The energy harvesting device described above may also include a DC-to-DC voltage converter circuit **28**. The DC-to-DC voltage converter circuit **28** is responsive to the main energy storage DC voltage and generates an output DC voltage in response thereto. The output connector **30** provides the output connector DC voltage in response to the output DC voltage generated by the DC-to-DC voltage converter circuit **28**, if such is included.

[0109] The energy harvesting device also preferably includes a watchdog monitor circuit **50**. The watchdog monitor circuit **50** is responsive to the temporarily stored DC voltage and selectively provides the temporarily stored DC voltage to the first power management circuit **46** if the temporarily stored DC voltage is greater than a predetermined voltage threshold.

[0110] In a more preferred form of the present invention, the energy harvesting device includes a solar cell circuit having a solar cell **40**. The solar cell circuit receives light energy and generates a solar cell DC output voltage in response thereto. The solar cell DC output voltage is provided to at least one of the main energy storage device **18**, the jump start DC voltage device **44** and the output connector **30**. Furthermore, a second power management circuit **48** may be included for the solar cell circuit. The second power management circuit **48** is responsive to the solar cell DC output voltage and generates a managed solar cell DC output voltage in response thereto. The managed solar cell DC output voltage is provided to at least one of the main energy storage device **18**, the jump start DC voltage device **44** and the output connector **30**.

[0111] It should be noted that at least some of the DC voltages generated by the multiplier and rectifier circuits **42** may be coupled to provide group DC voltages, and the group DC voltages may be combined in parallel to define the combined DC voltage provided to the temporary storage device **14**.

[0112] In a more specific form of the present invention, the plurality of antennas **2** of the energy harvesting device includes a first antenna **2***a*, a second antenna **2***b*, a third antenna **2***c* and a fourth antenna **2***d*, the first antenna **2***a* generating a first RF electrical signal, the second antenna **2***b* generating a second RF electrical signal, the third antenna **2***c* generating a third RF electrical signal, and the fourth antenna **2***d* generating a fourth RF electrical signal. Furthermore, the plurality of multiplier and rectifier circuits **42** of the energy harvesting device includes a first multiplier and rectifier circuit **42***a*, a second multiplier and rectifier circuit **42***b*, a third multiplier and rectifier circuit **42***c* and a fourth multiplier and rectifier circuit **42***d*, the first multiplier and rectifier circuit **42***a* being responsive to the first RF electrical signal generated by the first antenna **2***a* and generating a first multiplied and rectified DC voltage in response thereto, the second multiplier and rectifier circuit **42***b* being responsive to the second RF electrical signal generated by the second antenna **2***b* and generating a second multiplied and rectified DC voltage in response thereto, the third multiplier and rectifier circuit **42***c* being responsive to the third RF electrical signal generated by the third antenna **2***c* and generating a third multiplied and rectified DC voltage in response thereto, and the fourth multiplier and rectifier circuit **42***d* being responsive to the fourth RF electrical signal generated by the fourth antenna **2***d* and generating a fourth multiplied and rectified DC voltage in response thereto. The first multiplied and rectified DC voltage and the second multiplied and rectified DC voltage are preferably coupled in series to define a first group DC voltage. Similarly, the third multiplied and rectified DC voltage and the fourth multiplied and rectified DC voltage are preferably coupled in series to define a second group DC voltage. The first group DC voltage and the second group DC voltage are preferably combined in parallel to define the combined DC voltage which is provided to the temporary storage device **14**.

[0113] Each antenna **2** of the plurality of antennas may have either a broadband bandwidth characteristic or a narrowband bandwidth characteristic, or the antennas **2** may be a mix of broadband and narrowband bandwidth characteristics.

[0114] The energy harvesting device of the present invention preferably includes a housing **32**. The housing **32** has a base **36** defining an internal cavity for containing the electrical circuit of the energy harvesting device, and a cover **34** pivotally mounted on the base **36**. The base **36** has a top surface **38**, and the cover **34** is preferably pivotable between at least a closed position in which the cover **34** is in close proximity to the base **36** to at least partially cover the top surface **38**, and an open position in which the cover **34** is substantially perpendicular to the top surface **38** of the base.

[0115] The plurality of antennas **2** is preferably mounted on the pivotable cover **34** of the housing **32** and are positionable in at least a first position in which the plurality of antennas **2** is in close proximity to the top surface **38** of the base **36** when the cover **34** is in the closed position, and a second position in which the plurality of antennas **2** is in a raised position from the top surface **38** of the base **36** when the cover **34** is in the open position.

[0116] If a solar cell circuit is included in the energy harvesting device, where the solar cell circuit includes a solar cell **40** for receiving light energy, the solar cell **40** is preferably mounted on the top surface **38** of the base **36** and is selectively covered and uncovered by the cover **34** when the cover **34** is respectively pivoted between the closed position and the open position. The cover **34** is preferably at least partially translucent to at least partially allow light energy to pass therethrough and to be received by the solar cell **40** when the cover **34** is in the closed position.

[0117] Alternatively, the energy harvesting device includes a housing **32** having a top surface **38**, whereby the solar cell **40** is mounted on the top surface **38** of the housing. A cover need not be included. The housing **32** may include a base **36** defining an internal cavity for containing the electrical circuit of the energy harvesting device, the top surface **38** of the housing forming a portion of the base **36**. Additionally, a non-pivotable cover **34** may be mounted on the base **36**, the cover **34** at least partially covering the top surface **38** of the base **36** and the solar cell **40** mounted on the top surface **38**. In this embodiment, it is preferred if the cover **34** is at least partially translucent to allow light energy to pass therethrough and impinge on the solar cell **40** situated below the cover **34**.

[0118] Now, various forms of a method of harvesting energy in accordance with the present invention will further be described in detail.

[0119] In one form of the present invention, a method of harvesting energy includes the steps of receiving a transmitted radio frequency (RF) signal to provide a received RF signal, converting the received RF signal to an RF electrical signal, converting the RF electrical signal to a direct current (DC) voltage, temporarily storing the DC voltage to provide a temporarily stored DC voltage, monitoring the temporarily stored DC voltage and generating a pulsed output signal if the temporarily stored DC voltage is above a first predetermined voltage level, converting the temporarily stored DC voltage to a first converted DC voltage which is greater than the temporarily stored DC voltage, converting the first converted DC voltage to a second converted DC voltage which is less than the first converted DC voltage, storing the second converted DC voltage to provide a stored second converted DC voltage, converting the stored second converted DC voltage to a third converted DC voltage which is greater than the stored second converted DC voltage, and providing the third converted DC voltage to an output connector **30**, an electronic device **8** being connectable to the output connector **30** for receiving thereon the third converted DC voltage provided to the output connector **30**.

[0120] The method of harvesting energy described above may further include the step of connecting a source of external DC voltage to an input connector **24**, the input connector **24** receiving the external DC voltage, the first converted DC voltage being at least partially defined by the external DC voltage.

[0121] Furthermore, the method of harvesting energy described above may further include the step of monitoring the stored second converted DC voltage and not generating the pulsed output signal if the stored second converted DC voltage is above a second predetermined voltage level.

[0122] In another form of the present invention, a method of harvesting energy includes the steps of receiving a transmitted radio frequency (RF) signal to provide a received RF signal, converting the received RF signal to a direct current (DC) voltage, and providing the DC voltage on an output connector **30**, an electronic device **8** being connectable to the output connector **30** for receiving the DC voltage provided on the output connector **30**.

[0123] In yet another form of the present invention, a method of harvesting energy includes the steps of receiving by a plurality of antennas **2** a transmitted radio frequency (RF) signal and generating RF electrical signals in response to the transmitted RF signals received by the plurality of antennas **2**, multiplying and rectifying the RF electrical signals generated by the plurality of antennas **2** and generating a plurality of direct current (DC) voltages in response thereto, combining in series at least some of the DC voltages to provide a combined DC voltage, the combined DC voltage being greater in magnitude than each individual DC voltage, temporarily storing the combined DC voltage thereby providing a temporarily stored DC voltage, generating a DC charging voltage in response to the temporarily stored DC voltage, storing the DC charging voltage to provide a stored DC charging voltage, and generating a DC output connector voltage on an output connector **30** in response to the stored DC charging voltage, an electronic device **8** being connectable to the output connector **30** for receiving thereon the output connector DC voltage provided on the output connector **30**.

[0124] Furthermore, the method of harvesting energy described above may further include the step of generating a jump start DC voltage, the combined DC voltage including the at least some of the DC voltages generated by the multiplying and rectifying step connected in series and the jump start DC voltage, and may further include the step of connecting an external DC voltage to an input connector **24**, at least one of the stored DC charging voltage, the jump start DC voltage and the output connector DC voltage being defined at least partially by the external DC voltage.

[0125] Additionally, the method of harvesting energy described previously may include the step of converting the stored DC charging voltage to the output connector DC voltage provided on the output connector **30**, and may include the further step of monitoring the temporarily stored DC voltage and selectively generating the DC charging voltage if the temporarily stored DC voltage is greater than a predetermined voltage threshold.

[0126] If a solar cell circuit is included in the energy harvesting device of the present invention, then the method of harvesting energy described above may include the step of receiving light energy by the solar cell circuit and generating a solar cell DC output voltage in response thereto, the at least one of the stored DC charging voltage, the output connector DC voltage provided on the output connector and the jump start DC voltage being defined at least partially by the solar cell DC output voltage.

[0127] The method of harvesting energy may further include the steps of coupling in series at least some of the DC voltages generated by the multiplying and rectifying step to provide group DC voltages, and combining the group DC voltages in parallel to define the combined DC voltage.

[0128] The method of harvesting energy may further include the step of adjustably positioning the plurality of antennas **2** between a first position and a second position, wherein the plurality of antennas **2** exhibits better signal reception characteristics to receive a transmitted RF signal when in the first position than when in the second position.

[0129] Although illustrative embodiments of the present invention have been described herein with reference to the accompanying drawings, it is to be understood that the invention is not limited to those precise embodiments, and that various other changes and modifications may be effected therein by one skilled in the art without departing from the scope or spirit of the invention.

What is claimed is:

1. An energy harvesting device, which comprises:

an antenna for receiving a transmitted radio frequency (RF) signal;

a rectifier circuit operatively coupled to the antenna, the rectifier circuit converting the RF signal received by the antenna to a direct current (DC) voltage; and

an output connector operatively coupled to the rectifier circuit and providing the DC voltage thereon, an electronic device being connectable to the output connector for receiving the DC voltage provided on the output connector.

2. An energy harvesting device, which comprises:

an antenna for receiving a transmitted radio frequency (RF) signal, the antenna converting the received RF signal to an RF electrical signal;

an RF-to-DC rectifier circuit, the RF-to-DC rectifier circuit being operatively connected to the antenna and converting the RF electrical signal to a direct current (DC) voltage;

a first storage device, the first storage device being operatively connected to the RF-to-DC rectifier circuit and being provided with the DC voltage and being charged thereby;

a microcontroller, the microcontroller having a first signal input operatively connected to the first storage device to monitor the DC voltage on the first storage device, the microcontroller generating a pulsed output signal if the DC voltage on the first storage device is above a first pre-determined voltage level;

a first stage DC-to-DC voltage converter circuit, the first stage DC-to-DC voltage converter circuit being operatively connected to the first storage device and generating an output DC voltage which is greater than the DC voltage on the first storage device;

a second stage DC-to-DC voltage converter circuit, the second stage DC-to-DC voltage converter circuit being operatively connected to the first stage DC-to-DC voltage converter circuit and generating an output DC voltage which is less than the output DC voltage generated by the first stage DC-to-DC voltage converter circuit;

a second storage device, the second storage device being operatively connected to the second stage DC-to-DC voltage converter circuit and being provided with the output DC voltage generated by the second stage DC-to-DC voltage converter circuit and being charged thereby;

a third stage DC-to-DC voltage converter circuit, the third stage DC-to-DC voltage converter circuit being operatively connected to the second storage device and generating an output DC voltage which is greater than the DC voltage on the second storage device; and

an output connector, the output connector being operatively connected to the third stage DC-to-DC voltage converter circuit and providing the output DC voltage generated by the third stage DC-to-DC voltage converter circuit thereon, an electronic device being connectable to the output connector for receiving thereon the output DC voltage provided to the output connector.

3. An energy harvesting device as defined by claim **2**, which further comprises:

an input connector, the input connector being operatively connected to the second stage DC-to-DC voltage converter circuit, the input connector being connectable to an external source of DC voltage for providing the DC voltage from the external source to the second stage DC-to-DC voltage converter circuit, the second stage DC-to-DC voltage converter circuit generating the output DC voltage of the second stage DC-to-DC voltage converter circuit in response to the DC voltage from the external source provided on the input connector.

4. An energy harvesting device as defined by claim **2**, wherein the first storage device includes a capacitor.

5. An energy harvesting device as defined by claim **2**, wherein the second storage device includes a battery.

6. An energy harvesting device as defined by claim **2**, wherein the microcontroller includes a second signal input operatively connected to the second storage device to monitor the DC voltage on the second storage device, the microcontroller not generating the pulsed output signal if the DC voltage on the second storage device is above a second predetermined voltage level.

7. An energy harvesting device, which comprises:

a plurality of antennas for receiving a transmitted radio frequency (RF) signal, the antennas generating RF electrical signals in response to the transmitted RF signals received thereby;

a plurality of multiplier and rectifier circuits, each multiplier and rectifier circuit being responsive to a respective RF electrical signal generated by an antenna of the plurality of antennas and generating a direct current (DC) voltage in response thereto, at least some of the DC voltages being stacked in series to provide a combined DC voltage, the combined DC voltage being greater in magnitude than each individual DC voltage generated by a respective multiplier and rectifier circuit;

a temporary storage device, the temporary storage device being responsive to the combined DC voltage and providing a temporarily stored DC voltage in response thereto;

a first power management circuit, the first power management circuit being responsive to the temporarily stored DC voltage and generating a DC charging voltage in response thereto;

a main energy storage device, the main energy storage device being responsive to the DC charging voltage generated by the first power management circuit and providing a main energy storage DC voltage in response thereto; and

an output connector, the output connector providing an output connector DC voltage thereon in response to the main energy storage DC voltage, an electronic device being connectable to the output connector for receiving thereon the output connector DC voltage provided on the output connector.

8. An energy harvesting device as defined by claim **7**, which further comprises:

a jump start DC voltage device, the jump start DC voltage device providing a jump start DC voltage, the combined DC voltage including the at least some of the DC voltages generated by the plurality of multiplier and rectifier circuits stacked in series and the jump start DC voltage.

9. An energy harvesting device as defined by claim **7**, which further comprises:

an input connector, the input connector being connectable to a source of DC voltage external to the energy harvesting device, the DC voltage of the external source being provided to at least one of the main energy storage device and the output connector.

10. An energy harvesting device as defined by claim **7**, which further comprises:

a DC-to-DC voltage converter circuit, the DC-to-DC voltage converter circuit being responsive to the main energy storage DC voltage and generating an output DC voltage in response thereto, the output connector providing the output connector DC voltage in response to the output DC voltage generated by the DC-to-DC voltage converter circuit.

11. An energy harvesting device as defined by claim **7**, which further comprises:

a watchdog monitor circuit, the watchdog monitor circuit being responsive to the temporarily stored DC voltage and selectively providing the temporarily stored DC voltage to the first power management circuit if the temporarily stored DC voltage is greater than a predetermined voltage threshold.

12. An energy harvesting device as defined by claim **7**, which further comprises:

a solar cell circuit, the solar cell circuit receiving light energy and generating a solar cell DC output voltage in response thereto, the solar cell DC output voltage being provided to at least one of the main energy storage device and the output connector.

13. An energy harvesting device as defined by claim **8**, which further comprises:

a solar cell circuit, the solar cell circuit receiving light energy and generating a solar cell DC output voltage in response thereto, the solar cell DC output voltage being provided to at least one of the main energy storage device, the output connector and the jump start DC voltage device.

14. An energy harvesting device as defined by claim **7**, which further comprises:

a solar cell circuit, the solar cell circuit receiving light energy and generating a solar cell DC output voltage in response thereto; and

a second power management circuit, the second power management circuit being responsive to the solar cell

DC output voltage and generating a managed solar cell DC output voltage in response thereto, the managed solar cell DC output voltage being provided to at least one of the main energy storage device and the output connector.

15. An energy harvesting device as defined by claim **8**, which further comprises:

a solar cell circuit, the solar cell circuit receiving light energy and generating a solar cell DC output voltage in response thereto; and

a second power management circuit, the second power management circuit being responsive to the solar cell DC output voltage and generating a managed solar cell DC output voltage in response thereto, the managed solar cell DC output voltage being provided to at least one of the main energy storage device, the output connector and the jump start DC voltage device.

16. An energy harvesting device as defined by claim **7**, wherein at least some of the DC voltages generated by the multiplier and rectifier circuits are coupled in series to provide group DC voltages, and wherein the group DC voltages are combined in parallel to define the combined DC voltage.

17. An energy harvesting device as defined by claim **16**, wherein the plurality of antennas includes a first antenna, a second antenna, a third antenna and a fourth antenna, the first antenna generating a first RF electrical signal, the second antenna generating a second RF electrical signal, the third antenna generating a third RF electrical signal, the fourth antenna generating a fourth RF electrical signal;

wherein the plurality of multiplier and rectifier circuits includes a first multiplier and rectifier circuit, a second multiplier and rectifier circuit, a third multiplier and rectifier circuit and a fourth multiplier and rectifier circuit, the first multiplier and rectifier circuit being responsive to the first RF electrical signal generated by the first antenna and generating a first multiplied and rectified DC voltage in response thereto, the second multiplier and rectifier circuit being responsive to the second RF electrical signal generated by the second antenna and generating a second multiplied and rectified DC voltage in response thereto, the third multiplier and rectifier circuit being responsive to the third RF electrical signal generated by the third antenna and generating a third multiplied and rectified DC voltage in response thereto, the fourth multiplier and rectifier circuit being responsive to the fourth RF electrical signal generated by the fourth antenna and generating a fourth multiplied and rectified DC voltage in response thereto;

wherein the first multiplied and rectified DC voltage and the second multiplied and rectified DC voltage are coupled in series to define a first group DC voltage;

wherein the third multiplied and rectified DC voltage and the fourth multiplied and rectified DC voltage are coupled in series to define a second group DC voltage;

and wherein the first group DC voltage and the second group DC voltage are combined in parallel to define the combined DC voltage.

18. An energy harvesting device as defined by claim **7**, wherein each antenna of the plurality of antennas has a broadband bandwidth characteristic.

19. An energy harvesting device as defined by claim **7**, wherein each antenna of the plurality of antennas has a narrowband bandwidth characteristic.

20. An energy harvesting device as defined by claim **7**, which further comprises:

a housing, the housing having a base defining an internal cavity, and a cover pivotally mounted on the base, the base having a top surface, the cover being pivotable between at least a closed position in which the cover is in close proximity to the base to at least partially cover the top surface, and an open position in which the cover is substantially perpendicular to the top surface of the base.

21. An energy harvesting device as defined by claim **20**, wherein the plurality of antennas is mounted on the pivotable cover of the housing and being positionable in at least a first position in which the plurality of antennas is in close proximity to the top surface of the base when the cover is in the closed position, and a second position in which the plurality of antennas is in a raised position from the top surface of the base when the cover is in the open position.

22. An energy harvesting device as defined by claim **20**, which further comprises:

a solar cell circuit, the solar cell circuit including a solar cell for receiving light energy, the solar cell circuit generating a solar cell DC output voltage in response to the light energy received by the solar cell, the solar cell being mounted on the top surface of the base and selectively covered and uncovered by the cover when the cover is respectively pivoted between the closed position and the open position.

23. An energy harvesting device as defined by claim **7**, which further comprises:

a solar cell circuit, the solar cell circuit including a solar cell for receiving light energy, the solar cell circuit generating a solar cell DC output voltage in response to the light energy received by the solar cell; and

a housing, the housing having a top surface, the solar cell being mounted on the top surface of the housing

24. An energy harvesting device as defined by claim **23**, wherein the housing includes a base defining an internal cavity, and a cover mounted on the base, the top surface of the housing forming a portion of the base, the cover at least partially covering the top surface of the base and the solar cell mounted on the top surface.

25. An energy harvesting device as defined by claim **24**, wherein the cover is at least partially translucent to at least partially allow light energy pass therethrough and to be received by the solar cell.

26. An energy harvesting device as defined by claim **21**, wherein the cover is at least partially translucent to at least partially allow light energy to pass therethrough and to be received by the solar cell when the cover is in the closed position.

27. An energy harvesting device as defined by claim **7**, wherein the temporary storage device includes a capacitor.

28. An energy harvesting device as defined by claim **7**, wherein the main energy storage device includes a battery.

29. An energy harvesting device as defined by claim **8**, wherein the jump start DC voltage device includes a battery.

30. A method of harvesting energy, which comprises the steps of:

receiving a transmitted radio frequency (RF) signal to provide a received RF signal;

converting the received RF signal to an RF electrical signal;

converting the RF electrical signal to a direct current (DC) voltage;

temporarily storing the DC voltage to provide a temporarily stored DC voltage;

monitoring the temporarily stored DC voltage and generating a pulsed output signal if the temporarily stored DC voltage is above a first predetermined voltage level;

converting the temporarily stored DC voltage to a first converted DC voltage which is greater than the temporarily stored DC voltage;

converting the first converted DC voltage to a second converted DC voltage which is less than the first converted DC voltage;

storing the second converted DC voltage to provide a stored second converted DC voltage;

converting the stored second converted DC voltage to a third converted DC voltage which is greater than the stored second converted DC voltage; and

providing the third converted DC voltage to an output connector, an electronic device being connectable to the output connector for receiving thereon the third converted DC voltage provided to the output connector.

31. A method of harvesting energy as defined by claim **30**, which further comprises the step of:

connecting a source of external DC voltage to an input connector, the input connector receiving the external DC voltage, the first converted DC voltage being at least partially defined by the external DC voltage.

32. A method of harvesting energy as defined by claim **30**, wherein the step of temporarily storing the DC voltage is at least partially performed by using a capacitor.

33. A method of harvesting energy as defined by claim **30**, wherein the step of storing the second converted DC voltage is at least partially performed by using a battery.

34. A method of harvesting energy as defined by claim **30**, which further comprises the step of:

monitoring the stored second converted DC voltage and not generating the pulsed output signal if the stored second converted DC voltage is above a second predetermined voltage level.

35. A method of harvesting energy, which comprises the steps of:

receiving a transmitted radio frequency (RF) signal to provide a received RF signal;

converting the received RF signal to a direct current (DC) voltage; and

providing the DC voltage on an output connector, an electronic device being connectable to the output connector for receiving the DC voltage provided to the output connector.

36. A method of harvesting energy, which comprises the steps of:

receiving by a plurality of antennas a transmitted radio frequency (RF) signal and generating RF electrical signals in response to the transmitted RF signals received by the plurality of antennas;

multiplying and rectifying the RF electrical signals generated by the plurality of antennas and generating a plurality of direct current (DC) voltages in response thereto;

combining in series at least some of the DC voltages to provide a combined DC voltage, the combined DC voltage being greater in magnitude than each individual DC voltage;

temporarily storing the combined DC voltage thereby providing a temporarily stored DC voltage;

generating a DC charging voltage in response to the temporarily stored DC voltage;

storing the DC charging voltage to provide a stored DC charging voltage; and

generating a DC output connector voltage on an output connector in response to the stored DC charging voltage, an electronic device being connectable to the output connector for receiving thereon the output connector DC voltage provided on the output connector.

37. A method of harvesting energy as defined by claim **36**, which further comprises the step of:

generating a jump start DC voltage, the combined DC voltage including the at least some of the DC voltages generated by the multiplying and rectifying step connected in series and the jump start DC voltage.

38. A method of harvesting energy as defined by claim **36**, which further comprises the step of:

connecting an external DC voltage to an input connector, at least one of the stored DC charging voltage and the output connector DC voltage being defined at least partially by the external DC voltage.

39. A method of harvesting energy as defined by claim **36**, which further comprises the step of:

converting the stored DC charging voltage to the output connector DC voltage provided on the output connector.

40. A method of harvesting energy as defined by claim **36**, which further comprises the step of:

monitoring the temporarily stored DC voltage and selectively generating the DC charging voltage if the temporarily stored DC voltage is greater than a predetermined voltage threshold.

41. A method of harvesting energy as defined by claim **36**, which further comprises the step of:

receiving light energy by a solar cell circuit and generating a solar cell DC output voltage in response thereto, at least one of the stored DC charging voltage and the output connector DC voltage provided on the output connector being defined at least partially by the solar cell DC output voltage.

42. A method of harvesting energy as defined by claim **37**, which further comprises the step of:

receiving light energy by a solar cell circuit and generating a solar cell DC output voltage in response thereto, the at least one of the stored DC charging voltage, the output connector DC voltage provided on the output connector and the jump start DC voltage being defined at least partially by the solar cell DC output voltage.

43. A method of harvesting energy as defined by claim **36**, which further comprises the steps of:

coupling in series at least some of the DC voltages generated by the multiplying and rectifying step to provide group DC voltages; and

combining the group DC voltages in parallel to define the combined DC voltage.

44. A method of harvesting energy as defined by claim **36**, wherein each antenna of the plurality of antennas has a broadband bandwidth characteristic.

45. A method of harvesting energy as defined by claim **36**, wherein each antenna of the plurality of antennas has a narrowband bandwidth characteristic.

46. A method of harvesting energy as defined by claim **36**, which further comprises the steps of:

adjustably positioning the plurality of antennas between a first position and a second position, wherein the plurality of antennas exhibits better signal reception characteristics to receive a transmitted RF signal when in the first position than when in the second position.

47. A method of harvesting energy as defined by claim **41**, which further comprises the step of:

mounting the plurality of antennas and at least a portion of the solar cell circuit to a housing, the housing having a base defining an internal cavity, and a cover pivotally mounted on the base, the base having a top surface, the cover being pivotable between at least a closed position in which the cover is in close proximity to the base to at least partially cover the top surface, and an open position in which the cover is substantially perpendicular to the top surface of the base, the plurality of antennas being mounted on the pivotable cover of the housing and being positionable in at least a first position in which the plurality of antennas is in close proximity to the top surface of the base when the cover is in the closed position, and a second position in which the plurality of antennas is in a raised position from the top surface of the base when the cover is in the open position, the at least portion of the solar cell circuit being mounted on the top surface of the base and selectively covered and uncovered by the cover when the cover is respectively pivoted between the closed position and the open position.

48. A method of harvesting energy as defined by claim **47**, wherein the cover is at least partially translucent to at least partially allow light energy to pass therethrough and to be received by the at least portion of the solar cell circuit.

49. A method of harvesting energy as defined by claim **36**, wherein the step of temporarily storing the combined DC voltage is at least partially performed by using a capacitor.

50. A method of harvesting energy as defined by claim **36**, wherein the step of storing the DC charging voltage is at least partially performed by using a battery.

51. A method of harvesting energy as defined by claim **37**, wherein the step of generating a jump start DC voltage is at least partially performed by using a battery.

* * * * *

Phil's Old Radios: Home: Building

Radio Wrinkles

Build a Crystal Shortwave Radio

This project combines two popular themes from radio history—crystal radios and shortwave (SW) listening. It's designed from scratch by our non-resident engineer Walter Heskes.

Despite all of the advances in modern electronics, there are thousands of crystal sets in daily use throughout the world. The fidelity of tone and clarity of crystal rectification are hard to beat. What's more, crystal sets are absolutely free! A crystal receiver is powered solely by the radio waves that it pulls from the air.

We dubbed this inexpensive radio the "EconOceanic" in honor of the legendary Zenith TransOceanic family of shortwave radios. It tunes the AM broadcast band, from .5 MHz to 1.6 MHz, and most of the shortwave band, from 1.7 MHz up to about 17 MHz.

There's plenty of SW activity in these bands, so you'll hear lots of different voices, a mix of languages, and even music. In addition to powerhouse stations such as the BBC, you'll also hear non-commercial amateurs, some of whom broadcast using the dits and dahs of International Morse Code. And don't forget to set your wristwatch by the precise WWV time signals. They appear at exact multiples of 5.0 MHz across the SW bands. Listen for the calibration beeps and periodic time announcements.

We hope you enjoy this project. If you build an EconOceanic, send some email to let us know how it worked out. Walter can be reached at walterh@interworld.com.

Building Instructions

You're eager to start assembly, but a few minutes of preparation will help this project go smoothly. Here's how to get started.

Printing Instructions

Before you lay hands on any parts, print out this page and all of the diagrams in the following table.

Cabinet Decal	Component Layout
Coil Selector Circuit	Tuner/Detector Circuit
Coil Winding Diagram	Antenna Bracket Assembly

The large diagrams are designed to print on a standard 8.5 x 11-inch sheet of paper. Most of them are oriented horizontally for easier reading online. Choose your printer's "landscape" mode for the horizontal diagrams, to make sure they'll print on one sheet.

Be sure that your printer is set to print at a 1-to-1 ratio, so that the cabinet decal and layout diagram are the correct size.

Parts and Tools

We recommend getting *all* of the parts you need before you start assembly. Nothing is more frustrating than getting halfway through a project, then having to wait several days for some missing part to arrive in the mail.

The following parts list gives part numbers and sources for all major parts, plus some notes on particular items.

Most of the parts you need to build the EconOceanic Crystal Shortwave Receiver are available through Radio Shack. Two exceptions are the big tuning capacitor and the little trimmer capacitor. The big tuning capacitor can be obtained from Antique Electronic Supply. Or, you may be lucky enough to have a spare tuning capacitor in your junkbox. The small trimmer capacitor is available from All Electronics, allcorp@allcorp.com, 1-800-826-5432. Some notes on specific parts appear after this list.

Quantity	Component	Type	Part No.	Price
1	Diode	1N34A [1]	Radio Shack 276-1123	$ 1.19 pkg 10
1	Variable capacitor [2]	Dual-range: 40-460 pF/30-156 pF	AES CV300-145	$3.50
1	Trimmer capacitor	6-40pF	All Elect. VCAP-14	$1.50
1	Ceramic capacitors [3]	10 pF, 22 pF, 50 pF	Radio Shack 272-801	$2.49 pkg 100
2	Phone jacks	1/8-inch diameter	Radio Shack 274-251	$2.49 pkg 3
1	Speaker terminal strip [4]	Spring-loaded	Radio Shack 274-315	$ 1.29
4	Rotary switches	2-pole, 6-circuit	Radio Shack 275-1386	$1.69
1	Slide switch	DPDT	Radio Shack 275-407	$ 1.29 pkg 2
1	Enameled wire [5]	#22 and #26	Radio Shack 278-1345	$3.99 pkg 3
3	Knobs	1.0-inch diameter	Radio Shack 274-416	$2.29 pkg 4
2	Knobs	0.75-inch diameter	Radio Shack 274-415	$ 1.89 pkg 4
1	Project box	7.5" x 4.33" x 2.22" plastic	Radio Shack 270-224	$ 3.49
1	Cardboard tube	From paper towel roll		
1	Epoxy glue [6]	JB Weld or similar		
1	Household glue [7]	Elmer's Glueall or similar		
1	Rubber feet [8]	Heavy-duty self-stick	Radio Shack 64-2342	$ 1.49 pkg 8
Optional				
1		Battery powered	Radio Shack 277-	$11.99

	Amplified speaker [9]		1008	
1	Headphones	2,000-ohm	AES P-A466	$14.50
1	Antenna	6-foot telescoping	Radio Shack 270-1408B	$4.99
1	Fuse holder clips [10]	(for antenna swivel)	Radio Shack 270-739A	$ 1.19 pkg 2

Notes

[1] Any type of germanium diode will work.

[2] A variable air capacitor can be scrounged from any old AC/DC superheterodyne AM radio. Use the RF section for the wide range (40-460 pF) and the oscillator section for the narrow range (30-156 pF)

[3] Bags of ceramic capacitors are available in large quantities from Antique Electronic Supply and Radio Shack. These capacitors will work fine, but if you buy them in bulk, you may need to identify them using their markings or by reading them with a capacitor meter.

[4] The speaker terminal strip is used for the ground and antenna connections, as explained in the building instructions.

[5] Enamel-coated wire is available from Radio Shack and other suppliers. Radio Shack part no. 278-1345 is a package of three spools of enameled wire in the following sizes: #22 (gold enamel), #26 (green enamel), and #30 (red enamel), all for $3.99. If you don't see it on display, find it in the catalog and point it out to the clerks so they will order it for you.

[6] It's not critical which type of glue you use to fasten the metal frame of the tuning capacitor to the plastic surface of the instrument panel. Epoxy glue (JB weld or similar) will work, and so will silicon-based glues and perhaps others.

[7] Elmer's Glueall will work to coat the coils as you wind them. Or, you can use polyurethane varnish to seal the coils.

[8] Use four rubber feet to support the cabinet and two rubber feet to fasten the coil to the inside panel of the cabinet.

[9] If you choose to use an amplified speaker (highly recommended), use a short coaxial cable terminated with 1/8" plugs on both ends to connect the EconOceanic output to the speaker.

[10] One of the fuse holder clips is modified to form a swiveling support for the antenna mast, as explained in the building instructions. To fasten the antenna mast assembly to the cabinet, you'll also need a small screw, rubber washer, metal washer, lockwasher, and nut. If you don't have those items lying around somewhere, your house is cleaner than ours!

The tools that you'll need are pretty simple—a soldering iron, drill, wire stripper (or knife), screwdriver, and pliers or crescent wrench.

Preparing the Cabinet

Let's go! The first major step is to drill mounting holes for the controls in the plastic cabinet. The entire radio is mounted to the front panel, making it easy to remove for modifications or simply for showing a friend your handiwork.

The Component Layout Diagram and Cabinet Decal show you where to drill holes in the cabinet front panel for the components. Carefully trim the decal to fit the cabinet, then punch or cut holes in the decal where indicated. Glue on the decal. Don't put any glue under the places where you'll be drilling holes. Using a scribe or sharp knife, and guided by the scribe marks on the decal, make a little guide hole in the exact center of each place to drill. Using a sharp knife, carefully trim away the paper from each hole, so the drill won't tear the paper.

Regarding the decal, not all printers are created equal, so make sure it's printed at the right size before gluing it on. At the correct size, the decal measures 7.5 inches wide (horizontally) at the trim lines. If you can't get it to print correctly, you can try loading the image into a graphics program such as Photoshop or Paintshop Pro and resizing it for a new printout. The component layout diagram gives exact measurements for the holes, in any case.

Before you drill any holes, check clearances by placing the actual knobs in place atop the diagram. Now drill the holes. Drilling plastic is fast and easy. Wear protective eyegear to guard against flying fragments of plastic.

If you'd like to protect the paper decal against dirt and wear, spray on a thin coat of clear lacquer or poly finish, then set it aside to dry. A more permanent solution would be to have it laminated beforehand at a local copy shop.

Finally, mount the components in the positions shown in the component layout diagram. Pop the front panel into the case and admire your work. From the outside, this project looks almost finished—all that's left is to build the radio that goes inside!

Below are front and rear views of the author's prototype, after mounting the controls on the front panel. (This prototype did not have a decal.)

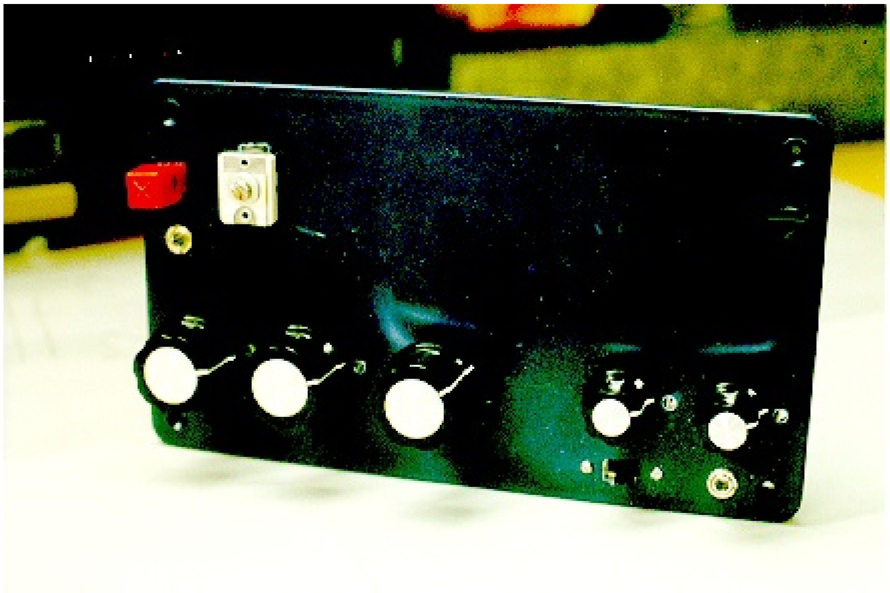

Building the Switchable Coils

Switchable coils are a unique feature of the EconOceanic. With many old shortwave radios, changing bands required you to unplug one coil and plug in another. With the EconOceanic, you simply turn switches to change coils in or out. That's more convenient, and it reduces wear and tear on the coils.

Making a Rigid Coil Form

In the EconOceanic, all of the BC and SW coils will be wound on a single core. The first step is to get something to wind your coils on. A standard roll of paper towels comes on a 1 5/8-inch diameter cardboard tube, which is just right.

Although you can wind the coils on a single tube, it's a good idea to reinforce the tube with a second stiffening layer, to help resist the tension of the wire as you wind the coils. To create the reinforcement, get an extra cardboard tube, slice it down its length, and remove a 3/8-inch wide strip along the cut, thus reducing its diameter. (Remember that this inner tube has to fit inside another tube of the same original diameter.) Apply a thin layer of Elmer's Glueall to the entire outer surface of the inner tube and insert it through the outer tube. Let it expand and gently press the tubes together to spread the glue. Hold them together for a couple of minutes, then set them aside for the glue to dry. Now you have a single, thick core.

Marking Eight Sections on the Coil Form

This design requires precise winding, which will be easier if you take a moment to mark your core with lines along its length to indicate its halves, quadrants, and eighth-sections. (In other words, draw eight equally-spaced straight lines, each one going the entire length of the tube.) When you need to wind a coil of 22 3/4 turns, you'll know exactly where to stop during the 23rd turn.

The more accurate your coils, the better they'll work. A coil with the wrong number of turns provides the wrong inductance and will not tune the desired range of frequencies.

Winding Your Coils

It's time to start winding, a task that requires care. Here's a little trick that simplifies the process. As you wind each coil, it helps to apply a thin layer of Elmer's Glueall to hold the most recently wound section together. Otherwise, the wires tend to spring apart (if that happens, you have to rewind the coil from scratch.) Wind about 10 turns, hold them tightly in place with your thumb, apply a pea-sized glob of glue to the wound section, and spread it with the middle finger of your other hand. Work the glue between the coils to encapsulate them. The thin layer dries in a few seconds and you can quickly continue winding. It's a technique you can develop pretty quickly. By the time you've wound all 101.5 turns of the big broadcast (BC) band coil, you'll feel like an expert!

Wind one coil at a time, following the Coil Winding Diagram. Use a full spool of wire when you start winding the big BC coil. It can be heartbreaking to run out of wire just before you reach the end.

Leave 10 inches of extra wire at each end of each coil. It is better to have extra length now which can be trimmed away later. These ends are the leads that you'll connect to the other components. Mark each lead wires using a little flag of masking tape (for instance, "250 uH Antenna" and "250 uH Ground"). The markers will simplify hookup later on.

The chart below lists the number of turns and size of wire to use for each coil. This radio tunes five bands. Each band requires a pair of windings: a small primary and a larger secondary.

Coil Section	Number of Turns	Wire gauge
250uH primary	17	26
250uH secondary	102	26
40uH primary	5.6	22
40uH secondary	33.5	22
20 uH primary	3.3	22
20 uH secondary	19.7	22
15 uH primary	2.3	22
15 uH secondary	14.5	22
10 uH primary	2.2	22
10 uH secondary	11.1	22

Each winding must be separated by a small gap from the next. Leave a 1/8-inch gap between windings for the 250uH, 40 uH, and 20 uH coils. Leave a 1/16-inch gap between the 15 uH and 10 uH coils.

Winding can be tedious, but it doesn't pay to rush. If you do this part correctly, you'll have a good-performing radio that performs virtually maintenance-free for many years.

When all the winding is done, check your work to make sure that everything is tight and secure. If necessary, apply more glue where needed. You can also apply a thin coat of clear varnish, if desired, for a more finished appearance.

This photo shows the author's prototype coil, ready for mounting.

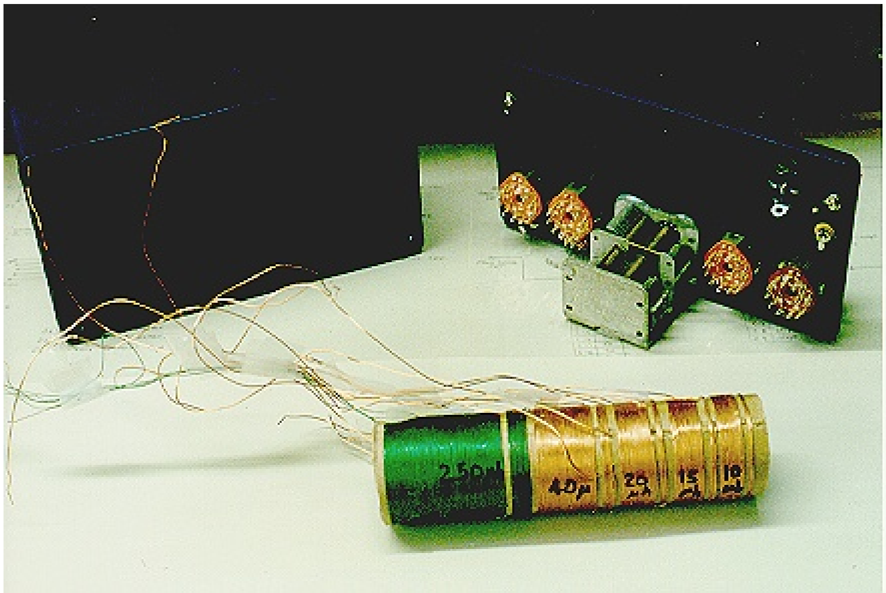

Mounting the Coils

Mount the coil assembly to the inside surface of the front panel. Check for adequate clearance when the tuning capacitor is fully open. Leave about 1/8-inch minimum clearance between the open rotor and the coil assembly. The rotor and coil must not touch. Mark the center of the coil form and mount self-sticking rubber feet to the centerline of the coil form. The feet serve as standoffs from the front panel, providing some clearance under the coil through which you can pass wires.

This photo shows the author's prototype coil mounted inside the front panel, which was placed upside down on the cabinet for the photo. Note the little plastic ties used to keep related leads together.

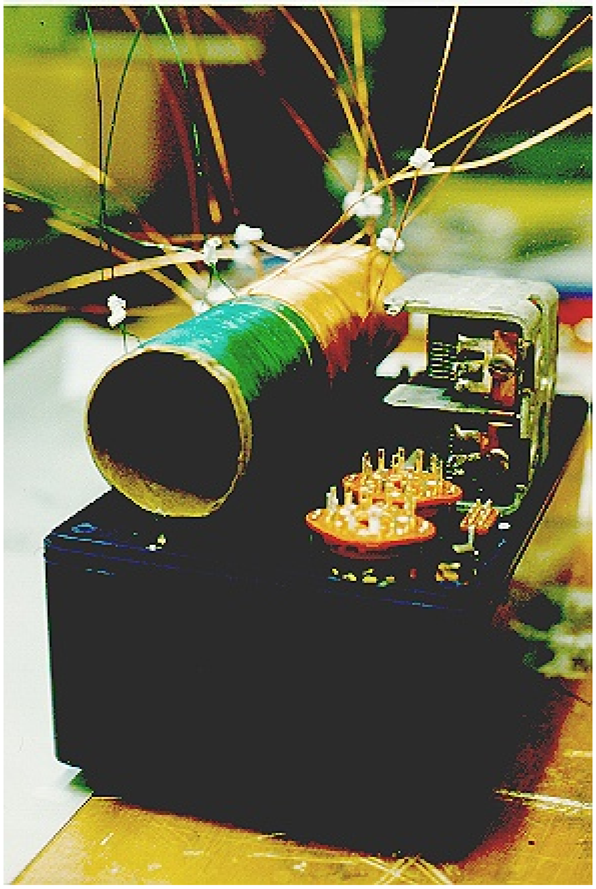

Wiring the chassis

Before you start hooking up components, take a few moments to study the schematic diagrams carefully. The ones you'll be using are the Coil Selector Circuit and Tuner/Detector Circuit.

Connecting Wires for Antenna and Ground

Let's start by attaching lead wires to the antenna and ground connections. These are shown in the Coil Selector Circuit. Connect wires to both antenna input jacks, the antenna trimmer, and the ground input jack. Check your connections for correctness and solder them.

Connecting Coils to the Coil Selector Switch

Identify the pinouts of the rotary switches. Use an ohmmeter to check continuity across the switch circuits and note the position of the switch pointer as each circuit is made. When you are certain you understand how the switch circuits are opened and closed, start wiring the antenna coils to the antenna coil selector switch.

Measure the length of wire needed to span the distance from the coil to the lug at the rotary selector switch. Straight runs are fine. Minimize the number of right angle bends, since those can reduce circuit sensitivity. Trim the coil takeoff leads to approximately 10 inches.

Using a razor blade or fine sandpaper, carefully scrape the enamel coating off each lead wire before you thread it through a lug on the rotary switch. Be careful not to break the fragile lead as you bend it around the lug.

Checking Coil Connections

Do not solder any connections until you have checked for proper continuity through each switch circuit and its coil. Refer to the Coil Selector Circuit as needed while checking your work.

Wiring Tuning Coils

After all the antenna coils are wired to the rotary switch, you can begin wiring the tuning coil circuits, referring to the Tuner/Detector Circuit diagram. Again, check for continuity through each switch circuit and its coil. Do not solder until all leads have been properly wired.

Next, connect the leads of the tuning coil to the tuning capacitor.

Connecting Capacitors

Connect the ceramic capacitors to the bandspread rotary switch. This is tedious work because the capacitor leads are short and the spacing between adjacent rotary switch lugs is small. Work slowly and carefully.

Connect the leads to the bandspread hi-lo selector switch. Since the DPDT switch has two circuits and only one circuit is needed, you can wire across both circuits if that makes wiring a bit easier for you. Or, use a larger SPDT switch.

Making Final Connections

Connect the 1N34 diode to the phone jack and to the band capacitor selector switch. Polarity is not important.

Connect the capacitors and leads to the tuning capacitor selector switch.

Connect the leads to the output jack.

Final Check and Soldering

You're almost done! Before soldering any connections, follow the schematic diagrams and recheck every connection for continuity. Correct any wiring errors that you find. When you're sure that everything's hooked up correctly, solder each lead.

Mounting the Antenna

The EconOceanic can use many different kinds of antennas. You'll get the best reception from a long outdoor antenna, of course. In many areas, however, you can get respectable performance from the 6-foot telescoping antenna listed in the Parts List.

To mount the telescoping antenna, modify one of the fuse holder clips as shown in the Antenna Bracket Assembly diagram. The clip forms a swiveling support for the antenna mast. To fasten the antenna mast assembly to the cabinet, you'll also need a small screw, rubber washer, metal washer, lockwasher, and nut.

Using Your Radio

Your EconOceanic radio is completely assembled. Now for the exciting part—using it! This description may sound harder at first than it really is in practice. We'll go through the process simply because it's a little more involved than using

Step 1. Connect antenna and ground. Connect your antenna and ground wires to the appropriate terminals on the front panel. If you are using the telescoping antenna, it is already connected, of course.

Step 2. Connect listening device. Plug in your 2000-ohm headphones or, better yet, use Radio Shack's amplified speaker so everyone can listen at the same time. The amplified speaker is a transistorized amplifier connected to an internal speaker. It also has a 1/8-inch phone jack, allowing you to listen without disturbing others.

The author connected the output of the amplified speaker to a good quality high fidelity speaker. Later, the author added a second amplified speaker to the radio. The second stage amplifies the weakest signals to audible levels and easily fills the room with sound. This diagram explains the connections:

Connect this . . .	To this
EconOceanic output jack	Input jack of first amplified speaker
Output jack of first amplified speaker	Input jack of second amplified speaker
Output jack of second amplified speaker	Hi-fi speaker

This photo shows the author's prototype, with temporary labels on the panel, connected to two amplified speakers in series.

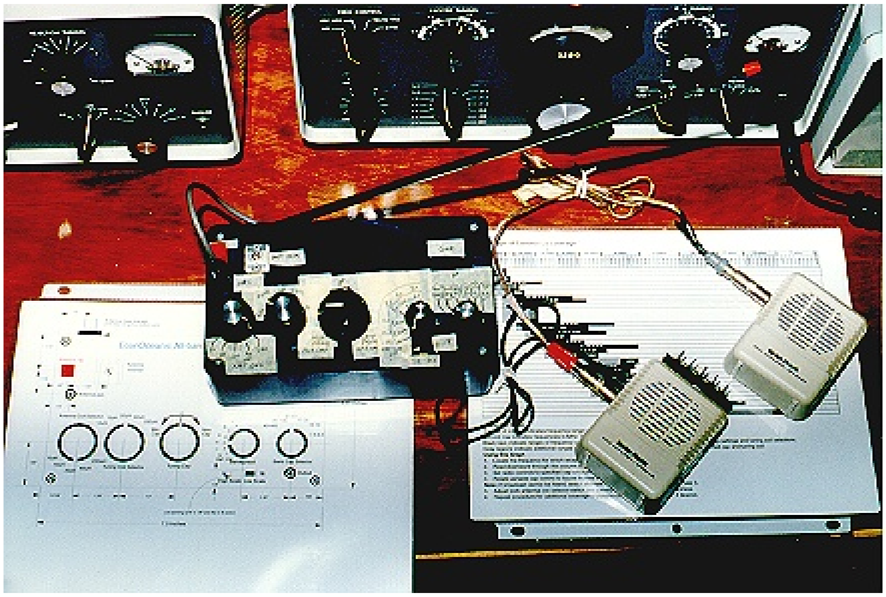

Step 3. Choose a Station. Depending upon the time of day and your location in the world, you can tune across the SW bands and listen in on broadcast activity. Some of the SW bands are more active in the evening hours, while others operate during the day.

Consult a schedule of shortwave broadcasts and note the time of day and frequency at which a desired broadcast will occur. Schedules are published in SW hobbyist magazines available at better newsstands. You can also find loads of schedule information on the Web, such as this resource page at SWLing.com. Another good resource is the book World Radio TV Handbook.

Step 4. Look up tuner settings. Use the Bar Graph of Continuous Coverage to find the correct capacitor and coil settings to tune in the desired frequency. Print this chart using your printer's landscape mode so that it will fit on a single 8.5x11-inch sheet.

As you become more familiar with the principles of resonance, you will find that you can tune the receiver to the desired frequency by thinking about the resonance formula and selecting coils and capacitors according to the following logic:

Low-frequency Broadcasts	High-frequency Broadcasts
Most capacitance	Least capacitance
Longest coil	Shortest coil

For example, to tune the AM broadcast band, you would select your longest coil (250uH) and your biggest capacitor setting (460 pF). Similarly, to tune the highest shortwave frequencies, you would select your shortest coil (10uH) and your smallest capacitor setting (7 - 9 pF).

It's possible to receive the same broadcast signal with different combinations of coils and capacitors. The best combination is the one at which the incoming signal is loudest. Experiment and learn the best settings. You can also combine dissimilar primary and secondary coils to receive radio signals. That is, you can set the antenna coil to 250uH and the tuning coil to 20uH and you will receive SW signals. This occurs because of the inductive coupling effects of the multiple coils on the same coil form.

Remember that the tuning coil alone determines the frequency at which radio signals will be received. The tuning coil resonates with the tuning capacitor. The antenna coil merely feeds the incoming radio waves to the tuning coil. Enjoy this advantage of mounting several adjacent coils on the same core.

Step 5. Turn the antenna coil selector switch to a desired primary coil.

Step 6. Rotate the tuning coil selector switch to the associated secondary coil.

Step 7. Turn the band capacitor selector switch to a desired setting. Now the band is selected.

Step 8. Rotate the tuning capacitor to select a broadcast signal within the selected band.

Step 9. Starting at the lowest setting of the "low" bandspread scale, adjust the bandspread switch for maximum broadcast signal loudness.

Step 10. If you are using the amplified speaker, adjust the volume to a comfortable setting. If you are using a pair of unamplified, 2000-ohm headphones, there is no loudness control.

Step 11. Log your settings. On a sheet of paper, write the time of day, the settings of your selector switches, and a description identifying the source of the broadast signal. For example, you might write:

```
9 P.M. EST, 20uH A coil, 20uH T coil, 13-19 pF, Radio Havana (Cuba) 8.9 MHz.
```

Logging your reception reports will help you locate your favorite radio broadcasts. Documenting your work also adds a hint of substance and credibility to your DX sessions. Above all, have fun!

Theory of Operation

This block diagram shows each stage of the EconOceanic crystal shortwave receiver. Click on each component to read more about it.

```
   Aerial
      |
      |
Antenna coil-->Tuning coil-->Tuning capacitor-->Detector-->Output
      |
      |
Earth ground
```

Aerial

The aerial's job is to capture electromagnetic waves from all sources and directions and funnel those charges into the antenna coil. The loudness of received signals is directly proportional to the height and length of the aerial. Mount the aerial as high as possible, preferably outdoors. A single wire 100 feet long makes an ideal aerial. The EconOceanic design provides for two antenna connections: one 1/8-inch phone jack and a spring-loaded clip (originally intended for fastening hi-fi speaker cables).

If your antenna is mounted outdoors, you should disconnect it whenever there is any chance of lightning in your vicinity.

Antenna coil

The antenna coil is the primary winding of the pair of coils that make up the antenna-tuning transformer. It is a fraction (1/6) of the number of turns that make up the tuning coil. The secondary coil multiplies the electrical field created by the antenna coil.

Earth ground

The earth ground connection effectively extends the antenna into the earth, which contains electrical signals. Keep the ground wire as short and straight as possible. There may be some sensitivity loss if the excess ground wire is coiled into a loop. Use either a metal cold water pipe that is buried at least several feet in the ground or an electrical conduit that is similarly buried or grounded. You can usually make a satisfactory ground connection using the screw that fastens the cover plate to an electrical outlet box.

Tuning coil

The tuning coil is the inductor which, in series with the tuning capacitor, determines the resonant frequency of the receiver according to the formula: **frequency = 1/[2*pi*sqroot(L*C)]**. Wind your coil carefully and space it close to the antenna coil (about 1/8" apart) to tune the specified frequencies.

The amount of coupling is directly proportional to the distance between the windings. If you slide the windings far apart, there is little coupling and a narrow resonance peak (e.g., 100 KHz). If you move the windings closer, there is more coupling and a broader resonance peak (e.g., 200 KHz). A broader resonance peak will mean less selective tuning; that is, several stations will be heard simultaneously at one tuner setting. Experiment, if you wish, with spacing your coils. For example, try 1/4-inch between the primary and secondary windings and test your design for selectivity. Have fun as you learn.

Tuning capacitor

The tuning capacitor is the variable component in the "tank" circuit which comprises the coil and the cap. By varying the capacitance, you can change the resonant frequency of the receiver within the limits of the coil inductance and the total capacitance available in the variable capacitor.

Detector (diode)

In this radio, a diode serves as the detector, separating the fluctuating direct current (containing the voices and music of the broadcast) from the amplitude-modulated, radio-frequency, alternating-current carrier wave that was transmitted from the radio station antenna. The direct current generated by the detector will drive the output device.

Output (headphones or amplified speaker)

The output device changes the electrical energy in the detector circuit into mechanical energy that moves air against our ears to create sound. For satisfactory sound output, this needs to have an input sensitivity of a few millivolts and an input impedance of a few thousand ohms, such as a pair of 2000-ohm headphones or Radio Shack's amplified speaker (part no. 277-1008). Don't try to use a permanent-magnet speaker, such as one scrounged from an old radio. Permanent-magnet speakers require more DC energy than a crystal set can generate.

If you completed this project and enjoy using your EconOceanic, please drop us a line and share your experiences. Walter's email address is walterh@interworld.com.

This radio construction project, including all descriptions, diagrams, photos, and the underlying electronic design, is published here for the noncommercial use of radio hobbyists. You may print and reproduce these project instructions for your personal use. Commercial use of this material is strictly forbidden.

©1995-2017 Philip I. Nelson, all rights reserved

CRYSTAL ENERGY RECEIVER©

Build Your Own Crystal Energy Receiver©

Crystal Energy Receiver©

FREE ENERGY
FROM THIN AIR!
INVOLVES CRYSTALS AND DANGEROUS VOLTAGES!
Copyright 2016. Drew Paul Designs. All Rights Reserved.

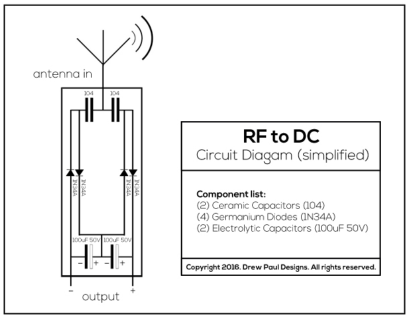
antenna in
104
104
1N34A
1N34A
1N34A
1N34A
100uF 50V
100uF 50V
-
+
-
+
-
output
+

RF to DC
Circuit Diagam (simplified)

Component list:
(2) Ceramic Capacitors (104)
(4) Germanium Diodes (1N34A)
(2) Electrolytic Capacitors (100uF 50V)

Copyright 2016. Drew Paul Designs. All rights reserved.

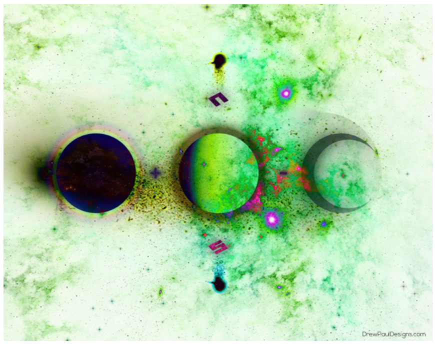
DrewPaulDesigns.com

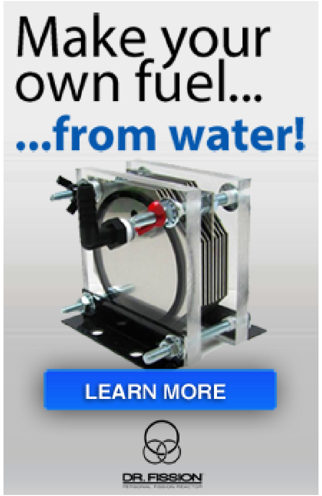
Make your own fuel...
...from water!
LEARN MORE
DR. FISSION

"Dendera Light"
circa 1500 BC
Image: creative commons, Wikipedia

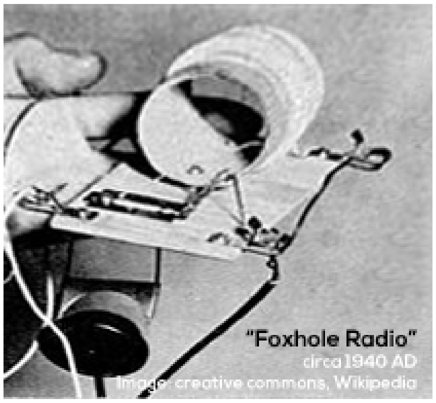

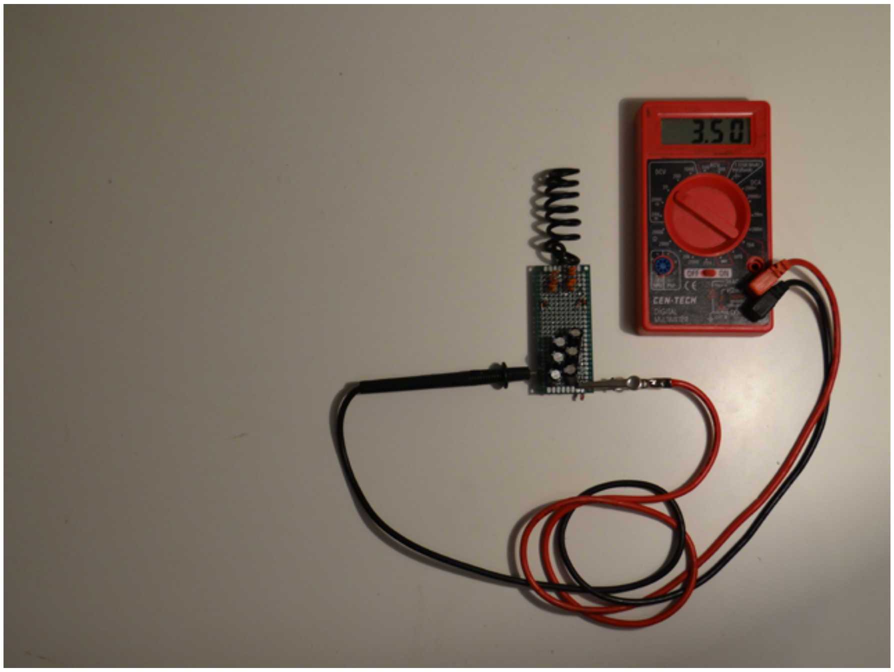

"Free energy from the air? Yea, right!"

Sardonic skepticism was my first reaction to this unusual concept, as well.

Though, its not so far out there, in fact. Light can be converted to DC current with solar panels, electricity can be converted to magnetism as I did in my last article, in a microphone sound waves are converted (by vibrating a magnet near a coil) to an electrical signal, solar rays can even be focused and converted to heat in devices like this. When we think about it, energy is all around us and can be harvested in an enumerable many of ways.

Today, we are going to take a rather novel approach to this venture. The earth is magnetic and anyone who has ever used a compass knows this. Magnetic bodies in motion produce electricity, we can see this in any alternator, like the one in your car. So, therefore the earth is electric as well as magnetic, by their very definitions.

Can we detect this energy? Yes, we sure can! Ever turn on a radio in the middle of nowhere and heard static? That is your radio picking up naturally occurring energy in the RF spectrum!

Can we use this energy to do work? Absolutely! This has been known for a long time. Crystal radios have been around since before the 1930's and can run with no input energy other than the radio signal. Even when completely isolated, but from the atmosphere, a crystal radio will produce a voltage in the earpiece resulting in a sound (albeit and undesirable one).

Well, this is where it gets interesting...

Can we replicate this effect? Yea, and with modern components like the high quality crystals found in germanium diodes, we can even increase efficiency. By applying this concept as a contemporary crystal energy receiver, we can also take advantage of a wide range of energetic frequencies rather than tuning in to just one. Can we scale it up? Definitely. Things like micro germanium diodes, high efficiency antennas and compact contemporary capacitors make the components that are required to build a crystal receiver fit in the palm of your hand. While there may or may not be a more efficient way, this renewable energy solution is simple to employ and can be scaled up, or down, indefinitely.

It sounds like we can build a Crystal Energy Receiver.© Let's give it a shot...

What You'll Need

Get the Kit

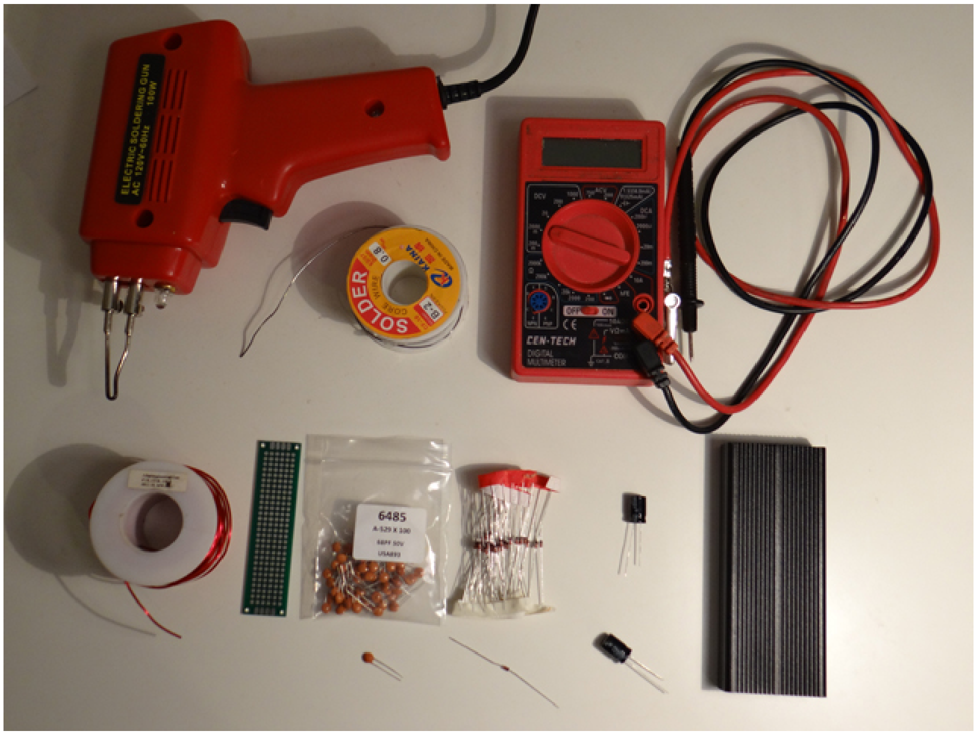
ELECTRIC SOLDERING GUN
AC 120V~60Hz 100W
SOLDER
CEN-TECH
DIGITAL MULTIMETER
6485
A-529 X 100

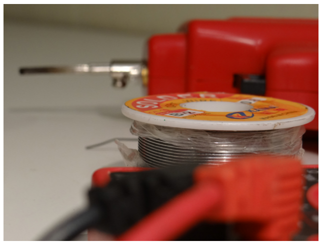

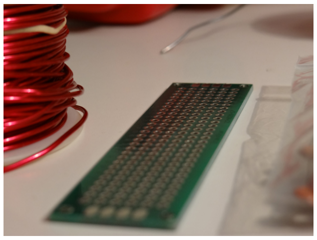

6485
A-529 X 100

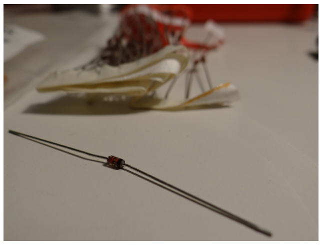

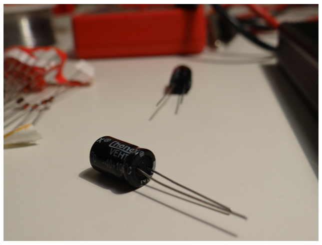

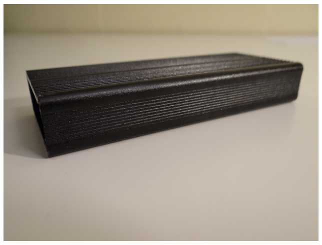

One of the reasons this particular renewable energy harvesting method is so viable is the relatively few and easy to obtain materials required.

The simplest crystal receiver design needs no power and can be built with only three parts: a coil, a crystal and a resistor. We're going to optimize that design in order to produce a cleaner and more reliable output signal by first polarizing the input amplitude, then rectifying and filtering the signal. Then we'll add an antenna, case and connections.

The parts for the circuit include:
(1) Circuit Board
(1) 10-18 gauge Copper Wire
(2-12+) Ceramic Capacitors (matched)
(2-6+) Electrolytic Capacitors (matched)
*note various types of capacitors can be used
(4) Germanium Crystal Diodes (1A+)
Total Cost: +/-**$0.40** (USD, scaled for volume)

In addition, you'll probably want to get:
(1) Project box (optional)
(1) Antenna (a loop antenna or elevated antenna is recommended and can be made with copper wire)

Get the Kit

The tools you'll need are:
Soldering Iron/ Solder (optional)
Multimeter
Oscilloscope (optional)

That's it. Yup, that's all. Once we've got it all, let's begin.

Build the Circuit

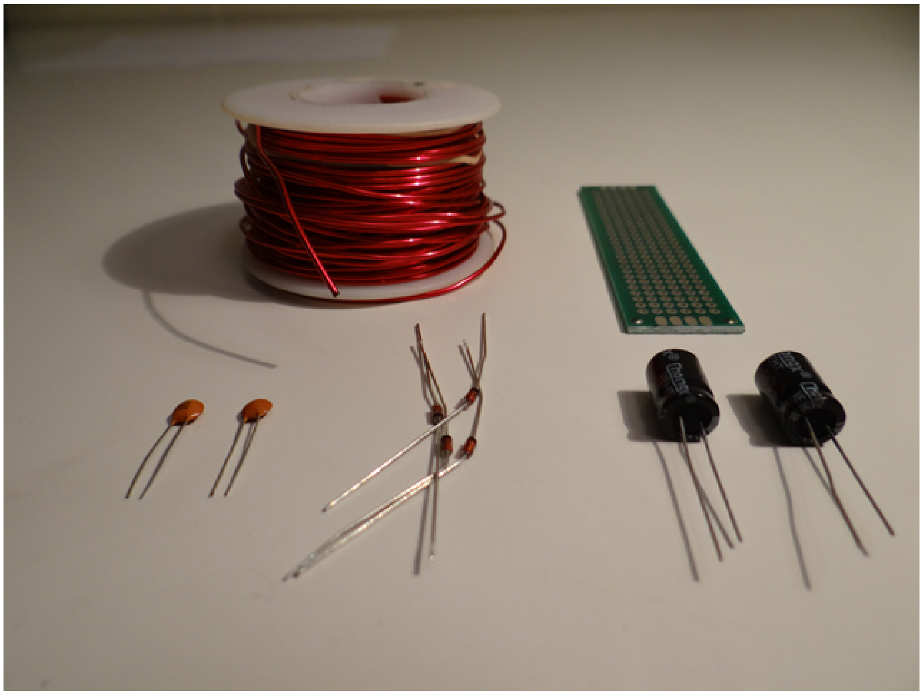

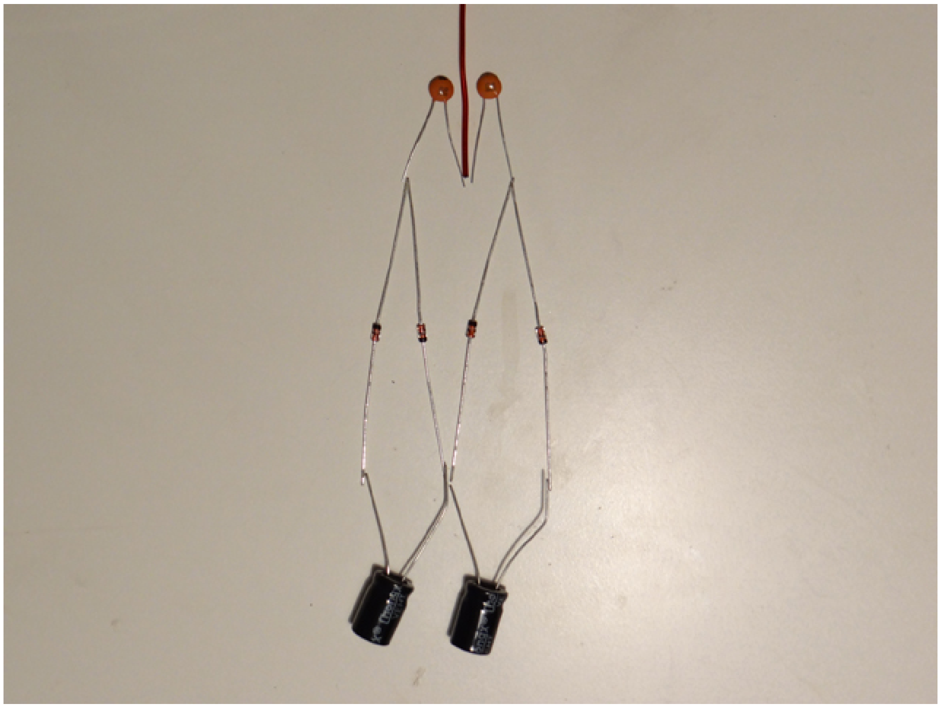

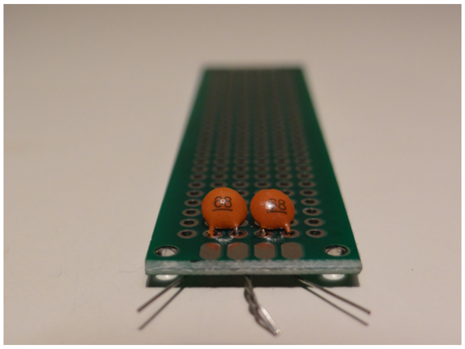

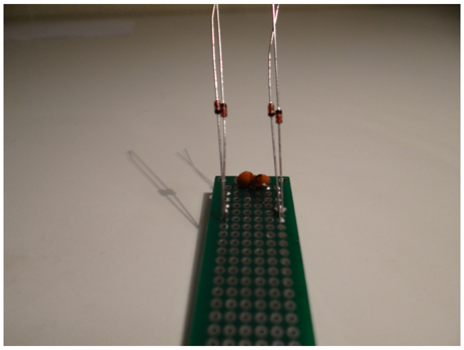

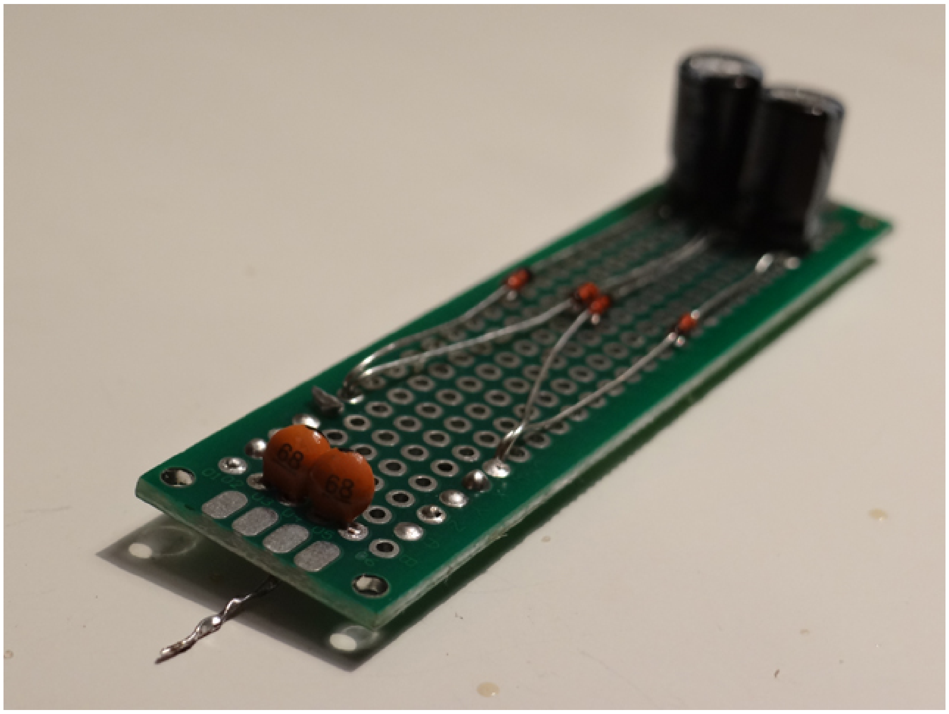

We're going to build the simplest version of this circuit variation in order to understand how each component interacts and as a proof of concept.

There are three simple systems at work in the circuit that are composed of capacitors, which store energy, and diodes that direct it.

Energy in the band of radio waves, among others, will vibrate a wire antenna on an atomic level, sending a discernible signal to its lead. This signal will then meet the junction between two ceramic capacitors wired in series. This junction will force positive charge from the wave to travel in one direction and negative charge in the other direction which, when collected again, makes the signal uniform and polar. Connecting the two capacitors in series creates leads on each end; the now positively charged side of one and the now negatively charged side of the other creates a two cell battery. This stage of the circuit takes a signal with a net value of zero, adds the absolute values of the positive and negative amplitudes with respect to the origin and produces a positive integer. This concept can be thought of as taking:

$$(+1) + (-1) + (+1) + (-1) = \mathbf{0}$$

and converting it to:

$$|\,[(+1) + (+1)]\,| = \mathbf{2}$$
$$\mathbf{+}$$
$$|\,[(-1) + (-1)]\,| = \mathbf{2}$$
$$\mathbf{= 4}$$

Isn't math fun?

To each of these leads from our two capacitors in series, we will connect two crystal diodes, one facing each direction, to form what is called a bridge rectifier. A bridge rectifier is a configuration which will convert an alternating current to a direct one by cleverly rerouting the signal.

By connecting the bridge rectifier as shown in the circuit diagram, this direct current from the diodes then charges the electrolytic capacitors. This stage normalizes the amplitude, making the current constant and usable.

Components can easily be twisted together for testing and then soldered to a circuit board to secure.

Test and Optimize

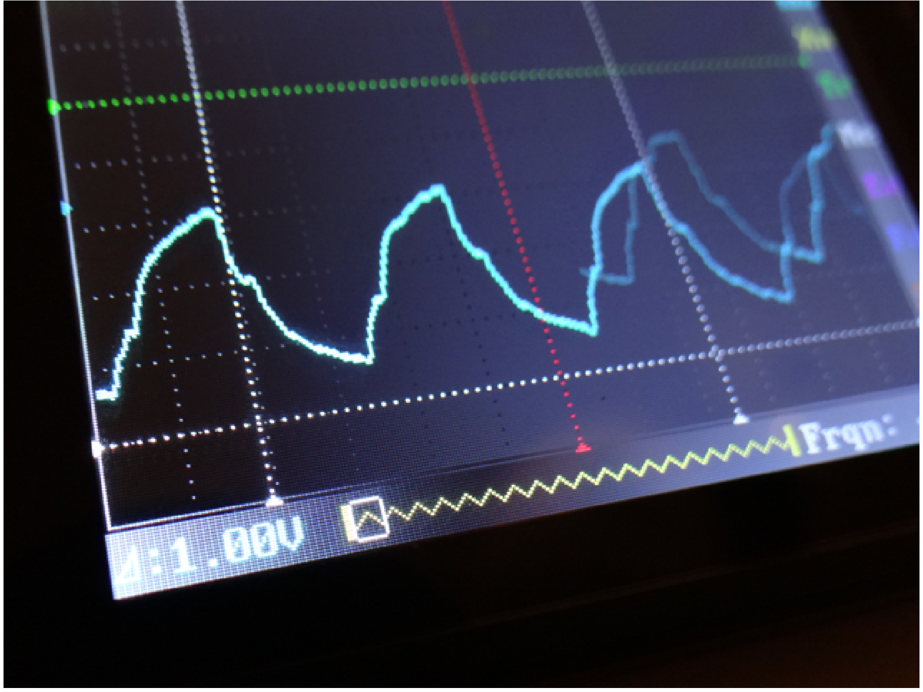

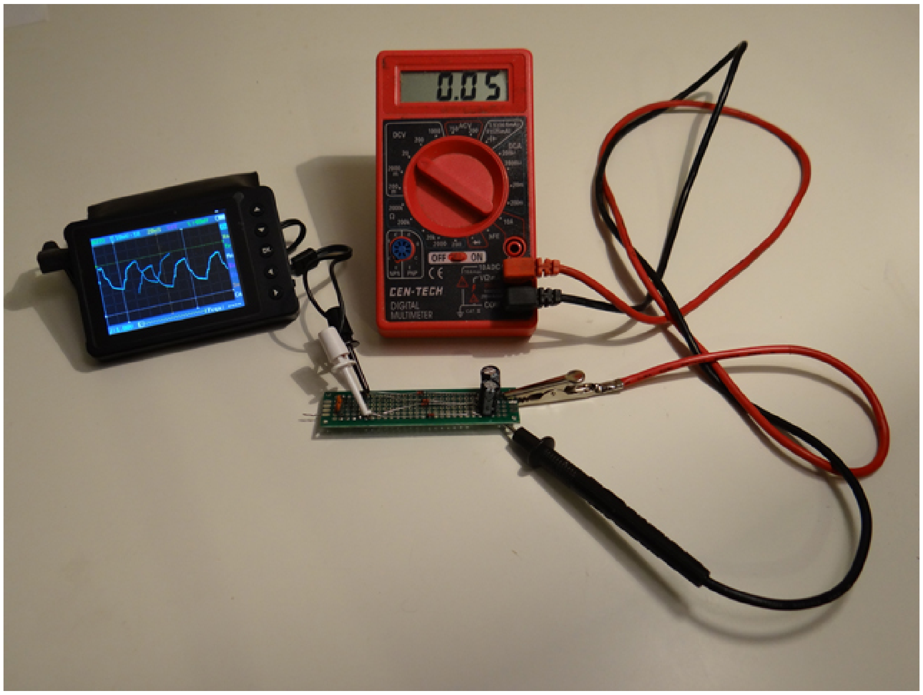
0.05
OFF ON
CEN-TECH
DIGITAL MULTIMETER

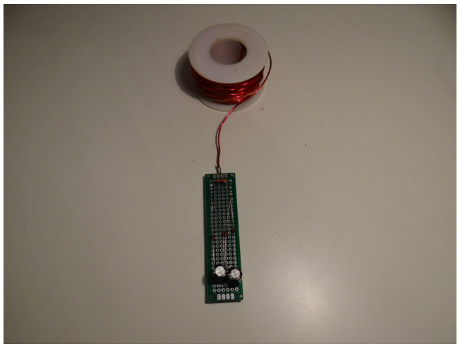

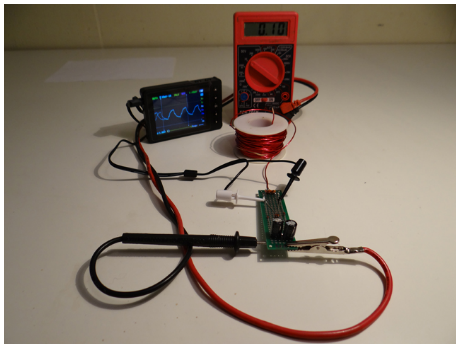
0.10

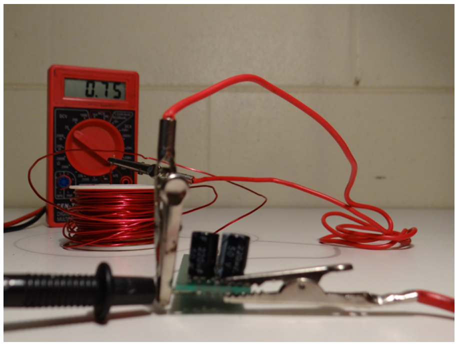
0.75

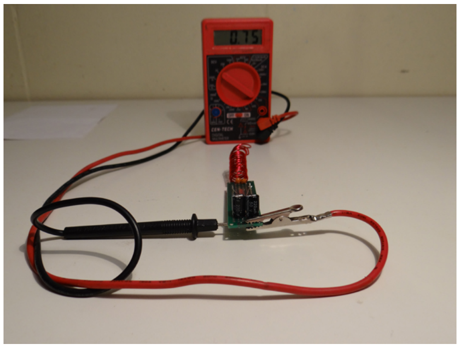
0.75

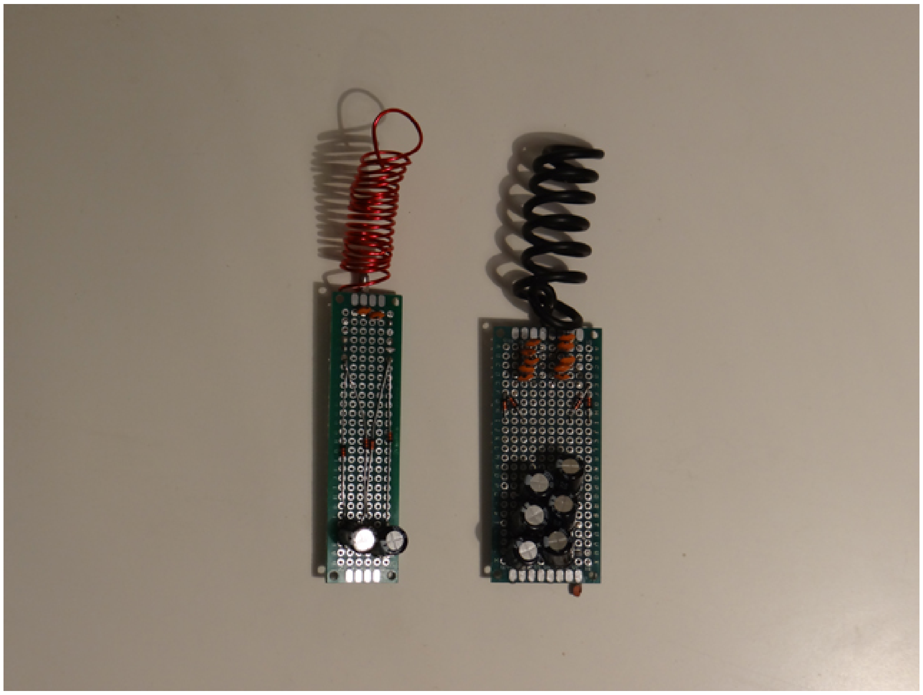

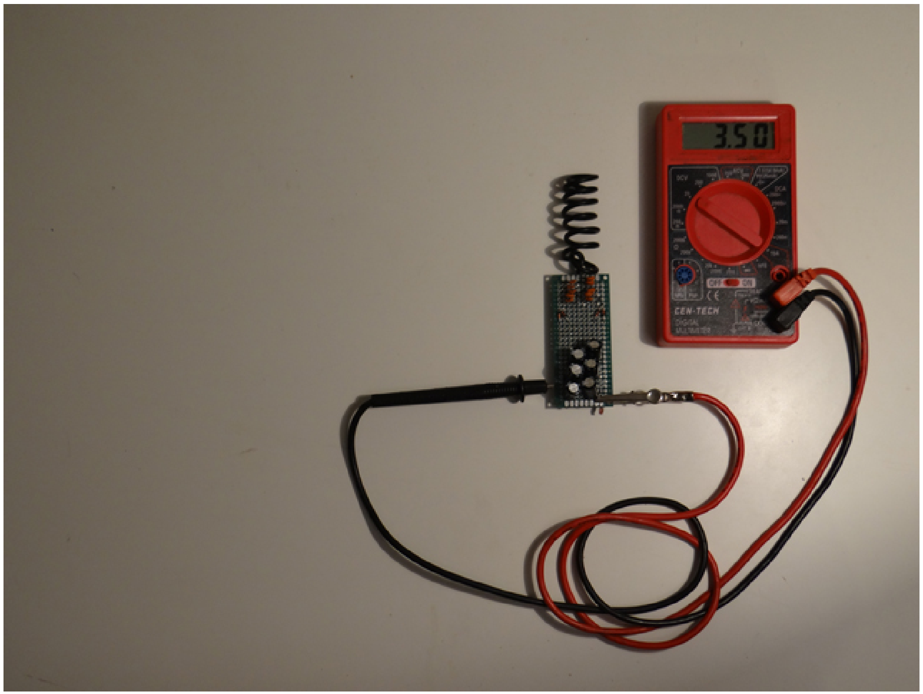

Note: Photos depict tests performed in an open-air urban environment.

To test and analyze our circuit, we'll be using a digital voltmeter and oscilloscope.

Then, by connecting a voltmeter to the output, we'll immediately begin to see a small voltage climbing in the 10-100mV range. If not, we'll want to check our connections and make sure the circuit is not isolated from the environment by taking it outside to a clear area.

By connecting an oscilloscope to the outside leads of our two ceramic capacitor bank, we will see the the polarized signal being captured from the air around us. We can then connect after the diodes to see our varying direct current and then to after the electrolytic capacitors to see a normalized, usable direct current at our output.

We can then optimize the input resistance in two ways. Firstly, we can add additional ceramic capacitors in parallel to our original two and make sure our soldered connections are consistent and thick in this area.

We can optimize the circuit's capacity by adding electrolytic capacitors in parallel to our original two which will allow this circuit to charge slightly when not in use. For this purpose, a charging circuit can also be added here in order to incorporate an optional battery bank.

We can optimize the antenna by attaching loops and coils of copper wire in various positions, store-bought antennas or by stringing some wire up to the highest point you can reach.

We don't have to stop there, either. We can now connect multiple circuits in series to increase voltage or in parallel to increase current. This can be done indefinitely.

Add a Case and Antenna

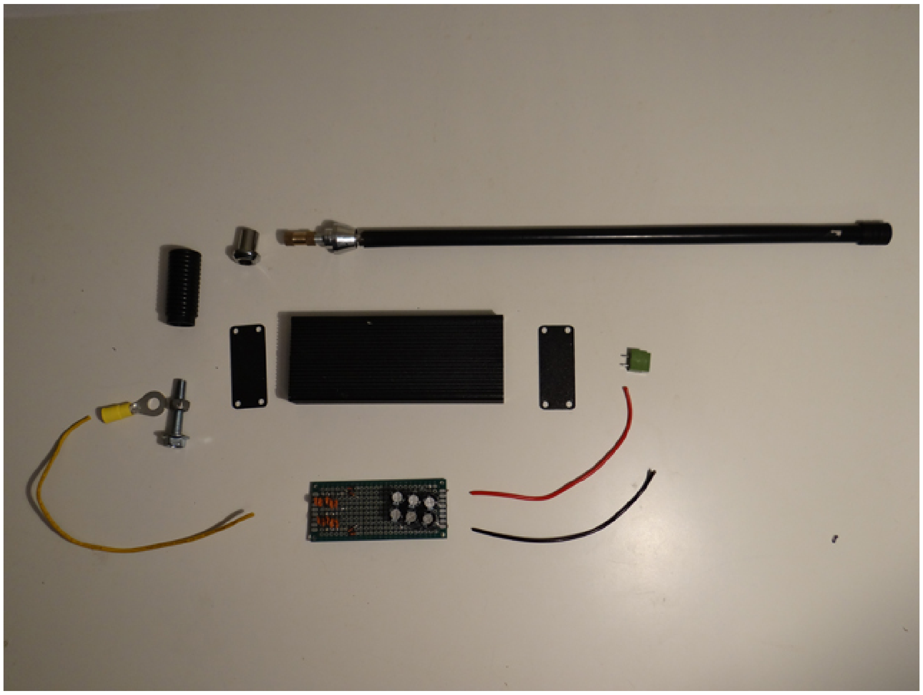

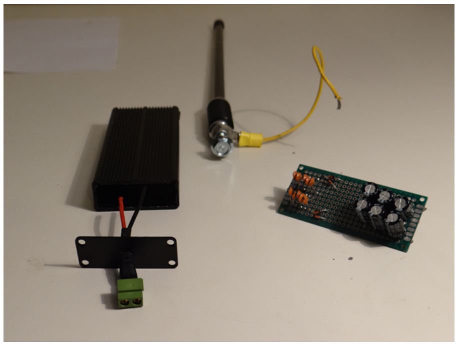

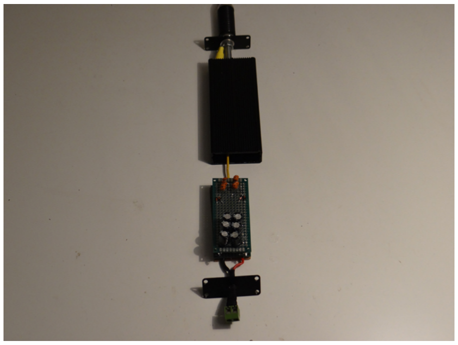

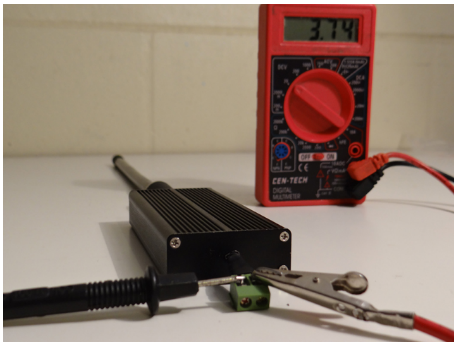
3.74
CEN-TECH

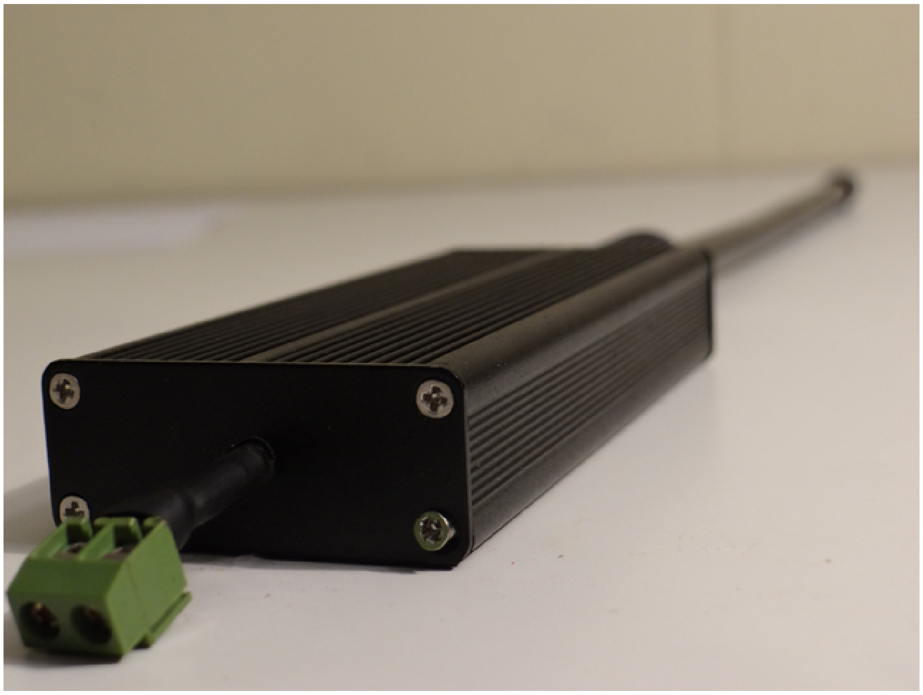

After choosing an antenna in the last step we'll now want to permanently wire it. Whether you choose a compact antenna for portability or a tall fixed antenna for power and range, we will wire it in the same manner according to the diagram in the previous step. Note that the input on the configuration here is grounded to the metallic case, and thus the users hand, and incorporation of longer antennas will require proportionally more substantial grounding.

We will then attach a terminal to the output to allow us to connect this circuit to an electrical device or charging circuit and battery bank.

Next, we will add a case, making sure to isolate exposed leads with non-conductive material especially if mounting in a conductive case. A piece of cardboard secured with glue is sufficient for the circuit's bottom and shrink wrap or electrical tape can be used in the case of any additional exposed leads. Drill two holes in your enclosure, one for the antenna or antenna lead and another for your output terminals. You can then insert your components, fasten the enclosure and your device is ready to use!

Your Crystal Energy Receiver© is Complete

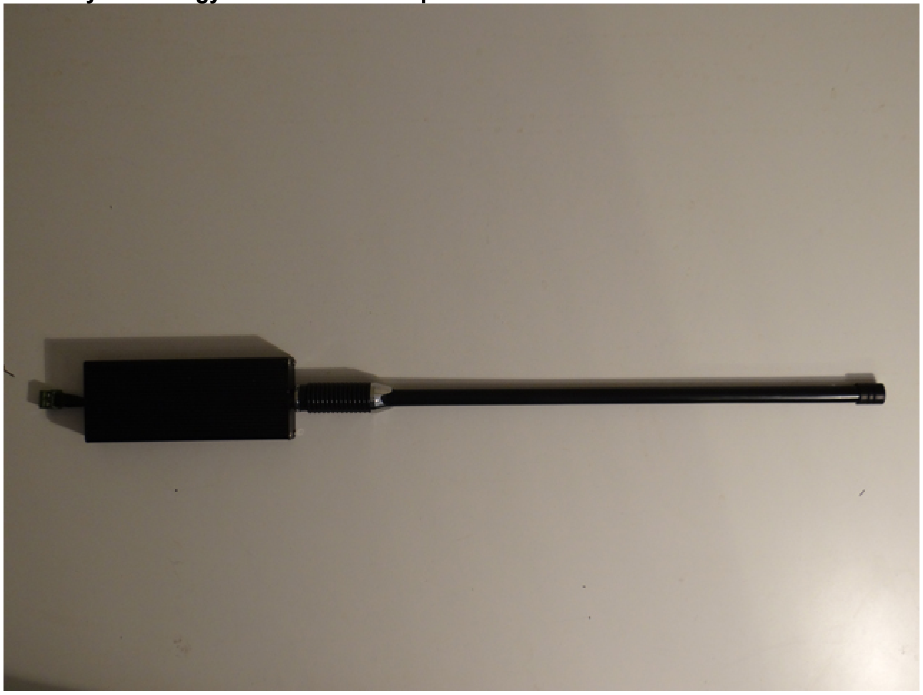

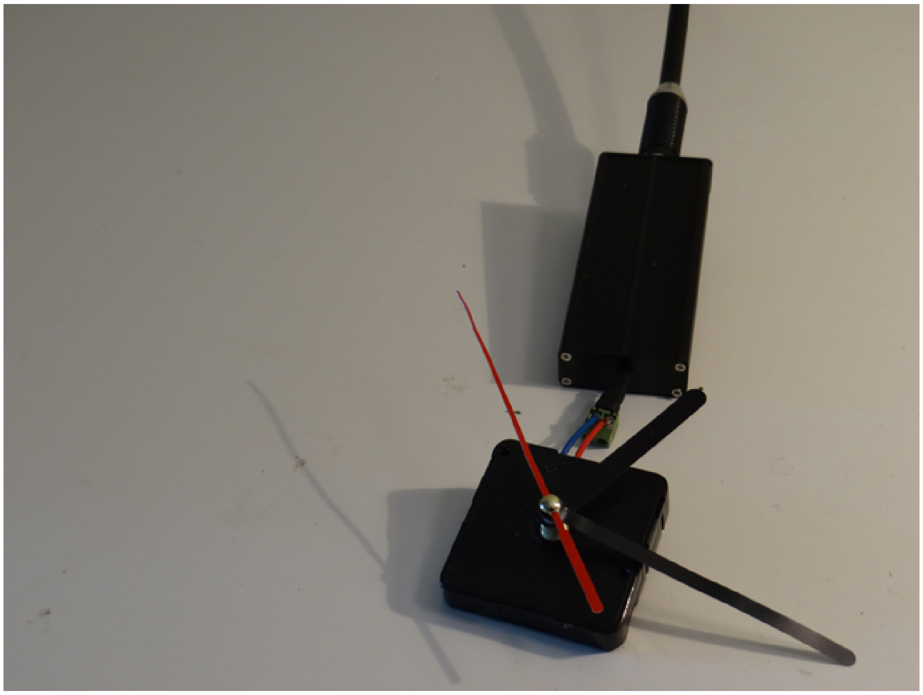

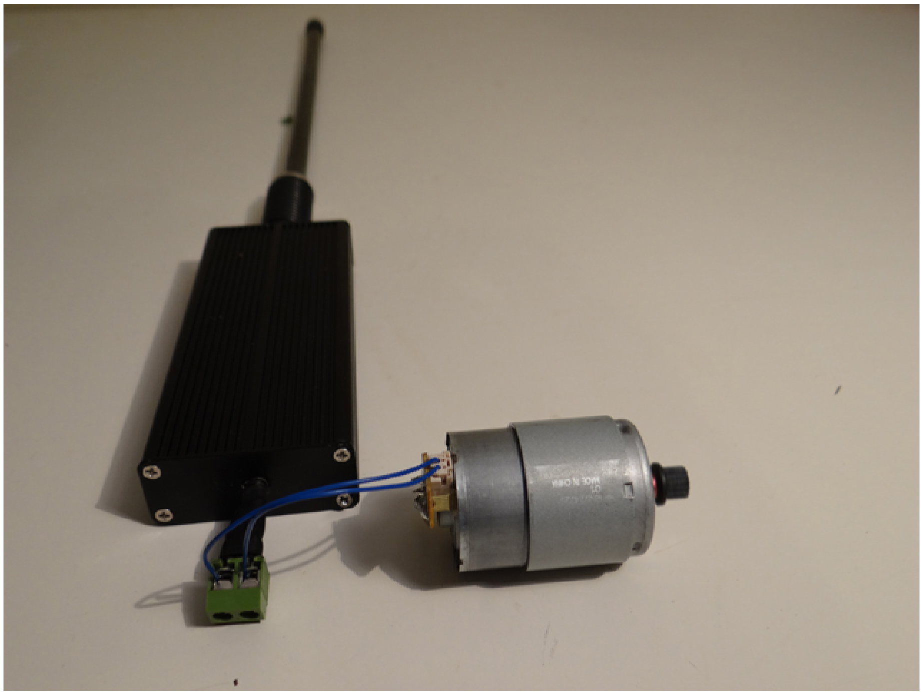

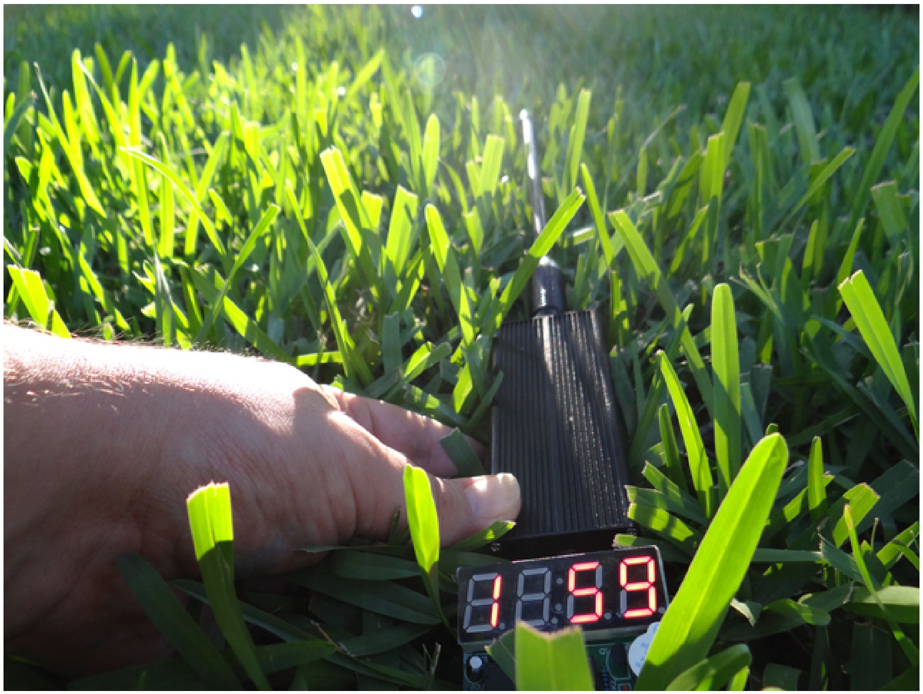

It is hour 11:59; the energy revolution is upon us!

Your Crystal Energy Receiver© is now complete and ready to use!

I built a portable version, for proof of concept and demonstration purposes. However, you can go as big as you want- to passively charge batteries or run equipment remotely; or go as small as you want- to power sensors, RFID devices, small electronics and more.

I used this harvested energy to easily power a low-consumption quartz clock, a digital chronograph with integrated circuits and LCD and was even able to momentarily rotate a small dc motor.

Because of its simplicity this device is a durable, efficient and reasonably effective method of harvesting radiant energy in a simple, replicable and sustainable way. I humbly hope that the contributions made here, and by those reading, can be one day used by people worldwide to conveniently capture free energy.

Want to build it? **Get the Kit!**

See more **Projects**

Copyright 2016. Drew Paul. All rights reserved.

Please take caution. User is responsible for safe operation of components.

DREW PAUL DESIGNS

WEB | PRINT | MARKETING | MEDIA
by
blabor.com

Crystal receiver set 11

Step 1   preparing the components

Back to the index   back to set 11

Preparation of the tuning capacitor

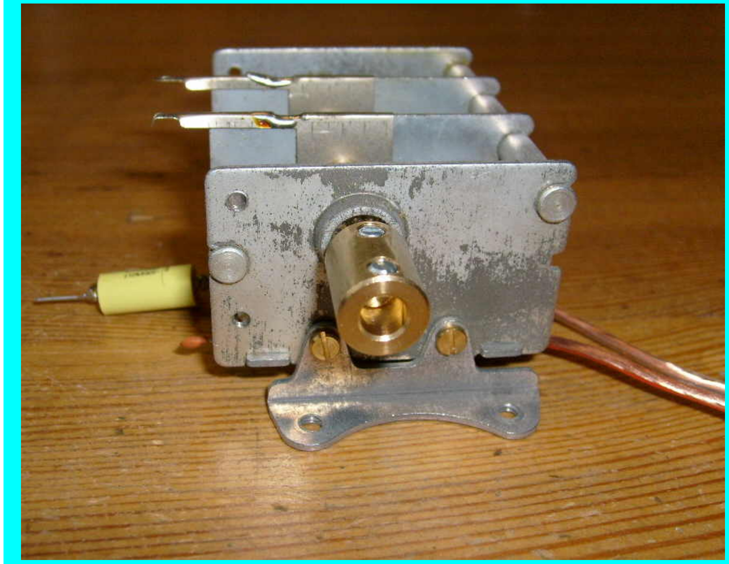

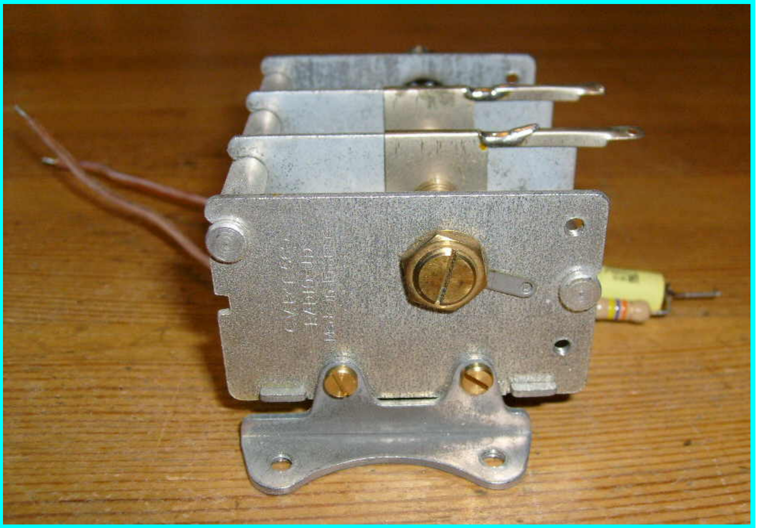

Mount the mounting supports on the tuning capacitor (they are delivered with the tuning capacitor).
Place the shaft coupler on the shaft of the tuning capacitor.

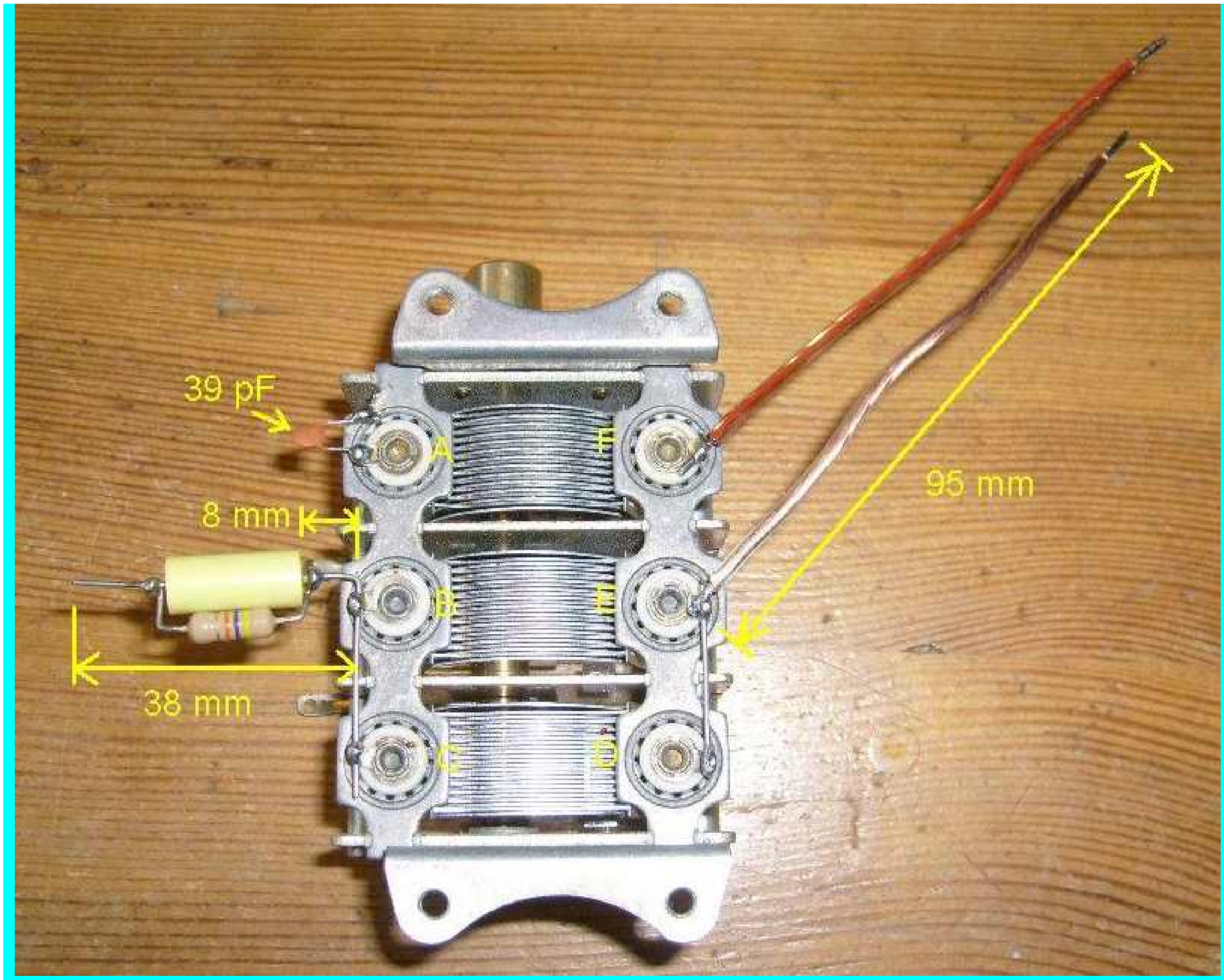

Lay the tuning capacitor upside down (in this picture, the shaft is at the upper side).
Bend the 6 solder tags (A,B,C,D,E and F) upwards.

Solder a 39 pF capacitor between solder tag A and the frame of the tuning capacitor.
Scrap (if necessary) the oxide from frame and solder tag, to make the soldering easier.

Mount the 47 kΩ resistor parallel to the 68 nF capacitor.
The wires of the resistor are turned around the wires of the capacitor, and then soldered together.

Bend the wire of the capacitor in a right angle, at 8 mm from the capacitor.
Solder the 68 nF capacitor with one wire to the points B and C of the tuning capacitor, as shown on the picture above.
Cut of the other wire as shown.

Solder the cut off wire between the points D and E of the tuning capacitor.

Take two pieces of wire with transparent insulation with a length of 95 mm.
Solder one wire to point E.
Solder the other wire (with the red line on it) to point F.
The wires must point in the direction as shown on the picture.

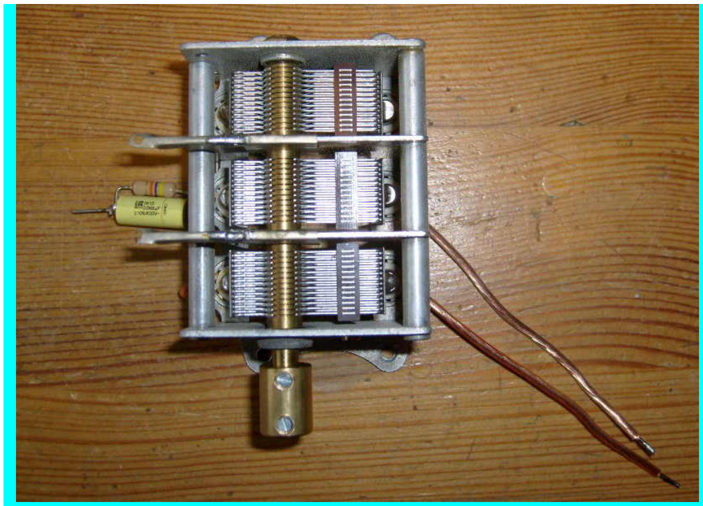

Top view of the tuning capacitor.

Preparations on the transformer.

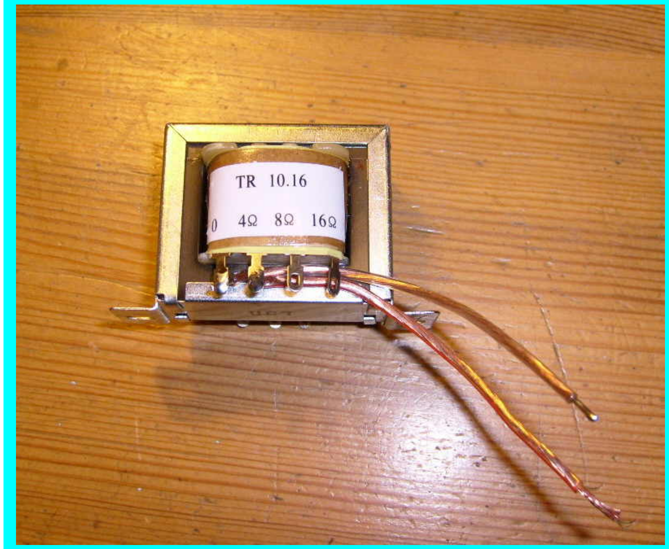

Solder two wires (length 95 mm) with transparent insulation to the transformer.
One wire to the "0" connection.
The other wire (with the red line), to the 4Ω connection.

Preparing the wooden sticks.

Saw 2 sticks with a diameter of 9 mm to a length of 100 mm.
Drill in each stick 2 holes of 3 mm.
One hole at 6 mm from the end, the other at 81 mm from the end.
Lacquer the blanc wooden sticks with brown lacquer, and then one time with blanc varnish.

Take 2 sticks of about 2.5 mm thickness and about 170 mm long.
Lacquer the sticks with brown lacquer.

Saw the sticks in two.

Preparation of the headphone socket.

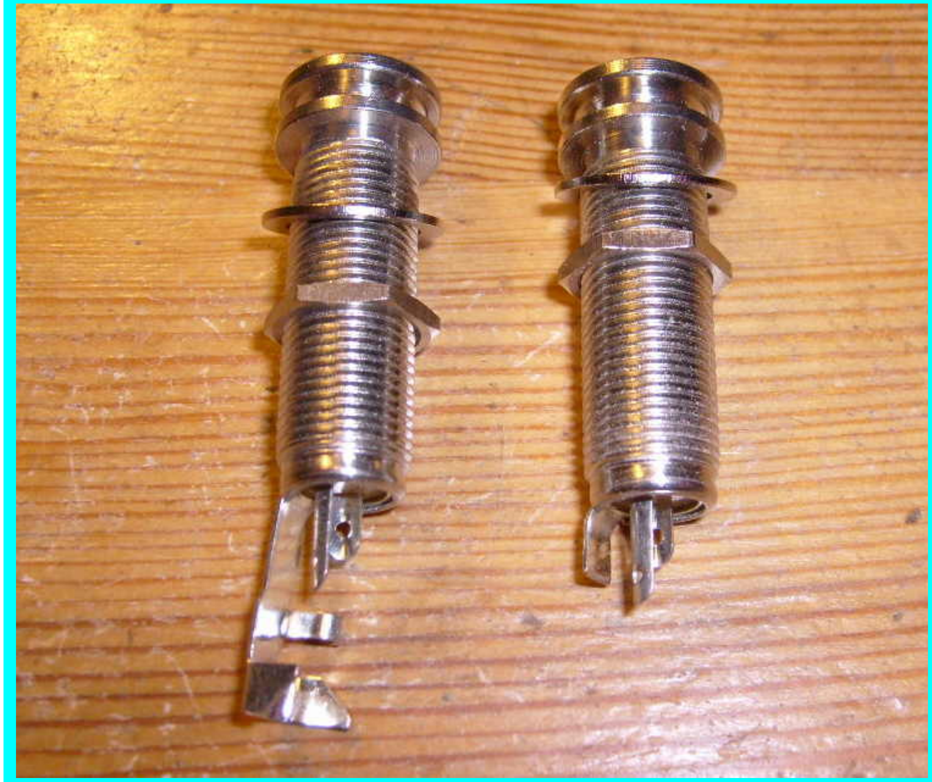

The socket for the headphone must be modified as follows:
remove the back part of the ground connection.
Left: the original socket.
Right: the ground connection is shortened.

[Back to the index](#) [back to set 11](#)

Essential Technology

Home About Contact Drupal Recent posts

Crystal Sets

Posted on September 4th, 2010

s a child, my father got me into building my first crystal set. I don't fully recall the details, but I seem to recall it was a kitset and I was in Nelson when I built it - If I remember I'll ask him. What I do remember is building a few of them, including one that fitted inside a tic-tac container and another that had no tuning cap, just an icecream stick with a screw through the end that scratched through to vary the coil windings in use.

Anyway, I've recently starting teaching my daughter how to make them, and how they work - and I thought it might be useful to someone to show what steps we went through (and are still going through), which sites were helpful, and what parts to use.

When I first started looking around for a kitset I was disappointed to find there was little or nothing available here in New Zealand... although Jaycar carry a kit (shown to the right) - but its not exactly educational - nor is it built in the simple spirit of crystal radios (ie: using a minimal amount of parts scrounged from a box of used bits and pieces and screwed to an old slab of wood). It also has a very narrow tuning range due to only using a small trimmer capacitor rather than a full tuning capacitor. It did tune in a station, however, and the look of amazement and excitement on my daughters face when she could actually hear it was enough to get me thinking, googling and researching.

I don't recall all the sites I came across while looking for more information, but some of the most useful are listed at the end. What I discovered was that spider coils were the new thing, where coils wound on a ferrite rod (similar to the Jaycar kitset one) were in-vogue when I was young. Spider coils - Higher 'Q' than a cylindrical coil due to less interwinding capacitance. Don't sweat the technobabble - but just know that spider coils work better, and are easier to wind - especially for small hands.

So, I made a former up by drawing it in TurboCAD, printing it, gluing it to some corrugated cardboard then cutting it out with a sharp hobby knife. I used a small jewellers screwdriver to make some holes to feed the wire in (24 gauge enamelled copper wire), and the tuning cap. I then soldered everything together in place.

Note: The tuning cap has two gangs - a 60pF and a 160pF one. By soldering the two outside wires together and using that as one connection, and the center wire as the other we get 220pF... not quite enough, but it works.

The actual winding details for the coil I calculated with Dan Petersens calculator (Version 5.7 is here - its all imperial, but it does the job). What I worked out was that around 56 turns would be good with the bog-standard 60-160pF variable capacitor that you can get at Jaycar. It should, by my calculation, cover the entire AM band.

The circuit I settled on is about the most basic possible - and the performance was pretty good - and coverage was excellent. With a little tweaking of the two trimmers on the back of the tuning cap (sometimes called a polyvaricon) I could tune both Radio Live (570ish at one end and Coast AM (1590ish) at the other... a pretty good spread for such a basic design. Whats even more impressive is how little aerial it needed (although we are in the city, and Coast in particular is transmitted very close to us).

The 47k resistor is there for the crystal headphones. If the resistor isn't there then you probably won't hear anything. If you have old-school vintage headphones then you can remove the resistor. Here are better pics of the front and back.

FM Crystal Set - Experimental

This is one I built as an experiment to see if it would work... no joy yet, but I'm still trying. Its based on this one: http://solomonsmusic.net/FM_CrystalRadio.html

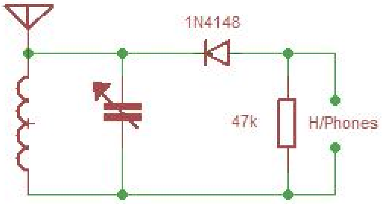

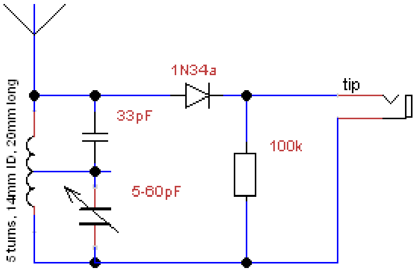

Site Search:

Android Arduino Clean URLs Computer Science Crystal Set Drupal Drupal 7 Error Google Google Maps LCD Linux Map MySQL PHP PIC PIC16F628A Python Radio Spider Coil Sun Tablet Ubuntu university

more tags

Recent posts

- Vintage Ekco RCX 'Ekcosprite' - Volume won't...
- Typewriter effect in Python
- Arduino Traffic Light Example
- Murphy (F&P) Service Information for various models
- Philco 'Alabama' - NZ Model
- Android App Inventor - Using "List From CSV ROW"...
- Christchurch House / Section For Sale Map (now New Zealand...
- First play with an Arduino Uno and an HC-SR04 Ultrasonic...

more

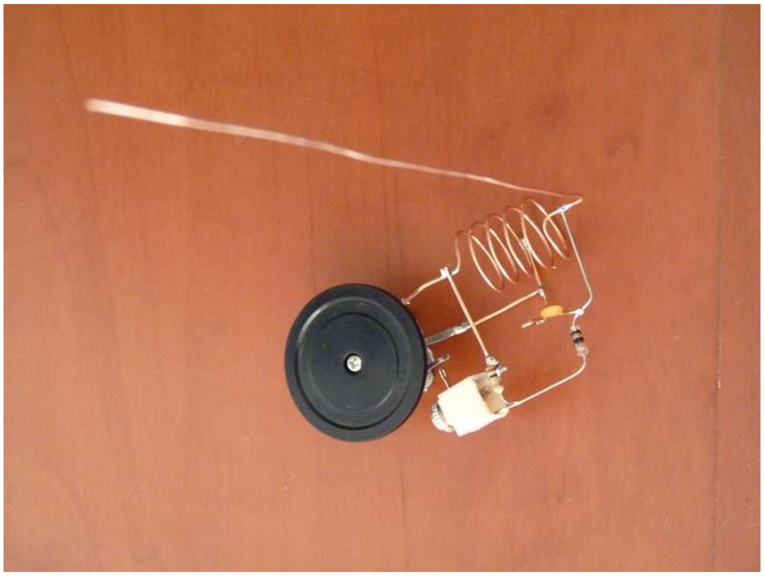

FM Crystal Set - version 2

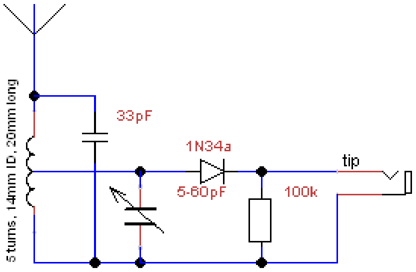

I built this version after a little advice on The Radio Board... but yet again, no joy. Still tinkering tho. (Oh, there is a typo - I used a 10pF cap over the coil, not 33pF).

Links:

http://makearadio.com - Dave Schmarders site - An absolute crystal set craftsman

http://crystalradio.us - The Birmingham Alabama Crystal Radio Group - lots of sets to look at

http://peeblesoriginals.com/ - Plenty of hard-to-find parts

http://bentongue.com/ - Incredibly clever guy, lots of good theory

http://theradioboard.com/rb/viewforum.php?f=2 - The Radio Board - Crystal Set Forums

http://www.daycounter.com/Calculators/Air-Core-Inductor-Calculator.phtml - air-cored L calc

http://electronbunker.sasktelwebsite.net/InductanceCalc.html - advanced air-cored L calc

http://electronbunker.sasktelwebsite.net/Bandspreading_3.html - Bandspreading

UPDATE: More info here

Tags: 1N34 Crystal Set General Radio Spider Coil TPR

Site design, layout and content © Essential Technology, 2012 - except where noted.

Simply Symmetry

Crystal Radio for Casual Listening

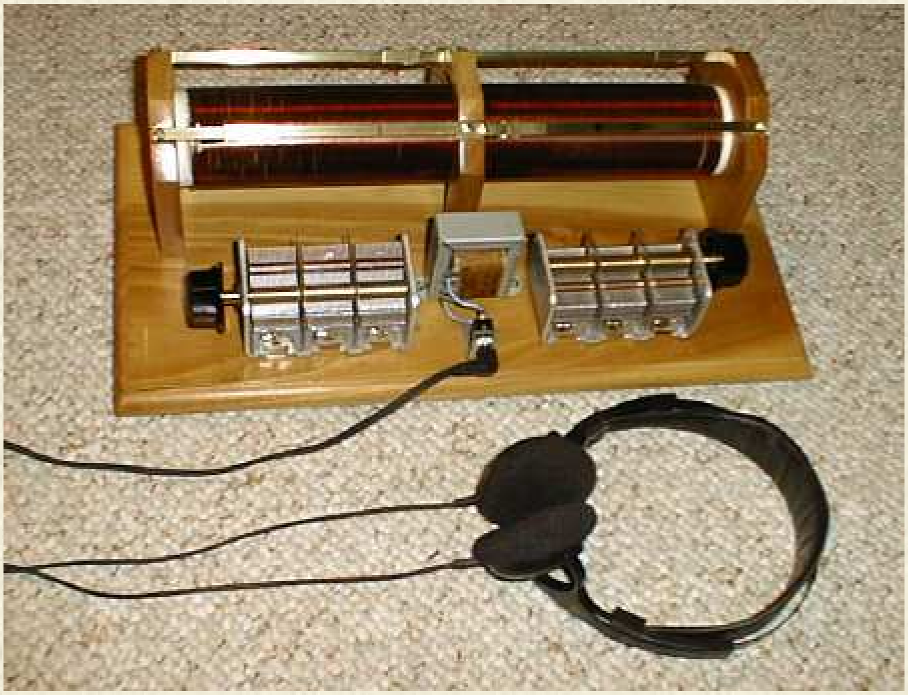

Having been inspired by a building contest, I built this radio in January and February 2001 for use in my workroom. This is a crystal radio, using no outside power or batteries. It derives its power from an outside antenna strung across the roof of my house. It is quite sensitive, and it's efficient enough to drive a speaker directly from the earphone jack (that's right, no amplifier), which never fails to astound me. With Walkman-style headphones, the fidelity is awesome, rivaling FM. Who would think AM could sound so good!

The radio is not the only thing that must be sensitive. I must be, too. Because my workroom is visible from the kitchen and living room, it's imperative that I include the SAF (Spousal Acceptance Factor) into my design specifications. I chose to use all new parts in this project. The wood is poplar, browned in the sun, covered in two coats of laquer, sanded and finished with paste wax. You can glimpse a hint of its beauty in the images.

Crystal Radio enthusiasts will note the striking absence of wiring. Crystal radios are so simple, they are often built with exposed wiring. I've seen enough of those I decided to try something that looked truly wireless. It also helped me get close to my design goal of perfect symmetry.

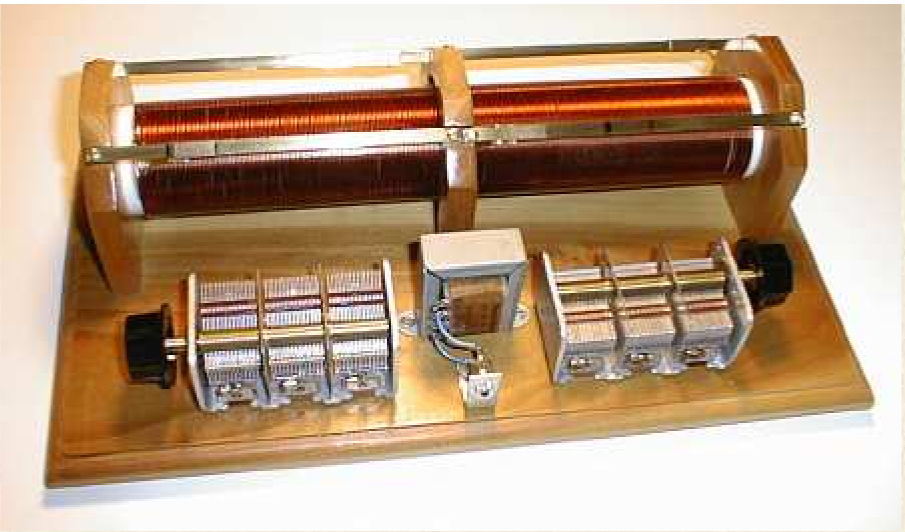

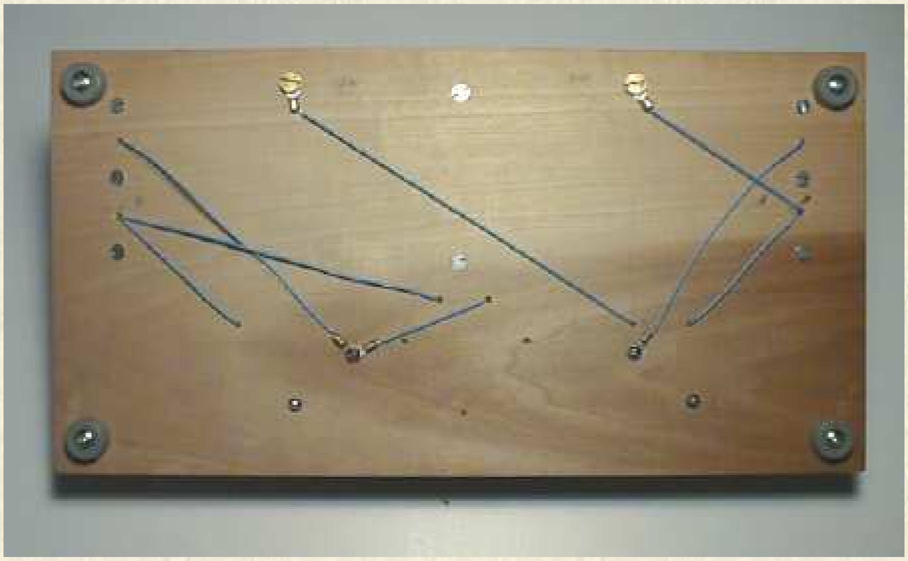

The tube is a 12 inch section of 2" PVC pipe. It was wound as one long coil, then cut in the middle. (Sorry, I don't recall how many turns, but don't worry, it's not critical.)

The wires to the slider bars run down internally through the wood supports. All other wiring is on the underside of the wood base. Wire emerges through small holes on the back sides of the variable capacitors and transformer.

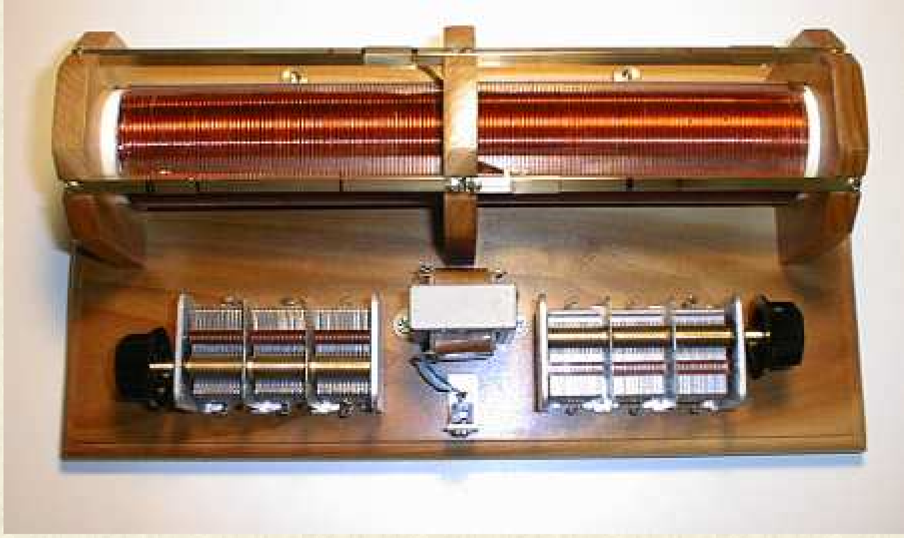

This view shows the antenna and ground binding posts at the back edge. These are made of brass 10-32 screws and big knurled nuts. The knurled nut bears down on a brass washer with its underside soldered to a 10-32 hex nut. The wood is countersunk just enough to receive this hex nut so that when the screw is tighened from the bottom, it draws the nut/washer assembly tightly together.

This view shows the slider assembly. It uses square brass tubing with a strip of sheet brass soldered to it. Note that the insulation has been removed from

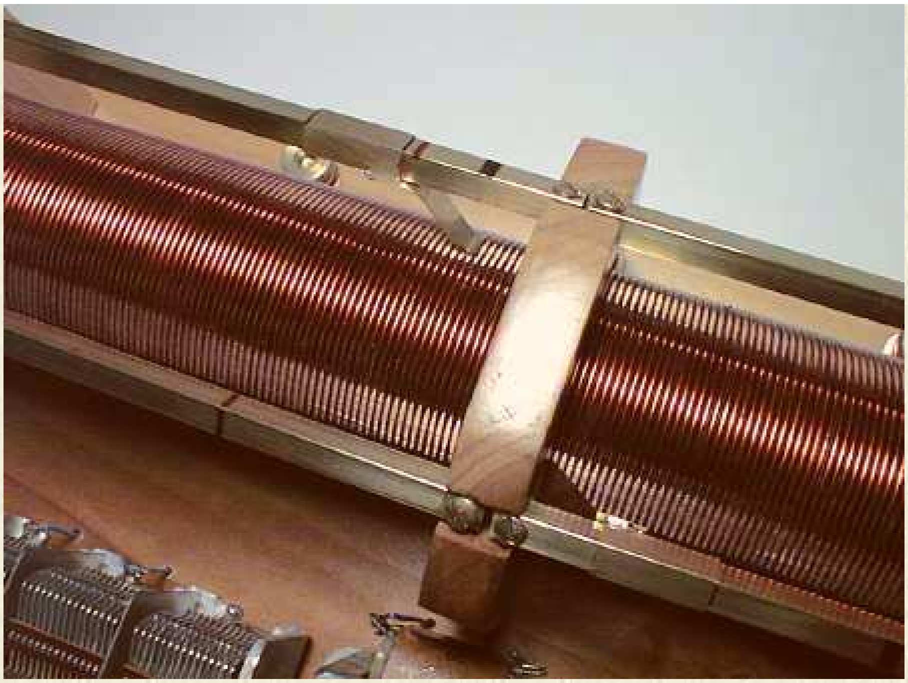

the 16 guage wire along the paths of the four sliders. Note the red and black ink on the square tubing, marking the best slider positions.

Barely visible at the bottom left and center are blue wires emerging from holes in the wood where they are soldered to the variable capacitor and the transformer. Visible toward the top left is the brass knurled knob for the antenna connection.

Schematic

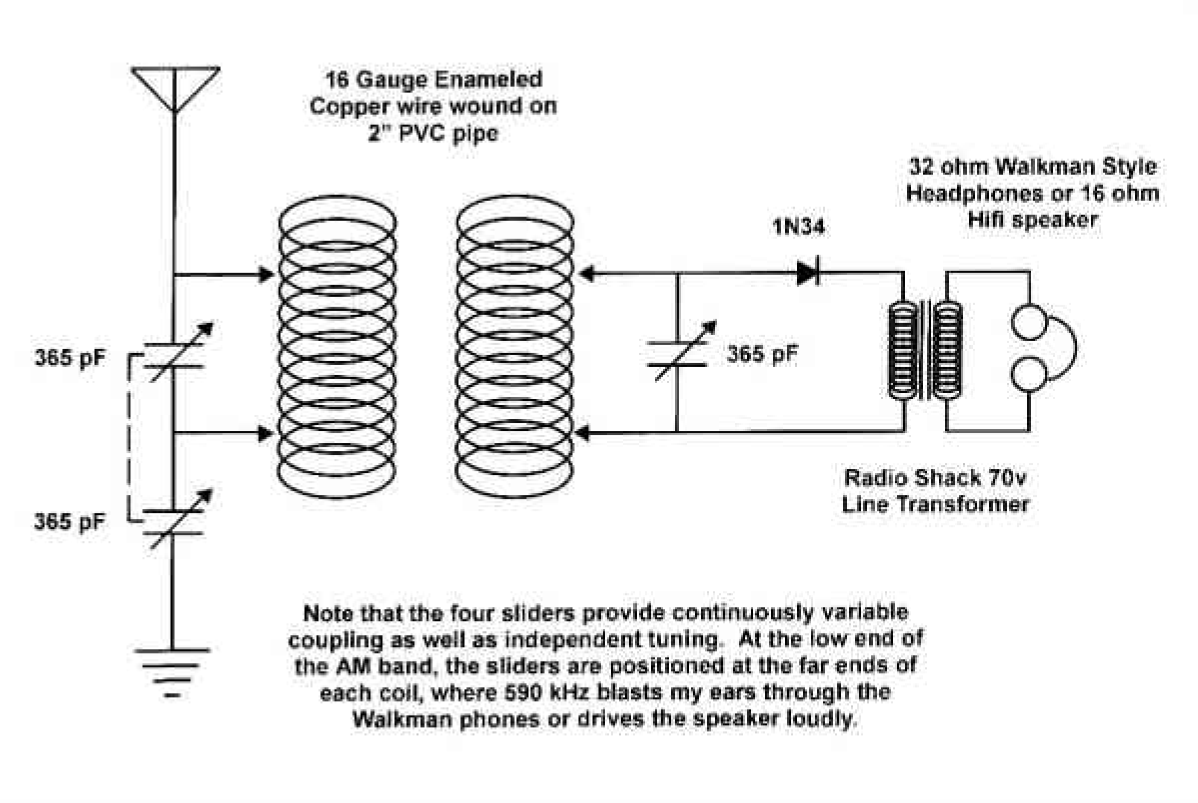
16 Gauge Enameled
Copper wire wound on
2" PVC pipe
32 ohm Walkman Style
Headphones or 16 ohm
Hifi speaker
1N34
365 pF
365 pF
365 pF
Radio Shack 70v
Line Transformer

Note that the four sliders provide continuously variable coupling as well as independent tuning. At the low end of the AM band, the sliders are positioned at the far ends of each coil, where 590 kHz blasts my ears through the Walkman phones or drives the speaker loudly.

Update: 11-29-03

I don't sell plans or instructions for this radio. I built it a while back, and just about everything I know about it is featured on this web page. If you want to build a radio like this, please read the text carefully. If you have a hard time figuring out how to build yours from what is provided here, a crystal radio kit (available elsewhere) is probably a better choice for you. Although this page is not intended to be used as a set of plans, it will have to suffice for

do-it-yourselfers. As an alternative, you might try building another type where I do provide instructions.

One more thing: For this radio to work you MUST have a good antenna and ground connection. Those are not covered here. There are serious safety considerations involved. You can find such infomation on the web.

Home

How to make a batteryless (crystal set) radio

By **Sagar Sapkota** - July 21, 2012

CRYSTAL SET

The simplest radio receiver, known as a Crystal Set, consists of nothing more than a coil, tuning capacitor, diode detector, and a pair of earphones. A typical circuit diagram for a Crystal Set Radio is given below where inductor or coil L1 is tuned by variable capacitor VC1 to the transmitter frequency. Diode D1 demodulates the signal, which is fed straight to the earphones. There is no amplification.

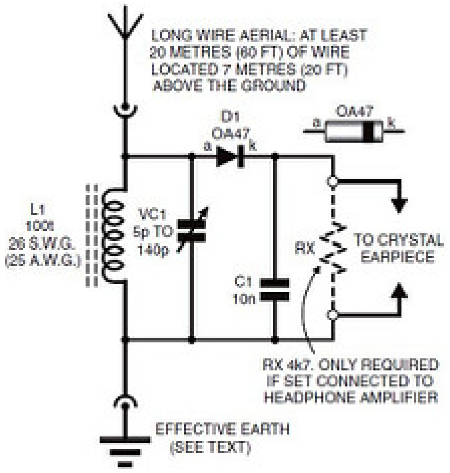
LONG WIRE AERIAL: AT LEAST
20 METRES (60 FT) OF WIRE
LOCATED 7 METRES (20 FT)
ABOVE THE GROUND
D1
OA47
a
k
OA47
a
k
L1
100t
26 S.W.G.
(25 A.W.G.)
VC1
5p TO
140p
RX
TO CRYSTAL
EARPIECE
C1
10n
RX 4k7. ONLY REQUIRED
IF SET CONNECTED TO
HEADPHONE AMPLIFIER
EFFECTIVE EARTH
(SEE TEXT)

CLICK HERE FOR BIGGER PHOTO

** You can use 1N34 Germanium diode in place of OA47*

A long (at least 20 metres), high (17 metres or more) aerial and a good earth (a buried biscuit tin or a metre of copper pipe driven into damp ground) are required in order to ensure audible headphone reception. The earphones originally used with these receivers had an impedance of around 4000 ohms and were very sensitive (and heavy and uncomfortable). They are no longer available, but a crystal earpiece, which relies on the piezoelectric effect, will give acceptable results.

Low impedance "Walkman" type earphones are NOT suitable.

Component details:

Resistor- RX-4.7k, 0.25 W- only required if set is connect to audio amplifier

Capacitors- C1- 10nF disc ceramic VC1- 5p to 140p, polythene dielectric variable capacitor.

Semiconductors: D1: OA47 or 1N34- Germanium Diode

Miscellaneous: L1- Ferrite Rod, 100mm(4 inch)x9mm/10mm dia., with coil.

Crystal earpiece and jack socket to suit; plastic control knob; plastic insulated flexible cable for aerial wire, downlead and earth connection, 30 meters minimum; buried biscuit tin or 1 meter of copper pipe for earth system; 50gm reel of 26SWG enamelled copper wire, for tuning coil; card and glue for coil former; multistrand connecting wire; crocodile clips or terminals for aerial and earth lead connection; solder, etc.

READ MY EXPERIENCES- HOW EASILY I MADE A BATTERYLESS RADIO

DRAWBACKS

Quite apart from the absence of amplification, two factors seriously limit the performance of crystal receivers. Germanium diodes become increasingly reluctant to conduct as the applied voltage falls below 0.2V, and this makes the receiver insensitive to weak signals. Silicon diodes have a threshold of around 0.6V, and are, therefore, unsuitable for circuits of this kind.

** Silicon diodes like 1N4001 are not suitable*

The earphone loading imposes heavy damping on the tuned circuit, hence, reduces its ability to separate signals. With such low selectivity insensitivity can be a blessing, and crystal sets are normally only capable of receiving a single, strong transmission on the long and medium wavebands. They will sometimes receive more than one if a shortwave coil is fitted.

The aerial and diode can be connected to tappings on the tuning coil in order to reduce damping, but the improvement in selectivity is usually at the expense of audio output. When valves cost a week's wages and had to be powered by large dry batteries and lead/acid accumulators, the construction of simple receivers of this kind could be justified. With high performance transistors now costing only a few pence or cents, crystal sets are now regarded as 'nostalgic pieces". Some readers may, however, wish to build one out of curiosity, or for the novelty of having a receiver that **does not require a power supply.**

CIRCUIT DETAILS

Ferrite loop aerial L1 and polythene dielectric variable capacitor VC1 form the tuned circuit. Point contact germanium diode D1 (1N34 or OA47) demodulates the signal; capacitor C1 bypasses residual r.f. (radio frequency) to earth and also exhibits a reservoir action, enabling the a.f. (audio frequency) output to approach its peak value. The recovered audio signal is fed directly to a crystal earpiece. Signal voltages introduced in the ferrite loop aerial by the radiated magnetic field are much too much to produce an output from the detector, and the component is used here simply as a tuning coil. The ferrite core does, however, reduce the number of turns required for the coil winding, thereby reducing its resistance and increasing its audio quality.

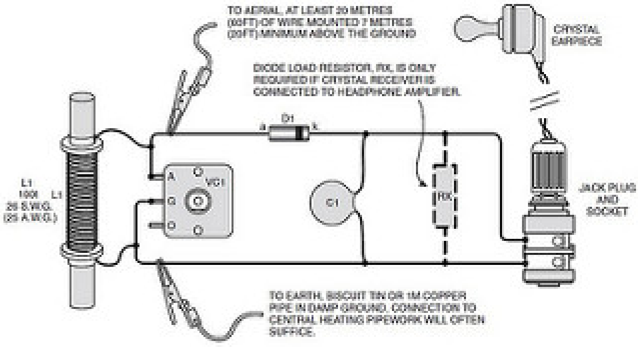
TO AERIAL, AT LEAST 20 METRES (60FT) OF WIRE MOUNTED 7 METRES (20FT) MINIMUM ABOVE THE GROUND
CRYSTAL EARPIECE
DIODE LOAD RESISTOR, RX, IS ONLY REQUIRED IF CRYSTAL RECEIVER IS CONNECTED TO HEADPHONE AMPLIFIER.
D1
a
k
L1 100t 26 S.W.G. (25 A.W.G.)
L1
A
VC1
G
O
C1
RX
JACK PLUG AND SOCKET
TO EARTH, BISCUIT TIN OR 1M COPPER PIPE IN DAMP GROUND. CONNECTION TO CENTRAL HEATING PIPEWORK WILL OFTEN SUFFICE.

GET BIGGER PICTURE

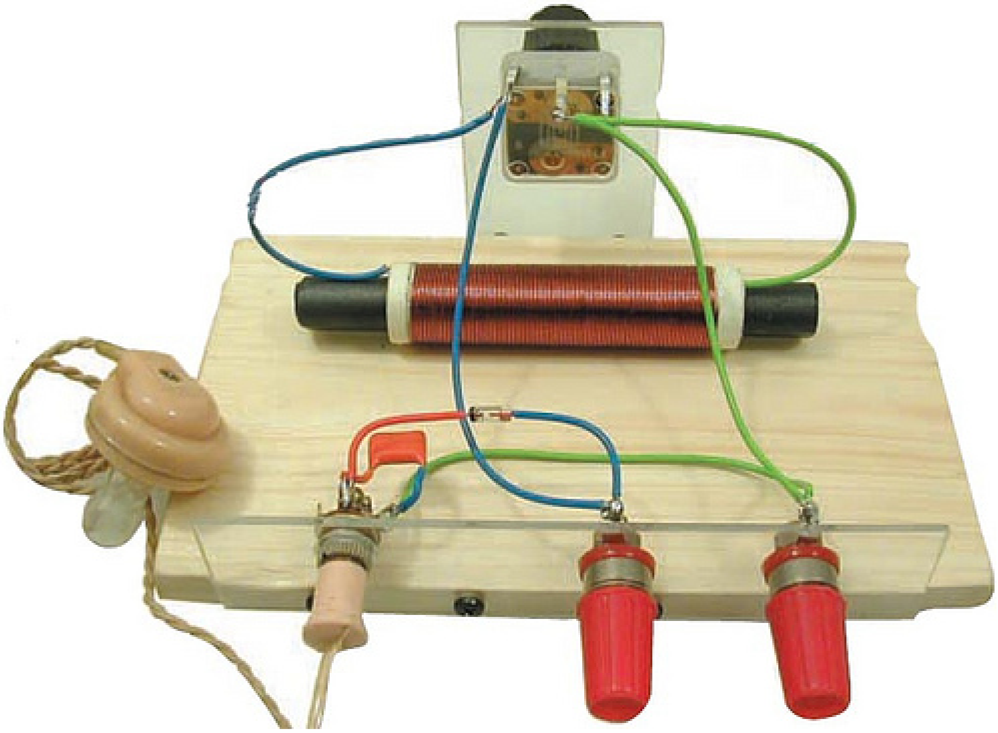

COIL DETAILS

Full construction and winding details for the ferrite tuning/ aerial coil L1 are shown in figure.

The coil is made from 26 SWG(25 AWG) enamelled copper wire, close wound on a cardboard former.

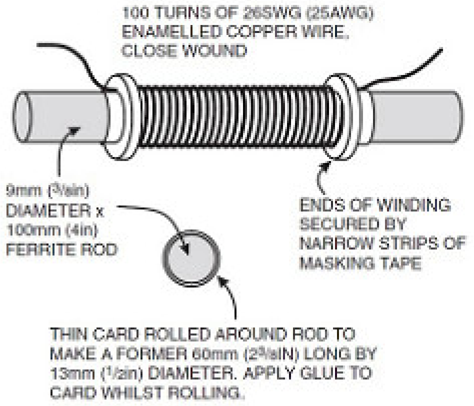
100 TURNS OF 26SWG (25AWG) ENAMELLED COPPER WIRE, CLOSE WOUND
9mm (3/8in) DIAMETER x 100mm (4in) FERRITE ROD
ENDS OF WINDING SECURED BY NARROW STRIPS OF MASKING TAPE
THIN CARD ROLLED AROUND ROD TO MAKE A FORMER 60mm (2³/8IN) LONG BY 13mm (1/2in) DIAMETER. APPLY GLUE TO CARD WHILST ROLLING.

The r.f. bypass capacitor C1 can in practice, be omitted with no noticeable reduction in performance. However, if the set is to be connected to either the headphone amplifier or speaker amplifier decribed next month, the component together with diode, load resistor, RX, must be included.

AMPLIFICATION Audio frequency amplification, after the diode detector will permit the use of low impedance Walkman type earphones or even loud speaker operation. It will do nothing, however, to overcome the diode's insensitivity to weak signals. For this we must have radio frequency amplification of the signals picked up by the aerial before they reach the detector(The standard circuit for a transistor portable receiver has three stages of radio frequency amplification ahead of the diode).

For amplification, you can use a simple audio amplifier using LM386.

READ MY EXPERIENCES ON MAKING A BATTERYLESS RADIO

WATCH THESE VIDEOS:

- **JLCPCB: Prototype 10 PCBs for $2 (2-layer,100*100mm)**
- **China's Largest PCB Prototype Manufacturer, 290,000+ Customers & 8,000+ Online Orders Per Day**
- **PCB Pricing: https://jlcpcb.com/quote**

US 20180069486A1

(19) **United States**

(12) **Patent Application Publication**

OUDA et al.

(10) **Pub. No.: US 2018/0069486 A1**

(43) **Pub. Date: Mar. 8, 2018**

(54) **RF-TO-DC POWER CONVERTERS FOR WIRELESS POWERING**

(71) Applicant: **KING ABDULLAH UNIVERSITY OF SCIENCE AND TECHNOLOGY**, Thuwal (SA)

(72) Inventors: **Mahmoud Hamdy OUDA**, Thuwal (SA); **Khaled Nabil SALAMA**, Thuwal (SA)

(21) Appl. No.: **15/556,118**

(22) PCT Filed: **May 26, 2016**

(86) PCT No.: **PCT/IB2016/053094**

§ 371 (c)(1),
(2) Date: **Sep. 6, 2017**

Related U.S. Application Data

(60) Provisional application No. 62/166,275, filed on May 26, 2015.

Publication Classification

(51) **Int. Cl.**

H02M 7/219	(2006.01)
H02J 50/80	(2006.01)

(52) **U.S. Cl.**

CPC ***H02M 7/219*** (2013.01); ***H02J 50/80*** (2016.02)

(57) **ABSTRACT**

Various examples are provided related to radio frequency (RF) to direct current (DC) power conversion. In one example, a RF-to-DC converter includes a fully cross-coupled rectification circuit including a pair of forward rectifying transistors and a feedback circuit configured to provide feedback bias signals to gates of the pair of forward rectifying transistors via feedback branch elements. In another example, a method includes receiving a radio frequency (RF) signal; rectifying the RF signal via a fully cross-coupled rectification circuit including a pair of forward rectifying transistors; and providing a DC output voltage from an out put connection of the fully cross-coupled rectification circuit, where gating of the pair of forward rectifying transistors is controlled by feedback bias signals provided to gates of the pair of forward rectifying transistors via feedback branch elements.

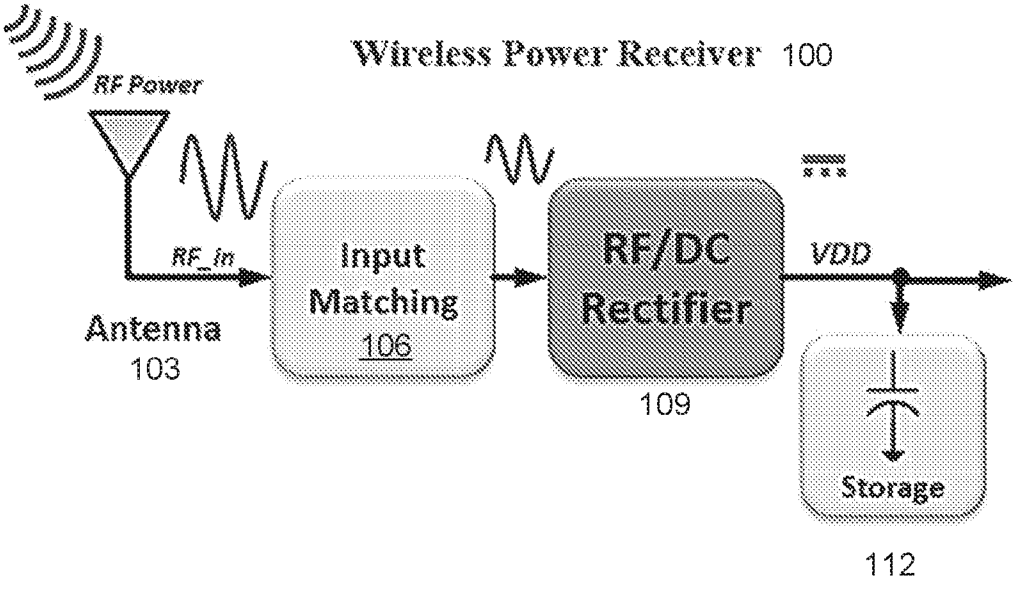

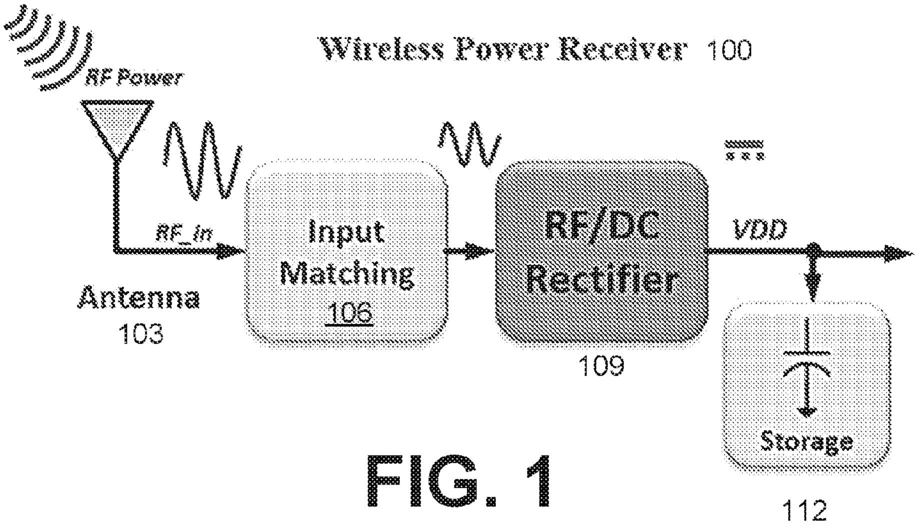
Wireless Power Receiver 100
RF Power
RF_In
Antenna
103
Input
Matching
106
RF/DC
Rectifier
109
VDD
Storage
112

FIG. 1

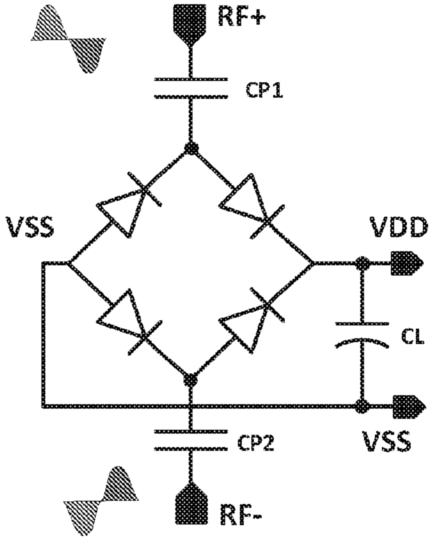
RF+
CP1
VSS
VDD
CL
CP2
VSS
RF-

FIG. 2A

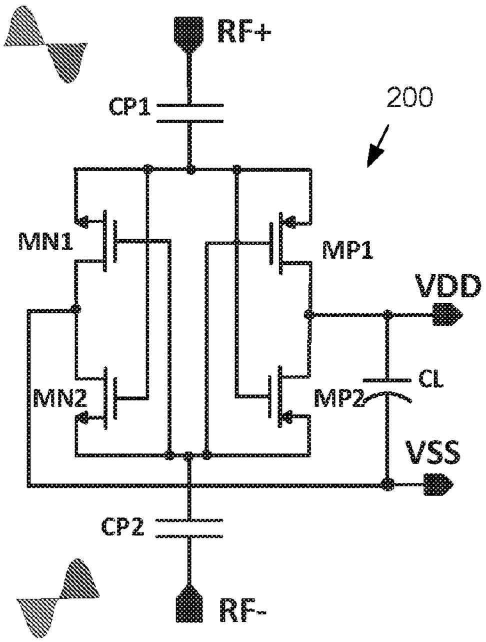
RF+
CP1
200
MN1
MP1
VDD
MN2
MP2
CL
VSS
CP2
RF-

FIG. 2B

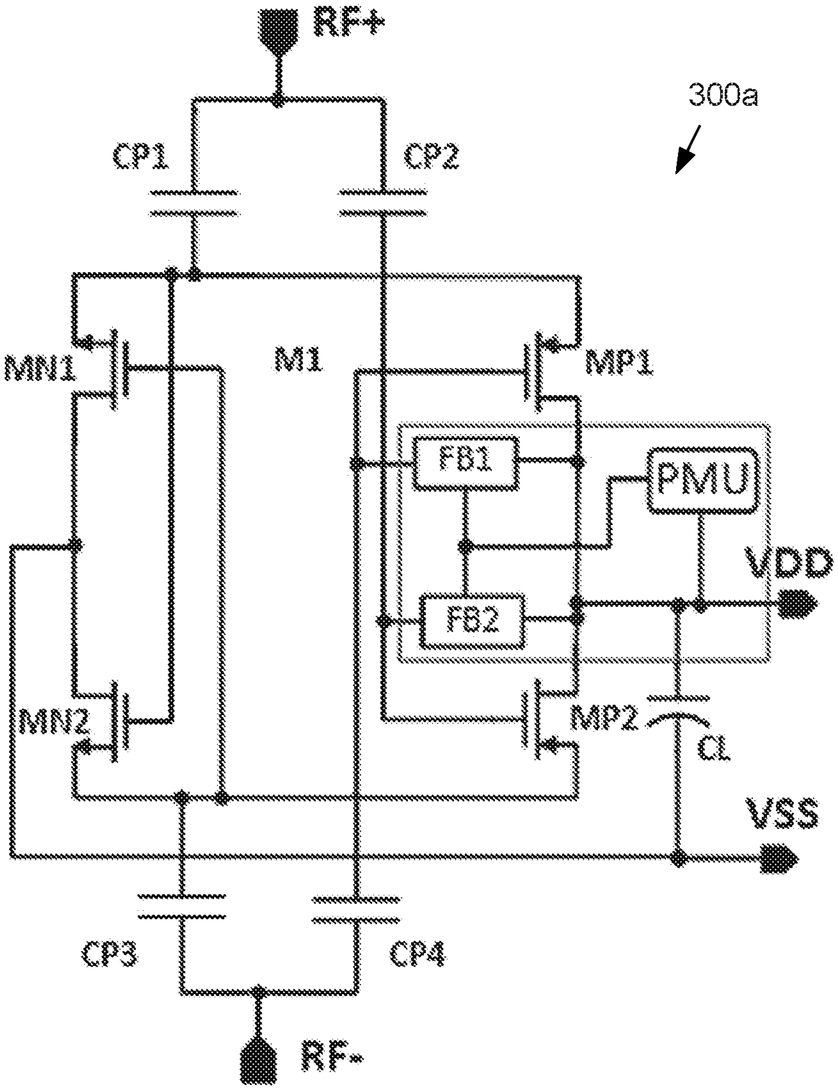
RF+
300a
CP1
CP2
MN1
M1
MP1
FB1
PMU
VDD
FB2
MN2
MP2
CL
VSS
CP3
CP4
RF-

FIG. 3A

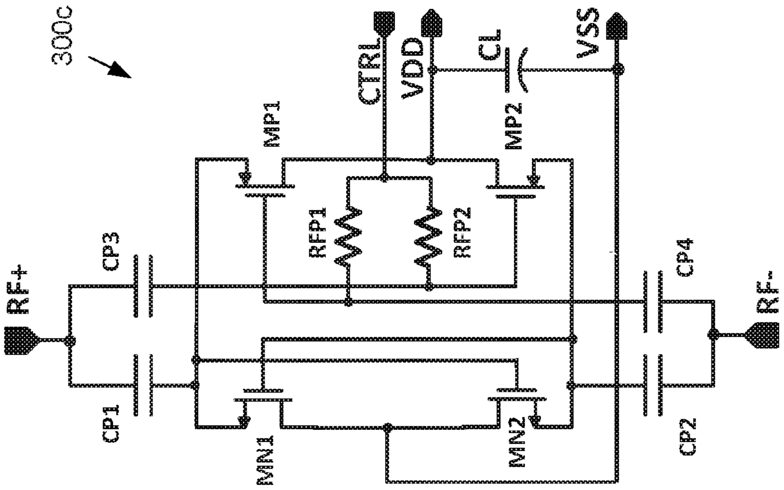
300c
RF+
CP3
CP1
MN1
MP1
RFP1
RFP2
CTRL
VDD
CL
MP2
VSS
CP4
CP2
MN2
RF-

FIG. 3C

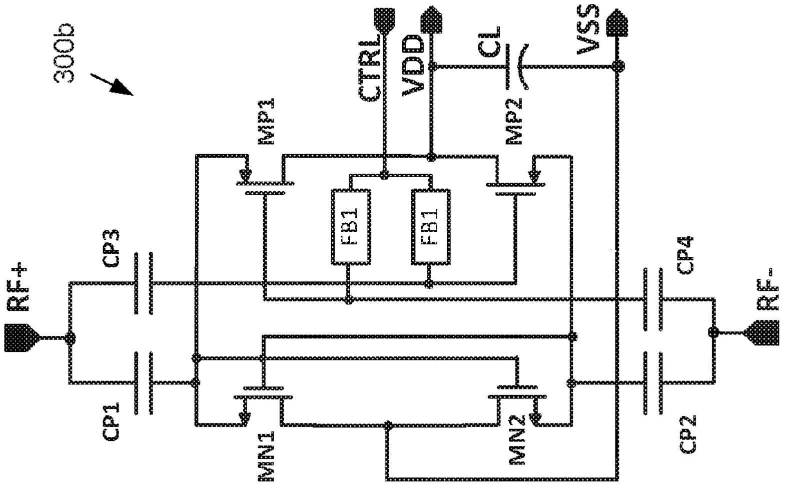
300b
RF+
CP3
CP1
MN1
MP1
FB1
FB1
CTRL
VDD
CL
MP2
VSS
CP4
CP2
MN2
RF-

FIG. 3B

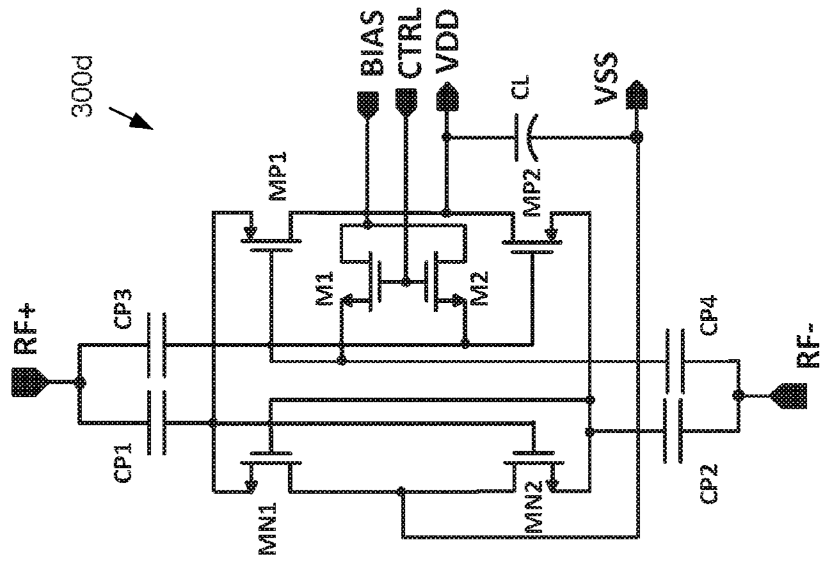
300d
RF+
CP3
CP1
MN1
MP1
M1
M2
BIAS
CTRL
VDD
CL
VSS
MP2
MN2
CP4
CP2
RF-

FIG. 3D

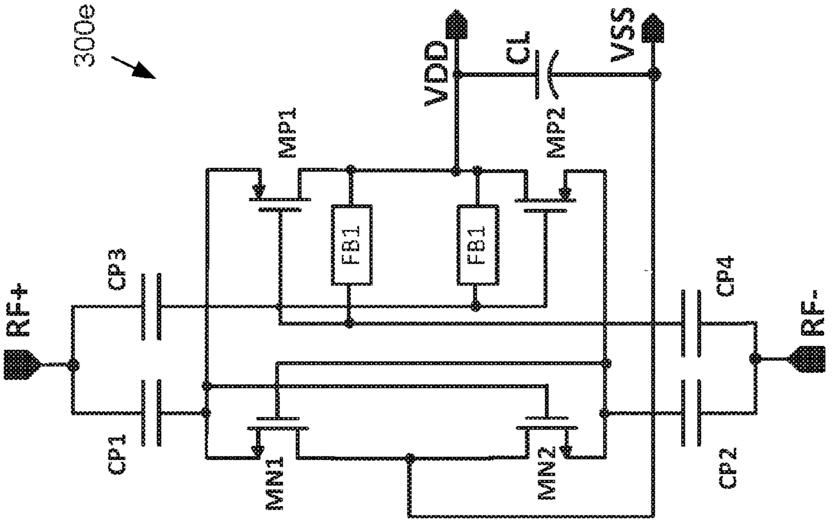
300e
RF+
CP3
CP1
MN1
MP1
FB1
FB1
VDD
CL
VSS
MP2
MN2
CP4
CP2
RF-

FIG. 3E

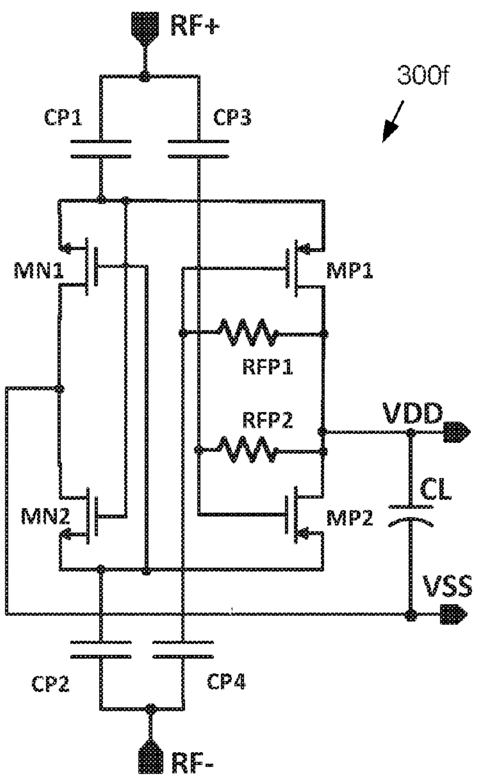
RF+
300f
CP1
CP3
MN1
MP1
RFP1
RFP2
VDD
CL
MN2
MP2
VSS
CP2
CP4
RF-

FIG. 3F

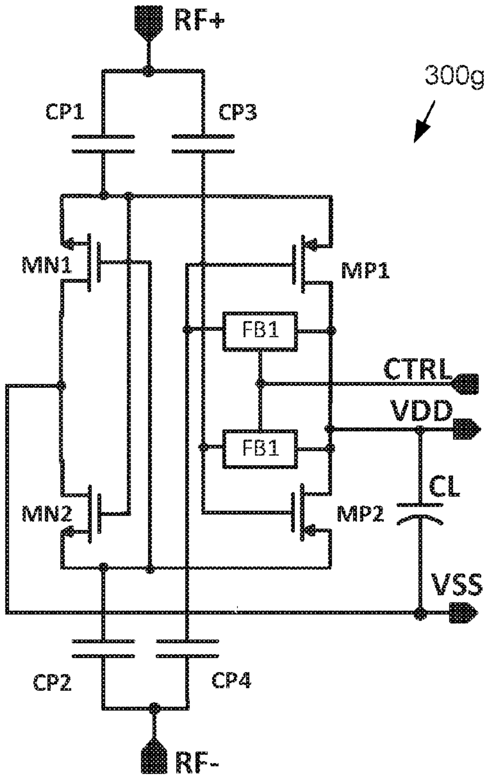
RF+
300g
CP1
CP3
MN1
MP1
FB1
CTRL
VDD
FB1
CL
MN2
MP2
VSS
CP2
CP4
RF-

FIG. 3G

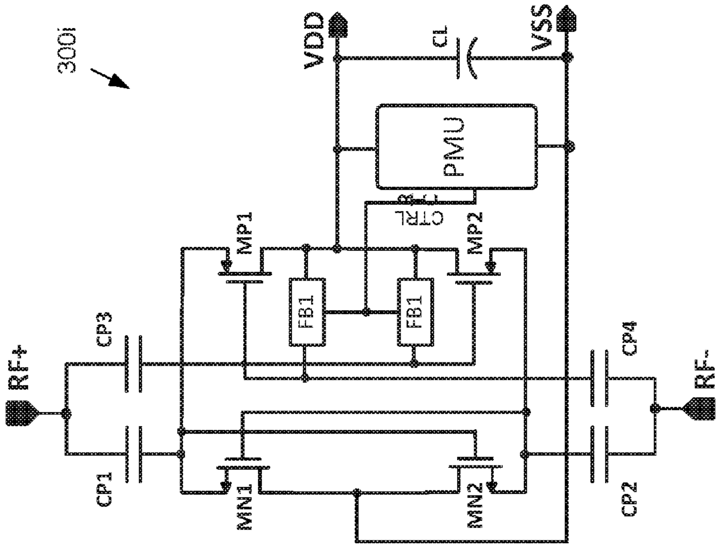
300i
RF+
CP3
CP1
MP1
FB1
FB1
MP2
MN1
MN2
CTRLB
PMU
VDD
CL
VSS
CP4
CP2
RF-

FIG. 3I

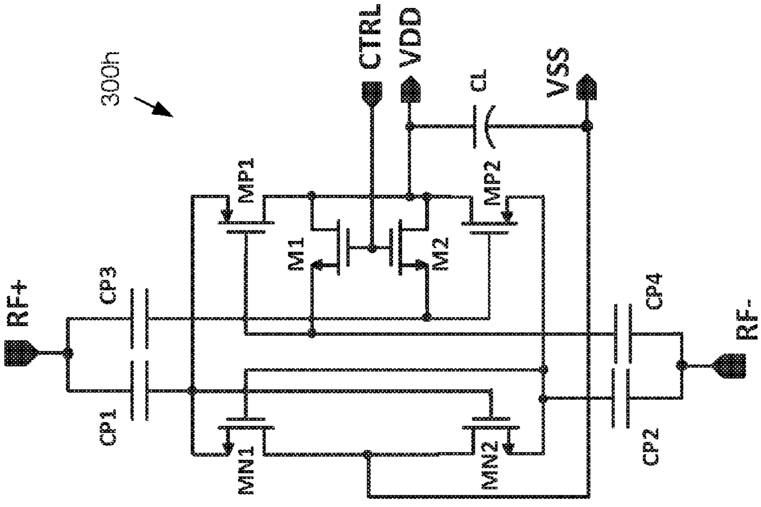
300h
RF+
CP3
CP1
MP1
M1
M2
MP2
MN1
MN2
CTRL
VDD
CL
VSS
CP4
CP2
RF-

FIG. 3H

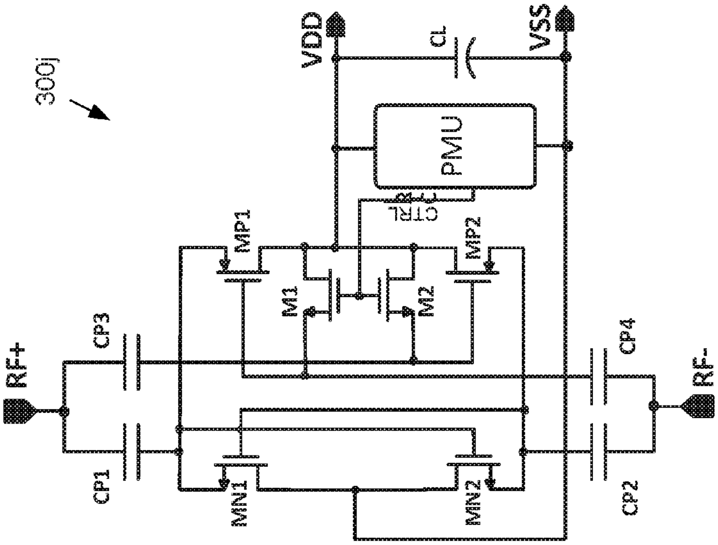
300j
RF+
CP1
CP3
MN1
MP1
M1
M2
MP2
MN2
CTRL
PMU
VDD
CL
VSS
CP2
CP4
RF-

FIG. 3J

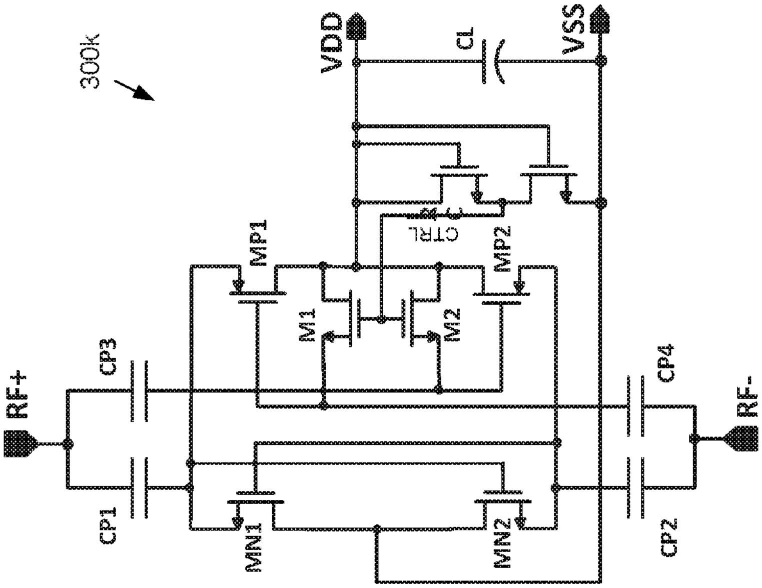
300k
RF+
CP1
CP3
MN1
MP1
M1
M2
MP2
MN2
CTRL
VDD
CL
VSS
CP2
CP4
RF-

FIG. 3K

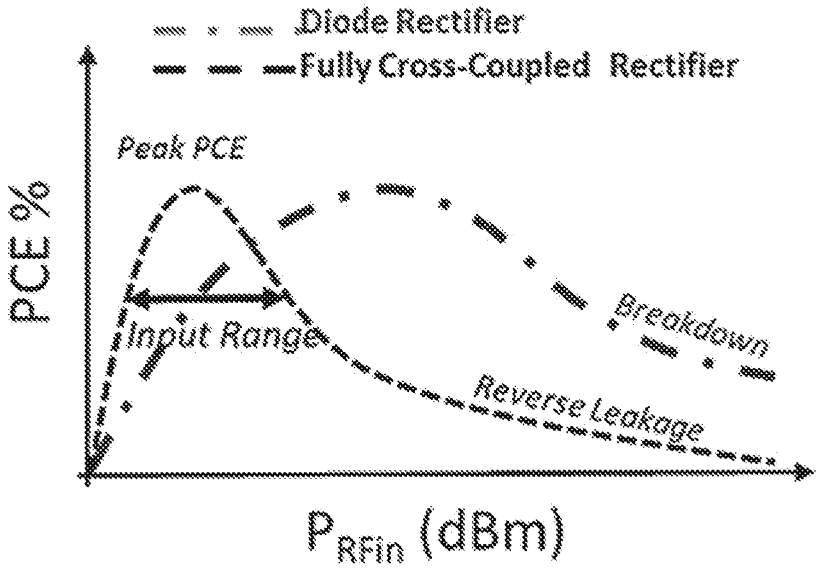
Diode Rectifier
Fully Cross-Coupled Rectifier
Peak PCE
PCE %
Input Range
Breakdown
Reverse Leakage
P_{RFin} (dBm)

FIG. 4A

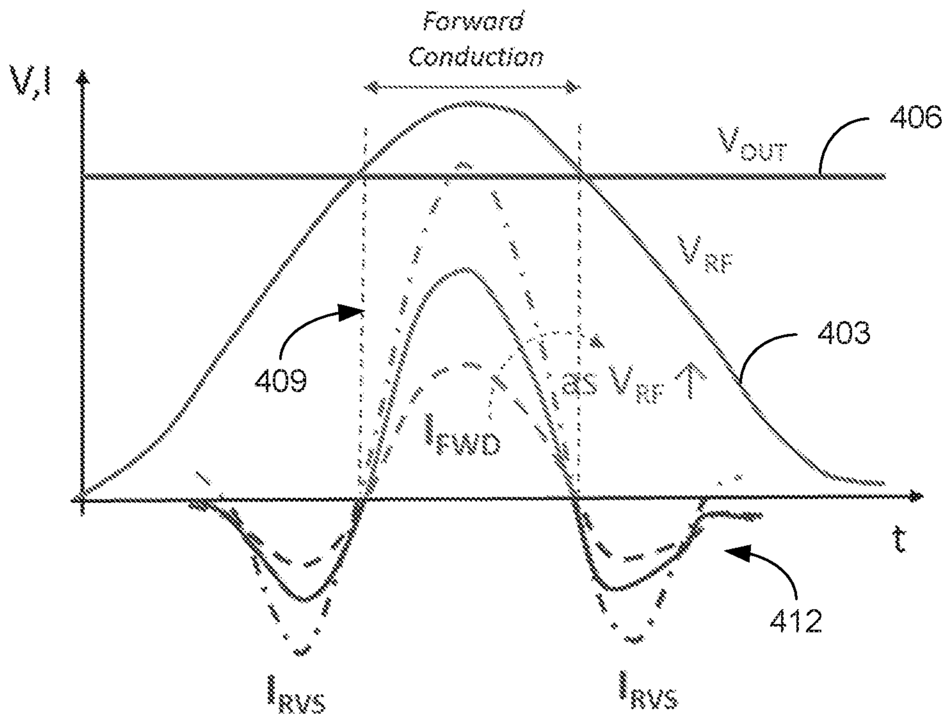
Forward Conduction
V,I
V_{OUT}
406
V_{RF}
403
409
as V_{RF} ↑
I_{FWD}
t
412
I_{RVS}
I_{RVS}

FIG. 4B

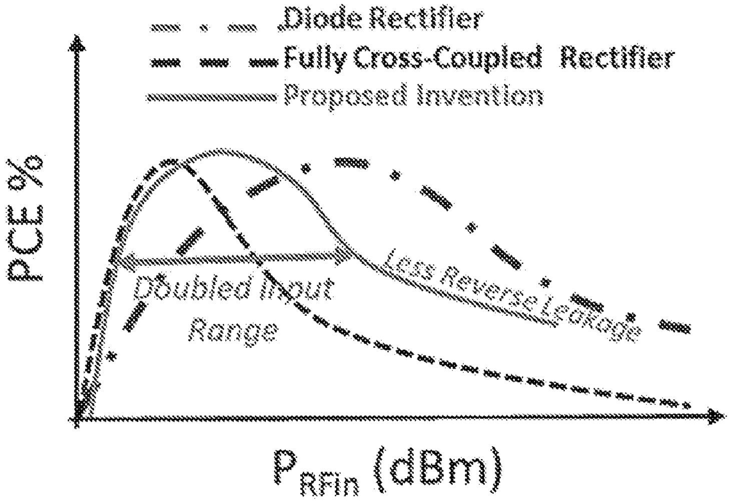
Diode Rectifier
Fully Cross-Coupled Rectifier
Proposed Invention
PCE %
Doubled Input Range
Less Reverse Leakage
P_{RFin} (dBm)

FIG. 4C

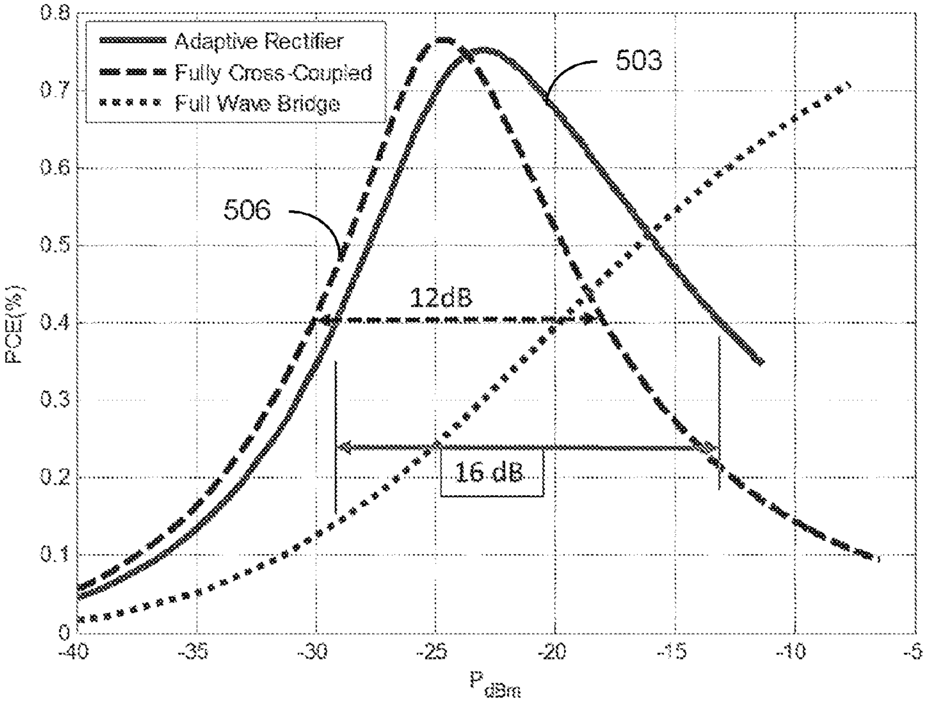
Adaptive Rectifier
Fully Cross-Coupled
Full Wave Bridge
503
506
12dB
16 dB
PCE(%)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
-40
-35
-30
-25
-20
-15
-10
-5
P_{dBm}

FIG. 5A

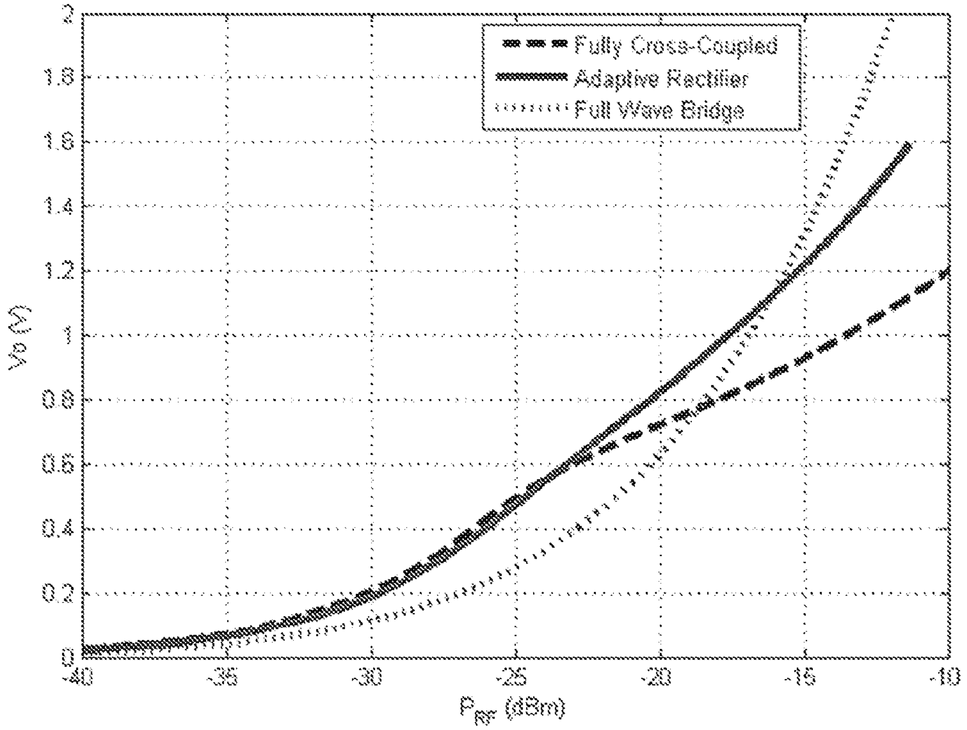
Fully Cross-Coupled
Adaptive Rectifier
Full Wave Bridge
Vo (V)
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-40
-35
-30
-25
-20
-15
-10
P_{RF} (dBm)

FIG. 5B

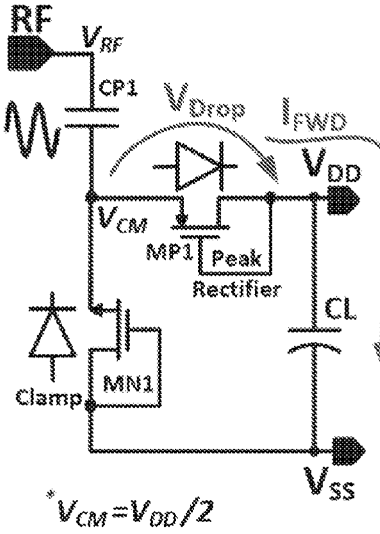
RF
V_{RF}
CP1
V_{Drop}
I_{FWD}
V_{DD}
V_{CM}
MP1
Peak
Rectifier
CL
Clamp
MN1
V_{SS}
$^*V_{CM}=V_{DD}/2$

FIG. 6A

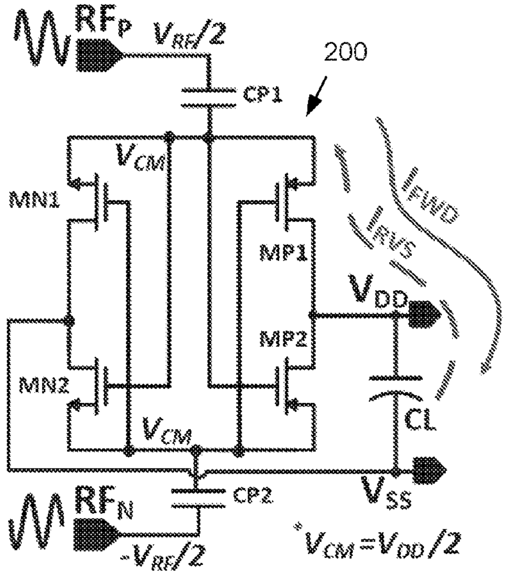
RF_P
$V_{RF}/2$
200
CP1
V_{CM}
MN1
I_{FWD}
I_{RVS}
MP1
V_{DD}
MP2
MN2
CL
V_{CM}
CP2
V_{SS}
RF_N
$-V_{RF}/2$
$^*V_{CM}=V_{DD}/2$

FIG. 6B

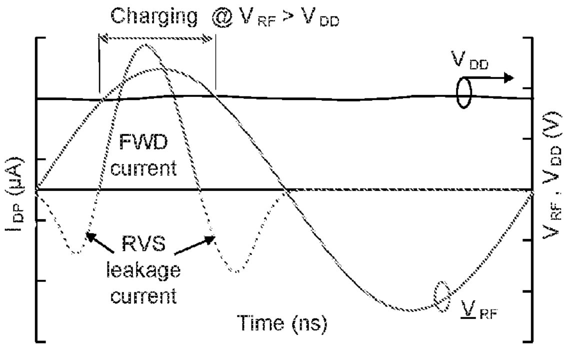

FIG. 7A

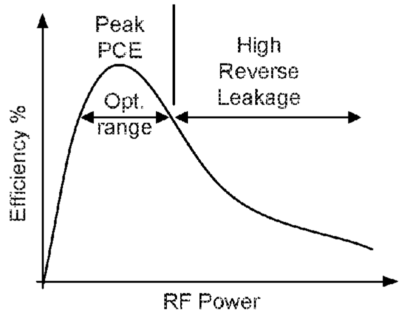

FIG. 7B

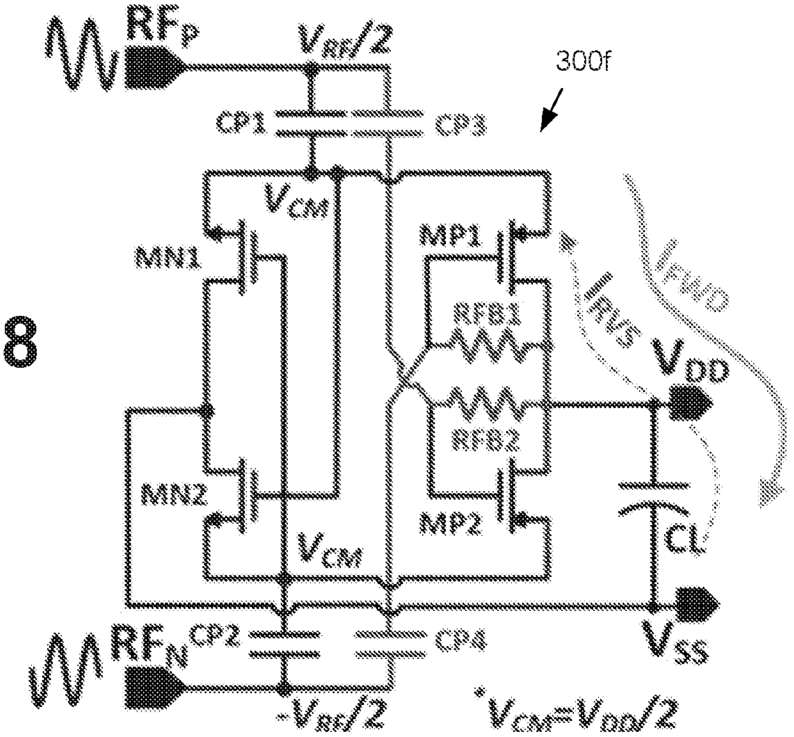
RF_P
$V_{RF}/2$
300f
CP1
CP3
V_{CM}
MN1
MP1
I_{FWD}
I_{RVS}
RFB1
V_{DD}
RFB2
MN2
MP2
CL
V_{CM}
RF_N
CP2
CP4
V_{SS}
$-V_{RF}/2$
$^*V_{CM}=V_{DD}/2$

FIG. 8

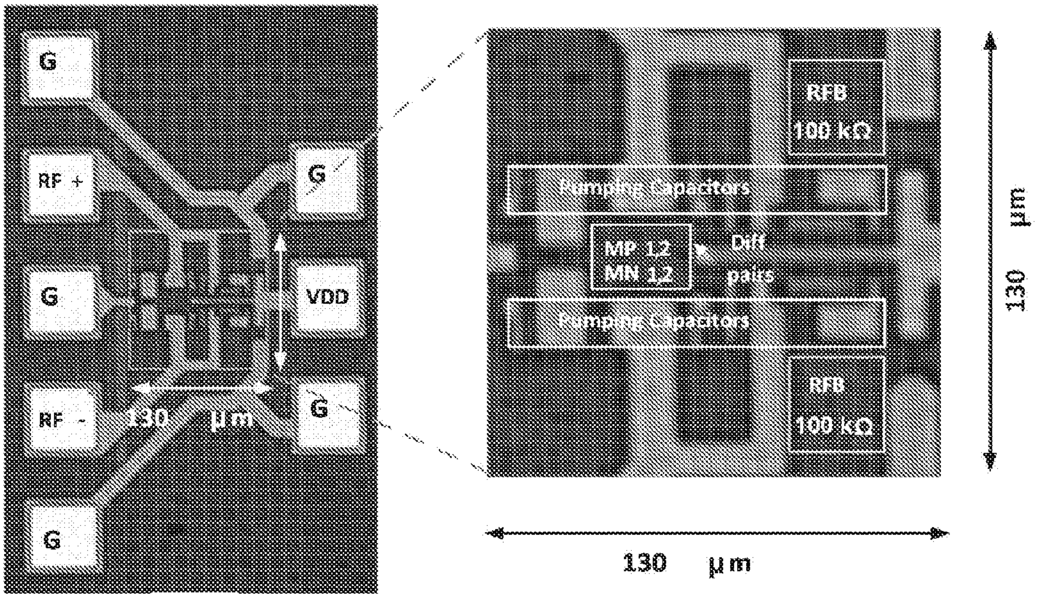
G
RF +
G
G
VDD
RF -
130 µm
G
G
RFB
100 kΩ
Pumping Capacitors
MP 1,2
MN 1,2
Diff
pairs
Pumping Capacitors
RFB
100 kΩ
130 µm
130 µm

FIG. 9

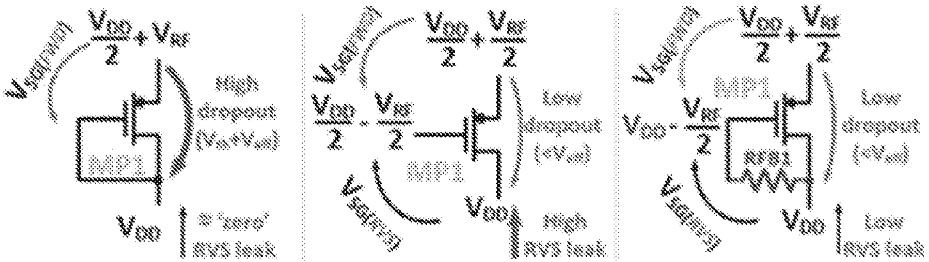
$\frac{V_{DD}}{2}+V_{RF}$
High dropout ($V_{th}+V_{sat}$)
MP1
V_{DD}
≈ 'zero' RVS leak
$\frac{V_{DD}}{2}+\frac{V_{RF}}{2}$
$\frac{V_{DD}}{2}-\frac{V_{RF}}{2}$
Low dropout ($<V_{sat}$)
MP1
V_{DD}
High RVS leak
$\frac{V_{DD}}{2}+\frac{V_{RF}}{2}$
$V_{DD}-\frac{V_{RF}}{2}$
MP1
RFB1
Low dropout ($<V_{sat}$)
V_{DD}
Low RVS leak

(a) Diode connected (Dickson multiplier)

(b) Cross-coupled (Conventional)

(c) Self-biased (Proposed)

FIG. 10

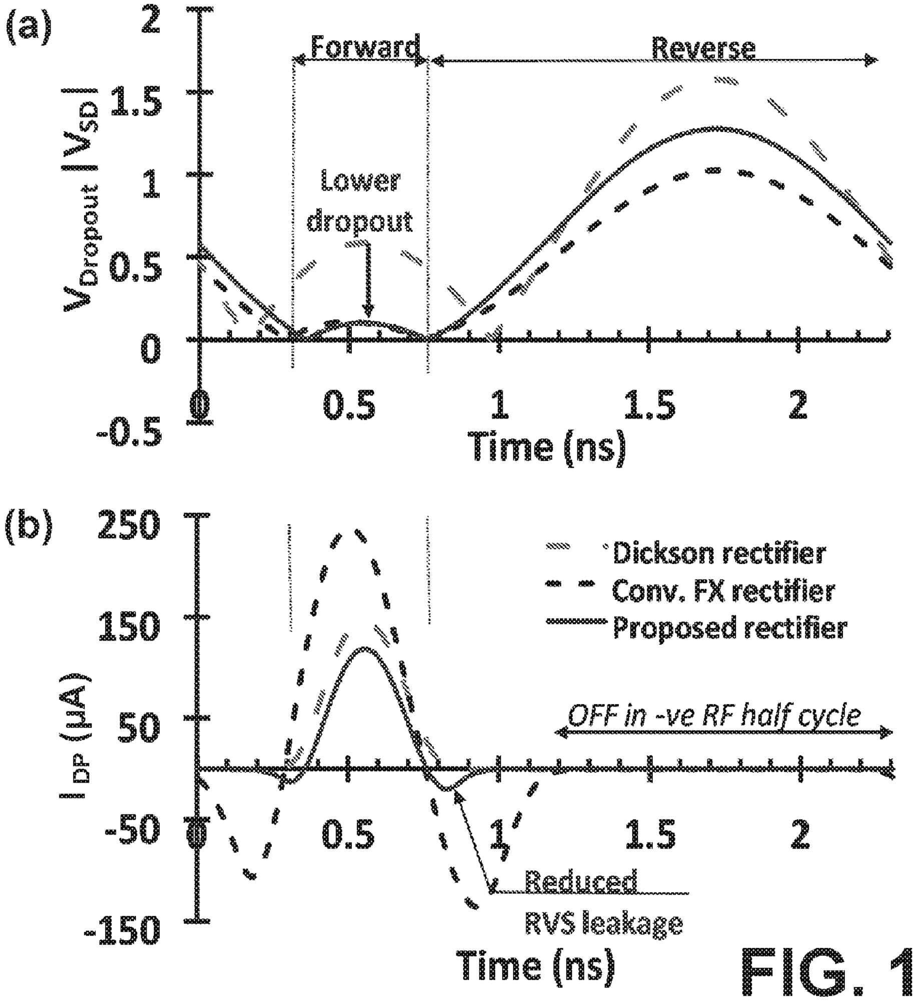

FIG. 11

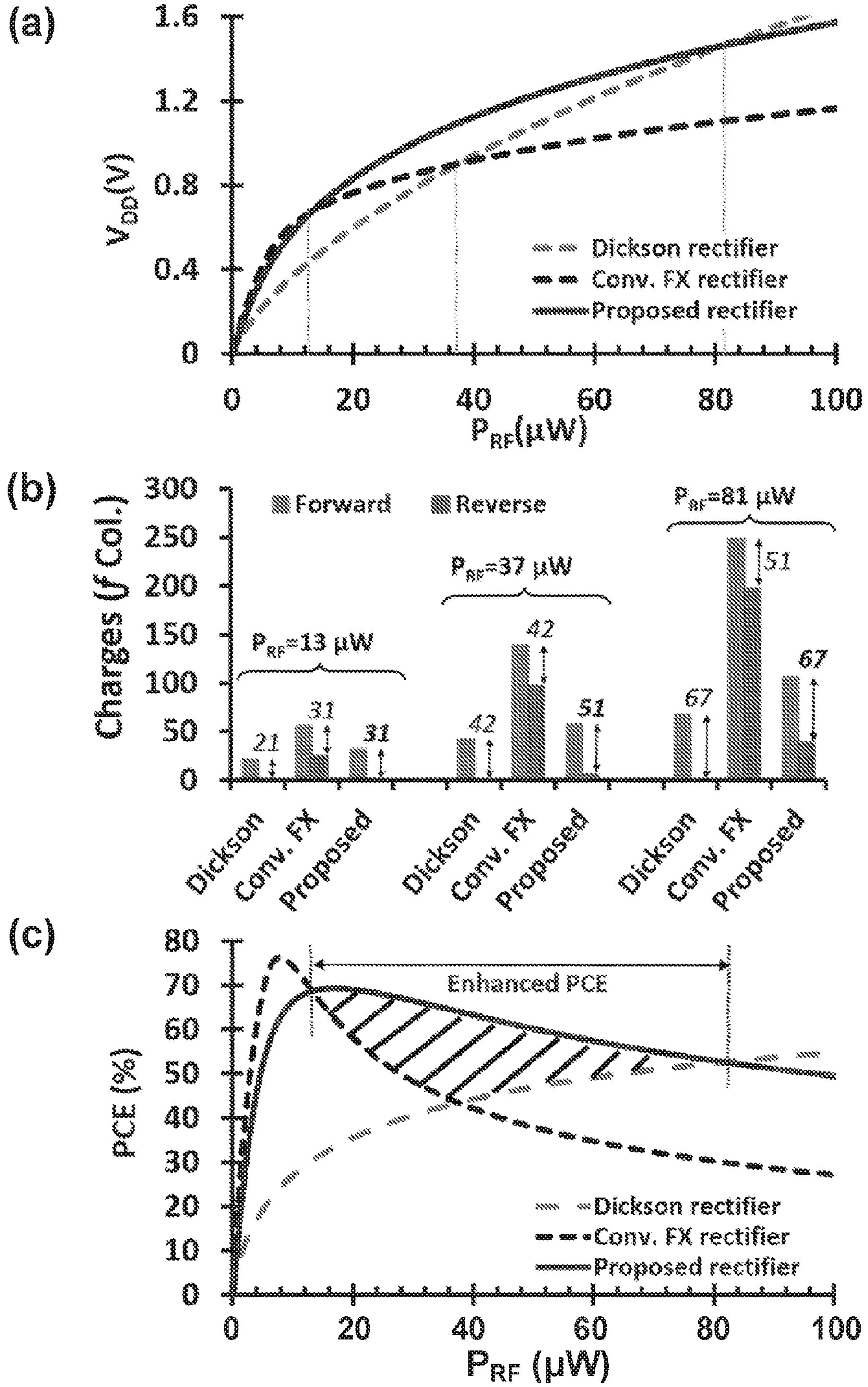
(a)
1.6
1.2
0.8
0.4
0
V_{DD}(V)
Dickson rectifier
Conv. FX rectifier
Proposed rectifier
0
20
40
60
80
100
P_{RF}(μW)
(b)
300
250
200
150
100
50
0
Charges (f Col.)
Forward
Reverse
P_{RF}=13 μW
P_{RF}=37 μW
P_{RF}=81 μW
21
31
31
42
42
51
67
51
67
Dickson
Conv. FX
Proposed
Dickson
Conv. FX
Proposed
Dickson
Conv. FX
Proposed
(c)
80
70
60
50
40
30
20
10
0
PCE (%)
Enhanced PCE
Dickson rectifier
Conv. FX rectifier
Proposed rectifier
0
20
40
60
80
100
P_{RF} (μW)

FIG. 12

COMPARISON OF RECTIFYING DEVICES

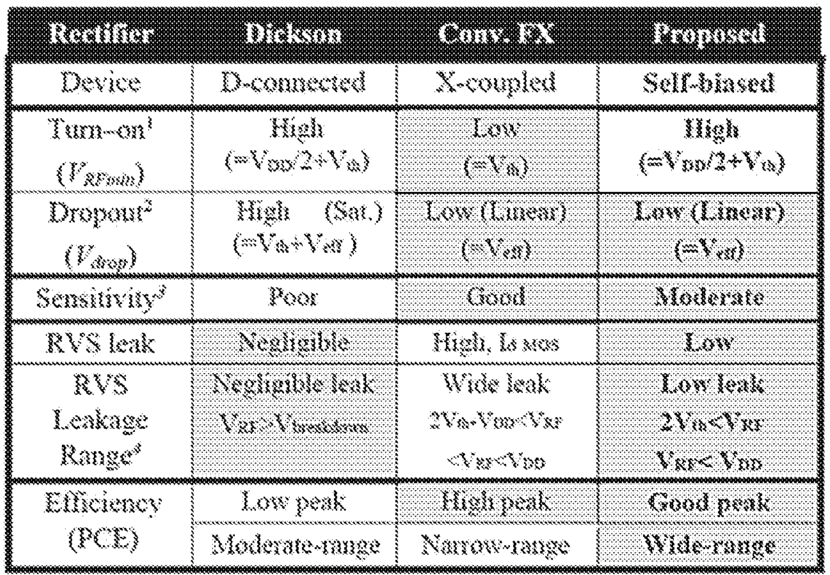

Rectifier	Dickson	Conv. FX	Proposed
Device	D-connected	X-coupled	Self-biased
Turn-on[1] (V_{RFmin})	High (=$V_{DD}/2+V_{th}$)	Low (=V_{th})	High (=$V_{DD}/2+V_{th}$)
Dropout[2] (V_{drop})	High (Sat.) (=$V_{th}+V_{eff}$)	Low (Linear) (=V_{eff})	Low (Linear) (=V_{eff})
Sensitivity[3]	Poor	Good	Moderate
RVS leak	Negligible	High, $I_{s\ MOS}$	Low
RVS Leakage Range[4]	Negligible leak $V_{RF}>V_{breakdown}$	Wide leak $2V_{th}-V_{DD}<V_{RF}$ $<V_{RF}<V_{DD}$	Low leak $2V_{th}<V_{RF}$ $V_{RF}<V_{DD}$
Efficiency (PCE)	Low peak	High peak	Good peak
	Moderate-range	Narrow-range	Wide-range

[1] *Turn-on at* $V_{SG} > V_{th}$.

[2] *Dropout* affects sensitivity, output voltage and efficiency.

[3] *Sensitivity: turn on at low* V_{RF} *& with low dropout voltage.*

[4] *Reverse leakage if* ($V_{SG(RVS)} > V_{th}$) & ($V_{SD(RVS)} > 0$).

FIG. 13

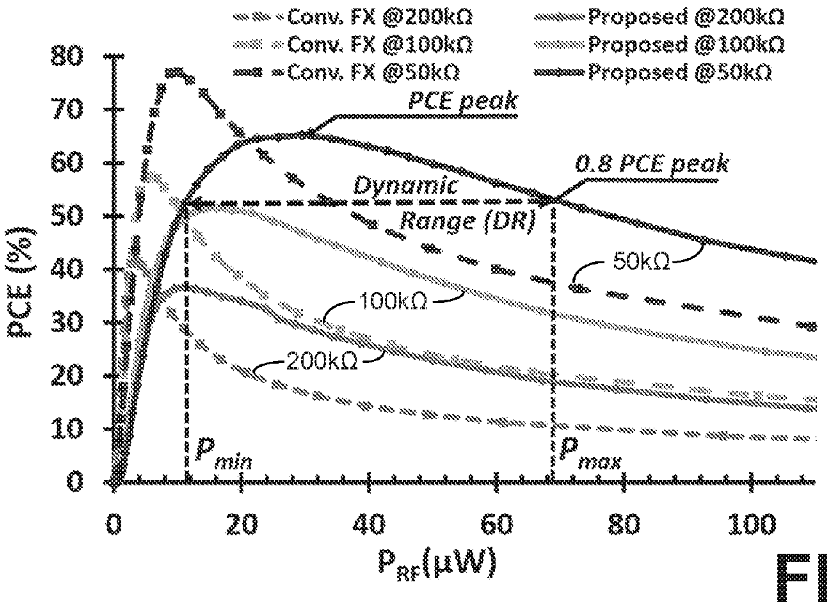

FIG. 14

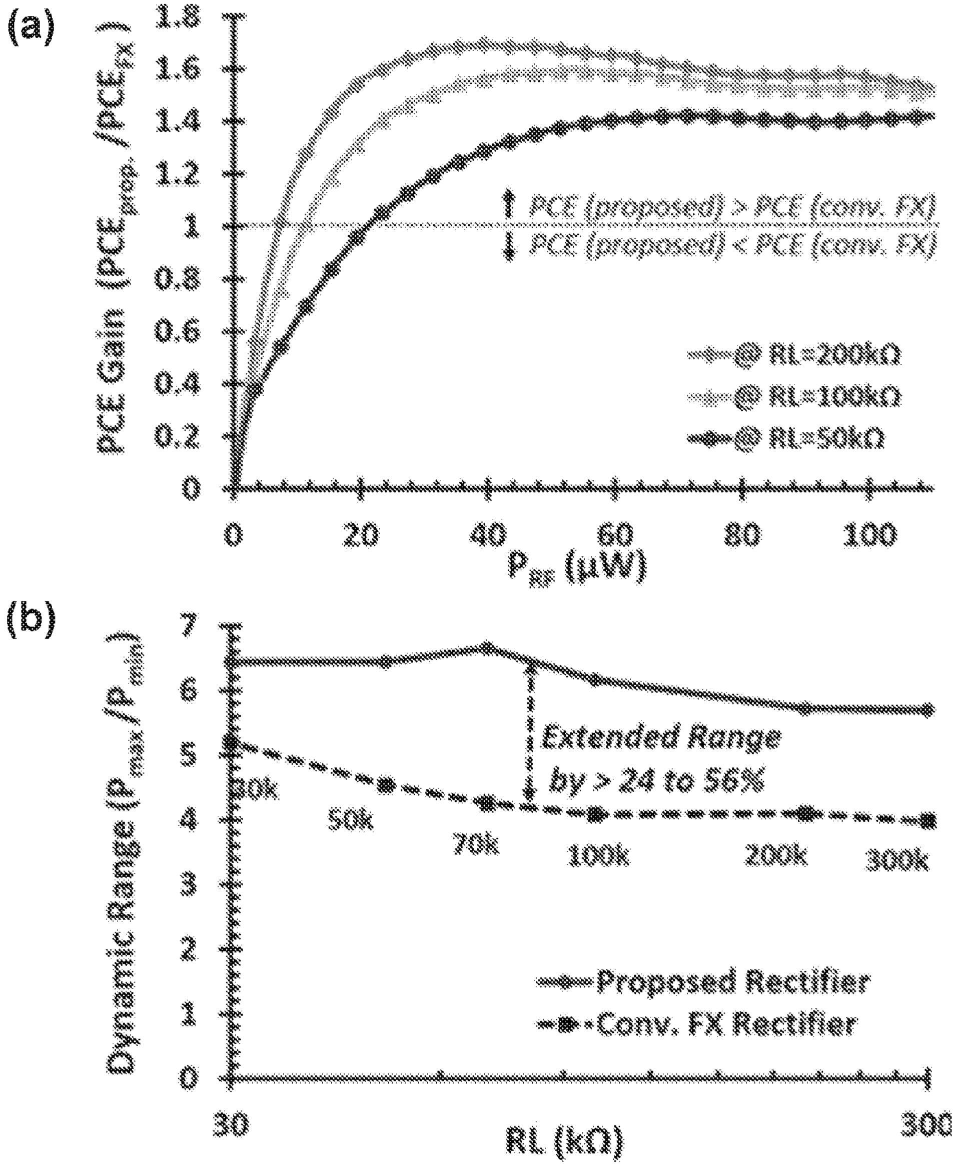
(a)
PCE Gain $(PCE_{prop.}/PCE_{FX})$
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
P_{RF} (μW)
↑ PCE (proposed) > PCE (conv. FX)
↓ PCE (proposed) < PCE (conv. FX)
@ RL=200kΩ
@ RL=100kΩ
@ RL=50kΩ
(b)
Dynamic Range (P_{max}/P_{min})
7
6
5
4
3
2
1
0
30
300
RL (kΩ)
Extended Range by > 24 to 56%
30k
50k
70k
100k
200k
300k
Proposed Rectifier
Conv. FX Rectifier

FIG. 15

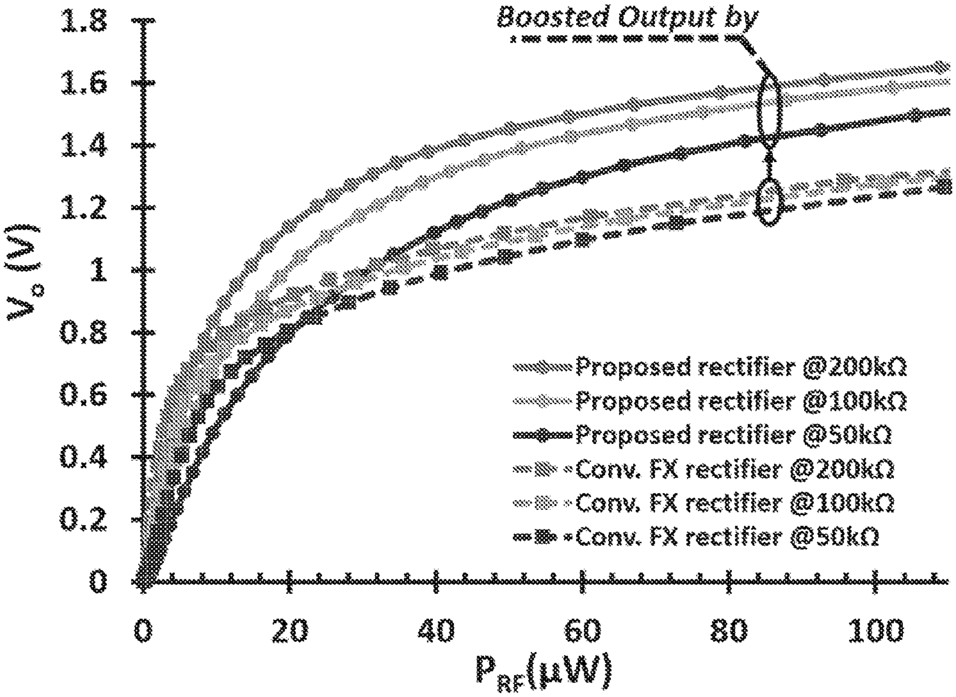
Boosted Output by
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
V_o (V)
0
20
40
60
80
100
P_{RF}(μW)
Proposed rectifier @200kΩ
Proposed rectifier @100kΩ
Proposed rectifier @50kΩ
Conv. FX rectifier @200kΩ
Conv. FX rectifier @100kΩ
Conv. FX rectifier @50kΩ

FIG. 16

PERFORMANCE COMPARISON.

CMOS Process	Architecture	# stages	Chip Area (mm^2)	Frequency	Sensitivity P_{RF}@Vo=1V	Load (R_L)	Peak PCE	Dynamic Range^{a} P_{max}/P_{min}
0.18 μm	Adaptive self-biased crossconnected rectifier (Proposed)	1-stage	0.017	433 MHz	-15.2 dBm	50 $k\Omega$	65.3 %	6.5
					-17 dBm	100 $k\Omega$	51.5 %	6.2
0.18 μm	Fully cross-coupled (FX) rectifier (Conventional)	1-stage	0.005	433 MHz	-13.6 dBm	50 $k\Omega$	77 %	4.5
					-14.5 dBm	100 $k\Omega$	57.5 %	4.1
0.18 μm	Fully cross-coupled with inter-stage RF injection	3-stage	0.088	433 MHz	-7.8 dBm	30 $k\Omega$	13.2 %	3.2
					-4.1 dBm	100 $k\Omega$	10 %	3
0.18 μm	Dickson rectifier with multiple V_{th} offset cancellation.	4-stage	0.15	433 MHz	-11.2 dBm	50 $k\Omega$	32 %	4.2
					-14 dBm	100 $k\Omega$	34 %	4.7
0.18 μm	Hybrid V_{th} cancellation using RF input & DC output voltage	1-stage	NA	433 MHz	> 0 dBm^{b}	30 $k\Omega$	12.5 %	3.3
					> 0 dBm^{b}	100 $k\Omega$	4.6 %	3.6

FIG. 17

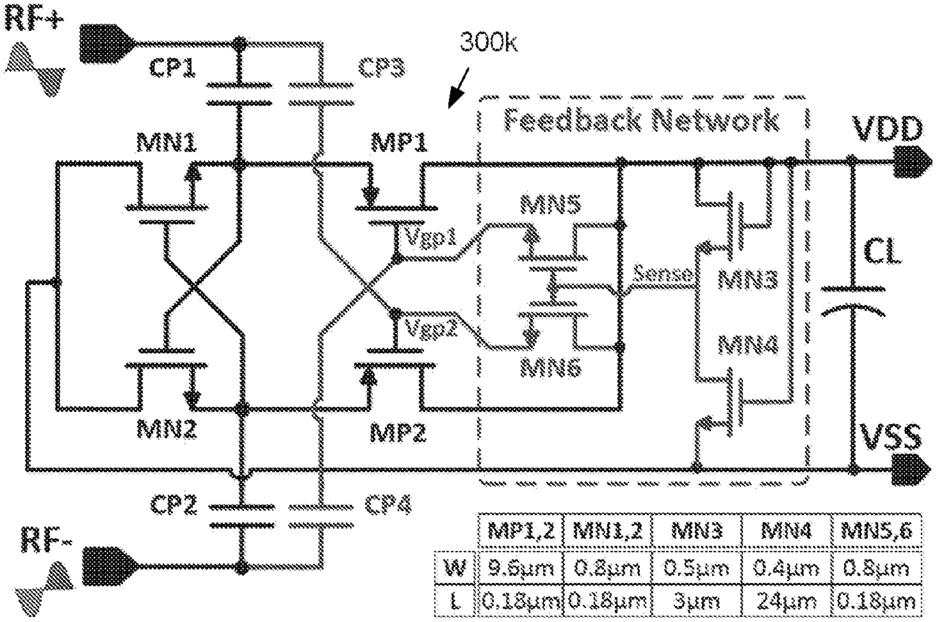
RF+
CP1
CP3
300k
Feedback Network
VDD
MN1
MP1
MN5
Vgp1
Sense
MN3
CL
Vgp2
MN4
MN6
MN2
MP2
VSS
CP2
CP4
RF-

	MP1,2	MN1,2	MN3	MN4	MN5,6
W	9.6μm	0.8μm	0.5μm	0.4μm	0.8μm
L	0.18μm	0.18μm	3μm	24μm	0.18μm

FIG. 18

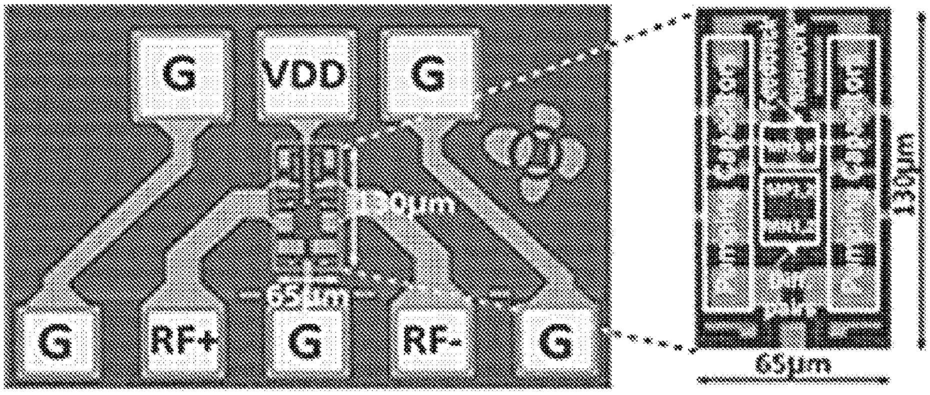
G
VDD
G
130μm
65μm
G
RF+
G
RF-
G
130μm
65μm

FIG. 19

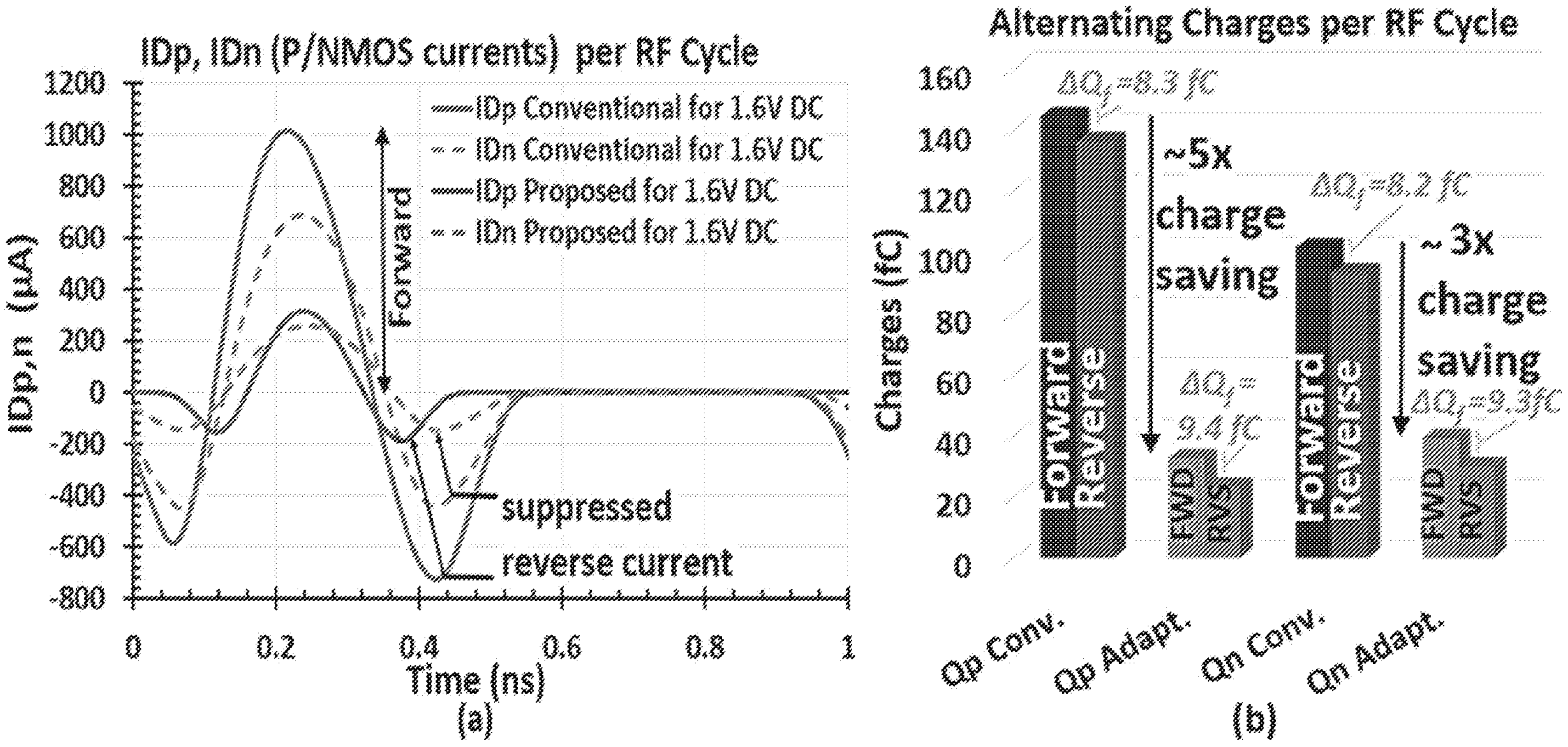
IDp, IDn (P/NMOS currents) per RF Cycle
IDp,n (µA)
1200
1000
800
600
400
200
0
-200
-400
-600
-800
IDp Conventional for 1.6V DC
IDn Conventional for 1.6V DC
IDp Proposed for 1.6V DC
IDn Proposed for 1.6V DC
Forward
suppressed
reverse current
0
0.2
0.4
0.6
0.8
1
Time (ns)
(a)
Alternating Charges per RF Cycle
Charges (fC)
160
140
120
100
80
60
40
20
0
ΔQ_f=8.3 fC
~5x
charge
saving
ΔQ_f=8.2 fC
~ 3x
charge
saving
Forward
Reverse
ΔQ_f=
9.4 fC
FWD
RVS
Forward
Reverse
ΔQ_f=9.3fC
FWD
RVS
Qp Conv.
Qp Adapt.
Qn Conv.
Qn Adapt.
(b)

FIG. 20

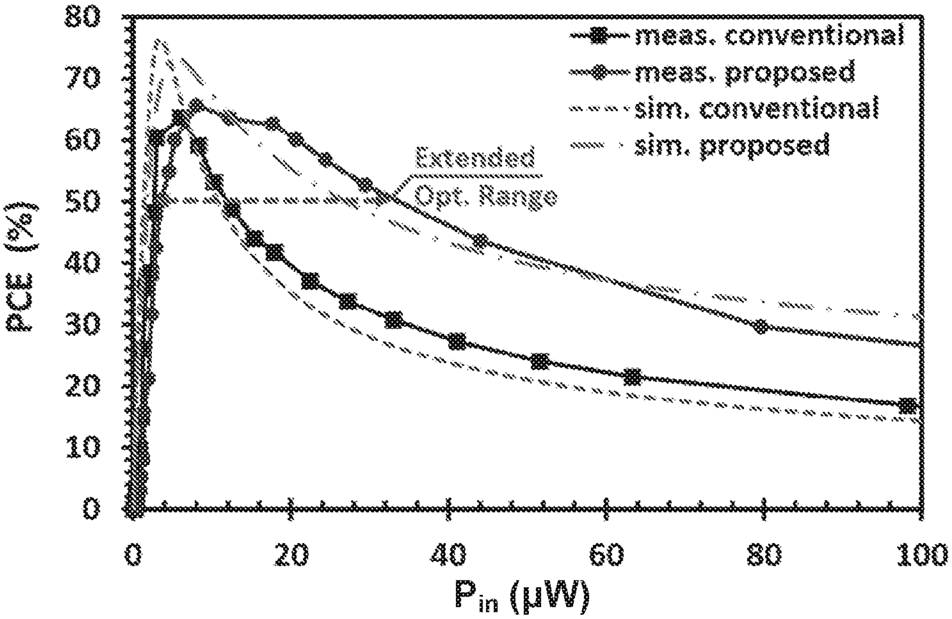
80
70
60
50
40
30
20
10
0
PCE (%)
0
20
40
60
80
100
P_{in} (μW)
meas. conventional
meas. proposed
sim. conventional
sim. proposed
Extended
Opt. Range

FIG. 21

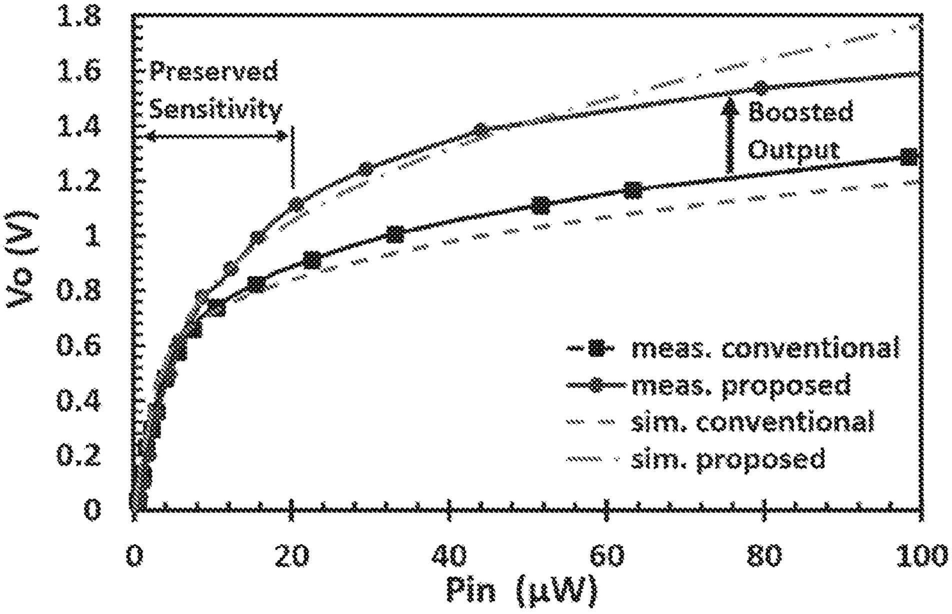
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Vo (V)
0
20
40
60
80
100
Pin (μW)
Preserved
Sensitivity
Boosted
Output
meas. conventional
meas. proposed
sim. conventional
sim. proposed

FIG. 22

PERFORMANCE COMPARISON.

Process	Configuration	Sensitivity	Peak PCE	Opt. Range[a] P_{max}/P_{min}
0.18μm	1-stage, Proposed Adaptive[b]	-18 dBm[c] @Vo=1V	65%[c]	4.6x[c]
0.18μm	1-stage, Standard cross-coupled	-14.8 dBm[c] @Vo=1V	63%[c]	2.7x[c]
0.18μm	1-stage, Standard cross-coupled	-12.8 dBm[c] @Vo=1V	82.6%[c]	2.7x[c]
0.25μm	36-stage, floating gate doubler	-14.6dBm[d] @Vo=1V	60.7%[d]	2x[d]
90nm	5-stage, Standard cross-coupled	-20 dBm[d] @Vo=1V	40%[d]	2.4x[d]
0.13μm	12-stage,Adaptive V_{th} compensated	-17dBm[f] @Vo=1V	32%[f]	2.9x[f]
0.13μm	4-stage, Reconfigurable	-21dBm[e] @Vo=2V	60%[e]	4.4x[e]

[a]Input power range of PCE>0.9 Peak PCE, [b]Intra-stage, [c]@100$k\Omega$, [d]@330$k\Omega$, [f]@500$k\Omega$, [e]@ tracking load (regulator with external ref.

FIG. 23

RF-TO-DC POWER CONVERTERS FOR WIRELESS POWERING

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to, and the benefit of, co-pending U.S. provisional application entitled "RF-TO-DC POWER CONVERTERS FOR WIRELESS POWER RECEIVERS" having Ser. No. 62/166,275, filed May 26, 2015, which is hereby incorporated by reference in its entirety.

BACKGROUND

[0002] Wireless power transfer is a vital part of radio frequency (RF) energy harvesting or wireless powering of systems. Applications range from radio frequency identification (RFID) tags and battery-less sensors for biomedical applications to wirelessly charging smart phones and consumer electronics, and even electric vehicles, or other environmental applications. One of the largest range application is wireless powering channel in the space based solar power systems.

SUMMARY

[0003] Embodiments of the present disclosure are related to radio frequency (RF) to direct current (DC) power conversion and applications thereof.

[0004] In one embodiment, among others, a radio frequency to direct current (RF-to-DC) converter comprises a fully cross-coupled rectification circuit including a pair of forward rectifying transistors; and a feedback circuit configured to provide feedback bias signals to gates of the pair of forward rectifying transistors via feedback branch elements. In one or more aspects of these embodiments, the feedback branch elements can comprise controlled elements. The controlled elements can be MOSFET transistors. The feedback branch elements can be resistors. The feedback bias signals can be supplied from a common control input via the feedback branch elements. The feedback bias signals can be supplied from a common bias input via the feedback branch elements, the feedback bias signals controlled via a common control input. In one or more aspects of these embodiments, the RF-to-DC converter can comprise a power management unit (PMU) coupled to the common control input. The PMU can provide a control signal to the common control input based at least in part upon an output voltage of the RF-to-DC converter. The feedback bias signals can be supplied from a common output connection of the pair of forward rectifying transistors via the feedback branch elements. The common output connection can couple drains of the pair of forward rectifying transistors. The RF-to-DC converter can comprise input pumping capacitors coupled between RF input connections and the fully cross-coupled rectification circuit.

[0005] In another embodiment, a method comprises receiving a radio frequency (RF) signal; rectifying the RF signal via a fully cross-coupled rectification circuit including a pair of forward rectifying transistors; and providing a DC output voltage from an output connection of the fully cross-coupled rectification circuit, where gating of the pair of forward rectifying transistors is controlled by feedback bias signals provided to gates of the pair of forward rectifying transistors via feedback branch elements. In one or more aspects of these embodiments, the method can comprise sensing the DC output voltage and providing a control signal to the feedback branch elements based upon the DC output voltage. The control signal can be provided by a power management unit (PMU) configured to sample the DC output voltage. The control signal can be provided by a transistor sensing network coupled to the output connection of the fully cross-coupled rectification circuit. The feedback branch elements can be resistors coupled between the gates of the pair of forward rectifying transistors and the output connection. The RE signal can be a UHF signal at, e.g., 433 MHz or 950 MHz.

[0006] Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims. In addition, all optional and preferred features and modifications of the described embodiments are usable in all aspects of the disclosure taught herein. Furthermore, the individual features of the dependent claims, as well as all optional and preferred features and modifications of the described embodiments are combinable and interchangeable with one another.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] Many aspects of the present disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

[0008] FIG. **1** is a block diagram of an example of a wireless power receiver in accordance with various embodiments of the present disclosure.

[0009] FIGS. **2**A and **2**B are schematic diagrams of a full wave bridge rectifier and a fully cross-coupled (FX) rectifier, respectively, in accordance with various embodiments of the present disclosure.

[0010] FIGS. **3**A through **3**K are schematic diagrams illustrating various embodiments of adaptive and/or reconfigurable wide range rectifiers in accordance with various embodiments of the present disclosure.

[0011] FIGS. **4**A through **4**C are plots illustrating examples of operational characteristics of full wave bridge, fully cross-coupled, and wide range rectifiers in accordance with various embodiments of the present disclosure.

[0012] FIGS. **5**A and **5**B are plots illustrating examples of simulation results of an adaptive and/or reconfigurable wide range rectifier in accordance with various embodiments of the present disclosure.

[0013] FIGS. **6**A and **6**B are schematic diagrams illustrating examples of a Dickson rectifier and a conventional fully cross-coupled (FX) rectifier, respectively, in accordance with various embodiments of the present disclosure.

[0014] FIG. **7**A and **7**B are plots illustrating examples of simulation results of a conventional FX rectifier in accordance with various embodiments of the present disclosure.

[0015] FIG. **8** is a schematic diagram illustrating an example of a self-biased rectifier of FIG. **3**F in accordance with various embodiments of the present disclosure.

[0016] FIG. **9** are die microphotographs of a fabricated self-biased rectifier of FIG. **8** in accordance with various embodiments of the present disclosure.

[0017] FIG. **10** includes schematic diagrams illustrating examples of operating points of the rectifying device (MP**1**) of the Dickson rectifier of FIG. **6**A, conventional FX rectifier of FIG. **6**B, and self-biased rectifier of FIG. **8** in accordance with various embodiments of the present disclosure.

[0018] FIGS. **11** and **12** includes plots illustrating examples of simulation results of the Dickson rectifier of FIG. **6**A, conventional FX rectifier of FIG. **6**B, and self-biased rectifier of FIG. **8** in accordance with various embodiments of the present disclosure.

[0019] FIG. **13** is a table comparing simulation results of the Dickson rectifier of FIG. **6**A, conventional FX rectifier of FIG. **6**B, and self-biased rectifier of FIG. **8** in accordance with various embodiments of the present disclosure.

[0020] FIGS. **14**, **15** and **16** include plots illustrating examples of evaluation results of the conventional FX rectifier of FIG. **6**B and self-biased rectifier of FIG. **8** in accordance with various embodiments of the present disclosure.

[0021] FIG. **17** is a table comparing performance characteristics of the self-biased rectifier of FIG. **8** and other conventional rectifiers in accordance with various embodiments of the present disclosure.

[0022] FIG. **18** is a schematic diagram illustrating an example of an adaptive rectifier of FIG. **3**K in accordance with various embodiments of the present disclosure.

[0023] FIG. **19** are die microphotographs of a fabricated adaptive rectifier of FIG. **18** in accordance with various embodiments of the present disclosure.

[0024] FIG. **20** includes plots illustrating examples of simulation results of the conventional FX rectifier of FIG. **6**B and and adaptive rectifier of FIG. **18** in accordance with various embodiments of the present disclosure.

[0025] FIGS. **21** and **22** include plots illustrating examples of evaluation results of the conventional FX rectifier of FIG. **6**B and adaptive rectifier of FIG. **18** in accordance with various embodiments of the present disclosure

[0026] FIG. **23** is a table comparing performance characteristics of the adaptive rectifier of FIG. **18** and other conventional rectifiers in accordance with various embodiments of the present disclosure.

DETAILED DESCRIPTION

[0027] Disclosed herein are various embodiments of methods, devices, apparatus and/or systems related to radio frequency (RF) to direct current (DC) power conversion. Reference will now be made in detail to the description of the embodiments as illustrated in the drawings, wherein like reference numbers indicate like parts throughout the several views.

[0028] Wireless powering can facilitate the transmission of electrical power wirelessly from generators to consumers by transmitting, e.g., solar power from space-based solar cells to a receiving station on the earth by a wireless powering microwave link. On the other scale, wireless powering can enable the use of battery-less devices such as RFID tags used in supply chain, access cards, logistics, tracking animals and goods. These passive, wirelessly powered tags can serve in biomedical implantable devices to provide improved health care. With the advance in the scaling of integrated circuits and the wide variety of the available sensors, wireless powering can be used to open doors for the Internet of the Things (IoT).

[0029] Research to make wireless powering links more efficient ranges from the selection of a suitable frequency and the design of the transmitting and receiving antennas, to work on the RF-to-DC power converters. These power converters are considered the core of a wireless power receiver and can define the efficiency of the wireless powering link. Mainly, AC-to-DC voltage rectifiers are used to convert the received RF power to a usable DC power that can be stored in a storage capacitor to provide a sustainable power supply. FIG. **1** shows block diagram of an example of a wireless power receiver **100**, which includes an antenna **103**, an input matching network **106** and an RF-to-DC power converter **109** that outputs DC power to an energy storage and power management unit **112**.

[0030] One of the most common implementations for the RF-to-DC power converter **109** is a full wave bridge rectifier. FIG. **2**A shows a schematic diagram of a full wave bridge rectifier. The operation depends on the rectification property of the electronic diodes and provides full wave current rectification of the received AC power. But, it requires a minimum input voltage higher than twice the diode built-in voltage to start the rectification and its four diodes still carry a voltage drop equals this built-in voltage. This degrades both the sensitivity and the efficiency of the AC-to-DC power converter, and thus the entire wireless power receiver. Sensitivity is related to the minimum RF input power (AC power) level needed by the rectifier to generate an acceptable (or usable) output DC voltage level to supply and operate the load circuits. A possible but expensive solution is to implement the full wave diode rectifier using Schottky diodes with a very low built-in voltage, but these diodes are not usually available in common low cost CMOS technologies. If they are available through high cost special CMOS processes, they generally have less accurate models for reliable design and development cycles. Other implementations have equivalent rectifier structures that may be achievable on conventional CMOS technologies.

[0031] One implementation of the AC-to-DC power converter can use fully cross-coupled MOSFET devices. FIG. **2**B shows a schematic diagram of a fully cross-coupled (FX) rectifier **200**. It can achieve a significant improvement in both the sensitivity and the efficiency of the power converter. But, includes a main limiting factor for the power conversion efficiency that makes it only efficient over a narrow range of input RF power levels. If the input RF power level exceeds a certain threshold, it starts leaking the stored energy from the storage to the input port. This reverse leakage deteriorates the power conversion efficiency at high RF power levels. Other implementations include an adaptive reconfigurable rectifier that utilizes fast comparators at 13 MHz, but that cannot be used at higher frequencies such as 950 MHz UHF RFID and a reconfigurable rectifier architecture on an intra-stage level using a reconfigurable connectivity between the rectifier stages.

[0032] Wide Range RF-to-DC Power Converter Architecture

[0033] An efficient, wide range RF-to-DC power converter stage architecture that can operate efficiently over more than double the RF input range of the fully cross-coupled rectifier and without degrading the sensitivity of the wireless power receiver will be presented in this disclosure. It mitigates the reverse leakage problem in fully cross-coupled rectifiers without degrading the sensitivity. It is designed and optimized for UHF 950 MHz RFID application and can be implemented using, e.g., AMS 0.18 μm CMOS technology or other suitable low cost CMOS technology process. It achieves a peak RF-to-DC power conversion efficiency of about 75% and maintains RF-to-DC power conversion efficiency (PCE) higher than 40% for a 16 dB dynamic range of input RF power levels. This dynamic range is more than double that of the standard fully cross-coupled rectifier. At same time, it can achieve a sensitivity of −17 dBm input RF power to provide output DC voltage of one Volt to a typical load of 100 kΩ.

[0034] Referring now to FIG. **3**A, shown is a schematic diagram illustrating an exam pe of a wide range fully cross-coupled rectifier architecture **300** *a* in accordance with various embodiments of the present disclosure. The wide range rectifier **300***a* can be designed and implemented using conventional, low cost, CMOS 0.18 μm technology. The PMOS transistors (MP**1** and MP**2**) and the NMOS transistors (MN**1** and MN**2**) are the rectifying devices and capacitors CP**1** through CP**4** are the input pumping capacitors. The pumping capacitors (CP**1**-CP**4**) provide two functions. They pump the charges from the RF input through the rectifying devices to the output storage capacitor, as well as decouple the DC bias of the gates of MP**1** and MP**2** from the gates of MN**1** and MN**2**.

[0035] At low input RF power levels, the fully cross-coupled rectifying devices (MP**1**, MP**2** MN**1**, MN**2**) start rectifying at a very sensitive level of RF input voltage in the same way as a standard fully cross-coupled rectifier. At high RF input power levels, the standard fully cross-coupled rectifier **200** shown in FIG. **2**B suffers from reverse leakage that deteriorates its power conversion efficiency. The power conversion efficiency (PCE) can be expressed as the ratio of the usable DC output power P_{DC} to the RF input power P_{RF} or the ratio of the usable DC output power P_{DC} to the total summation of the usable DC power and the total power loss P_{Loss} including the dissipation power loss P_D and the reverse leakage power loss P_{RVS}.

$$PCE = \frac{P_{DC}}{P_{RF}} = \frac{P_{DC}}{P_{DC} + P_D + P_{RVS}}. \quad (1)$$

The dissipation loss is the power dissipated internally in the rectifying devices of the RF-to-DC power converter and is directly related to the drop voltage over the rectifying devices times the conducting current in the devices. The reverse leakage loss is due to reverse leakage power that leaks from the charged output storage capacitor (CL) back to the antenna through the rectifying devices. As the input power increases, the leakage power loss P_{RVS} dominates the power loss and degrades the power conversion efficiency significantly. Other power losses can include mismatch loss where power is reflected due to an impedance mismatch between the antenna **103** (FIG. **1**) and the rectifier (**109**). As shown in FIG. **1**, input matching **106** can be used to reduce this effect.

[0036] On the other hand, the wide range rectifier **300***a* can start a smart DC biasing mechanism to the gates of the forward rectifying devices (MP**1** and MP**2**) through the feedback branch devices (FB**1** and FB**2**) to limit the undesired reverse leakage power loss from the energy storage capacitor (CL), which charges and holds the output DC voltage. The feedback branches include biasing elements to bias the gates of the forward rectifying devices (MP**1** and MP**2**). The biasing elements (FB**1** and FB**2**) can be implemented by, e.g., voltage controlled switches such as a MOSFET transistor operating in linear region (acting as a voltage controlled resistor) or by resistors (as there is no DC current flowing into the MOSFET gates). The feedback branches can provide a DC short circuit and an RF open circuit, similar to an RF Chock.

[0037] This DC biasing mechanism can be controlled by a power management unit (PMU) as shown in FIG. **3**A. The PMU can sense and/or sample the output voltage of the rectifier **300***a* , and provide a control signal (CTRL) to control the feedback branches based at least in part upon the output voltage. The PMU can comprise processing circuitry including a processor (or a system on chip (SoC) or microprocessor) or can be implemented with transistors. The output voltage sampling can be implemented by, e.g., a diode connected MOS transistor and a series resistor (or MOSFET in triode region). In other implementations, the microprocessor or SoC can include analog-to-digital converters for sampling the output voltage.

[0038] The gates of the PMOS transistors (MP**1** and MP**2**) are decoupled from the corresponding NMOS transistors (MN**1** and MN**2**), so that they can be biased and controlled independently. This smart active biasing is mainly controlled by the output voltage level. So that when the RF input power level increases, the feedback mechanism will start working smartly to limit the undesired reverse leakage power loss from the energy storage capacitor (CL). This boosts the RE-to-DC power conversion efficiency as expressed in Equation (1). By suppressing the reverse leakage power loss, the rectifier **300***a* achieves higher power conversion efficiency and also higher output DC voltage in the range of high RF input power levels, while maintaining the sensitive performance at low RF input power levels.

[0039] Referring next to FIGS. **3**B-**3**D, shown are schematic diagrams illustrating examples of a reconfigurable wide range rectifier architecture **300**. As shown in the rectifier **300***b* of FIG. **3**B, the DC voltage of the gates of the rectifying devices MP**1** and MP**2** are decoupled from the gates of devices MN**1** and MN**2** by adding two pumping capacitors (CP**3** and CP**4**). The rectifying devices (MP**1** and MP**2**) can be controlled by applying a DC voltage (CTRL) to their gates via feedback elements (FB**1** and FB**2**). By this control signal (CTRL), the rectifier **300** *b* can be reconfigured to optimize its efficiency for a wider range of operating RE input power levels. The feedback elements (FB**1** and FB**2**) should act as a DC short circuit and an AC (or RF) open circuit such as provided by an RE Choke (RFC). This can avoid disturbing the RF input power. In some implementations, the feedback elements (FB**1** and FB**2**) can be implemented in a rectifier **300** *c* by large resistors (RFP**1** and RFP**2**) as shown in the example of FIG. **3**C because, there is no dc current passing into the gates of the MOSFETs. In other implementations, the feedback elements (FB**1** and FB**2**) can be implemented in a rectifier **300** *d* by MOSFET transistors (M**1** and M**2**) as shown FIG. **3**D to avoid the parasitic capacitance of the large resistors (R**1** and R**2**) in an on-chip implementation.

[0040] Referring next to FIGS. **3**E and **3**F, shown are schematic diagrams illustrating examples of a self-adaptive wide range rectifier architecture **300**. The self-adaptive rectifiers **300** can work efficiently in various operating conditions. As shown in the rectifier **300** *e* of FIG. **3**E, the control signal (CTRL) is extracted from the rectifier **300***e* itself via the feedback elements (FB**1** and FB**2**), so the rectifier **300***e* is self-biased in the appropriate direction. When the RE input power increases, the self-bias mechanism will limit the reverse leakage of the rectifying devices (MP**1** and MP**2**) and enhance the overall power conversion efficiency. In some implementations, the feedback elements (FB**1** and FB**2**) can be implemented by large resistors (RFP**1** and RFP**2**) as shown in the rectifier **300** *f* of FIG. **3**F because there is no DC current passing into the gates of the MOSFETs.

[0041] Referring next to FIGS. **3**G and **3**H, shown are schematic diagrams illustrating examples of a reconfigurable, adaptive wide range rectifier architecture **300**. As shown by the rectifier **300** *g* of FIG. **3**G, the self-feedback biasing branches (FB**1** and FB**2**) can be controlled by a control signal (CTRL). In some implementations, the feedback elements (FB**1** and FB**2**) can be implemented by MOSFET transistors (M**1** and M**2**) in an on-chip implementation as shown the rectifier **300** *h* in FIG. **3**H. This architecture can provides better tracking to the rectifier condition and hence can provide the appropriate feedback to enhance the overall power conversion efficiency.

[0042] Referring next to FIGS. **3**I-**3**K, shown are schematic diagrams illustrating examples of a reconfigurable, adaptive wide range rectifier architecture **300** with a power management unit (PMU) as a fully integrated module for system on chip (SOC) solutions. The control signal (CTRL) is internally generated from a power management unit (PMU) as shown in the rectifier **300***i* of FIG. **3**I. In some implementations, the feedback elements (FB**1** and FB**2**) can be implemented by MOSFET transistors (M**1** and M**2**) as shown the rectifier **300***j* in FIG. **3**J. A simple method to implement the power management unit (PMU) with transistors is illustrated in the rectifier **300***k* of FIG. **3**K. In some embodiments, the PMU comprises a diode connected transistor and a resistive branch (implemented by a MOSFET operating in linear region).

[0043] It should be noted the described techniques are not limited to rectifying devices MP**1** and MP**2**. In various embodiments, the described techniques can be applied to devices MN**1** and MN**2** either separately or in combination with MP**1** and MP**2** For example, the techniques can be applied to all of the rectifying devices MP**1**, MP**2** MN**1** and MN**2** in a wide range of RF-to-DC power converters.

[0044] Simulation Results

[0045] The fully cross-coupled (FX) rectifier **200** of FIG. **2**B is commonly used for RF-to-DC power converter in RFID and wireless power receivers. Although it shows an improved sensitivity (better than the diode based rectifier of FIG. **2**A), it suffers from a severe drawback that degrades its overall power conversion efficiency (PCE) at high levels of input RF power. The plot of FIG. **4**A illustrates the degradation due to the reverse leakage. This degradation of PCE of the fully cross-coupled rectifier **200** in the high range of RF input power may be attributed to the reverse leakage from the output storage capacitor to the input of the FX rectifier **200**. This reverse leakage increases as the RF input power level increases.

[0046] While this loss mechanism is negligible in the diode based rectifier of FIG. **2**A in which the reverse current is very small, reverse leakage is very significant in fully cross-coupled rectifiers **200**. But, diode based rectifiers suffer from poor sensitivity and cannot work efficiently at low RF power levels as shown in FIG. **4**A. So, the reverse leakage limits the input range of the fully cross-coupled rectifier **200** which is a commonly used sensitive rectifier. This limits the operating range of the wireless power receiver in terms of RF power levels and distance.

[0047] The reverse leakage mainly comes from the fact that the MOSFET is originally a bidirectional device and remains bidirectional in the fully cross-coupled rectifier architecture **300**. FIG. **4**B shows a half cycle of the RF input voltage (**403**) and its corresponding output DC voltage (**406**). When the instantaneous RF input voltage is less than the output DC voltage over the output storage capacitor (CL), the stored energy in the output storage capacitor leaks through the MOSFET in the reverse direction and producing a periodic discharging current that passes in the rectifying devices.

[0048] As illustrated in FIG. **4**B, when RF input voltage **403** is higher than the output DC voltage **406**, the rectifying MOSFETS conduct in forward direction (**409**). When the instantaneous RF input voltage **403** is less than the output DC voltage **406**, the rectifying device conducts and leaks in reverse direction (**412**). Moreover, this reverse leakage current **412** increases as the RF input level increases and limits the overall PCE as shown in FIG. **4**A. As the RE input voltage **403** increases, the reverse leakage **412** increases more than the forward current **409** increase. So, the overall power conversion efficiency degrades as the RE input voltage increases. In contrast, the wide range rectifier architecture **300** extends the power conversion efficiency over a wider range of RF input power as shown in FIG. **4**C, which can enhance the PCE.

[0049] Simulation results of a wide range RE-to-DC power converter using the wide range rectifier architecture **300***a* of FIG. **3**A were implemented on CMOS 0.18 μm technology and are presented in FIGS. **5**A and **5**B. The simulations used an IC industry standard, Electronic Design Automation (EDA) Simulator from Cadence. The performance of the rectifier **300***a* is compared with the full wave bridge rectifier of FIG. **2**A and the fully cross coupled rectifier **200** of FIG. **2**B. The presented results are based on an RE input signal at 950 MHz and a 100 kΩ loading to test the performance of a single stage of the rectifier architecture **300** at the UHF range used for RFID applications.

[0050] The RF-to-DC power conversion efficiency (PCE) is shown in FIG. **5**A for the wide range rectifier **300***a* (curve **503**) and the PCE of the standard fully cross-coupled rectifier **200** (curve **506**) over the RF input power range. Both architectures show similar peak power conversion efficiency of about 75%. However, the wide range rectifier **300** is able to maintain a PCE that is higher than 40% for a 16 dB (39.8 ratio) range of RF input power levels, which is more than double the 12 dB range of the standard fully cross-coupled rectifier **200** (a 15.8 ratio).

[0051] At the same time, the wide range rectifier **300***a* provides a higher output voltage than the standard fully cross-coupled rectifier **200** as shown in FIG. **5**B. This boost in PCE performance is achieved while maintaining a superior sensitivity as illustrated in FIG. **5**B. The adaptive wide range rectifier **300** can provide a DC output voltage of one Volt from −17 dBm RF input power. On the other hand, the full wave diode bridge rectifier shows very poor sensitivity and efficiency at low and mid-range of input power.

[0052] In this disclosure, a highly efficient, fully cross-coupled rectifier **300** with a wide input range was presented. The architecture of the wide range rectifier **300** and its performance were discussed and compared with the standard fully cross-coupled rectifier **200**. Self-adaptive architectures were discussed by which no additional components and no external control are needed. The wide range rectifier architecture **300** achieves a peak RF-to-DC power conversion efficiency of 75% and maintains a power conversion efficiency higher than 40% for 16 dBm dynamic range of RF input power levels, which is more than double the input range of the fully cross-coupled rectifier **200**. At same time, the wide range rectifier **300** achieves sensitivity of −17 dBm RF input power to provide output DC voltage of one Volt to a typical load of 100 kΩ.

[0053] A self-biased, cross-coupled differential rectifier was examined with enhanced power-conversion efficiency over an extended range of input power. A prototype was designed for UHF 433 MHz RF power-harvesting applications and was implemented using 0.18 μm CMOS technology. The proposed self-biased rectifier architecture was compared to a conventional cross-coupled rectifier. Validation demonstrated an efficiency improvement of more than 40% and an input power range extension of more than 50%. A sensitivity of −15.2 dBm (30 μW) input power for 1V output voltage and a peak power-conversion efficiency of 65% are achieved for a 50 kΩ load.

[0054] Traditionally, AC-DC power converters have been realized using diode-based architectures such as the Greinacher cell, the Dickson multiplier, or a full-wave bridge rectifier. Diode-based rectifiers have simply been realized in CMOS technology using diode-connected transistors, such as the Dickson rectifier shown in FIG. **6**A. However, they suffer from poor sensitivity at low input power and from high dropout voltage that degrades their power conversion efficiency at mid-to-high input power. Enhanced sensitivity can be achieved using Schottky diodes that require additional fabrication steps and hence are seldom offered in conventional CMOS processes. On the other hand, the differential, fully cross-coupled (FX) rectifier **200**, shown in FIG. **6**B (and FIG. **2**B), is widely used in RFID applications due to its improved sensitivity and high peak efficiency. However, this improvement is gained at the expense of the reverse characteristic of the rectifying devices, because the rectifying cross-coupled transistor is still a bidirectional device—unlike diodes or diode-connected transistors. Hence, it conducts in a reverse direction I_{RVS} once the instantaneous value of the RF signal, V_{RF} becomes lower than the output DC voltage, V_{DD}, in every RF cycle, as shown in FIG. **7**A. This periodic reverse leakage is exacerbated as the RF power level grows, degrading the rectifier conversion efficiency at high input power, as shown in FIG. **7**B. Consequently, the rectifier operates efficiently within a limited range of input power. To evaluate different rectifier architectures, a dynamic range is defined as the input-power range at which the rectifier maintains a power conversion efficiency (PCE) higher than 80% of its peak efficiency.

[0055] A low MHz adaptive architecture can reconfigure the rectifier stage as a diode-connected voltage doubler or as a half cross-coupled rectifier based on the input power level. However, it utilizes active diodes driven by fast comparators that are not power efficient at high RF frequencies such as UHF 433 MHz or higher. A multistage configuration can be used to enhance the efficiency of the cross-coupled rectifier by rewiring stages as a series or parallel multistage rectifier. Though this technique extends the efficiency of the cross-coupled rectifier over a wide range of input power, it includes multiple stages with high input capacitance and low input resistance, thus complicating the matching network design. An adaptive offset calibration technique can be used to compensate the threshold voltage, V_{th}, in diode based rectifiers to improve their sensitivity. However, it utilizes two more auxiliary, multistage rectifiers to compensate the V_{th} of the main rectifier, thus occupying a larger area, adding more losses, and degrading the PCE.

[0056] A self-biased, cross-coupled, differential CMOS rectifier **300***f* (FIG. **3**F) was experimentally validated and found to have enhanced efficiency over wider input range than both the diode connected rectifier of FIG. **6**A and the cross-coupled rectifier **200** of FIG. **6**B. This facilitates reliable and efficient RF-to-DC power conversion at varying RF power levels, and it adds more spatial freedom between the wireless power transmitter and receiver. The self-biased cross-coupled architecture is illustrated by the schematic diagram in FIG. **8** and its operational concept and the experimental results are discussed and compared with measurements of a fabricated conventional FX rectifier **200**. The proposed self-biased rectifier **300***f* was implemented in a 0.18 μm CMOS process technology, and occupied a 130 μm×130 μm active area. A die microphotograph of the fabricated self-biased rectifier **300***f* is shown in FIG. **9**, with an enlarged view of the active area including the feedback resistors (RFB). A conventional FX rectifier was implemented on the same die for a fair comparison to the self-biased rectifier **300***f* using an identical test setup and conditions. The RF measurement setup included an Agilent's vector network analyzer (N5225A) and a digital multimeter (34420A).

[0057] As shown in the schematic diagram of FIG. **8**, the fabricated self-biased cross-coupled rectifier **300***f* utilizes the cross-coupled configuration with the differential-drive capability to maintain good sensitivity at low input power by operating in the linear region and holding a low dropout voltage. Moreover, a self-biasing mechanism is added to limit the reverse leakage when sensing high input power. The self-biasing mechanism can control the conduction of the rectifying devices by raising their effective turn-on voltage at high RF power levels. A simple implementation of this mechanism can be achieved by applying the output DC voltage directly to the controlling gates of the rectifying transistors (MP**1** and MP**2**) without disturbing the RF signal at the differential inputs. This can be achieved by decoupling the DC voltage of the rectifying PMOS MP**1**/MP**2** gates apart of their corresponding NMOS MN**1**/NM**2** transistors using capacitors CP**3** and CP**4** and then, applying the DC self-bias using an RF choke (RFC) coil (not shown) that is DC short circuit and AC open circuit. It is worth noting that the self-biasing branch is connected to a MOSFET gate: hence, no DC current passes through it, whereas the self-biasing RFC branches can be simply replaced by high feedback resistors (e.g., RFB**1** and RFB**2** with a resistance of about 100 kΩ) without loading the RF inputs, as shown in FIG. **8**.

[0058] The operating point of the self-biased cross-coupled rectifying device (MP**1** in FIG. **8**) was compared to the conventional cross-coupled and diode-connected rectifying devices (**MP1** in FIGS. **6**A and **6**B, respectively). FIG. **10** illustrates the operating point of the rectifying device (**MP1**) in the three configurations: a diode-connected transistor (FIG. **6**A); a cross-coupled transistor (FIG. **6**B); and a self-biased transistor (FIG. **8**). Each rectifying device **MP1** connects the AC-coupled RF node (RF superimposed on a common mode $V_{CM}=V_{DD}/2$) at its input terminal to the output V_{DD} point.

[0059] As shown in FIG. **10**(*a*), the gate and drain of the diode-connected device **MP1** are tied to the output V_{DD} to act as a two-terminal diode. In the case of the conventional cross-coupled device **MP1** shown in FIG. **10**(*b*), the gate is attached to the RF_N node ($=-V_{RF}/2$ superimposed on $V_{DD}/2$). However in the self-biased cross-coupled device **MP1** shown in FIG. **10**(*c*), the gate RF_Nsignal is superimposed on V_{DD}. Note that both the self-biased cross-coupled device of FIG. **10**(*b*) and the diode-connected device of FIG. **9**(*a*) need a minimum turn-on RF voltage, V_{RFmin}, equal to threshold voltage, V_{th}, plus $V_{DD}/2$. This is unlike the conventional cross-coupled device of FIG. **10**(*c*), which requires V_{th} only.

[0060] Moreover, the device of FIG. **10**(*c*) conducts in the linear region and holds a dropout voltage of less than or equal to its overdrive voltage, V_{eff}, passing a higher output voltage than the diode-connected device of FIG. **10**(*a*), which requires a dropout of $V_{th}+V_{eff}$. As a result, the self-biased cross-coupled device of FIG. **10**(*c*) retains moderate sensitivity (i.e., it generates a higher output voltage at the same RF input power) compared to the diode-connected device of FIG. **9**(*a*). This is derived from the operating points shown in FIG. **10**, and it matches the simulated dropout voltages of the three rectifiers at the same input power level of 37 μW, as shown by the simulation results of the dropout voltage (V_{SD}) in FIG. **11**(*a*).

[0061] During the periodic RF signal transitions, the reverse leakage of the diode-connected device of FIG. **10**(*a*) is negligible, while the conventional cross-coupled device of FIG. **10**(*b*) is strongly biased by $V_{SG(RVS)}=V_{RF}/2+V_{DD}/2$ in the reverse direction. Accordingly, the reverse current lobes increase at the 37 μW input power level, as shown by the simulation results of the drain current (I_{DP}) in the rectifying PMOS devices of the Dickson, conventional FX, and proposed self-biased cross-coupled rectifiers in FIG. **11**(*b*). On the other hand, as depicted in FIG. **10**(*c*), the device is biased only at $V_{SG(RVS)}=V_{RF}/2$, leading to reduced reverse current lobes, as shown in FIG. **11**(*b*). It worth noting that, during the negative RF half cycle, the gate of **MP1** in the conventional FX rectifier of FIG. **10**(*b*) and self-biased rectifier of FIG. **10**(*c*) is positively biased, shutting down its current completely, like the reverse-biased, diode-connected device.

[0062] FIG. **12**(*a*) shows the simulation of the output voltage V_{DD} versus the input power P_{RF} for all three rectifiers for a load of 50 kΩ. At low RF power, the proposed self-biased rectifier **300*f*** out performs the Dickson rectifier while it out performs the FX rectifier at high RF power. This may be explained by plotting the alternating charges flowing in **MP1** per RF cycle Q_P (with the left bar representing the forward harvested charges and the right bar representing the reverse leakage charges) at three levels of distinct RF power P_{RF}, as shown in FIG. **12**(*b*). The simulated PCE of the three rectifiers versus input power P_{RF} is plotted in FIG. **12**(*c*). As shown in FIG. **12**, the conventional FX rectifier **200** (FIG. **6**B) has a peak PCE at low RE power, but it falls off rapidly due to its high reverse leakage current, which is shown in FIG. **11**(*b*). On the other hand, the Dickson (diode-based) rectifier (FIG. **6**A) displays a moderate PCE at the high RF power levels but suffers from poor efficiency at low input power due to its high dropout voltage, as shown in FIG. **11**(*a*). Conversely, the proposed self-biased cross-coupled rectifier **300*f*** (FIG. **8**) achieves a wider dynamic range with a slight decrease in its peak PCE relative to the conventional FX rectifier **200**.

[0063] A comparison of the three rectifying devices is summarized in the table of FIG. **13**, where the self-biased rectifying device **300*f*** shows competitive advantages of low dropout voltage, as shown in FIG. **11**(*a*), and hence low dissipation loss and better sensitivity than the diode-connected device of FIG. **6**A. At the same time, the self-biased rectifying device **300*f*** shows lower reverse leakage and wider input range than the conventional cross-coupled device **200**, as shown in FIGS. **11**(*b*) and **12**, respectively.

[0064] The PCE of the fabricated rectifiers was measured with a single tone 433 MHz signal at different input power levels, and the output DC voltage was recorded. After de-embedding the reflection and transmission losses, the net input power was calculated and the power conversion efficiency for the proposed self-biased rectifier **300*f*** and the conventional FX rectifier **200** was plotted in FIG. **14** versus the input power at different loads. The measured PCE of the self-biased rectifier **300*f*** (solid curves) and the conventional FX rectifier **200** (dashed curves) was plotted for three loads: 50 kΩ, 100 kΩ, and 200 kΩ. Although the self-biased rectifier **300*f*** exhibited a lower peak PCE (about 65%) than the conventional rectifier **200**, whose peak PCE was about 75%, the self-biased architecture maintained its power conversion efficiency over a broader dynamic range of input power. Thus, the proposed self-biased rectifier **300*f*** can operate efficiently at different RE power levels, enabling robust wireless powering from varying transmission distances or within unstable environments.

[0065] To elucidate the relative PCE enhancement of the proposed rectifier, the ratio of the self-biased rectifier PCE over the conventional FX rectifier PCE is plotted in FIG. **15**(*a*) for variable loads of R_L=50 kΩ, 100 kΩ, and 200 kΩ. An approximately 40% to 70% efficiency improvement was achieved in the proposed self-biased rectifier **300*f*** relative to the conventional FX rectifier **200**. Moreover, the self-biased rectifier **300*f*** maintained the achieved improvement of PCE over a wider input range at different loading conditions, as depicted in FIGS. **14** and **15**(*a*). To further illustrate this point, the dynamic range of input power to [P_{min} to P_{max}] at which each rectifier maintained 80% of its peak efficiency was represented as the ratio of P_{max}/P_{min} and is compared in FIG. **15**(*b*) for the self-biased rectifier **300*f*** and conventional FX rectifier **200** over different loading conditions. Evidentially, the proposed self-biased rectifier **300*f*** achieved a higher dynamic range across a wide range of loading conditions (from 30 kΩ to 300 kΩ). FIG. **16** depicts the measured output DC voltage versus input power level for the self-biased rectifier **300*f*** and conventional FX rectifier **200** equally loaded by 50 kΩ, 100 kΩ and 200 kΩ. At the same input power, the proposed self-biased rectifier **300*f*** delivered higher DC output voltage than the conventional FX rectifier 200 by 20% to 30%.

[0066] A performance comparison of the self-biased rectifier **300*f*** versus recent ISM 433 MHz rectifiers is summarized in the table in FIG. **17**. The proposed self-biased rectifier **300*f*** shows 2× improvement in dynamic range and 5× enhancement in efficiency relative to the fully cross-coupled rectifier with inter-stage RF injection described in "A novel cascading scheme to improve the performance of voltage multiplier circuits" by Chouhan et al. (*Analog Integr. Circ. Sig. Process,* vol. 84, no. 3, pp. 373-381, 2015). More than 50% wider dynamic range and doubled efficiency was achieved compared to the Dickson rectifier with multiple V_{th} offset cancellation described in "Auto-calibrating threshold compensation technique for RF energy harvesters" by Gharehbaghi et al. (*IEEE RFIC Symposium,* May 2015, pp, 179-182) with a 9× chip area saving. The proposed self-biased, cross-coupled, differential rectifier **300***f* was shown to provide 50% enhanced PCE over a wide range of input power levels and under different loading conditions when compared to other rectifiers. A sensitivity of (P_{in}=30 μW at V_0=1V) and peak PCE=65.3% were achieved for a 50 kΩ load.

[0067] A wide-range, differential cross-coupled rectifier **300***k* (FIG. **3**K) was also examined with an extended dynamic range of input RF power that enables wireless powering from varying distances. The proposed self-adaptive architecture can mitigate the reverse-leakage problem in conventional, cross-coupled rectifiers without degrading sensitivity. A prototype was designed for UHF RFID applications, and was implemented using 0.18 μm CMOS technology. On-chip measurements demonstrate a sensitivity of −18 dBm for 1V output over a 100 kΩ load and a peak RE-to-DC power conversion efficiency of 65%. A conventional, fully cross-coupled rectifier was fabricated alongside for comparison and the proposed rectifier shows more than 2× increase in dynamic range and a 25% boosting in output voltage than the conventional rectifier.

[0068] A wide-range, differential RF-to-DC power converter stage using an adaptive, self-biasing technique was examined and found to extend the optimal input range at which the rectifier maintains 90% of its peak power conversion efficiency over a wider input range than conventional rectifiers. This can be achieved without trading off the sensitivity advantage of the cross-coupled scheme. The adaptive architecture is illustrated by the schematic diagram in FIG. **18** and the experimental results are presented. The proposed adaptive rectifier **300***k* was implemented in a 0.18 μm CMOS process technology and occupied a 65×130 μm active area including the two additional capacitors and transistors. A microphotograph of the module including the testing pads is shown in FIG. **19**, with an enlarged view of the active area. All tests are repeated for a single-stage, conventional, FX rectifier fabricated on the same die using the same test setup for comparison purposes to illustrate the advantage of the cross-coupled scheme.

[0069] Referring back to FIG. **18**, the operation principle of the proposed adaptive rectifier **300***k* is based on limiting the conduction of the rectifying devices at high input power range at which the reverse leakage becomes severe. This is achieved by raising the DC bias of the rectifying transistors (MP**1** and MP**2**) gates as the input RF power increases. The proposed rectifier is comprised of four driving MOSFETs (MP**1**, MP**2** MN**1** and MN**2**) as rectifying devices that are connected in a fully, cross-coupled manner to enhance the rectifier sensitivity. However, unlike the conventional cross-coupled rectifier shown in FIG. **2**B, the DC voltage of the gates of MP**1** and MP**2** are decoupled from their corresponding NMOS gates, while the AC voltages of the PMOS gates are coupled to the corresponding AC voltage of NMOS gates via two pumping capacitors CP**3** and CP**4**.

[0070] To control the conduction of the rectifying devices, self DC biasing can be dynamically applied to the PMOS gates only in the presence of high input RF power. While this self-biasing can be implemented in various ways, a simple yet dynamic approach is adopted by using two feedback transistors MN**5** and MN**6**, as shown in FIG. **18**. Here, MN**5** and MN**6** are controlled by a 'Sense' signal which can be derived from the output DC voltage (VDD) using a potential divider of a diode-connected transistor MN**3** and a linear transistor MN**4**.

[0071] FIG. **20**(*a*) shows the simulated currents of the rectifying PMOS devices (solid curves) and NMOS devices (dashed curves) in a conventional FX rectifier **200** (FIG. **2**B) and proposed adaptive rectifier **300***k* (FIG. **18**). In the presence of high input RF power, the output DC voltage and the corresponding 'Sense' signal will have high voltage levels. This high 'Sense' signal switches on the biasing transistors (MN**5** and MN**6**), thus applying dynamic biasing to the gates of the PMOS rectifying devices to limit their reverse leakage, as shown in FIG. **20**(*a*). Although the DC bias applied to the PMOS gate limits both its forward and reverse conductivity, it does not increase its conduction loss as it holds low voltage drop by operating in linear region while passing lower current than the conventional rectifier, as shown in FIG. **20**(*a*). The rectifying PMOS and NMOS currents and alternating charges per RF cycle that produce 1.6V output DC voltage from the proposed and conventional rectifiers are plotted in FIGS. **20**(*a*) and **20**(*b*), respectively. As expected, both rectifiers utilize the same net charges, ΔQ_f, to produce the same output voltage for equal loading. However, the magnitude of alternating charges flowing back and forth is much larger (5×) in the case of the conventional FX rectifier **200** compared to the proposed adaptive rectifier **300***k,* demanding higher RF power for the same output voltage. Although the adaptive bias circuit is applied to the PMOS transistors in the example of FIG. **18**, reverse leakage for NMOS can also be reduced due to the reduced effective bias leading to a 3× charge saving, as shown in FIG. **20**(*b*).

[0072] On the other hand, at low RF power, the sampled 'Sense' signal is low and MN5 and MN**6** are off. At the same time, the RF voltage can still be applied differentially to the gate-source terminals of the four rectifying devices as in the conventional, cross-coupled rectifier **200**. This preserves the main sensitivity advantage of the cross-coupled scheme to the low input RF power levels. Moreover, the proposed adaptive implementation has minimal loading effects. The sensing branch (MN**3** and MN**4**) can be designed such that it has a low quiescent current (e.g., about 1.3 μA at V_o=1.2V, R_L=100 kΩ) by increasing the length of the diode connected MN**3** and the linear-resistive MN**4**, while it is OFF at $V_o<V_{th}$. Also, the biasing branches (MN**5** and MN**6**) have no static power consumption as they are connected to gates of MOS devices (MP**1** and MP**2**) and to the decoupling capacitors.

[0073] During the evaluation, the input reflection coefficient was first measured using Agilent's vector network analyzer (N5225A) at different input RF power levels. Then, the input RF power (P_{in}) was deduced from the RF source power level after the reflection and transmission losses were de-embedded. The RF measurements were carried out using a 1 GHz sinusoidal signal fed to a single-stage adaptive rectifier loaded by a 100 kΩ. The measured RF-to-DC power conversion efficiency (PCE) is in good agreement with the simulation results for the proposed adaptive rectifier **300***k* and the conventional FX rectifier **200** as shown in FIG. **21**, where the simulated results are plotted with dashed curves and the measured results are plotted with solid curves. The adaptive rectifier **300***k* was verified to maintain a high PCE over an extended input range of RF power (more than 2× compared to the conventional FX rectifier **200**), as shown in FIG. **21**. This enables the proposed adaptive rectifier **300***k* to operate at both low RF power level (e.g., at a long wireless power range) with superior sensitivity and also at high RF power level with extended efficiency. The measured output DC voltage of the proposed adaptive rectifier **300***k* and the conventional FX rectifier **200** are shown in FIG. **22**, where the simulated results are plotted with dashed curves and the measured results are plotted with solid curves. At the same input RF power level, the adaptive rectifier **300***k* achieves a boosted output DC voltage which is 25% higher than that of the conventional FX rectifier **200** under the same loading conditions. Furthermore, voltage boosting ranges from 24.5% to 29% were measured over varying loads from 50 k to 500 kΩ. While the proposed adaptive rectifier **300***k* achieves extended optimal range, it still maintains a high peak efficiency of 65% and sensitivity of −18 dBm for 1V output over 100 kΩ.

[0074] A performance comparison of the adaptive rectifier **300***k* with prior art rectifiers is presented in the table in FIG. **23**. The optimal range is defined as the input power range over which the rectifier maintains power conversion efficiency higher than 90% of its peak efficiency. The proposed adaptive architecture achieved a much wider optimal range compared to previously reported architectures; including the most common, fully cross-coupled rectifier in "High-efficiency differential-drive CMOS rectifier for UHF RFIDs" by Kotani et al, (*IEEE J. Solid-State Circuits,* vol. 44, no. 11, pp. 3011-3018, Nov, 2009) with 5 dB better sensitivity. More than 2× improvement was achieved compared to the enhanced, 36-stage voltage doubler with floating gate devices in "Efficient far-field radio frequency energy harvesting for passively powered sensor networks" by Le et al. (*IEEE J. Solid State Circuits,* vol. 43, no. 5, pp. 1287-1302, May 2008). Also, the proposed architecture provides a wide optimal range without sacrificing neither the PCE as in "Codesign of a CMOS rectifier and small loop antenna for highly sensitive RF energy harvesters" by Stoopman et al. (*IEEE J. Solid-State Circuits,* vol. 49, no. 3, pp. 622-634, March 2014) nor sensitivity as in Kotani and Le. Although, a reconfigurable 4-stage rectifier in "A reconfigurable differential CMOS RF energy scavenger with 60% peak efficiency and −21 dBm sensitivity" by Scorcioni et al. (*IEEE Microw. Wireless Compon. Lett.,* 2013) shows comparable optimal range, it utilizes multi-stage re-configuration and was measured over a tracking load (a regulator with external reference). It is worth noting that the proposed single-stage adaptive architecture can be combined with this tracking load and multi-stage techniques in Scorcioni and "2.4-GHz High-efficiency adaptive power harvester" by Li et al. (*IEEE Trans, VLSI Syst.,* vol. PP, no. 99, p. 1, 2013) to build a wider range, multi-stage rectifier.

[0075] The self-adaptive, differential, cross-coupled rectifier **300***k* can operate with an extended input RF power range without trading off the sensitivity (P_{in}=−18 dBm at V_o=1V) or the peak power conversion efficiency (peak PCE=65%). The architecture of the adaptive rectifier **300***k* was presented, and its performance was verified and compared with the conventional, fully cross-coupled rectifier **200**.

[0076] It should be emphasized that the above-described embodiments of the present disclosure are merely possible examples of implementations set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and protected by the following claims.

[0077] It should be noted that ratios, concentrations, amounts, and other numerical data may be expressed herein in a range format. It is to be understood that such a range format is used for convenience and brevity, and thus, should be interpreted in a flexible manner to include not only the numerical values explicitly recited as the limits of the range, but also to include all the individual numerical values or sub-ranges encompassed within that range as if each numerical value and sub-range is explicitly recited. To illustrate, a concentration range of "about 0.1% to about 5%" should be interpreted to include not only the explicitly recited concentration of about 0.1 wt % to about 5 wt %, but also include individual concentrations (e.g., 1%, 2%, 3%, and 4%) and the sub-ranges (e.g., 0.5%, 1.1%, 2.2%, 3.3%, and 4.4%) within the indicated range. The term "about" can include traditional rounding according to significant figures of numerical values. In addition, the phrase "about 'x' to 'y'" includes "about 'x' to about 'y'".

1. A radio frequency to direct current (RF-to-DC) converter, comprising:

a fully cross-coupled rectification circuit including a pair of forward rectifying transistors; and

a feedback circuit configured to provide feedback bias signals to gates of the pair of forward rectifying transistors via feedback branch elements.

2. The RF-to-DC converter of claim **1**, wherein the feedback branch elements comprise controlled elements.

3. The RF-to-DC converter of claim **2**, wherein the controlled elements are MOSFET transistors.

4. The RF-to-DC converter of claim **1**, wherein the feedback branch elements are resistors.

5. The RF-to-DC converter of claim **1**, wherein the feedback bias signals are supplied from a common control input via the feedback branch elements.

6. The RF-to-DC converter of claim **1**, wherein the feedback bias signals are supplied from a common bias input via the feedback branch elements, the feedback bias signals controlled via a common control input.

7. The RF-to-DC converter of claim **6**, comprising a power management unit (PMU) coupled to the common control input.

8. The RF-to-DC converter of claim **7**, wherein the PMU provides a control signal to the common control input based at least in part upon an output voltage of the RF-to-DC converter.

9. The RF-to-DC converter of claim **1**, wherein the feedback bias signals are supplied from a common output connection of the pair of forward rectifying transistors via the feedback branch elements.

10. The RF-to-DC converter of claim **9**, wherein the common output connection couples drains of the pair of forward rectifying transistors.

11. A method, comprising:
receiving a radio frequency (RF) signal;
rectifying the RF signal via a fully cross-coupled rectification circuit including a pair of forward rectifying transistors; and
providing a DC output voltage from an output connection of the fully cross-coupled rectification circuit, where gating of the pair of forward rectifying transistors is controlled by feedback bias signals provided to gates of the pair of forward rectifying transistors via feedback branch elements.

12. The method of claim **11**, comprising:
sensing the DC output voltage; and
providing a control signal to the feedback branch elements based upon the DC output voltage.

13. The method of claim **12**, wherein the control signal is provided by a power management unit (PMU) configured to sample the DC output voltage.

14. The method of claim **12**, wherein the control signal is provided by a transistor sensing network coupled to the output connection of the fully cross-coupled rectification circuit.

15. The method of claim **12**, wherein the feedback branch elements are resistors coupled between the gates of the pair of forward rectifying transistors and the output connection.

* * * * *

US008489199B2

(12) **United States Patent**
Rofougaran

(10) **Patent No.: US 8,489,199 B2**
(45) **Date of Patent: Jul. 16, 2013**

(54) **BIO-MEDICAL UNIT WITH POWER HARVESTING MODULE AND RF COMMUNICATION**

(75) Inventor: **Ahmadreza (Reza) Rofougaran**, Newport Coast, CA (US)

(73) Assignee: **Broadcom Corporation**, Irvine, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 519 days.

(21) Appl. No.: **12/783,649**

(22) Filed: **May 20, 2010**

(65) **Prior Publication Data**

US 2011/0077714 A1 Mar. 31, 2011

Related U.S. Application Data

(60) Provisional application No. 61/247,060, filed on Sep. 30, 2009.

(51) **Int. Cl.**
A61N 1/378 (2006.01)

(52) **U.S. Cl.**
USPC **607/60**; 607/61

(58) **Field of Classification Search**
USPC 607/30, 32, 33, 60, 61
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,125,382	B2	10/2006	Zhou et al.
7,241,266	B2	7/2007	Zhou et al.
7,266,269	B2	9/2007	Koste et al.

OTHER PUBLICATIONS

Arjang Hassibi, et al., "A Spectral-Scanning Nuclear Magnetic Resonance Imaging (MRI) Transceiver", IEEE Journal of Solid-State Circuits, Jun. 2009, pp. 1805-1813, vol. 44, No. 6.

John E. Speich, et al., "Medical Robotics", Encyclopedia of Biomaterials and Biomedical Engineering, 2004, pp. 983-992, Marcel Dekker, Inc.

Eric Freudenthal, et al., "Evaluation of HF RFID for Implanted Medical Applications", Apr. 16, 2006, 4 pp.

Gibson, Ultrasound as a Proposed Drug Release Mechanism in Biomedical Microrobots, Arizona Space Grant Consortium, Univ. of Arizona Advanced Microsystems Laboratory, Dept. of Aerospace and Mechanical Engineering, Apr. 19, 2008, 17 pages.

Marin Soljacic et al., "WiTricity: Non-Radiactive Wireless Power Transfer," Oct. 9, 2008, 56 pages.

Robert E. Carlson, Ph.D., et al., "Development of an Implantable Glucose Sensor," Dec. 4, 2007, 16 pages.

Shekhar Bhansali, "Role of MEMS and Nanotechnology in Medical Technologies," University of South Florida, Mar. 1, 2004, 29 pages.

Primary Examiner — Carl H Layno
Assistant Examiner — Paula J Stice
(74) *Attorney, Agent, or Firm* — Garlick & Markison

(57) **ABSTRACT**

A bio-medical unit includes a power harvesting module and an RF communication module. The power harvesting module generates a power supply voltage from the electromagnetic signal. The RF communication module interprets a predetermined pattern of the transmitting of the electromagnetic signal to determine null intervals; receives a downstream RF communication signal during at least one of the null intervals; converts the downstream RF communication signal into downstream information; converts upstream information into an upstream RF communication signal; and/or transmits the upstream RF communication signal during one or more of the null intervals.

20 Claims, 27 Drawing Sheets

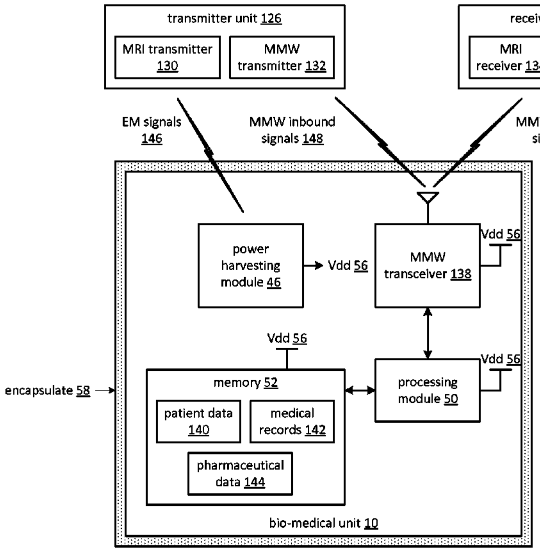
transmitter unit 126
MRI transmitter 130
MMW transmitter 132
receiver unit 128
MRI receiver 134
MMW receiver 136
EM signals 146
MMW inbound signals 148
MMW outbound signals 150
power harvesting module 46
Vdd 56
MMW transceiver 138
Vdd 56
Vdd 56
memory 52
patient data 140
medical records 142
pharmaceutical data 144
Vdd 56
processing module 50
encapsulate 58
bio-medical unit 10

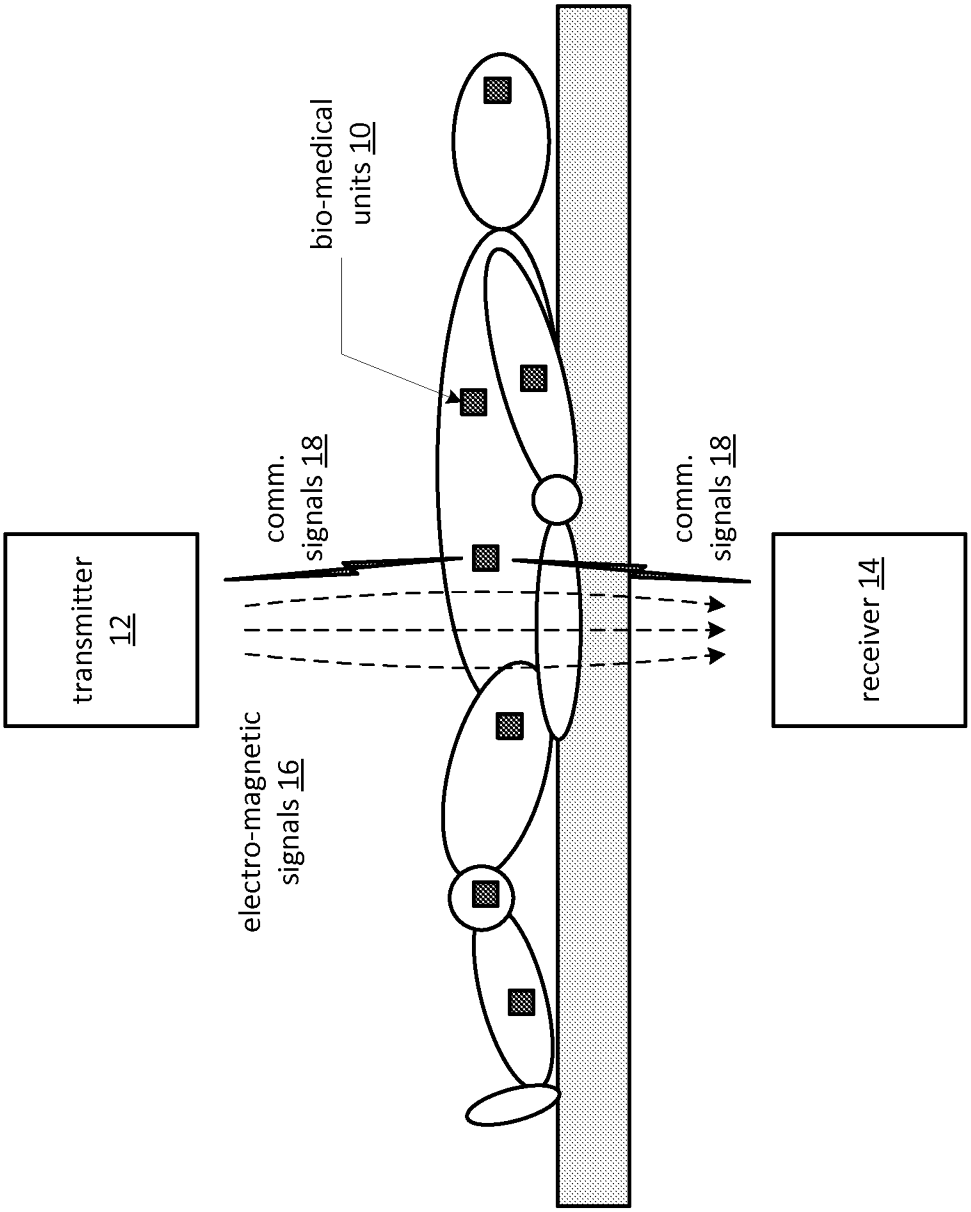
transmitter 12
receiver 14
bio-medical units 10
comm. signals 18
comm. signals 18
electro-magnetic signals 16

FIG. 1

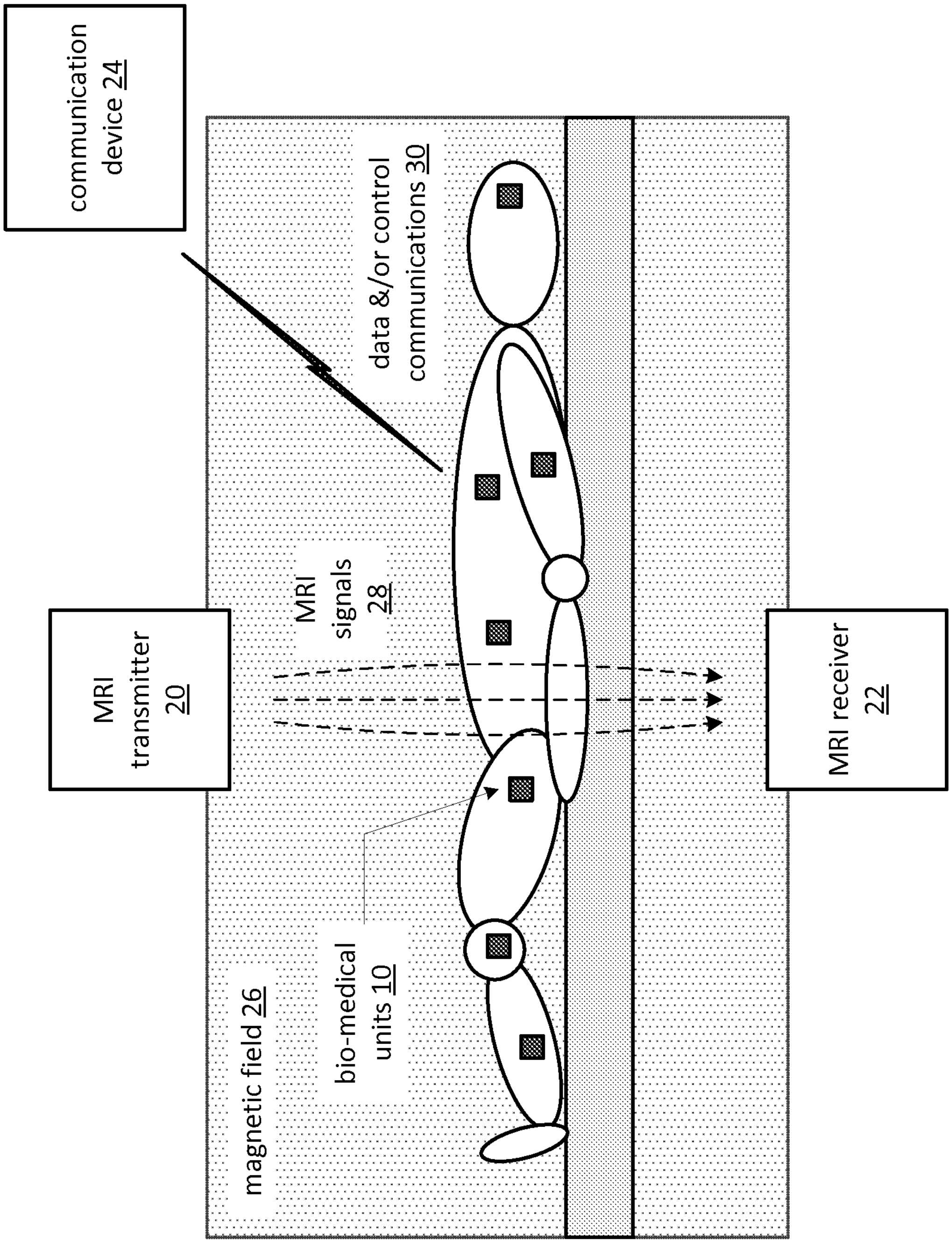
communication device 24
data &/or control communications 30
MRI transmitter 20
MRI signals 28
MRI receiver 22
bio-medical units 10
magnetic field 26

FIG. 2

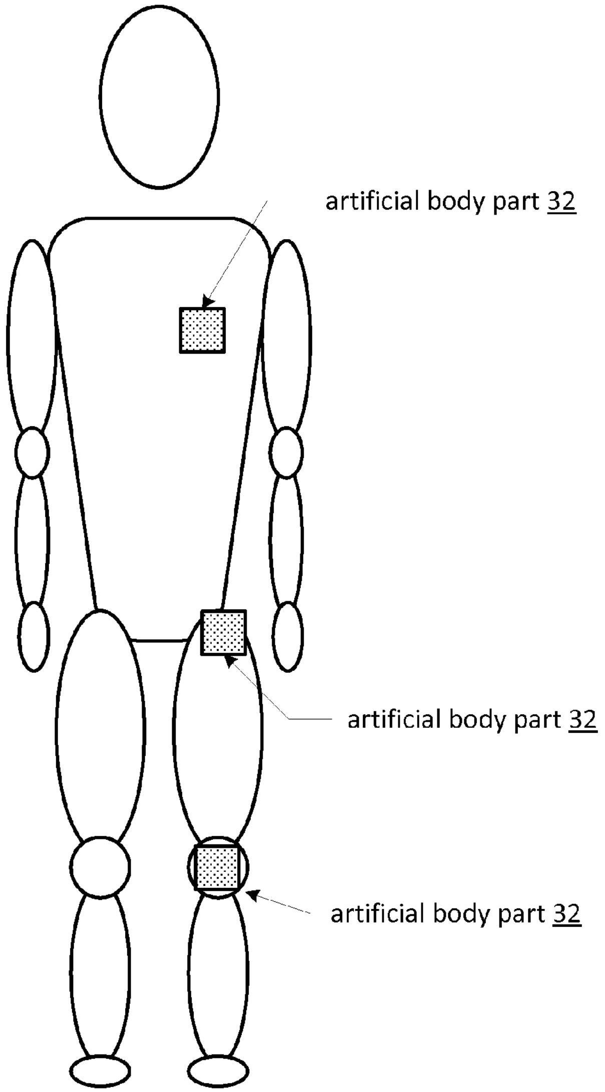
artificial body part 32
artificial body part 32
artificial body part 32

FIG. 3

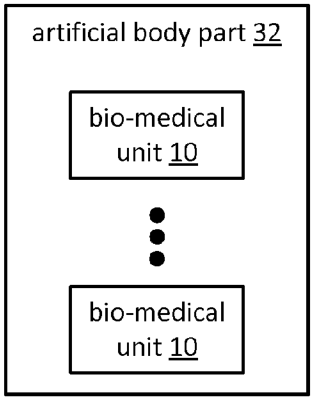
artificial body part 32
bio-medical unit 10
bio-medical unit 10

FIG. 4

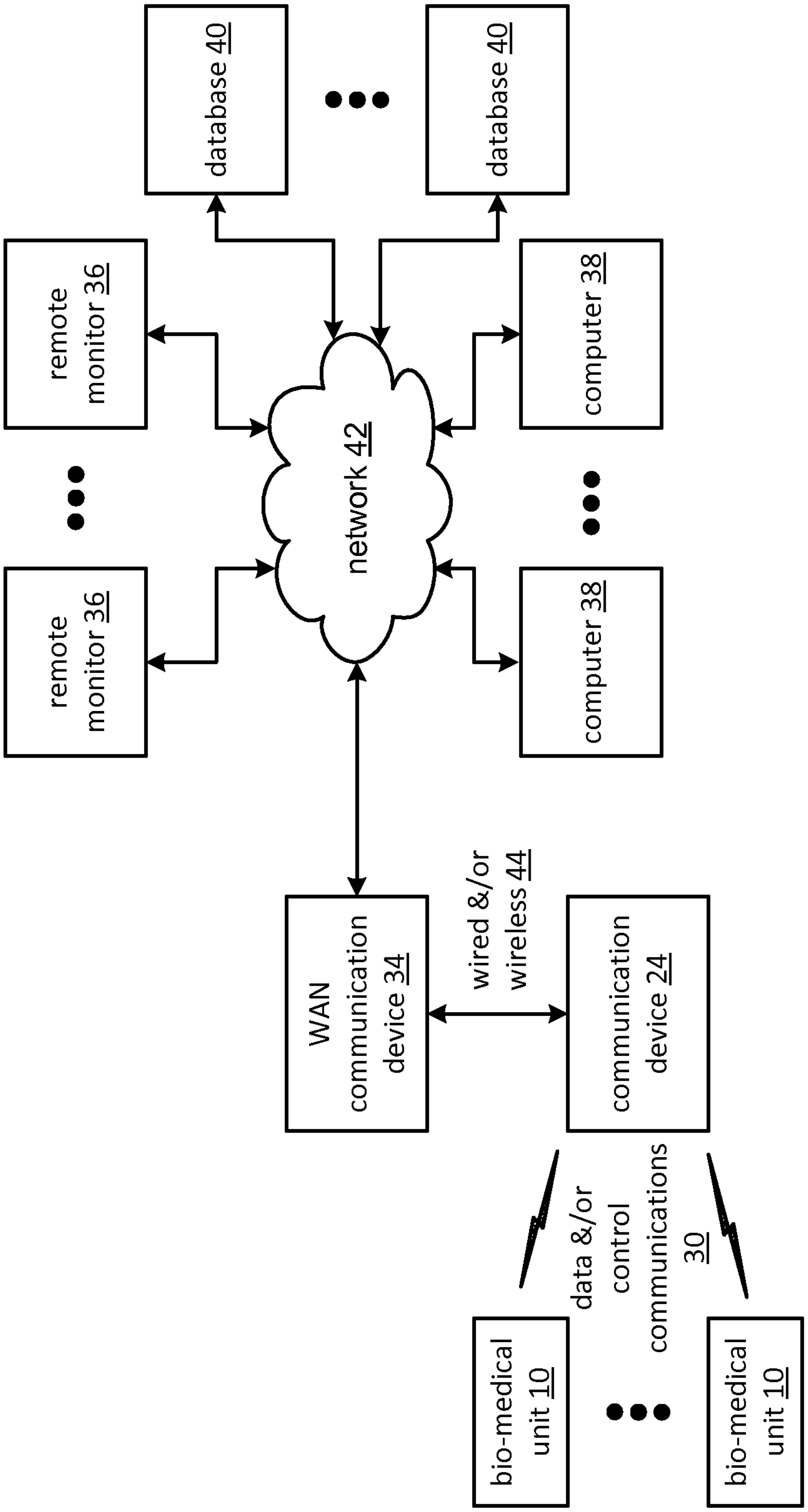
database 40
database 40
remote monitor 36
remote monitor 36
computer 38
computer 38
network 42
WAN communication device 34
wired &/or wireless 44
communication device 24
data &/or control communications 30
bio-medical unit 10
bio-medical unit 10

FIG. 5

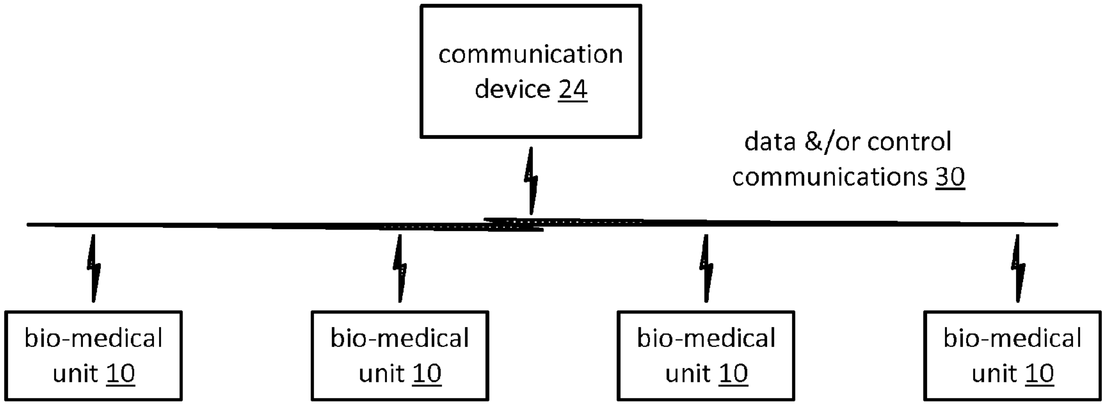
communication device 24
data &/or control communications 30
bio-medical unit 10
bio-medical unit 10
bio-medical unit 10
bio-medical unit 10

FIG. 6

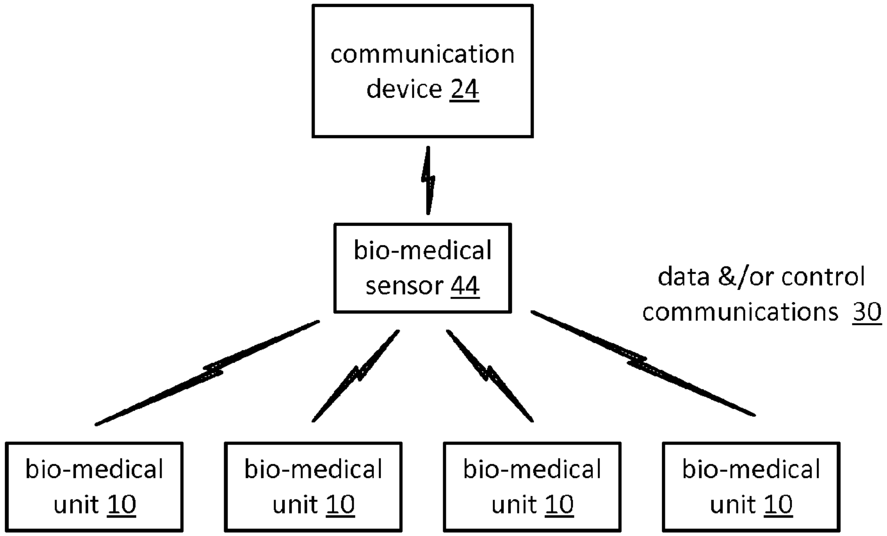
communication device 24
bio-medical sensor 44
data &/or control communications 30
bio-medical unit 10
bio-medical unit 10
bio-medical unit 10
bio-medical unit 10

FIG. 7

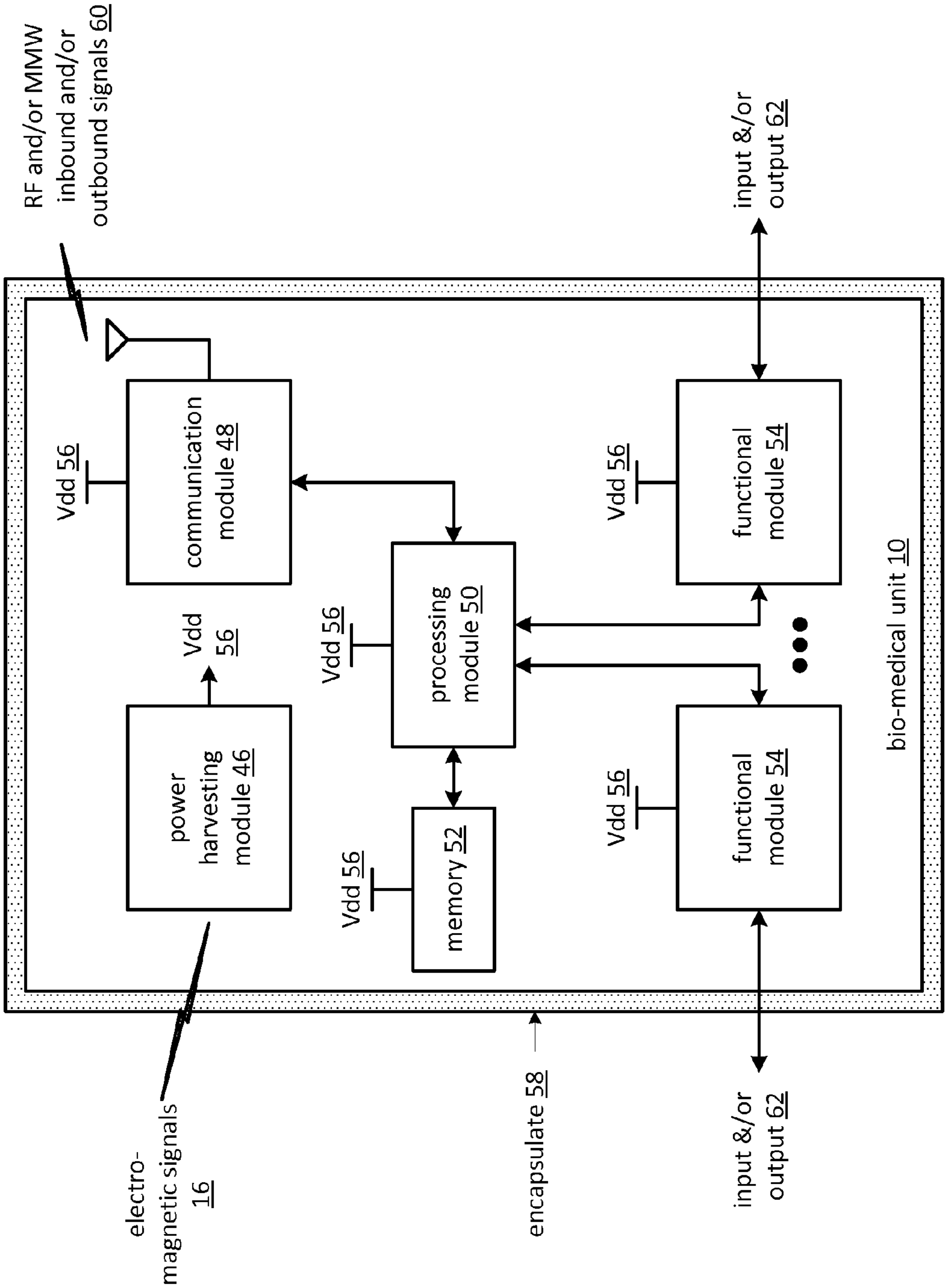
RF and/or MMW inbound and/or outbound signals 60
input &/or output 62
Vdd 56
communication module 48
Vdd 56
power harvesting module 46
Vdd 56
processing module 50
Vdd 56
memory 52
Vdd 56
functional module 54
Vdd 56
functional module 54
bio-medical unit 10
electro-magnetic signals 16
encapsulate 58
input &/or output 62

FIG. 8

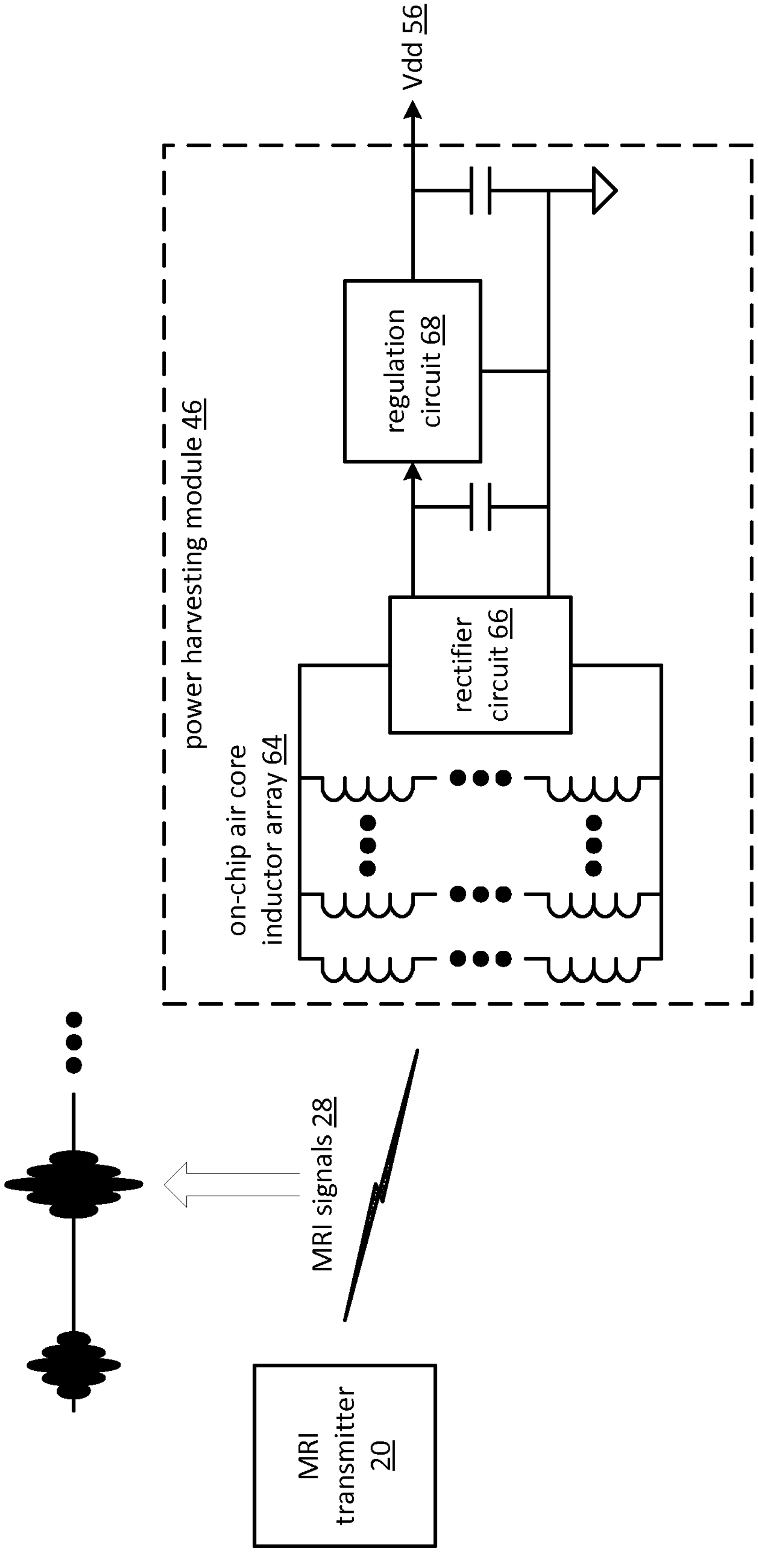
Vdd 56
power harvesting module 46
regulation circuit 68
rectifier circuit 66
on-chip air core inductor array 64
MRI signals 28
MRI transmitter 20

FIG. 9

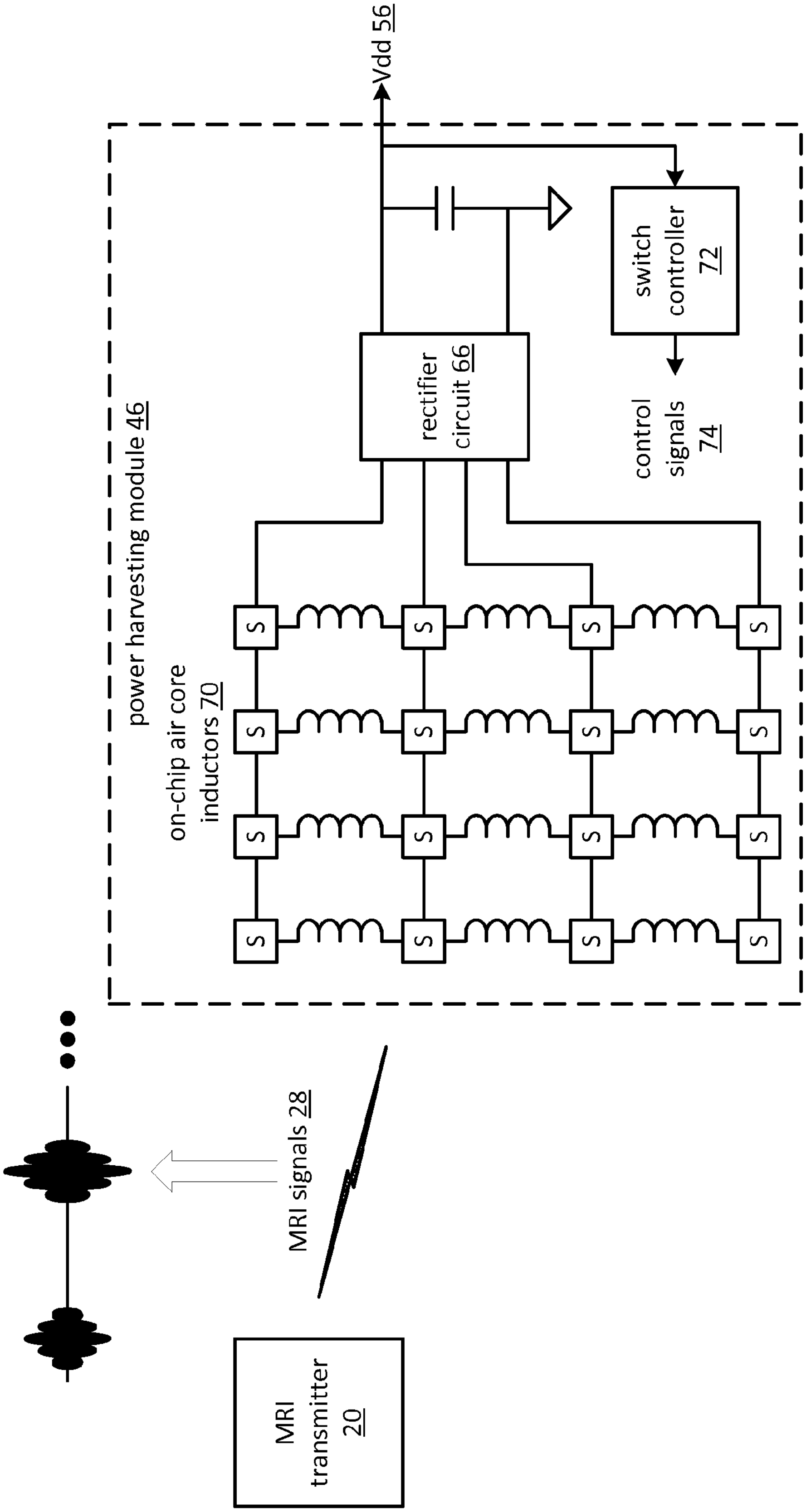
Vdd 56
power harvesting module 46
on-chip air core inductors 70
S
rectifier circuit 66
switch controller 72
control signals 74
MRI signals 28
MRI transmitter 20

FIG. 10

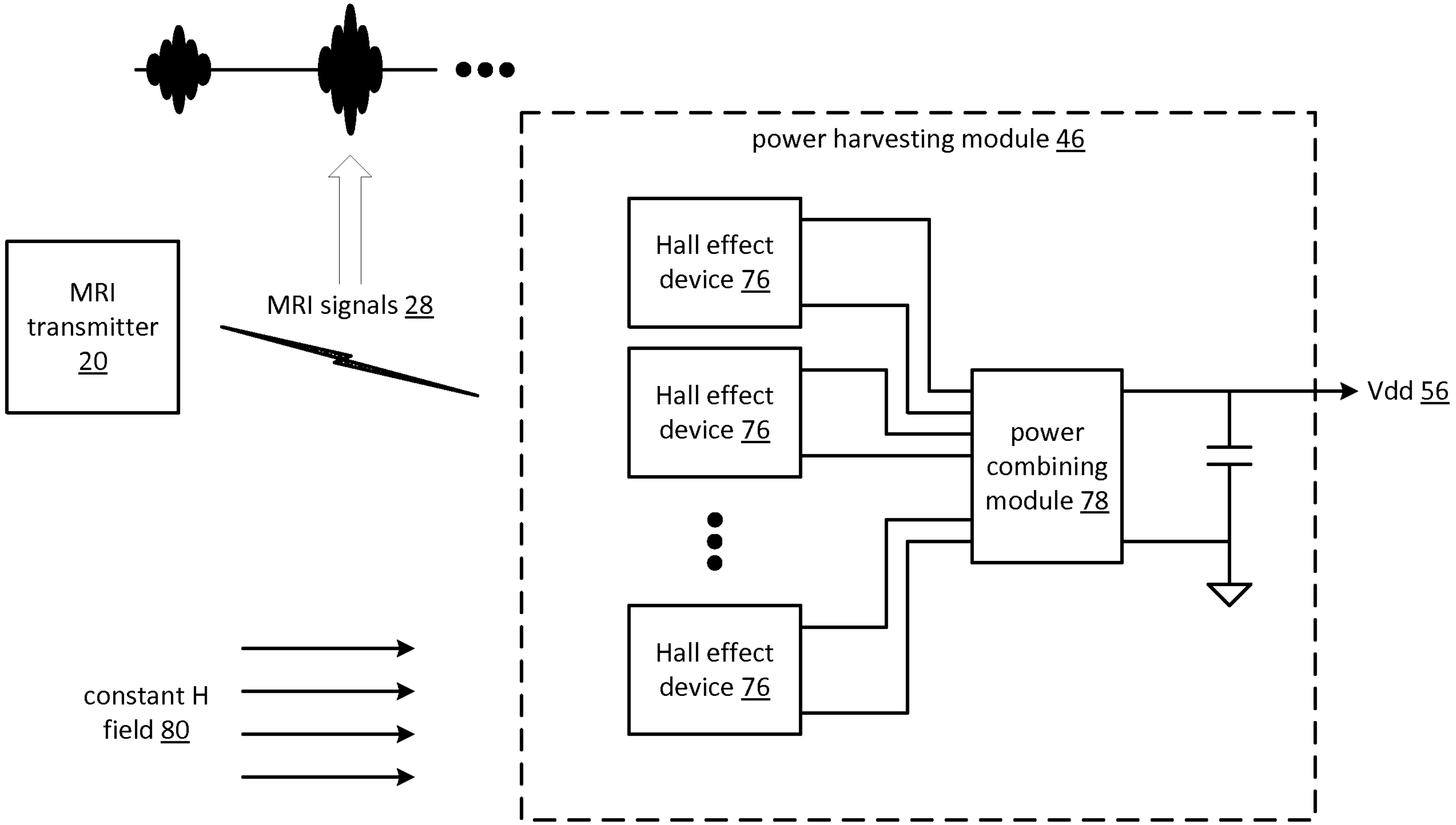
MRI transmitter 20
MRI signals 28
constant H field 80
power harvesting module 46
Hall effect device 76
Hall effect device 76
Hall effect device 76
power combining module 78
Vdd 56

FIG. 11

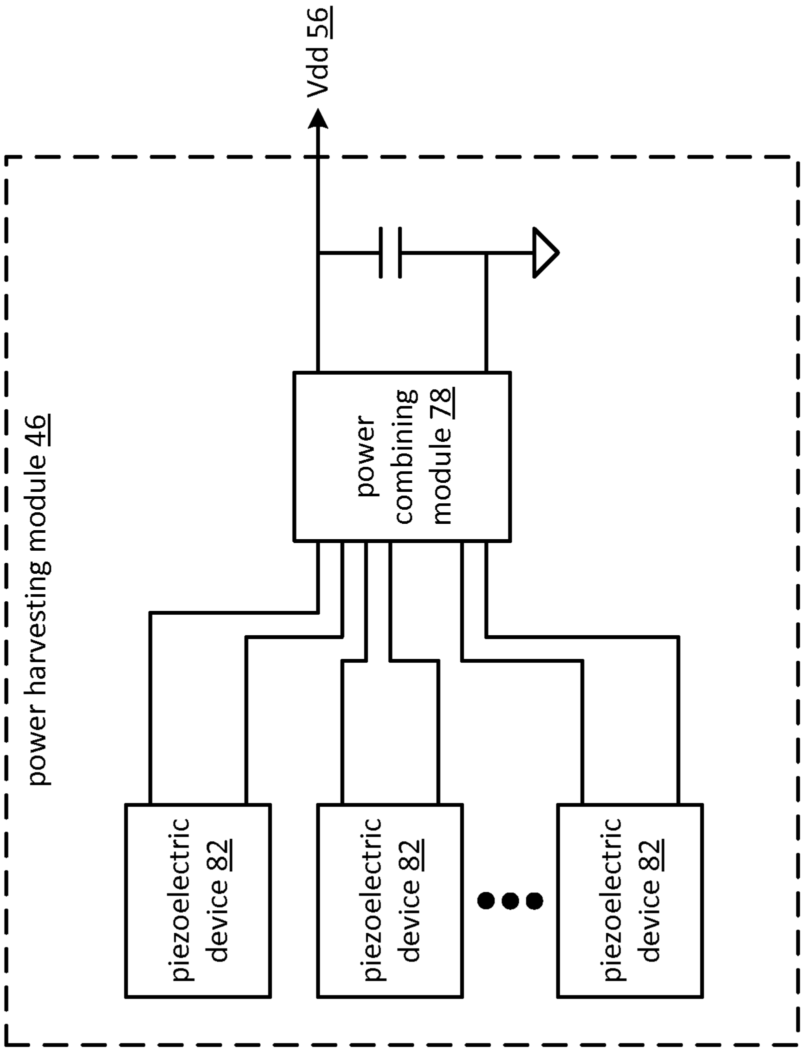
Vdd 56
power harvesting module 46
power combining module 78
piezoelectric device 82
piezoelectric device 82
piezoelectric device 82

FIG. 12

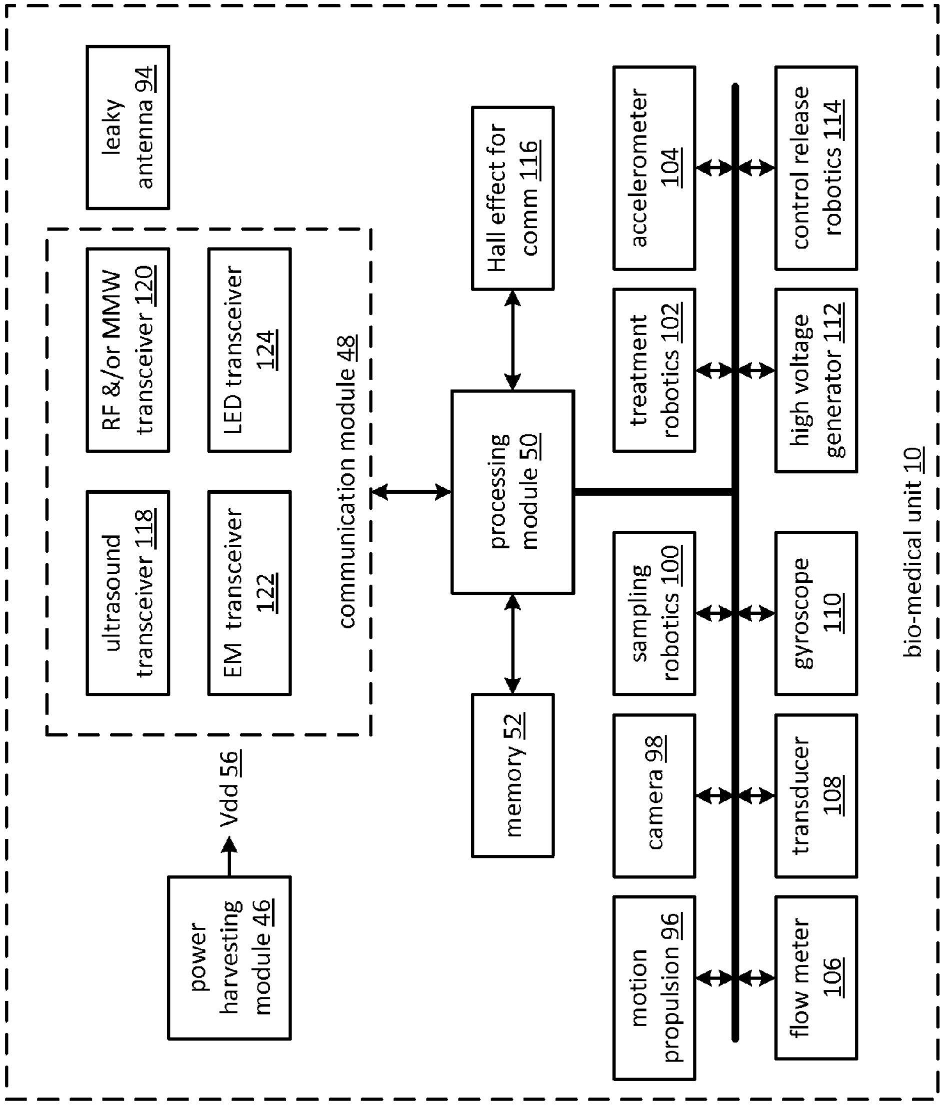
leaky antenna 94
RF &/or MMW transceiver 120
LED transceiver 124
ultrasound transceiver 118
EM transceiver 122
communication module 48
Vdd 56
power harvesting module 46
Hall effect for comm 116
processing module 50
memory 52
accelerometer 104
treatment robotics 102
sampling robotics 100
camera 98
motion propulsion 96
control release robotics 114
high voltage generator 112
gyroscope 110
transducer 108
flow meter 106
bio-medical unit 10

FIG. 13

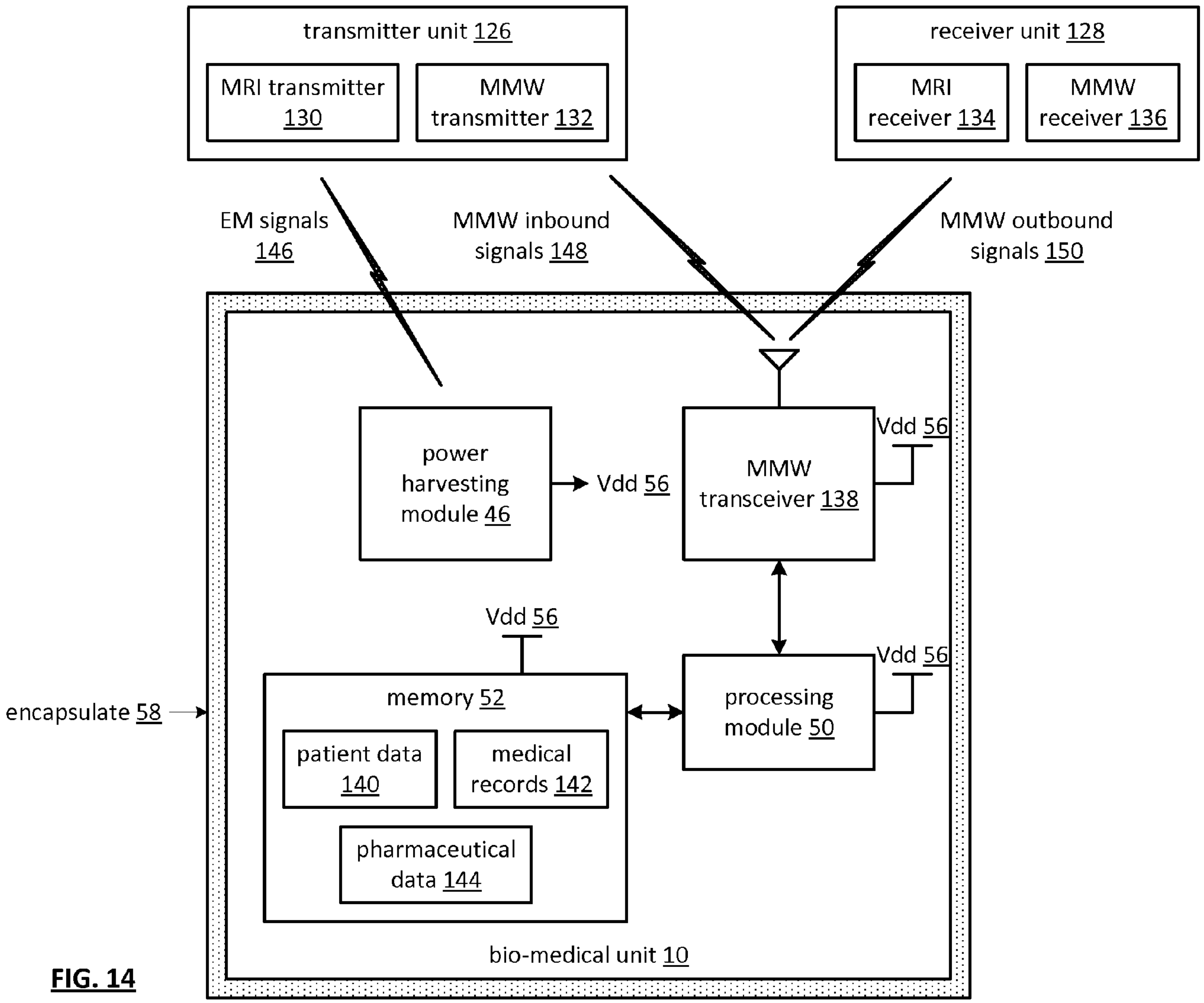
transmitter unit 126
MRI transmitter 130
MMW transmitter 132
receiver unit 128
MRI receiver 134
MMW receiver 136
EM signals 146
MMW inbound signals 148
MMW outbound signals 150
power harvesting module 46
Vdd 56
MMW transceiver 138
Vdd 56
Vdd 56
memory 52
patient data 140
medical records 142
pharmaceutical data 144
processing module 50
Vdd 56
encapsulate 58
bio-medical unit 10

FIG. 14

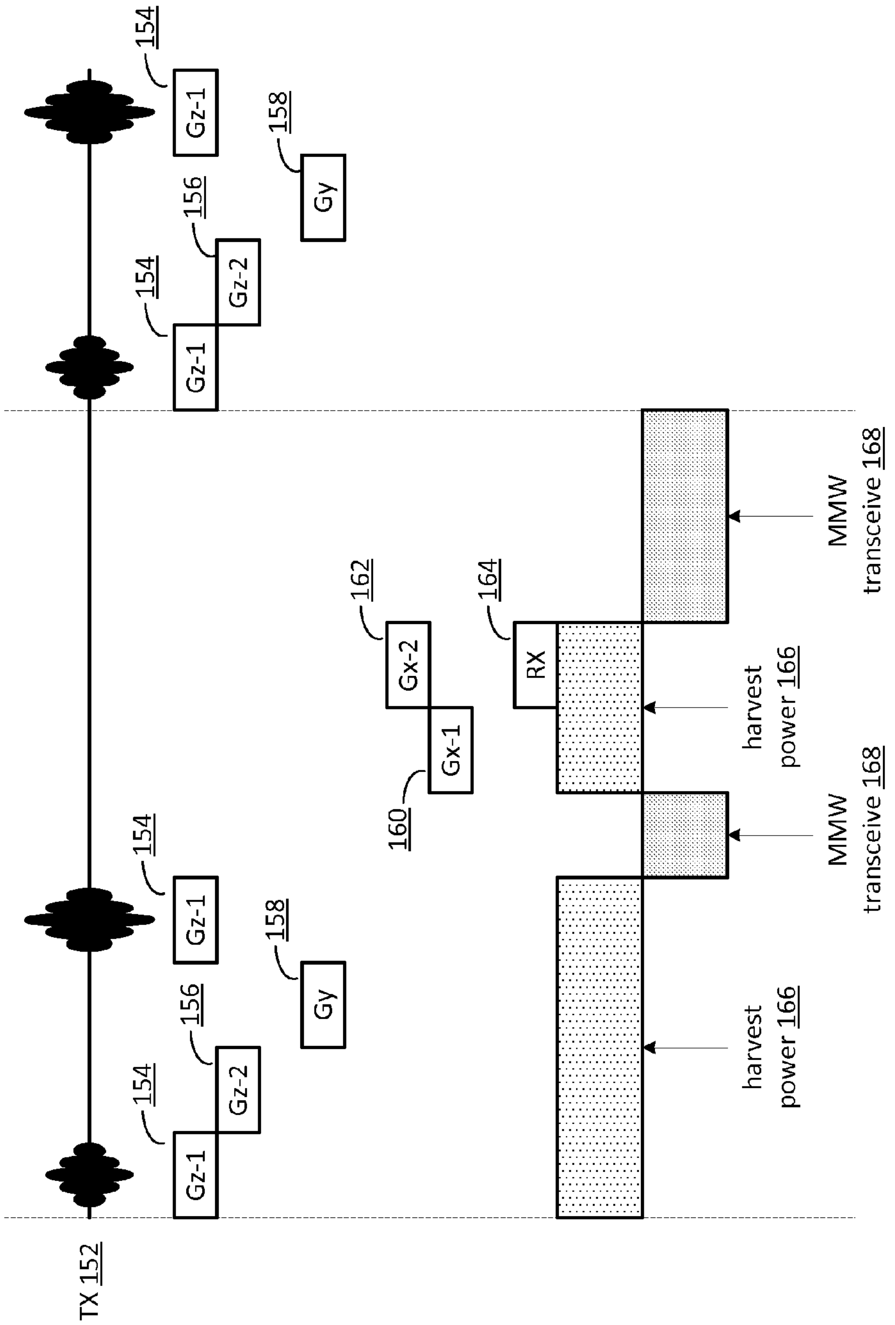
TX 152
Gz-1
154
Gz-2
156
Gy
158
Gz-1
154
Gx-1
160
Gx-2
162
RX
164
harvest power 166
MMW transceive 168
harvest power 166
MMW transceive 168
Gz-1
154
Gz-2
156
Gy
158
Gz-1
154

FIG. 15

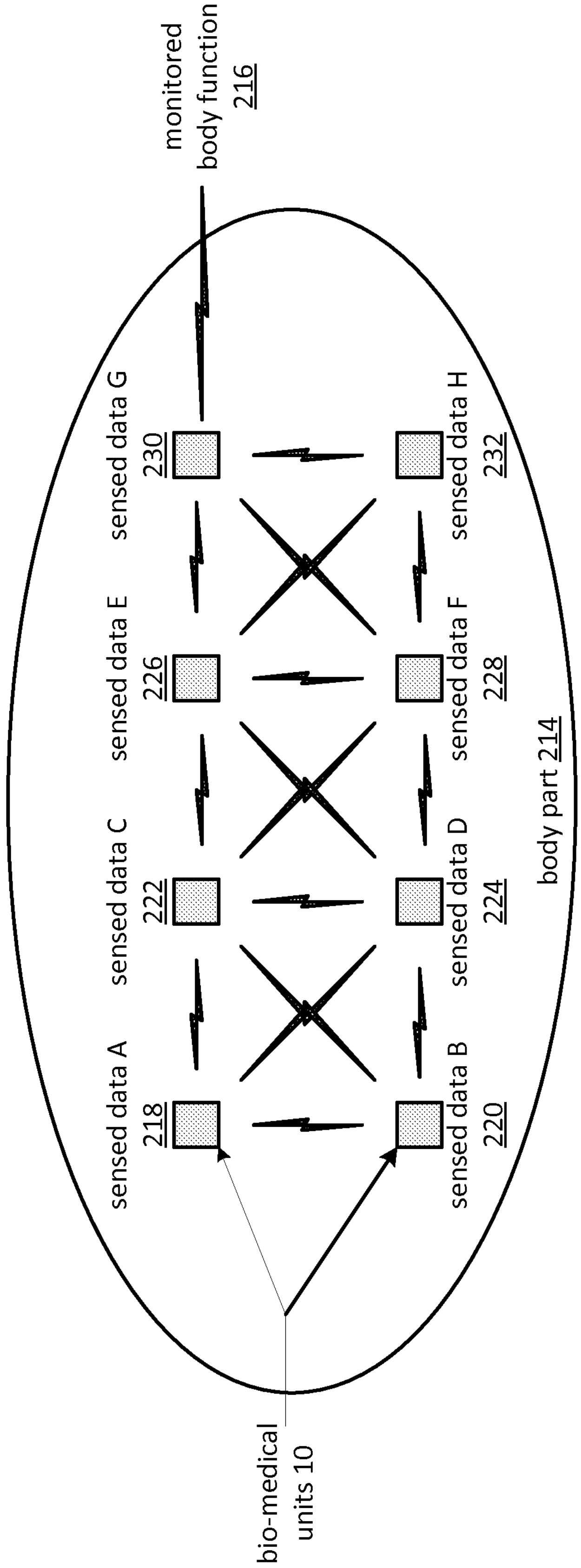
monitored body function 216
sensed data A 218
sensed data B 220
sensed data C 222
sensed data D 224
sensed data E 226
sensed data F 228
sensed data G 230
sensed data H 232
body part 214
bio-medical units 10

FIG. 16

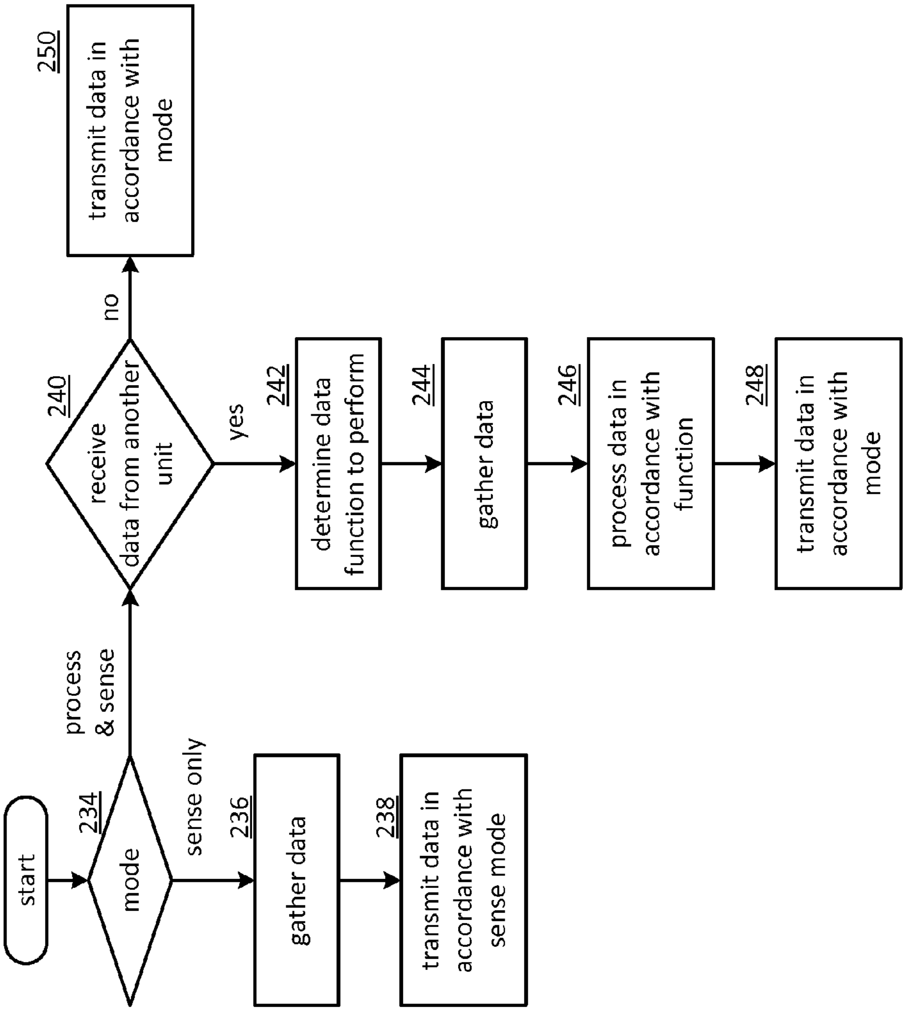
start
234
mode
sense only
236
gather data
238
transmit data in accordance with sense mode
process & sense
240
receive data from another unit
no
250
transmit data in accordance with mode
yes
242
determine data function to perform
244
gather data
246
process data in accordance with function
248
transmit data in accordance with mode

FIG. 17

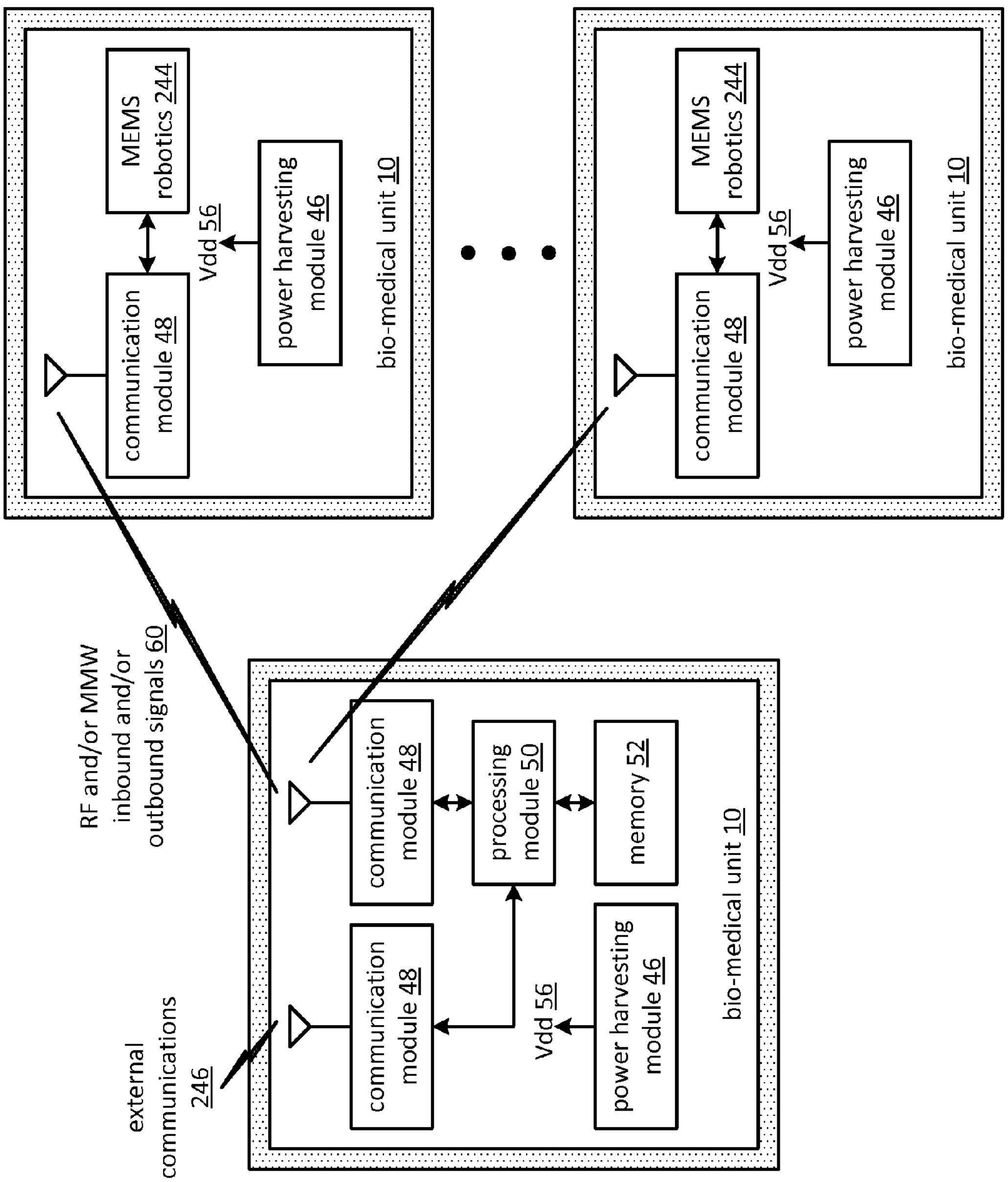
MEMS robotics 244
communication module 48
Vdd 56
power harvesting module 46
bio-medical unit 10
MEMS robotics 244
communication module 48
Vdd 56
power harvesting module 46
bio-medical unit 10
RF and/or MMW inbound and/or outbound signals 60
communication module 48
processing module 50
memory 52
communication module 48
Vdd 56
power harvesting module 46
bio-medical unit 10
external communications 246

FIG. 18

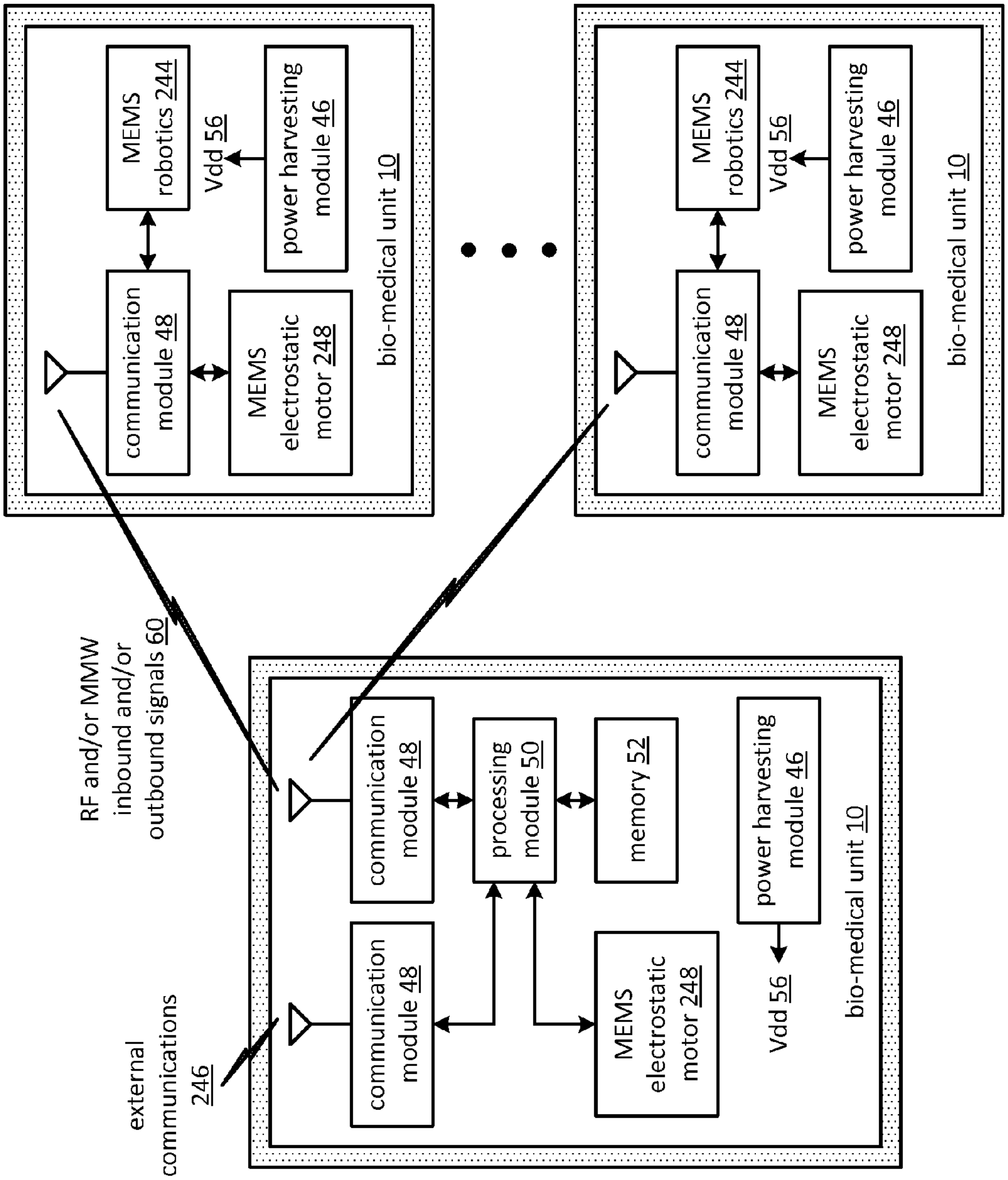
MEMS robotics 244
Vdd 56
power harvesting module 46
communication module 48
MEMS electrostatic motor 248
bio-medical unit 10
MEMS robotics 244
Vdd 56
power harvesting module 46
communication module 48
MEMS electrostatic motor 248
bio-medical unit 10
RF and/or MMW inbound and/or outbound signals 60
communication module 48
processing module 50
memory 52
power harvesting module 46
communication module 48
MEMS electrostatic motor 248
Vdd 56
bio-medical unit 10
external communications 246

FIG. 19

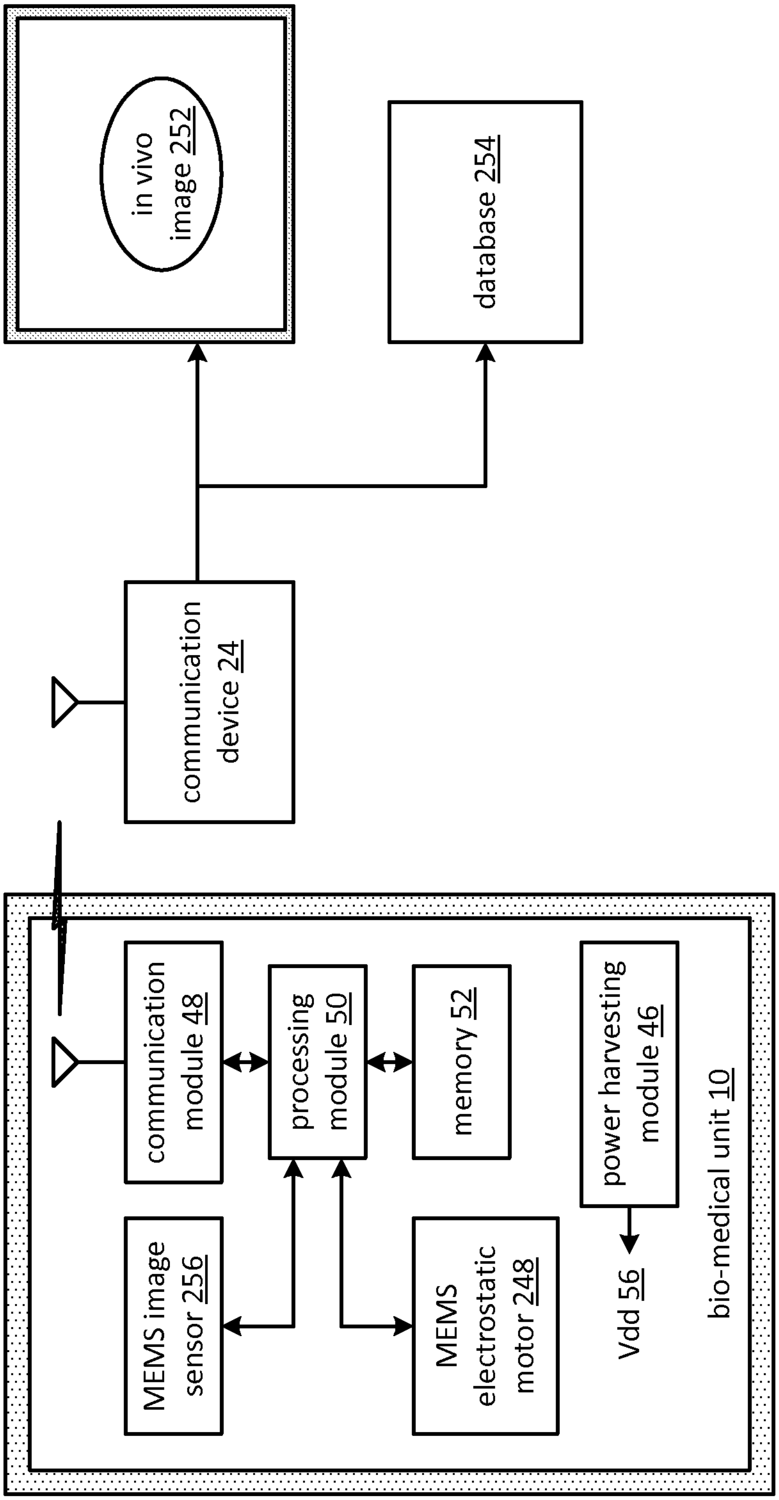
in vivo image 252
database 254
communication device 24
MEMS image sensor 256
communication module 48
processing module 50
memory 52
MEMS electrostatic motor 248
power harvesting module 46
Vdd 56
bio-medical unit 10

FIG. 20

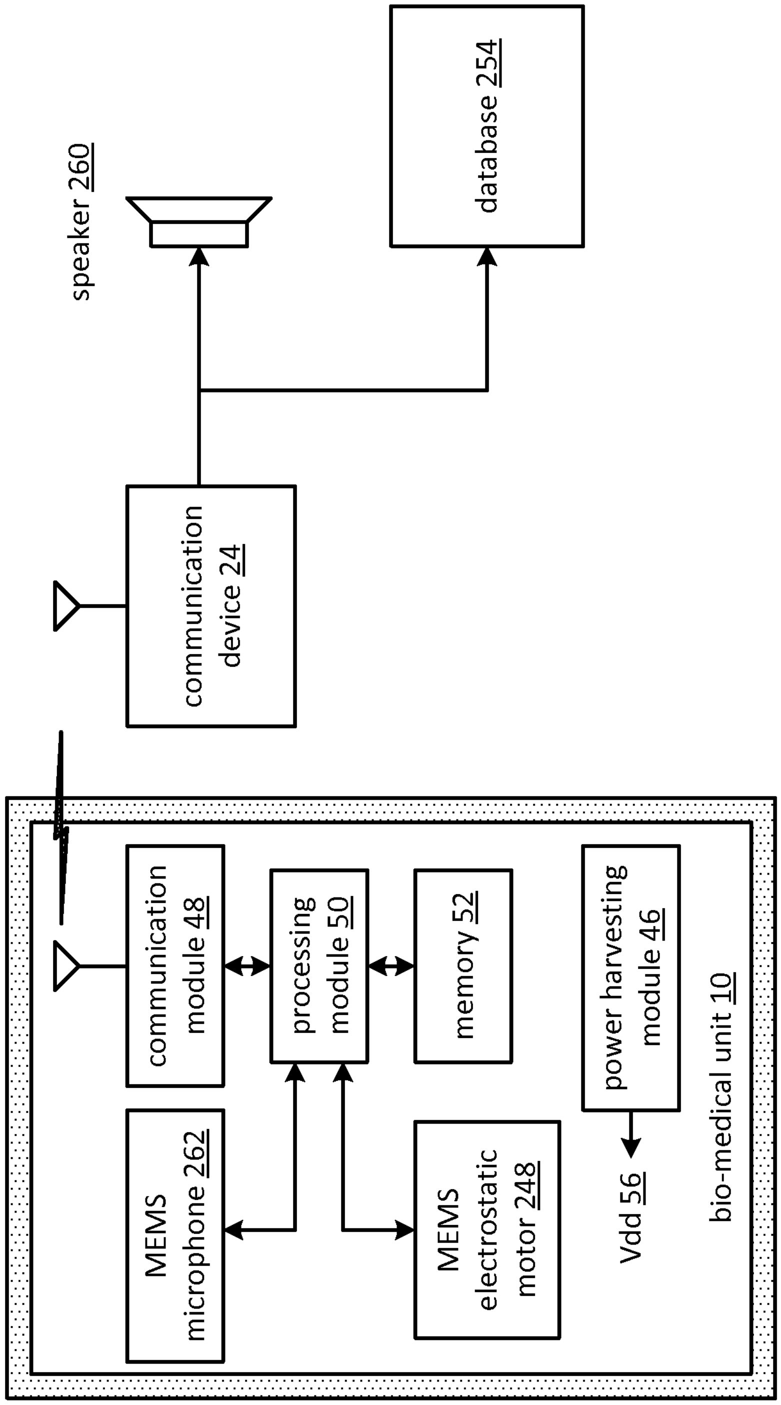
speaker 260
database 254
communication device 24
communication module 48
processing module 50
memory 52
power harvesting module 46
MEMS microphone 262
MEMS electrostatic motor 248
Vdd 56
bio-medical unit 10

FIG. 21

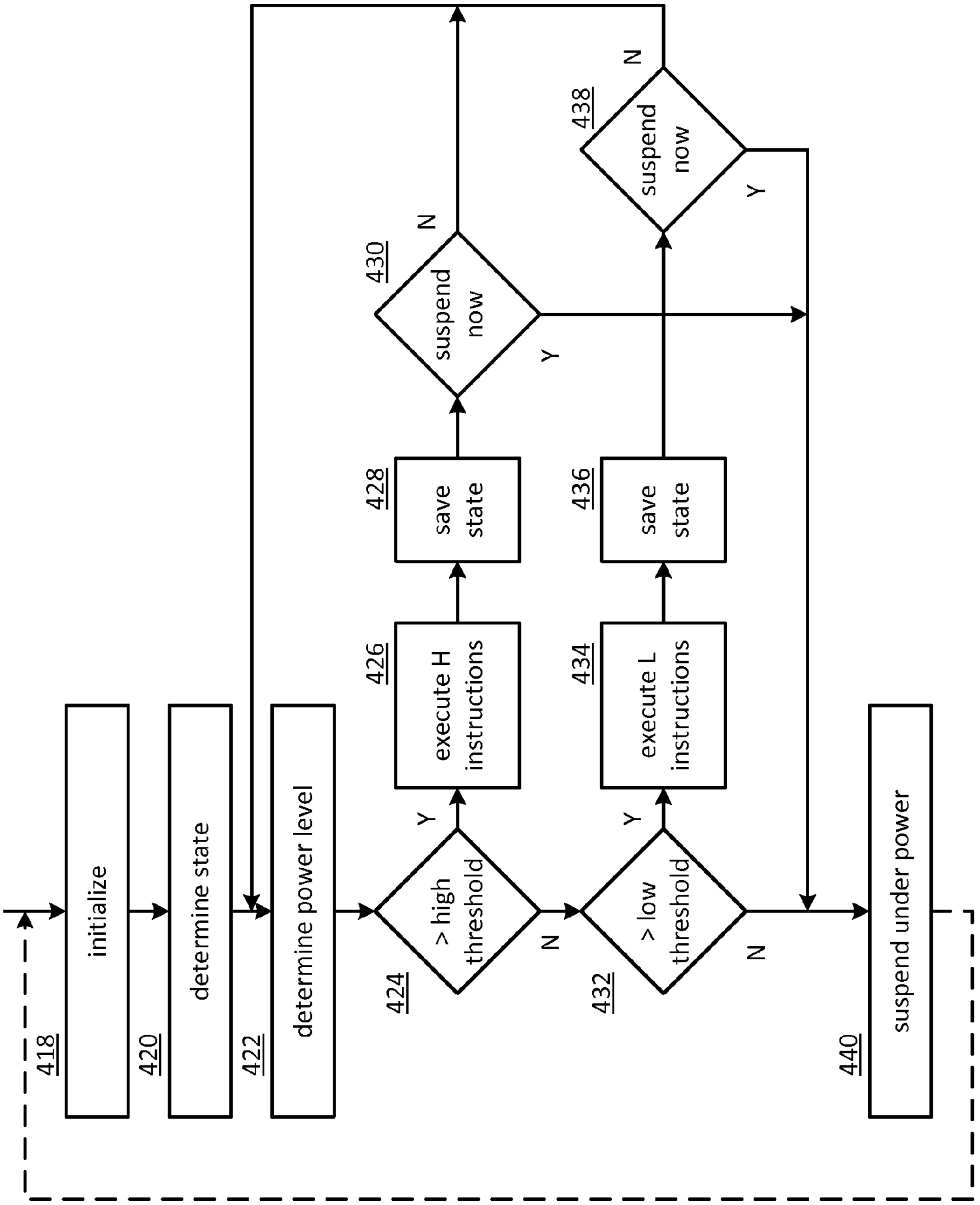
418
initialize
420
determine state
422
determine power level
424
> high threshold
Y
N
426
execute H instructions
428
save state
430
suspend now
Y
N
432
> low threshold
Y
N
434
execute L instructions
436
save state
438
suspend now
Y
N
440
suspend under power

FIG. 22

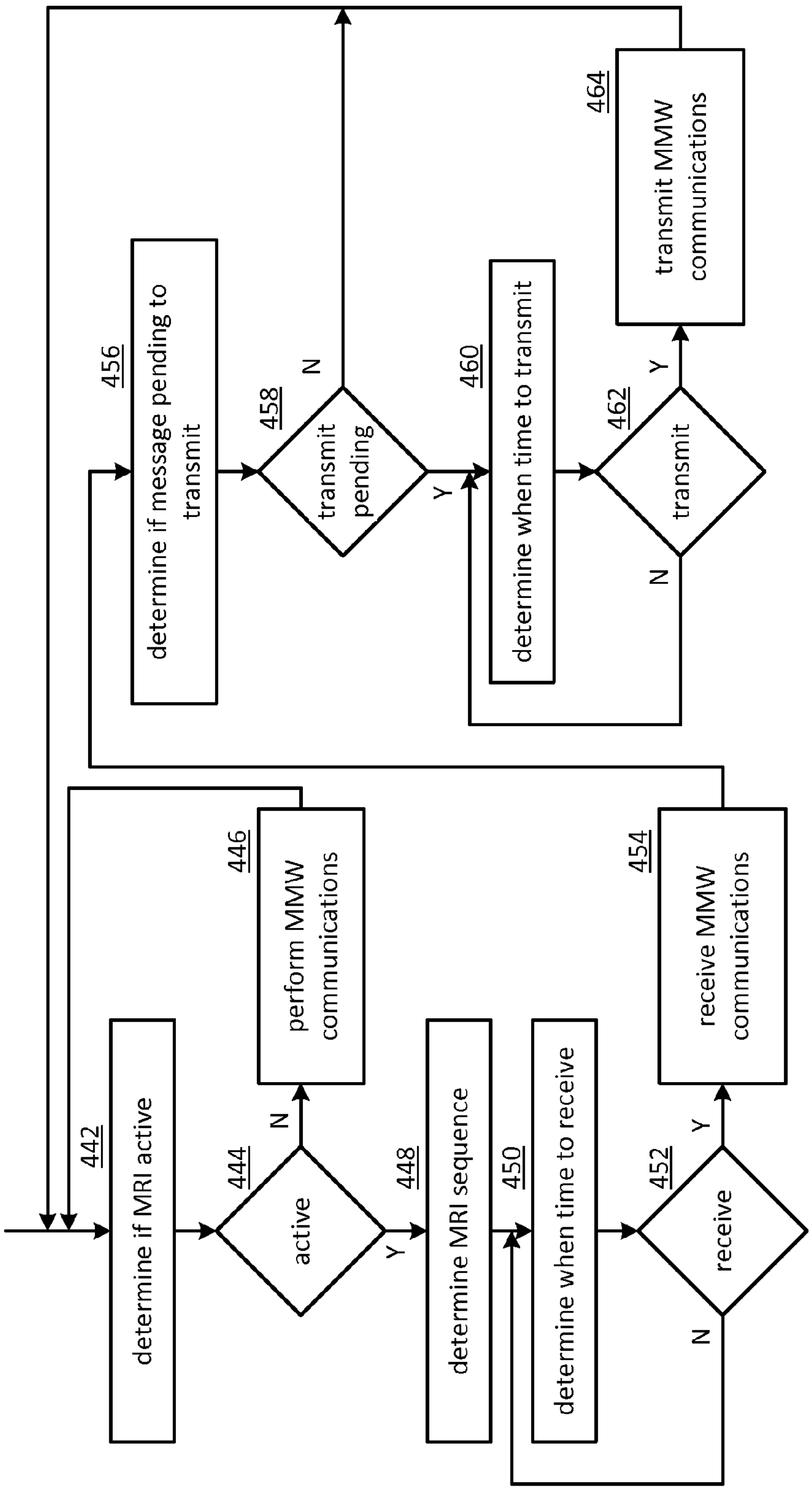
442
determine if MRI active
444
active
N
Y
446
perform MMW communications
448
determine MRI sequence
450
determine when time to receive
452
receive
Y
N
454
receive MMW communications
456
determine if message pending to transmit
458
transmit pending
N
Y
460
determine when time to transmit
462
transmit
Y
N
464
transmit MMW communications

FIG. 23

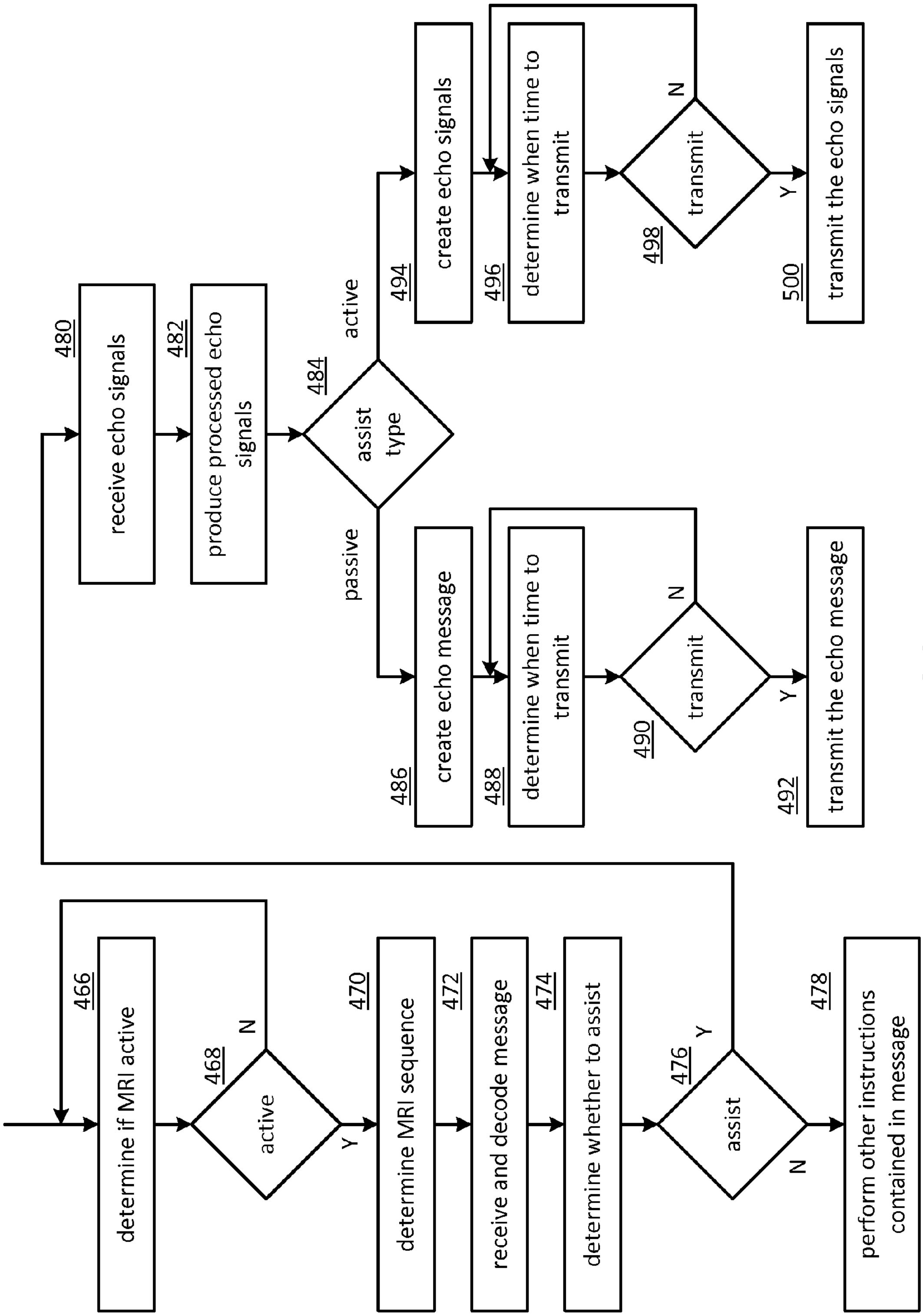
466
determine if MRI active
468
active
N
Y
470
determine MRI sequence
472
receive and decode message
474
determine whether to assist
476
assist
Y
N
478
perform other instructions contained in message
480
receive echo signals
482
produce processed echo signals
484
assist type
passive
active
486
create echo message
488
determine when time to transmit
490
transmit
N
Y
492
transmit the echo message
494
create echo signals
496
determine when time to transmit
498
transmit
N
Y
500
transmit the echo signals

FIG. 24

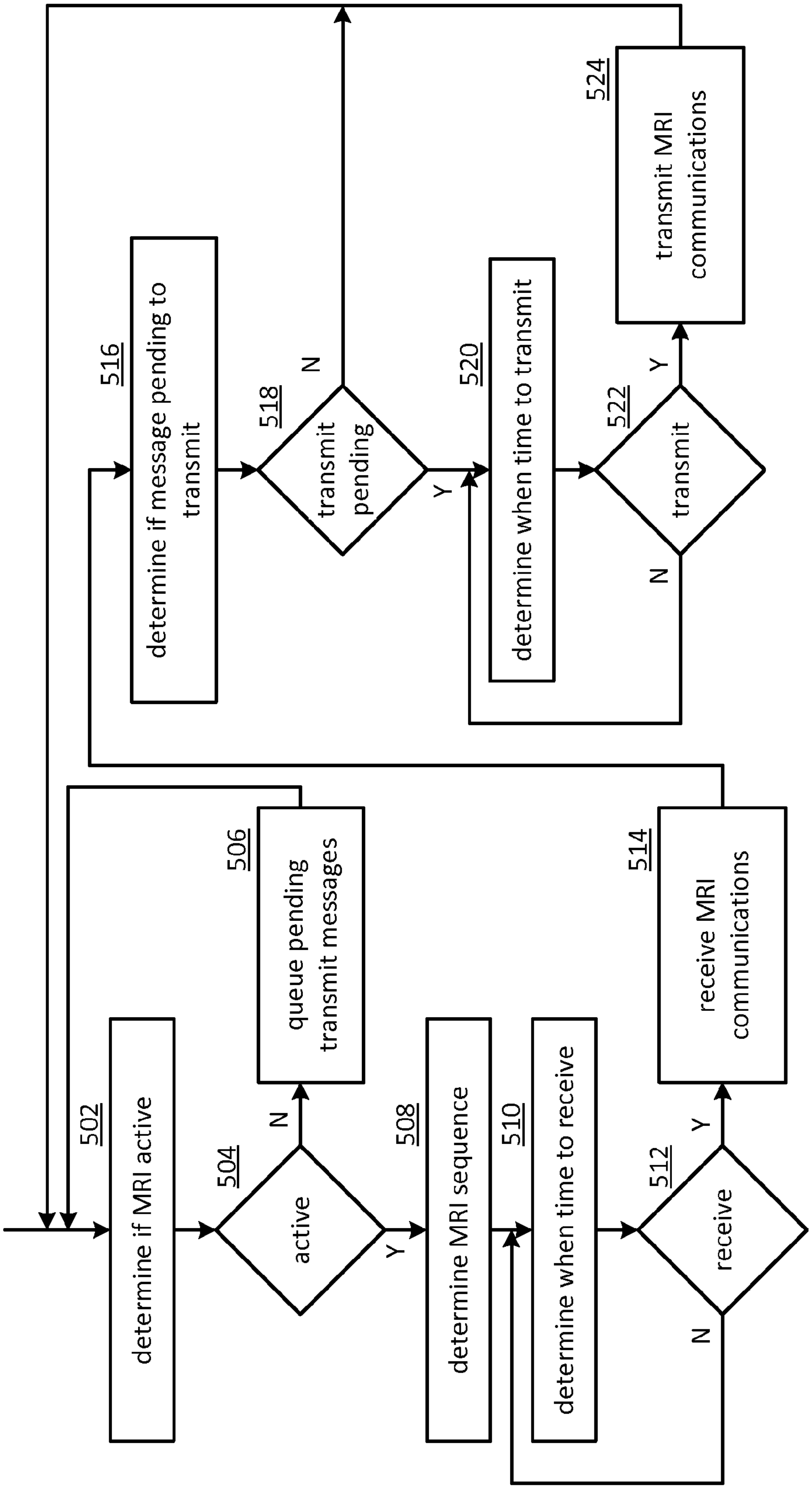
502
determine if MRI active
504
active
Y
N
506
queue pending transmit messages
508
determine MRI sequence
510
determine when time to receive
512
receive
Y
N
514
receive MRI communications
516
determine if message pending to transmit
518
transmit pending
N
Y
520
determine when time to transmit
522
transmit
Y
N
524
transmit MRI communications

FIG. 25

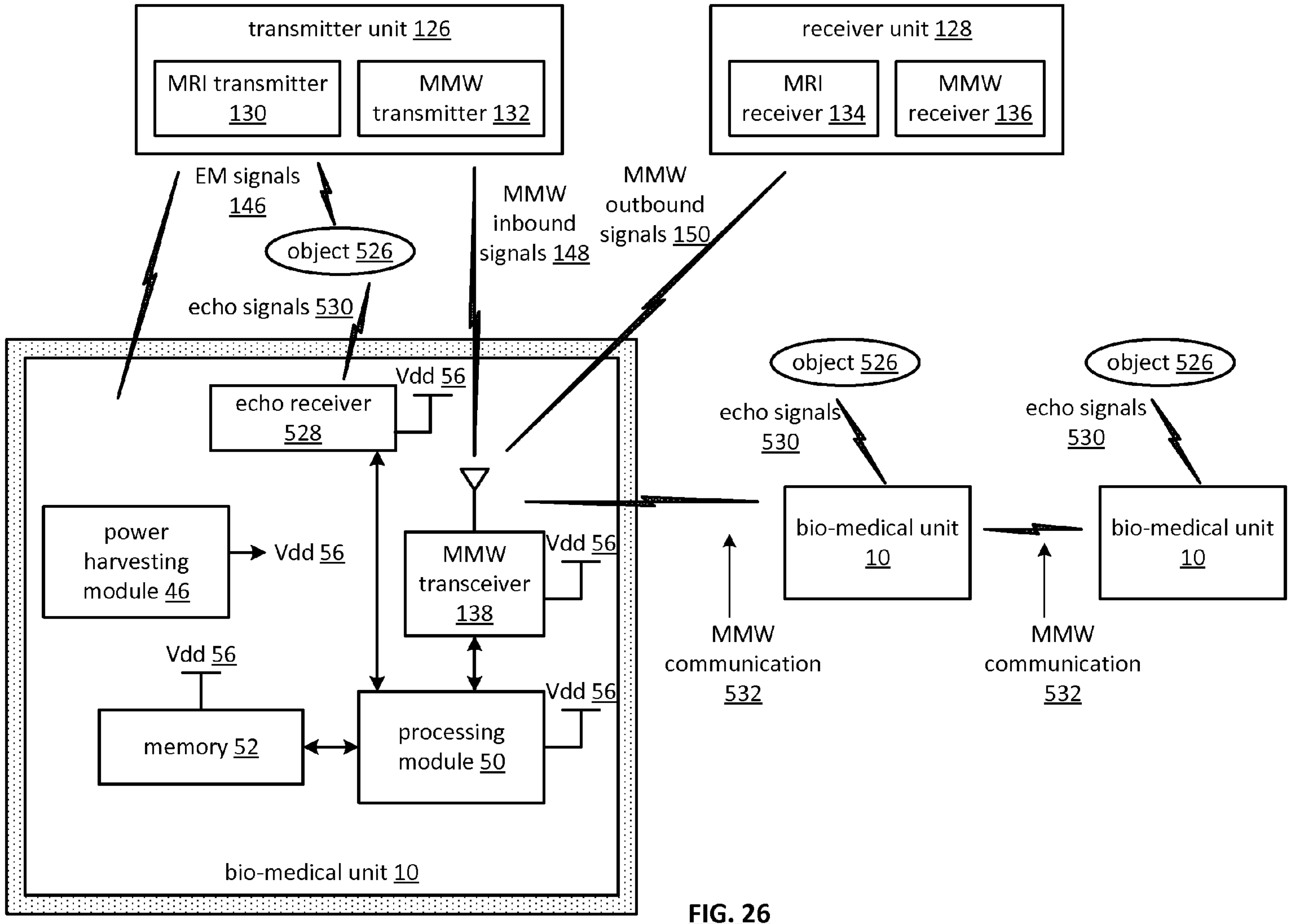
transmitter unit 126
MRI transmitter 130
MMW transmitter 132
receiver unit 128
MRI receiver 134
MMW receiver 136
EM signals 146
object 526
echo signals 530
MMW inbound signals 148
MMW outbound signals 150
echo receiver 528
Vdd 56
power harvesting module 46
Vdd 56
MMW transceiver 138
Vdd 56
Vdd 56
memory 52
processing module 50
Vdd 56
bio-medical unit 10
object 526
echo signals 530
bio-medical unit 10
MMW communication 532
object 526
echo signals 530
bio-medical unit 10
MMW communication 532

FIG. 26

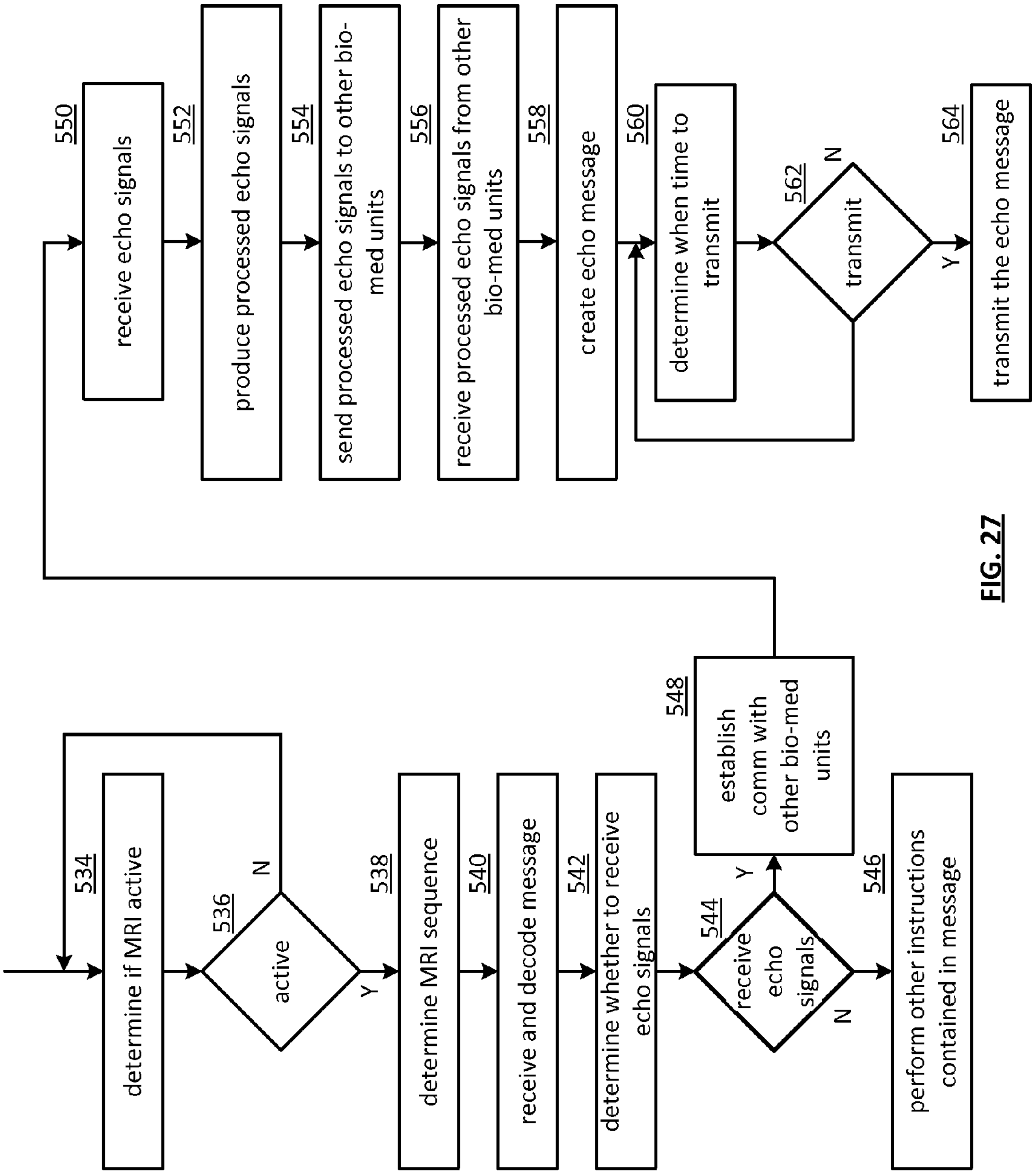
534
determine if MRI active
536
active
N
Y
538
determine MRI sequence
540
receive and decode message
542
determine whether to receive echo signals
544
receive echo signals
Y
N
548
establish comm with other bio-med units
546
perform other instructions contained in message
550
receive echo signals
552
produce processed echo signals
554
send processed echo signals to other bio-med units
556
receive processed echo signals from other bio-med units
558
create echo message
560
determine when time to transmit
562
transmit
N
Y
564
transmit the echo message

FIG. 27

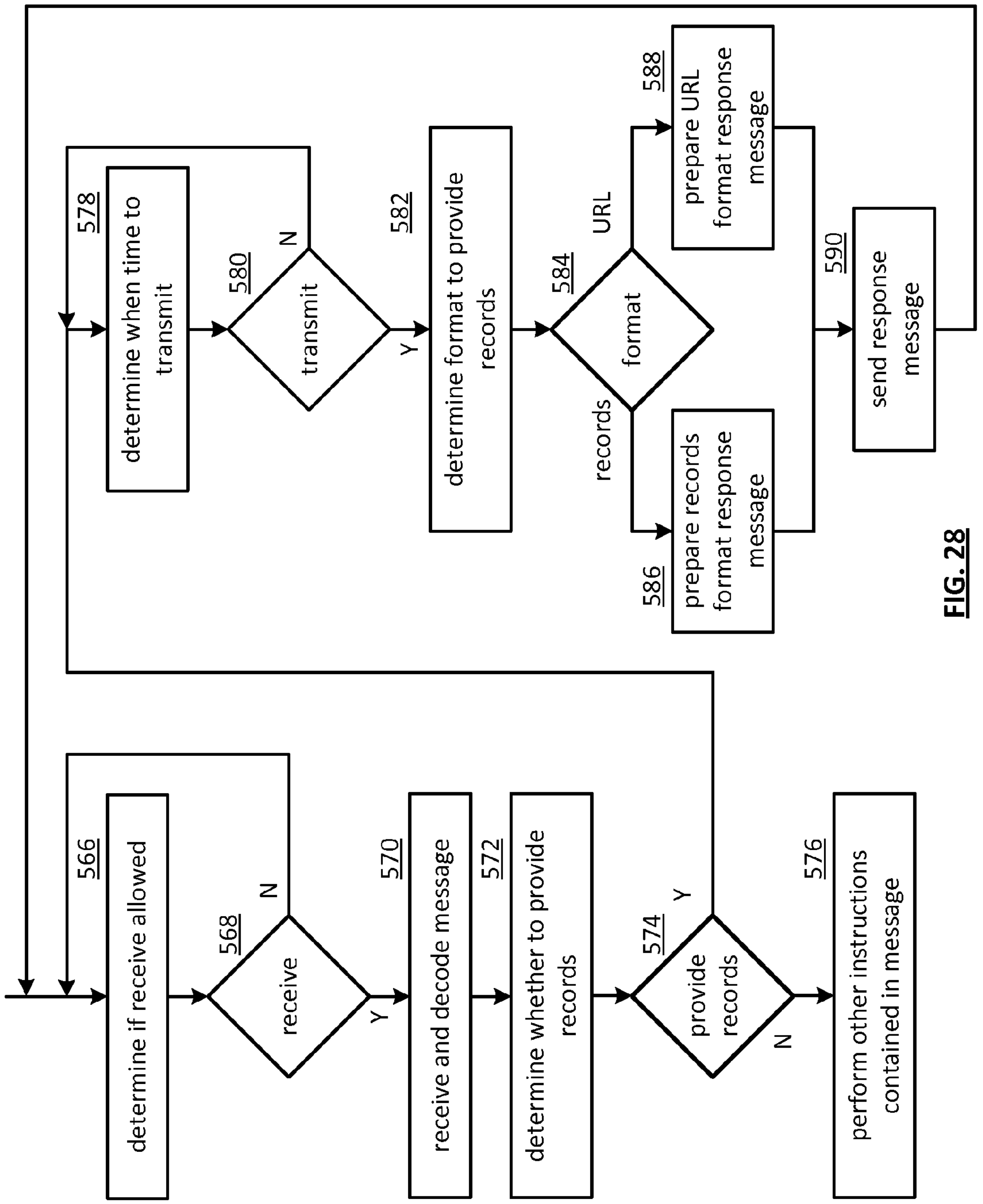
566
determine if receive allowed
568
receive
N
Y
570
receive and decode message
572
determine whether to provide records
574
provide records
Y
N
576
perform other instructions contained in message
578
determine when time to transmit
580
transmit
N
Y
582
determine format to provide records
584
format
URL
records
586
prepare records format response message
588
prepare URL format response message
590
send response message

FIG. 28

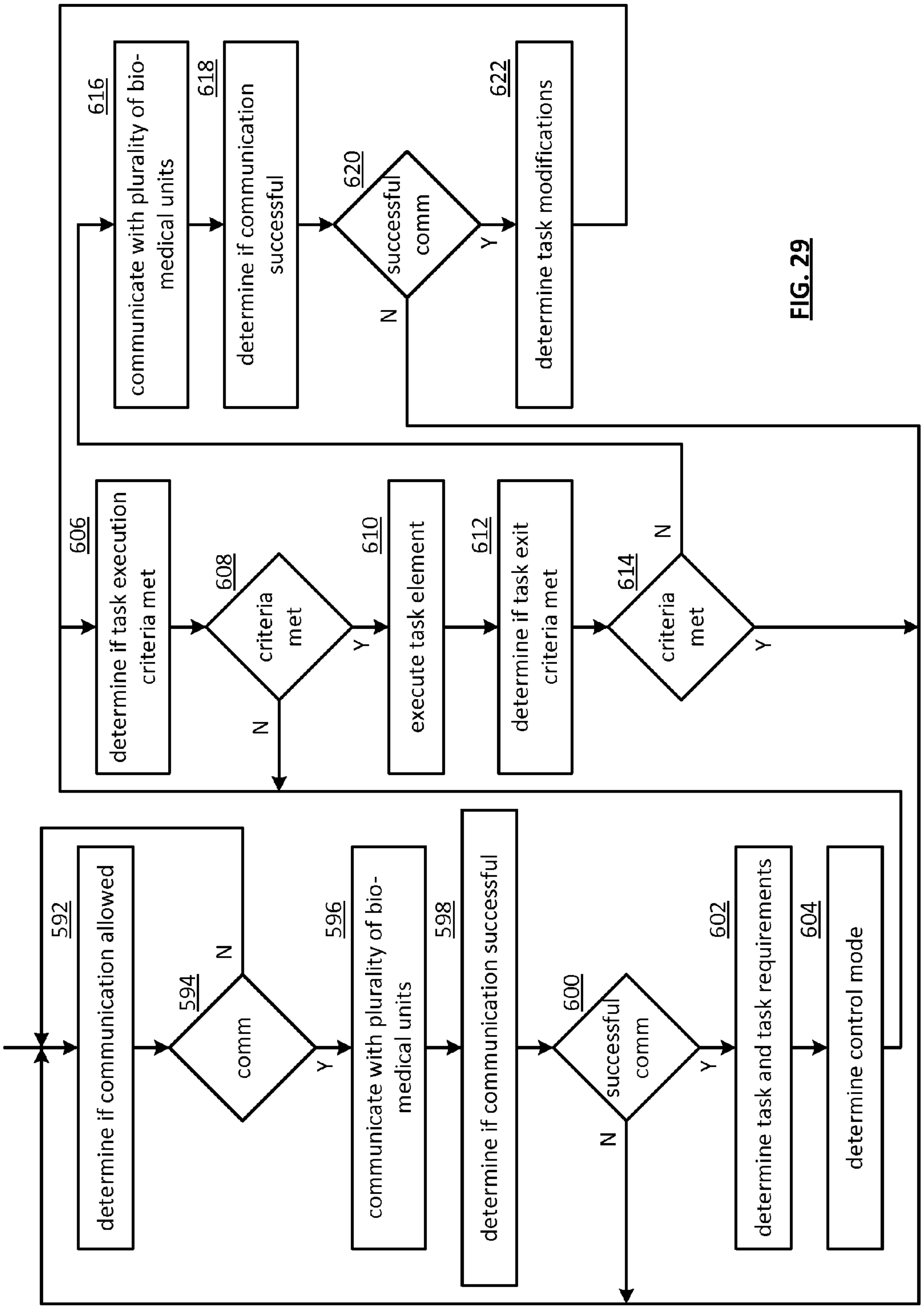

592
determine if communication allowed
594
comm
N
Y
596
communicate with plurality of bio-medical units
598
determine if communication successful
600
successful comm
N
Y
602
determine task and task requirements
604
determine control mode
606
determine if task execution criteria met
608
criteria met
N
Y
610
execute task element
612
determine if task exit criteria met
614
criteria met
N
Y
616
communicate with plurality of bio-medical units
618
determine if communication successful
620
successful comm
N
Y
622
determine task modifications

FIG. 29

BIO-MEDICAL UNIT WITH POWER HARVESTING MODULE AND RF COMMUNICATION

This patent application is claiming priority under 35 USC §119 to a provisionally filed patent application entitled BIO-MEDICAL UNIT AND APPLICATIONS THEREOF, having a provisional filing date of Sep. 30, 2009, and a provisional Ser. No. 61/247,060.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

Not Applicable

INCORPORATION-BY-REFERENCE OF MATERIAL SUBMITTED ON A COMPACT DISC

Not Applicable

BACKGROUND OF THE INVENTION

1. Technical Field of the Invention

This invention relates generally to medical equipment and more particularly to wireless medical equipment.

2. Description of Related Art

As is known, there is a wide variety of medical equipment that aids in the diagnosis, monitoring, and/or treatment of patients' medical conditions. For instances, there are diagnostic medical devices, therapeutic medical devices, life support medical devices, medical monitoring devices, medical laboratory equipment, etc. As specific exampled magnetic resonance imaging (MRI) devices produce images that illustrate the internal structure and function of a body.

The advancement of medical equipment is in step with the advancements of other technologies (e.g., radio frequency identification (RFID), robotics, etc.). Recently, RFID technology has been used for in vitro use to store patient information for easy access. While such in vitro applications have begun, the technical advancement in this area is in its infancy.

BRIEF SUMMARY OF THE INVENTION

The present invention is directed to apparatus and methods of operation that are further described in the following Brief Description of the Drawings, the Detailed Description of the Invention, and the claims. Other features and advantages of the present invention will become apparent from the following detailed description of the invention made with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING(S)

FIG. **1** is a diagram of an embodiment of a system in accordance with the present invention;

FIG. **2** is a diagram of another embodiment of a system in accordance with the present invention;

FIG. **3** is a diagram of an embodiment of an artificial body part including one or more bio-medical units in accordance with the present invention;

FIG. **4** is a schematic block diagram of an embodiment of an artificial body part in accordance with the present invention;

FIG. **5** is a diagram of another embodiment of a system in accordance with the present invention;

FIG. **6** is a diagram of another embodiment of a system in accordance with the present invention;

FIG. **7** is a diagram of another embodiment of a system in accordance with the present invention;

FIG. **8** is a schematic block diagram of an embodiment of a bio-medical unit in accordance with the present invention;

FIG. **9** is a schematic block diagram of an embodiment of a power harvesting module in accordance with the present invention;

FIG. **10** is a schematic block diagram of another embodiment of a power harvesting module in accordance with the present invention;

FIG. **11** is a schematic block diagram of another embodiment of a power harvesting module in accordance with the present invention;

FIG. **12** is a schematic block diagram of another embodiment of a power harvesting module in accordance with the present invention;

FIG. **13** is a schematic block diagram of another embodiment of a bio-medical unit in accordance with the present invention;

FIG. **14** is a diagram of another embodiment of a system in accordance with the present invention;

FIG. **15** is a diagram of an example of a communication protocol within a system in accordance with the present invention;

FIG. **16** is a diagram of an embodiment of a network of bio-medical units in accordance with the present invention;

FIG. **17** is a logic diagram of an embodiment of a method for bio-medical unit communications in accordance with the present invention;

FIG. **18** is a diagram of an embodiment of a network of bio-medical units that include MEMS robotics in accordance with the present invention;

FIG. **19** is a diagram of another embodiment of a network of bio-medical units that include MEMS robotics in accordance with the present invention;

FIG. **20** is a diagram of an embodiment of a bio-medical unit collecting image data in accordance with the present invention;

FIG. **21** is a diagram of an embodiment of a bio-medical unit collecting audio and/or ultrasound data in accordance with the present invention;

FIG. **22** is a logic diagram of an embodiment of a method for bio-medical unit communications in accordance with the invention;

FIG. **23** is a logic diagram of an embodiment of a method for MMW communications within a MRI sequence in accordance with the invention;

FIG. **24** is a logic diagram of an embodiment of a method for processing of MRI signals in accordance with the present invention;

FIG. **25** is a logic diagram of an embodiment of a method for communication utilizing MRI signals in accordance with the present invention;

FIG. **26** is a schematic block diagram of another embodiment of a bio-medical unit in accordance with the present invention;

FIG. **27** is a logic diagram of another embodiment of a method for bio-medical unit communications in accordance with the invention;

FIG. **28** is a logic diagram of another embodiment of a method for bio-medical unit communications in accordance with the invention; and

FIG. **29** is a logic diagram of an embodiment of a method for coordination of bio-medical unit task execution in accordance with the present invention.

DETAILED DESCRIPTION OF THE INVENTION

FIG. **1** is a diagram of an embodiment of a system that includes a plurality of bio-medical units **10** embedded within a body and/or placed on the surface of the body to facilitate diagnosis, treatment, and/or data collections. Each of the bio-medical units **10** is a passive device (e.g., it does not include a power source (e.g., a battery)) and, as such, includes a power harvesting module. The bio-medical units **10** may also include one or more of memory, a processing module, and functional modules. Alternatively, or in addition to, each of the bio-medical units **10** may include a power source.

In operation, a transmitter emits **12** electromagnetic signals **16** that pass through the body and are received by a receiver **14**. The transmitter **12** and receiver **14** may be part of a piece of medical diagnostic equipment (e.g., magnetic resonance imaging (MRI), X-ray, etc.) or independent components for stimulating and communicating with the network of bio-medical units in and/or on a body. One or more of the bio-medical units **10** receives the transmitted electromagnetic signals **16** and generates a supply voltage therefrom. Examples of this will be described in greater detail with reference to FIGS. **8-12**.

Embedded within the electromagnetic signals **16** (e.g., radio frequency (RF) signals, millimeter wave (MMW) signals, MRI signals, etc.) or via separate signals, the transmitter **12** communicates with one or more of the bio-medical units **10**. For example, the electromagnetic signals **16** may have a frequency in the range of a few MHz to 900 MHz and the communication with the bio-medical units **10** is modulated on the electromagnetic signals **16** at a much higher frequency (e.g., 5 GHz to 300 GHz). As another example, the communication with the bio-medical units **10** may occur during gaps (e.g., per protocol of medical equipment or injected for communication) of transmitting the electromagnetic signals **16**. As another example, the communication with the bio-medical units **10** occurs in a different frequency band and/or using a different transmission medium (e.g., use RF or MMW signals when the magnetic field of the electromagnetic signals are dominate, use ultrasound signals when the electromagnetic signals **16** are RF and/or MMW signals, etc.).

One or more of the bio-medical units **10** receives the communication signals **18** and processes them accordingly. The communication signals **18** may be instructions to collect data, to transmit collected data, to move the unit's position in the body, to perform a function, to administer a treatment, etc. If the received communication signals **18** require a response, the bio-medical unit **10** prepares an appropriate response and transmits it to the receiver **14** using a similar communication convention used by the transmitter **12**.

FIG. **2** is a diagram of another embodiment of a system that includes a plurality of bio-medical units **10** embedded within a body and/or placed on the surface of the body to facilitate diagnosis, treatment, and/or data collections. Each of the bio-medical units **10** is a passive device and, as such, includes a power harvesting module. The bio-medical units **10** may also include one or more of memory, a processing module, and functional modules. In this embodiment, the person is placed in an MRI machine (fixed or portable) that generates a magnetic field **26** through which the MRI transmitter **20** transmits MRI signals **28** to the MRI receiver **22**.

One or more of the bio-medical units **10** powers itself by harvesting energy from the magnetic field **26** or changes thereof as produced by gradient coils, from the magnetic fields of the MRI signals **28**, from the electrical fields of the MRI signals **28**, and/or from the electromagnetic aspects of the MRI signals **28**. A unit **10** converts the harvested energy into a supply voltage that supplies other components of the unit (e.g., a communication module, a processing module, memory, a functional module, etc.).

A communication device **24** communicates data and/or control communications **30** with one or more of the bio-medical units **10** over one or more wireless links. The communication device **24** may be a separate device from the MRI machine or integrated into the MRI machine. For example, the communication device **24**, whether integrated or separate, may be a cellular telephone, a computer with a wireless interface (e.g., a WLAN station and/or access point, Bluetooth, a proprietary protocol, etc.), etc. A wireless link may be one or more frequencies in the ISM band, in the 60 GHz frequency band, the ultrasound frequency band, and/or other frequency bands that supports one or more communication protocols (e.g., data modulation schemes, beamforming, RF or MMW modulation, encoding, error correction, etc.).

The composition of the bio-medical units **10** includes non-ferromagnetic materials (e.g., paramagnetic or diamagnetic) and/or metal alloys that are minimally affected by an external magnetic field **26**. In this regard, the units harvest power from the MRI signals **28** and communicate using RF and/or MMW electromagnetic signals with negligible chance of encountering the projectile or missile effect of implants that include ferromagnetic materials.

FIG. **3** is a diagram of an embodiment of an artificial body part **32** including one or more bio-medical units **10** that may be surgically implanted into a body. The artificial body part **32** may be a pace maker, a breast implant, a joint replacement, an artificial bone, splints, fastener devices (e.g., screws, plates, pins, sutures, etc.), artificial organ, etc. The artificial body part **32** may be permanently embedded in the body or temporarily embedded into the body.

FIG. **4** is a schematic block diagram of an embodiment of an artificial body part system **32** that includes an artificial body part and one or more bio-medical units **10**. The artificial body part (e.g., an electronic medical aid device (e.g., pacemaker, nerve stimulator, etc.), a metal artificial bone or portion thereof, a non-metal artificial bone or portion thereof, a metal screw, a non-metal screw, a metal plate, a non-metal plate, a metal splint, a non-metal splint, a metal suture, a non-metal suture, and/or an adhesive suture) includes a bio-medical mounting mechanism (e.g., a cavity, a surface mount area, etc.).

The bio-medical unit is mechanically associated with the artificial body part based on the bio-medical mounting mechanism. For example, if the artificial body part includes one or more cavities, the bio-medical unit **10** may positioned within the one or more cavities. As another example, if the artificial body part is an adhesive suture, the bio-medical unit **10** may be mounted on the surface of the suture and/or under a layer of the suture. As yet another example, if the artificial body part is molded, the bio-medical unit may be embedded within the artificial body part during the molding process.

The bio-medical unit includes a power harvesting module, a wireless communication module, and a functional module. The power harvesting module generates a supply voltage from a wireless source (e.g., MRI signals, magnetic fields, RF signals, MMW signals, etc.). The wireless communication module wirelessly communicates inbound data and outbound data with a communication device (e.g., device **24**). One or more embodiments of the power harvesting module and/or of the wireless communication module are discussed with reference to one or more of the preceding and/or following Figures.

The functional module performs at least one function regarding the artificial body part, wherein the functional module performs the at least one function in response to the inbound data and/or to generate the outbound data. For example, the at least one function may include a treatment function, an adjustment function, a diagnostic function, a sampling function, and/or a monitoring function. As a more specific example, the treatment function may include applying a medication, applying a laser treatment, applying an ultrasound treatment, and/or applying an electronic stimulation treatment. As another more specific example, the adjustment function may include adjusting force of the artificial body part (e.g., adjusting tension of a splint) and/or adjusting torque of the artificial body part (e.g., turning a screw).

In an embodiment, the bio-medical unit further includes a coil and/or an antenna. The power harvesting module may use the coil to generate the supply voltage from electromagnetic energy and/or use the antenna to generate the supply voltage from RF signals and/or MMW signals. The wireless communication module may use the coil for near field communication and/or use the antenna for RF and/or MMW communication. Note that the coil and/or antenna may be at least partially embedded within the artificial body part (e.g., at least partially in a cavity, forming threads of a screw, etc.). Alternatively, or in addition to, the artificial body part functions as the at least one of the coil and the antenna. For example, a metal screw, plate, splint, etc. may be formed to provide the coil and/or antenna, or include a section of the artificial body part to provide the coil and/or antenna.

In a further embodiment of the artificial body part system **32** includes the artificial body part and a plurality of bio-medical units **10**. The bio-medical units **10** may perform one or more functions. For example, a first bio-medical unit performs a treatment function; a second bio-medical unit performs an adjustment function; a third bio-medical unit performs a diagnostic function; a fourth bio-medical unit performs a sampling function; and a fifth bio-medical unit performs a monitoring function.

As another example, one bio-medical unit **10** may be used to detect infections, the body's acceptance of the artificial body part **32**, measure localized body temperature, monitor performance of the artificial body part **32**, and/or data gathering for other diagnostics. Another bio-medical unit **10** may be used for deployment of treatment (e.g., disperse medication, apply electrical stimulus, apply RF radiation, apply laser stimulus, etc.). Yet another bio-medical unit **10** may be used to adjust the position of the artificial body part **32** and/or a setting of the artificial body part **32**. For example, a bio-medical unit **10** may be used to mechanically adjust the tension of a splint, screws, etc. As another example, a bio-medical unit **10** may be used to adjust an electrical setting of the artificial body part **32**.

FIG. **5** is a diagram of another embodiment of a system that includes a plurality of bio-medical units **10** and one or more communication devices **24** coupled to a wide area network (WAN) communication device **34** (e.g., a cable modem, DSL modem, base station, access point, hot spot, etc.). The WAN communication device **34** is coupled to a network **42** (e.g., cellular telephone network, internet, etc.), which is coupled to it a plurality of remote monitors **36**, a plurality of databases **40**, and a plurality of computers **38**.

In an example of operation, one or more of the remote monitors **36** may receive images and/or other data **30** from one or more of the bio-medical units **10** via the communication device **24**, the WAN communication device **34**, and the network **42**. In this manner, a person(s) operating the remote monitors **36** may view images and/or the data **30** gathered by the bio-medical units **10**. This enables a specialist to be consulted without requiring the patient to travel to the specialist's office.

In another example of operation, one or more of the computers **38** may communicate with the bio-medical units **10** via the communication device **24**, the WAN communication device **34**, and the network **42**. In this example, the computer **36** may provide commands **30** to one or more of the bio-medical units **10** to gather data, to dispense a medication, to move to a new position in the body, to perform a mechanical function (e.g., cut, grasp, drill, puncture, stitch, patch, etc.), etc. As such, the bio-medical units **10** may be remotely controlled via one or more of the computers **36**.

In another example of operation, one or more of the bio-medical units **10** may read and/or write data from or to one or more of the databases **40**. For example, data (e.g., a blood sample analysis) generated by one or more of the bio-medical units **10** may be written to one of the databases **40**. The communication device **24** and/or one of the computers **36** may control the writing of data to or the reading of data from the database(s) **40**. The data may further include medical records, medical images, prescriptions, etc.

FIG. **6** is a diagram of another embodiment of a system that includes a plurality of bio-medical units **10**. In this embodiment, the bio-medical units **10** can communicate with each other directly and/or communicate with the communication device **24** directly. The communication medium may be an infrared channel(s), an RF channel(s), a MMW channel(s), and/or ultrasound. The units may use a communication protocol such as token passing, carrier sense, time division multiplexing, code division multiplexing, frequency division multiplexing, etc.

FIG. **7** is a diagram of another embodiment of a system that includes a plurality of bio-medical units **10**. In this embodiment, one of the bio-medical units **44** functions as an access point for the other units. As such, the designated unit **44** routes communications between the units **10** and between one or more units **10** and the communication device **24**. The communication medium may be an infrared channel(s), an RF channel(s), a MMW channel(s), and/or ultrasound. The units **10** may use a communication protocol such as token passing, carrier sense, time division multiplexing, code division multiplexing, frequency division multiplexing, etc.

FIG. **8** is a schematic block diagram of an embodiment of a bio-medical unit **10** that includes a power harvesting module **46**, a communication module **48**, a processing module **50**, memory **52**, and one or more functional modules **54**. The processing module **50** may be a single processing device or a plurality of processing devices. Such a processing device may be a microprocessor, micro-controller, digital signal processor, microcomputer, central processing unit, field programmable gate array, programmable logic device, state machine, logic circuitry, analog circuitry, digital circuitry, and/or any device that manipulates signals (analog and/or digital) based on hard coding of the circuitry and/or operational instructions. The processing module **50** may have an associated memory **52** and/or memory element, which may be a single memory device, a plurality of memory devices, and/or embedded circuitry of the processing module. Such a memory device **52** may be a read-only memory, random access memory, volatile memory, non-volatile memory, static memory, dynamic memory, flash memory, cache memory, and/or any device that stores digital information. Note that if the processing module **50** includes more than one processing device, the processing devices may be centrally located (e.g., directly coupled together via a wired and/or wireless bus structure) or may be distributedly located (e.g., cloud computing via indirect coupling via a local area network and/or a wide area network). Further note that when the processing module **50** implements one or more of its functions via a state machine, analog circuitry, digital circuitry, and/or logic circuitry, the memory and/or memory element storing the corresponding operational instructions may be embedded within, or external to, the circuitry comprising the state machine, analog circuitry, digital circuitry, and/or logic circuitry. Still further note that, the memory element stores, and the processing module executes, hard coded and/or operational instructions corresponding to at least some of the steps and/or functions illustrated in FIGS. **1-29**.

The power harvesting module **46** may generate one or more supply voltages **56** (Vdd) from one or more of MRI electromagnetic signals **16**, magnetic fields **26**, RF signals, MMW signals, and body motion. The power harvesting module **46** may be implemented as disclosed in U.S. Pat. No. 7,595,732 to generate one or more supply voltages from an RF signal. Alternatively, or in addition to, the power harvesting module **46** may be implemented as shown in one or more FIGS. **9-11** to generate one or more supply voltages **56** from an MRI signal **28** and/or magnetic field **26**. As yet another alternative, or a further addition, the power harvesting module **46** may be implemented as shown in FIG. **12** to generate one or more supply voltage **56** from body motion.

The communication module **48** may include a receiver section and a transmitter section. The transmitter section converts an outbound symbol stream into an outbound RF or MMW signal **60** that has a carrier frequency within a given frequency band (e.g., 900 MHz, 2.5 GHz, 5 GHz, 57-66 GHz, etc.). In an embodiment, this may be done by mixing the outbound symbol stream with a local oscillation to produce an up-converted signal. One or more power amplifiers and/or power amplifier drivers amplifies the up-converted signal, which may be RF or MMW bandpass filtered, to produce the outbound RF or MMW signal **60**. In another embodiment, the transmitter section includes an oscillator that produces an oscillation. The outbound symbol stream provides phase information (e.g., +/−Δθ[phase shift] and/or θ(t) [phase modulation]) that adjusts the phase of the oscillation to produce a phase adjusted RF or MMW signal, which is transmitted as the outbound RF signal **60**. In another embodiment, the outbound symbol stream includes amplitude information (e.g., A(t) [amplitude modulation]), which is used to adjust the amplitude of the phase adjusted RF or MMW signal to produce the outbound RF or MMW signal **60**.

In yet another embodiment, the transmitter section includes an oscillator that produces an oscillation. The outbound symbol provides frequency information (e.g., +/−Δf [frequency shift] and/or f(t) [frequency modulation]) that adjusts the frequency of the oscillation to produce a frequency adjusted RF or MMW signal, which is transmitted as the outbound RF or MMW signal **60**. In another embodiment, the outbound symbol stream includes amplitude information, which is used to adjust the amplitude of the frequency adjusted RF or MMW signal to produce the outbound RF or MMW signal **60**. In a further embodiment, the transmitter section includes an oscillator that produces an oscillation. The outbound symbol provides amplitude information (e.g., +/−ΔA [amplitude shift] and/or A(t) [amplitude modulation) that adjusts the amplitude of the oscillation to produce the outbound RF or MMW signal **60**.

The receiver section amplifies an inbound RF or MMW signal **60** to produce an amplified inbound RF or MMW signal. The receiver section may then mix in-phase (I) and quadrature (Q) components of the amplified inbound RF or MMW signal with in-phase and quadrature components of a local oscillation to produce a mixed I signal and a mixed Q signal. The mixed I and Q signals are combined to produce an inbound symbol stream. In this embodiment, the inbound symbol may include phase information (e.g., +/−Δθ[phase shift] and/or θ(t) [phase modulation]) and/or frequency information (e.g., +/−Δf [frequency shift] and/or f(t) [frequency modulation]). In another embodiment and/or in furtherance of the preceding embodiment, the inbound RF or MMW signal includes amplitude information (e.g., +/−ΔA [amplitude shift] and/or A(t) [amplitude modulation]). To recover the amplitude information, the receiver section includes an amplitude detector such as an envelope detector, a low pass filter, etc. Note that, as may be discussed herein, RF may include signals within the radio frequency band of 30 Hz to 3 GHz and/or within the microwave (or MMW) frequency band of 3 GHz to 300 GHz.

The processing module **50** generates the outbound symbol stream from outbound data and converts the inbound symbol stream into inbound data. For example, the processing module **50** converts the inbound symbol stream into inbound data (e.g., voice, text, audio, video, graphics, etc.) in accordance with one or more wireless communication standards (e.g., GSM, CDMA, WCDMA, HSUPA, HSDPA, WiMAX, EDGE, GPRS, IEEE 802.11, Bluetooth, ZigBee, universal mobile telecommunications system (UMTS), long term evolution (LTE), IEEE 802.16, evolution data optimized (EV-DO), etc.). Such a conversion may include one or more of: digital intermediate frequency to baseband conversion, time to frequency domain conversion, space-time-block decoding, space-frequency-block decoding, demodulation, frequency spread decoding, frequency hopping decoding, beamforming decoding, constellation demapping, deinterleaving, decoding, depuncturing, and/or descrambling.

As another example, the processing module **50** converts outbound data (e.g., voice, text, audio, video, graphics, etc.) into outbound symbol stream in accordance with one or more wireless communication standards (e.g., GSM, CDMA, WCDMA, HSUPA, HSDPA, WiMAX, EDGE, GPRS, IEEE 802.11, Bluetooth, ZigBee, universal mobile telecommunications system (UMTS), long term evolution (LTE), IEEE 802.16, evolution data optimized (EV-DO), etc.). Such a conversion includes one or more of: scrambling, puncturing, encoding, interleaving, constellation mapping, modulation, frequency spreading, frequency hopping, beamforming, space-time-block encoding, space-frequency-block encoding, frequency to time domain conversion, and/or digital baseband to intermediate frequency conversion.

Each of the one or more functional modules **54** provides a function to support treatment, data gathering, motion, repairs, and/or diagnostics. The functional modules **54** may be implemented using nanotechnology and/or microelectronic mechanical systems (MEMS) technology. Various examples of functional modules **54** are illustrated in one or more of FIGS. **13-29**.

The bio-medical unit **10** may be encapsulated by an encapsulate **58** that is non-toxic to the body. For example, the encapsulate **58** may be a silicon based product, a non-ferromagnetic metal alloy (e.g., stainless steel), etc. As another example, the encapsulate **58** may include a spherical shape and have a ferromagnetic liner that shields the unit from a magnetic field and to offset the forces of the magnetic field.

The bio-medical unit **10** may be implemented on a single die that has an area of a few millimeters or less. The die may be fabricated in accordance with CMOS technology, Gallium-Arsenide technology, and/or any other integrated circuit die fabrication process.

FIG. **9** is a schematic block diagram of an embodiment of a power harvesting module **46** that includes an array of on-chip air core inductors **64**, a rectifying circuit **66**, capacitors, and a regulation circuit **68**. The inductors **64** may each have an inductance of a few nano-Henries to a few micro-Henries and may be coupled in series, in parallel, or a series parallel combination.

In an example of operation, the MRI transmitter **20** transmits MRI signals **28** at a frequency of 3-45 MHz at a power level of up to 35 KWatts. The air core inductors **64** are electromagnetically coupled to generate a voltage from the magnetic and/or electric field generated by the MRI signals **28**. Alternatively, or in addition to, the air core inductors **64** may generate a voltage from the magnetic field **26** and changes thereof produced by the gradient coils. The rectifying circuit **66** rectifies the AC voltage produced by the inductors to produce a first DC voltage. The regulation circuit generates one or more desired supply voltages **56** from the first DC voltage.

The inductors **64** may be implemented on one more metal layers of the die and include one or more turns per layer. Note that trace thickness, trace length, and other physical properties affect the resulting inductance.

FIG. **10** is a schematic block diagram of another embodiment of a power harvesting module **46** that includes a plurality of on-chip air core inductors **70**, a plurality of switching units (S), a rectifying circuit **66**, a capacitor, and a switch controller **72**. The inductors **70** may each having an inductance of a few nano-Henries to a few micro-Henries and may be coupled in series, in parallel, or a series parallel combination.

In an example of operation, the MRI transmitter **20** transmits MRI signals **28** at a frequency of 3-45 MHz at a power level of up to 35 KWatts. The air core inductors **70** are electromagnetically coupled to generate a voltage from the magnetic and/or electric field generated by the MRI signals **28**. The switching module **72** engages the switches via control signals **74** to couple the inductors **70** in series and/or parallel to generate a desired AC voltage. The rectifier circuit **66** and the capacitor(s) convert the desired AC voltage into the one or more supply voltages **56**.

FIG. **11** is a schematic block diagram of another embodiment of a power harvesting module **46** that includes a plurality of Hall effect devices **76**, a power combining module **78**, and a capacitor(s). In an example of operation, the Hall effect devices **76** generate a voltage based on the constant magnetic field (H) and/or a varying magnetic field. The power combining module **78** (e.g., a wire, a switch network, a transistor network, a diode network, etc.) combines the voltages of the Hall effect devices **76** to produce the one or more supply voltages **56**.

FIG. **12** is a schematic block diagram of another embodiment of a power harvesting module **46** that includes a plurality of piezoelectric devices **82**, a power combining module **78**, and a capacitor(s). In an example of operation, the piezoelectric devices **82** generate a voltage based on body movement, ultrasound signals, movement of body fluids, etc. The power combining module **78** (e.g., a wire, a switch network, a transistor network, a diode network, etc.) combines the voltages of the Hall effect devices **82** to produce the one or more supply voltages **56**. Note that the piezoelectric devices **82** may include one or more of a piezoelectric motor, a piezoelectric actuator, a piezoelectric sensor, and/or a piezoelectric high voltage device.

The various embodiments of the power harvesting module **46** may be combined to generate more power, more supply voltages, etc. For example, the embodiment of FIG. **9** may be combined with one or more of the embodiments of FIGS. **11** and **12**.

FIG. **13** is a schematic block diagram of another embodiment of a bio-medical unit **10** that includes a power harvesting module **46**, a communication module **48**, a processing module **50**, memory **52**, and may include one or more functional modules **54** and/or a Hall effect communication module **116**. The communication module **48** may include one or more of an ultrasound transceiver **118**, an electromagnetic transceiver **122**, an RF and/or MMW transceiver **120**, and a light source (LED) transceiver **124**. Note that examples of the various types of communication modules **48** will be described in greater detail with reference to one or more of FIGS. **14-29**.

The one or more functional modules **54** may perform a repair function, an imaging function, and/or a leakage detection function, which may utilize one or more of a motion propulsion module **96**, a camera module **98**, a sampling robotics module **100**, a treatment robotics module **102**, an accelerometer module **104**, a flow meter module **106**, a transducer module **108**, a gyroscope module **110**, a high voltage generator module **112**, a control release robotics module **114**, and/or other functional modules described with reference to one or more other figures. The functional modules **54** may be implemented using MEMS technology and/or nanotechnology. For example, the camera module **98** may be implemented as a digital image sensor in MEMS technology. Example of these various modules will be described in greater detail with reference to one or more of FIGS. **14-29**.

The Hall effect communication module **116** utilizes variations in the magnetic field and/or electrical field to produce a plus or minus voltage, which can be encoded to convey information. For example, the charge applied to one or more Hall effect devices **76** may be varied to produce the voltage change. As another example, an MRI transmitter **20** and/or gradient unit may modulate a signal on the magnetic field **26** it generates to produce variations in the magnetic field **26**.

FIG. **14** is a diagram of another embodiment of a system that includes one or more bio-medical units **10**, a transmitter unit **126**, and a receiver unit **128**. Each of the bio-medical units **10** includes a power harvesting module **46**, an RF and/or MMW (RF) transceiver **138**, a processing module **50**, memory **52** and one or more functional modules (as shown in one or more of the previous figures). The transmitter unit **126** includes an electromagnetic (e.g., a MRI) transmitter **130** and a MMW transmitter **132**. The receiver unit **128** includes an electromagnetic (e.g., MRI) receiver **134** and a MMW receiver **136**. Note that the MMW transmitter **132** and MMW receiver **136** may be in the same unit (e.g., in the transmitter unit, in the receiver unit, or housed in a separate device).

In an example of operation, the electromagnetic transmitter **130** transmits an electromagnetic (EM) signal **146** in a predetermined pattern (e.g., a magnetic resonance imaging (MRI) signal having at least one of a radio frequency (RF) component and a gradient magnetic field component that has a pattern as shown in FIG. **15**). In addition, the RF communication transmitter **132** of the transmitter unit **126** transmits a downstream RF communication signal, which may contain a request for data, a command, etc.

The power harvesting module **46** receives the EM signal **146** and generates a power supply voltage therefrom. The RF communication module **138**, which is powered by the power supply voltage, receives the downstream RF communication signal during at least one of the null intervals of the transmission pattern. The RF communication module **138** then converts the downstream RF communication signal into downstream information, which includes at least one of downstream data and a downstream instruction.

The functional module performs a bio-medical function based on the at least one of downstream data and the downstream instruction. For example, the bio-medical function may be to take a temperature reading, to release a certain amount of a medication, to take a blood-oxygen reading, etc. As another example, the bio-medical function may cause the bio-medical unit to store the downstream data.

In response to the downstream instruction, or based on another stimulus (e.g., pre-programmed, timed interval, etc.), the bio-medical unit **10** may have data and/or instruction to transmit. In this instance, the function module generates at least one of an upstream instruction and upstream data. The processing module converts the upstream instruction and/or upstream data into upstream information.

The RF communication module **138** interprets the predetermined pattern of the transmitting of the electromagnetic signal to determine null intervals. In addition, the RF communication module **138** converts the upstream information into an upstream RF communication signal and transmits the upstream RF communication signal during one or more of the null intervals. The RF communication receiver **134** of the receiver unit **128** receives the upstream RF communication signal and processes it accordingly.

FIG. **15** is a diagram of an example of a communication protocol within the system of FIG. **14**. In this diagram, the MRI transmitter **20** transmits RF signals **152**, which have a frequency in the range of 3-45 MHz, at various intervals with varying signal strengths. The power harvesting module **46** of the bio-medical units **10** may use these signals to generate power for the bio-medical unit **10**.

In addition to the MRI transmitter **20** transmitting its signal, a constant magnetic field and various gradient magnetic fields **154**-**164** are created (one or more in the x dimension Gx, one or more in the y dimension Gy, and one or more in the z direction Gz). The power harvesting module **46** of the bio-medical unit **10** may further use the constant magnetic field and/or the varying magnetic fields **154**-**164** to create power for the bio-medical unit **10**.

During non-transmission periods of the cycle, the bio-medical unit **10** may communicate **168** with the MMW transmitter **132** and/or MMW receiver **136**. In this regard, the bio-medical unit **10** alternates from generating power to MMW communication in accordance with the conventional transmission-magnetic field pattern of an MRI machine.

FIG. **16** is a schematic block diagram of an embodiment of networked bio-medical units **10** that communicate with each other, perform sensing functions to produce sensed data **218**-**232**, process the sensed data to produce processed data, and transmit the processed data **216**. The bio-medical units **10** may be positioned in a body part to sense data across the body part and to transmit data to an external communication device. The transmitted data may be further processed or aggregated from sensed data.

The bio-medical units **10** may monitor various types of biological functions over a short term or a long term to produce the sensed data **218**-**232**. Note that the sensed data **218**-**232** may include blood flow rate, blood pressure, temperature, air flow, blood oxygen level, density, white cell count, red cell count, position information, etc.

The bio-medical unit **10** establishes communications with one or more other bio-medical units **10** to facilitate the communication of sensed data **218**-**232** and processed data **216**. The communication may include EM signals, MMW signals, optical signals, sound signals, and/or RF signals.

The bio-medical unit **10** may determine position information based on the sensed data **218**-**232** and include the position information in the communication. The bio-medical unit **10** may also determine a mode of operation based on one or more of a command, a list, a predetermination, sensed data, and/or processed data. For example, a bio-medical unit **10** at the center of the body part may be in a mode to sense temperature and a bio-medical unit **10** at the outside edge of the body part may sense blood flow.

The bio-medical unit **10** may receive processed data **218**-**232** from another bio-medical unit and re-send the same processed data **218**-**232** to yet another bio-medical unit **10**. The bio-medical unit **10** may produce processed data based on sensed data **218**-**232** from the bio-medical unit **10** and/or received processed data from another bio-medical unit **10**.

FIG. **17** is a flowchart illustrating the processing of networked bio-medical unit data where the bio-medical unit determines the sense mode based on one or more of a predetermination, a stored mode indicator in memory, a command, and/or a dynamic sensed data condition. The method begins at step **234** where the bio-medical unit **10** determines the mode. The method branches to step **240** when the bio-medical unit **10** determines that the mode is process and sense. The method continues to step **236** when the bio-medical unit **10** determines that the mode is sense only.

At step **236**, the bio-medical unit **10** gathers data from one or more of the functional modules **54** to produce sensed data. The bio-medical unit **10** may transmit the sensed data **238** to another bio-medical unit **10** and/or an external communication device in accordance with the sense mode. For example, the bio-medical unit **10** may transmit the sensed data at a specific time, to a specific bio-medical unit **10**, to a specific external communication device, after a certain time period, when the data is sensed, and/or when the sensed data compares favorably to a threshold (e.g., a temperature trip point).

The method continues at step **240** where the bio-medical unit **10** determines whether it has received data from another unit **10**. If not, the method continues to step **250**, where the bio-medical unit **10** transmits its sensed data to another bio-medical unit **10** and/or an external communication device in accordance with the sense mode.

When the bio-medical unit **10** has received data from another unit, the method continues at step **242**, where the bio-medical unit **10** determines a data function to perform based on one or more of the content of the received data, the sensed data, a command, and/or a predetermination. The data function may one or more of initialization, comparing, compiling, and/or performing a data analysis algorithm.

The method continues at step **244**, where the bio-medical unit **10** gathers data from the functional modules **54**, and/or the received data from one or more other bio-medical units **10**. The method continues at step **246**, where the bio-medical unit **10** processes the data in accordance with a function to produce processed data. In addition to the example provided above, the function may also include the functional assignment of the bio-medical unit **10** as determined by a predetermination, a command, sensed data, and/or processed data (e.g., measure blood pressure from the plurality of bio-medical units and summarize the high, low, and average).

The method continues at step **248**, where the bio-medical unit **10** transmits the processed data to another bio-medical unit **10** and/or to an external communication device in accordance with the sense mode. For example, the bio-medical unit **10** may transmit the sensed data at a specific time, to a specific bio-medical unit **10**, to a specific external communication device, after a certain time period, when the data is sensed, and/or when the sensed data compares favorably to a threshold (e.g., a temperature trip point). Note that the communication protocol may be the same or different between bio-medical units **10** and/or between the bio-medical unit **10** and the external communication device.

FIG. **18** is a schematic block diagram of an embodiment of a parent bio-medical unit (on the left) communicating with an external unit to coordinates the functions of one or more children bio-medical units **10** (on the right). The parent unit includes a communication module **48** for external communications, a communication module **48** for communication with the children units, the processing module **50**, the memory **52**, and the power harvesting module **46**. Note that the parent unit may be implemented one or more chips and may in the body or one the body.

Each of the child units includes a communication module **48** for communication with the parent unit and/or other children units, a MEMS robotics **244**, and the power harvesting module **46**. The MEMS robotics **244** may include one or more of a MEMS technology saw, drill, spreader, needle, injection system, and actuator. The communication module **48** may support RF and/or MMW inbound and/or outbound signals **60** to the parent unit such that the parent unit may command the child units in accordance with external communications commands.

In an example of operation, the patent bio-medical unit receives a communication from the external source, where the communication indicates a particular function the child units are to perform. The parent unit processes the communication and relays relative portions to the child units in accordance with a control mode. Each of the child units receives their respective commands and performs the corresponding functions to achieve the desired function.

FIG. **19** is a schematic block diagram of another embodiment of a plurality of task coordinated bio-medical units **10** including a parent bio-medical unit **10** (on the left) and one or more children bio-medical units **10** (on the right). The parent unit may be implemented one or more chips and may in the body or one the body. The parent unit may harvest power in conjunction with the power booster **84**.

The parent unit includes the communication module **48** for external communications, the communication module **48** for communication with the children units, the processing module **50**, the memory **52**, a MEMS electrostatic motor **248**, and the power harvesting module **46**. The child unit includes the communication module **48** for communication with the parent unit and/or other children units, a MEMS electrostatic motor **248**, the MEMS robotics **244**, and the power harvesting module **46**. Note that the child unit has fewer components as compared to the parent unit and may be smaller facilitating more applications where smaller bio-medical units **10** enhances their effectiveness.

The MEMS robotics **244** may include one or more of a MEMS technology saw, drill, spreader, needle, injection system, and actuator. The MEMS electrostatic motor **248** may provide mechanical power for the MEMS robotics **244** and/or may provide movement propulsion for the child unit such that the child unit may be positioned to optimize effectiveness. The child units may operate in unison to affect a common task. For example, the plurality of child units may operate in unison to saw through a tissue area.

The child unit communication module **48** may support RF and/or MMW inbound and/or outbound signals **60** to the parent unit such that the parent unit may command the children units in accordance with external communications commands.

The child unit may determine a control mode and operate in accordance with the control mode. The child unit determines the control mode based on one or more of a command from a parent bio-medical unit, external communications, a preprogrammed list, and/or in response to sensor data. Note that the control mode may include autonomous, parent (bio-medical unit), server, and/or peer as previously discussed.

FIG. **20** is a schematic block diagram of an embodiment of a bio-medical unit **10** based imaging system that includes the bio-medical unit **10**, the communication device **24**, a database **254**, and an in vivo image unit **252**. The bio-medical unit **10** may perform scans and provide the in vivo image unit **252** with processed image data for diagnostic visualization.

The bio-medical unit **10** includes a MEMS image sensor **256**, the communication module **48** for external communications with the communication device, the processing module **50**, the memory **52**, the MEMS electrostatic motor **248**, and the power harvesting module **46**. In an embodiment the bio-medical unit **10** and communication device **24** communicate directly. In another embodiment, the bio-medical unit **10** and communication device **24** communicate through one or more intermediate networks (e.g., wireline, wireless, cellular, local area wireless, Bluetooth, etc.). The MEMS image sensor **256** may include one or more sensors scan types for optical signals, MMW signals, RF signals, EM signals, and/or sound signals.

The in vivo unit **252** may send a command to the bio-medical unit **10** via the communication device **24** to request scan data. The request may include the scan type. The in vivo unit **252** may receive the processed image data from the bio-medical unit **10**, compare it to data in the database **254**, process the data further, and provide image visualization.

FIG. **21** is a schematic block diagram of an embodiment of a bio-medical unit **10** based sounding system that includes the bio-medical unit **10**, the communication device **24**, the database **254**, and a speaker **260**. The bio-medical unit **10** may perform scans and provide the speaker **260** with processed sounding data for diagnostic purposes via the communication device **24**.

The bio-medical unit **10** includes a MEMS microphone **262**, the communication module **48** for external communications with the communication device **24**, the processing module **50**, the memory **52**, the MEMS electrostatic motor **248**, and the power harvesting module **46**. In an embodiment the bio-medical unit **10** and communication device **24** communicate directly. In another embodiment, the bio-medical unit **10** and communication device **24** communicate through one or more intermediate networks (e.g., wireline, wireless, cellular, local area wireless, Bluetooth, etc.) The MEMS microphone **262** may include one or more sensors to detect audible sound signals, sub-sonic sound signals, and/or ultrasonic sound signals.

The processing module **50** may produce the processed sounding data based in part on the received sound signals and in part on data in the database **254**. The processing module **50** may retrieve data via the communication module **48** and communication device **24** link from the database **254** to assist in the processing of the signals (e.g., pattern matching, filter recommendations, sound field types). The processing module **50** may process the signals to detect objects, masses, air flow, liquid flow, tissue, distances, etc. The processing module **50** may provide the processed sounding data to the speaker **260** for audible interpretation. In another embodiment, the bio-medical unit **10** assists an ultrasound imaging system by relaying ultrasonic sounds from the MEMS microphone **262** to the ultrasound imaging system instead of to the speaker **260**.

FIG. **22** is a flowchart of an embodiment of a method for controlling power harvesting within a bio-medical unit **10**.

The method begins at step **418** wherein the processing module **50** of the bio-medical unit **10** initializes (e.g., when it is supplied power and wakes up) itself. For example, the processing module **50** executes an initialization boot sequence stored in the memory **52**. The initialization boot sequence includes operational instructions that cause the processing module to initialize its registers to accept further instructions. The initialization boot sequence may further include operational instructions to initialize one or more of the communication module **48**, the functional module(s) **54** initialized, etc.

The method continues at step **420** where the processing module **50** determines the state of the bio-medical unit (e.g., actively involved in a task, inactive, data gathering, performing a function, etc.). Such a determination may be based on one or more of previous state(s) (e.g., when the processing module was stopped prior to losing power), an input from the functional module **54**, a list of steps or elements of a task, the current step of a MRI sequence, and/or new tasks received via the communication module **48**.

The method continues at step **422** where the processing module **50** determines the bio-medical unit power level, which may be done by measuring the power harvesting module **46** output Vdd **56**. Note that voltage is one proxy for the power level and that other proxies may be utilized including estimation of milliWatt-hours available, a time of operation before loss of operating power estimate, a number of CPU instructions estimate, a number of task elements, a number of tasks estimate, and/or another other estimator to assist in determining how much the bio-medical unit **10** can accomplish prior to losing power. Further note that the processing module **50** may save historic records of power utilization in the memory **52** to assist in subsequent determinations of the power level.

The method continues at step **424** where the processing module **50** compares the power level to the high threshold (e.g., a first available power level that allows for a certain level of processing). If yes, the method continues to step **426** where the processing module **50** enables the execution of H number of instructions. The processing module **50** may utilize a predetermined static value of the H instructions or a dynamic value that changes as a result of the historic records. For example, the historic records may indicate that there was an average of 20% more power capacity left over after the last ten times of instruction execution upon initialization. The processing module **50** may adjust the value of H upward such that the on-going left over power is less than 20% in order to more fully utilize the available power each time the bio-medical unit **10** has power.

The method continues at step **428** where the processing module **50** saves the state in the memory **52** upon completion of the execution of the H instructions such that the processing module **50** can start in a state in accordance with this state upon the next initialization. The method then continues at step **430** where the processing module **50** determines whether it will suspend operations based on one or more of a redetermined power level (e.g., power left after executing the instructions), a predetermined list, a task priority, a task state, a priority indicator, a command, a message, and/or a functional module input. If not, the method repeats at step **422**. If yes, the method branches to step **440** where the processing module **50** suspends operations of the bio-medical unit.

If, at step **424**, the power level is not greater than the high threshold, the method continues at step **432** where the processing module **50** determines whether the power level compares favorably to a low threshold. If not, the method continues a step **440** where the processing module **50** suspends operations of the bio-medical unit.

If the comparison at step **432** was favorable, the method continues at step **434** where the processing module **50** executes L instructions. The processing module **50** may utilize a predetermined static value of the L instructions or a dynamic value that changes as a result of the historic records as discussed previously. For example, the historic records may indicate that there was an average of 10% more power capacity left over after the last ten times of instruction execution upon initialization. The processing module **50** may adjust the value of L downward such that the on-going left over power is less than 10% in order to more fully utilize the available power each time the bio-medical unit **10** has power.

The method continues at step **436** where the processing module **50** saves the state in the memory **52** upon completion of the execution of the L instructions such that the processing module **50** can start in a state in accordance with this state upon the next initialization. The method then continues at step **438** where the processing module **50** determines whether it will suspend operations based on one or more of a redetermined power level (e.g., power left after executing the instructions), a predetermined list, a task priority, a task state, a priority indicator, a command, a message, and/or a functional module input. If yes, the method branches to step **440**. If not, the method repeats at step **422**.

FIG. **23** is a flowchart illustrating MMW communications within a MRI sequence where the processing module **50** determines MMW communications in accordance with an MRI sequence. The method begins at step **442** where the processing module **50** determines whether the MRI is active based on receiving MRI EM signals. At step **444**, the method branches to step **446** or step **448**. When the MRI is active, the method continues at step **446** where the processing module **50** performs MMW communications as previously discussed.

The method continues at step **448** where the processing module **50** determines the MRI sequence based on received MRI EM signals (e.g., gradient pulses and/or MRI RF pulses as shown in one or more of the preceding figures). The method continues at step **450** where the processing module **50** determines whether it is time to perform receive MMW communication in accordance with the MRI sequence. For example, the MMW transceiver **138** may receive MMW inbound signals **148** between any of the MRI sequence steps. As another example, the MMW transceiver **138** may receive MMW inbound signals **148** between specific predetermined steps of the MRI sequence.

At step **452** the method branches back to step **450** or to step **454**. When it is time to receive, the method continues at step **454** where the processing module **50** coordinates the MMW transceiver **138** receiving the MMW inbound signals, which may include one or more of a status request, a records request, a sensor data request, a processed data request, a position request, a command, and/or a request for MRI echo signal data. The method then continues at step **456** where the processing module **50** determines whether there is at least one message pending to transmit (e.g., in a transmit queue). At step **458** the method branches back to step **442** or to step **460**.

At step **460**, the processing module **50** determines when it is time to transmit a MMW communication in accordance with the MRI sequence. For example, the MMW transceiver **138** may transmit MMW outbound signals **150** between any of the MRI sequence steps. As another example, the MMW transceiver **138** may transmit MMW outbound signals **150** between specific predetermined steps of the MRI sequence.

At step **462**, the method branches to back step **456** or to step **464**. The method continues at step **464** where the processing module **50** coordinates the MMW transceiver **138** transmitting the MMW outbound signals **150**, which may include one or more of a status request response, a records request response, a sensor data request response, a processed data request response, a position request response, a command response, and/or a request for MRI echo signal data response. The method then branches back to step **442**.

FIG. **24** is a flowchart illustrating the processing of MRI signals where the processing module **50** of the bio-medical unit **10** may assist the MRI in the reception and processing of MRI EM signals **146**. The method begins at step **466** where the processing module **50** determines if the MRI is active based on receiving MRI EM signals **146**. The method branches back to step **466** when the processing module **50** determines that the MRI is not active. For example, the MRI sequence may not start until the processing module **50** communicates to the MRI unit that it is available to assist. The method continues to step **470** when the processing module **50** determines that the MRI is active.

At step **470**, the processing module **50** determines the MRI sequence based on received MRI EM signals **146** (e.g., gradient pulses and/or MRI RF pulses). At step **472**, the processing module receives EM signals **146** and/or MMW communication **532** in accordance with the MRI sequence and decodes a message. For example, the MMW transceiver **138** may receive MMW inbound signals **148** between any of the MRI sequence steps. As another example, the MMW transceiver **138** may receive MMW inbound signals **148** between specific predetermined steps of the MRI sequence. In yet another example, the processing module **50** may receive EM signals **146** at any point of the MRI sequence such that the EM signals **146** contain a message for the processing module **50**.

At step **474**, the processing module **50** determines whether to assist in the MRI sequence based in part on the decoded message. The determination may be based on a comparison of the assist request to the capabilities of the bio-medical unit **10**. At step **476**, the method branches to step **480** when the processing module **50** determines to assist in the MRI sequence. The method continues at step **478** where the processing module **50** performs other instructions contained in the message and the method ends.

At step **480**, the processing module **50** begins the assist steps by receiving echo signals **530** during the MRI sequence. Note the echo signals **530** may comprise EM RF signals across a wide frequency band as reflected off of tissue during the MRI sequence. At step **482**, the processing module **50** processes the received echo signals **530** to produce processed echo signals. Note that this may be a portion of the overall processing required to lead to the desired MRI imaging.

At step **484**, the processing module **50** determines the assist type based on the decoded message from the MRI unit. The assist type may be at least passive or active where the passive type collects echo signal **530** information and sends it to the MRI unit via MMW outbound signals **150** and the active type collects echo signal information and re-generates a form of the echo signals **530** and sends the re-generated echo signals to the MRI unit via outbound modulated EM signals (e.g., the MRI unit interprets the re-generated echo signals as echo signals to improve the overall system gain and sensitivity).

The method branches to step **494** when the processing module **50** determines the assist type to be active. The method continues to step **486** when the processing module **50** determines the assist type to be passive. At step **486**, the processing module **50** creates an echo message based on the processed echo signals where the echo message contains information about the echo signals **530**.

At step **488**, the processing module **50** determines when it is time to transmit the echo message encoded as MMW outbound signals **150** via MMW communication in accordance with the MRI sequence. For example, the MMW transceiver **138** may transmit MMW outbound signals **150** between any of the MRI sequence steps. In another example, the MMW transceiver **138** may transmit MMW outbound signals **150** between specific predetermined steps of the MRI sequence.

At step **490**, the method branches back to step **488** when the processing module **50** determines that it is not time to transmit the echo message. At step **490**, the method continues to step **492** where the processing module **50** transmits the echo message encoded as MMW outbound signals **150**.

At step **494**, the processing module **50** creates echo signals based on the processed echo signals. At step **496**, the processing module **50** determines when it is time to transmit the echo signals as outbound modulated EM signals **180** in accordance with the MRI sequence. At step **498**, the method branches back to step **496** when the processing module **50** determines that it is not time to transmit the echo signals. At step **498**, the method continues to step **500** where the processing module **50** transmits the echo signals encoded as outbound modulated EM signals **180**. Note that the transmitted echo signals emulate the received echo signals **530** with improvements to overcome low MRI power levels and/or low MRI receiver sensitivity.

FIG. **25** is a flowchart illustrating communication utilizing MRI signals where the processing module **50** determines MMW signaling in accordance with an MRI sequence. The method begins at step **502** where the processing module **50** determines if the MRI is active based on receiving MRI EM signals **146**. At step **504**, the method branches to step **508** when the processing module **50** determines that the MRI is active. At step **504**, the method continues to step **506** when the processing module **50** determines that the MRI is not active. At step **506**, the processing module **50** queues pending transmit messages. The method branches to step **502**.

At step **508**, the processing module **50** determines the MRI sequence based on received MRI EM signals **146** (e.g., gradient pulses and/or MRI RF pulses). At step **510**, the processing module **50** determines when it is time to perform receive communication in accordance with the MRI sequence. For example, the EM transceiver **174** may receive inbound modulated EM signals **146** containing message information from any of the MRI sequence steps.

At step **512**, the method branches back to step **510** when the processing module **50** determines that it is not time to perform receive communication. At step **512**, the method continues to step **514** where the processing module **50** directs the EM transceiver **174** to receive the inbound modulated EM signals. The processing module **50** may decode messages from the inbound modulated EM signals **146** such that the messages include one or more of a echo signal collection assist request, a status request, a records request, a sensor data request, a processed data request, a position request, a command, and/or a request for MRI echo signal data. Note that the message may be decoded from the inbound modulated EM signals **146** in one or more ways including detection of the ordering of the magnetic gradient pulses, counting the number of gradient pulses, the slice pulse orderings, detecting small differences in the timing of the pulses, and/or demodulation of the MRI RF pulse.

At step **516** the processing module **50** determines if there is at least one message pending to transmit (e.g., in a transmit queue). At step **518**, the method branches back to step **502** when the processing module **50** determines that there is not at least one message pending to transmit. At step **518**, the method continues to step **520** where the processing module **50** determines when it is time to perform transmit communication in accordance with the MRI sequence. For example, the EM transceiver **174** may transmit outbound modulated EM signals **180** between any of the MRI sequence steps. In another example, the EM transceiver **174** may transmit the outbound modulated EM signals **180** between specific predetermined steps of the MRI sequence. In yet another example, the EM transceiver **174** may transmit the outbound modulated EM signals **180** in parallel with specific predetermined steps of the MRI sequence, but may utilize a different set of frequencies unique to the EM transceiver **174**.

At step **522**, the method branches back to step **520** when the processing module **50** determines that it is not time to perform transmit communication. At step **522**, the method continues to step **524** where the processing module **50** directs the EM transceiver **174** to prepare the outbound modulated EM signals **180** based on the at least one message pending to transmit. The processing module **50** may encode messages into the outbound modulated EM signals **180** such that the messages include one or more of a status request response, a records request response, a sensor data request response, a processed data request response, a position request response, a command response, and/or a request for MRI echo signal data response. The method branches back to step **502**.

FIG. **26** is a schematic block diagram of an embodiment of a MRI system utilizing a plurality of bio-medical units **10** to provide processed echo signal information based on received MRI echo signals **530** across the plurality of bio-medical units **10**. The MRI system includes the transmitter unit **126**, the receiver unit, **128** and a plurality of bio-medical units **10** equipped to communicate with each other and with the transmitter unit **126** and the receiver unit **128**. The bio-medical unit **10** includes an echo receiver **528**, the MMW transceiver **138**, the power harvesting module **46**, the memory **52**, and the processing module **50**. The echo receiver **528** receives echo signals **530** from an object **526** in accordance with the MRI sequence.

The processing module **50** produces processed echo signals based on the echo signals **530** received by the echo receiver **528**. The processing module **50** may determine to share the processed echo signal information with the plurality of bio-medical units **10** based in part on instructions received by the MMW transceiver **138** from the transmitter unit **126**. The processing module **50** may produce further processed echo signals based on the shared processed echo signals of the plurality of bio-medical units **10**. The processing module **50** may encode and send the further processed echo signal information as MMW outbound signals **150** to the receiver unit **128**. The method of processing and communicating the processed echo signals is discussed in greater detail with reference to FIG. **27**.

FIG. **27** is another flowchart illustrating the processing of MRI signals where the processing module **50** of the bio-medical unit **10** may communicate with other bio-medical units **10** to assist the MRI in the reception and processing of MRI EM signals **146**. The method begins at step **534** where the processing module **50** determines if the MRI is active based on receiving MRI EM signals **146** via the EM transceiver **174**. At step **536**, the method branches back to step **534** when the processing module **50** determines that the MRI is not active. For example, the MRI sequence may not start until the processing module **50** communicates to the MRI that it is available to assist. At step **536**, the method continues to step **538** when the processing module **50** determines that the MRI is active.

At step **538**, the processing module **50** determines the MRI sequence based on received MRI EM signals **146** (e.g., gradient pulses and/or MRI RF pulses). At step **540**, the processing module **50** receives EM signals **146** and/or MMW inbound signals **148** in accordance with the MRI sequence and decodes a message for the processing module **50**.

At step **542**, the processing module **50** determines whether to receive echo signals **530** produced by the MRI sequence based in part on the decoded message. The determination may be based on a comparison of the MRI echo signal data request to the capabilities of the bio-medical unit **10**. At step **544**, the method branches to step **548** when the processing module **50** determines to receive echo signals **530**. At step **544**, the method continues to step **546** when the processing module **50** determines to not receive echo signals **530**. At step **546**, the processing module **50** performs other instructions contained in the message. The method ends.

At step **548**, the processing module **50** establishes communications with other bio-medical units **10**. The processing module **50** may establish communications with other bio-medical units **10** utilizing MMW communication **532**. The processing module **50** may send a receive echo signals command to the other bio-medical units **10** such that the plurality of bio-medical units **10** will all receive subsequent echo signals **530**.

At step **550**, the processing module **50** receives echo signals **530** via the echo receiver **528**. Note the echo signals **530** may comprise EM RF signals across a wide frequency band as reflected off of the object **526** during the MRI sequence. At step **552**, the processing module **50** processes the received echo signals **530** to produce processed echo signals. Note that this may be a portion of the overall processing required to lead to the desired MRI imaging.

At step **554**, the processing module **50** sends the processed echo signals to other bio-medical units via MMW communication **532**. Note that bio-medical units **10** receiving the processed echo signals may re-send the processed echo signals to still other bio-medical units **10**. Further note that the one of more of the plurality of bio-medical units **10** may further process the processed echo signals to produce processed echo signals.

At step **556**, the processing module **50** may receive the processed echo signals from other bio-medical units via MMW communication **532**. Note that the process echo signals may include functional module **54** information including sensor data and location. The processing module **50** may process the processed echo signals to produce processed echo signals (e.g., refined signals, an aggregation, a composite, a result, etc.).

At step **558**, the processing module **50** creates an echo message based on the processed echo signals where the echo message contains information about the echo signals **530**. At step **560**, the processing module **50** determines when it is time to transmit the echo message encoded as MMW outbound signals **150** via MMW communication **532** in accordance with the MRI sequence. For example, the MMW transceiver **138** may transmit MMW outbound signals **150** between any of the MRI sequence steps. For example, the MMW transceiver **138** may transmit MMW outbound signals **150** between specific predetermined steps of the MRI sequence.

At step **562**, the method branches back to step **560** when the processing module **50** determines it is not time to transmit the echo message. At step **566**, the method continues to step **564** when the processing module **50** determines it is time to transmit the echo message. At step **564**, the processing module **50** transmits the echo message encoded as MMW outbound signals.

FIG. **28** is a flowchart illustrating the communication of records where the processing module **50** of the bio-medical unit **10** determines to provide medical records. The method begins at step **566** where the processing module **50** determines if receiving MMW communication is allowed. The determination may be based on one or more of a timer, a command, available power, a priority indicator, and/or interference indicator. For example, the MMW transceiver **138** may receive MMW inbound signals **148** for a 500 millisecond window every 3 minutes.

At step **568**, the method branches back to step **566** when the processing module **50** determines that receiving MMW communication is not allowed. At step **568**, the method continues to step **570** where the processing module **50** directs the MMW transceiver **138** to receive MMW inbound signals **148**. The processing module **50** may decode messages from the MMW inbound signals **148** such that the decoded message include one or more of a status request, a records request, a sensor data request, a processed data request, a position request, a command, and/or a request for MRI echo signal data.

At step **572**, the processing module **50** determines whether to provide records in response to the records request based in part on the decoded message. The determination may be based on a comparison of the records request to the capabilities of the bio-medical unit **10**. Note that records may include patient history, medications, alerts, allergies, personal information, contact information, age, weight, test results, etc.

At step **576**, the method branches to step **578** when the processing module **50** determines to provide records. At step **576**, the method continues to step **576** when the processing module **50** determines to not provide records. At step **576**, the processing module **50** performs other instructions contained in the message. The method ends.

At step **578**, the processing module **50** determines when it is time to transmit. The determination may be based on a timer, a command, available power, a priority indicator, a timeslot, and/or interference indicator. At step **580**, the method branches back to step **578** when the processing module **50** determines it is not time to transmit. At step **580**, the method continues to step **582** when the processing module **50** determines it is time to transmit.

At step **582**, the processing module **50** determines the format to provide records. The format determination may be based on one or more of a memory lookup, a command, available power, the type of records requested, an access ID of the requester, a priority indicator, a level of detail indicator, and/or a freshness indicator. Note that the format may include records format as stored in the bio-medical unit memory (e.g., all or a portion of the records) or a uniform resource locator (URL) to link to another memory in one or more of the service provider's computer, the database, and/or the server.

At step **584**, the method branches to step **588** when the processing module **50** determines the format to provide records is the URL format. At step **584**, the method continues to step **586** where the processing module **50** prepares the records format response message based on records information retrieved from the bio-medical unit memory **52**. The method branches to step **590**.

At step **588**, the processing module prepares the URL format response message based on retrieving the URL from the bio-medical unit memory **52**. At step **590**, the processing module **50** transmits the response message encoded as MMW outbound signals **150**. For example, the bio-medical unit **10** transmits the response message via a second wireless communications medium including one or more of infrared signals, ultrasonic signals, visible light signals, audible sound signals, and/or EM signals via one or more of the functional modules.

FIG. **29** is a flowchart illustrating the coordination of bio-medical unit task execution where the processing module **50** determines and executes tasks with at least one other bio-medical unit **10**. The method begins at step **592** where the processing module **50** determines if communication is allowed. The determination may be based on one or more of a timer, a command, available power, a priority indicator, an MRI sequence, and/or interference indicator.

At step **594**, the method branches back to step **592** when the processing module **50** determines that communication is not allowed. At step **594**, the method continues to step **596** when the processing module **50** determines that communication is allowed. At step **596**, the processing module **50** directs the communication module **48** to communicate with a plurality of bio-medical units **10** utilizing RF and/or MMW inbound and/or outbound signals. The processing module **50** may decode messages from the RF and/or MMW inbound and/or outbound signals inbound signals. At step **598**, the processing module **50** determines if communications with the plurality of bio-medical units **10** is successful based in part on the decoded messages.

At step **600**, the method branches back to step **592** when the processing module determines that communications with the plurality of bio-medical units **10** is not successful. Note that forming a network with the other bio-medical units **10** may be required to enable joint actions. At step **600**, the method continues to step **602** when the processing module **50** determines that communications with the plurality of bio-medical units **10** is successful.

At step **602**, the processing module **50** determines the task and task requirements. The task determination may be based on one or more of a command from a parent bio-medical unit **10**, external communications, a preprogrammed list, and/or in response to sensor data. The task requirements determination may be based on one or more of the task, a command from a parent bio-medical unit **10**, external communications, a preprogrammed list, and/or in response to sensor data. Note that the task may include actions including one or more of drilling, moving, sawing, jumping, spreading, sensing, lighting, pinging, testing, and/or administering medication.

At step **604**, the processing module **50** determines the control mode based on one or more of a command from a parent bio-medical unit **10**, external communications, a preprogrammed list, and/or in response to sensor data. Note that the control mode may include autonomous, parent (bio-medical unit), server, and/or peer.

At step **606**, the processing module **50** determines if task execution criteria are met based on sensor data, communication with other bio-medical units **10**, a command, a status indicator, a safety indicator, a stop indicator, and/or location information. Note that the task execution criteria may include one or more of safety checks, position information of the bio-medical unit **10**, position information of other bio-medical units **10**, and/or sensor data thresholds.

At step **608**, the method branches back to step **606** when the processing module **50** determines that the task execution criteria are not met. At step **608**, the method continues to step **610** when the processing module **50** determines that the task execution criteria are met. At step **610**, the processing module **50** executes a task element. A task element may include a portion or step of the overall task. For example, move one centimeter of a task to move three centimeters.

At step **612**, the processing module **50** determines if task exit criteria are met based on a task element checklist status, sensor data, communication with other bio-medical units **10**, a command, a status indicator, a safety indicator, a stop indicator, and/or location information. Note that the task exit criteria define successful completion of the task.

At step **614**, the method branches back to step **592** when the processing module **50** determines that the task exit criteria are met. In other words, the plurality of bio-medical units **10** is done with the current task and is ready for the next task. At step **614**, the method continues to step **616** when the processing module **50** determines that the task exit criteria are not met.

At step **616**, the processing module **50** directs the communication module **48** to communicate with the plurality of bio-medical units **10** utilizing RF and/or MMW inbound and/or outbound. The processing module **50** may decode messages from the RF and/or MMW inbound and/or outbound signals inbound signals. Note that the messages may include information in regards to task modifications (e.g., course corrections). At step **618**, the processing module **50** determines if communications with the plurality of bio-medical units **10** is successful based in part on the decoded messages.

At step **620**, the method branches back to step **592** when the processing module determines that communications with the plurality of bio-medical units is not successful (e.g., to potentially restart). Note that maintaining the network with the other bio-medical unit may be required to enable joint actions. At step **620**, the method continues to step **622** when the processing module determines that communications with the plurality of bio-medical units is successful.

At step **622**, the processing module **50** determines task modifications. The task modifications may be based on one or more of a command from a parent bio-medical unit **10**, and/or external communications. The task modifications determination may be based on one or more of the task, a command from a parent bio-medical unit **10**, external communications, a preprogrammed list, and/or in response to sensor data. The method branches back to step **606** to attempt to complete the current task.

As may be used herein, the terms “substantially” and “approximately” provides an industry-accepted tolerance for its corresponding term and/or relativity between items. Such an industry-accepted tolerance ranges from less than one percent to fifty percent and corresponds to, but is not limited to, component values, integrated circuit process variations, temperature variations, rise and fall times, and/or thermal noise. Such relativity between items ranges from a difference of a few percent to magnitude differences. As may also be used herein, the term(s) “coupled to” and/or “coupling” and/or includes direct coupling between items and/or indirect coupling between items via an intervening item (e.g., an item includes, but is not limited to, a component, an element, a circuit, and/or a module) where, for indirect coupling, the intervening item does not modify the information of a signal but may adjust its current level, voltage level, and/or power level. As may further be used herein, inferred coupling (i.e., where one element is coupled to another element by inference) includes direct and indirect coupling between two items in the same manner as “coupled to”. As may even further be used herein, the term “operable to” indicates that an item includes one or more of power connections, input(s), output (s), etc., to perform one or more its corresponding functions and may further include inferred coupling to one or more other items. As may still further be used herein, the term “associated with”, includes direct and/or indirect coupling of separate items and/or one item being embedded within another item. As may be used herein, the term “compares favorably”, indicates that a comparison between two or more items, signals, etc., provides a desired relationship. For example, when the desired relationship is that signal **1** has a greater magnitude than signal **2**, a favorable comparison may be achieved when the magnitude of signal **1** is greater than that of signal **2** or when the magnitude of signal **2** is less than that of signal **1**.

The present invention has also been described above with the aid of method steps illustrating the performance of specified functions and relationships thereof. The boundaries and sequence of these functional building blocks and method steps have been arbitrarily defined herein for convenience of description. Alternate boundaries and sequences can be defined so long as the specified functions and relationships are appropriately performed. Any such alternate boundaries or sequences are thus within the scope and spirit of the claimed invention.

The present invention has been described above with the aid of functional building blocks illustrating the performance of certain significant functions. The boundaries of these functional building blocks have been arbitrarily defined for convenience of description. Alternate boundaries could be defined as long as the certain significant functions are appropriately performed. Similarly, flow diagram blocks may also have been arbitrarily defined herein to illustrate certain significant functionality. To the extent used, the flow diagram block boundaries and sequence could have been defined otherwise and still perform the certain significant functionality. Such alternate definitions of both functional building blocks and flow diagram blocks and sequences are thus within the scope and spirit of the claimed invention. One of average skill in the art will also recognize that the functional building blocks, and other illustrative blocks, modules and components herein, can be implemented as illustrated or by discrete components, application specific integrated circuits, processors executing appropriate software and the like or any combination thereof.

What is claimed is:

1. A bio-medical unit for implanting in or on a body, the bio-medical unit comprising:

a power harvesting module operable to generate a power supply voltage from an electromagnetic signal transmitted from a signal source external to the bio-medical unit; and

a radio frequency (RF) communication module coupled to the power harvesting module and powered by the power supply voltage, in which the RF communication module is operable to:

interpret a pattern of the electromagnetic signal to determine null intervals in the electromagnetic signal;

receive a downstream RF communication signal from a communication device external to the bio-medical unit during at least one of the null intervals;

convert the downstream RF communication signal into downstream information for processing by the bio-medical unit;

convert upstream information of the bio-medical unit into an upstream RF communication signal; and

transmit the upstream RF communication signal to the communication device during one or more of the null intervals.

2. The bio-medical unit of claim **1**, wherein the electromagnetic signal comprises: a magnetic resonance imaging (MRI) signal having at least one of a radio frequency (RF) component and a gradient magnetic field component.

3. The bio-medical unit of claim **1** further comprises:

a processing module operable to:

convert the downstream information into at least one of downstream data and a downstream instruction for processing; and

convert at least one of an upstream instruction and upstream data into the upstream information of the bio-medical unit; and

a function module operable to perform at least one of:

generate the at least one of an upstream instruction and the upstream data; and

perform a bio-medical function based on the at least one of downstream data and the downstream instruction.

4. The bio-medical unit of claim **3** further comprises:

the processing module further operable to:

interpret the downstream instruction to determine whether the functional module is to be enabled to receive an echo signal, wherein the echo signal corresponds to the electromagnetic signal reflecting off of an internal portion of the body; and

enable the functional module when the functional module is to be enabled to receive an echo signal;

the functional module is further operable to:

receive the echo signal; and

convert the echo signal into echo information; and

the processing module is further operable to convert the echo information into the upstream information.

5. The bio-medical unit of claim **4**, wherein the processing module is further operable to perform at least one of:

passively convert the echo information into the upstream information; and

actively convert the echo information into the upstream information.

6. The bio-medical unit of claim **1**, wherein the RF communication module further functions to:

determine whether an electromagnetic function is active;

when the electromagnetic function is active, determine the predetermined pattern;

determine the at least one null interval for receiving the downstream RF communication signal;

determine whether the upstream RF communication is pending; and

determine the one or more of the null intervals, when the upstream RF communication is pending.

7. A bio-medical unit for implanting in or on a body, comprising:

a power harvesting module operable to generate a supply voltage from an electromagnetic signal transmitted from a signal source external to the bio-medical unit; and

a communication module coupled to the power harvesting module and powered by the supply voltage that is generated by the power harvesting module, in which the communication module is operable to determine null intervals in the electromagnetic signal and communicate with a communication device external to the bio-medical unit during one or more of the null intervals.

8. The bio-medical unit of claim **7**, wherein the electromagnetic signal is a magnetic resonance imaging (MRI) signal.

9. The bio-medical unit of claim **7**, wherein the communication module communicates with the communication device utilizing radio frequency (RF) signals.

10. The bio-medical unit of claim **7**, wherein the communication module communicates with the communication device utilizing millimeter wave (MMW) signals.

11. The bio-medical unit of claim **7**, wherein the communication module communicates with the communication device utilizing ultrasound signals.

12. The bio-medical unit of claim **7**, wherein the communication module communicates with the communication device utilizing optical signals.

13. The bio-medical unit of claim **7**, wherein the signal source and the communication device are located within a magnetic resonance imaging (MRI) device.

14. The bio-medical unit of claim **7**, wherein the signal source is within a magnetic resonance imaging (MRI) device and the communication device is located at a different location than the MRI device.

15. The bio-medical unit of claim **7** further comprising a processing module coupled to be powered by the supply voltage from the power harvesting module, and the processing module coupled to the communication module to process communication signals received from the communication device.

16. The bio-medical unit of claim **7** further comprising a functional module coupled to be powered by the supply voltage from the power harvesting module, and the functional module coupled to the communication module to respond to communication signals received from the communication device to perform a medical function.

17. The bio-medical unit of claim **7** further comprising:

a processing module coupled to be powered by the supply voltage from the power harvesting module, and the processing module coupled to the communication module to process communication signals received from the communication device; and

a functional module coupled to be powered by the supply voltage from the power harvesting module, and the functional module coupled to the processing module to respond to communication signals processed by the processing module to perform a medical function.

18. A method of communicating with a bio-medical unit implanted in or on a body, comprising:

subjecting the bio-medical unit to an electromagnetic signal transmitted from a signal source external to the bio-medical unit;

utilizing a power harvesting module of the bio-medical unit to harvest energy from the electromagnetic signal to generate a supply voltage;

powering a communication module of the bio-medical unit by use of the supply voltage;

interpreting a pattern of the electromagnetic signal to determine null intervals in the electromagnetic signal;

receiving, in the communication module, a downstream communication signal from a communication device external to the bio-medical unit during at least one of the null intervals; and

processing the downstream communication signal to perform a medical function of the bio-medical unit.

19. The method of claim **18**, wherein when subjecting the bio-medical unit to the electromagnetic signal, the electromagnetic signal is a magnetic resonance imaging (MRI) signal.

20. The method of claim **18** further comprising, transmitting an upstream communication signal from the communication module to the communication device during at least one of the null intervals.

* * * * *

US009021277B2

(12) **United States Patent**
Shearer et al.

(10) **Patent No.: US 9,021,277 B2**
(45) **Date of Patent: Apr. 28, 2015**

(54) **POWERING DEVICES USING RF ENERGY HARVESTING**

(71) Applicant: **Powercast Corporation**, Ligonier, PA (US)

(72) Inventors: **John G. Shearer**, Ligonier, PA (US); **Charles E. Greene**, Pittsburgh, PA (US); **Daniel W. Harrist**, Carnegie, PA (US)

(73) Assignee: **Powercast Corporation**, Ligonier, PA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **14/143,334**

(22) Filed: **Dec. 30, 2013**

(65) **Prior Publication Data**

US 2014/0111032 A1 Apr. 24, 2014

Related U.S. Application Data

(60) Continuation of application No. 12/499,618, filed on Jul. 8, 2009, now Pat. No. 8,621,245, which is a division of application No. 11/447,412, filed on Jun. 6, 2006, now abandoned.

(60) Provisional application No. 60/688,587, filed on Jun. 8, 2005.

(51) **Int. Cl.**

G06F 1/26	(2006.01)
H02J 17/00	(2006.01)
G06K 19/07	(2006.01)
H04B 1/16	(2006.01)

(52) **U.S. Cl.**
CPC ***H02J 17/00*** (2013.01); ***G06K 19/0707*** (2013.01); ***H04B 1/1607*** (2013.01)

(58) **Field of Classification Search**
CPC G06F 1/26; G06K 19/0707
USPC 713/300
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,297,618	B2	10/2001	Emori et al.
6,447,448	B1	9/2002	Ishikawa et al.
6,615,074	B2	9/2003	Mickle et al.
6,784,358	B2	8/2004	Kukulka
6,835,501	B2	12/2004	Morishita et al.
6,836,095	B2	12/2004	Fogg
6,858,970	B2	2/2005	Malkin et al.
6,894,467	B2	5/2005	Pons et al.
7,256,695	B2	8/2007	Hamel et al.
7,268,517	B2	9/2007	Rahmel et al.
7,561,866	B2	7/2009	Oliver et al.
7,956,572	B2	6/2011	Zane et al.
8,621,245	B2	12/2013	Shearer et al.

(Continued)

OTHER PUBLICATIONS

Non-Final Office Action for U.S. Appl. No. 11/447,412, mailed Jun. 3, 2008.

(Continued)

Primary Examiner — Albert Wang

(57) **ABSTRACT**

Disclosed is an apparatus for an application including a core device for the application. The apparatus includes a power (preferably RF energy) harvester connected to the core device to power the core device. Also disclosed is a method for an application. The method includes the steps of converting RF energy into usable energy. There is the step of powering the core device with the usable energy.

20 Claims, 90 Drawing Sheets

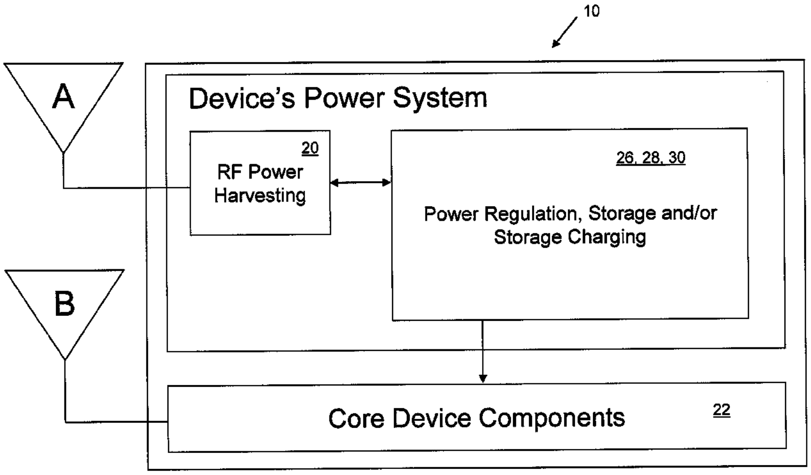

(56) References Cited

U.S. PATENT DOCUMENTS

2002/0072784	A1	6/2002	Sheppard, Jr. et al.	
2004/0150529	A1	8/2004	Benoit et al.	
2004/0259604	A1	12/2004	Mickle et al.	
2005/0240778	A1 *	10/2005	Saito	713/186
2006/0281435	A1	12/2006	Shearer et al.	

OTHER PUBLICATIONS

Final Office Action for U.S. Appl. No. 11/447,412, mailed Jan. 8, 2009.

Office Action for U.S. Appl. No. 12/499,618, mailed Jun. 22, 2011.

Office Action for U.S. Appl. No. 12/499,618, mailed Apr. 19, 2012.

Final Office Action for U.S. Appl. No. 12/499,618, mailed Feb. 21, 2013.

International Search Report and Written Opinion for International Application No. PCT/US06/21940, mailed Sep. 25, 2007, 9 pages.

Examination Report for Australian Application No. 2006255054, mailed Nov. 27, 2009, 2 pages.

Office Action for Chinese Application No. 200680020092.3, received on Sep. 7, 2009, 5 pages.

Examination Report for Singapore Application No. 200718208-2, mailed Sep. 28, 2009, 6 pages.

Office Action for Mexican Patent Application No. MX/a/2007/015229, received on Mar. 4, 2010, 2 pages.

Examination Report for Australian Application No. 2006255054, mailed Jul. 7, 2011, 3 pages.

European Search Report for European Patent Application No. 06772310.6, mailed Jul. 27, 2011; 7 pages.

* cited by examiner

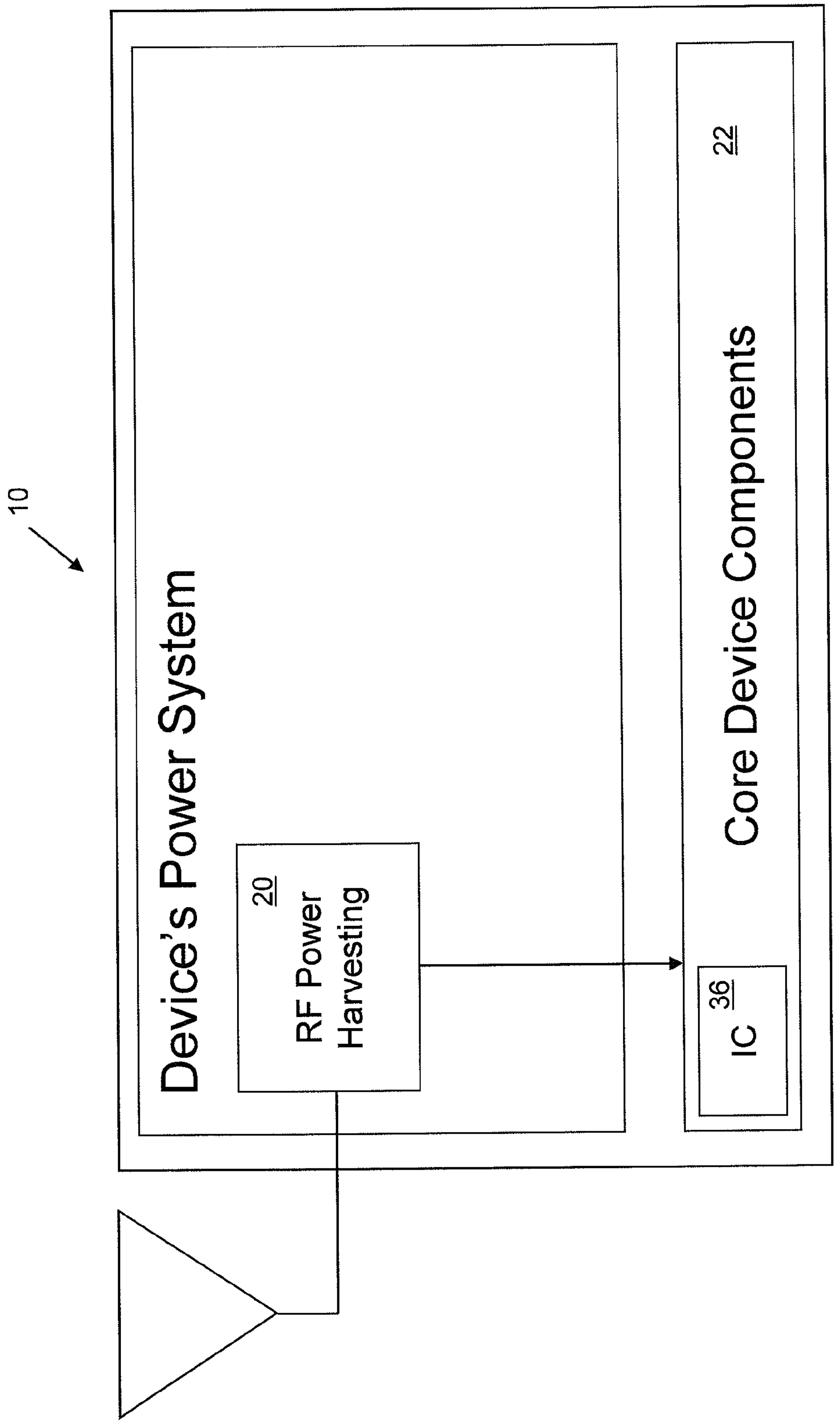
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
IC 36

Fig. 1

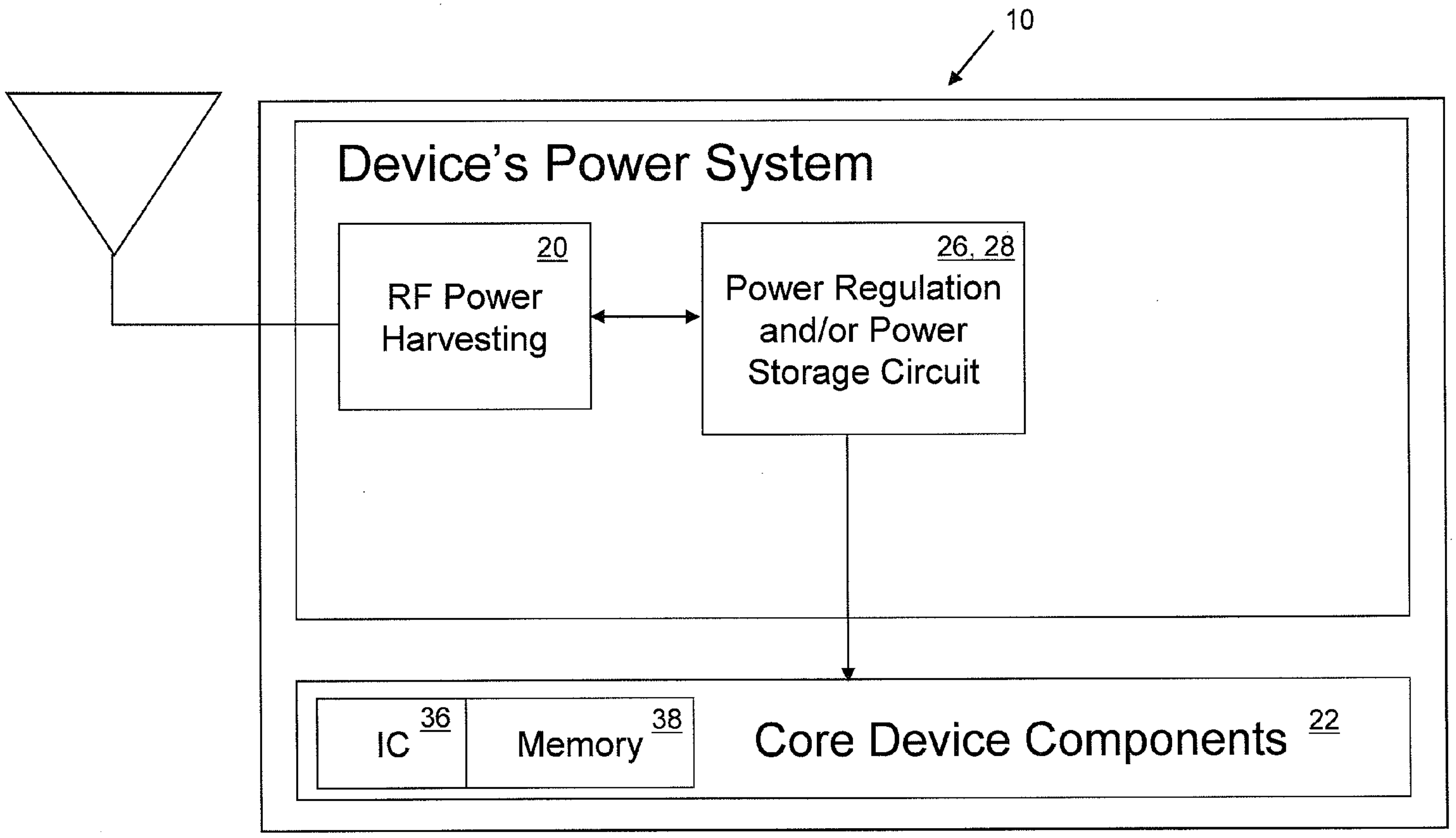
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
36
IC
38
Memory
Core Device Components
22

Fig. 2

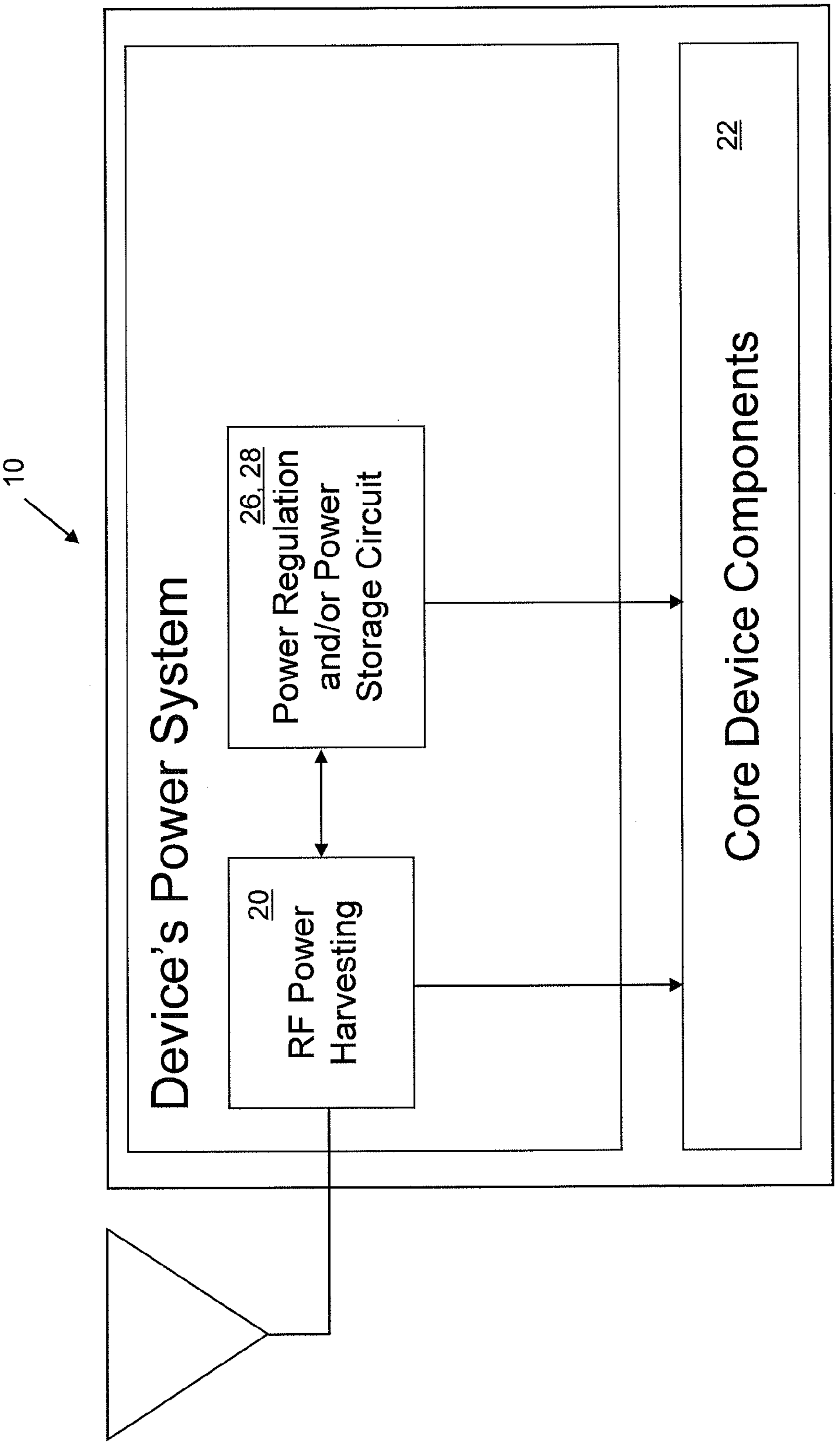
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Core Device Components
22

Fig. 3

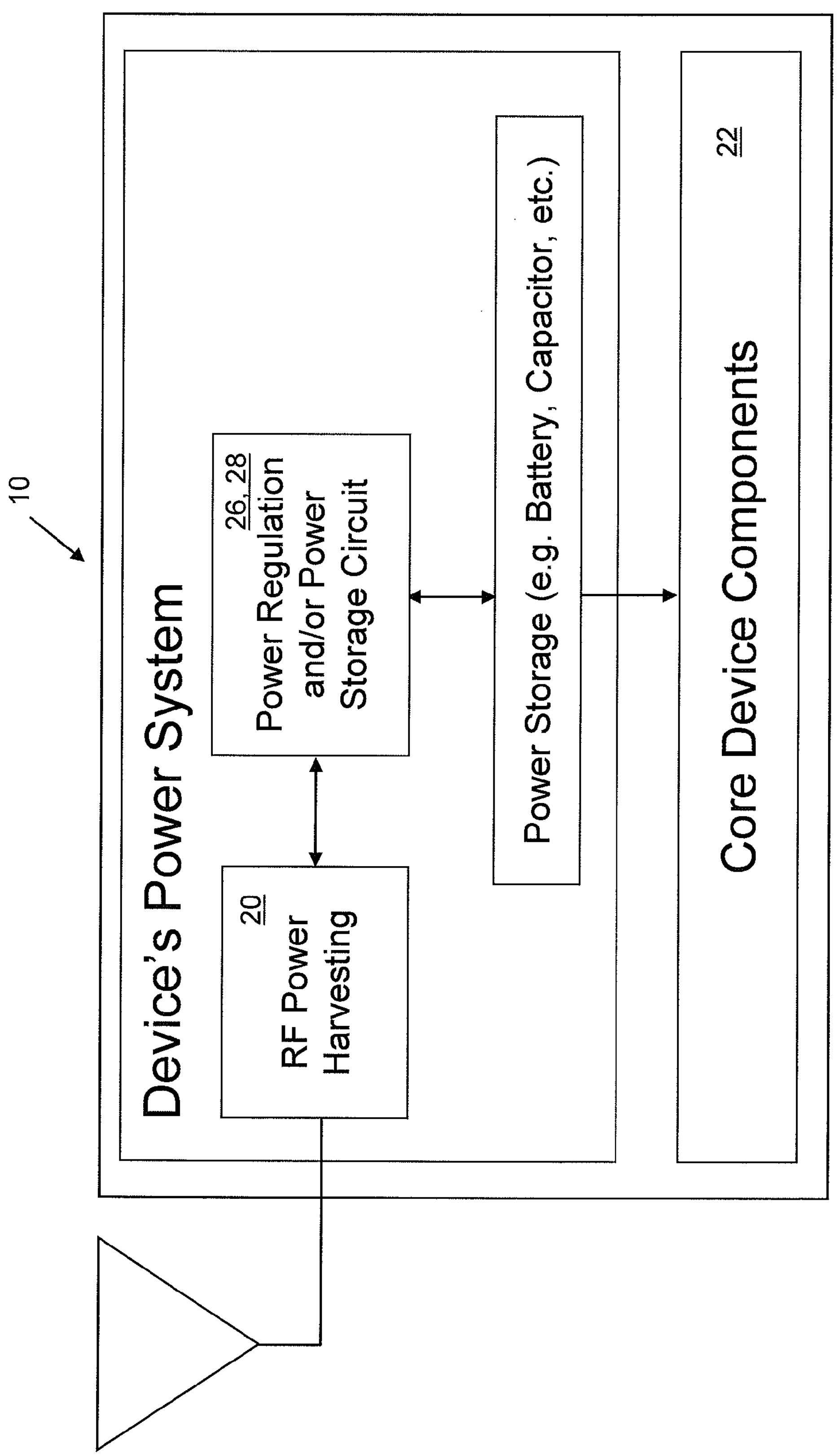
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 4

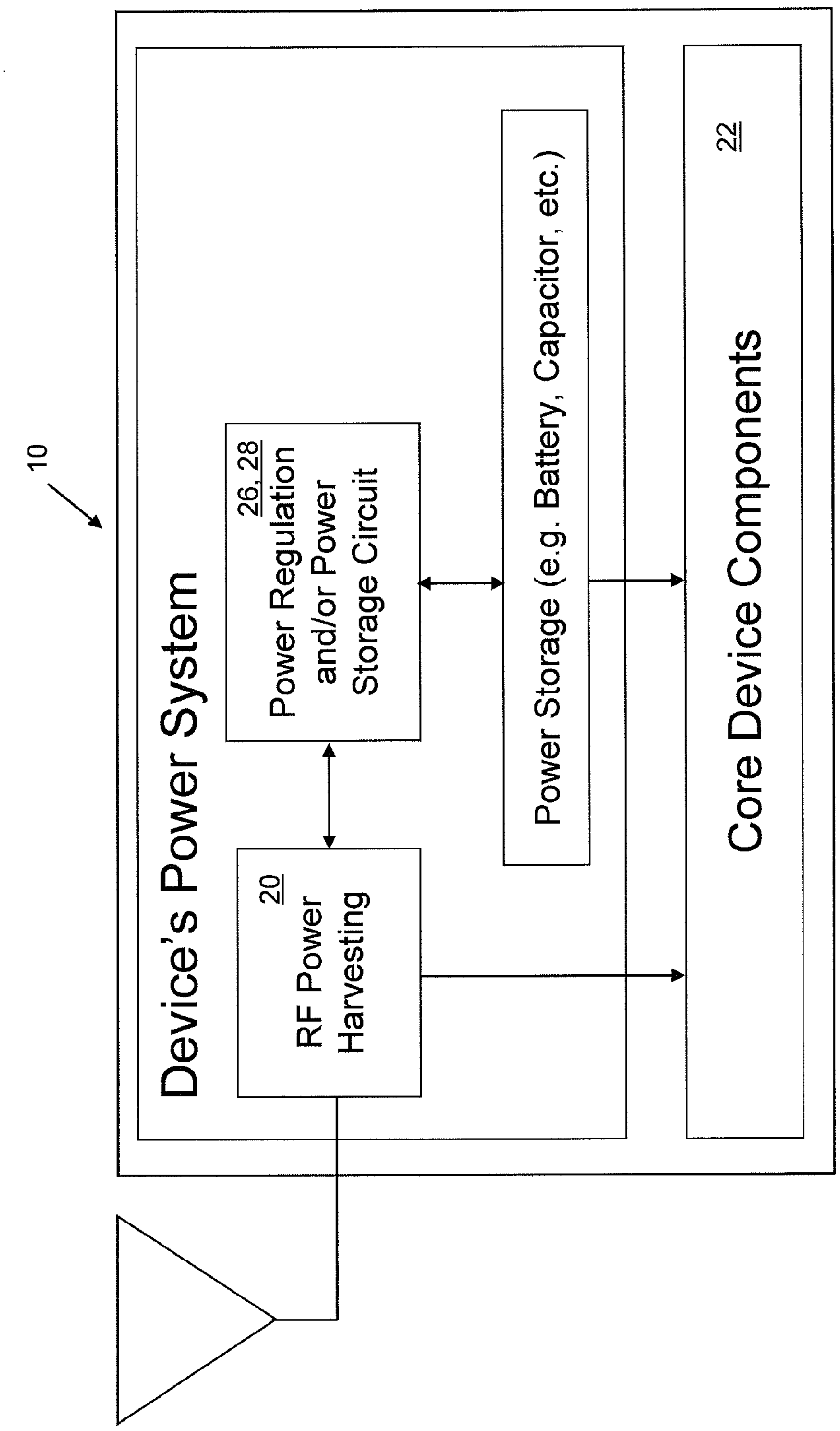
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 5

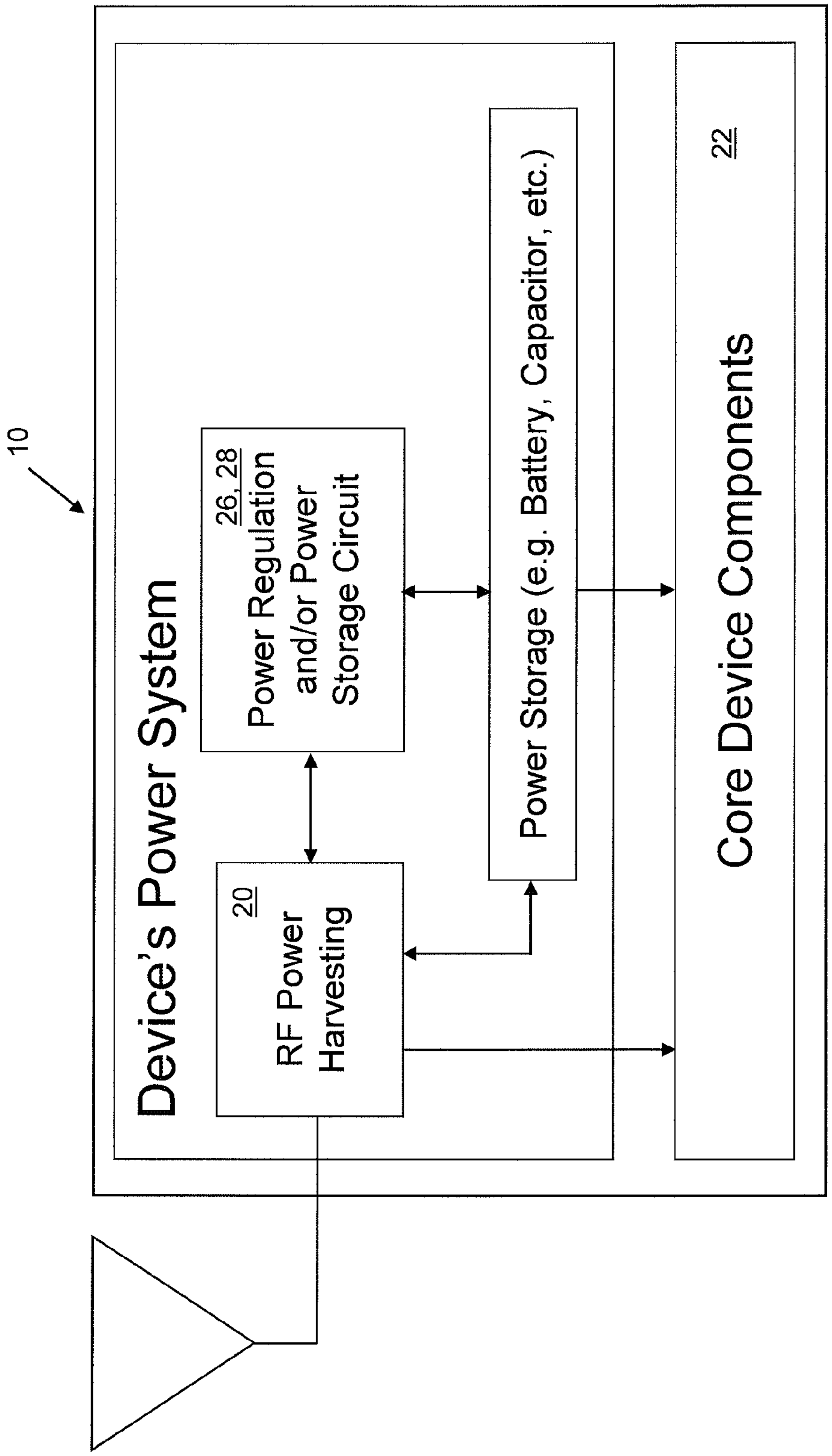
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 6

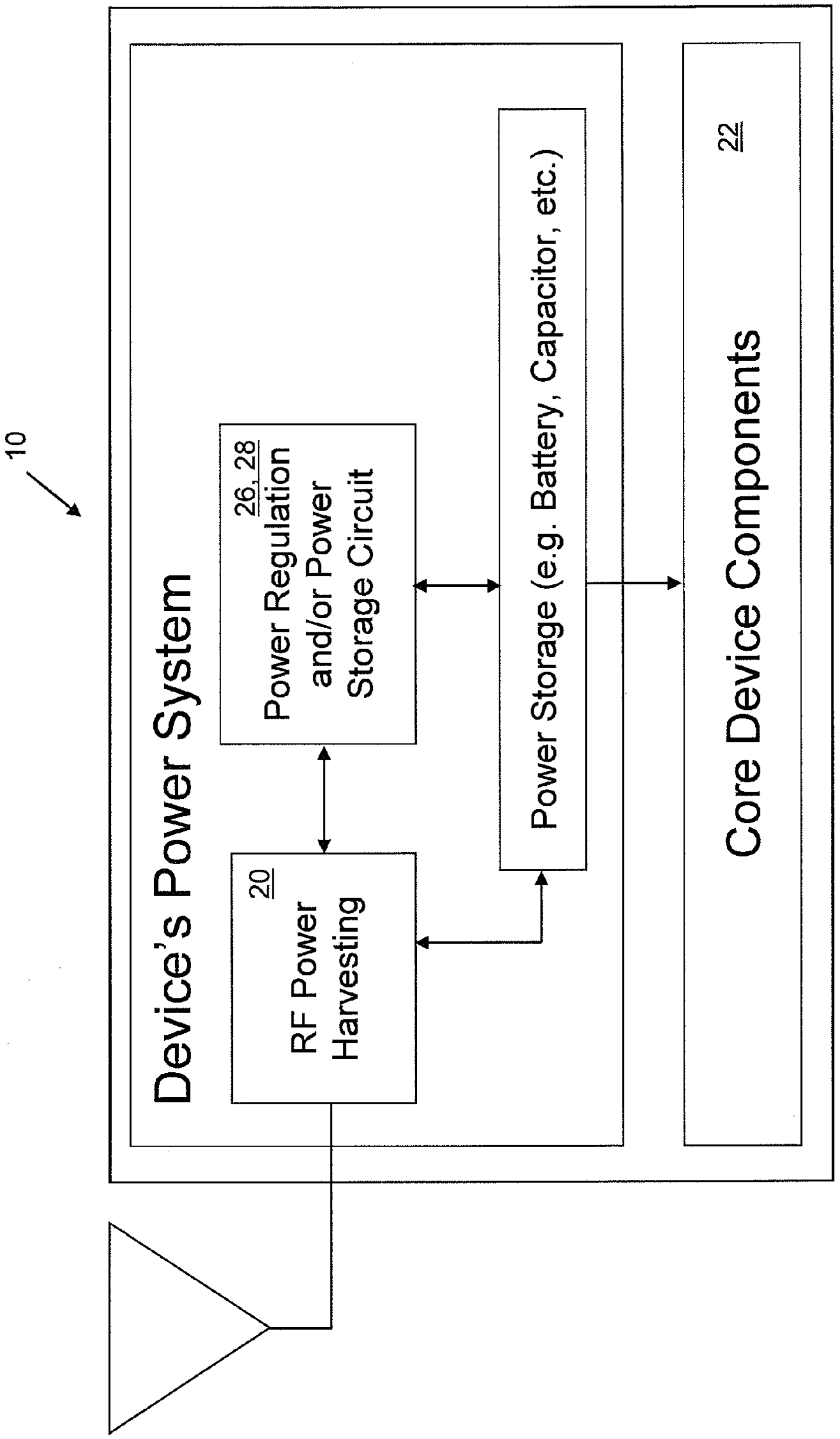
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 7

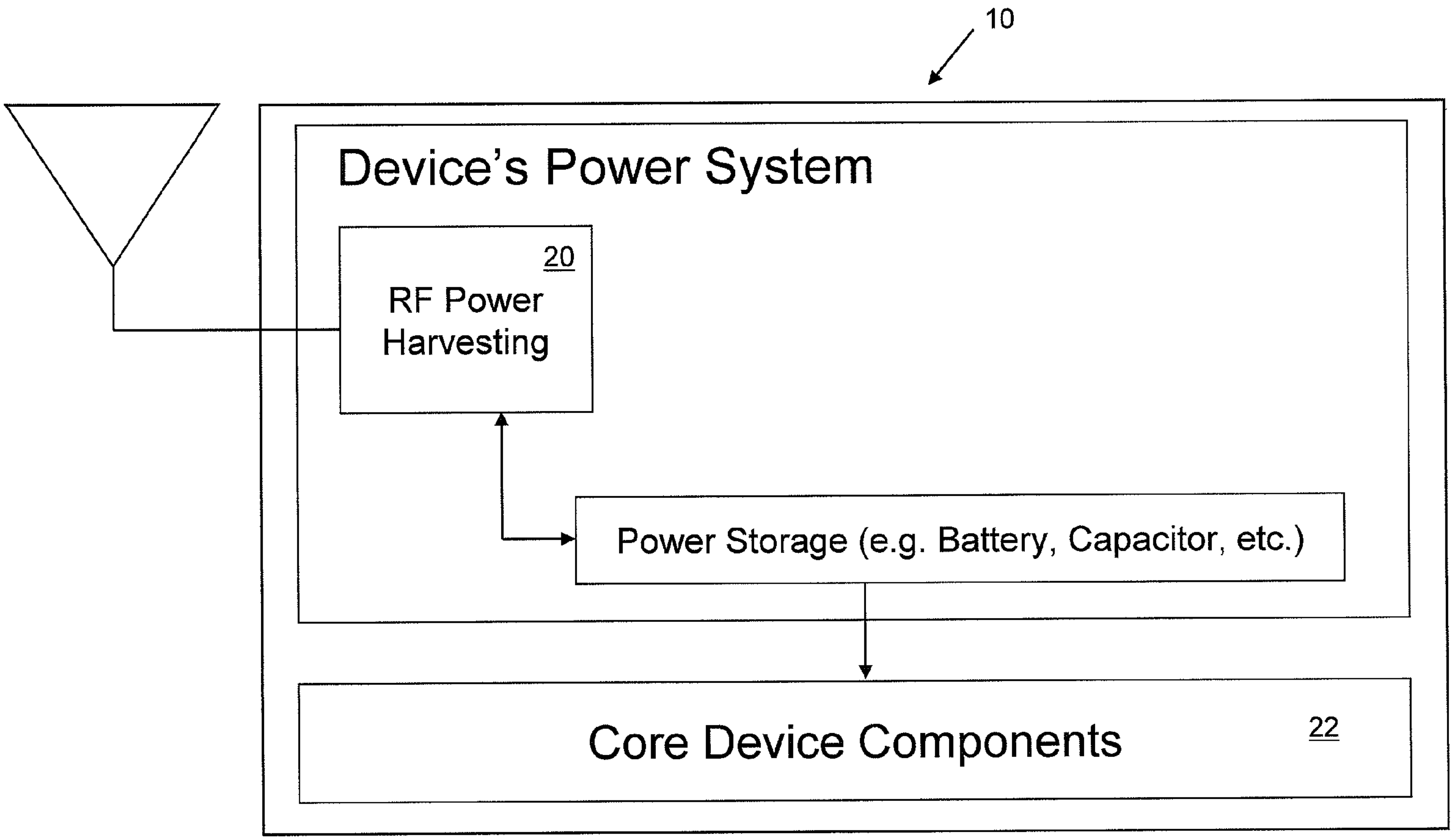
10
Device's Power System
20
RF Power Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 8

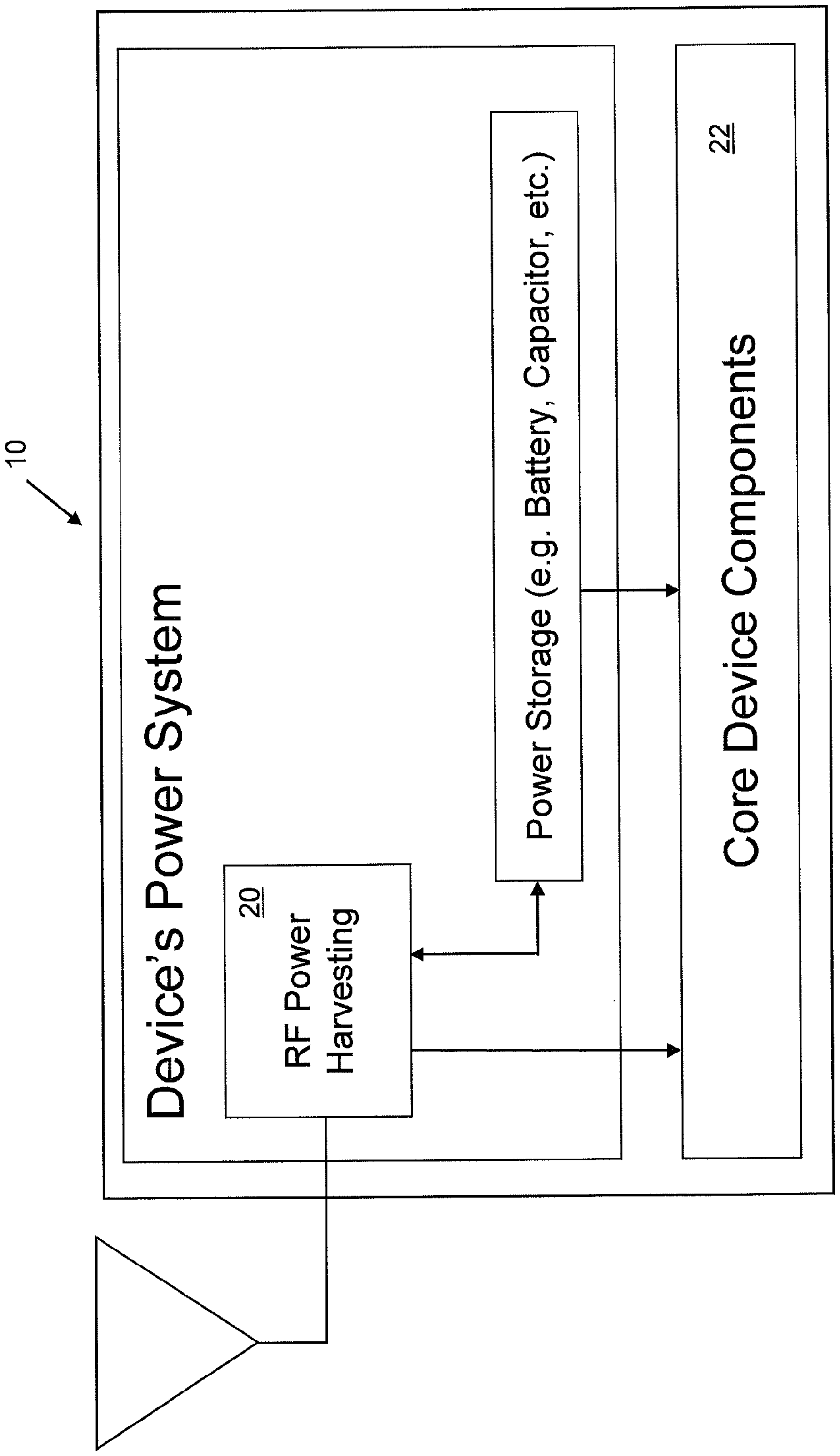
10
Device's Power System
20
RF Power Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 9

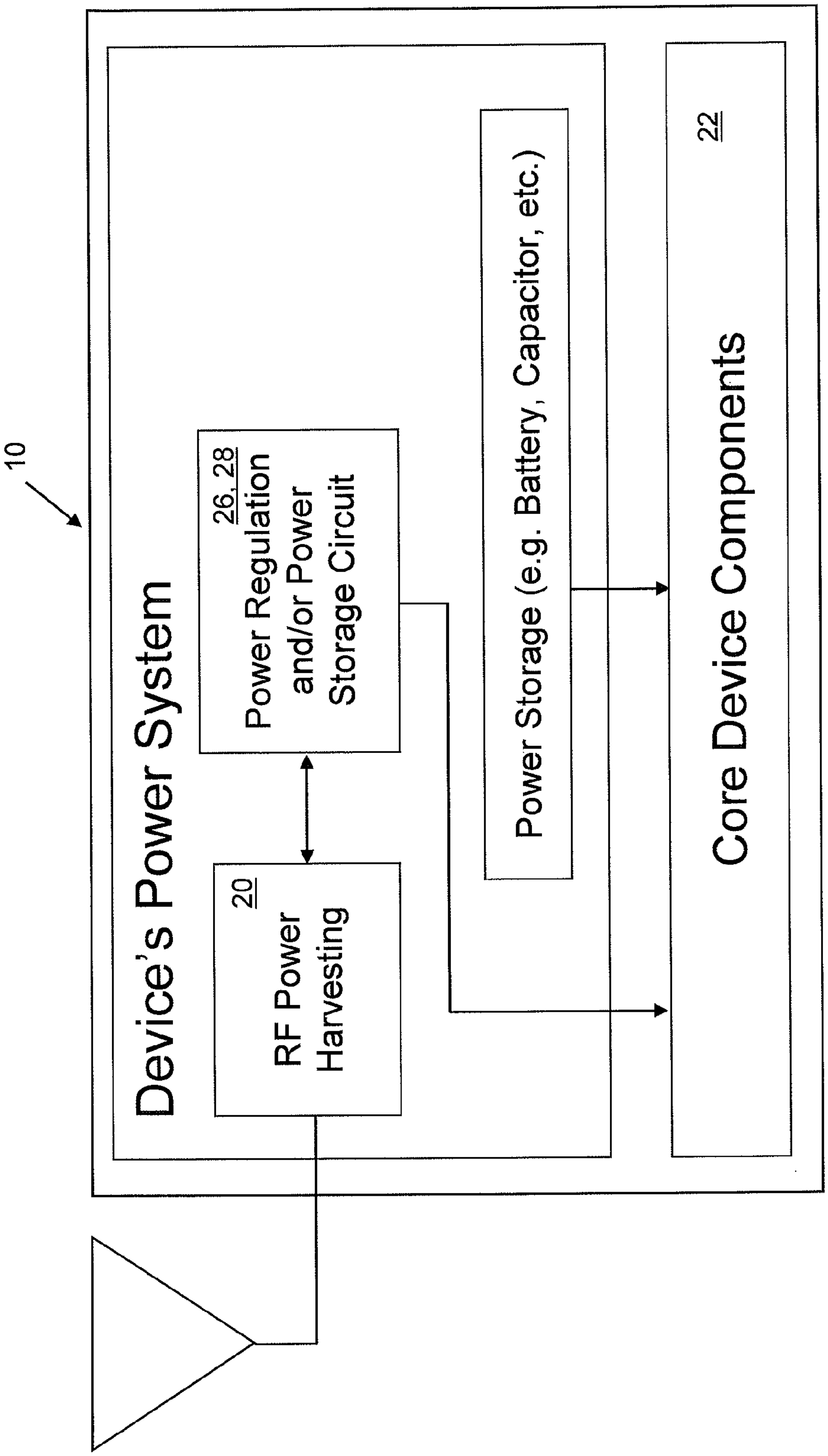
10
Device's Power System
Power Regulation and/or Power Storage Circuit
26, 28
RF Power Harvesting
20
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 10

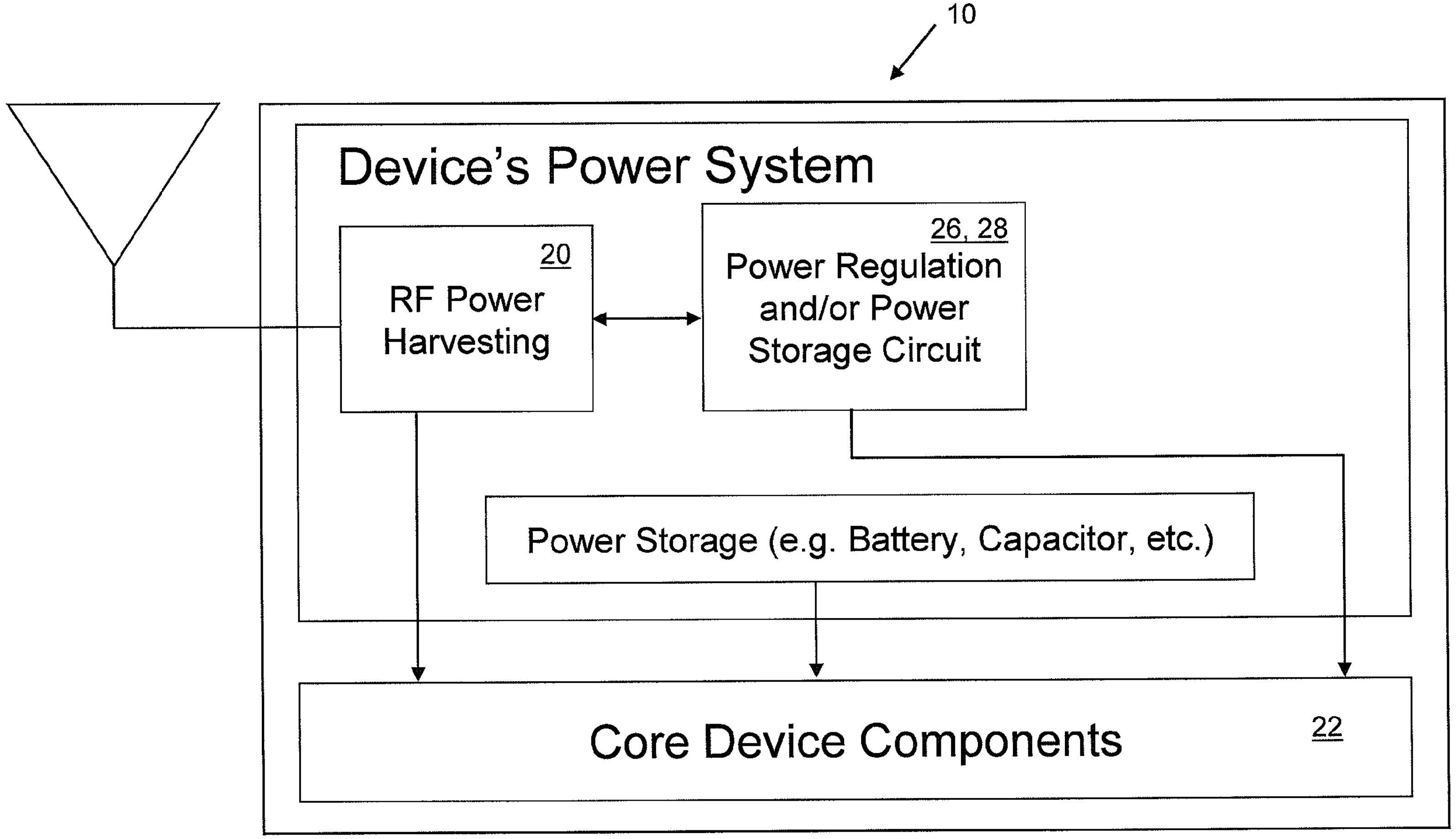
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 11

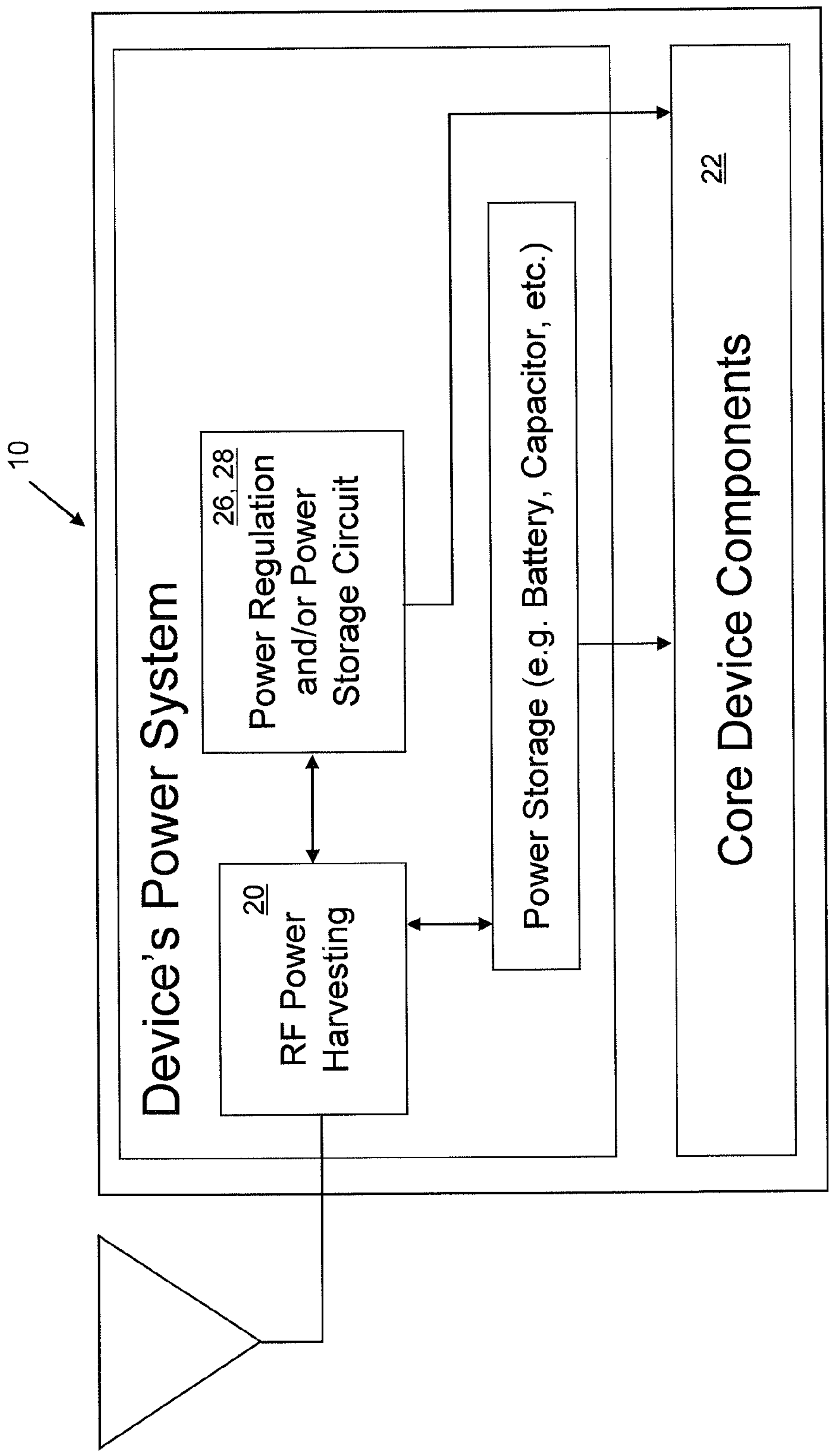
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 12

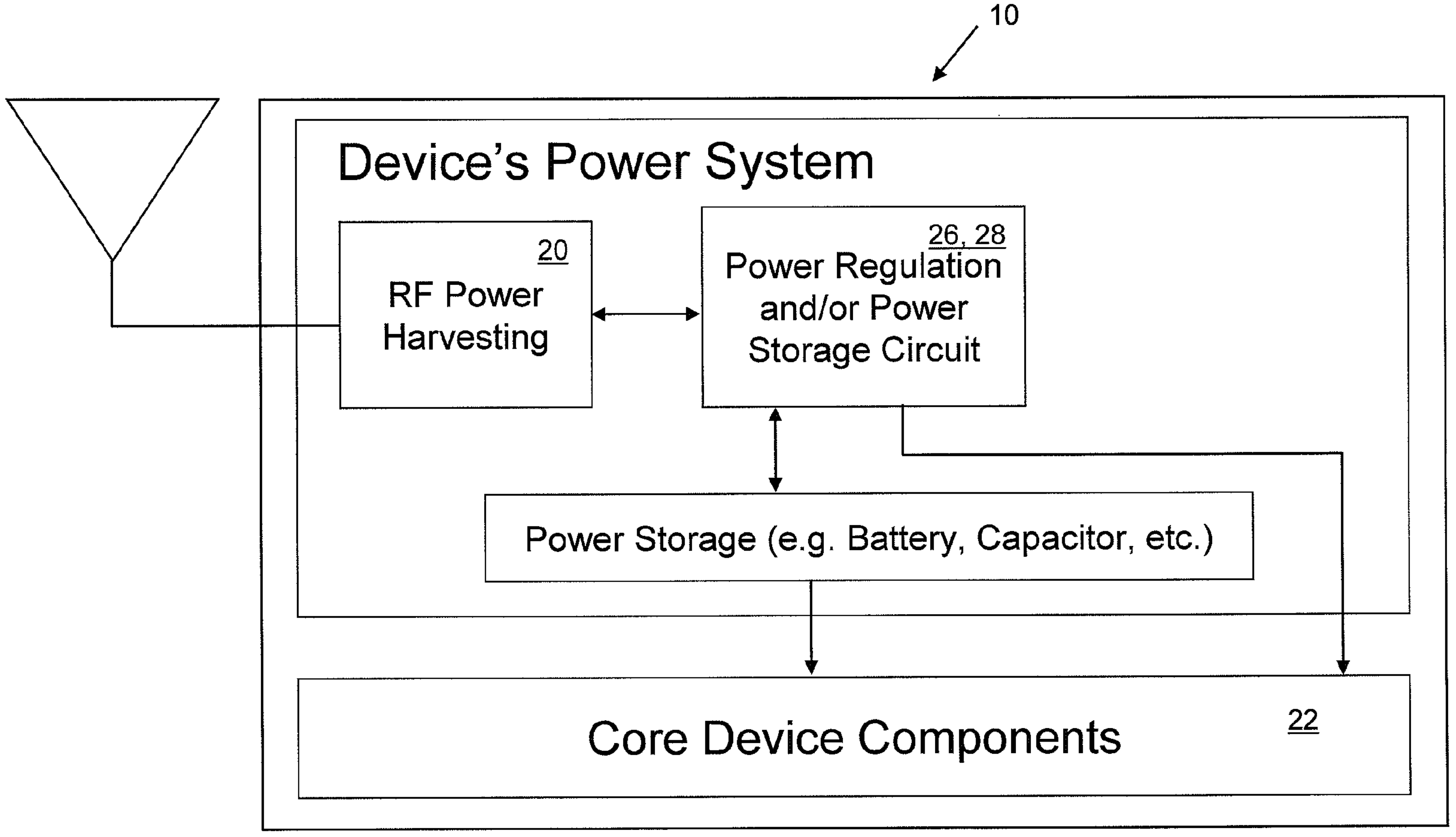
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 13

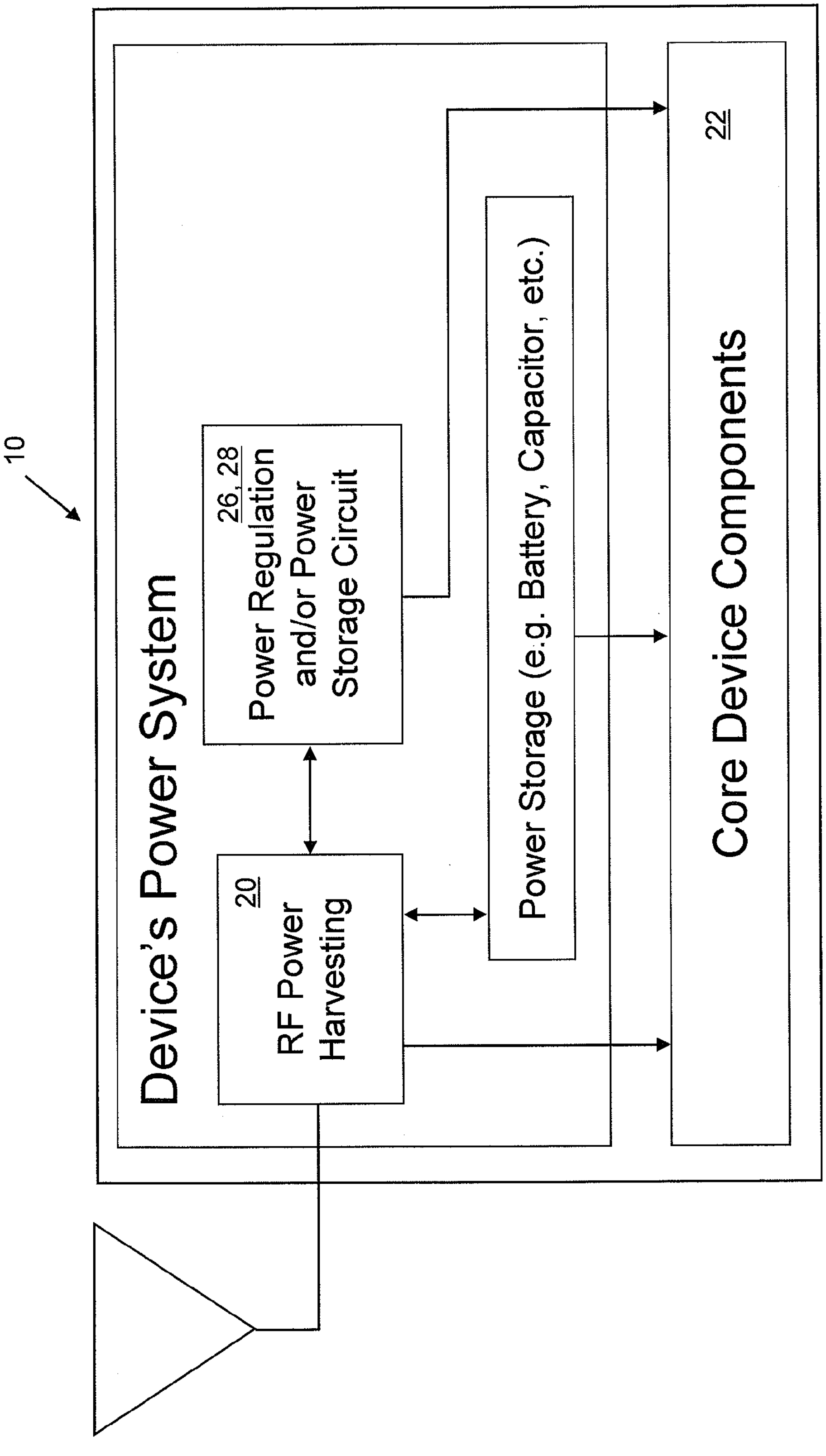
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 14

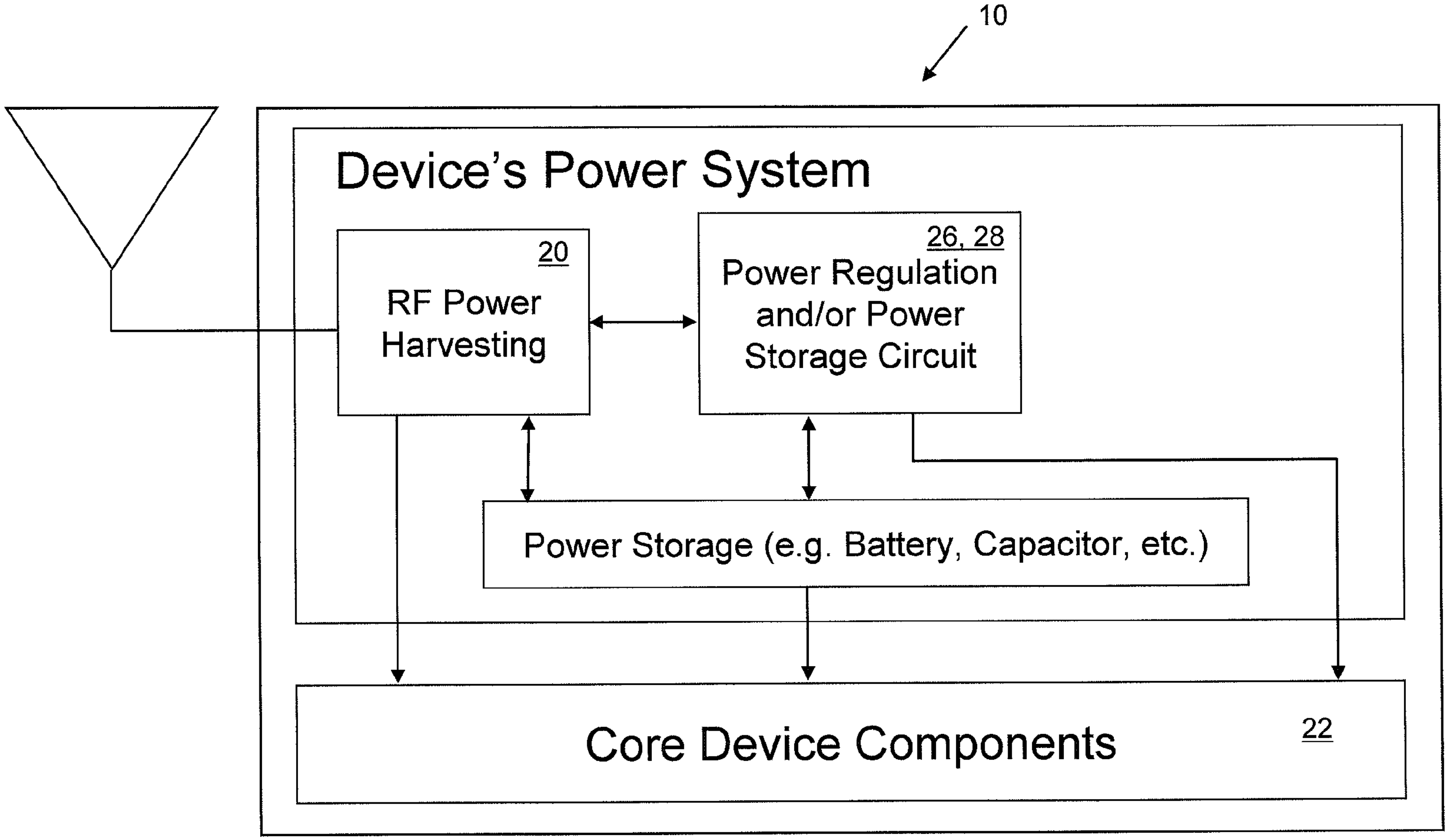
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 15

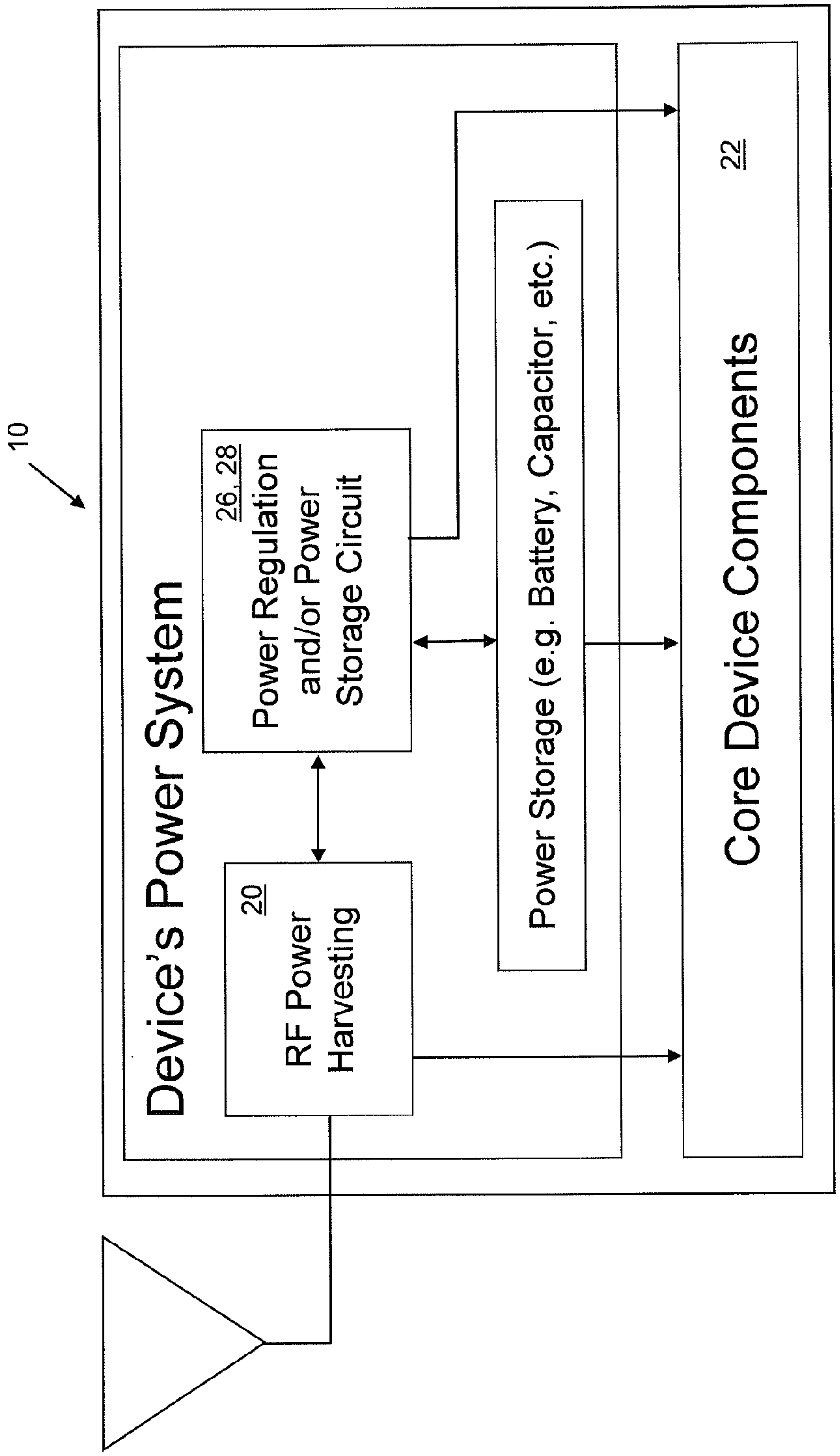
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 16

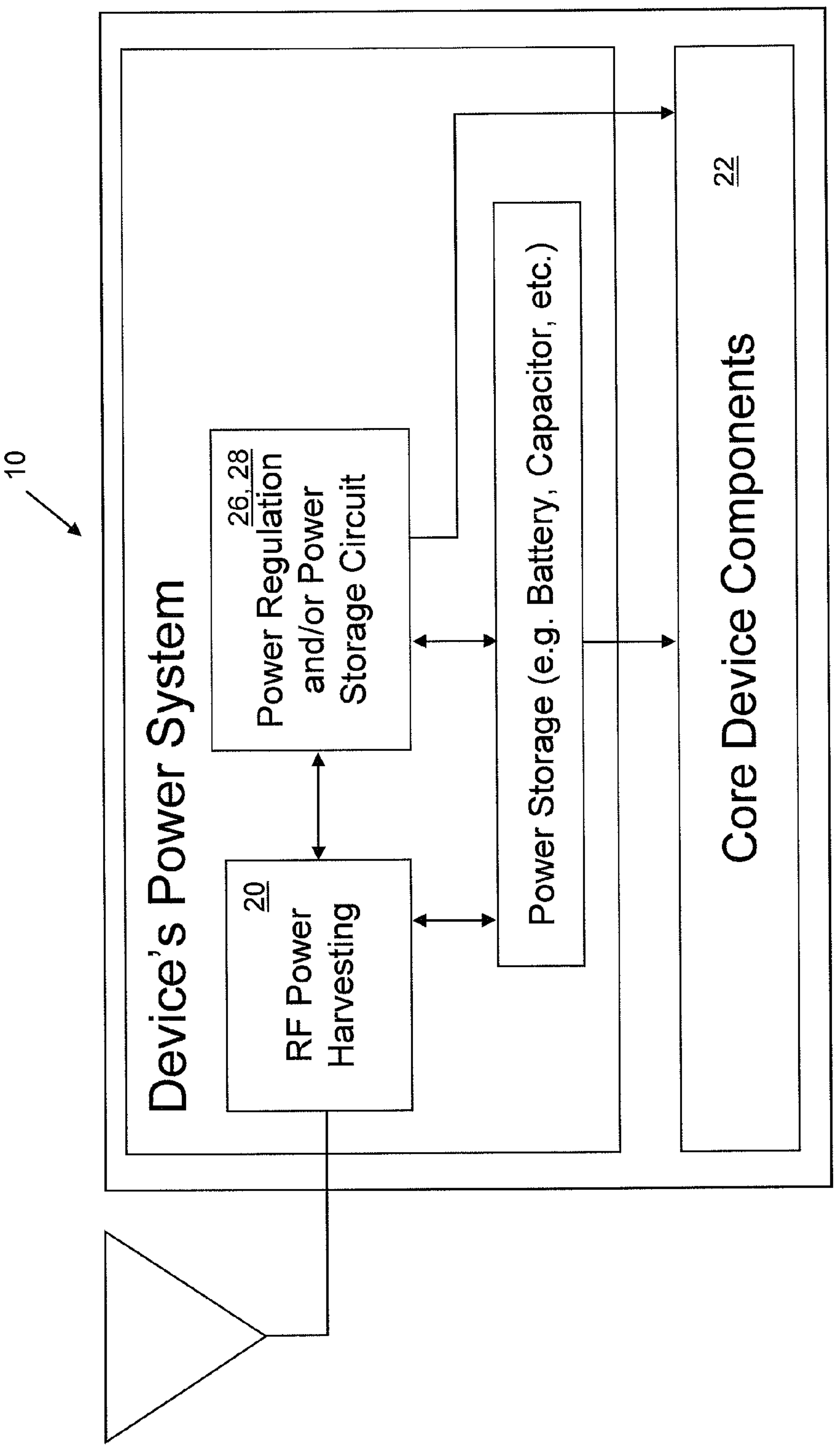
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 17

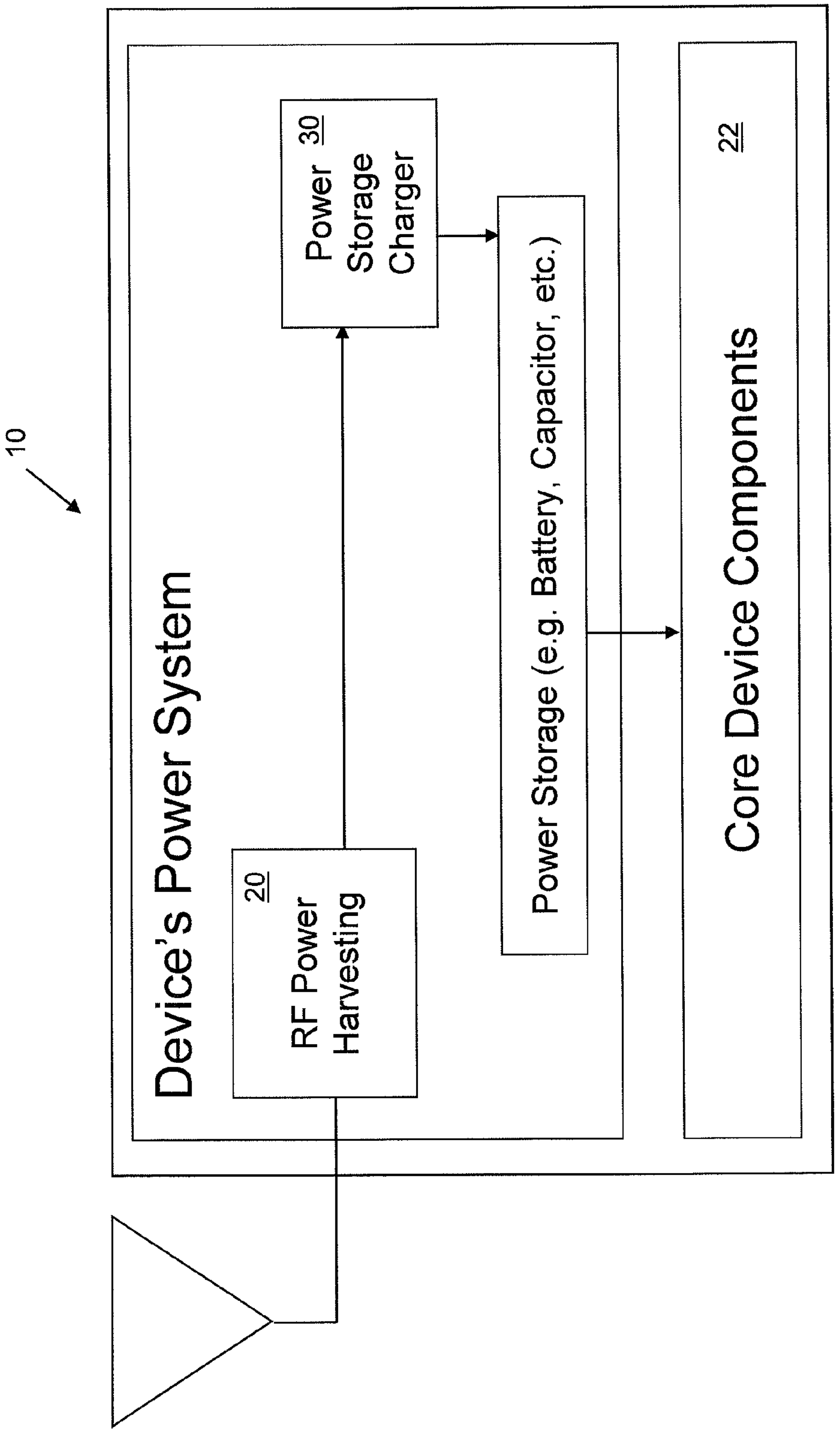
10
Device's Power System
RF Power Harvesting 20
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 18

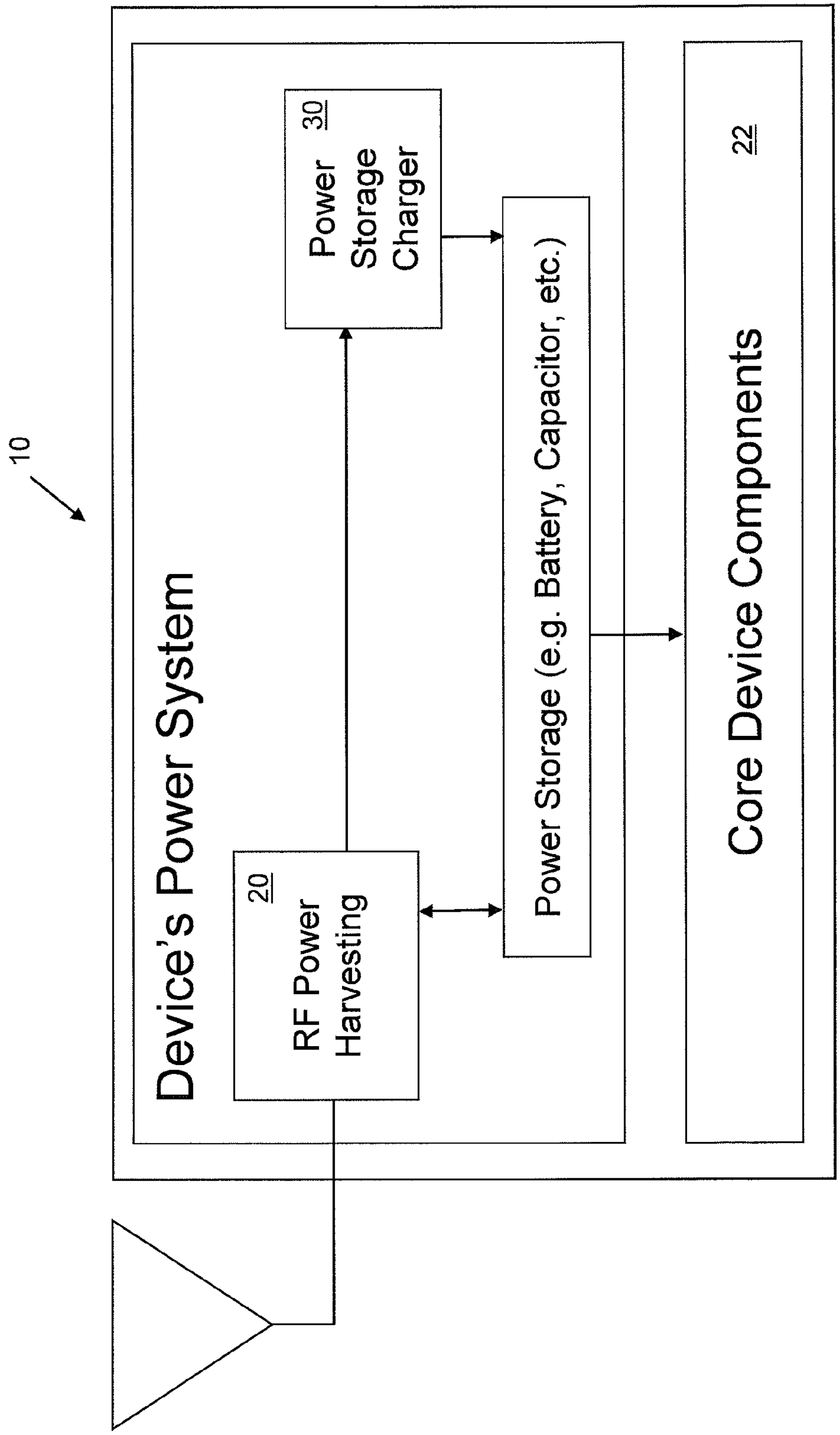
10
Device's Power System
20
RF Power Harvesting
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 19

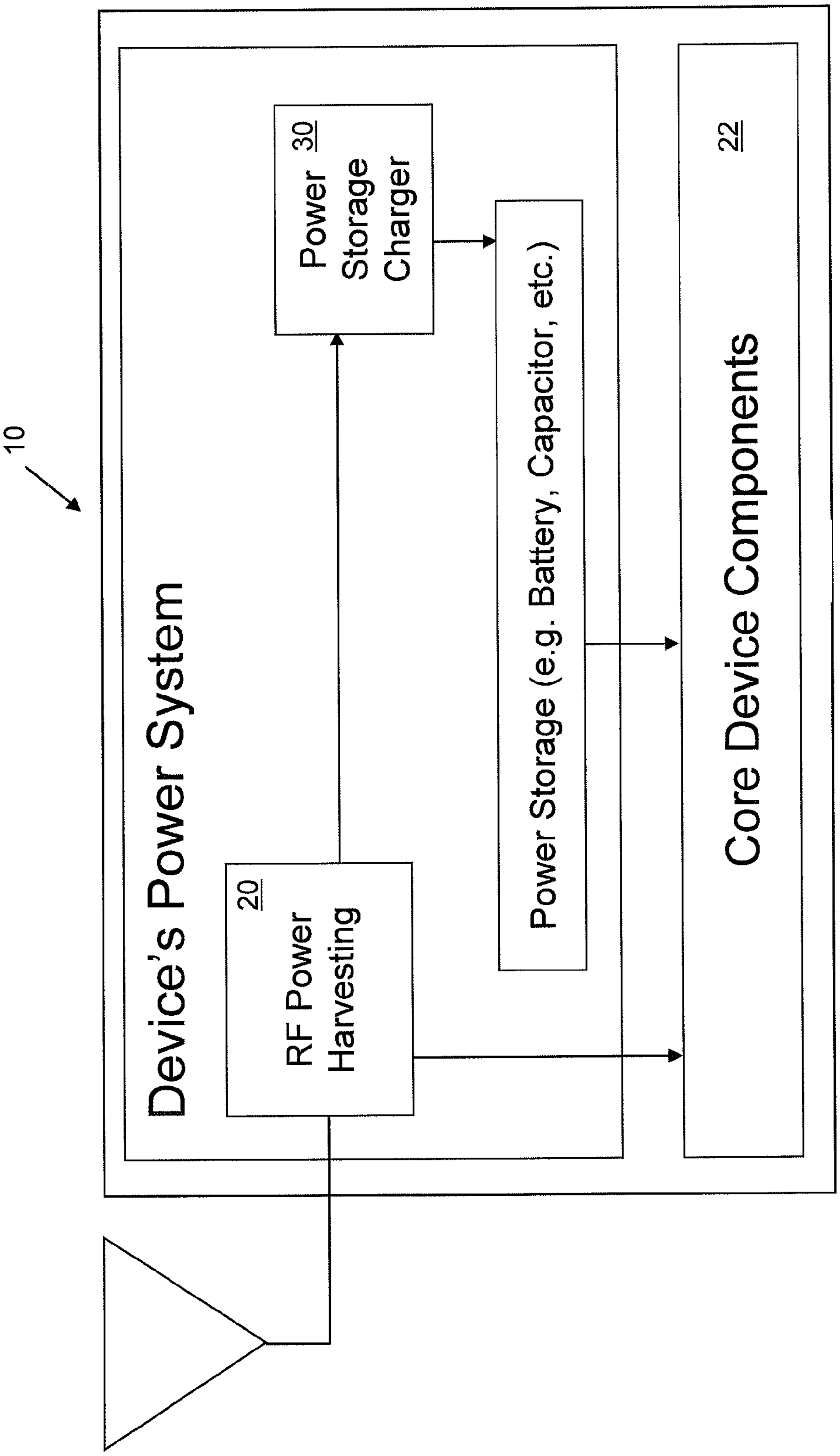
10
Device's Power System
RF Power Harvesting 20
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 20

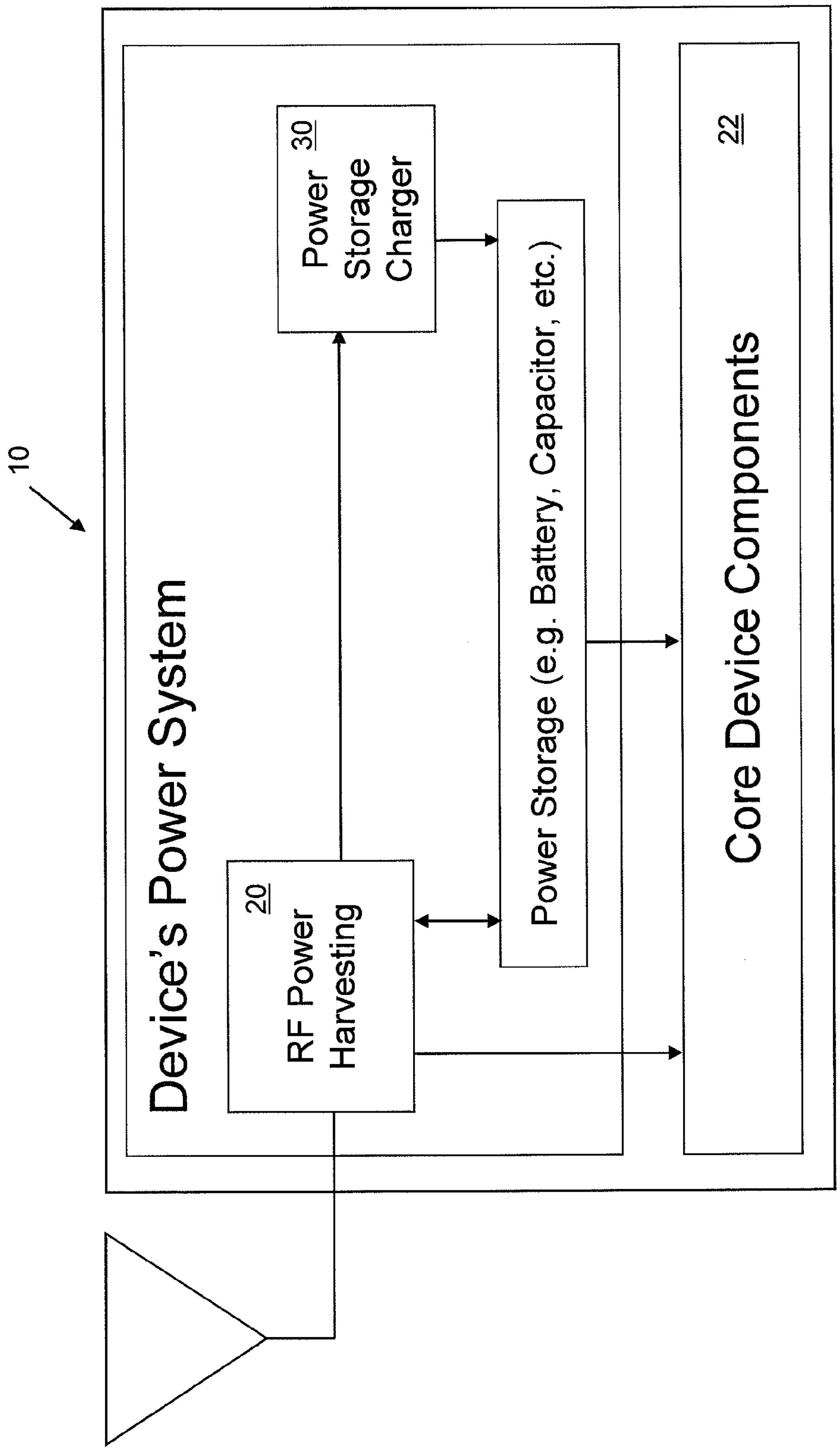
10
Device's Power System
RF Power Harvesting 20
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 21

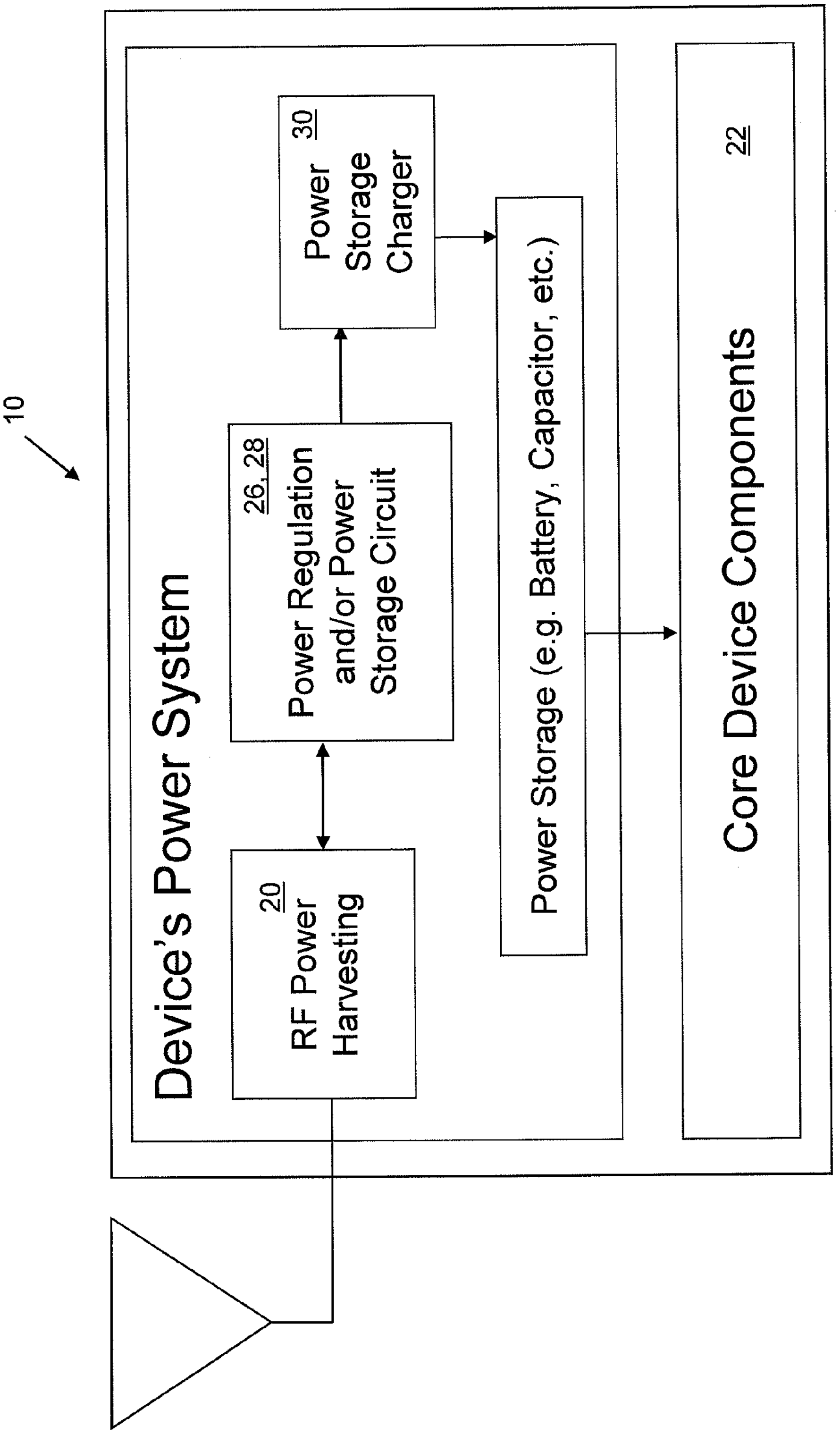
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 22

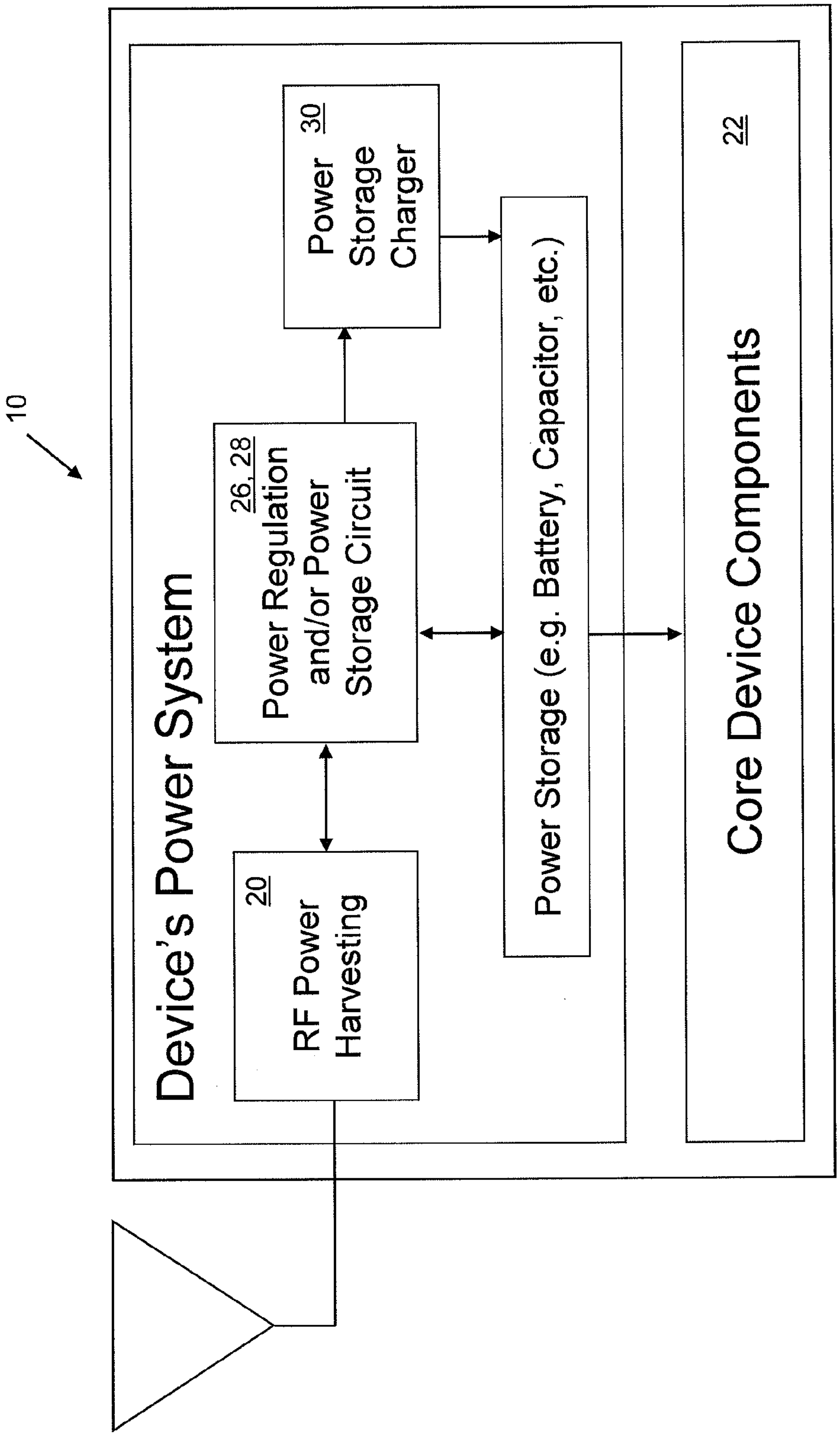
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 23

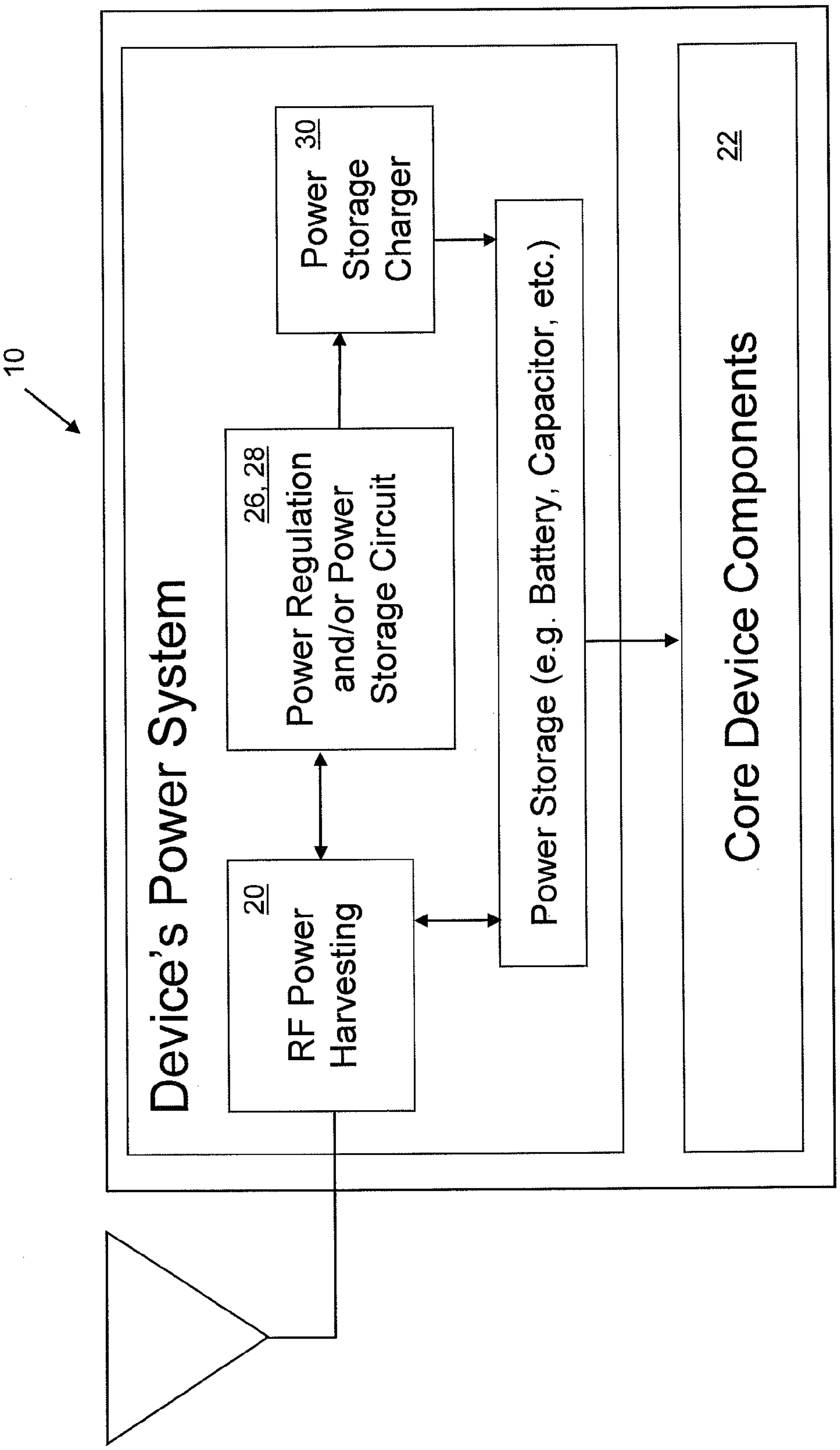
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 24

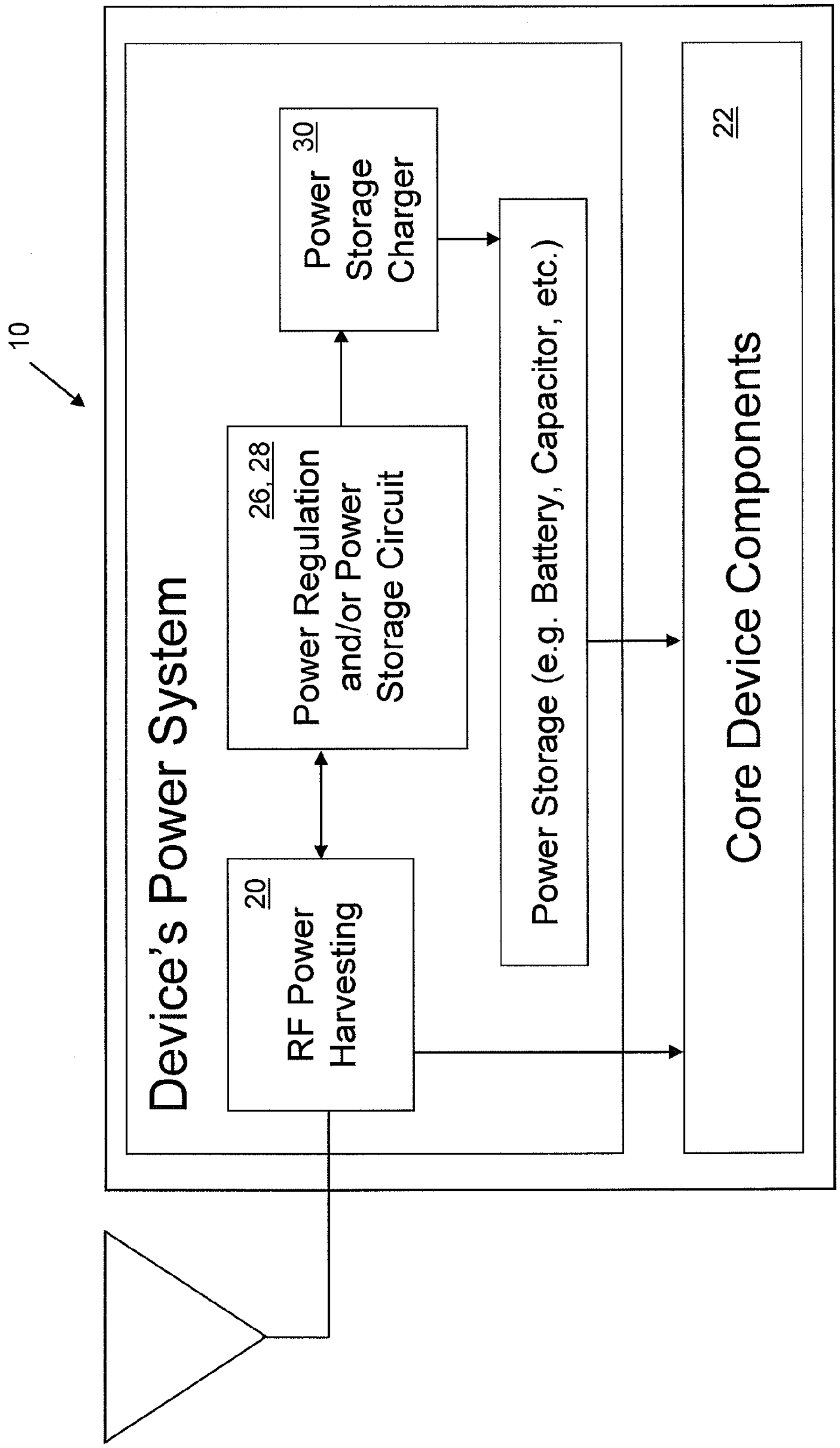
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 25

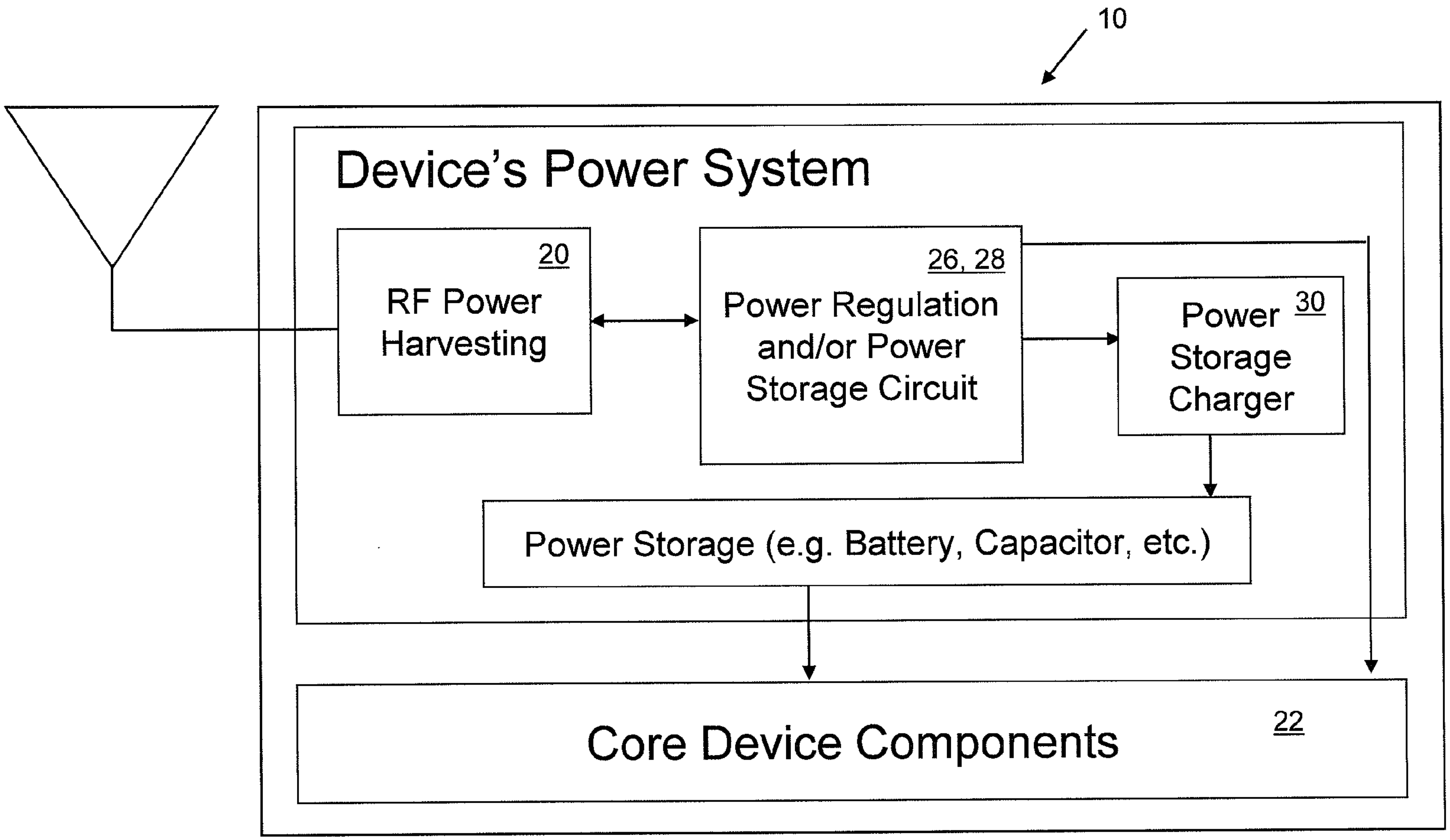
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 26

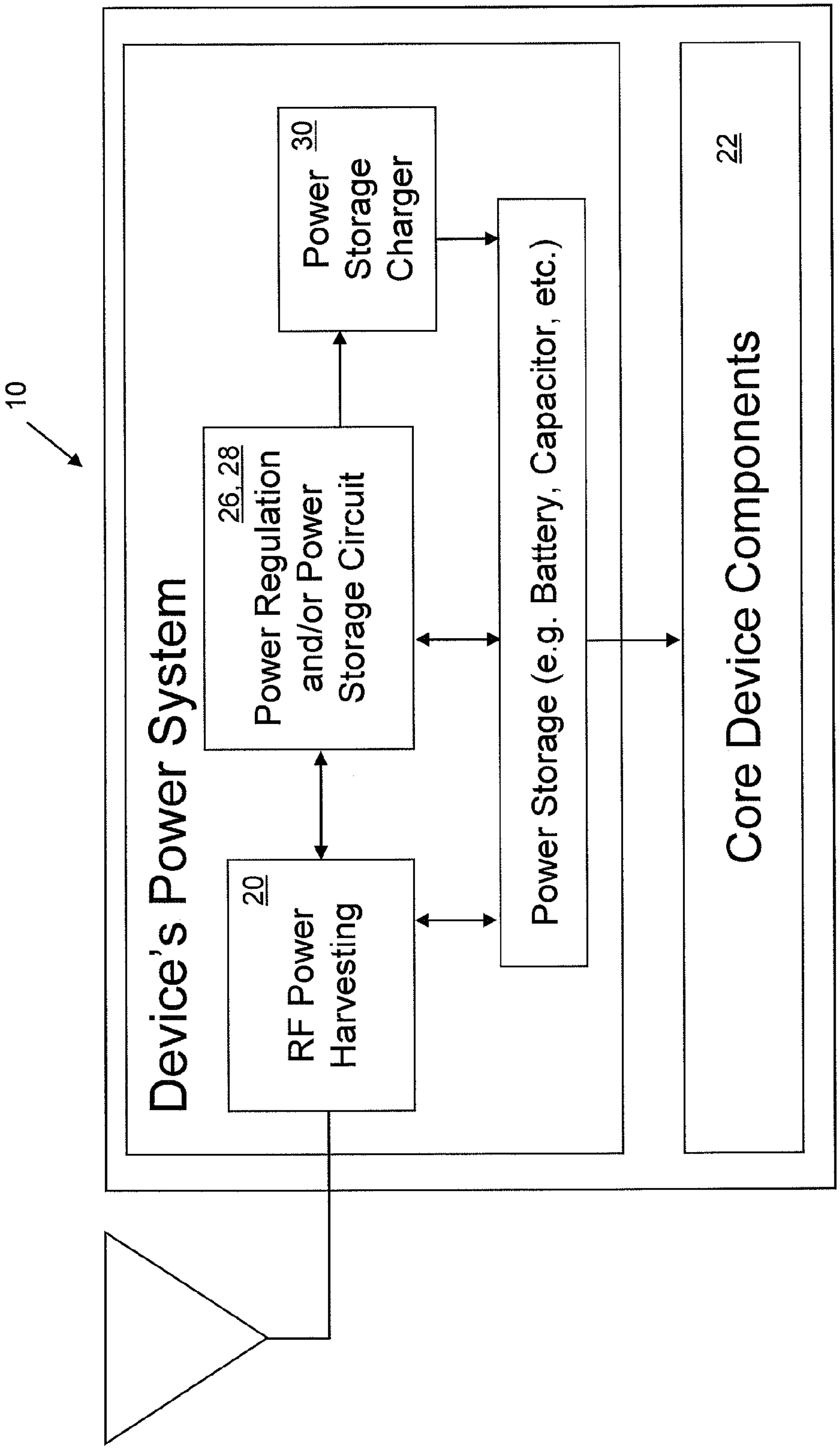
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 27

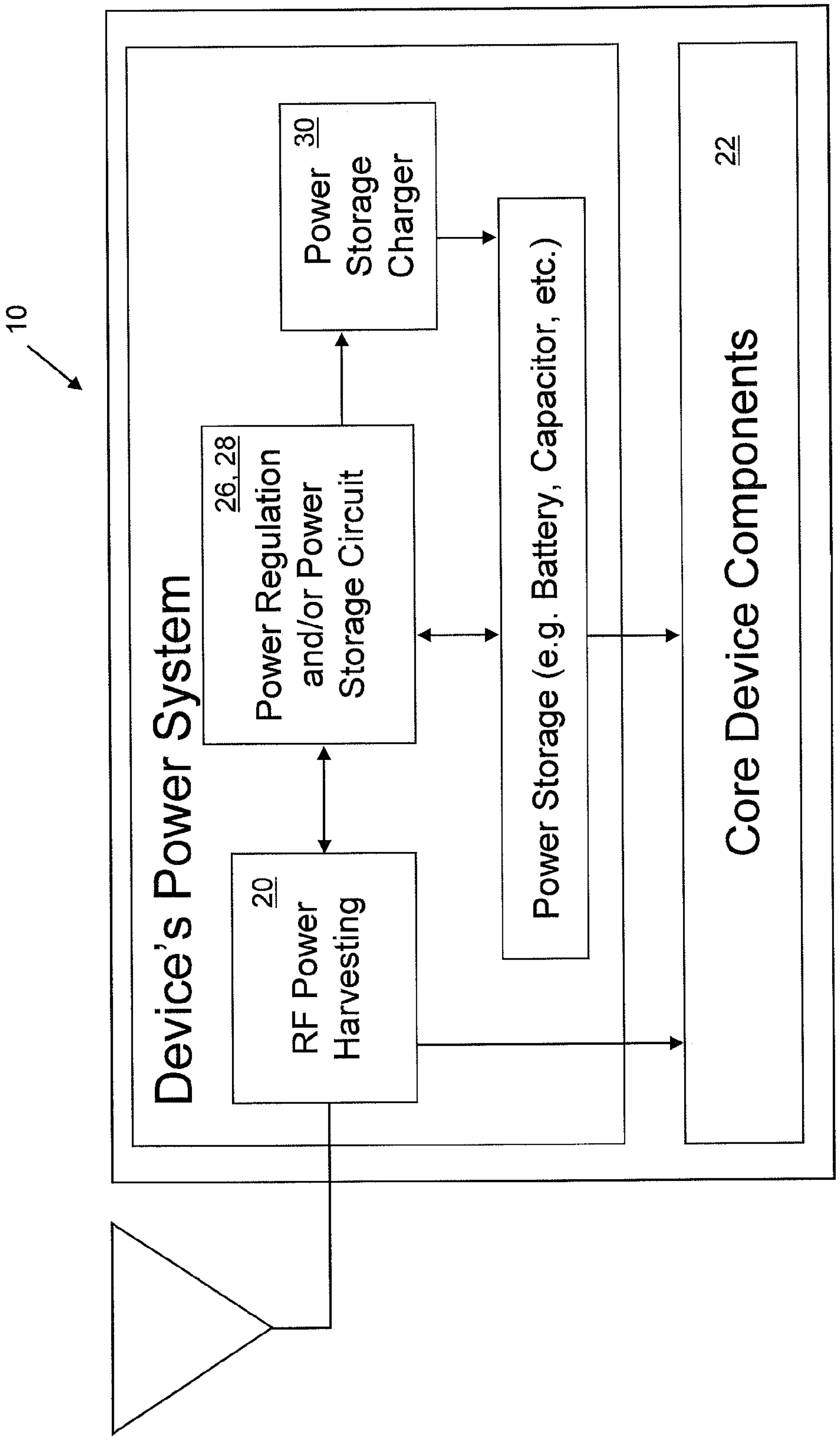
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 28

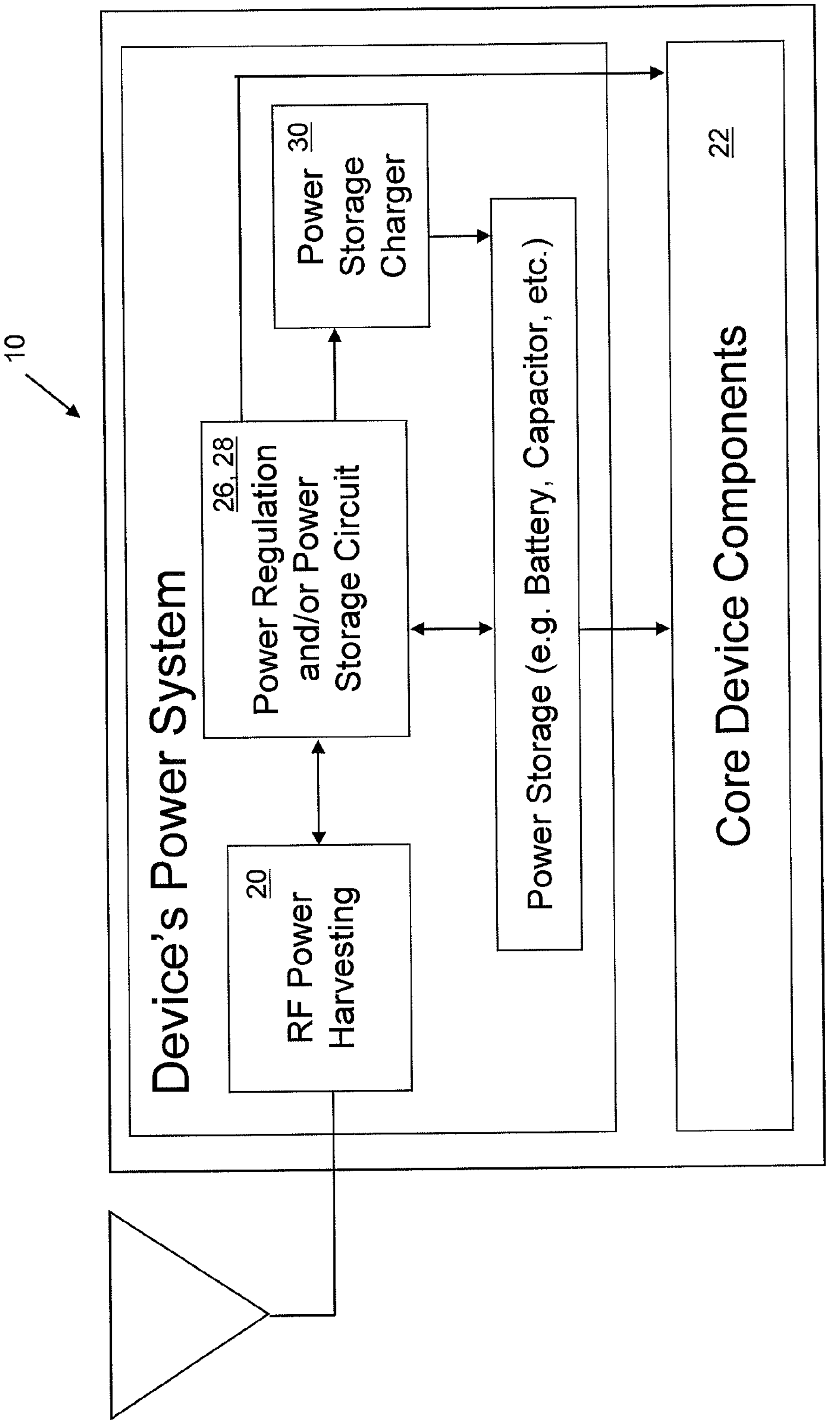
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 29

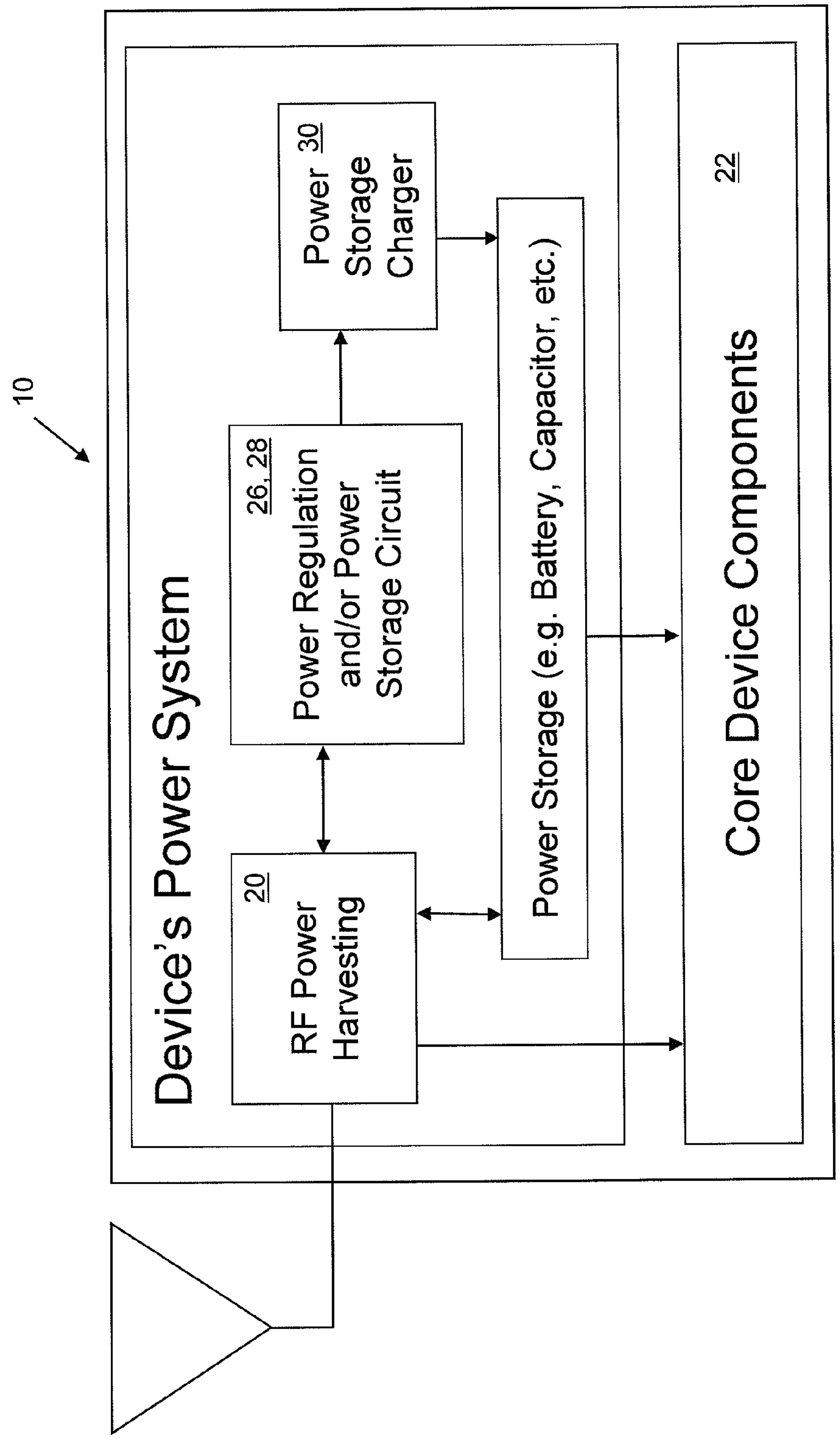
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 30

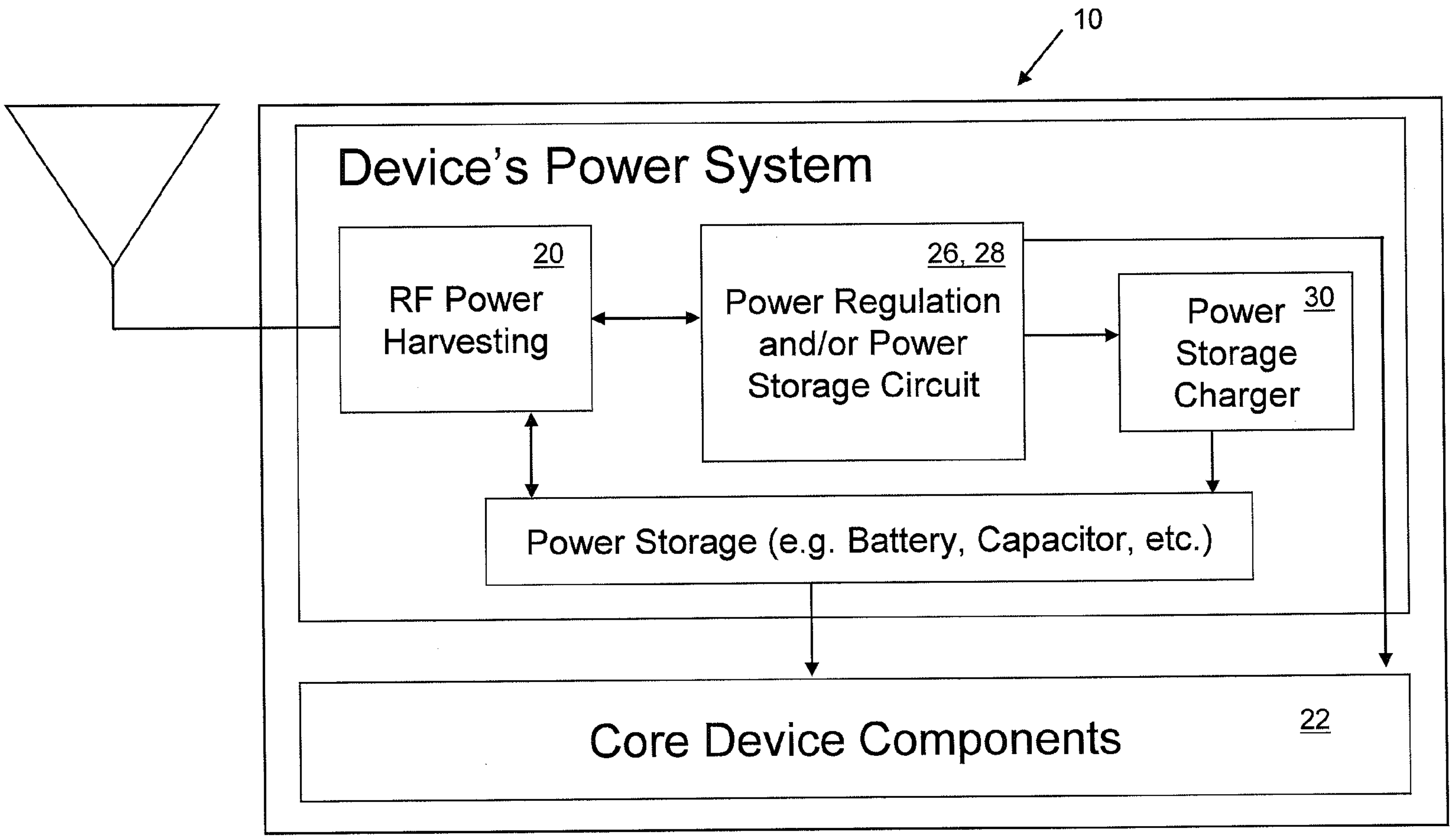
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 31

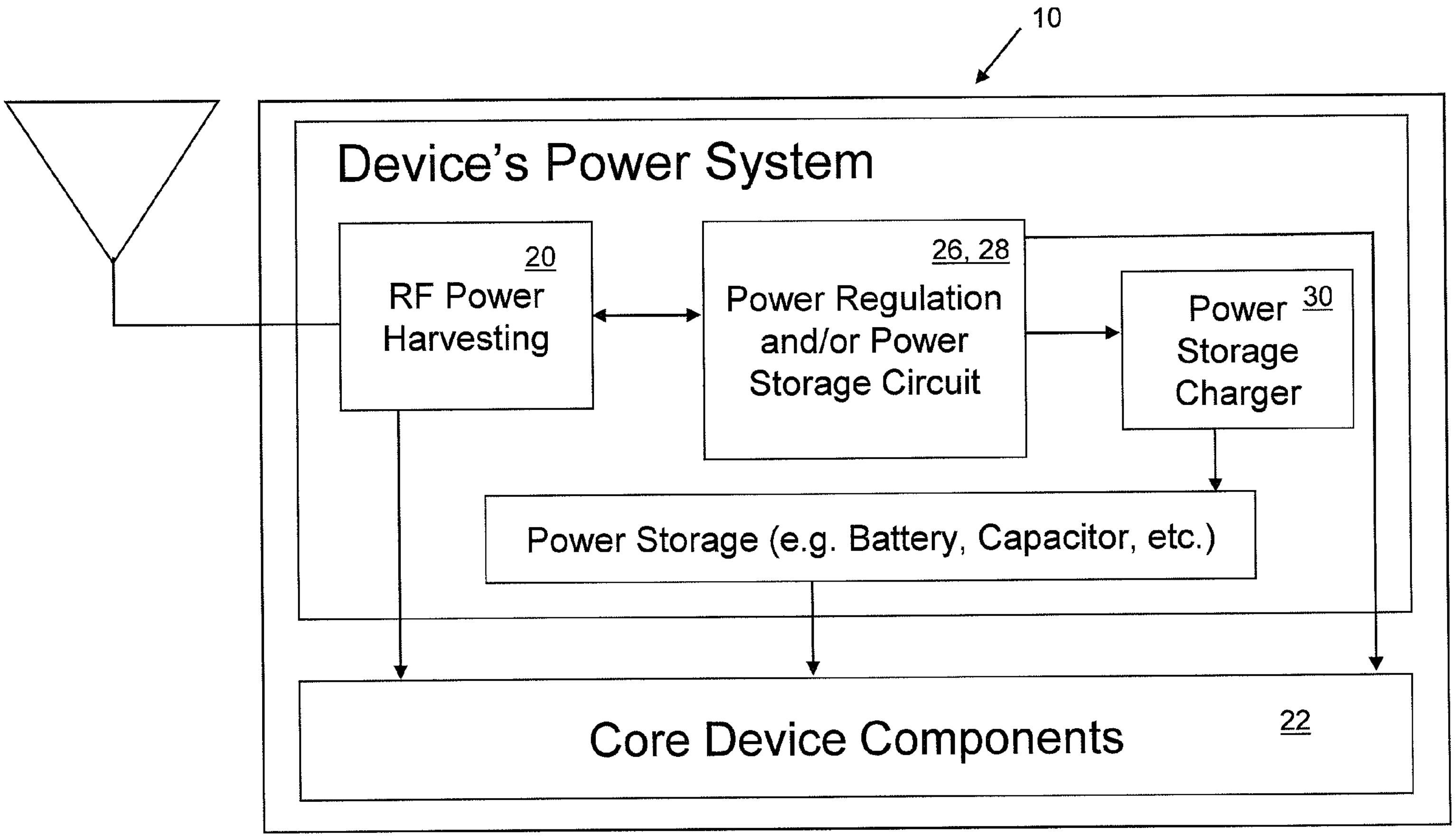
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 32

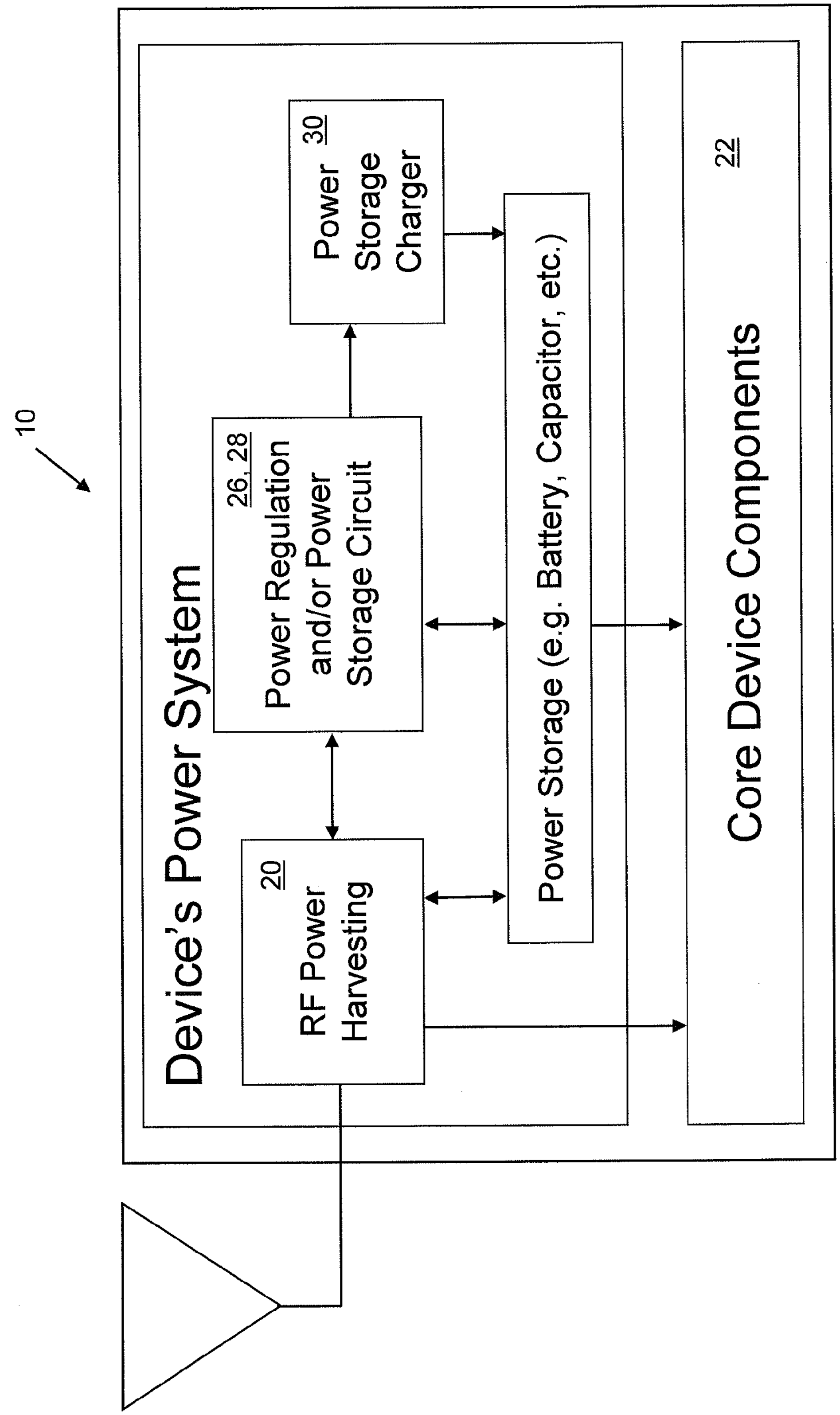
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 33

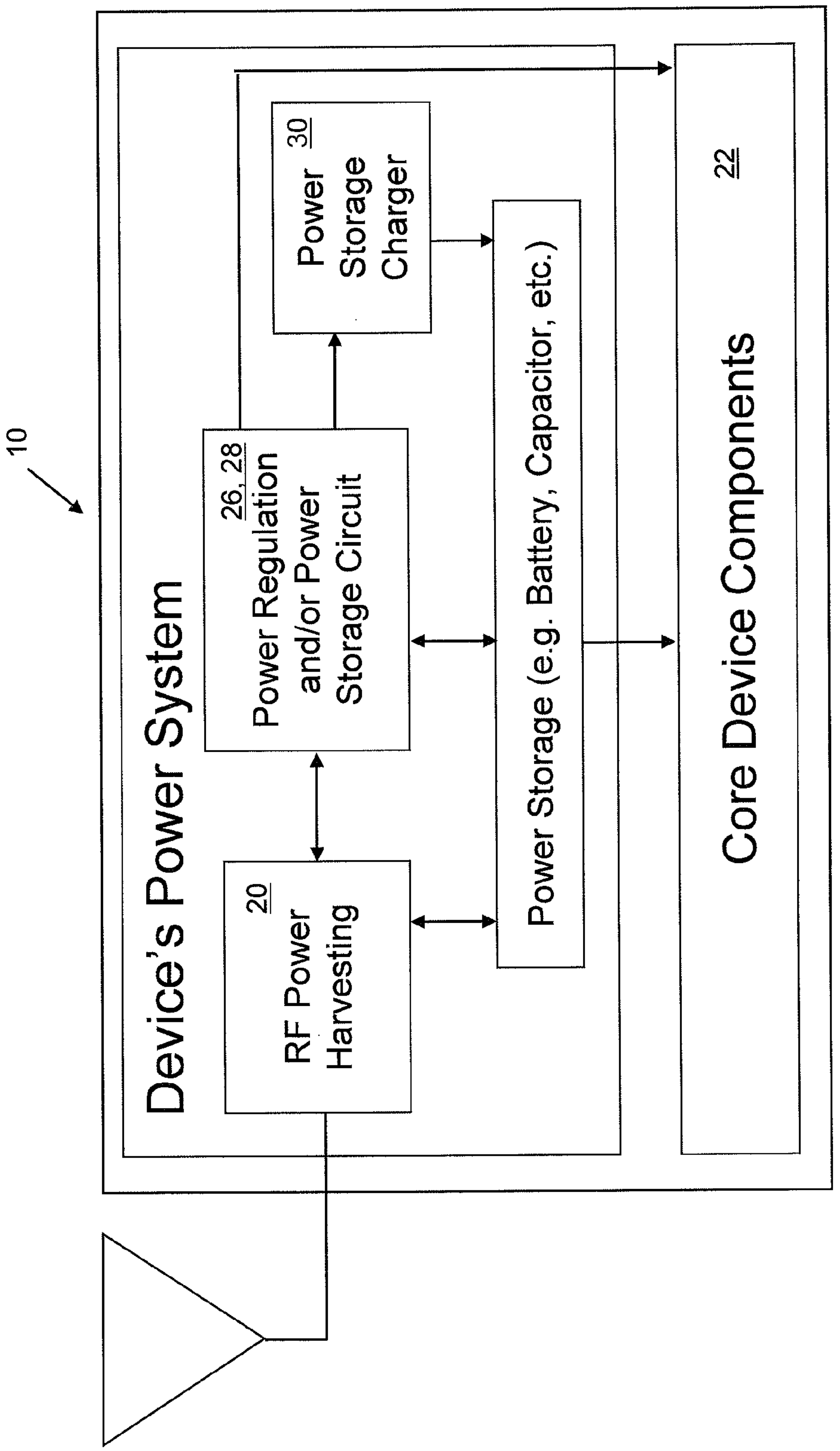
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 34

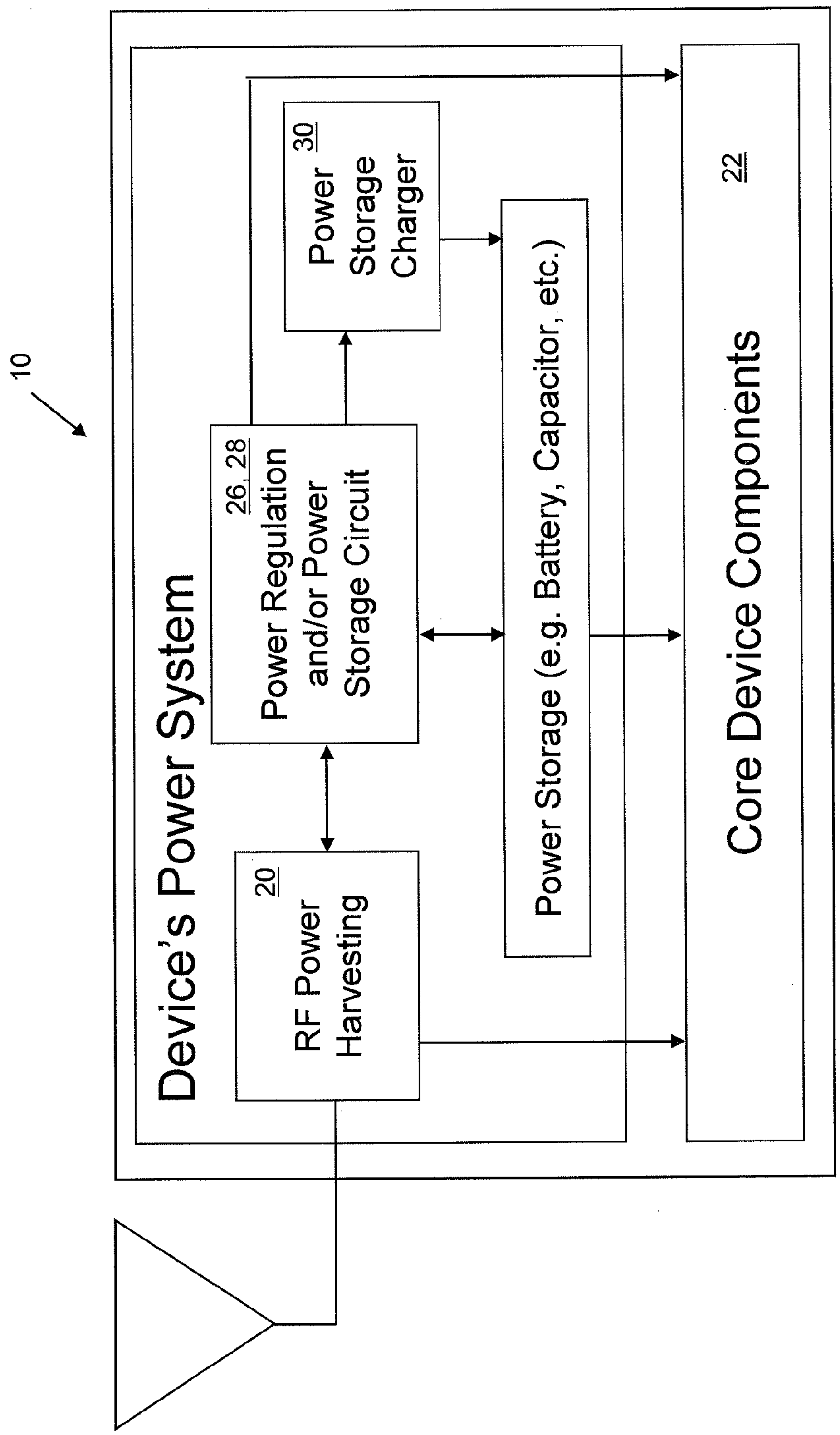
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 35

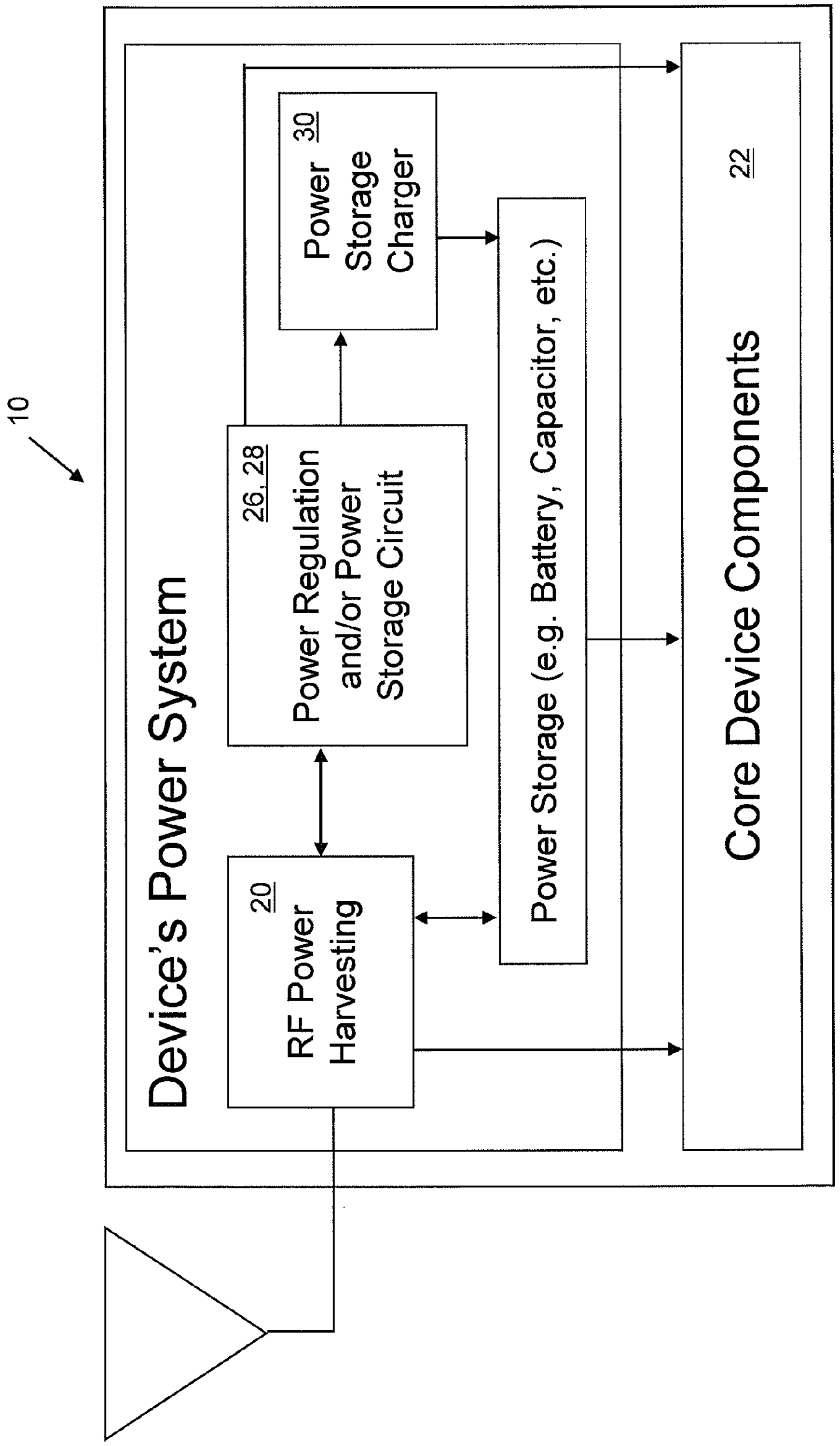
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 36

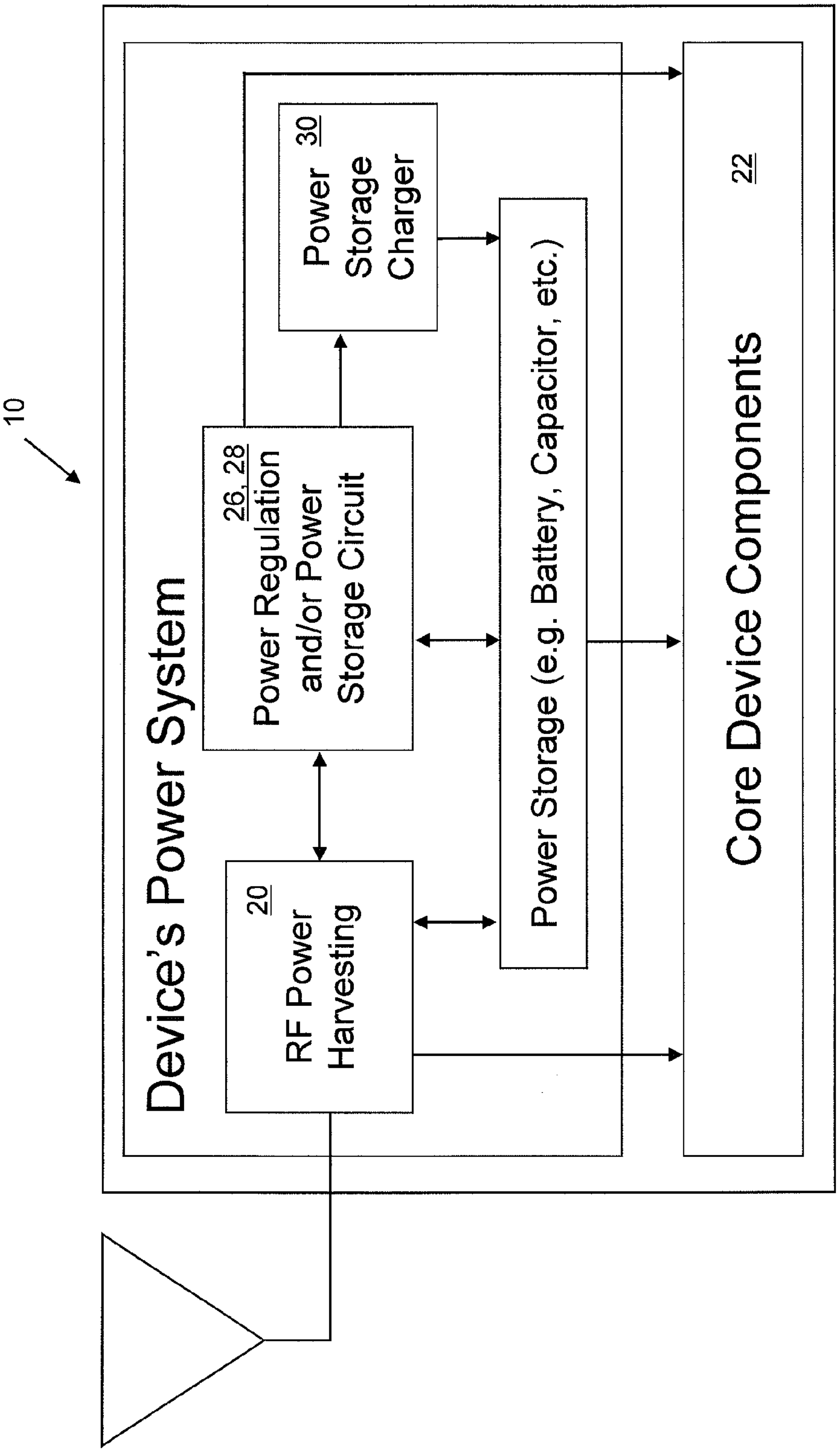
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 37

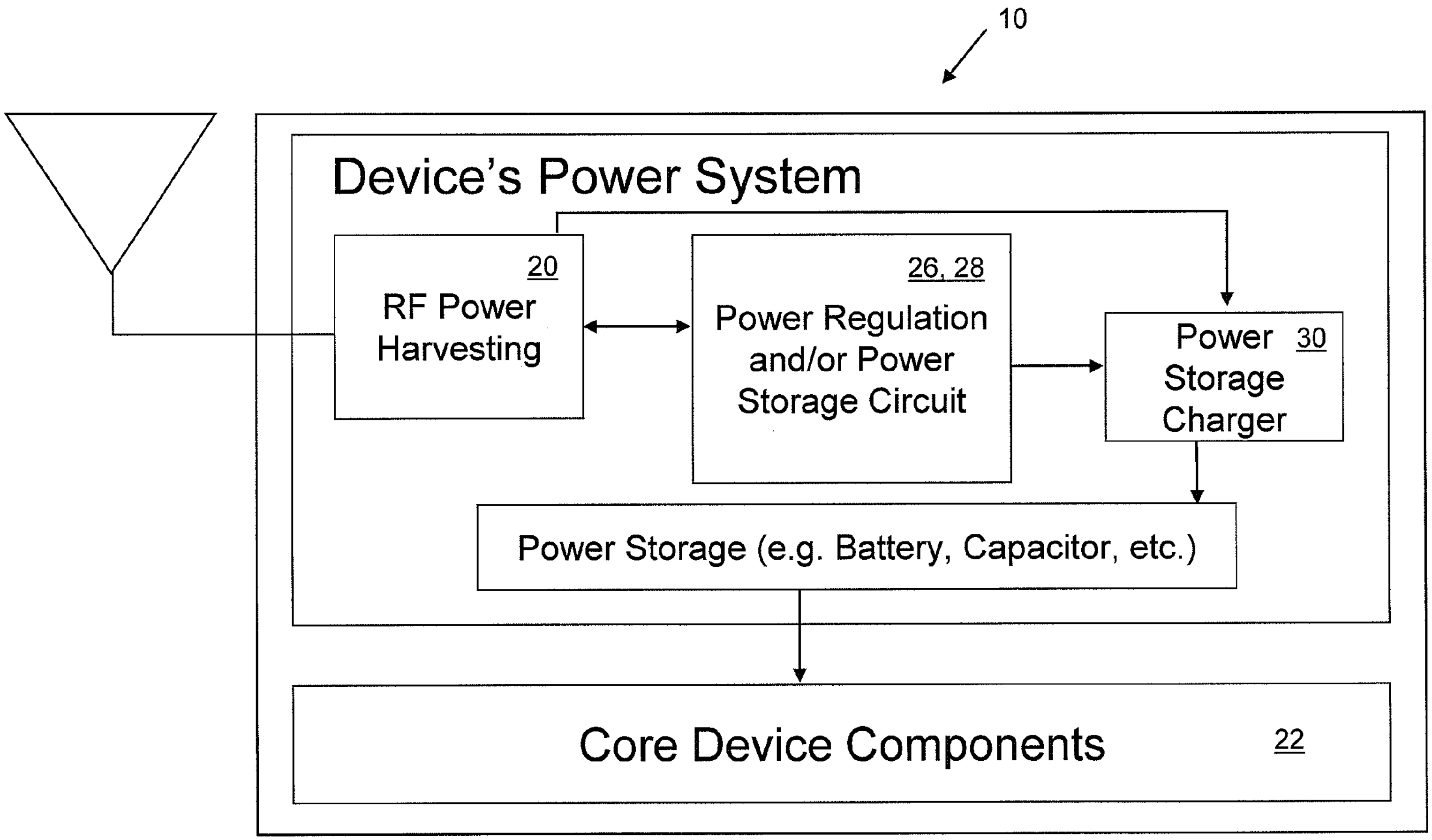
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 38

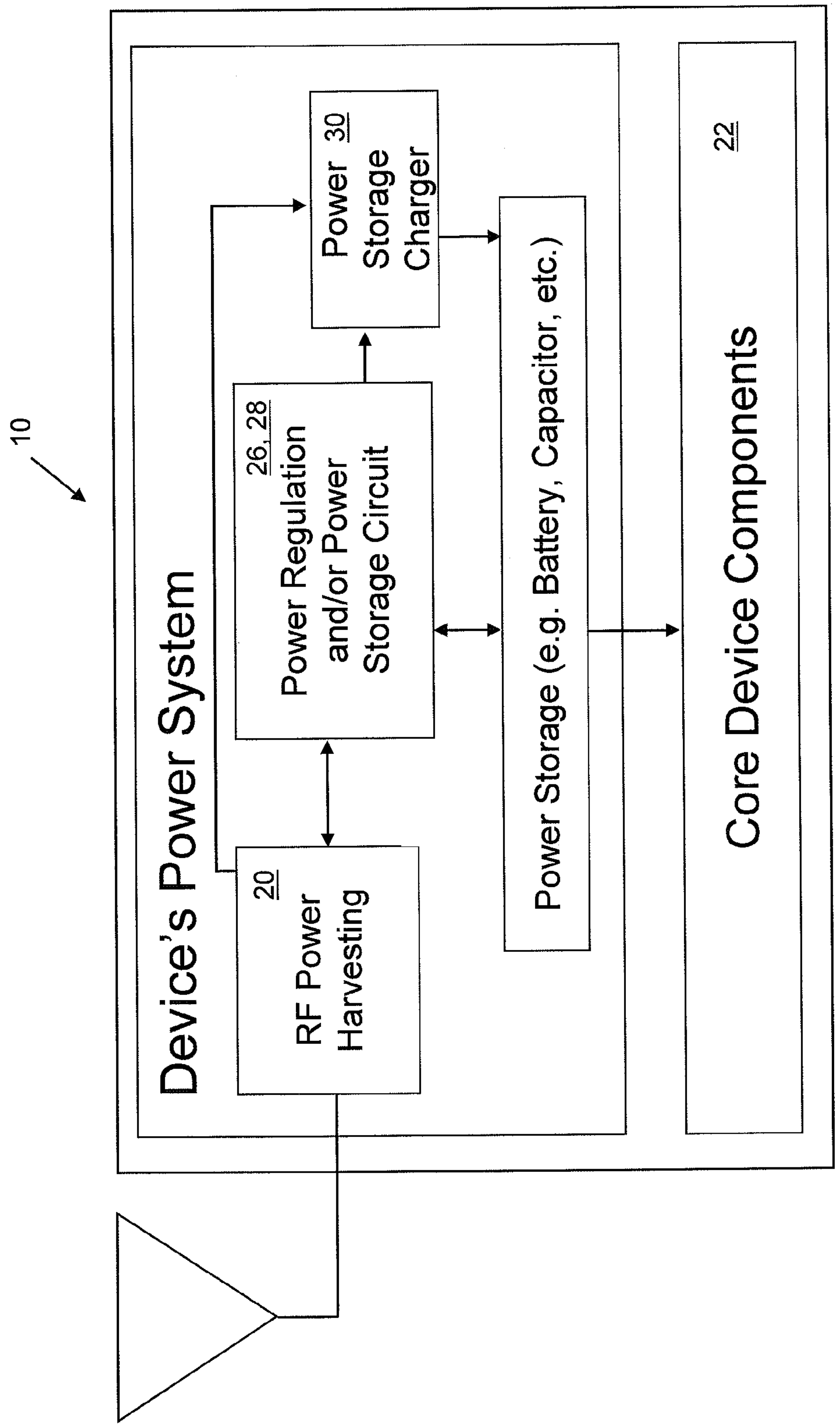
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 39

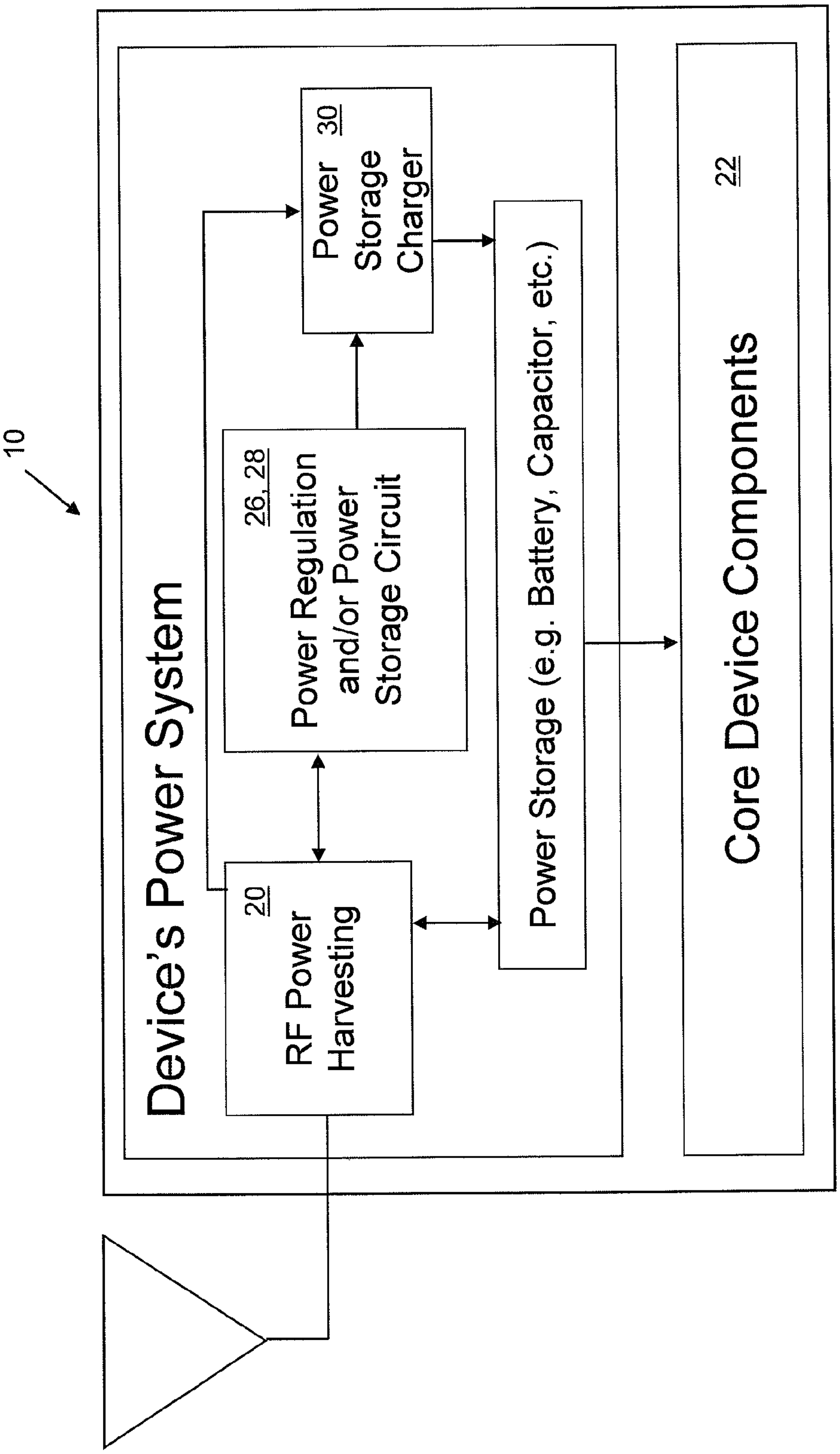
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 40

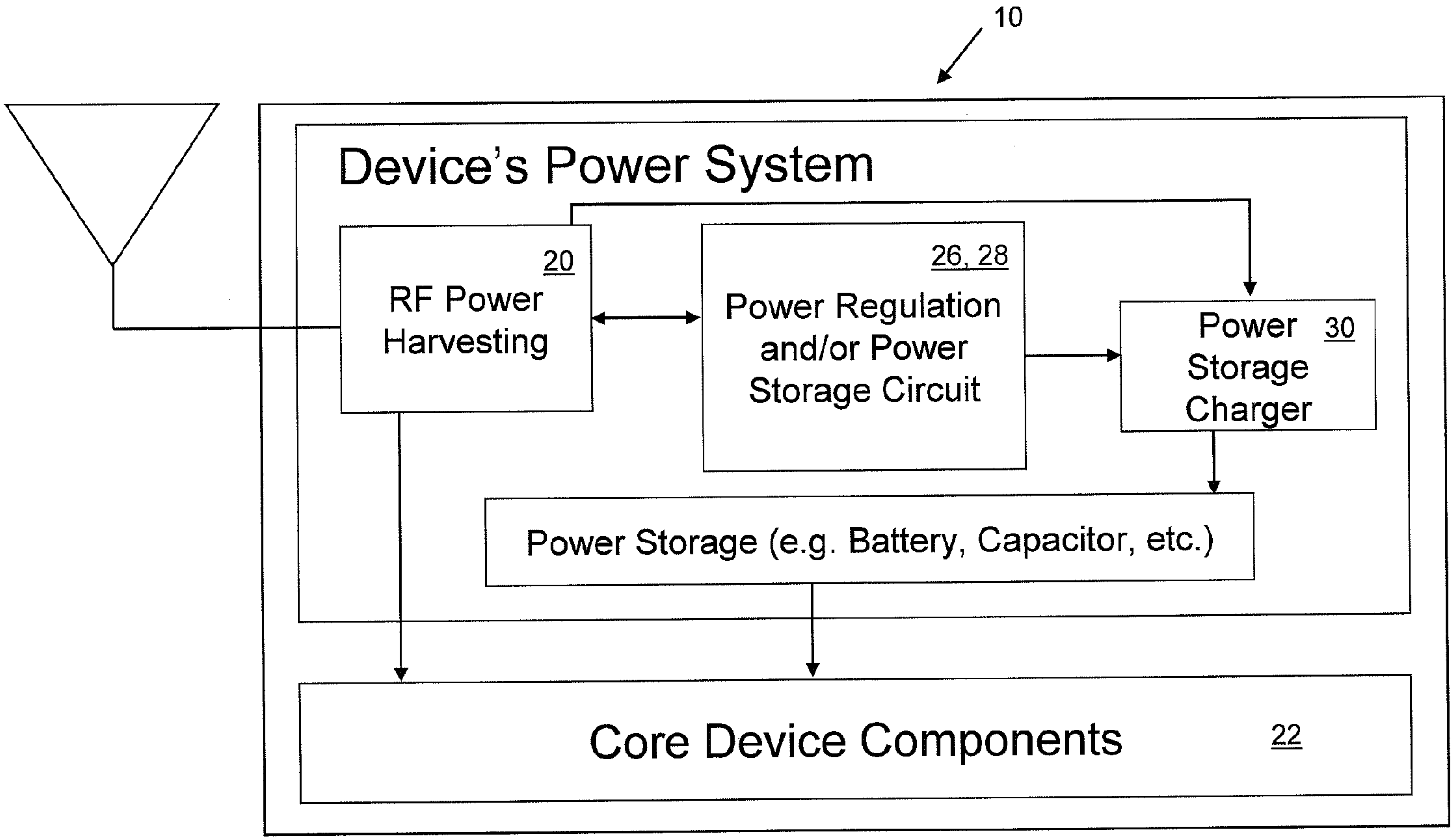
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 41

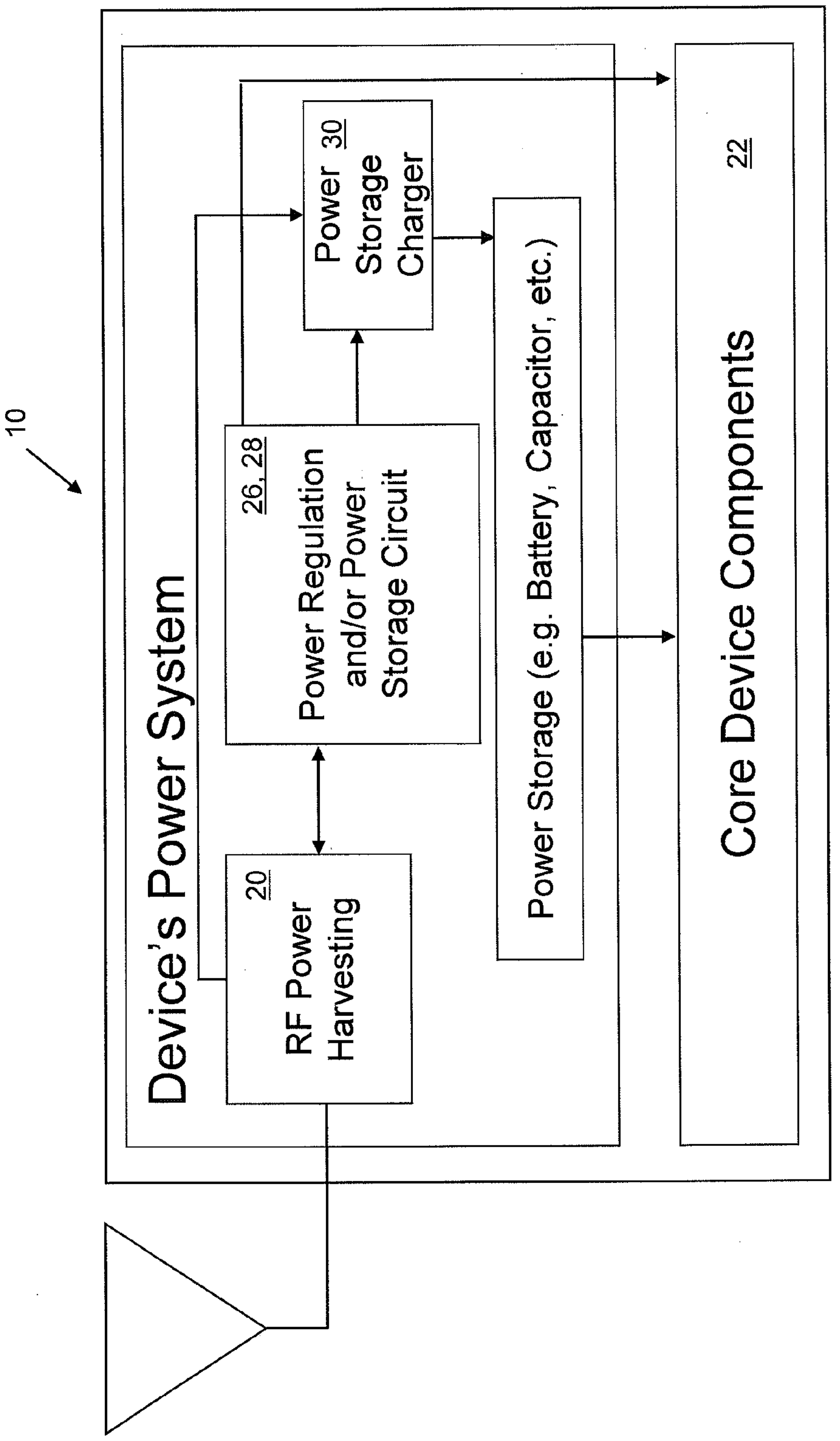
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 42

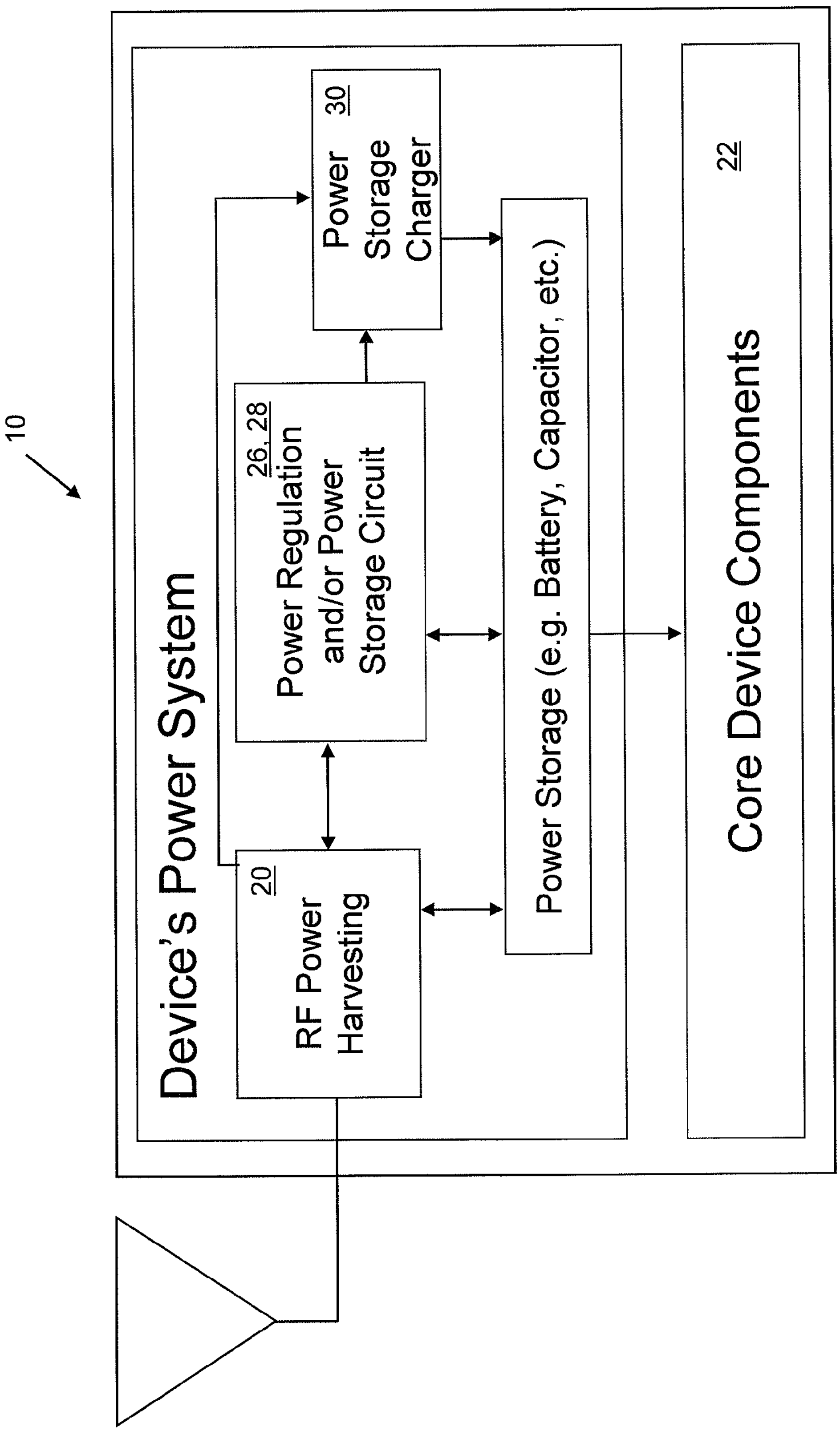
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 43

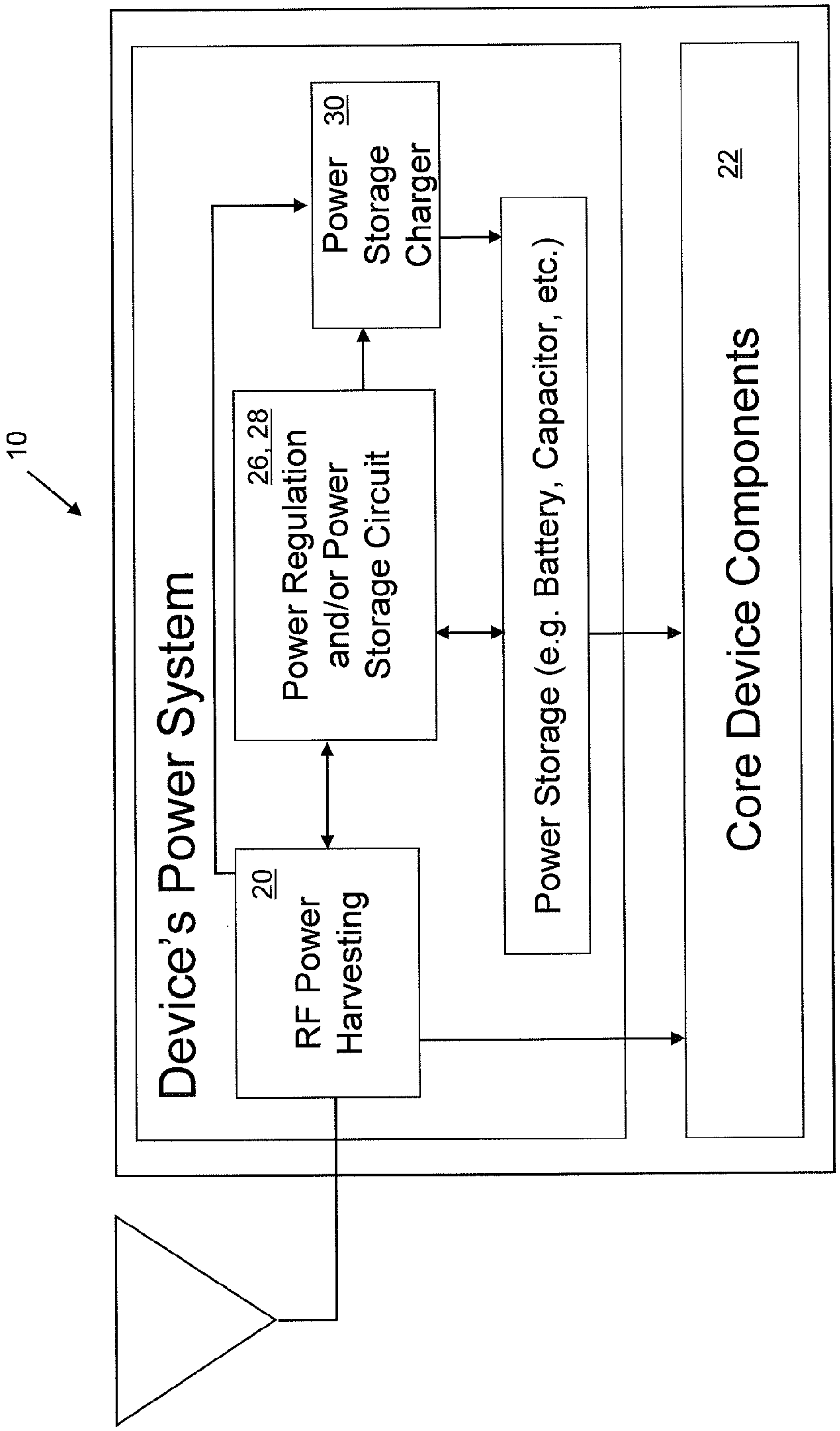
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 44

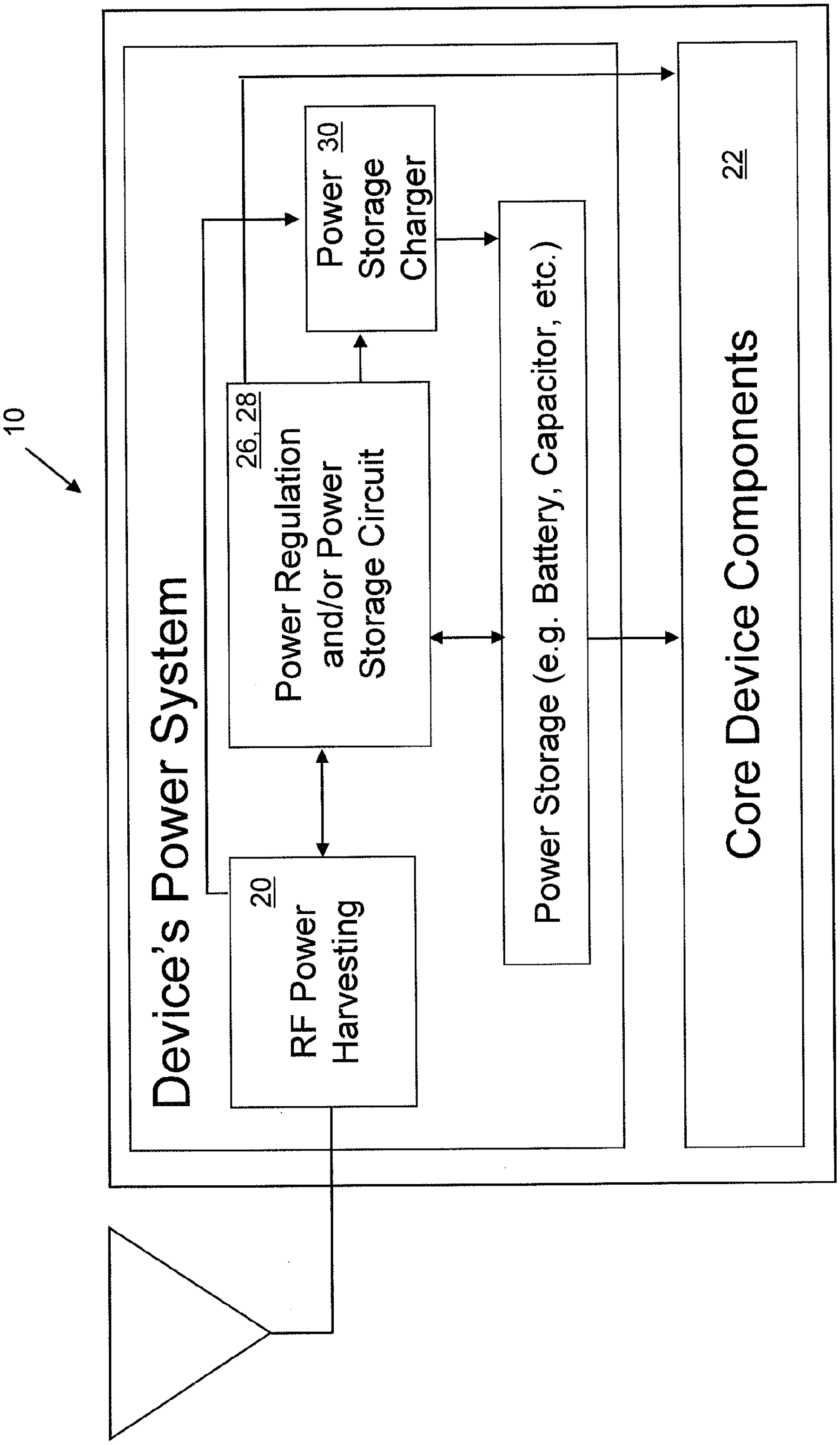
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 45

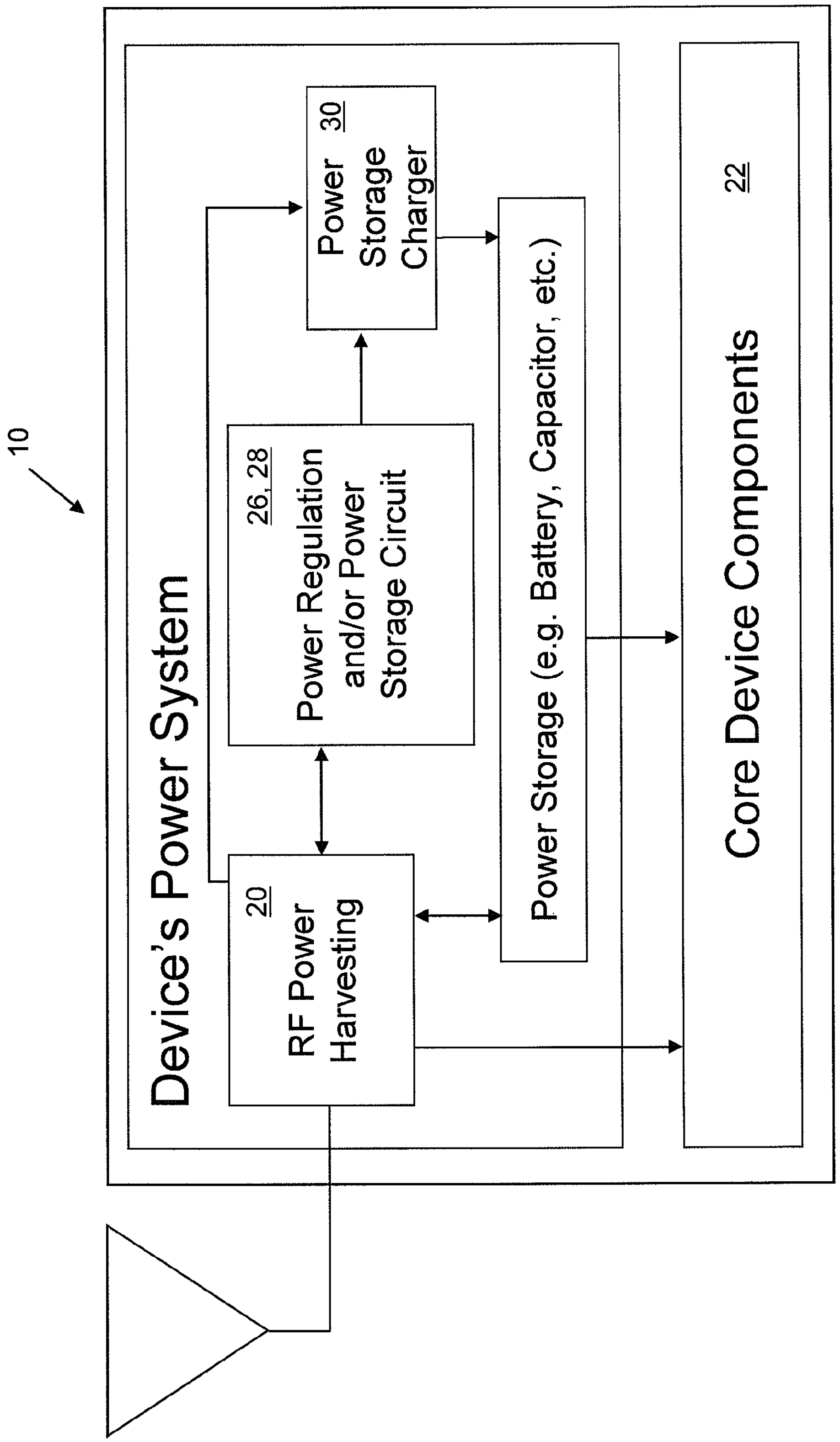
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 46

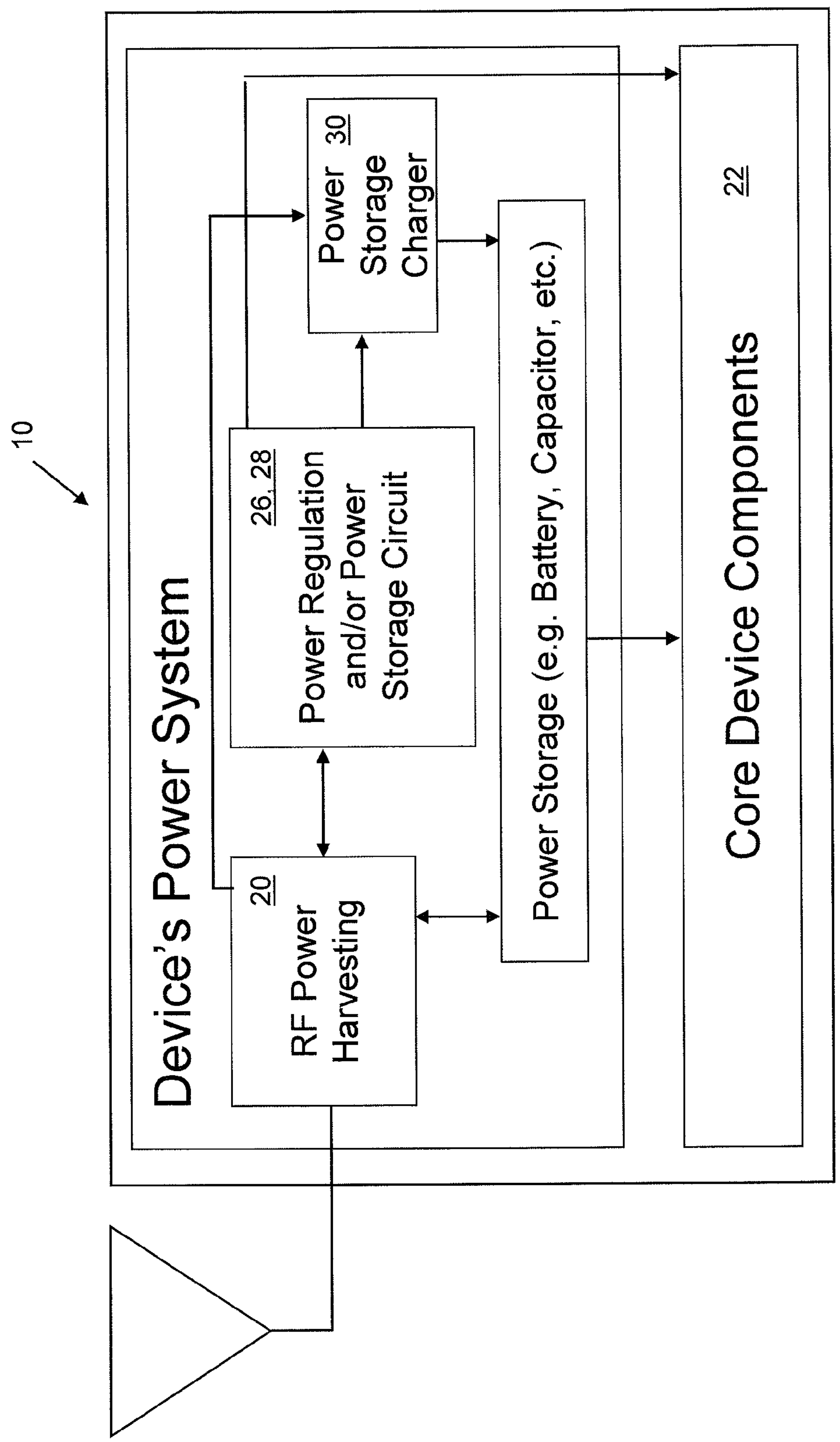
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 47

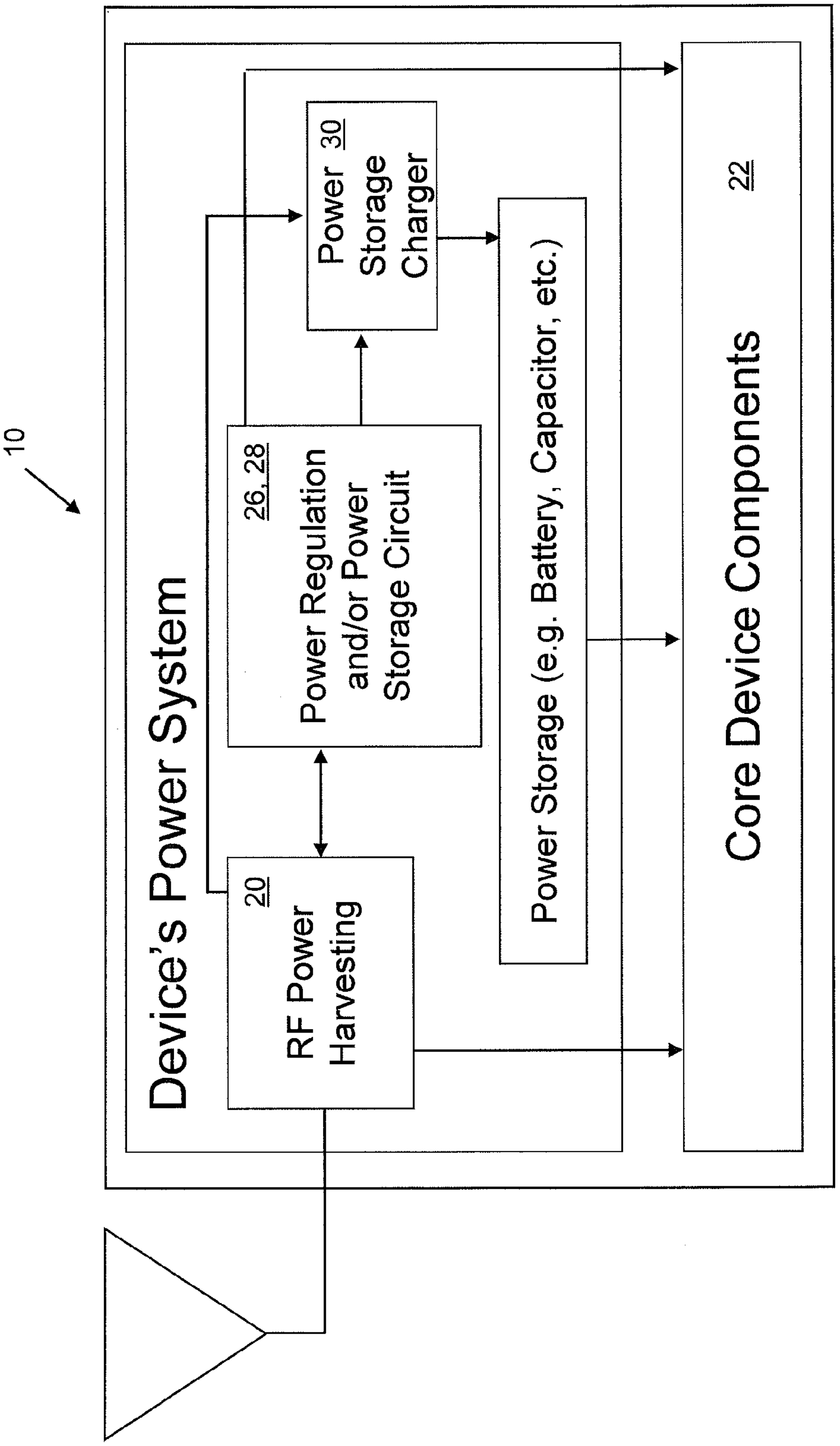
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 48

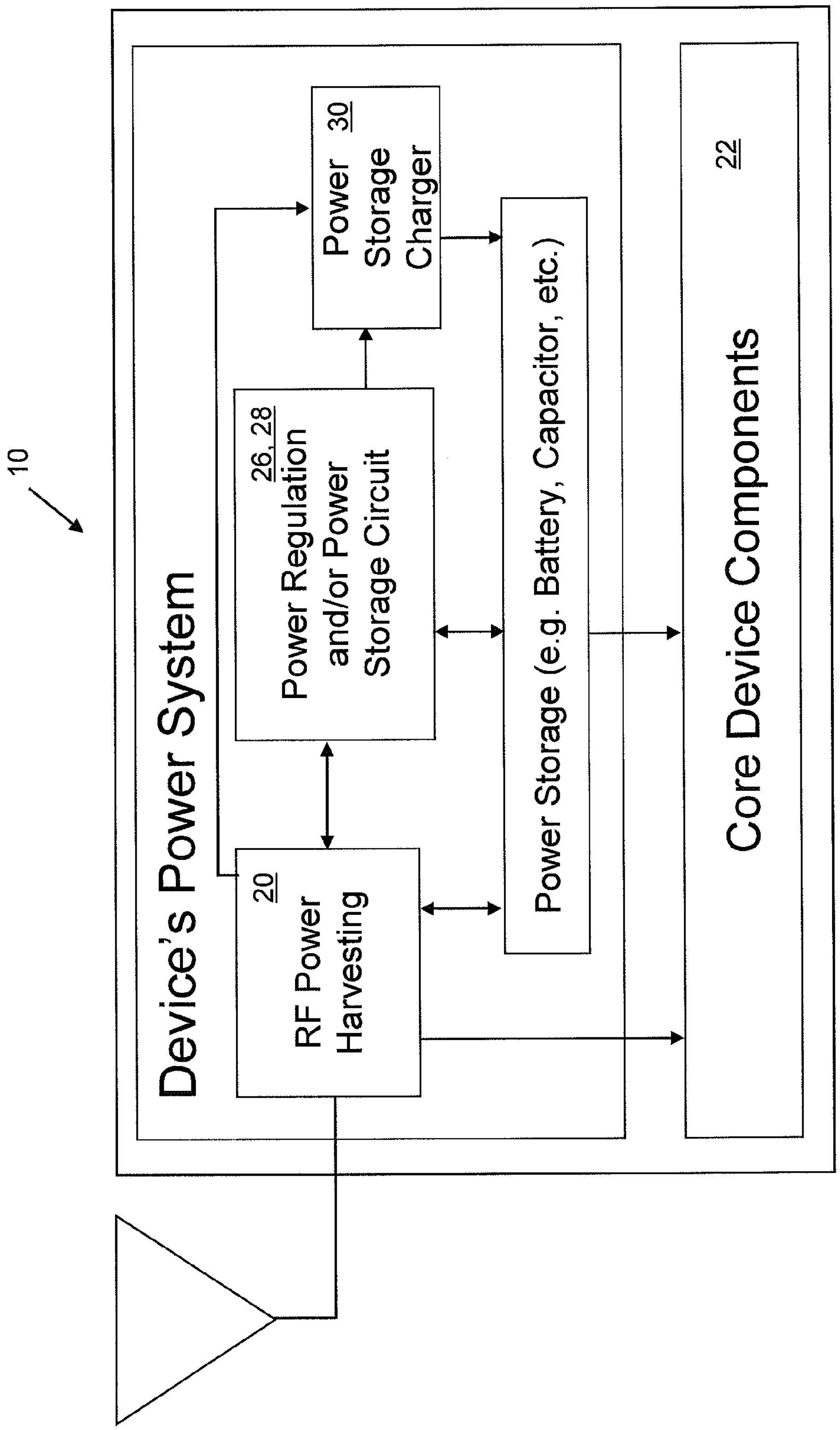
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 49

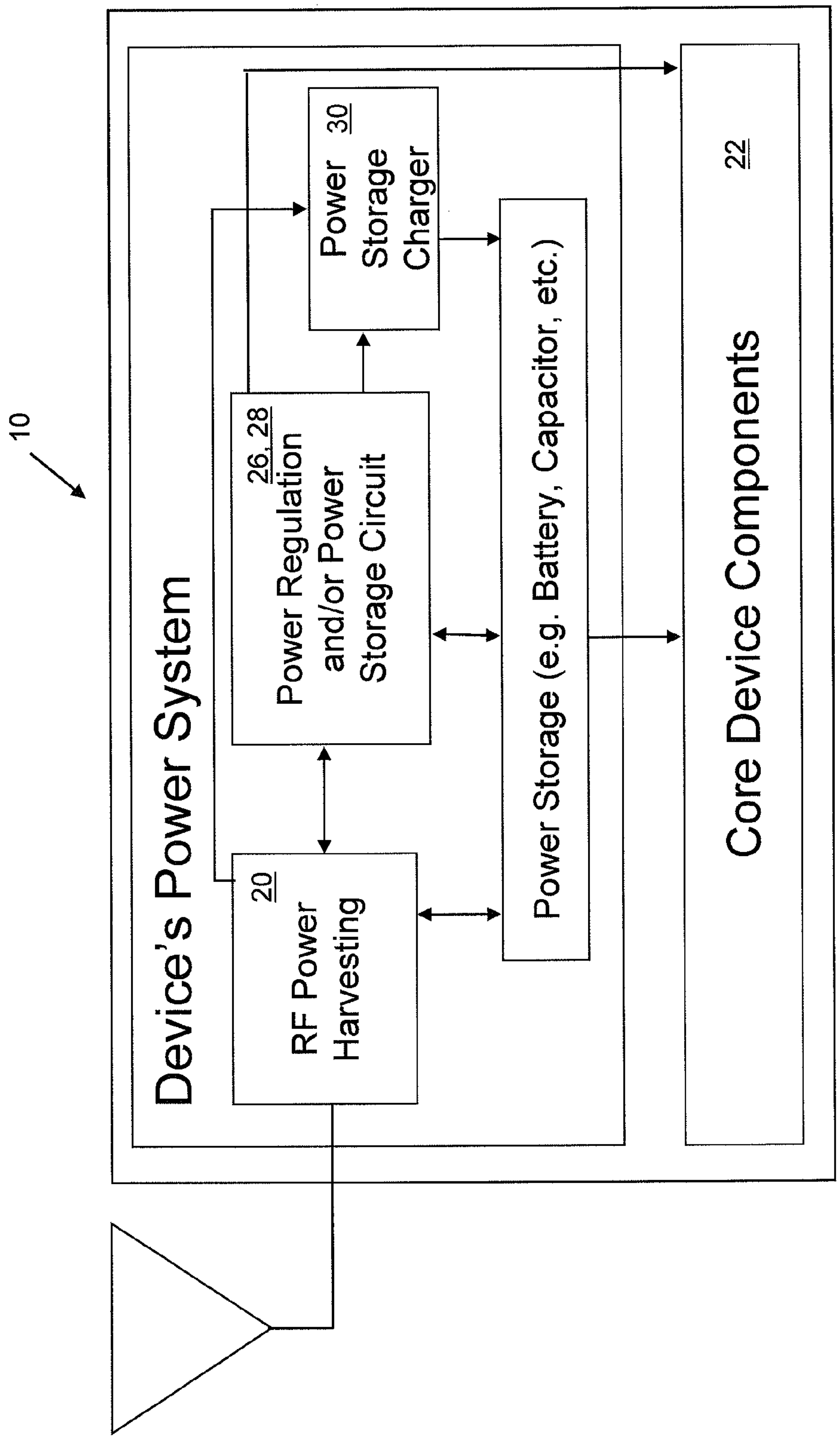
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 50

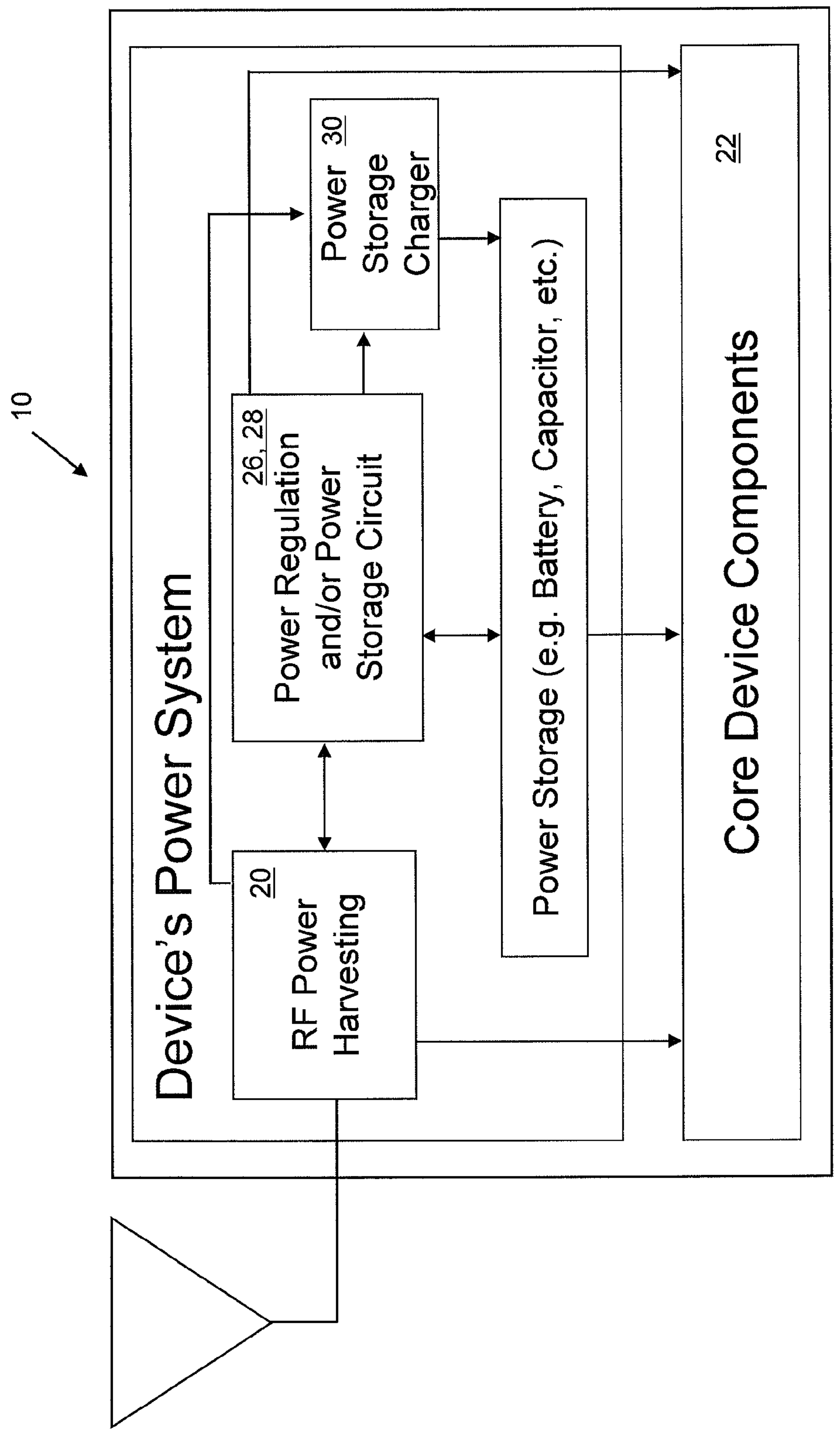
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 51

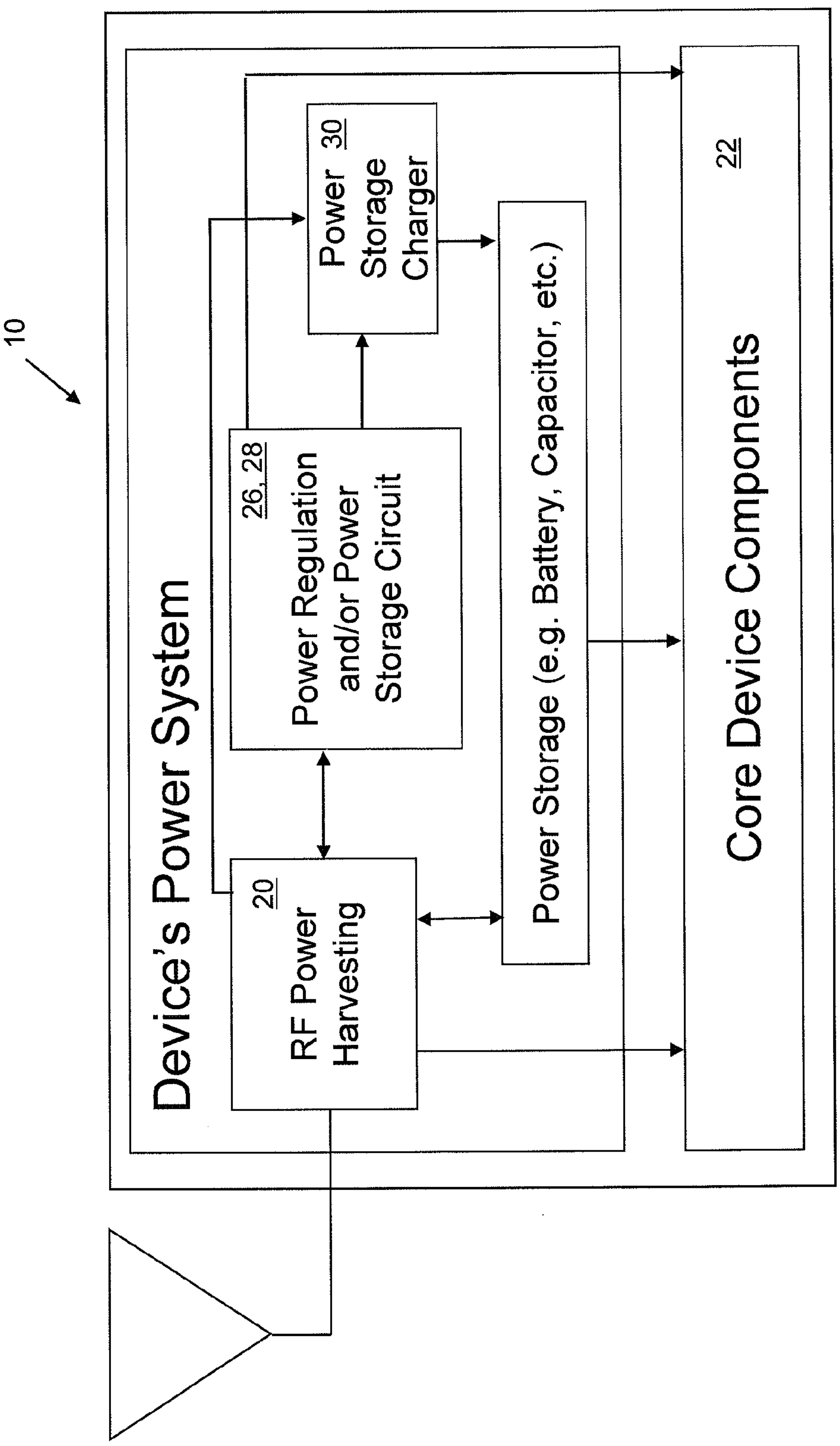
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 52

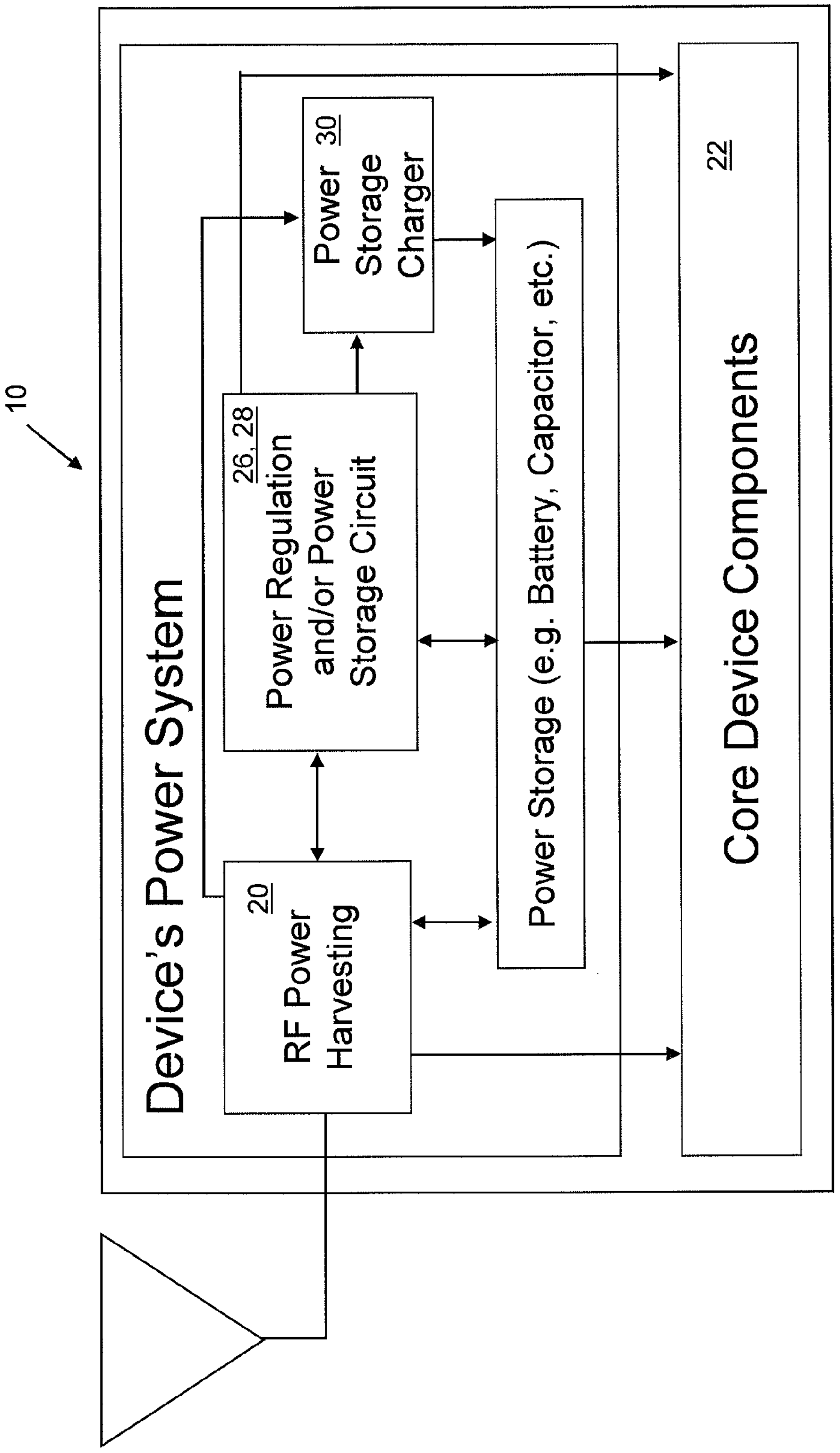
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30 Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 53

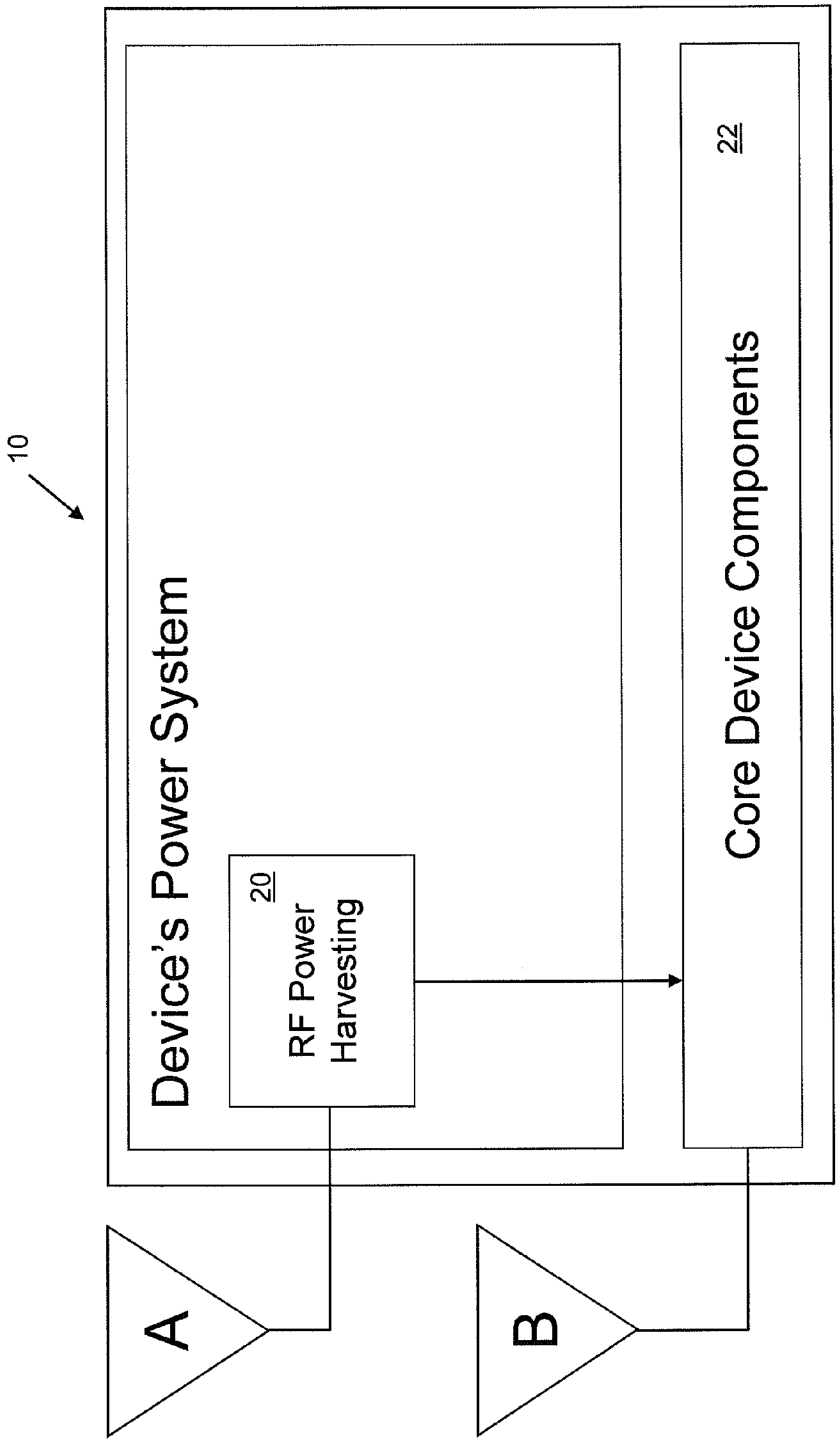
10
Device's Power System
RF Power Harvesting
20
Core Device Components
22
A
B

Fig. 54

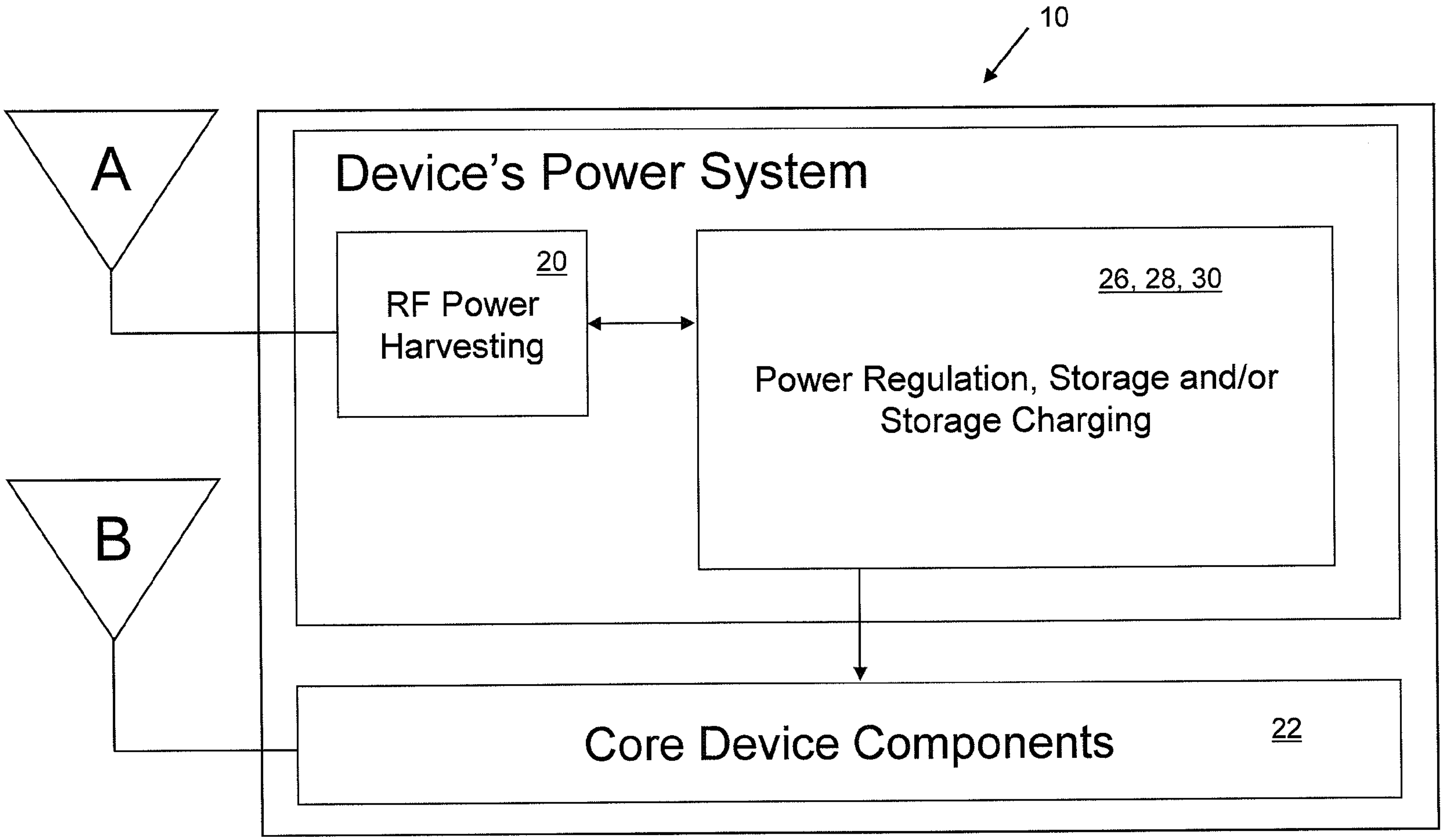
10
A
B
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22

Fig. 55

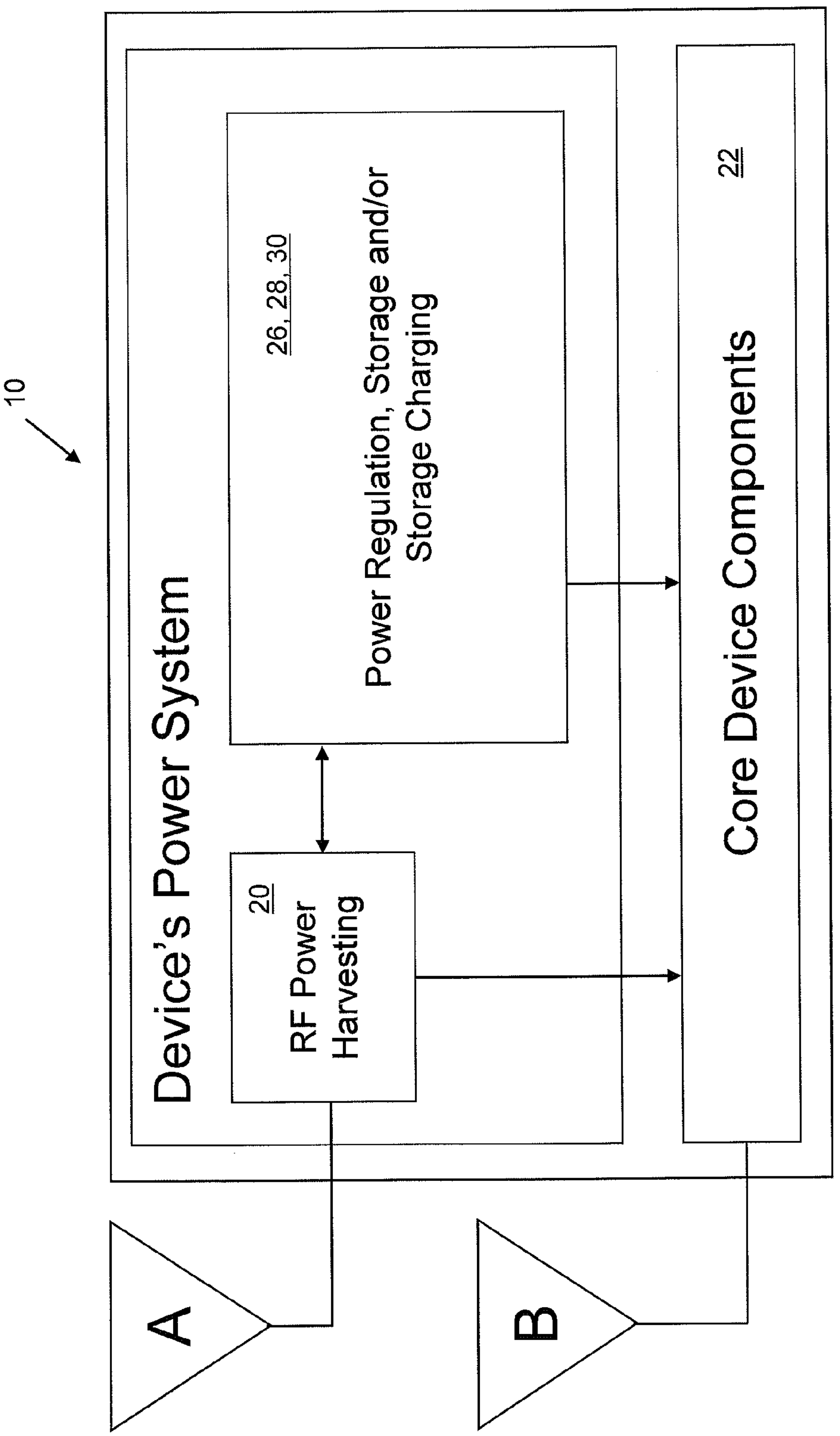
10
Device's Power System
RF Power Harvesting
20
Power Regulation, Storage and/or Storage Charging
26, 28, 30
Core Device Components
22
A
B

Fig. 56

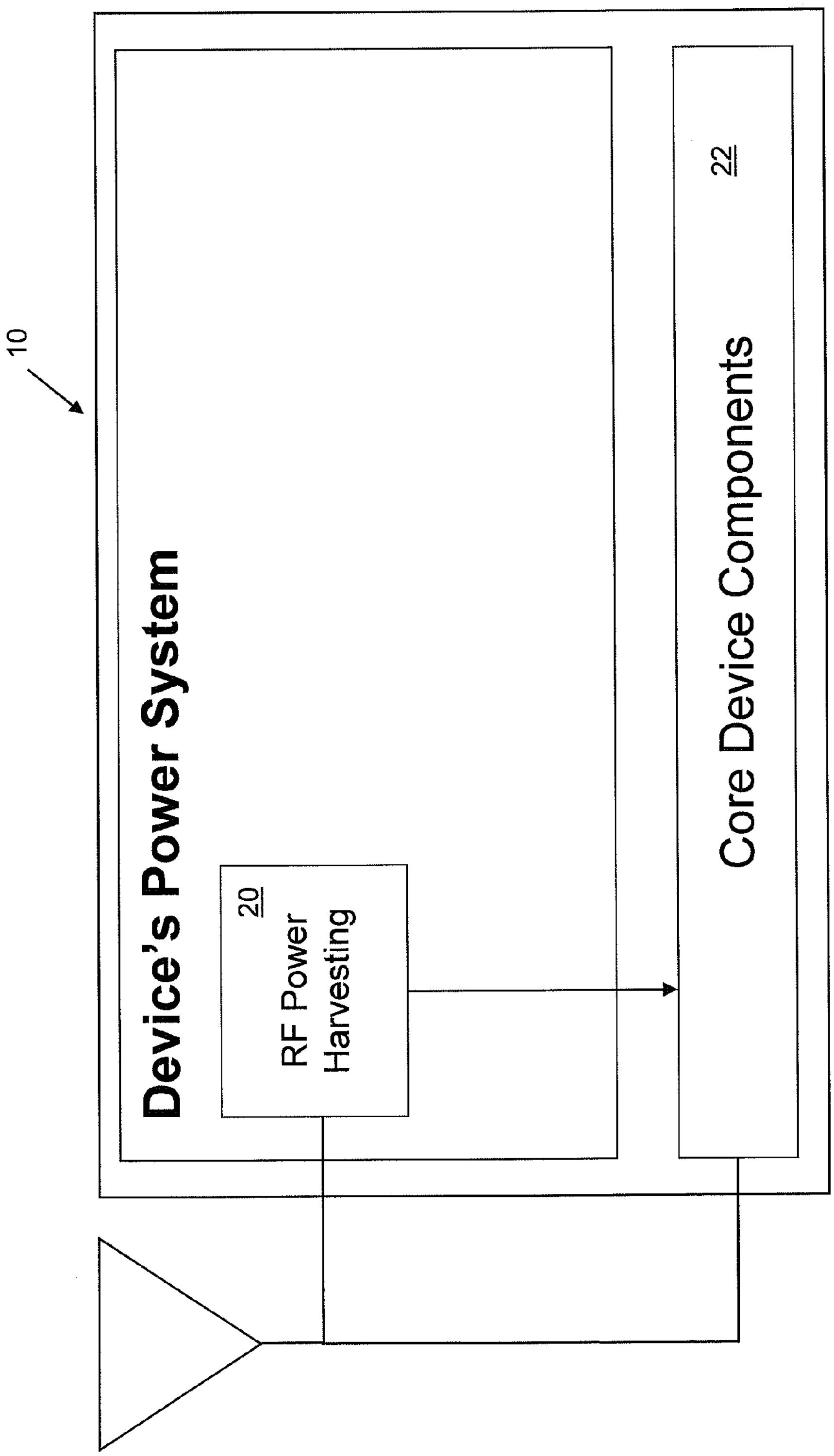
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22

Fig. 57

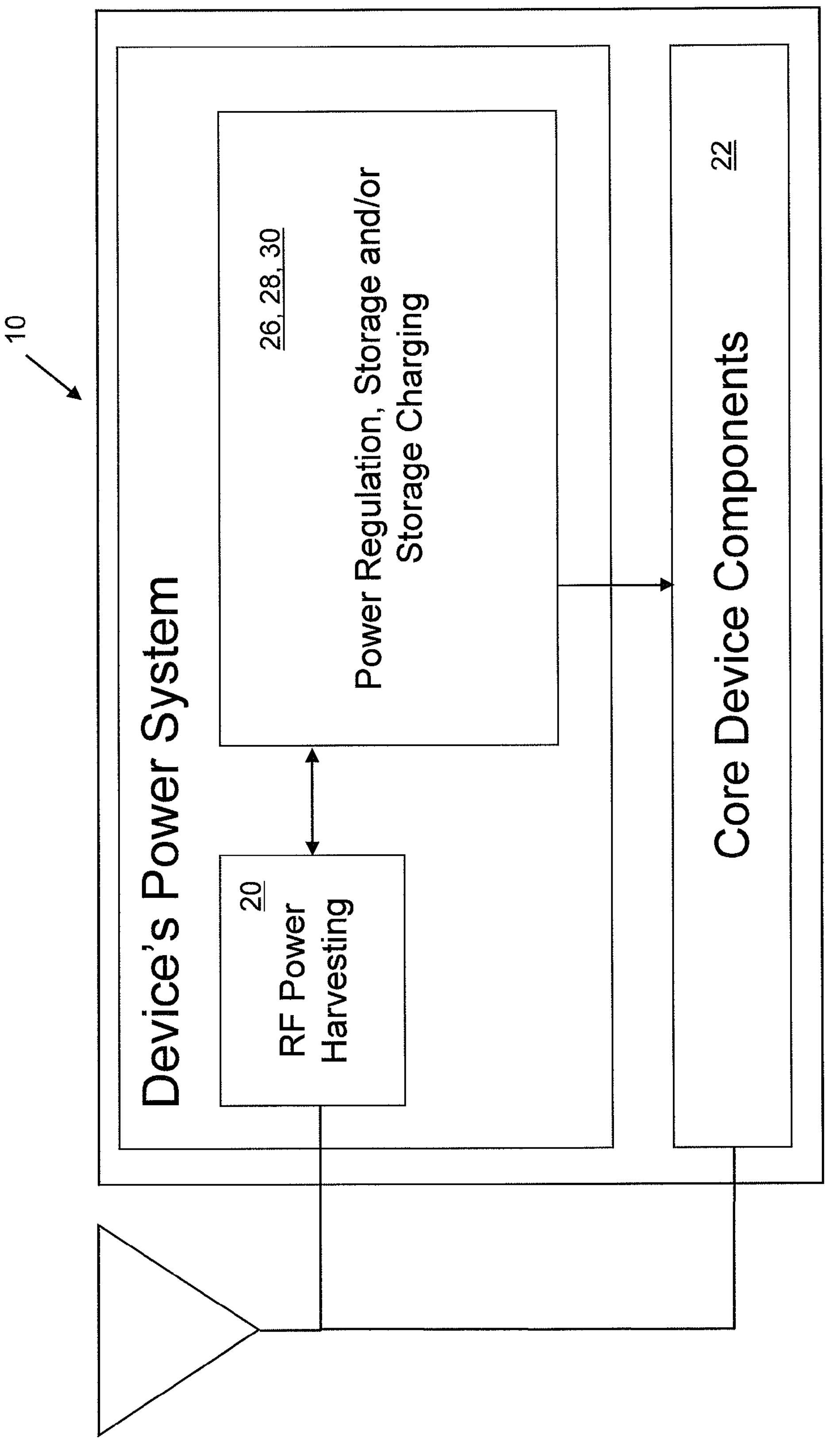
10
Device's Power System
RF Power Harvesting
20
Power Regulation, Storage and/or Storage Charging
26, 28, 30
Core Device Components
22

Fig. 58

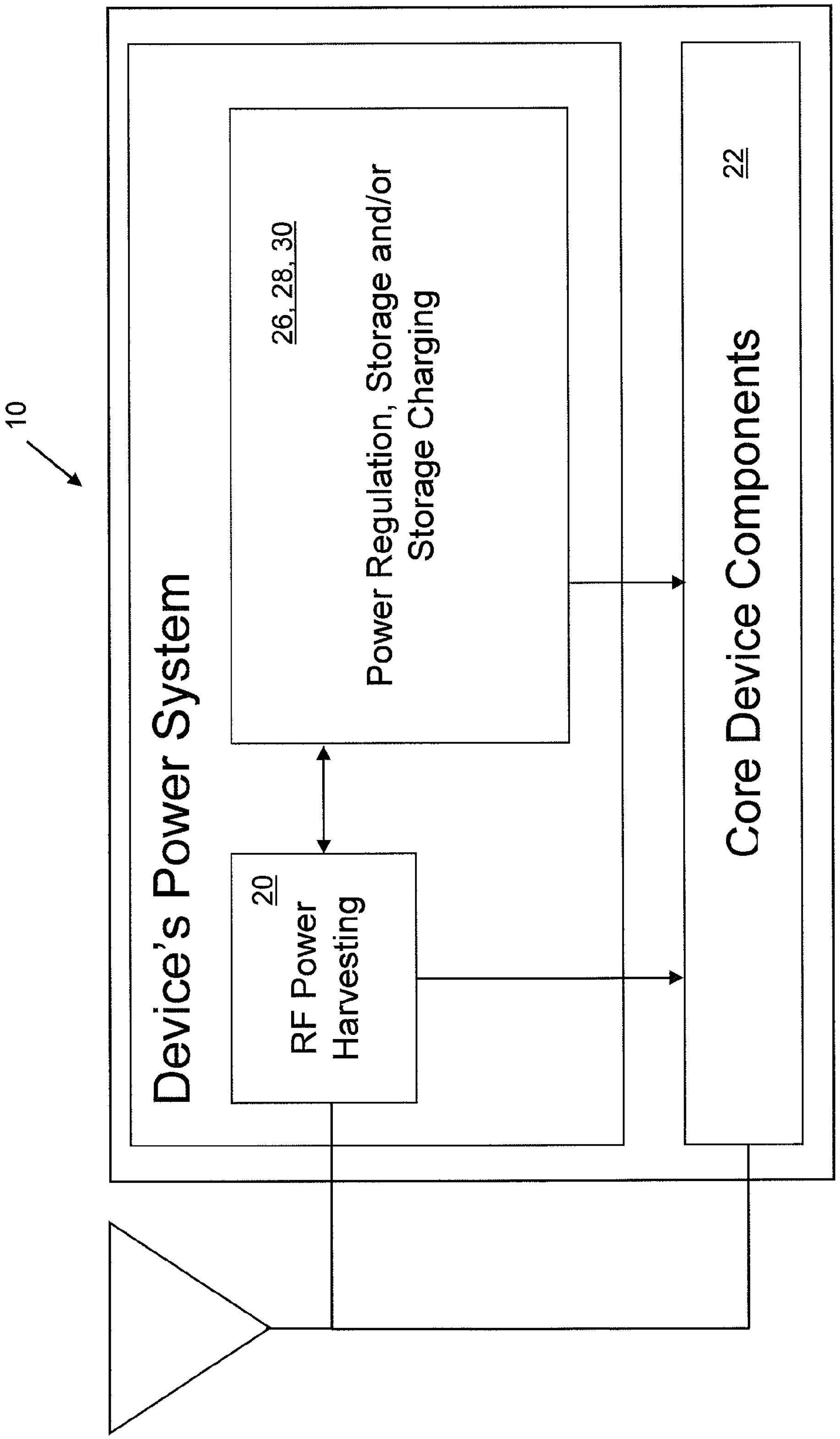
10
Device's Power System
Power Regulation, Storage and/or Storage Charging
26, 28, 30
RF Power Harvesting
20
Core Device Components
22

Fig. 59

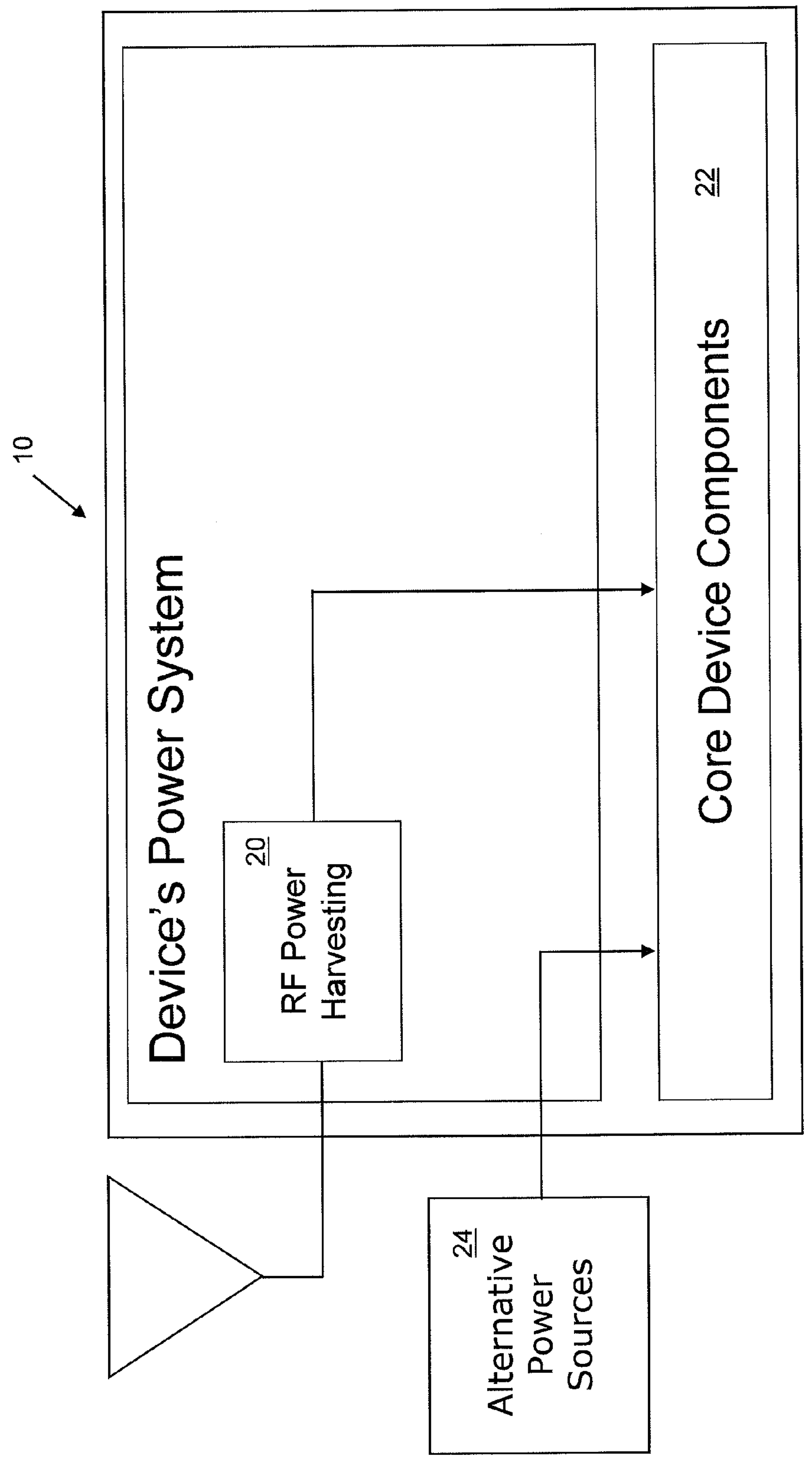
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 60

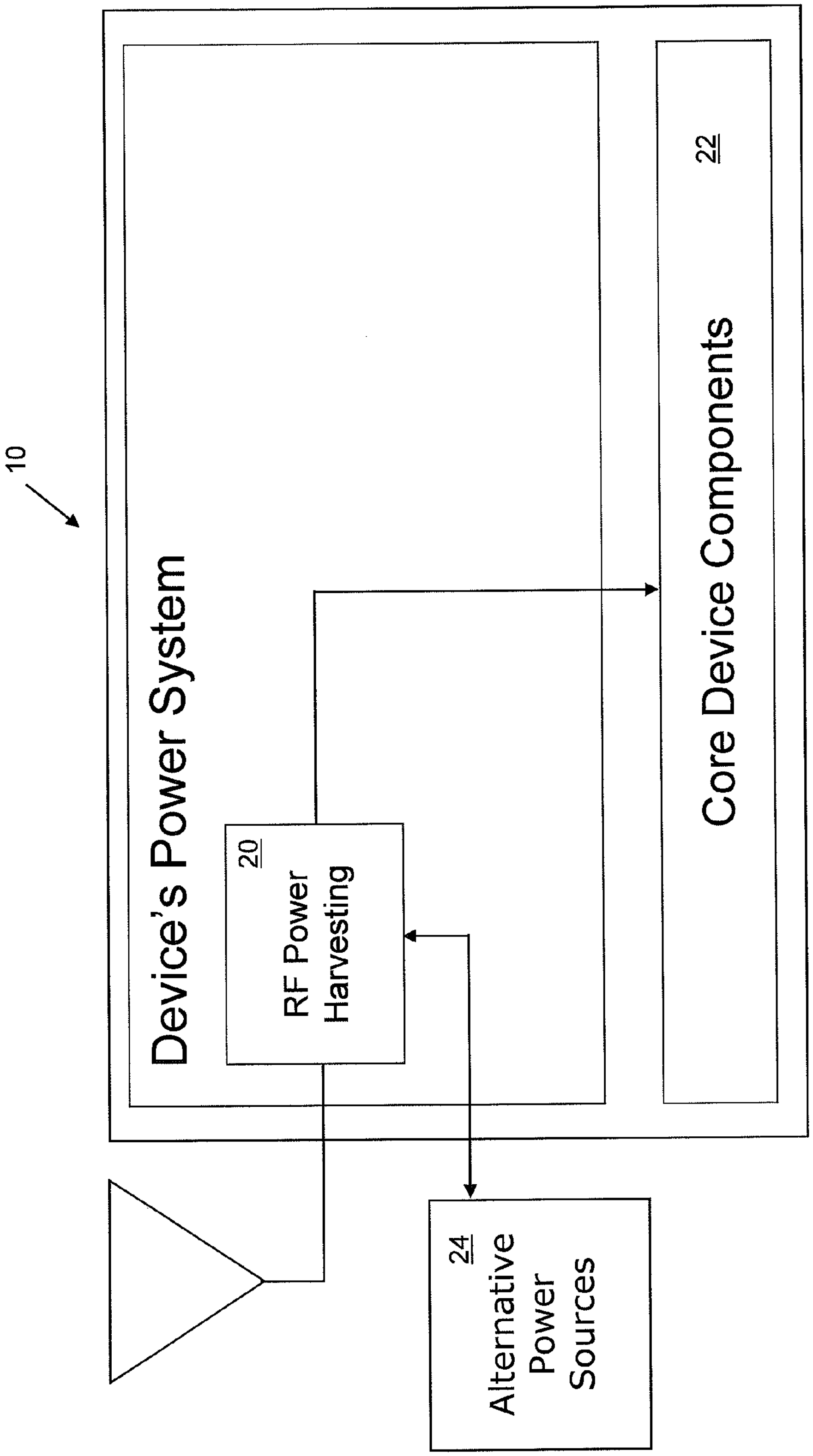
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 61

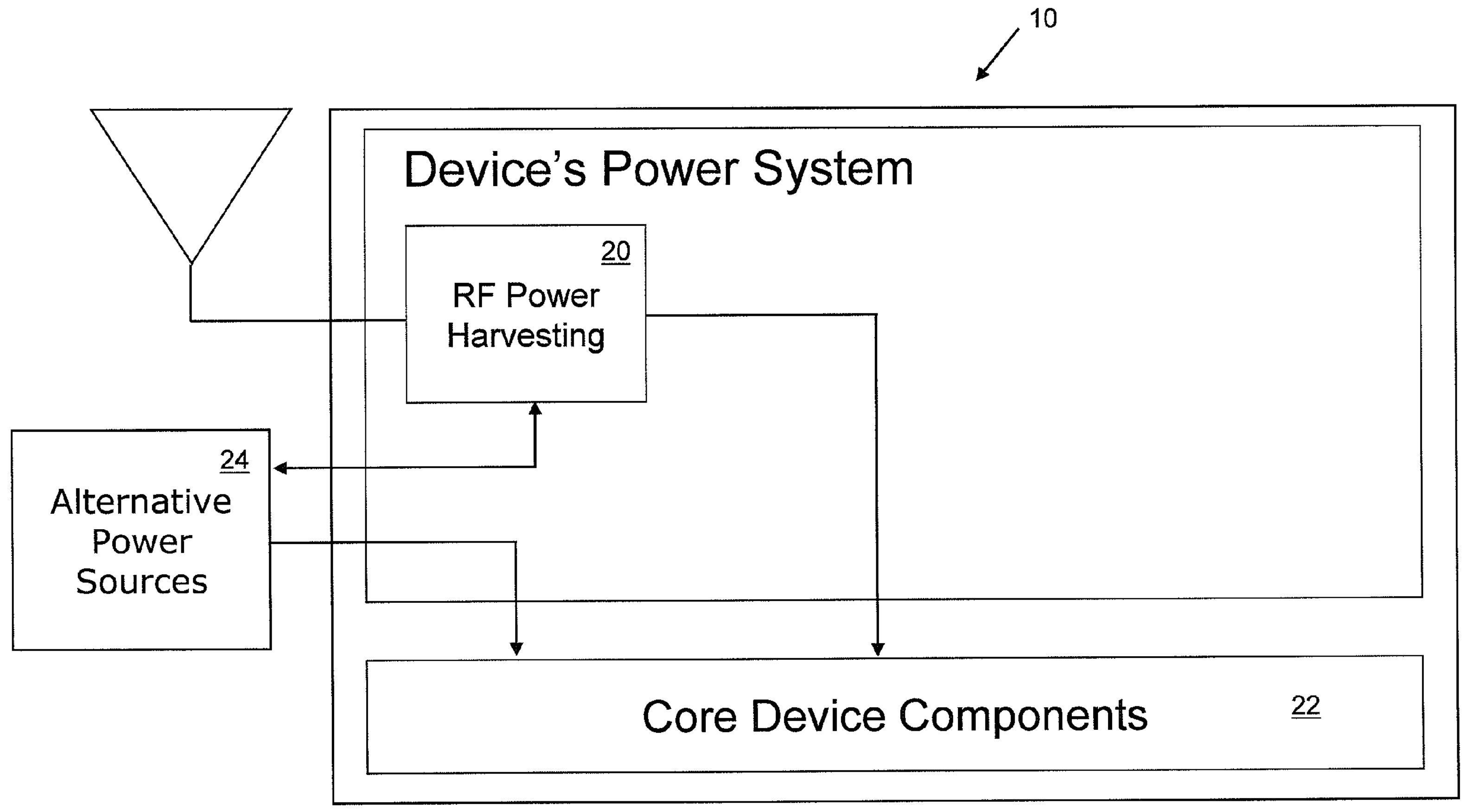
10
Device's Power System
20
RF Power Harvesting
24
Alternative Power Sources
Core Device Components
22

Fig. 62

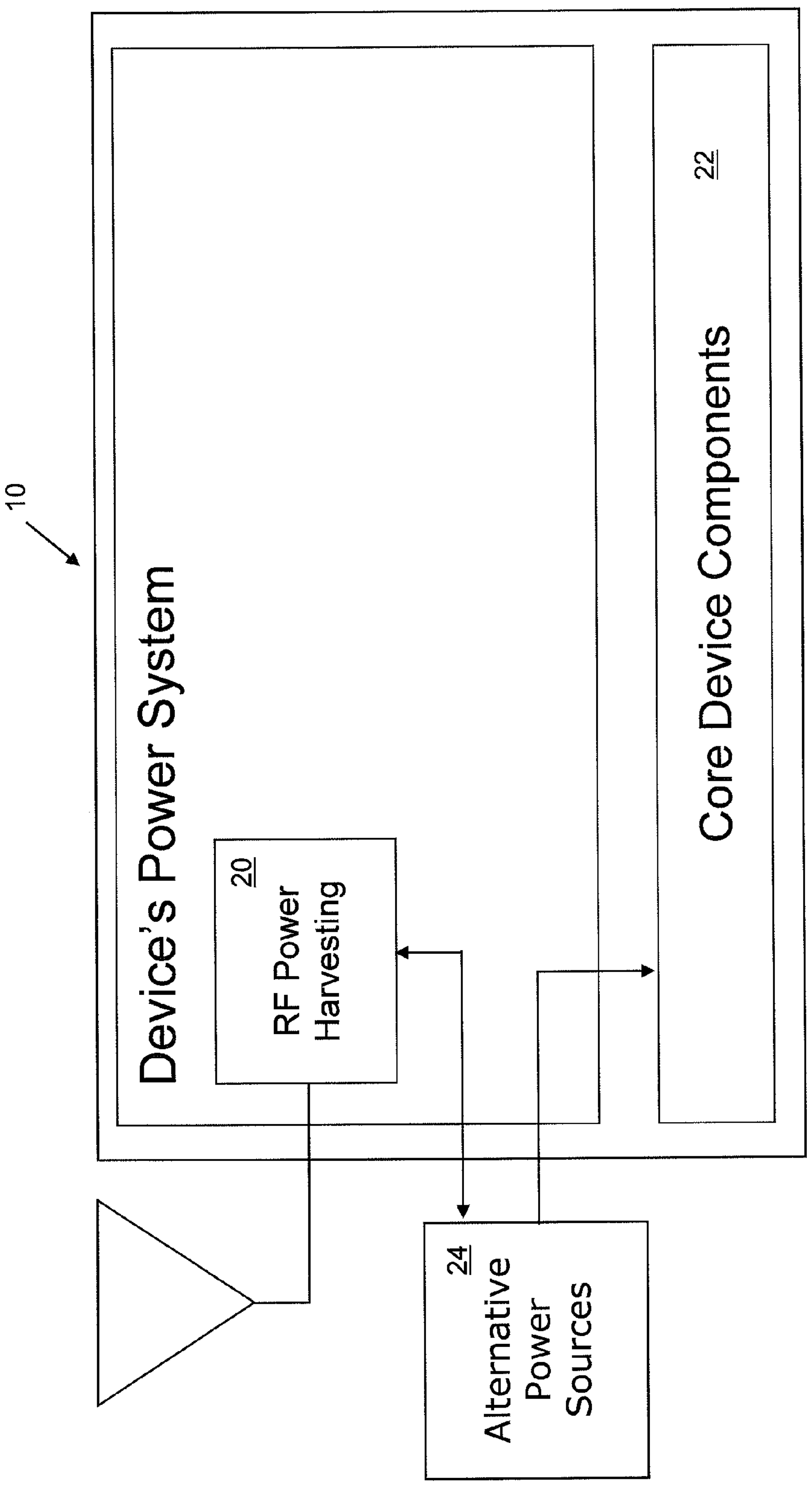
10
Device's Power System
20
RF Power Harvesting
22
Core Device Components
24
Alternative Power Sources

Fig. 63

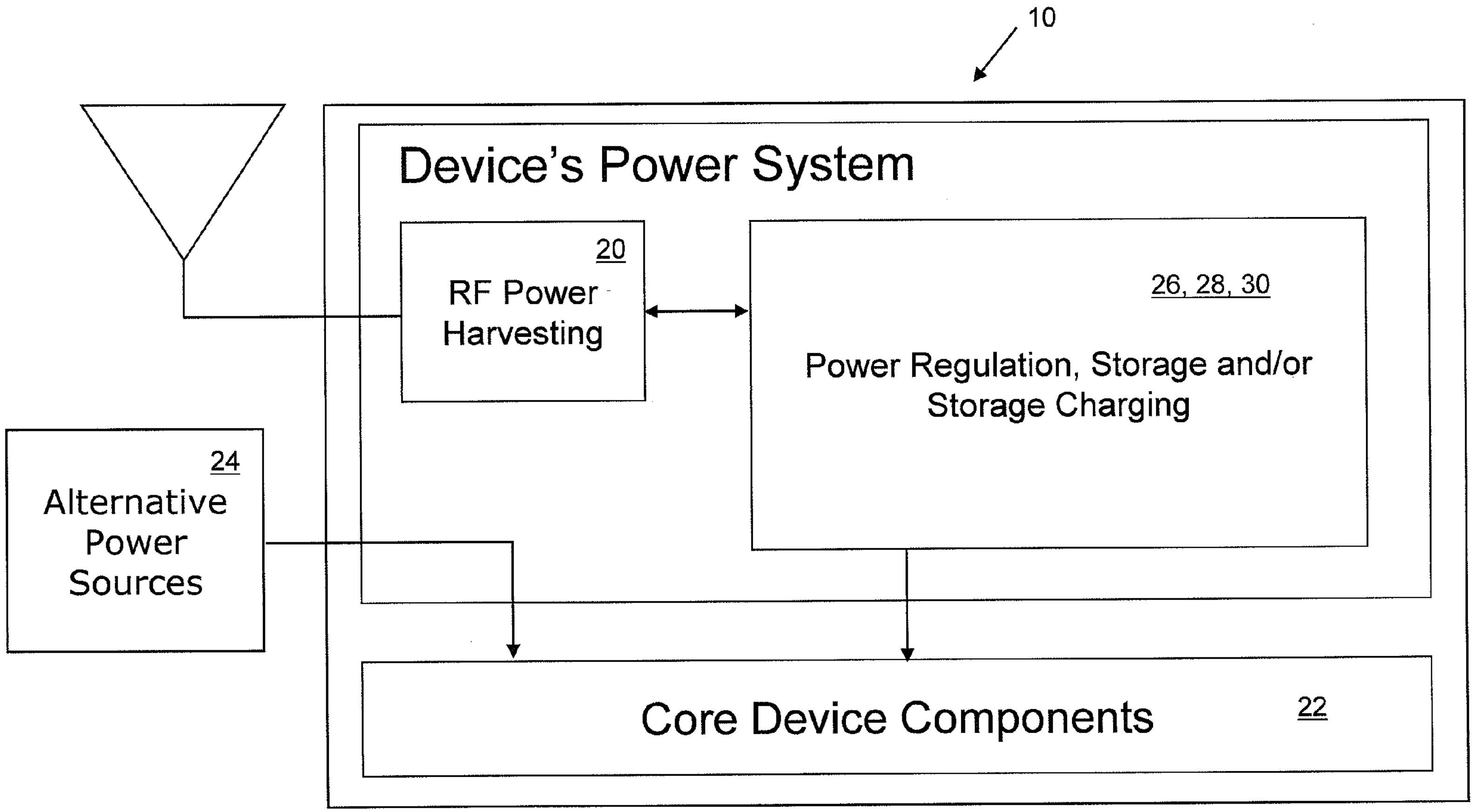
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 64

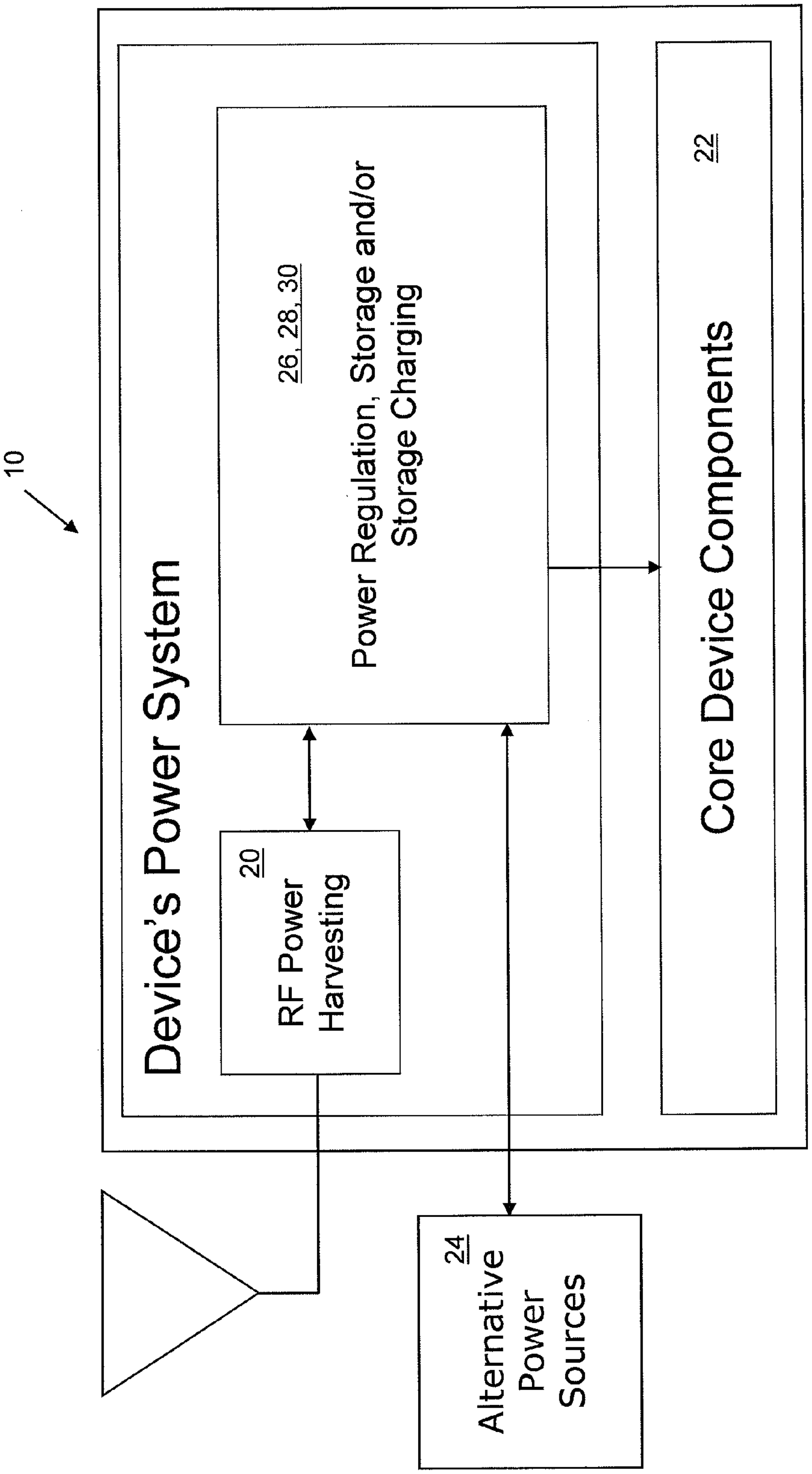
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 65

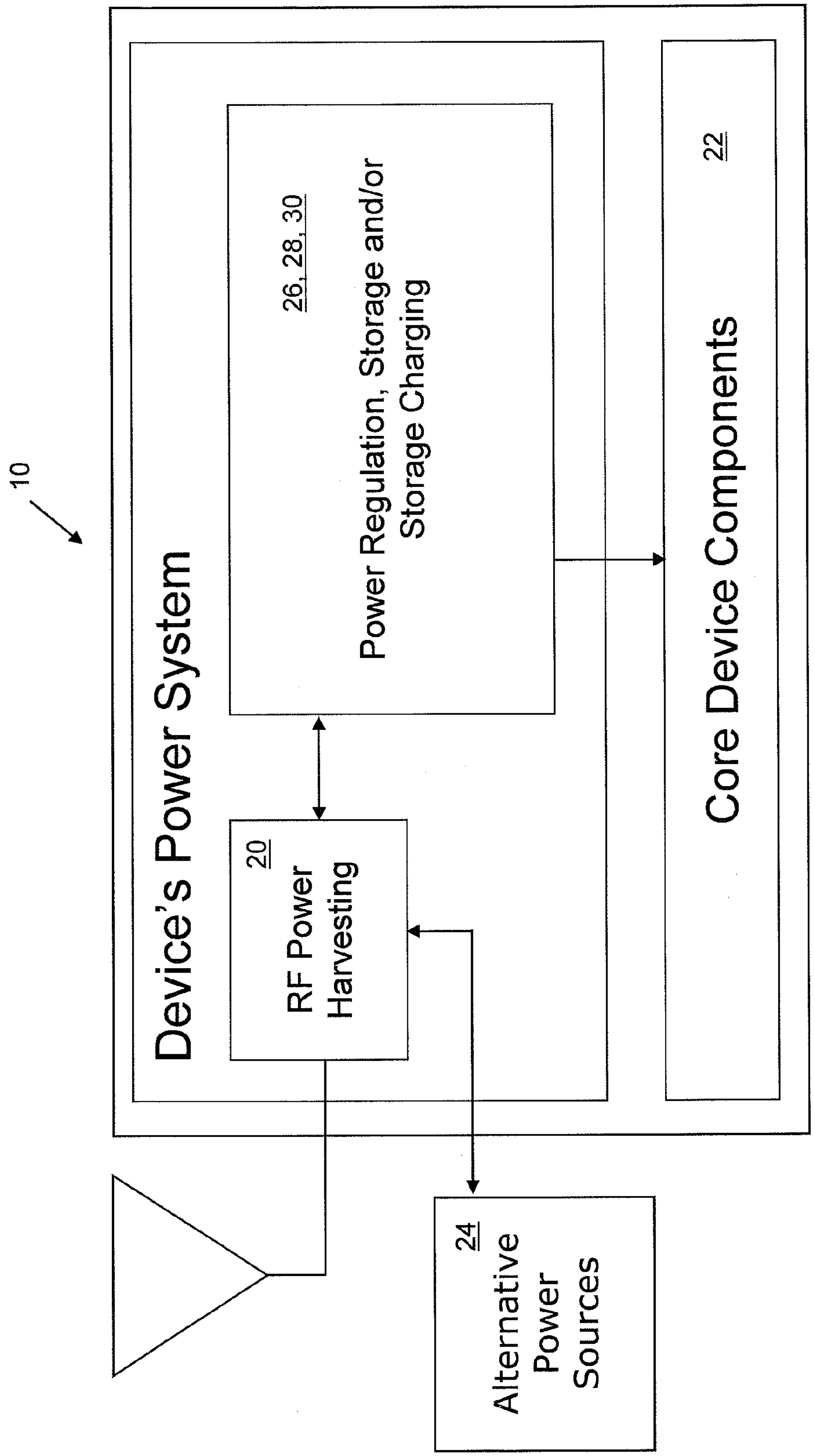
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 66

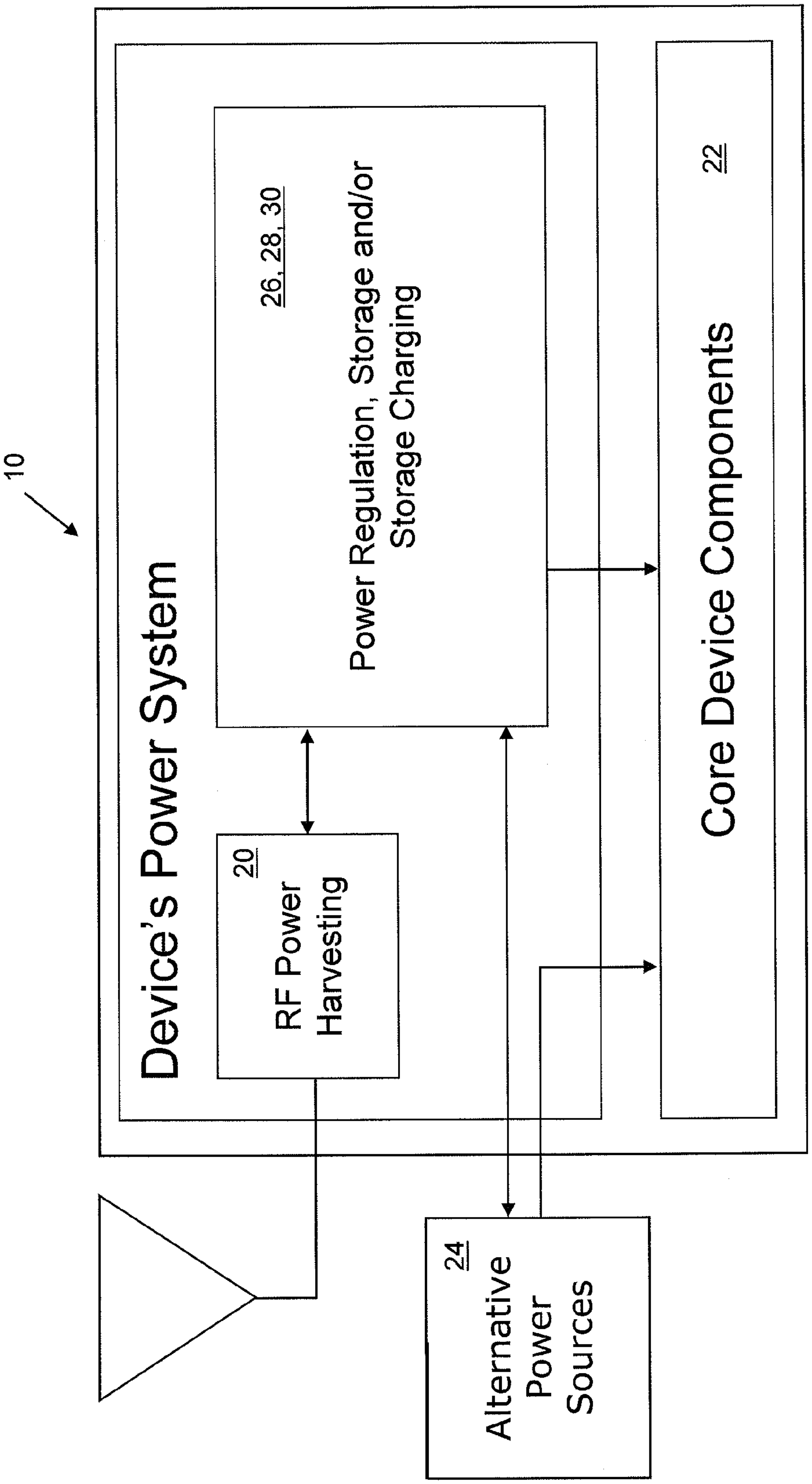
10
Device's Power System
Power Regulation, Storage and/or Storage Charging
26, 28, 30
RF Power Harvesting
20
Core Device Components
22
Alternative Power Sources
24

Fig. 67

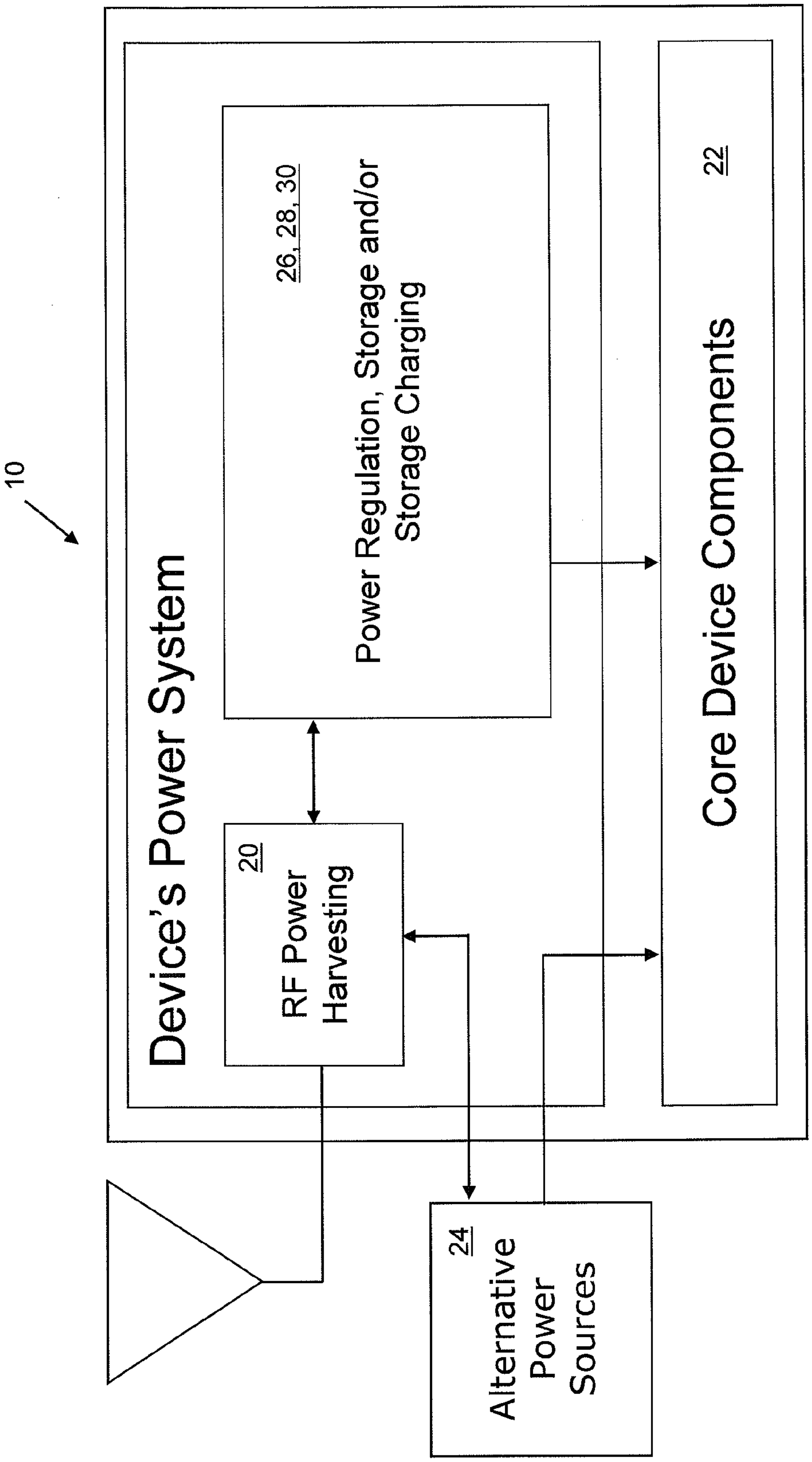
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
22
Core Device Components
24
Alternative Power Sources

Fig. 68

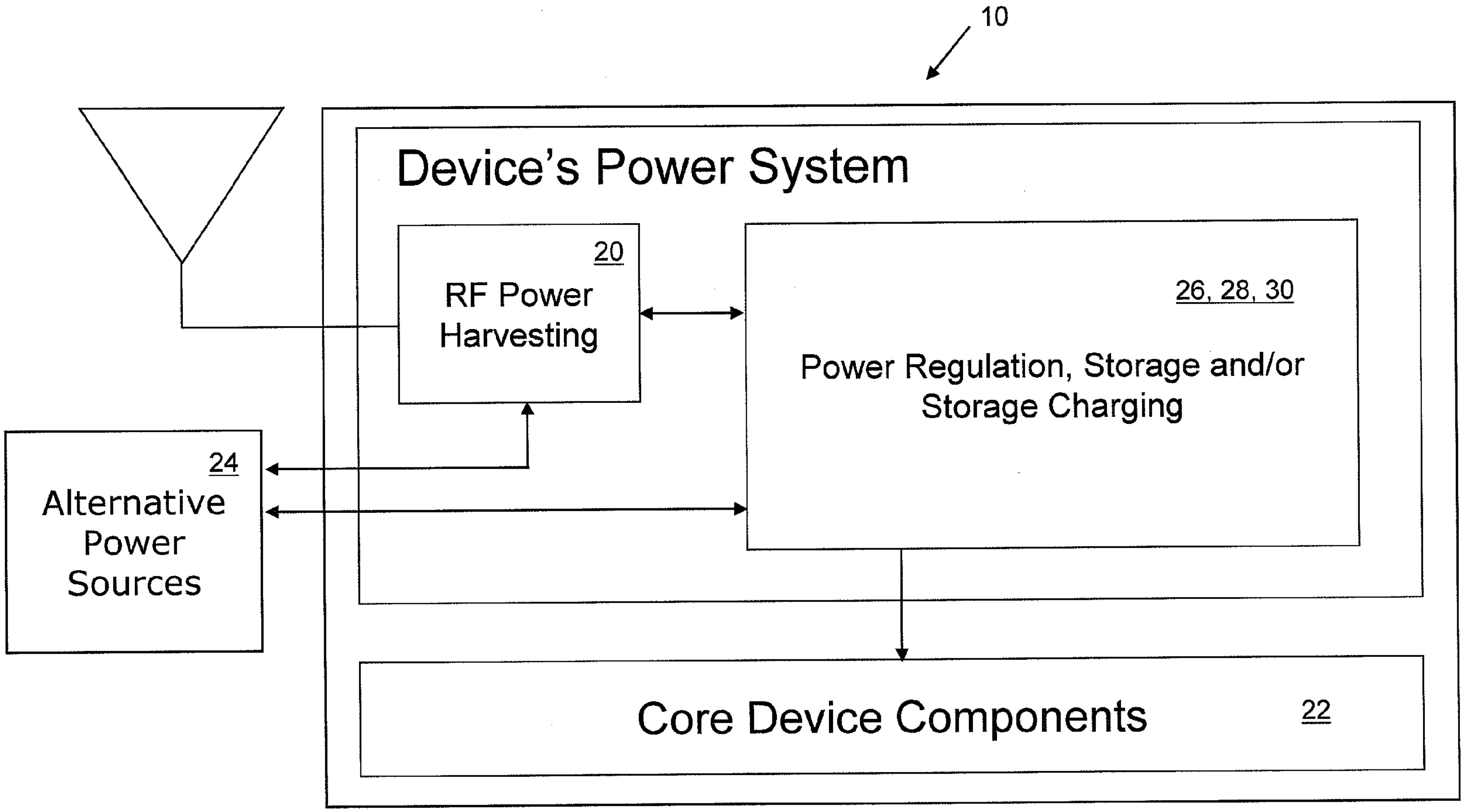
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 69

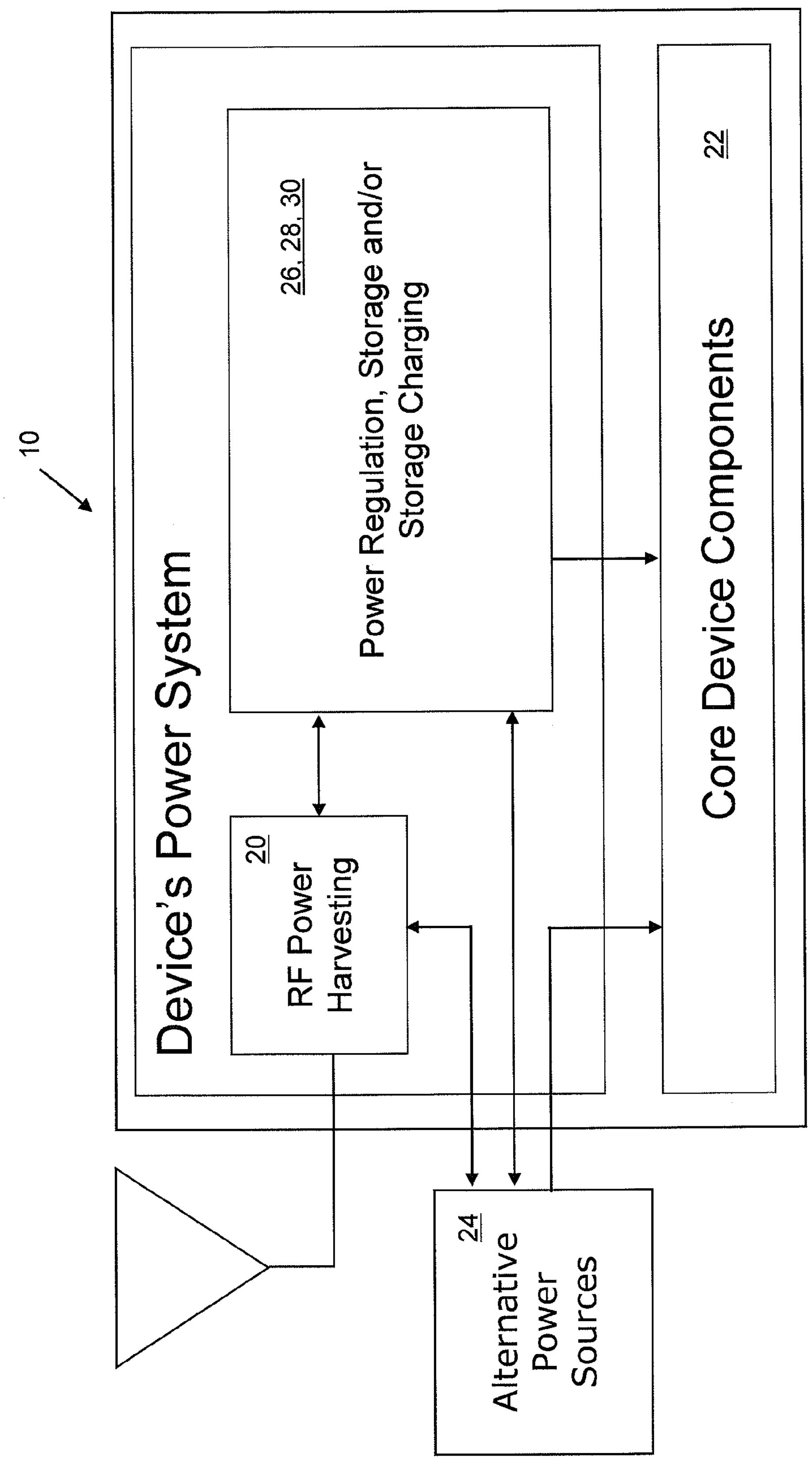
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
22
Core Device Components
24
Alternative Power Sources

Fig. 70

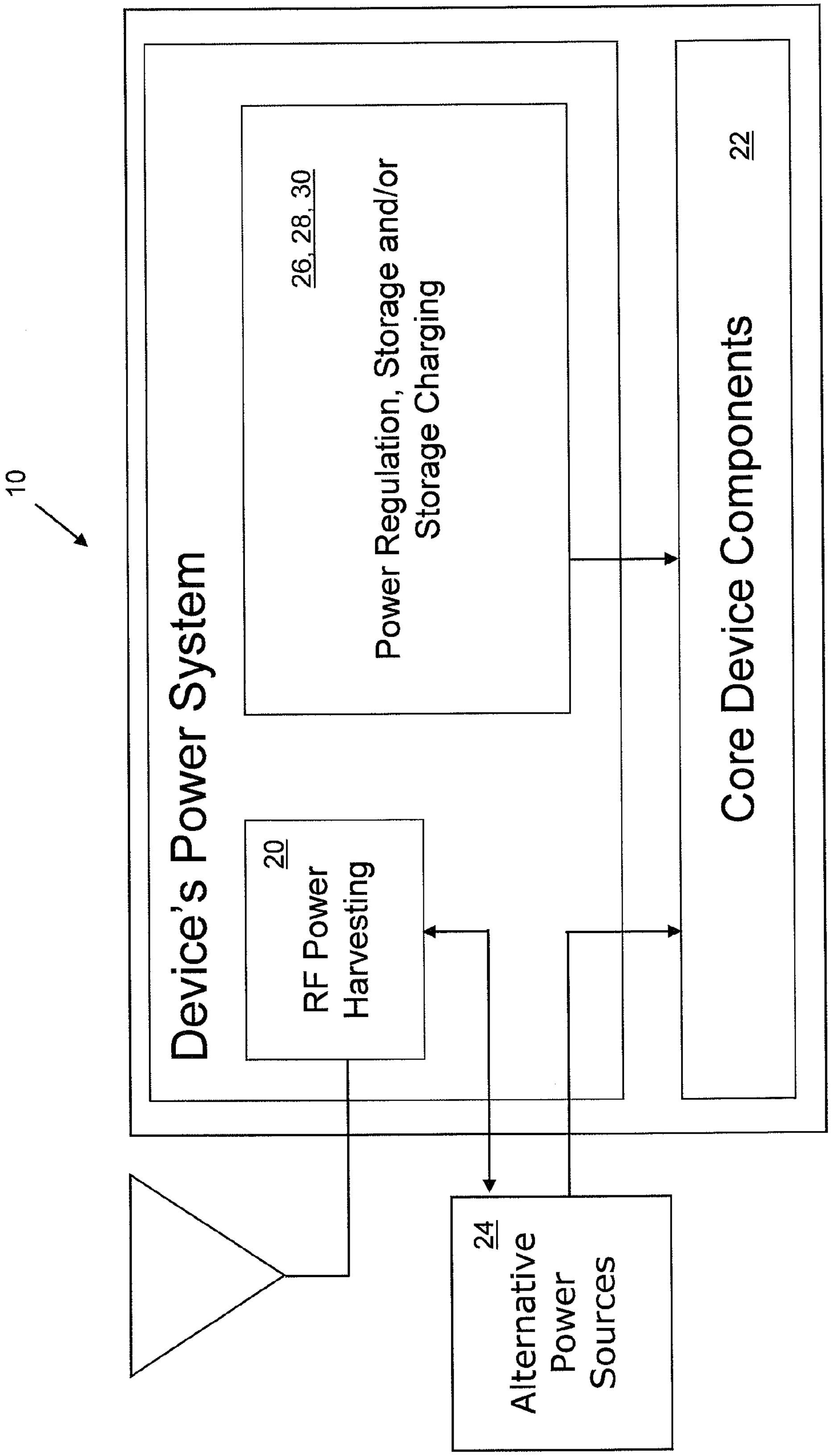
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 71

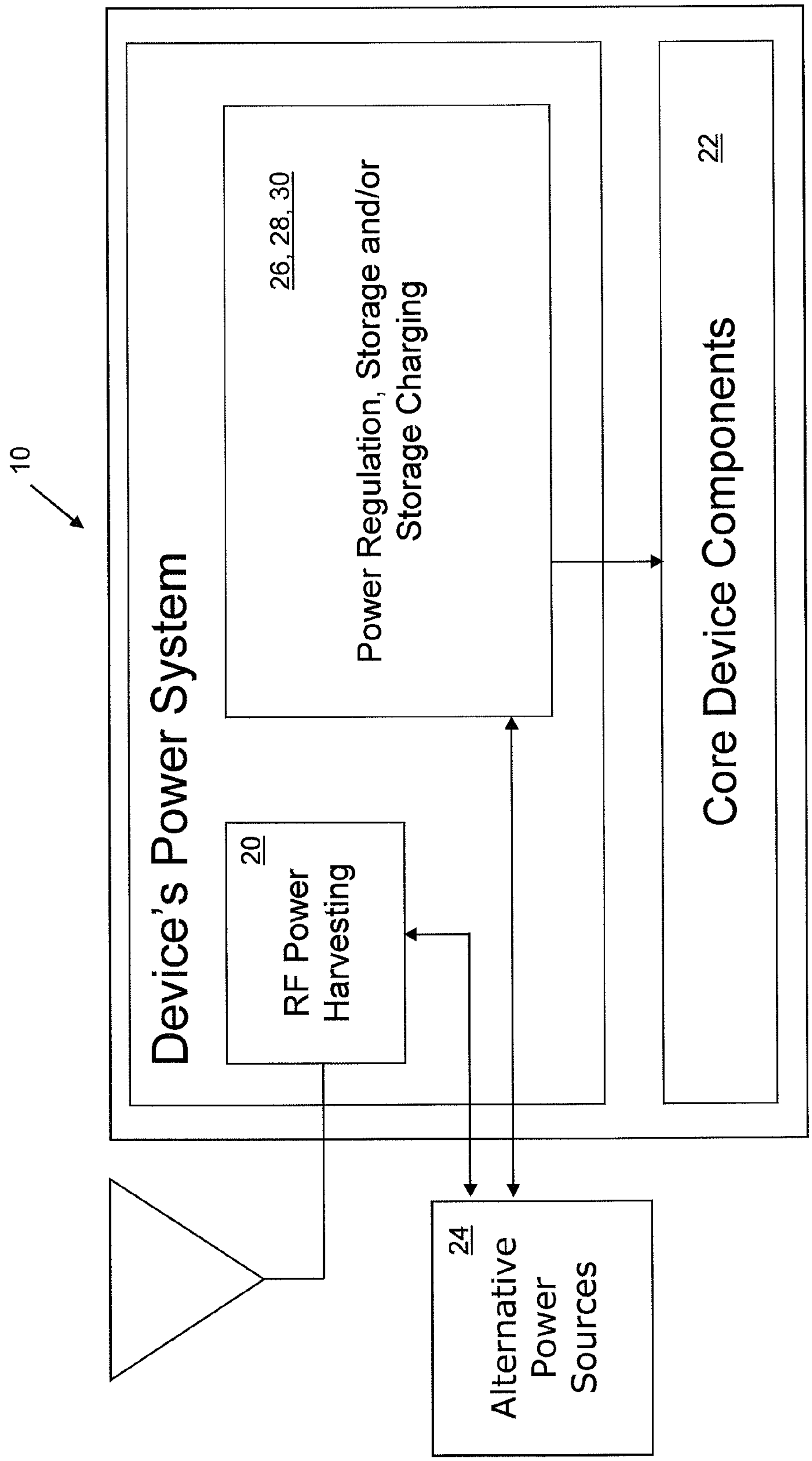
10
Device's Power System
RF Power Harvesting
20
Power Regulation, Storage and/or Storage Charging
26, 28, 30
Core Device Components
22
Alternative Power Sources
24

Fig. 72

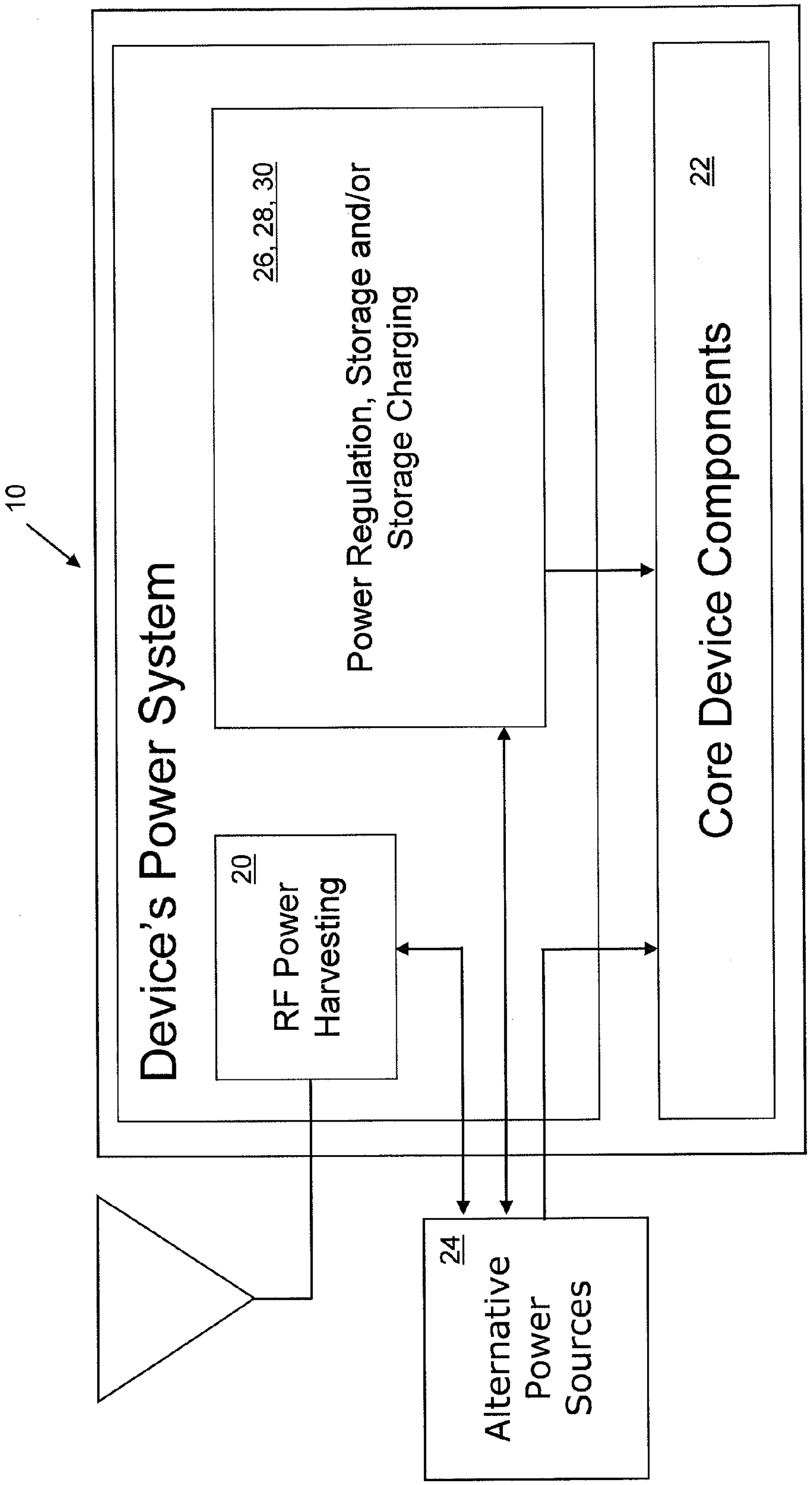
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 73

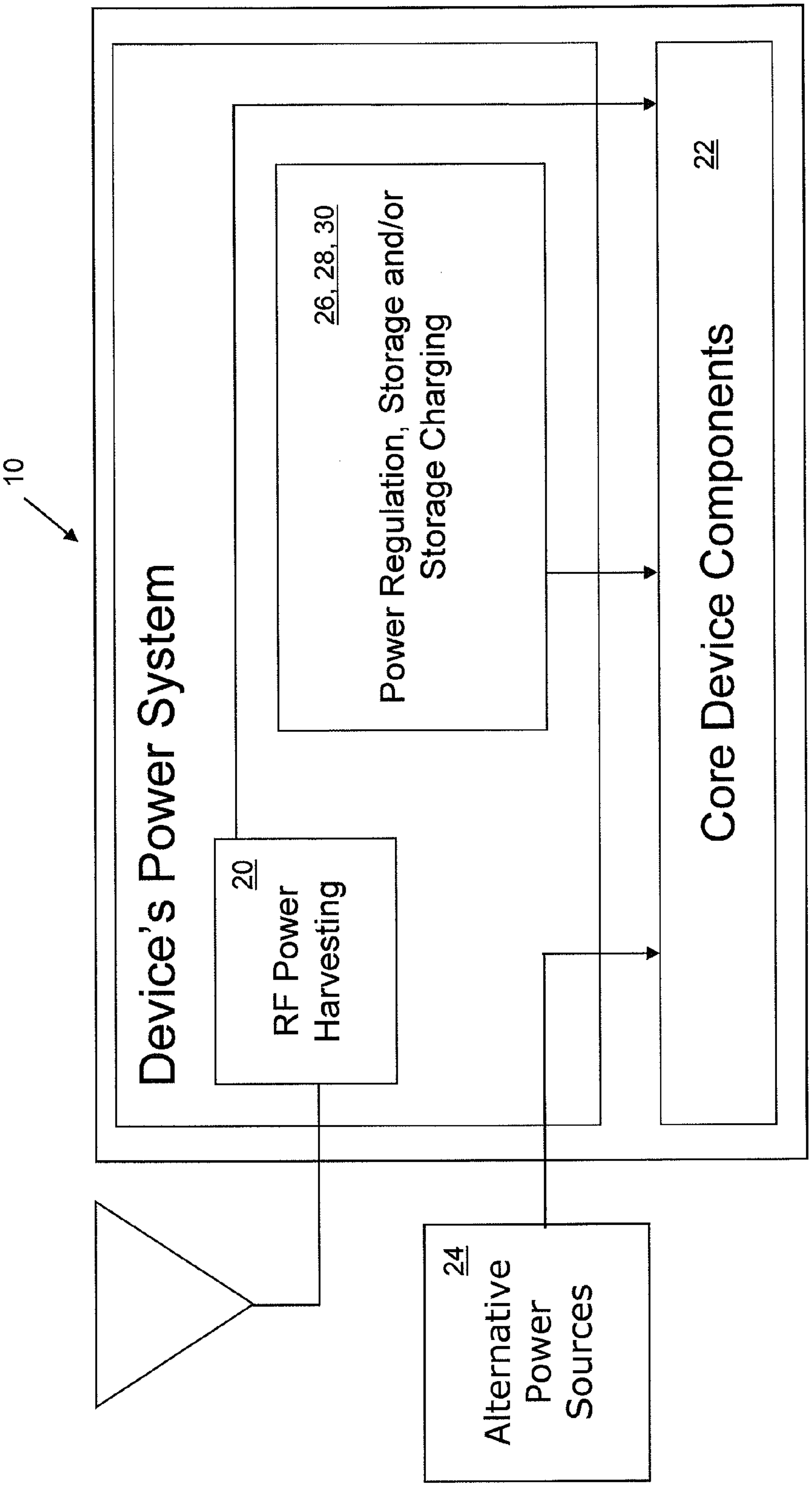
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 74

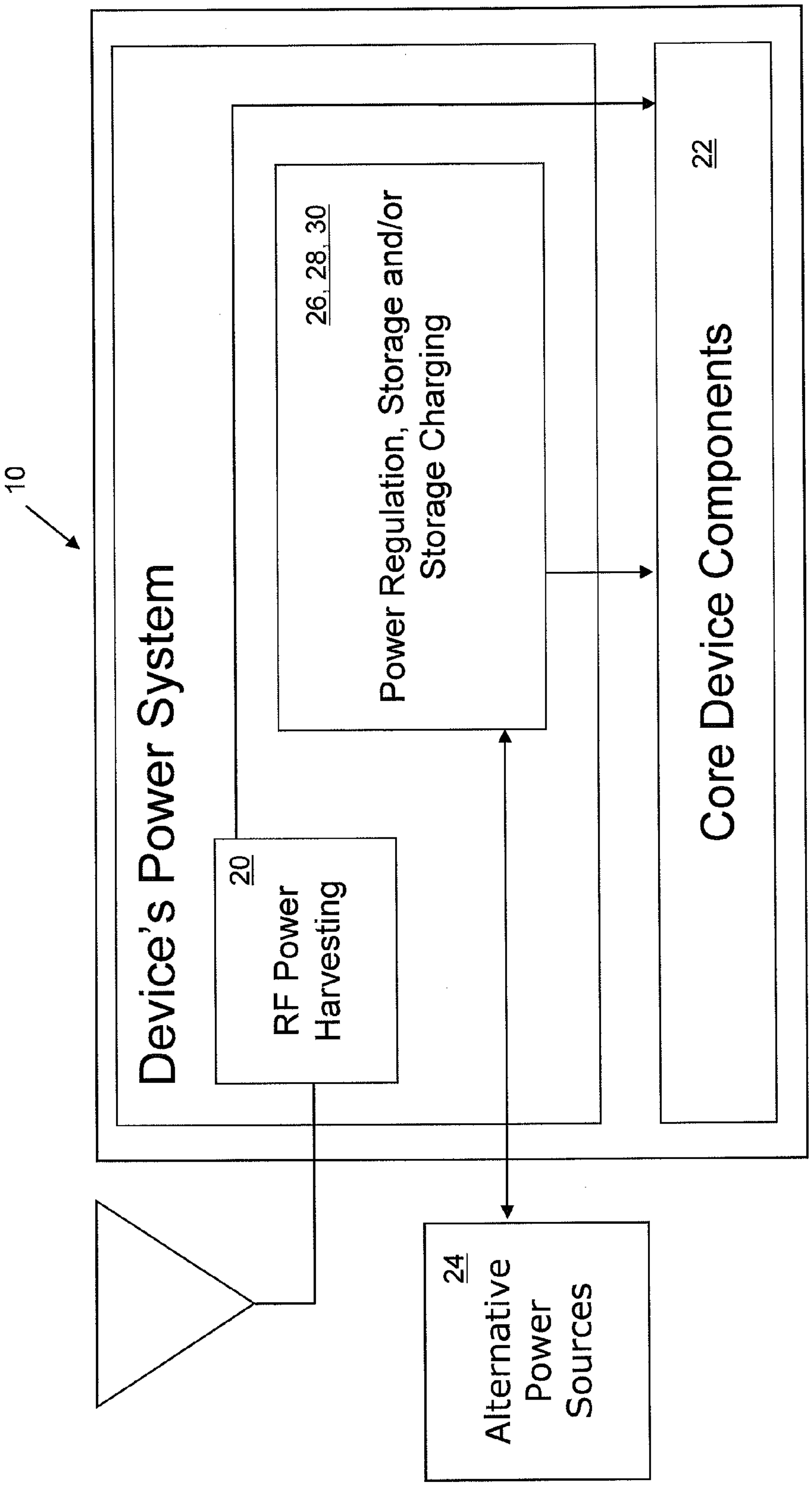
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 75

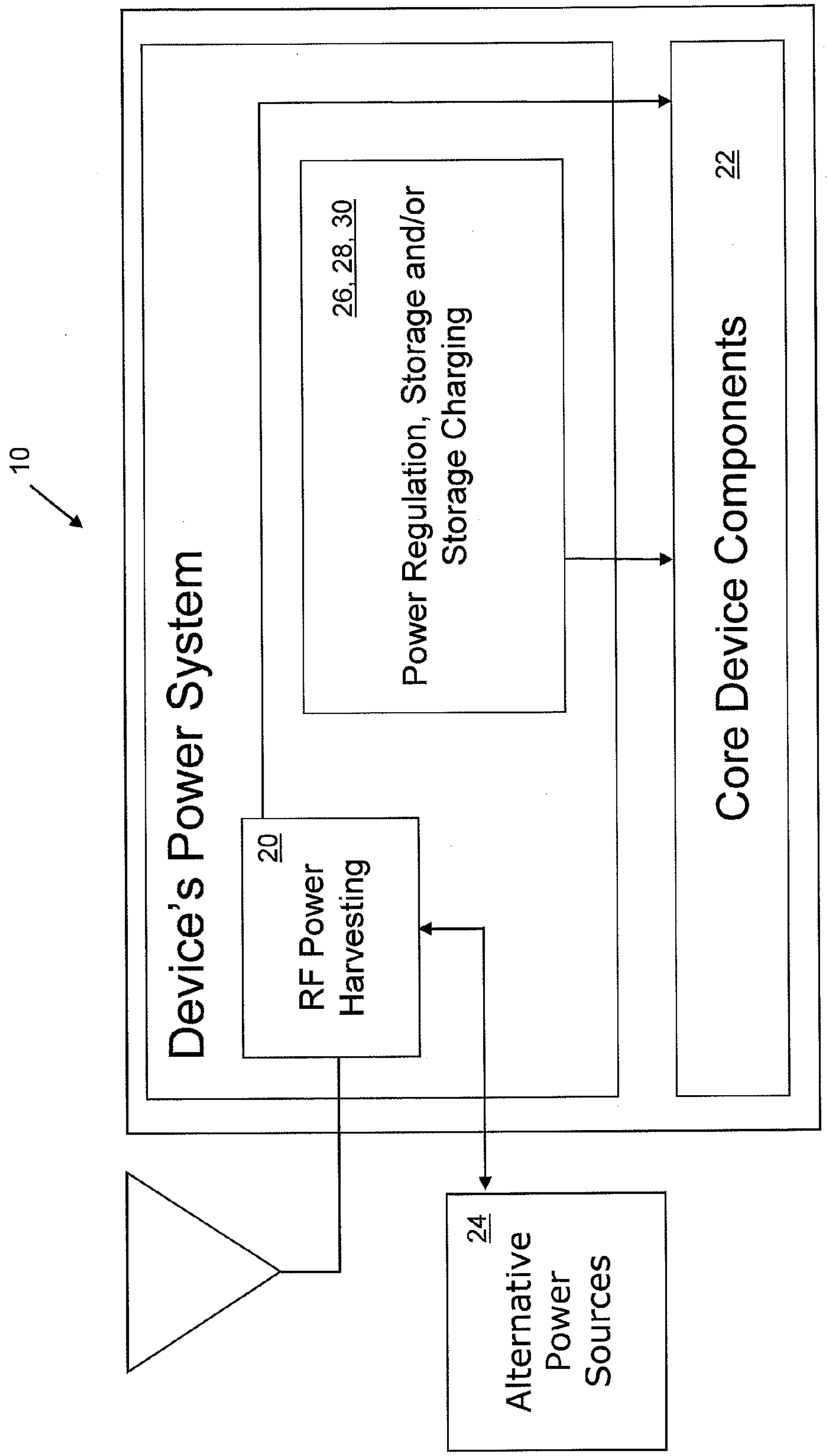
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 76

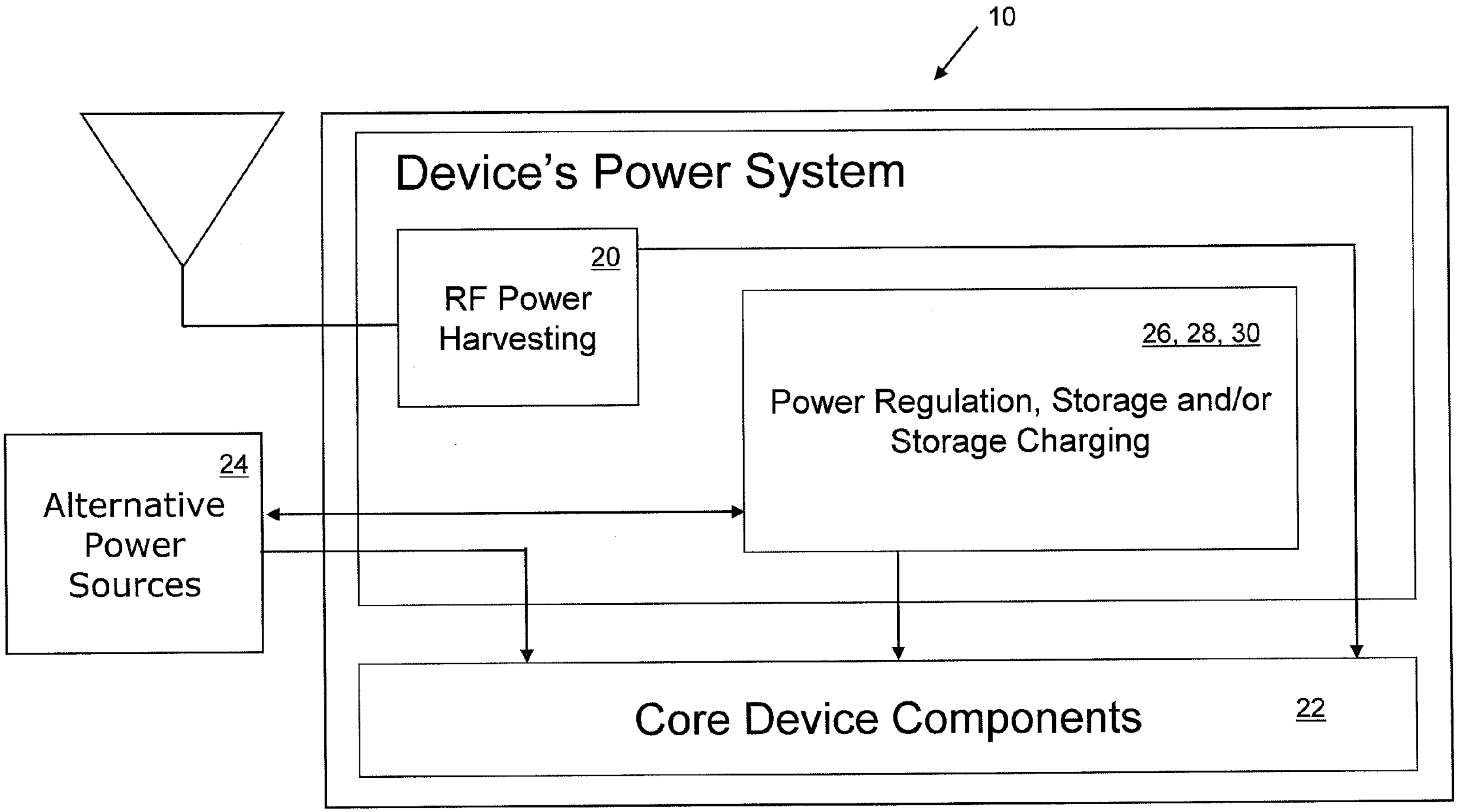
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 77

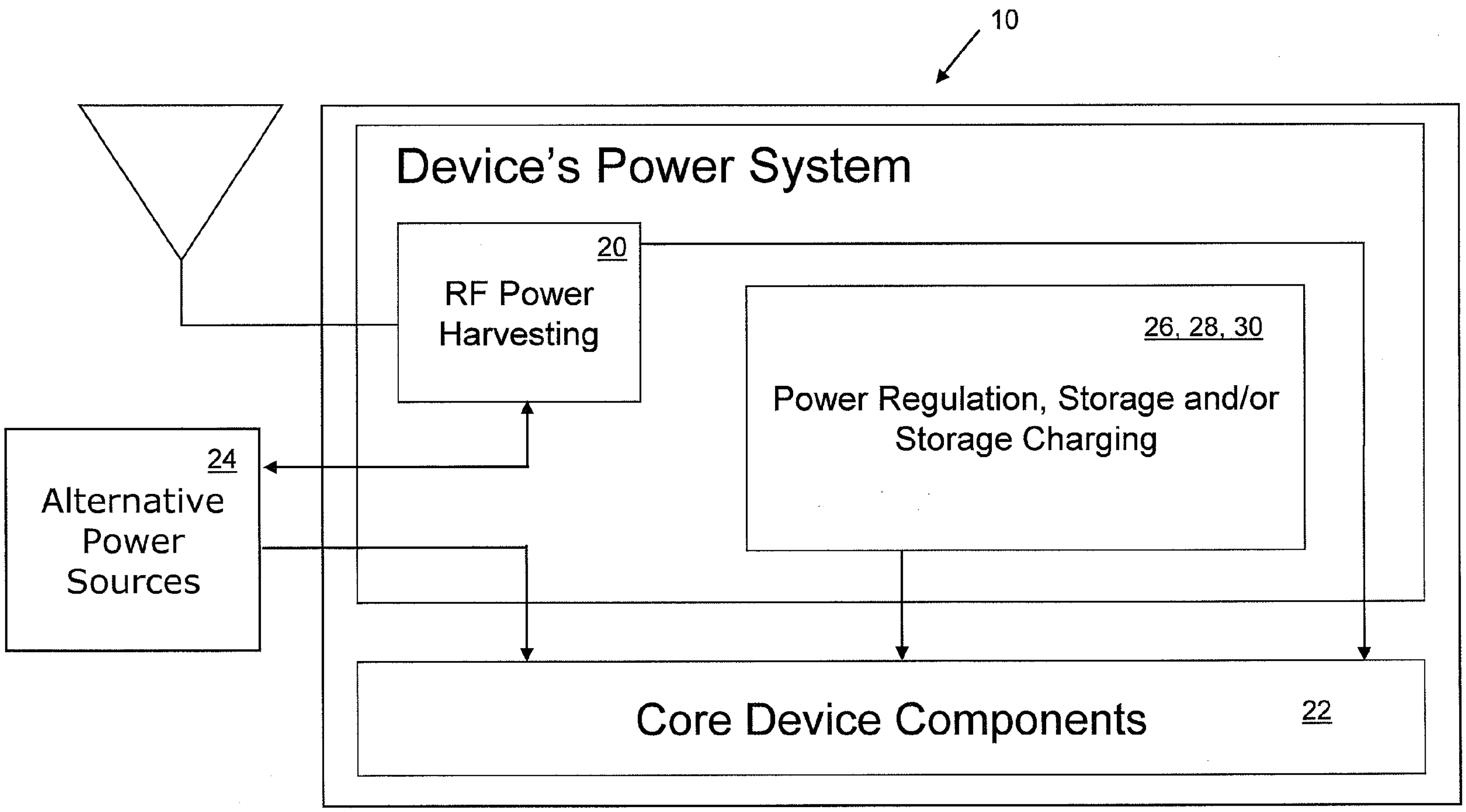
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 78

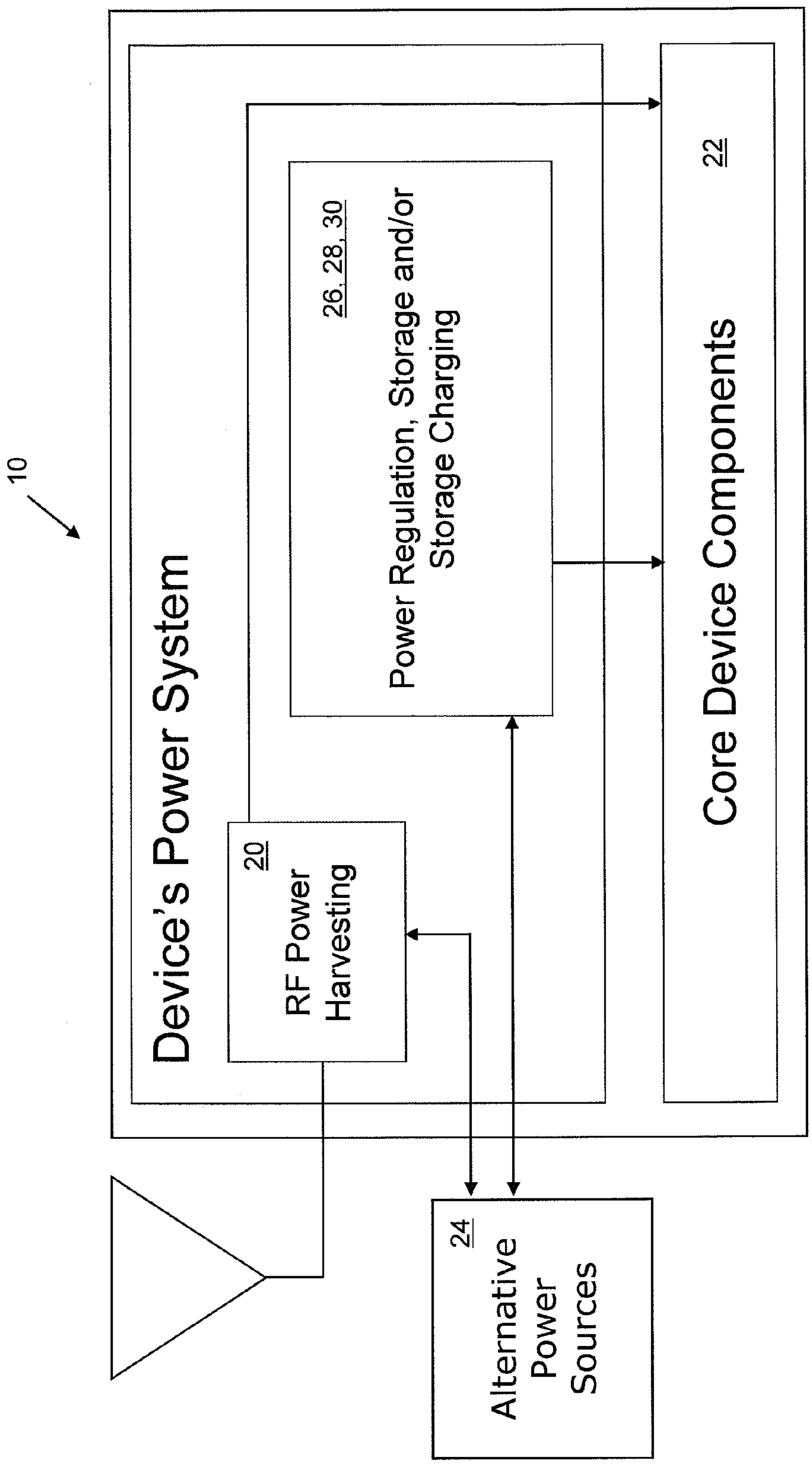
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 79

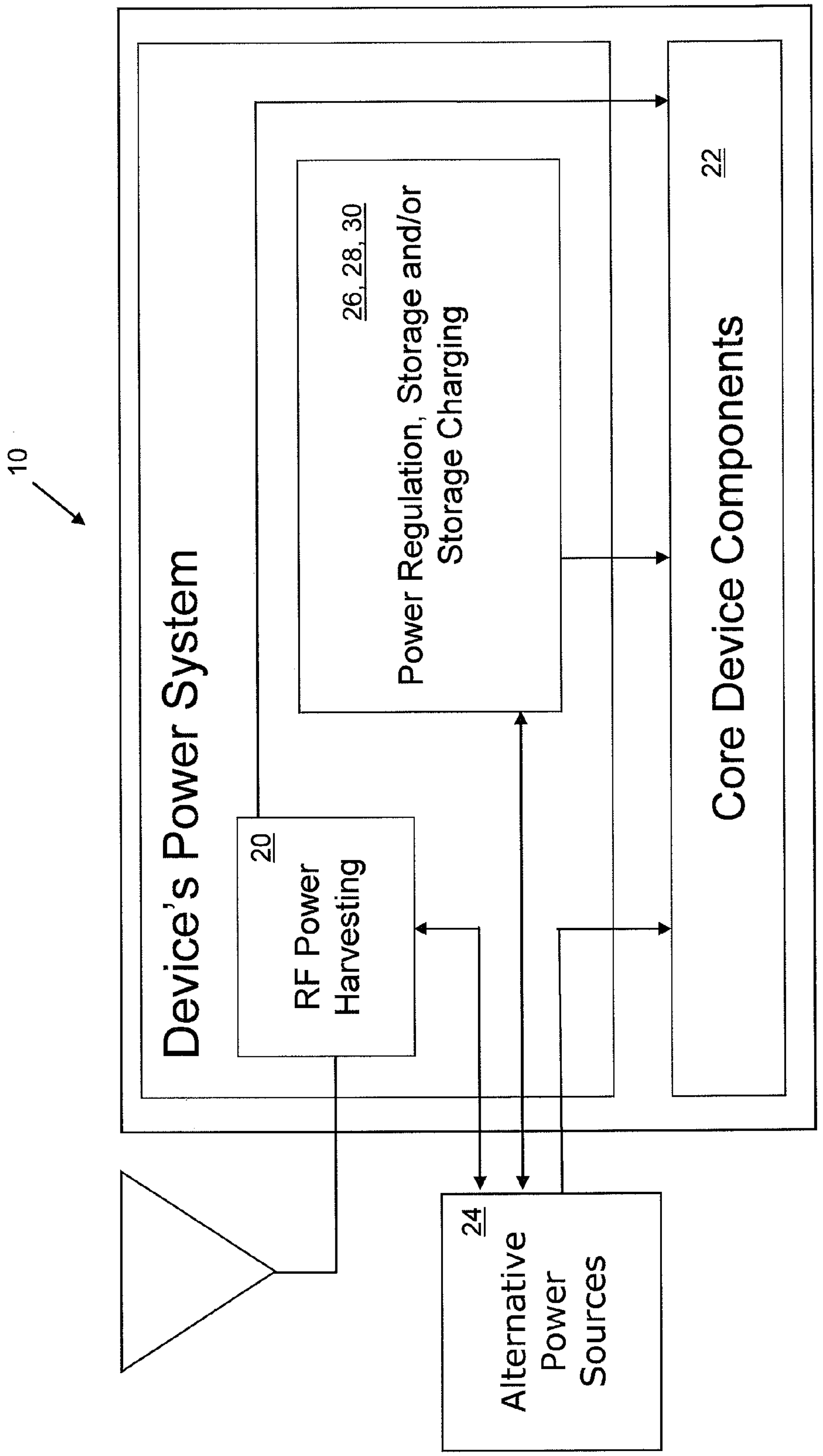
10
Device's Power System
Power Regulation, Storage and/or Storage Charging
26, 28, 30
Core Device Components
22
RF Power Harvesting
20
Alternative Power Sources
24

Fig. 80

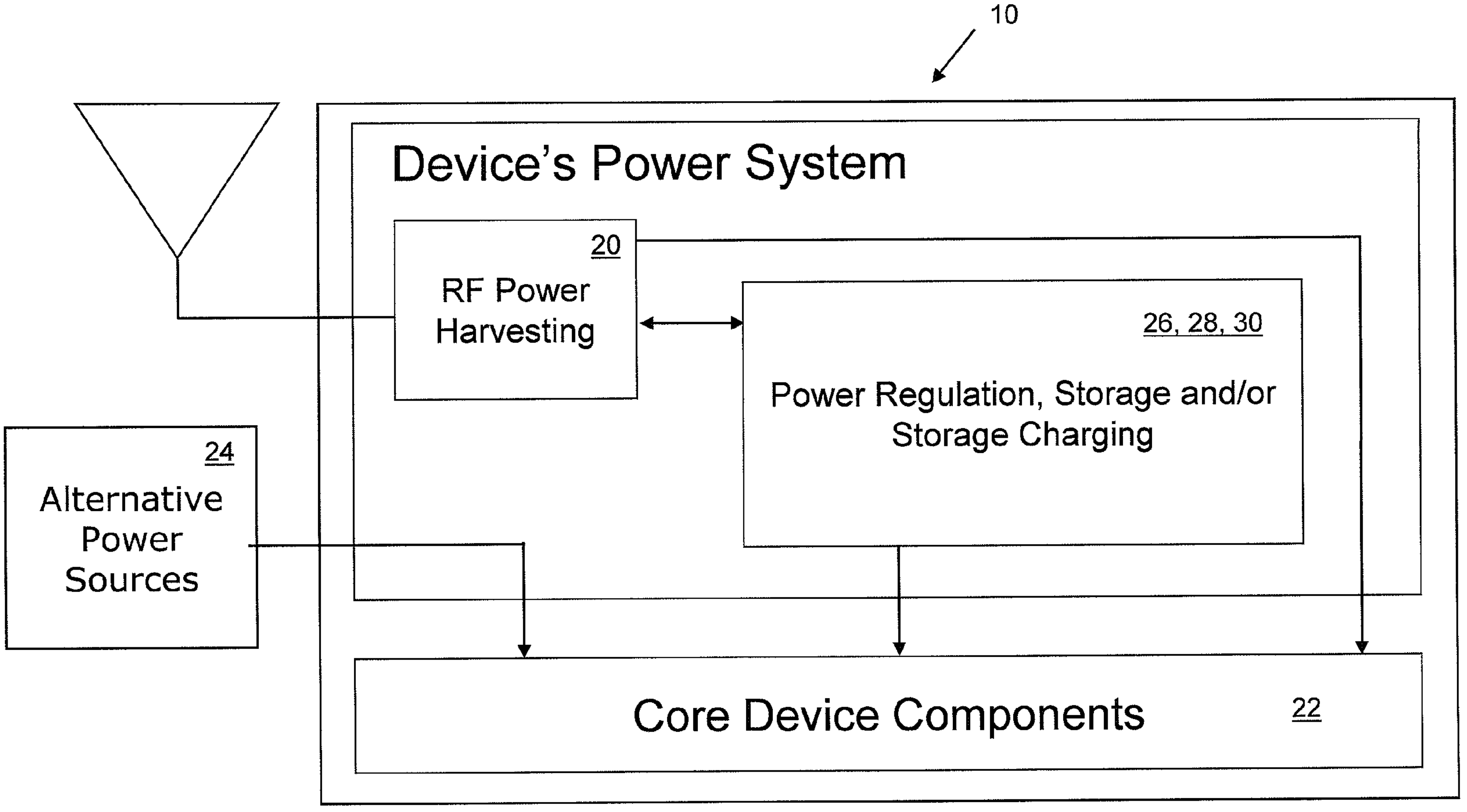
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 81

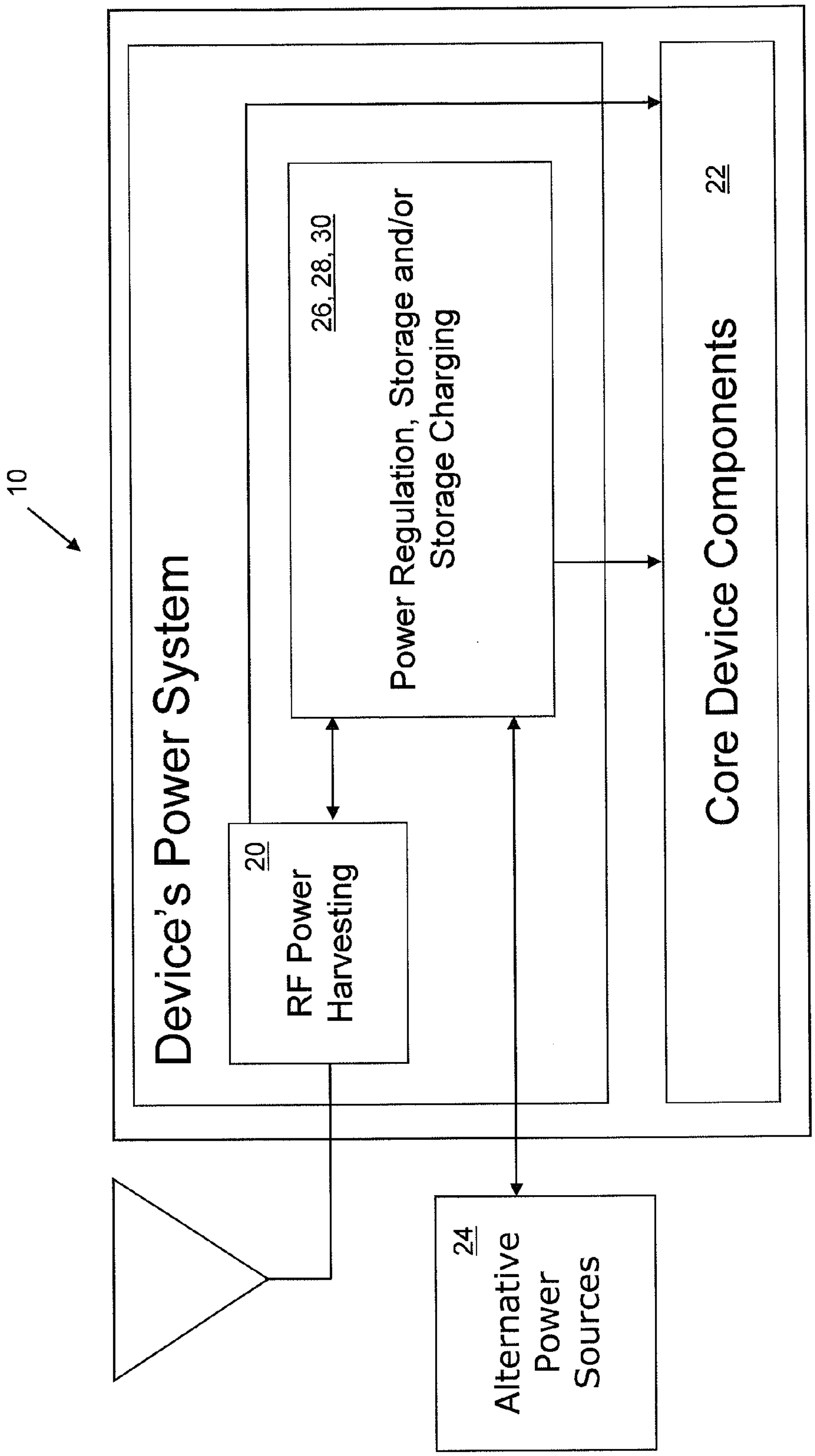
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 82

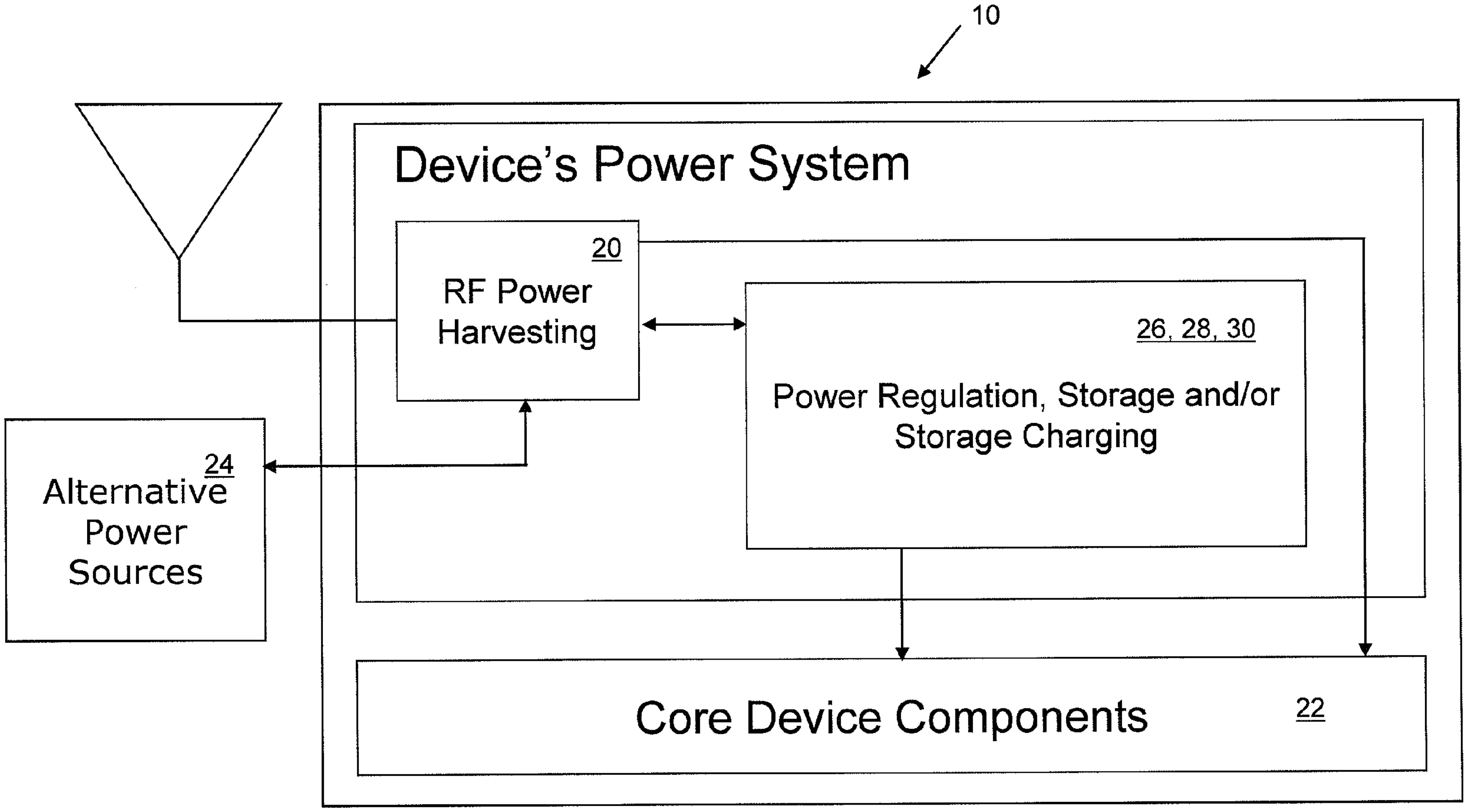
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 83

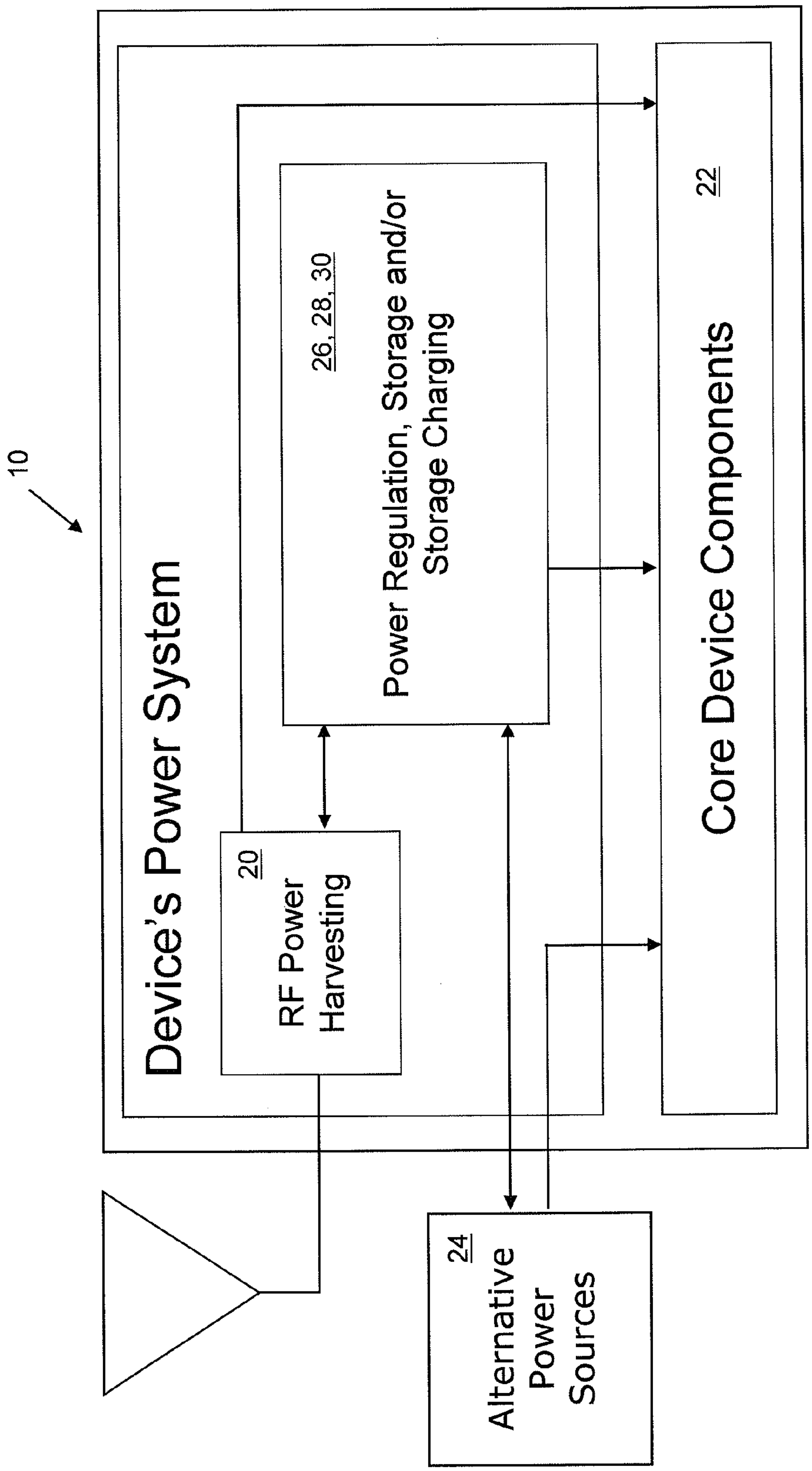
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 84

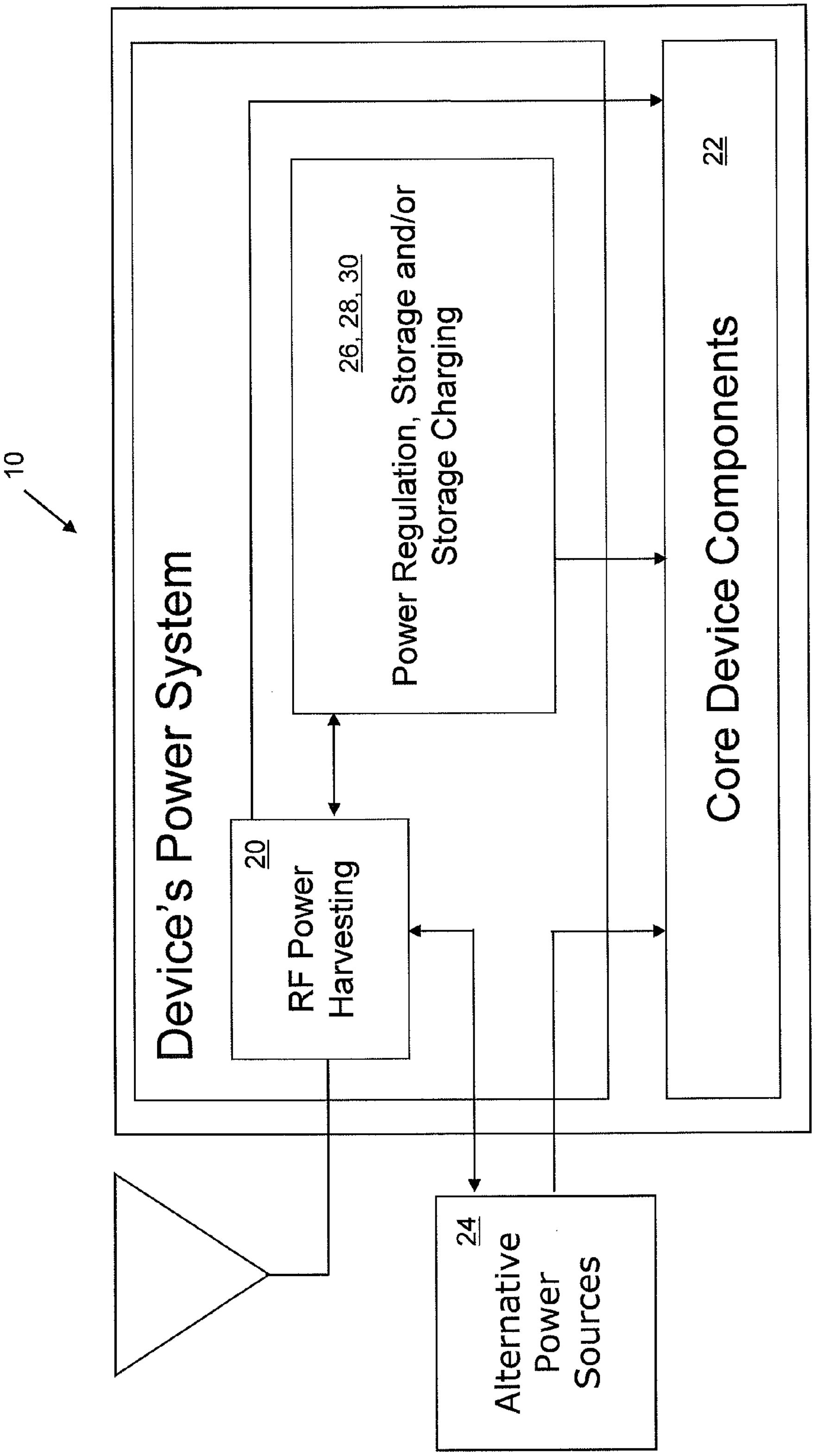
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 85

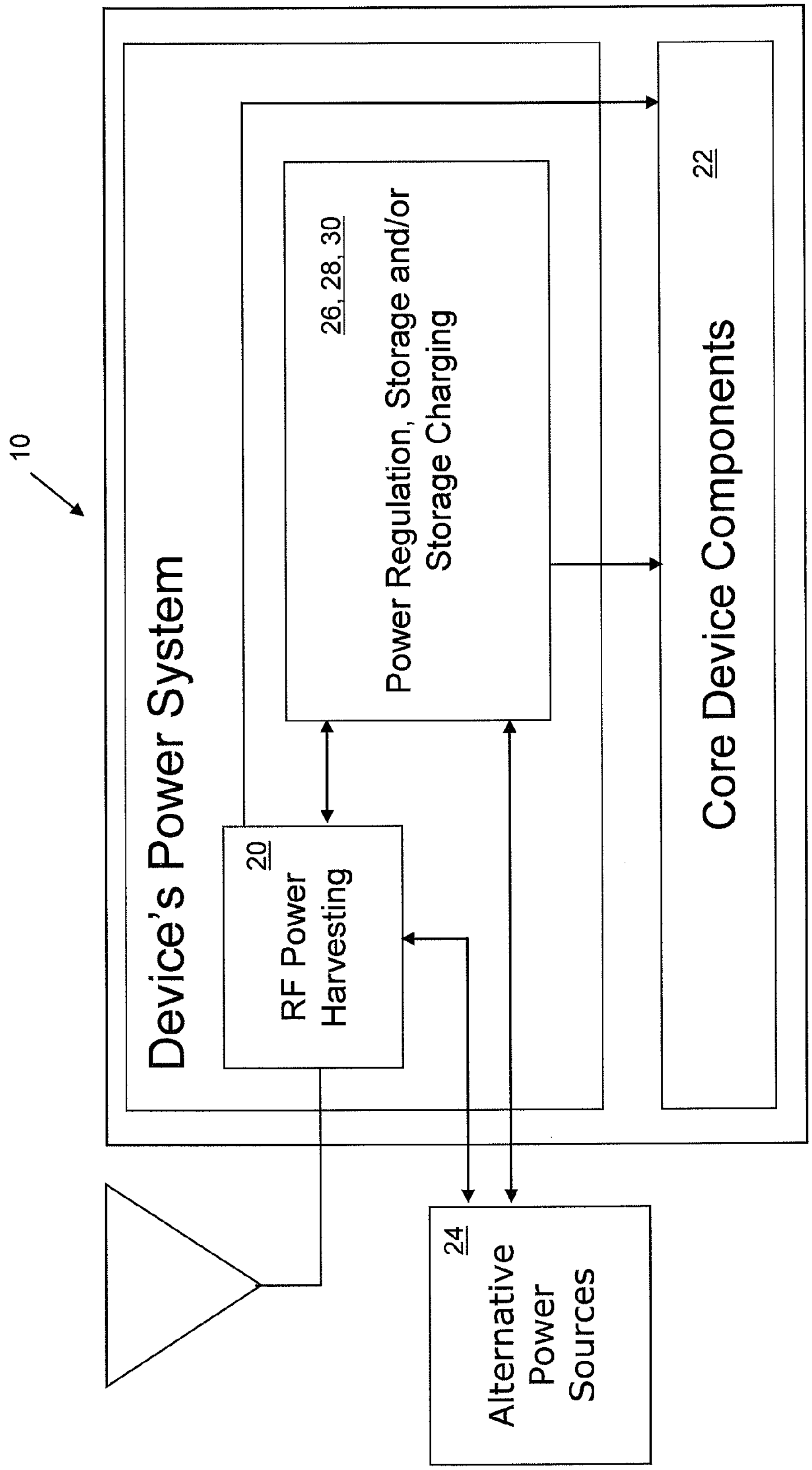
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 86

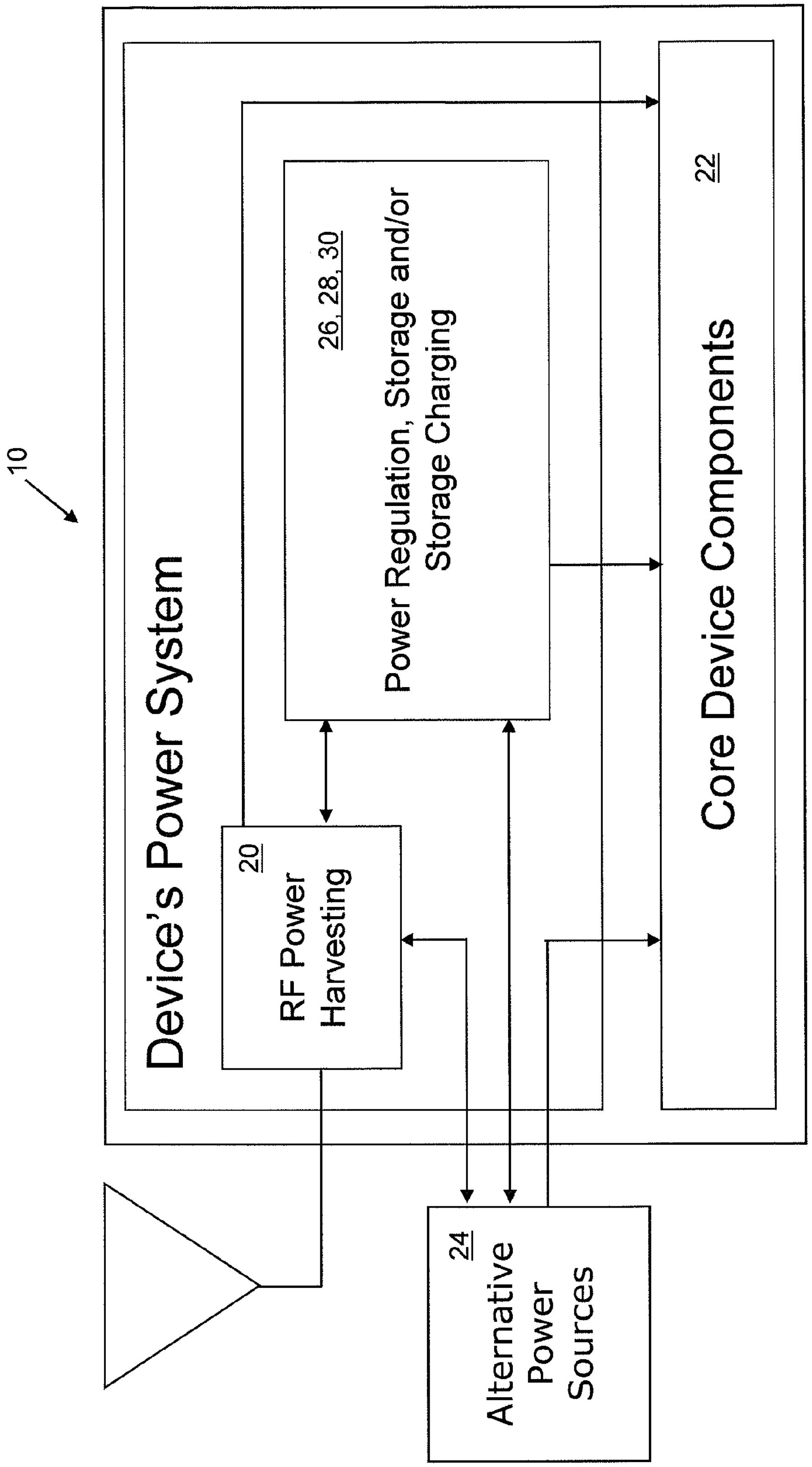
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 87

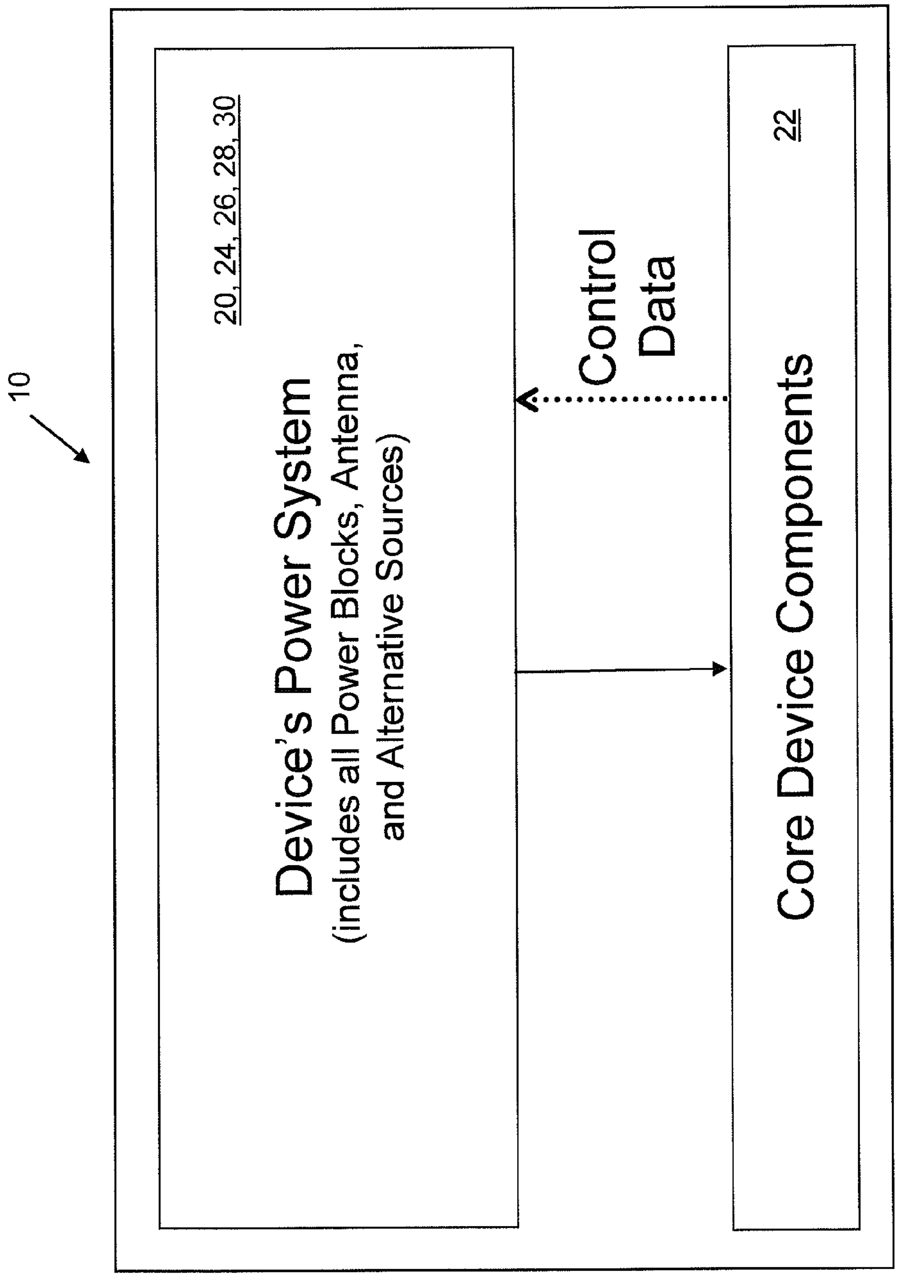
10
Device's Power System
(includes all Power Blocks, Antenna, and Alternative Sources)
20, 24, 26, 28, 30
Control Data
Core Device Components
22

Fig. 88

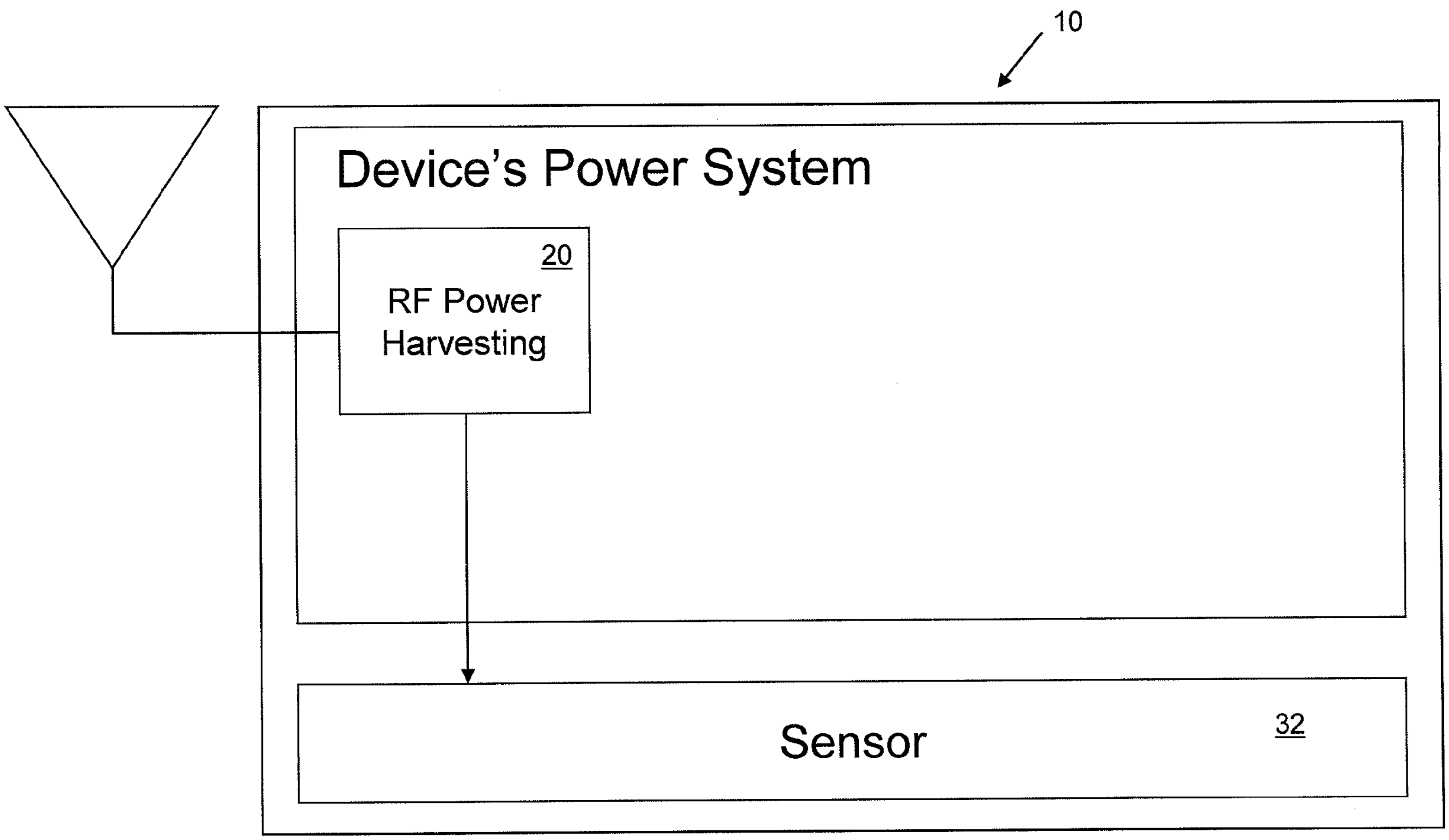
10
Device's Power System
20
RF Power Harvesting
Sensor
32

Fig. 89

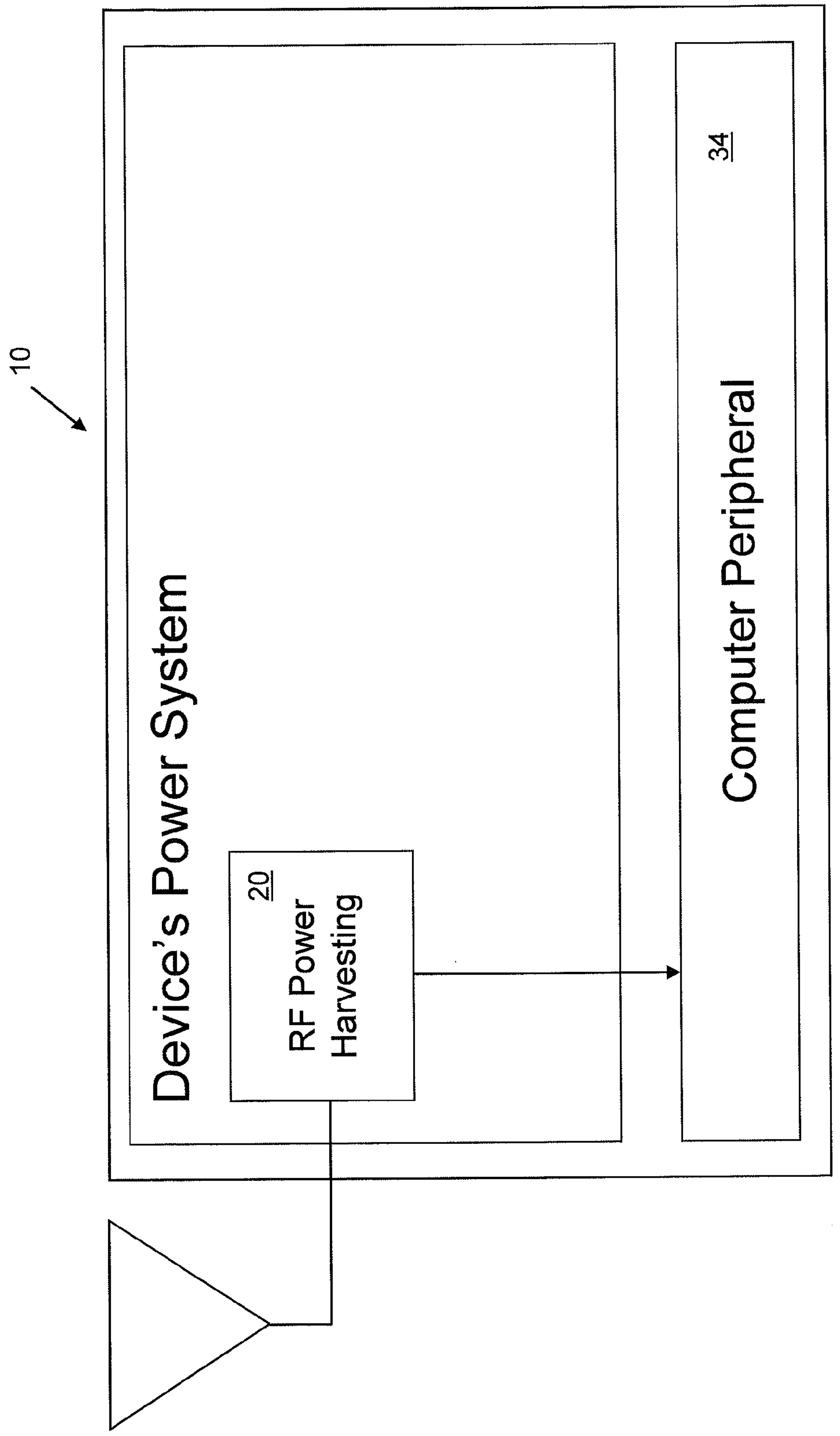
10
Device's Power System
20
RF Power Harvesting
Computer Peripheral
34

Fig. 90

POWERING DEVICES USING RF ENERGY HARVESTING

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to and is a continuation application of U.S. patent application Ser. No. 12/499,618, filed Jul. 8, 2009, now U.S. Pat. No. 8,621,245, which is a divisional application of U.S. patent application Ser. No. 11/447,412, filed Jun. 6, 2006, which claims priority to U.S. Provisional Patent Application No. 60/688,587, filed on Jun. 8, 2005; each of which is incorporated herein by reference in its entirety.

BACKGROUND

The present invention is related to the wireless powering of devices. More specifically, the present invention is related to the wireless powering of devices with a power harvester.

As processor capabilities have expanded and power requirements have decreased there has been an ongoing explosion of devices that operate completely independent of wires or power cords. These “untethered” devices range from cell phones, and wireless keyboards to building sensors and active RFID tags.

Engineers and designers of these untethered devices continue to have to deal with the limitations of portable power sources, primarily batteries as the key design parameter. While performance of processors and portable devices have been doubling every 18-24 months driven by Moore’s law, battery technology in terms of capacity has only been growing at measly 6% per year. Even with power conscious designs and the latest in battery technology, many devices do not provide the lifetime cost and maintenance requirements for applications that require a large number of untethered devices such as logistics, and building automation. Today’s devices that need two-way communication require scheduled maintenance every three to 18 months to replace or recharge the device’s power source (typically a battery). One-way devices simply broadcasting their status (one-way) such as automated utility meter readers have a better battery life, typically requiring replacement within 10 years. For both device types, scheduled power-source maintenance is costly and disruptive to the entire system that a device is intended to monitor and/or control. Unscheduled maintenance trips are even more costly and disruptive. On a macro level, the relatively high cost associated with the internal battery also reduces the practical, or economically viable, number of devices that can be deployed.

The ideal solution to the power problem for untethered devices is a device or system that can collect and harness sufficient energy from the external environment. The harnessed energy would then either directly power an untethered device or augment a battery or other storage component. Directly powering an untethered device enables the device to be constructed without the need for a battery. Augmenting a storage component could be along two lines: 1) increasing the overall life of the device or 2) by providing more power to the device to increase the functionality of the device. The other parameters for an ideal solution is that the harnessing device could be used in a wide range of environments including harsh and sealed environments (e.g. nuclear reactors), would be inexpensive to produce, would be safe for humans, and would have a minimal effect on the basic size, weight and other physical characteristics of the untethered device.

SUMMARY

The present invention pertains to an apparatus for an application. The apparatus comprises a core device preferably having an integrated circuit for the application. The apparatus comprises a power harvester (preferably a radio frequency (RF) energy harvester) connected to the core device to power the core device.

The present invention pertains to an apparatus for an application. The apparatus comprises a core device preferably having an integrated circuit for the application. The apparatus comprises means for receiving energy wirelessly and providing power from the energy to the core device to power the integrated circuit of the core device. The receiving means is connected to the core device.

The present invention pertains to a method for an application. The method comprises the steps of converting RF energy into usable energy. There is the step of preferably powering an integrated circuit of the core device with the usable energy.

This invention pertains to a technique that uses radio frequency (RF) energy as a source of energy to directly power a device or augment a power storage component in a device. The present invention meets the requirements described in the previous “Background of the Invention” section.

Traditional RF receiving devices have focused on maximizing selectivity of the frequency to isolate and to be coherent without interference from other sources. In contrast, while the present invention operates at a specific frequency or range of frequencies, the device accepts any interference to supplement the output power of the device. Also, the research related to power harvesting that uses RF energy as a source has primarily focused on devices in close proximity of the source (inductive or near-field energy). In most cases, prior research assumed a dedicated or directed source of RF to power the device.

The invention should not be confused with power transfer by inductive coupling, which requires the device to be relatively close to the power transmission source. The RFID Handbook by the author Klaus Finkenzeller defines the inductive coupling region as distance between the transmitter and receiver of less than 0.16 times lambda where lambda is the wavelength of the RF wave. The invention can be implemented in the near field (sometimes referred to as inductive) region as well as the far-field region. The far-field region is distances greater than 0.16 times lambda.

It is an object of this invention to provide a method and apparatus to

1. remotely energize an untethered device without using direct wiring
2. power or augment the life of the power storage component so it matches the life of the device and, ultimately, powers the off-grid device with or without the use of batteries
3. allow untethered devices to be virtually maintenance free
4. provide augmentation for other energy harvesting technologies (solar, piezoelectric, etc.)
5. provide backup power to tethered devices

It is a further object of this invention to directly power a device or augment a power storage component in a device in conjunction with other power harvesting technologies and storage elements.

With this method and apparatus a device’s power storage components do not require replacement, thus enabling the device to be permanently placed off-grid, where it may be physically impractical, costly, or dangerous (due to a harsh environment) to provide maintenance.

For devices on-grid (tethered) or with reliable power sources, RF power harvesting can be used as a backup in case the primary power source is lost.

BRIEF DESCRIPTION OF THE DRAWINGS

In the accompanying drawings, the preferred embodiment of the invention and preferred methods of practicing the invention are illustrated in which:

FIG. **1** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

FIG. **2** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

FIG. **3** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

FIG. **4** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

FIG. **5** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

FIG. **6** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

FIG. **7** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

FIG. **8** is a block diagram of an RF Power Harvesting block in communication with a Power Storage block.

FIG. **9** is a block diagram of an RF Power Harvesting block in communication with a Power Storage block and used to supply power to Core Device Components.

FIG. **10** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

FIG. **11** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

FIG. **12** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block.

FIG. **13** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

FIG. **14** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

FIG. **15** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

FIG. **16** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

FIG. **17** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block.

FIG. **18** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger.

FIG. **19** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and the RF Power Harvesting block in communication with a Power Storage block.

FIG. **20** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and Core Device Components.

FIG. **21** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and Core Device Components and the RF Power Harvesting block in communication with a Power Storage block.

FIG. **22** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

FIG. **23** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

FIG. **24** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

FIG. **25** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

FIG. **26** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

FIG. **27** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

FIG. **28** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the RF Power Harvesting block supplying power to Core Device Components.

FIG. **29** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

FIG. **30** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

FIG. **31** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

FIG. **32** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power Core Device Components.

FIG. **33** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

FIG. **34** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

FIG. **35** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

FIG. **36** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or

Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

FIG. **37** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

FIG. **38** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

FIG. **39** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

FIG. **40** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to the Power Storage Charger.

FIG. **41** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

FIG. **42** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

FIG. **43** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

FIG. **44** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the RF Power Harvesting block supplying power to Core Device Components and used to supply power to a Power Storage Charger.

FIG. **45** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

FIG. **46** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

FIG. **47** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

FIG. **48** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power Core Device Components and used to supply power to a Power Storage Charger.

FIG. **49** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

FIG. **50** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

FIG. **51** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

FIG. **52** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

FIG. **53** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

FIG. **54** is a block diagram of an RF Power Harvesting block using Antenna A to directly supply power to Core Device Components.

FIG. **55** is a block diagram of an RF Power Harvesting block using Antenna A to supply power to a Power Regulation, Storage and/or Storage Charging block.

FIG. **56** is a block diagram of an RF Power Harvesting block using Antenna A to supply power to a Power Regulation, Storage and/or Storage Charging block and used to supply power to Core Device Components.

FIG. **57** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

FIG. **58** is a block diagram of an RF Power Harvesting block used to supply power to a Power Regulation, Storage and/or Storage Charging block.

FIG. **59** is a block diagram of an RF Power Harvesting block used to supply power to a Power Regulation, Storage and/or Storage Charging block and used to supply power to Core Device Components.

FIG. **60** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

FIG. **61** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components and in communication with an Alternative Power Sources block.

FIG. **62** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and used to directly supply power to Core Device Components.

FIG. **63** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

FIG. **64** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

FIG. **65** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

FIG. **66** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

FIG. **67** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

FIG. **68** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

FIG. **69** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

FIG. **70** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

FIG. **71** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

FIG. **72** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

FIG. **73** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

FIG. **74** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

FIG. **75** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

FIG. **76** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

FIG. **77** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

FIG. **78** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

FIG. **79** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

FIG. **80** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources.

FIG. **81** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

FIG. **82** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

FIG. **83** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

FIG. **84** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

FIG. **85** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

FIG. **86** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

FIG. **87** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

FIG. **88** is a block diagram of an entire power system for the device.

FIG. **89** is a block diagram of a power harvesting block used to supply power to a core device having a sensor.

FIG. **90** is a block diagram of a power harvesting block used to supply power to a core device having a computer peripheral.

DETAILED DESCRIPTION

A complete understanding of the invention will be obtained from the following description when taken in connection with the accompanying drawing figures wherein like reference characters identify like parts throughout.

There is shown an apparatus **10** for an application. The apparatus **10** comprises a core device **22** preferably having an integrated circuit **36** for the application. The apparatus **10** comprises a power harvester **20** connected to the core device **22** to power the core device **22**.

The apparatus **10** preferably includes an alternative power source **24** connected to the core device **22** to power the core device **22** in conjunction with the power harvester **20**. Preferably, the apparatus **10** includes a power regulator **26** and/or power storage circuit **28** connected to the power harvester **20**. The apparatus **10** preferably includes a power storage charger **30** connected to the power harvester **20**. Preferably, the apparatus **10** includes a power storage connected to the power harvester **20**.

Preferably, the core device **22** includes a memory **38** connected to the integrated circuit **36** and to the power harvester **20** to power the memory **36**.

The core device **22** can include a sensor **32**, as shown in FIG. **89**. The sensor **32** can include a proximity sensor, an intrusion sensor, an environmental sensor, a chemical sensor, a biological sensor, a sensor in contact with an automobile, an occupancy sensor, a motion sensor, a position sensor, a metal detector, or a sensor **32** in contact with an aircraft. The sensor **32** can include an alarm connected to the power harvester **20** to power the alarm, a display connected to the power harvester **20** to power the display, a sensor **32** disposed in a building, an industrial automation sensor, a sensor **32** in contact with an elevator, a temperature sensor, a fire sensor, or an accelerometer.

The sensor **32** can include a level sensor, a gas level sensor, a fluid level sensor, a light level sensor. The sensor **32** can include a flow sensor, a gas flow sensor, a fluid flow sensor, a light flow sensor, or a plasma flow sensor.

The sensor **32** can include a pressure sensor, a gas pressure sensor, a fluid pressure sensor, a light sensor, an infrared light sensor, an ultraviolet light sensor, an x-ray sensor, a cosmic ray sensor, a visible light sensor, a gamma ray sensor, a stress sensor, a strain sensor, a depth sensor, or an electrical characteristic sensor.

The sensor **32** includes a voltage sensor, a current sensor, a viscosity sensor, an acoustical sensor, a sound sensor, a listening sensor, a thickness sensor, a density sensor, a surface quality sensor, a volume sensor, a physical sensor, a mass sensor, a weight sensor, a conductivity sensor, a distance sensor, an orientation sensor, or a vibration sensor.

The sensor **32** can include a radioactivity sensor, a field strength sensor, an electric field sensor or a magnetic field sensor, a smoke detector, a carbon monoxide detector, a radon detector, an air quality sensor, a humidity sensor, a glass breakage sensor, or a break beam detector. The sensor can include a thermal energy sensor, an electromagnetic sensor, a mechanical sensor, an optical sensor, a radiation sensor, a sensor in contact with a vehicle, or a sensor **32** in contact with a water craft.

The present invention pertains to an apparatus **10** for an application. The apparatus **10** comprises a core device **22** having an integrated circuit **36** for the application. The apparatus **10** comprises means for receiving energy wirelessly and providing power from the energy to the core device **22** to power the integrated circuit **36** of the core device **22**. The receiving means is connected to the core device **22**. Preferably, the core device **22** includes means for sensing.

The present invention pertains to a method for an application. The method comprises the steps of converting RF energy into usable energy. There is the step of preferably powering an integrated circuit **36** of the core device **22** with the usable energy.

Preferably, there is the step of regulating the usable energy provided to the core device **22**. There is preferably the step of storing the usable energy. Preferably, there is the step of providing power to the core device **22** from an alternative power source **24** in conjunction with the usable energy.

Alternatively, the core device **22** can include a computer peripheral **34**, as shown in FIG. **90**. The computer peripheral **34** can include a handheld game, a gaming system, a game controller, a controller, a keyboard, a mouse, a computer terminal, computer storage, or computer equipment.

The present invention can be implemented in numerous ways. A number of these ways are depicted in FIGS. **1-90**. These figures contain multiple blocks that are configured in multiple ways.

In the figures, an arrow represents the flow of power unless otherwise stated. Single-headed (one-way) arrows represent that the power is flowing from one block to another. The single-headed arrow may represent multiple wires that provide power from one block to multiple parts in the other block. Two-headed (two-way) arrows represent a single wire that can have power flow in either direction or multiple wires each having power flow in a single direction.

As an example, a two-headed arrow between an RF power harvesting **20** block and a power regulation **26** and/or power storage circuit **28** block can represent a single wire that allows harvested power to flow into a storage device such as a capacitor. The same block diagram can also represent two wires between the two blocks with the first wire allowing harvested power to flow into a voltage regulator **26**. The second wire can allow the regulated voltage to feedback to the RF power harvesting **20** block to provide power to internal components such as transistors to increase the performance of the RF power harvesting **20** block.

Each block is described in detail below. Each block represents the functionality described below associated with it. For instance, the RF power harvesting **20** block describes a power harvester **20**, and the power regulation **26** and/or power storage circuit **28** block describes a power regulator **26** and a power storage circuit **28**.

Device's Power System Block

The device's power system block includes all components/circuitry required to power the device. This block may include an RF power harvesting **20** block, a power regulation **26** and/or power storage circuit **28** block, a power storage block, and/or a power storage charger **30** block (all discussed below).

RF Power Harvesting **20** Block

The RF power harvesting **20** block is connected to an antenna. The antenna may or may not be used as the communications antenna for the core device **22** components. The RF power harvesting **20** block is used to convert the energy captured by the antenna into usable power, such as direct current (DC) voltage.

This block may include antenna matching, rectifying circuitry, voltage transforming circuitry, and/or other performance optimizing circuitry. The rectifying circuitry may include a diode(s), a transistor(s), or some other rectifying device or combination. Examples of the rectifying circuitry include, but are not limited to, half-wave, full-wave, and voltage doubling circuits.

The output of the RF power harvesting **20** block is a DC voltage or current. The RF power harvesting **20** block may accept feedback (or input) from other circuitry or blocks, which may be used to control the harvesting circuitry to improve the performance or vary the output. This feedback may include, but is not limited to, a DC voltage or a clock from the core device **22** components.

U.S. Pat. No. 6,615,074 (FIGS. **8**, **9**, **12***a*, **12***b*, **13**, **14**), incorporated by reference, herein, shows numerous examples of RF power harvesting circuits that can be used to implement the block and function described.

Power Regulation **26** and/or Power Storage Circuit **28** Block

It may be necessary to regulate the converted power (i.e., hold the power at a constant level) for specific devices. The devices that would need this block require a fairly constant voltage or current. Deviations from the required values may cause the device to not perform within its specifications.

The regulation can be implemented in many different ways. The block can be as simple as using a Zener diode, or as complicated as using an integrated circuit such as a linear voltage regulator **26** or switching regulator **26** to hold the voltage at a constant level.

Certain devices have a more tolerable power requirement. For these devices, the regulation stage may be excluded.

This block may also include, with or without the regulation, a storage device such as a capacitor, a battery, or some other device able to store charge. The output from the power regulation **26** and/or power storage circuit **28** block may be used as feedback to other blocks within the device's power system block or to the alternative power sources **24** block (described below), if they require a regulated supply voltage or stored power.

U.S. Pat. No. 6,894,467 (FIGS. **1**, **3**), Linear Voltage Regulator, incorporated by reference, herein, is an example of a practical application of implementing the regulation described in the block. U.S. Pat. No. 6,297,618 (FIGS. **1-4**), Power storage device and method of measuring voltage of storage battery, incorporated by reference, herein, is an example of a practical application of implementing the storage described in the block.

Power Storage Charger **30** Block

The power storage charger **30** block may be needed if a power storage component of the device requires a special charging mechanism, such as pulse charging or trickle charging. This block controls how the captured and converted power is supplied to the power storage component.

U.S. Pat. No. 6,836,095 (FIGS. **1-3**), Battery Charging Method and Apparatus, incorporated by reference herein, is an example of a practical application of implementing the special charging mechanism described in the block.

Power Storage Block

If a device has intermittent power requirements, it may be necessary to store the captured power for use at a later time. The power can be stored in the power storage block, which could include a battery, a capacitor, and/or another type of power storage component. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other power storage elements.

The output from the power storage block may be used as feedback to other blocks within the device's power system or to the alternative power sources **24** block, if they require a dedicated and predictable supply voltage.

U.S. Pat. No. 6,297,618 (FIGS. **1-4**), Power Storage Device and Method of Measuring Voltage of Storage Battery, incorporated by reference herein, is an example of a practical application of implementing the storage described in the block. U.S. Pat. No. 6,835,501, Alkaline Rechargeable Battery, incorporated by reference herein, is also an example of a practical application of implementing the storage described in the block.

Core Device **22** Components Block

The core device **22** components block is the portion of the device that is receiving power from the device's power system. This block may be, but is not limited to, the devices listed in the subsequent pages of this document. It may be advantageous for the core device **22** components block to communicate with any of the blocks that are supplying power to it. This communication can include, but is not limited to, a feedback control signal such as a clock or an ON/OFF command. As an example, the device may want to turn off the alternative power sources **24** block if it is receiving sufficient power from the RF power harvesting **20** block.

Alternative Power Sources **24** Block

RF energy harvesting also has the ability to be augmented by other types of power harvesting, storage components, or dedicated sources (e.g. power line). The alternative power sources **24** block shows how this type of system could be implemented. The augmenting power harvesting technologies include, but are not limited to, solar, light (visible and non-visible), piezoelectric, vibration, acoustic, thermal, microgenerators, wind, and other environmental elements. This block can work independently or have communication with other blocks.

U.S. Pat. No. 6,784,358, Solar Cell Structure Utilizing an Amorphous Silicon Discrete By-Pass Diode, incorporated by reference herein, is an example of a practical application of implementing an alternative power source **24** described by the block. U.S. Pat. No. 6,858,970, Multi-Frequency Piezoelectric Energy Harvester, incorporated by reference herein, is also an example of a practical application of implementing an alternative power source **24** described by the block.

Power Regulation **26**, Storage **28** and/or Storage Charging **30** Block

The power regulation **26**, storage **28** and/or storage charging **30** block contains all the combinations of the power regulation **26** and/or power storage circuit **28** block, power storage charger **30** block, and power storage block. This block is used in the later figures to reduce the number of figures needed to show how the blocks can interconnect.

The disclosed invention is an apparatus and method for an application for retrieving radio frequency (RF) energy by an antenna, converting that energy into direct current (DC) power, regulating that energy using an optimized circuit, storing that energy in an optimized component, and/or supplying the power for a specific device.

Retrieval of RF Energy

The RF energy is retrieved from the environment by the use of the antenna. The antenna can be shared or standalone with respect to an antenna used for the device's wireless communication. FIGS. **54-56** show a device that has an antenna A for use by the RF power harvester **20** and an antenna B used for wireless one-way or two-way communication. FIGS. **57-59** show a device where the antenna is shared by both a device's communication module (within the core device **22** components) and the RF power harvester **20**. In terms of form factor, the antenna used by the apparatus **10** can be a separate component or integrated directly into the form factor of the device.

The antenna is able to capture two types of available RF energy. The first type of energy exists as ambient RF energy. This type of RF surrounds us in our day-to-day lives and is usually generated to carry one way or two-way combinations of voice, video and data communications. The sources that the antenna can harvest from include, but are not limited to, medium-frequency AM radio broadcast, very-high-frequency (VHF) FM radio broadcast and television broadcast, ultra-high-frequency (UHF) broadcast, cellular base stations, wireless data access points, super-high-frequency (SHF) frequencies, and the industrial, scientific, and medical (ISM) bands. These sources cover transmission frequencies from 300 kHz to 30 GHz.

The second type of energy available is directed RF energy. This type of RF energy is directed from a transmitter specifically designed to deliver RF energy for harvesting by the antenna. The transmitter can be configured as a standalone device or integrated into an existing device.

Conversion of the Energy into DC

The RF energy captured by the antenna must be converted into a useful form of energy for the specific device. This conversion is shown in block form in all FIGS. (**1-90**) as the RF power harvesting **20** block. The most common form of useable energy is DC energy. To perform this conversion, the block includes circuitry to rectify the captured alternating current (AC) energy to create DC energy. The rectification in this block can be done with a diode(s), a transistor(s), or some other rectifying device or combination.

Regulation of the Energy

It may be necessary to regulate the converted power (hold the power at a constant level) for specific devices. FIGS. **2-7**, **10-17**, and **22-53** show how this regulation can be implemented using a power regulation **26** and/or power storage circuit **28** block. The devices that would need this block require a fairly constant voltage or current. Deviations from the required values may cause the device to not perform within its specifications. The regulation can be implemented in many different ways. The block can be as simple as using a Zener diode, or as complicated as using an integrated circuit such as a linear voltage regulator **26** or a switching regulator **26** to hold the voltage at a constant level. Certain devices have a more tolerable power requirement. For these devices, the regulation stage may be excluded.

Storage of the Energy

If a device has intermittent power requirements, such as the devices exampled by FIGS. **2-53**, it may be necessary to store the captured power for use at a later time. The power can be stored in the power storage block or the power regulation **26** and/or power storage circuit **28** block. Storage devices can include, but are not limited to, a battery, a capacitor, or another type of power storage component. In certain applications, it may be necessary to include additional circuitry that controls how the power is transferred to the storage device. The power storage charger **30** block is shown in FIGS. **18-53**. This may be needed if the power storage component requires a special charging mechanism such as pulse charging or trickle charging. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other storage elements. There are devices that will not require storage. These devices can run directly off of the converted power. These devices also may or may not require regulation of the captured power.

Supplying the Power

The captured DC power, which may or may not be regulated and/or stored, is supplied to the device, which is represented by the core device **22** components block in the figures. This may be a single connection or it may supply multiple parts of the device with power.

RF energy harvesting also has the ability to be augmented by other types of power harvesting or storage components. Other power harvesting technologies include, but are not limited to, solar, light (visible and non-visible), piezoelectric, vibration, acoustic, thermal, microgenerators, wind, and other environmental elements. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other storage elements. FIGS. **60-88** show how the alternative power sources **24** block can be connected to an RF energy harvesting system. These figure show how the RF energy harvesting **20** block and the alternative power sources **24** block can work independently or have communication with each other. The antenna configurations shown in FIGS. **54-59** are still applicable with the addition of an alternative power sources **24** block. These antenna configurations can be applied to FIGS. **60-88**.

RF energy harvesting also has the ability to provide a backup to devices on-grid (tethered) or with reliable power sources, which can be used in case the primary power source is lost. As an example, it may be mandated by regulations that a sensor has auxiliary power in case the primary supply is lost. It could be possible to use a rechargeable battery that obtains its charge from the primary supply when in operation. However, if the primary supply is lost for a time greater than the life of the rechargeable battery, the specification of uninterrupted power is not met. RF energy could be used to supply power to the described device while the primary supply is not available. The primary supply could include, but is not limited to, an on-grid connection, a generator, a battery, or other reliable power supply.

RF energy harvesting with or without alternative source augmentation is applicable to provide electric power directly or indirectly to a range of electronic components contained in any specific electrical or electronic device and includes, but is not limited to:

Passive electronic components, active electronic components;

Resistors, fixed resistors, variable resistors, thermistors, thyristor, thermocouple;

Capacitors, Electrolytic Capacitors, Tantalum Capacitors, Ceramic Capacitors, Multilayer Ceramic Capacitors, Polystyrene Film Capacitors, Electric Double Layer Capacitors (Super Capacitors), Polyester Film Capacitors, Polypropylene Capacitors, Mica Capacitors, Metallized Polyester Film Capacitors, Variable Capacitors;

Diodes, Voltage regulation diodes, light-emitting diodes, organic light-emitting diodes, Variable capacitance diodes, Rectification diodes, Switching diodes, Regulation Diodes, Diode bridges, Schottky barrier diodes, tunnel diodes, PIN diodes, Zener diodes, Avalanche diodes, TVSs;

Integrated circuits, microcontroller unit (MCU), microprocessor unit (MPU), logic circuits, memory, printed circuits, circuit boards, printed wiring boards;

Transistors, MOSFETs, FETs, BJTs, JFETs, IGBTs, Relays, Antennas, semiconductors, conductors, inductors, relays, diacs, triacs, SCRs, MOVs;

Fuses, circuit breakers;

Batteries, Non-rechargeable batteries, rechargeable batteries, coin cell batteries, button cell batteries, alkaline batteries, lithium batteries, lithium ion batteries, lithium polymer batteries, NIMH batteries, NICAD batteries, Lead acid batteries, Zinc air batteries, Manganese Lithium batteries, Niobium Titanium Lithium batteries, Vanadium Pentoxide Lithium batteries, Carbon Zinc batteries, Zinc Chloride batteries, Lithium Thionyl Chloride batteries, Manganese Dioxide batteries, Lithium Poly-Carbonmonofluoride batteries, Lithium Manganese Dioxide batteries, Lithium Chloride batteries, Lead Acid Calcium batteries, Lead Acid Tin batteries, Oxy Nickel batteries, Silver Oxide batteries, Magnesium batteries;

Inductors, Coils, High Frequency Coils, Toroidal Coils, Transformers, switches, chokes;

Motors, DC motors, stepper motor, AC motors, Fans;

Crystals, Oscillators, Clocks, Timers; and

Displays, LCDs, LED displays.

RF energy harvesting with or without alternative source augmentation is applicable across a range of markets and specific devices and includes but is not limited to:

Consumer electronics:

Electronic equipment, wired devices, battery powered devices, wireless communication devices, cell phones, telephones, phones, cordless phones, portable phones, Bluetooth devices, Bluetooth headsets, hands-free headsets, headsets, headphones, Wireless headsets, radios, AM/FM radios, shortwave radios, weather radios, Two-way radios, portable radio, lights, lanterns, portable lights, flashlights, nightlights, spotlights, search lights, calculators, graphing calculators, desk calculators, clocks, alarm clocks, wall clocks, desk clocks, travel clocks, watches, wristwatches, pocket watches, stop watches, timers, voice recorders, Dictaphones, laser pointers, power tools, cordless power tools, electronic razors, electric razors, handheld games, gaming systems, game controllers, wireless game controllers, remote controls, battery chargers, computers, portable computers, keyless entry, toys, toy guns, toy laser guns, games, microphones, musical instruments, musical effects processors, musical instrument tuners, metronomes, electronic chord charts, door openers, garage door openers, PDA, Cameras, Video recorders, Multimeter, electronic test equipment, hand-held electronics, portable electronics, wireless pens, sound generators, noise generators, language translators, electric toothbrushes, portable televisions, pagers, transceivers, toy vehicles, remote control vehicles, toy planes, remote control planes, pet containment systems, invisible fence pet sensors, memory backup, base station battery backups, appliance battery backups, uninterrupted power supplies, GPS devices, memory retention power supplies, metal detectors, stud finders, metal stud finders, stun guns, tazers, wearable devices, baby monitors, intercoms, doorbells, wireless doorbells, electronic office supplies, electronic staplers, radar jammers, radar detectors, digital scales, microfilm cassettes, video head testers, compasses, noise canceling headphones, air samplers, depth finders, barometers, weather measurement instruments, data transfer devices, automatic distress signaling unit, Wireless audio speakers, Satellite radios, Police scanners, Car navigation systems (GPS devices), Decorative lights, Christmas lights, garden lights, lawn lights, ornamental lights, porch lights;

Multi-media players: MP3, DVD, analog music players, CD players, tape players, digital music players, digital video players, minidisc; and

Computer: keyboards, mice, peripherals, computer equipment, electronic computers, computer storage, computer terminals;

Building/Industrial Automation:

Sensors: Position, elevator, temperature, fire, accelerometers, level, gas level, fluid level, light level, flow, gas flow, fluid flow, light flow, plasma flow, pressure, gas pressure, fluid pressure, motion, light, infrared light, ultraviolet light, X-rays, cosmic rays, visible light, gamma rays, chemical, stress, strain, depth, electrical characteristics, voltage, current, viscosity, acoustical, sound, listening, thickness, density, surface quality, volume, physical, mass, weight, force, conductivity, distance, orientation, vibration, radioactivity, field strength, electric field strength, magnetic field strength, occupancy, smoke detector, carbon monoxide detector, radon detector, air quality, humidity, glass breakage, break beam detector;

Controls: Position, elevator, temperature, fire, accelerometers, level, gas level, fluid level, light level, flow, gas flow, fluid flow, light flow, plasma flow, pressure, gas pressure, fluid pressure, motion, light, infrared light, ultraviolet light, X-rays, cosmic rays, visible light, gamma rays, chemical, stress, strain, depth, electrical characteristics, voltage, current, viscosity, acoustical, sound, listening, thickness, density, surface quality, volume, physical, mass, weight, force, conductivity, distance, orientation, vibration, radioactivity, field strength, electric field strength, magnetic field strength, occupancy, smoke detector, carbon monoxide detector, radon detector, air quality, humidity; and

Devices: Thermostats, light switches, door locks, smartcard door locks, lighting, emergency lighting, motion lighting, safety lighting, highway lighting, construction lighting, sign lighting, roadway sign lighting, construction sign lighting, automatic flushing units, automatic soap dispenser, automatic paper towel dispenser, automatic faucets, automatic door sensors, identification reader, fingerprint reader, credit card readers, card readers, valve actuators, gauges, analog gauges, digital gauges, fire extinguishers, wireless switches, remotely operated inspection equipment, gas/oil pipeline monitoring systems, robotic pipeline inspection gauges, “auto-reclosers” for electric power lines, sonar buoys, telemetry systems, electronic record tracking systems, robbery tracking devices, interrogators, programmers, emergency exit alarms, alarms, flood alarms, gas alarms, electronic entry systems, security keypads, silo transducers, data recorders, signal tracers, anti-static strap testers, radiosonde weather balloons, utilities load controllers, profilometers, noise cancellation equipment, infrared beacons;

Military/Government:

Tracking tags: Weapons, vehicle, soldier, gear/assets, staff, general population, security access badges;

Sensors: Proximity, intrusion, environmental, chemical/biological; and

Equipment: Battery charger, surveillance, card readers, identification reader, fingerprint reader, retinal scanners, satellites, rockets, space vehicles, search and rescue transponders (SARTs), emergency position-indicating rescue beacons (EPIRBs), emergency locator transmitters (ELTs), military radios, electronic toll collection systems, postal tracking systems, communications, thermal imaging, night vision, training targets, field medical equipment, house arrest monitors, laser tags, electronic parking meters, multiple integrated laser engagement system, munitions and mines, ship sensors;

Utility:

Gas consumption meters, water consumption meters, and electric consumption meters;

Logistics & Supply Chain Management:

Radio-frequency identification devices (RFID), RFID readers;

Tracking: Asset tags, cargo container location beacons, transponders, transceivers; and

Devices: Smart price tags, smart shelving, handheld barcode scanner, barcode scanners, credit card readers, card readers, retail signage, hotel door locks;

Homeland Security:

Sensors: Occupancy, proximity, environmental, chemical/biological, motion, position; and

Metal detector wand;

Medical:

Implantable: cochlear implants, neural stimulators, pace makers, medication administration, defibrillator;

Body function monitors: pressure, temperature, respiration, blood oxygenation, insulin, hearing aid, pulse, EKG, heart, Holter;

Tracking tags: Patient, baby identification, assets, supplies, staff, medication, instruments;

Devices: Home healthcare equipment, ambulatory infusion pumps, blood analyzers, biofeedback systems, bone growth stimulators, thermometers, digital thermometers, stimulators, galvanic stimulators, muscle stimulators, pediatric scales; and

Agriculture—livestock tracking and asset tracking:

Tracking: livestock, asset, wildlife tracking devices; and

Equipment: cattle prods;

Automotive:

Automotive antennas, Automotive Audio Systems, Automotive Lighting, Automotive Video Systems, Computers, Processors, Controls, Switches, Electric Motors, Actuators, Ignition Systems, Starter Systems, Injection Systems, Powertrain Electronics, Radar Detectors, Proximity Detectors, Safety Systems, Security Systems, Sensors, Regulators, Distributors, Vehicle Control, Wiper Systems, Washer Systems, Radio, Video Systems, Entertainment Systems, Navigation Systems, GPS systems, Power Mirror Systems, Emission control systems;

Appliances:

Monitoring systems and control systems for major and small appliances including washing machines, dryers, refrigerators, freezers, coolers, air conditioners, humidifiers, dehumidifiers, air purifiers, air filters, fans, furnaces, water heaters, boilers, space heaters, sowing machines, ice makers, microwave ovens, convection ovens, ovens, toaster ovens, ranges, range hoods, cooktops, stoves, stovetops, crock pots, hot plates, dishwashers, garbage disposals, can openers, vacuum cleaners, blenders, mixers, food processors, irons, coffee makers, toasters, grills, hair dryers, electric tooth brushes, electric razors, electric drills, electric screwdrivers, chainsaws, lawnmowers, push mowers, riding mowers, trimmers, brush cutters, pruners, edgers, vending machines;

Ventilation, Heating, Air-Conditioning, and Commercial Refrigeration Equipment:

Monitoring systems, control systems;

Engine, Turbine, and Power Transmission Equipment;

Monitoring systems, control systems;

Other General Purpose Machinery Manufacturing:
Monitoring systems, control systems;
Telecommunications:
Monitoring systems, control systems;
Portable;
Aircraft:
Monitoring systems, control systems, actuator systems, sensors.

It should be noted that devices within a specific category may be applicable across multiple areas even if they are not specifically listed. (e.g., temperature sensors apply to Industrial and Building Automation and Appliances).

To retrofit or redesign the devices listed, it is possible to implement the described systems in numerous ways. It may be advantageous to leave the device design as-is including the existing power supply. As an example, a device may use non-rechargeable batteries to operate. The device will most likely have a protection circuit to prevent damage if the batteries are installed incorrectly. The protection mechanism is commonly a diode inline with the positive terminal of the battery. In this case, the RF power harvester **20** with or without the alternative power source **24** could be inserted, with an antenna, into the device. The power generated by the RF power harvester **20** (and the alternative power source **24**, if applicable) could be connected to the device after the protection mechanism described to avoid potential charging of a non-rechargeable battery.

Another way to configure the device's power system is to replace the non-rechargeable batteries with rechargeable batteries. In this instance, the output from the RF power harvester **20** (and alternative power source **24**) could be connected to either side of the protections device. If the connection is before the protection mechanism, the RF power harvester **20** (and the alternative power source **24**, if applicable) will recharge the battery and supply power to the device. If the connection is after the protection mechanism, the RF power harvester **20** (and the alternative power source **24**, if applicable) will supply power to the device and the battery will supply any extra power needed that could not be supplied by the RF power harvester **20** (and the alternative power source **24**, if applicable). It should be noted that the protection device in this case is not needed for proper operation. Its only function would be to protect the batteries from being installed incorrectly. An antenna could be contained inside or placed on the outside of the device.

Another configuration of the device's power system is to remove the existing batteries and install the RF power harvester **20** (and the alternative power source **24**) in the enclosure provided for the batteries. An antenna could be contained inside or placed on the outside of the device.

Yet another method of configuring the device's power system would be to reduce the number of batteries and replace them with the RF power harvester **20** (and the alternative power source **24**). In this case, the output from the system would be connected to the batteries in series or parallel depending on the original battery configuration. An antenna could be contained inside or placed on the outside of the device.

An additional option, would be to completely redesign the product and integrate the required circuitry and storage components into the device. This method is probably the most advantageous because it can fully take advantage of the benefits offered by the RF power harvester **20** (and the alternative power source **24**). An antenna could be contained inside or placed on the outside of the device.

If the RF power harvester **20** (and the alternative power source **24**) is used as a backup to the primary power supply, a switch could be implemented into the system in order to switch the RF power harvester **20** (and the alternative power source **24**) on when the primary source is lost. In this case, an antenna could be contained inside or placed on the outside of the device.

To show the flexibility of RF energy harvesting, several products were retrofitted to include an RF power harvester **20** (energy harvesting circuitry). These products include a wireless keyboard, a wall clock, and a desk calculator.

The wireless keyboard is an example of recharging and augmenting a battery to supply power to a device. This system is shown in FIG. **13**. The output from the regulation circuitry recharges the battery and supplies power to the keyboard. The battery is also used to supply power to the keyboard. The keyboard also includes a separate antenna for receiving power and for data communications. The antenna configuration can be seen in FIG. **55**.

The wall clock is an example of a direct powering system. The wall clock was retrofitted to include an RF power harvester **20** and the internal AA battery was removed. This system is shown in FIG. **2**. The wall clock did not need regulation, but did require a capacitor for storage to supply the pulse of power to move the second hand.

The calculator is an example of using RF energy harvesting with another energy harvesting technology. The calculator had an internal 1.5V coin cell battery and a small solar panel. The internal battery was removed, however, the solar panel was left intact. This system is shown in FIG. **60**. In this system, the calculator can receive power from both the solar panel and the RF power harvester **20** to eliminate the need for a battery.

As an additional example, an RF power harvester **20** (energy harvesting circuit) similar to the ones shown in U.S. Pat. No. 6,615,074 (FIGS. **8**, **9**, **12***a*, **12***b*, **13**, **14**), was connected in series with a 0.5V solar cell. Individually, the solar cell was able to provide 0.480V to a 10 kilo-ohm resistor, which was being used to simulate the core device **22** components. This corresponds to 23 microwatts. The RF power harvesting circuit by itself was able to provide 2.093V across the 10 kilo-ohm resistor when being supplied by 1 milliwatt of RF power. This corresponds to 438 microwatts. The two circuit outputs were then combined in series by connecting the output from the RF energy harvesting circuit to the ground of the solar cell. The output of the solar cell was then connected to the resistor. The other end of the resistor was connected to the ground of the RF energy harvesting circuit. The voltage across the resistor with the circuits connected as shown in FIG. **63** was 2.445V. This corresponds to 598 microwatts.

As can be seen, the combination of the two technologies produces a result higher than the addition of the individual powers. From this, it can be determined that the two technologies can cooperate in a way that produces favorable results.

In the example given, the solar cell produces current to supply the load and helps to bias the RF rectifying diodes, which allows the RF energy harvesting circuit to operate a higher efficiency. The solar cell also changes the impedance seen by the RF energy harvesting circuit, which produces a beneficial result. To be more specific, when examining the power output of the individual circuit (solar and RF power harvesting), the sum of the power captured by the individual circuits was 23 μW+438 μW=461 μW. However, when the two circuits are combined and are allowed to work in conjunction with one another, the output power becomes 598 μW. This result shows that combining the two power-harvesting technologies produces a 30 percent increase in the output power for this example. This same technique can be applied to multiple energy harvesting technologies to produce even greater output power. The equations for this example are shown below.

Individual Circuits

$$P_I=P_1+P_2+\ldots+P_N$$

Combined Circuits

$$P_C>P_I=P_1+P_2+\ldots+P_N$$

where P_I is the sum of the individual output powers

P_C is the output of the combined circuit

P_1 is the output power from the first power harvesting technology

P_2 is the output power from the second power harvesting technology

P_N is the output power from the N^{th} power harvesting technology

N is the number of power harvesting technologies or circuits.

As can be seen by the previous examples, RF energy harvesting can be used alone or in conjunction with alternative power sources to power a wide range of devices. The addition of RF energy harvesting technology to the device allows for increased battery life, increased functionality, or the removal of the primary battery.

Although the invention has been described in detail in the foregoing embodiments for the purpose of illustration, it is to be understood that such detail is solely for that purpose and that variations can be made therein by those skilled in the art without departing from the spirit and scope of the invention except as it may be described by the following claims.

What is claimed is:

1. An apparatus, comprising:

a first antenna configured to wirelessly receive energy from a source that is remote from the apparatus and in a far-field region of the first antenna and a second antenna;

the second antenna different from and electrically isolated from the first antenna, the second antenna configured to wirelessly communicate with the source;

a power harvester electrically coupled to the first antenna and configured to convert the energy received at the first antenna to a DC voltage;

a radio-frequency identification (RFID) device electrically coupled to the power harvester and to the second antenna, the RFID device configured to wirelessly communicate with the source via the second antenna;

an integrated circuit configured to operate based on the DC voltage; and

a sensor electrically coupled to the integrated circuit, the sensor configured to operate based on the DC voltage and to detect at least one of an environmental condition associated with a location of the apparatus or a physical condition associated with an item coupled to the apparatus.

2. The apparatus of claim **1**, further comprising:

a regulator coupled to the power harvester and configured to regulate the DC voltage to produce a regulated DC voltage, the sensor configured to operate based on the regulated DC voltage received from the regulator.

3. The apparatus of claim **1**, wherein the integrated circuit is configured to perform an application associated with the at least one of the environmental condition or the physical condition detected by the sensor.

4. The apparatus of claim **1**, further comprising:

an electronic display configured to operate based on the DC voltage.

5. The apparatus of claim **1**, wherein the sensor includes at least one of a temperature sensor, a motion sensor, a vibration sensor, a humidity sensor, an orientation sensor, or a position sensor.

6. The apparatus of claim **1**, wherein the RFID device does not operate based on a battery, and the sensor does not operate based on a battery.

7. The apparatus of claim **1**, wherein the apparatus is a RFID tag.

8. An apparatus, comprising:

an antenna configured to wirelessly receive energy from a source that is remote from the apparatus and in a far-field region of the antenna;

a power harvester electrically coupled to the antenna and configured to convert the energy received at the antenna to a DC voltage;

a radio-frequency identification (RFID) device electrically coupled to the antenna and configured to wirelessly communicate with the source via the antenna, the RFID device operatively coupled to the power harvester and including an integrated circuit configured to operate based on the DC voltage; and

a sensor electrically coupled to the RFID device and the power harvester, the sensor configured to operate based on the DC voltage and to detect at least one of an environmental condition associated with a location of the apparatus or a physical condition associated with an item coupled to the apparatus,

the integrated circuit configured to perform an application associated with the at least one of the environmental condition or the physical condition detected by the sensor.

9. The apparatus of claim **8**, wherein the power harvester is configured to wirelessly communicate with the source via the antenna.

10. The apparatus of claim **8**, wherein the sensor is electrically coupled to the antenna and is configured to wirelessly communicate with the source via the antenna.

11. The apparatus of claim **8**, further comprising:

a regulator coupled to the power harvester and configured to regulate the DC voltage to produce a regulated DC voltage, the integrated circuit configured to operate based on the regulated DC voltage received from the regulator.

12. The apparatus of claim **8**, further comprising:

an electronic display configured to operate based on the DC voltage.

13. The apparatus of claim **8**, wherein the sensor includes at least one of a temperature sensor, a motion sensor, a vibration sensor, a humidity sensor, or a position sensor.

14. The apparatus of claim **8**, wherein the RFID device does not operate based on a battery, and the sensor does not operate based on a battery.

15. An apparatus, comprising:

an antenna configured to wirelessly receive energy from a source that is remote from the apparatus and in a far-field region of the antenna;

a power harvester electrically coupled to the antenna and configured to convert the energy received at the antenna to a DC voltage to produce a power associated with the DC voltage;

a regulator coupled to the power harvester and configured to regulate the power associated with the DC voltage to produce a regulated power;

a radio-frequency identification (RFID) device electrically coupled to the antenna and configured to wirelessly communicate with the source via the antenna, the RFID device electrically coupled to the power harvester and the regulator, the RFID device including an integrated circuit configured to operate based on at least one of the power received from the power harvester or the regulated power received from the regulator; and

an electronic display electrically coupled to the RFID device and configured to operate, at least in part, based on the regulated power received from the regulator.

16. The apparatus of claim **15**, wherein the electronic display is electrically coupled to the regulator, the electronic display is configured to operate based on the regulated power received from the regulator and the power received from the power harvester.

17. The apparatus of claim **15**, wherein the electronic display is one of a liquid-crystal display (LCD) or a light-emitting diode (LED) display.

18. The apparatus of claim **15**, further comprising:

a sensor configured to operate, at least in part, based on the regulated power received from the regulator, the sensor including at least one of a temperature sensor, a motion sensor, a vibration sensor, a humidity sensor, or a position sensor.

19. The apparatus of claim **15**, wherein the power harvester is configured to wirelessly communicate with the source via the antenna.

20. The apparatus of claim **15**, wherein the RFID device does not operate based on a battery, and the electronic display does not operate based on a battery.

* * * * *

US009768711B2

(12) **United States Patent**
Hameed et al.

(10) **Patent No.: US 9,768,711 B2**
(45) **Date of Patent: Sep. 19, 2017**

(54) **RF-DC POWER CONVERTER**

(71) Applicant: **THE GOVERNORS OF THE UNIVERSITY OF ALBERTA**, Edmonton (CA)

(72) Inventors: **Zohaib Hameed**, Edmonton (CA); **Kambiz Moez**, Edmonton (CA)

(73) Assignees: **Zohaib Hameed**, Saint Paul, MN (US); **Kambiz Moez**, Edmonton, Alberta (CA)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **14/740,021**

(22) Filed: **Jun. 15, 2015**

(65) **Prior Publication Data**

US 2015/0365013 A1 Dec. 17, 2015

Related U.S. Application Data

(60) Provisional application No. 62/011,957, filed on Jun. 13, 2014.

(51) **Int. Cl.**

H02M 7/25	(2006.01)
H02M 3/07	(2006.01)
H02M 7/10	(2006.01)

(52) **U.S. Cl.**
CPC ***H02M 7/25*** (2013.01); *H02M 3/07* (2013.01); *H02M 7/103* (2013.01)

(58) **Field of Classification Search**
CPC .. H02M 7/25; H02M 3/073; H02M 2003/078; H02M 7/103; H02M 3/07
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,847,518 A *	7/1989	Leidich	G01R 1/20 327/315
4,935,644 A *	6/1990	Tsujimoto	H02M 3/073 327/536
6,037,622 A *	3/2000	Lin	G11O 5/145 257/299
6,130,574 A *	10/2000	Bloch	G11C 16/30 327/536
6,172,886 B1 *	1/2001	Lauterbach	H02M 3/073 327/536
6,359,501 B2 *	3/2002	Lin	H02M 3/073 327/536
6,603,346 B2 *	8/2003	Sawada	H02M 3/073 327/536

(Continued)

Primary Examiner — Timothy J Dole
Assistant Examiner — Sisay G Tiku

(57) **ABSTRACT**

A rectifier comprising a chain of transistors for RF-DC conversion. In order to compensate for the thresholds of the transistors, each transistor can be connected to a junction earlier or later in the chain. By using both p-type and n-type transistors in the same chain, the different types of transistors can be compensated in different directions allowing more transistors to be compensated. Additional transistors connected to the gates of transistors of the main chain can allow the transistors of the main chain to be forward compensated at one part of the input cycle and backward compensated in another part to minimize both the voltage threshold of the rectifier and the leakage current. The line for compensation of the voltage threshold during forward conduction can comprise a solid line or a transistor, and if a transistor is used it may be diode-connected.

15 Claims, 14 Drawing Sheets

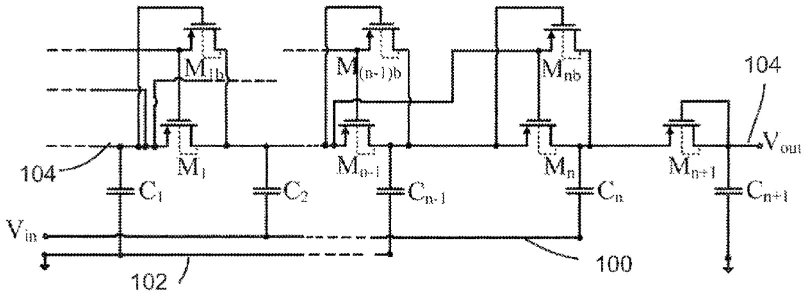

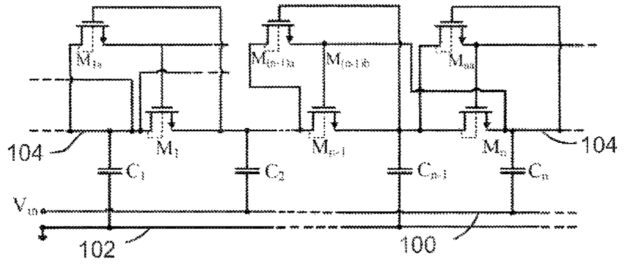

(56) References Cited

U.S. PATENT DOCUMENTS

6,677,806	B2 *	1/2004	Bloch	H02M 3/073 327/536
7,586,361	B2 *	9/2009	Fukuda	H02M 3/073 327/536
7,592,857	B2 *	9/2009	Wu	H02M 3/073 327/536
2014/0268964	A1 *	9/2014	Xia	H02M 7/25 363/127

* cited by examiner

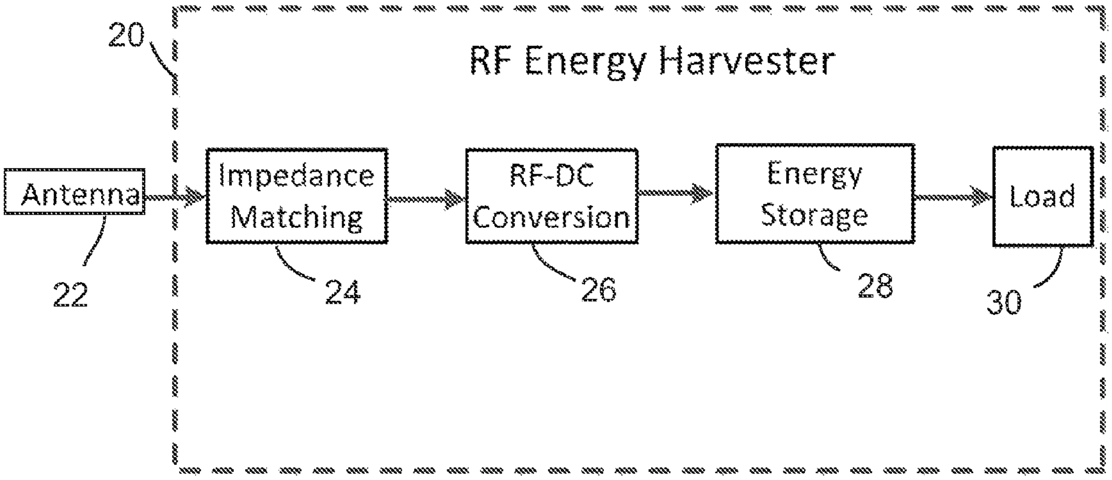
20
RF Energy Harvester
Antenna
Impedance Matching
RF-DC Conversion
Energy Storage
Load
22
24
26
28
30

Fig. 1 (Prior Art)

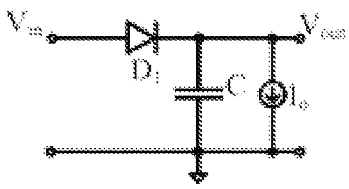
Vin
D1
C
Io
Vout

Fig. 2A (Prior Art)

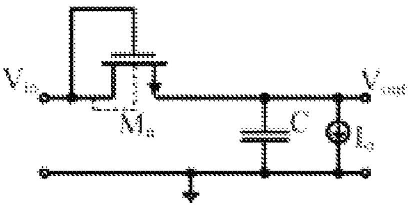
Vin
Mn
C
Io
Vout

Fig. 2B (Prior Art)

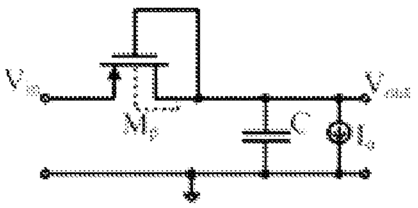
Vin
Mp
C
Io
Vout

Fig. 2C (Prior Art)

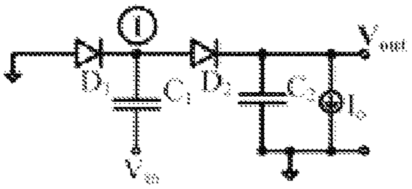
D1
C1
D2
C2
Io
Vout
Vin

Fig. 2D (Prior Art)

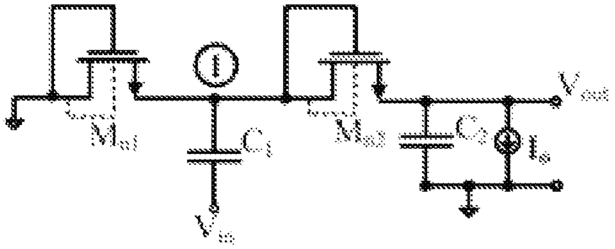
I
V_out
M_a1
C_1
M_a2
C_2
I_s
V_in

Fig. 2E (Prior Art)

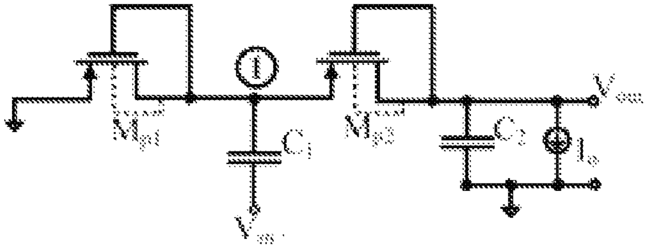
I
V_out
M_p1
C_1
M_p2
C_2
I_s
V_in

Fig. 2F (Prior Art)

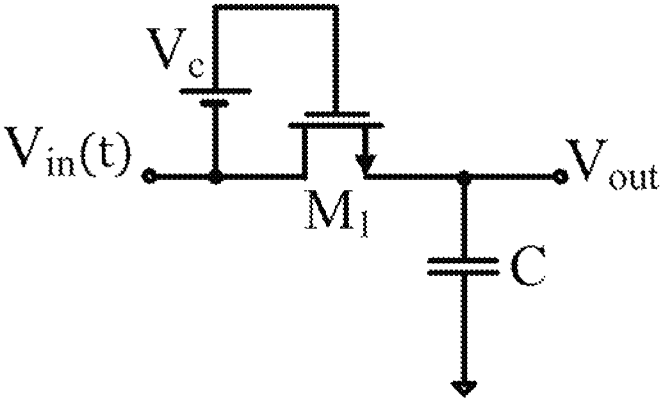
V_c
V_in(t)
M_1
V_out
C

Fig. 2G (Prior Art)

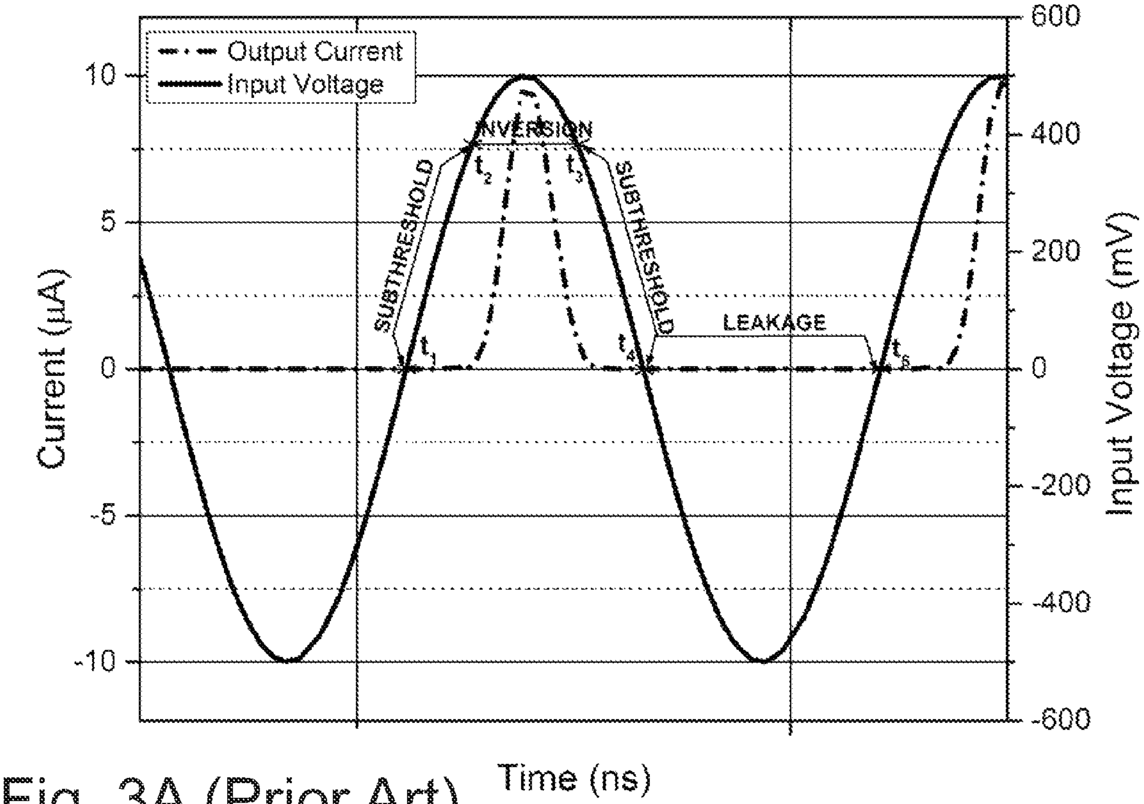

Fig. 3A (Prior Art)

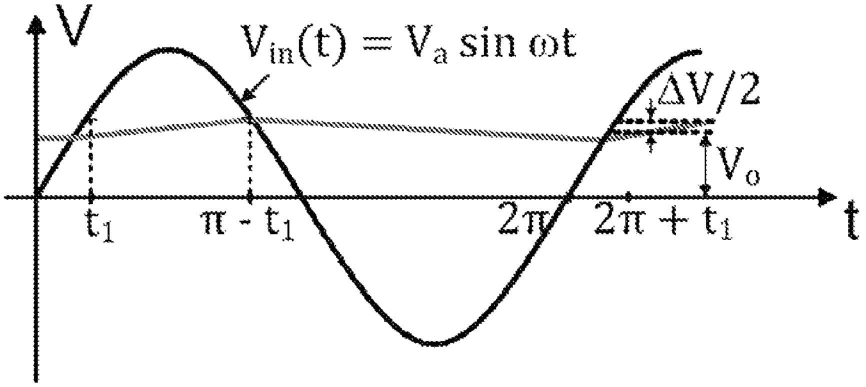
V
$V_{in}(t) = V_a \sin \omega t$
$\Delta V/2$
V_o
t_1
$\pi - t_1$
2π
$2\pi + t_1$
t

Fig. 3B (Prior Art)

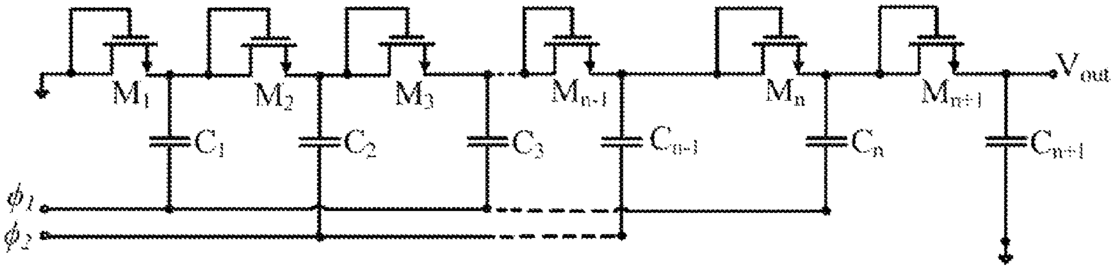

M_1
M_2
M_3
M_{n-1}
M_n
M_{n+1}
V_{out}
C_1
C_2
C_3
C_{n-1}
C_n
C_{n+1}
ϕ_1
ϕ_2

Fig. 4A (Prior Art)

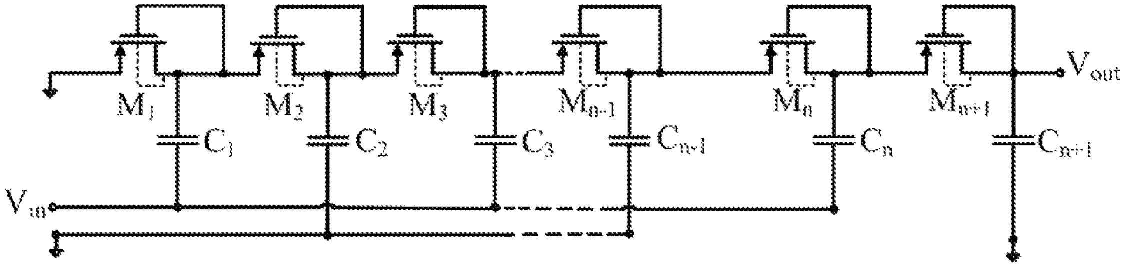

M_1
M_2
M_3
M_{n-1}
M_n
M_{n+1}
V_{out}
C_1
C_2
C_3
C_{n-1}
C_n
C_{n+1}
V_{in}

Fig. 4B (Prior Art)

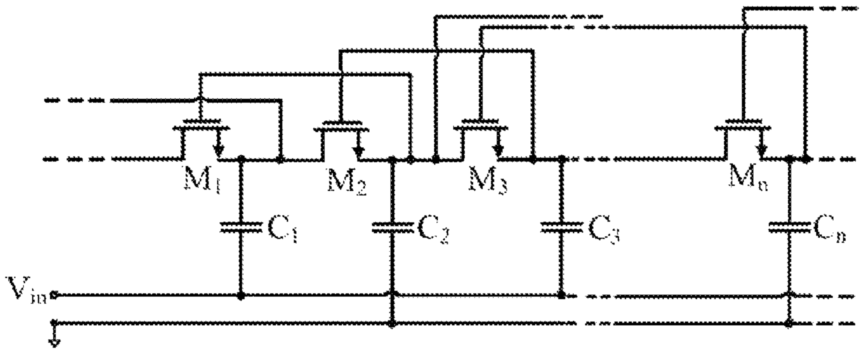
M_1
M_2
M_3
M_n
C_1
C_2
C_3
C_n
V_{in}

Fig. 5 (Prior Art)

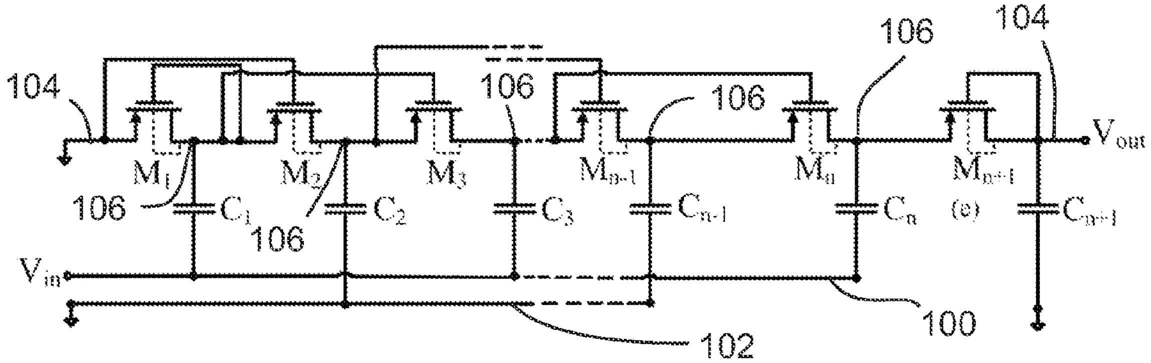
104
106
106
106
104
V_{out}
M_1
M_2
M_3
M_{n-1}
M_n
M_{n+1}
(e)
106
106
C_1
C_2
C_3
C_{n-1}
C_n
C_{n+1}
V_{in}
102
100

Fig. 6A

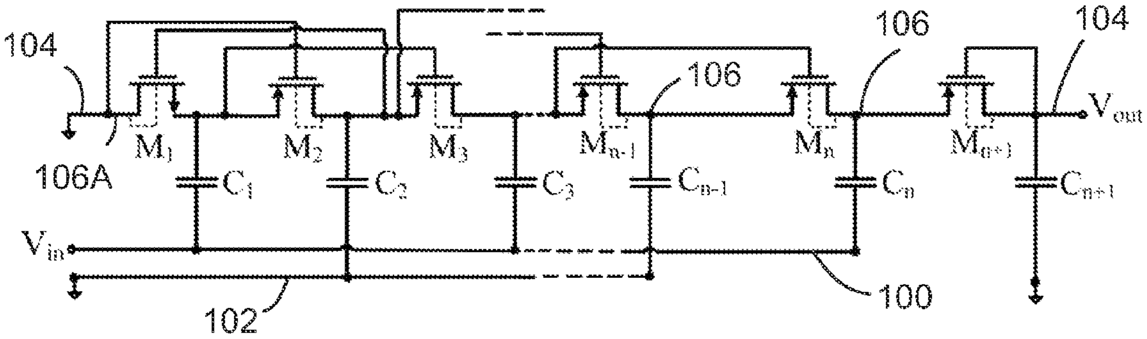
104
106A
M_1
M_2
M_3
M_{n-1}
106
M_n
106
M_{n+1}
104
V_{out}
C_1
C_2
C_3
C_{n-1}
C_n
C_{n+1}
V_{in}
102
100

Fig. 6B

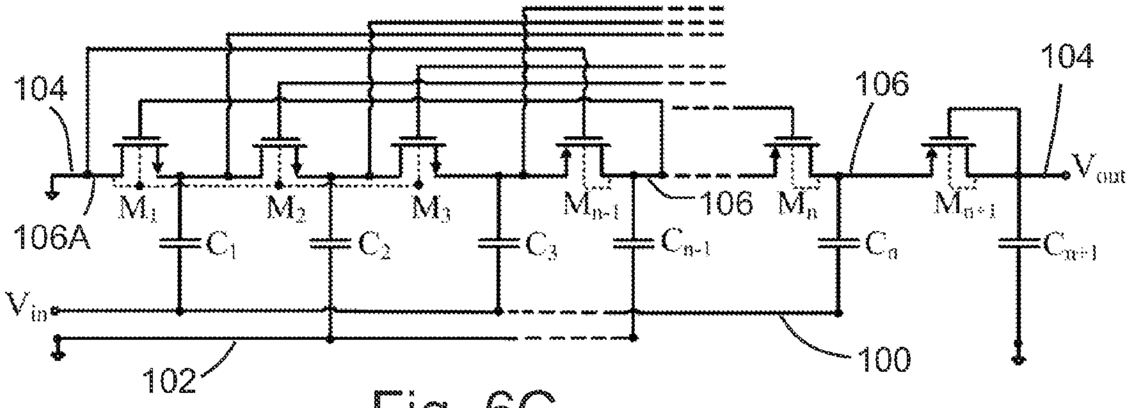
104
106A
M_1
M_2
M_3
M_{n-1}
106
M_n
106
M_{n+1}
104
V_{out}
C_1
C_2
C_3
C_{n-1}
C_n
C_{n+1}
V_{in}
102
100

Fig. 6C

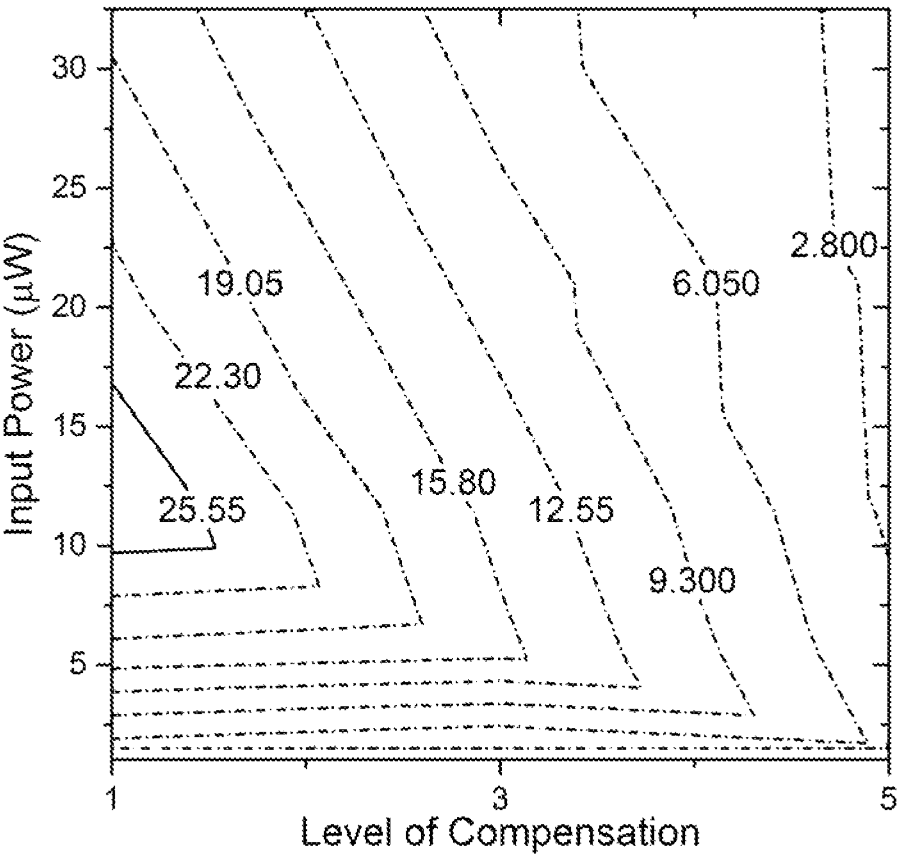
Input Power (μW)
30
25
20
15
10
5
Level of Compensation
1
3
5
19.05
22.30
25.55
15.80
12.55
9.300
6.050
2.800

Fig. 7

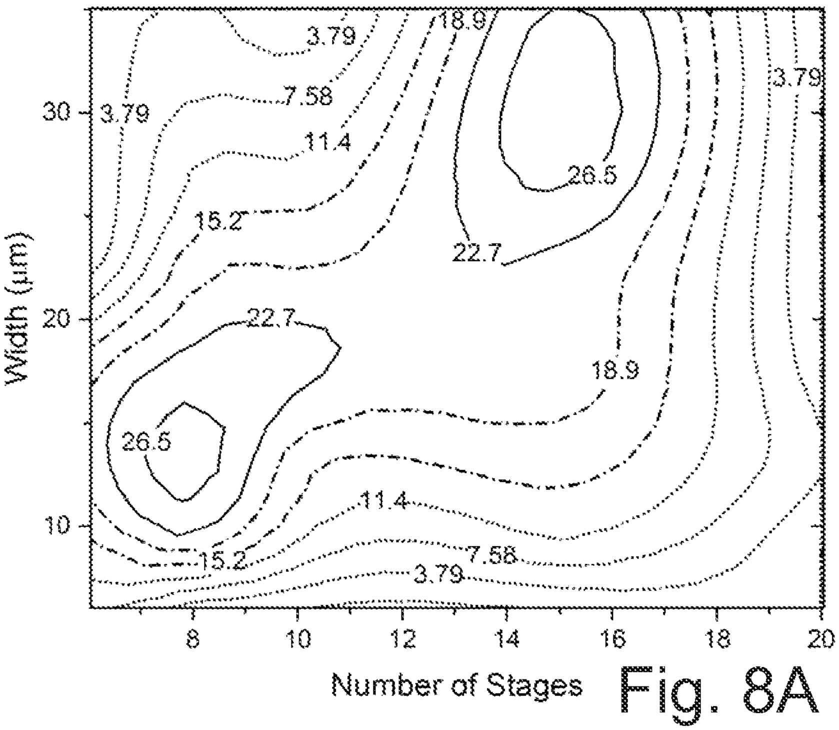
3.79
18.9
7.58
3.79
3.79
11.4
26.5
15.2
22.7
22.7
18.9
26.5
11.4
15.2
7.58
3.79
Width (μm)
30
20
10
8
10
12
14
16
18
20
Number of Stages

Fig. 8A

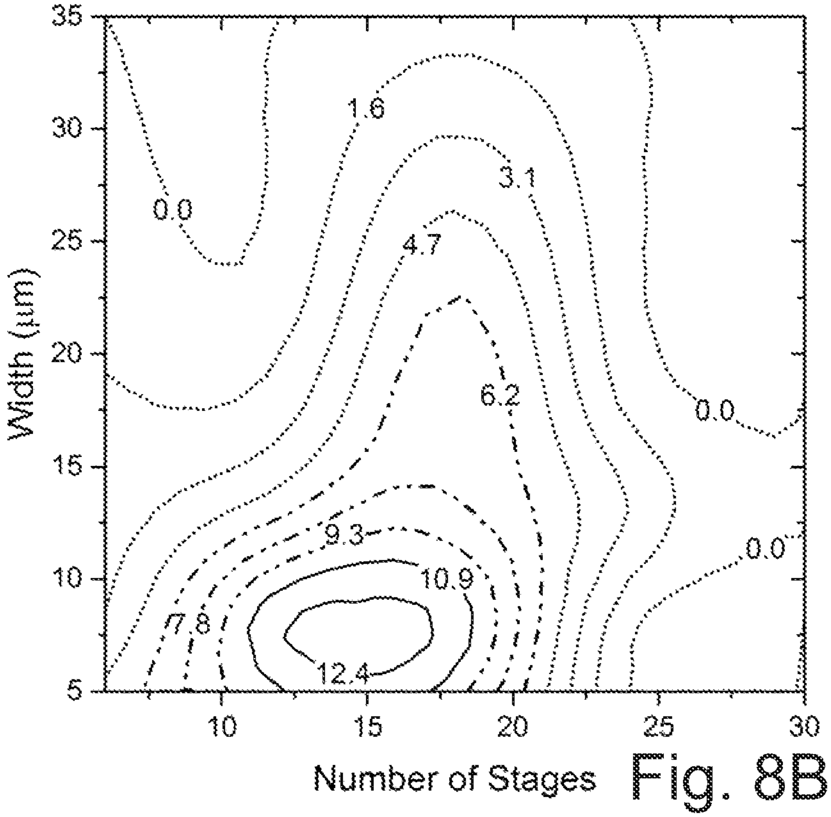
1.6
3.1
0.0
4.7
6.2
0.0
9.3
0.0
10.9
7.8
12.4
Width (μm)
35
30
25
20
15
10
5
10
15
20
25
30
Number of Stages

Fig. 8B

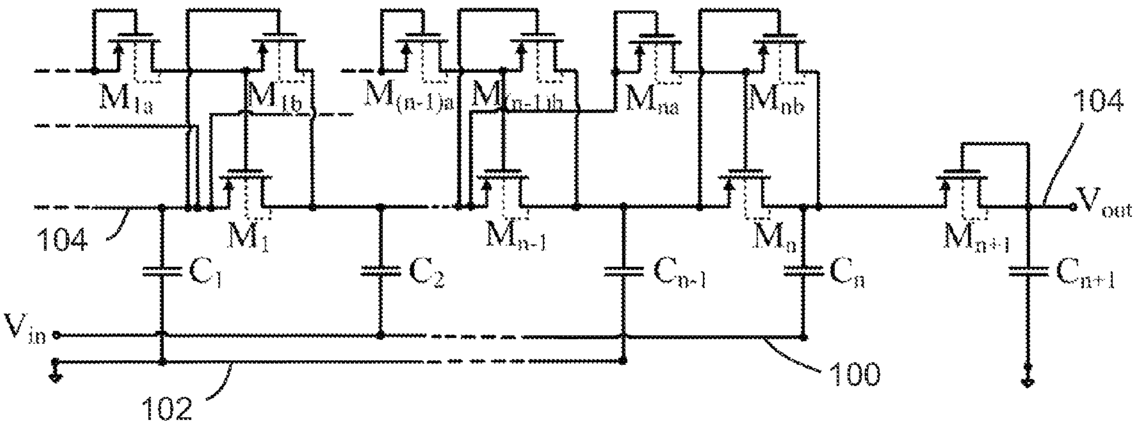
M1a
M1b
M(n-1)a
M(n-1)b
Mna
Mnb
104
Vout
104
M1
Mn-1
Mn
Mn+1
C1
C2
Cn-1
Cn
Cn+1
Vin
100
102

Fig. 9A

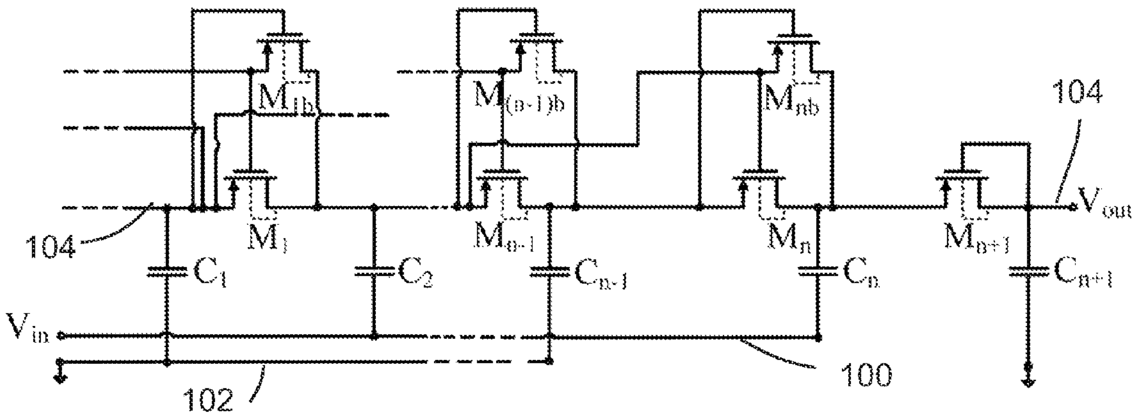
M1b
M(n-1)b
Mnb
104
Vout
104
M1
Mn-1
Mn
Mn+1
C1
C2
Cn-1
Cn
Cn+1
Vin
100
102

Fig. 9B

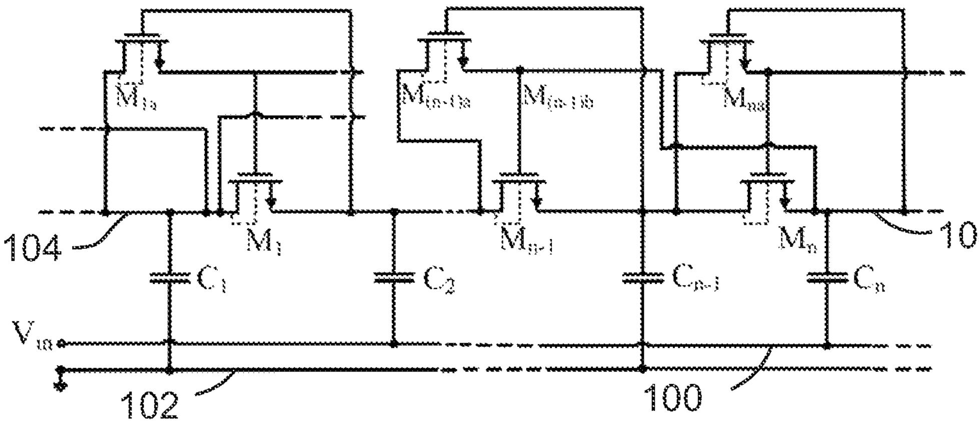
M_{1a}
$M_{(n-1)a}$
$M_{(n-1)b}$
M_{na}
104
M_1
M_{n-1}
M_n
104
C_1
C_2
C_{n-1}
C_n
V_{in}
102
100

Fig. 9C

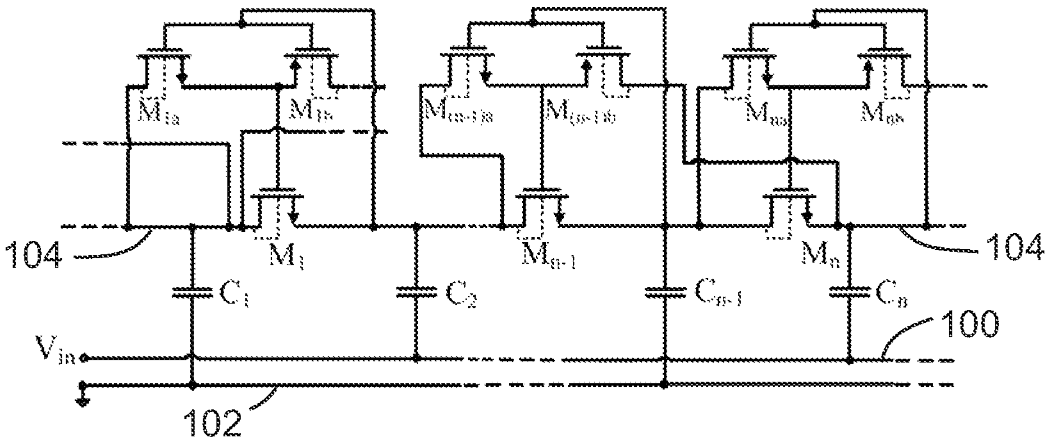
M_{1a}
M_{1b}
$M_{(n-1)a}$
$M_{(n-1)b}$
M_{na}
M_{nb}
104
M_1
M_{n-1}
M_n
104
C_1
C_2
C_{n-1}
C_n
100
V_{in}
102

Fig. 9D

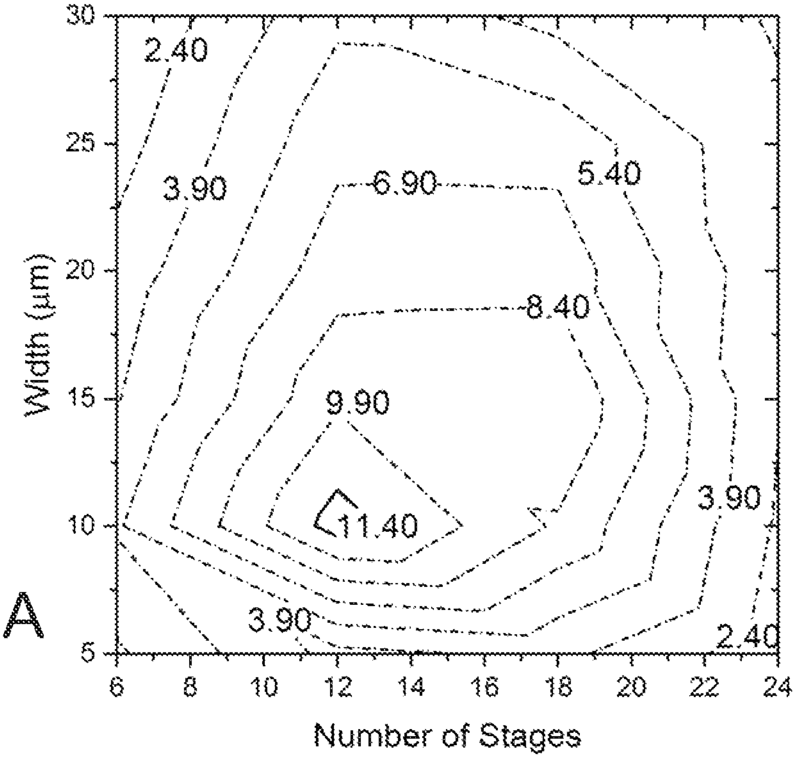

Fig. 10A

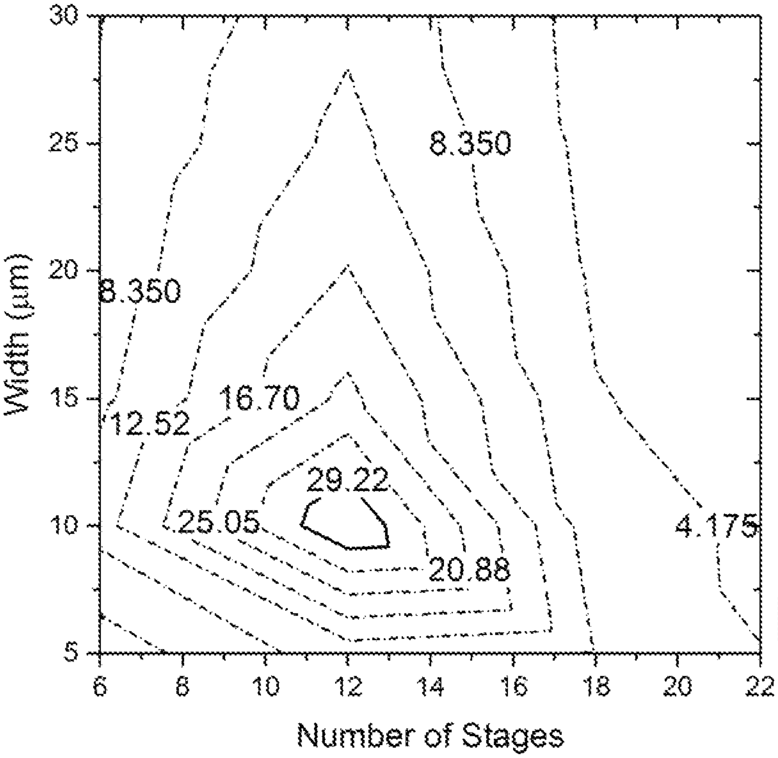

Fig. 10B

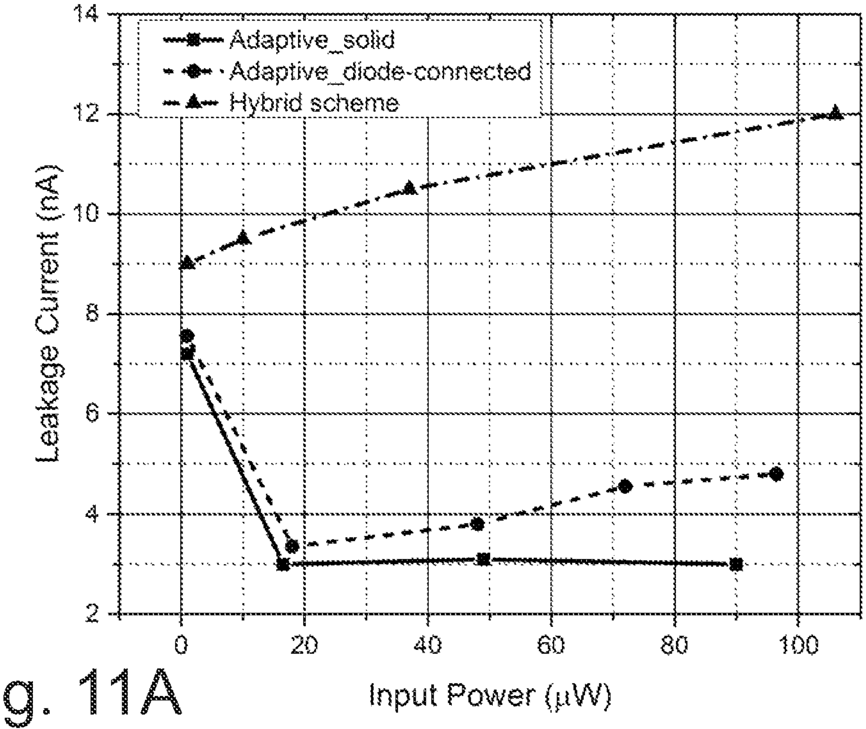

Fig. 11A

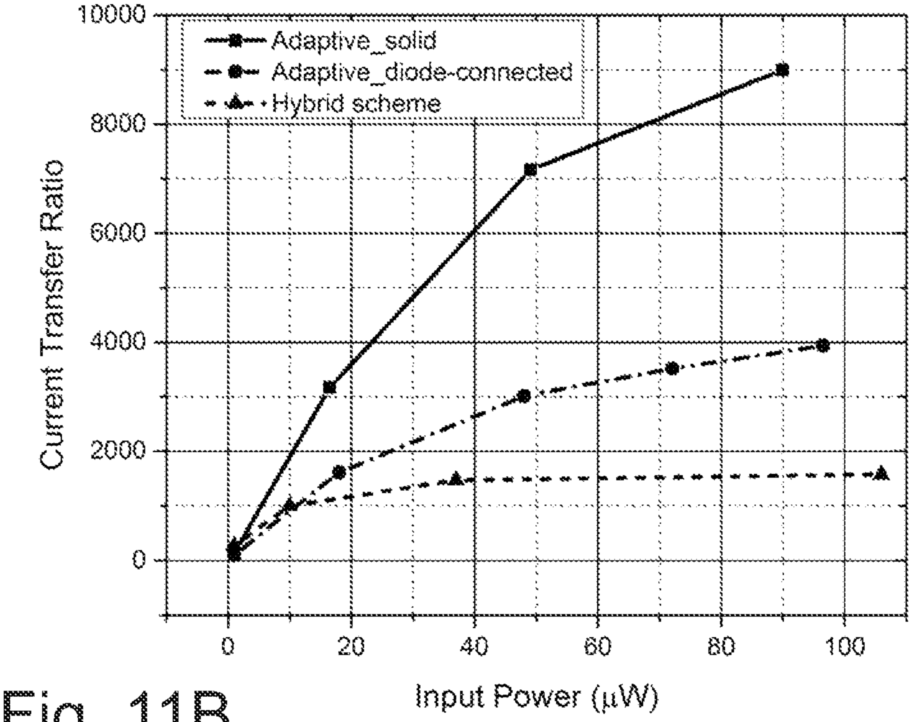

Fig. 11B

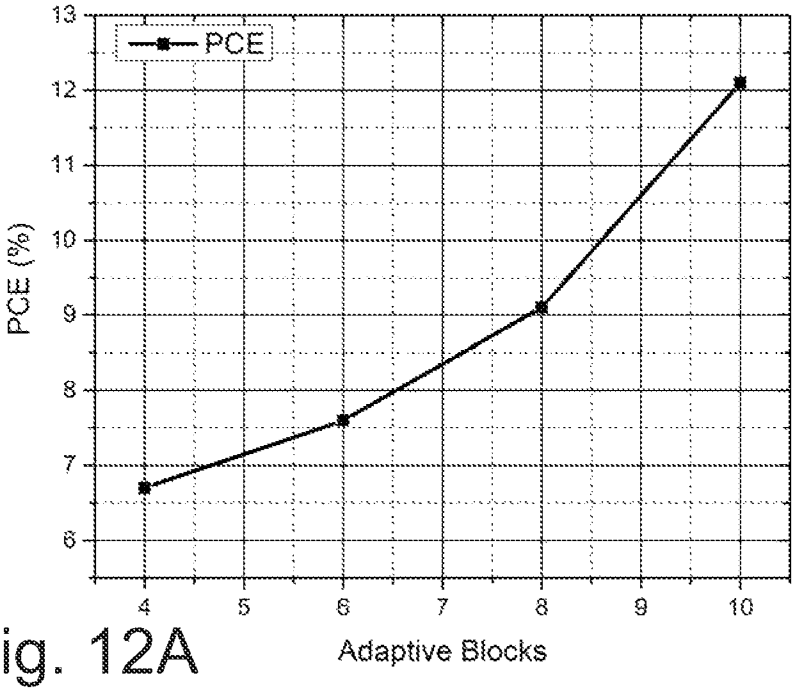

Fig. 12A

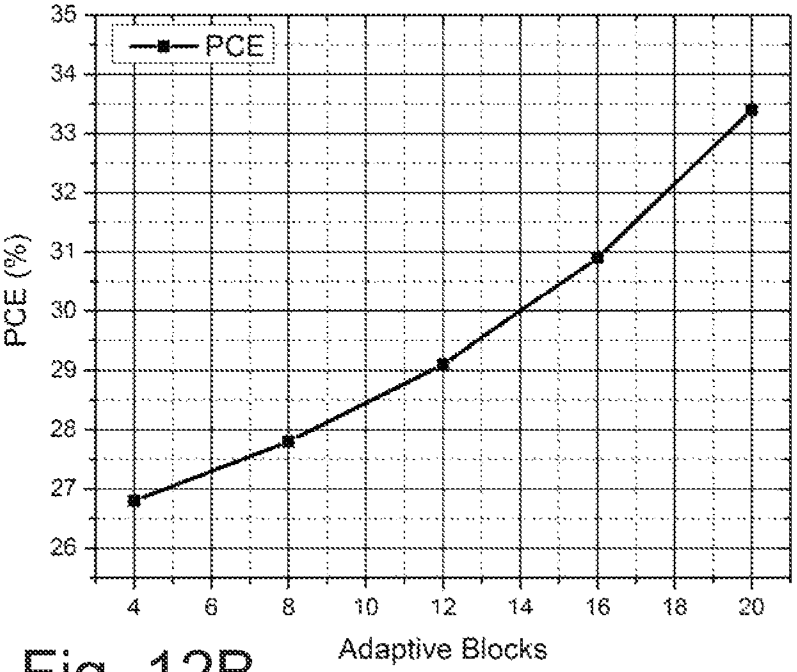

Fig. 12B

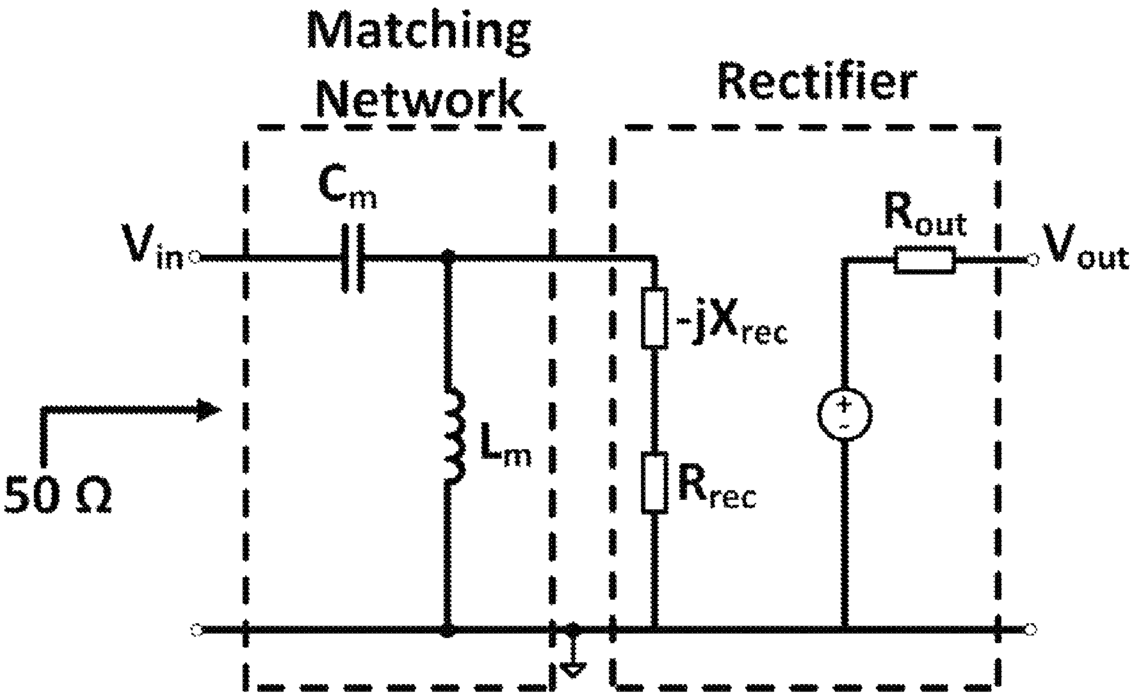
Matching Network
Rectifier
C_m
V_{in}
L_m
50 Ω
$-jX_{rec}$
R_{rec}
R_{out}
V_{out}
+
-

Fig. 13

RF-DC POWER CONVERTER

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit under 35 USC 119(e) of U.S. provisional application Ser. No. 62/011,957 filed Jun. 13, 2014.

TECHNICAL FIELD

Rectification, particularly for RF energy harvesting.

BACKGROUND

The vision of realizing Internet of Things (IoT) pervasively connecting large number of sensors and devices requires development of novel solutions that scavenges ambient energy to supply the power required for the operation of the sensors. Relying on batteries as the source of energy for wireless sensors impose several limitations including the need for routine maintenance/charging of batteries, operation interruption and cost involved in replacing batteries specially those employed in harsh environments, and challenges of scaling of battery-powered wireless sensors to millions of nodes. RF energy from ambient sources can be used to power efficiently the sensor networks with or without batteries. In addition, they can also be used to partially/fully charge other portable electronic devices to extend their battery life.

There is accordingly a growing interest in harvesting ambient energy to partially/fully supply the energy required for the operation of portable electronic devices. Scavenging energy from the ambient electromagnetic wave referred as RF energy harvesting is one of the most popular method for powering low-power wireless sensors. As most of today's integrated circuits are fabricated in CMOS technology, it is highly desirable to integrate RF energy harvesting systems with the rest of the system on a single CMOS chip for reduced cost and small form factor. RF powered devices are widely used in wireless sensor networks operating at power levels of 10 to 100's of μW. RF energy harvesting can be used in applications where the traditional energy sources such as light, vibrations are not available.

In an RF energy harvesting system **20** as shown in FIG. **1**, an antenna **22** receives the incident RF signal, an impedance matching circuit **24** maximizes the power transfer from the antenna to the power converter **26**, and a RF-DC power conversion circuit **26** converts the incident RF power to DC output power. The output DC voltage is stored in an energy storage component **28** (battery or capacitor) or can directly power the wireless sensors. The major challenge of scavenging RF energy is the limited signal strength of the RF waves and the low efficiency of the harvesting circuit at low input power.

The radiated signals are received by the RF energy harvesting system **20** and converted into a DC output voltage which is used to power-up the system. The radiated signals have to adhere to the regulation standards set by the federal communication commission (FCC). The unlicensed industrial, scientific and medical (ISM) band with a center frequency of 915 MHz operating at 902-928 MHz with a maximum effective isotropic radiated power (EIRP) of 4 W can be used for RF energy harvesting. The major challenge of scavenging RF energy is the limited signal strength of the RF waves and the low efficiency of the harvesting circuit. The signal strength is limited due to the attenuation in the

signal due to the free space loss. Also, the power at the transmitter level cannot be increased without violating the FCC guidelines. Hence modification has to be done at the receiver side. The power consumption for the commercial sensor network nodes vary based on the manufacturers and has been estimated by various authors to be around 10 to 100 μW depending on the sensing application and the radio protocol. A multi-volt supply voltage is typically required for the operation of the sensor circuitry.

The power harvester unit **20** comprising a multi-stage rectifier **26** is a key component in RF energy harvesting systems. It converts the incoming weak RF signal into a DC voltage. The performance of the rectifier unit **26** can be evaluated based on its power conversion efficiency (PCE) which is the ratio of power delivered to the load to the input power, based on sensitivity i.e. the minimum input power required for production of a DC voltage at the output and finally based on output DC voltage levels. To increase rectifier's PCE, the energy losses such as those introduced by the non-zero ON resistances of the rectifying devices must be reduced. To increase sensitivity and output voltage levels, rectifying devices with lower threshold voltages are required. Hence, these performance parameters of the power harvester are strongly affected by the threshold voltage of the rectifying devices. The power level has to exceed the threshold voltage of the rectifying device to turn on the rectifier unit in the RF-DC power conversion circuit. The minimum power is referred as the power-up threshold of the system. In terms of voltage level, a minimum input voltage is required for the circuit operation and is referred as the dead-zone of the rectifier.

The performance of an RF energy harvesting system **20** is significantly affected by the threshold voltage of the rectifying device **26**, the voltage that is required to turn on the semiconductor devices used as rectifying devices. A rectifying device **26** with lower threshold voltage enables the operation of RF-DC power converter **26** at low input power levels significantly reducing the power-up threshold of the rectifiers, and increasing the output voltage level for the same input power. The threshold voltage of device can be reduced using different technology-based approaches for the devices including silicon-on-sapphire (SOS), schottky diodes such as silicon-titanium schottky diodes or SMS and the HSMS diodes, special low-threshold-voltage transistors in CMOS process, and floating gate transistors which store a pre-charged voltage at the gate to lower the threshold voltage. The drawback of using technology-based approaches is additional fabrication steps that increases the production cost and prevents integration of RF energy harvester in mainstream Complementary Metal-Oxide-Silicon Integrated Circuits (CMOS ICs). Active/passive circuit techniques can be alternatively used to reduce the threshold voltage of the device. Active technique requires an external power source or secondary battery and is generally used in active sensors or active RFID. This enables more sophisticated applications at the price of increased cost and maintenance. Passive techniques do not require an additional source of energy but may require additional circuit where an auxiliary rectification chain is used to generate the compensating threshold voltage for the main RF-DC power conversion circuit. The auxiliary chain requires additional power and occupies additional area. An internal threshold voltage cancellation circuit was introduced in Kocer et al. where the compensating voltage was generated passively and stored in a capacitor that is applied at the gate-source terminal of the MOS transistor. This technique requires a large resistance and high capacitance value which leads to a large area on the chip. A self-biasing technique consisting of an off-chip high impedance resistive network was used by Li et al. to provide DC biasing voltages. Another technique can be using floating gate transistors storing a pre-charged voltage at the gate lowering its threshold voltage. The threshold voltage can be varied by changing the body-source potential of the transistor. This technique requires an additional pre-charge phase making it unsuitable for fully battery-less applications. The RF-DC power conversion circuit consisting of NMOS transistors with grounded body terminal leads to an increase in the threshold with the number of stages due to the body effect. This degrades the efficiency of the power conversion circuit. The body terminal of the transistors can be dynamically controlled using additional circuit or floating well devices. The floating well technique generates undesirable substrate current. Also, a triple-well source-body connected device has high parasitic capacitance leading to reduced efficiency. A cross-coupled differential scheme was used in Sciorcioni et al. consisting of triple-well NMOS and standard PMOS transistors to reduce the threshold voltage. A self-compensation scheme based on the Dickson topology was introduced in Curty et al. consisting of triple-well NMOS transistors to provide individual body biasing. The compensating voltage was provided by connecting the gate terminal to later stages. The design of the power converter in the works was focused on only reducing the threshold voltage neglecting the key role played by the reverse leakage current in introducing power losses. The reduction in threshold voltage increases the reverse conduction current causing additional loss.

To address the trade-off between the reduced threshold voltage and increased leakage current, Lee et al. proposed use of high speed comparators to control the reverse leakage current. The use of comparator increases the power consumption and limits the usefulness of this technique to low-frequency applications. Dynamic CMOS Dickson pump has been designed in previous works to eliminate the V_{th} drop while reducing the reverse leakage current. These circuits are designed for digital application with no emphasis on low-power operation. Differential-drive (4T-cell) architecture with the cross coupled bridge configuration and its variant has been used in previous works to reduce the threshold voltage as well as lower the leakage current. Differential circuit requires a PCB balun for the single-ended to differential conversion or differential antenna. Also, the differential circuit requires triple-well NMOS transistors and larger number of rectifying devices for the same number of stages compared with single-ended one.

SUMMARY

A power conversion circuit is provided having a first input line and a second input line, the first and second input lines configured to receive an alternating voltage differential between the first and second input lines, a multi-stage rectifier comprising transistors arranged in series, each transistor having a gate, a source and a drain, adjacent transistors of the series being connected so that for adjacent p-type transistors the drain of the left p-type transistor is connected to the right adjacent source of the right p-type transistor and source of the right p-type transistor is connected to the left adjacent drain of the left p-type transistor, for adjacent n-type transistors the source of the left n-type transistor is connected to the right adjacent drain of the right n-type transistor and drain of the right n-type transistor is connected to the left adjacent source of the left n-type transistor, for a p-type transistor adjacent to an n-type transistor the source of the right p-type transistor is connected to the source of the left adjacent n-type transistor and drain of the left p-type transistor is connected to the drain of the right adjacent n-type transistor to form a junction, each junction being connected to one of the first input line and the second input line via a capacitor, with adjacent junctions having one junction of the adjacent junctions connected to the first input line and the other junction of the adjacent junctions connected to the second input line, the gate of each transistor of the multi-stage rectifier being connected to a respective junction that is not a junction formed by the connection of that transistor to an adjacent transistor but a junction in the previous stage for the p-type transistor or later stage for the n-type transistor, at least one of the transistor in the multi-stage rectifier being p-type and at least one being n-type.

In various embodiments, there may be included any one or more of the following features: There may also be an auxiliary chain of p-type transistors, each auxiliary transistor having a gate, a source and a drain, each transistor of the auxiliary chain being connected to a p-type transistor of the multi-stage rectifier in the main chain, so that the gate of the respective auxiliary transistor is connected to the source of the respective transistor in the main chain, the source of the respective auxiliary transistor is connected to the gate of the transistor in the main chain and also connected to the previous stage junction of the main chain which is N transistors away from the respective transistor, and the drain of the respective auxiliary transistor is connected to the drain of the respective transistor or to the junction of the later stage transistor in the main chain. There may also be a diode connected transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of previous stages which is N transistors away from the respective transistor. There may also be an auxiliary chain of n-type transistors, each auxiliary transistor having a gate, a source and a drain, each transistor of the auxiliary chain being connected to an n-type transistor of the multi-stage rectifier in the main chain, so that the gate of the respective auxiliary transistor is connected to the source of the respective transistor in the main chain, the source of the respective auxiliary transistor is connected to the gate of the transistor in the main chain and also connected to the later stage junction of the main chain which is N transistors away from the respective transistor, and the drain of the respective auxiliary transistor is connected to the drain of the respective transistor or to the junction of the previous stage transistor in the main chain. There may also be a diode connected transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of later stages which is N transistors away from the respective transistor.

There is also provided an auxiliary chain of transistors, each auxiliary transistor having a gate, a source and a drain, each transistor of the auxiliary chain being connected to a respective transistor of a main chain of a multi-stage rectifier in the main chain, so that the gate of the respective auxiliary transistor is connected to the source of the respective transistor in the main chain, the source of the respective auxiliary transistor is connected to the gate of the transistor in the main chain and also connected to the later stage junction of the main chain which is N transistors away from the respective transistor in the case that the respective transistor is n-type, or to the earlier stage junction of the main chain which is N transistors away from the respective transistor in the case that the respective transistor is p-type, and the drain of the respective auxiliary transistor is connected to the drain of the respective transistor or to the junction of a previous stage transistor in the main chain, in the case that the respective transistor is n-type, or to the junction of a later stage transistor in the main chain, in the case that the respective transistor is p-type.

In various embodiments, there may be included any one or more of the following features: the transistors of the main chain and auxiliary chain may be n-type and there may also be a p-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of later stages, such that the gate of the additional p-type auxiliary transistor is connected to the gate of the n-type auxiliary transistor and the source of the n-type and p-type auxiliary transistors are connected, and the drain of the p-type auxiliary transistors is connected to junction of the main chain which is at 'N' later stages from the respective transistor. The transistors of the main chain and auxiliary chain may be p-type and there may also be an n-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of earlier stages, such that the gate of the additional n-type auxiliary transistor is connected to the gate of the p-type auxiliary transistor and the source of the n-type and p-type auxiliary transistors are connected, and the drain of the n-type auxiliary transistors is connected to junction of the main chain which is at 'N' earlier stages from the respective transistor. There may also be a diode connected p-type or n-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of later stages which is N transistors away from the respective transistor. There may also be a diode connected p-type or n-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of earlier stages which is N transistors away from the respective transistor.

In any of the above embodiments, there may be included any one or more of the following features: There may also be an additional transistor connected in series with the multistage rectifier, the additional transistor having a gate, a source and a drain, the gate of the additional transistor being connected to the drain of the additional transistor. The gate of each transistor in the main chain may be connected to the junction of previous stages for p-type transistor and later stages for n-type transistor. N of the transistors in the multistage rectifier arranged in series may be n-type and all but N of the transistors in multistage rectifier arranged in series may be p-type. The second input line may be grounded. The body terminal of each p-type transistor may be connected to the respective drain and the body terminal of each n-type transistor may be either grounded or connected to the respective drain terminal.

These and other aspects of the device and method are set out in the claims, which are incorporated here by reference.

BRIEF DESCRIPTION OF THE FIGURES

Embodiments will now be described with reference to the figures, in which like reference characters denote like elements, by way of example, and in which:

FIG. **1** is a block diagram showing energy flow from a receiver through a power harvester to a load;

FIG. **2**A is a schematic diagram of a prior art half-wave rectifier using a diode;

FIG. **2**B is a schematic diagram of a prior art half-wave rectifier using diode-connected NMOS;

FIG. **2**C is a schematic diagram of a prior art half-wave rectifier using diode connected PMOS;

FIG. **2**D is a schematic diagram of a prior art voltage doubler full wave rectifier using diodes;

FIG. **2**E is a schematic diagram of a prior art voltage doubler full wave rectifier using diode connected NMOS;

FIG. **2**F is a schematic diagram of a prior art voltage doubler full wave rectifier using diode connected PMOS;

FIG. **2**G is a schematic diagram of a prior art threshold-compensated transistor in a single-stage rectifier;

FIG. **3**A is a graph of output current with respect to input voltage for a prior art PMOS voltage doubler as shown in FIG. **2**F;

FIG. **3**B is a graph showing input and output voltage over time of the threshold-compensated transistor of FIG. **2**G;

FIG. **4**A is a schematic diagram of a prior art NMOS Dickson charge multiplier;

FIG. **4**B is a schematic diagram of a prior art PMOS Dickson charge multiplier;

FIG. **5** is a schematic diagram of a prior art forward compensated charge multiplier;

FIG. **6**A is a schematic diagram of a level-1 back compensated PMOS charge multiplier;

FIG. **6**B is a schematic diagram of a level-1 hybrid forward and back-compensated multiplier using NMOS and PMOS;

FIG. **6**C is a schematic diagram of a level-3 hybrid forward and back-compensated multiplier using NMOS and PMOS;

FIG. **7** is a contour plot of constant efficiency for input power versus level of compensation;

FIG. **8**A is a contour plot of constant efficiency for width versus number of stages for a level-1 compensated multiplier;

FIG. **8**B is a contour plot of constant efficiency for width versus number of stages for a level-3 compensated multiplier;

FIG. **9**A is a schematic diagram of a multiplier with an adaptive threshold-voltage compensation scheme using diode-connected transistors;

FIG. **9**B is a schematic diagram of a multiplier using p-type transistors with an adaptive threshold-voltage compensation scheme using solid-wired connections;

FIG. **9**C is a schematic diagram of a multiplier using n-type transistors with an adaptive threshold-voltage compensation scheme using solid wired connections;

FIG. **9**D is a schematic diagram of a multiplier using n-type transistors with an adaptive threshold compensation scheme using p-type and n-type transistors;

FIG. **10**A is a contour plot of constant efficiency for width versus number of stages for an adaptive diode-connected scheme as shown in FIG. **9**A at an operating frequency of 915 MHz;

FIG. **10**B is a contour plot of constant efficiency for width versus number of stages for an adaptive solid-wired scheme as shown in FIG. **9**B at an operating frequency of 915 MHz;

FIG. **11**A is a graph of leakage current versus input power with different schemes at 915 MHz;

FIG. **11**B is a graph of current transfer ratio versus input power with different schemes at 915 MHz;

FIG. **12**A is a graph of efficiency versus number of adaptive blocks for the adaptive diode-connected scheme;

FIG. **12**B is a graph of efficiency versus number of adaptive blocks for the adaptive solid-wired scheme; and

FIG. **13** is a schematic diagram showing an impedance matching circuit.

DETAILED DESCRIPTION

There is provided a hybrid forward and backward threshold voltage compensation scheme for RF-to-DC power conversion. In an embodiment, PMOS transistors are used as rectifying devices in all stages except for the first few stages eliminating the need for triple-well NMOS transistors. The compensating voltage is provided by connecting the gate terminal of the PMOS transistors to previous stages, also referred to as back-compensation, and also by connecting the gate terminal of the initial few NMOS transistors to later stages. Using PMOS transistors as a rectifying device in all the stages except for the first few stages eliminate the need for triple-well NMOS transistors. Also, all the MOS transistors in the main rectification chain are compensated in the proposed scheme. In addition to the hybrid scheme, an adaptive method suitable for rectifying devices operating at low power level is proposed to control the ON/OFF operation of the MOS transistors in the main rectification chain. An auxiliary block consisting of PMOS transistor is introduced to control the switching of the MOS transistors in the main chain. The PMOS transistors in the main chain are back-compensated in the conduction phase and forward-compensated when not conducting; increasing the gate-source potential and reducing the reverse leakage. The controlling voltage of the transistors in the auxiliary chain is derived from the local node of the main chain. Moreover, a procedure for designing matching network to provide passive voltage amplification and maximum power transfer is proposed. The dead-zone of the power converter unit is obtained using Y parameter analysis and the matching circuit is designed at the point of maximum power transfer. The RF-to-DC power conversion circuit may be fabricated using conventional methods such as IBM's 0.13 μm CMOS technology.

Power Harvester Circuit

FIG. **1** shows a block diagram of the power harvester unit of an RF energy harvesting system **20**. In the RF energy harvesting system shown, the signal is received by an antenna **22**. An impedance matching circuit **24** comprising a high quality factor (Q) resonator is used to ensure that maximum power is transferred from the receiver to the rectifier block **26**. However, the matching network itself introduces some energy losses because of the limited quality factors of on-chip passive inductors and capacitors required to match the input impedance of the rectifier to the output impedance of the antenna. The high Q resonator can also provide passive voltage amplification of the input signal. Depending on the embodiment, the high Q resonator may have a fixed frequency or may be tuneable. The next block is the RF-to-DC power conversion circuit **26** that converts the RF signal to DC voltage. The rectifier preferably operates at low input power levels while providing high power conversion efficiency. Several rectifier units may be cascaded to increase the overall output voltage, referred collectively as a voltage multiplier. There may be a power management block (not shown) to control the flow of power from the energy harvester to the load. The harvested power is then finally used to charge an energy storage device **28** such as a battery or a capacitor, which in turn powers a load **30**. Alternatively, the energy storage device may be omitted and the load powered directly by the power harvester. The power management block may also be omitted with or without the energy storage.

Rectifier Circuit

In half-wave rectifiers, the rectifying device is conducting for half of the cycle, passing the input voltage to the output. However, the output voltage is lower than the input by the amount of voltage drop needed for turning ON the rectifying device. A half-wave rectifier can be implemented using diodes as shown in FIG. **2**A. The voltage drop across the device depends on the threshold voltage of the diode. FIGS. **2**B and **2**C show half-wave rectification circuits implemented using diode-connected NMOS and PMOS transistors respectively in CMOS technology. The voltage drop in this case depends on the threshold voltage of the transistor. Both diode-connected PMOS and NMOS transistors can be used as rectifying devices. FIG. **2**D shows the voltage doubler unit which is a cascade of two half-wave rectifiers. The voltage doubler rectifies the AC input in both the positive and negative cycles. Similar to the half wave rectification, the diodes are implemented by connecting the drain and the gate terminal of the MOS transistor together such that the transistor is always in saturation region in forward bias condition, as shown in FIGS. **2**E and **2**F.

The voltage doubler may be used as a basic unit in a power conversion circuit. FIG. **2**F shows an example voltage doubler. In the voltage doubler, each of the transistors conducts only during one half of the input cycle. M_{P1} conducts during the negative phase while M_{P2} conducts during the positive phase of the input cycle. Each of the transistors goes through the sub-threshold region, inverse region and the leakage region. The maximum voltage at the output of voltage doubler is 2Vamp−|VTP1|−|VTP2|, where V_{amp} is the maximum amplitude of the input signal and V_{TP1}, V_{TP2} are the threshold voltages of the diode-connected transistors M_{P1} and M_{P2} respectively. Thus the maximum possible voltage is twice the RF signal's amplitude only when the threshold voltage of the transistors is equal to zero. To understand the operation of the voltage doubler in detail, we can look at the transient analysis of the doubler circuit of FIG. **2**F, as shown in FIG. **3**A. There are three regions of operation of the circuit as seen from FIG. **3**A. It is described as follows: the subthreshold operation extends from $V_{in}=0$ to $V_{in}=|V_{TP}|$, where V_{in} is the input voltage. The current in this region is an exponential function of the input voltage. The inversion region extends from $V_{in}=|V_{TP}|$ to $V_{in}=V_{amp}$. The current in the in this region is a square function of the input voltage. In the inversion region, the output current reaches its peak value when $V_{in}=V_{amp}$. Finally, the leakage region extends from $V_{in}=0$ to the next $V_{in}=0$ in the negative half-cycle. The current that flows through transistor M_{P2} during this time interval is referred as reverse leakage current. Thus, the actual charge transfer mechanism is only for a short duration while other parameters such as sub-threshold and reverse leakage currents have to be considered for the rest of the cycle.

Several voltage doublers can be cascaded to increase the overall output voltage. These voltage multipliers are also known as "charge pumps" and can generate a voltage several times larger than their input supply voltage. The Dickson multiplier with diode-connected transistors is commonly used for integrated applications. The Dickson multiplier of FIG. **4**A is modified for low power energy harvesting applications by grounding the ϕ_2 (out of phase) clock terminal and applying the input signal at the ϕ_1 terminal. FIG. **4**B shows the conventional Dickson multiplier with the PMOS voltage doubler as a rectifier unit. The body terminal of the transistors is connected as shown in FIG. **4**B to further reduce the threshold voltage while the transistors are conducting. For example, in the positive phase of the input cycle, transistors M_2, M_{n-1} and M_{n+1} are conducting. The source of the transistors is at a higher potential compared to the body terminal. This reduces the threshold voltage in the conducting region. At the same time, transistors M_1, M_3 and M_n are in the reverse region (leakage operation). The gate-source and the body-source potential is zero for these transistors. Thus, the threshold voltage of transistors M_2,

M_{n-1} and M_{n+1} while they are conducting is lower than the threshold voltage of the transistors M_1, M_3 and M_n which are in the reverse region in the positive input phase. Similarly, in the negative phase of the input cycle, transistors M_1, M_3 and M_n which are conducting have a lower threshold voltage than the transistors M_2, M_{n-1} and M_{n+1}.

To increase the output voltage of multi-stage rectifiers, the number of rectifier stages must be increased accordingly. However, as the number of stages increases the power conversion efficiency is reduced as larger number of the transistors dissipates more power. Also, the threshold voltage of the standard MOS transistor which lies around 450 mV in 130 nm technology, reduces the sensitivity of the power harvester block. As the number of stages in the rectifier increases, the power losses increases, reducing the overall efficiency. Too few of stages, leads to low output DC voltage even if the PCE is high. To achieve a high DC voltage, large transistors have to be used, leading to high leakage and parasitic losses whereas smaller transistor size affects the charge transfer leading to low DC voltage. The strategy while designing the multi-stage rectifier circuit of FIG. **4**B is as follows: The individual stages of the voltage doubler can be cascaded to increase the output voltage. As the number of stages increases, the output voltage increases. With the increase in the number of stages, the PCE decreases and the output voltage saturates. Hence the transistors as well as the pumping capacitors are scaled while increasing the stages. The scaling is done to maintain the incremental voltage per stage and the PCE relatively constant with the increase in the number of stages. Thus the PCE can be maintained with proper device scaling while increasing the DC voltage for a constant output power. The scaling of width and the pumping capacitance with the stages results in impedance looking into the rectifier to be unchanged so that the input power and thus the PCE remains constant. Thus, when designing a multistage rectifier unit one can optimize of number of stages and width of the transistors to maximize the PCE for a given output voltage.

The standard Dickson multiplier can be modified for designs involving energy harvesters where the input voltage is low. FIG. **4**A shows an N-stage conventional Dickson multiplier. A threshold self-compensation technique was described by Dickson. The gate of the transistor in this technique is connected to the adjacent source of the transistor instead of the traditional diode-connected structure. Thus, providing bias voltage equivalent to the incremental voltage across each stage. Based on this technique, a forward compensated topology was implemented by Papotto et al. where the bias voltage was increased by extending the gate length connection. FIG. **5** shows the basic forward compensated NMOS topology.

Threshold voltage reduction techniques requiring additional circuits are not suitable for integrated low power energy harvesting applications as these circuits occupy large areas and cause additional power dissipation. Passive threshold voltage reduction technique such as the self-compensation method of FIG. **5** does not require additional circuit and can reduce the threshold voltage. In the self-compensation method, the threshold voltage of the diode-connected NMOS transistors increases with the stages due to the body effect. The body effect can be dynamically controlled using additional circuits but generates undesirable substrate current. Another alternative is using triple-well process to individually bias the body terminal of NMOS transistors and reduce the threshold voltage variation. Additional parasitic capacitance is introduced due to the well structure increasing the losses at each node and affecting the overall efficiency.

Hybrid Forward and Back-threshold compensated power converter

In order to eliminate the need for triple-well processes, we propose a threshold voltage compensation technique scheme as described below. PMOS transistors are used instead of NMOS transistors as rectifying devices. As each PMOS transistor has its own n-well, the body of PMOS transistor can be biased individually without the necessity of a triple-well CMOS process. Unlike an NMOS transistor which needs a higher potential at the gate terminal to offset the threshold voltage, a PMOS transistor requires negative gate-source potential. Therefore, the threshold voltage of PMOS transistors can be reduced by connecting the gate potential to the previous node rather than later node. As shown in FIG. **6**A, alternating input voltage V_{in} is applied on first input line **100** to odd-numbered capacitors C_1, C_3, . . . C_n. Ground voltage is applied on second input line **102** to even numbered capacitors C_2, . . . C_{n-1}. Capacitor C_{n+1}, while also even-numbered, is shown as connected to ground by a separate line. A third line **104** accumulates voltage for the output and is connected to ground at one end and to output at the other. In this and other figures, the end of the line having more negative voltage, which in the embodiments shown is ground, is shown at the left of line **104** and the end of the line having more positive voltage, which in the embodiments shown is the output, is shown at the right of line **104**. The capacitors are each connected to line **104**, and are numbered in the order in which they are connected to line **104** from left to right. Each capacitor maintains a respective voltage differential between the respective input line it is connected to and the part of line **104** it is connected to. This voltage differential varies over time as current flows through line **104** between capacitors. Line **104** comprises transistors M_1, M_2, M_3, . . . M_{n-1}, M_n, M_{n+1} arranged in series, a transistor being arranged between each pair of adjacent capacitors and between capacitor C_1 and ground. Adjacent transistors define junctions **106**, the capacitors being connected to the junctions. The body terminal of each transistor is connected to the drain terminal of the same transistor, as shown by dotted lines. The gate of the first transistor (for n-level compensation, the first 'n' transistors) and the last transistor are connected to their respective drains. The gates of the other transistors are connected to the drain of the transistor n+1 transistors to the left of the respective transistor. The proposed scheme as shown in FIG. **6**A reduces the threshold voltages of all PMOS transistors except the first one leading to an increased output voltage. For an n-level compensation, for the proposed scheme, there will be 'n' initial PMOS transistors that will be uncompensated as seen from FIG. **6**A To solve this problem, 'n' uncompensated PMOS transistors are replaced by NMOS transistors with grounded body terminals as seen from FIG. **6**B. As shown in FIG. **6**B, the first transistor is an NMOS transistor with a source connected to the source of the PMOS transistor to its right to define a junction and has a gate connected to the junction n stages to the right of the junction defined by its source. That is, the gate is connected to the junction defined by the source of the PMOS transistor **2** (i.e, n+1) transistors to the right. The other transistors are PMOS and, apart from the last, each has a gate connected to a junction n stages to the left of the junction defined by its source. For the purposes of the preceding statement and statements about junctions elsewhere in this document, the grounded end of line **104** can be considered to be a junction **106**A one stage from the junction defined by the connection of M_1 and M_2. Thus the compensating voltage is provided by connecting the gate terminal to later stages for the first few

NMOS transistors and connecting the gate terminal to previous stages for PMOS transistors. The scheme shown in FIG. **6**A and FIG. **6**B is level-1 compensation. The compensation level can be increased by connecting the gate terminal of PMOS to the source terminals of the transistor of the following stage rather than the source of its immediate neighbor. The last transistor of the multiplier is left uncompensated to reduce the leakage. FIG. **6**C shows the proposed level-3 hybrid forward- and back-compensated multiplier. As shown in FIG. **6**C, the first **3** (i.e., n) transistors are NMOS transistors with grounded body terminals. Each NMOS (more generally, n-type) transistor has its source on the left and drain on the right, and each PMOS (more generally, p-type) transistor has its drain on the left and source on the right, where “left” and “right” are defined so that that the end of line **104** with more positive voltage is the “right”, as shown in the figures. “Earlier” and “later” can also be used to mean “left” and “right”. While ground is shown in the figures to the left and the output to the right, the ground could also be to the right and the output to the left. See below for a discussion of the definitions of “source” and “drain”. The gates of these first **3** NMOS transistors are connected to the respective junctions **106** three (that is, ‘n’) stages to the right of the junctions defined by their respective sources. The remaining transistors except the last are PMOS transistors with gates connected to the junctions **106** three (that is, ‘n’) stages to the left of the junctions defined by their respective sources. Increasing the level of compensation leads to reduction in the threshold voltage which improves the forward conduction but also leads to increased reverse leakage current degrading the rectifier’s PCE. Only odd level compensation is used as it maximizes the source-gate potential of PMOS transistors due to the alternating voltage phase with successive nodes.

To find the optimum number of stages and level of compensation, extensive simulations were conducted. FIG. **7** shows the efficiency contour plots with different level of compensation and input power level. Maximum PCE at the lowest input power level is obtained when the level of compensation is one. As the level of compensation increases, the reverse current increases which causes additional power loss and degradation in efficiency. Hence level-1 compensation gives the maximum efficiency while level-3 or higher is advantageous in reducing the threshold voltage of the RF-DC power conversion circuit. FIG. **8**A shows the constant efficiency contour plot as a function of transistor width and the number of stages for level-1 compensation. Eight-stage of doubler design equivalent to 16-stage rectifier with transistor width of 13-μm and fifteen-stage of doubler design equivalent to 30-stage rectifier with transistor width of 28 μm gives the highest efficiency contour. The former one is preferred due to lesser area on the chip. The level-1 and level-3 compensated multi-stage rectifier are referred to as efficiency circuit and voltage circuit respectively Similarly for the voltage circuit constant efficiency contour plot as a function of width and the number of stages is plotted as shown in FIG. **8**B. Twelve-stage of doubler design equivalent to 24-stage rectifier with transistor width of 8 μm is selected based on the plot. For level-3 compensation, more body-grounded NMOS transistors are required compared with the level-1. Also, the reverse leakage loss is higher for level-3 compensation due to the larger compensating voltage compared with the level-1 compensation. Increasing the level of compensation lowers the minimum input voltage requirement whereas increasing the number of stages while lowering the width of the transistors was based on the design strategy discussed earlier.

During the positive input phase, transistors M_2, M_{n-1} and M_{n+1} in FIGS. **6**A-**6**C are conducting. Since the gate terminals are connected to the previous node V_{SG} in the positive phase (conduction) is relatively high compared to the conventional Dickson multiplier. However, in the negative phase when the transistors M_2, M_{n-1} and M_{n+1} should ideally be not conducting, the V_{SG} is higher compared to the conventional case. Thus the hybrid forward and back-threshold compensated scheme is effective in increasing the forward conduction current but increases the reverse leakage current at the same time. As seen in FIG. **7**, with the increase in the level of compensation the PCE degrades. The simulated maximum PCE of 28% degrades to 11% from level-1 compensation to level-3 compensation respectively. To ameliorate this issue an adaptive scheme is proposed to control the reverse leakage current.

Especially at extremely low input power (e.g. micro-watts power level), the reverse leakage current has a significant adverse effect on the PCE and the output DC voltage of an RF rectifier as proven by the following analysis. Considering a threshold-compensated transistor of a single-stage rectifier shown in FIG. **2**G driven by an input source $V_{in}=V_a \sin \omega t$ assuming the compensation voltage is modeled by V_C and its input and the output steady-state voltage waveforms shown in FIG. **3**B, the overall PCE is defined as

$$PCE = \frac{P_{out,forward} - P_{leakage}}{P_{input}} \tag{1}$$

where $P_{out,forward}$ is the output power delivered to the load when transistor is forward-biased, $P_{leakage}$ is the output power lost due to leakage when transistor is reverse-biased, and P_{input} is the input power. The forward region extends from $t=t_1$ to $t=\pi-t_1$ with the current $I_d(t)$ through the MOS transistor M_1 as seen in FIG. **3**B.

$$\frac{P_{out,forward}}{P_{input}} = \frac{\frac{1}{\pi-2t_1}\int_{t_1}^{\pi-t_1} V_{out}\cdot I_d\cdot dt}{\frac{1}{\pi-2t_1}\int_{t_1}^{\pi-t_1} V_{in}\cdot I_d\cdot dt} = \frac{\frac{1}{\pi-2t_1}\int_{t_1}^{\pi-t_1} (V_a - V_{TH} + V_c)\cdot dt}{\frac{1}{\pi-2t_1}\int_{t_1}^{\pi-t_1} V_a \sin\omega t\cdot dt} \tag{2}$$

Assuming the ripple voltage variation ΔV is much smaller than the average output voltage V_o, the output voltage for the one-stage rectifier can be expressed as $V_o=V_a-V_{TH}+V_C$ where V_{TH} is the threshold voltage of the transistor. Performing integration on (2) gives the ratio of the output power in the forward region to the input power as

$$\frac{P_{out,forward}}{P_{input}} = \frac{\omega\cdot(\pi-2t_1)\cdot(V_a - V_{TH} + V_c)}{2V_a\cos\omega t_1} \tag{3}$$

The value of t_1 lies between $0<t_1<\pi/2$ based on the value of $V_{TH}-V_C$. The time t_1 indicates the onset of inversion region and is close to zero when the compensating voltage is near the threshold voltage and will be closer to $\pi/2$ when the compensating voltage is near-zero value. As seen in (3), the ratio of the output power in the forward region to the input power increases with increased voltage compensation in the forward region. In the reverse-biased region, the leakage current is expressed as

$$I_{leak}(t)=I_o\cdot(W/L)\cdot e^{(V_{gs}-V_{TH})/\eta V_T}\cdot(1-e^{-V_{ds}/\eta V_T}) \tag{4}$$

Replacing the gate-source bias voltage by V_C, source-drain bias by $V_{in}(t)-V_{out}$ and V_{in} by $V_a \sin \omega t$, the leakage current as a function of time can be expressed as

$$I_{leak}(t)=I_o\cdot(W/L)\cdot e^{(V_c-V_{TH})/\eta V_T}\cdot(1-e^{(V_a \sin\,\omega t-V_a+V_{TH}-V_c)/\eta V_T}) \tag{5}$$

where $(\pi-t_1)<t<(2\pi+t_1)$. With increase in the compensation, the power loss increases due to the higher leakage current. As seen in (3) and (5), the ratio of the output power in the forward region to the input power even though increases with larger threshold-voltage compensation, the losses in the leakage region is higher due to the increased compensation. This indicates the fundamental trade-off between the level of threshold-voltage compensation and the leakage current of the transistors.

In a field effect transistor, the "source" and "drain" terminals are defined by the direction of flow of charge carriers when the transistor is "on": charge carriers flow from the source to the drain. For an n-type transistor, the charge carriers are electrons and for a p-type transistor the charge carriers are holes. As electrons flow from negative to positive, the source of an n-type transistor is at a lower voltage than the drain and vice versa for a p-type transistor. FETs do not necessarily have any asymmetry between the source and drain terminals beyond this. In the present system, the relative voltages of the source and drain terminals reverse in the normal course of events, as the system is exposed to alternating current in order to rectify it. There are two possible approaches to defining which terminal is the "source" and which terminal is the "drain" in this circumstance. One possible approach is to say that which terminal is the "drain" and which terminal is the "source" varies over time depending on which terminal is at the higher voltage. Thus, by this first definition, the source of an n-type transistor is whichever of the source and drain terminals has the lower voltage at any given moment, and for a p-type transistor whichever has the higher voltage at a given momentA second approach, however, is possible where the desired conduction is in a particular direction only. In the present system, conduction in one direction is desirable and conduction in the other ("leakage") is undesirable, thus the source can be defined according to which terminal is the source when current flows through the transistor in the desired direction; which terminal is considered the "source" and which terminal is considered the "drain" does not change over time in this approach despite the change in voltages. In the discussion here concerning the source-gate voltage (V_{SG}), the first approach is used, i.e. the source-gate voltage is the voltage between the gate and whichever of the source or drain terminals that has the higher or lower voltage at a given time depending on if the transistor is p-type or n-type. In addition, while the sign of a source-gate voltage difference required to operate a transistor depends on whether the transistor is p-type or n-type, V_{SG} is discussed as if the voltage to operate the transistor is always positive for p-type transistor and negative for n-type transistor. For other purposes in this document including the claims, a version of the second approach to defining which terminal is the source and which is the drain is used in which the direction of desired current is based on the desired current direction through the device as a whole. Thus, throughout the figures, p-type transistors are shown as having their sources to the left and n-type transistors are shown as having their sources to the right.

A diode-connected transistor is a transistor in which one of the source and drain terminals is connected to the gate to cause the transistor to act as a diode. A diode connected FET allows charge carriers to flow from the terminal not connected to the gate to the terminal connected to the gate, if sufficient voltage is applied, and blocks charge carriers from flowing in the opposite direction. The diode connected transistor is diode connected because it is desired that current flow in one direction is allowed and in flow in the other direction blocked. Thus, following a version of the second approach above to defining which terminal is the source and which terminal is the drain based on desired current through the individual transistor, the source or drain terminal of a diode connected transistor that is connected to the gate would be by definition the drain of the diode connected transistor. However, in this document it is the desired current through the device as a whole that is used, thus in e.g. FIG. **9**A transistor M_{na} is diode-connected and the terminal connected to the gate is depicted as the source.

To investigate the effect of the level of threshold-compensation on the PCE of multi-stage hybrid threshold-compensated rectifiers, level-1 rectifier which is connecting the gate terminal to one previous node for the PMOS transistors (FIG. **6**A) and level-3 rectifier which is connecting the gate terminal to three previous nodes (FIG. **6**C) for the PMOS transistors are simulated to obtain constant maximum efficiency contour plots as shown in FIG. **7**. Increasing the level of compensation leads to reduction in the threshold voltage which improves the forward-conduction but also increases the reverse leakage current. In the multipliers as shown in FIGS. **6**A and **6**C, during the positive input phase, transistor M_n is forward-biased. Since the gate terminals are connected to the previous node, V_{sg} when forward-biased is higher compared to that of the conventional diode-connected case with no threshold-voltage compensation which increases the ratio of output power in the forward region to the input power. However, when the transistor M_n is reverse-biased, the V_{sg} is higher compared to the conventional diode-connected case increasing the reverse leakage current of the reverse-biased transistor. This can also be explained by the efficiency contour plots of the hybrid scheme as seen from FIG. **7**. The contour plots are simulated for constant efficiency with different level of compensation and input power levels. As seen from FIG. **7**, though with the increase in the level of compensation the threshold voltage reduces, the efficiency degrades at the same time which is due to the increased reverse leakage current. The simulated maximum PCE of 28% degrades to 11% from level-1 compensation to level-3 compensation respectively. Accordingly an adaptive scheme is proposed to control the threshold voltage and the reverse leakage current of the rectifying device dynamically to improve the PCE over a wide range of input power. Ideally, the threshold voltage of the rectifying device should be zero when the transistors are forward-biased while the threshold voltage should be high when the transistors are reverse-biased to prevent any leakage or the losses associated with it.

We propose an adaptive forward and backward threshold-voltage compensation scheme that use minimal additional circuitry to increase the threshold-voltage compensation when transistors are forward-biased and decrease the compensation voltage when they are reverse-biased. The PMOS transistors in the main rectification chain are back-compensated when forward-biased and forward-compensated when reverse-biased; increasing the forward-current and reducing the reverse leakage current dynamically.

Proposed Adaptive Threshold Voltage Compensation

The hybrid forward and back-compensated topology described above can be made adaptive using an auxiliary block controlling the gate-source voltage of the MOS transistors in the main rectification chain. The auxiliary structure may be realized using all PMOS transistors to allow for individual body biasing. Preferably the auxiliary blocks are designed using minimum number of PMOS transistors so that the power losses do not increase considerably due to the additional blocks. The controlling voltage of the transistors in the auxiliary chain is derived from the local node of the main rectification chain. Two possible implementation of the proposed adaptive threshold voltage compensation scheme are discussed in the following section, one using diode-connected PMOS transistors and the other solid-wired connection to adaptively adjust the level of the threshold-voltage compensation.

Using Diode-connected Transistor for Back-compensation

FIG. **9**A shows the hybrid forward and back-threshold compensated rectifier using diode-connected transistors for back-compensation. The back-compensation reduces the threshold voltage when the transistors are forward-biased and forward-compensation reduces the reverse leakage current when reverse-biased with the control signal derived from the local node. In the embodiment shown the last transistor M_{n+1} is left uncompensated to reduce the leakage. During the negative input phase, in which the transistor M_n is forward-biased, the transistor M_n is back-compensated by the diode-connected transistor M_{na} while the V_{SG} terminal voltage for the transistor M_{nb} lies below its threshold voltage resulting in transistor M_{nb} turned OFF. In comparison to the embodiment shown in FIG. **6**B, diode-connected transistor M_{na} lies on the line connecting to the gate of transistor M_n. Transistor M_{nb} lies on a different line not corresponding to a line in FIG. **6**B, providing a forward connection for the gate of transistor M_n when transistor M_{nb} is active. In the embodiment shown in FIG. **9**A p-type transistors are used, but n-type could also be used. Additional transistors of the opposite type may be used at one end as in FIG. **6**B. The back-compensation for the transistor M_n in the main chain enhances the forward current. During the positive input phase, in which the transistor M_n is reverse-biased, the V_{SG} terminal for transistor M_{nb} is enough to turn ON the forward connection thus reducing the V_{SG} bias of the transistor M_n to zero resulting in a reduced leakage current. FIG. **10**A shows the efficiency contour plot for the adaptive diode-connected scheme as a function of width and the number of stages of the rectifier. Generating contour plots are an effective way to optimize the number of stages and the width of the transistors to maximize the PCE while obtaining the required voltage. Twelve-stage of voltage doubler design equivalent to 24-stage rectifier with transistor width of 11 μm gives the highest efficiency contour. The width of the diode-connected transistor in the auxiliary chain should be comparable to the width of the transistors in the main rectification chain so that it provides low forward-resistance when the transistors are conducting. In an embodiment, the width of the diode connected transistor in the auxiliary chain may be 8 μm. The auxiliary transistor used for forward compensation to control the reverse-leakage is selected in an embodiment to be 480 nm, an order of magnitude smaller than the diode-connected transistor to minimize their power consumption and reduce the parasitic at the nodes of the main rectification chain. Larger transistor widths are avoided to reduce the parasitic at the nodes of the main rectification chain.

Also, for the diode-connected scheme, the auxiliary block is added for every alternate transistor starting from the later stages and adding the blocks towards the initial stages. The maximum PCE increases with the addition of the adaptive auxiliary blocks.

Using Solid Connection for Back-compensation

FIG. **9**B shows the adaptive hybrid topology using solid wired connection instead of the diode-connected transistor for reducing the threshold voltage. When the transistors are in the conduction phase (forward biased) they are back-compensated with solid wired connection instead of the diode-connected transistors which prevents the forward-losses associated with it. In the embodiment of FIG. **9**B, as compared to FIG. **9**A, the transistors M_{na}, $M_{n-1)a}$ etc. are omitted and replaced by a solid wire connection as in FIG. **6**B, but the transistors M_{nb}, $M_{(n-1)b}$ etc. are retained. In the embodiment shown in FIG. **9**A p-type transistors are used, but n-type could also be used. Additional transistors of the opposite type may be used at one end as in FIG. **6**B. When the transistor M_n is reverse-biased (not in the conduction phase), the V_{sg} terminal voltage for the auxiliary transistor M_{nb} is large to turn ON the forward-connection and reduce the source-gate bias of transistor M_n. This decreases the reverse leakage current greatly. FIG. **10**B shows an efficiency contour plot for the solid-wired scheme as a function of width and the number of stages of the rectifier. Twelve stages of doubler design equivalent to 24-stage rectifier with transistor width of 10 μm gives the maximum efficiency contour. The efficiency contour plot for the solid-connection follows similar trend as the diode-connected one. With the increase in the width of the transistors while maintaining the number of stages, the efficiency initially increases and then degrades due to the increased parasitic losses. Similarly, when the width of the transistor is kept constant and the number of stages is increased, the efficiency initially increases and then degrades as more number of stages is added due to the increased power loss with additional stages. Coupling capacitor value of 4 pF is selected as it has a very modest impact on the rectifier's efficiency.

Further embodiments are shown in FIG. **9**C and FIG. **9**D. The embodiment of FIG. **9**C operates in the same way as the embodiment of FIG. **9**B but uses n-type transistors instead of p-type. Accordingly, the solid wired connections to the gates of the transistors in the main chain are to the right instead of to the left. FIG. **9**D is an embodiment in which the connections to the gates of the transistors of the main chain are each controlled by two transistors. However, unlike in FIG. **9**A, one of these two transistors is p-type and the other is n-type. This allows the gates of these two transistors to be controlled by the same voltage, here using a line connecting to the junction to the right of the corresponding main transistor, and have at most one of the transistors ON at any given point in the input cycle. In addition, the transistors in the main chain are n-type in FIG. **9**D, whereas the transistors of the main chain are p-type in FIG. **9**A. However, either of the embodiments of FIG. **9**A or FIG. **9**D could use n-type or p-type transistors in the main chain, with corresponding changes to the direction of connections as seen in the differences between FIG. **9**B and FIG. **9**C. FIGS. **9**C and **9**D don't show the additional transistor at the end of the rectifier that is shown as M_{n+1} in FIGS. **9**A and **9**B, but such an additional transistor may also be present in these figures. The additional transistor can be p-type or n-type.

To investigate the effectiveness of the adaptive blocks in reducing the leakage current, leakage current as a function of input power is plotted as shown in FIG. **11**A. The leakage current is simulated for the last transistor with a 1 MΩ load for different input power levels. As shown in FIG. **11**A, comparing the schemes i.e. hybrid, adaptive solid and adaptive diode-connected for power levels of 1 μW-100 μW and at an operating frequency of 915 MHz, the reverse leakage current is maximum for the hybrid scheme and increases with the input power. At an input power of 105 μW, the leakage current is 12 nA. The adaptive solid and the adaptive diode-connected scheme have an auxiliary transistor to control the reverse leakage current. At an input power of 1 μW, the auxiliary transistor does not have enough input power to turn ON and provide the required forward-compensation. Hence the leakage current is comparable with the hybrid scheme. At 1 μW (−30 dBm), the output transistor's leakage current is 6.96 nA for the adaptive solid scheme and 7.56 nA for the adaptive diode-connected scheme with a 1 MΩ load. As the input power increases, the leakage current for the adaptive scheme is drastically reduced as seen in FIG. **11**A. Also, the reverse leakage current for the adaptive schemes is relatively constant with increase in the input power. The leakage current also depends on the load resistance and increases with decrease in the load resistance. Another performance measuring parameter we have defined is the current transfer ratio which is the ratio of the forward-current to the reverse leakage current. FIG. **11**B shows the current transfer ratio as a function of input power. The current transfer ratio at an input power of 1 μW (−30 dBm) for the hybrid scheme is 270, for the adaptive solid scheme is 140 and for the adaptive diode-connected one is 108. The current transfer ratio for the adaptive solid scheme increases rapidly and outperforms the hybrid scheme from 2 μW (−27 dBm) while for the adaptive diode-connected one, the current transfer ratio intersects the hybrid curve at 11 μW. The current transfer ratio for the hybrid scheme initially increases with the input power and saturates to approximately 1500 at 100 μW of input power. The current transfer ratio for the adaptive solid scheme rapidly increases with the input power. At an input power of 90 μW, the current transfer ratio for the adaptive solid scheme is 9000 as shown in FIG. **11**B. The current transfer ratio for the adaptive diode-connected scheme is 4000 at 97 μW of input power. The current transfer ratio for the adaptive-solid scheme at an input power of 100 μW (−10 dBm) is 6 times while for the diode-connected one is 2.5 times better than the hybrid scheme.

The proposed adaptive scheme is effective in increasing the PCE with the addition of auxiliary adaptive blocks. The increase in the PCE with the addition of the adaptive auxiliary blocks for the adaptive diode-connected scheme is shown in FIG. **12**A. With the addition of the auxiliary block the maximum PCE increases to 12%. The auxiliary block is added for every alternate transistor. However, the additional power consumed by the auxiliary diode-connected transistor when the rectifying device in the main rectification chain is forward-biased prevents the scheme from achieving higher PCEs. Unlike the adaptive diode-connected scheme, the auxiliary adaptive block is added to all the transistors for the adaptive solid scheme. Except for the first and the last stage, all the transistors for the solid adaptive scheme have the adaptive blocks. The additional power consumption and the parasitic capacitance introduced by the auxiliary chain are much lower than the former circuit as the solid-connected scheme uses only one transistor of width 480 nm per adaptive block. Hence the PCE is relatively higher at extremely low power levels. As seen from FIG. **12**B, with the addition of adaptive blocks, the efficiency increases. FIG. **12**B shows the improvement in PCE of the rectifier with the addition of auxiliary block. The maximum PCE reaches to 33.5% when the number of adaptive blocks is 20, a figure much higher than the adaptive diode-connected scheme which has a maximum PCE of 12%. The PCE and the output voltage performance at a relatively higher power level 0 dBm to −20 dBm is much better compared with only the hybrid scheme.

Impedance Matching Technique

The overall PCE of the RF energy harvesting system can be improved by increasing the passive voltage amplification due to the matching network, which increases the overall voltage at the input of the rectifier. The rectifier is modeled as $R_{rec}-jX_{rec}$ where R_{rec} is the real part of the rectifier's impedance and X_{rec} is the imaginary part. An impedance matching circuit is designed to match it to 50 Ω as shown in FIG. **13**.

The input capacitance is mainly associated with the parasitic capacitance of the transistors and the layout. As the number of stages increases, the input resistance decreases while the parasitic capacitance increases. Increase in the value of load current also decreases the input resistance. The major factor contributing to increase in load current is lower load resistance. Other factors affecting the input impedance is the aspect ratio (W/L) of the transistors, input power level. Increase in the aspect ratio, increases the parasitic capacitance whereas decreases the input resistance. At the output side, the rectifier can be modeled by a voltage source and an output resistance. Thus the non-linear characteristic of the rectifier circuit makes it difficult to analyze them. The input impedance of the rectifier can be assumed to be constant only when small signal input is applied. In this case, the transient signal is a large signal with a capacitive component at the input causing a phase difference component ϕ to appear. The derivation in appendix though considers the reactive impedance still assumes the linear VI relationship. In real case, the VI relationship is non-linear and the input resistance is to be found using CAD tools.

Experimental Results—Hybrid Scheme

Three rectifiers, named as "efficiency", "voltage" and "1-stage PMOS doubler" are designed and fabricated in a 0.13 μm 8-metal CMOS process. The active die areas for efficiency circuit, voltage circuit and the 1-stage PMOS doubler test circuit are 230 μm×810 μm, 230 μm×1050 μm and 160 μm×70 μm respectively. The chip is wire-bonded onto PCB board with FR4 substrate and tested with Agilent MXG-N5181 signal generator at a frequency of 915 MHz using a single-tone sinusoidal signal. The receiver power is calculated by finding the average power at the input of the rectifier. The performance of the designed efficiency and voltage circuit is measured for a range of input power levels.

Output DC voltage was measured for different peak-to-peak input voltages. For a 1 MΩ load, an input voltage of 170 mV results in 2.4 V and 2.8 V for efficiency and voltage circuit, respectively. The voltage multiplication ratio (VMR) which is the ratio of DC voltage to the peak-to-peak input voltage is 14 and 17 for efficiency and voltage circuit respectively. A 220 mV signal results in 3.1 V for efficiency circuit and 4.0 V for voltage circuit. Similarly for a 5 MΩ load, an output of 2.7 V (VMR=16) for efficiency circuit and 3.0 V (VMR=18) for voltage circuit was measured at an input peak-to-peak voltage of 170 mV. Thus, voltage circuit which has a higher level of compensation than efficiency circuit has a lower input voltage requirement.

Harvested power was measured for different received powers. From the results, it was observed that at low power-level, the harvested power has a higher dependence on load current. For a 1 MΩ load at power levels greater than −30 dBm, the rate of decay in the harvested power curve is higher than the 5 MΩ load. The roll-off in the harvested power for a 1 MΩ load starts at about −20 dBm while the roll-off point for a 5 MΩ load is approximately −30 dBm. Even for voltage circuit, the rate of decay in the harvested power curve is greater for a 1 MΩ load compared to a 5 MΩ load. A lower load value has a higher current requirement. The performance of both the circuits in terms of roll-off is similar for the same load value. The designed efficiency and voltage circuit outperforms the circuit in Papotto et al. especially at low power levels. The efficiency circuit delivers an output power of 4.7 μW at an input power of −16.8 dBm (20.9 μW) when loaded by 1 MΩ. With a 5 MΩ load, the output power is 1 μW for an input power of −17.5 dBm (17.7 μW). An output power of 3.4 μW at an input power of −14.8 dBm (33.1 μW) for a 1 MΩ load is supplied by voltage circuit. A graph of the measurements was found to be in close agreement with the simulation results.

Power efficiency was measured for different received power levels. The measured and the simulated power conversion efficiency for efficiency and voltage circuit for different load resistance values are further compared. The efficiency comparison is done while de-embedding the input reflections in Papotto et al. The output DC voltage was also measured for the efficiency and voltage circuits for different load resistance values. The PCE is optimized for low power levels. When loaded by 1 MΩ, efficiency circuit attains a maximum measured PCE of 22.6% at −16.8 dBm (20.9 μW) while delivering 2.2 V to the output. A maximum measured PCE of 21.6% is obtained by efficiency circuit while producing an output voltage of 1.1 V at an input power of −26.5 dBm (2.23 μW) for a 5 MΩ load. At an input power level of −14.8 dBm (33.1 μW), voltage circuit achieves a maximum measured PCE of 10.2% for a 1 MΩ load while delivering 1.8 V. Due to a lower load current requirement for a 5 MΩ load, the output voltage is higher compared to a 1 MΩ load. At an input power of −22.5 dBm (5.6 μW), voltage circuit has a measured output DC voltage of 1 V while efficiency circuit has a measured output DC voltage of 1.8 V at −24 dBm (4 μW) for a 5 MΩ load.

Experimental Results—Adaptive Scheme

Three RF-DC power converters named as “adaptive solid,” “adaptive diode-connected” and the “hybrid” circuits were designed and fabricated side by side in a 0.13 μm metal CMOS process with eight layers of metallization. The active die areas for adaptive solid and adaptive diode-connected is 0.25 mm^2 and hybrid circuit is 0.15 mm^2. The chip is wire-bonded onto a 2-layer FR-4 PCB board and tested with Agilent MXG-N5181 signal generator using frequency modulated continuous signal in the 902-928 MHz industrial, scientific and medical (ISM) band. An off-chip L-section impedance matching network is implemented on the PCB to convert the RF-DC power converter's input impedance to 50Ω. The output DC voltage was obtained with an oscilloscope or a digital multimeter. The measured and the simulated PCE for the adaptive and the hybrid scheme were compared for a load resistance of 1 MΩ at different input power levels. The output DC voltage was measured for the adaptive and the hybrid scheme for a load resistance of 1 MΩ. The simulation is performed at a frequency of 915 MHz which is the center frequency for the 902-928 MHz ISM band. The adaptive RF-DC power converters are designed to provide high PCE and a large output DC voltage for input power levels of 1 μW-100 μW (−30 dBm to −10 dBm). At larger input power, even with lower PCE the available output power is large hence designing for high input power levels is not so crucial. The adaptive solid scheme attains a maximum PCE of 32% at an input power of 32 μW (−15 dBm) with an output DC voltage of 3.2 V for a 1 MΩ load. For the input power of 32 μW and 1 MΩ load, the hybrid scheme has a PCE of 18% and delivers 2.6 V to the output. The adaptive diode-connected scheme has a maximum PCE of 11.3% at an input power of 118 μW while delivering 3.7 V to the output. The additional power consumed by the auxiliary diode-connected transistor when the rectifying device in the main rectification chain is forward-biased prevents the adaptive diode-connected scheme from achieving higher PCEs. The PCE was measured for different input power for the adaptive solid scheme at different load resistances. The output DC voltage for the adaptive solid scheme was also measured as a function of input power for different load resistances. As the load resistance decreases, the peak conversion efficiency curve shifts towards similar efficiencies at higher input power. The maximum measured PCE is 33.4% for a load resistance of 500 kΩ load at an input power of 83 μW. The PCE for a 500 kΩ load is larger than a 1 MΩ load for input power levels greater than 50 μW (−13 dBm). The peak power conversion efficiency curve is a function of the load resistance and can provide peak efficiency at a much lower power levels for larger load resistances. With decrease in load resistance, the circuit provides a smaller output voltage than with the high load resistance due to the low load current requirement at high load resistances. The hybrid scheme provides a larger output voltage compared to the adaptive schemes for input power levels lesser than 15 μW. The adaptive scheme outperforms the hybrid scheme once the power-up threshold-requirement is met. A DC output voltage of 3.2 V is obtained at an input power of 64 μW (−12 dBm) for a 500 kΩ load.

The harvested power was measured as a function of input power for the adaptive and the hybrid scheme. When loaded by 1 MΩ, an output power of 10 μW is harvested with an input power of 250 μW (−6 dBm) for the hybrid scheme while the adaptive solid scheme harvests 10 μW with only 30 μW (−15.3 dBm) of input power. With a load resistance of 500 kΩ, 10 μW of output power is harvested at an input power of 42 μW (−13.8 dBm). An output power of 20 μW is harvested at an input power of 64 μW (−12 dBm) for a 500 kΩ load using adaptive solid scheme. For a 1 MΩ load, at larger input power, the PCE degrades much more rapidly compared to a 500 kΩ load. For a 1 MΩ load, 20 μW of power is harvested by the adaptive solid scheme at 164 μW (−7.9 dBm). Even with an input power of few milliwatts, the hybrid scheme is not effective in harvesting 20 μW of output power.

The adaptive scheme is highly effective when the input power is above 10 μW (−20 dBm) and a large output voltage (more than 1.5 V) is desired. The performance of the adaptive solid scheme is similar to the hybrid scheme for power levels of 1-10 μW (−30 dBm to −20 dBm). The sensitivity of the RF-DC power converter for obtaining an output voltage of 1 V with a 1 MΩ load for the adaptive solid scheme is 8.9 μW (−20.5 dBm). The sensitivity of the circuit using the hybrid scheme is 6.9 μW (−21.6 dBm). The adaptive diode-connected scheme gives similar performance as the adaptive solid scheme at input power levels greater than 100 μW (−10 dBm). For low power applications, the adaptive solid scheme should be preferred over the adaptive diode-connected scheme.

The forgoing description pertains to circuitry using MOSFET enhancement-mode transistors. Nothing in the description should be taken to exclude using other types of transistors with corresponding changes to the circuitry that would be obvious to a person skilled in the art.

Immaterial modifications may be made to the embodiments described here without departing from what is covered by the claims.

In the claims, the word "comprising" is used in its inclusive sense and does not exclude other elements being present. The indefinite articles "a" and "an" before a claim feature do not exclude more than one of the feature being present. Each one of the individual features described here may be used in one or more embodiments and is not, by virtue only of being described here, to be construed as essential to all embodiments as defined by the claims.

APPENDIX

In section II, the average input power was suggested consisting of real and imaginary parts.

$$V = V_{in}\sin\omega t,\ I = I_{in}(\sin\omega t - \phi)$$

$$P_{mean} = \frac{1}{T_2 - T_1}\int_{T_1}^{T_2} V_{in}\sin\omega t \cdot I_{in}(\sin\omega t - \phi)dt$$

$$Z_{in} = \frac{V_{in}^2}{2P_{mean}} \cdot \cos\phi + \frac{V_{in}^2}{4P_{mean}} \cdot \frac{\sin[2\omega(T_2 - T_1) - \phi]}{\omega(T_2 - T_1)}$$

Z_{in} consists of real and imaginary component. Equating the imaginary component to zero by considering the phase difference $\phi=0$.

$$R_{in} = \frac{V_{in}^2}{2P_{mean}},\ C_{in} = \frac{4P_{mean}(T_2 - T_1)}{V_{in}^2\sin[2\omega(T_2 - T_1) - \phi]}$$

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

1. A power conversion circuit comprising:

a first input line and a second input line, the first and second input lines configured to receive an alternating voltage differential between the first and second input lines;

a multi-stage rectifier comprising n transistors arranged in series, where n is at least 2, each transistor having a gate, a source and a drain, adjacent transistors of the series being connected so that for adjacent p-type transistors, one of them being left p-type transistor positioned at the left-hand side and the other being right p-type transistor positioned at the right-hand side, the drain of the left p-type transistor is connected to the source of the right p-type transistor and source of the right p-type transistor is connected to the drain of the left p-type transistor, for adjacent n-type transistors, one of them being left n-type transistor positioned at the left-hand side and the other being right n-type transistor positioned at the right-hand side, the source of the left n-type transistor is connected to the drain of the right n-type transistor and drain of the right n-type transistor is connected to the source of the left n-type transistor, for a p-type transistor adjacent to an n-type transistor, n-type transistor positioned at the left-hand side and the p-type transistor positioned at the right-hand side, the source of the p-type transistor is connected to the source of the n-type transistor and drain of the p-type transistor is connected to the drain of the n-type transistor to form a junction, each junction being connected to one of the first input line and the second input line via a capacitor, with adjacent junctions having one junction of the adjacent junctions connected to the first input line and the other junction of the adjacent junctions connected to the second input line;

the gate of a k^{th} transistor of the multi-stage rectifier, where k is a number between 1 and n, is connected to the source of the $(k-1)^{th}$, or the $(k-2)^{th}$, . . . , or the 1^{st} main transistor for the p-type transistor or to the source of the $(k+1)^{th}$, or the $(k+2)^{th}$, . . . , or the n^{th} main transistor for the n-type transistor, at least one of the transistor in the multi-stage rectifier being p-type and at least one being n-type.

2. The power conversion circuit of claim **1** further comprising an auxiliary chain of n p-type transistors, each auxiliary transistor having a gate, a source and a drain, each transistor of the auxiliary chain being connected to a p-type transistor of the multi-stage rectifier in the main chain, so that the gate of the k^{th} main auxiliary transistor, where k is a number between 1 and n, is connected to the source of the k^{th} main transistor, the source of the k^{th} auxiliary transistor is connected to the gate of the k^{th} main transistor and also connected to the source of the $(k-1)^{th}$, or the $(k-2)^{th}$, . . . , or the 1^{st} main transistor, and the drain of the k^{th} auxiliary transistor is connected to the drain of the k^{th} main transistor or the drain of the $(k+1)^{th}$, or the $(k+2)^{th}$, . . . , or the n^{th} main transistor.

3. The power conversion circuit of claim **2** further comprising a diode connected transistor on a line connecting the gate of the k^{th} main transistor to the source of the $(k-1)^{th}$, or the $(k-2)^{th}$, . . . , or the 1^{st} main transistor.

4. The power conversion circuit of claim **1** further comprising an auxiliary chain of n n-type transistors, each auxiliary transistor having a gate, a source and a drain, each transistor of the auxiliary chain being connected to an n-type transistor of the multi-stage rectifier in the main chain, so that the gate of the k^{th} auxiliary transistor, where k is a number between 1 and n, is connected to the source of the k^{th} main transistor, the source of the k^{th} auxiliary transistor is connected to the gate of the k^{th} main transistor and also connected to the source of the $(k+1)^{th}$, or the $(k+2)^{th}$, . . . , or the n^{th} main transistor, and the drain of the k^{th} auxiliary transistor is connected to the drain of the k^{th} main transistor or to the drain of the $(k-1)^{th}$, or the $(k-2)^{th}$, . . . , or the 1^{st} main transistor.

5. The power conversion circuit of claim **4** further comprising a diode connected transistor on a line connecting the gate of the k^{th} main transistor to the source of the $(k+1)^{th}$, or the $(k+2)^{th}$, . . . , or the n^{th} main transistor.

6. The power conversion circuit of claim **1** further comprising an additional transistor connected in series with the multistage rectifier, the additional transistor having a gate, a source and a drain, the gate of the additional transistor being connected to the drain of the additional transistor.

7. The power conversion circuit of claim **1** in which the gate of the k^{th} transistor is connected to the source of the $(k-1)^{th}$, or the $(k-2)^{th}$, . . . , or the 1^{st} transistor for p-type transistor and to the source of the $(k-1)^{th}$, or the $(k-2)^{th}$, . . . , or the 1^{st} main transistor for n-type transistor.

8. The power conversion circuit of claim **7** in which N of the transistors in the multistage rectifier arranged in series are n-type and all but N of the transistors in multistage rectifier arranged in series are p-type.

9. The power conversion circuit of claim **1** in which the second input line is grounded.

10. The power conversion circuit of claim **1** in which the body terminal of each p-type transistor is connected to the respective drain and the body terminal of each n-type transistor is either grounded or connected to the respective drain terminal.

11. A power conversion circuit comprising:

an auxiliary chain of n auxiliary transistors, where n is at least 2, each auxiliary transistor having a gate, a source and a drain; a main chain of n main transistors connected in series, each main transistor having a gate, a source and a drain; each transistor of the auxiliary being connected to a respective transistor of the main chain of multi-stage rectifier;

wherein, if the auxiliary and main transistors are n-type, the gate of a k^{th} auxiliary transistor, where k is a number between 1 and n, is connected to the source of the k^{th} main transistor, the source of the k^{th} auxiliary transistor is connected to the gate of the k^{th} main transistor, the source of the k^{th} auxiliary transistor is also connected to the source of the $(k+1)^{th}$, or the $(k+2)^{th}$, . . . , or the n^{th} main transistor, and the drain of the k^{th} auxiliary transistor is connected to the drain of the k^{th} main transistor;

and

wherein, if the auxiliary and main transistors are p-type, the gate of a k^{th} auxiliary transistor is connected to the source of the k^{th} main transistor, the source of k^{th} auxiliary transistor is connected to the gate of the k_{th} main transistor, the source of the k^{th} auxiliary transistor is also connected to the source of the $(k-1)^{th}$, or the $(k-2)^{th}$, . . . , or the 1^{st} main transistor, and the drain of the k^{th} auxiliary transistor is connected to the drain of the k^{th} main transistor.

12. The power conversion circuit of claim **11** in which the transistors of the main chain and auxiliary chain are n-type and further comprising a p-type transistor on a line connecting the gate of the k^{th} main transistor to the source of the $(k+1)^{th}$, or the $(k+2)^{th}$, . . . , or the n^{th} main transistor, such that the gate of the additional p-type auxiliary transistor is connected to the gate of the k^{th} n-type auxiliary transistor, and the source of the additional p-type auxiliary transistor is connected to the source of the k^{th} n-type auxiliary transistor, and the drain of the additional p-type auxiliary transistors is connected to the source of the $(k+1)^{th}$, or the $(k+2)^{th}$, . . . , or the n^{th} main transistor.

13. The power conversion circuit of claim **11** in which the transistors of the main chain and auxiliary chain are p-type and further comprising an n-type transistor on a line connecting the gate of the k^{th} main transistor to the source of the $(k-1)^{th}$, or the$(k-2)^{th}$, . . . , or the 1^{st} main transistor, such that the gate of the additional n-type auxiliary transistor is connected to the gate of the k^{th} p-type auxiliary transistor and the source of the additional n-type auxiliary transistor is connected to the source of the k^{th} n-type auxiliary transistor, and the drain of the additional n-type auxiliary transistors is connected to the source of the $(k-1)^{th}$, or the$(k-2)^{th}$, . . . , or the 1^{st} main transistor.

14. The power conversion circuit of claim **11** in which the transistors of the main chain and auxiliary chain are n-type and further comprising a diode connected p-type or n-type transistor on a line connecting the gate of the k^{th} main transistor to the source of the $(k+1)^{th}$, or the $(k+2)^{th}$, . . . , or the n^{th} main transistor.

15. The power conversion circuit of claim **11** in which the transistors of the main chain and auxiliary chain are p-type and further comprising a diode connected p-type or n-type transistor on a line connecting the gate of the k^{th} main transistor to the source of the $(k-1)^{th}$, or the $(k-2)^{th}$, . . . , or the 1^{st} main transistor.

* * * * *

US 20060184209A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2006/0184209 A1**

John et al. (43) **Pub. Date: Aug. 17, 2006**

(54) **DEVICE FOR BRAIN STIMULATION USING RF ENERGY HARVESTING**

(76) Inventors: **Constance M. John**, San Francisco, CA (US); **Varghese John**, San Francisco, CA (US); **Martin H. Mickle**, Pittsburgh, PA (US)

Correspondence Address:
BUCHANAN INGERSOLL PC
(INCLUDING BURNS, DOANE, SWECKER & MATHIS)
POST OFFICE BOX 1404
ALEXANDRIA, VA 22313-1404 (US)

(21) Appl. No.: **11/219,404**

(22) Filed: **Sep. 2, 2005**

Related U.S. Application Data

(60) Provisional application No. 60/606,853, filed on Sep. 2, 2004.

Publication Classification

(51) **Int. Cl.**
A61N 1/34 (2006.01)

(52) **U.S. Cl.** **607/45**

(57) **ABSTRACT**

A device for brain stimulation using radio frequency harvesting is disclosed. The device includes a circuit implantable under a scalp of a patient, the circuit comprising a radio frequency harvesting power circuit and a stimulation circuit, and a plurality of electrodes coupled to the circuit, the plurality of electrodes providing brain stimulation to targeted areas of the brain. The electrodes may provide stimulation to targeted areas of the brain including deep brain stimulation for the treatment of Parkinson's disease and cortical stimulation for the treatment of stroke victims.

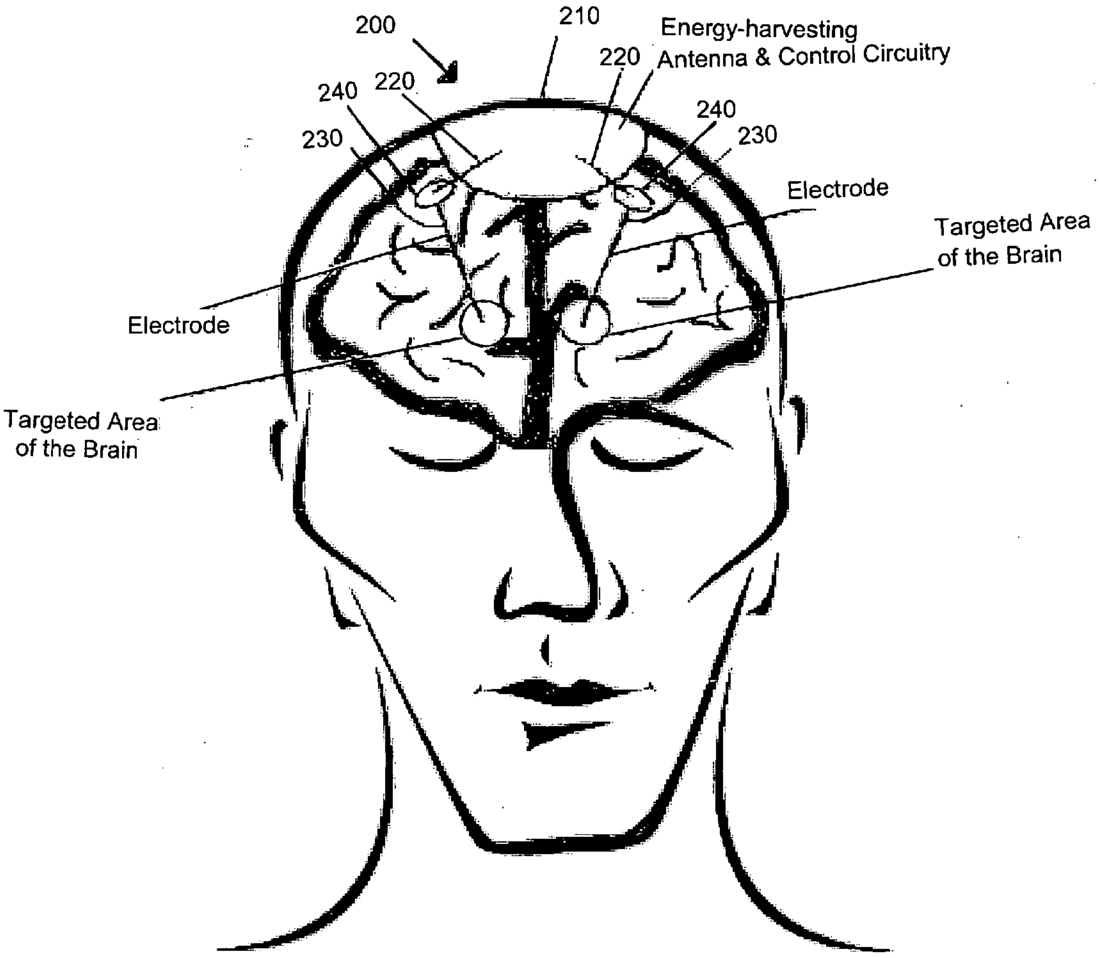
200
210
Energy-harvesting Antenna & Control Circuitry
220
240
230
220
240
230
Electrode
Targeted Area of the Brain
Electrode
Targeted Area of the Brain

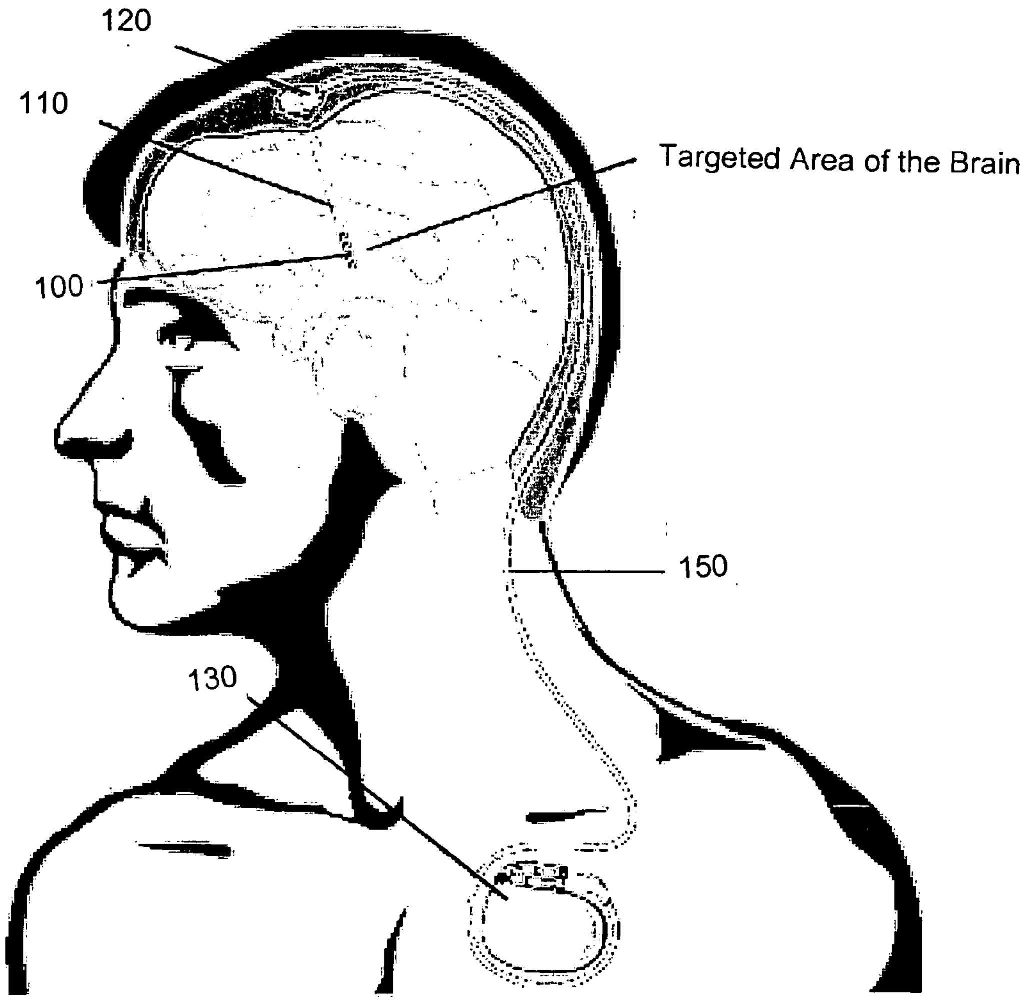
120
110
Targeted Area of the Brain
100
150
130

PRIOR ART

FIG. 1

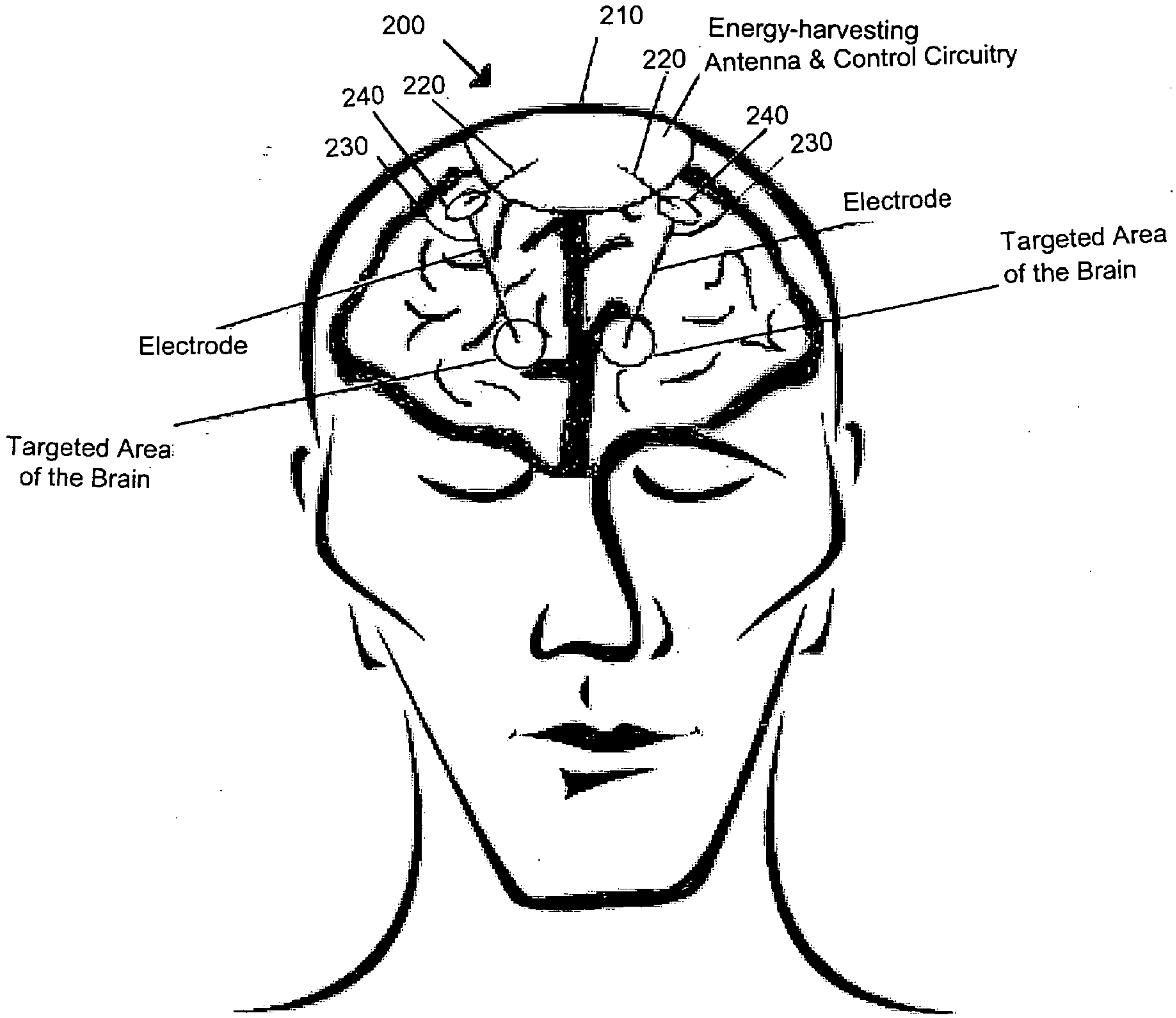
Energy-harvesting Antenna & Control Circuitry
Electrode
Targeted Area of the Brain
Electrode
Targeted Area of the Brain

FIG. 2A

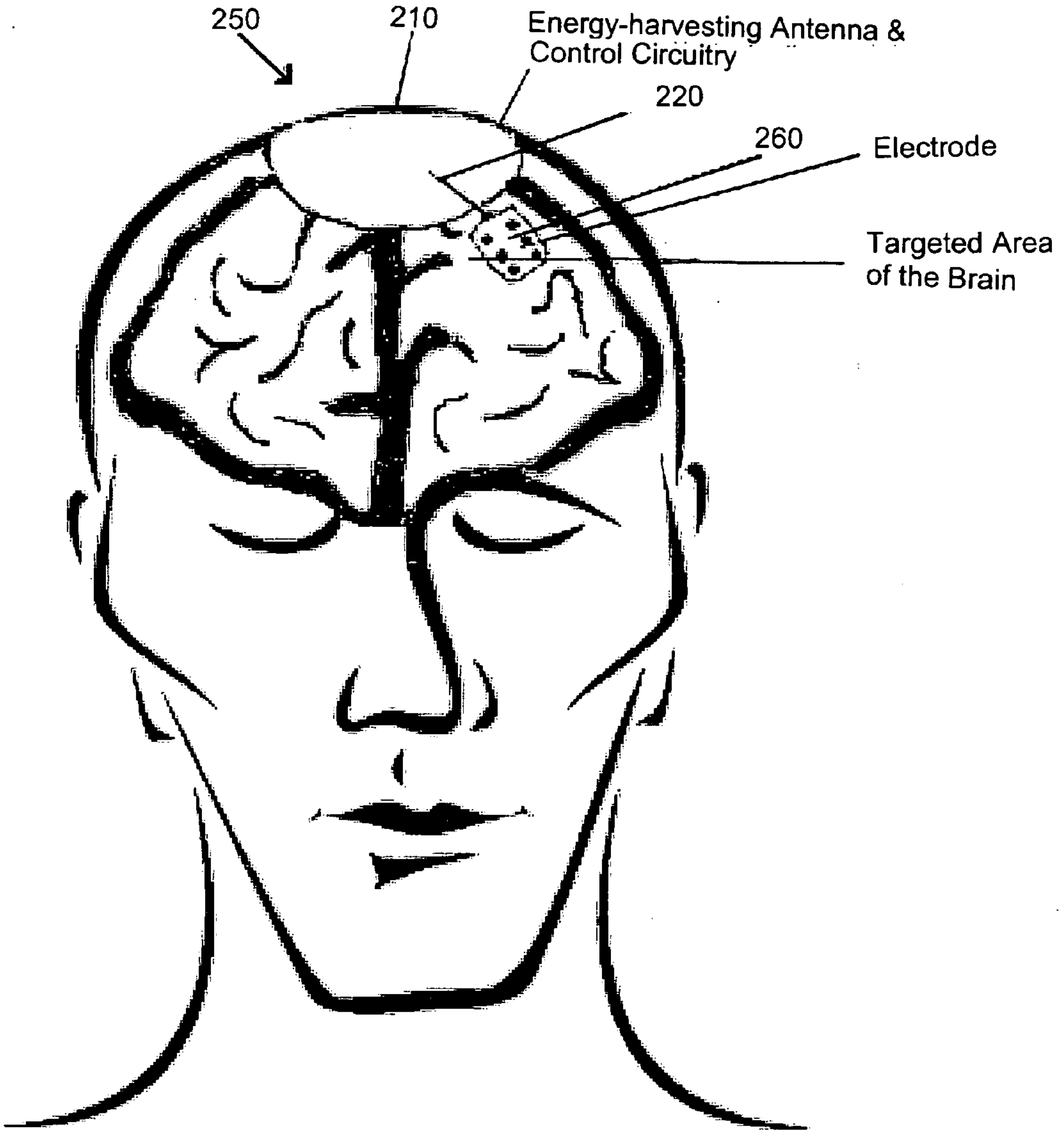
250
210
Energy-harvesting Antenna & Control Circuitry
220
260
Electrode
Targeted Area of the Brain

FIG. 2B

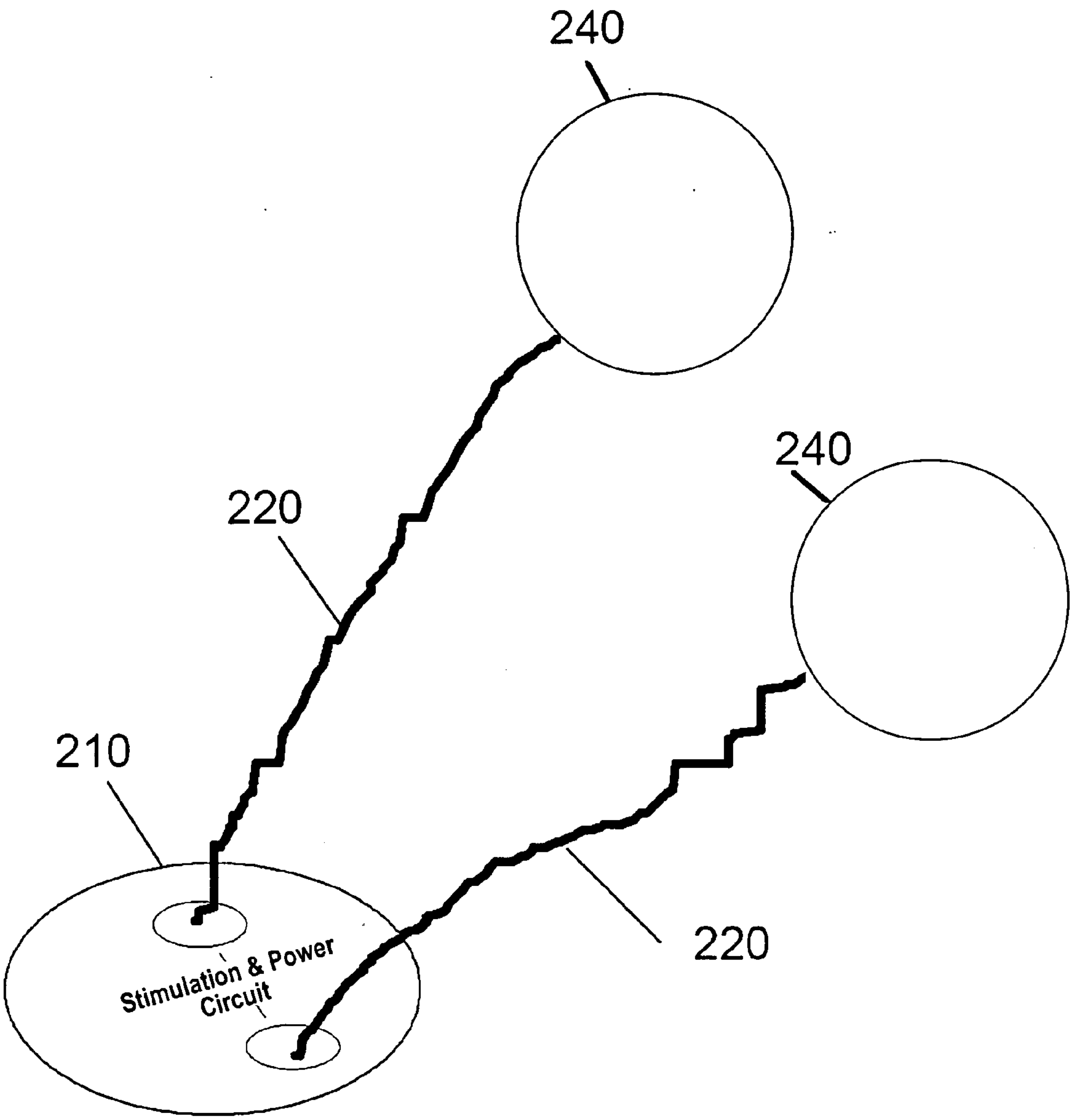
240
240
220
220
210
Stimulation & Power Circuit

FIG. 2C

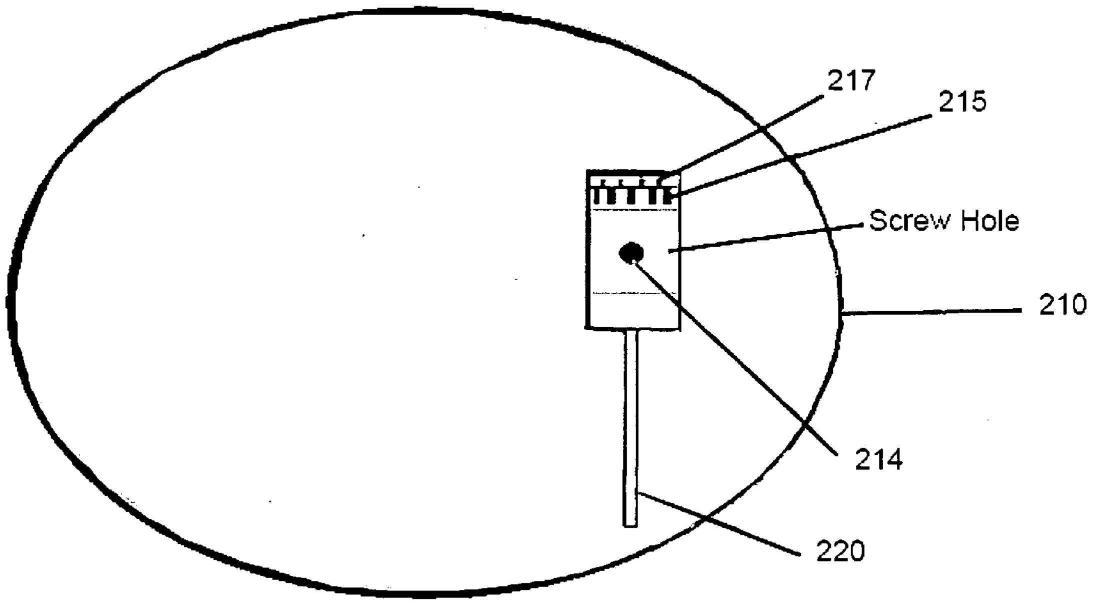
217
215
Screw Hole
210
214
220

FIG. 2D

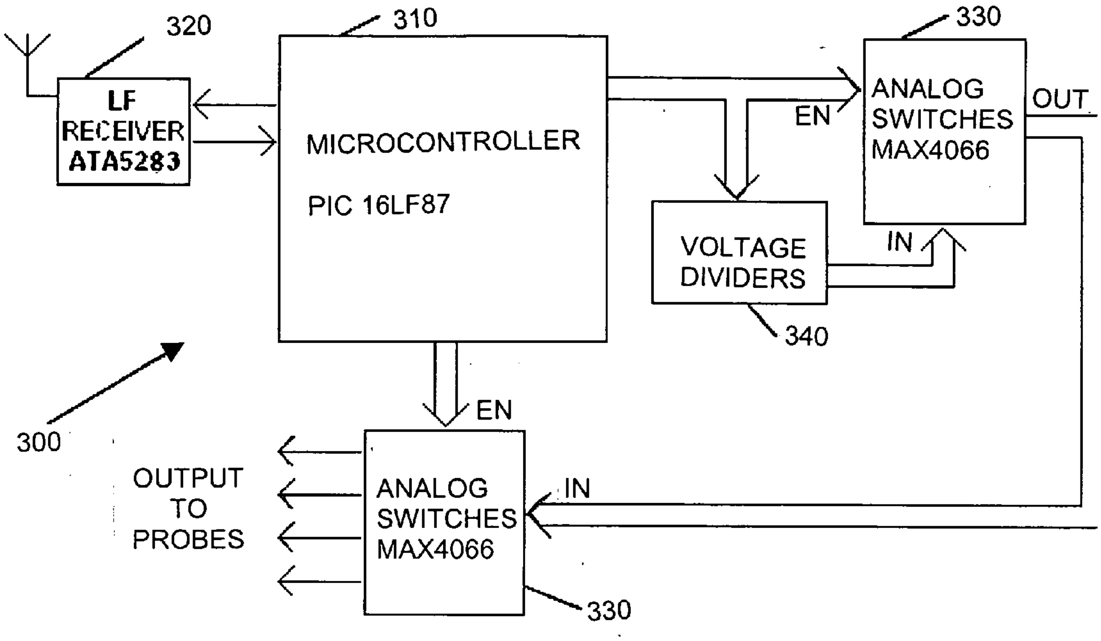
320
LF
RECEIVER
ATA5283
310
MICROCONTROLLER
PIC 16LF87
330
ANALOG
SWITCHES
MAX4066
OUT
EN
VOLTAGE
DIVIDERS
IN
340
300
EN
OUTPUT
TO
PROBES
ANALOG
SWITCHES
MAX4066
IN
330

FIG. 3

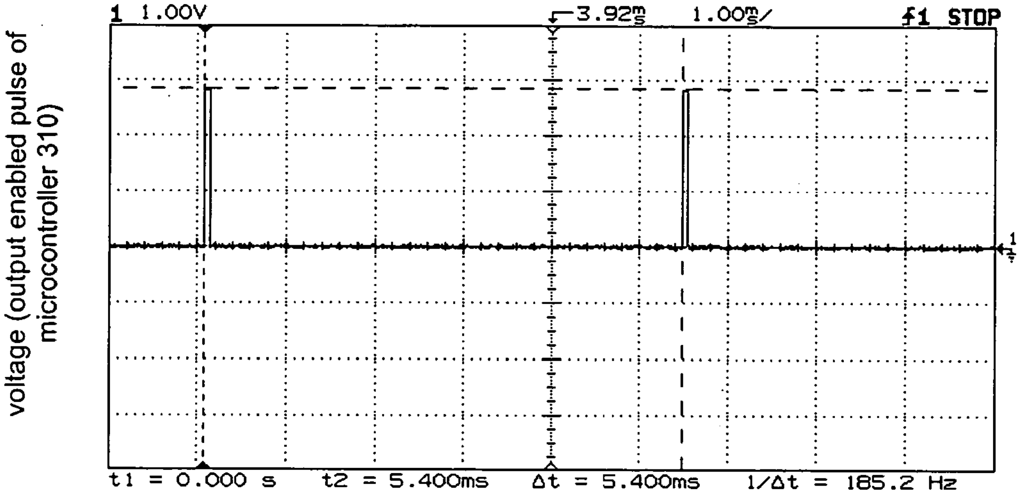
voltage (output enabled pulse of microcontroller 310)
1 1.00V
-3.92ms
1.00ms/
f1 STOP
1
t1 = 0.000 s
t2 = 5.400ms
Δt = 5.400ms
1/Δt = 185.2 Hz

FIG. 4

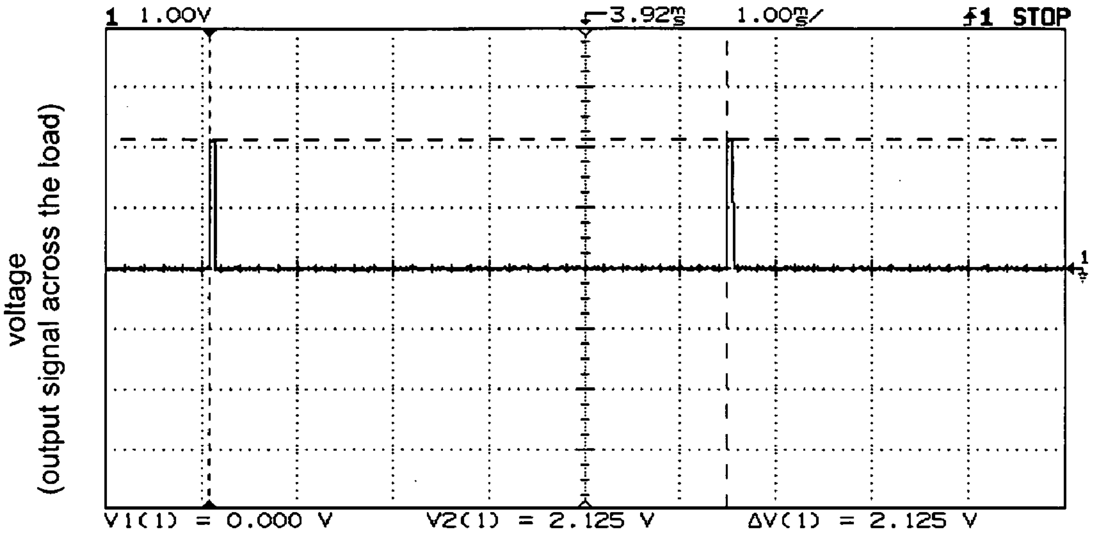
1 1.00V
-3.92ms
1.00ms/
f1 STOP
voltage
(output signal across the load)
V1(1) = 0.000 V
V2(1) = 2.125 V
ΔV(1) = 2.125 V

FIG. 5

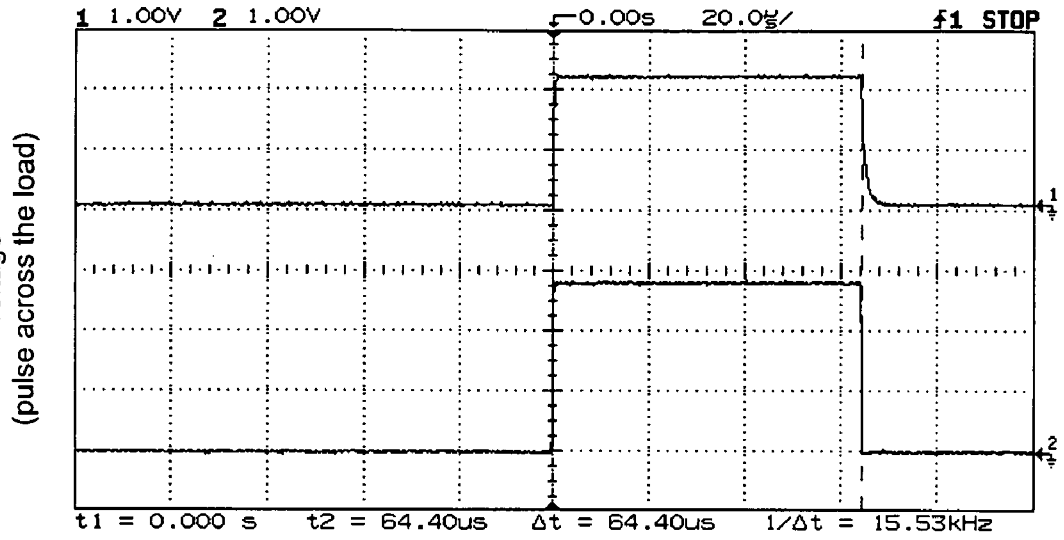
1 1.00V 2 1.00V
0.00s
20.0µs/
f1 STOP
voltage
(pulse across the load)
1
2
t1 = 0.000 s
t2 = 64.40us
Δt = 64.40us
1/Δt = 15.53kHz

FIG. 6

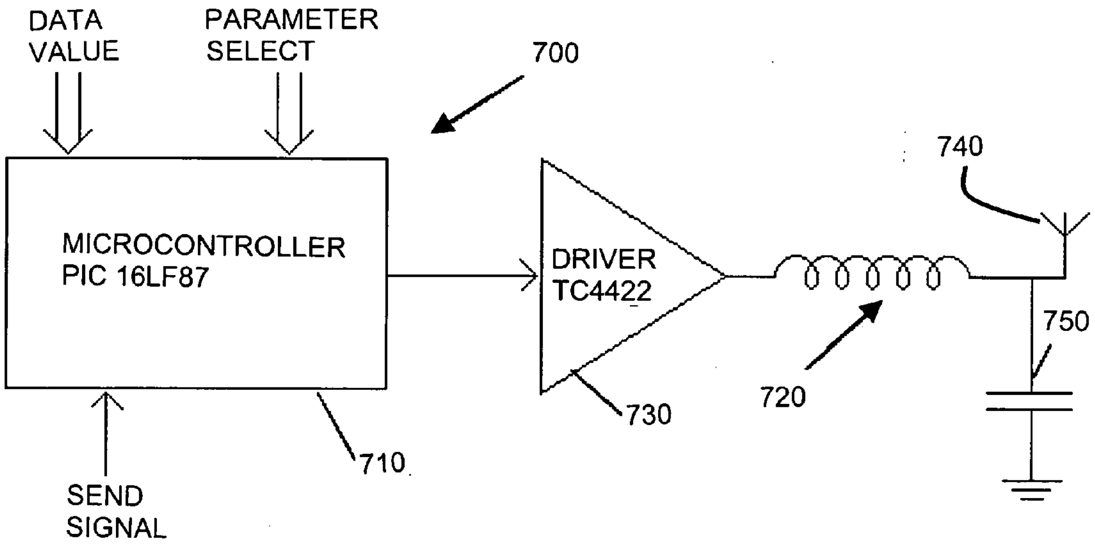
DATA
VALUE
PARAMETER
SELECT
700
740
MICROCONTROLLER
PIC 16LF87
DRIVER
TC4422
750
730
720
710
SEND
SIGNAL

FIG. 7

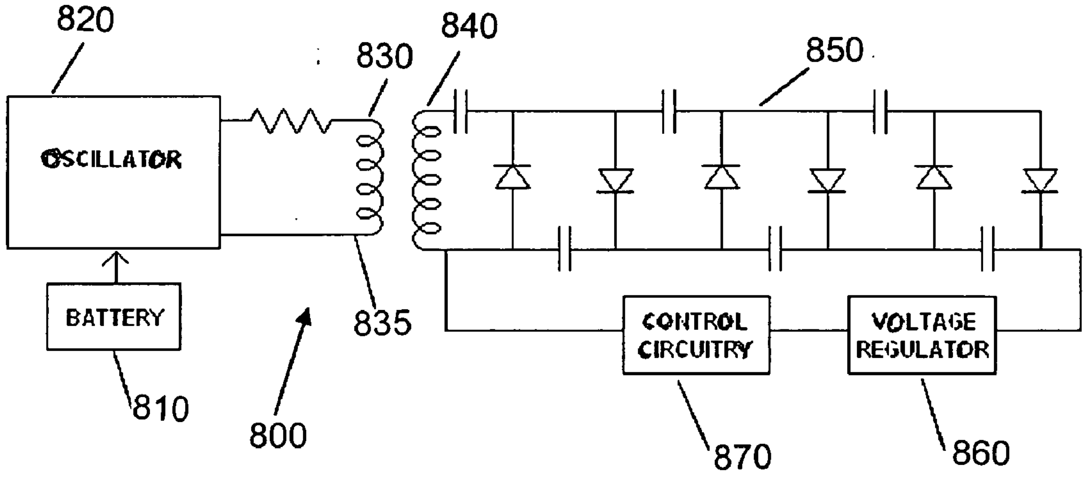
OSCILLATOR
BATTERY
CONTROL CIRCUITRY
VOLTAGE REGULATOR

FIG. 8A

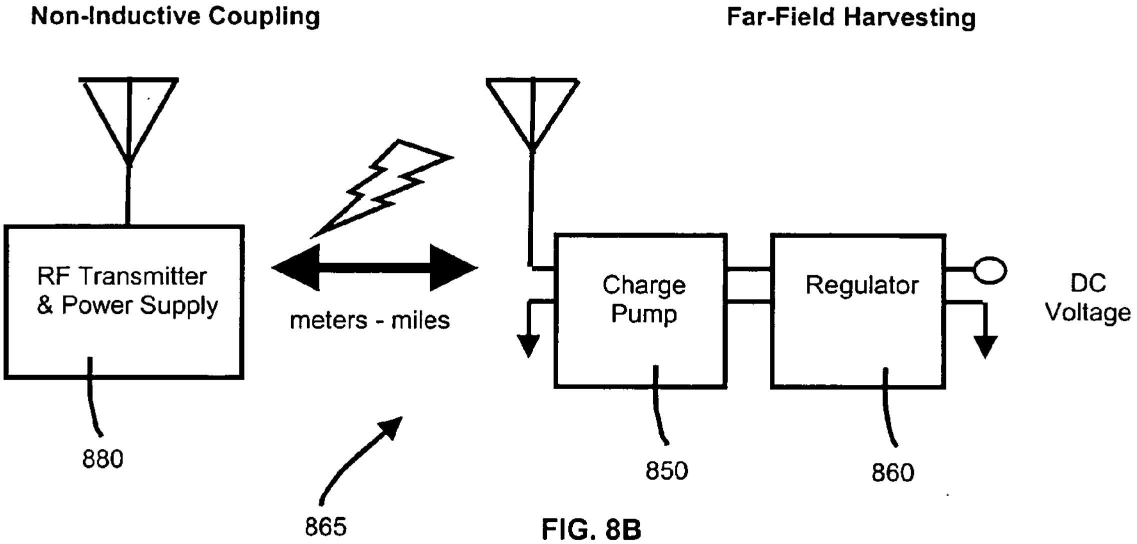
Non-Inductive Coupling
Far-Field Harvesting
RF Transmitter & Power Supply
meters - miles
Charge Pump
Regulator
DC Voltage
880
865
850
860

FIG. 8B

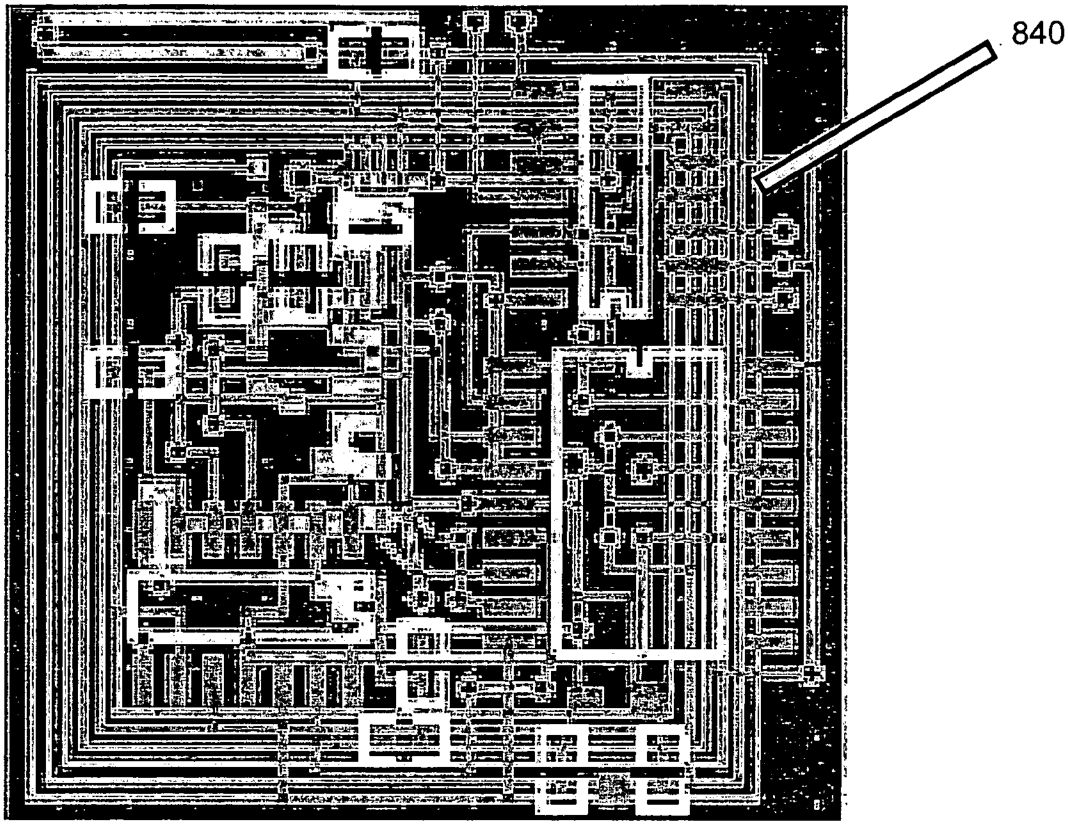

FIG. 9

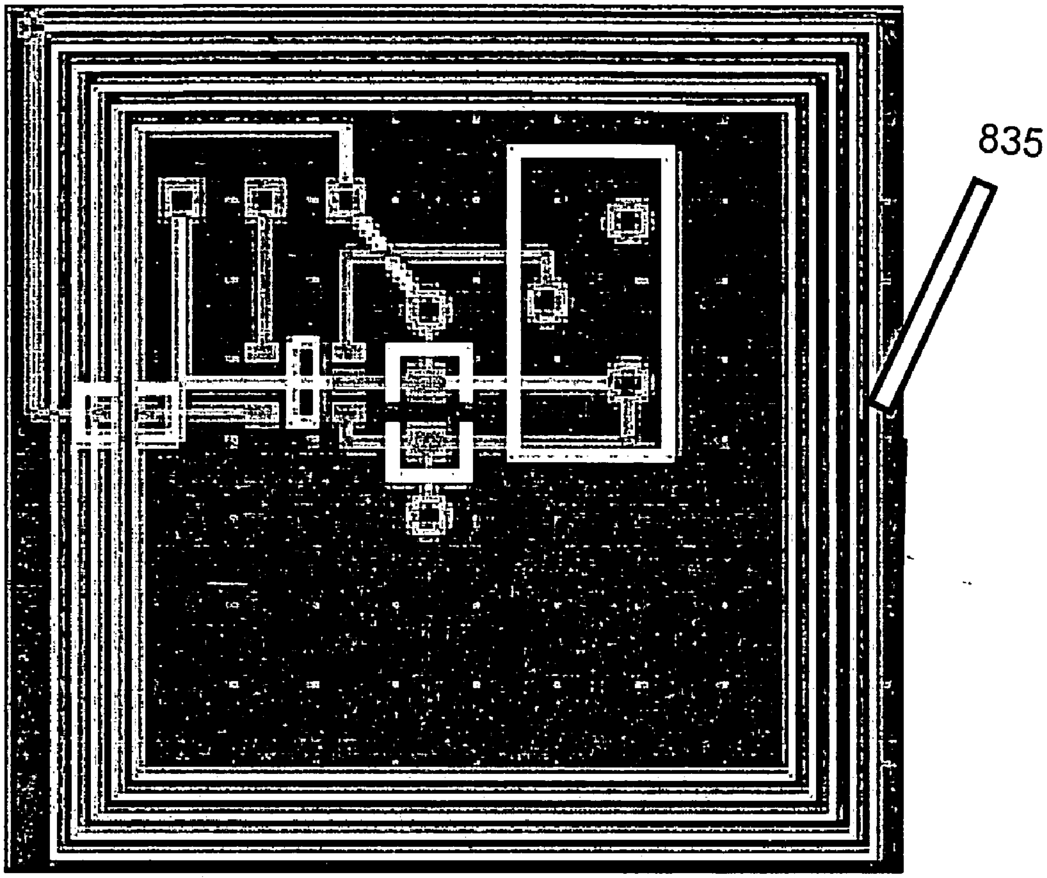

FIG. 10

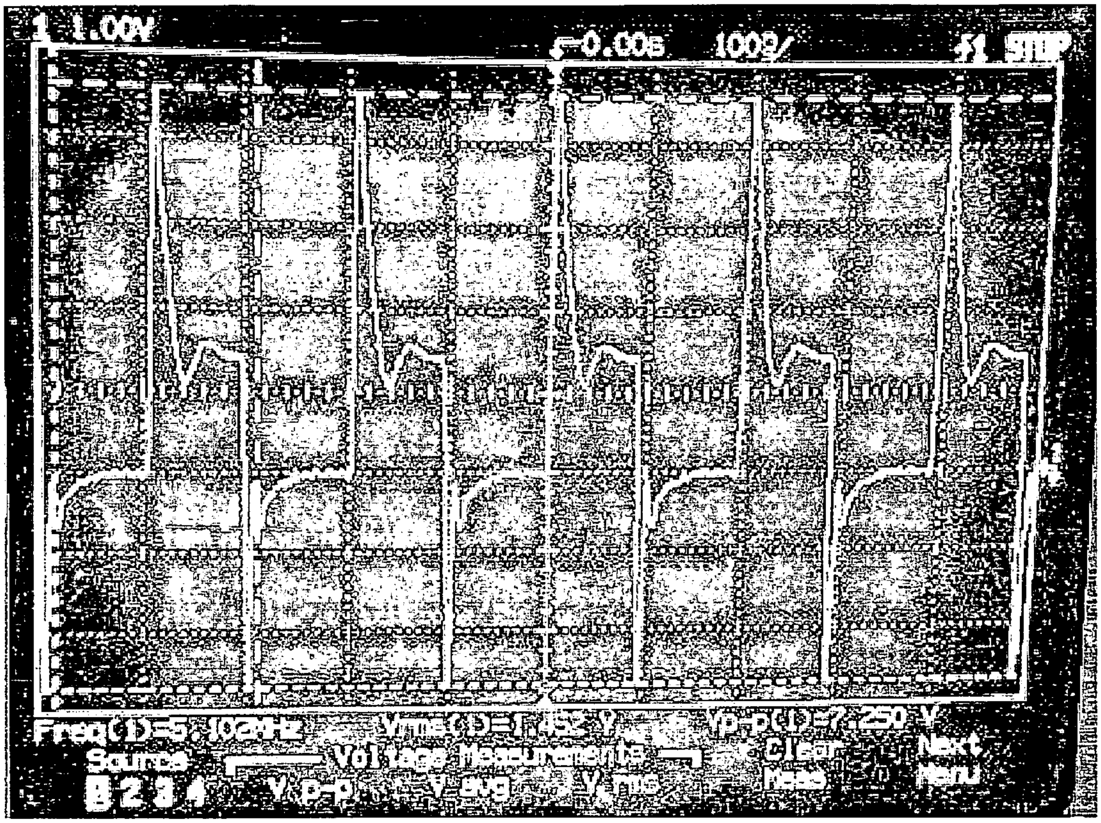
1 1.00V
0.00s
100?/
f1 STOP
Freq(1)=5.102MHz
Vrms(1)=1.452 V
Vp-p(1)=7.250 V
Source
Voltage Measurements
Clear Meas
Next Menu
1 2 3 4
V p-p
V avg
V rms

FIG. 11

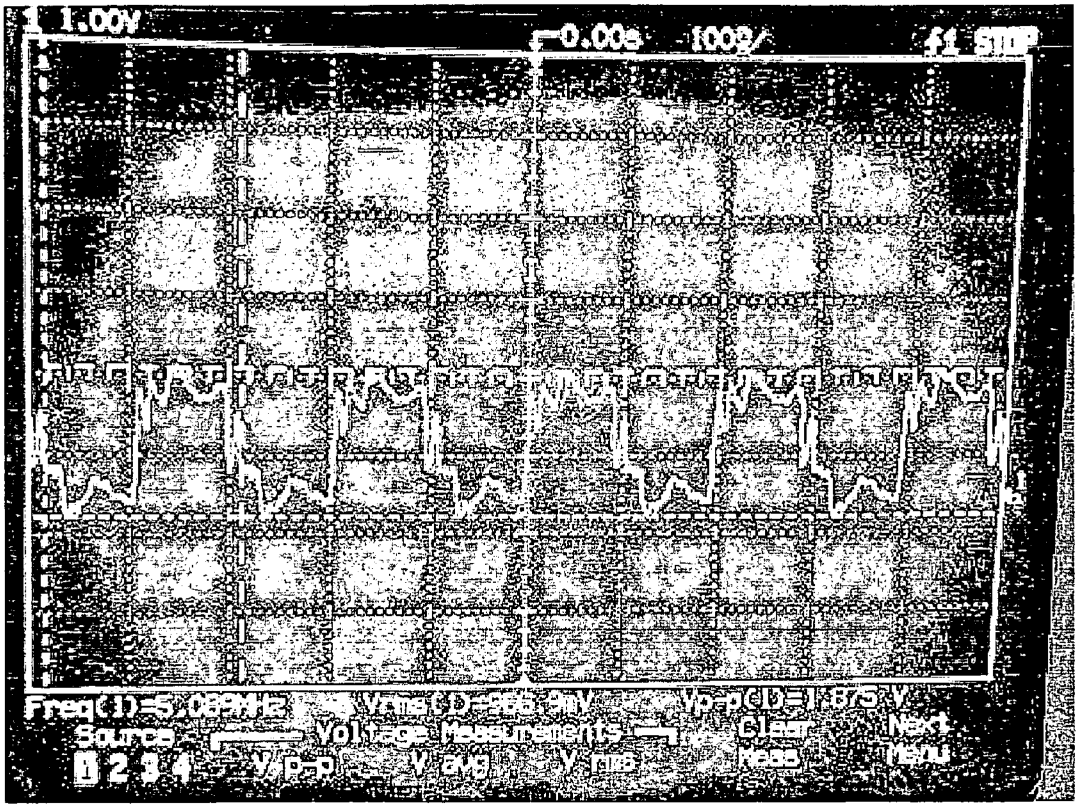
1 1.00V
0.00s
STOP
Freq(1)=5.08MHz
Vp-p(1)=1.875V
Source
Voltage Measurements
Clear Meas
Next Menu
1 2 3 4
V p-p
V avg
V rms

FIG. 12

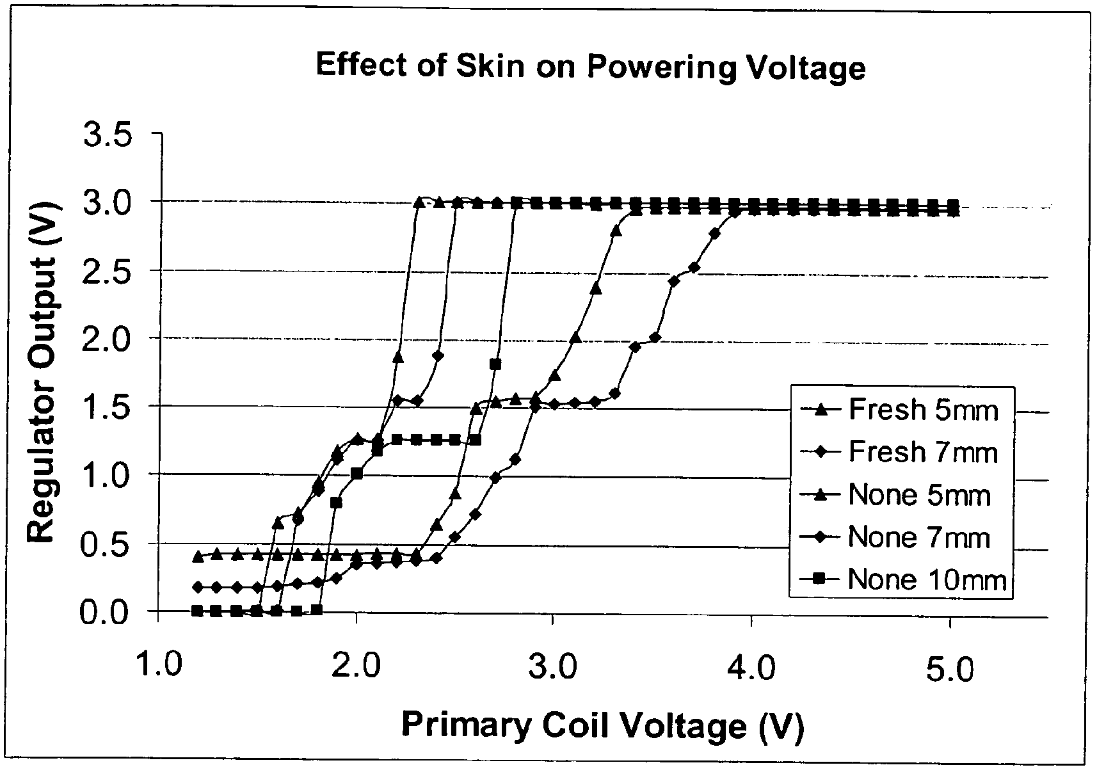
Effect of Skin on Powering Voltage
Regulator Output (V)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1.0
2.0
3.0
4.0
5.0
Primary Coil Voltage (V)
Fresh 5mm
Fresh 7mm
None 5mm
None 7mm
None 10mm

FIG. 13

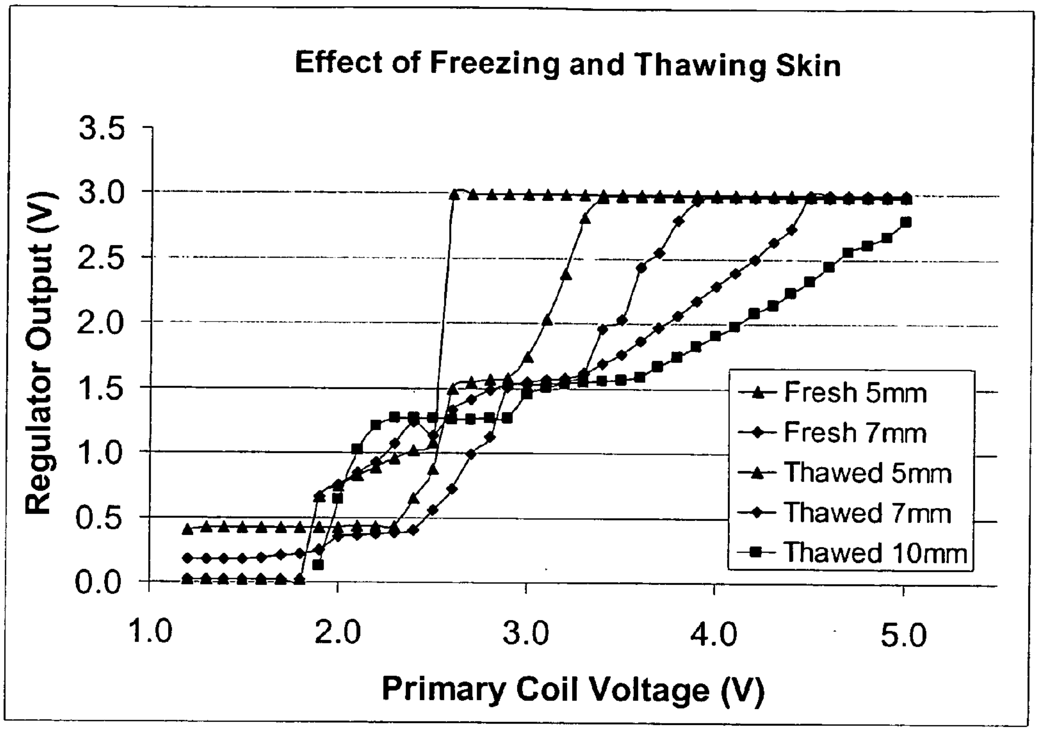
Effect of Freezing and Thawing Skin
Regulator Output (V)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1.0
2.0
3.0
4.0
5.0
Primary Coil Voltage (V)
Fresh 5mm
Fresh 7mm
Thawed 5mm
Thawed 7mm
Thawed 10mm

FIG. 14

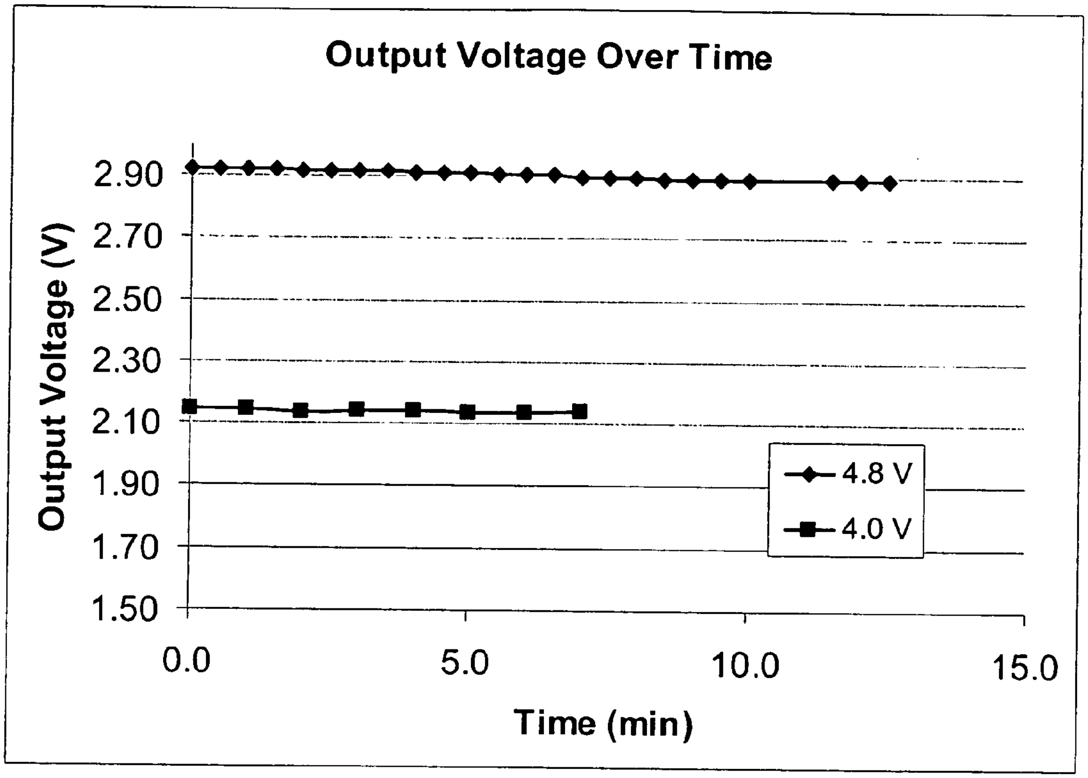
Output Voltage Over Time
Output Voltage (V)
2.90
2.70
2.50
2.30
2.10
1.90
1.70
1.50
0.0
5.0
10.0
15.0
Time (min)
4.8 V
4.0 V

FIG. 15

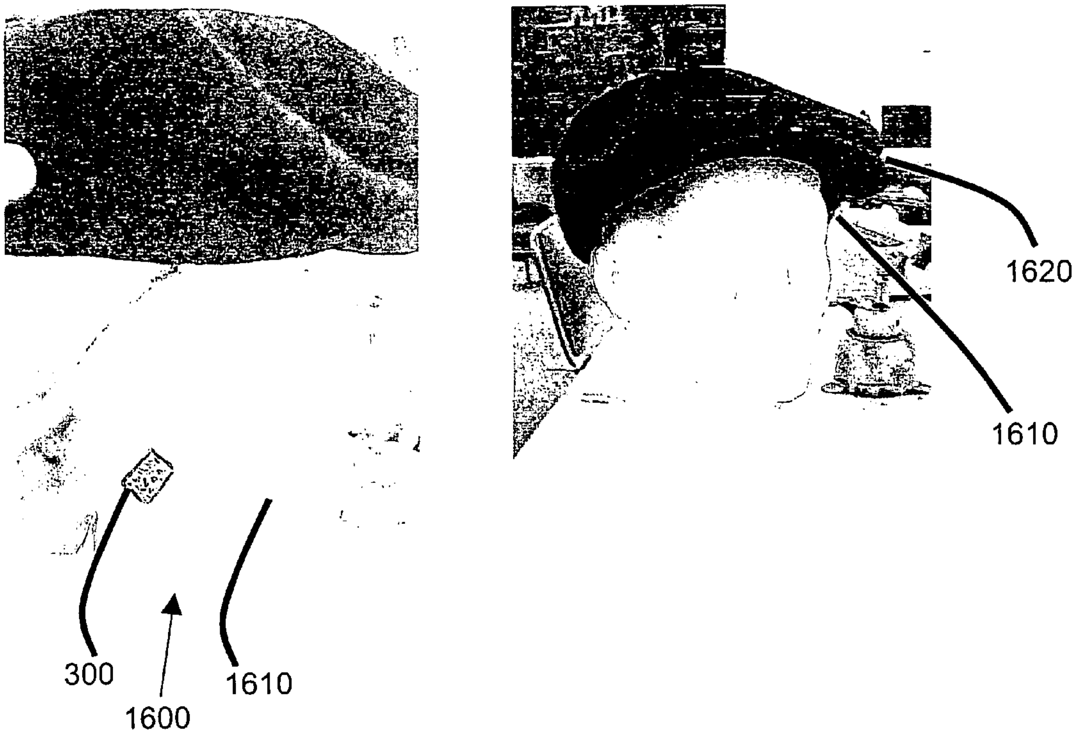
300
1600
1610
1620
1610

FIG. 16

DEVICE FOR BRAIN STIMULATION USING RF ENERGY HARVESTING

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims priority under 35 U.S.C. 119(e) from provisional patent application Ser. No. 60/606,853, entitled "Device For Deep Brain Stimulation (DBS) Using RF Energy Harvesting", filed on Sep. 2, 2004, the disclosure of which is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

[0002] The present invention relates generally to systems and apparatus for providing brain stimulation and more particularly to a device for harvesting radio frequency (RF) energy that can be implanted under a human scalp to produce stimulation in different regions of the brain, including deep brain stimulation (DBS).

BACKGROUND OF THE INVENTION

[0003] DBS is a surgical technique first used in humans over 25 years ago. DBS has been used in a wide variety of brain targets, including the thalamus, globus pallidus and the subthalamic nucleus. Diseases that have been effectively treated with DBS include movement disorders including essential tremor [Lyons K E, Pahwa R. Deep Brain Stimulation and Essential Tremor, J Clin Neurophysiol. 2004 January-February;21(1):2-5], Parkinson's disease [Byrd D L, Marks W J Jr, Starr P A. Deep brain stimulation for advanced Parkinson's disease. AORN J. 2000 September;72(3):387-90, 393-408] and dystonia [Vidailhet M. et al., Bilateral deep-brain stimulation of the globus pallidus in primary generalized dystonia. N Engl J Med. 2005 Feb. 3;352(5):459-67]. Other indications for DBS are being explored, including chronic pain, cluster headache, persistent vegatative state, epilepsy, Alzheimer's, and psychiatric disorders including obsessive-compulsive disorder and intractable depression.

[0004] Parkinson's disease (PD) is an idiopathic neurodegenerative disorder that is characterized by the presence of tremor, rigidity, akinesia or bradykinesia (slowness of movement) and postural instability. It is believed to be caused by the loss of a specific, localized population of neurons in a region of the brain called the substantia nigra. These cells normally produce dopamine, a neurotransmitter that allows brain cells to communicate with each other. These dopaminergic cells in the substantia nigra are part of an elaborate motor circuit in the brain that runs through a series of discrete brain nuclei known as the basal ganglia that control movement. It is believed that the symptoms of PD are caused by an imbalance of motor information flow through the basal ganglia.

[0005] Conventionally, a medication known as levodopa has been the mainstay of treatment for patients with Parkinson's disease. However, long-term therapy with levodopa has several well-known complications that limit the medications effectiveness and tolerability. The first of these is the development of involuntary movements known as dyskinesias. These movements can be violent at times and as or more disabling than the Parkinson's symptoms themselves. The other frequent complication is the development of "on-off" fluctuations, where patients cycle between periods of good function (the "on" period) and periods of poor function (the "off" period). These fluctuations can become very frequent, up to 7 or more cycles per day, and can cause patients to become suddenly and unpredictably "off" to the point where they cannot move.

[0006] Lesioning procedures such as pallidotomy were known to improve the motor symptoms of Parkinson's disease, presumably by disruption of the abnormal neuronal activity in the motor circuitry of the basal ganglia. The discovery that MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) produced a Parkinsonian-like state in non-human primates allowed electrophysiologic study of this phenomenon by numerous investigators. The discovery that high frequency stimulation could mimic the effect of lesioning led to the use of DBS for PD in humans in the early 1990's. DBS was found to improve all of the cardinal symptoms of Parkinson's disease while allowing the patient to decrease or sometimes even eliminate the amount of levodopa medication, therefore decreasing both dyskinesia and "on-off" fluctuations.

[0007] DBS is currently the surgical treatment of choice for medically refractory Parkinson's disease. Two brain targets have been found to provide clinical benefit when chronically stimulated; the subthalamic nucleus (STN) and the internal segment of the globus pallidus (GPi). In a recent prospective, double-blinded cross-over study involving 96 patients with STN DBS and 38 patients with GPi DBS, the STN group reported an improvement in the percentage of time spent during the day with good mobility and without dyskinesia from 27% to 74%. The GPi group also reported a significant improvement, from 28% to 64%.

[0008] Although the mechanism of action is not fully understood it is thought to act by either depolarization blockade, synaptic inhibition, synaptic depression, or stimulation-induced modulation of pathologic network activity [McIntyre C C, Savasta M, Walter B L, Vitek J L. How does deep brain stimulation work? Present understanding and future questions. J Clin Neurophysiol. 2004 January-February;21(1):40-50]. It is believed that DBS acts somehow to suppress the neuronal activity by the stimulation of the region of the brain immediately adjacent to the electrode. This hypothesis seems to be supported by the fact that lesioning a specific structure in the brain has the same clinical effect as stimulating that same structure at high (greater than 100-150 Hz) frequency. In fact, DBS has largely replaced the older lesioning procedures (such as pallidotomy and thalamotomy) that used to be the mainstay of surgical treatment for movement disorders such as Parkinson's disease. The high frequency stimulation may act to hyperpolarize immediately adjacent neurons such that they become incapable of producing normal action potentials. An alternative hypothesis is that DBS may be altering more distant structures or even fibers from far removed nerve cells that are passing through or near the area of stimulation. Whatever the mechanism of action, DBS has a distinct advantage over the older lesioning techniques because it is an adjustable therapy and does not involve destruction of the patient's brain tissue.

[0009] Prior art DBS devices have several limitations that can lead to adverse effects including infection, cutaneous erosion, and lead breaking or disconnection [Temel Y,

Ackermans L, Celik H, Spincemaille G H, Van Der Linden C, Walenkamp G H, Van De Kar T, Visser-Vandewalle V. Management of hardware infections following deep brain stimulation. Acta Neurochir (Wien) 2004;146(4):355-61; Putzke J D, Wharen R E, Jr., Wszolek Z K, Turk M F, Strongosky A J, Uitti R J. Thalamic deep brain stimulation for tremor-predominant Parkinson's disease. Parkinsonism Relat Disord 2003;10(2):81-8.; Umemura A, Jaggi J L, Hurtig H I, Siderowf A D, Colcher A, Stern M B, Baltuch G H. Deep brain stimulation for movement disorders: morbidity and mortality in 109 patients. J Neurosurg 2003;98(4):779-84; Hariz M I. Complications of deep brain stimulation surgery. Mov Disord 2002;17(Suppl 3):S162-**6**]. One study found that 27% of 66 patients with implanted DBS devices had hardware problems [Kondziolka D, Whiting D, Germanwala A, Oh M. Hardware-related complications after placement of thalamic deep brain stimulator systems. Stereotact Funct Neurosurg 2002;79(3-4):228-33. This relatively high incidence of hardware problems is similar to the results of a study where 20 (25.3%) of 79 patients who received 124 permanent DBS electrode implants had 26 hardware-related complications [Oh M Y, Abosch A, Kim S H, Lang A E, Lozano A M. Long-term hardware-related complications of deep brain stimulation. Neurosurgery 2002;50(6):1268-74; discussion 1274-**6**]. In addition, intracranial electrode implantation can induce a hematoma or contusion. Nonetheless, most authors agree that the benefit to risk ratio of DBS is favorable.

[0010] A prior art DBS device is shown in **FIG. 1** and includes an electrode **100** disposed in a targeted area of the brain. The electrode is coupled to a lead **110** held in place at the top of the skull by a securement device **120**. The lead **110** is coupled to a neurostimulator **130** powered by a battery-powered pulse generator **140** by means of a lead **150**. The lead **150**, which averages about 15 inches in length, is implanted under the scalp and traverses the length of the patient's neck to the chest where the neurostimulator **130** and battery **140** are implanted.

[0011] The pulse generator **140** is typically placed underneath the skin just below the collar bone and is capable of stimulating at one or any combination of the four contacts present on the end of the electrode **110** in the brain. The parameters of the stimulating current (voltage, frequency, pulse width) can also be selected by the treating physician or health care worker. The pulse generator **140** is programmed through the skin via a telemetry device that allows the practitioner to select the desired stimulation parameters and also perform diagnostic tests on the device.

[0012] Implantation of the conventional DBS device is costly as for implantation of a single electrode in the brain for treatment of one side of the body the procedure requires three incisions; one on the top of the head, one behind the ear and the third just below the collarbone where the leads are connected. The implantation of the electrode **110** and the implantable pulse generator **140** is sometimes performed on different days. The incisions can be prone to infection in the immediate postoperative period. In some elderly patients with thin skin, the pulse generator **140** or wire can erode through the skin and become exposed to potential contamination. Infection or erosion often results in the need to remove the entire device, as antibiotic treatment alone in this setting rarely will clear the infection adequately. The lead **150** restricts the patient's mobility in the neck region and may break. Furthermore, the battery **140** must be replaced every three to five years. Additional drawbacks of the DBS device include the risk of erosion of the leads or hardware, infection, and magnetic sensitivity.

[0013] A prior art deep brain stimulation system is disclosed in U.S. Pat. No. 6,920,359 entitled "Deep Brian Stimulation System for the Treatment of Parkinson's Disease or Other Disorders". The DBS system includes a small, implantable pulse generator implanted directly in the cranium of the patient, thereby eliminating the long lead wires conventionally used. The disclosed system does not provide for the harvesting of energy to power the pulse generator.

[0014] As noted in Table 1, there are several current and potential indications for deep brain stimulation.

[0015] Known systems for providing electrical stimulus to the motor cortex of the brain, such as the Northstar Stroke Recovery Treatment System available from Northstar Neuroscience, Inc., also include an implantable pulse generator implanted in the pectoral area of a patient. A cortical stimulation lead includes an electrode connected to the implantable pulse generator which is used to deliver stimulation to the cortex. The electrode is placed on top of the dura and coupled to the implantable pulse generator by means of a lead which traverses the length of the patient's neck to the patient's pectoral area.

[0016] Motor cortex stimulation (MCS) is a process involving the application of stimulation signals to the motor cortex in the brain of a patient during physical rehabilitation of the disabled region of the body. The MCS system includes a pulse generator connected to a strip electrode that is surgically implanted over a portion of only the motor cortex (precentral gyrus). Because MCS involves the application of stimulation signals to surface regions of the brain rather than deep neural structures, electrode implantation procedures for MCS are significantly less invasive and time consuming than those for DBS. The current evaluation of MCS is for stroke. Stroke-related disabilities affect more than 200,000 people in the U.S. each year. Good results have been reported in MCS treatment of stroke victims. With a MCS device, a stamp-sized electrode is placed on the surface of the brain. This is attached to a wire that goes through the neck to an implantable pulse generator in the pectoral area.

[0017] Neurostimulation and responsive neurostimulation (RNS) are also being tested for the treatment of medically refractory epilepsy. In treating epilepsy, the RNS system can be designed to detect abnormal electrical activity in the brain and respond by delivering electrical stimulation to normalize brain activity before the patient experiences seizure symptoms. For either neurostimulation or RNS for treatment of epilepsy the electrode or electrodes of the device deliver a short train of electrical pulses to the brain near the patient's seizure focus.

[0018] In order to obviate the need for long leads and batteries, attempts have been made in the prior art to transmit energy through space from a base station to a remote station. One such system is disclosed in U.S. Pat. No. 6,289,237 entitled "Apparatus for Energizing a Remote Station and Related Method". The base station transmits energy which may be RF power, light, acoustic, magnetic or other suitable forms of space transmitted or "radiant" energy to the remote station. Within the remote station, the received

energy is converted into DC power which serves to operate the remote station. The source of power for the remote station is the base station and, therefore, there is no need for the remote station to carry an electrical storage device such as a battery. It is suggested that this facilitates the remote station being encapsulated within a suitable protective material, such as a resinous plastic. Homopolymers, elastomers and silicon dioxide are also suggested as suitable materials for such purposes. Further, it is suggested that this facilitates miniaturization of the remote station and placing the remote station in functionally desirable locations which need not be readily accessible. The remote station, for example, could be implanted in a patient.

[0019] The use of a wireless communication link between a base station and transponders in a radio frequency identification system employing modulated back-scattered waves is also known. See Rao, An Overview of Bulk Scattered Radio Frequency Identification System (RFID) IEEE (1999). It has also been suggested to employ a silicon chip in a transponder having a change pump on voltage doubler current. Hornby, RFID Solutions for the Express Parcel and Airline Baggage Industry, Texas Instruments, Limited (Oct. 7, 1999).

[0020] For use in miniaturized electronic chip systems, an electronic article containing a microchip having at least one antenna structured to communicate with an antenna remotely disposed with respect to the microchip is disclosed in U.S. Pat. No. 6,615,074 entitled "Apparatus for Energizing a Remote Station and Related Method". Power enhancement is achieved using a voltage doubler. The antenna of the disclosed apparatus is comparable in volume to a Smart Dust device. Smart Dust is a combination MEMS/Electronic device on the order of 1 mm×1 mm×1 mm.

[0021] What is needed therefore is a brain stimulation device that overcomes the disadvantages of the prior art brain stimulation devices. What is needed is a brain stimulation device that requires a single implantation site and surgery. What is also needed is a brain stimulation device that uses RF energy as a power source. What is further needed is a brain stimulation device that converts RF energy and stores the converted RF energy. What is also needed is a brain stimulation device that is flexible and implantable under the scalp. What is needed is a brain stimulation device that does not require leads or a pulse generator to be placed outside of the head area that are subject to disconnection or breakage. What is also needed is a brain stimulation device for electrical stimulation in the brain that is smaller and more self-contained and that does not require a pulse generator to be implanted elsewhere in the body. What is further needed is a device that is less susceptible to hardware problems or complications. What is needed is a device that has less potential for erosion through the skin. What is also needed is a device that is has a power source that does not need to be replaced.

SUMMARY OF INVENTION

[0022] The device for brain stimulation using RF energy harvesting of the present invention overcomes the disadvantages of the prior art, fulfills the needs in the prior art, and accomplishes its various purposes by providing a brain stimulation device that harvests radio frequency energy and is implantable under the scalp. The brain stimulation device of the invention may include an electrode that penetrates into the brain to provide neurostimulation to the brain. The brain stimulation device may also include an electrode that is used to provide stimulation to the brain cortex.

[0023] In accordance with one aspect of the invention, a device for brain stimulation using radio frequency harvesting includes a circuit implantable under a scalp of a patient, the circuit comprising a radio frequency harvesting power circuit and a stimulation circuit, and a plurality of electrodes coupled to the circuit, the plurality of electrodes providing brain stimulation to targeted areas of the brain. An advantage of this system is that it may use "trickle charging" wherein the device is charged by the harvesting power circuit. Moreover, anther advantage of this invention is the power transmitter which sends power to the device can be used both to send power and to send information.

[0024] There has been outlined, rather broadly, the more important features of the invention in order that the detailed description thereof that follows may be better understood, and in order that the present contribution to the art may be better appreciated. There are, of course, additional features of the invention that will be described below and which will form the subject matter of the claims appended herein.

[0025] In this respect, before explaining at least one embodiment of the invention in detail, it is to be understood that the invention is not limited in its application to the details of design and to the arrangement of components set forth in the following description or illustrated in the drawings. The invention is capable of other embodiments and of being practiced and carried out in various ways. Also, it is to be understood that the phraseology and terminology employed herein, as well as the abstract, are for the purpose of description and should not be regarded as limiting.

[0026] As such, those skilled in the art will appreciate that the conception upon which this disclosure is based may readily be utilized as a basis for the designing of other methods and systems for carrying out the several purposes of the present invention. It is important, therefore, that the claims be regarded as including such equivalent methods and systems insofar as they do not depart from the spirit and scope of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0027] The accompanying drawings, which are incorporated in and form a part of the specification, illustrate the embodiments of the present invention and together with the description, serve to explain the principles of the invention. In the drawings:

[0028] **FIG. 1** is a schematic representation of a prior art DBS device;

[0029] **FIG. 2A** is a schematic representation of a device for deep brain stimulation using RF energy harvesting in accordance with the invention;

[0030] **FIG. 2B** is a schematic representation of a device for cortical stimulation using RF energy harvesting in accordance with the invention;

[0031] **FIG. 2C** is a schematic representation of the device of **FIG. 2A** illustrating lead securement devices;

[0032] **FIG. 2D** is a schematic representation of the device of **FIG. 2A** illustrating an attachment means for connecting a lead wire to a circuit of the device;

[0033] **FIG. 3** is a schematic representation of a stimulation circuit in accordance with the invention;

[0034] **FIG. 4** is a graph showing an output enable pulse from a microcontroller of the stimulation circuit shown in **FIG. 3** in accordance with the invention;

[0035] **FIG. 5** is a graph showing an output signal from the microcontroller of the stimulation circuit shown in **FIG. 3** applied across a resistive load in accordance with the invention.

[0036] **FIG. 6** is a graph showing pulses across the resistive load;

[0037] **FIG. 7** is a schematic representation of an external programming circuit in accordance with the invention;

[0038] **FIG. 8A** is a schematic representation of an external power circuit inductively coupled to a power circuit in accordance with the invention;

[0039] **FIG. 8B** is a schematic representation of an alternative embodiment of the external power circuit non-inductively coupled to the power circuit in accordance with the invention;

[0040] **FIG. 9** is an illustration of a PCB layout of the stimulating circuit in accordance with the invention;

[0041] **FIG. 10** is an illustration of the external power circuit in accordance with the invention/

[0042] **FIG. 11** is an oscilloscope screen showing the output voltage from an oscillator of the external power circuit in accordance with the invention;

[0043] **FIG. 12** is an oscilloscope screen showing a voltage across a primary coil series resistance of the external power circuit in accordance with the invention;

[0044] **FIG. 13** is a graph showing the effect of skin disposed between the primary coil and a secondary coil of the stimulating circuit in accordance with the invention;

[0045] **FIG. 14** is a graph showing the effect of freezing the thawing the skin in accordance with the invention;

[0046] **FIG. 15** is a graph of the output voltage over time in accordance with the invention; and

[0047] **FIG. 16** is a pictoral representation of a model having the stimulation circuit implanted in a scalp and the external powering circuit disposed in a hat in accordance with the invention.

DETAILED DESCRIPTION OF THE INVENTION

[0048] A device for deep brain stimulation using RF energy harvesting **200** of he invention is shown implanted under a human scalp in **FIG. 2A**. A flexible, implantable disc-shaped portion **210** having a diameter of about 6 cm and a thickness of between 3 and 4 mm may be formed of a biocompatible material and include circuitry as further described herein. Lead wires **220** may lead from the circuitry and be coupled to electrodes **230** disposed in targeted areas of the brain. Electrodes **230** may include conventional electrodes used for DBS. Neurostimulation lead securement devices **240** including burr hole caps may serve to secure the lead wires **220** to the electrodes. The circuitry may be operable to harvest and store RF energy, control the operation of the device **200** and provide neurostimulation pulses and signals to the targeted areas of the brain.

[0049] A device for cortical brain stimulation using RF energy harvesting **250** of the invention is shown in **FIG. 2B**. The flexible, implantable disc-shaped portion **210** is shown implanted under the scalp. Lead wires **270** may lead from the circuitry of the disc-shaped portion **260** and be coupled to electrodes **280** disposed on the cortical dura.

[0050] With reference to **FIGS. 2C and 2D**, lead securement devices **240** are shown. Lead securement devices **240** may include StimLoc devices available from ign. Lead securement devices **240** may minimize dislodgment of lead wires **220**.

[0051] Lead wires **220** may be coupled to the circuitry of the disc-shaped portion **210** by means of connectors **215**. Connectors **215** may include a plurality of male contacts **217** for providing electrical contact to corresponding female contacts of the circuitry (not shown). A screw hole **219** may be formed in the connector **215** for securing the connectors **215** to the disc-shaped portion **210** and for securing the disc-shaped portion **210** to the skull of the patient.

[0052] The circuitry may include a stimulation circuit **300** as shown in **FIG. 3** and a portion of the power circuit as shown in **FIG. 8A**. The stimulation circuit **300** may include a circuit printed onto the disc-shaped portion **210**. For purposes of illustration, the stimulation circuit **300** may be modeled using discreet components. The stimulation circuit **300** may include a PIC microcontroller **310** such as the PIC16LF87. The microcontroller **310** may manage the internal stimulation circuitry. A low frequency receiver chip **320** such as the ATA5283 may be coupled to the microcontroller **310** and may convert RF communications into programming commands which the microcontroller **310** interprets. An array of analog switches **330** such as the MAX4066 may be coupled to the microcontroller **310** and connect to voltage dividers **340** to output stimulation locations. Analog switches **330** may be coupled to electrodes **230** (**FIG. 2A**) and **260** (**FIG. 2B**).

[0053] According to internal parameters which can be modified via an external RF programming signal, the microcontroller **310** may control analog switch states to determine a voltage applied to any combination of four output locations including four output locations on electrodes **230** and **260**. The maximum possible voltage is determined by the supply voltage to the circuit **300**. A pulsing frequency, nominally 185 Hz, can be adjusted slightly as well as whether a stimulation pulse is applied or not. In order to conserve energy, the microcontroller **310** may enter a standby mode for 4 ms between pulses, greatly reducing power consumption. The microcontroller **310** may be operated with an internal clock frequency of 125 KHz, giving an efficient tradeoff between power conservation and proper functionality. This clock frequency allows pulse durations in increments of 32 micro-seconds. The output pulse duration can be adjusted between ~60 and ~180 microseconds. With reference to **FIG. 4**, **FIG. 5** and **FIG. 6**, the frequency output, varied voltage output and pulse duration of the microcontroller **310** are shown respectively.

[0054] Every pulsing cycle, the programming input from the low frequency receiver chip **320** may be checked. If a programming signal is present, an input code may be read

sequentially and the specified parameter adjusted to a new value, after which the program continues its pulsing routine.

[0055] The low frequency receiver chip **320** used for receiving external programming commands uses an amplitude shift keying (ASK) protocol. The state of a 125 KHz signal being received determines the output voltage of the low frequency receiver chip **320**: on-high, off-low. While waiting for a signal, the low frequency receiver chip **320** may remain in standby mode, conserving power. Upon the presence of a programming signal, the low frequency receiver chip **320** may wake up and send the coded data to the microcontroller **310**, after which the microcontroller **310** may tell the low frequency receiver chip **320** to enter standby mode again. The programming signal may include a preliminary “on” time to wake up the low frequency receiver chip **320**, a 4-bit header, a 3-bit parameter identifier, and a 4-bit data value. Each bit time is 2 milliseconds, allowing enough time for the microcontroller **310** to process the bit reception before the next bit arrives. An antenna attached to a coil input of the low frequency receiver chip **320** may be a short wire having a strong programming signal.

[0056] Eight analog switches **330** may be used to control the output pulsing. Four switches **330** may determine a path of the selected voltage to the four possible output locations. Each of these may be controlled by one of the microcontroller outputs, which are in turn enabled or disabled depending on the internal variable for output locations. The inputs of the four switches **330** may be attached to the outputs of the other four switches **330**. The inputs of these four switches **330** may all be attached to different voltage dividers **340**, providing four different voltage levels, ranging from three quarters of the supply voltage to the supply voltage maximum of 3V. Each switch **330** may be controlled by an individual microcontroller signal, which also drives the voltage divider **340** for its particular switch **330**. For every pulse, only one of these four microcontroller signals is active, enabling the voltage from its divider **340** to be sent to the output switches **330** and ultimately to the electrodes **230** and **260**. The use of static voltage dividers **340** to provide output voltage scaling may minimize power consumption. In an alternative embodiment of the invention, a custom digital-analog converter could be used to allow for a higher range of stimulation voltages.

[0057] When tested for power consumption, a ~1 Ω resistor was put in series with a powering circuitry described herein. The voltage measured across the resistor while in operation was approximately 17 µV, implying that the DC current required is ~17 µA. At a supply voltage of 3 V, this equates to a power consumption of 51 µW. If operated for 24 hours, the implant would consume a little over 4.4 Joules/day. Typical parameters of a stimulation signal provided for Parkinson’s disease are a series of pulses of 120 microsecond duration, 2.5 volts in strength at a repetition rate of 185 pulses per second. Assuming these typical parameters, there are:

185 pulses/second*60 seconds/minute*60 minutes/hour*24 hours/day=1.5984*10^7 pulses/day.

With pulse duration of 120 micro seconds, this gives a total energy application duration of 1.5984*10^7*120*$10^{6=1918.08}$ seconds. With 2.5 volts and 50 micro amps=120 micro watts, the total energy required for stimulation is 120*10^6 watts*1918.08 seconds=0.2302 joules per day. As disclosed herein, energy harvesting by the power circuit is on the order of 12-15 joules per day and the stimulation energy required is more than adequately provided by the power circuit.

[0058] An external programmer circuit **700** may include a microcontroller **710** including a PIC16LF87, an inductor/capacitor (LC) oscillating circuit **720** (125 KHz), and an intermediate MOSFET driver **730** including a TC4422 as shown in **FIG. 7**. The MOSFET driver **730** may supply enough energy for driving the LC circuit **720**. When a programming signal is to be sent, a button (not shown) may be pressed, telling the microcontroller **710** to read its inputs and stimulate the MOSFET driver **730** to oscillate the LC circuit **720** according to a communication protocol. Input voltages may be controlled by simple switches. Four switches may dictate the value to be sent, while five switches may dictate which parameter is to be changed. Only one of these switches should be on at one time. A Phidget RFID antenna **740** designed for 125 KHz may be attached to the high voltage side of a capacitor **750** of the LC circuit **720** for sending the programming signal. The circuit **700** may be powered via a 12-Volt wall supply. The 12 V drives the MOSFET driver **730** and is regulated to 5 V for the switches and microcontroller **710**.

[0059] An external powering circuit **800** may include a battery **810** for powering an oscillator **820** which drives a transformer-like setup **830** as shown in **FIG. 8A**. The coils **835** on one side of the transformer **830** may be disposed in a cap worn on the head of a patient, a headband worn on the head of the patient, or on a headboard of a bed in which the patient lies. The coils **840** on the other side of the transformer **830** may be coupled to the stimulation circuit **300** and may be disposed proximate the coils **835**. An AC signal coming from coil **840** may be amplified and rectified through a charge pump **850** having three stages, after which a voltage may be clamped with a regulator **860** to prevent spiking. A control circuit **870** may control operation of the voltage regulator **860**.

[0060] The oscillator **820** may include an LTC6900. This oscillator **820** produces a 50% duty cycle square wave to drive the primary coil **835** of the transformer **830** and requires only a potentiometer for adjusting the frequency. The charge pump **850** may be a Cockroft Walton voltage multiplier, utilizing a ladder of diodes and capacitors to rectify and amplify the signal. The amplification depends on the number of stages used. Three stages have been found to be enough for a substantial voltage multiplication across a load of 200 KΩ. The capacitors may be 0.1-µF each and the diodes may include BAT54SW surface mount diodes with a forward voltage drop of ~0.24-V. The regulator **860** may include an LT1521-3, which clamps a higher input voltage to 3 V.

[0061] Previous empirical testing showed square coils (both primary **835** and secondary **840**), 1 in.×1 in., with 5 turns each are effective for transferring enough energy to power the stimulation circuit **300**. Coils **840** are shown in PCB layout in **FIG. 9** and coils **835** are shown in PCB layout in **FIG. 10**.

[0062] The optimal frequency depends on the dielectric and distance between coils **835** and **840**. Frequencies in the range of 2 MHz to 15 MHz may be used. The oscillator **820** can be powered with 3 AAA batteries (4.5 V). In examining

the actual signal through the primary coil **835**, the voltage waveforms in **FIG. 11** and **FIG. 12** were obtained. **FIG. 11** shows the output voltage from the oscillator **820**. **FIG. 12** shows the voltage across the primary coil series resistance, from which the RMS current is calculated to be 29.36 mA_{RMS}.

[**0063**] The embodiment described above provides for near field harvesting and includes inductive coupling between coils **835** and **840**. With reference to **FIG. 8B** and in an alternative embodiment of the invention, the power circuit **865** for powering the stimulation circuit **300** may be non-inductively coupled to an external source of RF energy **880**. In this far field embodiment, the power circuit **865** may be disposed in a wrist band worn by the patient, in a room transmitter or in a transmitter disposed in a building occupied by the patient. In yet another alternative embodiment of the invention, the power circuit **865** may harvest ambient RF energy such as energy transmitted in space by using an inherently tuned antenna as described in U.S. Pat. No. 6,856,291, the description of which is incorporated by reference in its entirety herein. Furthermore, a rechargeable battery or other storage device (not shown) may be employed to store harvested energy. "Non-inductive" as described herein being directed RF.

[**0064**] To demonstrate the effectiveness of the powering and programming schemes through tissue, the device **200** was tested through swine skin. Clear tape was used to cover the conductive surfaces on the primary coil **835** and the secondary coil **840** to prevent interaction with the moisture on the skin. This tape had negligible effect on the inductive coupling.

[**0065**] Three different tests were performed, each following the same procedure. At a certain separation, the voltage powering the primary coil oscillator **820** was adjusted and the maximum output voltage from the secondary coil voltage regulator **860** was measured. The primary coil oscillator voltage started at 5 V and was decreased in increments of 0.1 V until the maximum regulator output voltage had reached a steady minimum.

[**0066**] The first test was performed with no skin between the transformer coils **835** and **840**. Data was acquired at separations of 5 mm, 7 mm, and 10 mm, values chosen based on the common range of human scalp thickness. The second test used fresh swine skin of thicknesses 5 mm and 7 mm between the transformer coils **835** and **840**. The test was interrupted, preventing the testing of 10 mm thick skin. The third test used the same pieces of swine skin, 5, 7, and 10 mm thick, after they had been frozen and thawed.

[**0067**] **FIG. 13** shows the results from the first two tests for comparison purposes. The presence of the skin reduced the inductive coupling between the coils, and hence the possible maximum output voltage in the range of 1.5-3.8 V. At 4.0 V and above, the maximum output voltage of ~3 V is obtainable even with the presence of the skin.

[**0068**] The same pieces of skin were tested a second time due to interruption of the first test. However, they had all been frozen and thawed in the interim, affecting the results slightly. **FIG. 14** shows the effect that the freezing and thawing of the skin had on the energy transfer of the transformer coils. Both the 7- and 10-mm thick pieces of skin reduced the inductive coupling, but the 5-mm skin actually improved in performance. This may be due to the presence of a layer of fat in both the 7- and 10-mm pieces that is absent in the 5-mm piece.

[**0069**] Another test was performed to find the effect of the skin over time. The stimulus for this test was the degrading performance of the 10-mm thick skin over time during the interrupted test mentioned above. For this test, the 7-mm thick piece of skin was used between the primary and secondary coil. The frequency was adjusted to produce a maximum output voltage, which was measured successively over a period of time. The results shown in **FIG. 15** support the fact that performance does not degrade over time. The slight drop in output voltage is likely due to the mechanical nature of the frequency-tuning potentiometer. Notice that the output voltage reaches a steady value and remains constant after that point.

[**0070**] In order to demonstrate the concept and functionality of the device **200**, a model **1600** was created as shown in **FIG. 16**. The stimulation circuit **300** was put on top of a Styrofoam head **1610** with wires running down through the bottom for power- and pulse-monitoring purposes. An ABS Plastic cap (not shown) was made to simulate the head's scalp, covering the stimulation circuit **300**. The primary powering coil **835** with the batteries **810** was secured in a hat **1620** over the position of the stimulation circuit **300** to provide for near field inductive coupling.

[**0071**] The device **200** was tested on a cadaver head to show that a signal may be generated through the scalp and to demonstrate the programmability of the device **200** during stimulation. First, an incision was made in the scalp of the cadaver head. The secondary coil **840** was inserted and placed on the skull and the incision was sewn, leaving the lead wires **220** of the circuit exposed. Six wires were used on the implanted circuitry, four representing the electrodes **230**, which were connected to an oscilloscope and two wires for power and ground. The primary coil **835** was then taped on the scalp directly on top of the implanted circuitry and connected to a power supply.

[**0072**] The experiment began by demonstrating the programmability of the stimulation circuit **300**. Four parameters were varied and displayed on the oscilloscope; pulse width (60, 120 and 180 micro seconds), amplitude (2.34V, 2.75V, 2.94V, 3.13 V), frequency (191 and 194 Hz) and the shifting from one stimulating probe to another (i.e. probe **1** to probe **2** or probe **1** to all four probes). A 10K OHM resistor was used to represent the brain resistance, however this resistance is higher than the resistance for the brain (900 to 1100 Ohms), but a 10 k Ohm resistor was used to ensure there was enough power.

[**0073**] Next, several voltages were tested to determine the output source voltage of the power circuit **865**. Initially, the power supply connected to the primary coil **835** was set at 5 V and was decremented by 0.1 V to 1.2 V. The voltage on the secondary coil **840** was clamped so as not to exceed 3V. As the voltage decreased on the power supply, the output voltage on the secondary coil **840** was steady at 3 V until it declined around 2.2 V. Once the voltage decreased from 3V, a potentiometer was adjusted to obtain the maximum voltage. The data obtained shows the when the voltage drops, the amplitude voltage and frequency drop off as well.

[**0074**] The device for brain stimulation using RF harvesting of the present invention provides a brain stimulation device that requires a single implantation site and surgery to thereby reduce both the cost and trauma to the patient of the implantation procedure. The brain stimulation device further uses RF energy as a power source to eliminate the need for a battery implanted in the pectoral area of the patient. The brain stimulation device further converts RF energy and

stores the converted RF energy for use in stimulation of targeted brain areas. The brain stimulation device is flexible and implantable under the scalp to minimize discomfort for the patient.

[0075] The foregoing description of the embodiments of the invention has been presented for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed. Many modifications and variations are possible in light of the above teaching. It is intended that the scope of the invention be limited not by this detailed description, but rather by the claims appended hereto.

TABLE 1

Indication	Prevalence	Reference for Prevalence
	FDA Approved Indications	
Essential tremor (1997)	1.5–3% population	Am J Med 115:134–42, 2003
Parkinson's (2002)	1% of population > 50	
Dystonia (2004)	~150,000 in US	Mov Disord 3:188–94, 1988
	Current Clinical Trials	
Obsessive-compulsive disorder	2–3% of population	J Clin Psych 53 Suppl: 4–10, 1992
Tourette's syndrome	~1–2% children	J Psychosom Res. 55:3–6, 2003; Can J Neurol Sci Suppl 1:S64–71, 2003
Intractable epilepsy	~100,000 in US	Neurology 56: 1445–52, 2001; Rev Neurol (Paris); 160 Spec No 1:5S31–5, 2004
Intractable depression	1,000,000	Psychiatr Clin North Am 19:179–200, 1996; http://www.mhsource.com/depconsult/june 2004.jhtml?_requestid = 605984

We claim:

1. A system for brain stimulation using radio frequency harvesting comprising:

a device implantable under a scalp of a patient, the device comprising a radio frequency harvesting power circuit and a stimulation circuit; and

at least one electrode coupled to the stimulation circuit, the at least one electrode providing brain stimulation to targeted areas of the brain.

2. The system of claim 1, wherein the device is fabricated from a biocompatible substrate.

3. The system of claim 1, wherein the device is flexible and conformable to a shape of the scalp.

4. The system of claim 1, wherein the power circuit comprises a charge pump inductively coupled to a primary coil of an external power circuit.

5. The system of claim 1, wherein the power circuit comprises an inherently tuned antenna for harvesting energy transmitted in space.

6. The system of claim 1, wherein the at least one electrode provides deep brain stimulation.

7. The system of claim 1, wherein the at least one electrode provides cortical stimulation.

8. The system of claim 1, further comprising a programming circuit operable to control the stimulation circuit.

9. The system of claim 8, wherein the programming circuit is operable to control a stimulation circuit voltage output.

10. The system of claim 8, wherein the programming circuit is operable to control a stimulation circuit output pulse width.

11. The system of claim 8, wherein the programming circuit is operable to control a stimulation circuit output frequency.

12. The system of claim 1, further comprising an energy storage device coupled to the device.

13. A system for brain stimulation comprising:

a device implantable under a scalp of a patient, the device comprising a coupled power circuit and a stimulation circuit; and

at least one electrode coupled to the stimulation circuit, the at least one electrode providing brain stimulation to targeted areas of the brain.

14. The system of claim 13, wherein the coupled power circuit is inductively coupled.

15. The system of claim 13, wherein the coupled power circuit is non-inductively coupled.

16. A system for brain stimulation comprising:

a device implantable under a scalp of a patient, the device comprising a power circuit powered by ambient radio frequency energy and a stimulation circuit; and

at least one electrode coupled to the stimulation circuit, the at least one electrode providing brain stimulation to targeted areas of the brain.

17. The system of claim 1, wherein the at least one electrode provides brain stimulation to targeted areas of the brain in response to signals received from the stimulation circuit.

18. A method of providing brain stimulation, comprising:

harvesting power in a power harvesting circuit in a device implantable under the scalp of a patient; and

providing brain stimulation to targeted areas of the brain with at least one electrode connected to the power harvesting circuit.

19. The method of claim 18, wherein the power harvesting circuit harvests radio frequency energy.

20. The method of claim 18, wherein the power harvesting circuit harvests energy by an inductively coupled power circuit in the device.

21. The method of claim 18, wherein the power harvesting circuit harvests energy by a non-inductively coupled power circuit in the device.

22. The method of claim 18, wherein the brain stimulation is used to treat Parkinson's disease.

23. The method of claim 18, wherein the brain stimulation is used to treat stroke patients.

* * * * *

US 20060281435A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2006/0281435 A1**

Shearer et al. (43) **Pub. Date: Dec. 14, 2006**

(54) **POWERING DEVICES USING RF ENERGY HARVESTING**

(75) Inventors: **John G. Shearer**, Ligonier, PA (US); **Charles E. Greene**, Pittsburgh, PA (US); **Daniel W. Harrist**, Carnegie, PA (US)

Correspondence Address:
Ansel M. Schwartz
Attorney at Law
Suite 304
201 N. Craig Street
Pittsburgh, PA 15213 (US)

(73) Assignee: **FireFly Power Technologies, Inc.**

(21) Appl. No.: **11/447,412**

(22) Filed: **Jun. 6, 2006**

Related U.S. Application Data

(60) Provisional application No. 60/688,587, filed on Jun. 8, 2005.

Publication Classification

(51) **Int. Cl.**
H04B 1/16 (2006.01)

(52) **U.S. Cl.** **455/343.1**

(57) **ABSTRACT**

Disclosed is an apparatus for an application including a core device for the application. The apparatus includes a power (preferably RF energy) harvester connected to the core device to power the core device. Also disclosed is a method for an application. The method includes the steps of converting RF energy into usable energy. There is the step of powering the core device with the usable energy.

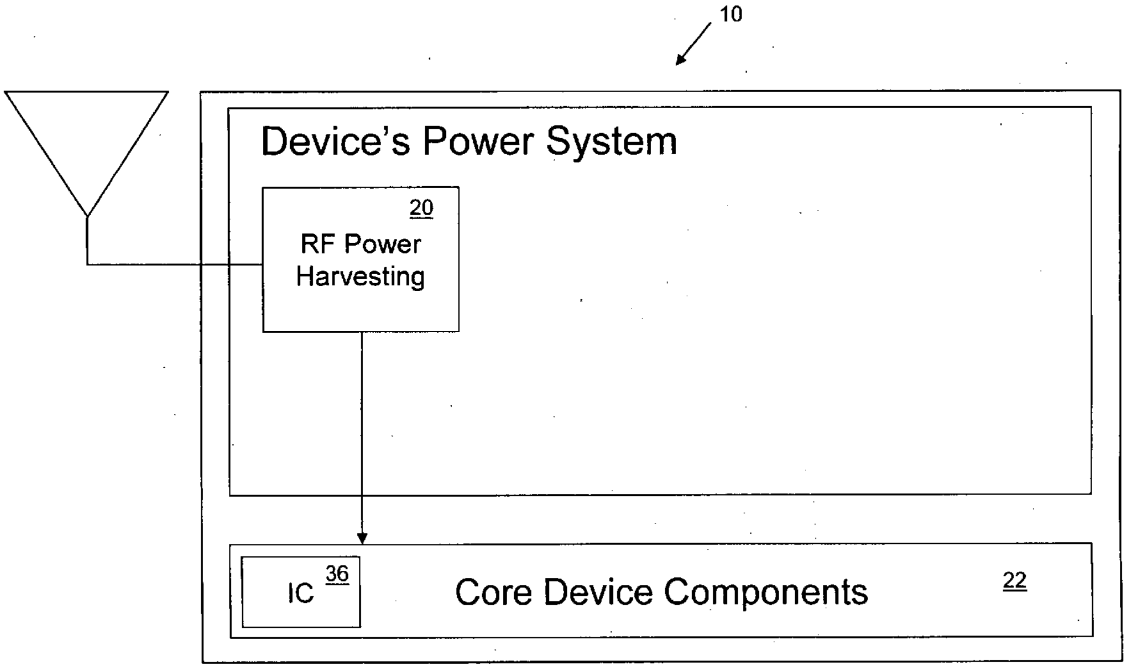

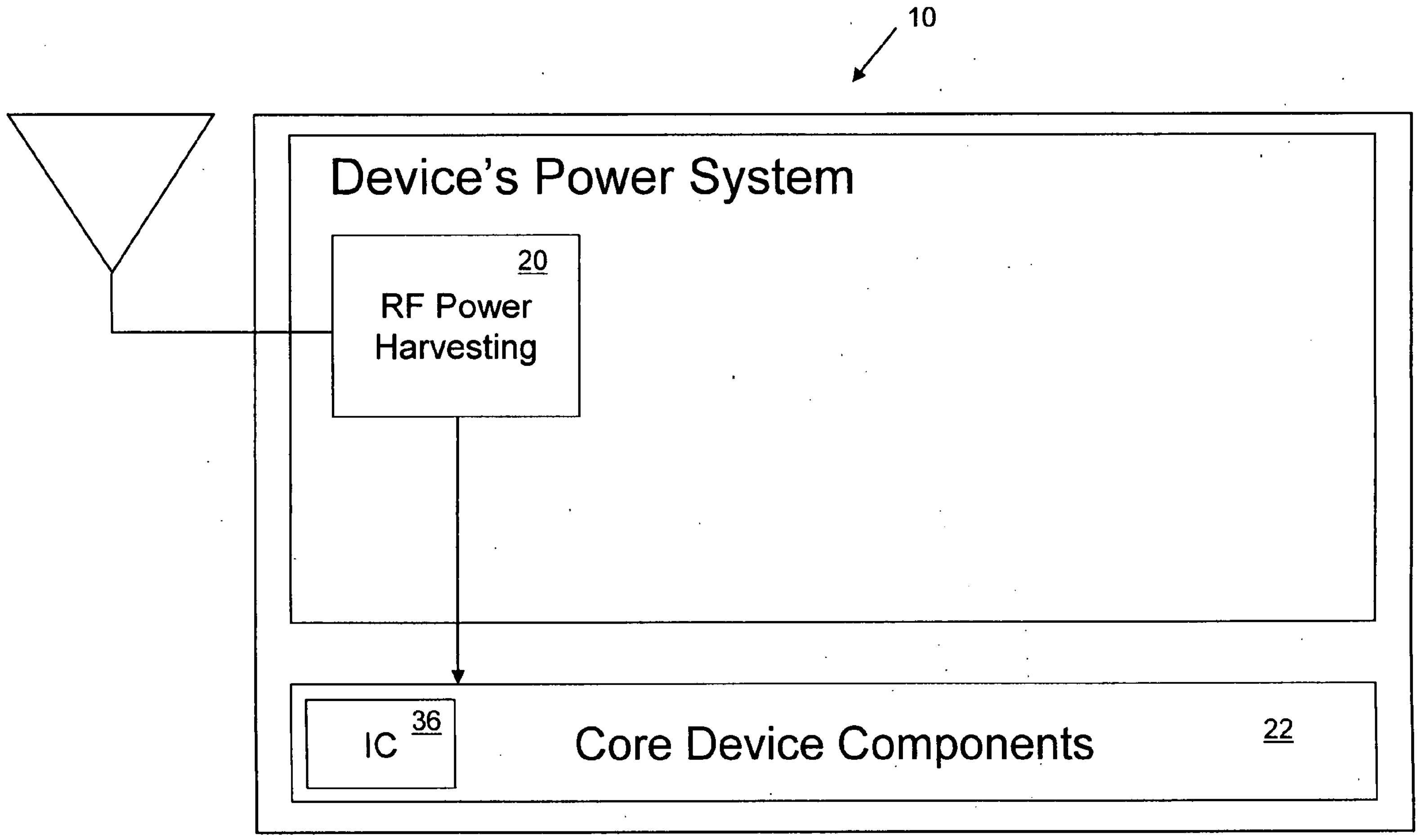
10
Device's Power System
20
RF Power
Harvesting
IC
36
Core Device Components
22

Fig. 1

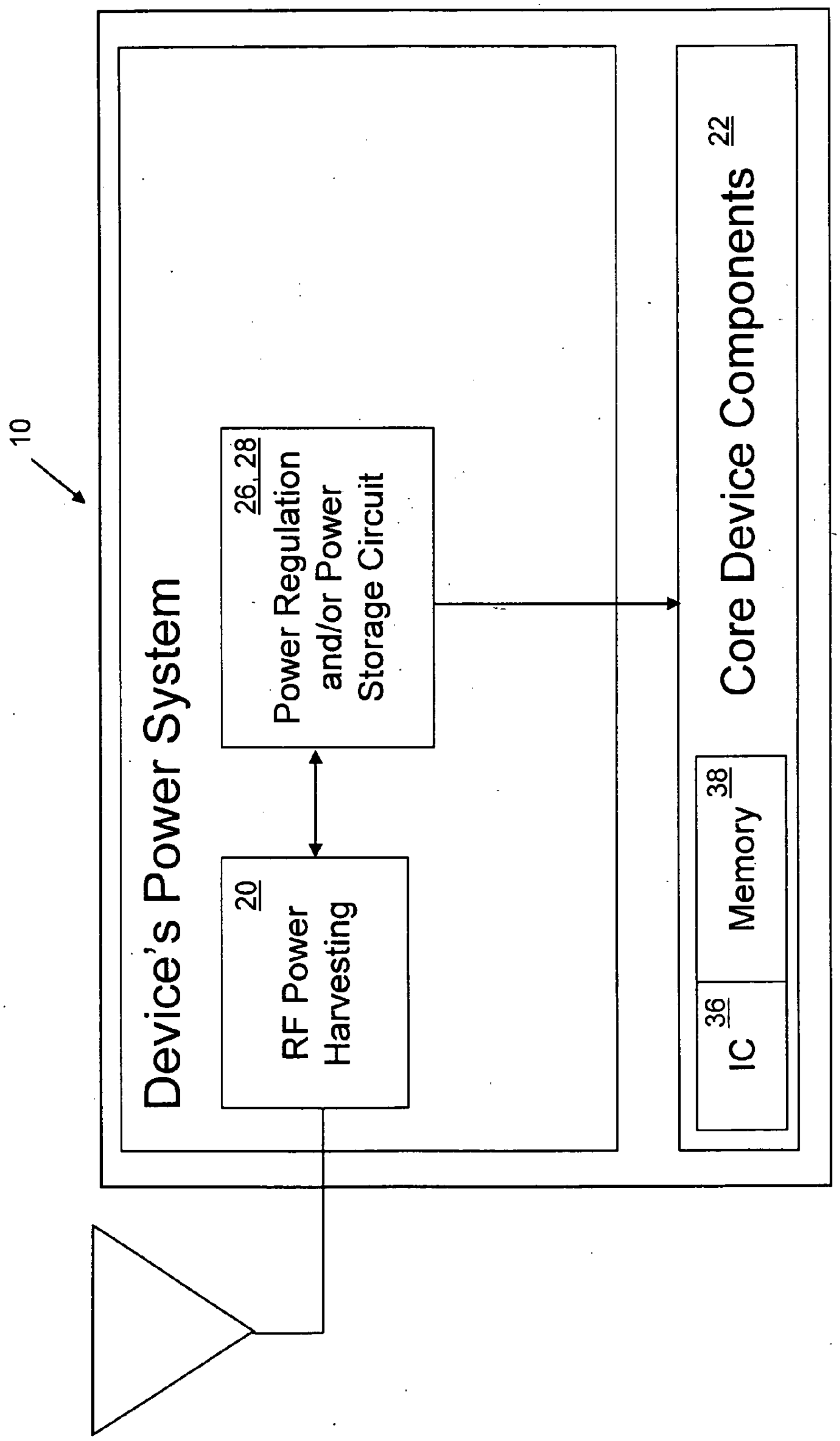
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Core Device Components 22
IC 36
Memory 38

Fig. 2

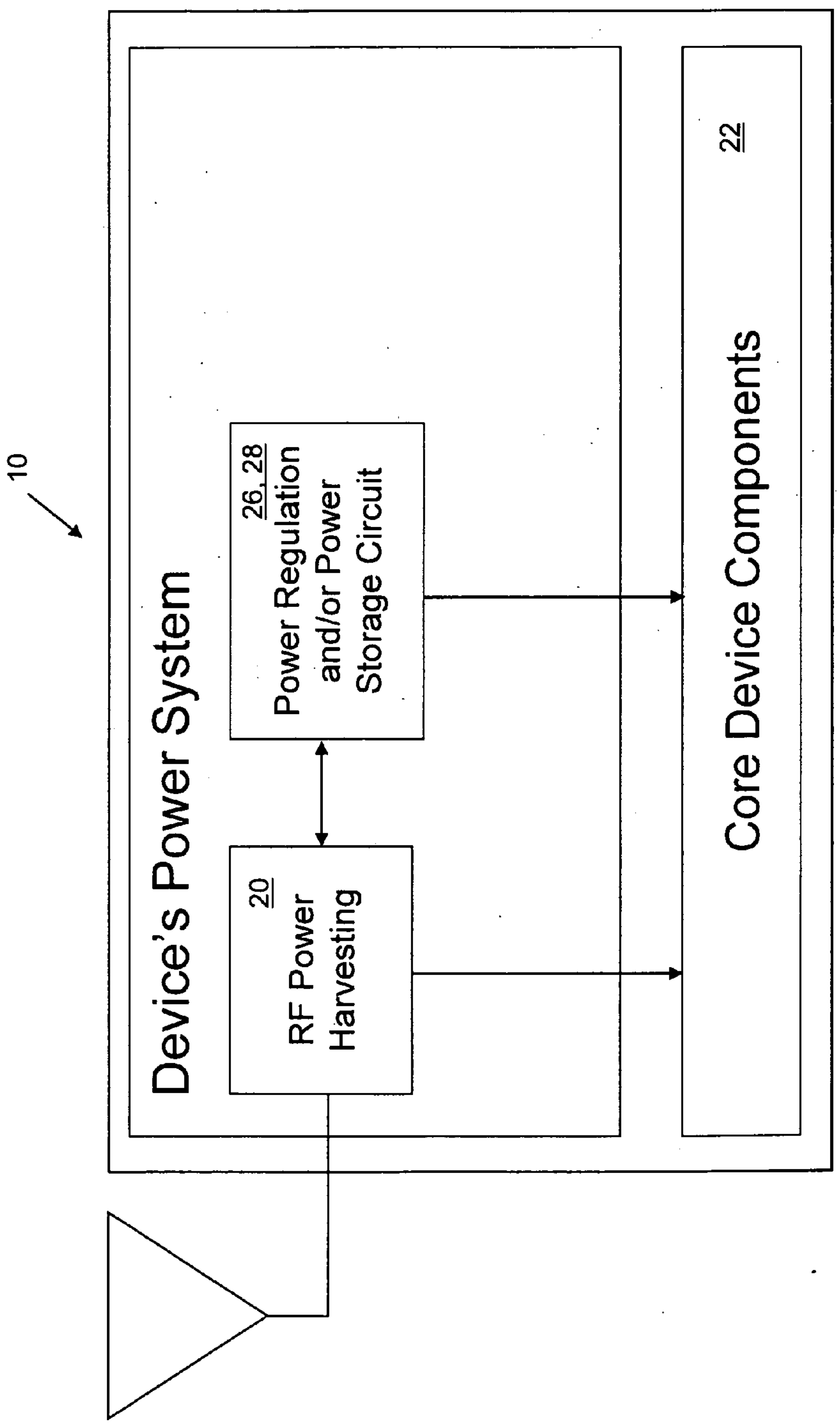
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Core Device Components
22

Fig. 3

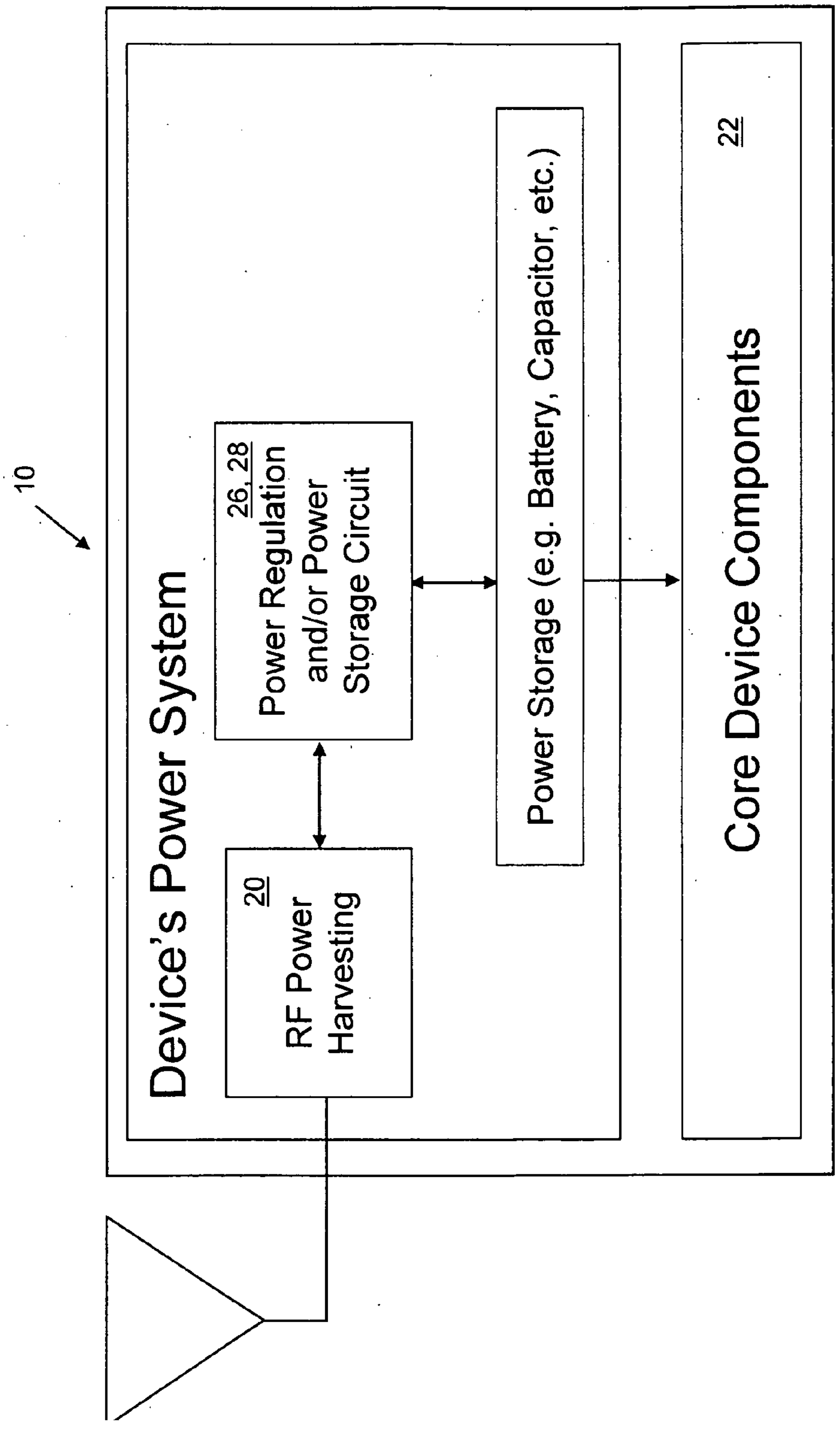
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 4

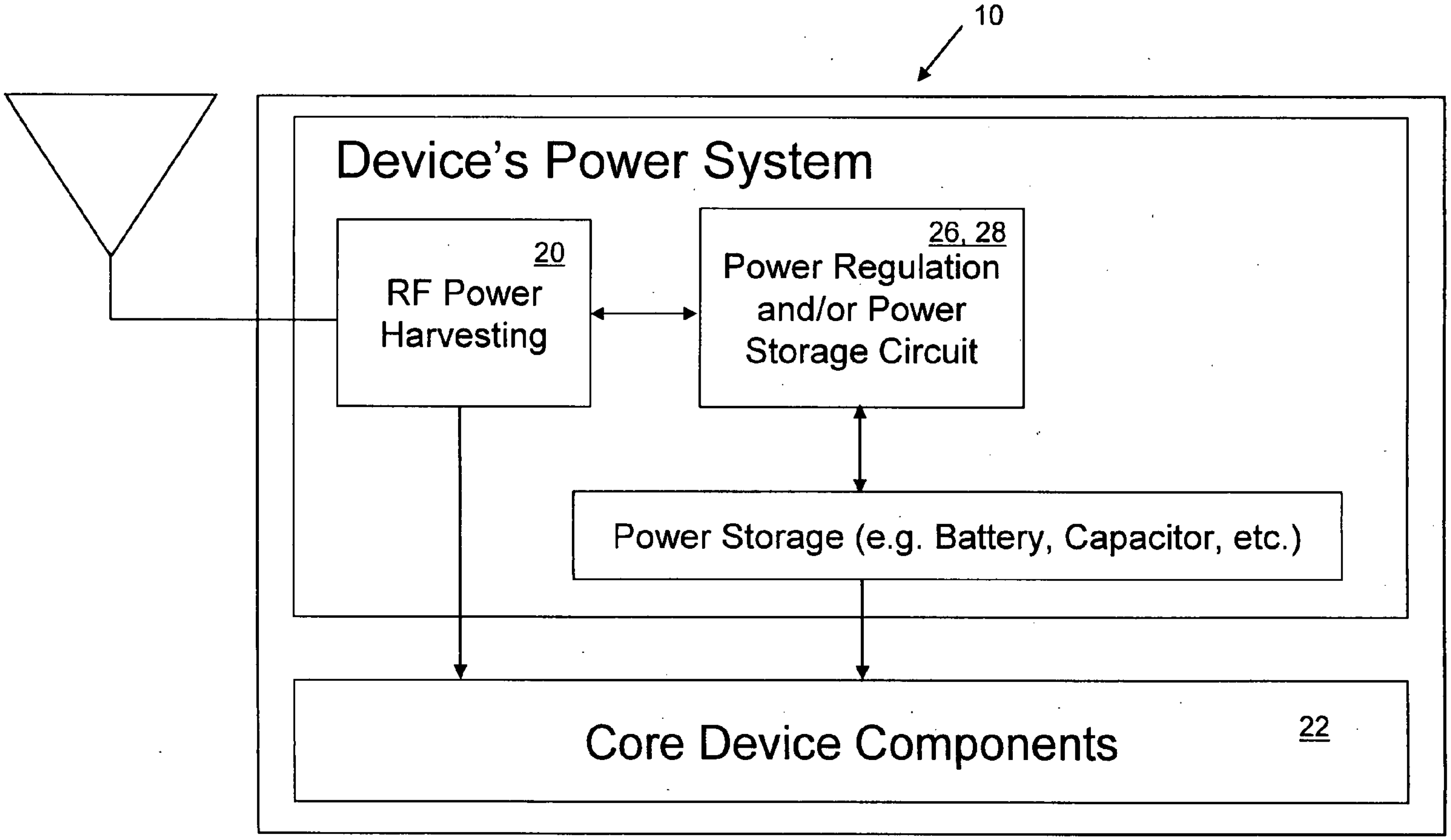
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 5

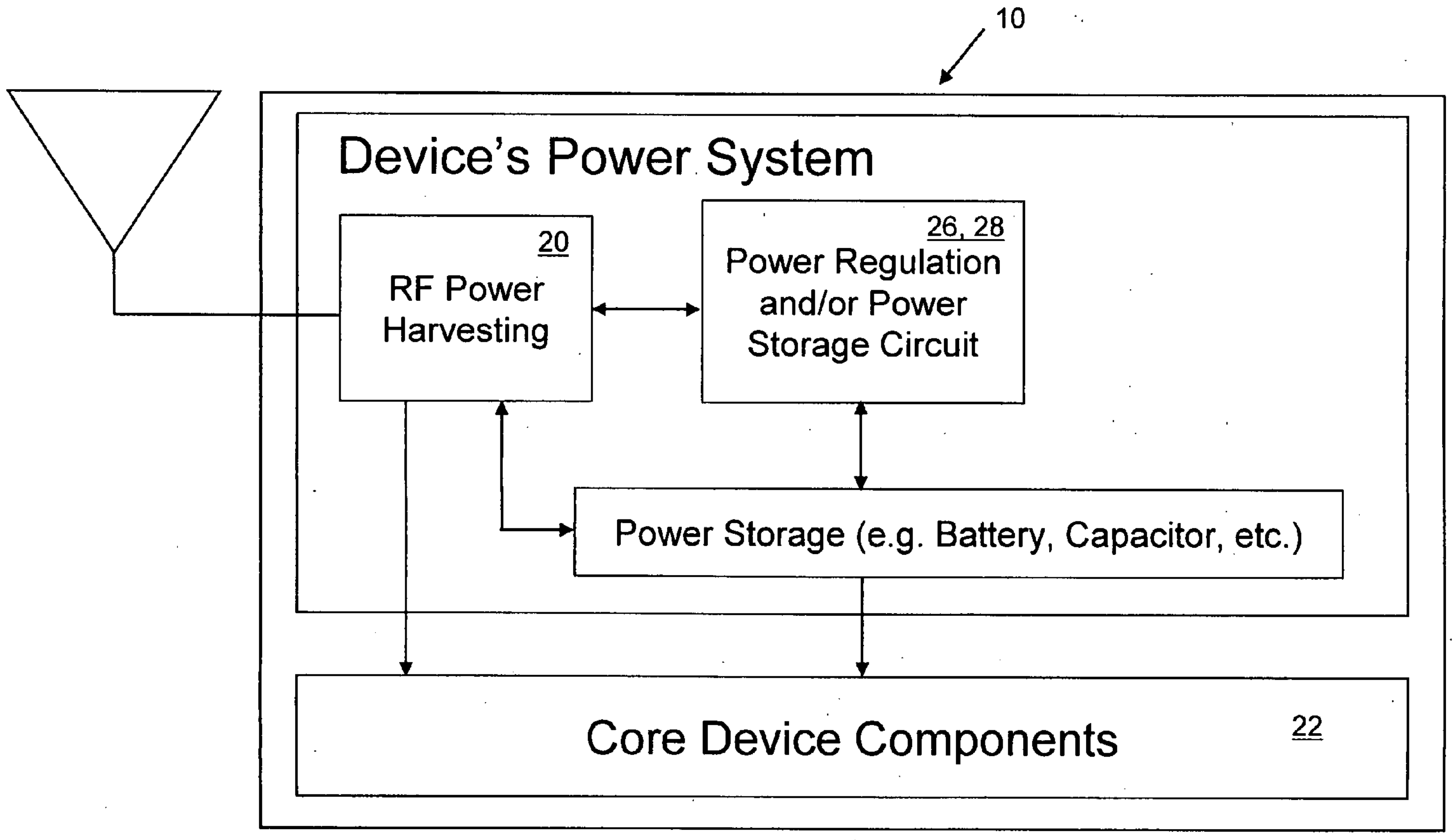
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 6

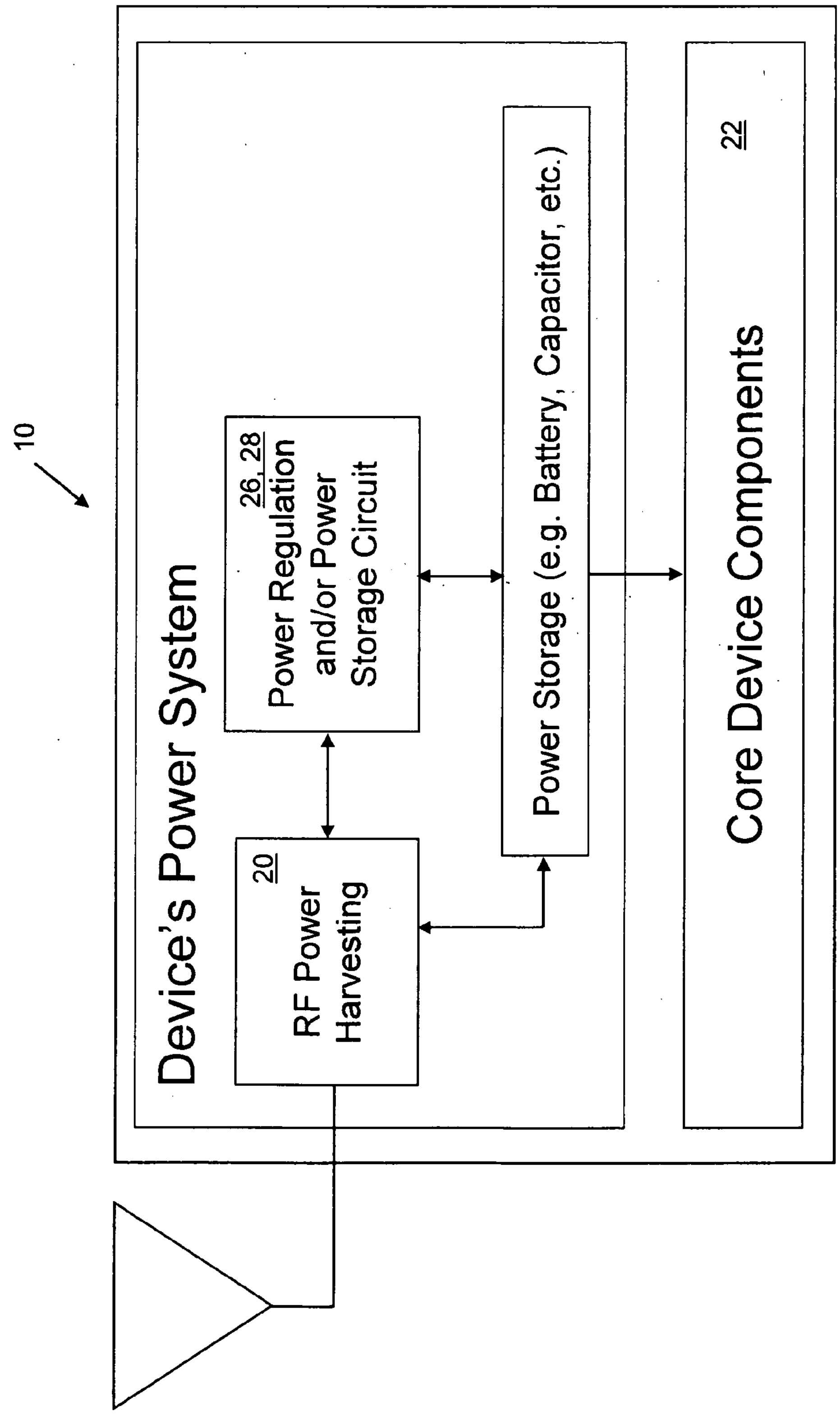
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 7

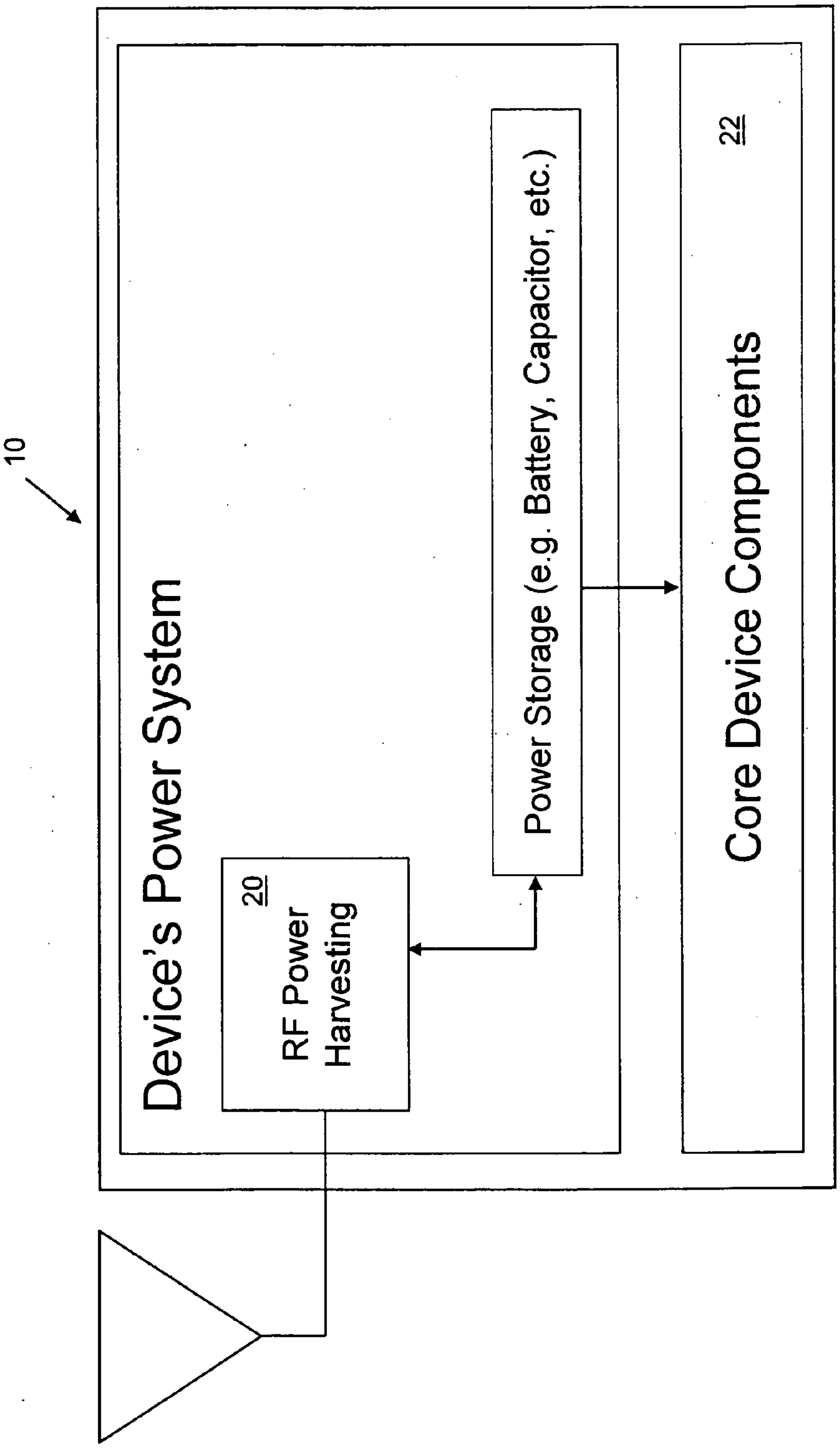
10
Device's Power System
20
RF Power Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 8

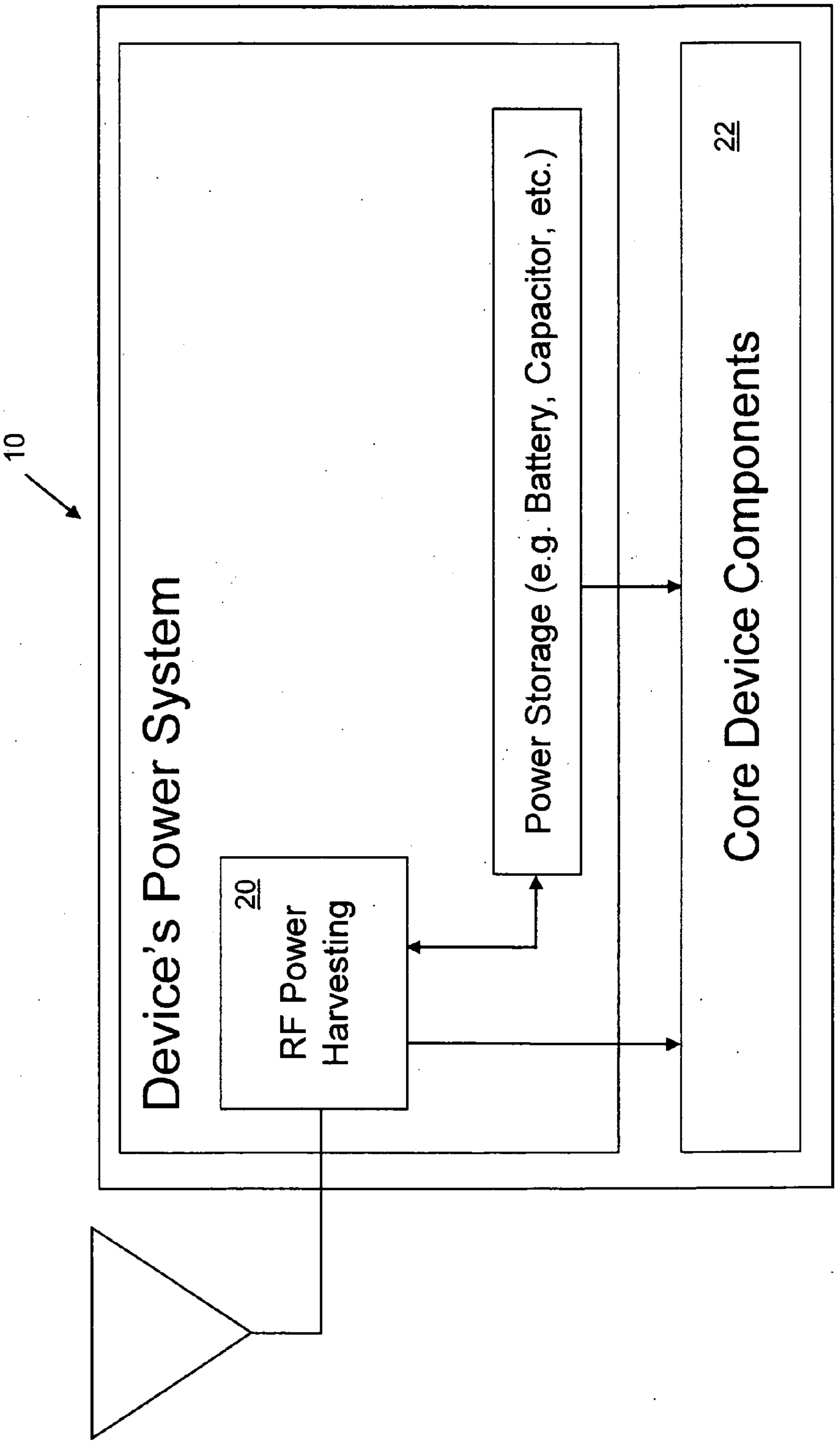
10
Device's Power System
20
RF Power Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 9

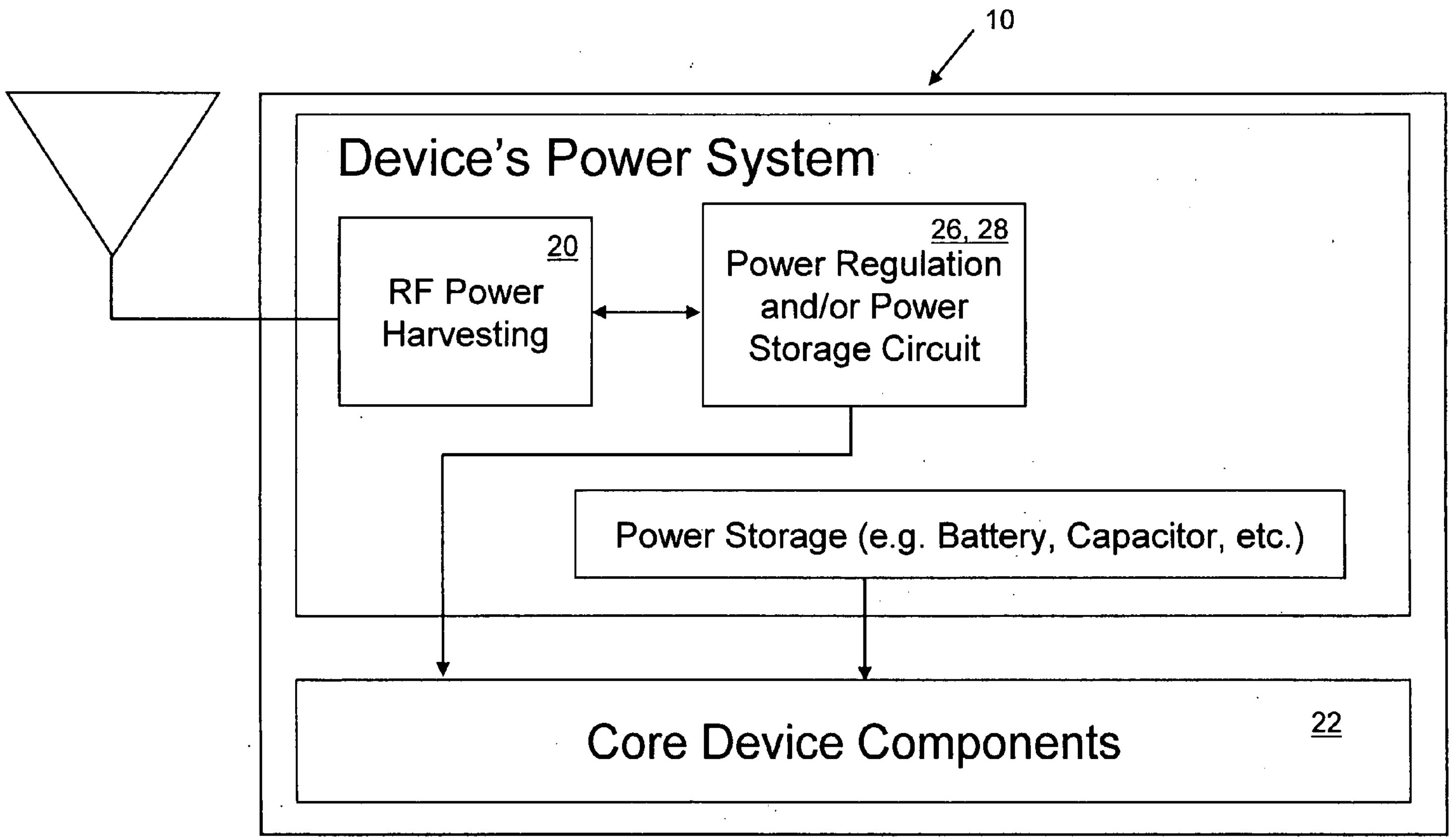
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 10

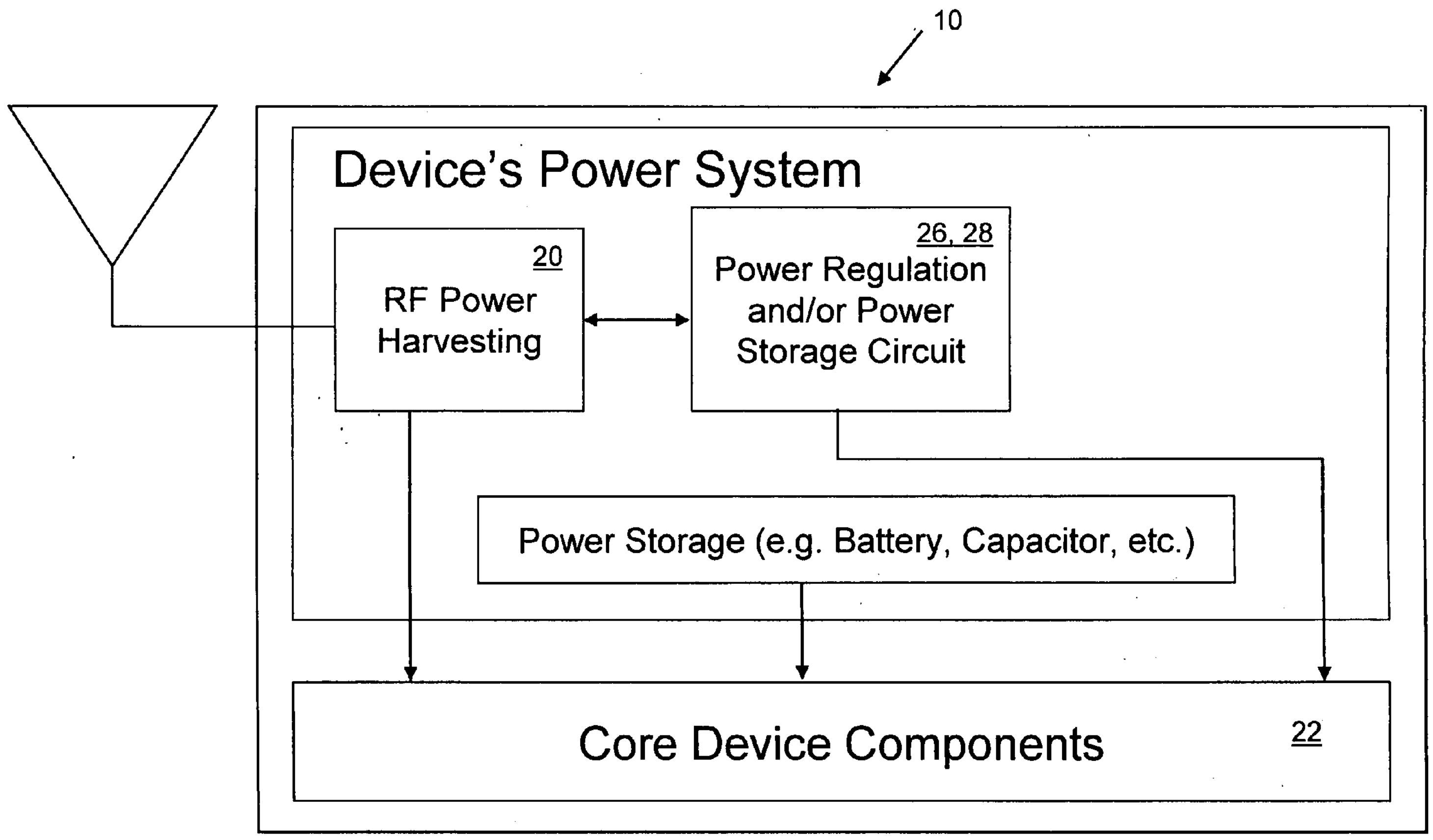
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 11

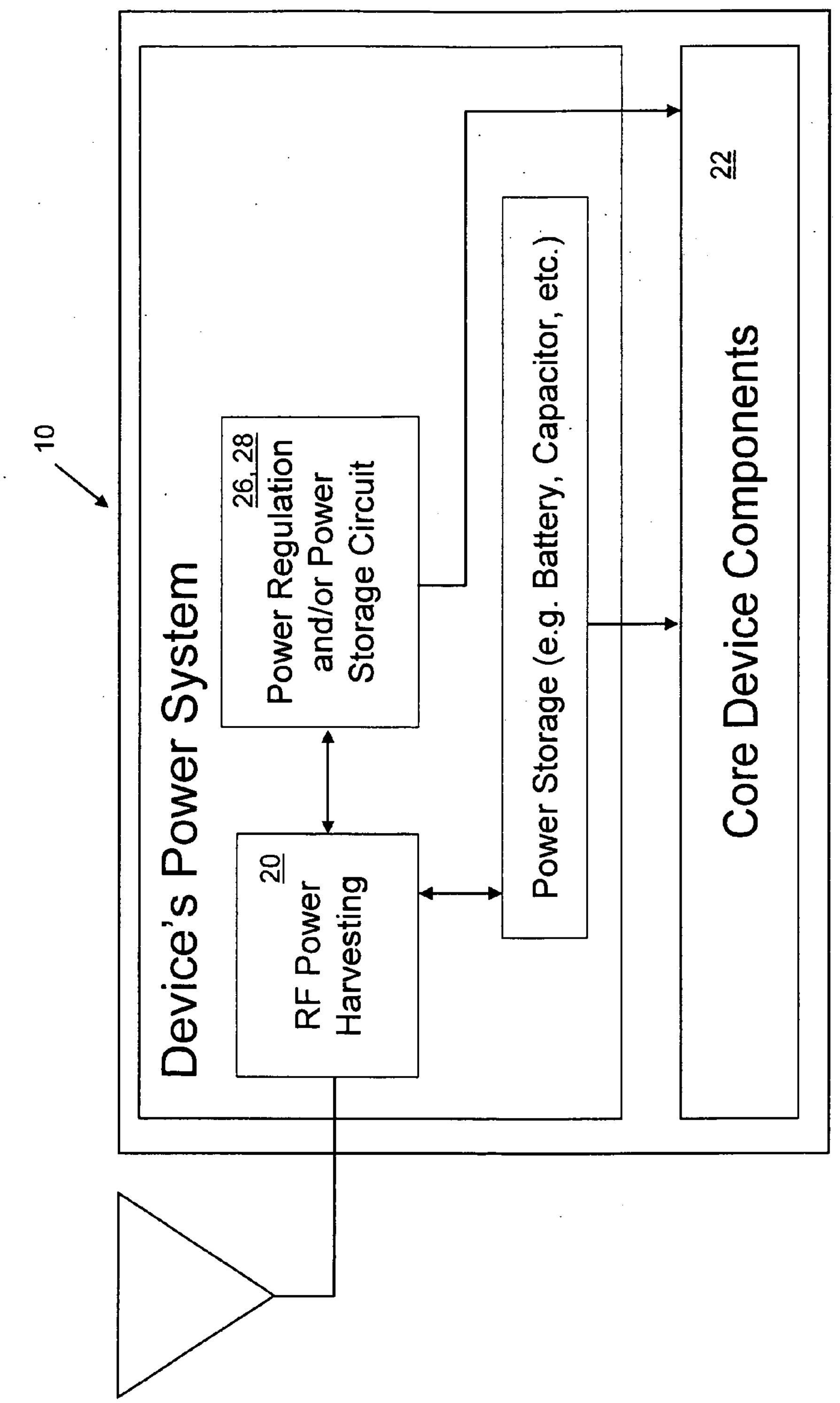
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 12

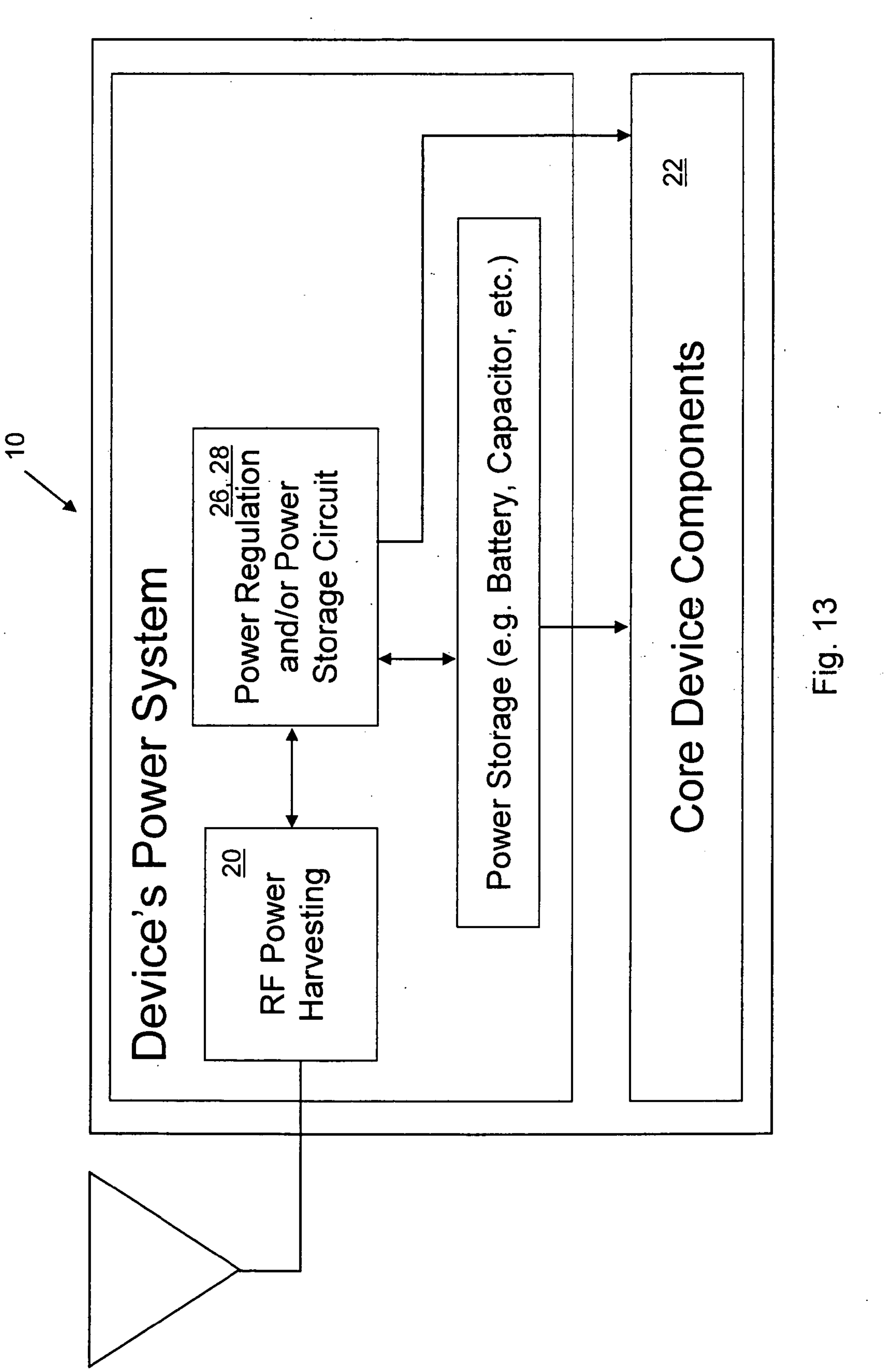
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 13

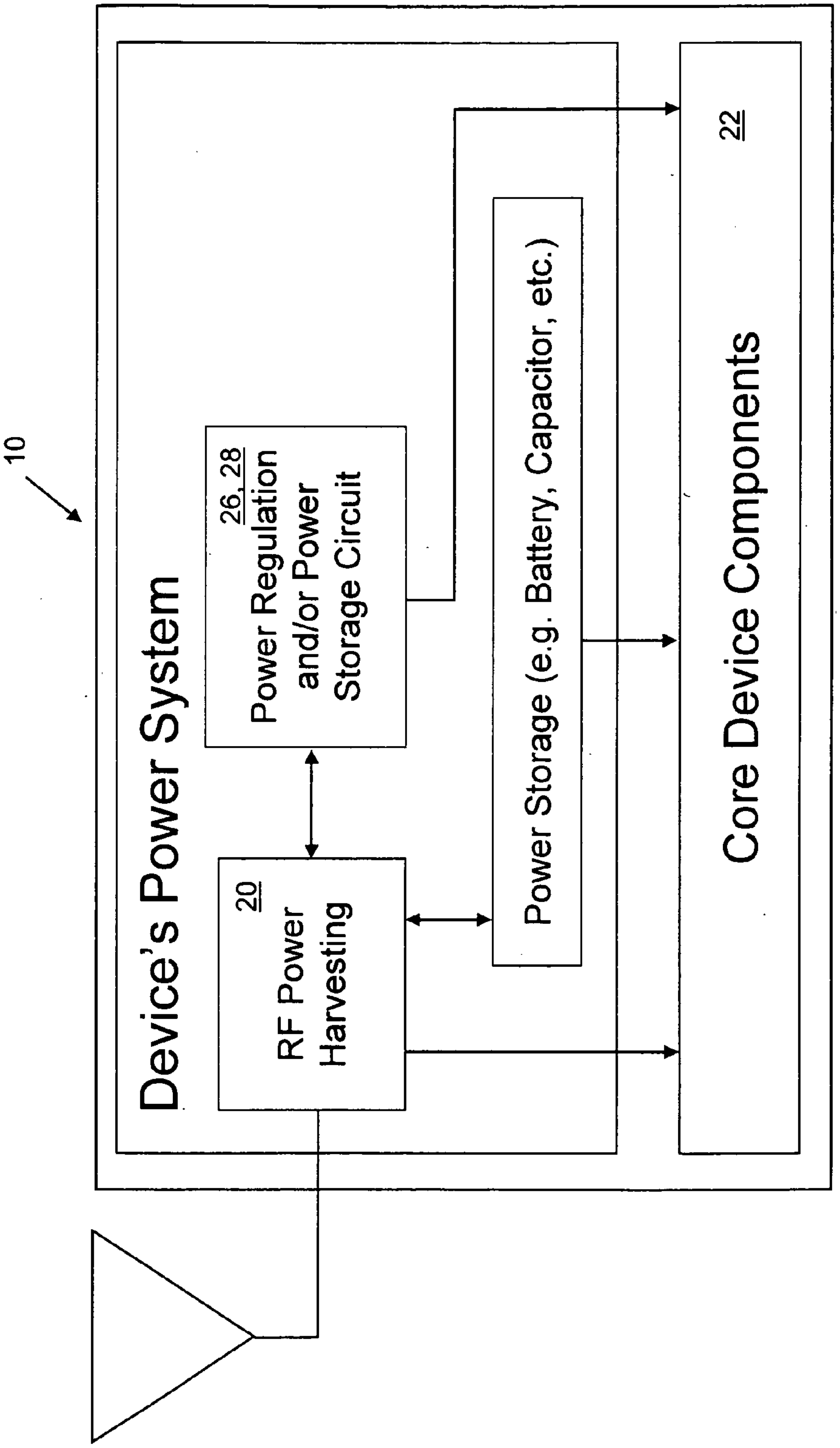
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 14

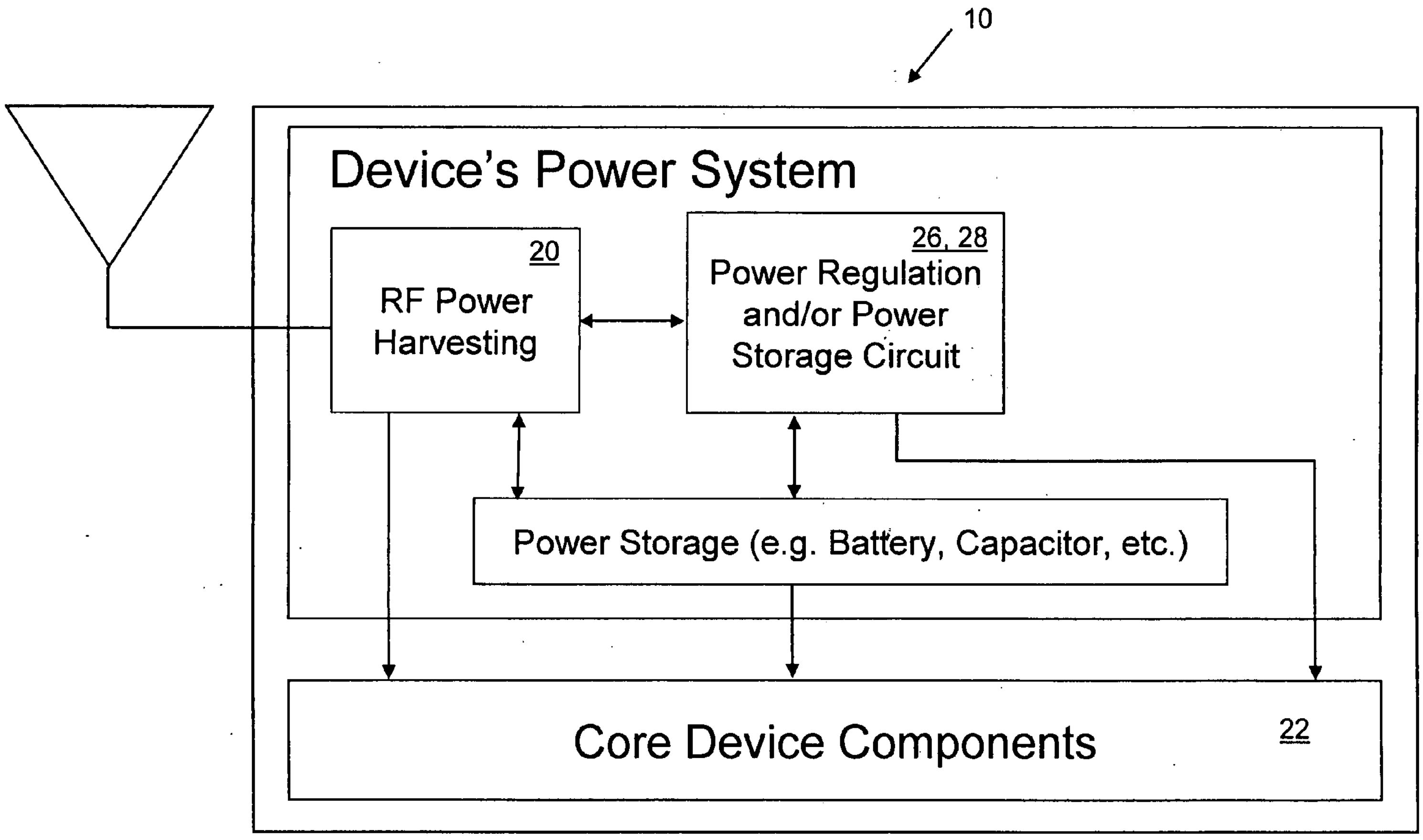
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 15

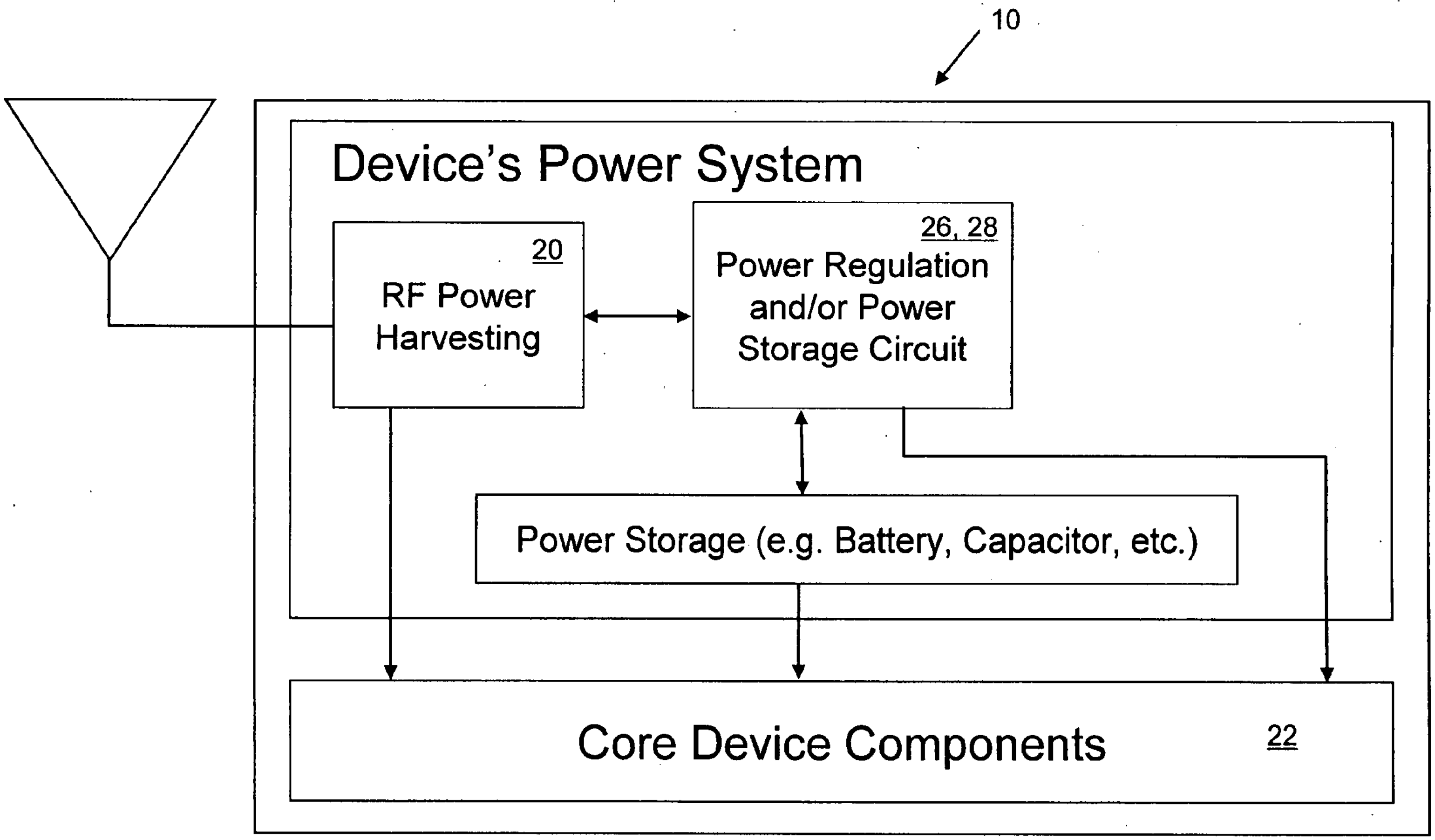
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 16

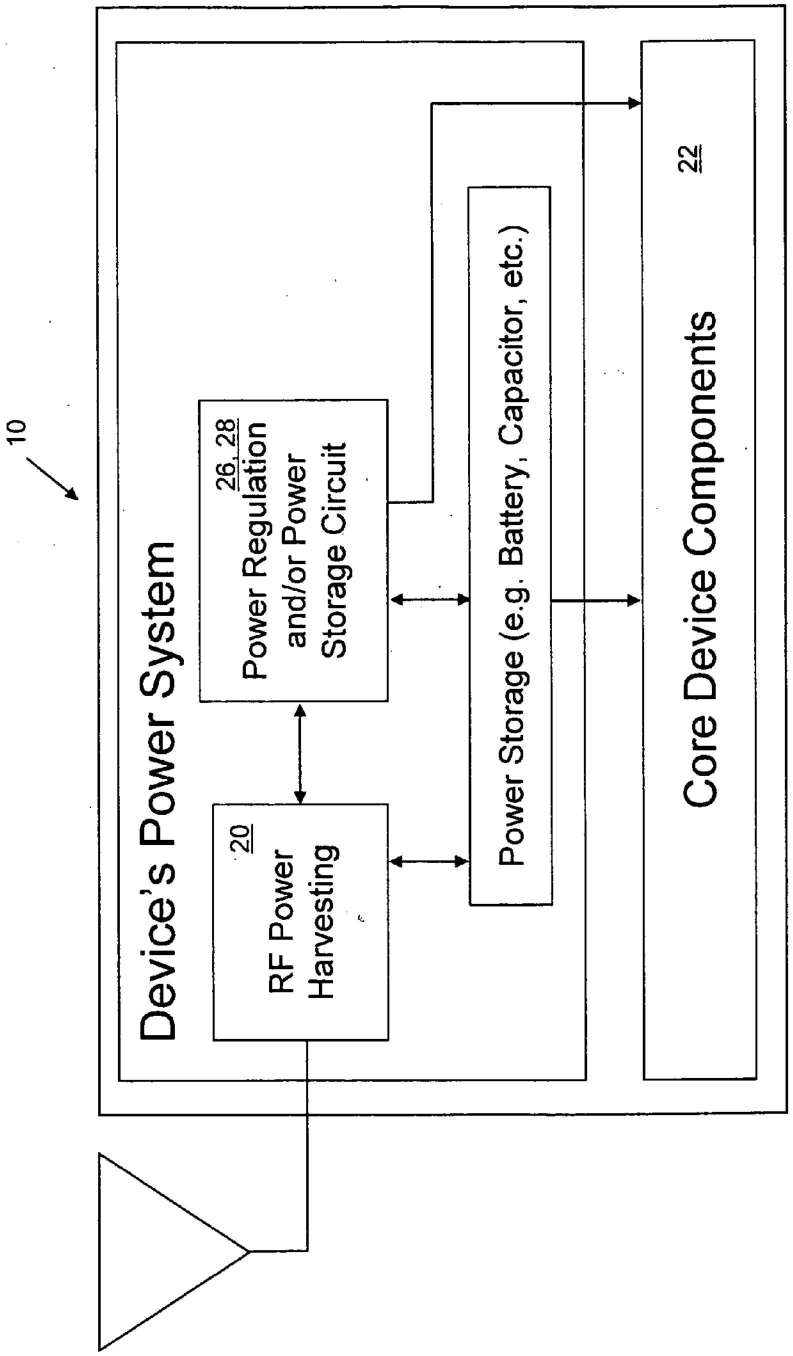
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 17

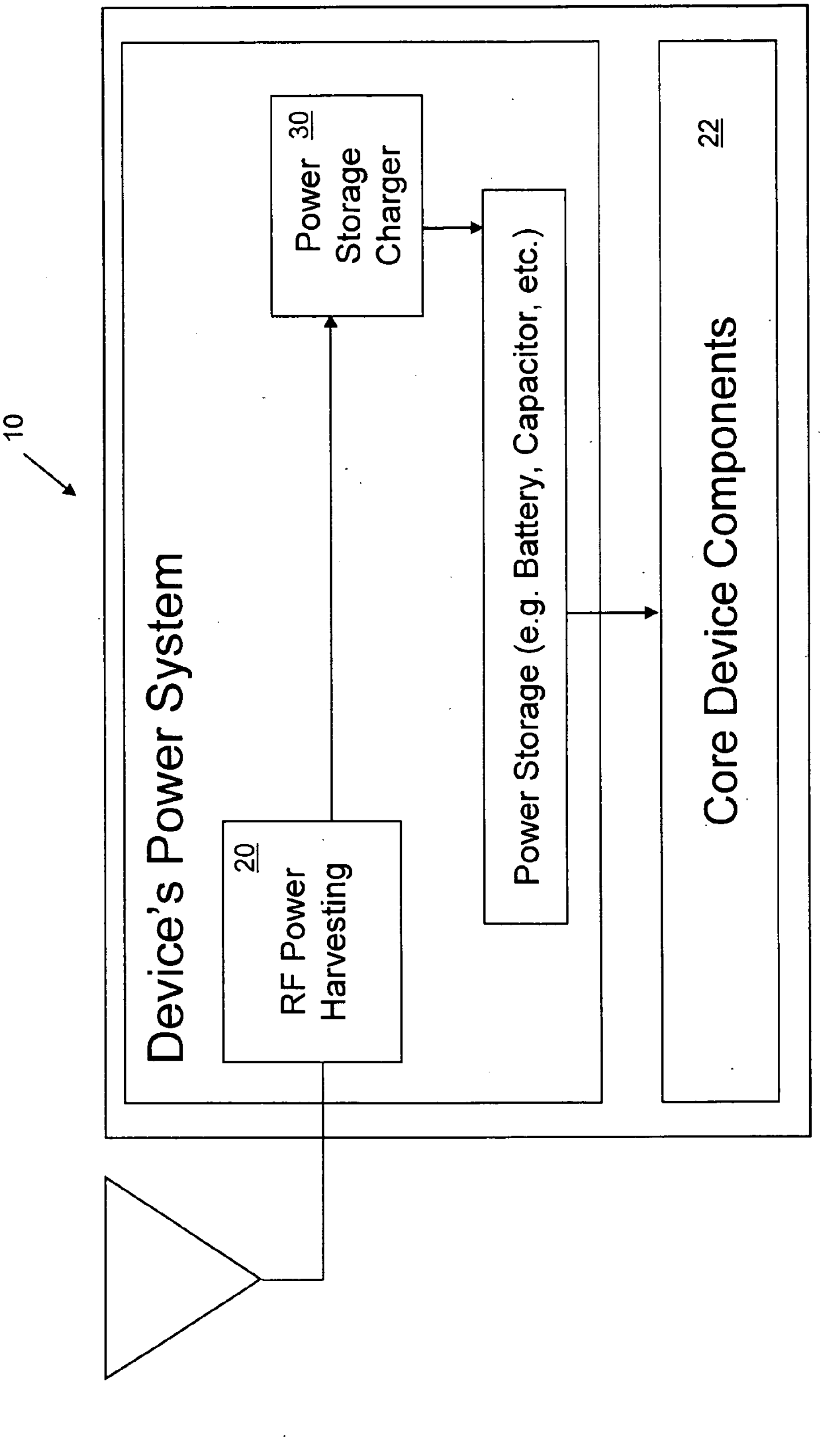
10
Device's Power System
20
RF Power Harvesting
Power 30
Storage
Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 18

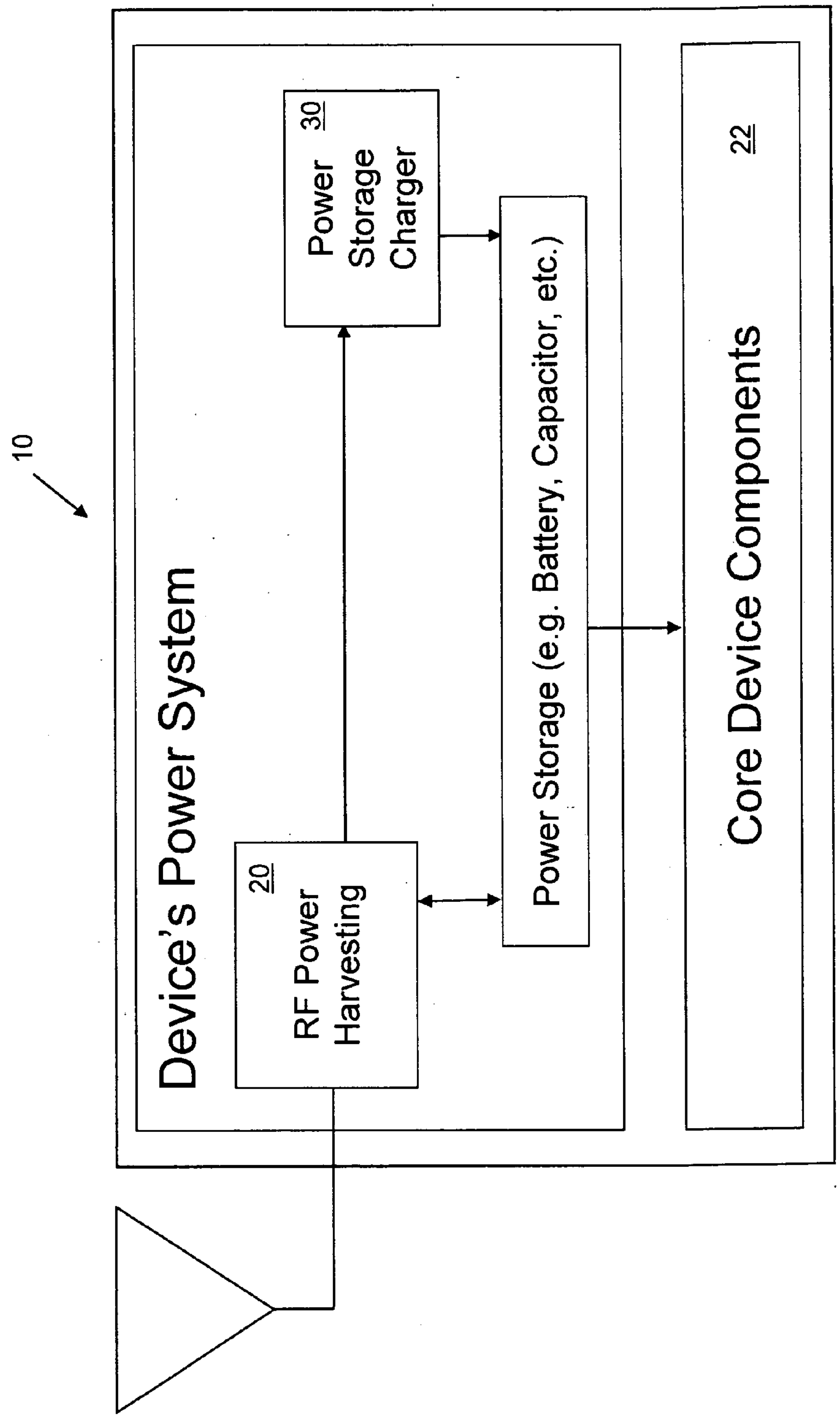
10
Device's Power System
RF Power Harvesting
20
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 19

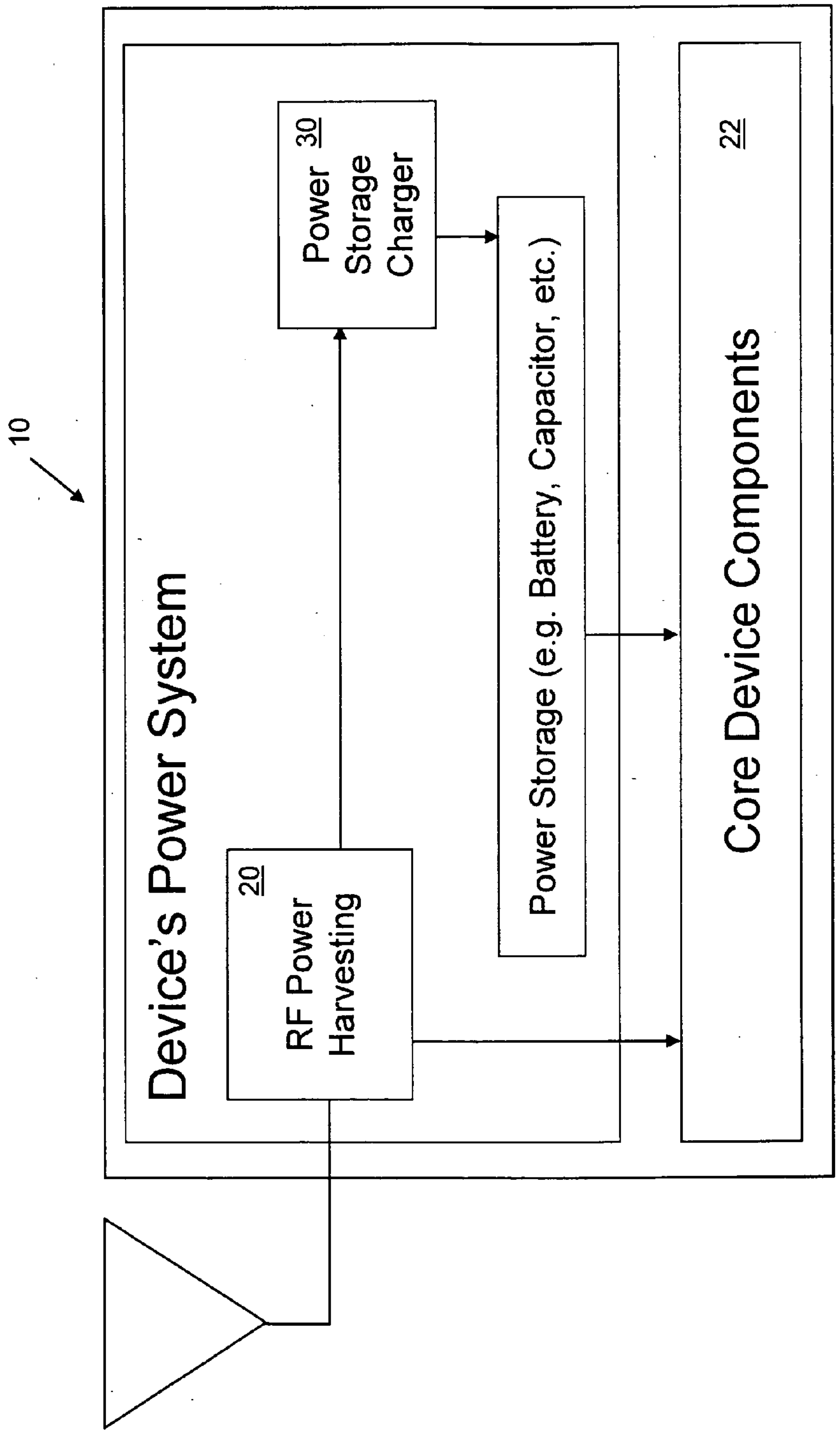
10
Device's Power System
RF Power Harvesting
20
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 20

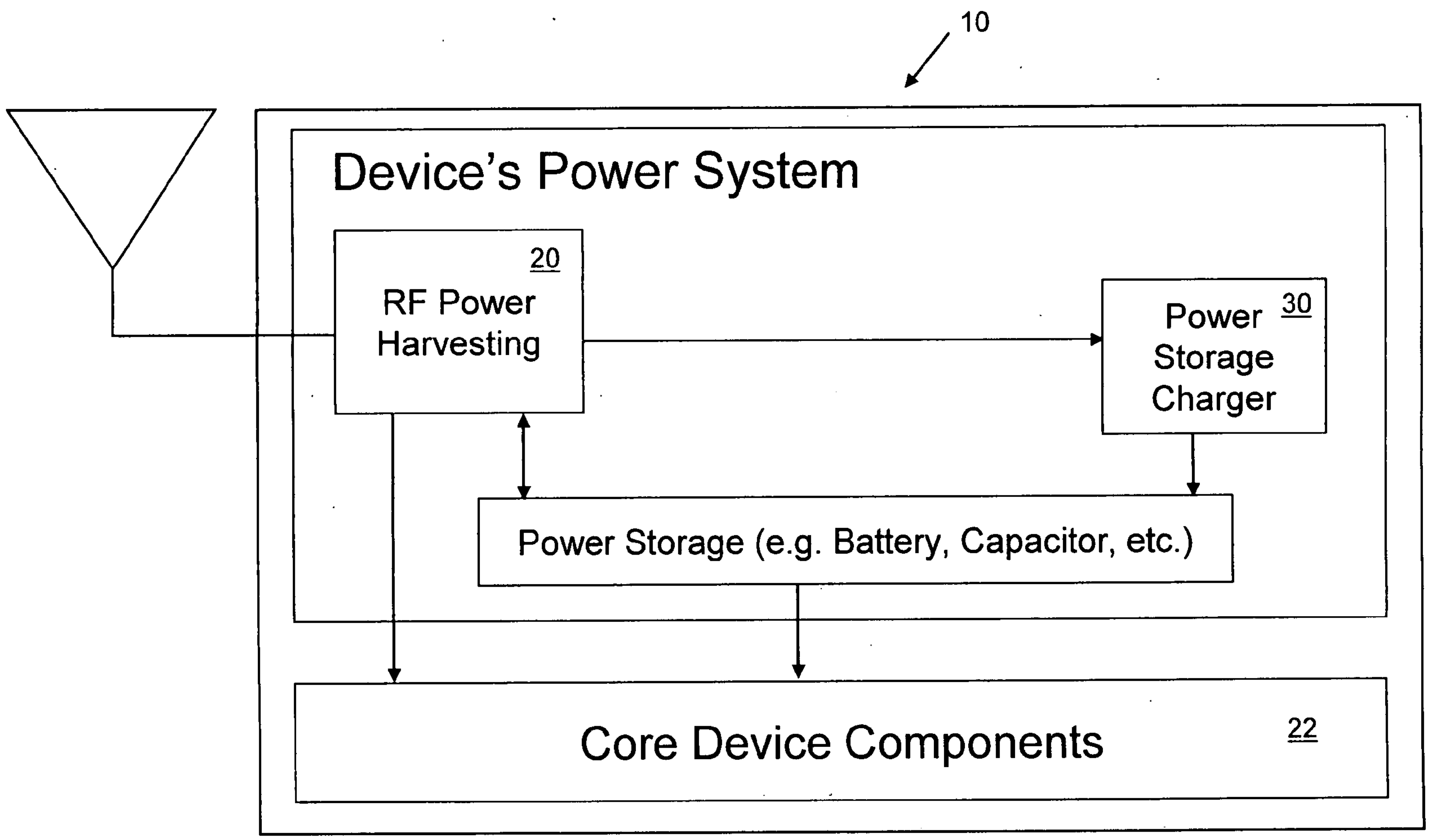
10
Device's Power System
20
RF Power Harvesting
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 21

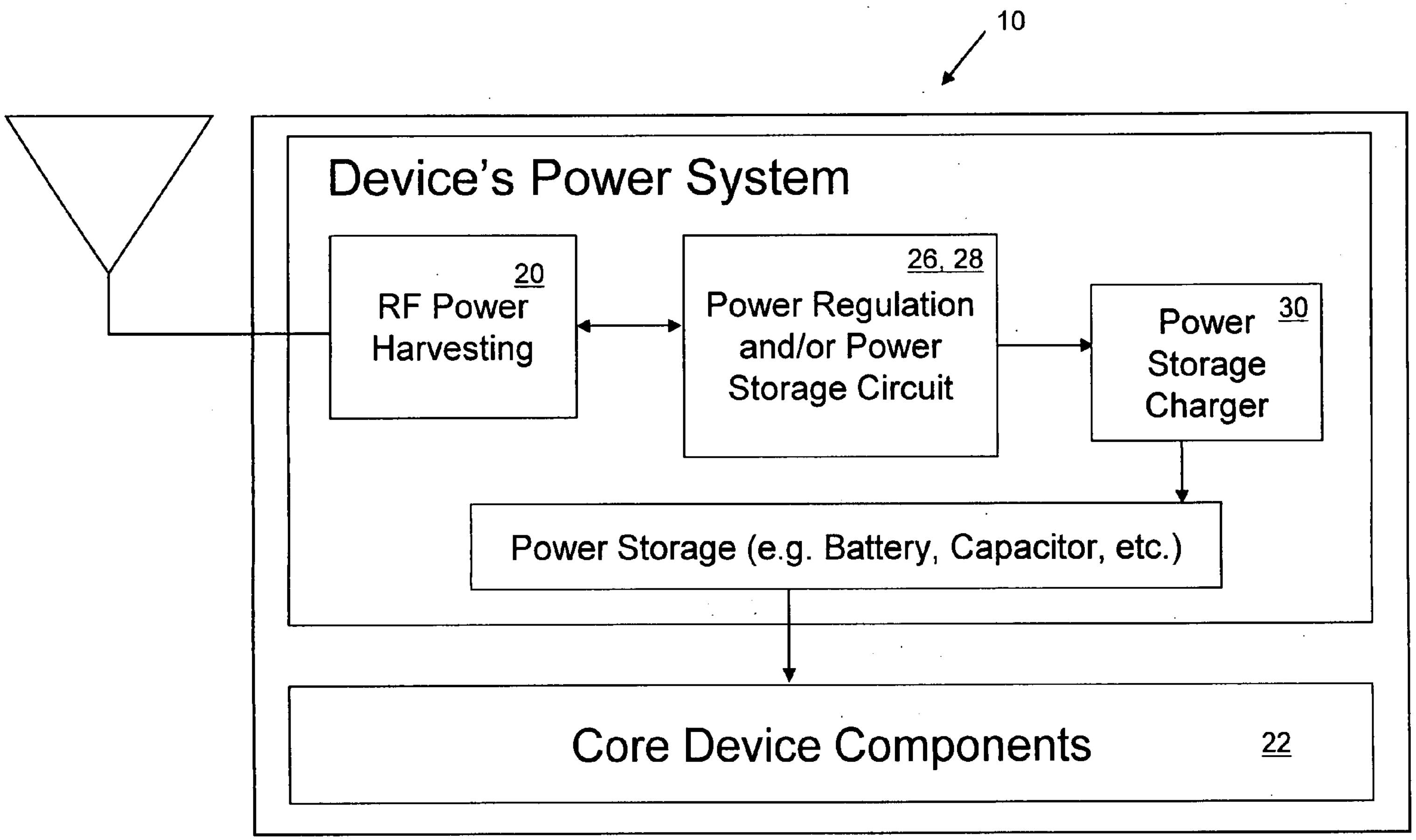
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 22

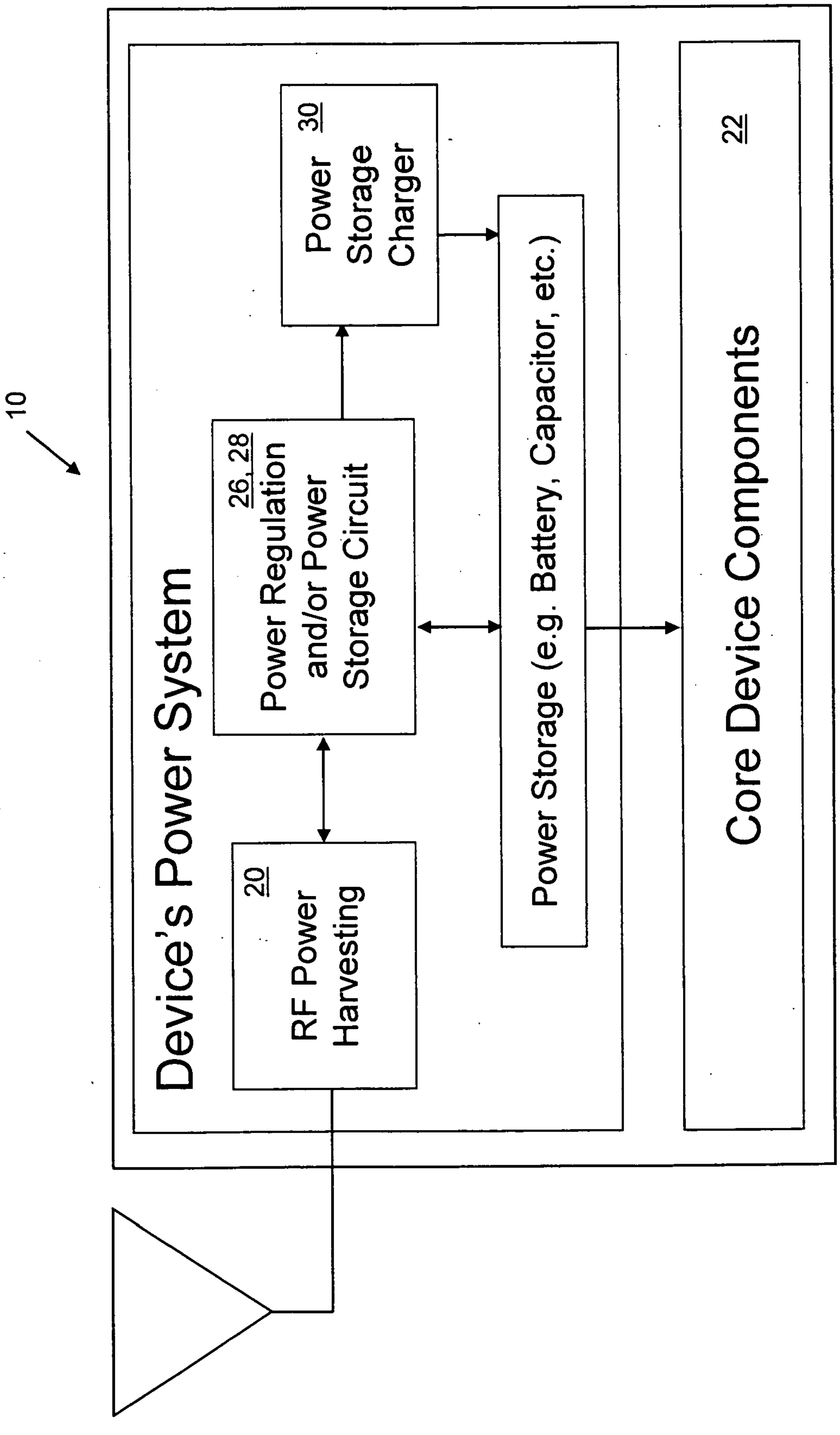
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 23

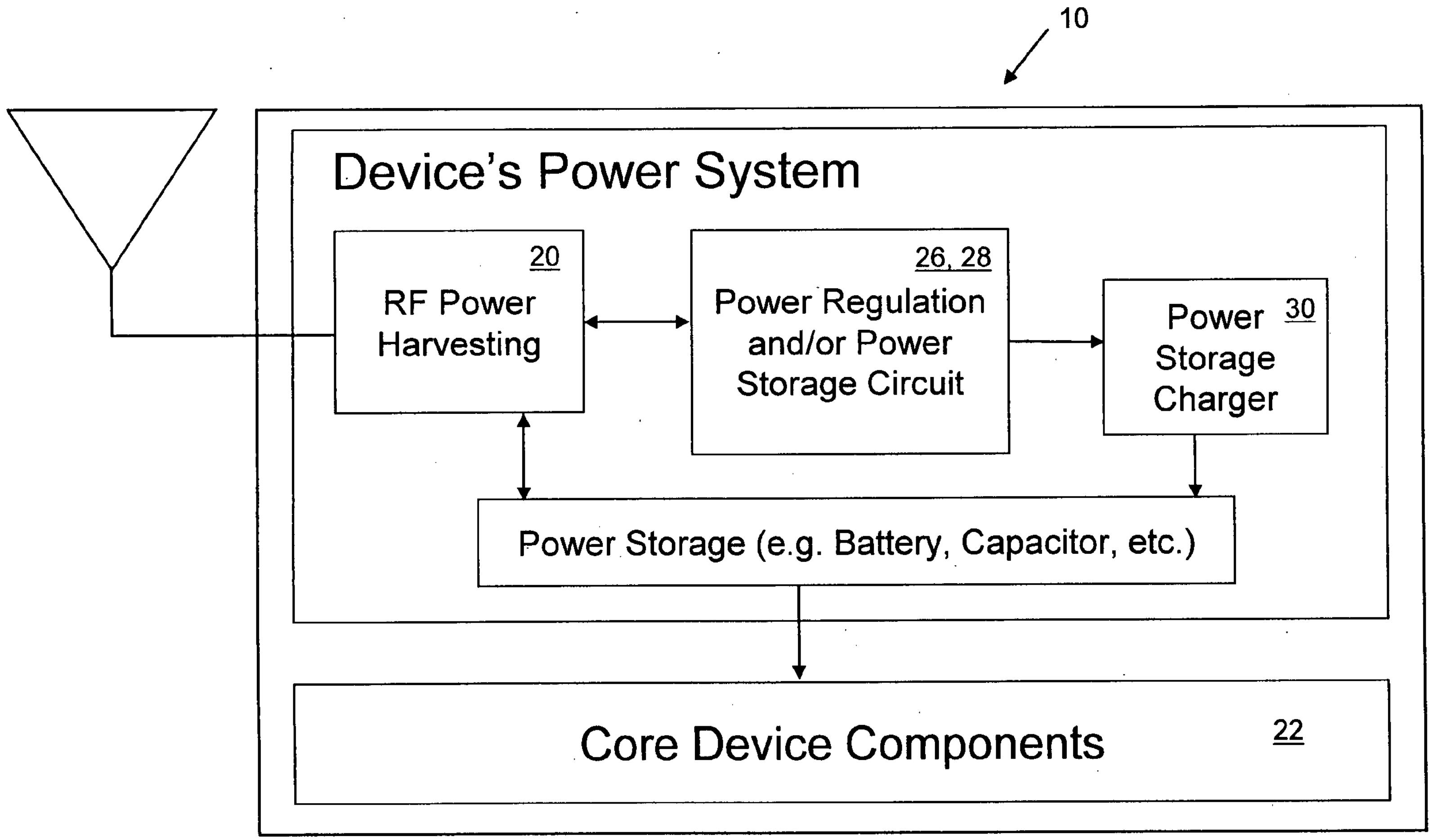
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 24

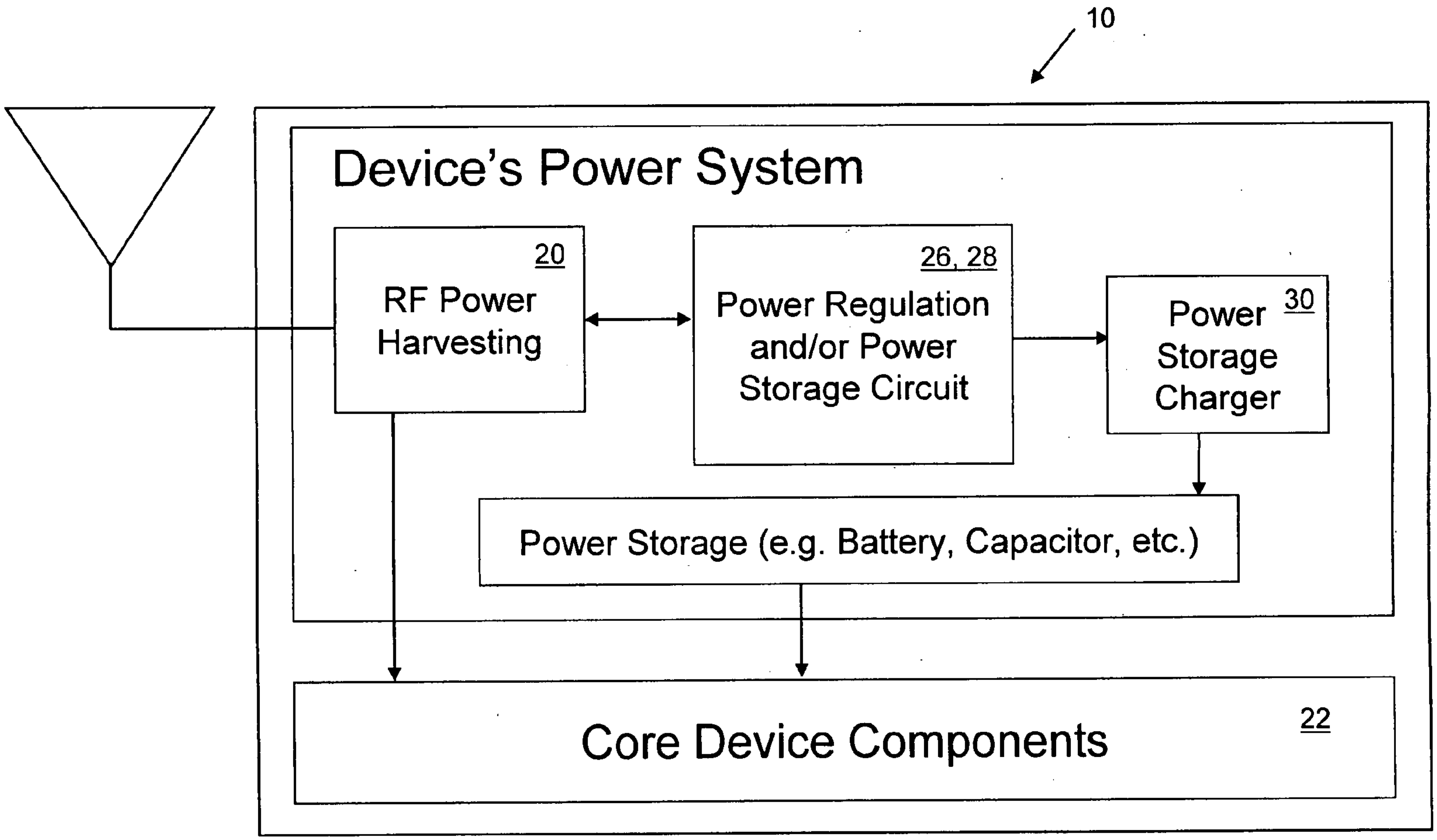
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 25

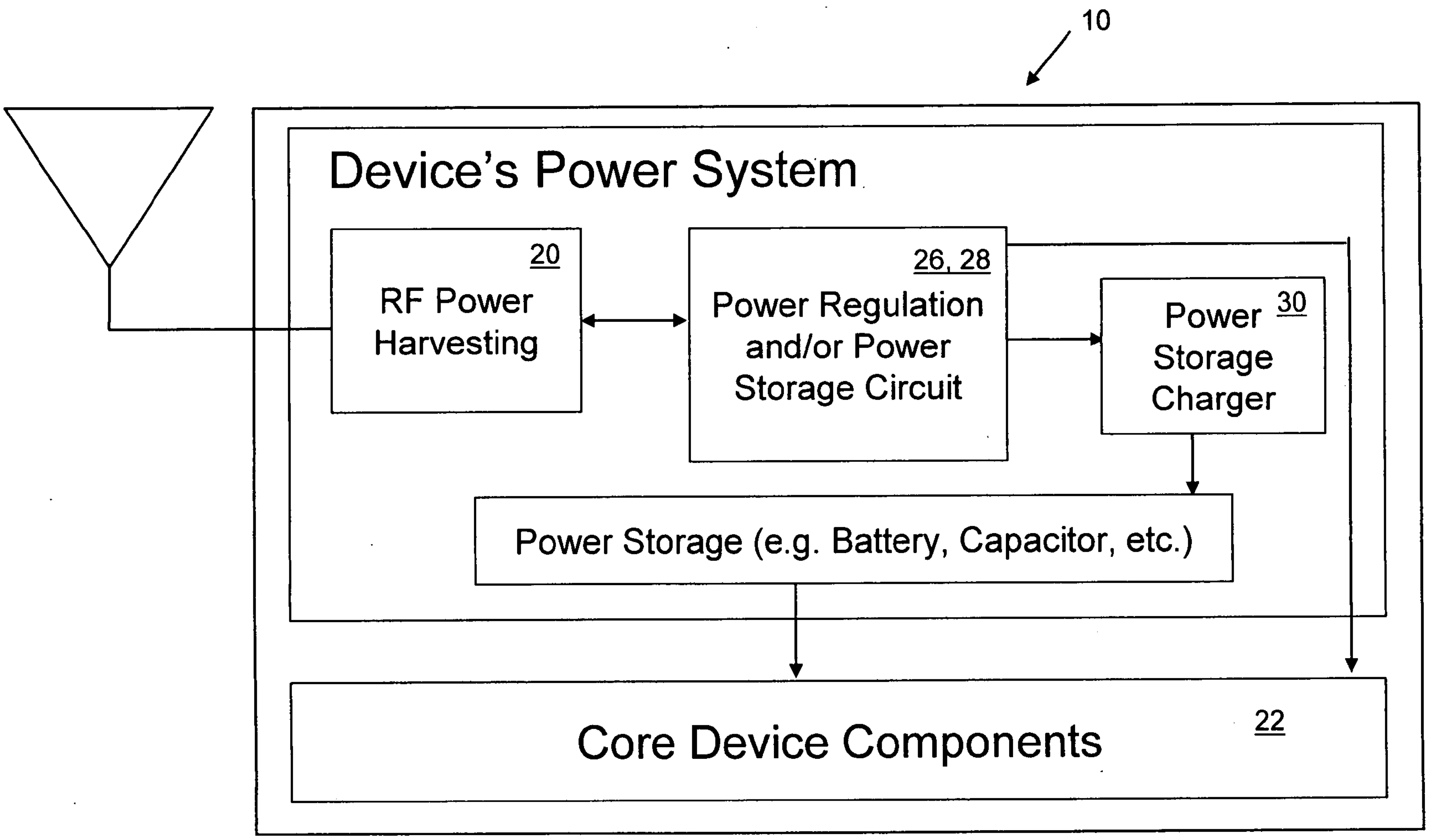
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 26

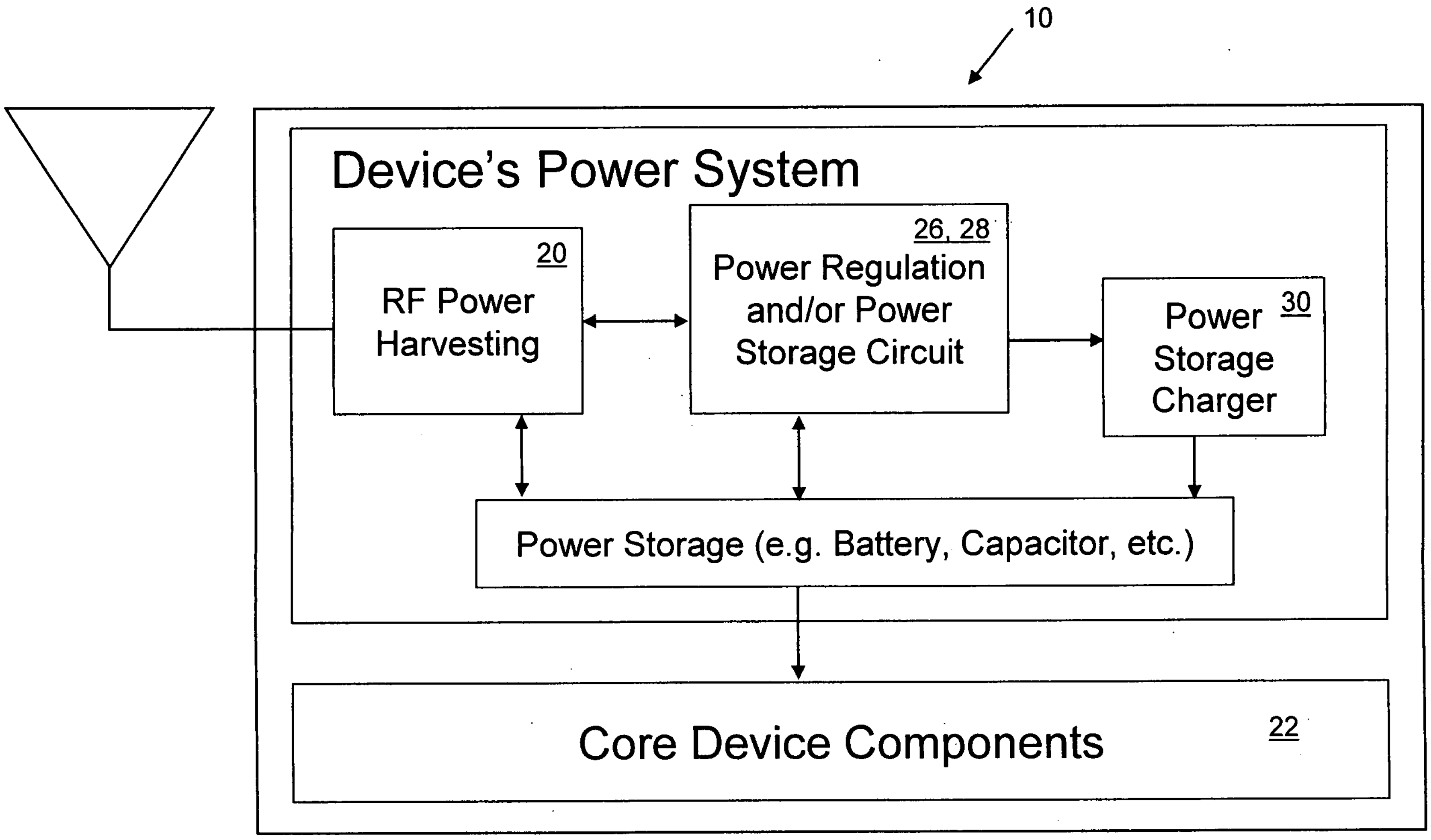
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 27

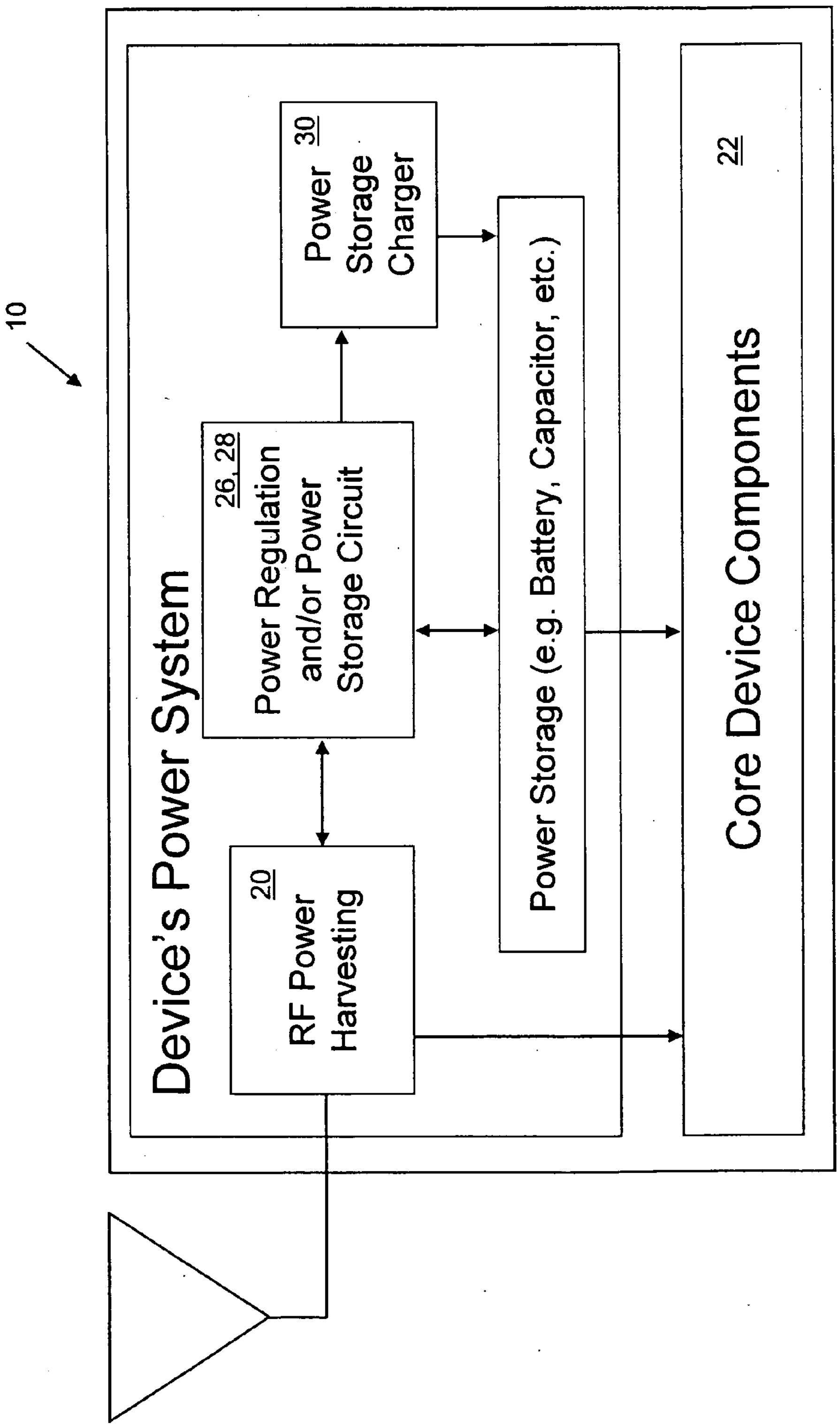
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 28

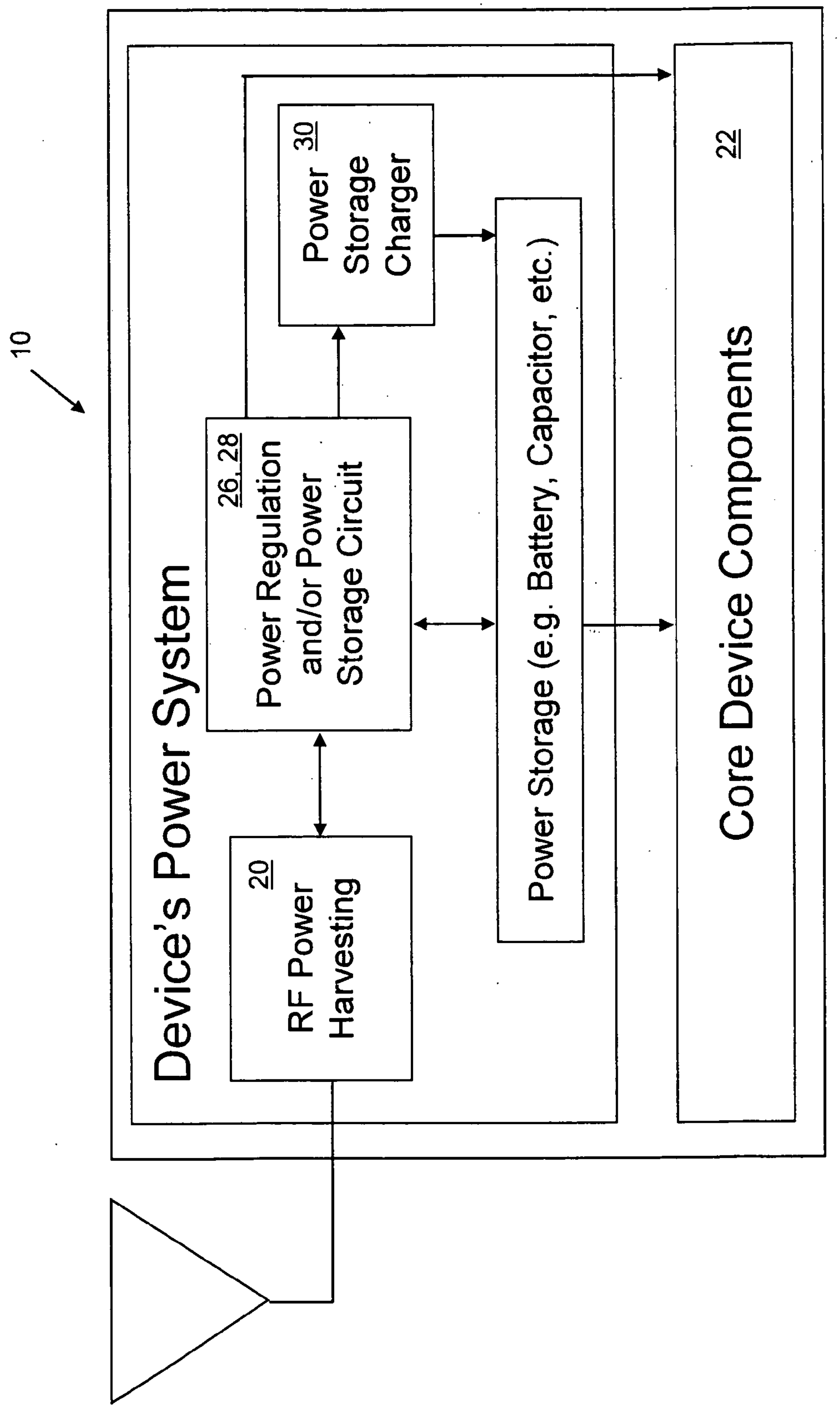
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 29

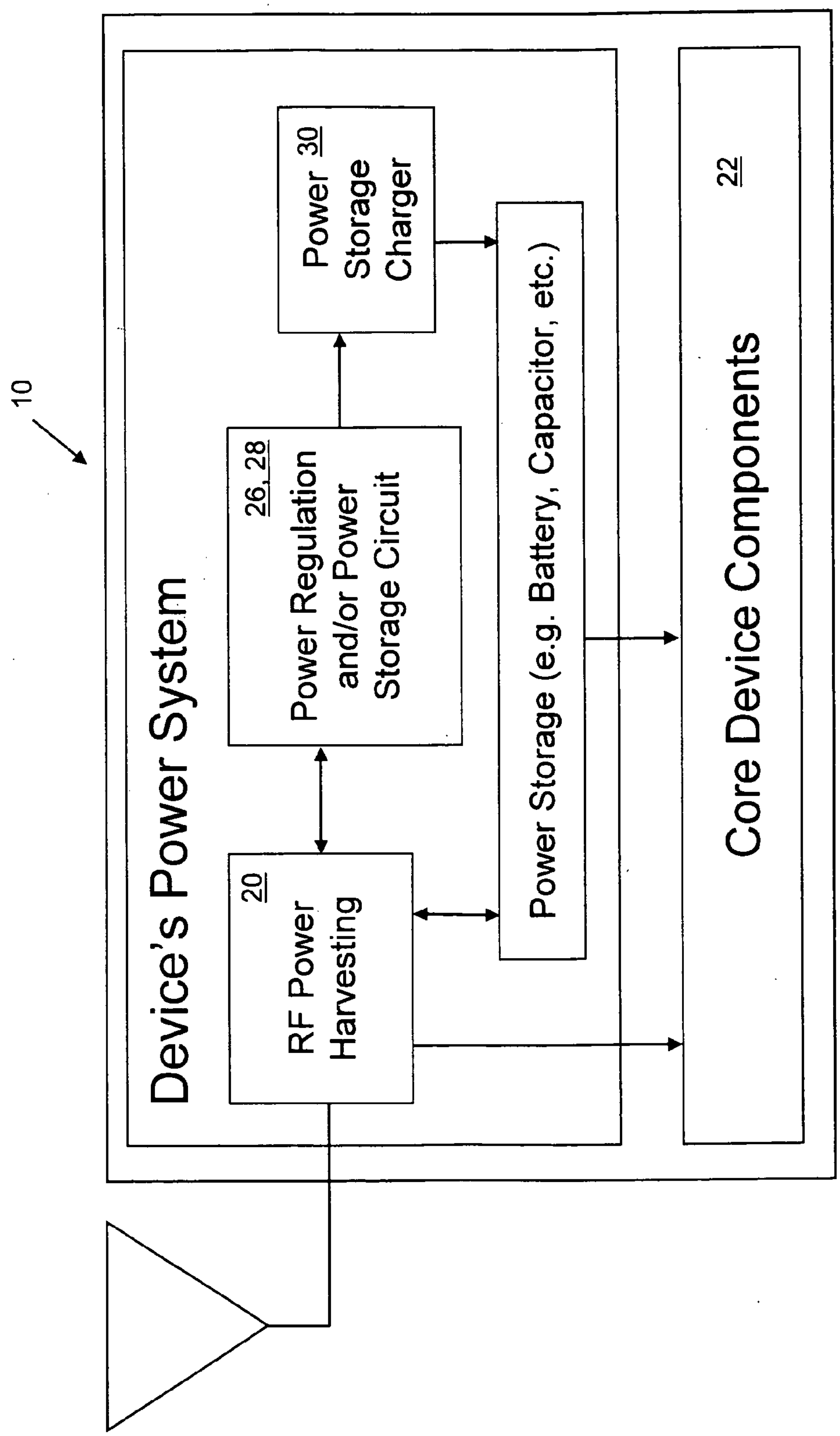
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 30

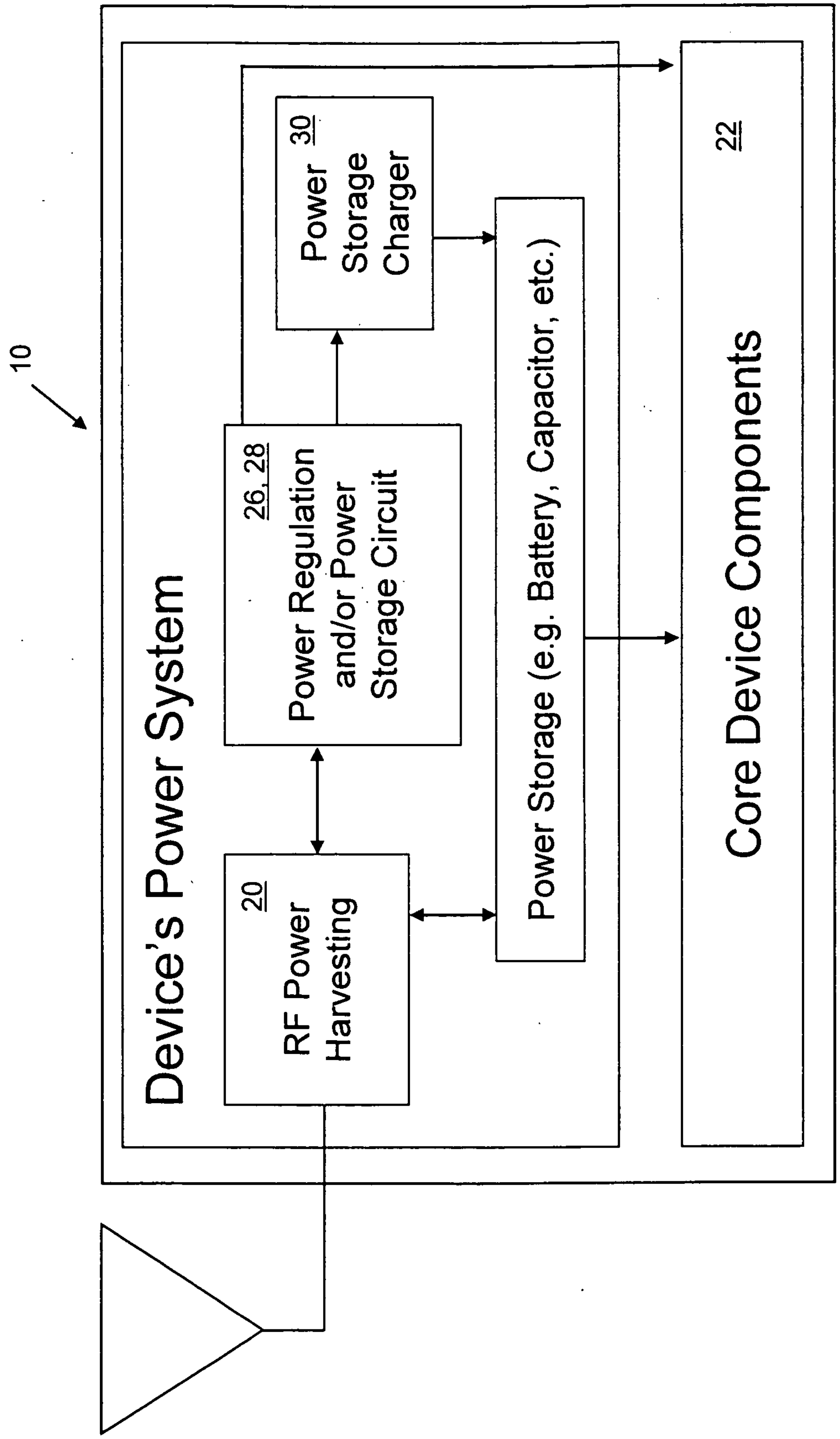
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 31

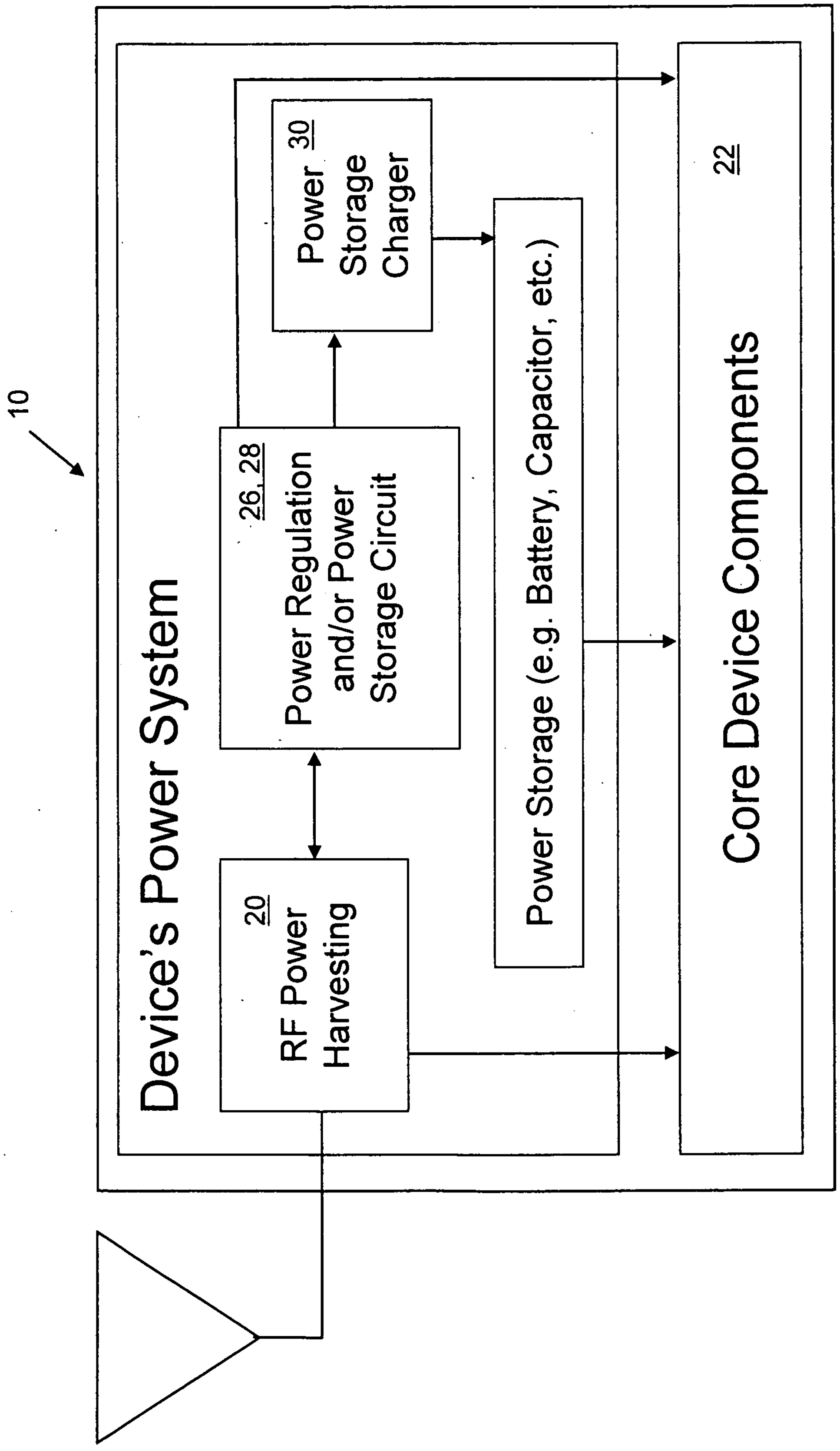
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 32

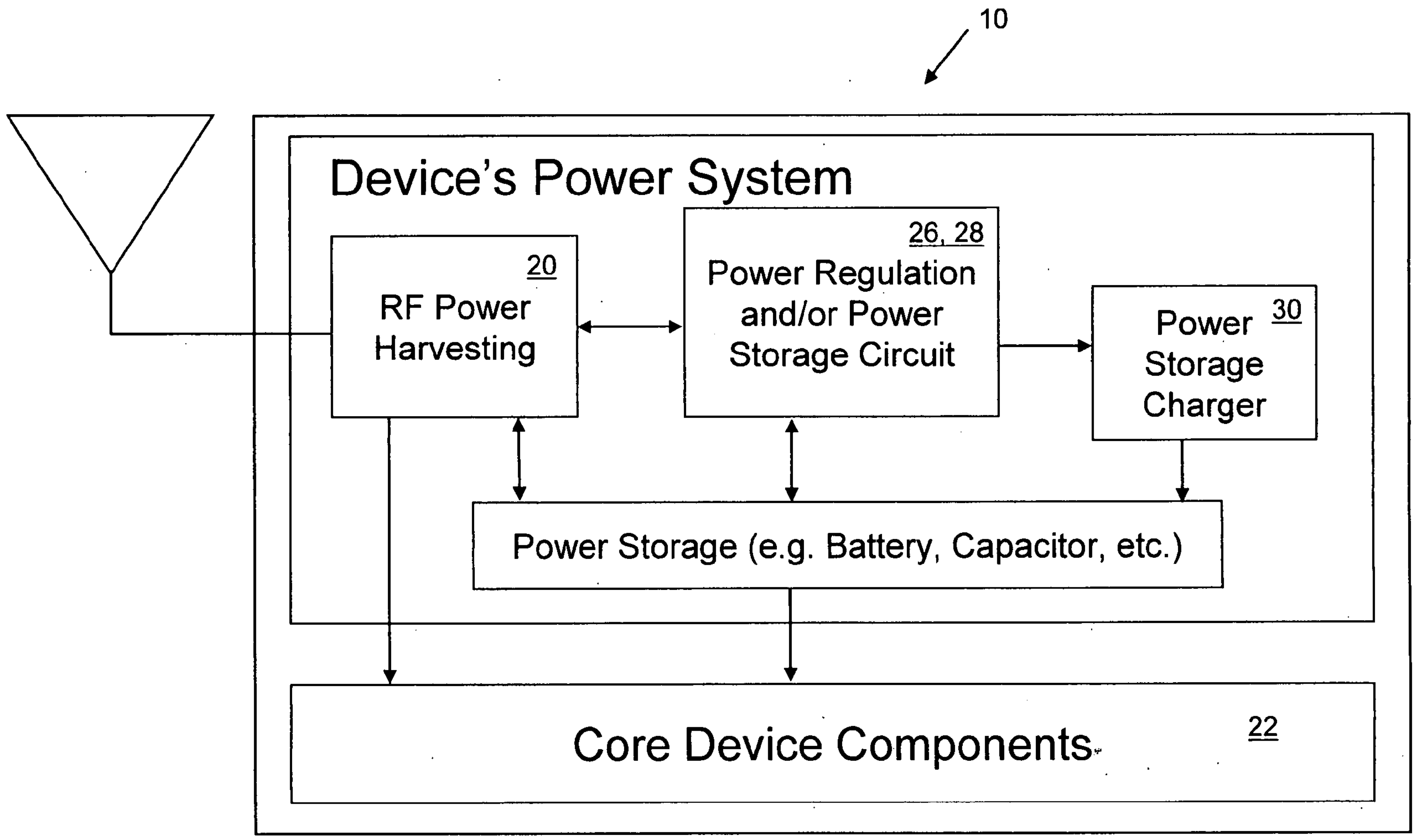
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 33

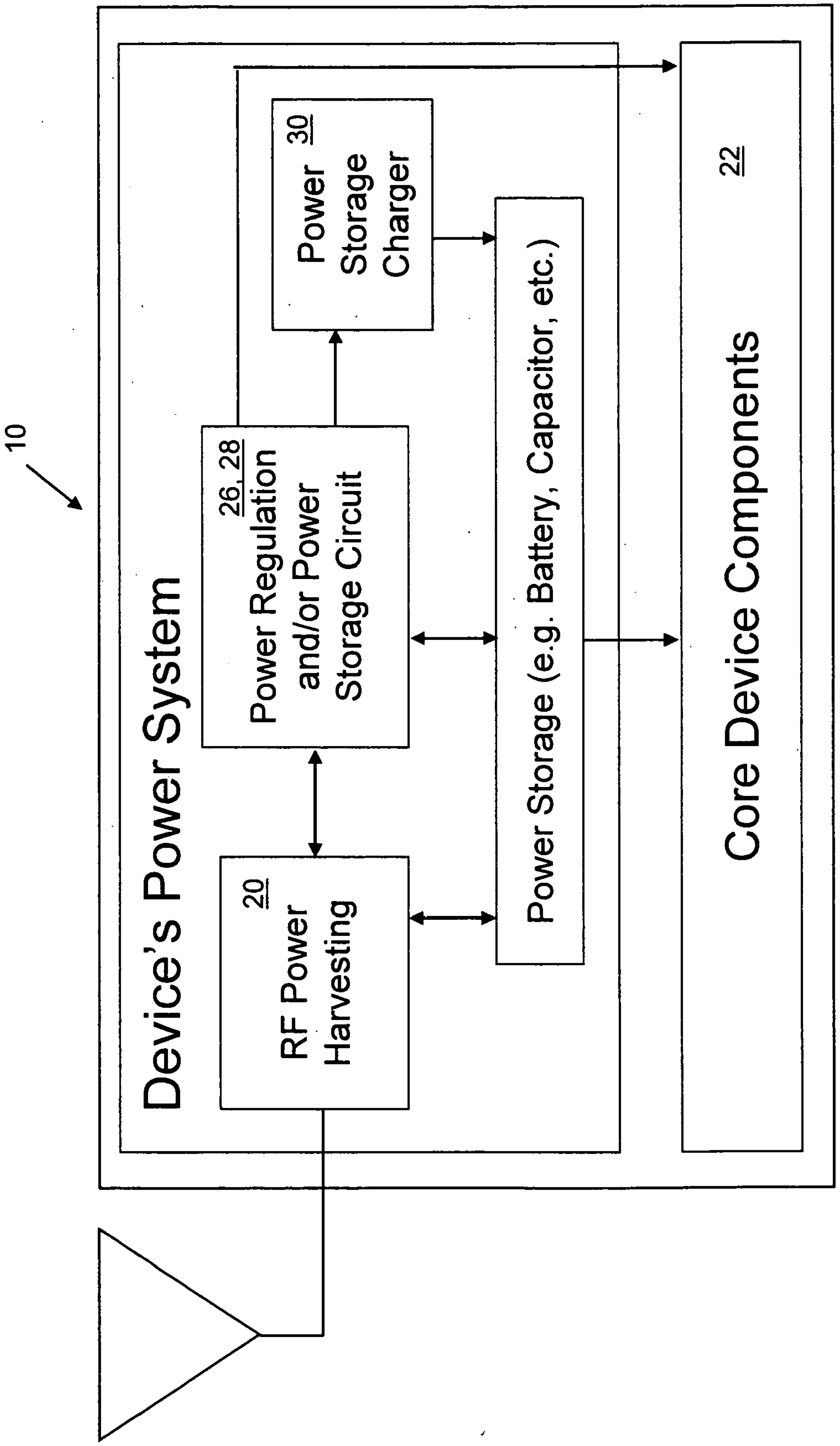
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 34

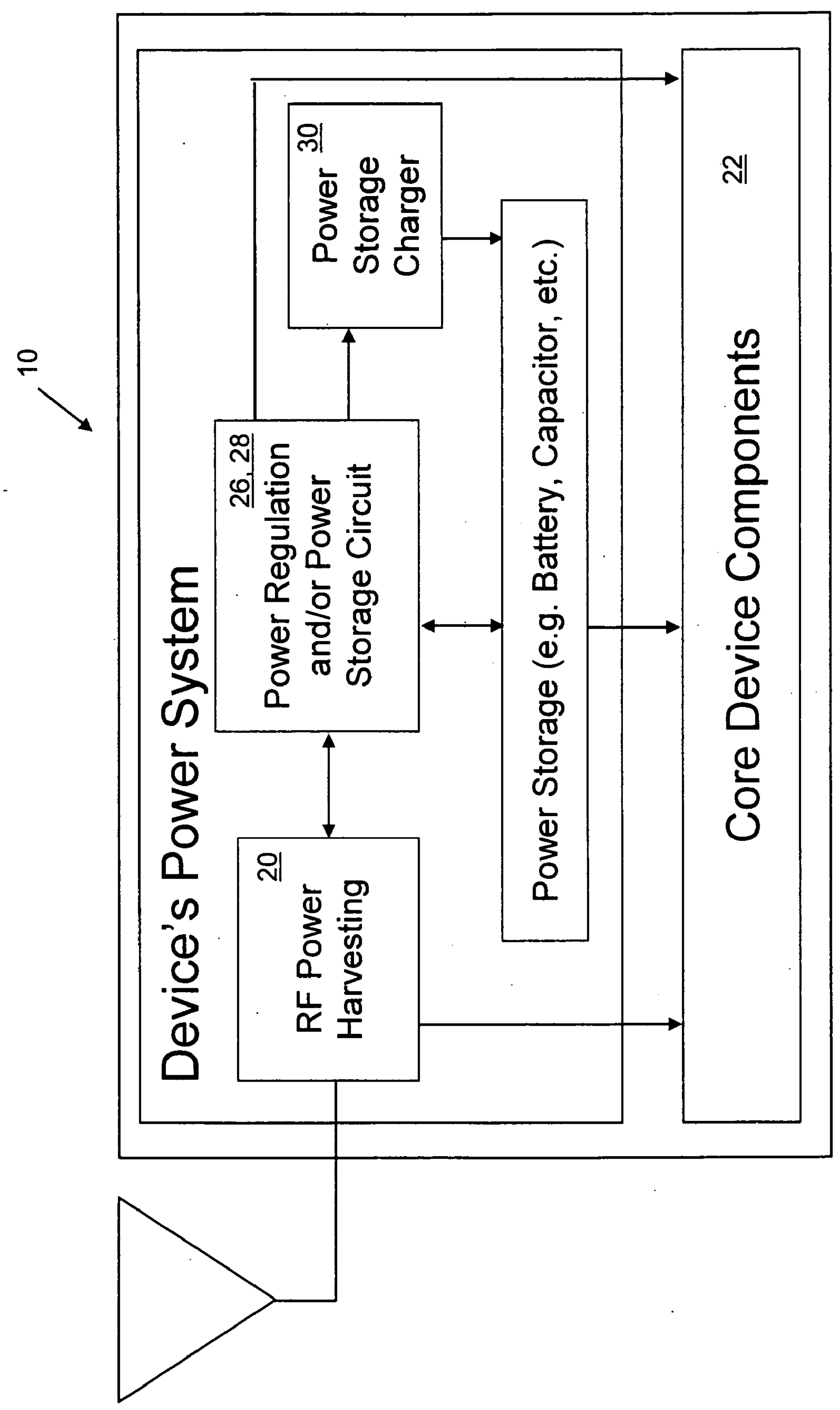
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 35

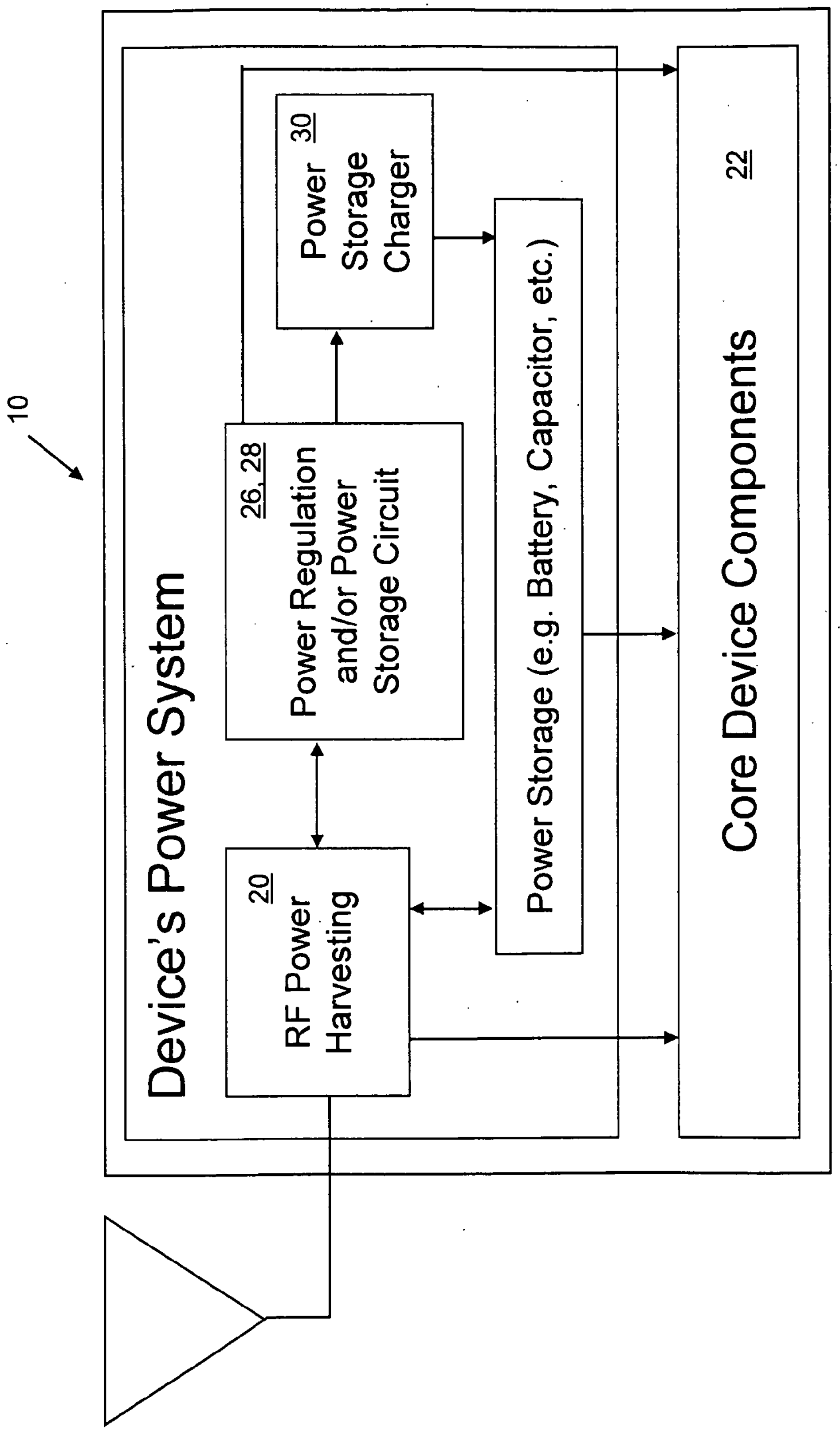
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 36

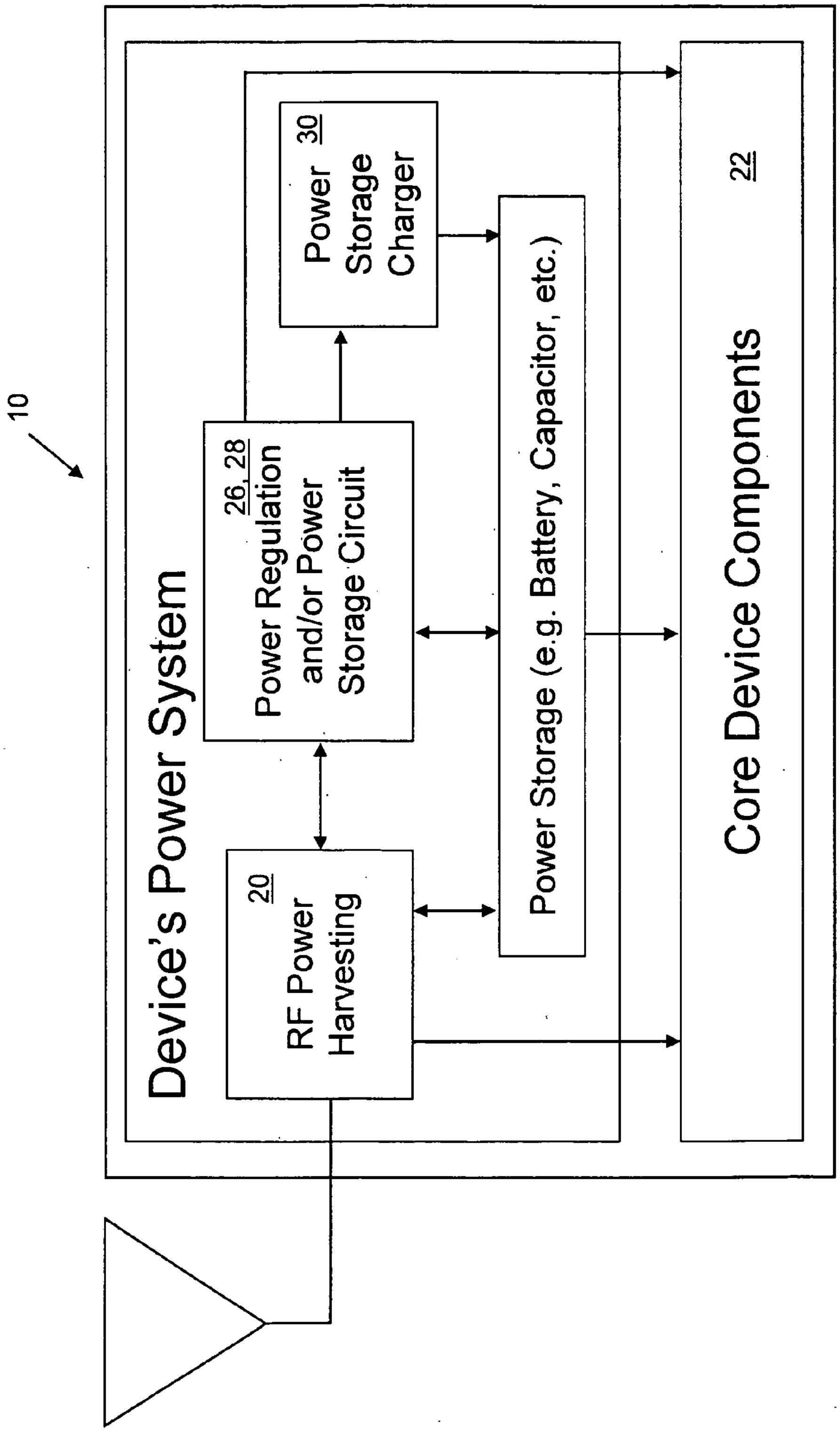
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 37

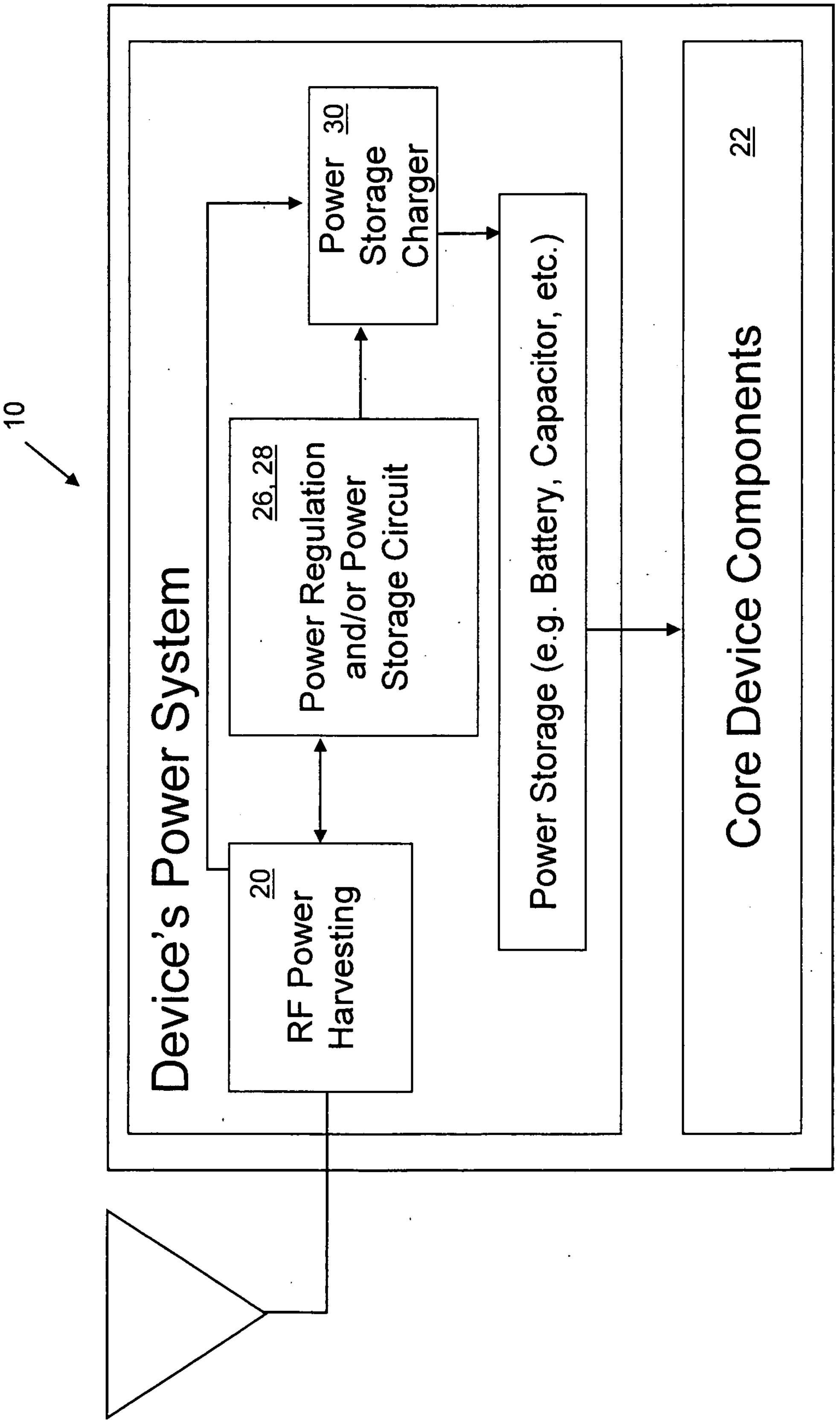
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 38

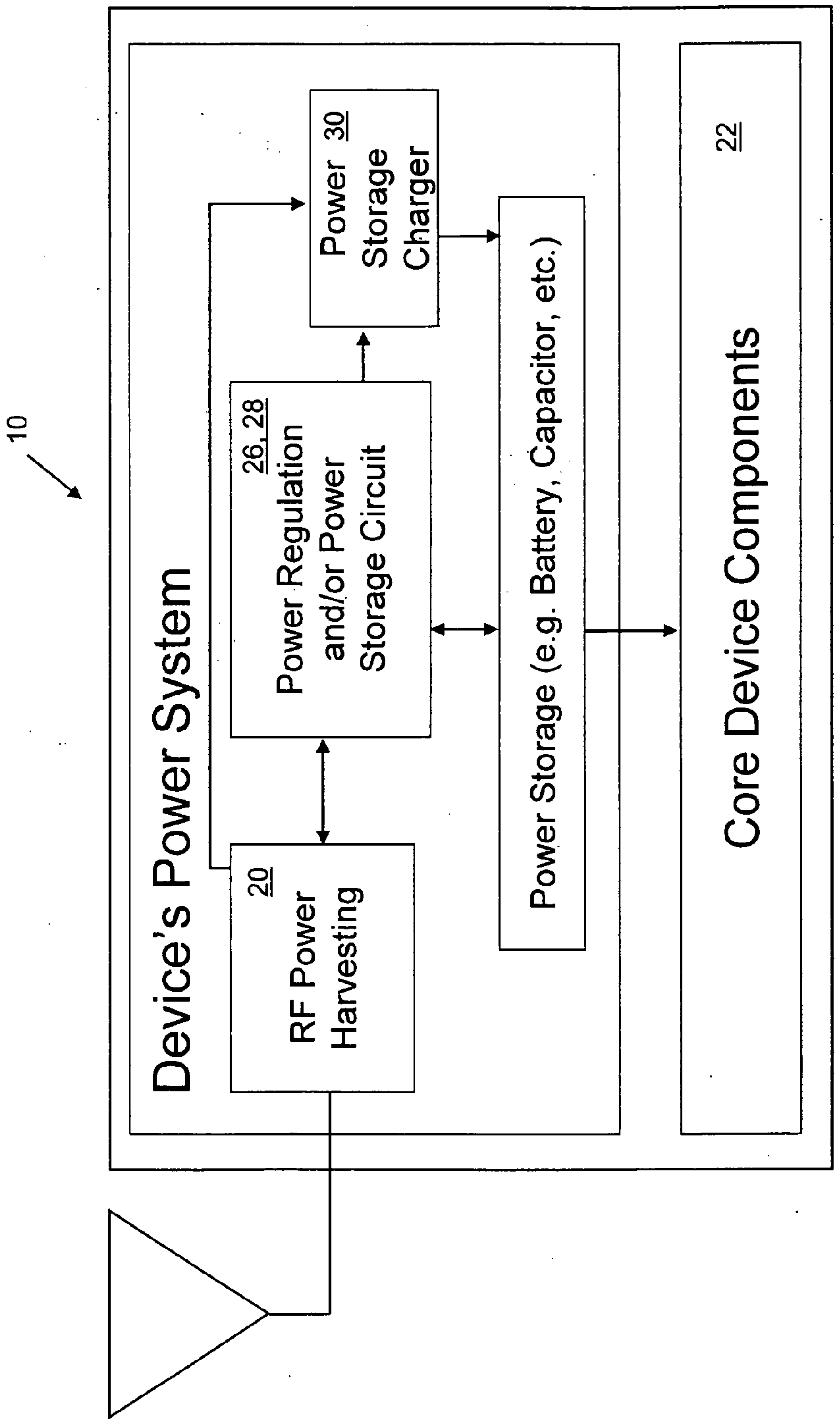
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 39

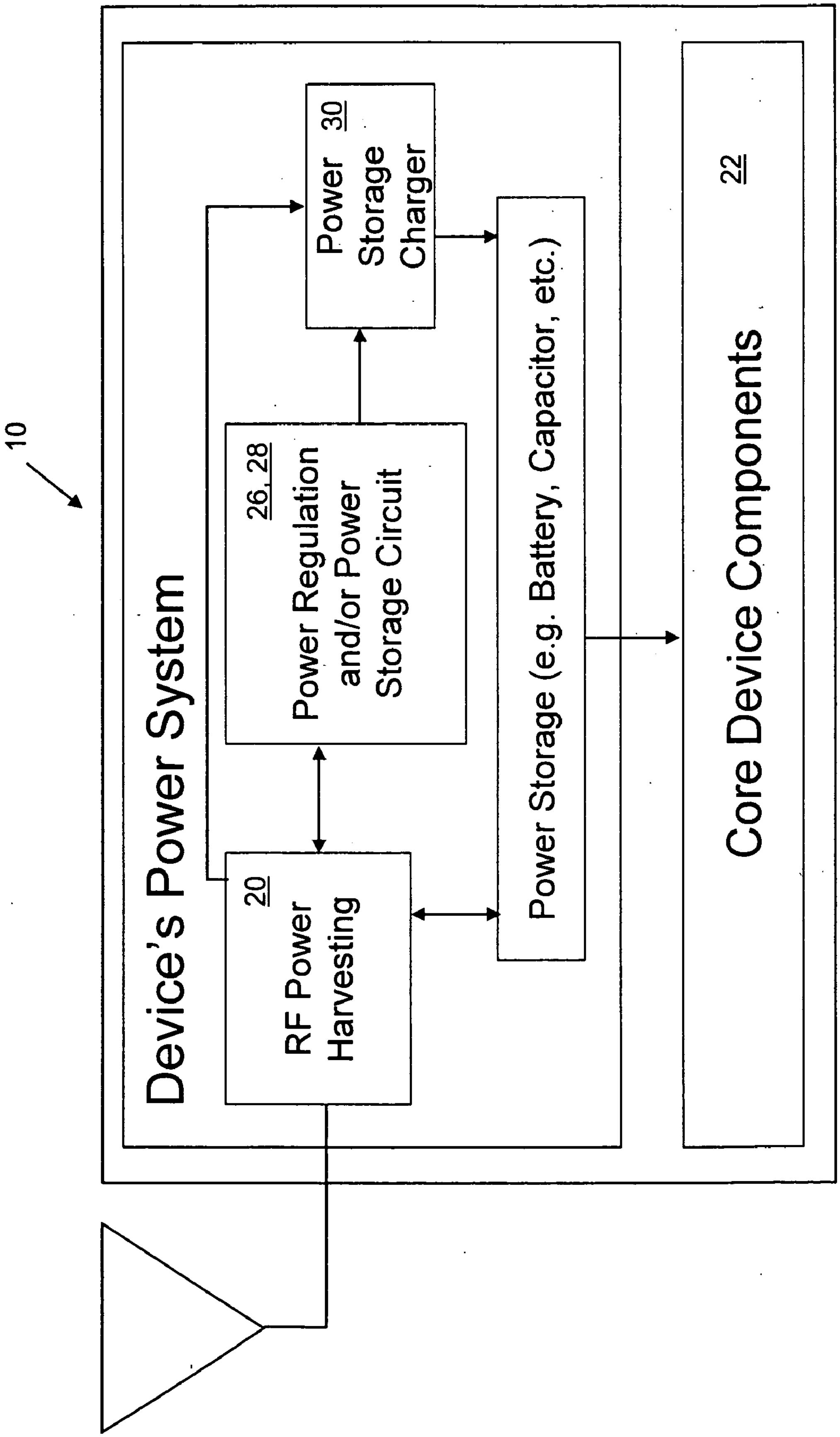
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 40

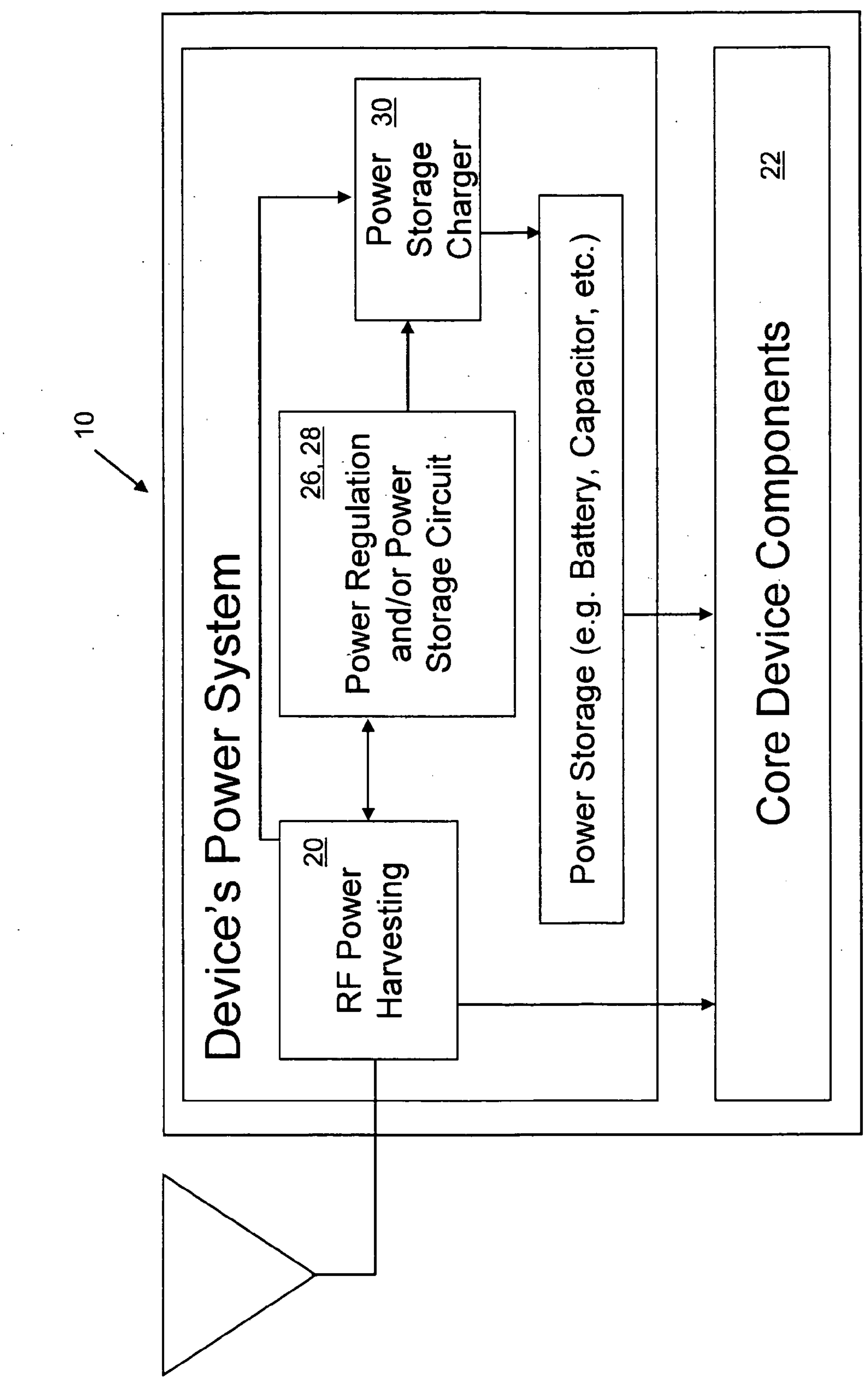
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 41

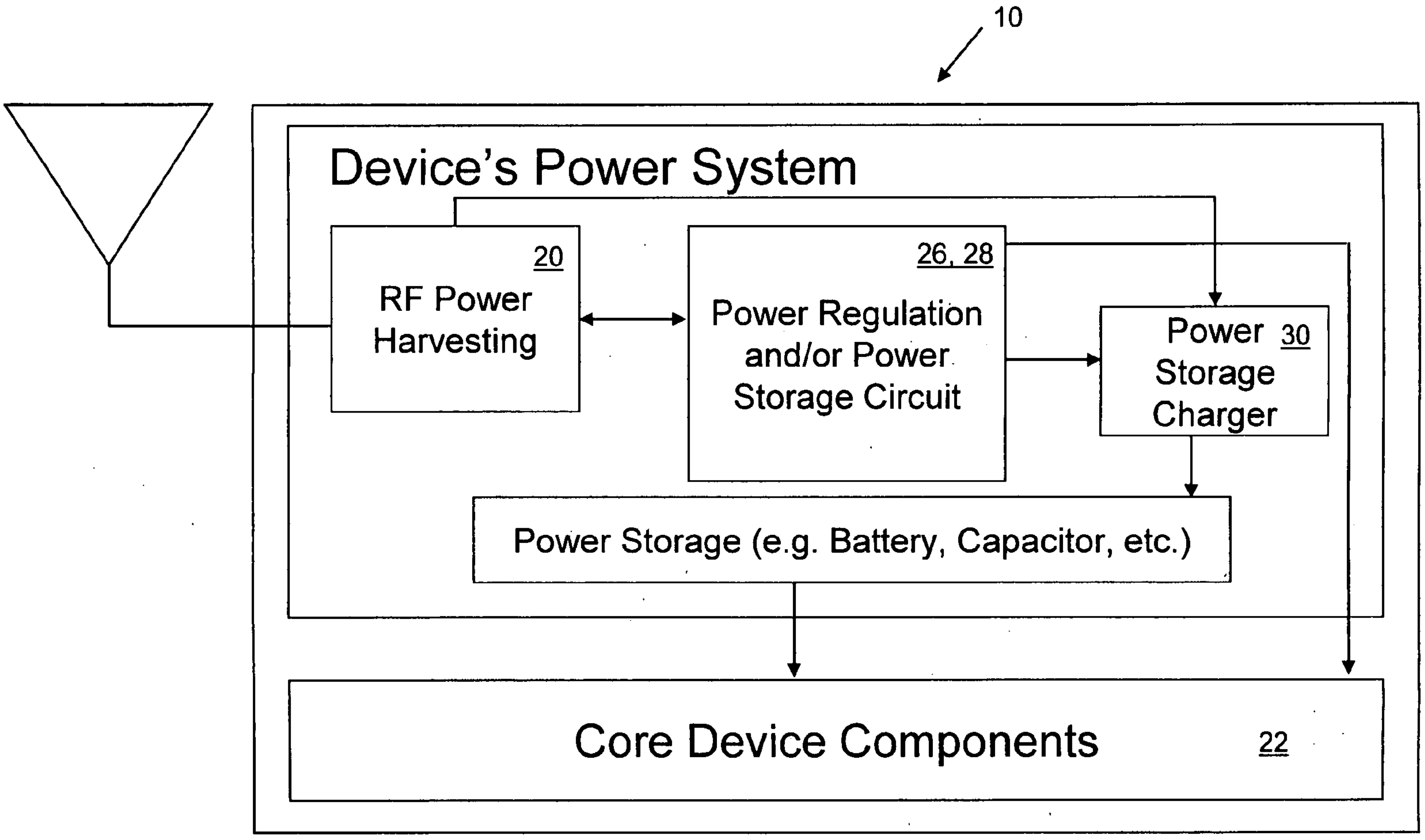
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30 Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 42

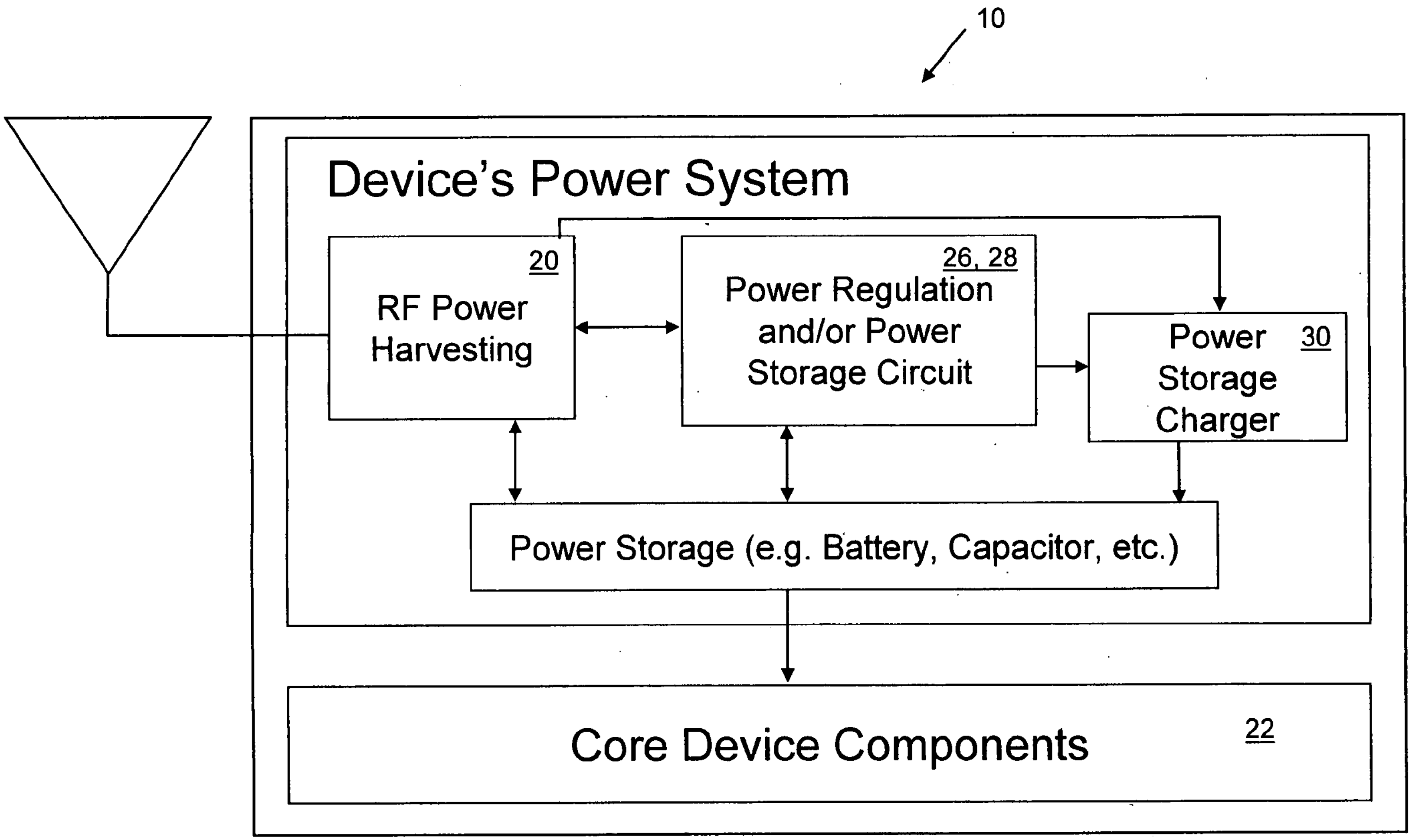
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 43

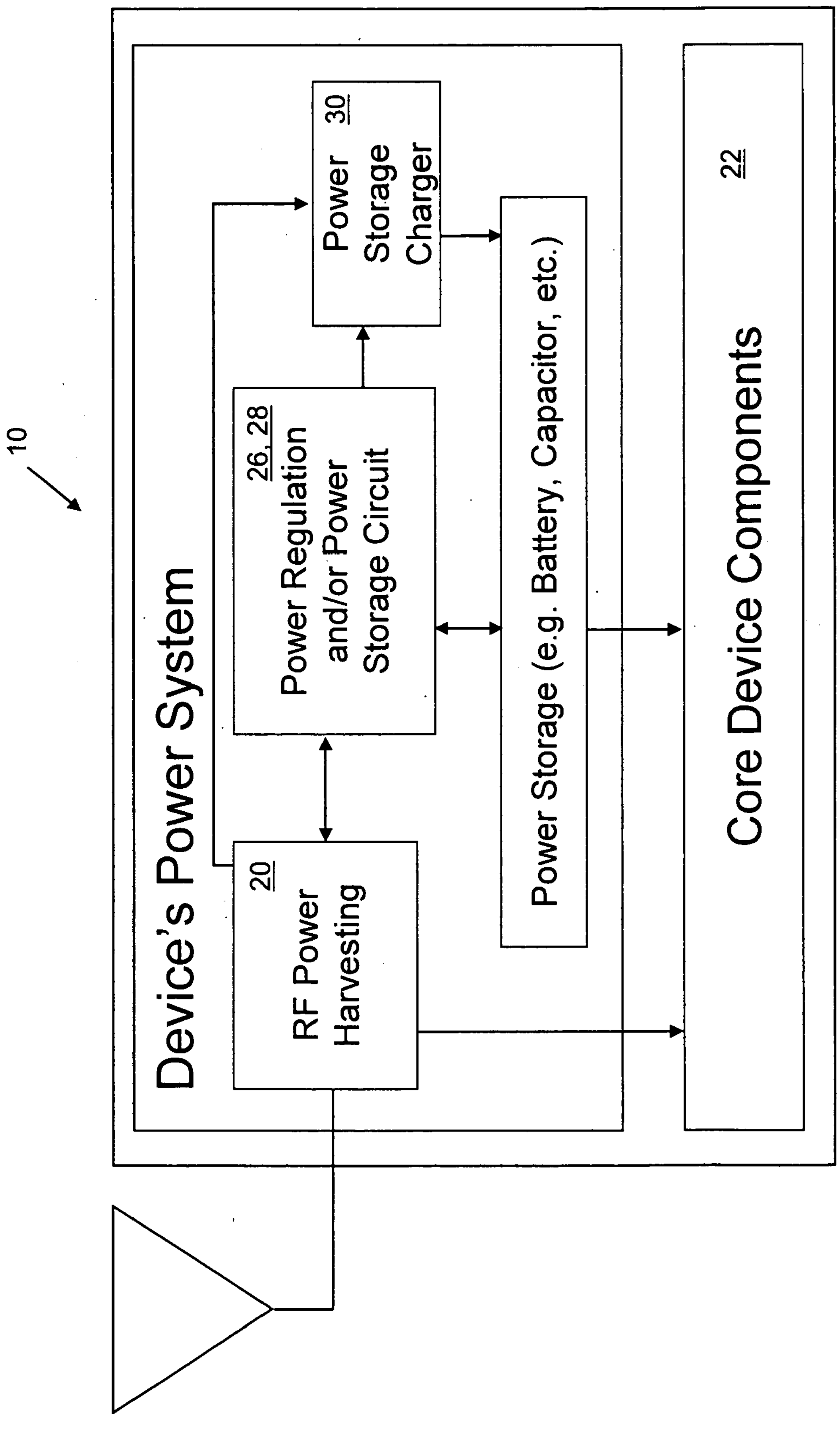
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 44

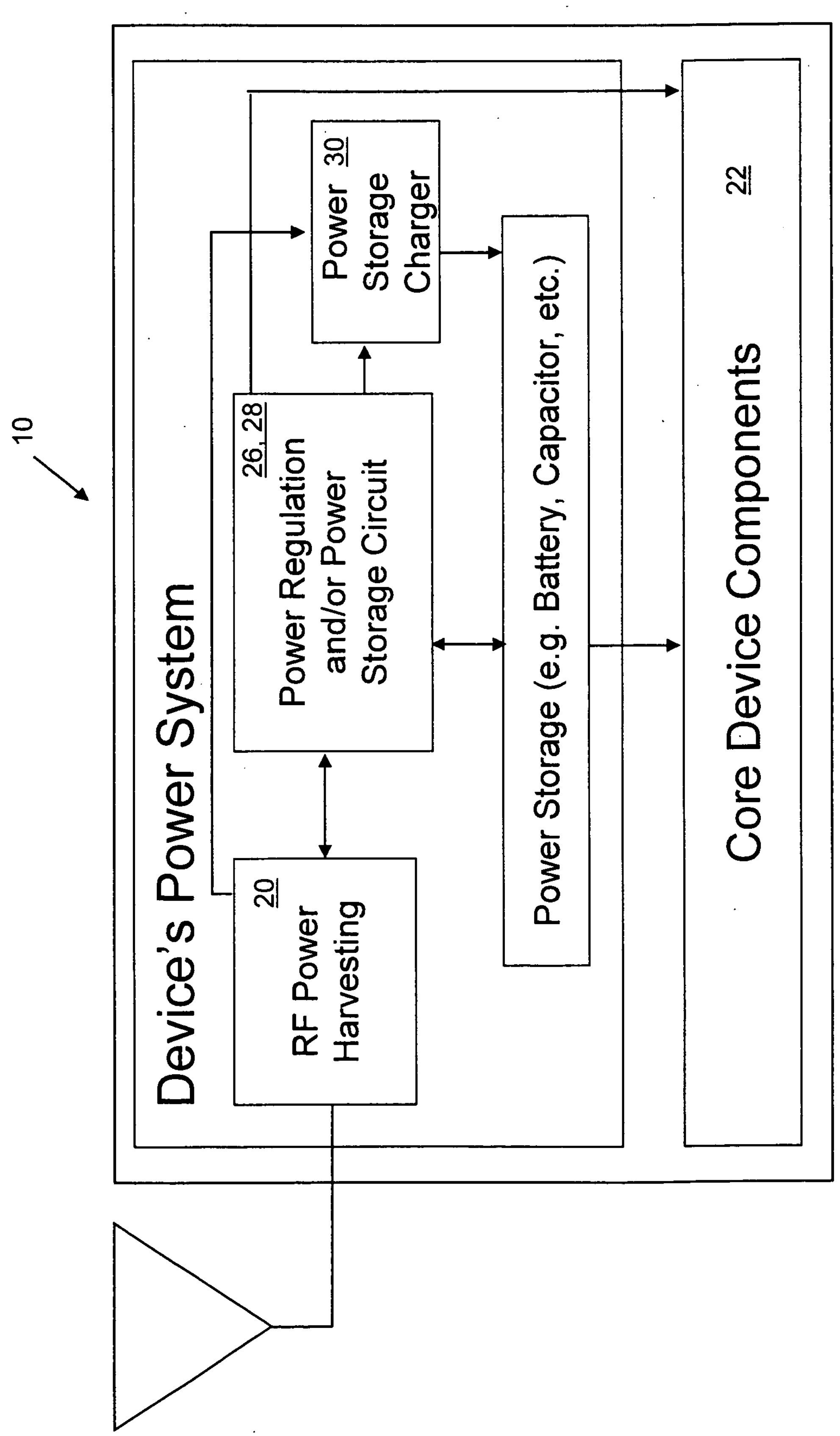
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30 Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 45

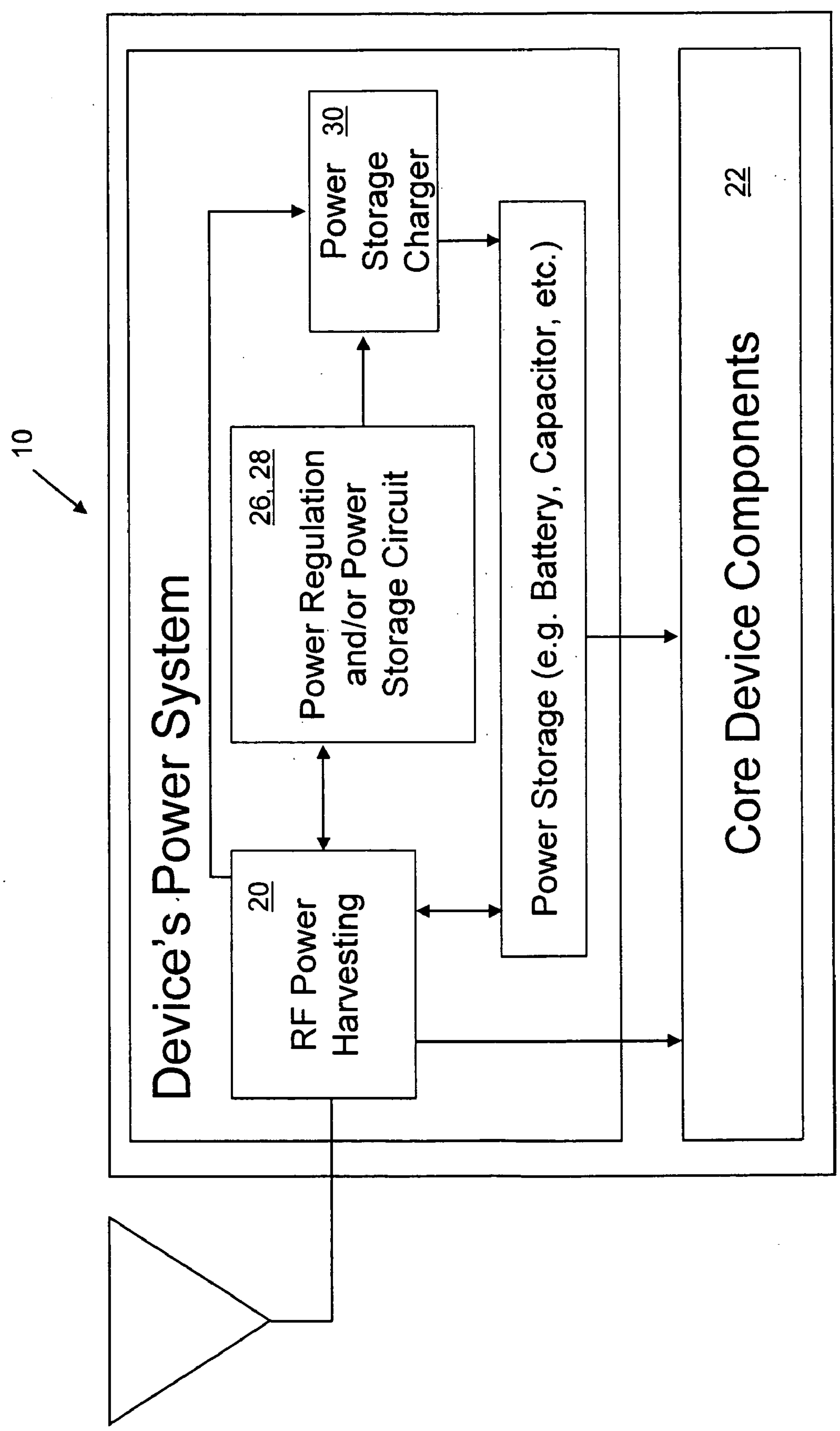
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 46

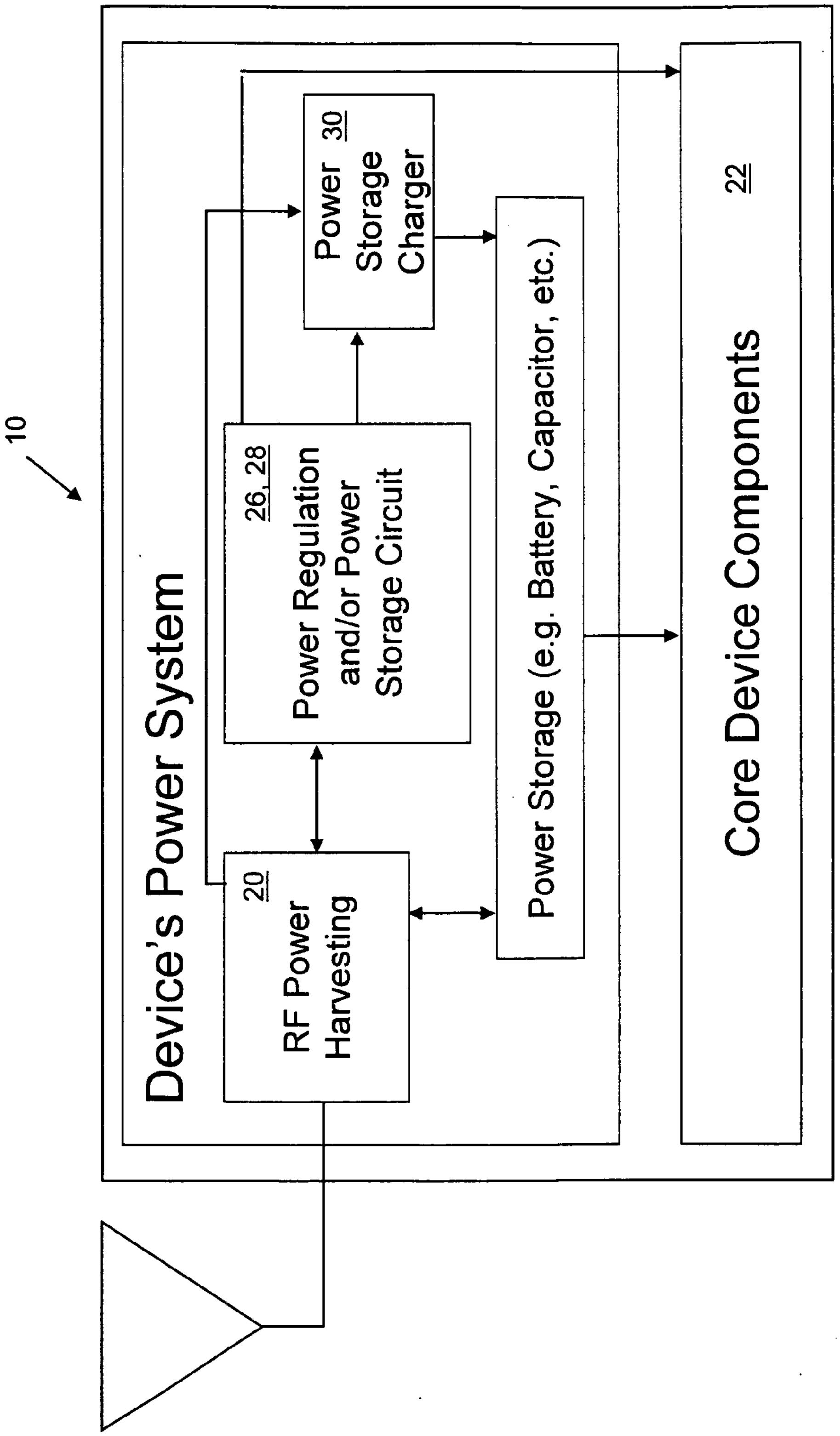
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage
Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 47

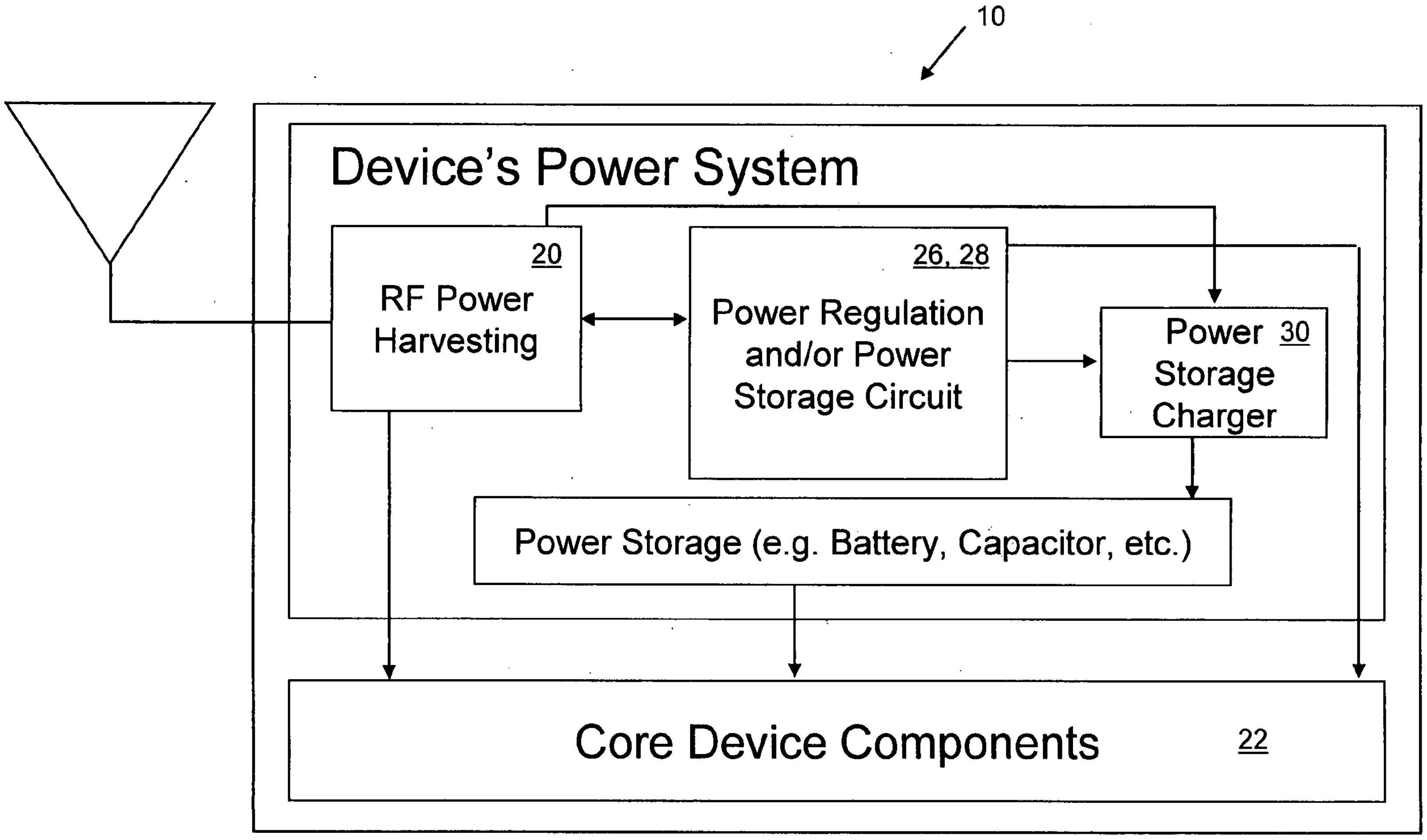
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 48

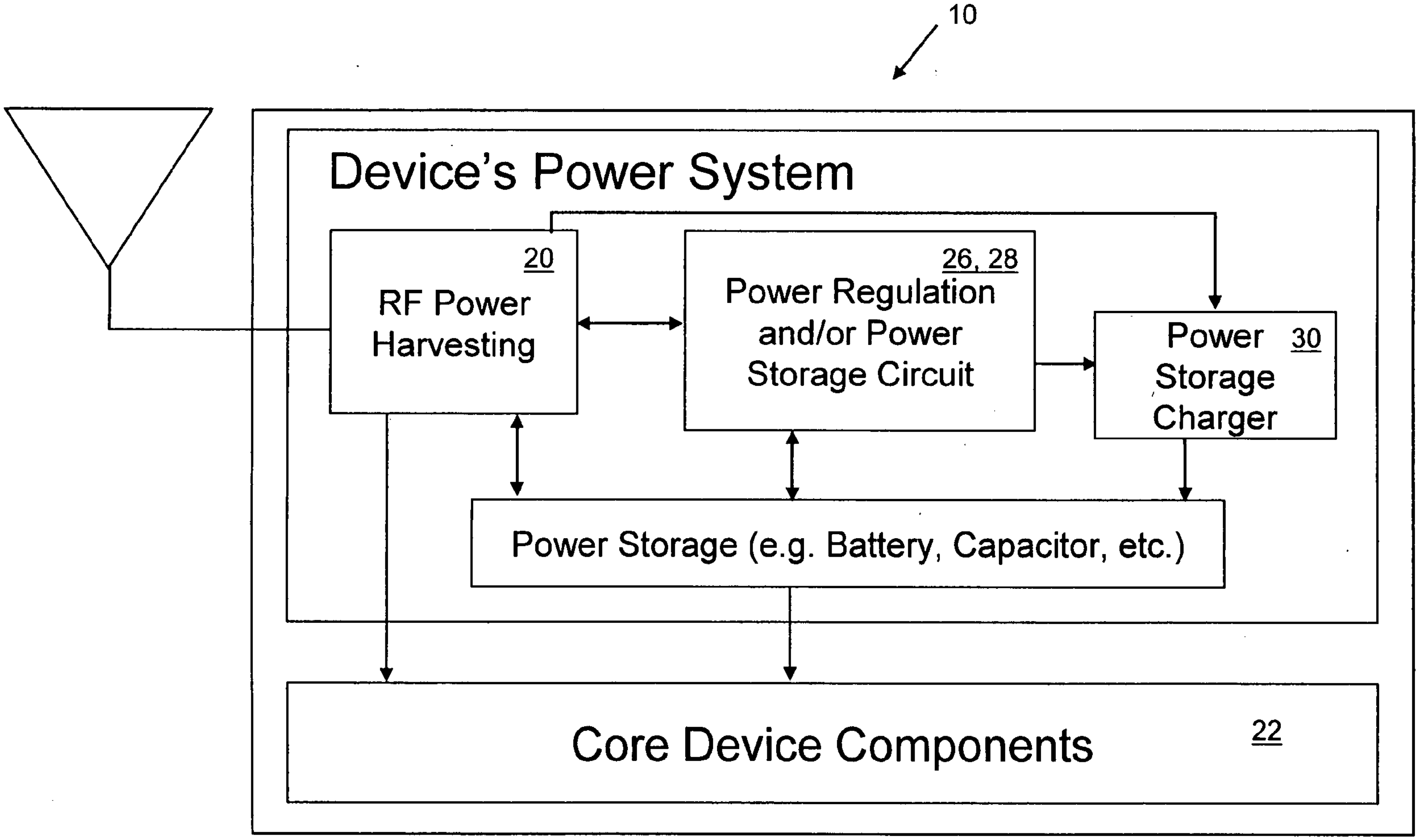
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 49

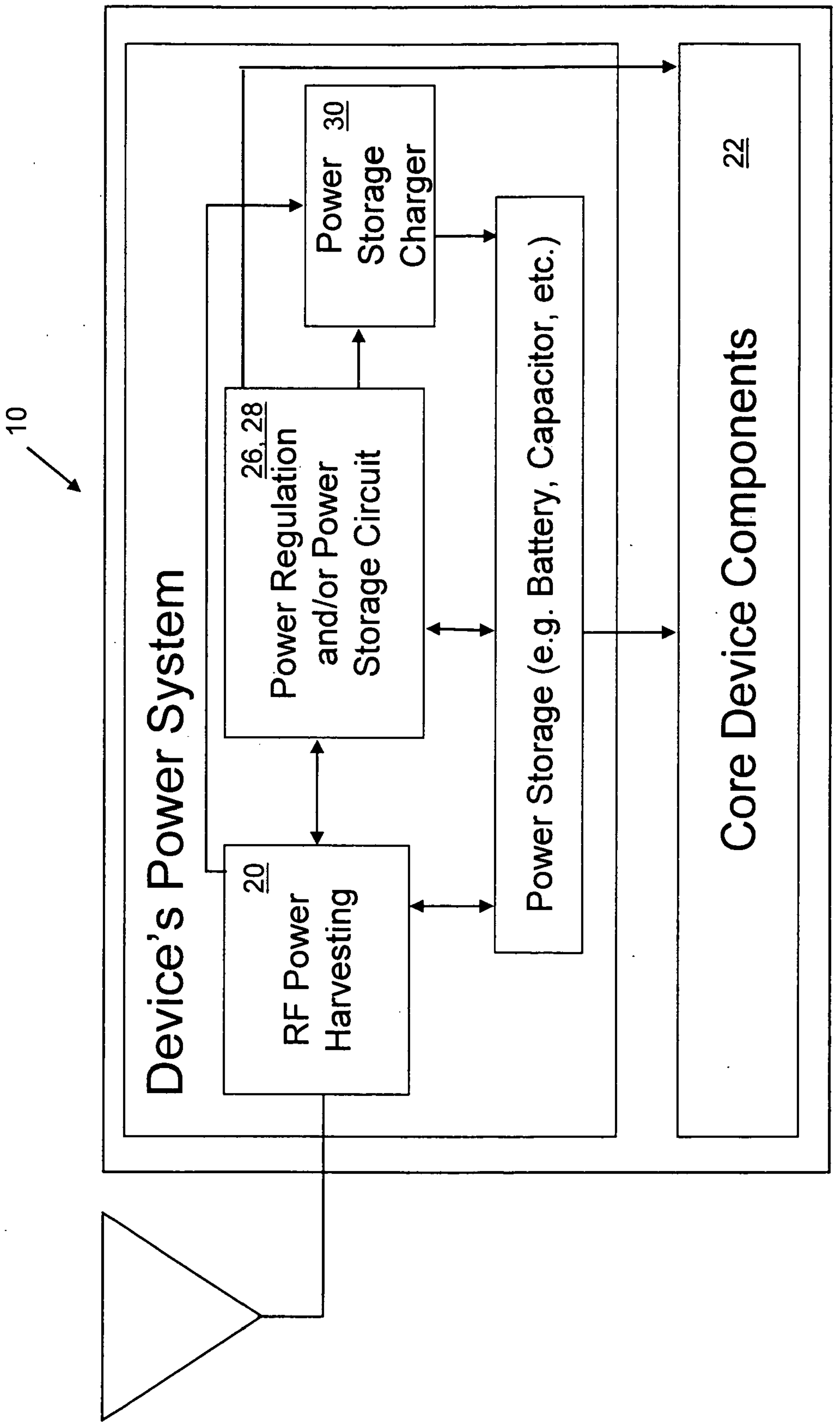
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 50

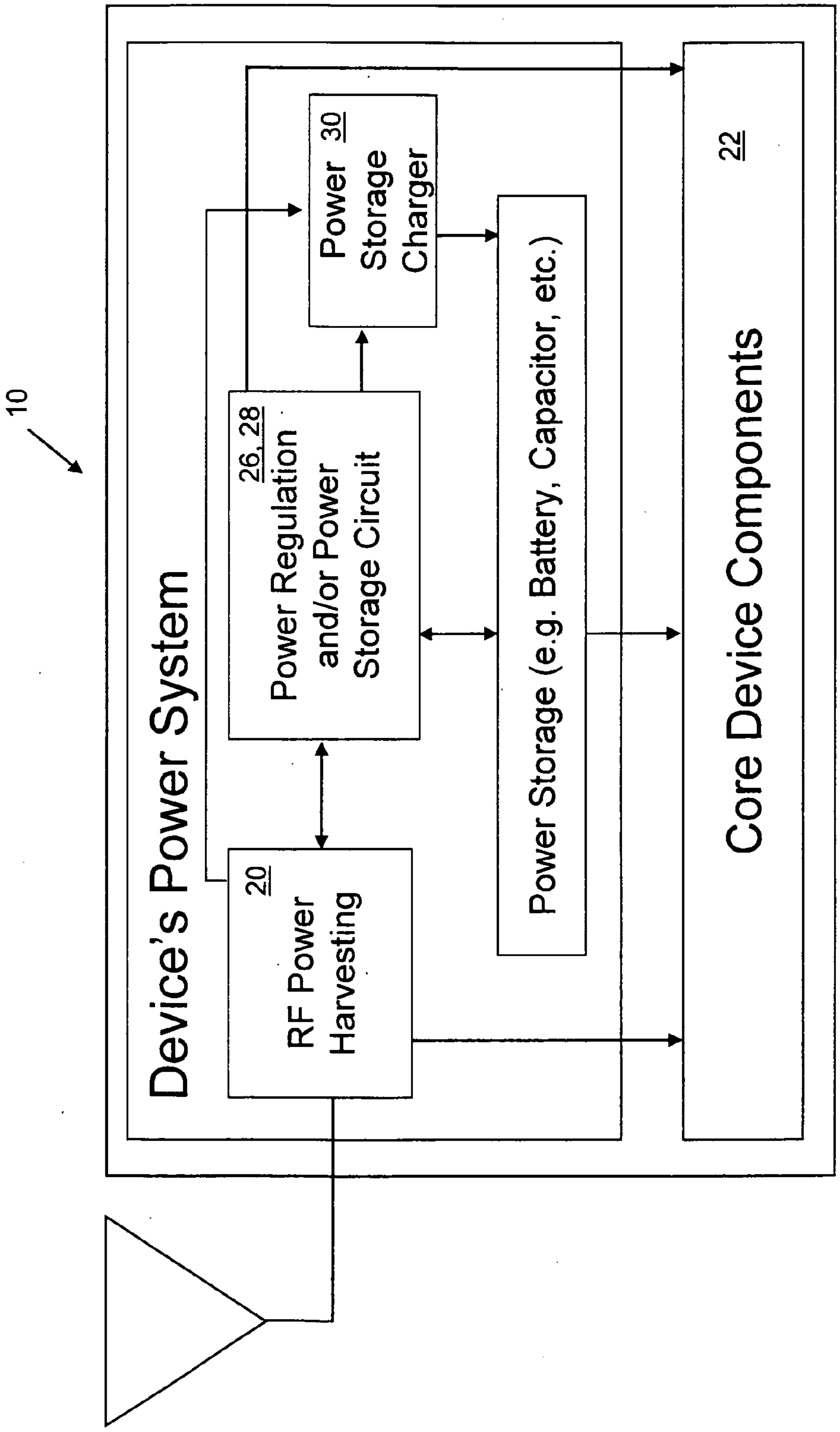
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 51

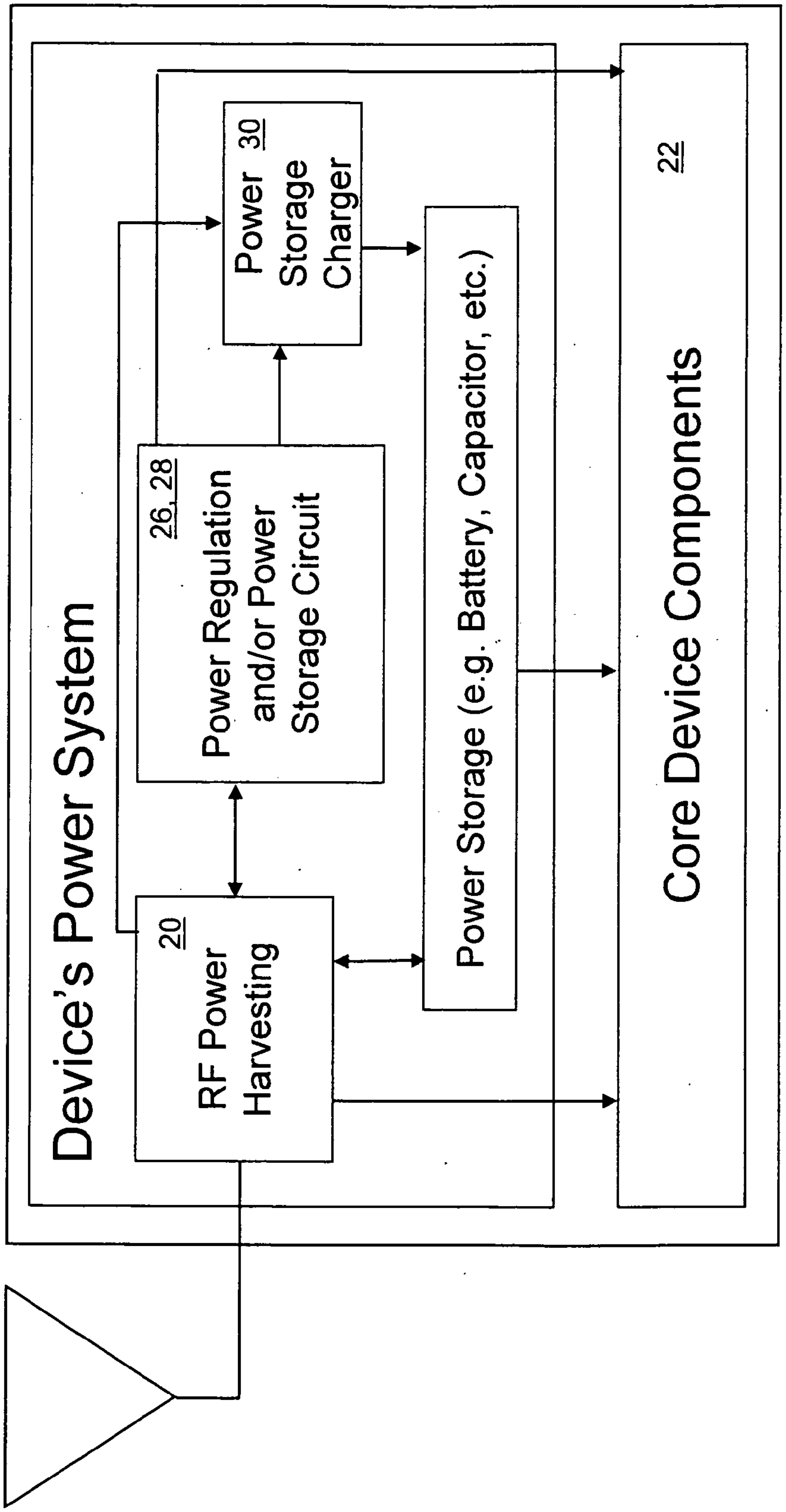
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 52

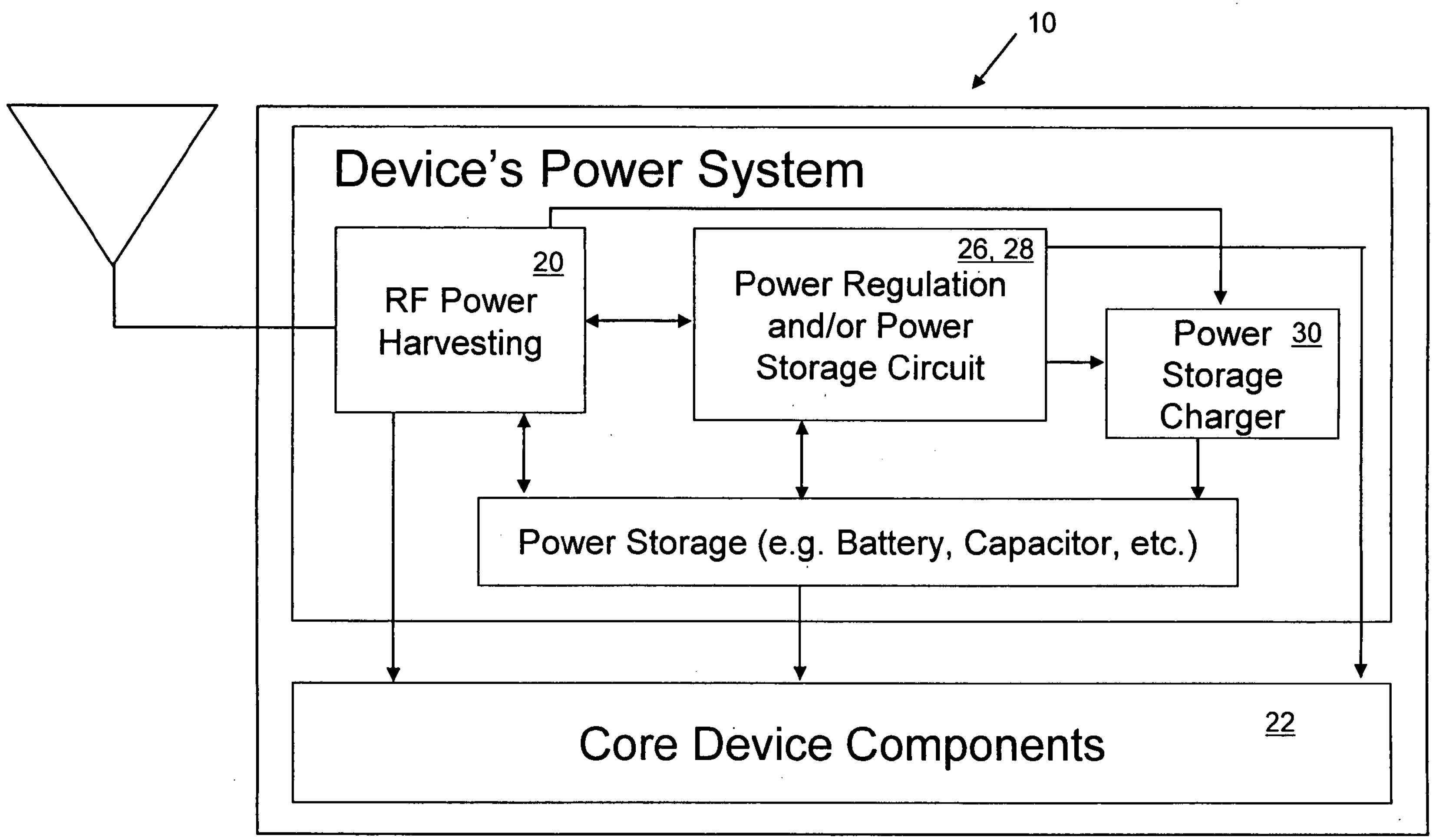
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage
Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 53

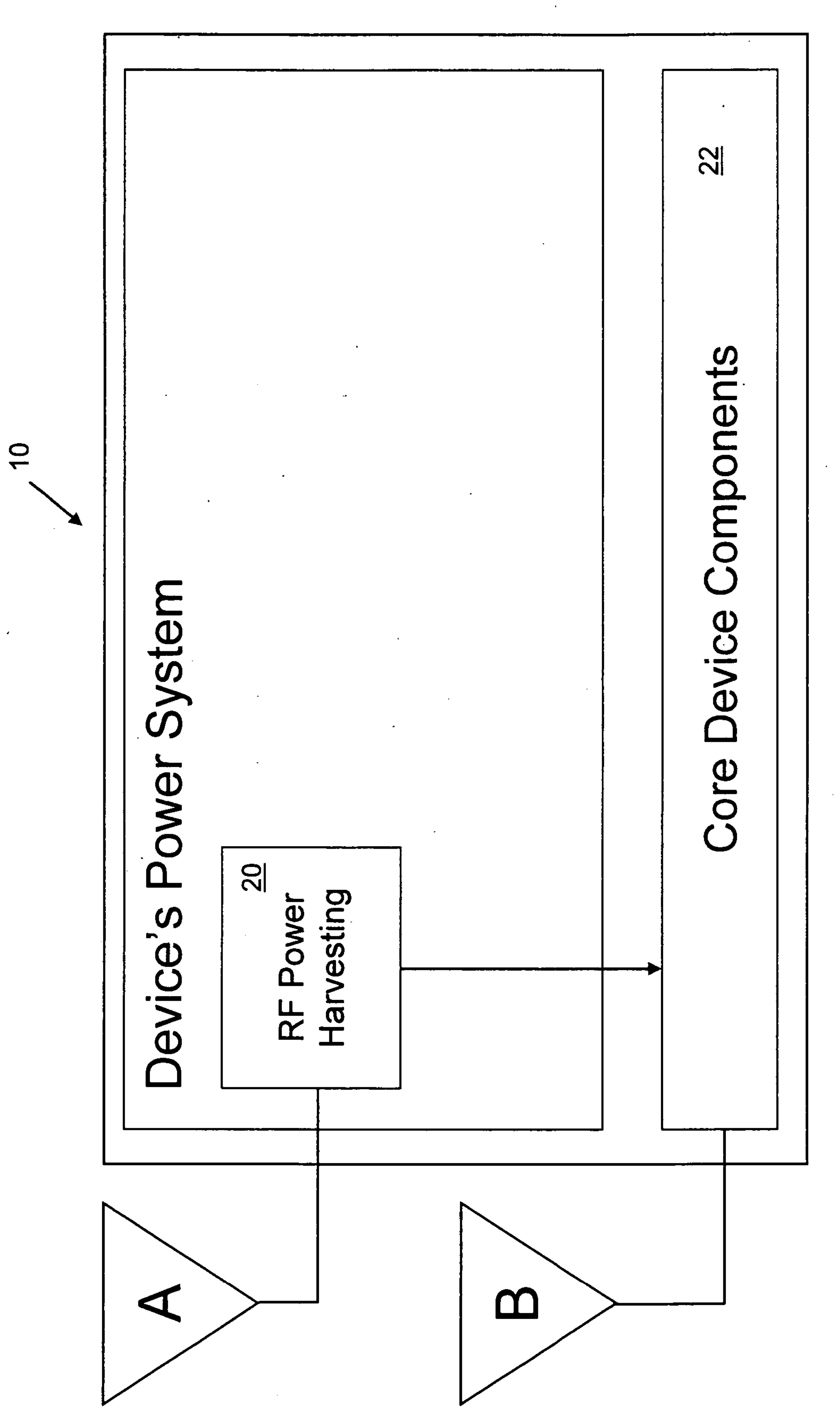
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
A
B

Fig. 54

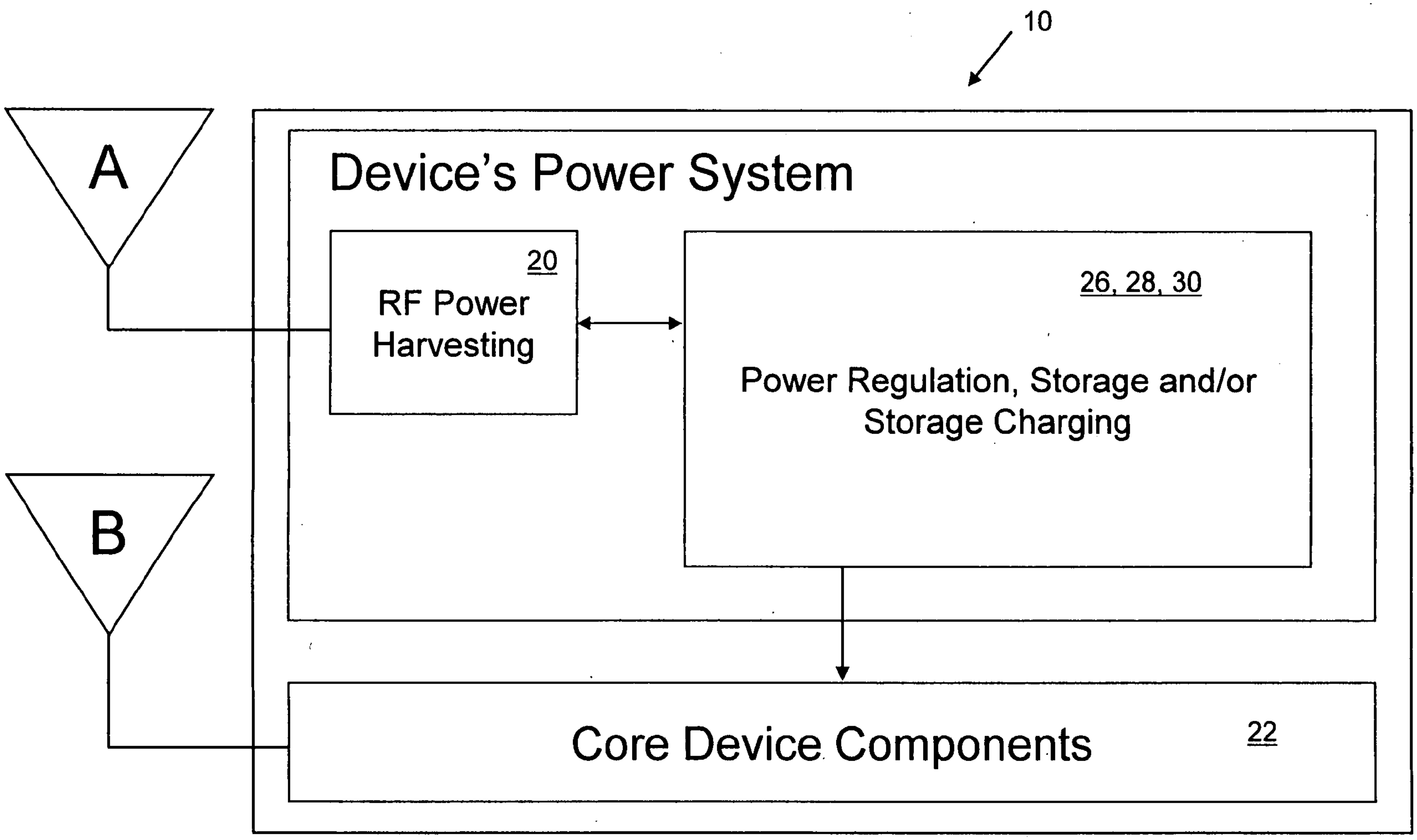
10
A
B
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22

Fig. 55

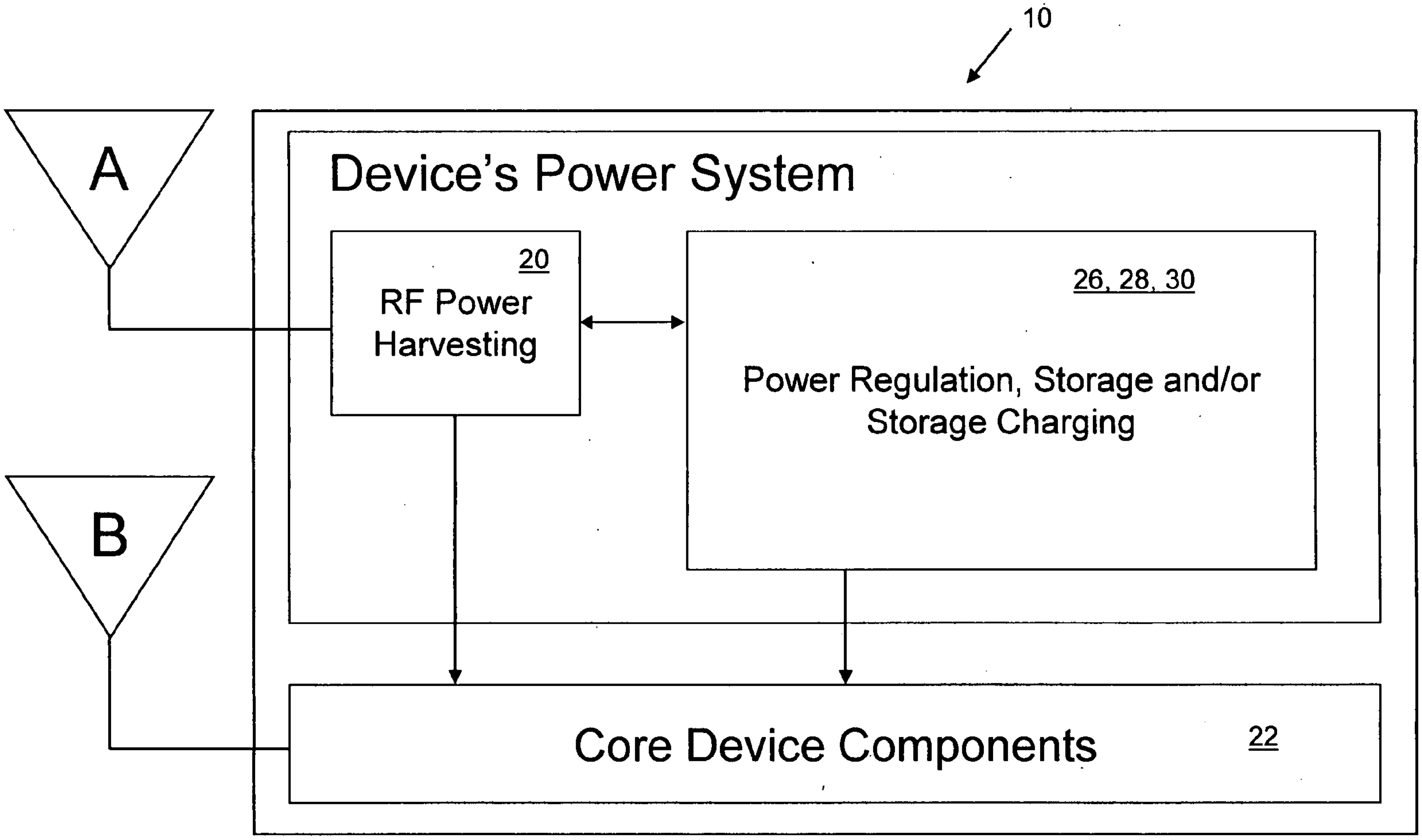
10
A
B
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22

Fig. 56

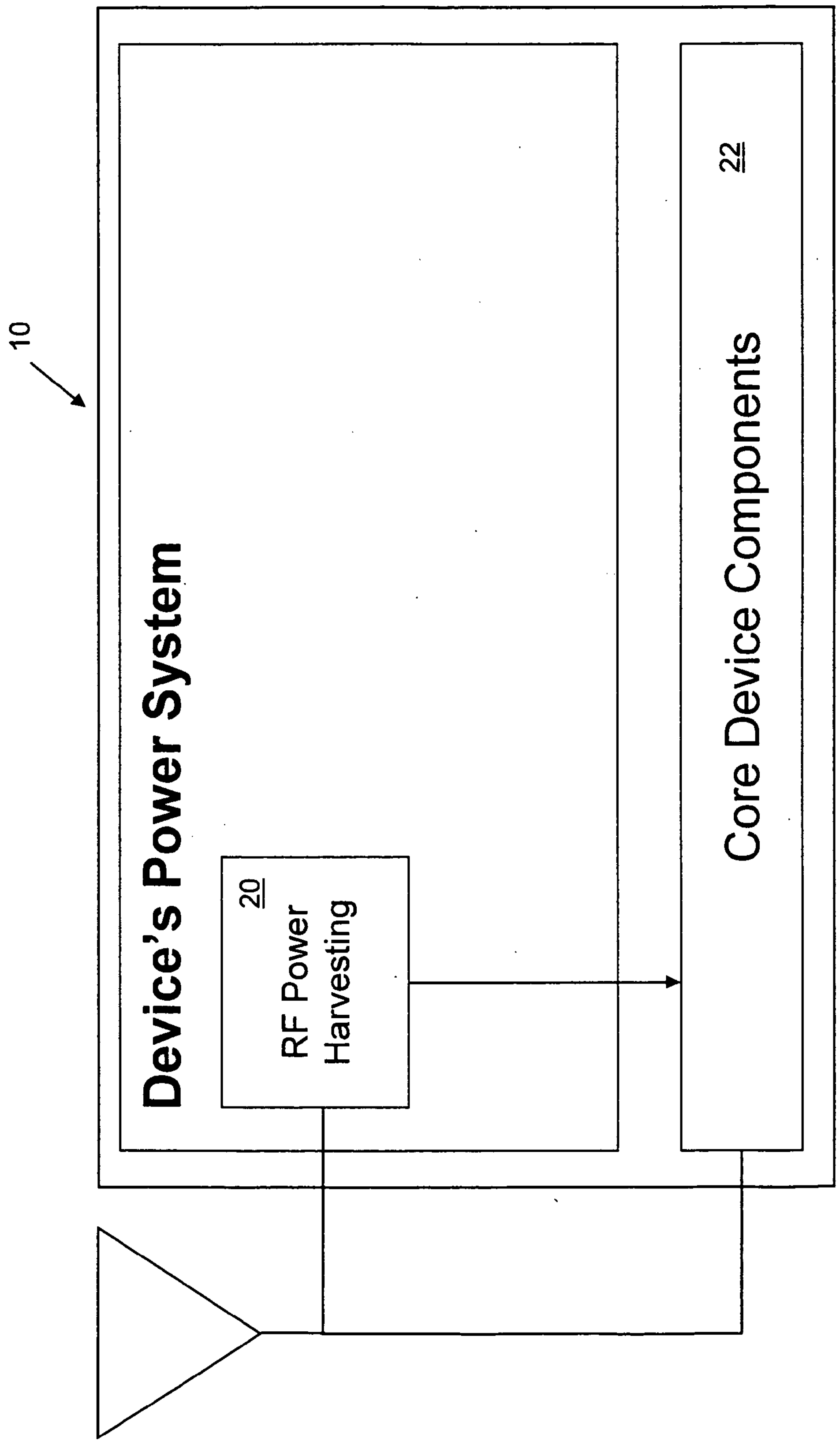
10
Device's Power System
20
RF Power
Harvesting
Core Device Components
22

Fig. 57

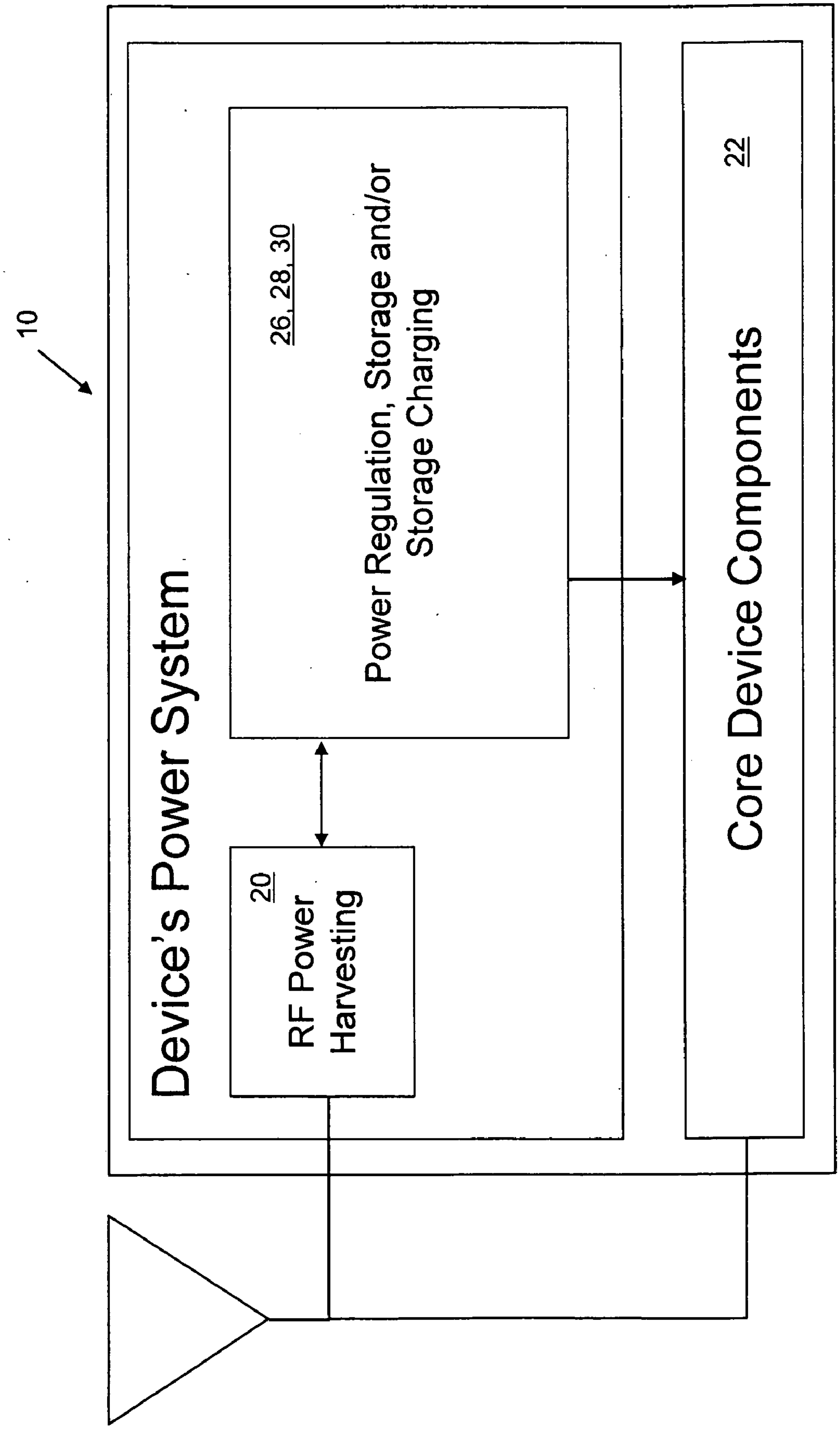
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components

Fig. 58

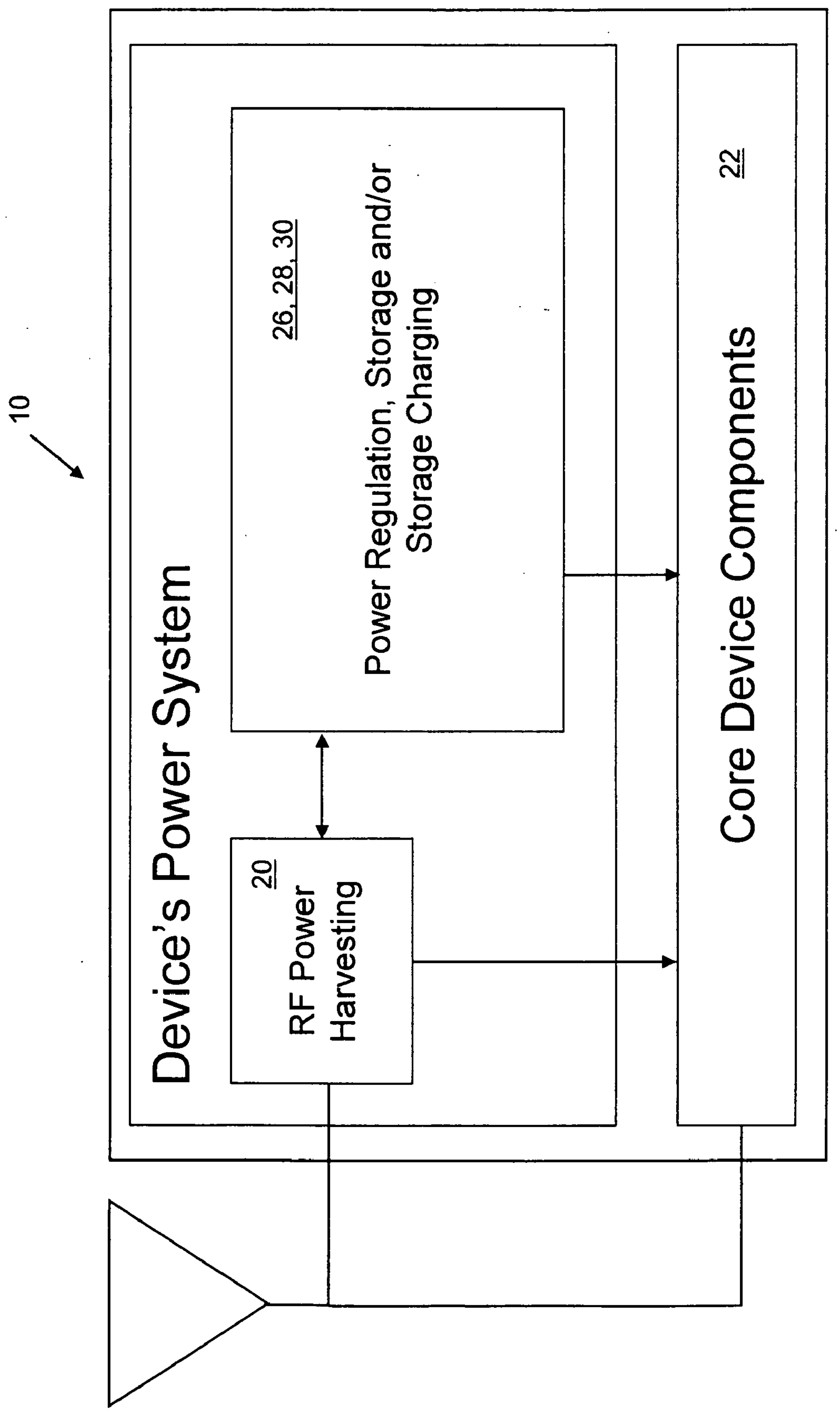
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22

Fig. 59

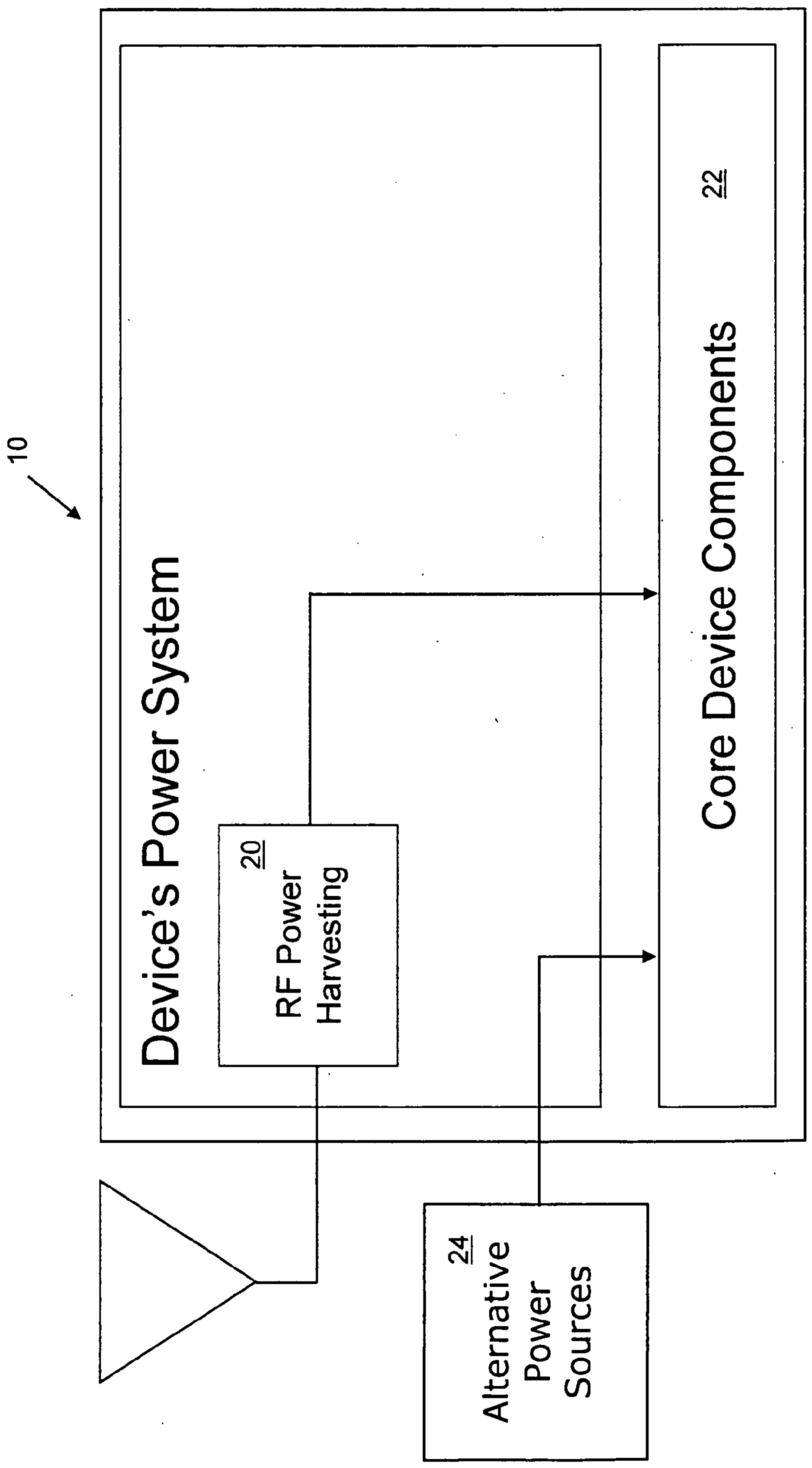
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 60

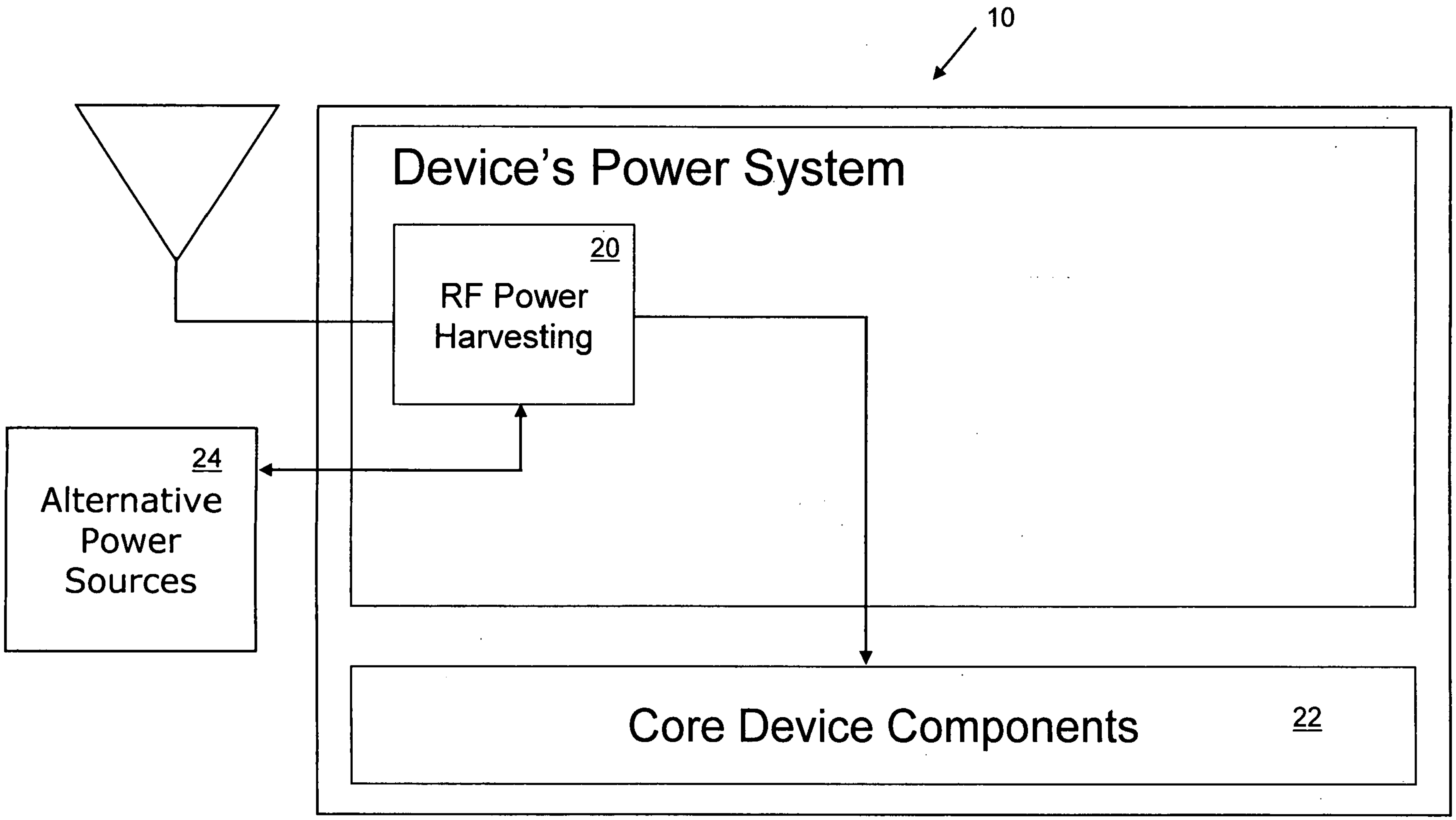
10
Device's Power System
20
RF Power Harvesting
24
Alternative Power Sources
Core Device Components
22

Fig. 61

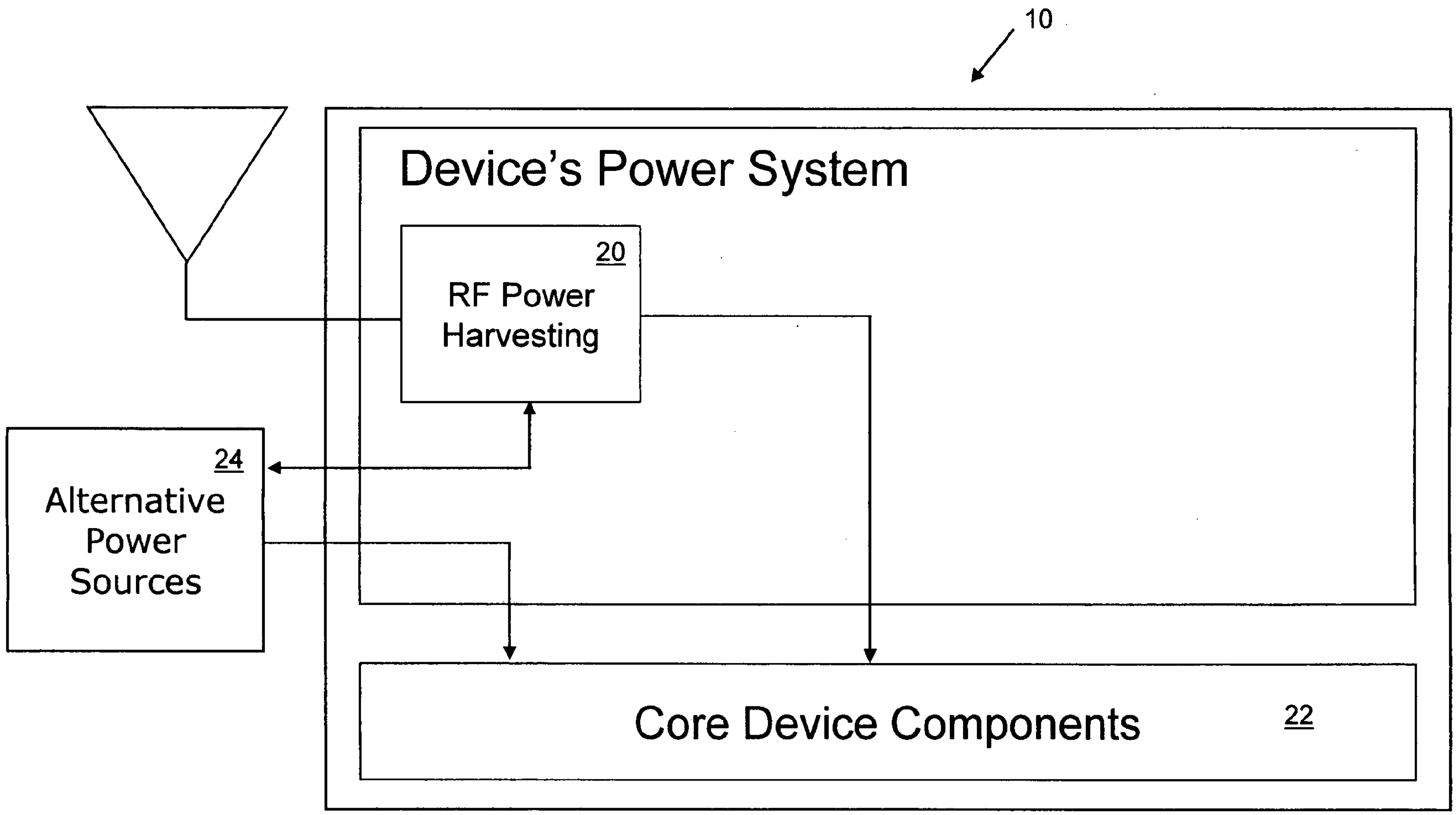
10
Device's Power System
20
RF Power Harvesting
24
Alternative Power Sources
Core Device Components
22

Fig. 62

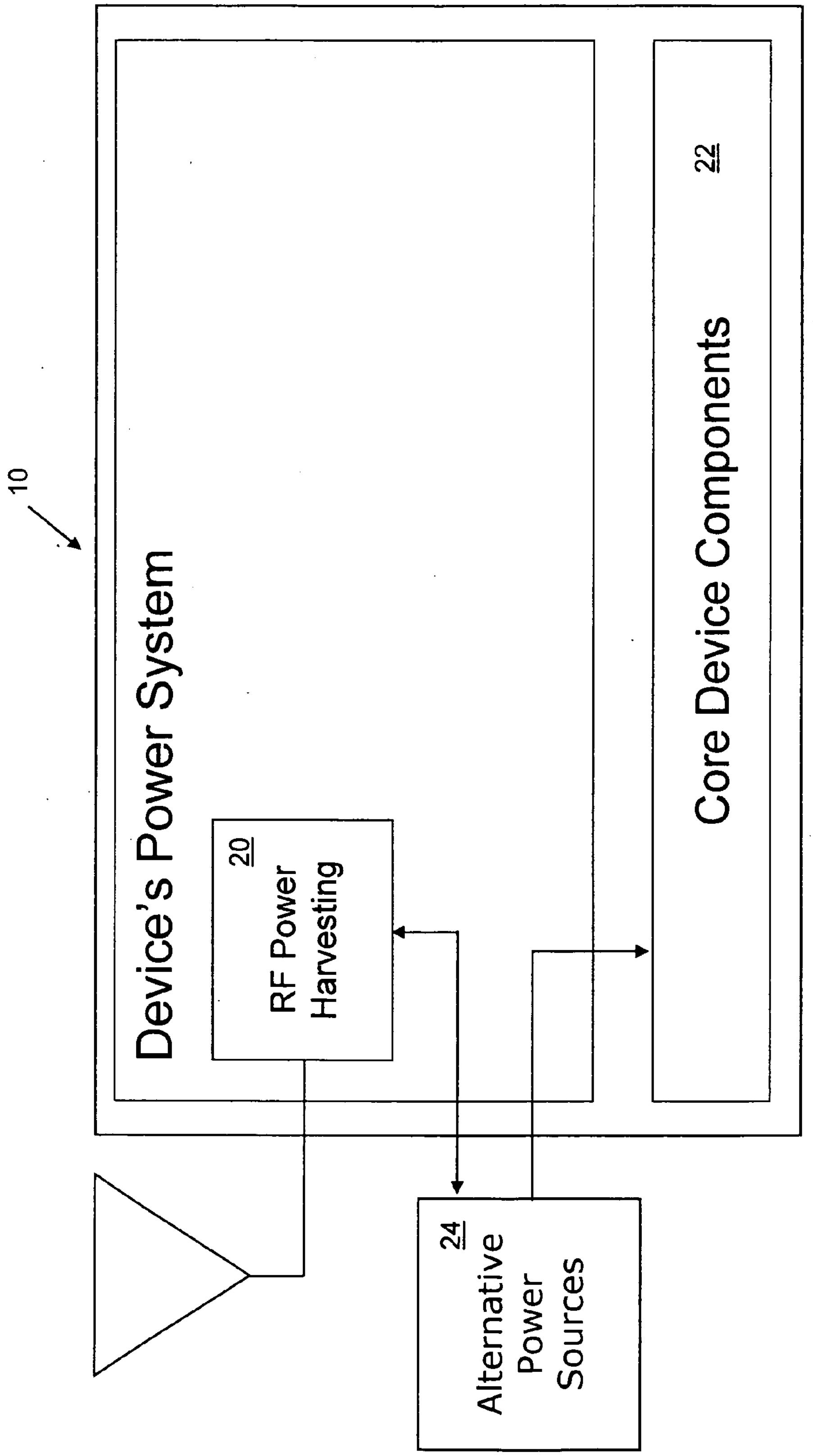
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 63

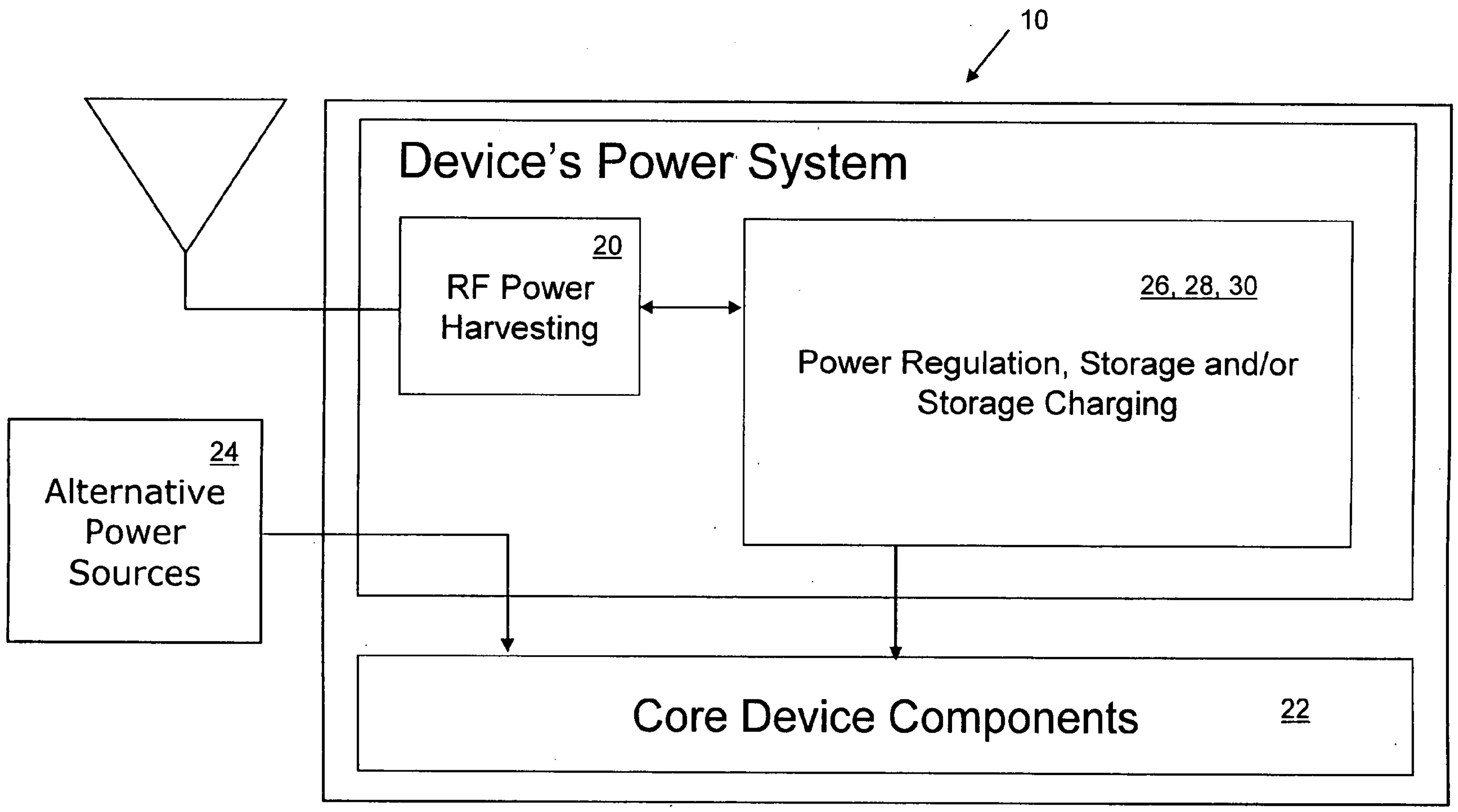
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 64

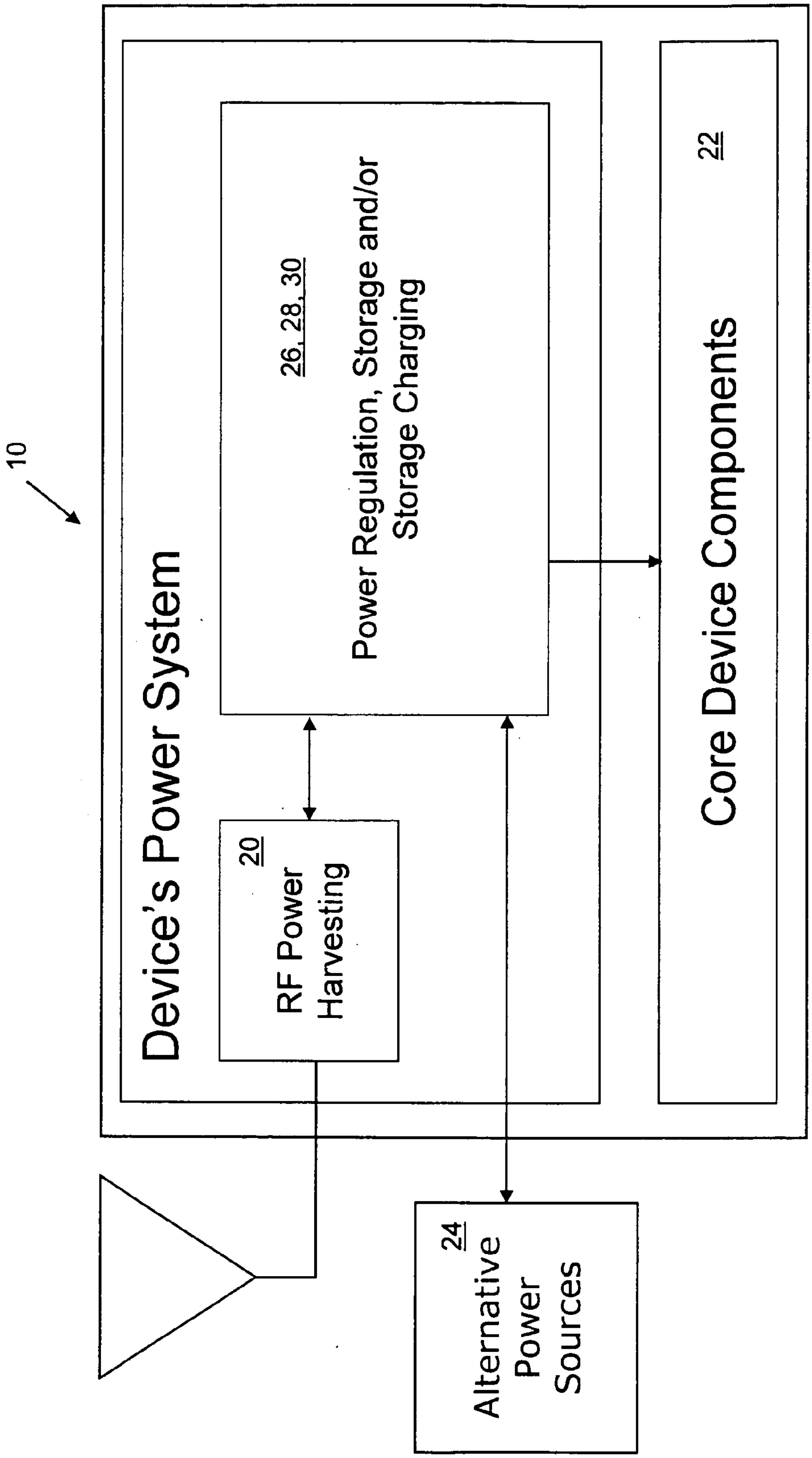
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 65

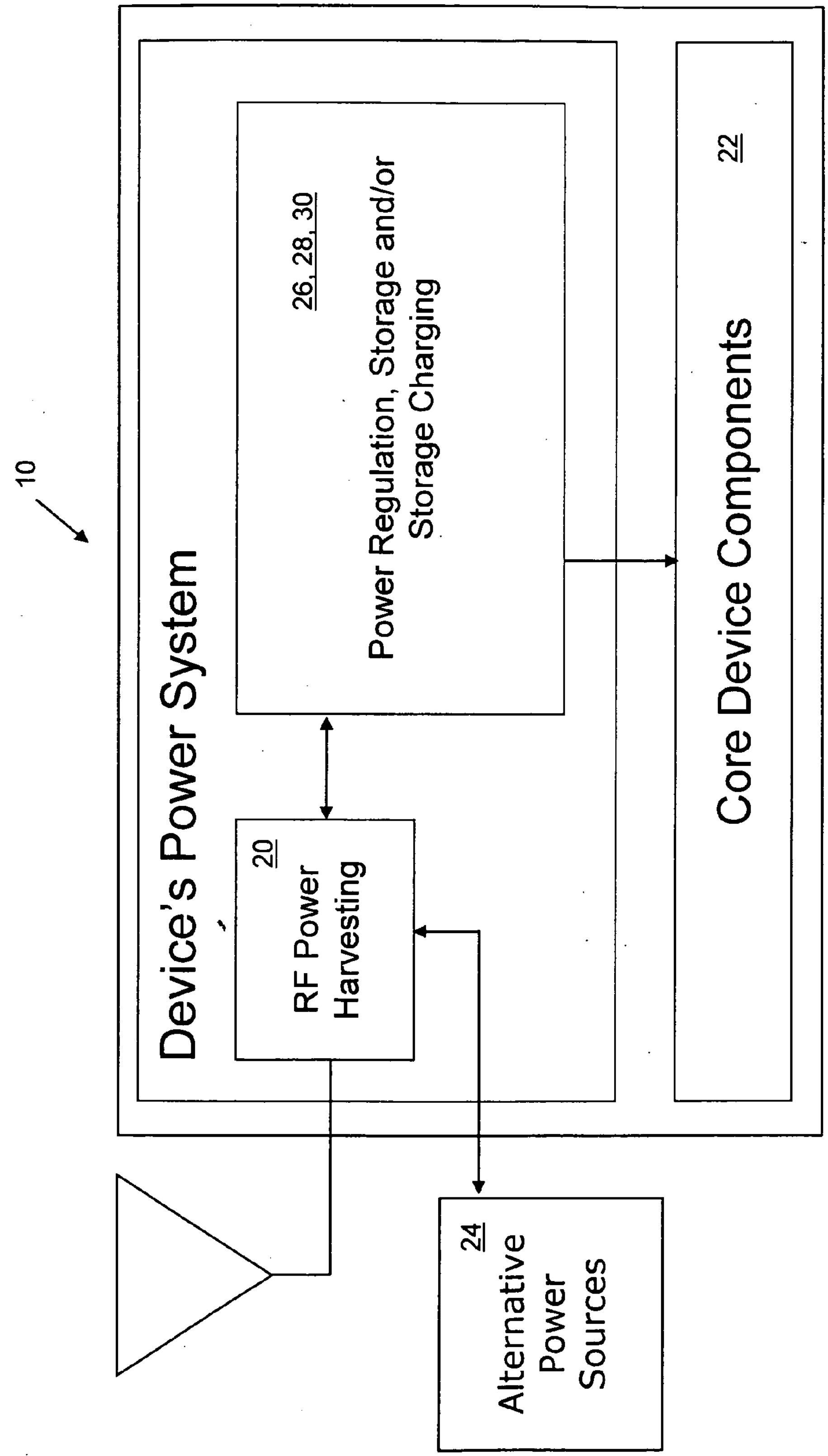
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 66

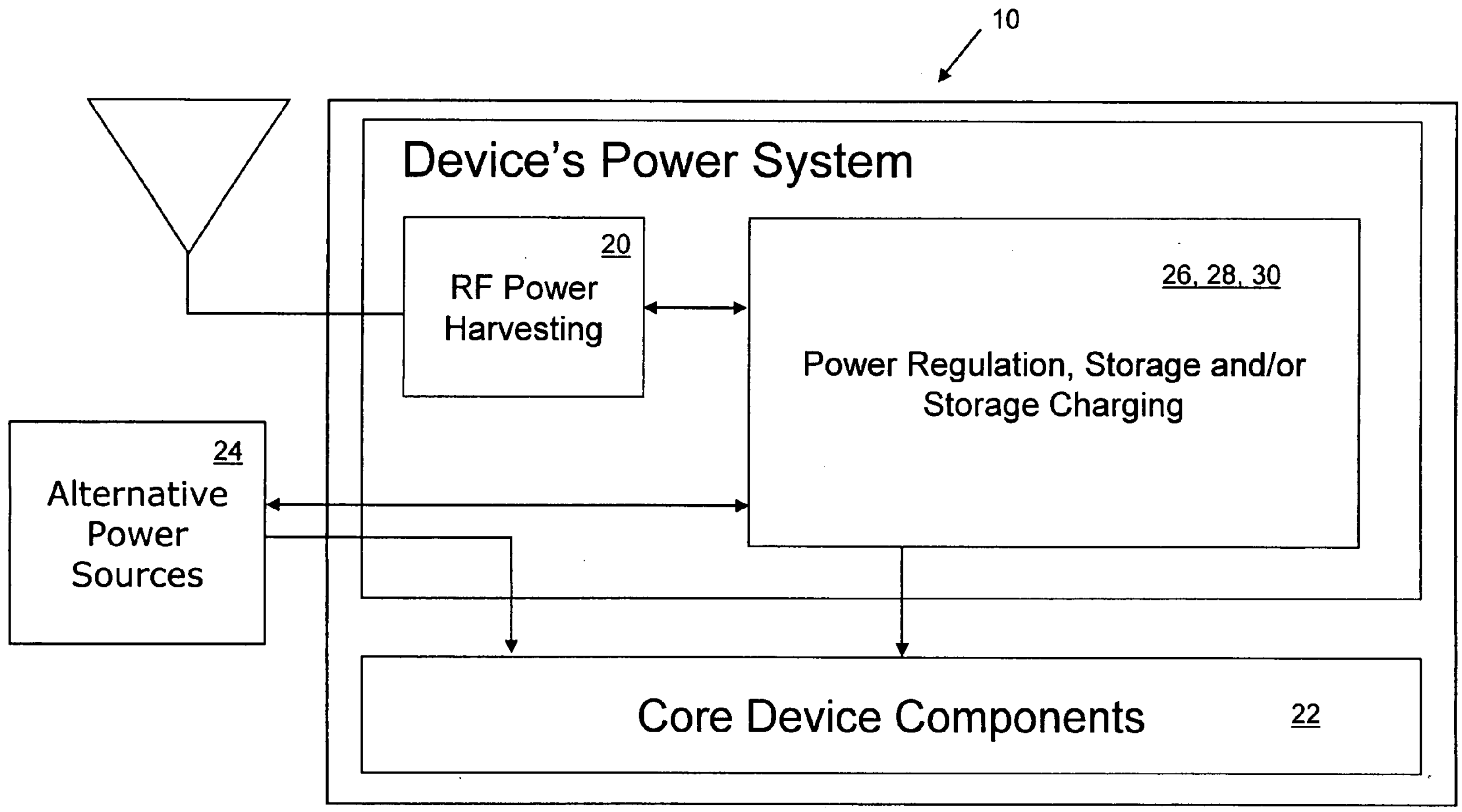
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 67

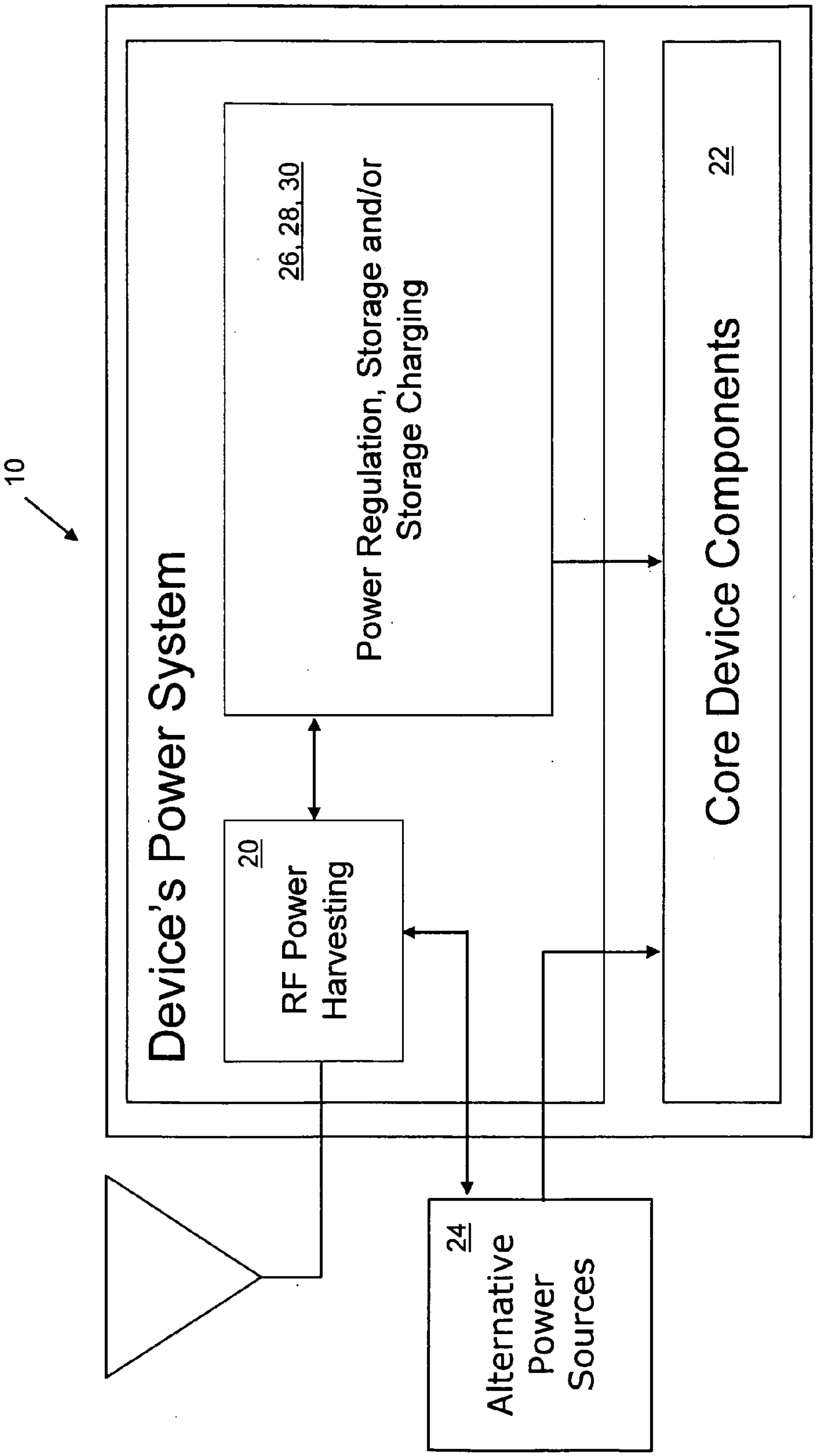
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 68

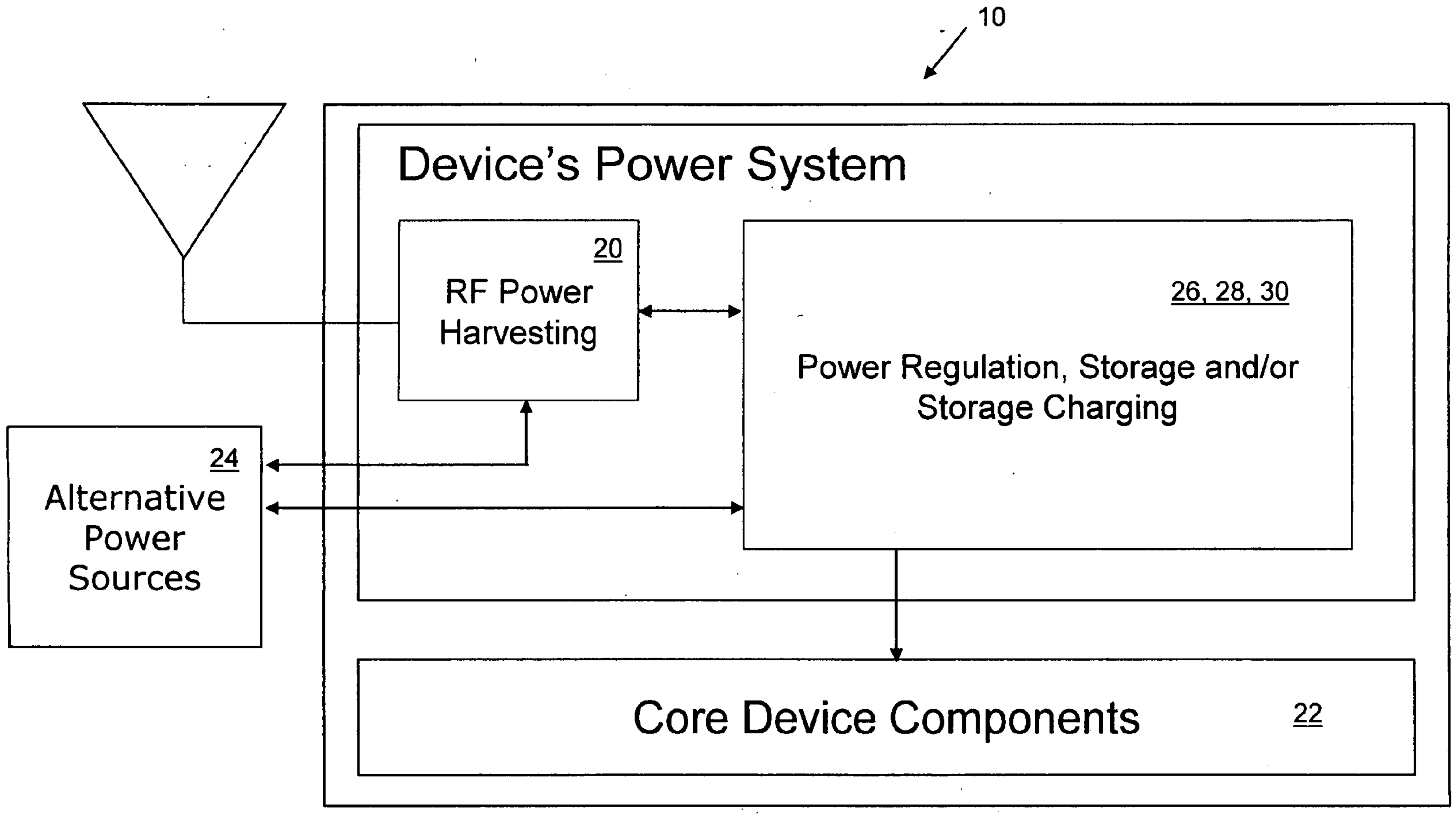
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 69

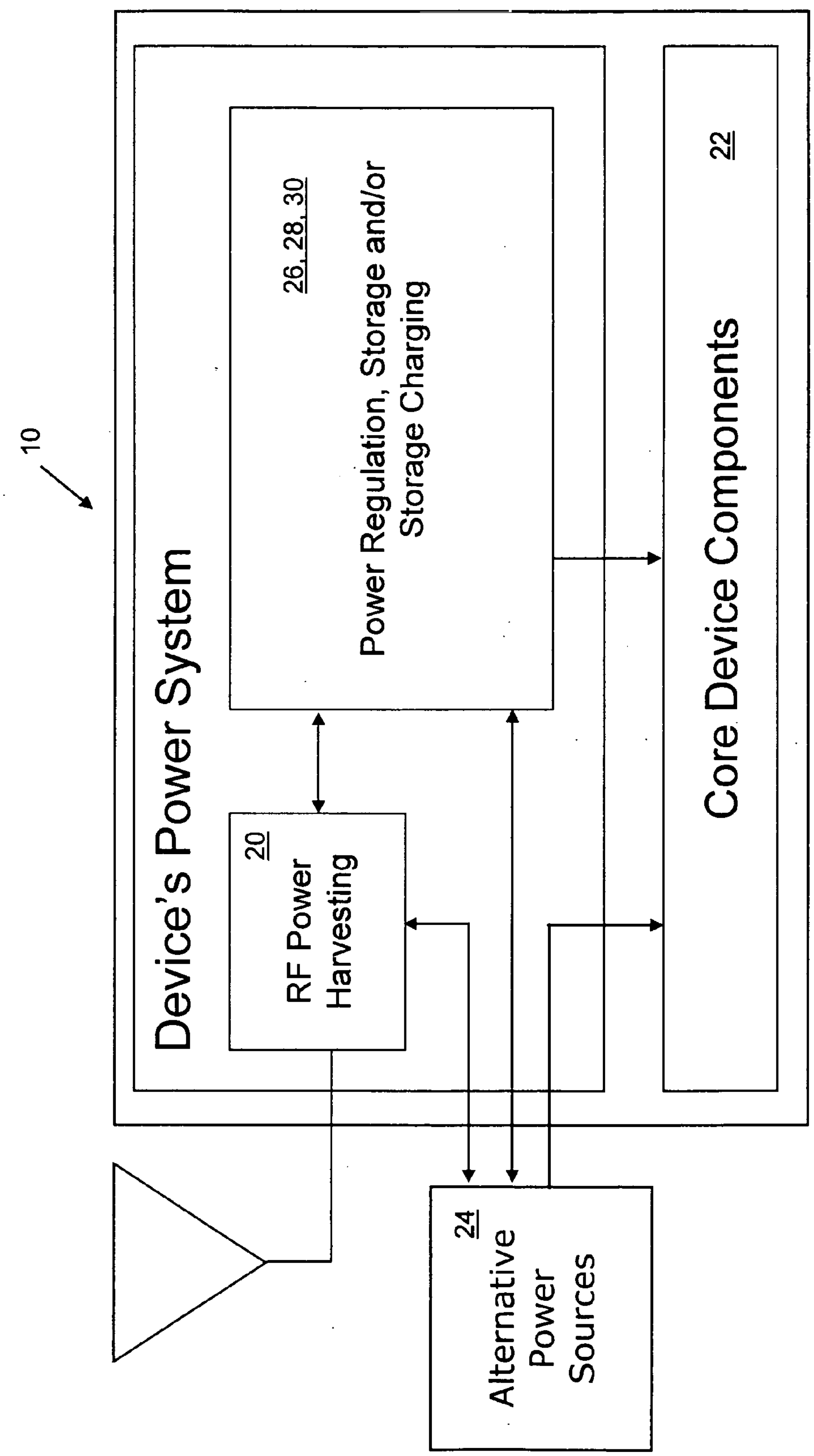
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 70

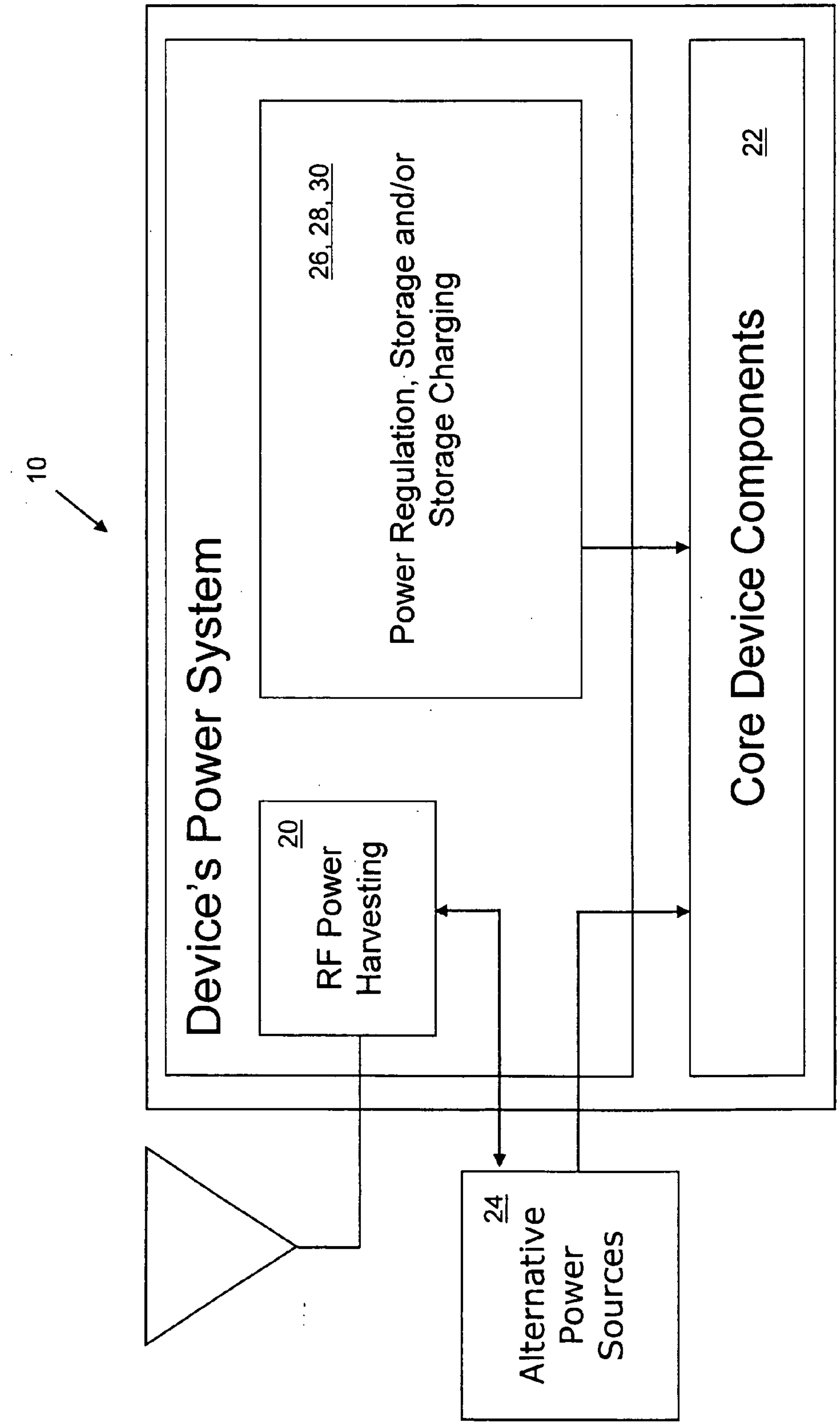
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 71

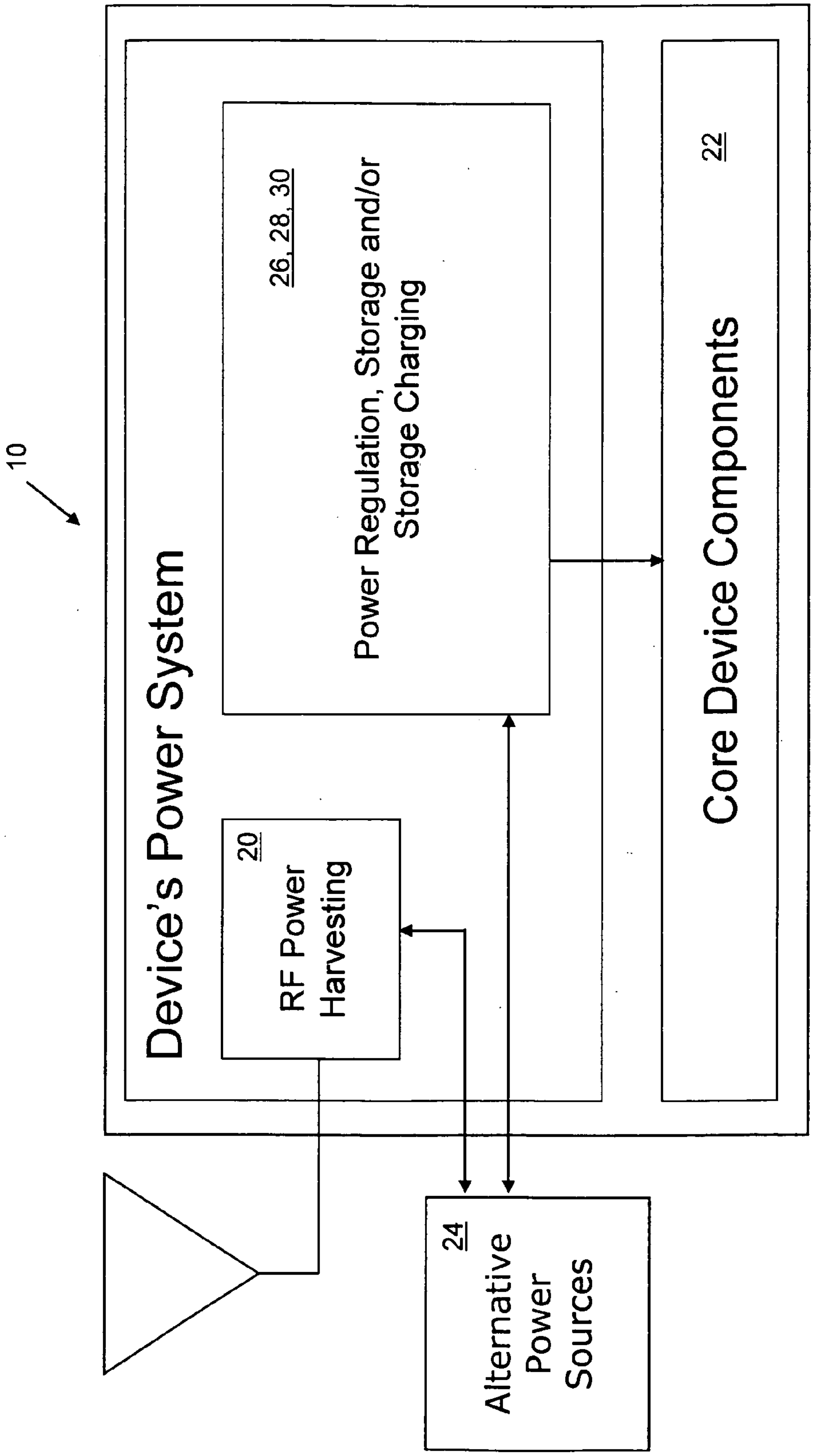
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 72

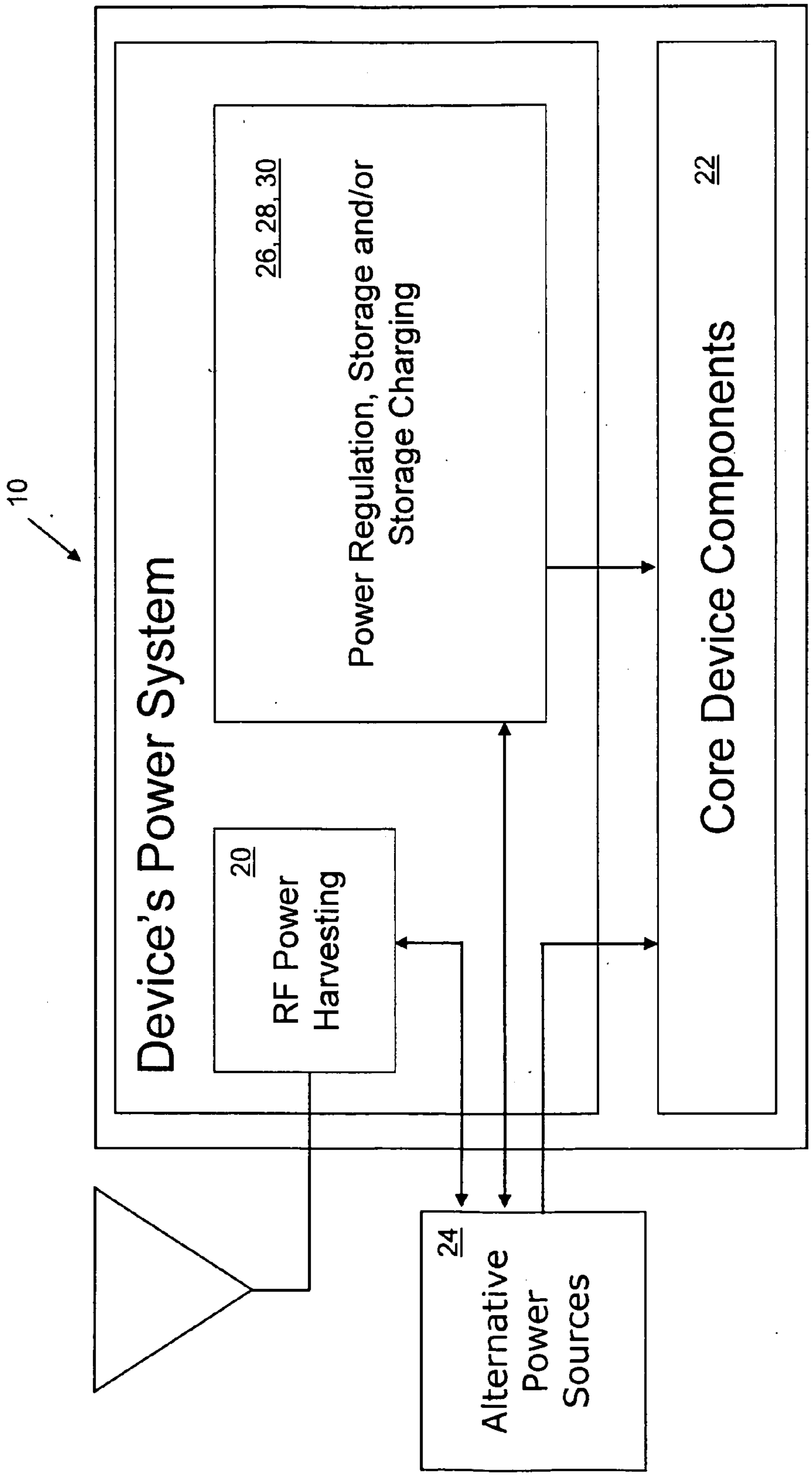
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 73

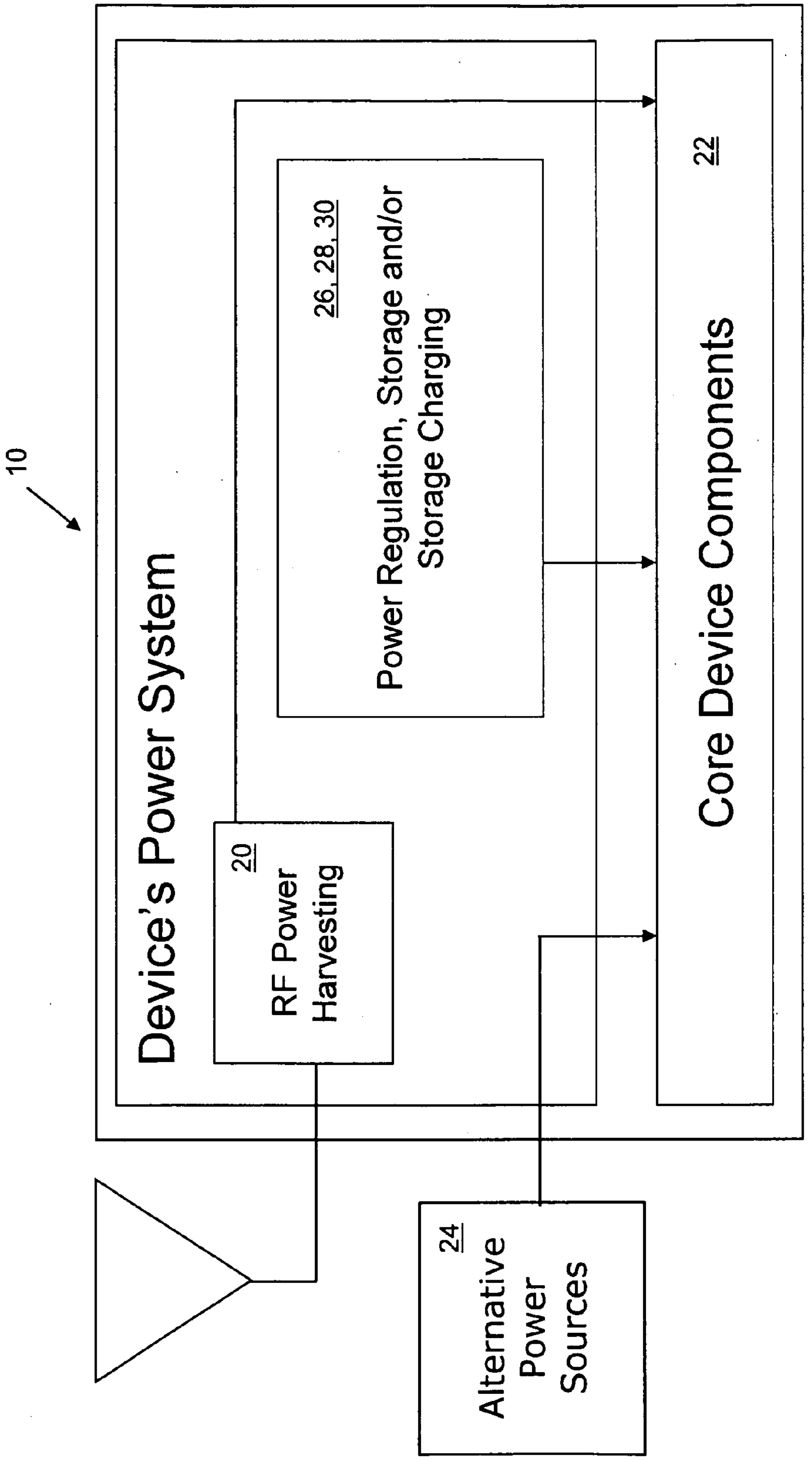
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 74

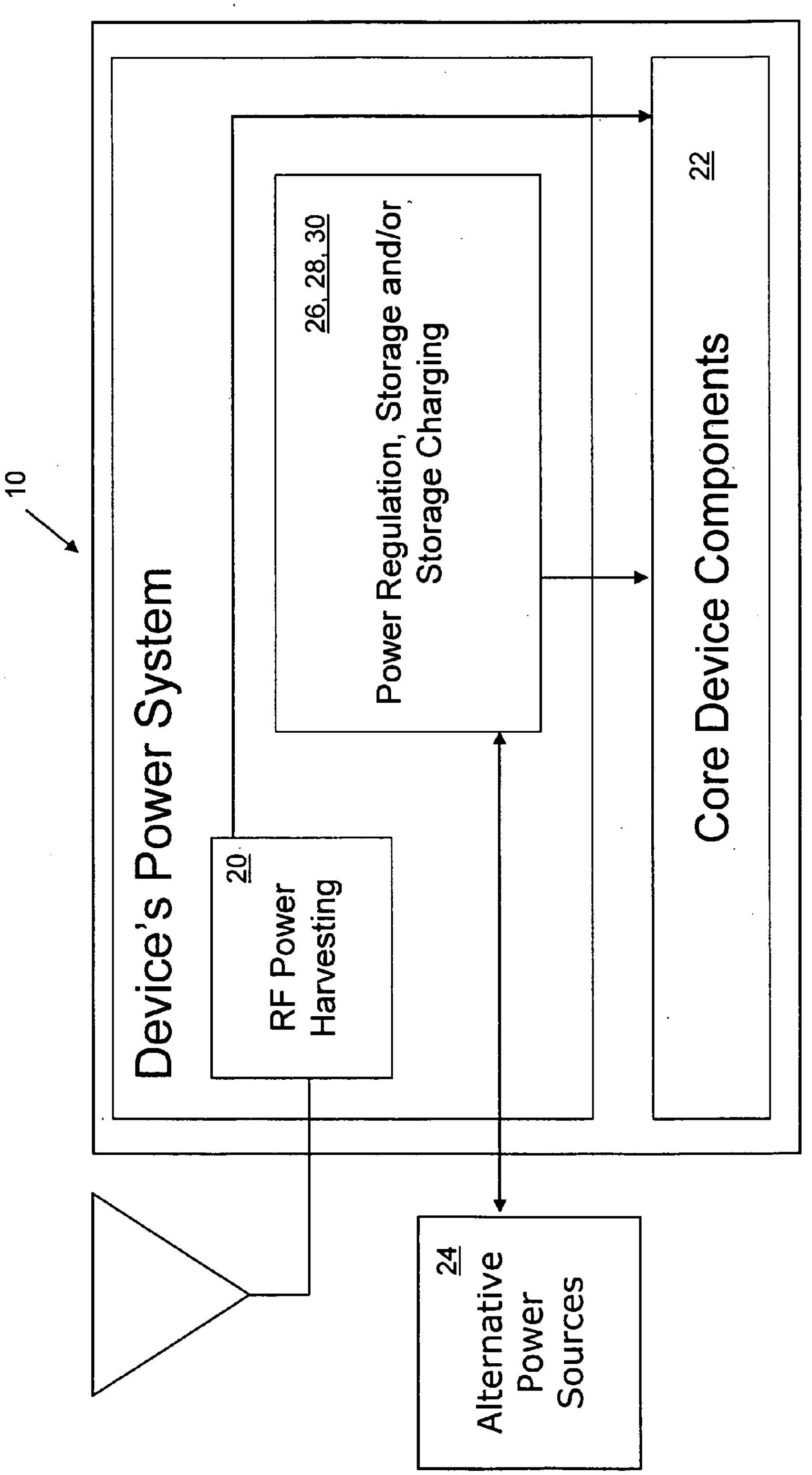
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 75

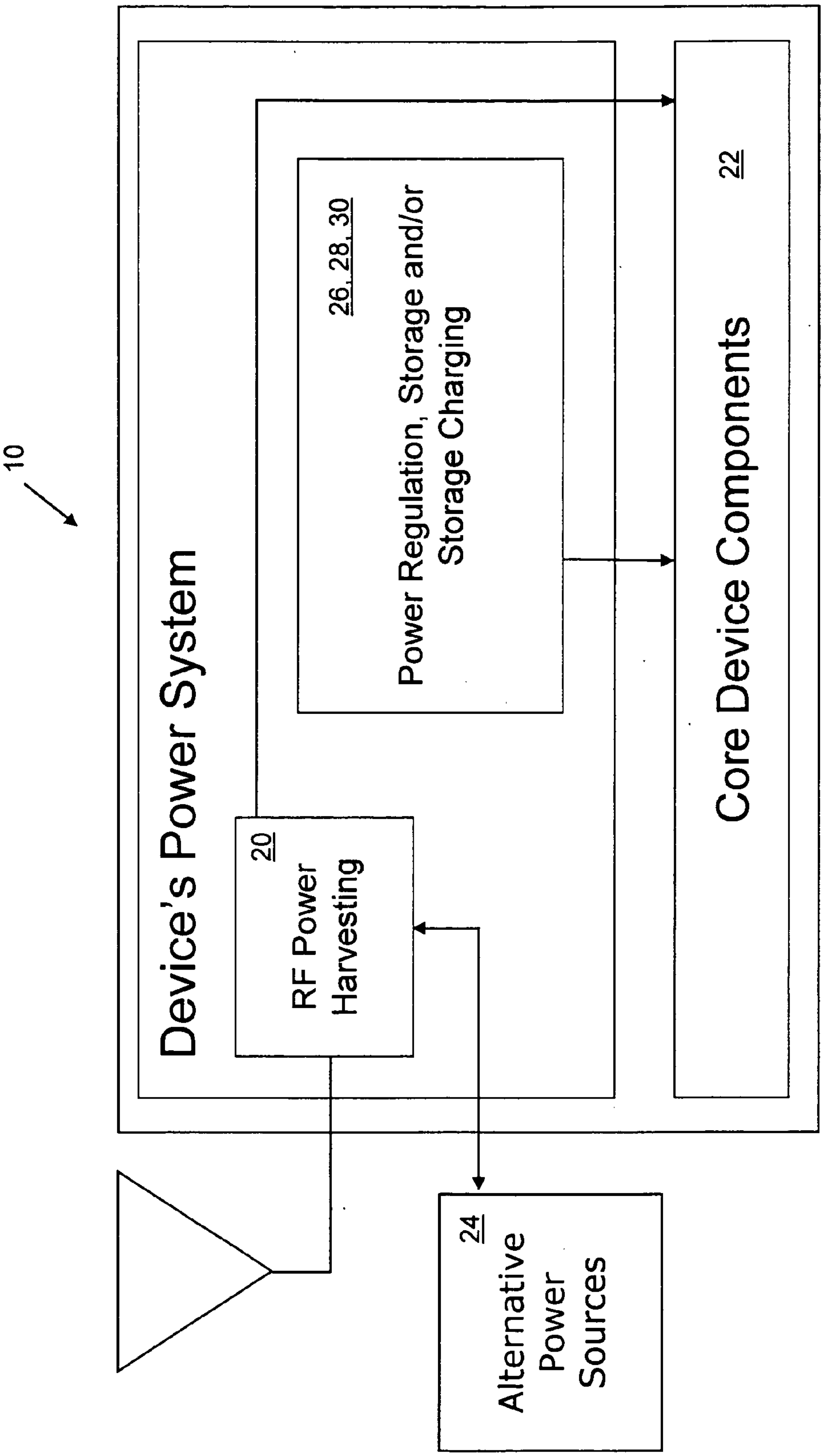
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 76

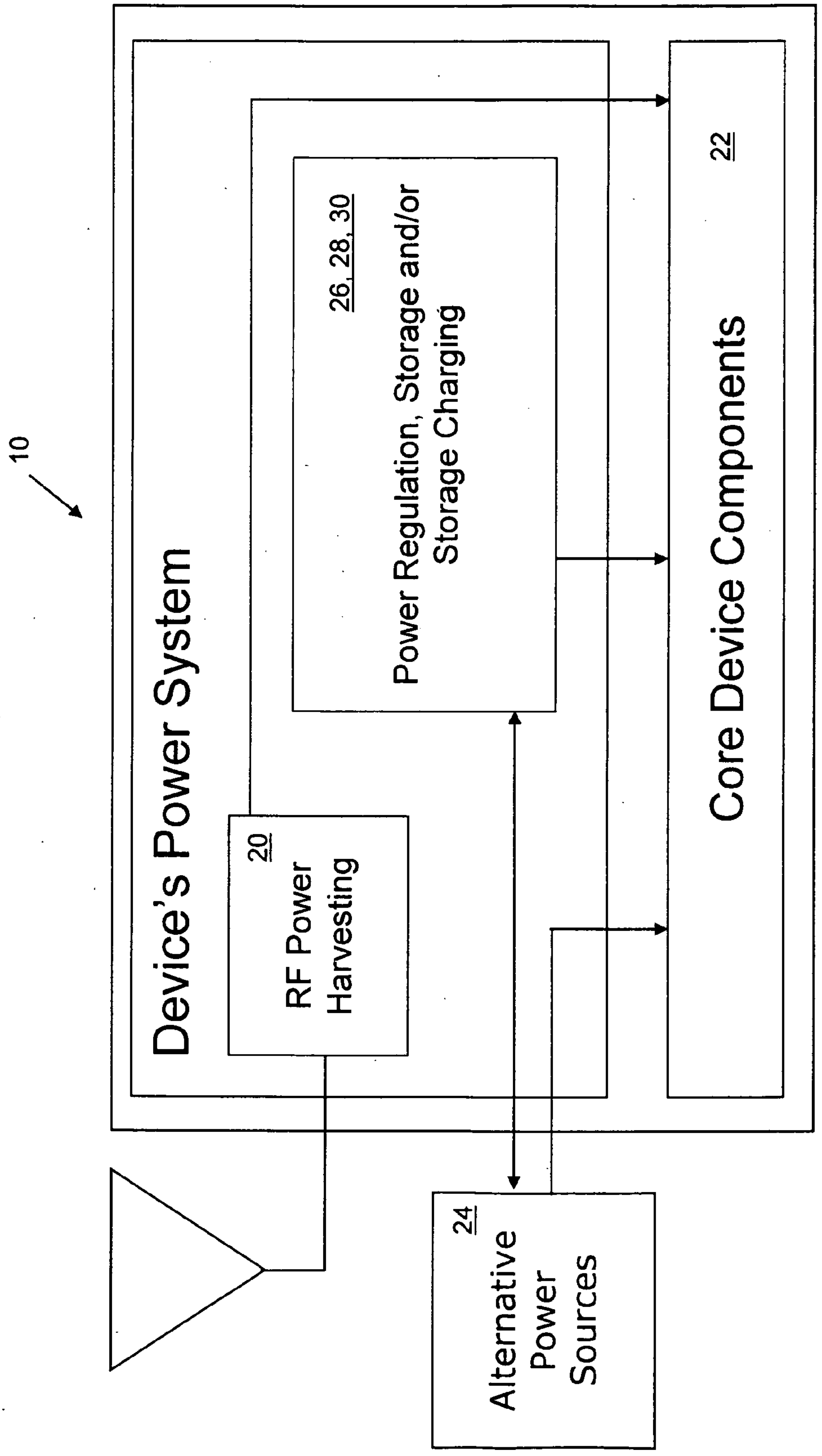
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 77

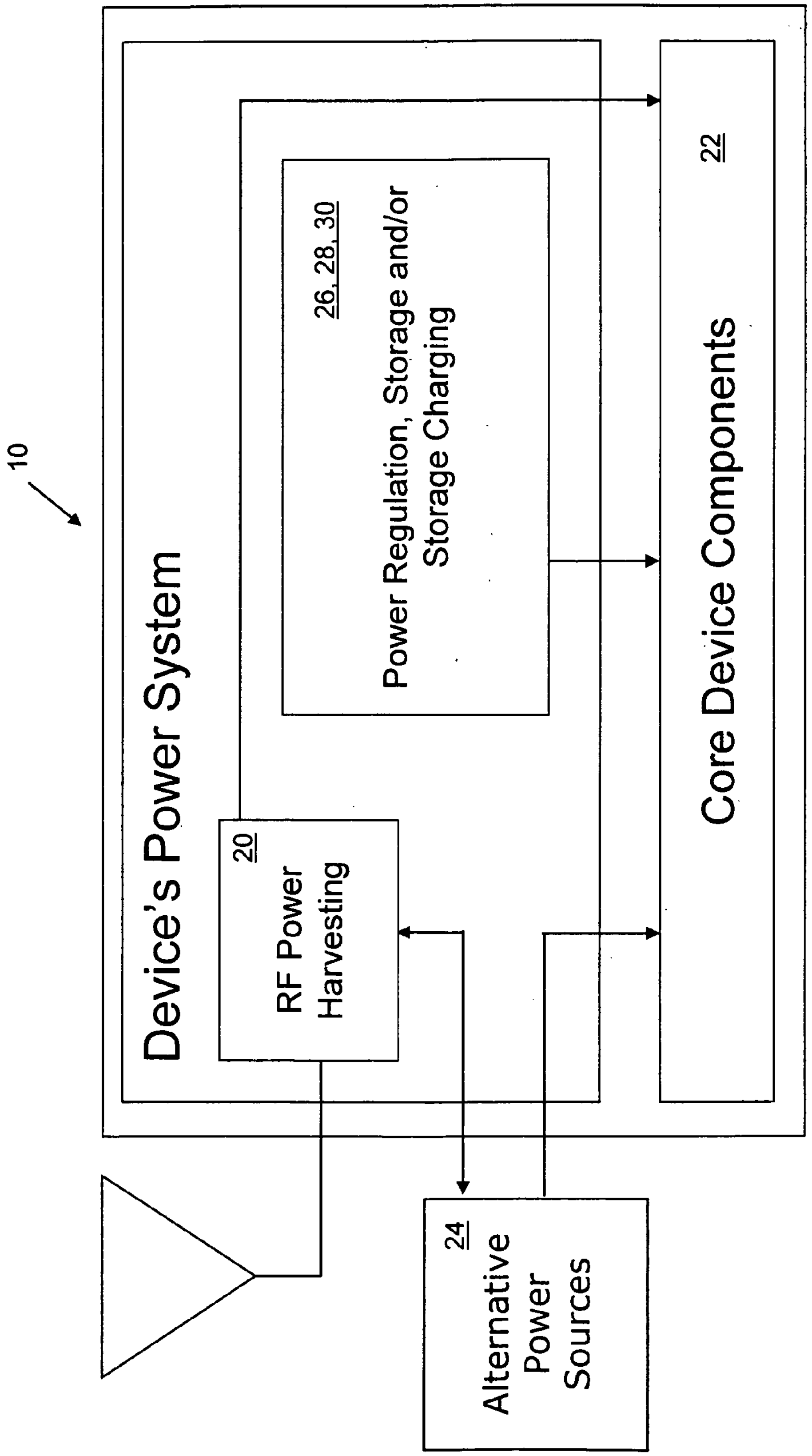
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 78

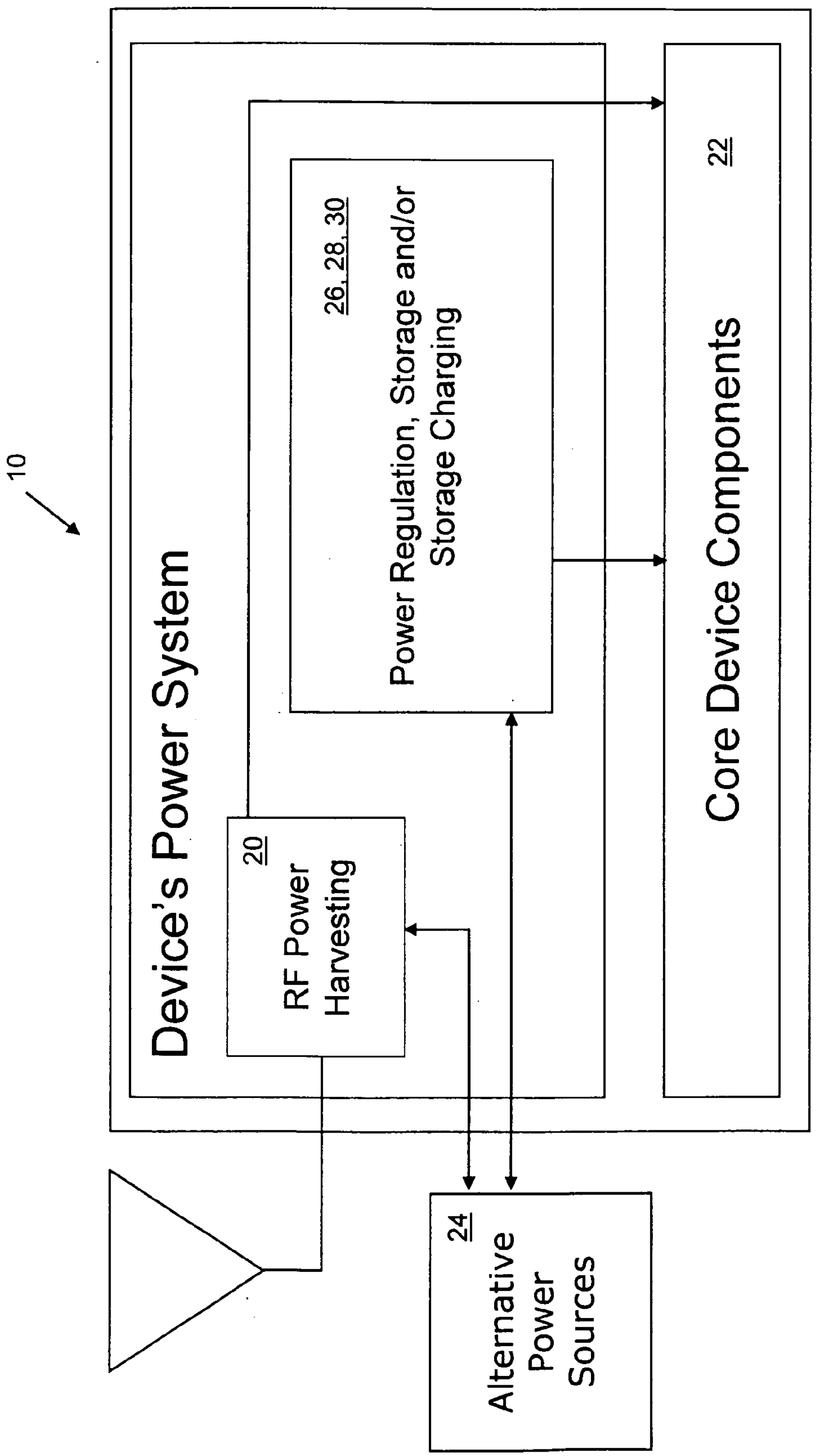
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 79

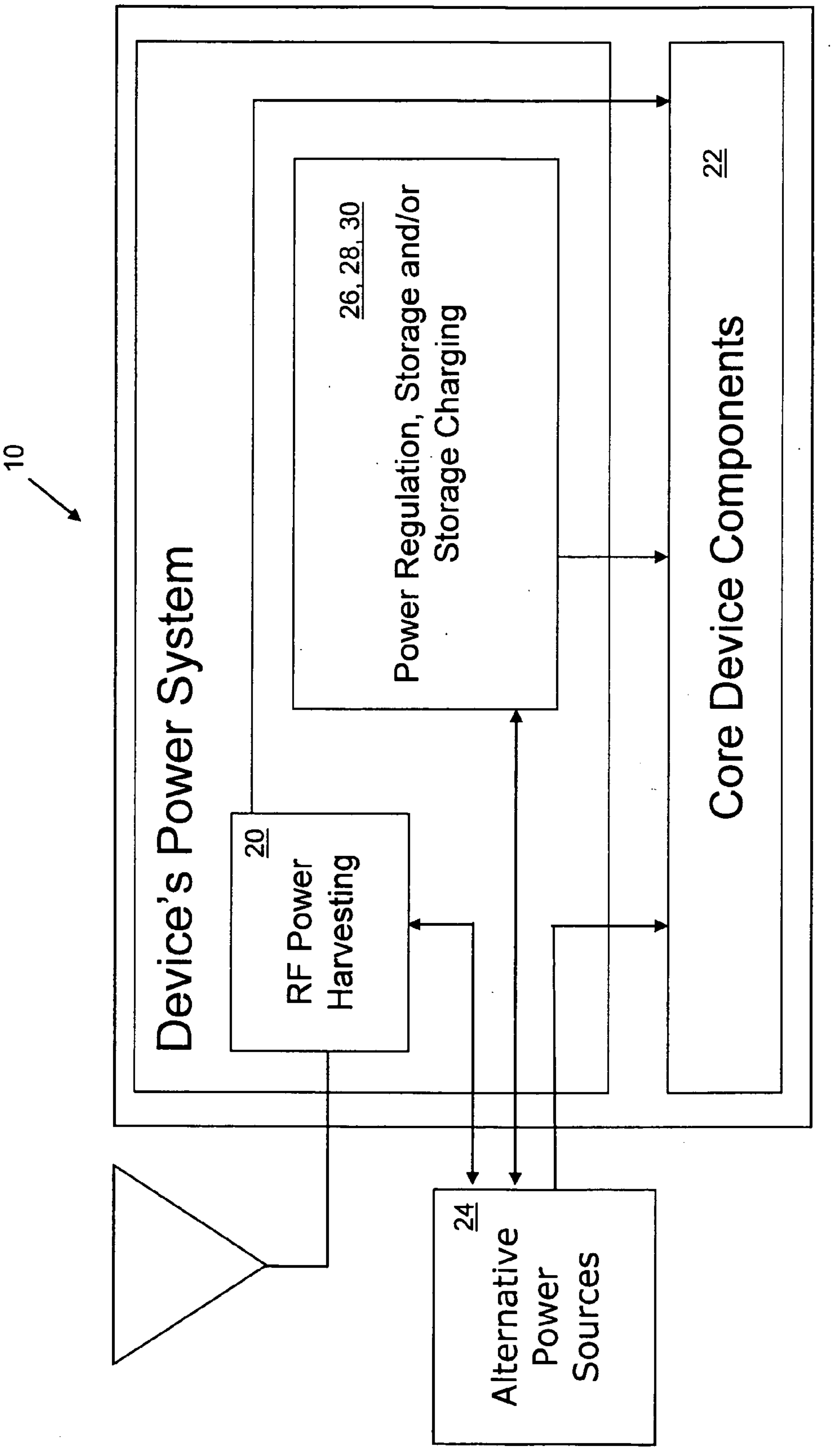
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 80

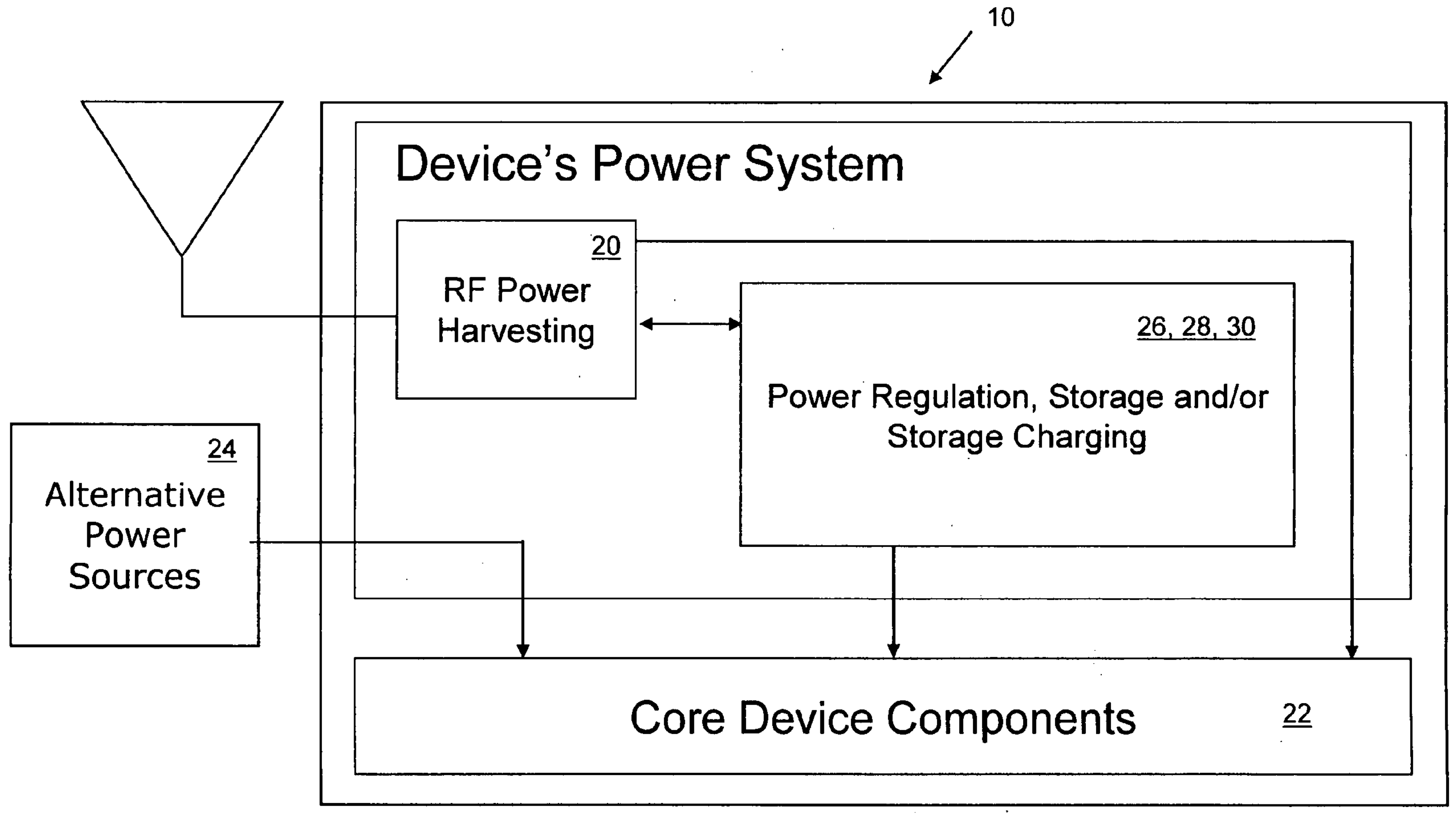
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 81

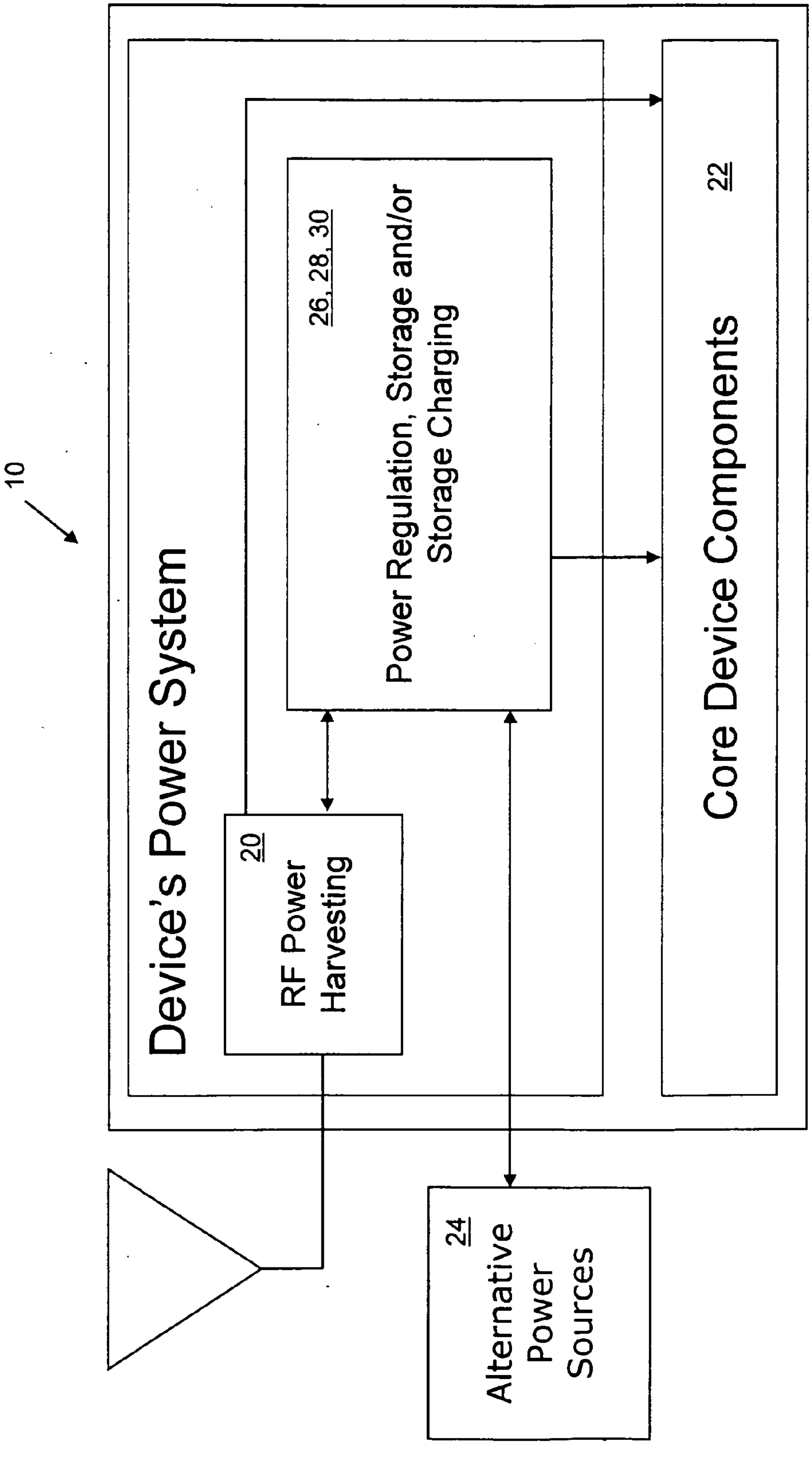
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 82

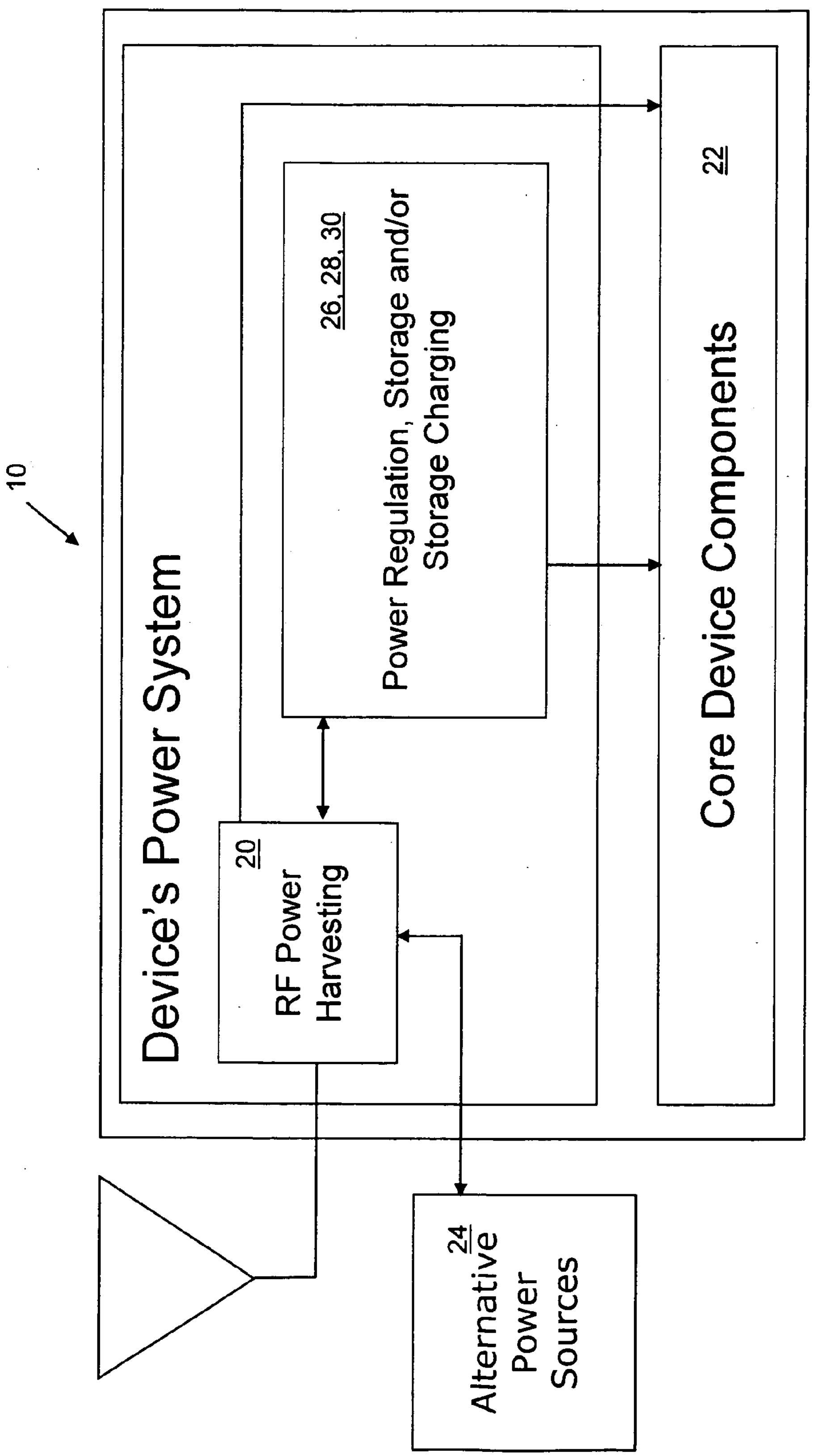
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 83

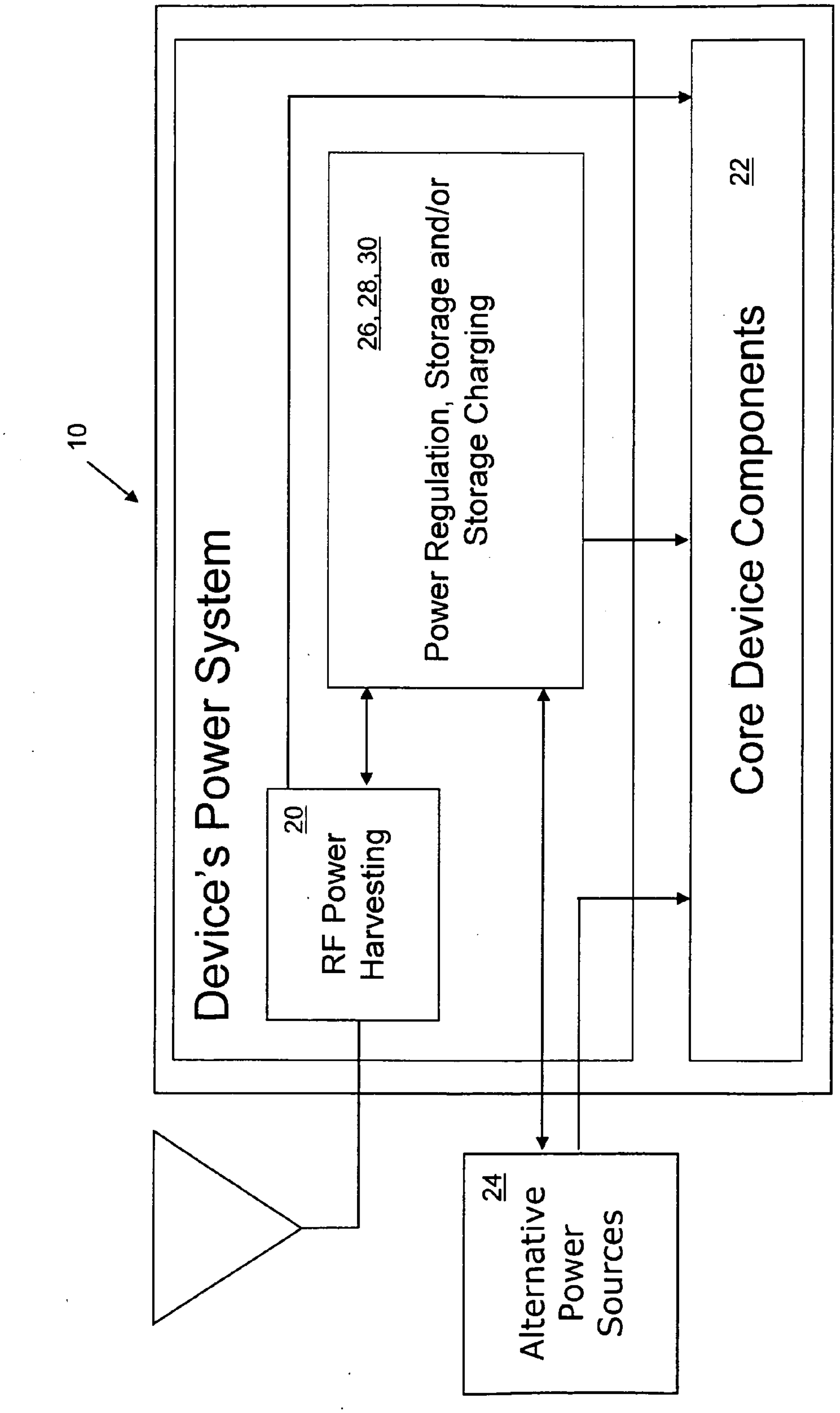
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 84

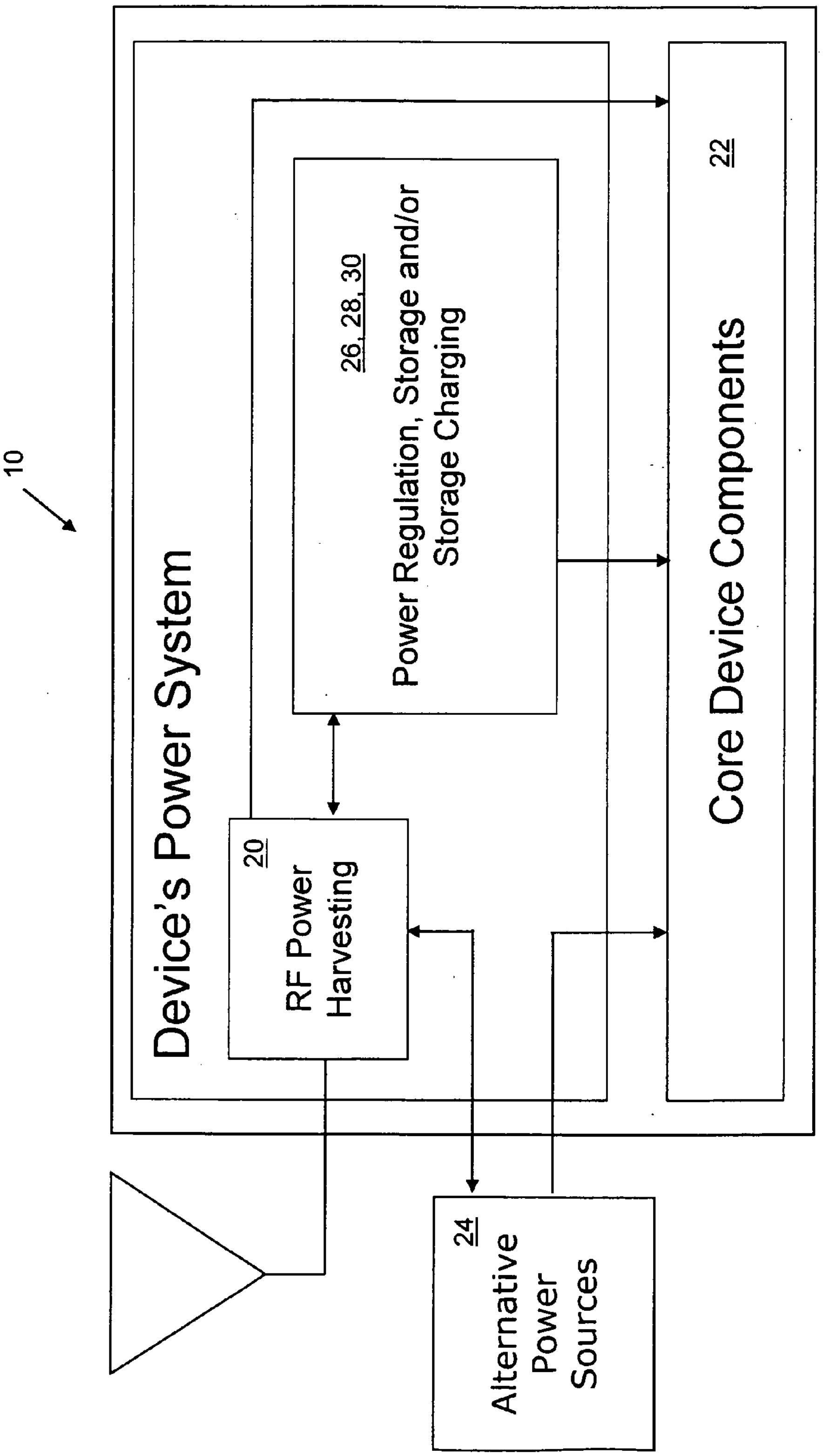
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 85

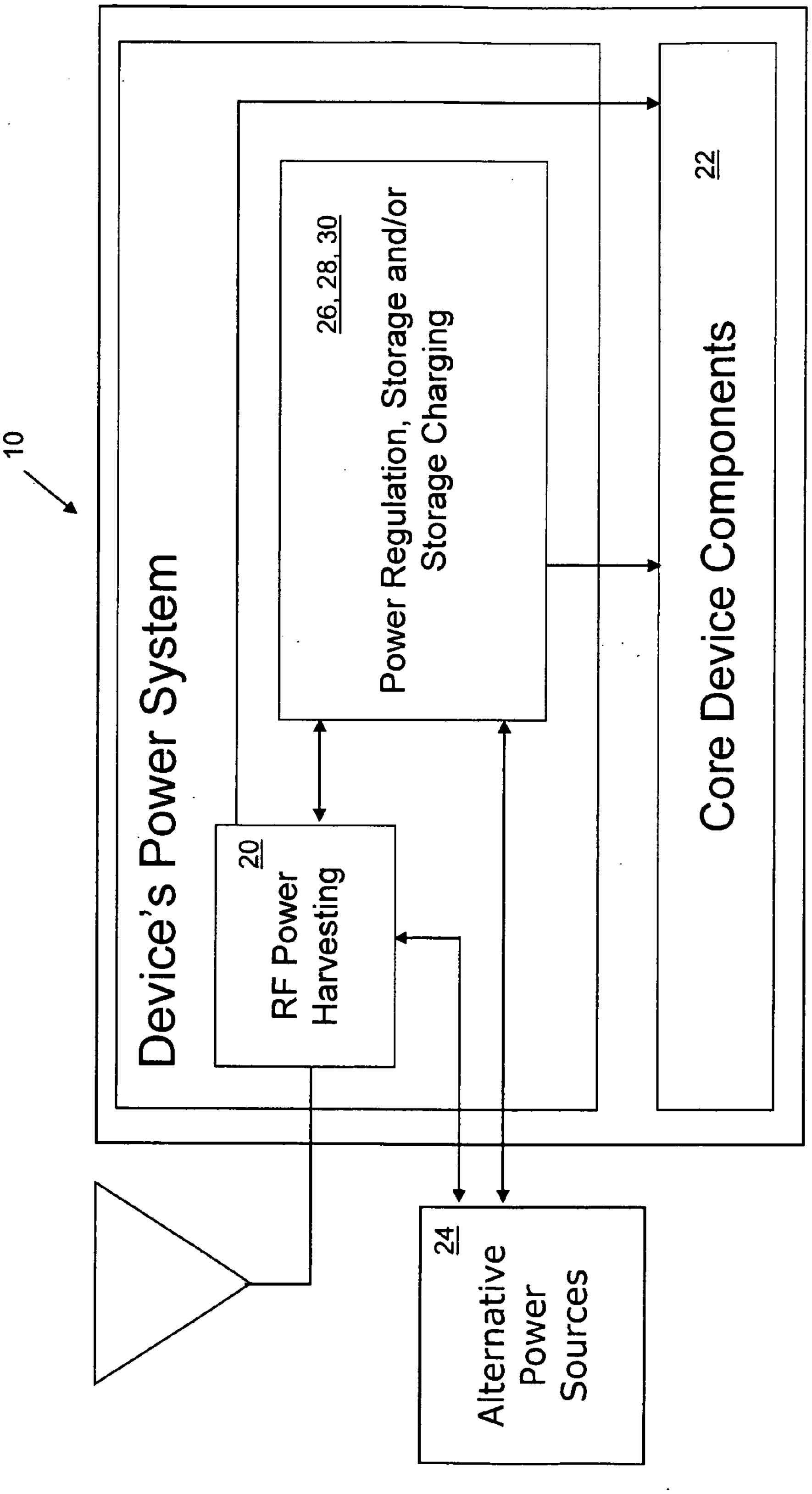
10
Device's Power System
RF Power Harvesting
20
Power Regulation, Storage and/or Storage Charging
26, 28, 30
Core Device Components
22
Alternative Power Sources
24

Fig. 86

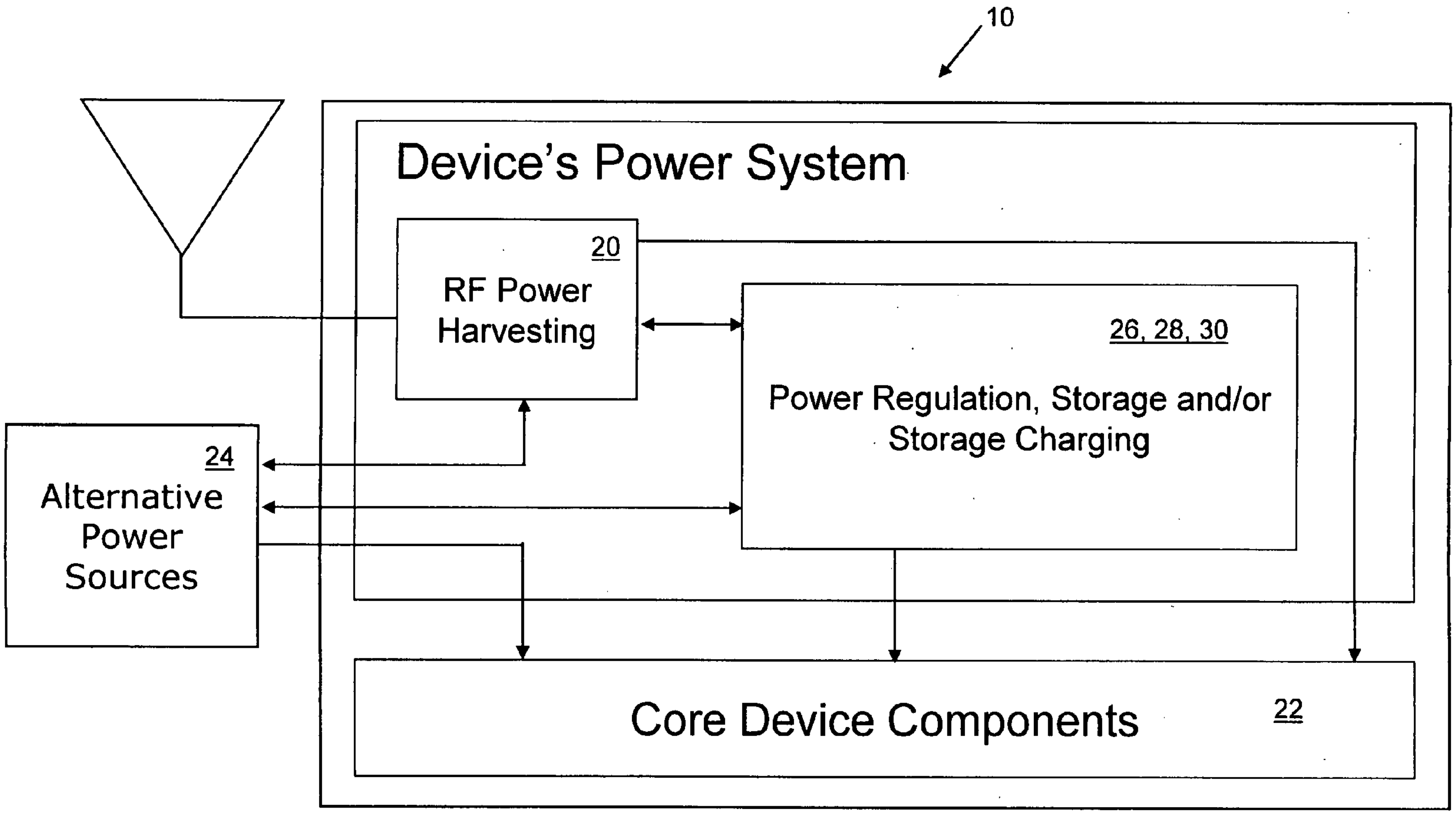
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 87

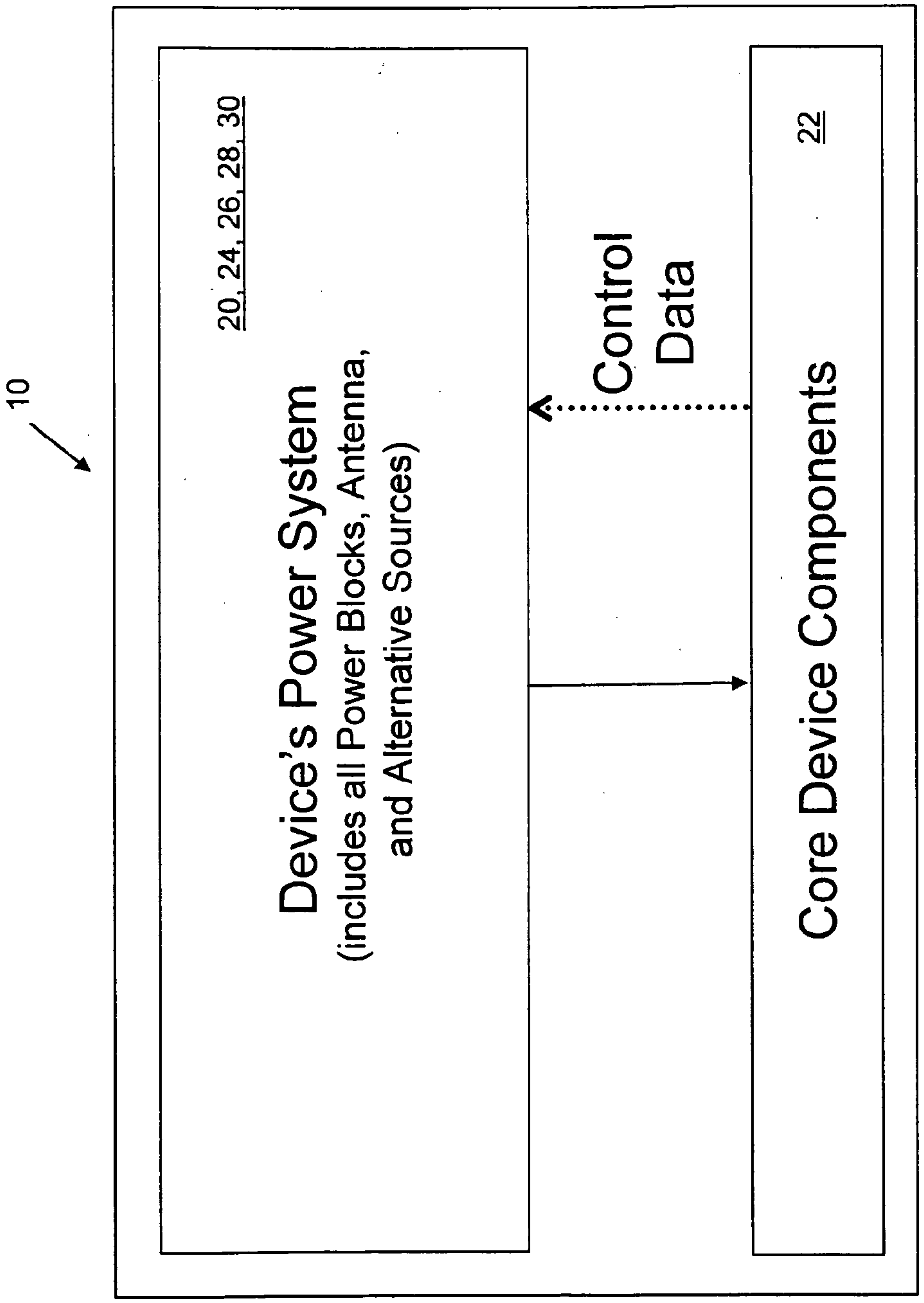
10
Device's Power System
(includes all Power Blocks, Antenna, and Alternative Sources)
20, 24, 26, 28, 30
Control Data
Core Device Components
22

Fig. 88

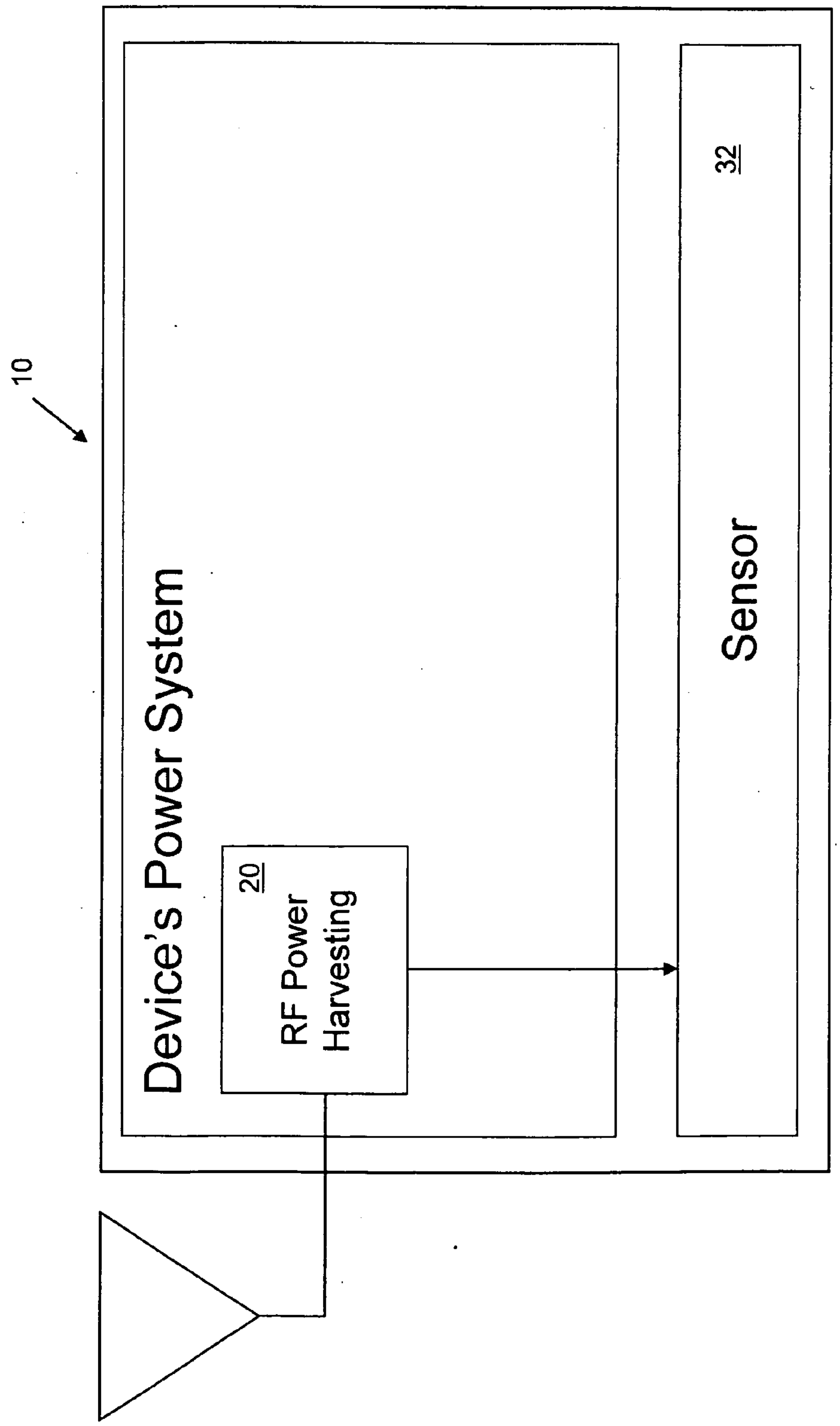
10
Device's Power System
20
RF Power Harvesting
32
Sensor

Fig. 89

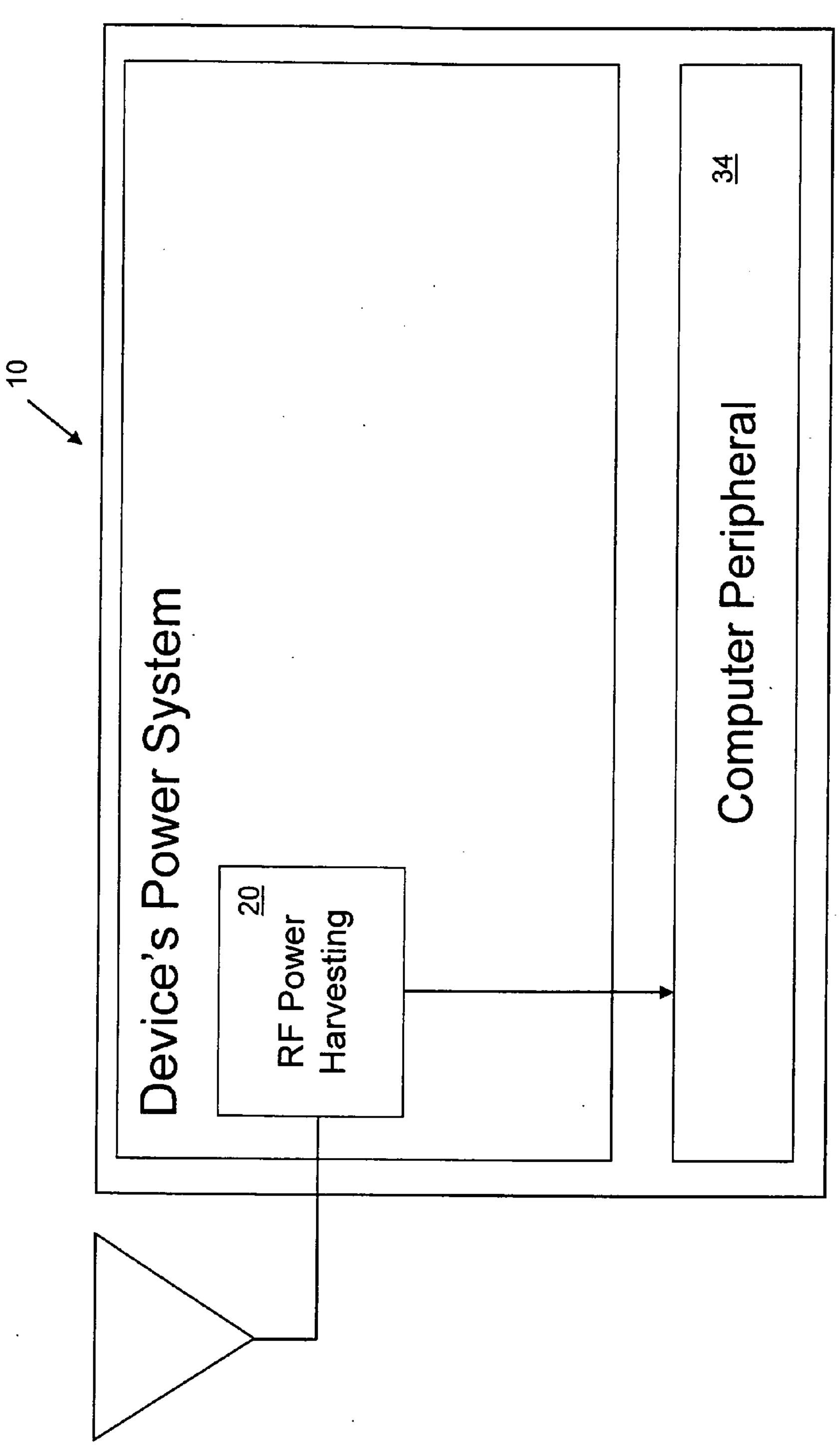
10
Device's Power System
20
RF Power Harvesting
Computer Peripheral
34

Fig. 90

POWERING DEVICES USING RF ENERGY HARVESTING

FIELD OF THE INVENTION

[0001] The present invention is related to the wireless powering of devices. More specifically, the present invention is related to the wireless powering of devices with a power harvester.

BACKGROUND OF THE INVENTION

[0002] As processor capabilities have expanded and power requirements have decreased there has been an ongoing explosion of devices that operate completely independent of wires or power cords. These "untethered" devices range from cell phones, and wireless keyboards to building sensors and active RFID tags.

[0003] Engineers and designers of these untethered devices continue to have to deal with the limitations of portable power sources, primarily batteries as the key design parameter. While performance of processors and portable devices have been doubling every 18-24 months driven by Moore's law, battery technology in terms of capacity has only been growing at measly 6% per year. Even with power conscious designs and the latest in battery technology, many devices do not provide the lifetime cost and maintenance requirements for applications that require a large number of untethered devices such as logistics, and building automation. Today's devices that need two-way communication require scheduled maintenance every three to 18 months to replace or recharge the device's power source (typically a battery). One-way devices simply broadcasting their status (one-way) such as automated utility meter readers have a better battery life, typically requiring replacement within 10 years. For both device types, scheduled power-source maintenance is costly and disruptive to the entire system that a device is intended to monitor and/or control. Unscheduled maintenance trips are even more costly and disruptive. On a macro level, the relatively high cost associated with the internal battery also reduces the practical, or economically viable, number of devices that can be deployed.

[0004] The ideal solution to the power problem for untethered devices is a device or system that can collect and harness sufficient energy from the external environment. The harnessed energy would then either directly power an untethered device or augment a battery or other storage component. Directly powering an untethered device enables the device to be constructed without the need for a battery. Augmenting a storage component could be along two lines: 1) increasing the overall life of the device or 2) by providing more power to the device to increase the functionality of the device. The other parameters for an ideal solution is that the harnessing device could be used in a wide range of environments including harsh and sealed environments (e.g. nuclear reactors), would be inexpensive to produce, would be safe for humans, and would have a minimal effect on the basic size, weight and other physical characteristics of the untethered device.

BRIEF SUMMARY OF THE INVENTION

[0005] The present invention pertains to an apparatus for an application. The apparatus comprises a core device preferably having an integrated circuit for the application. The apparatus comprises a power harvester (preferably a radio frequency (RF) energy harvester) connected to the core device to power the core device.

[0006] The present invention pertains to an apparatus for an application. The apparatus comprises a core device preferably having an integrated circuit for the application. The apparatus comprises means for receiving energy wirelessly and providing power from the energy to the core device to power the integrated circuit of the core device. The receiving means is connected to the core device.

[0007] The present invention pertains to a method for an application. The method comprises the steps of converting RF energy into usable energy. There is the step of preferably powering an integrated circuit of the core device with the usable energy.

[0008] This invention pertains to a technique that uses radio frequency (RF) energy as a source of energy to directly power a device or augment a power storage component in a device. The present invention meets the requirements described in the previous "Background of the Invention" section.

[0009] Traditional RF receiving devices have focused on maximizing selectivity of the frequency to isolate and to be coherent without interference from other sources. In contrast, while the present invention operates at a specific frequency or range of frequencies, the device accepts any interference to supplement the output power of the device. Also, the research related to power harvesting that uses RF energy as a source has primarily focused on devices in close proximity of the source (inductive or near-field energy). In most cases, prior research assumed a dedicated or directed source of RF to power the device.

[0010] The invention should not be confused with power transfer by inductive coupling, which requires the device to be relatively close to the power transmission source. The RFID Handbook by the author Klaus Finkenzeller defines the inductive coupling region as distance between the transmitter and receiver of less than 0.16 times lambda where lambda is the wavelength of the RF wave. The invention can be implemented in the near field (sometimes referred to as inductive) region as well as the far-field region. The far-field region is distances greater than 0.16 times lambda.

[0011] It is an object of this invention to provide a method and apparatus to

[0012] 1. remotely energize an untethered device without using direct wiring

[0013] 2. power or augment the life of the power storage component so it matches the life of the device and, ultimately, powers the off-grid device with or without the use of batteries

[0014] 3. allow untethered devices to be virtually maintenance free

[0015] 4. provide augmentation for other energy harvesting technologies (solar, piezoelectric, etc.)

[0016] 5. provide backup power to tethered devices

[0017] It is a further object of this invention to directly power a device or augment a power storage component in a device in conjunction with other power harvesting technologies and storage elements.

[0018] With this method and apparatus a device's power storage components do not require replacement, thus enabling the device to be permanently placed off-grid, where it may be physically impractical, costly, or dangerous (due to a harsh environment) to provide maintenance.

[0019] For devices on-grid (tethered) or with reliable power sources, RF power harvesting can be used as a backup in case the primary power source is lost.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING

[0020] In the accompanying drawings, the preferred embodiment of the invention and preferred methods of practicing the invention are illustrated in which:

[0021] **FIG. 1** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

[0022] **FIG. 2** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0023] **FIG. 3** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0024] **FIG. 4** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0025] **FIG. 5** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0026] **FIG. 6** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0027] **FIG. 7** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0028] **FIG. 8** is a block diagram of an RF Power Harvesting block in communication with a Power Storage block.

[0029] **FIG. 9** is a block diagram of an RF Power Harvesting block in communication with a Power Storage block and used to supply power to Core Device Components.

[0030] **FIG. 10** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0031] **FIG. 11** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0032] **FIG. 12** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block.

[0033] **FIG. 13** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0034] **FIG. 14** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

[0035] **FIG. 15** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

[0036] **FIG. 16** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0037] **FIG. 17** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block.

[0038] **FIG. 18** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger.

[0039] **FIG. 19** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and the RF Power Harvesting block in communication with a Power Storage block.

[0040] **FIG. 20** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and Core Device Components.

[0041] **FIG. 21** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and Core Device Components and the RF Power Harvesting block in communication with a Power Storage block.

[0042] **FIG. 22** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0043] **FIG. 23** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0044] **FIG. 24** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0045] **FIG. 25** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0046] **FIG. 26** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0047] **FIG. 27** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0048] **FIG. 28** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the RF Power Harvesting block supplying power to Core Device Components.

[0049] **FIG. 29** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0050] **FIG. 30** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

[0051] **FIG. 31** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0052] **FIG. 32** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power Core Device Components.

[0053] **FIG. 33** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0054] **FIG. 34** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0055] **FIG. 35** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0056] **FIG. 36** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0057] **FIG. 37** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0058] **FIG. 38** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0059] **FIG. 39** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0060] **FIG. 40** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to the Power Storage Charger.

[0061] **FIG. 41** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0062] **FIG. 42** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0063] **FIG. 43** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

[0064] **FIG. 44** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the RF Power Harvesting block supplying power to Core Device Components and used to supply power to a Power Storage Charger.

[0065] **FIG. 45** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0066] **FIG. 46** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0067] **FIG. 47** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

[0068] **FIG. 48** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power Core Device Components and used to supply power to a Power Storage Charger.

[0069] **FIG. 49** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0070] **FIG. 50** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

[0071] **FIG. 51** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0072] **FIG. 52** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0073] **FIG. 53** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0074] **FIG. 54** is a block diagram of an RF Power Harvesting block using Antenna A to directly supply power to Core Device Components.

[0075] **FIG. 55** is a block diagram of an RF Power Harvesting block using Antenna A to supply power to a Power Regulation, Storage and/or Storage Charging block.

[0076] **FIG. 56** is a block diagram of an RF Power Harvesting block using Antenna A to supply power to a Power Regulation, Storage and/or Storage Charging block and used to supply power to Core Device Components.

[0077] **FIG. 57** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

[0078] **FIG. 58** is a block diagram of an RF Power Harvesting block used to supply power to a Power Regulation, Storage and/or Storage Charging block.

[0079] **FIG. 59** is a block diagram of an RF Power Harvesting block used to supply power to a Power Regulation, Storage and/or Storage Charging block and used to supply power to Core Device Components.

[0080] **FIG. 60** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

[0081] **FIG. 61** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0082] **FIG. 62** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and used to directly supply power to Core Device Components.

[0083] **FIG. 63** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0084] **FIG. 64** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

[0085] **FIG. 65** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

[0086] **FIG. 66** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0087] **FIG. 67** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

[0088] **FIG. 68** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0089] **FIG. 69** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0090] **FIG. 70** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0091] **FIG. 71** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0092] **FIG. 72** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0093] **FIG. 73** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0094] **FIG. 74** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

[0095] **FIG. 75** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

[0096] **FIG. 76** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0097] **FIG. 77** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

[0098] **FIG. 78** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0099] **FIG. 79** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0100] **FIG. 80** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources.

[0101] **FIG. 81** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

[0102] **FIG. 82** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

[0103] **FIG. 83** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0104] **FIG. 84** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

[0105] **FIG. 85** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0106] **FIG. 86** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative

Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0107] **FIG. 87** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0108] **FIG. 88** is a block diagram of an entire power system for the device.

[0109] **FIG. 89** is a block diagram of a power harvesting block used to supply power to a core device having a sensor.

[0110] **FIG. 90** is a block diagram of a power harvesting block used to supply power to a core device having a computer peripheral.

DETAILED DESCRIPTION OF THE INVENTION

[0111] A complete understanding of the invention will be obtained from the following description when taken in connection with the accompanying drawing figures wherein like reference characters identify like parts throughout.

[0112] There is shown an apparatus **10** for an application. The apparatus **10** comprises a core device **22** preferably having an integrated circuit **36** for the application. The apparatus **10** comprises a power harvester **20** connected to the core device **22** to power the core device **22**.

[0113] The apparatus **10** preferably includes an alternative power source **24** connected to the core device **22** to power the core device **22** in conjunction with the power harvester **20**. Preferably, the apparatus **10** includes a power regulator **26** and/or power storage circuit **28** connected to the power harvester **20**. The apparatus **10** preferably includes a power storage charger **30** connected to the power harvester **20**. Preferably, the apparatus **10** includes a power storage connected to the power harvester **20**.

[0114] Preferably, the core device **22** includes a memory **38** connected to the integrated circuit **36** and to the power harvester **20** to power the memory **36**.

[0115] The core device **22** can include a sensor **32**, as shown in **FIG. 89**. The sensor **32** can include a proximity sensor, an intrusion sensor, an environmental sensor, a chemical sensor, a biological sensor, a sensor in contact with an automobile, an occupancy sensor, a motion sensor, a position sensor, a metal detector, or a sensor **32** in contact with an aircraft. The sensor **32** can include an alarm connected to the power harvester **20** to power the alarm, a display connected to the power harvester **20** to power the display, a sensor **32** disposed in a building, an industrial automation sensor, a sensor **32** in contact with an elevator, a temperature sensor, a fire sensor, or an accelerometer.

[0116] The sensor **32** can include a level sensor, a gas level sensor, a fluid level sensor, a light level sensor. The sensor **32** can include a flow sensor, a gas flow sensor, a fluid flow sensor, a light flow sensor, or a plasma flow sensor.

[0117] The sensor **32** can include a pressure sensor, a gas pressure sensor, a fluid pressure sensor, a light sensor, an infrared light sensor, an ultraviolet light sensor, an x-ray sensor, a cosmic ray sensor, a visible light sensor, a gamma ray sensor, a stress sensor, a strain sensor, a depth sensor, or an electrical characteristic sensor.

[0118] The sensor **32** includes a voltage sensor, a current sensor, a viscosity sensor, an acoustical sensor, a sound sensor, a listening sensor, a thickness sensor, a density sensor, a surface quality sensor, a volume sensor, a physical sensor, a mass sensor, a weight sensor, a conductivity sensor, a distance sensor, an orientation sensor, or a vibration sensor.

[0119] The sensor **32** can include a radioactivity sensor, a field strength sensor, an electric field sensor or a magnetic field sensor, a smoke detector, a carbon monoxide detector, a radon detector, an air quality sensor, a humidity sensor, a glass breakage sensor, or a break beam detector. The sensor can include a thermal energy sensor, an electromagnetic sensor, a mechanical sensor, an optical sensor, a radiation sensor, a sensor in contact with a vehicle, or a sensor **32** in contact with a water craft.

[0120] The present invention pertains to an apparatus **10** for an application. The apparatus **10** comprises a core device **22** having an integrated circuit **36** for the application. The apparatus **10** comprises means for receiving energy wirelessly and providing power from the energy to the core device **22** to power the integrated circuit **36** of the core device **22**. The receiving means is connected to the core device **22**. Preferably, the core device **22** includes means for sensing.

[0121] The present invention pertains to a method for an application. The method comprises the steps of converting RF energy into usable energy. There is the step of preferably powering an integrated circuit **36** of the core device **22** with the usable energy.

[0122] Preferably, there is the step of regulating the usable energy provided to the core device **22**. There is preferably the step of storing the usable energy. Preferably, there is the step of providing power to the core device **22** from an alternative power source **24** in conjunction with the usable energy.

[0123] Alternatively, the core device **22** can include a computer peripheral **34**, as shown in **FIG. 90**. The computer peripheral **34** can include a handheld game, a gaming system, a game controller, a controller, a keyboard, a mouse, a computer terminal, computer storage, or computer equipment.

[0124] The present invention can be implemented in numerous ways. A number of these ways are depicted in **FIGS. 1-90**. These figures contain multiple blocks that are configured in multiple ways.

[0125] In the figures, an arrow represents the flow of power unless otherwise stated. Single-headed (one-way) arrows represent that the power is flowing from one block to another. The single-headed arrow may represent multiple wires that provide power from one block to multiple parts in the other block. Two-headed (two-way) arrows represent a single wire that can have power flow in either direction or multiple wires each having power flow in a single direction.

[0126] As an example, a two-headed arrow between an RF power harvesting **20** block and a power regulation **26** and/or power storage circuit **28** block can represent a single wire that allows harvested power to flow into a storage device such as a capacitor. The same block diagram can also represent two wires between the two blocks with the first wire allowing harvested power to flow into a voltage regulator **26**. The second wire can allow the regulated voltage to feedback to the RF power harvesting **20** block to provide power to internal components such as transistors to increase the performance of the RF power harvesting **20** block.

[0127] Each block is described in detail below. Each block represents the functionality described below associated with it. For instance, the RF power harvesting **20** block describes a power harvester **20**, and the power regulation **26** and/or power storage circuit **28** block describes a power regulator **26** and a power storage circuit **28**.

[0128] Device's Power System Block

[0129] The device's power system block includes all components/circuitry required to power the device. This block may include an RF power harvesting **20** block, a power regulation **26** and/or power storage circuit **28** block, a power storage block, and/or a power storage charger **30** block (all discussed below).

[0130] RF Power Harvesting **20** Block

[0131] The RF power harvesting **20** block is connected to an antenna. The antenna may or may not be used as the communications antenna for the core device **22** components. The RF power harvesting **20** block is used to convert the energy captured by the antenna into usable power, such as direct current (DC) voltage.

[0132] This block may include antenna matching, rectifying circuitry, voltage transforming circuitry, and/or other performance optimizing circuitry. The rectifying circuitry may include a diode(s), a transistor(s), or some other rectifying device or combination. Examples of the rectifying circuitry include, but are not limited to, half-wave, full-wave, and voltage doubling circuits.

[0133] The output of the RF power harvesting **20** block is a DC voltage or current. The RF power harvesting **20** block may accept feedback (or input) from other circuitry or blocks, which may be used to control the harvesting circuitry to improve the performance or vary the output. This feedback may include, but is not limited to, a DC voltage or a clock from the core device **22** components.

[0134] U.S. Pat. No. 6,615,074 (**FIGS. 8, 9, 12***a*, **12***b*, **13**, **14**), incorporated by reference, herein, shows numerous examples of RF power harvesting circuits that can be used to implement the block and function described.

[0135] Power Regulation **26** and/or Power Storage Circuit **28** Block

[0136] It may be necessary to regulate the converted power (i.e., hold the power at a constant level) for specific devices. The devices that would need this block require a fairly constant voltage or current. Deviations from the required values may cause the device to not perform within its specifications.

[0137] The regulation can be implemented in many different ways. The block can be as simple as using a Zener diode, or as complicated as using an integrated circuit such as a linear voltage regulator **26** or switching regulator **26** to hold the voltage at a constant level.

[0138] Certain devices have a more tolerable power requirement. For these devices, the regulation stage may be excluded.

[0139] This block may also include, with or without the regulation, a storage device such as a capacitor, a battery, or some other device able to store charge. The output from the power regulation **26** and/or power storage circuit **28** block may be used as feedback to other blocks within the device's power system block or to the alternative power sources **24** block (described below), if they require a regulated supply voltage or stored power.

[0140] U.S. Pat. No. 6,894,467 (**FIGS. 1, 3**), Linear Voltage Regulator, incorporated by reference, herein, is an example of a practical application of implementing the regulation described in the block. U.S. Pat. No. 6,297,618 (**FIGS. 1-4**), Power storage device and method of measuring voltage of storage battery, incorporated by reference, herein, is an example of a practical application of implementing the storage described in the block.

[0141] Power Storage Charger **30** Block

[0142] The power storage charger **30** block may be needed if a power storage component of the device requires a special charging mechanism, such as pulse charging or trickle charging. This block controls how the captured and converted power is supplied to the power storage component.

[0143] U.S. Pat. No. 6,836,095 (**FIGS. 1-3**), Battery Charging Method and Apparatus, incorporated by reference herein, is an example of a practical application of implementing the special charging mechanism described in the block.

[0144] Power Storage Block

[0145] If a device has intermittent power requirements, it may be necessary to store the captured power for use at a later time. The power can be stored in the power storage block, which could include a battery, a capacitor, and/or another type of power storage component. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other power storage elements.

[0146] The output from the power storage block may be used as feedback to other blocks within the device's power system or to the alternative power sources **24** block, if they require a dedicated and predictable supply voltage.

[0147] U.S. Pat. No. 6,297,618 (**FIGS. 1-4**), Power Storage Device and Method of Measuring Voltage of Storage Battery, incorporated by reference herein, is an example of a practical application of implementing the storage described in the block. U.S. Pat. No. 6,835,501, Alkaline Rechargeable Battery, incorporated by reference herein, is also an example of a practical application of implementing the storage described in the block.

[0148] Core Device **22** Components Block

[0149] The core device **22** components block is the portion of the device that is receiving power from the device's power system. This block may be, but is not limited to, the devices listed in the subsequent pages of this document. It may be advantageous for the core device **22** components block to communicate with any of the blocks that are supplying power to it. This communication can include, but is not limited to, a feedback control signal such as a clock or an ON/OFF command. As an example, the device may want to turn off the alternative power sources **24** block if it is receiving sufficient power from the RF power harvesting **20** block.

[**0150**] Alternative Power Sources **24** Block

[**0151**] RF energy harvesting also has the ability to be augmented by other types of power harvesting, storage components, or dedicated sources (e.g. power line). The alternative power sources **24** block shows how this type of system could be implemented. The augmenting power harvesting technologies include, but are not limited to, solar, light (visible and non-visible), piezoelectric, vibration, acoustic, thermal, microgenerators, wind, and other environmental elements. This block can work independently or have communication with other blocks.

[**0152**] U.S. Pat. No. 6,784,358, Solar Cell Structure Utilizing an Amorphous Silicon Discrete By-Pass Diode, incorporated by reference herein, is an example of a practical application of implementing an alternative power source **24** described by the block. U.S. Pat. No. 6,858,970, Multi-Frequency Piezoelectric Energy Harvester, incorporated by reference herein, is also an example of a practical application of implementing an alternative power source **24** described by the block.

[**0153**] Power Regulation **26**, Storage **28** and/or Storage Charging **30** Block

[**0154**] The power regulation **26**, storage **28** and/or storage charging **30** block contains all the combinations of the power regulation **26** and/or power storage circuit **28** block, power storage charger **30** block, and power storage block. This block is used in the later figures to reduce the number of figures needed to show how the blocks can interconnect.

[**0155**] The disclosed invention is an apparatus and method for an application for retrieving radio frequency (RF) energy by an antenna, converting that energy into direct current (DC) power, regulating that energy using an optimized circuit, storing that energy in an optimized component, and/or supplying the power for a specific device.

[**0156**] Retrieval of RF Energy

[**0157**] The RF energy is retrieved from the environment by the use of the antenna. The antenna can be shared or standalone with respect to an antenna used for the device's wireless communication. **FIGS. 54-56** show a device that has an antenna A for use by the RF power harvester **20** and an antenna B used for wireless one-way or two-way communication. **FIGS. 57-59** show a device where the antenna is shared by both a device's communication module (within the core device **22** components) and the RF power harvester **20**. In terms of form factor, the antenna used by the apparatus **10** can be a separate component or integrated directly into the form factor of the device.

[**0158**] The antenna is able to capture two types of available RF energy. The first type of energy exists as ambient RF energy. This type of RF surrounds us in our day-to-day lives and is usually generated to carry one way or two-way combinations of voice, video and data communications. The sources that the antenna can harvest from include, but are not limited to, medium-frequency AM radio broadcast, very-high-frequency (VHF) FM radio broadcast and television broadcast, ultra-high-frequency (UHF) broadcast, cellular base stations, wireless data access points, super-high-frequency (SHF) frequencies, and the industrial, scientific, and medical (ISM) bands. These sources cover transmission frequencies from 300 kHz to 30 GHz.

[**0159**] The second type of energy available is directed RF energy. This type of RF energy is directed from a transmitter specifically designed to deliver RF energy for harvesting by the antenna. The transmitter can be configured as a standalone device or integrated into an existing device.

[**0160**] Conversion of the Energy into DC

[**0161**] The RF energy captured by the antenna must be converted into a useful form of energy for the specific device. This conversion is shown in block form in all FIGS. (**1-90**) as the RF power harvesting **20** block. The most common form of useable energy is DC energy. To perform this conversion, the block includes circuitry to rectify the captured alternating current (AC) energy to create DC energy. The rectification in this block can be done with a diode(s), a transistor(s), or some other rectifying device or combination.

[**0162**] Regulation of the Energy

[**0163**] It may be necessary to regulate the converted power (hold the power at a constant level) for specific devices. **FIGS. 2-7**, **10-17**, and **22-53** show how this regulation can be implemented using a power regulation **26** and/or power storage circuit **28** block. The devices that would need this block require a fairly constant voltage or current. Deviations from the required values may cause the device to not perform within its specifications. The regulation can be implemented in many different ways. The block can be as simple as using a Zener diode, or as complicated as using an integrated circuit such as a linear voltage regulator **26** or a switching regulator **26** to hold the voltage at a constant level. Certain devices have a more tolerable power requirement. For these devices, the regulation stage may be excluded.

[**0164**] Storage of the Energy

[**0165**] If a device has intermittent power requirements, such as the devices exampled by **FIGS. 2-53**, it may be necessary to store the captured power for use at a later time. The power can be stored in the power storage block or the power regulation **26** and/or power storage circuit **28** block. Storage devices can include, but are not limited to, a battery, a capacitor, or another type of power storage component. In certain applications, it may be necessary to include additional circuitry that controls how the power is transferred to the storage device. The power storage charger **30** block is shown in **FIGS. 18-53**. This may be needed if the power storage component requires a special charging mechanism such as pulse charging or trickle charging. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other storage elements. There are devices that will not require storage. These devices can run directly off of the converted power. These devices also may or may not require regulation of the captured power.

[**0166**] Supplying the Power

[**0167**] The captured DC power, which may or may not be regulated and/or stored, is supplied to the device, which is represented by the core device **22** components block in the figures. This may be a single connection or it may supply multiple parts of the device with power.

[0168] RF energy harvesting also has the ability to be augmented by other types of power harvesting or storage components. Other power harvesting technologies include, but are not limited to, solar, light (visible and non-visible), piezoelectric, vibration, acoustic, thermal, microgenerators, wind, and other environmental elements. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other storage elements. **FIGS. 60-88** show how the alternative power sources **24** block can be connected to an RF energy harvesting system. These figure show how the RF energy harvesting **20** block and the alternative power sources **24** block can work independently or have communication with each other. The antenna configurations shown in **FIGS. 54-59** are still applicable with the addition of an alternative power sources **24** block. These antenna configurations can be applied to **FIGS. 60-88**.

[0169] RF energy harvesting also has the ability to provide a backup to devices on-grid (tethered) or with reliable power sources, which can be used in case the primary power source is lost. As an example, it may be mandated by regulations that a sensor has auxiliary power in case the primary supply is lost. It could be possible to use a rechargeable battery that obtains its charge from the primary supply when in operation. However, if the primary supply is lost for a time greater than the life of the rechargeable battery, the specification of uninterrupted power is not met. RF energy could be used to supply power to the described device while the primary supply is not available. The primary supply could include, but is not limited to, an on-grid connection, a generator, a battery, or other reliable power supply.

[0170] RF energy harvesting with or without alternative source augmentation is applicable to provide electric power directly or indirectly to a range of electronic components contained in any specific electrical or electronic device and includes, but is not limited to:

[0171] Passive electronic components, active electronic components

[0172] Resistors, fixed resistors, variable resistors, thermistors, thyristor, thermocouple

[0173] Capacitors, Electrolytic Capacitors, Tantalum Capacitors, Ceramic Capacitors, Multilayer Ceramic Capacitors, Polystyrene Film Capacitors, Electric Double Layer Capacitors (Super Capacitors), Polyester Film Capacitors, Polypropylene Capacitors, Mica Capacitors, Metallized Polyester Film Capacitors, Variable Capacitors

[0174] Diodes, Voltage regulation diodes, light-emitting diodes, organic light-emitting diodes, Variable capacitance diodes, Rectification diodes, Switching diodes, Regulation Diodes, Diode bridges, Schottky barrier diodes, tunnel diodes, PIN diodes, Zener diodes, Avalanche diodes, TVSs

[0175] Integrated circuits, microcontroller unit (MCU), microprocessor unit (MPU), logic circuits, memory, printed circuits, circuit boards, printed wiring boards

[0176] Transistors, MOSFETs, FETs, BJTs, JFETs, IGBTs, Relays, Antennas, semiconductors, conductors, inductors, relays, diacs, triacs, SCRs, MOVs

[0177] Fuses, circuit breakers

[0178] Batteries, Non-rechargeable batteries, rechargeable batteries, coin cell batteries, button cell batteries, alkaline batteries, lithium batteries, lithium ion batteries, lithium polymer batteries, NIMH batteries, NICAD batteries, Lead acid batteries, Zinc air batteries, Manganese Lithium batteries, Niobium Titanium Lithium batteries, Vanadium Pentoxide Lithium batteries, Carbon Zinc batteries, Zinc Chloride batteries, Lithium Thionyl Chloride batteries, Manganese Dioxide batteries, Lithium Poly-Carbonmonofluoride batteries, Lithium Manganese Dioxide batteries, Lithium Chloride batteries, Lead Acid Calcium batteries, Lead Acid Tin batteries, Oxy Nickel batteries, Silver Oxide batteries, Magnesium batteries

[0179] Inductors, Coils, High Frequency Coils, Toroidal Coils, Transformers, switches, chokes

[0180] Motors, DC motors, stepper motor, AC motors, Fans

[0181] Crystals, Oscillators, Clocks, Timers

[0182] Displays, LCDs, LED displays

[0183] RF energy harvesting with or without alternative source augmentation is applicable across a range of markets and specific devices and includes but is not limited to:

[0184] Consumer electronics

[0185] Electronic equipment, wired devices, battery powered devices, wireless communication devices, cell phones, telephones, phones, cordless phones, portable phones, Bluetooth devices, Bluetooth headsets, hands-free headsets, headsets, headphones, Wireless headsets, radios, AM/FM radios, shortwave radios, weather radios, Two-way radios, portable radio, lights, lanterns, portable lights, flashlights, nightlights, spotlights, search lights, calculators, graphing calculators, desk calculators, clocks, alarm clocks, wall clocks, desk clocks, travel clocks, watches, wristwatches, pocket watches, stop watches, timers, voice recorders, Dictaphones, laser pointers, power tools, cordless power tools, electronic razors, electric razors, handheld games, gaming systems, game controllers, wireless game controllers, remote controls, battery chargers, computers, portable computers, keyless entry, toys, toy guns, toy laser guns, games, microphones, musical instruments, musical effects processors, musical instrument tuners, metronomes, electronic chord charts, door openers, garage door openers, PDA, Cameras, Video recorders, Multi-meter, electronic test equipment, hand-held electronics, portable electronics, wireless pens, sound generators, noise generators, language translators, electric toothbrushes, portable televisions, pagers, transceivers, toy vehicles, remote control vehicles, toy planes, remote control planes, pet containment systems, invisible fence pet sensors, memory backup, base station battery backups, appliance battery backups, uninterrupted power supplies, GPS devices, memory retention power supplies, metal detectors, stud finders, metal stud finders, stun guns, tazers, wearable devices, baby monitors, intercoms, doorbells, wireless doorbells, electronic office supplies, electronic staplers, radar jammers, radar detectors, digital scales, microfilm cassettes, video head testers, compasses, noise canceling headphones, air samplers, depth finders, barometers, weather measurement instruments, data transfer devices, automatic distress signaling unit, Wireless audio speakers, Satellite radios, Police scanners, Car navigation systems (GPS devices), Decorative lights, Christmas lights, garden lights, lawn lights, ornamental lights, porch lights

[0186] Multi-media players: MP3, DVD, analog music players, CD players, tape players, digital music players, digital video players, minidisc

[0187] Computer: keyboards, mice, peripherals, computer equipment, electronic computers, computer storage, computer terminals

[0188] Building/Industrial Automation

[0189] Sensors: Position, elevator, temperature, fire, accelerometers, level, gas level, fluid level, light level, flow, gas flow, fluid flow, light flow, plasma flow, pressure, gas pressure, fluid pressure, motion, light, infrared light, ultraviolet light, X-rays, cosmic rays, visible light, gamma rays, chemical, stress, strain, depth, electrical characteristics, voltage, current, viscosity, acoustical, sound, listening, thickness, density, surface quality, volume, physical, mass, weight, force, conductivity, distance, orientation, vibration, radioactivity, field strength, electric field strength, magnetic field strength, occupancy, smoke detector, carbon monoxide detector, radon detector, air quality, humidity, glass breakage, break beam detector

[0190] Controls: Position, elevator, temperature, fire, accelerometers, level, gas level, fluid level, light level, flow, gas flow, fluid flow, light flow, plasma flow, pressure, gas pressure, fluid pressure, motion, light, infrared light, ultraviolet light, X-rays, cosmic rays, visible light, gamma rays, chemical, stress, strain, depth, electrical characteristics, voltage, current, viscosity, acoustical, sound, listening, thickness, density, surface quality, volume, physical, mass, weight, force, conductivity, distance, orientation, vibration, radioactivity, field strength, electric field strength, magnetic field strength, occupancy, smoke detector, carbon monoxide detector, radon detector, air quality, humidity

[0191] Devices: Thermostats, light switches, door locks, smart-card door locks, lighting, emergency lighting, motion lighting, safety lighting, highway lighting, construction lighting, sign lighting, roadway sign lighting, construction sign lighting, automatic flushing units, automatic soap dispenser, automatic paper towel dispenser, automatic faucets, automatic door sensors, identification reader, fingerprint reader, credit card readers, card readers, valve actuators, gauges, analog gauges, digital gauges, fire extinguishers, wireless switches, remotely operated inspection equipment, gas/oil pipeline monitoring systems, robotic pipeline inspection gauges, "auto-reclosers" for electric power lines, sonar buoys, telemetry systems, electronic record tracking systems, robbery tracking devices, interrogators, programmers, emergency exit alarms, alarms, flood alarms, gas alarms, electronic entry systems, security keypads, silo transducers, data recorders, signal tracers, anti-static strap testers, radiosonde weather balloons, utilities load controllers, profilometers, noise cancellation equipment, infrared beacons

[0192] Military/Government

[0193] Tracking tags: Weapons, vehicle, soldier, gear/assets, staff, general population, security access badges

[0194] Sensors: Proximity, intrusion, environmental, chemical/biological

[0195] Equipment: Battery charger, surveillance, card readers, identification reader, fingerprint reader, retinal scanners, satellites, rockets, space vehicles, search and rescue transponders (SARTs), emergency position-indicating rescue beacons (EPIRBs), emergency locator transmitters (ELTs), military radios, electronic toll collection systems, postal tracking systems, communications, thermal imaging, night vision, training targets, field medical equipment, house arrest monitors, laser tags, electronic parking meters, multiple integrated laser engagement system, munitions and mines, ship sensors

[0196] Utility

[0197] Gas consumption meters, water consumption meters, and electric consumption meters

[0198] Logistics & Supply Chain Management

[0199] Radio-frequency identification devices (RFID), RFID readers

[0200] Tracking: Asset tags, cargo container location beacons, transponders, transceivers

[0201] Devices: Smart price tags, smart shelving, handheld barcode scanner, barcode scanners, credit card readers, card readers, retail signage, hotel door locks

[0202] Homeland Security

[0203] Sensors: Occupancy, proximity, environmental, chemical/biological, motion, position

[0204] Metal detector wand

[0205] Medical

[0206] Implantable: cochlear implants, neural stimulators, pace makers, medication administration, defibrillator

[0207] Body function monitors: pressure, temperature, respiration, blood oxygenation, insulin, hearing aid, pulse, EKG, heart, Holter,

[0208] Tracking tags: Patient, baby identification, assets, supplies, staff, medication, instruments

[0209] Devices: Home healthcare equipment, ambulatory infusion pumps, blood analyzers, biofeedback systems, bone growth stimulators, thermometers, digital thermometers, stimulators, galvanic stimulators, muscle stimulators, pediatric scales

[0210] Agriculture—livestock tracking and asset tracking.

[0211] Tracking: livestock, asset, wildlife tracking devices

[0212] Equipment: cattle prods

[0213] Automotive

[0214] Automotive antennas, Automotive Audio Systems, Automotive Lighting, Automotive Video Systems, Computers, Processors, Controls, Switches, Electric Motors, Actuators, Ignition Systems, Starter Systems, Injection Systems, Powertrain Electronics, Radar Detectors, Proximity Detectors, Safety Systems, Security Systems, Sensors, Regulators, Distributors, Vehicle Control, Wiper Systems, Washer Systems, Radio, Video Systems, Entertainment Systems, Navigation Systems, GPS systems, Power Mirror Systems, Emission control systems

[0215] Appliances

[0216] Monitoring systems and control systems for major and small appliances including washing machines, dryers, refrigerators, freezers, coolers, air conditioners, humidifiers, dehumidifiers, air purifiers, air filters, fans, furnaces, water heaters, boilers, space heaters, sowing machines, ice makers, microwave ovens, convection ovens, ovens, toaster ovens, ranges, range hoods, cooktops, stoves, stovetops, crock pots, hot plates, dishwashers, garbage disposals, can openers, vacuum cleaners, blenders, mixers, food processors, irons, coffee makers, toasters, grills, hair dryers, electric tooth brushes, electric razors, electric drills, electric screwdrivers, chainsaws, lawnmowers, push mowers, riding mowers, trimmers, brush cutters, pruners, edgers, vending machines,

[0217] Ventilation, Heating, Air-Conditioning, and Commercial Refrigeration Equipment

[0218] Monitoring systems, control systems

[0219] Engine, Turbine, and Power Transmission Equipment

[0220] Monitoring systems, control systems

[0221] Other General Purpose Machinery Manufacturing

[0222] Monitoring systems, control systems

[0223] Telecommunications

[0224] Monitoring systems, control systems

[0225] Portable

[0226] Aircraft

[0227] Monitoring systems, control systems, actuator systems, sensors

[0228] It should be noted that devices within a specific category may be applicable across multiple areas even if they are not specifically listed. (e.g., temperature sensors apply to Industrial and Building Automation and Appliances).

[0229] To retrofit or redesign the devices listed, it is possible to implement the described systems in numerous ways. It may be advantageous to leave the device design as-is including the existing power supply. As an example, a device may use non-rechargeable batteries to operate. The device will most likely have a protection circuit to prevent damage if the batteries are installed incorrectly. The protection mechanism is commonly a diode inline with the positive terminal of the battery. In this case, the RF power harvester **20** with or without the alternative power source **24** could be inserted, with an antenna, into the device. The power generated by the RF power harvester **20** (and the alternative power source **24**, if applicable) could be connected to the device after the protection mechanism described to avoid potential charging of a non-rechargeable battery.

[0230] Another way to configure the device's power system is to replace the non-rechargeable batteries with rechargeable batteries. In this instance, the output from the RF power harvester **20** (and alternative power source **24**) could be connected to either side of the protections device. If the connection is before the protection mechanism, the RF power harvester **20** (and the alternative power source **24**, if applicable) will recharge the battery and supply power to the device. If the connection is after the protection mechanism, the RF power harvester **20** (and the alternative power source **24**, if applicable) will supply power to the device and the battery will supply any extra power needed that could not be supplied by the RF power harvester **20** (and the alternative power source **24**, if applicable). It should be noted that the protection device in this case is not needed for proper operation. Its only function would be to protect the batteries from being installed incorrectly. An antenna could be contained inside or placed on the outside of the device.

[0231] Another configuration of the device's power system is to remove the existing batteries and install the RF power harvester **20** (and the alternative power source **24**) in the enclosure provided for the batteries. An antenna could be contained inside or placed on the outside of the device.

[0232] Yet another method of configuring the device's power system would be to reduce the number of batteries and replace them with the RF power harvester **20** (and the alternative power source **24**). In this case, the output from the system would be connected to the batteries in series or parallel depending on the original battery configuration. An antenna could be contained inside or placed on the outside of the device.

[0233] An additional option, would be to completely redesign the product and integrate the required circuitry and storage components into the device. This method is probably the most advantageous because it can fully take advantage of the benefits offered by the RF power harvester **20** (and the alternative power source **24**). An antenna could be contained inside or placed on the outside of the device.

[0234] If the RF power harvester **20** (and the alternative power source **24**) is used as a backup to the primary power supply, a switch could be implemented into the system in order to switch the RF power harvester **20** (and the alternative power source **24**) on when the primary source is lost. In this case, an antenna could be contained inside or placed on the outside of the device.

[0235] To show the flexibility of RF energy harvesting, several products were retrofitted to include an RF power harvester **20** (energy harvesting circuitry). These products include a wireless keyboard, a wall clock, and a desk calculator.

[0236] The wireless keyboard is an example of recharging and augmenting a battery to supply power to a device. This system is shown in **FIG. 13**. The output from the regulation circuitry recharges the battery and supplies power to the keyboard. The battery is also used to supply power to the keyboard. The keyboard also includes a separate antenna for receiving power and for data communications. The antenna configuration can be seen in **FIG. 55**.

[0237] The wall clock is an example of a direct powering system. The wall clock was retrofitted to include an RF power harvester **20** and the internal AA battery was removed. This system is shown in **FIG. 2**. The wall clock did not need regulation, but did require a capacitor for storage to supply the pulse of power to move the second hand.

[0238] The calculator is an example of using RF energy harvesting with another energy harvesting technology. The calculator had an internal 1.5V coin cell battery and a small solar panel. The internal battery was removed, however, the solar panel was left intact. This system is shown in **FIG. 60**. In this system, the calculator can receive power from both the solar panel and the RF power harvester **20** to eliminate the need for a battery.

[0239] As an additional example, an RF power harvester **20** (energy harvesting circuit) similar to the ones shown in U.S. Pat. No. 6,615,074 (**FIGS. 8, 9**, **12***a*, **12***b*, **13**, **14**), was connected in series with a 0.5V solar cell. Individually, the solar cell was able to provide 0.480V to a 10 kilo-ohm resistor, which was being used to simulate the core device **22** components. This corresponds to 23 microwatts. The RF power harvesting circuit by itself was able to provide 2.093V across the 10 kilo-ohm resistor when being supplied by 1 milliwatt of RF power. This corresponds to 438 microwatts. The two circuit outputs were then combined in series by connecting the output from the RF energy harvesting circuit to the ground of the solar cell. The output of the solar cell was then connected to the resistor. The other end of the resistor was connected to the ground of the RF energy harvesting circuit. The voltage across the resistor with the circuits connected as shown in **FIG. 63** was 2.445V. This corresponds to 598 microwatts.

[0240] As can be seen, the combination of the two technologies produces a result higher than the addition of the individual powers. From this, it can be determined that the two technologies can cooperate in a way that produces favorable results.

[0241] In the example given, the solar cell produces current to supply the load and helps to bias the RF rectifying diodes, which allows the RF energy harvesting circuit to operate a higher efficiency. The solar cell also changes the impedance seen by the RF energy harvesting circuit, which produces a beneficial result. To be more specific, when examining the power output of the individual circuit (solar and RF power harvesting), the sum of the power captured by the individual circuits was 23 uW+438 uW=**461** uW. However, when the two circuits are combined and are allowed to work in conjunction with one another, the output power becomes 598 uW. This result shows that combining the two power-harvesting technologies produces a 30 percent increase in the output power for this example. This same technique can be applied to multiple energy harvesting technologies to produce even greater output power. The equations for this example are shown below.

Individual Circuits

$$P_I = P_1 + P_2 + \ldots + P_N$$

Combined Circuits

$$P_C > P_I = P_1 + P_2 + \ldots + P_N$$

[0242] where P_I is the sum of the individual output powers

[0243] P_C is the output of the combined circuit

[0244] P_1 is the output power from the first power harvesting technology

[0245] P_2 is the output power from the second power harvesting technology

[0246] P_N is the output power from the Nth power harvesting technology

[0247] N is the number of power harvesting technologies or circuits

[0248] As can be seen by the previous examples, RF energy harvesting can be used alone or in conjunction with alternative power sources to power a wide range of devices. The addition of RF energy harvesting technology to the device allows for increased battery life, increased functionality, or the removal of the primary battery.

[0249] Although the invention has been described in detail in the foregoing embodiments for the purpose of illustration, it is to be understood that such detail is solely for that purpose and that variations can be made therein by those skilled in the art without departing from the spirit and scope of the invention except as it may be described by the following claims.

1. An apparatus for an application comprising:

a core device having an integrated circuit for the application; and

a power harvester connected to the core device to power the integrated circuit of the core device.

2. An apparatus as described in claim 1 including an alternative power source connected to the core device to power the core device in conjunction with the power harvester.

3. An apparatus as described in claim 1 including a power regulator and/or power storage circuit connected to the power harvester.

4. An apparatus as described in claim 1 including a power storage charger connected to the power harvester.

5. An apparatus as described in claim 1 including power storage connected to the power harvester.

6. An apparatus as described in claim 1 wherein the core device includes a memory connected to the integrated circuit and to the power harvester to power the memory.

7. An apparatus as described in claim 6 wherein the core device includes a sensor.

8. An apparatus as described in claim 7 wherein the sensor includes a proximity sensor.

9. An apparatus as described in claim 7 wherein the sensor includes an intrusion sensor.

10. An apparatus as described in claim 7 wherein the sensor includes an environmental sensor.

11. An apparatus as described in claim 7 wherein the sensor includes a chemical sensor.

12. An apparatus as described in claim 7 wherein the sensor includes a biological sensor.

13. An apparatus as described in claim 7 wherein the sensor includes a sensor in contact with an automobile.

14. An apparatus as described in claim 7 wherein the sensor includes an occupancy sensor.

15. An apparatus as described in claim 7 wherein the sensor includes a motion sensor.

16. An apparatus as described in claim 7 wherein the sensor includes a position sensor.

17. An apparatus as described in claim 7 wherein the sensor includes a metal detector.

18. An apparatus as described in claim 7 wherein the sensor includes a sensor in contact with an aircraft.

19. An apparatus as described in claim 7 wherein the sensor includes an alarm connected to the power harvester to power the alarm.

20. An apparatus as described in claim 7 wherein the sensor includes a display connected to the power harvester to power the display.

21. An apparatus as described in claim 7 wherein the sensor includes a sensor disposed in a building.

22. An apparatus as described in claim 7 wherein the sensor includes an industrial automation sensor.

23. An apparatus as described in claim 7 wherein the sensor includes a sensor in contact with an elevator.

24. An apparatus as described in claim 7 wherein the sensor includes a temperature sensor.

25. An apparatus as described in claim 7 wherein the sensor includes a fire sensor.

26. An apparatus as described in claim 7 wherein the sensor includes an accelerometer.

27. An apparatus as described in claim 7 wherein the sensor includes a level sensor.

28. An apparatus as described in claim 27 wherein the sensor includes a gas level sensor, a fluid level sensor, or a light level sensor.

29. An apparatus as described in claim 7 wherein the sensor includes a flow sensor.

30. An apparatus as described in claim 29 wherein the sensor includes a gas flow sensor, a fluid flow sensor, a light flow sensor, or a plasma flow sensor.

31. An apparatus as described in claim 7 wherein the sensor includes a pressure sensor.

32. An apparatus as described in claim 31 wherein the sensor includes a gas pressure sensor or a fluid pressure sensor.

33. An apparatus as described in claim 7 wherein the sensor includes a light sensor.

34. An apparatus as described in claim 33 wherein the sensor includes an infrared light sensor, an ultraviolet light sensor, an x-ray sensor, a cosmic ray sensor, a visible light sensor, or a gamma ray sensor.

35. An apparatus as described in claim 7 wherein the sensor includes a stress sensor.

36. An apparatus as described in claim 7 wherein the sensor includes a strain sensor.

37. An apparatus as described in claim 7 wherein the sensor includes a depth sensor.

38. An apparatus as described in claim 7 wherein the sensor includes an electrical characteristic sensor.

39. An apparatus as described in claim 7 wherein the sensor includes a voltage sensor.

40. An apparatus as described in claim 7 wherein the sensor includes a current sensor.

41. An apparatus as described in claim 7 wherein the sensor includes a viscosity sensor.

42. An apparatus as described in claim 7 wherein the sensor includes an acoustical sensor.

43. An apparatus as described in claim 7 wherein the sensor includes a sound sensor.

44. An apparatus as described in claim 7 wherein the sensor includes a listening sensor.

45. An apparatus as described in claim 7 wherein the sensor includes a thickness sensor.

46. An apparatus as described in claim 7 wherein the sensor includes a density sensor.

47. An apparatus as described in claim 7 wherein the sensor includes a surface quality sensor.

48. An apparatus as described in claim 7 wherein the sensor includes a volume sensor.

49. An apparatus as described in claim 7 wherein the sensor includes a physical sensor.

50. An apparatus as described in claim 7 wherein the sensor includes a mass sensor.

51. An apparatus as described in claim 7 wherein the sensor includes a weight sensor.

52. An apparatus as described in claim 7 wherein the sensor includes a conductivity sensor.

53. An apparatus as described in claim 7 wherein the sensor includes a distance sensor.

54. An apparatus as described in claim 7 wherein the sensor includes an orientation sensor.

55. An apparatus as described in claim 7 wherein the sensor includes a vibration sensor.

56. An apparatus as described in claim 7 wherein the sensor includes a radioactivity sensor.

57. An apparatus as described in claim 7 wherein the sensor includes a field strength sensor.

58. An apparatus as described in claim 57 wherein the sensor includes an electric field sensor or a magnetic field sensor.

59. An apparatus as described in claim 7 wherein the sensor includes a smoke detector.

60. An apparatus as described in claim 7 wherein the sensor includes a carbon monoxide detector.

61. An apparatus as described in claim 7 wherein the sensor includes a radon detector.

62. An apparatus as described in claim 7 wherein the sensor includes an air quality sensor.

63. An apparatus as described in claim 7 wherein the sensor includes a humidity sensor.

64. An apparatus as described in claim 7 wherein the sensor includes a glass breakage sensor.

65. An apparatus as described in claim 7 wherein the sensor includes a break beam detector.

66. An apparatus for an application comprising:

a core device having an integrated circuit for the application; and means for receiving energy wirelessly and providing power from the energy to the core device to power the integrated circuit of the core device, the receiving means connected to the core device.

67. A method for an application comprising the steps of:

converting RF energy into usable energy; and

powering an integrated circuit of a core device with the usable energy.

68. A method as described in claim 67 including the step of regulating the usable energy provided to the core device.

69. A method as described in claim 68 including the step of storing the usable energy.

70. A method as described in claim 69 including the step of providing power to the core device from an alternative power source in conjunction with the usable energy.

71. An apparatus as described in claim 6 wherein the core device includes a computer peripheral.

72. An apparatus as described in claim 71 wherein the computer peripheral includes a handheld game.

73. An apparatus as described in claim 71 wherein the computer peripheral includes a gaming system.

74. An apparatus as described in claim 71 wherein the computer peripheral includes a game controller.

75. An apparatus as described in claim 71 wherein the computer peripheral includes a controller.

76. An apparatus as described in claim 71 wherein the computer peripheral includes a keyboard.

77. An apparatus as described in claim 71 wherein the computer peripheral includes a mouse.

78. An apparatus as described in claim 71 wherein the computer peripheral includes a computer terminal.

79. An apparatus as described in claim 71 wherein the computer peripheral includes computer storage.

80. An apparatus as described in claim 71 wherein the computer peripheral includes computer equipment.

81. An apparatus as described in claim 7 wherein the sensor includes a thermal energy sensor.

82. An apparatus as described in claim 7 wherein the sensor includes an electromagnetic sensor.

83. An apparatus as described in claim 7 wherein the sensor includes a mechanical sensor.

84. An apparatus as described in claim 7 wherein the sensor includes an optical sensor.

85. An apparatus as described in claim 7 wherein the sensor includes a radiation sensor.

86. An apparatus as described in claim 7 wherein the sensor includes a sensor in contact with a vehicle.

87. An apparatus as described in claim 7 wherein the sensor includes a sensor in contact with a water craft.

* * * * *

US 20090152954A1

(19) **United States**

(12) **Patent Application Publication**

Le et al.

(10) **Pub. No.: US 2009/0152954 A1**

(43) **Pub. Date: Jun. 18, 2009**

(54) **RF ENERGY HARVESTING CIRCUIT**

(76) Inventors: **Triet Tu Le**, Portland, OR (US); **Terri S. Flez**, Corvallis, OR (US); **Kartikeya Mayaram**, Corvallis, OR (US)

Correspondence Address:
LUMEN PATENT FIRM
2345 YALE STREET, SECOND FLOOR
PALO ALTO, CA 94306 (US)

(21) Appl. No.: **12/218,761**

(22) Filed: **Jul. 17, 2008**

Related U.S. Application Data

(60) Provisional application No. 60/961,009, filed on Jul. 17, 2007, provisional application No. 60/993,260, filed on Sep. 10, 2007.

Publication Classification

(51) **Int. Cl.**
H02M 3/18 (2006.01)
H02J 17/00 (2006.01)

(52) **U.S. Cl.** **307/110**; 307/149

(57) **ABSTRACT**

A voltage doubler rectifier RF power harvesting system is provided having at least one power harvesting module with a voltage doubler rectifier that includes a DC voltage input and output, signal input, first and second floating gate transistors that are connected in series between the DC voltage input and output, and a gate control of the floating gate is connected to a drain of the transistor. The voltage double rectifier includes a first capacitor disposed between the input signal and the drain of the first floating gate transistor, and a second capacitor disposed between a ground and the drain of the second floating gate transistor. The system includes an input antenna and a transmitter with a transmitter antenna, where a full-wave peak-to-peak voltage of an incoming RF signal is rectified. The system also includes a powered device connected to the DC voltage output to utilize the rectified signal.

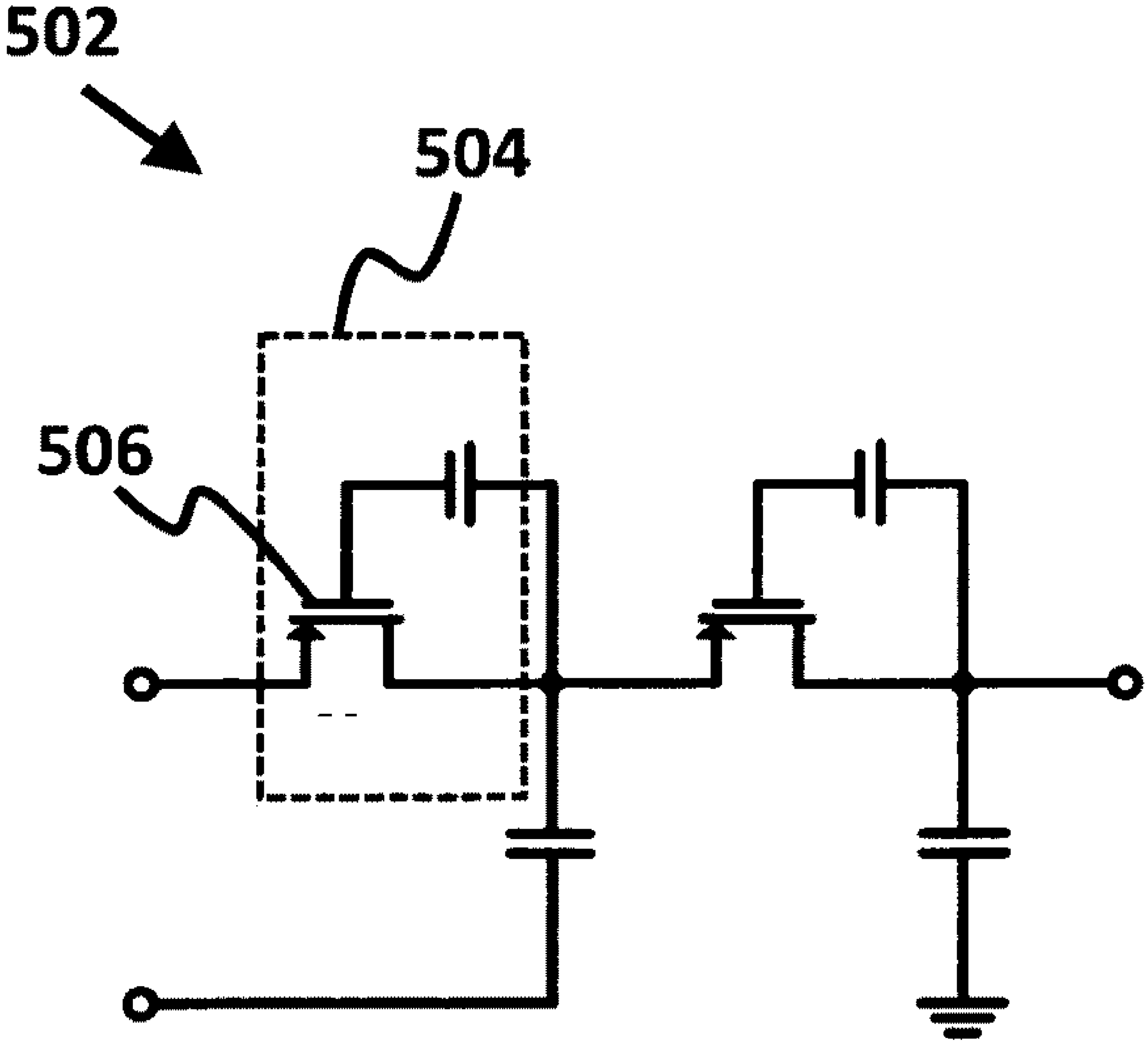

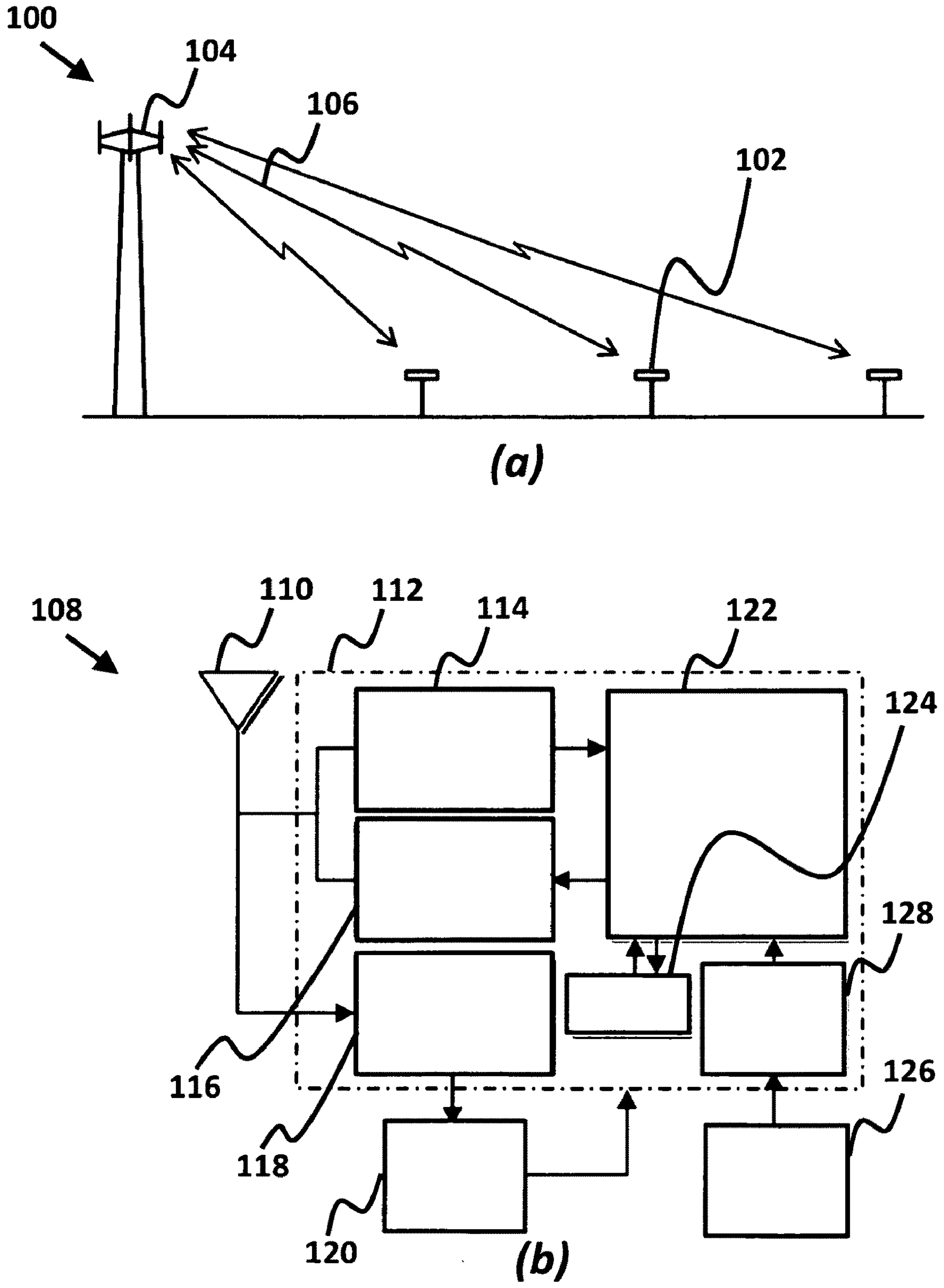
100
104
106
102
(a)
108
110
112
114
122
124
128
116
126
118
120
(b)

FIG. 1

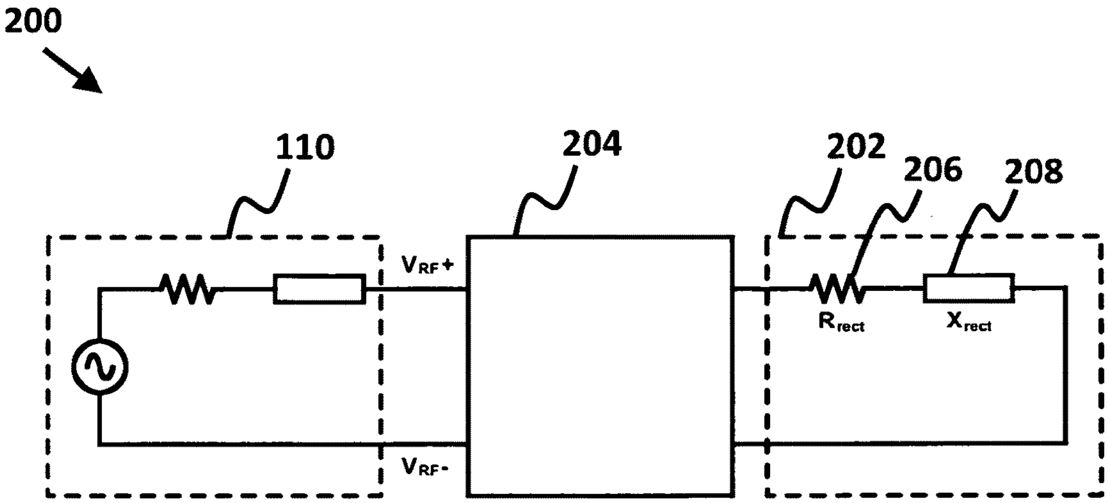
200
110
204
202
206
208
$V_{RF}+$
$V_{RF}-$
R_{rect}
X_{rect}

FIG. 2

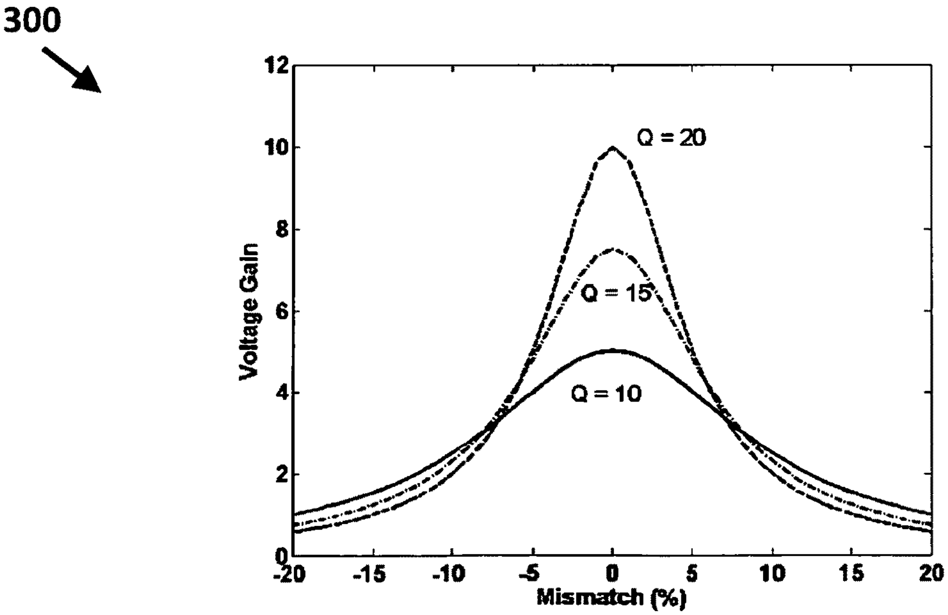
300
Q = 20
Q = 15
Q = 10
Voltage Gain
Mismatch (%)
0
2
4
6
8
10
12
-20
-15
-10
-5
0
5
10
15
20

FIG. 3

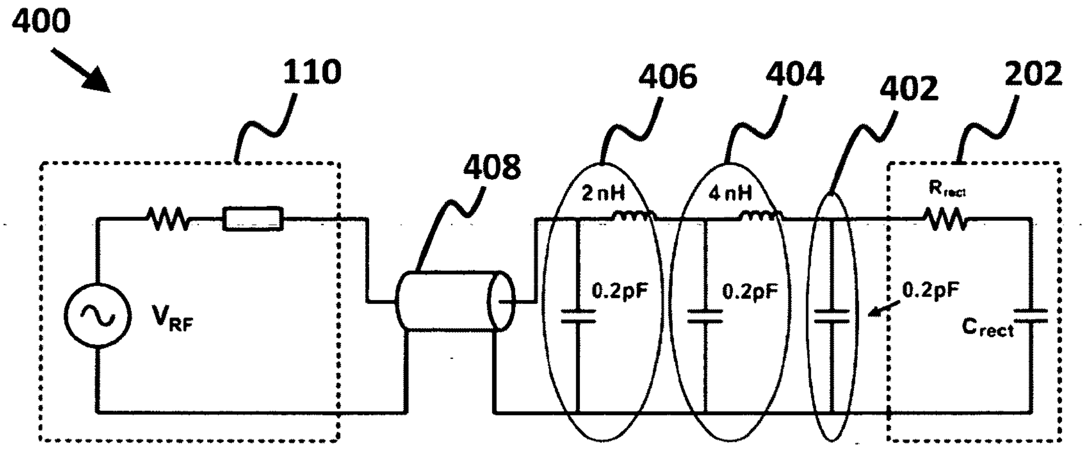
400
110
406
404
402
202
408
2 nH
4 nH
R_{rect}
V_{RF}
0.2pF
0.2pF
0.2pF
C_{rect}

FIG. 4

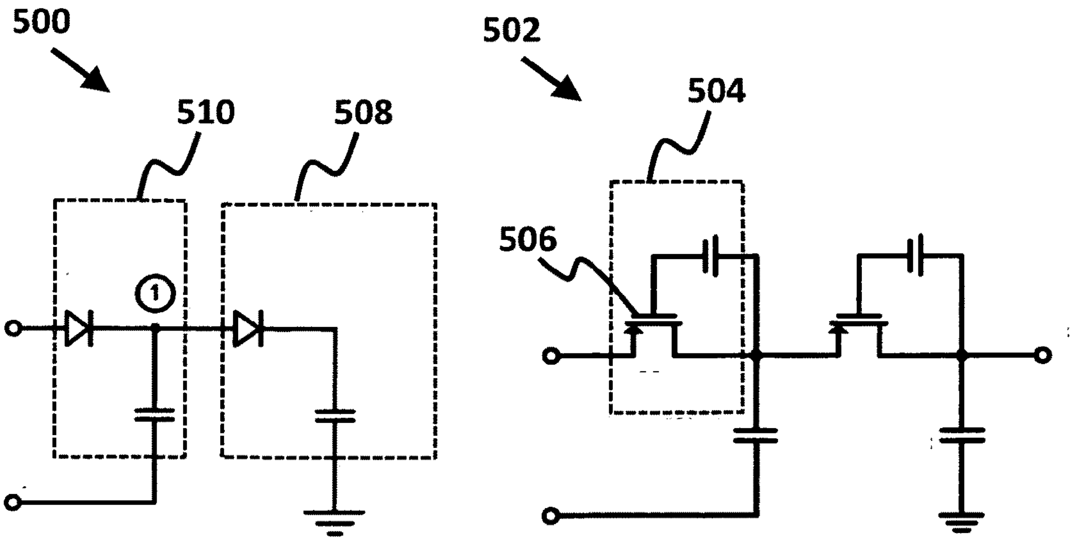
500
510
508
①
502
504
506

(a) prior art (b)

FIG. 5

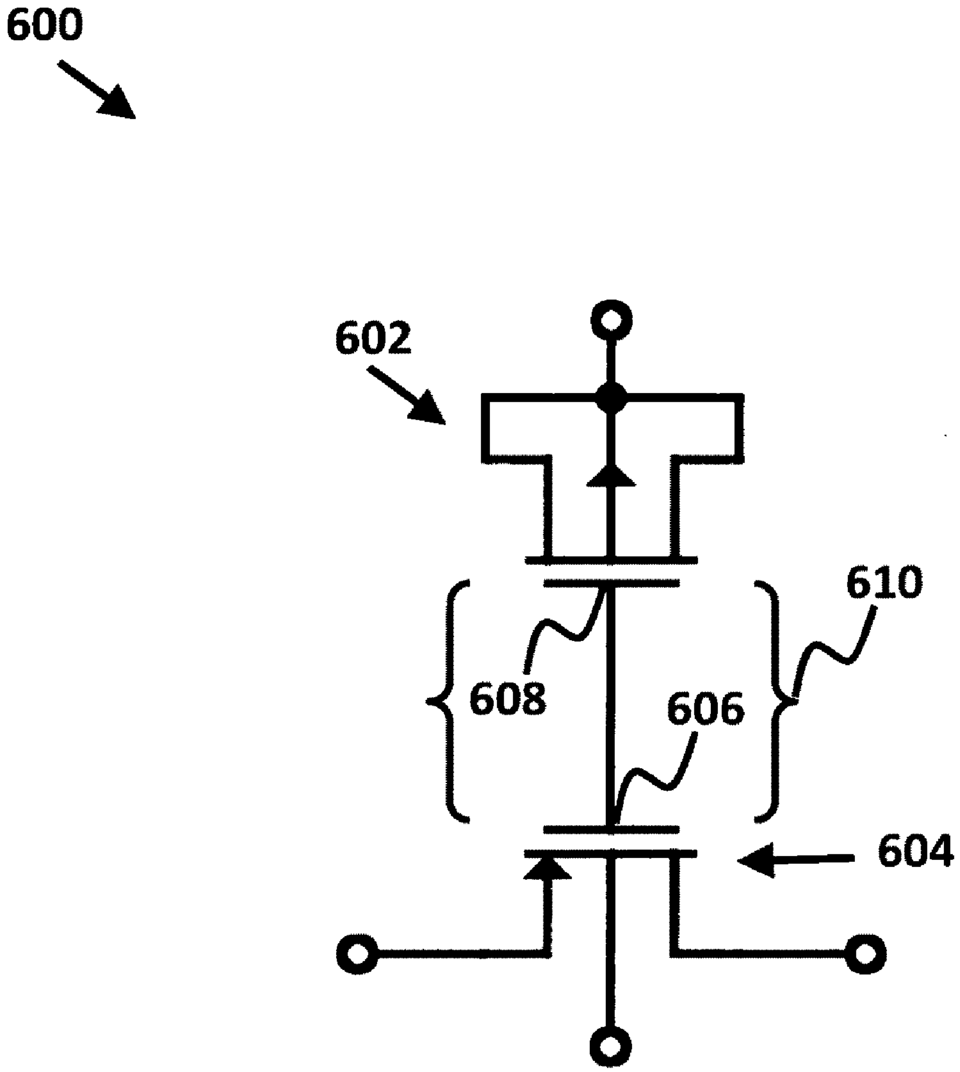
600
602
610
608
606
604

FIG. 6

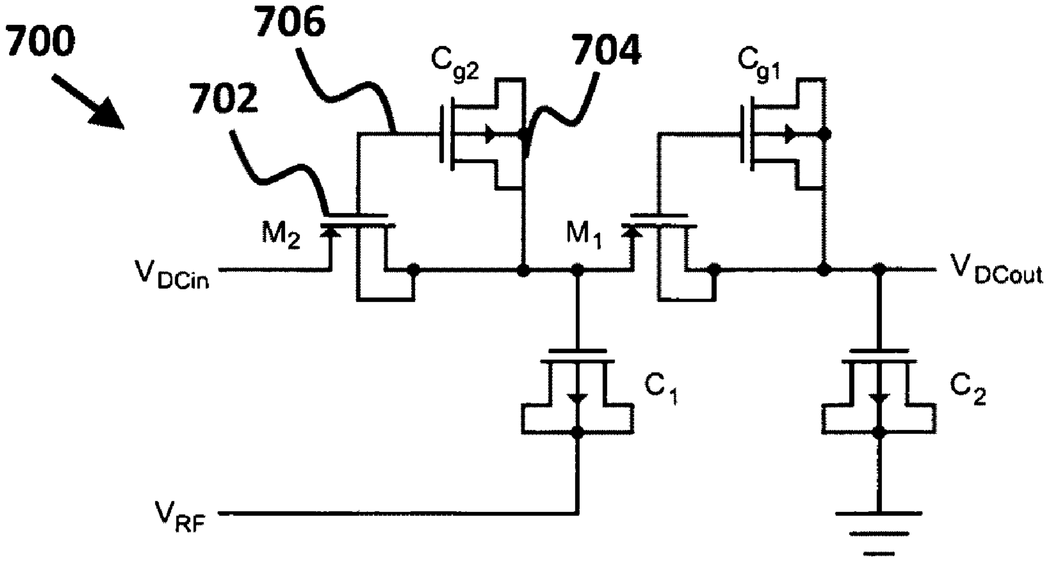
700
706
702
704
C_{g2}
C_{g1}
M_2
M_1
V_{DCin}
V_{DCout}
C_1
C_2
V_{RF}

FIG. 7

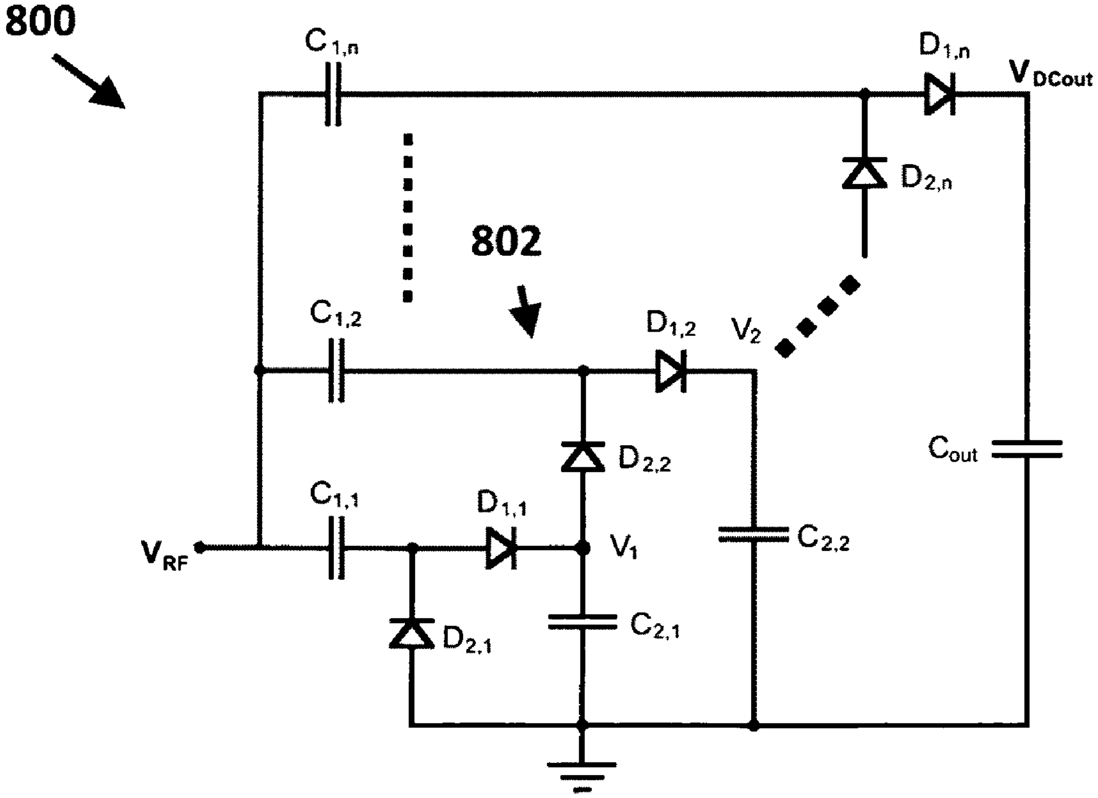
800
$C_{1,n}$
$D_{1,n}$
V_{DCout}
$D_{2,n}$
802
$C_{1,2}$
$D_{1,2}$
V_2
$D_{2,2}$
C_{out}
$C_{1,1}$
$D_{1,1}$
V_1
$C_{2,2}$
V_{RF}
$D_{2,1}$
$C_{2,1}$

FIG. 8

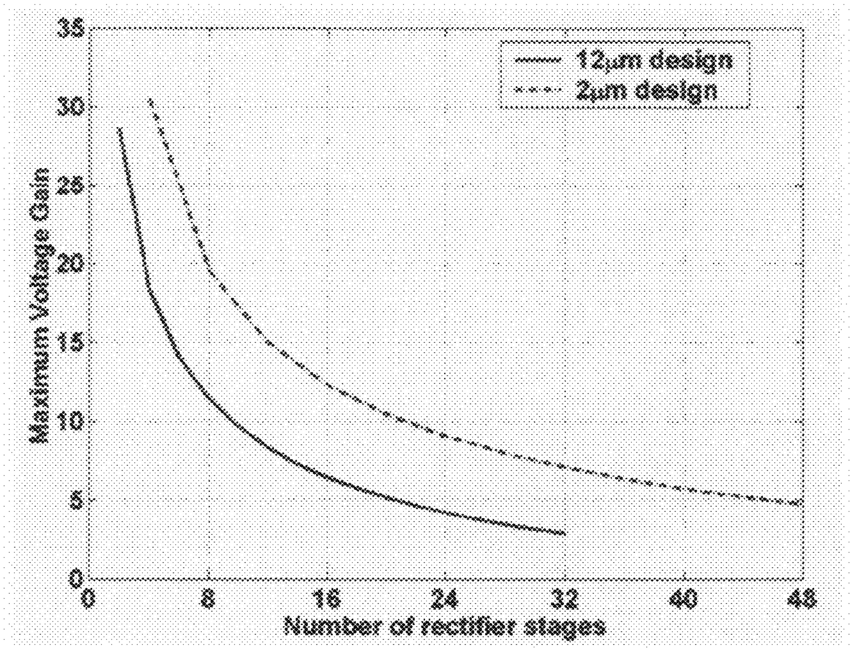
Maximum Voltage Gain
0
5
10
15
20
25
30
35
0
8
16
24
32
40
48
Number of rectifier stages
12μm design
2μm design

(a)

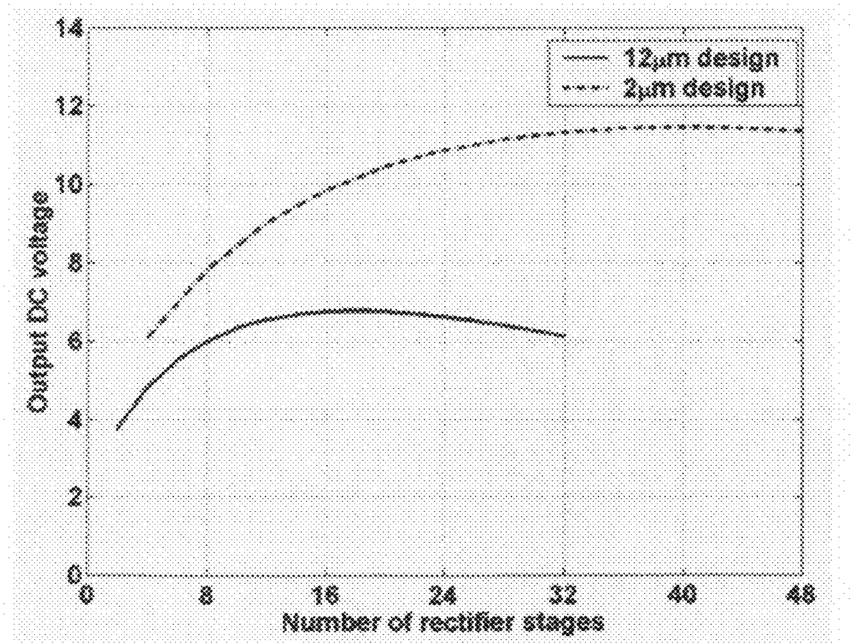
Output DC voltage
0
2
4
6
8
10
12
14
0
8
16
24
32
40
48
Number of rectifier stages
12μm design
2μm design

(b)

FIG. 9

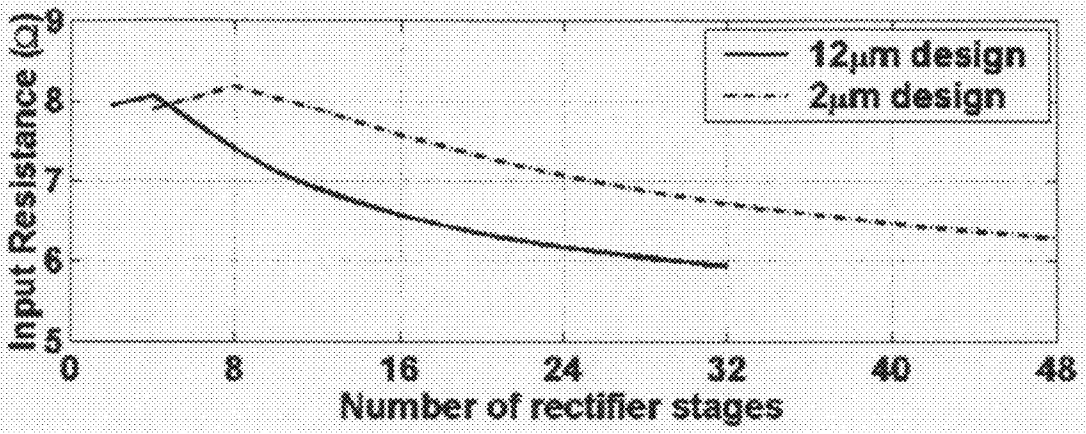
Input Resistance (Ω)
9
8
7
6
5
0
8
16
24
32
40
48
Number of rectifier stages
12μm design
2μm design

(c)

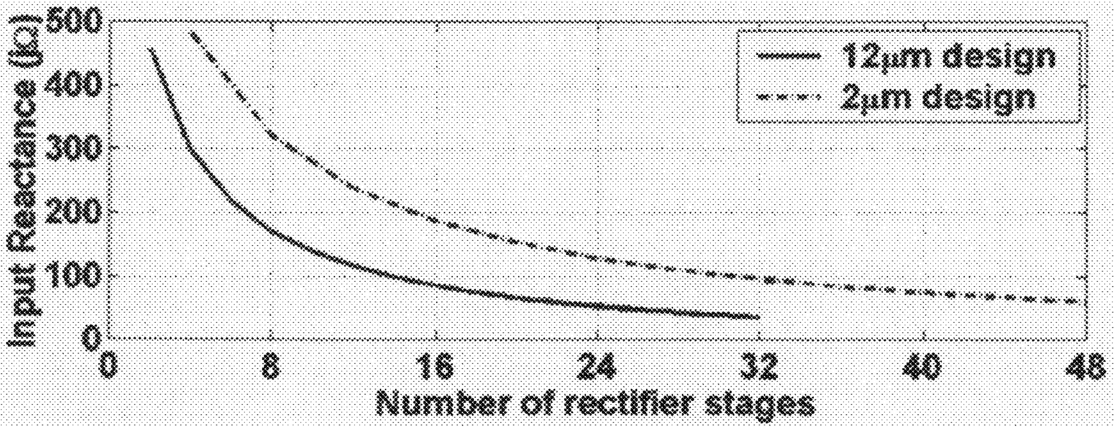
Input Reactance (jΩ)
500
400
300
200
100
0
0
8
16
24
32
40
48
Number of rectifier stages
12μm design
2μm design

(d)

FIG. 9 (cont.)

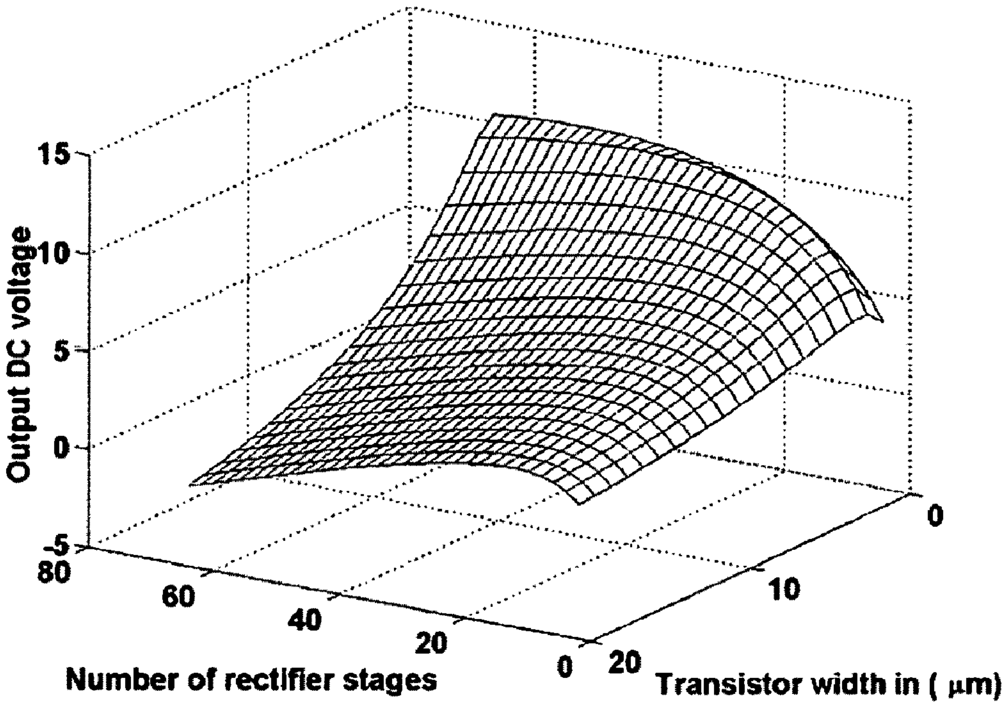
Output DC voltage
15
10
5
0
-5
80
60
40
20
0
Number of rectifier stages
0
10
20
Transistor width in (μm)

(a)

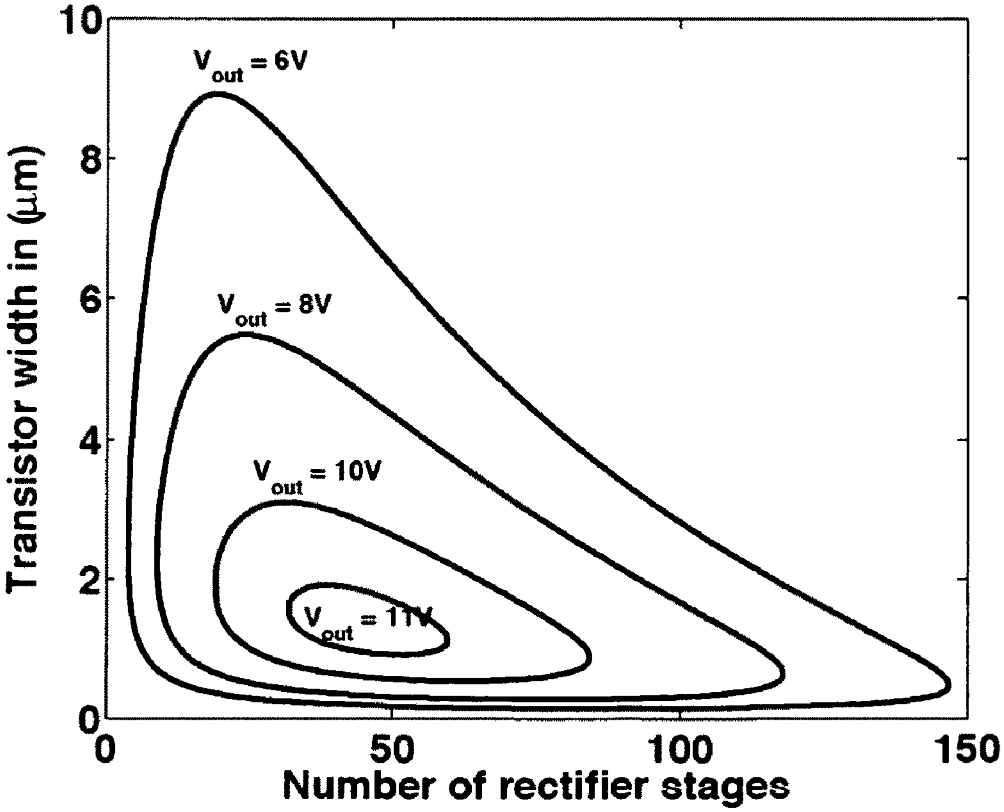
10
8
6
4
2
0
Transistor width in (μm)
0
50
100
150
Number of rectifier stages
V_out = 6V
V_out = 8V
V_out = 10V
V_out = 11V

(b)

FIG. 10

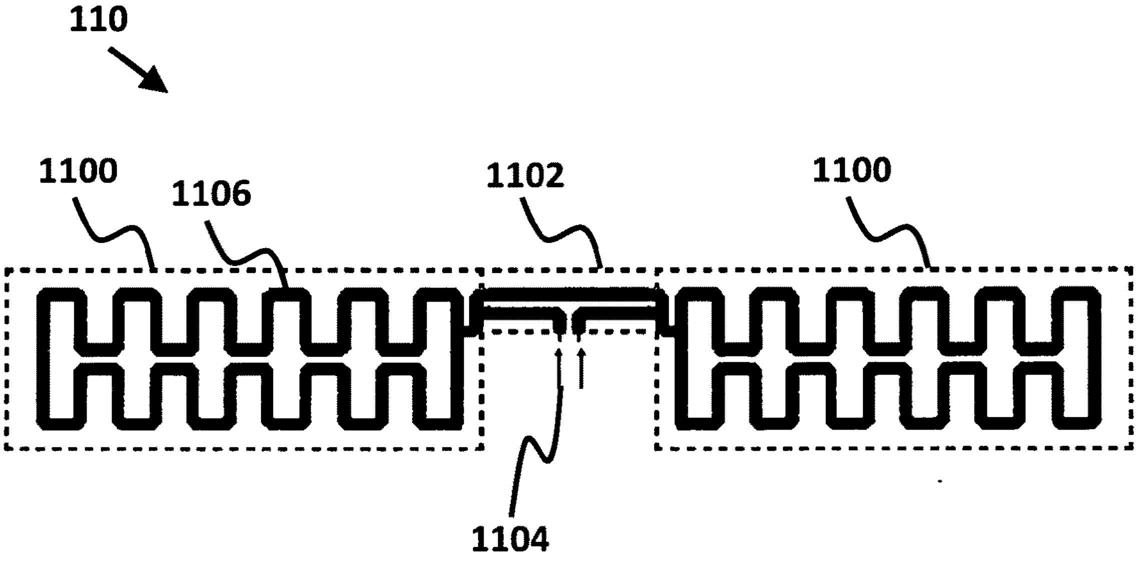
110
1100
1106
1102
1100
1104

FIG. 11

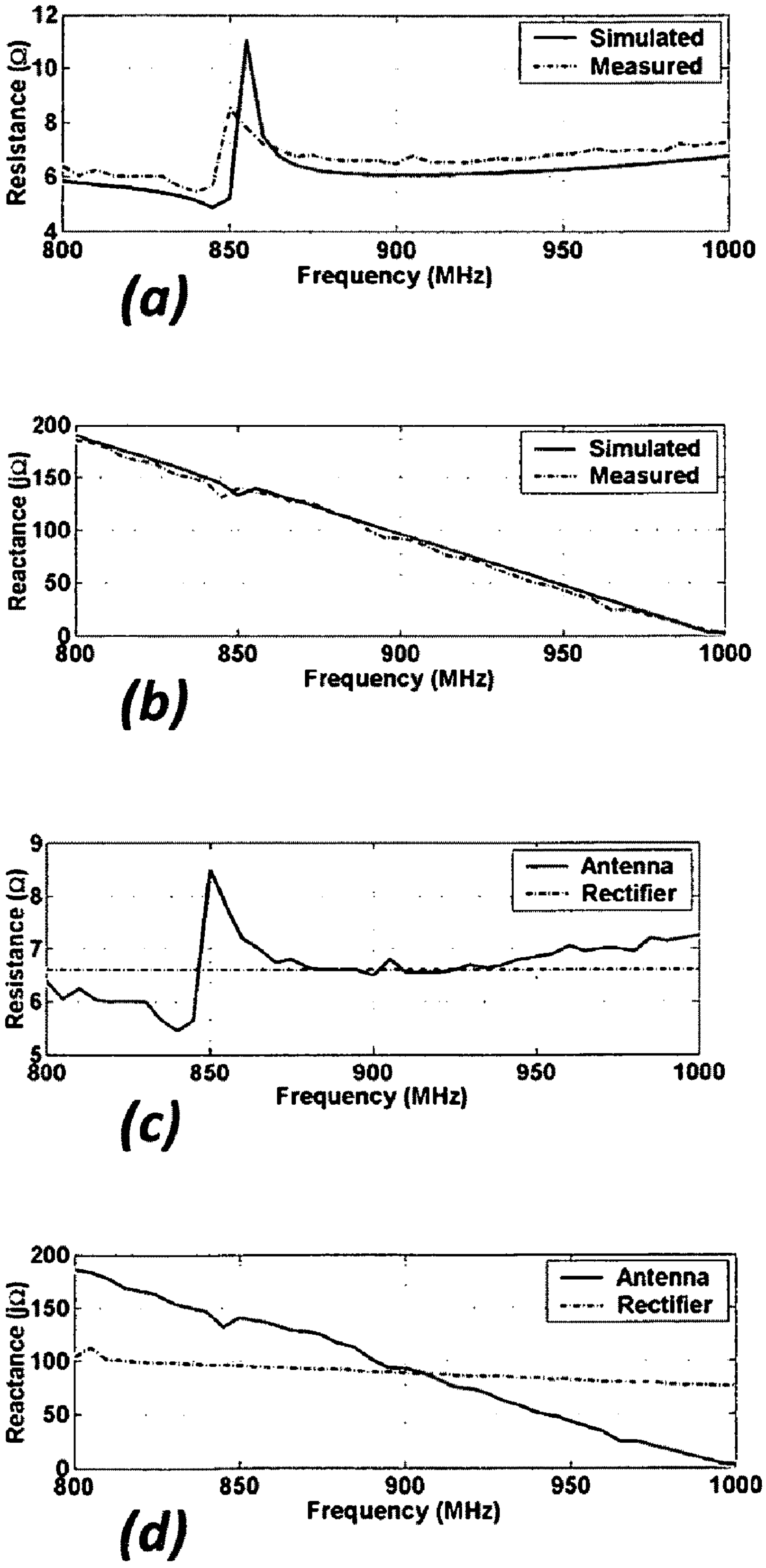
12
10
8
6
4
Resistance (Ω)
Simulated
Measured
800
850
900
950
1000
Frequency (MHz)
(a)
200
150
100
50
0
Reactance (jΩ)
Simulated
Measured
800
850
900
950
1000
Frequency (MHz)
(b)
9
8
7
6
5
Resistance (Ω)
Antenna
Rectifier
800
850
900
950
1000
Frequency (MHz)
(c)
200
150
100
50
0
Reactance (jΩ)
Antenna
Rectifier
800
850
900
950
1000
Frequency (MHz)
(d)

FIG. 12

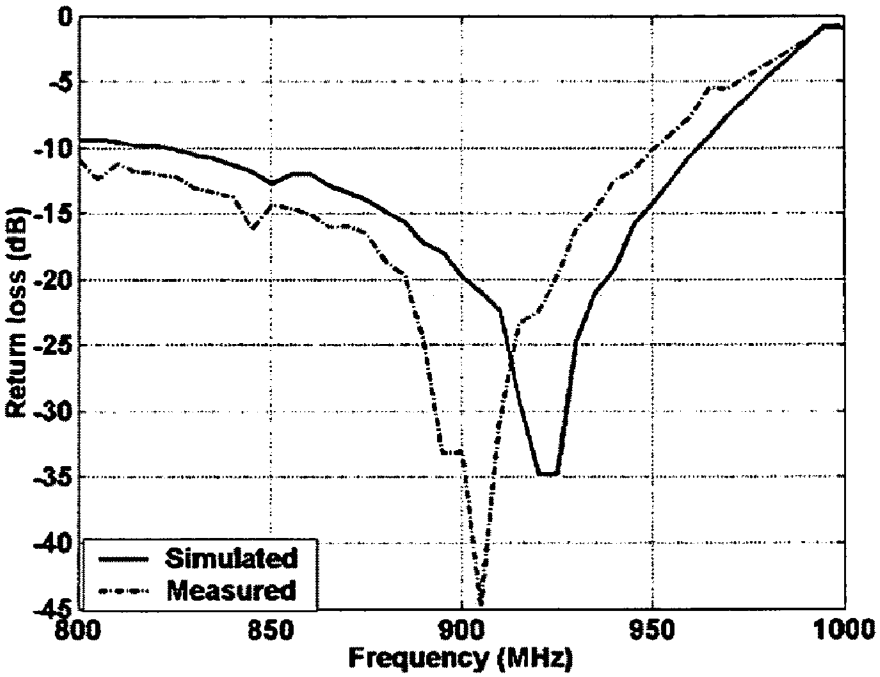
0
-5
-10
-15
-20
-25
-30
-35
-40
-45
Return loss (dB)
800
850
900
950
1000
Frequency (MHz)
Simulated
Measured

(a)

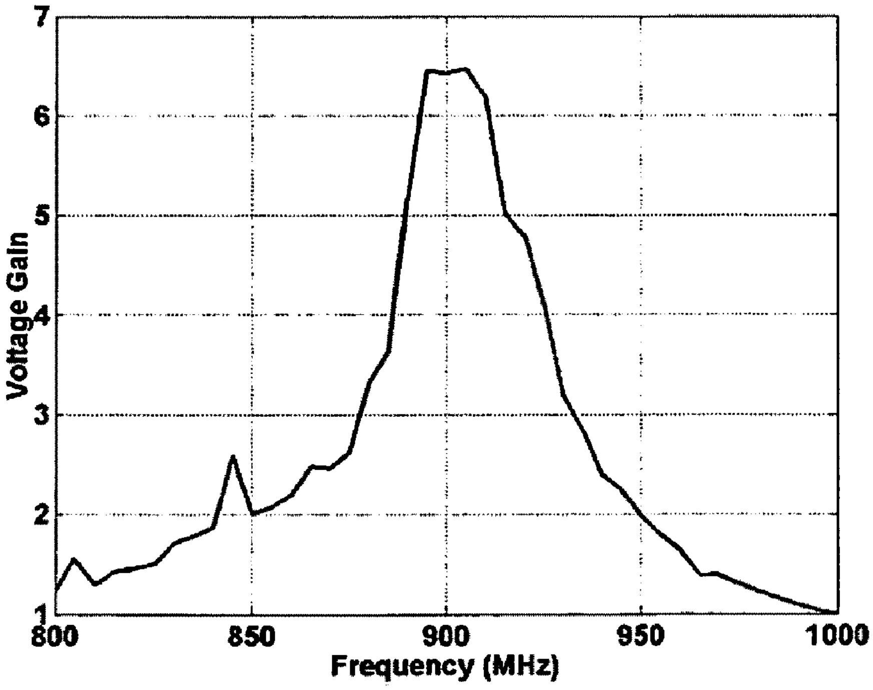
7
6
5
4
3
2
1
Voltage Gain
800
850
900
950
1000
Frequency (MHz)

(b)

FIG. 13

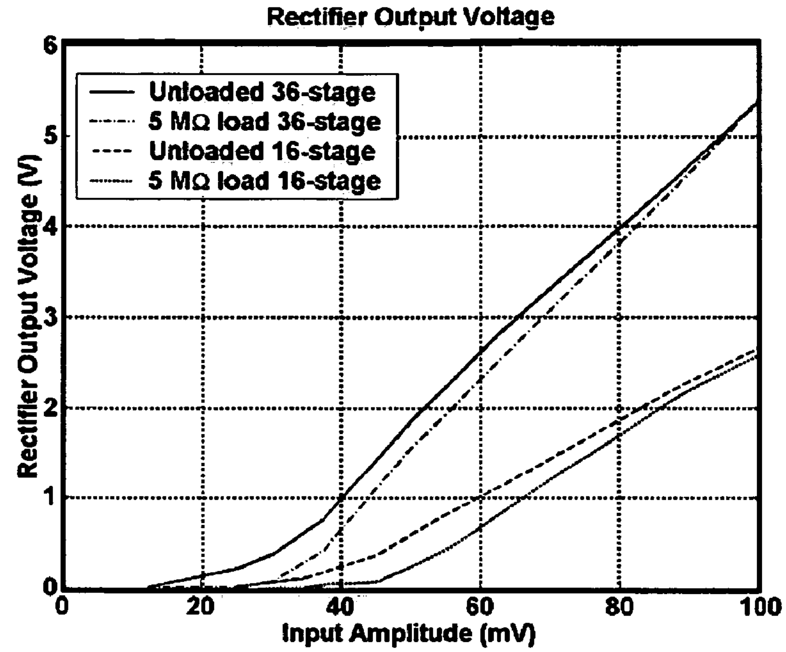
Rectifier Output Voltage
Unloaded 36-stage
5 MΩ load 36-stage
Unloaded 16-stage
5 MΩ load 16-stage
Rectifier Output Voltage (V)
0
1
2
3
4
5
6
0
20
40
60
80
100
Input Amplitude (mV)

(a)

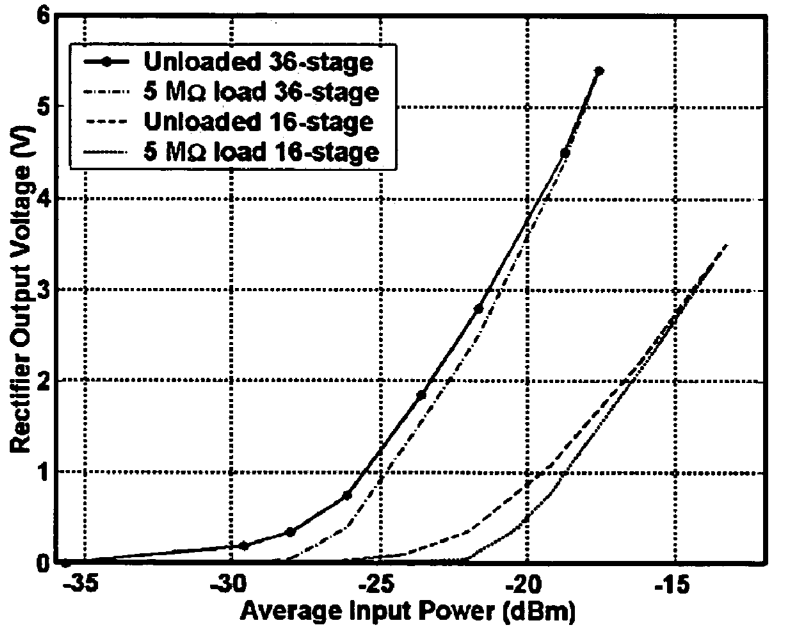
Unloaded 36-stage
5 MΩ load 36-stage
Unloaded 16-stage
5 MΩ load 16-stage
Rectifier Output Voltage (V)
0
1
2
3
4
5
6
-35
-30
-25
-20
-15
Average Input Power (dBm)

(b)

FIG. 14

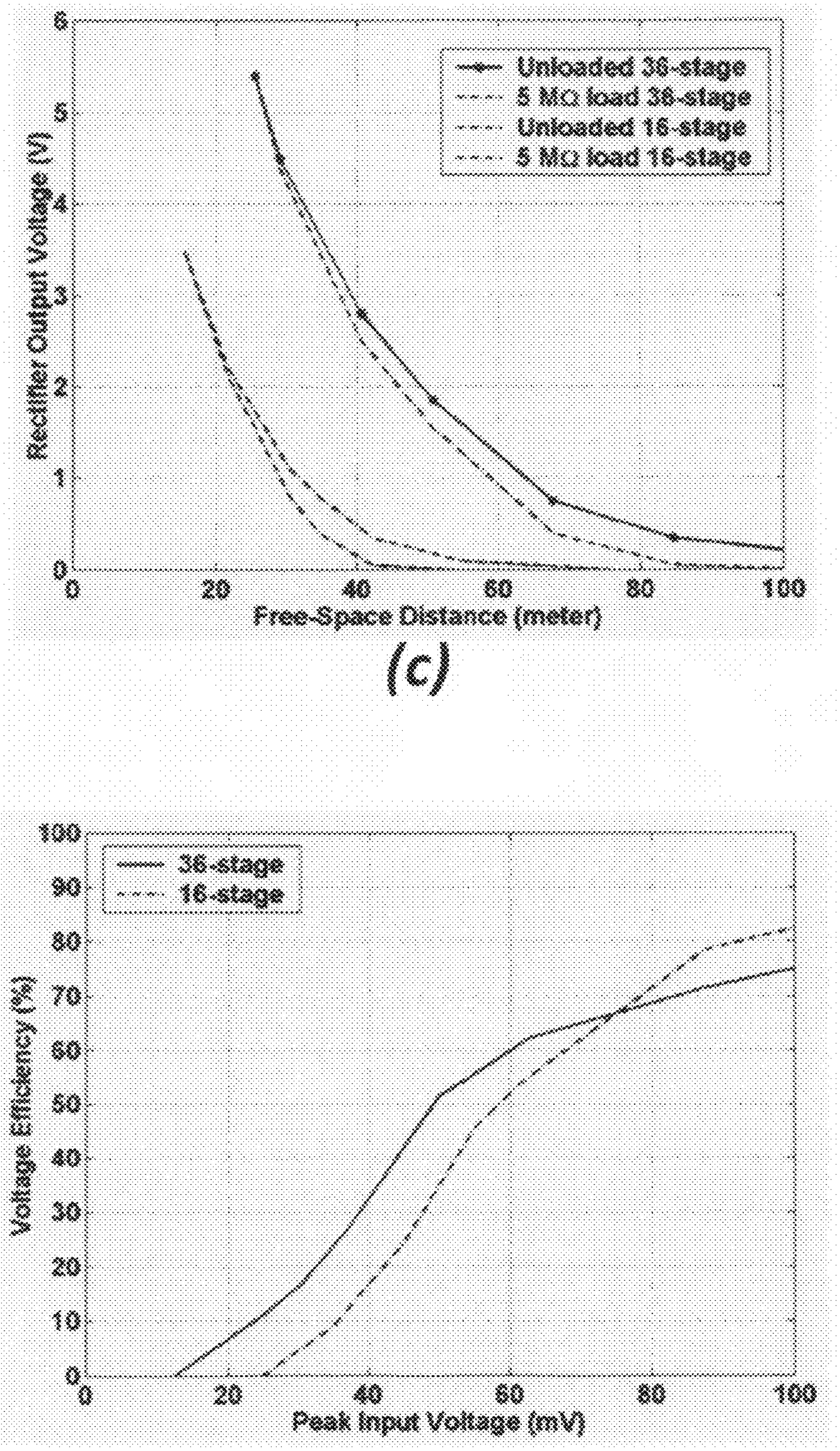

(c)

(d)

FIG. 14 (cont.)

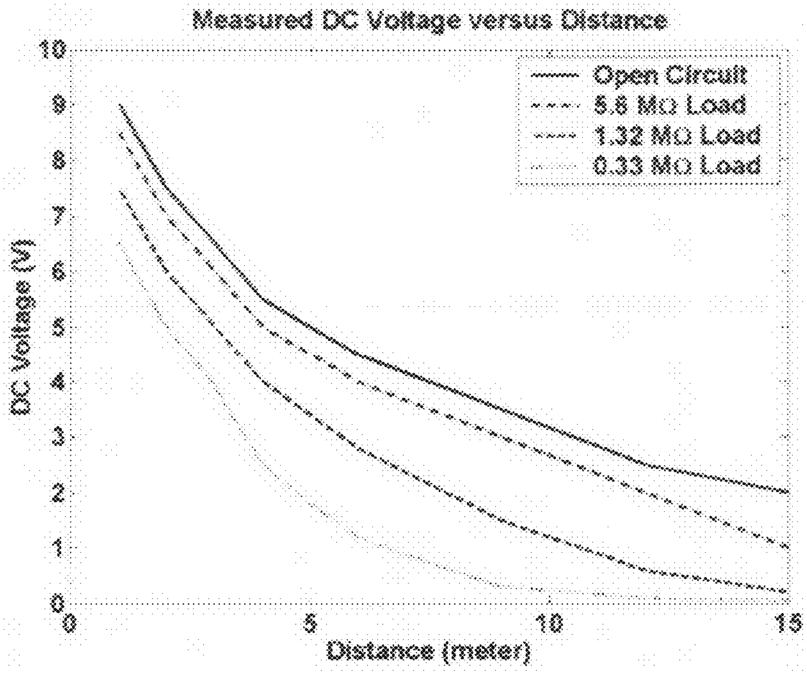
Measured DC Voltage versus Distance
Open Circuit
5.6 MΩ Load
1.32 MΩ Load
0.33 MΩ Load
DC Voltage (V)
Distance (meter)
0
1
2
3
4
5
6
7
8
9
10
0
5
10
15

(a)

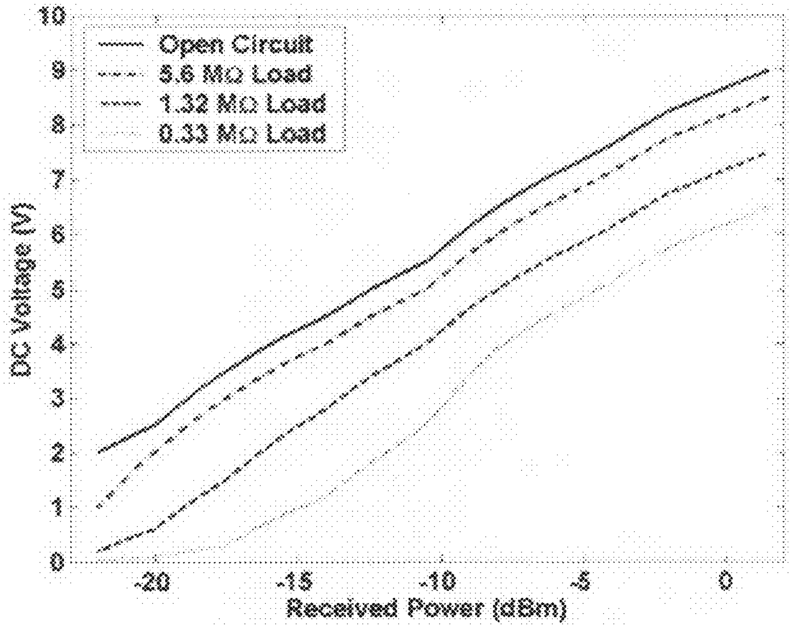
Open Circuit
5.6 MΩ Load
1.32 MΩ Load
0.33 MΩ Load
DC Voltage (V)
Received Power (dBm)
0
1
2
3
4
5
6
7
8
9
10
-20
-15
-10
-5
0

(b)

FIG. 15

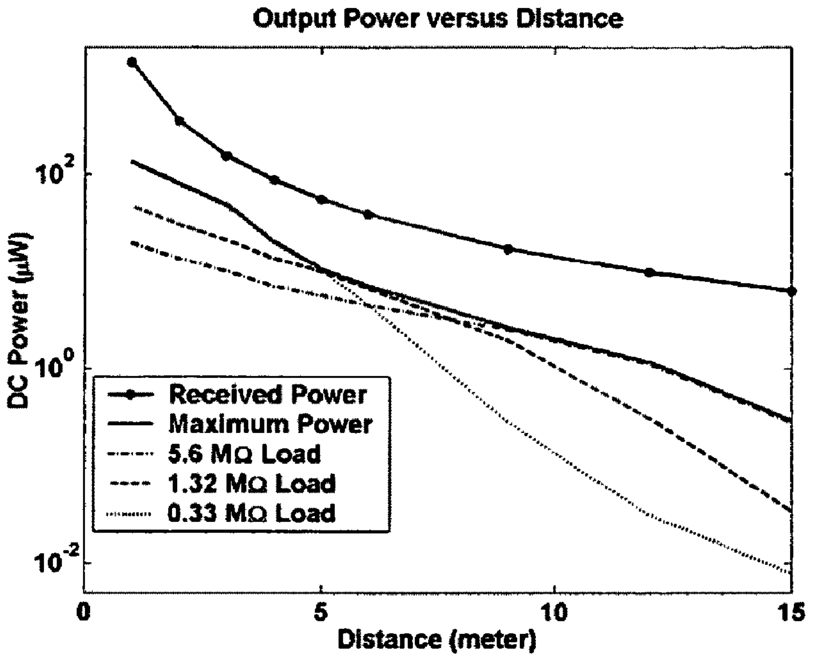

(c)

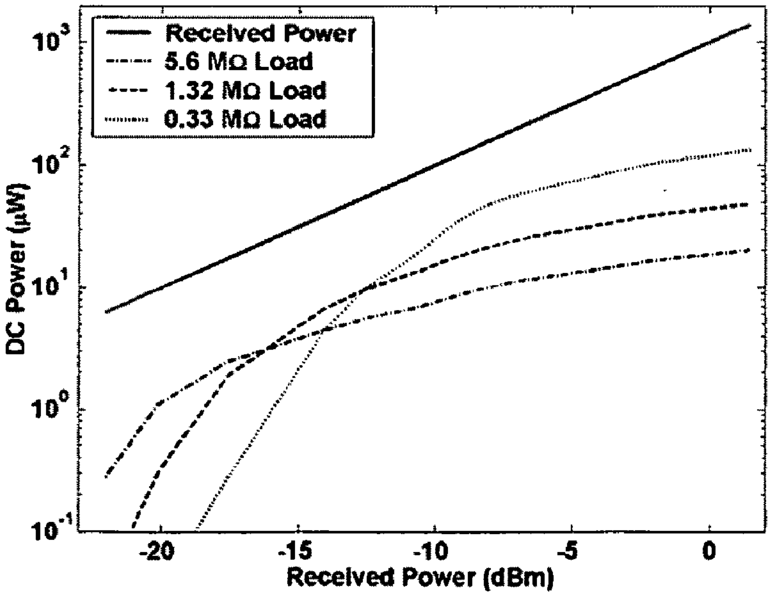

(d)

FIG. 15 (cont.)

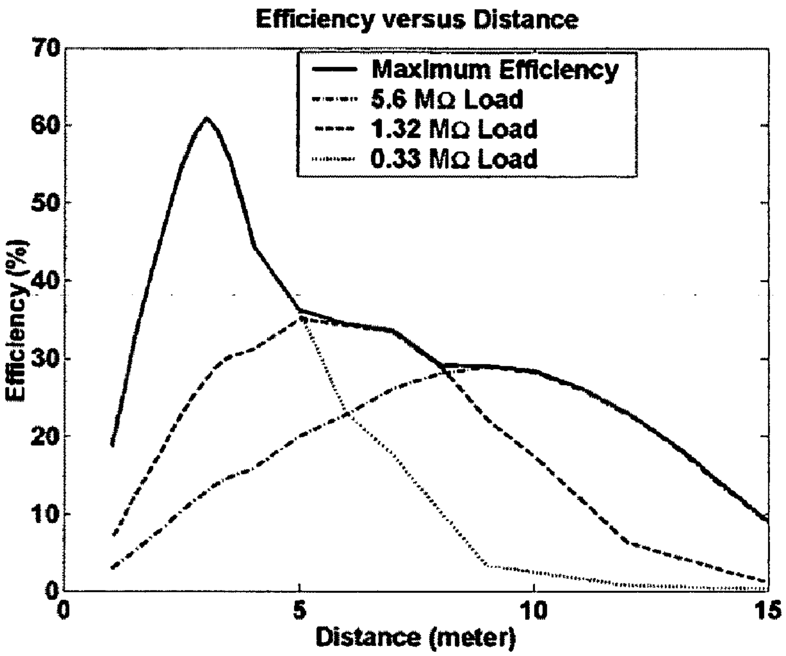

(e)

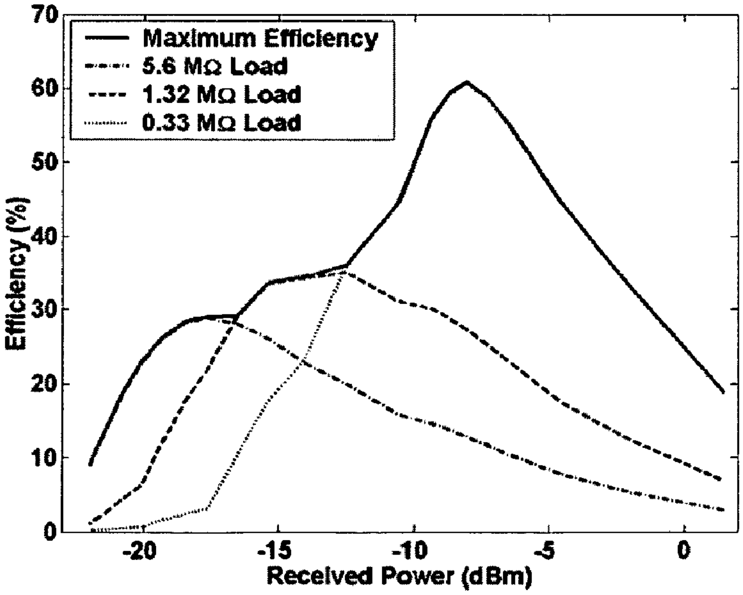

(f)

FIG. 15 (cont.)

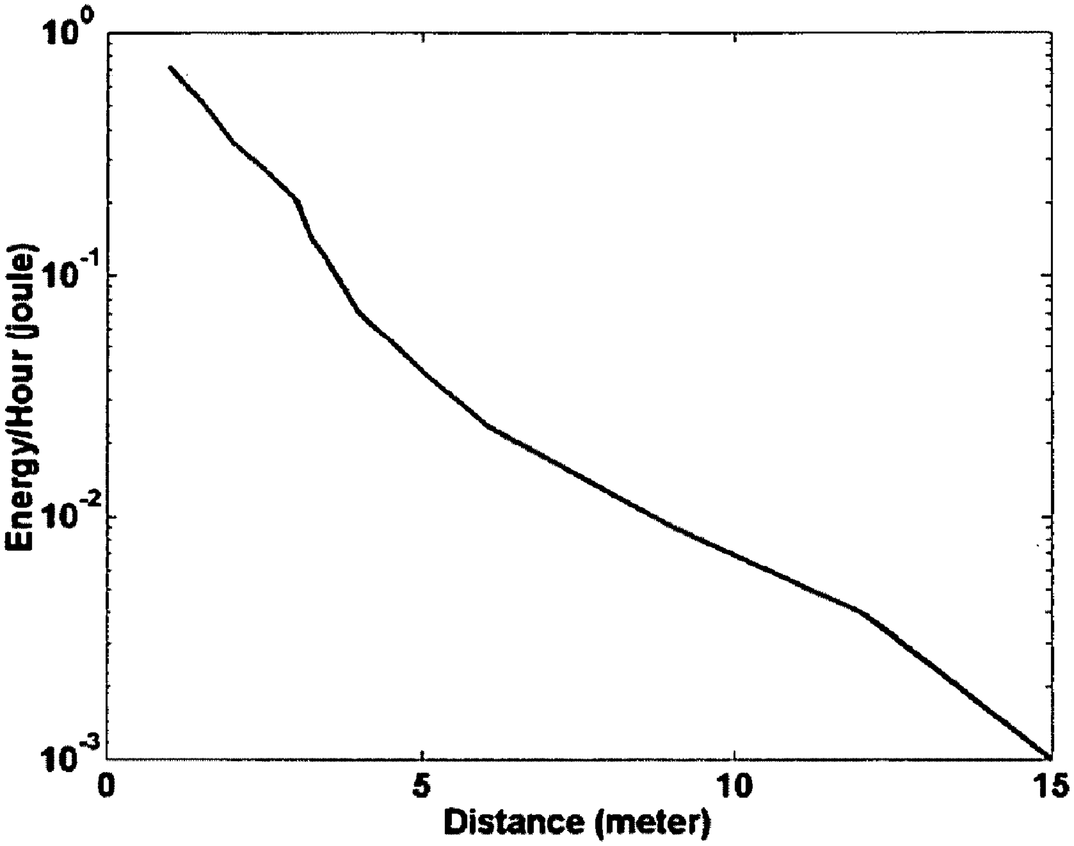

FIG. 16

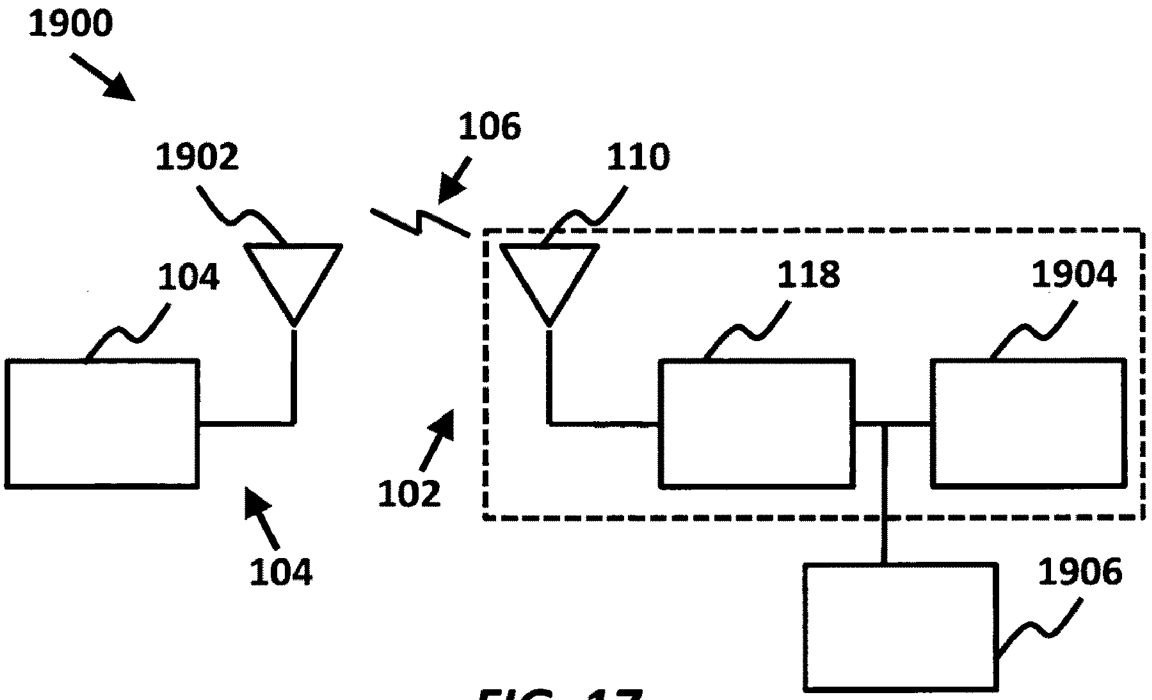
1900
1902
1904
1906

FIG. 17

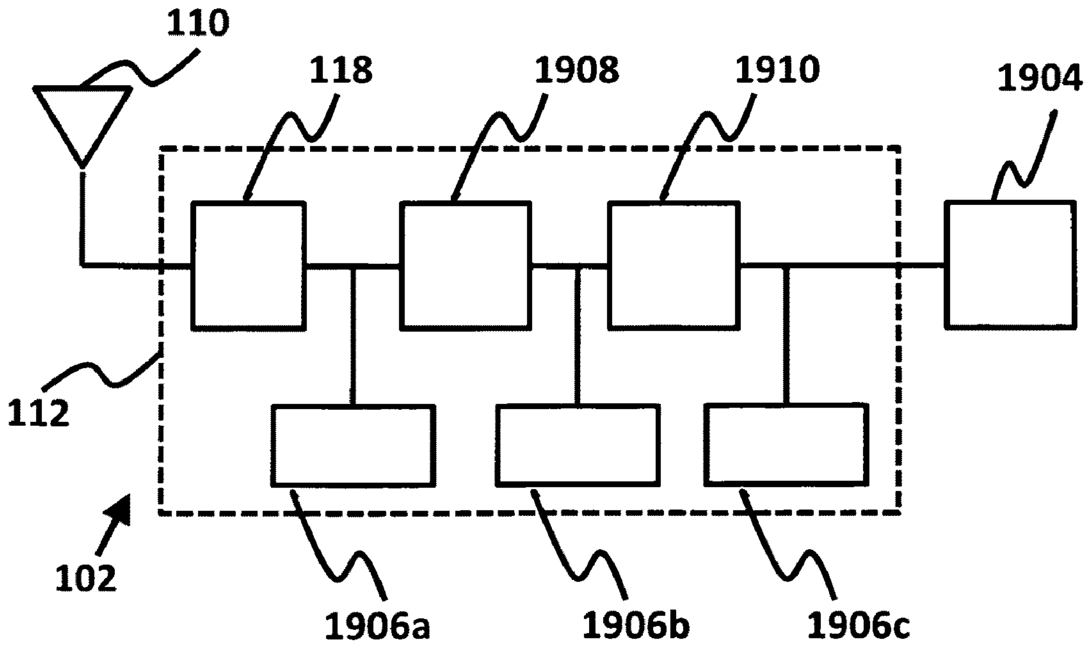
110
118
1908
1910
1904
112
102
1906a
1906b
1906c

FIG. 18

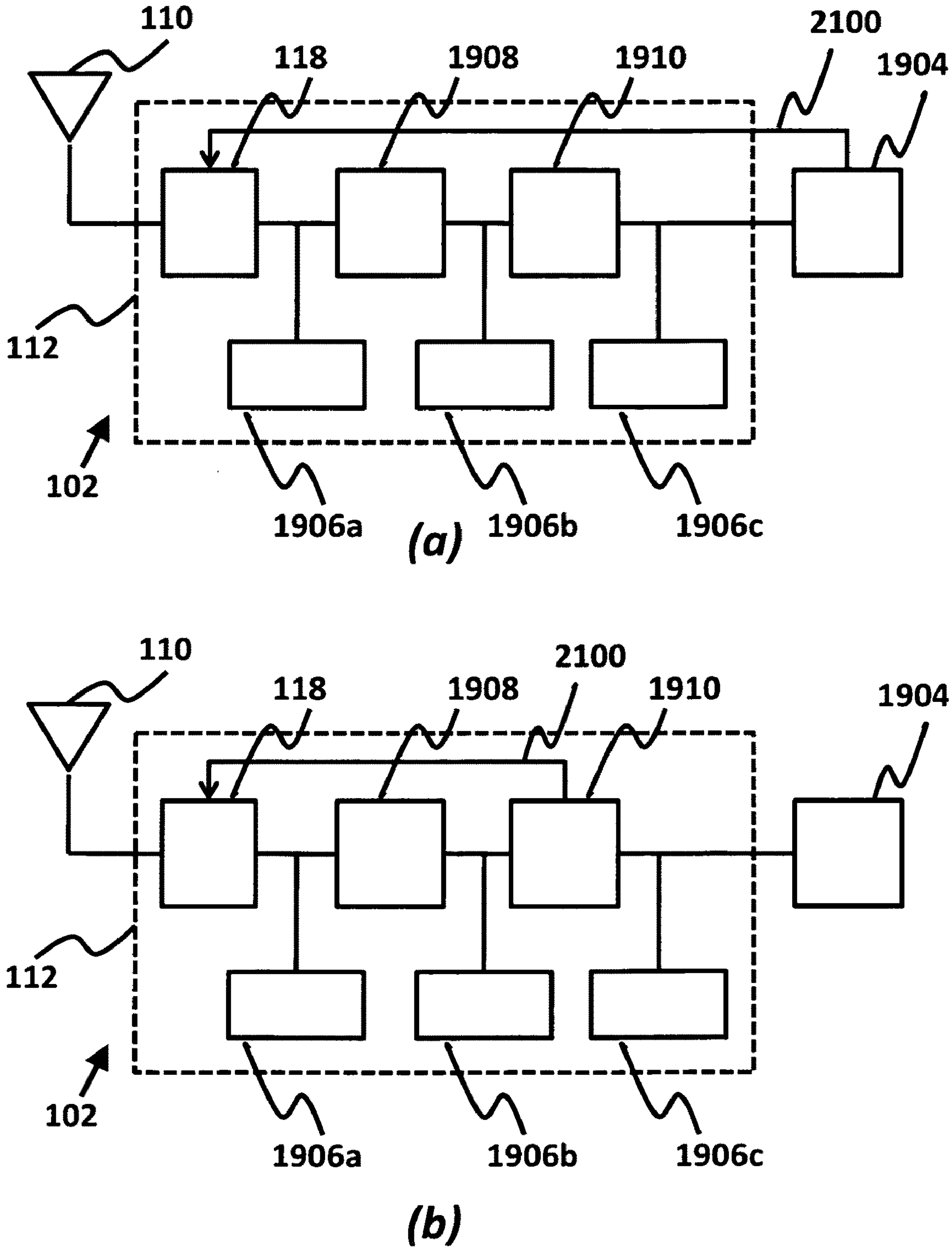
110
118
1908
1910
2100
1904
112
102
1906a
1906b
1906c
(a)
110
2100
118
1908
1910
1904
112
102
1906a
1906b
1906c
(b)

FIG. 19

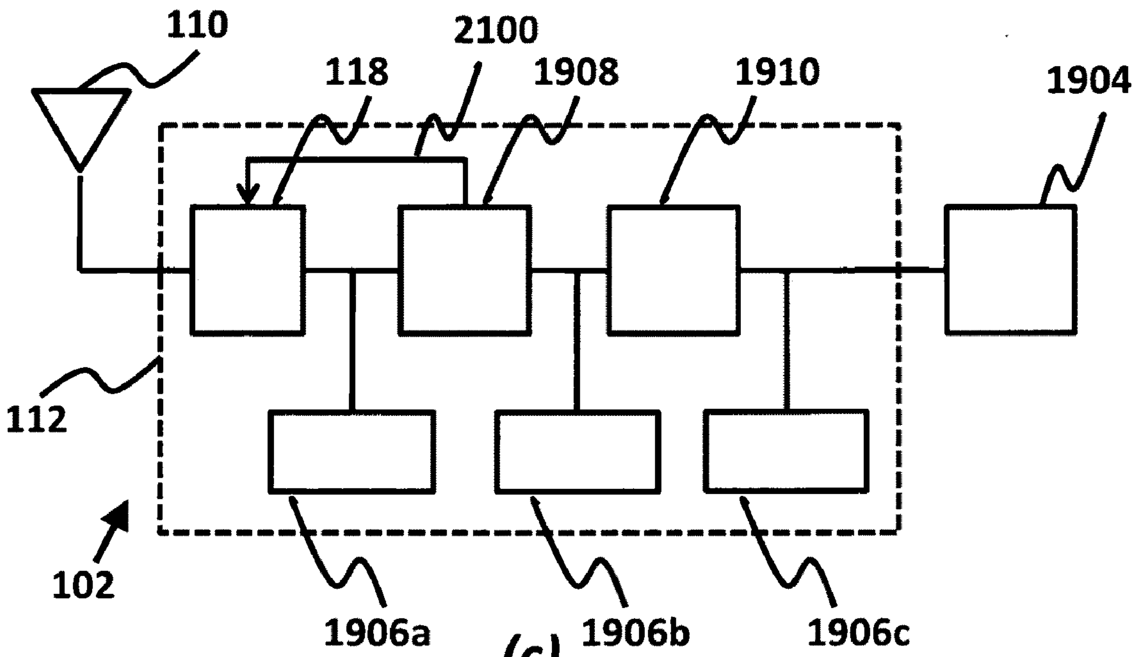
110
2100
118
1908
1910
1904
112
102
1906a
1906b
1906c

(c)

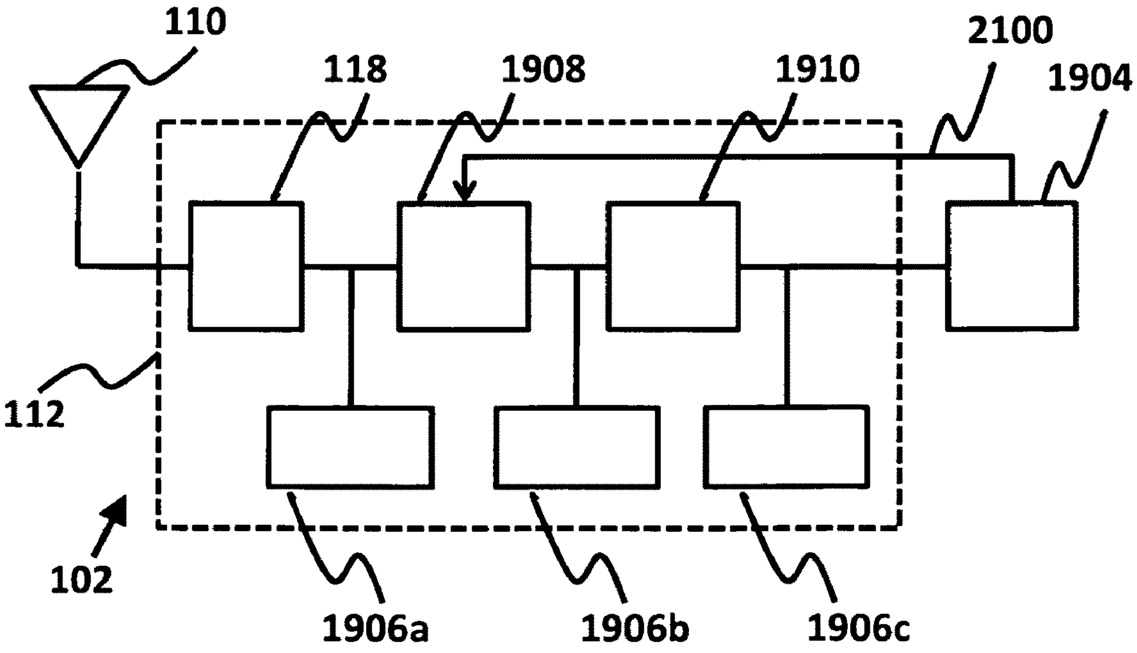
110
118
1908
1910
2100
1904
112
102
1906a
1906b
1906c

(d)

FIG. 19 (cont.)

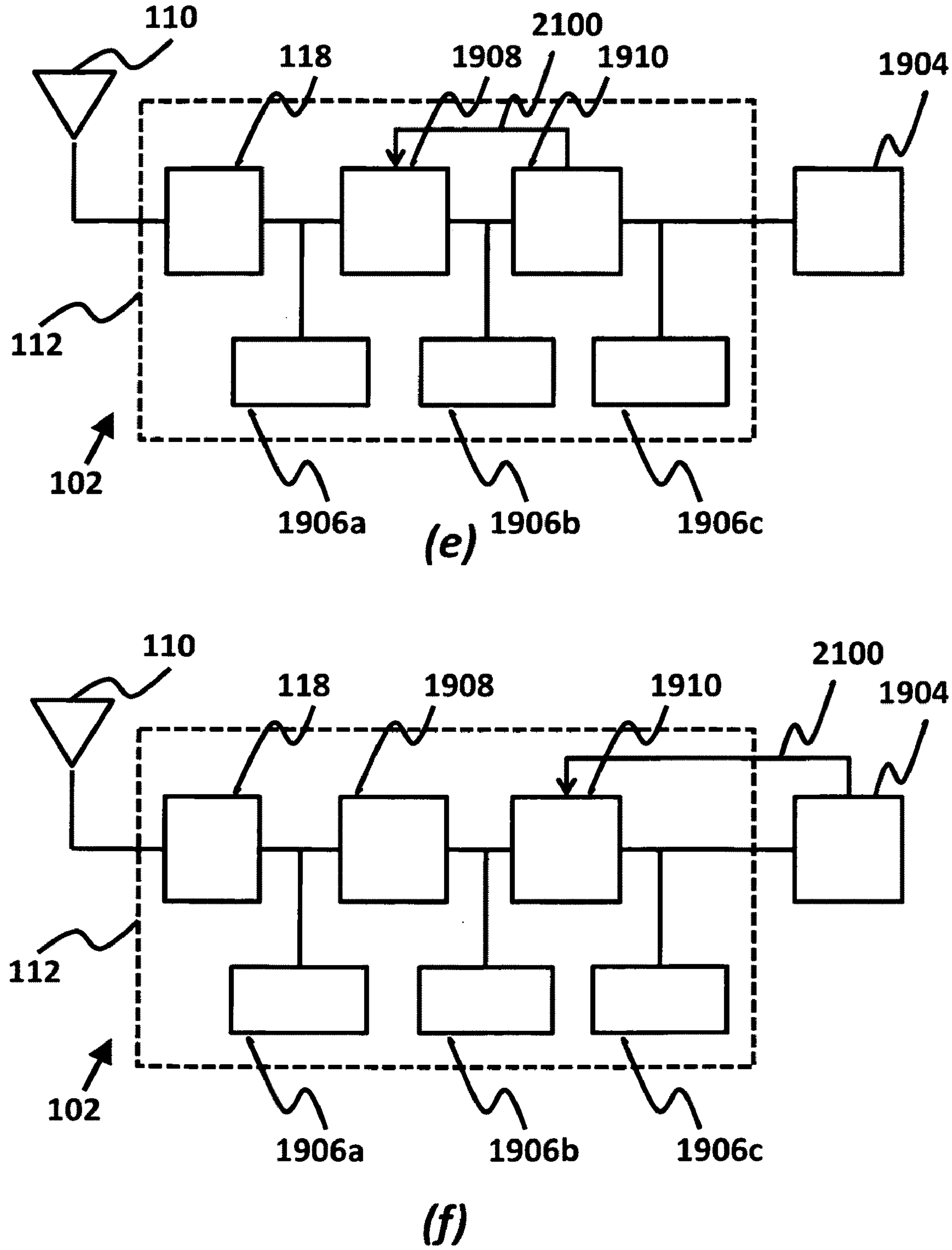
110
118
1908
2100
1910
1904
112
102
1906a
1906b
1906c
110
118
1908
1910
2100
1904
112
102
1906a
1906b
1906c

(e)

(f)

FIG. 19 (cont.)

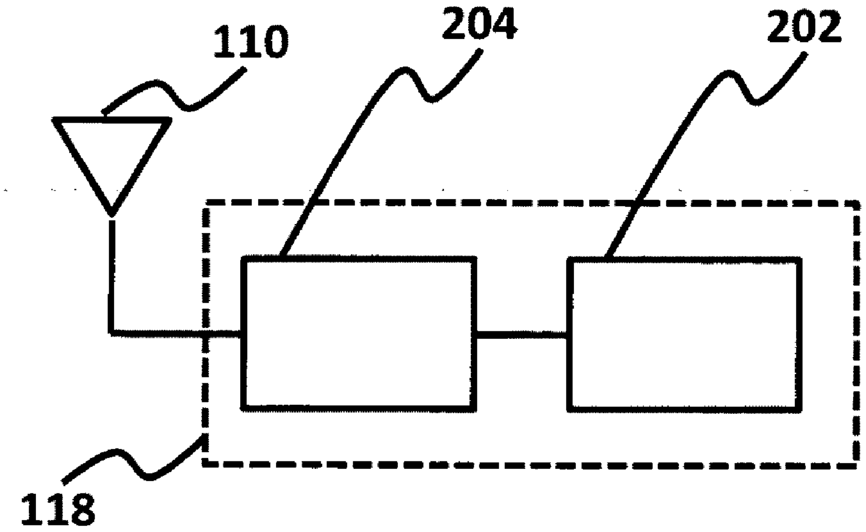
110
204
202
118

FIG. 20

RF ENERGY HARVESTING CIRCUIT

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is cross-referenced to and claims the benefit from U.S. Provisional Patent Application 60/961,009 filed Jul. 17, 2007, and from U.S. Provisional Patent Application 60/993,260 filed Sep. 10, 2007, which are hereby incorporated by reference.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] The present invention was supported in part by grant number DBI-0529223 from the National Science Foundation (NSF). The U.S. Government has certain rights in the invention.

FIELD OF THE INVENTION

[0003] The invention relates generally to energy harvesting. More particularly, the invention relates to the conversion of radio frequency electromagnetic power to electrical power.

BACKGROUND

[0004] Power extraction is an increasingly important technology, especially in applications relating to harvesting power from propagating radio frequency (RF) signals. RF powered devices are often used in applications where battery replacement is impossible, such as structural monitoring where the RF powered devices are embedded into a structure. Other applications for RF powered devices are in telemetry systems to remotely measure and report data back to a central processing unit, and in passive radio frequency identification (RFID) or passive RF tags to replace the bar code as a new form of data collection, where passive RFID tags are typically used in the range of 1-3 meters. Further, many modem biomedical implants are passively powered with radio waves to prolong the lifetime of the implanted device, and to reduce the chances of infection and chemical instability from the use of batteries. These biomedical implants generally operate within close proximity of the base station (typically 1-50 cm) and require robust designed since there is little tolerance for error in implanted devices. RF powered devices are also used in ultra-low power sensor networks in remote areas to eliminate the need for batteries in the sensor system to keep the sensor network free of maintenance, however they may have a backup battery in case the power provided by the RF radiation is insufficient. The applications for these sensor networks normally require an operating distance of 3 to 100 meters. Other applications for RF powered devices include access control, equipment monitoring and even personal identification.

[0005] RF powered devices require a power conversion circuit that can extract enough DC power from the incident electromagnetic waves for the passive device to operate. Far field RF powered devices are known to generally operate from distances of less than 10 meters from the RF source due to the high power loss from RF wave propagation at UHF frequencies. Other known devices achieve sufficient power but provide low output voltage with higher load current making them inadequate for use in passively powered wireless sensor networks.

[0006] For far-field RF energy harvesting applications, the RF energy is extracted from the air at a very low power density, since the propagation energy drops off rapidly as distance from the source is increased. More specifically, in free space, the electric field and power density drop off at the rate of $1/d^2$, where d is the distance from the radiating source. Here, the available power to a receiver in a far-field RF harvesting device decreases by 6 dB for every doubling of distance from the transmitter. Further, with multi-path fading, the power density drops off at a much faster rate than $1/d^2$. It is therefore critical that the power conversion circuit operate at very low receive power to achieve longer operating distance.

[0007] One of the major challenges to achieving this goal is the relatively high voltage requirement of rectifying circuits currently employed. When the available RF power to the receiver is under 100 μW, the available voltage for rectification in the RF to DC conversion system falls below 0.3V, which is much too low to overcome the threshold voltage (V_{th}) of conventional rectifier circuits. Currently, far-field RF energy harvesting circuits known in the art require a start-up circuit.

[0008] Accordingly, there is a need to develop an efficient rectifier circuit as well as improved system level design for RF to DC power conversion. What is needed is a highly efficient passive power conversion circuit for long distance passive sensing in distributed sensor networks. Further, new rectification circuits for far-field RF energy harvesting devices are needed to improve on the minimum power-threshold requirements for the system to operate. To overcome this power-threshold, significantly more efficient circuit and system level design is required over what is known in the art, especially in reducing or eliminating the need for a start-up circuit. Solutions must be found to circumvent or diminish the "dead-zone" in voltage rectification and otherwise reduce the effective threshold voltage in standard CMOS rectifier designs.

SUMMARY OF THE INVENTION

[0009] To address the current needs for a robust and efficient energy harvesting system, a voltage doubler rectifier RF power harvesting system is provided. The system includes at least one power harvesting module, where the power harvesting module has a voltage doubler rectifier structure circuit that includes a DC voltage input, a DC voltage output, a signal input, a first floating gate transistor and a second floating gate transistor, where the floating gate transistors are connected in series between the DC voltage input and the DC voltage output, and a gate control of the floating gate is connected to a drain of the transistor. The voltage double rectifier structure further includes a first capacitor disposed between the input signal and the drain of the first floating gate transistor, and a second capacitor disposed between a ground and the drain of the second floating gate transistor. The voltage double rectifier RF power harvesting system further includes a harvester antenna, where the harvester antenna is disposed to provide a signal to the signal input. Additionally included are a transmitter and a transmitter antenna, where the transmitter provides the signal for output from the transmitter antenna, and a full-wave peak-to-peak voltage of an incoming RF signal from the antenna to the signal input is rectified. The system also includes a powered device, connected to the DC voltage output to utilize the rectified signal.

[0010] In one embodiment of the current invention, the at least one power harvester further includes at least one charge storage device, where the charge storage device can include capacitors, rechargeable batteries, or non-rechargeable batteries, and the charge storage device is disposed between the voltage doubler rectifier structure and the powered device.

[0011] According to another embodiment of the invention, the at least one power harvester further includes a voltage limiting/regulation device, where the voltage limiting/regulating device is disposed between the voltage doubler rectifier structure and the powered device. Here, the embodiment can further include feedback loops where the feedback loops are feedback connections that can include the powered device to the voltage doubler rectifier structure, the powered device to the voltage limiting/regulation device, or the voltage/regulation device to the voltage doubler rectifier structure. This embodiment can also include at least one charge storage device disposed between the powered device and the voltage doubler rectifier.

[0012] According to another embodiment of the invention, the at least one power harvester further includes a power management device, where the power management device is disposed between the voltage doubler rectifier structure and the powered device. Here, the embodiment can further include feedback loops, where the feedback loops are feedback connections that can include the powered device to the voltage doubler rectifier structure, the powered device to the voltage limiting/regulation device, the powered device to the power management device, the power management device to the voltage limiting/regulation device, the power management device to the voltage doubler rectifier structure, or the voltage/regulation device to the voltage doubler rectifier structure. This embodiment can also include at least one charge storage device disposed between the powered device and the voltage doubler rectifier.

[0013] According to another embodiment, the at least one power harvester further can include an impedance matching device disposed between the harvester antenna and the voltage doubler rectifier. In this embodiment the impedance matching is an adjustable impedance matching, where the adjustment is programmable. Further, the adjustable impedance matching can be programmable.

[0014] In a further embodiment of the current invention, the at least one power harvester further includes passive components disposed between the signal input and the harvesting antenna, where a wider band match is provided along a cascade of the power harvesters.

[0015] In yet another embodiment, the floating gate transistor can be a PMOS-PMOScap, a PMOS-NMPScap, an NMOS-NMOScap, an NMOS-PMOScap, an NMOS-MiMcap, or a PMPS-MiMcap.

[0016] In a further embodiment, the harvesting antenna is a meanderline antenna, where the meanderline antenna includes an inner loop disposed between a pair of approximately matching outer loops, where the inner loop has a pair of antenna ports, and an input impedance of the meanderline antenna matches an in impedance of the voltage doubler rectifier. Here, the antenna outer loop can have a perimeter of 35 centimeters. Further, the antenna inner loop can have a perimeter to match the input impedance of the rectifier. Additionally, the antenna matching outer loops are disposed to divide the antenna into two identical parts as a virtual ground, where a fully differential signal is provided at the antenna ports.

[0017] In a further embodiment, the voltage doubler rectifier structure can be disposed in at least a two-stage rectifier cascade. According to one embodiment, the cascade is a 36-stage cascade. According to another embodiment the cascade is a 16 stage cascade.

BRIEF DESCRIPTION OF THE FIGURES

[0018] The objectives and advantages of the present invention will be understood by reading the following detailed description in conjunction with the drawing, in which:

[0019] FIG. **1**(*a*) shows a drawing of communication links between base station (hub) and passively powered sensors in a passively powered sensor network according to the present invention.

[0020] FIG. **1**(*b*) shows a diagram of an RF-DC power conversion system in a passively powered sensor according to the present invention.

[0021] FIG. **2** shows a block diagram of a passive RF-DC conversion circuit showing the equivalent circuit representation for the antenna and rectifier according to the present invention.

[0022] FIG. **3** shows a graph of the effect of impedance mismatch in high Q resonators.

[0023] FIG. **4** shows a diagram of parasitic components that affect performance of the RF-DC power conversion circuit.

[0024] FIG. **5**(*a*) shows a prior art conventional voltage doubler rectifier.

[0025] FIG. **5**(*b*) shows a diagram of a PMOS floating-gate rectifier according to the present invention.

[0026] FIG. **6** shows a diagram of a PMOS implementation of a floating gate transistor according to the present invention.

[0027] FIG. **7** shows a transistor level schematic of PMOS floating-gate rectifier according to the present invention.

[0028] FIG. **8** shows a diagram of a voltage doubler rectifier with N-stages in cascade according to the present invention.

[0029] FIGS. **9**(*a*)-(*d*) show the effect of a number of rectifier stages on (a) maximum voltage gain, (b) output DC voltage, and (c and d) input impedance of rectifier, according to the present invention.

[0030] FIGS. **10**(*a*)-(*b*) show (a) output voltage curves as function of the number of rectifier stages and the transistor width. (b) contour plot of constant output voltage according to the present invention.

[0031] FIG. **11** shows an antenna for an RF-DC converter according to the present invention.

[0032] FIGS. **12**(*a*)-(*d*) show (a) and (b) simulated and measured components of the antenna input impedance, and (c) and (d) measured input impedance of antenna and rectifier circuit according to the present invention.

[0033] FIGS. **13**(*a*)-(*b*) show (a) measured return loss, and (b) voltage gain for 36-stage rectifier design according to the present invention.

[0034] FIGS. **14**(*a*)-(*b*) show die images of (a) 16-stage rectifier circuit, and (b) 36-stage rectifier circuit fabricated in a 0.25 mm CMOS process according to the present invention.

[0035] FIGS. **15**(*a*)-(*d*) show measured output DC voltage (a) as a function of the sinusoidal input, (b) as function of input power, (c) as function of distance, and (d) measured voltage efficiency according to the present invention.

[0036] FIG. **16** shows PCB to establish performance of the RF-DC power converter at the system level according to the present invention.

[0037] FIGS. **17**(*a*)-(*f*) show measured DC output voltage (a) as function of distance and load for the 36-stage rectifier, and (b) as function of received power for the 36-stage rectifier. (c) Measured output power performance as function of distance and load for the 36-stage rectifier. (d) Measured output power as function of received power for the 36-stage rectifier. (e) Measured power conversion efficiency curves as function of distance and load for the 36-stage rectifier. (f) Measured power conversion efficiency curves as function of received power for the 36-stage rectifier according to the present invention.

[0038] FIG. **18** shows a graph of the measured energy extracted per hour as function of distance for the 36-stage rectifier according to the present invention.

[0039] FIG. **19** shows a shows a block diagram of one embodiment of the energy harvesting system according to the present invention.

[0040] FIG. **20** shows a diagram of one embodiment of a power harvester with voltage limiting/regulation, charge storage devices and power management according to the present invention.

[0041] FIGS. **21**(*a*)-(*i*) show diagrams of various feedback loops with a power harvester with voltage limiting/regulation, charge storage devices and power management according to the present invention.

[0042] FIG. **22** shows a block diagram of one embodiment of power harvester according to the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0043] Although the following detailed description contains many specifics for the purposes of illustration, anyone of ordinary skill in the art will readily appreciate that many variations and alterations to the following exemplary details are within the scope of the invention. Accordingly, the following preferred embodiment of the invention is set forth without any loss of generality to, and without imposing limitations upon, the claimed invention.

[0044] The present invention is directed to systems, circuitry and techniques for obtaining, recovering, acquiring and/or harvesting electrical energy from an environment having/including radio frequency (RF) signals (for example, signals in the is a frequency or rate of oscillation within the range of about 3 Hz and about 30 GHz). The RF signals may be periodic or non-periodic. The environment is generally described as being local to the system (for example, within a radius of 40 meters, and more preferably, within a radius of 3 meters). In one embodiment, the system consists of RF power harvester circuitry to obtain, recover, acquire and/or harvest (hereinafter collectively "harvest" or the like (for example, "harvests", "harvested", "harvesting")) the RF electromagnetic waves from the environment. The RF waves may come in the form of a power transmitter or ambience RF wave. In operation, an RF signal is received by the harvester circuitry, which converts the RF signal into a voltage to be used by devices (for example, a substantially DC voltage). In one embodiment, the power harvester circuitry employs floating-gate transistors to provide improved power conversion efficiency of the system.

[0045] An RF-DC power conversion system is provided, which efficiently converts far-field RF energy to DC voltages at very low received power and voltages. Passive rectifier circuits are provided in a 0.25 μm CMOS technology using floating gate transistors as rectifying diodes. A 36-stage rectifier disclosed, according to one embodiment, can rectify input voltages as low as 50 mV with a voltage gain of 6.4 and operate with received power as low as 5.5 μW (−22.6 dBm). Implemented for far field, the circuit operates at a distance of 44 meters from a 4 W EIRP source. The high voltage range achievable at low load current, makes it practical for use in passively powered sensor networks.

[0046] The current invention provides a highly sensitive and efficient rectifier circuit and system level design for RF to DC power conversion. Fully passive rectifier circuits are provided in a 0.25 μm CMOS technology disposed to operate at very low received power. A receive antenna is provided in a 4-layer FR4 board to maximize power transfer in the system.

[0047] The harvested electrical energy can then be used to power one or more electrical devices. After RF signals are converted to DC signals, the invention can power an electrical device directly or it can incorporate other components, such as charge storage devices (e.g. batteries and capacitors), voltage limiters/regulators, power management devices, or any combination thereof. In another embodiment, the system can also incorporate feedbacks between any two or more of the components including the electrical device to be powered, the RF to DC converter, the voltage regulator, and the power management device.

[0048] When RF powered devices harvest their power from RF wave radiation, a radiating source or base station is required to transmit a high intensity RF signal wirelessly through the air. The RF signals can be within the range of about 3 Hz and about 30 GHz and may be periodic or non-periodic. The environment is generally described as being local to the system within a radius range of 3 to 40 meters.

[0049] Referring now to the figures, FIG. **1**(*a*) shows one embodiment of how the system operates as a network **100** where multiple energy harvesting systems **102** receive their energy from the same power source (shown as a hub or base station) **104**, however the RF waves **106** may also come in the form of ambient RF waves **106**. FIG. **1**(*b*) shows a diagram of an RF-DC power conversion system **108** in a passively powered sensor **102** according to one embodiment of the current invention, where the RF-DC power conversion system **108** includes an antenna **110**, connected to an integrated circuit **112**. Here, the integrated circuit **112** includes a receiver **114** and a transmitter **116**, which are connected to the antenna **110** and, optionally, to each other. The integrated circuit further includes an RF-DC converter **118** connected to the antenna **110**. According to this embodiment, the integrated circuit **112** has a holding capacitor **120** connected to the RF-DC converter **116** for receiving and holding the converted power, and providing power back to the integrated circuit **112**. Further shown is a baseband processing block **122** connected to receive a signal from the receiver **114** and connected to provide a signal to the transmitter **116**. Further, the baseband processing block **122** is connected to receive and provide signals with an electrically erasable programmable read-only memory (EEPROM) **124**. Further shown in this embodiment is an ancillary sensor **126** connected to the integrated circuit **112** through an analog to digital converter interface **128**, which may provide an additional signal for further power conversion or other sensing capabilities.

[0050] FIG. **2** shows a diagram of a passive RF-DC conversion circuit **200** showing the equivalent circuit representation for the antenna **110**, rectifier **202** and impedance matching **204**. The RF to DC power conversion system of the current invention operates in UHF frequencies in the industrial, scientific and medical bands (ISM band) of 902-928 MHz. In this frequency range, RF power is transmitted more efficiently for longer distances and experiences lower propagational losses than higher frequency bands (i.e., 2.4 GHz). The system is optimized to operate at distances above 10 meters with the capacity to store charge for long periods of time. Given that the power density drops off at the rate of $1/d^2$ in free space, the propagational RF signal loss through the air at 915 MHz can be calculated to be 51.6 dB with the Friis equation for freespace loss. The maximum transmit power allowed by the FCC in the 902-928 MHz band is 36 dBm, equivalent isotropically radiated power (EIRP) (30 dBm maximum transmit power with 6 dB antenna gain), thus the received power for distances greater than 10 meters is below 27 μW, this translates to less than 75 mV in a 50Ω matched system. The power conversion circuit must be highly sensitive for long-range operation, thus the threshold voltage of the system must be greatly reduced to improve power conversion efficiency at distances greater than 10 meters. It is also essential to come up with design techniques at the system level that can increase the voltage available for rectification to further increase the power conversion efficiency.

[0051] An RF-DC power conversion system is provided to passively amplify the voltage available for rectification by forming a high-Q resonator. The system maximizes the voltage coming into the RF-DC power conversion system **108** so that it can provide a stable DC output voltage at ultra-low receive power. The block diagram **200** of the system **108**, shown in FIG. **2**, consists of an antenna **110** to pick up the power radiated by the RF waves **106**, an impedance matching network **204** to ensure maximum power transfer in the system **108**, and a rectifier circuit **202** to convert the RF signal **106** to a DC voltage. The passive amplification of voltage is done by matching the impedances between the receive antenna **110** and the rectifier circuit **202**. Due to the high-Q nature of the voltage rectification circuit **202** (see circuit schematic in FIG. **5***b*) the impedance matching creates a high-quality factor (Q factor) resonator between the receive antenna **110** and the rectifier circuit **202**.

[0052] A key aspect to improve the efficiency of the RF-DC power conversion at the system level, according to the current invention, is to maximize the input voltage to the rectifier **202**. This is done by forming a resonator (see circuit schematic in FIG. **5***b*) with high-Q factor between the impedances of the receive antenna **110** and the rectifier circuit **202**. By doing this, it passively amplifies the incoming RF signal **106**. To ensure the maximum possible power is transferred to the rectifier circuit **202**, the impedance of the receive antenna **110** is matched to the input impedance of the rectifier circuit **202**. Due to the fact that systems with high-Q resonate with greater amplitude at the resonant frequency than systems with low-Q, the high-Q resonator acts as a passive voltage-amplifier to increase the peak voltage coming into the input of the rectifier without dissipating additional power. The passive voltage gain from the high-Q resonator is directly proportional to its Q. With an increase in the amplitude of the voltage coming to the input of the rectifier **202**, the output voltage of the rectifier **202** also increases and, therefore, increases the overall power conversion efficiency of the system **108**. One drawback of the high-Q resonator is it can reduce the operating bandwidth of the RF-DC power conversion system **108** since

$$Q = \omega \cdot \left(\frac{\text{Energy Stored}}{\text{Average Power Dissipated}} \right) \text{ or } Q = \frac{f_c}{\Delta f} \tag{1}$$

where ω is the resonant frequency in radians/second, f_c is the center frequency of operation, and Δf is the bandwidth of the system. For the far-field RF-DC power conversion system **108** operating in the band 902-928 MHz, the maximum Q that can be attained without sacrificing bandwidth is 35. This limitation on the system Q does not cause much concern since on-chip components rarely have Q of more than 10, and the parasitic resistance from these components damp out the resonator to prevent the Q of the system from limiting the bandwidth. In the case of a series connected matched LC resonator, the reactive components are complex conjugates of each other and the resistive components are matched. The Q of the resonator is therefore

$$Q = \frac{1}{2}\frac{\omega L}{R} = \frac{X_L}{2R} \text{ or } Q = \frac{1}{2}\frac{1}{\omega RC} = \frac{X_C}{2R} \tag{2}$$

where X_L and X_C are the reactive components, and R is the resistive component of the LC resonator. The Q of the resonator is half the Q of the rectifier and antenna since the resistance in the series connected matched resonator is doubled. To achieve a high system Q, it is therefore desirable to increase the reactive components of the rectifier **202** and antenna **110** while reducing resistive components.

[0053] A matching network **204** between the receive antenna **110** and rectifier **202** is necessary to fine tune the impedance match between the antenna **110** and the rectifier **202** to further reduce transmission loss and increase the voltage gain. Coarse impedance matching is done through circuit and antenna design but fine impedance matching must also be done on the PCB for more accurate matching. FIG. **3** shows a graph of the simulated effect **300** of impedance mismatch for a typical high Q resonator. The maximum voltage gain that can be attained is equal to the Q of the resonator, or half that of the rectifier Q. With impedance mismatch greater than 7%, the passive voltage gain from the resonator is reduced to below 3 regardless of the resonator Q. With impedance mismatch greater than 15%, the high-Q matching network provides no voltage gain and even yields a voltage attenuation that becomes greater with increasing Q. The impedance mismatch between the antenna and rectifier must be minimized to obtain a high voltage gain.

[0054] From the top-level diagram of FIG. **2**, the rectifier circuit is modeled by impedance with a real part R_{rect} **206** and a reactive part X_{rect} **208**. From the system point of view, the rectifier circuit **202** must be designed to reduce threshold voltage loss (V_{th}) as much as possible to improve the efficiency of the RF-DC power conversion system **108**. The rectifier circuit (see FIG. **5**(*b*)) must also be provided so that the output voltage can be scaled by cascading multiple rectifier stages in series (see FIG. **8**). Improving the Q of the rectifier input impedance is essential to increasing the power conversion efficiency of the overall system **108**. The rectifier input impedance Q should also be kept as high as possible and parasitic components in the rectifier must be kept as low as possible to maintain a high overall system Q and thus power conversion efficiency.

[0055] The number of cascaded rectifier stages (see FIG. **8**) in the RF-DC conversion system **108** also has a significant effect on the rectifier input impedance. In a conventional voltage rectification circuit design (see FIG. **5**(*a*)) in CMOS technology, the rectifier impedance as seen from the input is capacitive and resistive due to the gate capacitance and the channel resistance (r_{ds}) of the MOS transistor. In general, cascading multiple rectifier stages in series causes capacitive components to increase linearly with the number of stages, while providing parallel paths causes the resistive component to decrease. With a large number of stages, the resistive component from the rectifier is so low that it is dominated by other sources of parasitic resistance (i.e. drain and source connection resistors) and hence, the Q of the system is reduced due to the linearly increased parasitic capacitance.

[0056] Conversely, if there are too few rectifier stages in cascade, the output voltage of the rectifier may not be high enough to operate the sensor node. As the number of rectifier stages increases, the DC output voltage increases until the number of rectifier stages reaches an optimal point. Adding more stages beyond the optimal point reduces the system Q and causes a reduction in the DC output voltage. Thus, it is critical for the number of rectifier stages to be selected through thorough circuit simulation so that the output DC voltage is maximized while maintaining a high system Q to achieve maximum power conversion efficiency.

[0057] The current invention provides a reduction of the unwanted parasitic components that affect system performance. Since the system is to be designed with a high-Q resonator, any additional parasitic components between the antenna and rectifier will greatly diminish the performance of the power conversion system. It is therefore critical to specify an accurate parasitic model for simulating the effect of all parasitic components that affect power conversion efficiency. In the design of the RF-DC power conversion system, the traces connected to inputs of the rectifier are most sensitive to parasitics since a high-Q resonance is required at these inputs. For example, FIG. **4** shows an equivalent circuit model **400** for the parasitic components that affect the power conversion efficiency of the rectifier circuit.

[0058] These parasitic components cannot be avoided altogether, but the value of each parasitic component can be reduced through careful layout and package selection. The bond pad **402** can be designed to be minimum size and only consisting of the top two metal layers to reduce bond pad capacitance to the substrate. The package for the integrated circuit is selected so that pin-parasitics are minimized and critical input pins **404** are placed in locations where the length of bond wires **406** is minimized. PCB traces **408** are made as short as possible and they are impedance controlled to reduce the parasitic capacitance and inductance.

[0059] The antenna **110** structure is critical in the RF-DC power conversion system since it must extract the power radiated by the RF waves **106**. The antenna **110** performs best when it is impedance matched to the rectifier circuit **202** at the operating frequency to reduce transmission loss from PCB traces. Also, the antenna **110** must be small in area and must have a bandwidth large enough to cover the frequency band from 902-928 MHz.

[0060] FIG. **5**(*a*) shows the conventional voltage doubling rectification circuit **500**, and FIG. **5**(*b*) shows a floating-gate rectification circuit **502** according to the current invention. For the floating-gate rectification circuit **502**, floating-gate devices **504** are used to create a gate-source bias to reduce the threshold voltage loss of the MOS transistor **506**.

[0061] The voltage doubler rectifier structure is provided for the RF-DC power conversion system because it rectifies the full-wave peak-to-peak voltage of the incoming RF signal and it can be arranged in cascade to increase the output voltage. The voltage conventional doubler rectifier **500** in FIG. **5**(*a*) consists of a peak rectifier **508** formed by D_1 and C_2 and a voltage clamp **510** formed by D_2 and C_1. The voltage clamp **510** and the peak rectifier **508** are arranged in cascade configuration to provide a passive level shift in voltage before rectification.

[0062] In the negative phase of the input, current flows through diode D_2 while D_1 is cutoff. The voltage across diode D_2 stays constant around its threshold voltage and the voltage at node (1) is charged to $-V_{th2}$. At the negative peak, the voltage across capacitor C_1 is $V_{amp}-V_{th2}$, where V_{amp} is the amplitude of the input signal. In the positive phase of the input, current flows through diode D_1 while D_2 is in cutoff. The voltage across capacitor C_1 remains the same as the previous phase because it has no way to discharge. At the positive peak, the voltage across D_2 is $2V_{amp}-V_{th2}$. Since D_1 is conducting current to charge C_2, the voltage at the output is a threshold voltage below that across D_2, i.e., the voltage at the output V_{out} is $2V_{amp}-V_{th2}-V_{th1}$.

[0063] For the floating-gate rectifier circuit **502** of the current invention, the diodes D_1 and D_2 of the conventional rectifier circuit **500** are replaced by diode-tied floating gate transistors **504**. The floating gate devices **504** are provided to passively reduce the threshold voltage of the rectifier circuit **502**. In a floating gate device **504**, when charge is injected into the floating gate of the transistor **506**, it remains in the gate oxide because of the high impedance provided by the oxide layer. The gate oxide is a very good insulator, which keeps the charge from leaking off in the floating gate **504**.

[0064] To provide a floating gate device **600** in a standard CMOS process, a MOS capacitor **602** is placed in series with the gate **606** of the diode-tied transistor **604** as shown in FIG. **6**. The gate of the diode-tied transistor **606** and the MOS capacitor gate **608** are connected together to form a high impedance node to trap charges in the floating gate **610**. The charge in the floating gate **610** is therefore fixed which results in a fixed voltage bias across the MOS capacitor **602**. The charges that are trapped inside the floating gate device **600** act as a gate-source bias to passively reduce the effective threshold voltage of the transistor **604**.

[0065] The charge on the floating gate **610** can be injected via Fowler-Norheim (F-N) tunneling when the rectifier is not operating, or it can be charged by injecting a relatively large sinusoidal signal to the input of the rectifier at any time. The Fowler-Norheim tunneling technique charges the floating gate **610** to the desired voltage much faster, but the amount of charge is harder to control and also, additional circuitry is needed to inject or remove charge from the floating gate **610**. Referring again to FIG. **5**(*b*), the rectifier circuit **502** of the current invention is provided to have charge injected to the floating gate **504** with a relatively large sinusoidal input with amplitude larger than the threshold voltage of the transistor **506** used for rectification. With the large sinusoidal voltage injection, charge is injected to the floating gate **504** via the parasitic capacitance between the gate-source and gate-drain junction of the transistor **506**. With the floating-gate device **504**, the threshold voltages of the diode-tied transistors M_1 and M_2 are reduced by creating a gate-source bias. The gates of transistors M_1 and M_2 in FIG. **5**(*b*) are high impedance nodes so any charge trapped in the floating gates **504** can be retained for a long time. Retaining charge in floating-gate devices **504** is critical to the useful lifetime for the power conversion circuit in the current invention. With the 70-angstrom oxide thickness in the 0.25 μm CMOS process, the device retains charge in the floating gate **504** in excess of 10 years for normal operation at room temperature. However, the performance of the rectifier circuit **502** may reduce slightly as charge is leaked from the floating gate **504**. During fabrication, the residual charge trapped in the floating gate **504** may also affect the threshold voltage of the rectifier circuit **502**, hence the floating gate **504** must be programmed to account for these residual charges.

[0066] The transistor level schematic of the 36-stage floating-gate rectifier **700** is shown in FIG. **7**. V_{RF} represents the input signal extracted from the RF wave **106** (not shown), V_{DCin} is the input DC voltage coming from the previous rectifier stage (not shown) and V_{DCout} is the output DC voltage of the rectifier **700**. The 36-stage embodiment **700** uses diode-tied PMOS transistors **702** as rectifying devices and a MOS capacitor (MOSCAP) **704** is used to create the gate-source bias for each individual diode-tied transistor **702** in the rectifier **700**. The MOSCAPs **704** (C_{g1} and C_{g2}) generally operate in the depletion region and their capacitance is a function of the applied voltage. To create the gate-source bias, a large sinusoidal signal is applied at the input of the rectifier **700** and charge is injected over time through the parasitic capacitance of the diode-tied transistors **702**. When the gate-source bias reaches a potential close to V_{th}, the capacitance reaches a flat point and remains constant. The amount of charge stored in the floating gate **708** reaches equilibrium and the MOSCAP **704** operates in the inversion region, the source-gate bias voltage is, therefore, approximately the threshold voltage of the diode-tied transistors **702** (M_1 and M_2).

[0067] The individual stages of the floating-gate voltage doubler rectifier circuit can be arranged in cascade to increase the output voltage of the rectifier. FIG. **8** shows N stages of a voltage doubler rectifier in cascade **800**. When the rectifier stages **802** are cascaded, each rectifier stage **802** acts as a passive voltage level shifter in addition to the voltage shift in voltage clamp and peak rectification. The number of rectifier stages **602** used in the design is important since too few rectifier stages yields insufficient output voltage and too many rectifier stages damps out the effect of the high-Q resonator.

[0068] One of the important tradeoffs in the voltage doubler rectifier is the size of the transistor versus parasitic capacitance. The smaller the transistor size, the less parasitic capacitance it has, however, rectification efficiency is lowered by the smaller transistor size since smaller transistor can deliver less current to the load. The transistor sizes can be reduced to a few times the minimum width to reduce parasitic capacitance as seen from the input of the rectifier, however, the reduction in channel width may cause a decrease in the performance of the rectifier due to the increase in the channel resistance of the diode-tied transistors. As an example, two different sample designs are provided to compare the tradeoffs between a reduction in the parasitics and the rectifier performance.

[0069] The first sample design uses a relatively large device size of 12 mm/0.24 μm (NMOS) and the second uses 2 μm/0.24 μm PMOS devices. FIG. **9**(*a*) shows the maximum voltage gain that can be achieved in the rectifier with different numbers of floating-gate rectifier stages arranged in cascade. As the number of rectifier stages increase, the resonator Q between the antenna and rectifier impedance is reduced. In fact, the resonator Q is inversely proportional to the number of rectifier stages.

[0070] The output voltage of the rectifier for both designs with an input voltage with 300 mV amplitude is shown in FIG. **9**(*b*). The output voltage initially increases as more rectifier stages are added until an optimal point then it reduces as the resonator Q is decreased. The impact of the number of rectifier stages on the input impedance is shown in FIGS. **9**(*c*) and **9**(*d*). As the number of rectifier stages increase, the capacitive component in the rectifier input impedance increases thus reducing the reactive component of the rectifier input impedance. The reactive component of the rectifier input impedance is found to be inversely proportional to the number of rectifier stages. With the decrease in the reactive component, the maximum voltage gain that can be achieved at the input is also decreased at the same rate since the resistive component in the input impedance stays fairly constant. In this example, the optimized number of cascaded rectifier stages for the 12 μm (NMOS) and 2 μm (PMOS) designs is 16 and 36 stages, respectively. The simulated input impedance for the two samples is 6.6-j85Ω and 6.6-j84Ω, almost identical to each other.

[0071] To establish the tradeoff between the transistor size used for rectification versus the number of rectifier stages, a model is extracted by curve fitting of the output voltage data for the two different designs. FIG. **10**(*a*) shows a 3-dimensional surface plot for the rectifier output voltage as a function of the number of rectifier stages and the transistor width. It can be observed that a width of 2 μm is the best design choice for the highest output voltage when the number of rectifier stages is below 40. Any increase in device size from 2 μm results in a reduction in the output voltage. Reducing the device size below 2 μm could lead to a slight improvement in the output voltage. However, the number of rectifier stages required is significantly increased which results in a larger die size. FIG. **10**(*b*) shows the constant output voltage contour plot as function of the transistor width and the number of rectifier stages. The 36-stage design with a 2 μm device size is very close to the highest voltage contour, thus, this design is optimal for the output voltage.

[0072] The antenna **110** for the RF-DC conversion circuit **108** is designed with meander lines on a printed circuit board to reduce the area of the antenna **110** and to provide the desired antenna input impedance to the impedance matching network. FIG. **11** shows one embodiment of the antenna with two outer loops **1100** on each side connected to an inner loop **1102** in the center. The antenna ports **1104** are in the middle and connect to the antenna **110** through the inner loop **1102** to transform the impedance as seen from the outer loops **1100** of the antenna **110**. To achieve resonance at the desired frequency, the antenna **110** is provided to match with the input impedance of the floating-gate rectifier circuit **502**, **700**, **800**). The outer loop **1100** of the antenna **110** is designed to tune the antenna **110** to the correct operating frequency and the perimeter of the loop **1100** is 35 centimeters, roughly equal to one wavelength at 916 MHz, according to this embodiment. The perimeter of the inner loop **1102** varies depending on the value of the input impedance of the rectifier (**502**, **700**, **800**). The line of symmetry between the two sides of the antenna **110** splits the antenna **110** into two identical parts and this acts as a virtual ground to create the fully differential signal at the antenna ports **1104**. The dimension of the drawn area of the antenna is 15 cm×2 cm (6×0.8 inches). The antenna **110** is tuned for maximum received power at 916 MHz and the antenna is designed on the top layer of a 4-layer FR4 board with carefully controlled impedance to ensure the impedance of PCB traces does not vary from the intended values. According to one embodiment, the traces **1106** in the antenna **110** are exposed so that they can be tapped to fine tune the operating frequency. The impedance matching network **204**

(see FIG. **2**) can be placed anywhere along the antenna **110** to fine-tune the matching between the antenna **110** and the rectifier circuit (**502**, **700**, **800**). The antenna **110** has a bandwidth of about 40 MHz to maintain a high antenna Q. Due to the fact that the antenna **110** must be impedance matched to the rectifier (**502**, **700**, **800**), the loop configuration provides little antenna gain in the antenna **110** throughout the frequency bandwidth it operates.

[0073] FIG. **12**(*a*) shows the simulated and FIG. **12**(*b*) shows measured input impedance of the antenna **110** around the operating frequency band. The measured resistive component shows less than 1Ω variation from the simulated value and the reactive component also correlates well with the simulated results. The simulated impedance shown in FIG. **12**(*c*) and the measured input impedance shown in FIG. **12**(*d*) corresponds to the components of the receive antenna **110** for the 36-stage rectifier configuration (see FIGS. **7** and **8**). From the measurement data, the measured resistive components of the antenna **110** and rectifier (**502**, **700**, **800**) match well throughout the intended bandwidth for both embodiments. The reactive component of the 36-stage design matches at 910 MHz.

[0074] FIG. **13**(*a*) shows the plot of the return loss (S_{11}) and FIG. **13**(*b*) shows the voltage gain for the 36-stage configuration. The antenna **110** matches so that less than 1% of the transmitted signal is reflected back to the antenna (i.e. S_{11} below −20 dB) in the frequency band of operation 902-928 MHz. The configuration achieves voltage gain above 5 in a 30 MHz bandwidth, however the peak voltage gain is shifted down by approximately 10 MHz from the intended 915 MHz center frequency.

[0075] As an example of the efficacy of the current invention, rectifier devices with 16 NMOS stages and 36 PMOS stages were fabricated in a 0.25 μm 5 metal single-poly CMOS process as shown in FIGS. **14**(*a*) and (*b*), respectively. The active die areas are 350 μm×600 μm and 400 μm×1000 μm, respectively. The majority of the active area is consumed by the MOSCAPs for floating gate devices and the holding capacitors. The 16-stage and 36-stage rectifier layouts are arranged in two rows of 8 stages each, and six rows of 6 stages each, respectively. Both dies have a pair of 1-stage rectifiers included so that the functionality of the NMOS and PMOS diode-tied floating-gate devices can be individually characterized. Each rectifier is packaged in a 0.5 mm pin pitch QFN-32 package to reduce bond wire parasitics and trace lengths on the PCB.

[0076] In order to test the rectifier circuit, a small-amplitude 1 MHz sinusoidal signal is applied directly to the input. The low frequency signal minimizes parasitic effects introduced by the board as well as voltage losses in the PCB traces. FIG. **15**(*a*) shows the measured output DC voltage unloaded and with a 5 MW load for both rectifier designs as a function of the input sinusoidal amplitude, FIG. **15**(*b*). From a straight-line extrapolation of the output voltage curves, the threshold voltage for the 16-stage rectifier is 33 mV compared to 30 mV for the 36-stage rectifier. Since the difference in threshold voltages is very small between the two rectifier designs, the improvement in the output voltage of the 36-stage rectifier is primarily due to the higher number of cascaded rectifier stages used in the design.

[0077] The performance of the rectifier circuit can be specified by the sensitivity of the rectifier or the minimum input power required for voltage rectification. The average input power can be expressed as:

$$P = \frac{QV_{rms}^2}{R} = \frac{QV_{amp}^2}{2R} \tag{3}$$

where Q is the passive voltage gain from the antenna to the input of the rectifier, VMS is the input rms voltage to the rectifier, V_{amp} is the rectifier input amplitude and R is the input impedance of the rectifier circuit. The voltage gain of the 16-stage and 36-stage rectifiers, measured from the return loss of the matching between the rectifier and antenna is 4.8 and 6.4, respectively. Assuming a rectifier input impedance of 7Ω, the equivalent input power is derived and shown in FIG. **15**(*b*). In the case of the 16-stage rectifier, 2V is achieved for 80 mV input, which is equivalent to a power 20.95 μW (−16.8 dBm). Using the Friis formula for free-space wave propagation, this corresponds to an operating distance 23 meters (FIG. **15**(*c*)) assuming a 4 W radiating source. This is the maximum theoretical distance that can be achieved with this design while keeping radiated power and antenna gain under the FCC emission limit. For the 36-stage rectifier circuit, the rectifier performs better with low voltages due to the increased number of voltage doubler rectifier stages and the higher passive voltage gain. The 36-stage rectifier circuit achieves 2V for an input voltage amplitude as low as 50 mV. Using (3), these values are equivalent to 4.7 μW of power (−23.3 dBm) at 2.0 V output. In free-space, this corresponds to 49 meters in distance.

[0078] Defining the voltage efficiency as the percentage of the theoretical voltage achieved, the theoretical output for the voltage doubler architecture is the number of cascaded rectifier stages times the peak-to-peak voltage into the rectifier. In the case of the 16-stage rectifier, the theoretical output voltage is 16 times the input peak-to-peak voltage. For input voltages above 100 mV, the rectifier efficiency is over 80% as shown in FIG. **15**(*d*) and, hence, it is more efficient at higher input voltages. The voltage efficiency for the 36-stage design is higher as compared to the 16-stage rectifier design for low voltages.

[0079] To test the overall performance of the RF-DC rectifier in actual applications, a custom printed circuit board was fabricated that includes the antenna with exposed metal traces for impedance and frequency tuning as shown in FIG. **16**. The chip is directly soldered to the board and physically abuts to the antenna to minimize any stray inductance. The output DC voltage is observed via the SMA connector with an oscilloscope or a digital multimeter (DMM).

[0080] The performance of the RF-DC power conversion system can be measured wirelessly with propagating electromagnetic wave radiation. The output of an HP/Agilent 8665A signal generator is amplified by a Mini Circuits LZY-2 power amplifier to generate signals with 6 W of power in the frequency range of 902-928 MHz. The signal at the output of the power amplifier is then fed to an antenna through several series connected SMA cables. The losses in the cable combined with the return loss between the power amplifier and the transmit antenna are measured to be around 1.7 dB. Thus, the maximum radiated power is about 36 dBm (4 Watts).

[0081] Only the 36-stage rectifier design is wirelessly measured since it has superior performance in terms of received power sensitivity. Before measuring the overall performance, it is necessary to tune the frequency of the resonator circuit. The resonant frequency can be tuned to the 902-928 MHz range by placing a shorted stub tuning on the bare metal trace at the end of the antenna on the PCB. The tuning of the resonant frequency does affect the output voltage of the RF-DC conversion circuit, since the shorted stub tuning slightly modifies the physical shape of the antenna. The decrease in the output voltage is greater when the original resonant frequency is further from the desired (i.e., longer length stub tuning) band and the decrease is smaller for a resonant frequency closer to the desired band. Although the frequency tuning does affect the output voltage, this change in output voltage is much smaller compared to the change in output voltage caused by impedance mismatch. For the 36-stage rectifier design, the resonant frequency is originally 860 MHz so the size of the loop is reduced only slightly to increase the resonant frequency to 906 MHz. For the wireless measurement of the RF-DC power conversion circuit, the center frequency is set at 906 MHz for the 36-stage rectifier design.

[0082] FIGS. **17**(*a*) and (*b*) show the measured output voltage of the 36-stage rectifier as a function of distance and input power for various resistive loads. The 36-stage conversion circuit is capable of outputting DC signal levels of 2V at distances up to 15 meters with a 36 dBm radiating source when it is not loaded as shown in FIG. **17**(*a*). The measured output DC voltage decreases exponentially as distance increases and also decreases as the load resistance is decreased. With a 0.33MΩ load, the conversion circuit operates within 7 meters from the radiating source while providing 1 volt DC. From observation, the operating distance of the rectifier reduces quickly as the load current is increased while at closer distances, the dependence on load current is much less. FIG. **17**(*b*) shows the output voltage curves of the 36-stage design as function of received power calculated from the Friis equation for free-space propagation loss. The slope of the output voltage curves are relatively constant over the full range since the output voltage is inversely proportional to the log of the operating distance. The 36-stage design operates well when the received power is higher than −22 dBm.

[0083] FIG. **17**(*c*) shows the measured output power as a function of distance. The top curve represents the theoretical received power in free-space as calculated by the Friis formula for free-space propagation loss given a 36 dBm radiating source. With a 0.33MΩ resistive load, the current requirement at the load is higher so more current is drawn from the output of the rectifier to drive the load hence the curve shows a steeper slope. The slope of the roll off is directly proportional to the load current at the output of the rectifier. With a 1.32 MΩ load, 1 μW at 10 meters is possible and up to 12.3 meters with the 5.6 MΩ load. FIG. **17**(*d*) shows the output power as function of received power. From this plot, the point where the power curve starts to roll off from the straight (open circuit) line can be seen more clearly. For the 36-stage rectifier design, the output power starts to roll off at about −8 dBm for the 0.33 MΩ load, −13 dBm for the 1.32 MΩ load, and −18 dBm for the 5.6 MΩ load.

[0084] The measured power conversion efficiency versus distance curves are shown in FIGS. **17**(*e*) and (*f*). The power conversion efficiency curve for each individual load is shown as well as the maximum efficiency that can be achieved at a particular distance. For a perfectly matched resonator network, the power available for rectification is half of the received power in the antenna since only half of the power is dissipated in the rectifier. The power conversion efficiency of the RF-DC power conversion system is defined as the ratio between the DC output power and available power for rectification (half of the received power) from the antenna. The maximum efficiency measured is 60% with the 0.33 Mb load. For lower load currents, the output power is higher at longer distances so the power conversion efficiency is also sufficient at longer distances from the RF source. The configuration is most efficient at 3 meters and the power conversion efficiency curves rolls off as the inverse of distance from the RF source. The peak efficiency with the 5.6 MΩ and 1.32 MΩ load is 29% at about 9 meters and 35% at 5 meters, respectively.

[0085] FIG. **18** shows the maximum energy that can be extracted from the RF signal in one hour at various distances from the radiating source. From the logarithmic plot of extracted energy against distance, the extracted energy curve is approximately a straight line. The maximum extractable energy in an hour is 0.8 J at 1 meter, 40 mJ at 5 meters, 7 mJ at 10 meters and about 1 mJ at 15 meters from the source. With one-hour charge time, the RF-DC power conversion circuit acquires and replenishes enough energy to operate a circuit that dissipates 1 mW for one second or 100 mW for 10 ms at distances up to 15 meter. For an increase of every meter in distance, the amount of energy that can be extracted from the RF signal is reduced by a factor of 2.6. This establishes the trade off between power dissipation and operating distance of the integrated circuit powered by the RF-DC power conversion circuit.

[0086] According to one aspect of the present invention, acquiring and/or harvesting electrical energy is done from an environment having RF frequency, or rate of oscillation, within the range of about 3 Hz and about 30 GHz. The RF signals may be periodic or non-periodic. FIG. **19** shows a block diagram of one embodiment of the energy harvesting system **1900**, where the diagram shows how the RF waves **106** may come from a power transmitter **104** having a transmitter antenna **1902** however it is within the scope of the invention that the RF waves may be ambient RF waves, and be received by the energy harvester system **102**. In operation, an RF signal **106** is received by the harvester circuitry **112** (see FIG. **2**), which converts the RF signal **106** into a voltage to be used by devices **1904**, for example, a substantially DC voltage. FIG. **19** is a diagram of an RF power harvesting system **102** that harvests RF signal **104** from ambient RF energy via one or more RF signal sources **104**. The RF signal source **104** provides one or more RF signals **106**, which may be continuous or non-continuous, and/or periodic or non-periodic, and/or wideband/multi-band or narrowband, where the power transmitter **104** can be powered by a wall outlet, a battery or other power sources. The harvesting antenna **110** collects the RF signal(s) **106** and the power harvester **102** converts the RF signal(s) **106** to a DC supply. The DC voltage may be used by various devices **1902**, or stored on charge storage devices **1906**, where the charge storage device may be external or internal to the system **102**. The charge storage device **1906** stores the charge output by the power harvester **102** and, as needed, supplies the charge to the device **1904** for example, when its power dissipation exceed the power harvested by the power harvester **102**. The charge storage **1906** device may also extend the operational usage time of the device by providing extra charge.

[0087] According to one embodiment of the invention, the power harvester **102** can communicate back to the transmitter **104** via radio link or via backscattering of the incident RF wave **106**. The charge storage device may be considered external to the power harvester.

[0088] FIG. **20** is a diagram of an exemplary embodiment of a power harvester **102** with voltage limiting/regulation **1908**, charge storage devices **1906** and power management **1910**. The antenna **110** collects the RF signal **106** (not shown), the RF-DC rectifies **118** the RF signal **106** to a DC voltage that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage device **1906***a* and its output charges a different charge storage device **1906***b*. Power management **1910**, in this embodiment, controls the conduction path between the output of voltage regulator **1908** and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed or instructed. Charge storage **1906***c* holds the excess charge from the power harvester **102** when device is in operation and gives off charge when device is drawing more current than the power harvester circuit **112** can provide. Notably, the power management **1910** may include a switch or circuitry, which controls, monitors and/or analyzes:

[0089] (i) the power harvesting operation of the power harvester and/or components thereof, for example, the power conversion circuitry;

[0090] (ii) the operating characteristics of the power harvester and/or components thereof;

[0091] (iii) the characteristics of the output power of the power harvester, for example current; voltage and temporal characteristics thereof;

[0092] (iv) the storage operation of one or more of the charge storage devices and/or charge supplied thereto (via, for example, the power harvester); and/or

[0093] (v) the characteristics of the output power of one or more of the charge storage devices, for example, current, voltage and temporal characteristics thereof.

[0094] The voltage limiting/regulation **1908** limits the voltage that can be charged to circuits and charge storage devices **1906** to prevent it from breaking down. The voltage regulation suppresses the voltage supply voltage variation to a small range, preferably under 100 mV of variation, over wide range of temperature, supply voltage and process variations.

[0095] The power management circuitry **1910** controls the charge transfer from the power harvester **102** to the device for safe operation. The power management **1910** may turn on the conduction path for charge transfer when power at the device **1904** is needed and turns off the conduction path when power is no longer needed. The power management **1910** may connect the power harvester **102** to the external device **1904** continuously, at a regular interval, or only once, depending on the operational power requirement of the device **1904**.

[0096] FIGS. **21**(*a*)-(*f*) are diagrams of exemplary embodiments of a power harvester **102** with voltage limiting/regulation **1908**, charge storage devices **1906** and power management **1910**. In the embodiment shown in FIG. **21**(*a*), a feedback **2100** between the device **1904** and the RF-DC power conversion **118** gives the device **1904** the control of the power harvester **102**. The feedback loop **2100** can be used to partially or fully enabling and disabling the RF-DC power conversion **118**. It can be used for tuning the RF-DC power conversion **118** for desirable performance and it can also be used to select between different circuit structures in the RF-DC power conversion **118**. Feedback **2100** is preferred when the RF-DC power conversion **118** requires tuning for enhanced and/or optimized matching, or when communication to the transmitter is required. The device **1904** can trigger the impedance matching network in the RF-DC **118** from matched to total mismatch. For example, this exemplary embodiment of the power harvester **102** may enable a communication link to the transmitter via backscattering of the incident RF signal. On a system where the antenna is shared, the feedback can disconnect and isolate the RFDC block from the antenna.

[0097] In the embodiment shown in FIG. **21**(*b*), the feedback **2100** between the power management **1910** and the RF-DC power conversion **118** allows the power management **1910** control of power harvester **102**. The feed back loop **2100** can be used to partially or fully enabling and disabling the RF-DC power conversion **118**. It can be used for tuning the RF-DC power conversion **118** for desirable performance and it can also be used to select between different circuit structures in the RF-DC power conversion **118**. This feedback configuration may be preferred when the RF-DC power conversion **118** requires tuning for enhanced and/or optimized matching that must be controlled by the power management **1910**. This configuration may be used to adjust the output voltage level of the RF-DC power conversion circuit **112**, which in turn varies the voltage at regulator output to flexibly accommodate for the voltage requirement of the device **1904**. On a system where the antenna is shared, the feedback may be employed to disconnect and isolate the RF-DC block from the antenna so it can be used for other purposes. The feedback loop **2100** can also be used for the purpose of triggering the charging the floating gate devices in certain rectifier circuits to enhance power conversion efficiency or minimize or prevent degradation in the performance of the power harvester.

[0098] FIG. **21**(*c*) shows a diagram of a power harvester **102** with voltage limiting/regulation **1908**, charge storage devices **1906** and power management **1910**. The antenna **110** collects the RF signal **104** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage device **1906***a* and its output charges a different charge storage device **1906***b*. Power management **1910** controls the conduction path between the output of voltage regulator **1908** and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed. Charge storage **1906***c* holds the excess charge from the power harvester **102** when device **1904** is in operation and gives off charge when device **1904** is drawing more current than the power harvester circuit **112** (not shown) can provide. The feedback **2100** between the voltage limiting/regulation **1908** and the RF-DC power conversion **118** allows the voltage limiting/regulation **1908** control of the power harvester **102**. The feed back loop **2100** can be used to partially or fully enabling and disabling the RF-DC power conversion **118**. It can be used for tuning the RF-DC power conversion **118** for desirable performance and it can also be used to select between different circuit structures in the RF-DC power conversion **118**. This feedback configuration is preferred when the voltage limiter/regulation **1908** needs to directly adjust the output voltage level of the RF-DC power conversion **118** to prevent the voltage regulator circuit from getting over saturated in voltage or breaking down. The configuration can be use to aid in the tuning of the impedance matching networks in the RF-DC power conversion **118**. It can also be used for the purpose of triggering the charging the floating gate devices in the rectifier circuits to enhance power conversion efficiency or prevent degradation in the performance of the power harvester.

[0099] FIG. **21**(*d*) shows a diagram another embodiment of the power harvester **102** with voltage limiting/regulation **1908**, charge storage devices **1906** and power management **1910**. The antenna **110** collects the RF signal **106** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage device **1906***a* and its output charges a different charge storage device **1906***b*. Power management **1910** controls the conduction path between the output of voltage regulator **1908** and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed. Charge storage **1906***c* holds the excess charge from the power harvester when device is in operation and gives off charge when device **1904** is drawing more current than the power harvester circuit **112** (not shown) can provide. The feedback **2100** between the device **1904** and the voltage limiting/regulation **1908** allows the device **1904** to control the regulated output voltage level of the power harvester **102**. The feed back loop **2100** can be used to partially or fully enabling and disabling the voltage limiter/regulation **1908**. It can be to select between different devices **1904** or reference voltages in the voltage limiter/regulation **1908** to robustly vary the voltage level as seen from the device **1908**. This configuration is preferred when the device does not operate on a wide range power supply and requires the voltage regulation to be very accurate, preferably to within less than 50 mV of the desire operating voltage.

[0100] FIG. **21**(*e*) shows a diagram of a power harvester with voltage limiting/regulation **1908**, charge storage devices **1906** and power management **1910**. The antenna **110** collects the RF signal **106** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage **1906***a* device and its output charges a different charge storage device **1906***b*. Power management **1910** controls the conduction path between the output of voltage regulator **1908** and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed. Charge storage **1906***c* holds the excess charge from the power harvester **102** when device **1904** is in operation and gives off charge when device **1904** is drawing more current than the power harvester circuit **112** (not shown) can provide. The feedback **2100** between the power management **1910** and the voltage limiting/regulation **1908** allows power management circuits to control the regulated output voltage level of the power harvester **102**. The feed back loop **2100** can be used to partially or fully enabling and disabling the voltage limiter/regulation **1908**. It can be to select between different devices or reference voltages in the voltage limiter/regulation block to robustly vary the voltage level as seen from the device.

[0101] FIG. **21**(*f*) shows a diagram of a power harvester with voltage limiting/regulation **1908**, charge storage devices **1906** and power management **1910**. The antenna **110** collects the RF signal **106** (not shown), the RF-DC rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage device **1906***a* and its output charges a different charge storage device **1906***b*. Power management **1910** controls the conduction path between the output of voltage regulator and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed. Charge storage **1906***c* holds the excess charge from the power harvester **102** when device **1904** is in operation and gives off charge when device **1904** is drawing more current than the power harvester circuit can provide. The feedback **2100** between the device **1904** and the power management **1910** allows device **1904** some control of the power management circuits. The feed back loop **2100** can be used to partially or fully enabling and disabling the power management **1910**. This is desirable when the device **1904** operates frequently such that, power can be saved by turning off the power management circuit. The feedback **2100** also allows the control of how much charge is transferred from the power harvester **201** to the device **1904**.

[0102] It should be obvious that the above embodiments can be provided without the power management block **1910**, and without charge storage device **1906***c*, where the antenna **110** collects the RF signal **106** (not shown), the RF-DC rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage **1906***a* device and its output charges a different charge storage device **1906***b*. Without the power management circuit, the device **1904** is directly powered by the power from the voltage limiting/regulation **1908** and the charge stored on charge storage **1906***b*.

[0103] It should be obvious that the feedback **2100** provided in FIGS. **21**(*a*)-(*f*) can be applied in a similar manner without the power management block **1910**, and without charge storage device **1906***c*. The antenna **110** collects the RF signal **106** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage device **1906***a* and its output charges a different charge storage device **1906***b*. Without the power management **1910**, the device **1904** is directly powered by the power from the voltage limiting/regulation **1908** and the charge stored on charge storage **1906***b*. The feedback **2100** between the device **1904** and the RF-DC power conversion **118** gives the device **1904** the control the power harvester **102**. The feed back loop **2100** can be used to partially or fully enabling and disabling the RF-DC power conversion **118**. It can be used for tuning the RF-DC power conversion **118** for desirable performance and it can also be used to select between different circuit structures in the RF-DC power conversion **118**. This feedback **2100** configuration is preferred when the RF-DC power conversion **118** requires tuning for optimized matching or communication to the transmitter **104** (not shown) is required. The device **1904** can trigger the impedance matching network **204** (not shown) in the RF-DC **118** from matched to total mismatch. This enables a communication link to the transmitter **104** via backscattering of the incident RF signal. On a system where the antenna is shared, the feedback can disconnect and isolate the RF-DC block from the antenna.

[0104] In another embodiment provided without the power management block **1910**, and without charge storage device **1906***c* and having feedback **2100**, the antenna **110** collects the RF signal **106** (not shown), the RF-DC **118** rectifies the RF signal **106** to a DC voltage, the voltage limiter/regulator **1908** regulates the DC voltage and its output is available to an external device **1904**. The feedback **2100** between the voltage limiting/regulation **1908** and the RF-DC power conversion **118** allows the voltage limiting/regulation **1908** control of power harvester **102**. The feedback loop **2100** can be used to partially or fully enabling and disabling the RF-DC power conversion **1910**. It can be used for tuning the RF-DC power conversion **118** for desirable performance and it can also be used to select between different circuit structures in the RF-DC power conversion **118**. This feedback **2100** configuration may be preferred when the voltage limiter/regulation **1908** needs to directly adjust the output voltage level of the RF-DC power conversion **118** to prevent the voltage regulator **1908** from getting over saturated in voltage or breaking down. The configuration can be use to aid in the tuning of the impedance matching networks in the RF-DC power conversion **118**. It can also be used for the purpose of triggering the charging of the floating gate devices in the rectifier circuits to enhance power conversion efficiency or prevent degradation in the performance of the power harvester.

[0105] In another embodiment provided without the power management block **1910**, and without charge storage device **1906***c* and having feedback **2100**, the antenna **110** collects the RF signal **106** (not shown), the RF-DC **118** rectifies the RF signal **106** to a DC voltage, that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage device **1906***a* and its output charges a different charge storage device **1906***b*. Without the power management circuit, the device is directly powered by the power from the voltage limiting/regulation **1908** and the charge stored on charge storage **1906***b*. The feedback between the voltage limiting/regulation **1908** and the RF-DC power conversion **118** allows the voltage limiting/regulation **1908** control of power harvester **102**. The feed back loop **2100** can be used to partially or fully enabling and disabling the RF-DC power conversion **118**. It can be used for tuning the RF-DC power conversion **118** for desirable performance and it can also be used to select between different circuit structures in the RF-DC power conversion **118**. This feedback **2100** configuration is preferred when the voltage limiter/regulation **1908** needs to directly adjust the output voltage level of the RF-DC power conversion **118** to prevent the voltage regulator circuit **1908** from getting over saturated in voltage or breaking down. The configuration can be use to aid in the tuning of the impedance matching networks in the RF-DC power conversion **118**. It can also be used for the purpose of triggering the charging of the floating gate devices in the rectifier circuits to enhance power conversion efficiency or prevent degradation in the performance of the power harvester.

[0106] In another embodiment provided without the power management block **1910**, and without charge storage device **1906***c* and having feedback **2100**, the antenna **110** collects the RF signal **106** (not shown), the RF-DC **118** rectifies the RF signal **106** to a DC voltage, that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage device **1906***a* and its output charges a different charge storage device **1906***b*. Without the power management circuit, the device **1904** is directly powered by the power from the voltage limiting/regulation circuit **1908** and the charge stored on charge storage **1906***b*. The feedback **2100** between the device **1904** and the voltage limiting/regulation **1908** allows the device **1904** to control the regulated output voltage level of the power harvester **102**. The feed back loop **2100** can be used to partially or fully enabling and disabling the voltage limiter/regulation **1908**. It can select between different devices **1904** or reference voltages in the voltage limiter/regulation **1908** to robustly vary the voltage level as seen from the device **1904**. This configuration is preferred when the device **1904** does not operate on a wide range power supply and requires the voltage regulation to be very accurate, preferably to within less than 50 mV of the desire operating voltage.

[0107] It should be obvious that the above embodiments can be provided without the power without charge storage device **1906***c*. Charge storage after the power management block **1910** is not necessary when the conduction path between the power harvester output (not shown) and device **1904** is strongly on, and sufficient power can be delivered through this conduction path. The antenna **110** collects the RF signal **104** (not shown), the RF-DC **188** rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage device **1906***a* and its output charges a different charge storage device **1906***b*. Power management **1910** controls the conduction path between the output of voltage regulator **1908** and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed.

[0108] Charge storing **1906***b* after the voltage limiting/regulation **1908** is not necessary when the power management **1910** is turned on often, thus charging charge **1906***b*. The antenna **110** collects the RF signal, the RF-DC rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage device **1906***a* and its output is feeding to the power management **1910**. Power management **1910** controls the conduction path between the output of voltage regulator **1908** and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed. Charge storage **1906***c* holds the excess charge from the power harvester **102** when device **1904** is in operation and gives off charge when device **1904** is drawing more current than the power harvester **102** can provide.

[0109] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks and without charge storage **1096***b* after regulator **1908** are possible.

[0110] Charge storage **1906***a* is not needed after the RF-DC power conversion **118** when the power received by the RF-DC power conversion **118** is stable or when the power requirement of the voltage limiting/regulation **1908** does not exceed the power harvested up to RF-DC conversion. The antenna **110** collects the RF signal **104** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage and is fed directly to the voltage limiter/regulation block **1908**, the voltage limiter/regulator **1908** regulates the DC voltage from the output of RF-DC **118** and its output charges a charge storage device **1908***b*. Power management controls **1910** the conduction path between the output of voltage regulator **1908** and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed. Charge storage **1906***c* holds the excess charge from the power harvester when device **1904** is in operation and gives off charge when device **1904** is drawing more current than the power harvester circuit **102** can provide.

[0111] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks and without charge storage **1096***a* after regulator **1908** are possible.

[0112] In cases where the device **1904** operates at peak power that does not exceed harvested power and/or the regulated power, charge storage device **1906***b* after the voltage limiter/regulator **1908** and charge storage device **1906***c* after regulation and power management **1910** are not require as part of the power harvester **102**. The antenna **110** collects the RF signal **104** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*, the voltage regulator **1908** regulates the DC voltage on charge storage **1906***a*. Power management **1910** controls the conduction path between the output of voltage regulator to device and turn on the conduction switch when power is needed at the device **1904**.

[0113] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks and without charge storage **1096***b* after regulator **1908** and **1906***c* after power management **1910** are possible.

[0114] In another embodiment, the charge storage devices **1906***a* and **1906***c* are not implemented. The antenna **110** collects the RF signal **104** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage that is then applied to the voltage limiting/regulation **1908**. The voltage limiter/regulator **1908** regulates the DC voltage from the RF-DC power conversion **118** and its output charges charge storage device **1906***b*. Power management **1910** controls the conduction path between the output of voltage regulator **1908** and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed.

[0115] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks and without charge storage **1096***a* after RF converter **118** and **1906***c* after power management **1910** are possible.

[0116] In another embodiment, the charge storage devices **1906***a* and **1906***b* are not implemented. The antenna **110** collects the RF signal **104** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage that is then applied to the input of the voltage limiter/regulation **1908**. The voltage limiter/regulator **1908** regulates the DC voltage from the output of the RF-DC power conversion circuit **118** and its output is fed to the power management **1910**. Power management **1910** controls the conduction path between the output of voltage regulator **1908** and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed. Charge storage **1906***c* holds the excess charge from the power harvester **210** when device **1904** is in operation and gives off charge when device **1904** is drawing more current than the power harvester circuit **102** can provide.

[0117] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks and without charge storage **1096***a* after RF converter **118** and **1906***b* after voltage limiter/regulator **1908** are possible.

[0118] In another embodiment, the charge storage devices **1906***a*, **1906***b* and **1906***c* are not implemented. In cases that device **1904** operates once within a time interval and/or at power level lower than or equal to the harvested power from the RF environment, charge storage devices **1906***a* **1906***b* and **1906***c* are not require as part of the power harvester **102**. The antenna **101** collects the RF signal **104** (not shown), the RF-DC rectifies the RF signal to a DC voltage, the voltage regulator regulates the DC voltage from the output of RF-DC. Power management **1910** controls the conduction path between the output of voltage regulator **1908** to device **1904** and turn on the conduction switch when power is needed at the device **1904**.

[0119] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks and without charge storage **1096***a* after RF converter **118**, **1906***b* after voltage limiter/regulator **1908**, and **1906***c* after power manager **1910** are possible.

[0120] In another embodiment, the charge storage devices **1906***b* and **1906***c* are not implemented, and power management **1910** is not implemented. In cases where the device **1904** operates at peak power that does not exceed harvested power and/or the regulated power, power management **19019** and charge storage device **1906***c* and charge storage device **1906***b* after voltage regulation **1908** is not require as part of the power harvester **102**. The antenna **110** collects the RF signal **104** (not shown), the RF-DC rectifies the RF signal to a DC voltage that is stored on a charge storage device (charge storage **1**), the voltage regulator regulates the DC voltage on **1906***a*. The voltage regulator supplies **1908** the power directly to the device **1904**.

[0121] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks **2100** and without the charge storage devices **1906***b* and **1906***c*, and power management **1910** are possible.

[0122] In another embodiment, the charge storage devices **1906***a* and **1906***c* are not implemented, and power management **1910** is not implemented. In cases that device **1904** operates once within a time interval and/or at power level lower than or equal to the harvested power from the RF environment, power management **1910** and charge storage **1906***c* after power management **1910**, and charge storage **1906***a* after RF-DC **118** are not require as part of the power harvester **102**. The antenna **110** collects the RF signal **104** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage, the voltage regulator **1908** regulates the DC voltage from the output of RF-DC **118**, that is provided to device **1904**.

[0123] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks **2100** and without the charge storage devices **1906***a* and **1906***c*, and power management **1910** are possible.

[0124] In another embodiment, the charge storage devices **1906***a*, **1906***b* and **1906***c* are not implemented, and power management **1910** is not implemented. In cases that device **1904** operates once within a time interval and/or at power level lower than or equal to the harvested power from the RF environment, charge storage devices **1906***a*, **1906***b* and **1906***c* and power management **1910** is not require as part of the power harvester **102**. The antenna **110** collects the RF signal **104** (not shown), the RF-DC rectifies the RF signal to a DC voltage, the voltage regulator regulates the DC voltage from the output of RF-DC and provides it to device **1904** and turns on a conduction switch (not shown) when power is needed at the device **1904**.

[0125] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks **2100** and without the charge storage devices **1906***a*, **1906***b* and **1906***c*, and power management **1910** are possible.

[0126] In another embodiment, the charge storage device **1906***b* and voltage limiter/regulator **1908** are not implemented. In cases that device **1904** operates once within a time interval and/or at power level lower than or equal to the harvested power from the RF environment, charge storage device **1906***b* is not require as part of the power harvester. The antenna collects the RF signal, the RF-DC rectifies the RF signal to a DC voltage, the voltage is provided to power manager **1910** which can store needed power in storage device **1906***c* for powering device **1904**. The power management **1910** also supplies the power directly to the device **1904**.

[0127] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks **2100** and without the charge storage devices **1906***b* after voltage limiter/regulator **1908** and voltage limiter/regulator **1908** are possible.

[0128] In another embodiment, the charge storage devices **1906***b* and **1906***c*, and voltage limiter/regulator **1908** are not implemented. In cases where the device **1904** is able to tolerate wide power supply ranges, charge storage devices **1906***b* and **1906***c*, and voltage limiter/regulator **1908** are not require as part of the power harvester **102**. The antenna **110** collects the RF signal **104** (not shown), the RF-DC rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*. Power management **1910** controls the conduction path between the output of RF-DC **118** to device **1904** and turn on a conduction switch (not shown) when power is needed at the device **1904**.

[0129] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks **2100** and without the charge storage devices **1906***b* after voltage limiter/regulator **1908** and charge storage devices **1906***b* after power manager **1910**, and voltage limiter/regulator **1908** are possible.

[0130] In another embodiment, the charge storage devices **1906***a* and **1906***b*, and voltage limiter/regulator **1908** are not implemented. In cases where the device **1904** is able to tolerate wide power supply ranges, charge storage devices **1906***a* and **1906***b*, and voltage regulator **1908** are not require as part of the power harvester **102**. The antenna **110** collects the RF signal **104** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***c*. Power management **1910** controls the conduction path between the output of RF-DC **118** to device **1908** and turn on a conduction switch (not shown) when power is needed at the device **1904**.

[0131] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks **2100** and without the charge storage devices **1906***a* and **1906***b*, and voltage limiter/regulator **1908** are possible.

[0132] In another embodiment, the charge storage devices **1906***a*, **1906***b* and **1906***c*, and voltage limiter/regulator **1908** are not implemented. In cases where the device **1904** is able to tolerate wide power supply ranges, with peak power less than the harvestable energy, the charge storage devices **1906***a*, **1906***b* and **1906***c*, and voltage limiter/regulator **1908** are not require as part of the power harvester **102**. The antenna **110** collects the RF signal **104** (not shown), the RF-DC rectifies the RF signal to a DC. In this embodiment, power management **1910** may control the conduction path between the output of the RF-DC **118** to external device **1904**. The power management **1910** may responsively "turn on" the conduction when power is needed at the device **1904** or as instructed (for example, by the external device and/or other device).

[0133] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks **2100** and without charge storage devices **1906***a*, **1906***b* and **1906***c*, and voltage limiter/regulator **1908** are possible.

[0134] In another embodiment, the charge storage devices **1906***b* and **1906***c*, and voltage limiter/regulator **1908** and power management **1910** are not implemented. In cases where the device **1904** is able to tolerate wide power supply range, the charge storage devices **1906***b* and **1906***c*, and voltage limiter/regulator **1908** and power management **1910** are not require as part of the power harvester **102**. The antenna **110** collects the RF signal **104** (not shown), the RF-DC rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*. The device **1904** is powered by the charge stored on charge storage **1906***a*.

[0135] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks **2100** and without the charge storage devices **1906***b* and **1906***c*, and voltage limiter/regulator **1908** and power management **1910** are possible.

[0136] In another embodiment, the charge storage devices **1906***a*, **1906***b* and **1906***c*, and voltage limiter/regulator **1908** and power management **1910** are not implemented. This embodiment can be advantageous where the external device **1904** is able to tolerate wide power supply range a, with peak power less than the harvestable energy, the charge storage devices **1906***a*, **1906***b* and **1906***c*, and voltage limiter/regulator **1908** and power management **1910** are not require as part of the power harvester **102**. In this embodiment, the antenna **110** collects the RF signal **104** (not shown), the RF-DC rectifies the RF signal to a DC. The device **1904** draws power directly from the output of RF-DC power conversion circuitry **118**.

[0137] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks **2100** and without the charge storage devices **1906***a*, **1906***b* and **1906***c*, and voltage limiter/regulator **1908** and power management **1910** are possible.

[0138] FIG. **22** shows a block diagram of one embodiment of power harvester, according to the current invention, where shown is the antenna **110** connected to the impedance match **204**, which is connected to the rectifier **202** of the RF-DC converter **118**. In cases where the device **1904** (not shown) is able to tolerate wide power supply range, with peak power less than the harvestable energy, the charge storage devices **1906***a*, **1906***b* and **1906***c*, and voltage limiter/regulator **1908** and power management **1910** not require as part of the power harvester. The RF signal **104** (not shown) is collected at the antenna **110** and passes through impedance matching networks **204** before reaching the rectifier **202**. In this embodiment, the impedance matching circuitry **204** may provide an enhanced and/or predetermined impedance match between the output of the antenna **110** and the input of the RF-DC converter **118**. For example, the impedance matching circuitry may be high-Q to passively increase the voltage amplitude to the rectifier or it can be low-Q to increase the bandwidth of harvestable RF signal. In certain embodiment, it may be preferred that the impedance be matched to within 5% of the ideal impedance value to achieve higher passive voltage gain from high-Q matching and higher power from low-Q matching.

[0139] The present invention has now been described in accordance with several exemplary embodiments, which are intended to be illustrative in all aspects, rather than restrictive. Thus, the present invention is capable of many variations in detailed implementation, which may be derived from the description contained herein by a person of ordinary skill in the art.

[0140] All such variations are considered to be within the scope and spirit of the present invention as defined by the following claims and their legal equivalents.

What is claimed:

1. A voltage doubler rectifier RF power harvesting system comprising:
- a. at least one power harvesting module, wherein said power harvesting module comprises:
  - i. a voltage doubler rectifier structure circuit, wherein said voltage doubler rectifier structure comprises:
    - 1. a DC voltage input;
    - 2. a DC voltage output;
    - 3. a signal input;
    - 4. a first floating gate transistor and a second floating gate transistor, wherein said floating gate transistors are connected in series between said DC voltage input and said DC voltage output, wherein a gate control of said floating gate is connected to a drain of said transistor;
    - 5. a first capacitor disposed between said input signal and said drain of said first floating gate transistor; and
    - 6. a second capacitor disposed between a ground and said drain of said second floating gate transistor;
- b. a harvester antenna, wherein said harvester antenna is disposed to provide a signal to said signal input;
- c. a transmitter;
- d. a transmitter antenna, wherein said transmitter provides said signal for output from said transmitter antenna, wherein a full-wave peak-to-peak voltage of an incoming RF signal from said transmitter antenna to said signal input is rectified; and
- e. a powered device, wherein said powered device is connected to said DC voltage output to utilize said rectified signal.

2. The RF power harvesting system of claim **1**, wherein said at least one power harvesting module further comprises at least one charge storage device, wherein said charge storage device is selected from a group consisting of capacitors, rechargeable batteries, and non-rechargeable batteries.

3. The RF power harvesting system of claim **1**, wherein said at least one power harvester further comprises a voltage limiting/regulation device, wherein said voltage limiting/regulation device is disposed between said voltage double rectifier structure and said powered device.

4. The RF power harvesting system of claim **3**, wherein said at least one power harvesting module further comprises at least one charge storage device, wherein said charge storage device is selected from a group consisting of capacitors, rechargeable batteries, and non-rechargeable batteries.

5. The RF power harvesting system of claim **3**, further comprises feedback loops wherein said feedback loops are feedback connections selected from a group consisting of said powered device to said voltage doubler rectifier structure, said powered device to said voltage limiting/regulation device, and said voltage doubler rectifier structure to said voltage limiting/regulation device.

6. The RF power harvesting system of claim **5**, wherein said at least one power harvesting module further comprises at least one charge storage device, wherein said charge storage device is selected from a group consisting of capacitors, rechargeable batteries, and non-rechargeable batteries.

7. The RF power harvesting system of claim **1**, wherein said at least one power harvesting module further comprises a power management device, wherein said power management device is disposed between said voltage doubler rectifier structure and said powered device.

8. The RF power harvesting system of claim **7**, wherein said at least one power harvesting module further comprises at least one charge storage device, wherein said charge storage device is selected from a group consisting of capacitors, rechargeable batteries, and non-rechargeable batteries.

9. The RF power harvesting system of claim **7**, wherein said at least one power harvesting module further comprises feedback loops wherein said feedback loops are feedback connections selected from a group consisting of said powered device to said voltage doubler rectifier structure, said powered device to said power management device, and said power management device to said voltage doubler rectifier structure.

10. The RF power harvesting system of claim **7**, wherein said at least one power harvesting module further comprises at least one charge storage device, wherein said charge storage device is selected from a group consisting of capacitors, rechargeable batteries, and non-rechargeable batteries.

11. The RF power harvesting system of claim **1**, wherein said at least one power harvesting module further comprises a power management device and a voltage limiting/regulation device, wherein said power management device and said voltage limiting/regulation device are disposed between said voltage doubler rectifier structure and said powered device.

12. The RF power harvesting system of claim **11**, wherein said at least one power harvesting module further comprises at least one charge storage device, wherein said charge storage device is selected from a group consisting of capacitors, rechargeable batteries, and non-rechargeable batteries.

13. The RF power harvesting system of claim **10** further comprises feedback loops wherein said feedback loops are feedback connections selected from a group consisting of said powered device to said voltage doubler rectifier structure, said powered device to said voltage limiting/regulation device, said powered device to said power management device, said power management device to said voltage limiting/regulation device, said power management device to said voltage doubler rectifier structure, and said voltage/regulation device to said voltage doubler rectifier structure.

14. The RF power harvesting system of claim **11**, wherein said at least one power harvesting module further comprises at least one charge storage device, wherein said charge storage device is selected from a group consisting of capacitors, rechargeable batteries, and non-rechargeable batteries.

15. The RF power harvesting system of claim **1**, wherein said at least one power harvester further comprises an impedance matching device disposed between said harvester antenna and said voltage doubler rectifier.

16. The RF power harvesting system of claim **15**, wherein said impedance matching is an adjustable impedance matching, wherein said adjustment is programmable.

17. The RF power harvesting system of claim **16**, wherein said adjustable impedance matching is programmable.

18. The RF power harvesting system of claim **1**, wherein said at least one power harvester further comprises passive components disposed between said signal input and said harvesting antenna, whereby a wider band match is provided along a cascade of said power harvesters.

19. The RF power harvesting system of claim **1**, wherein said floating gate transistor is selected from a group consisting of a PMOS-PMOScap, a PMOS-NMPScap, an NMOS-NMOScap, an NMOS-PMOScap, an NMOS-MiMcap, and a PMPS-MiMcap.

20. The RF power harvesting system of claim **1**, wherein said harvesting antenna is a meanderline antenna, wherein said meanderline antenna comprises an inner loop disposed between a pair of approximately matching outer loops, wherein said inner loop comprises a pair of antenna ports, wherein an input impedance of said meanderline antenna matches an in impedance of said voltage doubler rectifier.

21. The RF power harvesting system of claim **20** wherein said antenna outer loop comprises a perimeter of 35 centimeters.

22. The RF power harvesting system of claim **22** wherein said antenna inner loop comprises a perimeter to match said input impedance of said rectifier.

23. The RF power harvesting system of claim **23** wherein said antenna matching outer loops are disposed to divide said antenna into two identical parts as a virtual ground, wherein a fully differential signal is provided at said antenna ports.

24. The RF power harvesting system of claim **1**, wherein said voltage doubler rectifier structure is disposed in at least a two-stage rectifier cascade.

25. The RF power harvesting system of claim **24**, wherein said cascade is a 36-stage cascade.

26. The RF power harvesting system of claim **24**, wherein said cascade is a 16-stage cascade.

* * * * *

US 20100090656A1

(19) **United States**
(12) **Patent Application Publication** (10) **Pub. No.: US 2010/0090656 A1**
Shearer et al. (43) **Pub. Date: Apr. 15, 2010**

(54) **POWERING DEVICES USING RF ENERGY HARVESTING**

(76) Inventors: **John G. Shearer**, Ligonier, PA (US); **Charles E. Greene**, Pittsburgh, PA (US); **Daniel W. Harrist**, Carnegie, PA (US)

Correspondence Address:
COOLEY GODWARD KRONISH LLP
ATTN: Patent Group
Suite 1100, 777 - 6th Street, NW
WASHINGTON, DC 20001 (US)

(21) Appl. No.: **12/499,618**

(22) Filed: **Jul. 8, 2009**

Related U.S. Application Data

(62) Division of application No. 11/447,412, filed on Jun. 6, 2006, now abandoned.

(60) Provisional application No. 60/688,587, filed on Jun. 8, 2005.

Publication Classification

(51) **Int. Cl.**
H02J 7/04 (2006.01)
H02J 7/00 (2006.01)
G05F 3/00 (2006.01)

(52) **U.S. Cl.** **320/139**; 320/137; 307/151

(57) **ABSTRACT**

Disclosed is an apparatus for an application including a core device for the application. The apparatus includes a power (preferably RF energy) harvester connected to the core device to power the core device. Also disclosed is a method for an application. The method includes the steps of converting RF energy into usable energy. There is the step of powering the core device with the usable energy.

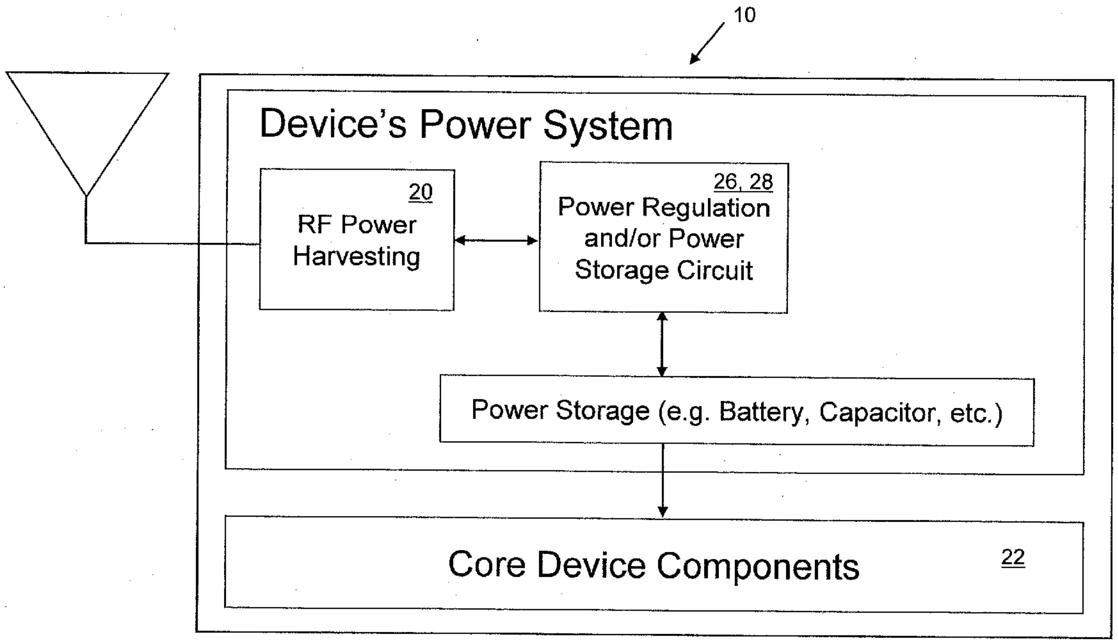
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

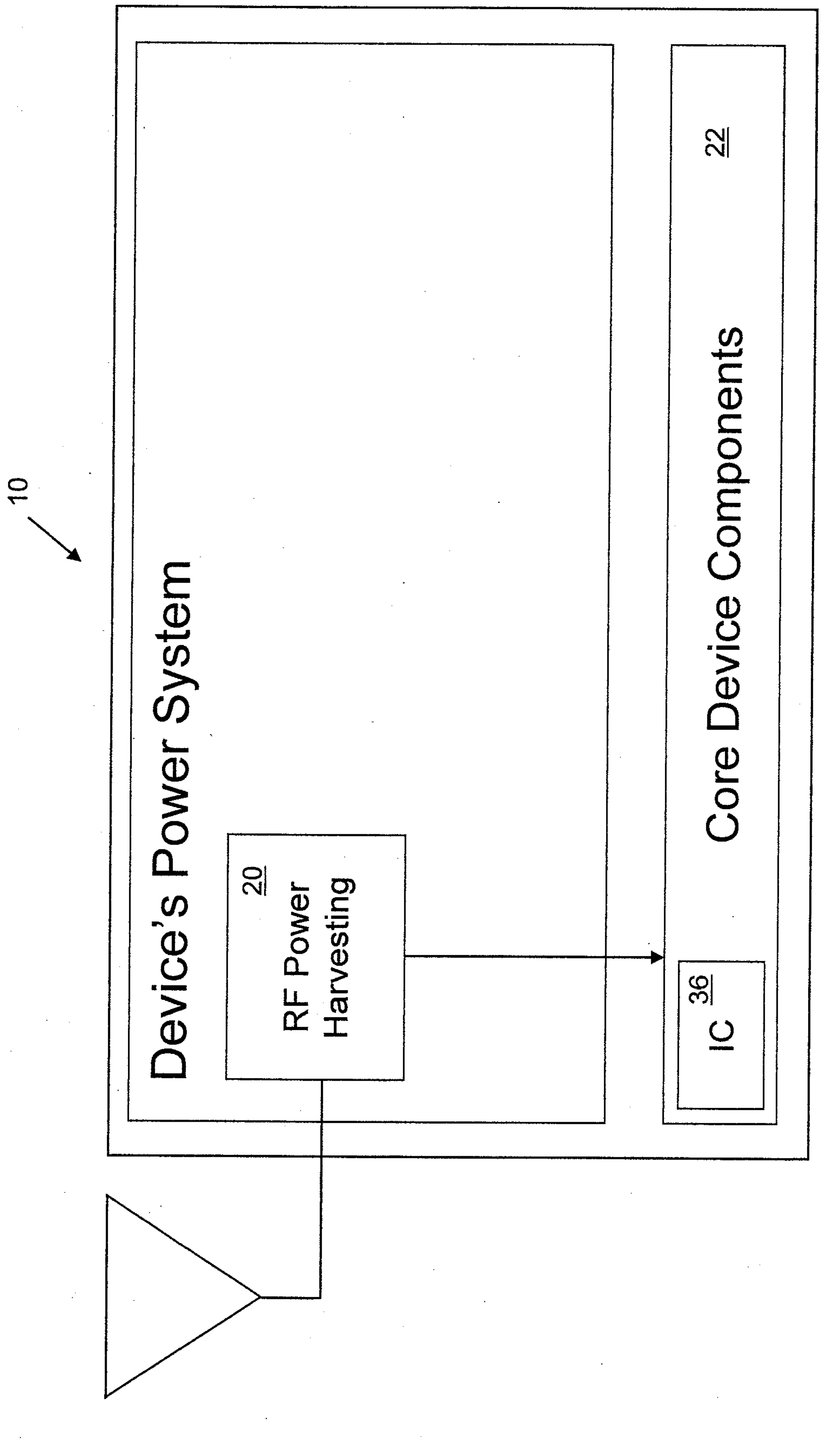
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
IC 36

Fig. 1

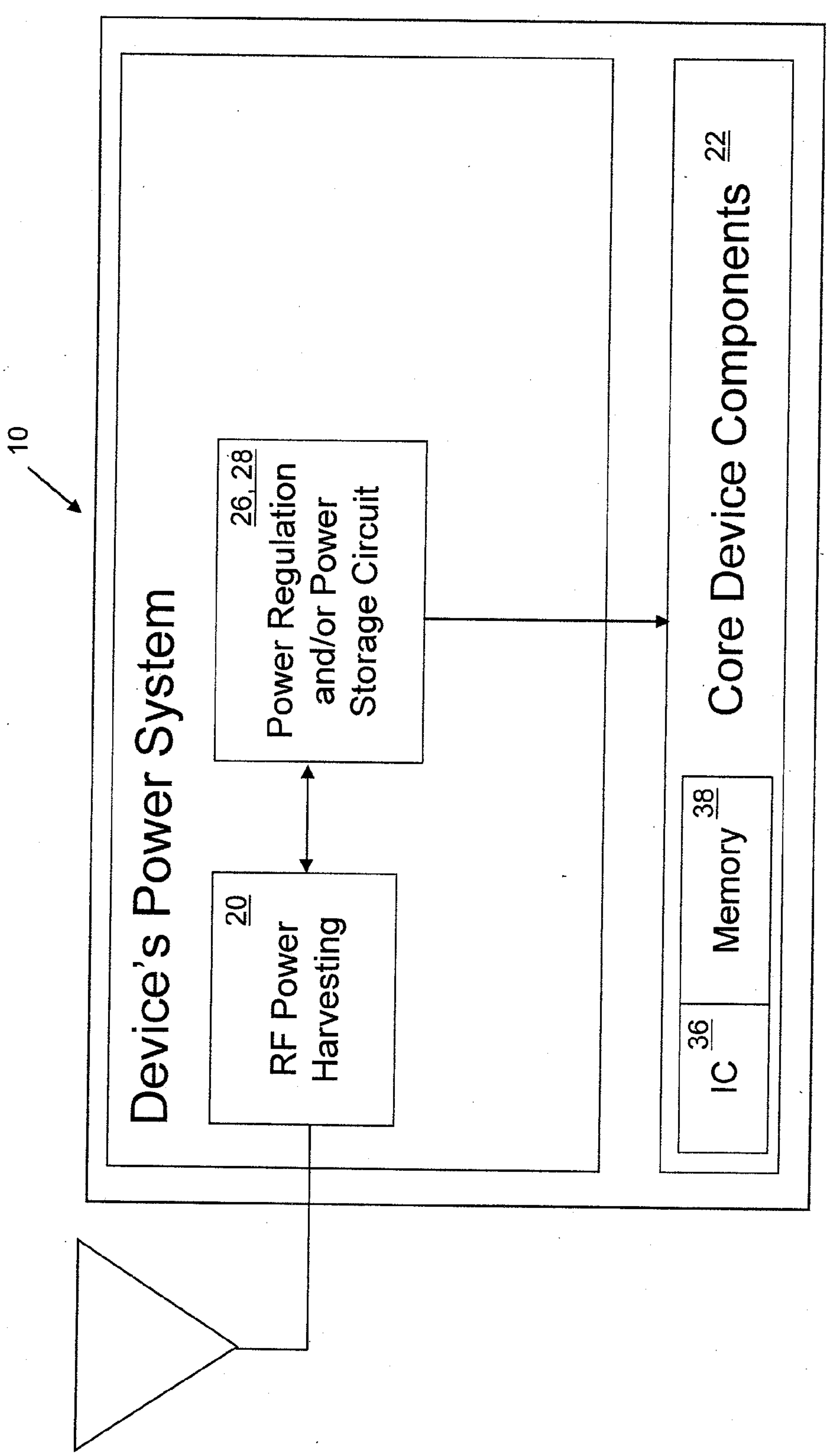
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Core Device Components 22
IC 36
Memory 38

Fig. 2

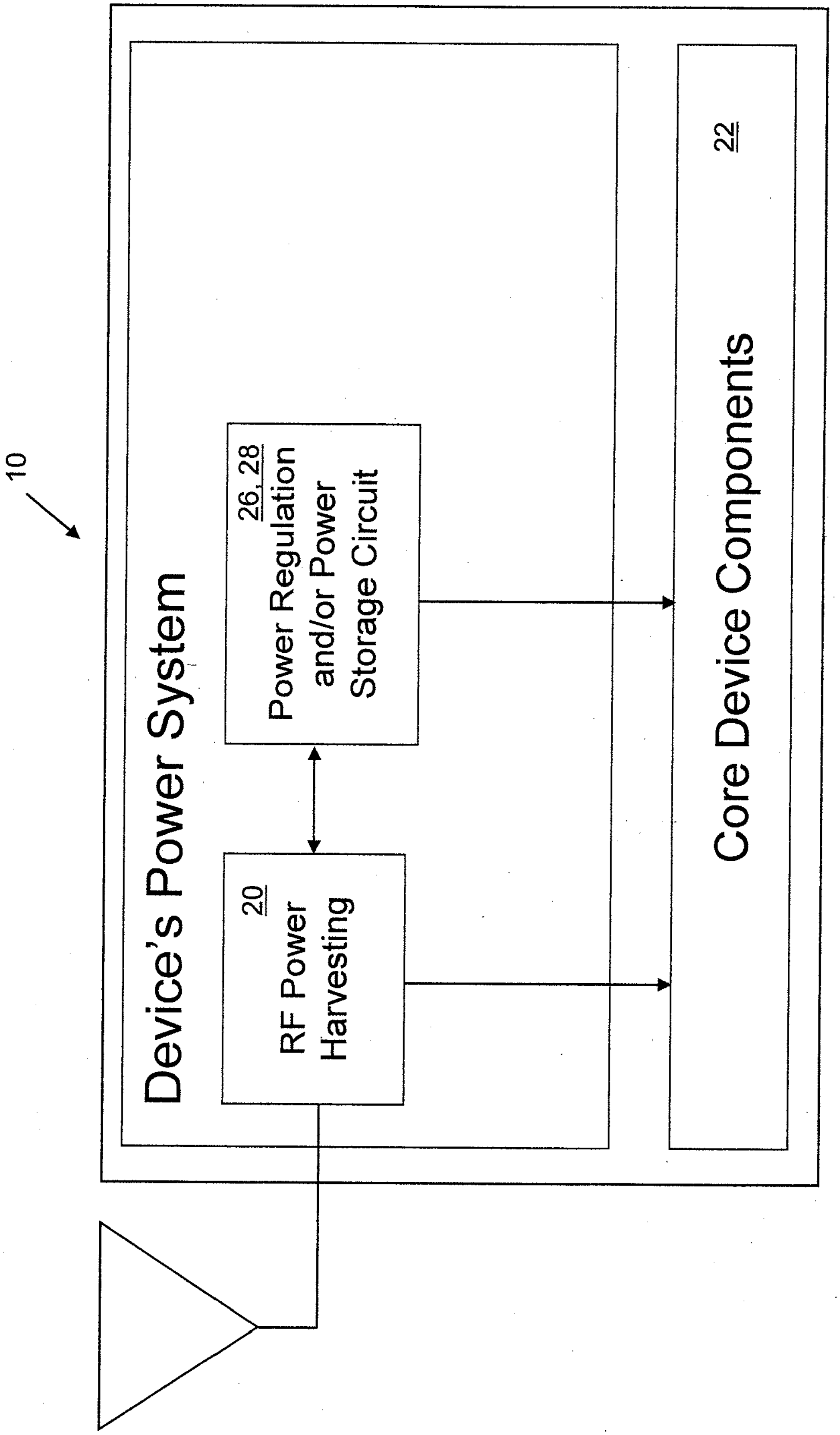
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Core Device Components
22

Fig. 3

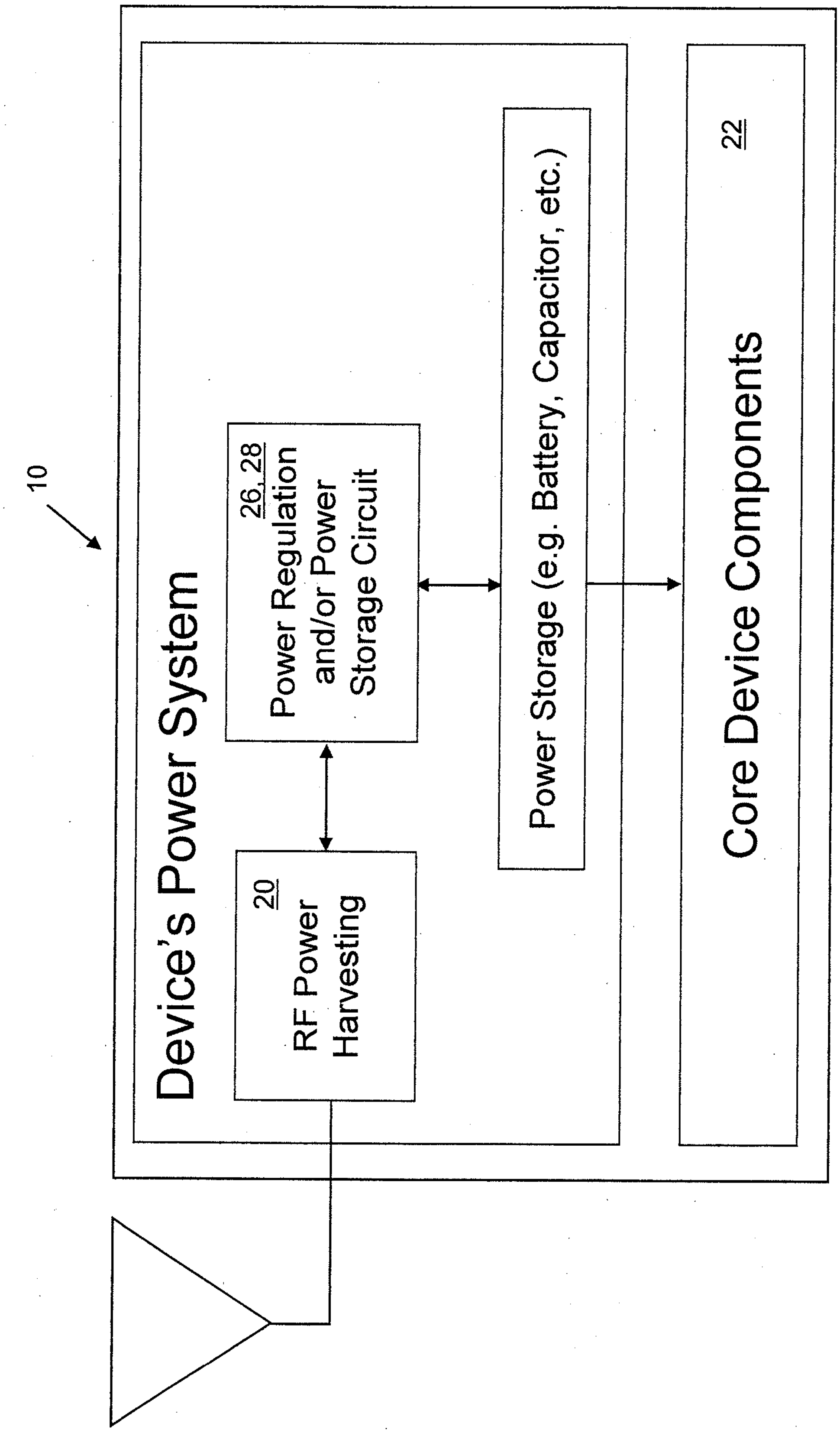
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 4

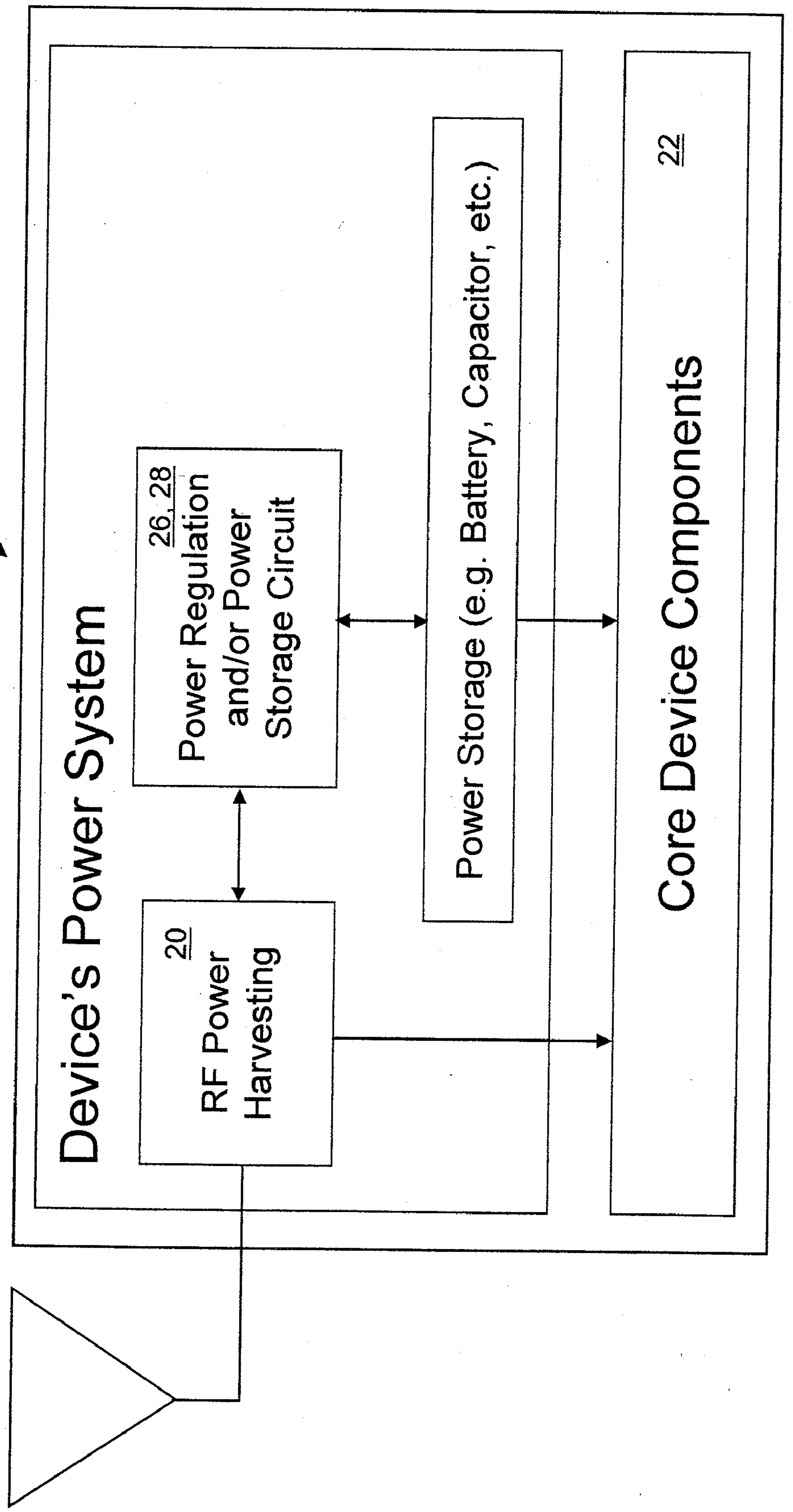
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 5

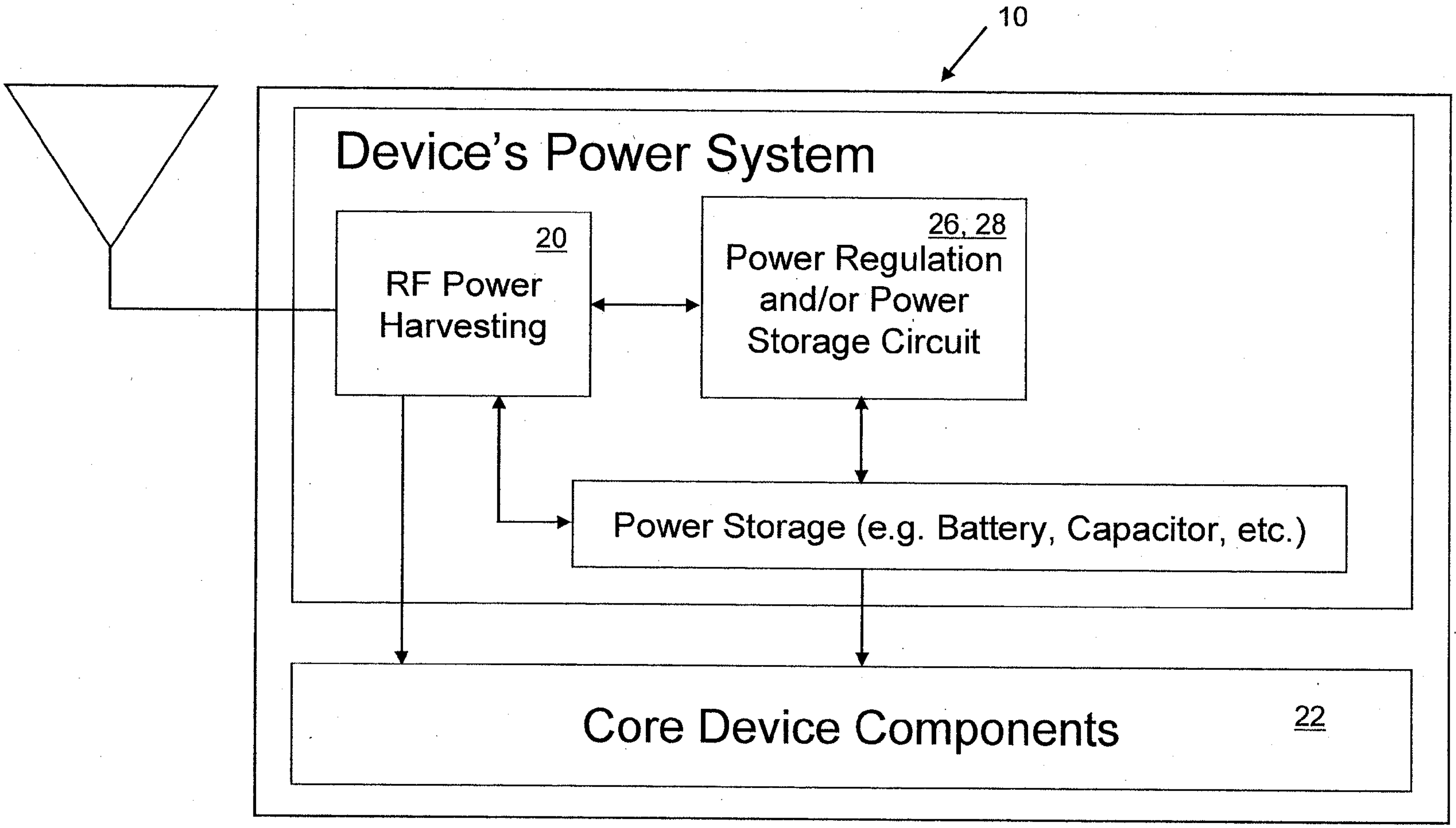
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 6

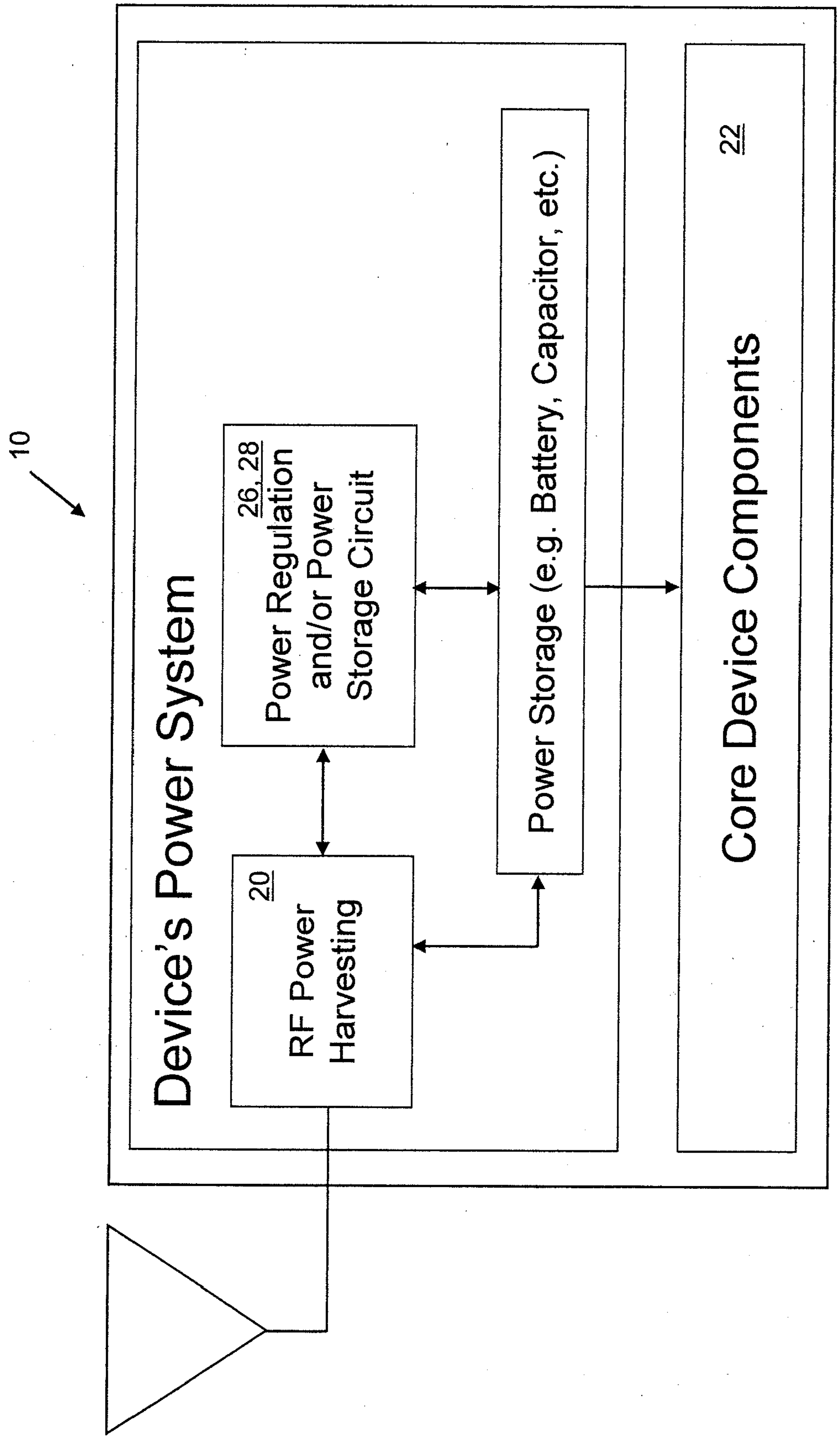
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 7

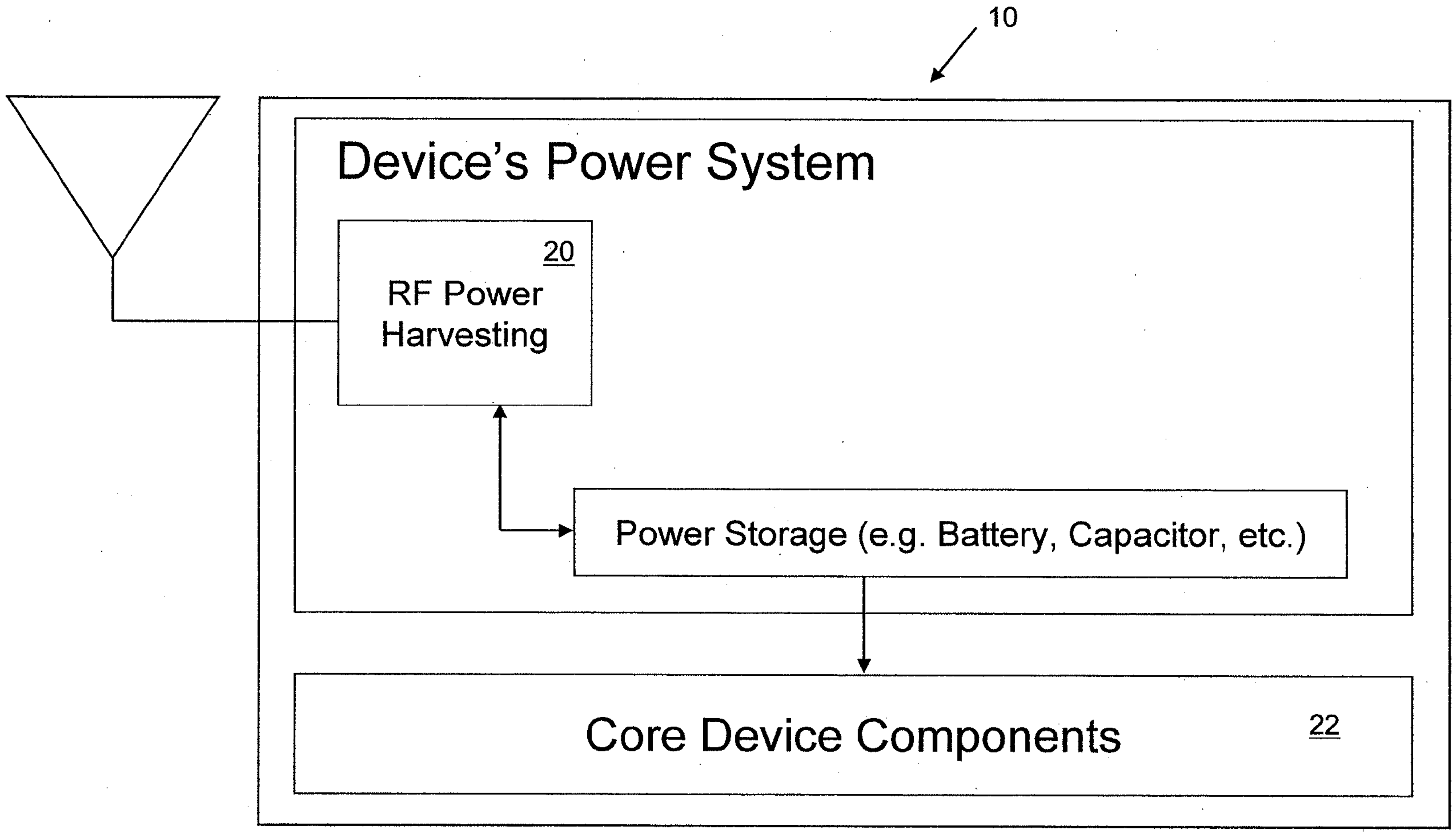
10
Device's Power System
20
RF Power Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 8

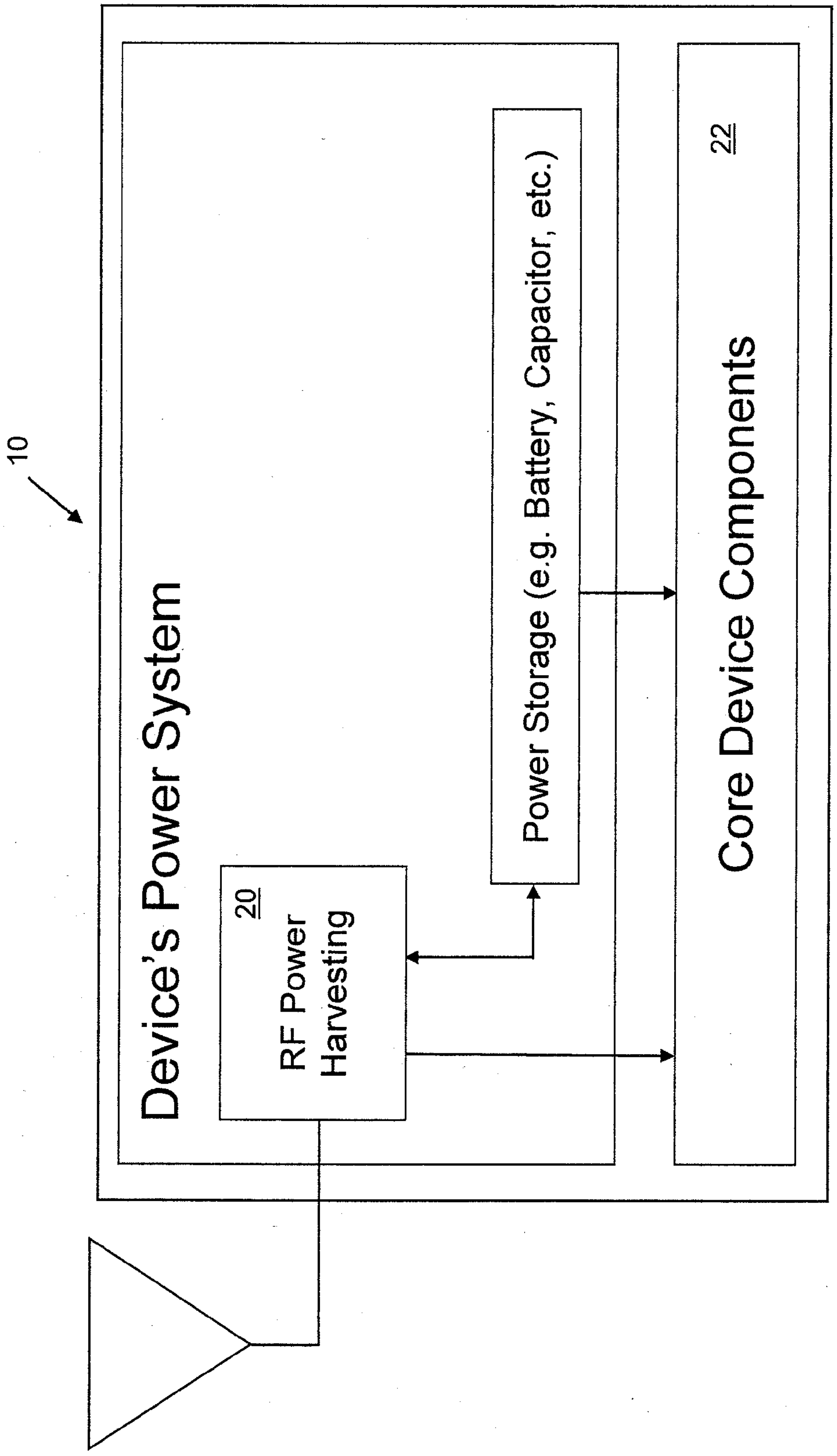
10
Device's Power System
20
RF Power Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 9

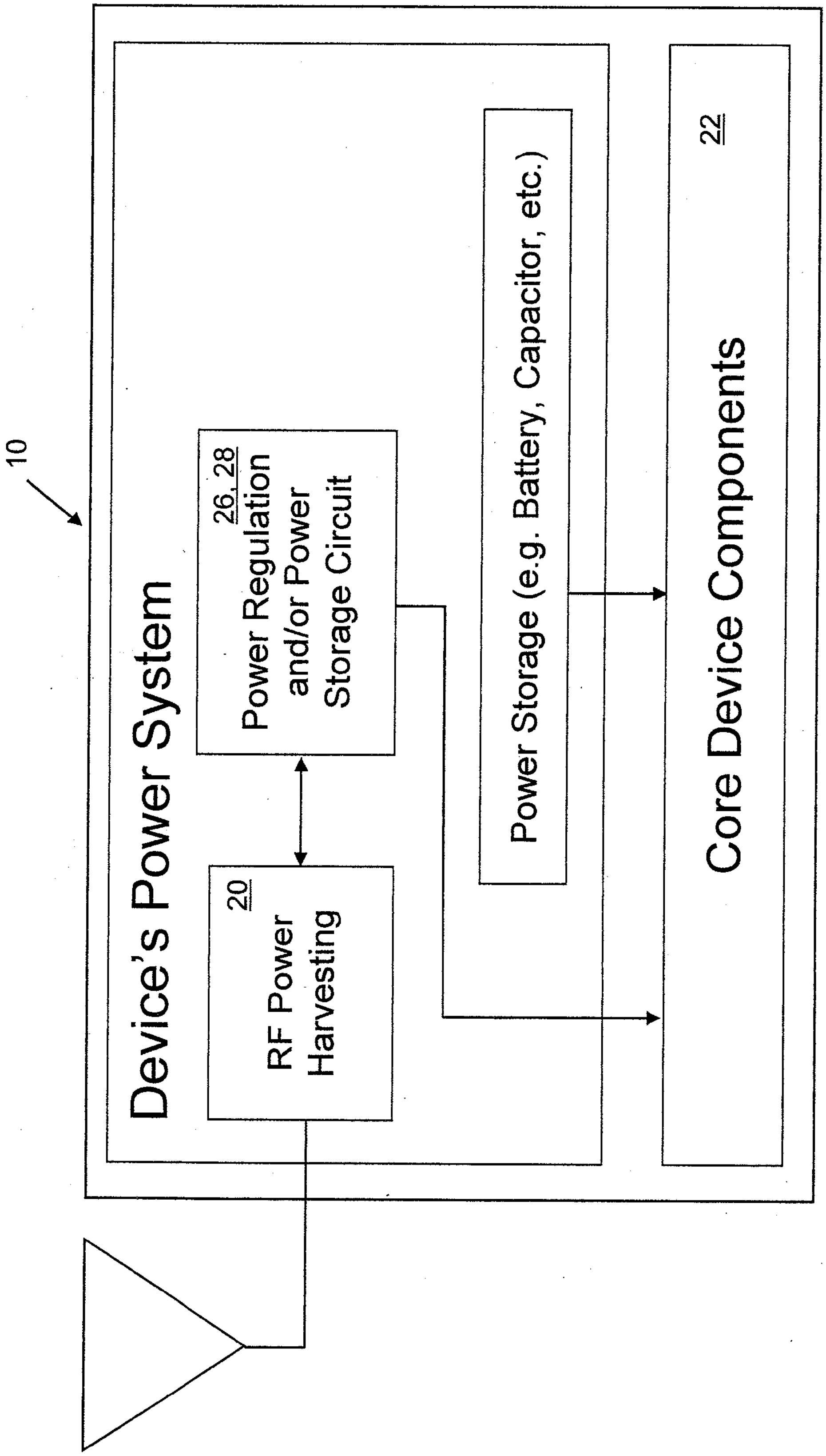
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 10

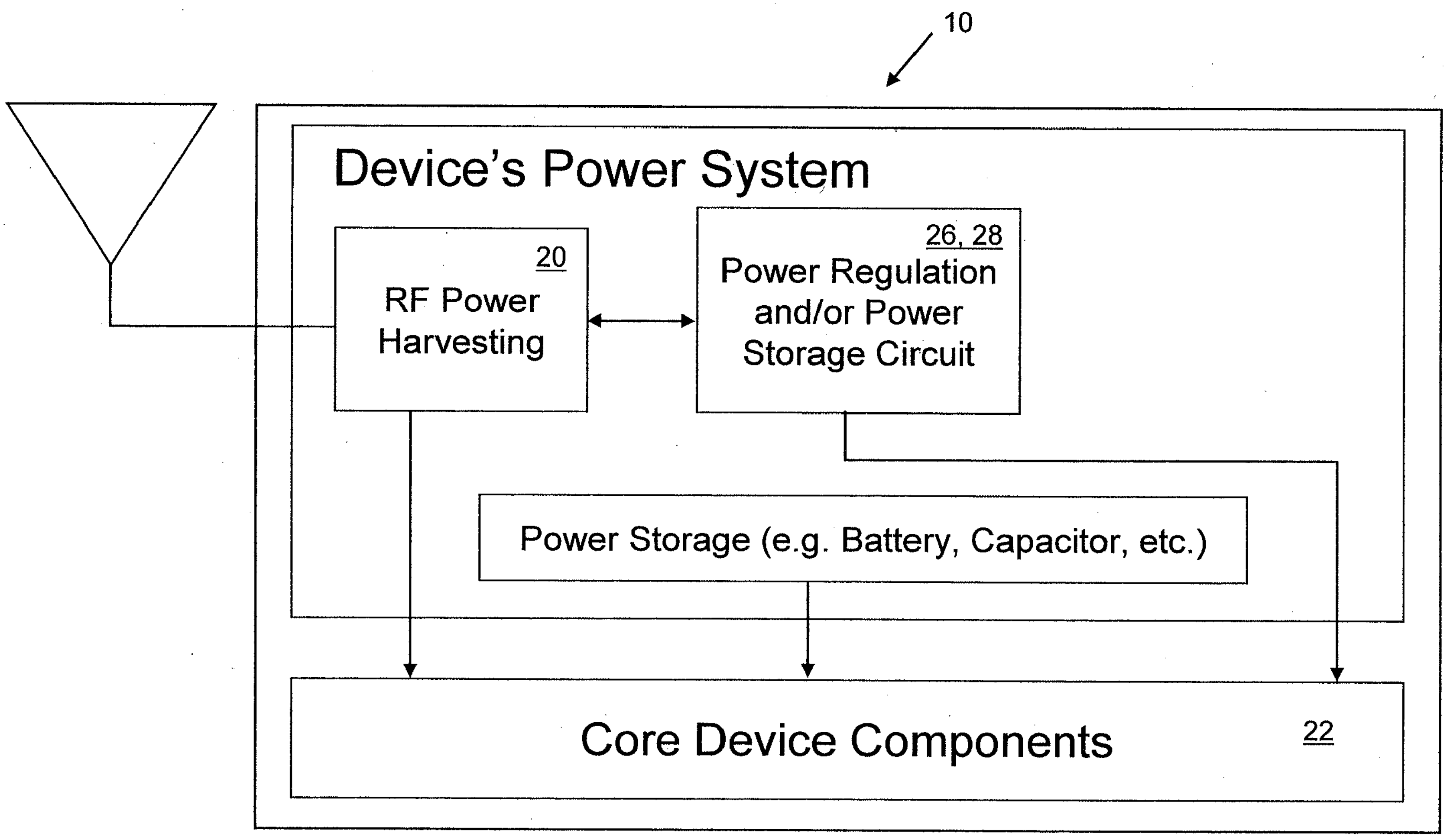
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 11

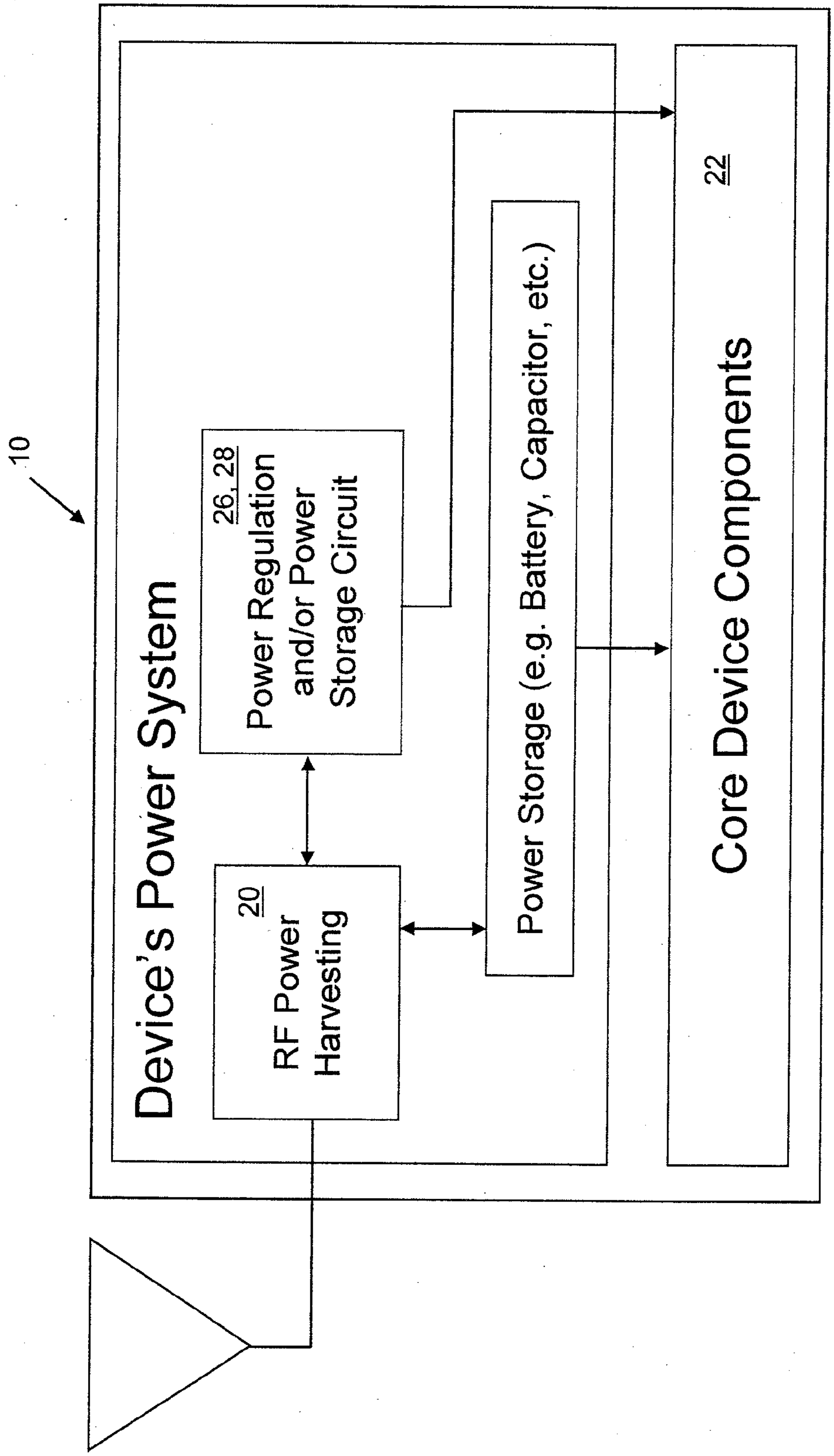
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 12

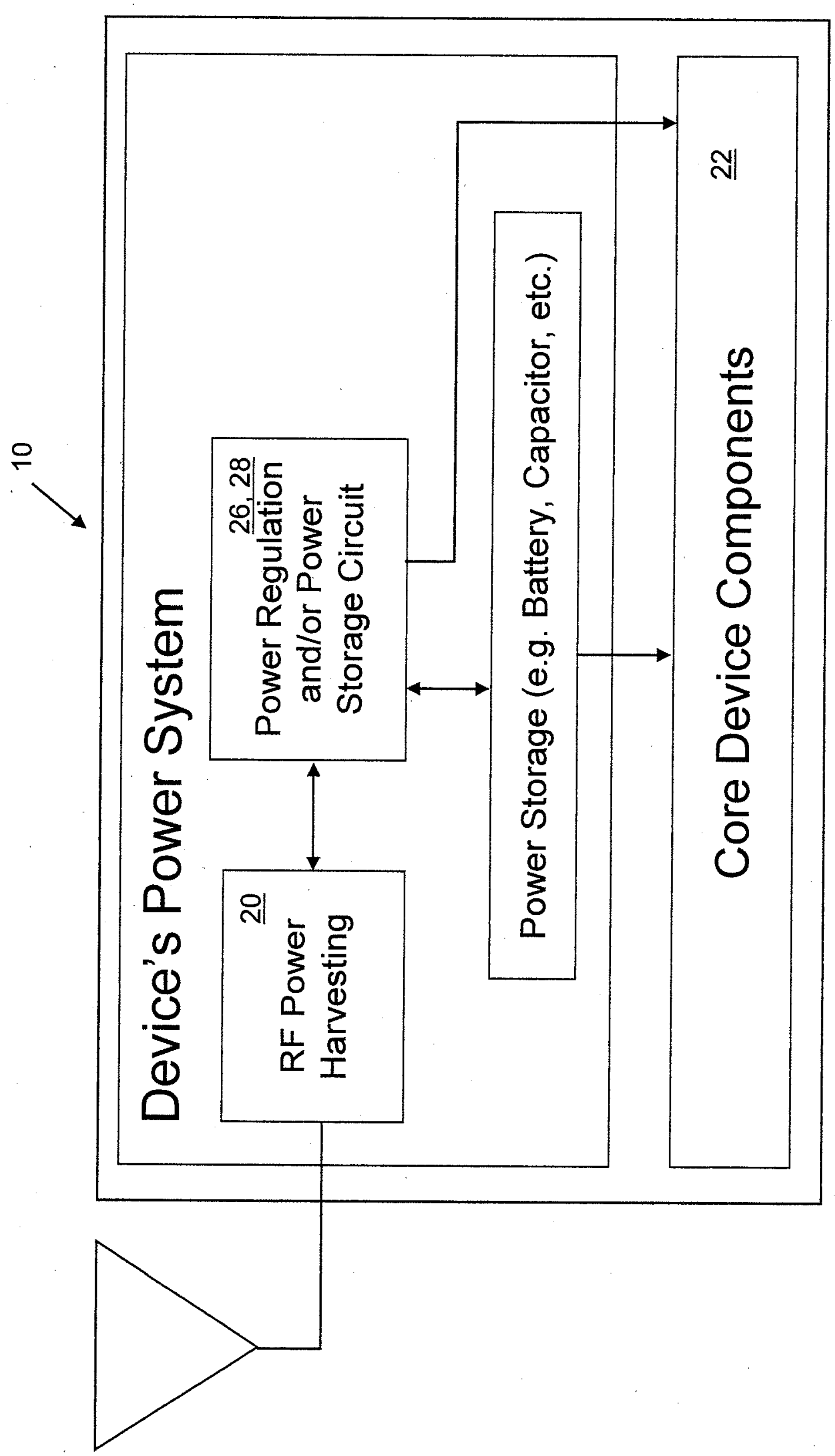
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 13

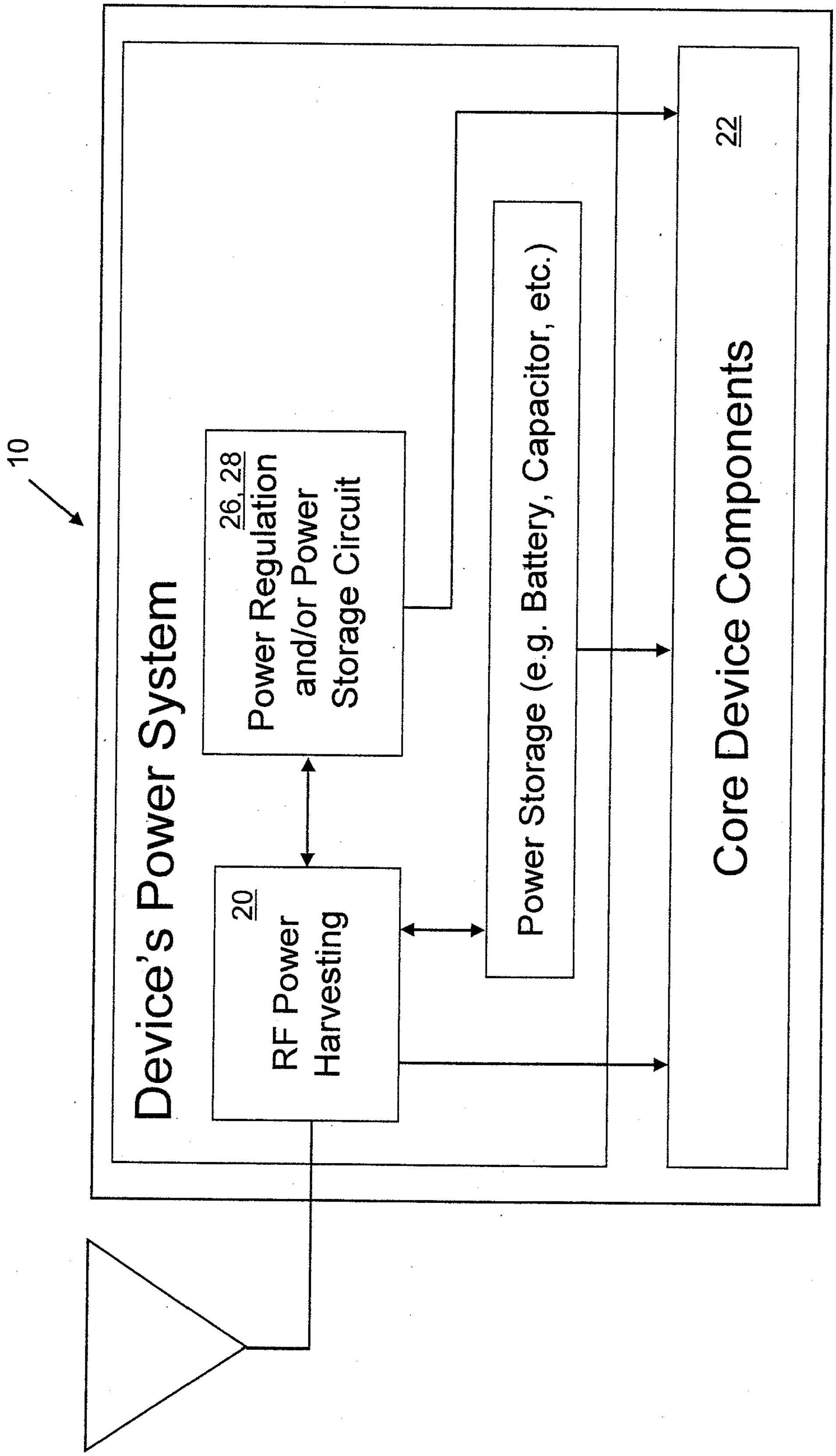
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 14

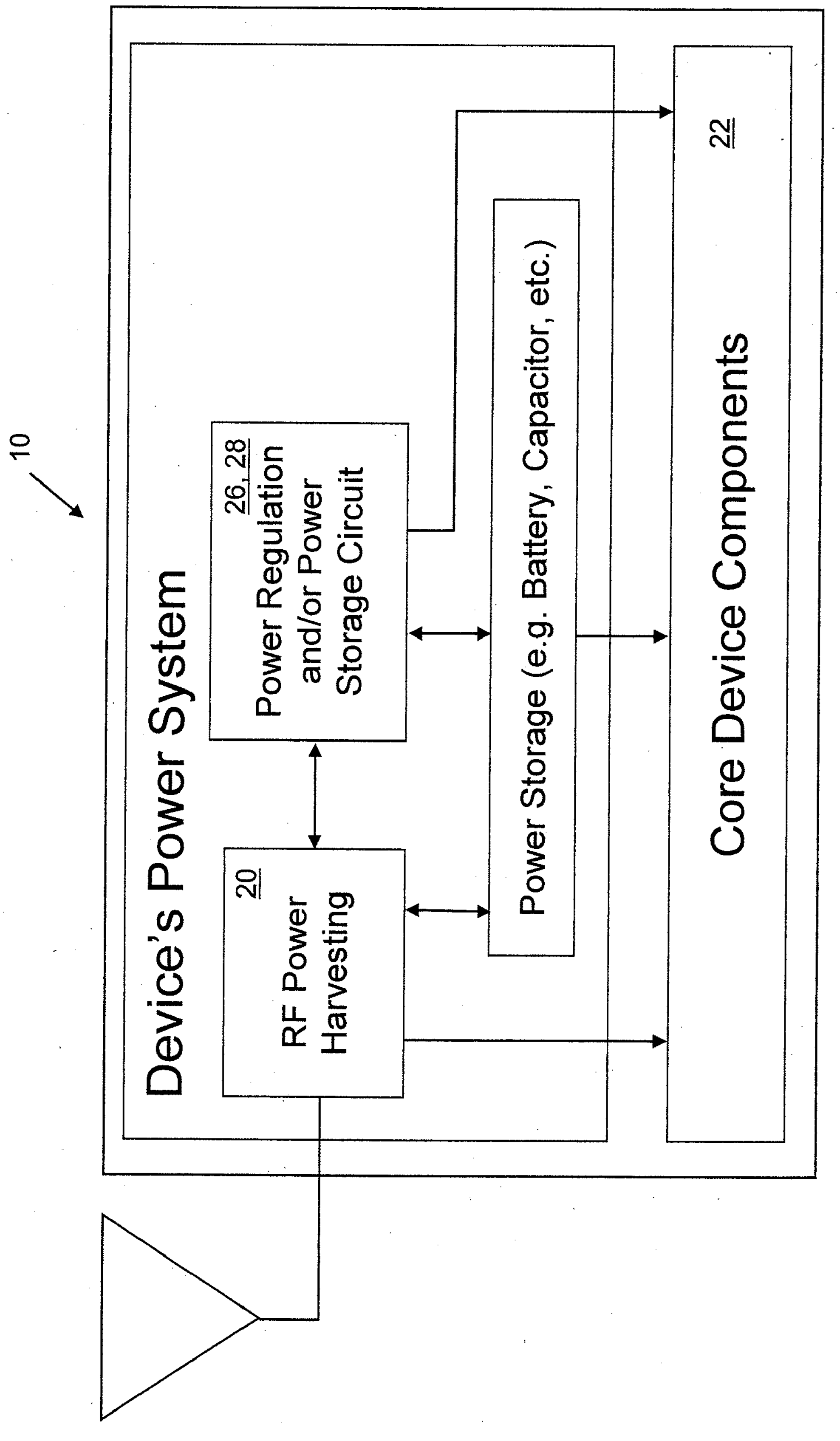
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 15

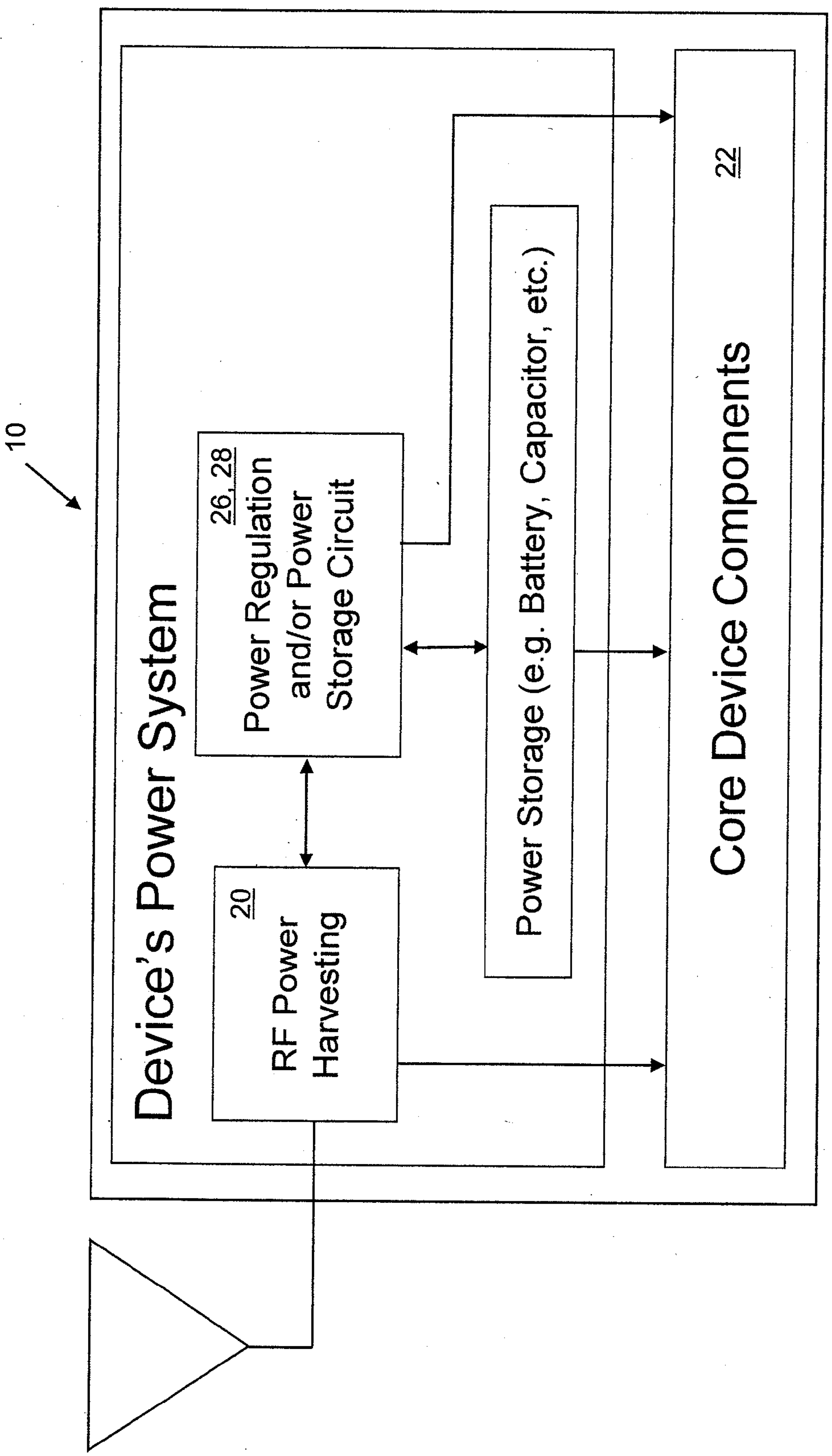
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 16

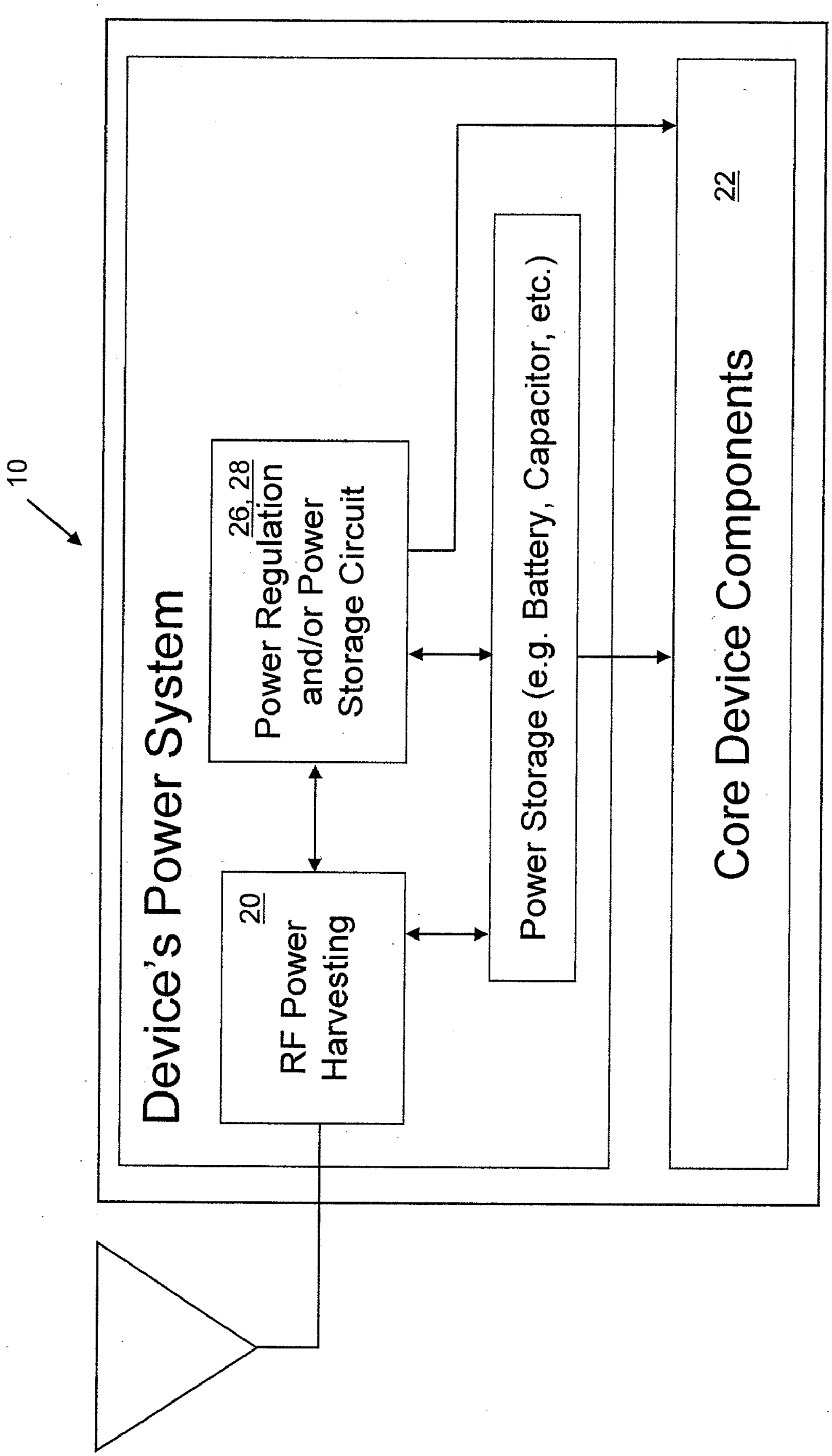
10
Device's Power System
Power Regulation and/or Power Storage Circuit
26, 28
RF Power Harvesting
20
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 17

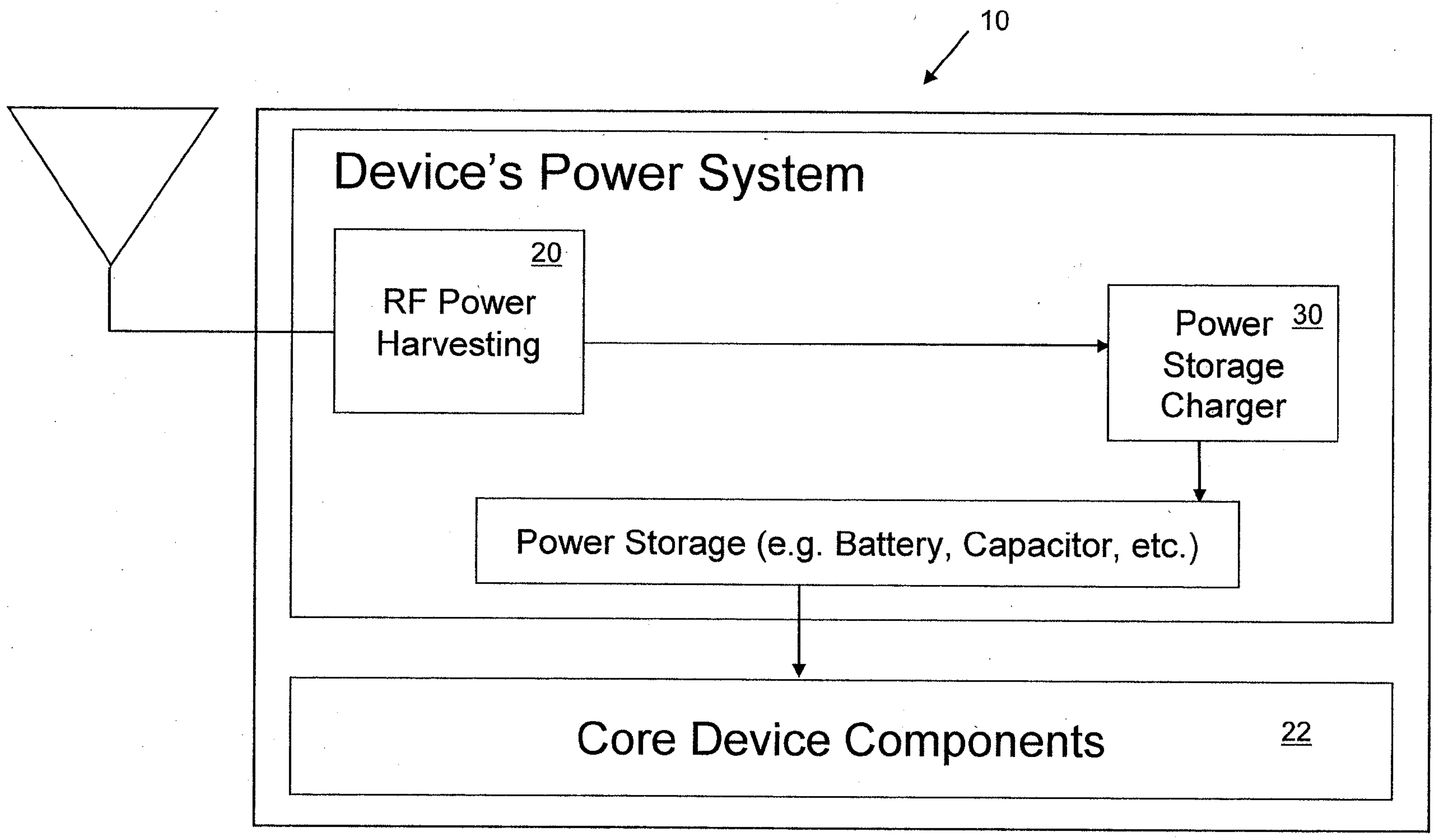
10
Device's Power System
20
RF Power Harvesting
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 18

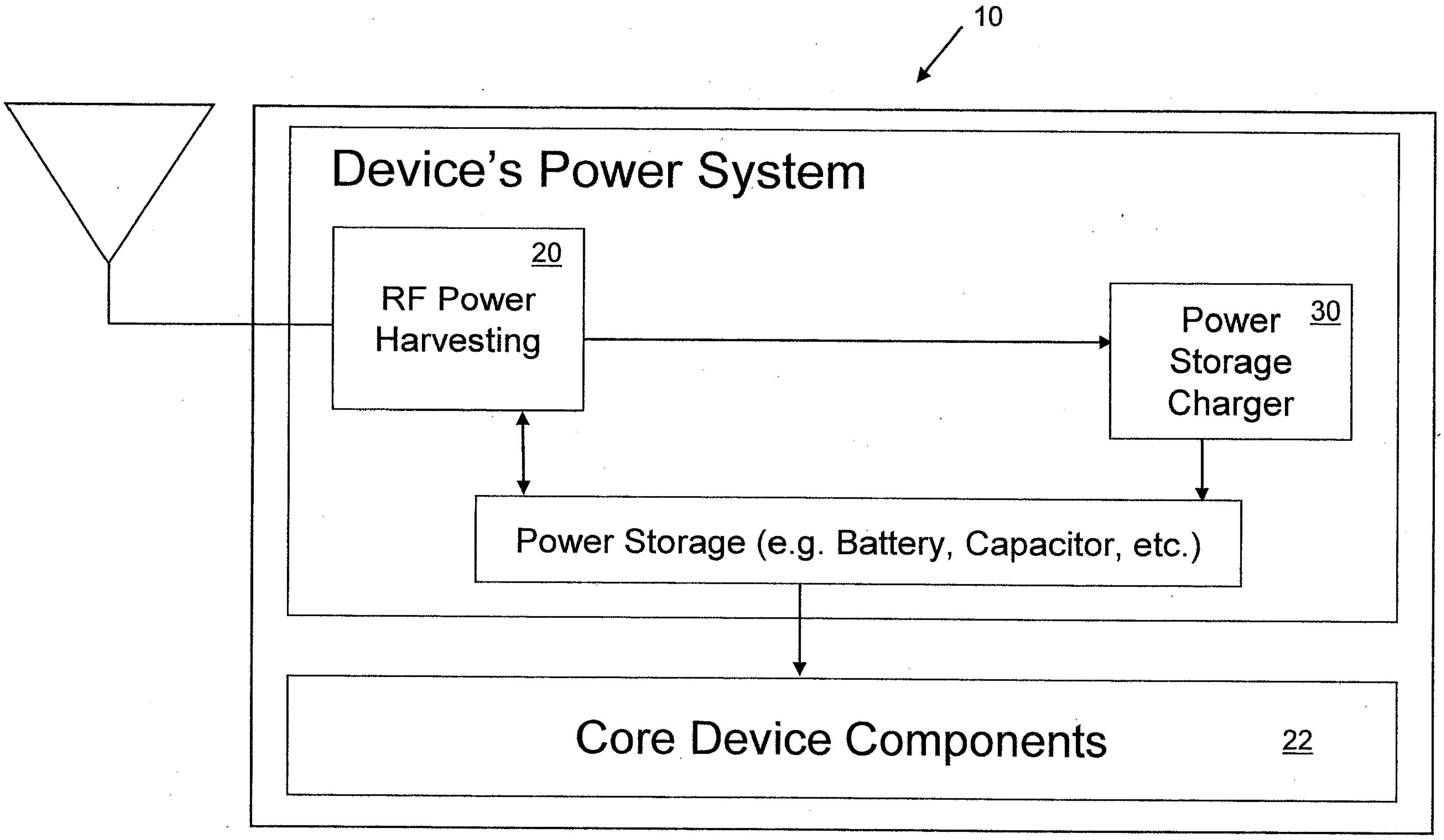
10
Device's Power System
20
RF Power Harvesting
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 19

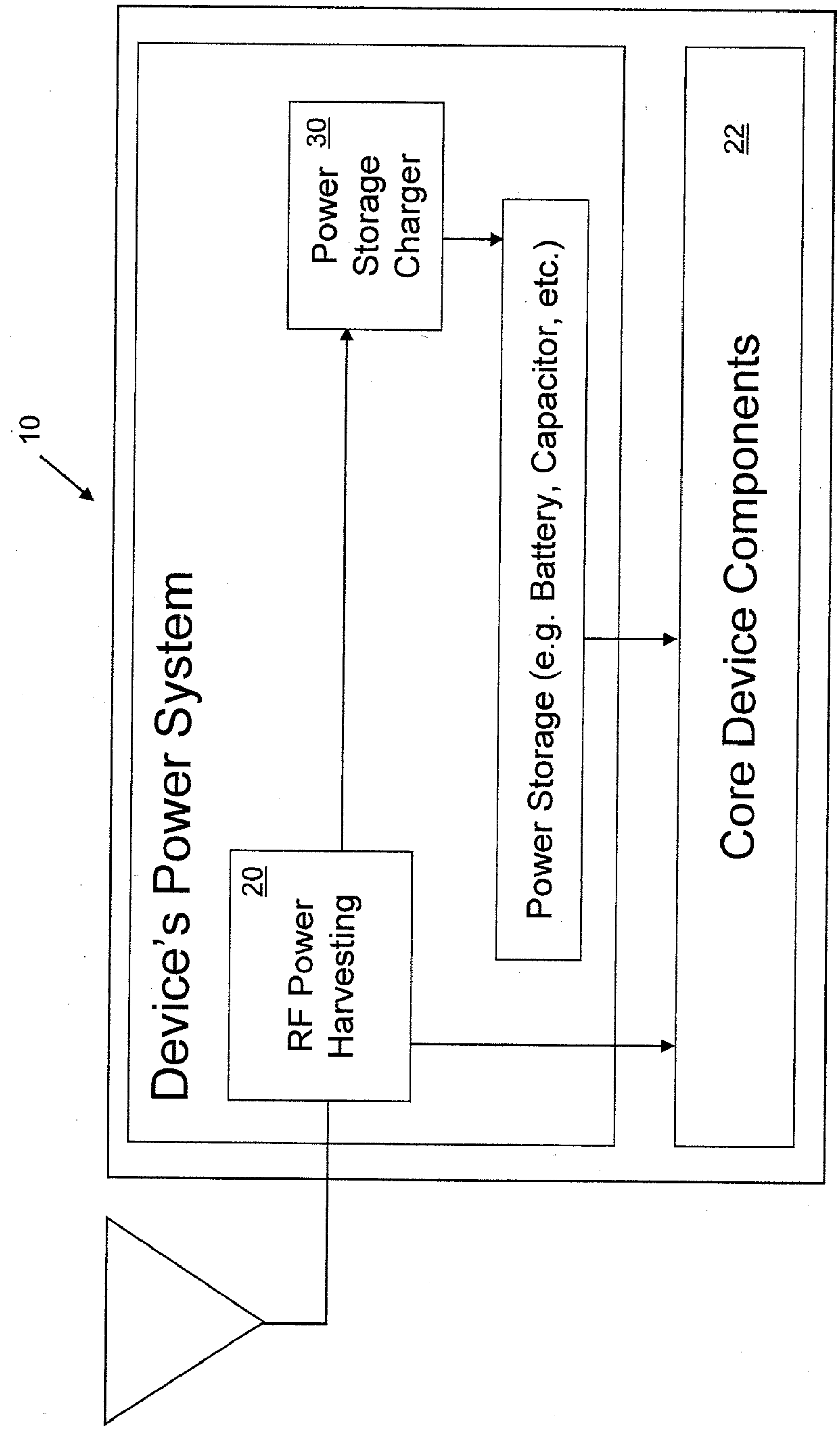
10
Device's Power System
RF Power Harvesting 20
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 20

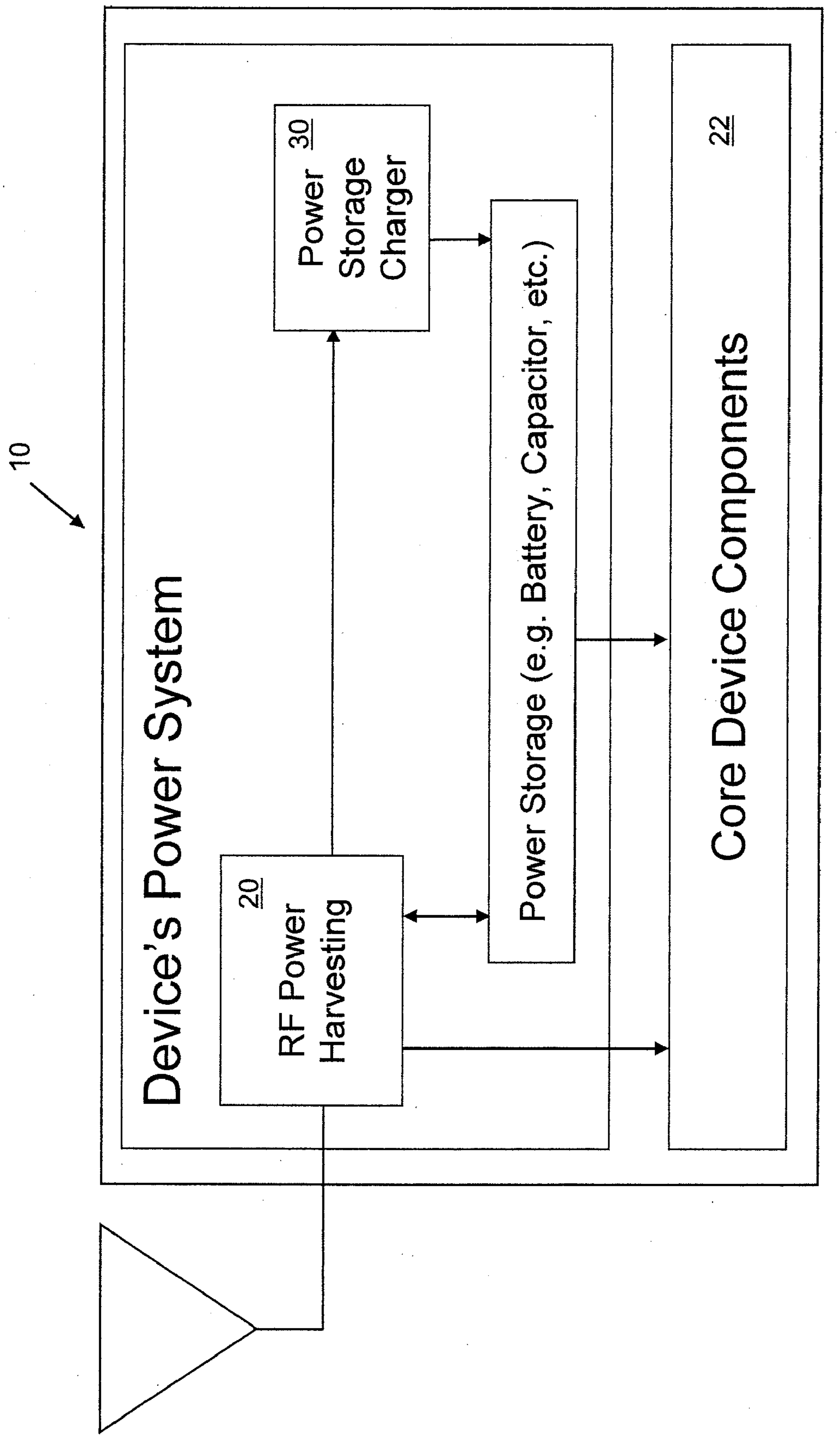
10
Device's Power System
RF Power Harvesting
20
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 21

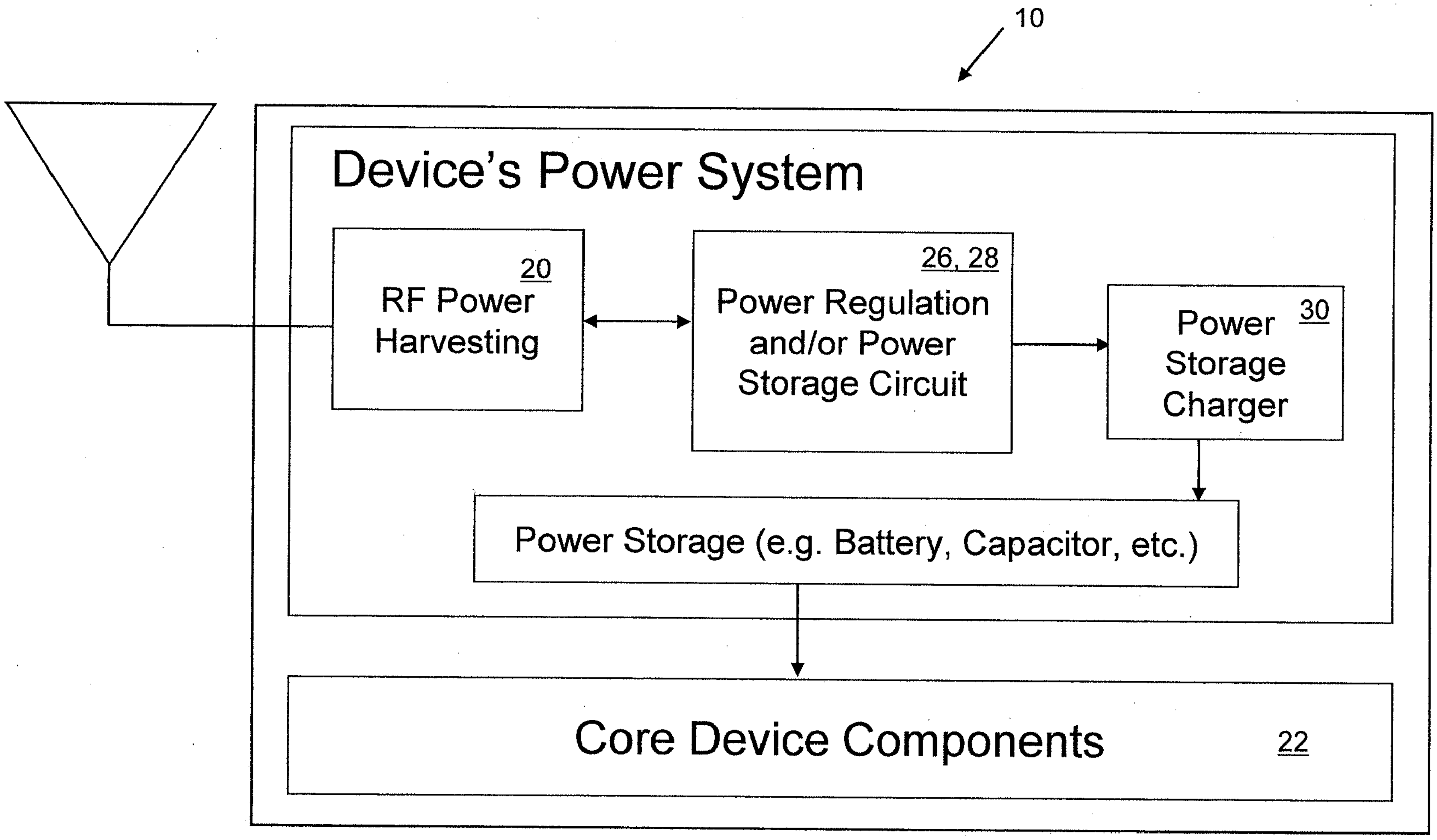
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 22

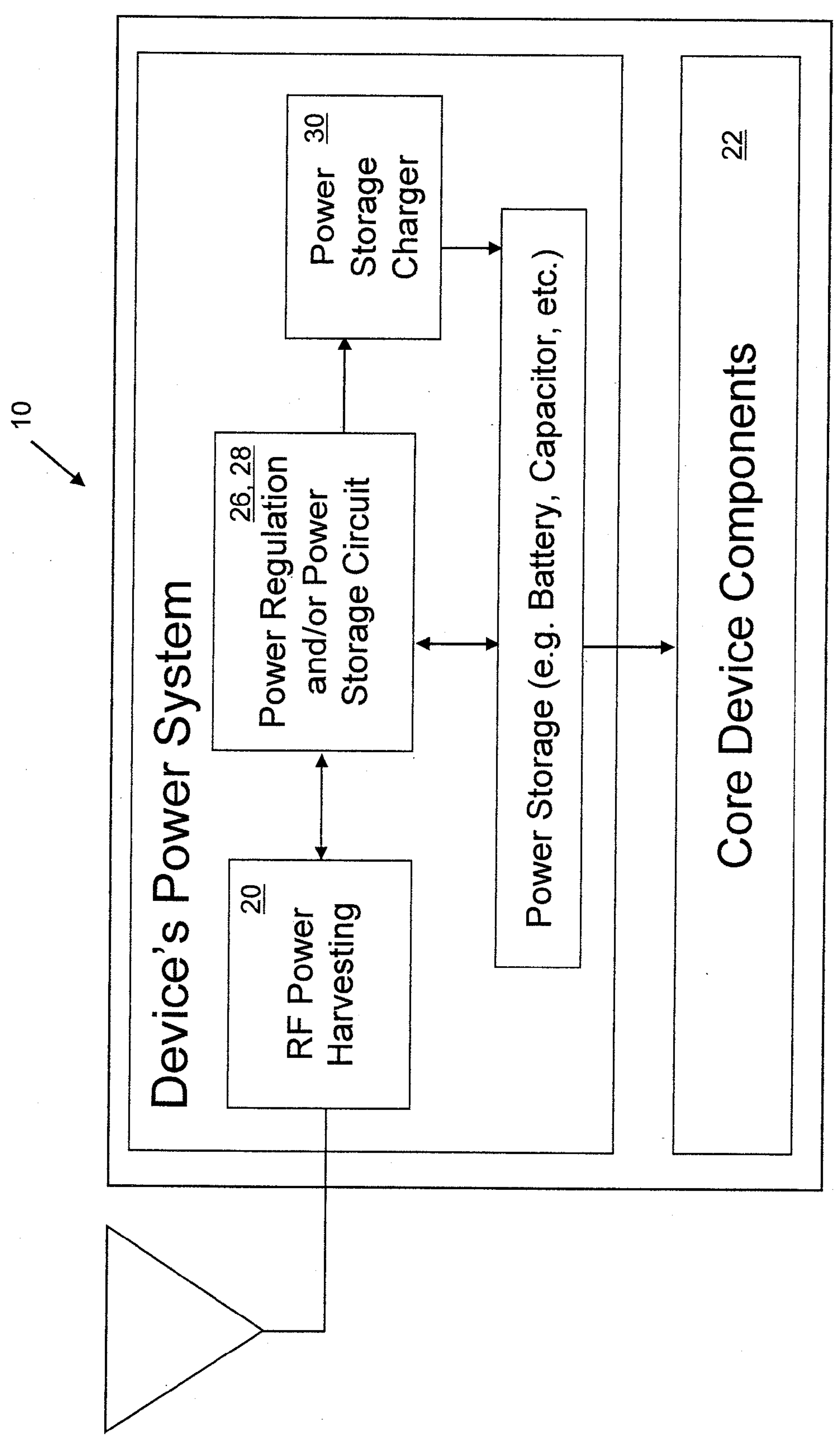
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 23

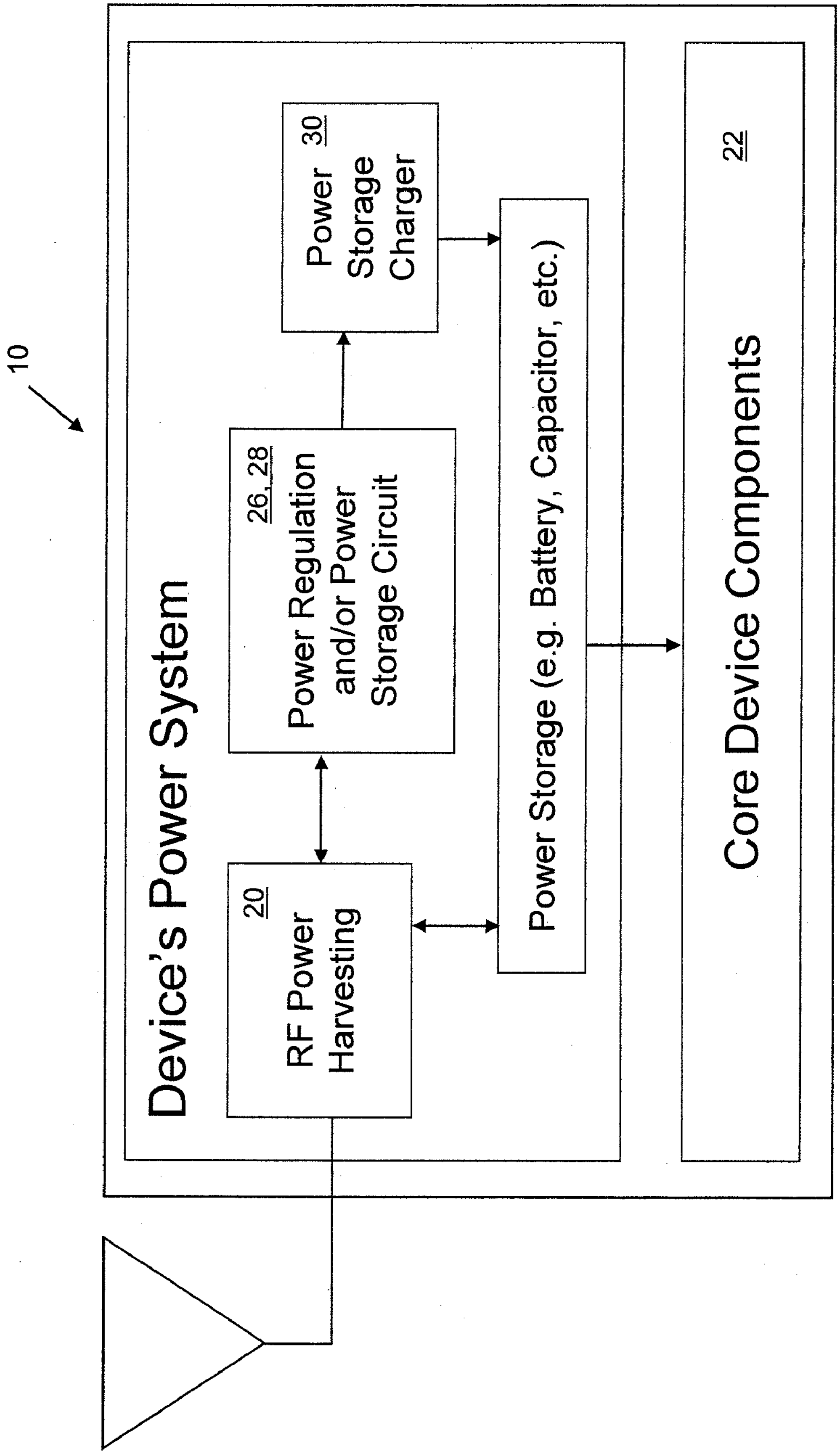
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 24

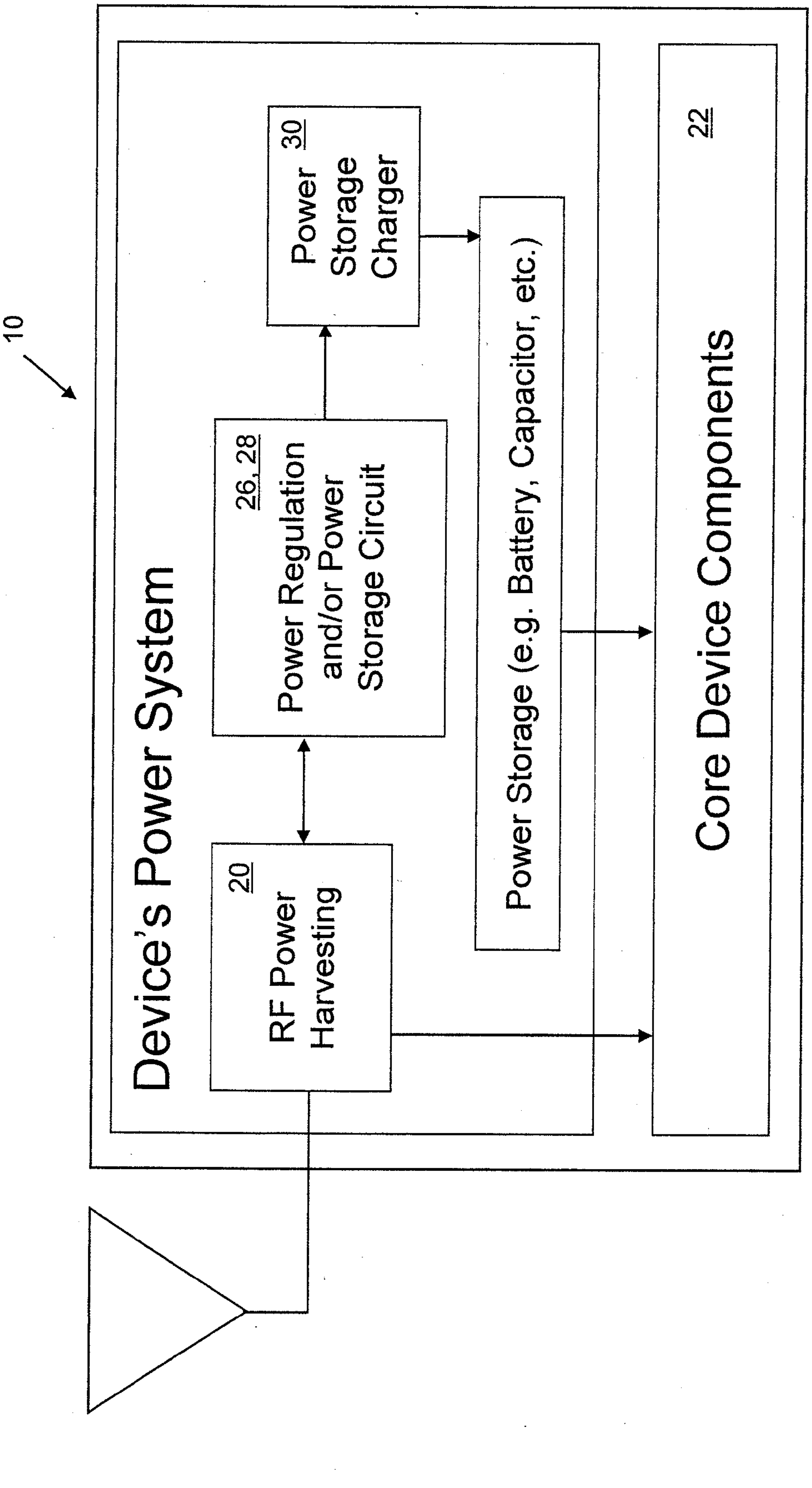
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 25

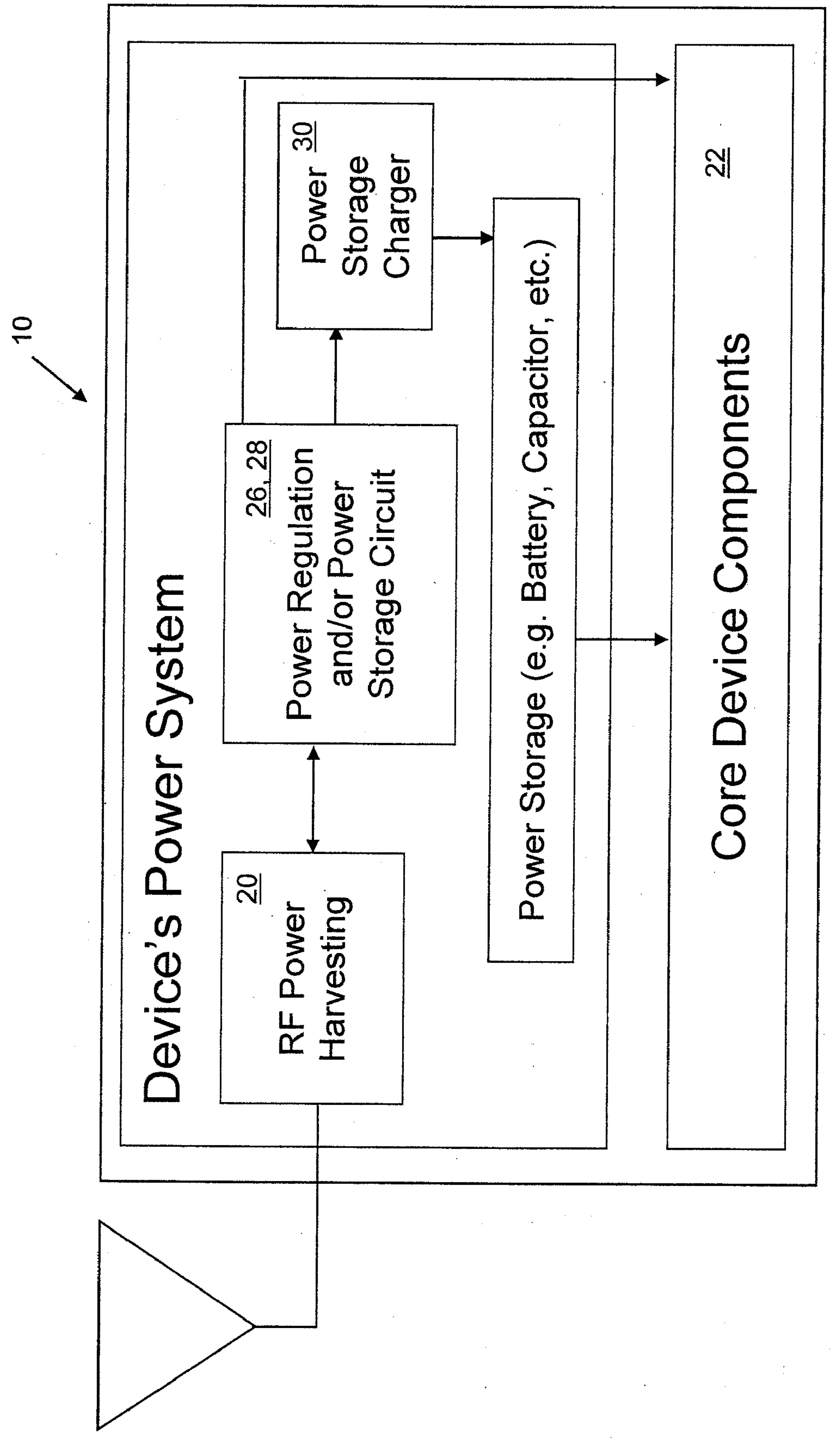
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 26

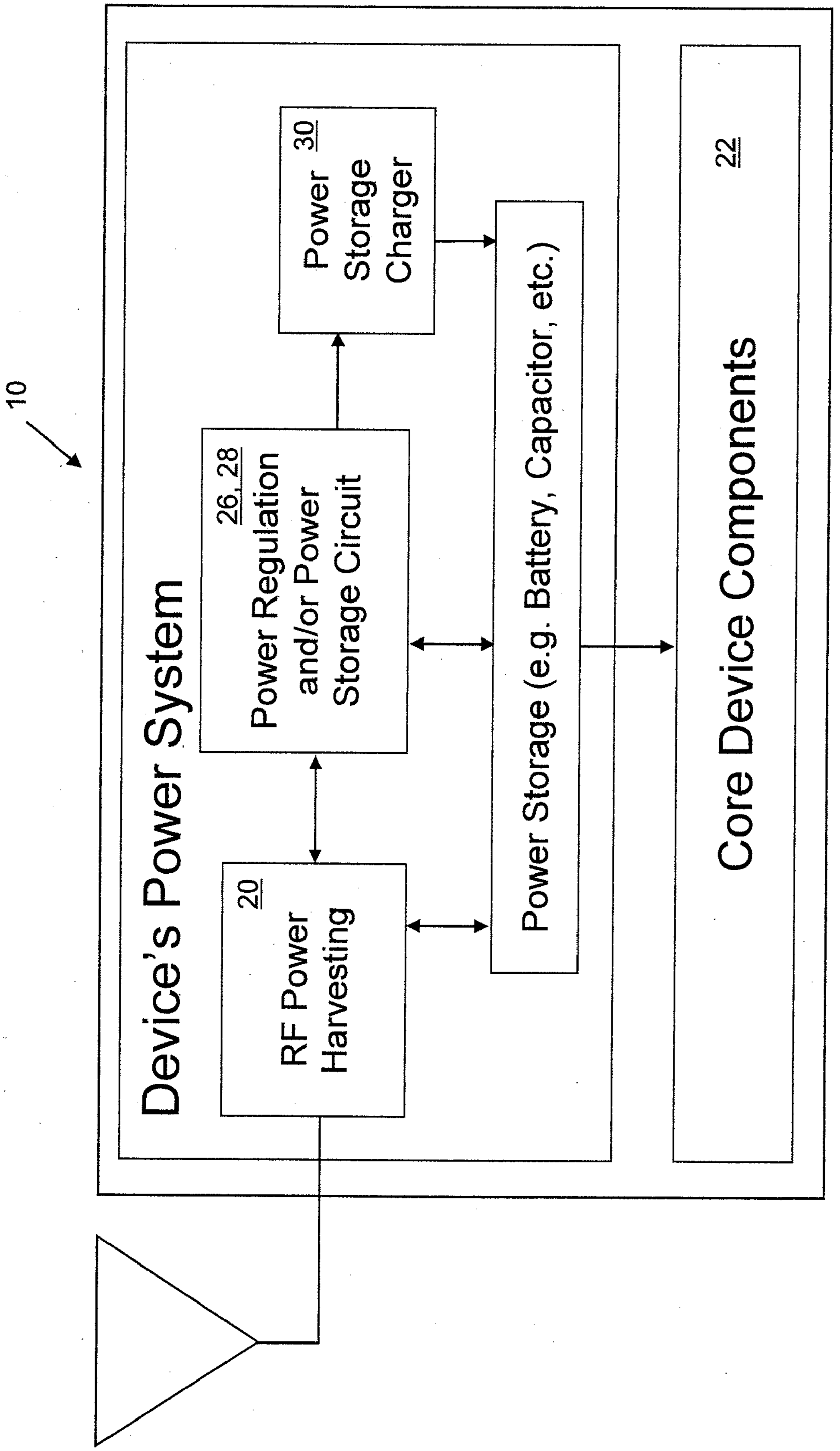
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 27

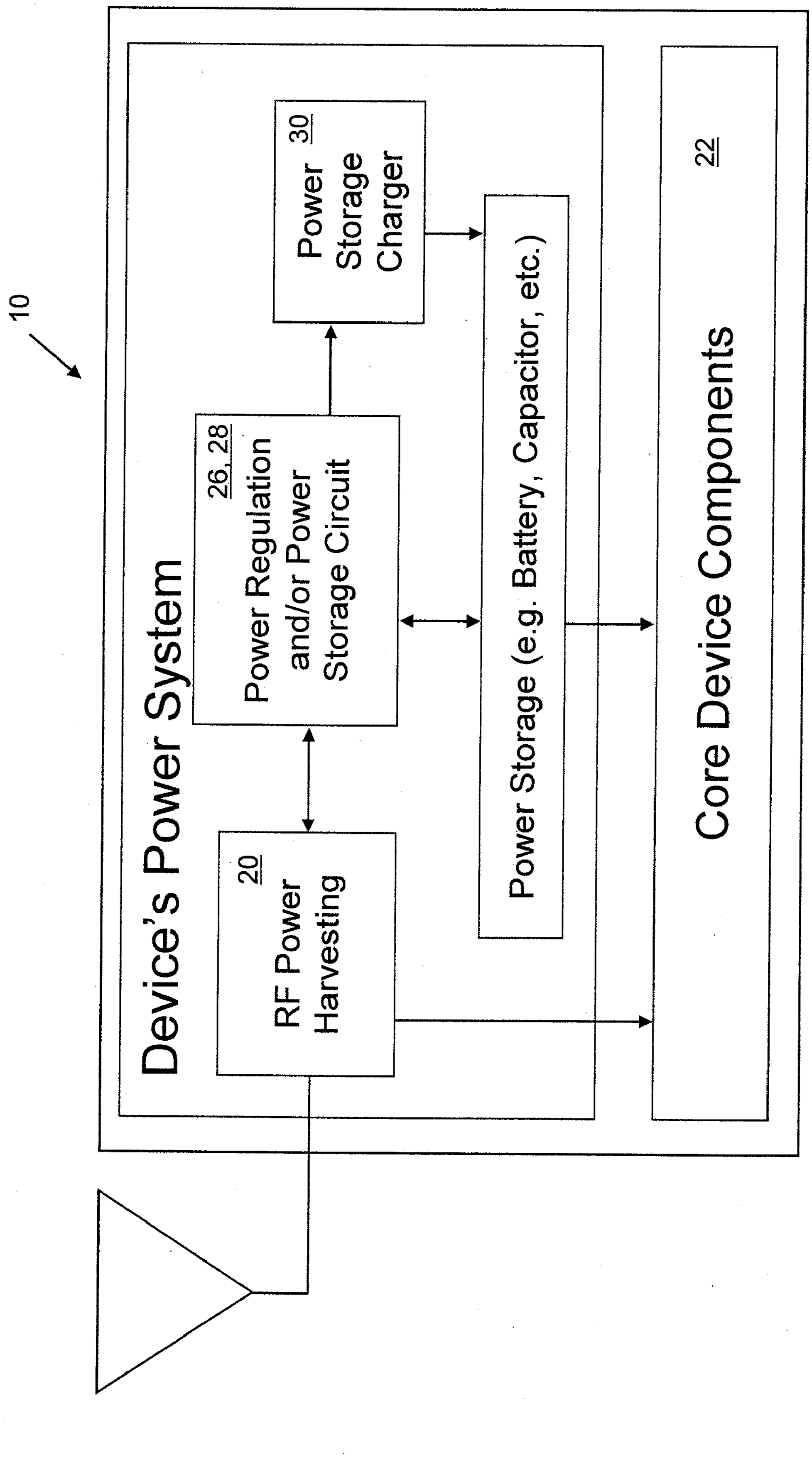
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 28

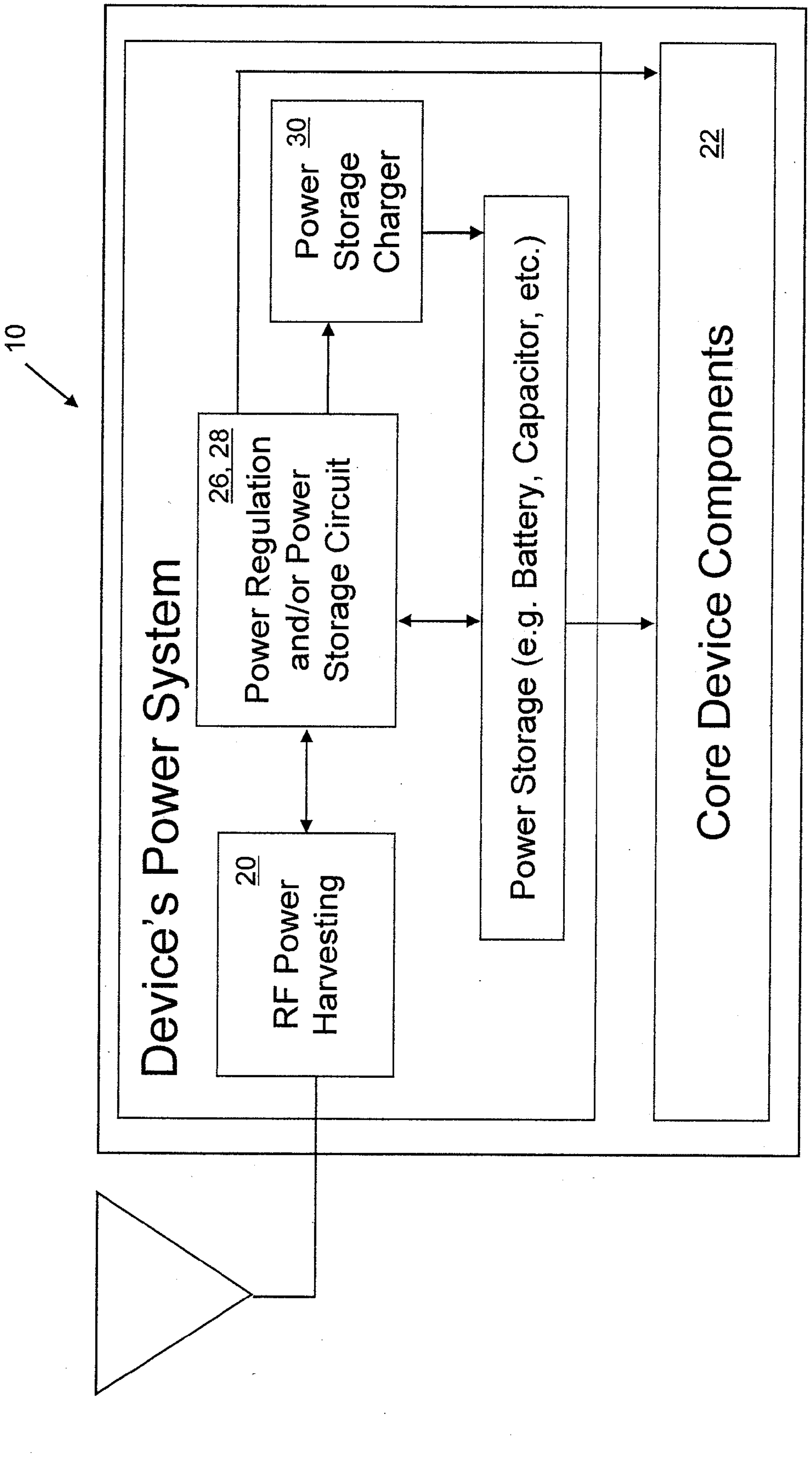
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 29

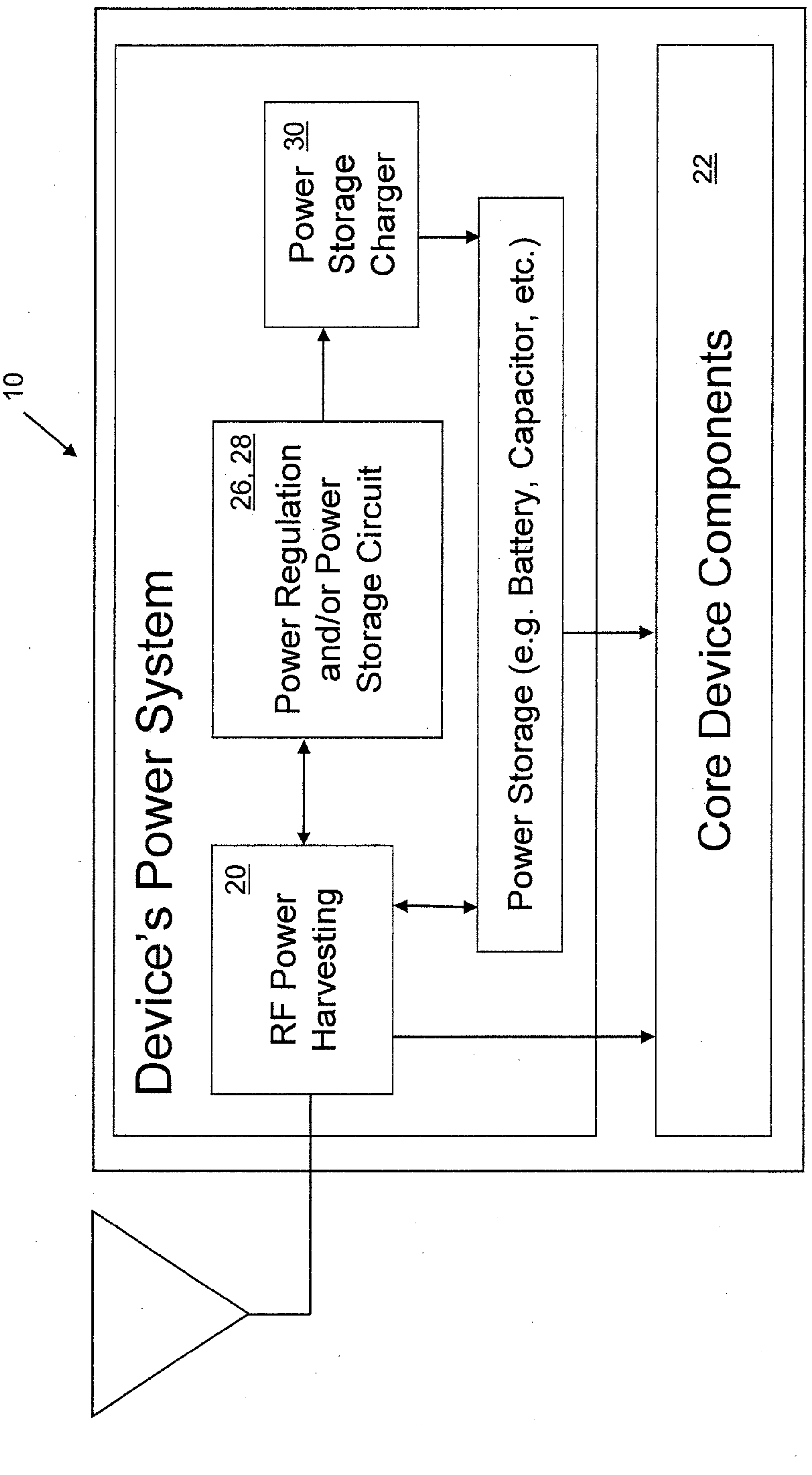
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 30

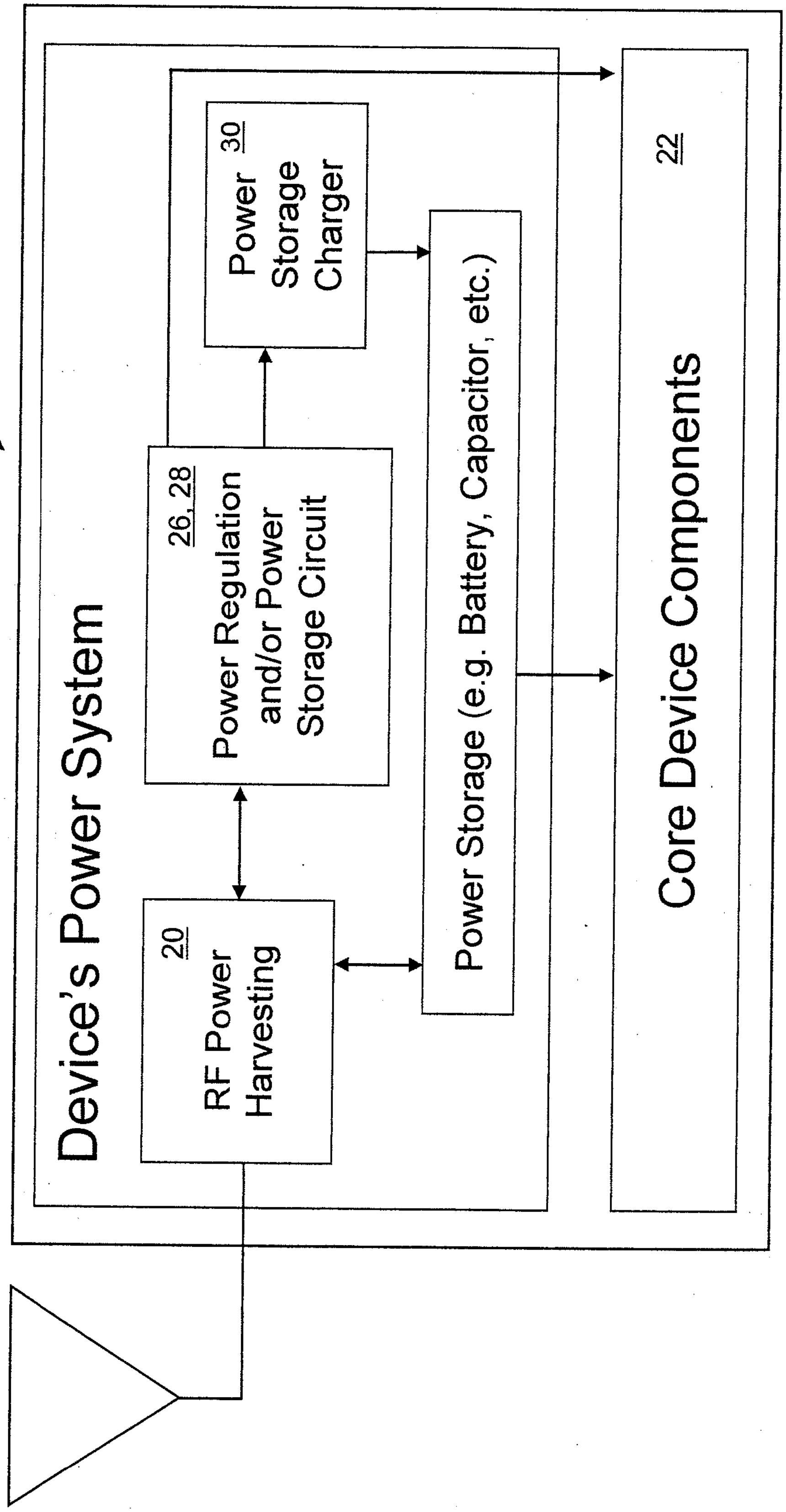
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 31

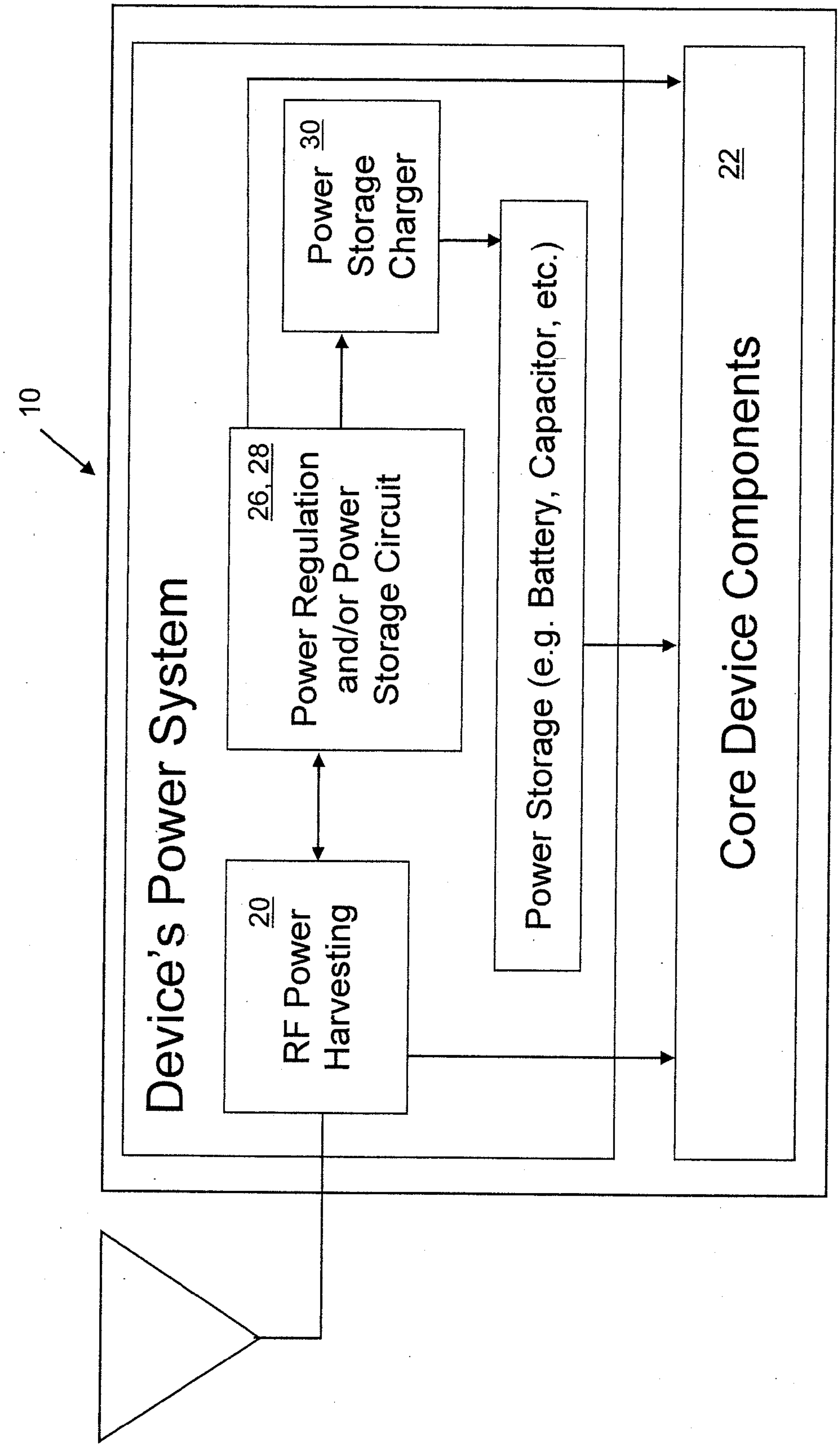
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 32

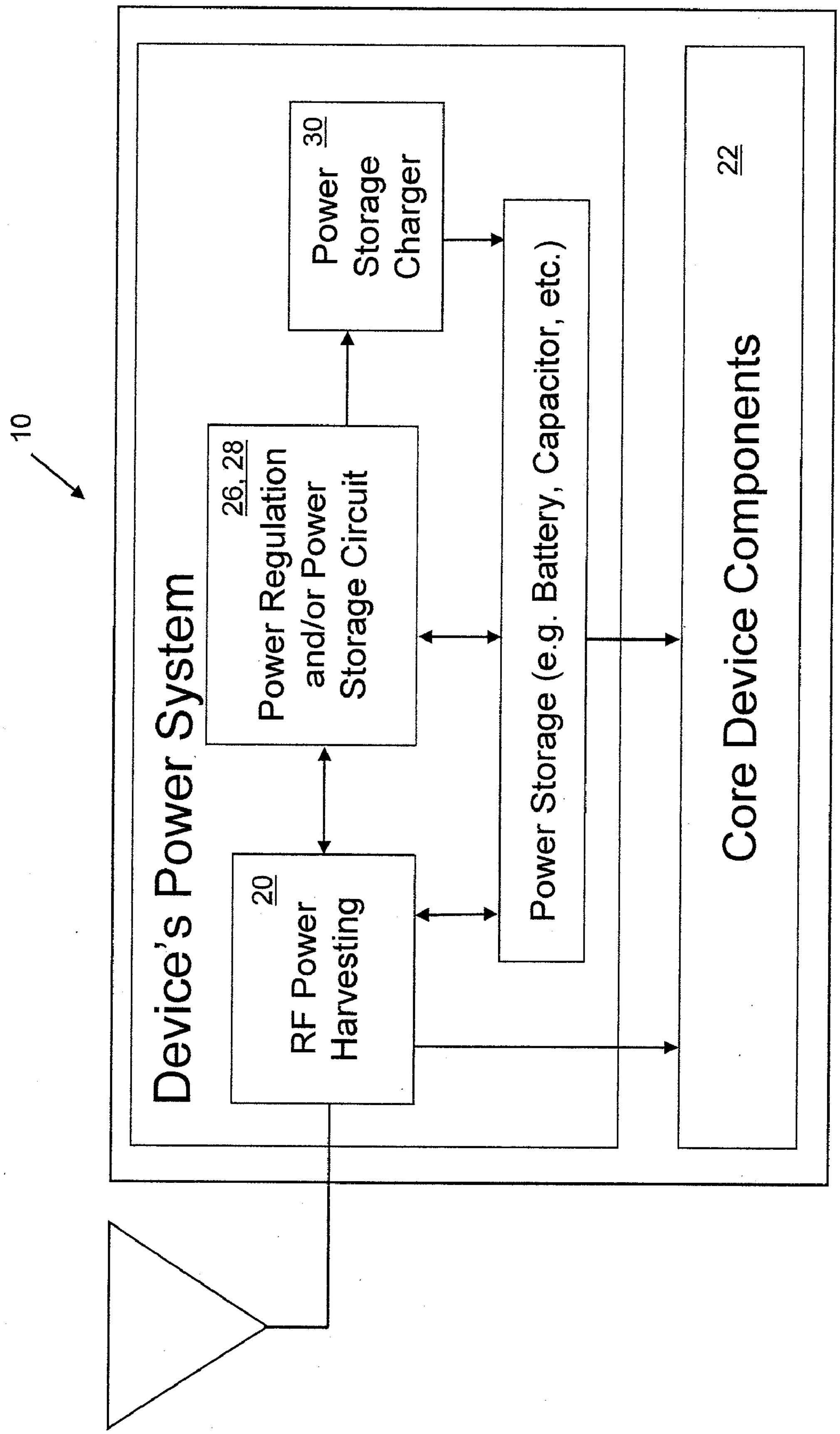
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 33

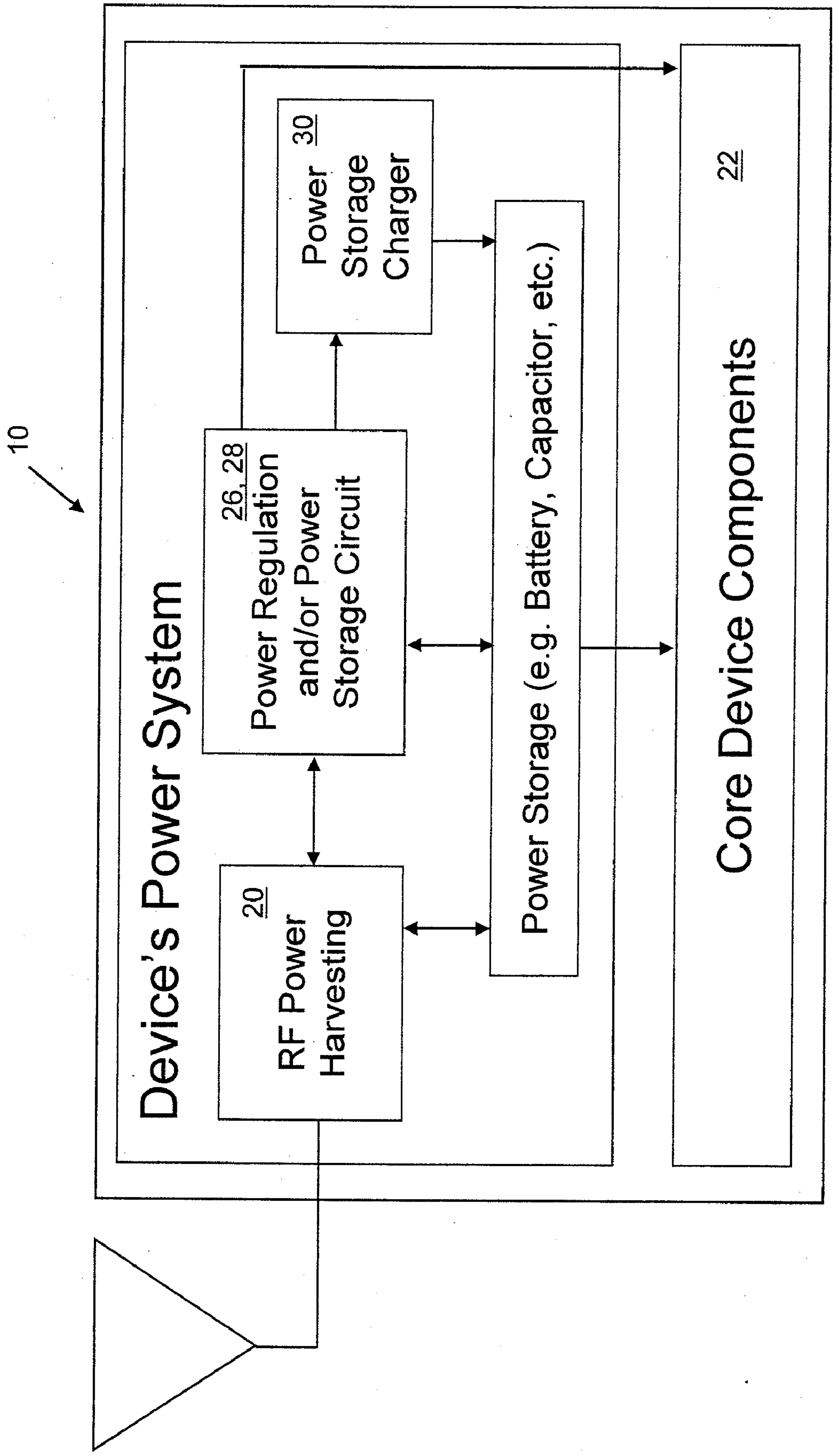
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 34

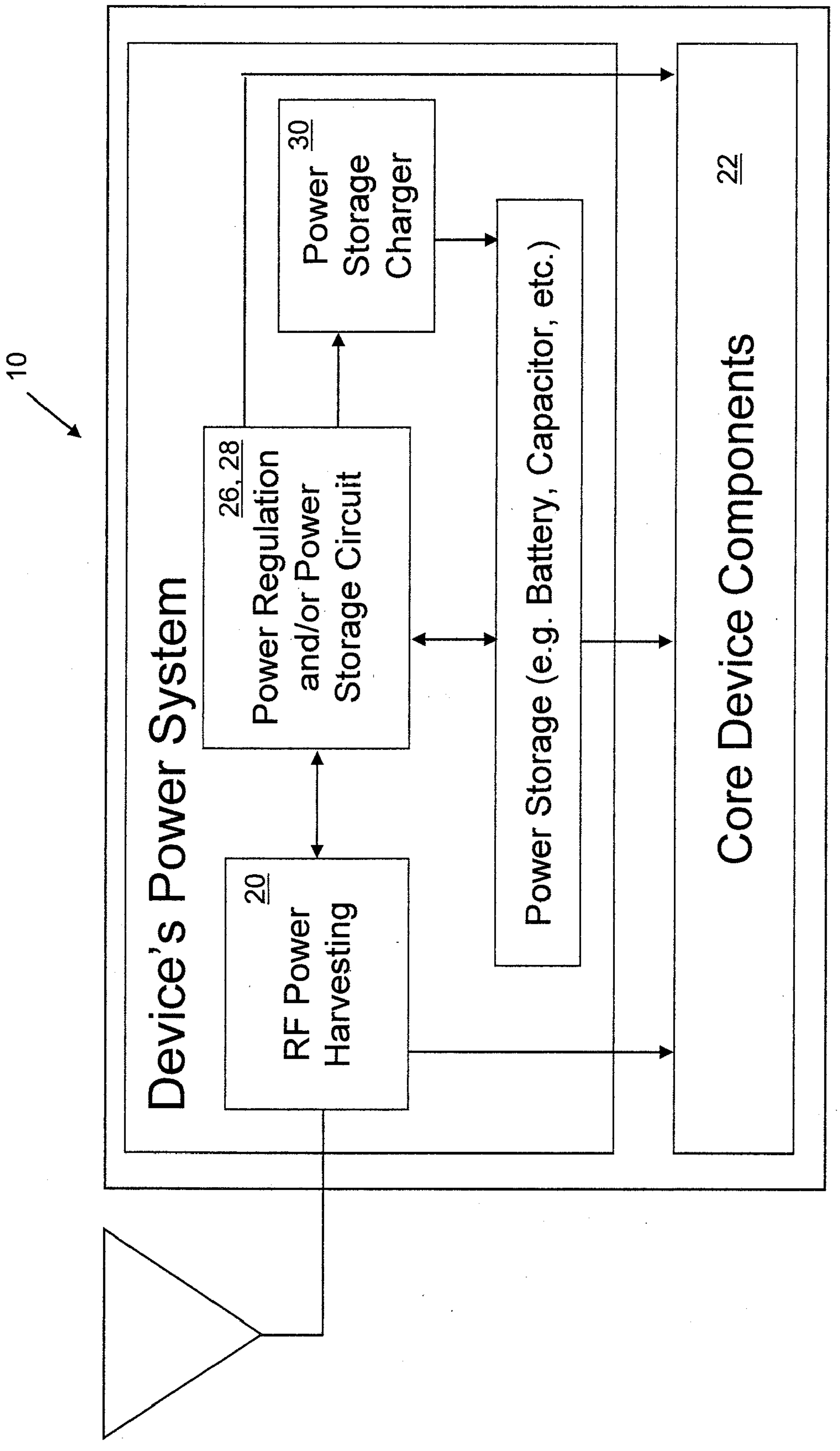
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 35

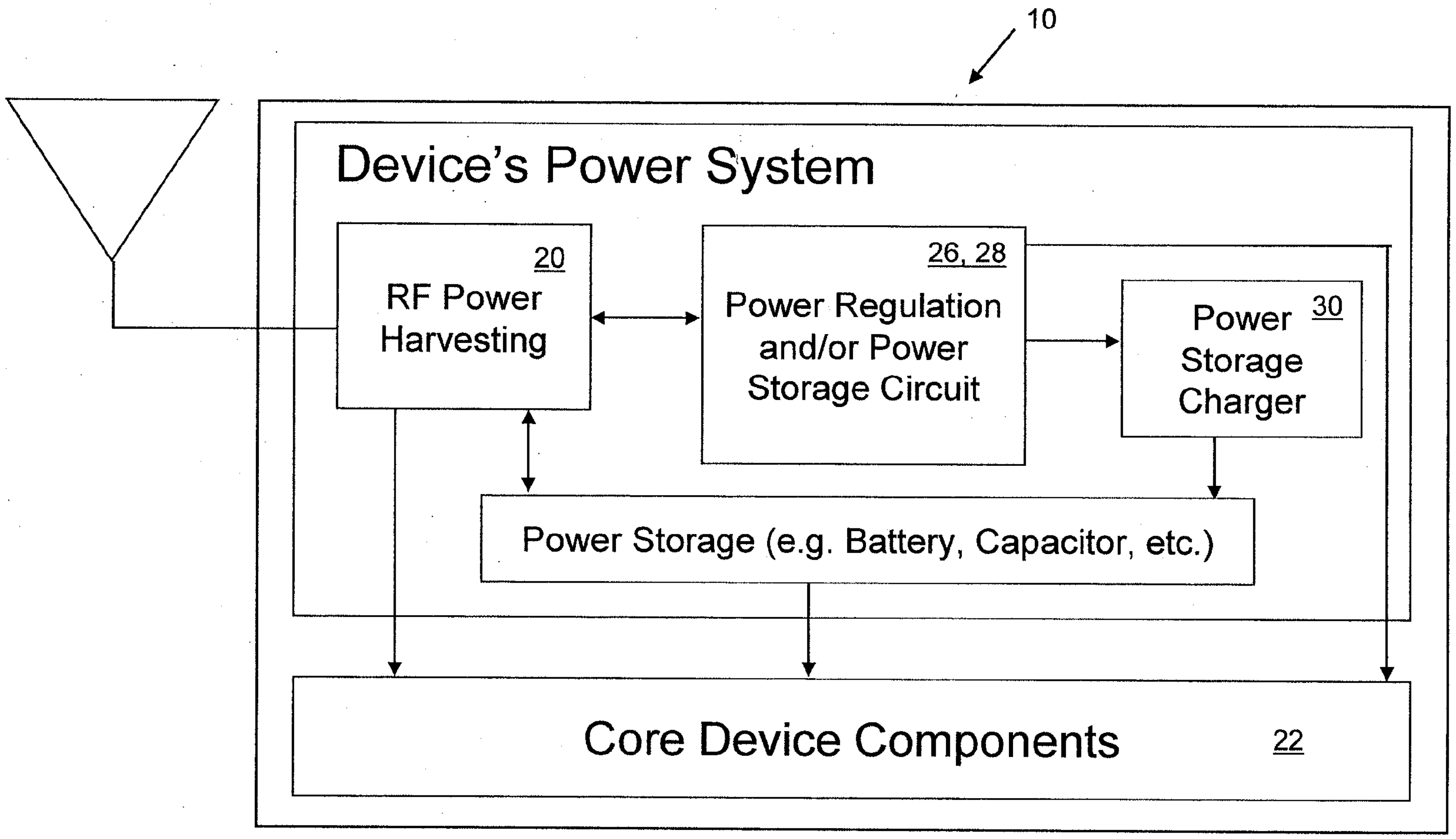
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 36

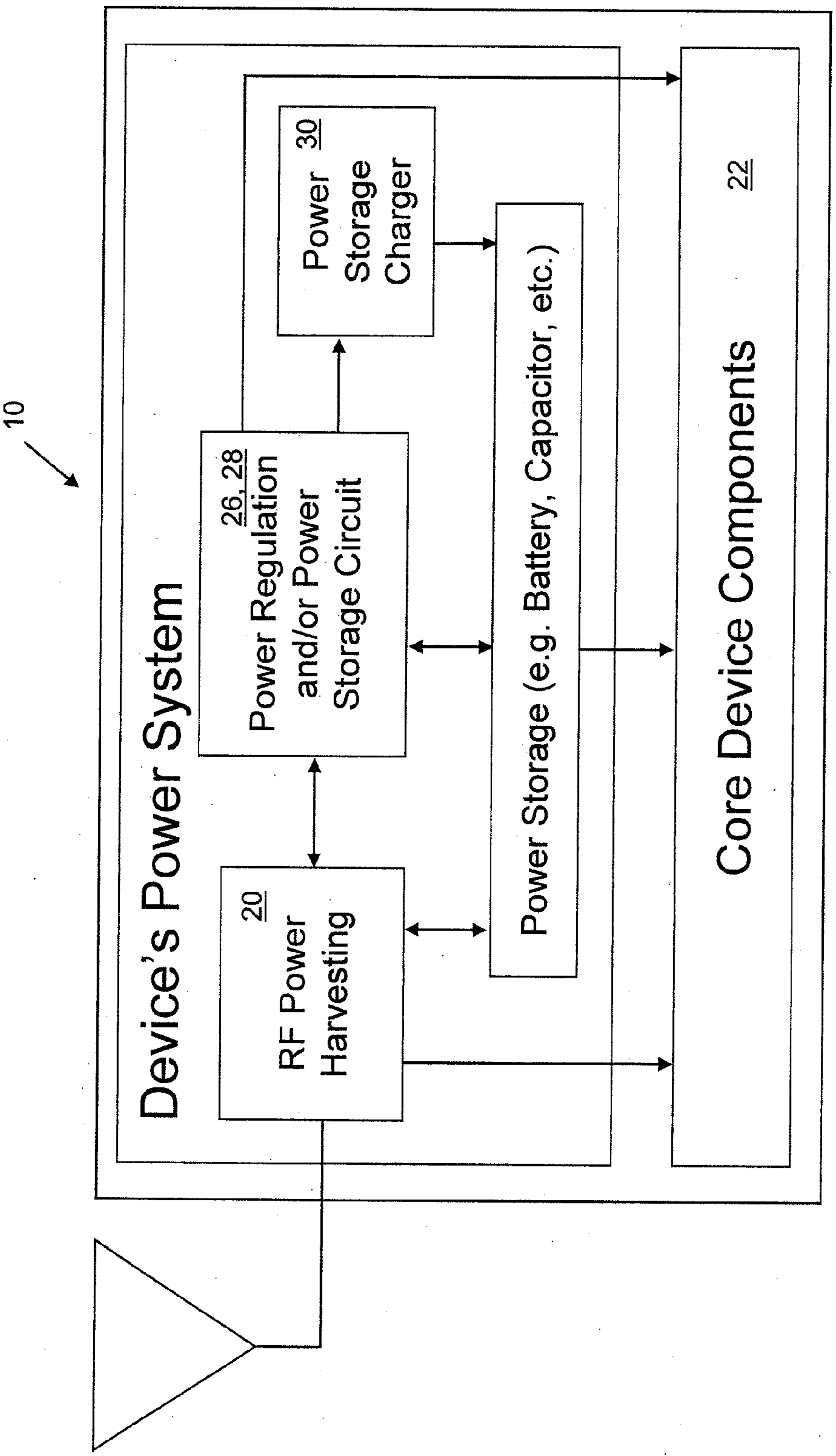
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 37

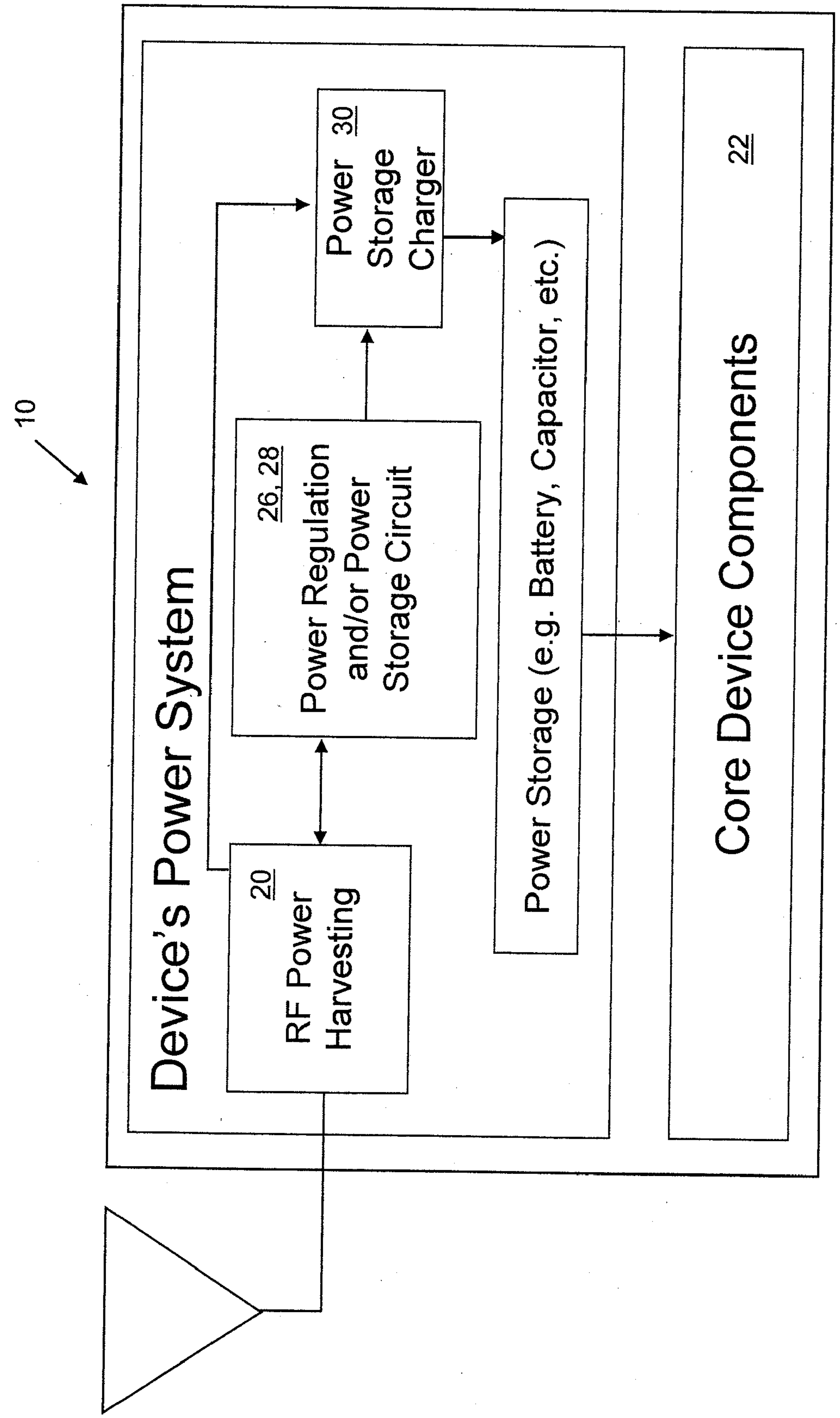
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 38

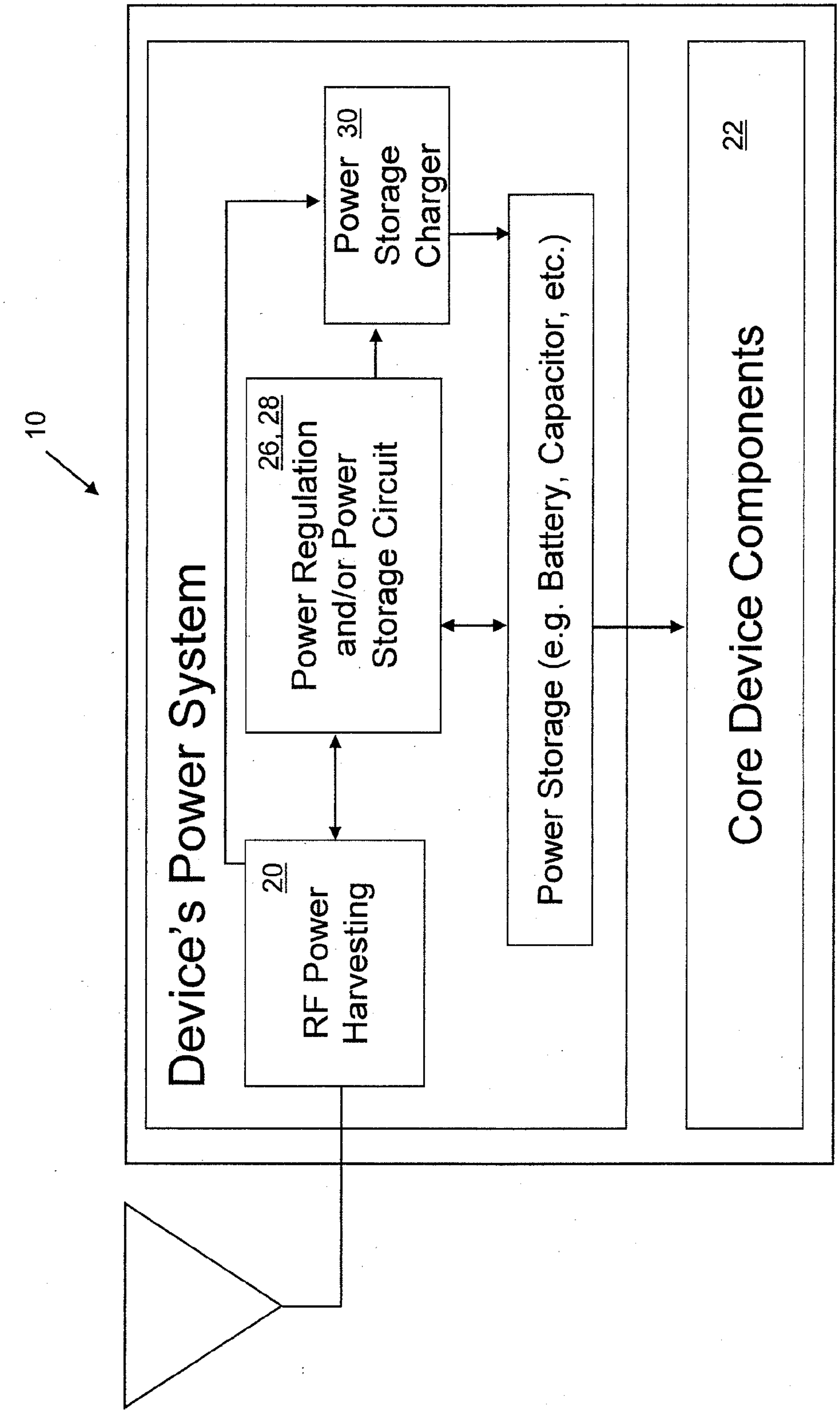
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 39

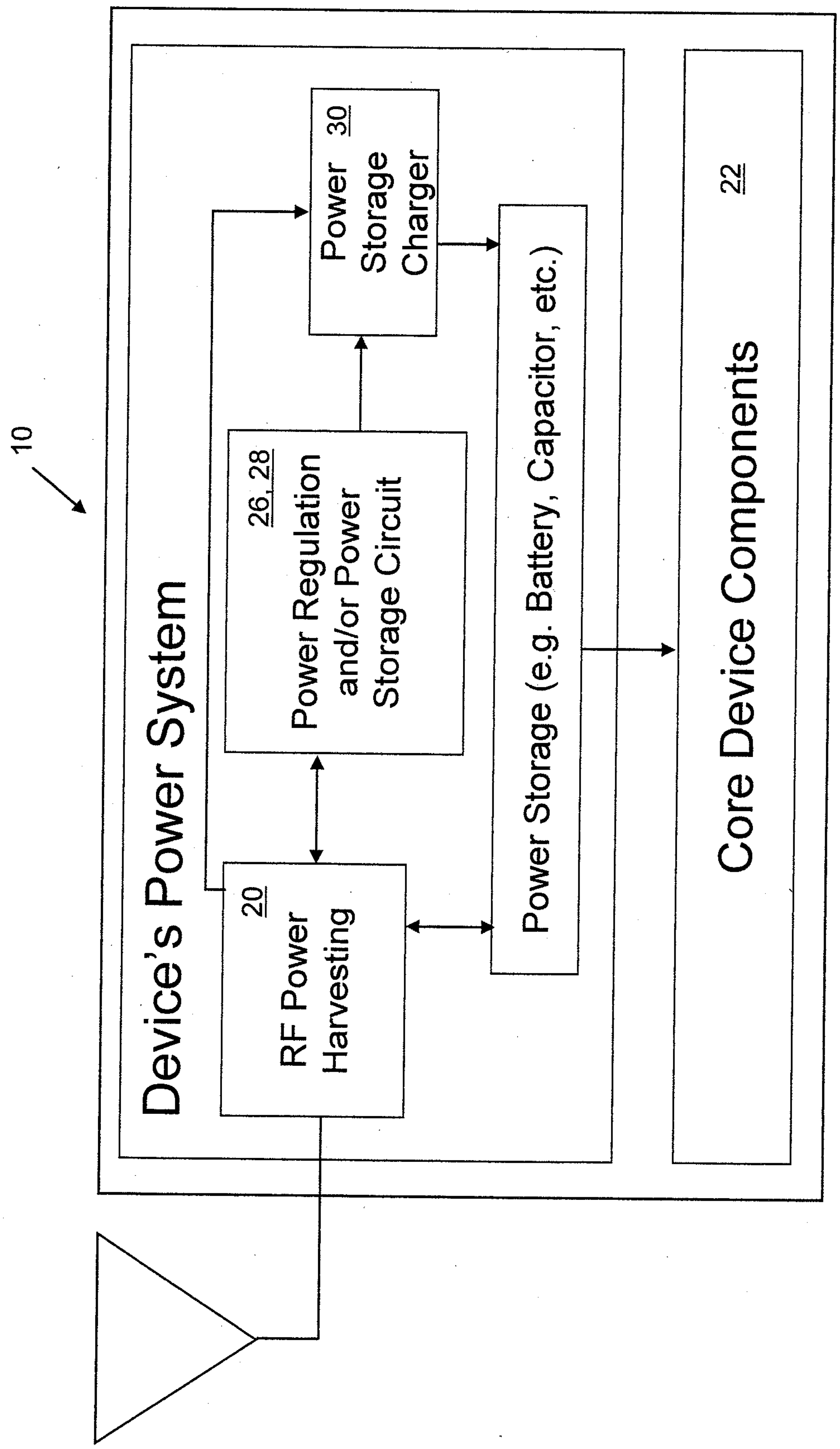
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 40

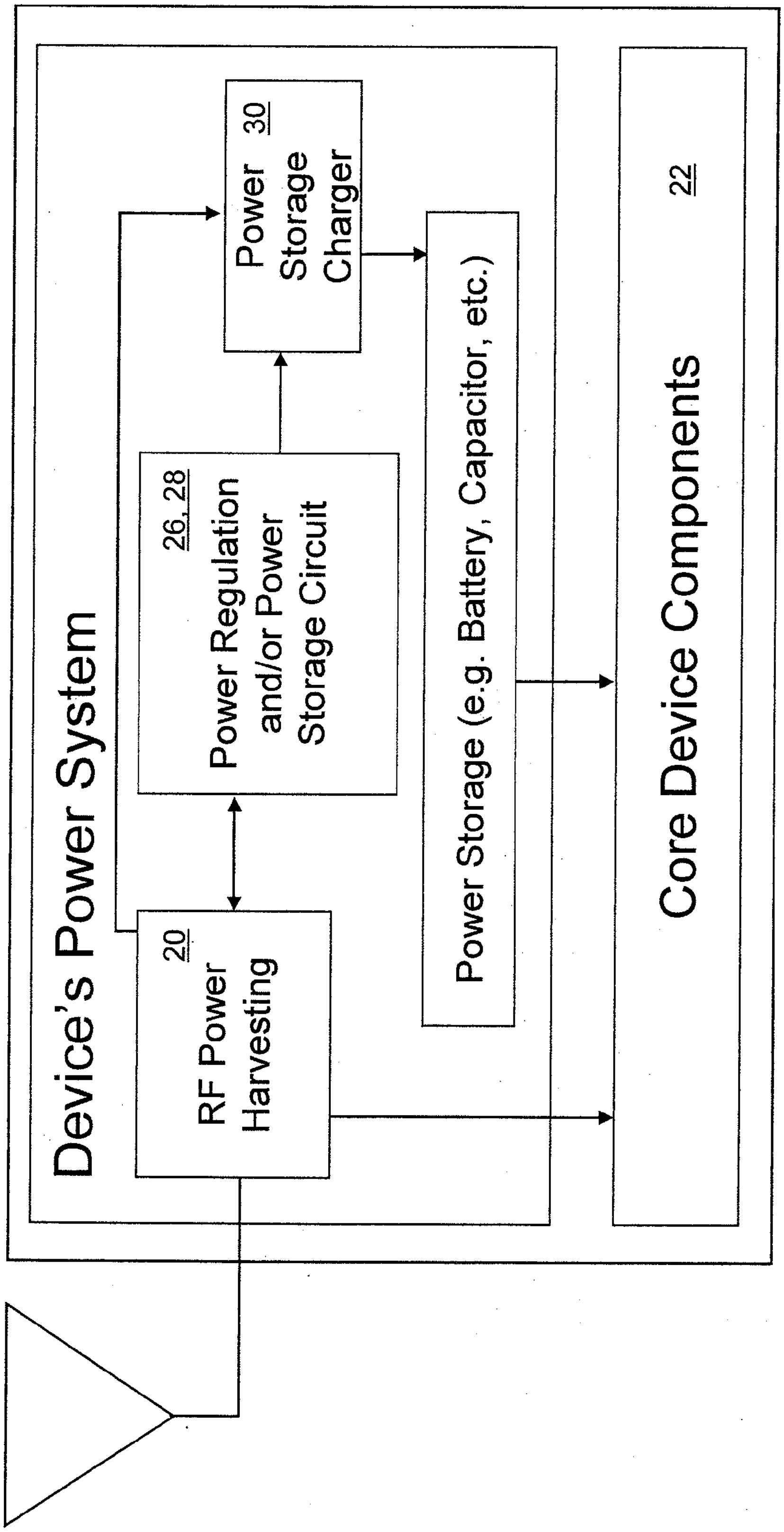
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 41

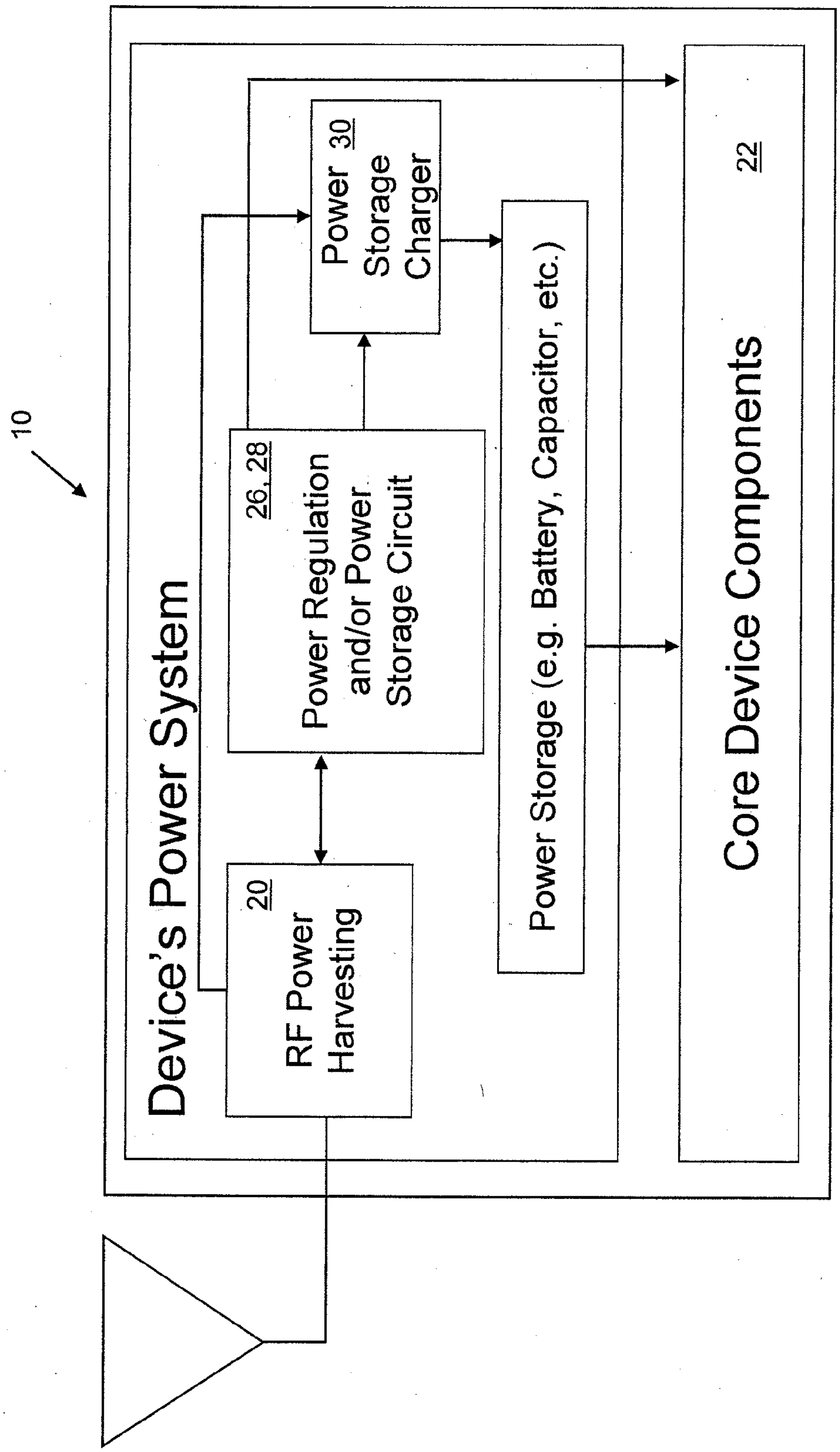
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 42

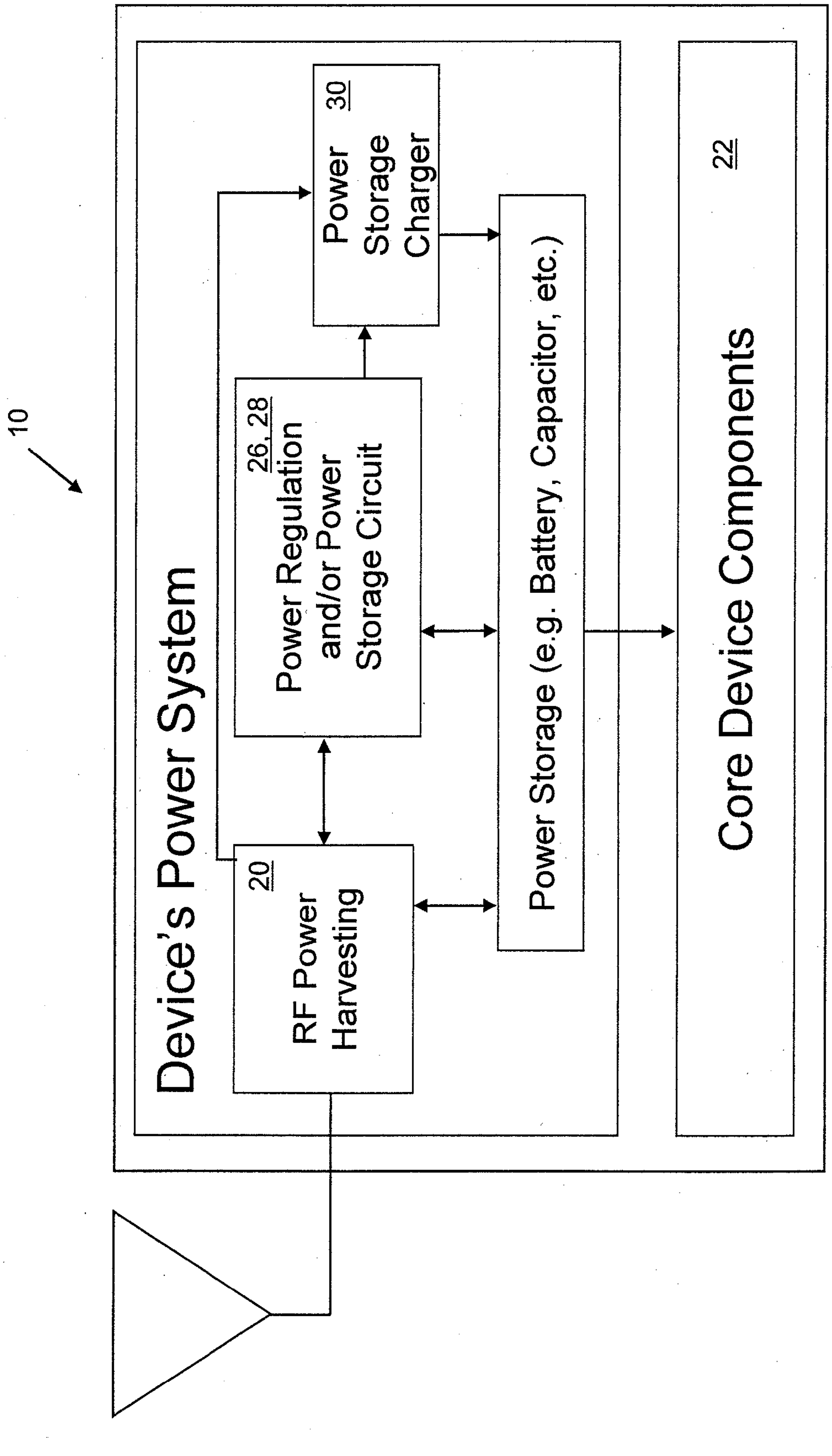
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 43

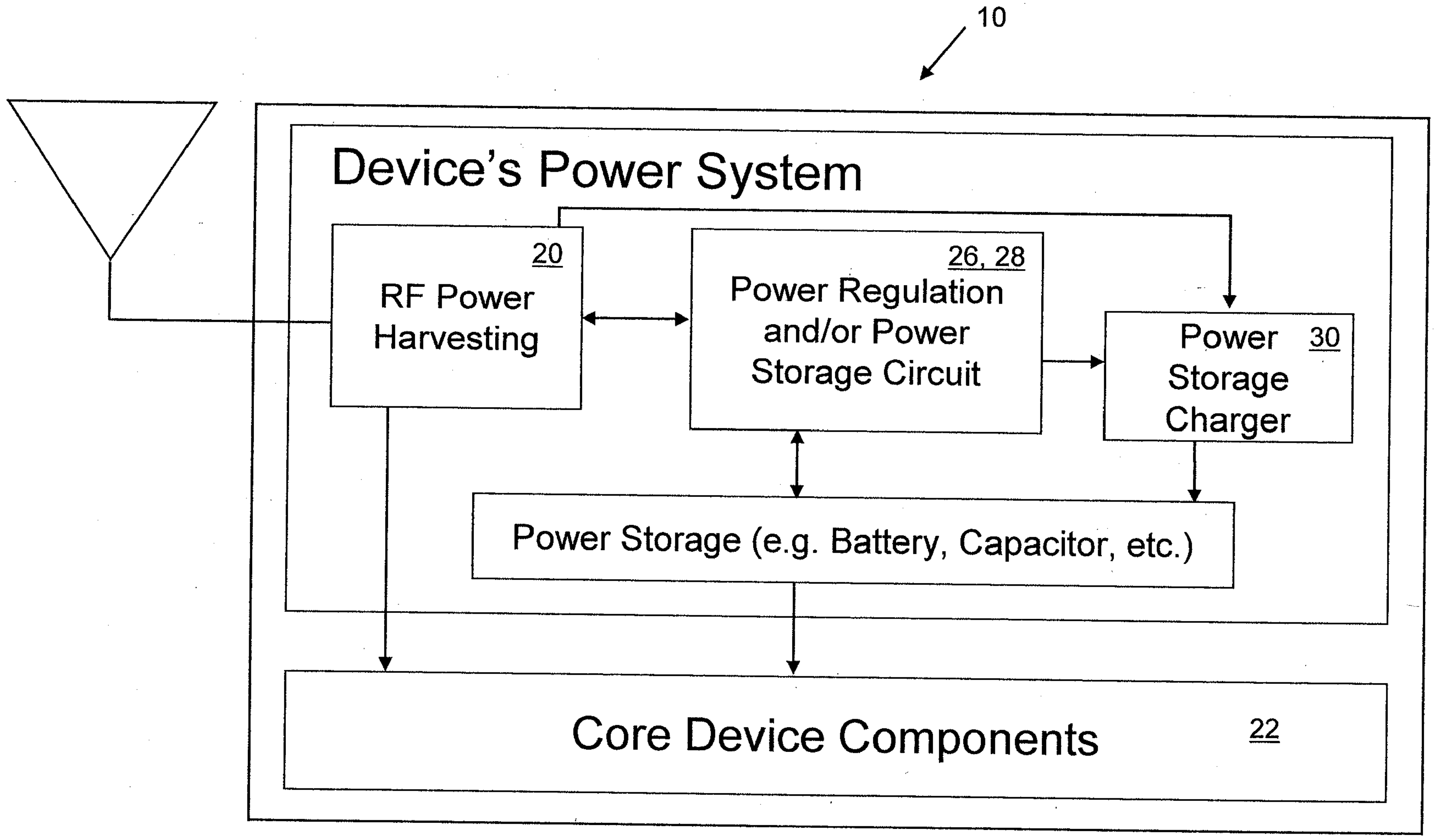
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 44

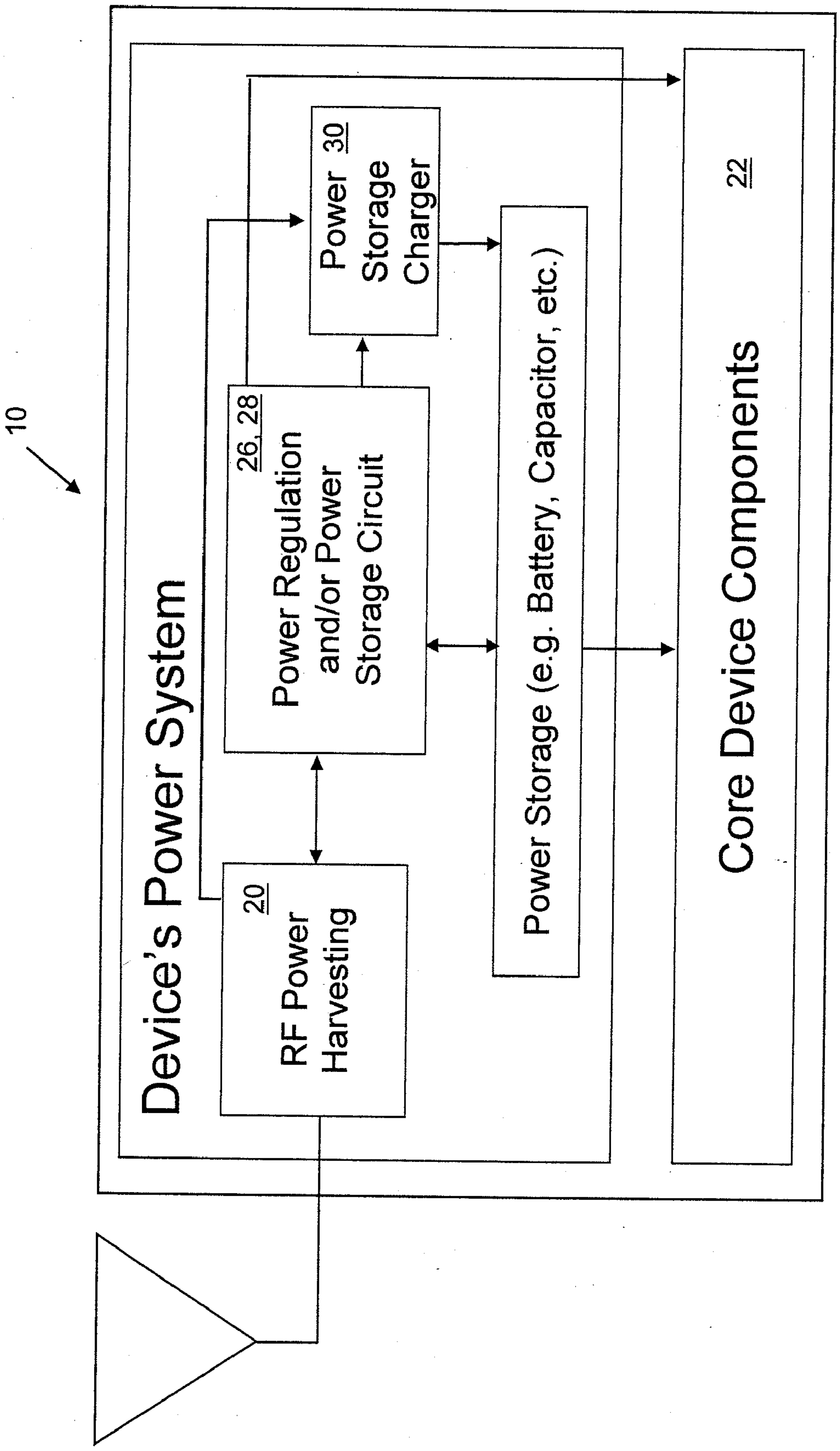
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 45

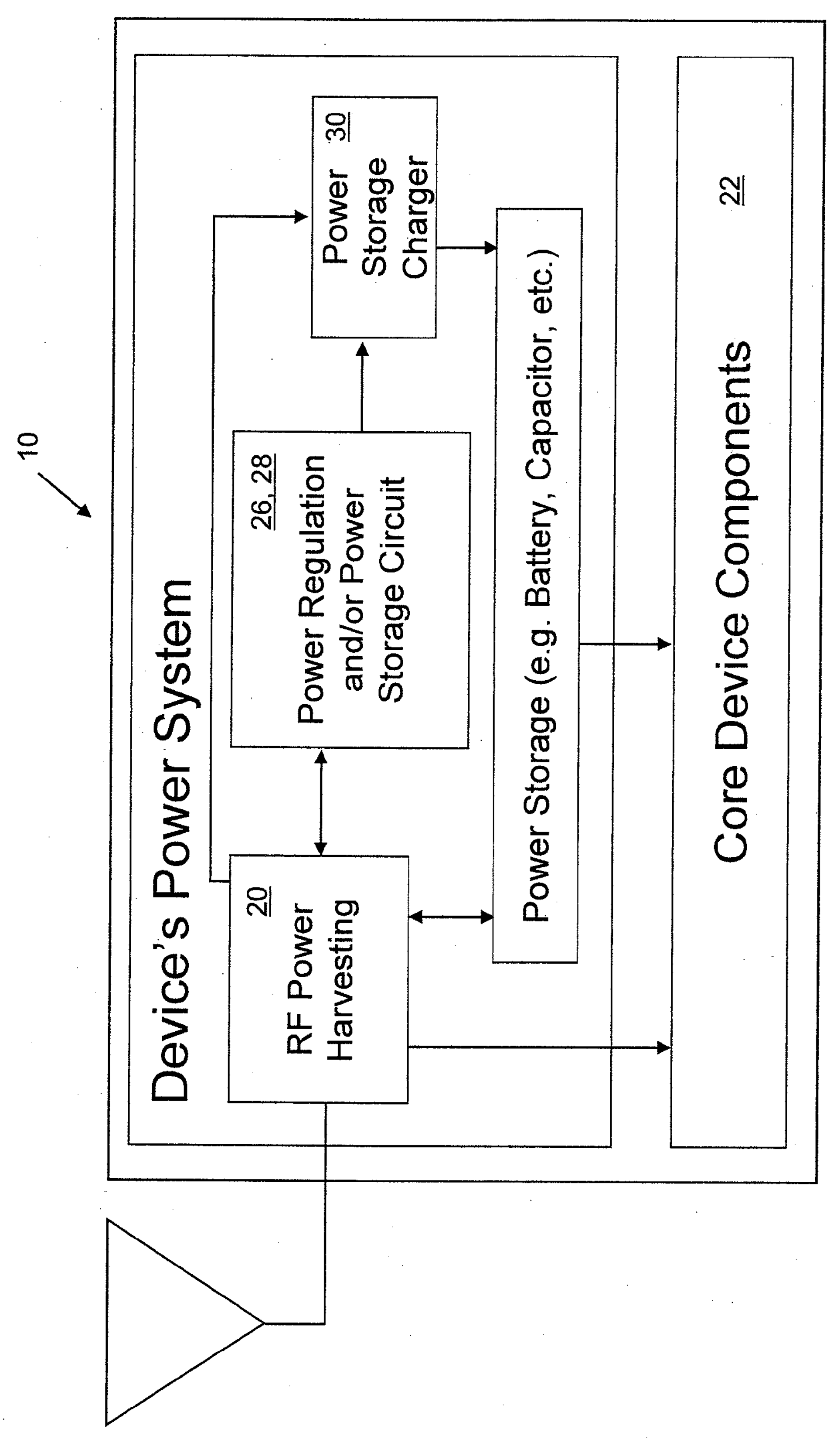
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 46

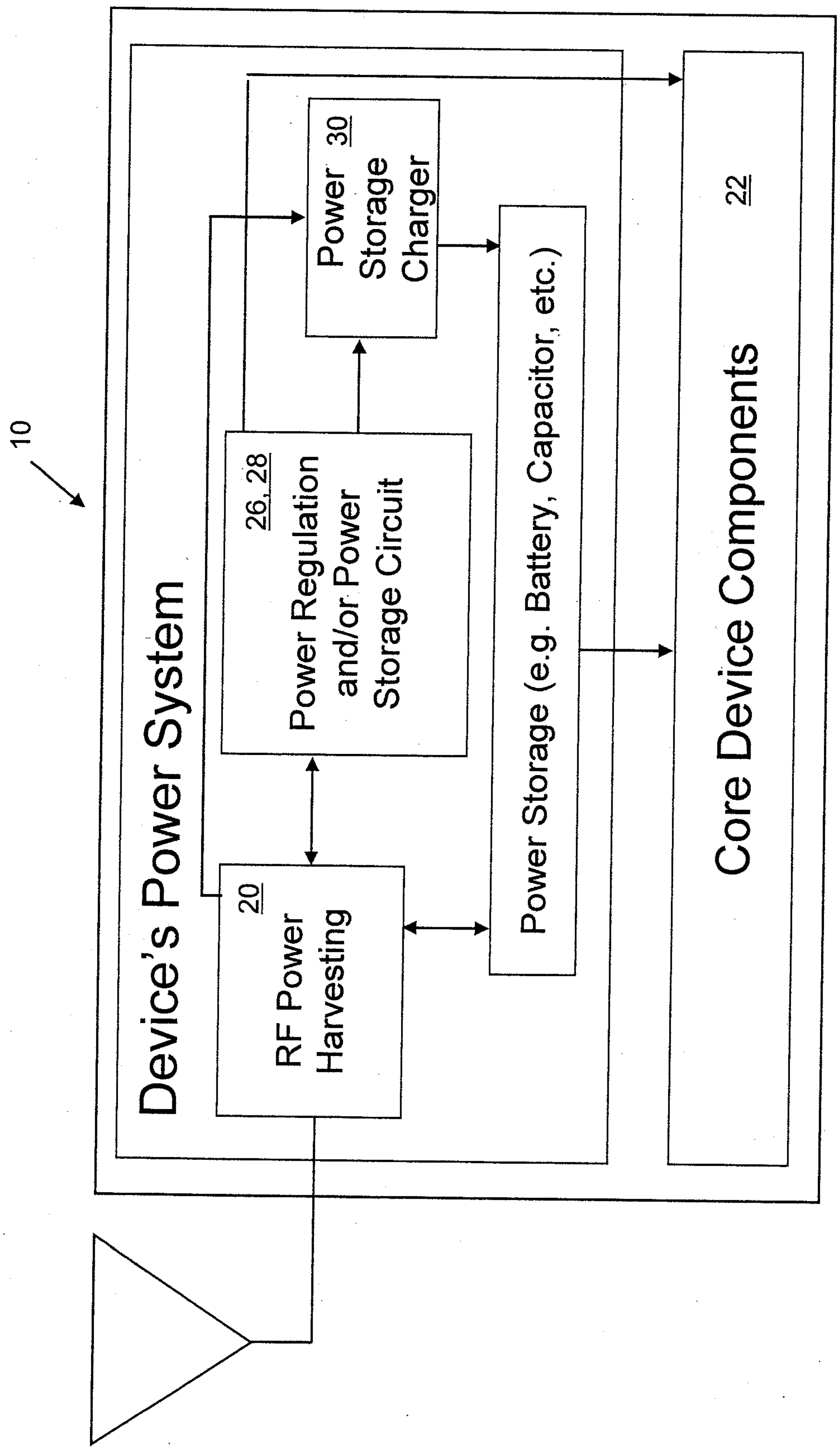
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 47

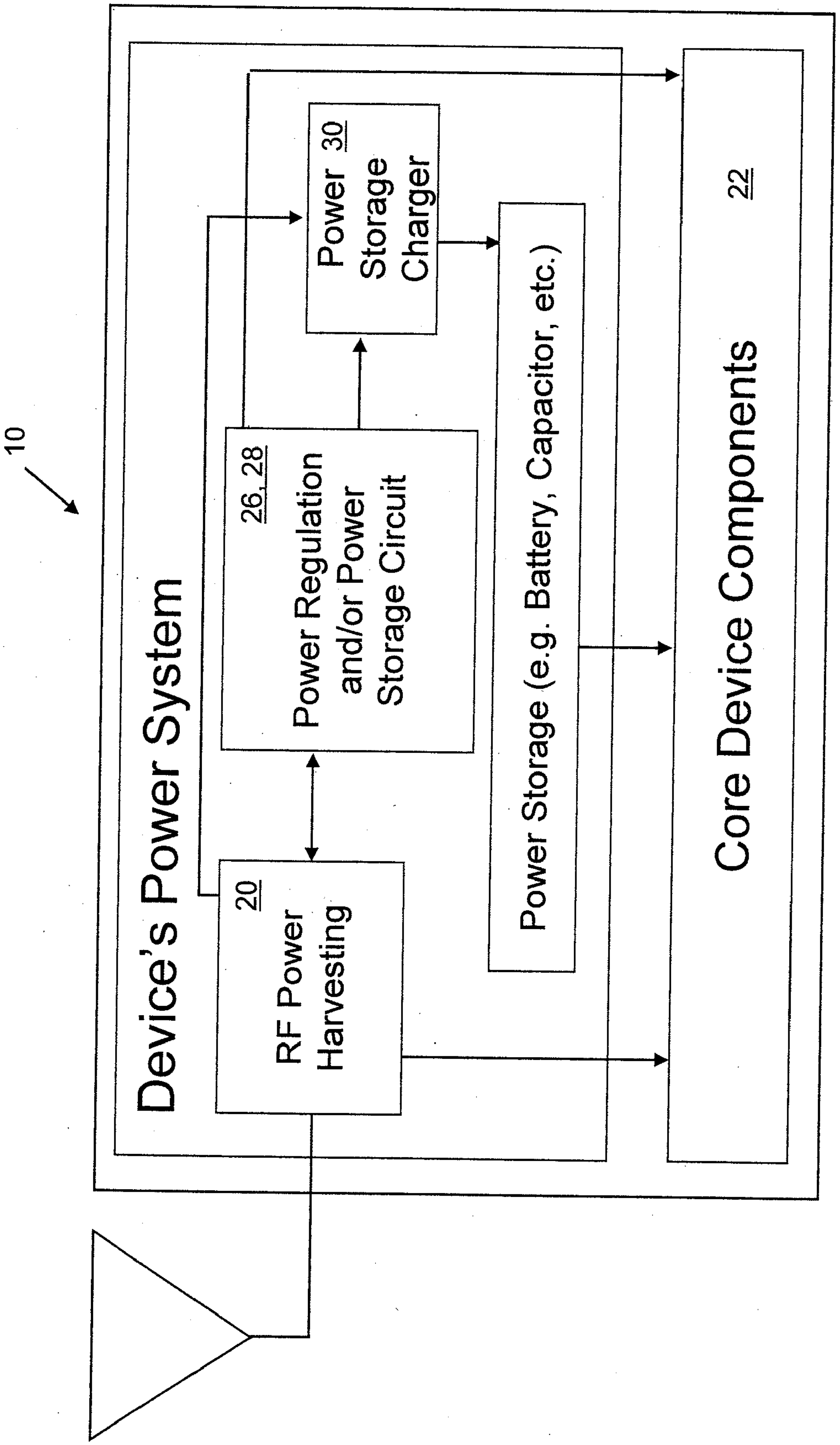
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30 Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 48

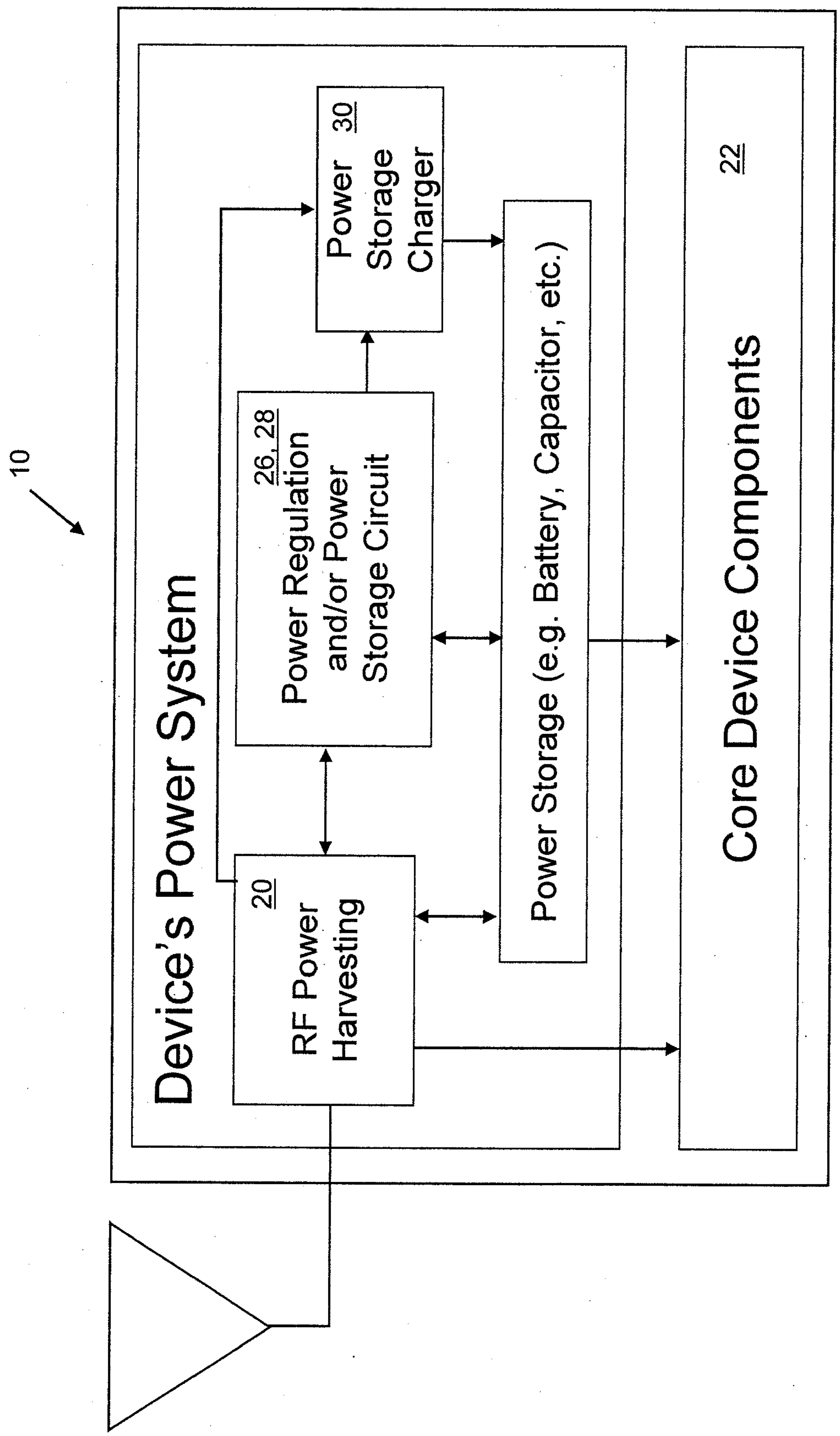
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 49

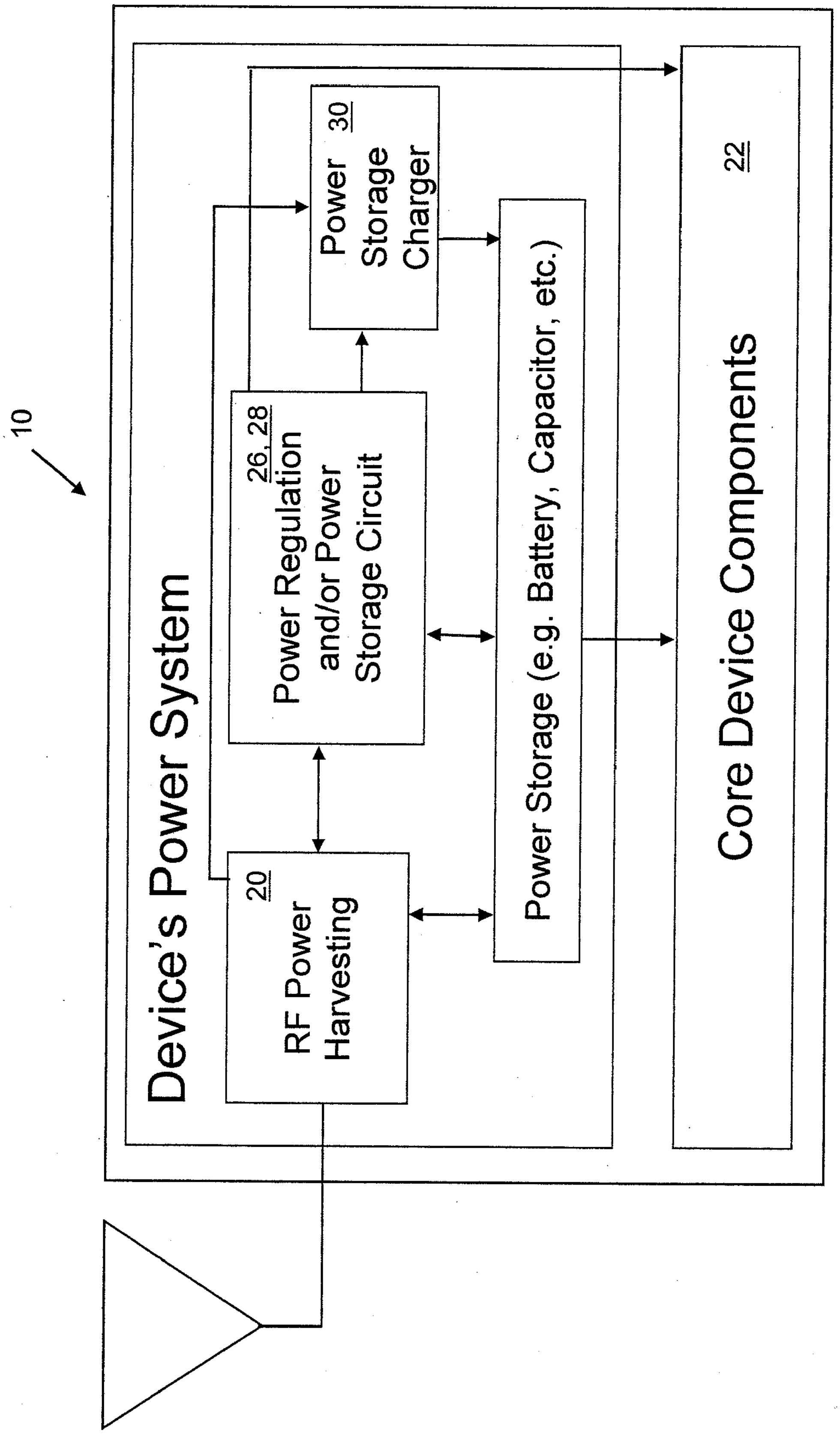
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 50

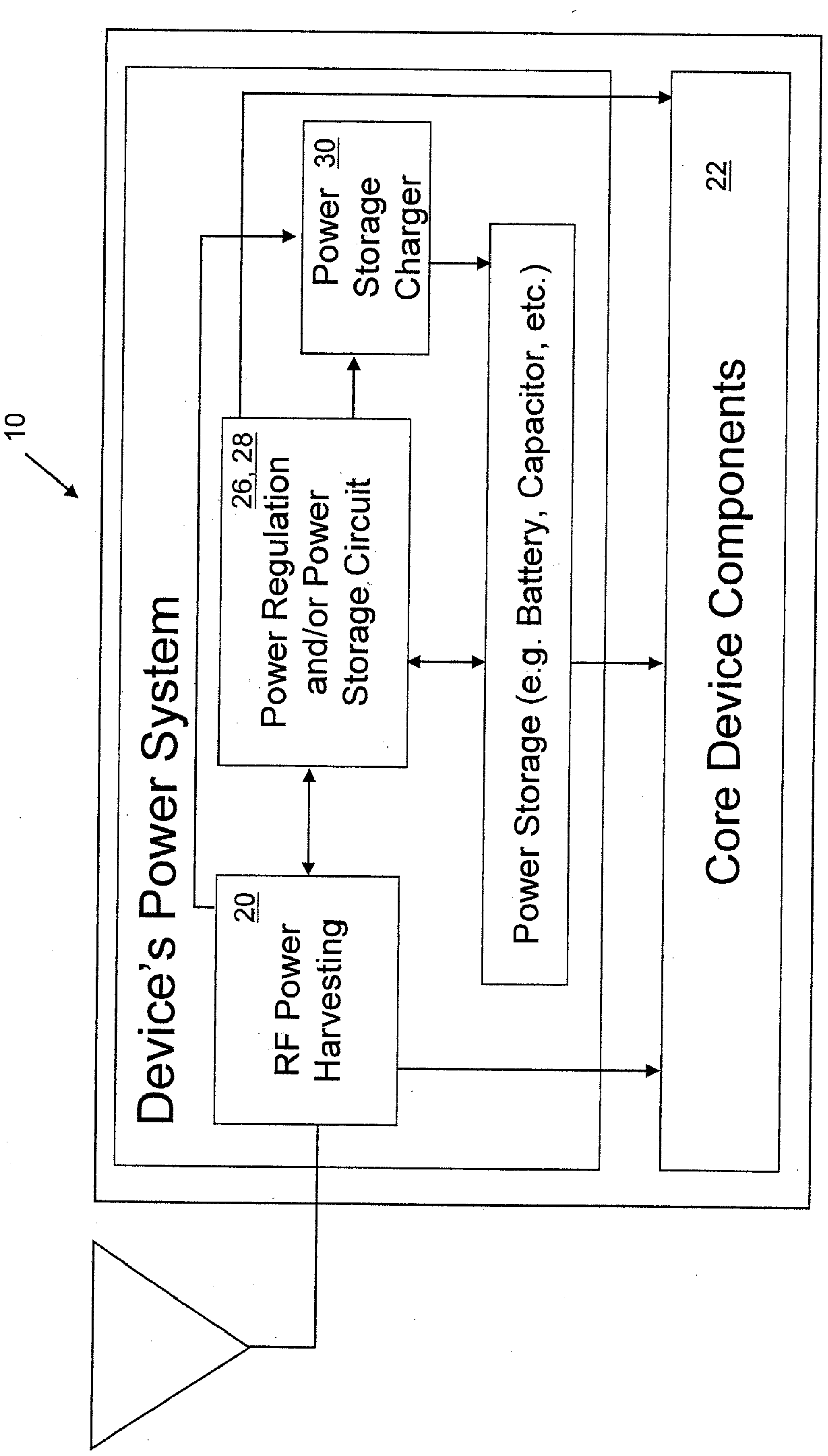
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 51

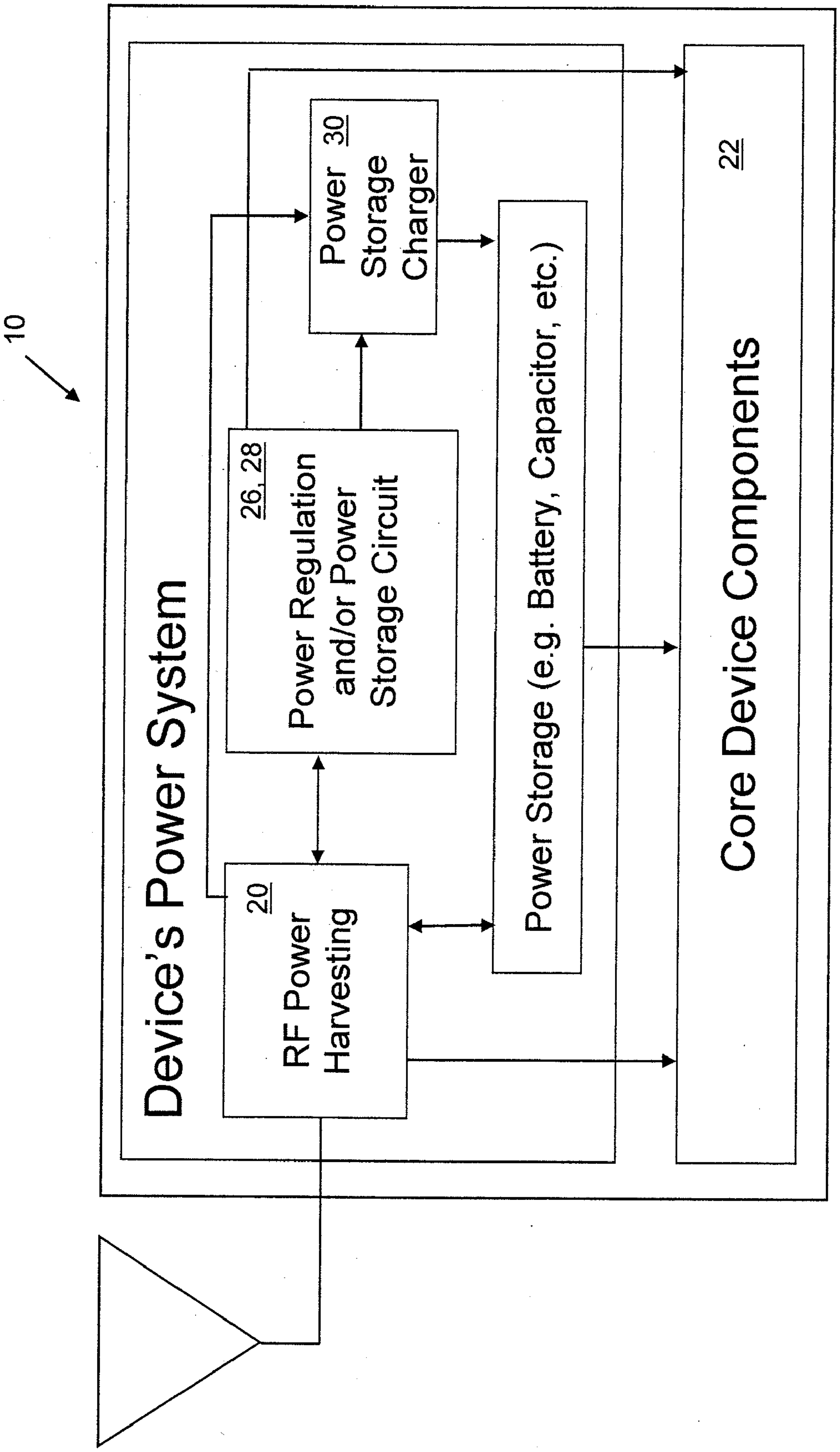
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage
Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 52

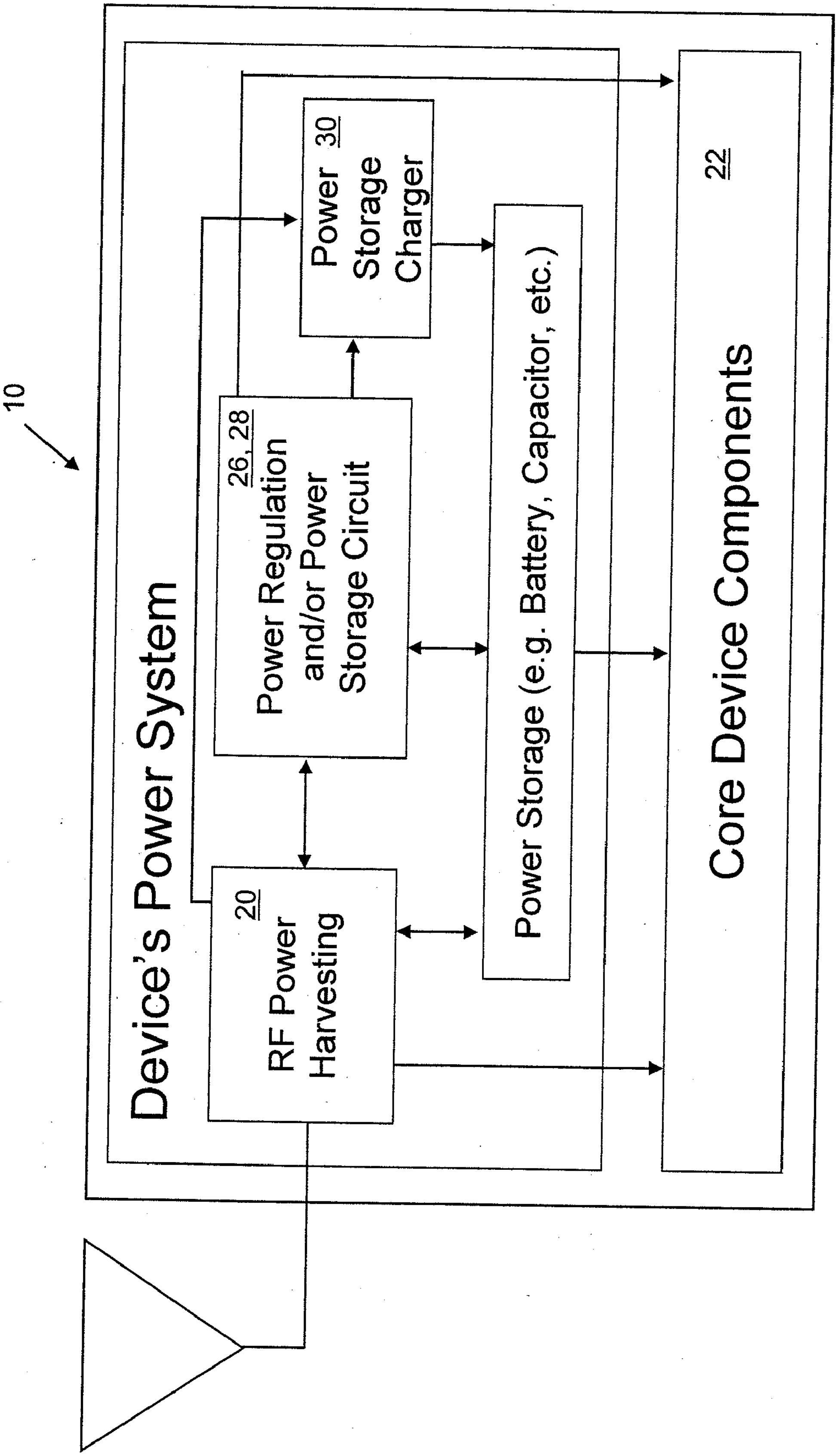
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 53

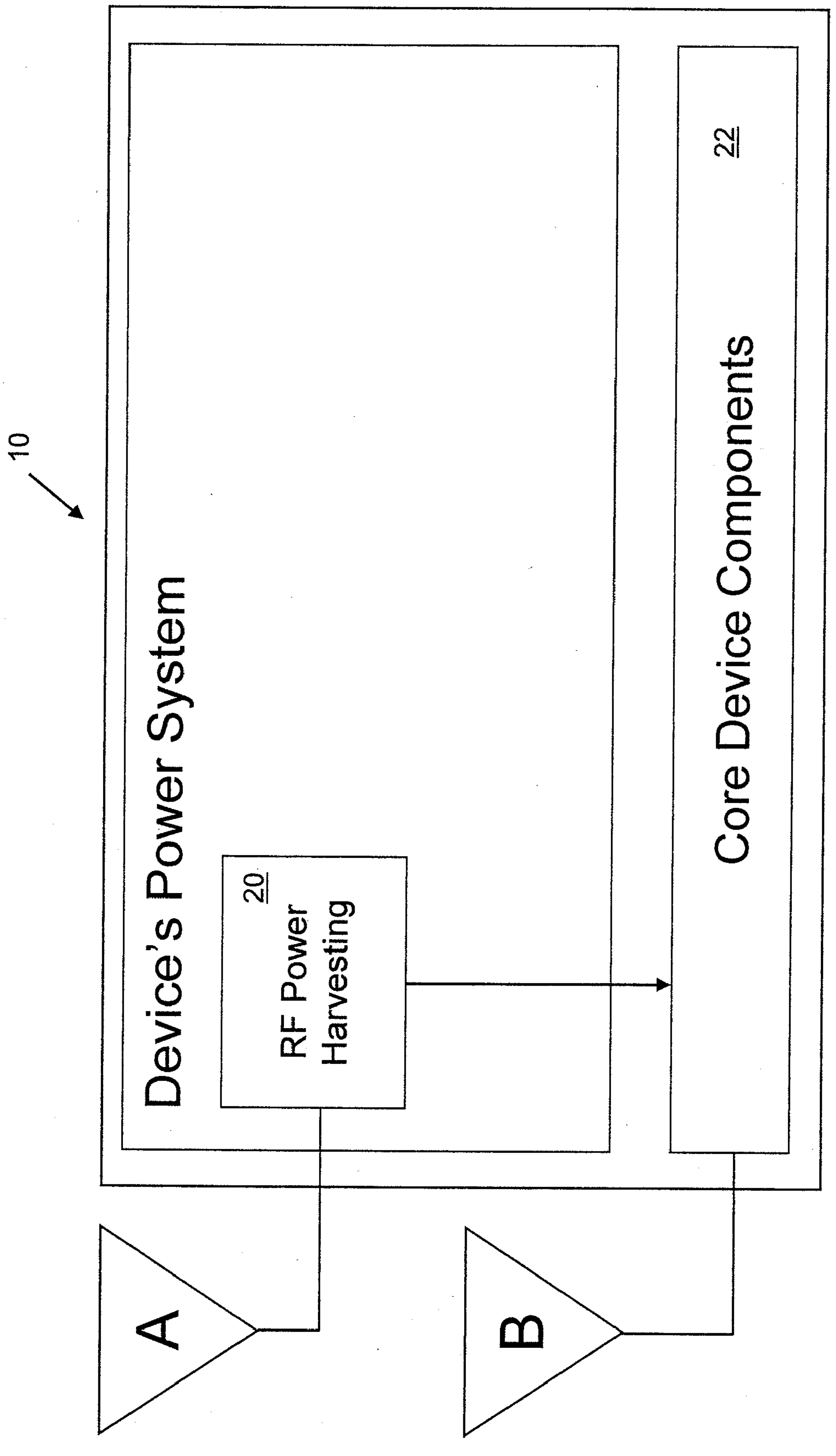
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
A
B

Fig. 54

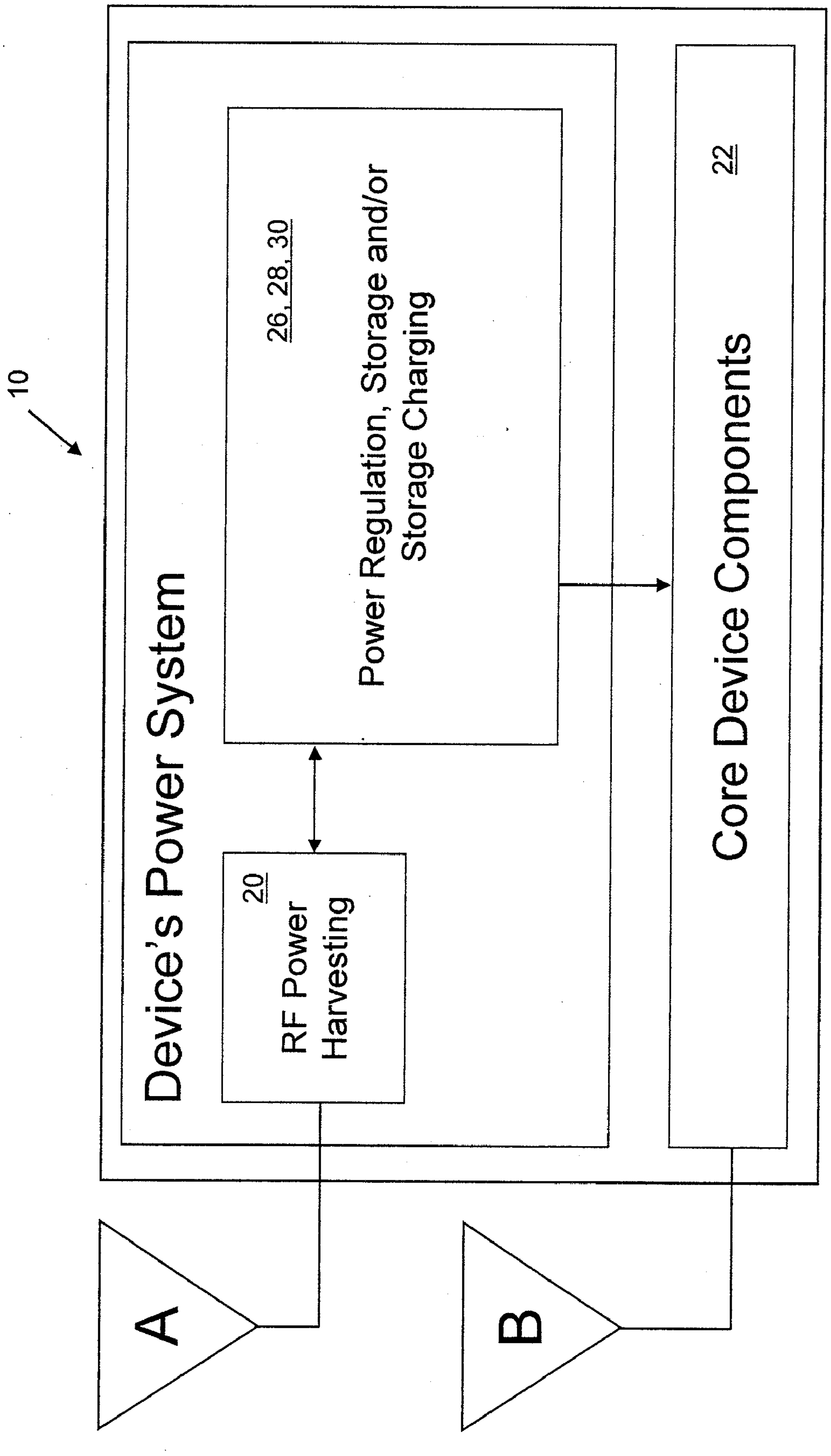
10
Device's Power System
RF Power Harvesting
20
Power Regulation, Storage and/or Storage Charging
26, 28, 30
Core Device Components
22
A
B

Fig. 55

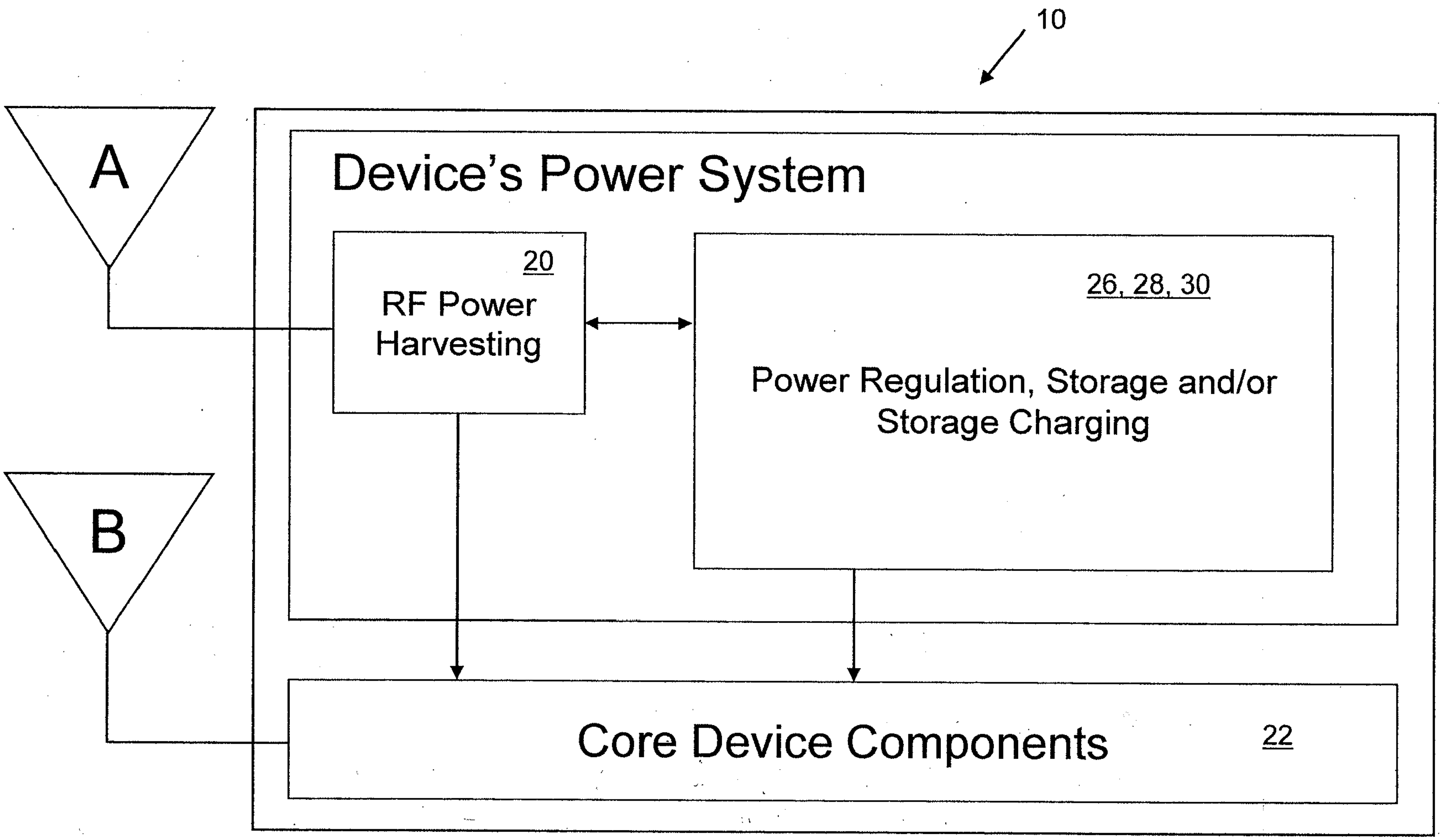
10
A
B
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22

Fig. 56

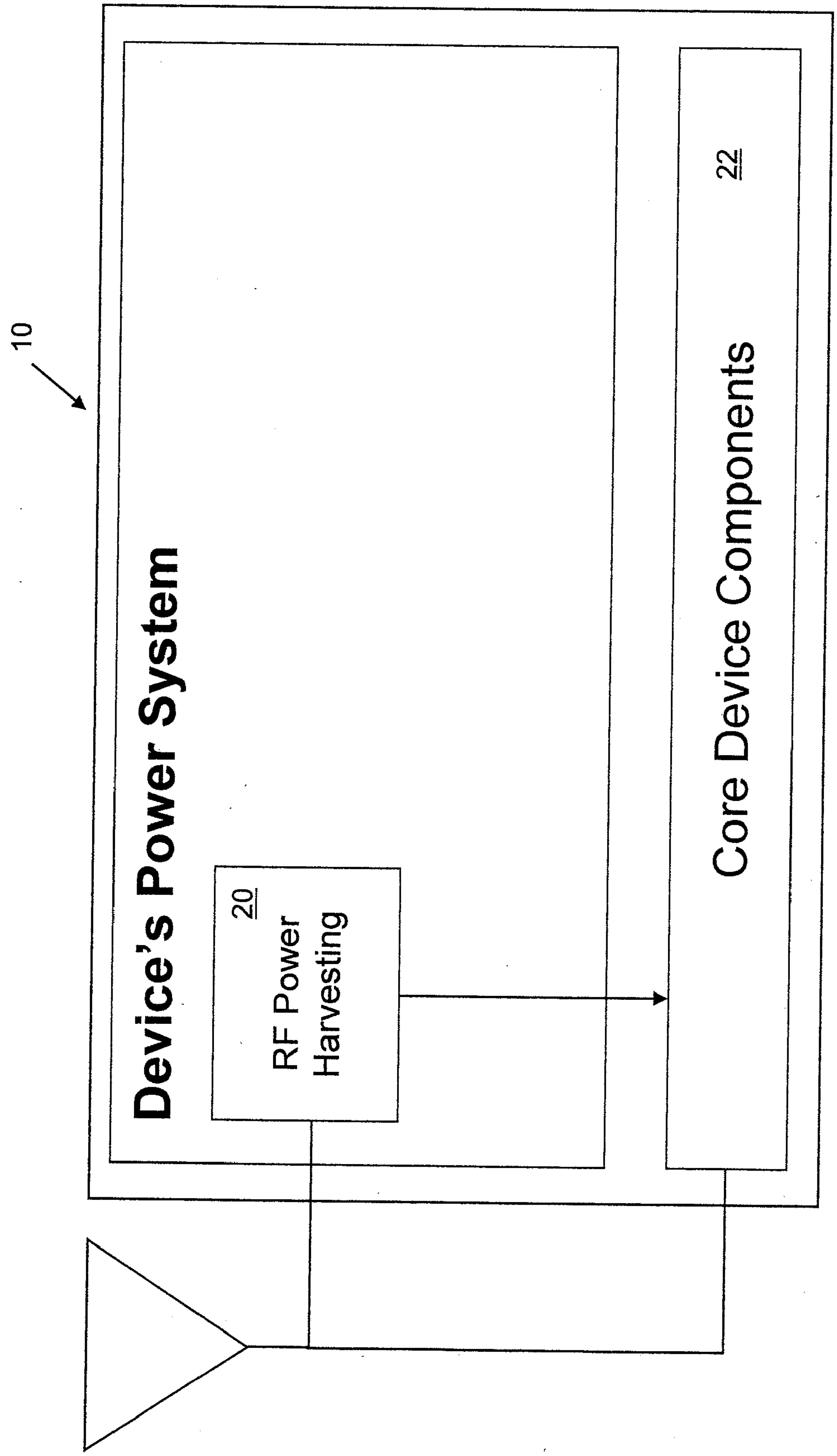
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22

Fig. 57

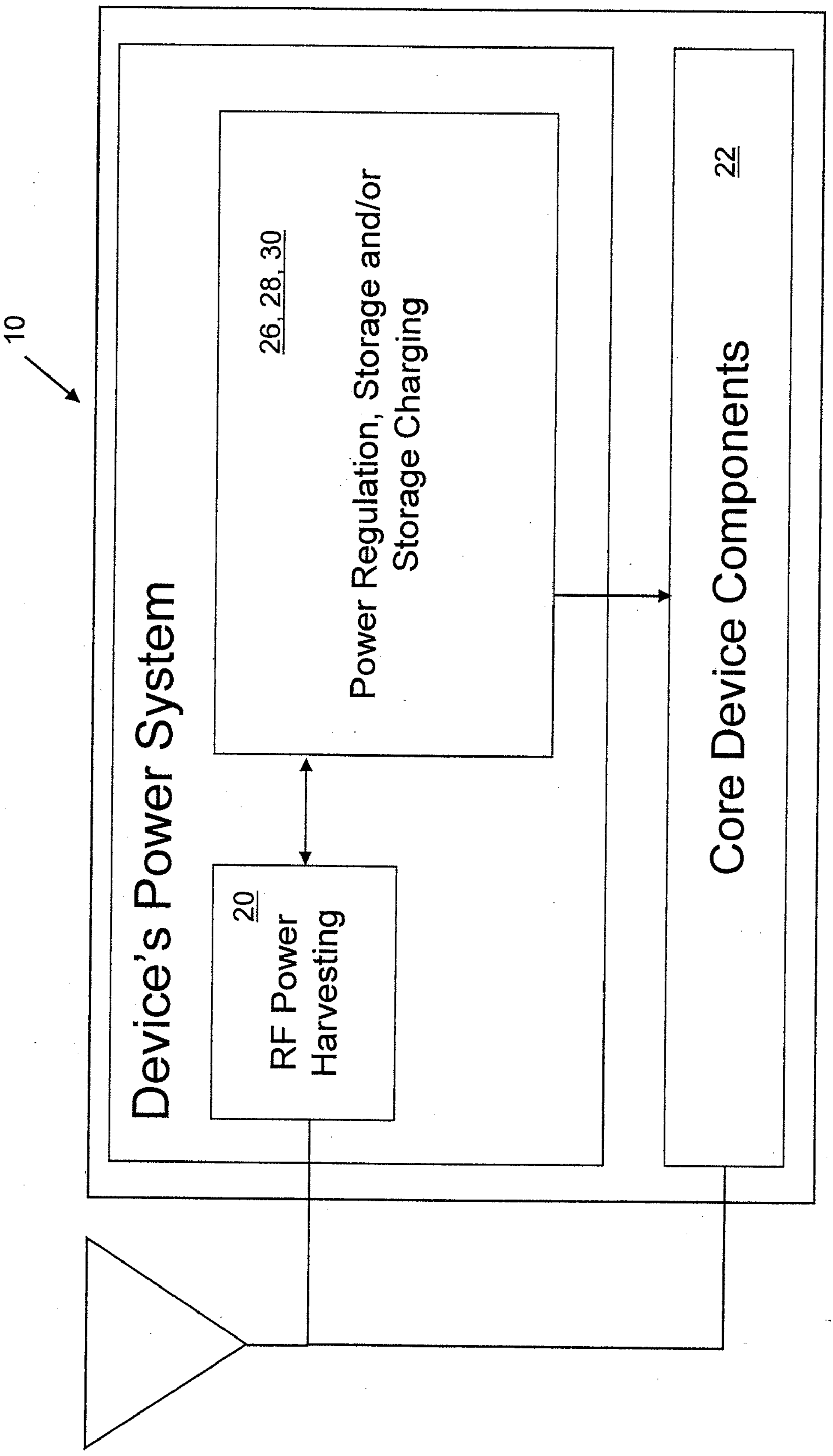
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22

Fig. 58

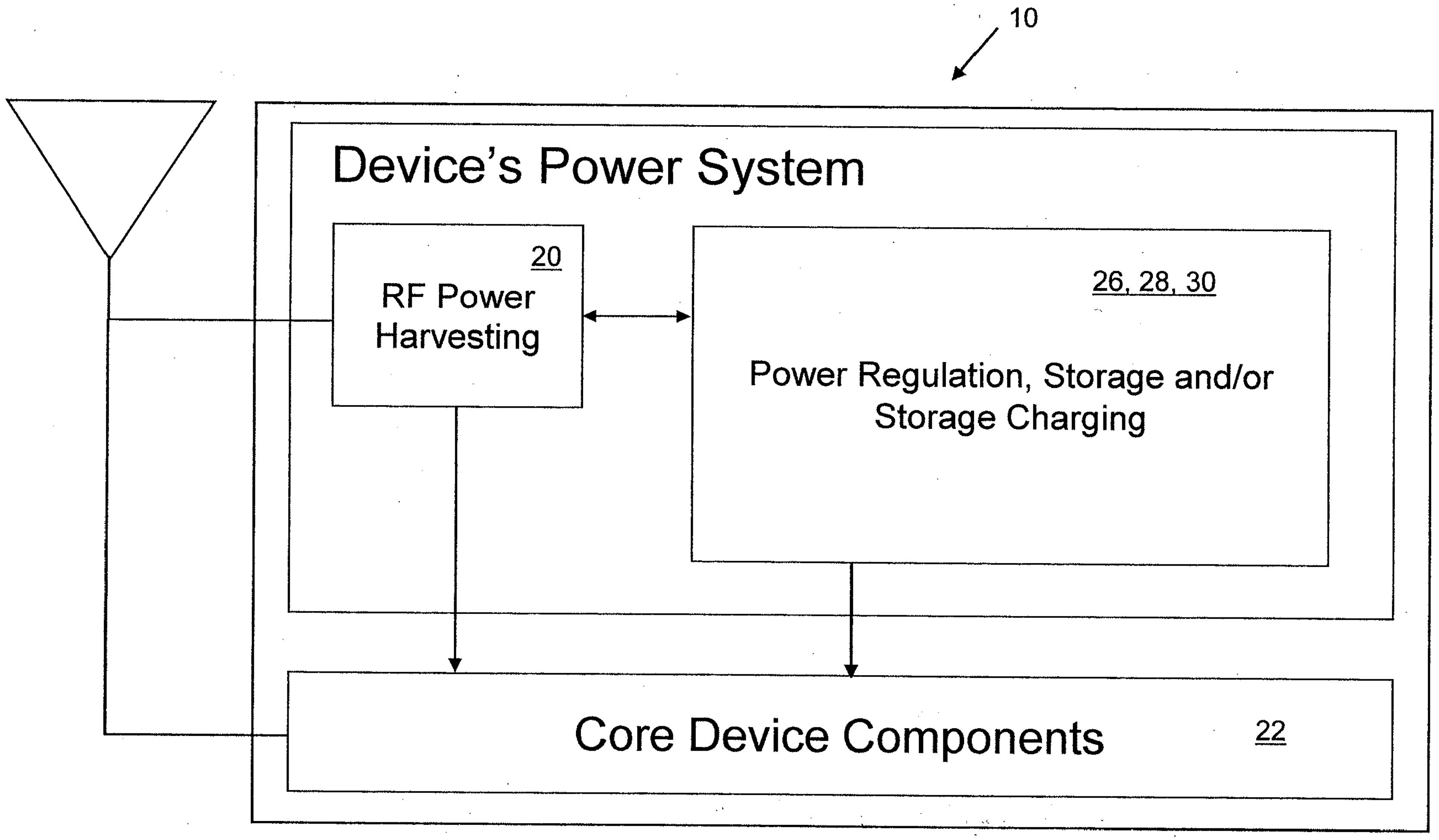
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22

Fig. 59

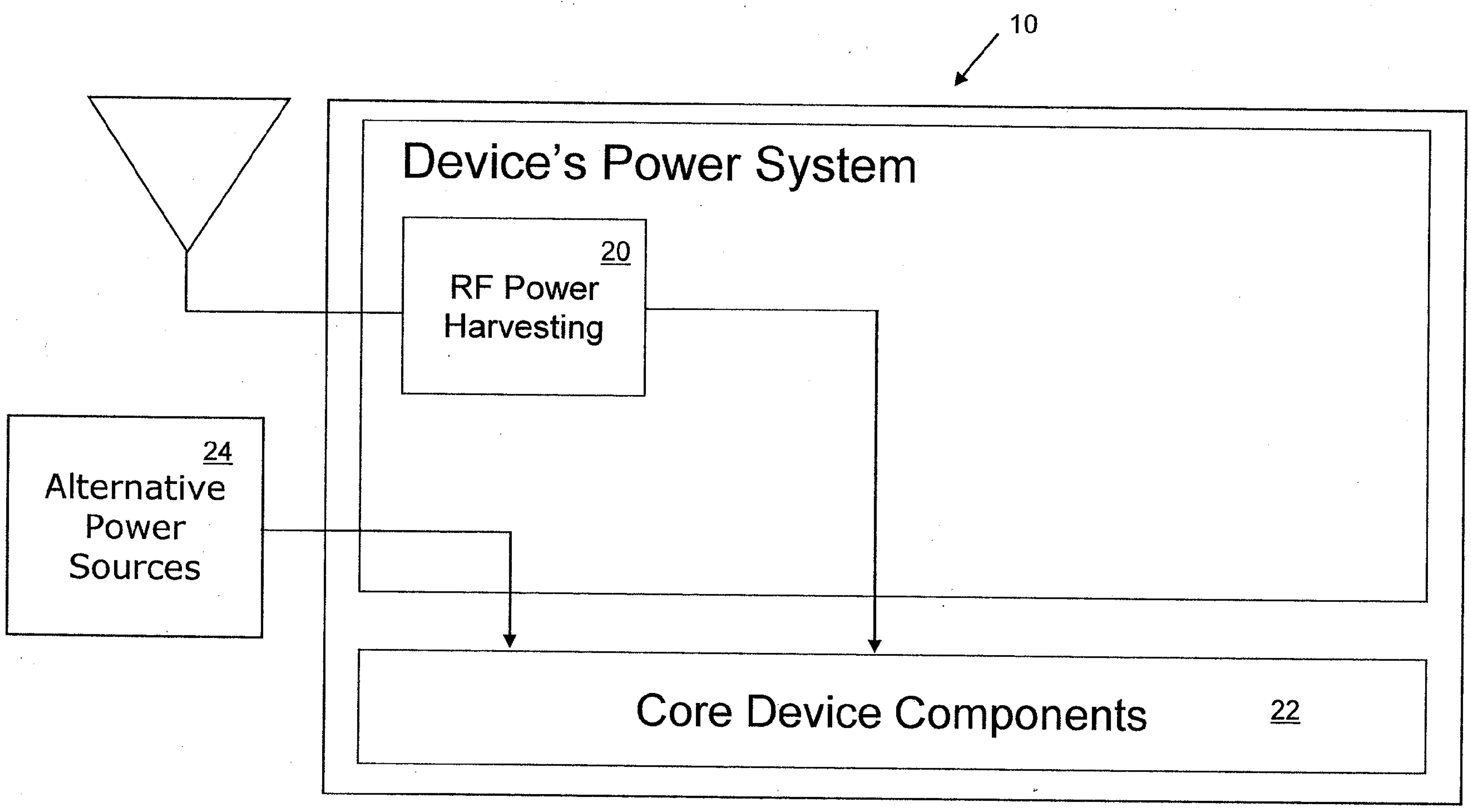
10
Device's Power System
20
RF Power
Harvesting
24
Alternative
Power
Sources
Core Device Components
22

Fig. 60

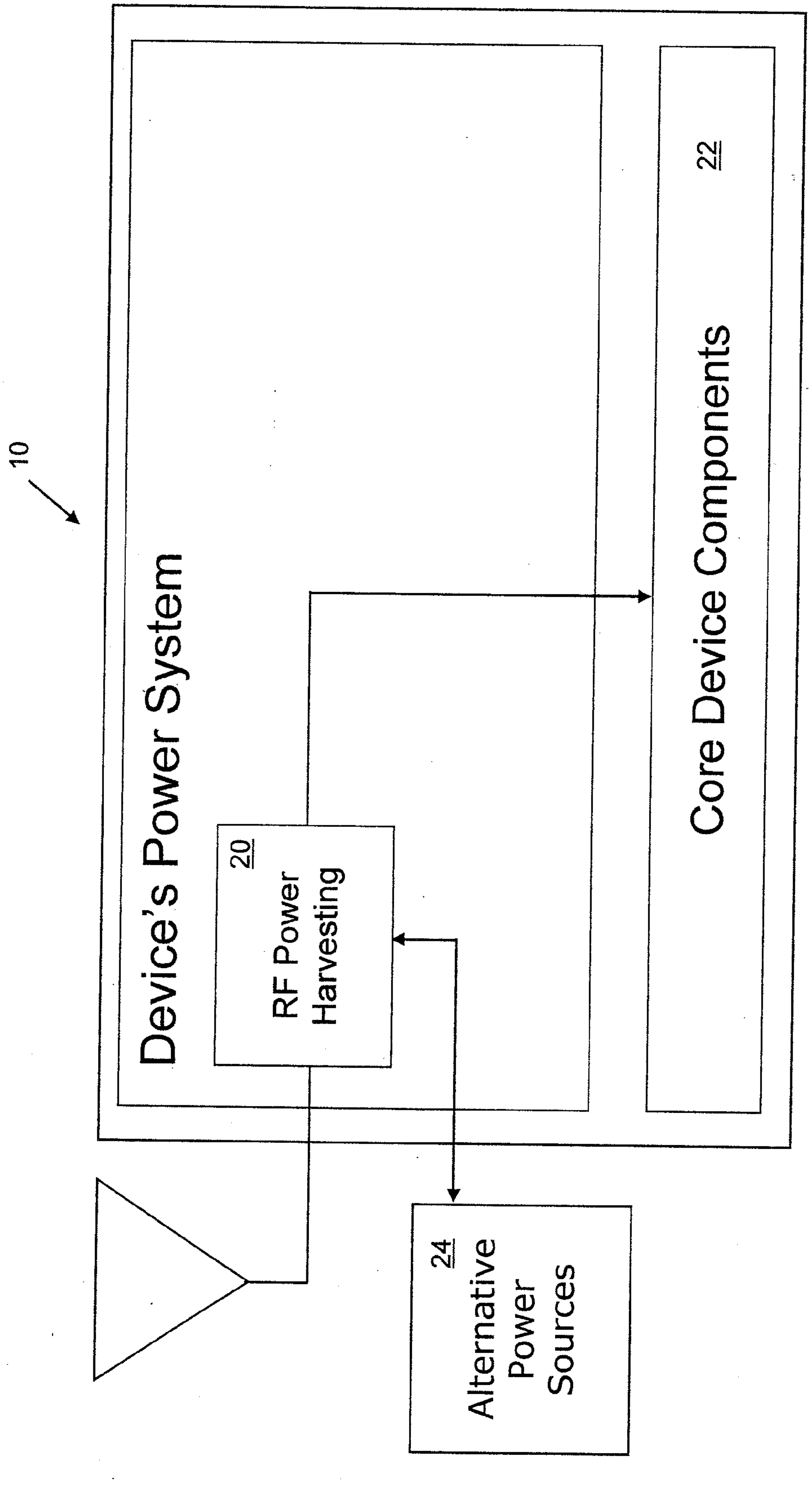
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 61

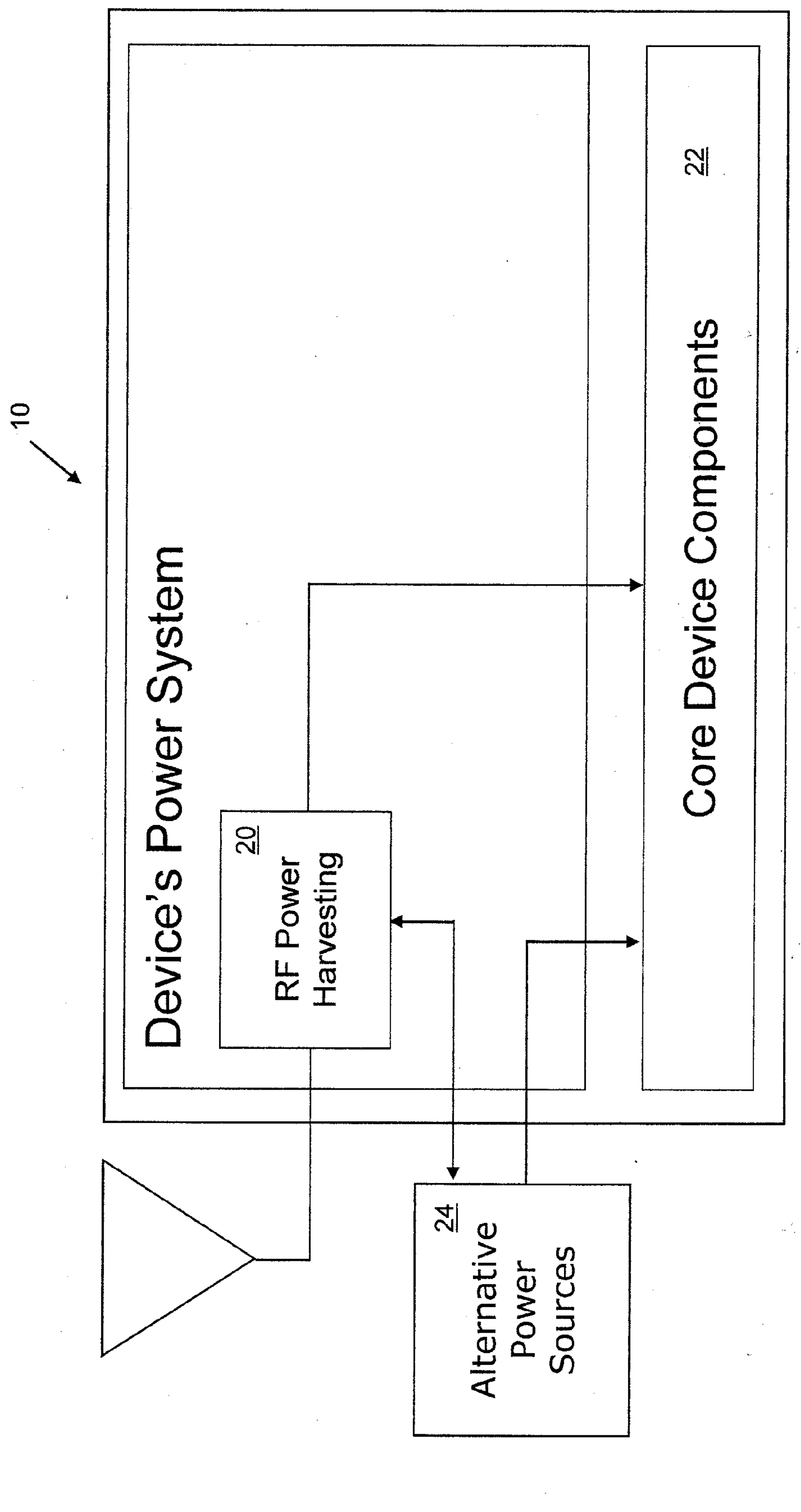
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 62

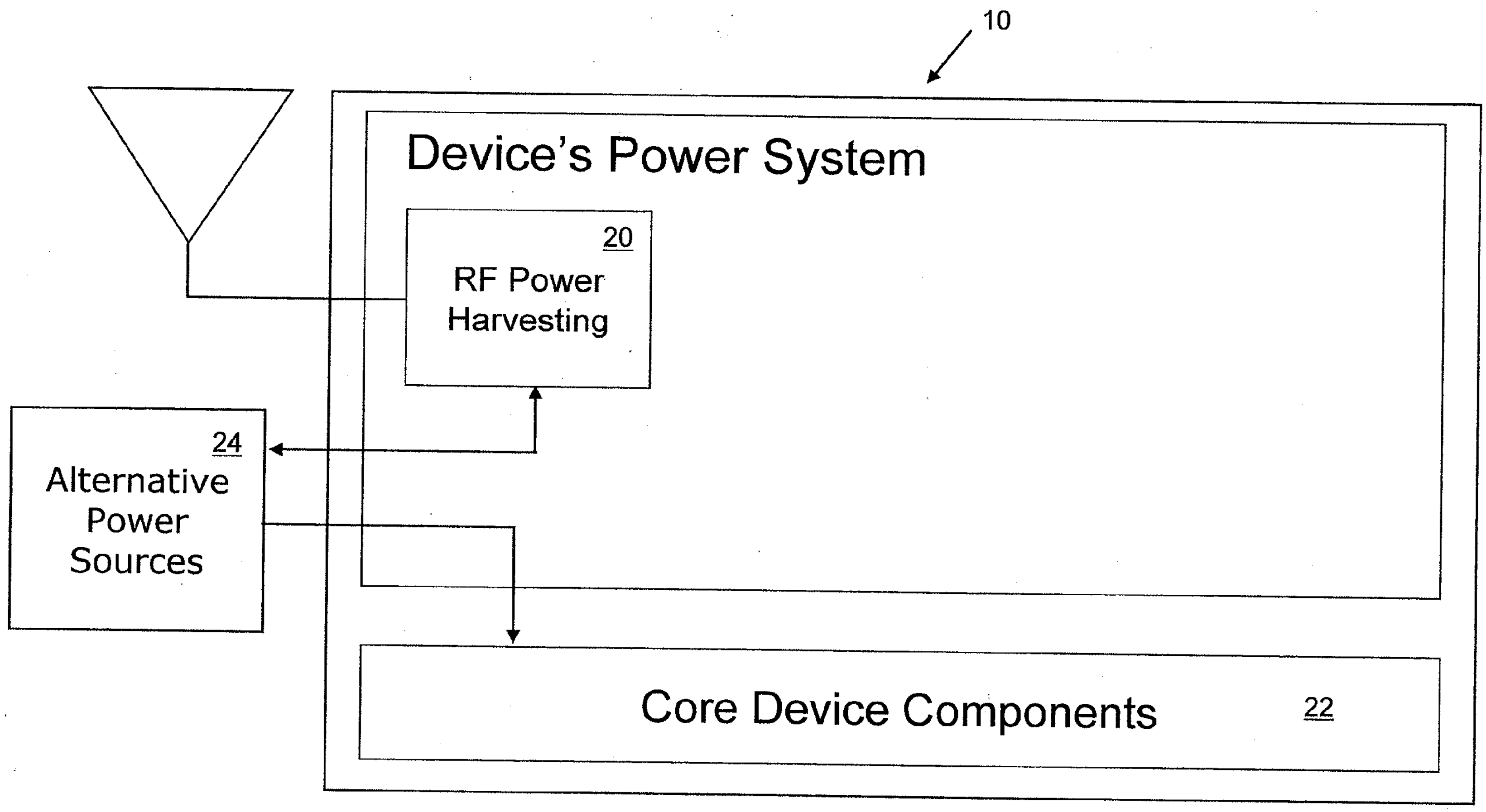
10
Device's Power System
20
RF Power Harvesting
24
Alternative Power Sources
Core Device Components
22

Fig. 63

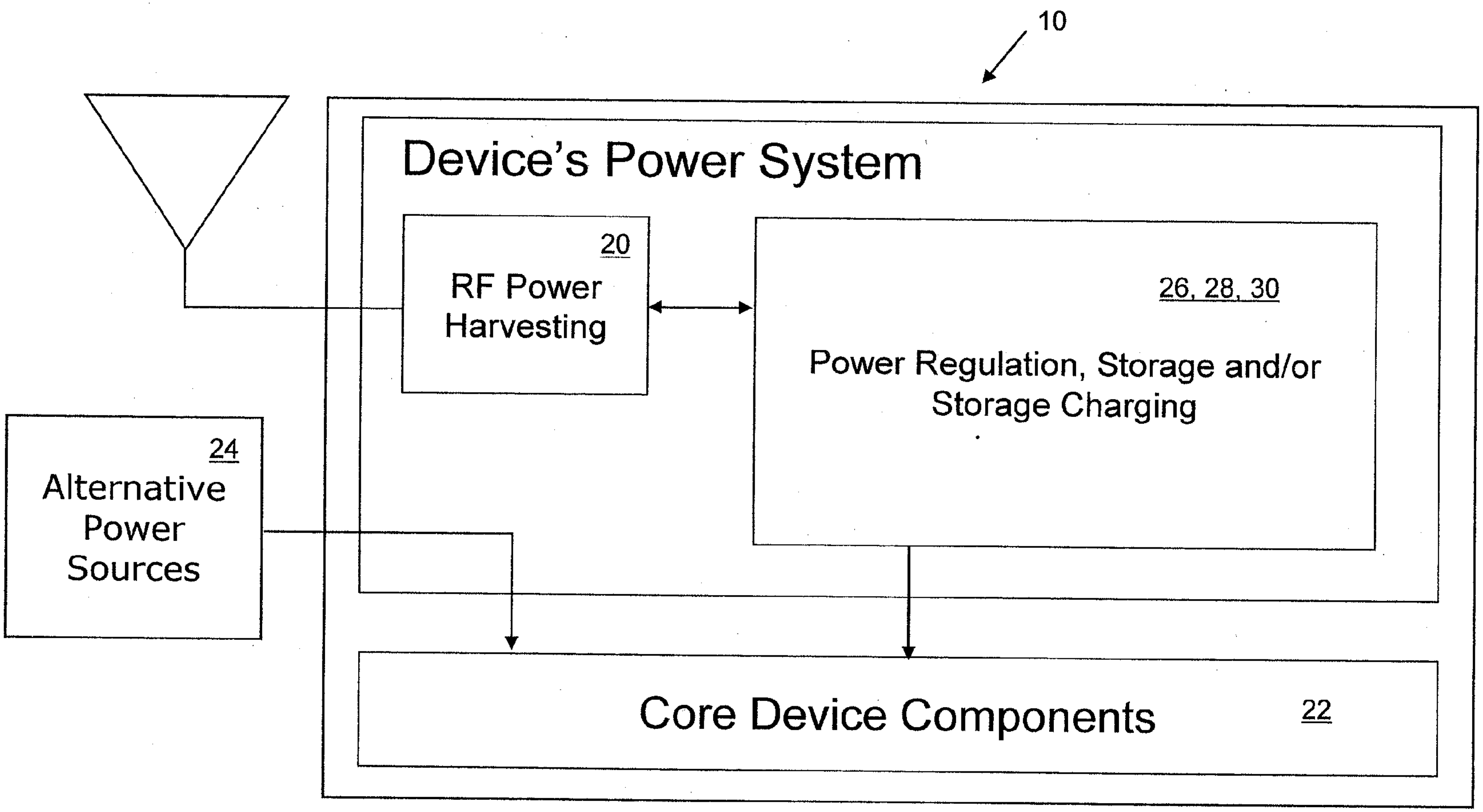
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 64

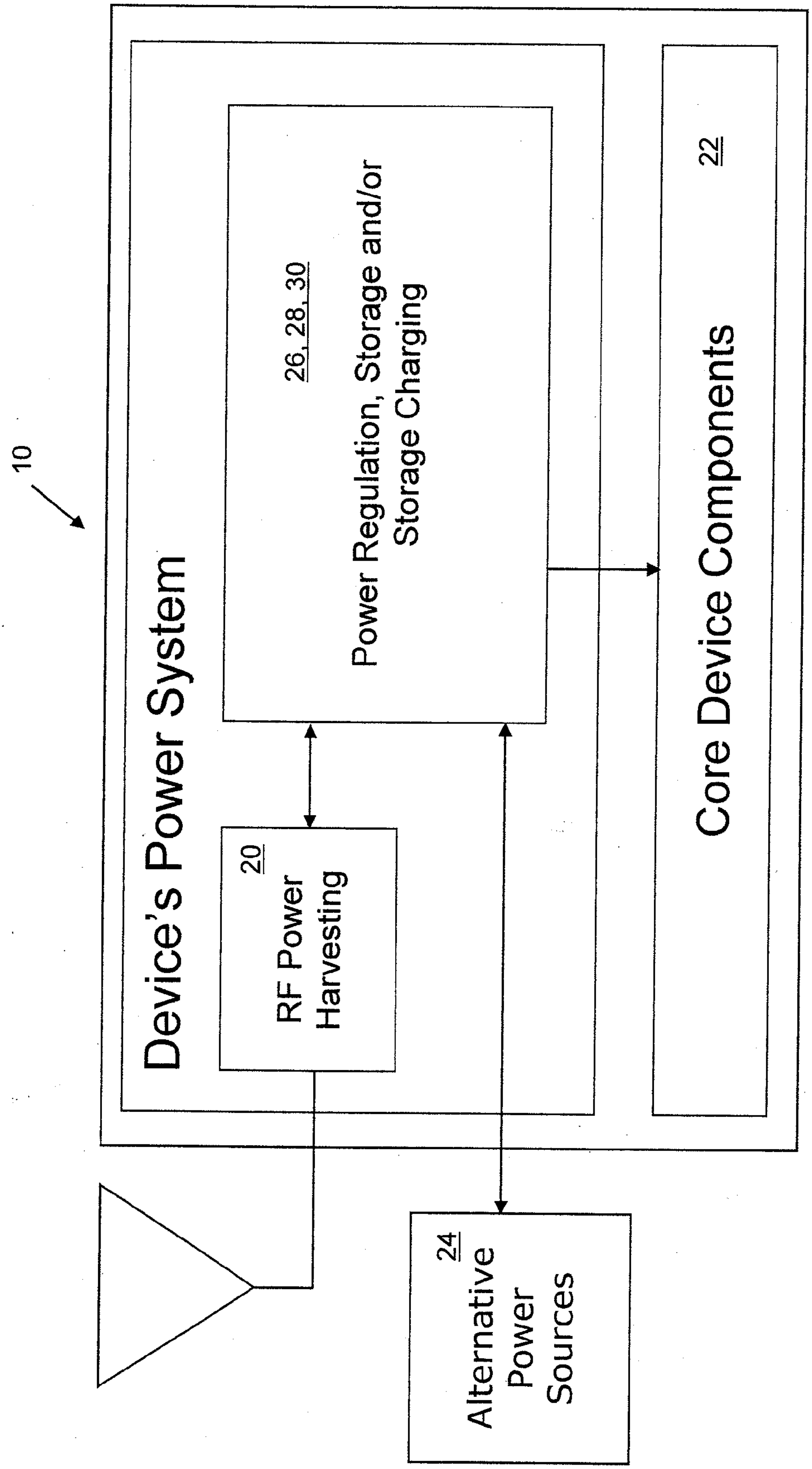
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 65

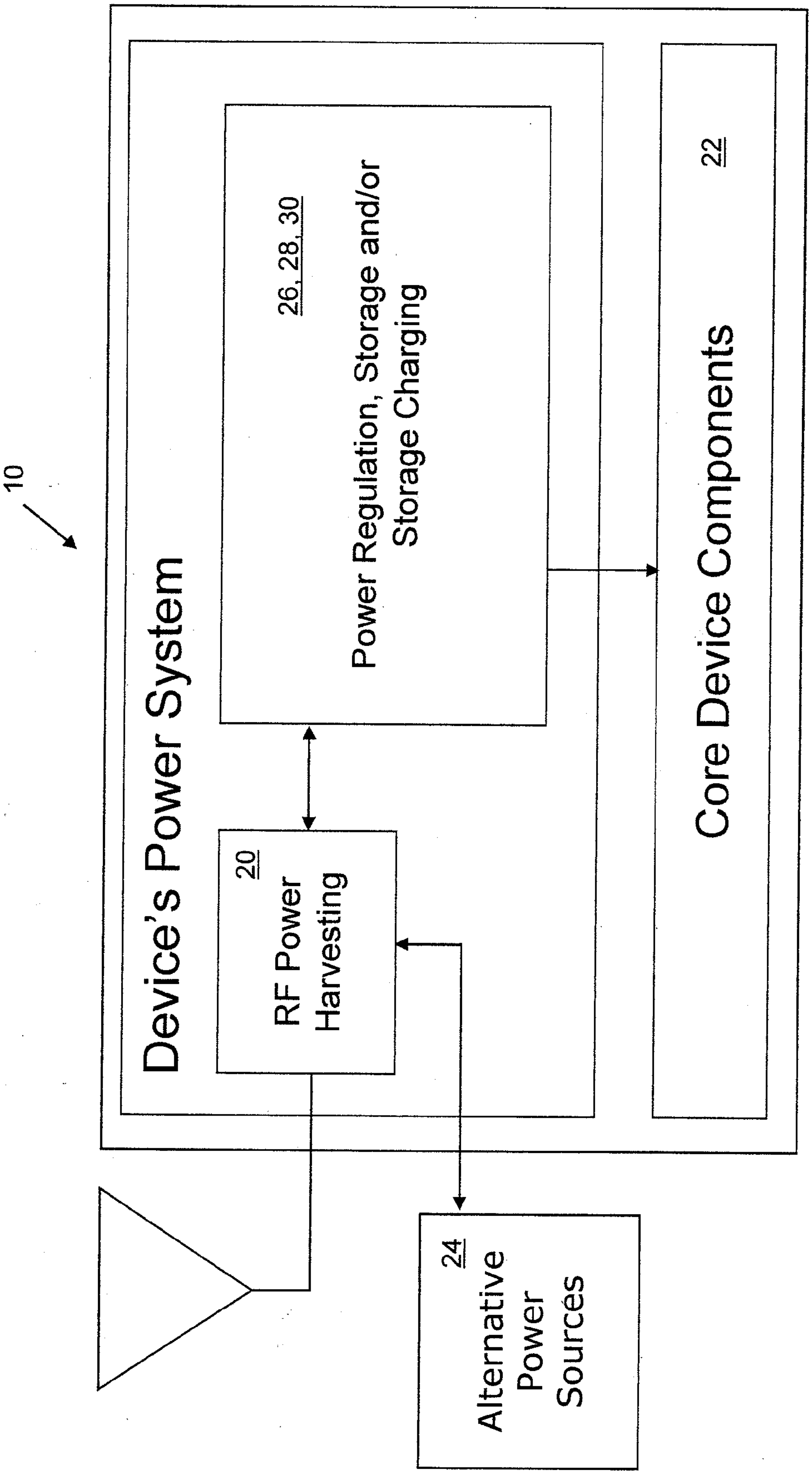
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 66

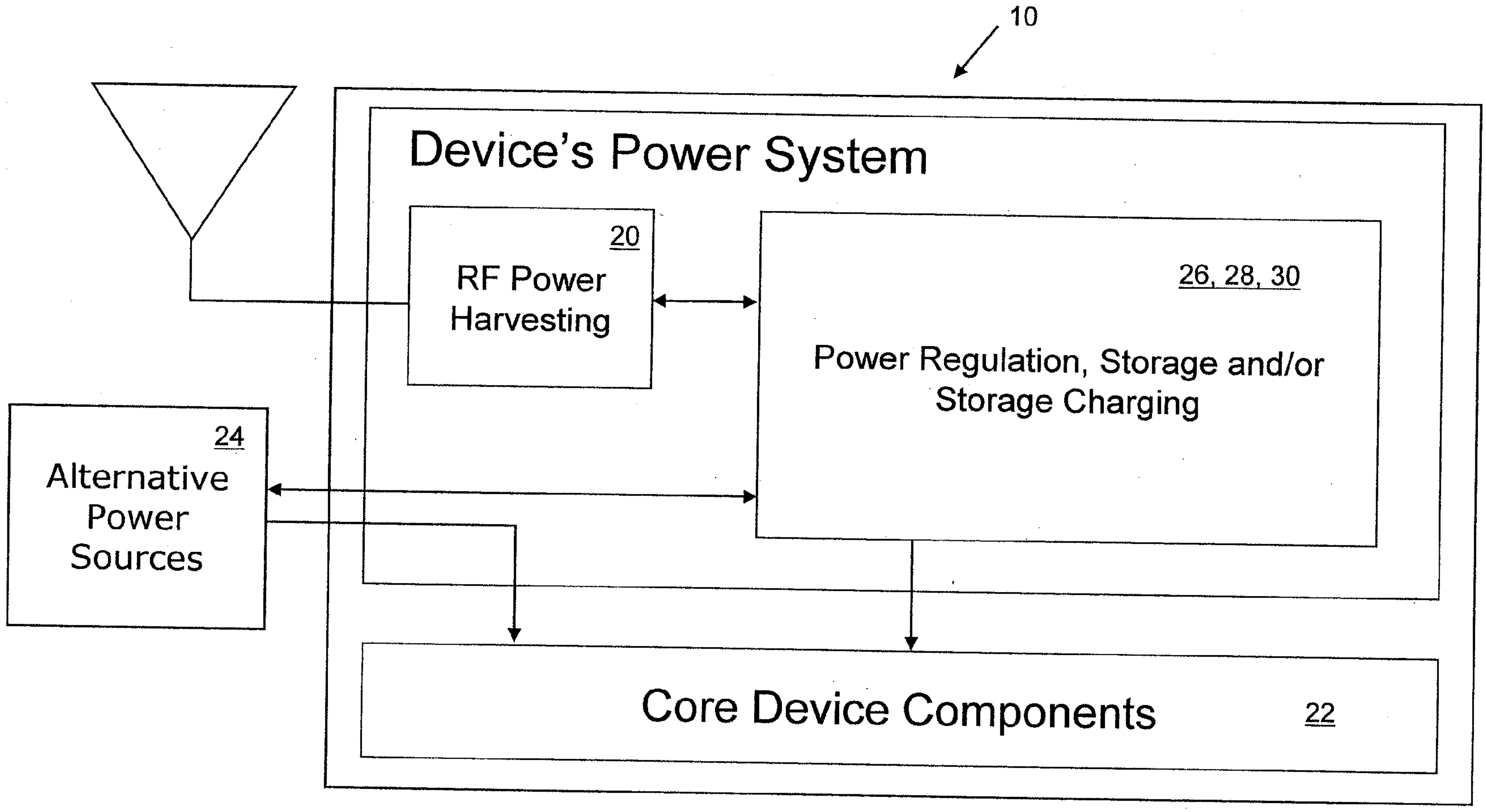
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 67

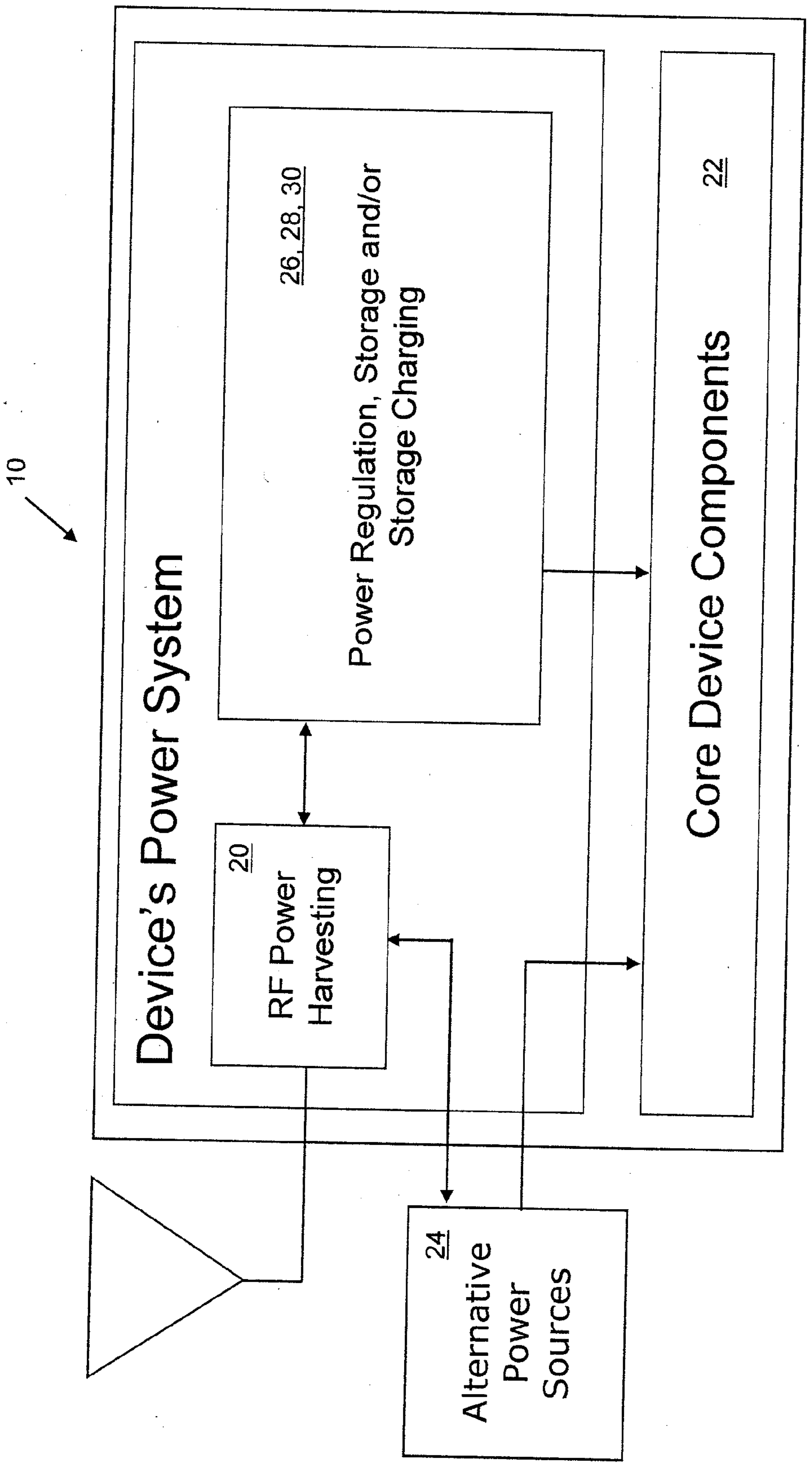
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
22
Core Device Components
24
Alternative Power Sources

Fig. 68

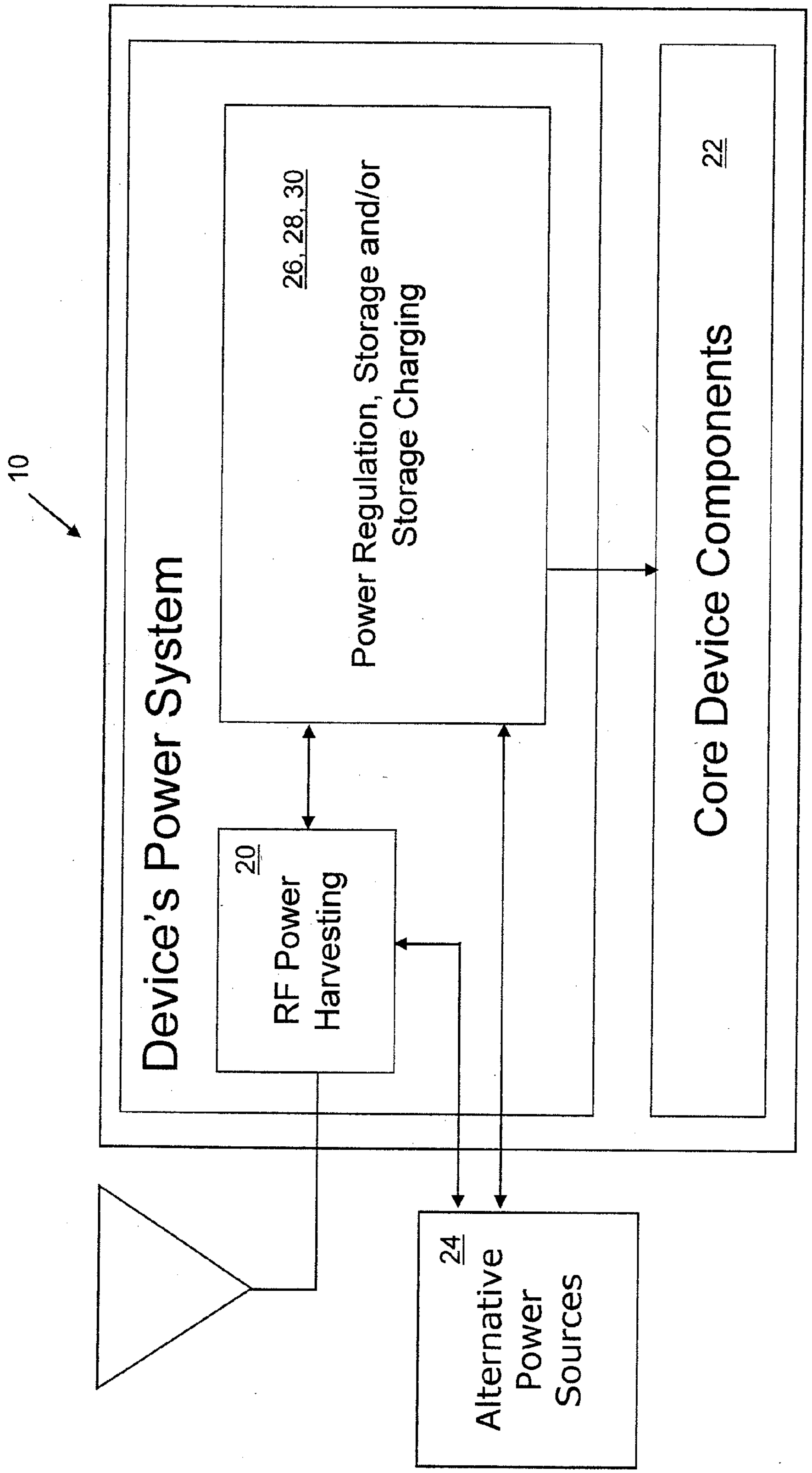
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 69

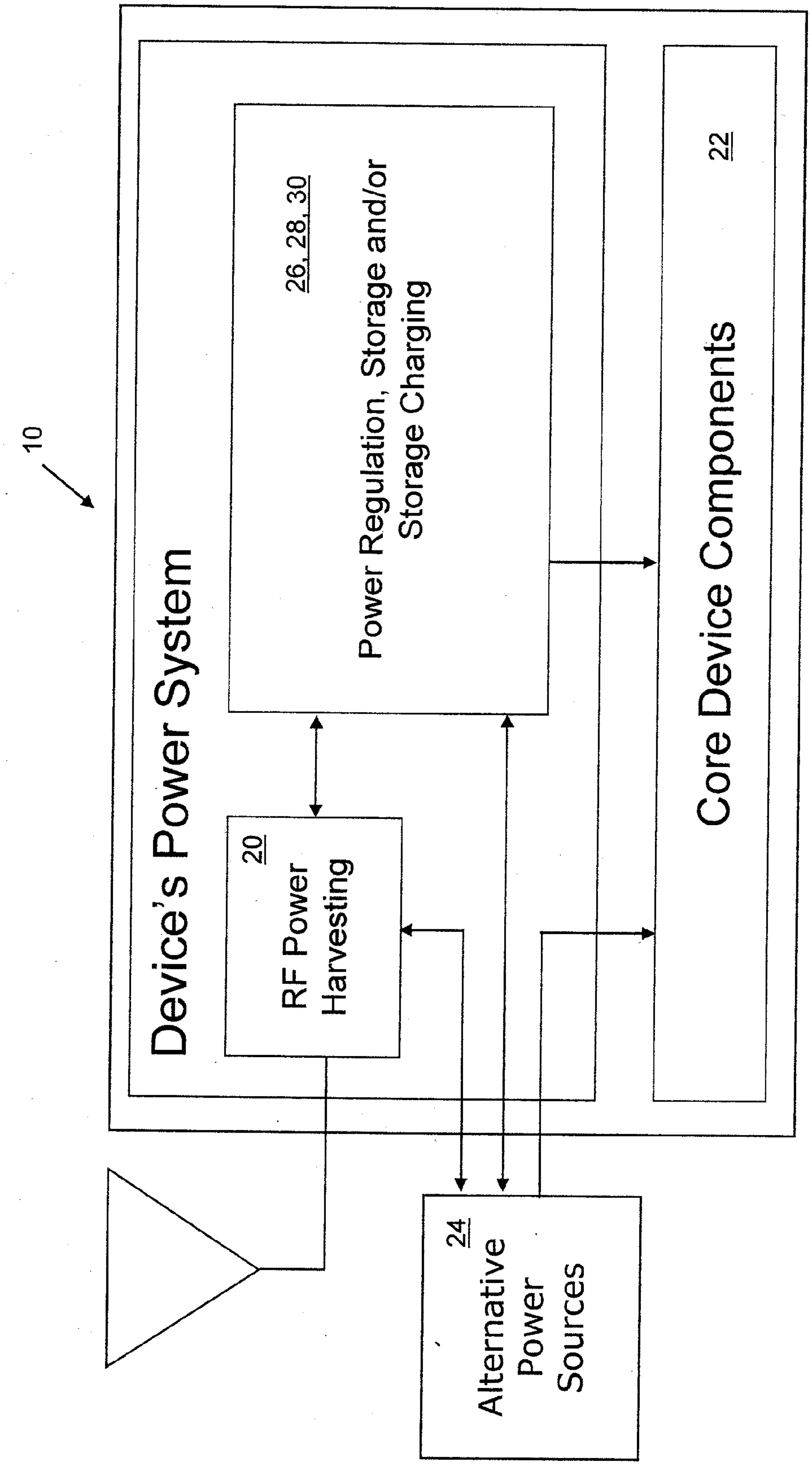
10
Device's Power System
Power Regulation, Storage and/or Storage Charging
26, 28, 30
Core Device Components
22
RF Power Harvesting
20
Alternative Power Sources
24

Fig. 70

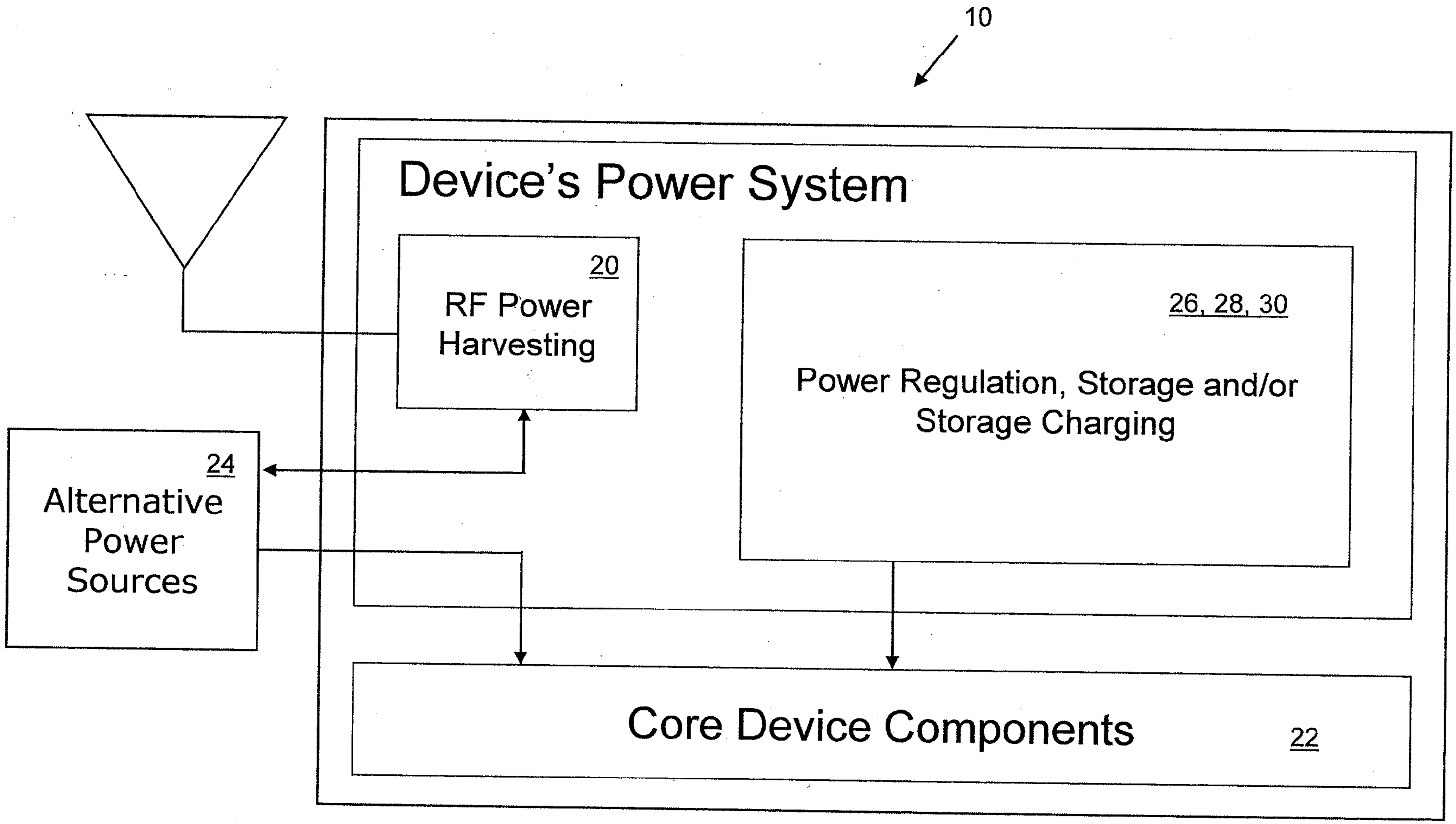
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 71

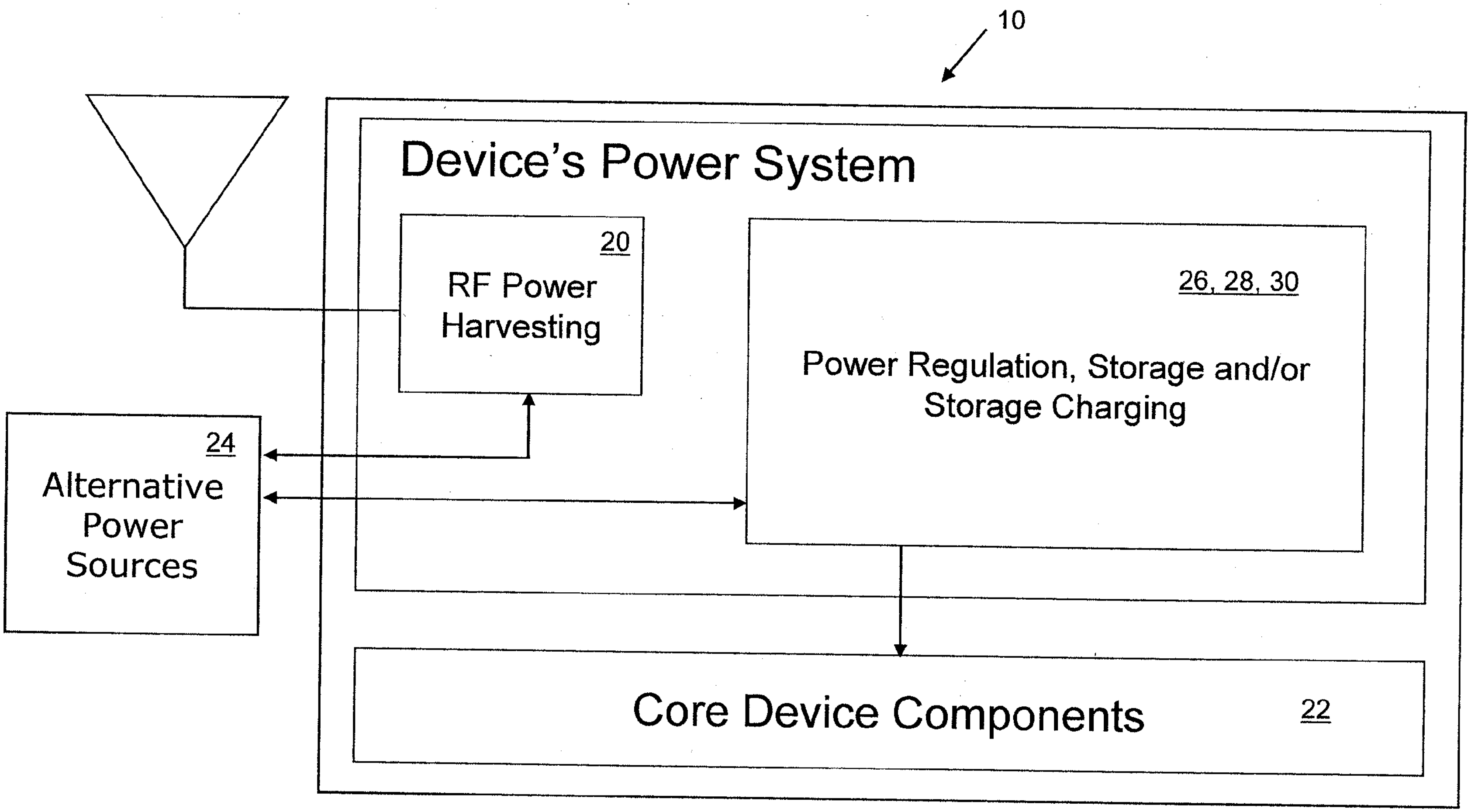
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 72

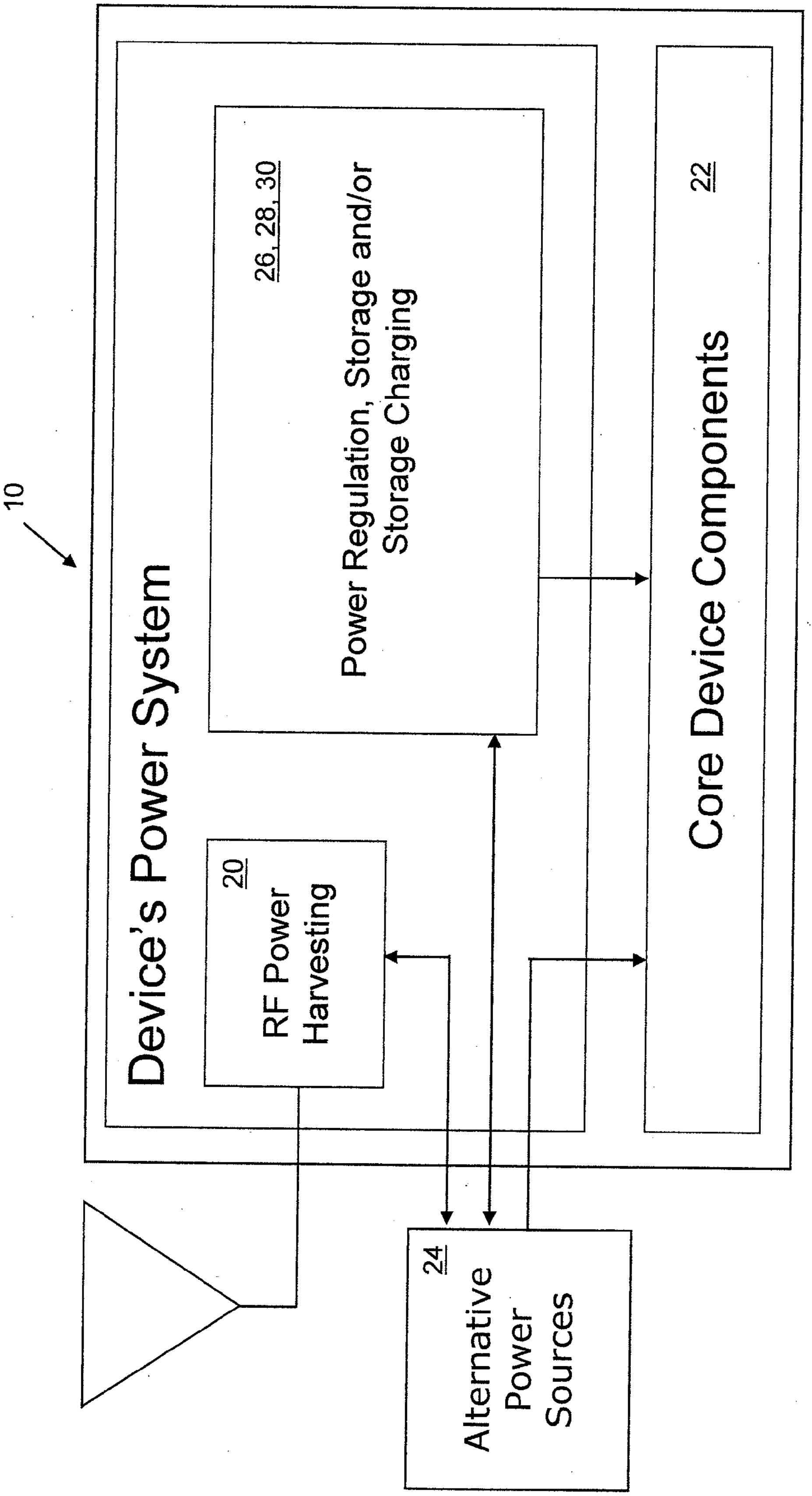
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 73

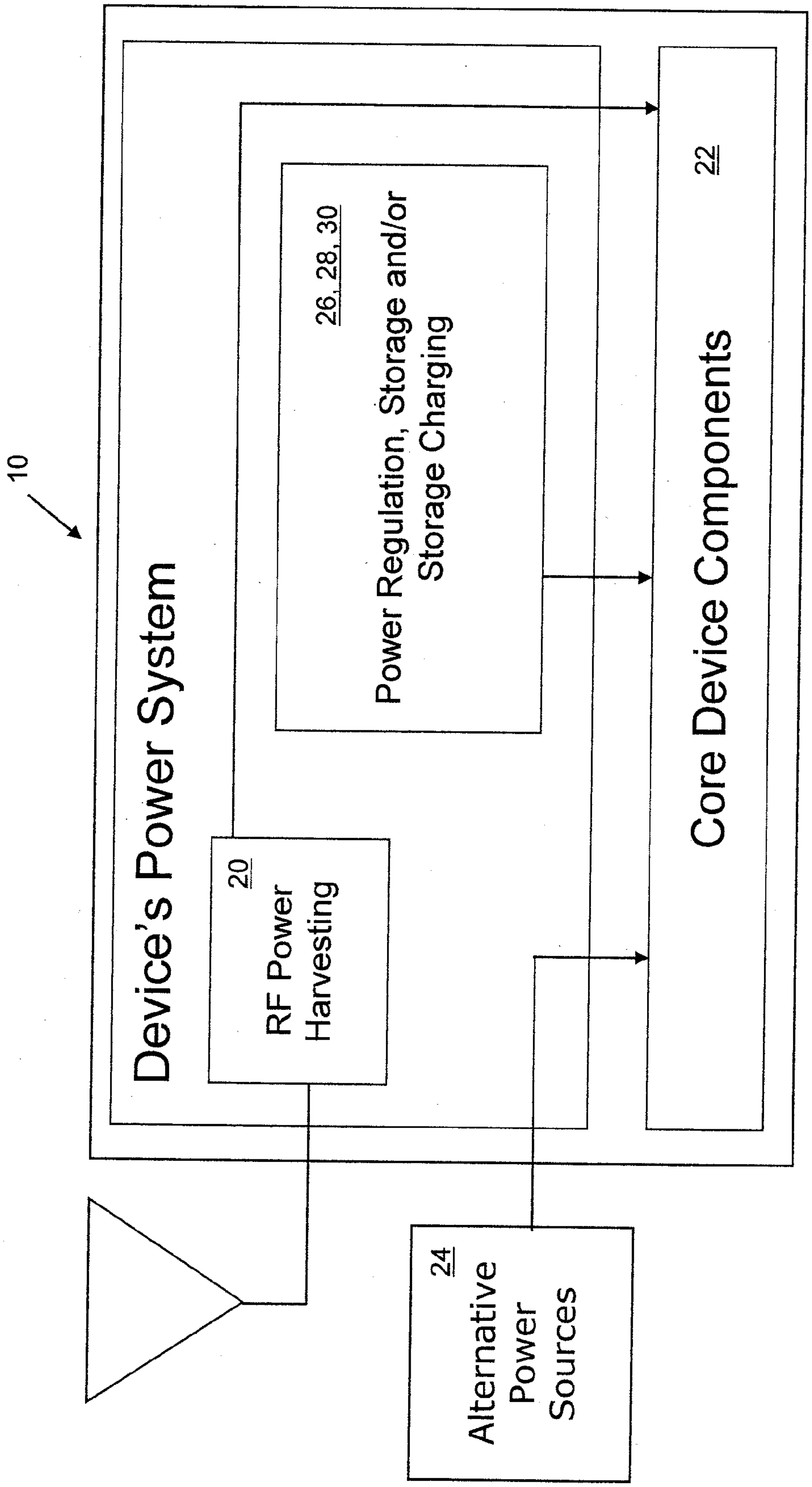
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 74

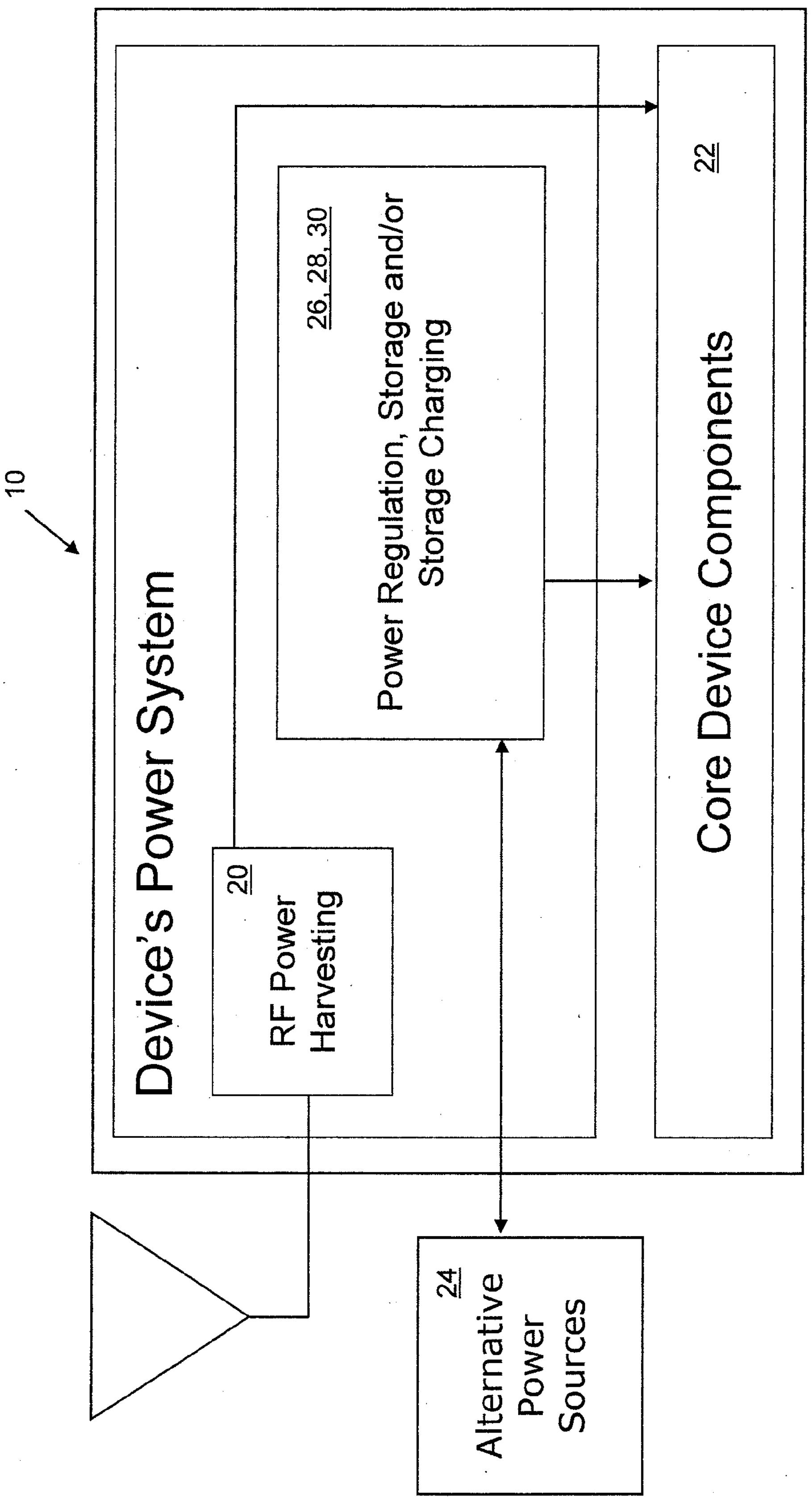
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 75

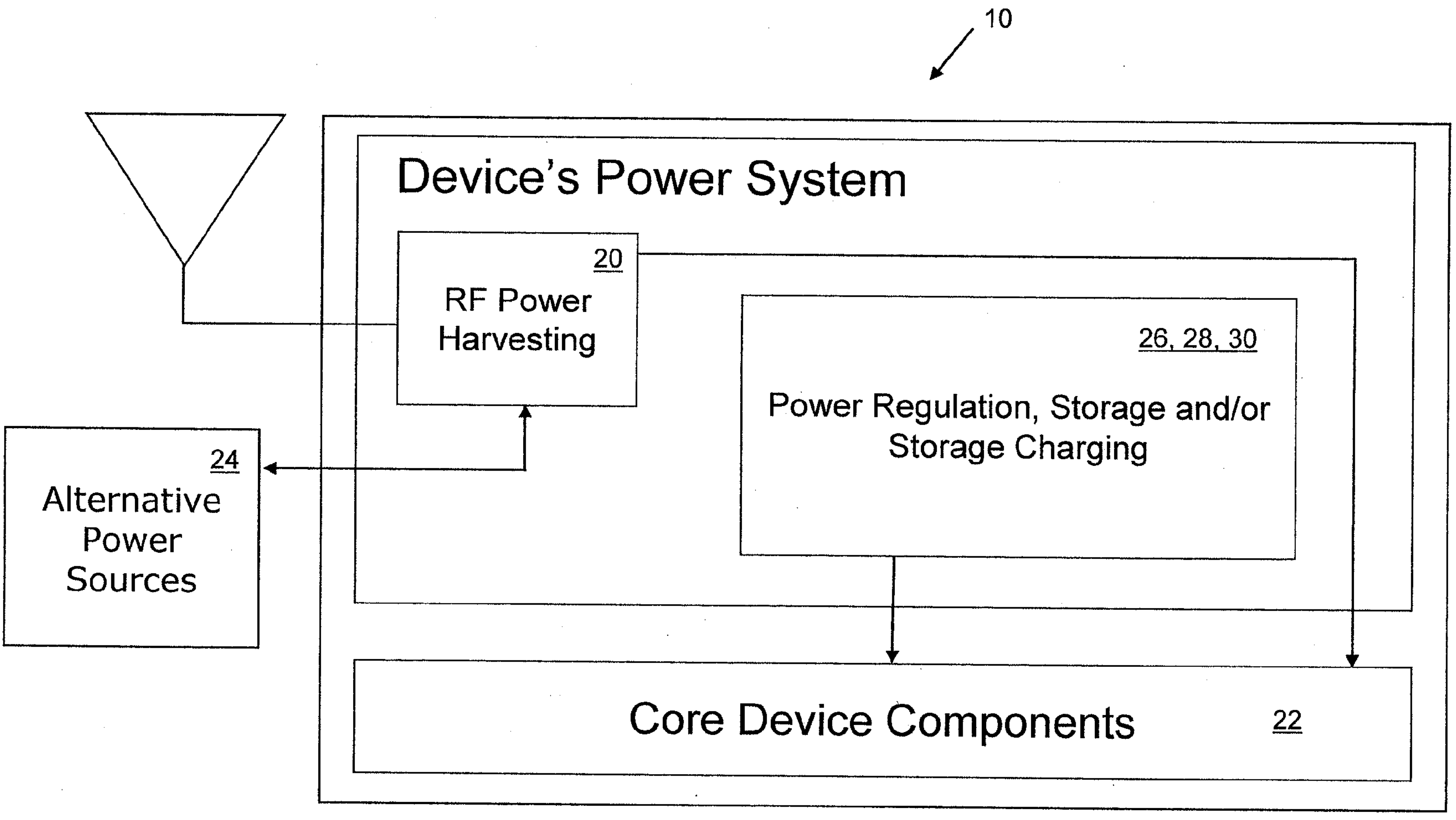
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 76

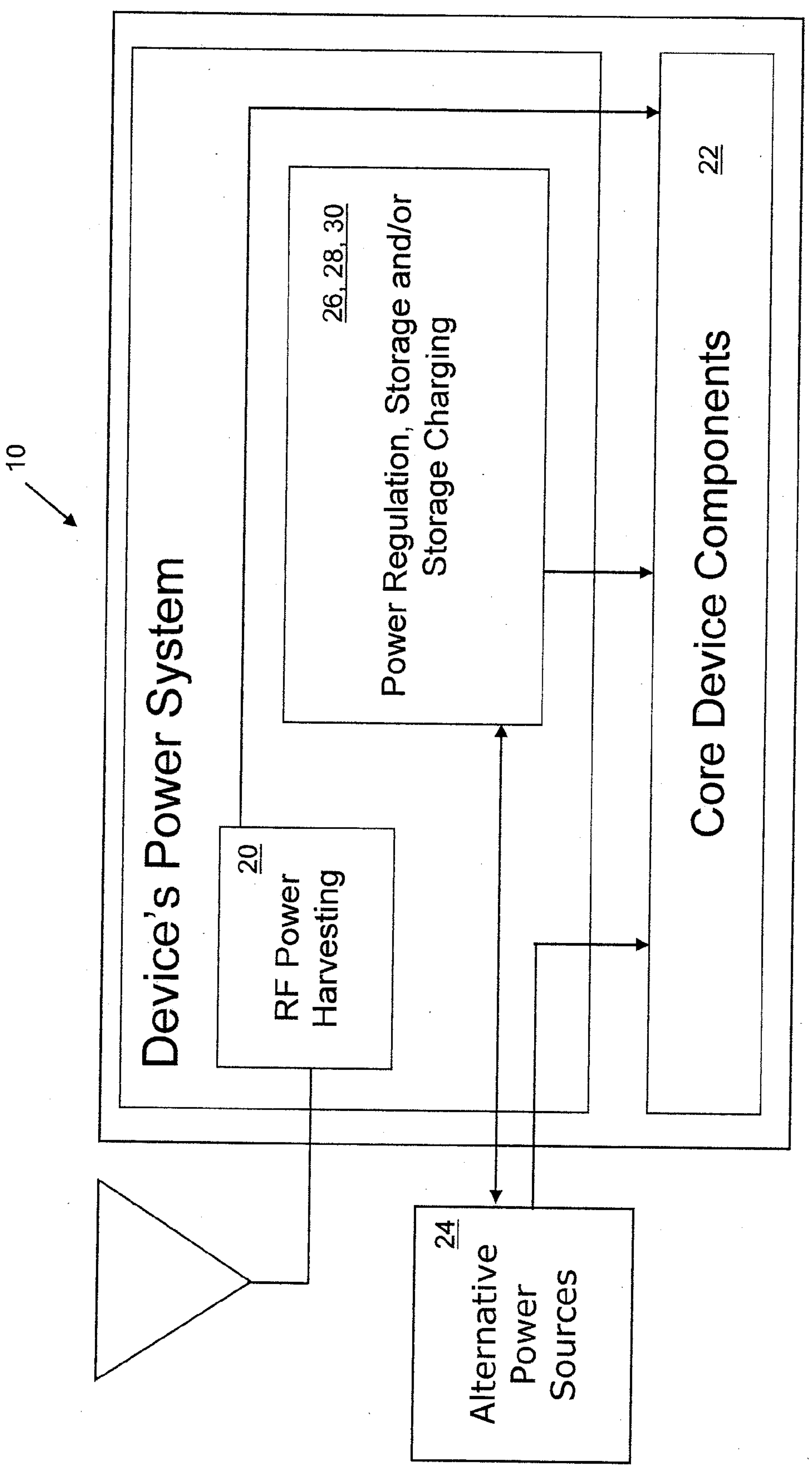
10
Device's Power System
RF Power Harvesting
20
Power Regulation, Storage and/or Storage Charging
26, 28, 30
Core Device Components
22
Alternative Power Sources
24

Fig. 77

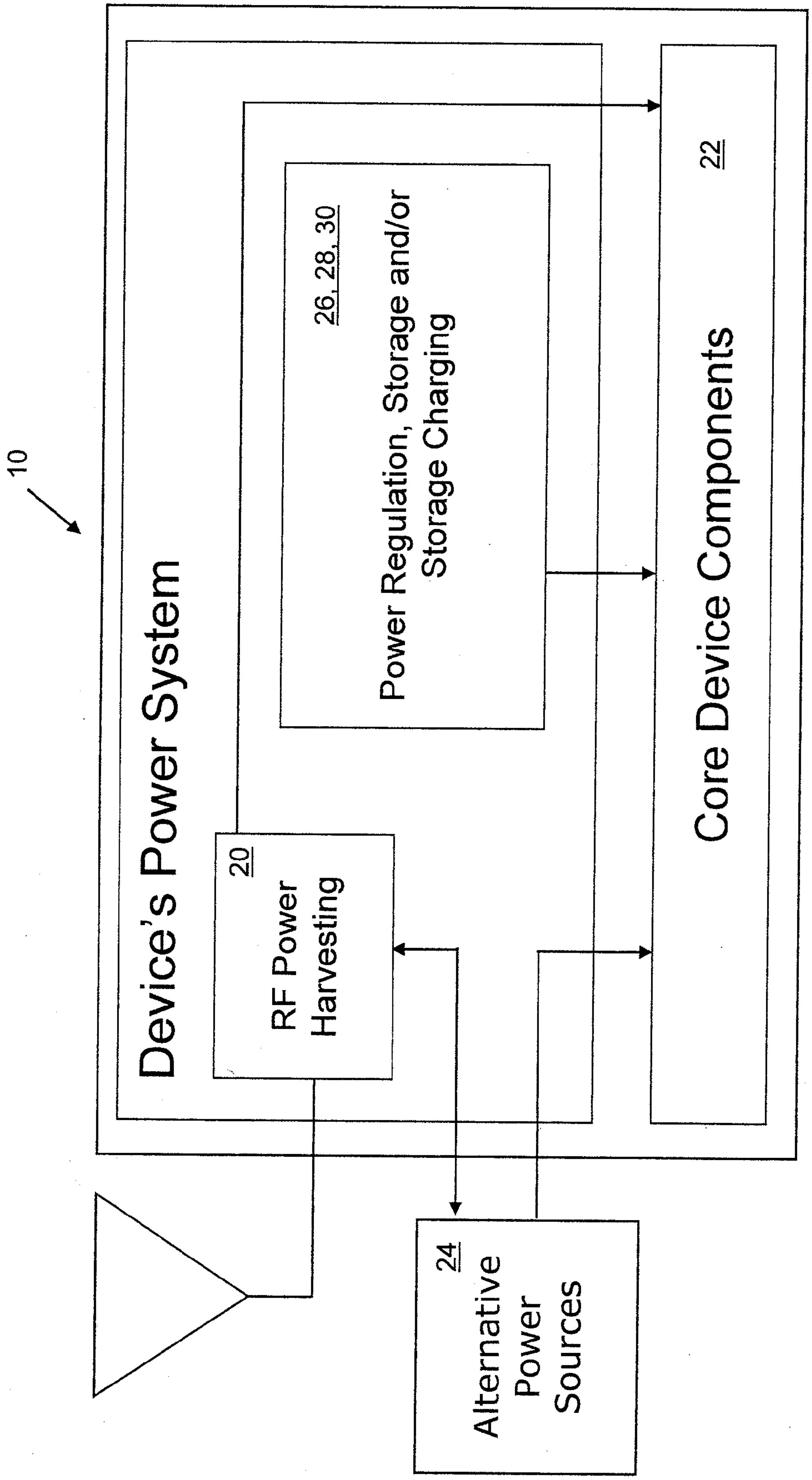
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 78

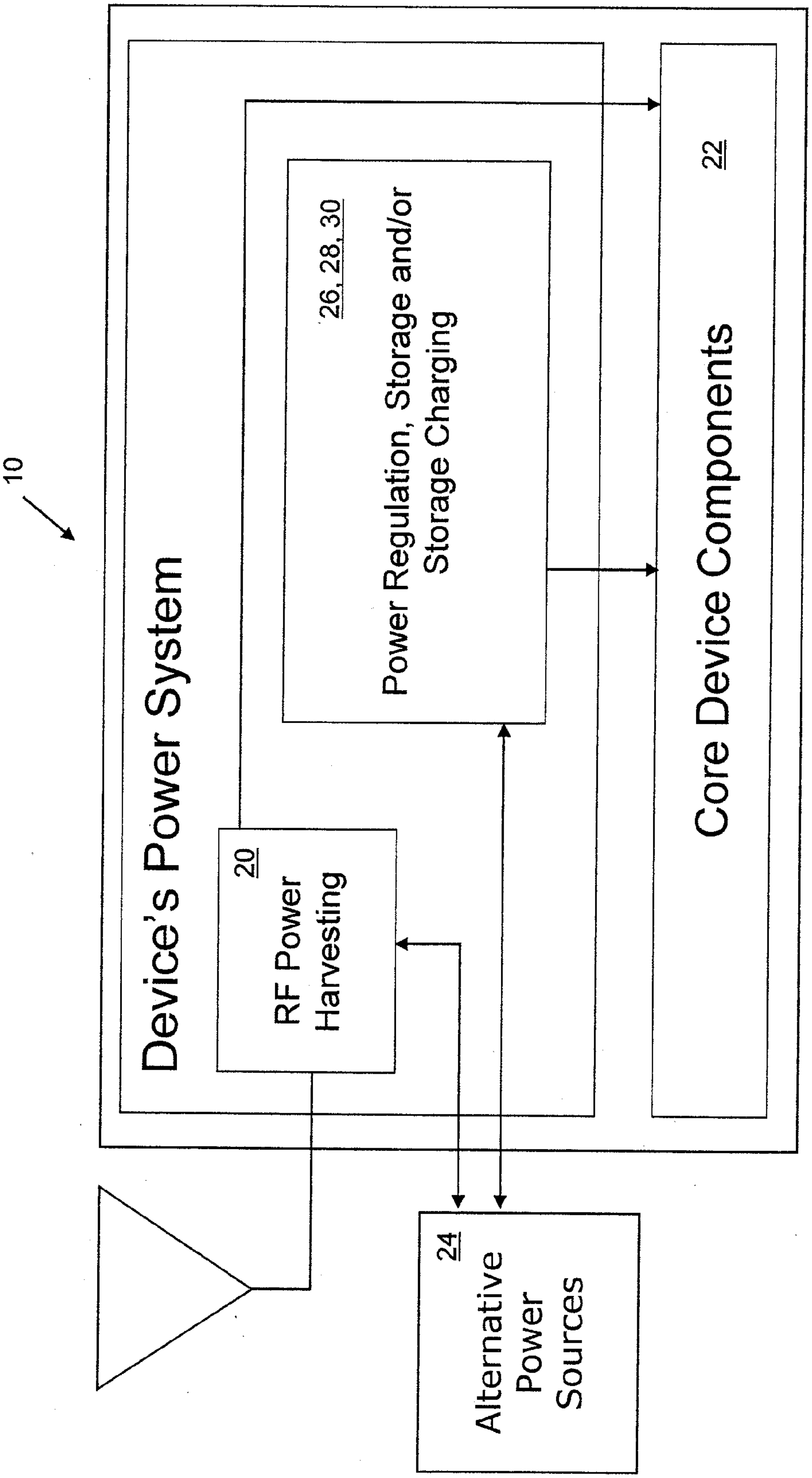
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 79

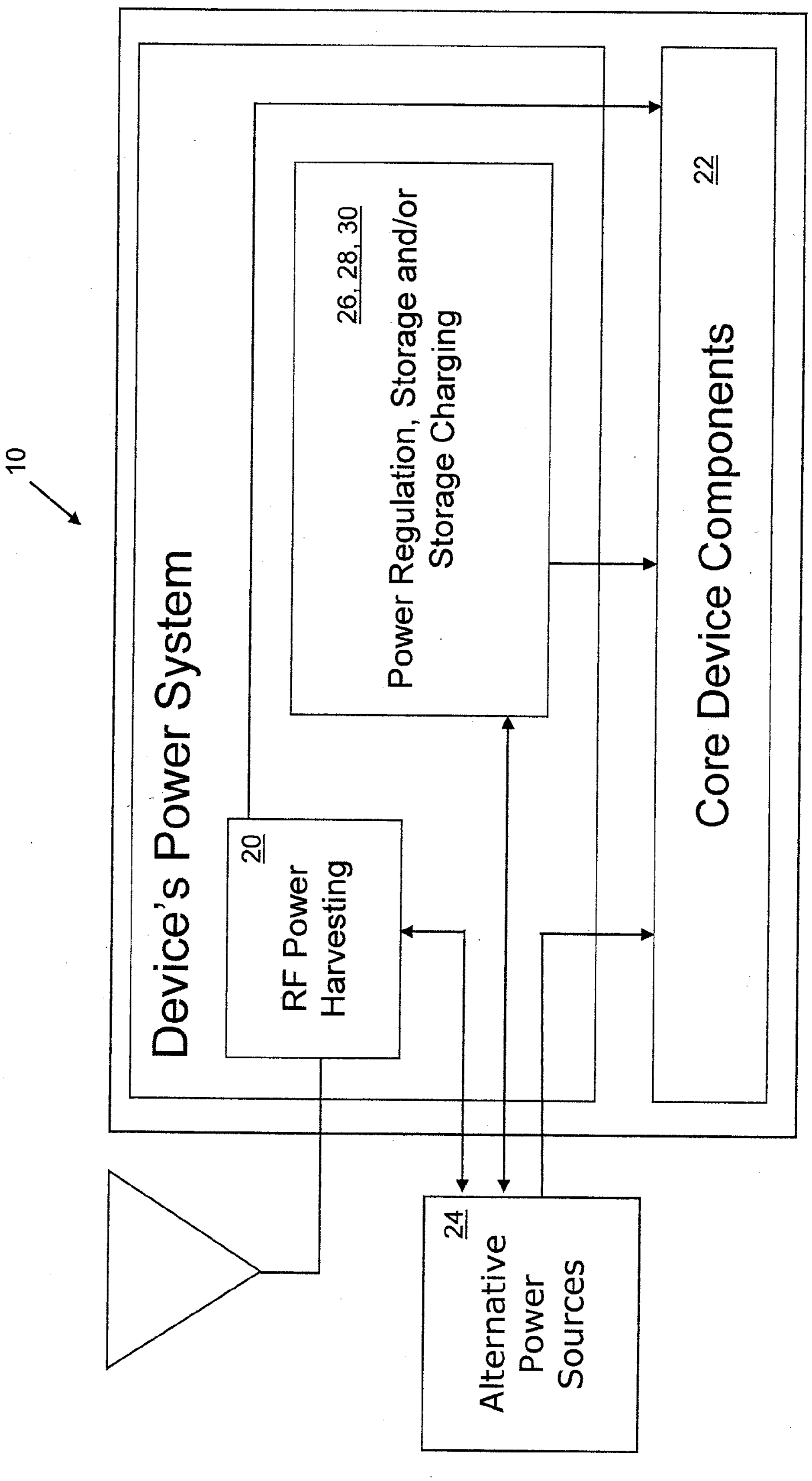
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 80

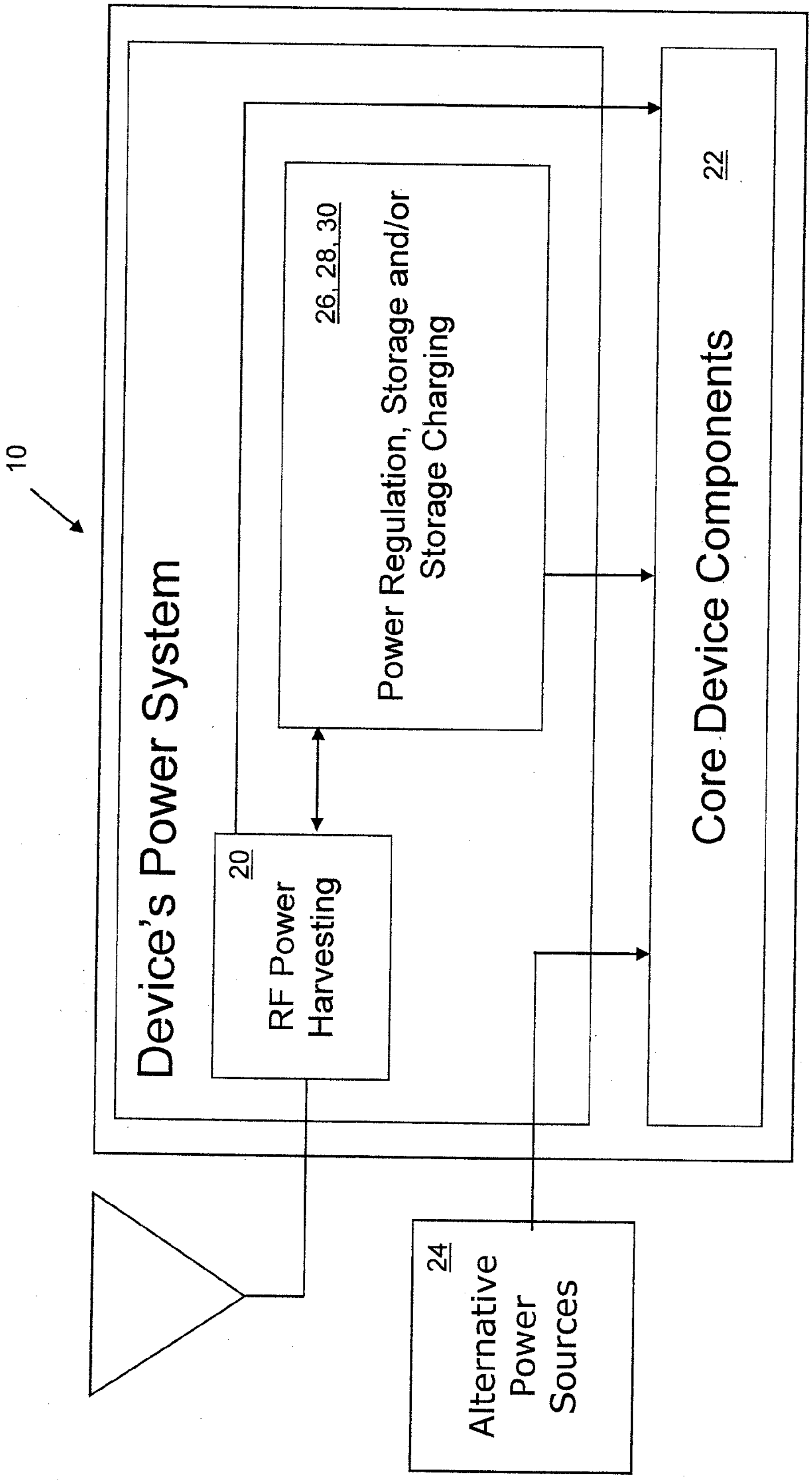
10
Device's Power System
RF Power Harvesting
20
Power Regulation, Storage and/or Storage Charging
26, 28, 30
Core Device Components
22
Alternative Power Sources
24

Fig. 81

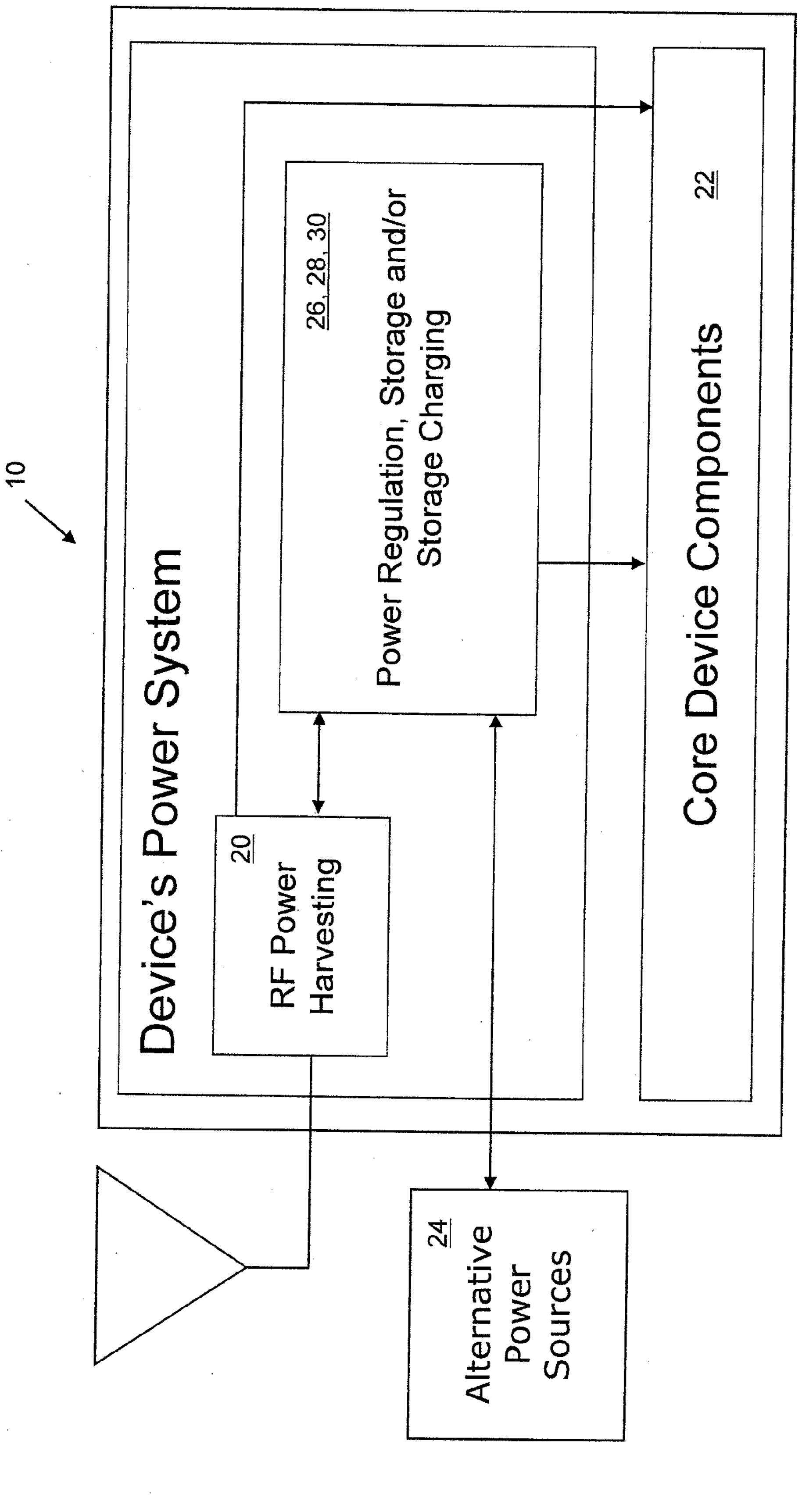
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 82

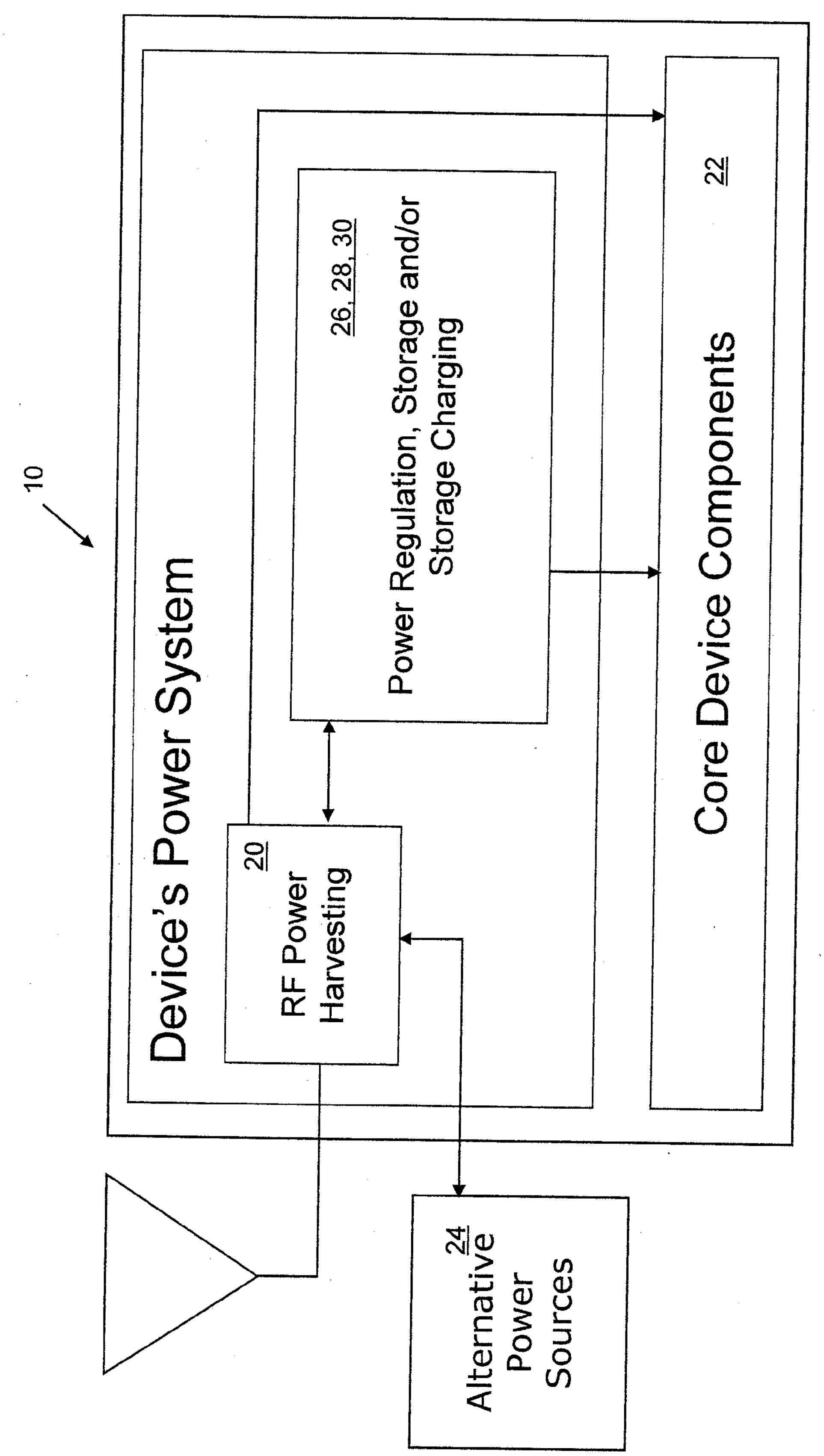
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 83

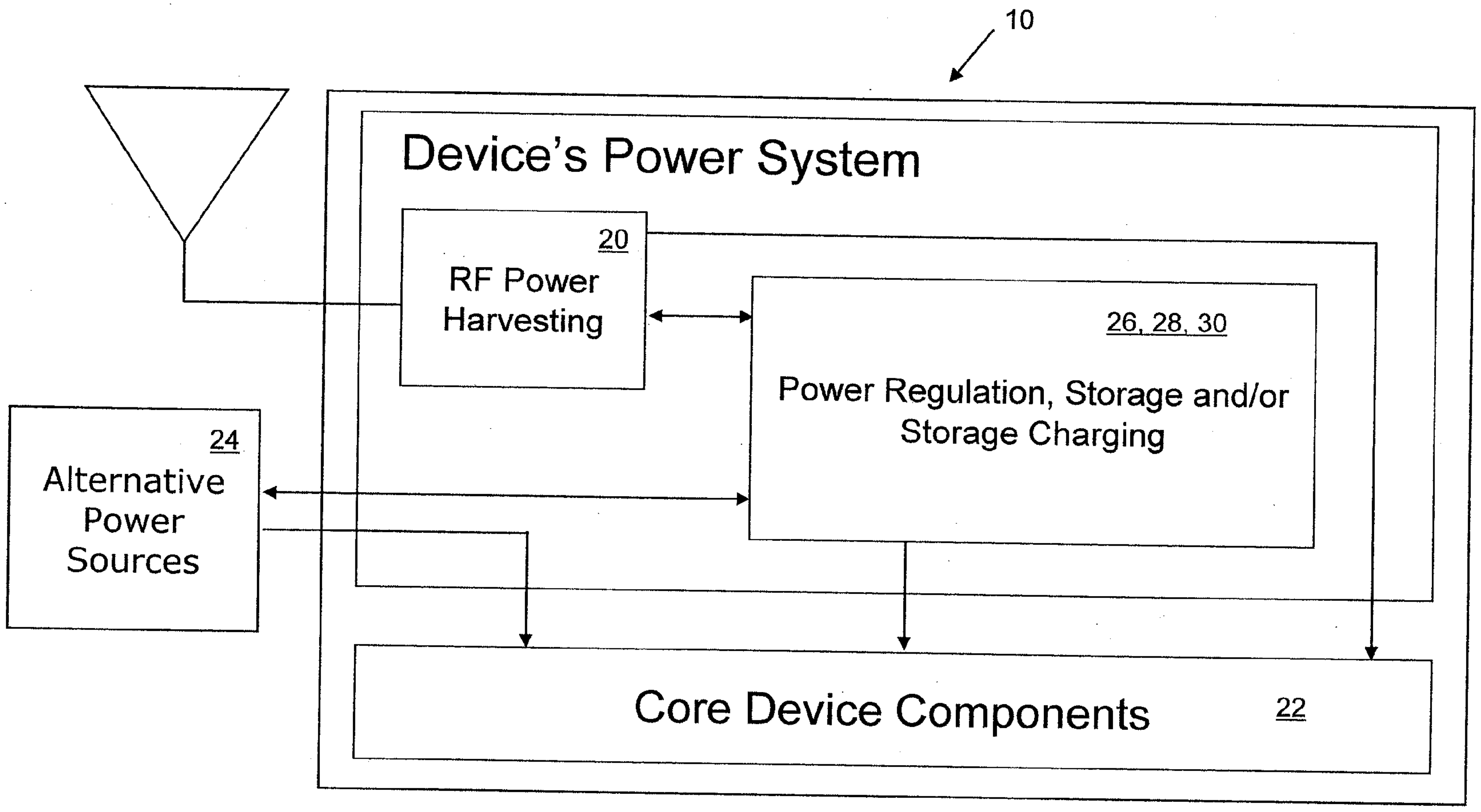
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 84

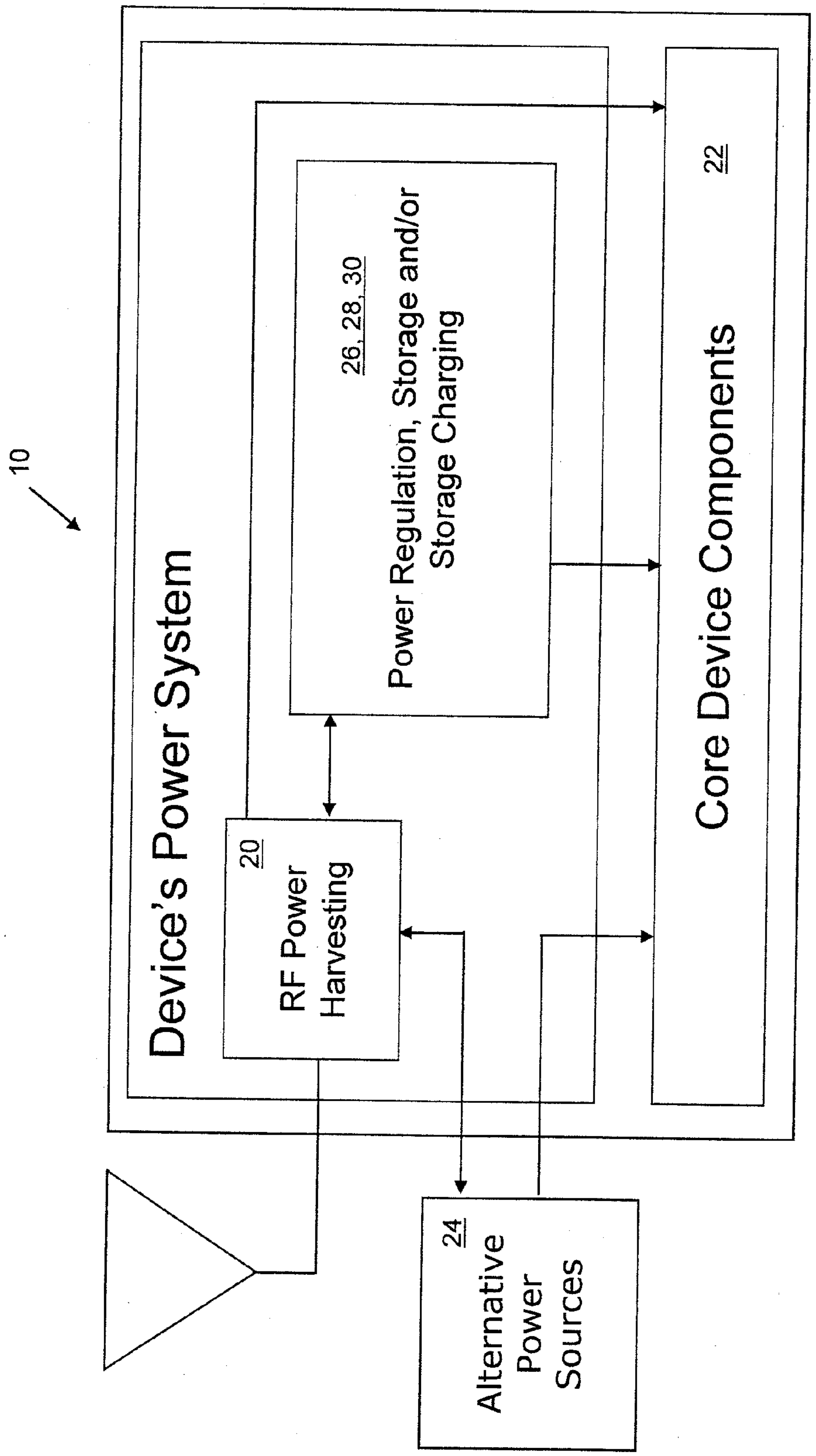
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 85

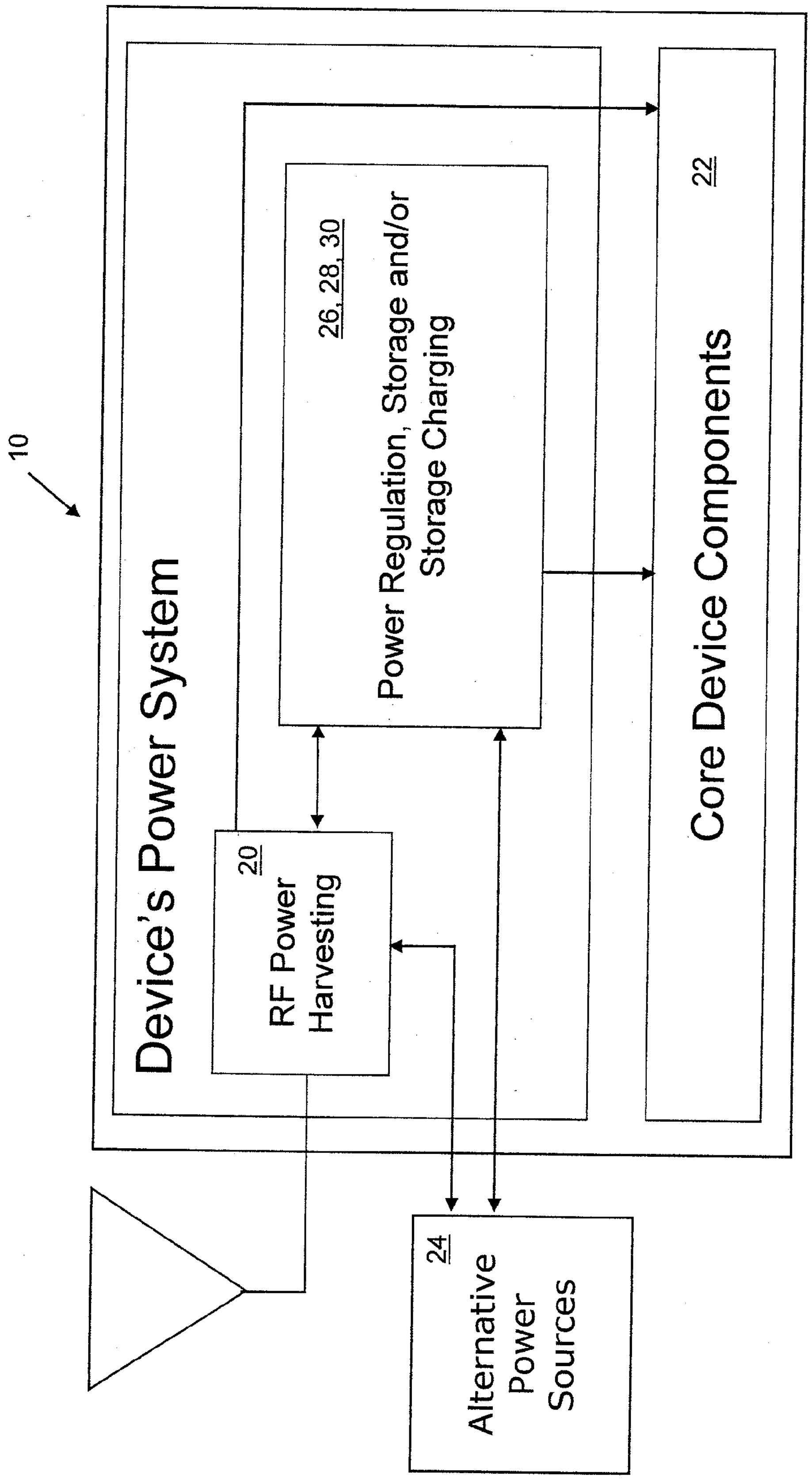
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 86

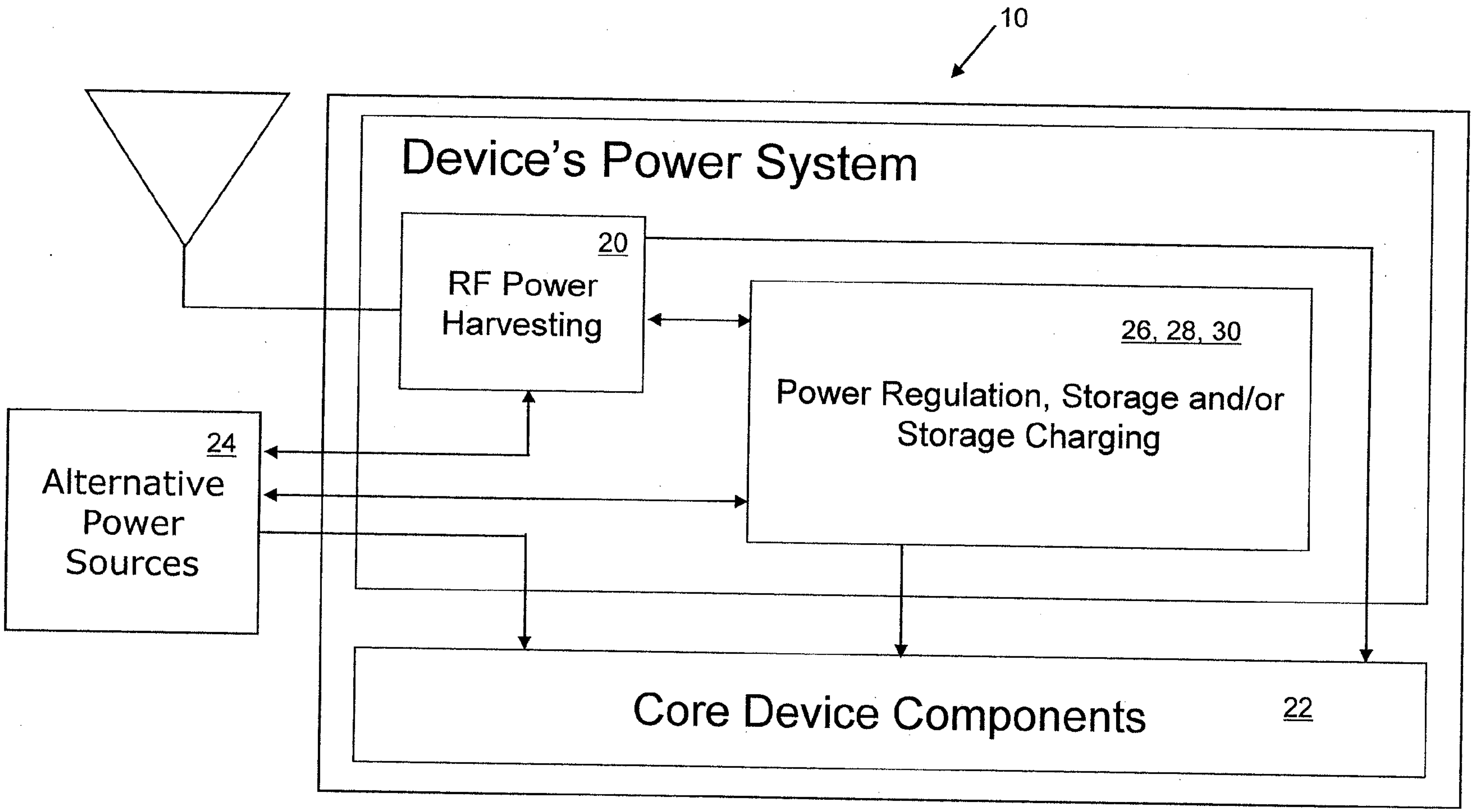
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 87

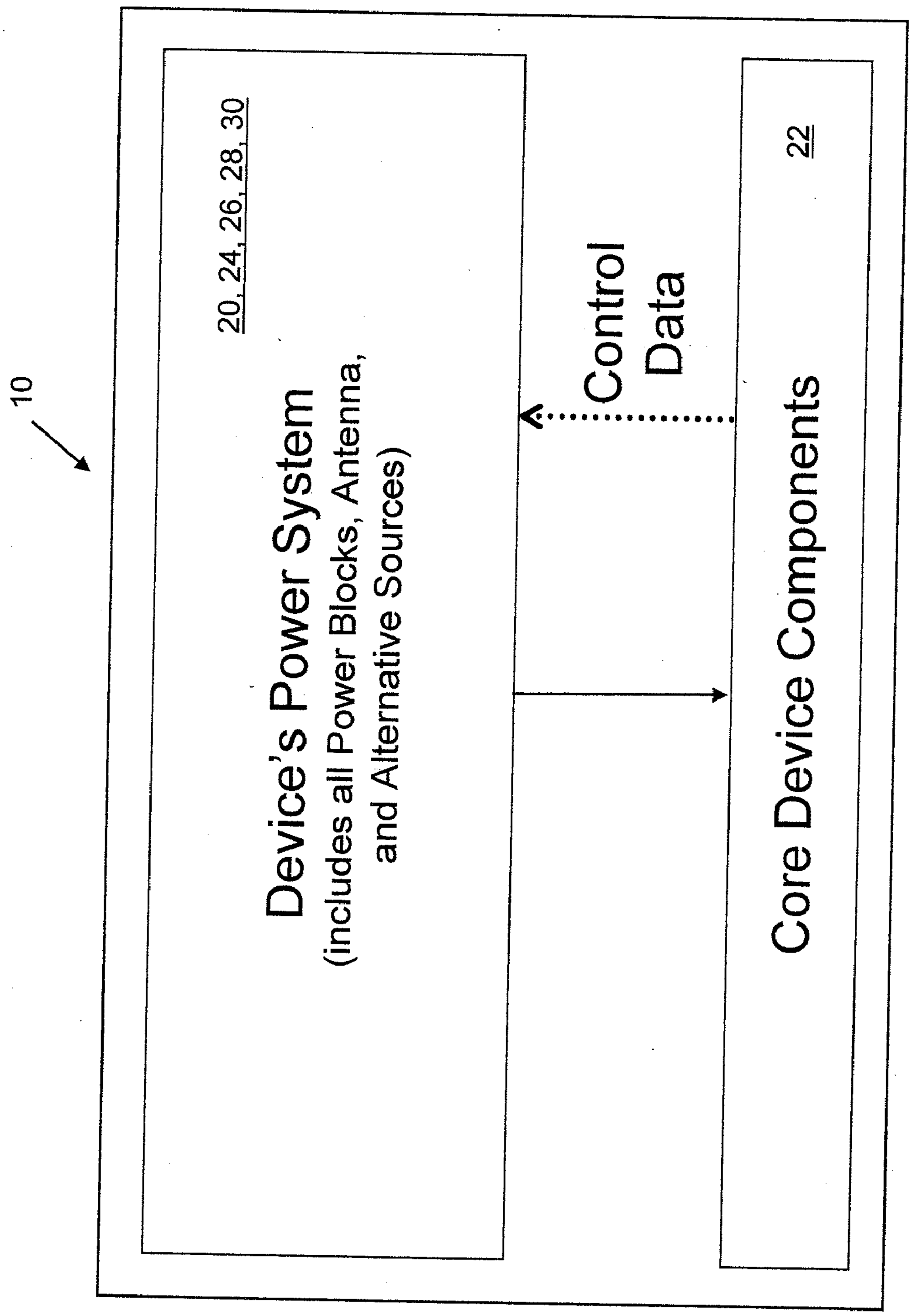
10
Device's Power System
(includes all Power Blocks, Antenna, and Alternative Sources)
20, 24, 26, 28, 30
Control Data
Core Device Components
22

Fig. 88

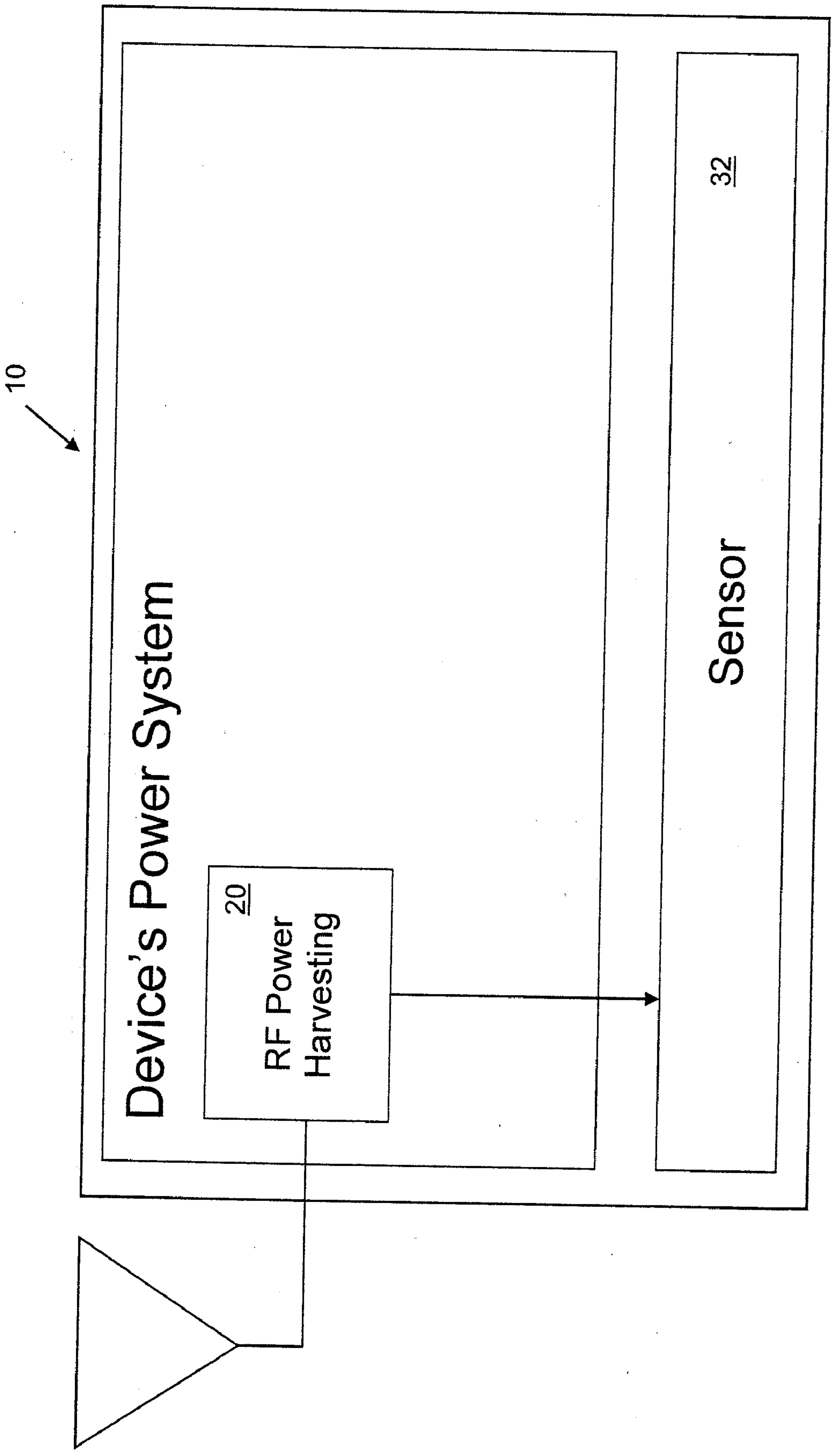
10
Device's Power System
20
RF Power Harvesting
Sensor
32

Fig. 89

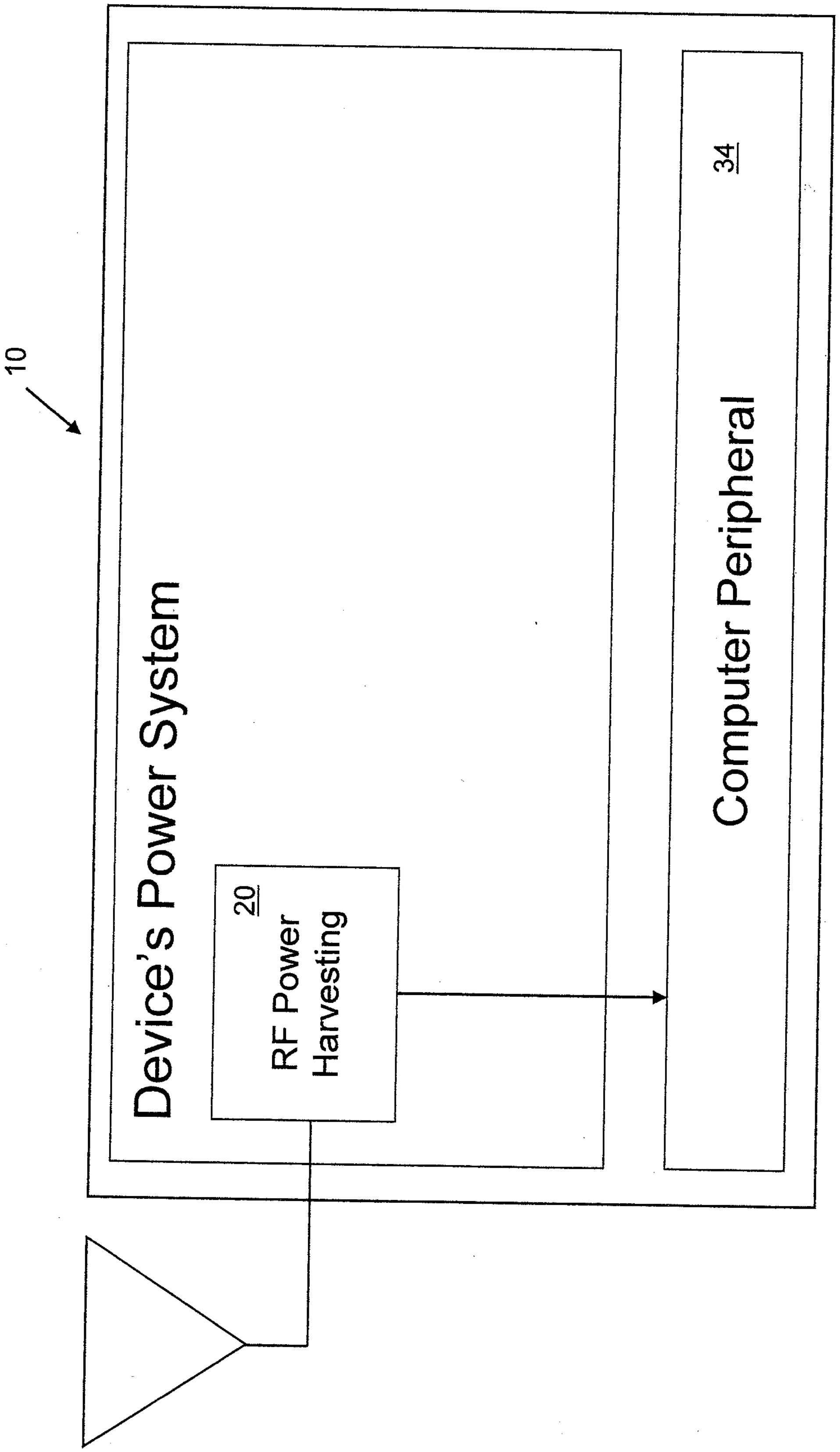
10
Device's Power System
20
RF Power Harvesting
Computer Peripheral
34

Fig. 90

POWERING DEVICES USING RF ENERGY HARVESTING

[0001] This application claims priority to and is a divisional application of U.S. patent application Ser. No. 11/447,412, filed Jun. 6, 2006, and entitled "POWERING DEVICES USING RF ENERGY HARVESTING," which claims priority to U.S. Provisional Patent Application No. 60/688,587, filed on Jun. 8, 2005, and entitled "POWERING DEVICES USING RF ENERGY HARVESTING;" both of which are incorporated herein by references in their entireties.

FIELD OF THE INVENTION

[0002] The present invention is related to the wireless powering of devices. More specifically, the present invention is related to the wireless powering of devices with a power harvester.

BACKGROUND OF THE INVENTION

[0003] As processor capabilities have expanded and power requirements have decreased there has been an ongoing explosion of devices that operate completely independent of wires or power cords. These "untethered" devices range from cell phones, and wireless keyboards to building sensors and active RFID tags.

[0004] Engineers and designers of these untethered devices continue to have to deal with the limitations of portable power sources, primarily batteries as the key design parameter. While performance of processors and portable devices have been doubling every 18-24 months driven by Moore's law, battery technology in terms of capacity has only been growing at measly 6% per year. Even with power conscious designs and the latest in battery technology, many devices do not provide the lifetime cost and maintenance requirements for applications that require a large number of untethered devices such as logistics, and building automation. Today's devices that need two-way communication require scheduled maintenance every three to 18 months to replace or recharge the device's power source (typically a battery). One-way devices simply broadcasting their status (one-way) such as automated utility meter readers have a better battery life, typically requiring replacement within 10 years. For both device types, scheduled power-source maintenance is costly and disruptive to the entire system that a device is intended to monitor and/or control. Unscheduled maintenance trips are even more costly and disruptive. On a macro level, the relatively high cost associated with the internal battery also reduces the practical, or economically viable, number of devices that can be deployed.

[0005] The ideal solution to the power problem for untethered devices is a device or system that can collect and harness sufficient energy from the external environment. The harnessed energy would then either directly power an untethered device or augment a battery or other storage component. Directly powering an untethered device enables the device to be constructed without the need for a battery. Augmenting a storage component could be along two lines: 1) increasing the overall life of the device or 2) by providing more power to the device to increase the functionality of the device. The other parameters for an ideal solution is that the harnessing device could be used in a wide range of environments including harsh and sealed environments (e.g. nuclear reactors), would be inexpensive to produce, would be safe for humans, and would have a minimal effect on the basic size, weight and other physical characteristics of the untethered device.

BRIEF SUMMARY OF THE INVENTION

[0006] The present invention pertains to an apparatus for an application. The apparatus comprises a core device preferably having an integrated circuit for the application. The apparatus comprises a power harvester (preferably a radio frequency (RF) energy harvester) connected to the core device to power the core device.

[0007] The present invention pertains to an apparatus for an application. The apparatus comprises a core device preferably having an integrated circuit for the application. The apparatus comprises means for receiving energy wirelessly and providing power from the energy to the core device to power the integrated circuit of the core device. The receiving means is connected to the core device.

[0008] The present invention pertains to a method for an application. The method comprises the steps of converting RF energy into usable energy. There is the step of preferably powering an integrated circuit of the core device with the usable energy.

[0009] This invention pertains to a technique that uses radio frequency (RF) energy as a source of energy to directly power a device or augment a power storage component in a device. The present invention meets the requirements described in the previous "Background of the Invention" section.

[0010] Traditional RF receiving devices have focused on maximizing selectivity of the frequency to isolate and to be coherent without interference from other sources. In contrast, while the present invention operates at a specific frequency or range of frequencies, the device accepts any interference to supplement the output power of the device. Also, the research related to power harvesting that uses RF energy as a source has primarily focused on devices in close proximity of the source (inductive or near-field energy). In most cases, prior research assumed a dedicated or directed source of RF to power the device.

[0011] The invention should not be confused with power transfer by inductive coupling, which requires the device to be relatively close to the power transmission source. The RFID Handbook by the author Klaus Finkenzeller defines the inductive coupling region as distance between the transmitter and receiver of less than 0.16 times lambda where lambda is the wavelength of the RF wave. The invention can be implemented in the near field (sometimes referred to as inductive) region as well as the far-field region. The far-field region is distances greater than 0.16 times lambda.

[0012] It is an object of this invention to provide a method and apparatus to

- **[0013]** 1. remotely energize an untethered device without using direct wiring
- **[0014]** 2. power or augment the life of the power storage component so it matches the life of the device and, ultimately, powers the off-grid device with or without the use of batteries
- **[0015]** 3. allow untethered devices to be virtually maintenance free
- **[0016]** 4. provide augmentation for other energy harvesting technologies (solar, piezoelectric, etc.)
- **[0017]** 5. provide backup power to tethered devices

[0018] It is a further object of this invention to directly power a device or augment a power storage component in a device in conjunction with other power harvesting technologies and storage elements.

[0019] With this method and apparatus a device's power storage components do not require replacement, thus enabling the device to be permanently placed off-grid, where it may be physically impractical, costly, or dangerous (due to a harsh environment) to provide maintenance.

[0020] For devices on-grid (tethered) or with reliable power sources, RF power harvesting can be used as a backup in case the primary power source is lost.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING

[0021] In the accompanying drawings, the preferred embodiment of the invention and preferred methods of practicing the invention are illustrated in which:

[0022] FIG. **1** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

[0023] FIG. **2** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0024] FIG. **3** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0025] FIG. **4** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0026] FIG. **5** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0027] FIG. **6** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0028] FIG. **7** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0029] FIG. **8** is a block diagram of an RF Power Harvesting block in communication with a Power Storage block.

[0030] FIG. **9** is a block diagram of an RF Power Harvesting block in communication with a Power Storage block and used to supply power to Core Device Components.

[0031] FIG. **10** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0032] FIG. **11** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0033] FIG. **12** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block.

[0034] FIG. **13** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0035] FIG. **14** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

[0036] FIG. **15** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

[0037] FIG. **16** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0038] FIG. **17** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block.

[0039] FIG. **18** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger.

[0040] FIG. **19** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and the RF Power Harvesting block in communication with a Power Storage block.

[0041] FIG. **20** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and Core Device Components.

[0042] FIG. **21** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and Core Device Components and the RF Power Harvesting block in communication with a Power Storage block.

[0043] FIG. **22** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0044] FIG. **23** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0045] FIG. **24** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0046] FIG. **25** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0047] FIG. **26** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0048] FIG. **27** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0049] FIG. **28** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the RF Power Harvesting block supplying power to Core Device Components.

[0050] FIG. **29** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0051] FIG. **30** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

[0052] FIG. **31** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0053] FIG. **32** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power Core Device Components.

[0054] FIG. **33** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or

Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0055] FIG. **34** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0056] FIG. **35** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0057] FIG. **36** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0058] FIG. **37** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0059] FIG. **38** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0060] FIG. **39** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0061] FIG. **40** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to the Power Storage Charger.

[0062] FIG. **41** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0063] FIG. **42** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0064] FIG. **43** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

[0065] FIG. **44** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the RF Power Harvesting block supplying power to Core Device Components and used to supply power to a Power Storage Charger.

[0066] FIG. **45** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0067] FIG. **46** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0068] FIG. **47** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

[0069] FIG. **48** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power Core Device Components and used to supply power to a Power Storage Charger.

[0070] FIG. **49** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0071] FIG. **50** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

[0072] FIG. **51** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0073] FIG. **52** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0074] FIG. **53** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0075] FIG. **54** is a block diagram of an RF Power Harvesting block using Antenna A to directly supply power to Core Device Components.

[0076] FIG. **55** is a block diagram of an RF Power Harvesting block using Antenna A to supply power to a Power Regulation, Storage and/or Storage Charging block.

[0077] FIG. **56** is a block diagram of an RF Power Harvesting block using Antenna A to supply power to a Power Regulation, Storage and/or Storage Charging block and used to supply power to Core Device Components.

[0078] FIG. **57** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

[0079] FIG. **58** is a block diagram of an RF Power Harvesting block used to supply power to a Power Regulation, Storage and/or Storage Charging block.

[0080] FIG. **59** is a block diagram of an RF Power Harvesting block used to supply power to a Power Regulation, Storage and/or Storage Charging block and used to supply power to Core Device Components.

[0081] FIG. **60** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

[0082] FIG. **61** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0083] FIG. **62** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and used to directly supply power to Core Device Components.

[0084] FIG. **63** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0085] FIG. **64** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

[0086] FIG. **65** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

[0087] FIG. **66** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0088] FIG. **67** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

[0089] FIG. **68** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0090] FIG. **69** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0091] FIG. **70** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0092] FIG. **71** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0093] FIG. **72** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0094] FIG. **73** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0095] FIG. **74** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

[0096] FIG. **75** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

[0097] FIG. **76** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0098] FIG. **77** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

[0099] FIG. **78** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0100] FIG. **79** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0101] FIG. **80** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources.

[0102] FIG. **81** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

[0103] FIG. **82** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

[0104] FIG. **83** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0105] FIG. **84** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

[0106] FIG. **85** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0107] FIG. **86** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0108] FIG. **87** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0109] FIG. **88** is a block diagram of an entire power system for the device.

[0110] FIG. **89** is a block diagram of a power harvesting block used to supply power to a core device having a sensor.

[0111] FIG. **90** is a block diagram of a power harvesting block used to supply power to a core device having a computer peripheral.

DETAILED DESCRIPTION OF THE INVENTION

[0112] A complete understanding of the invention will be obtained from the following description when taken in connection with the accompanying drawing figures wherein like reference characters identify like parts throughout.

[0113] There is shown an apparatus **10** for an application. The apparatus **10** comprises a core device **22** preferably having an integrated circuit **36** for the application. The apparatus **10** comprises a power harvester **20** connected to the core device **22** to power the core device **22**.

[0114] The apparatus **10** preferably includes an alternative power source **24** connected to the core device **22** to power the core device **22** in conjunction with the power harvester **20**. Preferably, the apparatus **10** includes a power regulator **26** and/or power storage circuit **28** connected to the power harvester **20**. The apparatus **10** preferably includes a power storage charger **30** connected to the power harvester **20**. Preferably, the apparatus **10** includes a power storage connected to the power harvester **20**.

[0115] Preferably, the core device **22** includes a memory **38** connected to the integrated circuit **36** and to the power harvester **20** to power the memory **36**.

[0116] The core device **22** can include a sensor **32**, as shown in FIG. **89**. The sensor **32** can include a proximity sensor, an intrusion sensor, an environmental sensor, a chemical sensor, a biological sensor, a sensor in contact with an automobile, an occupancy sensor, a motion sensor, a position sensor, a metal detector, or a sensor **32** in contact with an aircraft. The sensor **32** can include an alarm connected to the power harvester **20** to power the alarm, a display connected to the power harvester **20** to power the display, a sensor **32** disposed in a building, an industrial automation sensor, a sensor **32** in contact with an elevator, a temperature sensor, a fire sensor, or an accelerometer.

[0117] The sensor **32** can include a level sensor, a gas level sensor, a fluid level sensor, a light level sensor. The sensor **32** can include a flow sensor, a gas flow sensor, a fluid flow sensor, a light flow sensor, or a plasma flow sensor.

[0118] The sensor **32** can include a pressure sensor, a gas pressure sensor, a fluid pressure sensor, a light sensor, an infrared light sensor, an ultraviolet light sensor, an x-ray sensor, a cosmic ray sensor, a visible light sensor, a gamma ray sensor, a stress sensor, a strain sensor, a depth sensor, or an electrical characteristic sensor.

[0119] The sensor **32** includes a voltage sensor, a current sensor, a viscosity sensor, an acoustical sensor, a sound sensor, a listening sensor, a thickness sensor, a density sensor, a surface quality sensor, a volume sensor, a physical sensor, a mass sensor, a weight sensor, a conductivity sensor, a distance sensor, an orientation sensor, or a vibration sensor.

[0120] The sensor **32** can include a radioactivity sensor, a field strength sensor, an electric field sensor or a magnetic field sensor, a smoke detector, a carbon monoxide detector, a radon detector, an air quality sensor, a humidity sensor, a glass breakage sensor, or a break beam detector. The sensor can include a thermal energy sensor, an electromagnetic sensor, a mechanical sensor, an optical sensor, a radiation sensor, a sensor in contact with a vehicle, or a sensor **32** in contact with a water craft.

[0121] The present invention pertains to an apparatus **10** for an application. The apparatus **10** comprises a core device **22** having an integrated circuit **36** for the application. The apparatus **10** comprises means for receiving energy wirelessly and providing power from the energy to the core device **22** to power the integrated circuit **36** of the core device **22**. The receiving means is connected to the core device **22**. Preferably, the core device **22** includes means for sensing.

[0122] The present invention pertains to a method for an application. The method comprises the steps of converting RF energy into usable energy. There is the step of preferably powering an integrated circuit **36** of the core device **22** with the usable energy.

[0123] Preferably, there is the step of regulating the usable energy provided to the core device **22**. There is preferably the step of storing the usable energy. Preferably, there is the step of providing power to the core device **22** from an alternative power source **24** in conjunction with the usable energy.

[0124] Alternatively, the core device **22** can include a computer peripheral **34**, as shown in FIG. **90**. The computer peripheral **34** can include a handheld game, a gaming system, a game controller, a controller, a keyboard, a mouse, a computer terminal, computer storage, or computer equipment.

[0125] The present invention can be implemented in numerous ways. A number of these ways are depicted in FIGS. **1-90**. These figures contain multiple blocks that are configured in multiple ways.

[0126] In the figures, an arrow represents the flow of power unless otherwise stated. Single-headed (one-way) arrows represent that the power is flowing from one block to another. The single-headed arrow may represent multiple wires that provide power from one block to multiple parts in the other block. Two-headed (two-way) arrows represent a single wire that can have power flow in either direction or multiple wires each having power flow in a single direction.

[0127] As an example, a two-headed arrow between an RF power harvesting **20** block and a power regulation **26** and/or power storage circuit **28** block can represent a single wire that allows harvested power to flow into a storage device such as a capacitor. The same block diagram can also represent two wires between the two blocks with the first wire allowing harvested power to flow into a voltage regulator **26**. The second wire can allow the regulated voltage to feedback to the RF power harvesting **20** block to provide power to internal components such as transistors to increase the performance of the RF power harvesting **20** block.

[0128] Each block is described in detail below. Each block represents the functionality described below associated with it. For instance, the RF power harvesting **20** block describes a power harvester **20**, and the power regulation **26** and/or power storage circuit **28** block describes a power regulator **26** and a power storage circuit **28**.

[0129] Device's Power System Block

[0130] The device's power system block includes all components/circuitry required to power the device. This block may include an RF power harvesting **20** block, a power regulation **26** and/or power storage circuit **28** block, a power storage block, and/or a power storage charger **30** block (all discussed below).

[0131] RF Power Harvesting **20** Block

[0132] The RF power harvesting **20** block is connected to an antenna. The antenna may or may not be used as the communications antenna for the core device **22** components. The RF power harvesting **20** block is used to convert the energy captured by the antenna into usable power, such as direct current (DC) voltage.

[0133] This block may include antenna matching, rectifying circuitry, voltage transforming circuitry, and/or other performance optimizing circuitry. The rectifying circuitry may include a diode(s), a transistor(s), or some other rectifying device or combination. Examples of the rectifying circuitry include, but are not limited to, half-wave, full-wave, and voltage doubling circuits.

[0134] The output of the RF power harvesting **20** block is a DC voltage or current. The RF power harvesting **20** block may accept feedback (or input) from other circuitry or blocks, which may be used to control the harvesting circuitry to improve the performance or vary the output. This feedback may include, but is not limited to, a DC voltage or a clock from the core device **22** components.

[0135] U.S. Pat. No. 6,615,074 (FIGS. 8, 9, 12a, 12b, 13, 14), incorporated by reference, herein, shows numerous examples of RF power harvesting circuits that can be used to implement the block and function described.

[0136] Power Regulation **26** and/or Power Storage Circuit **28** Block

[0137] It may be necessary to regulate the converted power (i.e., hold the power at a constant level) for specific devices. The devices that would need this block require a fairly constant voltage or current. Deviations from the required values may cause the device to not perform within its specifications.

[0138] The regulation can be implemented in many different ways. The block can be as simple as using a Zener diode, or as complicated as using an integrated circuit such as a linear voltage regulator **26** or switching regulator **26** to hold the voltage at a constant level.

[0139] Certain devices have a more tolerable power requirement. For these devices, the regulation stage may be excluded.

[0140] This block may also include, with or without the regulation, a storage device such as a capacitor, a battery, or some other device able to store charge. The output from the power regulation **26** and/or power storage circuit **28** block may be used as feedback to other blocks within the device's power system block or to the alternative power sources **24** block (described below), if they require a regulated supply voltage or stored power.

[0141] U.S. Pat. No. 6,894,467 (FIGS. 1, 3), Linear Voltage Regulator, incorporated by reference, herein, is an example of a practical application of implementing the regulation described in the block. U.S. Pat. No. 6,297,618 (FIGS. 1-4), Power storage device and method of measuring voltage of storage battery, incorporated by reference, herein, is an example of a practical application of implementing the storage described in the block.

[0142] Power Storage Charger **30** Block

[0143] The power storage charger **30** block may be needed if a power storage component of the device requires a special charging mechanism, such as pulse charging or trickle charging. This block controls how the captured and converted power is supplied to the power storage component.

[0144] U.S. Pat. No. 6,836,095 (FIGS. 1-3), Battery Charging Method and Apparatus, incorporated by reference herein, is an example of a practical application of implementing the special charging mechanism described in the block.

[0145] Power Storage Block

[0146] If a device has intermittent power requirements, it may be necessary to store the captured power for use at a later time. The power can be stored in the power storage block, which could include a battery, a capacitor, and/or another type of power storage component. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other power storage elements.

[0147] The output from the power storage block may be used as feedback to other blocks within the device's power system or to the alternative power sources **24** block, if they require a dedicated and predictable supply voltage.

[0148] U.S. Pat. No. 6,297,618 (FIGS. 1-4), Power Storage Device and Method of Measuring Voltage of Storage Battery, incorporated by reference herein, is an example of a practical application of implementing the storage described in the block. U.S. Pat. No. 6,835,501, Alkaline Rechargeable Battery, incorporated by reference herein, is also an example of a practical application of implementing the storage described in the block.

[0149] Core Device **22** Components Block

[0150] The core device **22** components block is the portion of the device that is receiving power from the device's power system. This block may be, but is not limited to, the devices listed in the subsequent pages of this document. It may be advantageous for the core device **22** components block to communicate with any of the blocks that are supplying power to it. This communication can include, but is not limited to, a feedback control signal such as a clock or an ON/OFF command. As an example, the device may want to turn off the alternative power sources **24** block if it is receiving sufficient power from the RF power harvesting **20** block.

[0151] Alternative Power Sources **24** Block

[0152] RF energy harvesting also has the ability to be augmented by other types of power harvesting, storage components, or dedicated sources (e.g. power line). The alternative power sources **24** block shows how this type of system could be implemented. The augmenting power harvesting technologies include, but are not limited to, solar, light (visible and non-visible), piezoelectric, vibration, acoustic, thermal, microgenerators, wind, and other environmental elements. This block can work independently or have communication with other blocks.

[0153] U.S. Pat. No. 6,784,358, Solar Cell Structure Utilizing an Amorphous Silicon Discrete By-Pass Diode, incorporated by reference herein, is an example of a practical application of implementing an alternative power source **24** described by the block. U.S. Pat. No. 6,858,970, Multi-Frequency Piezoelectric Energy Harvester, incorporated by reference herein, is also an example of a practical application of implementing an alternative power source **24** described by the block.

[0154] Power Regulation **26**, Storage **28** and/or Storage Charging **30** Block

[0155] The power regulation **26**, storage **28** and/or storage charging **30** block contains all the combinations of the power regulation **26** and/or power storage circuit **28** block, power storage charger **30** block, and power storage block. This block is used in the later figures to reduce the number of figures needed to show how the blocks can interconnect.

[0156] The disclosed invention is an apparatus and method for an application for retrieving radio frequency (RF) energy by an antenna, converting that energy into direct current (DC) power, regulating that energy using an optimized circuit, storing that energy in an optimized component, and/or supplying the power for a specific device.

[0157] Retrieval of RF Energy

[0158] The RF energy is retrieved from the environment by the use of the antenna. The antenna can be shared or standalone with respect to an antenna used for the device's wireless communication. FIGS. **54-56** show a device that has an antenna A for use by the RF power harvester **20** and an antenna B used for wireless one-way or two-way communication. FIGS. **57-59** show a device where the antenna is shared by both a device's communication module (within the core device **22** components) and the RF power harvester **20**. In terms of form factor, the antenna used by the apparatus **10** can be a separate component or integrated directly into the form factor of the device.

[0159] The antenna is able to capture two types of available RF energy. The first type of energy exists as ambient RF energy. This type of RF surrounds us in our day-to-day lives and is usually generated to carry one way or two-way combinations of voice, video and data communications. The sources that the antenna can harvest from include, but are not limited to, medium-frequency AM radio broadcast, very-high-frequency (VHF) FM radio broadcast and television broadcast, ultra-high-frequency (UHF) broadcast, cellular base stations, wireless data access points, super-high-frequency (SHF) frequencies, and the industrial, scientific, and medical (ISM) bands. These sources cover transmission frequencies from 300 kHz to 30 GHz.

[0160] The second type of energy available is directed RF energy. This type of RF energy is directed from a transmitter specifically designed to deliver RF energy for harvesting by the antenna. The transmitter can be configured as a standalone device or integrated into an existing device.

[0161] Conversion of the Energy into DC

[0162] The RF energy captured by the antenna must be converted into a useful form of energy for the specific device. This conversion is shown in block form in all FIGS. **1-90**) as the RF power harvesting **20** block. The most common form of useable energy is DC energy. To perform this conversion, the block includes circuitry to rectify the captured alternating current (AC) energy to create DC energy. The rectification in this block can be done with a diode(s), a transistor(s), or some other rectifying device or combination.

[0163] Regulation of the Energy

[0164] It may be necessary to regulate the converted power (hold the power at a constant level) for specific devices. FIGS. **2-7**, **10-17**, and **22-53** show how this regulation can be implemented using a power regulation **26** and/or power storage circuit **28** block. The devices that would need this block require a fairly constant voltage or current. Deviations from the required values may cause the device to not perform within its specifications. The regulation can be implemented in many different ways. The block can be as simple as using a Zener diode, or as complicated as using an integrated circuit such as a linear voltage regulator **26** or a switching regulator **26** to hold the voltage at a constant level. Certain devices have a more tolerable power requirement. For these devices, the regulation stage may be excluded.

[0165] Storage of the Energy

[0166] If a device has intermittent power requirements, such as the devices exampled by FIGS. **2-53**, it may be necessary to store the captured power for use at a later time. The power can be stored in the power storage block or the power regulation **26** and/or power storage circuit **28** block. Storage devices can include, but are not limited to, a battery, a capacitor, or another type of power storage component. In certain applications, it may be necessary to include additional circuitry that controls how the power is transferred to the storage device. The power storage charger **30** block is shown in FIGS. **18-53**. This may be needed if the power storage component requires a special charging mechanism such as pulse charging or trickle charging. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other storage elements. There are devices that will not require storage. These devices can run directly off of the converted power. These devices also may or may not require regulation of the captured power.

[0167] Supplying the Power

[0168] The captured DC power, which may or may not be regulated and/or stored, is supplied to the device, which is represented by the core device **22** components block in the figures. This may be a single connection or it may supply multiple parts of the device with power.

[0169] RF energy harvesting also has the ability to be augmented by other types of power harvesting or storage components. Other power harvesting technologies include, but are not limited to, solar, light (visible and non-visible), piezoelectric, vibration, acoustic, thermal, microgenerators, wind, and other environmental elements. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other storage elements. FIGS. **60-88** show how the alternative power sources **24** block can be connected to an RF energy harvesting system. These figure show how the RF energy harvesting **20** block and the alternative power sources **24** block can work independently or have communication with each other. The antenna configurations shown in FIGS. **54-59** are still applicable with the addition of an alternative power sources **24** block. These antenna configurations can be applied to FIGS. **60-88**.

[0170] RF energy harvesting also has the ability to provide a backup to devices on-grid (tethered) or with reliable power sources, which can be used in case the primary power source is lost. As an example, it may be mandated by regulations that a sensor has auxiliary power in case the primary supply is lost. It could be possible to use a rechargeable battery that obtains its charge from the primary supply when in operation. However, if the primary supply is lost for a time greater than the life of the rechargeable battery, the specification of uninterrupted power is not met. RF energy could be used to supply power to the described device while the primary supply is not available. The primary supply could include, but is not limited to, an on-grid connection, a generator, a battery, or other reliable power supply.

[0171] RF energy harvesting with or without alternative source augmentation is applicable to provide electric power directly or indirectly to a range of electronic components contained in any specific electrical or electronic device and includes, but is not limited to:

[0172] Passive electronic components, active electronic components;

[0173] Resistors, fixed resistors, variable resistors, thermistors, thyristor, thermocouple;

[0174] Capacitors, Electrolytic Capacitors, Tantalum Capacitors, Ceramic Capacitors, Multilayer Ceramic Capacitors, Polystyrene Film Capacitors, Electric Double Layer Capacitors (Super Capacitors), Polyester Film Capacitors, Polypropylene Capacitors, Mica Capacitors, Metallized Polyester Film Capacitors, Variable Capacitors;

[0175] Diodes, Voltage regulation diodes, light-emitting diodes, organic light-emitting diodes, Variable capacitance diodes, Rectification diodes, Switching diodes, Regulation Diodes, Diode bridges, Schottky barrier diodes, tunnel diodes, PIN diodes, Zener diodes, Avalanche diodes, TVSs;

[0176] Integrated circuits, microcontroller unit (MCU), microprocessor unit (MPU), logic circuits, memory, printed circuits, circuit boards, printed wiring boards;

[0177] Transistors, MOSFETs, FETs, BJTs, JFETs, IGBTs, Relays, Antennas, semiconductors, conductors, inductors, relays, diacs, triacs, SCRs, MOVs;

[0178] Fuses, circuit breakers;

[0179] Batteries, Non-rechargeable batteries, rechargeable batteries, coin cell batteries, button cell batteries, alkaline batteries, lithium batteries, lithium ion batteries, lithium polymer batteries, NIMH batteries, NICAD batteries, Lead acid batteries, Zinc air batteries, Manganese Lithium batteries, Niobium Titanium Lithium batteries, Vanadium Pentoxide Lithium batteries, Carbon Zinc batteries, Zinc Chloride batteries, Lithium Thionyl Chloride batteries, Manganese Dioxide batteries, Lithium Poly-Carbonmonofluoride batteries, Lithium Manganese Dioxide batteries, Lithium Chloride batteries, Lead Acid Calcium batteries, Lead Acid Tin batteries, Oxy Nickel batteries, Silver Oxide batteries, Magnesium batteries;

[0180] Inductors, Coils, High Frequency Coils, Toroidal Coils, Transformers, switches, chokes;

[0181] Motors, DC motors, stepper motor, AC motors, Fans;

[0182] Crystals, Oscillators, Clocks, Timers; and

[0183] Displays, LCDs, LED displays.

[0184] RF energy harvesting with or without alternative source augmentation is applicable across a range of markets and specific devices and includes but is not limited to:

[0185] Consumer Electronics:

[0186] Electronic equipment, wired devices, battery powered devices, wireless communication devices, cell phones, telephones, phones, cordless phones, portable phones, Bluetooth devices, Bluetooth headsets, hands-free headsets, headsets, headphones, Wireless headsets, radios, AM/FM radios, shortwave radios, weather radios, Two-way radios, portable radio, lights, lanterns, portable lights, flashlights, nightlights, spotlights, search lights, calculators, graphing calculators, desk calculators, clocks, alarm clocks, wall clocks, desk clocks, travel clocks, watches, wristwatches, pocket watches, stop watches, timers, voice recorders, Dictaphones, laser pointers, power tools, cordless power tools, electronic razors, electric razors, handheld games, gaming systems, game controllers, wireless game controllers, remote controls, battery chargers, computers, portable computers, keyless entry, toys, toy guns, toy laser guns, games, microphones, musical instruments, musical effects processors, musical instrument tuners, metronomes, electronic chord charts, door openers, garage door openers, PDA, Cameras, Video recorders, Multimeter, electronic test equipment, hand-held electronics, portable electronics, wireless pens, sound generators, noise generators, language translators, electric toothbrushes, portable televisions, pagers, transceivers, toy vehicles, remote control vehicles, toy planes, remote control planes, pet containment systems, invisible fence pet sensors, memory backup, base station battery backups, appliance battery backups, uninterrupted power supplies, GPS devices, memory retention power supplies, metal detectors, stud finders, metal stud finders, stun guns, tazers, wearable devices, baby monitors, intercoms, doorbells, wireless doorbells, electronic office supplies, electronic staplers, radar jammers, radar detectors, digital scales, microfilm cassettes, video head testers, compasses, noise canceling headphones, air samplers, depth finders, barometers, weather measurement instruments, data transfer devices, automatic distress signaling unit, Wireless audio speakers, Satellite radios, Police scanners, Car navigation systems (GPS devices), Decorative lights, Christmas lights, garden lights, lawn lights, ornamental lights, porch lights;

[0187] Multi-media players: MP3, DVD, analog music players, CD players, tape players, digital music players, digital video players, minidisc; and

[0188] Computer: keyboards, mice, peripherals, computer equipment, electronic computers, computer storage, computer terminals;

[0189] Building/Industrial Automation:

[0190] Sensors: Position, elevator, temperature, fire, accelerometers, level, gas level, fluid level, light level, flow, gas flow, fluid flow, light flow, plasma flow, pressure, gas pressure, fluid pressure, motion, light, infrared light, ultraviolet light, X-rays, cosmic rays, visible light, gamma rays, chemical, stress, strain, depth, electrical characteristics, voltage, current, viscosity, acoustical, sound, listening, thickness, density, surface quality, volume, physical, mass, weight, force, conductivity, distance, orientation, vibration, radioactivity, field strength, electric field strength, magnetic field strength, occupancy, smoke detector, carbon monoxide detector, radon detector, air quality, humidity, glass breakage, break beam detector;

[0191] Controls: Position, elevator, temperature, fire, accelerometers, level, gas level, fluid level, light level, flow, gas flow, fluid flow, light flow, plasma flow, pressure, gas pressure, fluid pressure, motion, light, infrared light, ultraviolet light, X-rays, cosmic rays, visible light, gamma rays, chemical, stress, strain, depth, electrical characteristics, voltage, current, viscosity, acoustical, sound, listening, thickness, density, surface quality, volume, physical, mass, weight, force, conductivity, distance, orientation, vibration, radioactivity, field strength, electric field strength, magnetic field strength, occupancy, smoke detector, carbon monoxide detector, radon detector, air quality, humidity; and

[0192] Devices: Thermostats, light switches, door locks, smart-card door locks, lighting, emergency lighting, motion lighting, safety lighting, highway lighting, construction lighting, sign lighting, roadway sign lighting, construction sign lighting, automatic flushing units, automatic soap dispenser, automatic paper towel dispenser, automatic faucets, automatic door sensors, identification reader, fingerprint reader, credit card readers, card readers, valve actuators, gauges, analog gauges, digital gauges, fire extinguishers, wireless switches, remotely operated inspection equipment, gas/oil pipeline monitoring systems, robotic pipeline inspection gauges, “auto-reclosers” for electric power lines, sonar buoys, telemetry systems, electronic record tracking systems, robbery tracking devices, interrogators, programmers, emergency exit alarms, alarms, flood alarms, gas alarms, electronic entry systems, security keypads, silo transducers, data recorders, signal tracers, anti-static strap testers, radiosonde weather balloons, utilities load controllers, profilometers, noise cancellation equipment, infrared beacons;

[0193] Military/Government:

[0194] Tracking tags: Weapons, vehicle, soldier, gear/assets, staff, general population, security access badges;

[0195] Sensors: Proximity, intrusion, environmental, chemical/biological; and

[0196] Equipment: Battery charger, surveillance, card readers, identification reader, fingerprint reader, retinal scanners, satellites, rockets, space vehicles, search and rescue transponders (SARTs), emergency position-indicating rescue beacons (EPIRBs), emergency locator transmitters (ELTs), military radios, electronic toll collection systems, postal tracking systems, communications, thermal imaging, night vision, training targets, field medical equipment, house arrest monitors, laser tags, electronic parking meters, multiple integrated laser engagement system, munitions and mines, ship sensors;

[0197] Utility:

[0198] Gas consumption meters, water consumption meters, and electric consumption meters;

[0199] Logistics & Supply Chain Management:

[0200] Radio-frequency identification devices (RFID), RFID readers;

[0201] Tracking: Asset tags, cargo container location beacons, transponders, transceivers; and

[0202] Devices: Smart price tags, smart shelving, hand-held barcode scanner, barcode scanners, credit card readers, card readers, retail signage, hotel door locks;

[0203] Homeland Security:

- **[0204]** Sensors: Occupancy, proximity, environmental, chemical/biological, motion, position; and
- **[0205]** Metal detector wand;

[0206] Medical:

- **[0207]** Implantable: cochlear implants, neural stimulators, pace makers, medication administration, defibrillator;
- **[0208]** Body function monitors: pressure, temperature, respiration, blood oxygenation, insulin, hearing aid, pulse, EKG, heart, Holter;
- **[0209]** Tracking tags: Patient, baby identification, assets, supplies, staff, medication, instruments;
- **[0210]** Devices: Home healthcare equipment, ambulatory infusion pumps, blood analyzers, biofeedback systems, bone growth stimulators, thermometers, digital thermometers, stimulators, galvanic stimulators, muscle stimulators, pediatric scales; and

[0211] Agriculture—Livestock Tracking and Asset Tracking:

- **[0212]** Tracking: livestock, asset, wildlife tracking devices; and
- **[0213]** Equipment: cattle prods;

[0214] Automotive:

- **[0215]** Automotive antennas, Automotive Audio Systems, Automotive Lighting, Automotive Video Systems, Computers, Processors, Controls, Switches, Electric Motors, Actuators, Ignition Systems, Starter Systems, Injection Systems, Powertrain Electronics, Radar Detectors, Proximity Detectors, Safety Systems, Security Systems, Sensors, Regulators, Distributors, Vehicle Control, Wiper Systems, Washer Systems, Radio, Video Systems, Entertainment Systems, Navigation Systems, GPS systems, Power Mirror Systems, Emission control systems;

[0216] Appliances:

- **[0217]** Monitoring systems and control systems for major and small appliances including washing machines, dryers, refrigerators, freezers, coolers, air conditioners, humidifiers, dehumidifiers, air purifiers, air filters, fans, furnaces, water heaters, boilers, space heaters, sowing machines, ice makers, microwave ovens, convection ovens, ovens, toaster ovens, ranges, range hoods, cooktops, stoves, stovetops, crock pots, hot plates, dishwashers, garbage disposals, can openers, vacuum cleaners, blenders, mixers, food processors, irons, coffee makers, toasters, grills, hair dryers, electric tooth brushes, electric razors, electric drills, electric screwdrivers, chainsaws, lawnmowers, push mowers, riding mowers, trimmers, brush cutters, pruners, edgers, vending machines;

[0218] Ventilation, Heating, Air-Conditioning, and Commercial Refrigeration Equipment:

- **[0219]** Monitoring systems, control systems;

[0220] Engine, Turbine, and Power Transmission Equipment;

[0221] Monitoring systems, control systems;

[0222] Other General Purpose Machinery Manufacturing:

- **[0223]** Monitoring systems, control systems;

[0224] Telecommunications:

- **[0225]** Monitoring systems, control systems;
- **[0226]** Portable;

[0227] Aircraft:

- **[0228]** Monitoring systems, control systems, actuator systems, sensors.

[0229] It should be noted that devices within a specific category may be applicable across multiple areas even if they are not specifically listed. (e.g., temperature sensors apply to Industrial and Building Automation and Appliances).

[0230] To retrofit or redesign the devices listed, it is possible to implement the described systems in numerous ways. It may be advantageous to leave the device design as-is including the existing power supply. As an example, a device may use non-rechargeable batteries to operate. The device will most likely have a protection circuit to prevent damage if the batteries are installed incorrectly. The protection mechanism is commonly a diode inline with the positive terminal of the battery. In this case, the RF power harvester **20** with or without the alternative power source **24** could be inserted, with an antenna, into the device. The power generated by the RF power harvester **20** (and the alternative power source **24**, if applicable) could be connected to the device after the protection mechanism described to avoid potential charging of a non-rechargeable battery.

[0231] Another way to configure the device's power system is to replace the non-rechargeable batteries with rechargeable batteries. In this instance, the output from the RF power harvester **20** (and alternative power source **24**) could be connected to either side of the protections device. If the connection is before the protection mechanism, the RF power harvester **20** (and the alternative power source **24**, if applicable) will recharge the battery and supply power to the device. If the connection is after the protection mechanism, the RF power harvester **20** (and the alternative power source **24**, if applicable) will supply power to the device and the battery will supply any extra power needed that could not be supplied by the RF power harvester **20** (and the alternative power source **24**, if applicable). It should be noted that the protection device in this case is not needed for proper operation. Its only function would be to protect the batteries from being installed incorrectly. An antenna could be contained inside or placed on the outside of the device.

[0232] Another configuration of the device's power system is to remove the existing batteries and install the RF power harvester **20** (and the alternative power source **24**) in the enclosure provided for the batteries. An antenna could be contained inside or placed on the outside of the device.

[0233] Yet another method of configuring the device's power system would be to reduce the number of batteries and replace them with the RF power harvester **20** (and the alternative power source **24**). In this case, the output from the system would be connected to the batteries in series or parallel depending on the original battery configuration. An antenna could be contained inside or placed on the outside of the device.

[0234] An additional option, would be to completely redesign the product and integrate the required circuitry and storage components into the device. This method is probably the most advantageous because it can fully take advantage of the benefits offered by the RF power harvester **20** (and the alternative power source **24**). An antenna could be contained inside or placed on the outside of the device.

[0235] If the RF power harvester **20** (and the alternative power source **24**) is used as a backup to the primary power supply, a switch could be implemented into the system in order to switch the RF power harvester **20** (and the alternative power source **24**) on when the primary source is lost. In this case, an antenna could be contained inside or placed on the outside of the device.

[0236] To show the flexibility of RF energy harvesting, several products were retrofitted to include an RF power harvester **20** (energy harvesting circuitry). These products include a wireless keyboard, a wall clock, and a desk calculator.

[0237] The wireless keyboard is an example of recharging and augmenting a battery to supply power to a device. This system is shown in FIG. **13**. The output from the regulation circuitry recharges the battery and supplies power to the keyboard. The battery is also used to supply power to the keyboard. The keyboard also includes a separate antenna for receiving power and for data communications. The antenna configuration can be seen in FIG. **55**.

[0238] The wall clock is an example of a direct powering system. The wall clock was retrofitted to include an RF power harvester **20** and the internal AA battery was removed. This system is shown in FIG. **2**. The wall clock did not need regulation, but did require a capacitor for storage to supply the pulse of power to move the second hand.

[0239] The calculator is an example of using RF energy harvesting with another energy harvesting technology. The calculator had an internal 1.5V coin cell battery and a small solar panel. The internal battery was removed, however, the solar panel was left intact. This system is shown in FIG. **60**. In this system, the calculator can receive power from both the solar panel and the RF power harvester **20** to eliminate the need for a battery.

[0240] As an additional example, an RF power harvester **20** (energy harvesting circuit) similar to the ones shown in U.S. Pat. No. 6,615,074 (FIGS. 8, 9, 12a, 12b, 13, 14), was connected in series with a 0.5V solar cell. Individually, the solar cell was able to provide 0.480V to a 10 kilo-ohm resistor, which was being used to simulate the core device **22** components. This corresponds to 23 microwatts. The RF power harvesting circuit by itself was able to provide 2.093V across the 10 kilo-ohm resistor when being supplied by 1 milliwatt of RF power. This corresponds to 438 microwatts. The two circuit outputs were then combined in series by connecting the output from the RF energy harvesting circuit to the ground of the solar cell. The output of the solar cell was then connected to the resistor. The other end of the resistor was connected to the ground of the RF energy harvesting circuit. The voltage across the resistor with the circuits connected as shown in FIG. **63** was 2.445V. This corresponds to 598 microwatts.

[0241] As can be seen, the combination of the two technologies produces a result higher than the addition of the individual powers. From this, it can be determined that the two technologies can cooperate in a way that produces favorable results.

[0242] In the example given, the solar cell produces current to supply the load and helps to bias the RF rectifying diodes, which allows the RF energy harvesting circuit to operate a higher efficiency. The solar cell also changes the impedance seen by the RF energy harvesting circuit, which produces a beneficial result. To be more specific, when examining the power output of the individual circuit (solar and RF power harvesting), the sum of the power captured by the individual circuits was 23 μW+438 μW=461 μW. However, when the two circuits are combined and are allowed to work in conjunction with one another, the output power becomes 598 μW. This result shows that combining the two power-harvesting technologies produces a 30 percent increase in the output power for this example. This same technique can be applied to multiple energy harvesting technologies to produce even greater output power. The equations for this example are shown below.

[0243] Individual Circuits

$$P_I=P_1+P_2+\ldots+P_N$$

[0244] Combined Circuits

$$P_C>P_I=P_1+P_2+\ldots+P_N$$

[0245] where P_I is the sum of the individual output powers

[0246] P_C is the output of the combined circuit

[0247] P_1 is the output power from the first power harvesting technology

[0248] P_2 is the output power from the second power harvesting technology

[0249] P_N is the output power from the N^{th} power harvesting technology

[0250] N is the number of power harvesting technologies or circuits.

[0251] As can be seen by the previous examples, RF energy harvesting can be used alone or in conjunction with alternative power sources to power a wide range of devices. The addition of RF energy harvesting technology to the device allows for increased battery life, increased functionality, or the removal of the primary battery.

[0252] Although the invention has been described in detail in the foregoing embodiments for the purpose of illustration, it is to be understood that such detail is solely for that purpose and that variations can be made therein by those skilled in the art without departing from the spirit and scope of the invention except as it may be described by the following claims.

1-87. (canceled)

88. An apparatus, comprising:

a power harvester configured to convert energy received via an antenna to a DC voltage;

a power storage device configured to store a power associated with the DC voltage;

a sensor configured to operate based on the DC voltage, the sensor configured to detect an environmental condition associated with a location of the sensor, the sensor being collocated with the power harvester; and

an integrated circuit configured to operate based on the DC voltage, the integrated circuit configured to perform an application associated with the environmental condition detected by the sensor.

89. The apparatus of claim **88**, wherein:

the DC voltage is a first DC voltage, the integrated circuit is configured to operate based on the first DC voltage or a second DC voltage from a power source.

90. The apparatus of claim **88**, wherein:

the power associated with the DC voltage is a first power associated with the DC voltage; and

the integrated circuit is configured to produce a control signal that causes a second power associated with the DC voltage to be transferred from the power harvester to the integrated circuit.

91. The apparatus of claim **88**, wherein:

the sensor and the integrated circuit are each configured to operate based on the power associated with the DC voltage stored in the power storage device; and the power harvester is configured to be disabled when the power associated with the DC voltage stored in the power storage device is sufficient for at least one of the sensor or the integrated circuit to operate.

92. The apparatus of claim **88**, further comprising:

a power storage charger configured to control storage of the power associated with the DC voltage at the power storage device, the power storage charger configured to operate in a power pulse charging mode.

93. The apparatus of claim **88**, further comprising:

a power storage charger configured to control storage of the power associated with the DC voltage at the power storage device, the power storage charger configured to operate in a power pulse charging mode at a first time, the power storage charger configured to operate in a power trickle charging mode at a second time.

94. The apparatus of claim **88**, wherein:

the power harvester is configured to receive at least one of an ambient radio frequency energy, a directed radio frequency energy, or an energy from a power source connected to the power harvester.

95. The apparatus of claim **88**, further comprising:

a power storage charger operatively coupled to the power harvester, the power harvester configured to provide an activation signal to power storage charger to trigger the storage of the power associated with the DC voltage at the power storage device.

96. The apparatus of claim **88**, wherein:

the integrated circuit is configured to provide a deactivation signal to a power source when the power associated with the DC voltage is more than a predetermined power value, the deactivation signal configured to impede output of a power from the power source, the power source different from the power harvester and the power storage device.

97. The apparatus of claim **88**, wherein the power storage device is a first power storage device, the apparatus further comprising:

a second power storage device operatively coupled to the first power storage device and the integrated circuit, the second power storage device configured to store a portion of the power associated with the DC voltage and provide the portion of the power associated with the DC voltage to at least one of the sensor or the integrated circuit.

98. An apparatus, comprising:

a power harvester configured to convert energy received via an antenna to a first DC voltage, the power harvester configured to be biased by a second DC voltage produced by a power source;

a sensor configured to operate based on the first DC voltage, the sensor and the power harvester being collocated; and

an integrated circuit configured to operate based on the first DC voltage, the integrated circuit configured to perform an application associated with the sensor.

99. The apparatus of claim **98**, further comprising:

a power storage device configured to store a power associated with the first DC voltage.

100. The apparatus of claim **98**, further comprising:

a power storage charger operatively coupled to the power harvester, the power harvester configured to provide an activation signal to the power storage charger to trigger the storage of a power associated with the first DC voltage at a power storage device.

101. The apparatus of claim **98**, wherein:

the integrated circuit is configured to provide a deactivation signal to the power source when a power associated with the first DC voltage is more than a predetermined power value.

102. The apparatus of claim **98**, wherein:

the integrated circuit is configured to provide an activation signal to the power source when a power associated with the DC voltage is less than a predetermined power value, the activation signal configured to initiate output of a power from the power source.

103. An apparatus, comprising:

a power harvester configured to convert energy received via an antenna to a DC voltage to produce a power associated with the DC voltage;

a regulator coupled to the power harvester and configured to regulate the power associated with the DC voltage, the regulator configured to transfer the regulated power to a power storage charger;

a power storage device configured to store power;

a power storage charger coupled to the power harvester and regulator, the power storage charger configured to control the storage of the regulated power associated with the DC voltage to the power storage device; and

an integrated circuit configured to operate based on the power received from the power harvester or from the power storage device,

the power harvester configured to transfer power associated with the DC voltage to the regulator when the power harvesting is in a first mode, the power harvester configured to transfer power associated with the DC voltage to the power storage device when the power harvester is in a second mode, the power harvester is configured to transfer power associated with the DC voltage to the integrated circuit when the power harvester is in a third mode.

104. The apparatus of claim **103**, wherein:

the power storage charger is configured to operate in a power pulse charging mode at a first time and a power trickle charging mode at a second time.

105. The apparatus of claim **103**, further comprising:

a sensor configured to operate based on the power received from the power harvester or from the power storage device, the sensor configured to detect an environmental condition associated with a location of the sensor or a physical condition associated with an item coupled to the sensor, the integrated circuit configured to perform an application associated with the environmental condition or the physical condition detected by the sensor.

106. The apparatus of claim **103**, wherein:

the regulator is configured to send a control signal that causes the power associated with the DC voltage to be transferred from the power harvester to the regulator.

107. The apparatus of claim **103**, wherein:

the power storage device is configured to send a control signal that causes the power associated with the DC voltage to be transferred from the power harvester to the power storage device.

* * * * *

US 20110077714A1

(19) **United States**

(12) **Patent Application Publication**

Rofougaran

(10) **Pub. No.: US 2011/0077714 A1**

(43) **Pub. Date: Mar. 31, 2011**

(54) **BIO-MEDICAL UNIT WITH POWER HARVESTING MODULE AND RF COMMUNICATION**

(75) Inventor: **Ahmadreza (Reza) Rofougaran**, Newport Coast, CA (US)

(73) Assignee: **BROADCOM CORPORATION**, Irvine, CA (US)

(21) Appl. No.: **12/783,649**

(22) Filed: **May 20, 2010**

Related U.S. Application Data

(60) Provisional application No. 61/247,060, filed on Sep. 30, 2009.

Publication Classification

(51) **Int. Cl.**

A61N 1/378	(2006.01)
A61N 1/08	(2006.01)
A61N 1/372	(2006.01)

(52) **U.S. Cl.** **607/60**; 607/61

(57) **ABSTRACT**

A bio-medical unit includes a power harvesting module and an RF communication module. The power harvesting module generates a power supply voltage from the electromagnetic signal. The RF communication module interprets a predetermined pattern of the transmitting of the electromagnetic signal to determine null intervals; receives a downstream RF communication signal during at least one of the null intervals; converts the downstream RF communication signal into downstream information; converts upstream information into an upstream RF communication signal; and/or transmits the upstream RF communication signal during one or more of the null intervals.

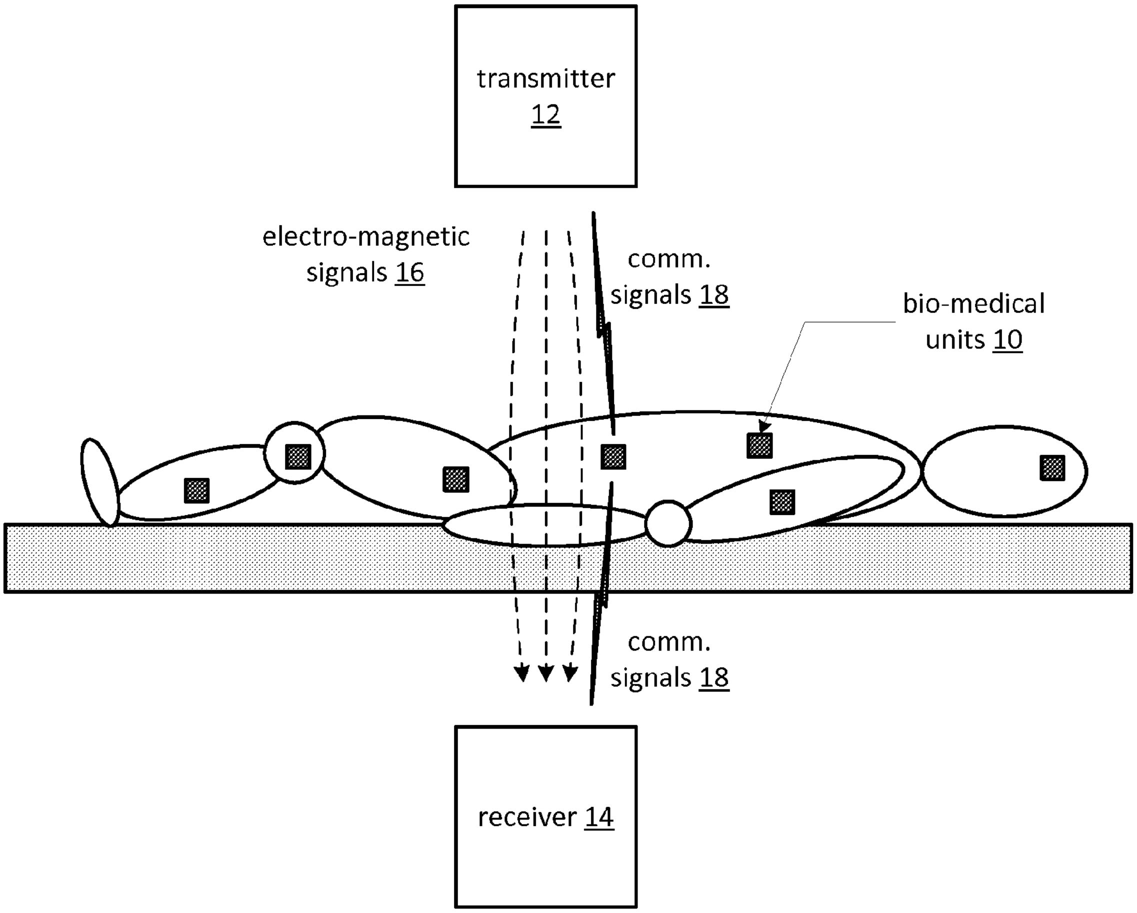

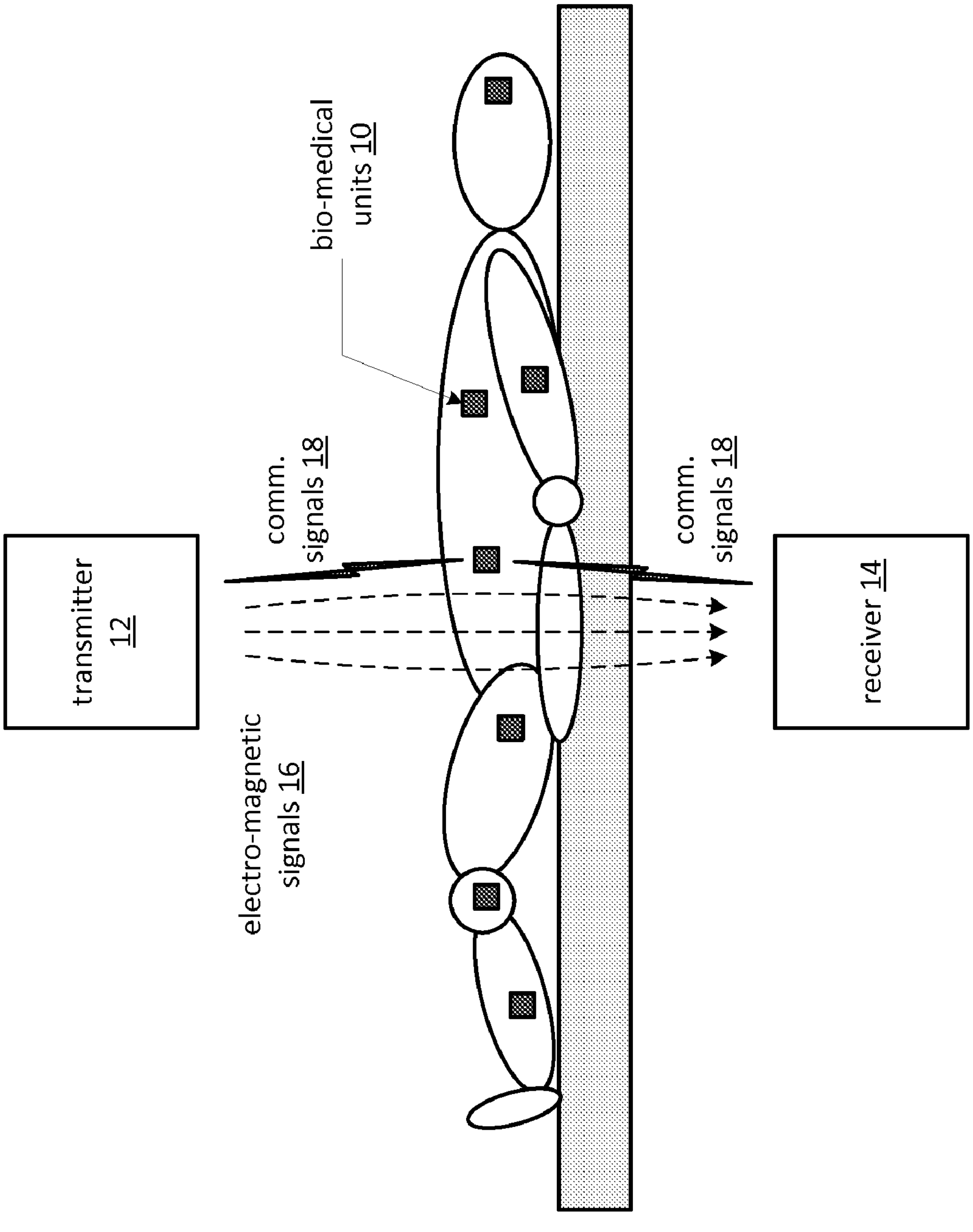
transmitter 12
electro-magnetic signals 16
comm. signals 18
bio-medical units 10
comm. signals 18
receiver 14

FIG. 1

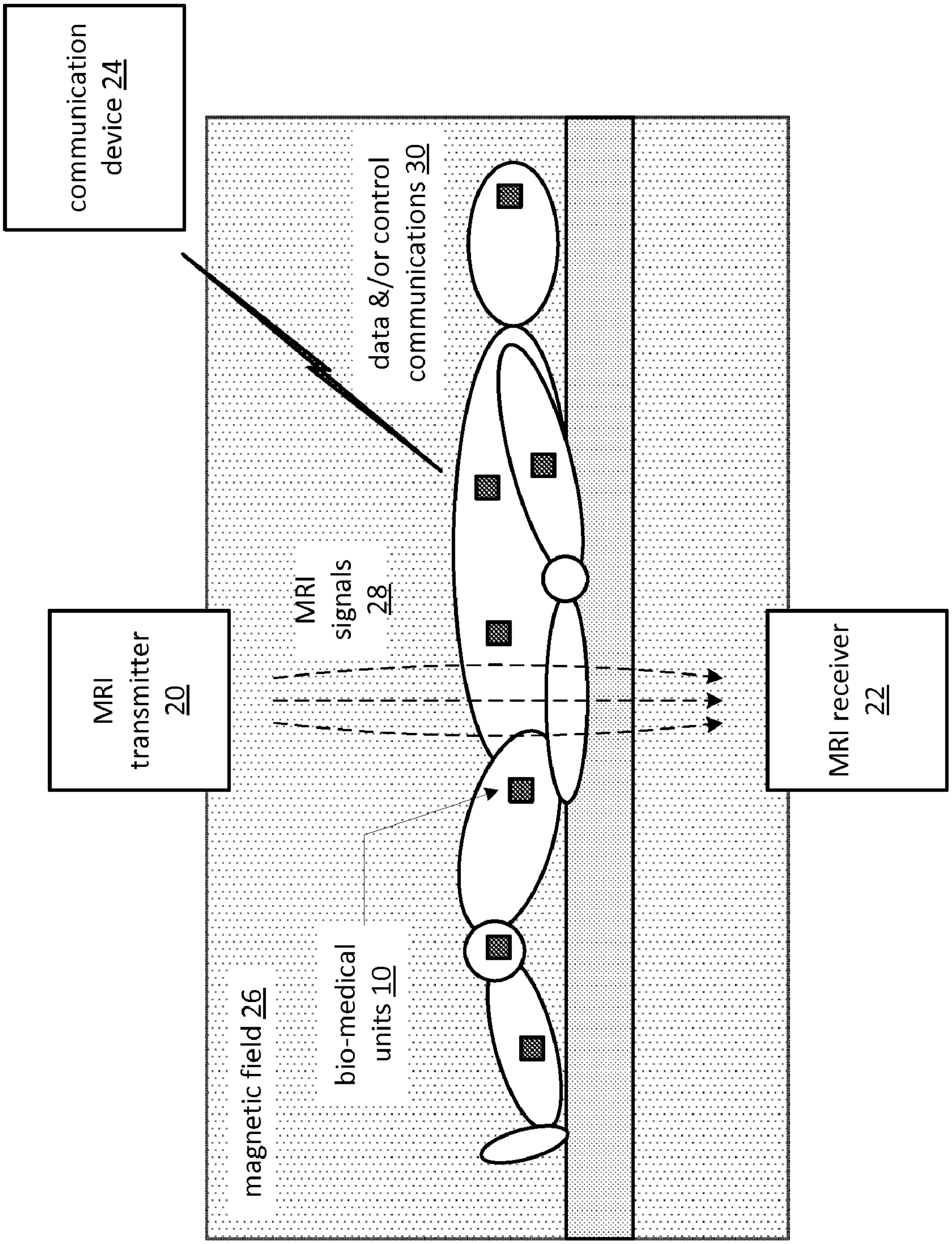
communication device 24
data &/or control communications 30
MRI signals 28
MRI transmitter 20
MRI receiver 22
magnetic field 26
bio-medical units 10

FIG. 2

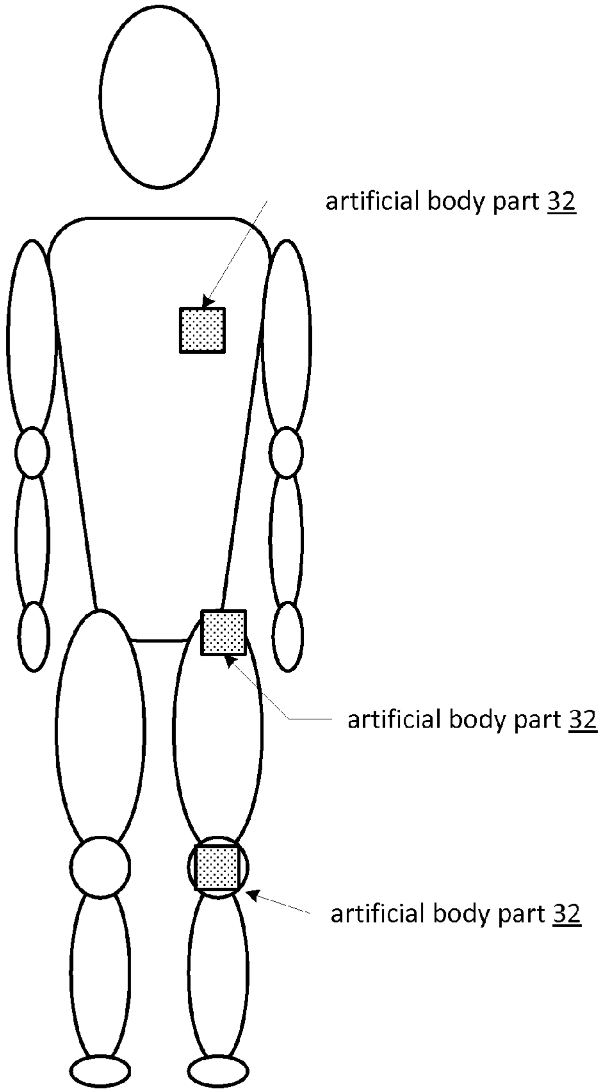
artificial body part 32
artificial body part 32
artificial body part 32

FIG. 3

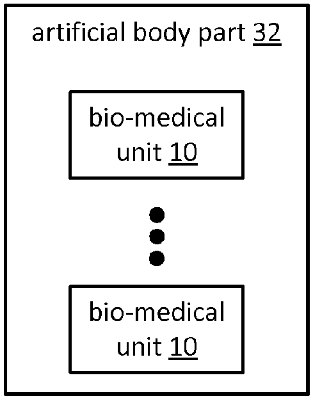
artificial body part 32
bio-medical unit 10
bio-medical unit 10

FIG. 4

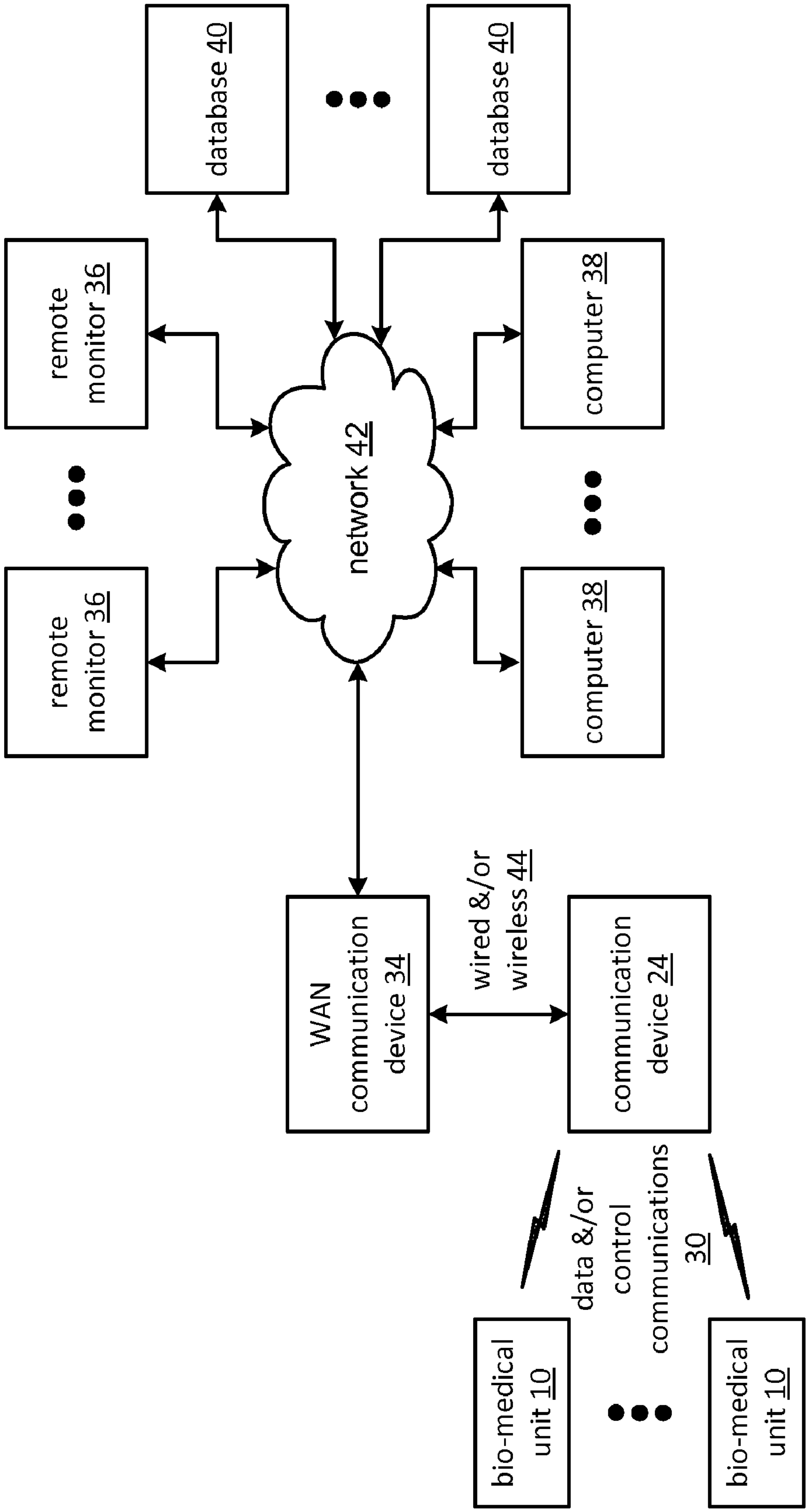
bio-medical unit 10
bio-medical unit 10
data &/or control communications 30
communication device 24
wired &/or wireless 44
WAN communication device 34
network 42
remote monitor 36
remote monitor 36
computer 38
computer 38
database 40
database 40

FIG. 5

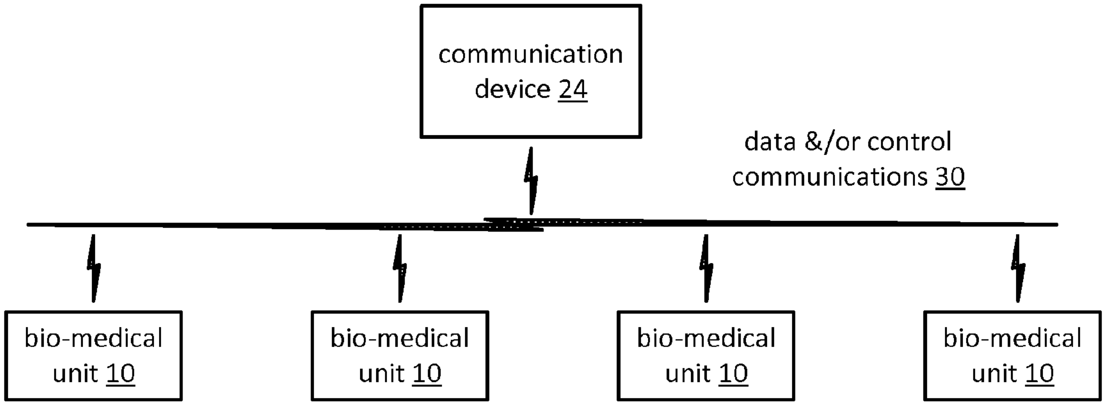
communication device 24
data &/or control communications 30
bio-medical unit 10
bio-medical unit 10
bio-medical unit 10
bio-medical unit 10

FIG. 6

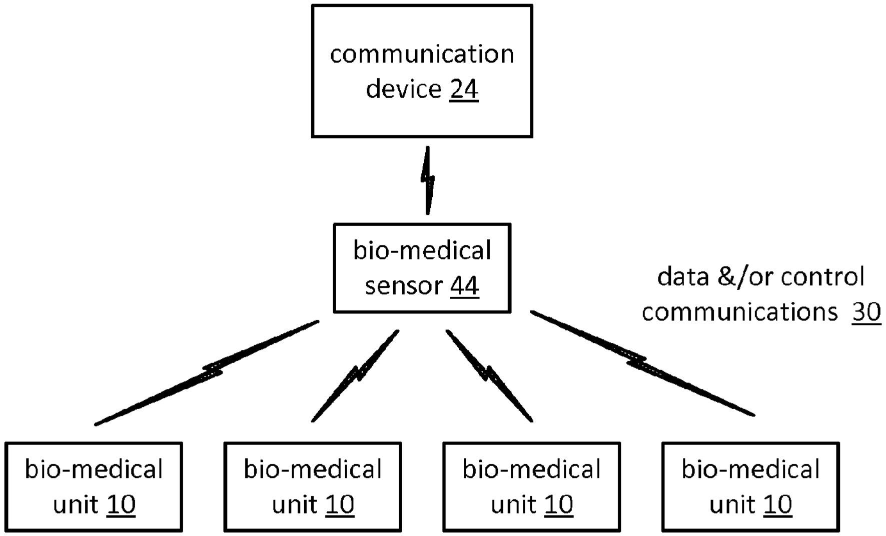
communication device 24
bio-medical sensor 44
data &/or control communications 30
bio-medical unit 10
bio-medical unit 10
bio-medical unit 10
bio-medical unit 10

FIG. 7

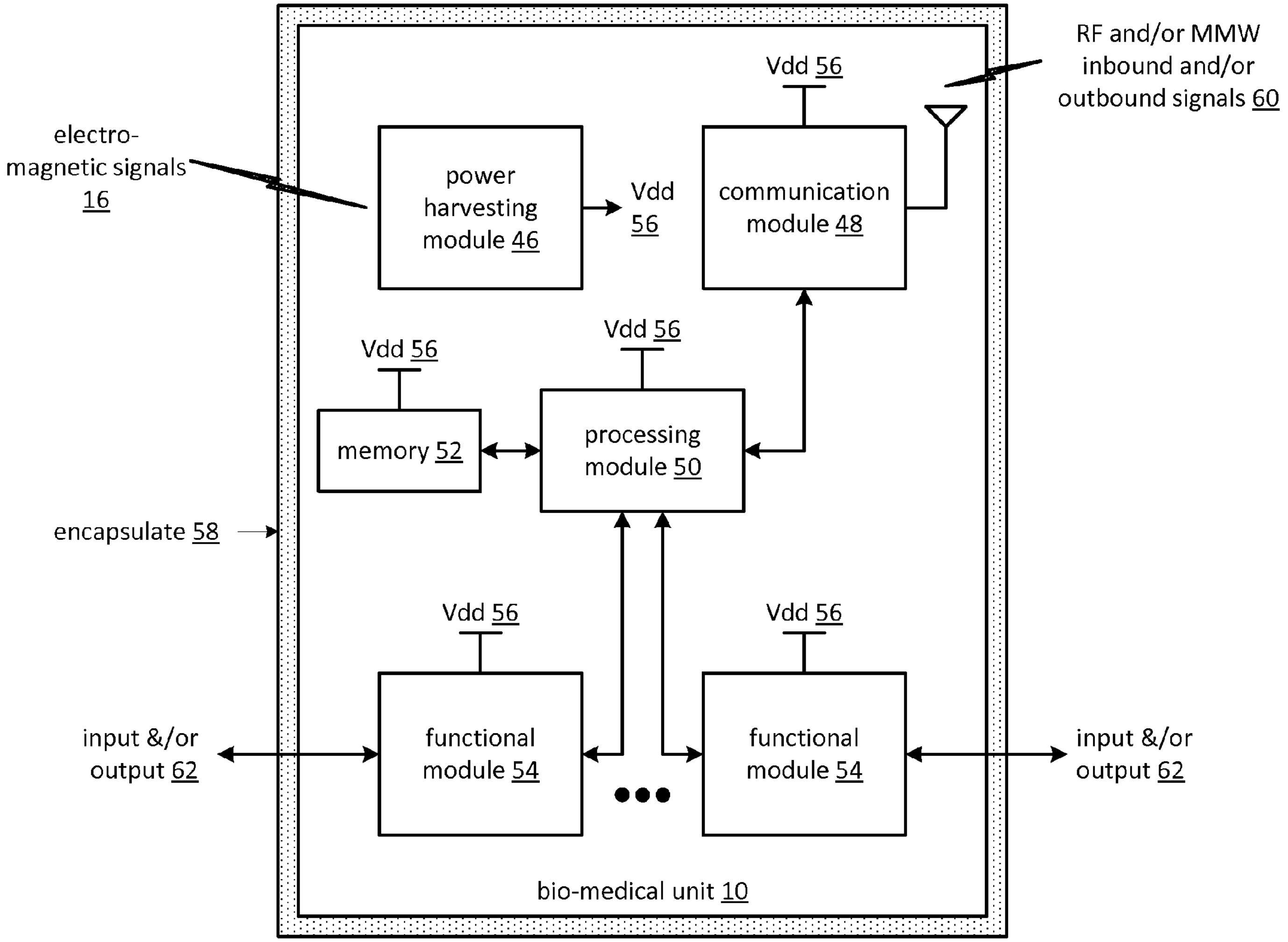
electro-
magnetic signals
16
power
harvesting
module 46
Vdd
56
Vdd 56
communication
module 48
RF and/or MMW
inbound and/or
outbound signals 60
Vdd 56
memory 52
Vdd 56
processing
module 50
encapsulate 58
Vdd 56
Vdd 56
input &/or
output 62
functional
module 54
functional
module 54
input &/or
output 62
bio-medical unit 10

FIG. 8

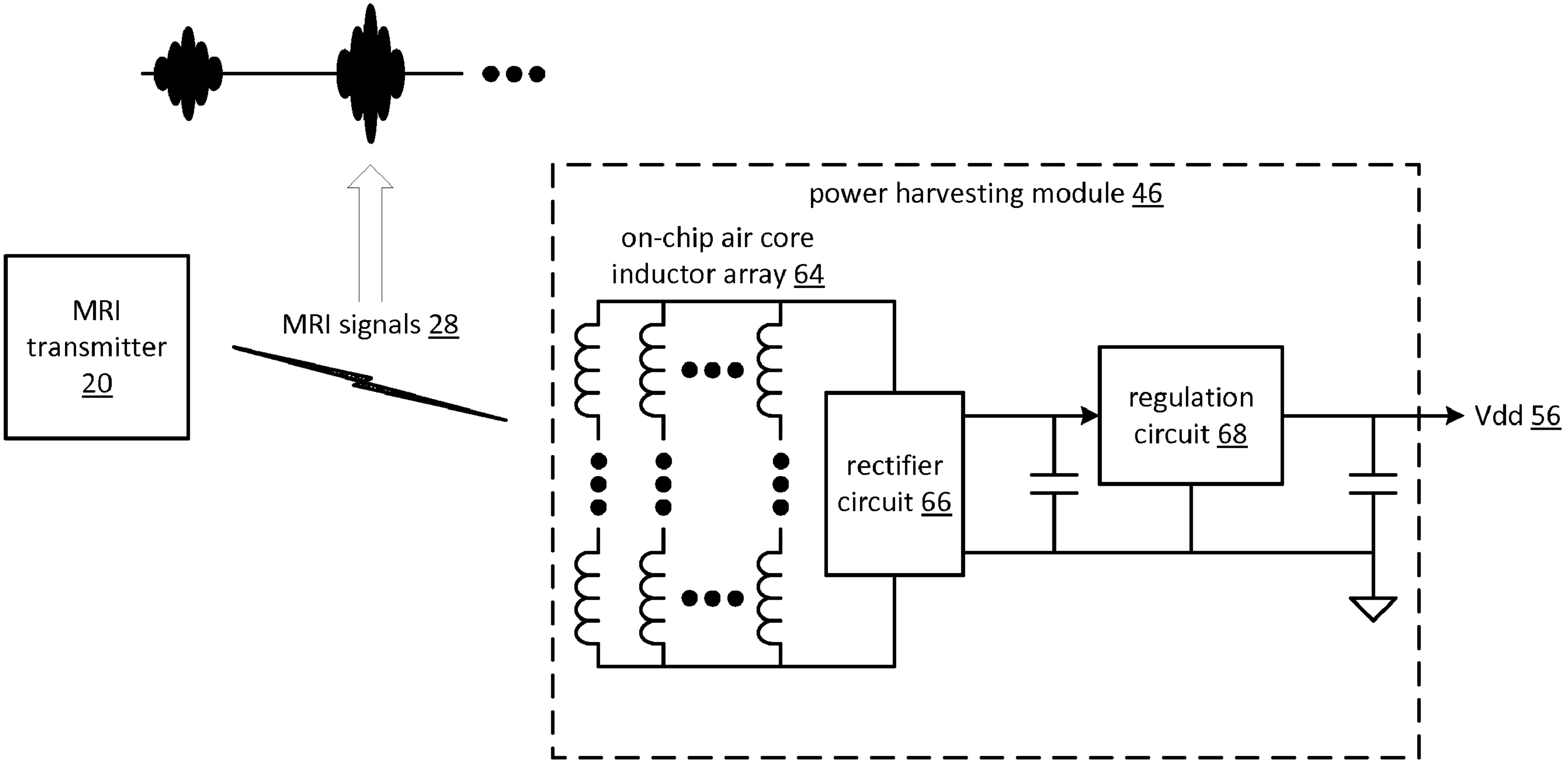
MRI transmitter 20
MRI signals 28
power harvesting module 46
on-chip air core inductor array 64
rectifier circuit 66
regulation circuit 68
Vdd 56

FIG. 9

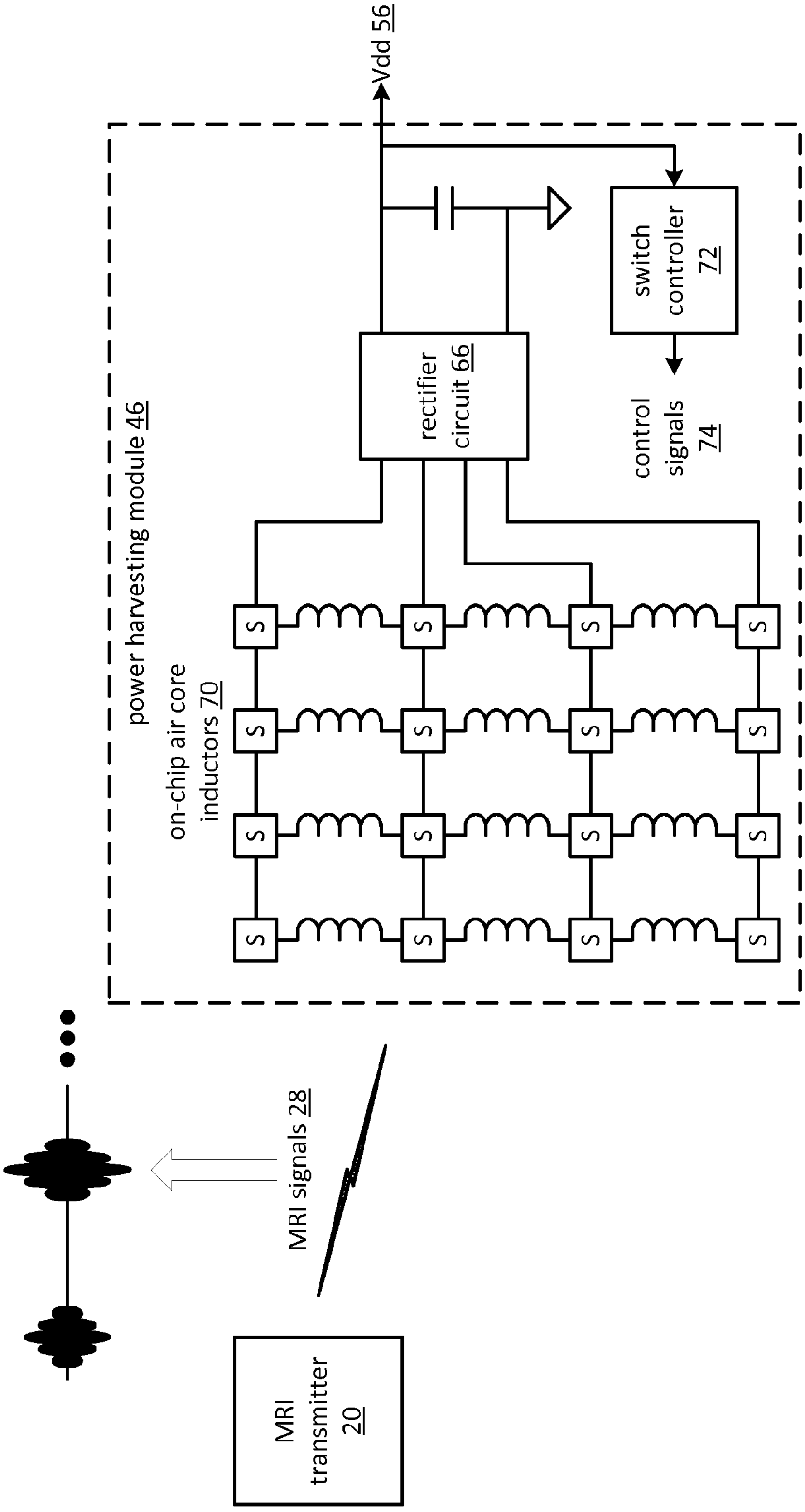
Vdd 56
power harvesting module 46
on-chip air core inductors 70
rectifier circuit 66
switch controller 72
control signals 74
S
MRI signals 28
MRI transmitter 20

FIG. 10

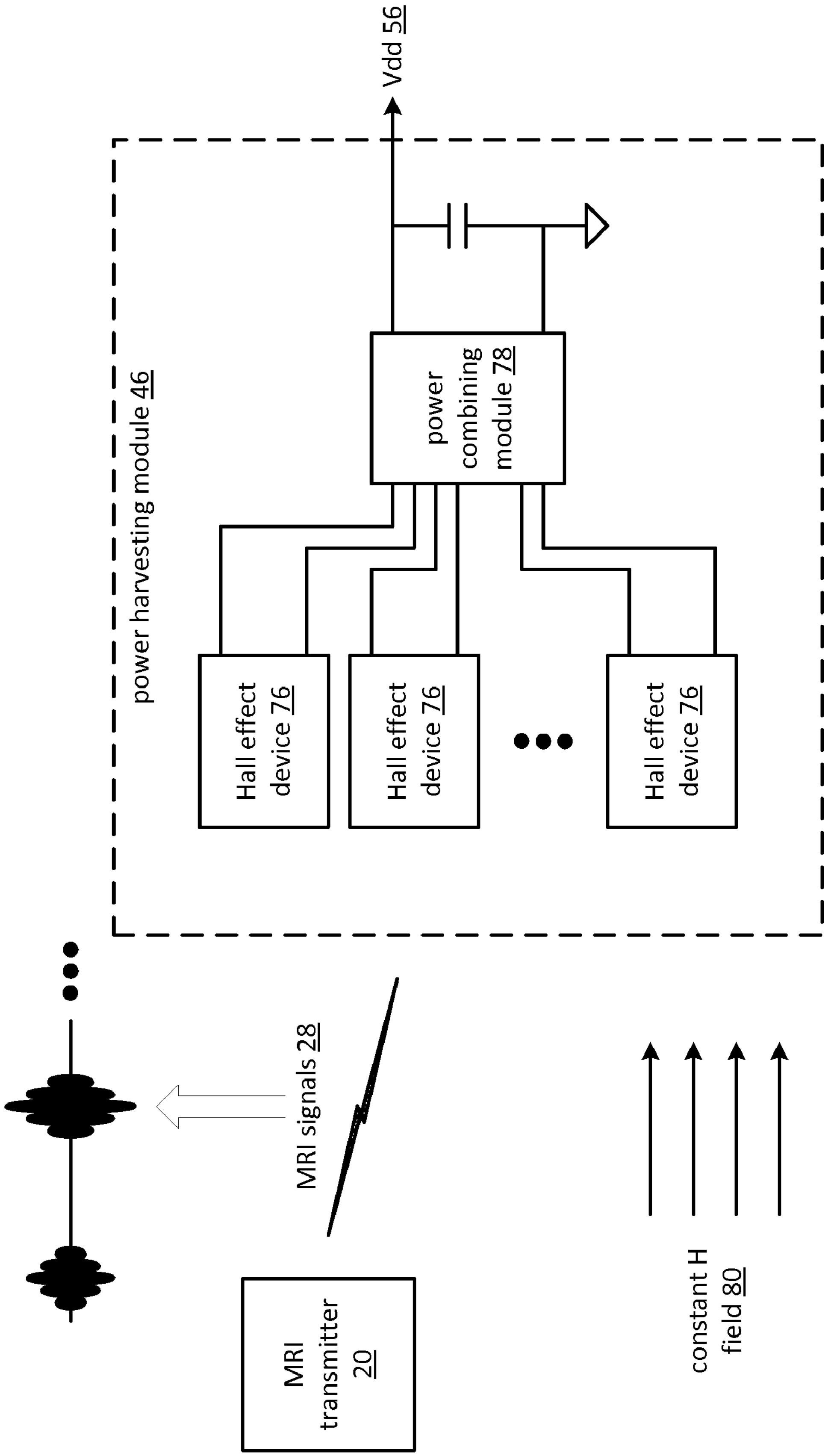
Vdd 56
power harvesting module 46
power combining module 78
Hall effect device 76
Hall effect device 76
Hall effect device 76
MRI signals 28
MRI transmitter 20
constant H field 80

FIG. 11

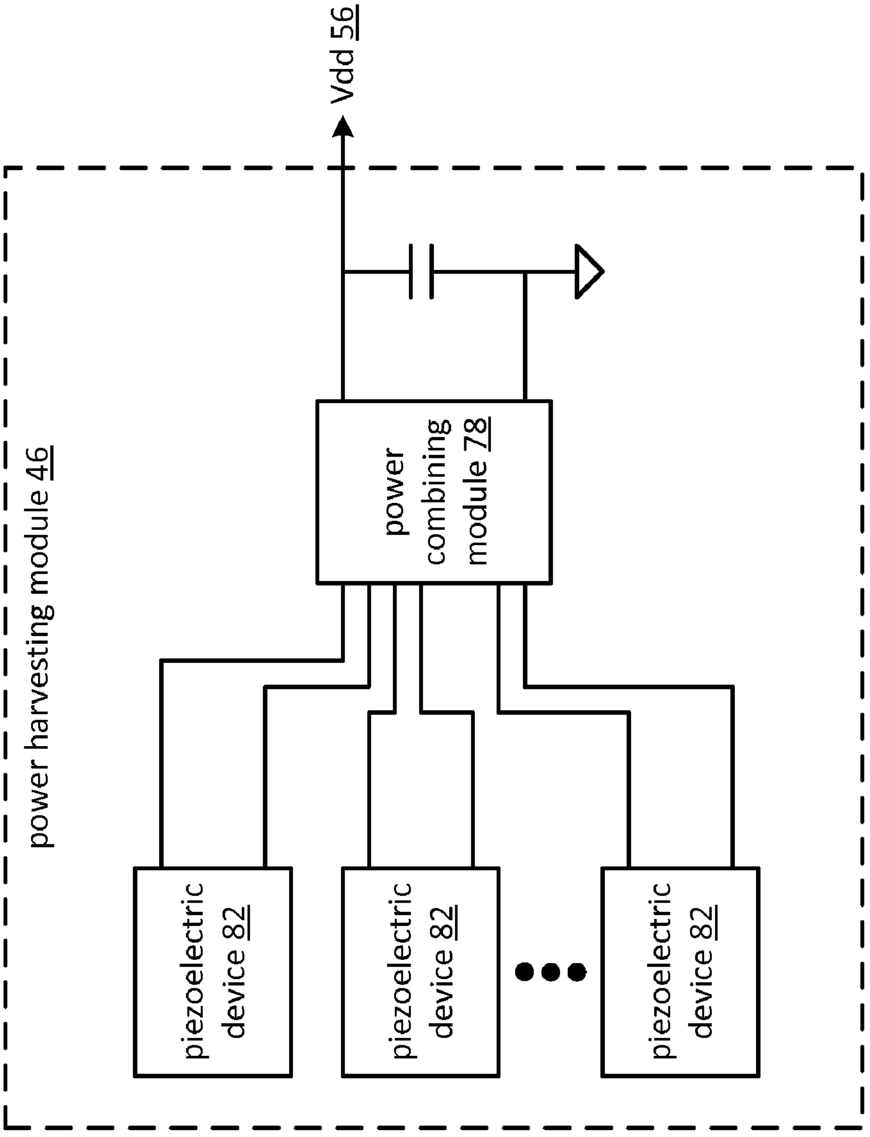
Vdd 56
power harvesting module 46
power combining module 78
piezoelectric device 82
piezoelectric device 82
piezoelectric device 82

FIG. 12

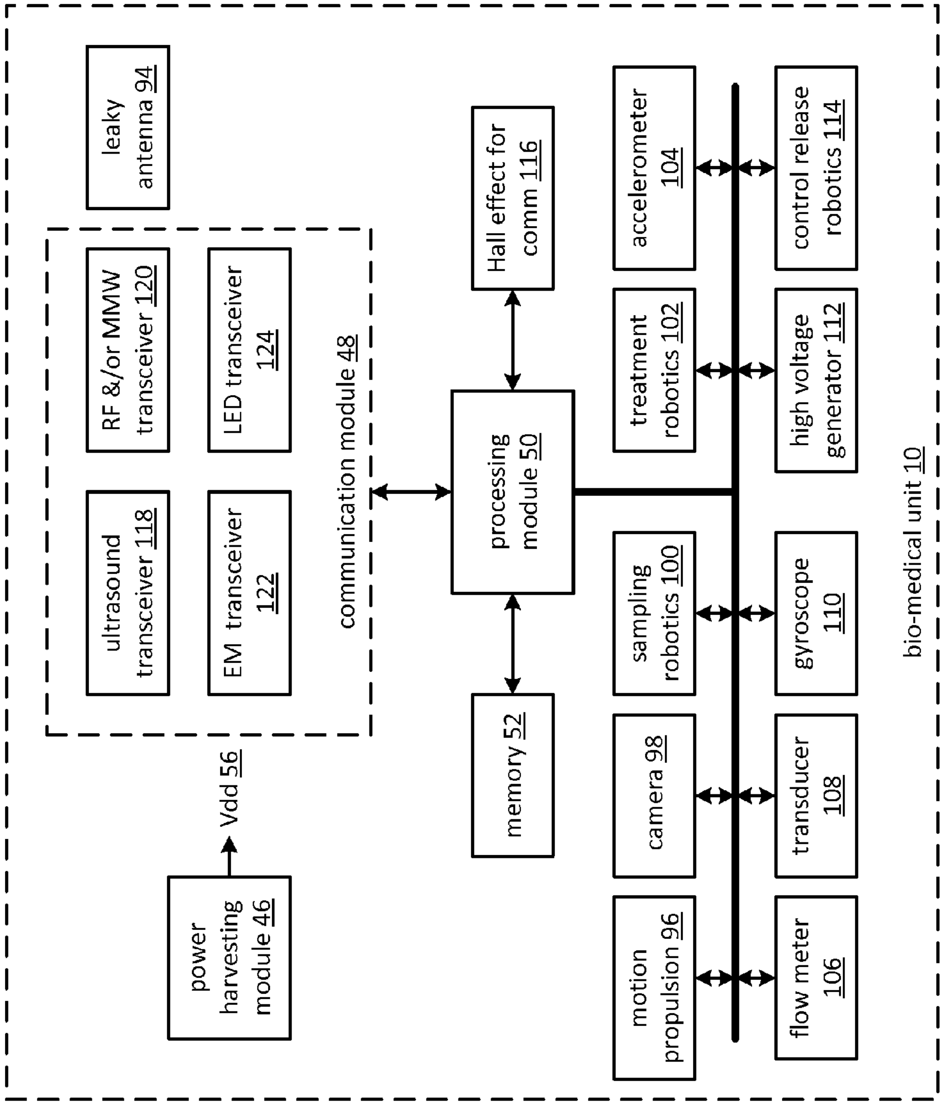
power harvesting module 46
Vdd 56
ultrasound transceiver 118
RF &/or MMW transceiver 120
EM transceiver 122
LED transceiver 124
communication module 48
leaky antenna 94
memory 52
processing module 50
Hall effect for comm 116
motion propulsion 96
camera 98
sampling robotics 100
treatment robotics 102
accelerometer 104
flow meter 106
transducer 108
gyroscope 110
high voltage generator 112
control release robotics 114
bio-medical unit 10

FIG. 13

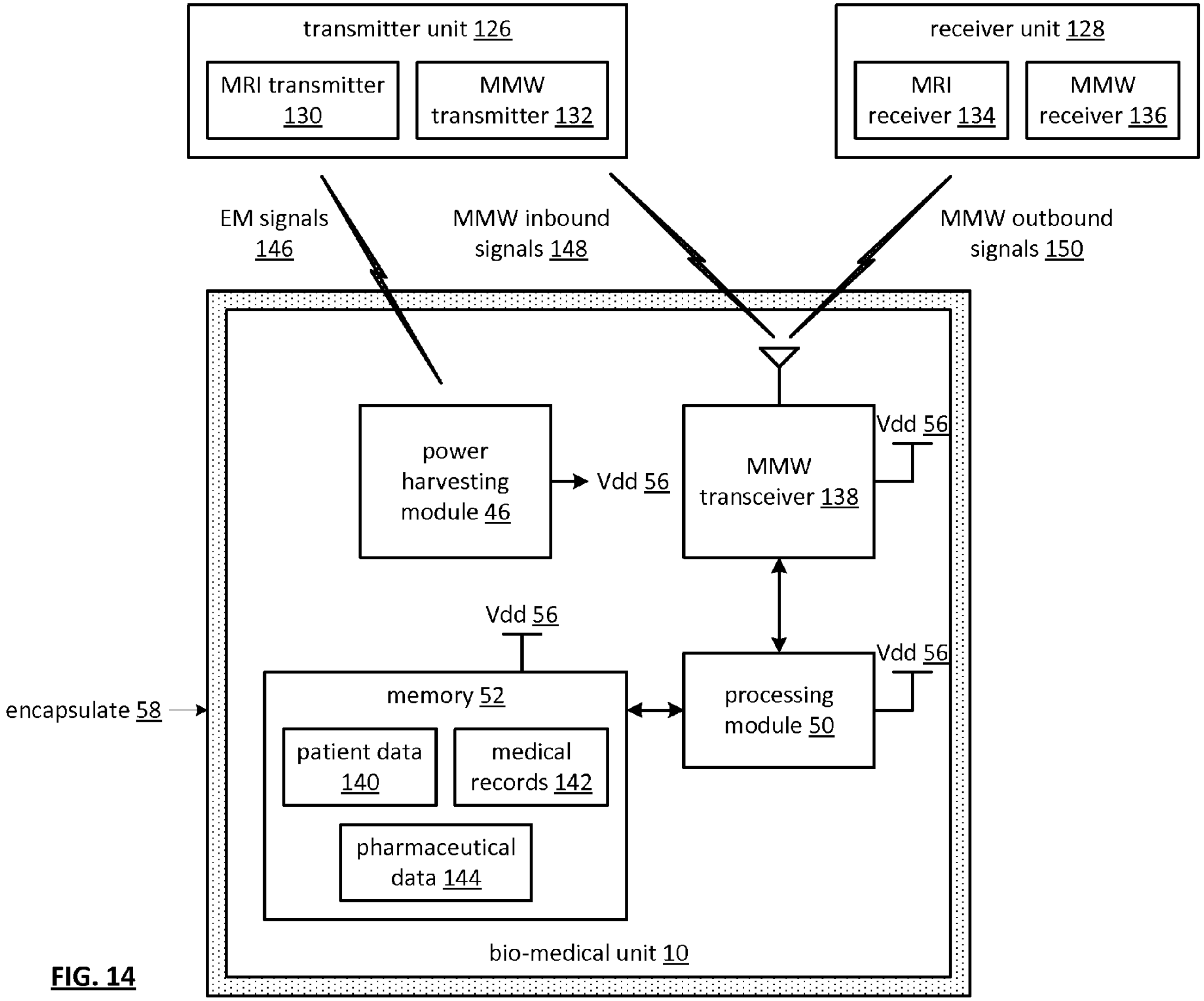
transmitter unit 126
MRI transmitter 130
MMW transmitter 132
receiver unit 128
MRI receiver 134
MMW receiver 136
EM signals 146
MMW inbound signals 148
MMW outbound signals 150
power harvesting module 46
Vdd 56
MMW transceiver 138
Vdd 56
Vdd 56
memory 52
patient data 140
medical records 142
pharmaceutical data 144
processing module 50
Vdd 56
encapsulate 58
bio-medical unit 10

FIG. 14

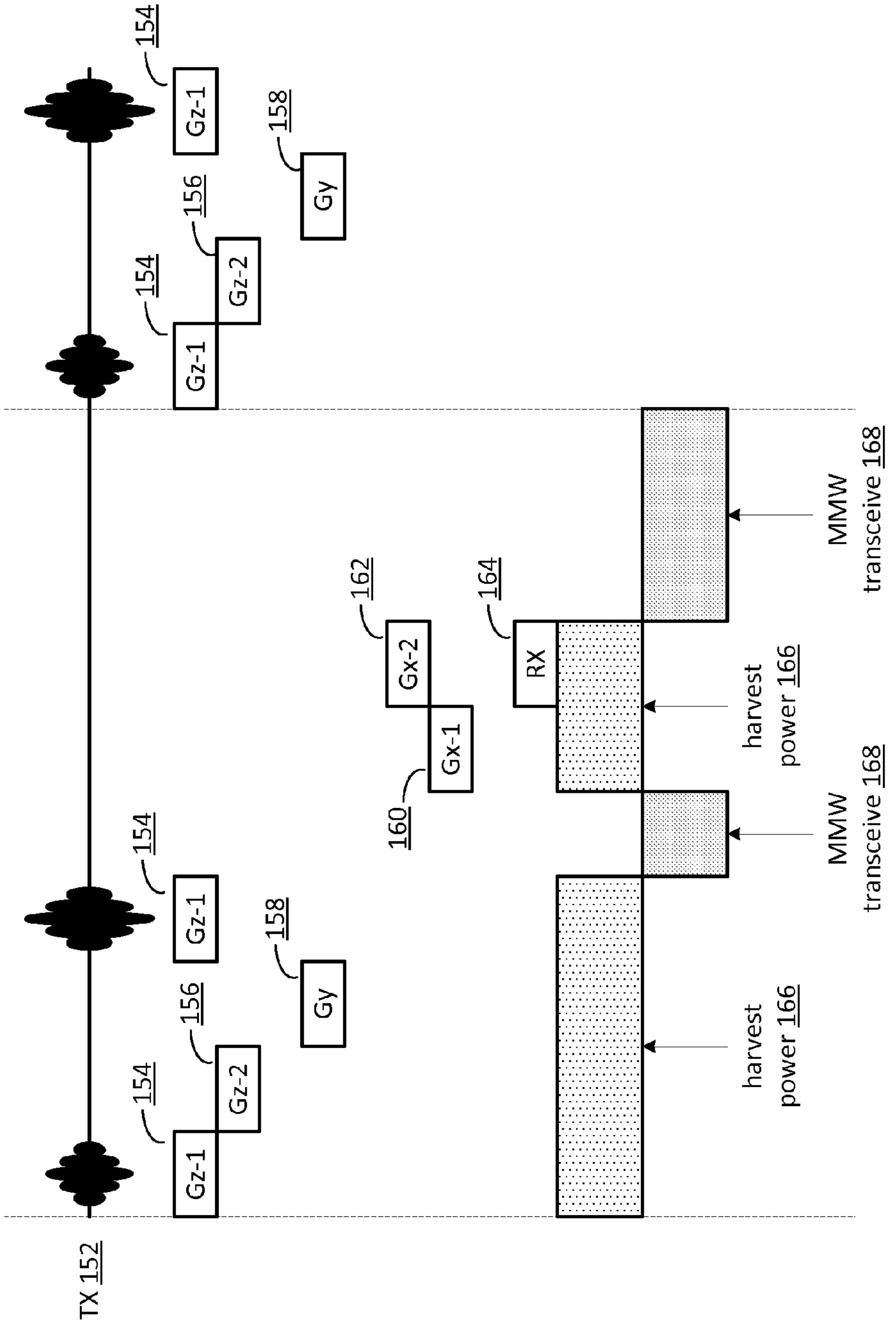
TX 152
Gz-1
154
Gz-2
156
Gz-1
154
Gy
158
Gz-1
154
Gz-2
156
Gz-1
154
Gy
158
Gx-1
160
Gx-2
162
RX
164
harvest power 166
MMW transceive 168
harvest power 166
MMW transceive 168

FIG. 15

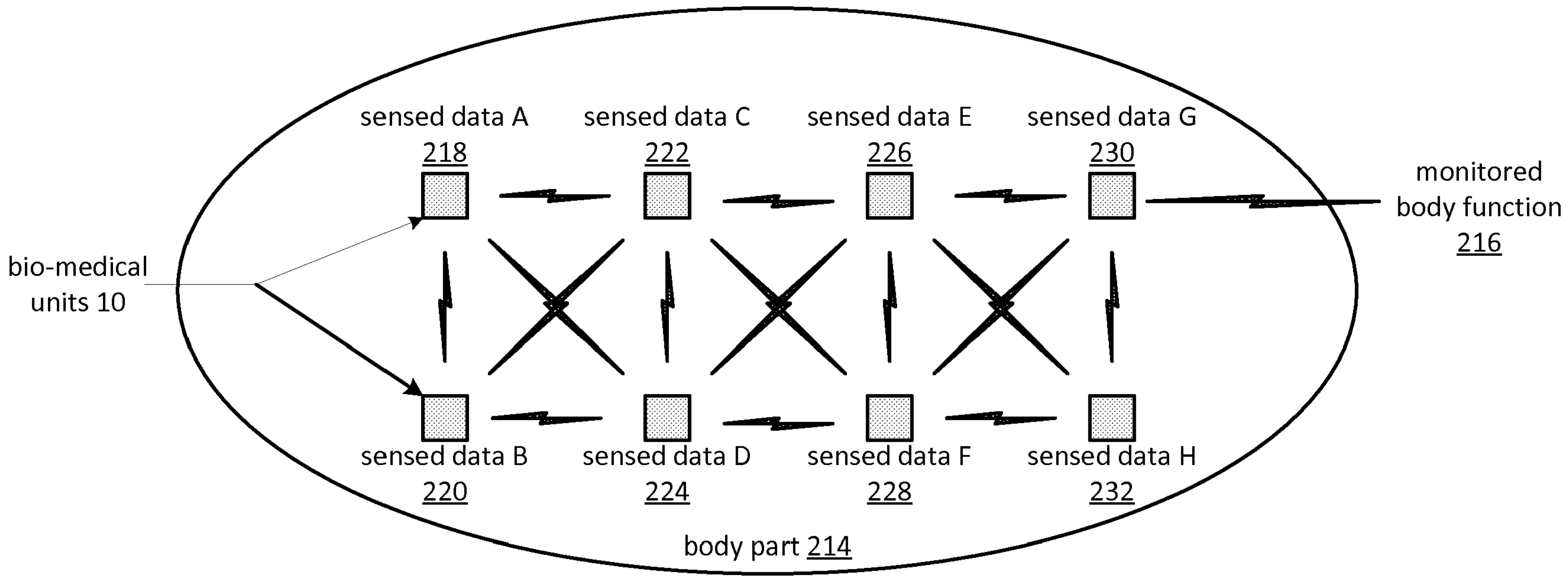
sensed data A
218
sensed data C
222
sensed data E
226
sensed data G
230
monitored
body function
216
bio-medical
units 10
sensed data B
220
sensed data D
224
sensed data F
228
sensed data H
232
body part 214

FIG. 16

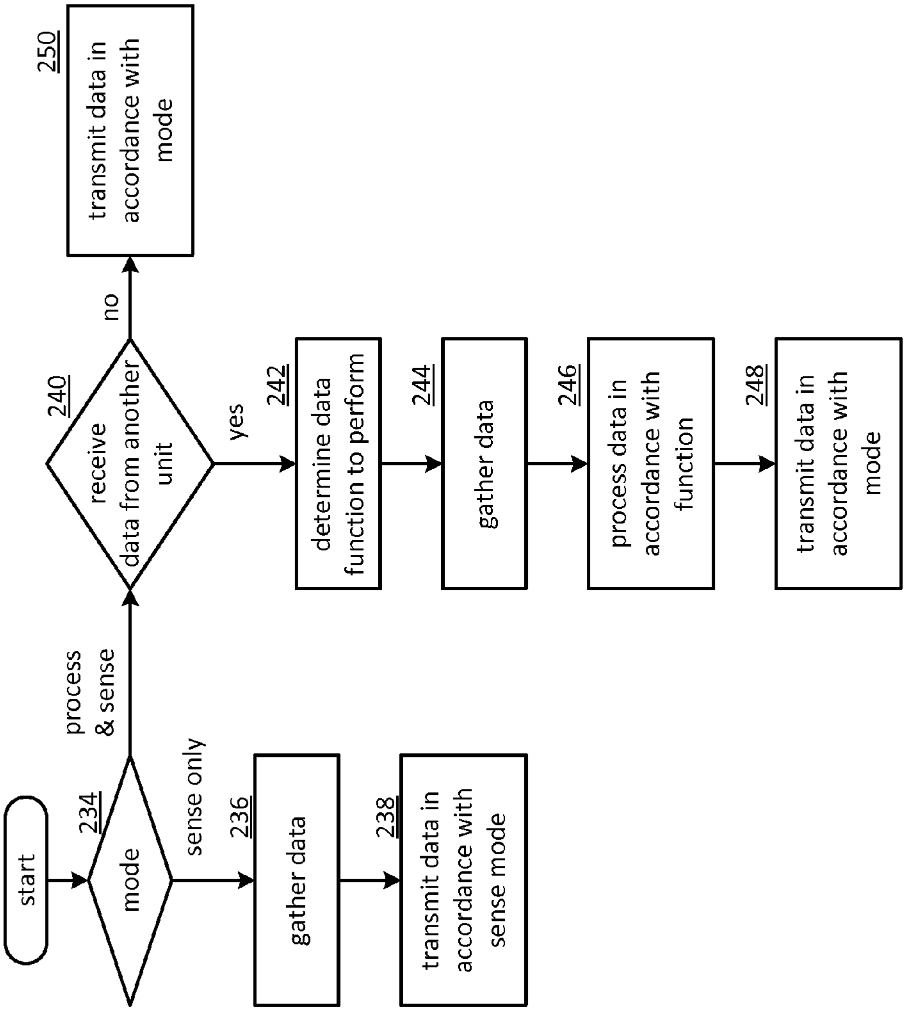
start
mode
234
sense only
process & sense
gather data
236
transmit data in accordance with sense mode
238
receive data from another unit
240
yes
no
determine data function to perform
242
gather data
244
process data in accordance with function
246
transmit data in accordance with mode
248
transmit data in accordance with mode
250

FIG. 17

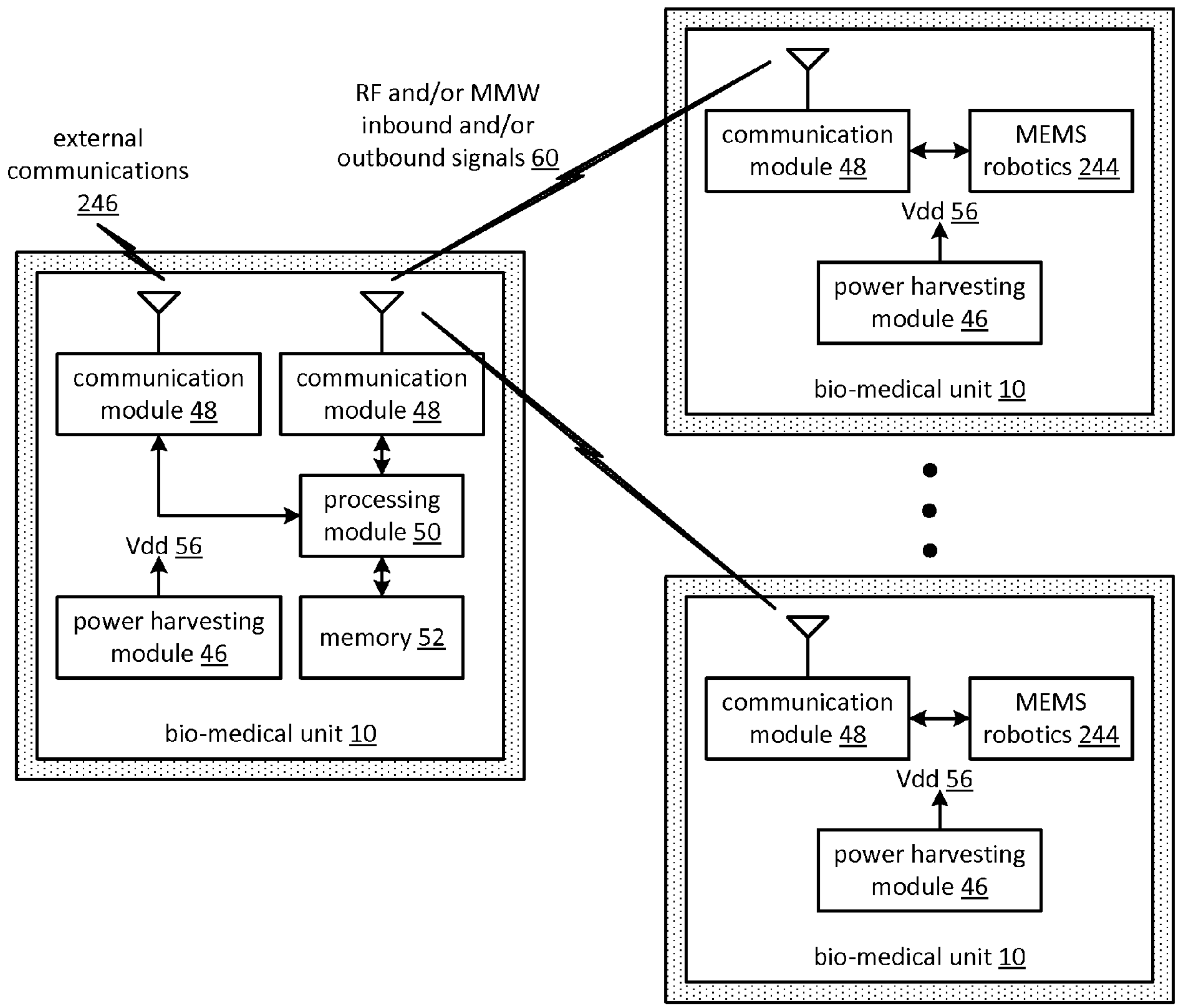
external
communications
246
RF and/or MMW
inbound and/or
outbound signals 60
communication
module 48
communication
module 48
processing
module 50
Vdd 56
power harvesting
module 46
memory 52
bio-medical unit 10
communication
module 48
MEMS
robotics 244
Vdd 56
power harvesting
module 46
bio-medical unit 10
communication
module 48
MEMS
robotics 244
Vdd 56
power harvesting
module 46
bio-medical unit 10

FIG. 18

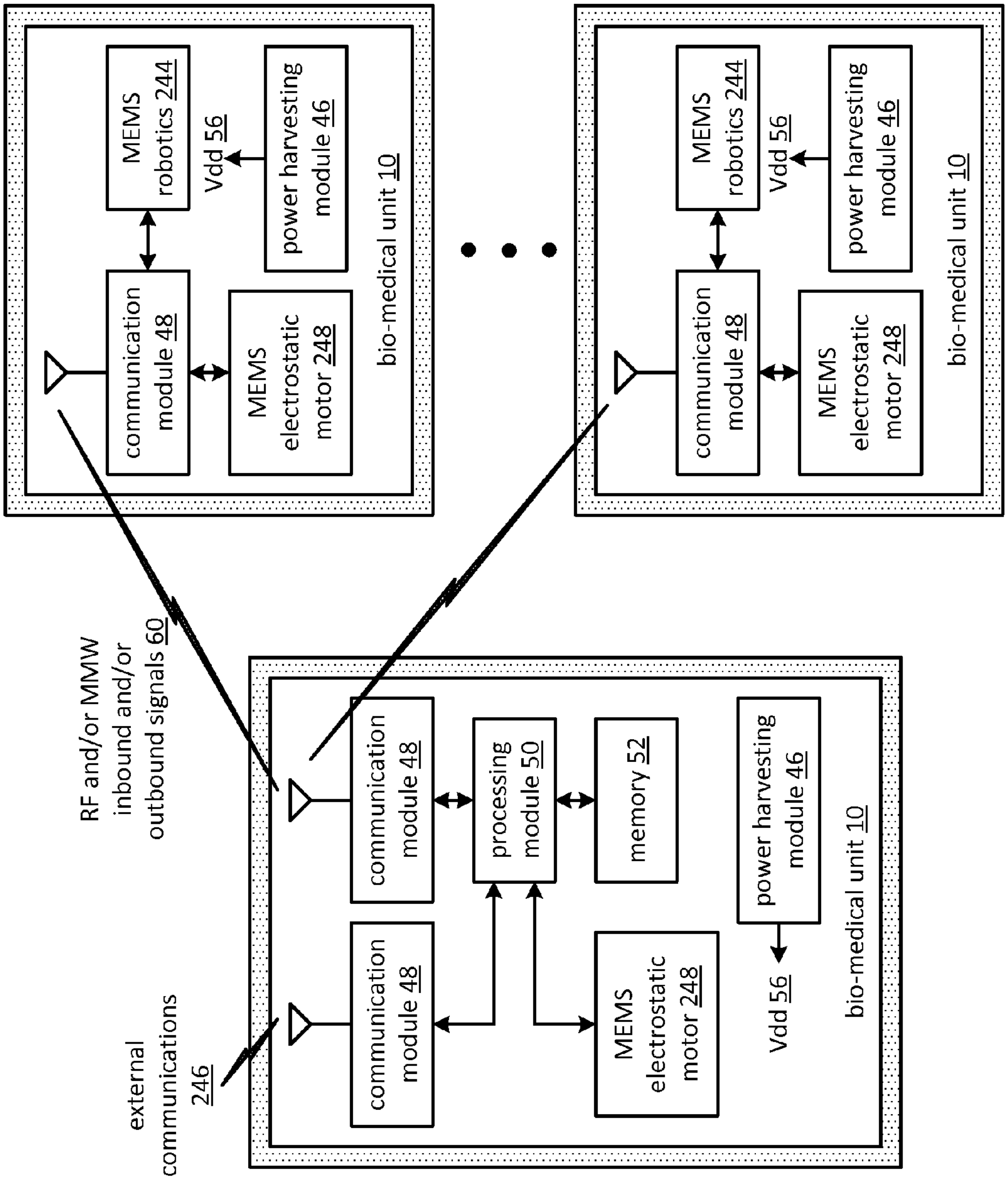
MEMS robotics 244
Vdd 56
power harvesting module 46
communication module 48
MEMS electrostatic motor 248
bio-medical unit 10
MEMS robotics 244
Vdd 56
power harvesting module 46
communication module 48
MEMS electrostatic motor 248
bio-medical unit 10
RF and/or MMW inbound and/or outbound signals 60
external communications 246
communication module 48
communication module 48
processing module 50
memory 52
power harvesting module 46
MEMS electrostatic motor 248
Vdd 56
bio-medical unit 10

FIG. 19

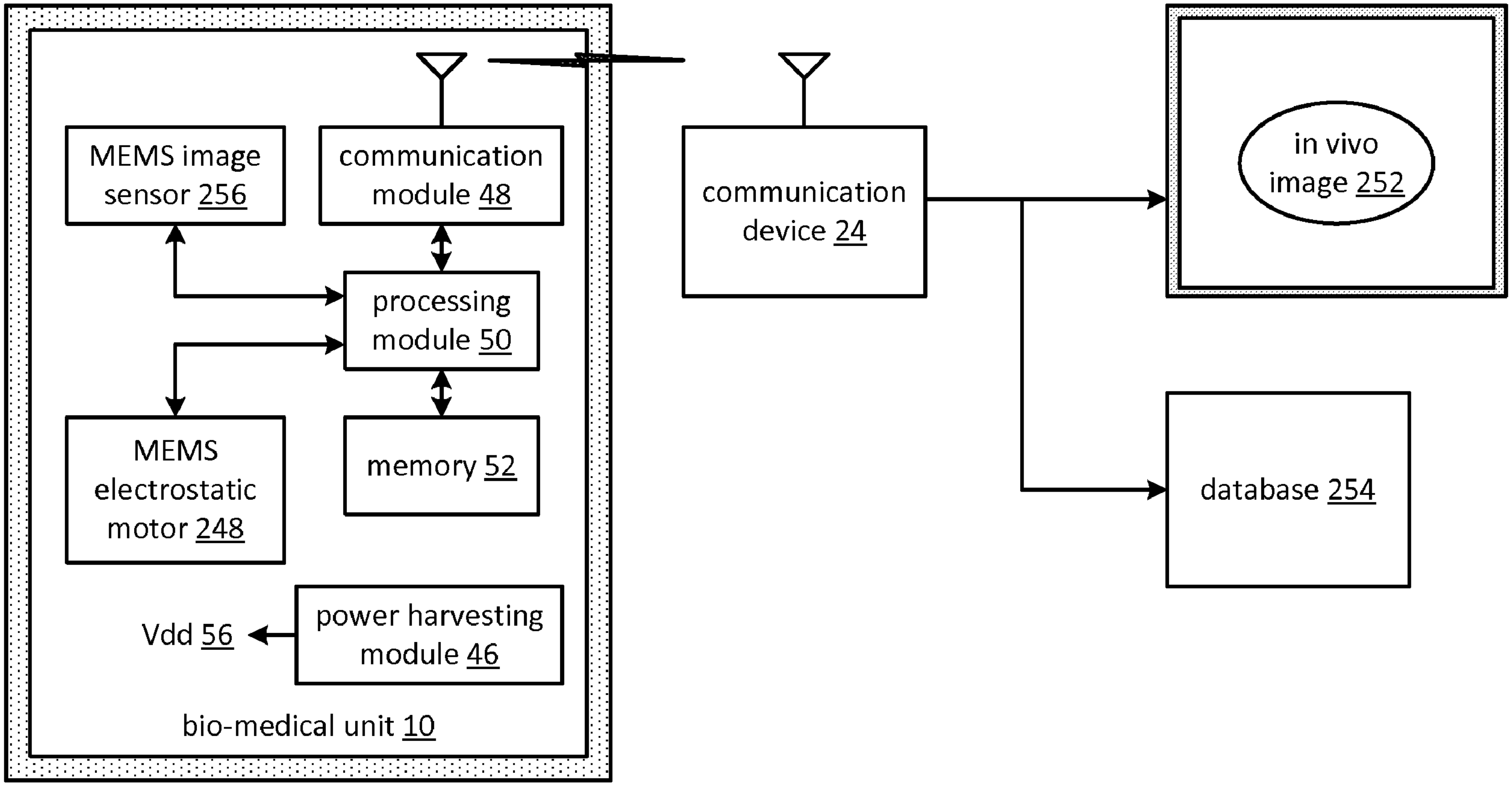
MEMS image sensor 256
communication module 48
processing module 50
MEMS electrostatic motor 248
memory 52
Vdd 56
power harvesting module 46
bio-medical unit 10
communication device 24
in vivo image 252
database 254

FIG. 20

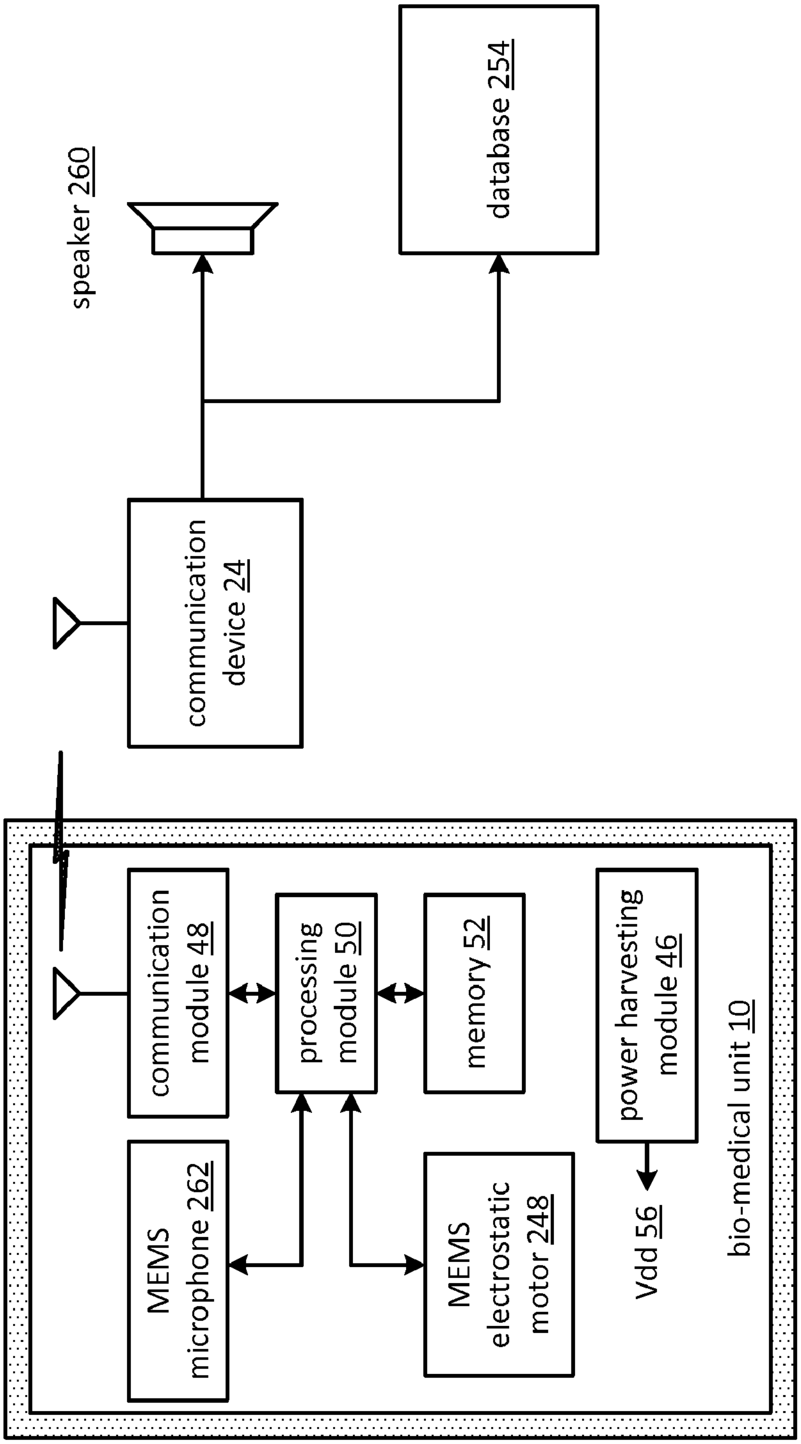
speaker 260
database 254
communication device 24
communication module 48
processing module 50
memory 52
power harvesting module 46
MEMS microphone 262
MEMS electrostatic motor 248
Vdd 56
bio-medical unit 10

FIG. 21

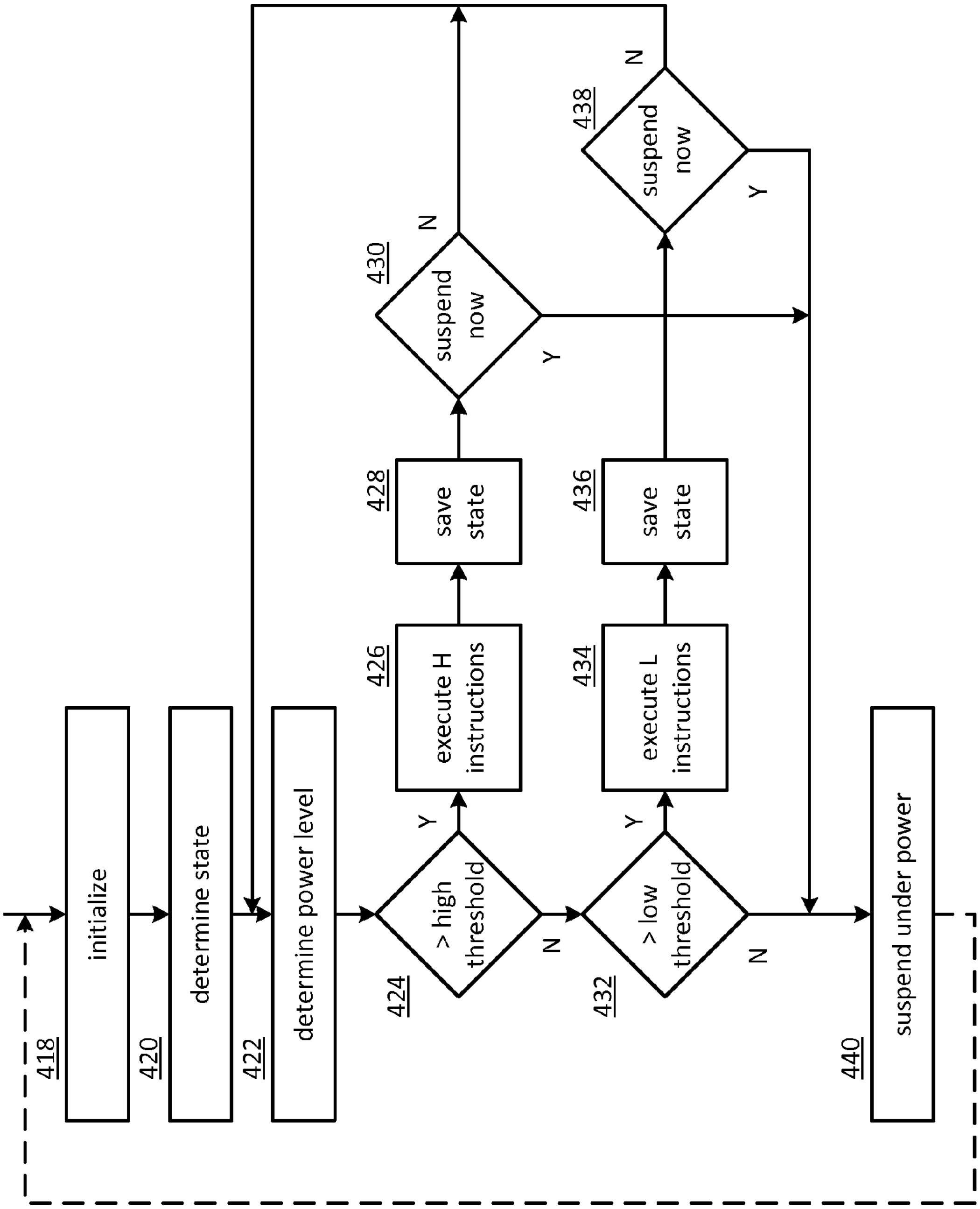
418
initialize
420
determine state
422
determine power level
424
> high threshold
Y
N
426
execute H instructions
428
save state
430
suspend now
Y
N
432
> low threshold
Y
N
434
execute L instructions
436
save state
438
suspend now
Y
N
440
suspend under power

FIG. 22

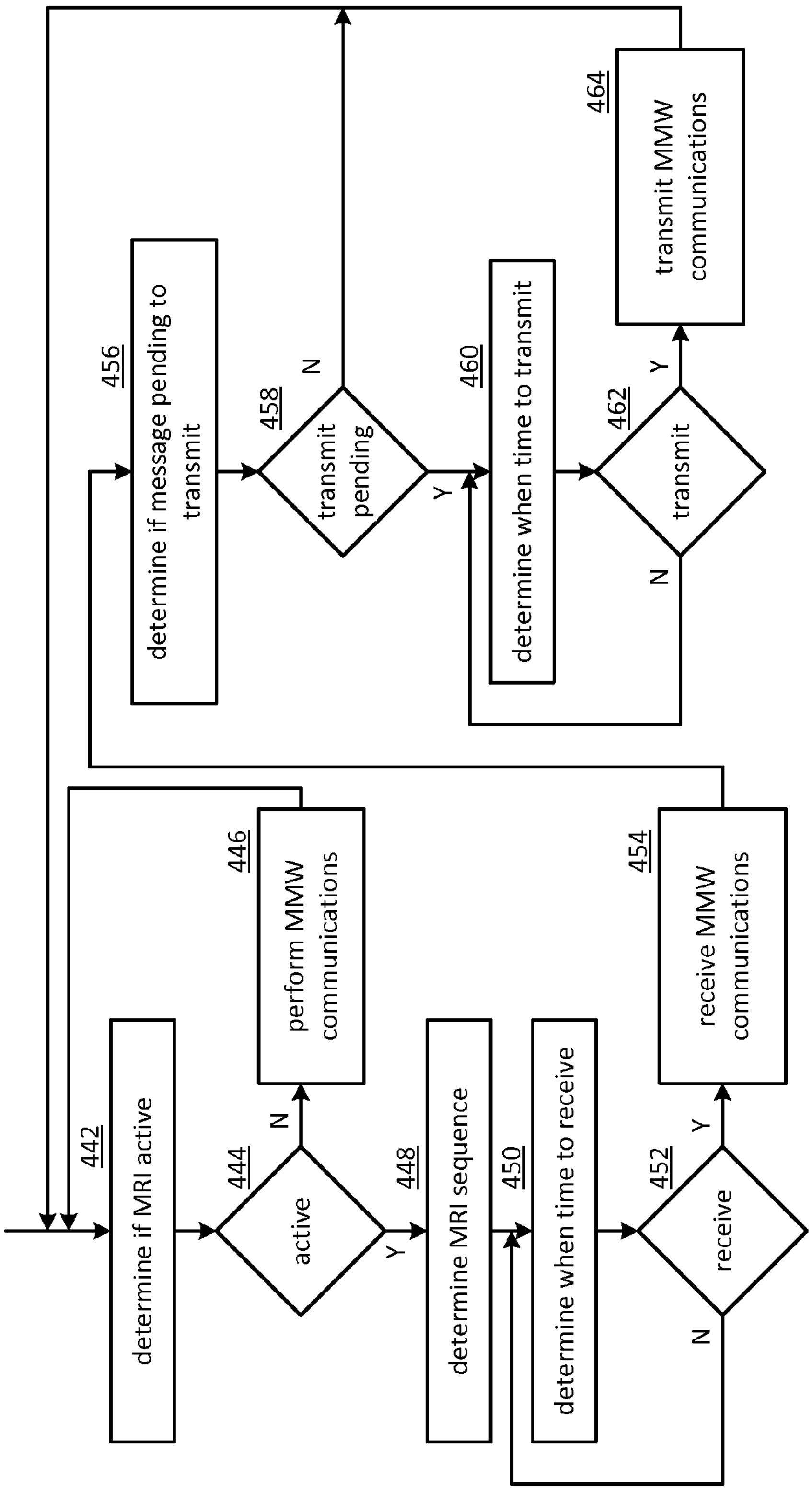
442
determine if MRI active
444
active
N
Y
446
perform MMW communications
448
determine MRI sequence
450
determine when time to receive
452
receive
Y
N
454
receive MMW communications
456
determine if message pending to transmit
458
transmit pending
N
Y
460
determine when time to transmit
462
transmit
Y
N
464
transmit MMW communications

FIG. 23

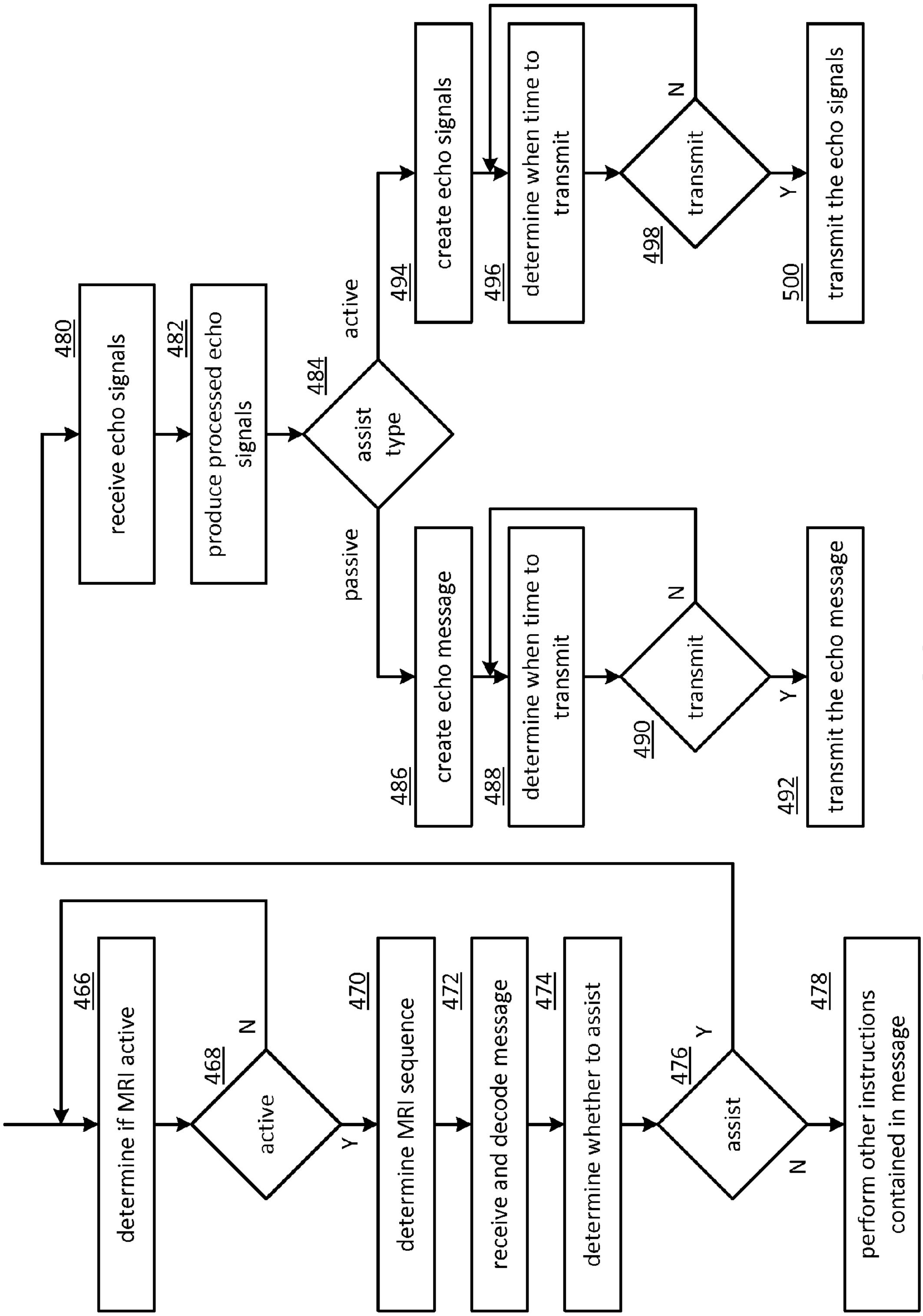
466
determine if MRI active
468
active
N
Y
470
determine MRI sequence
472
receive and decode message
474
determine whether to assist
476
assist
Y
N
478
perform other instructions contained in message
480
receive echo signals
482
produce processed echo signals
484
assist type
passive
active
486
create echo message
488
determine when time to transmit
490
transmit
N
Y
492
transmit the echo message
494
create echo signals
496
determine when time to transmit
498
transmit
N
Y
500
transmit the echo signals

FIG. 24

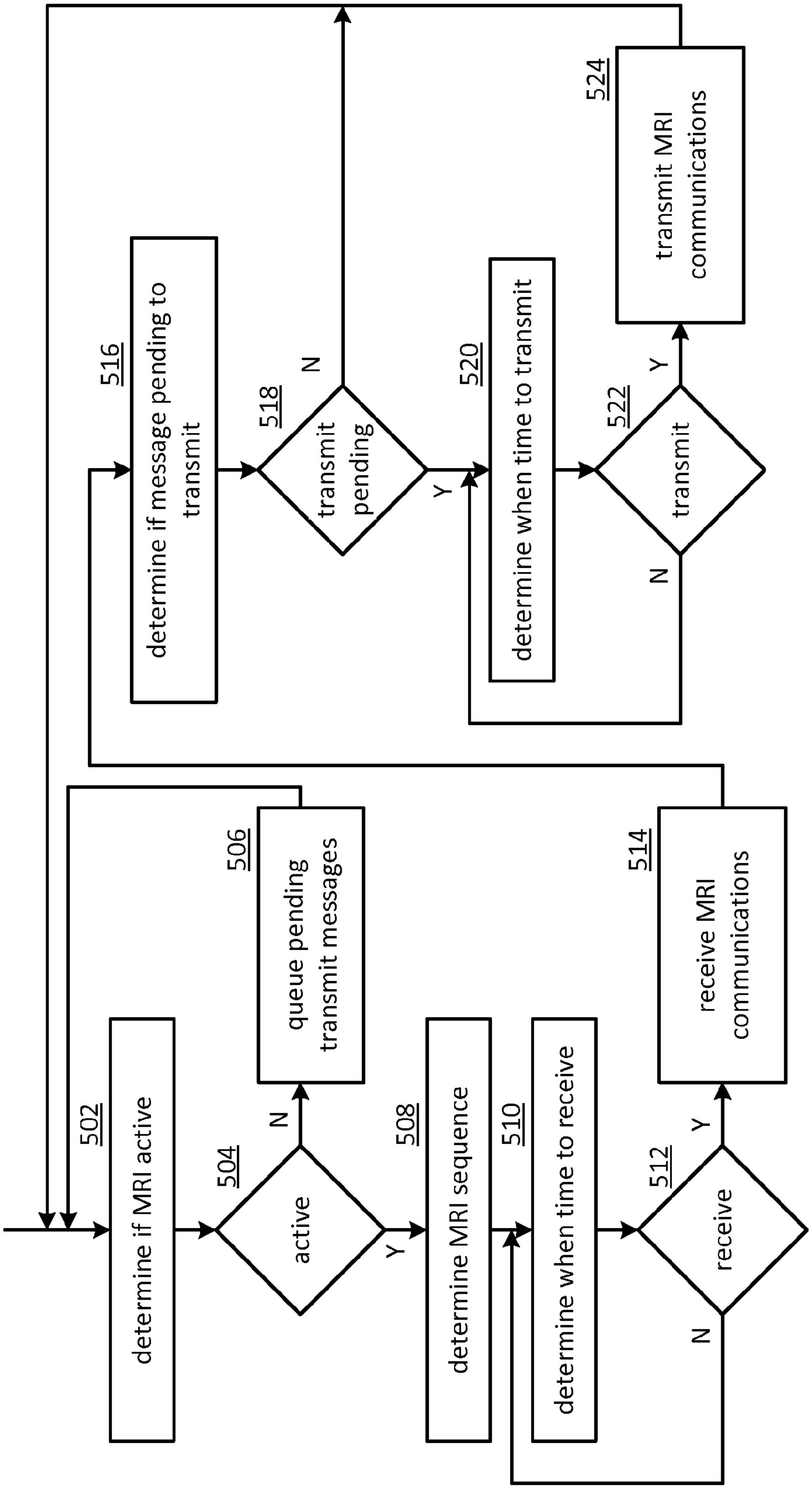
502
determine if MRI active
504
active
Y
N
506
queue pending transmit messages
508
determine MRI sequence
510
determine when time to receive
512
receive
Y
N
514
receive MRI communications
516
determine if message pending to transmit
518
transmit pending
Y
N
520
determine when time to transmit
522
transmit
Y
N
524
transmit MRI communications

FIG. 25

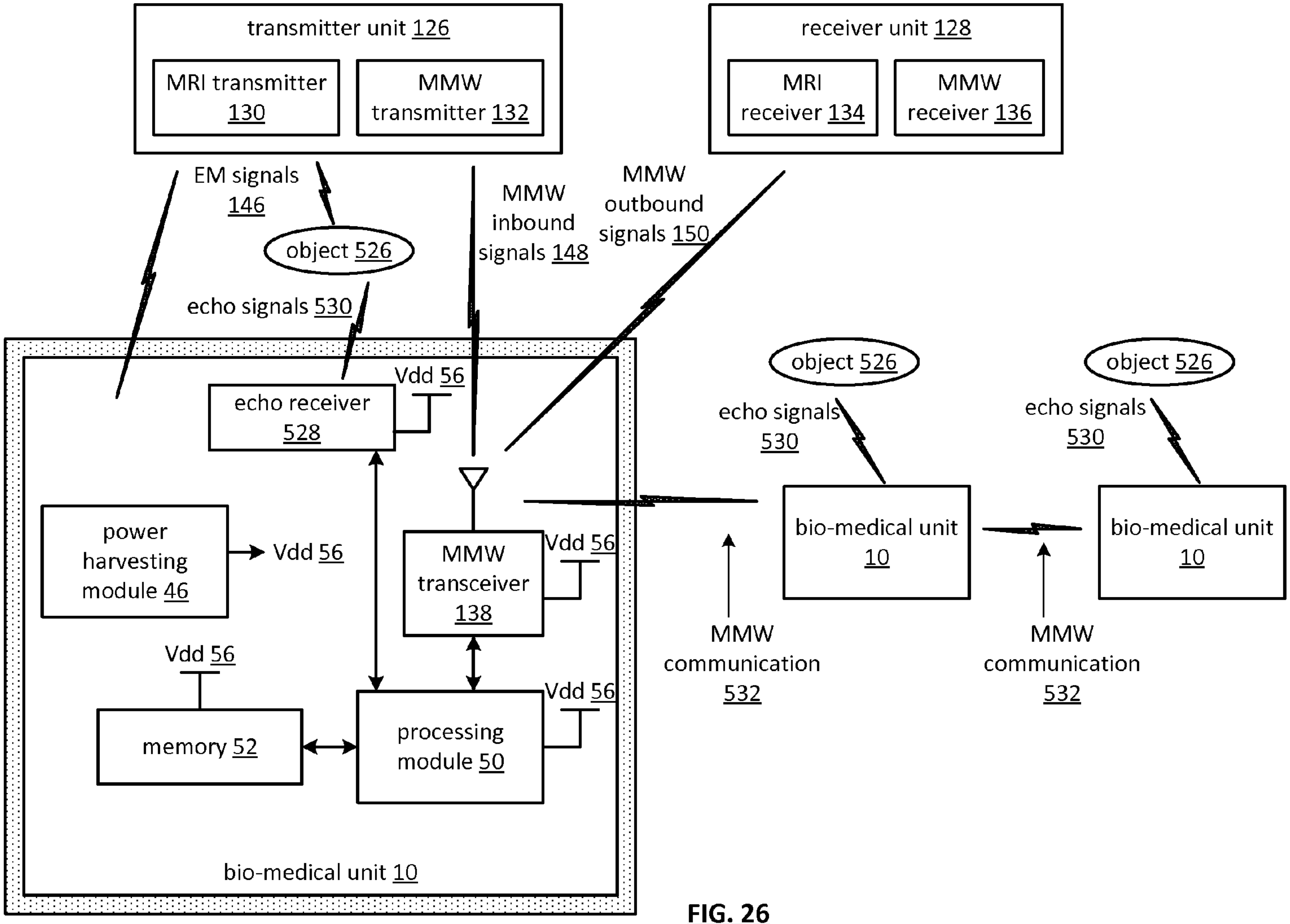
transmitter unit 126
MRI transmitter 130
MMW transmitter 132
receiver unit 128
MRI receiver 134
MMW receiver 136
EM signals 146
object 526
MMW inbound signals 148
MMW outbound signals 150
echo signals 530
echo receiver 528
Vdd 56
power harvesting module 46
Vdd 56
MMW transceiver 138
Vdd 56
Vdd 56
memory 52
processing module 50
Vdd 56
bio-medical unit 10
object 526
echo signals 530
bio-medical unit 10
MMW communication 532
object 526
echo signals 530
bio-medical unit 10
MMW communication 532

FIG. 26

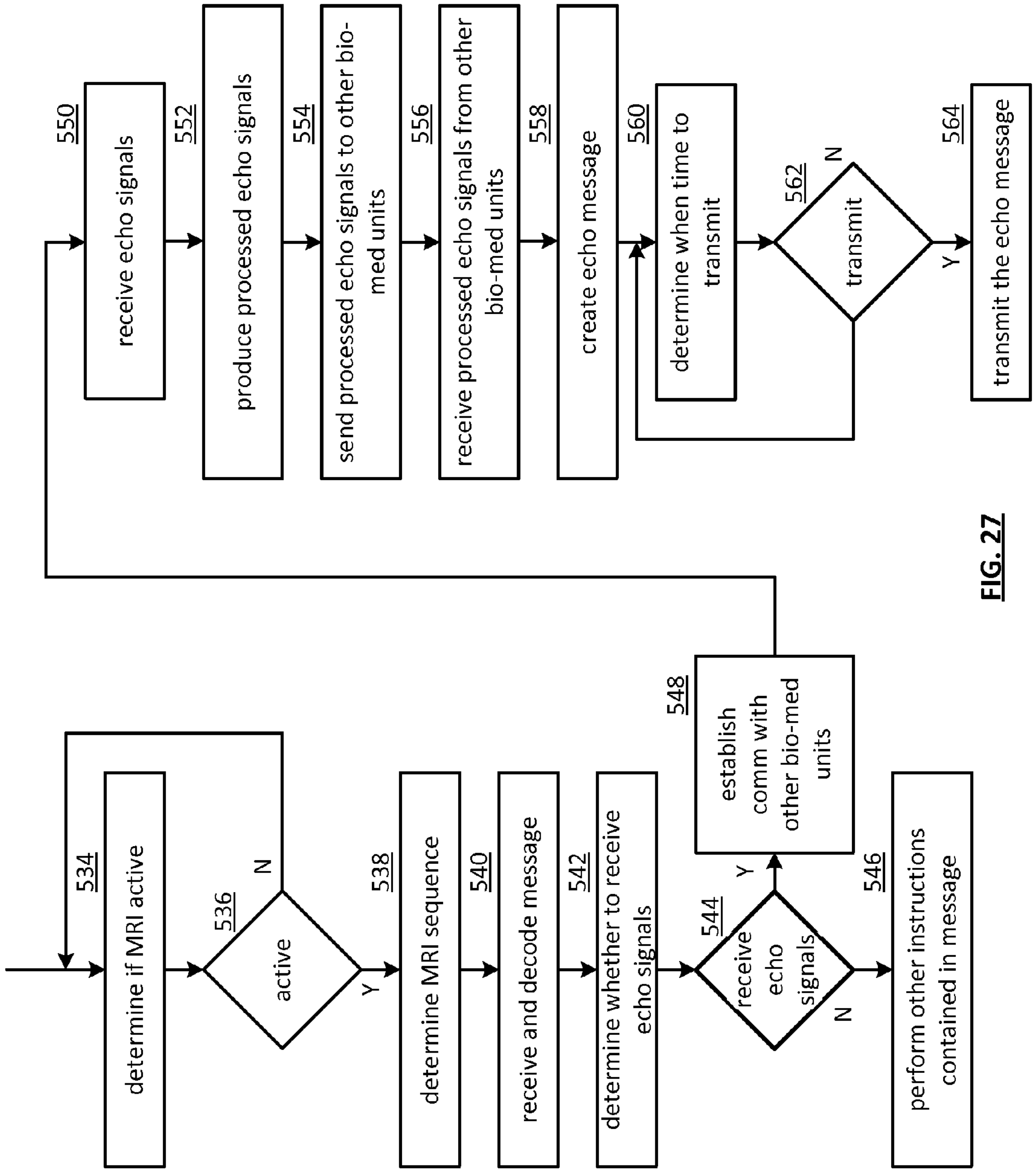
534
determine if MRI active
536
active
N
Y
538
determine MRI sequence
540
receive and decode message
542
determine whether to receive echo signals
544
receive echo signals
Y
N
548
establish comm with other bio-med units
546
perform other instructions contained in message
550
receive echo signals
552
produce processed echo signals
554
send processed echo signals to other bio-med units
556
receive processed echo signals from other bio-med units
558
create echo message
560
determine when time to transmit
562
transmit
N
Y
564
transmit the echo message

FIG. 27

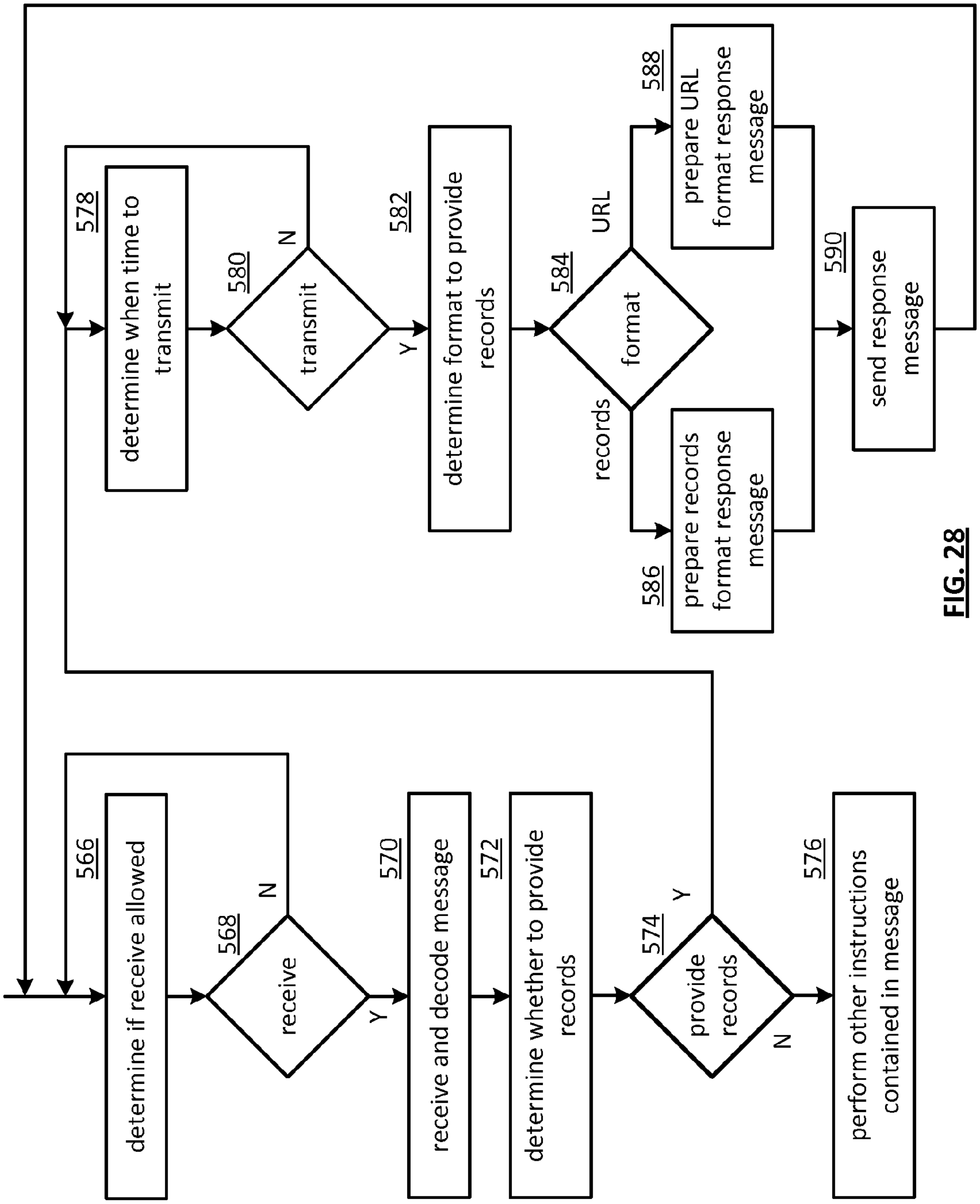
566
determine if receive allowed
568
receive
N
Y
570
receive and decode message
572
determine whether to provide records
574
provide records
Y
N
576
perform other instructions contained in message
578
determine when time to transmit
580
transmit
N
Y
582
determine format to provide records
584
format
records
URL
586
prepare records format response message
588
prepare URL format response message
590
send response message

FIG. 28

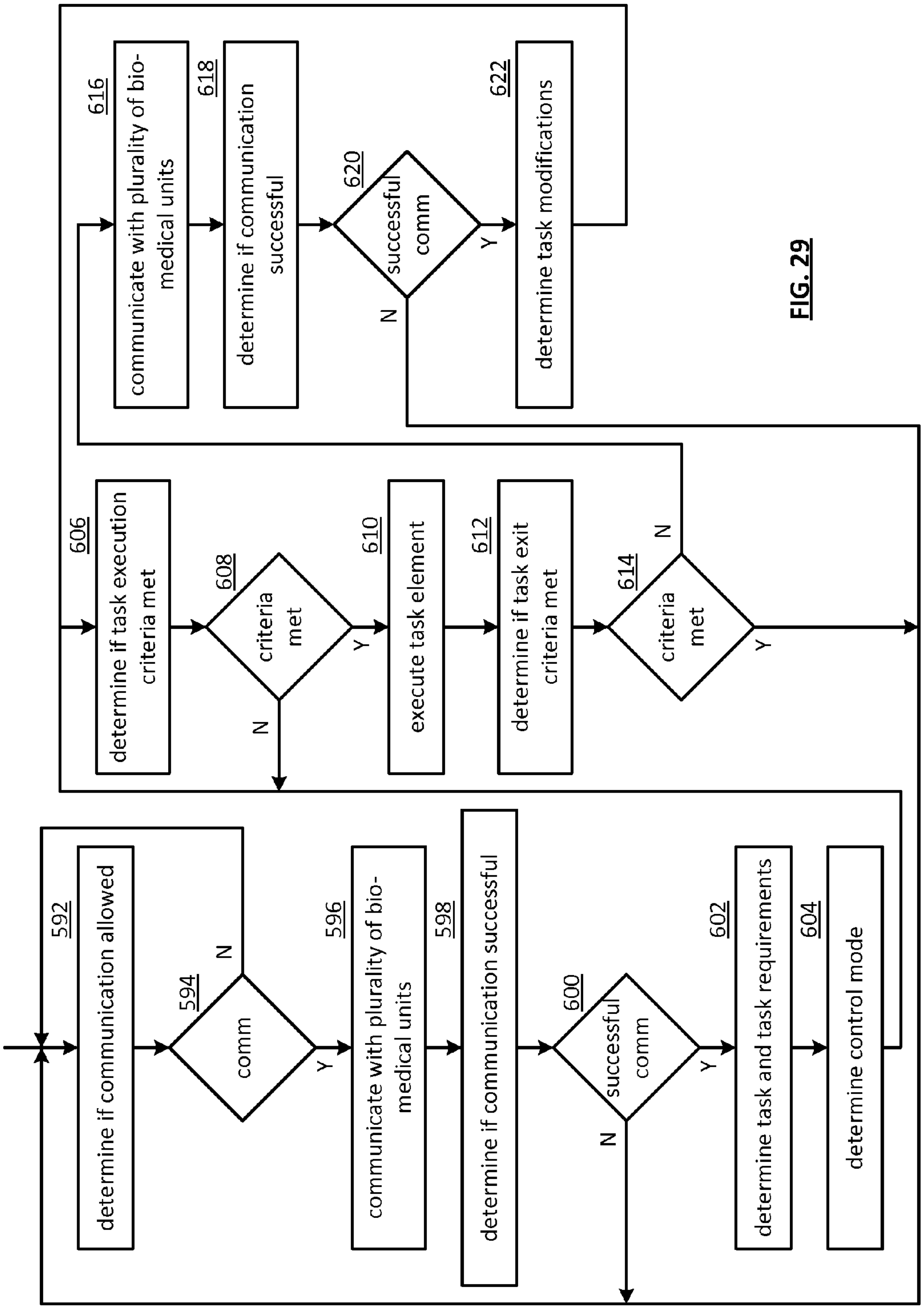
592
determine if communication allowed
594
comm
N
Y
596
communicate with plurality of bio-medical units
598
determine if communication successful
600
successful comm
N
Y
602
determine task and task requirements
604
determine control mode
606
determine if task execution criteria met
608
criteria met
N
Y
610
execute task element
612
determine if task exit criteria met
614
criteria met
N
Y
616
communicate with plurality of bio-medical units
618
determine if communication successful
620
successful comm
N
Y
622
determine task modifications

FIG. 29

BIO-MEDICAL UNIT WITH POWER HARVESTING MODULE AND RF COMMUNICATION

[0001] This patent application is claiming priority under 35 USC §119 to a provisionally filed patent application entitled BIO-MEDICAL UNIT AND APPLICATIONS THEREOF, having a provisional filing date of Sep. 30, 2009, and a provisional Ser. No. 61/247,060.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] Not Applicable

INCORPORATION-BY-REFERENCE OF MATERIAL SUBMITTED ON A COMPACT DISC

[0003] Not Applicable

BACKGROUND OF THE INVENTION

[0004] 1. Technical Field of the Invention

[0005] This invention relates generally to medical equipment and more particularly to wireless medical equipment.

[0006] 2. Description of Related Art

[0007] As is known, there is a wide variety of medical equipment that aids in the diagnosis, monitoring, and/or treatment of patients' medical conditions. For instances, there are diagnostic medical devices, therapeutic medical devices, life support medical devices, medical monitoring devices, medical laboratory equipment, etc. As specific exampled magnetic resonance imaging (MRI) devices produce images that illustrate the internal structure and function of a body.

[0008] The advancement of medical equipment is in step with the advancements of other technologies (e.g., radio frequency identification (RFID), robotics, etc.). Recently, RFID technology has been used for in vitro use to store patient information for easy access. While such in vitro applications have begun, the technical advancement in this area is in its infancy.

BRIEF SUMMARY OF THE INVENTION

[0009] The present invention is directed to apparatus and methods of operation that are further described in the following Brief Description of the Drawings, the Detailed Description of the Invention, and the claims. Other features and advantages of the present invention will become apparent from the following detailed description of the invention made with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING(S)

[0010] FIG. **1** is a diagram of an embodiment of a system in accordance with the present invention;

[0011] FIG. **2** is a diagram of another embodiment of a system in accordance with the present invention;

[0012] FIG. **3** is a diagram of an embodiment of an artificial body part including one or more bio-medical units in accordance with the present invention;

[0013] FIG. **4** is a schematic block diagram of an embodiment of an artificial body part in accordance with the present invention;

[0014] FIG. **5** is a diagram of another embodiment of a system in accordance with the present invention;

[0015] FIG. **6** is a diagram of another embodiment of a system in accordance with the present invention;

[0016] FIG. **7** is a diagram of another embodiment of a system in accordance with the present invention;

[0017] FIG. **8** is a schematic block diagram of an embodiment of a bio-medical unit in accordance with the present invention;

[0018] FIG. **9** is a schematic block diagram of an embodiment of a power harvesting module in accordance with the present invention;

[0019] FIG. **10** is a schematic block diagram of another embodiment of a power harvesting module in accordance with the present invention;

[0020] FIG. **11** is a schematic block diagram of another embodiment of a power harvesting module in accordance with the present invention;

[0021] FIG. **12** is a schematic block diagram of another embodiment of a power harvesting module in accordance with the present invention;

[0022] FIG. **13** is a schematic block diagram of another embodiment of a bio-medical unit in accordance with the present invention;

[0023] FIG. **14** is a diagram of another embodiment of a system in accordance with the present invention;

[0024] FIG. **15** is a diagram of an example of a communication protocol within a system in accordance with the present invention;

[0025] FIG. **16** is a diagram of an embodiment of a network of bio-medical units in accordance with the present invention;

[0026] FIG. **17** is a logic diagram of an embodiment of a method for bio-medical unit communications in accordance with the present invention;

[0027] FIG. **18** is a diagram of an embodiment of a network of bio-medical units that include MEMS robotics in accordance with the present invention;

[0028] FIG. **19** is a diagram of another embodiment of a network of bio-medical units that include MEMS robotics in accordance with the present invention;

[0029] FIG. **20** is a diagram of an embodiment of a bio-medical unit collecting image data in accordance with the present invention;

[0030] FIG. **21** is a diagram of an embodiment of a bio-medical unit collecting audio and/or ultrasound data in accordance with the present invention;

[0031] FIG. **22** is a logic diagram of an embodiment of a method for bio-medical unit communications in accordance with the invention;

[0032] FIG. **23** is a logic diagram of an embodiment of a method for MMW communications within a MRI sequence in accordance with the invention;

[0033] FIG. **24** is a logic diagram of an embodiment of a method for processing of MRI signals in accordance with the present invention;

[0034] FIG. **25** is a logic diagram of an embodiment of a method for communication utilizing MRI signals in accordance with the present invention;

[0035] FIG. **26** is a schematic block diagram of another embodiment of a bio-medical unit in accordance with the present invention;

[0036] FIG. **27** is a logic diagram of another embodiment of a method for bio-medical unit communications in accordance with the invention;

[0037] FIG. **28** is a logic diagram of another embodiment of a method for bio-medical unit communications in accordance with the invention; and

[0038] FIG. **29** is a logic diagram of an embodiment of a method for coordination of bio-medical unit task execution in accordance with the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0039] FIG. **1** is a diagram of an embodiment of a system that includes a plurality of bio-medical units **10** embedded within a body and/or placed on the surface of the body to facilitate diagnosis, treatment, and/or data collections. Each of the bio-medical units **10** is a passive device (e.g., it does not include a power source (e.g., a battery)) and, as such, includes a power harvesting module. The bio-medical units **10** may also include one or more of memory, a processing module, and functional modules. Alternatively, or in addition to, each of the bio-medical units **10** may include a power source.

[0040] In operation, a transmitter emits **12** electromagnetic signals **16** that pass through the body and are received by a receiver **14**. The transmitter **12** and receiver **14** may be part of a piece of medical diagnostic equipment (e.g., magnetic resonance imaging (MRI), X-ray, etc.) or independent components for stimulating and communicating with the network of bio-medical units in and/or on a body. One or more of the bio-medical units **10** receives the transmitted electromagnetic signals **16** and generates a supply voltage therefrom. Examples of this will be described in greater detail with reference to FIGS. **8-12**.

[0041] Embedded within the electromagnetic signals **16** (e.g., radio frequency (RF) signals, millimeter wave (MMW) signals, MRI signals, etc.) or via separate signals, the transmitter **12** communicates with one or more of the bio-medical units **10**. For example, the electromagnetic signals **16** may have a frequency in the range of a few MHz to 900 MHz and the communication with the bio-medical units **10** is modulated on the electromagnetic signals **16** at a much higher frequency (e.g., 5 GHz to 300 GHz). As another example, the communication with the bio-medical units **10** may occur during gaps (e.g., per protocol of medical equipment or injected for communication) of transmitting the electromagnetic signals **16**. As another example, the communication with the bio-medical units **10** occurs in a different frequency band and/or using a different transmission medium (e.g., use RF or MMW signals when the magnetic field of the electromagnetic signals are dominate, use ultrasound signals when the electromagnetic signals **16** are RF and/or MMW signals, etc.).

[0042] One or more of the bio-medical units **10** receives the communication signals **18** and processes them accordingly. The communication signals **18** may be instructions to collect data, to transmit collected data, to move the unit's position in the body, to perform a function, to administer a treatment, etc. If the received communication signals **18** require a response, the bio-medical unit **10** prepares an appropriate response and transmits it to the receiver **14** using a similar communication convention used by the transmitter **12**.

[0043] FIG. **2** is a diagram of another embodiment of a system that includes a plurality of bio-medical units **10** embedded within a body and/or placed on the surface of the body to facilitate diagnosis, treatment, and/or data collections. Each of the bio-medical units **10** is a passive device and, as such, includes a power harvesting module. The bio-medical units **10** may also include one or more of memory, a processing module, and functional modules. In this embodiment, the person is placed in an MRI machine (fixed or portable) that generates a magnetic field **26** through which the MRI transmitter **20** transmits MRI signals **28** to the MRI receiver **22**.

[0044] One or more of the bio-medical units **10** powers itself by harvesting energy from the magnetic field **26** or changes thereof as produced by gradient coils, from the magnetic fields of the MRI signals **28**, from the electrical fields of the MRI signals **28**, and/or from the electromagnetic aspects of the MRI signals **28**. A unit **10** converts the harvested energy into a supply voltage that supplies other components of the unit (e.g., a communication module, a processing module, memory, a functional module, etc.).

[0045] A communication device **24** communicates data and/or control communications **30** with one or more of the bio-medical units **10** over one or more wireless links. The communication device **24** may be a separate device from the MRI machine or integrated into the MRI machine. For example, the communication device **24**, whether integrated or separate, may be a cellular telephone, a computer with a wireless interface (e.g., a WLAN station and/or access point, Bluetooth, a proprietary protocol, etc.), etc. A wireless link may be one or more frequencies in the ISM band, in the 60 GHz frequency band, the ultrasound frequency band, and/or other frequency bands that supports one or more communication protocols (e.g., data modulation schemes, beamforming, RF or MMW modulation, encoding, error correction, etc.).

[0046] The composition of the bio-medical units **10** includes non-ferromagnetic materials (e.g., paramagnetic or diamagnetic) and/or metal alloys that are minimally affected by an external magnetic field **26**. In this regard, the units harvest power from the MRI signals **28** and communicate using RF and/or MMW electromagnetic signals with negligible chance of encountering the projectile or missile effect of implants that include ferromagnetic materials.

[0047] FIG. **3** is a diagram of an embodiment of an artificial body part **32** including one or more bio-medical units **10** that may be surgically implanted into a body. The artificial body part **32** may be a pace maker, a breast implant, a joint replacement, an artificial bone, splints, fastener devices (e.g., screws, plates, pins, sutures, etc.), artificial organ, etc. The artificial body part **32** may be permanently embedded in the body or temporarily embedded into the body.

[0048] FIG. **4** is a schematic block diagram of an embodiment of an artificial body part system **32** that includes an artificial body part and one or more bio-medical units **10**. The artificial body part (e.g., an electronic medical aid device (e.g., pacemaker, nerve stimulator, etc.), a metal artificial bone or portion thereof, a non-metal artificial bone or portion thereof, a metal screw, a non-metal screw, a metal plate, a non-metal plate, a metal splint, a non-metal splint, a metal suture, a non-metal suture, and/or an adhesive suture) includes a bio-medical mounting mechanism (e.g., a cavity, a surface mount area, etc.).

[0049] The bio-medical unit is mechanically associated with the artificial body part based on the bio-medical mounting mechanism. For example, if the artificial body part includes one or more cavities, the bio-medical unit **10** may positioned within the one or more cavities. As another example, if the artificial body part is an adhesive suture, the bio-medical unit **10** may be mounted on the surface of the suture and/or under a layer of the suture. As yet another example, if the artificial body part is molded, the bio-medical unit may be embedded within the artificial body part during the molding process.

[0050] The bio-medical unit includes a power harvesting module, a wireless communication module, and a functional module. The power harvesting module generates a supply voltage from a wireless source (e.g., MRI signals, magnetic fields, RF signals, MMW signals, etc.). The wireless communication module wirelessly communicates inbound data and outbound data with a communication device (e.g., device **24**). One or more embodiments of the power harvesting module and/or of the wireless communication module are discussed with reference to one or more of the preceding and/or following Figures.

[0051] The functional module performs at least one function regarding the artificial body part, wherein the functional module performs the at least one function in response to the inbound data and/or to generate the outbound data. For example, the at least one function may include a treatment function, an adjustment function, a diagnostic function, a sampling function, and/or a monitoring function. As a more specific example, the treatment function may include applying a medication, applying a laser treatment, applying an ultrasound treatment, and/or applying an electronic stimulation treatment. As another more specific example, the adjustment function may include adjusting force of the artificial body part (e.g., adjusting tension of a splint) and/or adjusting torque of the artificial body part (e.g., turning a screw).

[0052] In an embodiment, the bio-medical unit further includes a coil and/or an antenna. The power harvesting module may use the coil to generate the supply voltage from electromagnetic energy and/or use the antenna to generate the supply voltage from RF signals and/or MMW signals. The wireless communication module may use the coil for near field communication and/or use the antenna for RF and/or MMW communication. Note that the coil and/or antenna may be at least partially embedded within the artificial body part (e.g., at least partially in a cavity, forming threads of a screw, etc.). Alternatively, or in addition to, the artificial body part functions as the at least one of the coil and the antenna. For example, a metal screw, plate, splint, etc. may be formed to provide the coil and/or antenna, or include a section of the artificial body part to provide the coil and/or antenna.

[0053] In a further embodiment of the artificial body part system **32** includes the artificial body part and a plurality of bio-medical units **10**. The bio-medical units **10** may perform one or more functions. For example, a first bio-medical unit performs a treatment function; a second bio-medical unit performs an adjustment function; a third bio-medical unit performs a diagnostic function; a fourth bio-medical unit performs a sampling function; and a fifth bio-medical unit performs a monitoring function.

[0054] As another example, one bio-medical unit **10** may be used to detect infections, the body's acceptance of the artificial body part **32**, measure localized body temperature, monitor performance of the artificial body part **32**, and/or data gathering for other diagnostics. Another bio-medical unit **10** may be used for deployment of treatment (e.g., disperse medication, apply electrical stimulus, apply RF radiation, apply laser stimulus, etc.). Yet another bio-medical unit **10** may be used to adjust the position of the artificial body part **32** and/or a setting of the artificial body part **32**. For example, a bio-medical unit **10** may be used to mechanically adjust the tension of a splint, screws, etc. As another example, a bio-medical unit **10** may be used to adjust an electrical setting of the artificial body part **32**.

[0055] FIG. **5** is a diagram of another embodiment of a system that includes a plurality of bio-medical units **10** and one or more communication devices **24** coupled to a wide area network (WAN) communication device **34** (e.g., a cable modem, DSL modem, base station, access point, hot spot, etc.). The WAN communication device **34** is coupled to a network **42** (e.g., cellular telephone network, internet, etc.), which is coupled to it a plurality of remote monitors **36**, a plurality of databases **40**, and a plurality of computers **38**.

[0056] In an example of operation, one or more of the remote monitors **36** may receive images and/or other data **30** from one or more of the bio-medical units **10** via the communication device **24**, the WAN communication device **34**, and the network **42**. In this manner, a person(s) operating the remote monitors **36** may view images and/or the data **30** gathered by the bio-medical units **10**. This enables a specialist to be consulted without requiring the patient to travel to the specialist's office.

[0057] In another example of operation, one or more of the computers **38** may communicate with the bio-medical units **10** via the communication device **24**, the WAN communication device **34**, and the network **42**. In this example, the computer **36** may provide commands **30** to one or more of the bio-medical units **10** to gather data, to dispense a medication, to move to a new position in the body, to perform a mechanical function (e.g., cut, grasp, drill, puncture, stitch, patch, etc.), etc. As such, the bio-medical units **10** may be remotely controlled via one or more of the computers **36**.

[0058] In another example of operation, one or more of the bio-medical units **10** may read and/or write data from or to one or more of the databases **40**. For example, data (e.g., a blood sample analysis) generated by one or more of the bio-medical units **10** may be written to one of the databases **40**. The communication device **24** and/or one of the computers **36** may control the writing of data to or the reading of data from the database(s) **40**. The data may further include medical records, medical images, prescriptions, etc.

[0059] FIG. **6** is a diagram of another embodiment of a system that includes a plurality of bio-medical units **10**. In this embodiment, the bio-medical units **10** can communicate with each other directly and/or communicate with the communication device **24** directly. The communication medium may be an infrared channel(s), an RF channel(s), a MMW channel(s), and/or ultrasound. The units may use a communication protocol such as token passing, carrier sense, time division multiplexing, code division multiplexing, frequency division multiplexing, etc.

[0060] FIG. **7** is a diagram of another embodiment of a system that includes a plurality of bio-medical units **10**. In this embodiment, one of the bio-medical units **44** functions as an access point for the other units. As such, the designated unit **44** routes communications between the units **10** and between one or more units **10** and the communication device **24**. The communication medium may be an infrared channel (s), an RF channel(s), a MMW channel(s), and/or ultrasound. The units **10** may use a communication protocol such as token passing, carrier sense, time division multiplexing, code division multiplexing, frequency division multiplexing, etc.

[0061] FIG. **8** is a schematic block diagram of an embodiment of a bio-medical unit **10** that includes a power harvesting module **46**, a communication module **48**, a processing module **50**, memory **52**, and one or more functional modules **54**. The processing module **50** may be a single processing device or a plurality of processing devices. Such a processing device may be a microprocessor, micro-controller, digital signal processor, microcomputer, central processing unit, field programmable gate array, programmable logic device, state machine, logic circuitry, analog circuitry, digital circuitry, and/or any device that manipulates signals (analog and/or digital) based on hard coding of the circuitry and/or operational instructions. The processing module **50** may have an associated memory **52** and/or memory element, which may be a single memory device, a plurality of memory devices, and/or embedded circuitry of the processing module. Such a memory device **52** may be a read-only memory, random access memory, volatile memory, non-volatile memory, static memory, dynamic memory, flash memory, cache memory, and/or any device that stores digital information. Note that if the processing module **50** includes more than one processing device, the processing devices may be centrally located (e.g., directly coupled together via a wired and/or wireless bus structure) or may be distributedly located (e.g., cloud computing via indirect coupling via a local area network and/or a wide area network). Further note that when the processing module **50** implements one or more of its functions via a state machine, analog circuitry, digital circuitry, and/or logic circuitry, the memory and/or memory element storing the corresponding operational instructions may be embedded within, or external to, the circuitry comprising the state machine, analog circuitry, digital circuitry, and/or logic circuitry. Still further note that, the memory element stores, and the processing module executes, hard coded and/or operational instructions corresponding to at least some of the steps and/or functions illustrated in FIGS. **1-29**.

[0062] The power harvesting module **46** may generate one or more supply voltages **56** (Vdd) from one or more of MRI electromagnetic signals **16**, magnetic fields **26**, RF signals, MMW signals, and body motion. The power harvesting module **46** may be implemented as disclosed in U.S. Pat. No. 7,595,732 to generate one or more supply voltages from an RF signal. Alternatively, or in addition to, the power harvesting module **46** may be implemented as shown in one or more FIGS. **9-11** to generate one or more supply voltages **56** from an MRI signal **28** and/or magnetic field **26**. As yet another alternative, or a further addition, the power harvesting module **46** may be implemented as shown in FIG. **12** to generate one or more supply voltage **56** from body motion.

[0063] The communication module **48** may include a receiver section and a transmitter section. The transmitter section converts an outbound symbol stream into an outbound RF or MMW signal **60** that has a carrier frequency within a given frequency band (e.g., 900 MHz, 2.5 GHz, 5 GHz, 57-66 GHz, etc.). In an embodiment, this may be done by mixing the outbound symbol stream with a local oscillation to produce an up-converted signal. One or more power amplifiers and/or power amplifier drivers amplifies the up-converted signal, which may be RF or MMW bandpass filtered, to produce the outbound RF or MMW signal **60**. In another embodiment, the transmitter section includes an oscillator that produces an oscillation. The outbound symbol stream provides phase information (e.g., +/−Δθ[phase shift] and/or θ(t) [phase modulation]) that adjusts the phase of the oscillation to produce a phase adjusted RF or MMW signal, which is transmitted as the outbound RF signal **60**. In another embodiment, the outbound symbol stream includes amplitude information (e.g., A(t) [amplitude modulation]), which is used to adjust the amplitude of the phase adjusted RF or MMW signal to produce the outbound RF or MMW signal **60**.

[0064] In yet another embodiment, the transmitter section includes an oscillator that produces an oscillation. The outbound symbol provides frequency information (e.g., +/−Δf [frequency shift] and/or f(t) [frequency modulation]) that adjusts the frequency of the oscillation to produce a frequency adjusted RF or MMW signal, which is transmitted as the outbound RF or MMW signal **60**. In another embodiment, the outbound symbol stream includes amplitude information, which is used to adjust the amplitude of the frequency adjusted RF or MMW signal to produce the outbound RF or MMW signal **60**. In a further embodiment, the transmitter section includes an oscillator that produces an oscillation. The outbound symbol provides amplitude information (e.g., +/−ΔA [amplitude shift] and/or A(t) [amplitude modulation) that adjusts the amplitude of the oscillation to produce the outbound RF or MMW signal **60**.

[0065] The receiver section amplifies an inbound RF or MMW signal **60** to produce an amplified inbound RF or MMW signal. The receiver section may then mix in-phase (I) and quadrature (Q) components of the amplified inbound RF or MMW signal with in-phase and quadrature components of a local oscillation to produce a mixed I signal and a mixed Q signal. The mixed I and Q signals are combined to produce an inbound symbol stream. In this embodiment, the inbound symbol may include phase information (e.g., +/−Δθ[phase shift] and/or θ(t) [phase modulation]) and/or frequency information (e.g., +/−Δf [frequency shift] and/or f(t) [frequency modulation]). In another embodiment and/or in furtherance of the preceding embodiment, the inbound RF or MMW signal includes amplitude information (e.g., +/−ΔA [amplitude shift] and/or A(t) [amplitude modulation]). To recover the amplitude information, the receiver section includes an amplitude detector such as an envelope detector, a low pass filter, etc. Note that, as may be discussed herein, RF may include signals within the radio frequency band of 30 Hz to 3 GHz and/or within the microwave (or MMW) frequency band of 3 GHz to 300 GHz.

[0066] The processing module **50** generates the outbound symbol stream from outbound data and converts the inbound symbol stream into inbound data. For example, the processing module **50** converts the inbound symbol stream into inbound data (e.g., voice, text, audio, video, graphics, etc.) in accordance with one or more wireless communication standards (e.g., GSM, CDMA, WCDMA, HSUPA, HSDPA, WiMAX, EDGE, GPRS, IEEE 802.11, Bluetooth, ZigBee, universal mobile telecommunications system (UMTS), long term evolution (LTE), IEEE 802.16, evolution data optimized (EV-DO), etc.). Such a conversion may include one or more of: digital intermediate frequency to baseband conversion, time to frequency domain conversion, space-time-block decoding, space-frequency-block decoding, demodulation, frequency spread decoding, frequency hopping decoding, beamforming decoding, constellation demapping, deinterleaving, decoding, depuncturing, and/or descrambling.

[0067] As another example, the processing module **50** converts outbound data (e.g., voice, text, audio, video, graphics, etc.) into outbound symbol stream in accordance with one or more wireless communication standards (e.g., GSM, CDMA, WCDMA, HSUPA, HSDPA, WiMAX, EDGE, GPRS, IEEE 802.11, Bluetooth, ZigBee, universal mobile telecommunications system (UMTS), long term evolution (LTE), IEEE

802.16, evolution data optimized (EV-DO), etc.). Such a conversion includes one or more of: scrambling, puncturing, encoding, interleaving, constellation mapping, modulation, frequency spreading, frequency hopping, beamforming, space-time-block encoding, space-frequency-block encoding, frequency to time domain conversion, and/or digital baseband to intermediate frequency conversion.

[0068] Each of the one or more functional modules **54** provides a function to support treatment, data gathering, motion, repairs, and/or diagnostics. The functional modules **54** may be implemented using nanotechnology and/or microelectronic mechanical systems (MEMS) technology. Various examples of functional modules **54** are illustrated in one or more of FIGS. **13**-**29**.

[0069] The bio-medical unit **10** may be encapsulated by an encapsulate **58** that is non-toxic to the body. For example, the encapsulate **58** may be a silicon based product, a non-ferromagnetic metal alloy (e.g., stainless steel), etc. As another example, the encapsulate **58** may include a spherical shape and have a ferromagnetic liner that shields the unit from a magnetic field and to offset the forces of the magnetic field.

[0070] The bio-medical unit **10** may be implemented on a single die that has an area of a few millimeters or less. The die may be fabricated in accordance with CMOS technology, Gallium-Arsenide technology, and/or any other integrated circuit die fabrication process.

[0071] FIG. **9** is a schematic block diagram of an embodiment of a power harvesting module **46** that includes an array of on-chip air core inductors **64**, a rectifying circuit **66**, capacitors, and a regulation circuit **68**. The inductors **64** may each have an inductance of a few nano-Henries to a few micro-Henries and may be coupled in series, in parallel, or a series parallel combination.

[0072] In an example of operation, the MRI transmitter **20** transmits MRI signals **28** at a frequency of 3-45 MHz at a power level of up to 35 KWatts. The air core inductors **64** are electromagnetically coupled to generate a voltage from the magnetic and/or electric field generated by the MRI signals **28**. Alternatively, or in addition to, the air core inductors **64** may generate a voltage from the magnetic field **26** and changes thereof produced by the gradient coils. The rectifying circuit **66** rectifies the AC voltage produced by the inductors to produce a first DC voltage. The regulation circuit generates one or more desired supply voltages **56** from the first DC voltage.

[0073] The inductors **64** may be implemented on one more metal layers of the die and include one or more turns per layer. Note that trace thickness, trace length, and other physical properties affect the resulting inductance.

[0074] FIG. **10** is a schematic block diagram of another embodiment of a power harvesting module **46** that includes a plurality of on-chip air core inductors **70**, a plurality of switching units (S), a rectifying circuit **66**, a capacitor, and a switch controller **72**. The inductors **70** may each having an inductance of a few nano-Henries to a few micro-Henries and may be coupled in series, in parallel, or a series parallel combination.

[0075] In an example of operation, the MRI transmitter **20** transmits MRI signals **28** at a frequency of 3-45 MHz at a power level of up to 35 KWatts. The air core inductors **70** are electromagnetically coupled to generate a voltage from the magnetic and/or electric field generated by the MRI signals **28**. The switching module **72** engages the switches via control signals **74** to couple the inductors **70** in series and/or parallel to generate a desired AC voltage. The rectifier circuit **66** and the capacitor(s) convert the desired AC voltage into the one or more supply voltages **56**.

[0076] FIG. **11** is a schematic block diagram of another embodiment of a power harvesting module **46** that includes a plurality of Hall effect devices **76**, a power combining module **78**, and a capacitor(s). In an example of operation, the Hall effect devices **76** generate a voltage based on the constant magnetic field (H) and/or a varying magnetic field. The power combining module **78** (e.g., a wire, a switch network, a transistor network, a diode network, etc.) combines the voltages of the Hall effect devices **76** to produce the one or more supply voltages **56**.

[0077] FIG. **12** is a schematic block diagram of another embodiment of a power harvesting module **46** that includes a plurality of piezoelectric devices **82**, a power combining module **78**, and a capacitor(s). In an example of operation, the piezoelectric devices **82** generate a voltage based on body movement, ultrasound signals, movement of body fluids, etc. The power combining module **78** (e.g., a wire, a switch network, a transistor network, a diode network, etc.) combines the voltages of the Hall effect devices **82** to produce the one or more supply voltages **56**. Note that the piezoelectric devices **82** may include one or more of a piezoelectric motor, a piezoelectric actuator, a piezoelectric sensor, and/or a piezoelectric high voltage device.

[0078] The various embodiments of the power harvesting module **46** may be combined to generate more power, more supply voltages, etc. For example, the embodiment of FIG. **9** may be combined with one or more of the embodiments of FIGS. **11** and **12**.

[0079] FIG. **13** is a schematic block diagram of another embodiment of a bio-medical unit **10** that includes a power harvesting module **46**, a communication module **48**, a processing module **50**, memory **52**, and may include one or more functional modules **54** and/or a Hall effect communication module **116**. The communication module **48** may include one or more of an ultrasound transceiver **118**, an electromagnetic transceiver **122**, an RF and/or MMW transceiver **120**, and a light source (LED) transceiver **124**. Note that examples of the various types of communication modules **48** will be described in greater detail with reference to one or more of FIGS. **14**-**29**.

[0080] The one or more functional modules **54** may perform a repair function, an imaging function, and/or a leakage detection function, which may utilize one or more of a motion propulsion module **96**, a camera module **98**, a sampling robotics module **100**, a treatment robotics module **102**, an accelerometer module **104**, a flow meter module **106**, a transducer module **108**, a gyroscope module **110**, a high voltage generator module **112**, a control release robotics module **114**, and/or other functional modules described with reference to one or more other figures. The functional modules **54** may be implemented using MEMS technology and/or nanotechnology. For example, the camera module **98** may be implemented as a digital image sensor in MEMS technology. Example of these various modules will be described in greater detail with reference to one or more of FIGS. **14**-**29**.

[0081] The Hall effect communication module **116** utilizes variations in the magnetic field and/or electrical field to produce a plus or minus voltage, which can be encoded to convey information. For example, the charge applied to one or more Hall effect devices **76** may be varied to produce the voltage change. As another example, an MRI transmitter **20** and/or gradient unit may modulate a signal on the magnetic field **26** it generates to produce variations in the magnetic field **26**.

[0082] FIG. **14** is a diagram of another embodiment of a system that includes one or more bio-medical units **10**, a transmitter unit **126**, and a receiver unit **128**. Each of the bio-medical units **10** includes a power harvesting module **46**, an RF and/or MMW (RF) transceiver **138**, a processing module **50**, memory **52** and one or more functional modules (as shown in one or more of the previous figures). The transmitter unit **126** includes an electromagnetic (e.g., a MRI) transmitter **130** and a MMW transmitter **132**. The receiver unit **128** includes an electromagnetic (e.g., MRI) receiver **134** and a MMW receiver **136**. Note that the MMW transmitter **132** and MMW receiver **136** may be in the same unit (e.g., in the transmitter unit, in the receiver unit, or housed in a separate device).

[0083] In an example of operation, the electromagnetic transmitter **130** transmits an electromagnetic (EM) signal **146** in a predetermined pattern (e.g., a magnetic resonance imaging (MRI) signal having at least one of a radio frequency (RF) component and a gradient magnetic field component that has a pattern as shown in FIG. **15**). In addition, the RF communication transmitter **132** of the transmitter unit **126** transmits a downstream RF communication signal, which may contain a request for data, a command, etc.

[0084] The power harvesting module **46** receives the EM signal **146** and generates a power supply voltage therefrom. The RF communication module **138**, which is powered by the power supply voltage, receives the downstream RF communication signal during at least one of the null intervals of the transmission pattern. The RF communication module **138** then converts the downstream RF communication signal into downstream information, which includes at least one of downstream data and a downstream instruction.

[0085] The functional module performs a bio-medical function based on the at least one of downstream data and the downstream instruction. For example, the bio-medical function may be to take a temperature reading, to release a certain amount of a medication, to take a blood-oxygen reading, etc. As another example, the bio-medical function may cause the bio-medical unit to store the downstream data.

[0086] In response to the downstream instruction, or based on another stimulus (e.g., pre-programmed, timed interval, etc.), the bio-medical unit **10** may have data and/or instruction to transmit. In this instance, the function module generates at least one of an upstream instruction and upstream data. The processing module converts the upstream instruction and/or upstream data into upstream information.

[0087] The RF communication module **138** interprets the predetermined pattern of the transmitting of the electromagnetic signal to determine null intervals. In addition, the RF communication module **138** converts the upstream information into an upstream RF communication signal and transmits the upstream RF communication signal during one or more of the null intervals. The RF communication receiver **134** of the receiver unit **128** receives the upstream RF communication signal and processes it accordingly.

[0088] FIG. **15** is a diagram of an example of a communication protocol within the system of FIG. **14**. In this diagram, the MRI transmitter **20** transmits RF signals **152**, which have a frequency in the range of 3-45 MHz, at various intervals with varying signal strengths. The power harvesting module **46** of the bio-medical units **10** may use these signals to generate power for the bio-medical unit **10**.

[0089] In addition to the MRI transmitter **20** transmitting its signal, a constant magnetic field and various gradient magnetic fields **154-164** are created (one or more in the x dimension Gx, one or more in the y dimension Gy, and one or more in the z direction Gz). The power harvesting module **46** of the bio-medical unit **10** may further use the constant magnetic field and/or the varying magnetic fields **154-164** to create power for the bio-medical unit **10**.

[0090] During non-transmission periods of the cycle, the bio-medical unit **10** may communicate **168** with the MMW transmitter **132** and/or MMW receiver **136**. In this regard, the bio-medical unit **10** alternates from generating power to MMW communication in accordance with the conventional transmission-magnetic field pattern of an MRI machine.

[0091] FIG. **16** is a schematic block diagram of an embodiment of networked bio-medical units **10** that communicate with each other, perform sensing functions to produce sensed data **218-232**, process the sensed data to produce processed data, and transmit the processed data **216**. The bio-medical units **10** may be positioned in a body part to sense data across the body part and to transmit data to an external communication device. The transmitted data may be further processed or aggregated from sensed data.

[0092] The bio-medical units **10** may monitor various types of biological functions over a short term or a long term to produce the sensed data **218-232**. Note that the sensed data **218-232** may include blood flow rate, blood pressure, temperature, air flow, blood oxygen level, density, white cell count, red cell count, position information, etc.

[0093] The bio-medical unit **10** establishes communications with one or more other bio-medical units **10** to facilitate the communication of sensed data **218-232** and processed data **216**. The communication may include EM signals, MMW signals, optical signals, sound signals, and/or RF signals.

[0094] The bio-medical unit **10** may determine position information based on the sensed data **218-232** and include the position information in the communication. The bio-medical unit **10** may also determine a mode of operation based on one or more of a command, a list, a predetermination, sensed data, and/or processed data. For example, a bio-medical unit **10** at the center of the body part may be in a mode to sense temperature and a bio-medical unit **10** at the outside edge of the body part may sense blood flow.

[0095] The bio-medical unit **10** may receive processed data **218-232** from another bio-medical unit and re-send the same processed data **218-232** to yet another bio-medical unit **10**. The bio-medical unit **10** may produce processed data based on sensed data **218-232** from the bio-medical unit **10** and/or received processed data from another bio-medical unit **10**.

[0096] FIG. **17** is a flowchart illustrating the processing of networked bio-medical unit data where the bio-medical unit determines the sense mode based on one or more of a predetermination, a stored mode indicator in memory, a command, and/or a dynamic sensed data condition. The method begins at step **234** where the bio-medical unit **10** determines the mode. The method branches to step **240** when the bio-medical unit **10** determines that the mode is process and sense. The method continues to step **236** when the bio-medical unit **10** determines that the mode is sense only.

[0097] At step **236**, the bio-medical unit **10** gathers data from one or more of the functional modules **54** to produce sensed data. The bio-medical unit **10** may transmit the sensed data **238** to another bio-medical unit **10** and/or an external communication device in accordance with the sense mode. For example, the bio-medical unit **10** may transmit the sensed data at a specific time, to a specific bio-medical unit **10**, to a specific external communication device, after a certain time period, when the data is sensed, and/or when the sensed data compares favorably to a threshold (e.g., a temperature trip point).

[0098] The method continues at step **240** where the bio-medical unit **10** determines whether it has received data from another unit **10**. If not, the method continues to step **250**, where the bio-medical unit **10** transmits its sensed data to another bio-medical unit **10** and/or an external communication device in accordance with the sense mode.

[0099] When the bio-medical unit **10** has received data from another unit, the method continues at step **242**, where the bio-medical unit **10** determines a data function to perform based on one or more of the content of the received data, the sensed data, a command, and/or a predetermination. The data function may one or more of initialization, comparing, compiling, and/or performing a data analysis algorithm.

[0100] The method continues at step **244**, where the bio-medical unit **10** gathers data from the functional modules **54**, and/or the received data from one or more other bio-medical units **10**. The method continues at step **246**, where the bio-medical unit **10** processes the data in accordance with a function to produce processed data. In addition to the example provided above, the function may also include the functional assignment of the bio-medical unit **10** as determined by a predetermination, a command, sensed data, and/or processed data (e.g., measure blood pressure from the plurality of bio-medical units and summarize the high, low, and average).

[0101] The method continues at step **248**, where the bio-medical unit **10** transmits the processed data to another bio-medical unit **10** and/or to an external communication device in accordance with the sense mode. For example, the bio-medical unit **10** may transmit the sensed data at a specific time, to a specific bio-medical unit **10**, to a specific external communication device, after a certain time period, when the data is sensed, and/or when the sensed data compares favorably to a threshold (e.g., a temperature trip point). Note that the communication protocol may be the same or different between bio-medical units **10** and/or between the bio-medical unit **10** and the external communication device.

[0102] FIG. **18** is a schematic block diagram of an embodiment of a parent bio-medical unit (on the left) communicating with an external unit to coordinates the functions of one or more children bio-medical units **10** (on the right). The parent unit includes a communication module **48** for external communications, a communication module **48** for communication with the children units, the processing module **50**, the memory **52**, and the power harvesting module **46**. Note that the parent unit may be implemented one or more chips and may in the body or one the body.

[0103] Each of the child units includes a communication module **48** for communication with the parent unit and/or other children units, a MEMS robotics **244**, and the power harvesting module **46**. The MEMS robotics **244** may include one or more of a MEMS technology saw, drill, spreader, needle, injection system, and actuator. The communication module **48** may support RF and/or MMW inbound and/or outbound signals **60** to the parent unit such that the parent unit may command the child units in accordance with external communications commands.

[0104] In an example of operation, the patent bio-medical unit receives a communication from the external source, where the communication indicates a particular function the child units are to perform. The parent unit processes the communication and relays relative portions to the child units in accordance with a control mode. Each of the child units receives their respective commands and performs the corresponding functions to achieve the desired function.

[0105] FIG. **19** is a schematic block diagram of another embodiment of a plurality of task coordinated bio-medical units **10** including a parent bio-medical unit **10** (on the left) and one or more children bio-medical units **10** (on the right). The parent unit may be implemented one or more chips and may in the body or one the body. The parent unit may harvest power in conjunction with the power booster **84**.

[0106] The parent unit includes the communication module **48** for external communications, the communication module **48** for communication with the children units, the processing module **50**, the memory **52**, a MEMS electrostatic motor **248**, and the power harvesting module **46**. The child unit includes the communication module **48** for communication with the parent unit and/or other children units, a MEMS electrostatic motor **248**, the MEMS robotics **244**, and the power harvesting module **46**. Note that the child unit has fewer components as compared to the parent unit and may be smaller facilitating more applications where smaller bio-medical units **10** enhances their effectiveness.

[0107] The MEMS robotics **244** may include one or more of a MEMS technology saw, drill, spreader, needle, injection system, and actuator. The MEMS electrostatic motor **248** may provide mechanical power for the MEMS robotics **244** and/or may provide movement propulsion for the child unit such that the child unit may be positioned to optimize effectiveness. The child units may operate in unison to affect a common task. For example, the plurality of child units may operate in unison to saw through a tissue area.

[0108] The child unit communication module **48** may support RF and/or MMW inbound and/or outbound signals **60** to the parent unit such that the parent unit may command the children units in accordance with external communications commands.

[0109] The child unit may determine a control mode and operate in accordance with the control mode. The child unit determines the control mode based on one or more of a command from a parent bio-medical unit, external communications, a preprogrammed list, and/or in response to sensor data. Note that the control mode may include autonomous, parent (bio-medical unit), server, and/or peer as previously discussed.

[0110] FIG. **20** is a schematic block diagram of an embodiment of a bio-medical unit **10** based imaging system that includes the bio-medical unit **10**, the communication device **24**, a database **254**, and an in vivo image unit **252**. The bio-medical unit **10** may perform scans and provide the in vivo image unit **252** with processed image data for diagnostic visualization.

[0111] The bio-medical unit **10** includes a MEMS image sensor **256**, the communication module **48** for external communications with the communication device, the processing module **50**, the memory **52**, the MEMS electrostatic motor **248**, and the power harvesting module **46**. In an embodiment the bio-medical unit **10** and communication device **24** communicate directly. In another embodiment, the bio-medical unit **10** and communication device **24** communicate through one or more intermediate networks (e.g., wireline, wireless, cellular, local area wireless, Bluetooth, etc.). The MEMS image sensor **256** may include one or more sensors scan types for optical signals, MMW signals, RF signals, EM signals, and/or sound signals.

[0112] The in vivo unit **252** may send a command to the bio-medical unit **10** via the communication device **24** to request scan data. The request may include the scan type. The in vivo unit **252** may receive the processed image data from the bio-medical unit **10**, compare it to data in the database **254**, process the data further, and provide image visualization.

[0113] FIG. **21** is a schematic block diagram of an embodiment of a bio-medical unit **10** based sounding system that includes the bio-medical unit **10**, the communication device **24**, the database **254**, and a speaker **260**. The bio-medical unit **10** may perform scans and provide the speaker **260** with processed sounding data for diagnostic purposes via the communication device **24**.

[0114] The bio-medical unit **10** includes a MEMS microphone **262**, the communication module **48** for external communications with the communication device **24**, the processing module **50**, the memory **52**, the MEMS electrostatic motor **248**, and the power harvesting module **46**. In an embodiment the bio-medical unit **10** and communication device **24** communicate directly. In another embodiment, the bio-medical unit **10** and communication device **24** communicate through one or more intermediate networks (e.g., wireline, wireless, cellular, local area wireless, Bluetooth, etc.) The MEMS microphone **262** may include one or more sensors to detect audible sound signals, sub-sonic sound signals, and/or ultrasonic sound signals.

[0115] The processing module **50** may produce the processed sounding data based in part on the received sound signals and in part on data in the database **254**. The processing module **50** may retrieve data via the communication module **48** and communication device **24** link from the database **254** to assist in the processing of the signals (e.g., pattern matching, filter recommendations, sound field types). The processing module **50** may process the signals to detect objects, masses, air flow, liquid flow, tissue, distances, etc. The processing module **50** may provide the processed sounding data to the speaker **260** for audible interpretation. In another embodiment, the bio-medical unit **10** assists an ultrasound imaging system by relaying ultrasonic sounds from the MEMS microphone **262** to the ultrasound imaging system instead of to the speaker **260**.

[0116] FIG. **22** is a flowchart of an embodiment of a method for controlling power harvesting within a bio-medical unit **10**. The method begins at step **418** wherein the processing module **50** of the bio-medical unit **10** initializes (e.g., when it is supplied power and wakes up) itself. For example, the processing module **50** executes an initialization boot sequence stored in the memory **52**. The initialization boot sequence includes operational instructions that cause the processing module to initialize its registers to accept further instructions. The initialization boot sequence may further include operational instructions to initialize one or more of the communication module **48**, the functional module(s) **54** initialized, etc.

[0117] The method continues at step **420** where the processing module **50** determines the state of the bio-medical unit (e.g., actively involved in a task, inactive, data gathering, performing a function, etc.). Such a determination may be based on one or more of previous state(s) (e.g., when the processing module was stopped prior to losing power), an input from the functional module **54**, a list of steps or elements of a task, the current step of a MRI sequence, and/or new tasks received via the communication module **48**.

[0118] The method continues at step **422** where the processing module **50** determines the bio-medical unit power level, which may be done by measuring the power harvesting module **46** output Vdd **56**. Note that voltage is one proxy for the power level and that other proxies may be utilized including estimation of milliWatt-hours available, a time of operation before loss of operating power estimate, a number of CPU instructions estimate, a number of task elements, a number of tasks estimate, and/or another other estimator to assist in determining how much the bio-medical unit **10** can accomplish prior to losing power. Further note that the processing module **50** may save historic records of power utilization in the memory **52** to assist in subsequent determinations of the power level.

[0119] The method continues at step **424** where the processing module **50** compares the power level to the high threshold (e.g., a first available power level that allows for a certain level of processing). If yes, the method continues to step **426** where the processing module **50** enables the execution of H number of instructions. The processing module **50** may utilize a predetermined static value of the H instructions or a dynamic value that changes as a result of the historic records. For example, the historic records may indicate that there was an average of 20% more power capacity left over after the last ten times of instruction execution upon initialization. The processing module **50** may adjust the value of H upward such that the on-going left over power is less than 20% in order to more fully utilize the available power each time the bio-medical unit **10** has power.

[0120] The method continues at step **428** where the processing module **50** saves the state in the memory **52** upon completion of the execution of the H instructions such that the processing module **50** can start in a state in accordance with this state upon the next initialization. The method then continues at step **430** where the processing module **50** determines whether it will suspend operations based on one or more of a re-determined power level (e.g., power left after executing the instructions), a predetermined list, a task priority, a task state, a priority indicator, a command, a message, and/or a functional module input. If not, the method repeats at step **422**. If yes, the method branches to step **440** where the processing module **50** suspends operations of the bio-medical unit.

[0121] If, at step **424**, the power level is not greater than the high threshold, the method continues at step **432** where the processing module **50** determines whether the power level compares favorably to a low threshold. If not, the method continues a step **440** where the processing module **50** suspends operations of the bio-medical unit.

[0122] If the comparison at step **432** was favorable, the method continues at step **434** where the processing module **50** executes L instructions. The processing module **50** may utilize a predetermined static value of the L instructions or a dynamic value that changes as a result of the historic records as discussed previously. For example, the historic records may indicate that there was an average of 10% more power capacity left over after the last ten times of instruction execution upon initialization. The processing module **50** may adjust the value of L downward such that the on-going left over power is less than 10% in order to more fully utilize the available power each time the bio-medical unit **10** has power.

[0123] The method continues at step **436** where the processing module **50** saves the state in the memory **52** upon completion of the execution of the L instructions such that the processing module **50** can start in a state in accordance with this state upon the next initialization. The method then continues at step **438** where the processing module **50** determines whether it will suspend operations based on one or more of a re-determined power level (e.g., power left after executing the instructions), a predetermined list, a task priority, a task state, a priority indicator, a command, a message, and/or a functional module input. If yes, the method branches to step **440**. If not, the method repeats at step **422**.

[0124] FIG. **23** is a flowchart illustrating MMW communications within a MRI sequence where the processing module **50** determines MMW communications in accordance with an MRI sequence. The method begins at step **442** where the processing module **50** determines whether the MRI is active based on receiving MRI EM signals. At step **444**, the method branches to step **446** or step **448**. When the MRI is active, the method continues at step **446** where the processing module **50** performs MMW communications as previously discussed.

[0125] The method continues at step **448** where the processing module **50** determines the MRI sequence based on received MRI EM signals (e.g., gradient pulses and/or MRI RF pulses as shown in one or more of the preceding figures). The method continues at step **450** where the processing module **50** determines whether it is time to perform receive MMW communication in accordance with the MRI sequence. For example, the MMW transceiver **138** may receive MMW inbound signals **148** between any of the MRI sequence steps. As another example, the MMW transceiver **138** may receive MMW inbound signals **148** between specific predetermined steps of the MRI sequence.

[0126] At step **452** the method branches back to step **450** or to step **454**. When it is time to receive, the method continues at step **454** where the processing module **50** coordinates the MMW transceiver **138** receiving the MMW inbound signals, which may include one or more of a status request, a records request, a sensor data request, a processed data request, a position request, a command, and/or a request for MRI echo signal data. The method then continues at step **456** where the processing module **50** determines whether there is at least one message pending to transmit (e.g., in a transmit queue). At step **458** the method branches back to step **442** or to step **460**.

[0127] At step **460**, the processing module **50** determines when it is time to transmit a MMW communication in accordance with the MRI sequence. For example, the MMW transceiver **138** may transmit MMW outbound signals **150** between any of the MRI sequence steps. As another example, the MMW transceiver **138** may transmit MMW outbound signals **150** between specific predetermined steps of the MRI sequence.

[0128] At step **462**, the method branches to back step **456** or to step **464**. The method continues at step **464** where the processing module **50** coordinates the MMW transceiver **138** transmitting the MMW outbound signals **150**, which may include one or more of a status request response, a records request response, a sensor data request response, a processed data request response, a position request response, a command response, and/or a request for MRI echo signal data response. The method then branches back to step **442**.

[0129] FIG. **24** is a flowchart illustrating the processing of MRI signals where the processing module **50** of the bio-medical unit **10** may assist the MRI in the reception and processing of MRI EM signals **146**. The method begins at step **466** where the processing module **50** determines if the MRI is active based on receiving MRI EM signals **146**. The method branches back to step **466** when the processing module **50** determines that the MRI is not active. For example, the MRI sequence may not start until the processing module **50** communicates to the MRI unit that it is available to assist. The method continues to step **470** when the processing module **50** determines that the MRI is active.

[0130] At step **470**, the processing module **50** determines the MRI sequence based on received MRI EM signals **146** (e.g., gradient pulses and/or MRI RF pulses). At step **472**, the processing module receives EM signals **146** and/or MMW communication **532** in accordance with the MRI sequence and decodes a message. For example, the MMW transceiver **138** may receive MMW inbound signals **148** between any of the MRI sequence steps. As another example, the MMW transceiver **138** may receive MMW inbound signals **148** between specific predetermined steps of the MRI sequence. In yet another example, the processing module **50** may receive EM signals **146** at any point of the MRI sequence such that the EM signals **146** contain a message for the processing module **50**.

[0131] At step **474**, the processing module **50** determines whether to assist in the MRI sequence based in part on the decoded message. The determination may be based on a comparison of the assist request to the capabilities of the bio-medical unit **10**. At step **476**, the method branches to step **480** when the processing module **50** determines to assist in the MRI sequence. The method continues at step **478** where the processing module **50** performs other instructions contained in the message and the method ends.

[0132] At step **480**, the processing module **50** begins the assist steps by receiving echo signals **530** during the MRI sequence. Note the echo signals **530** may comprise EM RF signals across a wide frequency band as reflected off of tissue during the MRI sequence. At step **482**, the processing module **50** processes the received echo signals **530** to produce processed echo signals. Note that this may be a portion of the overall processing required to lead to the desired MRI imaging.

[0133] At step **484**, the processing module **50** determines the assist type based on the decoded message from the MRI unit. The assist type may be at least passive or active where the passive type collects echo signal **530** information and sends it to the MRI unit via MMW outbound signals **150** and the active type collects echo signal information and re-generates a form of the echo signals **530** and sends the re-generated echo signals to the MRI unit via outbound modulated EM signals (e.g., the MRI unit interprets the re-generated echo signals as echo signals to improve the overall system gain and sensitivity).

[0134] The method branches to step **494** when the processing module **50** determines the assist type to be active. The method continues to step **486** when the processing module **50** determines the assist type to be passive. At step **486**, the processing module **50** creates an echo message based on the processed echo signals where the echo message contains information about the echo signals **530**.

[0135] At step **488**, the processing module **50** determines when it is time to transmit the echo message encoded as MMW outbound signals **150** via MMW communication in accordance with the MRI sequence. For example, the MMW transceiver **138** may transmit MMW outbound signals **150**

between any of the MRI sequence steps. In another example, the MMW transceiver **138** may transmit MMW outbound signals **150** between specific predetermined steps of the MRI sequence.

[0136] At step **490**, the method branches back to step **488** when the processing module **50** determines that it is not time to transmit the echo message. At step **490**, the method continues to step **492** where the processing module **50** transmits the echo message encoded as MMW outbound signals **150**.

[0137] At step **494**, the processing module **50** creates echo signals based on the processed echo signals. At step **496**, the processing module **50** determines when it is time to transmit the echo signals as outbound modulated EM signals **180** in accordance with the MRI sequence. At step **498**, the method branches back to step **496** when the processing module **50** determines that it is not time to transmit the echo signals. At step **498**, the method continues to step **500** where the processing module **50** transmits the echo signals encoded as outbound modulated EM signals **180**. Note that the transmitted echo signals emulate the received echo signals **530** with improvements to overcome low MRI power levels and/or low MRI receiver sensitivity.

[0138] FIG. **25** is a flowchart illustrating communication utilizing MRI signals where the processing module **50** determines MMW signaling in accordance with an MRI sequence. The method begins at step **502** where the processing module **50** determines if the MRI is active based on receiving MRI EM signals **146**. At step **504**, the method branches to step **508** when the processing module **50** determines that the MRI is active. At step **504**, the method continues to step **506** when the processing module **50** determines that the MRI is not active. At step **506**, the processing module **50** queues pending transmit messages. The method branches to step **502**.

[0139] At step **508**, the processing module **50** determines the MRI sequence based on received MRI EM signals **146** (e.g., gradient pulses and/or MRI RF pulses). At step **510**, the processing module **50** determines when it is time to perform receive communication in accordance with the MRI sequence. For example, the EM transceiver **174** may receive inbound modulated EM signals **146** containing message information from any of the MRI sequence steps.

[0140] At step **512**, the method branches back to step **510** when the processing module **50** determines that it is not time to perform receive communication. At step **512**, the method continues to step **514** where the processing module **50** directs the EM transceiver **174** to receive the inbound modulated EM signals. The processing module **50** may decode messages from the inbound modulated EM signals **146** such that the messages include one or more of a echo signal collection assist request, a status request, a records request, a sensor data request, a processed data request, a position request, a command, and/or a request for MRI echo signal data. Note that the message may be decoded from the inbound modulated EM signals **146** in one or more ways including detection of the ordering of the magnetic gradient pulses, counting the number of gradient pulses, the slice pulse orderings, detecting small differences in the timing of the pulses, and/or demodulation of the MRI RF pulse.

[0141] At step **516** the processing module **50** determines if there is at least one message pending to transmit (e.g., in a transmit queue). At step **518**, the method branches back to step **502** when the processing module **50** determines that there is not at least one message pending to transmit. At step **518**, the method continues to step **520** where the processing module **50** determines when it is time to perform transmit communication in accordance with the MRI sequence. For example, the EM transceiver **174** may transmit outbound modulated EM signals **180** between any of the MRI sequence steps. In another example, the EM transceiver **174** may transmit the outbound modulated EM signals **180** between specific predetermined steps of the MRI sequence. In yet another example, the EM transceiver **174** may transmit the outbound modulated EM signals **180** in parallel with specific predetermined steps of the MRI sequence, but may utilize a different set of frequencies unique to the EM transceiver **174**.

[0142] At step **522**, the method branches back to step **520** when the processing module **50** determines that it is not time to perform transmit communication. At step **522**, the method continues to step **524** where the processing module **50** directs the EM transceiver **174** to prepare the outbound modulated EM signals **180** based on the at least one message pending to transmit. The processing module **50** may encode messages into the outbound modulated EM signals **180** such that the messages include one or more of a status request response, a records request response, a sensor data request response, a processed data request response, a position request response, a command response, and/or a request for MRI echo signal data response. The method branches back to step **502**.

[0143] FIG. **26** is a schematic block diagram of an embodiment of a MRI system utilizing a plurality of bio-medical units **10** to provide processed echo signal information based on received MRI echo signals **530** across the plurality of bio-medical units **10**. The MRI system includes the transmitter unit **126**, the receiver unit, **128** and a plurality of bio-medical units **10** equipped to communicate with each other and with the transmitter unit **126** and the receiver unit **128**. The bio-medical unit **10** includes an echo receiver **528**, the MMW transceiver **138**, the power harvesting module **46**, the memory **52**, and the processing module **50**. The echo receiver **528** receives echo signals **530** from an object **526** in accordance with the MRI sequence.

[0144] The processing module **50** produces processed echo signals based on the echo signals **530** received by the echo receiver **528**. The processing module **50** may determine to share the processed echo signal information with the plurality of bio-medical units **10** based in part on instructions received by the MMW transceiver **138** from the transmitter unit **126**. The processing module **50** may produce further processed echo signals based on the shared processed echo signals of the plurality of bio-medical units **10**. The processing module **50** may encode and send the further processed echo signal information as MMW outbound signals **150** to the receiver unit **128**. The method of processing and communicating the processed echo signals is discussed in greater detail with reference to FIG. **27**.

[0145] FIG. **27** is another flowchart illustrating the processing of MRI signals where the processing module **50** of the bio-medical unit **10** may communicate with other bio-medical units **10** to assist the MRI in the reception and processing of MRI EM signals **146**. The method begins at step **534** where the processing module **50** determines if the MRI is active based on receiving MRI EM signals **146** via the EM transceiver **174**. At step **536**, the method branches back to step **534** when the processing module **50** determines that the MRI is not active. For example, the MRI sequence may not start until the processing module **50** communicates to the MRI that it is available to assist. At step **536**, the method continues to step **538** when the processing module **50** determines that the MRI is active.

[0146] At step **538**, the processing module **50** determines the MRI sequence based on received MRI EM signals **146** (e.g., gradient pulses and/or MRI RF pulses). At step **540**, the processing module **50** receives EM signals **146** and/or MMW inbound signals **148** in accordance with the MRI sequence and decodes a message for the processing module **50**.

[0147] At step **542**, the processing module **50** determines whether to receive echo signals **530** produced by the MRI sequence based in part on the decoded message. The determination may be based on a comparison of the MRI echo signal data request to the capabilities of the bio-medical unit **10**. At step **544**, the method branches to step **548** when the processing module **50** determines to receive echo signals **530**. At step **544**, the method continues to step **546** when the processing module **50** determines to not receive echo signals **530**. At step **546**, the processing module **50** performs other instructions contained in the message. The method ends.

[0148] At step **548**, the processing module **50** establishes communications with other bio-medical units **10**. The processing module **50** may establish communications with other bio-medical units **10** utilizing MMW communication **532**. The processing module **50** may send a receive echo signals command to the other bio-medical units **10** such that the plurality of bio-medical units **10** will all receive subsequent echo signals **530**.

[0149] At step **550**, the processing module **50** receives echo signals **530** via the echo receiver **528**. Note the echo signals **530** may comprise EM RF signals across a wide frequency band as reflected off of the object **526** during the MRI sequence. At step **552**, the processing module **50** processes the received echo signals **530** to produce processed echo signals. Note that this may be a portion of the overall processing required to lead to the desired MRI imaging.

[0150] At step **554**, the processing module **50** sends the processed echo signals to other bio-medical units via MMW communication **532**. Note that bio-medical units **10** receiving the processed echo signals may re-send the processed echo signals to still other bio-medical units **10**. Further note that the one of more of the plurality of bio-medical units **10** may further process the processed echo signals to produce processed echo signals.

[0151] At step **556**, the processing module **50** may receive the processed echo signals from other bio-medical units via MMW communication **532**. Note that the process echo signals may include functional module **54** information including sensor data and location. The processing module **50** may process the processed echo signals to produce processed echo signals (e.g., refined signals, an aggregation, a composite, a result, etc.).

[0152] At step **558**, the processing module **50** creates an echo message based on the processed echo signals where the echo message contains information about the echo signals **530**. At step **560**, the processing module **50** determines when it is time to transmit the echo message encoded as MMW outbound signals **150** via MMW communication **532** in accordance with the MRI sequence. For example, the MMW transceiver **138** may transmit MMW outbound signals **150** between any of the MRI sequence steps. For example, the MMW transceiver **138** may transmit MMW outbound signals **150** between specific predetermined steps of the MRI sequence.

[0153] At step **562**, the method branches back to step **560** when the processing module **50** determines it is not time to transmit the echo message. At step **566**, the method continues to step **564** when the processing module **50** determines it is time to transmit the echo message. At step **564**, the processing module **50** transmits the echo message encoded as MMW outbound signals.

[0154] FIG. **28** is a flowchart illustrating the communication of records where the processing module **50** of the bio-medical unit **10** determines to provide medical records. The method begins at step **566** where the processing module **50** determines if receiving MMW communication is allowed. The determination may be based on one or more of a timer, a command, available power, a priority indicator, and/or interference indicator. For example, the MMW transceiver **138** may receive MMW inbound signals **148** for a 500 millisecond window every 3 minutes.

[0155] At step **568**, the method branches back to step **566** when the processing module **50** determines that receiving MMW communication is not allowed. At step **568**, the method continues to step **570** where the processing module **50** directs the MMW transceiver **138** to receive MMW inbound signals **148**. The processing module **50** may decode messages from the MMW inbound signals **148** such that the decoded message include one or more of a status request, a records request, a sensor data request, a processed data request, a position request, a command, and/or a request for MRI echo signal data.

[0156] At step **572**, the processing module **50** determines whether to provide records in response to the records request based in part on the decoded message. The determination may be based on a comparison of the records request to the capabilities of the bio-medical unit **10**. Note that records may include patient history, medications, alerts, allergies, personal information, contact information, age, weight, test results, etc.

[0157] At step **576**, the method branches to step **578** when the processing module **50** determines to provide records. At step **576**, the method continues to step **576** when the processing module **50** determines to not provide records. At step **576**, the processing module **50** performs other instructions contained in the message. The method ends.

[0158] At step **578**, the processing module **50** determines when it is time to transmit. The determination may be based on a timer, a command, available power, a priority indicator, a timeslot, and/or interference indicator. At step **580**, the method branches back to step **578** when the processing module **50** determines it is not time to transmit. At step **580**, the method continues to step **582** when the processing module **50** determines it is time to transmit.

[0159] At step **582**, the processing module **50** determines the format to provide records. The format determination may be based on one or more of a memory lookup, a command, available power, the type of records requested, an access ID of the requester, a priority indicator, a level of detail indicator, and/or a freshness indicator. Note that the format may include records format as stored in the bio-medical unit memory (e.g., all or a portion of the records) or a uniform resource locator (URL) to link to another memory in one or more of the service provider's computer, the database, and/or the server.

[0160] At step **584**, the method branches to step **588** when the processing module **50** determines the format to provide records is the URL format. At step **584**, the method continues to step **586** where the processing module **50** prepares the records format response message based on records information retrieved from the bio-medical unit memory **52**. The method branches to step **590**.

[0161] At step **588**, the processing module prepares the URL format response message based on retrieving the URL from the bio-medical unit memory **52**. At step **590**, the processing module **50** transmits the response message encoded as MMW outbound signals **150**. For example, the bio-medical unit **10** transmits the response message via a second wireless communications medium including one or more of infrared signals, ultrasonic signals, visible light signals, audible sound signals, and/or EM signals via one or more of the functional modules.

[0162] FIG. **29** is a flowchart illustrating the coordination of bio-medical unit task execution where the processing module **50** determines and executes tasks with at least one other bio-medical unit **10**. The method begins at step **592** where the processing module **50** determines if communication is allowed. The determination may be based on one or more of a timer, a command, available power, a priority indicator, an MRI sequence, and/or interference indicator.

[0163] At step **594**, the method branches back to step **592** when the processing module **50** determines that communication is not allowed. At step **594**, the method continues to step **596** when the processing module **50** determines that communication is allowed. At step **596**, the processing module **50** directs the communication module **48** to communicate with a plurality of bio-medical units **10** utilizing RF and/or MMW inbound and/or outbound signals. The processing module **50** may decode messages from the RF and/or MMW inbound and/or outbound signals inbound signals. At step **598**, the processing module **50** determines if communications with the plurality of bio-medical units **10** is successful based in part on the decoded messages.

[0164] At step **600**, the method branches back to step **592** when the processing module determines that communications with the plurality of bio-medical units **10** is not successful. Note that forming a network with the other bio-medical units **10** may be required to enable joint actions. At step **600**, the method continues to step **602** when the processing module **50** determines that communications with the plurality of bio-medical units **10** is successful.

[0165] At step **602**, the processing module **50** determines the task and task requirements. The task determination may be based on one or more of a command from a parent bio-medical unit **10**, external communications, a preprogrammed list, and/or in response to sensor data. The task requirements determination may be based on one or more of the task, a command from a parent bio-medical unit **10**, external communications, a preprogrammed list, and/or in response to sensor data. Note that the task may include actions including one or more of drilling, moving, sawing, jumping, spreading, sensing, lighting, pinging, testing, and/or administering medication.

[0166] At step **604**, the processing module **50** determines the control mode based on one or more of a command from a parent bio-medical unit **10**, external communications, a preprogrammed list, and/or in response to sensor data. Note that the control mode may include autonomous, parent (bio-medical unit), server, and/or peer.

[0167] At step **606**, the processing module **50** determines if task execution criteria are met based on sensor data, communication with other bio-medical units **10**, a command, a status indicator, a safety indicator, a stop indicator, and/or location information. Note that the task execution criteria may include one or more of safety checks, position information of the bio-medical unit **10**, position information of other bio-medical units **10**, and/or sensor data thresholds.

[0168] At step **608**, the method branches back to step **606** when the processing module **50** determines that the task execution criteria are not met. At step **608**, the method continues to step **610** when the processing module **50** determines that the task execution criteria are met. At step **610**, the processing module **50** executes a task element. A task element may include a portion or step of the overall task. For example, move one centimeter of a task to move three centimeters.

[0169] At step **612**, the processing module **50** determines if task exit criteria are met based on a task element checklist status, sensor data, communication with other bio-medical units **10**, a command, a status indicator, a safety indicator, a stop indicator, and/or location information. Note that the task exit criteria define successful completion of the task.

[0170] At step **614**, the method branches back to step **592** when the processing module **50** determines that the task exit criteria are met. In other words, the plurality of bio-medical units **10** is done with the current task and is ready for the next task. At step **614**, the method continues to step **616** when the processing module **50** determines that the task exit criteria are not met.

[0171] At step **616**, the processing module **50** directs the communication module **48** to communicate with the plurality of bio-medical units **10** utilizing RF and/or MMW inbound and/or outbound. The processing module **50** may decode messages from the RF and/or MMW inbound and/or outbound signals inbound signals. Note that the messages may include information in regards to task modifications (e.g., course corrections). At step **618**, the processing module **50** determines if communications with the plurality of bio-medical units **10** is successful based in part on the decoded messages.

[0172] At step **620**, the method branches back to step **592** when the processing module determines that communications with the plurality of bio-medical units is not successful (e.g., to potentially restart). Note that maintaining the network with the other bio-medical unit may be required to enable joint actions. At step **620**, the method continues to step **622** when the processing module determines that communications with the plurality of bio-medical units is successful.

[0173] At step **622**, the processing module **50** determines task modifications. The task modifications may be based on one or more of a command from a parent bio-medical unit **10**, and/or external communications. The task modifications determination may be based on one or more of the task, a command from a parent bio-medical unit **10**, external communications, a preprogrammed list, and/or in response to sensor data. The method branches back to step **606** to attempt to complete the current task.

[0174] As may be used herein, the terms "substantially" and "approximately" provides an industry-accepted tolerance for its corresponding term and/or relativity between items. Such an industry-accepted tolerance ranges from less than one percent to fifty percent and corresponds to, but is not limited to, component values, integrated circuit process variations, temperature variations, rise and fall times, and/or thermal noise. Such relativity between items ranges from a difference of a few percent to magnitude differences. As may also be used herein, the term(s) "coupled to" and/or "coupling" and/or includes direct coupling between items and/or indirect coupling between items via an intervening item (e.g., an item includes, but is not limited to, a component, an element, a circuit, and/or a module) where, for indirect coupling, the intervening item does not modify the information of a signal but may adjust its current level, voltage level, and/or power level. As may further be used herein, inferred coupling (i.e., where one element is coupled to another element by inference) includes direct and indirect coupling between two items in the same manner as "coupled to". As may even further be used herein, the term "operable to" indicates that an item includes one or more of power connections, input(s), output(s), etc., to perform one or more its corresponding functions and may further include inferred coupling to one or more other items. As may still further be used herein, the term "associated with", includes direct and/or indirect coupling of separate items and/or one item being embedded within another item. As may be used herein, the term "compares favorably", indicates that a comparison between two or more items, signals, etc., provides a desired relationship. For example, when the desired relationship is that signal **1** has a greater magnitude than signal **2**, a favorable comparison may be achieved when the magnitude of signal **1** is greater than that of signal **2** or when the magnitude of signal **2** is less than that of signal **1**.

[0175] The present invention has also been described above with the aid of method steps illustrating the performance of specified functions and relationships thereof. The boundaries and sequence of these functional building blocks and method steps have been arbitrarily defined herein for convenience of description. Alternate boundaries and sequences can be defined so long as the specified functions and relationships are appropriately performed. Any such alternate boundaries or sequences are thus within the scope and spirit of the claimed invention.

[0176] The present invention has been described above with the aid of functional building blocks illustrating the performance of certain significant functions. The boundaries of these functional building blocks have been arbitrarily defined for convenience of description. Alternate boundaries could be defined as long as the certain significant functions are appropriately performed. Similarly, flow diagram blocks may also have been arbitrarily defined herein to illustrate certain significant functionality. To the extent used, the flow diagram block boundaries and sequence could have been defined otherwise and still perform the certain significant functionality. Such alternate definitions of both functional building blocks and flow diagram blocks and sequences are thus within the scope and spirit of the claimed invention. One of average skill in the art will also recognize that the functional building blocks, and other illustrative blocks, modules and components herein, can be implemented as illustrated or by discrete components, application specific integrated circuits, processors executing appropriate software and the like or any combination thereof.

What is claimed is:

1. A medical system comprises:
a transmitter unit that includes:
an electromagnetic transmitter operable to transmit an electromagnetic signal in a predetermined pattern; and
a radio frequency (RF) communication transmitter operable to transmit a downstream RF communication signal;
a receiver unit that includes:
an electromagnetic receiver operable to receive the electromagnetic signal; and
a RF communication receiver operable to receive an upstream RF communication signal; and
a bio-medical unit for implanting into a body, wherein the bio-medical unit includes:
a power harvesting module operable to generate a power supply voltage from the electromagnetic signal;
a RF communication module powered by the power supply voltage and operable to:
interpret the predetermined pattern of the transmitting of the electromagnetic signal to determine null intervals;
receive the downstream RF communication signal during at least one of the null intervals;
convert the downstream RF communication signal into downstream information;
convert upstream information into the upstream RF communication signal; and
transmit the upstream RF communication signal during one or more of the null intervals.

2. The medical system of claim **1**, wherein the electromagnetic signal comprises:
a magnetic resonance imaging (MRI) signal having at least one of a radio frequency (RF) component and a gradient magnetic field component.

3. The medical system of claim **1**, wherein the bio-medical unit further comprises:
a processing module operable to:
convert the downstream information into at least one of downstream data and a downstream instruction; and
convert at least one of an upstream instruction and upstream data into the upstream information; and
a function module operable to:
generate the at least one of an upstream instruction and the upstream data; and
perform a bio-medical function based on the at least one of downstream data and the downstream instruction.

4. The medical system of claim **3** further comprises:
the processing module further operable to:
interpret the downstream instruction to determine whether the functional module is to be enabled to receive an echo signal, wherein the echo signal corresponds to the electromagnetic signal reflecting off of an internal portion of the body; and
enable the functional module when the functional module is to be enabled to receive an echo signal;
the functional module is further operable to:
receive the echo signal; and
convert the echo signal into echo information; and
the processing module is further operable to convert the echo information into the upstream information.

5. The medical system of claim **4**, wherein the processing module is further operable to perform at least one of:
passively convert the echo information into the upstream information; and
actively convert the echo information into the upstream information.

6. The medical system of claim **1**, wherein the RF communication module further functions to:
determine whether an electromagnetic function is active;
when the electromagnetic function is active, determine the predetermined pattern;

determine the at least one null interval for receiving the downstream RF communication signal;

determine whether the upstream RF communication is pending; and

when the upstream RF communication is pending, determine the one or more of the null intervals.

7. An electromagnetic based medical device comprises:

a transmitter unit that includes:

an electromagnetic transmitter operable to transmit an electromagnetic signal in a predetermined pattern; and

a radio frequency (RF) communication transmitter operable to transmit a downstream RF communication signal to one or more bio-medical units that are implanted within a body during at least one of a plurality of nulls of the predetermined pattern; and

a receiver unit that includes:

an electromagnetic receiver operable to receive the electromagnetic signal; and

a RF communication receiver operable to receive an upstream RF communication signal from the one or more bio-medical units during one or more of the plurality of nulls of the predetermined pattern.

8. The electromagnetic based medical device of claim **7**, wherein the electromagnetic signal comprises:

a magnetic resonance imaging (MRI) signal having at least one of a radio frequency (RF) component and a gradient magnetic field component.

9. The electromagnetic based medical device of claim **7** further comprises:

a processing module operable to:

convert at least one of downstream data and a downstream instruction into downstream information; and

convert upstream information into at least one of an upstream instruction and upstream data;

the transmitter unit further operable to convert the downstream information into the downstream RF communication signal; and

the receiver unit further operable to convert the upstream RF communication signal into the upstream information.

10. The electromagnetic based medical device of claim **9** further comprises:

the processing module further operable to:

interpret the upstream data to obtain an echo signal, wherein the echo signal corresponds to the electromagnetic signal reflecting off of an internal portion of the body; and

utilize the echo signal in conjunction with the received electromagnetic signal to generate an internal image of the body.

11. A bio-medical unit for implanting into a body, the bio-medical unit comprises:

a power harvesting module operable to generate a power supply voltage from the electromagnetic signal; and

a radio frequency (RF) communication module powered by the power supply voltage and operable to:

interpret the predetermined pattern of the transmitting of the electromagnetic signal to determine null intervals;

receive the downstream RF communication signal during at least one of the null intervals;

convert the downstream RF communication signal into downstream information;

convert upstream information into the upstream RF communication signal; and

transmitting the upstream RF communication signal during one or more of the null intervals.

12. The bio-medical unit of claim **11**, wherein the electromagnetic signal comprises:

a magnetic resonance imaging (MRI) signal having at least one of a radio frequency (RF) component and a gradient magnetic field component.

13. The bio-medical unit of claim **11** further comprises:

a processing module operable to:

convert the downstream information into at least one of downstream data and a downstream instruction; and

convert at least one of an upstream instruction and upstream data into the upstream information; and

a function module operable to perform at least one of:

generate the at least one of an upstream instruction and the upstream data; and

perform a bio-medical function based on the at least one of downstream data and the downstream instruction.

14. The bio-medical unit of claim **13** further comprises:

the processing module further operable to:

interpret the downstream instruction to determine whether the functional module is to be enabled to receive an echo signal, wherein the echo signal corresponds to the electromagnetic signal reflecting off of an internal portion of the body; and

enable the functional module when the functional module is to be enabled to receive an echo signal;

the functional module is further operable to:

receive the echo signal; and

convert the echo signal into echo information; and

the processing module is further operable to convert the echo information into the upstream information.

15. The bio-medical unit of claim **14**, wherein the processing module is further operable to perform at least one of:

passively convert the echo information into the upstream information; and

actively convert the echo information into the upstream information.

16. The bio-medical unit of claim **11**, wherein the RF communication module further functions to:

determine whether an electromagnetic function is active;

when the electromagnetic function is active, determine the predetermined pattern;

determine the at least one null interval for receiving the downstream RF communication signal;

determine whether the upstream RF communication is pending; and

when the upstream RF communication is pending, determine the one or more of the null intervals.

* * * * *

US 20110101789A1

(19) **United States**

(12) **Patent Application Publication**

Salter, JR. et al.

(10) **Pub. No.: US 2011/0101789 A1**

(43) **Pub. Date: May 5, 2011**

(54) **RF POWER HARVESTING CIRCUIT**

(76) Inventors: **Thomas Steven Salter, JR.**, Ellicott City, MD (US); **George M. Metze**, Millersville, MD (US); **Neil Goldsman**, Takoma Park, MD (US); **Kwangsik Choi**, College Park, MD (US); **Yves Ngu**, Essex, VT (US); **Zeynep Dilli**, Hyattsville, MD (US); **Martin Peckerar**, Silverspring, MD (US); **Li Bo**, College Park, MD (US)

(21) Appl. No.: **12/628,585**

(22) Filed: **Dec. 1, 2009**

Related U.S. Application Data

(60) Provisional application No. 61/118,808, filed on Dec. 1, 2008, provisional application No. 61/119,848, filed on Dec. 4, 2008.

Publication Classification

(51) **Int. Cl.**
H01F 38/14 (2006.01)

(52) **U.S. Cl.** **307/104**

(57) **ABSTRACT**

Provided is an RF power harvesting circuit with improved sensitivity to RF energy. The RF power harvesting device includes an inductor, a first capacitor connected to the inductor, a first MOSFET connected to a first node, and a second MOSFET connected to the first node. The inductor or the first capacitor are connected to the first node.

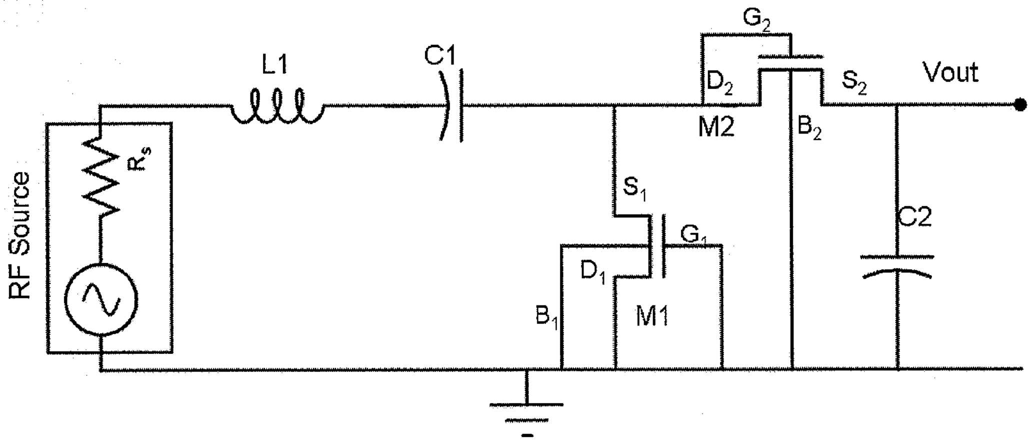

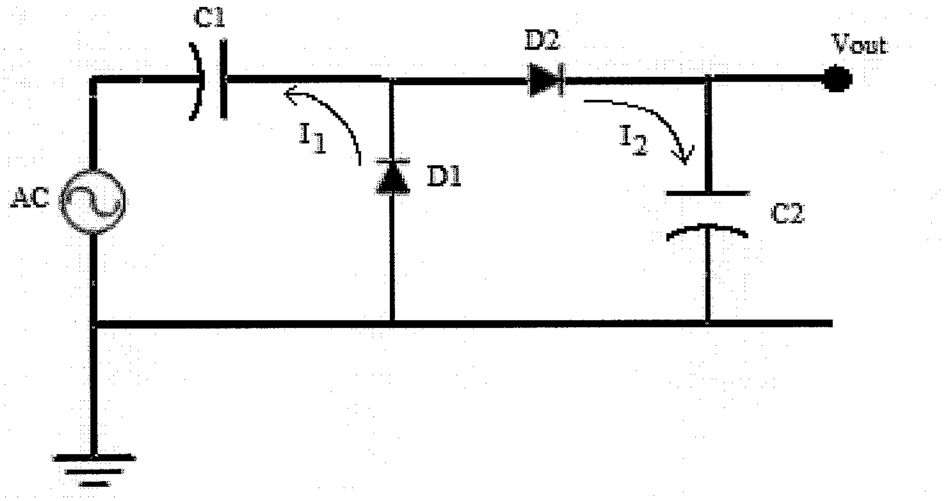
C1
D2
Vout
I_1
I_2
D1
AC
C2

FIG. 1

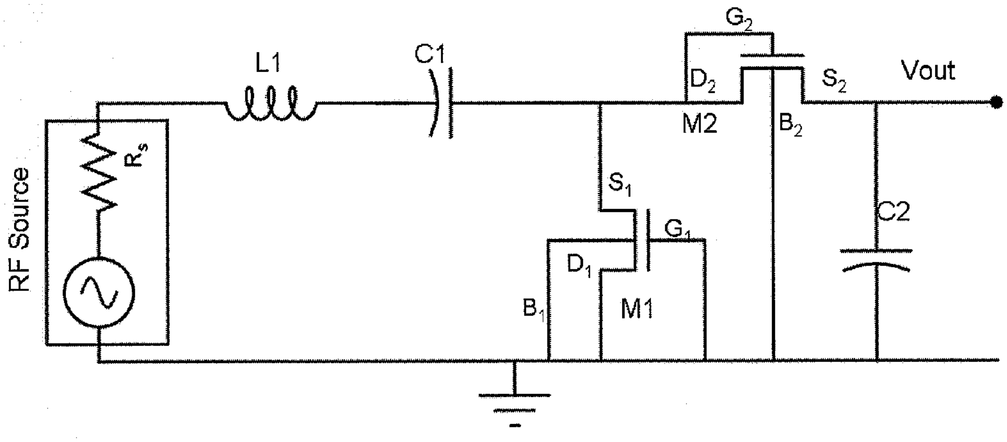
G_2
L1
C1
D_2
S_2
Vout
M2
B_2
R_s
RF Source
S_1
C2
G_1
D_1
B_1
M1

FIG. 2

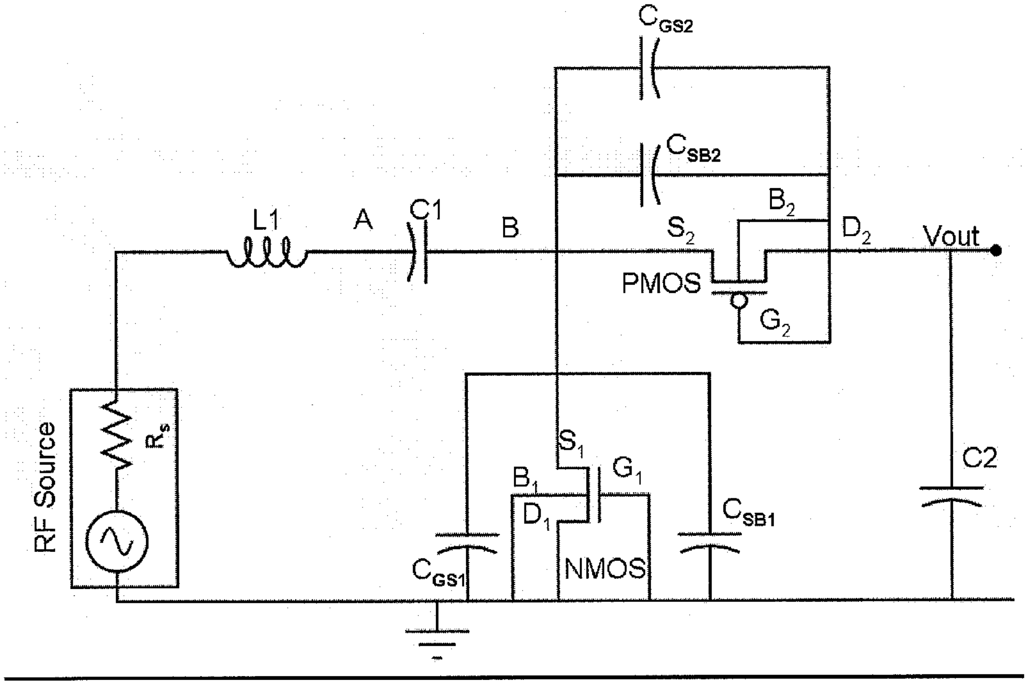
C_GS2
C_SB2
B_2
L1
A
C1
B
S_2
D_2
Vout
PMOS
G_2
RF Source
R_s
S_1
G_1
B_1
D_1
C_SB1
C2
C_GS1
NMOS

FIG. 3

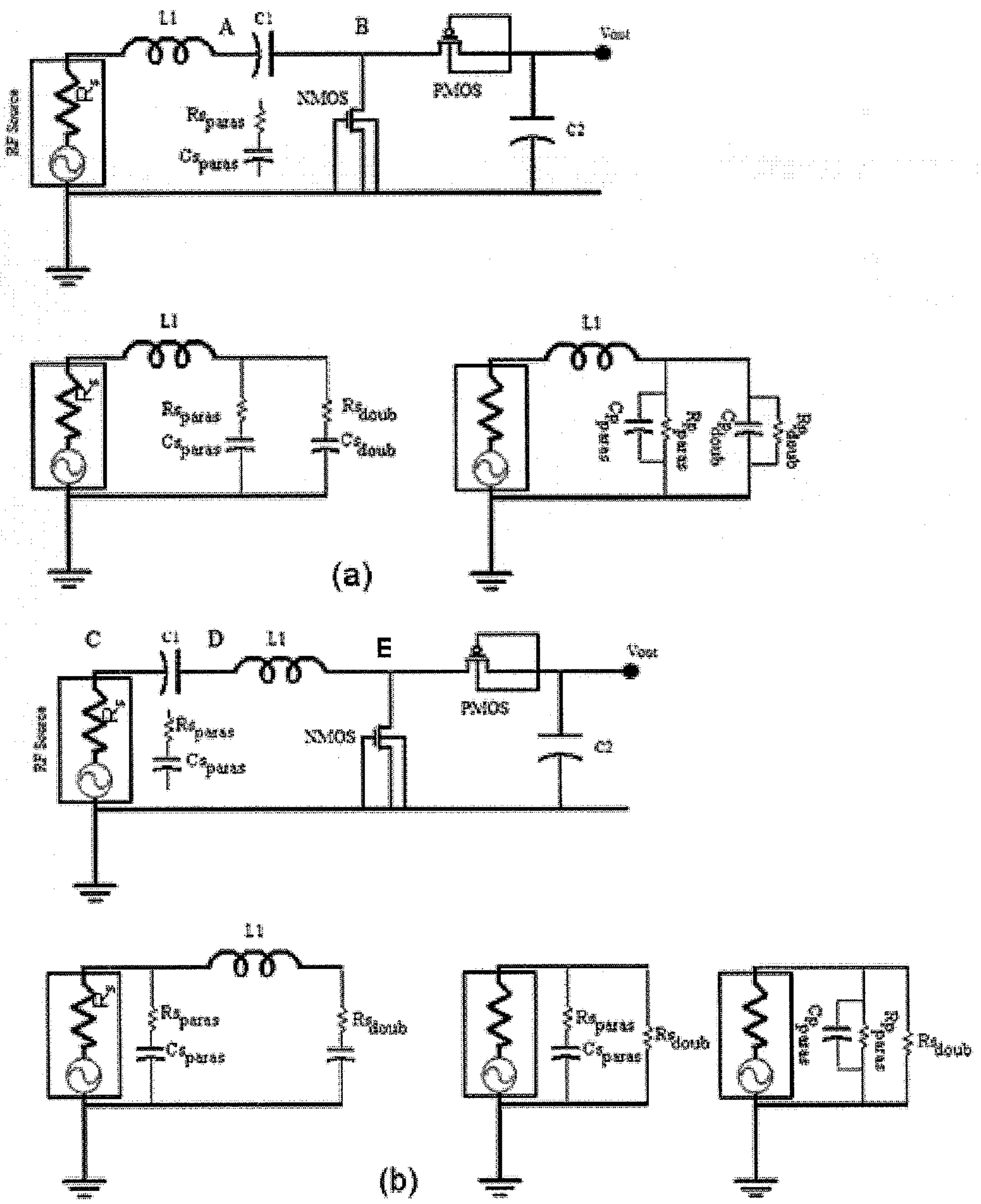
L1
A
C1
B
Vout
PMOS
NMOS
RF Source
R_s
Rs_{paras}
Cs_{paras}
C2
Rs_{doub}
Cs_{doub}
Cp_{paras}
Rp_{paras}
Cp_{doub}
Rp_{doub}
(a)
C
D
E
(b)

FIG. 4(a) and 4(b)

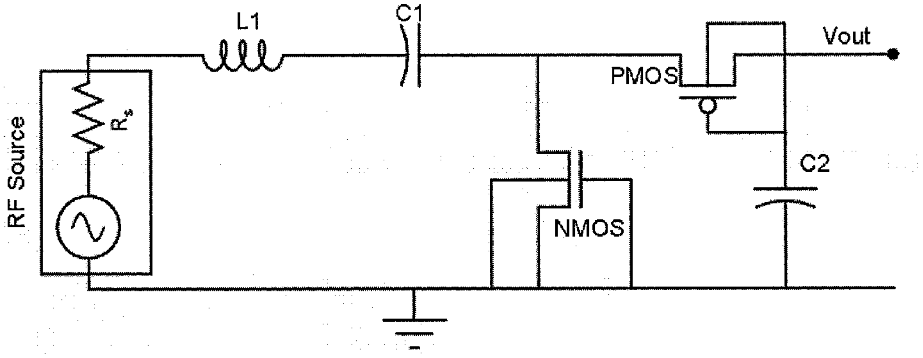
L1
C1
Vout
PMOS
RF Source
R_s
C2
NMOS

FIG. 5

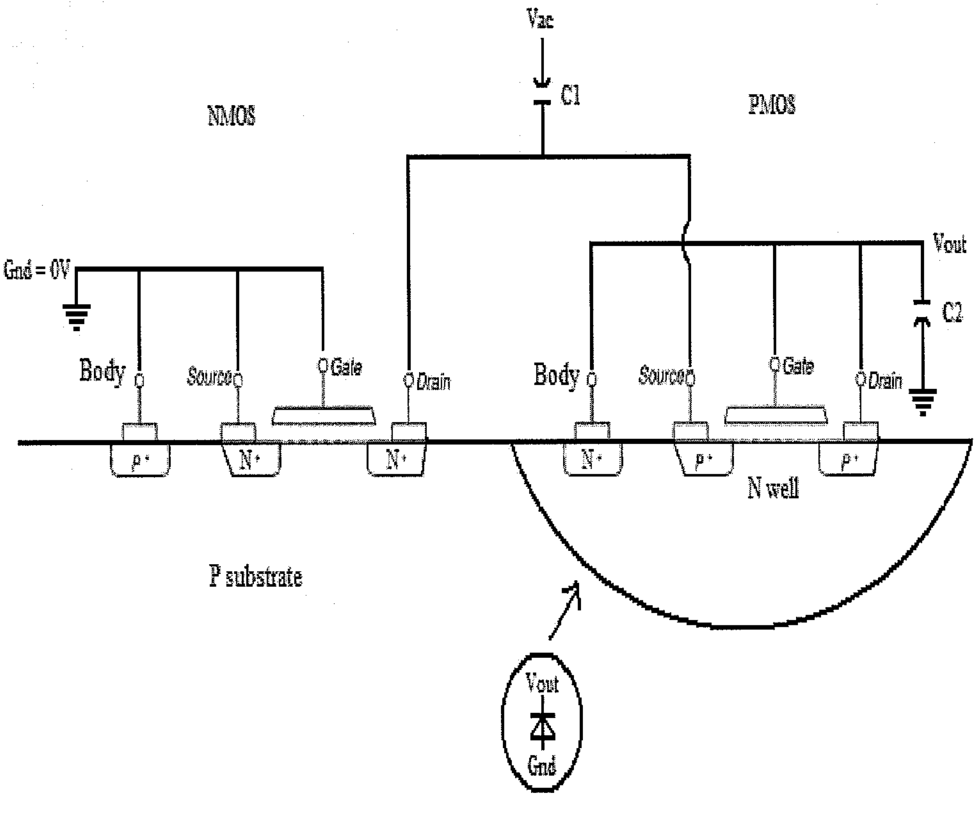
Vac
C1
NMOS
PMOS
Gnd = 0V
Vout
C2
Body
Source
Gate
Drain
Body
Source
Gate
Drain
P+
N+
N+
N+
P+
P+
N well
P substrate
Vout
Gnd

FIG. 6

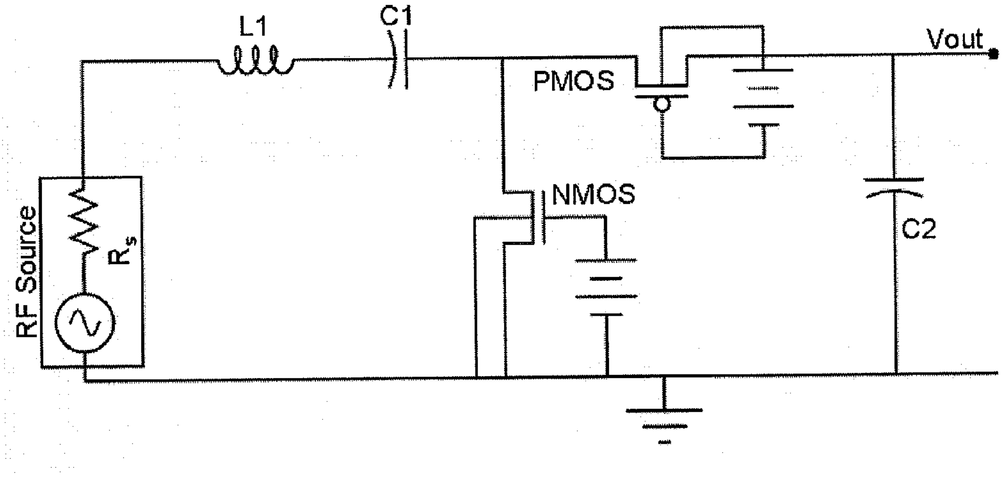
L1
C1
Vout
PMOS
NMOS
C2
RF Source
R_s

FIG. 7

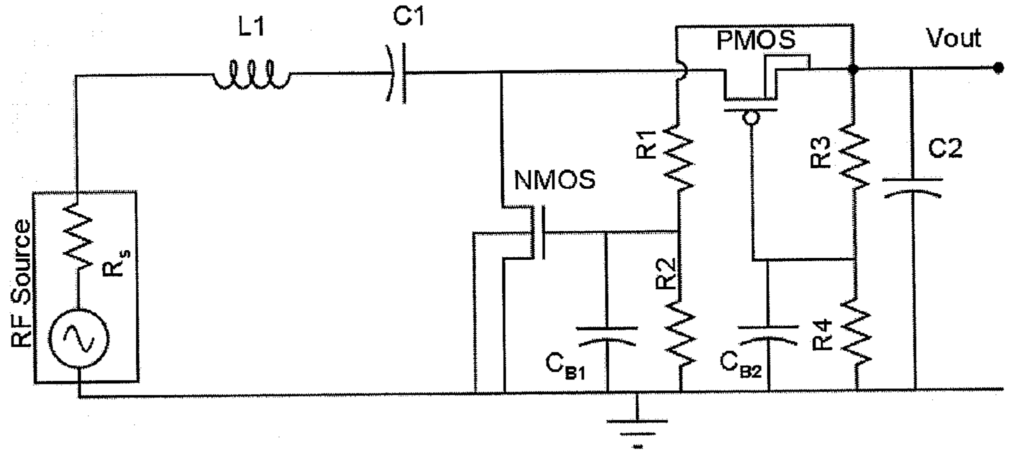
L1
C1
PMOS
Vout
R1
R3
C2
NMOS
RF Source
R_s
R2
R4
C_{B1}
C_{B2}

FIG. 8

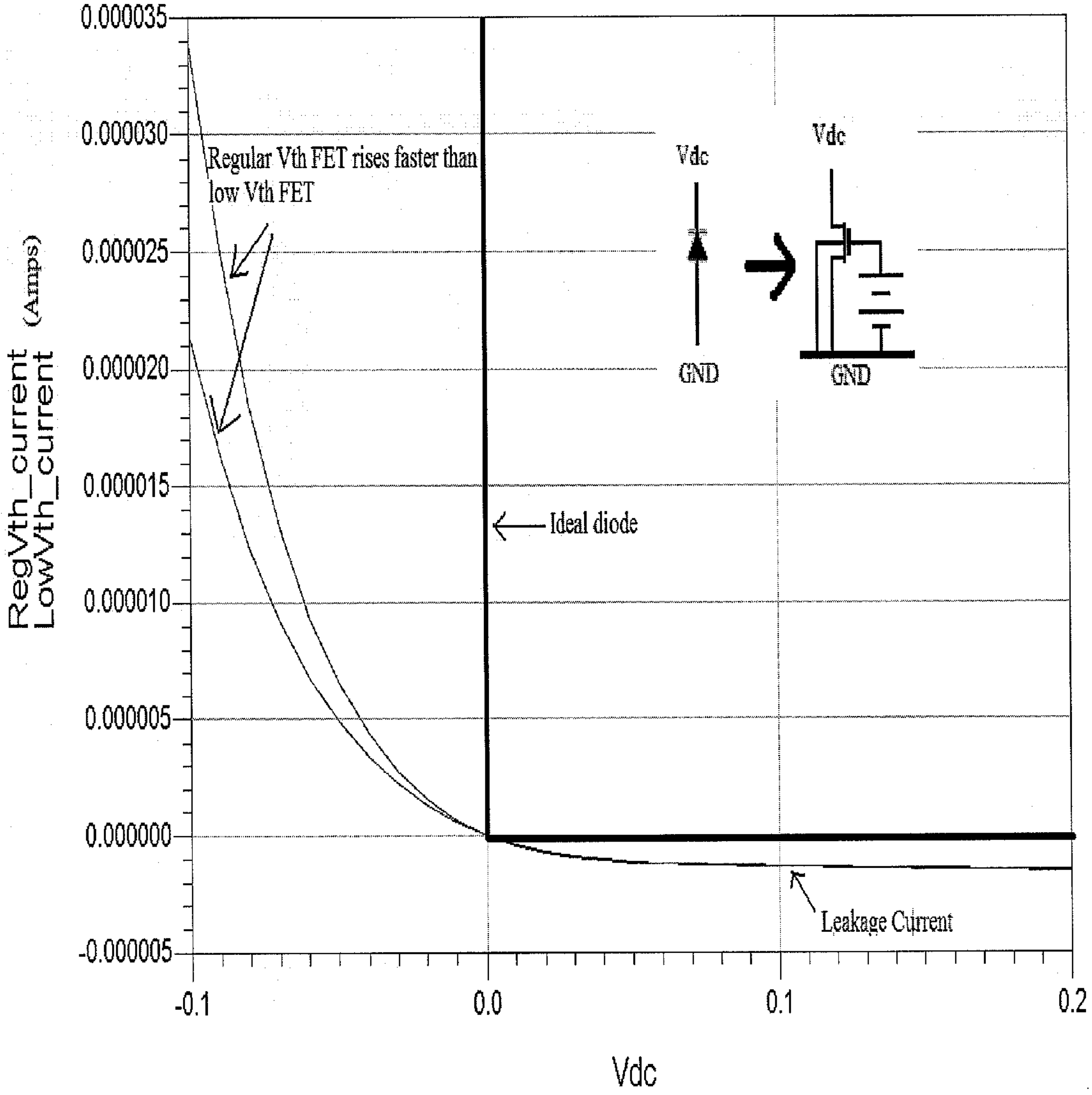
0.000035
0.000030
0.000025
0.000020
0.000015
0.000010
0.000005
0.000000
-0.000005
RegVth_current
LowVth_current
(Amps)
Regular Vth FET rises faster than low Vth FET
Vdc
GND
Vdc
GND
Ideal diode
Leakage Current
-0.1
0.0
0.1
0.2
Vdc

FIG. 9

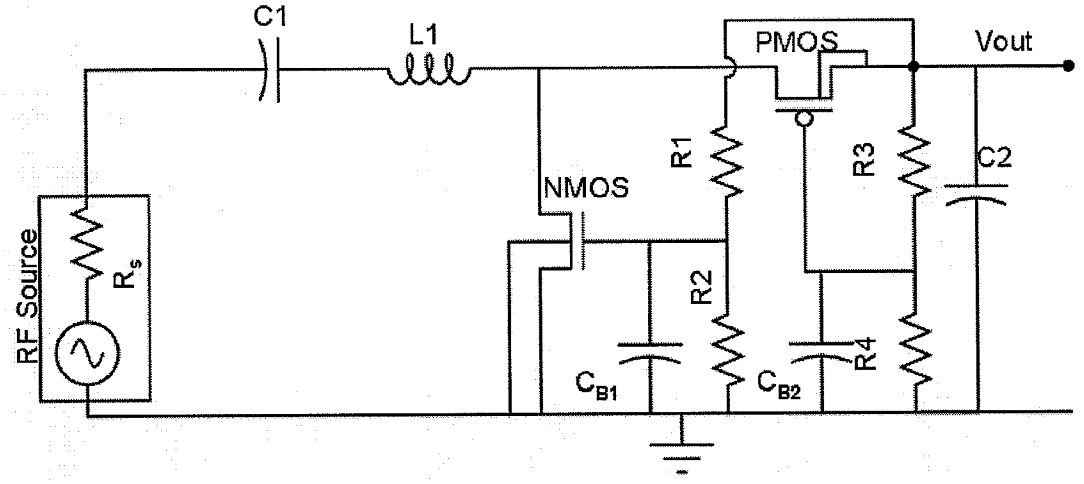
C1
L1
PMOS
Vout
R1
R3
C2
NMOS
RF Source
R_s
R2
C_{B1}
C_{B2}
R4

FIG. 10

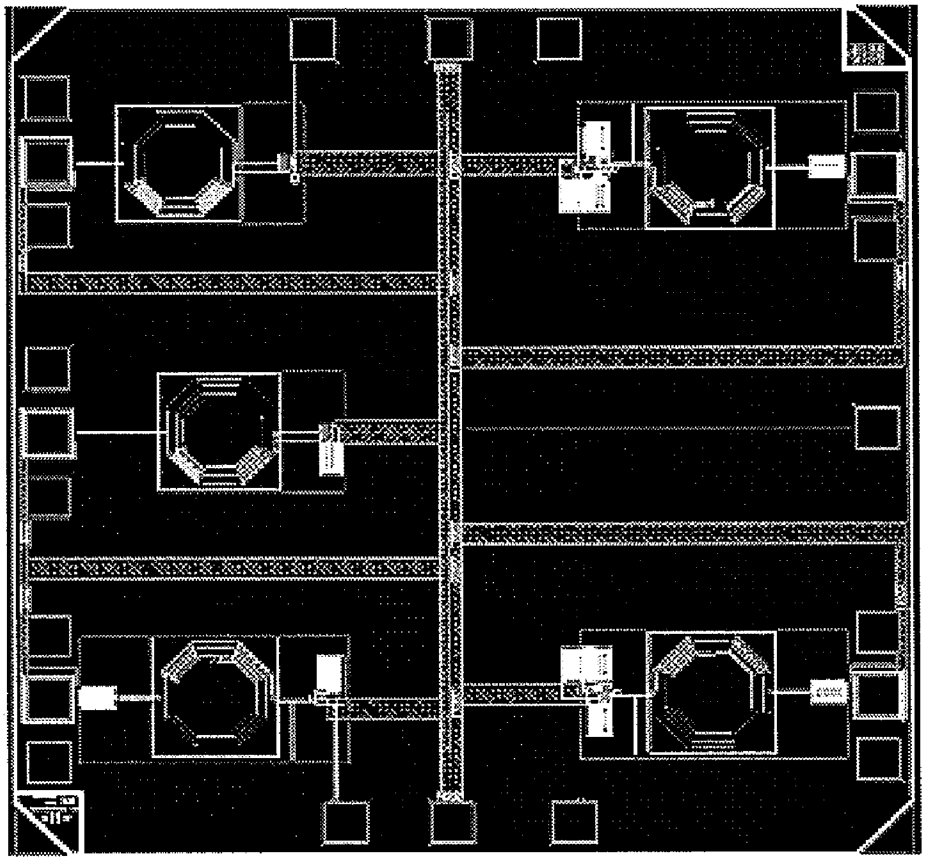

FIG. 11

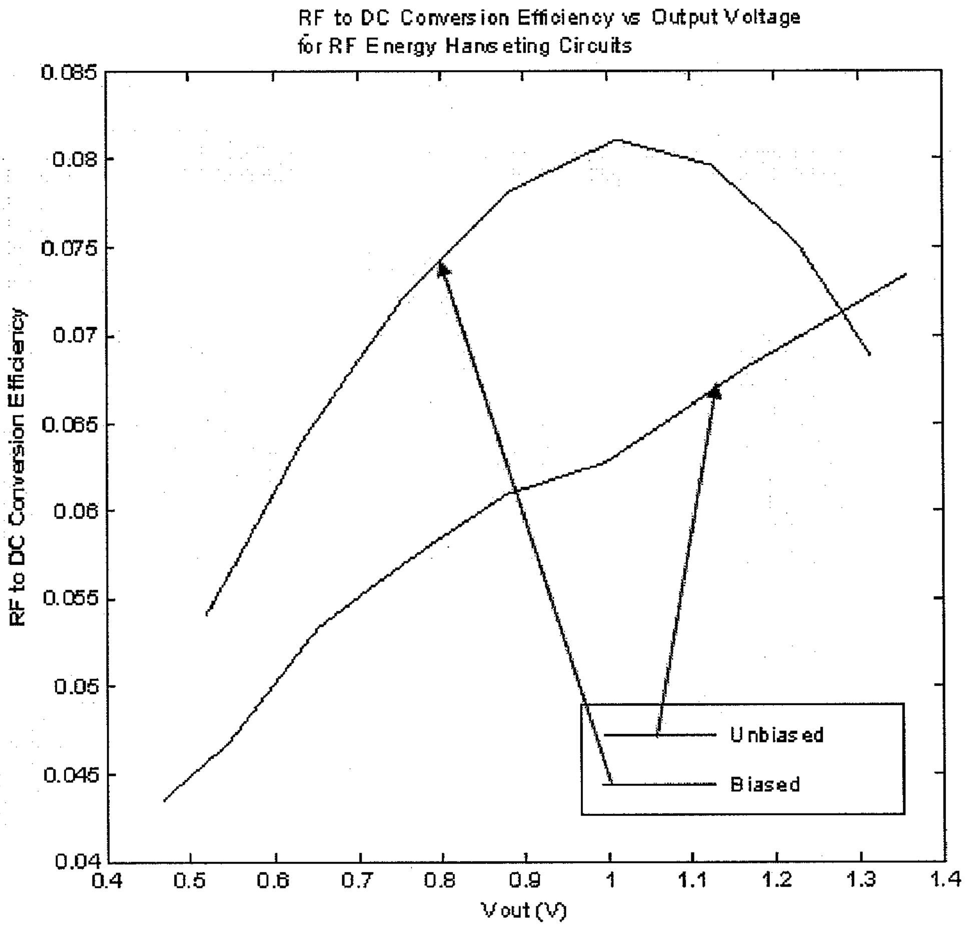
RF to DC Conversion Efficiency vs Output Voltage
for RF Energy Harvseting Circuits
RF to DC Conversion Efficiency
0.085
0.08
0.075
0.07
0.065
0.06
0.055
0.05
0.045
0.04
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
1.4
Vout (V)
Unbiased
Biased

FIG. 12

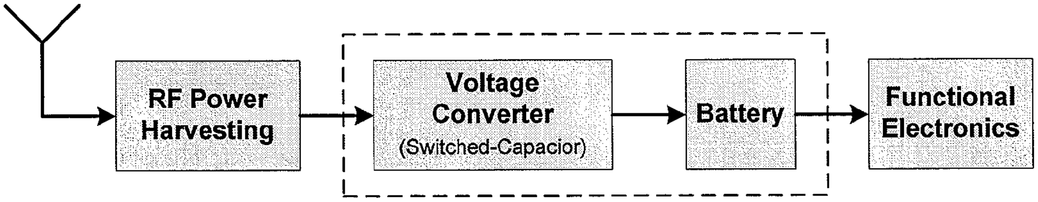
RF Power Harvesting
Voltage Converter
(Switched-Capacior)
Battery
Functional Electronics

FIG. 13

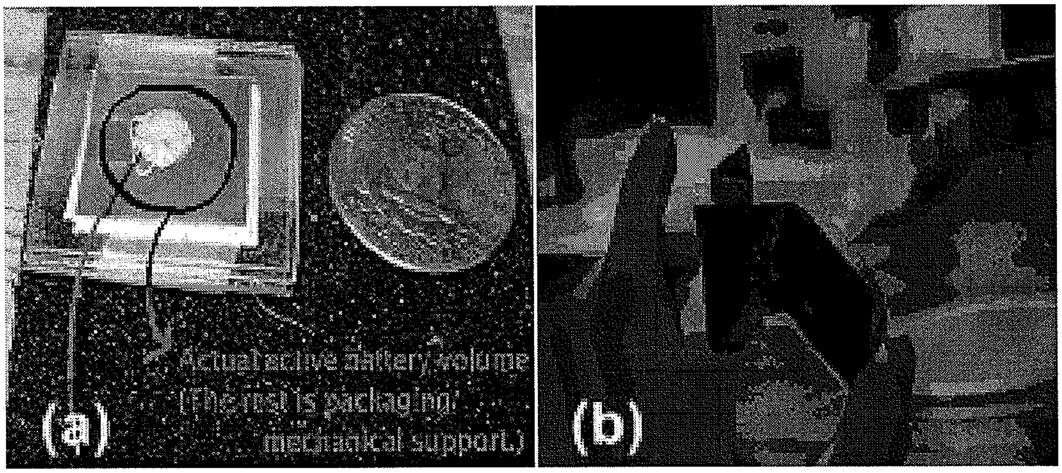
Actual active battery volume
(The rest is packaging/
mechanical support.)
(a)
(b)

FIGS. 14(a) and 14(b)

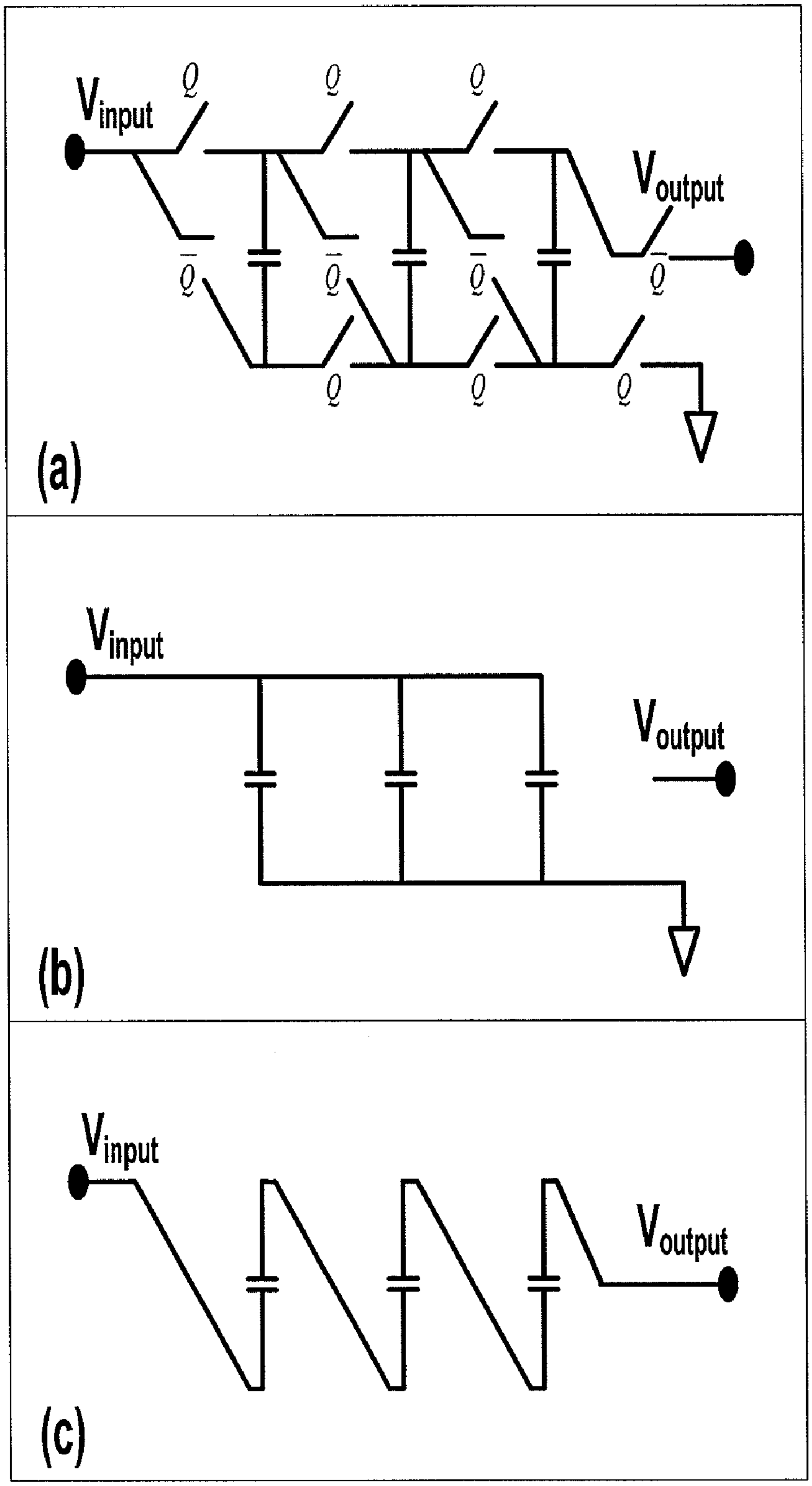
V_{input}
Q
Q
Q
V_{output}
$\bar{Q}$
$\bar{Q}$
$\bar{Q}$
$\bar{Q}$
Q
Q
Q
(a)
V_{input}
V_{output}
(b)
V_{input}
V_{output}
(c)

FIG. 15

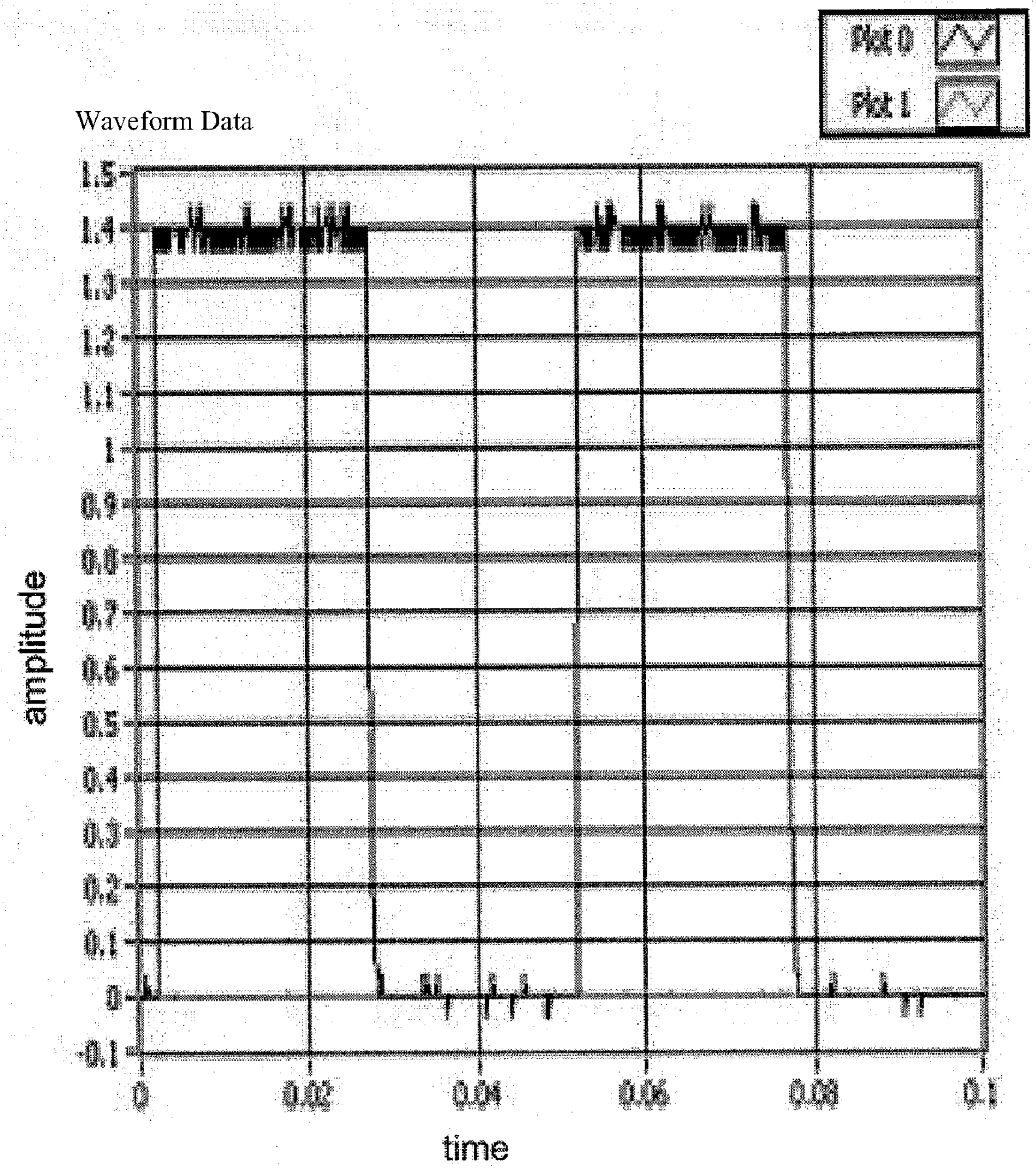
Waveform Data
Plot 0
Plot 1
amplitude
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
0
0.02
0.04
0.06
0.08
0.1
time

FIG. 16

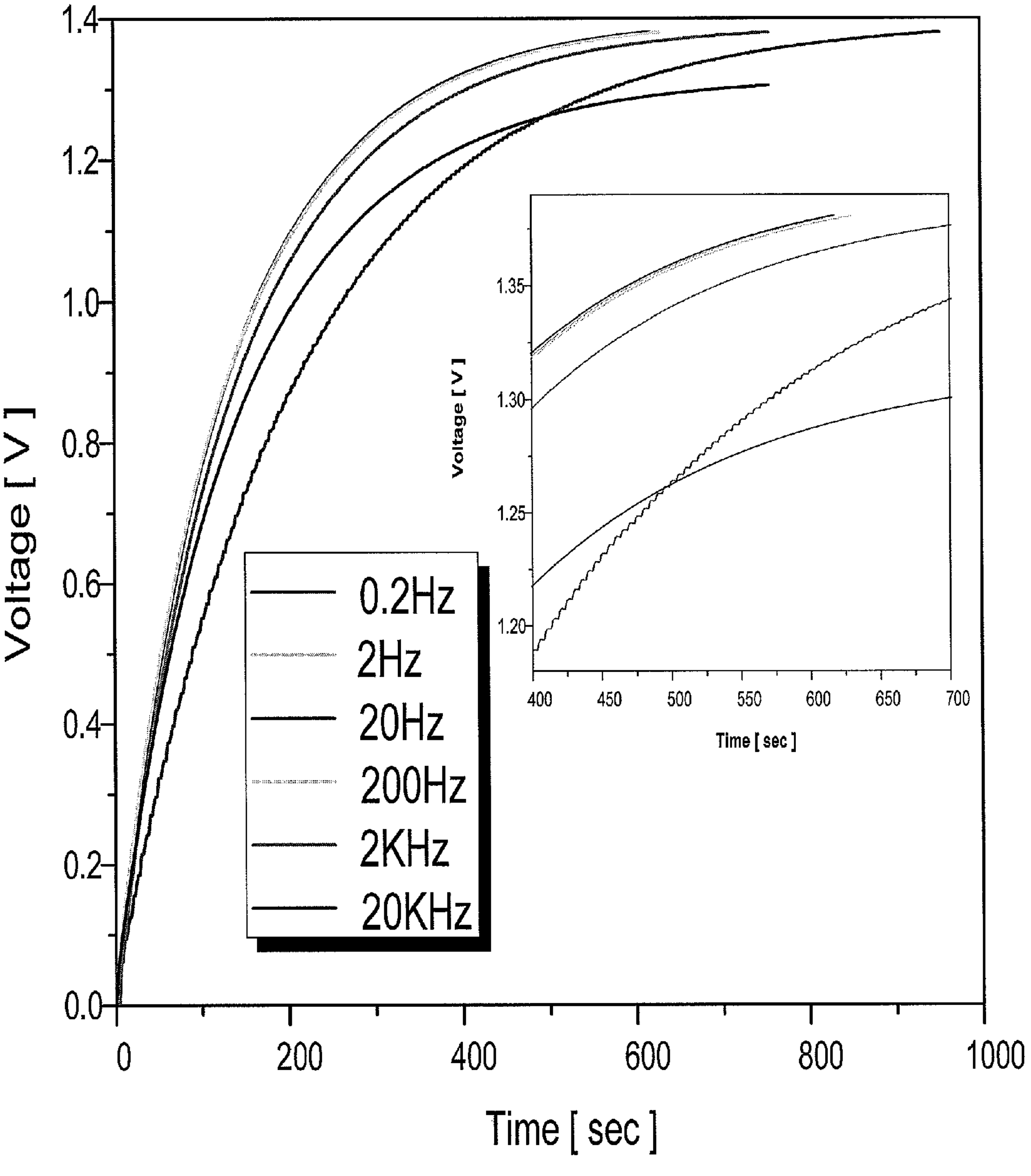

1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Voltage [V]
0
200
400
600
800
1000
Time [sec]
0.2Hz
2Hz
20Hz
200Hz
2KHz
20KHz
1.35
1.30
1.25
1.20
Voltage [V]
400
450
500
550
600
650
700
Time [sec]

FIG. 17

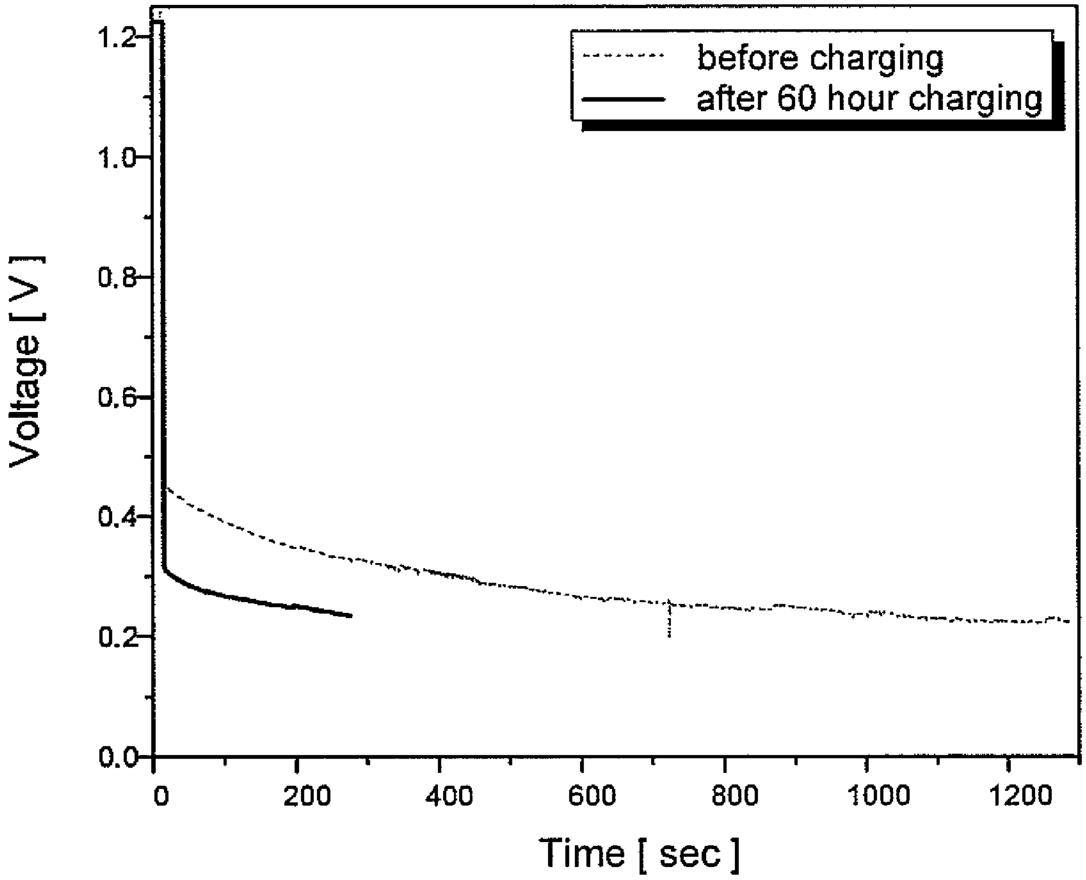
before charging
after 60 hour charging
Voltage [V]
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Time [sec]
0
200
400
600
800
1000
1200

FIG. 18

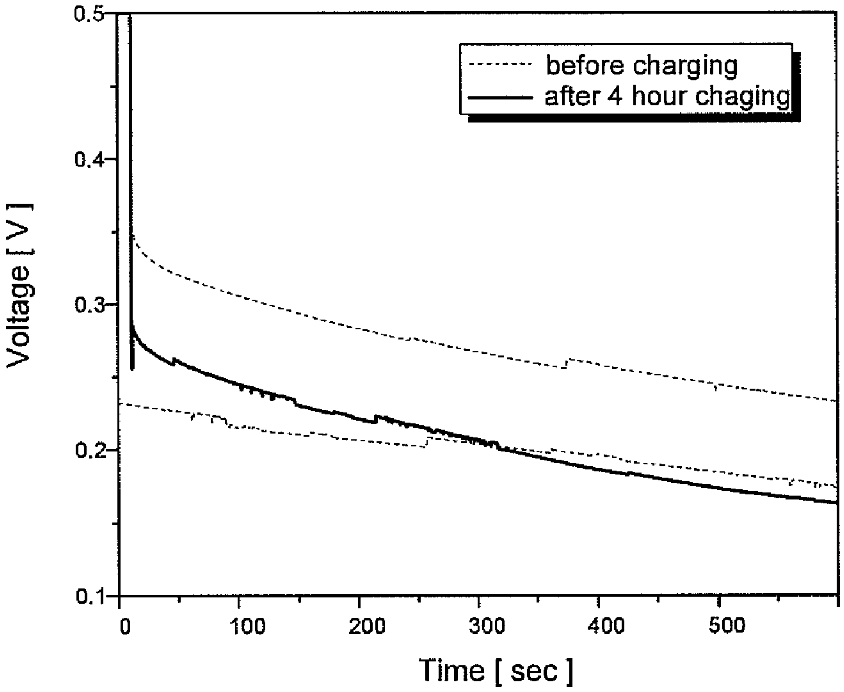
0.5
0.4
0.3
0.2
0.1
Voltage [V]
0
100
200
300
400
500
Time [sec]
before charging
after 4 hour chaging

FIG. 19

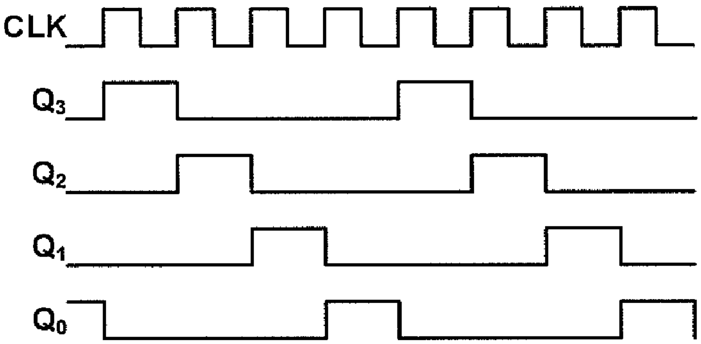
CLK
Q_3
Q_2
Q_1
Q_0

FIG. 20

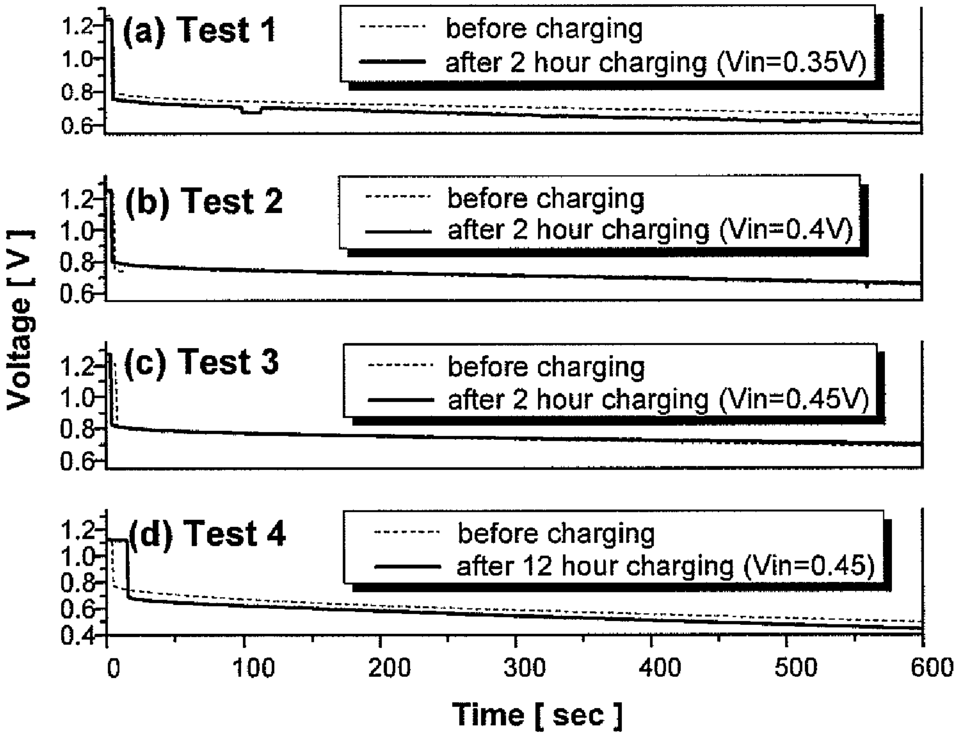
(a) Test 1
before charging
after 2 hour charging (Vin=0.35V)
(b) Test 2
before charging
after 2 hour charging (Vin=0.4V)
(c) Test 3
before charging
after 2 hour charging (Vin=0.45V)
(d) Test 4
before charging
after 12 hour charging (Vin=0.45)
Voltage [V]
1.2
1.0
0.8
0.6
0.4
0
100
200
300
400
500
600
Time [sec]

FIG. 21

RF POWER HARVESTING CIRCUIT

CROSS-REFERENCE TO RELATED PATENT APPLICATION

[0001] This application claims priority from U.S. Provisional Application No. 61/118,808 filed on Dec. 1, 2008, and U.S. Provisional Application No. 61/119,848 filed on Dec. 4, 2008, the disclosures of which are incorporated herein by reference in their entirety.

STATEMENT REGARDING FEDERALLY SPONSORED R&D

[0002] This invention was made with support under H9823004C0490 awarded by the National Security Agency. The government has certain rights in the invention.

BACKGROUND

[0003] 1. Technical Field

[0004] The present invention relates to an RF power harvesting circuit design, and more particularly, to an RF power harvesting circuit design which efficiently harvests power from RF energy.

[0005] 2. Description of the Related Art

[0006] Electronic devices are ubiquitous throughout the world today, with immeasurable applications and devices permeating all aspects of society. Energy that would have been wasted by such devices can now be recycled. For example, in some related art electronic devices (e.g., a mobile device) that generate radio frequency (RF), the RF is harvested back into the electronic device. All electronic devices require power to operate. Ambient energy in any form is an attractive source of power, particularly as power becomes more costly, or in areas where it is scarce. Ambient energy is also desirable where there is a need for electronic devices to operate for longer times between connection to readily available sources of power, or when there is a need to remotely recharge electronic devices, or to improve the overall efficiency of electronic devices. Of course, instead of scavenging for the ambient RF energy, intentional beaming of RF energy to a target device is also possible.

[0007] In this context, radio frequency (RF) energy represents an attractive source of power. Many electronic devices (e.g., communications devices) emit RF energy. For example, indoor power densities greater than 0.5 uW/cm^2 can be detected even a kilometer away from an FM radio tower. Comparable power densities can be detected at higher frequencies both domestically and internationally including GSM and ISM bands. Harvesting this energy from the environment can provide many benefits to electronic circuit and device designers. These benefits include, but are not limited to, extending operational life of electronic devices, providing new benefits such as remote recharging, reducing size of the electronic devices, and improving overall device efficiency.

[0008] RF power harvesting devices/RF energy scavenging devices have been constructed to capture this energy. An example of a related art power harvesting circuit is shown in FIG. **1**. FIG. **1** illustrates the basic Villard voltage doubler circuit. An intuitive understanding of this circuit can be gained by first examining what occurs when current flows in the direction of I_1. The diode D_2 blocks the flow of current through the capacitor C_2.

[0009] Therefore, all of the current goes across the capacitor C_1. This charges the capacitor C_1 up to roughly the same level as the peak of the AC voltage. Once the up-swing of the AC cycle (I_2) has been reached, the diode D_1 turns off, the diode D_2 turns on, and the voltage across both the AC source and C_1 drops across the capacitor C_2, charging it to approximately twice the peak voltage of the AC signal.

[0010] The related art RF scavenging circuits, however, require a minimum or threshold amount of incident RF energy to "turn on" and begin providing useful energy to other circuits and devices. Such devices can be devices that immediately perform a function or energy storage devices such as batteries and capacitors. Incident energy that falls below the minimum threshold is not captured, thereby reducing the efficiency and effectiveness of the scavenging circuit. Thus, there is a need for an RF power harvesting device that can capture and utilize ambient energy that falls below the minimum threshold for turning on the device.

SUMMARY

[0011] Embodiments of the disclosed RF energy harvesting circuit improve the sensitivity of RF energy harvesting circuits over the related art power harvesting circuits.

[0012] According to an aspect of the present invention, there is provided an RF power harvesting device including an inductor, a first capacitor connected to the inductor, a first MOSFET connected to a first node, and a second MOSFET connected to the first node, and the inductor or the first capacitor are connected to the first node.

[0013] In the RF power harvesting device, values of the first MOSFET and the second MOSFET are such that intrinsic capacitances of the first MOSFET and the second MOSFET and the inductor constitute a substantially resonant circuit during operation of the RF power harvesting device.

[0014] The RF power harvesting device of claim **1** further includes a second capacitor connected to a second node, and the second MOSFET is connected to the second node, wherein the second MOSFET is a PMOSFET.

[0015] The RF power harvesting device further includes a receiver which detects and receives ambient RF energy, and the first capacitor and the receiver are connected at a third node, the first capacitor and the inductor are connected in series at a fourth node, and the inductor is connected to the first node.

[0016] The RF power harvesting device further includes a first set of resisting elements connected to a first gate of the first MOSFET to apply a first DC bias to the first gate, and a second set of resisting elements connected to a second gate of the second MOSFET to apply a second DC bias to the second gate.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] The above and other features and advantages of the present invention will become more apparent by describing in detail exemplary embodiments thereof with reference to the attached drawings in which:

[0018] FIG. **1** illustrates a single stage of a related art Villard voltage doubler circuit.

[0019] FIG. **2**. illustrates a power harvesting device, according to an exemplary embodiment of the present invention.

[0020] FIG. **3** illustrates a power harvesting device, according to an exemplary embodiment of the present invention.

[0021] FIGS. **4**(*a*)-(*b*) illustrate different placements of a matching inductor and a DC blocking capacitor and the associated lossy parasitic capacitance and resistance, according to an exemplary embodiment of the present invention.

[0022] FIG. **5** illustrates a layout of a power harvesting device, according to an exemplary embodiment of the present invention.

[0023] FIG. **6** illustrates a power harvesting device, according to an exemplary embodiment of the present invention.

[0024] FIG. **7** illustrates biasing the gate of diode-connected MOSFETs, according to an exemplary embodiment of the present invention.

[0025] FIG. **8** illustrates a circuit design implementing sacrificial current biasing, according to an exemplary embodiment of the present invention.

[0026] FIG. **9** illustrates an ideal diode response of a diode connected MOSFET with regular FETs and a diode connected MOSFET using low threshold voltage FETs.

[0027] FIG. **10** shows a power harvesting device, according to an exemplary embodiment of the present invention.

[0028] FIG. **11** illustrates the layout of RF harvesting circuits to form a silicon integrated circuit, the layout including five RF power harvesting circuits, which are the physical implementations of the harvesting circuits shown in the circuit schematics shown in other figures, according to an exemplary embodiment of the present invention.

[0029] FIG. **12** illustrates a comparison of measured RF to DC conversion efficiency versus output voltage between a power harvesting device with sacrificial biasing, and a power harvesting device without biasing.

[0030] FIG. **13** illustrates a self-powered system.

[0031] FIGS. **14**(*a*)-(*b*) illustrate different types of batteries implemented in the system shown in FIG. **13**.

[0032] FIG. **15** illustrates a schematic of a switched capacitor DC-DC converter.

[0033] FIG. **16** illustrates a three stage converter operation for 0.35V input at 20 Hz switching using 100 uF external capacitors.

[0034] FIG. **17** illustrates the frequency dependency of a 1000 uF capacitor charging with 0.35V input and three 100 uF external storing capacitors.

[0035] FIG. **18** illustrates the discharging test results for 60 hour charging (1 KΩ) load, 0.2 Hz clock signal) with a rigid battery.

[0036] FIG. **19** illustrates the discharging test results for 4 hour charging (1 KΩ load, 25 Hz clock signal) using two chips with a rigid battery.

[0037] FIG. **20** illustrates a switching signal generated by a ring counter using four flip-flops.

[0038] FIG. **21** illustrates four converter charging results—(a) Vin=0.35V for a rigid battery, (b) Vin=0.4V for a rigid battery, (c) Vin=0.45V for a rigid battery, and (d) Vin=0.45V for a flexible battery.

DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS

[0039] The present invention will now be described more fully with reference to the accompanying drawings, in which exemplary embodiments of the invention are shown.

[0040] One embodiment of the present invention provides a modified power matched Villard voltage doubler circuit, compatible with modern complementary metal-oxide-semiconductor (CMOS) processes, which demonstrates the ability to harvest RF electromagnetic energy that is available at power levels as low as in the microwatt range. Revisions to the Villard voltage doubler have been made and accessory circuitry has been added to provide it with the ability to harvest energy from power levels as extremely low as tens of μWatts (micro) low power levels. In the related art, relatively large values of input power of hundreds of μWatts or even higher are required to utilize a Villard voltage doubler topology. The reason for this it that the general Villard voltage doubler topology utilizes rectifying diodes which need to be turned on in order to transform AC currents and voltages into DC currents and voltages, so that their intrinsic energy can be stored. However, turning on these diodes requires several tenths of a volt, and a relatively large amount of input power is typically required to generate these voltages.

[0041] The present invention provides several modifications to the Villard voltage doubler topology, which modifications allow it to harvest energy from very low power levels and go beyond the limitation previously set by the threshold level of the rectifying diodes. For instance, one embodiment of the present invention provides resonant reactive elements are incorporated in a CMOS circuit that allows for the generation of relatively large voltages, without the need for large power, in order to turn on the rectifying diodes. The Villard circuit topology is implemented in an actual CMOS process making it realizable for fabrication as an integrated circuit in modern technology. Certain parasitic elements of the CMOS process are utilized to increase voltage levels and improve circuit performance (typically, parasitic elements limit circuit performance).

[0042] Application specific circuit layout designs are disclosed so that the effect of other parasitic elements would be minimized. Diode connected CMOS devices are used instead of standard diodes to implement the Villard topology. Furthermore, in one embodiment, instead of using an n-type metal-oxide-semiconductor field-effect transistor (NMOSFET), one of the diode functions is implemented using a p-type MOSFET (PMOSFET) polarity. This embodiment mostly eliminates the body effect and allows for the diode connected MOSFETs to be turned on at a lower voltage, thereby giving rise to a lower threshold power in order to obtain AC to DC conversion and allow for energy storage at lower power levels. In another embodiment, a set of sacrificial bias resistors is used to increase the gate voltage of the MOSFETs, and thereby turn them on at a lower voltage which allows for further sensitivity and increases energy harvesting. In addition, in yet another embodiment, circuit blocks are stacked in series to generate even high DC voltage levels from microwatt levels of electromagnetic RF energy.

[0043] It is desirable to integrate the power harvesting circuits of the disclosed embodiments onto CMOS integrated circuits. To accomplish this, the diodes can be replaced with diode connected MOSFETS M**1** and M**2**, as illustrated in FIG. **2**. This modification has an advantage in that it can be naturally fabricated using a CMOS process. Furthermore, using diode connected MOSFETs gives added flexibility for establishing an adjustable effective diode turn on voltage.

[0044] An inductor L**1** can be added near the input power source (RF source) as shown in FIG. **2**. This helps give rise to a matching circuit. In addition to the standard concept of matching, which allows for maximum power transfer from a fixed source impedance, the inductor works in concert with the parasitic intrinsic capacitances usually associated with a MOSFET gate to generate a very large voltage across the reactive components (discussed below in further detail). The generation of such a large voltage is one advantage of using the power harvesting circuit shown in FIG. **1**. Specifically, the generation of the large voltage facilitates the turning on of the CMOS diode connected devices at low incident power levels in order to achieve efficient AC to DC conversion, and ultimately energy storage. Thus, instead of the parasitic intrinsic capacitances functioning to reduce the frequency response and be detrimental to CMOS circuit performance, their reactive nature when working in concordance with the added input inductor L**1** allows for generation of large voltages which will ultimately make the power harvesting device more sensitive to harvesting RF energy. This circuit has a complex impedance that mainly consists mainly of a reactive part.

[0045] A. Modification to Matching Inductor and use of Intrinsic Capacitance for Large Voltage Generation and Diode Turn-On.

[0046] As mentioned previously, in order to be able to utilize the Villard voltage doubler topology, it is necessary to achieve voltages sufficiently large to turn on the diodes (or diode connected MOSFETs) D**1** and D**2**. However, this is difficult if the power of the input signal is very low, in the microwatt range, for example. To overcome this problem, reactive components are utilized in the design which are able to accumulate relatively large amounts of AC power over time, and thus generate voltages that are sufficiently large to turn on the diodes, and thus harvest RF energy even at power levels as low as microwatts.

[0047] The reactive elements that generate the large voltages for turn on are shown in FIG. **3** (e.g., the inductor L**1**, and the intrinsic MOSFET capacitors C_{GS1}, C_{SB1}, C_{GS2}, and C_{SB2}). The inductor L**1** is used to resonate with the capacitive structures intrinsic to the MOSFETs themselves and the fabrication process. The inductor L**1** is an accessory element constructed on the chip. However, instead of adding an accessory capacitor network to generate the complementary reactive impedance, the capacitors that are typically deemed as parasitic and detrimental to circuit performance are utilized in the current embodiment. More specifically, the intrinsic source-gate capacitance (C_{GS1}, C_{GS2}), and the source-body capacitance (C_{SB1}, C_{SB2}) associated with the two diode connected MOSFETS (M**1** and M**2**), are used.

[0048] From FIG. **3**, it can be seen that that accessory inductor L**1** and the intrinsic capacitors C_{GS1}, C_{SB1}, C_{GS2}, and C_{SB2} together form a series resonant circuit. The values of these intrinsic capacitors are in the Femto-Farad range. Furthermore, the capacitors C**1** and C**2** are orders of magnitude larger than the aforementioned CMOS intrinsic capacitors, and thus have negligible impedance at the frequencies of interest (hundreds of MHz to GHz). As such, impedances of the capacitors C**1** and C**1** can be neglected. Considering the relative magnitudes of the intrinsic capacitors with respect to C**1** and C**2**, the intrinsic capacitors can be considered to be at AC ground. Since this is the case, the intrinsic capacitors are in parallel with the gate-source junction of the diode connected MOSFETs. Thus, when large voltages are generated across the capacitive reactive elements, large voltages are also generated across the gate-source voltages (V_{GS}) of the MOSFETs, thereby turning them on and allowing them to operate as rectifiers.

[0049] The optimal response of the circuit is obtained at an input frequency where the accessory matching circuit (comprising of L**1** and the intrinsic capacitors) is in resonance. This condition maximizes the voltage at the input (point B in FIG. **3**) of the circuit, and matches the source impedance of an attached antenna (or any power source) to the complex conjugate of the input impedance of the circuit. Assuming the input power source (i.e., the RF source) has a real source impedance, it can be derived that as long as the following expression (1) is true then the voltage at the input is maximized when the complex impedance (Z_{load}) at the input is matched to the source impedance Rs in FIG. **3** (Rsource in expression 1).

$$\sqrt{\frac{|Z_{load}|}{R_{source}\cos(\phi)}} > 2 \tag{1}$$

[0050] In expression 1, ϕ is the phase difference between the voltage and current wave form. This is equivalent to the phase difference between the real and imaginary components of the load impedance.

[0051] While achieving resonance is optimal for performance, the power harvesting circuit also performs well even when not at resonance. This is because, even at frequencies that are as much as 20% away from resonance, large voltages are still generated across the reactive components, specifically the intrinsic capacitors. These voltages are also in parallel with the gate-source voltage (V_{GS}) of the diode connected MOSFETs, thereby turning them on and allowing for energy harvesting even at power levels in the microwatt range.

[0052] B. Performance Improvement with Parasitic Aware Circuit Layout

[0053] While a schematic diagram of an electronic circuit is an abstraction, once that circuit is constructed into a silicon chip the performance of the current changes. One reason for this is that the physical implementation of passive structures, including inductors, capacitors and interconnects, gives rise to unintentional or parasitic elements. For example, laying out of an on-chip inductor gives rise to unintentional capacitors from the inductor to the substrate ground.

[0054] In one embodiment of the disclosed power harvesting device, performance enhancement can be achieved by switching the order of the impedance matching inductor L**1** and the DC blocking capacitor C**1**, as illustrated in FIG. **4*b***. While theoretically the schematic shown in FIG. **3** is independent of the order of the L**1**-C**1** pair, it becomes apparent after examination of the parasitics incurred during layout, that the schematics in FIG. **3** (also see FIGS. **4*a***) and **4*b*** are different. The schematic produced by the layout shown in FIG. **4*b*** is less sensitive to the negative parasitics associated with the capacitor C**1**.

[0055] In particular, the impedance to voltage doubler at point A in FIG. **3** is dominated by the intrinsic capacitances of the diode connected MOSFETs (C_{GS1}, C_{SB1}, C_{GS2} and C_{SB2}). Therefore, it appears as a short at the RF frequency and the impedance at point B is also highly reactive. The parasitic capacitance of C**1** results in an imaginary impedance parallel to the input impedance at point A and B. This imaginary impedance serves to reduce the magnitude of the complex impedance and reduce the phase difference between the imaginary and real part of the complex input impedance. It can be seen from the expression (2) below that this will reduce the magnitude of the voltage at the input to the voltage doubler.

$$|V_{load}| = \frac{V_{source}}{2}\sqrt{\frac{|Z_{load}|}{R_{source}\cos(\phi)}} \tag{2}$$

[0056] By switching the order of the capacitor C**1** and the inductor L**1**, as in FIG. **4***b*, the reactive part of the impedance at points D and E is tuned out generating a large resonant voltage across the MOSFET capacitors, resulting in a real impedance on the order of 10 to 100 ohms. This is orders of magnitude lower than any parasitic impedance to the substrate due to C**1**. Therefore, the circuit in FIG. **4***b* shows superior performance to the circuit in FIGS. **3** and **4***a* in the presence of parasitic capacitances from C**1**.

[0057] This improvement can also be understood as follows. If one considers the operation of the resonant voltage generator circuit, one realizes that the largest voltages are generated at point B and the lowest voltage exist at the node connecting inductor L**1** and the source resistor Rs in FIG. **3**. Therefore, in the current embodiment, the parasitic capacitor and parasitic series resistor are moved away from the highest potential in the circuit to the lowest potential in the power harvesting circuit by switching the order of the inductor L**1** and the capacitor C**1**. While this has no effect on the theoretical operation of the circuit with respect to a schematic (done without regard to actual layout and fabrication), it greatly improves actual operation by placing the parasitic elements at a lower voltage where they will draw substantially less current, and thereby minimize their negative impact on power harvesting circuit's operation.

[0058] Another implementation of this concept is performed with the layout of inductor L**1**. The inductor L**1** is implemented on two metal layers of the CMOS process. The lower layer, which is closer to the substrate, gives rise to larger parasitic capacitance and substrate resistance. Therefore, the inductor L**1** is fabricated with the lower layer location placed at the location where the circuit voltage is lowest. This location is the side of inductor L**1** that is closer to the input and specified by point D in FIG. **4***b*.

[0059] It should also be noted that modern processes such as the IBM 8 RF process require that a certain ratio of gate tiedowns to metal area be maintained. Connecting the inductor L**1** directly to the gate of NMOS **2** (e.g., M**2** in FIG. **2**) in a Villard voltage doubler would require an unrealistic number of tiedown contacts. Therefore, this technique is best suited for designs that use a PMOS in place of NMOS **2** as discussed below.

[0060] C. Use of PMOS to Reduce Body Effect

[0061] FIG. **5** illustrates a power harvesting circuit with an NMOSFET (e.g., M**2** in FIG. **2**) replaced with a PMOSFET (PMOS in FIG. **5**). In a steady state, an NMOS **2** (e.g., M in FIG. **2**) will have a source potential that is significantly higher than the body potential. This gives rise to an increase in the threshold voltage of NMOS **2** through the body effect. This causes the diode connected NMOS **2** to turn on at a higher voltage, thereby reducing efficiency. To minimize the body effect in the current embodiment, NMOS **2** is replaced with a diode connected PMOS, as shown in FIGS. **3** and **5**. The PMOS has the body, gate, and drain node connected to Vout and the source node connected to the voltage at the source terminal of the diode connected NMOS **1**. The body effect of NMOS **1** (e.g., M**1** in FIG. **2**) is already minimized and is treated similar to M**1** in FIG. **2**. The PMOS only conducts when the gate voltage is below the source voltage. Since the gate has been connected to Vout, this occurs when the AC voltage at the input is positive with respect to ground. Since both the body and the source of the PMOS are connected to Vout, no threshold voltage increase occurs due to the body effect. This improvement can be combined with the other embodiments discussed above, including the embodiment shown in FIG. **3** where the order of the capacitor C**1** and the inductor L**1** are transposed to reduce parasitic capacitive and resistive losses.

[0062] In order for the design to work, the PMOS needs to be sufficiently isolated to allow connecting the body of the PMOS to Vout without any current flowing from the PMOS body to the bulk substrate. Connecting the body of a diode connected PMOS to Vout is possible due to the fact that PMOS FETs are placed in an n doped well inside of the p substrate.

[0063] As seen in FIG. **6**, the PMOS is built in an N well. Since the junction from the N well to the P substrate acts as a diode, if the voltage potential in the N well is higher than the voltage potential in the P substrate, current will not flow. This is what allows the body to be connected to Vout which is higher than the potential of the substrate. The junction capacitance between the N well and the substrate is parallel to and several orders of magnitude lower than the output capacitance C**2** of the power harvesting circuit. Therefore, the junction capacitance has no significant effect on the circuit performance.

[0064] D. Sacrificial Biasing

[0065] The output voltage of the related art voltage doubler is two times the input voltage minus the threshold voltage of both diodes. Therefore, minimizing the threshold voltage of the diodes maximizes the voltage at the output of the voltage doubler. Previous works in the related art have attempted to improve RF power harvesting efficiency by using special MOSFETS with reduced gate threshold voltages.

[0066] In an exemplary embodiment of the present invention shown in FIGS. **7** and **8**, an alternative technique uses biasing at the gate combined with the use of a PMOS in place of the output NMOSFET to eliminate the threshold voltage for both diode connected MOSFETs.

[0067] For example, in the embodiment shown in FIG. **7**, voltage sources are directly connected to the gates of the diode connected MOSFETs in the RF power harvesting circuit.. An alternative approach is to use the desired output voltage to create the bias voltages through a voltage divider network, as shown in FIG. **8**. This technique sacrifices some current to create an overall improvement to system efficiency. In this embodiment, the NMOS **2** is replaced by a PMOS IN FIG. **8** for two reasons. First, the threshold voltage of the output diode connected MOSFET must be below Vout to be obtainable through a voltage divider network. Second, the gate is connected to a DC node in the circuit. Sufficiently larger resistors R**1**, R**2**, R**3**, and R**4** are used in the divider network IN FIG. **8** to reduce power dissipated due to the bias current. The resistors are chosen to be 1 MΩ to 10 MΩ for 10 μW input power. Generally, the smaller the input power, the higher the resistances. This power dissipation can be designed to be orders or magnitude below the increase in output power due to the bias voltages. Practical limitations are placed on the size of the resistors used in the divider network due to a physical limitation of the resistor size in the circuit layout.

[0068] In the current embodiment, traditional MOSFETs outperform low threshold voltage MOSFETs when sacrificial current biasing is utilized. This is in contrast to related art literature that has concluded that the use of low threshold voltage MOSFETs provides the best performance for RF power harvesting. So, instead of using low threshold voltage MOSFETs, we have introduced sacrificial biasing while using regular MOSFETs and have achieved improved results. The improvement is because we achieve the effect of easily turning on the MOSFETs, while retaining the improved performance of regular MOSFETs with respect to the turn on characteristics as shown in FIG. **9**. Regular MOSFETs, have a sharper turn-on transition than low threshold voltage MOSFETs resulting in a faster turn on when transitioning to the open state, showing better RF conversion efficiency.

[0069] FIG. **10** illustrates a comprehensive design of a power harvesting device which includes the transposition of the inductor L**1** and the capacitor C**1** of FIG. **4***b*, the implementation of the PMOS device in place of M**2** as shown in FIG. **5**, inclusion of sacrificial bias resistors R**1**-R**4** as shown in FIG. **8**, as well as implicit utilization of MOSFET intrinsic capacitors for achieving maximum turn on voltages with low power input as shown in FIG. **3**.

[0070] FIG. **11** is the layout of the harvesters to form a silicon integrated circuit. The layout shows five RF power harvesting circuits, which are the physical implementations of the harvesting circuits shown in the circuit schematics shown in other figures. The circuits have been used to compare and test the various designs of the energy harvester. The testing provided the data that has lead to the conclusion and claims in this patent application.

[0071] FIG. **12** illustrates a comparison of measured RF to DC conversion efficiency versus output voltage between a power harvesting device with sacrificial biasing, and a power harvesting device without biasing. The harvester with sacrificial biasing achieves high efficiency compared with unbiased design for 1V output voltage. The higher curve represents a higher conversion efficiency for RF energy into DC energy. That is, the results show that for the given output voltages, the sacrificial biasing provides improved efficiency over the unbiased design.

[0072] FIGS. **13**-**21** relate to self-powered systems. An RF power harvesting device discussed with respect to the above embodiments is incorporated. Available RF power is sometimes very weak, so it is not easy to generate a sufficient output for direct use. Using a combination block, converter, and battery provides an attractive solution.

[0073] FIG. **13** shows a block diagram of self-powered system including an RF power harvesting block **1**, a voltage converter **2**, a battery **3**, a combined block **4**, and functional electronics **5**. Here, the RF power harvesting block **1** could be one of the RF power harvesting devices discussed in the above embodiments.

[0074] In one embodiment, the converter **2** is a switched-capacitor DC-DC voltage converter. The converter is designed to generate maximum six times of input voltage at output using external capacitors. Through capacitor charging tests, it shows up to 40.5% energy transfer efficiency using 0.35V input for a 1000 uF capacitor charging to 1.4V during 10 minute with a three stage converter. The battery **3** can be a rigid type battery or a flexible type battery. Both types of batteries are successfully recharged using the converter **2**. In one embodiment, two hour charging with four parallel converters recharges a rigid battery fully with 0.4V input, so it returns the battery potential to the initial condition before discharging of a 9.85 KΩ load resistor during 10 minutes.

[0075] The rigid type electrochemical battery comprises hydrated Ruthenium Oxide ($RuO_2.xH_2O$) and activated Carbon (CA). These cells have high current capacity, rechargeability, and even flexibility. One of the targets of the battery **3** is playing a role of the power source for distributed sensor networks, or Smart-dust nodes, for achieving a stand-alone system. Much research in the related art has been focused on implementing self-powered systems such as a node of low-power ad hoc distributed networks using an RF power harvesting device. However, when the input power at RF harvesting block **1** is very small, the output is not enough to drive a system or node (e.g., the functional electronics **5**). This problem can be solved by using a voltage converter **2** and rechargeable battery **3** combination block **4**, as shown in FIG. **13**. For example, a DC voltage converter **2** stacks the small voltage of RF energy harvesting block **1** up to the necessary charging voltage for battery cell and charges a battery **3**. Then, the battery **3** provides enough bias or power for the system (e.g., functional electronics **5**) only when it needs to be working. In the current embodiment, a switched-capacitor DC voltage converter is used as the converter **2** for battery charging. This converter **2** and battery **3** combined block **4** bridges a gap between RF power harvesting block **1** and functional electronics **5**, e.g., in insufficient RF power environments. So, it can help to implement a self-powered or stand alone system more effectively.

[0076] Usually, the developed battery cells have 1.1V~1.2V built in potential after fabrication without any other charging or discharging steps, and slightly higher charging voltage than the built in potential was fine for battery recharging. Actually, charging voltage and time for constant voltage charging scheme show a trade-off relation. A 1.4 charging voltage of DC power supply is used during two hours for full recharging. FIG. **14** shows two types of batteries, a rigid battery (FIG. **14**(*a*)), and a flexible battery (FIG. **14**(*b*)).

[0077] Switched-Capacitor DC-DC Converter

[0078] There are two reasons why we choose a switched-capacitor scheme for DC-DC converter among several different types. The first is its simplicity of implementation, and the second is it needs capacitors for storing energy. The second factor is advantageous because the battery cells can be used as a capacitor also by some modifications. So the switched-capacitor method is a very attractive one.

[0079] FIG. **15**(*a*) shows the schematic of a switched capacitor DC-DC converter with three stages. A single chip is implemented using IBM8RF 0.13 um CMOS process including every switch for a five stage converter, except external capacitors. That is, maximum available output voltage is six times of input voltage through five external capacitors, and the output can be selected by a user from minimum (input×1) to maximum (input×6) just connecting an external capacitor's plus pad to input.

[0080] In the schematic, the top switches are made by nMOS only, but the middle and bottom switches are made by transmission gate, combining nMOS & pMOS to transfer the accumulated potential on each capacitor to the next stage without any loss. The converter works on two different phases. For capacitor storing phase ($Q=1$, $\overline{Q}=0$ FIG. **15**(*b*)), the input node is connected to every external capacitor, so all capacitors are in parallel. However, the output is disconnected from the converter. In this step, external capacitors are charged up through input. Then, during battery charging phase (Q=0, $\overline{Q}$=1,

[0081] FIG. **15**(*c*)), the parallel configuration of capacitors is changed into series, and the output is connected to the last capacitor's plus node. Note the input can be used in both phases without any waiting or sleeping mode. So, it can decrease the charging time by fully using the input source. FIG. **16** shows the output (~1.4V) of the converter using a three stage converter for 0.35V input at 20 Hz clock signal using 100 uF three external capacitors. Note the body contacts of pMOS in converter switches are connected into external 1.5V DC power supply. This external source can be replaced with a battery for fully self-powered system implementation.

[0082] Big Capacitor Charging

[0083] At the initial tests of the converter, a big capacitor, 1000 uF, replaces a battery for charging test to verify functionality of the converter, just using three stages. Also, frequency dependency is checked for different clock signals of 50% duty cycle. The external capacitors, storing energy from input voltage, and input voltage are fixed as 100 uF and 0.35V, respectively. FIG. **17** shows the results of capacitor charging for different clock signals, and the inner graph is a magnified version of 400~700 sec section. All tests are stopped when the capacitor voltages reach 1.38V, except 20 KHz test. The plots for 2 Hz, 20 Hz, and 200 Hz test are almost overlapped. As listed in Table 1, 2 Hz and 20 Hz test have almost same results for energy transfer efficiency, around 40%, and charging time about 10 minutes, where the supplied energy is calculated by integrating the measured current from power supply and the stored energy is just calculated value using $\frac{1}{2}CV^2$. As the clock frequency goes higher, the saturation voltage goes lower. It is caused by frequency dependency of dynamic power consumption of CMOS switches [6]. By frequency dependency tests, a 0.2 Hz~20 Hz clock frequency range is used for battery charging tests.

TABLE 1

Frequency dependency of 1000 uF capacitor charging properties

Clock Frequency [Hz]	Charging Time [sec]	Supplied Energy [mJ]	Stored Energy [mJ]	Energy Transfer Efficiency [%]
0.2	953.5	2.352	0.952	40.48
2	611.5	2.349	0.952	40.53
20	617.5	2.394	0.952	39.77
200	629.5	2.555	0.953	37.30
2K	755.5	2.755	0.953	34.59
20K*	755.5	3.010	0.852	28.31

*For 20 KHz test, 1.3 V saturated voltage is used for calculations.

[0084] Battery Charging

[0085] The test procedure follows three steps for battery charging: 1) initial discharging, 2) cell charging, and 3) discharging.

[0086] The first charging test set up is 0.3V input from power supply, five 100 uF external storing capacitors for a five stage converter, which provides 1.8V output without load, and 50% duty cycle 0.2 Hz clock signal. Actually, a test with 0.25V input, making 1.5V output by five stages, was performed, but the charging time took over several days. So, to reduce charging time, input voltage is increased from 0.25V to 0.3V. FIG. **18** shows the test results. After the initial discharging (dashed line) during 1280 seconds, the load voltage was dropped from 0.485V to 0.223V for 1 KΩ load resistor. The black solid line shows the discharging after 60 hours charging. 60 hours charging gives 0.316V initial load voltage, when 1 KΩ load is connected to the battery after charging. Though the charging time, 60 hours, is a little long, we demonstrate a DC-DC converter is working properly for the battery charging.

[0087] To reduce charging time, the input voltage is increased from 0.3V to 0.34V and two converters are used in parallel for another test. Clock signals are applied for two converters to operate in opposite mode at the same time. That is, if a converter is in capacitor storing mode, the other is in battery charging mode. The test setup is 0.34V input, generating 2V output without load due to using five stage, and 25 Hz clock signal with 50% duty cycle. The battery cell is discharged with 1 KΩ load during 20 minute before charging, as shown by the dashed line in FIG. **19**, and load voltage drop is measured from 0.358V to 0.174V. Then, two converters charge the cell for 4 hours. This leads to 0.29V initial load voltage at discharging after charging (solid line in FIG. **19**). Using higher input voltage and two converters, the charging time is effectively reduced from 60 hour to 4 hour.

[0088] Another charging test is performed using four converters for further decreasing charging time. To run four converters together, switching signals are generated using a ring counter, which is made by four D Flip-Flops. FIG. **20** shows the generated timing diagram of the counter, Q_0~Q_3. Simply, inverters inverse each timing signal for the reverse switching signal inputs, $\overline{Q_0}$~$\overline{Q_3}$. Using the switching signals, if Q_0 is high, only one output of converters is connected to battery cell, and the others are in capacitor storing mode.

[0089] During capacitor storing mode, external capacitors are in parallel, so it gives a high total capacitance. Accordingly, it makes a long RC time constant for capacitor storing. Previous tests used five stages with 100 uF external capacitor, which gave 500 uF total capacitance for capacitor storing mode. For quick charging up capacitors, only three stages are used for other tests, which need three external capacitors for each converter. However, reducing the number of stage affects the output voltage directly. To compensate for this, higher input voltage is used. Also, higher input voltage can reduce the charging time further by providing more charges on storing capacitor, if the current capacity of the input signal is not limited. In the following tests, 0.35V, 0.4V, and 0.45V input are used for comparison.

[0090] Four charging tests are performed using a rigid (Test 1~3) and flexible (Test 4) type battery. For a rigid battery, charging time is limited to 2 hour, but the flexible cell is charged for 12 hours due to the cell's own property. Usually, the flexible battery consumed more current at the same charging voltage compared with rigid one. Furthermore, 9.85 KΩ and 100.2Ω load resistor are used at discharging for rigid and flexible battery, respectively. The test results and setup are shown in FIG. **21** and Table 2.

[0091] Every charging test shows energy recovery of battery. Although 2 hour charging with 0.35V input is not sufficient for full recovery before discharging, other two tests, Test 2 & 3, show the fully recovered battery potential compared with discharging before charging tests. However, Test 3 does not give any improvement in spite of higher input voltage, 0.45V. This may be due to the aging of the cell through several charging and discharging tests. Test 4 shows 12 hour charging result for flexible cell.

TABLE 2

Charging test setup and results

							Discharging during 10 minutes			
			Charging (three stage converter)				Before charging [V]		After charging [V]	
	Battery Cell		Input	Clock	Charging time	Load resistor	Initial load	Final load	Initial load	Final load
Test #	Type	Size	[V]	[Hz]	[Hour]	[Ω]	voltage	voltage	voltage	voltage
1	Rigid	4 cm^2	0.35	20	2	9.85K	0.804	0.659	0.759	0.612
2	Rigid	4 cm^2	0.4	20	2	9.85K	0.790*	0.648	0.804	0.659
3	Rigid	4 cm^2	0.45	20	2	9.85K	0.818	0.688	0.828	0.701
4	Flexible	4 cm^2	0.45	2	12	100.2	0.775	0.497	0.689	0.445

*Initial load voltage for Test 3 before charging is estimated value because of a slacken load connection.

[0092] As described above with respect to FIGS. **13-21**, a switched-capacitor DC-DC converter is implemented using a CMOS process for a battery cell charging application. Up to 40% energy transfer efficiency is obtained through capacitor charging tests. For battery cell charging tests, the functionality of the converter is verified for both rigid and flexible type battery cells using much smaller input voltage compared with general high charging voltage, more than 1.2V for the batteries discussed with respect to FIGS. **13-21**.

[0093] The techniques described above show a dramatic increase in power harvesting efficiency as compared to existing RF Power Harvesting designs and, enable recharging portable wireless electric using ambient RF energy sources. Utilizing these improvements, a design goal of generating 1V output voltage with a greater than 20% RF to DC conversion efficiency from RF energy levels measured in the environment (66 uW) was met. This represents better than double the RF to DC conversion efficiency of the related art power matched RF energy harvesting circuit based on a Villard voltage doubler.

[0094] Several design improvements, novel to RF power harvesting circuits, have been disclosed. These improvement include RF to DC conversion efficiency through a reduction in the body effect of diode connected MOSFETs, reduction in the threshold voltage and by reducing the affects of circuit parasitics. These circuit improvements have been simulated to show a better than 60% improvement to RF to DC conversion efficiency.

[0095] While the present invention has been particularly shown and described with reference to exemplary embodiments thereof, it will be understood by one of ordinary skill in the art that various changes in form and details may be made therein without departing from the spirit and scope of the present invention as defined by the following claims.

1. An RF power harvesting device comprising:
an inductor;
a first capacitor connected to the inductor;
a first MOSFET connected to a first node, and
a second MOSFET connected to the first node,
wherein the inductor or the first capacitor are connected to the first node.

2. The RF power harvesting device of claim **1**, wherein values of the first MOSFET and the second MOSFET are such that intrinsic capacitances of the first MOSFET and the second MOSFET and the inductor constitute a substantially resonant circuit during operation of the RF power harvesting device.

3. The RF power harvesting device of claim **1**, further comprising:
a second capacitor connected to a second node, and the second MOSFET is connected to the second node,
wherein the second MOSFET is a PMOSFET.

4. The RF power harvesting device of claim **1**, further comprising:
a receiver which detects and receives ambient RF energy,
wherein the first capacitor and the receiver are connected at a third node, the first capacitor and the inductor are connected in series at a fourth node, and the inductor is connected to the first node.

5. The RF power harvesting device of claim **1**, further comprising:
a first set of resisting elements connected to a first gate of the first MOSFET to apply a first DC bias to the first gate; and
a second set of resisting elements connected to a second gate of the second MOSFET to apply a second DC bias to the second gate.

6. A receiver comprising:
an impedance matching circuit connected to a first node;
a first MOSFET connected to the first node; and
a second MOSFET connected to the first node, wherein
the first MOSFET is a diode-connected p type MOSFET.

7. The receiver of claim **6**, wherein the impedance matching circuit comprises an inductor and a capacitor in series, the inductor connected to the first node.

8. The receiver of claim **6**, wherein a resistive biasing network provides a bias voltage to gate terminals of the first MOSFET and the second MOSFET.

9. A system comprising:
a receiver;
a voltage converter; and
an energy storage device, wherein
the receiver is configured to receive a high frequency signal and supply a first voltage to the voltage converter, and
the voltage converter is configured to amplify the first voltage supplied by the receiver and supply the amplified voltage to the energy storage device.

10. The system of claim **9**, wherein the voltage converter includes a switched capacitor circuit.

* * * * *

US 20140111032A1

(19) **United States**
(12) **Patent Application Publication**
Shearer et al.
(10) **Pub. No.: US 2014/0111032 A1**
(43) **Pub. Date: Apr. 24, 2014**

(54) **POWERING DEVICES USING RF ENERGY HARVESTING**

(71) Applicant: **POWERCAST CORPORATION**, Ligonier, PA (US)

(72) Inventors: **John G. Shearer**, Ligonier, PA (US); **Charles E. Greene**, Pittsburgh, PA (US); **Daniel W. Harrist**, Carnegie, PA (US)

(73) Assignee: **POWERCAST CORPORATION**, Ligonier, PA (US)

(21) Appl. No.: **14/143,334**

(22) Filed: **Dec. 30, 2013**

Related U.S. Application Data

(60) Continuation of application No. 12/499,618, filed on Jul. 8, 2009, now Pat. No. 8,621,245, which is a division of application No. 11/447,412, filed on Jun. 6, 2006, now abandoned.

(60) Provisional application No. 60/688,587, filed on Jun. 8, 2005.

Publication Classification

(51) **Int. Cl.**
H02J 17/00 (2006.01)

(52) **U.S. Cl.**
CPC *H02J 17/00* (2013.01)
USPC **307/149**

(57) **ABSTRACT**

Disclosed is an apparatus for an application including a core device for the application. The apparatus includes a power (preferably RF energy) harvester connected to the core device to power the core device. Also disclosed is a method for an application. The method includes the steps of converting RF energy into usable energy. There is the step of powering the core device with the usable energy.

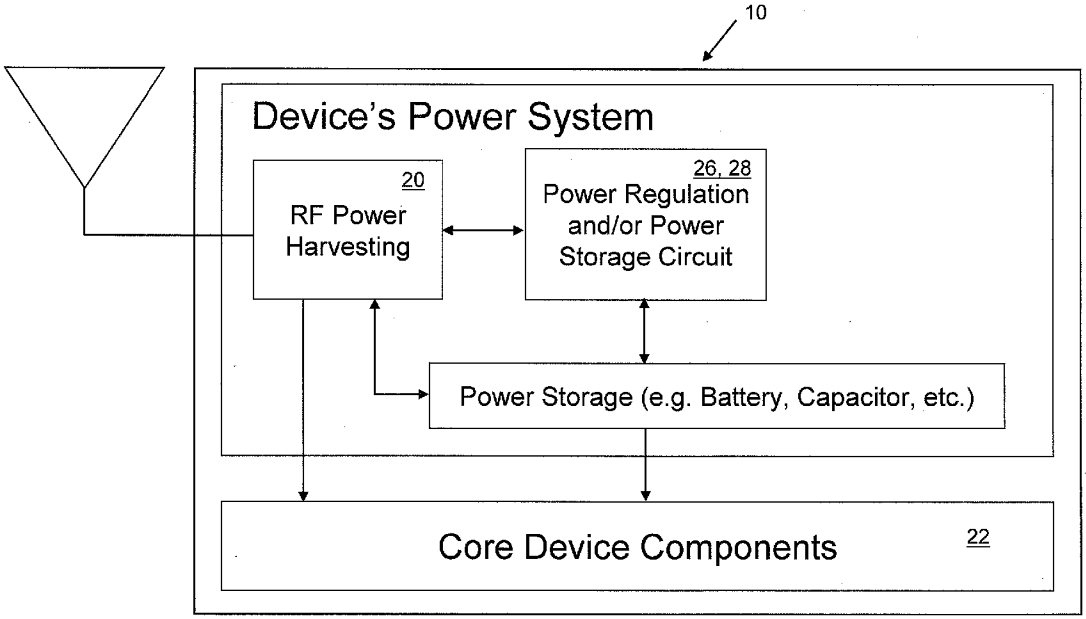
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

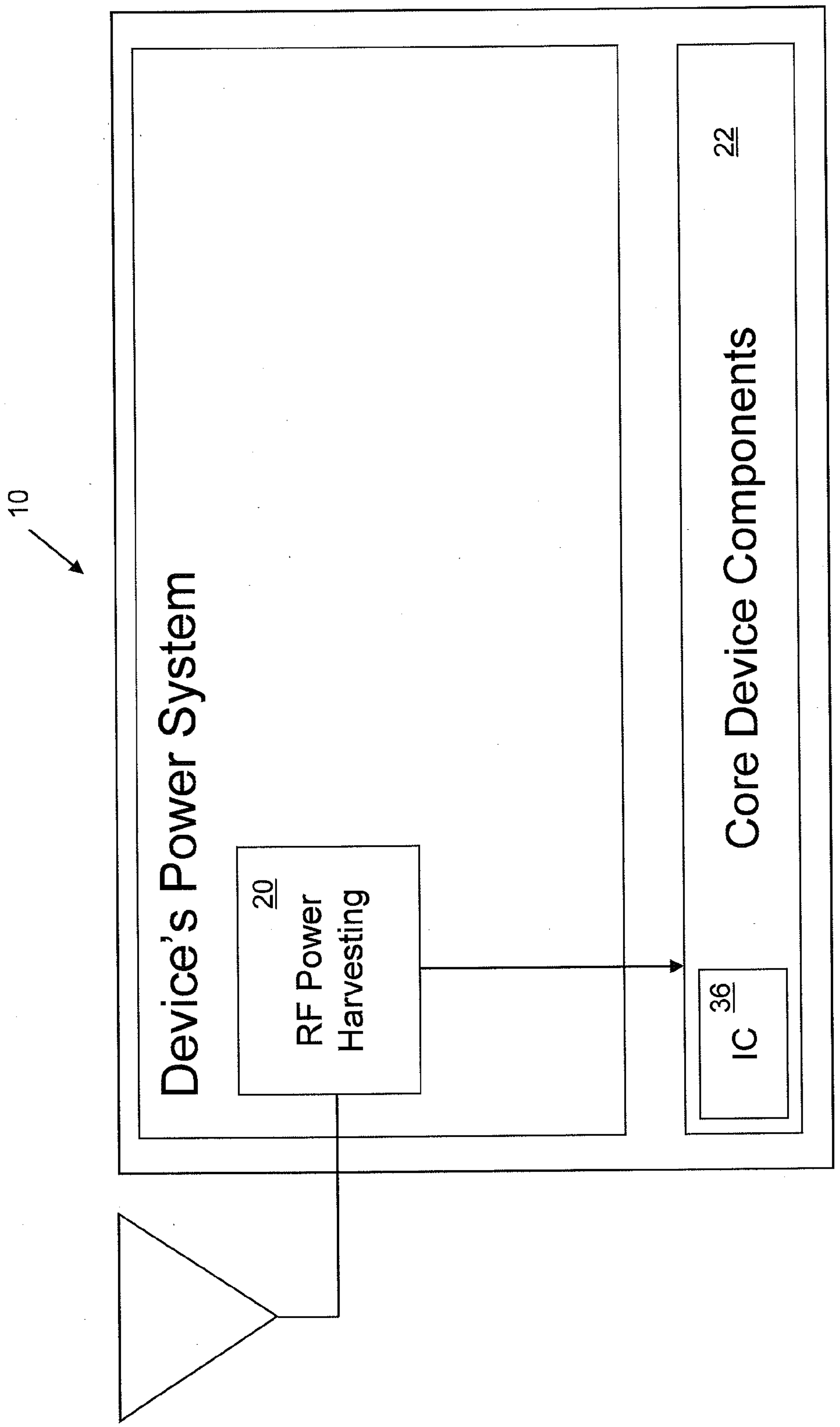
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
IC
36

Fig. 1

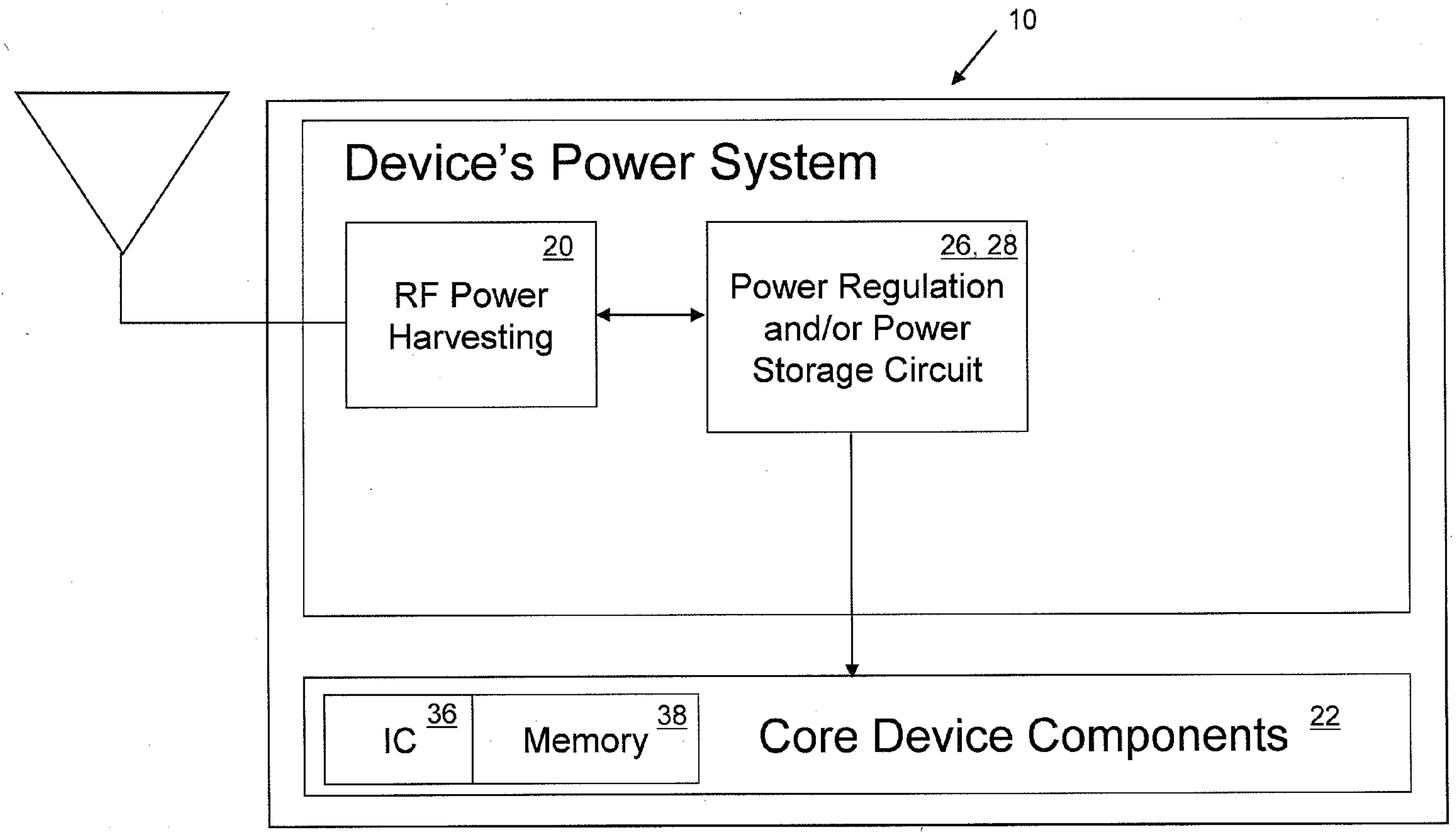
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
IC 36
Memory 38
Core Device Components 22

Fig. 2

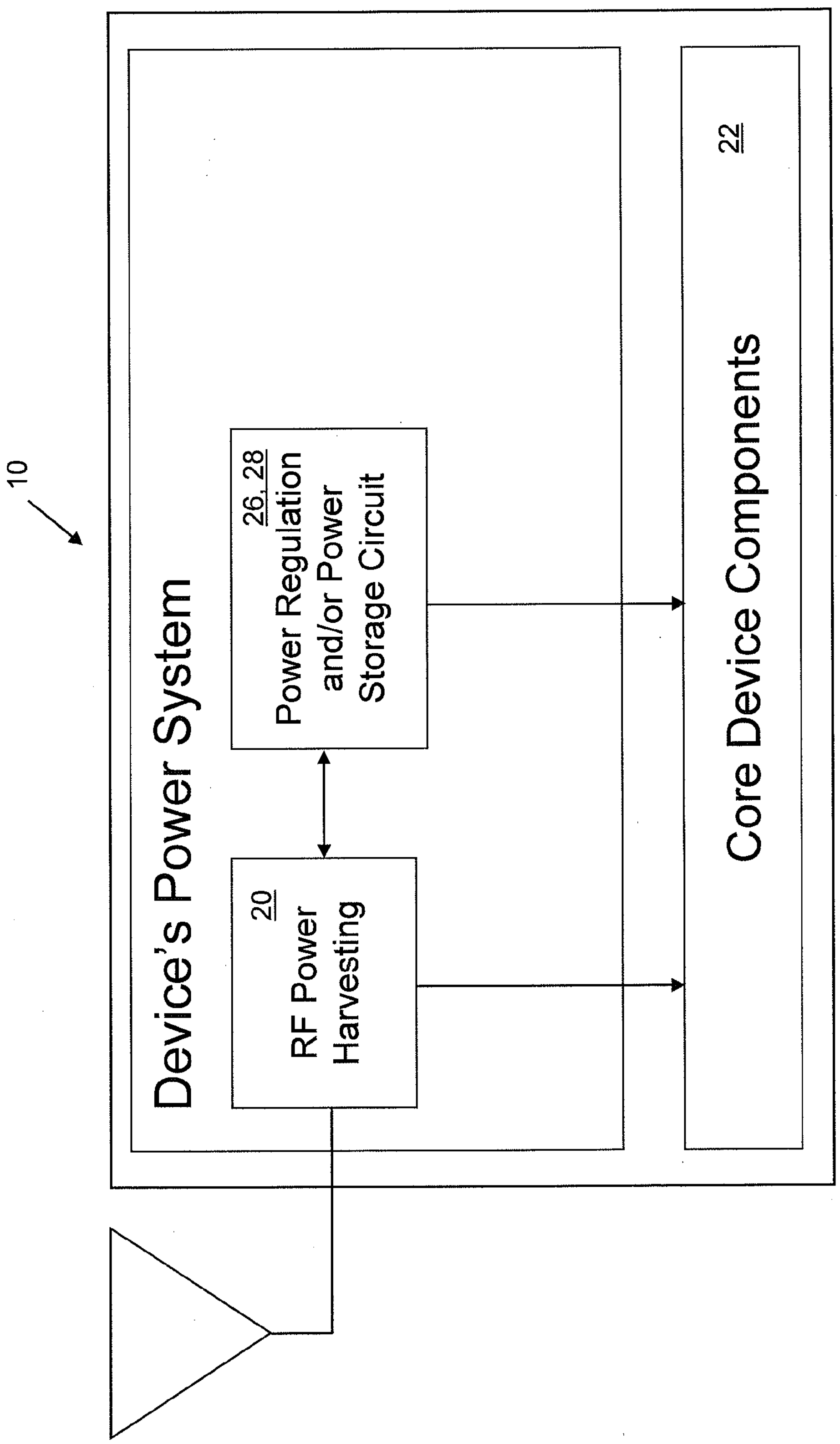
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Core Device Components
22

Fig. 3

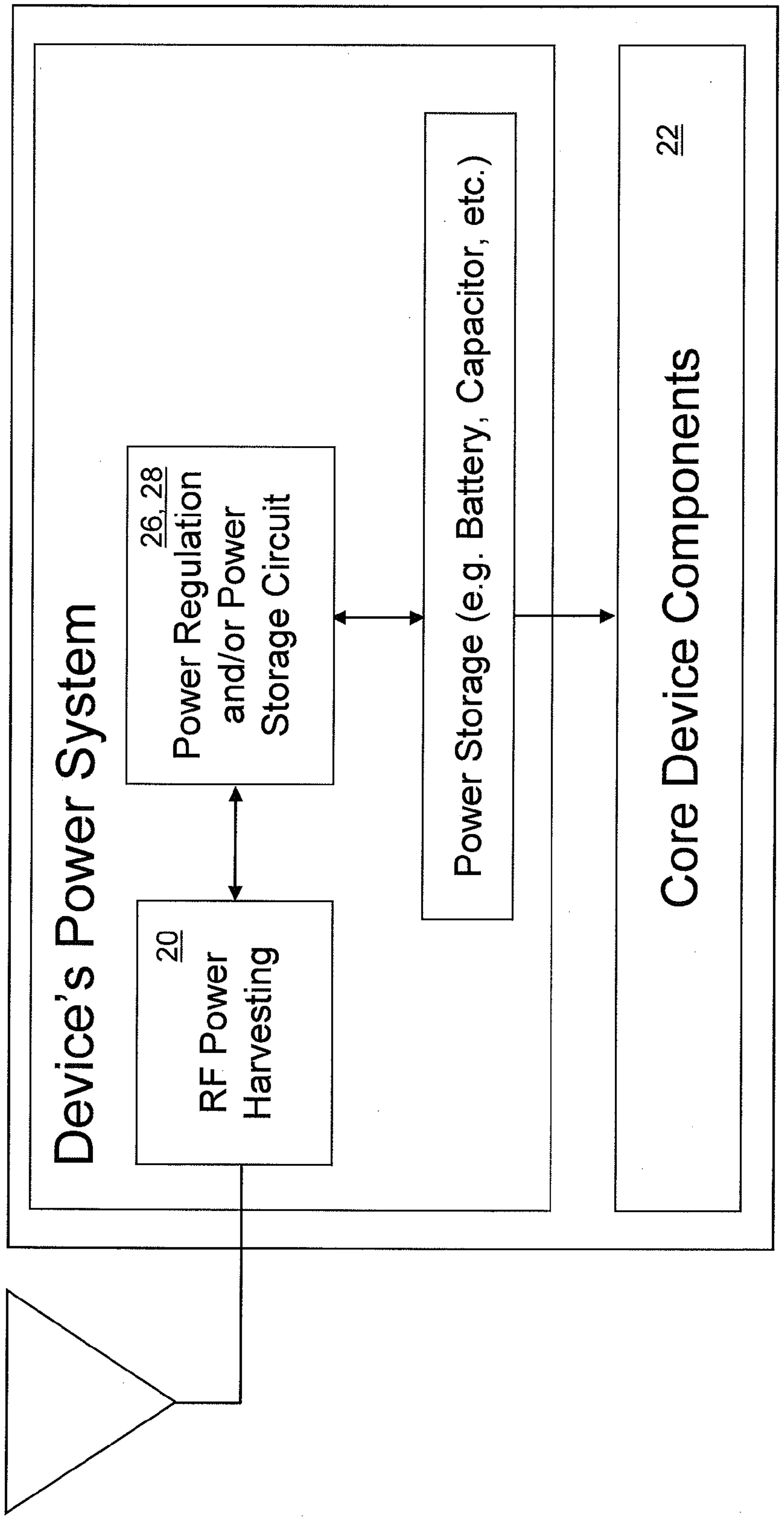
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 4

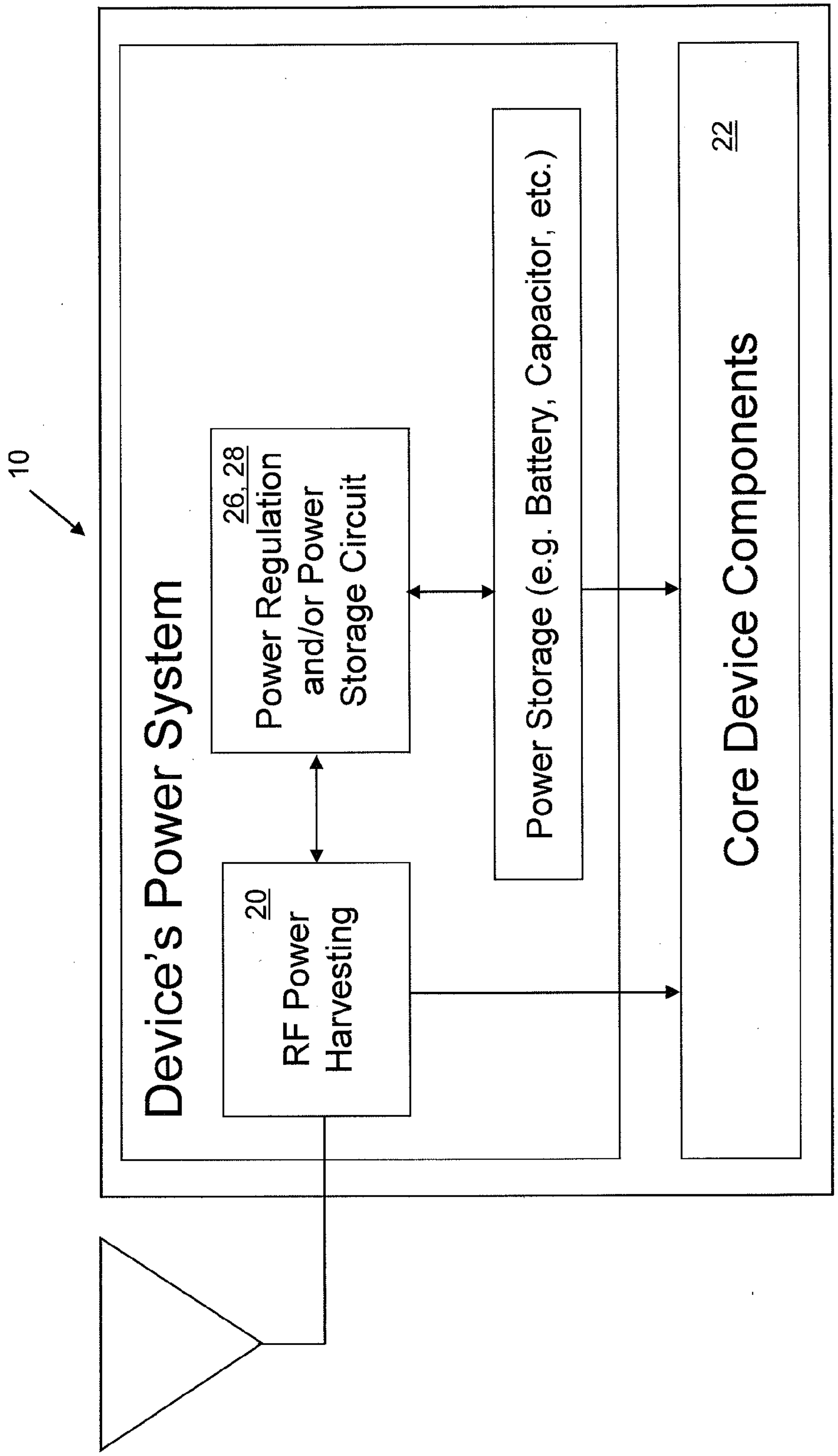
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 5

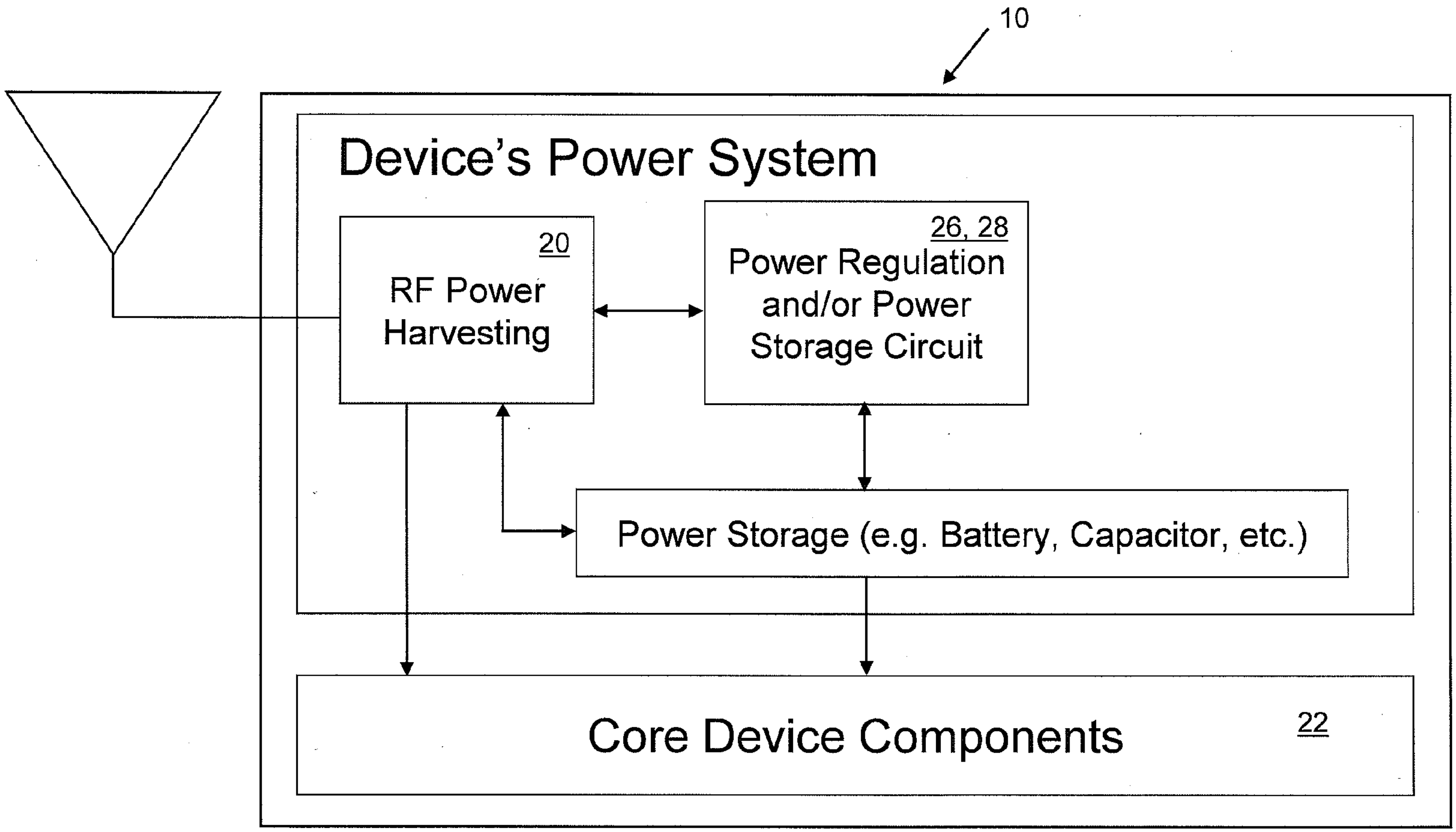
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 6

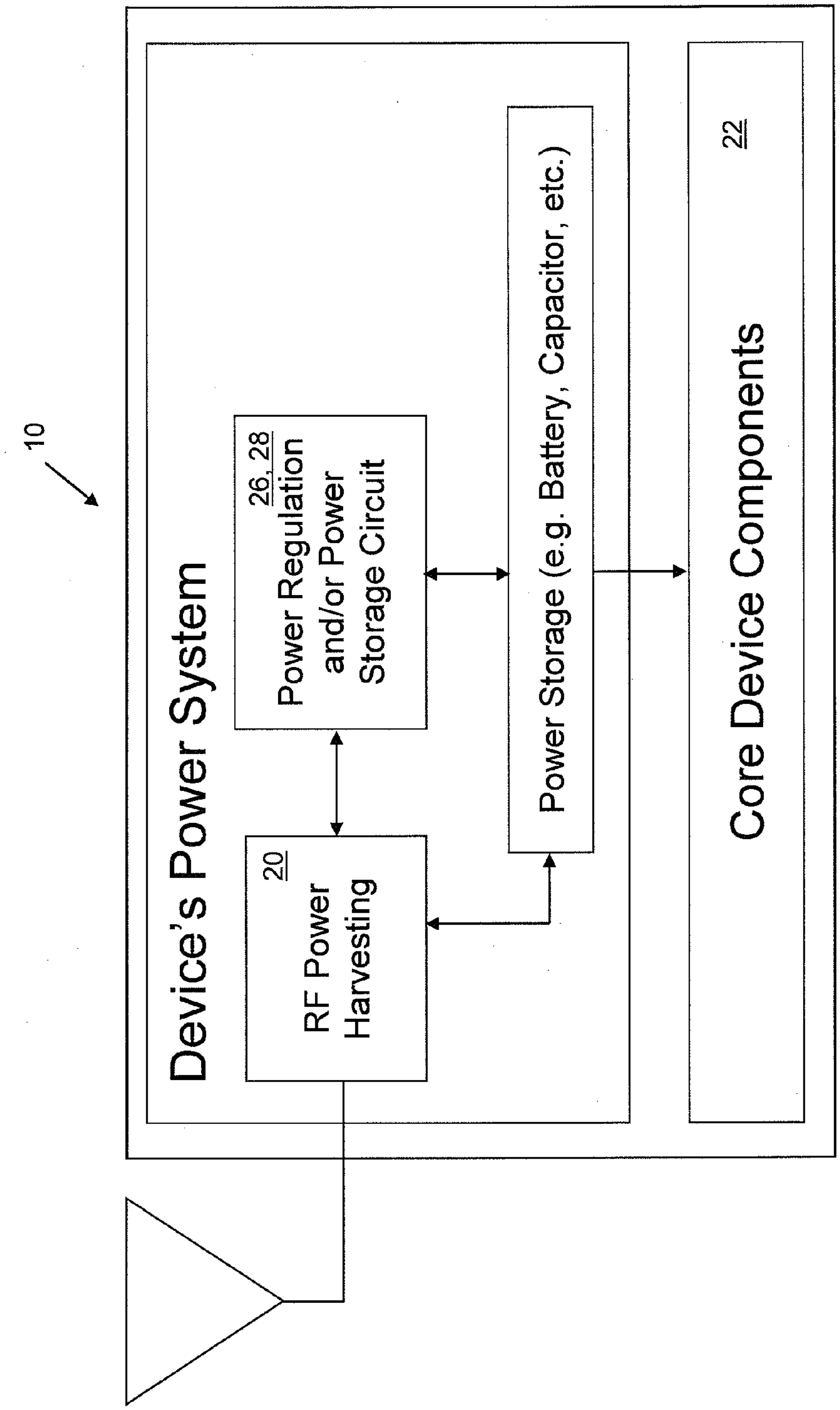
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 7

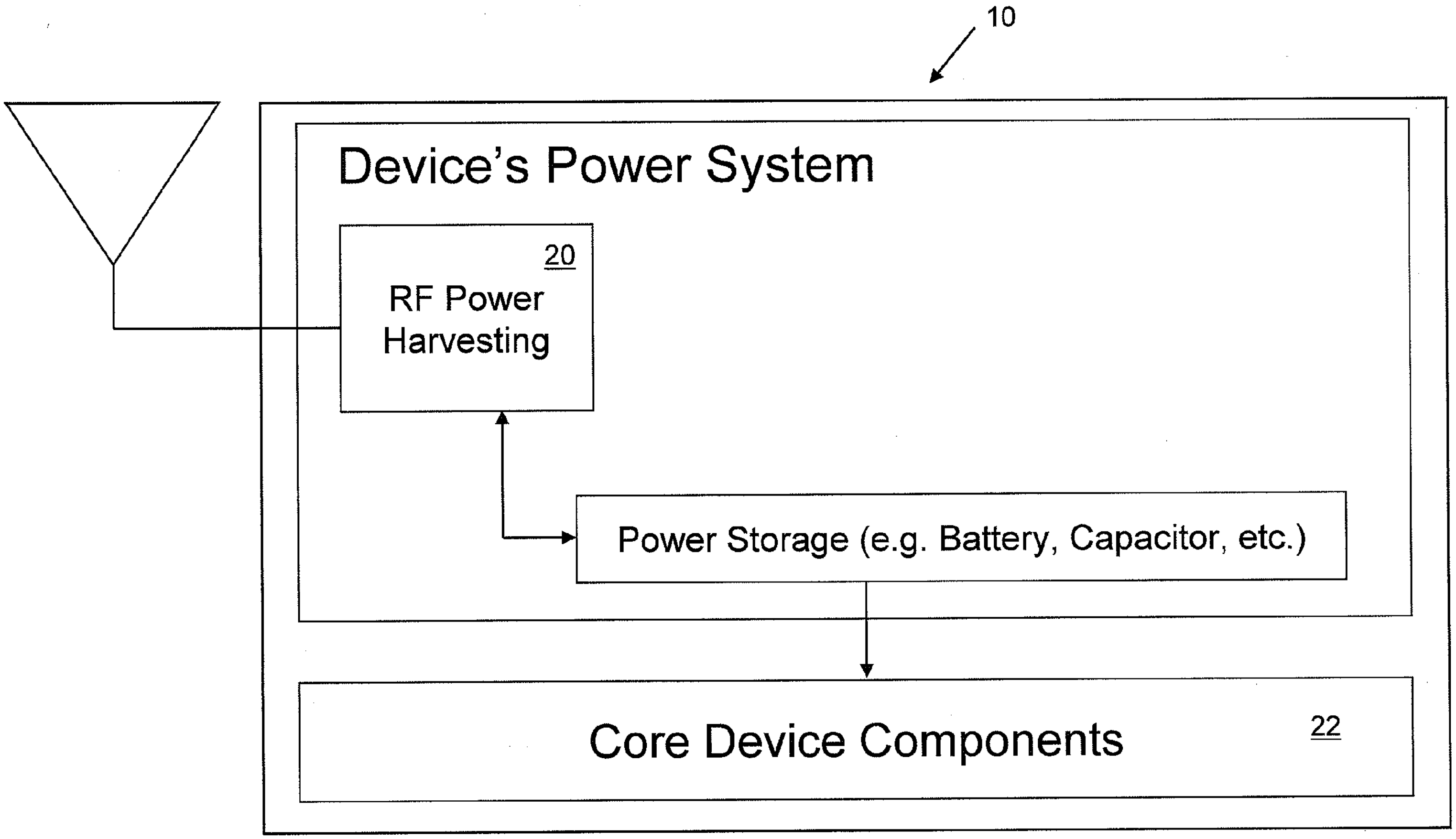
10
Device's Power System
20
RF Power Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 8

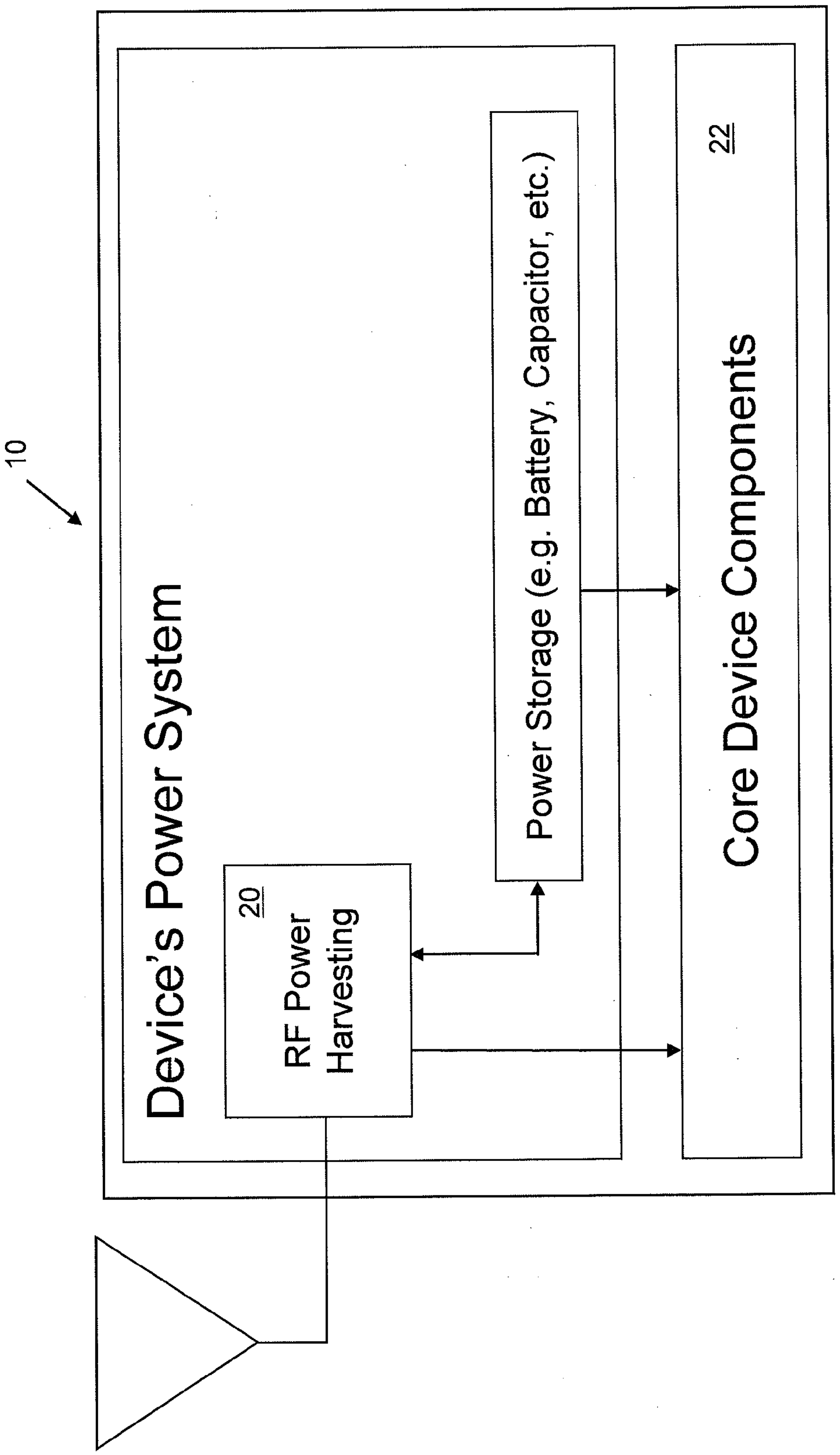
10
Device's Power System
20
RF Power Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 9

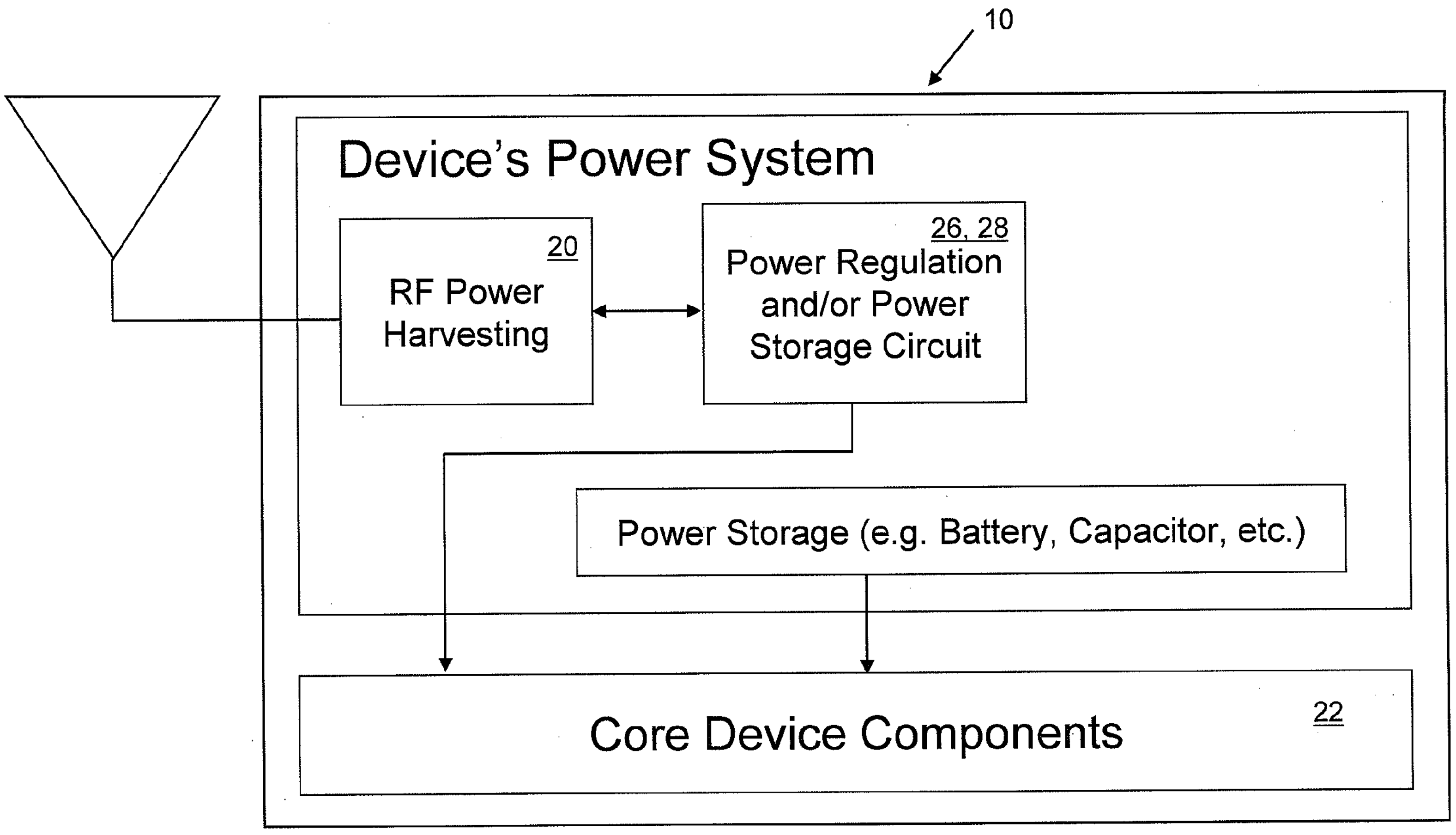
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 10

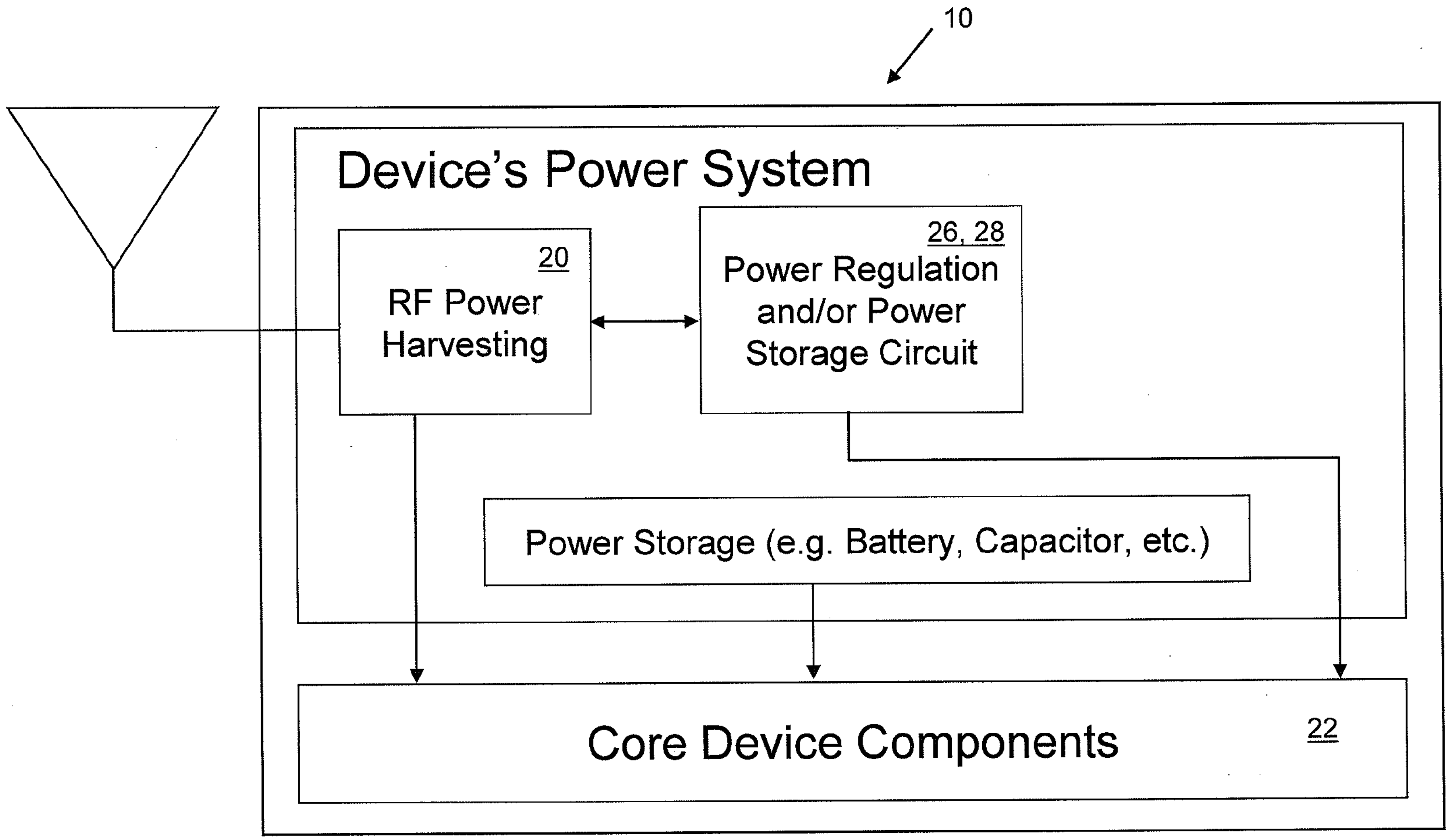
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 11

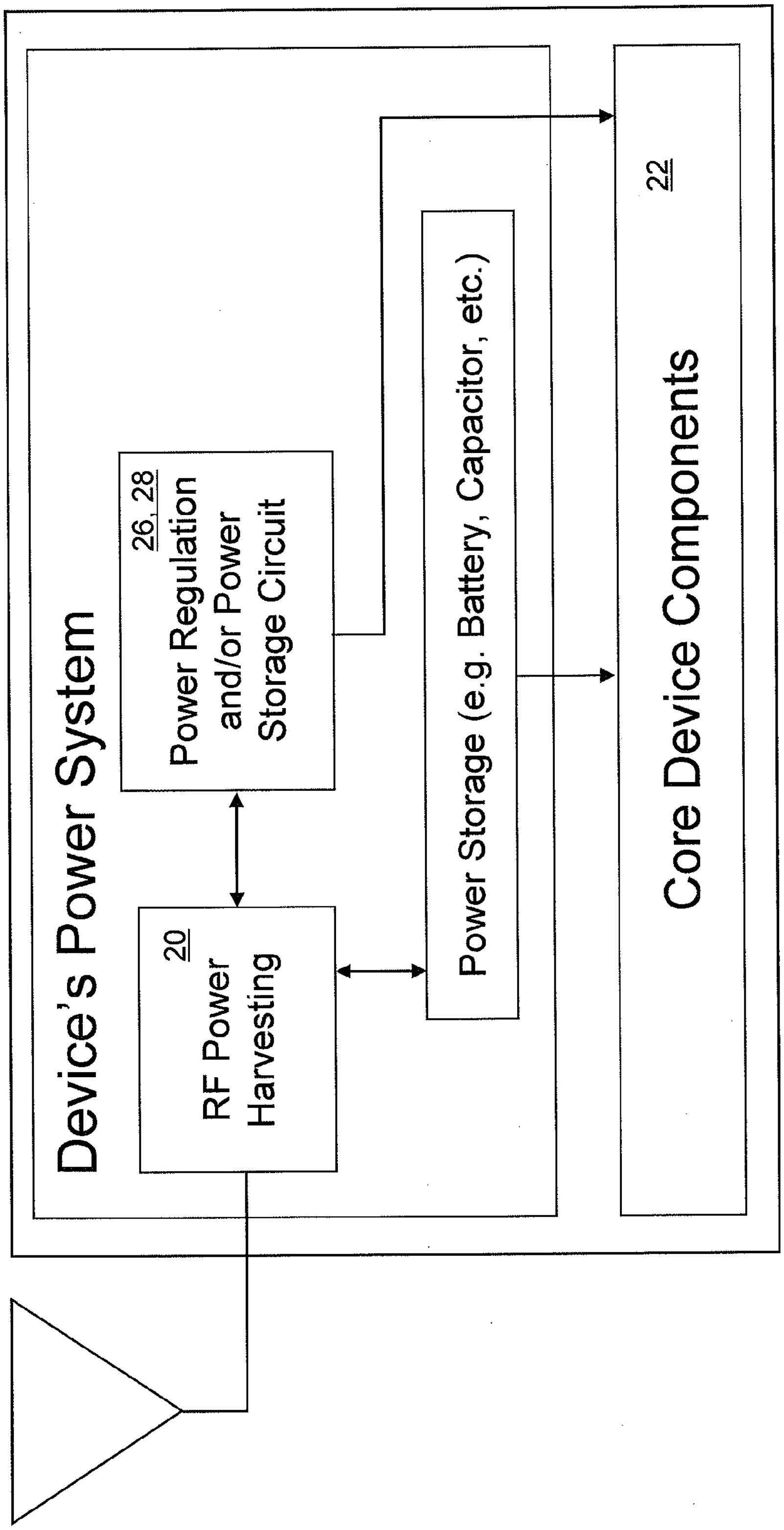
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 12

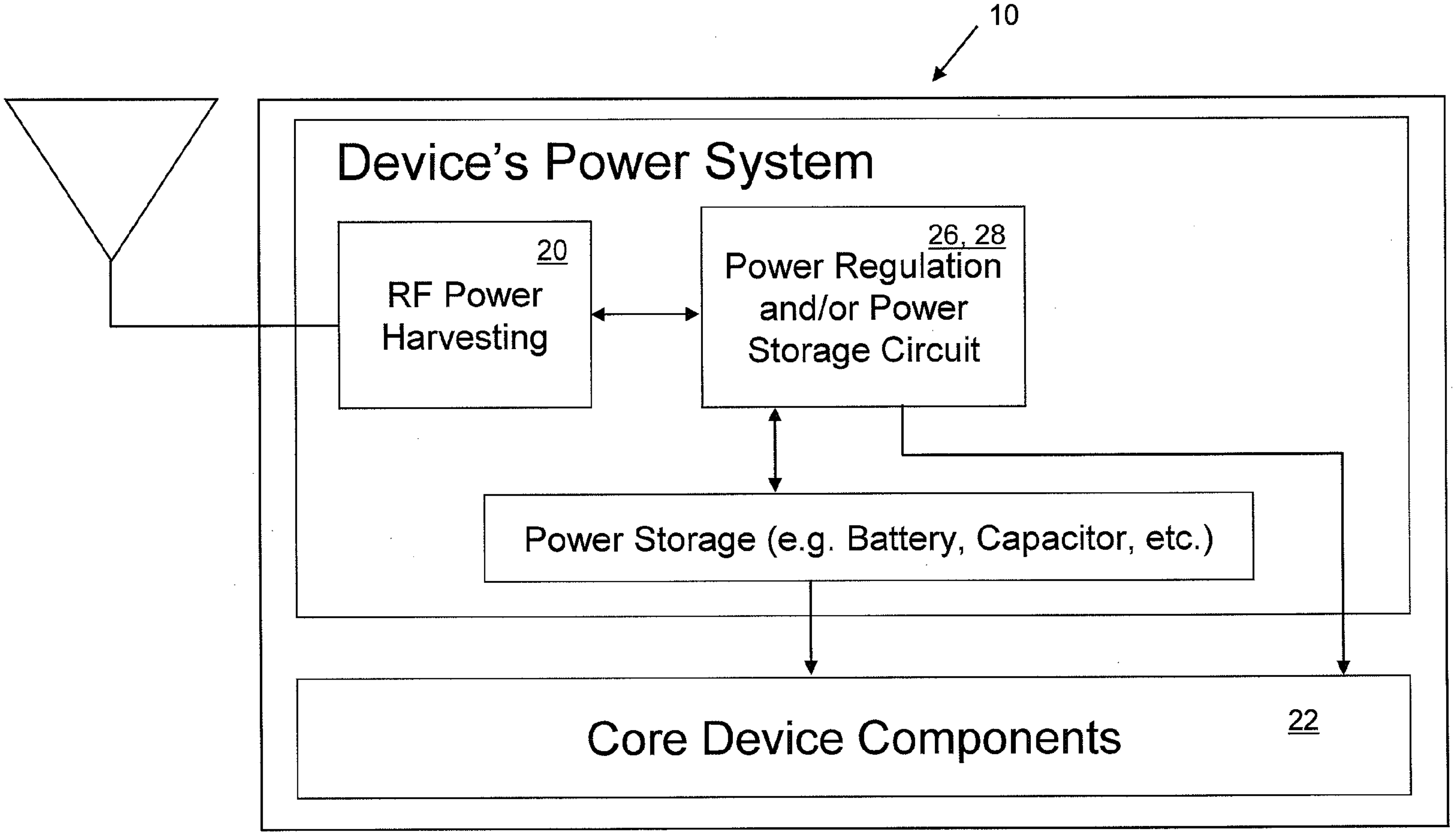
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 13

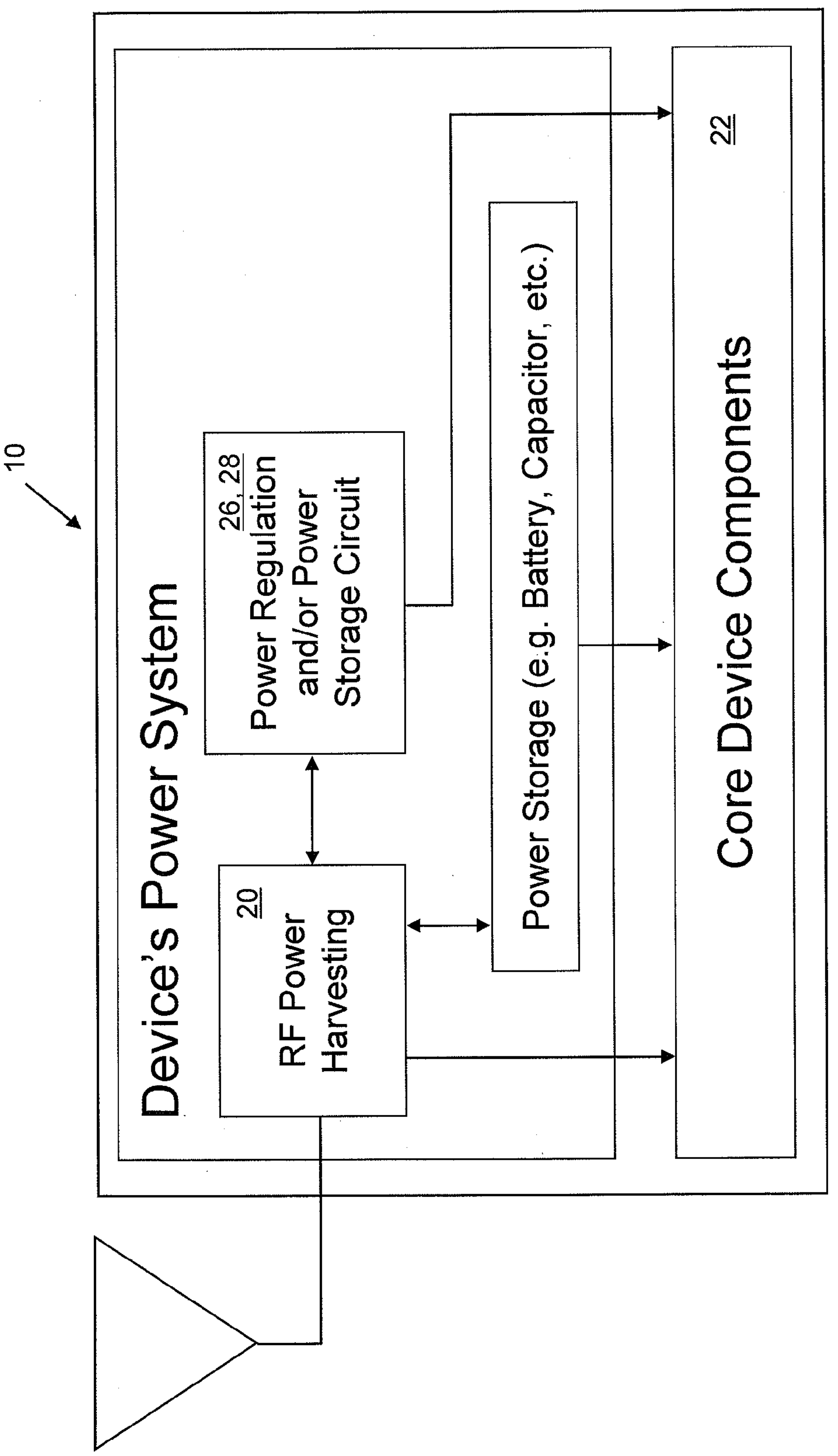
10
Device's Power System
Power Regulation and/or Power Storage Circuit
26, 28
20
RF Power Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 14

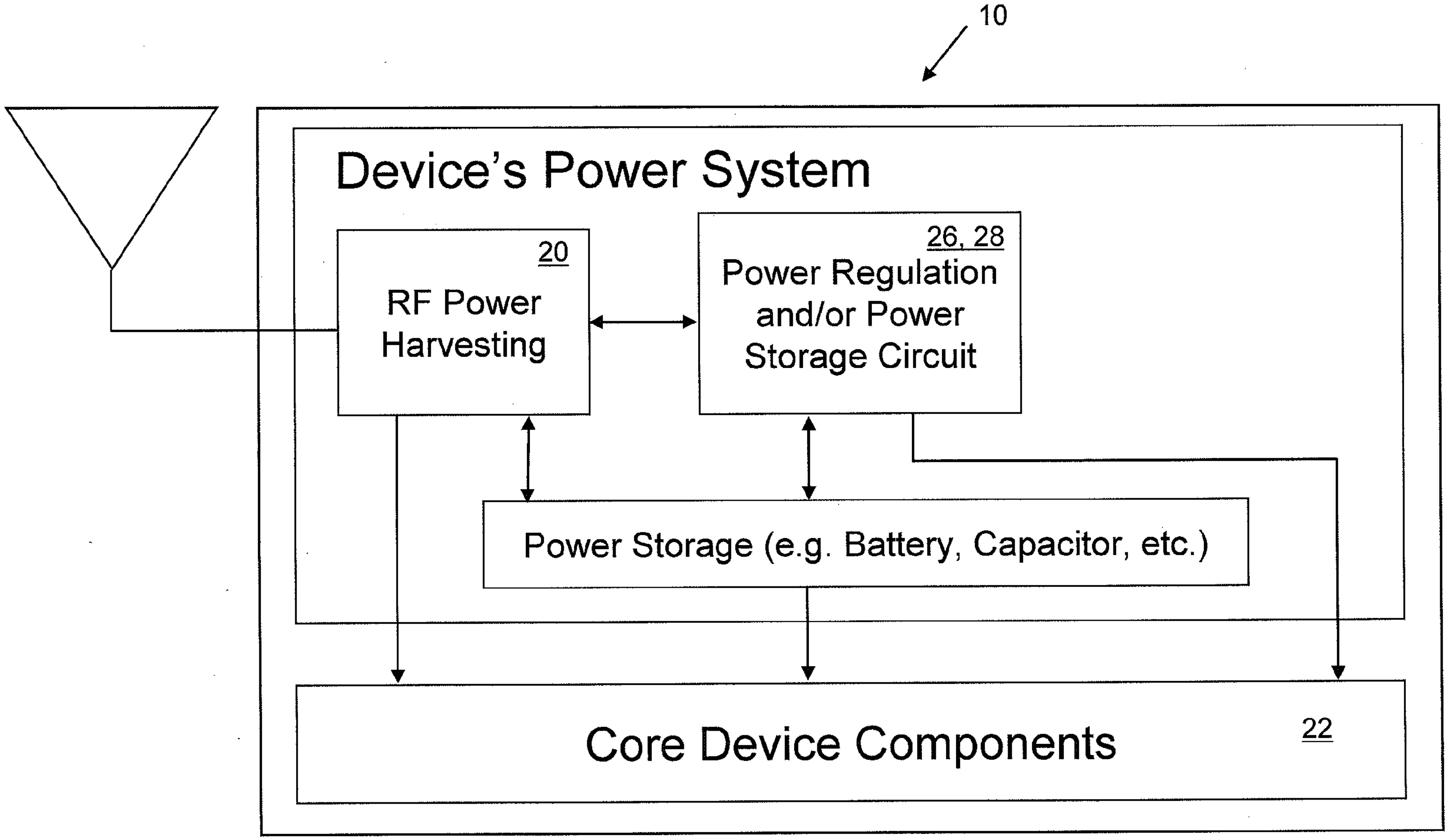
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 15

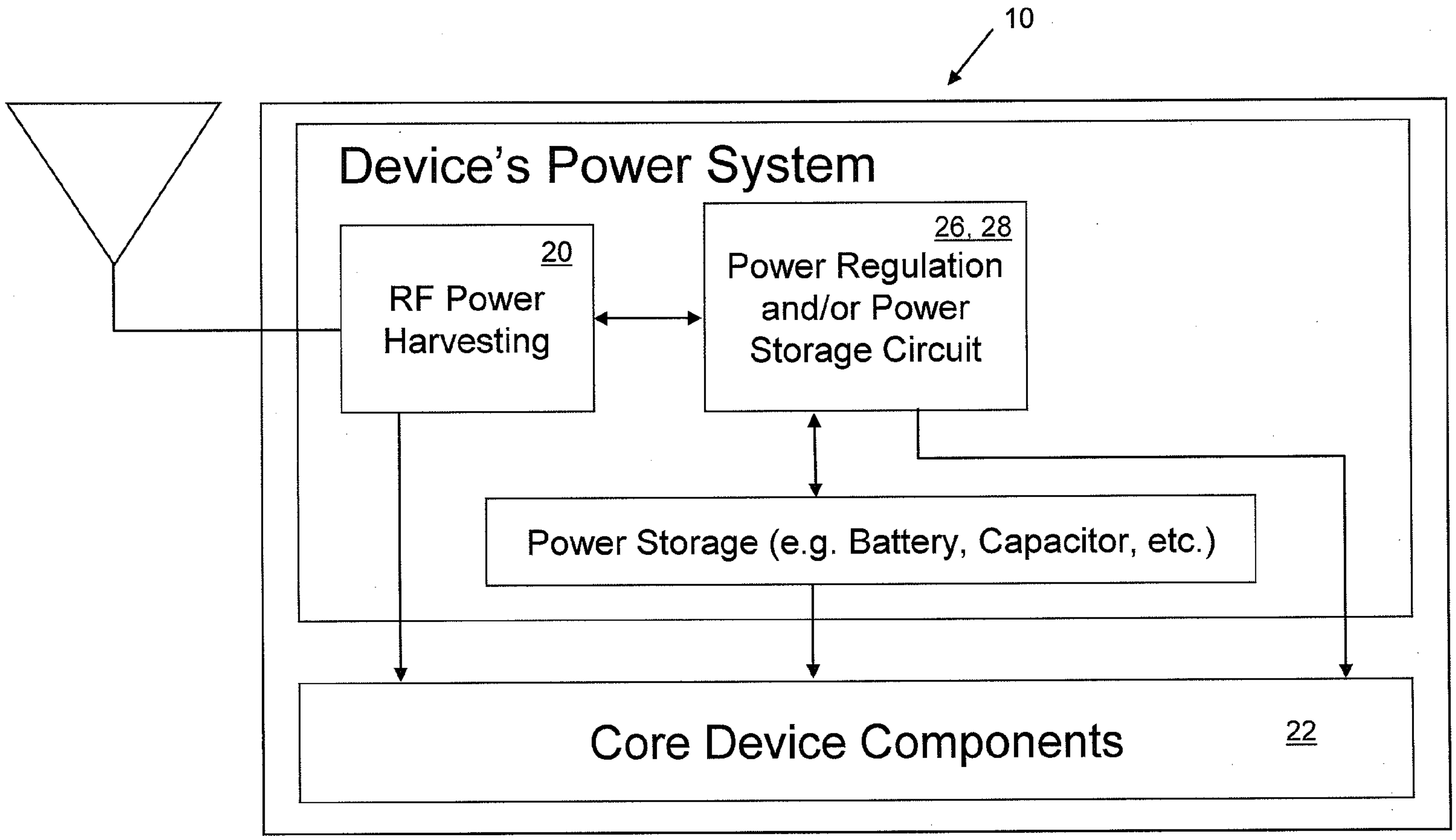
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 16

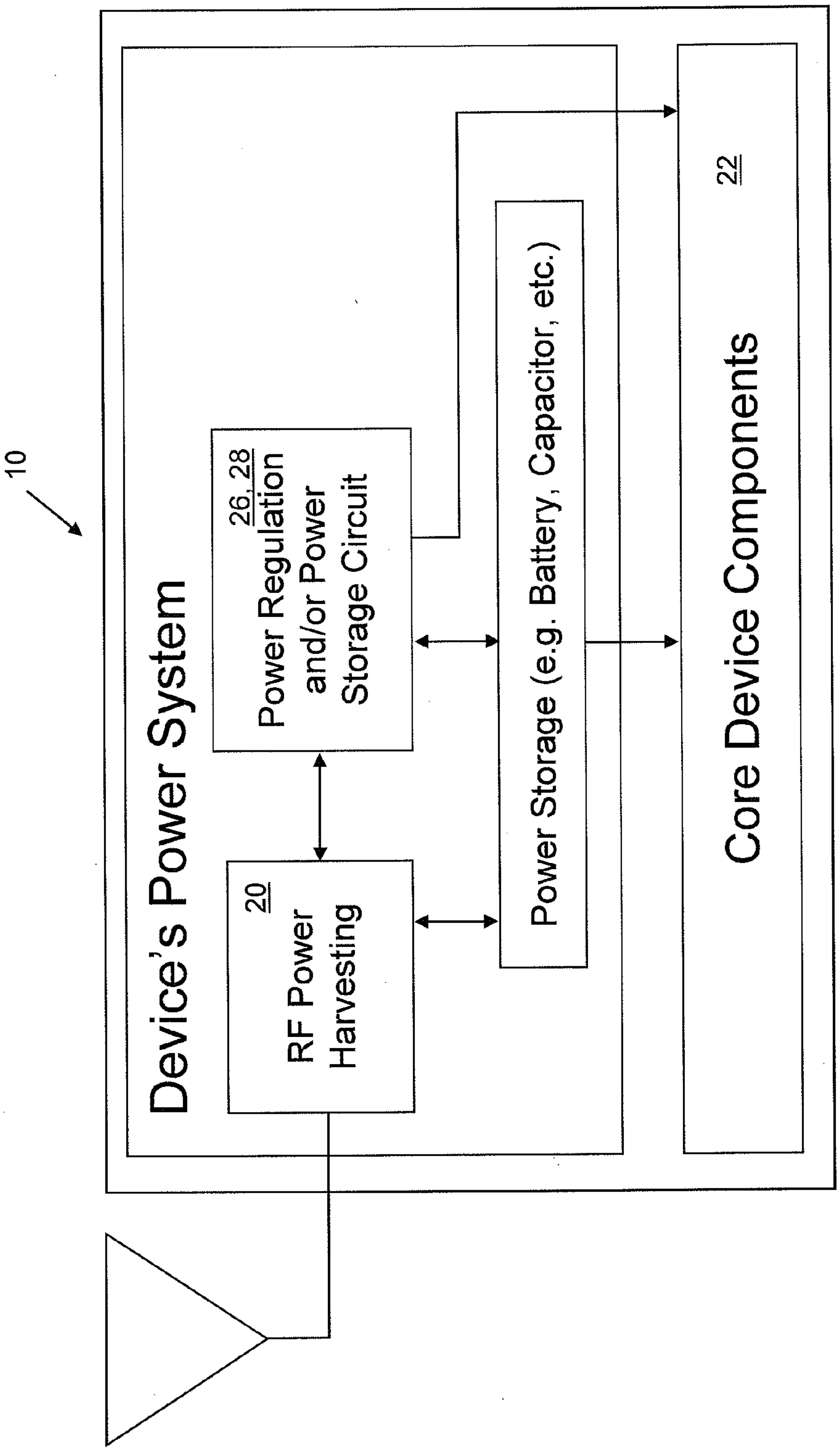
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 17

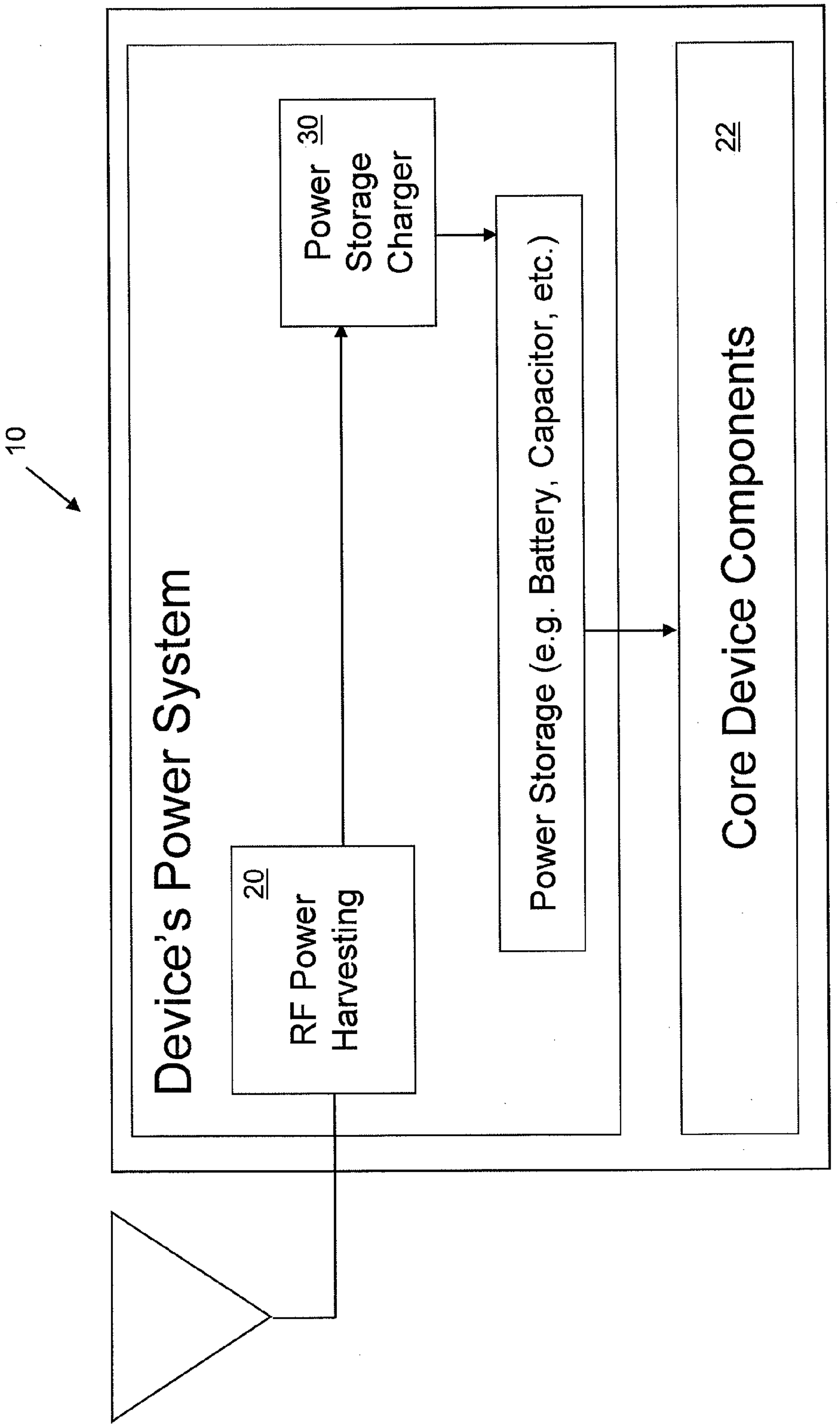
10
Device's Power System
20
RF Power Harvesting
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 18

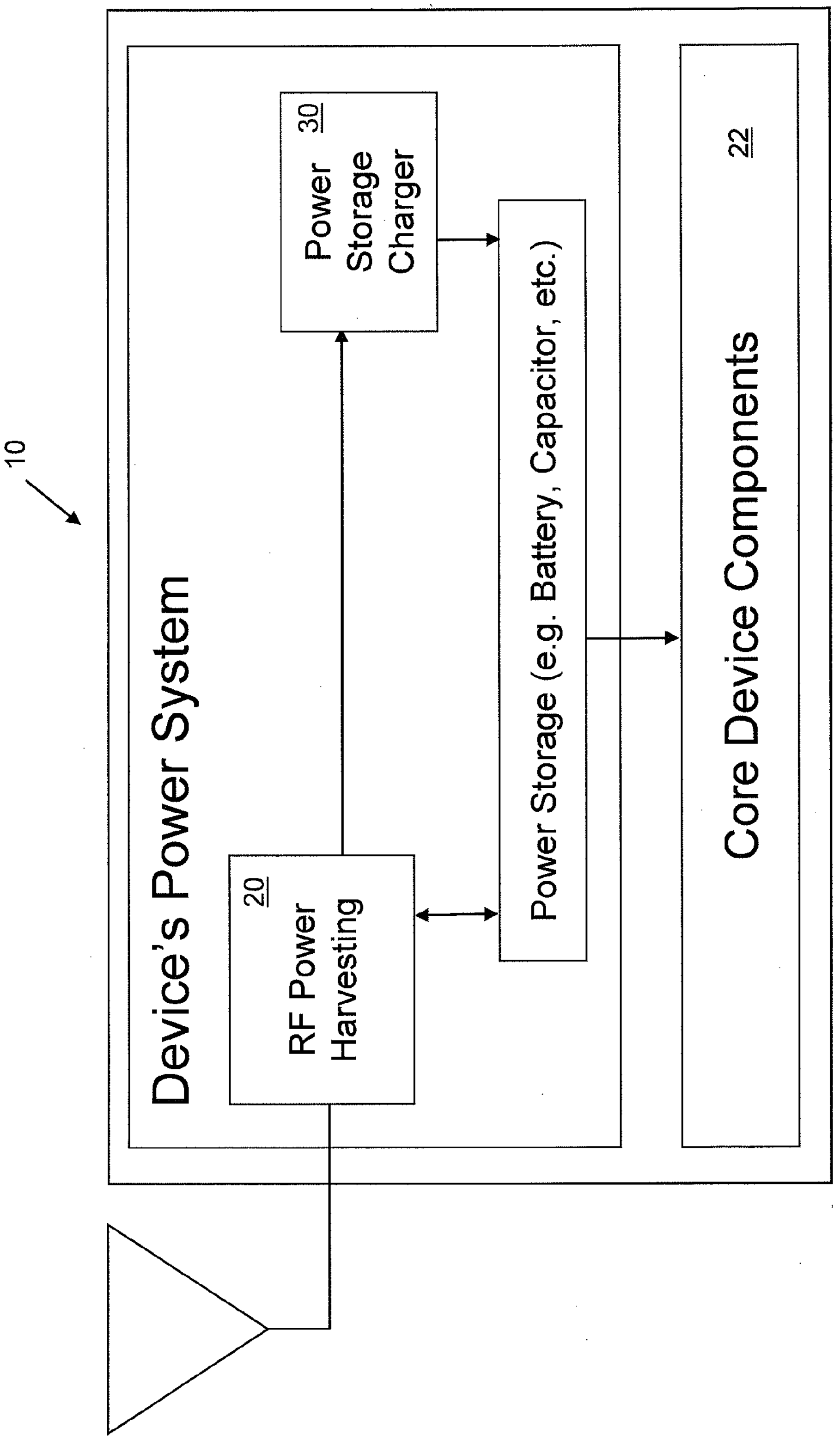
10
Device's Power System
20
RF Power Harvesting
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 19

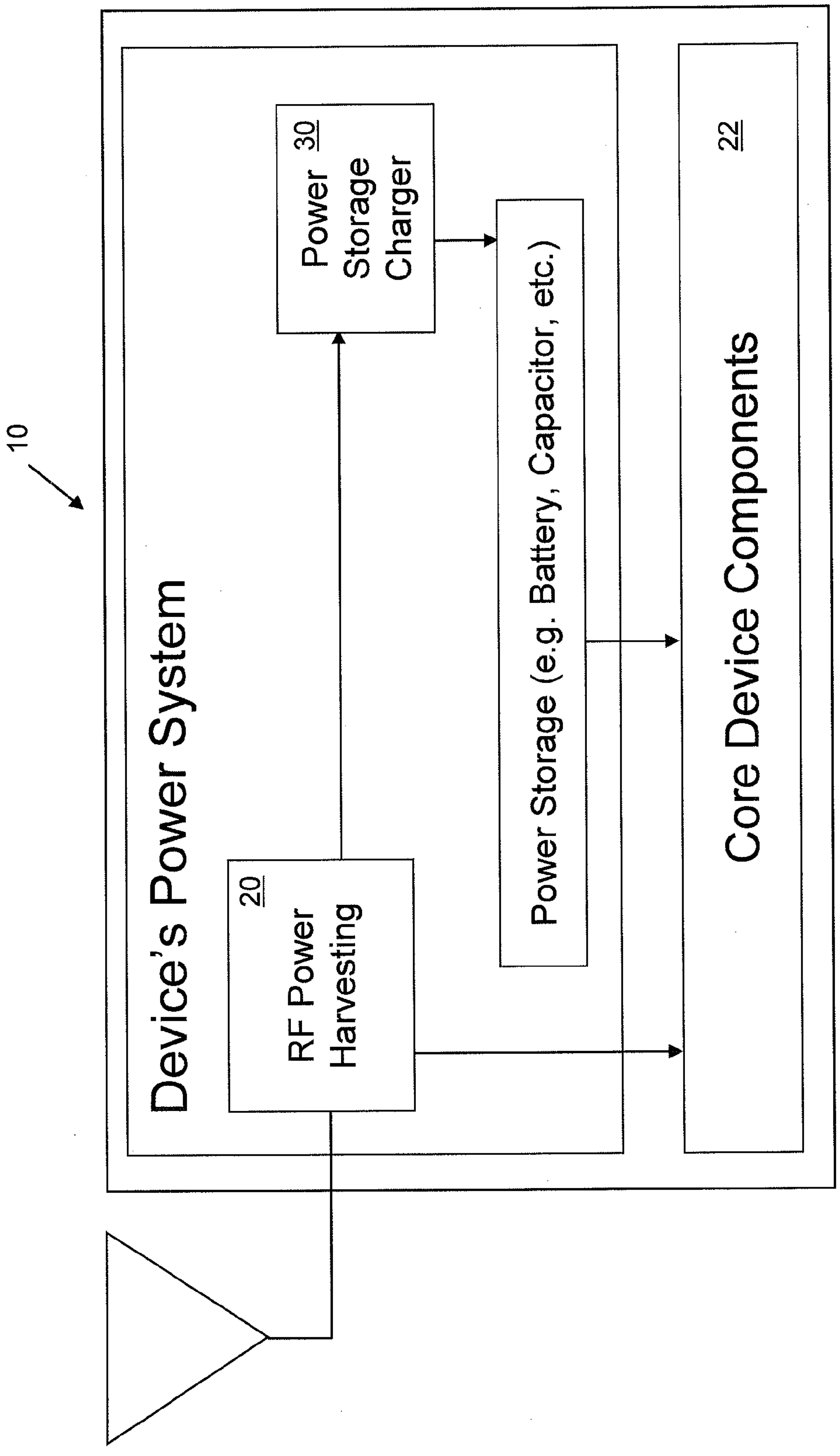
10
Device's Power System
20
RF Power
Harvesting
Power 30
Storage
Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 20

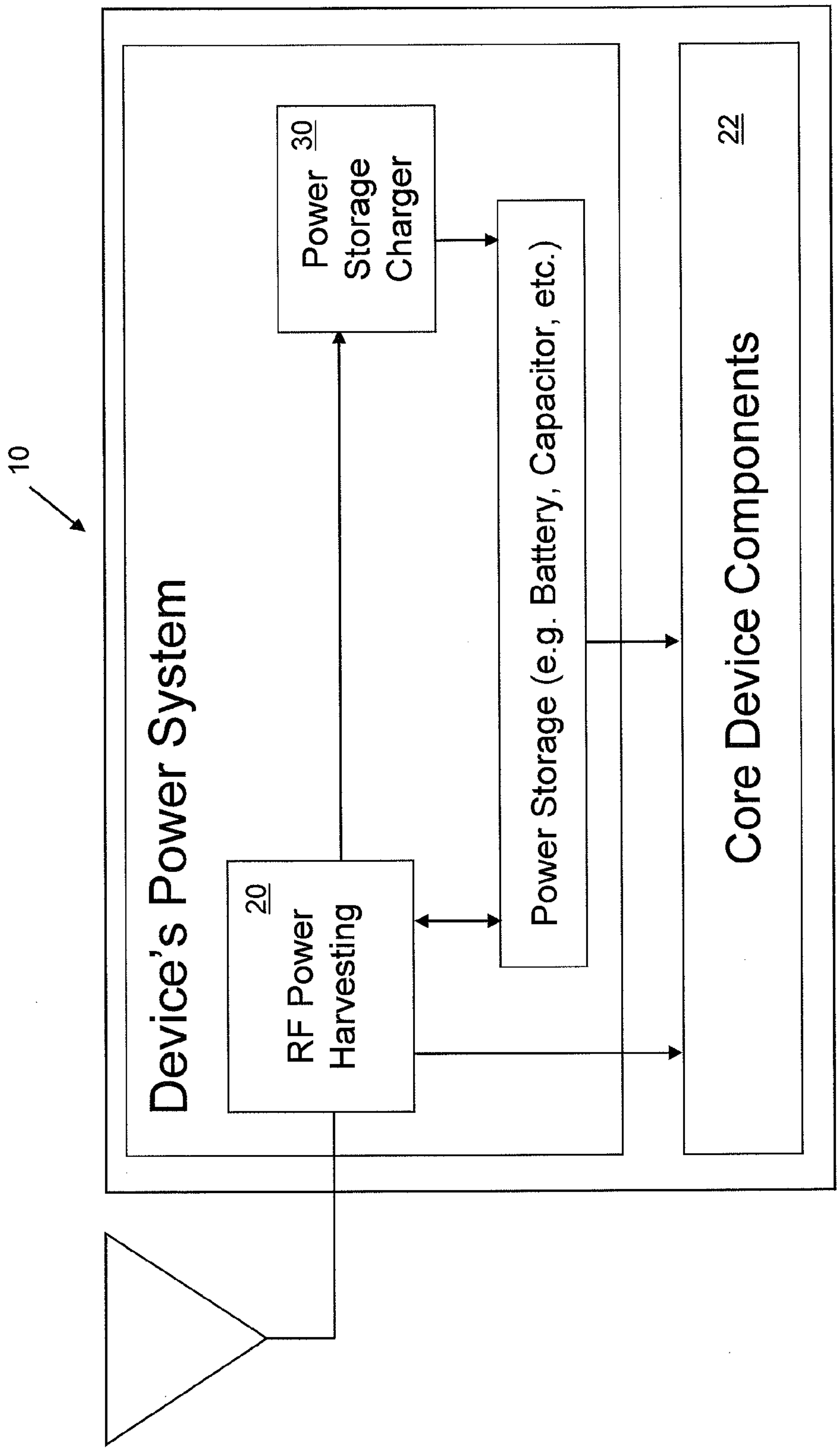
10
Device's Power System
20
RF Power Harvesting
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 21

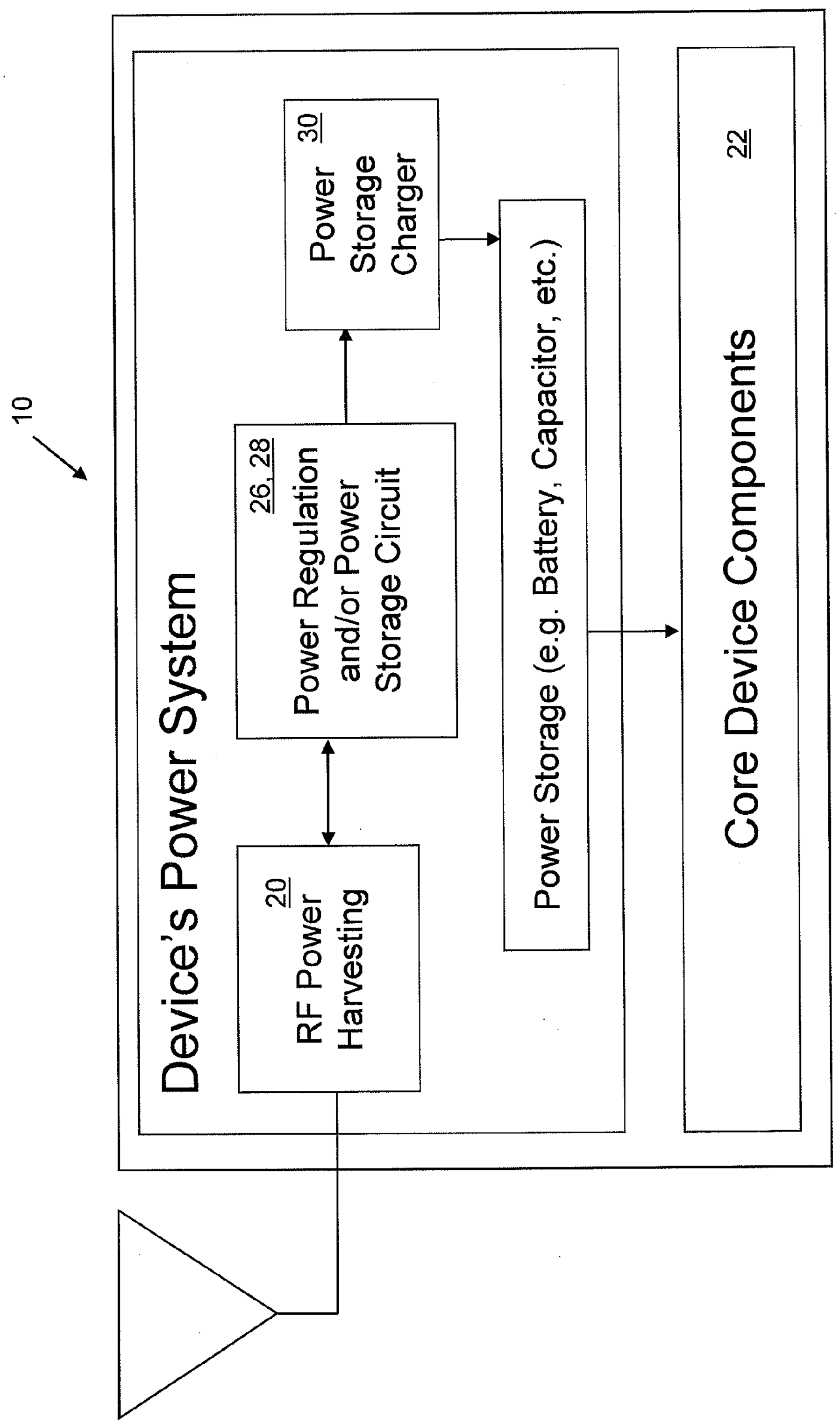
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 22

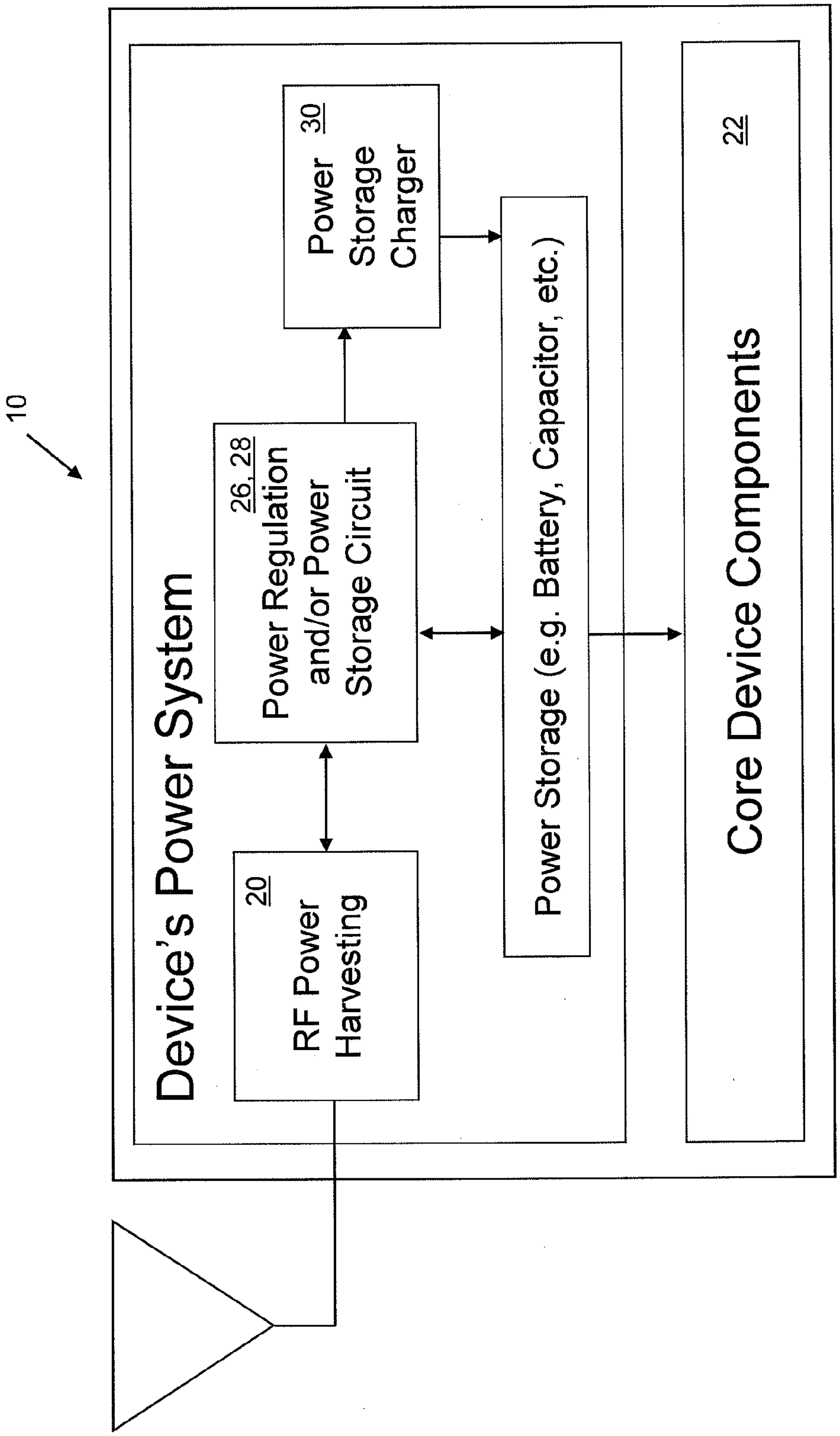
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 23

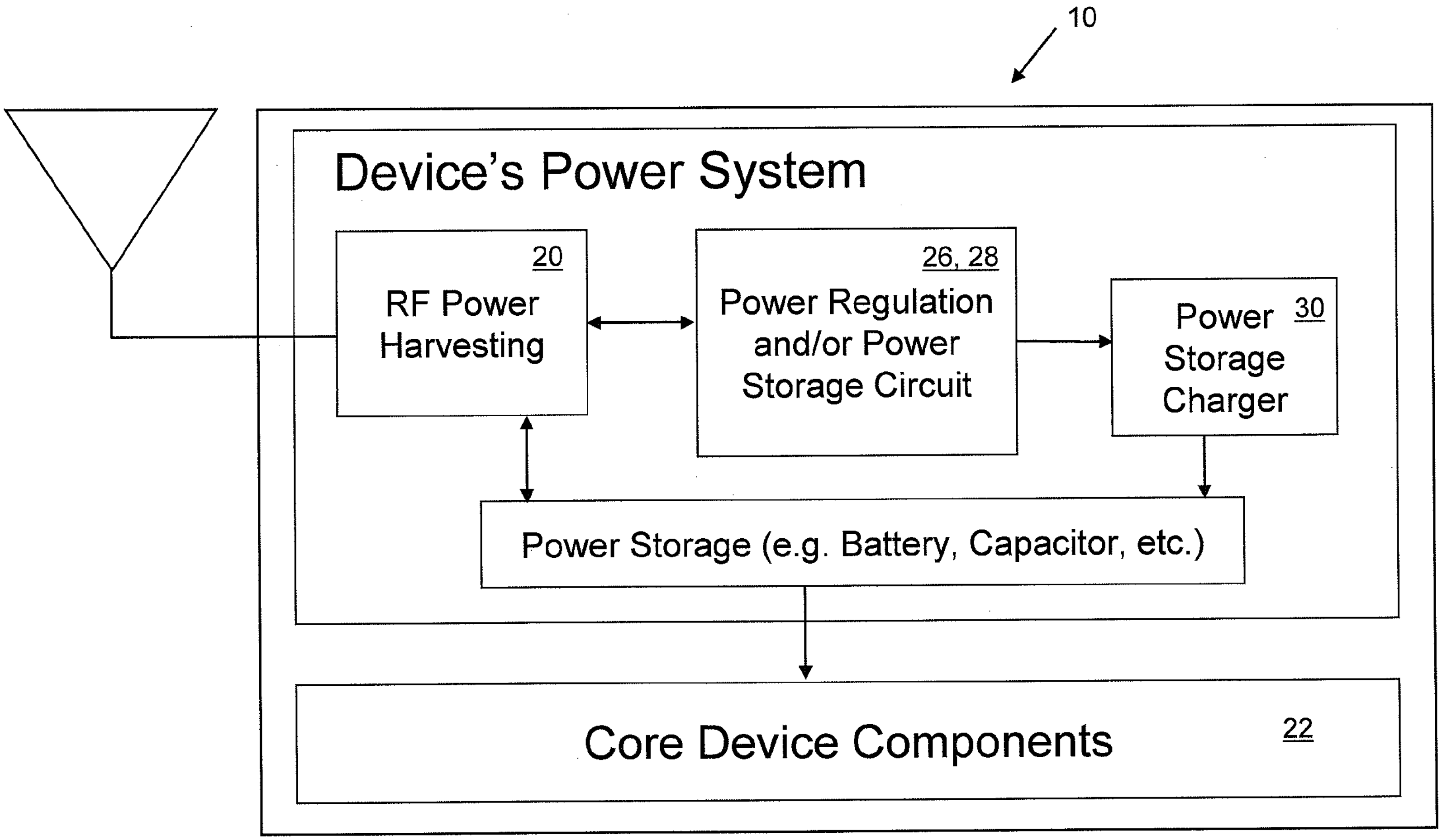
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage
Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 24

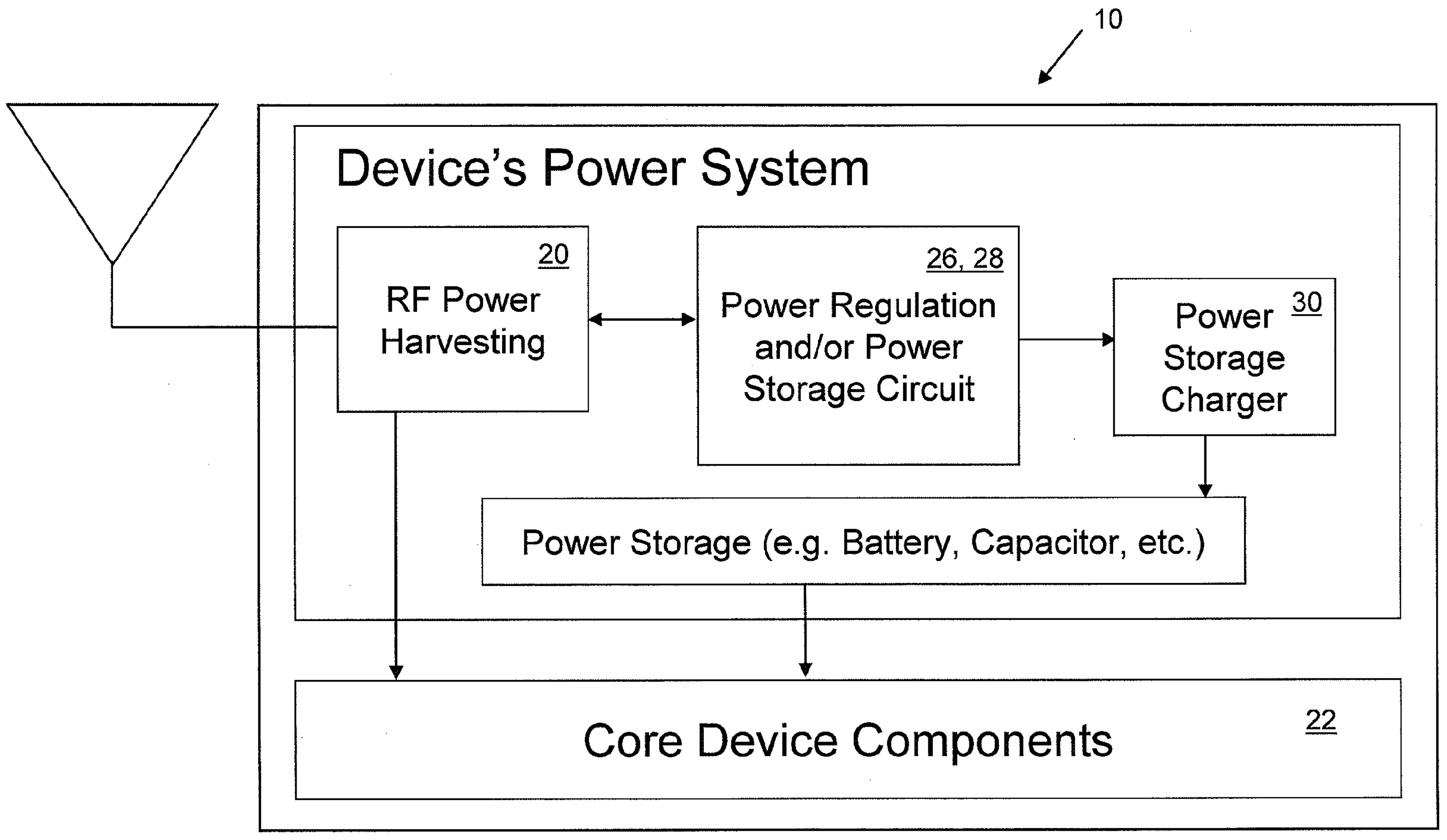
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 25

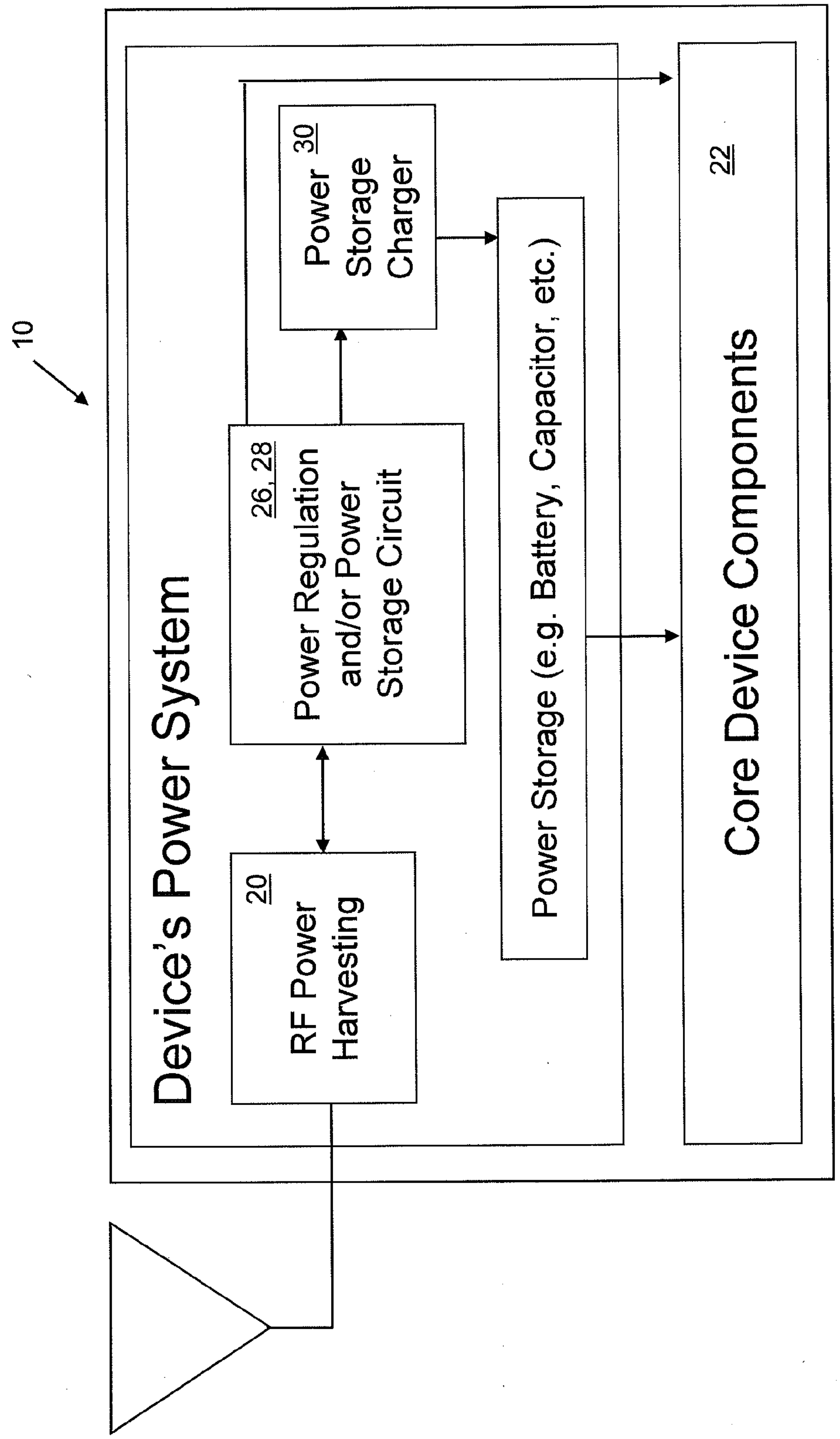
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 26

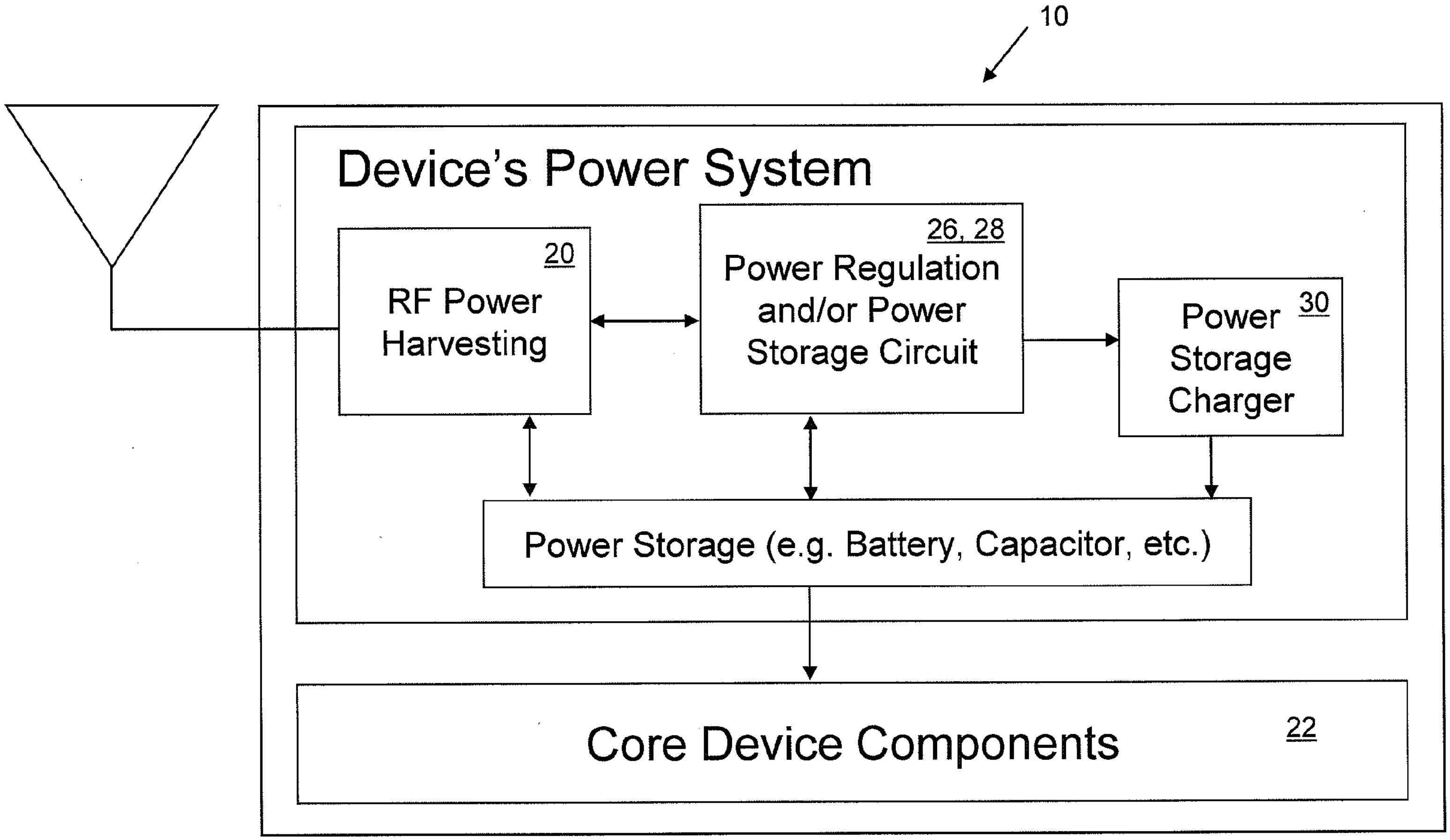
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 27

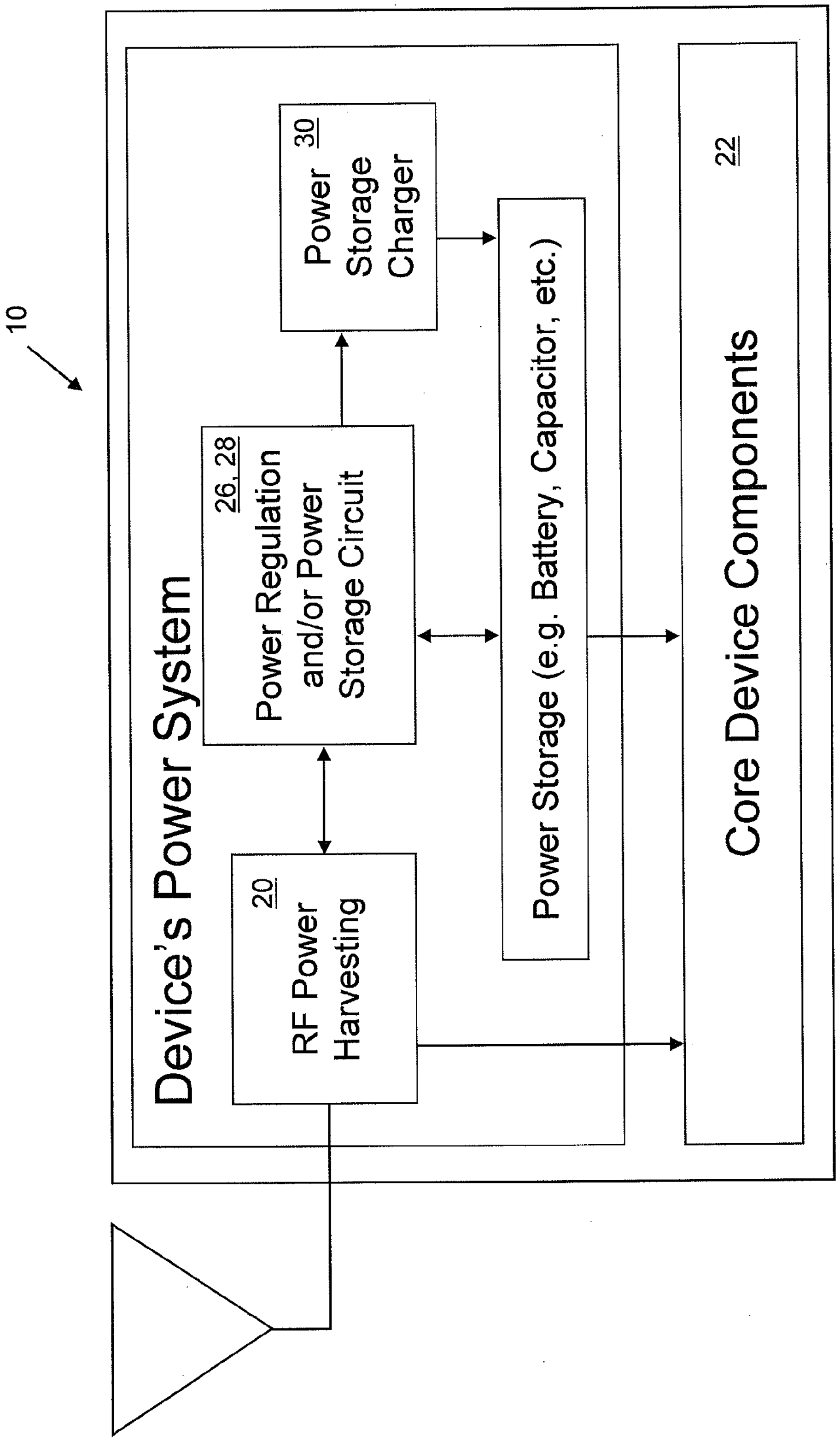
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 28

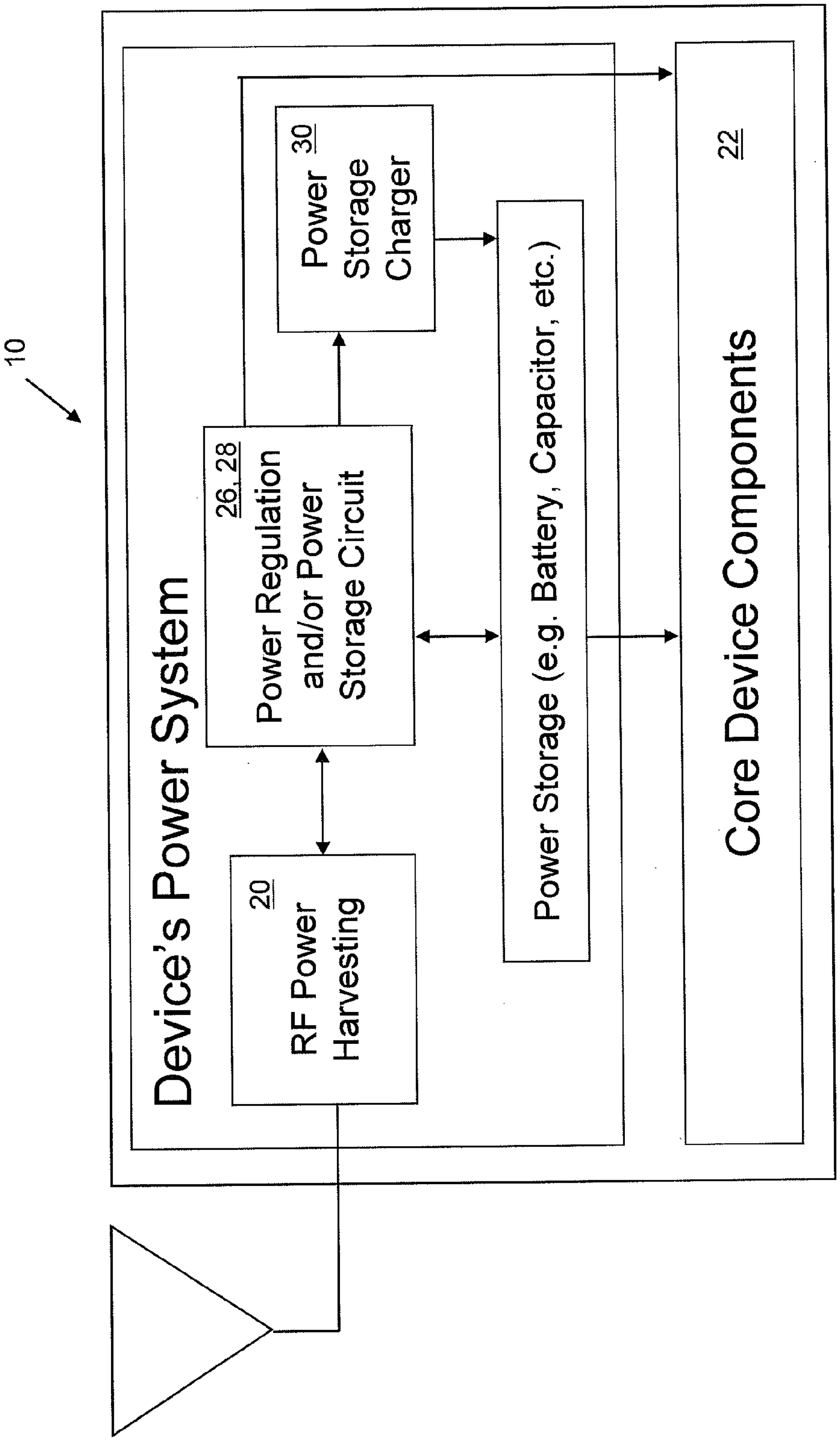
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 29

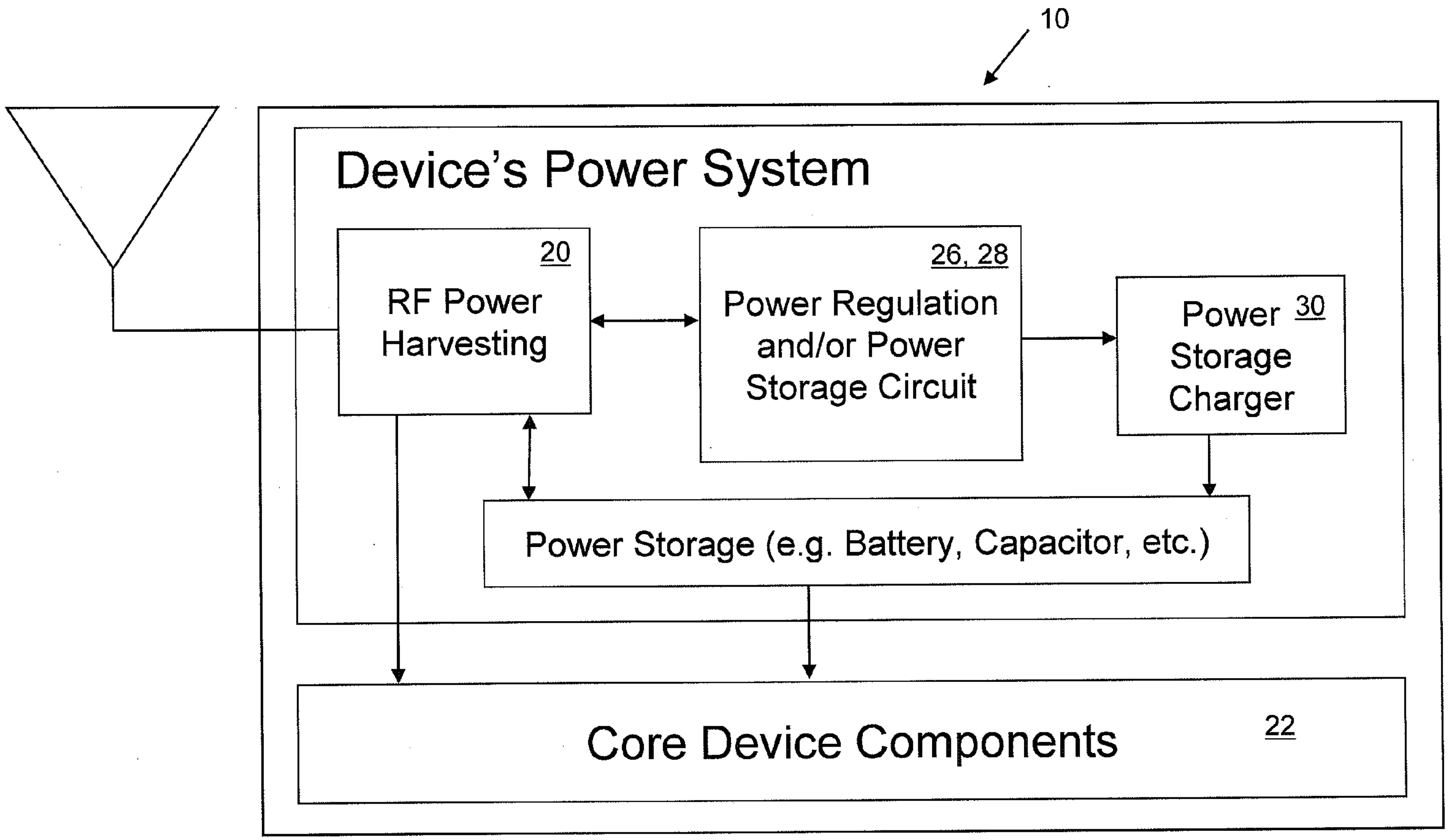
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30 Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 30

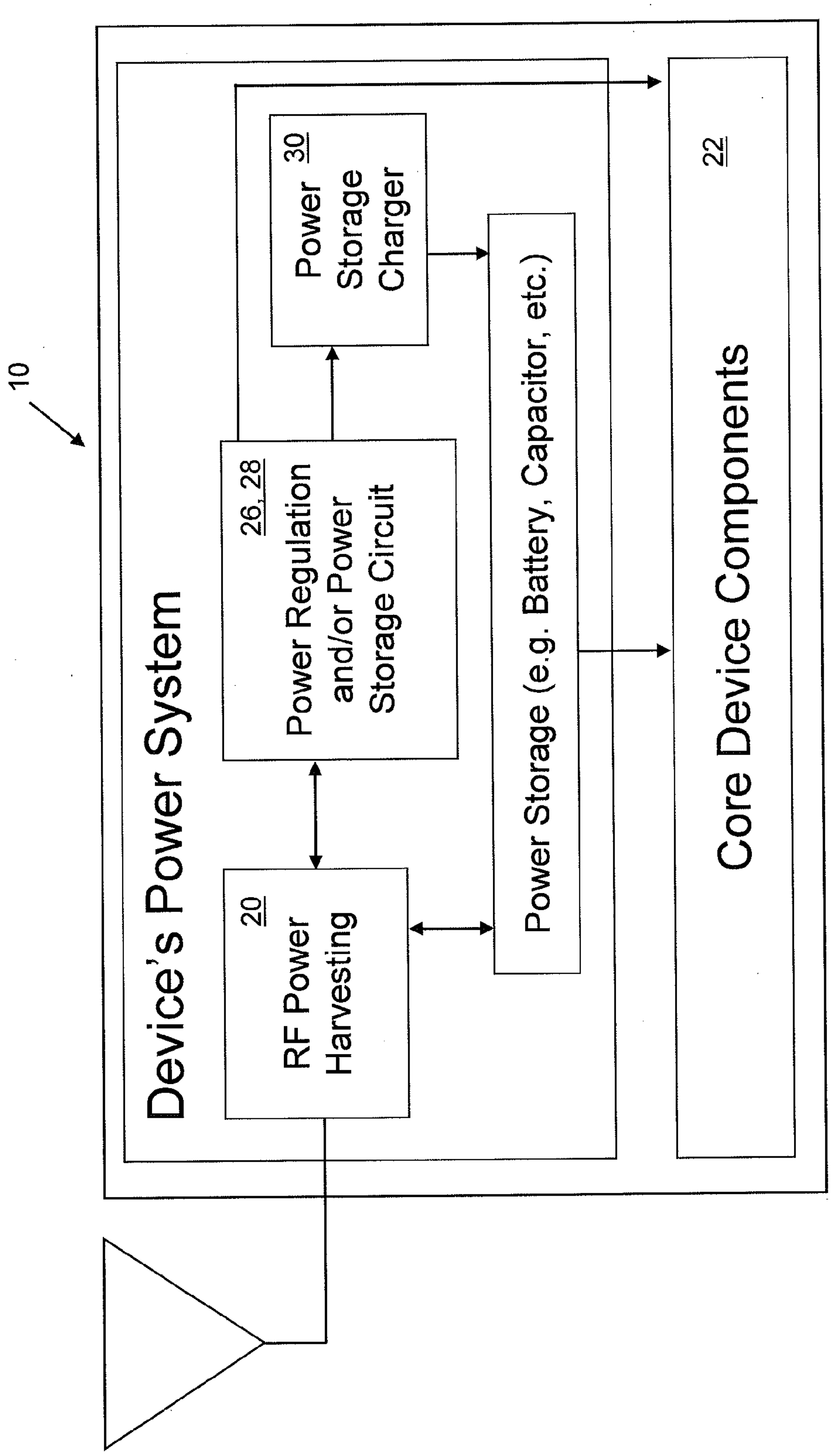
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 31

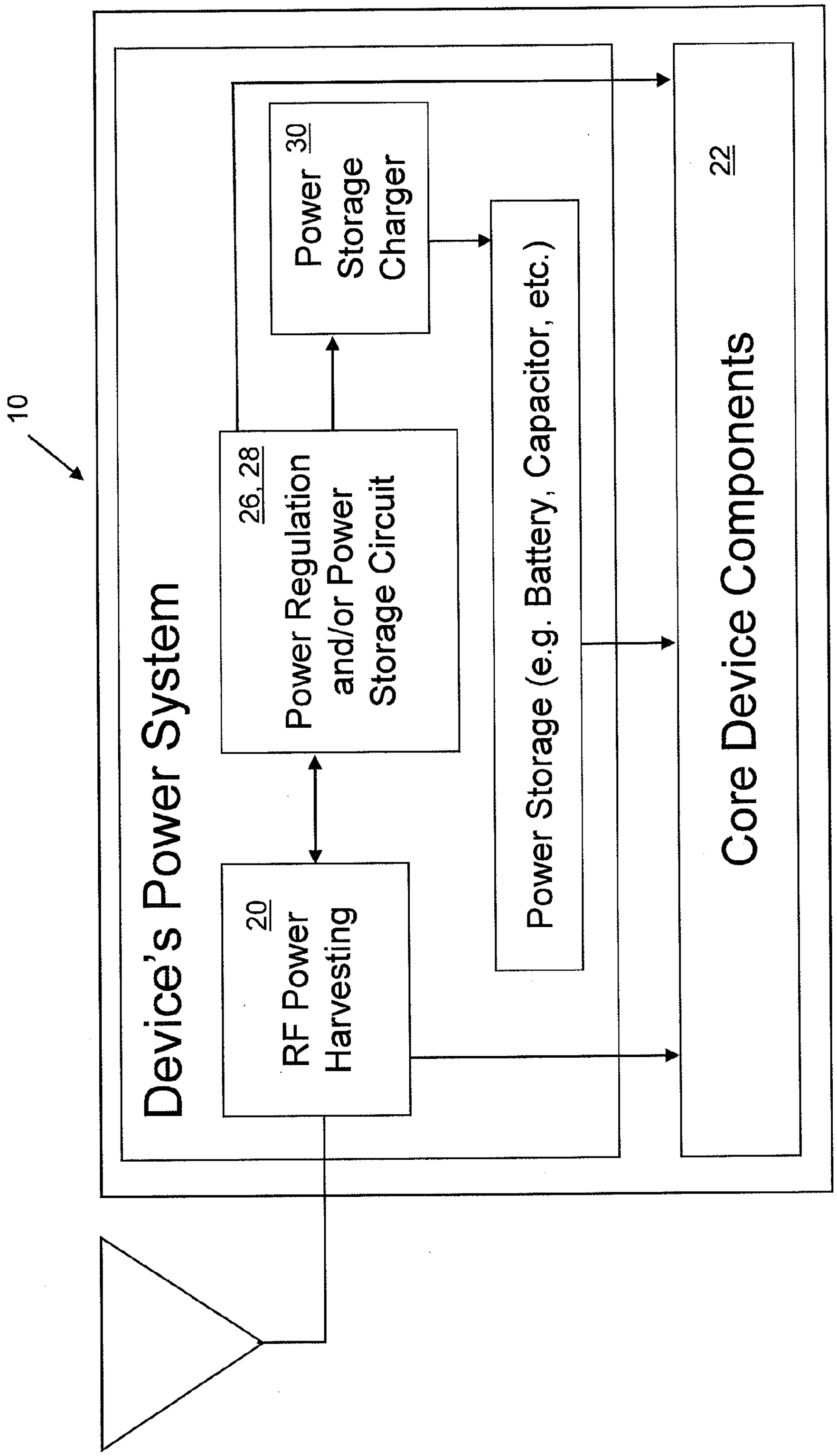
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 32

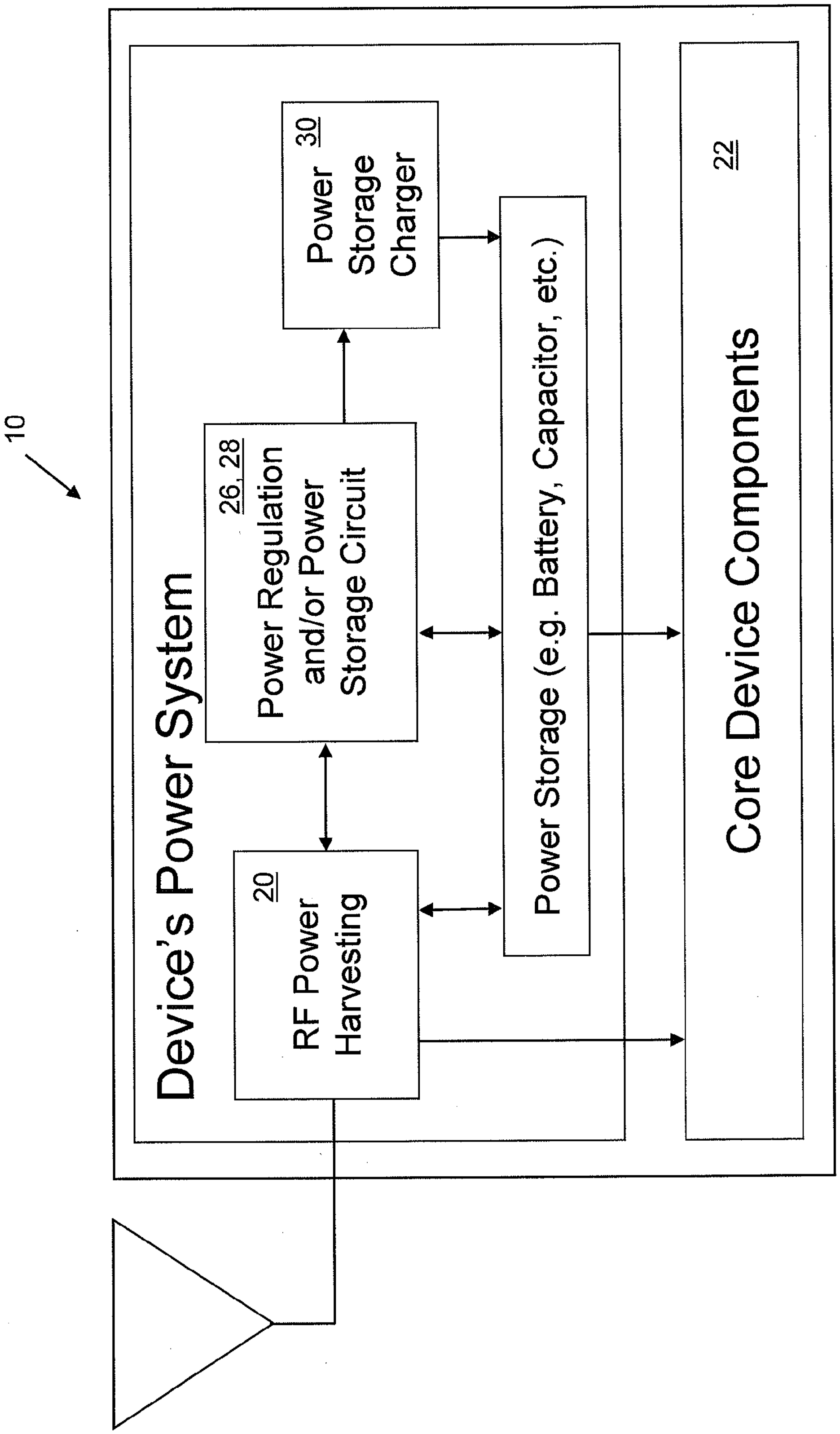
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 33

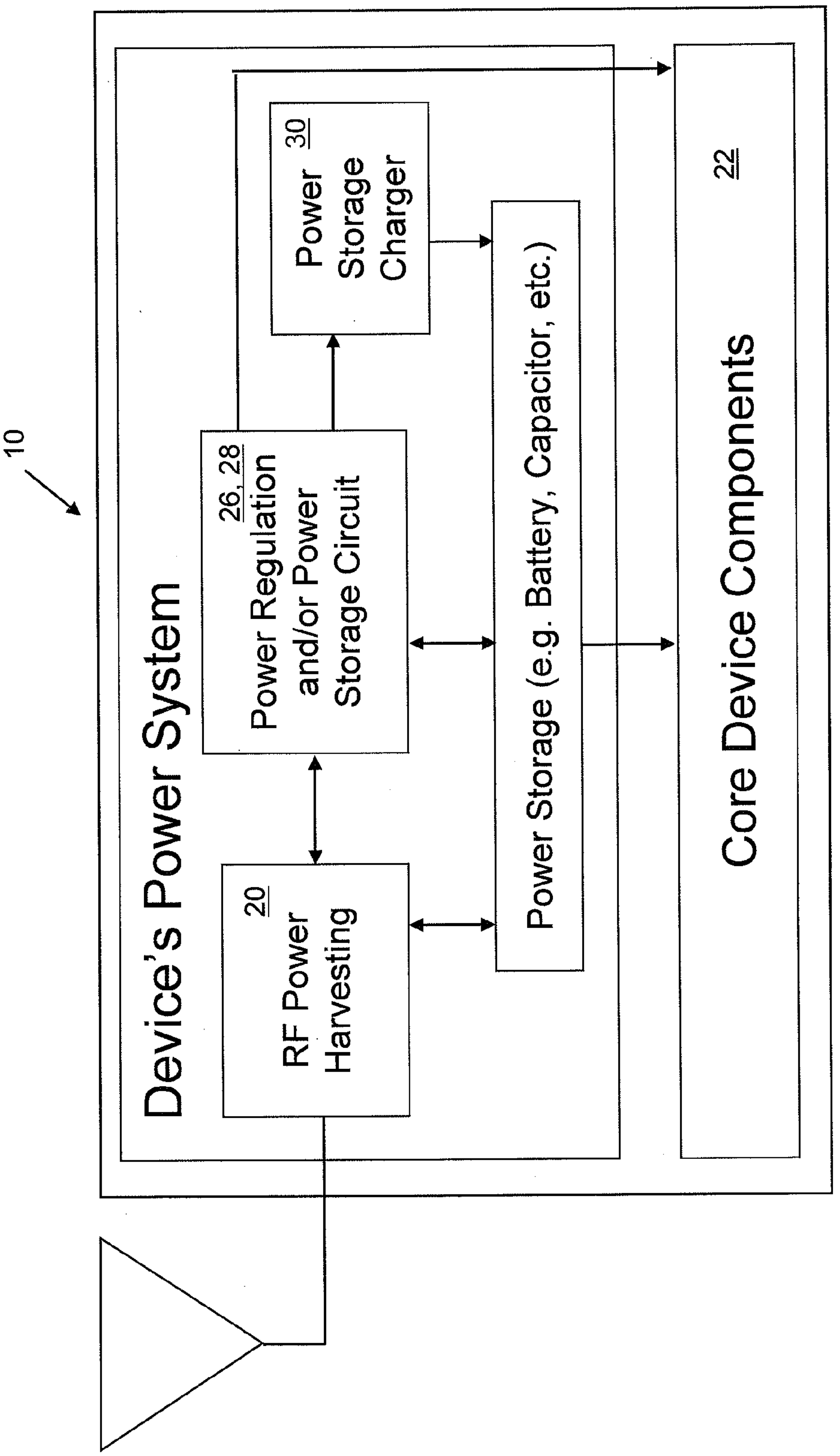
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 34

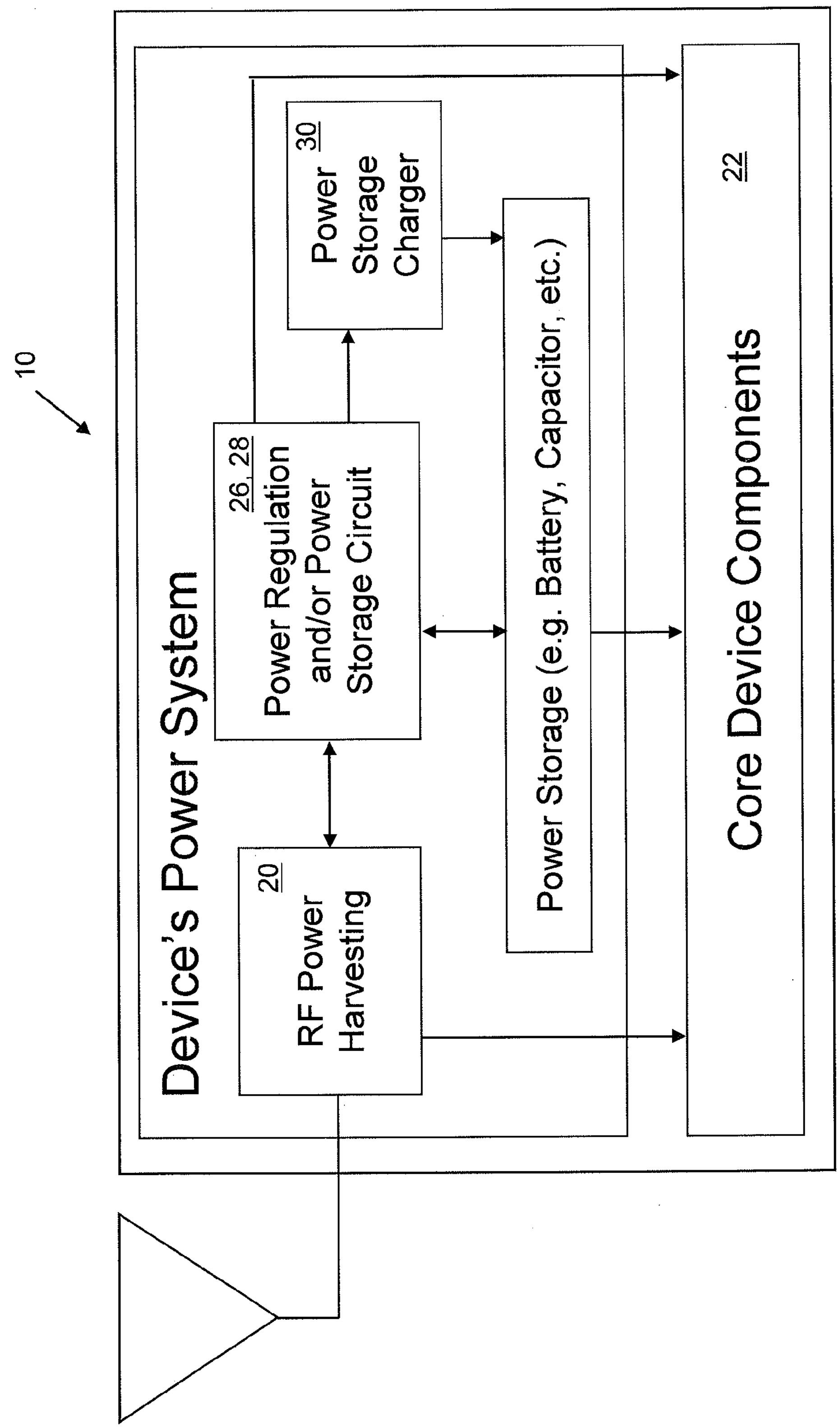
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 35

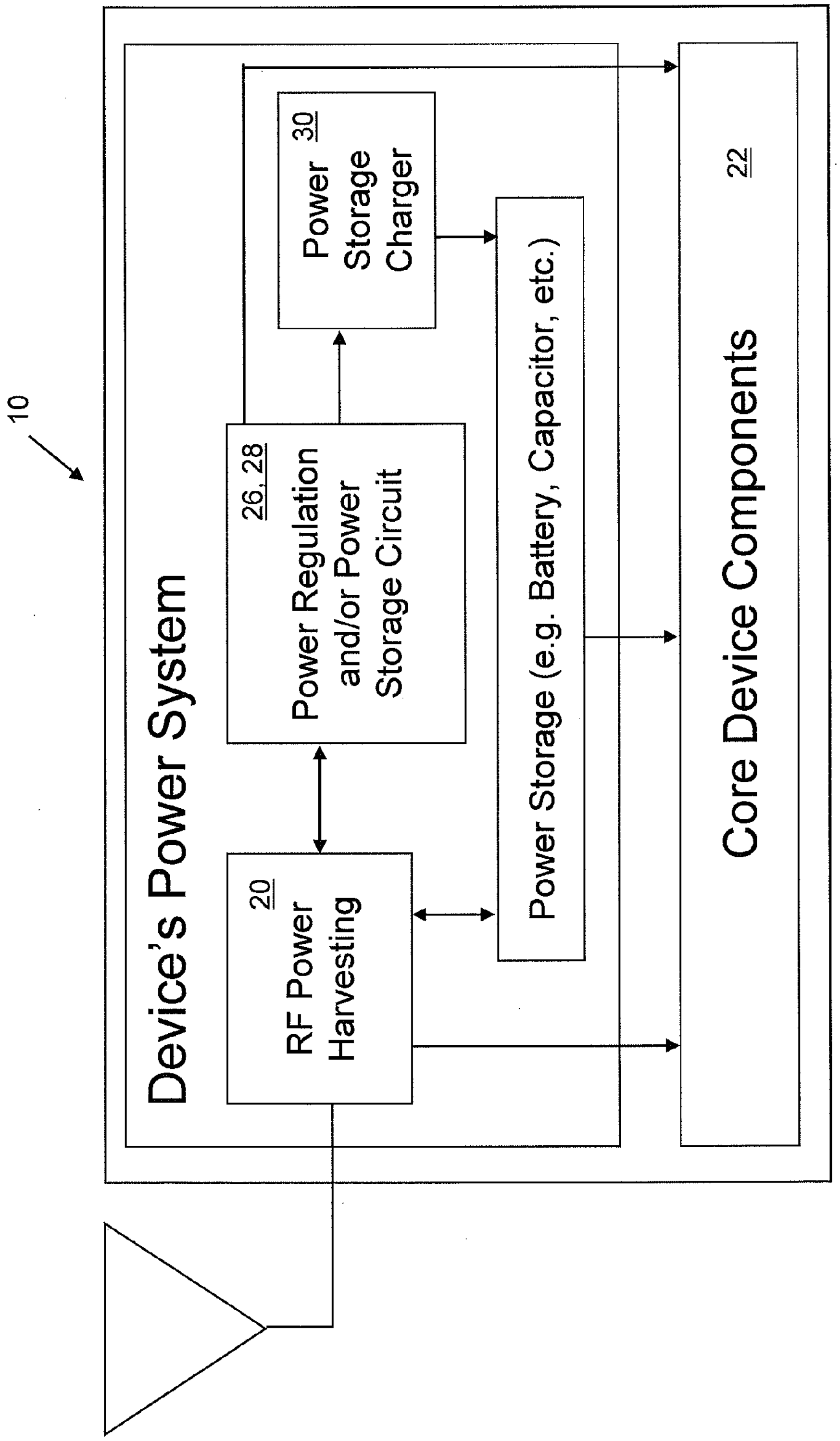
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 36

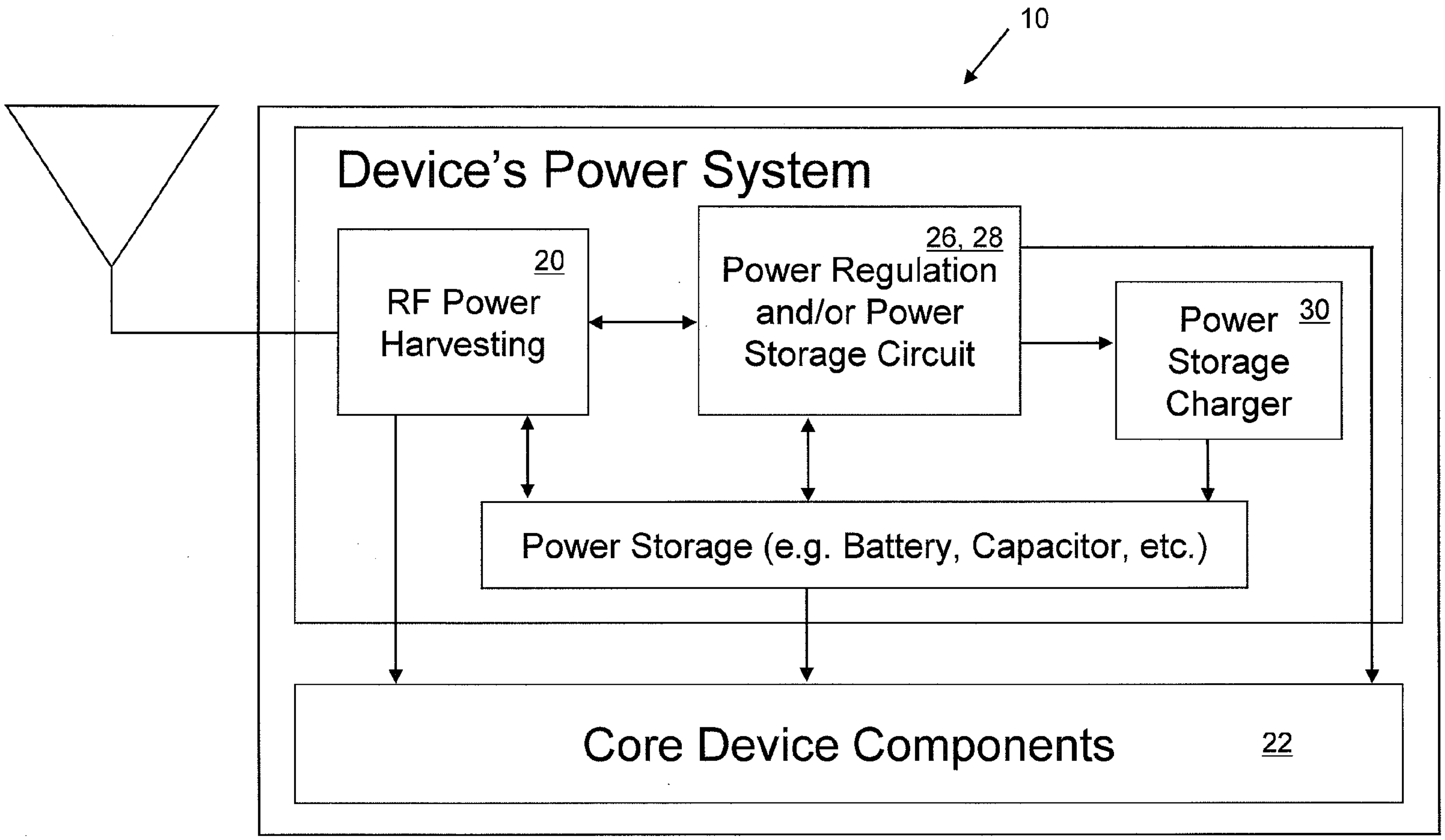
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 37

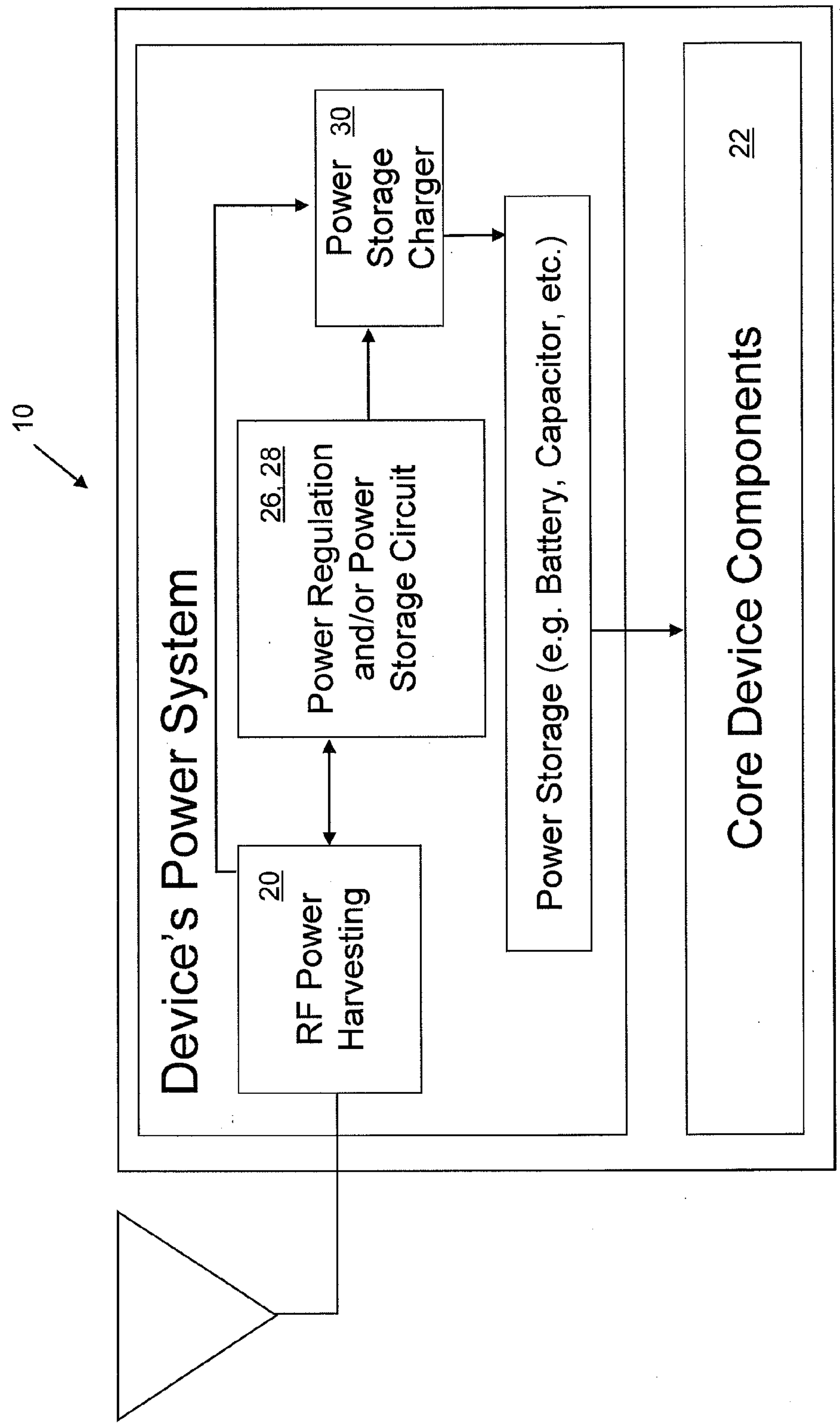
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 38

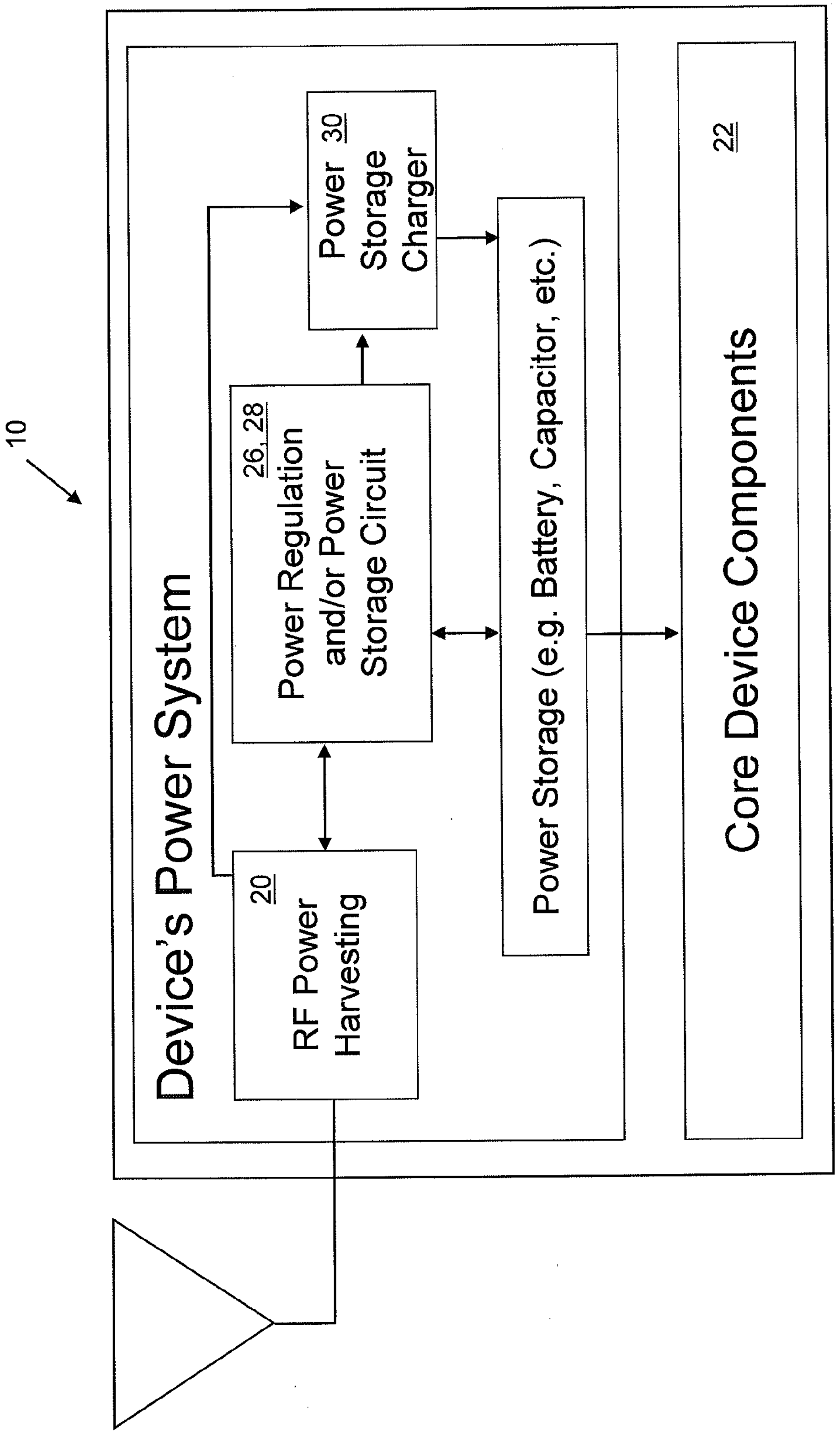
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30 Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 39

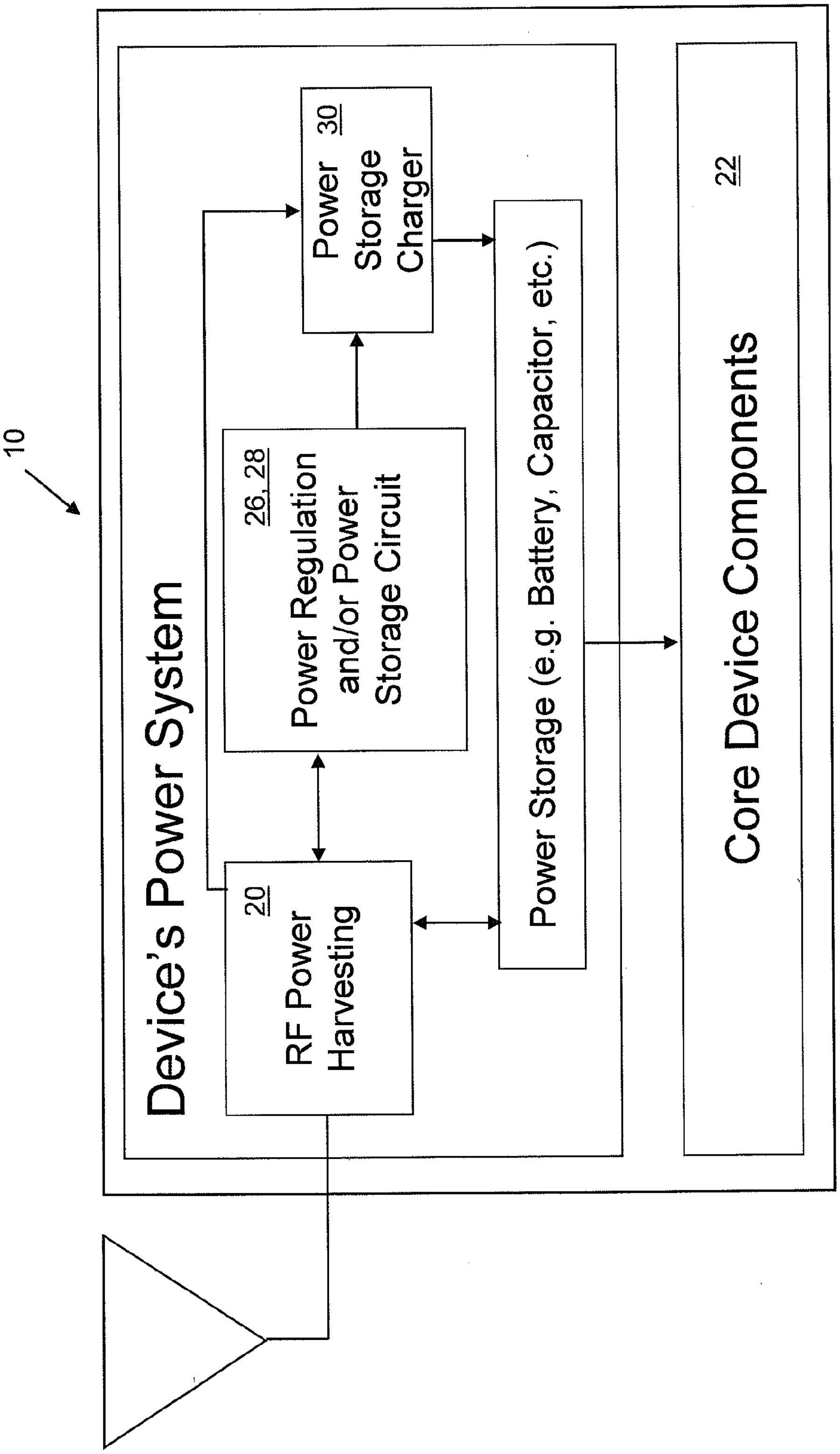
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 40

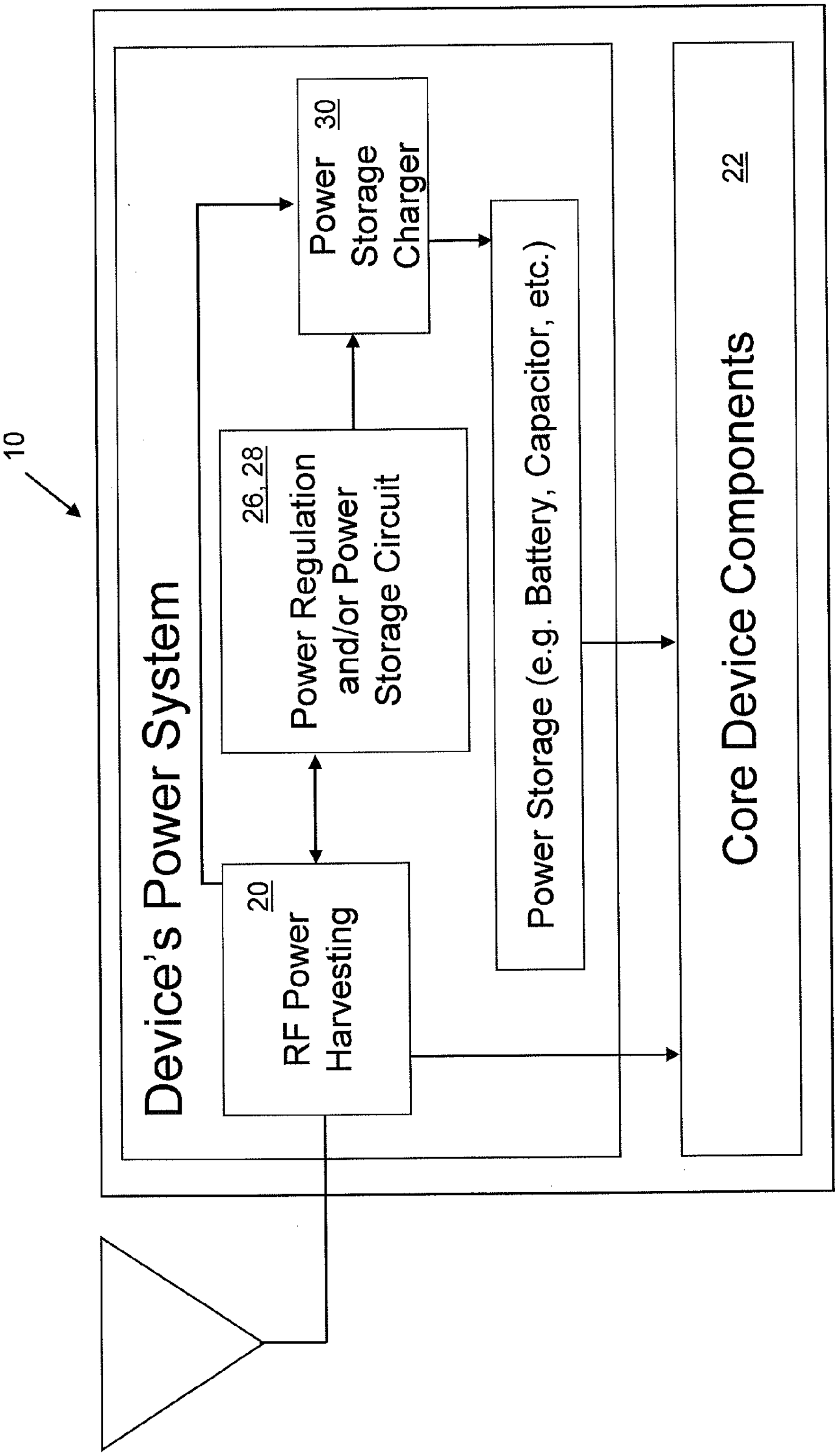
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 41

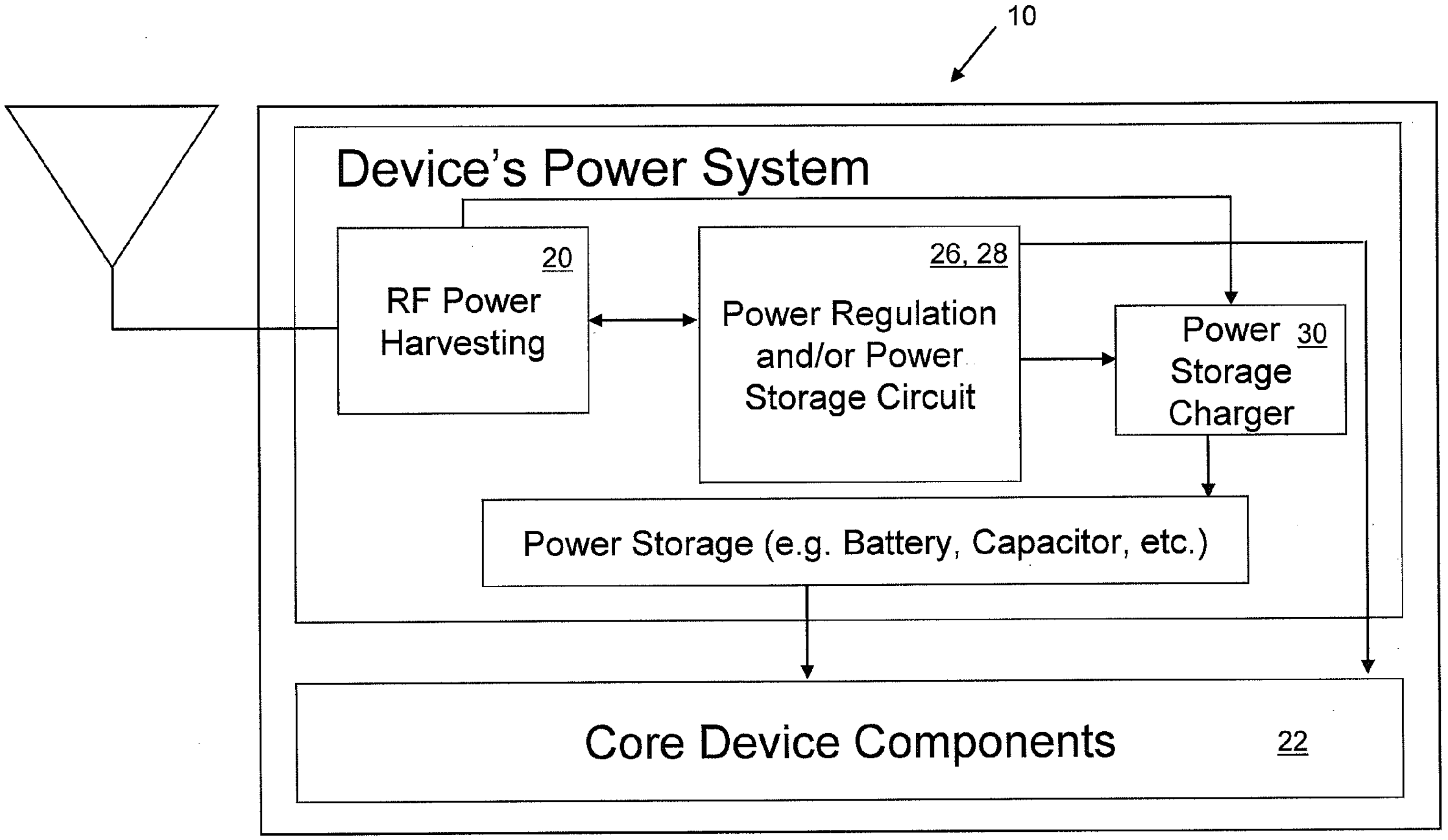
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 42

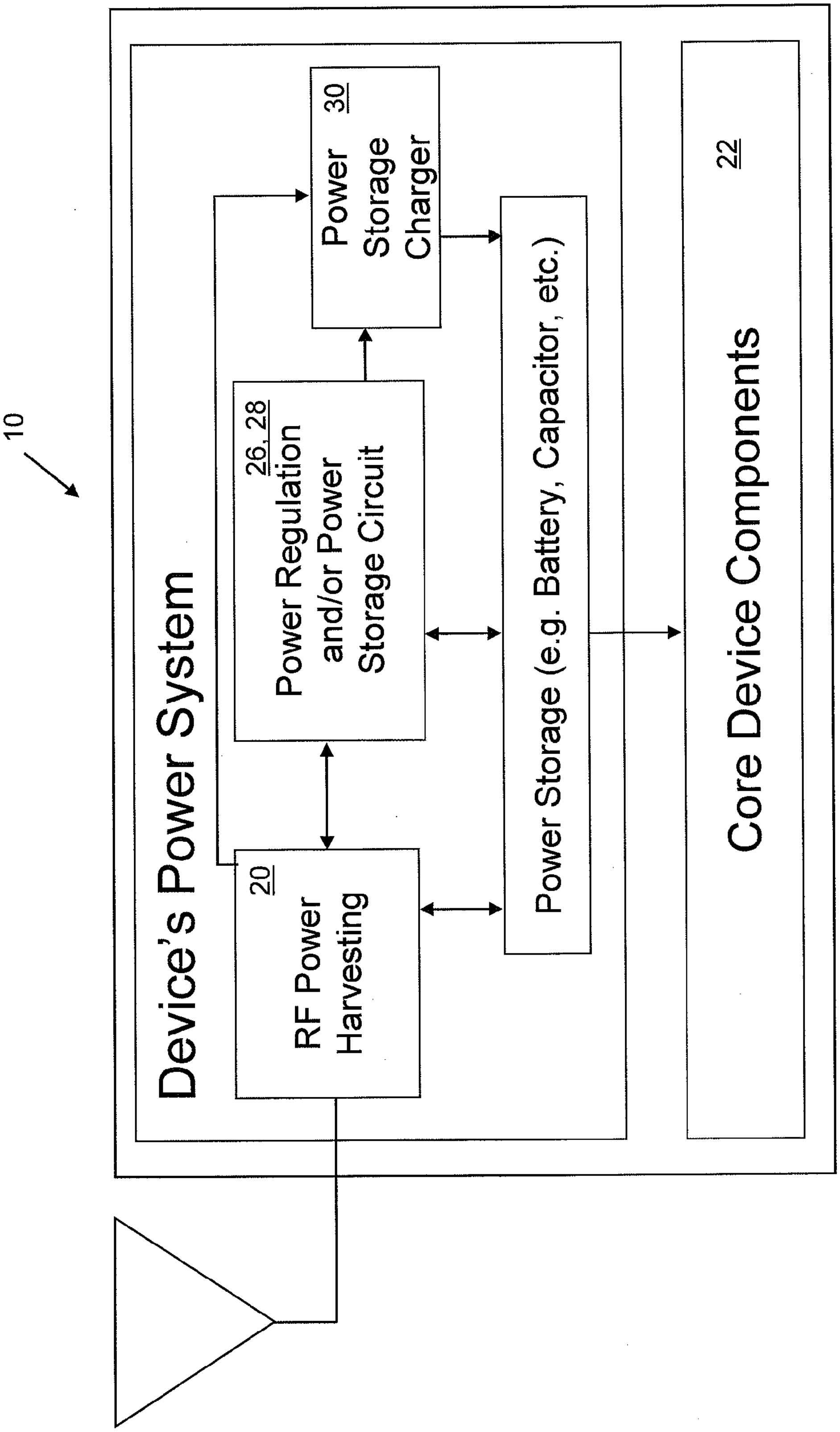
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 43

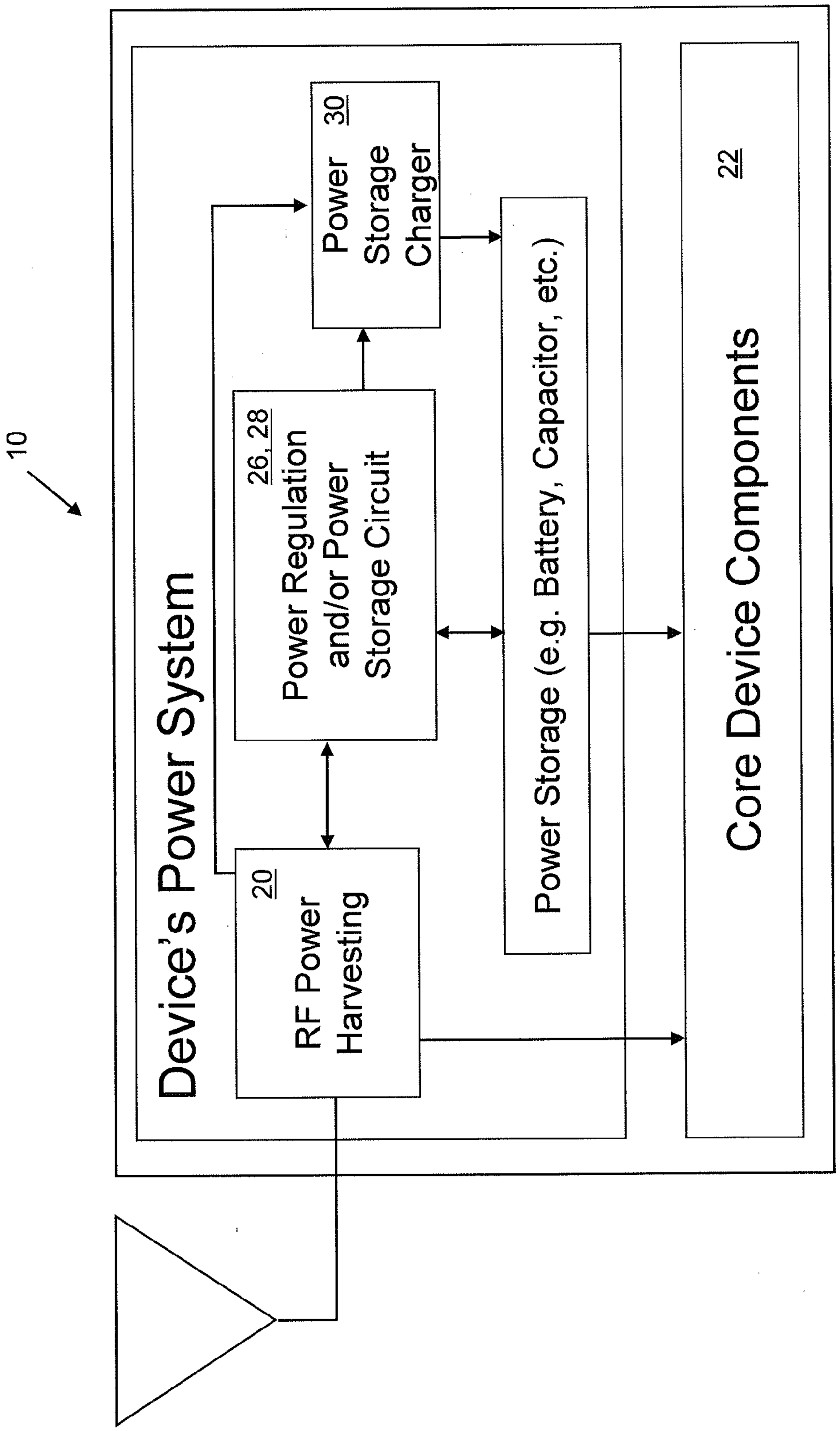
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 44

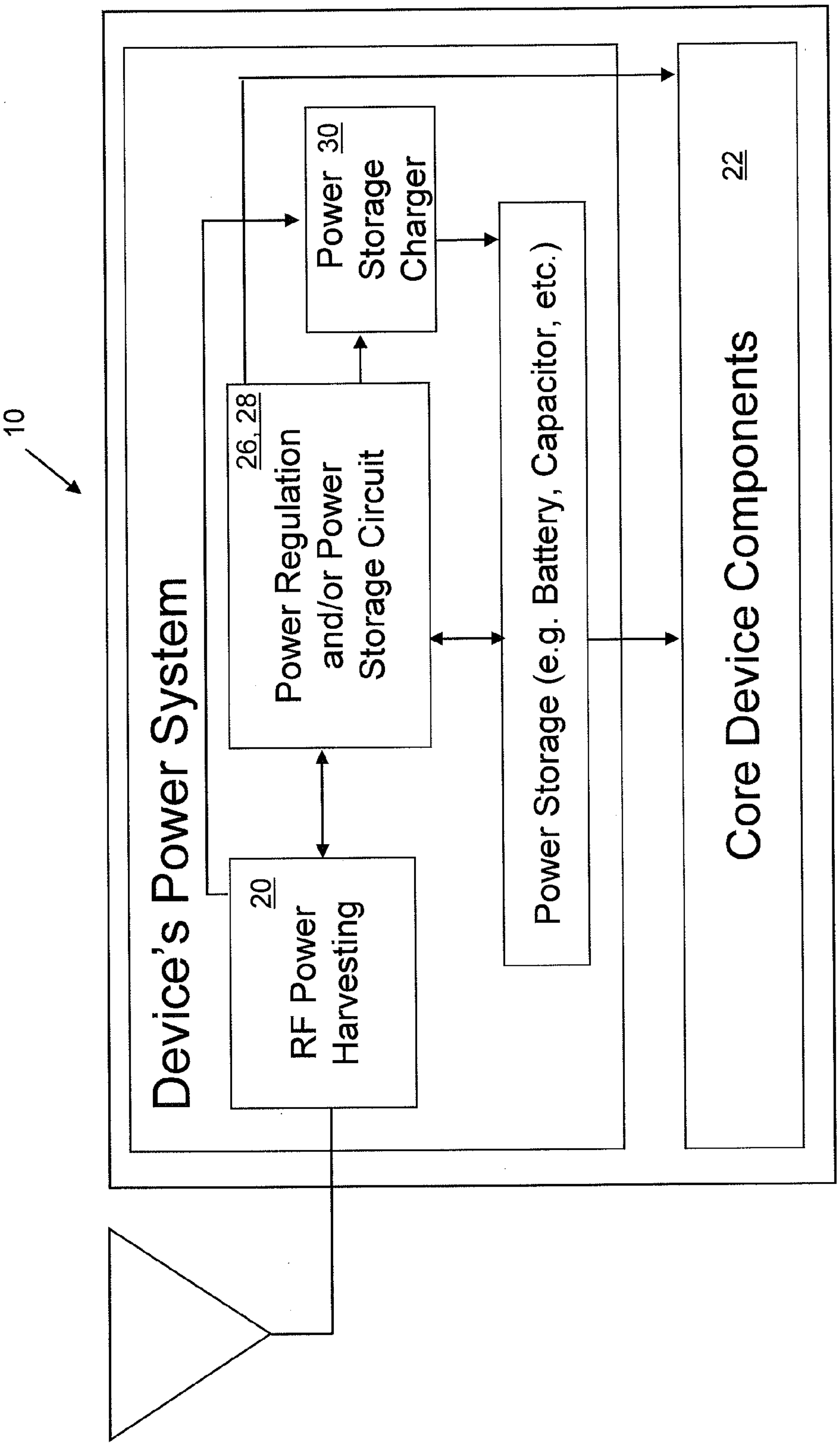
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30 Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 45

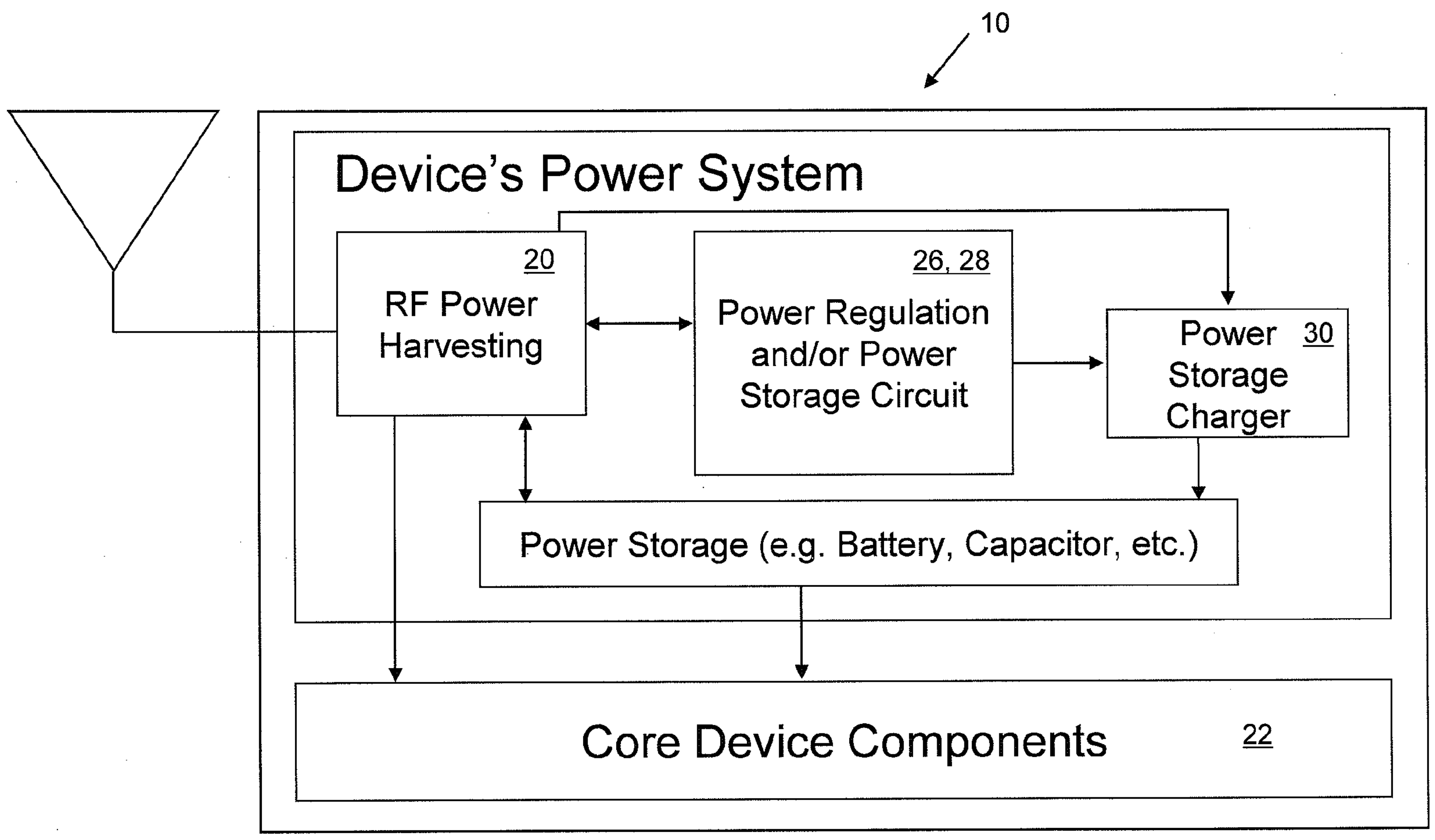
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 46

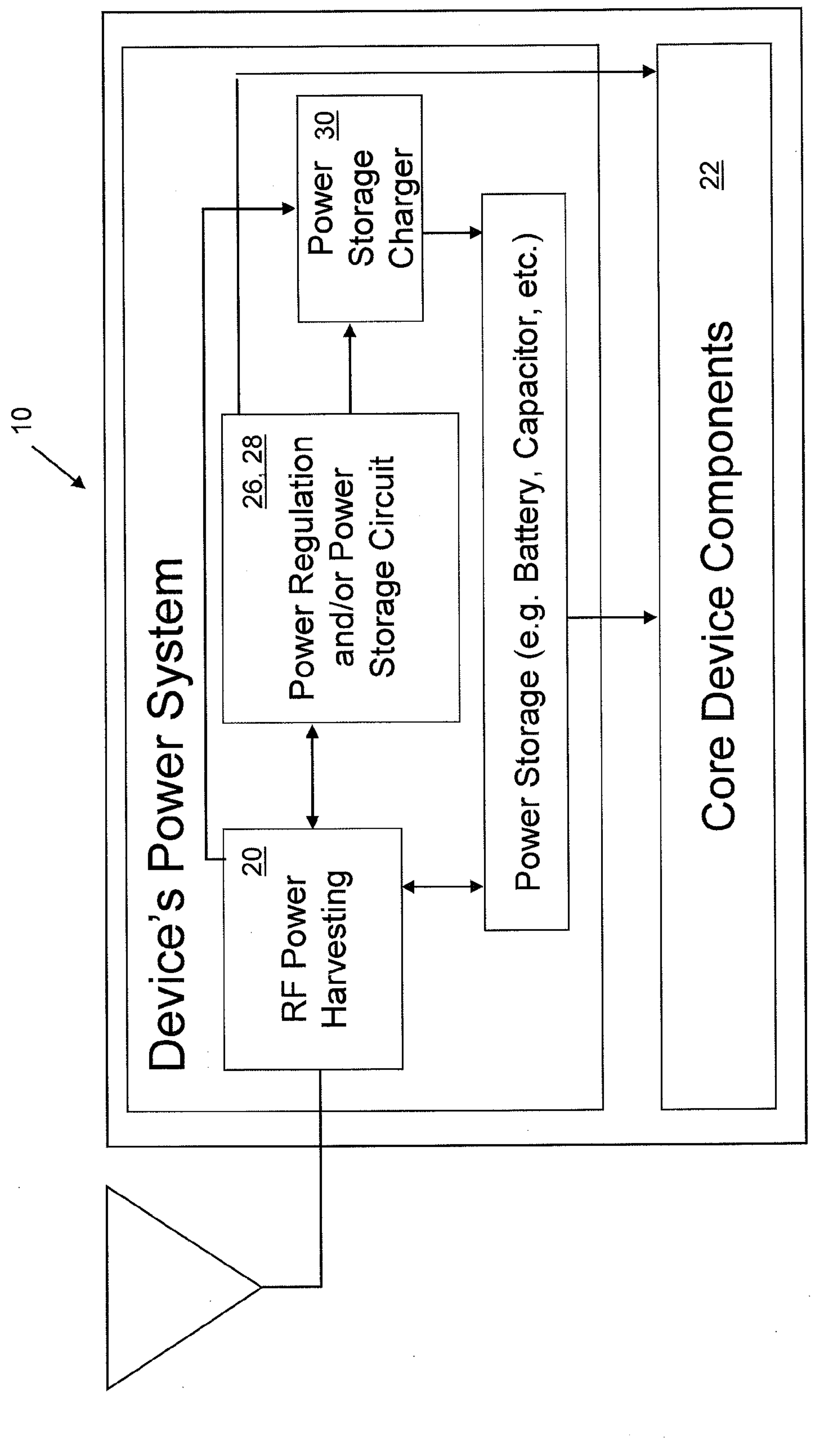
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 47

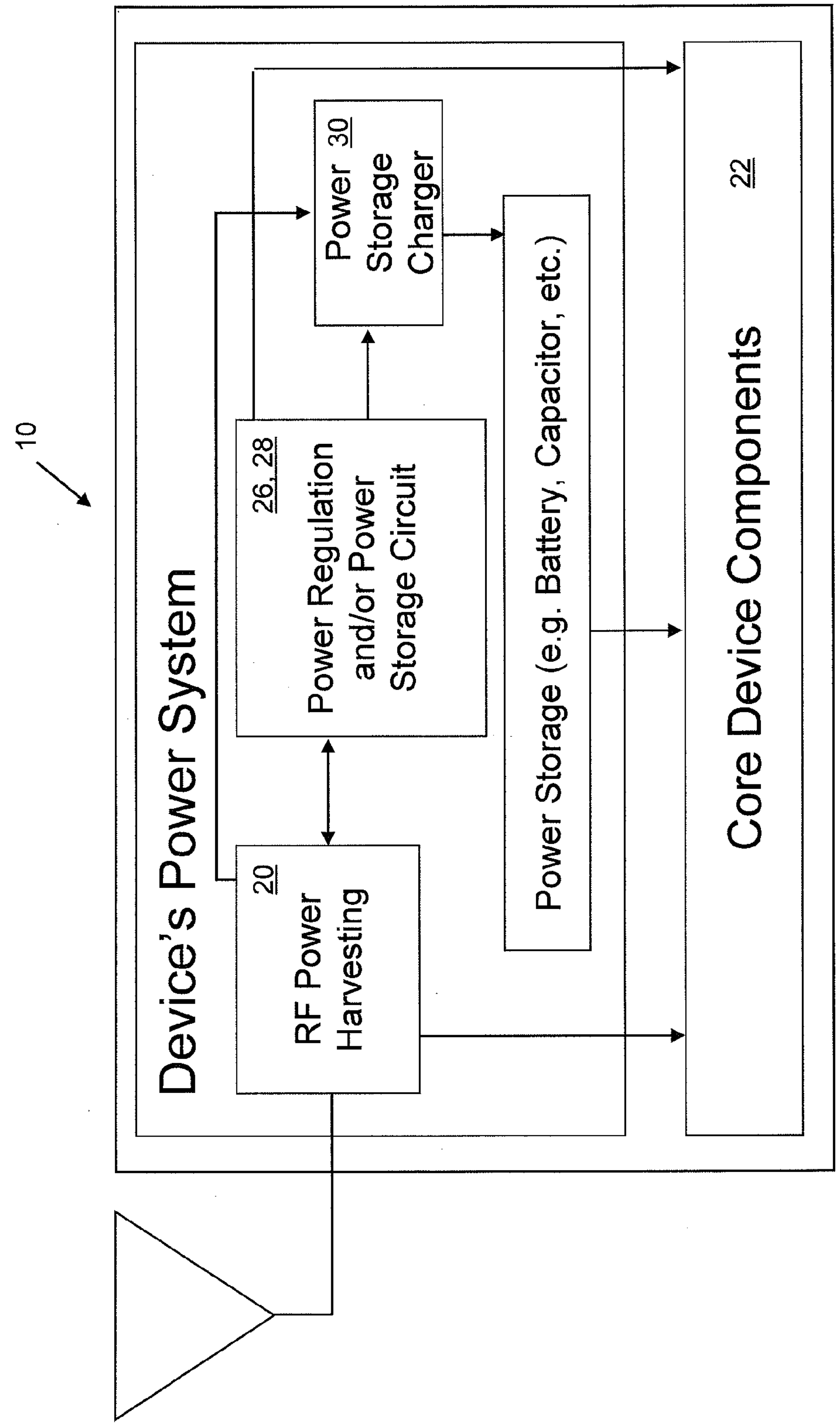
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 48

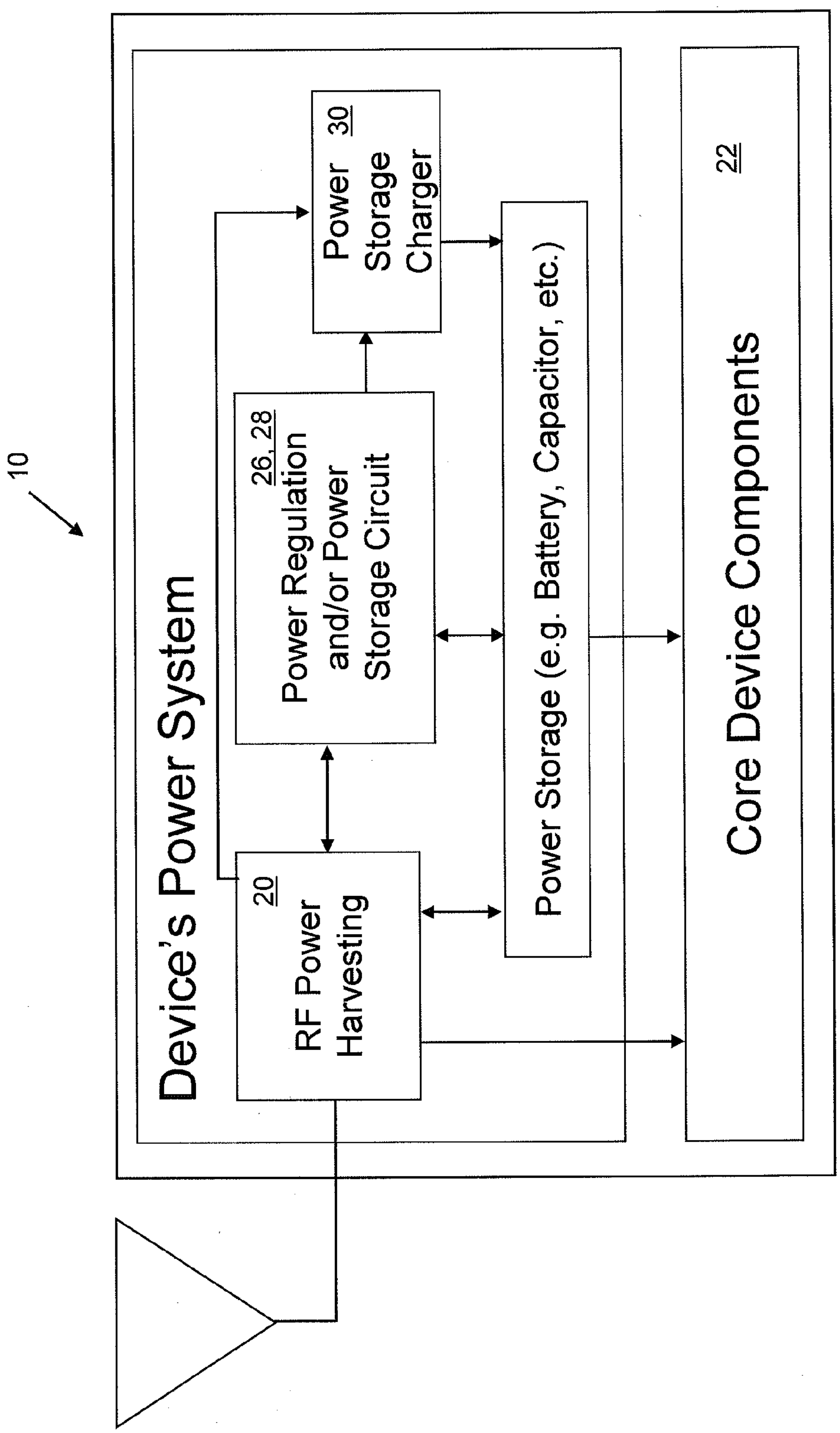
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 49

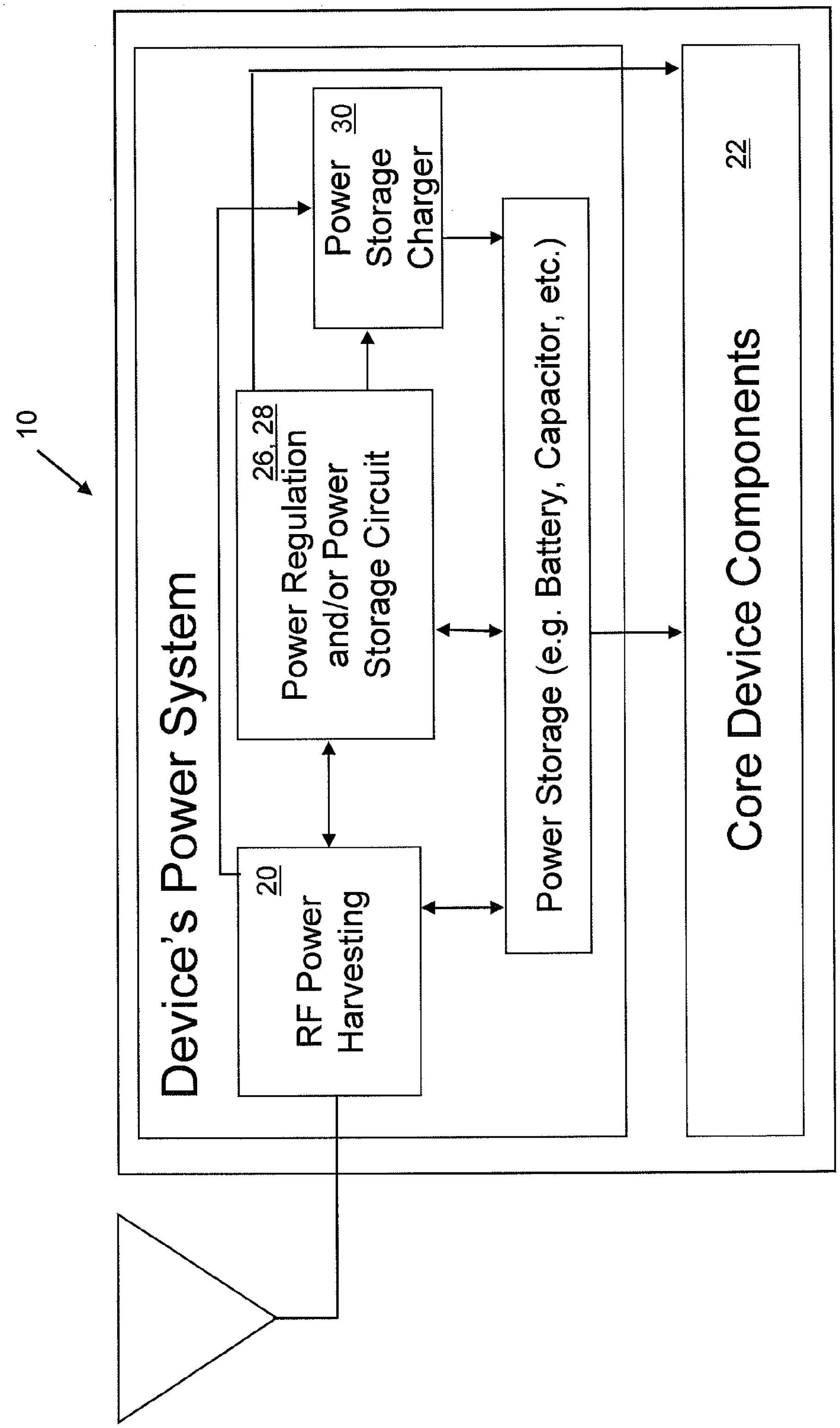
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 50

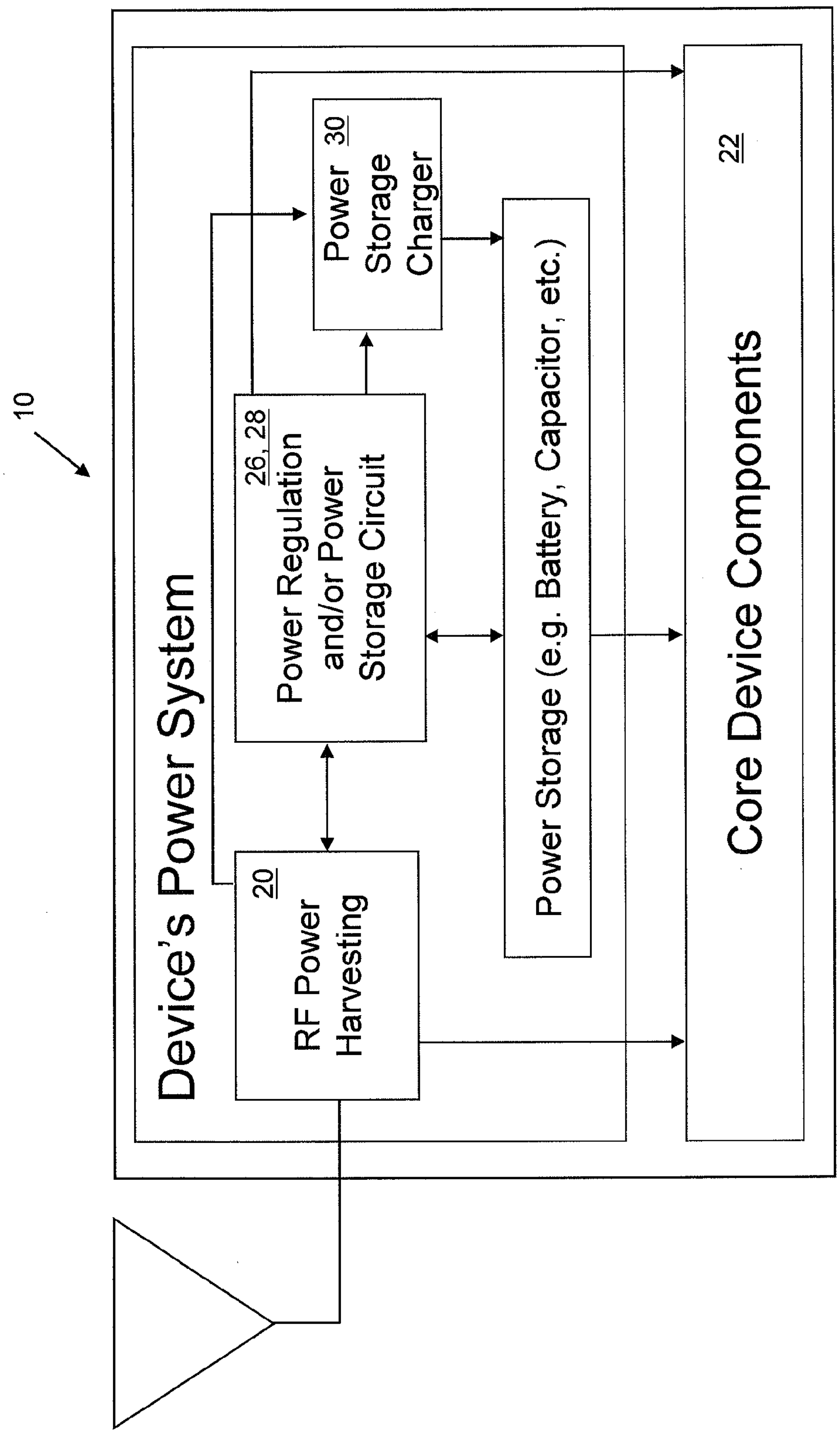
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 51

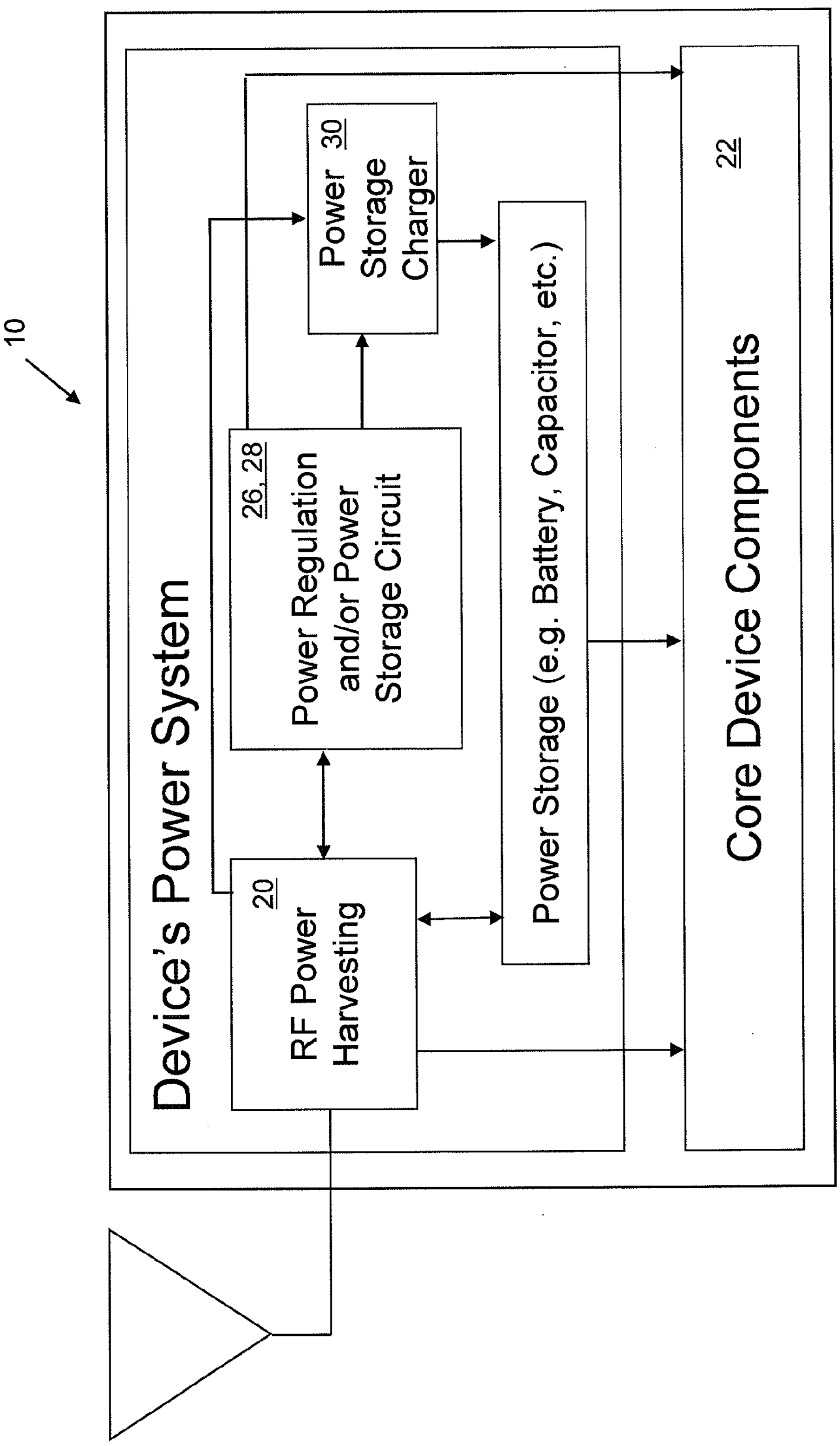
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 52

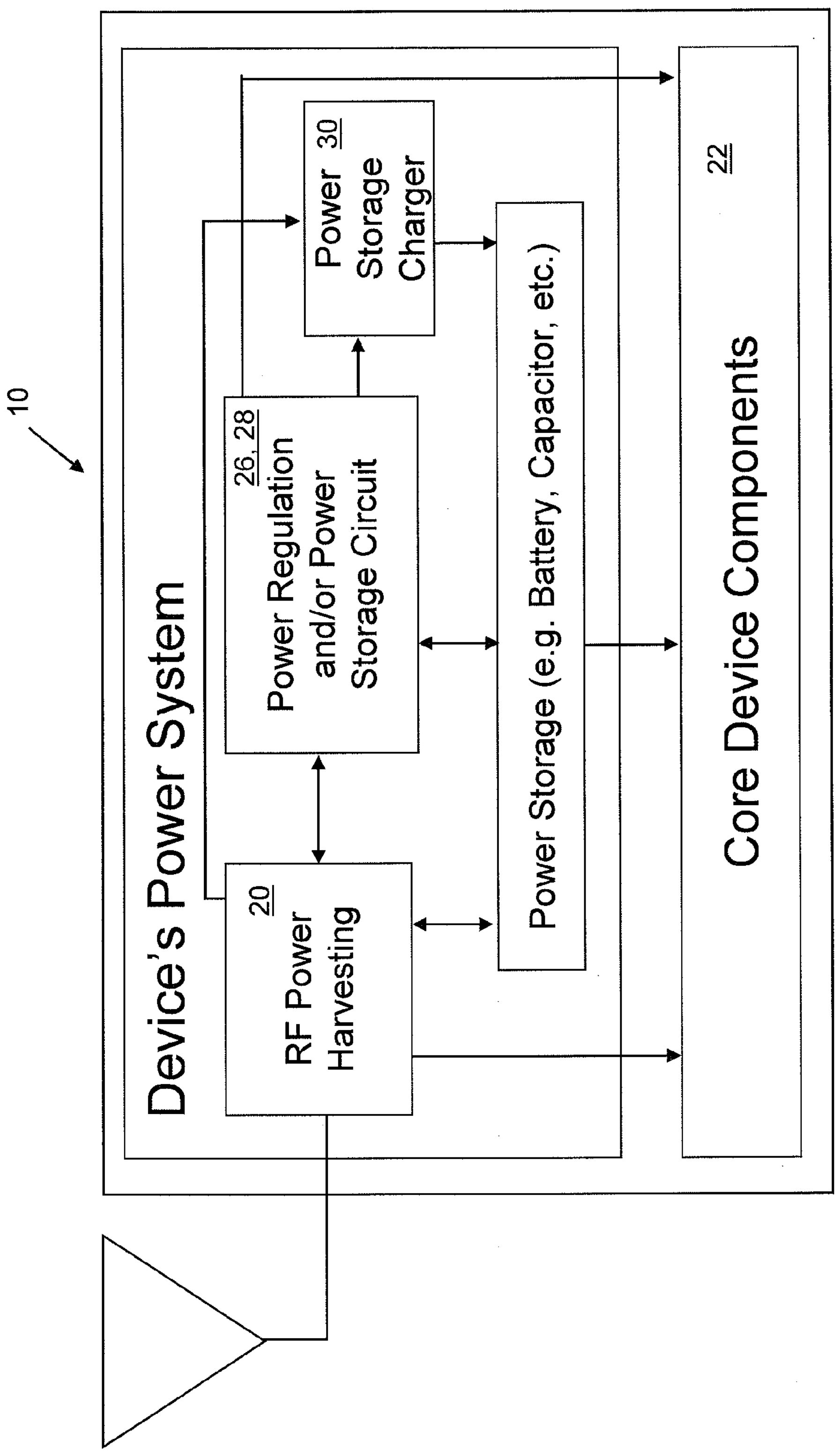
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 53

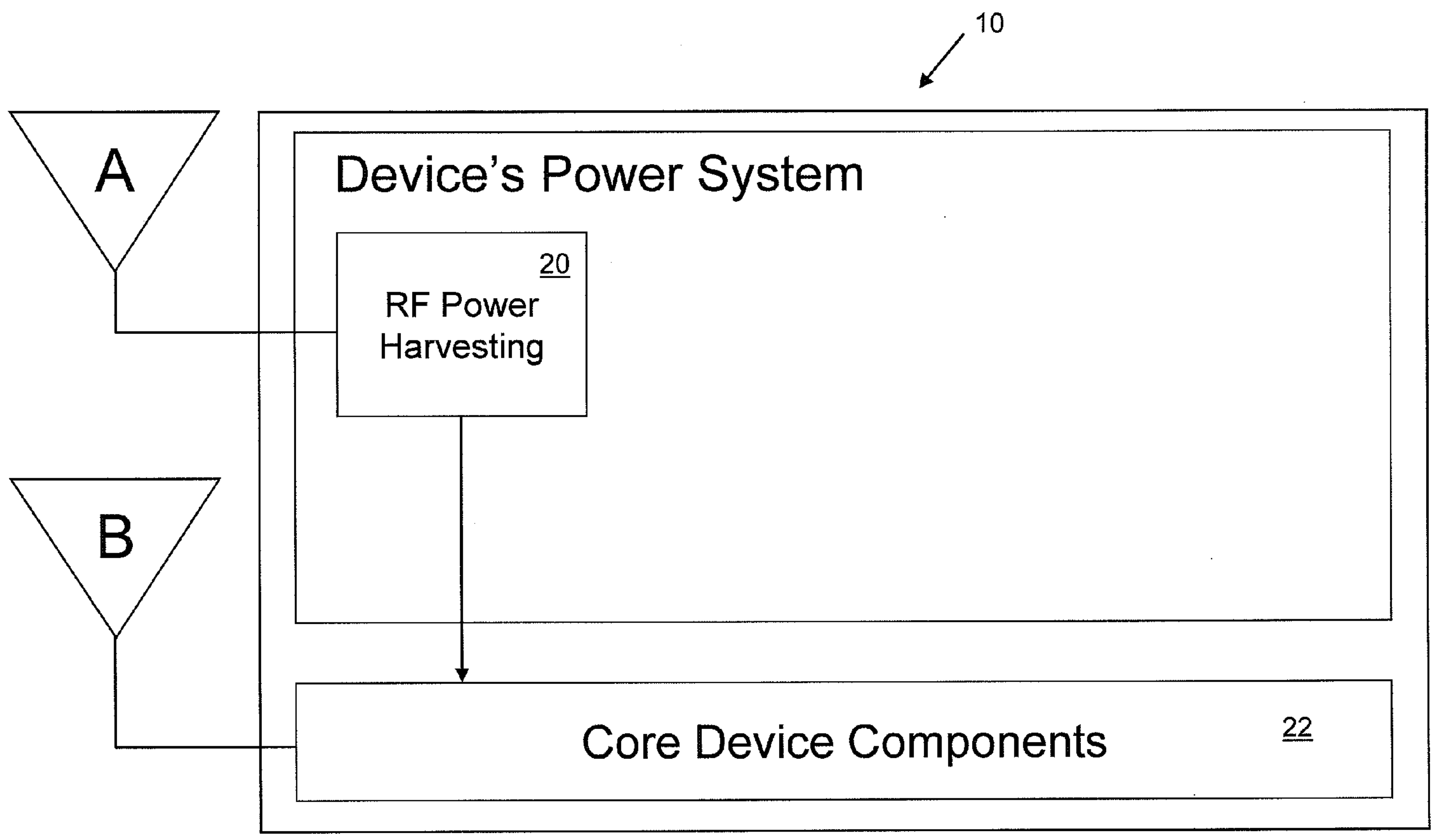
10
A
B
Device's Power System
20
RF Power Harvesting
Core Device Components
22

Fig. 54

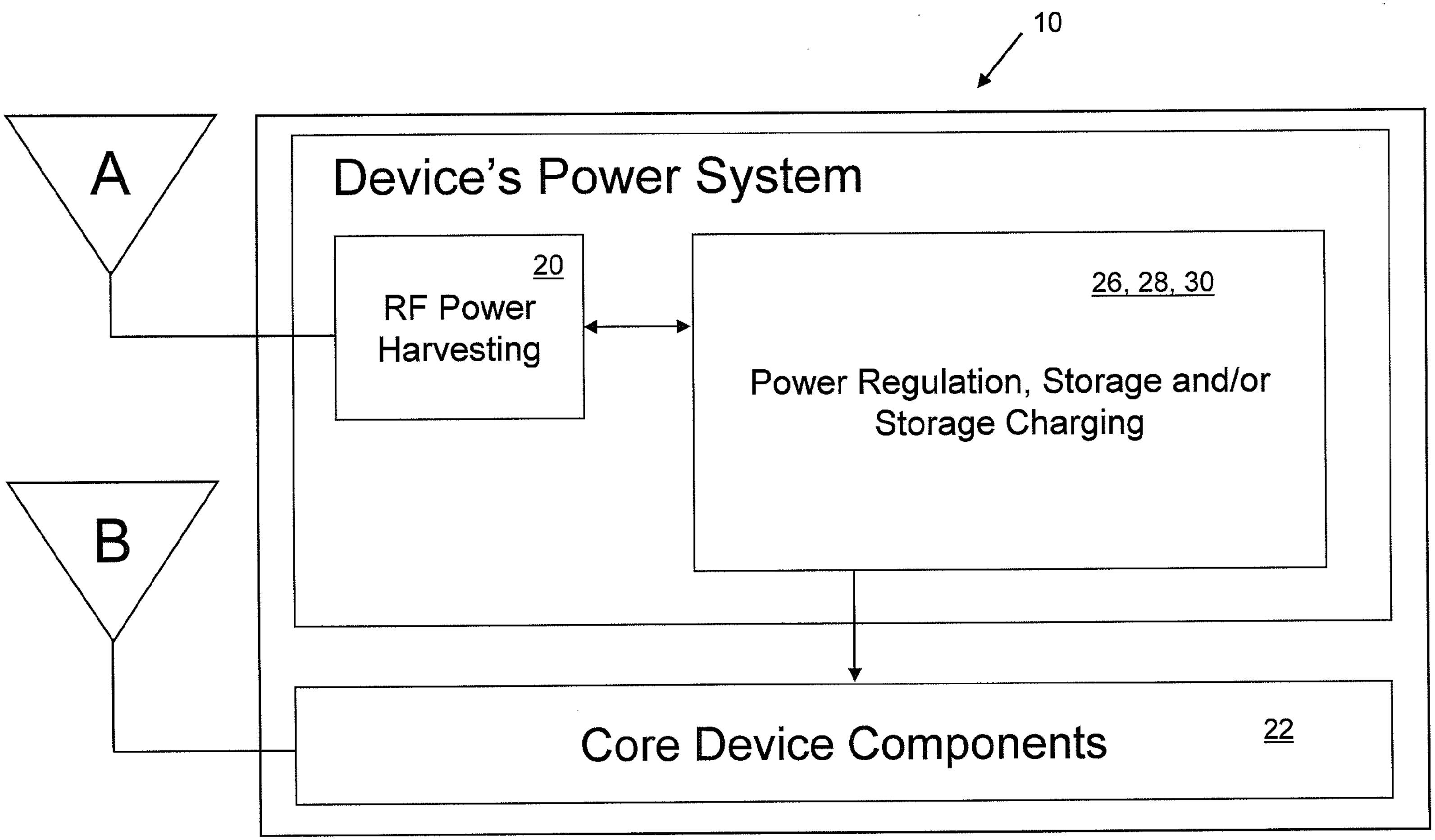
10
A
B
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22

Fig. 55

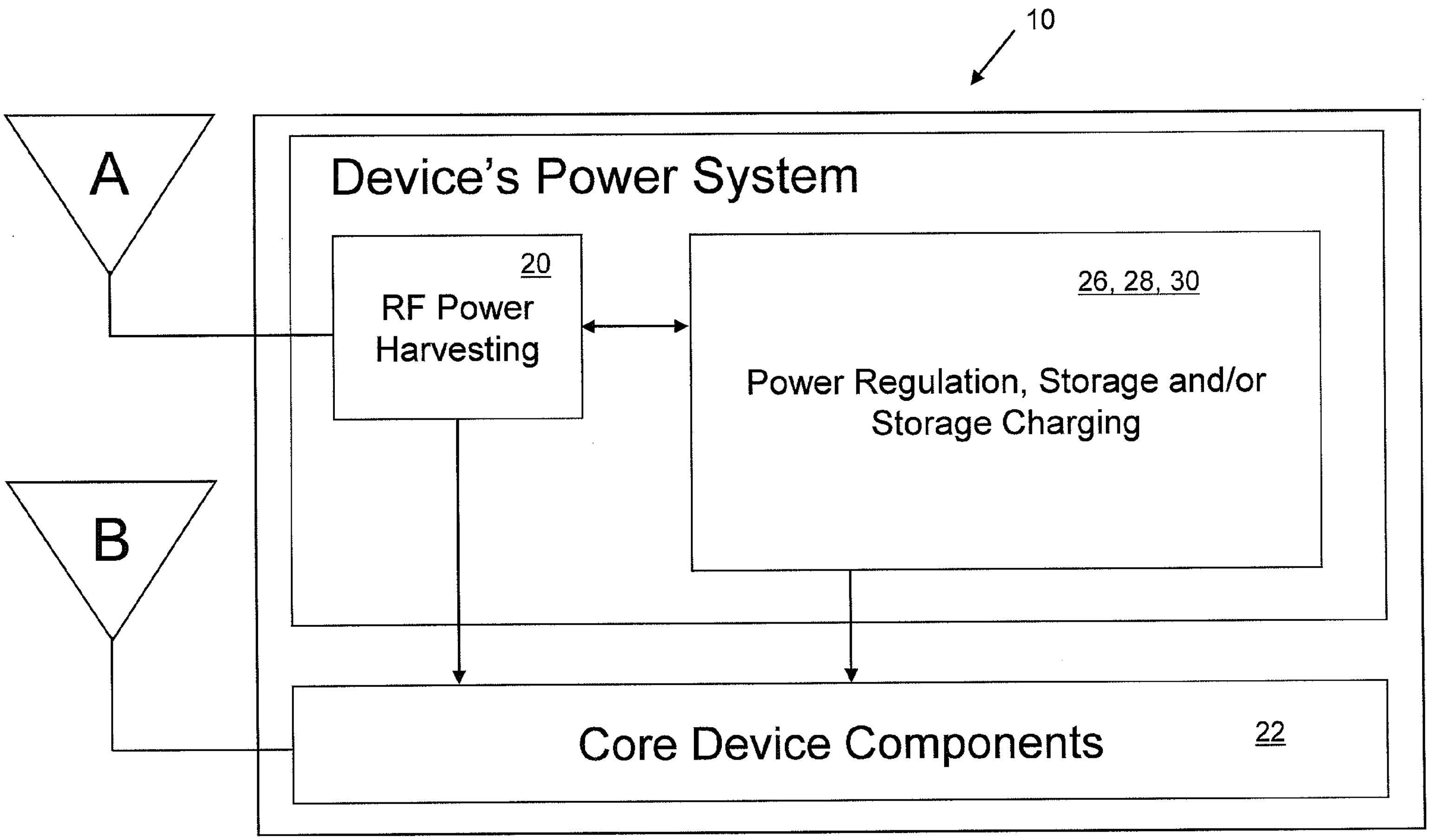
10
A
B
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22

Fig. 56

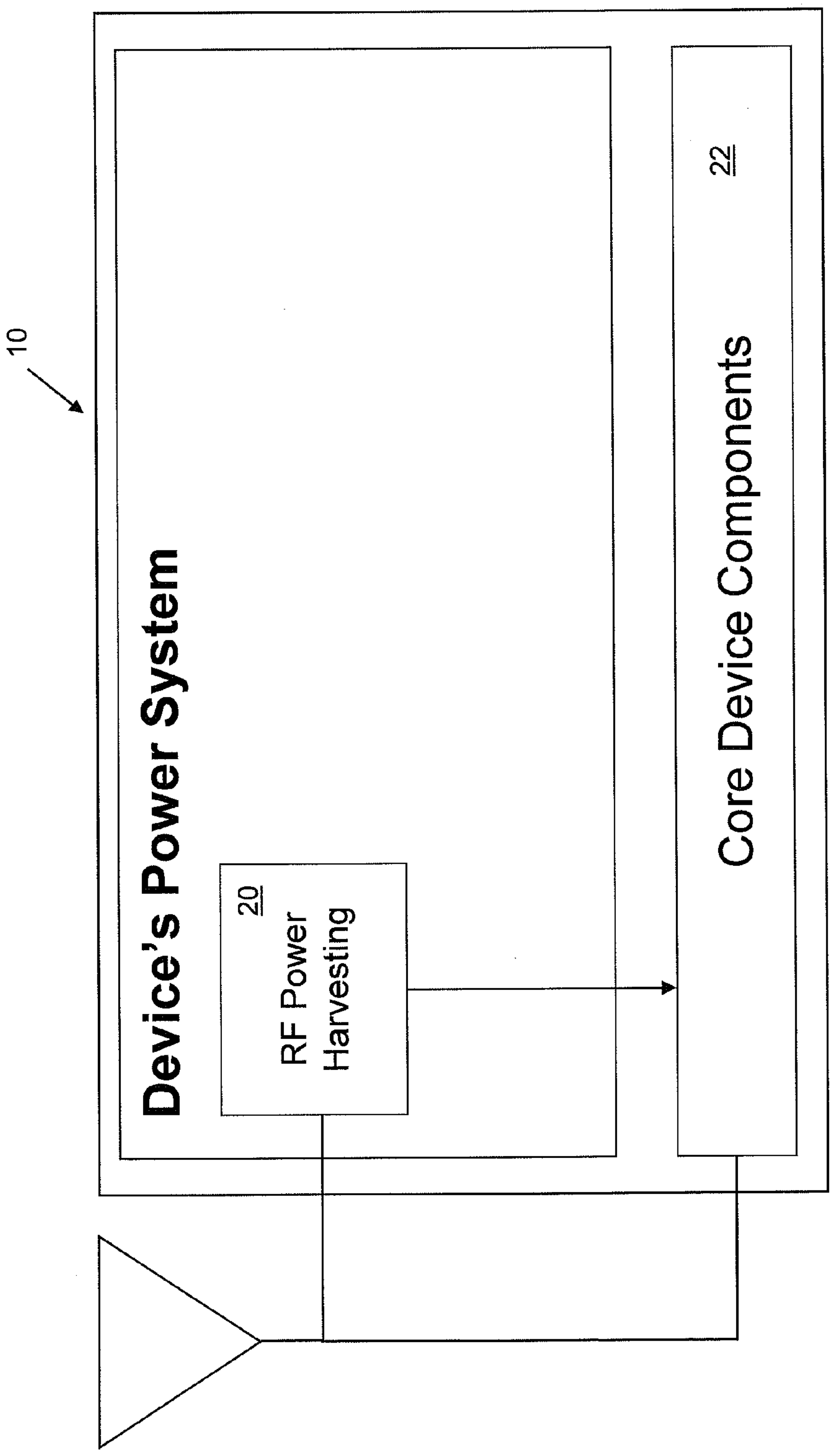
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22

Fig. 57

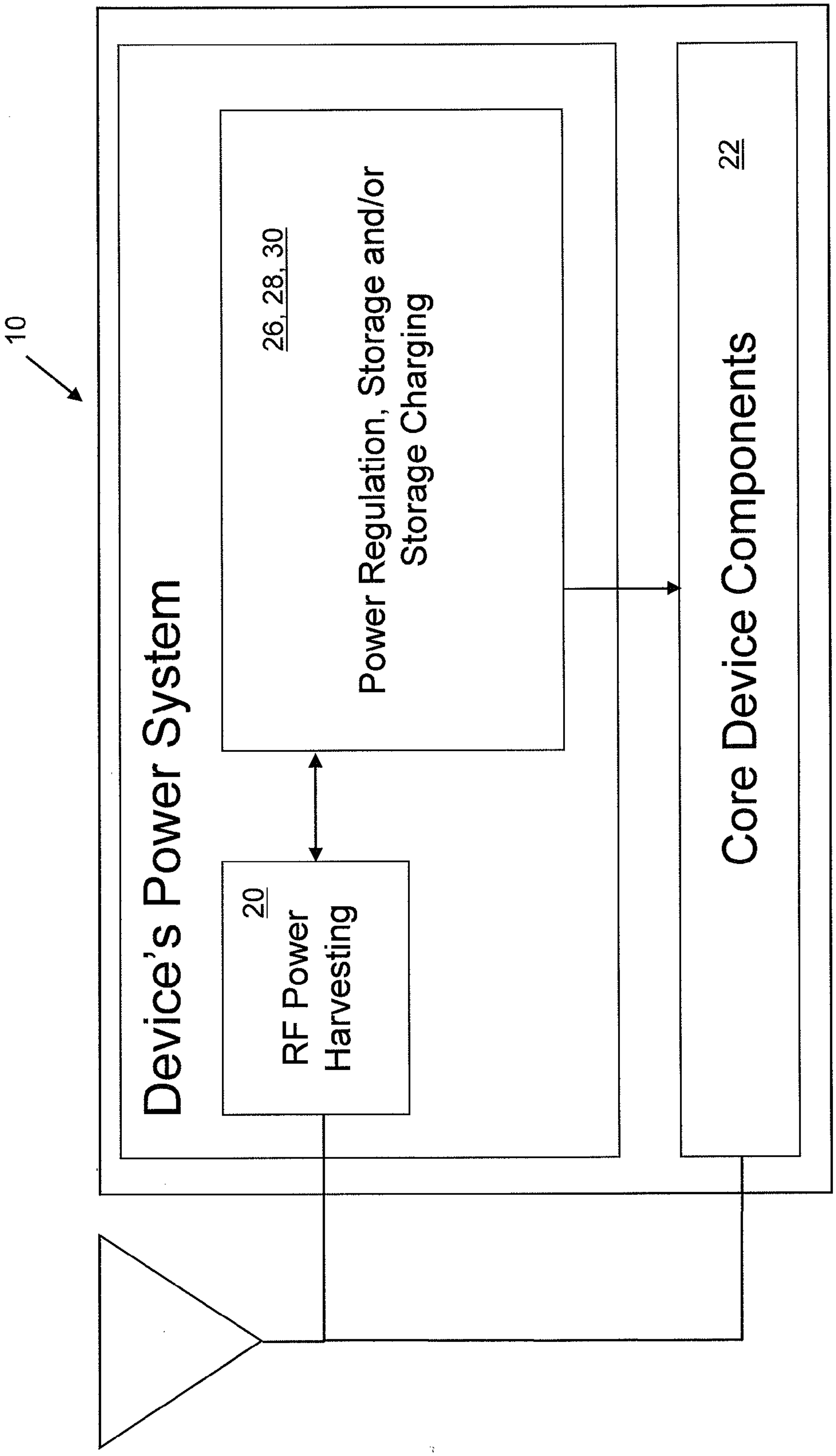
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22

Fig. 58

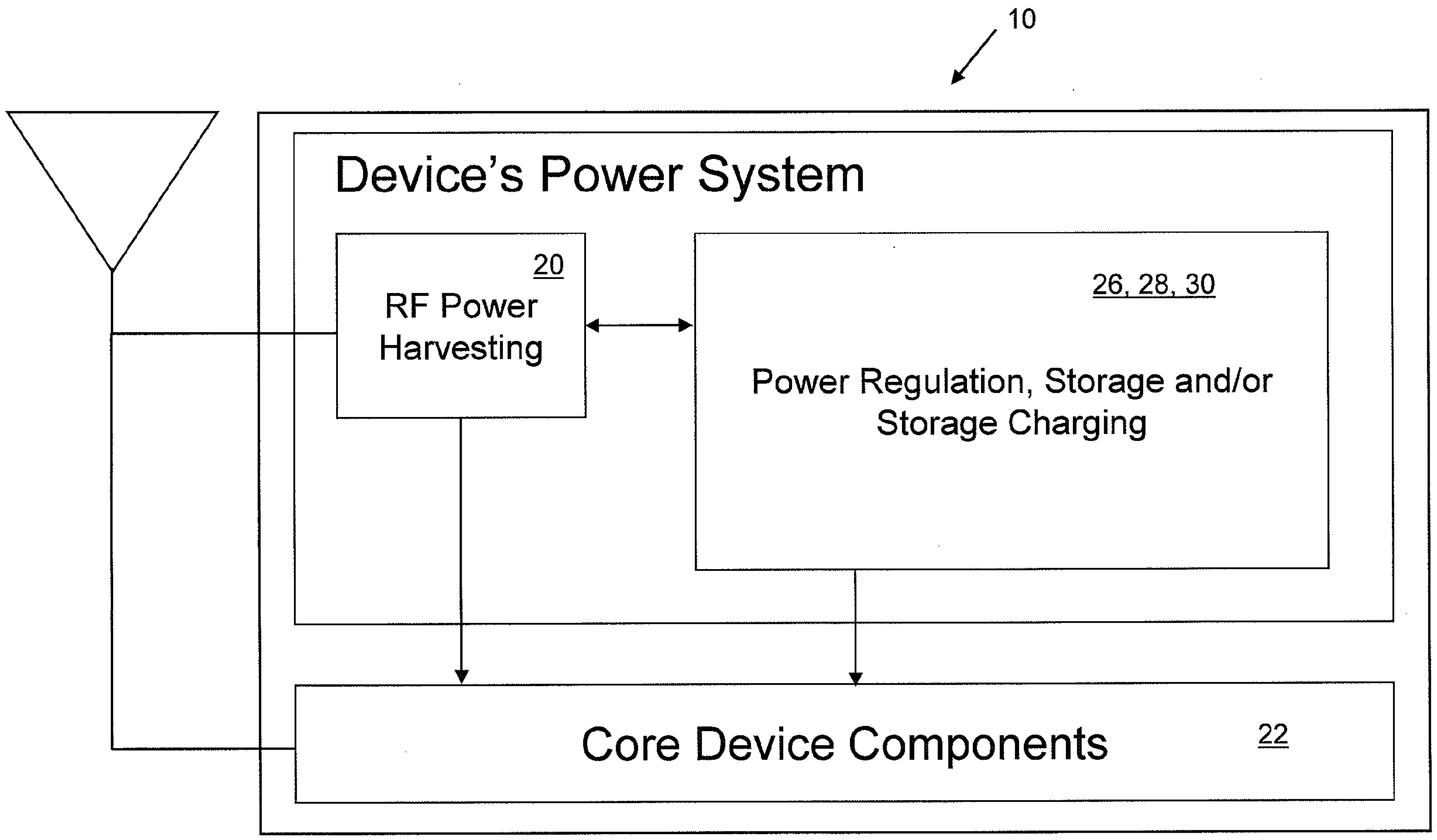
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22

Fig. 59

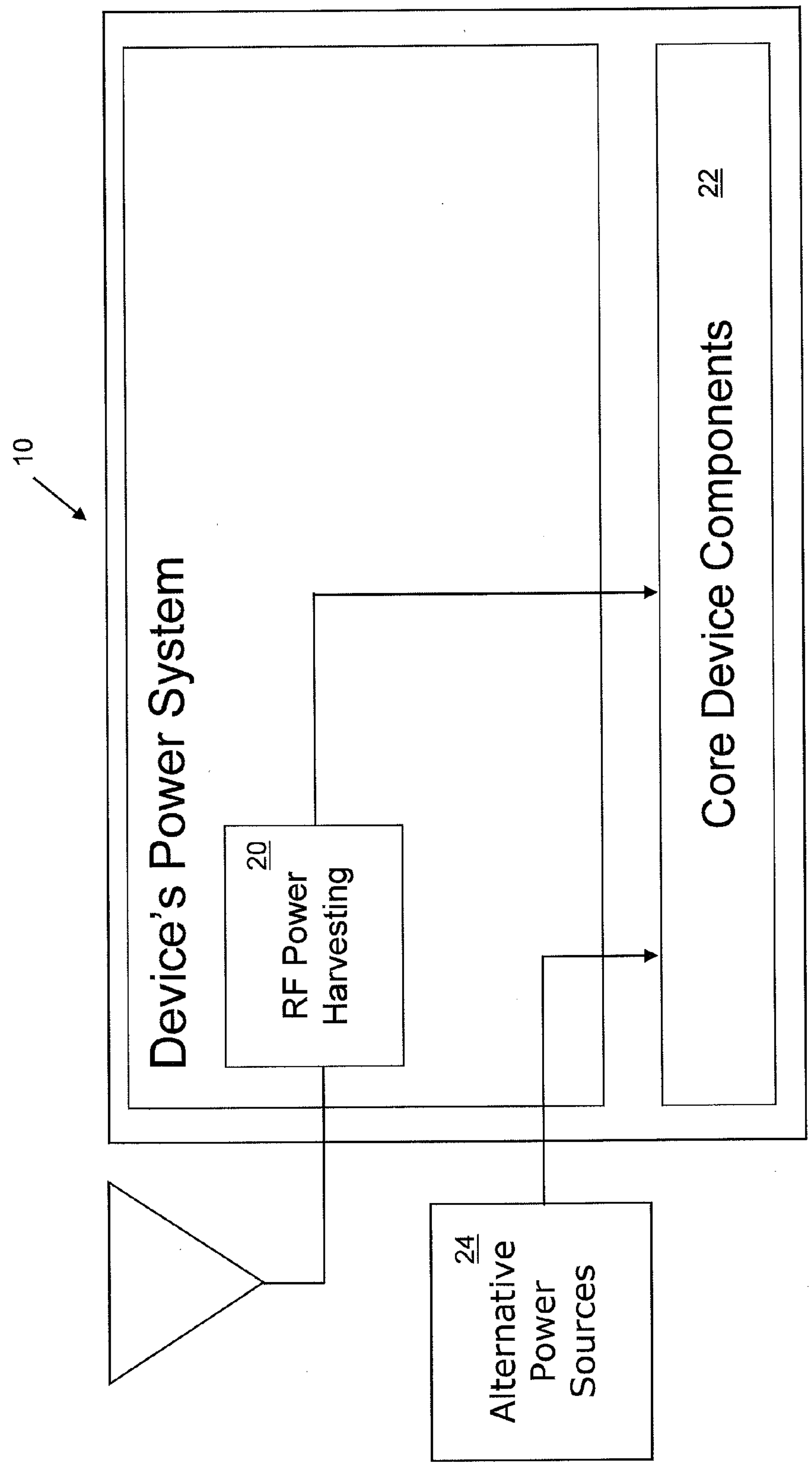
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 60

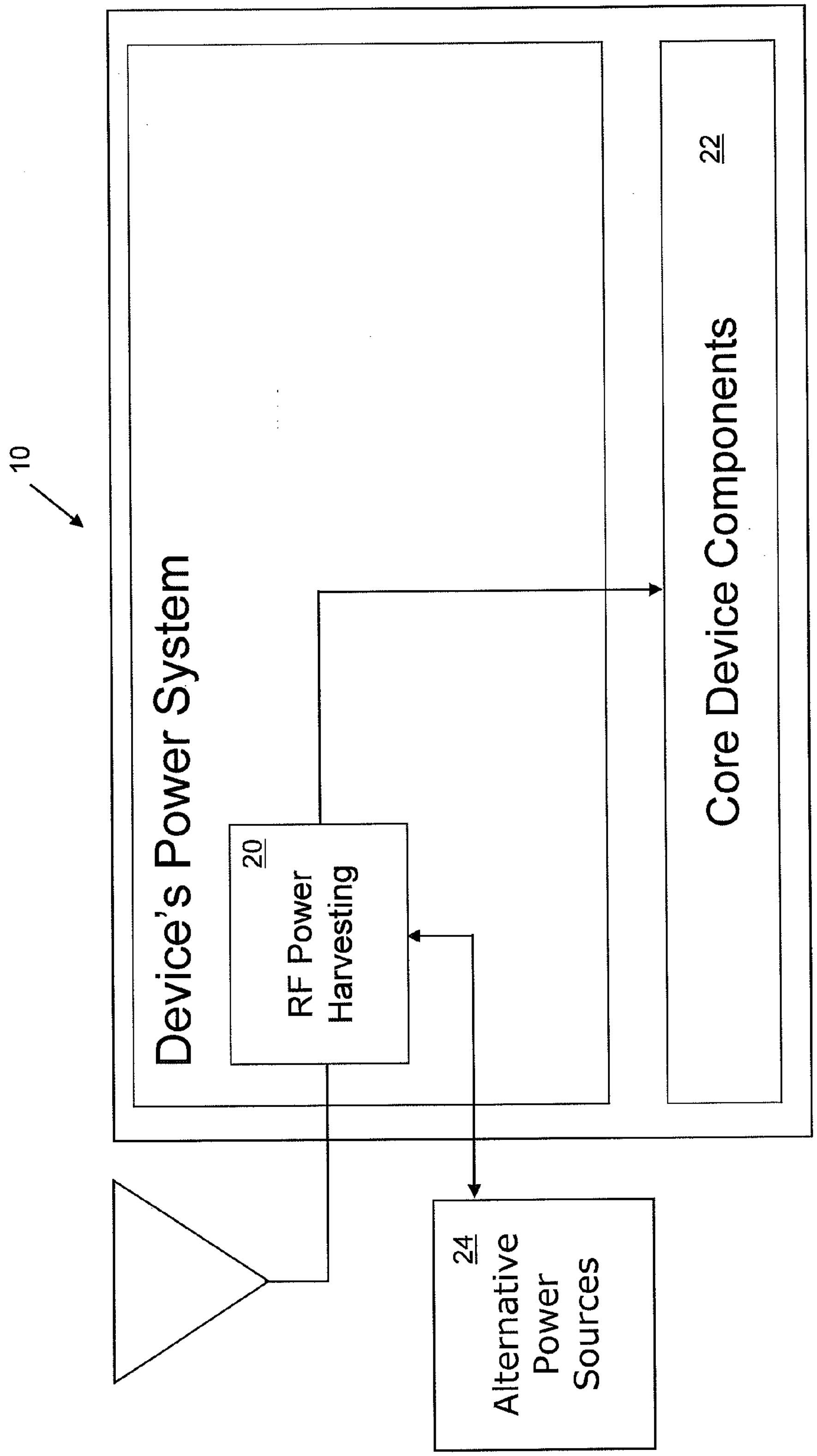
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 61

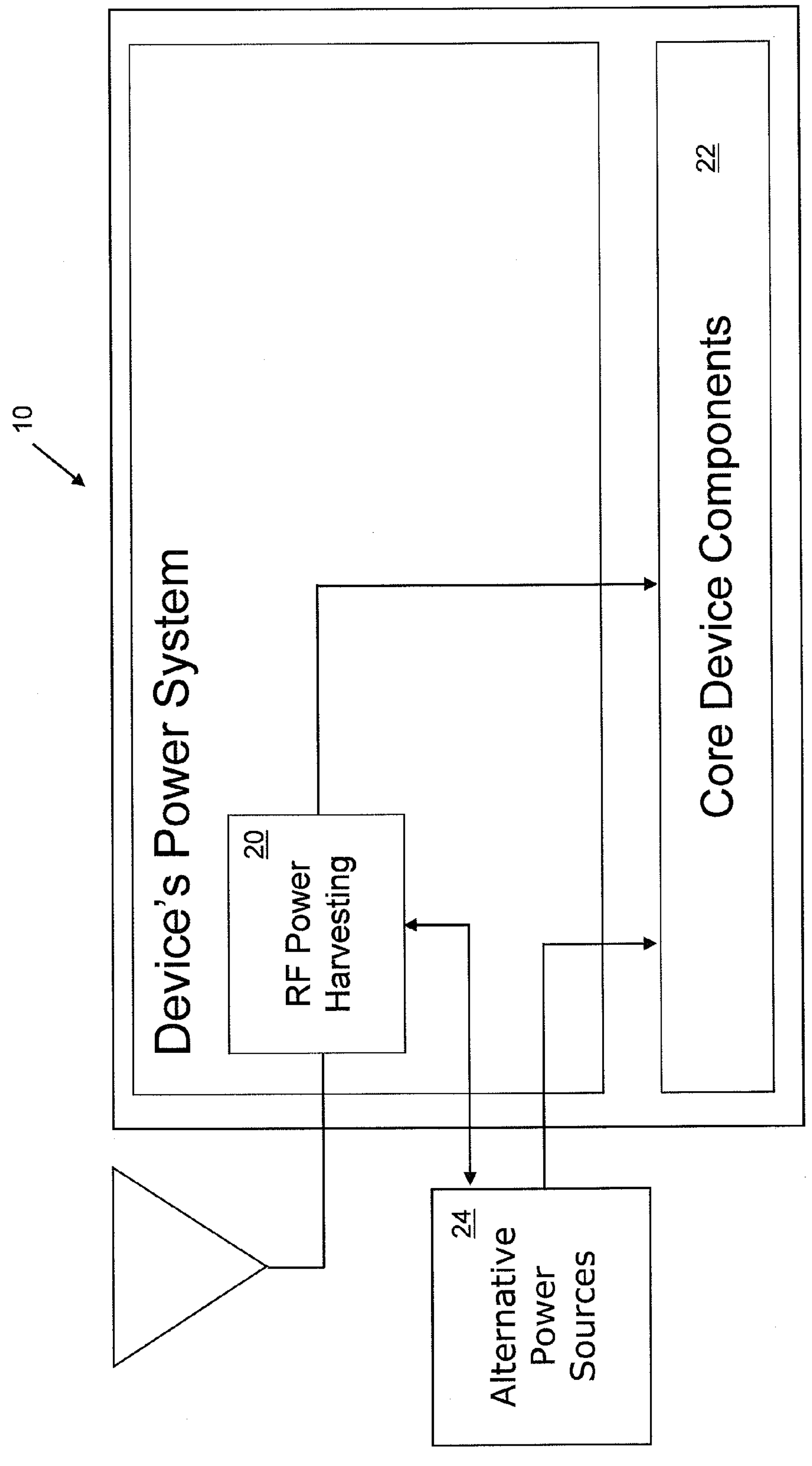
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 62

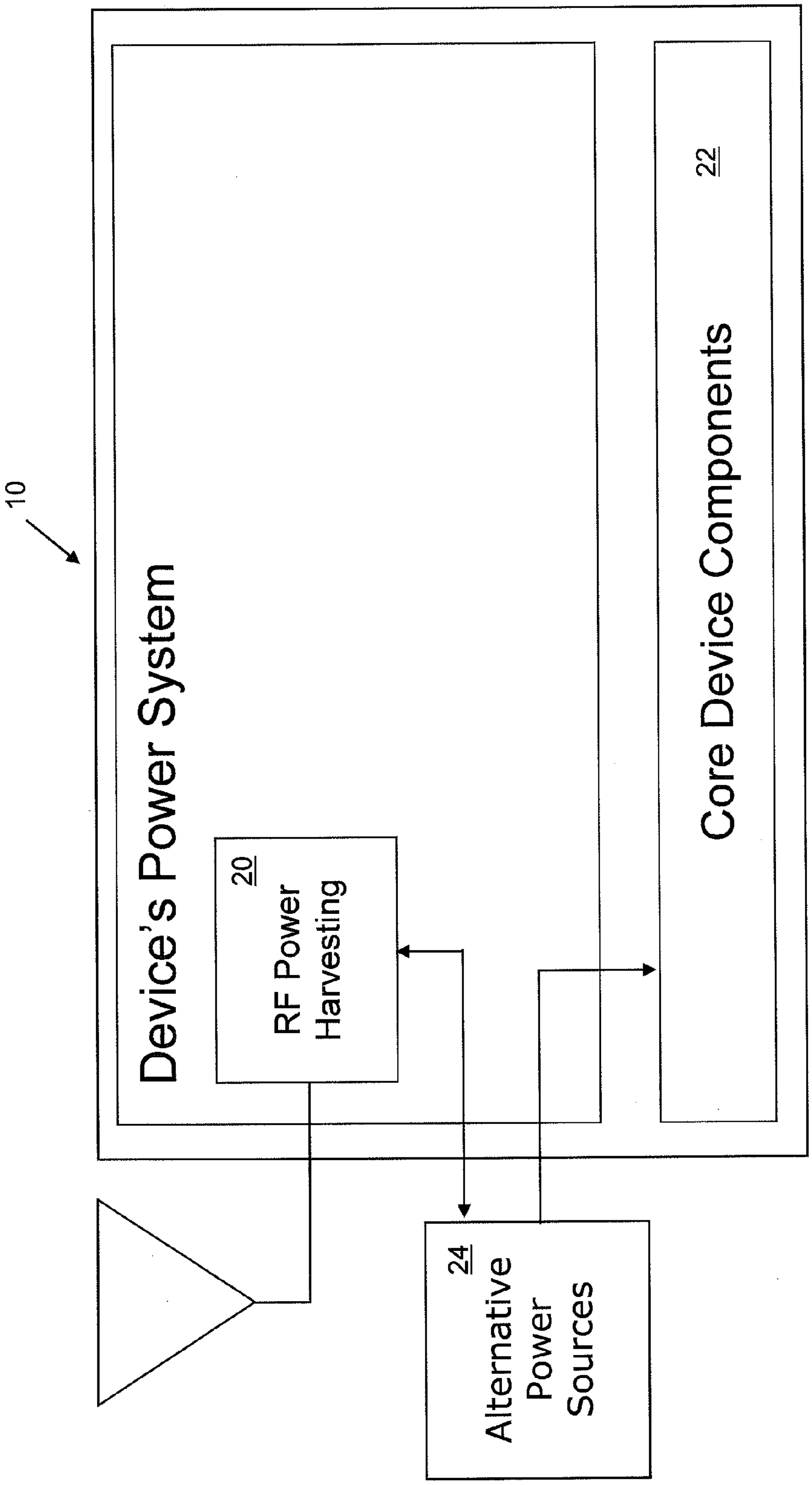
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 63

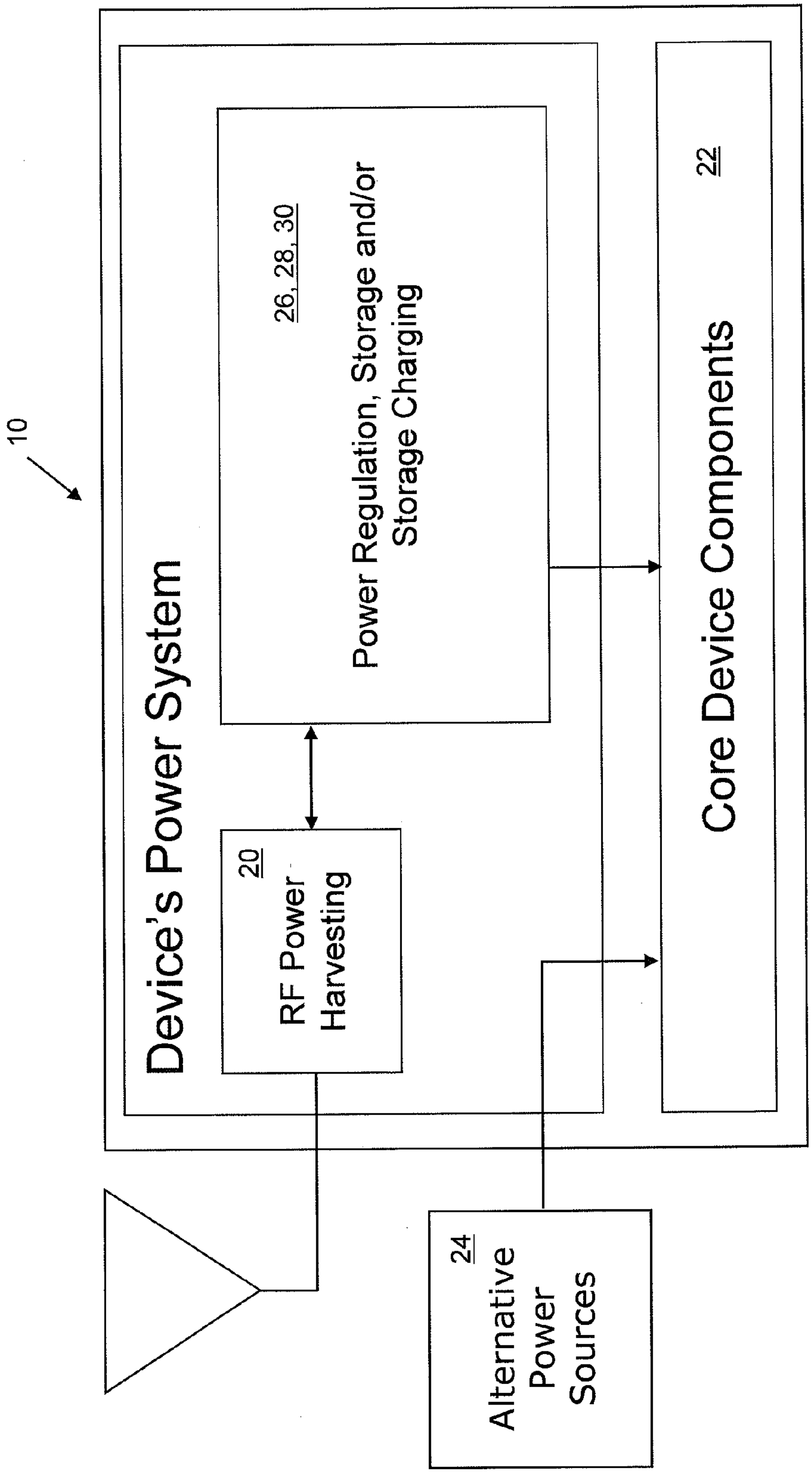
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 64

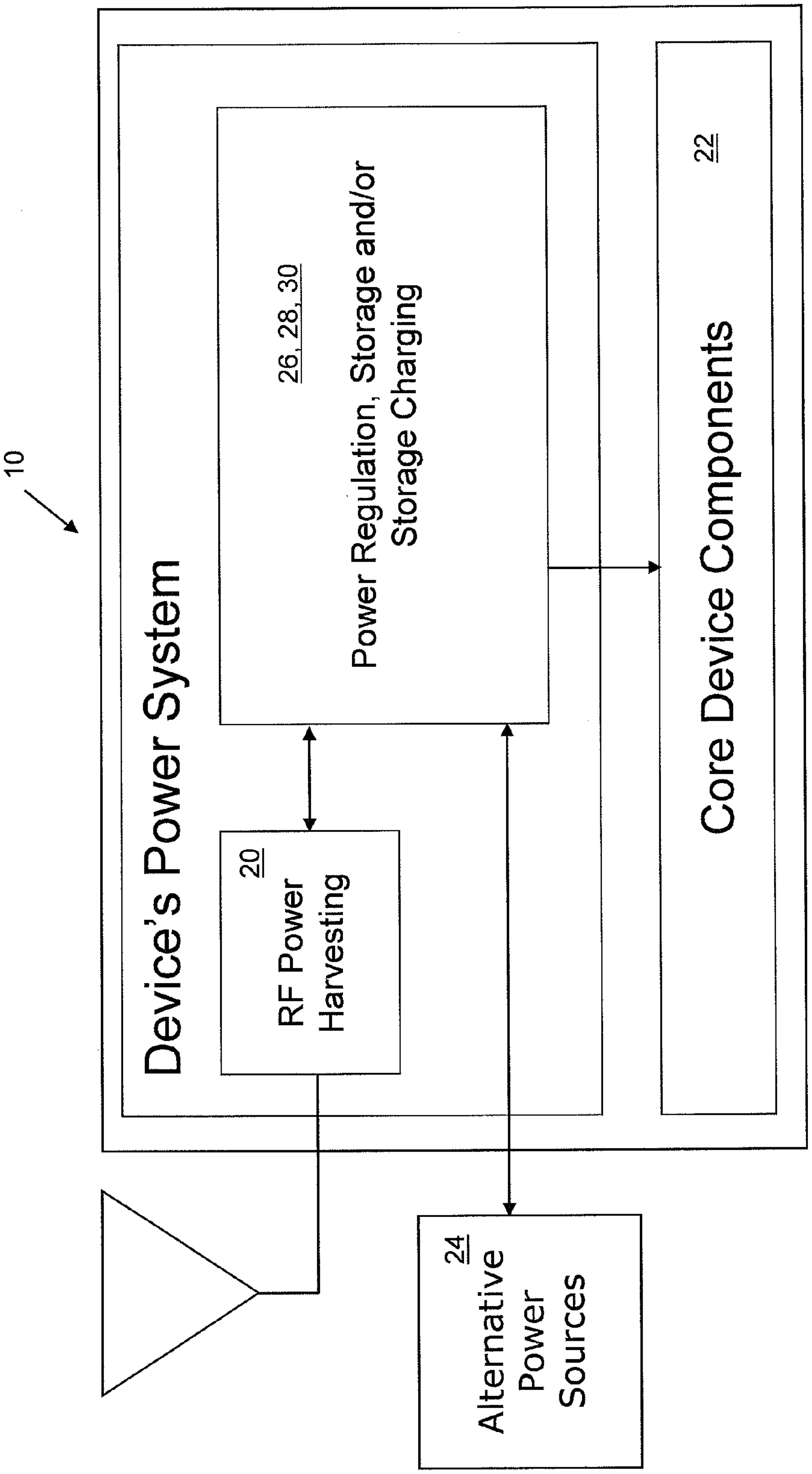
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 65

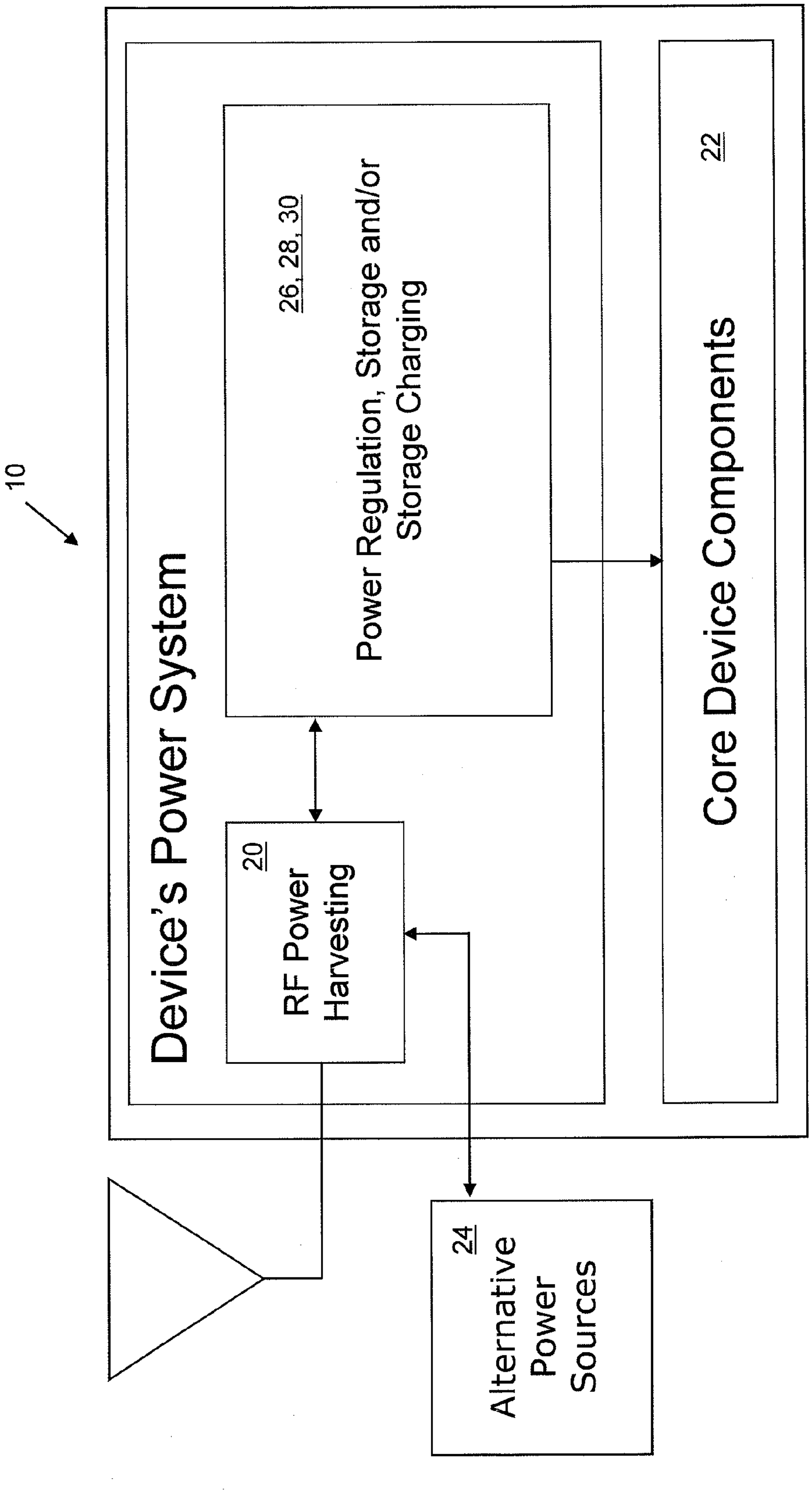
10
Device's Power System
Power Regulation, Storage and/or Storage Charging
26, 28, 30
RF Power Harvesting
20
Core Device Components
22
Alternative Power Sources
24

Fig. 66

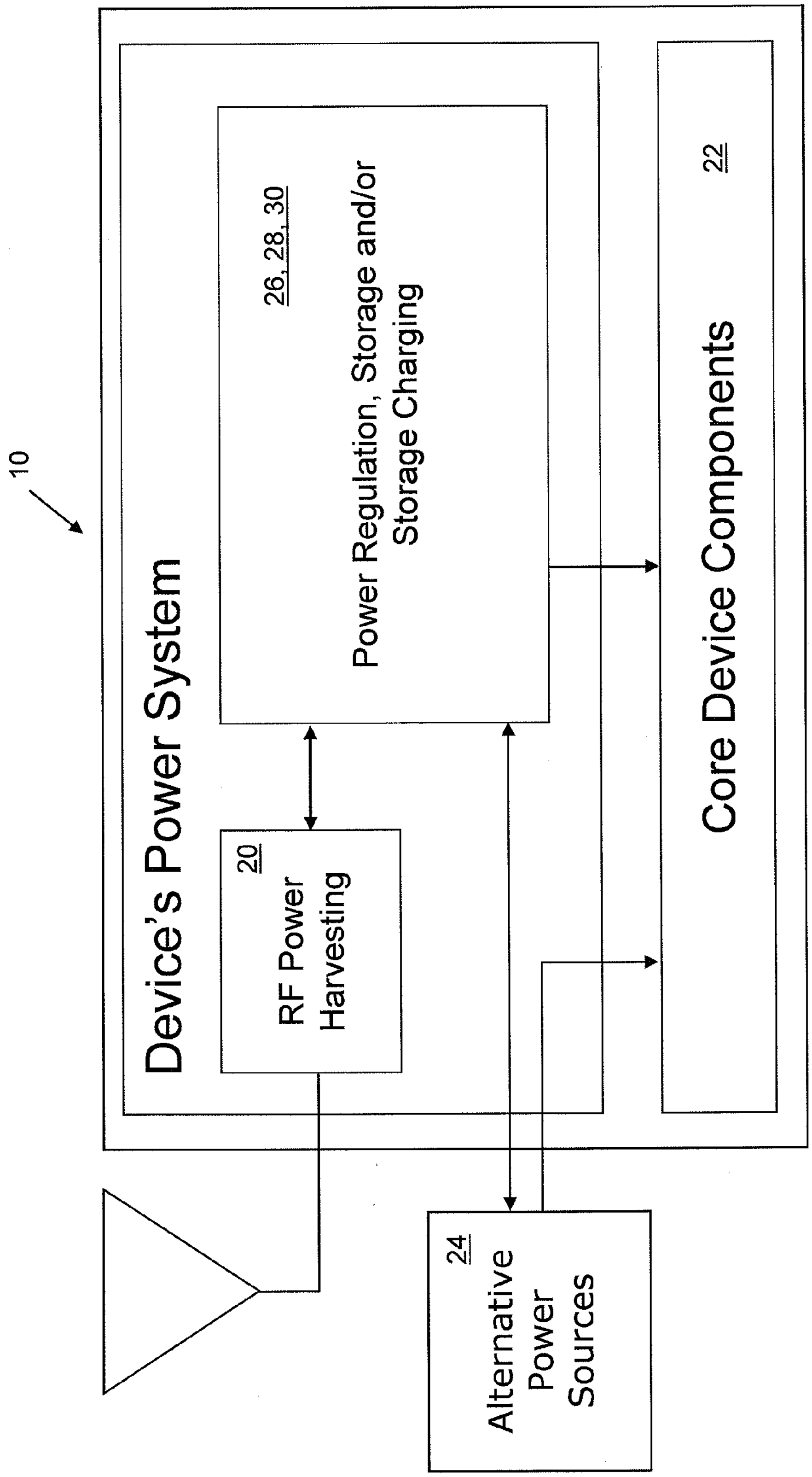
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 67

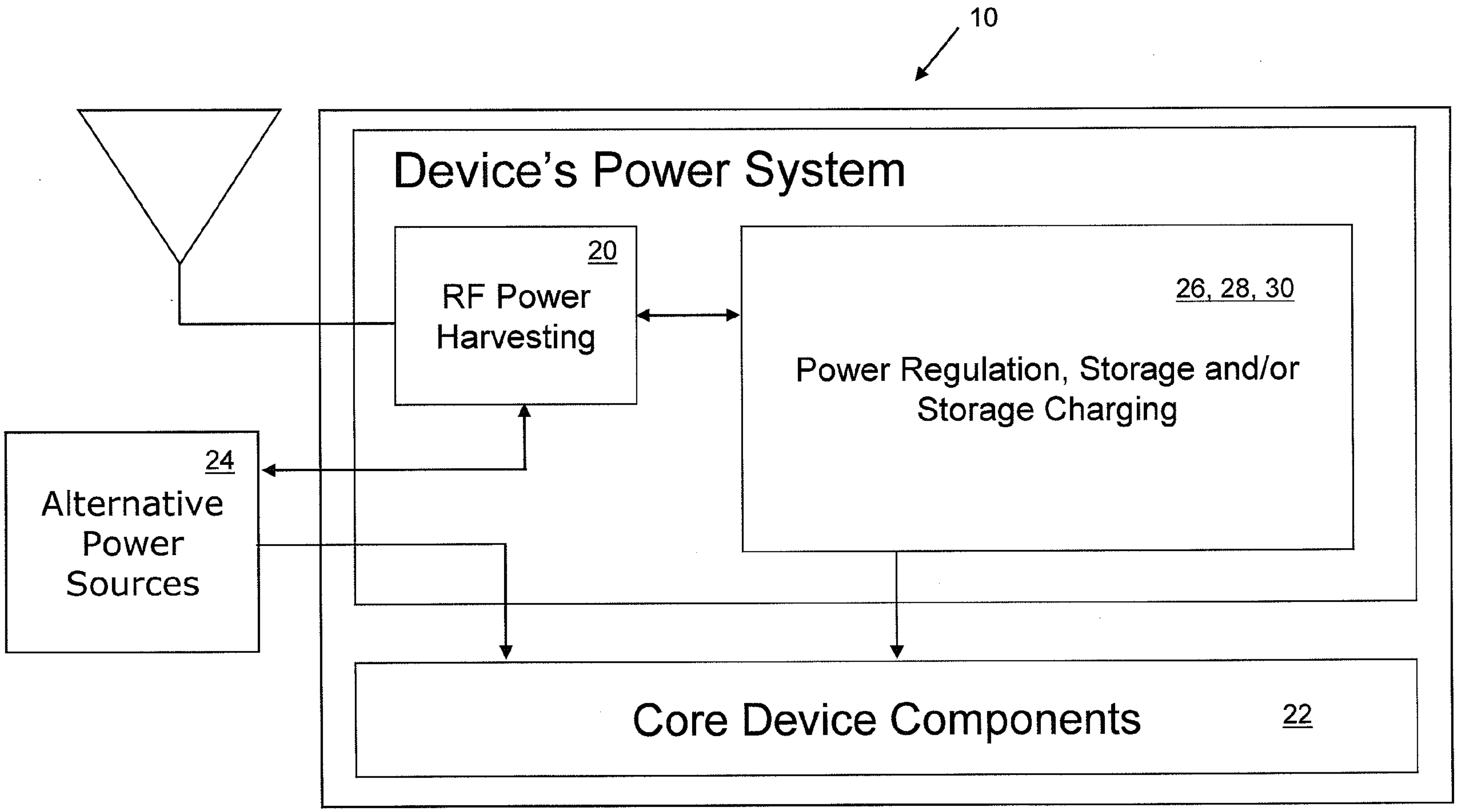
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 68

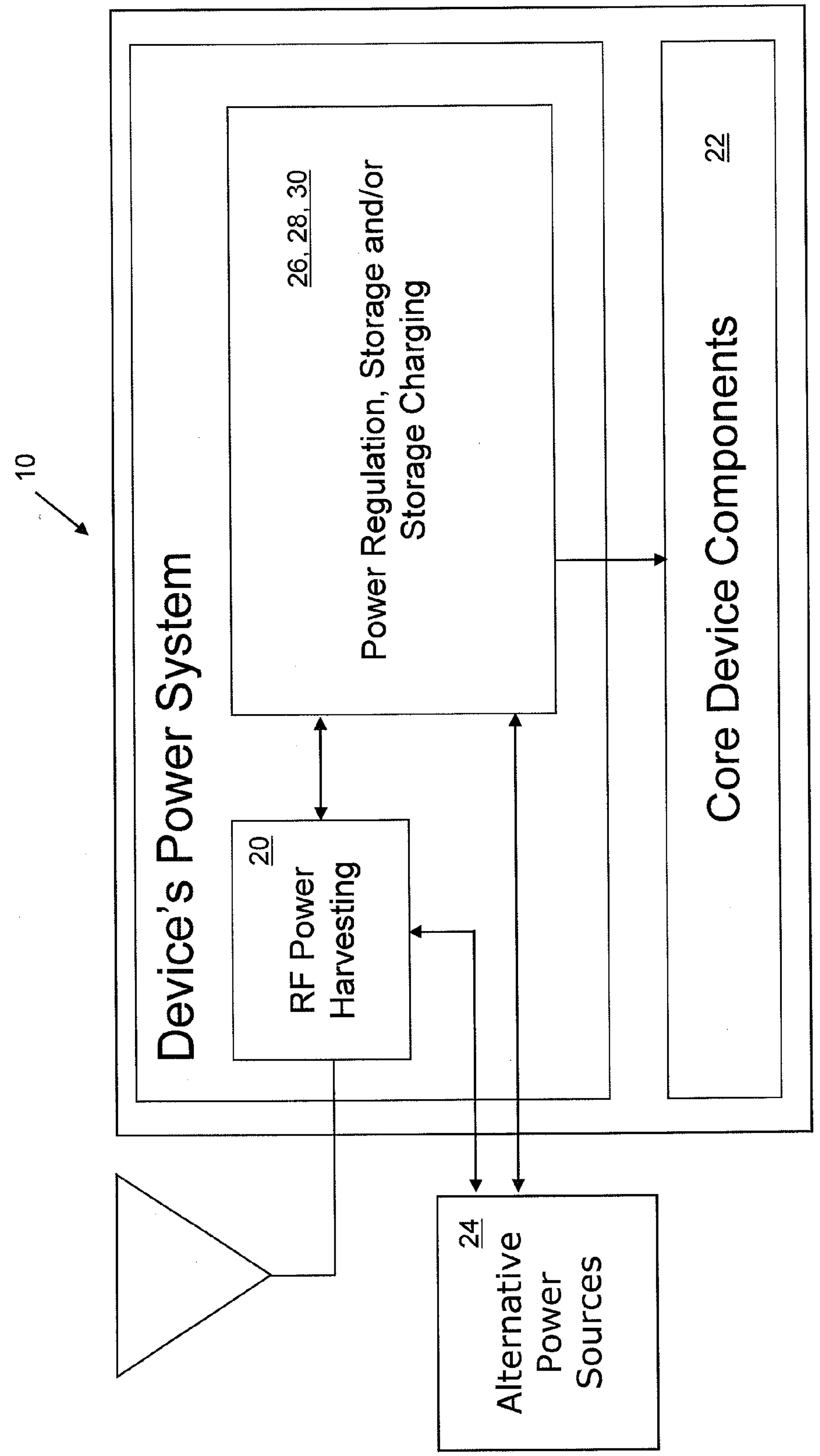
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 69

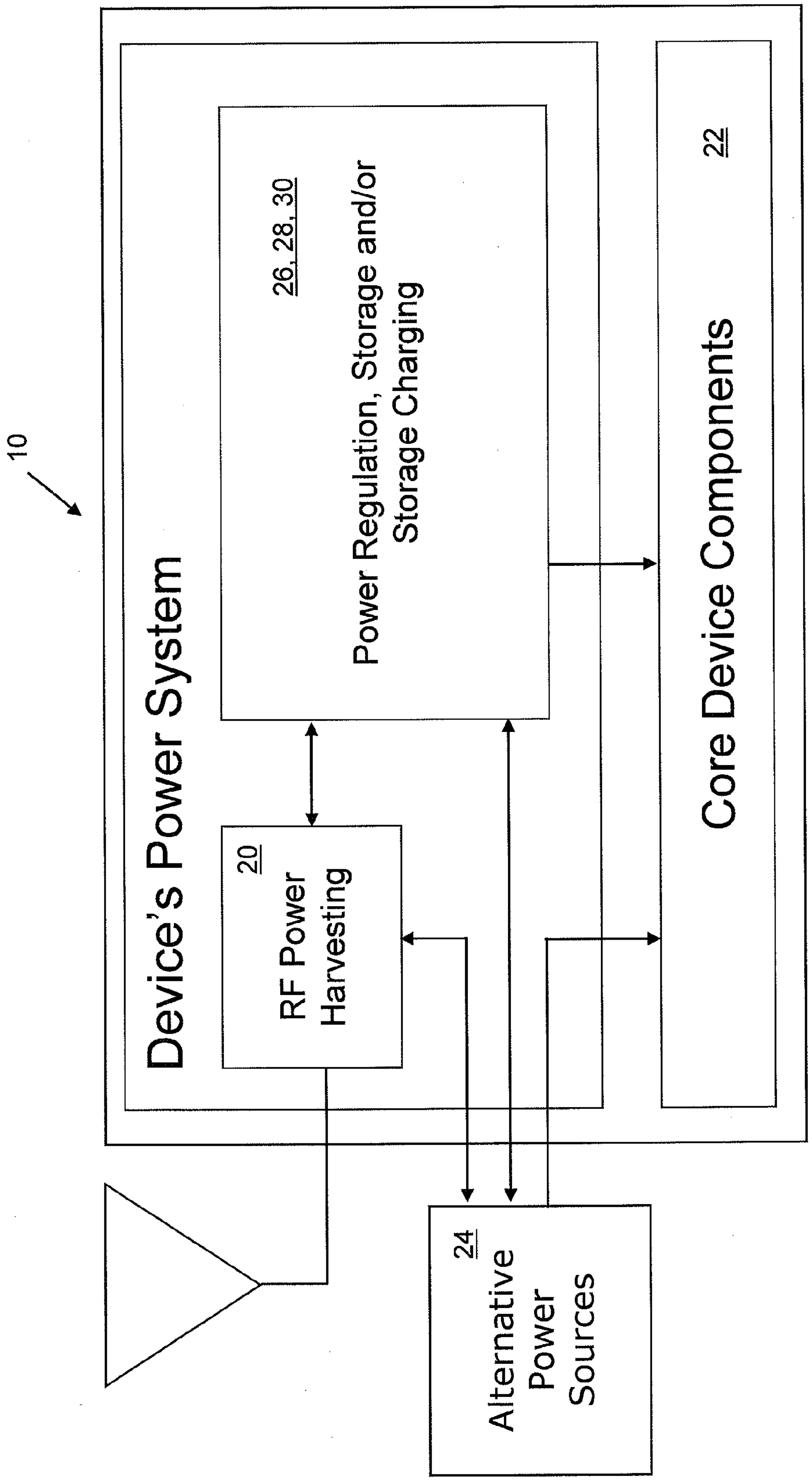
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 70

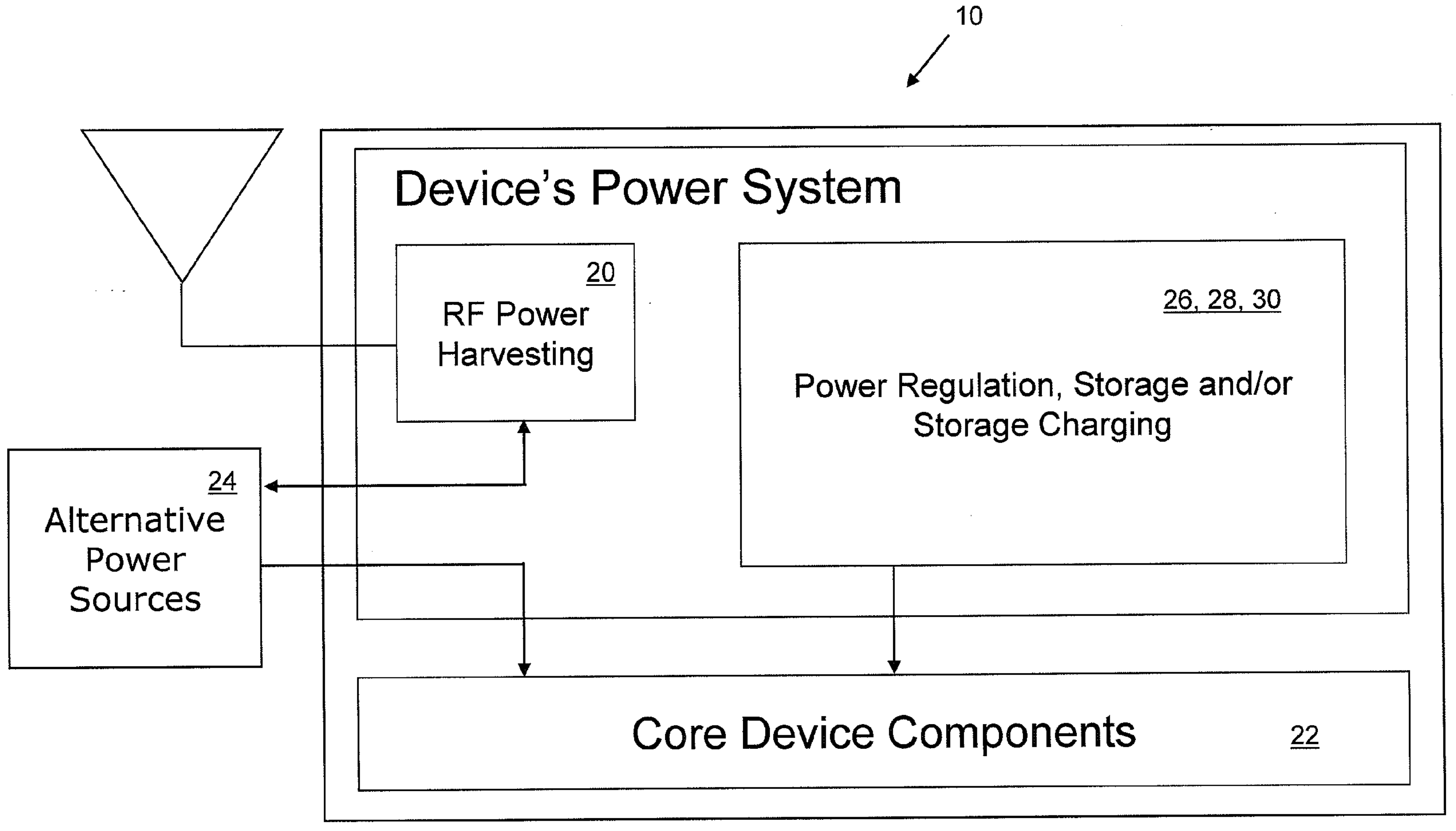
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 71

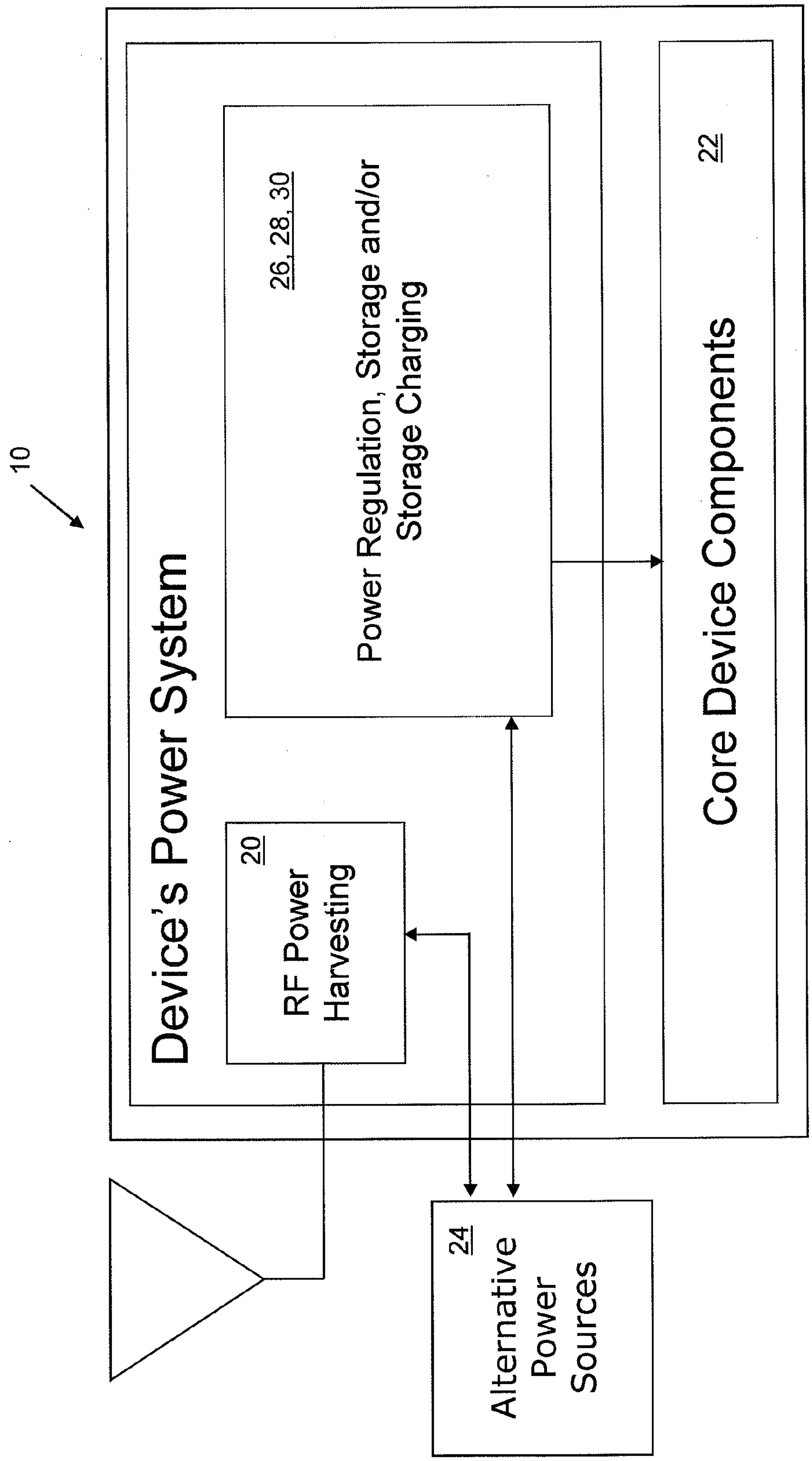
10
Device's Power System
RF Power Harvesting
20
Power Regulation, Storage and/or Storage Charging
26, 28, 30
Core Device Components
22
Alternative Power Sources
24

Fig. 72

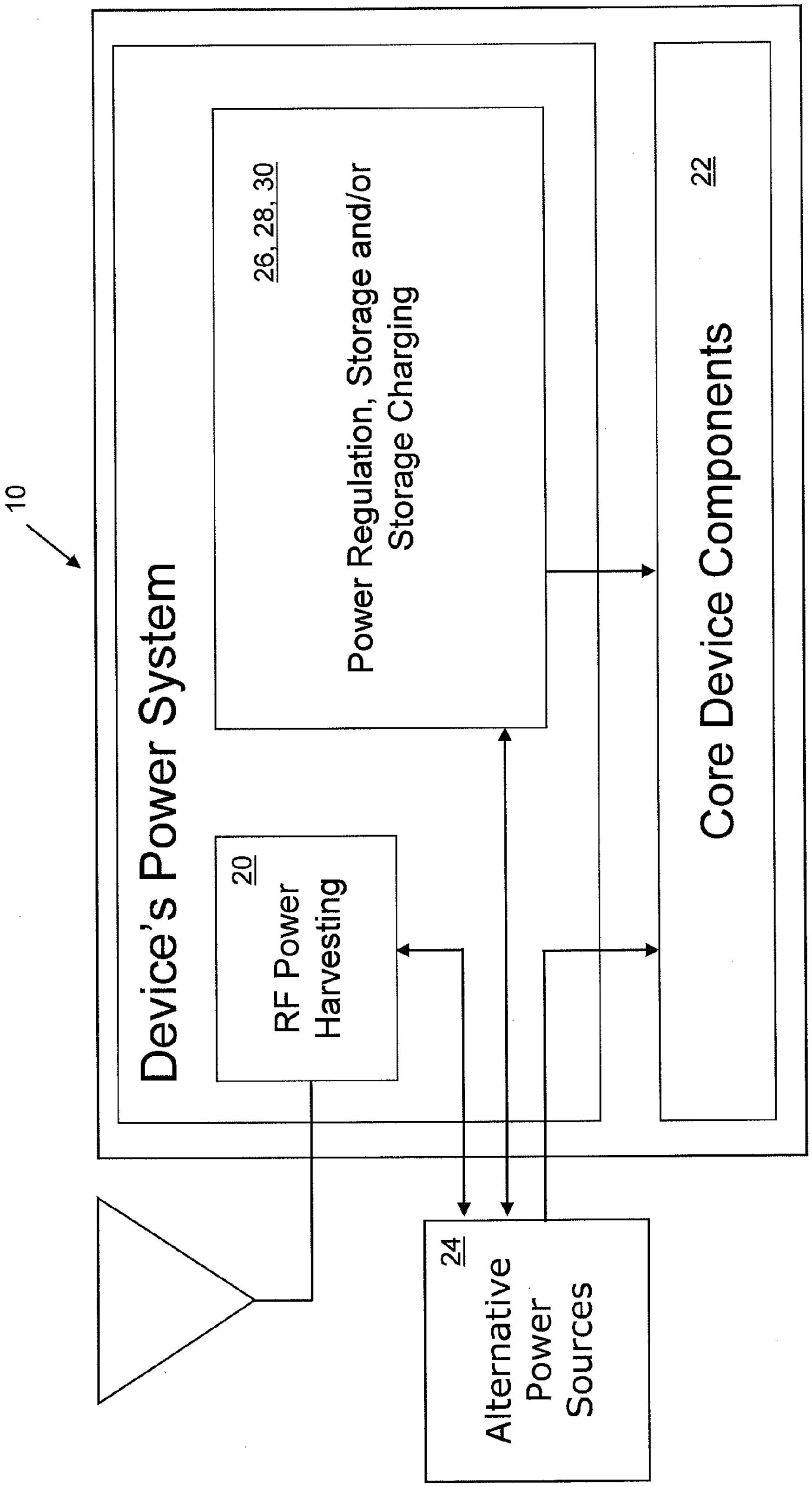
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 73

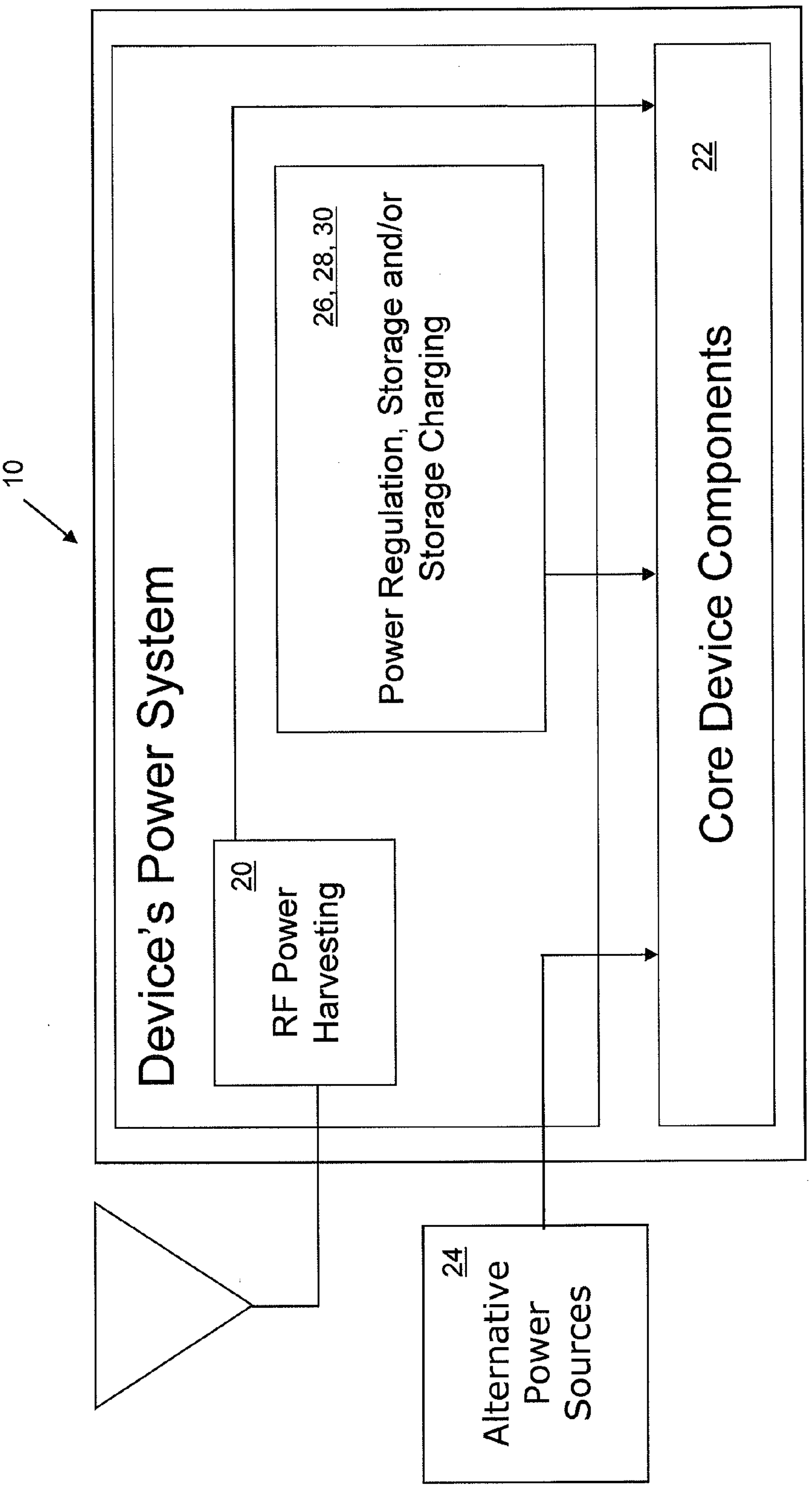
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 74

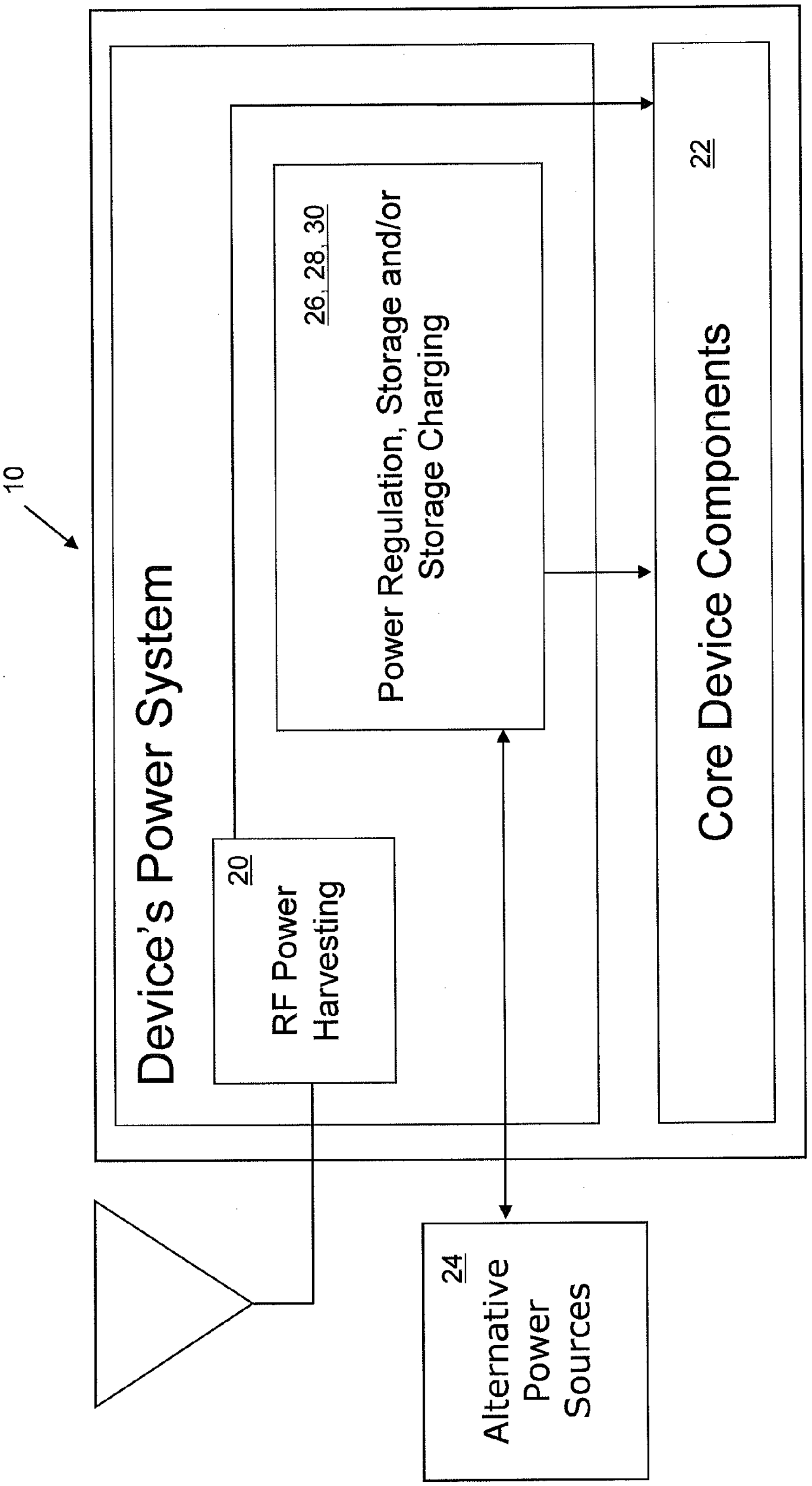
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 75

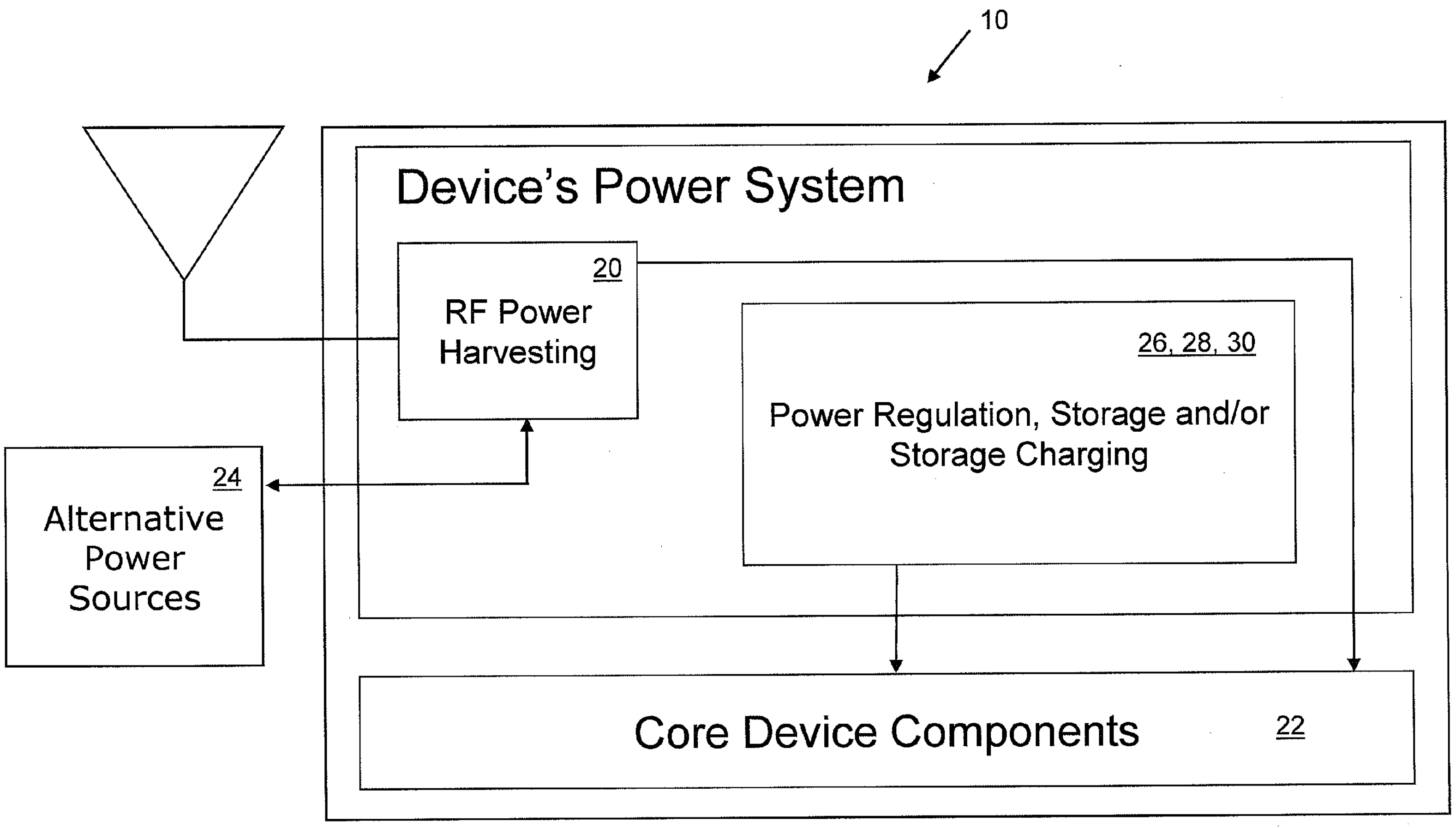
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 76

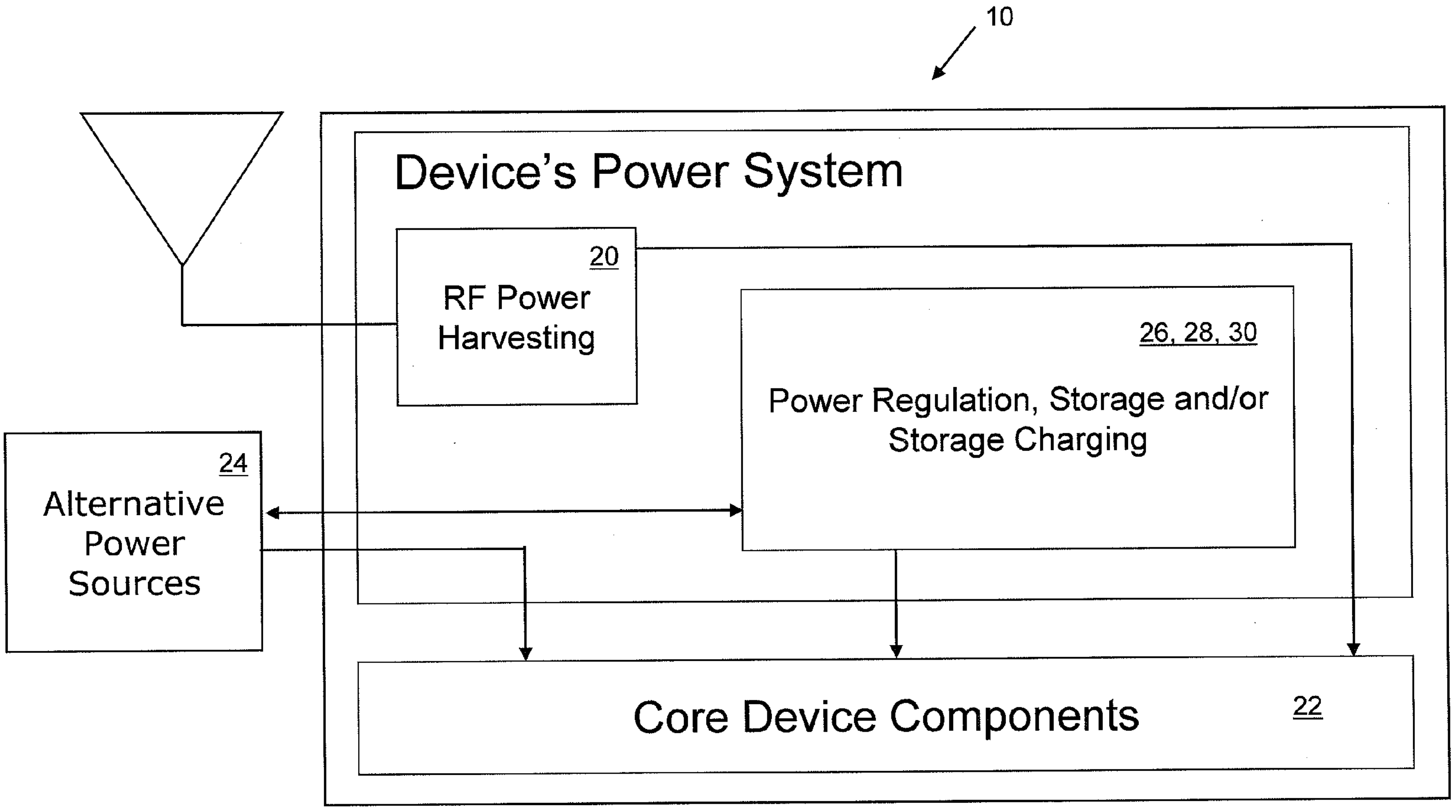
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 77

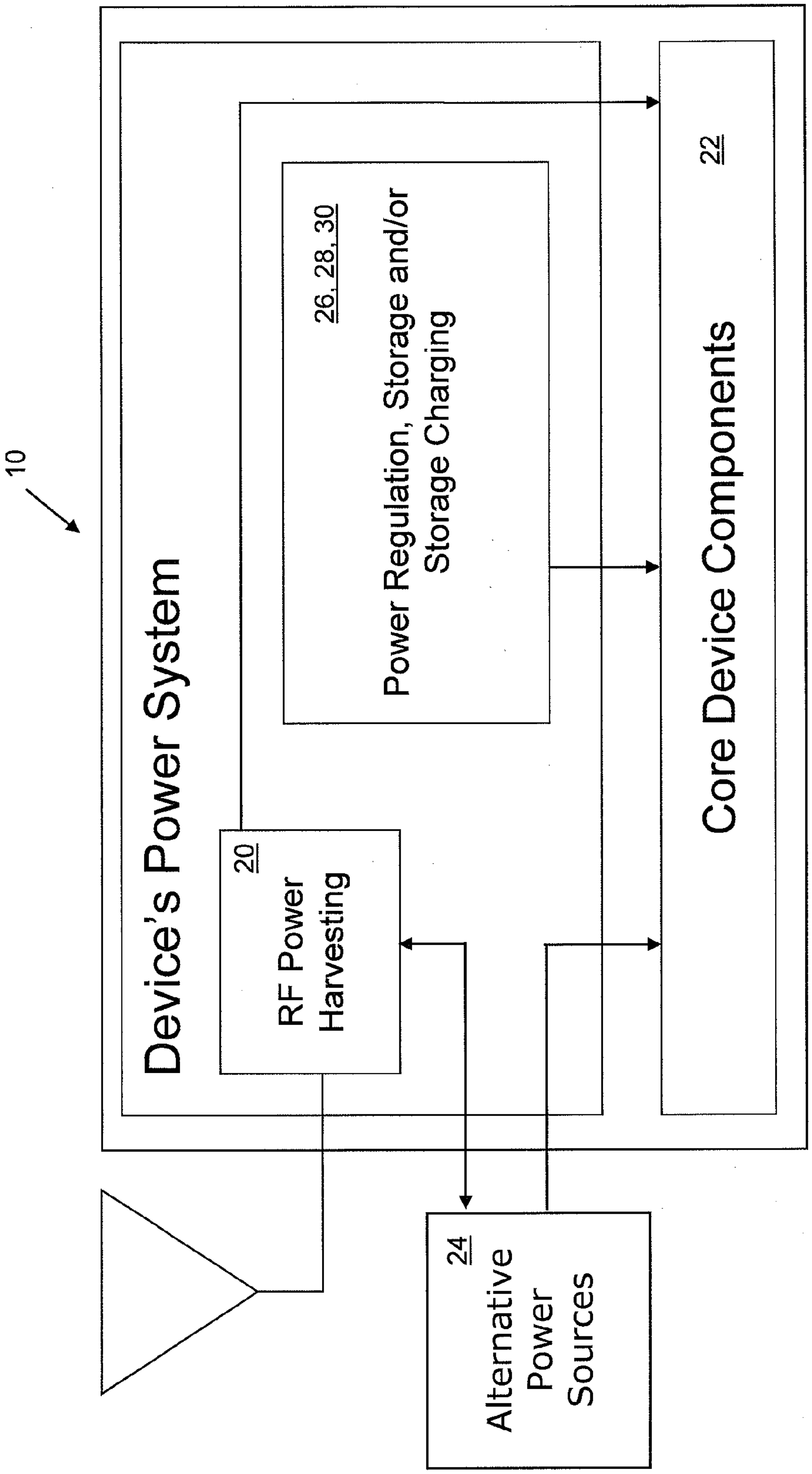
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 78

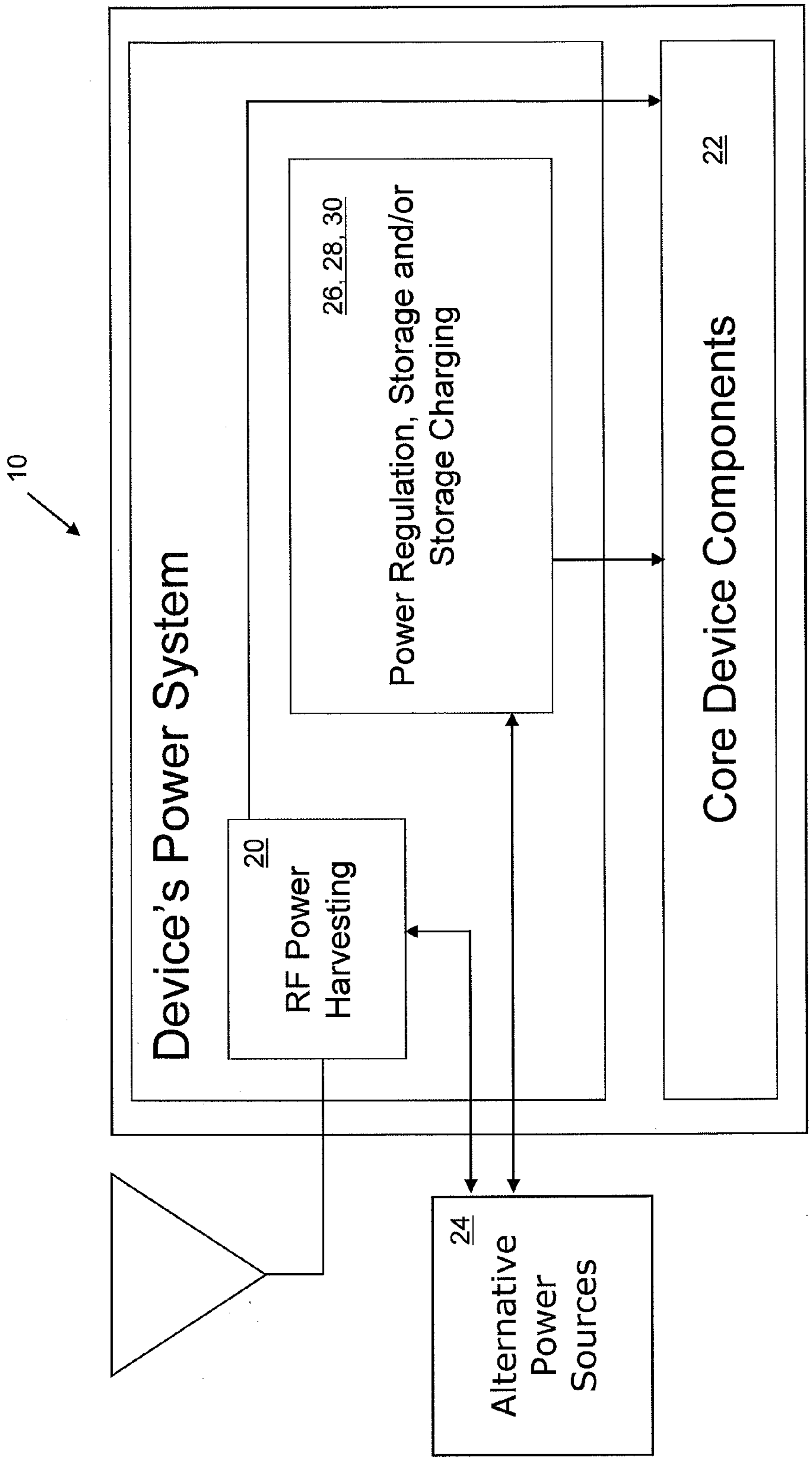
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 79

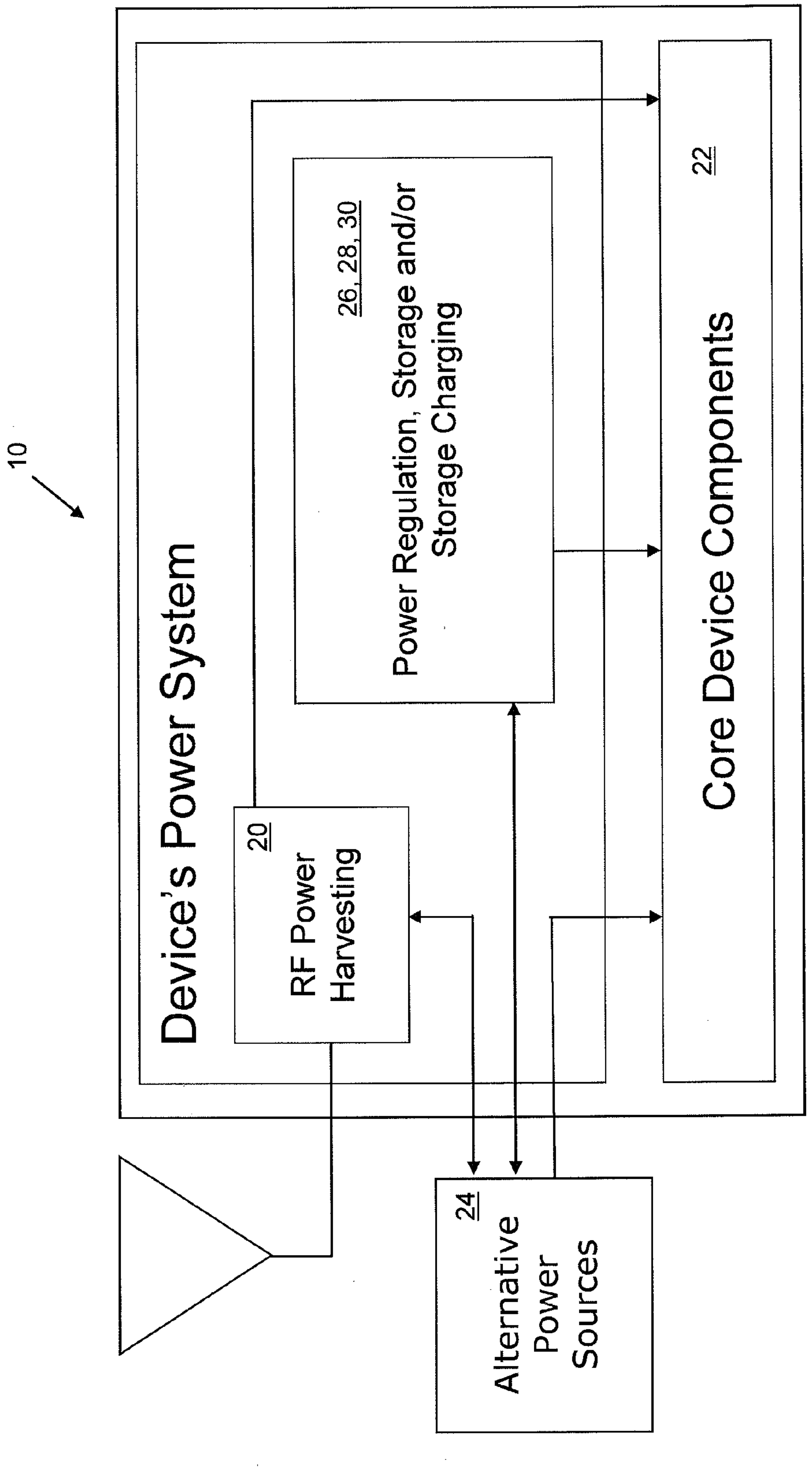
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 80

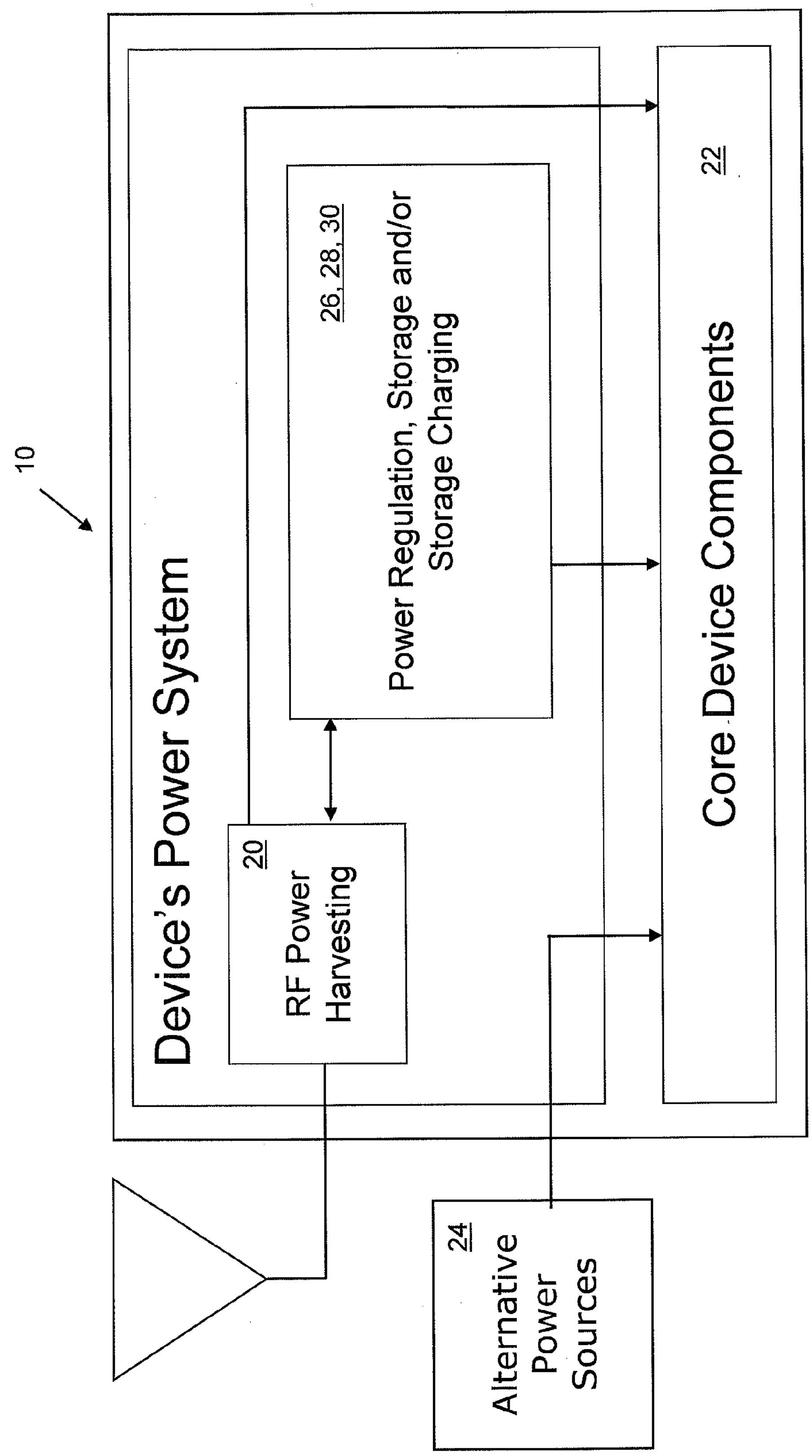
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 81

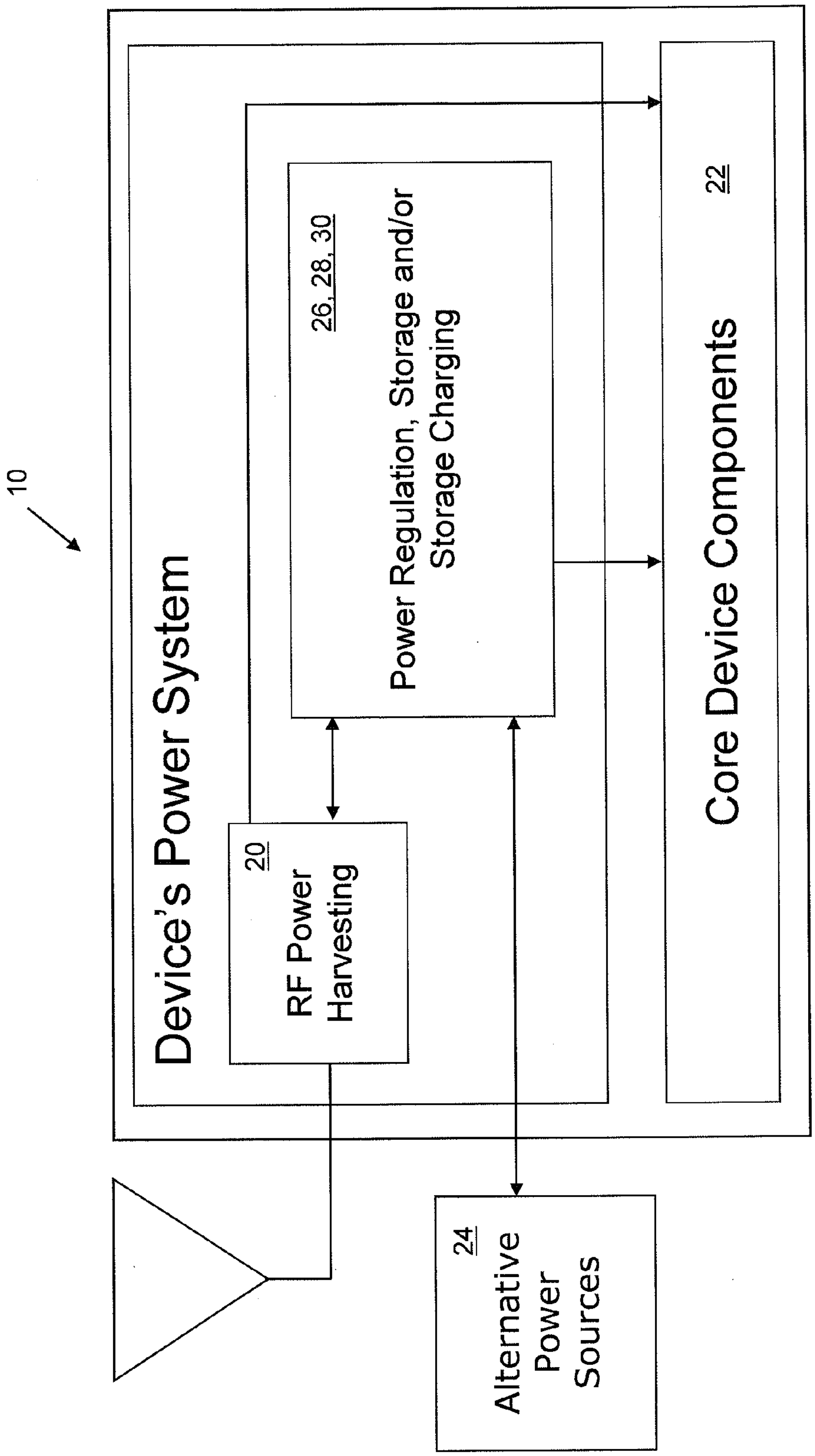
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 82

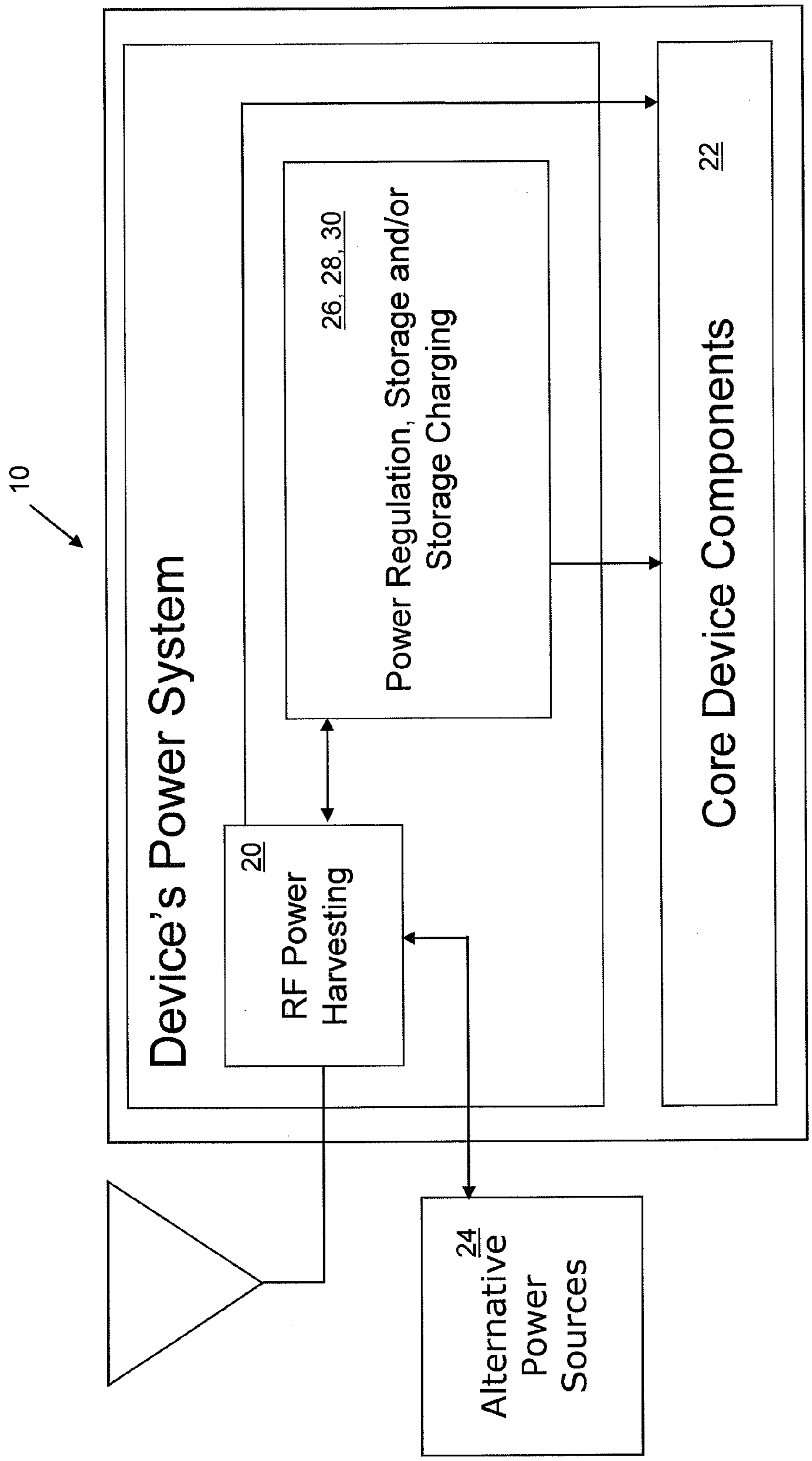
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 83

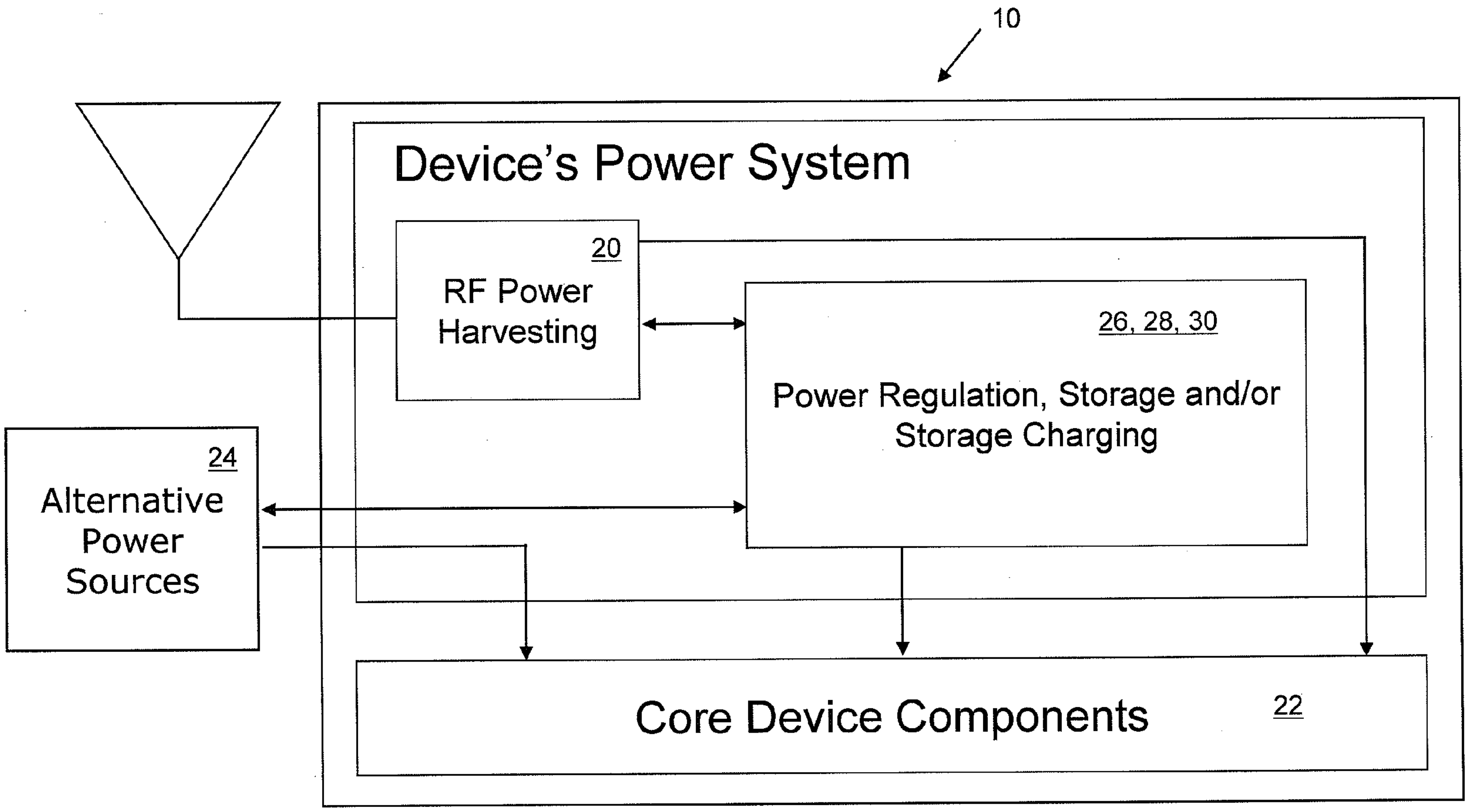
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 84

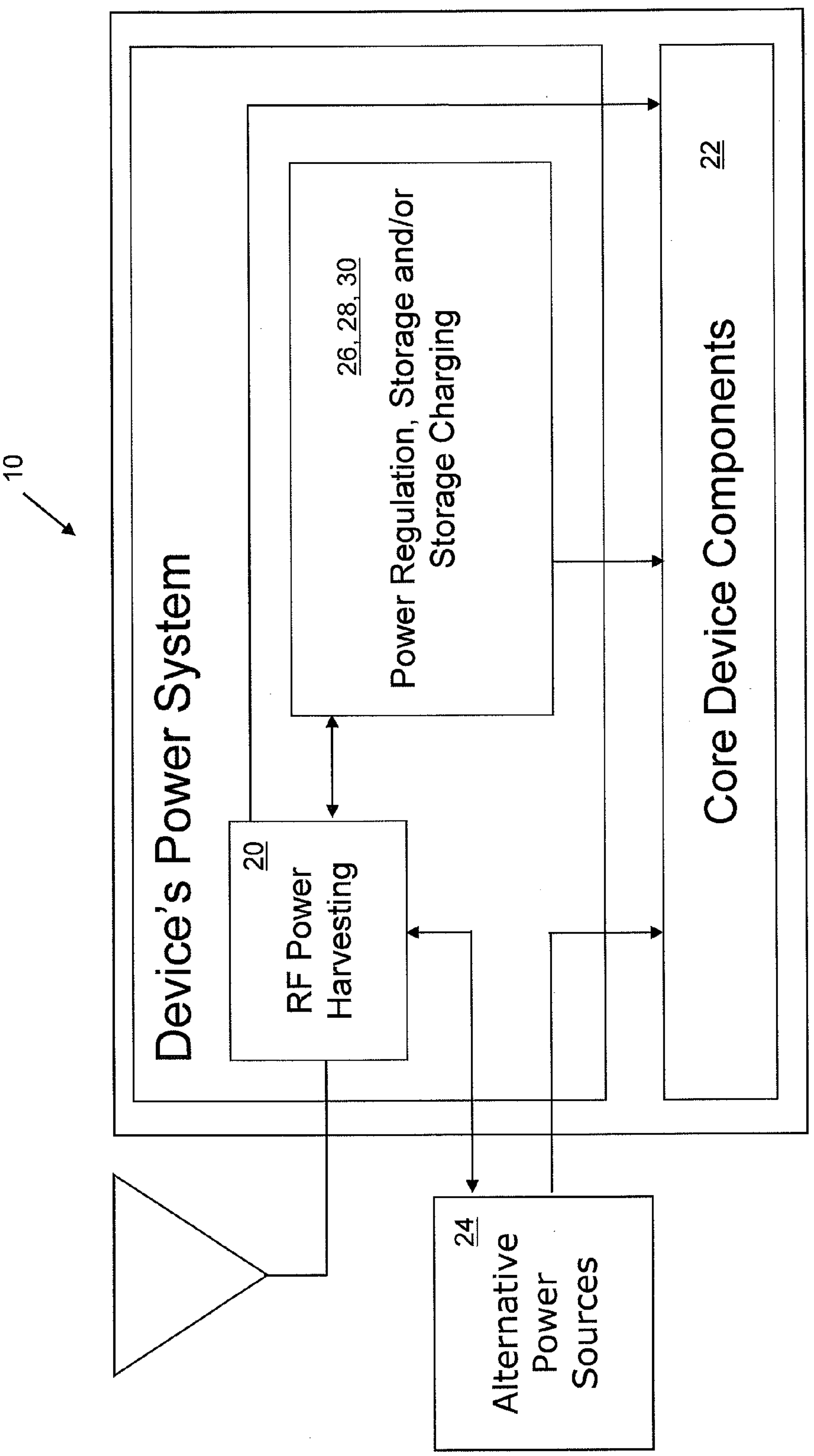
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 85

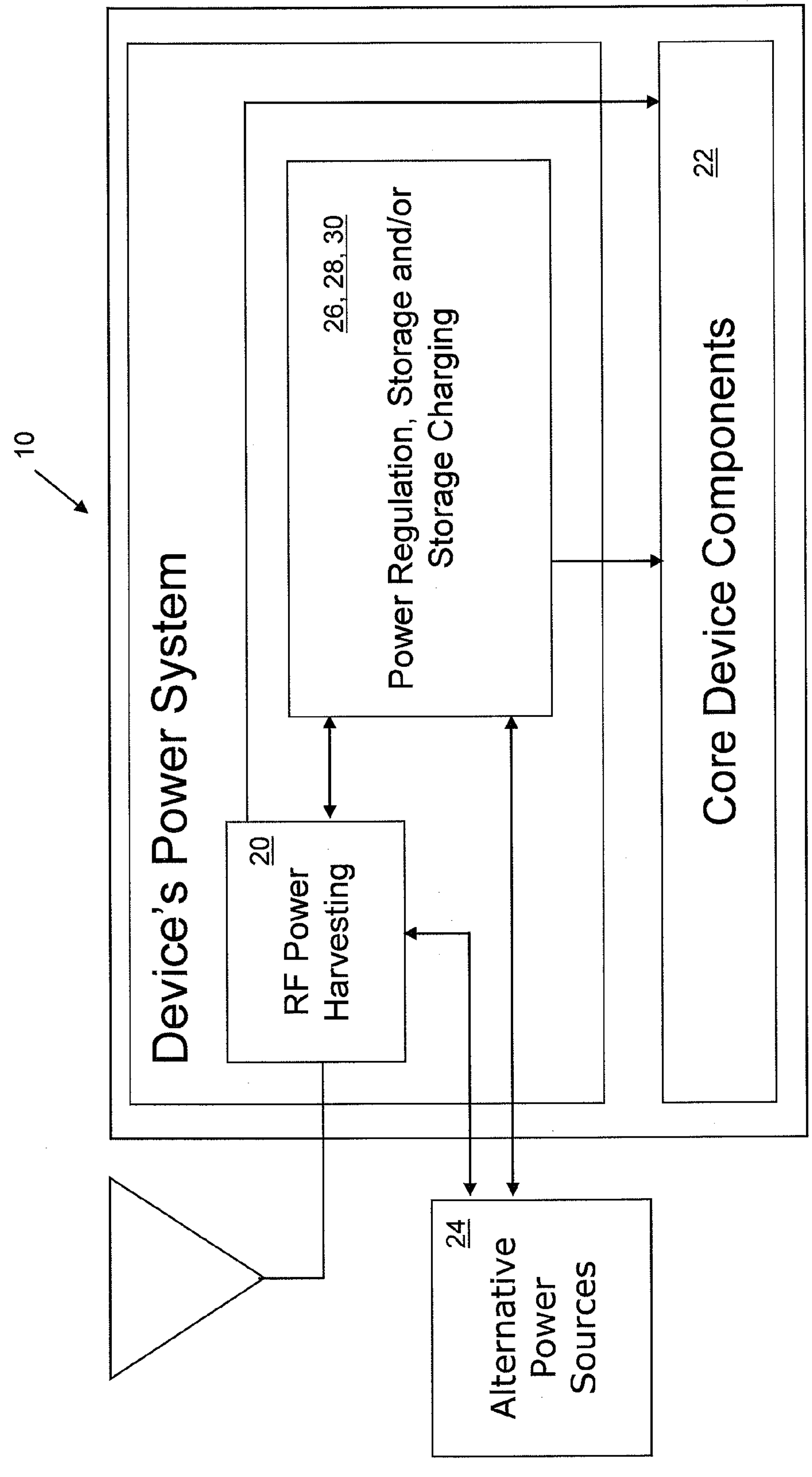
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 86

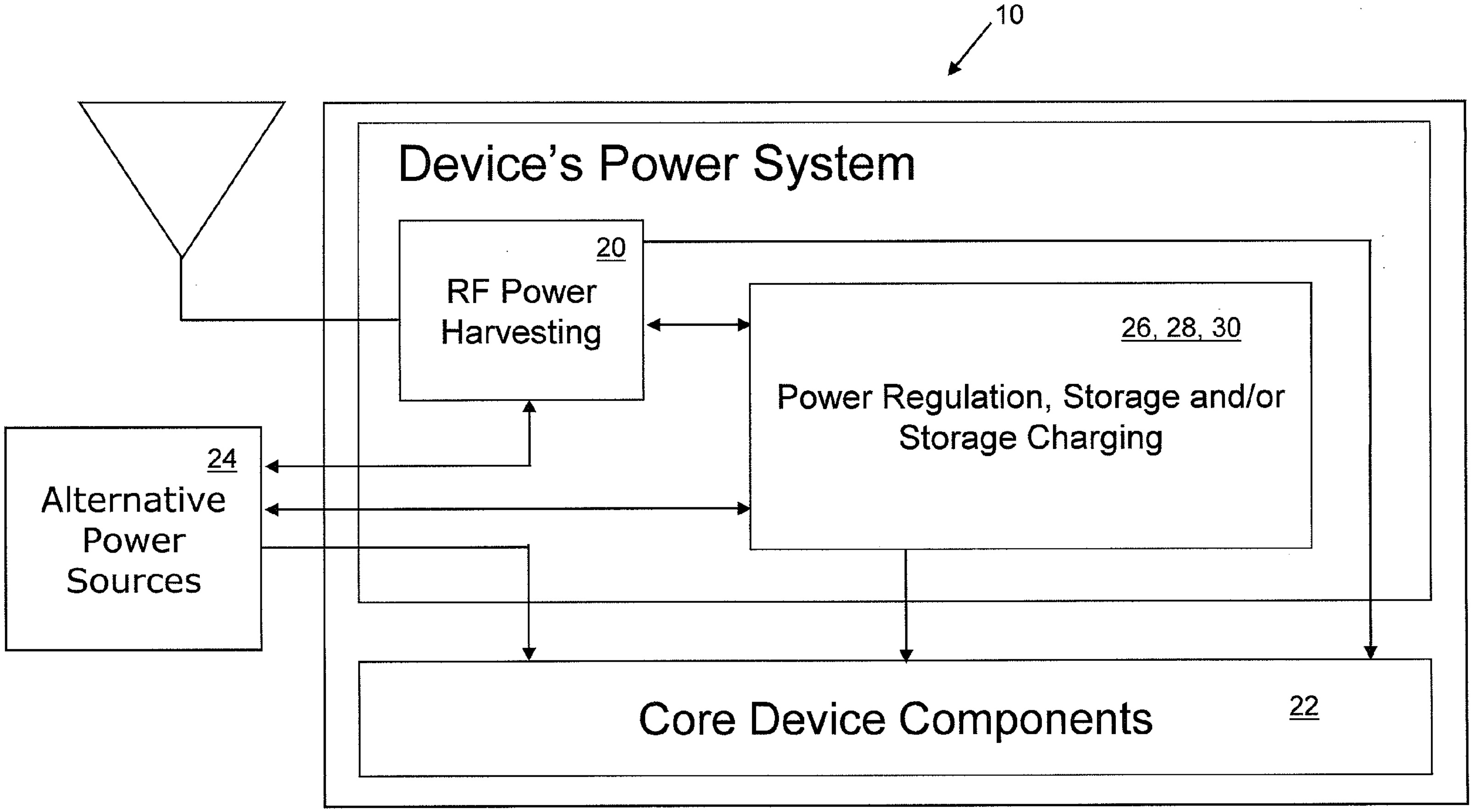
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 87

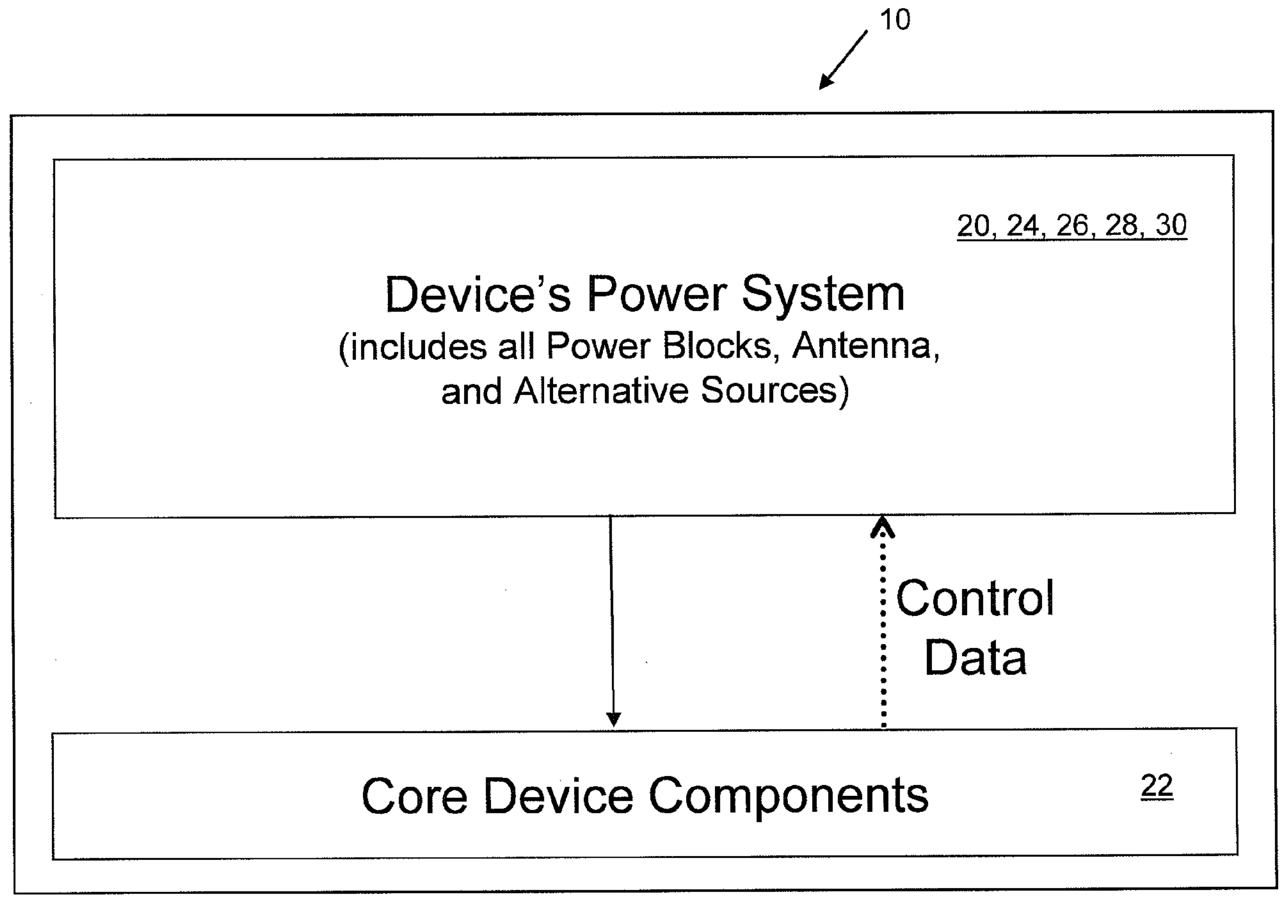
10
20, 24, 26, 28, 30
Device's Power System
(includes all Power Blocks, Antenna, and Alternative Sources)
Control Data
Core Device Components
22

Fig. 88

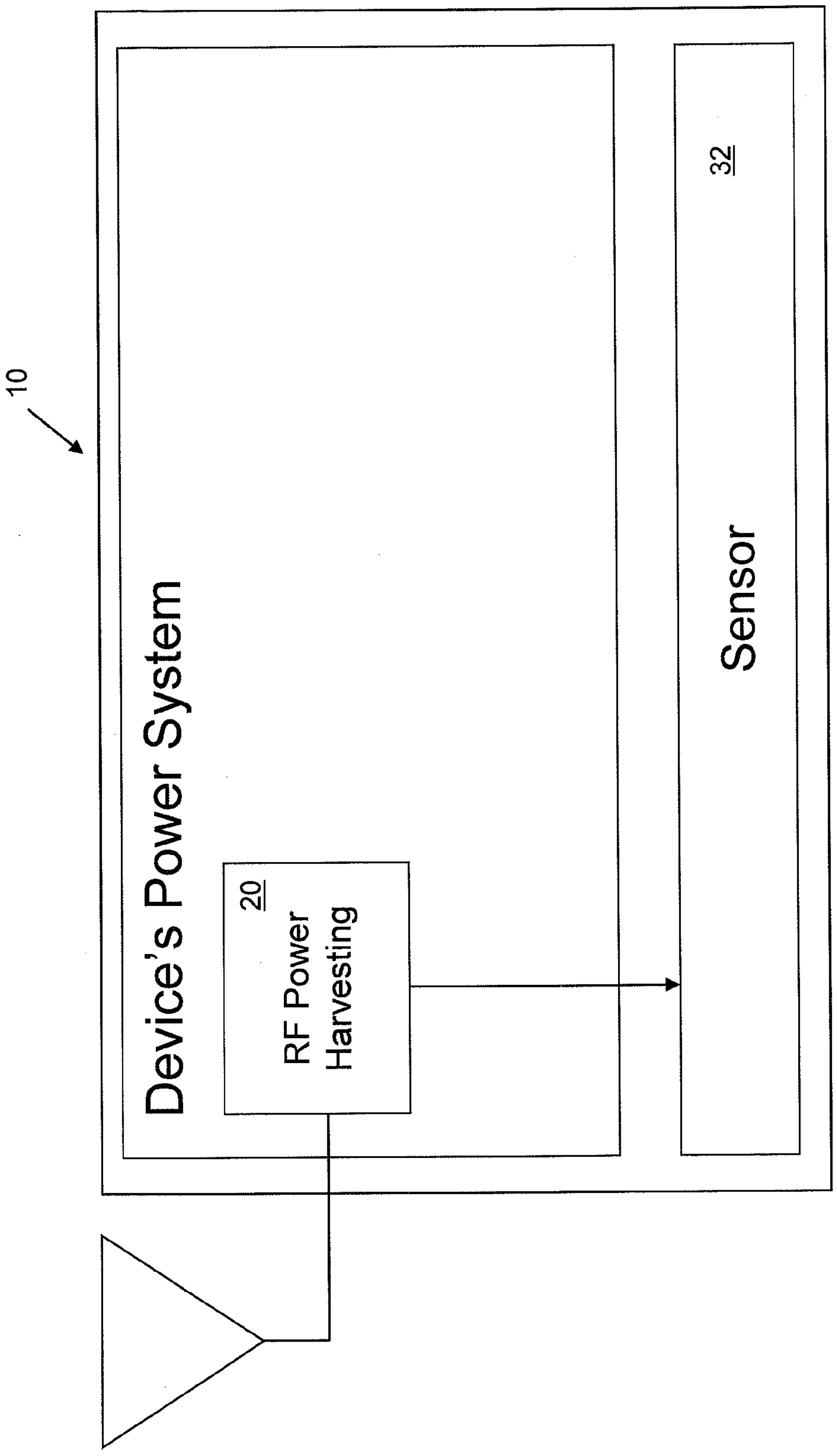
10
Device's Power System
20
RF Power Harvesting
Sensor
32

Fig. 89

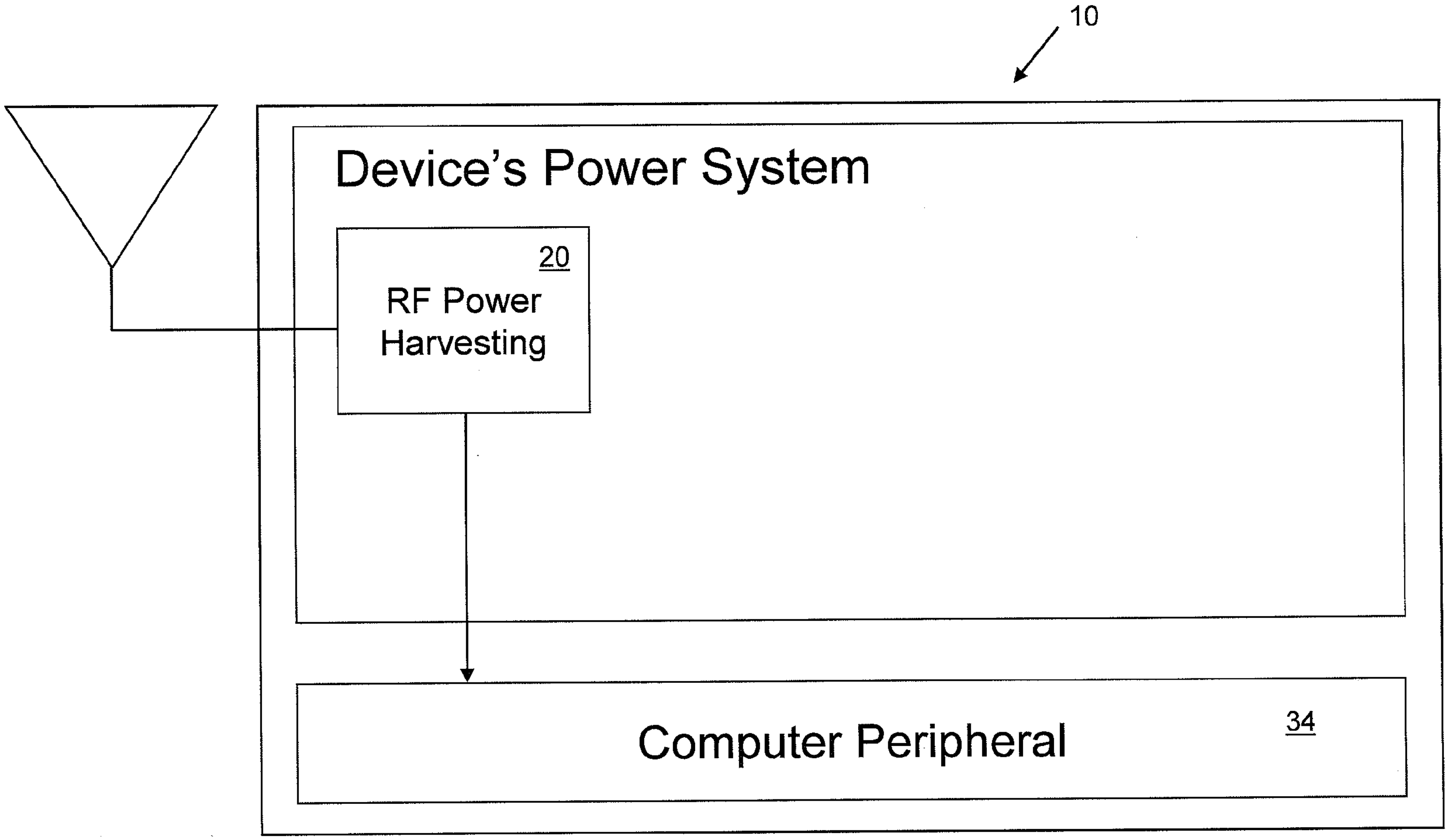
10
Device's Power System
20
RF Power Harvesting
Computer Peripheral
34

Fig. 90

POWERING DEVICES USING RF ENERGY HARVESTING

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to and is a continuation application of U.S. patent application Ser. No. 12/499,618, filed Jul. 8, 2009, which is a divisional application of U.S. patent application Ser. No. 11/447,412, filed Jun. 6, 2006, which claims priority to U.S. Provisional Patent Application No. 60/688,587, filed on Jun. 8, 2005; each of which is incorporated herein by reference in its entirety.

BACKGROUND

[0002] The present invention is related to the wireless powering of devices. More specifically, the present invention is related to the wireless powering of devices with a power harvester.

[0003] As processor capabilities have expanded and power requirements have decreased there has been an ongoing explosion of devices that operate completely independent of wires or power cords. These "untethered" devices range from cell phones, and wireless keyboards to building sensors and active RFID tags.

[0004] Engineers and designers of these untethered devices continue to have to deal with the limitations of portable power sources, primarily batteries as the key design parameter. While performance of processors and portable devices have been doubling every 18-24 months driven by Moore's law, battery technology in terms of capacity has only been growing at measly 6% per year. Even with power conscious designs and the latest in battery technology, many devices do not provide the lifetime cost and maintenance requirements for applications that require a large number of untethered devices such as logistics, and building automation. Today's devices that need two-way communication require scheduled maintenance every three to 18 months to replace or recharge the device's power source (typically a battery). One-way devices simply broadcasting their status (one-way) such as automated utility meter readers have a better battery life, typically requiring replacement within 10 years. For both device types, scheduled power-source maintenance is costly and disruptive to the entire system that a device is intended to monitor and/or control. Unscheduled maintenance trips are even more costly and disruptive. On a macro level, the relatively high cost associated with the internal battery also reduces the practical, or economically viable, number of devices that can be deployed.

[0005] The ideal solution to the power problem for untethered devices is a device or system that can collect and harness sufficient energy from the external environment. The harnessed energy would then either directly power an untethered device or augment a battery or other storage component. Directly powering an untethered device enables the device to be constructed without the need for a battery. Augmenting a storage component could be along two lines: 1) increasing the overall life of the device or 2) by providing more power to the device to increase the functionality of the device. The other parameters for an ideal solution is that the harnessing device could be used in a wide range of environments including harsh and sealed environments (e.g. nuclear reactors), would be inexpensive to produce, would be safe for humans, and would have a minimal effect on the basic size, weight and other physical characteristics of the untethered device.

SUMMARY

[0006] The present invention pertains to an apparatus for an application. The apparatus comprises a core device preferably having an integrated circuit for the application. The apparatus comprises a power harvester (preferably a radio frequency (RF) energy harvester) connected to the core device to power the core device.

[0007] The present invention pertains to an apparatus for an application. The apparatus comprises a core device preferably having an integrated circuit for the application. The apparatus comprises means for receiving energy wirelessly and providing power from the energy to the core device to power the integrated circuit of the core device. The receiving means is connected to the core device.

[0008] The present invention pertains to a method for an application. The method comprises the steps of converting RF energy into usable energy. There is the step of preferably powering an integrated circuit of the core device with the usable energy.

[0009] This invention pertains to a technique that uses radio frequency (RF) energy as a source of energy to directly power a device or augment a power storage component in a device. The present invention meets the requirements described in the previous "Background of the Invention" section.

[0010] Traditional RF receiving devices have focused on maximizing selectivity of the frequency to isolate and to be coherent without interference from other sources. In contrast, while the present invention operates at a specific frequency or range of frequencies, the device accepts any interference to supplement the output power of the device. Also, the research related to power harvesting that uses RF energy as a source has primarily focused on devices in close proximity of the source (inductive or near-field energy). In most cases, prior research assumed a dedicated or directed source of RF to power the device.

[0011] The invention should not be confused with power transfer by inductive coupling, which requires the device to be relatively close to the power transmission source. The RFID Handbook by the author Klaus Finkenzeller defines the inductive coupling region as distance between the transmitter and receiver of less than 0.16 times lambda where lambda is the wavelength of the RF wave. The invention can be implemented in the near field (sometimes referred to as inductive) region as well as the far-field region. The far-field region is distances greater than 0.16 times lambda.

[0012] It is an object of this invention to provide a method and apparatus to

- **[0013]** 1. remotely energize an untethered device without using direct wiring
- **[0014]** 2. power or augment the life of the power storage component so it matches the life of the device and, ultimately, powers the off-grid device with or without the use of batteries
- **[0015]** 3. allow untethered devices to be virtually maintenance free
- **[0016]** 4. provide augmentation for other energy harvesting technologies (solar, piezoelectric, etc.)
- **[0017]** 5. provide backup power to tethered devices

[0018] It is a further object of this invention to directly power a device or augment a power storage component in a device in conjunction with other power harvesting technologies and storage elements.

[0019] With this method and apparatus a device's power storage components do not require replacement, thus enabling the device to be permanently placed off-grid, where it may be physically impractical, costly, or dangerous (due to a harsh environment) to provide maintenance.

[0020] For devices on-grid (tethered) or with reliable power sources, RF power harvesting can be used as a backup in case the primary power source is lost.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] In the accompanying drawings, the preferred embodiment of the invention and preferred methods of practicing the invention are illustrated in which:

[0022] FIG. **1** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

[0023] FIG. **2** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0024] FIG. **3** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0025] FIG. **4** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0026] FIG. **5** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0027] FIG. **6** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0028] FIG. **7** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0029] FIG. **8** is a block diagram of an RF Power Harvesting block in communication with a Power Storage block.

[0030] FIG. **9** is a block diagram of an RF Power Harvesting block in communication with a Power Storage block and used to supply power to Core Device Components.

[0031] FIG. **10** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0032] FIG. **11** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0033] FIG. **12** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block.

[0034] FIG. **13** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0035] FIG. **14** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

[0036] FIG. **15** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

[0037] FIG. **16** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0038] FIG. **17** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block.

[0039] FIG. **18** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger.

[0040] FIG. **19** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and the RF Power Harvesting block in communication with a Power Storage block.

[0041] FIG. **20** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and Core Device Components.

[0042] FIG. **21** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and Core Device Components and the RF Power Harvesting block in communication with a Power Storage block.

[0043] FIG. **22** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0044] FIG. **23** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0045] FIG. **24** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0046] FIG. **25** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0047] FIG. **26** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0048] FIG. **27** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0049] FIG. **28** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the RF Power Harvesting block supplying power to Core Device Components.

[0050] FIG. **29** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0051] FIG. **30** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

[0052] FIG. **31** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0053] FIG. **32** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power Core Device Components.

[0054] FIG. **33** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or

Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0055] FIG. **34** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0056] FIG. **35** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0057] FIG. **36** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0058] FIG. **37** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0059] FIG. **38** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0060] FIG. **39** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0061] FIG. **40** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to the Power Storage Charger.

[0062] FIG. **41** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0063] FIG. **42** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0064] FIG. **43** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

[0065] FIG. **44** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the RF Power Harvesting block supplying power to Core Device Components and used to supply power to a Power Storage Charger.

[0066] FIG. **45** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0067] FIG. **46** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0068] FIG. **47** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

[0069] FIG. **48** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power Core Device Components and used to supply power to a Power Storage Charger.

[0070] FIG. **49** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0071] FIG. **50** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

[0072] FIG. **51** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0073] FIG. **52** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0074] FIG. **53** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0075] FIG. **54** is a block diagram of an RF Power Harvesting block using Antenna A to directly supply power to Core Device Components.

[0076] FIG. **55** is a block diagram of an RF Power Harvesting block using Antenna A to supply power to a Power Regulation, Storage and/or Storage Charging block.

[0077] FIG. **56** is a block diagram of an RF Power Harvesting block using Antenna A to supply power to a Power Regulation, Storage and/or Storage Charging block and used to supply power to Core Device Components.

[0078] FIG. **57** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

[0079] FIG. **58** is a block diagram of an RF Power Harvesting block used to supply power to a Power Regulation, Storage and/or Storage Charging block.

[0080] FIG. **59** is a block diagram of an RF Power Harvesting block used to supply power to a Power Regulation, Storage and/or Storage Charging block and used to supply power to Core Device Components.

[0081] FIG. **60** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

[0082] FIG. **61** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0083] FIG. **62** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and used to directly supply power to Core Device Components.

[0084] FIG. **63** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0085] FIG. **64** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

[0086] FIG. **65** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

[0087] FIG. **66** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0088] FIG. **67** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

[0089] FIG. **68** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0090] FIG. **69** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0091] FIG. **70** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0092] FIG. **71** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0093] FIG. **72** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0094] FIG. **73** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0095] FIG. **74** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

[0096] FIG. **75** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

[0097] FIG. **76** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0098] FIG. **77** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

[0099] FIG. **78** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0100] FIG. **79** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0101] FIG. **80** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources.

[0102] FIG. **81** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

[0103] FIG. **82** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

[0104] FIG. **83** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0105] FIG. **84** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

[0106] FIG. **85** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0107] FIG. **86** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0108] FIG. **87** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0109] FIG. **88** is a block diagram of an entire power system for the device.

[0110] FIG. **89** is a block diagram of a power harvesting block used to supply power to a core device having a sensor.

[0111] FIG. **90** is a block diagram of a power harvesting block used to supply power to a core device having a computer peripheral.

DETAILED DESCRIPTION

[0112] A complete understanding of the invention will be obtained from the following description when taken in connection with the accompanying drawing figures wherein like reference characters identify like parts throughout.

[0113] There is shown an apparatus **10** for an application. The apparatus **10** comprises a core device **22** preferably having an integrated circuit **36** for the application. The apparatus **10** comprises a power harvester **20** connected to the core device **22** to power the core device **22**.

[0114] The apparatus **10** preferably includes an alternative power source **24** connected to the core device **22** to power the core device **22** in conjunction with the power harvester **20**. Preferably, the apparatus **10** includes a power regulator **26** and/or power storage circuit **28** connected to the power harvester **20**. The apparatus **10** preferably includes a power storage charger **30** connected to the power harvester **20**. Preferably, the apparatus **10** includes a power storage connected to the power harvester **20**.

[0115] Preferably, the core device **22** includes a memory **38** connected to the integrated circuit **36** and to the power harvester **20** to power the memory **36**.

[0116] The core device **22** can include a sensor **32**, as shown in FIG. **89**. The sensor **32** can include a proximity sensor, an intrusion sensor, an environmental sensor, a chemical sensor, a biological sensor, a sensor in contact with an automobile, an occupancy sensor, a motion sensor, a position sensor, a metal detector, or a sensor **32** in contact with an aircraft. The sensor **32** can include an alarm connected to the power harvester **20** to power the alarm, a display connected to the power harvester **20** to power the display, a sensor **32** disposed in a building, an industrial automation sensor, a sensor **32** in contact with an elevator, a temperature sensor, a fire sensor, or an accelerometer.

[0117] The sensor **32** can include a level sensor, a gas level sensor, a fluid level sensor, a light level sensor. The sensor **32** can include a flow sensor, a gas flow sensor, a fluid flow sensor, a light flow sensor, or a plasma flow sensor.

[0118] The sensor **32** can include a pressure sensor, a gas pressure sensor, a fluid pressure sensor, a light sensor, an infrared light sensor, an ultraviolet light sensor, an x-ray sensor, a cosmic ray sensor, a visible light sensor, a gamma ray sensor, a stress sensor, a strain sensor, a depth sensor, or an electrical characteristic sensor.

[0119] The sensor **32** includes a voltage sensor, a current sensor, a viscosity sensor, an acoustical sensor, a sound sensor, a listening sensor, a thickness sensor, a density sensor, a surface quality sensor, a volume sensor, a physical sensor, a mass sensor, a weight sensor, a conductivity sensor, a distance sensor, an orientation sensor, or a vibration sensor.

[0120] The sensor **32** can include a radioactivity sensor, a field strength sensor, an electric field sensor or a magnetic field sensor, a smoke detector, a carbon monoxide detector, a radon detector, an air quality sensor, a humidity sensor, a glass breakage sensor, or a break beam detector. The sensor can include a thermal energy sensor, an electromagnetic sensor, a mechanical sensor, an optical sensor, a radiation sensor, a sensor in contact with a vehicle, or a sensor **32** in contact with a water craft.

[0121] The present invention pertains to an apparatus **10** for an application. The apparatus **10** comprises a core device **22** having an integrated circuit **36** for the application. The apparatus **10** comprises means for receiving energy wirelessly and providing power from the energy to the core device **22** to power the integrated circuit **36** of the core device **22**. The receiving means is connected to the core device **22**. Preferably, the core device **22** includes means for sensing.

[0122] The present invention pertains to a method for an application. The method comprises the steps of converting RF energy into usable energy. There is the step of preferably powering an integrated circuit **36** of the core device **22** with the usable energy.

[0123] Preferably, there is the step of regulating the usable energy provided to the core device **22**. There is preferably the step of storing the usable energy. Preferably, there is the step of providing power to the core device **22** from an alternative power source **24** in conjunction with the usable energy.

[0124] Alternatively, the core device **22** can include a computer peripheral **34**, as shown in FIG. **90**. The computer peripheral **34** can include a handheld game, a gaming system, a game controller, a controller, a keyboard, a mouse, a computer terminal, computer storage, or computer equipment.

[0125] The present invention can be implemented in numerous ways. A number of these ways are depicted in FIGS. **1-90**. These figures contain multiple blocks that are configured in multiple ways.

[0126] In the figures, an arrow represents the flow of power unless otherwise stated. Single-headed (one-way) arrows represent that the power is flowing from one block to another. The single-headed arrow may represent multiple wires that provide power from one block to multiple parts in the other block. Two-headed (two-way) arrows represent a single wire that can have power flow in either direction or multiple wires each having power flow in a single direction.

[0127] As an example, a two-headed arrow between an RF power harvesting **20** block and a power regulation **26** and/or power storage circuit **28** block can represent a single wire that allows harvested power to flow into a storage device such as a capacitor. The same block diagram can also represent two wires between the two blocks with the first wire allowing harvested power to flow into a voltage regulator **26**. The second wire can allow the regulated voltage to feedback to the RF power harvesting **20** block to provide power to internal components such as transistors to increase the performance of the RF power harvesting **20** block.

[0128] Each block is described in detail below. Each block represents the functionality described below associated with it. For instance, the RF power harvesting **20** block describes a power harvester **20**, and the power regulation **26** and/or power storage circuit **28** block describes a power regulator **26** and a power storage circuit **28**.

Device's Power System Block

[0129] The device's power system block includes all components/circuitry required to power the device. This block may include an RF power harvesting **20** block, a power regulation **26** and/or power storage circuit **28** block, a power storage block, and/or a power storage charger **30** block (all discussed below).

RF Power Harvesting **20** Block

[0130] The RF power harvesting **20** block is connected to an antenna. The antenna may or may not be used as the communications antenna for the core device **22** components. The RF power harvesting **20** block is used to convert the energy captured by the antenna into usable power, such as direct current (DC) voltage.

[0131] This block may include antenna matching, rectifying circuitry, voltage transforming circuitry, and/or other performance optimizing circuitry. The rectifying circuitry may include a diode(s), a transistor(s), or some other rectifying device or combination. Examples of the rectifying circuitry include, but are not limited to, half-wave, full-wave, and voltage doubling circuits.

[0132] The output of the RF power harvesting **20** block is a DC voltage or current. The RF power harvesting **20** block may accept feedback (or input) from other circuitry or blocks, which may be used to control the harvesting circuitry to improve the performance or vary the output. This feedback may include, but is not limited to, a DC voltage or a clock from the core device **22** components.

[0133] U.S. Pat. No. 6,615,074 (FIGS. **8**, **9**, **12***a*, **12***b*, **13**, **14**), incorporated by reference, herein, shows numerous examples of RF power harvesting circuits that can be used to implement the block and function described.

Power Regulation **26** and/or Power Storage Circuit **28** Block

[0134] It may be necessary to regulate the converted power (i.e., hold the power at a constant level) for specific devices. The devices that would need this block require a fairly constant voltage or current. Deviations from the required values may cause the device to not perform within its specifications.

[0135] The regulation can be implemented in many different ways. The block can be as simple as using a Zener diode, or as complicated as using an integrated circuit such as a linear voltage regulator **26** or switching regulator **26** to hold the voltage at a constant level.

[0136] Certain devices have a more tolerable power requirement. For these devices, the regulation stage may be excluded.

[0137] This block may also include, with or without the regulation, a storage device such as a capacitor, a battery, or some other device able to store charge. The output from the power regulation **26** and/or power storage circuit **28** block may be used as feedback to other blocks within the device's power system block or to the alternative power sources **24** block (described below), if they require a regulated supply voltage or stored power.

[0138] U.S. Pat. No. 6,894,467 (FIGS. **1**, **3**), Linear Voltage Regulator, incorporated by reference, herein, is an example of a practical application of implementing the regulation described in the block. U.S. Pat. No. 6,297,618 (FIGS. **1-4**), Power storage device and method of measuring voltage of storage battery, incorporated by reference, herein, is an example of a practical application of implementing the storage described in the block.

Power Storage Charger **30** Block

[0139] The power storage charger **30** block may be needed if a power storage component of the device requires a special charging mechanism, such as pulse charging or trickle charging. This block controls how the captured and converted power is supplied to the power storage component.

[0140] U.S. Pat. No. 6,836,095 (FIGS. **1-3**), Battery Charging Method and Apparatus, incorporated by reference herein, is an example of a practical application of implementing the special charging mechanism described in the block.

Power Storage Block

[0141] If a device has intermittent power requirements, it may be necessary to store the captured power for use at a later time. The power can be stored in the power storage block, which could include a battery, a capacitor, and/or another type of power storage component. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other power storage elements.

[0142] The output from the power storage block may be used as feedback to other blocks within the device's power system or to the alternative power sources **24** block, if they require a dedicated and predictable supply voltage.

[0143] U.S. Pat. No. 6,297,618 (FIGS. **1-4**), Power Storage Device and Method of Measuring Voltage of Storage Battery, incorporated by reference herein, is an example of a practical application of implementing the storage described in the block. U.S. Pat. No. 6,835,501, Alkaline Rechargeable Battery, incorporated by reference herein, is also an example of a practical application of implementing the storage described in the block.

Core Device **22** Components Block

[0144] The core device **22** components block is the portion of the device that is receiving power from the device's power system. This block may be, but is not limited to, the devices listed in the subsequent pages of this document. It may be advantageous for the core device **22** components block to communicate with any of the blocks that are supplying power to it. This communication can include, but is not limited to, a feedback control signal such as a clock or an ON/OFF command. As an example, the device may want to turn off the alternative power sources **24** block if it is receiving sufficient power from the RF power harvesting **20** block.

Alternative Power Sources **24** Block

[0145] RF energy harvesting also has the ability to be augmented by other types of power harvesting, storage components, or dedicated sources (e.g. power line). The alternative power sources **24** block shows how this type of system could be implemented. The augmenting power harvesting technologies include, but are not limited to, solar, light (visible and non-visible), piezoelectric, vibration, acoustic, thermal, microgenerators, wind, and other environmental elements. This block can work independently or have communication with other blocks.

[0146] U.S. Pat. No. 6,784,358, Solar Cell Structure Utilizing an Amorphous Silicon Discrete By-Pass Diode, incorporated by reference herein, is an example of a practical application of implementing an alternative power source **24** described by the block. U.S. Pat. No. 6,858,970, Multi-Frequency Piezoelectric Energy Harvester, incorporated by reference herein, is also an example of a practical application of implementing an alternative power source **24** described by the block.

Power Regulation **26**, Storage **28** and/or Storage Charging **30** Block

[0147] The power regulation **26**, storage **28** and/or storage charging **30** block contains all the combinations of the power regulation **26** and/or power storage circuit **28** block, power storage charger **30** block, and power storage block. This block is used in the later figures to reduce the number of figures needed to show how the blocks can interconnect.

[0148] The disclosed invention is an apparatus and method for an application for retrieving radio frequency (RF) energy by an antenna, converting that energy into direct current (DC) power, regulating that energy using an optimized circuit, storing that energy in an optimized component, and/or supplying the power for a specific device.

Retrieval of RF Energy

[0149] The RF energy is retrieved from the environment by the use of the antenna. The antenna can be shared or standalone with respect to an antenna used for the device's wireless communication. FIGS. **54-56** show a device that has an antenna A for use by the RF power harvester **20** and an antenna B used for wireless one-way or two-way communication. FIGS. **57-59** show a device where the antenna is shared by both a device's communication module (within the core device **22** components) and the RF power harvester **20**. In terms of form factor, the antenna used by the apparatus **10** can be a separate component or integrated directly into the form factor of the device.

[0150] The antenna is able to capture two types of available RF energy. The first type of energy exists as ambient RF energy. This type of RF surrounds us in our day-to-day lives and is usually generated to carry one way or two-way combinations of voice, video and data communications. The sources that the antenna can harvest from include, but are not limited to, medium-frequency AM radio broadcast, very-high-frequency (VHF) FM radio broadcast and television broadcast, ultra-high-frequency (UHF) broadcast, cellular base stations, wireless data access points, super-high-frequency (SHF) frequencies, and the industrial, scientific, and medical (ISM) bands. These sources cover transmission frequencies from 300 kHz to 30 GHz.

[0151] The second type of energy available is directed RF energy. This type of RF energy is directed from a transmitter specifically designed to deliver RF energy for harvesting by the antenna. The transmitter can be configured as a standalone device or integrated into an existing device.

Conversion of the Energy into DC

[0152] The RF energy captured by the antenna must be converted into a useful form of energy for the specific device. This conversion is shown in block form in all FIGS. (**1-90**) as the RF power harvesting **20** block. The most common form of useable energy is DC energy. To perform this conversion, the block includes circuitry to rectify the captured alternating current (AC) energy to create DC energy. The rectification in this block can be done with a diode(s), a transistor(s), or some other rectifying device or combination.

Regulation of the Energy

[0153] It may be necessary to regulate the converted power (hold the power at a constant level) for specific devices. FIGS. **2-7**, **10-17**, and **22-53** show how this regulation can be implemented using a power regulation **26** and/or power storage circuit **28** block. The devices that would need this block require a fairly constant voltage or current. Deviations from the required values may cause the device to not perform within its specifications. The regulation can be implemented in many different ways. The block can be as simple as using a Zener diode, or as complicated as using an integrated circuit such as a linear voltage regulator **26** or a switching regulator **26** to hold the voltage at a constant level. Certain devices have a more tolerable power requirement. For these devices, the regulation stage may be excluded.

Storage of the Energy

[0154] If a device has intermittent power requirements, such as the devices exampled by FIGS. **2-53**, it may be necessary to store the captured power for use at a later time. The power can be stored in the power storage block or the power regulation **26** and/or power storage circuit **28** block. Storage devices can include, but are not limited to, a battery, a capacitor, or another type of power storage component. In certain applications, it may be necessary to include additional circuitry that controls how the power is transferred to the storage device. The power storage charger **30** block is shown in FIGS. **18-53**. This may be needed if the power storage component requires a special charging mechanism such as pulse charging or trickle charging. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other storage elements. There are devices that will not require storage. These devices can run directly off of the converted power. These devices also may or may not require regulation of the captured power.

Supplying the Power

[0155] The captured DC power, which may or may not be regulated and/or stored, is supplied to the device, which is represented by the core device **22** components block in the figures. This may be a single connection or it may supply multiple parts of the device with power.

[0156] RF energy harvesting also has the ability to be augmented by other types of power harvesting or storage components. Other power harvesting technologies include, but are not limited to, solar, light (visible and non-visible), piezoelectric, vibration, acoustic, thermal, microgenerators, wind, and other environmental elements. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other storage elements. FIGS. **60-88** show how the alternative power sources **24** block can be connected to an RF energy harvesting system. These figure show how the RF energy harvesting **20** block and the alternative power sources **24** block can work independently or have communication with each other. The antenna configurations shown in FIGS. **54-59** are still applicable with the addition of an alternative power sources **24** block. These antenna configurations can be applied to FIGS. **60-88**.

[0157] RF energy harvesting also has the ability to provide a backup to devices on-grid (tethered) or with reliable power sources, which can be used in case the primary power source is lost. As an example, it may be mandated by regulations that a sensor has auxiliary power in case the primary supply is lost. It could be possible to use a rechargeable battery that obtains its charge from the primary supply when in operation. However, if the primary supply is lost for a time greater than the life of the rechargeable battery, the specification of uninterrupted power is not met. RF energy could be used to supply power to the described device while the primary supply is not available. The primary supply could include, but is not limited to, an on-grid connection, a generator, a battery, or other reliable power supply.

[0158] RF energy harvesting with or without alternative source augmentation is applicable to provide electric power directly or indirectly to a range of electronic components contained in any specific electrical or electronic device and includes, but is not limited to:

[0159] Passive electronic components, active electronic components;

[0160] Resistors, fixed resistors, variable resistors, thermistors, thyristor, thermocouple;

[0161] Capacitors, Electrolytic Capacitors, Tantalum Capacitors, Ceramic Capacitors, Multilayer Ceramic Capacitors, Polystyrene Film Capacitors, Electric Double Layer Capacitors (Super Capacitors), Polyester Film Capacitors, Polypropylene Capacitors, Mica Capacitors, Metallized Polyester Film Capacitors, Variable Capacitors;

[0162] Diodes, Voltage regulation diodes, light-emitting diodes, organic light-emitting diodes, Variable capacitance diodes, Rectification diodes, Switching diodes, Regulation Diodes, Diode bridges, Schottky barrier diodes, tunnel diodes, PIN diodes, Zener diodes, Avalanche diodes, TVSs;

[0163] Integrated circuits, microcontroller unit (MCU), microprocessor unit (MPU), logic circuits, memory, printed circuits, circuit boards, printed wiring boards;

[0164] Transistors, MOSFETs, FETs, BJTs, JFETs, IGBTs, Relays, Antennas, semiconductors, conductors, inductors, relays, diacs, triacs, SCRs, MOVs;

[0165] Fuses, circuit breakers;

[0166] Batteries, Non-rechargeable batteries, rechargeable batteries, coin cell batteries, button cell batteries, alkaline batteries, lithium batteries, lithium ion batteries, lithium polymer batteries, NIMH batteries, NICAD batteries, Lead acid batteries, Zinc air batteries, Manganese Lithium batteries, Niobium Titanium Lithium batteries, Vanadium Pentoxide Lithium batteries, Carbon

Zinc batteries, Zinc Chloride batteries, Lithium Thionyl Chloride batteries, Manganese Dioxide batteries, Lithium Poly-Carbonmonofluoride batteries, Lithium Manganese Dioxide batteries, Lithium Chloride batteries, Lead Acid Calcium batteries, Lead Acid Tin batteries, Oxy Nickel batteries, Silver Oxide batteries, Magnesium batteries;

[0167] Inductors, Coils, High Frequency Coils, Toroidal Coils, Transformers, switches, chokes;

[0168] Motors, DC motors, stepper motor, AC motors, Fans;

[0169] Crystals, Oscillators, Clocks, Timers; and

[0170] Displays, LCDs, LED displays.

[0171] RF energy harvesting with or without alternative source augmentation is applicable across a range of markets and specific devices and includes but is not limited to:

[0172] Consumer electronics:

[0173] Electronic equipment, wired devices, battery powered devices, wireless communication devices, cell phones, telephones, phones, cordless phones, portable phones, Bluetooth devices, Bluetooth headsets, handsfree headsets, headsets, headphones, Wireless headsets, radios, AM/FM radios, shortwave radios, weather radios, Two-way radios, portable radio, lights, lanterns, portable lights, flashlights, nightlights, spotlights, search lights, calculators, graphing calculators, desk calculators, clocks, alarm clocks, wall clocks, desk clocks, travel clocks, watches, wristwatches, pocket watches, stop watches, timers, voice recorders, Dictaphones, laser pointers, power tools, cordless power tools, electronic razors, electric razors, handheld games, gaming systems, game controllers, wireless game controllers, remote controls, battery chargers, computers, portable computers, keyless entry, toys, toy guns, toy laser guns, games, microphones, musical instruments, musical effects processors, musical instrument tuners, metronomes, electronic chord charts, door openers, garage door openers, PDA, Cameras, Video recorders, Multimeter, electronic test equipment, hand-held electronics, portable electronics, wireless pens, sound generators, noise generators, language translators, electric toothbrushes, portable televisions, pagers, transceivers, toy vehicles, remote control vehicles, toy planes, remote control planes, pet containment systems, invisible fence pet sensors, memory backup, base station battery backups, appliance battery backups, uninterrupted power supplies, GPS devices, memory retention power supplies, metal detectors, stud finders, metal stud finders, stun guns, tazers, wearable devices, baby monitors, intercoms, doorbells, wireless doorbells, electronic office supplies, electronic staplers, radar jammers, radar detectors, digital scales, microfilm cassettes, video head testers, compasses, noise canceling headphones, air samplers, depth finders, barometers, weather measurement instruments, data transfer devices, automatic distress signaling unit, Wireless audio speakers, Satellite radios, Police scanners, Car navigation systems (GPS devices), Decorative lights, Christmas lights, garden lights, lawn lights, ornamental lights, porch lights;

[0174] Multi-media players: MP3, DVD, analog music players, CD players, tape players, digital music players, digital video players, minidisc; and

[0175] Computer: keyboards, mice, peripherals, computer equipment, electronic computers, computer storage, computer terminals;

[0176] Building/Industrial Automation:

[0177] Sensors: Position, elevator, temperature, fire, accelerometers, level, gas level, fluid level, light level, flow, gas flow, fluid flow, light flow, plasma flow, pressure, gas pressure, fluid pressure, motion, light, infrared light, ultraviolet light, X-rays, cosmic rays, visible light, gamma rays, chemical, stress, strain, depth, electrical characteristics, voltage, current, viscosity, acoustical, sound, listening, thickness, density, surface quality, volume, physical, mass, weight, force, conductivity, distance, orientation, vibration, radioactivity, field strength, electric field strength, magnetic field strength, occupancy, smoke detector, carbon monoxide detector, radon detector, air quality, humidity, glass breakage, break beam detector;

[0178] Controls: Position, elevator, temperature, fire, accelerometers, level, gas level, fluid level, light level, flow, gas flow, fluid flow, light flow, plasma flow, pressure, gas pressure, fluid pressure, motion, light, infrared light, ultraviolet light, X-rays, cosmic rays, visible light, gamma rays, chemical, stress, strain, depth, electrical characteristics, voltage, current, viscosity, acoustical, sound, listening, thickness, density, surface quality, volume, physical, mass, weight, force, conductivity, distance, orientation, vibration, radioactivity, field strength, electric field strength, magnetic field strength, occupancy, smoke detector, carbon monoxide detector, radon detector, air quality, humidity; and

[0179] Devices: Thermostats, light switches, door locks, smart-card door locks, lighting, emergency lighting, motion lighting, safety lighting, highway lighting, construction lighting, sign lighting, roadway sign lighting, construction sign lighting, automatic flushing units, automatic soap dispenser, automatic paper towel dispenser, automatic faucets, automatic door sensors, identification reader, fingerprint reader, credit card readers, card readers, valve actuators, gauges, analog gauges, digital gauges, fire extinguishers, wireless switches, remotely operated inspection equipment, gas/oil pipeline monitoring systems, robotic pipeline inspection gauges, “auto-reclosers” for electric power lines, sonar buoys, telemetry systems, electronic record tracking systems, robbery tracking devices, interrogators, programmers, emergency exit alarms, alarms, flood alarms, gas alarms, electronic entry systems, security keypads, silo transducers, data recorders, signal tracers, antistatic strap testers, radiosonde weather balloons, utilities load controllers, profilometers, noise cancellation equipment, infrared beacons;

[0180] Military/Government:

[0181] Tracking tags: Weapons, vehicle, soldier, gear/assets, staff, general population, security access badges;

[0182] Sensors: Proximity, intrusion, environmental, chemical/biological; and

[0183] Equipment: Battery charger, surveillance, card readers, identification reader, fingerprint reader, retinal scanners, satellites, rockets, space vehicles, search and rescue transponders (SARTs), emergency position-indicating rescue beacons (EPIRBs), emergency locator transmitters (ELTs), military radios, electronic toll collection systems, postal tracking systems, communica-

tions, thermal imaging, night vision, training targets, field medical equipment, house arrest monitors, laser tags, electronic parking meters, multiple integrated laser engagement system, munitions and mines, ship sensors;

[0184] Utility:

- **[0185]** Gas consumption meters, water consumption meters, and electric consumption meters;

[0186] Logistics & Supply Chain Management:

- **[0187]** Radio-frequency identification devices (RFID), RFID readers;
- **[0188]** Tracking: Asset tags, cargo container location beacons, transponders, transceivers; and
- **[0189]** Devices: Smart price tags, smart shelving, handheld barcode scanner, barcode scanners, credit card readers, card readers, retail signage, hotel door locks;

[0190] Homeland Security:

- **[0191]** Sensors: Occupancy, proximity, environmental, chemical/biological, motion, position; and
- **[0192]** Metal detector wand;

[0193] Medical:

- **[0194]** Implantable: cochlear implants, neural stimulators, pace makers, medication administration, defibrillator;
- **[0195]** Body function monitors: pressure, temperature, respiration, blood oxygenation, insulin, hearing aid, pulse, EKG, heart, Holter;
- **[0196]** Tracking tags: Patient, baby identification, assets, supplies, staff, medication, instruments;
- **[0197]** Devices: Home healthcare equipment, ambulatory infusion pumps, blood analyzers, biofeedback systems, bone growth stimulators, thermometers, digital thermometers, stimulators, galvanic stimulators, muscle stimulators, pediatric scales; and

[0198] Agriculture—livestock tracking and asset tracking:

- **[0199]** Tracking: livestock, asset, wildlife tracking devices; and
- **[0200]** Equipment: cattle prods;

[0201] Automotive:

- **[0202]** Automotive antennas, Automotive Audio Systems, Automotive Lighting, Automotive Video Systems, Computers, Processors, Controls, Switches, Electric Motors, Actuators, Ignition Systems, Starter Systems, Injection Systems, Powertrain Electronics, Radar Detectors, Proximity Detectors, Safety Systems, Security Systems, Sensors, Regulators, Distributors, Vehicle Control, Wiper Systems, Washer Systems, Radio, Video Systems, Entertainment Systems, Navigation Systems, GPS systems, Power Mirror Systems, Emission control systems;

[0203] Appliances:

- **[0204]** Monitoring systems and control systems for major and small appliances including washing machines, dryers, refrigerators, freezers, coolers, air conditioners, humidifiers, dehumidifiers, air purifiers, air filters, fans, furnaces, water heaters, boilers, space heaters, sowing machines, ice makers, microwave ovens, convection ovens, ovens, toaster ovens, ranges, range hoods, cooktops, stoves, stovetops, crock pots, hot plates, dishwashers, garbage disposals, can openers, vacuum cleaners, blenders, mixers, food processors, irons, coffee makers, toasters, grills, hair dryers, electric tooth brushes, electric razors, electric drills, electric screwdrivers, chainsaws, lawnmowers, push mowers, riding mowers, trimmers, brush cutters, pruners, edgers, vending machines;

[0205] Ventilation, Heating, Air-Conditioning, and Commercial Refrigeration Equipment:

- **[0206]** Monitoring systems, control systems;

[0207] Engine, Turbine, and Power Transmission Equipment;

[0208] Monitoring systems, control systems;

[0209] Other General Purpose Machinery Manufacturing:

- **[0210]** Monitoring systems, control systems;

[0211] Telecommunications:

- **[0212]** Monitoring systems, control systems;
- **[0213]** Portable;

[0214] Aircraft:

- **[0215]** Monitoring systems, control systems, actuator systems, sensors.

[0216] It should be noted that devices within a specific category may be applicable across multiple areas even if they are not specifically listed. (e.g., temperature sensors apply to Industrial and Building Automation and Appliances).

[0217] To retrofit or redesign the devices listed, it is possible to implement the described systems in numerous ways. It may be advantageous to leave the device design as-is including the existing power supply. As an example, a device may use non-rechargeable batteries to operate. The device will most likely have a protection circuit to prevent damage if the batteries are installed incorrectly. The protection mechanism is commonly a diode inline with the positive terminal of the battery. In this case, the RF power harvester **20** with or without the alternative power source **24** could be inserted, with an antenna, into the device. The power generated by the RF power harvester **20** (and the alternative power source **24**, if applicable) could be connected to the device after the protection mechanism described to avoid potential charging of a non-rechargeable battery.

[0218] Another way to configure the device's power system is to replace the non-rechargeable batteries with rechargeable batteries. In this instance, the output from the RF power harvester **20** (and alternative power source **24**) could be connected to either side of the protections device. If the connection is before the protection mechanism, the RF power harvester **20** (and the alternative power source **24**, if applicable) will recharge the battery and supply power to the device. If the connection is after the protection mechanism, the RF power harvester **20** (and the alternative power source **24**, if applicable) will supply power to the device and the battery will supply any extra power needed that could not be supplied by the RF power harvester **20** (and the alternative power source **24**, if applicable). It should be noted that the protection device in this case is not needed for proper operation. Its only function would be to protect the batteries from being installed incorrectly. An antenna could be contained inside or placed on the outside of the device.

[0219] Another configuration of the device's power system is to remove the existing batteries and install the RF power harvester **20** (and the alternative power source **24**) in the enclosure provided for the batteries. An antenna could be contained inside or placed on the outside of the device.

[0220] Yet another method of configuring the device's power system would be to reduce the number of batteries and replace them with the RF power harvester **20** (and the alternative power source **24**). In this case, the output from the system would be connected to the batteries in series or par-

allel depending on the original battery configuration. An antenna could be contained inside or placed on the outside of the device.

[0221] An additional option, would be to completely redesign the product and integrate the required circuitry and storage components into the device. This method is probably the most advantageous because it can fully take advantage of the benefits offered by the RF power harvester **20** (and the alternative power source **24**). An antenna could be contained inside or placed on the outside of the device.

[0222] If the RF power harvester **20** (and the alternative power source **24**) is used as a backup to the primary power supply, a switch could be implemented into the system in order to switch the RF power harvester **20** (and the alternative power source **24**) on when the primary source is lost. In this case, an antenna could be contained inside or placed on the outside of the device.

[0223] To show the flexibility of RF energy harvesting, several products were retrofitted to include an RF power harvester **20** (energy harvesting circuitry). These products include a wireless keyboard, a wall clock, and a desk calculator.

[0224] The wireless keyboard is an example of recharging and augmenting a battery to supply power to a device. This system is shown in FIG. **13**. The output from the regulation circuitry recharges the battery and supplies power to the keyboard. The battery is also used to supply power to the keyboard. The keyboard also includes a separate antenna for receiving power and for data communications. The antenna configuration can be seen in FIG. **55**.

[0225] The wall clock is an example of a direct powering system. The wall clock was retrofitted to include an RF power harvester **20** and the internal AA battery was removed. This system is shown in FIG. **2**. The wall clock did not need regulation, but did require a capacitor for storage to supply the pulse of power to move the second hand.

[0226] The calculator is an example of using RF energy harvesting with another energy harvesting technology. The calculator had an internal 1.5V coin cell battery and a small solar panel. The internal battery was removed, however, the solar panel was left intact. This system is shown in FIG. **60**. In this system, the calculator can receive power from both the solar panel and the RF power harvester **20** to eliminate the need for a battery.

[0227] As an additional example, an RF power harvester **20** (energy harvesting circuit) similar to the ones shown in U.S. Pat. No. 6,615,074 (FIGS. **8**, **9**, **12***a*, **12***b*, **13**, **14**), was connected in series with a 0.5V solar cell. Individually, the solar cell was able to provide 0.480V to a 10 kilo-ohm resistor, which was being used to simulate the core device **22** components. This corresponds to 23 microwatts. The RF power harvesting circuit by itself was able to provide 2.093V across the 10 kilo-ohm resistor when being supplied by 1 milliwatt of RF power. This corresponds to 438 microwatts. The two circuit outputs were then combined in series by connecting the output from the RF energy harvesting circuit to the ground of the solar cell. The output of the solar cell was then connected to the resistor. The other end of the resistor was connected to the ground of the RF energy harvesting circuit. The voltage across the resistor with the circuits connected as shown in FIG. **63** was 2.445V. This corresponds to 598 microwatts.

[0228] As can be seen, the combination of the two technologies produces a result higher than the addition of the individual powers. From this, it can be determined that the two technologies can cooperate in a way that produces favorable results.

[0229] In the example given, the solar cell produces current to supply the load and helps to bias the RF rectifying diodes, which allows the RF energy harvesting circuit to operate a higher efficiency. The solar cell also changes the impedance seen by the RF energy harvesting circuit, which produces a beneficial result. To be more specific, when examining the power output of the individual circuit (solar and RF power harvesting), the sum of the power captured by the individual circuits was 23 μW+438 μW=461 μW. However, when the two circuits are combined and are allowed to work in conjunction with one another, the output power becomes 598 μW. This result shows that combining the two power-harvesting technologies produces a 30 percent increase in the output power for this example. This same technique can be applied to multiple energy harvesting technologies to produce even greater output power. The equations for this example are shown below.

[0230] Individual Circuits

$$P_I=P_1+P_2+\ldots+P_N$$

[0231] Combined Circuits

$$P_C>P_I=P_1+P_2+\ldots+P_N$$

[0232] where P_I is the sum of the individual output powers

[0233] P_C is the output of the combined circuit

[0234] P_1 is the output power from the first power harvesting technology

[0235] P_2 is the output power from the second power harvesting technology

[0236] P_N is the output power from the N^{th} power harvesting technology

[0237] N is the number of power harvesting technologies or circuits.

[0238] As can be seen by the previous examples, RF energy harvesting can be used alone or in conjunction with alternative power sources to power a wide range of devices. The addition of RF energy harvesting technology to the device allows for increased battery life, increased functionality, or the removal of the primary battery.

[0239] Although the invention has been described in detail in the foregoing embodiments for the purpose of illustration, it is to be understood that such detail is solely for that purpose and that variations can be made therein by those skilled in the art without departing from the spirit and scope of the invention except as it may be described by the following claims.

What is claimed is:

1. An apparatus, comprising:

a first antenna configured to wirelessly receive energy from a source, the source being remote from the apparatus;

a second antenna different from and electrically isolated from the first antenna, the second antenna configured to wirelessly communicate with the source;

a power harvester electrically coupled to the first antenna and configured to convert the energy received at the first antenna to a DC voltage;

a radio-frequency identification (RFID) device electrically coupled to the power harvester and to the second antenna, the RFID device configured to wirelessly communicate with the source via the second antenna;

an integrated circuit configured to operate based on the DC voltage; and a sensor electrically coupled to the integrated circuit, the sensor configured to operate based on the DC voltage and to detect at least one of an environmental condition associated with a location of the apparatus or a physical condition associated with an item coupled to the apparatus.

2. The apparatus of claim **1**, further comprising:

a regulator coupled to the power harvester and configured to regulate the DC voltage to produce a regulated DC voltage, the sensor configured to operate based on the regulated DC voltage received from the regulator.

3. The apparatus of claim **1**, wherein the integrated circuit is configured to perform an application associated with the at least one of the environmental condition or the physical condition detected by the sensor.

4. The apparatus of claim **1**, further comprising:

an electronic display configured to operate based on the DC voltage.

5. The apparatus of claim **1**, wherein the sensor includes at least one of a temperature sensor, a motion sensor, a vibration sensor, a humidity sensor, an orientation sensor, or a position sensor.

6. The apparatus of claim **1**, wherein the RFID device does not operate based on a battery, and the sensor does not operate based on a battery.

7. The apparatus of claim **1**, wherein the apparatus is a RFID tag.

8. An apparatus, comprising:

an antenna configured to wirelessly receive energy from a source, the source being remote from the apparatus;

a power harvester electrically coupled to the antenna and configured to convert the energy received at the antenna to a DC voltage;

a radio-frequency identification (RFID) device electrically coupled to the antenna and configured to wirelessly communicate with the source via the antenna, the RFID device operatively coupled to the power harvester and including an integrated circuit configured to operate based on the DC voltage; and

a sensor electrically coupled to the RFID device and the power harvester, the sensor configured to operate based on the DC voltage and to detect at least one of an environmental condition associated with a location of the apparatus or a physical condition associated with an item coupled to the apparatus,

the integrated circuit configured to perform an application associated with the at least one of the environmental condition or the physical condition detected by the sensor.

9. The apparatus of claim **8**, wherein the power harvester is configured to wirelessly communicate with the source via the antenna.

10. The apparatus of claim **8**, wherein the sensor is electrically coupled to the antenna and is configured to wirelessly communicate with the source via the antenna.

11. The apparatus of claim **8**, further comprising:

a regulator coupled to the power harvester and configured to regulate the DC voltage to produce a regulated DC voltage, the integrated circuit configured to operate based on the regulated DC voltage received from the regulator.

12. The apparatus of claim **8**, further comprising:

an electronic display configured to operate based on the DC voltage.

13. The apparatus of claim **8**, wherein the sensor includes at least one of a temperature sensor, a motion sensor, a vibration sensor, a humidity sensor, or a position sensor.

14. The apparatus of claim **8**, wherein the RFID device does not operate based on a battery, and the sensor does not operate based on a battery.

15. An apparatus, comprising:

an antenna configured to wirelessly receive energy from a source, the source being remote from the apparatus;

a power harvester electrically coupled to the antenna and configured to convert the energy received at the antenna to a DC voltage to produce a power associated with the DC voltage;

a regulator coupled to the power harvester and configured to regulate the power associated with the DC voltage to produce a regulated power;

a radio-frequency identification (RFID) device electrically coupled to the antenna and configured to wirelessly communicate with the source via the antenna, the RFID device electrically coupled to the power harvester and the regulator, the RFID device including an integrated circuit configured to operate based on at least one of the power received from the power harvester or the regulated power received from the regulator; and

an electronic display electrically coupled to the RFID device and configured to operate, at least in part, based on the regulated power received from the regulator.

16. The apparatus of claim **15**, wherein the electronic display is electrically coupled to the regulator, the electronic display is configured to operate based on the regulated power received from the regulator and the power received from the power harvester.

17. The apparatus of claim **15**, wherein the electronic display is one of a liquid-crystal display (LCD) or a light-emitting diode (LED) display.

18. The apparatus of claim **15**, further comprising:

a sensor configured to operate, at least in part, based on the regulated power received from the regulator, the sensor including at least one of a temperature sensor, a motion sensor, a vibration sensor, a humidity sensor, or a position sensor.

19. The apparatus of claim **15**, wherein the power harvester is configured to wirelessly communicate with the source via the antenna.

20. The apparatus of claim **15**, wherein the RFID device does not operate based on a battery, and the electronic display does not operate based on a battery.

* * * * *

US 20150236551A1

(19) **United States**

(12) **Patent Application Publication**

Shearer et al.

(10) **Pub. No.: US 2015/0236551 A1**

(43) **Pub. Date: Aug. 20, 2015**

(54) **POWERING DEVICES USING RF ENERGY HARVESTING**

(71) Applicant: **POWERCAST CORPORATION**, Ligonier, PA (US)

(72) Inventors: **John G. Shearer**, Ligonier, PA (US); **Charles E. Greene**, Cabot, PA (US); **Daniel W. Harrist**, Carnegie, PA (US)

(73) Assignee: **POWERCAST CORPORATION**, Ligonier, PA (US)

(21) Appl. No.: **14/697,053**

(22) Filed: **Apr. 27, 2015**

Related U.S. Application Data

(60) Continuation of application No. 14/143,334, filed on Dec. 30, 2013, now Pat. No. 9,021,277, which is a continuation of application No. 12/499,618, filed on Jul. 8, 2009, now Pat. No. 8,621,245, which is a division of application No. 11/447,412, filed on Jun. 6, 2006, now abandoned.

(60) Provisional application No. 60/688,587, filed on Jun. 8, 2005.

Publication Classification

(51) **Int. Cl.**
H02J 17/00 (2006.01)

(52) **U.S. Cl.**
CPC *H02J 17/00* (2013.01)

(57) **ABSTRACT**

Disclosed is an apparatus for an application including a core device for the application. The apparatus includes a power (preferably RF energy) harvester connected to the core device to power the core device. Also disclosed is a method for an application. The method includes the steps of converting RF energy into usable energy. There is the step of powering the core device with the usable energy.

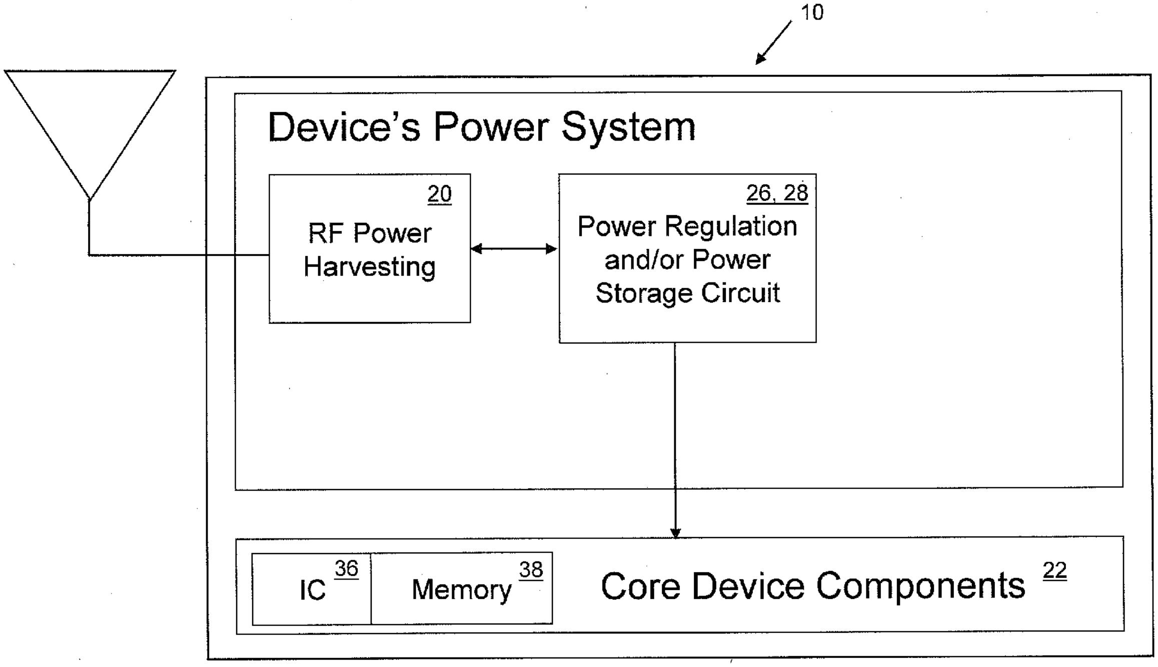
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
IC 36
Memory 38
Core Device Components 22

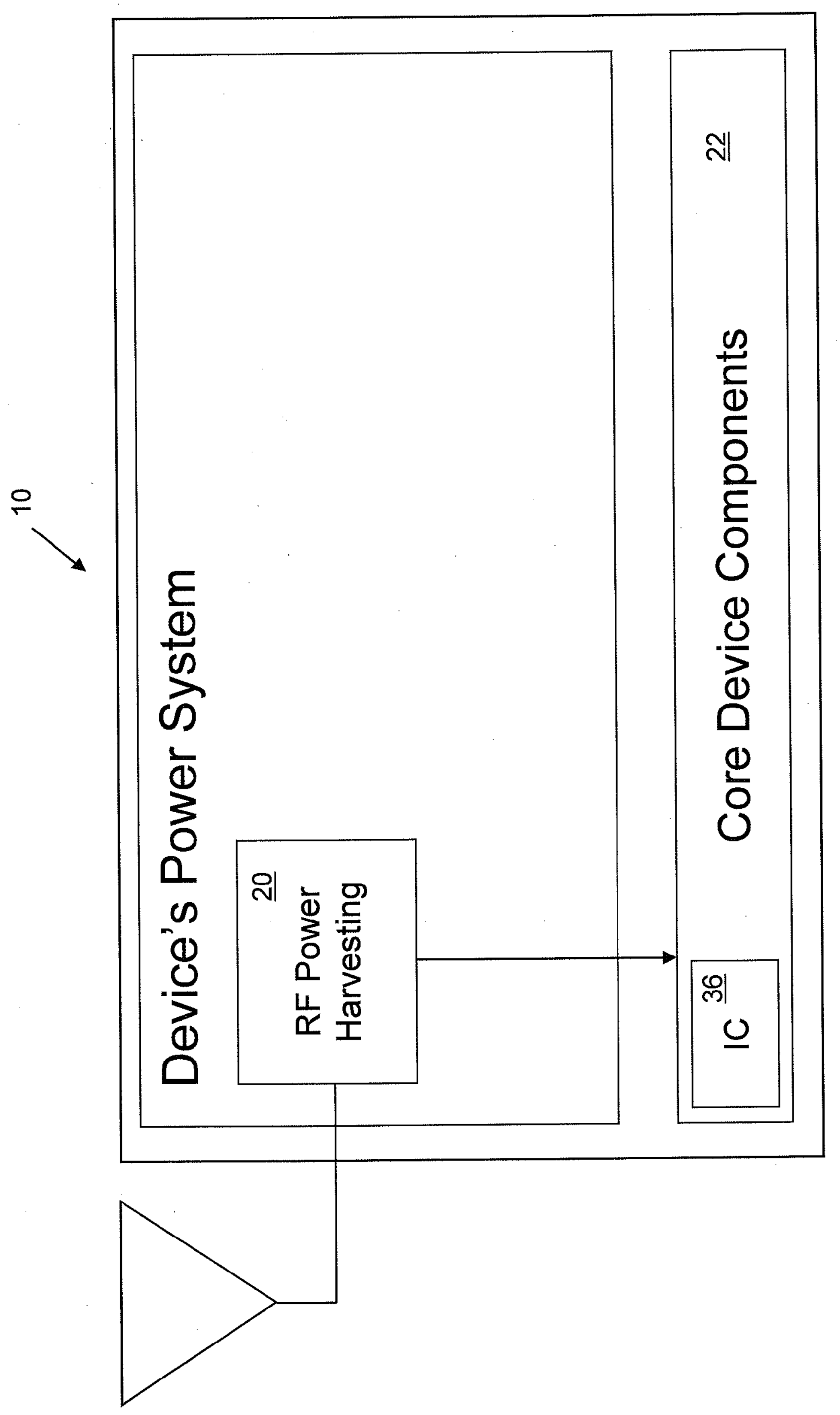
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
IC 36

Fig. 1

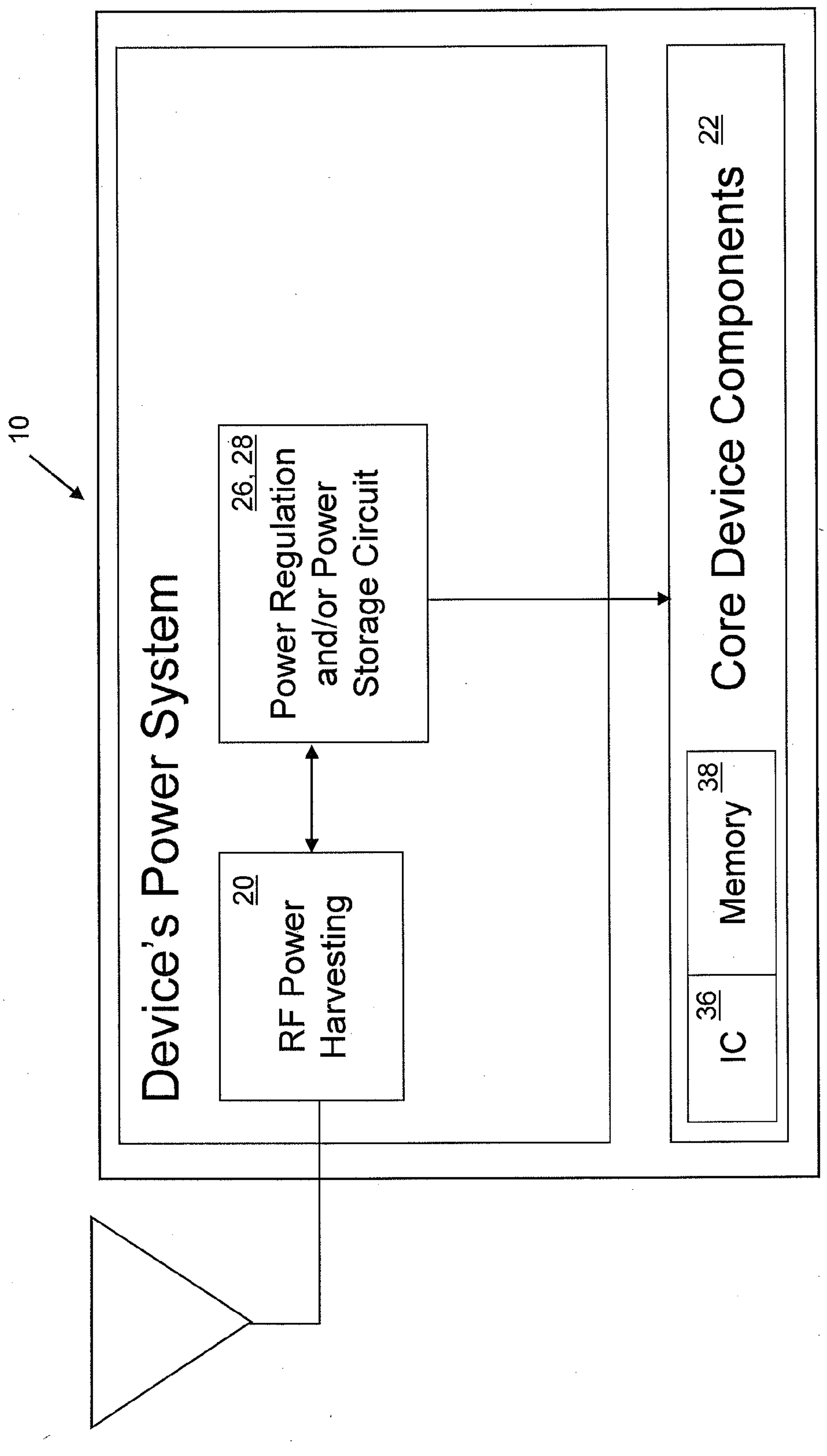
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Core Device Components
22
IC
36
Memory
38

Fig. 2

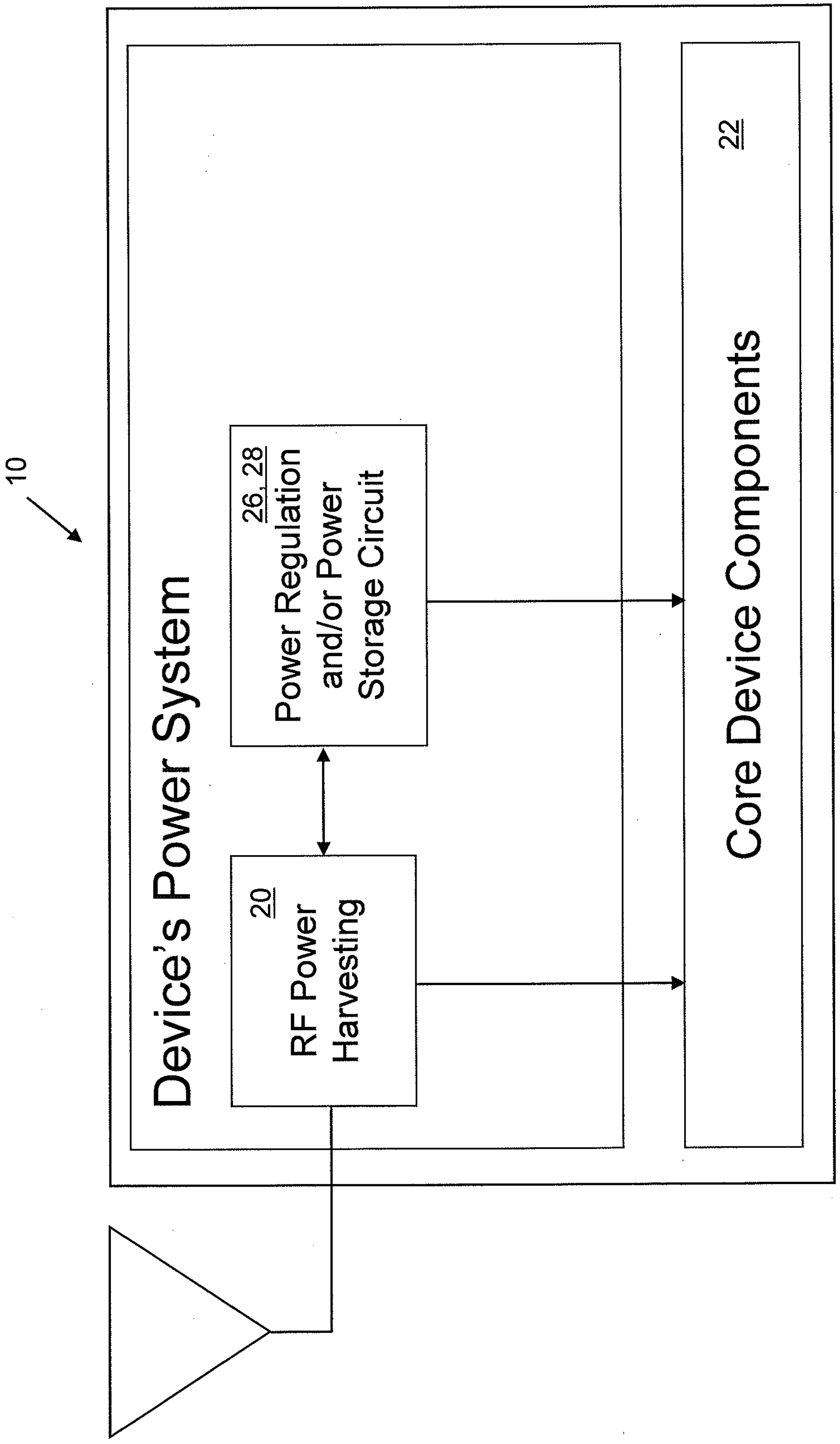
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Core Device Components
22

Fig. 3

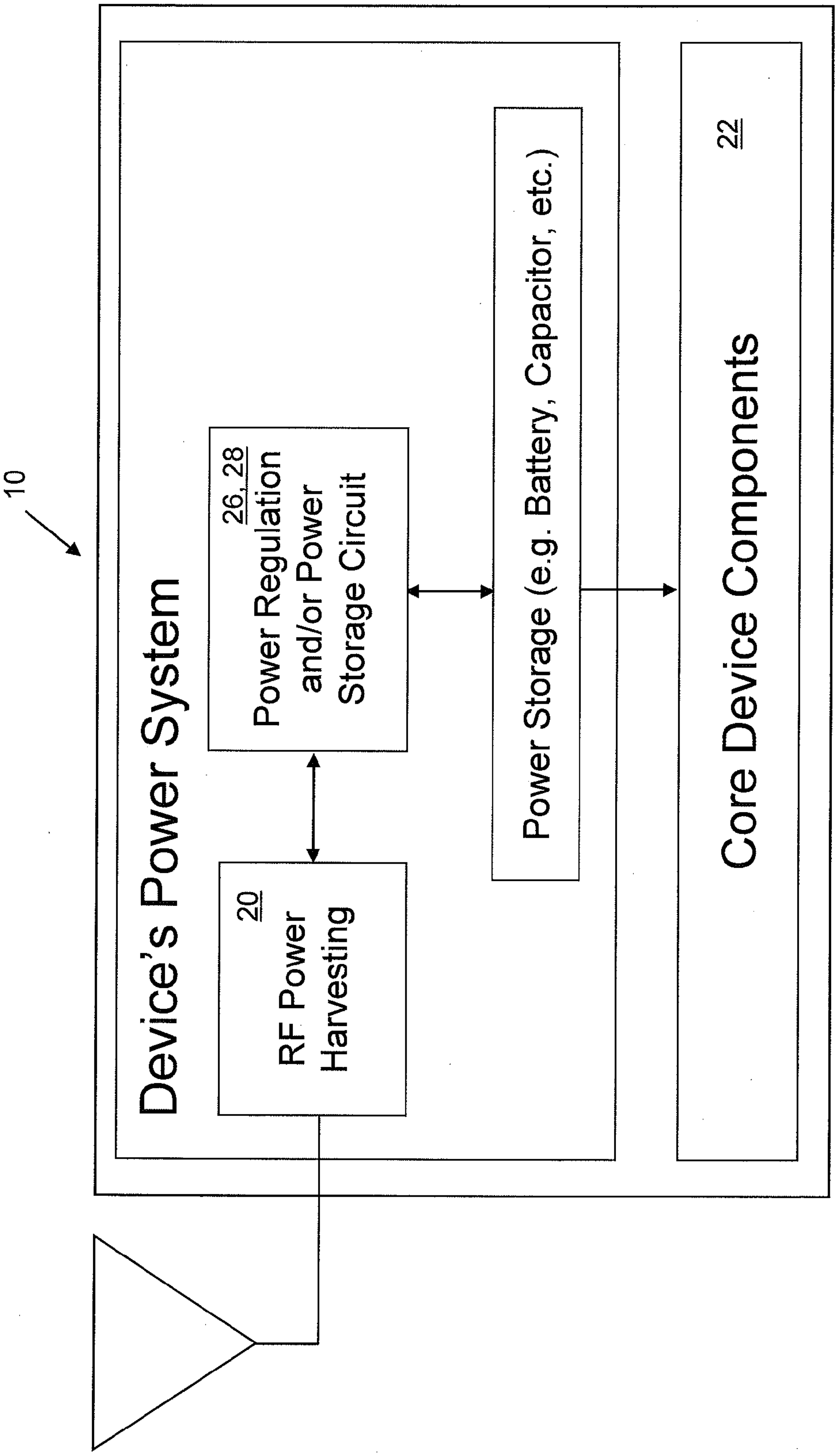
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 4

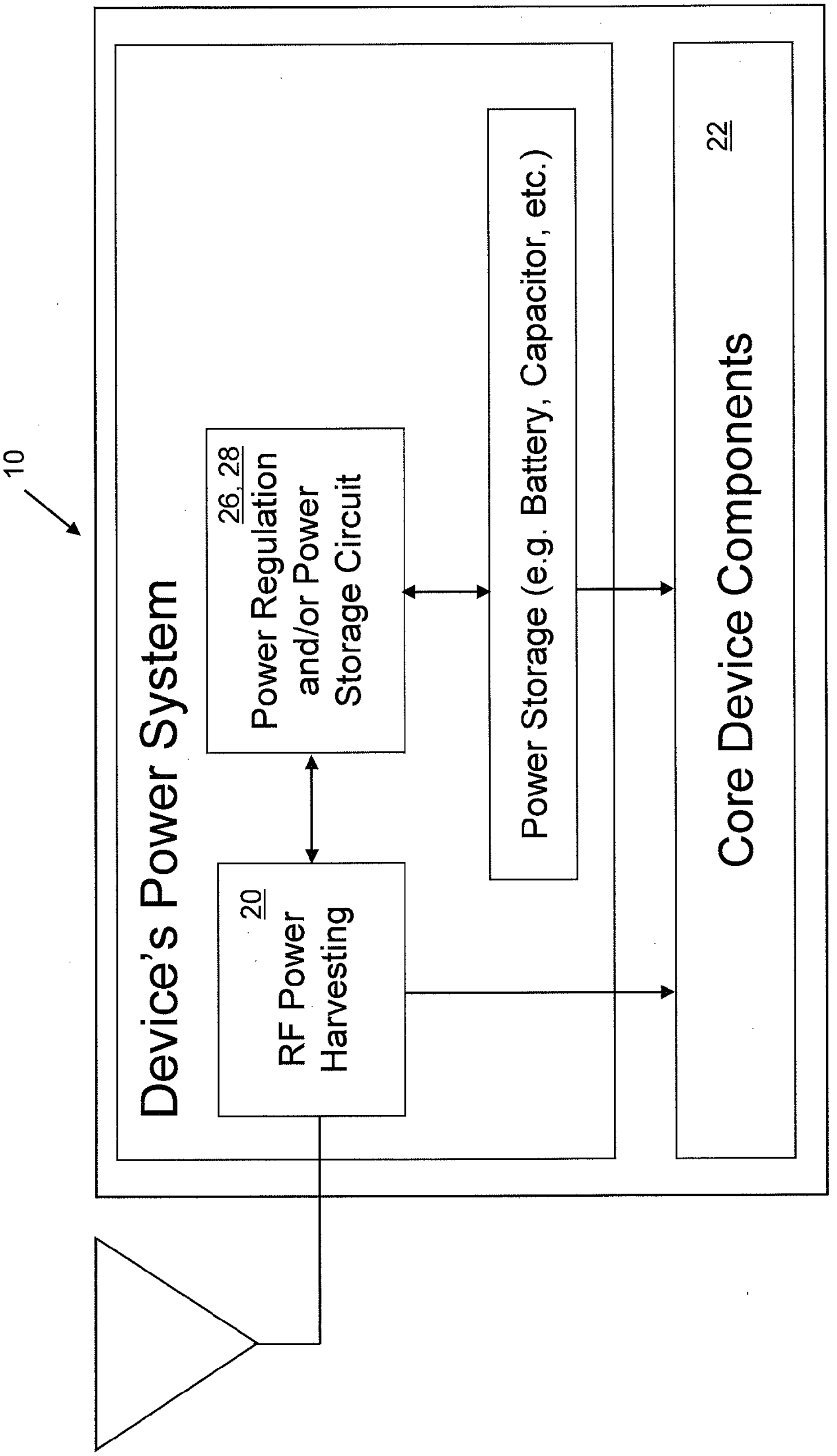
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 5

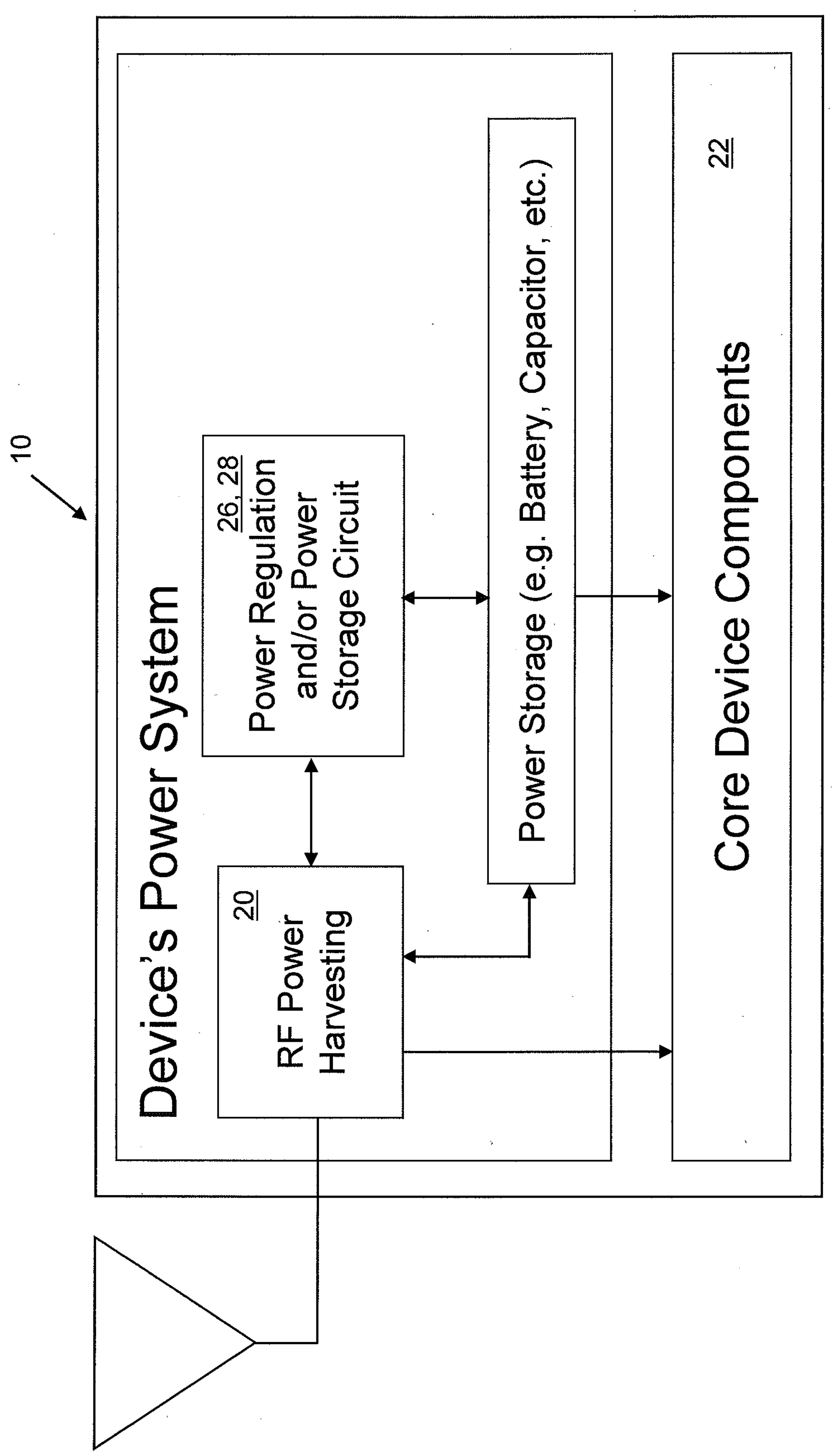
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 6

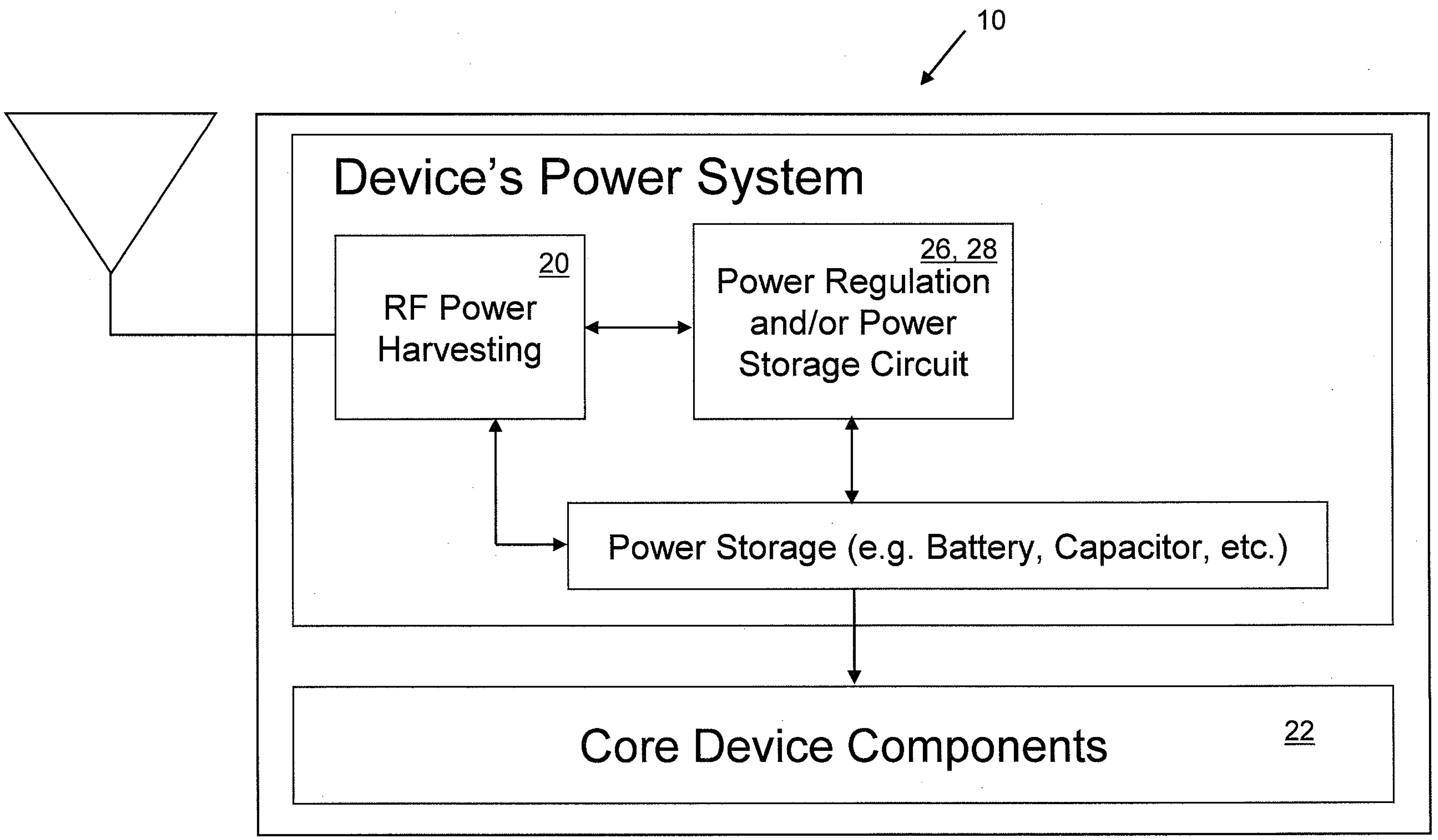
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 7

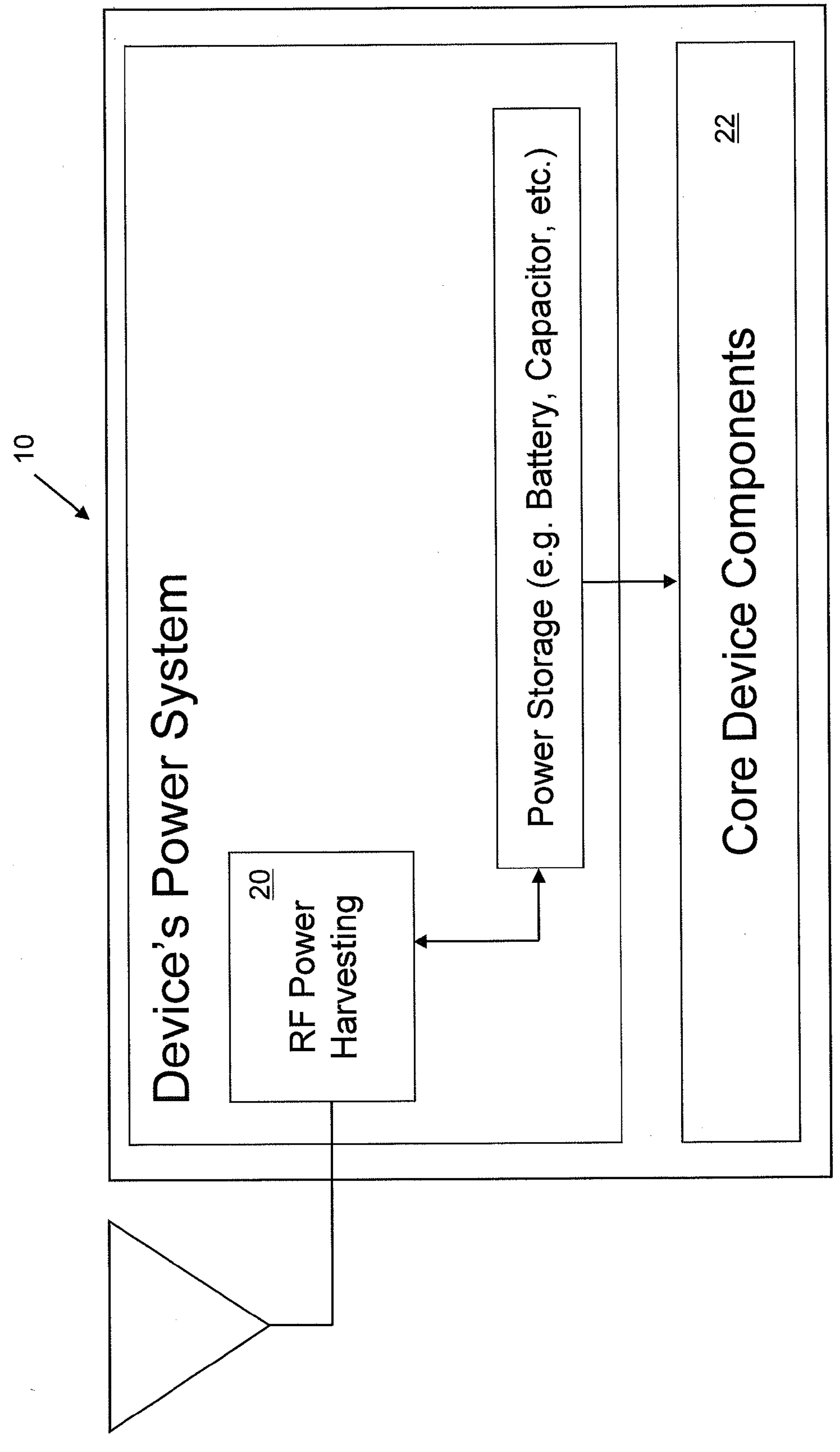
10
Device's Power System
RF Power Harvesting
20
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 8

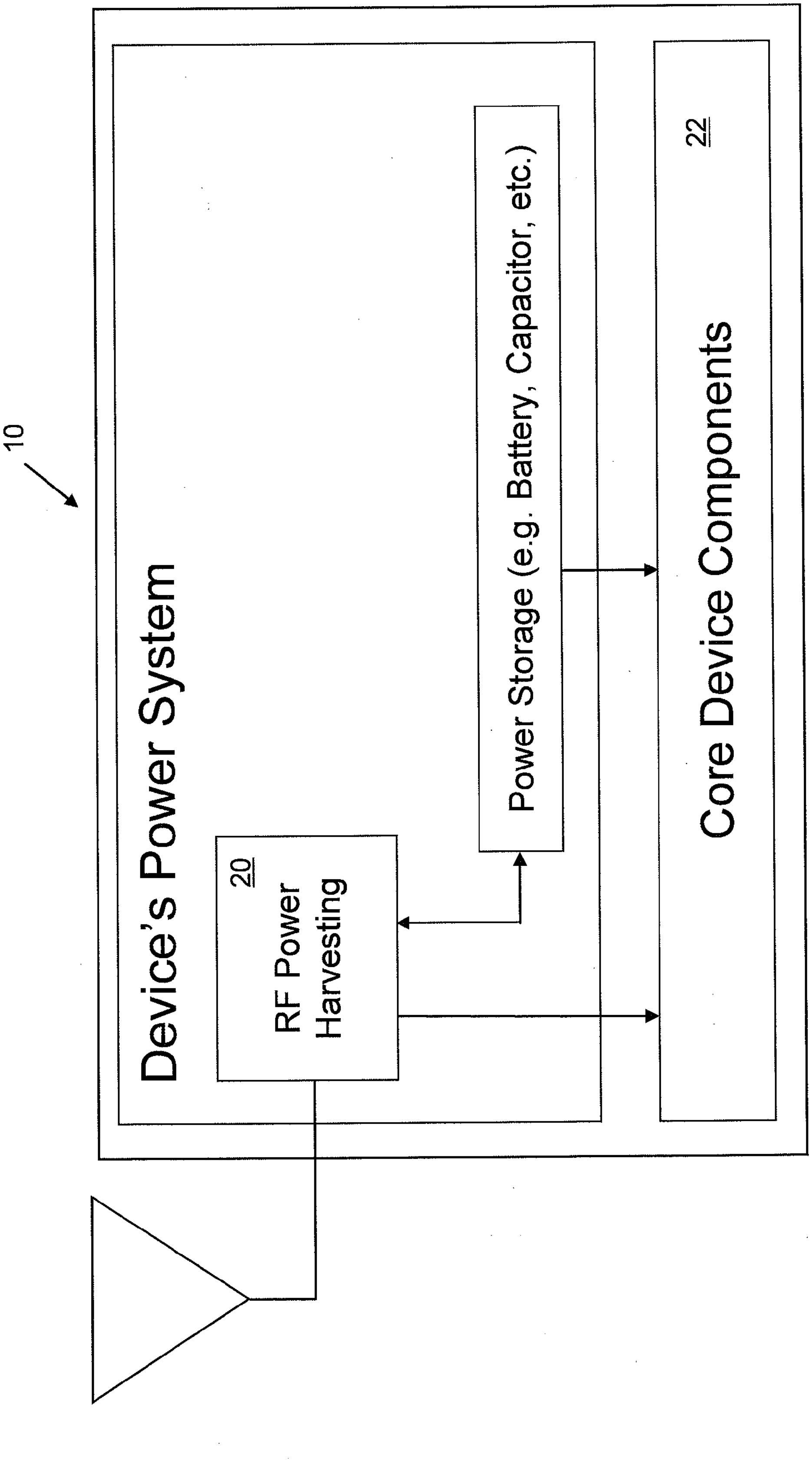
10
Device's Power System
20
RF Power Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 9

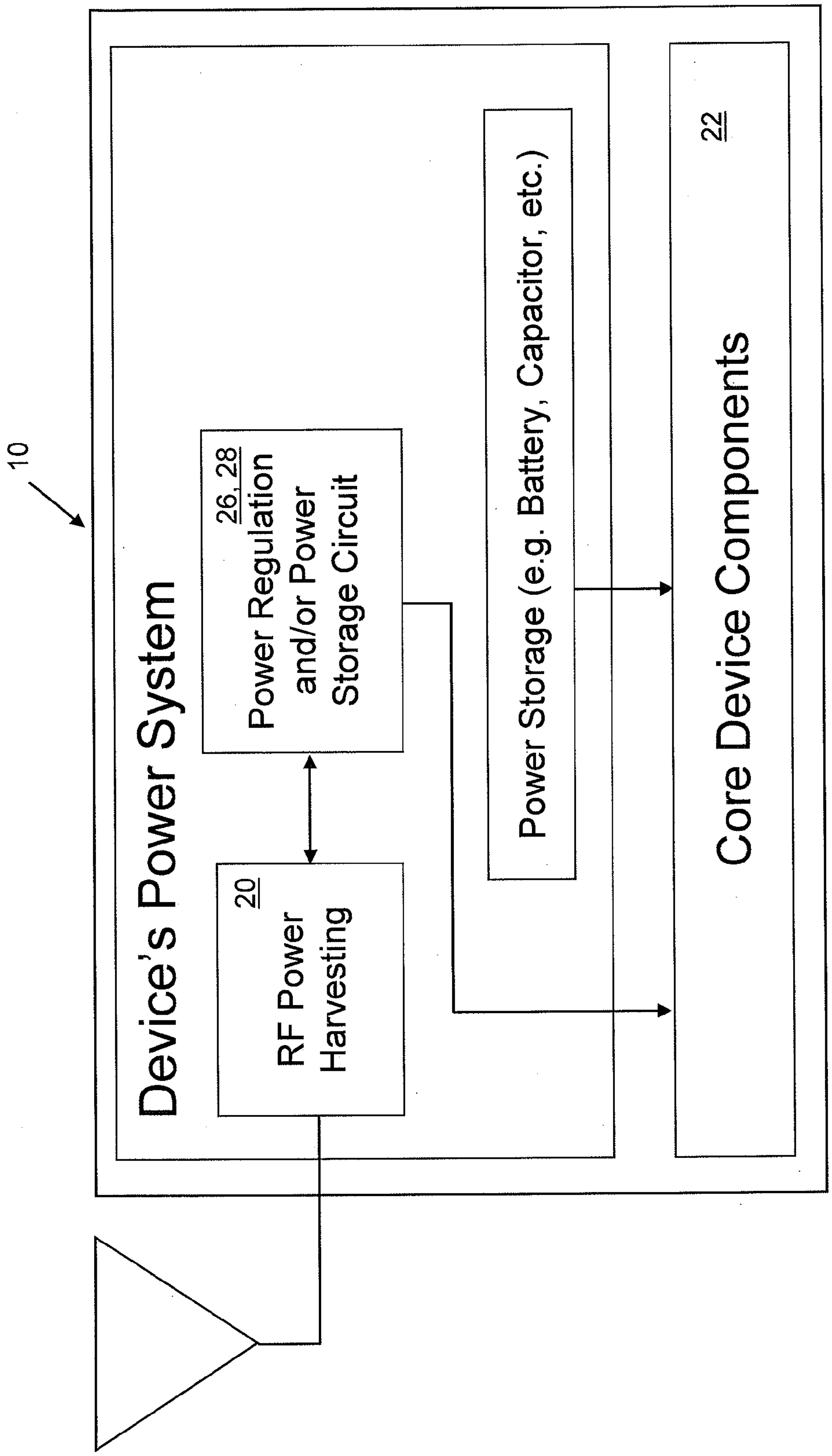
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 10

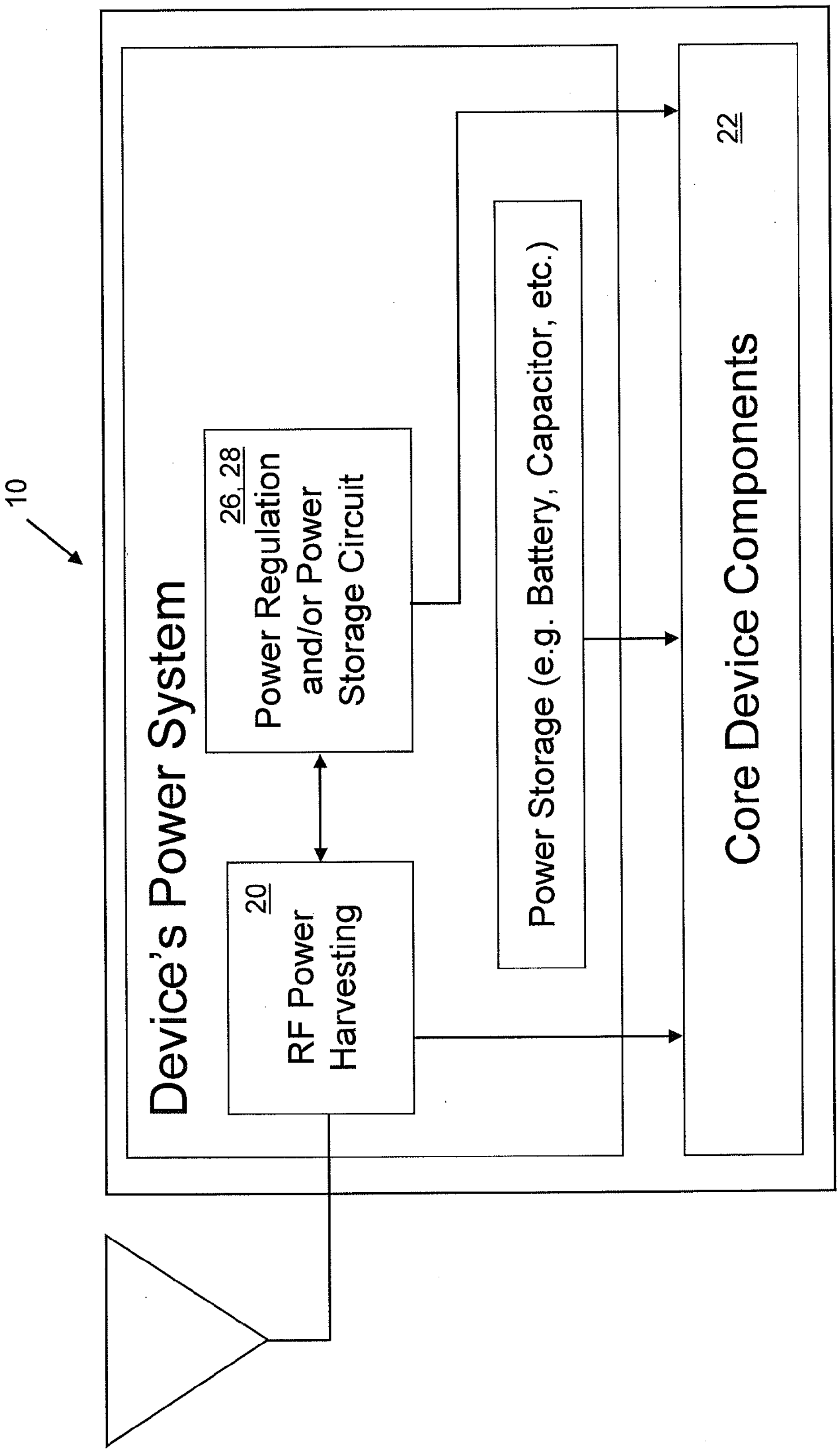
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 11

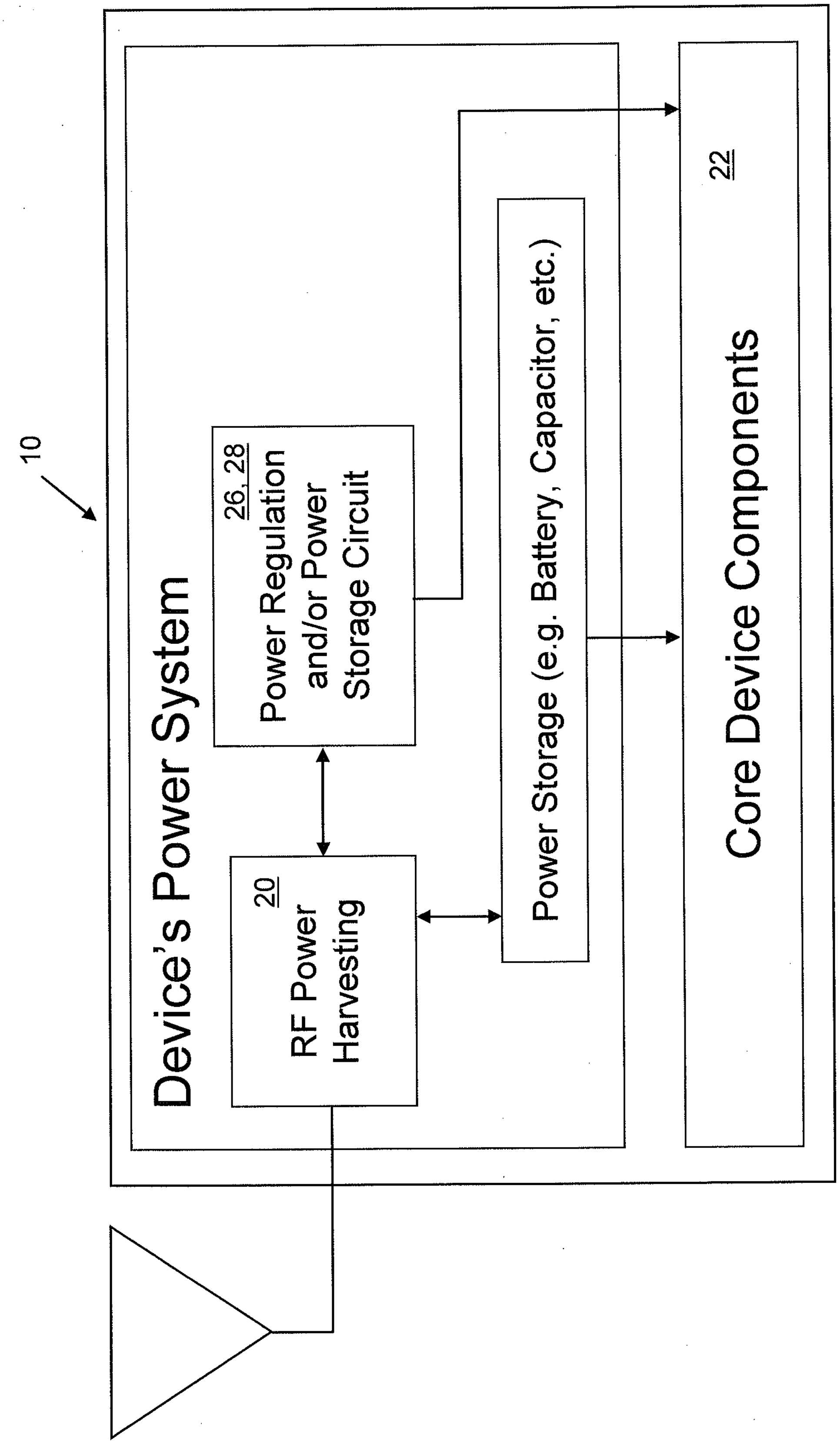
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 12

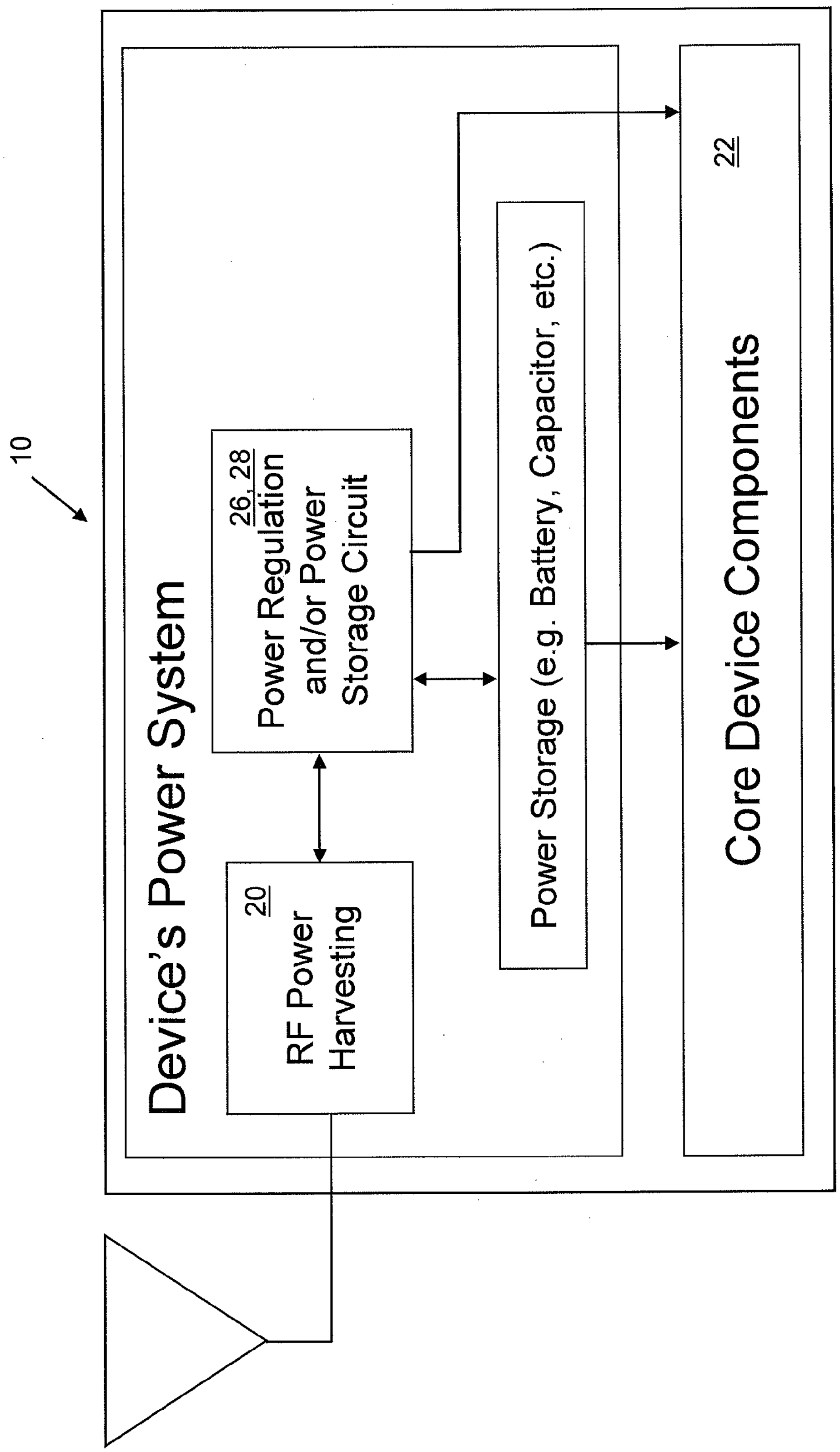
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 13

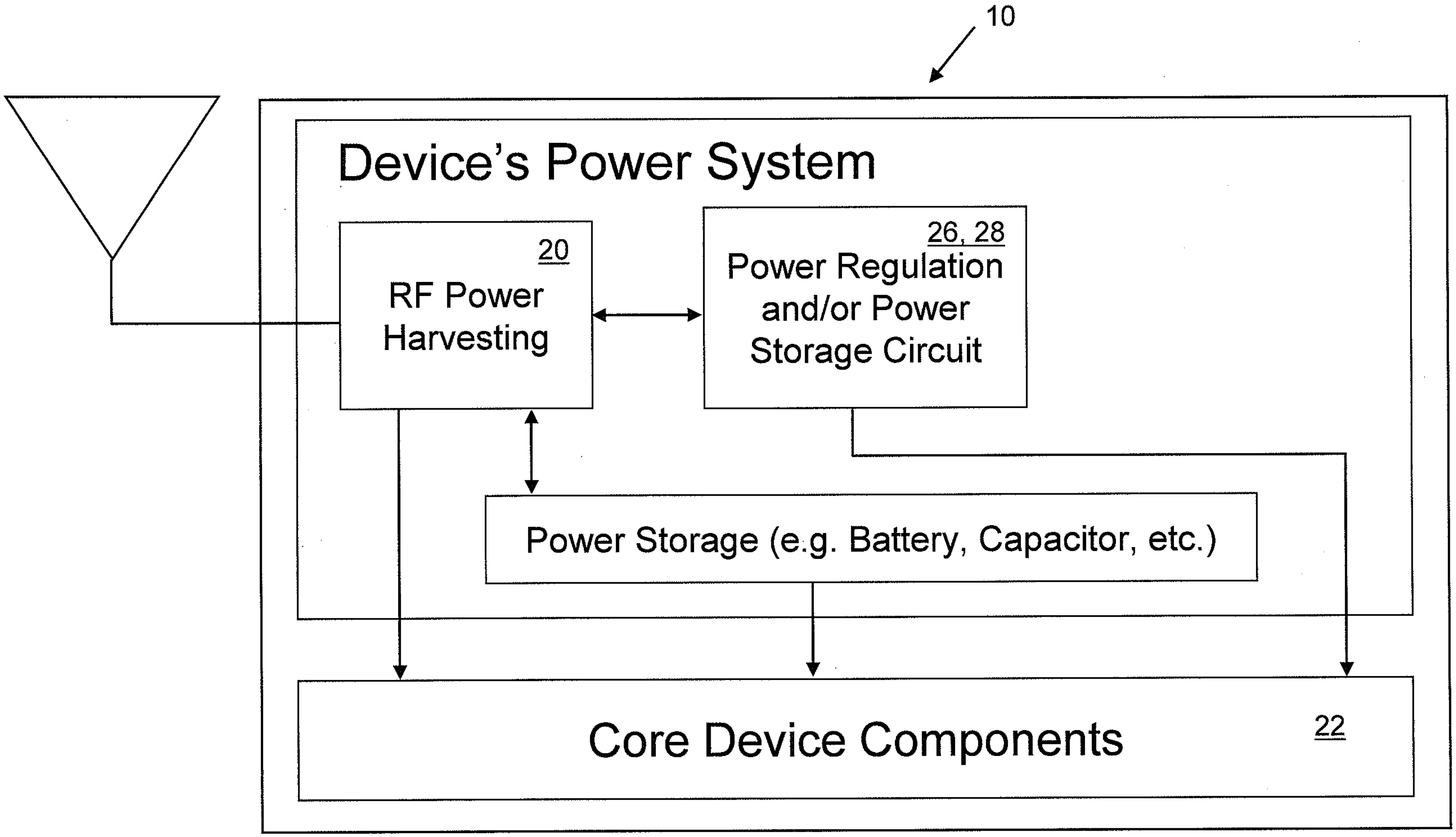
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 14

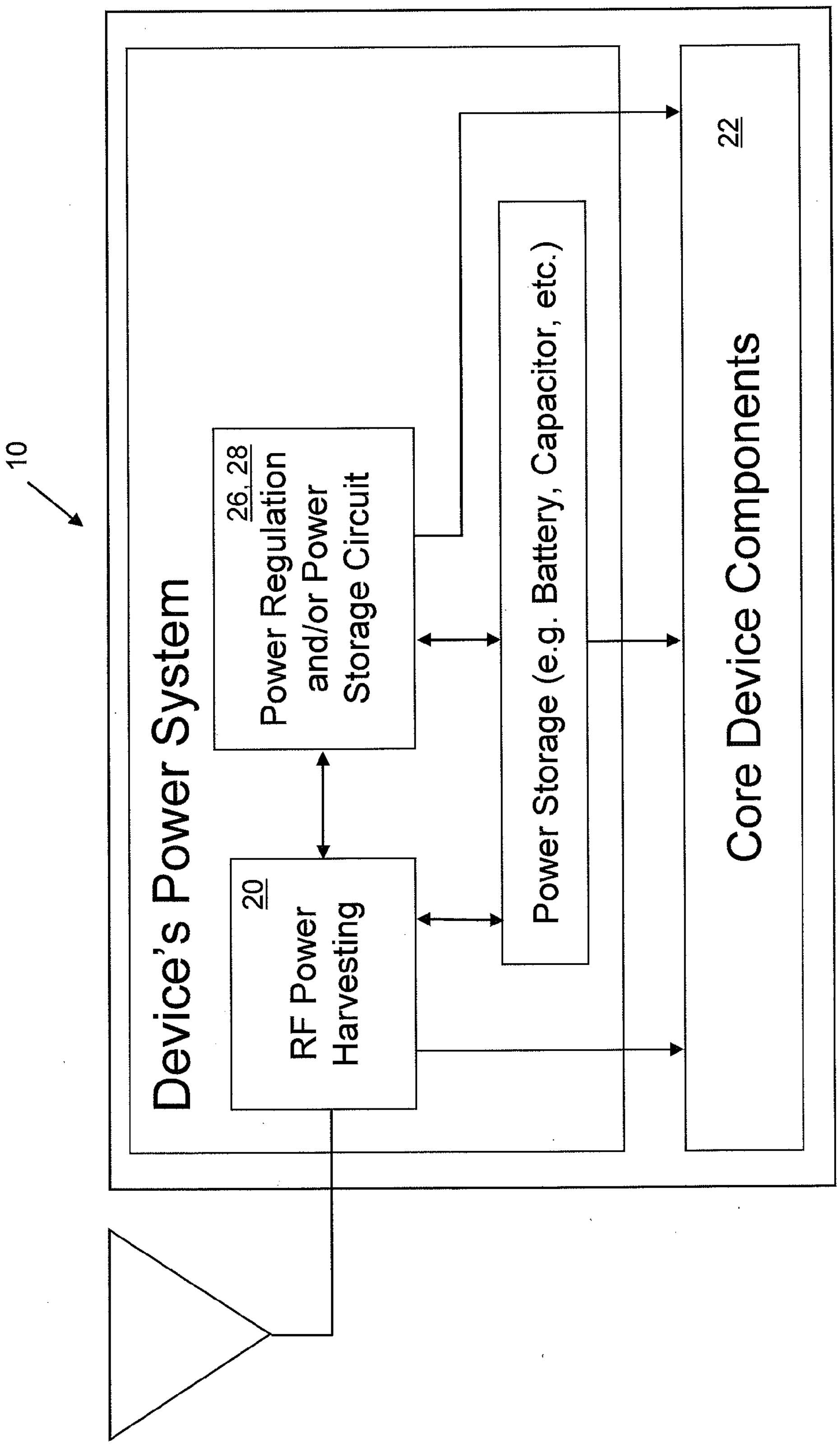
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 15

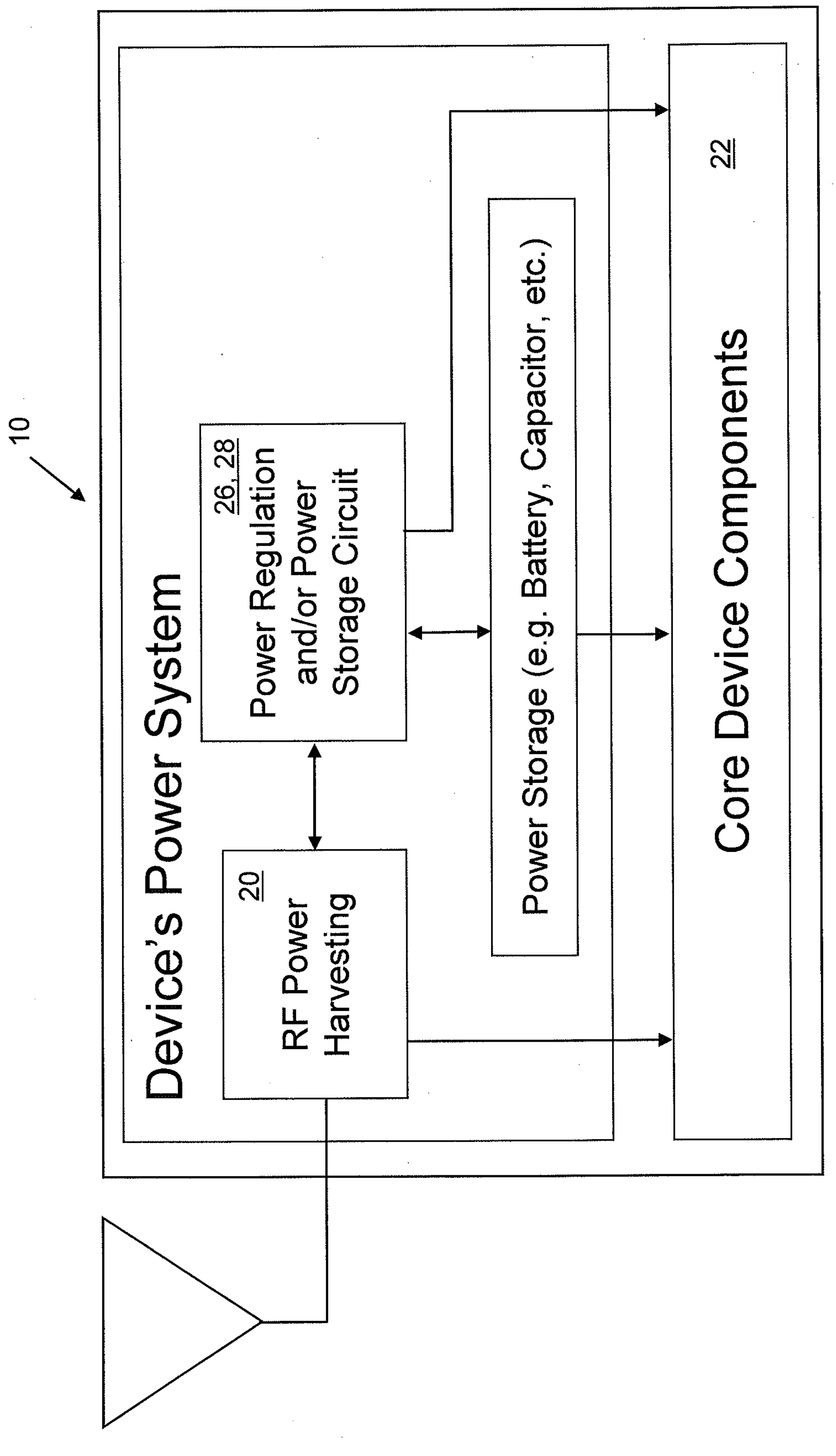
10
Device's Power System
26, 28
Power Regulation and/or Power Storage Circuit
20
RF Power Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 16

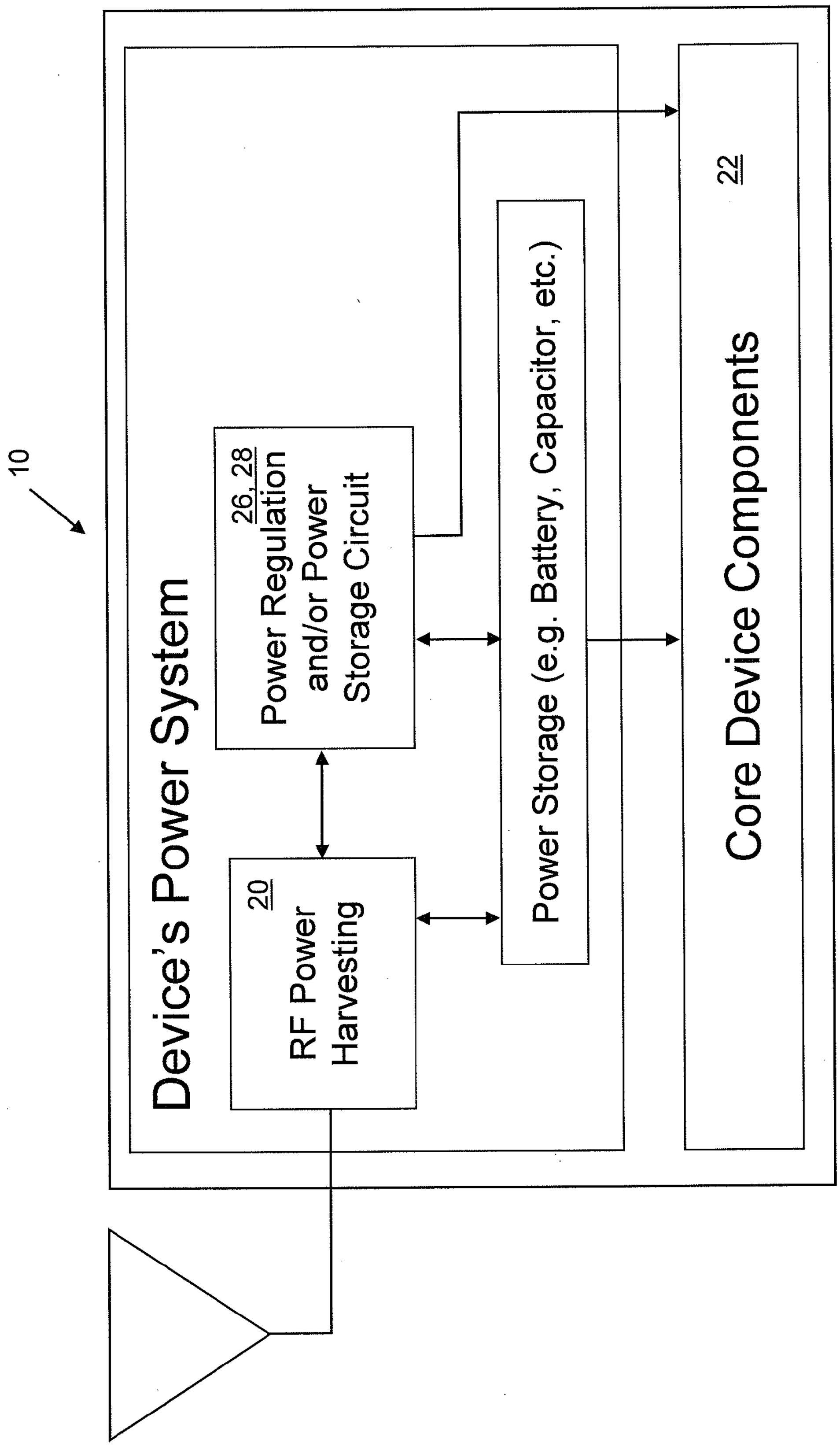
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 17

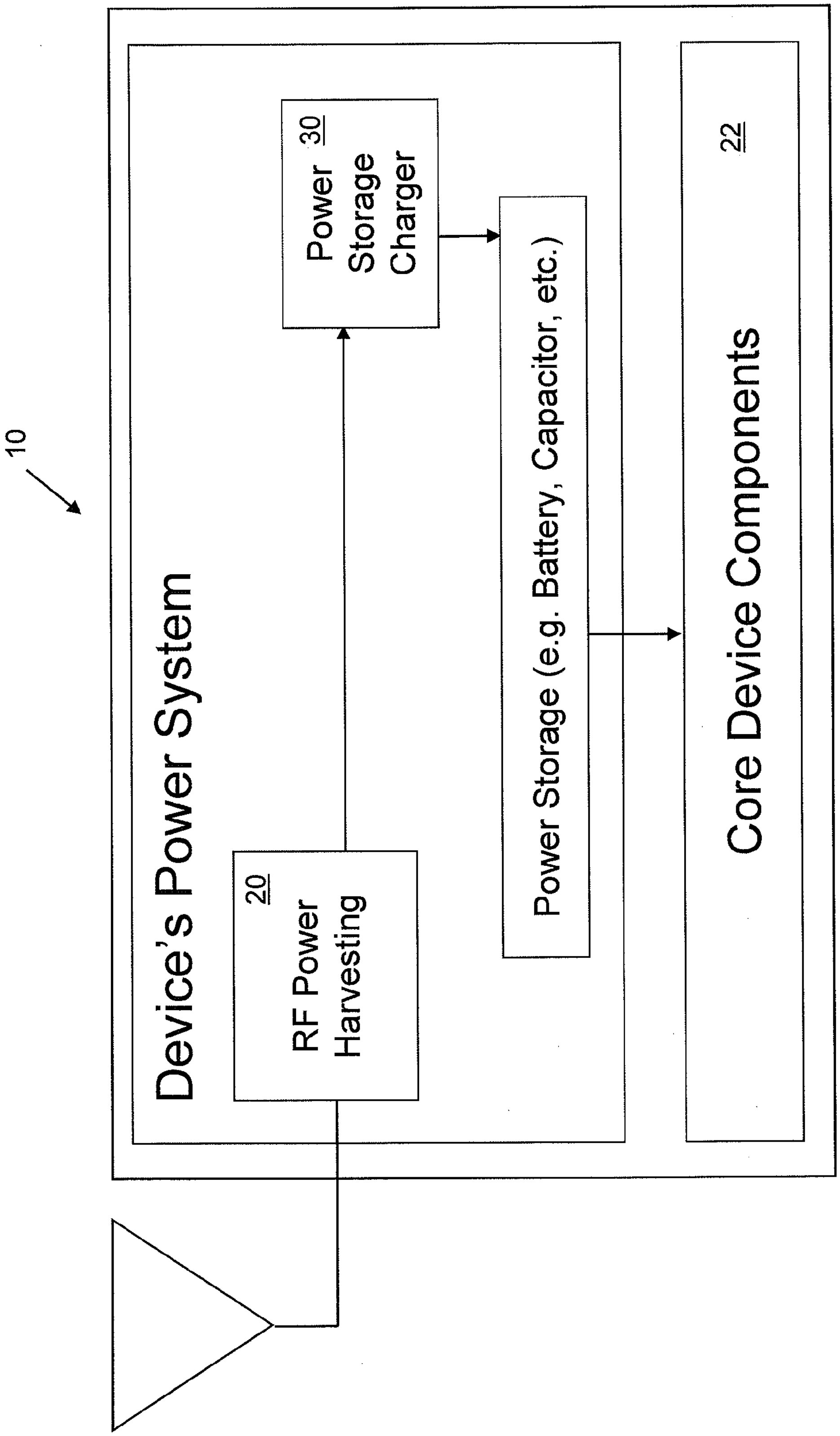
10
Device's Power System
20
RF Power Harvesting
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 18

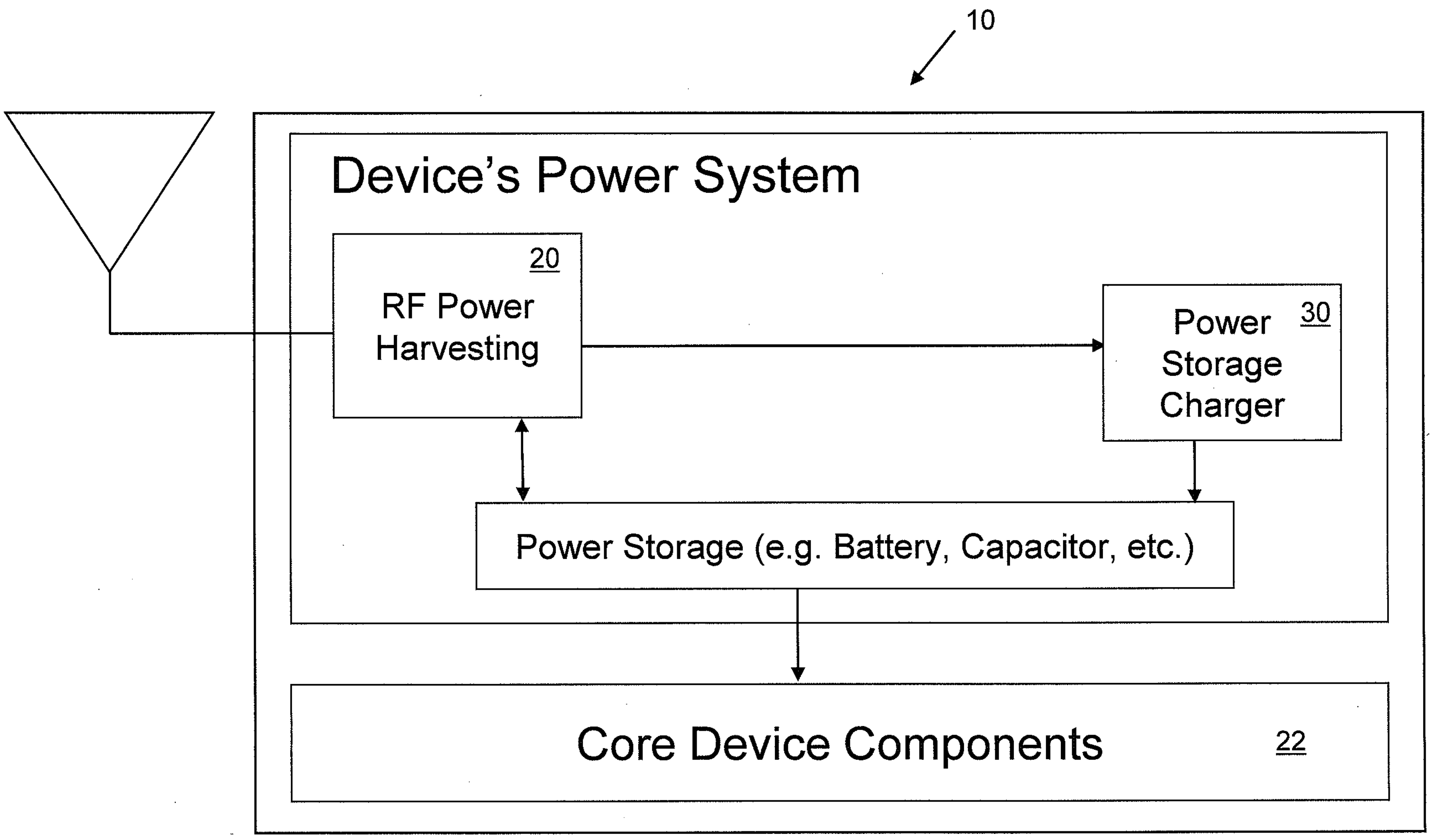
10
Device's Power System
20
RF Power Harvesting
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 19

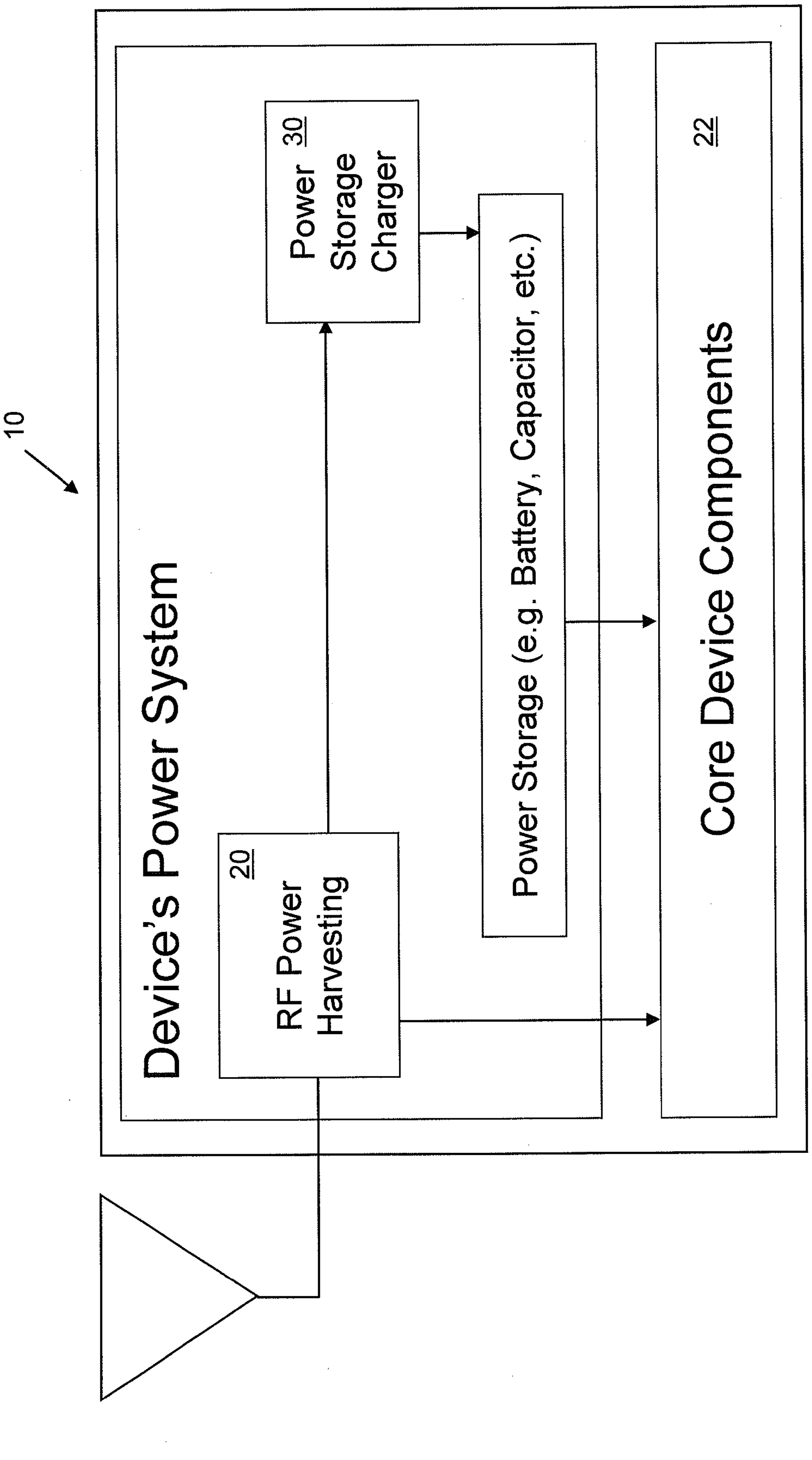
10
Device's Power System
20
RF Power Harvesting
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 20

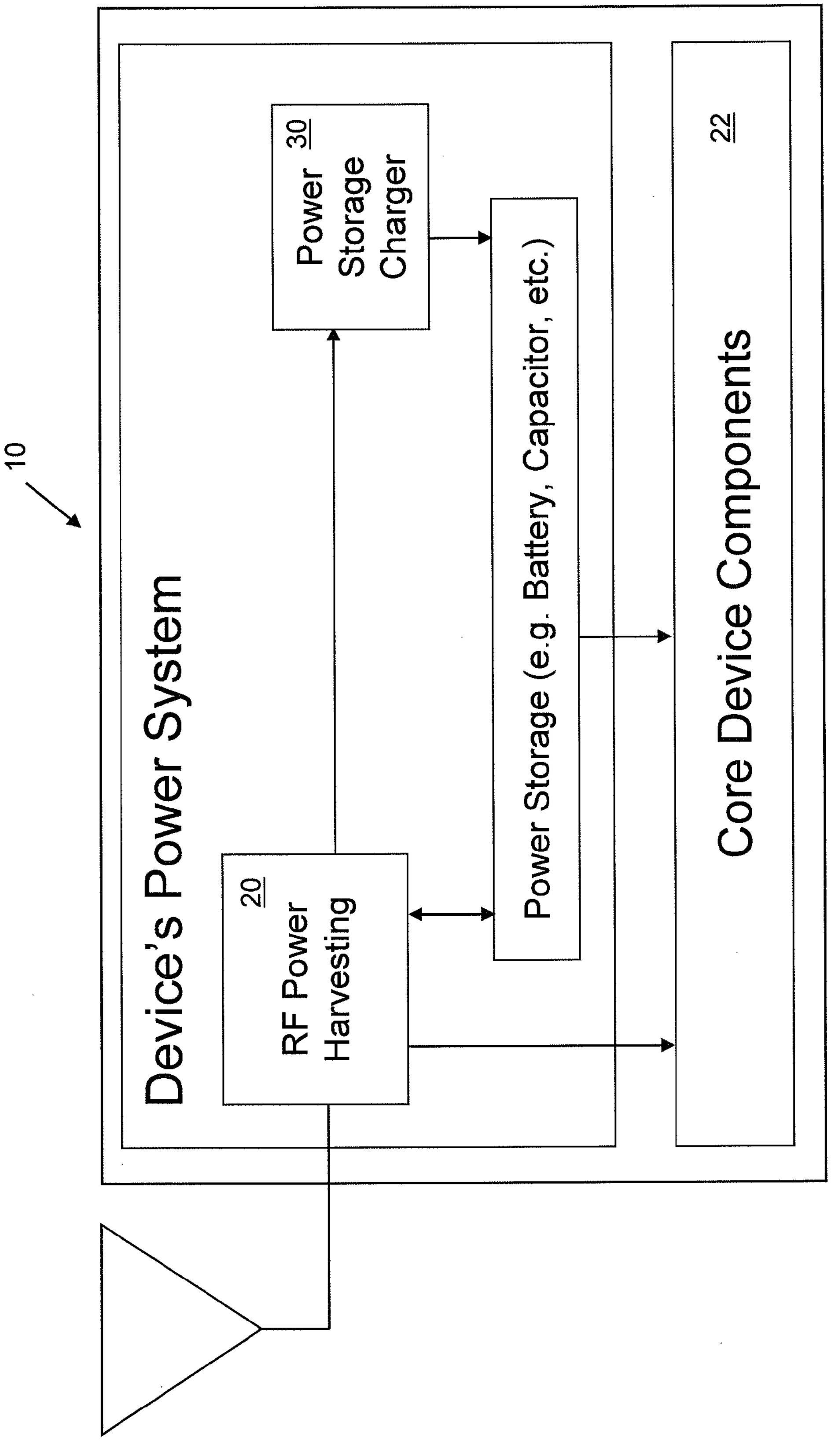
10
Device's Power System
20
RF Power Harvesting
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 21

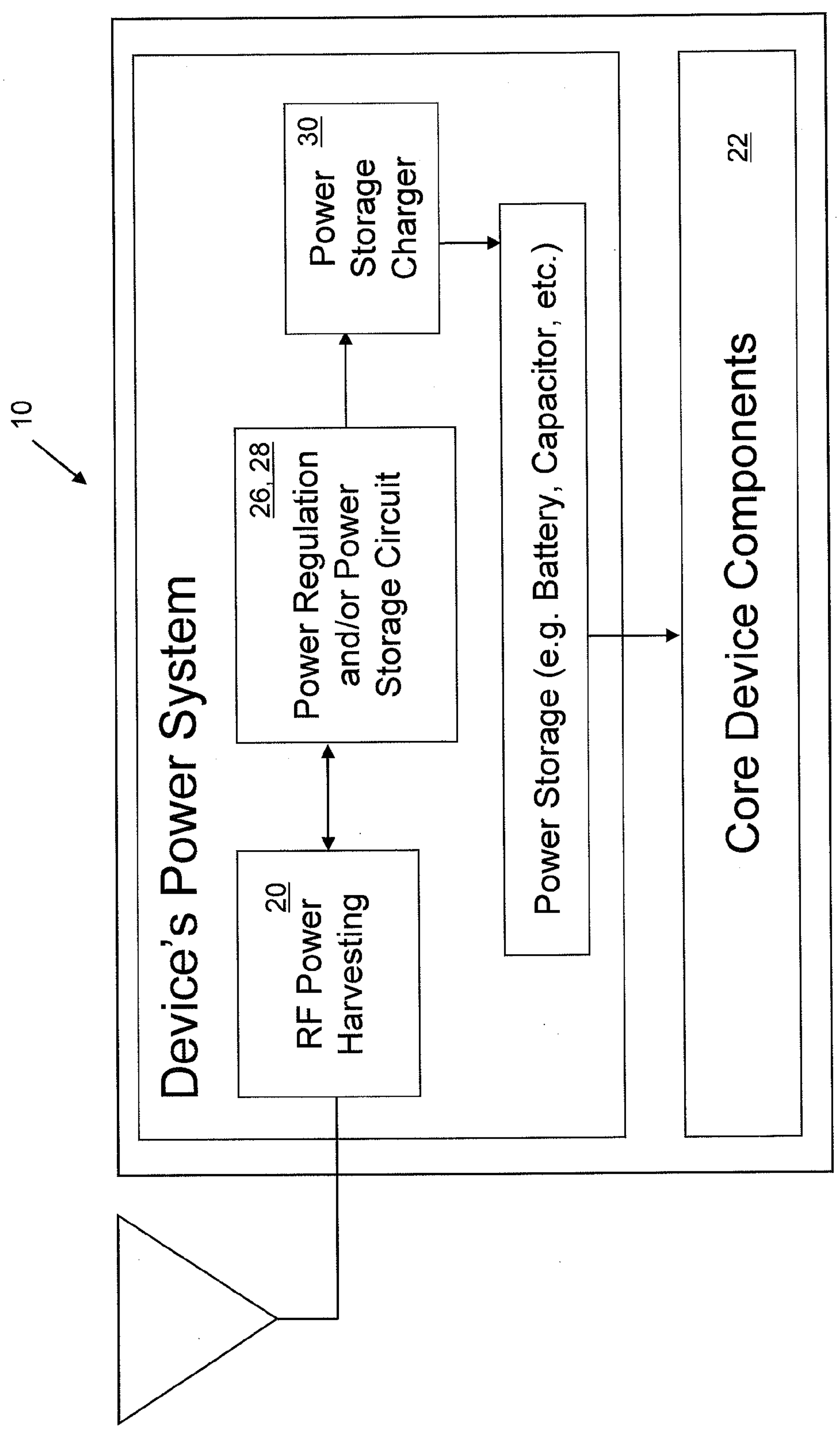
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 22

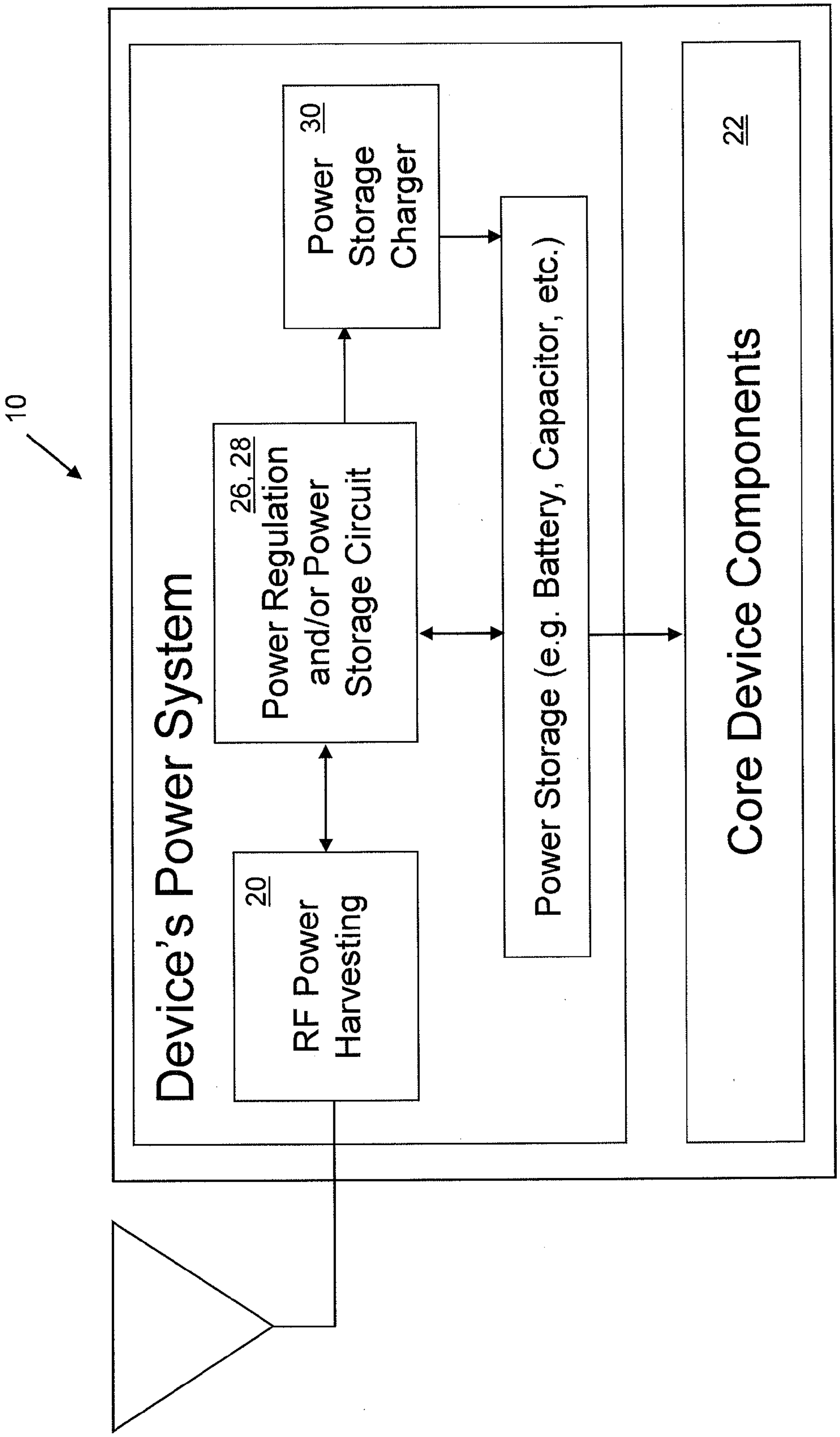
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 23

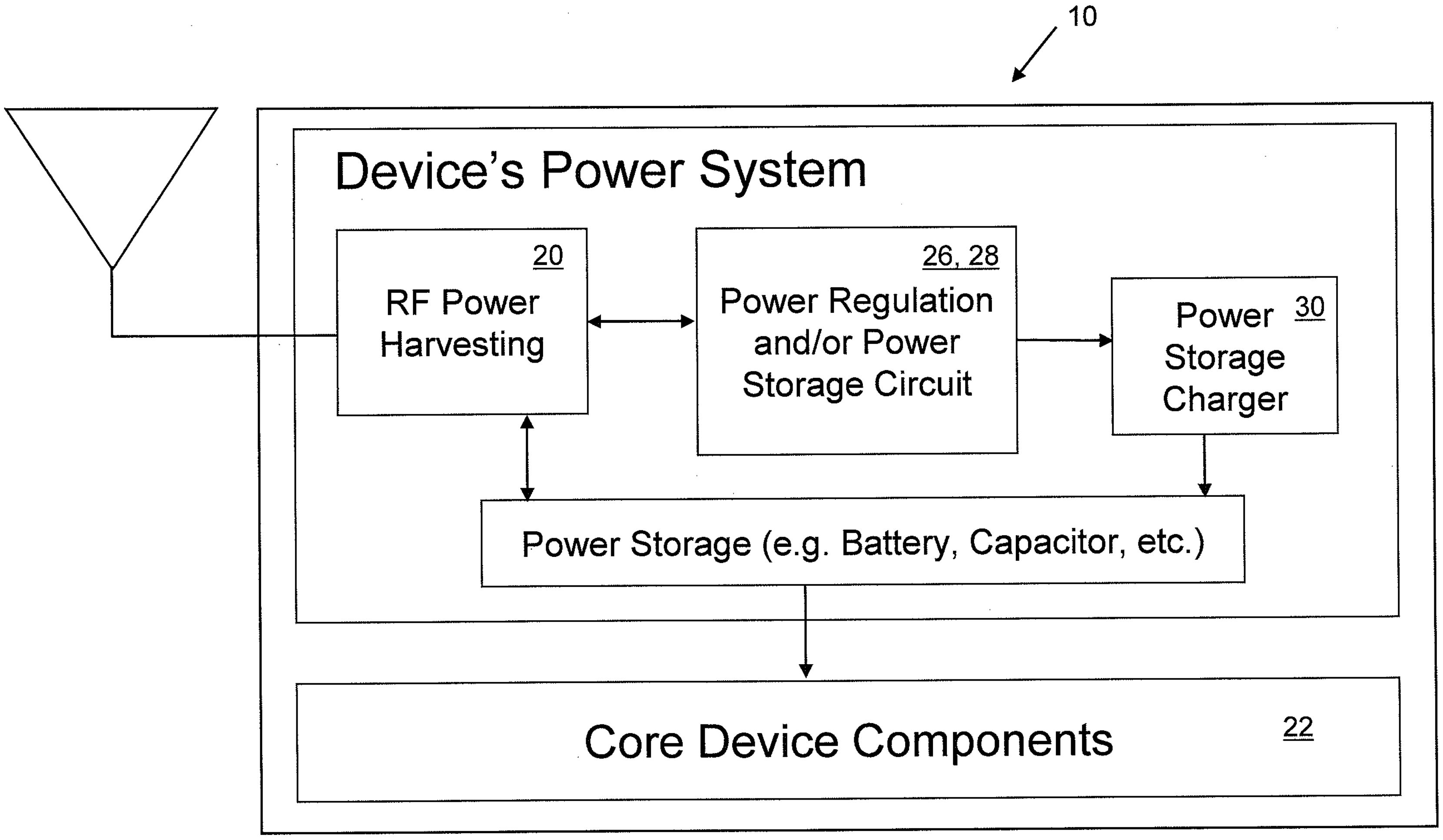
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage
Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 24

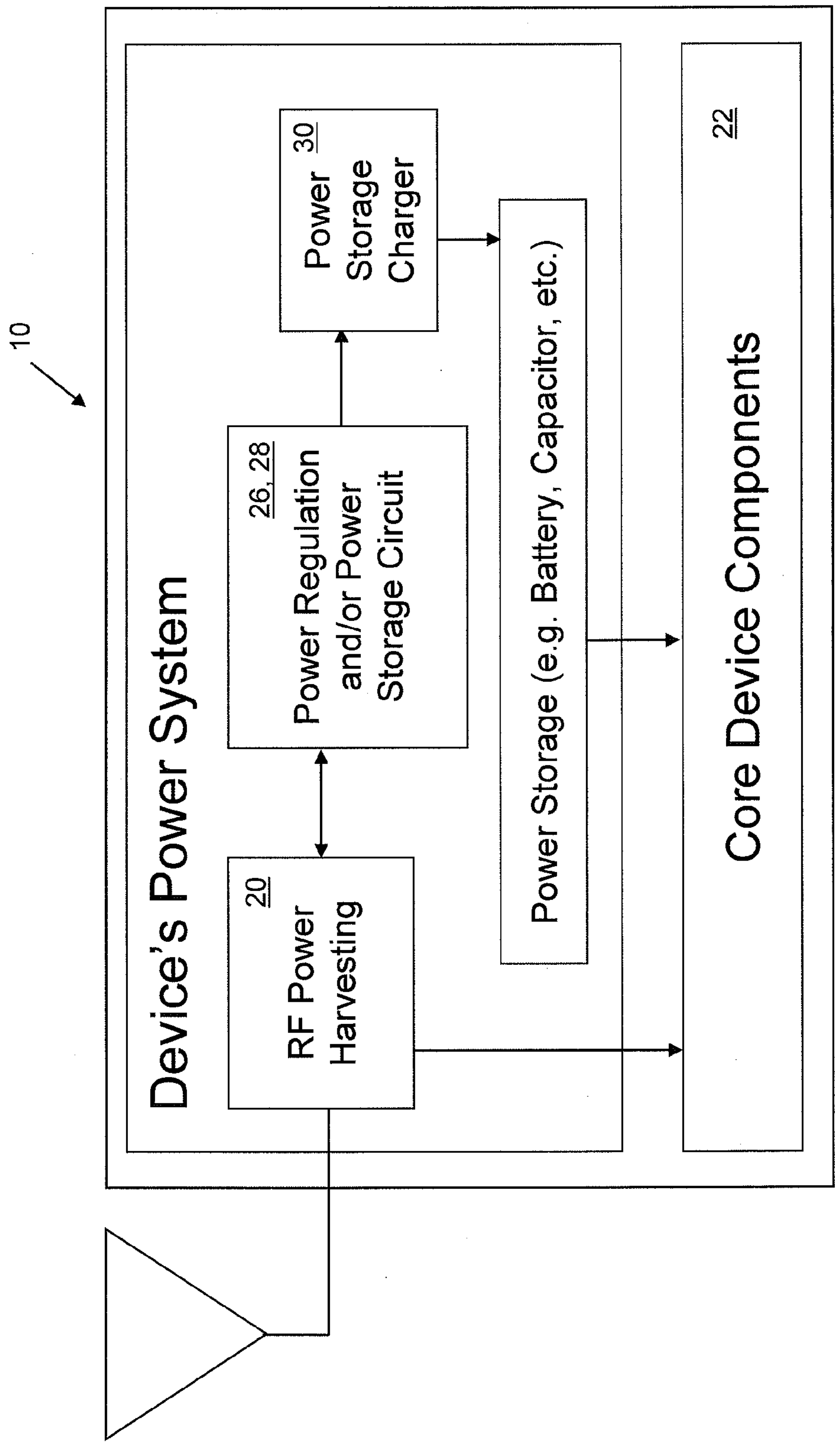
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 25

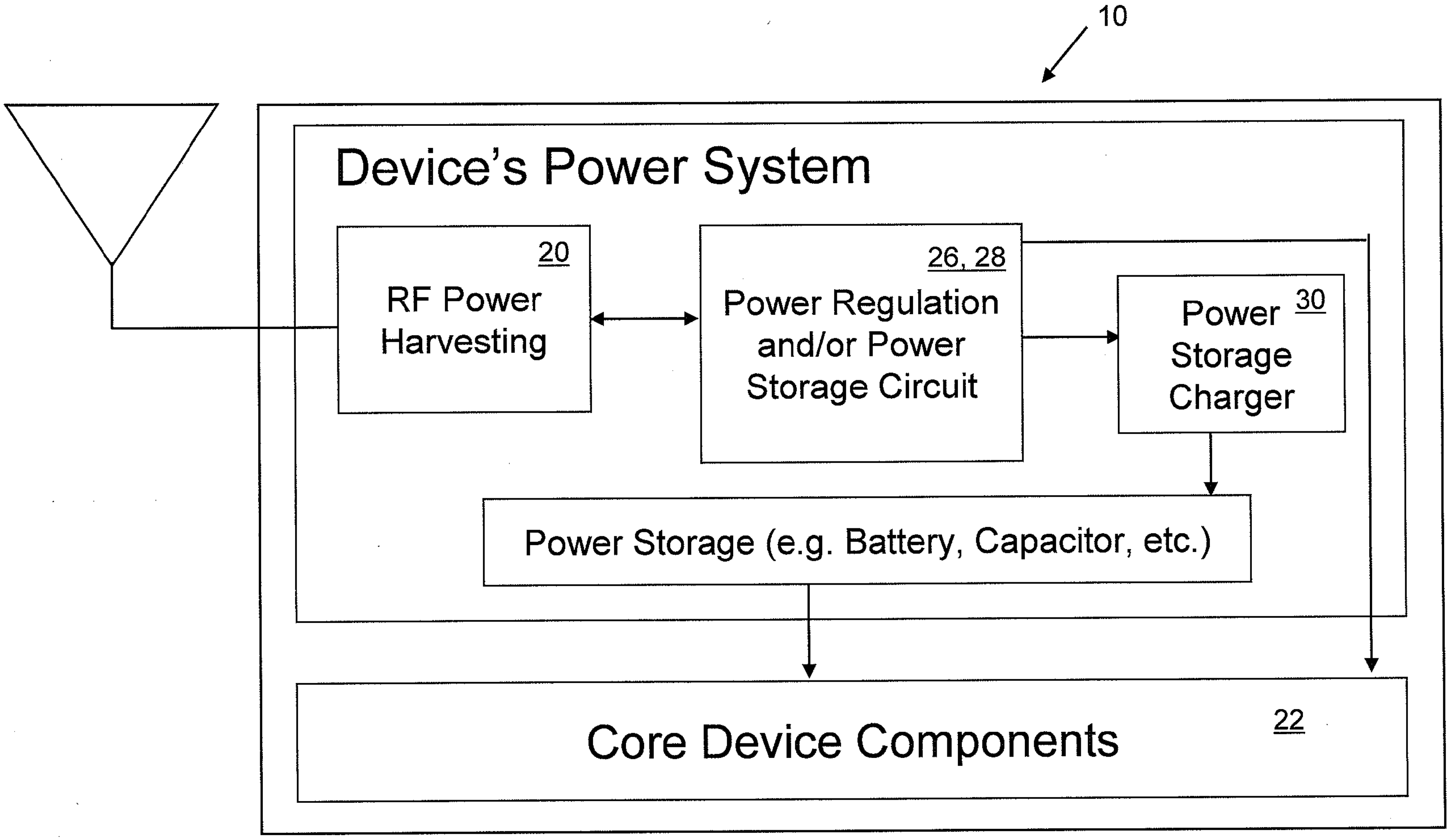
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 26

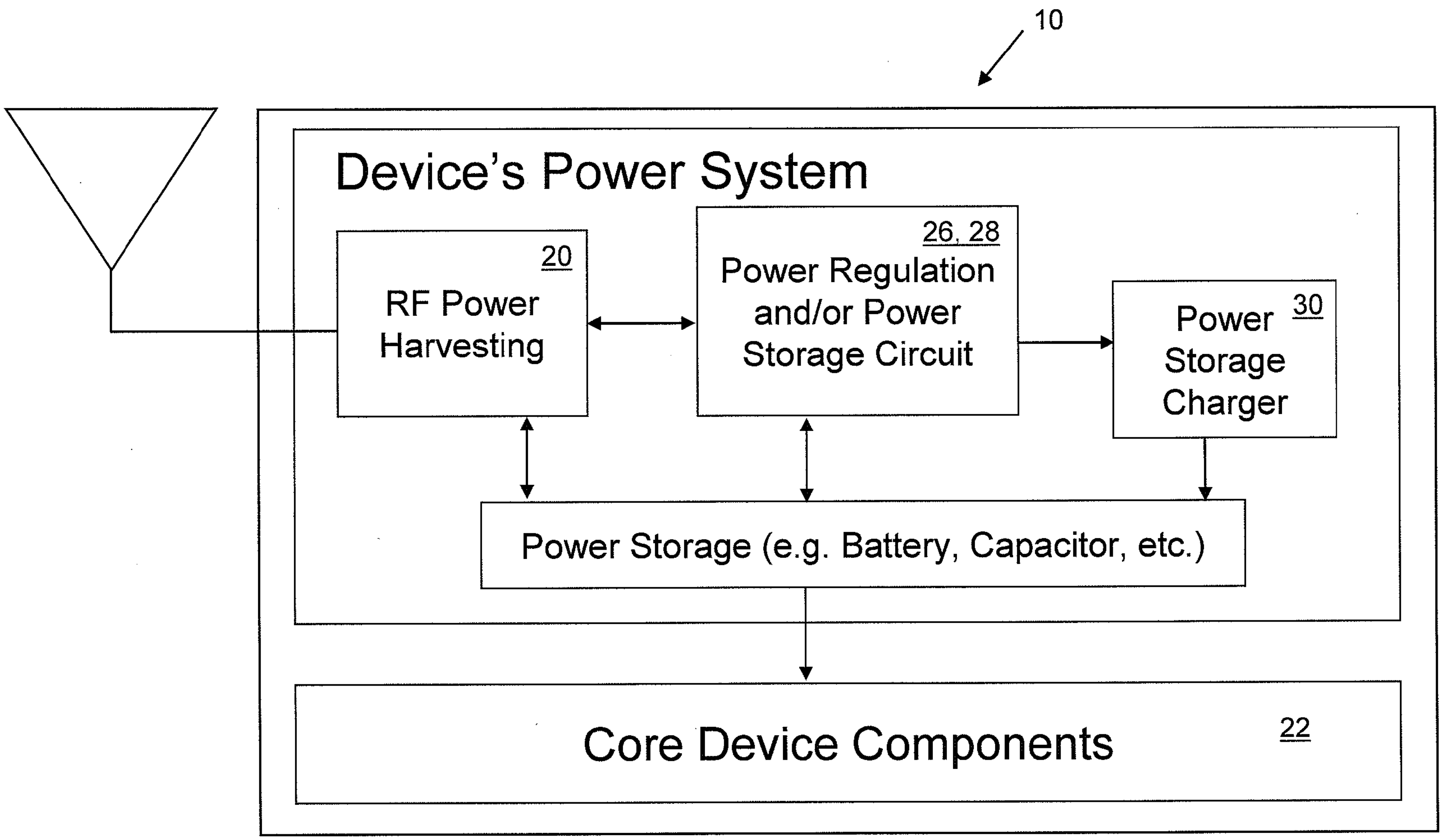
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 27

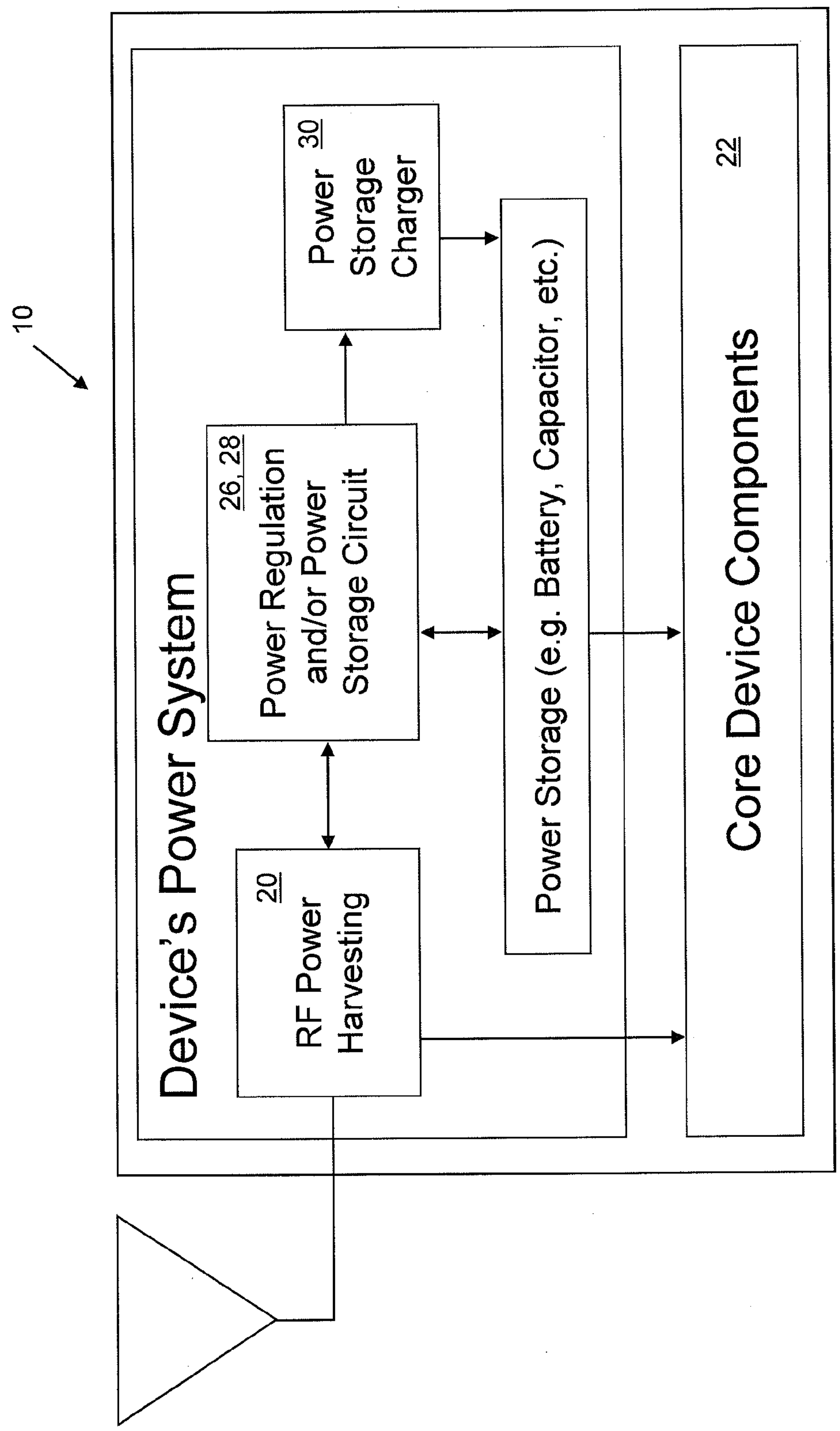
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 28

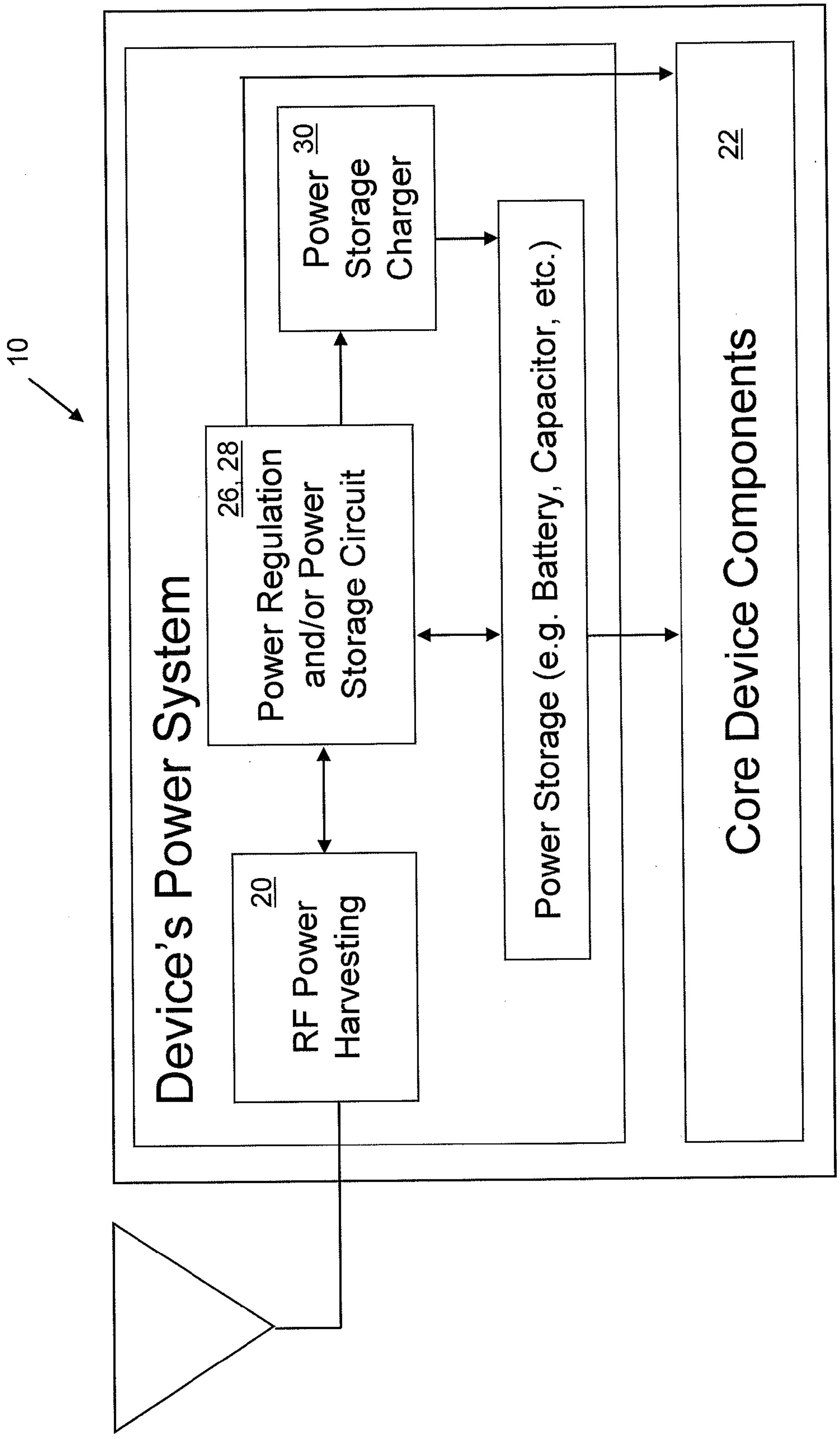
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 29

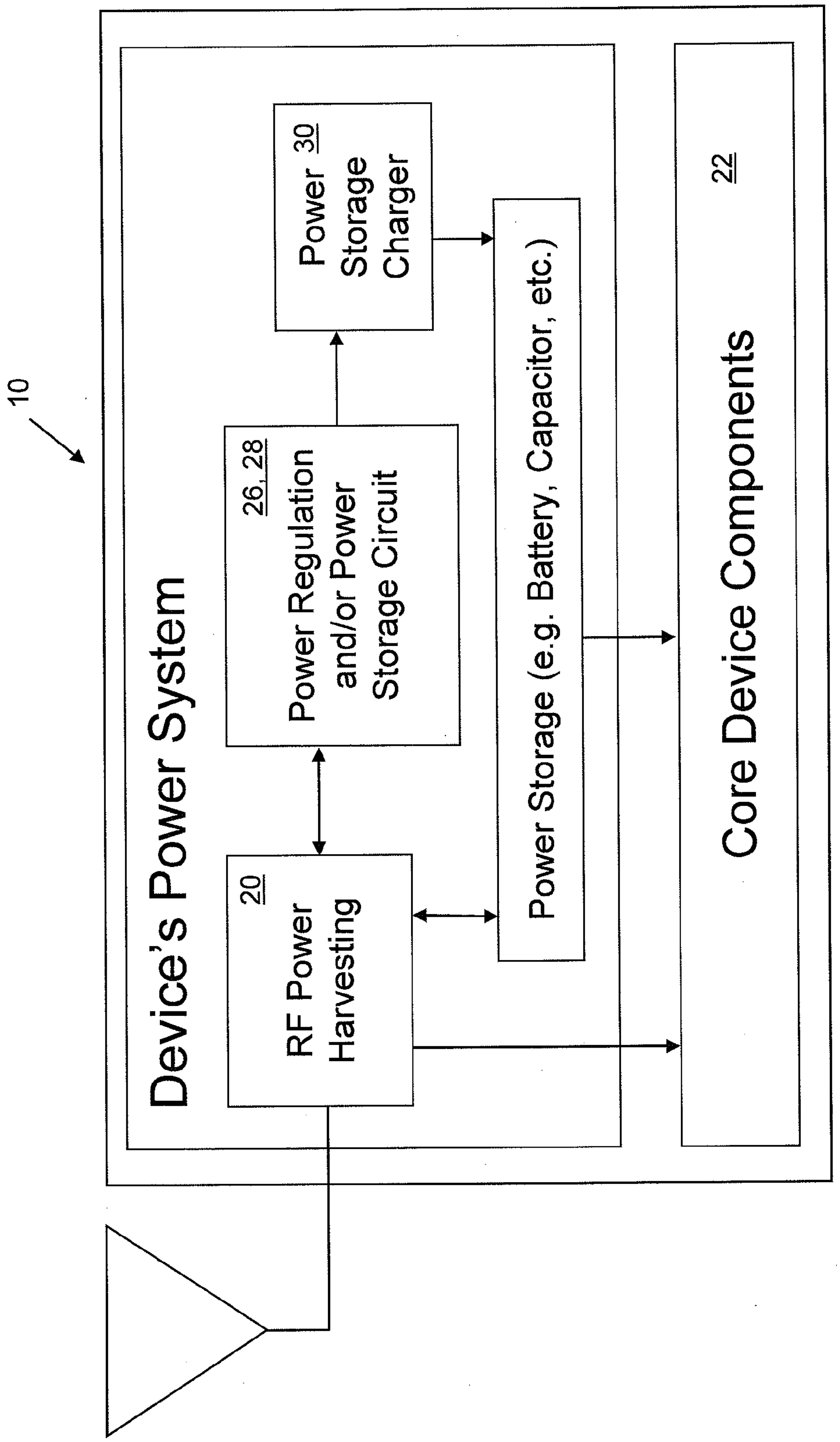
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30 Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 30

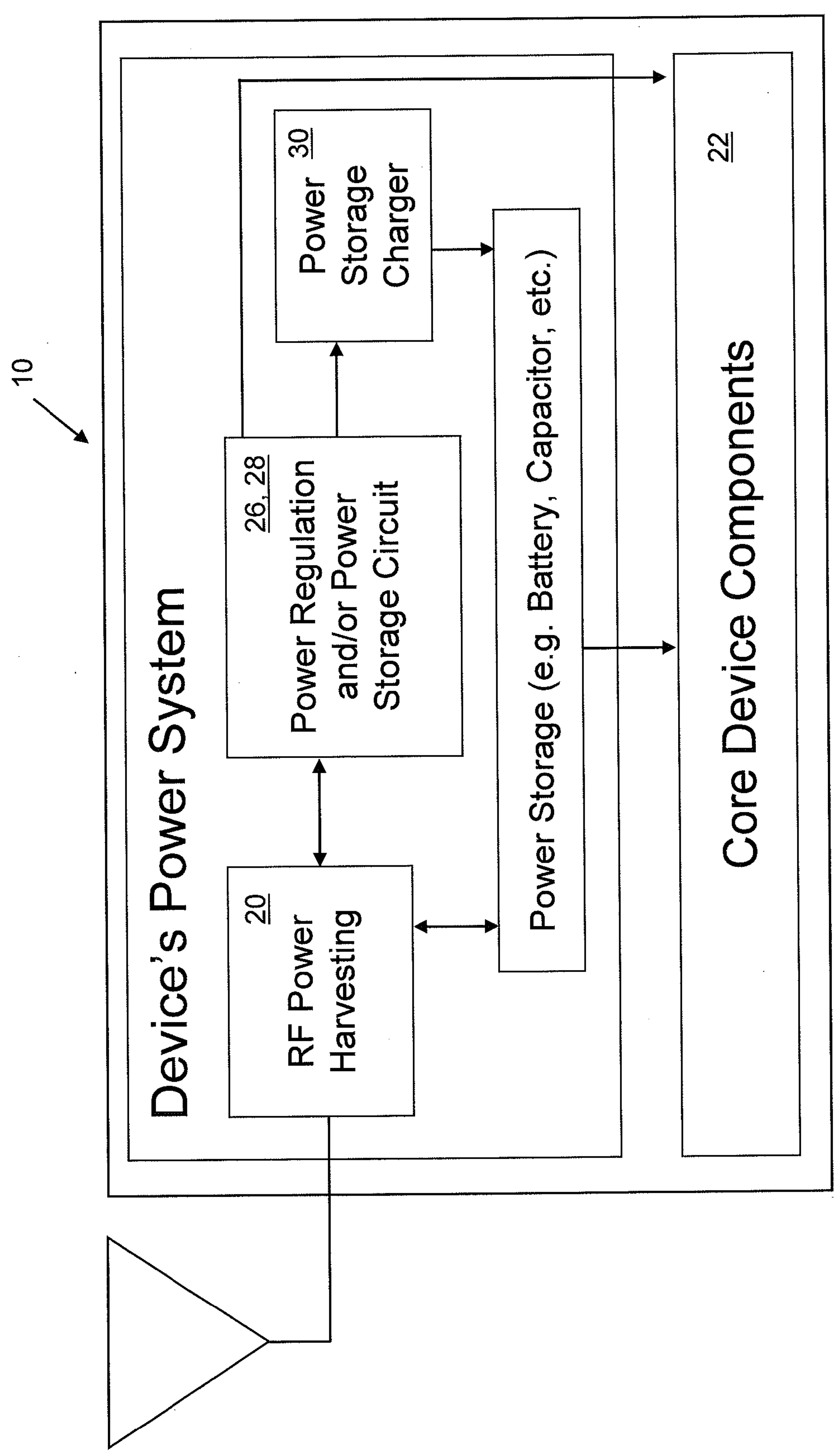
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 31

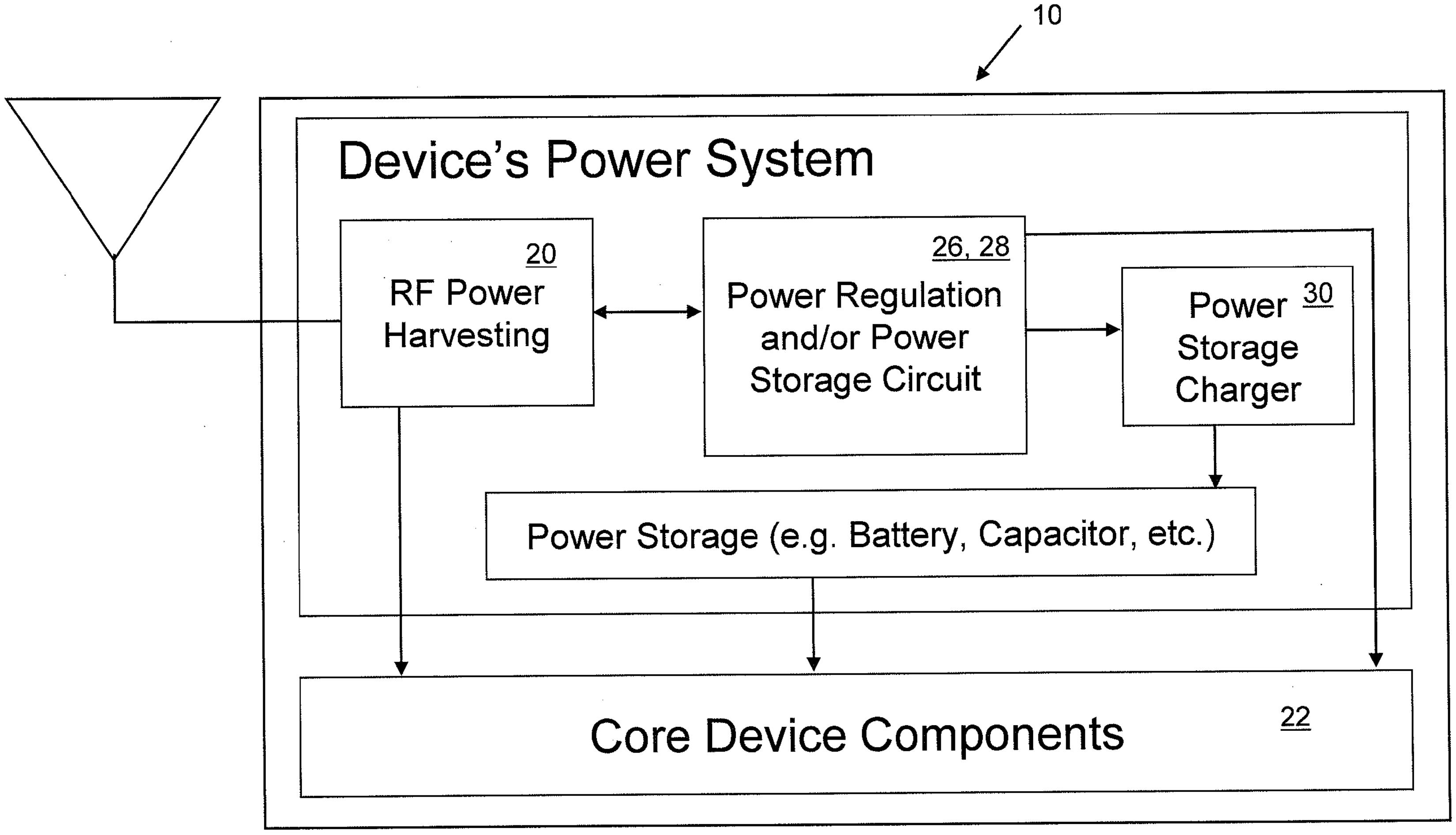
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 32

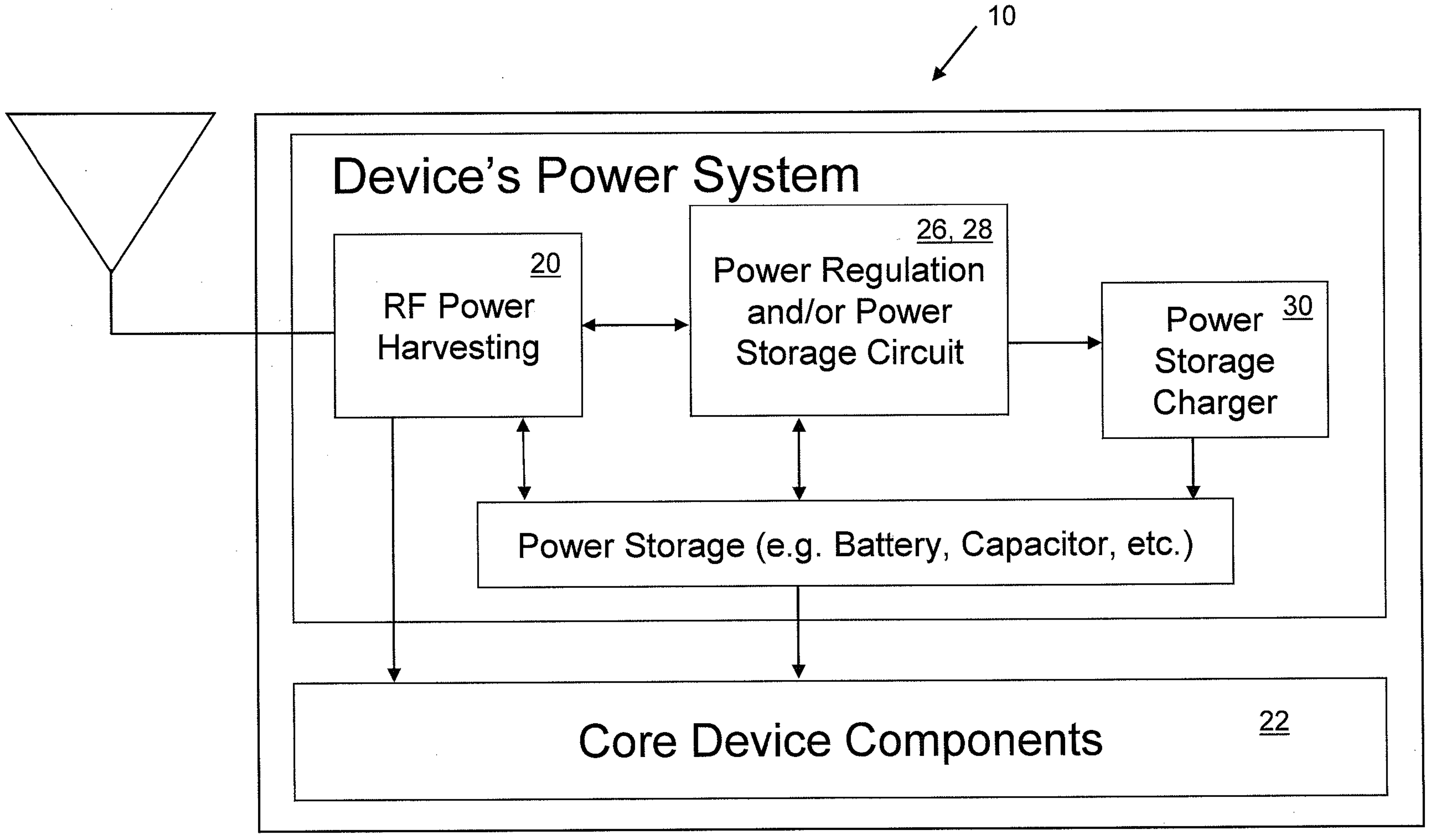
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 33

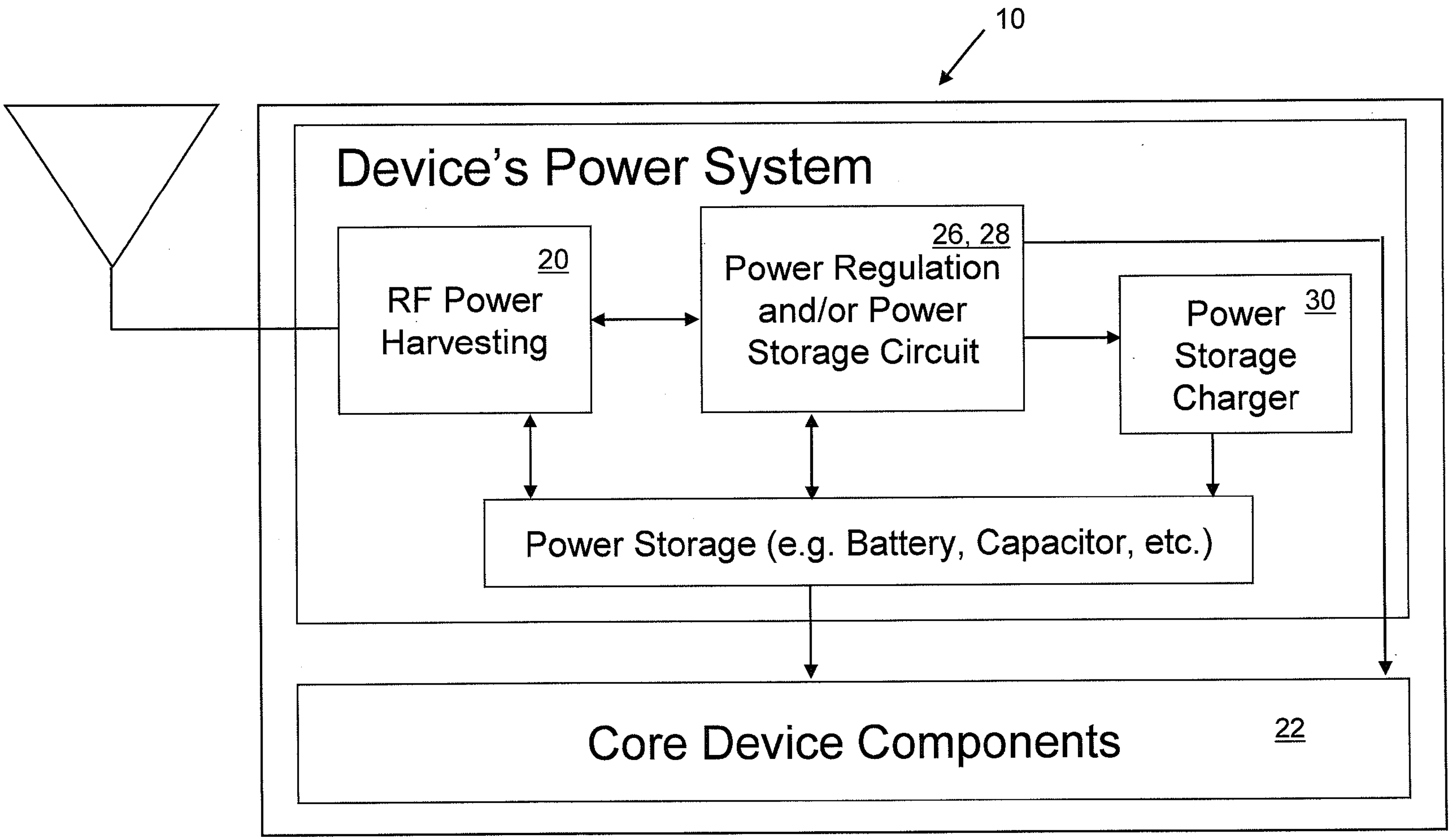
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 34

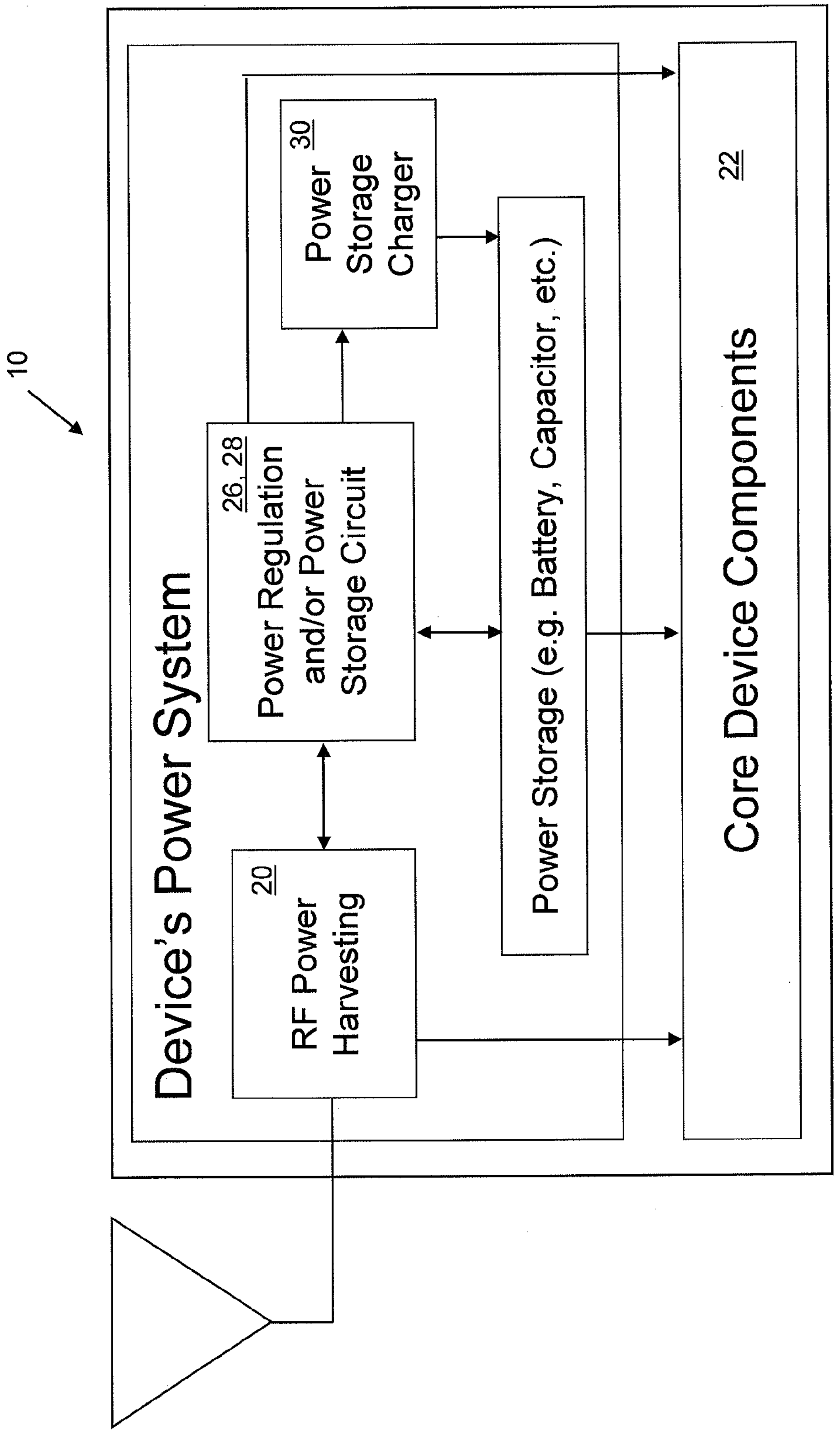
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 35

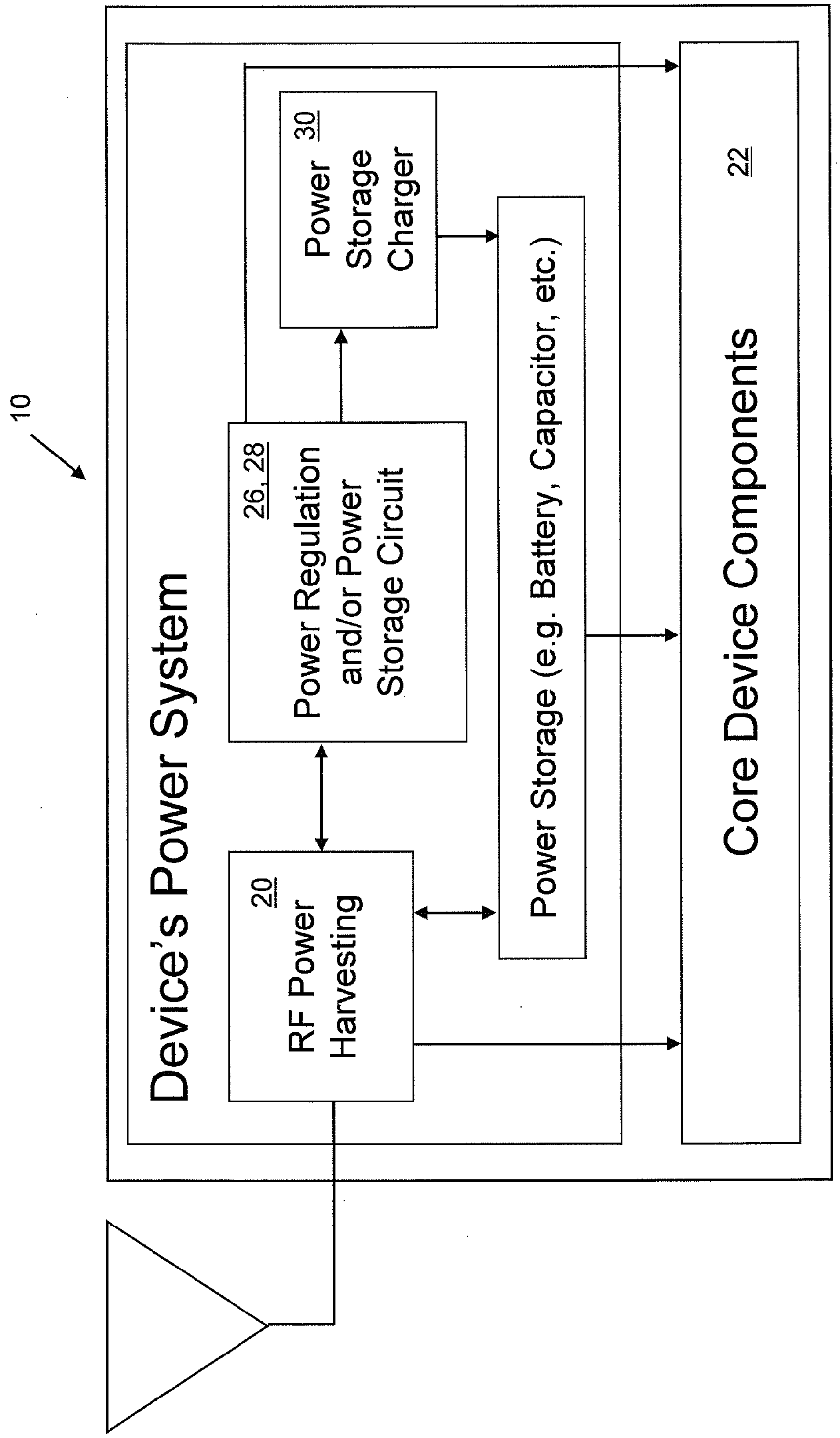
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 36

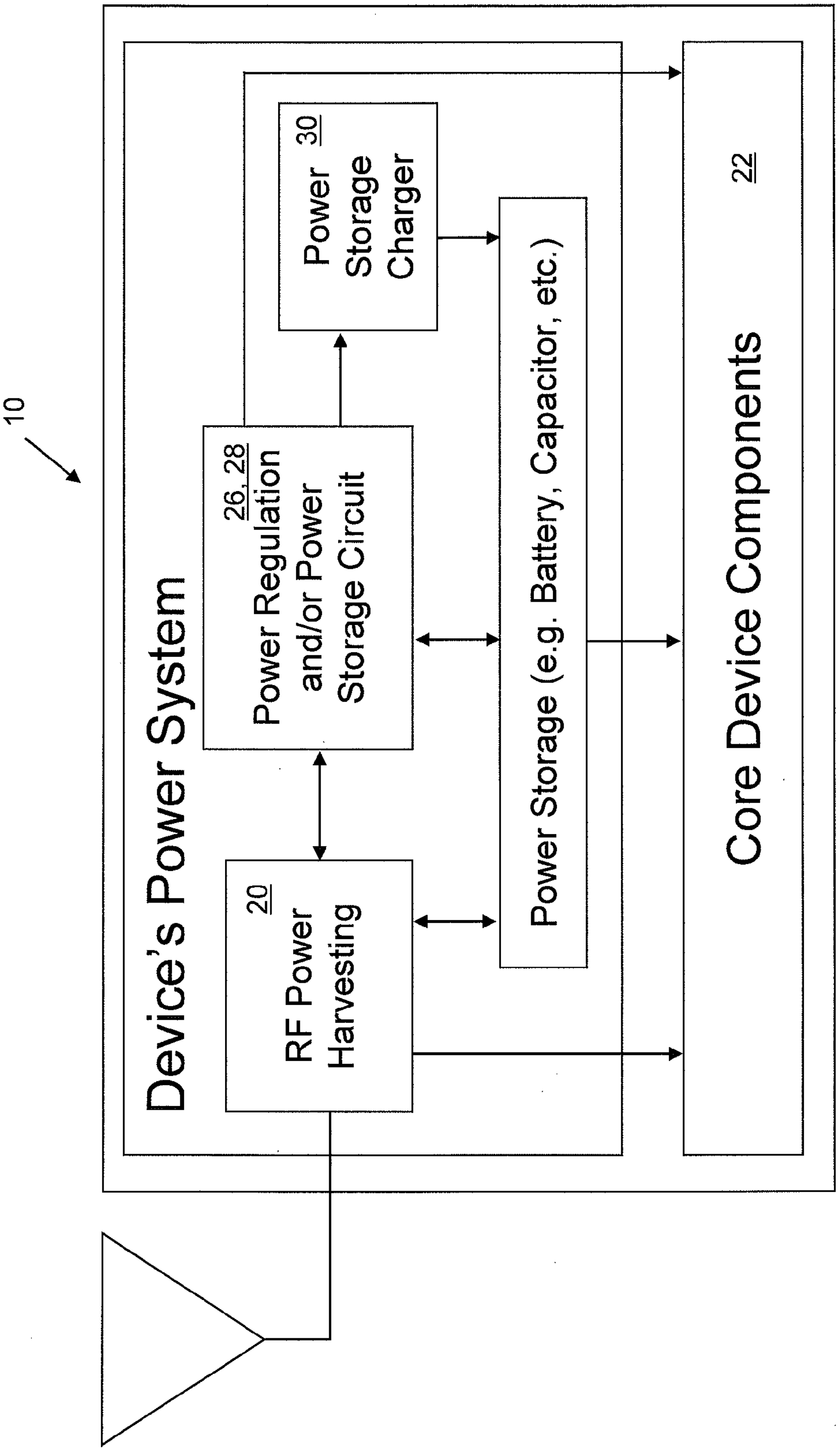
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 37

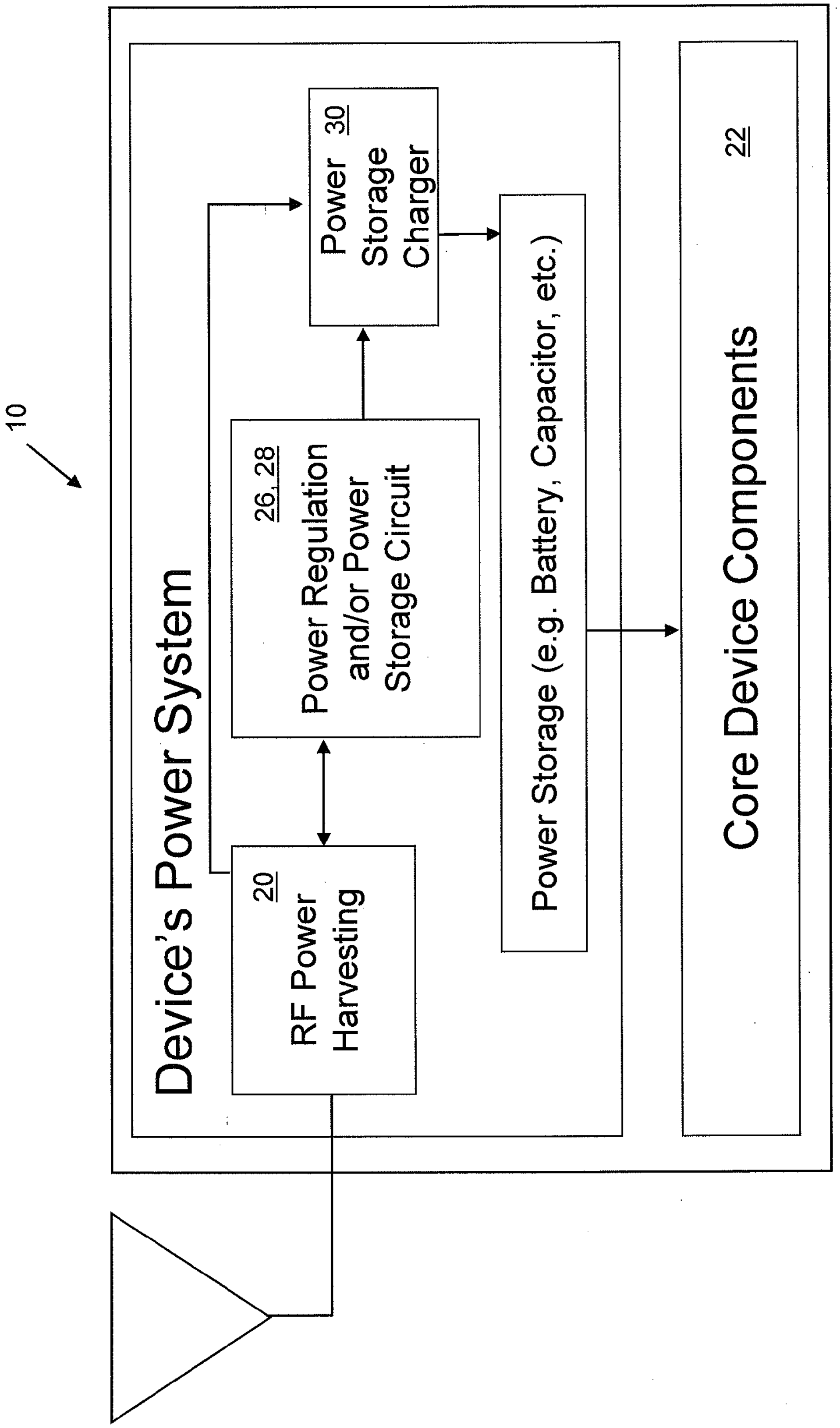
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 38

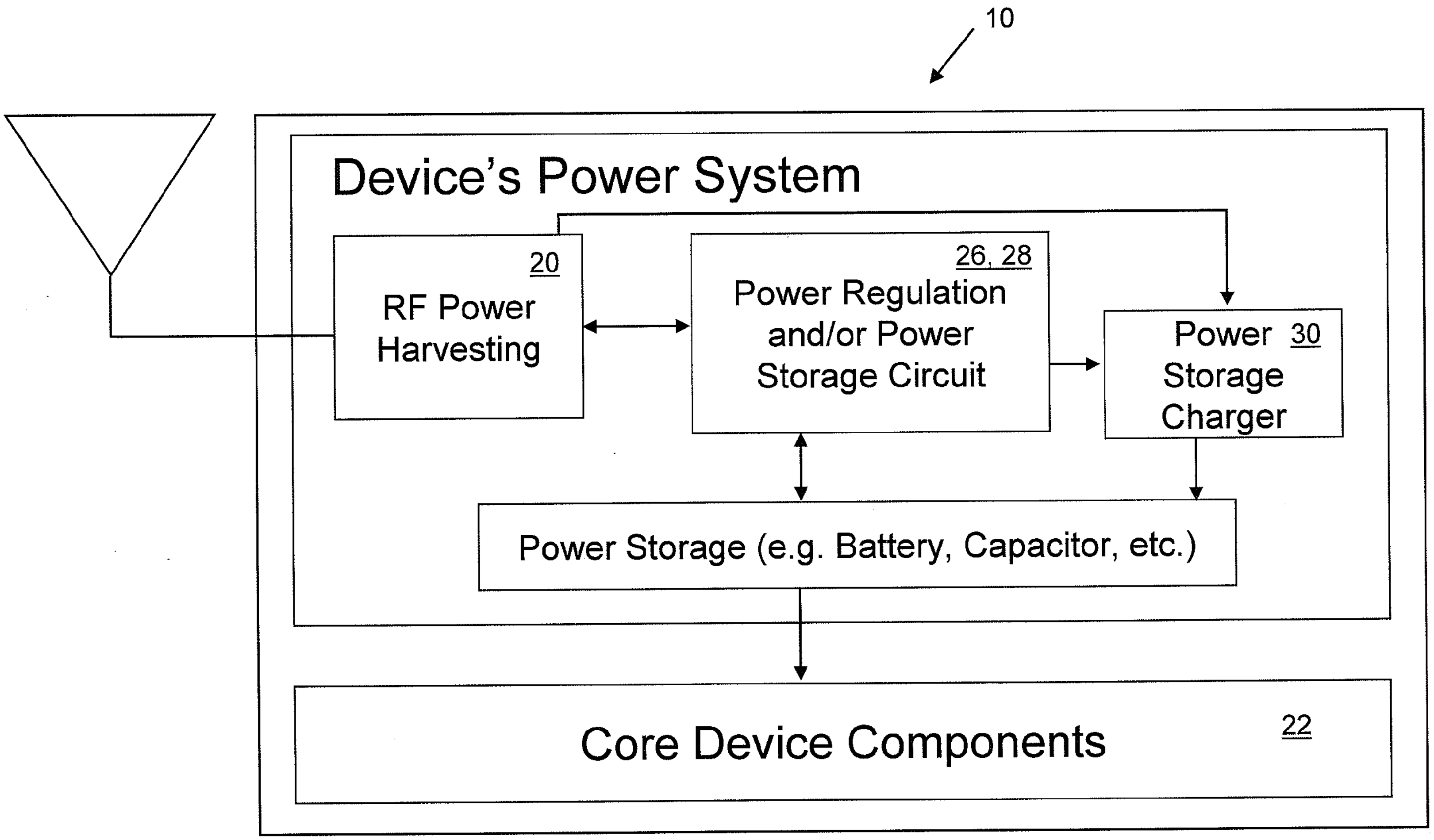
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage
Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 39

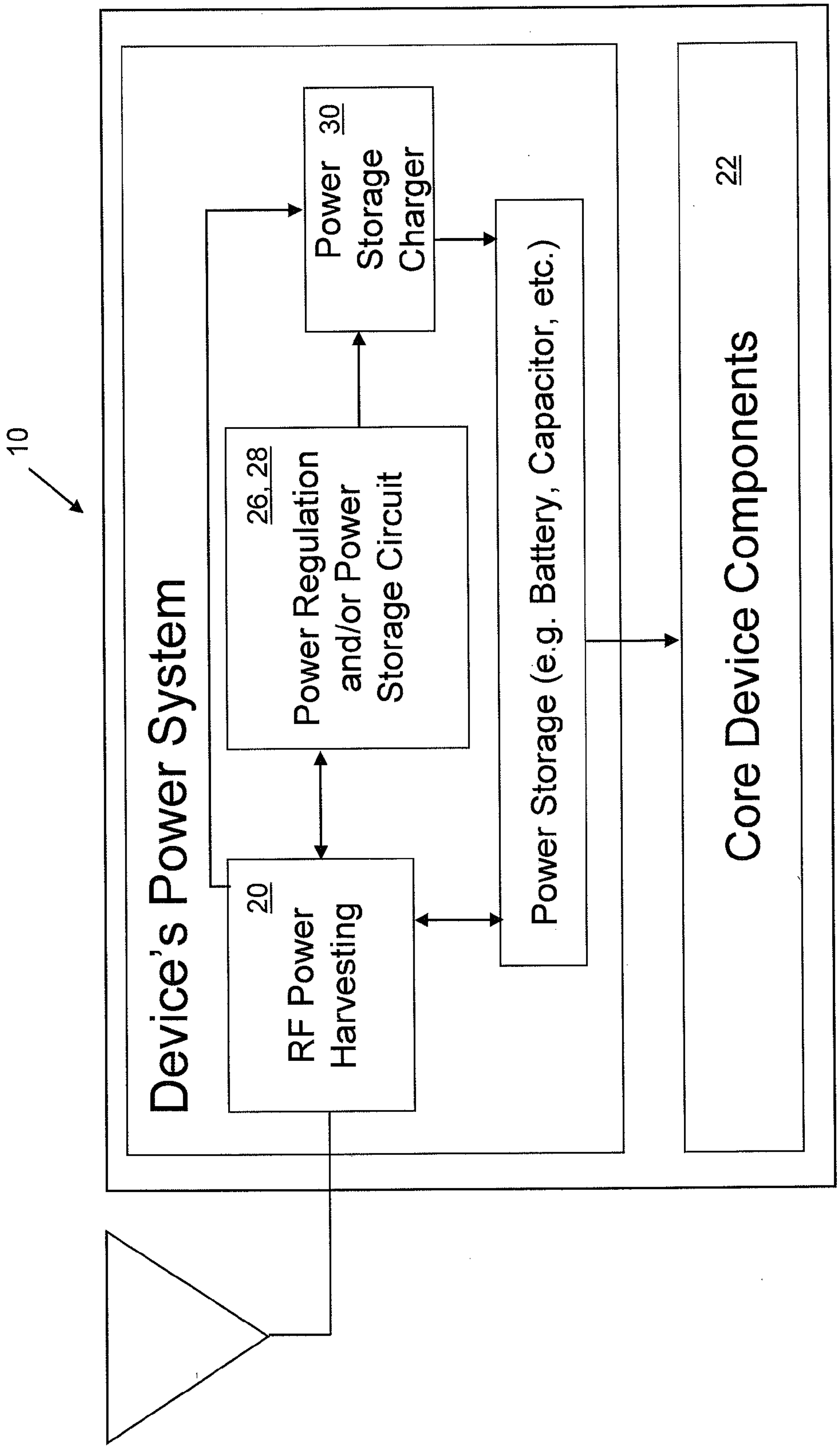
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 40

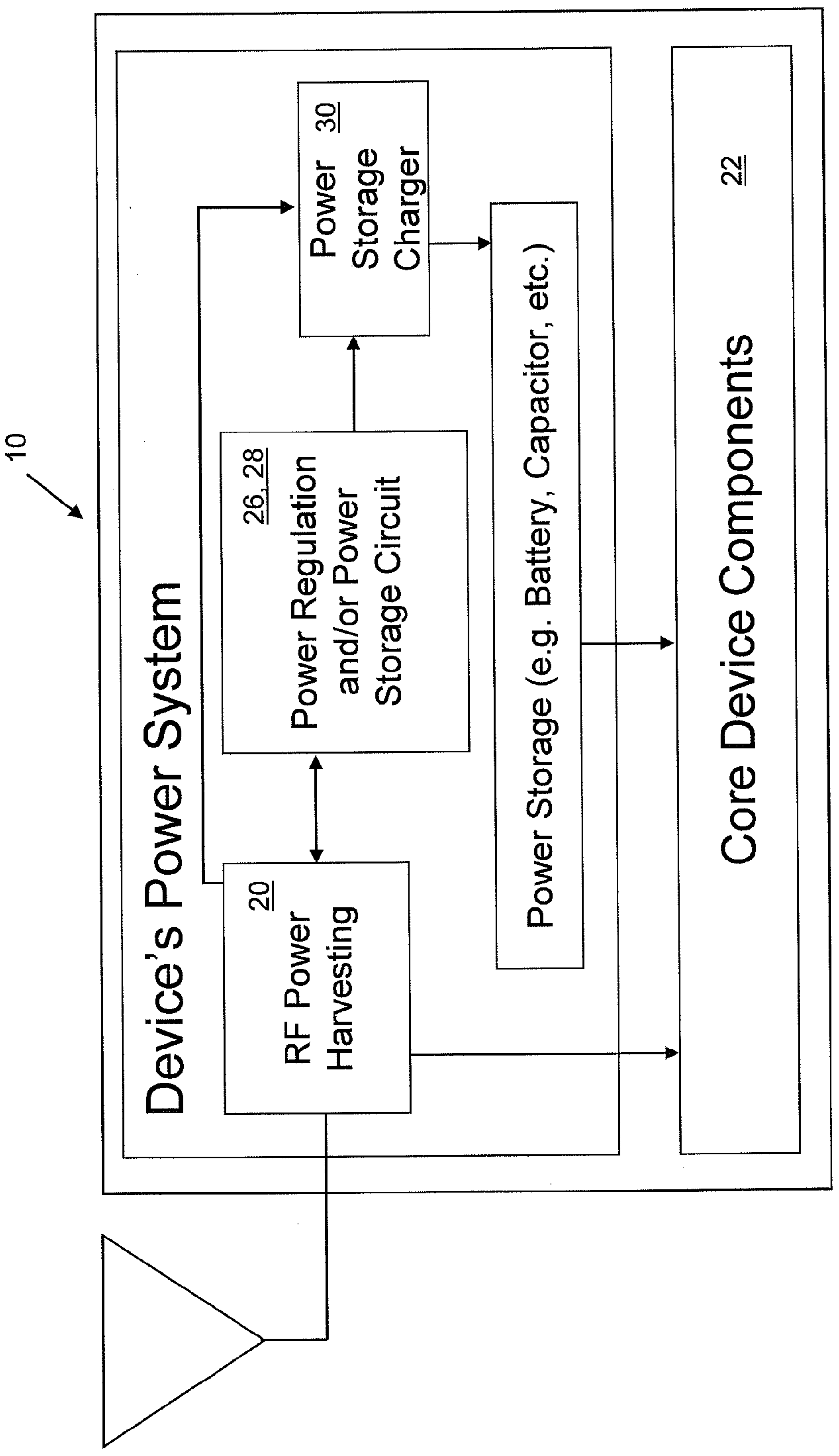
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 41

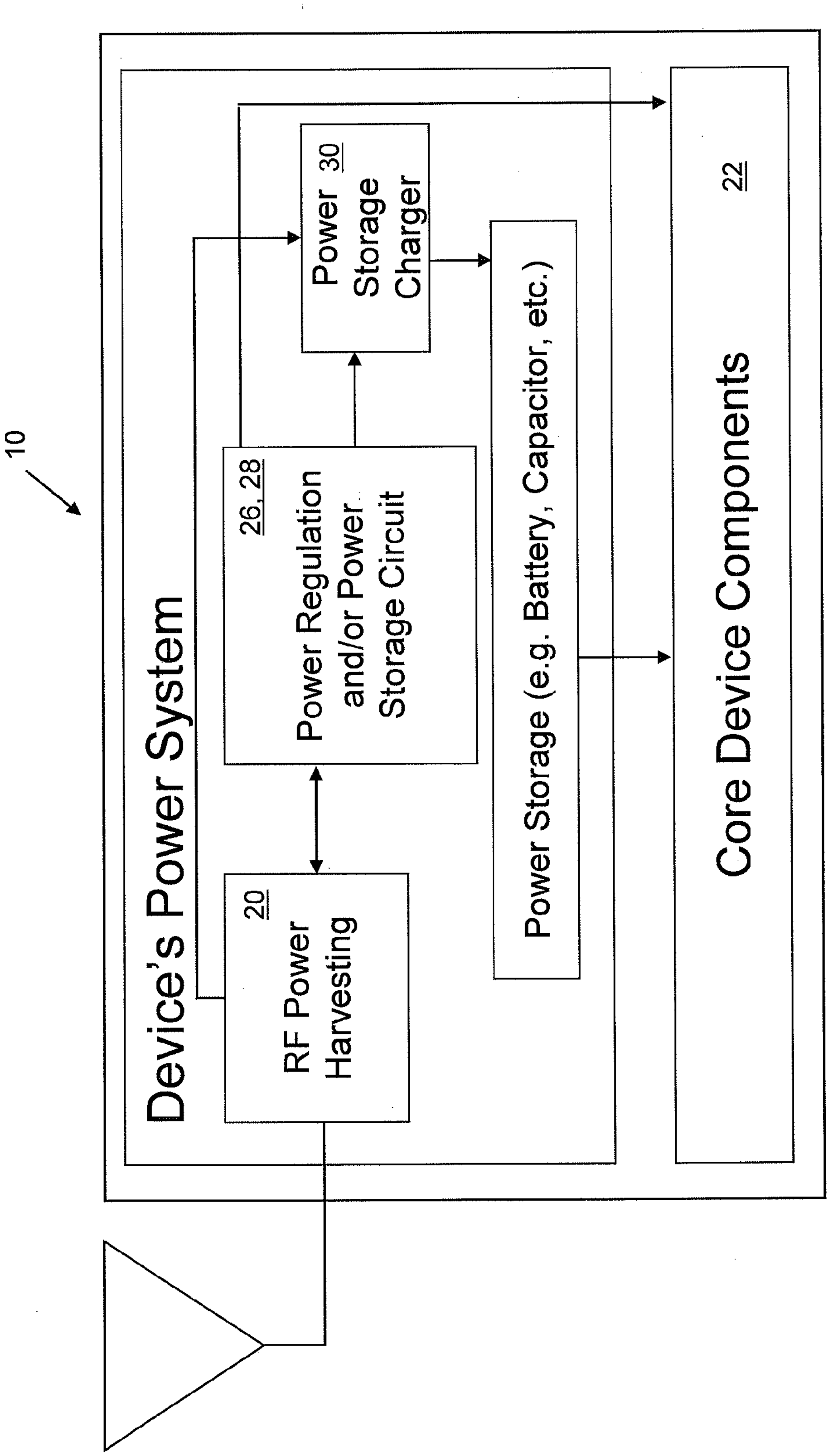
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 42

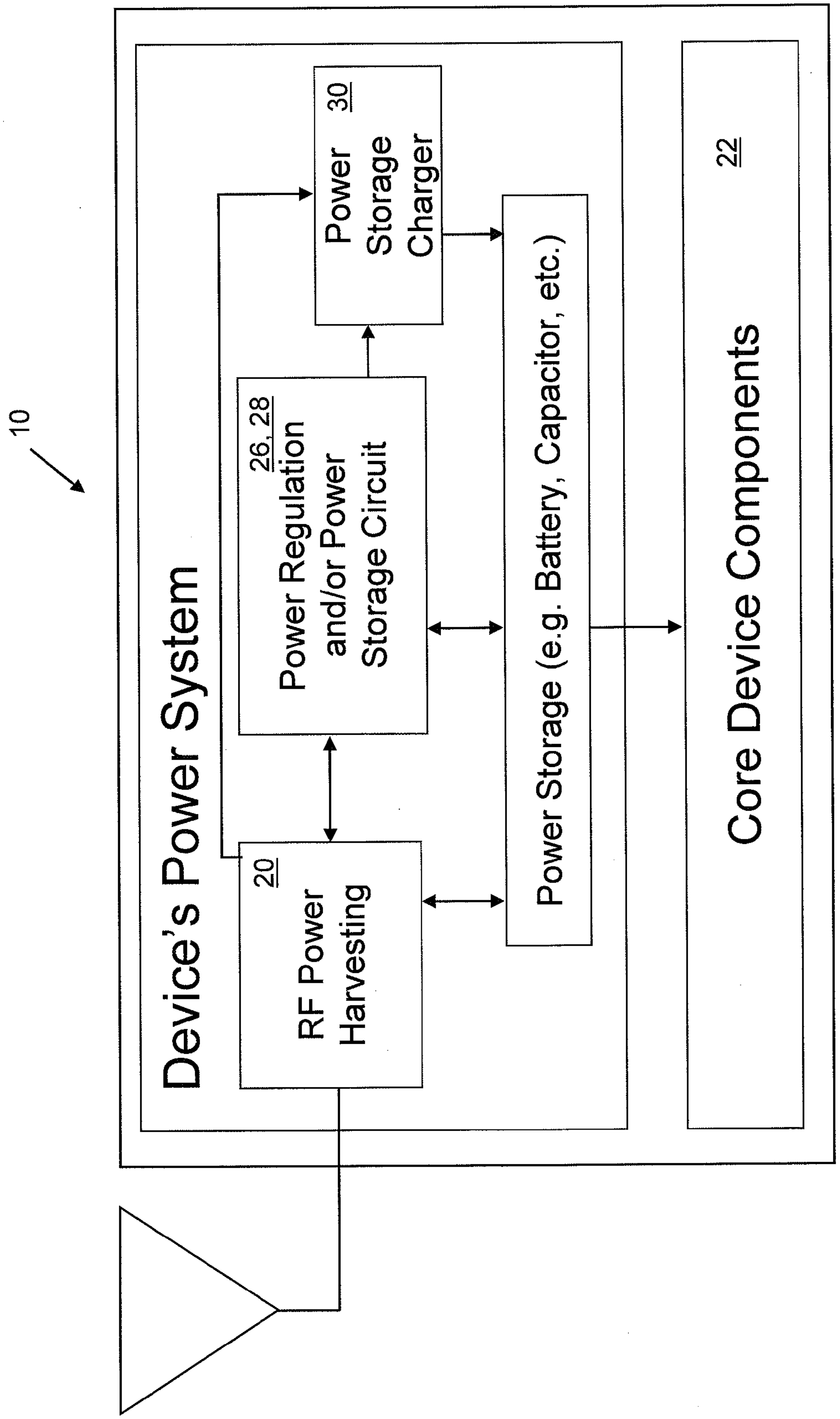
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 43

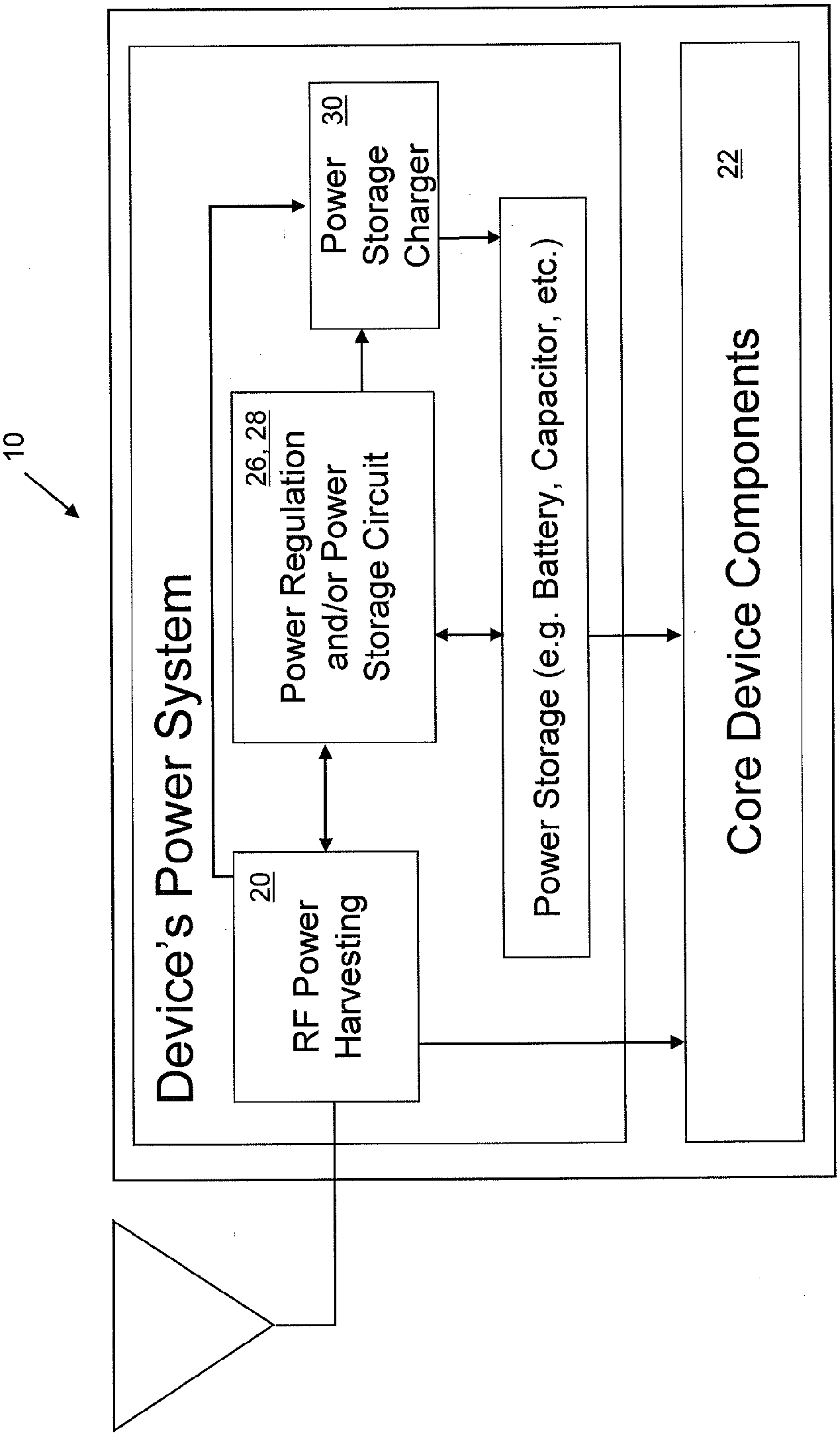
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 44

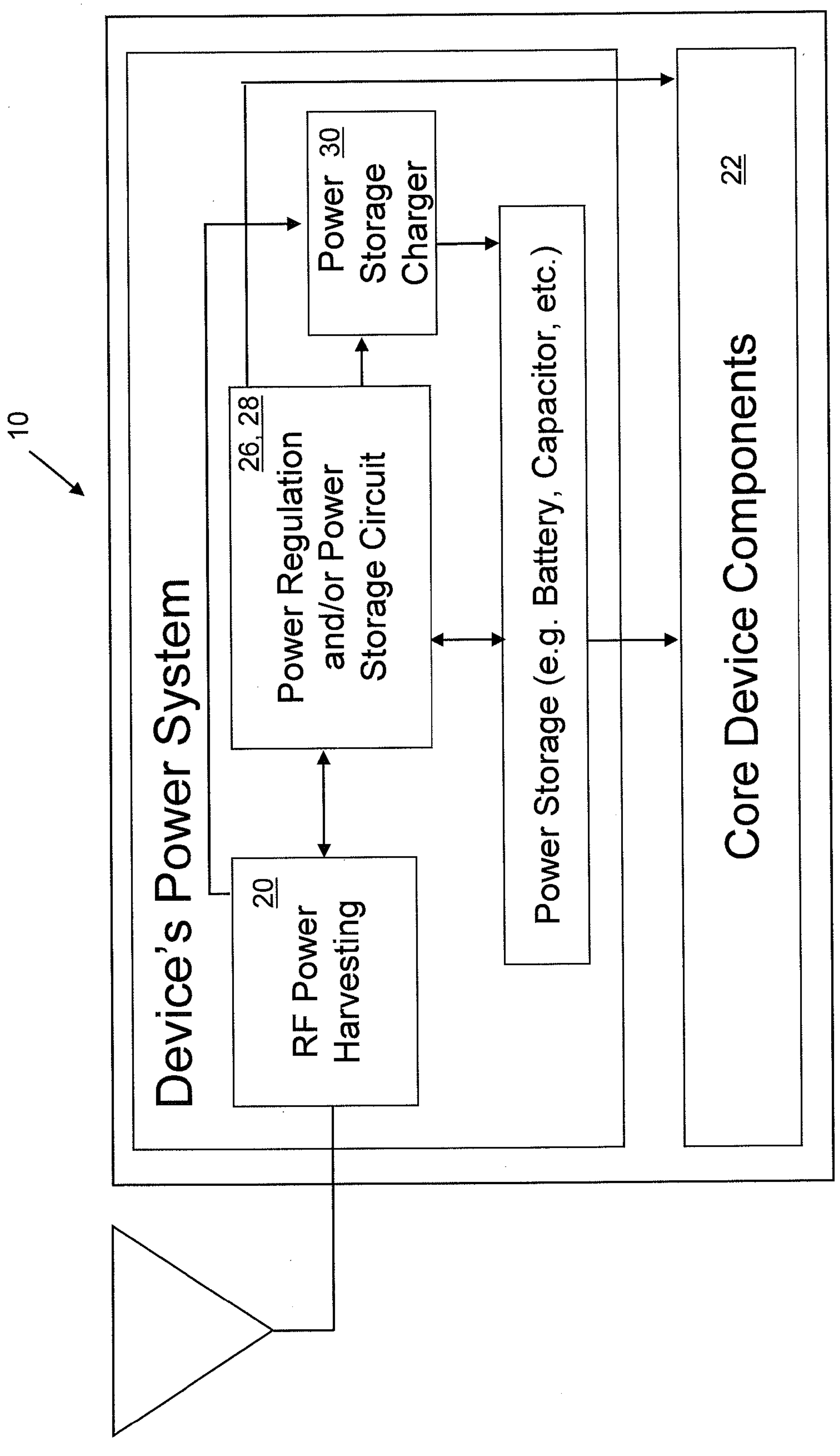
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 45

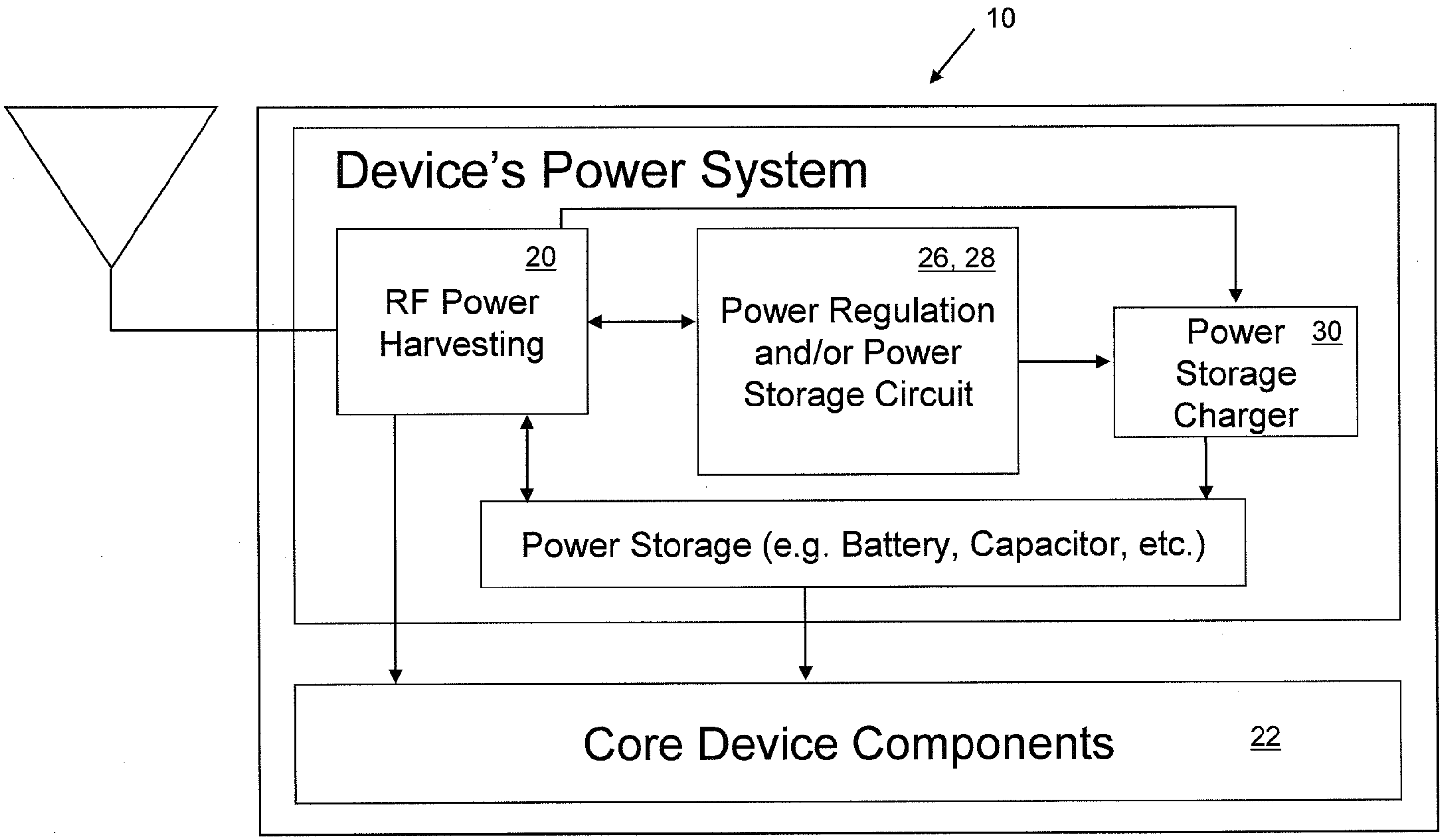
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 46

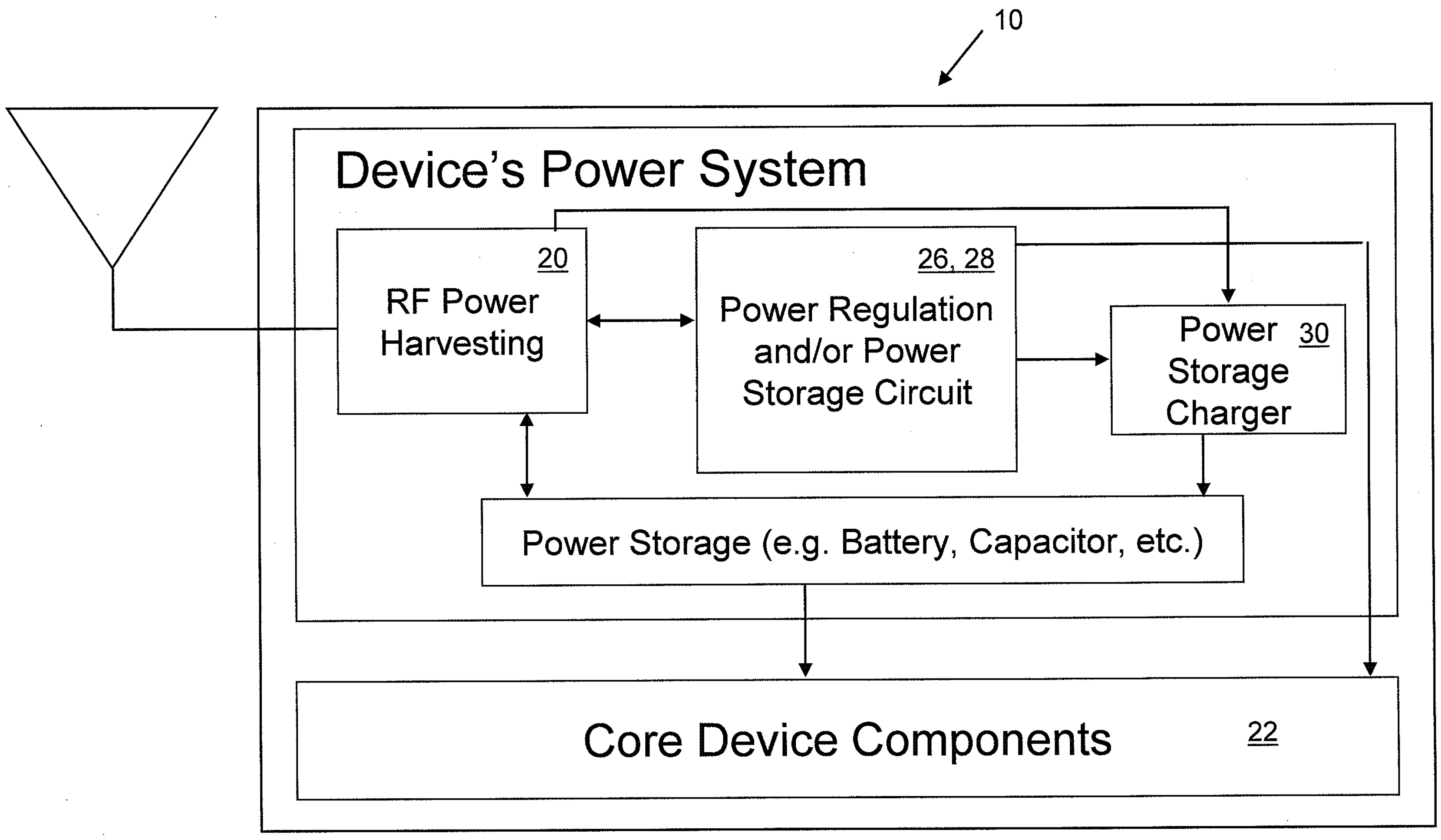
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 47

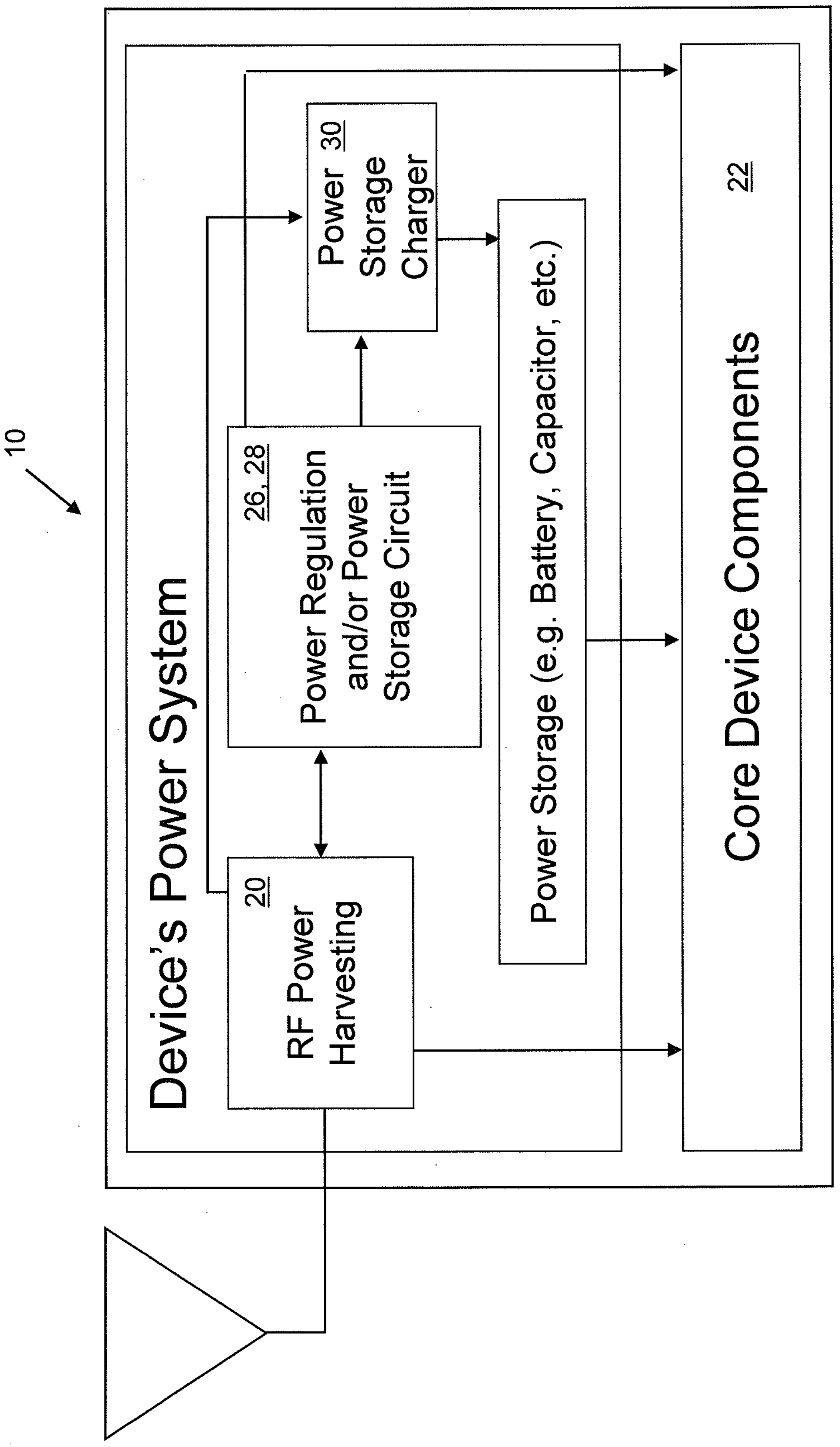
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage
Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 48

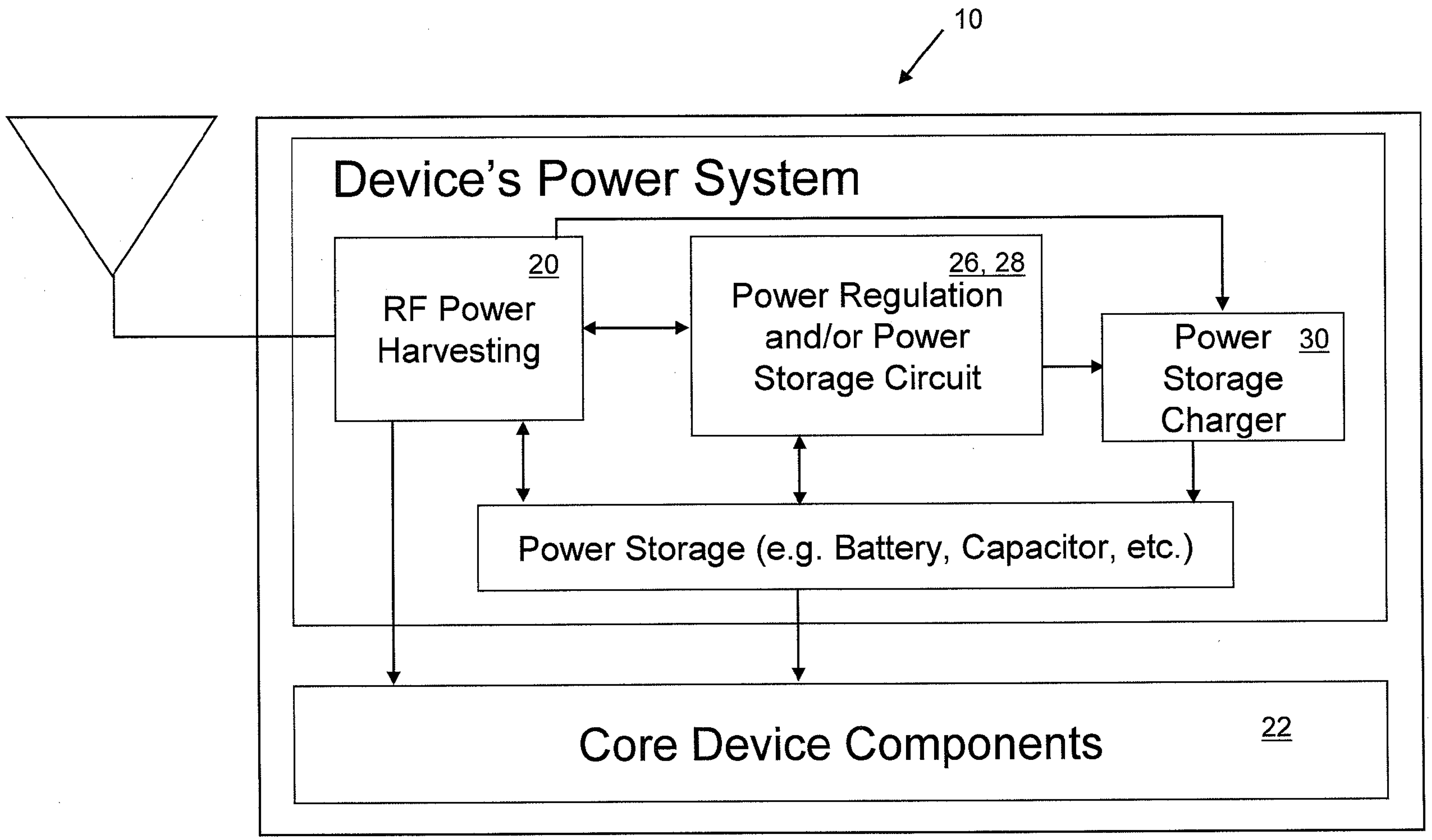
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 49

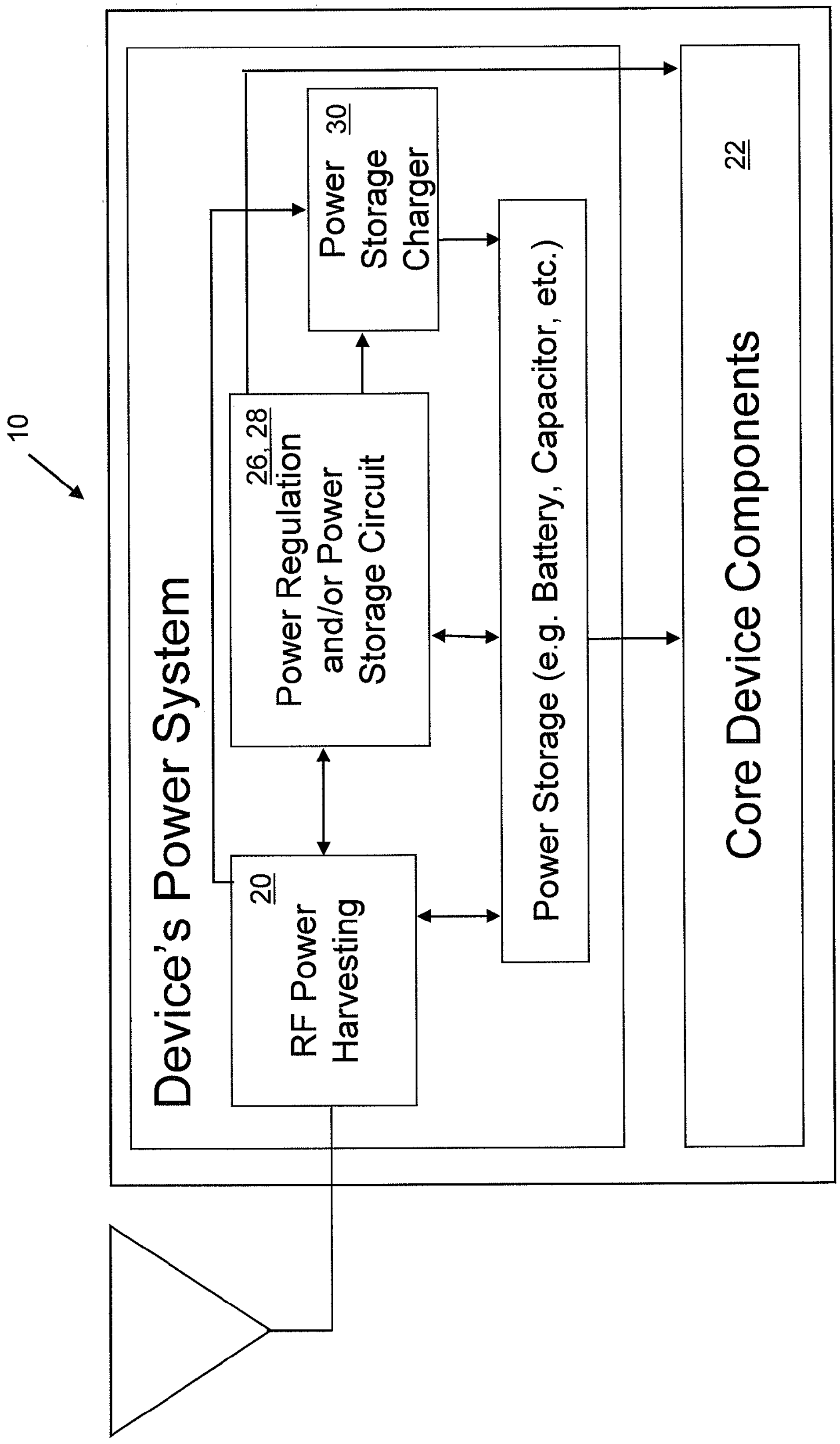
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 50

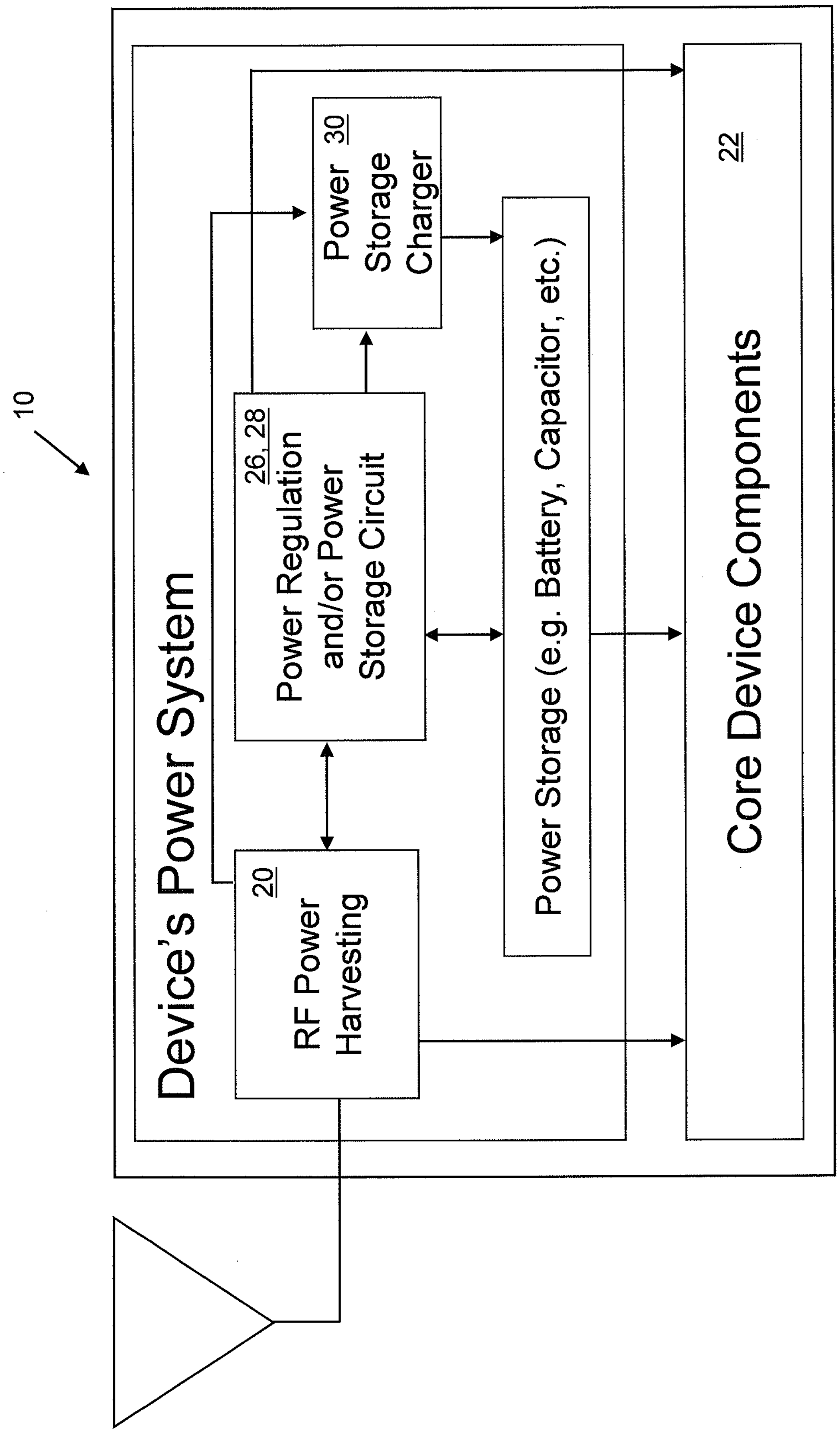
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 51

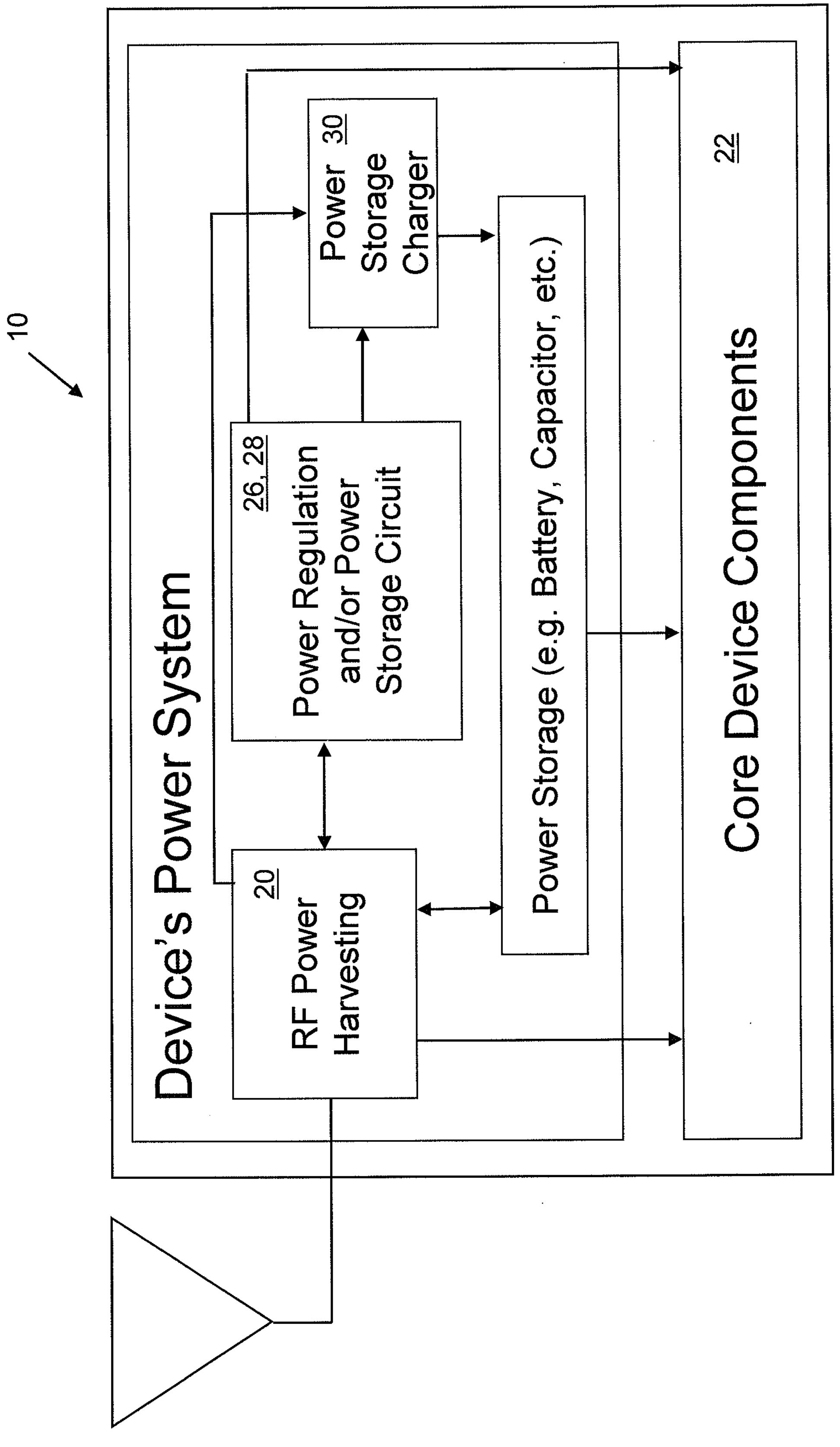
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 52

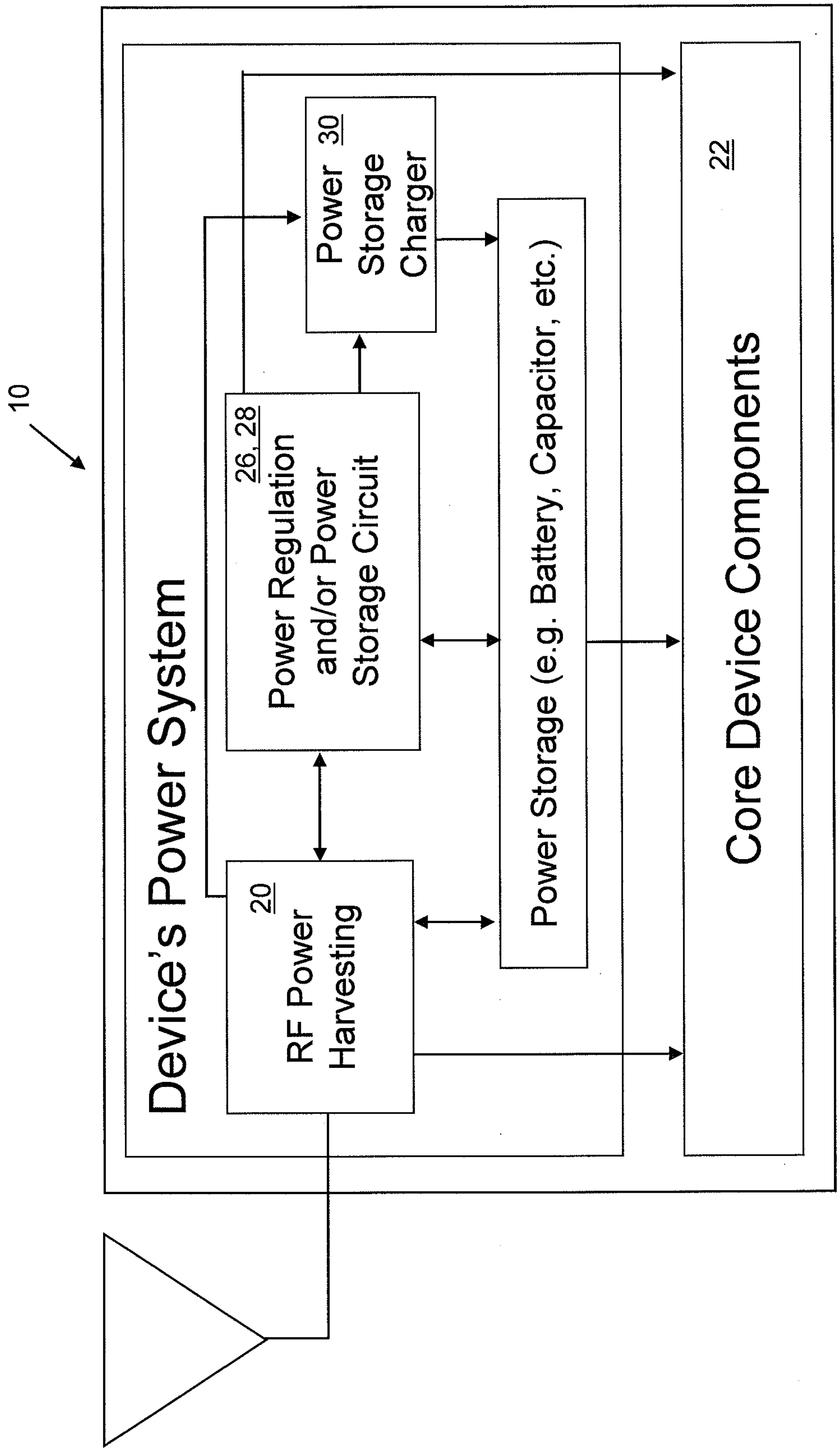
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 53

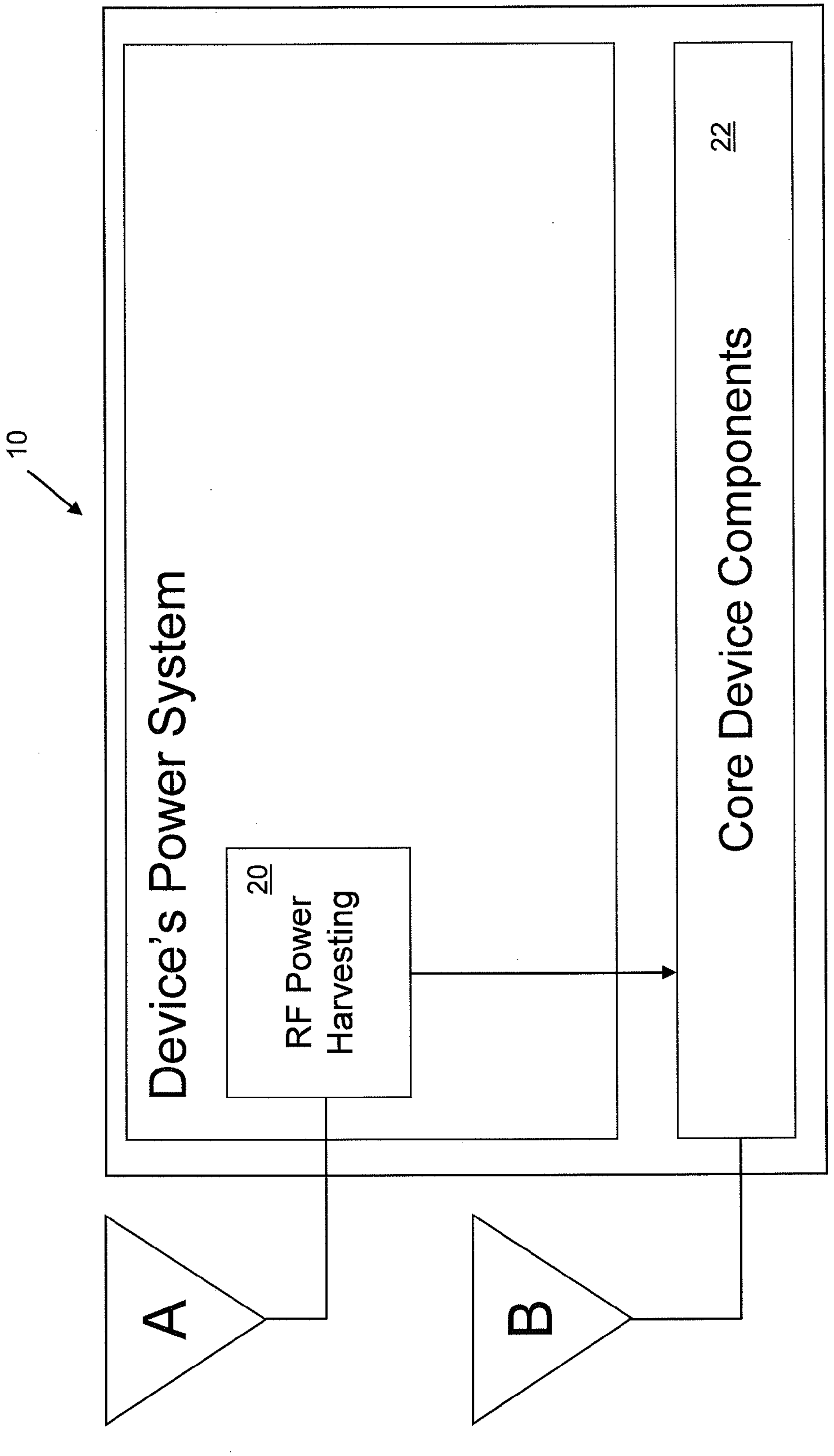
10
Device's Power System
20
RF Power Harvesting
22
Core Device Components
A
B

Fig. 54

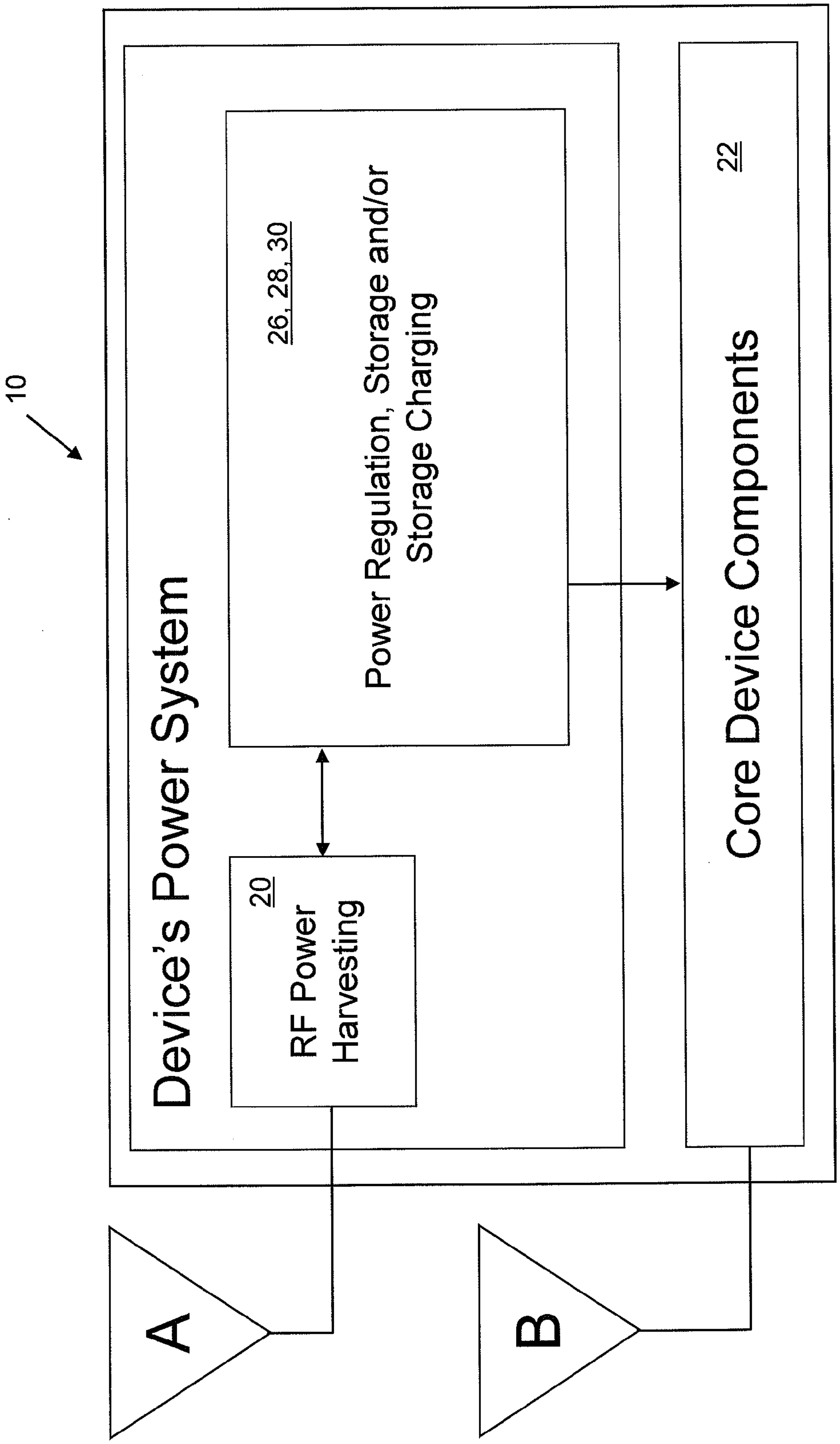
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
A
B

Fig. 55

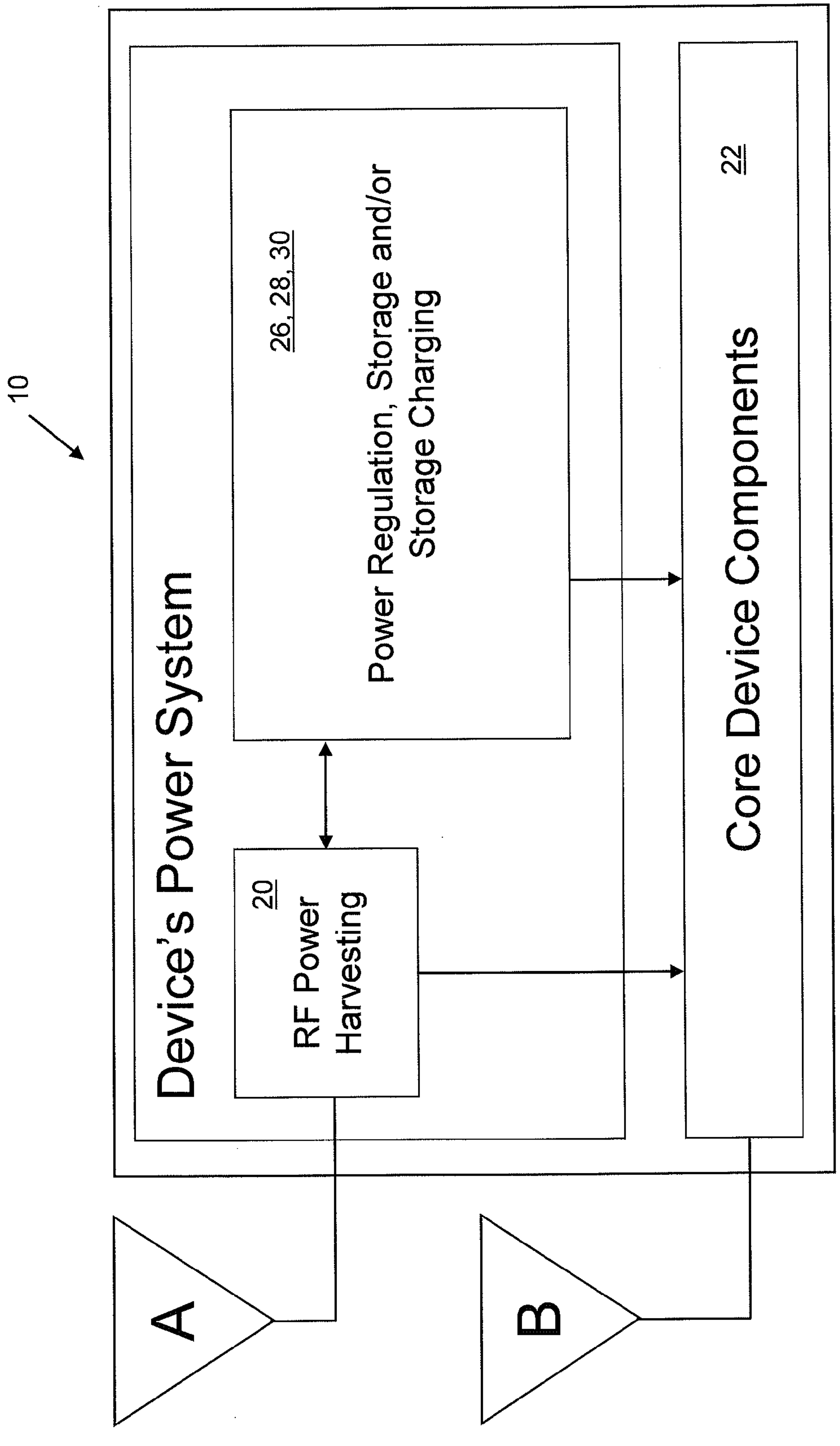
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
A
B

Fig. 56

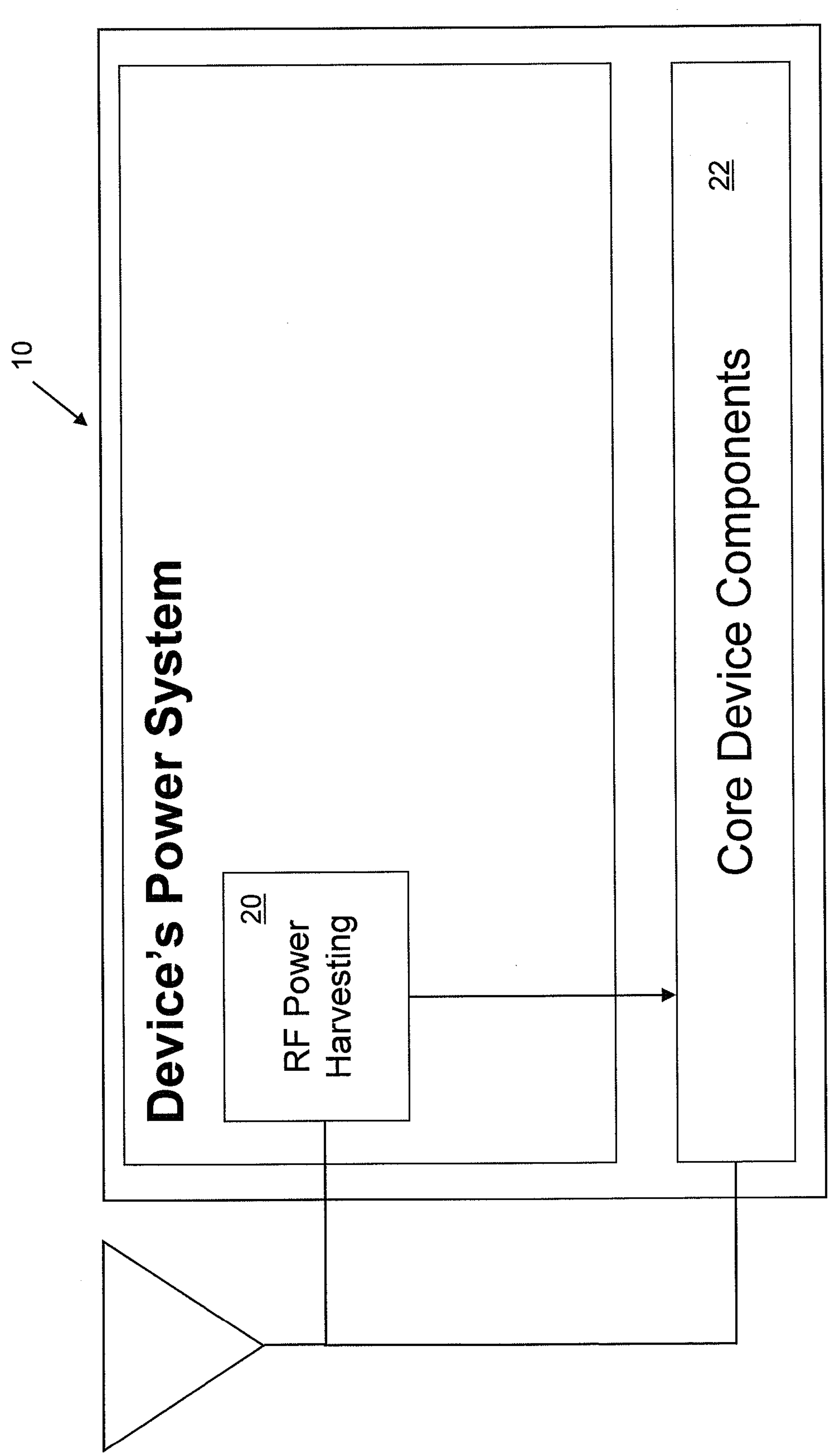
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22

Fig. 57

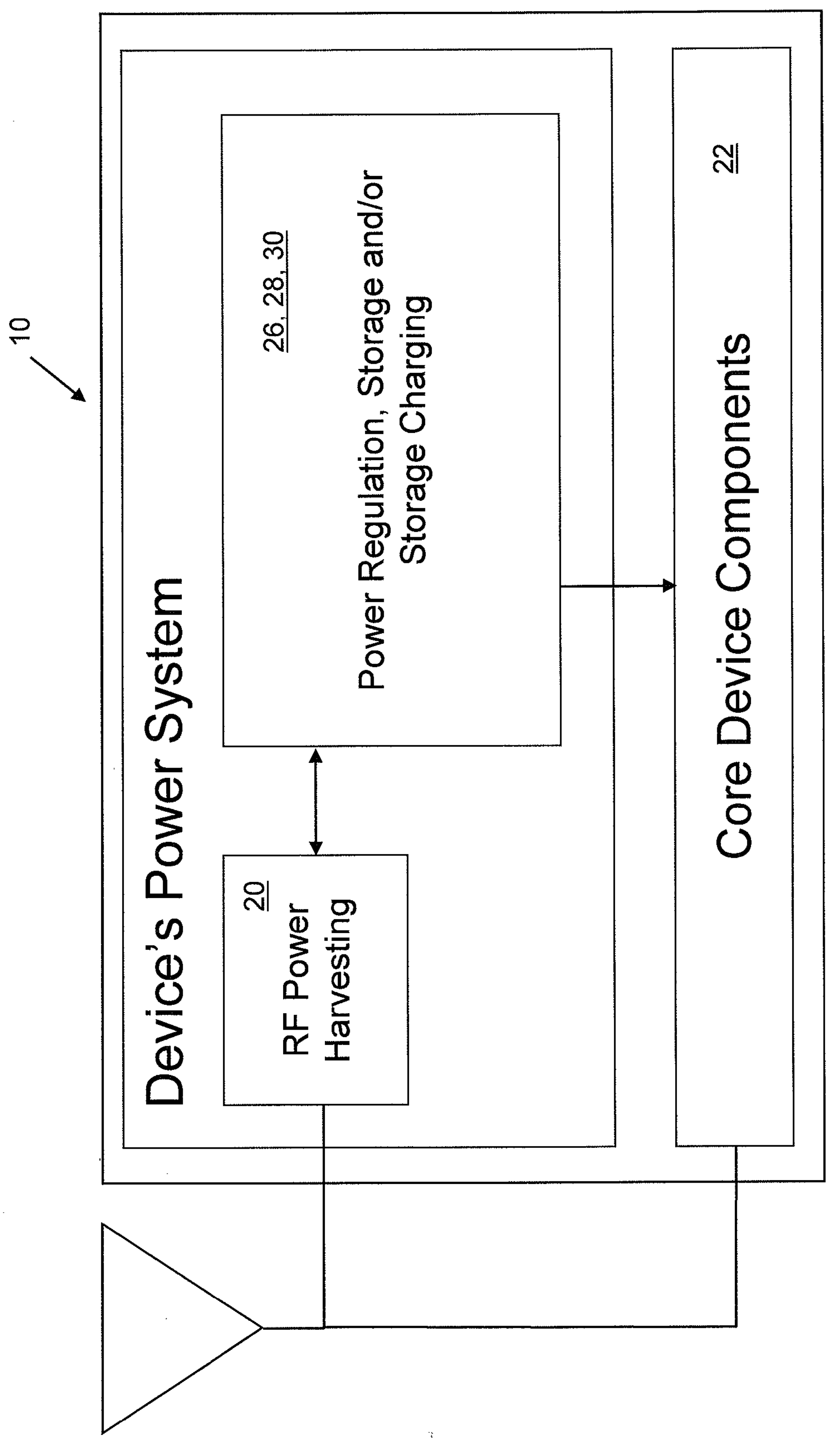
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components

Fig. 58

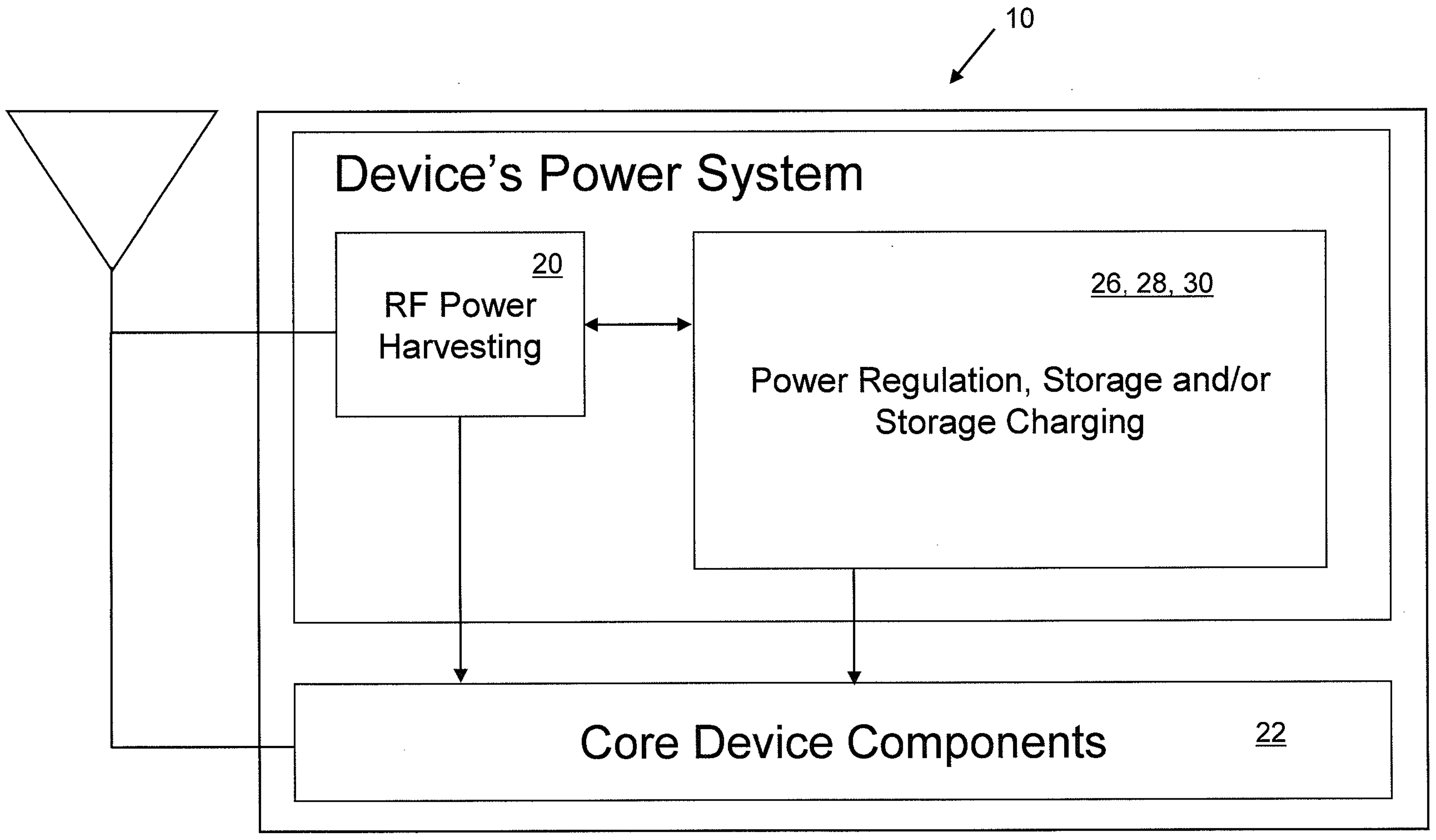
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22

Fig. 59

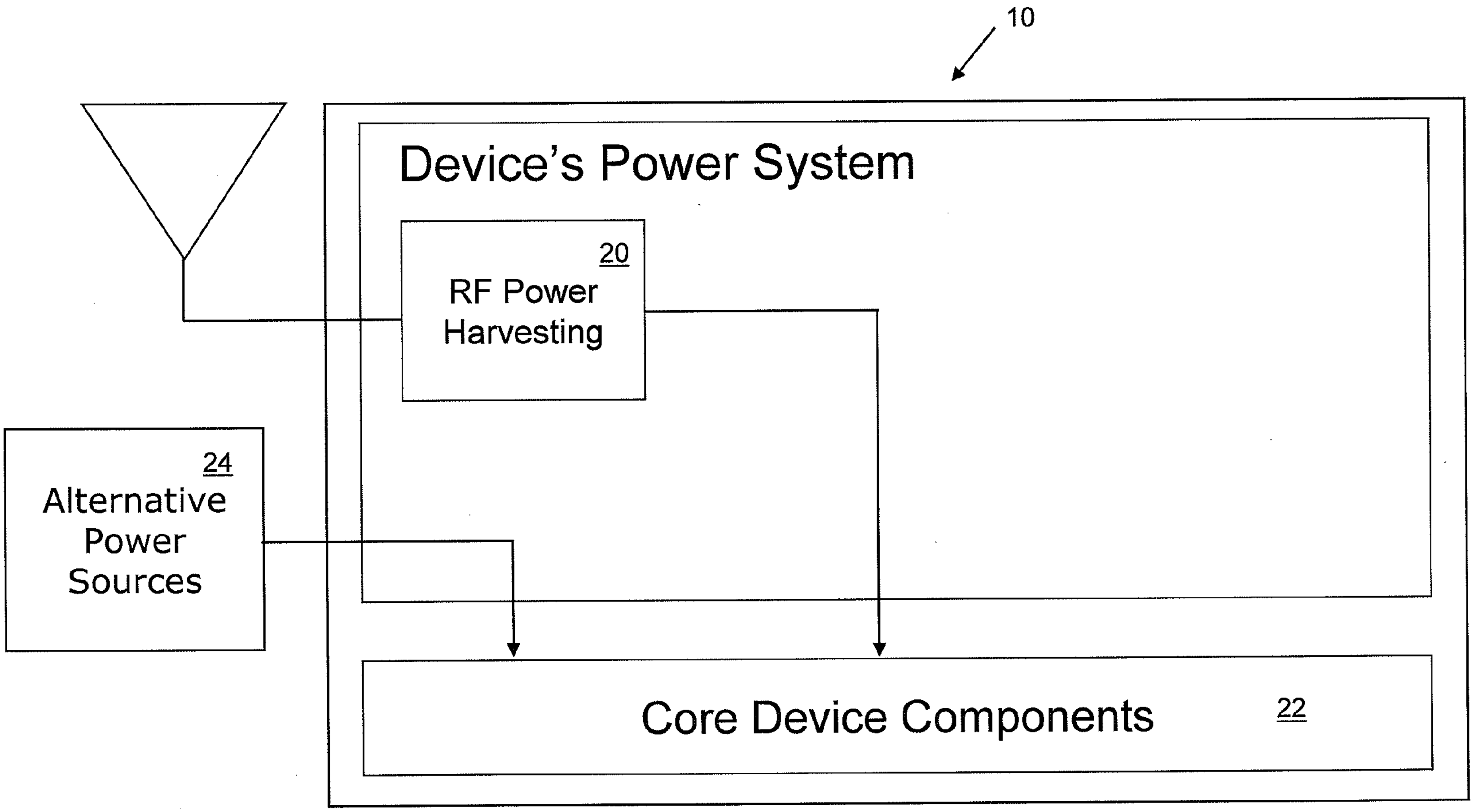
10
Device's Power System
20
RF Power Harvesting
24
Alternative Power Sources
Core Device Components
22

Fig. 60

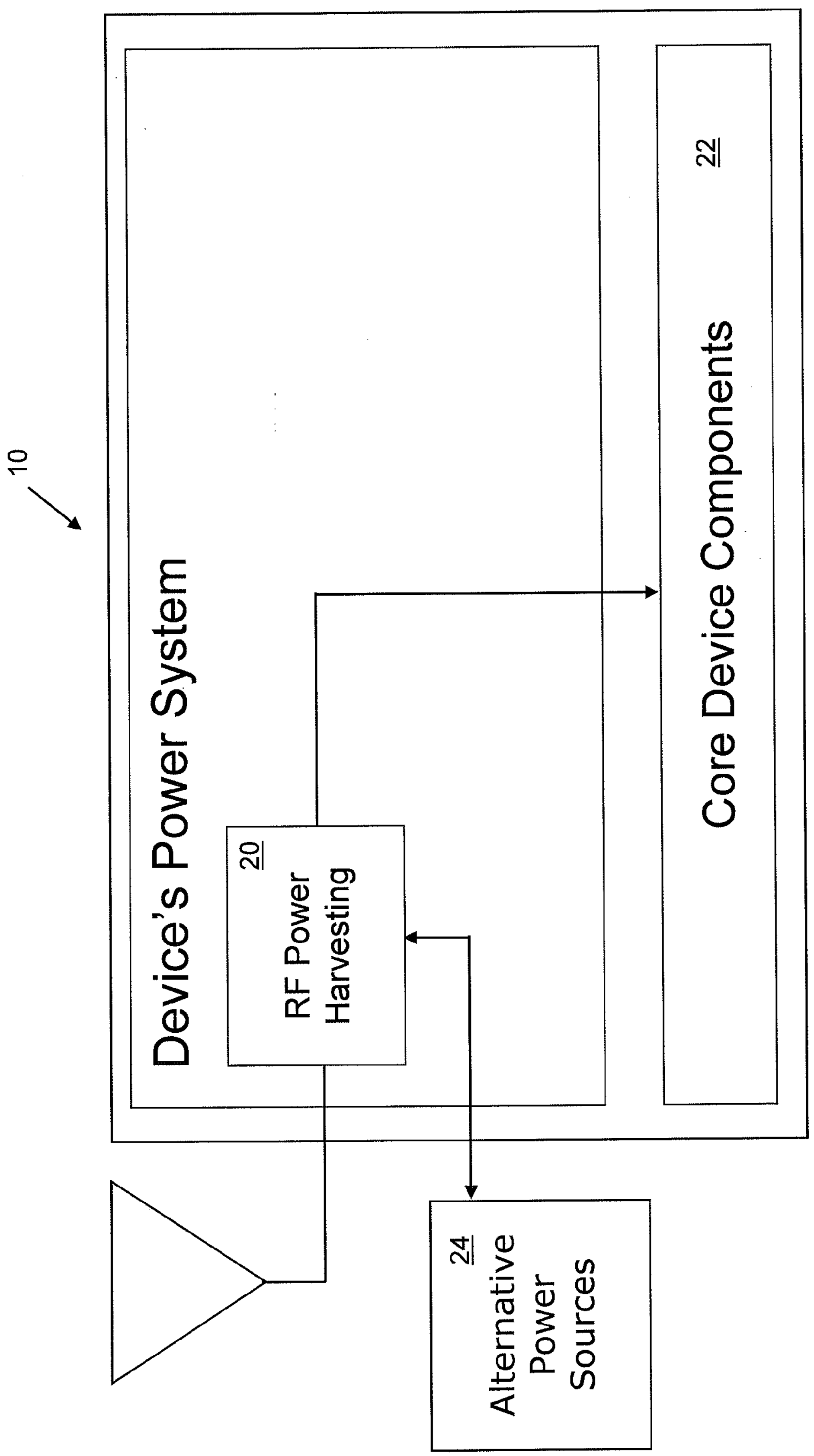
10
Device's Power System
20
RF Power Harvesting
22
Core Device Components
24
Alternative Power Sources

Fig. 61

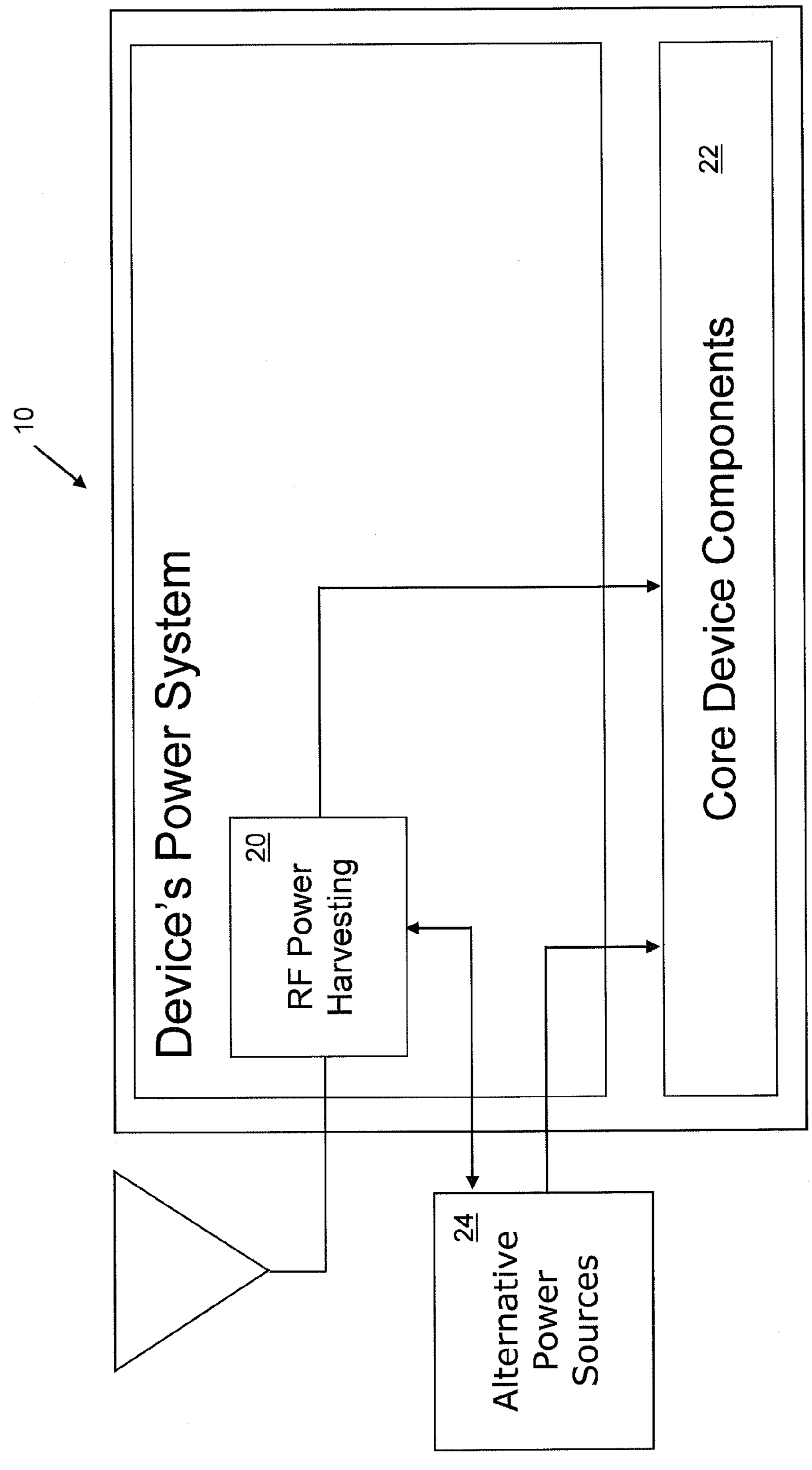
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 62

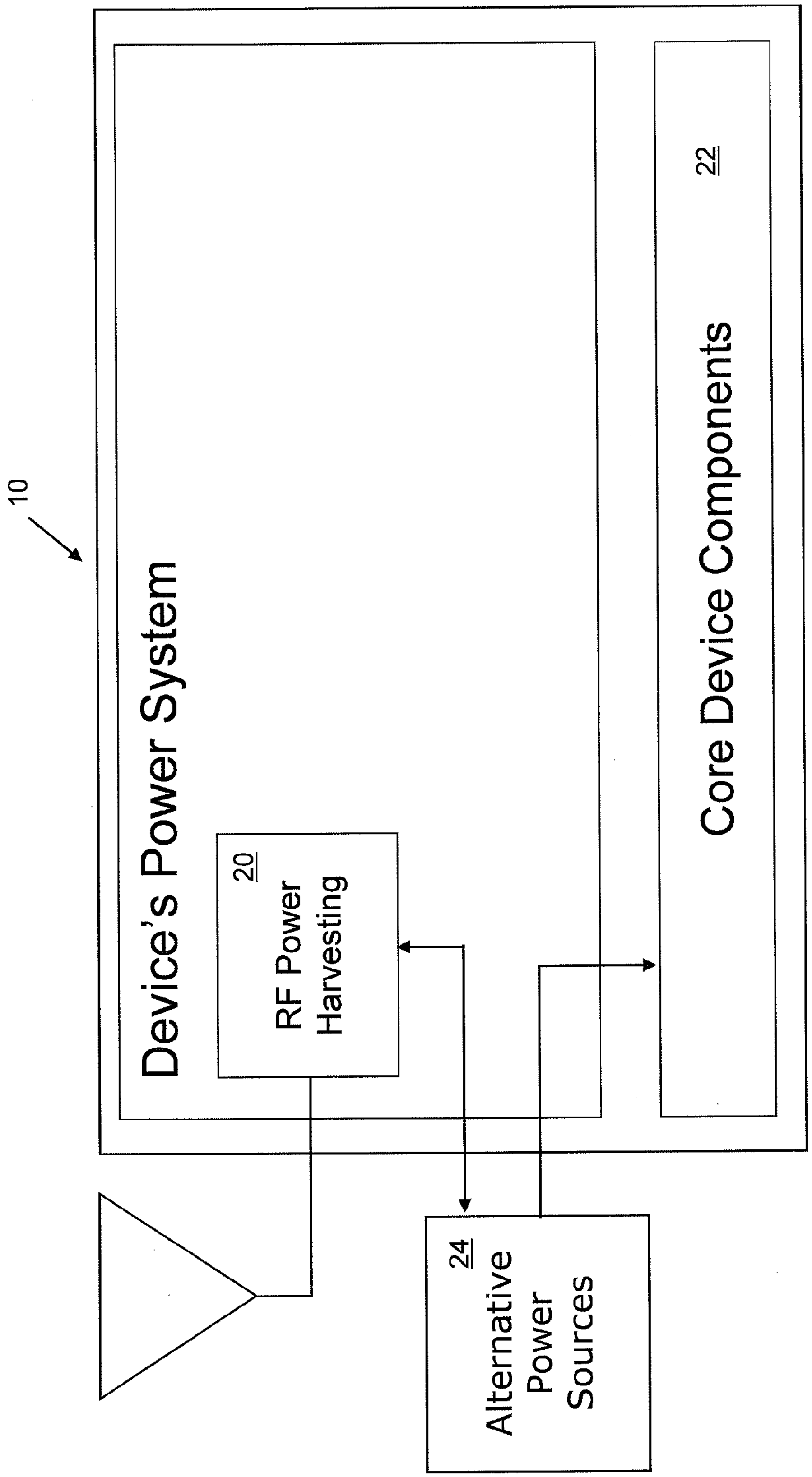
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 63

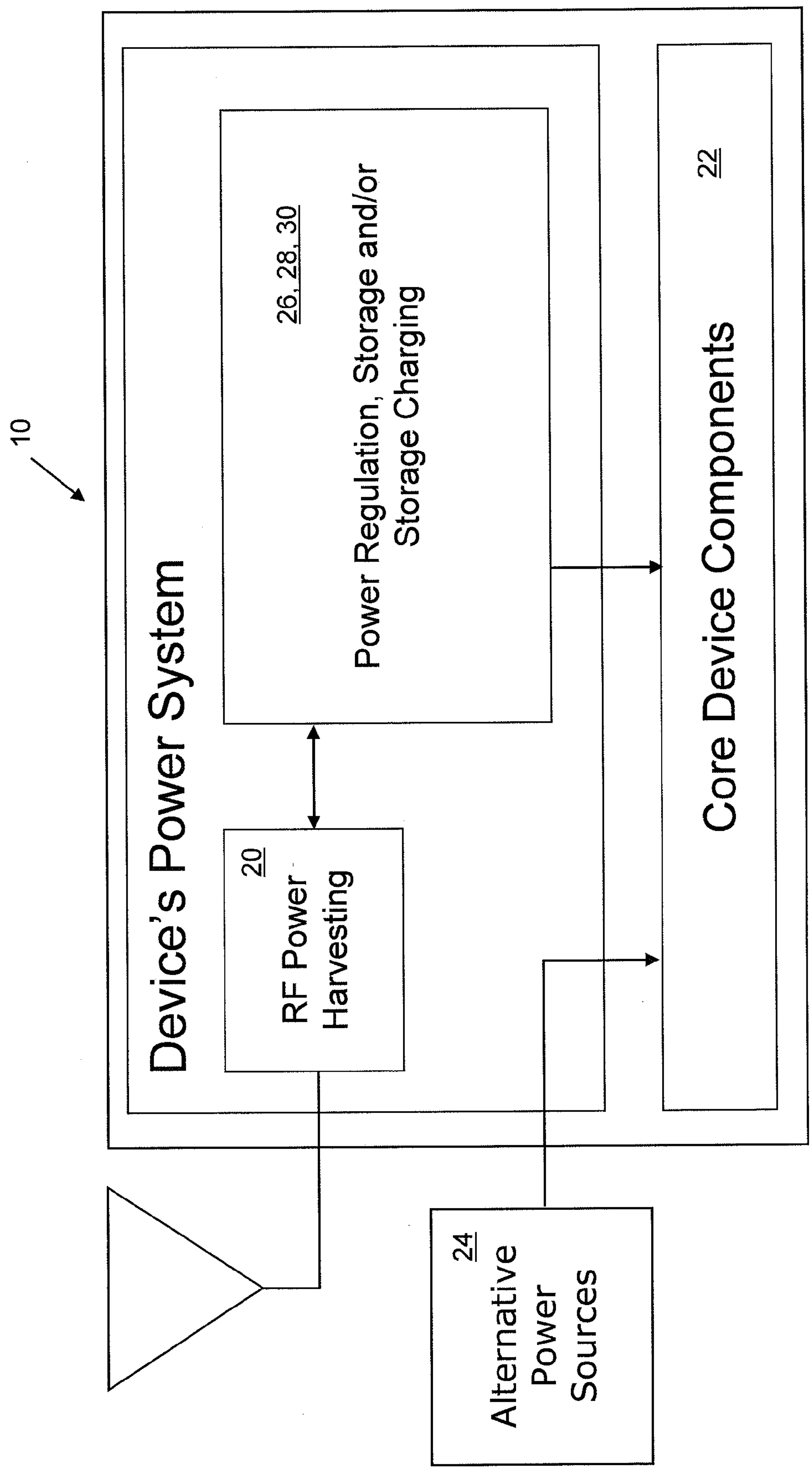
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 64

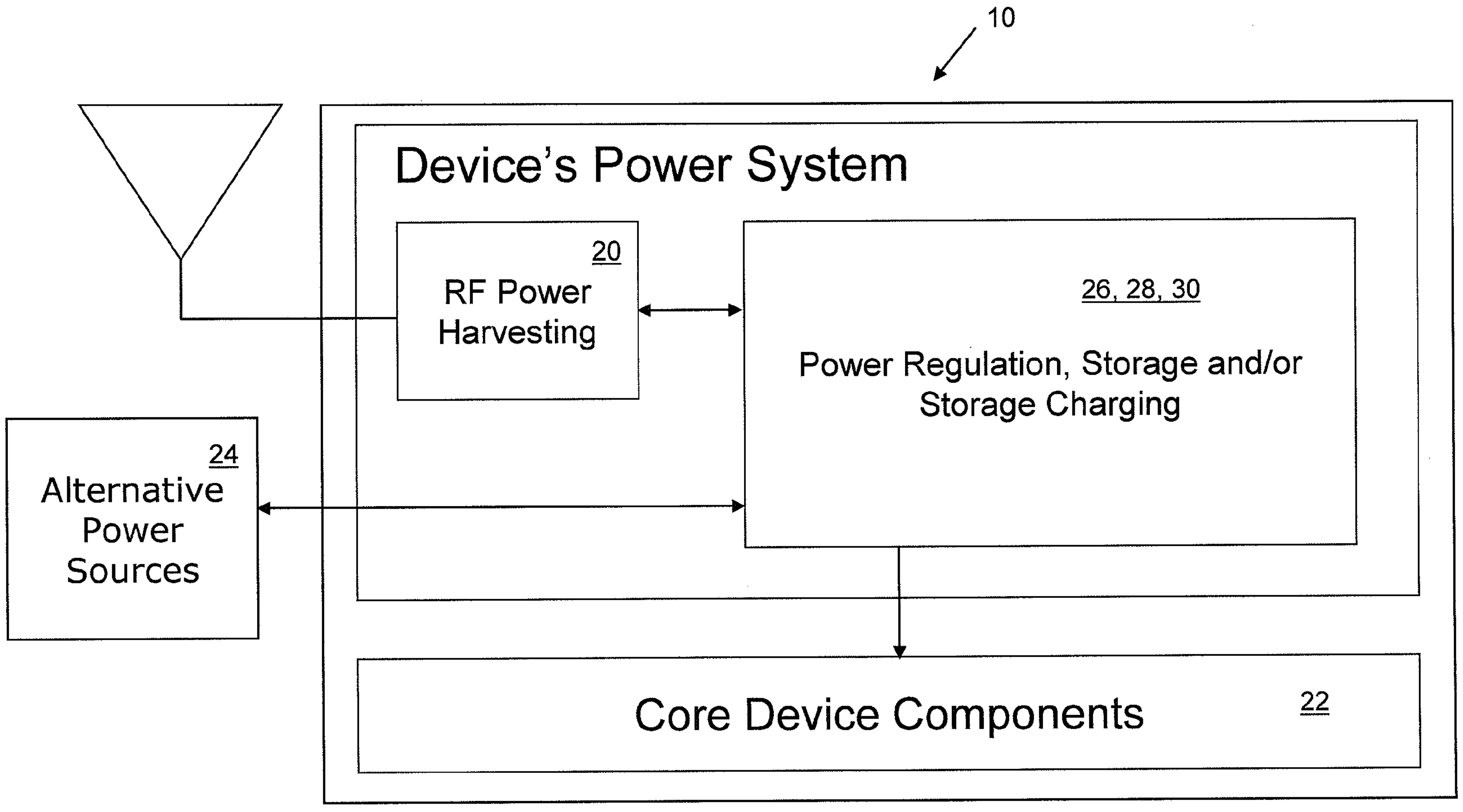
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 65

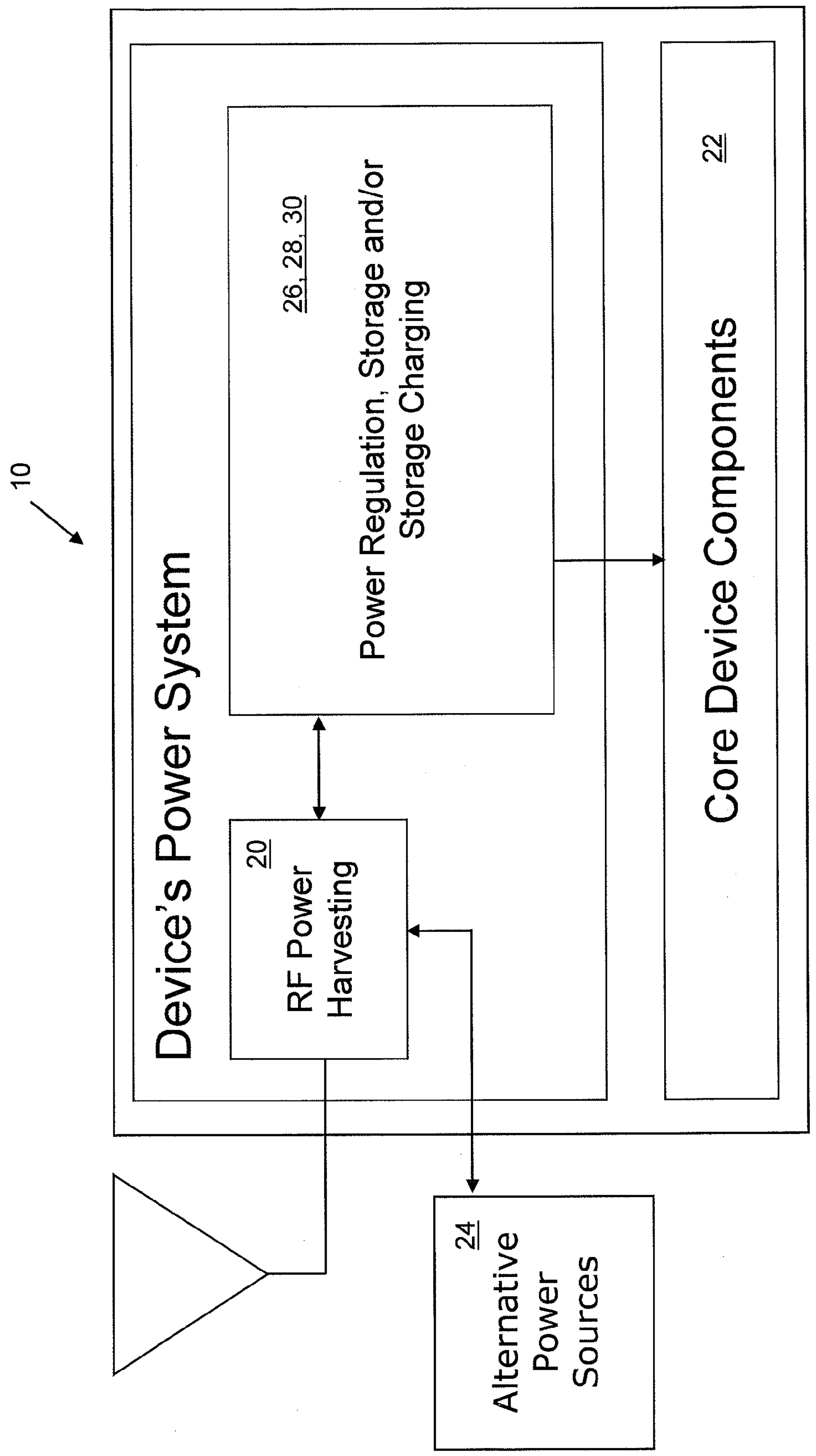
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
22
Core Device Components
24
Alternative Power Sources

Fig. 66

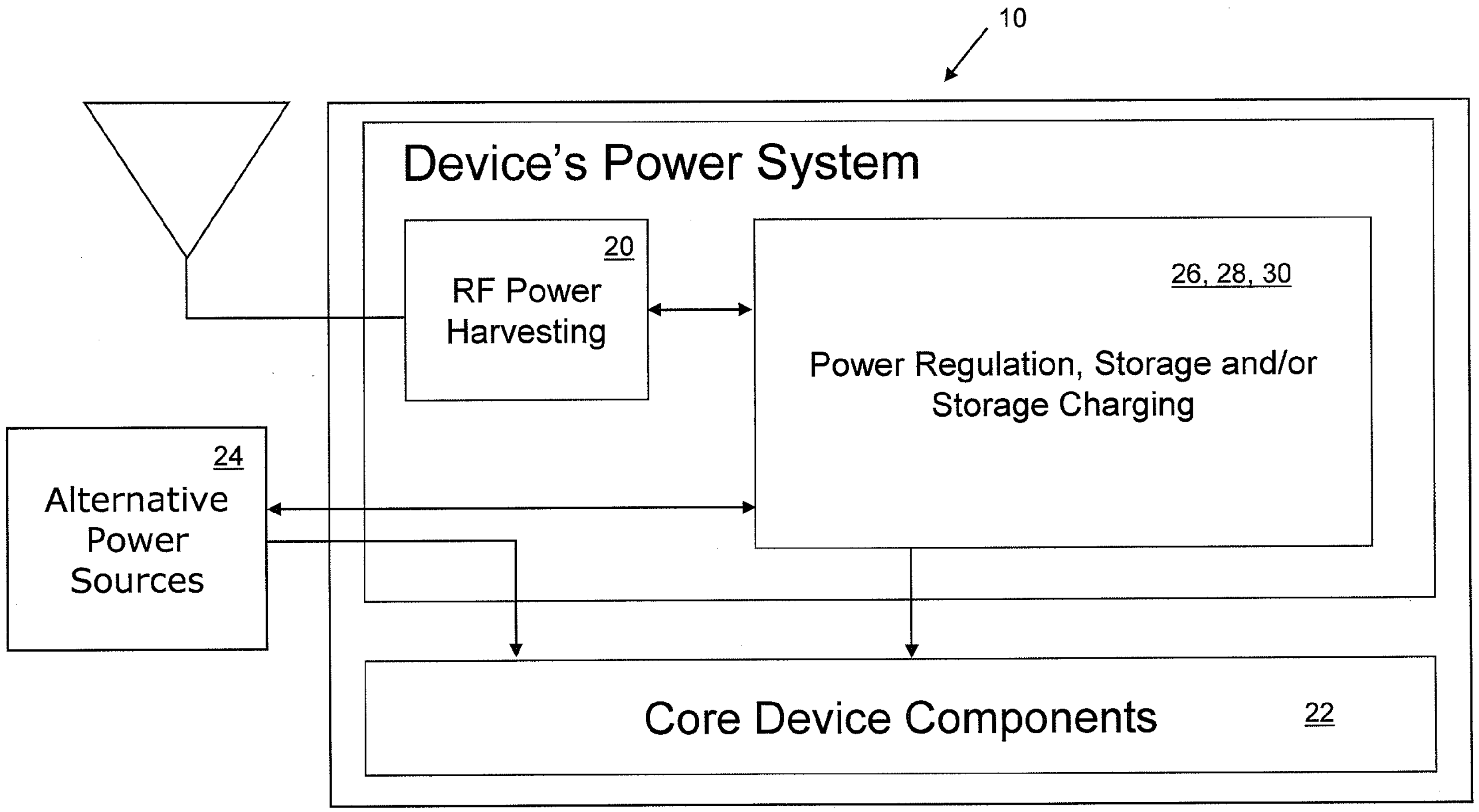
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 67

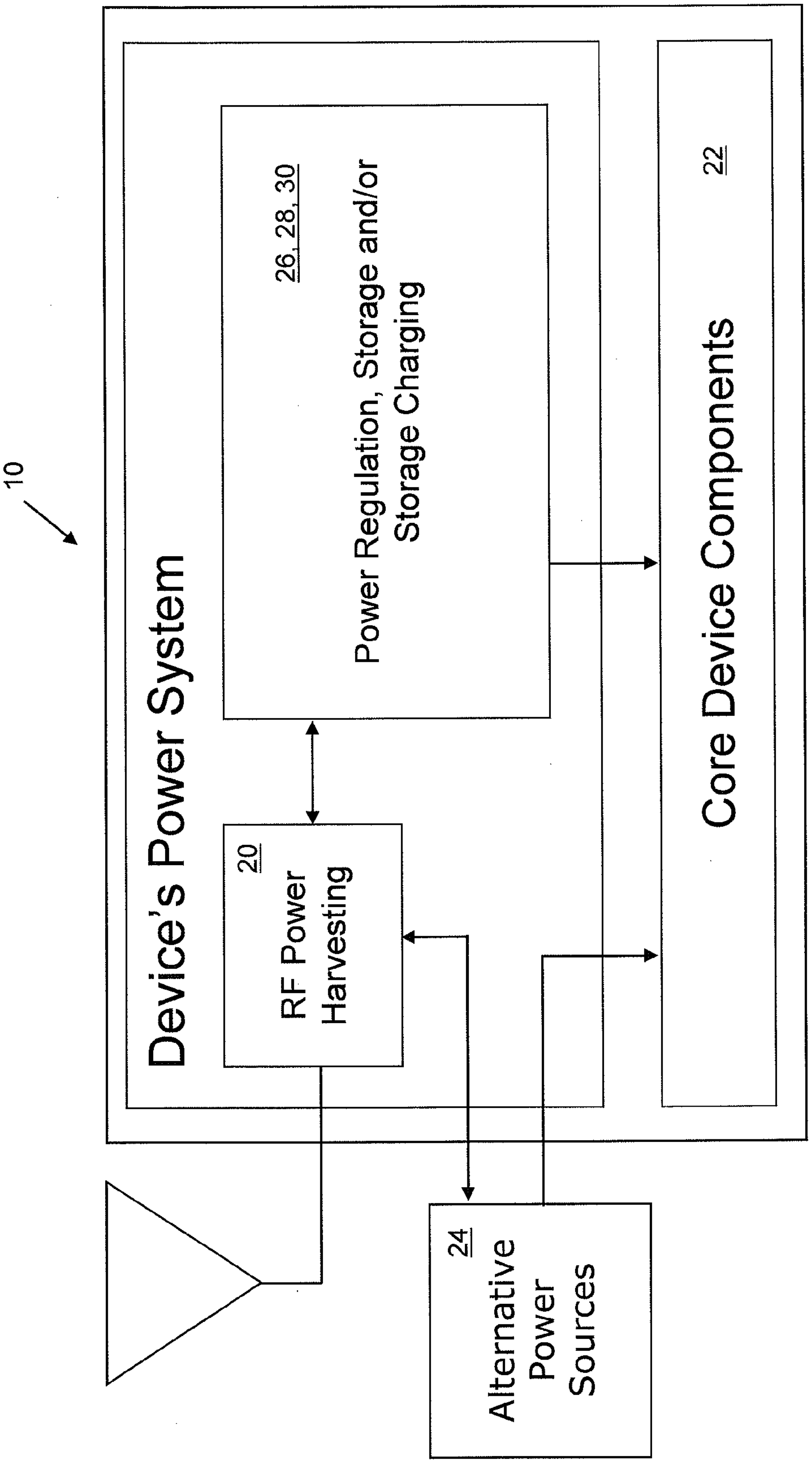
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 68

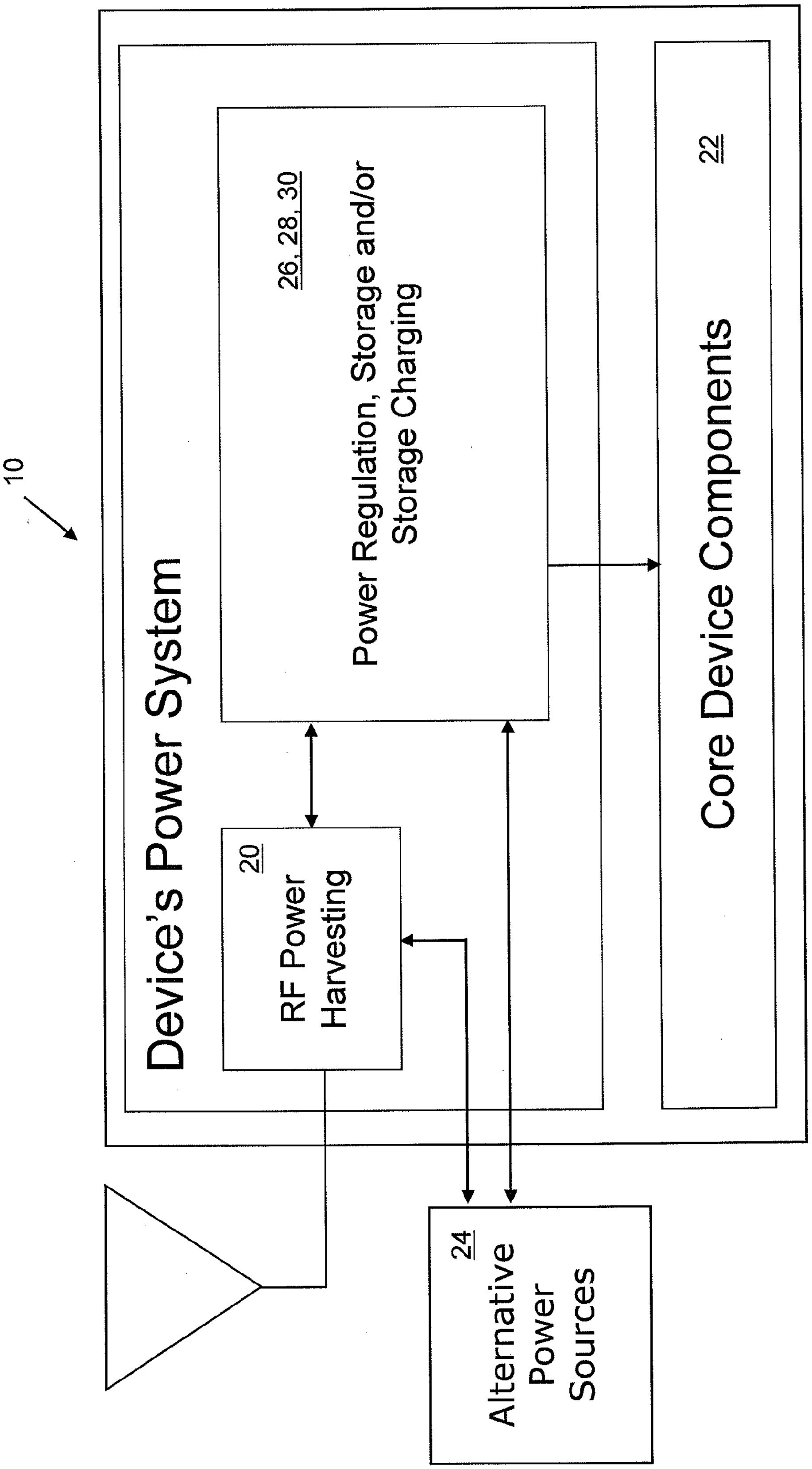
10
Device's Power System
RF Power Harvesting
20
Power Regulation, Storage and/or Storage Charging
26, 28, 30
Core Device Components
22
Alternative Power Sources
24

Fig. 69

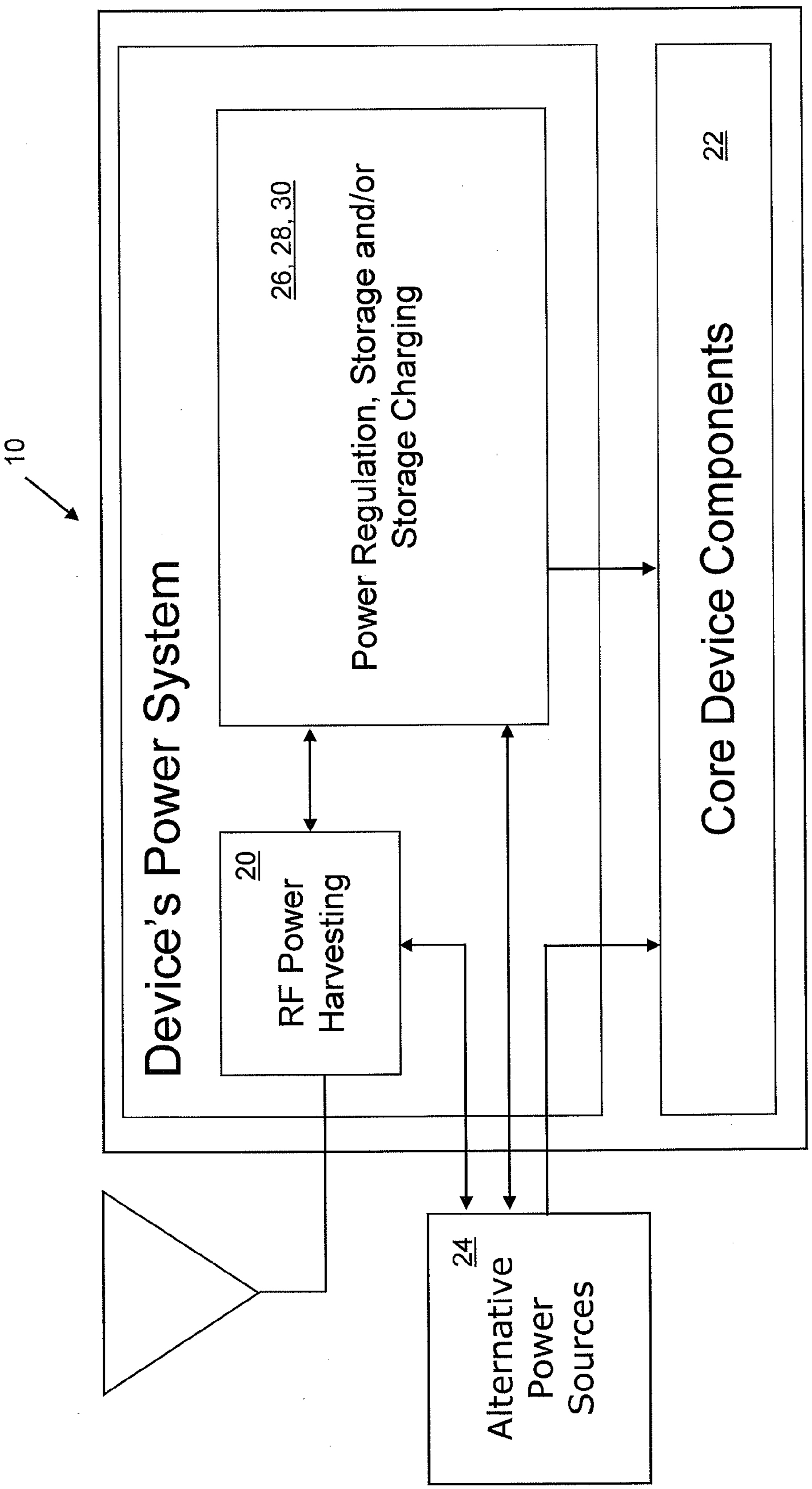
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 70

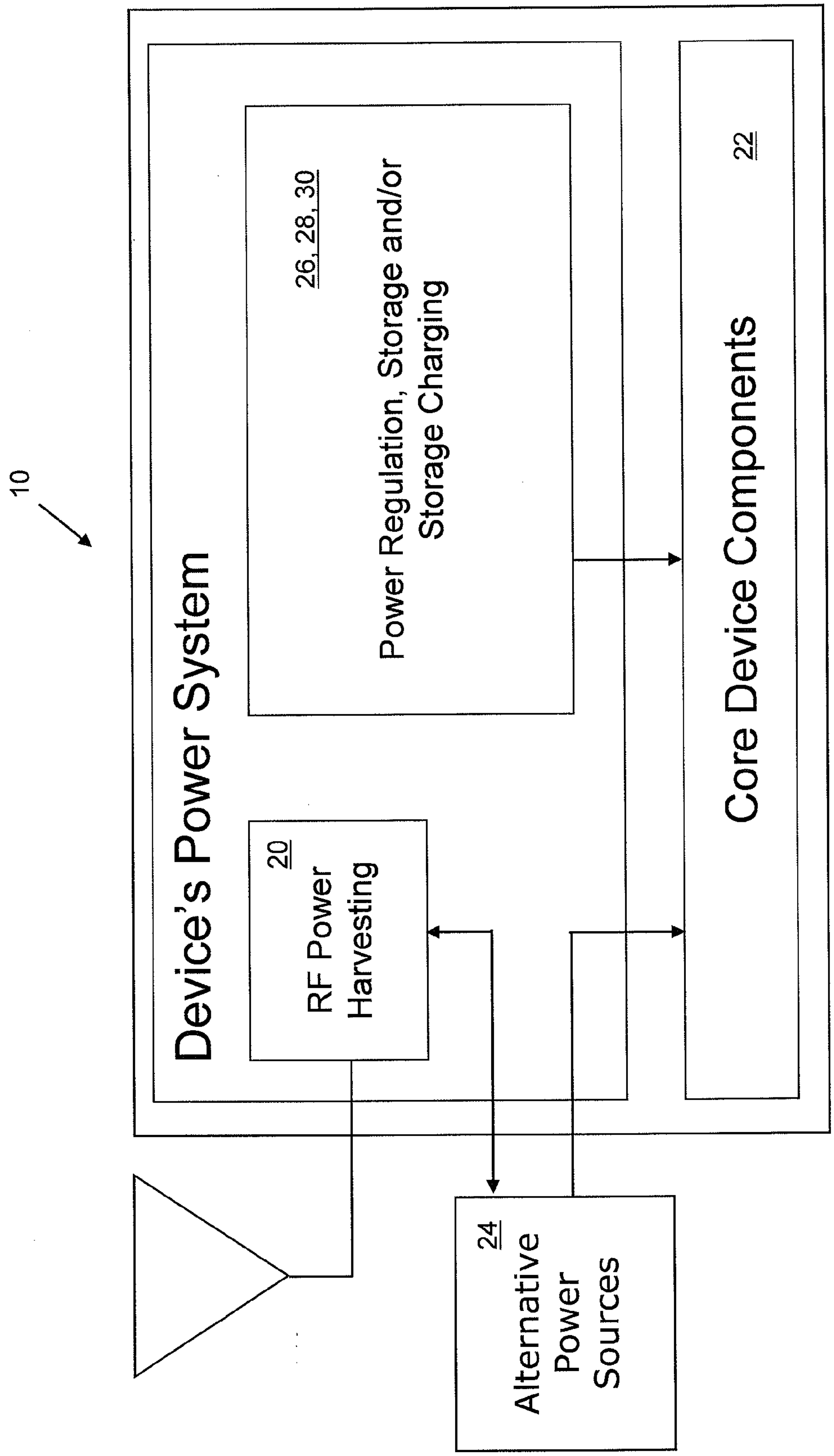
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 71

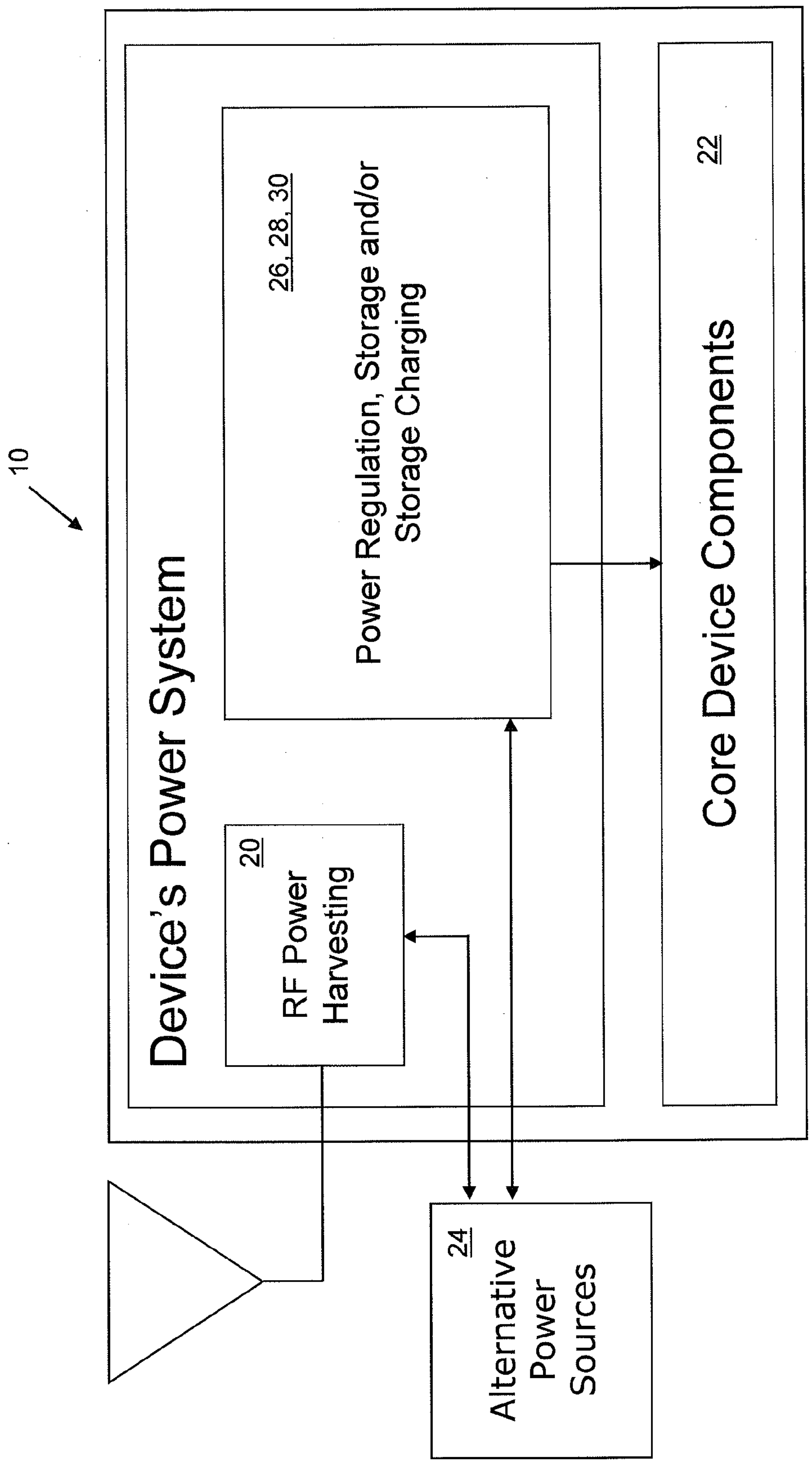
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 72

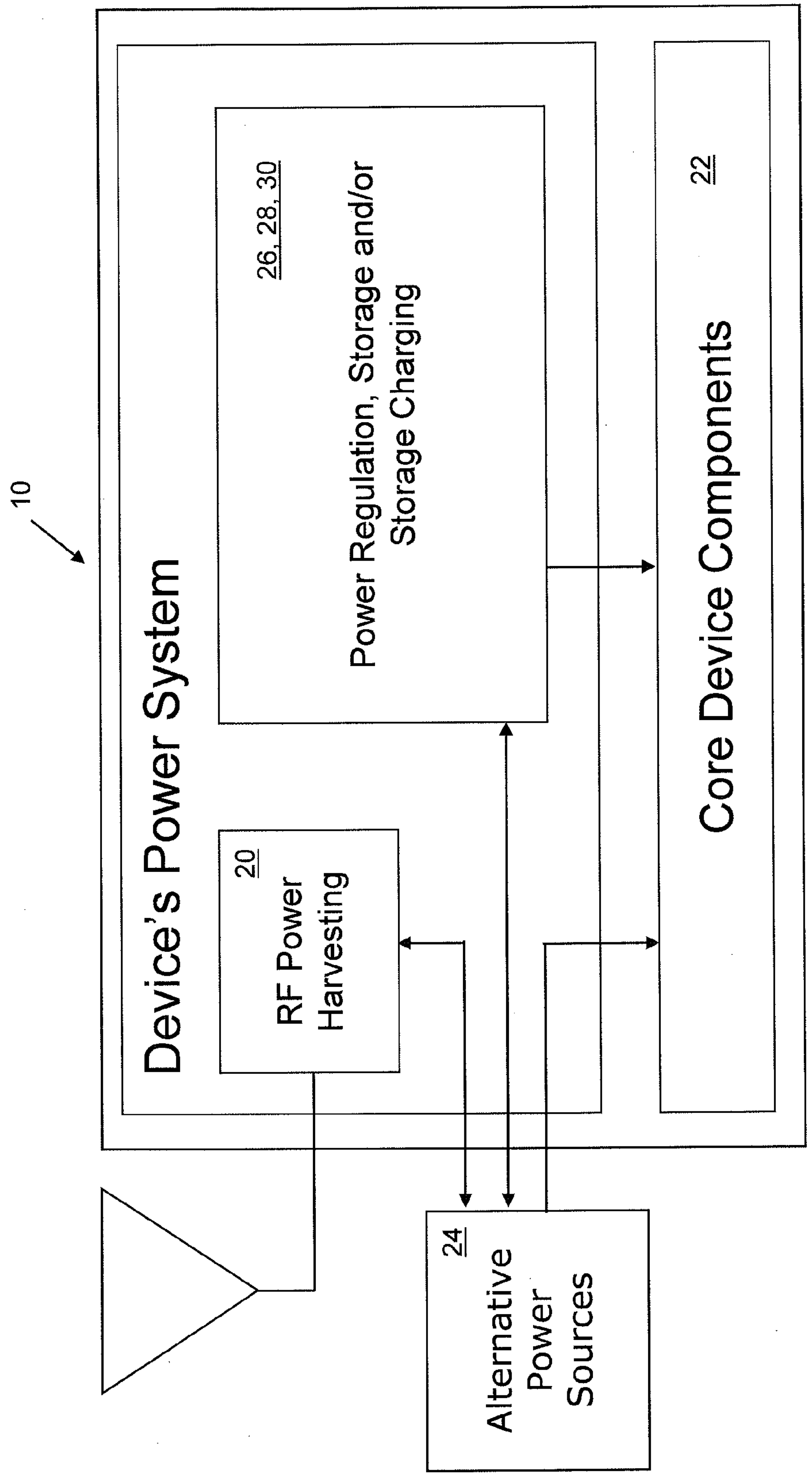
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 73

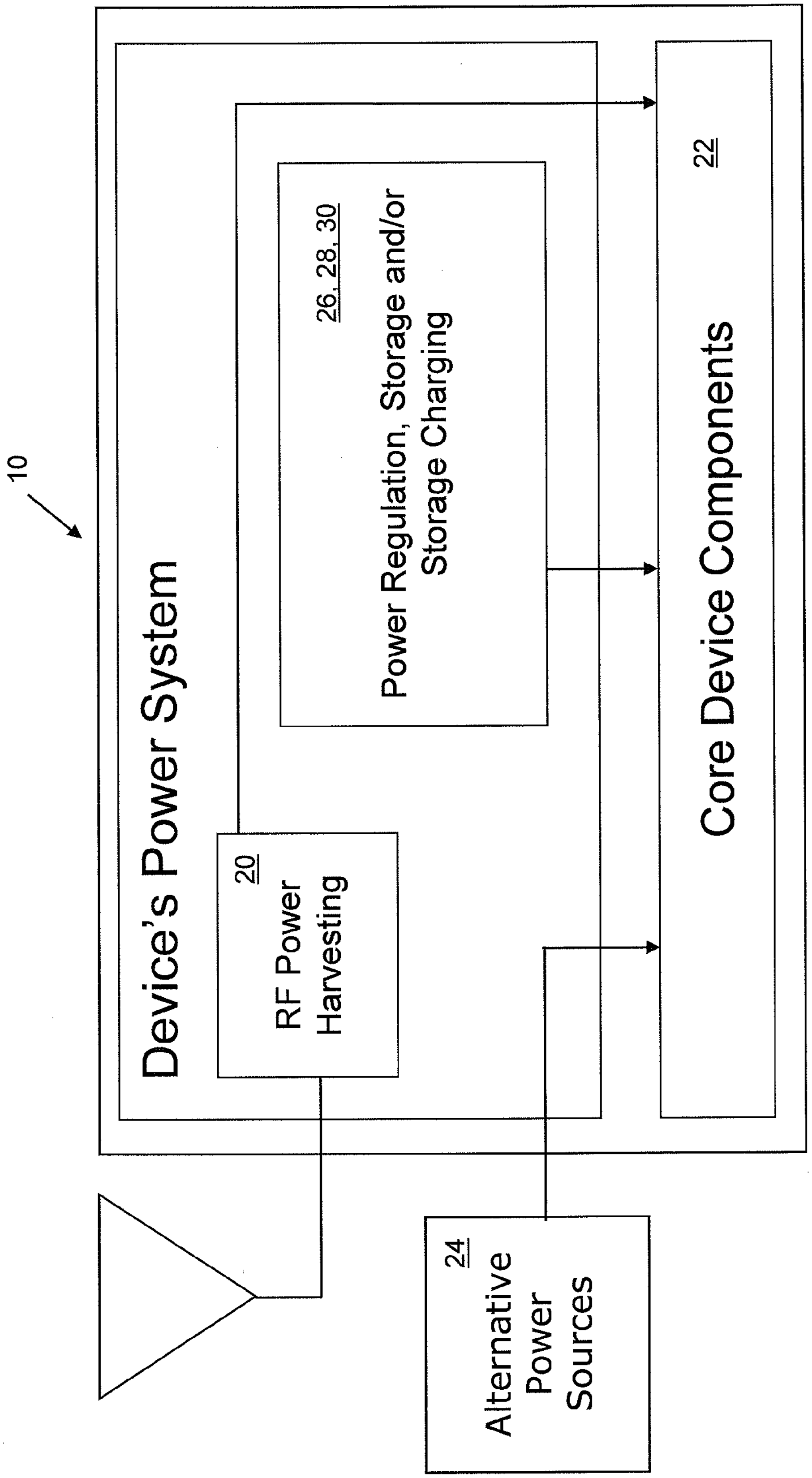
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 74

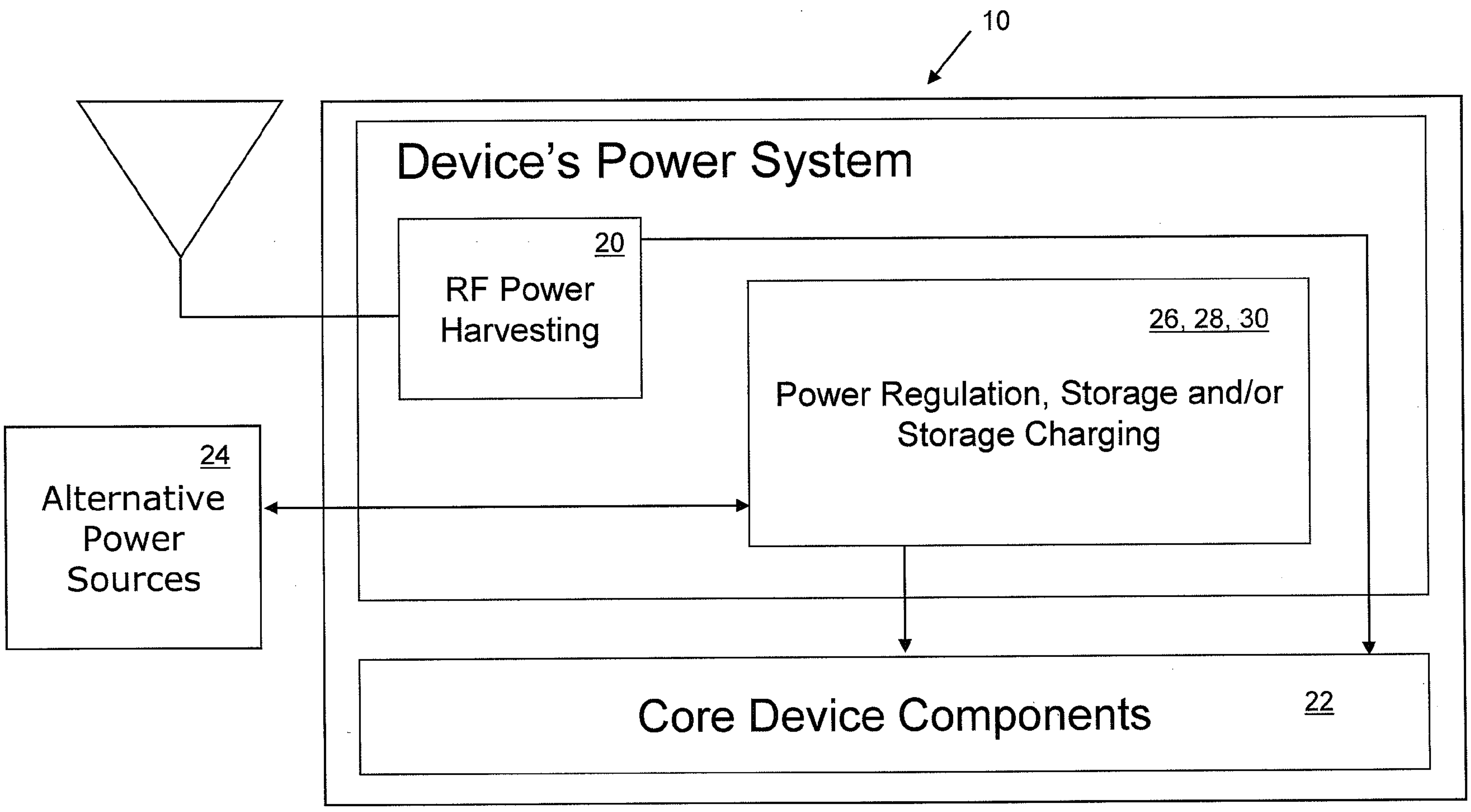
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 75

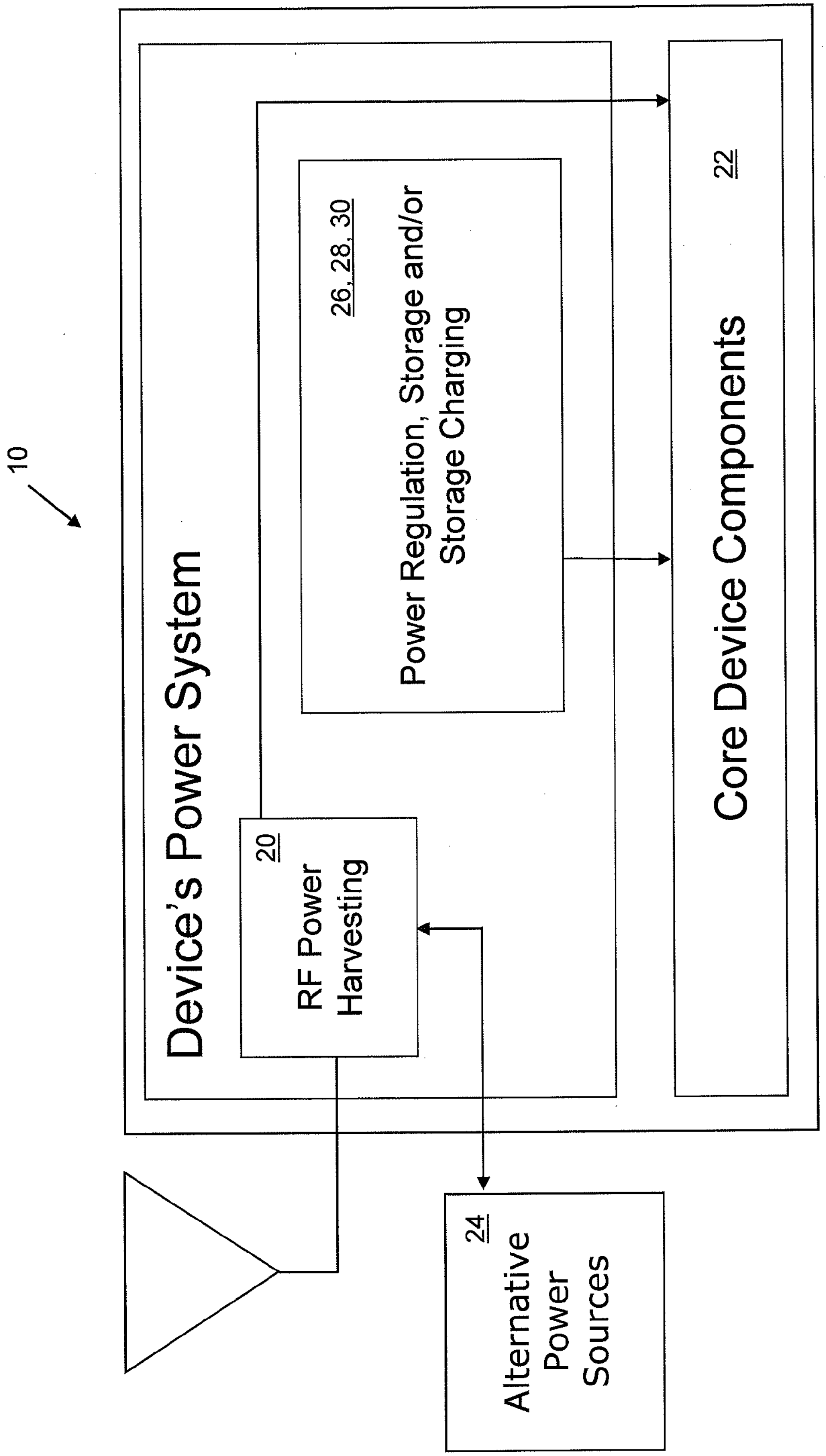
10
Device's Power System
RF Power Harvesting
20
Power Regulation, Storage and/or Storage Charging
26, 28, 30
Core Device Components
22
Alternative Power Sources
24

Fig. 76

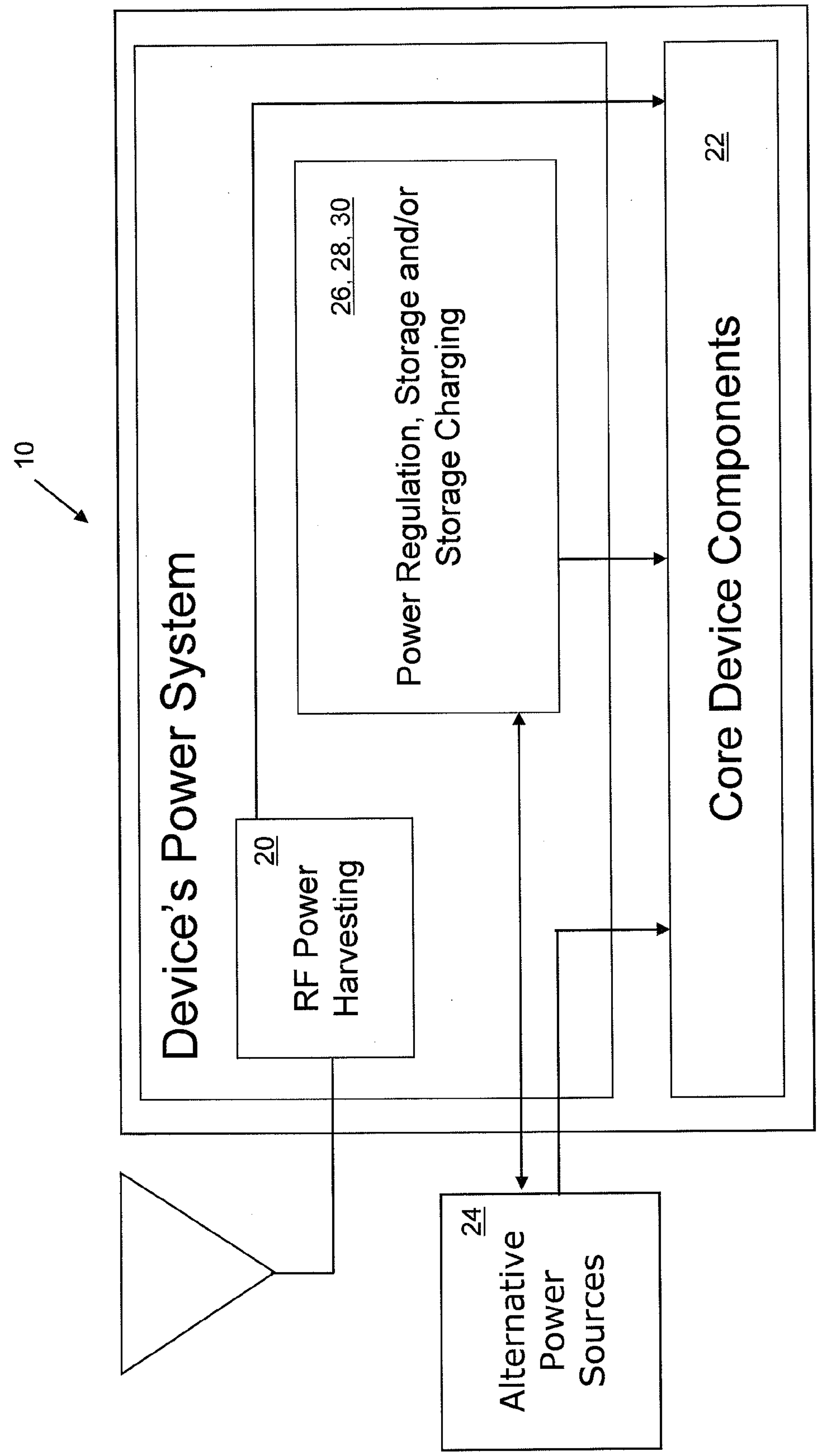
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 77

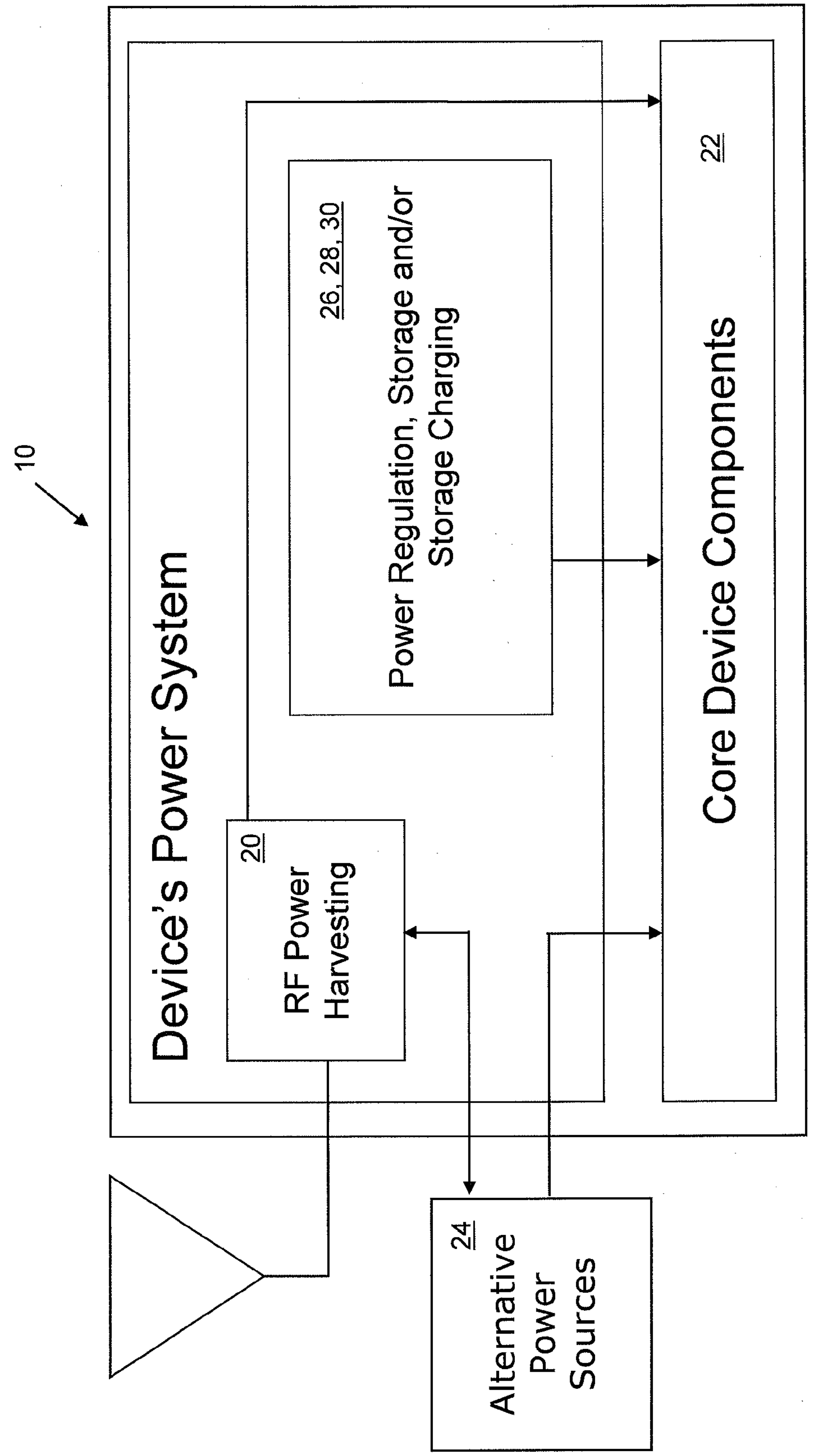
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 78

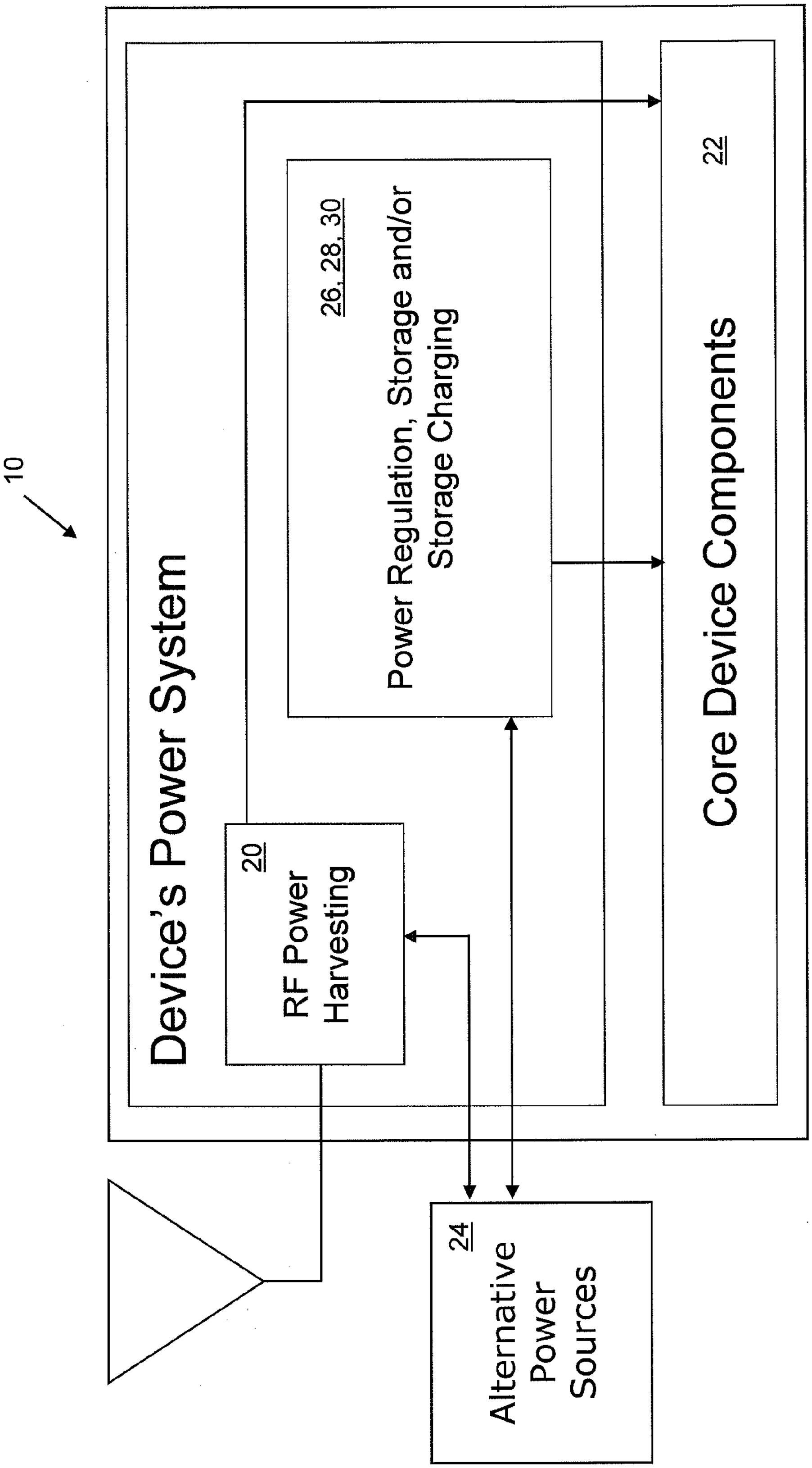
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 79

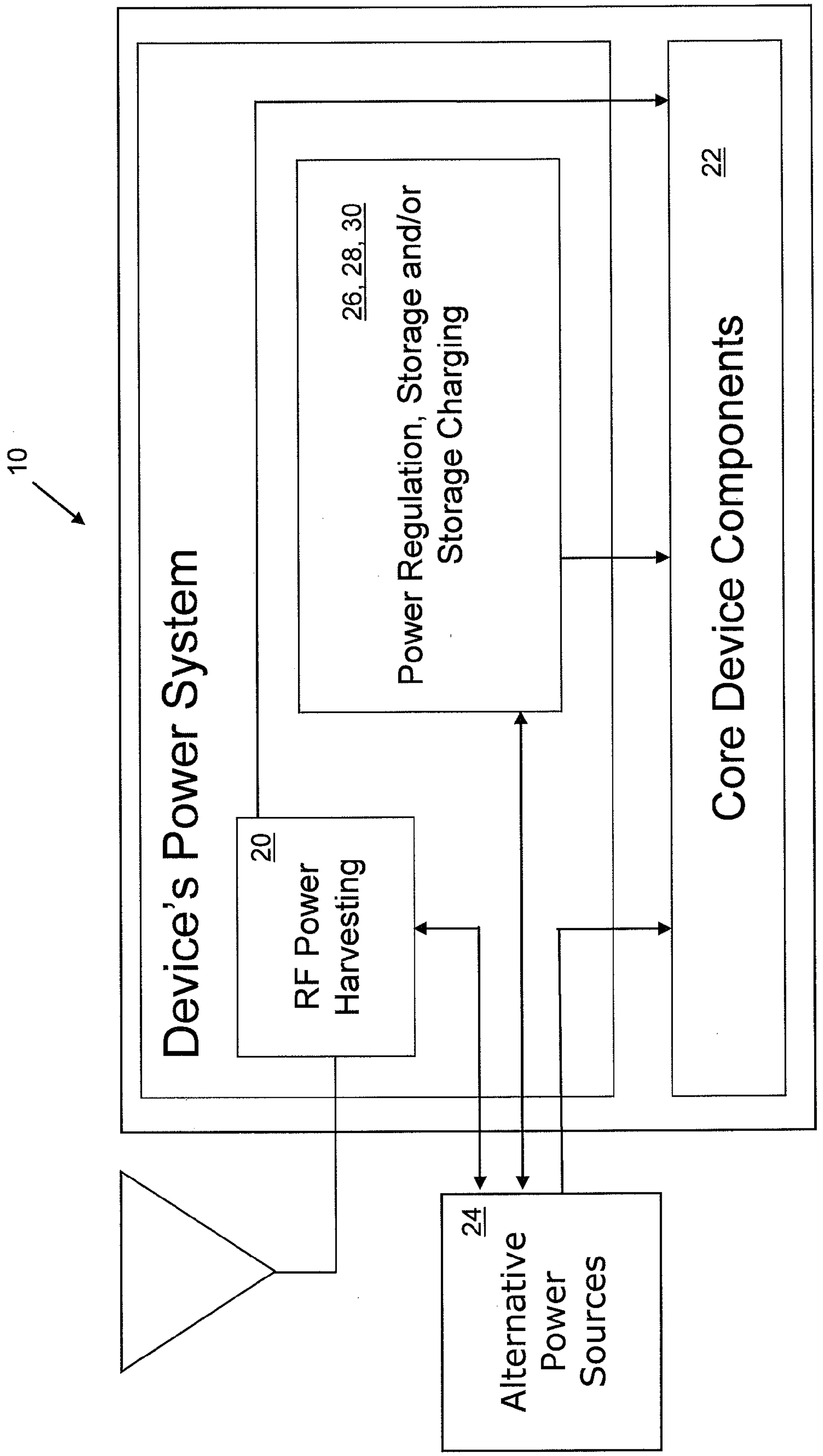
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 80

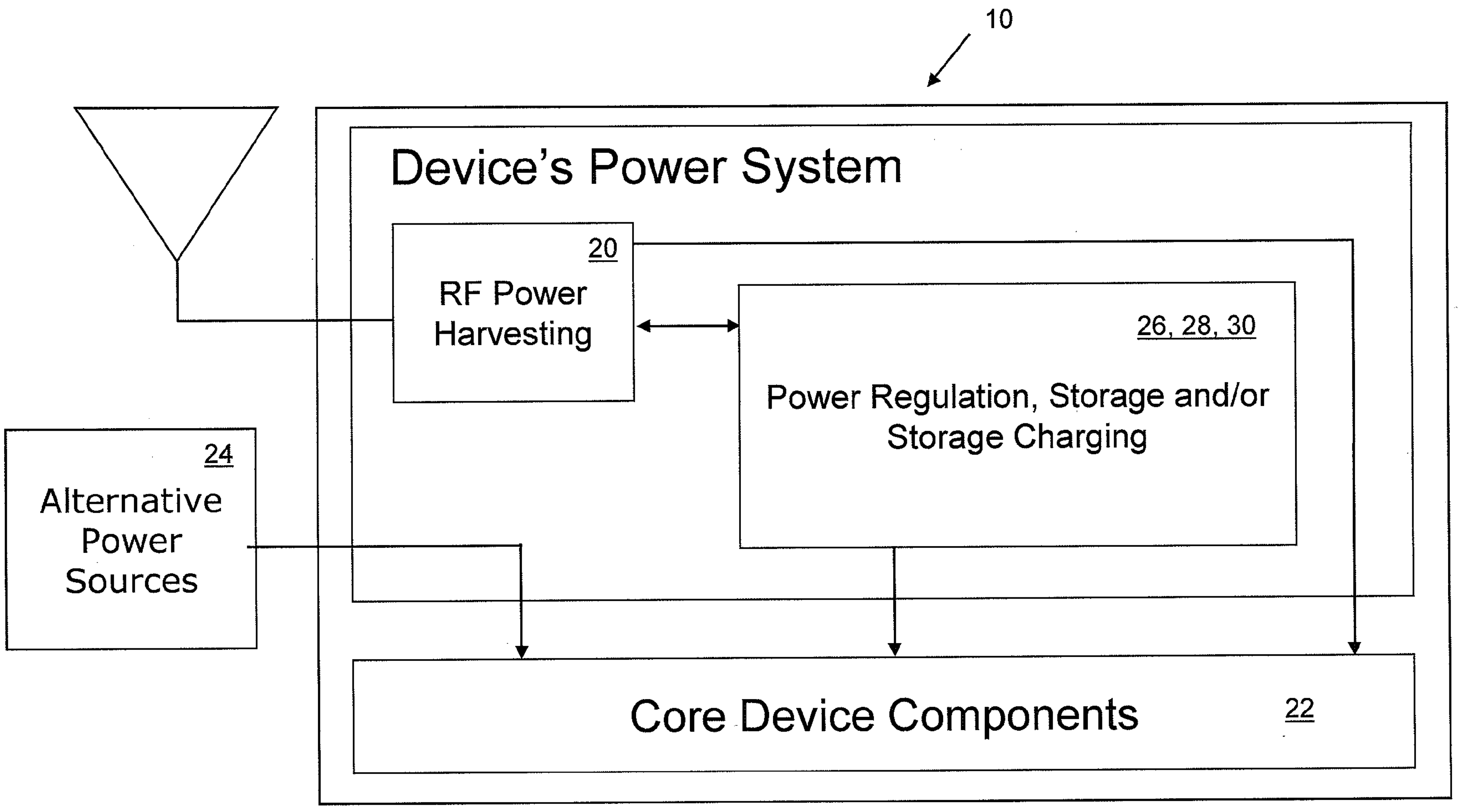
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 81

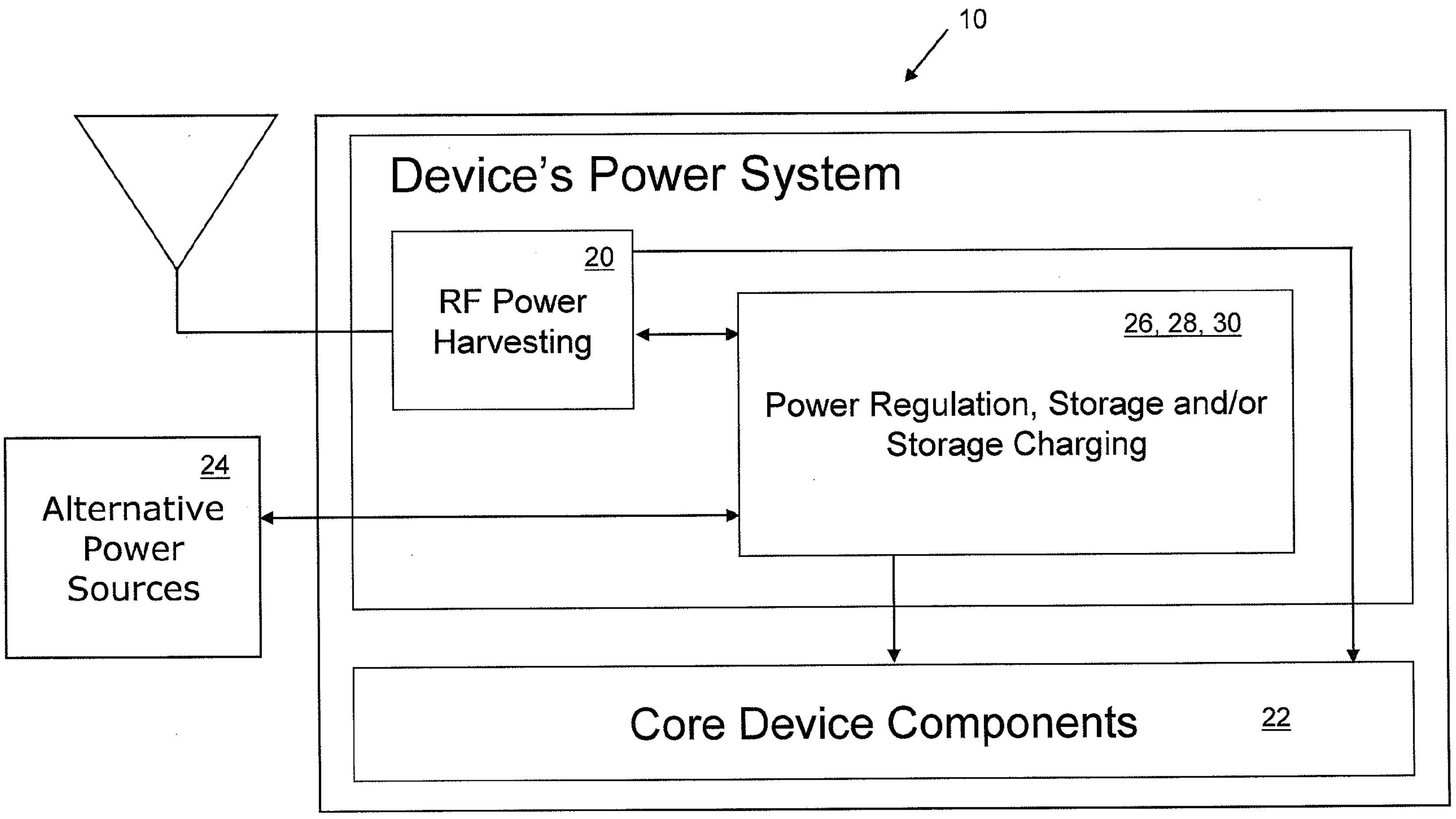
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 82

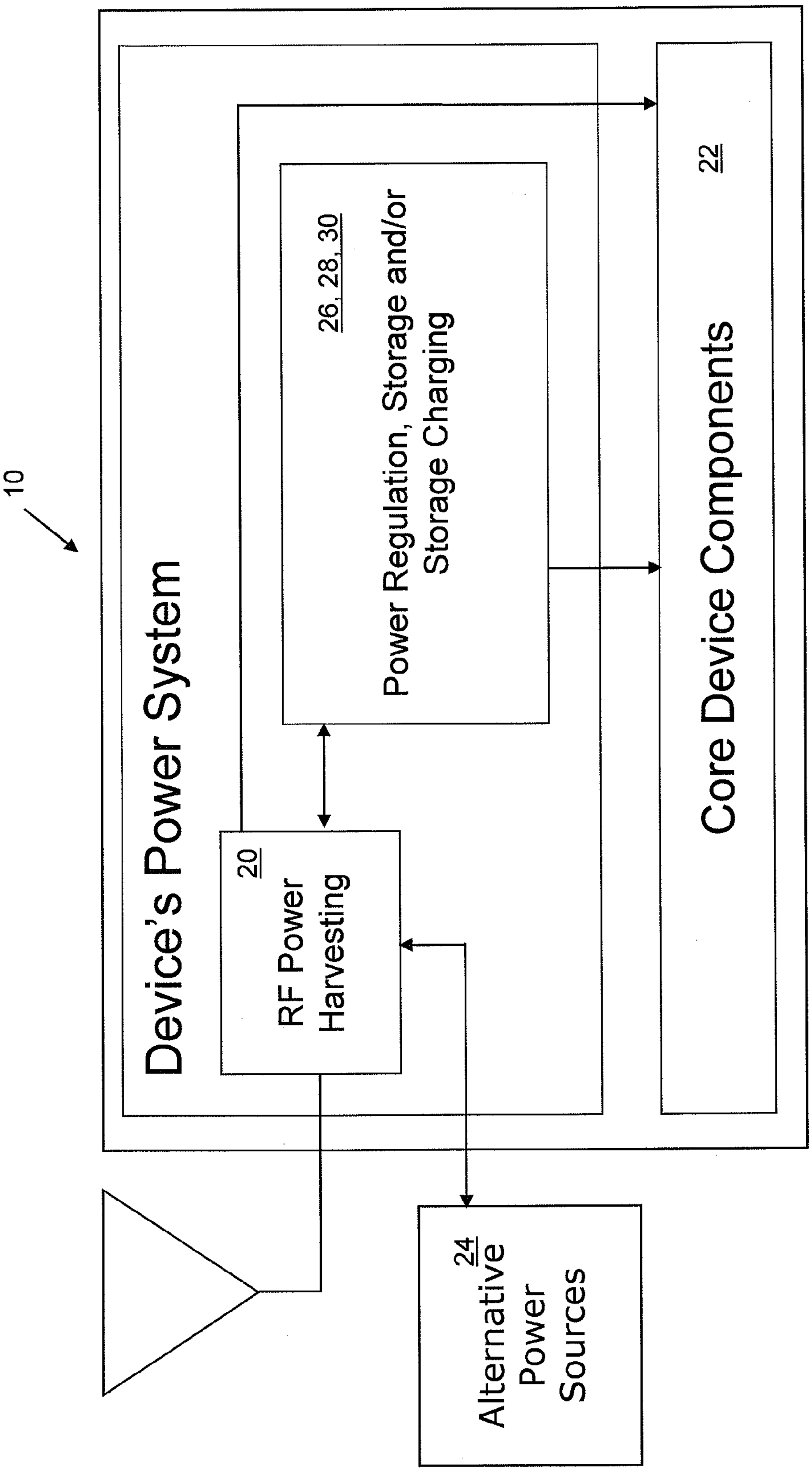
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 83

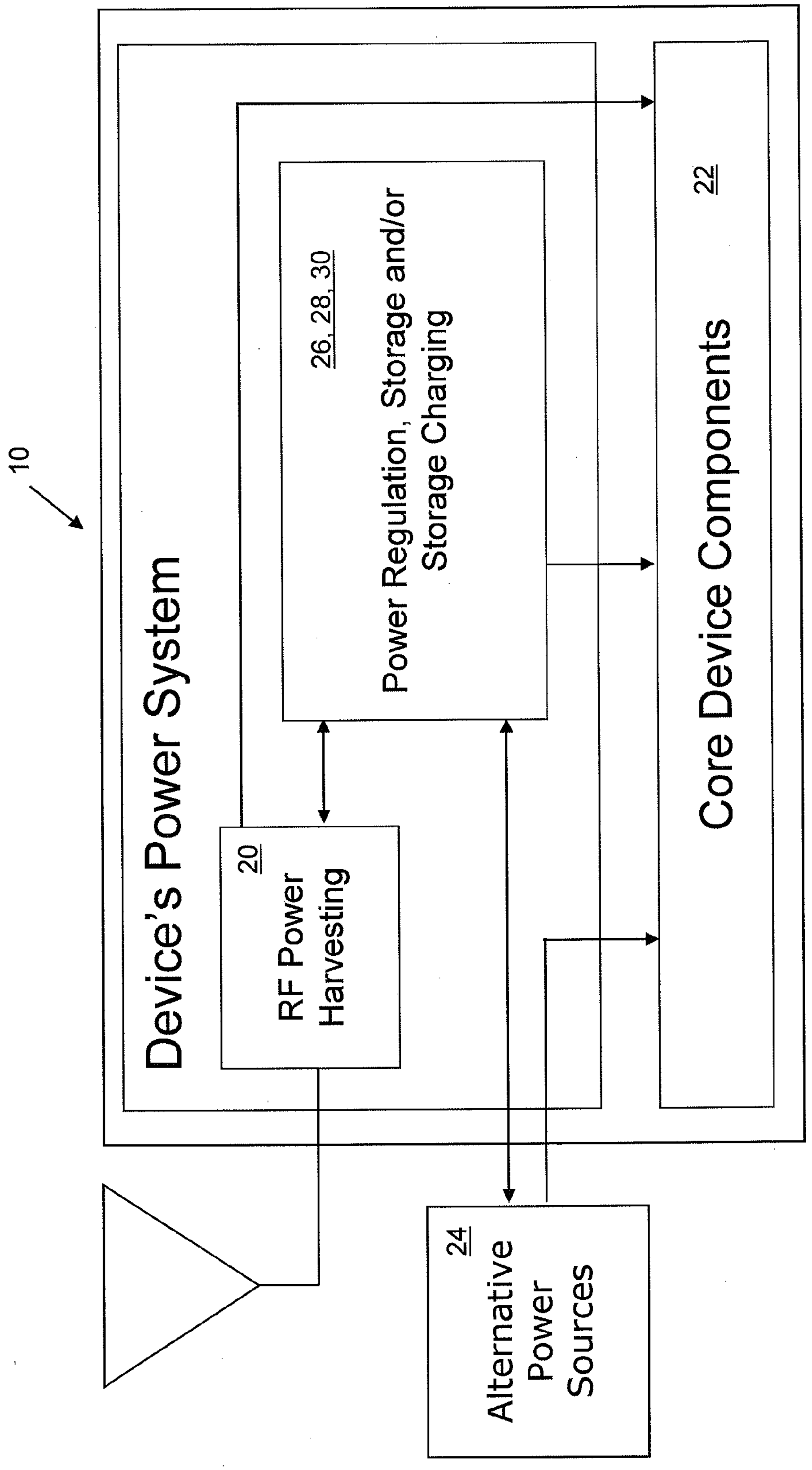
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 84

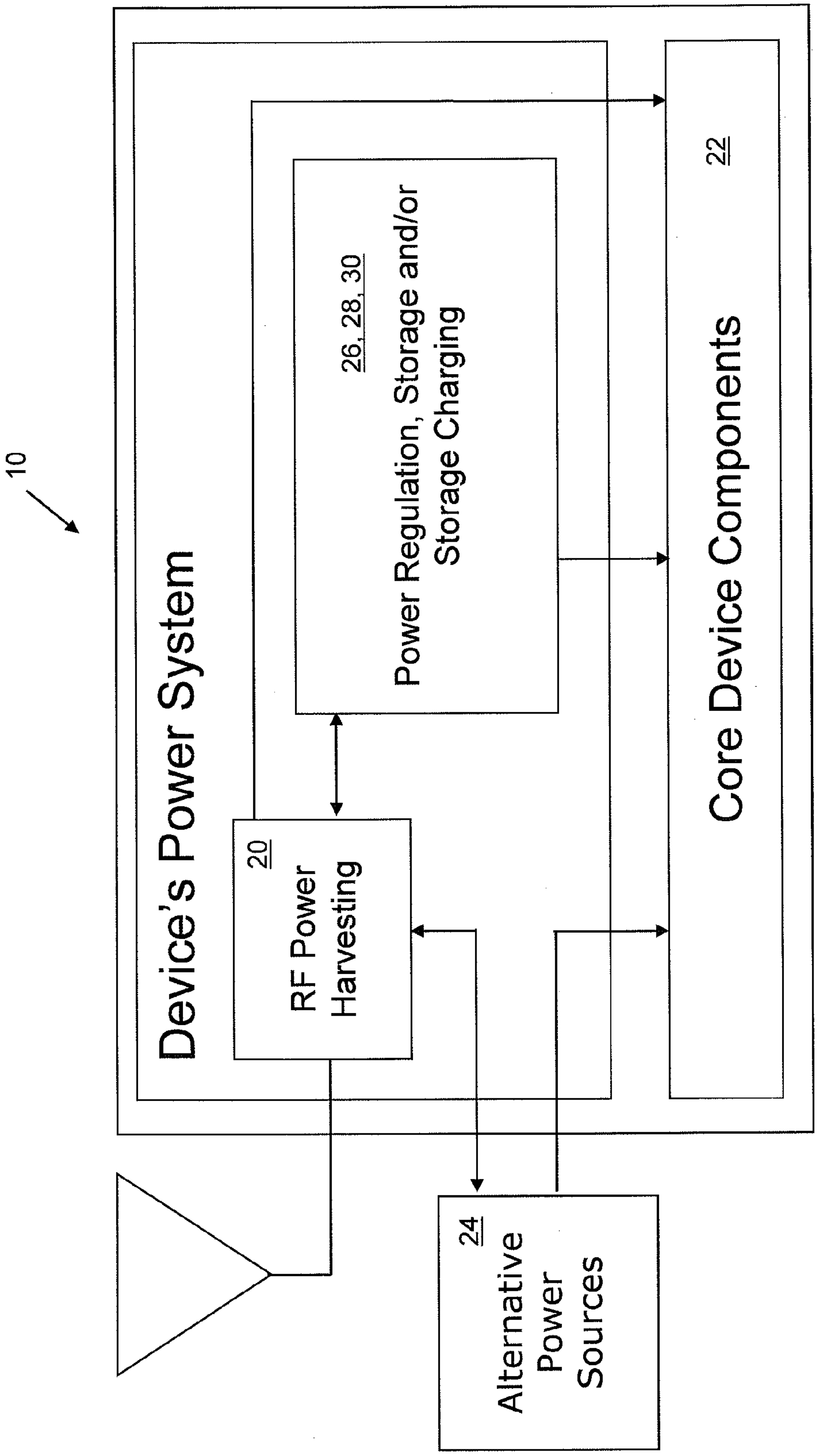
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 85

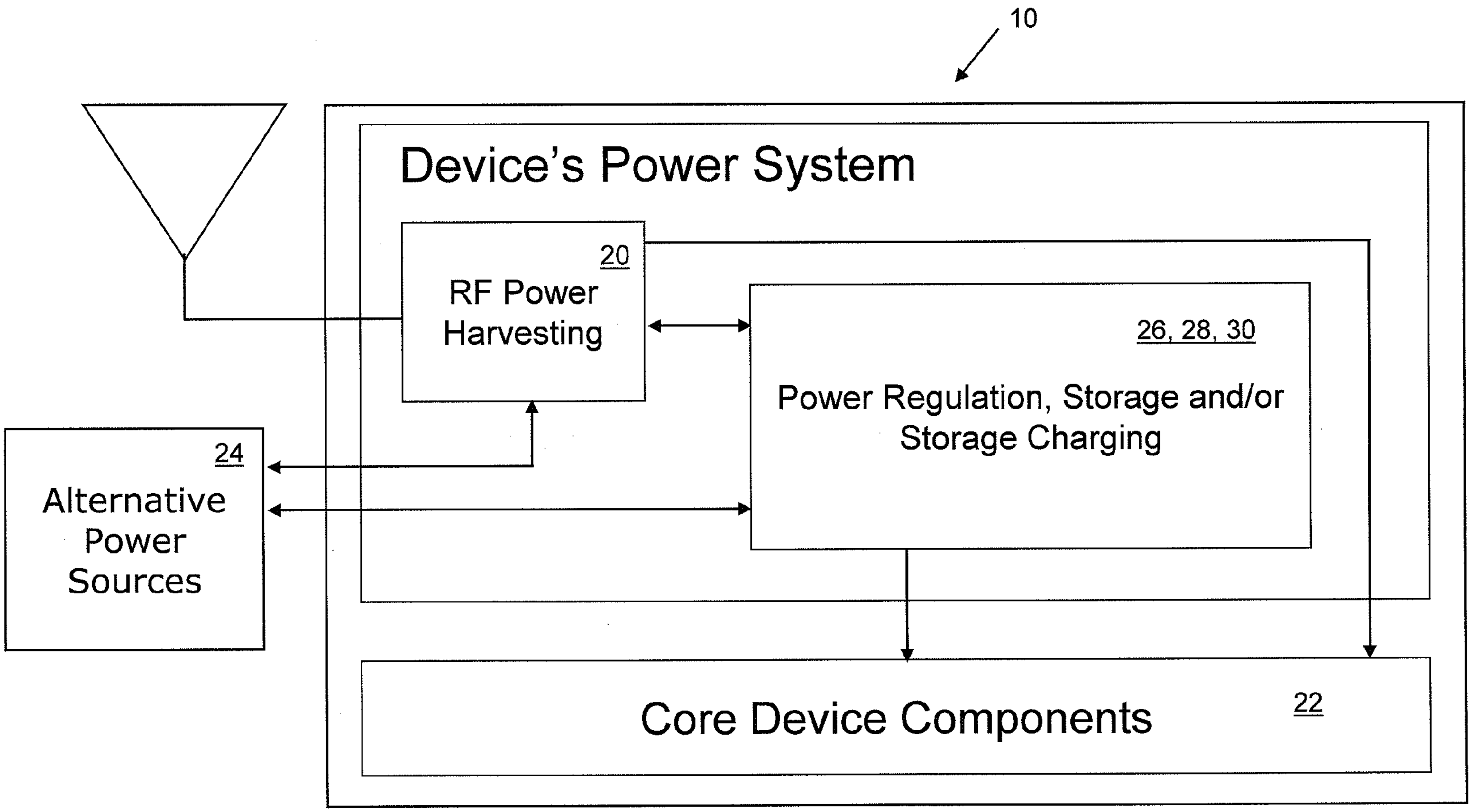
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 86

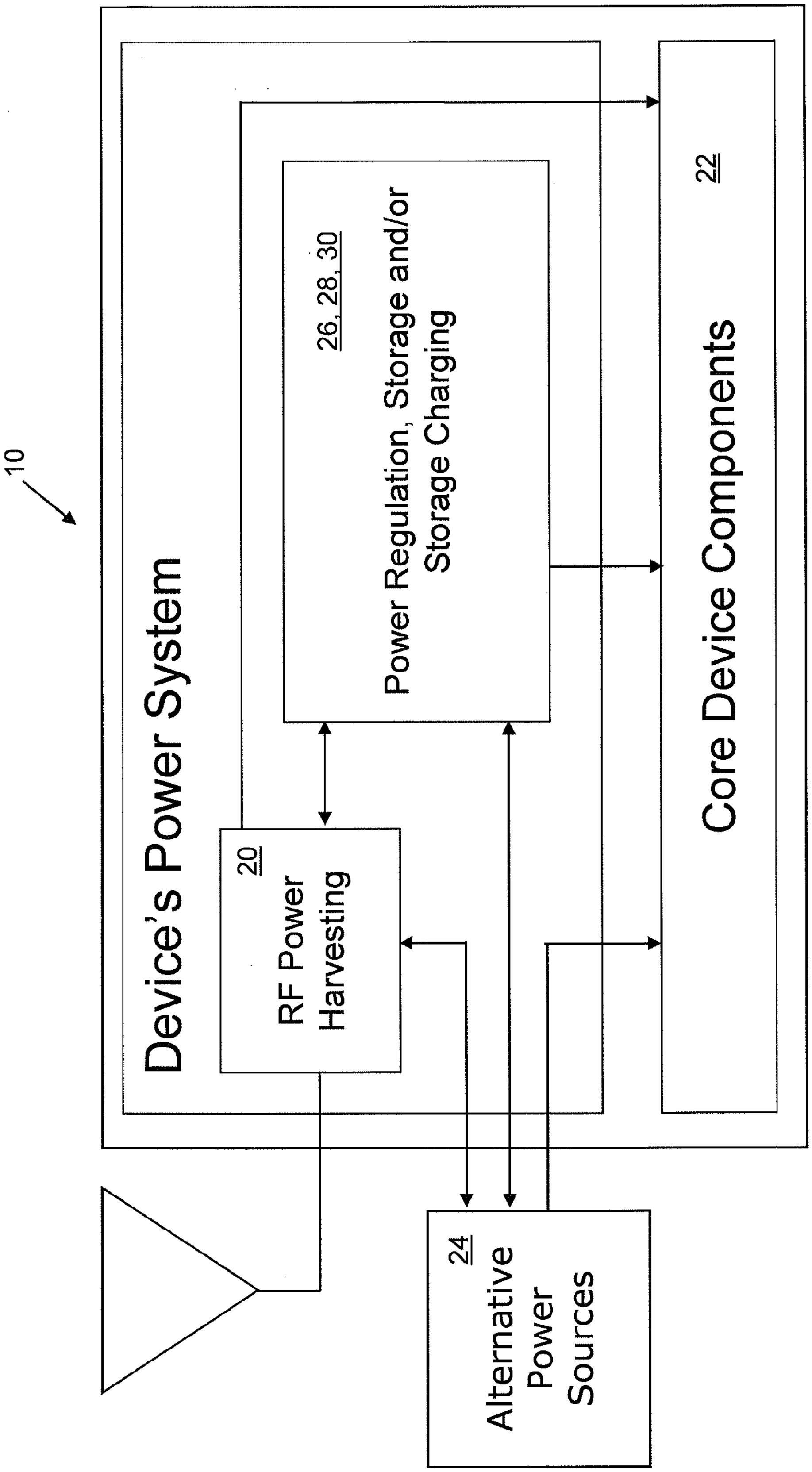
10
Device's Power System
RF Power Harvesting
20
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
Alternative Power Sources
24

Fig. 87

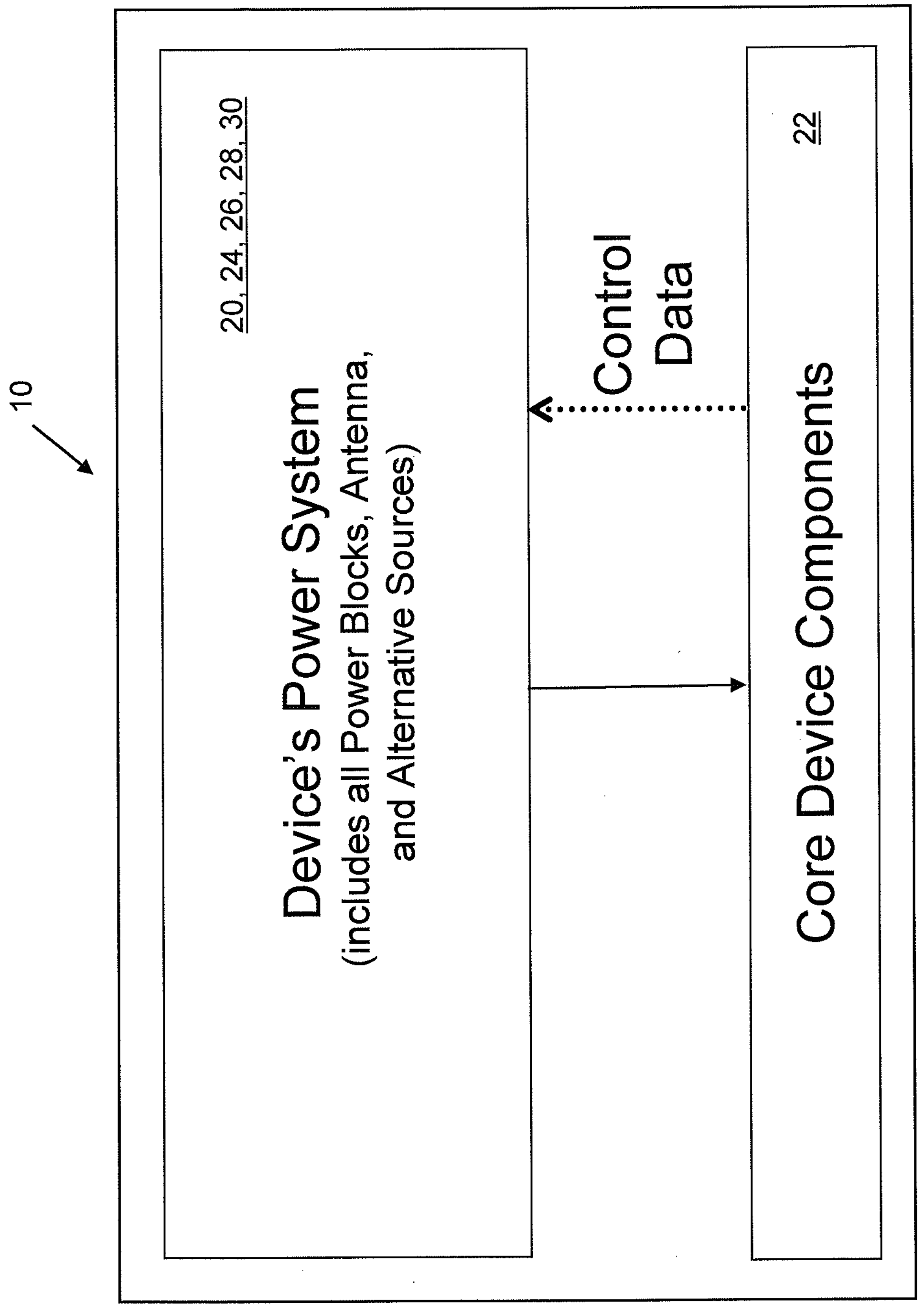
10
Device's Power System
(includes all Power Blocks, Antenna, and Alternative Sources)
20, 24, 26, 28, 30
Control Data
Core Device Components
22

Fig. 88

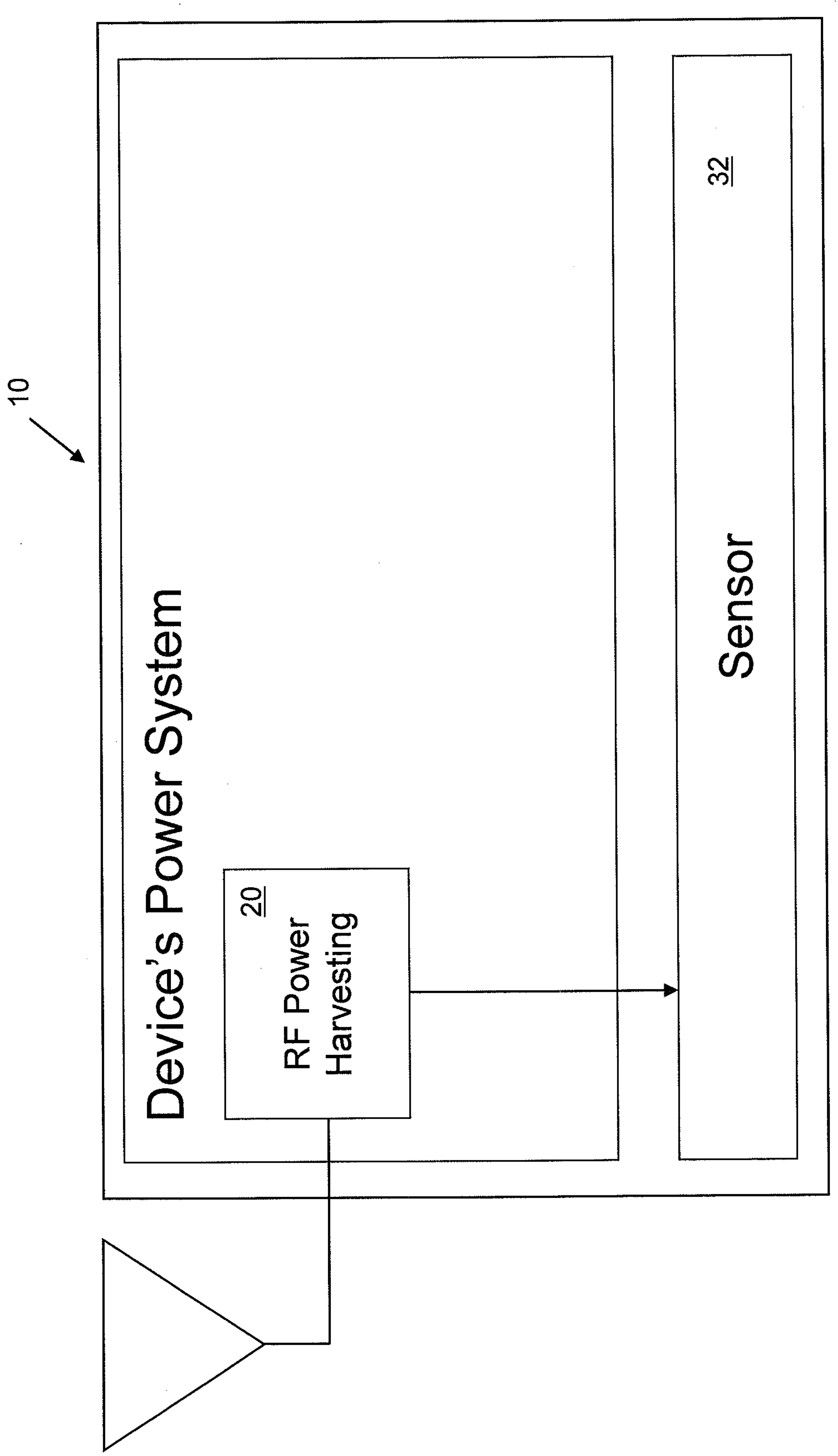
10
Device's Power System
20
RF Power Harvesting
Sensor
32

Fig. 89

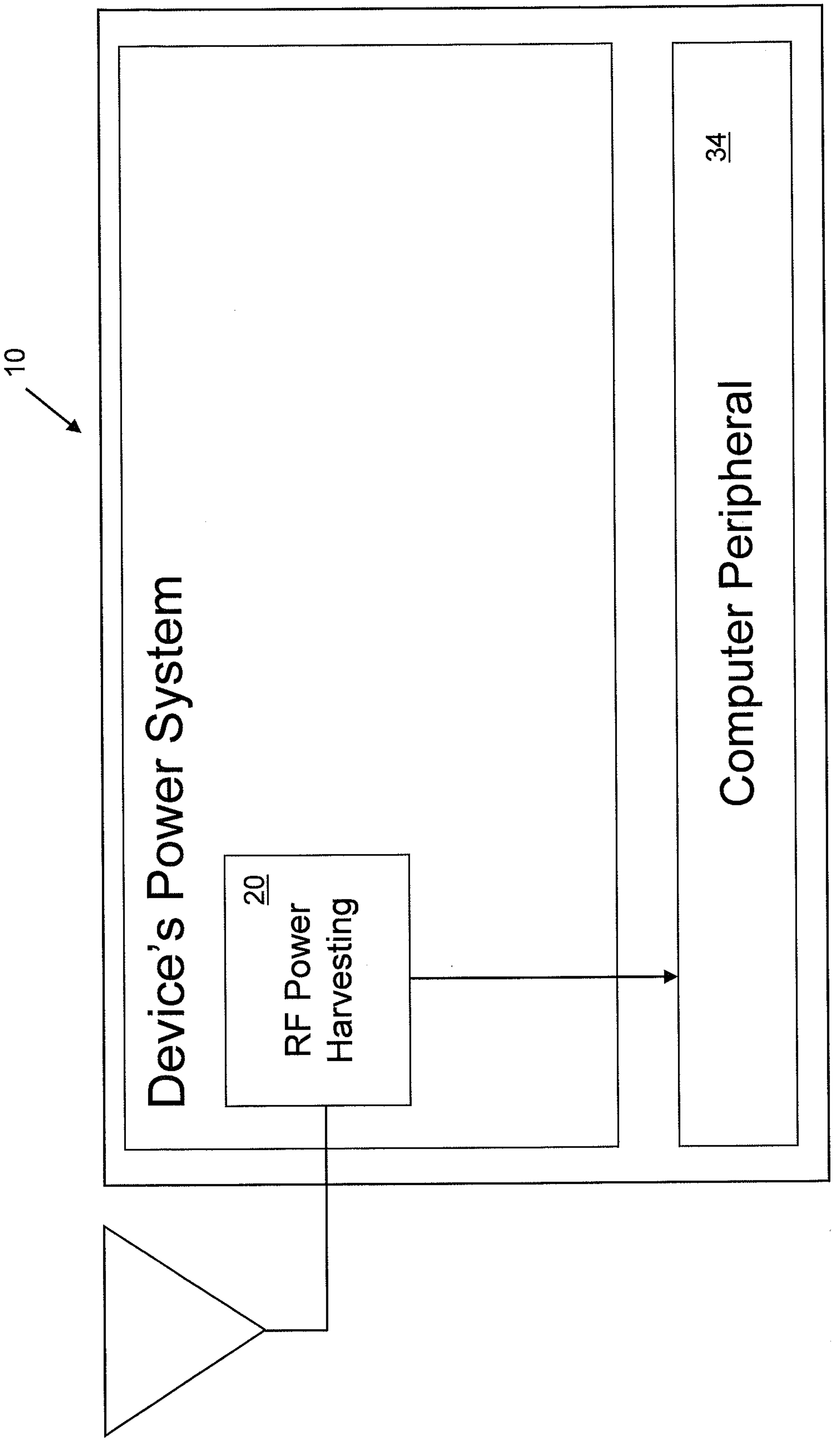
10
Device's Power System
20
RF Power Harvesting
Computer Peripheral
34

Fig. 90

POWERING DEVICES USING RF ENERGY HARVESTING

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to and is a continuation application of U.S. patent application Ser. No. 14/143,334, filed Dec. 30, 2013, which claims priority to and is a continuation application of U.S. patent application Ser. No. 12/499,618, filed Jul. 8, 2009 (now U.S. Pat. No. 8,621,245), which is a divisional application of U.S. patent application Ser. No. 11/447,412, filed Jun. 6, 2006, which claims priority to U.S. Provisional Patent Application No. 60/688,587, filed on Jun. 8, 2005; each of which is incorporated herein by reference in its entirety.

BACKGROUND

[0002] The present invention is related to the wireless powering of devices. More specifically, the present invention is related to the wireless powering of devices with a power harvester.

[0003] As processor capabilities have expanded and power requirements have decreased there has been an ongoing explosion of devices that operate completely independent of wires or power cords. These "untethered" devices range from cell phones, and wireless keyboards to building sensors and active RFID tags.

[0004] Engineers and designers of these untethered devices continue to have to deal with the limitations of portable power sources, primarily batteries as the key design parameter. While performance of processors and portable devices have been doubling every 18-24 months driven by Moore's law, battery technology in terms of capacity has only been growing at measly 6% per year. Even with power conscious designs and the latest in battery technology, many devices do not provide the lifetime cost and maintenance requirements for applications that require a large number of untethered devices such as logistics, and building automation. Today's devices that need two-way communication require scheduled maintenance every three to 18 months to replace or recharge the device's power source (typically a battery). One-way devices simply broadcasting their status (one-way) such as automated utility meter readers have a better battery life, typically requiring replacement within 10 years. For both device types, scheduled power-source maintenance is costly and disruptive to the entire system that a device is intended to monitor and/or control. Unscheduled maintenance trips are even more costly and disruptive. On a macro level, the relatively high cost associated with the internal battery also reduces the practical, or economically viable, number of devices that can be deployed.

[0005] The ideal solution to the power problem for untethered devices is a device or system that can collect and harness sufficient energy from the external environment. The harnessed energy would then either directly power an untethered device or augment a battery or other storage component. Directly powering an untethered device enables the device to be constructed without the need for a battery. Augmenting a storage component could be along two lines: 1) increasing the overall life of the device or 2) by providing more power to the device to increase the functionality of the device. The other parameters for an ideal solution is that the harnessing device could be used in a wide range of environments including harsh and sealed environments (e.g. nuclear reactors), would be inexpensive to produce, would be safe for humans, and would have a minimal effect on the basic size, weight and other physical characteristics of the untethered device.

SUMMARY

[0006] The present invention pertains to an apparatus for an application. The apparatus comprises a core device preferably having an integrated circuit for the application. The apparatus comprises a power harvester (preferably a radio frequency (RF) energy harvester) connected to the core device to power the core device.

[0007] The present invention pertains to an apparatus for an application. The apparatus comprises a core device preferably having an integrated circuit for the application. The apparatus comprises means for receiving energy wirelessly and providing power from the energy to the core device to power the integrated circuit of the core device. The receiving means is connected to the core device.

[0008] The present invention pertains to a method for an application. The method comprises the steps of converting RF energy into usable energy. There is the step of preferably powering an integrated circuit of the core device with the usable energy.

[0009] This invention pertains to a technique that uses radio frequency (RF) energy as a source of energy to directly power a device or augment a power storage component in a device. The present invention meets the requirements described in the previous "Background of the Invention" section.

[0010] Traditional RF receiving devices have focused on maximizing selectivity of the frequency to isolate and to be coherent without interference from other sources. In contrast, while the present invention operates at a specific frequency or range of frequencies, the device accepts any interference to supplement the output power of the device. Also, the research related to power harvesting that uses RF energy as a source has primarily focused on devices in close proximity of the source (inductive or near-field energy). In most cases, prior research assumed a dedicated or directed source of RF to power the device.

[0011] The invention should not be confused with power transfer by inductive coupling, which requires the device to be relatively close to the power transmission source. The RFID Handbook by the author Klaus Finkenzeller defines the inductive coupling region as distance between the transmitter and receiver of less than 0.16 times lambda where lambda is the wavelength of the RF wave. The invention can be implemented in the near field (sometimes referred to as inductive) region as well as the far-field region. The far-field region is distances greater than 0.16 times lambda.

[0012] It is an object of this invention to provide a method and apparatus to

[0013] 1. remotely energize an untethered device without using direct wiring

[0014] 2. power or augment the life of the power storage component so it matches the life of the device and, ultimately, powers the off-grid device with or without the use of batteries

[0015] 3. allow untethered devices to be virtually maintenance free

[0016] 4. provide augmentation for other energy harvesting technologies (solar, piezoelectric, etc.)

[0017] 5. provide backup power to tethered devices

[0018] It is a further object of this invention to directly power a device or augment a power storage component in a device in conjunction with other power harvesting technologies and storage elements.

[0019] With this method and apparatus a device's power storage components do not require replacement, thus enabling the device to be permanently placed off-grid, where it may be physically impractical, costly, or dangerous (due to a harsh environment) to provide maintenance.

[0020] For devices on-grid (tethered) or with reliable power sources, RF power harvesting can be used as a backup in case the primary power source is lost.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] In the accompanying drawings, the preferred embodiment of the invention and preferred methods of practicing the invention are illustrated in which:

[0022] FIG. **1** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

[0023] FIG. **2** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0024] FIG. **3** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0025] FIG. **4** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0026] FIG. **5** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0027] FIG. **6** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0028] FIG. **7** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0029] FIG. **8** is a block diagram of an RF Power Harvesting block in communication with a Power Storage block.

[0030] FIG. **9** is a block diagram of an RF Power Harvesting block in communication with a Power Storage block and used to supply power to Core Device Components.

[0031] FIG. **10** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0032] FIG. **11** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0033] FIG. **12** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block.

[0034] FIG. **13** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0035] FIG. **14** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

[0036] FIG. **15** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

[0037] FIG. **16** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0038] FIG. **17** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block.

[0039] FIG. **18** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger.

[0040] FIG. **19** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and the RF Power Harvesting block in communication with a Power Storage block.

[0041] FIG. **20** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and Core Device Components.

[0042] FIG. **21** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and Core Device Components and the RF Power Harvesting block in communication with a Power Storage block.

[0043] FIG. **22** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0044] FIG. **23** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0045] FIG. **24** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0046] FIG. **25** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0047] FIG. **26** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0048] FIG. **27** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0049] FIG. **28** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the RF Power Harvesting block supplying power to Core Device Components.

[0050] FIG. **29** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0051] FIG. **30** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

[0052] FIG. **31** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0053] FIG. **32** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power Core Device Components.

[0054] FIG. **33** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or

Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0055] FIG. **34** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0056] FIG. **35** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0057] FIG. **36** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0058] FIG. **37** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0059] FIG. **38** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0060] FIG. **39** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0061] FIG. **40** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to the Power Storage Charger.

[0062] FIG. **41** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0063] FIG. **42** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0064] FIG. **43** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

[0065] FIG. **44** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the RF Power Harvesting block supplying power to Core Device Components and used to supply power to a Power Storage Charger.

[0066] FIG. **45** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0067] FIG. **46** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0068] FIG. **47** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

[0069] FIG. **48** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power Core Device Components and used to supply power to a Power Storage Charger.

[0070] FIG. **49** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0071] FIG. **50** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

[0072] FIG. **51** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0073] FIG. **52** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0074] FIG. **53** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0075] FIG. **54** is a block diagram of an RF Power Harvesting block using Antenna A to directly supply power to Core Device Components.

[0076] FIG. **55** is a block diagram of an RF Power Harvesting block using Antenna A to supply power to a Power Regulation, Storage and/or Storage Charging block.

[0077] FIG. **56** is a block diagram of an RF Power Harvesting block using Antenna A to supply power to a Power Regulation, Storage and/or Storage Charging block and used to supply power to Core Device Components.

[0078] FIG. **57** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

[0079] FIG. **58** is a block diagram of an RF Power Harvesting block used to supply power to a Power Regulation, Storage and/or Storage Charging block.

[0080] FIG. **59** is a block diagram of an RF Power Harvesting block used to supply power to a Power Regulation, Storage and/or Storage Charging block and used to supply power to Core Device Components.

[0081] FIG. **60** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

[0082] FIG. **61** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0083] FIG. **62** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and used to directly supply power to Core Device Components.

[0084] FIG. **63** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0085] FIG. **64** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

[0086] FIG. **65** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

[0087] FIG. **66** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0088] FIG. **67** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

[0089] FIG. **68** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0090] FIG. **69** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0091] FIG. **70** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0092] FIG. **71** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0093] FIG. **72** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0094] FIG. **73** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0095] FIG. **74** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

[0096] FIG. **75** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

[0097] FIG. **76** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0098] FIG. **77** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

[0099] FIG. **78** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0100] FIG. **79** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0101] FIG. **80** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources.

[0102] FIG. **81** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

[0103] FIG. **82** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

[0104] FIG. **83** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0105] FIG. **84** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

[0106] FIG. **85** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0107] FIG. **86** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0108] FIG. **87** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0109] FIG. **88** is a block diagram of an entire power system for the device.

[0110] FIG. **89** is a block diagram of a power harvesting block used to supply power to a core device having a sensor.

[0111] FIG. **90** is a block diagram of a power harvesting block used to supply power to a core device having a computer peripheral.

DETAILED DESCRIPTION

[0112] A complete understanding of the invention will be obtained from the following description when taken in connection with the accompanying drawing figures wherein like reference characters identify like parts throughout.

[0113] There is shown an apparatus **10** for an application. The apparatus **10** comprises a core device **22** preferably having an integrated circuit **36** for the application. The apparatus **10** comprises a power harvester **20** connected to the core device **22** to power the core device **22**.

[0114] The apparatus **10** preferably includes an alternative power source **24** connected to the core device **22** to power the core device **22** in conjunction with the power harvester **20**. Preferably, the apparatus **10** includes a power regulator **26** and/or power storage circuit **28** connected to the power harvester **20**. The apparatus **10** preferably includes a power storage charger **30** connected to the power harvester **20**. Preferably, the apparatus **10** includes a power storage connected to the power harvester **20**.

[0115] Preferably, the core device **22** includes a memory **38** connected to the integrated circuit **36** and to the power harvester **20** to power the memory **36**.

[0116] The core device **22** can include a sensor **32**, as shown in FIG. **89**. The sensor **32** can include a proximity sensor, an intrusion sensor, an environmental sensor, a chemical sensor, a biological sensor, a sensor in contact with an automobile, an occupancy sensor, a motion sensor, a position sensor, a metal detector, or a sensor **32** in contact with an aircraft. The sensor **32** can include an alarm connected to the power harvester **20** to power the alarm, a display connected to the power harvester **20** to power the display, a sensor **32** disposed in a building, an industrial automation sensor, a sensor **32** in contact with an elevator, a temperature sensor, a fire sensor, or an accelerometer.

[0117] The sensor **32** can include a level sensor, a gas level sensor, a fluid level sensor, a light level sensor. The sensor **32** can include a flow sensor, a gas flow sensor, a fluid flow sensor, a light flow sensor, or a plasma flow sensor.

[0118] The sensor **32** can include a pressure sensor, a gas pressure sensor, a fluid pressure sensor, a light sensor, an infrared light sensor, an ultraviolet light sensor, an x-ray sensor, a cosmic ray sensor, a visible light sensor, a gamma ray sensor, a stress sensor, a strain sensor, a depth sensor, or an electrical characteristic sensor.

[0119] The sensor **32** includes a voltage sensor, a current sensor, a viscosity sensor, an acoustical sensor, a sound sensor, a listening sensor, a thickness sensor, a density sensor, a surface quality sensor, a volume sensor, a physical sensor, a mass sensor, a weight sensor, a conductivity sensor, a distance sensor, an orientation sensor, or a vibration sensor.

[0120] The sensor **32** can include a radioactivity sensor, a field strength sensor, an electric field sensor or a magnetic field sensor, a smoke detector, a carbon monoxide detector, a radon detector, an air quality sensor, a humidity sensor, a glass breakage sensor, or a break beam detector. The sensor can include a thermal energy sensor, an electromagnetic sensor, a mechanical sensor, an optical sensor, a radiation sensor, a sensor in contact with a vehicle, or a sensor **32** in contact with a water craft.

[0121] The present invention pertains to an apparatus **10** for an application. The apparatus **10** comprises a core device **22** having an integrated circuit **36** for the application. The apparatus **10** comprises means for receiving energy wirelessly and providing power from the energy to the core device **22** to power the integrated circuit **36** of the core device **22**. The receiving means is connected to the core device **22**. Preferably, the core device **22** includes means for sensing.

[0122] The present invention pertains to a method for an application. The method comprises the steps of converting RF energy into usable energy. There is the step of preferably powering an integrated circuit **36** of the core device **22** with the usable energy.

[0123] Preferably, there is the step of regulating the usable energy provided to the core device **22**. There is preferably the step of storing the usable energy. Preferably, there is the step of providing power to the core device **22** from an alternative power source **24** in conjunction with the usable energy.

[0124] Alternatively, the core device **22** can include a computer peripheral **34**, as shown in FIG. **90**. The computer peripheral **34** can include a handheld game, a gaming system, a game controller, a controller, a keyboard, a mouse, a computer terminal, computer storage, or computer equipment.

[0125] The present invention can be implemented in numerous ways. A number of these ways are depicted in FIGS. **1-90**. These figures contain multiple blocks that are configured in multiple ways.

[0126] In the figures, an arrow represents the flow of power unless otherwise stated. Single-headed (one-way) arrows represent that the power is flowing from one block to another. The single-headed arrow may represent multiple wires that provide power from one block to multiple parts in the other block. Two-headed (two-way) arrows represent a single wire that can have power flow in either direction or multiple wires each having power flow in a single direction.

[0127] As an example, a two-headed arrow between an RF power harvesting **20** block and a power regulation **26** and/or power storage circuit **28** block can represent a single wire that allows harvested power to flow into a storage device such as a capacitor. The same block diagram can also represent two wires between the two blocks with the first wire allowing harvested power to flow into a voltage regulator **26**. The second wire can allow the regulated voltage to feedback to the RF power harvesting **20** block to provide power to internal components such as transistors to increase the performance of the RF power harvesting **20** block.

[0128] Each block is described in detail below. Each block represents the functionality described below associated with it. For instance, the RF power harvesting **20** block describes a power harvester **20**, and the power regulation **26** and/or power storage circuit **28** block describes a power regulator **26** and a power storage circuit **28**.

Device's Power System Block

[0129] The device's power system block includes all components/circuitry required to power the device. This block may include an RF power harvesting **20** block, a power regulation **26** and/or power storage circuit **28** block, a power storage block, and/or a power storage charger **30** block (all discussed below).

RF Power Harvesting **20** Block

[0130] The RF power harvesting **20** block is connected to an antenna. The antenna may or may not be used as the communications antenna for the core device **22** components. The RF power harvesting **20** block is used to convert the energy captured by the antenna into usable power, such as direct current (DC) voltage.

[0131] This block may include antenna matching, rectifying circuitry, voltage transforming circuitry, and/or other performance optimizing circuitry. The rectifying circuitry may include a diode(s), a transistor(s), or some other rectifying device or combination. Examples of the rectifying circuitry include, but are not limited to, half-wave, full-wave, and voltage doubling circuits.

[0132] The output of the RF power harvesting **20** block is a DC voltage or current. The RF power harvesting **20** block may accept feedback (or input) from other circuitry or blocks, which may be used to control the harvesting circuitry to improve the performance or vary the output. This feedback may include, but is not limited to, a DC voltage or a clock from the core device **22** components.

[0133] U.S. Pat. No. 6,615,074 (FIGS. 8, 9, 12*a*, 12*b*, 13, 14), incorporated by reference, herein, shows numerous examples of RF power harvesting circuits that can be used to implement the block and function described.

Power Regulation **26** and/or Power Storage Circuit **28** Block

[0134] It may be necessary to regulate the converted power (i.e., hold the power at a constant level) for specific devices. The devices that would need this block require a fairly constant voltage or current. Deviations from the required values may cause the device to not perform within its specifications.

[0135] The regulation can be implemented in many different ways. The block can be as simple as using a Zener diode, or as complicated as using an integrated circuit such as a linear voltage regulator **26** or switching regulator **26** to hold the voltage at a constant level.

[0136] Certain devices have a more tolerable power requirement. For these devices, the regulation stage may be excluded.

[0137] This block may also include, with or without the regulation, a storage device such as a capacitor, a battery, or some other device able to store charge. The output from the power regulation **26** and/or power storage circuit **28** block may be used as feedback to other blocks within the device's power system block or to the alternative power sources **24** block (described below), if they require a regulated supply voltage or stored power.

[0138] U.S. Pat. No. 6,894,467 (FIGS. 1, 3), Linear Voltage Regulator, incorporated by reference, herein, is an example of a practical application of implementing the regulation described in the block. U.S. Pat. No. 6,297,618 (FIGS. 1-4), Power storage device and method of measuring voltage of storage battery, incorporated by reference, herein, is an example of a practical application of implementing the storage described in the block.

Power Storage Charger **30** Block

[0139] The power storage charger **30** block may be needed if a power storage component of the device requires a special charging mechanism, such as pulse charging or trickle charging. This block controls how the captured and converted power is supplied to the power storage component.

[0140] U.S. Pat. No. 6,836,095 (FIGS. 1-3), Battery Charging Method and Apparatus, incorporated by reference herein, is an example of a practical application of implementing the special charging mechanism described in the block.

Power Storage Block

[0141] If a device has intermittent power requirements, it may be necessary to store the captured power for use at a later time. The power can be stored in the power storage block, which could include a battery, a capacitor, and/or another type of power storage component. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other power storage elements.

[0142] The output from the power storage block may be used as feedback to other blocks within the device's power system or to the alternative power sources **24** block, if they require a dedicated and predictable supply voltage.

[0143] U.S. Pat. No. 6,297,618 (FIGS. 1-4), Power Storage Device and Method of Measuring Voltage of Storage Battery, incorporated by reference herein, is an example of a practical application of implementing the storage described in the block. U.S. Pat. No. 6,835,501, Alkaline Rechargeable Battery, incorporated by reference herein, is also an example of a practical application of implementing the storage described in the block.

Core Device **22** Components Block

[0144] The core device **22** components block is the portion of the device that is receiving power from the device's power system. This block may be, but is not limited to, the devices listed in the subsequent pages of this document. It may be advantageous for the core device **22** components block to communicate with any of the blocks that are supplying power to it. This communication can include, but is not limited to, a feedback control signal such as a clock or an ON/OFF command. As an example, the device may want to turn off the alternative power sources **24** block if it is receiving sufficient power from the RF power harvesting **20** block.

Alternative Power Sources **24** Block

[0145] RF energy harvesting also has the ability to be augmented by other types of power harvesting, storage components, or dedicated sources (e.g. power line). The alternative power sources **24** block shows how this type of system could be implemented. The augmenting power harvesting technologies include, but are not limited to, solar, light (visible and non-visible), piezoelectric, vibration, acoustic, thermal, microgenerators, wind, and other environmental elements. This block can work independently or have communication with other blocks.

[0146] U.S. Pat. No. 6,784,358, Solar Cell Structure Utilizing an Amorphous Silicon Discrete By-Pass Diode, incorporated by reference herein, is an example of a practical application of implementing an alternative power source **24** described by the block. U.S. Pat. No. 6,858,970, Multi-Frequency Piezoelectric Energy Harvester, incorporated by reference herein, is also an example of a practical application of implementing an alternative power source **24** described by the block.

Power Regulation **26**, Storage **28** and/or Storage Charging **30** Block

[0147] The power regulation **26**, storage **28** and/or storage charging **30** block contains all the combinations of the power regulation **26** and/or power storage circuit **28** block, power storage charger **30** block, and power storage block. This block is used in the later figures to reduce the number of figures needed to show how the blocks can interconnect.

[0148] The disclosed invention is an apparatus and method for an application for retrieving radio frequency (RF) energy by an antenna, converting that energy into direct current (DC) power, regulating that energy using an optimized circuit, storing that energy in an optimized component, and/or supplying the power for a specific device.

Retrieval of RF Energy

[0149] The RF energy is retrieved from the environment by the use of the antenna. The antenna can be shared or standalone with respect to an antenna used for the device's wireless communication. FIGS. **54-56** show a device that has an antenna A for use by the RF power harvester **20** and an antenna B used for wireless one-way or two-way communication. FIGS. **57-59** show a device where the antenna is shared by both a device's communication module (within the core device **22** components) and the RF power harvester **20**. In terms of form factor, the antenna used by the apparatus **10** can be a separate component or integrated directly into the form factor of the device.

[0150] The antenna is able to capture two types of available RF energy. The first type of energy exists as ambient RF energy. This type of RF surrounds us in our day-to-day lives and is usually generated to carry one way or two-way combinations of voice, video and data communications. The sources that the antenna can harvest from include, but are not limited to, medium-frequency AM radio broadcast, very-high-frequency (VHF) FM radio broadcast and television broadcast, ultra-high-frequency (UHF) broadcast, cellular base stations, wireless data access points, super-high-frequency (SHF) frequencies, and the industrial, scientific, and medical (ISM) bands. These sources cover transmission frequencies from 300 kHz to 30 GHz.

[0151] The second type of energy available is directed RF energy. This type of RF energy is directed from a transmitter specifically designed to deliver RF energy for harvesting by the antenna. The transmitter can be configured as a standalone device or integrated into an existing device.

Conversion of the Energy into DC

[0152] The RF energy captured by the antenna must be converted into a useful form of energy for the specific device. This conversion is shown in block form in all FIGS. **1-90**) as the RF power harvesting **20** block. The most common form of useable energy is DC energy. To perform this conversion, the block includes circuitry to rectify the captured alternating current (AC) energy to create DC energy. The rectification in this block can be done with a diode(s), a transistor(s), or some other rectifying device or combination.

Regulation of the Energy

[0153] It may be necessary to regulate the converted power (hold the power at a constant level) for specific devices. FIGS. **2-7**, **10-17**, and **22-53** show how this regulation can be implemented using a power regulation **26** and/or power storage circuit **28** block. The devices that would need this block require a fairly constant voltage or current. Deviations from the required values may cause the device to not perform within its specifications. The regulation can be implemented in many different ways. The block can be as simple as using a Zener diode, or as complicated as using an integrated circuit such as a linear voltage regulator **26** or a switching regulator **26** to hold the voltage at a constant level. Certain devices have a more tolerable power requirement. For these devices, the regulation stage may be excluded.

Storage of the Energy

[0154] If a device has intermittent power requirements, such as the devices exampled by FIGS. **2-53**, it may be necessary to store the captured power for use at a later time. The power can be stored in the power storage block or the power regulation **26** and/or power storage circuit **28** block. Storage devices can include, but are not limited to, a battery, a capacitor, or another type of power storage component. In certain applications, it may be necessary to include additional circuitry that controls how the power is transferred to the storage device. The power storage charger **30** block is shown in FIGS. **18-53**. This may be needed if the power storage component requires a special charging mechanism such as pulse charging or trickle charging. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other storage elements. There are devices that will not require storage. These devices can run directly off of the converted power. These devices also may or may not require regulation of the captured power.

Supplying the Power

[0155] The captured DC power, which may or may not be regulated and/or stored, is supplied to the device, which is represented by the core device **22** components block in the figures. This may be a single connection or it may supply multiple parts of the device with power.

[0156] RF energy harvesting also has the ability to be augmented by other types of power harvesting or storage components. Other power harvesting technologies include, but are not limited to, solar, light (visible and non-visible), piezoelectric, vibration, acoustic, thermal, microgenerators, wind, and other environmental elements. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other storage elements. FIGS. **60-88** show how the alternative power sources **24** block can be connected to an RF energy harvesting system. These figure show how the RF energy harvesting **20** block and the alternative power sources **24** block can work independently or have communication with each other. The antenna configurations shown in FIGS. **54-59** are still applicable with the addition of an alternative power sources **24** block. These antenna configurations can be applied to FIGS. **60-88**.

[0157] RF energy harvesting also has the ability to provide a backup to devices on-grid (tethered) or with reliable power sources, which can be used in case the primary power source is lost. As an example, it may be mandated by regulations that a sensor has auxiliary power in case the primary supply is lost. It could be possible to use a rechargeable battery that obtains its charge from the primary supply when in operation. However, if the primary supply is lost for a time greater than the life of the rechargeable battery, the specification of uninterrupted power is not met. RF energy could be used to supply power to the described device while the primary supply is not available. The primary supply could include, but is not limited to, an on-grid connection, a generator, a battery, or other reliable power supply.

[0158] RF energy harvesting with or without alternative source augmentation is applicable to provide electric power directly or indirectly to a range of electronic components contained in any specific electrical or electronic device and includes, but is not limited to:

[0159] Passive electronic components, active electronic components;

[0160] Resistors, fixed resistors, variable resistors, thermistors, thyristor, thermocouple;

[0161] Capacitors, Electrolytic Capacitors, Tantalum Capacitors, Ceramic Capacitors, Multilayer Ceramic Capacitors, Polystyrene Film Capacitors, Electric Double Layer Capacitors (Super Capacitors), Polyester Film Capacitors, Polypropylene Capacitors, Mica Capacitors, Metallized Polyester Film Capacitors, Variable Capacitors ;

[0162] Diodes, Voltage regulation diodes, light-emitting diodes, organic light-emitting diodes, Variable capacitance diodes, Rectification diodes, Switching diodes, Regulation Diodes, Diode bridges, Schottky barrier diodes, tunnel diodes, PIN diodes, Zener diodes, Avalanche diodes, TVSs;

[0163] Integrated circuits, microcontroller unit (MCU), microprocessor unit (MPU), logic circuits, memory, printed circuits, circuit boards, printed wiring boards;

[0164] Transistors, MOSFETs, FETs, BJTs, JFETs, IGBTs, Relays, Antennas, semiconductors, conductors, inductors, relays, diacs, triacs, SCRs, MOVs;

[0165] Fuses, circuit breakers;

[0166] Batteries, Non-rechargeable batteries, rechargeable batteries, coin cell batteries, button cell batteries, alkaline batteries, lithium batteries, lithium ion batteries, lithium polymer batteries, NIMH batteries, NICAD batteries, Lead acid batteries, Zinc air batteries, Manganese Lithium batteries, Niobium Titanium Lithium batteries, Vanadium Pentoxide Lithium batteries, Carbon

Zinc batteries, Zinc Chloride batteries, Lithium Thionyl Chloride batteries, Manganese Dioxide batteries, Lithium Poly-Carbonmonofluoride batteries, Lithium Manganese Dioxide batteries, Lithium Chloride batteries, Lead Acid Calcium batteries, Lead Acid Tin batteries, Oxy Nickel batteries, Silver Oxide batteries, Magnesium batteries;

[0167] Inductors, Coils, High Frequency Coils, Toroidal Coils, Transformers, switches, chokes;

[0168] Motors, DC motors, stepper motor, AC motors, Fans;

[0169] Crystals, Oscillators, Clocks, Timers; and

[0170] Displays, LCDs, LED displays.

[0171] RF energy harvesting with or without alternative source augmentation is applicable across a range of markets and specific devices and includes but is not limited to:

[0172] Consumer electronics:

[0173] Electronic equipment, wired devices, battery powered devices, wireless communication devices, cell phones, telephones, phones, cordless phones, portable phones, Bluetooth devices, Bluetooth headsets, hands-free headsets, headsets, headphones, Wireless headsets, radios, AM/FM radios, shortwave radios, weather radios, Two-way radios, portable radio, lights, lanterns, portable lights, flashlights, nightlights, spotlights, search lights, calculators, graphing calculators, desk calculators, clocks, alarm clocks, wall clocks, desk clocks, travel clocks, watches, wristwatches, pocket watches, stop watches, timers, voice recorders, Dictaphones, laser pointers, power tools, cordless power tools, electronic razors, electric razors, handheld games, gaming systems, game controllers, wireless game controllers, remote controls, battery chargers, computers, portable computers, keyless entry, toys, toy guns, toy laser guns, games, microphones, musical instruments, musical effects processors, musical instrument tuners, metronomes, electronic chord charts, door openers, garage door openers, PDA, Cameras, Video recorders, Multimeter, electronic test equipment, hand-held electronics, portable electronics, wireless pens, sound generators, noise generators, language translators, electric toothbrushes, portable televisions, pagers, transceivers, toy vehicles, remote control vehicles, toy planes, remote control planes, pet containment systems, invisible fence pet sensors, memory backup, base station battery backups, appliance battery backups, uninterrupted power supplies, GPS devices, memory retention power supplies, metal detectors, stud finders, metal stud finders, stun guns, tazers, wearable devices, baby monitors, intercoms, doorbells, wireless doorbells, electronic office supplies, electronic staplers, radar jammers, radar detectors, digital scales, microfilm cassettes, video head testers, compasses, noise canceling headphones, air samplers, depth finders, barometers, weather measurement instruments, data transfer devices, automatic distress signaling unit, Wireless audio speakers, Satellite radios, Police scanners, Car navigation systems (GPS devices), Decorative lights, Christmas lights, garden lights, lawn lights, ornamental lights, porch lights;

[0174] Multi-media players: MP3, DVD, analog music players, CD players, tape players, digital music players, digital video players, minidisc; and

[0175] Computer: keyboards, mice, peripherals, computer equipment, electronic computers, computer storage, computer terminals;

[0176] Building/Industrial Automation:

[0177] Sensors: Position, elevator, temperature, fire, accelerometers, level, gas level, fluid level, light level, flow, gas flow, fluid flow, light flow, plasma flow, pressure, gas pressure, fluid pressure, motion, light, infrared light, ultraviolet light, X-rays, cosmic rays, visible light, gamma rays, chemical, stress, strain, depth, electrical characteristics, voltage, current, viscosity, acoustical, sound, listening, thickness, density, surface quality, volume, physical, mass, weight, force, conductivity, distance, orientation, vibration, radioactivity, field strength, electric field strength, magnetic field strength, occupancy, smoke detector, carbon monoxide detector, radon detector, air quality, humidity, glass breakage, break beam detector;

[0178] Controls: Position, elevator, temperature, fire, accelerometers, level, gas level, fluid level, light level, flow, gas flow, fluid flow, light flow, plasma flow, pressure, gas pressure, fluid pressure, motion, light, infrared light, ultraviolet light, X-rays, cosmic rays, visible light, gamma rays, chemical, stress, strain, depth, electrical characteristics, voltage, current, viscosity, acoustical, sound, listening, thickness, density, surface quality, volume, physical, mass, weight, force, conductivity, distance, orientation, vibration, radioactivity, field strength, electric field strength, magnetic field strength, occupancy, smoke detector, carbon monoxide detector, radon detector, air quality, humidity; and

[0179] Devices: Thermostats, light switches, door locks, smart-card door locks, lighting, emergency lighting, motion lighting, safety lighting, highway lighting, construction lighting, sign lighting, roadway sign lighting, construction sign lighting, automatic flushing units, automatic soap dispenser, automatic paper towel dispenser, automatic faucets, automatic door sensors, identification reader, fingerprint reader, credit card readers, card readers, valve actuators, gauges, analog gauges, digital gauges, fire extinguishers, wireless switches, remotely operated inspection equipment, gas/oil pipeline monitoring systems, robotic pipeline inspection gauges, “auto-reclosers” for electric power lines, sonar buoys, telemetry systems, electronic record tracking systems, robbery tracking devices, interrogators, programmers, emergency exit alarms, alarms, flood alarms, gas alarms, electronic entry systems, security keypads, silo transducers, data recorders, signal tracers, anti-static strap testers, radiosonde weather balloons, utilities load controllers, profilometers, noise cancellation equipment, infrared beacons;

[0180] Military/Government:

[0181] Tracking tags: Weapons, vehicle, soldier, gear/assets, staff, general population, security access badges;

[0182] Sensors: Proximity, intrusion, environmental, chemical/biological; and

[0183] Equipment: Battery charger, surveillance, card readers, identification reader, fingerprint reader, retinal scanners, satellites, rockets, space vehicles, search and rescue transponders (SARTs), emergency position-indicating rescue beacons (EPIRBs), emergency locator transmitters (ELTs), military radios, electronic toll collection systems, postal tracking systems, communica-

tions, thermal imaging, night vision, training targets, field medical equipment, house arrest monitors, laser tags, electronic parking meters, multiple integrated laser engagement system, munitions and mines, ship sensors;

[0184] Utility:

[0185] Gas consumption meters, water consumption meters, and electric consumption meters;

[0186] Logistics & Supply Chain Management:

[0187] Radio-frequency identification devices (RFID), RFID readers;

[0188] Tracking: Asset tags, cargo container location beacons, transponders, transceivers; and

[0189] Devices: Smart price tags, smart shelving, hand-held barcode scanner, barcode scanners, credit card readers, card readers, retail signage, hotel door locks;

[0190] Homeland Security:

[0191] Sensors: Occupancy, proximity, environmental, chemical/biological, motion, position; and

[0192] Metal detector wand;

[0193] Medical:

[0194] Implantable: cochlear implants, neural stimulators, pace makers, medication administration, defibrillator;

[0195] Body function monitors: pressure, temperature, respiration, blood oxygenation, insulin, hearing aid, pulse, EKG, heart, Holter;

[0196] Tracking tags: Patient, baby identification, assets, supplies, staff, medication, instruments;

[0197] Devices: Home healthcare equipment, ambulatory infusion pumps, blood analyzers, biofeedback systems, bone growth stimulators, thermometers, digital thermometers, stimulators, galvanic stimulators, muscle stimulators, pediatric scales; and

[0198] Agriculture—livestock tracking and asset tracking:

[0199] Tracking: livestock, asset, wildlife tracking devices; and

[0200] Equipment: cattle prods;

[0201] Automotive:

[0202] Automotive antennas, Automotive Audio Systems, Automotive Lighting, Automotive Video Systems, Computers, Processors, Controls, Switches, Electric Motors, Actuators, Ignition Systems, Starter Systems, Injection Systems, Powertrain Electronics, Radar Detectors, Proximity Detectors, Safety Systems, Security Systems, Sensors, Regulators, Distributors, Vehicle Control, Wiper Systems, Washer Systems, Radio, Video Systems, Entertainment Systems, Navigation Systems, GPS systems, Power Mirror Systems, Emission control systems;

[0203] Appliances:

[0204] Monitoring systems and control systems for major and small appliances including washing machines, dryers, refrigerators, freezers, coolers, air conditioners, humidifiers, dehumidifiers, air purifiers, air filters, fans, furnaces, water heaters, boilers, space heaters, sowing machines, ice makers, microwave ovens, convection ovens, ovens, toaster ovens, ranges, range hoods, cooktops, stoves, stovetops, crock pots, hot plates, dishwashers, garbage disposals, can openers, vacuum cleaners, blenders, mixers, food processors, irons, coffee makers, toasters, grills, hair dryers, electric tooth brushes, electric razors, electric drills, electric screwdrivers, chainsaws, lawnmowers, push mowers, riding mowers, trimmers, brush cutters, pruners, edgers, vending machines;

[0205] Ventilation, Heating, Air-Conditioning, and Commercial Refrigeration Equipment:

[0206] Monitoring systems, control systems;

[0207] Engine, Turbine, and Power Transmission Equipment;

[0208] Monitoring systems, control systems;

[0209] Other General Purpose Machinery Manufacturing:

[0210] Monitoring systems, control systems;

[0211] Telecommunications:

[0212] Monitoring systems, control systems;

[0213] Portable;

[0214] Aircraft:

[0215] Monitoring systems, control systems, actuator systems, sensors.

[0216] It should be noted that devices within a specific category may be applicable across multiple areas even if they are not specifically listed. (e.g., temperature sensors apply to Industrial and Building Automation and Appliances).

[0217] To retrofit or redesign the devices listed, it is possible to implement the described systems in numerous ways. It may be advantageous to leave the device design as-is including the existing power supply. As an example, a device may use non-rechargeable batteries to operate. The device will most likely have a protection circuit to prevent damage if the batteries are installed incorrectly. The protection mechanism is commonly a diode inline with the positive terminal of the battery. In this case, the RF power harvester **20** with or without the alternative power source **24** could be inserted, with an antenna, into the device. The power generated by the RF power harvester **20** (and the alternative power source **24**, if applicable) could be connected to the device after the protection mechanism described to avoid potential charging of a non-rechargeable battery.

[0218] Another way to configure the device's power system is to replace the non-rechargeable batteries with rechargeable batteries. In this instance, the output from the RF power harvester **20** (and alternative power source **24**) could be connected to either side of the protections device. If the connection is before the protection mechanism, the RF power harvester **20** (and the alternative power source **24**, if applicable) will recharge the battery and supply power to the device. If the connection is after the protection mechanism, the RF power harvester **20** (and the alternative power source **24**, if applicable) will supply power to the device and the battery will supply any extra power needed that could not be supplied by the RF power harvester **20** (and the alternative power source **24**, if applicable). It should be noted that the protection device in this case is not needed for proper operation. Its only function would be to protect the batteries from being installed incorrectly. An antenna could be contained inside or placed on the outside of the device.

[0219] Another configuration of the device's power system is to remove the existing batteries and install the RF power harvester **20** (and the alternative power source **24**) in the enclosure provided for the batteries. An antenna could be contained inside or placed on the outside of the device.

[0220] Yet another method of configuring the device's power system would be to reduce the number of batteries and replace them with the RF power harvester **20** (and the alternative power source **24**). In this case, the output from the system would be connected to the batteries in series or par-

allel depending on the original battery configuration. An antenna could be contained inside or placed on the outside of the device.

[0221] An additional option, would be to completely redesign the product and integrate the required circuitry and storage components into the device. This method is probably the most advantageous because it can fully take advantage of the benefits offered by the RF power harvester **20** (and the alternative power source **24**). An antenna could be contained inside or placed on the outside of the device.

[0222] If the RF power harvester **20** (and the alternative power source **24**) is used as a backup to the primary power supply, a switch could be implemented into the system in order to switch the RF power harvester **20** (and the alternative power source **24**) on when the primary source is lost. In this case, an antenna could be contained inside or placed on the outside of the device.

[0223] To show the flexibility of RF energy harvesting, several products were retrofitted to include an RF power harvester **20** (energy harvesting circuitry). These products include a wireless keyboard, a wall clock, and a desk calculator.

[0224] The wireless keyboard is an example of recharging and augmenting a battery to supply power to a device. This system is shown in FIG. **13**. The output from the regulation circuitry recharges the battery and supplies power to the keyboard. The battery is also used to supply power to the keyboard. The keyboard also includes a separate antenna for receiving power and for data communications. The antenna configuration can be seen in FIG. **55**.

[0225] The wall clock is an example of a direct powering system. The wall clock was retrofitted to include an RF power harvester **20** and the internal AA battery was removed. This system is shown in FIG. **2**. The wall clock did not need regulation, but did require a capacitor for storage to supply the pulse of power to move the second hand.

[0226] The calculator is an example of using RF energy harvesting with another energy harvesting technology. The calculator had an internal 1.5V coin cell battery and a small solar panel. The internal battery was removed, however, the solar panel was left intact. This system is shown in FIG. **60**. In this system, the calculator can receive power from both the solar panel and the RF power harvester **20** to eliminate the need for a battery.

[0227] As an additional example, an RF power harvester **20** (energy harvesting circuit) similar to the ones shown in U.S. Pat. No. 6,615,074 (FIGS. 8, 9, 12*a*, 12*b*, 13, 14), was connected in series with a 0.5V solar cell. Individually, the solar cell was able to provide 0.480V to a 10 kilo-ohm resistor, which was being used to simulate the core device **22** components. This corresponds to 23 microwatts. The RF power harvesting circuit by itself was able to provide 2.093V across the 10 kilo-ohm resistor when being supplied by 1 milliwatt of RF power. This corresponds to 438 microwatts. The two circuit outputs were then combined in series by connecting the output from the RF energy harvesting circuit to the ground of the solar cell. The output of the solar cell was then connected to the resistor. The other end of the resistor was connected to the ground of the RF energy harvesting circuit. The voltage across the resistor with the circuits connected as shown in FIG. **63** was 2.445V. This corresponds to 598 microwatts.

[0228] As can be seen, the combination of the two technologies produces a result higher than the addition of the individual powers. From this, it can be determined that the two technologies can cooperate in a way that produces favorable results.

[0229] In the example given, the solar cell produces current to supply the load and helps to bias the RF rectifying diodes, which allows the RF energy harvesting circuit to operate a higher efficiency. The solar cell also changes the impedance seen by the RF energy harvesting circuit, which produces a beneficial result. To be more specific, when examining the power output of the individual circuit (solar and RF power harvesting), the sum of the power captured by the individual circuits was 23 μW+438 μW=461 μW. However, when the two circuits are combined and are allowed to work in conjunction with one another, the output power becomes 598 μW. This result shows that combining the two power-harvesting technologies produces a 30 percent increase in the output power for this example. This same technique can be applied to multiple energy harvesting technologies to produce even greater output power. The equations for this example are shown below.

[0230] Individual Circuits

$$P_1=P_1+P_2+\ldots+P_N$$

[0231] Combined Circuits

$$P_C>P_1=P_1+P_2+\ldots+P_N$$

[0232] where P_1 is the sum of the individual output powers

[0233] P_C is the output of the combined circuit

[0234] P_1 is the output power from the first power harvesting technology

[0235] P_2 is the output power from the second power harvesting technology

[0236] P_N is the output power from the N^{th} power harvesting technology

[0237] N is the number of power harvesting technologies or circuits.

[0238] As can be seen by the previous examples, RF energy harvesting can be used alone or in conjunction with alternative power sources to power a wide range of devices. The addition of RF energy harvesting technology to the device allows for increased battery life, increased functionality, or the removal of the primary battery.

[0239] Although the invention has been described in detail in the foregoing embodiments for the purpose of illustration, it is to be understood that such detail is solely for that purpose and that variations can be made therein by those skilled in the art without departing from the spirit and scope of the invention except as it may be described by the following claims.

What is claimed is:

1. An apparatus, comprising:

an antenna configured to wirelessly receive energy from a source that is remote from the apparatus;

a power harvester electrically coupled to the antenna and configured to convert the energy received at the antenna to a DC voltage to produce a first power associated with the DC voltage;

a regulator coupled to the power harvester and configured to regulate the first power associated with the DC voltage to produce a regulated power;

an alternative power source configured to produce a second power different from the first power and the regulated power; and

a radio-frequency identification (RFID) device electrically coupled to the antenna and configured to wirelessly communicate with the source via the antenna, the RFID device electrically coupled to the alternative power source, the power harvester and the regulator, the RFID device including an integrated circuit configured to operate based on at least one of the first power received from the power harvester, the regulated power received from the regulator, or the second power received from the alternative power source.

2. The apparatus of claim **1**, wherein the alternative power source is a rechargeable battery.

3. The apparatus of claim **1**, wherein the antenna is configured to wirelessly receive energy from the source, which is located in a far-field region of the antenna.

4. The apparatus of claim **1**, wherein the alternative power source is electrically coupled to the power harvester and is configured to supply at least a portion of the second power to the power harvester to operate the power harvester.

5. The apparatus of claim **1**, wherein the alternative power source is configured to supply the second power to the RFID device when at least one of the first power or regulated power being supplied to the RFID device is below a threshold level.

6. The apparatus of claim **1**, wherein the alternative power source is a non-rechargeable battery, the apparatus further comprising:

a protection mechanism electrically coupled to the non-rechargeable battery and configured to prevent charging of the non-rechargeable battery when the RFID device is receiving at least one of the first power from the power harvester or the regulated power from the regulator.

7. The apparatus of claim **6**, wherein the protection mechanism is a diode electrically connected to a positive terminal of the non-rechargeable battery.

8. An apparatus, comprising:

an antenna configured to wirelessly receive energy from a source that is remote from the apparatus;

a power harvester electrically coupled to the antenna and configured to convert the energy received at the antenna to a DC voltage to produce a first power associated with the DC voltage;

an alternative power source electrically coupled to the power harvester and configured to produce a second power different from the first power, the alternative power source configured to supply the second power to the power harvester such that the power harvester operates, at least in part, based on the second power; and

an integrated circuit electrically coupled to the power harvester and configured to operate based, at least in part, on the first power received from the power harvester.

9. The apparatus of claim **8**, further comprising:

a regulator electrically coupled to the power harvester and the integrated circuit, and configured to regulate the first power from the power harvester to produce a regulated power, the integrated circuit configured to operate based on the regulated power received from the regulator.

10. The apparatus of claim **8**, wherein the alternative power source is electrically coupled to the integrated circuit such that at least a portion of the second power produced by the alternative power source is supplied to the integrated circuit to operate the integrated circuit.

11. The apparatus of claim **10**, wherein the alternative power source supplies a first portion of the second power to the power harvester, and a second portion of the second power to the integrated circuit when the first power received from the power harvester at the integrated circuit is below a threshold level.

12. The apparatus of claim **8**, wherein the alternative power source is a non-rechargeable battery, the apparatus further comprising:

a protection mechanism electrically coupled to the non-rechargeable battery and configured to prevent charging of the non-rechargeable battery when the integrated circuit is receiving the first power from the power harvester.

13. The apparatus of claim **12**, wherein the protection mechanism is a diode electrically connected to a positive terminal of the non-rechargeable battery.

14. The apparatus of claim **8**, wherein the alternative power source is a rechargeable battery.

15. The apparatus of claim **8**, wherein the antenna is configured to wirelessly receive energy from the source, which is located in a far-field region of the antenna.

16. An apparatus, comprising:

an antenna configured to wirelessly receive energy from a source that is remote from the apparatus;

a power harvester electrically coupled to the antenna and configured to convert the energy received at the antenna to a DC voltage to produce a first power associated with the DC voltage;

an alternative power source electrically coupled to the power harvester and configured to receive the first power from the power harvester, the alternative power source configured to produce a second power that includes the first power; and

an integrated circuit electrically coupled to the alternative power source and configured to operate based on the second power received from the alternative power source.

17. The apparatus of claim **16**, wherein the alternative power source is a solar cell connected in series with the power harvester.

18. The apparatus of claim **16**, wherein the alternative power source is a solar cell configured to bias the power harvester.

19. The apparatus of claim **16**, wherein the alternative power source is configured to produce a third power independent of the first power,

the second power produced by the alternative power source being greater than the sum of the first power and the third power.

20. The apparatus of claim **16**, wherein the antenna is configured to wirelessly receive energy from the source, which is located in a far-field region of the antenna.

* * * * *

US 20150365013A1

(19) **United States**

(12) **Patent Application Publication**

Hameed et al.

(10) **Pub. No.: US 2015/0365013 A1**

(43) **Pub. Date: Dec. 17, 2015**

(54) **RF-DC POWER CONVERTER**

(71) Applicant: **THE GOVERNORS OF THE UNIVERSITY OF ALBERTA**, Edmonton (CA)

(72) Inventors: **Zohaib Hameed**, Edmonton (CA); **Kambiz Moez**, Edmonton (CA)

(21) Appl. No.: **14/740,021**

(22) Filed: **Jun. 15, 2015**

Related U.S. Application Data

(60) Provisional application No. 62/011,957, filed on Jun. 13, 2014.

Publication Classification

(51) **Int. Cl.**
H02M 7/217 (2006.01)

(52) **U.S. Cl.**
CPC *H02M 7/217* (2013.01)

(57) **ABSTRACT**

A rectifier comprising a chain of transistors for RF-DC conversion. In order to compensate for the thresholds of the transistors, each transistor can be connected to a junction earlier or later in the chain. By using both p-type and n-type transistors in the same chain, the different types of transistors can be compensated in different directions allowing more transistors to be compensated. Additional transistors connected to the gates of transistors of the main chain can allow the transistors of the main chain to be forward compensated at one part of the input cycle and backward compensated in another part to minimize both the voltage threshold of the rectifier and the leakage current. The line for compensation of the voltage threshold during forward conduction can comprise a solid line or a transistor, and if a transistor is used it may be diode-connected.

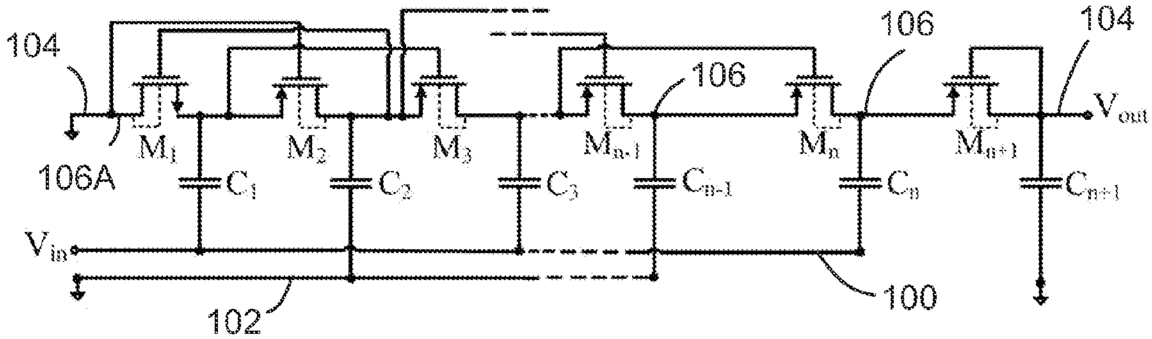

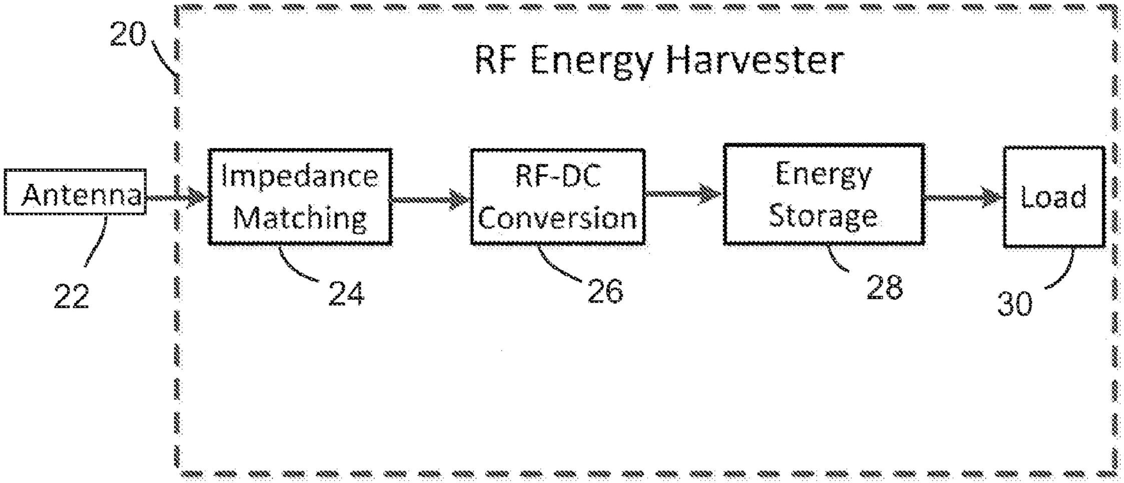
20
RF Energy Harvester
Antenna
Impedance Matching
RF-DC Conversion
Energy Storage
Load
22
24
26
28
30

Fig. 1 (Prior Art)

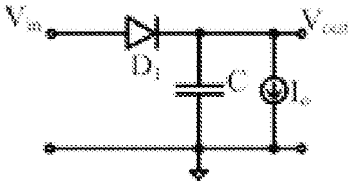
Vin
D1
C
Vout
Io

Fig. 2A (Prior Art)

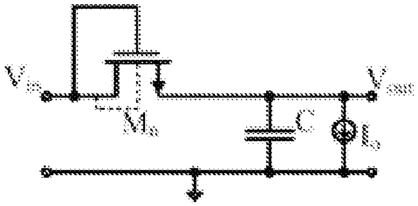
Vin
Mn
C
Vout
Io

Fig. 2B (Prior Art)

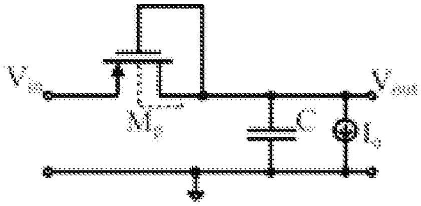
Vin
Mp
C
Vout
Io

Fig. 2C (Prior Art)

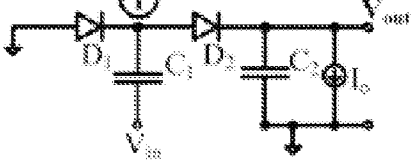
D1
C1
Vin
D2
C2
Vout
Io

Fig. 2D (Prior Art)

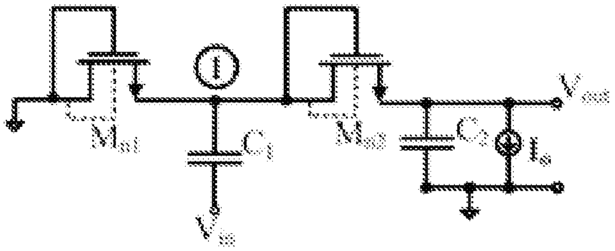
1
M_{n1}
C_1
V_{in}
M_{n2}
C_2
I_o
V_{out}

Fig. 2E (Prior Art)

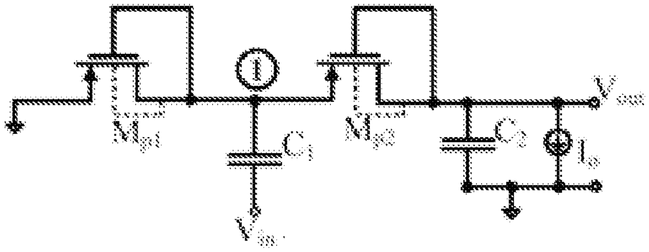
1
M_{p1}
C_1
V_{in}
M_{p2}
C_2
I_o
V_{out}

Fig. 2F (Prior Art)

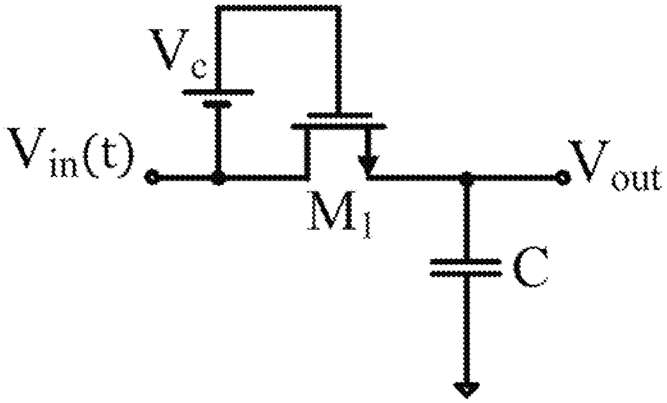
V_c
$V_{in}(t)$
M_1
V_{out}
C

Fig. 2G (Prior Art)

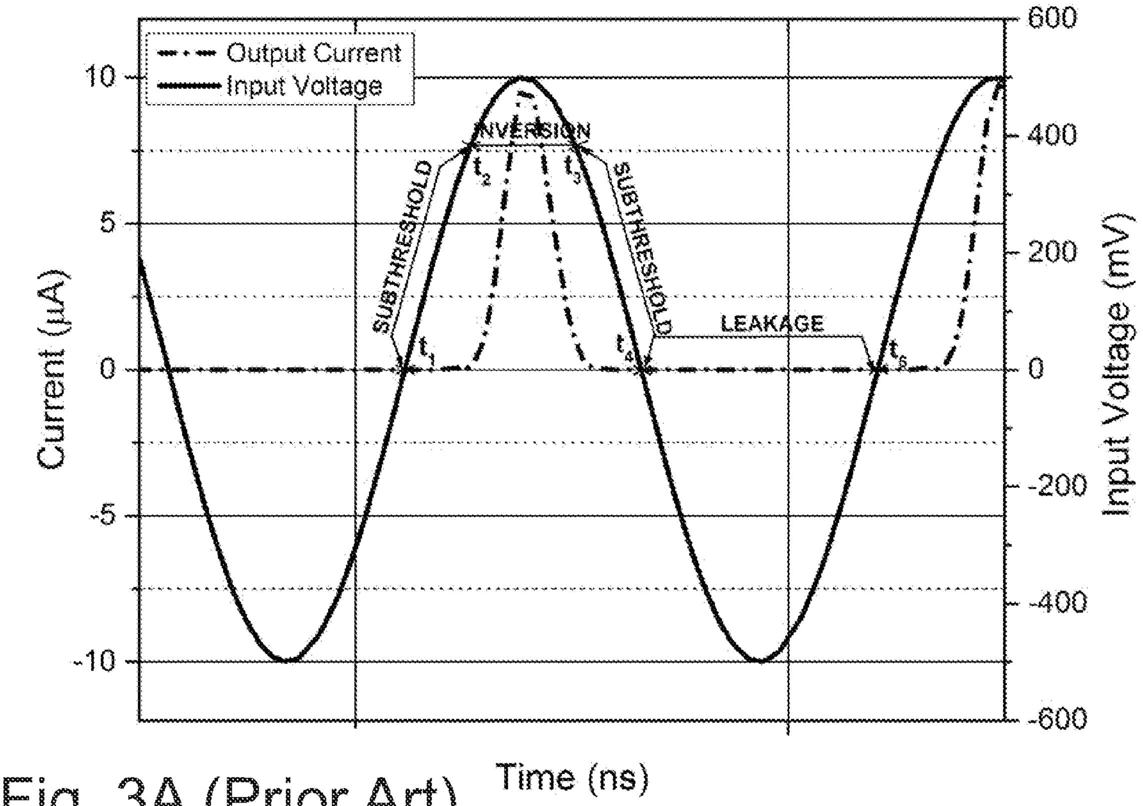

Fig. 3A (Prior Art)

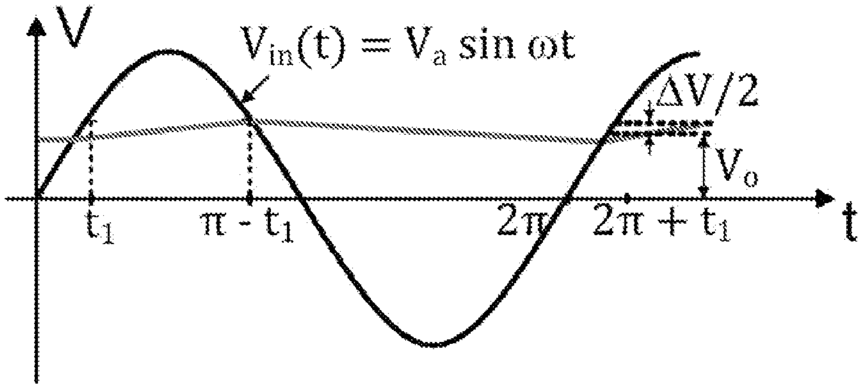
V
$V_{in}(t) = V_a \sin \omega t$
$\Delta V/2$
V_o
t_1
$\pi - t_1$
2π
$2\pi + t_1$
t

Fig. 3B (Prior Art)

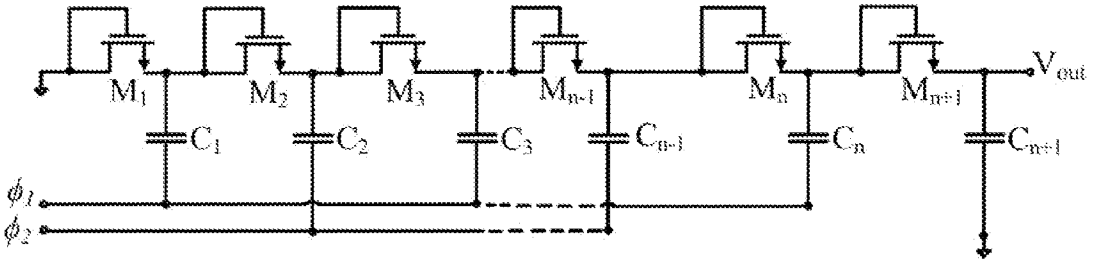
V_{out}
M_1
M_2
M_3
M_{n-1}
M_n
M_{n+1}
C_1
C_2
C_3
C_{n-1}
C_n
C_{n+1}
ϕ_1
ϕ_2

Fig. 4A (Prior Art)

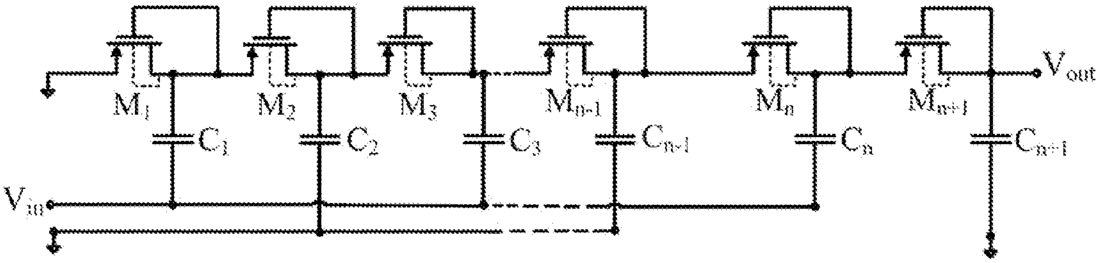
V_{out}
M_1
M_2
M_3
M_{n-1}
M_n
M_{n+1}
C_1
C_2
C_3
C_{n-1}
C_n
C_{n+1}
V_{in}

Fig. 4B (Prior Art)

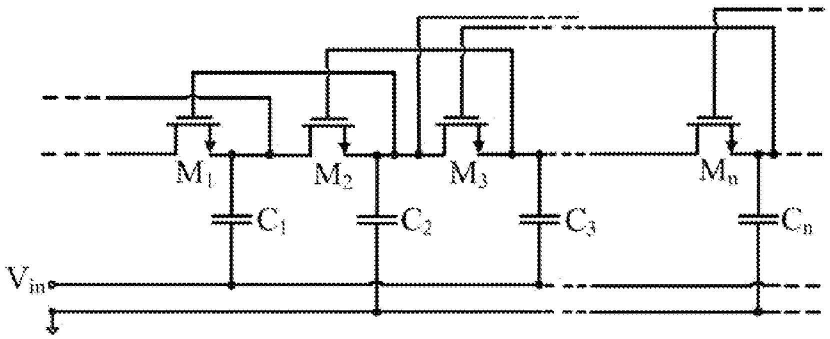
M$_1$
M$_2$
M$_3$
M$_n$
C$_1$
C$_2$
C$_3$
C$_n$
V$_{in}$

Fig. 5 (Prior Art)

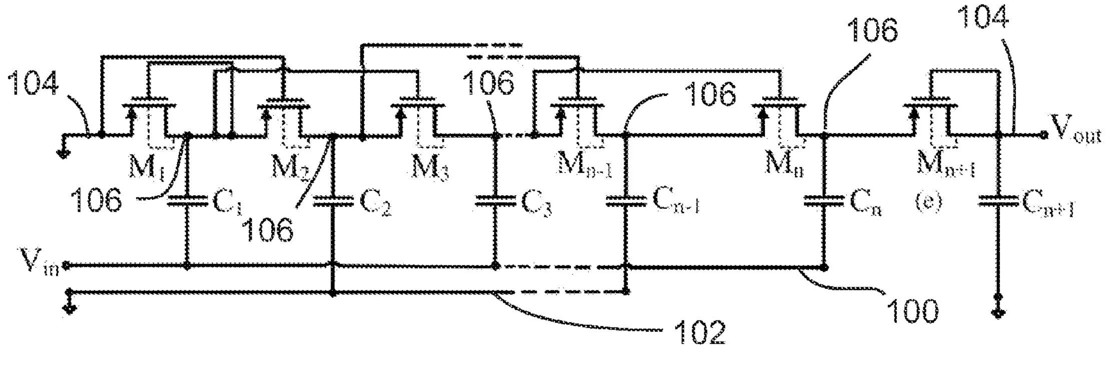
104
106
106
106
104
V$_{out}$
M$_1$
M$_2$
M$_3$
M$_{n-1}$
M$_n$
M$_{n+1}$
(e)
106
C$_1$
106
C$_2$
C$_3$
C$_{n-1}$
C$_n$
C$_{n+1}$
V$_{in}$
100
102

Fig. 6A

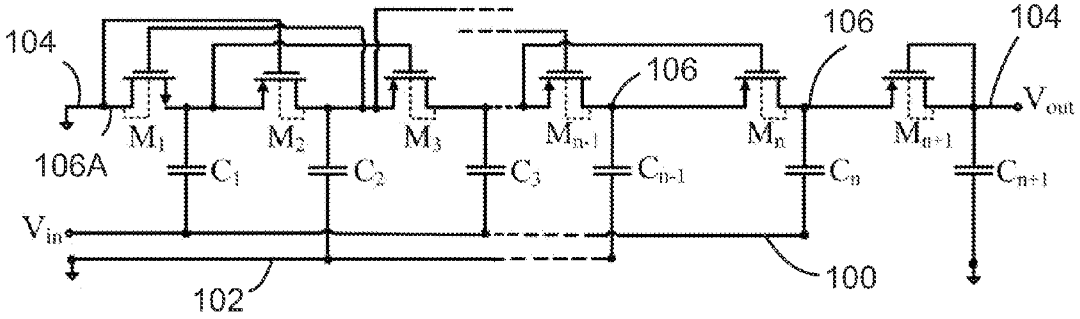
104
106A
106
106
104
M_1
M_2
M_3
M_{n-1}
M_n
M_{n+1}
C_1
C_2
C_3
C_{n-1}
C_n
C_{n+1}
V_{in}
V_{out}
102
100

Fig. 6B

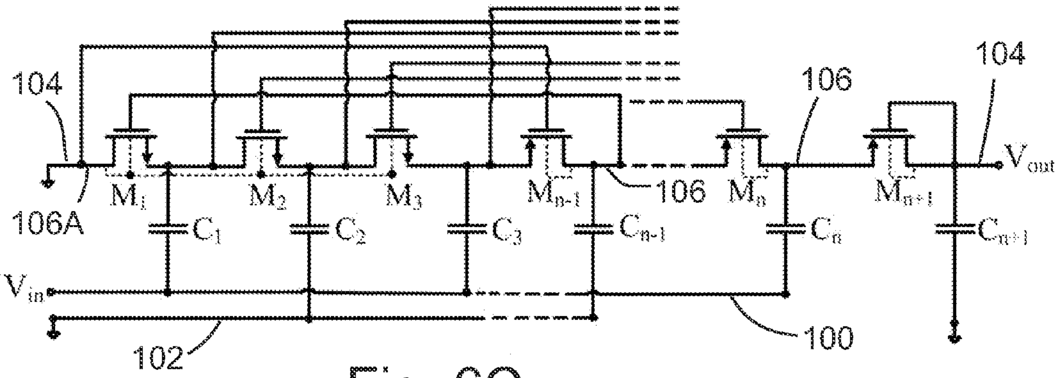
104
106A
106
106
104
M_1
M_2
M_3
M_{n-1}
M_n
M_{n+1}
C_1
C_2
C_3
C_{n-1}
C_n
C_{n+1}
V_{in}
V_{out}
102
100

Fig. 6C

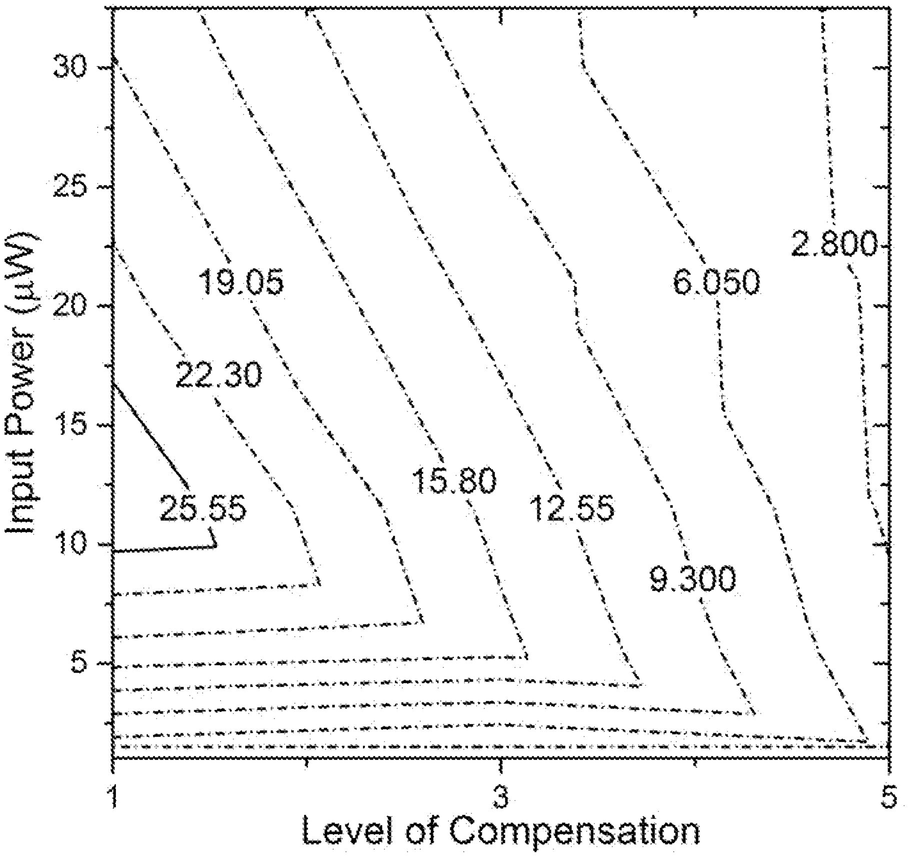
30
25
20
15
10
5
Input Power (μW)
1
3
5
Level of Compensation
19.05
22.30
25.55
15.80
12.55
9.300
6.050
2.800

Fig. 7

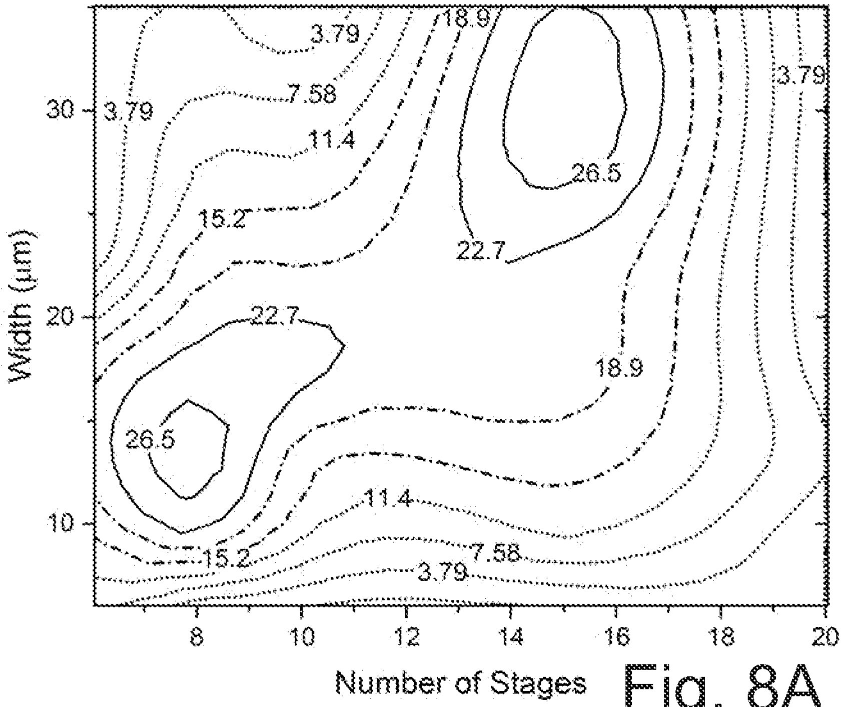
3.79
18.9
3.79
7.58
3.79
11.4
26.5
15.2
22.7
22.7
18.9
26.5
11.4
7.58
15.2
3.79
30
20
10
Width (μm)
8
10
12
14
16
18
20
Number of Stages

Fig. 8A

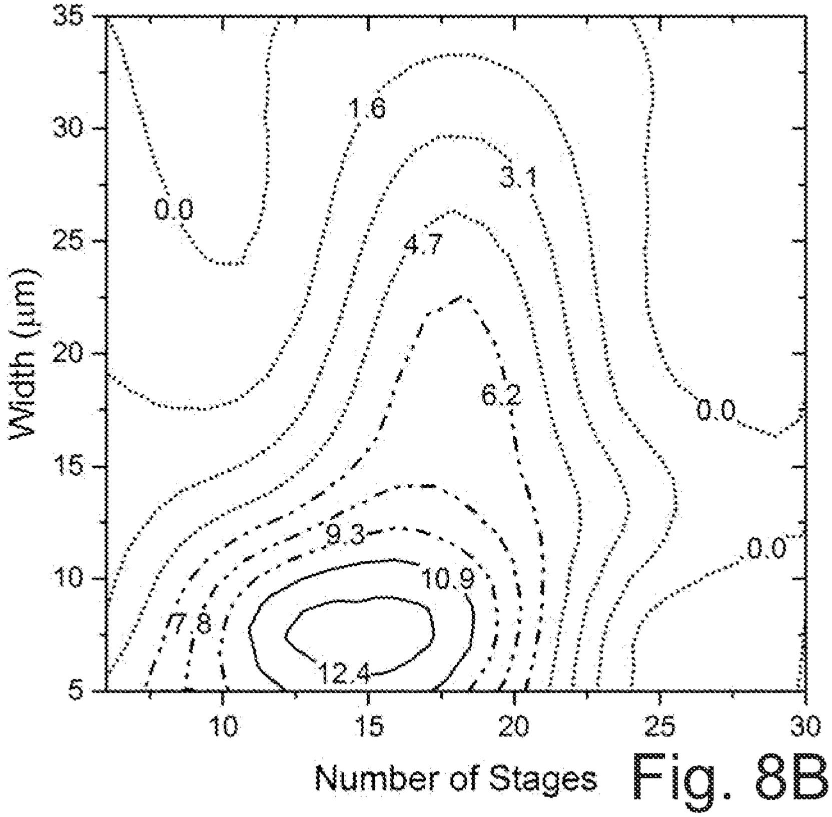
35
30
25
20
15
10
5
1.6
0.0
3.1
4.7
6.2
0.0
9.3
0.0
10.9
7.8
12.4
Width (μm)
10
15
20
25
30
Number of Stages

Fig. 8B

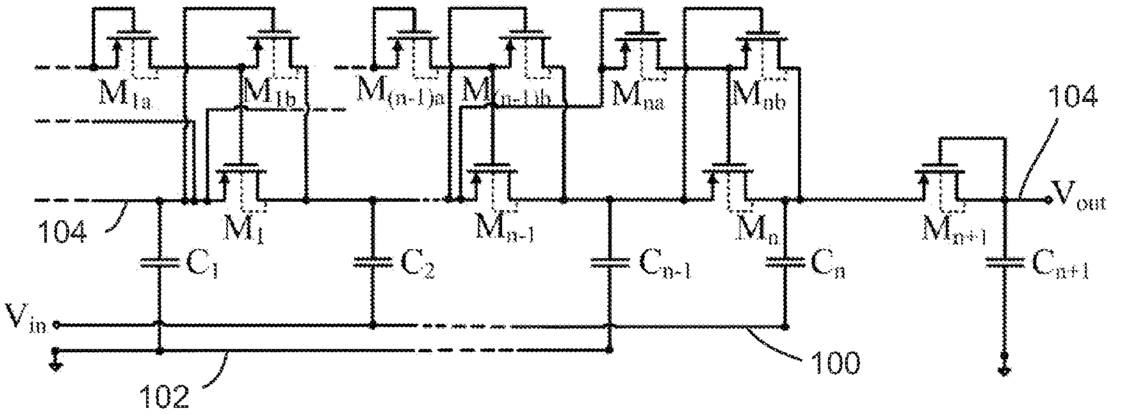

M_{1a}
M_{1b}
$M_{(n-1)a}$
$M_{(n-1)b}$
M_{na}
M_{nb}
104
V_{out}
104
M_1
M_{n-1}
M_n
M_{n+1}
C_1
C_2
C_{n-1}
C_n
C_{n+1}
V_{in}
100
102

Fig. 9A

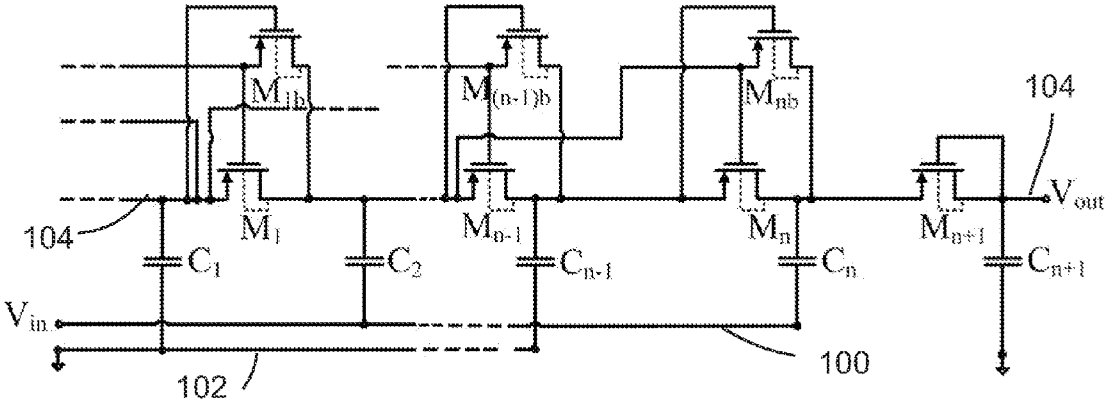

M_{1b}
$M_{(n-1)b}$
M_{nb}
104
V_{out}
104
M_1
M_{n-1}
M_n
M_{n+1}
C_1
C_2
C_{n-1}
C_n
C_{n+1}
V_{in}
102
100

Fig. 9B

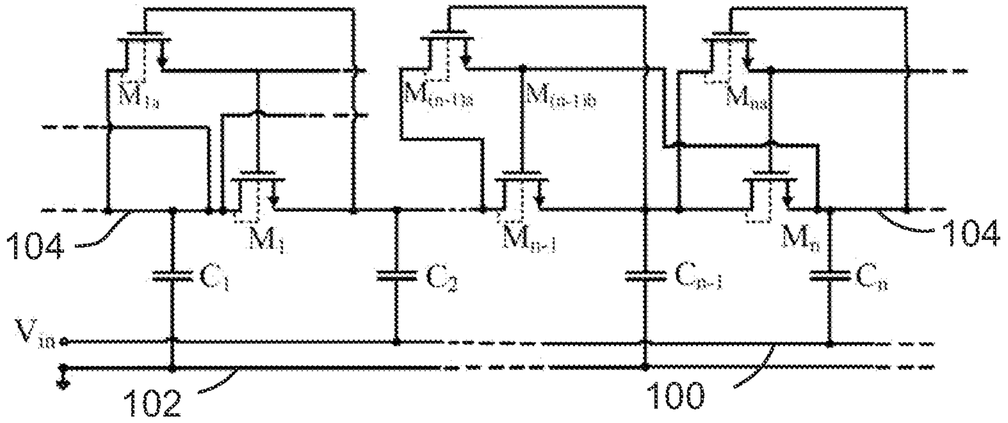
M_{1a}
$M_{(n-1)a}$
$M_{(n-1)b}$
M_{na}
M_1
M_{n-1}
M_n
104
104
C_1
C_2
C_{n-1}
C_n
V_{in}
102
100

Fig. 9C

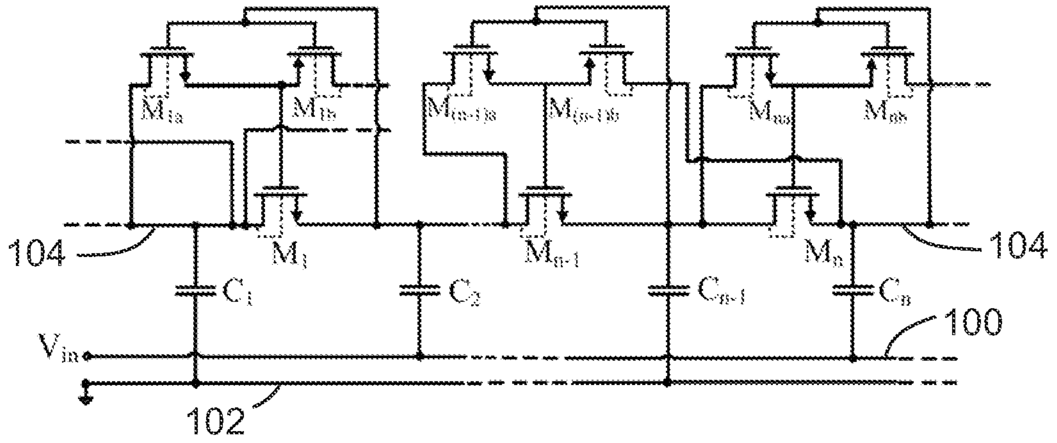
M_{1a}
M_{1b}
$M_{(n-1)a}$
$M_{(n-1)b}$
M_{na}
M_{nb}
M_1
M_{n-1}
M_n
104
104
C_1
C_2
C_{n-1}
C_n
100
V_{in}
102

Fig. 9D

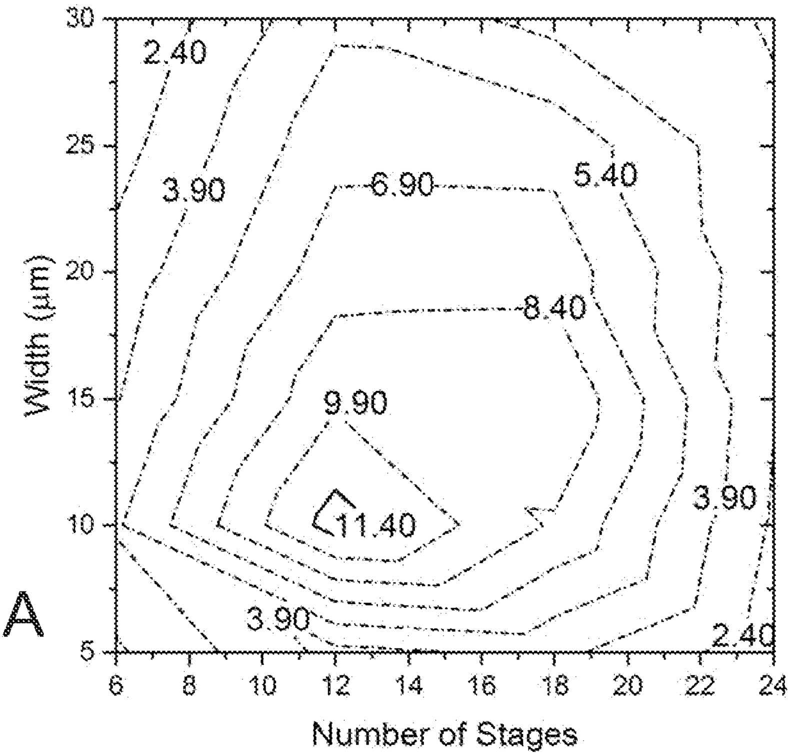

Fig. 10A

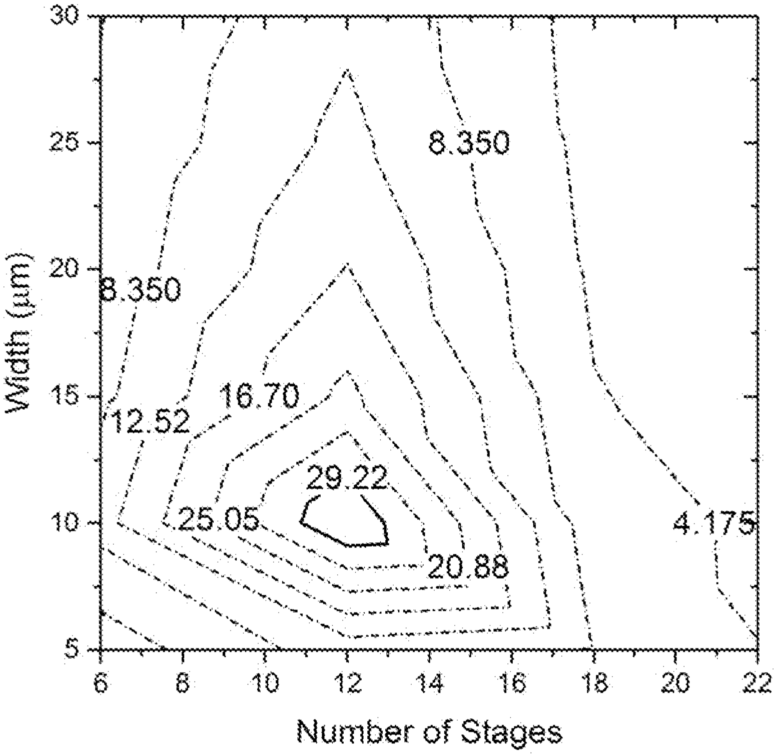

Fig. 10B

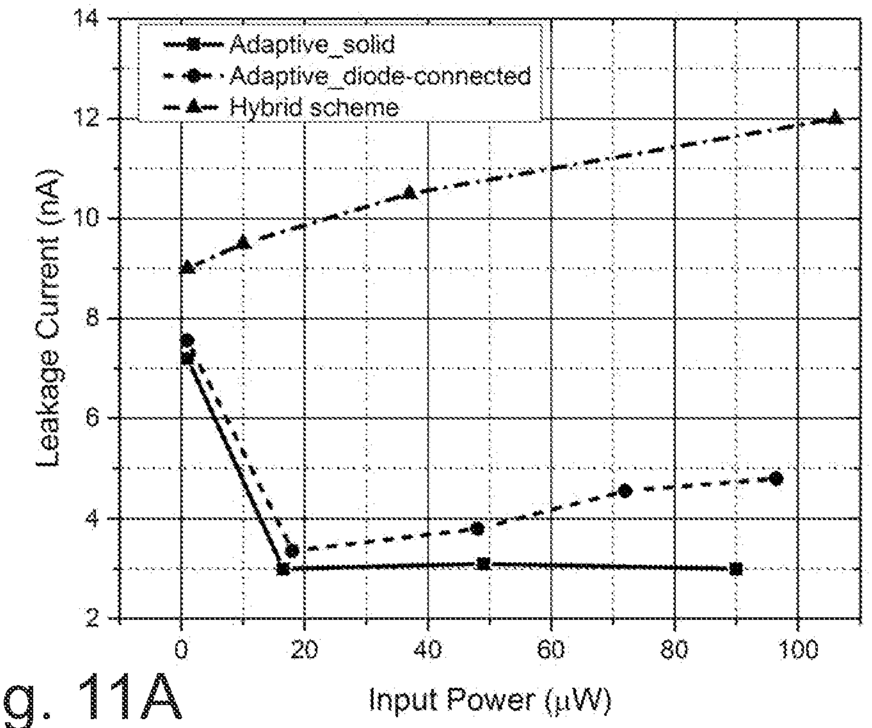

Fig. 11A

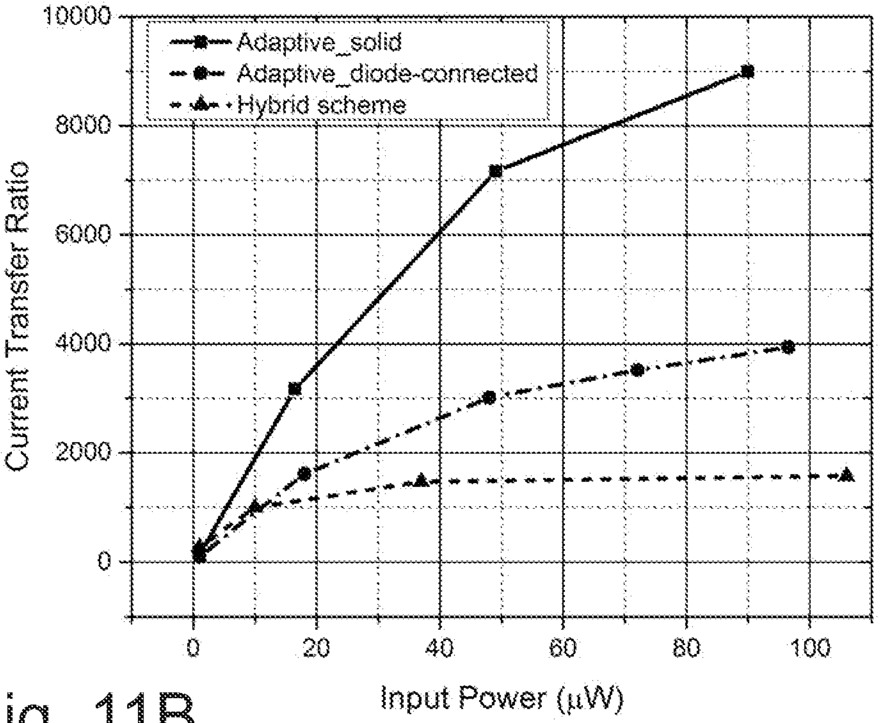

Fig. 11B

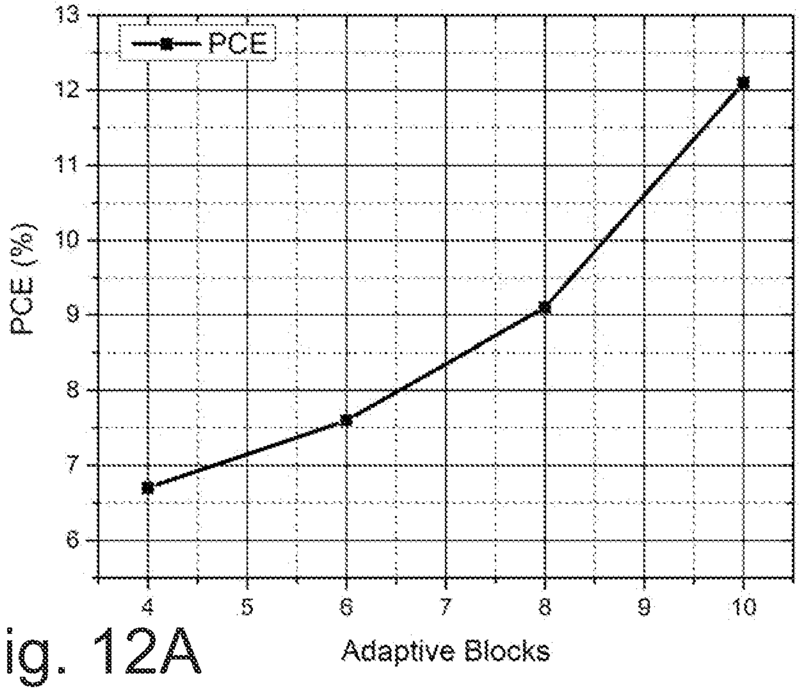

Fig. 12A

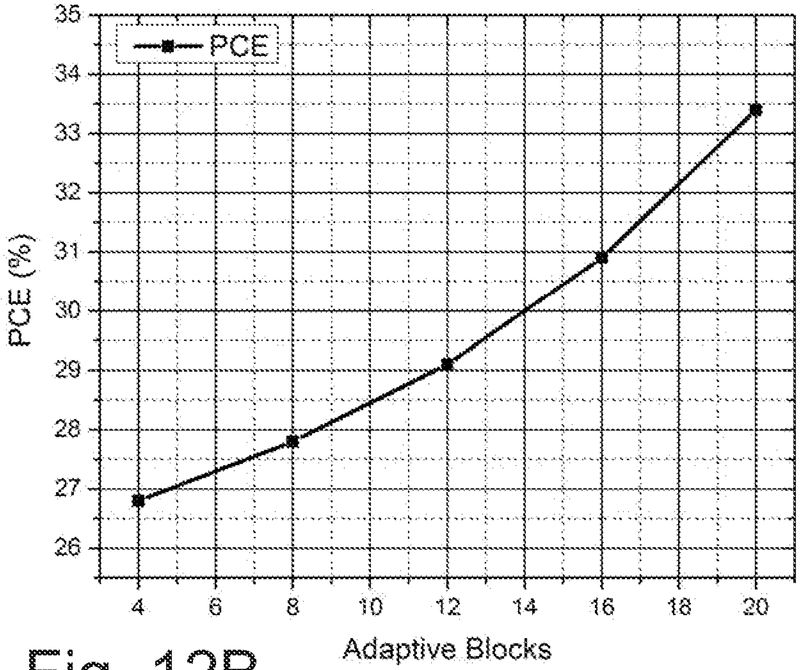

Fig. 12B

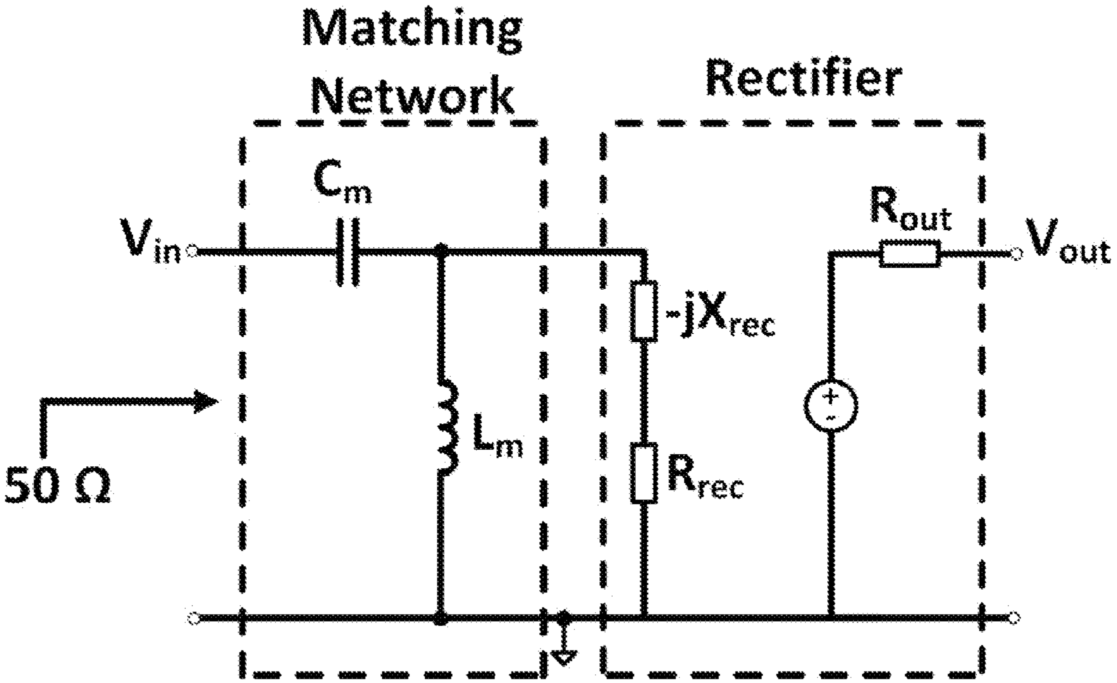
Matching Network
Rectifier
C_m
V_{in}
L_m
50 Ω
$-jX_{rec}$
R_{rec}
R_{out}
V_{out}
+
-

Fig. 13

RF-DC POWER CONVERTER

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit under 35 USC 119(e) of U.S. provisional application Ser. No. 62/011,957 filed Jun. 13, 2014.

TECHNICAL FIELD

[0002] Rectification, particularly for RF energy harvesting.

BACKGROUND

[0003] The vision of realizing Internet of Things (IoT) pervasively connecting large number of sensors and devices requires development of novel solutions that scavenges ambient energy to supply the power required for the operation of the sensors. Relying on batteries as the source of energy for wireless sensors impose several limitations including the need for routine maintenance/charging of batteries, operation interruption and cost involved in replacing batteries specially those employed in harsh environments, and challenges of scaling of battery-powered wireless sensors to millions of nodes. RF energy from ambient sources can be used to power efficiently the sensor networks with or without batteries. In addition, they can also be used to partially/fully charge other portable electronic devices to extend their battery life.

[0004] There is accordingly a growing interest in harvesting ambient energy to partially/fully supply the energy required for the operation of portable electronic devices. Scavenging energy from the ambient electromagnetic wave referred as RF energy harvesting is one of the most popular method for powering low-power wireless sensors. As most of today's integrated circuits are fabricated in CMOS technology, it is highly desirable to integrate RF energy harvesting systems with the rest of the system on a single CMOS chip for reduced cost and small form factor. RF powered devices are widely used in wireless sensor networks operating at power levels of 10 to 100's of μW. RF energy harvesting can be used in applications where the traditional energy sources such as light, vibrations are not available.

[0005] In an RF energy harvesting system **20** as shown in FIG. **1**, an antenna **22** receives the incident RF signal, an impedance matching circuit **24** maximizes the power transfer from the antenna to the power converter **26**, and a RF-DC power conversion circuit **26** converts the incident RF power to DC output power. The output DC voltage is stored in an energy storage component **28** (battery or capacitor) or can directly power the wireless sensors. The major challenge of scavenging RF energy is the limited signal strength of the RF waves and the low efficiency of the harvesting circuit at low input power.

[0006] The radiated signals are received by the RF energy harvesting system **20** and converted into a DC output voltage which is used to power-up the system. The radiated signals have to adhere to the regulation standards set by the federal communication commission (FCC). The unlicensed industrial, scientific and medical (ISM) band with a center frequency of 915 MHz operating at 902-928 MHz with a maximum effective isotropic radiated power (EIRP) of 4 W can be used for RF energy harvesting. The major challenge of scavenging RF energy is the limited signal strength of the RF waves and the low efficiency of the harvesting circuit. The signal strength is limited due to the attenuation in the signal due to the free space loss. Also, the power at the transmitter level cannot be increased without violating the FCC guidelines. Hence modification has to be done at the receiver side. The power consumption for the commercial sensor network nodes vary based on the manufacturers and has been estimated by various authors to be around 10 to 100 μW depending on the sensing application and the radio protocol. A multivolt supply voltage is typically required for the operation of the sensor circuitry.

[0007] The power harvester unit **20** comprising a multistage rectifier **26** is a key component in RF energy harvesting systems. It converts the incoming weak RF signal into a DC voltage. The performance of the rectifier unit **26** can be evaluated based on its power conversion efficiency (PCE) which is the ratio of power delivered to the load to the input power, based on sensitivity i.e. the minimum input power required for production of a DC voltage at the output and finally based on output DC voltage levels. To increase rectifier's PCE, the energy losses such as those introduced by the non-zero ON resistances of the rectifying devices must be reduced. To increase sensitivity and output voltage levels, rectifying devices with lower threshold voltages are required. Hence, these performance parameters of the power harvester are strongly affected by the threshold voltage of the rectifying devices. The power level has to exceed the threshold voltage of the rectifying device to turn on the rectifier unit in the RF-DC power conversion circuit. The minimum power is referred as the power-up threshold of the system. In terms of voltage level, a minimum input voltage is required for the circuit operation and is referred as the dead-zone of the rectifier.

[0008] The performance of an RF energy harvesting system **20** is significantly affected by the threshold voltage of the rectifying device **26**, the voltage that is required to turn on the semiconductor devices used as rectifying devices. A rectifying device **26** with lower threshold voltage enables the operation of RF-DC power converter **26** at low input power levels significantly reducing the power-up threshold of the rectifiers, and increasing the output voltage level for the same input power. The threshold voltage of device can be reduced using different technology-based approaches for the devices including silicon-on-sapphire (SOS), schottky diodes such as silicon-titanium schottky diodes or SMS and the HSMS diodes, special low-threshold-voltage transistors in CMOS process, and floating gate transistors which store a precharged voltage at the gate to lower the threshold voltage. The drawback of using technology-based approaches is additional fabrication steps that increases the production cost and prevents integration of RF energy harvester in mainstream Complementary Metal-Oxide-Silicon Integrated Circuits (CMOS ICs). Active/passive circuit techniques can be alternatively used to reduce the threshold voltage of the device. Active technique requires an external power source or secondary battery and is generally used in active sensors or active RFID. This enables more sophisticated applications at the price of increased cost and maintenance. Passive techniques do not require an additional source of energy but may require additional circuit where an auxiliary rectification chain is used to generate the compensating threshold voltage for the main RF-DC power conversion circuit. The auxiliary chain requires additional power and occupies additional area. An internal threshold voltage cancellation circuit was introduced in Kocer et al. where the compensating voltage was generated passively and stored in a capacitor that is applied at the gate-source terminal of the MOS transistor. This technique requires a large resistance and high capacitance value which leads to a large area on the chip. A self-biasing technique consisting of an off-chip high impedance resistive network was used by Li et al. to provide DC biasing voltages. Another technique can be using floating gate transistors storing a pre-charged voltage at the gate lowering its threshold voltage. The threshold voltage can be varied by changing the body-source potential of the transistor. This technique requires an additional pre-charge phase making it unsuitable for fully battery-less applications. The RF-DC power conversion circuit consisting of NMOS transistors with grounded body terminal leads to an increase in the threshold with the number of stages due to the body effect. This degrades the efficiency of the power conversion circuit. The body terminal of the transistors can be dynamically controlled using additional circuit or floating well devices. The floating well technique generates undesirable substrate current. Also, a triple-well source-body connected device has high parasitic capacitance leading to reduced efficiency. A cross-coupled differential scheme was used in Sciorcioni et al. consisting of triple-well NMOS and standard PMOS transistors to reduce the threshold voltage. A self-compensation scheme based on the Dickson topology was introduced in Curty et al. consisting of triple-well NMOS transistors to provide individual body biasing. The compensating voltage was provided by connecting the gate terminal to later stages. The design of the power converter in the works was focused on only reducing the threshold voltage neglecting the key role played by the reverse leakage current in introducing power losses. The reduction in threshold voltage increases the reverse conduction current causing additional loss.

[0009] To address the trade-off between the reduced threshold voltage and increased leakage current, Lee et al. proposed use of high speed comparators to control the reverse leakage current. The use of comparator increases the power consumption and limits the usefulness of this technique to low-frequency applications. Dynamic CMOS Dickson pump has been designed in previous works to eliminate the V_{th} drop while reducing the reverse leakage current. These circuits are designed for digital application with no emphasis on low-power operation. Differential-drive (**4**T-cell) architecture with the cross coupled bridge configuration and its variant has been used in previous works to reduce the threshold voltage as well as lower the leakage current. Differential circuit requires a PCB balun for the single-ended to differential conversion or differential antenna. Also, the differential circuit requires triple-well NMOS transistors and larger number of rectifying devices for the same number of stages compared with single-ended one.

SUMMARY

[0010] A power conversion circuit is provided having a first input line and a second input line, the first and second input lines configured to receive an alternating voltage differential between the first and second input lines, a multi-stage rectifier comprising transistors arranged in series, each transistor having a gate, a source and a drain, adjacent transistors of the series being connected so that for adjacent p-type transistors the drain of the left p-type transistor is connected to the right adjacent source of the right p-type transistor and source of the right p-type transistor is connected to the left adjacent drain of the left p-type transistor, for adjacent n-type transistors the source of the left n-type transistor is connected to the right adjacent drain of the right n-type transistor and drain of the right n-type transistor is connected to the left adjacent source of the left n-type transistor, for a p-type transistor adjacent to an n-type transistor the source of the right p-type transistor is connected to the source of the left adjacent n-type transistor and drain of the left p-type transistor is connected to the drain of the right adjacent n-type transistor to form a junction, each junction being connected to one of the first input line and the second input line via a capacitor, with adjacent junctions having one junction of the adjacent junctions connected to the first input line and the other junction of the adjacent junctions connected to the second input line, the gate of each transistor of the multi-stage rectifier being connected to a respective junction that is not a junction formed by the connection of that transistor to an adjacent transistor but a junction in the previous stage for the p-type transistor or later stage for the n-type transistor, at least one of the transistor in the multi-stage rectifier being p-type and at least one being n-type.

[0011] In various embodiments, there may be included any one or more of the following features: There may also be an auxiliary chain of p-type transistors, each auxiliary transistor having a gate, a source and a drain, each transistor of the auxiliary chain being connected to a p-type transistor of the multi-stage rectifier in the main chain, so that the gate of the respective auxiliary transistor is connected to the source of the respective transistor in the main chain, the source of the respective auxiliary transistor is connected to the gate of the transistor in the main chain and also connected to the previous stage junction of the main chain which is N transistors away from the respective transistor, and the drain of the respective auxiliary transistor is connected to the drain of the respective transistor or to the junction of the later stage transistor in the main chain. There may also be a diode connected transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of previous stages which is N transistors away from the respective transistor. There may also be an auxiliary chain of n-type transistors, each auxiliary transistor having a gate, a source and a drain, each transistor of the auxiliary chain being connected to an n-type transistor of the multi-stage rectifier in the main chain, so that the gate of the respective auxiliary transistor is connected to the source of the respective transistor in the main chain, the source of the respective auxiliary transistor is connected to the gate of the transistor in the main chain and also connected to the later stage junction of the main chain which is N transistors away from the respective transistor, and the drain of the respective auxiliary transistor is connected to the drain of the respective transistor or to the junction of the previous stage transistor in the main chain. There may also be a diode connected transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of later stages which is N transistors away from the respective transistor.

[0012] There is also provided an auxiliary chain of transistors, each auxiliary transistor having a gate, a source and a drain, each transistor of the auxiliary chain being connected to a respective transistor of a main chain of a multi-stage rectifier in the main chain, so that the gate of the respective auxiliary transistor is connected to the source of the respective transistor in the main chain, the source of the respective auxiliary transistor is connected to the gate of the transistor in the main chain and also connected to the later stage junction of the main chain which is N transistors away from the respective transistor in the case that the respective transistor is n-type, or to the earlier stage junction of the main chain which is N transistors away from the respective transistor in the case that the respective transistor is p-type, and the drain of the respective auxiliary transistor is connected to the drain of the respective transistor or to the junction of a previous stage transistor in the main chain, in the case that the respective transistor is n-type, or to the junction of a later stage transistor in the main chain, in the case that the respective transistor is p-type.

[0013] In various embodiments, there may be included any one or more of the following features: the transistors of the main chain and auxiliary chain may be n-type and there may also be a p-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of later stages, such that the gate of the additional p-type auxiliary transistor is connected to the gate of the n-type auxiliary transistor and the source of the n-type and p-type auxiliary transistors are connected, and the drain of the p-type auxiliary transistors is connected to junction of the main chain which is at 'N' later stages from the respective transistor. The transistors of the main chain and auxiliary chain may be p-type and there may also be an n-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of earlier stages, such that the gate of the additional n-type auxiliary transistor is connected to the gate of the p-type auxiliary transistor and the source of the n-type and p-type auxiliary transistors are connected, and the drain of the n-type auxiliary transistors is connected to junction of the main chain which is at 'N' earlier stages from the respective transistor. There may also be a diode connected p-type or n-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of later stages which is N transistors away from the respective transistor. There may also be a diode connected p-type or n-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of earlier stages which is N transistors away from the respective transistor.

[0014] In any of the above embodiments, there may be included any one or more of the following features: There may also be an additional transistor connected in series with the multistage rectifier, the additional transistor having a gate, a source and a drain, the gate of the additional transistor being connected to the drain of the additional transistor. The gate of each transistor in the main chain may be connected to the junction of previous stages for p-type transistor and later stages for n-type transistor. N of the transistors in the multistage rectifier arranged in series may be n-type and all but N of the transistors in multistage rectifier arranged in series may be p-type. The second input line may be grounded. The body terminal of each p-type transistor may be connected to the respective drain and the body terminal of each n-type transistor may be either grounded or connected to the respective drain terminal.

[0015] These and other aspects of the device and method are set out in the claims, which are incorporated here by reference.

BRIEF DESCRIPTION OF THE FIGURES

[0016] Embodiments will now be described with reference to the figures, in which like reference characters denote like elements, by way of example, and in which:

[0017] FIG. **1** is a block diagram showing energy flow from a receiver through a power harvester to a load;

[0018] FIG. **2**A is a schematic diagram of a prior art half-wave rectifier using a diode;

[0019] FIG. **2**B is a schematic diagram of a prior art half-wave rectifier using diode-connected NMOS;

[0020] FIG. **2**C is a schematic diagram of a prior art half-wave rectifier using diode connected PMOS;

[0021] FIG. **2**D is a schematic diagram of a prior art voltage doubler full wave rectifier using diodes;

[0022] FIG. **2**E is a schematic diagram of a prior art voltage doubler full wave rectifier using diode connected NMOS;

[0023] FIG. **2**F is a schematic diagram of a prior art voltage doubler full wave rectifier using diode connected PMOS;

[0024] FIG. **2**G is a schematic diagram of a prior art threshold-compensated transistor in a single-stage rectifier;

[0025] FIG. **3**A is a graph of output current with respect to input voltage for a prior art PMOS voltage doubler as shown in FIG. **2**F;

[0026] FIG. **3**B is a graph showing input and output voltage over time of the threshold-compensated transistor of FIG. **2**G;

[0027] FIG. **4**A is a schematic diagram of a prior art NMOS Dickson charge multiplier;

[0028] FIG. **4**B is a schematic diagram of a prior art PMOS Dickson charge multiplier;

[0029] FIG. **5** is a schematic diagram of a prior art forward compensated charge multiplier;

[0030] FIG. **6**A is a schematic diagram of a level-1 back compensated PMOS charge multiplier;

[0031] FIG. **6**B is a schematic diagram of a level-1 hybrid forward and back-compensated multiplier using NMOS and PMOS;

[0032] FIG. **6**C is a schematic diagram of a level-3 hybrid forward and back-compensated multiplier using NMOS and PMOS;

[0033] FIG. **7** is a contour plot of constant efficiency for input power versus level of compensation;

[0034] FIG. **8**A is a contour plot of constant efficiency for width versus number of stages for a level-1 compensated multiplier;

[0035] FIG. **8**B is a contour plot of constant efficiency for width versus number of stages for a level-3 compensated multiplier;

[0036] FIG. **9**A is a schematic diagram of a multiplier with an adaptive threshold-voltage compensation scheme using diode-connected transistors;

[0037] FIG. **9**B is a schematic diagram of a multiplier using p-type transistors with an adaptive threshold-voltage compensation scheme using solid-wired connections;

[0038] FIG. **9**C is a schematic diagram of a multiplier using n-type transistors with an adaptive threshold-voltage compensation scheme using solid wired connections;

[0039] FIG. **9**D is a schematic diagram of a multiplier using n-type transistors with an adaptive threshold compensation scheme using p-type and n-type transistors;

[0040] FIG. **10**A is a contour plot of constant efficiency for width versus number of stages for an adaptive diode-connected scheme as shown in FIG. **9**A at an operating frequency of 915 MHz;

[0041] FIG. **10**B is a contour plot of constant efficiency for width versus number of stages for an adaptive solid-wired scheme as shown in FIG. **9**B at an operating frequency of 915 MHz;

[0042] FIG. **11**A is a graph of leakage current versus input power with different schemes at 915 MHz;

[0043] FIG. **11**B is a graph of current transfer ratio versus input power with different schemes at 915 MHz;

[0044] FIG. **12**A is a graph of efficiency versus number of adaptive blocks for the adaptive diode-connected scheme;

[0045] FIG. **12**B is a graph of efficiency versus number of adaptive blocks for the adaptive solid-wired scheme; and

[0046] FIG. **13** is a schematic diagram showing an impedance matching circuit.

DETAILED DESCRIPTION

[0047] There is provided a hybrid forward and backward threshold voltage compensation scheme for RF-to-DC power conversion. In an embodiment, PMOS transistors are used as rectifying devices in all stages except for the first few stages eliminating the need for triple-well NMOS transistors. The compensating voltage is provided by connecting the gate terminal of the PMOS transistors to previous stages, also referred to as back-compensation, and also by connecting the gate terminal of the initial few NMOS transistors to later stages. Using PMOS transistors as a rectifying device in all the stages except for the first few stages eliminate the need for triple-well NMOS transistors. Also, all the MOS transistors in the main rectification chain are compensated in the proposed scheme. In addition to the hybrid scheme, an adaptive method suitable for rectifying devices operating at low power level is proposed to control the ON/OFF operation of the MOS transistors in the main rectification chain. An auxiliary block consisting of PMOS transistor is introduced to control the switching of the MOS transistors in the main chain. The PMOS transistors in the main chain are back-compensated in the conduction phase and forward-compensated when not conducting; increasing the gate-source potential and reducing the reverse leakage. The controlling voltage of the transistors in the auxiliary chain is derived from the local node of the main chain. Moreover, a procedure for designing matching network to provide passive voltage amplification and maximum power transfer is proposed. The dead-zone of the power converter unit is obtained using Y parameter analysis and the matching circuit is designed at the point of maximum power transfer. The RF-to-DC power conversion circuit may be fabricated using conventional methods such as IBM's 0.13 μm CMOS technology.

[0048] Power Harvester Circuit

[0049] FIG. **1** shows a block diagram of the power harvester unit of an RF energy harvesting system **20**. In the RF energy harvesting system shown, the signal is received by an antenna **22**. An impedance matching circuit **24** comprising a high quality factor (Q) resonator is used to ensure that maximum power is transferred from the receiver to the rectifier block **26**. However, the matching network itself introduces some energy losses because of the limited quality factors of on-chip passive inductors and capacitors required to match the input impedance of the rectifier to the output impedance of the antenna. The high Q resonator can also provide passive voltage amplification of the input signal. Depending on the embodiment, the high Q resonator may have a fixed frequency or may be tuneable. The next block is the RF-to-DC power conversion circuit **26** that converts the RF signal to DC voltage. The rectifier preferably operates at low input power levels while providing high power conversion efficiency. Several rectifier units may be cascaded to increase the overall output voltage, referred collectively as a voltage multiplier. There may be a power management block (not shown) to control the flow of power from the energy harvester to the load. The harvested power is then finally used to charge an energy storage device **28** such as a battery or a capacitor, which in turn powers a load **30**. Alternatively, the energy storage device may be omitted and the load powered directly by the power harvester. The power management block may also be omitted with or without the energy storage.

[0050] Rectifier Circuit

[0051] In half-wave rectifiers, the rectifying device is conducting for half of the cycle, passing the input voltage to the output. However, the output voltage is lower than the input by the amount of voltage drop needed for turning ON the rectifying device. A half-wave rectifier can be implemented using diodes as shown in FIG. **2**A. The voltage drop across the device depends on the threshold voltage of the diode. FIGS. **2**B and **2**C show half-wave rectification circuits implemented using diode-connected NMOS and PMOS transistors respectively in CMOS technology. The voltage drop in this case depends on the threshold voltage of the transistor. Both diode-connected PMOS and NMOS transistors can be used as rectifying devices. FIG. **2**D shows the voltage doubler unit which is a cascade of two half-wave rectifiers. The voltage doubler rectifies the AC input in both the positive and negative cycles. Similar to the half wave rectification, the diodes are implemented by connecting the drain and the gate terminal of the MOS transistor together such that the transistor is always in saturation region in forward bias condition, as shown in FIGS. **2**E and **2**F.

[0052] The voltage doubler may be used as a basic unit in a power conversion circuit. FIG. **2**F shows an example voltage doubler. In the voltage doubler, each of the transistors conducts only during one half of the input cycle. M_{P1} conducts during the negative phase while M_{P2} conducts during the positive phase of the input cycle. Each of the transistors goes through the sub-threshold region, inverse region and the leakage region. The maximum voltage at the output of voltage doubler is $2\,\mathrm{Vamp}-|\mathrm{VTP1}|-|\mathrm{VTP2}|$, where V_{amp} is the maximum amplitude of the input signal and V_{TP1}, V_{TP2} are the threshold voltages of the diode-connected transistors M_{P1} and M_{P2} respectively. Thus the maximum possible voltage is twice the RF signal's amplitude only when the threshold voltage of the transistors is equal to zero. To understand the operation of the voltage doubler in detail, we can look at the transient analysis of the doubler circuit of FIG. **2**F, as shown in FIG. **3**A. There are three regions of operation of the circuit as seen from FIG. **3**A. It is described as follows: the sub-threshold operation extends from $V_{in}=0$ to $V_{in}=|V_{TP}|$, where V_{in} is the input voltage. The current in this region is an exponential function of the input voltage. The inversion region extends from $V_{in}=|V_{TP}|$ to $V_{in}=V_{amp}$. The current in the in this region is a square function of the input voltage. In the inversion region, the output current reaches its peak value when $V_{in}=V_{amp}$. Finally, the leakage region extends from $V_{in}=0$ to the next $V_{in}=0$ in the negative half-cycle. The current that flows through transistor M_{P2} during this time interval is referred as reverse leakage current. Thus, the actual charge transfer mechanism is only for a short duration while other parameters such as subthreshold and reverse leakage currents have to be considered for the rest of the cycle.

[0053] Several voltage doublers can be cascaded to increase the overall output voltage. These voltage multipliers are also known as "charge pumps" and can generate a voltage several times larger than their input supply voltage. The Dickson multiplier with diode-connected transistors is commonly used for integrated applications. The Dickson multiplier of

FIG. **4**A is modified for low power energy harvesting applications by grounding the ϕ_2 (out of phase) clock terminal and applying the input signal at the ϕ_1 terminal. FIG. **4**B shows the conventional Dickson multiplier with the PMOS voltage doubler as a rectifier unit. The body terminal of the transistors is connected as shown in FIG. **4**B to further reduce the threshold voltage while the transistors are conducting. For example, in the positive phase of the input cycle, transistors M_2, M_{n-1} and M_{n+1} are conducting. The source of the transistors is at a higher potential compared to the body terminal. This reduces the threshold voltage in the conducting region. At the same time, transistors M_1, M_3 and M_n are in the reverse region (leakage operation). The gate-source and the body-source potential is zero for these transistors. Thus, the threshold voltage of transistors M_2, M_{n-1} and M_{n+1} while they are conducting is lower than the threshold voltage of the transistors M_1, M_3 and M_n which are in the reverse region in the positive input phase. Similarly, in the negative phase of the input cycle, transistors M_1, M_3 and M_n which are conducting have a lower threshold voltage than the transistors M_2, M_{n-1} and M_{n+1}.

[0054] To increase the output voltage of multi-stage rectifiers, the number of rectifier stages must be increased accordingly. However, as the number of stages increases the power conversion efficiency is reduced as larger number of the transistors dissipates more power. Also, the threshold voltage of the standard MOS transistor which lies around 450 mV in 130 nm technology, reduces the sensitivity of the power harvester block. As the number of stages in the rectifier increases, the power losses increases, reducing the overall efficiency. Too few of stages, leads to low output DC voltage even if the PCE is high. To achieve a high DC voltage, large transistors have to be used, leading to high leakage and parasitic losses whereas smaller transistor size affects the charge transfer leading to low DC voltage. The strategy while designing the multi-stage rectifier circuit of FIG. **4**B is as follows: The individual stages of the voltage doubler can be cascaded to increase the output voltage. As the number of stages increases, the output voltage increases. With the increase in the number of stages, the PCE decreases and the output voltage saturates. Hence the transistors as well as the pumping capacitors are scaled while increasing the stages. The scaling is done to maintain the incremental voltage per stage and the PCE relatively constant with the increase in the number of stages. Thus the PCE can be maintained with proper device scaling while increasing the DC voltage for a constant output power. The scaling of width and the pumping capacitance with the stages results in impedance looking into the rectifier to be unchanged so that the input power and thus the PCE remains constant. Thus, when designing a multistage rectifier unit one can optimize of number of stages and width of the transistors to maximize the PCE for a given output voltage.

[0055] The standard Dickson multiplier can be modified for designs involving energy harvesters where the input voltage is low. FIG. **4**A shows an N-stage conventional Dickson multiplier. A threshold self-compensation technique was described by Dickson. The gate of the transistor in this technique is connected to the adjacent source of the transistor instead of the traditional diode-connected structure. Thus, providing bias voltage equivalent to the incremental voltage across each stage. Based on this technique, a forward compensated topology was implemented by Papotto et al. where the bias voltage was increased by extending the gate length connection. FIG. **5** shows the basic forward compensated NMOS topology.

[0056] Threshold voltage reduction techniques requiring additional circuits are not suitable for integrated low power energy harvesting applications as these circuits occupy large areas and cause additional power dissipation. Passive threshold voltage reduction technique such as the self-compensation method of FIG. **5** does not require additional circuit and can reduce the threshold voltage. In the self-compensation method, the threshold voltage of the diode-connected NMOS transistors increases with the stages due to the body effect. The body effect can be dynamically controlled using additional circuits but generates undesirable substrate current. Another alternative is using triple-well process to individually bias the body terminal of NMOS transistors and reduce the threshold voltage variation. Additional parasitic capacitance is introduced due to the well structure increasing the losses at each node and affecting the overall efficiency.

[0057] Hybrid Forward and Back-threshold compensated power converter

[0058] In order to eliminate the need for triple-well processes, we propose a threshold voltage compensation technique scheme as described below. PMOS transistors are used instead of NMOS transistors as rectifying devices. As each PMOS transistor has its own n-well, the body of PMOS transistor can be biased individually without the necessity of a triple-well CMOS process. Unlike an NMOS transistor which needs a higher potential at the gate terminal to offset the threshold voltage, a PMOS transistor requires negative gate-source potential. Therefore, the threshold voltage of PMOS transistors can be reduced by connecting the gate potential to the previous node rather than later node. As shown in FIG. **6**A, alternating input voltage V_{in} is applied on first input line **100** to odd-numbered capacitors $C_1, C_3, \ldots C_n$. Ground voltage is applied on second input line **102** to even numbered capacitors $C_2, \ldots C_{n-1}$. Capacitor C_{n+1}, while also even-numbered, is shown as connected to ground by a separate line. A third line **104** accumulates voltage for the output and is connected to ground at one end and to output at the other. In this and other figures, the end of the line having more negative voltage, which in the embodiments shown is ground, is shown at the left of line **104** and the end of the line having more positive voltage, which in the embodiments shown is the output, is shown at the right of line **104**. The capacitors are each connected to line **104**, and are numbered in the order in which they are connected to line **104** from left to right. Each capacitor maintains a respective voltage differential between the respective input line it is connected to and the part of line **104** it is connected to. This voltage differential varies over time as current flows through line **104** between capacitors. Line **104** comprises transistors $M_1, M_2, M_3, \ldots M_{n-1}, M_n, M_{n+1}$ arranged in series, a transistor being arranged between each pair of adjacent capacitors and between capacitor C_1 and ground. Adjacent transistors define junctions **106**, the capacitors being connected to the junctions. The body terminal of each transistor is connected to the drain terminal of the same transistor, as shown by dotted lines. The gate of the first transistor (for n-level compensation, the first 'n' transistors) and the last transistor are connected to their respective drains. The gates of the other transistors are connected to the drain of the transistor n+1 transistors to the left of the respective transistor. The proposed scheme as shown in FIG. **6**A reduces the threshold voltages of all PMOS transistors except the first one leading to an increased output voltage. For an n-level compensation, for the proposed scheme, there will be 'n' initial PMOS transistors that will be uncompensated as seen from FIG. **6**A To solve this problem, 'n' uncompensated PMOS transistors are replaced by NMOS transistors with grounded body terminals as seen from FIG. **6**B. As shown in FIG. **6**B, the first transistor is an NMOS transistor with a source connected to the source of the PMOS transistor to its right to define a junction and has a gate connected to the junction n stages to the right of the junction defined by its source. That is, the gate is connected to the junction defined by the source of the PMOS transistor **2** (i.e, n+1) transistors to the right. The other transistors are PMOS and, apart from the last, each has a gate connected to a junction n stages to the left of the junction defined by its source. For the purposes of the preceding statement and statements about junctions elsewhere in this document, the grounded end of line **104** can be considered to be a junction **106**A one stage from the junction defined by the connection of M_1 and M_2. Thus the compensating voltage is provided by connecting the gate terminal to later stages for the first few NMOS transistors and connecting the gate terminal to previous stages for PMOS transistors. The scheme shown in FIG. **6**A and FIG. **6**B is level-1 compensation. The compensation level can be increased by connecting the gate terminal of PMOS to the source terminals of the transistor of the following stage rather than the source of its immediate neighbor. The last transistor of the multiplier is left uncompensated to reduce the leakage. FIG. **6**C shows the proposed level-3 hybrid forward- and back-compensated multiplier. As shown in FIG. **6**C, the first **3** (i.e., n) transistors are NMOS transistors with grounded body terminals. Each NMOS (more generally, n-type) transistor has its source on the left and drain on the right, and each PMOS (more generally, p-type) transistor has its drain on the left and source on the right, where "left" and "right" are defined so that that the end of line **104** with more positive voltage is the "right", as shown in the figures. "Earlier" and "later" can also be used to mean "left" and "right". While ground is shown in the figures to the left and the output to the right, the ground could also be to the right and the output to the left. See below for a discussion of the definitions of "source" and "drain". The gates of these first **3** NMOS transistors are connected to the respective junctions **106** three (that is, 'n') stages to the right of the junctions defined by their respective sources. The remaining transistors except the last are PMOS transistors with gates connected to the junctions **106** three (that is, 'n') stages to the left of the junctions defined by their respective sources. Increasing the level of compensation leads to reduction in the threshold voltage which improves the forward conduction but also leads to increased reverse leakage current degrading the rectifier's PCE. Only odd level compensation is used as it maximizes the source-gate potential of PMOS transistors due to the alternating voltage phase with successive nodes.

[0059] To find the optimum number of stages and level of compensation, extensive simulations were conducted. FIG. **7** shows the efficiency contour plots with different level of compensation and input power level. Maximum PCE at the lowest input power level is obtained when the level of compensation is one. As the level of compensation increases, the reverse current increases which causes additional power loss and degradation in efficiency. Hence level-1 compensation gives the maximum efficiency while level-3 or higher is advantageous in reducing the threshold voltage of the RF-DC power conversion circuit. FIG. **8**A shows the constant efficiency contour plot as a function of transistor width and the number of stages for level-1 compensation. Eight-stage of doubler design equivalent to 16-stage rectifier with transistor width of 13-μm and fifteen-stage of doubler design equivalent to 30-stage rectifier with transistor width of 28 μm gives the highest efficiency contour. The former one is preferred due to lesser area on the chip. The level-1 and level-3 compensated multi-stage rectifier are referred to as efficiency circuit and voltage circuit respectively Similarly for the voltage circuit constant efficiency contour plot as a function of width and the number of stages is plotted as shown in FIG. **8**B. Twelve-stage of doubler design equivalent to 24-stage rectifier with transistor width of 8 μm is selected based on the plot. For level-3 compensation, more body-grounded NMOS transistors are required compared with the level-1. Also, the reverse leakage loss is higher for level-3 compensation due to the larger compensating voltage compared with the level-1 compensation. Increasing the level of compensation lowers the minimum input voltage requirement whereas increasing the number of stages while lowering the width of the transistors was based on the design strategy discussed earlier.

[0060] During the positive input phase, transistors M_2, M_{n-1} and M_{n+1} in FIGS. **6**A-**6**C are conducting. Since the gate terminals are connected to the previous node V_{SG} in the positive phase (conduction) is relatively high compared to the conventional Dickson multiplier. However, in the negative phase when the transistors M_2, M_{n-1} and M_{n+1} should ideally be not conducting, the V_{SG} is higher compared to the conventional case. Thus the hybrid forward and back-threshold compensated scheme is effective in increasing the forward conduction current but increases the reverse leakage current at the same time. As seen in FIG. **7**, with the increase in the level of compensation the PCE degrades. The simulated maximum PCE of 28% degrades to 11% from level-1 compensation to level-3 compensation respectively. To ameliorate this issue an adaptive scheme is proposed to control the reverse leakage current.

[0061] Especially at extremely low input power (e.g. micro-watts power level), the reverse leakage current has a significant adverse effect on the PCE and the output DC voltage of an RF rectifier as proven by the following analysis. Considering a threshold-compensated transistor of a single-stage rectifier shown in FIG. **2**G driven by an input source $V_{in}=V_a \sin \omega t$ assuming the compensation voltage is modeled by V_C and its input and the output steady-state voltage waveforms shown in FIG. **3**B, the overall PCE is defined as

$$PCE = \frac{P_{out,forward} - P_{leakage}}{P_{input}} \tag{1}$$

[0062] where $P_{out,forward}$ is the output power delivered to the load when transistor is forward-biased, $P_{leakage}$ is the output power lost due to leakage when transistor is reverse-biased, and P_{input} is the input power. The forward region extends from $t=t_1$ to $t=\pi-t_1$ with the current $I_d(t)$ through the MOS transistor M_1 as seen in FIG. **3**B.

$$\frac{P_{out,forward}}{P_{input}} = \tag{2}$$

-continued

$$\frac{\frac{1}{\pi-2t_1}\int_{t_1}^{\pi-t_1} V_{out}\cdot I_d\cdot dt}{\frac{1}{\pi-2t_1}\int_{t_1}^{\pi-t_1} V_{in}\cdot I_d\cdot dt}=\frac{\frac{1}{\pi-2t_1}\int_{t_1}^{\pi-t_1}(V_a-V_{TH}+V_c)\cdot dt}{\frac{1}{\pi-2t_1}\int_{t_1}^{\pi-t_1} V_a\sin\omega t\cdot dt}$$

Assuming the ripple voltage variation A V is much smaller than the average output voltage V_o, the output voltage for the one-stage rectifier can be expressed as $V_o=V_a-V_{TH}+V_C$ where V_{TH} is the threshold voltage of the transistor. Performing integration on (2) gives the ratio of the output power in the forward region to the input power as

$$\frac{P_{out,forward}}{P_{input}}=\frac{\omega\cdot(\pi-2t_1)\cdot(V_a-V_{TH}+V_c)}{2V_a\cos\omega t_1} \tag{3}$$

[0063] The value of t_1 lies between $0<t_1<\pi/2$ based on the value of $V_{TH}-V_C$. The time t_1 indicates the onset of inversion region and is close to zero when the compensating voltage is near the threshold voltage and will be closer to $\pi/2$ when the compensating voltage is near-zero value. As seen in (3), the ratio of the output power in the forward region to the input power increases with increased voltage compensation in the forward region. In the reverse-biased region, the leakage current is expressed as

$$I_{leak}(t)=I_o\cdot(W/L)\cdot e^{(V_{gs}-V_{TH})\eta V_T}\cdot(1-e^{-V_{ds}/\eta V_T}) \tag{4}$$

[0064] Replacing the gate-source bias voltage by V_C, source-drain bias by $V_{in}(t)-V_{out}$ and V_{in} by V_a sin wt, the leakage current as a function of time can be expressed as

$$I_{leak}(t)=I_o\cdot(W/L)\cdot e^{(V_c-V_{TH})/\eta V_T}\cdot(1-e^{(V_a\sin\omega t-V_a+V_{TH}-V_c)/\eta V_T}) \tag{5}$$

[0065] where $(\pi-t_1)<t<(2\pi+t_1)$. With increase in the compensation, the power loss increases due to the higher leakage current. As seen in (3) and (5), the ratio of the output power in the forward region to the input power even though increases with larger threshold-voltage compensation, the losses in the leakage region is higher due to the increased compensation. This indicates the fundamental trade-off between the level of threshold-voltage compensation and the leakage current of the transistors.

[0066] In a field effect transistor, the "source" and "drain" terminals are defined by the direction of flow of charge carriers when the transistor is "on": charge carriers flow from the source to the drain. For an n-type transistor, the charge carriers are electrons and for a p-type transistor the charge carriers are holes. As electrons flow from negative to positive, the source of an n-type transistor is at a lower voltage than the drain and vice versa for a p-type transistor. FETs do not necessarily have any asymmetry between the source and drain terminals beyond this. In the present system, the relative voltages of the source and drain terminals reverse in the normal course of events, as the system is exposed to alternating current in order to rectify it. There are two possible approaches to defining which terminal is the "source" and which terminal is the "drain" in this circumstance. One possible approach is to say that which terminal is the "drain" and which terminal is the "source" varies over time depending on which terminal is at the higher voltage. Thus, by this first definition, the source of an n-type transistor is whichever of the source and drain terminals has the lower voltage at any given moment, and for a p-type transistor whichever has the higher voltage at a given momentA second approach, however, is possible where the desired conduction is in a particular direction only. In the present system, conduction in one direction is desirable and conduction in the other ("leakage") is undesirable, thus the source can be defined according to which terminal is the source when current flows through the transistor in the desired direction; which terminal is considered the "source" and which terminal is considered the "drain" does not change over time in this approach despite the change in voltages. In the discussion here concerning the source-gate voltage (V_{SG}), the first approach is used, i.e. the source-gate voltage is the voltage between the gate and whichever of the source or drain terminals that has the higher or lower voltage at a given time depending on if the transistor is p-type or n-type. In addition, while the sign of a source-gate voltage difference required to operate a transistor depends on whether the transistor is p-type or n-type, V_{SG} is discussed as if the voltage to operate the transistor is always positive for p-type transistor and negative for n-type transistor. For other purposes in this document including the claims, a version of the second approach to defining which terminal is the source and which is the drain is used in which the direction of desired current is based on the desired current direction through the device as a whole. Thus, throughout the figures, p-type transistors are shown as having their sources to the left and n-type transistors are shown as having their sources to the right.

[0067] A diode-connected transistor is a transistor in which one of the source and drain terminals is connected to the gate to cause the transistor to act as a diode. A diode connected FET allows charge carriers to flow from the terminal not connected to the gate to the terminal connected to the gate, if sufficient voltage is applied, and blocks charge carriers from flowing in the opposite direction. The diode connected transistor is diode connected because it is desired that current flow in one direction is allowed and in flow in the other direction blocked. Thus, following a version of the second approach above to defining which terminal is the source and which terminal is the drain based on desired current through the individual transistor, the source or drain terminal of a diode connected transistor that is connected to the gate would be by definition the drain of the diode connected transistor. However, in this document it is the desired current through the device as a whole that is used, thus in e.g. FIG. **9**A transistor M_{na} is diode-connected and the terminal connected to the gate is depicted as the source.

[0068] To investigate the effect of the level of threshold-compensation on the PCE of multi-stage hybrid threshold-compensated rectifiers, level-1 rectifier which is connecting the gate terminal to one previous node for the PMOS transistors (FIG. **6**A) and level-3 rectifier which is connecting the gate terminal to three previous nodes (FIG. **6**C) for the PMOS transistors are simulated to obtain constant maximum efficiency contour plots as shown in FIG. **7**. Increasing the level of compensation leads to reduction in the threshold voltage which improves the forward-conduction but also increases the reverse leakage current. In the multipliers as shown in FIGS. **6**A and **6**C, during the positive input phase, transistor M_n is forward-biased. Since the gate terminals are connected to the previous node, V_{sg} when forward-biased is higher compared to that of the conventional diode-connected case with no threshold-voltage compensation which increases the ratio of output power in the forward region to the input power. However, when the transistor M_n is reverse-biased, the V_{sg} is higher compared to the conventional diode-connected case increasing the reverse leakage current of the reverse-biased transistor. This can also be explained by the efficiency contour plots of the hybrid scheme as seen from FIG. **7**. The contour plots are simulated for constant efficiency with different level of compensation and input power levels. As seen from FIG. **7**, though with the increase in the level of compensation the threshold voltage reduces, the efficiency degrades at the same time which is due to the increased reverse leakage current. The simulated maximum PCE of 28% degrades to 11% from level-1 compensation to level-3 compensation respectively. Accordingly an adaptive scheme is proposed to control the threshold voltage and the reverse leakage current of the rectifying device dynamically to improve the PCE over a wide range of input power. Ideally, the threshold voltage of the rectifying device should be zero when the transistors are forward-biased while the threshold voltage should be high when the transistors are reverse-biased to prevent any leakage or the losses associated with it.

[0069] We propose an adaptive forward and backward threshold-voltage compensation scheme that use minimal additional circuitry to increase the threshold-voltage compensation when transistors are forward-biased and decrease the compensation voltage when they are reverse-biased. The PMOS transistors in the main rectification chain are back-compensated when forward-biased and forward-compensated when reverse-biased; increasing the forward-current and reducing the reverse leakage current dynamically.

[0070] Proposed Adaptive Threshold Voltage Compensation

[0071] The hybrid forward and back-compensated topology described above can be made adaptive using an auxiliary block controlling the gate-source voltage of the MOS transistors in the main rectification chain. The auxiliary structure may be realized using all PMOS transistors to allow for individual body biasing. Preferably the auxiliary blocks are designed using minimum number of PMOS transistors so that the power losses do not increase considerably due to the additional blocks. The controlling voltage of the transistors in the auxiliary chain is derived from the local node of the main rectification chain. Two possible implementation of the proposed adaptive threshold voltage compensation scheme are discussed in the following section, one using diode-connected PMOS transistors and the other solid-wired connection to adaptively adjust the level of the threshold-voltage compensation.

[0072] Using Diode-Connected Transistor for Back-Compensation

[0073] FIG. **9**A shows the hybrid forward and back-threshold compensated rectifier using diode-connected transistors for back-compensation. The back-compensation reduces the threshold voltage when the transistors are forward-biased and forward-compensation reduces the reverse leakage current when reverse-biased with the control signal derived from the local node. In the embodiment shown the last transistor M_{n+1} is left uncompensated to reduce the leakage. During the negative input phase, in which the transistor M_n is forward-biased, the transistor M_n is back-compensated by the diode-connected transistor M_{na} while the V_{SG} terminal voltage for the transistor M_{nb} lies below its threshold voltage resulting in transistor M_{nb} turned OFF. In comparison to the embodiment shown in FIG. **6**B, diode-connected transistor M_{na} lies on the line connecting to the gate of transistor M_n. Transistor M_{nb} lies on a different line not corresponding to a line in FIG. **6**B, providing a forward connection for the gate of transistor M_n when transistor M_{nb} is active. In the embodiment shown in FIG. **9**A p-type transistors are used, but n-type could also be used. Additional transistors of the opposite type may be used at one end as in FIG. **6**B. The back-compensation for the transistor M_n in the main chain enhances the forward current. During the positive input phase, in which the transistor M_n is reverse-biased, the V_{SG} terminal for transistor M_{nb} is enough to turn ON the forward connection thus reducing the V_{SG} bias of the transistor M_n to zero resulting in a reduced leakage current. FIG. **10**A shows the efficiency contour plot for the adaptive diode-connected scheme as a function of width and the number of stages of the rectifier. Generating contour plots are an effective way to optimize the number of stages and the width of the transistors to maximize the PCE while obtaining the required voltage. Twelve-stage of voltage doubler design equivalent to 24-stage rectifier with transistor width of 11 μm gives the highest efficiency contour. The width of the diode-connected transistor in the auxiliary chain should be comparable to the width of the transistors in the main rectification chain so that it provides low forward-resistance when the transistors are conducting. In an embodiment, the width of the diode connected transistor in the auxiliary chain may be 8 μm. The auxiliary transistor used for forward compensation to control the reverse-leakage is selected in an embodiment to be 480 nm, an order of magnitude smaller than the diode-connected transistor to minimize their power consumption and reduce the parasitic at the nodes of the main rectification chain. Larger transistor widths are avoided to reduce the parasitic at the nodes of the main rectification chain.

[0074] Also, for the diode-connected scheme, the auxiliary block is added for every alternate transistor starting from the later stages and adding the blocks towards the initial stages. The maximum PCE increases with the addition of the adaptive auxiliary blocks.

[0075] Using Solid Connection for Back-Compensation

[0076] FIG. **9**B shows the adaptive hybrid topology using solid wired connection instead of the diode-connected transistor for reducing the threshold voltage. When the transistors are in the conduction phase (forward biased) they are back-compensated with solid wired connection instead of the diode-connected transistors which prevents the forward-losses associated with it. In the embodiment of FIG. **9**B, as compared to FIG. **9**A, the transistors M_{na}, $M_{n-1})_a$ etc. are omitted and replaced by a solid wire connection as in FIG. **6**B, but the transistors M_{nb}, $M_{(n-1)b}$ etc. are retained. In the embodiment shown in FIG. **9**A p-type transistors are used, but n-type could also be used. Additional transistors of the opposite type may be used at one end as in FIG. **6**B. When the transistor M_n is reverse-biased (not in the conduction phase), the V_{sg} terminal voltage for the auxiliary transistor M_{nb} is large to turn ON the forward-connection and reduce the source-gate bias of transistor M_n. This decreases the reverse leakage current greatly. FIG. **10**B shows an efficiency contour plot for the solid-wired scheme as a function of width and the number of stages of the rectifier. Twelve stages of doubler design equivalent to 24-stage rectifier with transistor width of 10 μm gives the maximum efficiency contour. The efficiency contour plot for the solid-connection follows similar trend as the diode-connected one. With the increase in the width of the transistors while maintaining the number of stages, the efficiency initially increases and then degrades due to the increased parasitic losses. Similarly, when the width of the transistor is kept constant and the number of stages is increased, the efficiency initially increases and then degrades as more number of stages is added due to the increased power loss with additional stages. Coupling capacitor value of 4 pF is selected as it has a very modest impact on the rectifier's efficiency.

[0077] Further embodiments are shown in FIG. **9**C and FIG. **9**D. The embodiment of FIG. **9**C operates in the same way as the embodiment of FIG. **9**B but uses n-type transistors instead of p-type. Accordingly, the solid wired connections to the gates of the transistors in the main chain are to the right instead of to the left. FIG. **9**D is an embodiment in which the connections to the gates of the transistors of the main chain are each controlled by two transistors. However, unlike in FIG. **9**A, one of these two transistors is p-type and the other is n-type. This allows the gates of these two transistors to be controlled by the same voltage, here using a line connecting to the junction to the right of the corresponding main transistor, and have at most one of the transistors ON at any given point in the input cycle. In addition, the transistors in the main chain are n-type in FIG. **9**D, whereas the transistors of the main chain are p-type in FIG. **9**A. However, either of the embodiments of FIG. **9**A or FIG. **9**D could use n-type or p-type transistors in the main chain, with corresponding changes to the direction of connections as seen in the differences between FIG. **9**B and FIG. **9**C. FIGS. **9**C and **9**D don't show the additional transistor at the end of the rectifier that is shown as M_{n+1} in FIGS. **9**A and **9**B, but such an additional transistor may also be present in these figures. The additional transistor can be p-type or n-type.

[0078] To investigate the effectiveness of the adaptive blocks in reducing the leakage current, leakage current as a function of input power is plotted as shown in FIG. **11**A. The leakage current is simulated for the last transistor with a 1 MΩ load for different input power levels. As shown in FIG. **11**A, comparing the schemes i.e. hybrid, adaptive solid and adaptive diode-connected for power levels of 1 μW-100 μW and at an operating frequency of 915 MHz, the reverse leakage current is maximum for the hybrid scheme and increases with the input power. At an input power of 105 μW, the leakage current is 12 nA. The adaptive solid and the adaptive diode-connected scheme have an auxiliary transistor to control the reverse leakage current. At an input power of 1 μW, the auxiliary transistor does not have enough input power to turn ON and provide the required forward-compensation. Hence the leakage current is comparable with the hybrid scheme. At 1 μW (−30 dBm), the output transistor's leakage current is 6.96 nA for the adaptive solid scheme and 7.56 nA for the adaptive diode-connected scheme with a 1 MΩ load. As the input power increases, the leakage current for the adaptive scheme is drastically reduced as seen in FIG. **11**A. Also, the reverse leakage current for the adaptive schemes is relatively constant with increase in the input power. The leakage current also depends on the load resistance and increases with decrease in the load resistance. Another performance measuring parameter we have defined is the current transfer ratio which is the ratio of the forward-current to the reverse leakage current. FIG. **11**B shows the current transfer ratio as a function of input power. The current transfer ratio at an input power of 1 μW (−30 dBm) for the hybrid scheme is 270, for the adaptive solid scheme is 140 and for the adaptive diode-connected one is 108. The current transfer ratio for the adaptive solid scheme increases rapidly and outperforms the hybrid scheme from 2 μW (−27 dBm) while for the adaptive diode-connected one, the current transfer ratio intersects the hybrid curve at 11 μW. The current transfer ratio for the hybrid scheme initially increases with the input power and saturates to approximately 1500 at 100 μW of input power. The current transfer ratio for the adaptive solid scheme rapidly increases with the input power. At an input power of 90 μW, the current transfer ratio for the adaptive solid scheme is 9000 as shown in FIG. **11**B. The current transfer ratio for the adaptive diode-connected scheme is 4000 at 97 μW of input power. The current transfer ratio for the adaptive-solid scheme at an input power of 100 μW (−10 dBm) is 6 times while for the diode-connected one is 2.5 times better than the hybrid scheme.

[0079] The proposed adaptive scheme is effective in increasing the PCE with the addition of auxiliary adaptive blocks. The increase in the PCE with the addition of the adaptive auxiliary blocks for the adaptive diode-connected scheme is shown in FIG. **12**A. With the addition of the auxiliary block the maximum PCE increases to 12%. The auxiliary block is added for every alternate transistor. However, the additional power consumed by the auxiliary diode-connected transistor when the rectifying device in the main rectification chain is forward-biased prevents the scheme from achieving higher PCEs. Unlike the adaptive diode-connected scheme, the auxiliary adaptive block is added to all the transistors for the adaptive solid scheme. Except for the first and the last stage, all the transistors for the solid adaptive scheme have the adaptive blocks. The additional power consumption and the parasitic capacitance introduced by the auxiliary chain are much lower than the former circuit as the solid-connected scheme uses only one transistor of width 480 nm per adaptive block. Hence the PCE is relatively higher at extremely low power levels. As seen from FIG. **12**B, with the addition of adaptive blocks, the efficiency increases. FIG. **12**B shows the improvement in PCE of the rectifier with the addition of auxiliary block. The maximum PCE reaches to 33.5% when the number of adaptive blocks is 20, a figure much higher than the adaptive diode-connected scheme which has a maximum PCE of 12%. The PCE and the output voltage performance at a relatively higher power level 0 dBm to −20 dBm is much better compared with only the hybrid scheme.

[0080] Impedance Matching Technique

[0081] The overall PCE of the RF energy harvesting system can be improved by increasing the passive voltage amplification due to the matching network, which increases the overall voltage at the input of the rectifier. The rectifier is modeled as $R_{rec}-jX_{rec}$ where R_{rec} is the real part of the rectifier's impedance and X_{rec} is the imaginary part. An impedance matching circuit is designed to match it to 50Ω as shown in FIG. **13**.

[0082] The input capacitance is mainly associated with the parasitic capacitance of the transistors and the layout. As the number of stages increases, the input resistance decreases while the parasitic capacitance increases. Increase in the value of load current also decreases the input resistance. The major factor contributing to increase in load current is lower load resistance. Other factors affecting the input impedance is the aspect ratio (W/L) of the transistors, input power level. Increase in the aspect ratio, increases the parasitic capacitance whereas decreases the input resistance. At the output side, the rectifier can be modeled by a voltage source and an output resistance. Thus the non-linear characteristic of the rectifier circuit makes it difficult to analyze them. The input impedance of the rectifier can be assumed to be constant only when small signal input is applied. In this case, the transient signal is a large signal with a capacitive component at the input causing a phase difference component ϕ to appear. The derivation in appendix though considers the reactive impedance still assumes the linear VI relationship. In real case, the VI relationship is non-linear and the input resistance is to be found using CAD tools.

[0083] Experimental Results—Hybrid Scheme

[0084] Three rectifiers, named as "efficiency", "voltage" and "1-stage PMOS doubler" are designed and fabricated in a 0.13 μm 8-metal CMOS process. The active die areas for efficiency circuit, voltage circuit and the 1-stage PMOS doubler test circuit are 230 μm×810 μm, 230 μm×1050 μm and 160 μm×70 μm respectively. The chip is wire-bonded onto PCB board with FR4 substrate and tested with Agilent MXG-N5181 signal generator at a frequency of 915 MHz using a single-tone sinusoidal signal. The receiver power is calculated by finding the average power at the input of the rectifier. The performance of the designed efficiency and voltage circuit is measured for a range of input power levels.

[0085] Output DC voltage was measured for different peak-to-peak input voltages. For a 1 MΩ load, an input voltage of 170 mV results in 2.4 V and 2.8 V for efficiency and voltage circuit, respectively. The voltage multiplication ratio (VMR) which is the ratio of DC voltage to the peak-to-peak input voltage is 14 and 17 for efficiency and voltage circuit respectively. A 220 mV signal results in 3.1 V for efficiency circuit and 4.0 V for voltage circuit. Similarly for a 5 MΩ load, an output of 2.7 V (VMR=16) for efficiency circuit and 3.0 V (VMR=18) for voltage circuit was measured at an input peak-to-peak voltage of 170 mV. Thus, voltage circuit which has a higher level of compensation than efficiency circuit has a lower input voltage requirement.

[0086] Harvested power was measured for different received powers. From the results, it was observed that at low power-level, the harvested power has a higher dependence on load current. For a 1 MΩ load at power levels greater than −30 dBm, the rate of decay in the harvested power curve is higher than the 5 MΩ load. The roll-off in the harvested power for a 1 MΩ load starts at about −20 dBm while the roll-off point for a 5 MΩ load is approximately −30 dBm. Even for voltage circuit, the rate of decay in the harvested power curve is greater for a 1 MΩ load compared to a 5 MΩ load. A lower load value has a higher current requirement. The performance of both the circuits in terms of roll-off is similar for the same load value. The designed efficiency and voltage circuit outperforms the circuit in Papotto et al. especially at low power levels. The efficiency circuit delivers an output power of 4.7 μW at an input power of −16.8 dBm (20.9 μW) when loaded by 1 MΩ. With a 5 MΩ load, the output power is 1 μW for an input power of −17.5 dBm (17.7 μW). An output power of 3.4 μW at an input power of −14.8 dBm (33.1 μW) for a 1 MΩ load is supplied by voltage circuit. A graph of the measurements was found to be in close agreement with the simulation results.

[0087] Power efficiency was measured for different received power levels. The measured and the simulated power conversion efficiency for efficiency and voltage circuit for different load resistance values are further compared. The efficiency comparison is done while de-embedding the input reflections in Papotto et al. The output DC voltage was also measured for the efficiency and voltage circuits for different load resistance values. The PCE is optimized for low power levels. When loaded by 1 MΩ, efficiency circuit attains a maximum measured PCE of 22.6% at −16.8 dBm (20.9 μW) while delivering 2.2 V to the output. A maximum measured PCE of 21.6% is obtained by efficiency circuit while producing an output voltage of 1.1 V at an input power of −26.5 dBm (2.23 μW) for a 5 MΩ load. At an input power level of −14.8 dBm (33.1 μW), voltage circuit achieves a maximum measured PCE of 10.2% for a 1 MΩ load while delivering 1.8 V. Due to a lower load current requirement for a 5 MΩ load, the output voltage is higher compared to a 1 MΩ load. At an input power of −22.5 dBm (5.6 μW), voltage circuit has a measured output DC voltage of 1 V while efficiency circuit has a measured output DC voltage of 1.8 V at −24 dBm (4 μW) for a 5 MΩ load.

[0088] Experimental Results—Adaptive Scheme

[0089] Three RF-DC power converters named as "adaptive solid," "adaptive diode-connected" and the "hybrid" circuits were designed and fabricated side by side in a 0.13 μm metal CMOS process with eight layers of metallization. The active die areas for adaptive solid and adaptive diode-connected is 0.25 mm^2 and hybrid circuit is 0.15 mm^2. The chip is wire-bonded onto a 2-layer FR-4 PCB board and tested with Agilent MXG-N5181 signal generator using frequency modulated continuous signal in the 902-928 MHz industrial, scientific and medical (ISM) band. An off-chip L-section impedance matching network is implemented on the PCB to convert the RF-DC power converter's input impedance to 50Ω. The output DC voltage was obtained with an oscilloscope or a digital multimeter. The measured and the simulated PCE for the adaptive and the hybrid scheme were compared for a load resistance of 1 MΩ at different input power levels. The output DC voltage was measured for the adaptive and the hybrid scheme for a load resistance of 1 MΩ. The simulation is performed at a frequency of 915 MHz which is the center frequency for the 902-928 MHz ISM band. The adaptive RF-DC power converters are designed to provide high PCE and a large output DC voltage for input power levels of 1 μW-100 μW (−30 dBm to −10 dBm). At larger input power, even with lower PCE the available output power is large hence designing for high input power levels is not so crucial. The adaptive solid scheme attains a maximum PCE of 32% at an input power of 32 μW (−15 dBm) with an output DC voltage of 3.2 V for a 1 MΩ load. For the input power of 32 μW and 1 MΩ load, the hybrid scheme has a PCE of 18% and delivers 2.6 V to the output. The adaptive diode-connected scheme has a maximum PCE of 11.3% at an input power of 118 μW while delivering 3.7 V to the output. The additional power consumed by the auxiliary diode-connected transistor when the rectifying device in the main rectification chain is forward-biased prevents the adaptive diode-connected scheme from achieving higher PCEs. The PCE was measured for different input power for the adaptive solid scheme at different load resistances. The output DC voltage for the adaptive solid scheme was also measured as a function of input power for different load resistances. As the load resistance decreases, the peak conversion efficiency curve shifts towards similar efficiencies at higher input power. The maximum measured PCE is 33.4% for a load resistance of 500 kΩ load at an input power of 83 μW. The PCE for a 500 kΩ load is larger than a 1 MΩ load for input power levels greater than 50 μW (−13 dBm). The peak power conversion efficiency curve is a function of the load resistance and can provide peak efficiency at a much lower power levels for larger load resistances. With decrease in load resistance, the circuit provides a smaller output voltage than with the high load resistance due to the low load current requirement at high load resistances. The hybrid scheme provides a larger output voltage compared to the adaptive schemes for input power levels lesser than 15 μW. The adaptive scheme outperforms the hybrid scheme once the power-up threshold-requirement is met. A DC output voltage of 3.2 V is obtained at an input power of 64 μW (−12 dBm) for a 500 kΩ load.

[0090] The harvested power was measured as a function of input power for the adaptive and the hybrid scheme. When loaded by 1 MΩ, an output power of 10 μW is harvested with an input power of 250 μW (−6 dBm) for the hybrid scheme while the adaptive solid scheme harvests 10 μW with only 30 μW (−15.3 dBm) of input power. With a load resistance of 500 kΩ, 10 μW of output power is harvested at an input power of 42 μW (−13.8 dBm). An output power of 20 μW is harvested at an input power of 64 μW (−12 dBm) for a 500 kΩ load using adaptive solid scheme. For a 1 MΩ load, at larger input power, the PCE degrades much more rapidly compared to a 500 kΩ load. For a 1 MΩ load, 20 μW of power is harvested by the adaptive solid scheme at 164 μW (−7.9 dBm). Even with an input power of few milliwatts, the hybrid scheme is not effective in harvesting 20 μW of output power.

[0091] The adaptive scheme is highly effective when the input power is above 10 μW (−20 dBm) and a large output voltage (more than 1.5 V) is desired. The performance of the adaptive solid scheme is similar to the hybrid scheme for power levels of 1-10 μW (−30 dBm to −20 dBm). The sensitivity of the RF-DC power converter for obtaining an output voltage of 1 V with a 1 MΩ load for the adaptive solid scheme is 8.9 μW (−20.5 dBm). The sensitivity of the circuit using the hybrid scheme is 6.9 μW (−21.6 dBm). The adaptive diode-connected scheme gives similar performance as the adaptive solid scheme at input power levels greater than 100 μW (−10 dBm). For low power applications, the adaptive solid scheme should be preferred over the adaptive diode-connected scheme.

[0092] The forgoing description pertains to circuitry using MOSFET enhancement-mode transistors. Nothing in the description should be taken to exclude using other types of transistors with corresponding changes to the circuitry that would be obvious to a person skilled in the art.

[0093] Immaterial modifications may be made to the embodiments described here without departing from what is covered by the claims.

[0094] In the claims, the word “comprising” is used in its inclusive sense and does not exclude other elements being present. The indefinite articles “a” and “an” before a claim feature do not exclude more than one of the feature being present. Each one of the individual features described here may be used in one or more embodiments and is not, by virtue only of being described here, to be construed as essential to all embodiments as defined by the claims.

APPENDIX

[0095] In section II, the average input power was suggested consisting of real and imaginary parts.

$$V = V_{in}\sin\omega t,\ I = I_{in}(\sin\omega t - \phi)$$

$$P_{mean} = \frac{1}{T_2 - T_1}\int_{T_1}^{T_2} V_{in}\sin\omega t \cdot I_{in}(\sin\omega t - \phi)dt$$

$$Z_{in} = \frac{V_{in}^2}{2P_{mean}} \cdot \cos\phi + \frac{V_{in}^2}{4P_{mean}} \cdot \frac{\sin[2\omega(T_2 - T_1) - \phi]}{\omega(T_2 - T_1)}$$

[0096] Z_{in} consists of real and imaginary component. Equating the imaginary component to zero by considering the phase difference $\phi=0$.

$$R_{in} = \frac{V_{in}^2}{2P_{mean}},\ C_{in} = \frac{4P_{mean}(T_2 - T_1)}{V_{in}^2\sin[2\omega(T_2 - T_1) - \phi]}$$

1. A power conversion circuit comprising:

a first input line and a second input line, the first and second input lines configured to receive an alternating voltage differential between the first and second input lines;

a multi-stage rectifier comprising transistors arranged in series, each transistor having a gate, a source and a drain, adjacent transistors of the series being connected so that for adjacent p-type transistors the drain of the left p-type transistor is connected to the right adjacent source of the right p-type transistor and source of the right p-type transistor is connected to the left adjacent drain of the left p-type transistor, for adjacent n-type transistors the source of the left n-type transistor is connected to the right adjacent drain of the right n-type transistor and drain of the right n-type transistor is connected to the left adjacent source of the left n-type transistor, for a p-type transistor adjacent to an n-type transistor the source of the right p-type transistor is connected to the source of the left adjacent n-type transistor and drain of the left p-type transistor is connected to the drain of the right adjacent n-type transistor to form a junction, each junction being connected to one of the first input line and the second input line via a capacitor, with adjacent junctions having one junction of the adjacent junctions connected to the first input line and the other junction of the adjacent junctions connected to the second input line;

the gate of each transistor of the multi-stage rectifier being connected to a respective junction that is not a junction formed by the connection of that transistor to an adjacent transistor but a junction in the previous stage for the p-type transistor or later stage for the n-type transistor, at least one of the transistor in the multi-stage rectifier being p-type and at least one being n-type.

2. The power conversion circuit of claim **1** further comprising an auxiliary chain of p-type transistors, each auxiliary transistor having a gate, a source and a drain, each transistor of the auxiliary chain being connected to a p-type transistor of the multi-stage rectifier in the main chain, so that the gate of the respective auxiliary transistor is connected to the source of the respective transistor in the main chain, the source of the respective auxiliary transistor is connected to the gate of the transistor in the main chain and also connected to the previous stage junction of the main chain which is N transistors away from the respective transistor, and the drain of the respective auxiliary transistor is connected to the drain of the respective transistor or to the junction of the later stage transistor in the main chain.

3. The power conversion circuit of claim **1** further comprising an auxiliary chain of n-type transistors, each auxiliary transistor having a gate, a source and a drain, each transistor of the auxiliary chain being connected to an n-type transistor of the multi-stage rectifier in the main chain, so that the gate of the respective auxiliary transistor is connected to the source of the respective transistor in the main chain, the source of the respective auxiliary transistor is connected to the gate of the transistor in the main chain and also connected to the later stage junction of the main chain which is N transistors away from the respective transistor, and the drain of the respective auxiliary transistor is connected to the drain of the respective transistor or to the junction of the previous stage transistor in the main chain.

4. An auxiliary chain of transistors, each auxiliary transistor having a gate, a source and a drain, each transistor of the auxiliary chain being connected to a respective transistor of a main chain of a multi-stage rectifier in the main chain, so that the gate of the respective auxiliary transistor is connected to the source of the respective transistor in the main chain, the source of the respective auxiliary transistor is connected to the gate of the transistor in the main chain and also connected to the later stage junction of the main chain which is N transistors away from the respective transistor in the case that the respective transistor is n-type, or to the earlier stage junction of the main chain which is N transistors away from the respective transistor in the case that the respective transistor is p-type, and the drain of the respective auxiliary transistor is connected to the drain of the respective transistor or to the junction of a previous stage transistor in the main chain, in the case that the respective transistor is n-type, or to the junction of a later stage transistor in the main chain, in the case that the respective transistor is p-type.

5. The power conversion circuit of claim **4** in which the transistors of the main chain and auxiliary chain are n-type and further comprising a p-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of later stages, such that the gate of the additional p-type auxiliary transistor is connected to the gate of the n-type auxiliary transistor and the source of the n-type and p-type auxiliary transistors are connected, and the drain of the p-type auxiliary transistors is connected to junction of the main chain which is at 'N' later stages from the respective transistor.

6. The power conversion circuit of claim **4** in which the transistors of the main chain and auxiliary chain are p-type and further comprising an n-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of earlier stages, such that the gate of the additional n-type auxiliary transistor is connected to the gate of the p-type auxiliary transistor and the source of the n-type and p-type auxiliary transistors are connected, and the drain of the n-type auxiliary transistors is connected to junction of the main chain which is at 'N' earlier stages from the respective transistor.

7. The power conversion circuit of claim **2** further comprising a diode connected transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of previous stages which is N transistors away from the respective transistor.

8. The power conversion circuit of claim **3** further comprising a diode connected transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of later stages which is N transistors away from the respective transistor.

9. The power conversion circuit of claim **4** further comprising a diode connected p-type or n-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of later stages which is N transistors away from the respective transistor.

10. The power conversion circuit of claim **4** further comprising a diode connected p-type or n-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of earlier stages which is N transistors away from the respective transistor.

11. The power conversion circuit of claim **1** further comprising an additional transistor connected in series with the multistage rectifier, the additional transistor having a gate, a source and a drain, the gate of the additional transistor being connected to the drain of the additional transistor.

12. The power conversion circuit of claim **1** in which the gate of each transistor in the main chain is connected to the junction of previous stages for p-type transistor and later stages for n-type transistor.

13. The power conversion circuit of claim **12** in which N of the transistors in the multistage rectifier arranged in series are n-type and all but N of the transistors in multistage rectifier arranged in series are p-type.

14. The power conversion circuit of claim **1** in which the second input line is grounded.

15. The power conversion circuit of claim **1** in which the body terminal of each p-type transistor is connected to the respective drain and the body terminal of each n-type transistor is either grounded or connected to the respective drain terminal.

* * * * *

US 20170111073A1

(19) **United States**
(12) **Patent Application Publication**
Hajimiri et al.
(10) **Pub. No.: US 2017/0111073 A1**
(43) **Pub. Date: Apr. 20, 2017**

(54) **RF RECEIVER**

(71) Applicant: **California Institute of Technology**, Pasadena, CA (US)

(72) Inventors: **Seyed Ali Hajimiri**, La Canada Flintridge, CA (US); **Florian Bohn**, Pasadena, CA (US); **Behrooz Abiri**, Pasadena, CA (US)

(21) Appl. No.: **15/273,633**

(22) Filed: **Sep. 22, 2016**

Related U.S. Application Data

(63) Continuation-in-part of application No. 14/078,489, filed on Nov. 12, 2013.

(60) Provisional application No. 61/724,638, filed on Nov. 9, 2012, provisional application No. 62/222,106, filed on Sep. 22, 2015.

Publication Classification

(51) **Int. Cl.**

H04B 1/16	(2006.01)
H02J 7/02	(2006.01)
H02J 50/20	(2006.01)

(52) **U.S. Cl.**
CPC ***H04B 1/1607*** (2013.01); ***H02J 50/20*** (2016.02); ***H02J 7/025*** (2013.01)

(57) **ABSTRACT**

A device includes, in part, an antenna adapted to receive an RF signal that includes modulated data, a splitter/coupler adapted to split the received RF signal, a receiver adapted to demodulate the data from a first portion of the RF signal, and a power recovery unit adapted to convert to a DC power a second portion of the RF signal. The splitter/coupler is optionally adjustable to split the RF signal in accordance with a value that may be representative of a number of factors, such as the target data rate, the DC power requirement of the device, and the like. The device optionally includes a switch and/or a power combiner adapted to deliver all the received RF power to the receiver depending on any number of operation conditions of the device or the device's distance from an RF transmitting device.

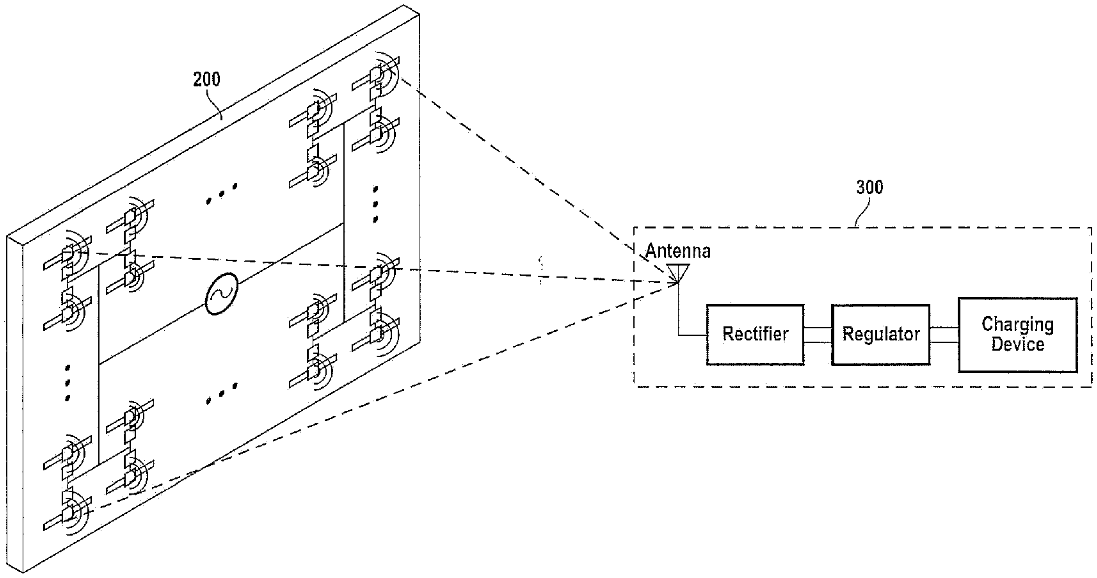

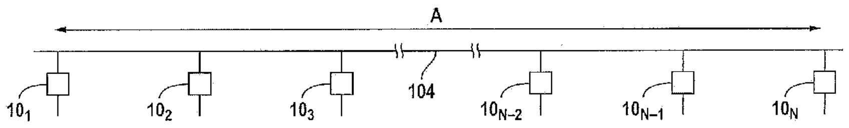
A
104
10_1
10_2
10_3
10_{N-2}
10_{N-1}
10_N

FIG. 1

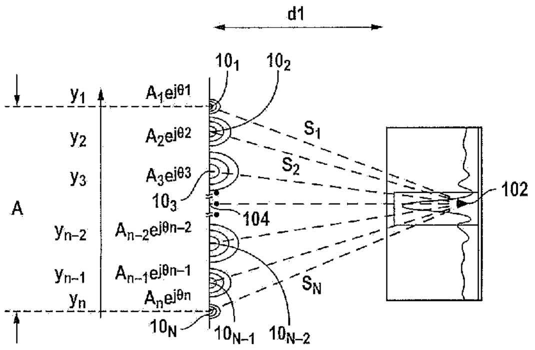
d1
10_1
10_2
y_1
$A_1e^{j\theta1}$
y_2
$A_2e^{j\theta2}$
S_1
S_2
y_3
$A_3e^{j\theta3}$
10_3
102
A
104
y_{n-2}
$A_{n-2}e^{j\theta n-2}$
y_{n-1}
$A_{n-1}e^{j\theta n-1}$
S_N
y_n
$A_ne^{j\theta n}$
10_N
10_{N-1}
10_{N-2}

FIG. 2

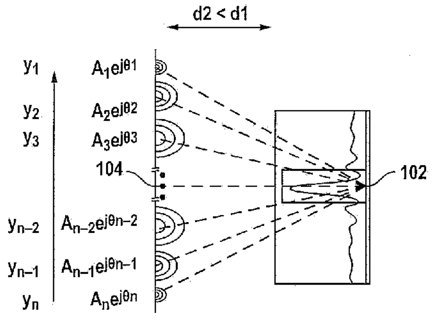
d2 < d1
y_1
$A_1e^{j\theta1}$
y_2
$A_2e^{j\theta2}$
y_3
$A_3e^{j\theta3}$
104
102
y_{n-2}
$A_{n-2}e^{j\theta n-2}$
y_{n-1}
$A_{n-1}e^{j\theta n-1}$
y_n
$A_ne^{j\theta n}$

FIG. 3

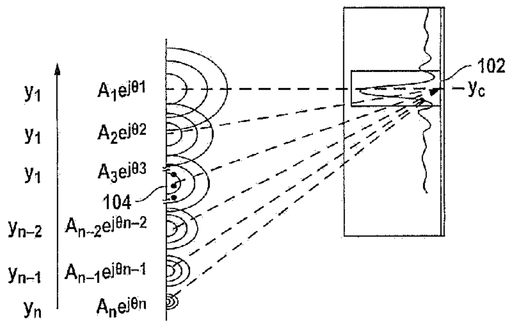
y_1
$A_1e^{j\theta 1}$
y_1
$A_2e^{j\theta 2}$
y_1
$A_3e^{j\theta 3}$
104
y_{n-2}
$A_{n-2}e^{j\theta n-2}$
y_{n-1}
$A_{n-1}e^{j\theta n-1}$
y_n
$A_ne^{j\theta n}$
102
y_c

FIG. 4

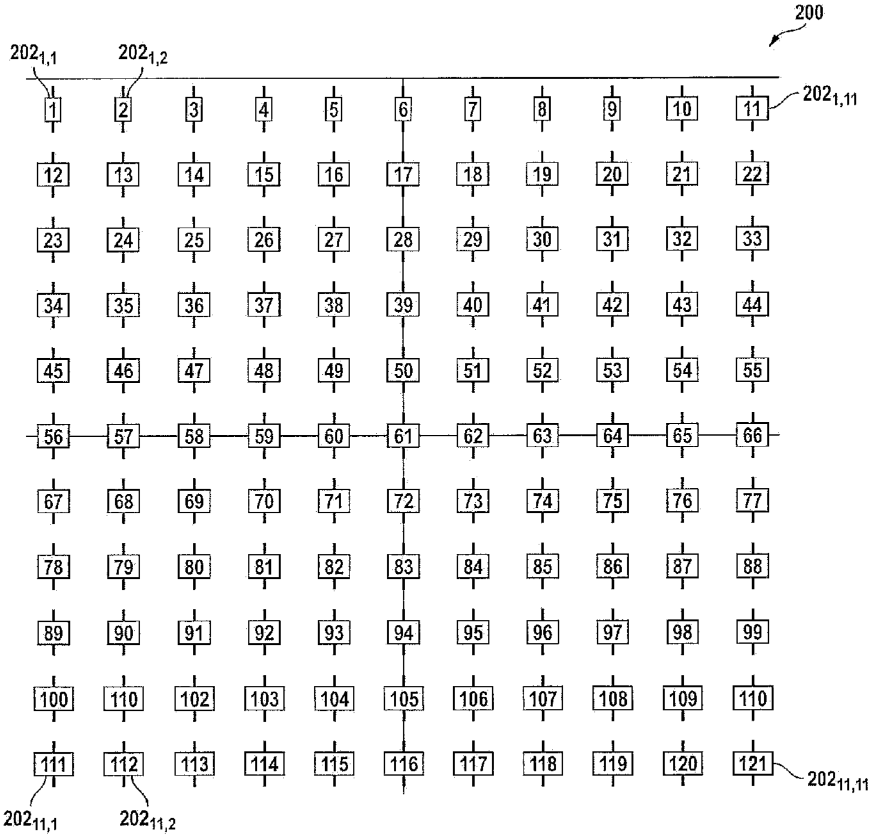
200
$202_{1,1}$
$202_{1,2}$
$202_{1,11}$
1 2 3 4 5 6 7 8 9 10 11
12 13 14 15 16 17 18 19 20 21 22
23 24 25 26 27 28 29 30 31 32 33
34 35 36 37 38 39 40 41 42 43 44
45 46 47 48 49 50 51 52 53 54 55
56 57 58 59 60 61 62 63 64 65 66
67 68 69 70 71 72 73 74 75 76 77
78 79 80 81 82 83 84 85 86 87 88
89 90 91 92 93 94 95 96 97 98 99
100 110 102 103 104 105 106 107 108 109 110
111 112 113 114 115 116 117 118 119 120 121
$202_{11,11}$
$202_{11,1}$
$202_{11,2}$

FIG. 5

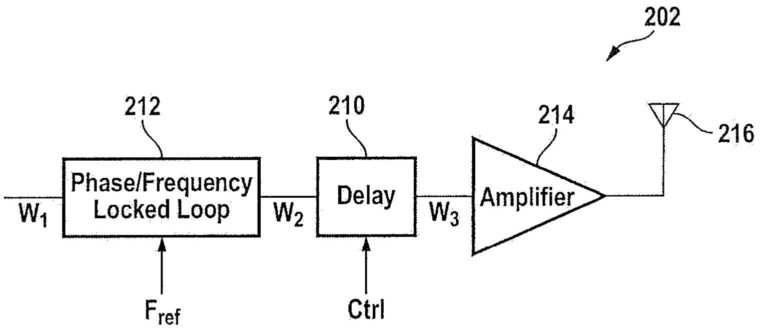
202
212
210
214
216
Phase/Frequency Locked Loop
Delay
Amplifier
W1
W2
W3
Fref
Ctrl

FIG. 6A

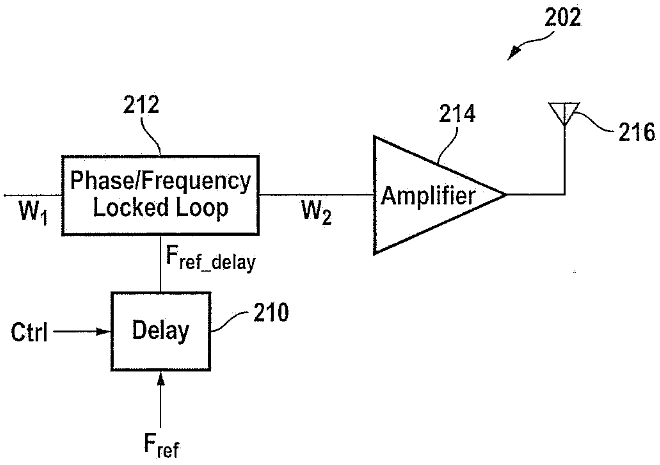
202
212
214
216
Phase/Frequency Locked Loop
Amplifier
W1
W2
Fref_delay
Ctrl
Delay
210
Fref

FIG. 6B

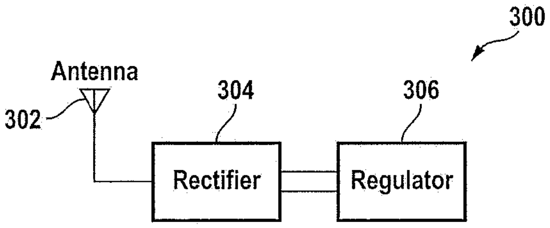
300
Antenna
302
304
306
Rectifier
Regulator

FIG. 7

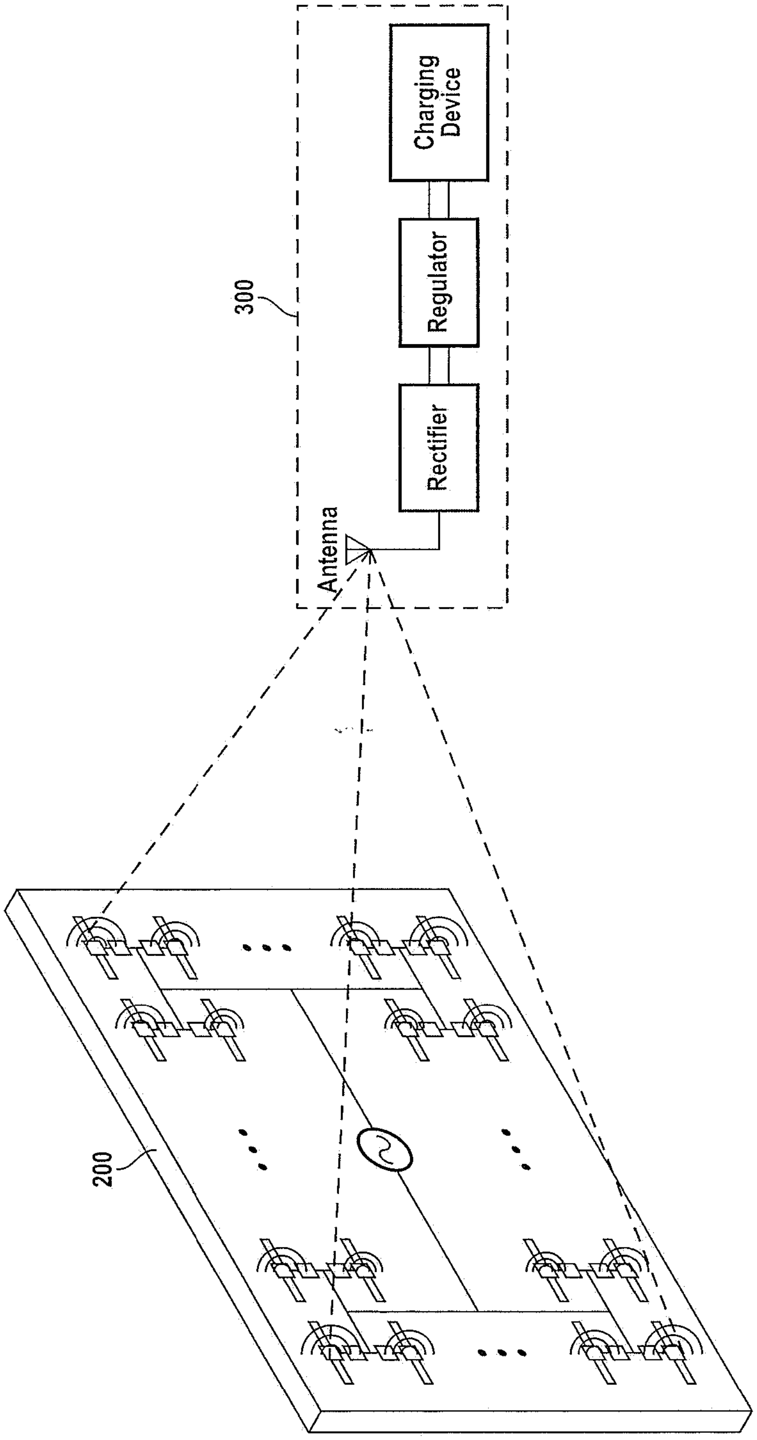
300
Charging Device
Regulator
Rectifier
Antenna
200

FIG. 8

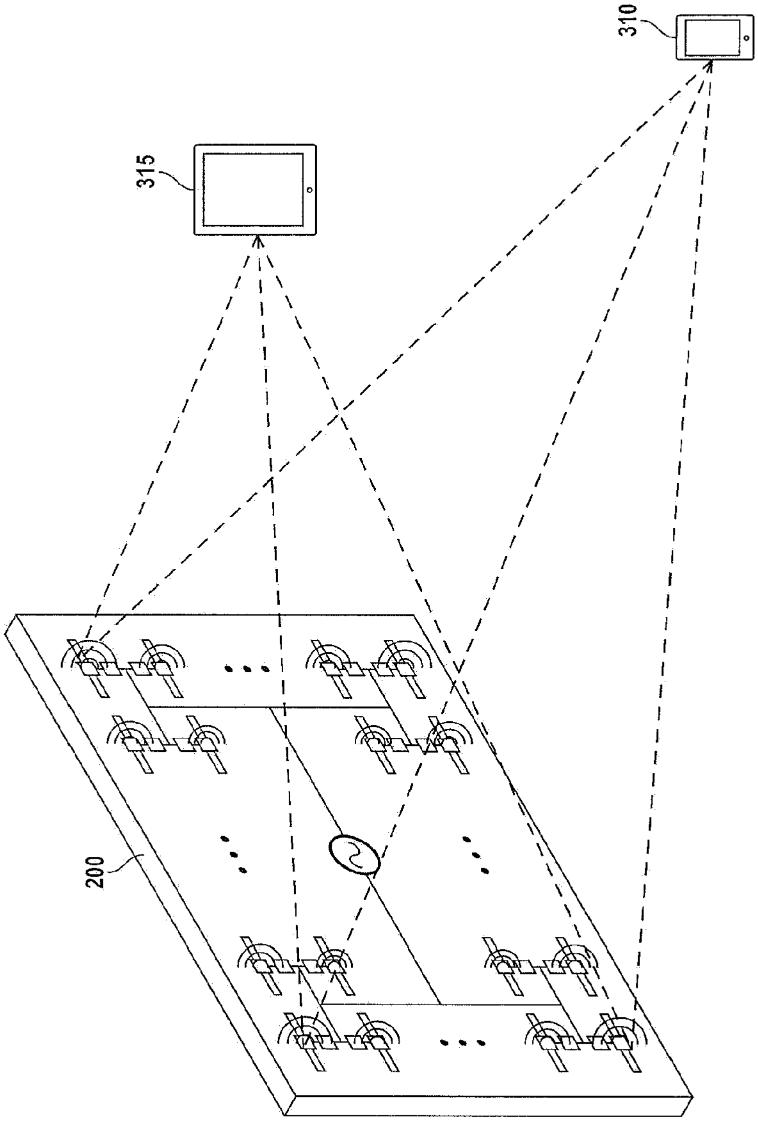

FIG. 9

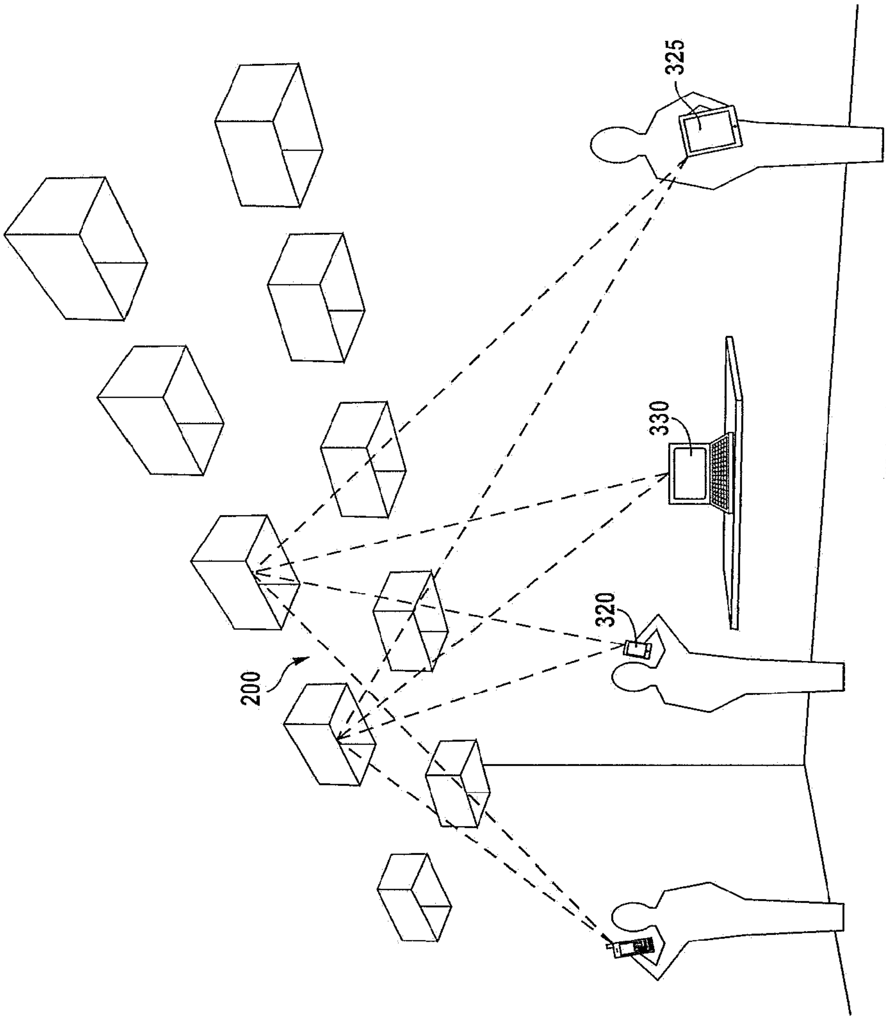
325
330
320
200

FIG. 10

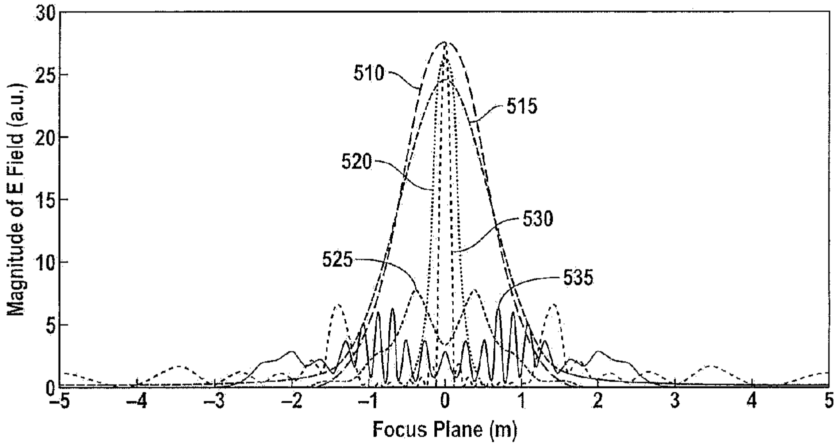

FIG. 11A

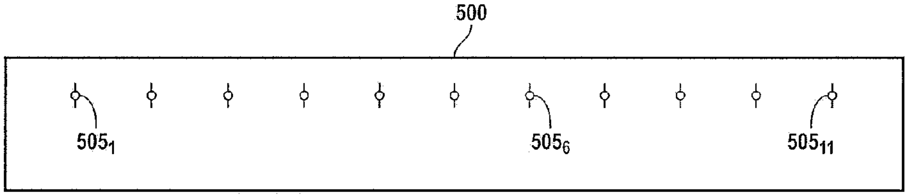
500
505_1
505_6
505_{11}

FIG. 11B

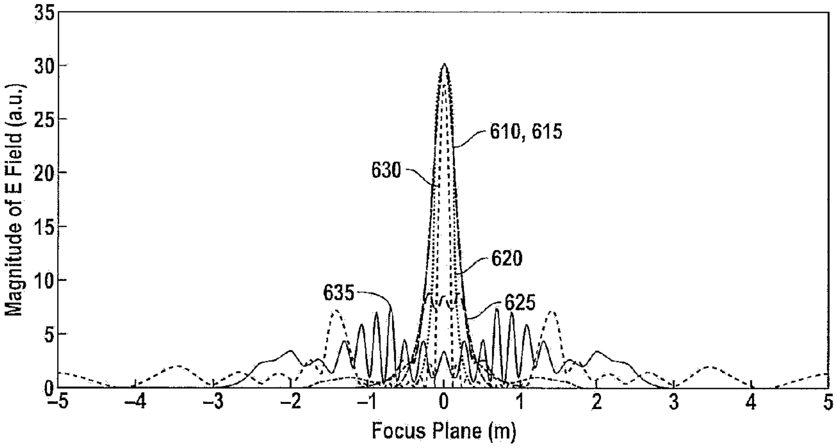

FIG. 12

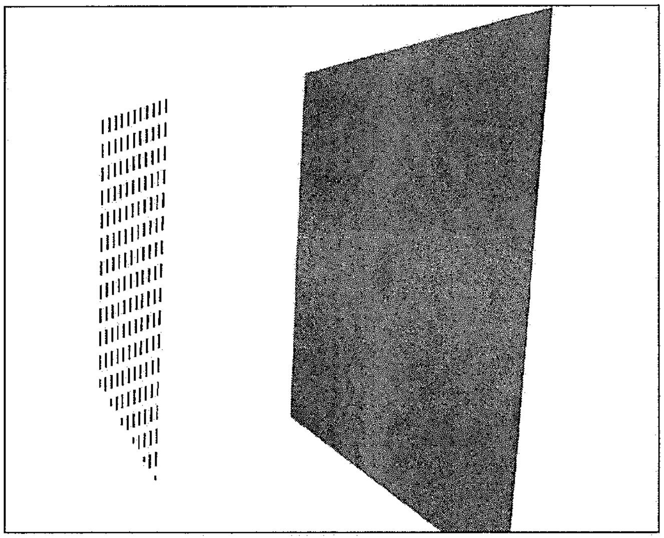

FIG. 13A

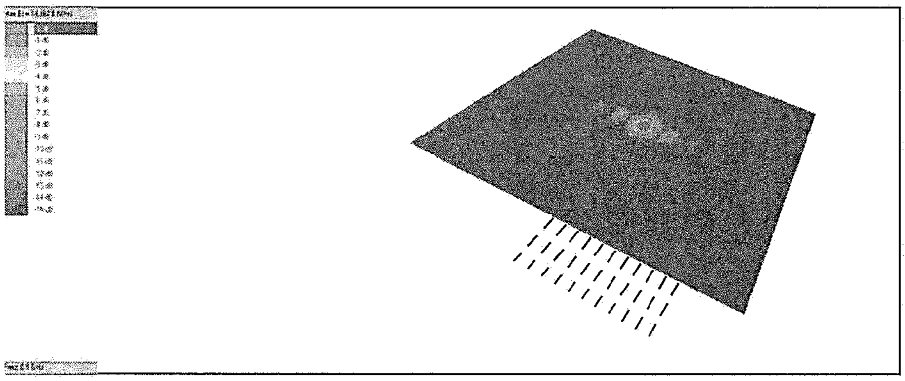

FIG. 13B

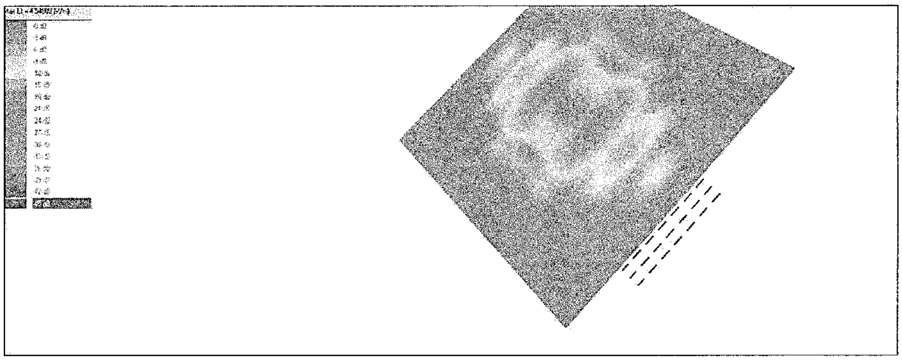

FIG. 14A

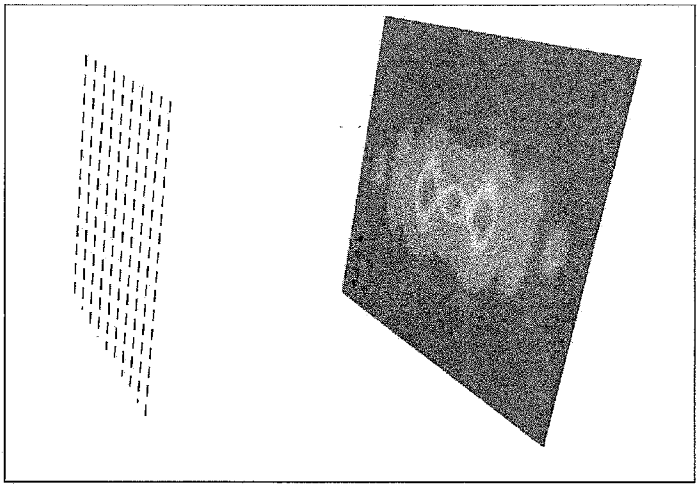

FIG. 14B

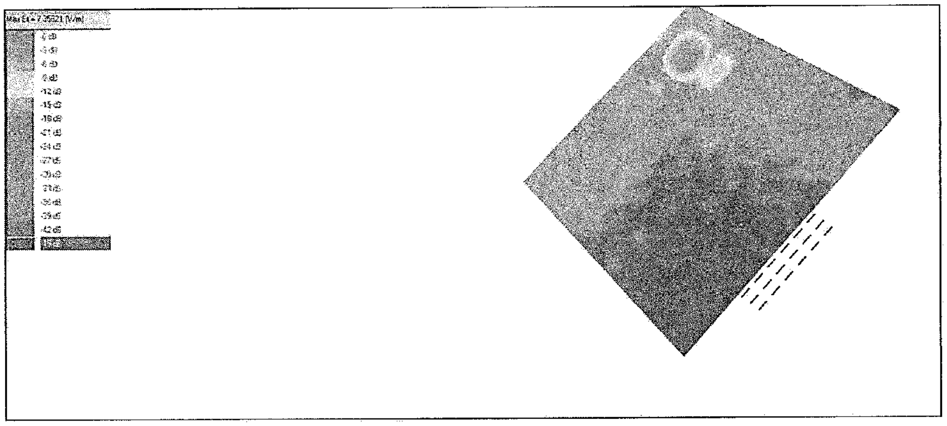

FIG. 15A

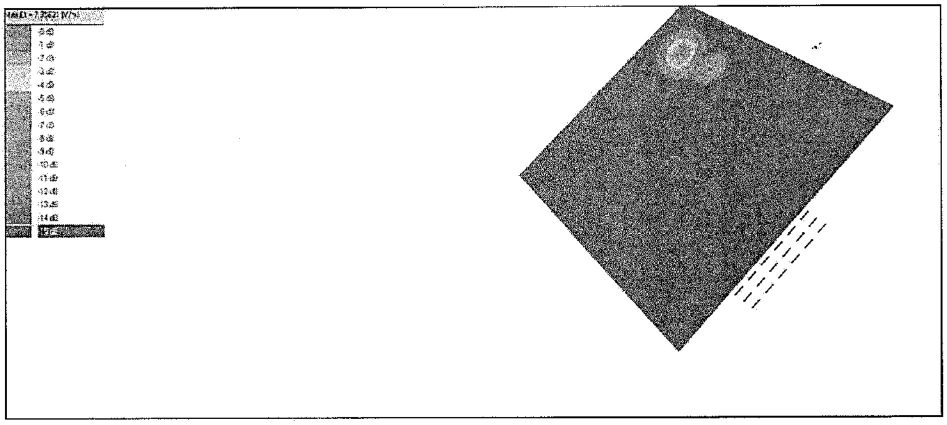

FIG. 15B

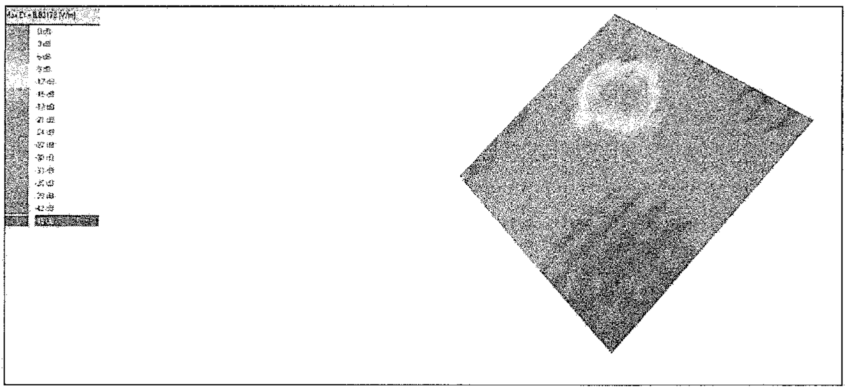

FIG. 16A

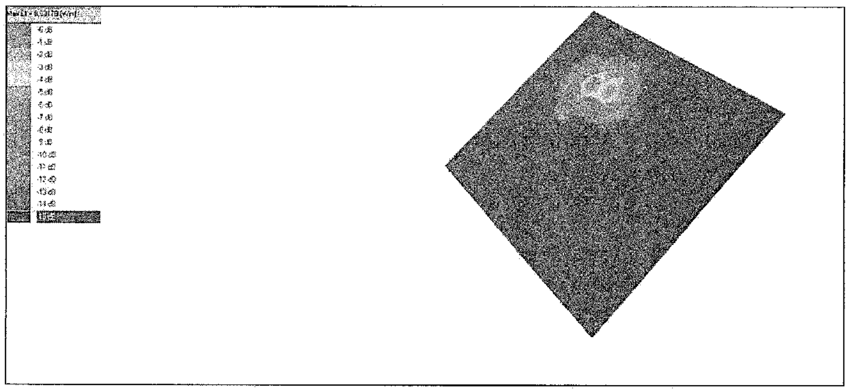

FIG. 16B

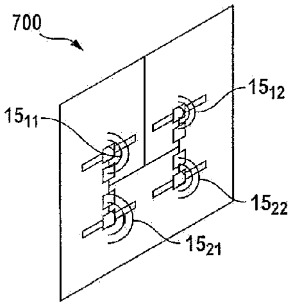
700
15_{11}
15_{12}
15_{22}
15_{21}

FIG. 17A

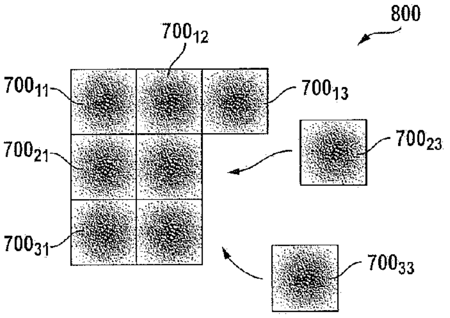
800
700_{12}
700_{11}
700_{13}
700_{21}
700_{23}
700_{31}
700_{33}

FIG. 17B

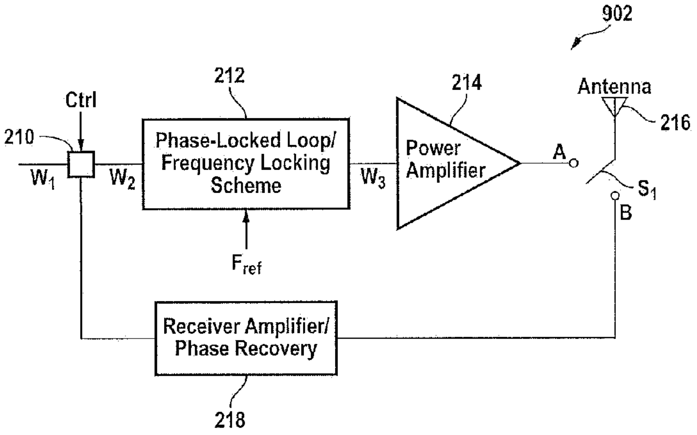
902
212
214
Ctrl
Antenna
216
210
Phase-Locked Loop/
Frequency Locking
Scheme
Power
Amplifier
A
W_1
W_2
W_3
S_1
B
F_{ref}
Receiver Amplifier/
Phase Recovery
218

FIG. 18

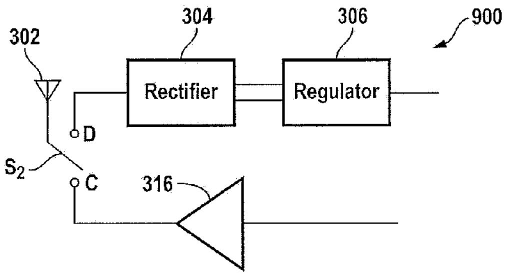
304
306
900
302
Rectifier
Regulator
D
S_2
C
316

FIG. 19

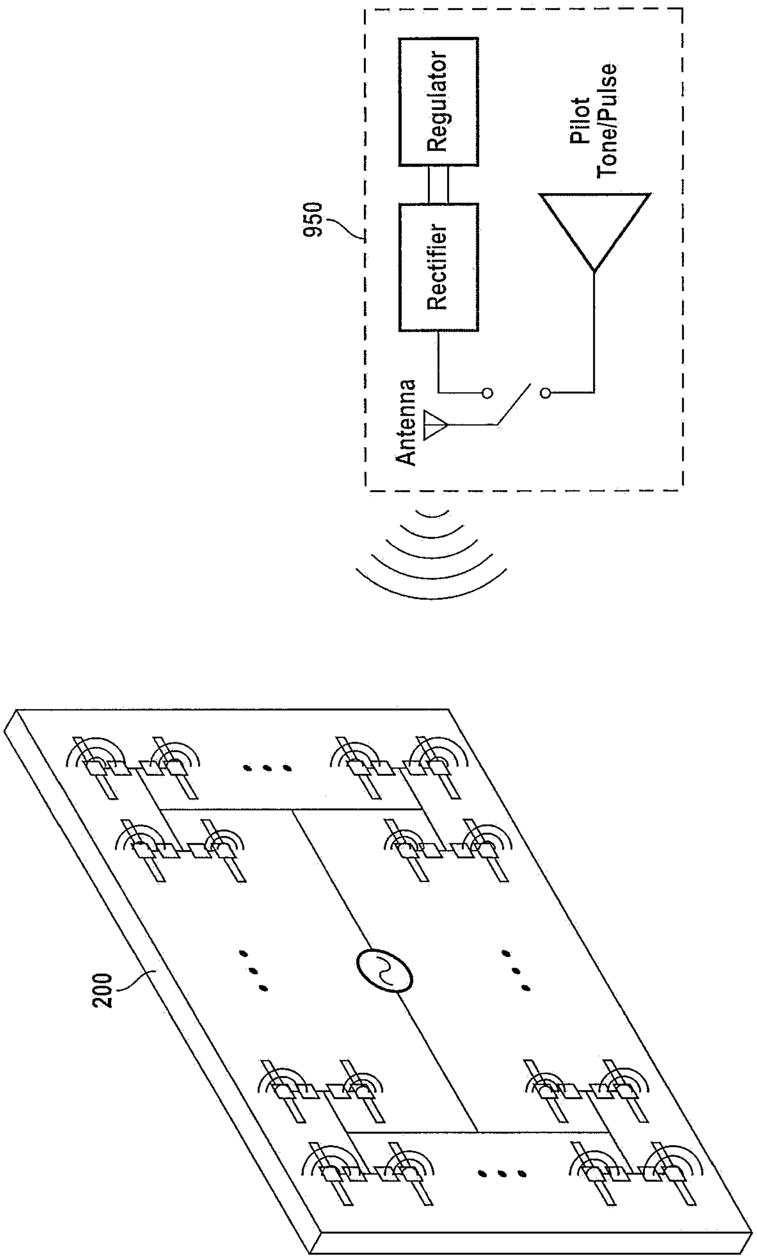
950
Regulator
Rectifier
Pilot
Tone/Pulse
Antenna
200

FIG. 20

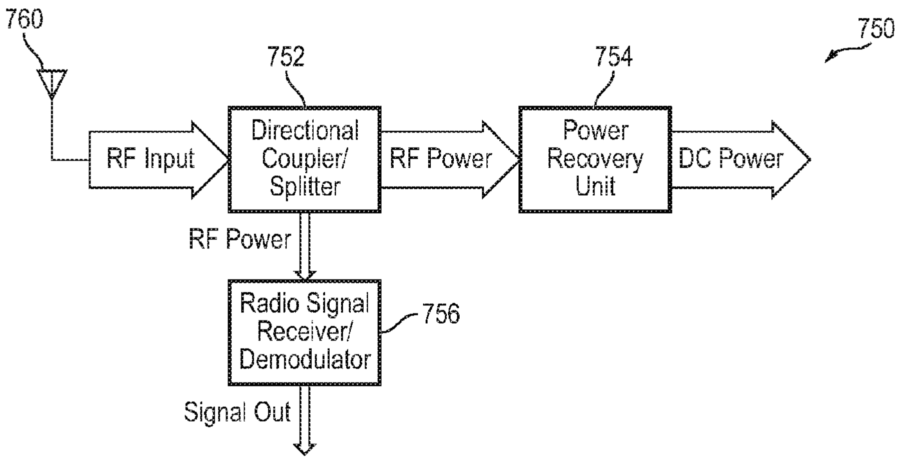
760
752
754
750
RF Input
Directional Coupler/ Splitter
RF Power
Power Recovery Unit
DC Power
RF Power
Radio Signal Receiver/ Demodulator
756
Signal Out

FIG. 21A

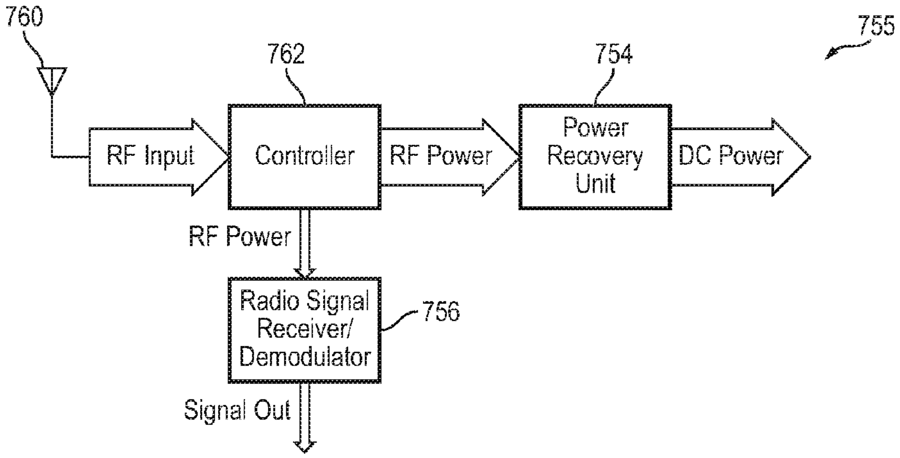
760
762
754
755
RF Input
Controller
RF Power
Power Recovery Unit
DC Power
RF Power
Radio Signal Receiver/ Demodulator
756
Signal Out

FIG. 21B

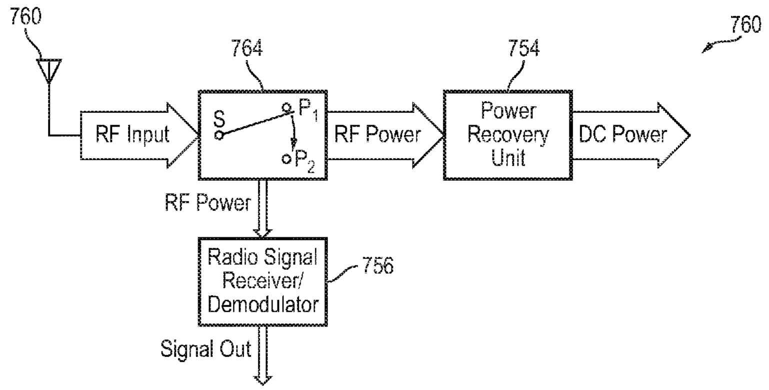
760
764
754
760
RF Input
S
P1
P2
RF Power
Power Recovery Unit
DC Power
RF Power
Radio Signal Receiver/ Demodulator
756
Signal Out

FIG. 21C

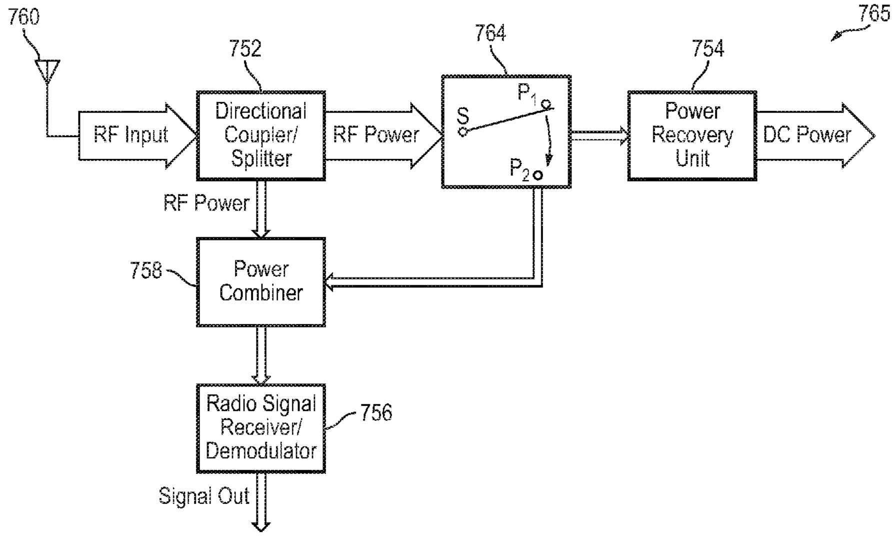
760
752
764
754
765
RF Input
Directional Coupler/ Splitter
RF Power
S
P1
P2
Power Recovery Unit
DC Power
RF Power
758
Power Combiner
Radio Signal Receiver/ Demodulator
756
Signal Out

FIG. 21D

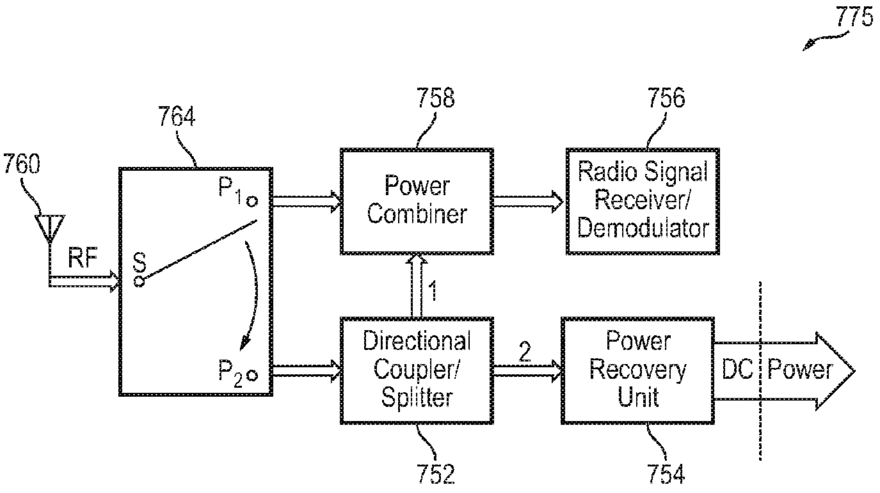
775
764
758
756
760
P_1
Power Combiner
Radio Signal Receiver/ Demodulator
RF
S
1
Directional Coupler/ Splitter
2
Power Recovery Unit
DC Power
P_2
752
754

FIG. 21E

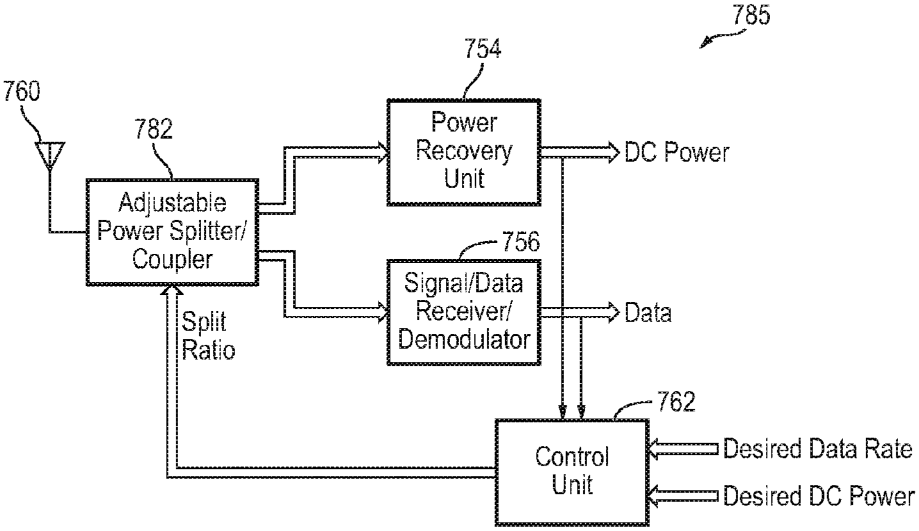
785
754
760
782
Power Recovery Unit
DC Power
Adjustable Power Splitter/ Coupler
756
Signal/Data Receiver/ Demodulator
Data
Split Ratio
762
Control Unit
Desired Data Rate
Desired DC Power

FIG. 21F

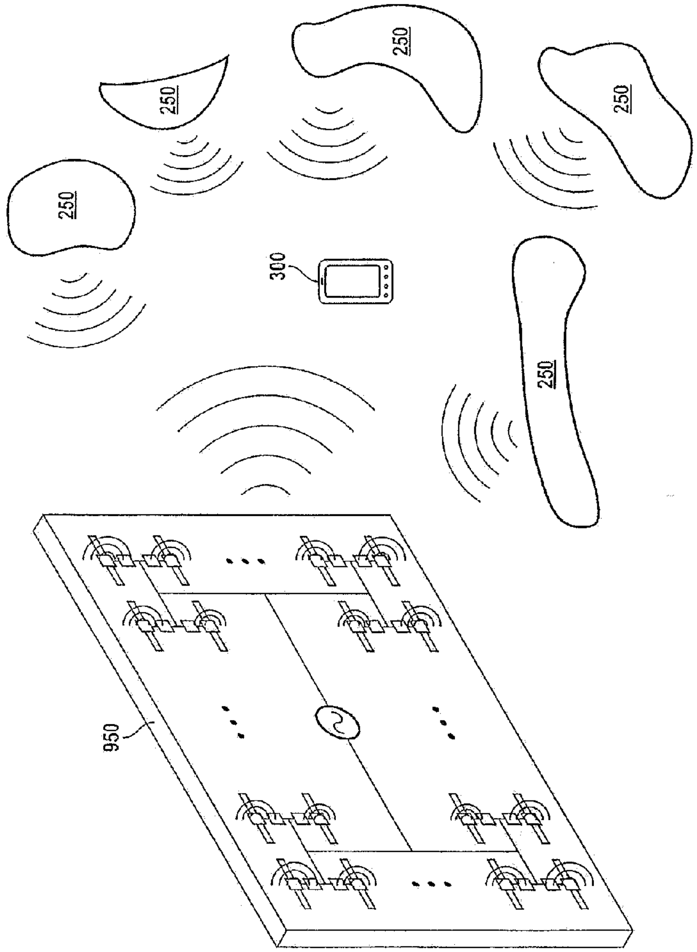
250
250
250
250
250
300
250
950

FIG. 22

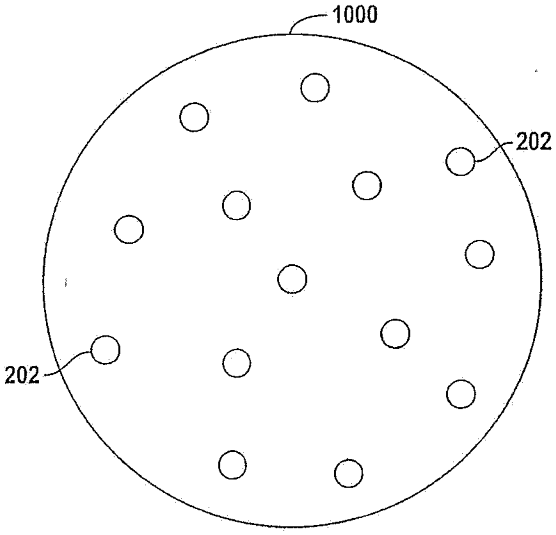
1000
202
202

FIG. 23A

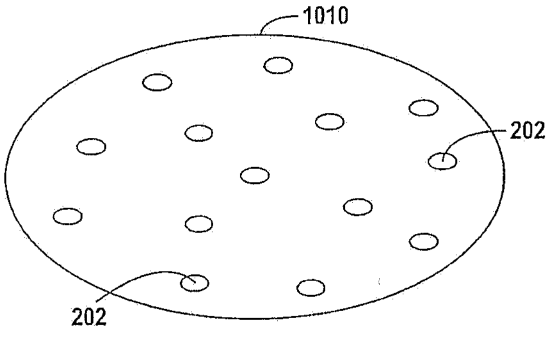
1010
202
202

FIG. 23B

RF RECEIVER

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present application is a continuation-in-part of application Ser. No. 14/078,489, filed Nov. 12, 2013, entitled "Smart RF Lensing: Efficient, Dynamic And Mobile Wireless Power Transfer" which claims benefit under 35 §U.S.C 119(e) of U.S. Provisional Patent Application No. 61/724,638, filed Nov. 9, 2012, entitled "Smart RF Lensing: Efficient, Dynamic And Mobile Wireless Power Transfer", the contents of which are incorporated herein by reference in its entirety.

[0002] The present application claims benefit under 35 §U.S.C 119(e) of U.S. Provisional Patent Application No. 62/222,106, filed Sep. 22, 2015, entitled "Smart RF Lensing: Efficient, Dynamic And Mobile Wireless Power Transfer", the content of which is incorporated herein by reference in its entirety.

[0003] The present application is related to the following US applications:

[0004] application Ser. No. 14/552,249, filed Nov. 24, 2014, entitled "Active CMOS Recovery Units For Wireless Power Transmission";

[0005] application Ser. No. 14/552,414, filed Nov. 24, 2014, entitled "Generator Unit For Wireless Power Transfer"; and

[0006] application Ser. No. 14/830,692, filed Aug. 19, 2015, entitled "Wireless Power Transfer";

the contents of all of which applications are incorporated herein by reference in their entirety.

BACKGROUND OF THE INVENTION

[0007] The present invention relates to wireless communication, and more particularly to wireless power and data transfer.

BACKGROUND OF THE INVENTION

[0008] Electrical energy used in powering electronic devices comes predominantly from wired sources. Conventional wireless power transfer relies on magnetic inductive effect between two coils placed in close proximity of one another. To increase its efficiency, the coil size is selected to be less than the wavelength of the radiated electromagnetic wave. The transferred power diminishes strongly as the distance between the source and the charging device is increased.

BRIEF SUMMARY OF THE INVENTION

[0009] An RF lens, in accordance with one embodiment of the present invention, includes, in part, a multitude of radiators adapted to radiate electromagnetic waves to power a device positioned away from the RF lens. Each of the multitude of radiators operates at the same frequency. The phase of the electromagnetic wave radiated by each of the multitude of radiators is selected to be representative of the distance between that radiator and the device.

[0010] In one embodiment, the multitude of radiators are formed in an array. In one embodiment, the array is a one-dimensional array. In another embodiment, the array is a two-dimensional array. In one embodiment, the amplitudes of the electromagnetic waves radiated by the radiators is variable. In one embodiment, each of the multitude of radiators includes, in part, a variable delay element, a control circuit adapted to lock the phase or frequency of the electromagnetic wave radiated by that radiator to the phase or frequency of a reference signal, an amplifier, and an antenna.

[0011] In one embodiment, the multitude of radiators are formed in a first radiator tile adapted to receive a second radiator tile having disposed therein another multitude of radiators. In one embodiment, the RF lens is further adapted to track a position of the device. In one embodiment, each of a first subset of the radiators includes a circuit for receiving an electromagnetic wave transmitted by the device thus enabling the RF lens to determine the position of the device in accordance with the phases of the electromagnetic wave received by the first subset of the radiators.

[0012] In one embodiment, each of at least a first subset of the radiators includes a circuit for receiving an electromagnetic wave transmitted by the device thereby enabling the RF lens to determine a position of the device in accordance with a travel time of the electromagnetic wave from the device to each of the first subset of the radiators and a travel time of a response electromagnetic wave transmitted from the RF lens to the device. In one embodiment, the RF lens is formed in a semiconductor substrate.

[0013] A method of wirelessly powering a device, in accordance with one embodiment of the present invention, includes, in part, transmitting a multitude of electromagnetic waves having the same frequency from a multitude of radiators to the device, selecting a phase of each of the multitude of radiators in accordance with a distance between that radiator and the device, and charging the device using the electromagnetic waves received by the device.

[0014] In one embodiment, the method further includes, in part, forming the radiators in an array. In one embodiment, the radiators are formed in a one-dimensional array. In another embodiment, the radiators are formed in a two-dimensional array. In one embodiment, the method further includes, in part, varying the amplitude of the electromagnetic wave radiated by each of the radiators.

[0015] In one embodiment, each radiators includes, in part, a variable delay element, a controlled locked circuit adapted to lock the phase or the frequency of the electromagnetic wave radiated by the radiator to the phase or frequency of a reference signal, an amplifier, and an antenna. In one embodiment, the radiators are formed in a first radiator tile adapted to receive a second radiator tile having disposed therein another multitude of radiators.

[0016] In one embodiment, the method further includes, in part, tracking the position of the device. In one embodiment, the method further includes, in part, determining the position of the device in accordance with relative phases of an electromagnetic wave transmitted by the device and received by each of at least a subset of the radiators. In one embodiment, the method further includes, in part, determining the position of the device in accordance with a travel time of an electromagnetic wave transmitted by the device and received by each of at least a subset of the radiators, and further in accordance with a travel time of a response electromagnetic wave transmitted from the RF lens to the device. In one embodiment, the method further includes, in part, forming the RF lens in a semiconductor substrate.

[0017] A device, in accordance with one embodiment of the present invention, includes, in part, an antenna adapted to receive an RF signal that includes modulated data, a splitter/coupler adapted to split the received RF signal into first and second portions, a receiver adapted to demodulate the data from the first portion of the received RF signal, and a power recovery unit adapted to convert the second portion of the RF signal to a DC power to power the device. In one embodiment, the splitter/coupler is an adjustable splitter/coupler

[0018] A device, in accordance with one embodiment of the present invention, includes, in part, an antenna adapted to receive an RF signal that includes modulated data, a receiver adapted to demodulate the data from a first portion of the RF signal, a power recovery unit adapted to convert a second portion of the RF signal to a DC power to power the device, and a controller adapted to receive the RF signal from the antenna and generate the first and second portions of the RF signal in accordance with impedance values of the receiver and the power recovery unit.

[0019] A device, in accordance with one embodiment of the present invention, includes, in part, an antenna adapted to receive an RF signal that includes modulated data, a switch adapted to receive the RF signal from the antenna, a power recovery unit adapted to convert the RF signal to a DC power to power the device when the switch is in a first position, and a receiver adapted to demodulate the data from the received RF signal when the switch is in a second position.

[0020] A device, in accordance with one embodiment of the present invention, includes, in part, an antenna adapted to receive an RF signal that includes modulated data, a splitter/coupler adapted to split the received RF signal into first and second portions, a switch adapted to receive the second portion of the RF signal from the splitter/coupler, a power recovery unit adapted to convert the second portion of the RF signal to a DC power to power the device when the switch is in a first position, and a power combiner adapted to receive the first portion of the RF signal from the splitter/coupler and further to receive the second portion of the RF signal when the switch is in a second position, and a receiver adapted to demodulate the data from an output signal of the power combiner. In one embodiment, the device further includes a controller adapted to cause the switch to be in the first position when a power of the received RF signal is less than a first threshold value. In one embodiment, the device further includes a controller adapted to cause the switch to be in the first position when the device indicates that its DC power exceeds a second threshold value.

[0021] A device, in accordance with one embodiment of the present invention, includes, in part, an antenna adapted to receive an RF signal that includes modulated data, a switch adapted to receive the RF signal from the antenna, a power combiner coupled to a first output terminal of the switch to receive the RF signal when the switch is in a first position, a splitter/coupler coupled to a second output terminal of the switch to receive the RF signal, when the switch is in a second position, to split the RF signal into a first portion and a second portion and deliver the first portion of the RF signal to the power combiner, a power recovery unit adapted to convert the first portion of the RF signal to a DC power to charge the device when the switch is in the second position, and a receiver adapted to demodulate the data from an output signal of the power combiner. In one embodiment, the device further includes, in part, a controller adapted to cause the switch to be in the first position when a power of the received RF signal is less than a threshold value. In yet another embodiment, the device further includes a controller adapted to cause the switch to be in the first position when the device indicates that its DC power exceeds a threshold value.

[0022] A device, in accordance with one embodiment of the present invention, includes, in part, an antenna adapted to receive an RF signal that includes modulated data, a controller, an adjustable splitter/coupler adapted to split the received RF signal into first and second portions in accordance with a value the adjustable splitter/coupler receives from the controller, a receiver adapted to demodulate the data from the first portion of the RF signal, and a power recovery unit adapted to convert the second portion of the RF signal to a DC power to charge the device. In one embodiment, the value supplied by the controller is defined by a target data rate of the device. In another embodiment, the value supplied by the controller is defined by a DC power requirement of the device.

[0023] A method, in accordance with one embodiment of the present invention, includes, in part, receiving an RF signal that includes modulated data, splitting the received RF signal to first and second portions, demodulating the data from the first portion of the RF signal, and converting the second portion of the RF signal to a DC power.

[0024] A method, in accordance with one embodiment of the present invention, includes, in part, receiving an RF signal that includes modulated data, demodulating the data from a first portion of the received RF signal via a receiver, converting a second portion of the RF signal to a DC power via a power recovery unit, and generating the first and second portions of the RF signal in accordance with impedance values of the receiver and the power recovery unit.

[0025] A method, in accordance with one embodiment of the present invention, includes, in part, receiving an RF signal that includes modulated data, converting the RF signal to a DC power when a switch is in a first position, and demodulating the data from the received RF signal when the switch is in a second position.

[0026] A method, in accordance with one embodiment of the present invention, includes, in part, receiving an RF signal that includes modulated data, demodulating the data using either a first portion of the RF signal or the RF signal, and converting a second portion of the RF signal to a DC power when the first portion of the RF signal is used for demodulating the data. In one embodiment, the data is modulated using the RF signal when a power of the received RF signal is less than a first threshold value. In one embodiment, the data is demodulated using the RF signal when an indication is received that a battery charge exceeds a threshold value.

[0027] A device, in accordance with one embodiment of the present invention, includes, in part, an antenna adapted to receive an RF signal that includes modulated data, a controller, an adjustable splitter/coupler adapted to split the received RF signal into first and second portions in accordance with a value the adjustable splitter/coupler receives from the controller, a receiver adapted to demodulate the data from the first portion of the RF signal, and a power recovery unit adapted to convert the second portion of the RF signal to a DC power to charge the device. In one embodiment, the value supplied by the controller is defined by a target data rate of the device. In another embodiment, the value supplied by the controller is defined by a DC power requirement of the device.

[0028] A method, in accordance with one embodiment of the present invention, includes, in part, receiving an RF signal, splitting the received RF signal to first and second portions in accordance with a received value, demodulating the data from the first portion of the RF signal, and converting the second portion of the RF signal to a DC power. In one embodiment, the value is defined by a target data rate. In one embodiment, the value is defined by a DC power requirement.

BRIEF DESCRIPTION OF THE DRAWINGS

[0029] FIG. **1** shows a one-dimensional array of radiators forming an RF lens, in accordance with one embodiment of the present invention.

[0030] FIG. **2** is a side view of the RF lens of FIG. **1** wirelessly delivering power to a device at a first location, in accordance with one exemplary embodiment of the present invention.

[0031] FIG. **3** is a side view of the RF lens of FIG. **1** wirelessly delivering power to a device at a second location, in accordance with one exemplary embodiment of the present invention.

[0032] FIG. **4** is a side view of the RF lens of FIG. **1** wirelessly delivering power to a device at a third location, in accordance with one exemplary embodiment of the present invention.

[0033] FIG. **5** shows a two-dimensional array of radiators forming an RF lens, in accordance with one exemplary embodiment of the present invention.

[0034] FIG. **6**A is a simplified block diagram of a radiator disposed in an RF lens, in accordance with one exemplary embodiment of the present invention.

[0035] FIG. **6**B is a simplified block diagram of a radiator disposed in an RF lens, in accordance with another exemplary embodiment of the present invention.

[0036] FIG. **7** shows a number of electronic components of a device adapted to be charged wirelessly, in accordance with one exemplary embodiment of the present invention.

[0037] FIG. **8** is a schematic diagram of an RF lens wirelessly charging a device, in accordance with one exemplary embodiment of the present invention.

[0038] FIG. **9** is a schematic diagram of an RF lens concurrently charging a pair of devices, in accordance with one exemplary embodiment of the present invention.

[0039] FIG. **10** is a schematic diagram of an RF lens concurrently charging a pair of mobile devices and a stationary device, in accordance with one exemplary embodiment of the present invention.

[0040] FIG. **11**A shows computer simulations of the electromagnetic field profiles of a one-dimensional RF lens, in accordance with one exemplary embodiment of the present invention.

[0041] FIG. **11**B is a simplified schematic view of an RF lens used in generating the electromagnetic field profiles of FIG. **11**A.

[0042] FIG. **12** shows the variations in computer simulated electromagnetic field profiles generated by the RF lens of FIG. **11**B as a function of the spacing between each adjacent pair of radiators disposed therein.

[0043] FIG. **13**A is an exemplary computer-simulated electromagnetic field profile of an RF lens and using a scale of −15 dB to 0 dB, in accordance with one exemplary embodiment of the present invention.

[0044] FIG. **13**B shows the computer-simulated electromagnetic field profile of FIG. **13**A using a scale of −45 dB to 0 dB.

[0045] FIG. **14**A is an exemplary computer-simulated electromagnetic field profile of the RF lens of FIG. **13**A and using a scale of −15 dB to 0 dB, in accordance with one exemplary embodiment of the present invention.

[0046] FIG. **14**B shows the computer-simulated electromagnetic field profile of FIG. **14**A using a scale of −45 dB to 0 dB, in accordance with one exemplary embodiment of the present invention.

[0047] FIG. **15**A is an exemplary computer-simulated electromagnetic field profile of an RF lens and using a scale of −15 dB to 0 dB, in accordance with one exemplary embodiment of the present invention.

[0048] FIG. **15**B shows the computer-simulated electromagnetic field profile of FIG. **15**A using a scale of −45 dB to 0 dB, in accordance with one exemplary embodiment of the present invention.

[0049] FIG. **16**A is an exemplary computer-simulated electromagnetic field profile of the RF lens of FIG. **15**A using a scale of −15 dB to 0 dB, in accordance with one exemplary embodiment of the present invention.

[0050] FIG. **16**B shows the computer-simulated electromagnetic field profile of FIG. **16**A using a scale of −45 dB to 0 dB, in accordance with one exemplary embodiment of the present invention.

[0051] FIG. **17**A shows an exemplary radiator tile having disposed therein four radiators, in accordance with one exemplary embodiment of the present invention.

[0052] FIG. **17**B shows an RF lens formed using a multitude of the radiator tiles of FIG. **17**A, in accordance with one exemplary embodiment of the present invention.

[0053] FIG. **18** is a simplified block diagram of a radiator disposed in an RF lens, in accordance with another exemplary embodiment of the present invention.

[0054] FIG. **19** shows a number of electronic components disposed in a device adapted to be charged wirelessly, in accordance with another exemplary embodiment of the present invention.

[0055] FIG. **20** shows an RF lens tracking a device using a signal transmitted by the device, in accordance with another exemplary embodiment of the present invention.

[0056] FIG. **21**A is a simplified high-level block diagram of a device, in accordance with one embodiment of the present invention.

[0057] FIG. **21**B is a simplified high-level block diagram of a device, in accordance with one embodiment of the present invention.

[0058] FIG. **21**C is a simplified high-level block diagram of a device, in accordance with one embodiment of the present invention.

[0059] FIG. **21**D is a simplified high-level block diagram of a device, in accordance with one embodiment of the present invention.

[0060] FIG. **21**E is a simplified high-level block diagram of a device, in accordance with one embodiment of the present invention.

[0061] FIG. **21**F is a simplified high-level block diagram of a device, in accordance with one embodiment of the present invention.

[0062] FIG. **22** shows an RF lens transferring power to a device in the presence of a multitude of scattering objects, in accordance with another exemplary embodiment of the present invention.

[0063] FIG. **23**A shows an RF lens formed using a multitude of radiators arranged in a circular shape, in accordance with one embodiment of the present invention.

[0064] FIG. **23**B shows an RF lens formed using a multitude of radiators arranged in an elliptical shape, in accordance with one embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0065] An RF lens, in accordance with one embodiment of the present invention, includes a multitude of radiators adapted to transmit radio frequency electromagnetic EM waves (hereinafter alternatively referred to as EM waves, or waves) whose phases and amplitudes are modulated so as to concentrate the radiated power in a small volume of space (hereinafter alternatively referred to as focus point or target zone) in order to power an electronic device positioned in that space. Accordingly, the waves emitted by the radiators are caused to interfere constructively at the focus point. Although the description below is provided with reference to wireless power transfer, the following embodiments of the present invention may be used to transfer any other kind of information wirelessly.

[0066] FIG. **1** shows a multitude of radiators, arranged in an array **100**, forming an RF lens, in accordance with one embodiment of the present invention. Array **100** is shown as including N radiators $\mathbf{10}_1$, $\mathbf{10}_2$, $\mathbf{10}_3$ . . . $\mathbf{10}_{N-1}$, $\mathbf{10}_N$ each adapted to radiate an EM wave whose amplitude and phase may be independently controlled in order to cause constructive interference of the radiated EM waves at a focus point where a device to be charged is located, where N is integer greater than 1. FIG. **2** is a side view of the array **100** when the relative phases of the waves generated by radiators $\mathbf{10}_i$ (i is an integer ranging from 1 to N) are selected so as to cause constructive interference between the waves to occur near region **102** where a device being wirelessly charged is positioned, i.e., the focus point. Region **102** is shown as being positioned at approximately distance d_1 from center **104** of array **100**. The distance between the array center and the focus point is alternatively referred to herein as the focal length. Although the following description of an RF lens is provided with reference to a one or two dimensional array of radiators, it is understood that an RF lens in accordance with the present invention may have any other arrangement of the radiators, such as a circular arrangement **1000** of radiators **202** as shown in FIG. **22**A, or the elliptical arrangement **1010** of radiators **202** shown in FIG. **22**B.

[0067] As seen from FIG. **2**, each radiator $\mathbf{10}_i$ is assumed to be positioned at distance y_i from center **104** of array **100**. The amplitude and phase of the wave radiated by radiator $\mathbf{10}_i$ are assumed to be represented by A_i and θ_i respectively. Assume further that the wavelength of the waves being radiated is represented by) λ. To cause the waves radiated by the radiators to interfere constructively in region **102** (i.e., the desired focus point), the following relationship is satisfied between various phases θ_i and distances y_i:

$$\theta_1 + \frac{2\pi}{\lambda}\sqrt{d_1^2 + y_1^2} = \theta_2 + \frac{2\pi}{\lambda}\sqrt{d_1^2 + y_2^2} = \cdots = \theta_N + \frac{2\pi}{\lambda}\sqrt{d_1^2 + y_N^2}$$

[0068] Since the phase of an RF signal may be accurately controlled, power radiated from multiple sources may be focused, in accordance with the present invention, onto a target zone where a device to be wirelessly charged is located. Furthermore, dynamic phase control enables the tracking of the device as it moves from its initial location. For example, as shown in FIG. **3**, if the device moves to a different position—along the focal plane—located at a distance d_2 from center point **104** of the array, in order to ensure that the target zone is also located at distance d_2, the phases of the sources may be adjusted in accordance with the following relationship:

$$\theta_1 + \frac{2\pi}{\lambda}\sqrt{d_2^2 + y_1^2} = \theta_2 + \frac{2\pi}{\lambda}\sqrt{d_2^2 + y_2^2} = \cdots = \theta_N + \frac{2\pi}{\lambda}\sqrt{d_2^2 + y_N^2} \qquad (1)$$

[0069] Referring to FIG. **4**, if the device moves to a different position away from the focal plane (e.g., to a different point along the y-axis) the radiators' phases are dynamically adjusted, as described below, so as to track and maintain the target zone focused on the device. Parameter y_c represents the y-component of the device's new position, as shown in FIG. **4**, from the focal plane of the array (i.e, the plane perpendicular to the y-axis and passing through center **104** of array **100**).

$$\begin{aligned}\theta_1 + \frac{2\pi}{\lambda}\sqrt{d_2^2 + (y_1 - y_c)^2} &= \\ \theta_2 + \frac{2\pi}{\lambda}\sqrt{d_2^2 + (y_2 - y_c)^2} &= \cdots = \theta_N + \frac{2\pi}{\lambda}\sqrt{d_2^2 + (y_N - y_c)^2}\end{aligned} \qquad (2)$$

[0070] The amount of power transferred is defined by the wavelength λ of the waves being radiated by the radiators, the array span or array aperture A as shown in FIG. **1**, and the focal length, i.e. (λF/A).

[0071] In one embodiment, the distance between each pair of radiators is of the order of the wavelength of the signal being radiated. For example, if the frequency of the radiated wave is 2.4 GHz (i.e., the wavelength is 12.5 cm), the distance between each two radiators may be a few tenths to a few tens of the wavelengths, that may vary depending on the application.

[0072] An RF lens, in accordance with the present invention, is operative to transfer power wirelessly in both near-field and far field regions. In the optical domain, a near field region is referred to as the Fresnel region and is defined as a region in which the focal length is of the order of the aperture size. In the optical domain, a far field region is referred to as the Fraunhofer region and is defined as a region in which the focal length (F) is substantially greater than ($2\ A^2/\lambda$).

[0073] To transfer power wirelessly to a device, in accordance with the present invention, the radiator phases are selected so as to account for differences in distances between the target point and the radiators. For example, assume that the focal length d_1 in FIG. **2** is of the order of the aperture size A. Therefore, since distances S_1, S_2, S_3 . . . S_N are different from one another, corresponding phases θ_1, θ_2, θ_3 . . . θ_N of radiators $\mathbf{10}_1$, $\mathbf{10}_2$, $\mathbf{10}_3$ . . . $\mathbf{10}_N$ are varied so as to satisfy expression (1), described above. The size of the focus point (approximately $\lambda F/A$) is relatively small for such regions because of the diffraction limited length.

[0074] A radiator array, in accordance with the present invention, is also operative to transfer power wirelessly to a target device in the far field region where the focal length is greater than ($2A^2/\lambda$). For such regions, the distances from the different array elements to the focus spot are assumed be to be the same. Accordingly, for such regions, $S_1=S_2=S_3$ . . . $=S_N$, and $\theta_1=\theta_2=\theta_3$ . . . $=\theta_N$. The size of the focus point is relatively larger for such regions and thus is more suitable for wireless charging of larger appliances.

[0075] FIG. **5** shows an RF lens **200**, in accordance with another embodiment of the present invention. RF lens **200** is shown as including a two dimensional array of radiators $\mathbf{202}_{i,j}$ arranged along rows and columns Although RF lens **200** is shown as including 121 radiators $\mathbf{202}_{i,j}$ disposed along 11 rows and 11 columns (integers i and j are indices ranging from 1 to 11) it is understood that an RF lens in accordance with embodiments of the present invention may have any number of radiators disposed along U rows and V columns, where U and V are integers greater one. In the following description, radiators $\mathbf{202}_{i,j}$ may be collectively or individually referred to as radiators **202**.

[0076] As described further blow, the array radiators are locked to a reference frequency, which may be a subharmonic (n=1, 2, 3 . . .) of the radiated frequency, or at the same frequency as the radiated frequency. The phase of the wave radiated by each radiator are controlled independently in order to enable the radiated waves to constructively interfere and concentrate their power onto a target zone within any region in space.

[0077] FIG. **6**A is a simplified block diagram of a radiator **202** disposed in RF lens **200**, in accordance with one embodiment of the present invention. As seen, radiator **202** is shown as including, in part, a programmable delay element (also referred to herein as phase modulator) **210**, a phase/frequency locked loop **212**, a power amplifier **214**, and an antenna **216**. Programmable delay element **210** is adapted to delay signal W_2 to generate signal W_3.

[0078] The delay between signals W_2 and W_3 is determined in accordance with control signal Ctrl applied to the delay element. In one embodiment, phase/frequency locked loop **212** receives signal W_1 as well as a reference clock signal having a frequency F_{ref} to generate signal W_2 whose frequency is locked to the reference frequency Ref. In another embodiment, signal W_2 generated by phase/frequency locked loop **212** has a frequency defined by a multiple of the reference frequency F_{ref}. Signal W_3 is amplified by power amplifier **214** and transmitted by antenna **216**. Accordingly and as described above, the phase of the signal radiated by each radiator **202** may be varied by an associated programmable delay element **210** disposed in the radiator.

[0079] FIG. **6**B is a simplified block diagram of a radiator **202** disposed in RF lens **200**, in accordance with another embodiment of the present invention. As seen, radiator **202** is shown as including, in part, a programmable delay element **210**, a phase/frequency locked loop **212**, a power amplifier **214**, and an antenna **216**. Programmable delay element **210** is adapted to delay the reference clock signal F_{ref} thereby to generate a delayed reference clock signal F_{ref_Delay}. The delay between signals F_{ref} and F_{ref_Delay} is determined in accordance with control signal Ctrl applied to the delay element **210**. Signal W_2 generated by phase/frequency locked loop **212** has a frequency locked to the frequency of signal F_{ref_Delay} or a multiple of the frequency of signal F_{ref_Delay}. In other embodiments (not shown), the delay element is disposed in and is part of phase/frequency locked loop **212**. In yet other embodiments (not shown), the radiators may not have an amplifier.

[0080] FIG. **7** shows a number of components of a device **300** adapted to be charged wirelessly, in accordance with one embodiment of the present invention. Device **300** is shown as including, in part, an antenna **302**, a rectifier **304**, and a regulator **306**. Antenna **302** receives the electromagnetic waves radiated by a radiator, in accordance with the present invention. Rectifier **304** is adapted to convert the received AC power to a DC power. Regulator **306** is adapted to regulate the voltage signal received from rectifier **304** and apply the regulated voltage to the device. High power transfer efficiency is obtained, in one embodiment, if the aperture area of the receiver antenna is comparable to the size of the target zone of the electromagnetic field. Since most of the radiated power is concentrated in a small volume forming the target zone, such a receiver antenna is thus optimized to ensure that most of the radiated power is utilized for charging up the device. In one embodiment, the device may be retro-fitted externally with components required for wireless charging. In another embodiment, existing circuitry present in the charging device, such as antenna, receivers, and the like, may be used to harness the power.

[0081] FIG. **8** is a schematic diagram of RF lens **200** wirelessly charging device **300**. In some embodiments, RF lens **200** wirelessly charges multiple devices concurrently. FIG. **9** shows RF lens **200** concurrently charging devices **310**, and **315** using focused waves of similar or different strengths. FIG. **10** shows RF lens **200** wireless charging mobile devices **320**, **325** and stationary device **330** all of which are assumed to be indoor.

[0082] FIG. **11**A shows computer-simulated electromagnetic field profiles generated by a one-dimensional RF lens at a distance 2 meters away from the RF lens having an array of 11 isotropic radiators. The beam profiles are generated for three different frequencies, namely 200 MHz (wavelength 150 cm), 800 MHz (wavelength 37.5 cm), and 2400 MHz (wavelength 12.50 cm). Since the distance between each pair of adjacent radiators of the RF lens is assumed to be 20 cm, the RF lens has an aperture of 2 m. Therefore, the wavelengths are of the order of aperture size and focal length of the radiator. FIG. **11**B is a simplified schematic view of such an RF lens **500** having 11 radiators **505***k* that are spaced 20 cm apart from one another, where K is an integer ranging from 1 to 11.

[0083] Plots **510**, **520** and **530** are computer simulations of the electromagnetic field profiles respectively for 200 MHz, 800 MHz, and 2400 signals radiated by radiator **500** when the relative phases of the various radiators are selected so as to account for the path differences from each of radiators $\mathbf{505}_k$ to the point located 2 meters away from radiator $\mathbf{505}_6$ in accordance with expression (1) above. For each of these profiles, the diffraction limited focus size is of the order of the wavelengths of the radiated signal. Plots **515**, **525** and **535** are computer simulations of the electromagnetic field profiles at a distance 2 meters away from the radiator array for 200 MHz, 800 MHz, and 2400 signals respectively when the phases of radiators **505**$_k$ were set equal to one another.

[0084] As seen from these profiles, for the larger wavelength having a frequency of 200 MHz (i.e, plots **510**, **515**), because the path differences from the individual radiators to the focus point are not substantially different, the difference between profiles **510** and **515** is relatively unpronounced. However, for each of 800 MHz and 2400 MHz frequencies, the EM confinement (focus) is substantially more when the relative phases of the various radiators are selected so as to account for the path differences from the radiators **505**$_k$ to the focus point than when radiator phases are set equal to one another. Although the above examples are provided with reference to operating frequencies of 200 MHz, 800 MHz, and 2400 MHz, it is understood that the embodiments of the present may be used in any other operating frequency, such as 5.8 GHz, 10 GHz, and 24 GHz.

[0085] FIG. **12** shows the variations in computer simulated electromagnetic field profiles generated by RF lens **500**—at a distance of 2 meters away from the RF lens—as a function of the spacing between each adjacent pair of radiators. The RF lens is assumed to operate at a frequency of 2400 MHz. Plots **610**, **620**, and **630** are computer simulations of the field profiles generated respectively for radiator spacings of 5 cm, 10 cm, and 20 cm after selecting the relative phases of the various radiators to account for the path differences from various radiators **505**$_k$ to the point 2 meters away from the RF lens, in accordance with expression (1) above. Plots **615**, **625**, and **650** are computer simulations of the field profiles generated respectively for radiator spacings of 5 cm, 10 cm, and 20 cm assuming all radiators disposed in RF lens **500** have equal phases. As is seen from these plots, as the distance between the radiators increases—thus resulting in a larger aperture size—the EM confinement also increases thereby resulting in a smaller focus point.

[0086] FIG. **13**A is the computer simulation of the EM profile of an RF lens at a distance 3 meters away from an RF lens having disposed therein a two-dimensional array of Hertzian dipoles operating at a frequency of 900 MHz, such as RF lens **200** shown in FIG. **5**. The spacing between the dipole radiators are assumed to be 30 cm. The relative phases of the radiators were selected so as to account for the path differences from the radiators to the focal point, assumed to be located 3 meters away from the RF lens. In other words, the relative phases of the radiators is selected to provide the RF lens with a focal length of approximately 3 meters. The scale used in generating FIG. **13**A is −15 dB to 0 dB. FIG. **13**B shows the EM profile of FIG. **13**A using a scale of −45 dB to 0 dB.

[0087] FIG. **14**A is the computer simulation of the EM profile of the RF lens of FIGS. **13**A/**13**B at a distance 2 meters away from the focal point, i.e., 5 meters away from the RF lens. As is seen from FIG. **14**A, the radiated power is diffused over a larger area compared to those shown in FIGS. **13**A and **13**B. The scale used in generating FIG. **14**A is −15 dB to 0 dB. FIG. **14**B shows the EM profile of FIG. **14**A using a scale of −45 dB to 0 dB.

[0088] FIG. **15**A is the computer simulation of the EM profile of an RF lens at a distance 3 meters away from the RF lens having disposed therein a two-dimensional array of Hertzian dipoles operating at a frequency of 900 MHz. The spacing between the dipole radiators are assumed to be 30 cm. The relative phases of the radiators are selected so as to account for the path differences from the radiators to the focal point, assumed to be located 3 meters away from the RF lens and at an offset of 1.5 m from the focal plane of the RF lens, i.e., the focus point has a y-coordinate of 1.5 meters from the focal plane (see FIG. **4**). The scale used in generating FIG. **15**A is −15 dB to 0. FIG. **15**B shows the EM profile of FIG. **15**A using a scale of −45 dB to 0 dB.

[0089] FIG. **16**A is the computer simulation of the EM profile of the RF lens of FIGS. **15**A/**15**B at a distance 2 meters away from the focal point, i.e., 5 meters away from the x-y plane of the RF lens. As is seen from FIG. **16**A, the radiated power is diffused over a larger area compared to that shown in FIG. **15**A. The scale used in generating FIG. **16**A is −15 dB to 0 dB. FIG. **16**B shows the EM profile of FIG. **16**A using a scale of −45 dB to 0 dB. The EM profiles shown in FIGS. **13**A, **13**B, **14**A, **14**B **15**A, **15**B, **16**A, **16**B demonstrate the versatility of an RF lens, in accordance with the present invention, in focusing power at any arbitrary point in 3D space.

[0090] In accordance with one aspect of the present invention, the size of the array forming an RF lens is configurable and may be varied by using radiator tiles each of which may include one or more radiators. FIG. **17**A shows an example of a radiator tile **700** having disposed therein four radiators **15**$_{11}$, **15**$_{12}$, **15**$_{21}$, and **15**$_{22}$. Although radiator tile **700** is shown as including four radiators, it is understood that a radiator tile, in accordance with one aspect of the present invention, may have fewer (e.g., one) or more than (e.g., 6) four radiators. FIG. **17**B shown an RF lens **800** initially formed using 7 radiator tiles, namely radiator tiles **700**$_{11}$, **700**$_{12}$, **700**$_{13}$, **700**$_{21}$, **700**$_{22}$, **700**$_{31}$, **700**$_{31}$—each of which is similar to radiator tile **700** shown in FIG. **17**A—and being provided with two more radiator tiles **700**$_{23}$ and **700**$_{33}$. Although not shown, it is understood that each radiator tile includes the electrical connections necessary to supply power to the radiators and deliver information from the radiators as necessary. In one embodiment, the radiators formed in the tiles are similar to radiator **202** shown in FIG. **6**.

[0091] In accordance with one aspect of the present invention, the RF lens is adapted to track the position of a mobile device in order to continue the charging process as the mobile device changes position. To achieve this, in one embodiment, a subset or all of the radiators forming the RF lens include a receiver. The device being charged also includes a transmitter adapted to radiate a continuous signal during the tracking phase. By detecting the relative differences between the phases (arrival times) of such a signal by at least three different receivers formed on the RF lens, the position of the charging device is tracked.

[0092] FIG. **18** is a simplified block diagram of a radiator **902** disposed in an RF lens, such as RF lens **200** shown in FIG. **5**, in accordance with one embodiment of the present invention. Radiator **902** is similar to radiator **202** shown in FIG. **6**, except that radiator **902** has a receiver amplifier and phase recovery circuit **218**, and a switch S_1. During power transfer, switch S_1 couples antenna **216** via node A to power amplifier **214** disposed in the transmit path. During tracking, switch S_1 couples antenna **216** via node B to receiver amplifier and phase recovery circuit **218** disposed in the receive path to receive the signal transmitted by the device being charged.

[0093] FIG. **19** shows a number of components of a device **900** adapted to be charged wirelessly, in accordance with one embodiment of the present invention. Device **900** is similar to device **300** shown in FIG. **7**, except that device **900** has a transmit amplifier **316**, and a switch S_2. During power transfer, switch S_2 couples antenna **302** via node D to rectifier **304** disposed in receive path. During tracking, switch S_2 couples antenna **302** via node C to transmit amplifier **316** to enable the transmission of a signal subsequently used by the RF lens to detect the position of device **300**. FIG. **20** shows RF lens **200** tracking device **900** by receiving the signal transmitted by device **900**.

[0094] A radiator, in accordance with any of the embodiments of the present invention, in addition to transferring RF power to a device wirelessly, may also wirelessly transfer modulated data to such a device. For example, a radiator, in accordance with the embodiments of the present invention, may operate as a wireless local area network (WLAN) router to direct signal power toward other such routers or receivers (WLAN or otherwise) to increase the received signal power by orders of magnitude and thereby increase the range, coverage and wireless data rates, as well as reduce the effect of multi-path propagation of the RF signal and power. Transferring wireless signal and/or power, in accordance with embodiments of the present invention and in conformity with any communication standards, such as WiFi, Zigbee, Bluetooth, GSM, GPRS, Edge, and UMTS, to a receiving device, and tracking the location of the device as it moves through physical space, greatly increases the range of the data link and achievable data rates.

[0095] Because the power levels required for data transmission are typically orders of magnitude lower than the power levels for wireless power transfer, in accordance with embodiments of the present invention, both power and data transmission may be performed concurrently. In other words, for any given amount of emitted power by a generation unit, in accordance with embodiments of the present invention, the range of wireless power transfer is smaller than the range for wireless data transmission. Therefore, a device that is in range for wireless power transfer, is also in range for wireless data transmission. Furthermore, the amount of power siphoned off in detecting the transmitted data signal is typically much smaller compared to the power available for concurrently powering the device. In addition, a device that may be charged wirelessly can also operate over the relatively narrow range of frequencies around the center frequency spanned by most signal modulation schemes used in data transmission.

[0096] FIG. **21**A is a simplified high-level block diagram of a device **750** (e.g., cellular phone, camera, wearable device, toy), in accordance with one embodiment of the present invention. Device **750** is shown as including, in part, an antenna **760**, a directional power splitter/coupler (alternatively referred to herein as coupler) **752**, a power recovery unit **754**, and a radio signal receiver/demodulator (alternatively referred to herein as receiver) **756**. Antenna **760** is adapted to receive a signal from a radiator or an array of radiators, as described for example above, and deliver the received signal to coupler **752**. The signal received by the antenna is a modulated signal that carries information/data as well as RF power, both of which are transmitted using the same frequency. Coupler **752** is adapted to transfer a first portion of the received RF signal to receiver **756** and a second portion of the received signal to power recovery unit **754**. Receiver **756** demodulates the received signal to recover the transmitted data. Power recovery unit **754** is adapted to convert the received RF signal to a DC power to power device **750** or charge device **750**'s battery. Although not shown, it is understood that device **750** may use filters and/or transistors instead of coupler **752** to distribute the received RF power between power recovery unit **754** and receiver **756**. U.S. application Ser. No. 14/552,249, filed Nov. 24, 2014, entitled “Active CMOS Recovery Units For Wireless Power Transmission”, the content of which is incorporated herein by reference in its entirety, describes in detail various embodiments of power recovery unit **754**. In one embodiment, directional coupler/splitter **752** is adjustable (via a controller not shown in FIG. **21**A) that can vary the split ratio so as to change the amount of power coupler/splitter **752** provides to receiver **756** relative to the power it provides to power recovery unit **754**.

[0097] FIG. **21**B is a simplified high-level block diagram of a device **755**, in accordance with one embodiment of the present invention. Device **750** is shown as including, in part, an antenna **760**, a controller **762**, a power recovery unit **754**, and a radio signal receiver/demodulator (hereinafter receiver) **756**. Antenna **760** is adapted to receive a signal from a radiator or an array of radiators, as described above, and deliver the received signal to controller **762**. Controller **762** is adapted to distribute the received RF power in accordance with an inverse ratio of the impedances of receiver **756** and power recovery unit **754**. Controller **762** may be further adapted to vary the impedances seen at its output terminals such that the output impedance having a closer value to the impedance seen at the input of controller **762** receives a higher amount of the received RF power.

[0098] FIG. **21**C is a simplified high-level block diagram of a device **760**, in accordance with another embodiment of the present invention. Device **760** is similar to device **750** except that device **760** uses a switch **764** in place of coupler **752**. When switch **764** is in position P_1, all of the received RF power is delivered to power recovery unit **754** thereby to charge device **760**. When switch **764** is in position P_2, all of the received RF power is delivered to receiver **756** to receive the transmitted data. For example, if device **760** is determined as being out-of-range for wireless power charging, switch **764** is placed in position P_2 to supply all of the received wireless power to receiver **756** to improve, e.g., its signal to noise ratio. Likewise, if device **760** is detected as being within range but requiring no additional charge, switch **764** is placed in position P_2 to supply all of the received wireless power to receiver **756**. If, for example, data reception is not of interest, switch **764** is placed in position P_1 to supply all of the received RF power to power recovery unit **754**.

[0099] For example, assume that an RF lens or a radiator array (a power generating unit) broadcasts an RF signal to determine if the received RF signal is strong enough to wirelessly charge a device receiving the RF signal. If the received RF signal is detected as not being strong enough to wirelessly charge the device, a controller (not shown in FIG. **21**C) controlling switch **764** may switch to radio receiving mode by placing switch **764** to position P_2. Once in the radio receiving mode, the device may then communicate with the radiator array to enable the radiator array to focus its RF power on the device. After the radiator array focuses its transmitted RF signal on device **760**, thereby causing the received RF signal to be sufficient for wireless charging, the controller may place switch **764** in position P_1 to enable power recovery unit **754** to charge device **760**. When device

760 is charged to a sufficient level, it may so inform the controller thus causing switch **764** to be placed in switch position P_2. It is understood that any number of conditions and criteria may be used to change the position of switch **764**.

[0100] FIG. **21**D is a simplified high-level block diagram of a device **765**, in accordance with another embodiment of the present invention. Device **765** is shown as including, in part, an antenna **760**, a coupler **752**, a switch **764**, a power combiner **758**, a power recovery unit **754**, and a receiver **756**. Antenna **760** is adapted to receive a modulated signal that carries information as well as RF power, both of which are transmitted using the same frequency. Coupler **752** is adapted to transfer a first portion of the received RF signal to power combiner **758**, and a second portion of the received signal to switch **764**. When switch **764** is position P_1, (i) the first portion of the power supplied by coupler **752** is delivered to receiver **756** via power combiner **758**, and (ii) the second portion of the power supplied by coupler **752** is delivered to power recovery unit **754** to wirelessly charge device **765**. Receiver **756** demodulates the received signal to recover the transmitted data.

[0101] When switch **764** is position P_2, power combiner **758** combines the first portion of the power it receives from coupler **752** with the second portion of power it receives from switch **764** and delivers the combined power to receiver **756**. Accordingly, in such embodiments, receiver **756** continues to receive the transmitted RF signal regardless of the switch position. For example, when the received RF power is detected as being below a threshold value, switch **764** is placed in position P_2 (via a controller not shown) so that all the received RF power is used for signal reception. When the received RF power is detected as being above a threshold value and/or device **765** requests to be charged, switch **764** is placed in position P_1 (via the controller) so that a relatively small fraction of the received RF power is used for signal detection by receiver **756**, and the remainder of the RF power is used by power recovery unit **754** to charge device **765**.

[0102] FIG. **21**E is a simplified high-level block diagram of a device **775**, in accordance with another embodiment of the present invention. Device **775** is shown as including, in part, an antenna **760**, a coupler **752**, a switch **764**, a power combiner **758**, a power recovery unit **754**, and a receiver **756**. Antenna **760** is adapted to receive a modulated signal that carries information/data as well as RF power, both of which are transmitted using the same frequency. Receiver **756** demodulates the received signal to recover the transmitted data.

[0103] When switch **764** is position P_2, the received RF signal is supplied to directional coupler **752**, which in turn, delivers (i) a first portion of the received power to receiver **756** via power combiner **758**, and (ii) a second portion of the received power to power recovery unit **754** for wirelessly charging device **775**. When switch **764** is in position P_1, substantially all of the received RF power is delivered to receiver **756** via power combiner **758**. Accordingly, device **775** is adapted to continuously deliver the RF signal to receiver **756**.

[0104] FIG. **21**F is a simplified high-level block diagram of a device **785**, in accordance with another embodiment of the present invention. Device **785** is shown as including, in part, an antenna **760**, an adjustable power splitter/coupler **782**, a control unit **762**, a power recovery unit **754**, and a receiver **756**. Antenna **760** is adapted to receive a modulated signal that carries information as well as RF power, both of which are transmitted using the same frequency. Adjustable power splitter/coupler **782** is adapted to split the RF signal it receives from antenna **760** into first and second portions in accordance with a split ratio adjustable power splitter/coupler **782** receives from control unit **762**. The split ratio supplied by control unit **762** may be defined by any number of factors, such as the target date rate set by receiver **756**, the DC power requested by the power recovery unit **754**, the power requested by receiver **756**, and the like.

[0105] Referring to FIGS. **21**A-**21**E, it is understood that power combiner **758** may be formed using any arrangement or configuration of active and/or passive devices capable of combining power from two or more sources. Typical examples include lossless power combiners such as coupled transmission lines, their lumped circuit equivalent using capacitors or inductors, magnetic transformers, and other circuits, as is well known. Directional coupler **752** may include coupled transmission lines, transformers and/or their lumped circuit equivalent. Switch **764** adapted to switch RF power may include pass gates, electronically tunable impedance networks, relays and the like, as is also well known.

[0106] In accordance with another embodiment of the present invention, a pulse based measurement technique is used to track the position of the mobile device. To achieve this, one or more radiators forming the RF lens transmit a pulse during the tracking phase. Upon receiving the pulse, the device being tracked sends a response which is received by the radiators disposed in the array. The travel time of the pulse from the RF lens to the device being tracked together with the travel times of the response pulse from the device being tracked to the RF lens is representative of the position of the device being tracked. In the presence of scatterers, the position of the device could be tracked using such estimation algorithms as maximum likelihood, or least-square, Kalman filtering, a combination of these techniques, or the like. The position of the device may also be determined and tracked using WiFi and GPS signals.

[0107] The presence of scattering objects, reflectors and absorbers may affect the RF lens' ability to focus the beam efficiently on the device undergoing wireless charging. For example, FIG. **22** shows an RF lens **950** transferring power to device **300** in the presence of a multitude of scattering objects **250**. To minimize such effects, the amplitude and phase of the individual radiators of the array may be varied to increase power transfer efficiency. Any one of a number of techniques may be used to vary the amplitude or phase of the individual radiators.

[0108] In accordance with one such technique, to minimize the effect of scattering, a signal is transmitted by one or more of the radiators disposed in the RF lens. The signal(s) radiated from the RF lens is scattered by the scattering objects and received by the radiators (see FIG. **18**). An inverse scattering algorithm is then used to construct the scattering behavior of the environment. Such a construction may be performed periodically to account for any changes that may occur with time. In accordance with another technique, a portion or the entire radiator array may be used to electronically beam-scan the surroundings to construct the scattering behavior from the received waves. In accordance with yet another technique, the device undergoing wireless charging is adapted to periodically send information about the power it receives to the radiator. An optimization algorithm then uses the received information to account for scattering so as to maximize the power transfer efficiency.

[0109] In some embodiments, the amplitude/phase of the radiators or the orientation of the RF lens may be adjusted to take advantage of the scattering media. This enable the scattering objects to have the proper phase, amplitude and polarization in order to be used as secondary sources of radiation directing their power towards the device to increase the power transfer efficiency.

[0110] The above embodiments of the present invention are illustrative and not limitative. Embodiments of the present invention are not limited by number of radiators disposed in an RF lens, nor are they limited by the number of dimensions of an array used in forming the RF lens. Embodiments of the present invention are not limited by the type of radiator, its frequency of operation, and the like. Embodiments of the present invention are not limited by the type of device that may be wirelessly charged. Embodiments of the present invention are not limited by the type of substrate, semiconductor, flexible or otherwise, in which various components of the radiator may be formed. Other additions, subtractions or modifications are obvious in view of the present disclosure and are intended to fall within the scope of the appended claims.

What is claimed is:

1. A device comprising:

an antenna adapted to receive an RF signal comprising modulated data;

a splitter/coupler coupled to the antenna and adapted to split the received RF signal into first and second portions;

a receiver adapted to demodulate the data from the first portion of the RF signal; and

a power recovery unit adapted to convert the second portion of the RF signal to a DC power to power the device.

2. A device comprising:

an antenna adapted to receive an RF signal comprising modulated data;

a receiver adapted to demodulate the data from a first portion of the RF signal;

a power recovery unit adapted to convert a second portion of the RF signal to a DC power to power the device; and

a controller adapted to receive the RF signal from the antenna and generate the first and second portions of the RF signal in accordance with impedance values of the receiver and the power recovery unit.

3. A device comprising:

an antenna adapted to receive an RF signal comprising modulated data;

a switch adapted to receive the RF signal from the antenna;

a power recovery unit adapted to convert the RF signal to a DC power to power the device when the switch is in a first position; and

a receiver adapted to demodulate the data from the received RF signal when the switch is in a second position.

4. A device comprising:

an antenna adapted to receive an RF signal comprising modulated data;

a splitter/coupler coupled to the antenna and adapted to split the received RF signal into first and second portions;

a switch adapted to receive the second portion of the RF signal from the splitter/coupler;

a power recovery unit adapted to convert the second portion of the RF signal to a DC power to power the device when the switch is in a first position; and

a power combiner adapted to receive the first portion of the RF signal from the splitter/coupler and further to receive the second portion of the RF signal when the switch is in a second position; and

a receiver adapted to demodulate the data from an output signal of the power combiner.

5. A device comprising:

an antenna adapted to receive an RF signal comprising modulated data;

a switch coupled to the antenna to receive the RF signal therefrom;

a power combiner coupled to a first output terminal of the switch to receive the RF signal when the switch is in a first position;

a splitter/coupler coupled to a second output terminal of the switch to receive the RF signal when the switch is in a second position, said splitter adapted to split the RF signal into a first portion and a second portion, said splitter/coupler delivering the first portion of the RF signal to the power combiner;

a power recovery unit adapted to convert the second portion of the RF signal to a DC power to charge the device when the switch is in the second position;

a receiver adapted to demodulate the data from an output signal of the power combiner.

6. The device of claim **5** further comprising:

a controller adapted to cause the switch to be in the first position when a power of the received RF signal is less than a threshold value.

7. The device of claim **5** further comprising:

a controller adapted to cause the switch to be in the first position when the device indicates that its DC power exceeds a threshold value.

8. The device of claim **4** further comprising:

a controller adapted to cause the switch to be in the first position when a power of the received RF signal is less than a first threshold value.

9. The device of claim **4** further comprising:

a controller adapted to cause the switch to be in the first position when the device indicates that its DC power exceeds a second threshold value.

10. A method comprising:

receiving an RF signal comprising modulated data;

splitting the received RF signal into first and second portions;

demodulating the data from the first portion of the RF signal; and

converting the second portion of the RF signal to a DC power.

11. A method comprising:

receiving an RF signal comprising modulated data;

demodulating the data from a first portion of the received RF signal via a receiver;

converting a second portion of the RF signal to a DC power via a power recovery unit;

generating the first and second portions of the RF signal in accordance with impedance values of the receiver and the power recovery unit

12. A method comprising:

receiving an RF signal comprising modulated data;

converting the RF signal to a DC power when a switch is in a first position; and

demodulating the data from the received RF signal when the switch is in a second position.

13. A method comprising:

receiving an RF signal comprising modulated data;

demodulating the data using either a first portion of the received RF signal or the received RF signal; and

converting a second portion of the RF signal to a DC power when the first portion of the RF signal is used for demodulating the data.

14. The method of claim **13** further comprising:

demodulating the data using the received RF signal when a power of the received RF signal is less than a first threshold value.

15. The method of claim **14** further comprising:

demodulating the data using the received RF signal when an indication is received that a battery charge exceeds a threshold value.

16. A device comprising:

an antenna adapted to receive an RF signal comprising modulated data;

a controller;

an adjustable splitter/coupler adapted to split the received RF signal into first and second portions in accordance with a value said adjustable splitter/coupler receives from the controller;

a receiver adapted to demodulate the data from the first portion of the RF signal; and

a power recovery unit adapted to convert the second portion of the RF signal to a DC power to power the device.

17. The device of claim **16** wherein said value is defined by a target data rate.

18. The device of claim **16** wherein said value is defined by a DC power requirement of the device.

19. A method comprising:

receiving an RF signal;

splitting the received RF signal into first and second portions in accordance with a received value;

demodulating the data from the first portion of the RF signal; and

converting the second portion of the RF signal to a DC power.

20. The method of claim **19** wherein said value is defined by a target data rate.

21. The device of claim **16** wherein said value is defined by a DC power requirement.

22. The device of claim **1** wherein said splitter/coupler is an adjustable splitter/coupler.

* * * * *

THE SMITHSONIAN INSTITUTION

The Smithsonian Institution is home to more than 141 million objects, ranging in size from insects and diamonds to locomotives and spacecraft. It is the world's largest museum complex, comprising 15 museums and galleries and the National Zoo in Washington DC, and two additional museums in New York City. Millions of visitors each year visit the nation's capital to view such treasures as the Hope Diamond, the Star Spangled Banner, and the Wright Flyer. A broad range of exhibits ensures a fun and educational experience for young and old alike.

One of the world's leading scientific research centers, the Institution has facilities in eight states and the Republic of Panama. Research projects in the arts, history and science are carried out by the Smithsonian all over the world. Some of the Smithsonian's research centers include the Smithsonian Astrophysical Observatory in Cambridge, Massachusetts, the Smithsonian Marine Station at Link Port, in Florida, and the Smithsonian Tropical Research Institute, in Panama.

For membership information of pre-visit planning material, write or call the Visitor Information and Associates Reception Center, Smithsonian Institution, Washington, D. C., 20560, (202) 357-2700 (voice), (202) 357-1729 (TTY). You may also visit the Smithsonian through our web site, www.si.edu.

HISTORY

James Smithson (1765 –1829), a British scientist, drew up his will in 1826 naming his nephew, Henry James Hungerford, as beneficiary. Smithson stipulated that, should the nephew die without heirs (as he did in 1835), the estate would go to the United States to found "at Washington, under the name of the Smithsonian Institution, an establishment for the increase and diffusion of knowledge..."

On July 1, 1836, Congress accepted the legacy bequeathed to the nation by James Smithson, and pledged the faith of the United States to the charitable trust. In 1838, following approval of the bequest by the British courts, the United States received Smithson's estate—bags of gold sovereigns—then the equivalent of $515,169. Eight years later, on August 10, 1846, an Act of Congress signed by President James K. Polk, established the Smithsonian Institution in its present form and provided for the administration of the trust, independent of the government itself, by a Board of Regents and Secretary of the Smithsonian.

SMITHSONIAN MUSEUMS, GALLERIES AND ZOOS

Smithsonian Institution [illegible]
Anacostia Museum
Arthur M. Sackler Gallery
Arts and Industries Building
Cooper-Hewitt, National Design Museum
Freer Gallery of Art
Hirshhorn Museum and Sculpture Garden
National Air and Space Museum
National Museum of African Art
National Museum of American History, Behring Center
National Museum of the American Indian
National Museum of Natural History
National Portrait Gallery
National Postal Museum
National Zoological Park
Renwick Gallery
S. Dillon Ripley Center
Smithsonian American Art Museum

SMITHSONIAN®

CRYSTAL RADIO™

WARNING!
ONLY FOR USE BY CHILDREN
OVER 8 YEARS OLD.

READ THE INSTRUCTIONS BEFORE USE,
FOLLOW THEM AND
KEEP THEM FOR REFERENCE.
STORE THE SET OUT OF REACH
OF SMALL CHILDREN.

PLEASE KEEP A NOTE OF OUR NAME AND ADDRESS DETAILS FOR FUTURE REFERENCE. IN EUROPE CONTACT:

CE

NSI SIMM GAmbH
D 76162 KARLSRUHE
GERMANY
49-0721-9584-116

PLEASE BE SURE TO READ THE ADVISE FOR SUPERVISING ADULTS AND THE SAFETY RULES CONTAINED IN THIS BOOKLET.

©2002 SMITHSONIAN® INSTITUTION
NATURAL SCIENCE INDUSTRIES, LTD.
910 ORLANDO AVE.
WEST HEMPSTEAD, N.Y. 11552

PRINTED IN USA

ITEM NO. 3261-08

ADVICE FOR SUPERVISING ADULTS:

- READ AND FOLLOW THESE SAFETY INSTRUCTIONS, THE SAFETY RULES AND KEEP THEM FOR REFERENCE.

- SUPERVISING ADULTS SHOULD DISCUSS ANY WARNINGS AND SAFETY INFORMATION WITH THE CHILD BEFORE COMMENCING THE ACTIVITIES.

SAFETY RULES:

- DO READ THE INSTRUCTIONS BEFORE USE, FOLLOW THEM AND KEEP THEM FOR REFERENCE.
- DO KEEP YOUNG CHILDREN AND ANIMALS AWAY FROM THE ACTIVITY AREA.
- DO STORE THE SET OUT OF REACH OF YOUNG CHILDREN.

Facts about early crystal development

Several inventors receive credit for developing a "crystal detector", a device which can pass current better in one direction than the other in an electrical circuit. In Germany, in the 1870's, Karl F. Braun noticed this property in certain mineral substances. Commercial development required a technology that would be suitable. This came with the advent of wireless communication and radio after 1900.

In Japan, Wichi Torikata investigated many minerals, including zincite, pyrolusite, iron pyrites and galena. In the United States, Greenleaf W. Pickard, associated with the Wireless Specialty Apparatus Co. of Boston, conducted extensive experiments as well. He became well known for the PERIKON Detector which employed zincite and chalcopyrite. The early 20th century work was generally done between 1900 and 1912. Included among the scientists is General H.H.C. Dunwoody of the U.S.Army who developed a carborundum detector.

Until radio broadcasting began in 1920 (Station KDKA Pittsburgh) experimenters utilized the crystal detector for radiotelephone communications and for reception of certain Morse Code signals. There was also extensive use by the U.S.Navy and other maritime services.

The modern crystal diode, such as provided with your set, is a spin-off of radar technology developed during World War II. The basic principle, however, remains the same, i.e. to remove the AUDIO (speech, music or Morse-Code) from the radio frequency carrier wave. In one early Pickard crystal design (left) a stiff metal point is adjustable over the crystal surface. In a later version a "Catswhisker" spring impinged on a piece of galena (right) and was varied to give a louder signal.

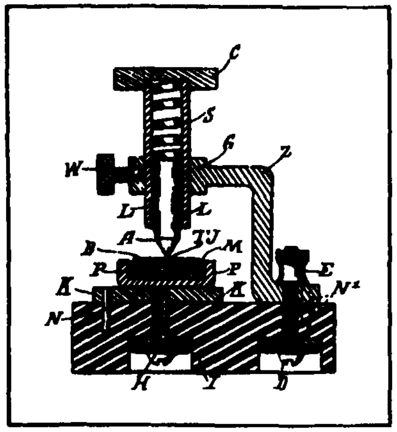

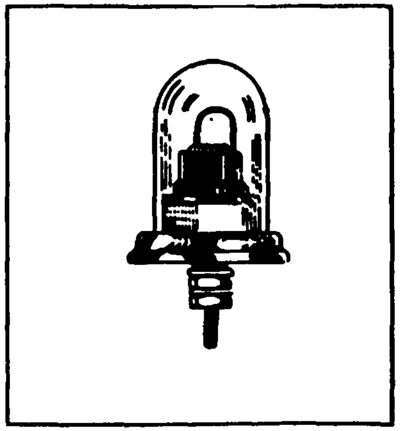

Your Smithsonian Smart Lab Crystal Radio Set contains the following items

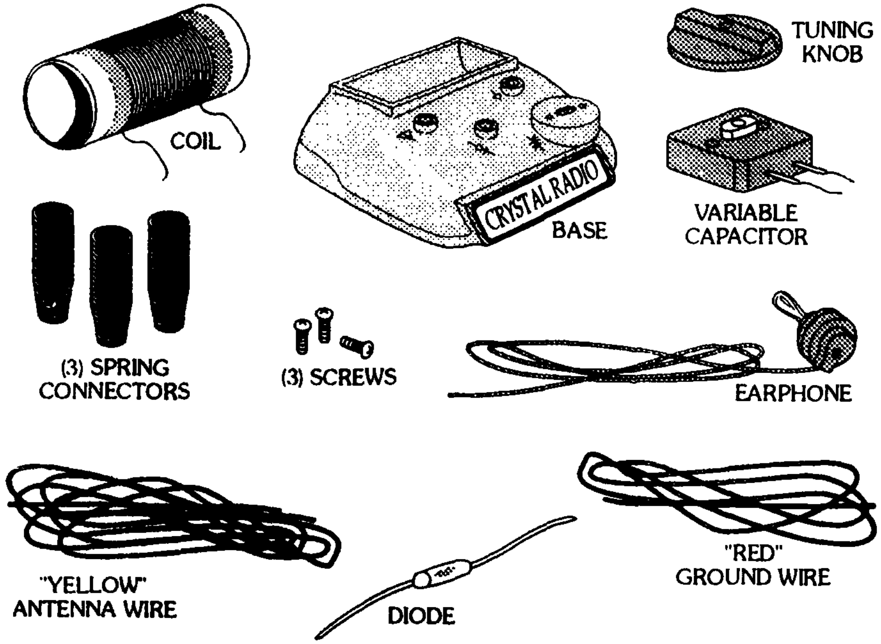
COIL
TUNING KNOB
CRYSTAL RADIO
BASE
VARIABLE CAPACITOR
(3) SPRING CONNECTORS
(3) SCREWS
EARPHONE
"YELLOW" ANTENNA WIRE
DIODE
"RED" GROUND WIRE

Shown below is a typical wiring diagram of your crystal radio.

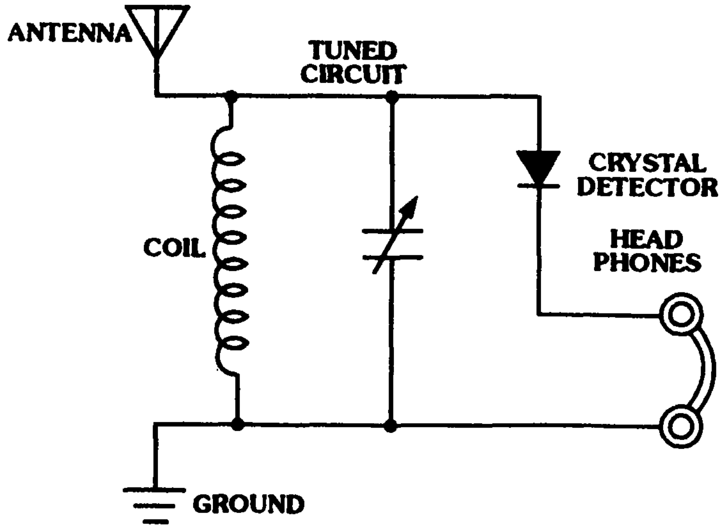
ANTENNA
TUNED CIRCUIT
CRYSTAL DETECTOR
HEAD PHONES
COIL
GROUND

Radio Technology (cont'd)

When these two different energy waves are put together by the mixer stage of the transmitter, the resulting wave, which is the signal that the station broadcasts, looks like this:

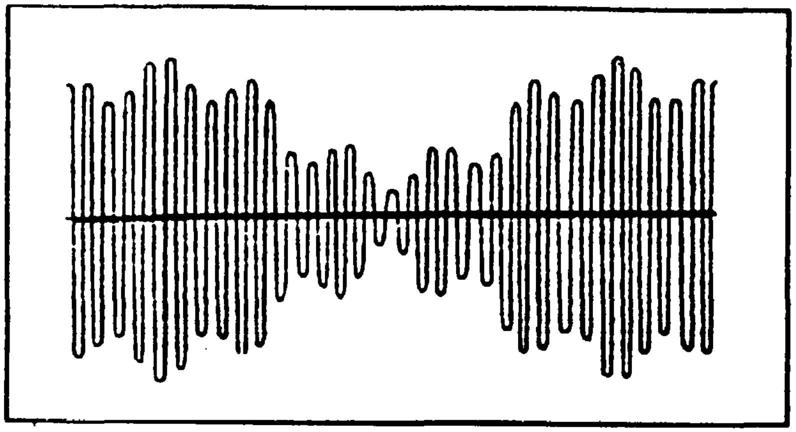

AM Broadcast Wave

The signal is called a MODULATED wave. You will notice that each of the waves is the same length, but the heights vary. Since the height of a wave is called its AMPLITUDE, The type of transmission that WXYZ uses is called AMPLITUDE MODULATION. That's why WXYZ is called an AM station.

The crystal radio that you built works in just the opposite way as the radio station. The modulated signal broadcast by WXYZ is "picked up" by the antenna on your radio.

The antenna is connected to the tuner circuit of the radio, which is made up of a coil of wire and a variable capacitor connected together. As you move the tuning knob, you are able to select, or isolate, the particular

This signal is then DEMODULATED by the diode, separating the voice (audio) wave from the carrier wave.

The voice (audio) wave then moves to the earphone, where it is changed back into sound waves that you can hear.

Learn the basic electronic symbols

Before starting to assemble your Smart Lab Crystal Radio please learn the basic electronic symbols that are shown on the plastic base.

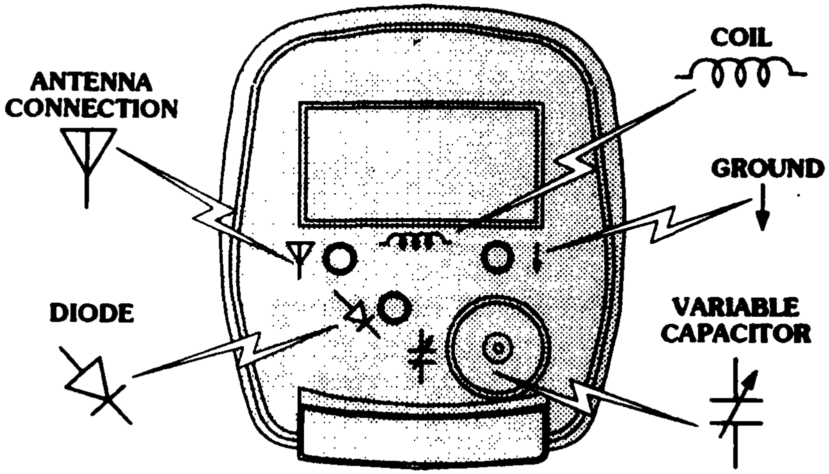

Explanation of components

Variable Capacitor - It is used to tune the radio to a station. The leads that are soldered to the lugs are used to connect it to the circuit.

Diode - A small crystal is sealed inside with leads connected to it. Pay careful attention to the position of the black bands painted on the diode close to one end.

Coil- This is a radio-tuning COIL. It was made by winding enameled copper wire around a paper core 80 times. The leads have been stripped and tinned so it can be connected to the circuit.

Earphone - It contains a small crystal that can make enough electricity to drive a metal diaphragm to produce sound. The leads have been stripped and tinned so it can be connected to the circuit.

Antenna - This is a wire used for radiating or receiving radio waves.

Ground - This is a wire used to make a electric connection with the earth or other type of grounding source to create a common return for an electric circuit.

DEAR CUSTOMER:

If we made an error and left something out of this set, or if something is damaged, we are sorry and wish to correct our error.
Please DO NOT return the set to the store where purchased, as the store does not have replacement parts. Instead, write us:

Please include:
1. **Date of Purchase**
2. **Where purchased**
3. **Price paid**
4. **Model Number**
5. **Name of item**
6. **Brief Description of Problem**
7. **Include Sales Slip**

We will do our best to satisfy you. Send your letter to:

NATURAL SCIENCE INDUSTRIES, LTD.
910 ORLANDO AVENUE
WEST HEMPSTEAD, N.Y. 11552-3942
ATTENTION: QUALITY CONTROL DEPT.

How to assemble your Crystal Radio

Before starting, it's important to assemble the parts correctly. Carefully read the following instructions step by step and have fun . . . half the fun is knowing you made your very own crystal radio.

Step ①

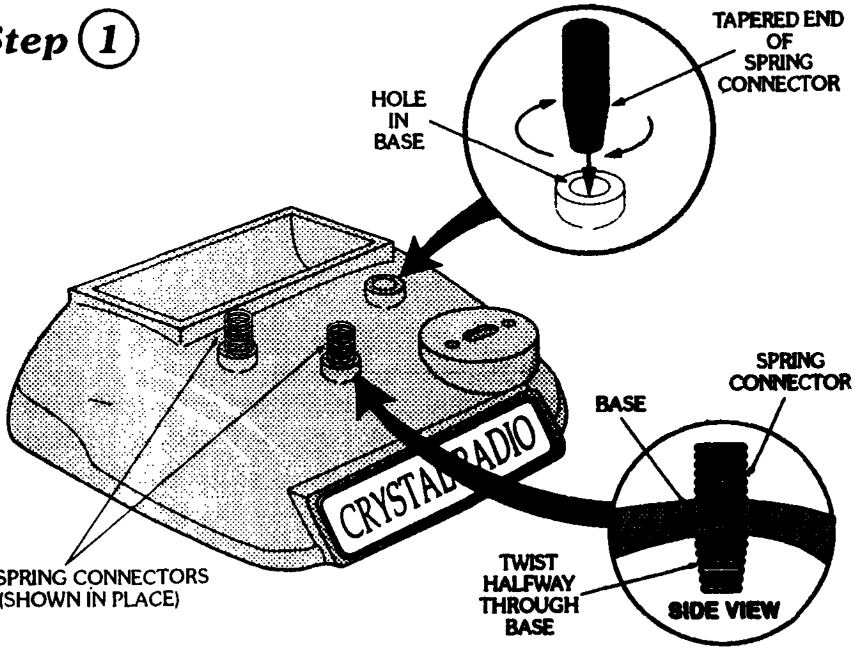
TAPERED END OF SPRING CONNECTOR
HOLE IN BASE
SPRING CONNECTOR
BASE
CRYSTAL RADIO
(2) SPRING CONNECTORS (SHOWN IN PLACE)
TWIST HALFWAY THROUGH BASE
SIDE VIEW

A. Locate the three Spring Connectors.
B. Insert one of the Spring Connectors (tapered end first) into one of the three holes on the Base.
C. Push down on the Spring Connector and twist to the right until the Connector is approximately half way through the base.

NOTE: Check to see if the Spring Connecter is inserted halfway through the Base by turning Base upside down (see insert).
D. Insert the remaining two Spring Connectors into the holes on Base the same way.

Radio Technology

Now that you have your crystal radio assembled and working, it is time to take a brief look at how it works. To do this you are going to take an imaginary trip to a radio station WXYZ, in anytown,USA.

When you arrive at WXYZ, you are met by the station's general manager, Mr. Smith, who is going to show you around.

First, Mr. Smith takes you to a STUDIO, a special room that a program comes from. There is a lot of equipment, including dials and switches, record players, lots and lots of records, and microphones.

Mr. Smith tells you that when you talk into one of the microphones, your words go into the station's electronic equipment. There, it is mixed with the station's carrier wave, and sent out into the air through the station's transmitting antenna - that tall tower you saw on top of the radio station building.

Then Mr. Smith takes a piece of paper and a pencil, and draws diagrams to show you what happens. He begins by saying that all energy travels in waves, and since the sound we make when we talk is a form of energy, it might look something like this:

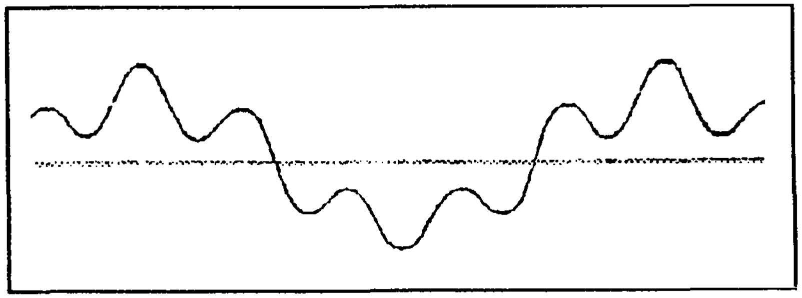

Sound Wave

The station's carrier wave is a radio wave, which is also a form of energy, which looks something like this:

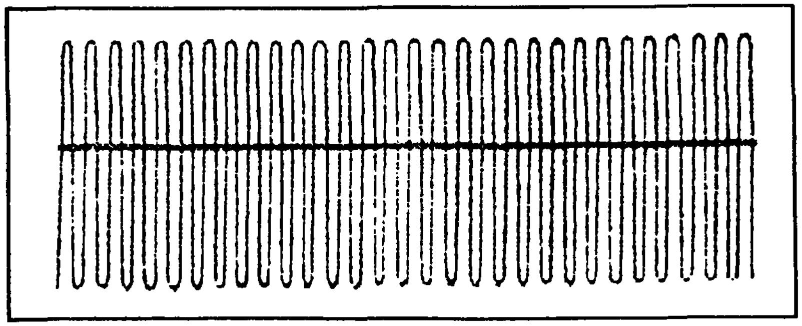

Radio Carrier Wave

How to properly operate your Smart Lab Crystal Radio and troubleshoot

After you have attached all the wires, carefully read the following instructions to make sure your crystal radio operates properly.

A. Place the earphone in your ear and turn the Tuning Knob clockwise and counterclockwise until you pick up the strongest signal.
Note: If you do not pick up a strong signal or do not hear anything, please read some of the troubleshooting solutions explained below.

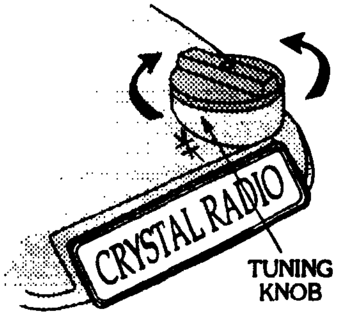

TROUBLESHOOTING HINTS

1. There may be a bad connection caused by improper assembly of your radio. Check all the spring connectors to make sure the wires are attached correctly. If a wire is loose or the bare wire is not making good contact with the spring connector, the radio may not work.
2. If the uninsulated part of the Antenna wire touches anything other than the Antenna Connection on the radio, the radio may work improperly or not at all..
3. If the bare stripped end of the Ground wire is not wrapped tightly around a water pipe (so that it makes good contact), try taping the wire to the shiny part of the pipe using duct tape. If the pipe is dull or rusty, use a piece of sandpaper to gently sand the area where the wire makes contact. Also make sure the wire doesn't touch anything other than the Ground Connection or the radio may not work properly.
4. You may live in an area where radio reception is generally poor. Instead of trying to use your radio during the day, try at night when many radio stations are received better.
5. The Antenna on your radio is a very important component. Make the Antenna as long and high above the ground as possible. If you live in a multi-story dwelling the highest floor should be used when operating the radio. Insulated bell-wire, which is available in hardware stores, makes a very good substitution for your present Antenna.

How to assemble your Crystal Radio (cont'd)

For this step you will need the variable capacitor, (3) screws and turning knob. Be carefull not to misplace the small screws when assembling.

Step ②

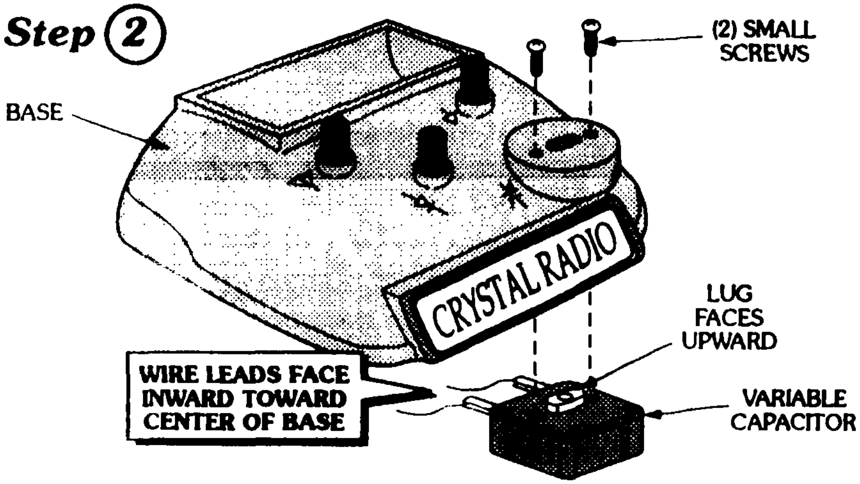

A. First place the Variable Capacitor up into the base so the holes on the capacitor line up with the holes on the base.
Note: When you do this be sure the wire leads face inward toward the base as shown and lug faces upward.

B. Next insert the two small screws into the holes on top of the Base and tighten using a small phillips screwdriver.

Step ③

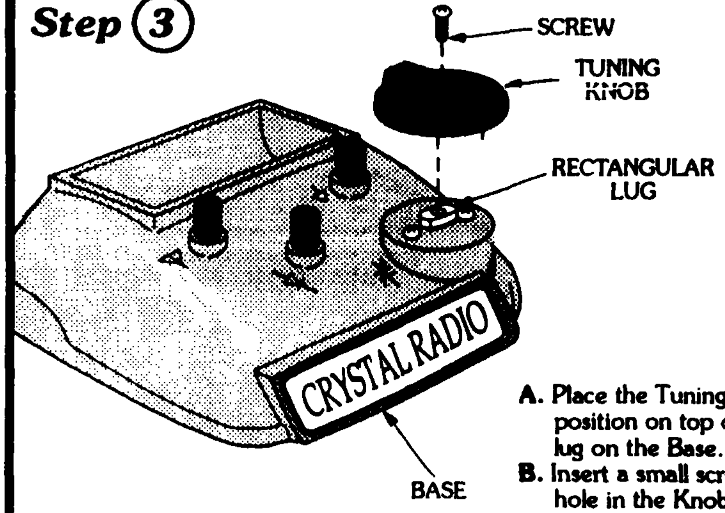

A. Place the Tuning Knob into position on top of the rectangular lug on the Base.
B. Insert a small screw into the hole in the Knob and tighten using a phillips screwdriver.

How to assemble your Crystal Radio (cont'd)

Step 4 For this step turn the Base upside down.

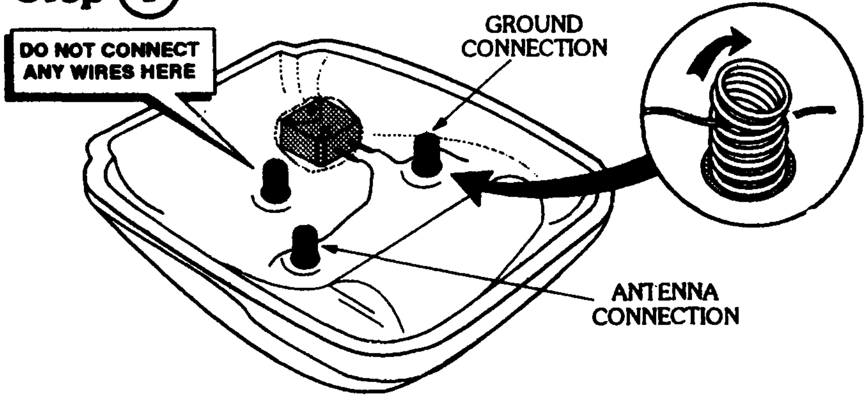
DO NOT CONNECT ANY WIRES HERE
GROUND CONNECTION
ANTENNA CONNECTION

A. Turn the Base upside down.
B. Connect one wire to the Antenna spring connector by bending the spring slightly to one side and inserting the wire through as shown in inset.
C. Connect the other wire to the Ground spring connector in the same manner.

NOTE: Do not connect any of the wires to the Diode spring connector or your crystal set will not work properly.

Step 5

NOTE: Make sure the Plated Ends of the Coil wires are making contact to the Spring Connections.

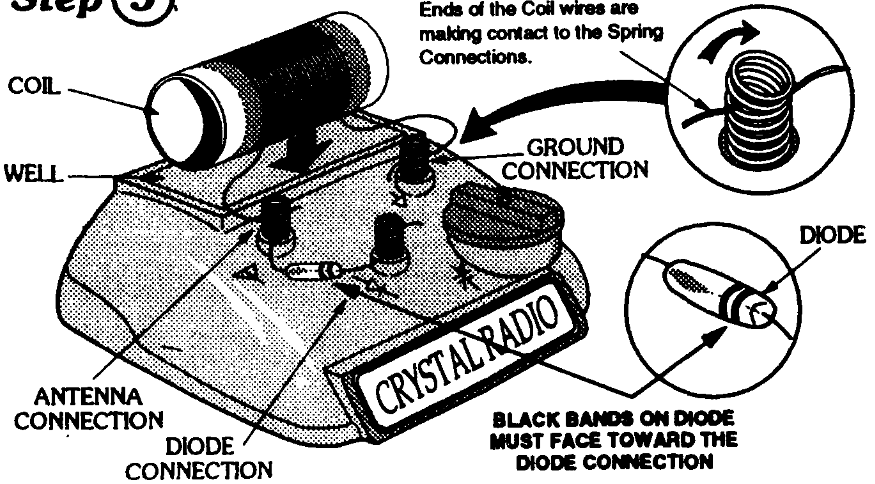
COIL
WELL
GROUND CONNECTION
DIODE
CRYSTAL RADIO
ANTENNA CONNECTION
DIODE CONNECTION

BLACK BANDS ON DIODE MUST FACE TOWARD THE DIODE CONNECTION

A. Next place the Coil in the well of Base with the wires hanging loosely toward the connectors .
B. Connect one of the plated ends of Coil wires to the Ground Connector by bending the Spring Connector as shown in inset. Connect the other Coil wire to the Antenna Connector.
C. Place the Diode between the Diode and Antenna Connections carefully noting the black bands face toward the Diode Connection.
D. Connect ends of Diode wires to Antenna and Diode spring connectors shown.

How to attach the earphones

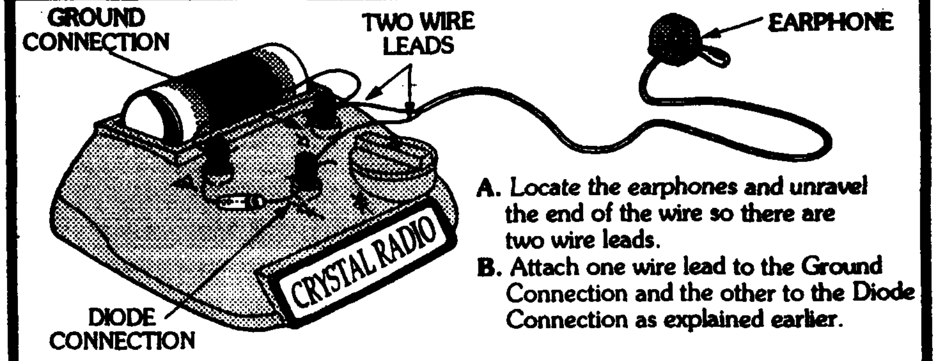
GROUND CONNECTION
TWO WIRE LEADS
EARPHONE
CRYSTAL RADIO
DIODE CONNECTION

A. Locate the earphones and unravel the end of the wire so there are two wire leads.
B. Attach one wire lead to the Ground Connection and the other to the Diode Connection as explained earlier.

Final assembly and operation

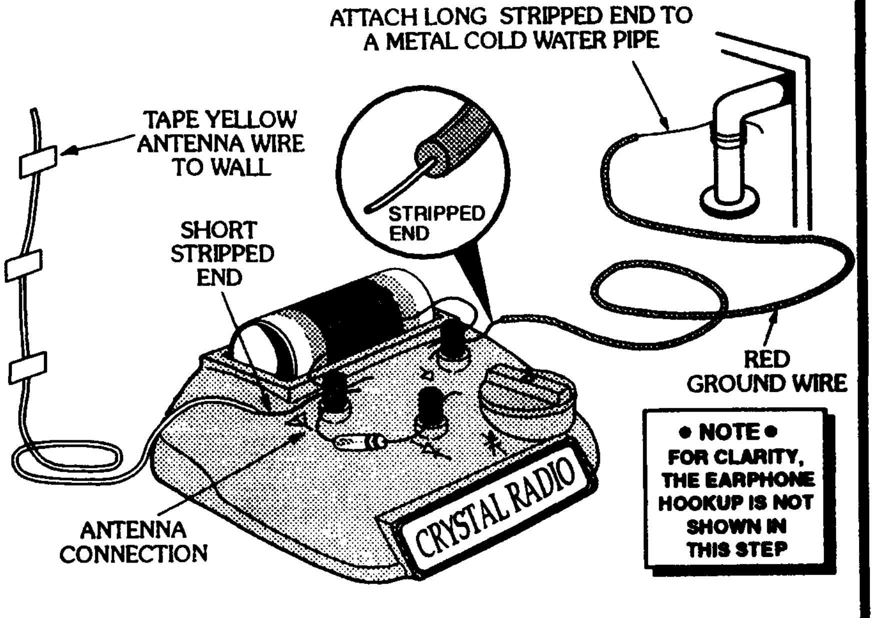
ATTACH LONG STRIPPED END TO A METAL COLD WATER PIPE
TAPE YELLOW ANTENNA WIRE TO WALL
STRIPPED END
SHORT STRIPPED END
RED GROUND WIRE
CRYSTAL RADIO
ANTENNA CONNECTION

**• NOTE •
FOR CLARITY, THE EARPHONE HOOKUP IS NOT SHOWN IN THIS STEP**

A. Locate the Red Ground and Yellow Antenna wire.
B. Attach the stripped end of Yellow Antenna wire to the Antenna Connection on the Base.

Note: The Antenna should be strung in a straight line away from power lines and large metal objects. We suggest taping it high up on a wall using masking tape as shown.

C. Next attach the short stripped end of the Ground wire to the Ground Connection.
D. Attach the Long stripped end to a metal cold water pipe, metal radiator pipe, etc., near where the radio will be used. Have a adult help you with this since some pipes may be hot.

Note: If there is not enough stripped, bare wire to wrap around a pipe then ask an adult to strip more insulation off the end using a pair of wire strippers.

Almighty Crystal Radio

A Triple-Ganged Crazy Crystal Radio

copyright © 2006 by Larry J Solomon

This circuit, of my own design, is my current winner for sensitivity and selectivity. Performance is truly astonshing. When used as a receiver connected to a fine sound system it outperforms several commercial AM radios that I have tested for sound quality.

Pay especial attention to the polarity of the coils when building this set. They are wound clockwise from top to bottom looking down on the coil. All the wire is 24 AWG.

It is also important to use 500pF variables to tune the entire broadcast band. The more common 365pF units will work but will not tune the entire band. Note that the rotor sides of the caps go to ground (green clip in prototype). The stators are the ones with lugs attached (white, red, and yellow alligator clips in prototype picture).

I added a level control to audio output because the output was too loud on most stations using the low level mag input on a receiver, causing distortion. The pot can adjust the level for best fidelity and pulls in almost any local station. An alternative is to use a high level input on the receiver. In that case, omit the pot and P2.

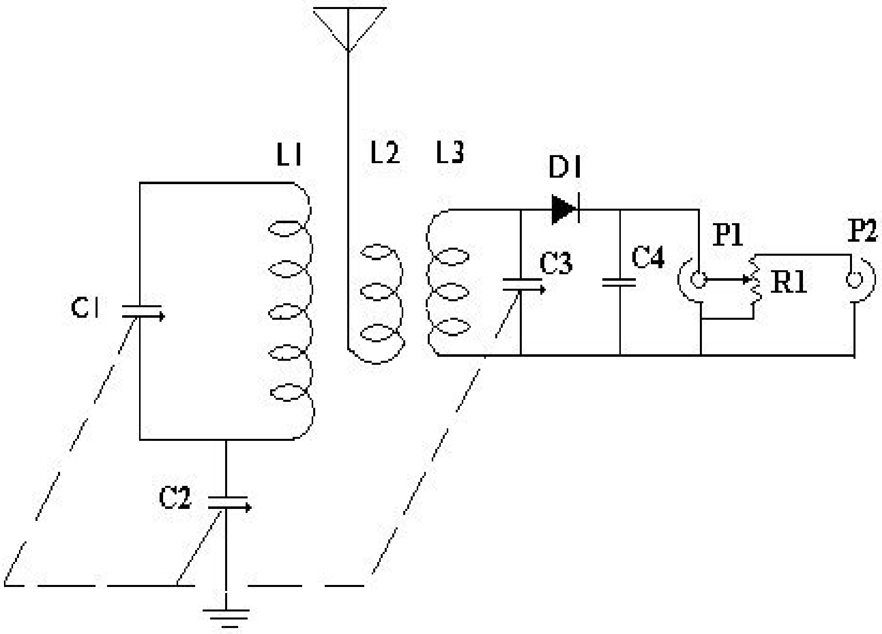

L1 - 50 turns 24 gauge magnet wire

L2 - 15 turns 24 gauge magnet wire wound on top of L1 and L3 separated by a layer of varnish

L3 - 25 turns 24 gauge magnet wire wound parallel to L1 after 12th turn of L1

C1, C2, C3 - 500 pF each, triple ganged air variable capacitors

C4 - 001.uF capacitor

D1 - 1N34 diode or rock crystal

R1 - 400K potentiometer

Coil form 3 inch diameter x 3 inch long

Here is the prototype. The coil was wound with my coil winder, of course.

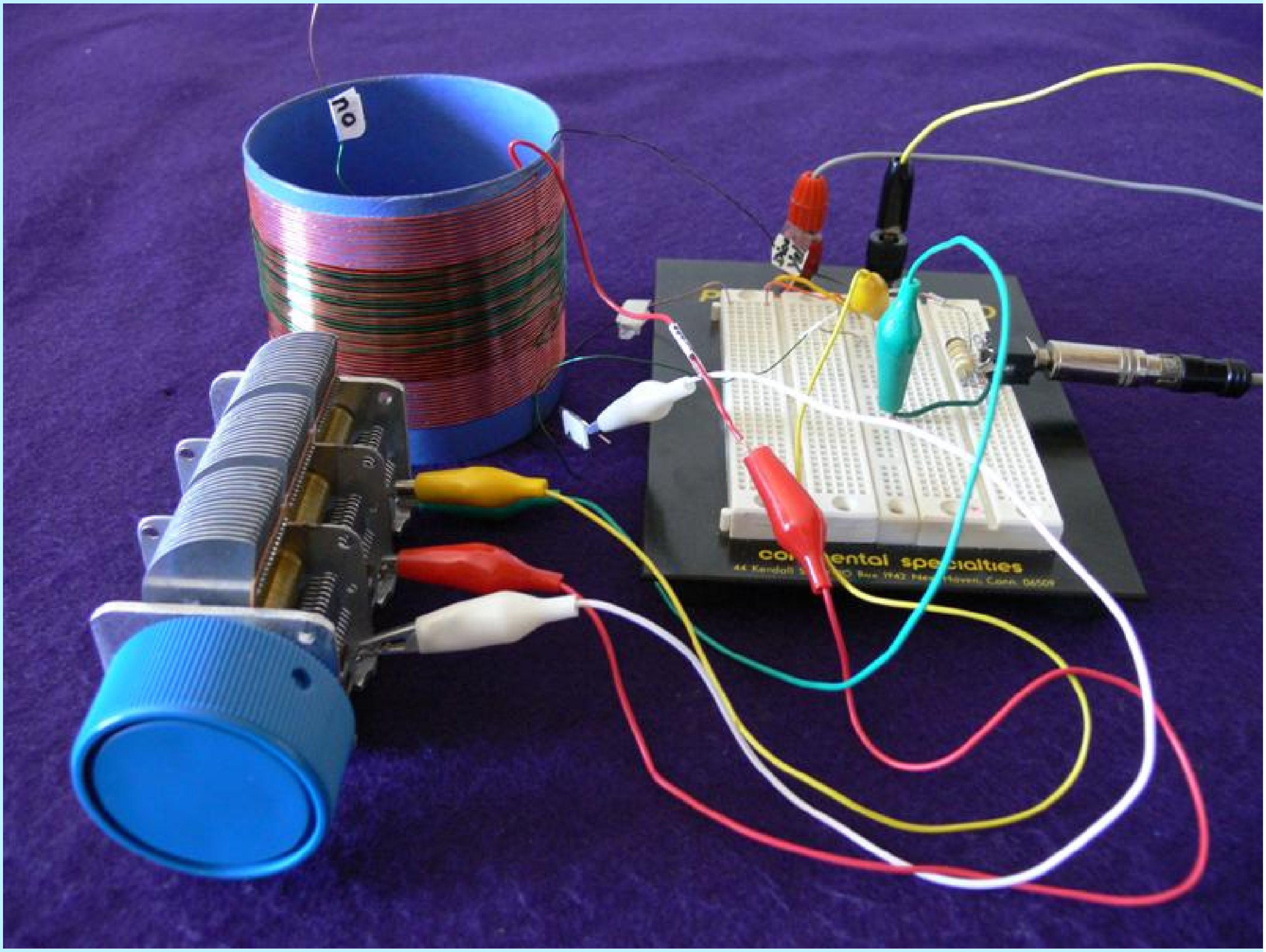
no
continental specialties
44 Kendall St. PO Box 1942 New Haven, Conn. 06509

Now some photos:

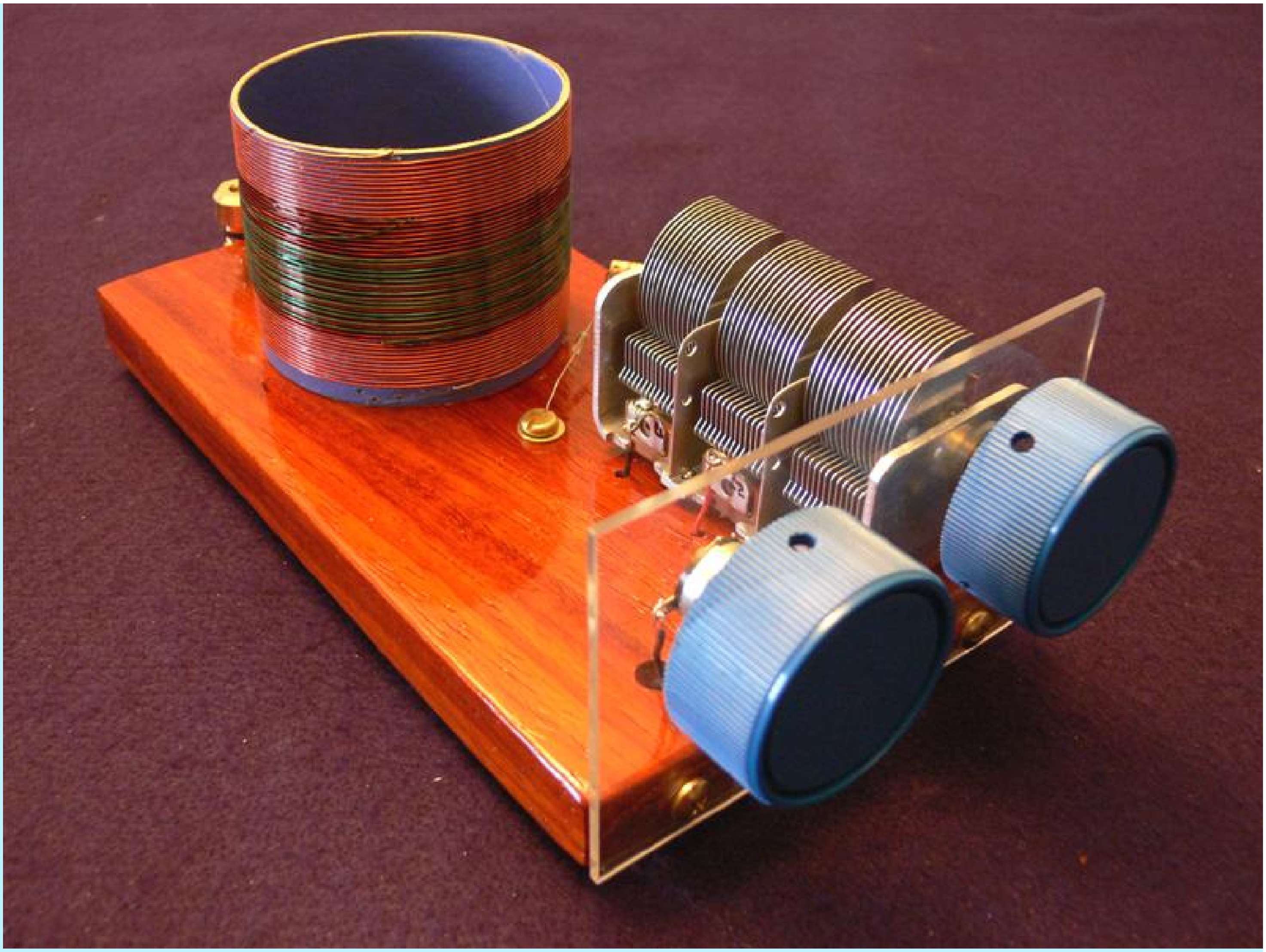

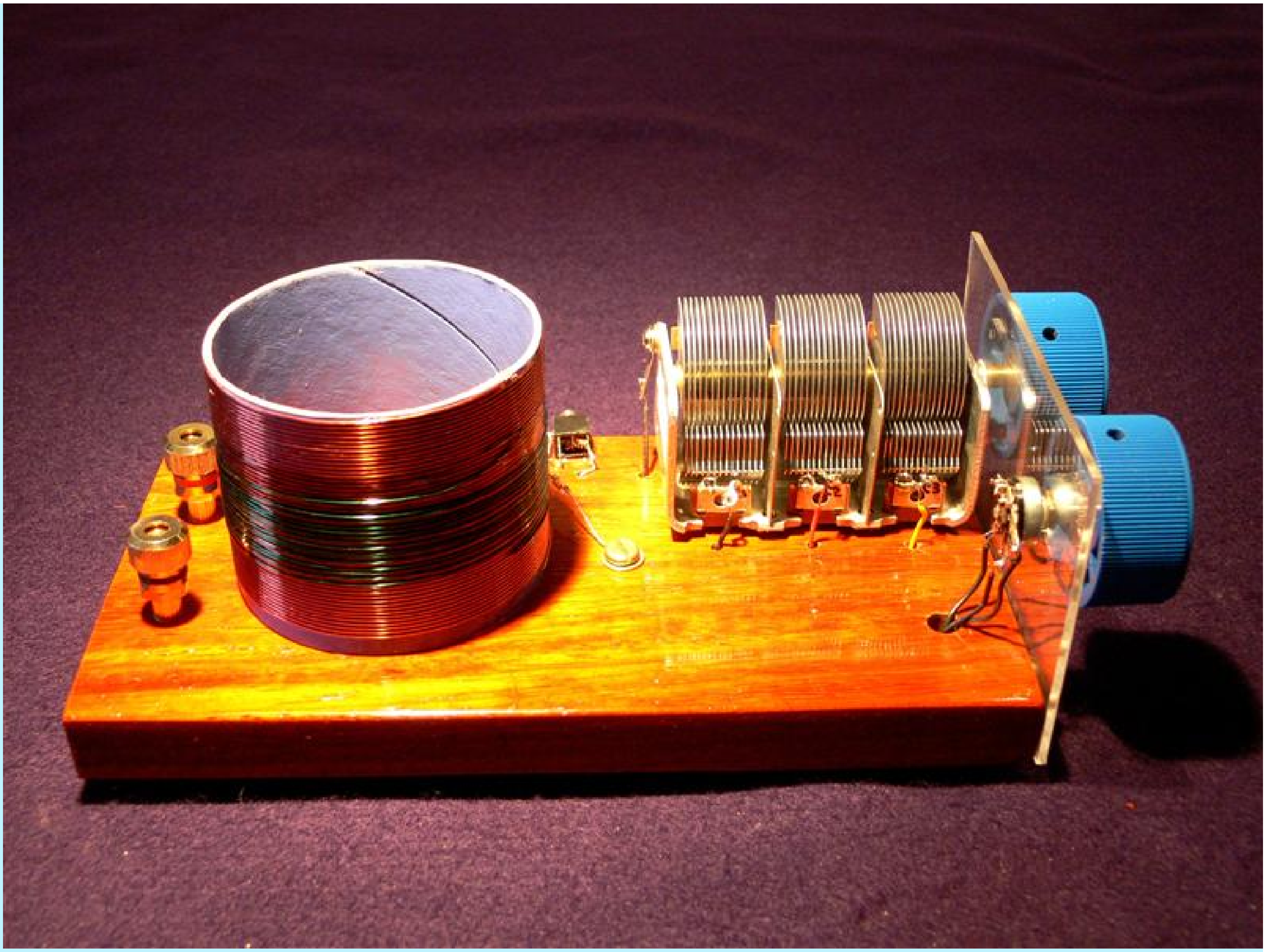

Home

US009564800B2

(12) United States Patent
Ramorini et al.

(10) **Patent No.: US 9,564,800 B2**
(45) **Date of Patent: *Feb. 7, 2017**

---

(54) **HIGH-EFFICIENCY ENERGY HARVESTING INTERFACE AND CORRESPONDING ENERGY HARVESTING SYSTEM**

(71) Applicant: **STMICROELECTRONICS S.R.L.**, Agrate Brianza (IT)

(72) Inventors: **Stefano Ramorini**, Arluno (IT); **Alessandro Gasparini**, Cusano Milanino (IT); **Alberto Cattani**, Cislago (IT)

(73) Assignee: **STMICROELECTRONICS S.R.L.**, Agrate Brianza (IT)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **15/163,394**

(22) Filed: **May 24, 2016**

(65) **Prior Publication Data**

US 2016/0268887 A1 Sep. 15, 2016

Related U.S. Application Data

(63) Continuation of application No. 14/749,254, filed on Jun. 24, 2015, now Pat. No. 9,379,615.

(30) **Foreign Application Priority Data**

Sep. 17, 2014 (IT) TO2014A0735

(51) **Int. Cl.**

H02M 3/02	(2006.01)
H02J 3/38	(2006.01)
H02J 7/35	(2006.01)

(52) **U.S. Cl.**
CPC ***H02M 3/02*** (2013.01); ***H02J 3/385*** (2013.01); ***H02J 7/35*** (2013.01)

(58) **Field of Classification Search**
CPC H02M 3/02; H02J 7/35
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,255,789 A	3/1981	Hartford et al.
5,903,130 A	5/1999	Rice et al.

(Continued)

OTHER PUBLICATIONS

Bandyopadhyay et al., "Platform Architecture for Solar, Thermal, and Vibration Energy Combining with MPPT and Single Inductor," IEEE Journal of Solid-State Circuits, vol. 47, No. 9, Sep. 2012. 17 pages.

(Continued)

Primary Examiner — Timothy J Dole
Assistant Examiner — Yusef Ahmed
(74) *Attorney, Agent, or Firm* — Seed IP Law Group LLP

(57) **ABSTRACT**

An electrical-energy harvesting system envisages a transducer for converting energy from an environmental energy source into a transduced signal, an electrical energy harvesting interface for receiving the transduced signal and for supplying a harvesting signal, and an energy storage element coupled to the electrical energy harvesting interface for receiving the harvesting signal. The electrical-energy harvesting system also includes a voltage converter connected to the electrical energy harvesting interface for generating a regulated voltage. The harvesting interface samples an open-circuit voltage value of the transduced signal, generates an optimized voltage value starting from the open-circuit voltage value, and generates an upper threshold voltage and a lower threshold voltage on the basis of the optimized voltage value. The harvesting interface controls the voltage converter in switching mode so that the harvesting signal has a value between the upper and lower threshold voltages in at least one operating condition.

20 Claims, 8 Drawing Sheets

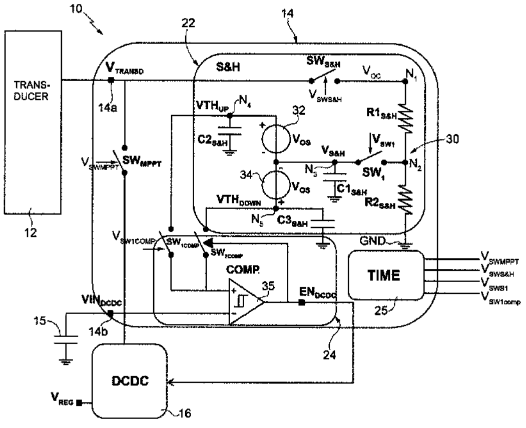

(56) References Cited

U.S. PATENT DOCUMENTS

5,949,054	A *	9/1999	Karpen	G06K 7/10851 235/462.01
5,949,197	A	9/1999	Kastner	
6,111,767	A *	8/2000	Handleman	G05F 1/67 136/293
8,193,798	B1	6/2012	Pace et al.	
8,749,218	B2	6/2014	Mateu et al.	
2004/0135562	A1	7/2004	Oden	
2006/0061342	A1	3/2006	Bernacchia et al.	
2010/0026479	A1	2/2010	Tran	
2011/0260536	A1	10/2011	Roy et al.	
2011/0298442	A1	12/2011	Waltisperger et al.	
2012/0109399	A1	5/2012	Tran	
2012/0119569	A1	5/2012	Karalis et al.	
2012/0274134	A1	11/2012	Gasparini et al.	
2012/0274144	A1*	11/2012	Gasparini	B60C 23/041 307/80
2012/0293021	A1*	11/2012	Teggatz	H02J 7/35 307/151
2014/0084876	A1*	3/2014	Ramorini	H02M 3/1588 320/166
2014/0084900	A1*	3/2014	Ramorini	H02M 3/155 323/355
2014/0097814	A1*	4/2014	Brewster	H02M 3/156 323/282
2014/0142876	A1	5/2014	John et al.	
2014/0239917	A1	8/2014	Bottarel et al.	
2014/0247020	A1	9/2014	Stankovic et al.	
2014/0292236	A1	10/2014	Ortiz	
2015/0214841	A1	7/2015	Ramorini et al.	
2015/0234416	A1*	8/2015	Ramorini	G05F 5/00 323/299

OTHER PUBLICATIONS

Brunelli et al., "Photovoltaic scavenging systems: Modeling and optimization," *Microelectronics Journal 40*:1337-1344, 2009.

Cho et al., "A Coreless Maximum Power Point Tracking Circuit of Thermoelectric Generators for Battery Charging Systems," IEEE Asian Solid-State Circuits Conference, Beijing, Nov. 8-10, 2010, 4 pages.

Dini et al., "A Fully-Autonomous Integrated RF Energy Harvesting System for Wearable Applications," Proceedings of the 43rd European Microwave Conference (EuMC), Nuremberg, Oct. 6-10, 2013, pp. 987-990.

Huang et al., "Thermoelectric Energy Harvesting with lmV Low Input Voltage and 390nA Quiescent Current for 99.6% Maximum Power Point Tracking," 38th European Solid-State Circuits Conference (ESSCIRC), Bordeaux, Sep. 17-21, 2012, pp. 105-108.

Italian Search Report dated May 13, 2015, for corresponding IT Application No. T020140735, 10 pages.

Kadirvel et al., "A 330nA Energy-Harvesting Charger with Battery Management for Solar and Thermoelectric Energy Harvesting," IEEE International Solid-State Circuits Conference, Feb. 20, 2012. 3 pages.

Kim et al., "A Seamless Mode Transfer Maximum Power Point Tracking Controller for Thermoelectric Generator Applications," *IEEE Transactions on Power Electronics 23*(5):2310-2318, Sep. 2008.

Kim et al., "Aggregated Modeling and Control of a Boost-Buck Cascade Converter for Maximum Power Point Tracking of a Thermoelectric Generator," IEEE Applied Power Electronics Conference and Exposition, 2008. APEC 2008. 7 pages.

Ni et al., "A Power Conditioning System for Thermoelectric Generator Based on Interleaved Boost Converter with MPPT Control," Electrical Machines and Systems (ICEMS), 2011 International Conference on. IEEE, 2011. 6 pages.

Weddell et al., "Photovoltaic Sample-and-Hold Circuit Enabling MPPT Indoors for Low-Power Systems," *IEEE Transactions on Circuits and Systems—I: Regular Papers 59*(6):1196-1204, Jun. 2012.

Weddell et al., "Ultra Low-Power Photovoltaic MPPT Technique for Indoor and Outdoor Wireless Sensor Nodes," Design, Automation & Test in Europe Conference & Exhibition, Grenoble, Mar. 14-18, 2011, 4 pages.

Yu et al., "Indoor Micro-Light Energy Harvesting System for Low Power Wireless Sensor Node," Proceedings Power MEMS, Nov. 30, 2010, pp. 21-24.

* cited by examiner

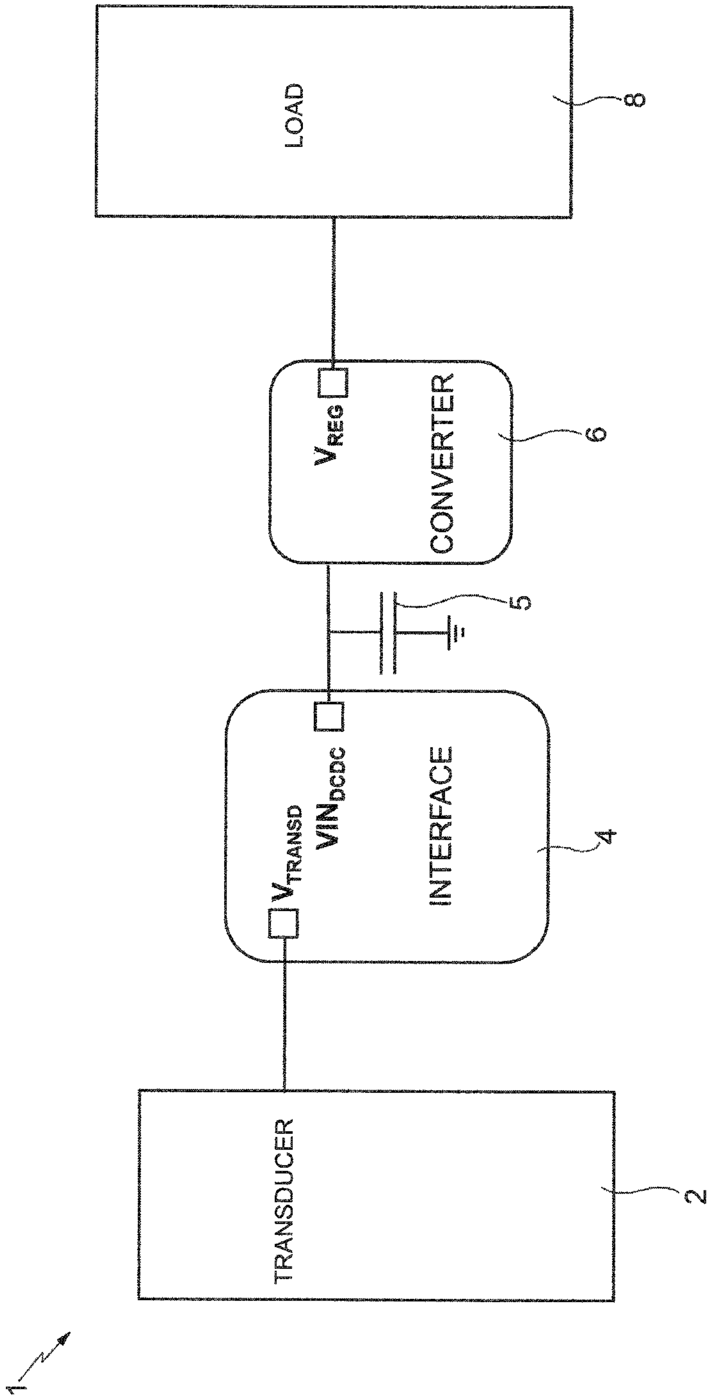
LOAD
8
V_{REG}
CONVERTER
6
5
V_{TRANSD}
VIN_{DCDC}
INTERFACE
4
TRANSDUCER
2
1

FIG. 1
(Prior Art)

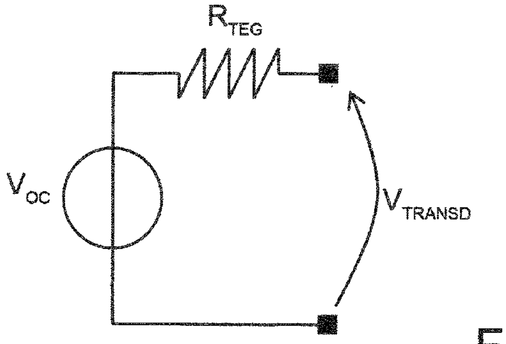
R_{TEG}
V_{OC}
V_{TRANSD}

FIG. 2A
(Prior Art)

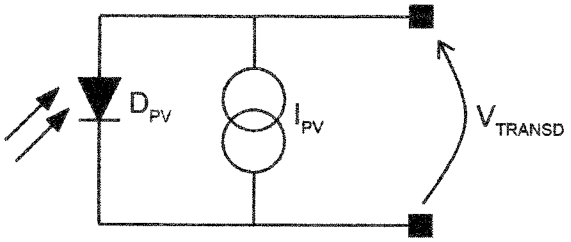
D_{PV}
I_{PV}
V_{TRANSD}

FIG. 2B
(Prior Art)

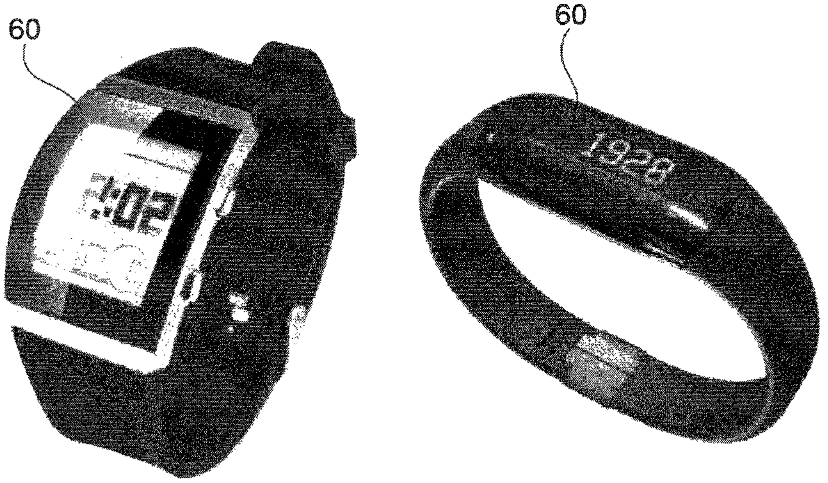
60
60
:02
1928

FIG. 9

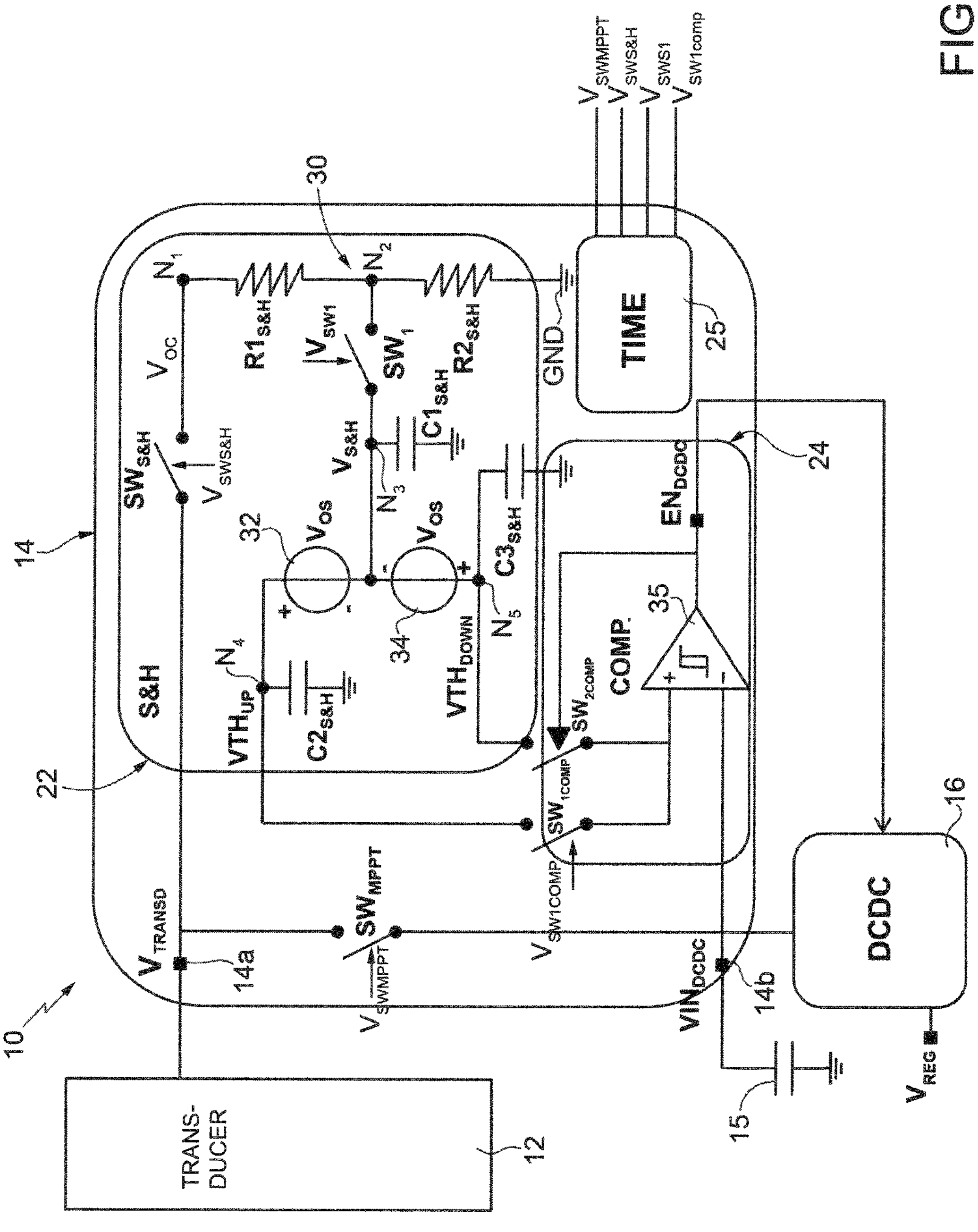
10
12
TRANS-
DUCER
14
14a
14b
15
16
22
24
25
30
32
34
35
V_{TRANSD}
SW_{MPPT}
V_{SWMPPT}
VIN_{DCDC}
DCDC
V_{REG}
S&H
$SW_{S\&H}$
$V_{SWS\&H}$
V_{OC}
N_1
N_2
N_3
N_4
N_5
$R1_{S\&H}$
$R2_{S\&H}$
SW_1
V_{SW1}
$V_{S\&H}$
$C1_{S\&H}$
$C2_{S\&H}$
$C3_{S\&H}$
V_{OS}
V_{OS}
VTH_{UP}
VTH_{DOWN}
GND
TIME
COMP.
SW_{1COMP}
SW_{2COMP}
$V_{SW1COMP}$
EN_{DCDC}
V_{SWMPPT}
$V_{SWS\&H}$
V_{SWS1}
$V_{SW1comp}$

FIG. 3

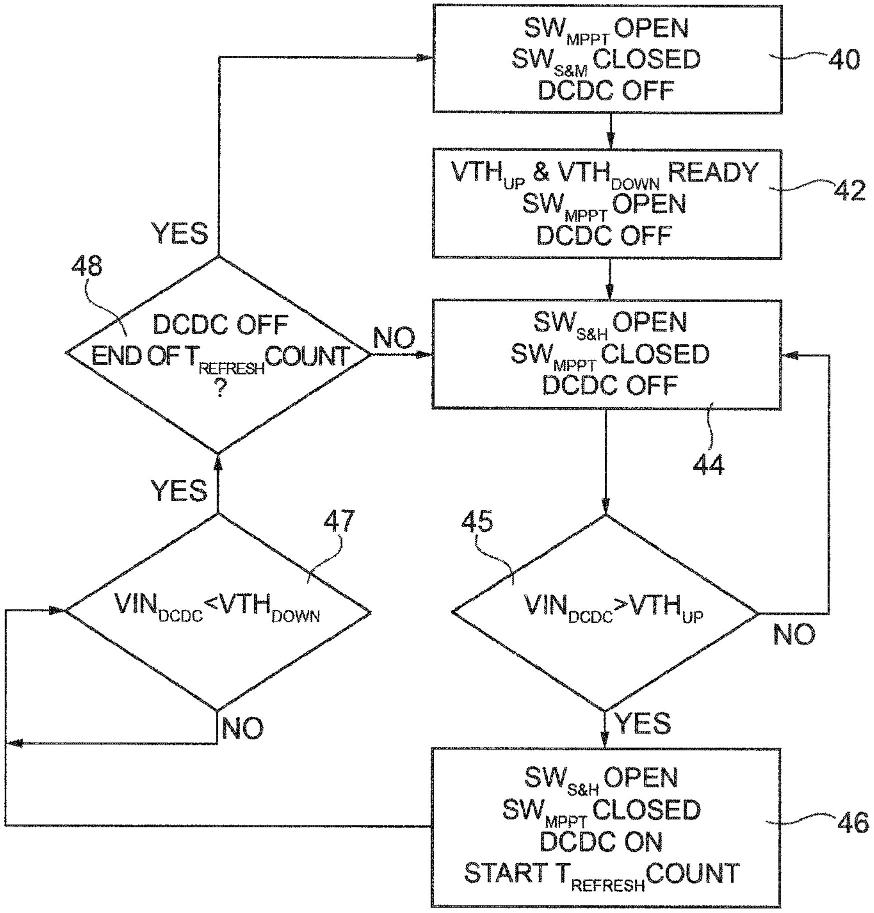
SW_{MPPT} OPEN
$SW_{S\&M}$ CLOSED
DCDC OFF
40
VTH_{UP} & VTH_{DOWN} READY
SW_{MPPT} OPEN
DCDC OFF
42
$SW_{S\&H}$ OPEN
SW_{MPPT} CLOSED
DCDC OFF
44
45
$VIN_{DCDC}>VTH_{UP}$
NO
YES
$SW_{S\&H}$ OPEN
SW_{MPPT} CLOSED
DCDC ON
START $T_{REFRESH}$ COUNT
46
47
$VIN_{DCDC}<VTH_{DOWN}$
NO
YES
48
DCDC OFF
END OF $T_{REFRESH}$ COUNT
?
YES
NO

FIG. 4

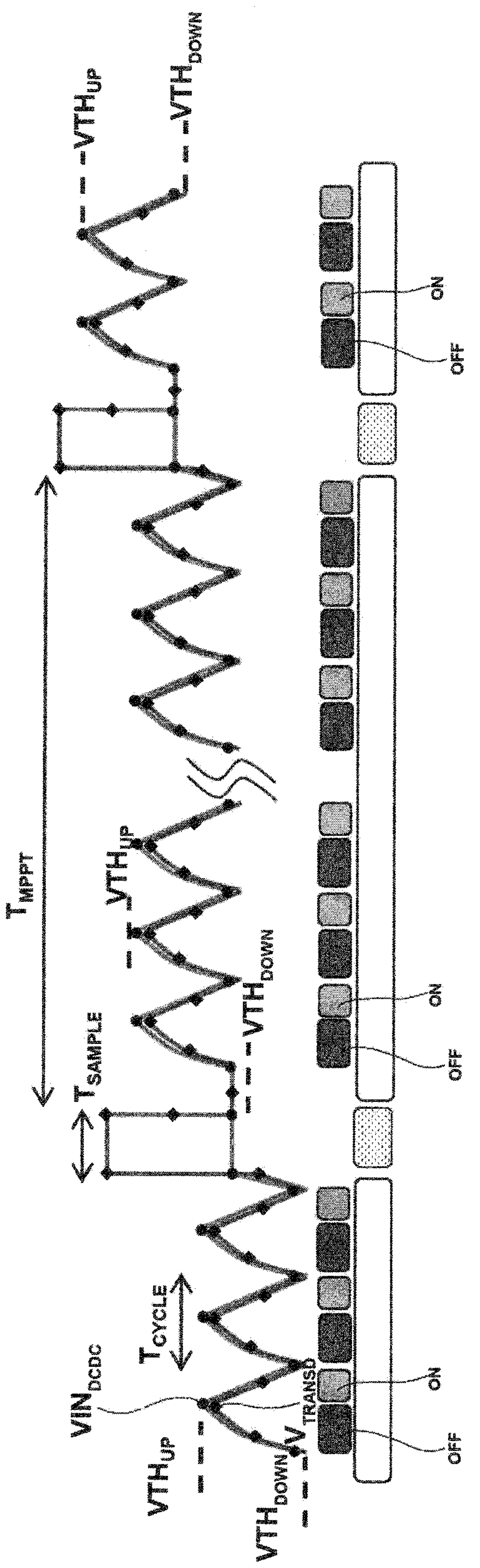
VIN_DCDC
VTH_UP
VTH_DOWN
V_TRANSD
T_CYCLE
T_SAMPLE
T_MPPT
VTH_UP
VTH_DOWN
VTH_UP
VTH_DOWN
ON
OFF
ON
OFF
ON
OFF

FIG. 5

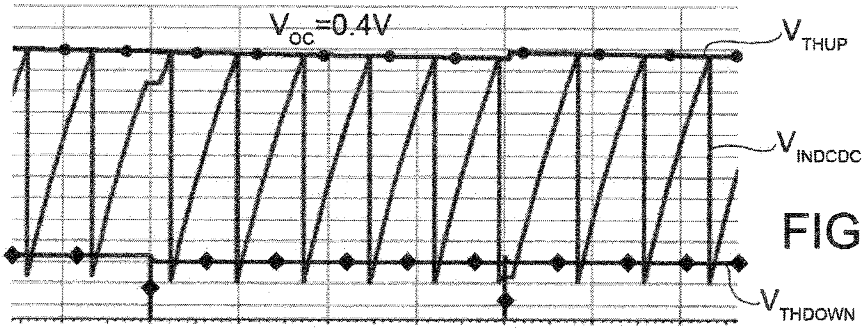

FIG. 6A

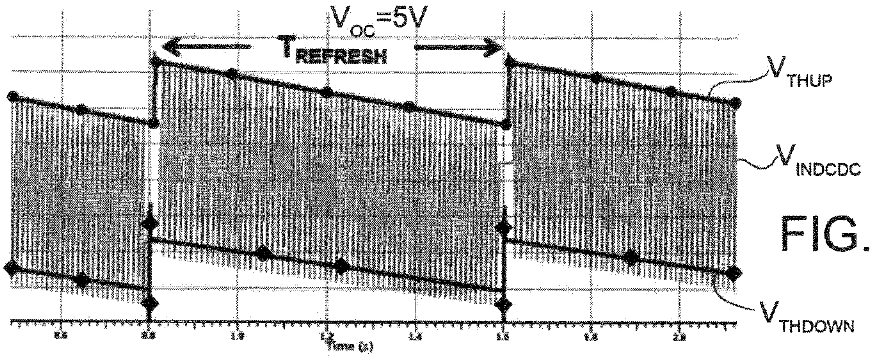

FIG. 6B

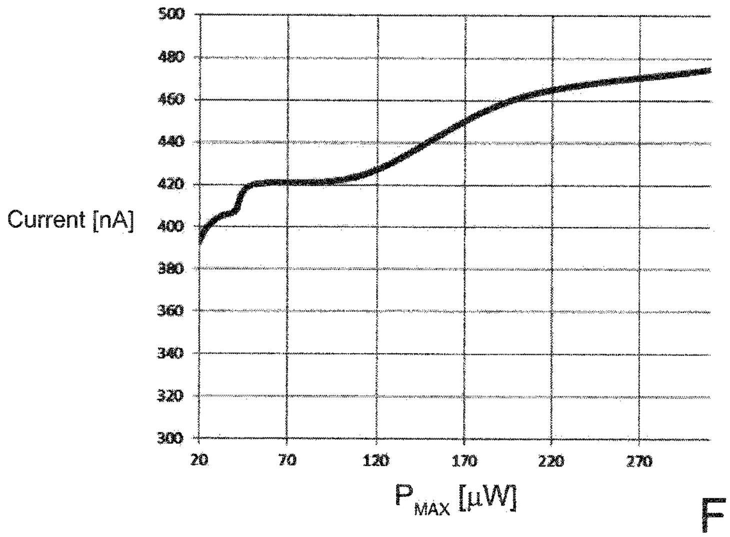

FIG. 8A

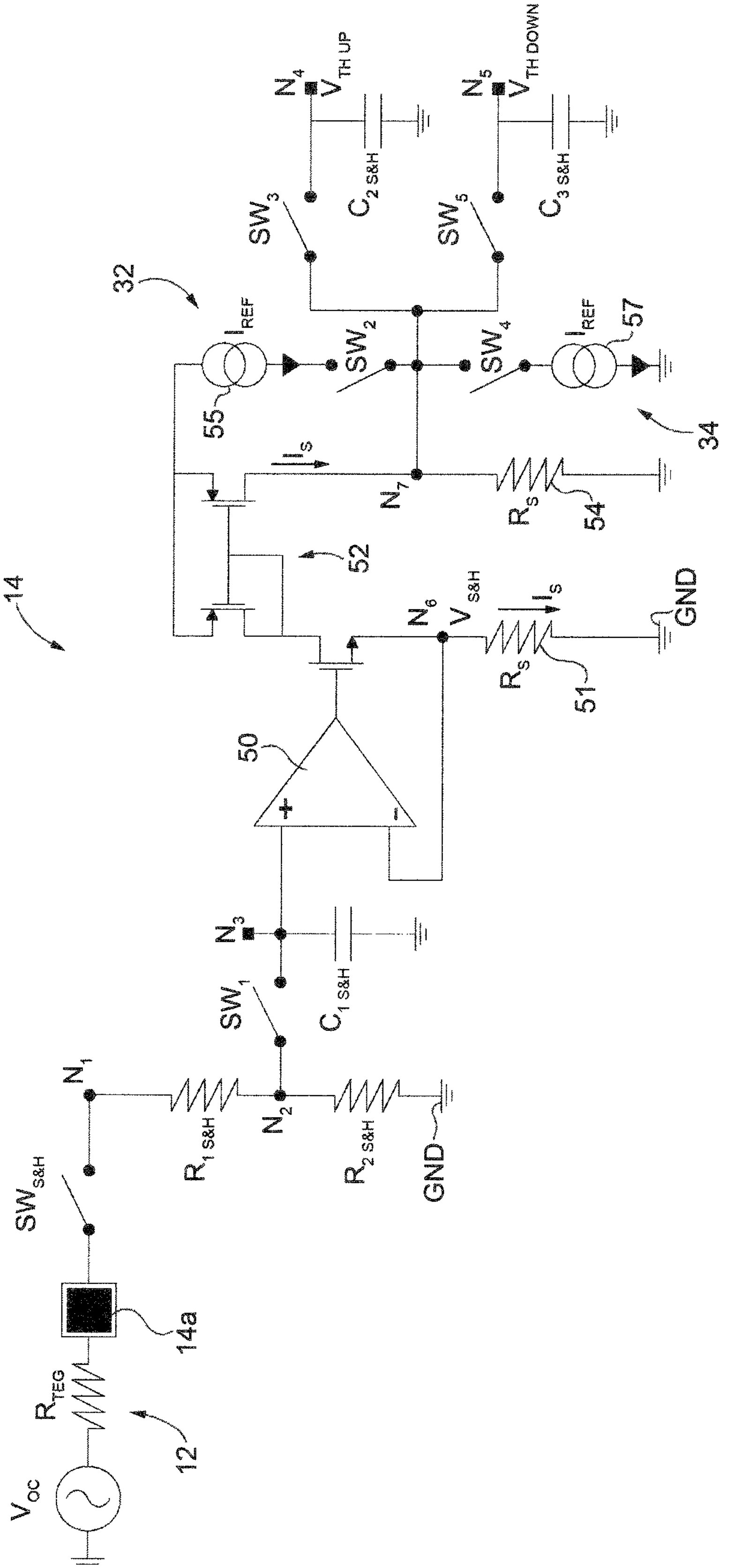
V_OC
R_TEG
12
14a
SW_S&H
N_1
R_1 S&H
N_2
R_2 S&H
GND
SW_1
C_1 S&H
N_3
50
+
-
14
52
N_6
V_S&H
I_S
R_S
51
GND
55
I_REF
32
SW_2
I_S
N_7
R_S
54
SW_4
I_REF
57
34
SW_3
C_2 S&H
N_4
V_TH UP
SW_5
C_3 S&H
N_5
V_TH DOWN

FIG. 7

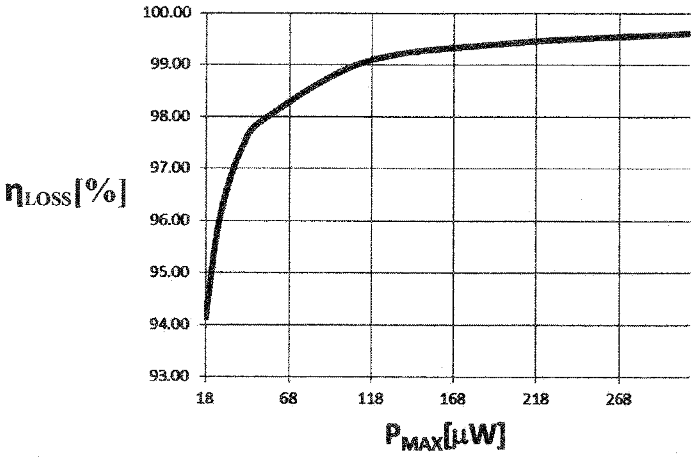

FIG. 8B

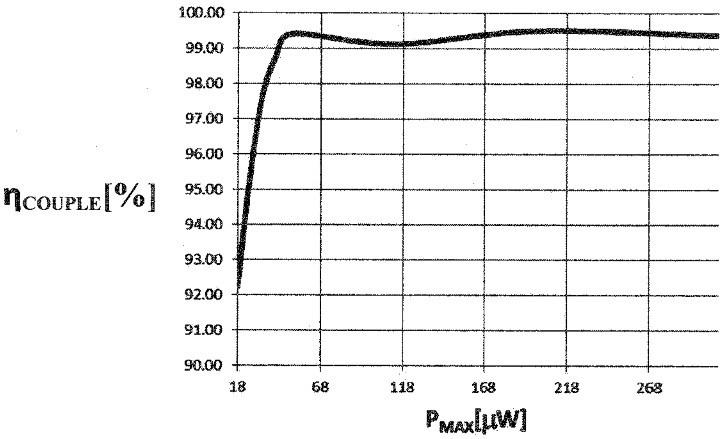

FIG. 8C

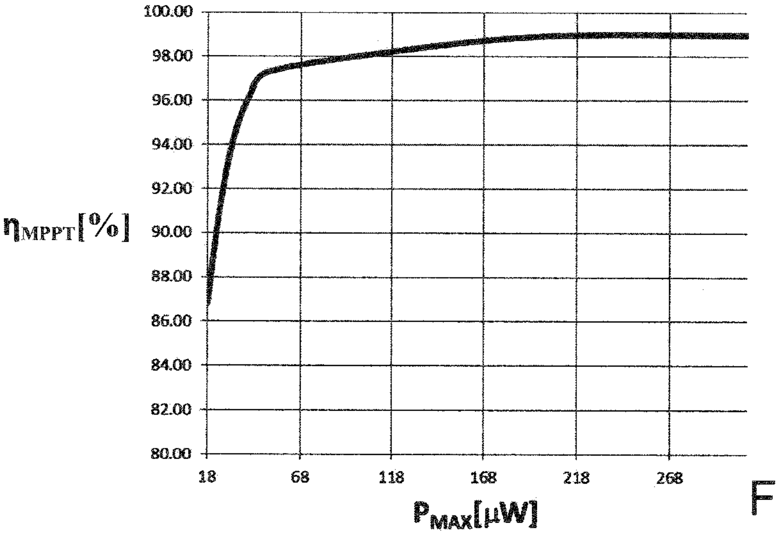

FIG. 8D

HIGH-EFFICIENCY ENERGY HARVESTING INTERFACE AND CORRESPONDING ENERGY HARVESTING SYSTEM

CROSS-REFERENCE TO RELATED APPLICATIONS

The present application is a Continuation of U.S. patent application Ser. No. 14/749,254, filed on Jun. 24, 2015, which is incorporated by reference in its entirety, and which, along with the present application, claims priority to Italian Patent Application TO2014A000735, filed on Sep. 17, 2014.

BACKGROUND

Technical Field

The present disclosure relates to a high-efficiency energy harvesting interface and to a corresponding energy harvesting system.

Description of the Related Art

As is known, systems for harvesting (or scavenging) energy from mechanical or environmental energy sources arouse considerable interest in a wide range of technological fields, for example in the field of portable electronic devices or in the automotive field.

Typically, energy harvesting systems are designed to harvest, store, and transfer energy generated by mechanical or environmental sources to a generic electrical load, which may be supplied, or, in the case of a battery, recharged. These systems thus enable production of electronic apparatuses that operate without a battery, or with a considerable increase in the duration of batteries in the case of apparatuses provided therewith.

For harvesting environmental energy, solar or thermoelectric generators may be used, which convert solar energy and thermal energy, respectively, into electrical energy.

FIG. **1** shows schematically and by functional blocks, an energy harvesting system of a known type.

The energy harvesting system **1** comprises a transducer **2**, for example a photovoltaic or thermoelectric generator that includes a plurality of cells (of a known type, not described in detail herein), which converts solar energy or thermal energy into electrical energy, typically into a DC voltage or, in any case, into a voltage that varies slowly in time (with respect to the electrical constants of the circuit), generating a transduction signal V_{TRANSD}.

The energy harvesting system **1** further comprises a harvesting interface **4**, designed to provide a condition of coupling with the transducer **2** of the MPPT (Maximum Power Point Tracking) type, in order to maximize extraction of power. The harvesting interface **4** is configured to receive at input the transduction signal V_{TRANSD} generated by the transducer **2** and supply at output a harvesting signal VIN_{DCDC}.

The energy harvesting system **1** further comprises: a storage capacitor **5**, which is connected to the output of the harvesting interface **4** and receives the harvesting signal VIN_{DCDC}, which determines charging thereof and consequent storage of energy; and a DC-DC converter **6**, connected to the storage capacitor **5** for receiving at input the stored electrical energy and generating at output a regulated signal V_{REG}, with an appropriate value so that it may be supplied to an electrical load **8**, for its supply or its recharging.

The global efficiency η_{TOT} of the energy harvesting system **1** is given by the expression:

$$\eta_{TOT}=\eta_{TRANSD}\cdot\eta_{MPPT}\cdot\eta_{DCDC}$$

where: η_{TRANSD} is the efficiency of the transducer **2**, indicating the amount of environmental energy, effectively converted by the transducer **2** into electrical energy; η_{MPPT} is the efficiency of the harvesting interface **4**, indicating the amount of converted electrical energy that is effectively used for charging the storage capacitor **5**; and η_{DCDC} is the efficiency of the DC-DC converter **6**.

In particular, the efficiency η_{MPPT} of the harvesting interface **4** indicates the ratio between the power effectively transferred onto the storage capacitor **5** and the maximum power that could theoretically be supplied, P_{MAX}.

In detail, this efficiency η_{MPPT} is given by the following expression:

$$\eta_{MPPT}=\eta_{COUPLE}\cdot\eta_{LOSS}$$

where η_{COUPLE} is the coupling factor between the transducer **2** and the harvesting interface **4** (indicating the impedance matching between the same transducer **2** and the harvesting interface **4**), and η_{LOSS} is the loss of power due to consumption by the harvesting interface **4**.

It has been shown that, in the case of a thermoelectric cell, which may be represented schematically, as illustrated in FIG. **2***a*, as an equivalent voltage generator V_{OC} connected to a series resistance R_{TEG}, the efficiency η_{MPPT} is maximized in the case where the transduction signal V_{TRANSD} has an optimized value V_{MPPT} equal to $V_{OC}/2$ (i.e., a value equal to one half of the load-less, or open-circuit, voltage supplied by the corresponding equivalent voltage generator).

Likewise, in the case of a photovoltaic cell, which may be represented schematically, as illustrated in FIG. **2***b*, as an equivalent current generator I_{PV} connected in parallel to a diode D_{PV} (in the figure, the equivalent series resistance of the generator is not represented), the efficiency η_{MPPT} is maximized in the case where the transduction signal V_{TRANSD} has an optimized value V_{MPPT} comprised between $0.75\cdot V_{OC}$ and $0.9\cdot V_{OC}$ (according to the constructional parameters of the photovoltaic cell and the material of which it is made), for example equal to $0.8\cdot V_{OC}$, where V_{OC} is once again the open-circuit voltage supplied by the photovoltaic cell.

It is consequently required that the harvesting interface **4** of the energy harvesting system **1** be configured in such a way that the transducer **2** operates in, or around, a working point that ensures the aforesaid condition of maximum efficiency.

For this purpose, a wide range of circuit configurations have been proposed for providing the harvesting interface **4**.

For instance, in the document entitled "A Seamless Mode Transfer Maximum Power Point Tracking Controller for Thermoelectric Generator Applications" by Rae-Young Kim, Jih-Sheng Lai, IEEE Transactions on Power Electronics, vol. 23, No. 5, September 2008, an interface circuit has been proposed, comprising a dual voltage conversion stage, formed by the cascade of a boost converter and a buck converter, the latter being designed to regulate the value of the output voltage. Tracking of the MPPT condition is obtained with a continuous-time control of the duty cycle of the boost converter.

The present Applicant has, however, realized that this solution involves a high power consumption, which is due to the fact that the control is of a continuous-time type, which does not render it suited to energy harvesting applications. Further, this solution does not prove flexible, being suited only to a specific type of transducer and to precise values of the electrical parameters associated thereto, further depending upon the tolerance in the values assumed by the same electrical parameters. In general, this solution also involves a large number of external components, which may not be made with integrated technology.

Another possible circuit implementation is described in the document entitled "Thermoelectric Energy Harvesting with 1 mV Low Input Voltage and 390 nA Quiescent Current for 99.6% Maximum Power Point Tracking" by Chao-Jen Huang, Wei-Chung Chen, Chia-Lung Ni, Ke-Horng Chen, Chien-Chun Lu, Yuan-Hua Chu, and Ming-Ching Kuo, 38th European Solid-State Circuits Conference (ESSCIRC), September 2012. This solution envisages a boost converter and a continuous-time algorithm, the so-called perturbation and observation algorithm, to achieve the MPPT condition; in particular, the duty cycle of the converter is perturbed, and the trend of the output voltage is measured: the MPPT condition corresponds to a maximum positive trend.

The present Applicant has, however, realized that also this solution has some disadvantages, amongst which: a high power consumption, intrinsic in a continuous-time perturbation and observation algorithm, which renders it difficult to use in energy harvesting applications; and a poor efficiency, when combined to a low-power transducer.

The document entitled "A Coreless Maximum Power Point Tracking Circuit of Thermoelectric Generators for Battery Charging Systems", by S. Cho, N. Kim, S. Park, S. Kim, IEEE Asian Solid-State Circuits Conference, Nov. 8-10, 2010, Beijing, China, describes yet a further solution for providing the harvesting interface. This solution envisages two conversion stages, with the cascade of a boost conversion stage and a buck conversion stage, the latter for regulation of the output voltage; the MPPT condition is achieved by a control of the switch in the boost stage.

The present Applicant has realized that also this solution, albeit presenting a simpler algorithm to achieve the MPPT condition, does not have a high efficiency, on account of the presence of two conversion stages. Further, also this solution requires a large number of external components, not made with integrated technology.

The subject matter discussed in the Background section is not necessarily prior art and should not be assumed to be prior art merely as a result of its discussion in the Background section. Along these lines, any recognition of problems in the prior art discussed in Background section or associated with such subject matter should not be treated as prior art unless expressly stated to be prior art. Instead, the discussion of any subject matter in the Background section should be treated as part of the inventor's approach to the particular problem, which in and of itself may also be inventive.

BRIEF SUMMARY

The present disclosure provides an energy harvesting interface that will enable the aforementioned problems and disadvantages to be overcome, in full or in part, and in particular that will provide a high efficiency.

According to the present disclosure, an energy harvesting system provided with an energy harvesting interface is consequently provided, as defined in the annexed claims.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

Non-limiting and non-exhaustive embodiments are described with reference to the following drawings, wherein like labels refer to like parts throughout the various views unless otherwise specified. For a better understanding of the present disclosure, preferred embodiments thereof are now

described purely by way of non-limiting example and with reference to the attached drawings, wherein:

FIG. **1** shows an energy harvesting system according to a known embodiment;

FIGS. **2***a* and **2***b* show the equivalent circuit diagrams of a transducer, including a thermoelectric cell or a photovoltaic cell, respectively;

FIG. **3** shows a circuit diagram of an energy harvesting system according to one embodiment of the present solution;

FIG. **4** shows a flowchart relating to operations performed by a harvesting interface in the system of FIG. **3**;

FIGS. **5** and **6***a*-**6***b* show plots of electrical quantities associated with the harvesting interface;

FIG. **7** shows a more detailed circuit diagram of one embodiment of the harvesting interface;

FIGS. **8***a*-**8***d* show plots relating to the electrical performance of the harvesting interface; and

FIG. **9** shows, by way of example, an electronic device, in particular a bracelet or an electronic watch, including the energy harvesting system.

DETAILED DESCRIPTION

As illustrated in FIG. **3**, an energy harvesting system **10** according to one embodiment of the present solution, comprises in general, and substantially as has been described previously: a transducer **12**, in particular a photovoltaic or thermoelectric generator, which generates a transduction signal V_{TRANSD}, in particular a DC voltage or a slowly varying voltage; a harvesting interface **14**, which provides a MPPT coupling, receives the transduction signal V_{TRANSD}, on an input terminal **14***a*, and supplies a harvesting signal VIN_{DCDC}, on an output terminal **14***b*; a storage capacitor **15**, which is connected to the output terminal **14***b* of the harvesting interface **14** and receives the harvesting signal VIN_{DCDC}; and a DC-DC converter **16**, connected to the storage capacitor **15** for receiving the electrical energy stored and generating at output a regulated signal V_{REG}, which is then supplied to an electrical load (here not illustrated), for its supply or recharge.

According to one aspect of the present solution, and as will also be described in detail hereinafter, the harvesting interface **14** comprises: a tracking switch SW_{MPPT}, which is connected between the input terminal **14***a* and the output terminal **14***b* of the harvesting interface **14** and is controlled by a control signal V_{SWMPPT}; a sample-and-hold (S&H) stage **22**, configured to sample the value Voc of the transduction signal V_{TRANSD}, generated by the transducer **12** in an open-circuit or loadless condition, at appropriate time intervals, to generate the optimized value V_{MPPT} (see the foregoing discussion) starting from the value V_{OC}, and to generate an upper threshold voltage VTH_{UP} and a lower threshold voltage VTH_{DOWN}, which satisfy the relation $VTH_{DOWN}<V_{MPPT}<VTH_{UP}$, on the basis of the optimized value V_{MPPT}; a comparator stage **24**, with hysteretic voltage control, on the basis of the upper and lower threshold voltages VTH_{UP}, VTH_{DOWN}, which generates at output an enabling signal EN_{DCDC} for the DC-DC converter **16**; and a timing stage **25**, which generates appropriate control and timing signals for operation of the harvesting interface **14**, amongst which the aforesaid control signal V_{SWMPPT}.

In general, operation of the harvesting interface **14** envisages, upon opening of the tracking switch SW_{MPPT}, sampling of the value of the transduction signal V_{TRANSD}, with the transducer **12** operating in loadless conditions, and corresponding generation of the upper and lower threshold voltages VTH_{UP}, VTH_{DOWN}; and subsequently, upon closing of the tracking switch SW_{MPPT}, generation of the harvesting signal VIN_{DCDC}, with a value comprised between the upper and lower threshold voltages VTH_{UP}, VTH_{DOWN} (thanks to the hysteretic control by the comparator stage **24**), and thus about the optimized value V_{MPPT}. Given that, for closing of the tracking switch SW_{MPPT}, the value of the harvesting signal VIN_{DCDC} coincides with the value of the transduction signal V_{TRANSD}, the condition of energy transfer into the storage capacitor **15** thus occurs in an MPPT condition, with substantially maximum efficiency and substantially maximum coupling between the transducer **12** and the harvesting interface **14**.

Conveniently, the S&H stage **22** is controlled by the timing stage **25** for sampling and periodically refreshing the value V_{OC} and consequently the values of the upper and lower threshold voltages VTH_{UP}, VTH_{DOWN} in such a way as to react promptly and adapt to possible variations of the operating conditions of the transducer **12**.

In detail, the S&H stage **22** of the harvesting interface **14** comprises: a sampling switch $SW_{S\&H}$, which is connected between the input terminal **14***a* of the harvesting interface **14** and a first internal node **N1** and receives a control signal $V_{SWS\&H}$; a voltage divider **30**, formed by a first dividing resistor $R1_{S\&H}$, connected between the first internal node **N1** and a second internal node **N2**, and a second dividing resistor $R2_{S\&H}$, connected between the second internal node **N2** and a reference terminal, or ground, GND (both dividing resistors $R1_{S\&H}$, $R2_{S\&H}$ have a resistance value much higher than the value of the series resistance of the equivalent generator of the transducer **12**); a first decoupling switch **SW1**, which is connected between the second internal node **N2** and a third internal node **N3** and receives a control signal V_{SW1}; a first holding capacitor $C1_{S\&H}$, connected between the third internal node **N3** and the reference terminal GND; a first voltage-generator module **32**, which is connected between the third internal node **N3** and a fourth internal node **N4** and is designed to generate an offset voltage V_{OS}; a second voltage-generator module **34**, which is connected between the third internal node **N3** and a fifth internal node **N5** and is designed to generate the same offset voltage V_{OS}; a second holding capacitor $C2_{S\&H}$, connected between the fourth internal node **N4** and the reference terminal GND; and a third holding capacitor $C3_{S\&H}$, connected between the fifth internal node **N5** and the reference terminal GND.

The comparator stage **24** comprises: a comparator **35**, including in a per se known manner an appropriately configured operational amplifier, having a first input terminal connected to the output terminal **14***b* of the harvesting interface **14**, a second input terminal, and an output terminal, which is connected to an enabling input of the DC-DC converter **16** and is designed to supply the enabling signal EN_{DCDC}; a first comparison switch $SW1_{COMP}$, which is connected between the fourth internal node **N4** of the harvesting interface **14** and the second input terminal of the comparator **35** and receives a control signal $V_{SW1COMP}$; and a second comparison switch $SW2_{COMP}$, which is connected between the fifth internal node **N5** of the harvesting interface **14** and the second input terminal of the comparator **35**, and receives, as a control signal, the enabling signal EN_{DCDC}.

The timing stage **25**, including in a per se known manner (not described in detail herein), an oscillator circuit, is configured to generate the control signals V_{SWMPPT}, $V_{SWMS\&H}$, V_{SW1}, $V_{SW1COMP}$ for the switches SW_{MPPT}, $SW_{MS\&H}$, **SW1**, $SW1_{COMP}$, according to a timing algorithm described in detail hereinafter. As will be discussed hereinafter, the timing stage **25** may supply further control signals for further switches that may be present in the circuit.

With reference also to the flowchart of FIG. **4**, operation of the harvesting interface **14** envisages that in an initial step **40** the value of the loadless or open-circuit voltage V_{OC} of the transduction signal V_{TRANSD} supplied by the transducer **12** is sampled and held.

For this purpose, the tracking switch SW_{MPPT} is driven into the opening condition, and the sampling switch $SW_{S\&H}$ is driven into the closing condition; in this step, the DC-DC converter **16** is turned off, and the first decoupling switch **SW1** is further driven into the closing condition.

The transducer **12** operates substantially in an open-circuit condition, given that the resistance as a whole supplied by the voltage divider **30** is much higher than its own equivalent series resistance, so that the value of the transduction signal V_{TRANSD} that is supplied and that is present on the first internal node **N1** substantially coincides with the loadless or open-circuit voltage V_{OC}.

In this situation, the voltage divider **30**, by an appropriate choice of the division ratio, generates on the second internal node **N2** a sampled voltage $V_{S\&H}$ having a value substantially equal to $V_{OC}/2$, in the case where the transducer **12** implements a thermoelectric cell, or comprised between $0.75\cdot V_{OC}$ and $0.9\cdot V_{OC}$, for example substantially equal to $0.8\cdot V_{OC}$, in the case where the transducer **12** implements, instead, a photovoltaic cell.

In any case, the value of the sampled voltage $V_{S\&H}$ corresponds to the value that the transduction signal V_{TRANSD} supplied by the transducer **12** assumes in a maximum efficiency or maximum coupling operating condition, i.e., to the optimized value V_{MPPT}, thus depending upon the electrical and constructional characteristics of the same transducer **12**.

The first holding capacitor $C1_{S\&H}$ is consequently charged to the aforesaid optimized value V_{MPPT} assumed by the sampled voltage $V_{S\&H}$.

In detail, the sampled voltage $V_{S\&H}$ is given by the following expression:

$$V_{S\&H}=V_{OC}\cdot R2_{S\&H}/(R1_{S\&H}+R2_{S\&H})$$

Consequently, the values of resistance of the dividing resistors $R1_{S\&H}$, $R2_{S\&H}$ are set, or regulated, in such a way that:

$$R2_{S\&H}=R1_{S\&H};\ R2_{S\&H}/(R1_{S\&H}+R2_{S\&H})=\tfrac{1}{2}$$

in the case where the transducer **12** implements a thermoelectric cell, and for example:

$$R2_{S\&H}=4\cdot R1_{S\&H};\ R2_{S\&H}/(R1_{S\&H}+R2_{S\&H})=0.8$$

in the case where the transducer **12** implements a photovoltaic cell, having a maximum efficiency in the condition $V_{MPPT}=0.8\cdot V_{OC}$.

It is noted that it is thus advantageous to provide at least one, or both, of the dividing resistors $R1_{S\&H}$, $R2_{S\&H}$ so that their resistance is configurable for generating the optimal value for the voltage V_{MPPT}.

Next (step **42**), once again with the tracking switch SW_{MPPT} open and the DC-DC converter **16** turned off, the values of the upper and lower threshold voltages VTH_{UP}, VTH_{DOWN}, at which the second and third holding capacitors $C2_{S\&H}$, $C3_{S\&H}$, respectively, are charged, are generated

$$VTH_{UP}=V_{S\&H}+V_{OS};\text{ and}$$

$$VTH_{DOWN}=V_{S\&H}-V_{OS}.$$

Then (step **44**), the sampling switch $SW_{S\&H}$ is driven into the opening condition, as likewise the first decoupling switch **SW1**. In this way, the voltage values stored in the holding capacitors $C2_{S\&H}$, $C3_{S\&H}$ are held, but for the leakage currents, which are in any case minimized with an appropriate design of the switching elements in the circuit.

Furthermore, the tracking switch SW_{MPPT} is driven into the closing condition, thus starting the step of tracking of the value of the transduction signal V_{TRANSD}, which enables substantially maximum efficiency and substantially maximum coupling to be obtained. The first comparison switch $SW1_{COMP}$ is further driven into the closing condition so that the second input terminal of the comparator **35** is at the upper threshold voltage VTH_{UP}.

As mentioned previously, the tracking step envisages that the DC-DC converter **16** is turned on/turned off via the hysteretic control of the comparator **24**, which generates the enabling signal EN_{DCDC}, causing the harvesting signal VIN_{DCDC} (and consequently the transduction signal V_{TRANSD}, given the presence of the short circuit defined by the tracking switch SW_{MPPT} in a closing condition) to have a variable trend between the upper and lower threshold voltages VTH_{UP}, VTH_{DOWN}, thus around the sampled voltage $V_{S\&H}$, i.e., the optimized value V_{MPPT}.

In detail, the DC-DC converter **16** stays off as long as the value of the harvesting signal VIN_{DCDC} is lower than the upper threshold voltage VTH_{UP}, as verified in step **45**.

As soon as the value of the harvesting signal VIN_{DCDC} exceeds the upper threshold voltage VTH_{UP}, step **46** (subsequent to step **45**), the enabling signal EN_{DCDC} switches (going, for example, to the high state), enabling the DC-DC converter **16**, which is consequently turned on.

It should be noted that switching of the same enabling signal EN_{DCDC} further controls closing of the second comparison switch $SW2_{COMP}$, so that the second input terminal of the comparator **35** goes to the lower threshold voltage VTH_{DOWN}, thus guaranteeing hysteretic operation of the comparator **25**.

Furthermore, activation of the DC-DC converter **16** entails a decrease in the value of the harvesting signal VIN_{DCDC}, since the average current in the DC-DC converter **16** (when it is on) is higher than the current supplied by the transducer **12**. In this step, the capacitance of the storage capacitor **15** is thus discharged with substantially constant current (when, instead, the DC-DC converter **16** is off, the same capacitance is charged by the transducer **12**).

In the same step **46**, count of a refresh time interval is started, following upon which, as described in detail hereinafter, the values of the sampled voltage $V_{S\&H}$ and of the threshold voltages VTH_{UP}, VTH_{DOWN} will have to be updated.

The DC-DC converter **16** remains on as long as the value of the harvesting signal VIN_{DCDC} is higher than the lower threshold voltage VTH_{DOWN}, as verified in step **47**.

As soon as the value of the harvesting signal VIN_{DCDC} drops below the lower threshold voltage VTH_{DOWN}, the enabling signal EN_{DCDC} switches again (going, for example, to the low state), disabling the DC-DC converter **16**, which is consequently turned off (step **48**).

Switching of the same enabling signal EN_{DCDC} further controls opening of the second comparison switch $SW2_{COMP}$ so that the second input terminal of the comparator **35** once again goes to the upper threshold voltage VTH_{UP}.

Furthermore, deactivation of the DC-DC converter **16**, entails an increase in the voltage of the harvesting signal VIN_{DCDC} on account of the current drawn by the transducer **12**.

The aforesaid steps of increase and decrease of the value of the harvesting signal VIN_{DCDC} (and consequently of the transduction signal V_{TRANSD}) repeats one after the other until the refresh time interval reaches a desired value (this value may conveniently be regulated, also during operation of the circuit), as verified in the same step **48**.

In this case, from step **48**, control returns to the initial step **40**, for a new sampling of the open-circuit voltage V_{OC} supplied by the transducer **12**, and refresh of the sampled voltage values and of the threshold values, in a manner altogether similar to what has been described previously.

The operation described will be better understood with reference also to the diagrams of FIG. **5**, which show the plots of the harvesting signal VIN_{DCDC} and of the transduction signal V_{TRANSD}, during the steps of tracking of the MPPT condition, with duration T_{MPPT}, and of refresh and sampling, with duration T_{SAMPLE}; further indicated are the time intervals in which the DC-DC converter **16** is on and those in which the DC-DC converter **16** is off, in each on/off cycle, having a duration T_{CYCLE}.

In particular, the time interval T_{MPPT} is much longer than the time interval T_{SAMPLE}: $T_{MPPT} >> T_{SAMPLE}$.

In general, time interval T_{MPPT} depends on the field of application, in particular on the fact that the environmental conditions to which the transducer **12** is subjected change rapidly; for example, the time interval T_{SAMPLE} may be of the order of some tens of milliseconds (for example, 25 ms) and the time interval T_{MPPT} may range from some seconds to some tens of seconds.

The above characteristic advantageously enables a considerable reduction of the average current consumption. During the tracking step, in effect, only the comparator **25** is on, with a resulting extremely low current consumption.

Purely by way of example, in the aforesaid FIG. **5**, the open-circuit voltage V_{OC} supplied by the transducer **12** increases with each refresh of the values sampled. It is, however, altogether evident that, upon start of the refresh time interval, the value of the transduction signal V_{TRANSD} may vary arbitrarily (for example, as a function of the changed environmental conditions, in terms of temperature or conditions of lighting), thus being able to increase, or even decrease, with respect to the immediately preceding step, according to the environmental stimulus.

As illustrated in FIGS. **6***a*, **6***b*, corresponding to an increase of the open-circuit voltage V_{OC} supplied by the transducer **12** is a decrease in the time interval T_{CYCLE}, defined previously; thus, a larger number of cycles during the step of tracking of the MPPT condition occur.

By way of example, FIG. **6***a* refers to a voltage V_{OC} of 0.4 V, whereas FIG. **6***b* refers to a voltage V_{OC} of 5 V.

FIG. **7** shows a possible circuit implementation for providing the first and second voltage-generator modules **32**, **34**, which enables substantially negligible leakage currents to be obtained.

In particular, the harvesting interface **14** (of which elements already described previously with reference to FIG. **3** are not described again) comprises in this case a voltage-tracking amplifier **50**, having its input connected to the third internal node N_3 and its output defining a sixth internal node N_6, onto which the sampled voltage $V_{S\&H}$ is thus brought.

Connected between the sixth internal node N_6 and the reference terminal GND is a first mirroring resistor **51** with resistance R_S, through which a mirroring current I_S, equal to $V_{S\&H}/R_S$, is consequently generated.

The harvesting interface **14** further comprises a current mirror **52** (obtained in a per se known manner, not described in detail herein), having a mirroring branch connected to the sixth internal node N_6 and a mirrored branch connected to a seventh internal node N_7, on which the mirroring current I_S is mirrored.

Connected between the seventh internal node N_7 and the reference terminal GND is a second mirroring resistor **54** with the same resistance R_S, so that on the seventh internal node N_7 there is the same sampled voltage $V_{S\&H}$.

The first voltage-generator module **32** is formed in this case by: a first current generator **55**, which may be selectively connected to the seventh internal node N_7 by a second decoupling switch SW**2** and generates a reference current I_{ref}; and a third decoupling switch SW**3**, designed to connect the seventh internal node N_7 selectively to the fourth internal node N_4, on which the upper threshold voltage VTH_{UP} is present during operation. The first current generator **55** may be obtained in any known way.

The second voltage-generator module **34** is in turn formed by: a second current generator **57**, which may be selectively connected to the same seventh internal node N_7, as an alternative to the first current generator **55**, via a fourth decoupling switch SW**4**, and generates the same reference current I_{ref}; and a fifth decoupling switch SW**5**, designed to connect the seventh internal node N_7 selectively to the fifth internal node N_5, on which the lower threshold voltage VTH_{DOWN} is present during operation. Also the second current generator **55** may be obtained in any known way.

The second, third, fourth, and fifth decoupling switches receive respective control signals from the same timing stage **25** (in a way not illustrated here, and as it has been discussed previously).

Operation of the circuit described envisages, as mentioned previously, in an initial step, sampling of the open-circuit voltage of the transducer **12** by closing of the sampling switch $SW_{S\&H}$ and of the first decoupling switch SW**1**, and consequent generation on the third internal node N**3** of the sampled voltage $V_{S\&H}$, having a value corresponding to the optimized value V_{MPPT}.

Next, the sampling switch $SW_{S\&H}$ and the first decoupling switch SW**1** are both opened, and the sampled voltage $V_{S\&H}$ is held on the first holding capacitor $C\mathbf{1}_{S\&H}$.

Next, but once again within the sampling time interval T_{SAMPLE}, the second and third decoupling switches SW**2**, SW**3** are first closed (with the fourth and fifth decoupling switches SW**4**, SW**5** open) so that on the seventh internal node N**7**, and consequently on the fourth internal node N**4**, the voltage $VTH_{UP}=V_{S\&H}+I_{REF}\cdot R_S$ is generated (note that the aforesaid offset voltage V_{OS} consequently corresponds here to $I_{REF}\cdot R_S$).

Next, once again within the sampling time interval T_{SAMPLE}, the second and third decoupling switches SW**2**, SW**3** are opened, and the fourth and fifth decoupling switches SW**4**, SW**5** are closed so that the voltage $VTH_{DOWN}=V_{S\&H}-I_{REF}\cdot R_S$ is generated on the seventh internal node N**7**, and consequently on the fifth internal node N**5**.

Next, also the fourth and fifth decoupling switches SW**4**, SW**5** are opened so that the upper and lower threshold voltages VTH_{UP}, VTH_{DOWN} are held on the respective second and third holding capacitors $C\mathbf{2}_{S\&H}$, $C\mathbf{3}_{S\&H}$ (for the entire duration of the subsequent tracking time interval T_{MPPT}), with minimal leakage currents through the open switches.

In this regard, the present Applicant has realized: a substantially maximum dispersion by the holding capacitors $C\mathbf{2}_{S\&H}$, $C\mathbf{3}_{S\&H}$ equal to 10 mV/s with an open-circuit voltage V_{OC} of 0.4 V (that is, equal to 5% of the optimal voltage V_{MPPT}, in the example 0.2 V, considering a duration of the time interval T_{MPPT} of one second); and a substantially maximum dispersion by the same holding capacitors $C\mathbf{2}_{S\&H}$, $C\mathbf{3}_{S\&H}$ equal to 100 mV/s with an open-circuit voltage V_{OC} of 5V (equal approximately to 2% of the optimal voltage V_{MPPT}, in the example 2.5V, considering once again a duration of the time interval T_{MPPT} of one second).

The performance achieved by the harvesting interface **14** is further highlighted by the plots of FIGS. **8***a*-**8***d*, which refer to use with a transducer **12** of a thermoelectric type, having the following characteristics: series resistance R_{TEG} of 20 kΩ, open-circuit voltage V_{OC} comprised between 1 V (minimum value) and 5 V (maximum value).

In detail, FIG. **8***a* shows the average consumption of the S&H stage **22** as the average power delivered by the transducer **12** (available power P_{MAX}) varies. It should be noted that this consumption is very contained and barely dependent upon the average power delivered, and thus barely dependent upon the environmental conditions in which the transducer **12** operates and upon the electrical characteristics of the same transducer **12**.

FIG. **8***b* shows the plot of the factor η_{LOSS} (see the foregoing discussion) as a function of the same power P_{MAX}.

FIG. **8***c* shows the plot of the factor η_{COUPLE} (see the foregoing discussion), once again as a function of the power P_{MAX}.

FIG. **8***d* shows the plot of the efficiency η_{MPPT} (see the foregoing discussion) once again as a function of the power P_{MAX}.

In particular, it is to be noted that the efficiency η_{MPPT} is higher than 90% with an available power of 20 μW, and rises above 98% with an available power higher than 100 μW.

The advantages of the proposed solution emerge clearly from the foregoing description.

In particular, the hysteretic voltage control for tracking of the MPPT condition enables a considerable saving in power consumption, for example as compared to a continuous-time control solution.

For instance, the condition $T_{MPPT}>>T_{SAMPLE}$ enables considerable accuracy in tracking of the aforesaid MPPT condition and a very high value of the factor η_{COUPLE} to be achieved.

Furthermore, the harvesting interface **14** does not envisage the use of further DC/DC converters to ensure the MPPT condition, thus reducing the number of external components required; a voltage control is in fact implemented, instead of a duty-cycle control, as in many known solutions.

In this regard, it is further pointed out that in the solution described, the holding capacitors are advantageously all obtained using integrated technology, not as external components (in fact, thanks to the reduction of the leakage currents, the value of the same capacitors may not be high).

The solution described further proves very flexible, enabling easy adaptation to use with different photovoltaic or thermoelectric cells; in particular, it is sufficient to regulate the division factor of the sampled voltage $V_{S\&H}$ via the voltage divider **30** to obtain a condition of improved matching.

Furthermore, once the operating mode has been selected, with a thermoelectric cell or a photovoltaic cell, the efficiency of tracking of the MPPT condition is found to be independent of the electrical parameters of the transducer **12**, for example of the corresponding equivalent resistance or the corresponding open-circuit voltage.

As mentioned previously, the energy harvesting system may advantageously be used for electrical supply of a device, which may even be without any battery, or equipped with a rechargeable battery.

By way of example, FIG. **9** shows an electronic bracelet or watch **60**, which incorporates the energy harvesting system for generation of electrical energy by exploiting the difference of temperature between the environment and the human body (the transducer **12** is in this case of a thermoelectric type). The electronic bracelet **60** may advantageously be used in the field of fitness, for example for counting the steps made by a user.

Finally, it is clear that modifications and variations may be made to what has been described and illustrated herein, without thereby departing from the scope of the present disclosure, as defined in the annexed claims.

In particular, the energy harvesting system **10** may comprise a plurality of transducers **12**, all of the same type or of a type different from one another.

Furthermore, it is evident that the energy harvesting system **10** may advantageously be used for other applications and other electronic devices, for example in the automotive field, or also in a mobile electronic device or in a garment or other article of clothing, for example in footwear, in the consumer electronics field (any mobile application), the industrial field (for example, in controlling processes that involve environments with high thermal gradients), or in the field of home automation (for example, in combination with photovoltaic generators).

The various embodiments described above can be combined to provide further embodiments. These and other changes can be made to the embodiments in light of the above-detailed description. In general, in the following claims, the terms used should not be construed to limit the claims to the specific embodiments disclosed in the specification and the claims, but should be construed to include all possible embodiments along with the full scope of equivalents to which such claims are entitled. Accordingly, the claims are not limited by the disclosure.

The invention claimed is:

1. An electrical-energy harvesting system, comprising:

a transducer to convert energy from an environmental energy source into a transduced signal;

a harvesting interface to receive the transduced signal on an input terminal and to supply an electrical-energy harvesting signal on an output terminal;

an energy storage element coupled to the output terminal of the harvesting interface to receive the electrical-energy harvesting signal;

a voltage converter coupled to the output terminal to generate a regulated voltage, wherein the harvesting interface is configured to:

sample an open-circuit voltage value of the transduced signal;

couple the output terminal to the input terminal during a tracking operating condition;

generate an optimized voltage value starting from the open-circuit voltage value;

decouple the output terminal from the input terminal during a refresh operating condition of sampling said open-circuit voltage value;

generate an upper threshold voltage based on the optimized voltage value during said refresh operating condition;

generate a lower threshold voltage based on the optimized voltage value during said refresh operating condition; and

control the voltage converter in a switching mode to thereby maintain the electrical-energy harvesting signal between the upper and lower threshold voltages in said tracking operating condition.

2. The system according to claim **1**, wherein the tracking operating condition is a condition of maximum power-point tracking.

3. The system according to claim **2**, wherein the harvesting interface further comprises:

a timing stage to generate control and timing signals to first determine said tracking operating condition and then to determine said refresh operating condition.

4. The system according to claim **3**, wherein said timing stage is configured to:

cyclically determine execution of said refresh operating condition to direct updating said open-circuit voltage value and said upper and lower threshold voltages and execution of said tracking operating condition following upon each new execution of said refresh operating condition.

5. The system according to claim **1**, wherein said optimized voltage value corresponds to a condition of substantially maximum coupling between said transducer and said harvesting interface.

6. The system according to claim **1**, wherein:

$$VTH_{DOWN}<V_{MPPT}<VTH_{UP},$$

where V_{MPPT} is the optimized voltage value, VTH_{DOWN} is the lower threshold voltage, and VTH_{UP} is the upper threshold voltage.

7. The system according to claim **1**, wherein the harvesting interface comprises

a voltage divider to generate the optimized voltage value starting from a sampled value of said open-circuit voltage, said optimized voltage being substantially equal to 50 percent of the sampled value of said open circuit voltage supplied by the transducer.

8. The system according to claim **1**, wherein the harvesting interface comprises:

a voltage divider to generate the optimized voltage value starting from a sampled value of said open-circuit voltage, said optimized voltage being between about 75 percent and about 90 percent of the sampled value of said open circuit voltage supplied by the transducer.

9. An electronic device, comprising:

a transducer to convert energy from an environmental or mechanical energy source into a transduced signal;

a harvesting interface to receive the transduced signal on an input terminal and to supply an electrical-energy harvesting signal on an output terminal;

an energy storage element coupled to the output terminal of the harvesting interface to receive the electrical-energy harvesting signal;

a voltage converter coupled to the output terminal to generate a regulated voltage, wherein the harvesting interface is configured to:

sample an open-circuit voltage value of the transduced signal;

couple the output terminal to the input terminal during a tracking operating condition;

control the voltage converter in a switching mode to thereby maintain the electrical-energy harvesting signal between an upper threshold voltage and a lower threshold voltages in said tracking operating condition;

decouple the output terminal from the input terminal during a refresh operating condition of sampling said open-circuit voltage value;

generate said upper and lower threshold voltages during said refresh operating condition; and

cyclically execute said tracking operating condition following upon each new execution of said refresh operating condition; and

an electrical load to receive the regulated voltage.

10. The electronic device according to claim **9**, wherein said electrical load includes a rechargeable battery.

11. The electronic device according to claim **9**, wherein said transducer includes at least one of a photovoltaic and a thermoelectric device.

12. The electronic device according to claim **9**, wherein said electronic device is a mobile electronic device, an automotive device, an industrial device, an article of clothing, or a fitness device.

13. The electronic device according to claim **9**, wherein said tracking operating condition is a condition of maximum power-point tracking.

14. The electronic device according to claim **9**, wherein:

$$VTH_{DOWN}<V_{MPPT}<VTH_{UP},$$

where V_{MPPT} is the optimized voltage value, VTH_{DOWN} is a lower threshold voltage, and VTH_{UP} is an upper threshold voltage.

15. An electrical-energy harvesting method, comprising:

by a transducer, converting energy from an environmental energy source into a transduced signal;

by a harvesting interface:

receiving the transduced signal on an input terminal;

supplying an electrical-energy harvesting signal on an output terminal;

sampling an open-circuit voltage value of the transduced signal;

coupling the output terminal to the input terminal during a tracking operating condition;

generating an optimized voltage value starting from the open-circuit voltage value;

decoupling the output terminal from the input terminal during a refresh operating condition of sampling said open-circuit voltage value;

generating an upper threshold voltage based on the optimized voltage value during said refresh operating condition;

generating a lower threshold voltage based on the optimized voltage value during said refresh operating condition; and

controlling a voltage converter in a switching mode to maintain the electrical-energy harvesting signal between the upper and lower threshold voltages in said tracking operating condition;

by an energy storage element coupled to the output terminal of the harvesting interface, receiving the electrical-energy harvesting signal and storing electrical energy; and

by the voltage converter coupled to the output terminal, generating a regulated voltage.

16. The method according to claim **15**, wherein said tracking operating condition is a condition of maximum power-point tracking.

17. The method according to claim **16**, further comprising, by a timing stage of the harvesting interface:

generating control and timing signals to first determine said tracking operating condition and then to determine said refresh operating condition.

18. The method according to claim **17**, further comprising, by the timing stage:

cyclically determining execution of said refresh operating condition;

updating said open-circuit voltage value and said upper and lower threshold voltages; and

executing said tracking operating condition following upon each new execution of said refresh operating condition.

19. The method according to claim **15**, wherein said transducer includes at least one of a photovoltaic and a thermoelectric device.

20. The method according to claim **15**, wherein said transducer is integrated in a mobile electronic device, an automotive device, an industrial device, an article of clothing, or a fitness device.

* * * * *

US 20150234416A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2015/0234416 A1**

Ramorini et al. (43) **Pub. Date: Aug. 20, 2015**

(54) **ENERGY HARVESTING INTERFACE WITH IMPROVED IMPEDANCE MATCHING, METHOD FOR OPERATING THE ENERGY HARVESTING INTERFACE, AND ENERGY HARVESTING SYSTEM USING THE ENERGY HARVESTING INTERFACE**

(71) Applicant: **STMicroelectronics S.r.l.**, Agrate Brianza (IT)

(72) Inventors: **Stefano Ramorini**, Arluno (IT); **Alessandro Gasparini**, Cusano Milanino (IT); **Alberto Cattani**, Cislago (IT); **Federica Papotti**, San Giorgio Di Mantova (IT)

(73) Assignee: **STMicroelectronics S.r.l.**, Agrate Brianza (IT)

(21) Appl. No.: **14/616,119**

(22) Filed: **Feb. 6, 2015**

(30) **Foreign Application Priority Data**

Feb. 14, 2014 (IT) TO2014A000128

Publication Classification

(51) **Int. Cl.**
G05F 5/00 (2006.01)

(52) **U.S. Cl.**
CPC *G05F 5/00* (2013.01)

(57) **ABSTRACT**

An energy harvesting interface receives an electrical signal from an inductive transducer and supplies a supply signal. The interface includes an input branch with a first switch and a second switch connected together in series between a first input terminal and an output terminal. The interface further includes a third switch and a fourth switch connected together in series between a second input terminal and the output terminal. A first electrical-signal-detecting device, coupled across the second switch, detects a first threshold value of an electric storage current in the inductor of the transducer. A second electrical-signal-detecting device, coupled across the fourth switch, detects whether the electric supply current that flows through the fourth switch reaches a second threshold value lower than the first threshold value. The second threshold is derived from the electric storage current.

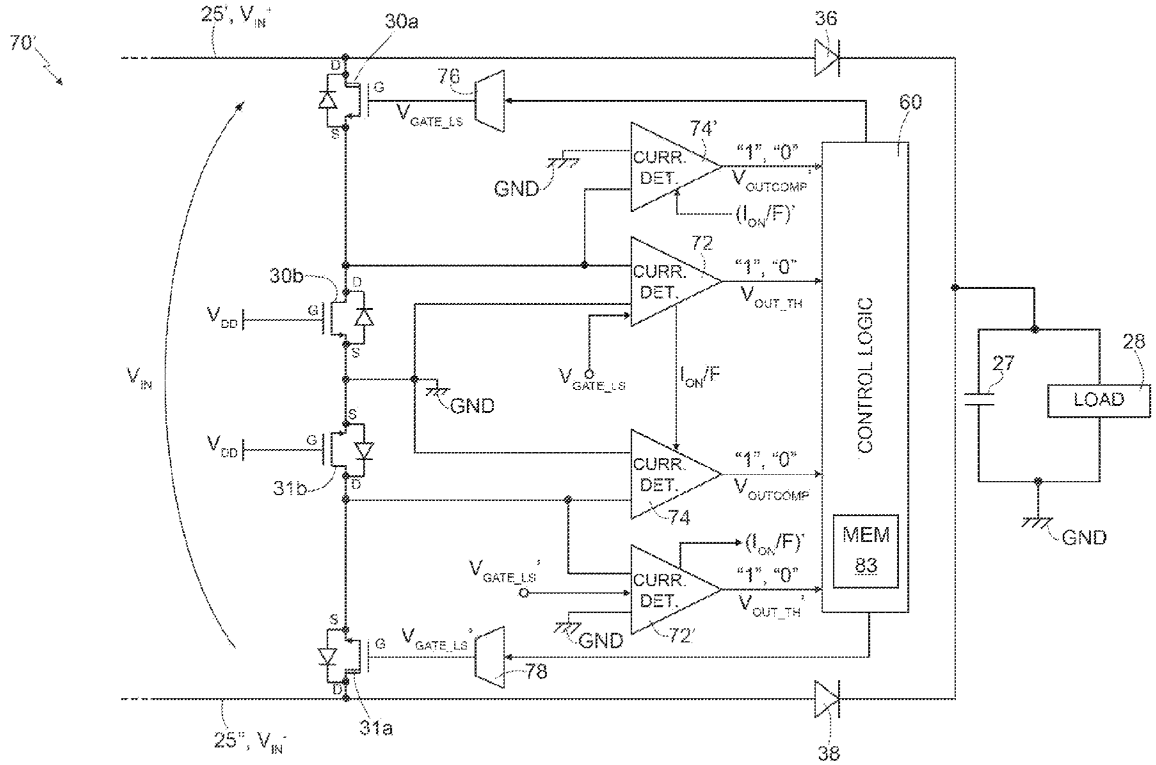
70'
25', V_{IN}^{+}
30a
36
76
D
G
S
V_{GATE_LS}
60
74'
CURR. DET.
"1", "0"
$V_{OUTCOMP}'$
GND
$(I_{ON}/F)'$
72
CURR. DET.
"1", "0"
V_{OUT_TH}
30b
V_{DD}
V_{GATE_LS}
I_{ON}/F
GND
V_{IN}
CONTROL LOGIC
27
28
LOAD
V_{DD}
31b
CURR. DET.
"1", "0"
$V_{OUTCOMP}$
74
MEM
83
GND
$(I_{ON}/F)'$
V_{GATE_LS}'
CURR. DET.
"1", "0"
V_{OUT_TH}'
GND
72'
V_{GATE_LS}'
78
31a
25", V_{IN}^{-}
38

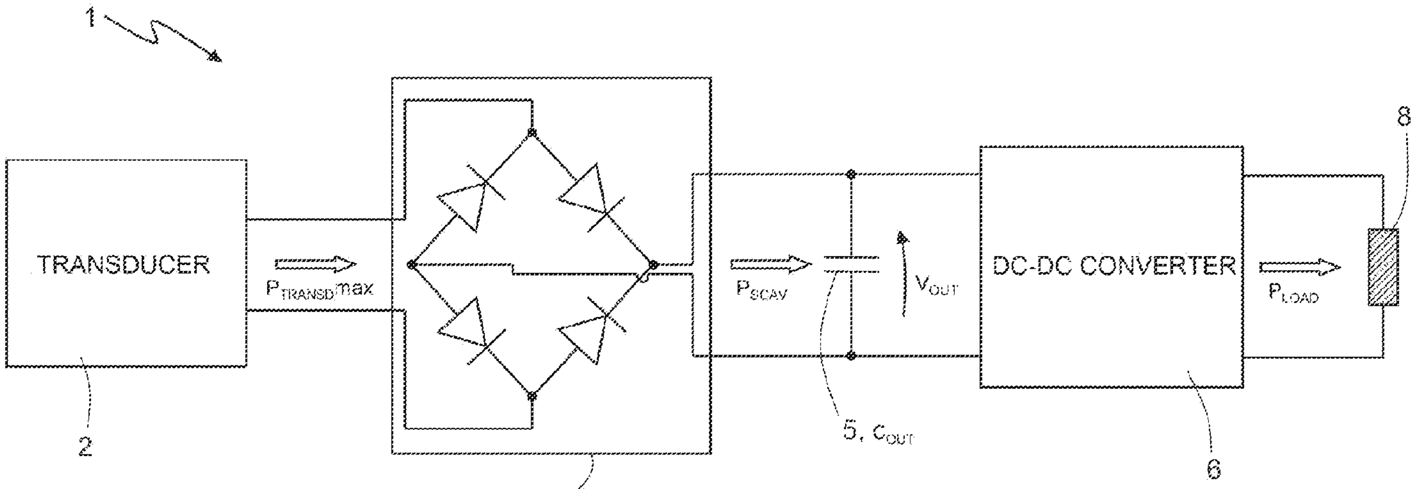
1
TRANSDUCER
2
$P_{TRANSD}max$
4
P_{SCAV}
5, C_{OUT}
V_{OUT}
DC-DC CONVERTER
6
P_{LOAD}
8

FIG. 1

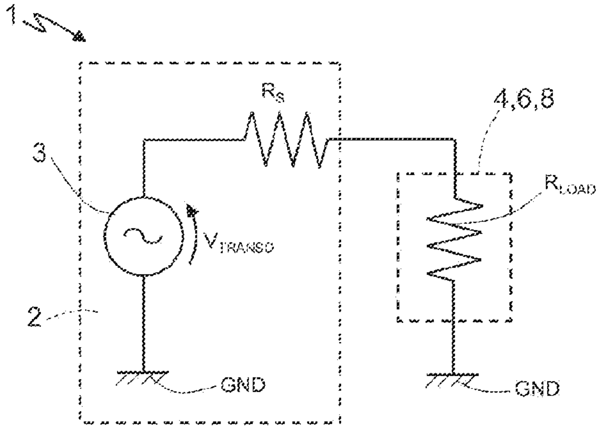
1
3
R_S
V_{TRANSD}
2
GND
4,6,8
R_{LOAD}
GND

FIG. 2

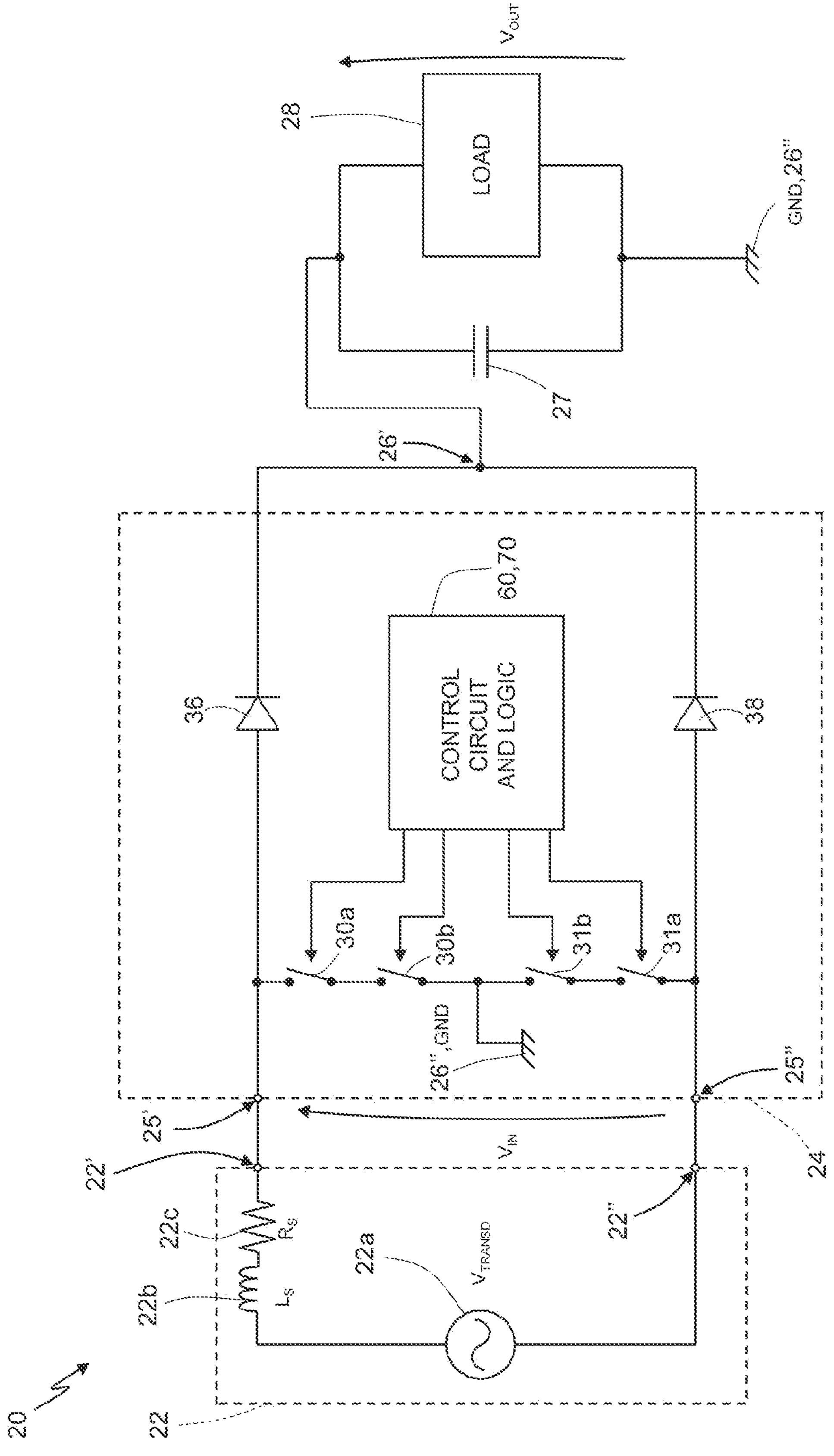
V_OUT
28
LOAD
GND,26''
27
26'
60,70
CONTROL CIRCUIT AND LOGIC
36
38
30a
30b
31b
31a
26'',GND
25'
25''
24
V_IN
22'
22''
22c
R_S
22b
L_S
22a
V_TRANSD
20
22

FIG. 3A

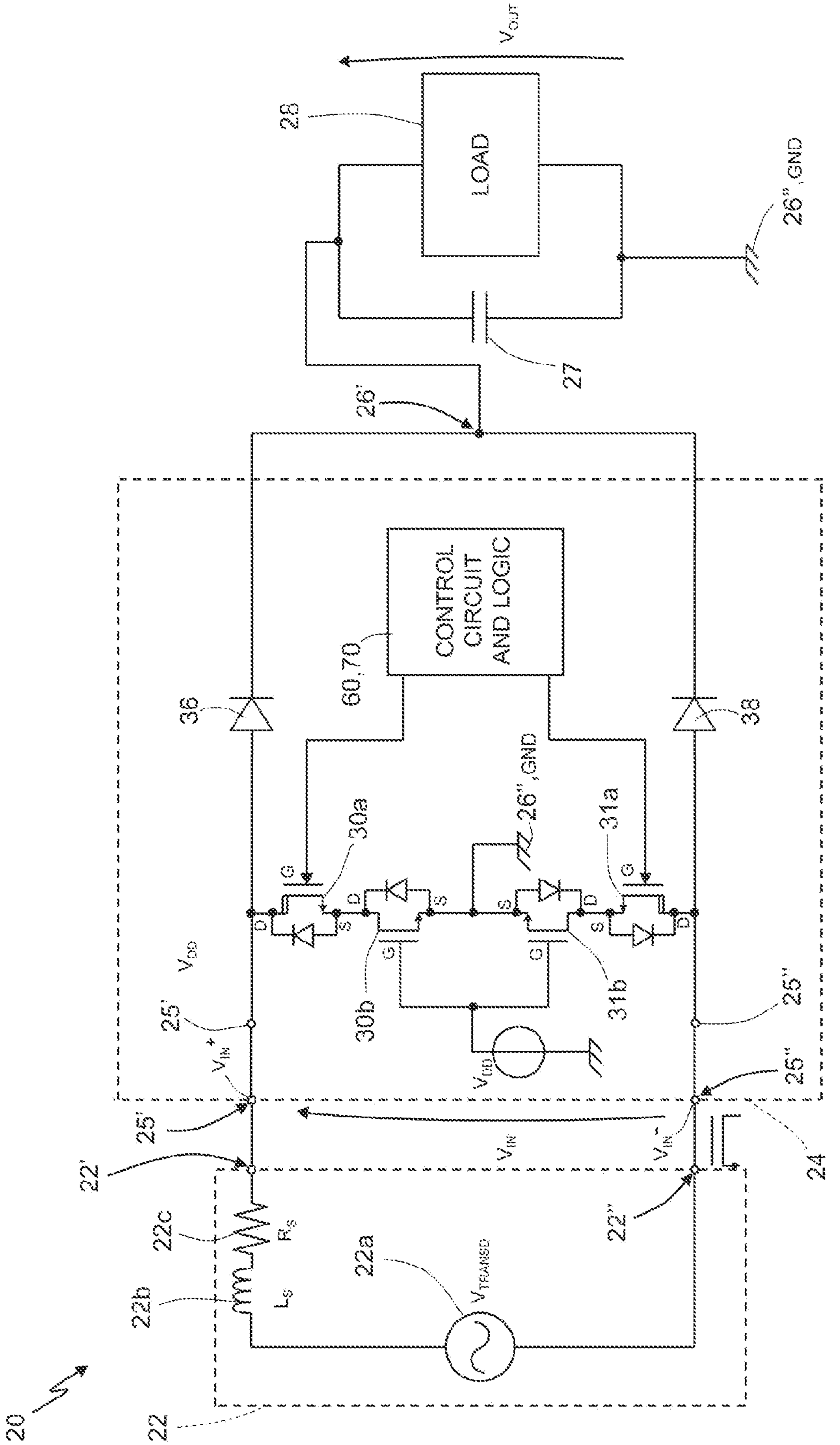
20
22
22a
V_{TRANSD}
22b
L_S
22c
R_S
22'
22''
V_{IN}
V_{IN}^+
V_{IN}^-
25'
25''
24
V_{DD}
V_{DD}
30a
30b
31a
31b
G
D
S
26'',GND
36
38
60,70
CONTROL CIRCUIT AND LOGIC
26'
27
28
LOAD
V_{OUT}
26'',GND

FIG. 3B

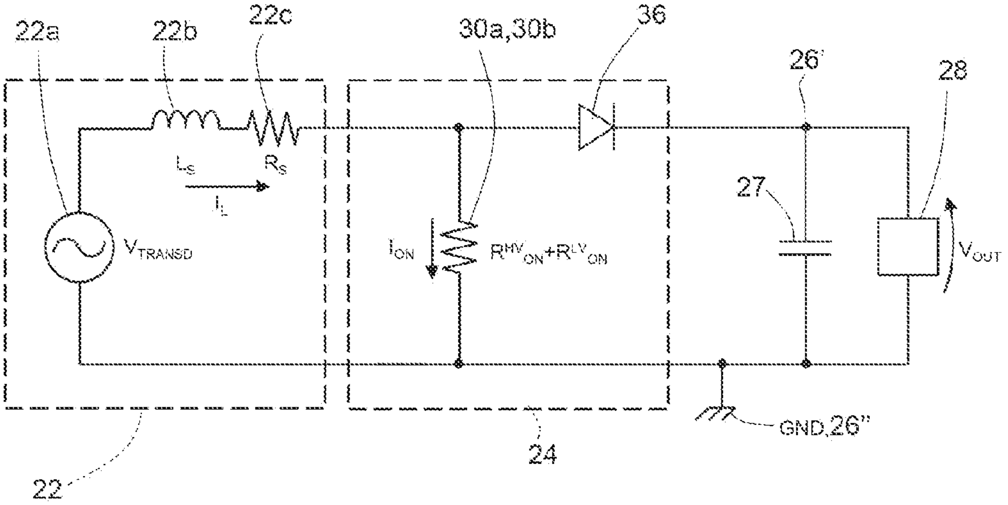
22a
22b
22c
30a,30b
36
26'
28
L_S
R_S
I_L
27
V_{TRANSD}
I_{ON}
$R^{HV}_{ON}+R^{LV}_{ON}$
V_{OUT}
GND,26''
24
22

FIG. 4A

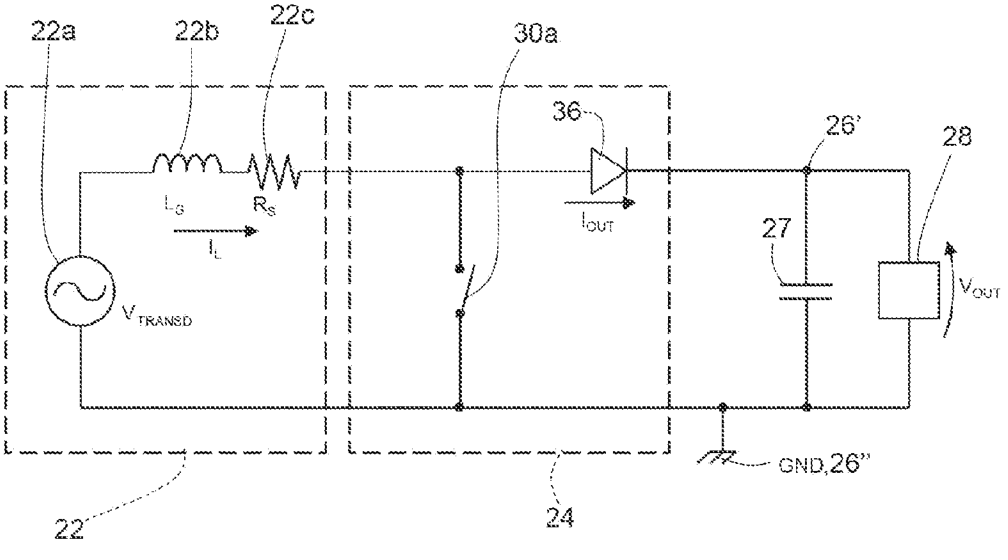
22a
22b
22c
30a
36
26'
28
L_S
R_S
I_L
I_{OUT}
27
V_{TRANSD}
V_{OUT}
GND,26''
24
22

FIG. 4B

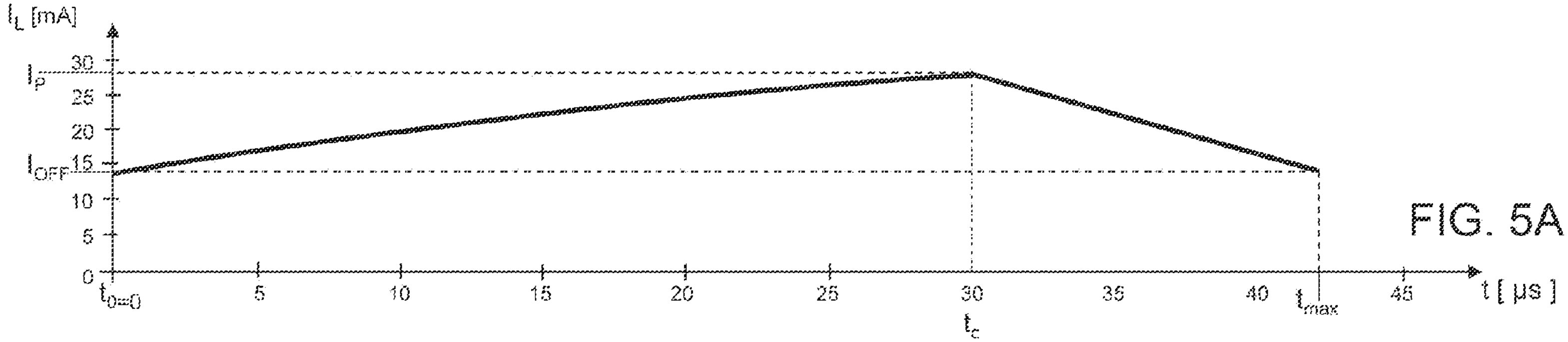

FIG. 5A

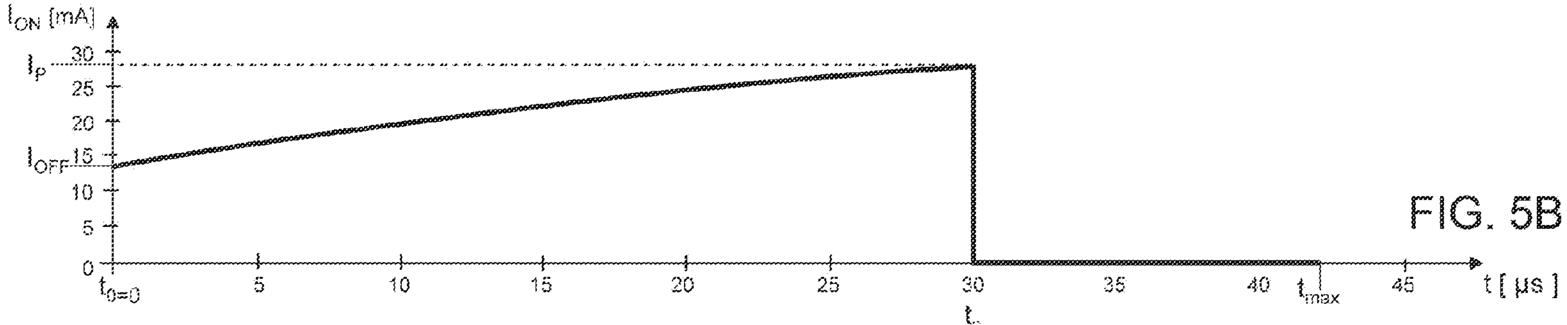

FIG. 5B

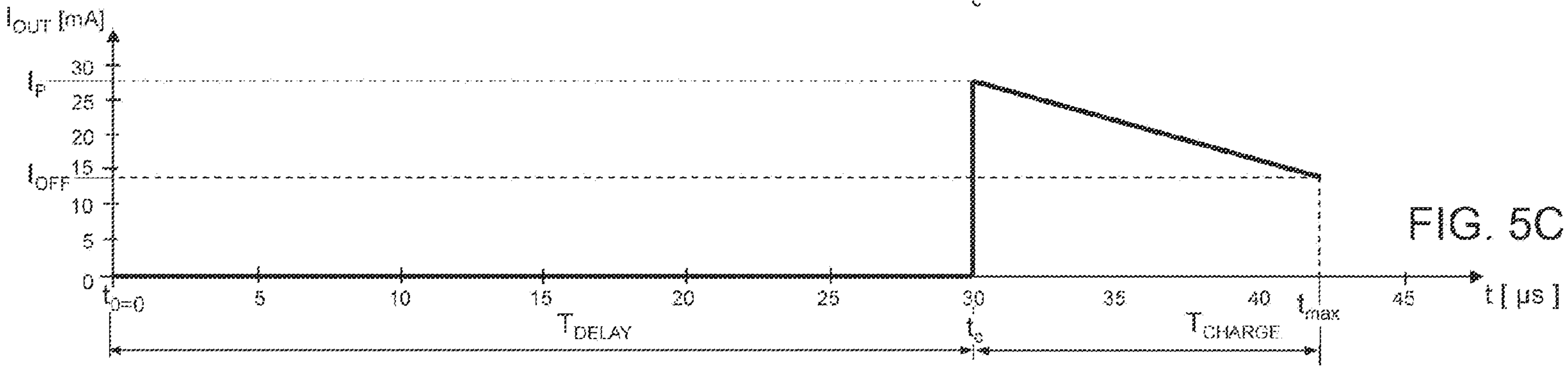

FIG. 5C

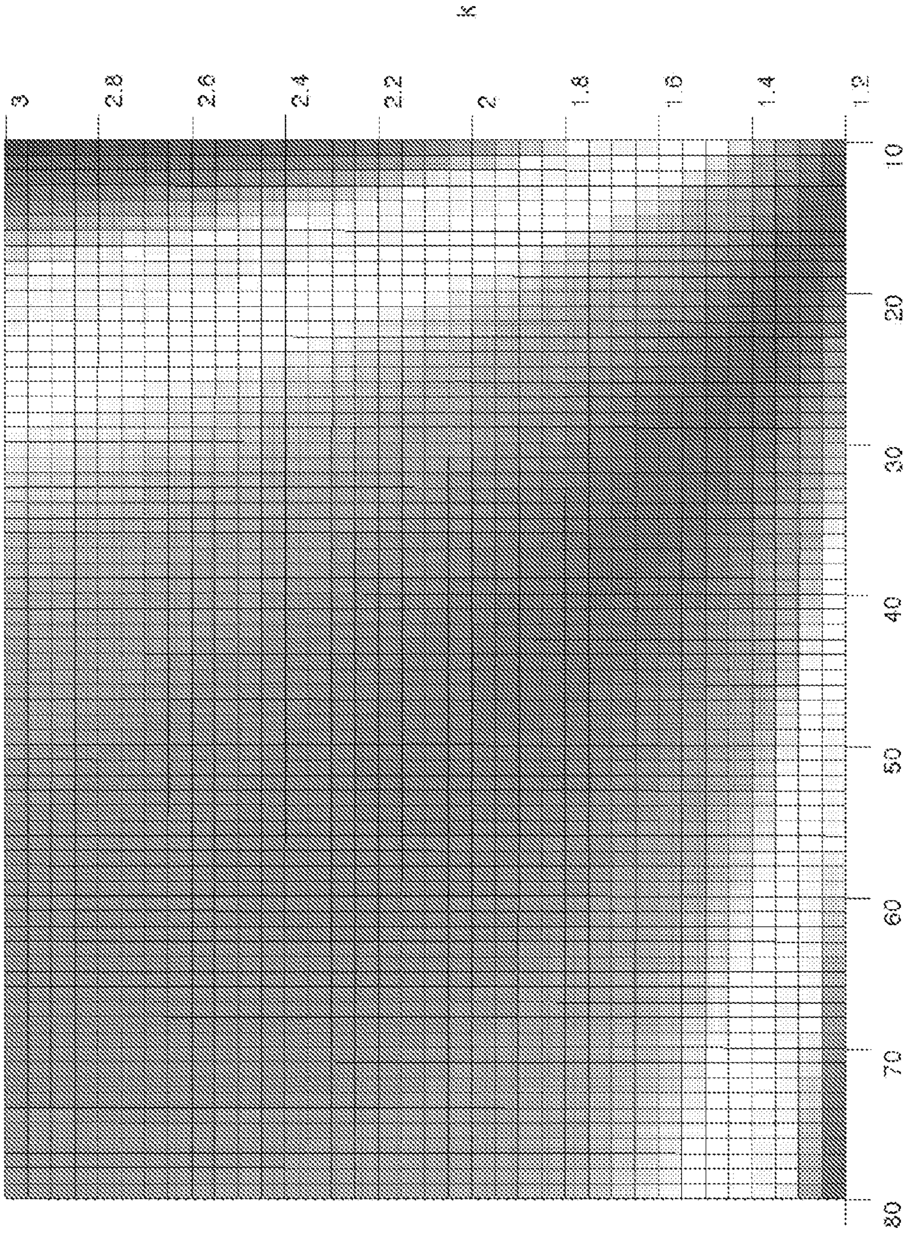

FIG. 6

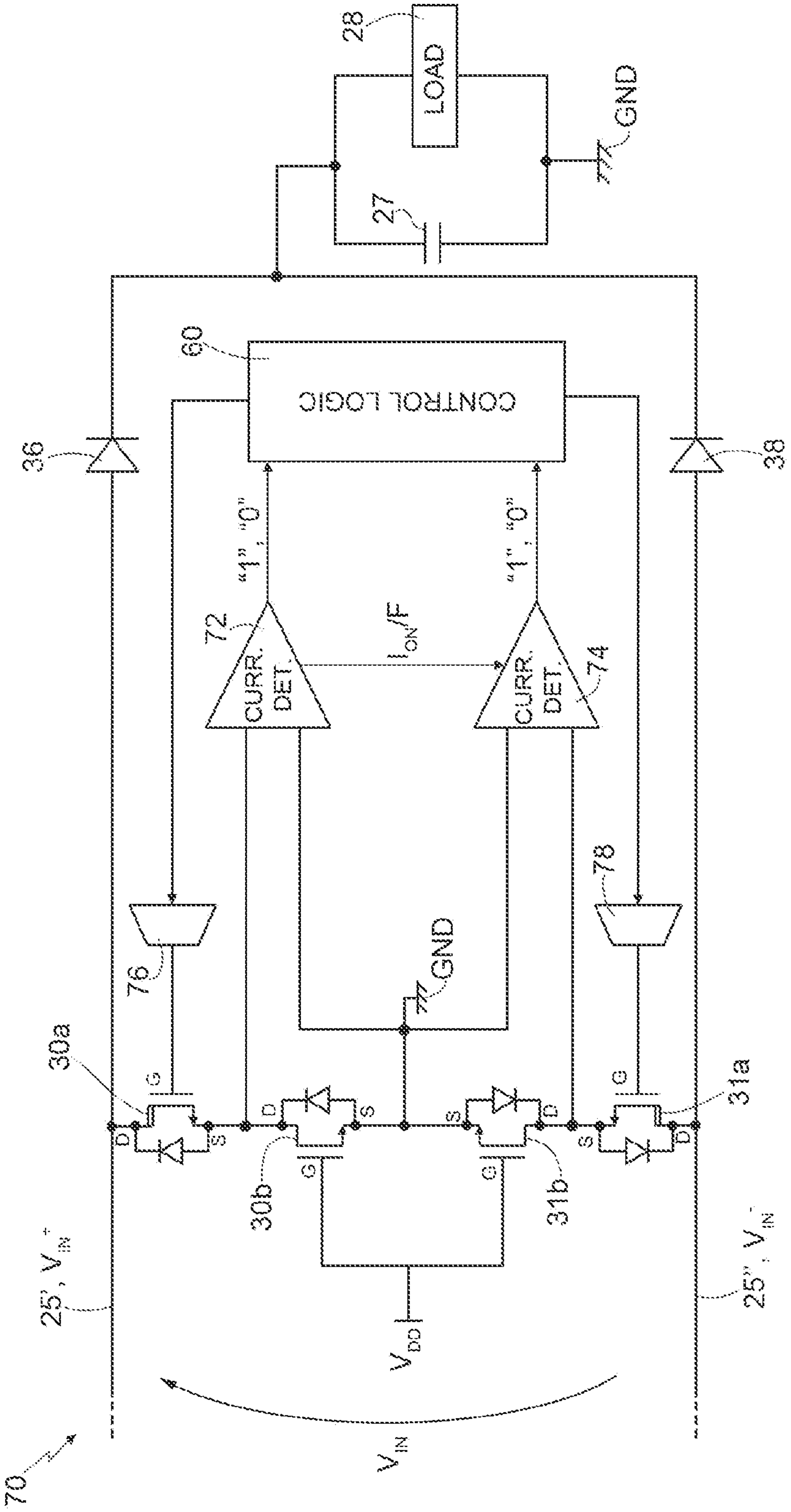
28
LOAD
GND
27
60
CONTROL LOGIC
36
38
"1", "0"
"1", "0"
72
CURR. DET.
I_{ON}/F
CURR. DET.
74
76
78
GND
30a
G
D
S
D
S
S
D
G
31a
30b
G
G
31b
D
S
V_{DD}
25', V_{IN}^+
25", V_{IN}^-
V_{IN}
70

FIG. 7

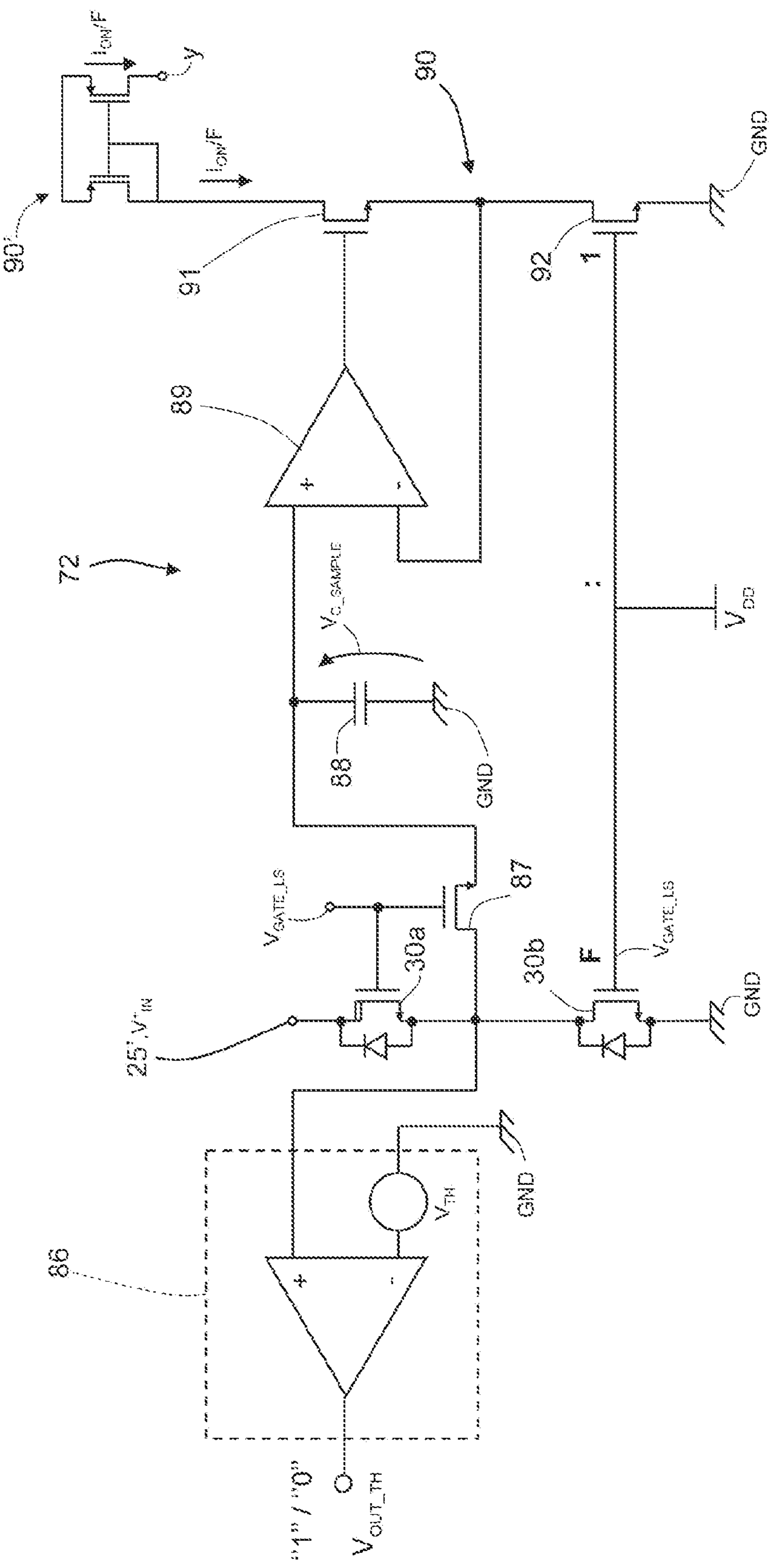
90'
I_ON/F
y
I_ON/F
90
GND
91
92
1
89
+
-
72
V_C_SAMPLE
88
GND
:
V_DD
V_GATE_LS
87
30a
25' V'_IN
F
30b
V_GATE_LS
GND
86
+
-
V_TH
GND
"1" / "0"
V_OUT_TH

FIG. 8

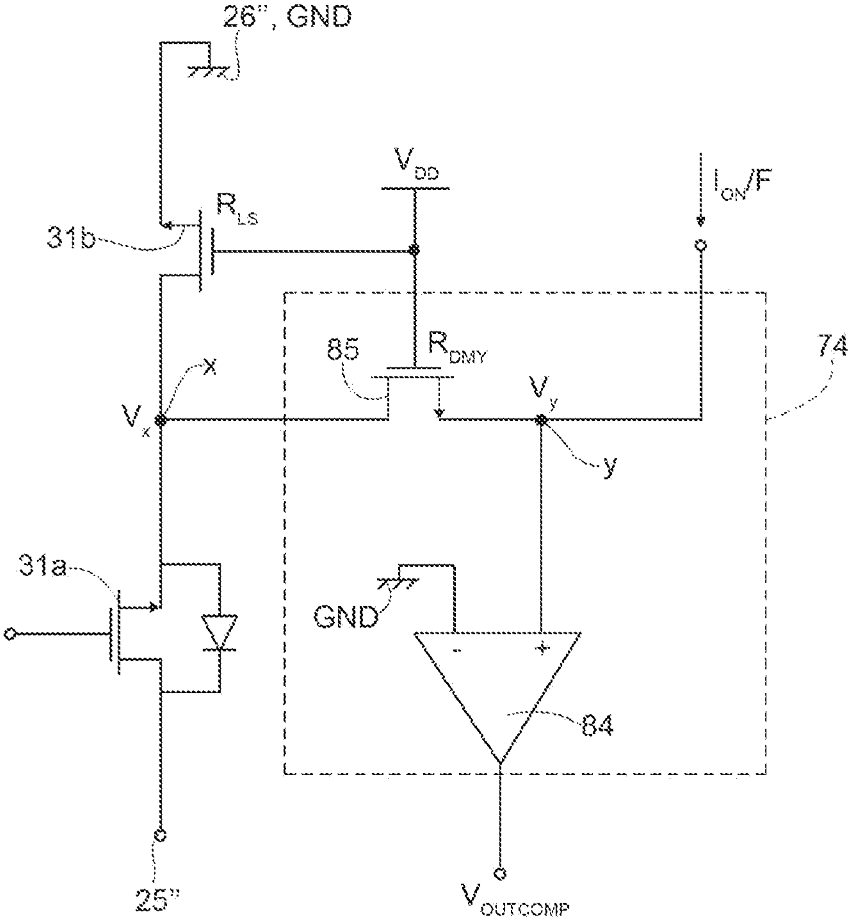
26", GND
V_{DD}
I_{ON}/F
R_{LS}
31b
85
R_{DMY}
74
x
V_x
V_y
y
31a
GND
-
+
84
25"
$V_{OUTCOMP}$

FIG. 9

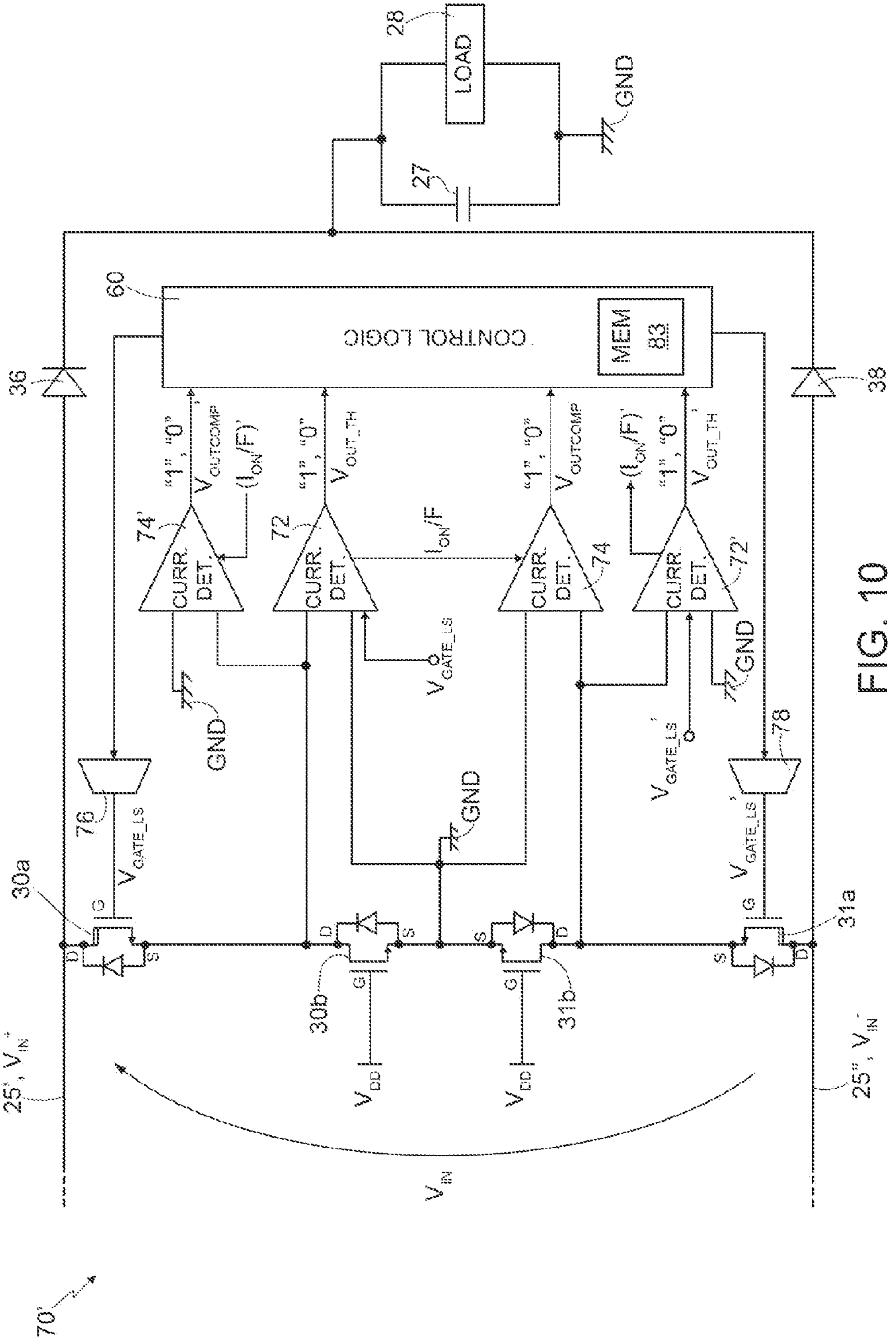
70'
25', V_{IN}^{+}
25", V_{IN}^{-}
V_{IN}
30a
30b
31a
31b
D
G
S
V_{DD}
V_{DD}
V_{GATE_LS}
V_{GATE_LS}'
76
78
GND
GND
GND
CURR. DET.
72
72'
74
74'
V_{GATE_LS}
V_{GATE_LS}'
I_{ON}/F
"1", "0"
V_{OUT_TH}
V_{OUT_TH}'
$V_{OUTCOMP}$
$V_{OUTCOMP}$'
$(I_{ON}/F)'$
CONTROL LOGIC
60
MEM
83
36
38
27
LOAD
28
GND

FIG. 10

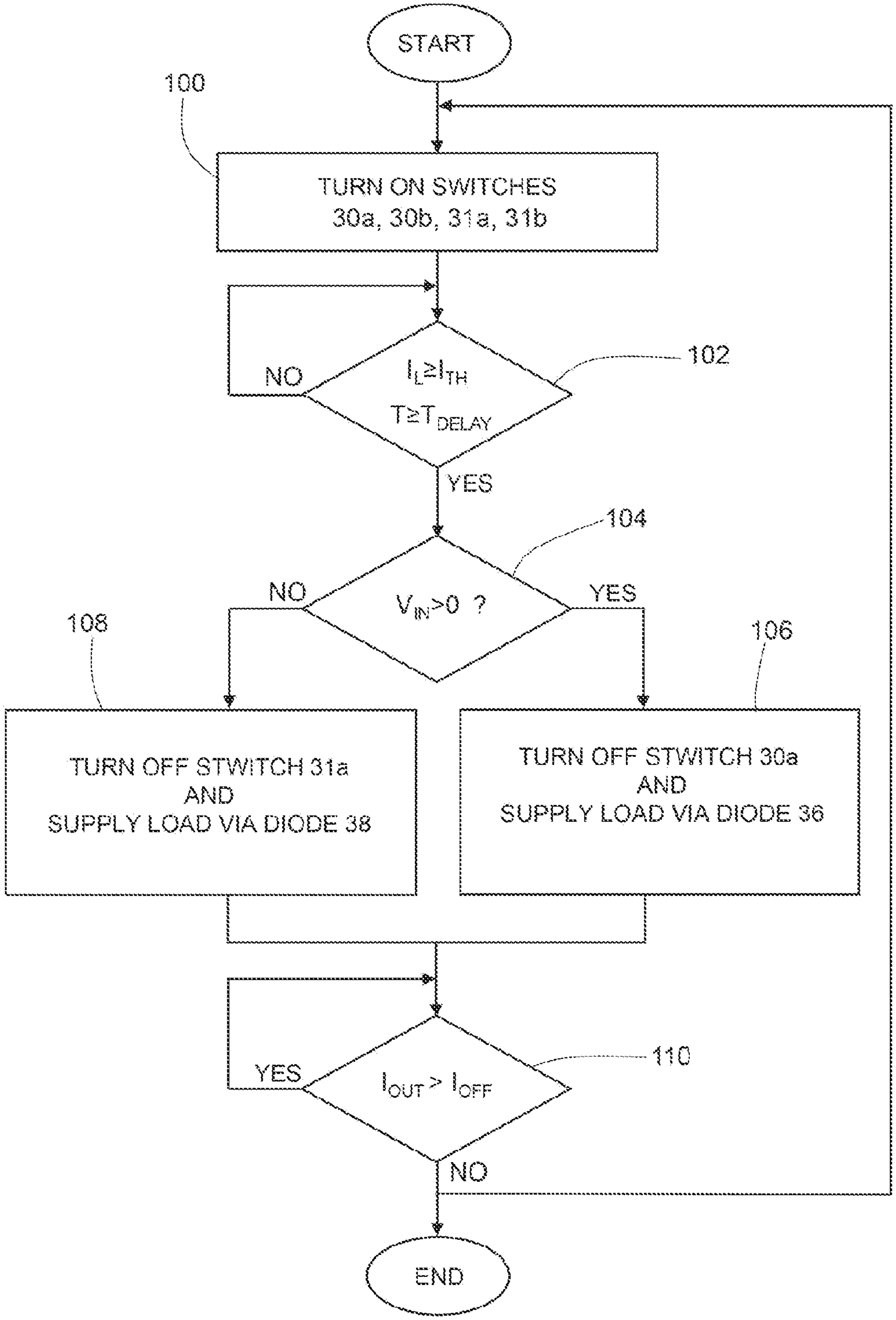
START
100
TURN ON SWITCHES
30a, 30b, 31a, 31b
102
NO
$I_L \geq I_{TH}$
$T \geq T_{DELAY}$
YES
104
NO
$V_{IN} > 0$?
YES
108
106
TURN OFF STWITCH 31a
AND
SUPPLY LOAD VIA DIODE 38
TURN OFF STWITCH 30a
AND
SUPPLY LOAD VIA DIODE 36
110
YES
$I_{OUT} > I_{OFF}$
NO
END

FIG. 11

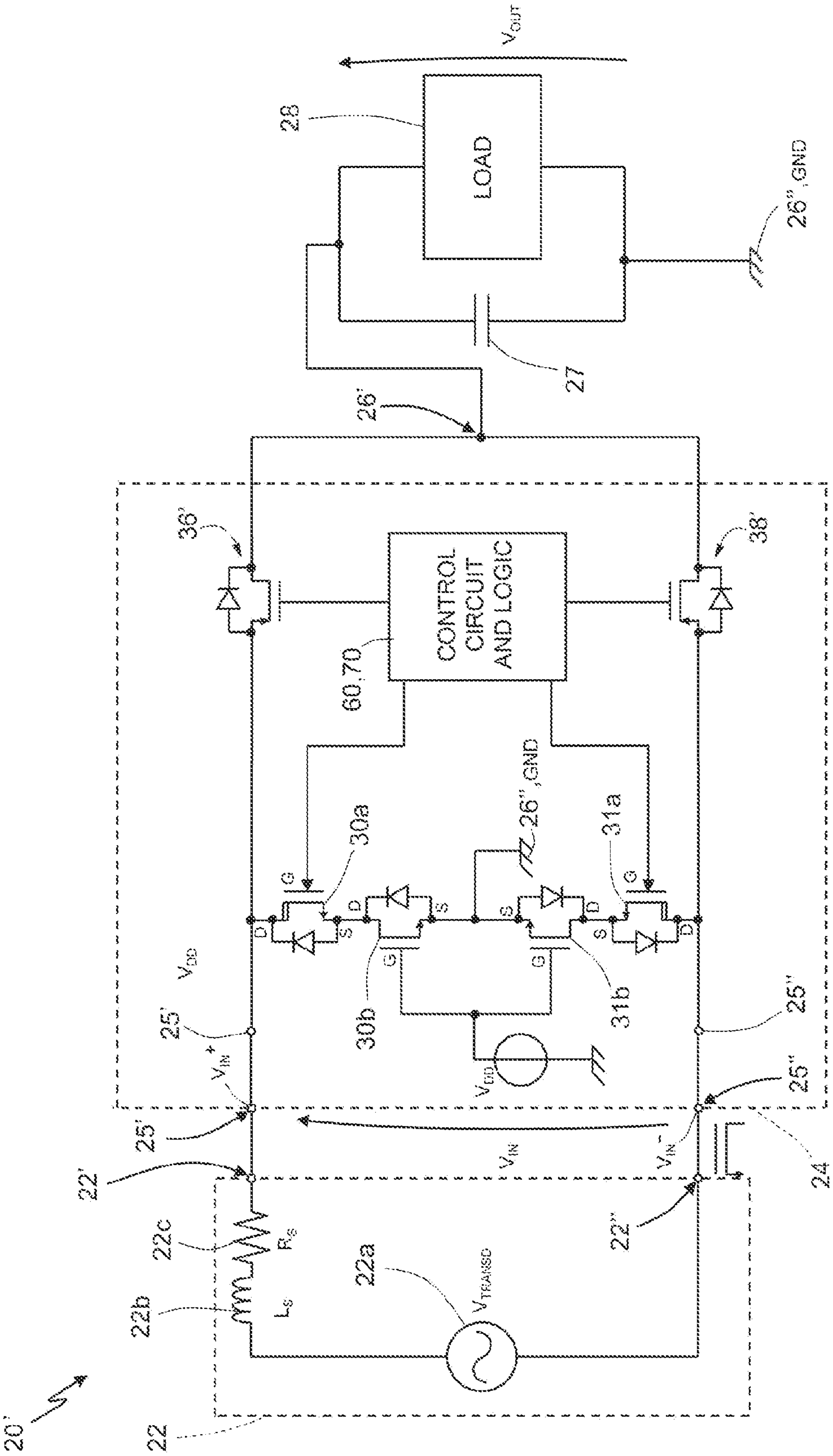
20'
22
22a
V_{TRANSD}
22b
L_S
22c
R_S
22'
22''
V_{IN}
V_{IN}^+
V_{IN}^-
25'
25''
24
V_{DD}
30a
30b
31a
31b
D
S
G
26'',GND
36'
38'
CONTROL CIRCUIT AND LOGIC
60,70
26'
27
28
LOAD
V_{OUT}

FIG. 12

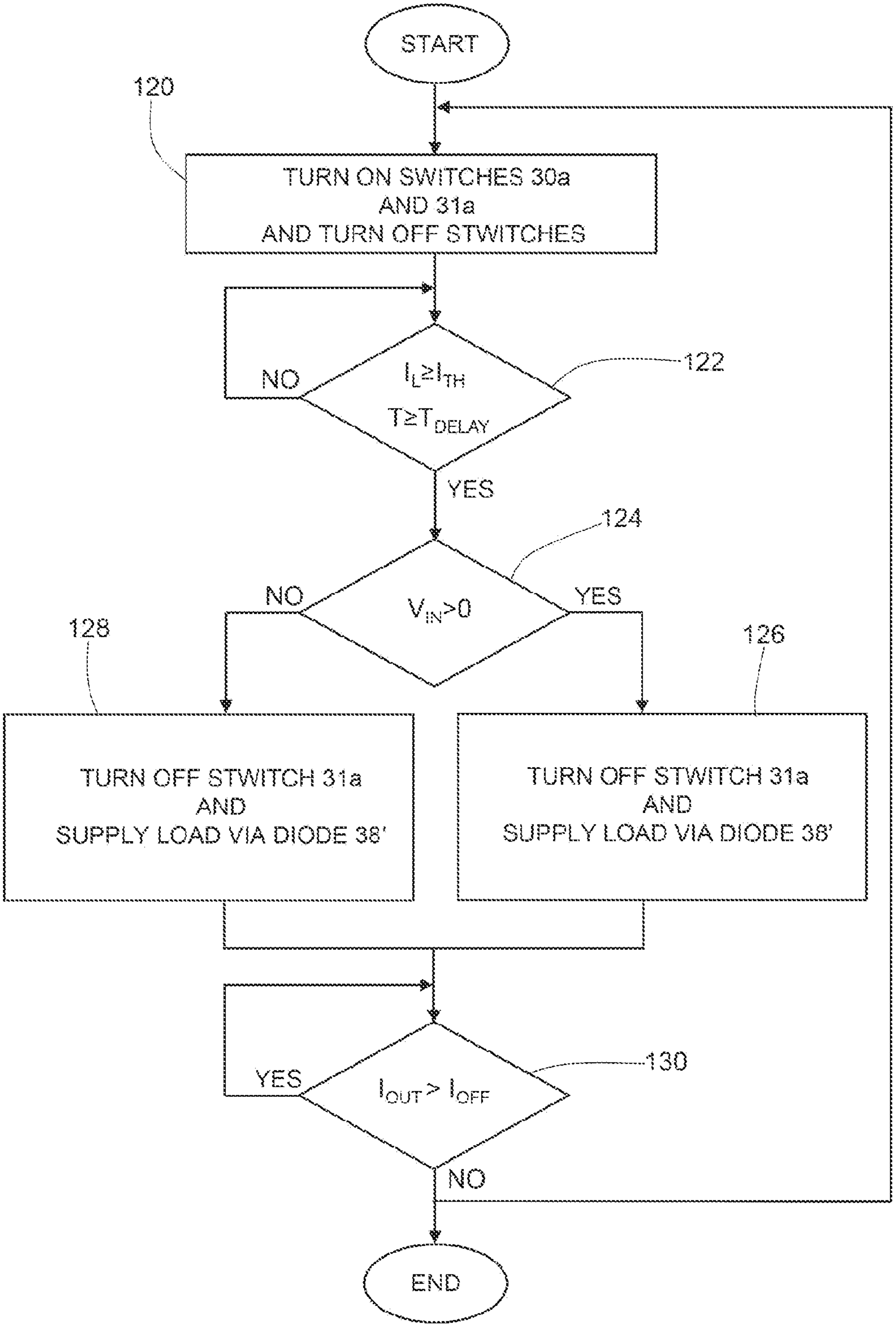
START
120
TURN ON SWITCHES 30a
AND 31a
AND TURN OFF STWITCHES
122
NO
$I_L \geq I_{TH}$
$T \geq T_{DELAY}$
YES
124
NO
$V_{IN} > 0$
YES
128
126
TURN OFF STWITCH 31a
AND
SUPPLY LOAD VIA DIODE 38'
TURN OFF STWITCH 31a
AND
SUPPLY LOAD VIA DIODE 38'
130
YES
$I_{OUT} > I_{OFF}$
NO
END

FIG. 13

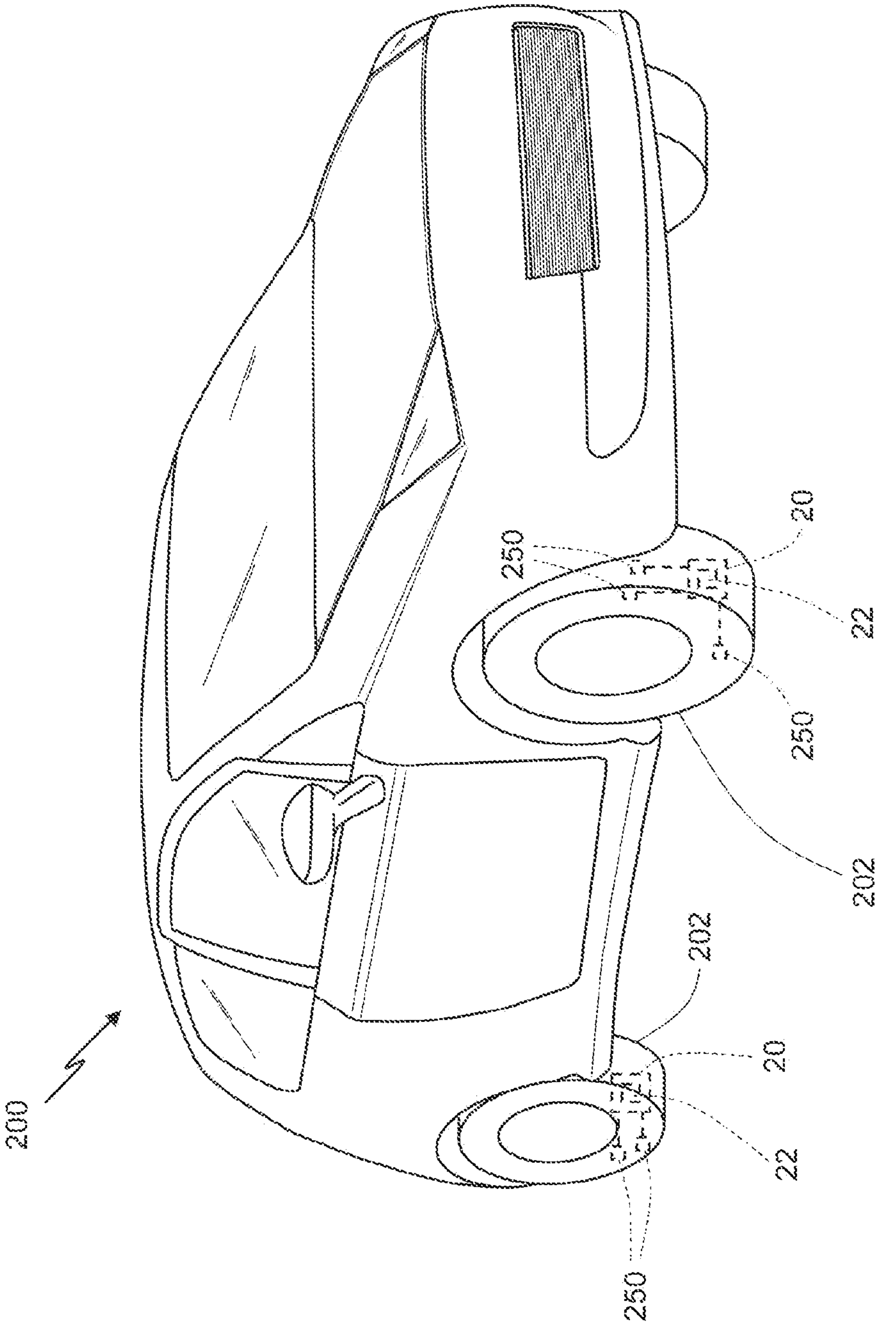
200
202
202
20
22
250
250
20
22
250

FIG. 14

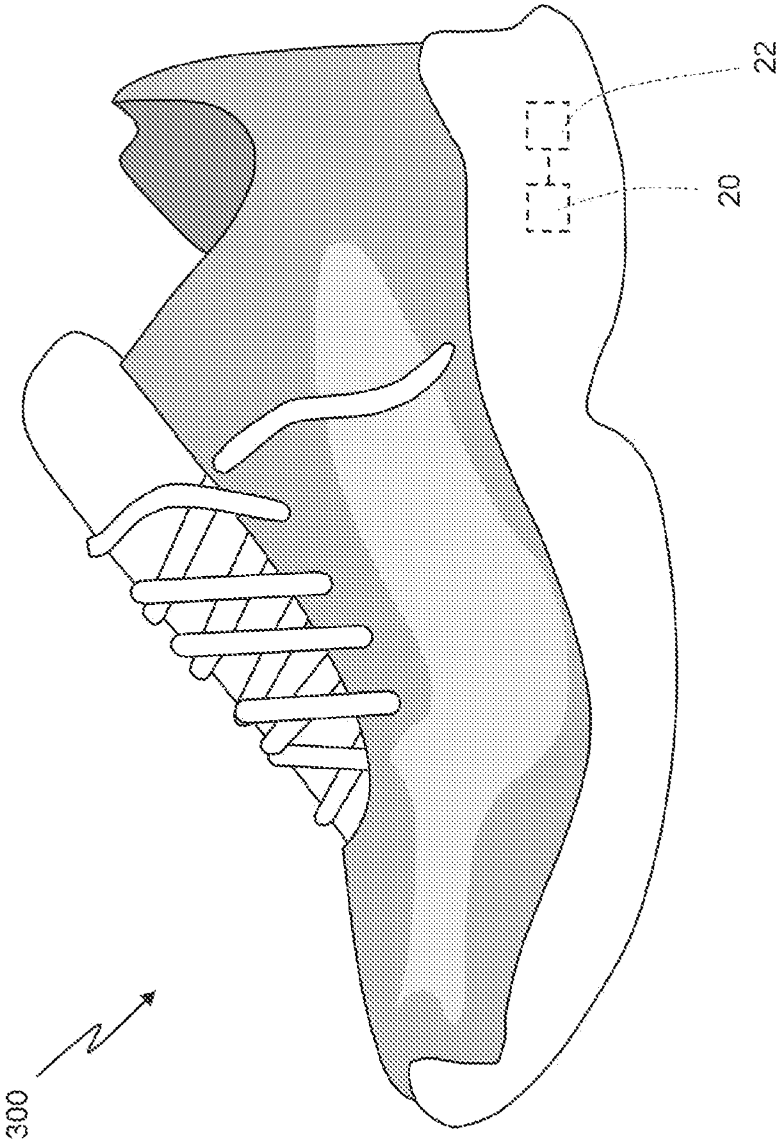

FIG. 15

ENERGY HARVESTING INTERFACE WITH IMPROVED IMPEDANCE MATCHING, METHOD FOR OPERATING THE ENERGY HARVESTING INTERFACE, AND ENERGY HARVESTING SYSTEM USING THE ENERGY HARVESTING INTERFACE

PRIORITY CLAIM

[0001] This application claims priority to Italian Application for Patent No. TO2014A000128 filed Feb. 14, 2014, the disclosure of which is incorporated by reference.

TECHNICAL FIELD

[0002] The present invention relates to an energy harvesting interface with improved impedance matching, a method for operating the energy harvesting interface, and an energy harvesting system comprising the energy harvesting interface.

BACKGROUND

[0003] As is known, systems for harvesting energy (also known as "energy-scavenging systems") from intermittent environmental-energy sources (i.e., ones that supply energy in an irregular way) have aroused and continue to arouse considerable interest in a wide range of technological fields. Typically, energy harvesting systems are configured to harvest, store, and transfer energy generated by mechanical or thermal sources to a generic load of an electrical type.

[0004] Low-frequency vibrations, such as, for example, mechanical vibrations of disturbance in systems with moving parts may be a valid source of energy. The mechanical energy is converted, by one or more appropriate transducers (for example, piezoelectric or electromagnetic devices) into electrical energy, which may be used for supplying an electrical load. In this way, the electrical load does not require batteries or other supply systems, which are cumbersome and present a low resistance in regard to mechanical stresses.

[0005] FIG. **1** is a schematic illustration in the form of functional blocks of an energy harvesting system of a known type.

[0006] FIG. **2** shows, according to a simplified circuit representation, the energy harvesting system of FIG. **1**.

[0007] The energy harvesting system of FIG. **1** comprises: a transducer **2**, for example of an electromagnetic or piezoelectric type, subject in use to environmental mechanical vibrations and configured to convert mechanical energy into electrical energy, typically into AC (alternating current) voltages; a scavenging interface **4**, for example comprising a diode-bridge rectifier circuit (also known as Graetz bridge), configured to receive at input the AC signal generated by the transducer **2** and supplying at output a DC (direct current) signal for charging a capacitor **5** connected to the output of the rectifier circuit **4**; and a DC-DC converter **6**, connected to the capacitor **5** for receiving at input the electrical energy stored by the capacitor **5** and supplying it to an electrical load **8**. Thus, the capacitor **5** has the function of an element for storing energy, which is made available, when required, to the electrical load **8** for operation of the latter.

[0008] This type of interface, which operates as a peak detector, presents some drawbacks. The efficiency of the system **1** of FIG. **1** is markedly dependent upon the signal generated by the transducer **2**. In the absence of the DC-DC converter **6**, the efficiency rapidly drops to zero (i.e., the system **1** is unable to harvest environmental energy) when the amplitude of the signal of the transducer **2** (signal V_{TRANSD}) assumes a value lower, in absolute value, than $V_{OUT}+2V_{TH\_D}$ where V_{OUT} is the voltage across the capacitor **5**, and V_{TH_D} is the threshold voltage of the diodes that form the energy harvesting interface **4**. As a consequence of this, the maximum energy that may be stored in the capacitor **5** is limited to the value $E_{max}=0.5\cdot C_{OUT}\cdot(V_{TRANSD}{}^{MAX}-2V_{TH\_D})^2$. If the amplitude of the signal V_{TRANSD} of the transducer **2** is lower than twice the threshold voltage V_{TH_D} of the diodes of the rectifier of the energy harvesting interface **4** (i.e., $V_{TRANSD}<2V_{TH\_D}$), then the efficiency of the system **1** is zero, the voltage across the output capacitor **5** is zero, the environmental energy is not harvested, and the electrical load **8** is not supplied.

[0009] When the DC-DC converter **6** (of a boost type) is set between the output capacitor **5** and the electrical load **8**, it is possible to make up for the drop in efficiency. However, in this situation, the current supplied by the transducer and rectified by the diode bridge is not regulated and is not actively controlled. Consequently, the impedance R_{LOAD} represented schematically in FIG. **2** cannot be matched to the series impedance R_S of the transducer **2**. This in any case causes a global loss of efficiency of the system **1**.

[0010] A further solution, which enables active control of the current supplied by the transducer **2**, envisages use of an AC-DC converter. This solution, for example proposed by IEEE TRANSACTIONS ON POWER ELECTRONICS, Vol. 25, No. 8, August 2010, pp. 2188-2199 (incorporated by reference), envisages the use of a closed-loop boost converter that exploits directly the series inductance of the transducer and generates a regulated voltage that charges the output capacitor. It is thus possible to supply the electrical load **8** directly by the output capacitor, without the aid of a DC-DC converter **6** of the type illustrated in FIG. **1**. The control loop enables the voltage on the output capacitor for being kept substantially constant. However, this solution presents some disadvantages. For instance, if the load requires a supply power that exceeds the maximum power that the output capacitor may supply, the regulated output voltage drops to zero in a substantially immediate way. Furthermore, in this condition, the AC-DC converter is unable to make up immediately for the voltage drop on the output capacitor, harvesting further energy for supply of the electrical load. The energy harvesting efficiency is thus jeopardized.

SUMMARY

[0011] Embodiments provide an energy harvesting interface, a method for operating the energy harvesting interface, a system for harvesting environmental energy comprising the energy harvesting interface, and an apparatus comprising the environmental-energy harvesting system that will enable the aforementioned problems and disadvantages for being overcome.

[0012] The energy harvesting interface (in particular, having the configuration of a rectifier circuit) may be connected between an input-signal source (in particular, a variable voltage signal) and an electrical load (with the optional interposition of a DC-DC converter configured to supply to the electrical load a voltage signal having a voltage level accepted by the electrical load). The energy harvesting interface comprises, according to one embodiment, a first switch and a second switch, set in series with respect to one another, connected between an input terminal of the interface and a reference terminal of the interface, set at constant voltage (e.g., ground voltage, in particular 0 V). The interface further comprises a third switch and a fourth switch, set in series with respect to one another, connected between an input terminal of the interface and the reference terminal of the interface on which the energy is harvested.

[0013] The energy harvesting interface further comprises a control logic, coupled to the control terminals of the first and second switches, configured to open/close the first and second switches by an appropriate control signal.

[0014] The energy harvesting interface further comprises, as has been said, additional, third and fourth, switches, each having a control terminal. In use, for a polarity of the transduced input signal, the third and fourth switches are kept closed and are used during steps of detection of the current that flows through them, by the control logic. These detection steps define passage from a condition of storage of energy harvested by the transducer (exploiting the inductor integrated in the transducer itself) to a condition of transfer of said energy at output, for example to a storage capacitor and/or to an electrical load.

[0015] The storage capacitor is coupled to the output of the energy harvesting interface, for example, via diodes or controlled switches. The electrical load may be coupled in parallel to the storage capacitor, said electrical load being supplied by the energy stored in the capacitor.

[0016] As has already been said, a DC-DC converter, of buck, or boost, or buck/boost type may be optionally set between the capacitor and the electrical load.

[0017] In a first operating condition, and for a first polarity of the transduced voltage, the first and second switches are closed and the energy harvesting interface stores electrical energy; the diode guarantees that during this operating condition the energy will not flow to the storage capacitor.

[0018] In a second operating condition, and for the first polarity of the transduced voltage, the first switch is opened and the second switch is kept closed; the capacitor is charged by the electrical energy previously stored during the first operating condition and transferred through the diode.

[0019] In the aforementioned first and second operating conditions, the third and fourth switches are kept closed (i.e., ON).

[0020] Passage from the first operating condition to the second operating condition, and vice versa, is cyclic. When the transduced voltage has a second polarity opposite to the first polarity (e.g., the first polarity is positive, and the second polarity is negative), the above operations are carried out in a similar way by appropriately controlling the third and fourth switches and keeping the first and second switches closed (ON).

[0021] The temporal duration of the first and second operating conditions is calculated by appropriate blocks for detecting the current that flows between the inputs of the energy harvesting interface and the reference terminal. These values are then supplied to the control logic that controls the switches according to one aspect.

[0022] The energy harvesting interface is described in detail with reference to an application thereof, in particular as rectifier circuit of an energy harvesting system set between a voltage source and a storage element and/or an electrical load.

BRIEF DESCRIPTION OF THE DRAWINGS

[0023] For a better understanding of the present invention, preferred embodiments thereof are now described, purely by way of non-limiting example and with reference to the attached drawings, wherein:

[0024] FIG. **1** shows an energy harvesting system according to a known embodiment;

[0025] FIG. **2** shows a circuit equivalent to the energy harvesting system of FIG. **1**;

[0026] FIGS. **3**A and **3**B show an energy harvesting system comprising a circuit of an energy harvesting interface that may operate according to the steps of the method of FIG. **11**, according to respective embodiments;

[0027] FIGS. **4**A and **4**B show the energy harvesting system of FIG. **3**A or **3**B in respective operating conditions, following one another in time;

[0028] FIGS. **5**A-**5**C show, using the same time scale, the time plot of current signals of the energy harvesting system of FIG. **3**A or FIG. **3**B in the operating conditions of FIGS. **4**A and **4**B;

[0029] FIG. **6** shows the plot of the coupling factor between the transducer and the harvesting-interface circuit of FIG. **3**A or FIG. **3**B, as operating parameters vary;

[0030] FIG. **7** shows the energy harvesting system of FIG. **3**A or FIG. **3**B in greater detail, according to one embodiment;

[0031] FIG. **8** shows a current detector used in the energy harvesting system of FIG. **7**;

[0032] FIG. **9** shows a further current detector used in the energy harvesting system of FIG. **7**;

[0033] FIG. **10** shows the energy harvesting system of FIG. **3**A or FIG. **3**B in greater detail, according to a further embodiment;

[0034] FIG. **11** shows, by a flowchart, steps of a method for operating the energy harvesting system of FIG. **3**A, **3**B, **7** or **10**;

[0035] FIG. **12** shows an energy harvesting system according to a further embodiment;

[0036] FIG. **13** shows, by a flowchart, steps of a method for operating the energy harvesting system of FIG. **12**;

[0037] FIG. **14** shows a vehicle comprising the energy harvesting system according to any one of the embodiments; and

[0038] FIG. **15** shows an item of footwear comprising the energy harvesting system according to any one of the embodiments.

DETAILED DESCRIPTION OF THE DRAWINGS

[0039] FIG. **3**A shows an energy harvesting system **20** comprising a rectifier circuit **24**, according to one embodiment.

[0040] In general, the energy harvesting system **20** comprises: a transducer **22** (similar to the transducer **2** of FIG. **1**) including its own output terminals **22'**, **22"**; the rectifier circuit **24**, including a first input terminal **25'** and a second input terminal **25"**, electrically coupled, respectively, to the output terminals **22'**, **22"** of the transducer **22**, a first output terminal **26'**, and a reference terminal **26"**; and a storage element **27**, for example a capacitor, connected between the first output terminal **26'** and the reference terminal **26"** of the rectifier circuit **24**, and configured to store electrical charge supplied at output from the rectifier circuit **24**. According to one embodiment, the reference terminal **26"** is a ground-voltage reference terminal GND, for example equal to approximately 0 V. Other reference voltages may be used.

[0041] The transducer **22** is, for example, an electromagnetic transducer, and is represented schematically for including a voltage generator **22***a*, configured to supply a voltage V_{TRANSD}, an inductor **22***b* (typical of the electromagnetic transducer) having an inductance L_S, and a resistor **22***c* having a resistance R_S, connected in series to the inductor **22***b*.

[0042] On the output of the rectifier circuit **24**, in parallel with the storage element **27**, there may be connected an electrical load **28**, configured for being supplied by the charge stored in the storage element **27** or by means, for example, of a DC-DC converter (not illustrated in the figures) in the case where the electrical load requires a voltage value different from the voltage generated at output from the rectifier circuit **24**.

[0043] Connected in series together between the first input terminal **25'** and the reference terminal **26"** of the rectifier circuit **24** are a high-voltage (HV) switch **30***a* and a low-voltage (LV) switch **30***b*, in particular of the voltage-controlled type. The switches **30***a* and **30***b* are, for example, N-channel field-effect transistors (FETs).

[0044] The HV switch **30***a* is a device that is able to withstand high voltages. According to one embodiment, the HV switch **30***a* is a DMOS transistor configured to operate with gate-to-drain voltages (V_{GD}) and drain-to-source voltages (V_{DS}) ranging between 30 and 50 V, for example 40 V.

[0045] In addition to DMOSs, it is likewise possible to use drift MOSs and drain extension MOSs, which are transistors that may withstand high voltages between drain and source terminals, and between gate and drain terminals. It may be noted that the range of voltages indicated is purely indicative and non-limiting. Technologies configured to withstand voltages higher than 50 V are known and under development, and may likewise be used in the context of the present circuit.

[0046] The LV switch **30***b* is a low-voltage device. According to one embodiment, the LV switch **30***b* is a CMOS transistor configured to operate with gate-to-source voltages (V_{GS}) ranging, for example, between 1 and 5 V, in particular 2.5-3.6 V, for example 3.3 V. Other technologies for low-voltage transistors envisage slightly higher operating voltages, for example in the region of 4-5 V.

[0047] It is evident that the values appearing indicate a possible embodiment, and vary according to the technology used for the transistors and to the specific application.

[0048] The first input terminal **25'** is electrically coupled to the first output terminal **26'** by a diode **36**. The diode **36** is chosen for having a low forward threshold voltage, in the region of 0.6-0.7 V, for maximizing the efficiency of the rectifier, above all in the steps where the voltage stored on the output capacitor is low.

[0049] According to an alternative embodiment (not illustrated), the diode **36** may be replaced by a MOSFET, for example, of the N-channel type. As is known, a MOSFET has an internal diode (parasitic diode). In this case, the MOSFET may be operated in an active way (by actively controlling turning-on and turning-off of the MOSFET), or in a passive way (by turning off the MOSFET and exploiting the internal parasitic diode).

[0050] Furthermore, the rectifier circuit **24** comprises a further HV switch **31***a* and a further LV switch **31***b*, connected together in series and electrically coupled between the second input terminal **25"** and the reference terminal **26"** of the rectifier circuit **24**. The switches **31***a* and **31***b* are similar to the switches **30***a* and **30***b*, and such that the HV switch **31***a* is a device that is able to withstand high gate-to-drain voltages and drain-to-source voltages (for example 30-50 V, in particular 40 V), whereas the LV switch **31***b* is a low-voltage device, for example a CMOS, which is able to withstand low gate-to-source voltages (for example, 1-5 V, in particular 2.5-3.6V, even more in particular 3.3 V). Other technologies for low-voltage transistors envisage slightly higher operating voltages, for example in the region of 4-5 V.

[0051] The second input terminal **25"** is electrically coupled to the first output terminal **26'** by a diode **38**, similar to the diode **36**. According to an alternative embodiment (not illustrated), the diode **38** may be replaced by a MOSFET, for example an N-channel MOSFET. As is known, a MOSFET has an internal diode (parasitic diode). In this case, said MOSFET may be operated in a active way (by actively controlling turning-on and turning-off of the MOSFET), or in a passive way (by turning off the MOSFET and exploiting the internal parasitic diode).

[0052] For simplicity of description, the HV switches **30***a* and **31***a* will hereinafter be referred to, respectively, as "high-voltage (HV) transistors **30***a* and **31***a*", without this implying any loss of generality, and the LV switches **30***b* and **31***b* be referred to, respectively, as "low-voltage (LV) transistors **30***b* and **31***b*", without this implying any loss of generality.

[0053] Likewise, the terms "transistor closed" or "transistor ON" will hereinafter refer to a transistor biased so to enable conduction of electric current between its source and drain terminals, i.e., configured to behave as a closed switch, and the terms "transistor open" or "transistor OFF" will hereinafter refer to a transistor biased so not to enable conduction of electric current between its source and drain terminals, i.e., configured to behave ideally as an open or inhibited switch.

[0054] FIG. **3**B shows the energy harvesting system **20** of FIG. **3**A, where the switches have been replaced by respective transistors. Each transistor is further represented with its own internal diode (parasitic diode).

[0055] With reference to FIG. **3**B, the drain terminal D of the HV transistor **30***a* is connected to the first input terminal **25'** of the rectifier circuit **24**, whereas the source terminal S of the HV transistor **30***a* is connected to the drain terminal D of the LV transistor **30***b*; the source terminal S of the LV transistor **30***b* is instead connected to the reference terminal **26"**. The diode **36** has its cathode electrically coupled to the output terminal **26'** of the rectifier circuit **24** and its anode electrically coupled to the first input terminal **25'** of the rectifier circuit **24**.

[0056] As regards the HV transistor **31***a* and LV transistor **31***b*, these are connected between the second input terminal **25"** and the reference terminal **26"** of the rectifier circuit **24** so that the source terminal S of the LV transistor **31***b* is connected to the reference terminal **26"**, the drain terminal D of the HV transistor **31***a* is connected to the second input terminal **25"**, and the remaining drain terminal D of the LV transistor **31***b* and source terminal S of the HV transistor **31***a* are connected together.

[0057] The diode **38** has its cathode electrically coupled to the output terminal **26'** of the rectifier circuit **24** and its anode electrically coupled to the second input terminal **25"** of the rectifier circuit **24**. During positive half-cycles of the input voltage V_{IN}, the voltage is rectified by appropriately driving the HV transistor **30***a*, keeping the transistors **30***b*, **31***a*, **31***b* in the ON state. Instead, during negative half-cycles of the input voltage V_{IN}, the voltage is rectified by appropriately driving the HV transistor **31***a*, keeping the transistors **31***b*, **30***a*, **30***b* in the ON state.

[0058] According to one embodiment, the rectifier circuit **24** further comprises a control circuit and a control logic, designated in FIG. **3**A or FIG. **3**B by the reference numbers **60** and **70**, and better described with reference to FIGS. **8** and **9**. Furthermore, the control logic **60** implements the steps of the method of FIG. **11**.

[0059] In use, for example for positive values of the voltage V_{IN}, the HV transistor **30***a* and the LV transistor **30***b* are kept ON for at least a time interval T_{DELAY} for storing energy in the inductor **22***b* (situation illustrated schematically in FIG. **4**A). During this step, storage of the energy in the inductor **22***b* is guaranteed by the fact that the diode **36** does not conduct. Furthermore, also the transistors **31***a* and **31***b* are kept in the ON state.

[0060] Then, when the time interval T_{DELAY} has elapsed and the energy stored (current flowing) in the inductor **22***b* has reached a minimum threshold value I_{TH}, the HV transistor **30***a* is turned off. A current may thus flow from the inductor **22***b*, through the diode **36**, to the storage element **27**/electrical load **28**. This situation is illustrated schematically in FIG. **4**B.

[0061] As has been said, the input signal V_{IN} is a variable signal, i.e., having a polarity that varies in time. For negative polarities of V_{IN}, what has been described with reference to FIGS. **4**A and **4**B is in any case valid by controlling in a similar way the transistors **31***a* and **31***b*. The steps of control of these transistors are not herein described for brevity, but are apparent to the person skilled in the branch on the basis of what has so far been described.

[0062] According to one embodiment, in both of the operating conditions of FIGS. **4**A and **4**B, for positive polarities of the input voltage V_{IN}, the LV transistor **30***b* is always kept closed, and the control logic **60** drives just the HV transistor **30***a* into the open/closed state. Likewise, for negative polarities of the input voltage V_{IN}, the control logic **60** drives just the HV transistor **31***a* into the open/closed state, whereas the LV transistor **31***b* is always kept closed. This situation is schematically represented in FIG. **3**B, which shows a voltage generator, configured to generate a voltage V_{DD}, coupled to the control terminals G of the LV transistors **30***b* and **31***b*. The voltage V_{DD} is chosen with a value such as to drive the LV transistors **30***b* and **31***b* into the closed state.

[0063] During the step of FIG. **4**B, where the current stored in the inductor **22***b* is transferred at output on the storage element **27** by the diode **36** (or alternatively the diode **38**, according to the polarity of the input voltage V_{IN}), an increase in the output voltage V_{OUT} is noted.

[0064] Hereinafter, operation of the rectifier **24** is described more fully with reference to a circuit model valid for one polarity (in particular, the positive polarity) of the input signal V_{IN}, for greater simplicity and clarity of description. As has been said, what has been described may in any case be readily applicable to control of the transistors **31***a* and **31***b*, in the case of negative polarity of the input signal V_{IN}.

[0065] FIG. **4**A shows a circuit equivalent to the circuit of FIG. **3**A or FIG. **3**B, for positive half-waves of the input voltage V_{IN}. The diode **36**, in this condition, does not conduct. The transistors **30***a* and **30***b* are ON. In this operating condition, the transistors **30***a* and **30***b* are ideally replaced by respective resistors which have a resistance R^{HV}_{ON} and, respectively, R^{LV}_{ON} (ON-state resistance of the transistors **30***a* and **30***b*).

[0066] The current I_L that flows in the inductor **22***b* is equal to the current I_{ON} that traverses the transistors **30***a* and **30***b* in the ON state. The value of the current I_L increases until it reaches a maximum, or peak, value I_p (see the graph of FIG. **5**A).

[0067] The current I_{ON} reaches the peak value I_p at time $t=t_c=T_{DELAY}$. For simplicity, it is assumed that the starting instant t_0 is 0 μs.

[0068] Once the time interval T_{DELAY} has elapsed and assuming that the current I_L that flows in the inductor **22***b* has reached a value equal to, or higher than, the threshold value I_{TH}, the operating condition represented schematically in FIG. **4**B is reached.

[0069] The time interval T_{DELAY} is the interval between the instant of closing of the HV transistor **30***a* (at time t_0) and the instant of opening of the HV transistor **30***a* (at time t_c).

[0070] The value of threshold current I_{TH} is chosen on the basis of the values of the maximum or of the minimum short-circuit current that the transducer supplies. It is the ratio between the maximum or minimum voltage V_{TRANSD} and the series resistance R_S, which depends on the application of the rectifier circuit **24**.

[0071] These values depend upon the characteristics of the transducer **22** and upon the environmental stresses to which the transducer **22** is subjected. In particular, the value of threshold current I_{TH} is chosen much lower than the peak value I_p that is expected for being reached in the application in which the rectifier circuit **24** is used. For instance, assuming that peak values I_p of approximately 150 mA are reached, the threshold I_{TH} may be chosen between approximately 5-10 mA. It is for being noted that the choice of a threshold current I_{TH} too close to the peak value I_p entails a low efficiency. In fact, according to what has been described, current is transferred at output only when the threshold I_{TH} is exceeded; all the portions of the signal V_{TRANSD} that generate a current with peak value $I_p<I_{TH}$ do not give a contribution of charge transferred at output.

[0072] With reference to FIG. **4**B, at time t_c, the HV transistor **30***a* is opened, and the diode **36** starts to conduct. The current I_L that flows from the inductor **22***b* to the output **26'** of the rectifier **24** is the current I_{OUT} that charges the capacitor **27**. In this step, the current in the inductor **22***b* decreases with a constant slope, until it reaches the predefined value I_{OFF} (at time t_{max}, see again FIG. **5**A).

[0073] I_{OFF} is a constant value, given by I_p/K, where K is a constant higher than 1 (chosen as explained hereinafter). FIG. **5***a* shows the plot of the current I_L at time t (in microseconds). The curve of the current I_L reaches the peak value I_p at the instant t_c, where the HV transistor **30***a* is opened (see FIG. **5**B).

[0074] Then, between t_c and t_{max} (time interval T_{CHARGE}) the current I_L decreases until it reaches the value $I_{OFF}=I_p/K$.

[0075] FIG. **5**B shows, using the same time scale as that of FIG. **5**A, the plot of the current I_{ON} that flows through the HV transistor **30***a* during the step of FIG. **4**A of charging of the inductor **22***b*. In the time interval t_0-t_c, the current I_{ON} follows the same plot as that of the current I_L; at the instant t_c, the HV transistor **30***a* is opened and consequently the current I_{ON} drops to zero.

[0076] FIG. **5**C shows, using the same time scale as that of FIGS. **5**A and **5**B, the plot of the output current I_{OUT}. The current I_{OUT} remains at a zero value in the time interval t_0–t_c to reach the peak value I_p at the instant t_c in which the capacitor **27** is electrically coupled to the inductor **22***b*. Then, between t_c and t_{max} (within the time interval T_{CHARGE}), the energy stored in the inductor **22***b* supplies and charges the capacitor **27**.

[0077] At time t_{max}, when the current that flows to the capacitor **27** reaches the threshold value I_{OFF}, the HV transistor **30***a* is closed so that the inductor **22***b* charges again, as has already been described. The steps of charge and discharge of the inductor **22***b* (and, consequently, of supply of the capacitor **27**/load **28**) repeat in a cyclic way.

[0078] The integral of the curve of I_{OUT} (FIG. **5**C) between time t_c and time t_{max} indicates the charge Q_{CYCLE} transferred between the input and the output of the rectifier **24** in the time interval T_{CHARGE}. In order to maximize the efficiency of transfer of charge between the input and the output of the rectifier **24**, it is expedient to maximize the value of the power P_{CYCLE} transferred at output from the rectifier circuit **24** in each cycle of charge/discharge of the inductor **22***b*. The power P_{CYCLE} is defined as $P_{CYCLE}=V_{OUT}\cdot I_{CYCLE}$, where I_{CYCLE} is given by $I_{CYCLE}=Q_{CYCLE}/T_{CYCLE}$, where T_{CYCLE} is the time interval between t_0 and t_{max} ($T_{CYCLE}=T_{DELAY}+T_{CHARGE}$).

[0079] It is noted that P_{CYCLE} is given by the following relation (where I_{ON} assumes the peak value I_p)

$$P_{CYCLE} = \frac{\frac{I_{ON} + I_{OFF}}{2} \cdot T_{CHARGE}}{T_{DELAY} + T_{CHARGE}} \cdot V_{OUT}$$

[0080] From the foregoing relation it may be noted how the power P_{CYCLE} is a function of the design parameters T_{DELAY} and K, and of the external variables V_{TRANSD} (voltage of the transducer, not predictable) and V_{OUT} (voltage across the capacitor **27**, which is not predictable either). Maximizing the value of P_{CYCLE} thus means finding the optimal values of T_{DELAY} and K such that the curve of P_{CYCLE} reaches a maximum value, or a value close to the maximum, or an optimal value definable according to the particular application and design requirements.

[0081] The curve of P_{CYCLE} reaches an optimal value when the output of the transducer **22** and the input of the rectifier circuit **24** show the same impedance (they are, that is, matched). The best coupling efficiency η_{COUPLE} between the transducer **22** and the rectifier circuit **24** is given by $P_{CYCLE}^{OPT}/P_{TRANSD}^{MAX}$, where P_{CYCLE}^{OPT} is the value of P_{CYCLE} calculated with optimal values of T_{DELAY} and K, and P_{TRANSD}^{MAX} is given by $(V_{TRANSD})^2/4R_S$.

[0082] Optimization of the value of P_{CYCLE} enables an optimal value of the time interval T_{DELAY} and of the factor K for being obtained (and vice versa) as a function of the value of V_{TRANSD} and V_{OUT}. However, the dependence of T_{DELAY} upon V_{TRANSD} and V_{OUT} is irrelevant for practical purposes, and the value of coupling efficiency η_{COUPLE} reaches values higher than 95% for values of V_{TRANSD} and V_{OUT} of practical interest.

[0083] FIG. **6** shows the plot of the coupling efficiency η_{COUPLE} as the values T_{DELAY} and K vary. The graph of FIG. **6** may be easily obtained from the expression of P_{CYCLE} by varying the parameters T_{DELAY} and K (fixing the values of the external variables V_{TRANSD} and V_{OUT}). To each value of η_{COUPLE} there corresponds a pair of values T_{DELAY} and K. It is thus possible to obtain automatically the pair of optimal values T_{DELAY} and K to obtain a desired value of coupling efficiency η_{COUPLE}. In the graph of FIG. **6**, the darker areas are the ones in which the value of coupling efficiency η_{COUPLE} is higher; instead, the lighter areas are the ones in which the value of coupling efficiency η_{COUPLE} is lower (low values of T_{DELAY} and high values of K, or high values of T_{DELAY} and low values of K).

[0084] In the specific case, a good compromise for the choice of the values of T_{DELAY} and K, in order to have contained consumption levels and good coupling efficiency, is obtained by choosing $T_{DELAY}=40$ μs and K=1.75. It is, however, evident that the choice of the values of T_{DELAY} and K depends upon the field of application, and these values may thus be chosen freely according to the need (in general with K≥1).

[0085] FIG. **7** shows, by functional blocks, a control circuit **70** for driving the HV transistor **30***a* in order to implement the operating conditions of FIGS. **4***a* and **4***b*. The control circuit **70** operates, in particular, for positive half-waves (V_{IN}^+) of the input signal V_{IN}. The LV transistors **30***b* and **31***b* are biased at constant voltage V_{DD}, for being kept always in the ON state. The value of the voltage V_{DD} is thus chosen on the basis of the characteristics of these transistors, with the purpose of driving them into the ON state.

[0086] In order to drive the HV transistor **31***a* for negative half-waves of the input signal V_{IN}, a circuit architecture is used similar to the one illustrated for the control circuit **70** (see, for example, FIG. **10**).

[0087] In greater detail, the control circuit **70** comprises a first current detector **72**, coupled between the source terminal S and the drain terminal D of the LV transistor **30***b*, for detecting (during the step of FIG. **4**A) the instant in which the current I_{ON} that flows through the LV transistor **30***b* (and, consequently, also through the HV transistor **30***a*) exceeds the threshold I_{TH}. When this condition is detected, and the interval T_{DELAY} has elapsed, the control logic **60** drives the HV transistor **30***a* into inhibition, thus controlling passage to the operating condition of FIG. **4**B. In addition, the current detector **72** participates in generation, in the step of FIG. **4**A, of a scaled copy I_{ON}/K of the current that flows in the LV transistor **30***b*, as illustrated more clearly hereinafter.

[0088] FIG. **8** shows in greater detail the first current detector **72**, according to one embodiment. With reference to FIG. **8**, a first portion of the current detector **72** comprises a comparator **86** configured to receive at input, on its non-inverting terminal, the voltage signal present on the source terminal S of the HV transistor **30***a* (or likewise on the drain terminal of the LV transistor **30***b*) and, on its inverting terminal, a voltage signal V_{TH} identifying the threshold I_{TH} (by an appropriate voltage-to-current conversion, in itself obvious). The comparator **86** generates at output a digital signal V_{OUT_TH}, which assumes the low logic level "0" when $I_{ON}<I_{TH}$ and the high logic level "1" when $I_{ON}\geq I_{TH}$ (or vice versa). In greater detail, the comparator **86** is configured to receive, at input on its non-inverting terminal, the voltage signal present on the source terminal of the HV transistor **30***a* (signal V_{IN}^+), and, at input on its inverting terminal, a threshold-voltage signal V_{TH} such that $V_{TH}=I_{TH}\cdot(R^{HV}_{ON}+R^{LV}_{ON})$ where, as has already been said, R^{HV}_{ON} is the ON-state resistance of the HV transistor **30***a*, and R^{LV}_{ON} is the ON-state resistance of the LV transistor **30***b*. When the voltage on the source terminal of the HV transistor **30***a* exceeds the threshold V_{TH}, the output of the comparator **86** changes state to signal that the threshold has been exceeded (and thus to indicate that $I_L=I_{ON}\geq I_{TH}$).

[0089] The signal at output from the comparator **86** is supplied to the control logic **60**. The control logic **60** monitors the duration of the time interval T_{DELAY} and, when the time interval T_{DELAY} has elapsed, turns off the HV transistor **30***a*.

[0090] Passage of the time interval T_{DELAY} may alternatively be monitored by the comparator **86**. In this case, the signal at output from the comparator **86** assumes a high logic level when $I_{ON} \geq I_{TH}$ and $t \geq T_{DELAY}$, and the control logic **60** turns off the HV transistor **30***a* at the rising edge of the digital signal generated by the comparator **86**.

[0091] A second portion of the current detector **72** comprises a negative feedback loop including a comparator **89**, which controls the current that flows on an output branch **90** of the current detector **72**, by acting on the control terminal of a transistor **91** belonging to the output branch **90**. The output branch **90** further comprises an additional transistor **92**, connected in series to the transistor **91**. Note that the transistor **92** is a low-voltage transistor, for example a CMOS. In particular, the transistor **92** is configured to operate with gate-to-source voltages in the range 1-5 V, in particular 2.5 V-3.6 V, for example at 3.3 V. Other low-voltage-transistor technologies envisage slightly higher operating voltages, for example in the region of 4-5 V.

[0092] In particular, the transistor **92** is of the same type as the LV transistor **30***b*, but is sized so that it has dimensions (measured in terms of width-to-length aspect ratio W/L) F times lower than the LV transistor **30***b* and is configured to conduct a current F times lower than the value assumed by I_{ON} (current that flows through the LV transistor **30***b*). The LV transistor **30***b* and the transistor **92** further have their respective control terminals connected together and biased at the voltage V_{DD}.

[0093] The negative feedback loop of the current detector **72** controls the gate voltage of the transistor **91** so that the drain voltage of the transistor **92** will be equal to the voltage across the capacitor **88**. In use, current always flows in the output branch **90**. In the step of FIG. **4**A, the current is variable and equal to I_{ON}/F, whereas in the step of FIG. **4**B the current is constant and equal to I_p/F. Sizing of the transistor **92** guarantees that the current that flows in the output branch **90** is a fraction **1**/F of the current I_{ON} (or of its peak value I_p, as has been said).

[0094] The negative feedback, obtained by the comparator **89** and the transistor **91**, ensures that the drain voltages of the transistors **30***b* and **92** will be identical. Consequently, the current that flows through the transistor **92** assumes values equal to the value of I_{ON} scaled by the factor F (when I_{ON} reaches the peak value I_p, this current will be equal to I_p/F). There is thus generated a scaled copy of the factor F of the peak current I_p. Since both of the transistors **30***b* and **92** are low-voltage transistors (e.g., CMOSs) they provide excellent matching properties so that the factor F is minimally affected by problems of mismatch between the transistors **30***b* and **92** (as, instead, would be the case, where the transistors **30***b* and **92** were high-voltage transistors). This enables a scaled copy of the peak current I_p for being obtained that is stable and with reproducible value.

[0095] The negative feedback provided by the comparator **89** ensures that the signal at input to the non-inverting terminal of the comparator **89** and the signal at input to the inverting terminal of the comparator **89** are equal so that the LV transistor **30***b* and the transistor **92** have the same source-to-drain and drain-to-gate voltages.

[0096] A current mirror **90'**, made in a per se known manner, is used for repeating the current I_{ON}/F that flows in the branch **90** and supplying it at output from the current detector **72**.

[0097] The first current detector **72** further comprises a transistor **87** having a drain terminal common to the source terminal of the HV transistor **30***a*, and its source terminal coupled to a capacitor **88** (the second terminal of the capacitor **88** is connected to the reference voltage GND). The control terminal G of the transistor **87** is connected to the control terminal G of the HV transistor **30***a* and to a biasing terminal at the voltage V_{GATE_LS}. In this way, the HV transistor **30***a* and the transistor **87** are driven into the ON/OFF state at the same time, using the same signal V_{GATE_LS} (in particular, with reference to FIG. **7**, the signal generated at output from a first driving device **76**).

[0098] During the time interval T_{DELAY} (situation of FIG. **4**A), the HV transistor **30***a* is ON (the signal V_{GATE_LS} has a value such as to drive the HV transistor **30***a* into the ON state). In the same way, also the transistor **87** is ON. The capacitor **88** is consequently charged at the voltage present on the first input terminal **25'** of the rectifier circuit **24** (in the figure, the voltage across the capacitor **88** is designated by V_{C_SAMPLE}).

[0099] The comparator **89** is connected to the source terminal of the transistor **87** and, when the transistor **87** is ON, it receives at input (on its non-inverting terminal) the voltage of the drain terminal of the LV transistor **30***b*, and at input (on its inverting terminal) the signal present on the drain terminal of the transistor **92** and on the source terminal of the transistor **91**; the output of the comparator **89** is coupled to the control terminal G of the transistor **91**. The transistor **91** is always ON; the comparator **89** biases the control terminal of the transistor **91** so that the current I_{ON}/F flows in the branch **90**, as is desired.

[0100] When the HV transistor **30***a* is OFF, also the transistor **87** is OFF, and the capacitor **88** is in the floating state, ensuring, during the time interval T_{CHARGE}, a current having a practically constant value through the transistor **92** and equal to I_p/F.

[0101] In fact, during the step of supply of the capacitor **27**/load **28**, the capacitor **88** ensures maintenance of the voltage V_{C_SAMPLE} across it, guaranteeing a substantially constant input signal (but for the losses of the capacitor **88**) on the non-inverting input of the comparator **89**. This enables generation of the current I_{ON}/F for being kept unaltered on the output branch **90** of the first current detector **72** during the step of FIG. **4**B (in this step, the current I_{ON} has reached the peak value I_p, and consequently a current I_p/F flows in the output branch **90** of the first current detector **72**).

[0102] To return to FIG. **7**, the control circuit **70** further comprises a second current detector **74**, coupled between the source terminal S and the drain terminal D of the LV transistor **31***b*.

[0103] The second current detector **74** is configured to detect the value of current I_{OUT} that flows through the LV transistor **31***b* during the operating step of FIG. **4**B, i.e., following upon charging of the capacitor **27**. In particular, the second current detector **74** co-operates with the control logic **60** in order to detect whether the current $I_L = I_{OUT}$ through the LV transistor **31***b* reaches the minimum value envisaged $I_{OFF} = I_p/K$. The output signal of the second current detector **74**, which indicates the value of current through the LV transistor **31***b*, is supplied at input to the control logic **60**.

[0104] The second current detector **74** receives at input the current I_{ON}/F (generated by the first current detector **72**, as described previously), and switches when the current through the LV transistor **31***b* reaches the minimum value envisaged, given by $I_{OFF}=I_p/K$.

[0105] The control circuit **70** further comprises the first driving device **76** and a second driving device **78**, which are coupled, respectively, between the control logic **60** and the control terminal G of the HV transistor **30***a* and between the control logic **60** and the control terminal G of the HV transistor **31***a*. The first driving device **76** and the second driving device **78** are, in themselves, of a known type, and are configured to drive the transistors **30***a*, **31***a* into the opening/closing condition on the basis of a respective control signal received from the control logic **60**. In particular, in the operating condition of FIG. **4**A (positive half-wave of the transduced signal V_{TRANSD}), the control logic **60** drives, via the first driving device **76**, the HV transistor **30***a* into the ON state and, via the second driving device **78**, the HV transistor **31***a* into the ON state. On the basis of the signal generated at output from the first current detector **72**, the control logic **60** detects whether the current $I_L=I_{ON}$ has reached (and/or exceeded) the threshold value I_{TH} and whether the time T_{DELAY} has elapsed: if so, the control logic **60** drives, via the first driving device **76**, the HV transistor **30***a* into the OFF state and, via the second driving device **78**, maintains the HV transistor **31***a* in the ON state. Then, the control logic **60** monitors, on the basis of the signal received from the second current detector **74**, the value of the current $I_L=I_{OUT}$ through the LV transistor **31***b* to control passage from the operating condition of supply of the capacitor **27**/load **28** (FIG. **4**B) to the operating condition of storage of energy in the inductor **22***b* (FIG. **4***a*).

[0106] FIG. **9** shows the second current detector **74** in greater detail.

[0107] The current detector **74** is electrically coupled to a node X set between the drain terminal D of the LV transistor **31***b* and the source terminal S of the HV transistor **31***a*, for receiving at input an intermediate voltage signal V_X present on said node X. The current detector **74** includes a coupling transistor **85**, of an N type, having its drain terminal coupled to the node X and its gate terminal biased at voltage V_{DD}. As in the case mentioned previously, the voltage V_{DD} is chosen with a value such as to drive into the ON state the coupling transistor **85**, which thus remains always ON during the operating steps of the energy harvesting system.

[0108] The current detector **74** further includes a comparator **84**, having an inverting input electrically coupled to the reference voltage GND and a non-inverting input electrically coupled to the source terminal of the coupling transistor **85**. In other words, the non-inverting input of the comparator **84** and the source terminal of the coupling transistor **85** are coupled to the same node Y, having a voltage V_Y. The node Y further receives the current signal I_{ON}/F generated at output by the first current detector **72**.

[0109] The comparator **84** generates at output a signal $V_{OUTCOMP}$ that is configured to assume alternatively a high logic value "1" and a low logic value "0" according to the value assumed by the signal V_Y.

[0110] The coupling transistor **85** presents in use an internal resistance (channel resistance, or ON-state resistance) $R_{DMY}=G\cdot R_{LS}$, where R_{LS} is the internal resistance (channel resistance, or ON-state resistance) of the LV transistor **31***b*. In other words, the resistance R_{DMY} is chosen equal to a multiple G of R_{LS}.

[0111] In use, it is found that the voltage V_X on the node X is given by:

$$V_X=-I_L\cdot R_{LS}$$

[0112] and the voltage V_Y on the node Y is given by

$$V_Y=V_X+G\cdot R_{LS}\cdot I_{ON}/F$$

Thus,

$$V_Y=-I_L\cdot R_{LS}+G/F\cdot I_{ON}\cdot R_{LS}$$

[0113] The output $V_{OUTCOMP}$ of the comparator **84** changes its logic value when the voltage V_Y reaches the threshold defined by the reference voltage GND, in this example chosen equal to 0 V. We obtain that the output of the comparator **84** changes its logic value when $V_Y=0$. From this condition it follows that the output $V_{OUTCOMP}$ identifies the fact that, the value of scaled copy I_{ON}/K has been reached by the output current $I_L=I_{OUT}$. In fact, setting $V_Y=0$ in the previous equation, we have that the threshold current I_L (i.e., the threshold current I_{OUT}) is equal to $(G/F)\cdot I_{ON}$. The constant K is consequently equal to F/G. It is pointed out that, as illustrated previously, in use, the value of I_{ON} at which there occurs passage from the step of FIG. **4**A to the step of FIG. **4**B is equal to I_p. Consequently, we find that the change of logic value of the signal $V_{OUTCOMP}$ identifies that the threshold $(G/F)\cdot I_p=I_p/K$ has been reached.

[0114] With reference once again to FIG. **7**, the signal $V_{OUTCOMP}$ of the comparator **84** is received by the control logic **60**, which controls, on the basis of the value of $V_{OUTCOMP}$ received, passage from the step of FIG. **4**B to the step of FIG. **4**A. For instance, a value $V_{OUTCOMP}$="0" identifies a situation in which the current I_{OUT} has not yet reached the threshold I_{OFF}; instead, a value $V_{OUTCOMP}$="1" identifies a situation in which $V_Y=0$ and the current I_{OUT} has reached the threshold I_{OFF}.

[0115] Preferably, the transistors **31***b*, **92** and **85** are low-voltage transistors manufactured with the same technology (e.g., CMOS technology) so that they guarantee optimal matching properties such that the factor G is minimally affected by problems of mismatch between the transistors **31***b*, **92**, and **85** (as instead would be the case, where both of the transistors were high-voltage transistors). Stabilizing G around a desired value corresponds to stabilizing the values of K and F around the values chosen. The parameter K thus has a minimal spread around the chosen and desired value.

[0116] According to one embodiment, the transistors **92** (FIG. **8**) and **85** (FIG. **9**) are provided in the form of modular transistors. The transistors **92** and **85** may be obtained as a series or parallel combination of the same basic module for minimizing the mismatch between the factors F and G, and thus render the parameter K "stable".

[0117] For instance, the transistor **92** is formed by connecting, in parallel to one another, a plurality of basic modules (each module being a low-voltage MOSFET with aspect ratio W_b/L_b) so that the source terminals of each basic module are electrically connected together to a common source node, and the drain terminals of each basic module are electrically connected together to a common drain node. The gate terminals of each basic module are selectively driven into the ON state or the OFF state to form, in use, the transistor **92**, the aspect ratio W/L of which is a multiple of the aspect ratio W_b/L_b of each basic module. In this way, by turning on/turning off selectively one or more basic modules, it is possible to regulate the total amount of current carried by the transistor **92** and consequently regulate the value of the ratio 1:F between the transistor **92** and the transistor **30***b*.

[0118] A similar solution may be applied to form the transistor **85**, with variable value of G. In this case, it is possible to connect, in parallel to one another, a plurality of series of basic modules of low-voltage MOSFETs, which have the same aspect ratio (for example, W_b/L_b). Each series of basic modules presents, in use, a respective electrical resistance to the passage of the current. By selectively activating/deactivating the series of the basic modules that form the transistor **85**, it is thus possible to regulate the value of electrical resistance represented by the transistor **85** in use. The transistor **85** behaves as a resistor with a variable resistance that may be selected according to the requirement.

[0119] Consequently, the values of F and G may be chosen according to the need, as a function of the value of the parameter K that it is desired to use for the specific application.

[0120] Turning-on/turning-off of the basic modules of the transistors **85** and **92** is performed by the control logic during use. For this purpose, the control logic includes a memory **83** that stores the information regarding which and/or how many basic modules of the transistors **85** and **92** are for being turned on. If the memory **83** is of the re-writeable type, this information may be updated/modified according to the need.

[0121] The transistors **30***b*, **92** and **85** are the components via which we it is possible to control I_{TH}, the factors F and G, and consequently the factor K. By providing them in a modular form, as has been described, they may be readily configured via the memory **83** and appropriate driving devices, for enabling/disabling a certain number of basic modules thus obtaining respective equivalent transistors, which have a desired respective aspect ratio W/L.

[0122] Thus, with just one device, appropriately configured via the information stored in the integrated memory **83**, it is possible to vary freely I_{TH}, T_{DELAY} and K and thus adapt to a very wide range of transducers available on the market.

[0123] With reference to FIG. **10**, a control circuit **70'** is illustrated for driving both of the HV transistors **30***a* and **31***a* into the ON/OFF state in order to implement the operating conditions of charging of the inductor **22***b* and supply of the capacitor **27** (and/or load **28**) for positive half-waves (V_{IN}^+) and negative half-waves (V_{IN}^-) of the input signal V_{IN}.

[0124] Elements of the control circuit **70'** that are similar to elements of the control circuit **70** of FIG. **7** are designated by the same reference numbers and will not be described any further.

[0125] The control circuit **70'** comprises, in addition to what has already been described with reference to the control circuit **70** of FIG. **7**, a third current detector **72'**, and a fourth current detector **74'**.

[0126] The third current detector **72'** is similar to the first current detector **72**, and consequently is not described and illustrated any further in the figures. The third current detector **72'**, unlike the first current detector **72**, is electrically coupled between the reference terminal GND (corresponding to the source terminal of the LV transistor **31***b*) and the drain terminal of the LV transistor **31***b*, and generates at output a current signal $(I_{ON}/F)'$.

[0127] The fourth current detector **74'** is similar to the second current detector **74**, and consequently is not described and illustrated any further in the figures. The fourth current detector **74'**, unlike the second current detector **74**, is electrically coupled between the reference terminal GND (corresponding to the source terminal of the LV transistor **30***b*) and the drain terminal of the LV transistor **30***b*, and further receives at input the current signal $(I_{ON}/F)'$ generated by the third current detector **72'**.

[0128] Operation of the third and fourth current detectors is altogether similar to what has already been described with reference to the first and second current detectors **72** and **74** and consequently is immediately evident to a person skilled in the branch.

[0129] In use, when a positive half-wave of the input signal V_{IN} is detected, the control logic **60** monitors just the signals generated at output from the first signal detector **72** (V_{OUT_TH}) and from the second signal detectors **74** ($V_{OUTCOMP}$) to evaluate passage from the step of charging of the inductor **22***b* to the step of supply of the capacitor **27**/load **28**, and vice versa. Instead, when a negative half-wave of the input signal V_{IN} is detected, the control logic **60** monitors just the signals generated at output from the third signal detector **72'** (V_{OUT_TH}') and from the fourth signal detectors **74'** ($V_{OUTCOMP}'$) to evaluate passage from the step of charging of the inductor **22***b* to the step of supply of the capacitor **27**/load **28**, and vice versa.

[0130] The control logic **60** implements the method for control of the HV transistors **30***a*, **30***b*, **31***a* and **31***b* described previously and illustrated schematically in FIG. **11**, by a flowchart.

[0131] With reference to FIG. **11** (step **100**), the HV transistors **30***a* and **31***a* are closed. It is considered in the sequel of the description that the LV transistors **30***b* and **31***b* will always be in the closed state (situation of FIG. **3**B).

[0132] In this way, the inductor **22***b* is charged via the current $I_L=I_{ON}$ that flows through the HV transistors **30***a* and **31***a*.

[0133] The value of current $I_L=I_{ON}$ is monitored (step **102**) for detecting whether it reaches (or exceeds) the required threshold value I_{TH}. At the same time, the control logic **60** monitors the time interval T_{DELAY}. In this case, the time t**0** of start of the time interval T_{DELAY} corresponds to the instant of closing of the HV transistors **30***a*, **31** *a*, according to step **100**.

[0134] In the case where the current I_L has not reached the threshold I_{TH} or the time T_{DELAY} has not elapsed (output NO from step **102**), it is necessary to wait for both of these conditions for being satisfied and the control logic **60** maintain the system **20** in the states **100**, **102** until the condition $I_L \geq I_{TH}$ is satisfied. Otherwise (output YES from step **102**), control passes to the next step **104**.

[0135] In step **104** a check is made to verify whether the input voltage V_{IN} has positive or negative polarity. This operation may be performed by the comparator **86**, which receives the signal V_{IN}^+ at input.

[0136] As has already been said, a circuit equivalent to what is illustrated in FIG. **8** is coupled (in a way not illustrated in the figure) to the HV transistor **31***a*, and used in a similar way for receiving the signal with negative polarity V_{IN}^-.

[0137] In the case where the input voltage V_{IN} has a positive polarity, control passes to step **106** (output YES from step **104**), where the HV transistor **30***a* is opened, thus supplying the capacitor **27**/load **28** via the diode **36**.

[0138] In the case where the input voltage V_{IN} has a negative polarity, control passes instead to step **108** (output NO from step **104**), where the capacitor **27**/load **28** is supplied via the diode **38**.

[0139] From steps **106** and **108** control passes to step **110**, where the control logic **60** monitors just one between the signals $V_{OUTCOMP}$ and $V_{OUTCOMP}'$ (from the second current detector **74** and, respectively, fourth current detector **74'**, according to the polarity of the input signal) to detect whether the current I_{OUT} assumes a value equal to I_{OFF}. As long as $I_{OUT}>I_{OFF}$, the control logic **60** keeps the system **20** in the step of charging of the capacitor **27**/supply of the load **28**. When $I_{OUT}\leq I_{OFF}$, control returns to step **100**. Steps **100-104** are carried out, as described with reference to FIGS. **5***a*-**5***c*, in a time interval equal to at least T_{DELAY} until the current in the inductor reaches the threshold I_{TH}, whereas steps **106-110** are carried out in a time interval equal to T_{CHARGE}.

[0140] The control logic **60** is, for example, a microcontroller, or finite-state machine, configured to drive the HV transistors **30***a* and **31***a* in order to execute the steps of the method of FIG. **11**. In particular, the control logic is integrated in the same device that forms the energy harvesting interface according to the present invention (i.e., it is not an external component).

[0141] According to an embodiment alternative to the one illustrated in FIGS. **3**A and **3**B, the diodes **36** and **38** may be replaced by a respective N-channel MOSFET, or by any active component appropriately controlled for implementing the steps described. This embodiment is illustrated in FIG. **12** (energy harvesting system **20'**). These transistors are designated by the reference numbers **36'** and **38'**, respectively. The transistors **36'** and **38'** are controlled in an active way, by actively driving them in conduction or inhibition. This control is carried out by the control logic **60**, possibly by interposition of respective driving blocks. More in particular, the control logic **60** turns on the transistor **36'** at positive half-waves of the input voltage V_{IN} and as it verifies that the current I_{ON} has reached the value I_p for getting energy to flow to the capacitor **27**/load **28**. Likewise, the control logic **60** turns on the transistor **38'** at negative half-waves of the input voltage V_{IN} and at as it verifies that the current I_{ON} has reached a value, in modulus, equal to I_p for getting energy to flow to the capacitor **27**/load **28**.

[0142] FIG. **13** shows, by a flowchart, a method for control of the HV transistors **30***a*, **30***b*, **31***a*, **31***b*, **36'** and **38'**.

[0143] With reference to FIG. **13** (step **120**), the HV transistors **30***a* and **31***a* are closed. The transistors **36'** and **38'** instead, are opened. It is assumed, in the sequel of the description, that the LV transistors **30***b* and **31***b* are always in the closed state. In this way, the inductor **22***b* is charged via the current $I_L=I_{ON}$ that flows through the HV transistors **30***a* and **31***a*.

[0144] The value of current $I_L=I_{ON}$ is monitored (step **122**) for detecting whether it reaches (or exceeds) the threshold value I_{TH} required. At the same time, the control logic **60** monitors the time interval T_{DELAY}. In this case, the time t_0 of start of the time interval T_{DELAY} corresponds to the instant of closing of the HV transistors **30***a*, **31** *a*, according to step **120**.

[0145] In the case where the current I_L has not reached the threshold I_{TH} or the time T_{DELAY} has not elapsed (output NO from step **122**), it is necessary to wait for both of these conditions for being satisfied, and the control logic **60** keeps the system **20** in the states **120**, **122** until the condition $I_L\geq I_{TH}$ is satisfied. Otherwise (output YES from step **122**), control passes to the next step **124**.

[0146] In step **124**, a check is made to verify whether the input voltage V_{IN} has positive or negative polarity.

[0147] In the case where the input voltage V_{IN} has a positive polarity, control passes to step **126** (output YES from step **124**), where the HV transistor **30***a* is turned off and the transistor **36'** is turned on, thus supplying the capacitor **27**/load **28** via the transistor **36'**.

[0148] In the case where the input voltage V_{IN} has a negative polarity, control passes instead to step **128** (output NO from step **124**), where the capacitor **27**/load **28** is supplied via the transistor **38'**.

[0149] From steps **126** and **128** control passes to step **130**, where the control logic **60** monitors just one between the signals $V_{OUTCOMP}$ and $V_{OUTCOMP}'$ (according to the polarity of the input signal) to detect whether the current I_{OUT} assumes a value equal to I_{OFF}. As long as $I_{OUT}>I_{OFF}$, the control logic **60** keeps the system **20** in the step of charging of the capacitor **27**/supply of the load **28**. When $I_{OUT}\leq I_{OFF}$, control returns to step **120**.

[0150] FIG. **14** shows a vehicle **200** comprising the energy harvesting system **20** of FIGS. **3**A, **3**B or the energy harvesting system **20'** of FIG. **12**. The vehicle **200** is, in particular, an automobile. It is evident, however, that the energy harvesting system **20**, **20'** may be used in any vehicle **200** or in systems or apparatuses different from a vehicle. In particular, the energy harvesting system **20**, **20'** may find application in generic systems in which it is desirable to harvest, store, and use environmental energy, in particular by conversion of mechanical energy into electrical energy.

[0151] With reference to FIG. **14**, the vehicle **200** comprises one or more transducers **22** coupled in a per se known manner to a portion of the vehicle **200** subject to mechanical stress and/or vibrations, for converting said mechanical stress and/or vibrations into electric current.

[0152] The energy harvesting system **20**, **20'** is connected to one or more electrical loads **28***a*, . . . , **28***n*, for example via interposition of a DC-DC converter. In particular, according to an application of the present invention, the electrical loads **28***a*, . . . , **28***n* comprise TPM (tyre-parameter monitoring) sensors **250** for monitoring tire parameters **202**. In this case, the TPM sensors **250** are coupled to an internal portion of the tires **202** of the vehicle **200**. Likewise, also the transducers **22** (for example, of an electromagnetic or piezoelectric type) are coupled to an internal portion of the tires **202**. The stress of the transducers **22** while the vehicle **200** is travelling causes production of an electric current/voltage signal at output from the transducer **22** by conversion of mechanical energy into electrical energy. The electrical energy thus produced is stored, as described previously, in the storage element **27** and supplied, via the DC-DC converter that may possibly be present, to the TPM sensors **250**.

[0153] According to one embodiment, the energy harvesting system **20**, **20'**, comprising one or more transducers and the TPM sensors **250**, are glued inside one or more tires **202**. The impact of the tire **202** on the ground during motion of the vehicle **200** enables production of electrical energy.

[0154] As an alternative to what is illustrated in FIG. **14**, the energy harvesting system **20**, **20'** may be set in any other portion of the vehicle **200**, and/or used for supplying an electrical load different from, or additional to, the TPM sensors **250**.

[0155] Another possible application of the energy harvesting system **20**, **20'** is generation of electrical energy by exploiting the mechanical energy produced by an individual when he is walking or running. In this case, the energy harvesting system **20**, **20'** is set inside the shoes **300** of said individual (for example, in the sole) as illustrated schematically in FIG. **15**. In systems configured for fitness purposes, where it is of particular interest to count the number of steps, it is useful to recover energy from the vibrations induced by walking or running for being able to supply, without the use of batteries, acceleration sensors and/or RFID transmitters that are able to communicate with cellphones, music-playing devices, or with any other apparatus involved in processing information regarding the steps.

[0156] From an examination of the characteristics provided according to the present disclosure the advantages that it affords emerge clearly.

[0157] In particular, the parameter K has a highly reproducible value (minimal spread) for increasing the performance, sturdiness, and efficiency of the system **20**, **20'**, minimizing the mismatch between the positive polarity and negative polarity of the signal of the transducer and preventing phenomena of reversal of the flow of current from the capacitor **27** to the input terminals **25'**, **25"** of the rectifier circuit **24**.

[0158] The scavenging efficiency is likewise high even when the amplitude of the signal V_{TRANSD} of the transducer **22** is lower than the voltage value stored in the capacitor **27**, thus overcoming a limitation of the diode-bridge rectifier architecture.

[0159] Furthermore, since in the case of a transducer **22** of an electromagnetic type the rectifier **24** exploits the inductor **22***b* of the transducer **22**, the scavenging efficiency is high even when the amplitude of the signal of the transducer is low. In this way, the limitation imposed by the diode-bridge rectifiers, which require a voltage of the transducer V_{TRANSD} higher than $2V_{TH\_D}$, where V_{TH_D} is the threshold voltage of the diodes of the rectifier, is overcome.

[0160] Using a high-voltage (HV) technology for the capacitor **27** and for the energy harvesting interface, it is possible to store high voltages, and thus high energy, in the capacitor, consequently increasing the autonomy of operation of the TPM sensors **250**.

[0161] The method described enables implementation of an active control (of the mean value and of the ripple) of the current supplied by the transducer, and enables an optimal impedance matching between the transducer **22** and the energy harvesting interface **24**. This ensures a high efficiency η_{SCAV} of the energy harvesting interface **24***b* irrespective of the speed of rotation of the tyres **202** and of the conditions of storage of the energy in the capacitor **27**.

[0162] Furthermore, as has been said, the value of the interval T_{DELAY} may be varied according to the particular application in which the rectifier **24** operates. The rectifier **24** thus finds use in systems different from the energy harvesting system **20**, **20'**, based upon electromagnetic transducers of any type.

[0163] In addition, the rectifier circuit **24** may be used with transducers of some other type, with interposition of an appropriate circuit between the transducer and the rectifier circuit configured to provide an energy accumulator similar to the inductor **22***b*.

[0164] Further, the rectifier **24** and the energy harvesting system **20**, **20'** are of a fully integrated type, and consequently require minimal installation space.

[0165] Finally, environmental-energy harvesting is obtained even when the signal of the transducer is lower than the voltage value stored on the output capacitor, which is not possible using a diode-bridge interface of a known type as illustrated in FIG. **1**. According to the present invention, the energy harvesting interface **24** is thus able to harvest energy even when the power supplied by the transducer is very low.

[0166] Finally, it is clear that modifications and variations may be made to what has been described and illustrated herein, without thereby departing from the scope of the present invention, as defined in the annexed claims.

[0167] In particular, according to one embodiment, the rectifier circuit **24** may comprise a number of transistors different from what has been described. For instance, the rectifier circuit **24** may be a half-wave rectifier, comprising just the current detectors **72** and **74** or, alternatively, just the current detectors **72'** and **74'**. Use of a half-wave rectifier may be advantageous in the case where the input signal V_{IN} is of a known type and comprises only positive (or negative) half-waves. Its use is, however, not recommended (albeit possible) in energy harvesting systems in so far as part of the input signal would be lost, at the expense of the efficiency of the system as a whole.

[0168] Furthermore, it is not always necessary for both the conditions $t>T_{DELAY}$ and $I_L>I_{TH}$ expressed with reference to the operating condition of FIG. **4**A for being satisfied. In particular, for voltage signals generated by transducers **22** of a known type, the voltage value always reaches peaks such as to enable the threshold I_{TH} for being exceeded within the time T_{DELAY}. Furthermore, an appropriate choice of T_{DELAY} always guarantees, for practical purposes, reaching of a minimum acceptable threshold I_{TH}.

[0169] In addition, there may be present a plurality of transducers **22**, indifferently all of the same type or of different types. For instance, the transducer/transducers may be chosen in the group comprising: electrochemical transducers (configured to convert chemical energy into an electrical signal), electromechanical transducers (configured to convert mechanical energy into an electrical signal), electro-acoustic transducers (configured to convert variations of acoustic pressure into an electrical signal), electromagnetic transducers (configured to convert a magnetic field into an electrical signal), photo-electric transducers (configured to convert light energy into an electrical signal), electrostatic transducers, thermoelectric transducers, piezoelectric transducers, thermo-acoustic transducers, thermomagnetic transducers, and thermo-ionic transducers.

1. An energy harvesting interface, comprising:

a first switch and a second switch coupled together in series between a first input terminal and a first output terminal;

a third switch and a fourth switch connected together in series between a second input terminal and the first output terminal;

a first charge-transfer element coupled between the first input terminal and the second output terminal;

a second charge-transfer element coupled between the second input terminal and the second output terminal; and

a first electrical-signal-detecting device coupled across the second switch and configured to detect whether an electric storage current that flows through the second switch reaches a first threshold value, said electric storage current being associated to an electrical charge stored in an input storage element;

a second electrical-signal-detecting device coupled across the fourth switch and configured to detect whether an electric supply current that flows through the fourth switch reaches a second threshold value lower than the first threshold value; and

a control logic configured to:
  close the first, second, third, and fourth switches until the electric storage current reaches the first threshold value;
  open the first switch, thus generating the electric supply current for supplying an output second storage element through the first charge-transfer element by discharging the input storage element; and
  keep the first switch open as long as the electric supply current is higher than said second threshold value.

2. The energy harvesting interface according to claim **1**, wherein the control logic is further configured to:
keep closed the first, second, third, and fourth switches for a time interval during which the input storage element stores electrical charge through the electric storage current.

3. The energy harvesting interface according to claim **1**, wherein, when the first threshold value has been reached, the first and second electrical-signal detecting devices co-operate for generating said second threshold value.

4. The energy harvesting interface according to claim **3**,
wherein the first electrical-signal-detecting device comprises:
  a first comparator coupled across the second switch for comparing the value of the electric storage current with the first threshold value; and
  a current mirror configured to generate a scaled copy, scaled by a first proportionality factor, of the electric storage current; and
wherein the second electrical-signal-detecting device comprises:
  a scaling resistor having a value of resistance that is equal to the value of resistance of the fourth switch multiplied by a second proportionality factor that is coupled between the fourth switch and an intermediate node and configured to generate, on the intermediate node, a sum voltage signal given by the sum of the voltage signal across the fourth switch, which is a function of the output current, and the voltage signal across the scaling resistor, which is a function of said scaled copy of the electric storage current; and
  a second comparator, having a first conduction terminal biased at a reference voltage of a zero value and a second conduction terminal coupled to the intermediate node, configured to generate a comparison signal identifying a condition in which the sum voltage signal has a substantially zero value.

5. The energy harvesting interface according to claim **4**, wherein the condition whereby the sum voltage signal has a substantially zero value corresponds to an output electric supply current signal that flows through the third and fourth switches equal to the peak value of the electric storage current reduced by a third proportionality factor that is equal to the first proportionality factor divided by the second proportionality factor.

6. The energy harvesting interface according to claim **5**, wherein said peak value reduced by the third proportionality factor is said second threshold value.

7. The energy harvesting interface according to claim **4**, wherein said second switch is formed by a low-voltage transistor, the current mirror comprising the second switch and a mirror transistor, which have their respective gate terminals coupled together, said second switch having a width-to-length aspect ratio W/L that is equal to the aspect ratio W/L of the mirror transistor multiplied by the first proportionality factor.

8. The energy harvesting interface according to claim **4**, wherein the scaling resistor is a low-voltage transistor having a width-to-length aspect ratio W/L that is equal to the aspect ratio W/L of the fourth switch multiplied by said second proportionality factor.

9. The energy harvesting interface according to claim **4**, wherein the transistor that forms the scaling resistor is a modular transistor having a value of the width-to-length aspect ratio W/L that may be selected according to the need, said energy harvesting interface further comprising a memory that stores information for driving said transistor that forms the scaling resistor for enabling/disabling one or more modules of the transistor that forms the scaling resistor in order to select the respective width-to-length aspect ratio W/L.

10. The energy harvesting interface according to claim **7**, wherein said mirror transistors and said second switch are modular transistors which have a respective value of the width-to-length aspect ratio W/L that may be selected according to the need, said energy harvesting interface further comprising a memory that stores information for driving said mirror transistor and said second switch for enabling/disabling one or more respective modules of the mirror transistor and of the second switch in order to select the respective width-to-length aspect ratios W/L.

11. The energy harvesting interface according to claim **4**, wherein the first proportionality factor and the second proportionality factor are chosen with a value such as to optimize the coupling efficiency to the input port of the energy harvesting interface.

12. The energy harvesting interface according to claim **1**, wherein the input port may be coupled to a transducer of an electromagnetic type including an inductor, said input storage element being the inductor of said transducer of an electromagnetic type.

13. An energy harvesting system configured to supply an electrical load, comprising:
  a transducer configured to convert energy coming from an energy source external to said system into a transduced electrical signal;
  an electrical-energy harvesting interface, configured to receive at input the transduced electrical signal and supply at output an output signal; and
  an electrical-energy storage element coupled to the output terminals of the electrical-energy harvesting interface for receiving at input the output signal, and configured to store electrical energy configured to supply the electrical load;
  wherein the electrical-energy harvesting interface comprises:
    a first switch and a second switch coupled together in series between a first input terminal and a first output terminal;
    a third switch and a fourth switch connected together in series between a second input terminal and the first output terminal;
    a first charge-transfer element coupled between the first input terminal and the second output terminal;
    a second charge-transfer element coupled between the second input terminal and the second output terminal; and

a first electrical-signal-detecting device coupled across the second switch and configured to detect whether an electric storage current that flows through the second switch reaches a first threshold value, said electric storage current being associated to an electrical charge stored in said transducer;

a second electrical-signal-detecting device coupled across the fourth switch and configured to detect whether an electric supply current that flows through the fourth switch reaches a second threshold value lower than the first threshold value; and

a control logic configured to:

close the first, second, third, and fourth switches until the electric storage current reaches the first threshold value;

open the first switch, thus generating the electric supply current for supplying the electrical-energy storage element through the first charge-transfer element by discharging the transducer; and

keep the first switch open as long as the electric supply current is higher than said second threshold value.

14. The system according to claim **13**, further comprising a DC-DC converter connected between the electrical-energy storage element and the electrical load, said DC-DC converter being configured to supply the electrical load via the electrical energy stored in the electrical-energy storage element.

15. The system according to claim **13** as installed in an apparatus.

16. The system according to claim **15**, wherein the apparatus comprises one of a vehicle and an item of sports footwear.

17. A method for harvesting energy, comprising:

closing first, second, third, and fourth switches of an energy harvesting interface, wherein the first switch and the second switch are connected together in series between a first input terminal and a first output terminal, and wherein the third switch and the fourth switch are connected together in series between a second input terminal and the first output terminal;

detecting electric storage current that flows through the closed second switch;

keeping the first, second, third and fourth switches closed until the detected electric storage current reaches a first threshold value, said electric storage current being associated to electrical charge stored in an input storage element;

opening the first switch thus generating an electric supply current for supplying an output storage element through a first charge-transfer element by discharging the input storage element;

detecting the electric supply current that flows through the closed fourth switch;

keeping the first switch open as long as the electric supply current is higher than a second threshold value that is lower than the first threshold value.

18. The energy-scavenging method according to claim **17**, wherein detecting the electric supply current that flows through the closed fourth switch comprises:

comparing the electric current that flows through the second switch with the first threshold value and, in the case where said electric current through the second switch is equal to or higher than the first threshold value, then:

generating a scaled copy, reduced by a first proportionality factor, of the electric current that flows through the second switch; and

generating a multiple of said scaled copy by multiplying said scaled copy by a second proportionality factor to obtain said second threshold value.

* * * * *

US 20060184209A1

(19) **United States**

(12) **Patent Application Publication**
John et al.

(10) **Pub. No.: US 2006/0184209 A1**

(43) **Pub. Date: Aug. 17, 2006**

(54) DEVICE FOR BRAIN STIMULATION USING RF ENERGY HARVESTING

(76) Inventors: **Constance M. John**, San Francisco, CA (US); **Varghese John**, San Francisco, CA (US); **Martin H. Mickle**, Pittsburgh, PA (US)

Correspondence Address:
BUCHANAN INGERSOLL PC
(INCLUDING BURNS, DOANE, SWECKER & MATHIS)
POST OFFICE BOX 1404
ALEXANDRIA, VA 22313-1404 (US)

(21) Appl. No.: **11/219,404**

(22) Filed: **Sep. 2, 2005**

Related U.S. Application Data

(60) Provisional application No. 60/606,853, filed on Sep. 2, 2004.

Publication Classification

(51) **Int. Cl.**
A61N 1/34 (2006.01)

(52) **U.S. Cl.** **607/45**

(57) ABSTRACT

A device for brain stimulation using radio frequency harvesting is disclosed. The device includes a circuit implantable under a scalp of a patient, the circuit comprising a radio frequency harvesting power circuit and a stimulation circuit, and a plurality of electrodes coupled to the circuit, the plurality of electrodes providing brain stimulation to targeted areas of the brain. The electrodes may provide stimulation to targeted areas of the brain including deep brain stimulation for the treatment of Parkinson's disease and cortical stimulation for the treatment of stroke victims.

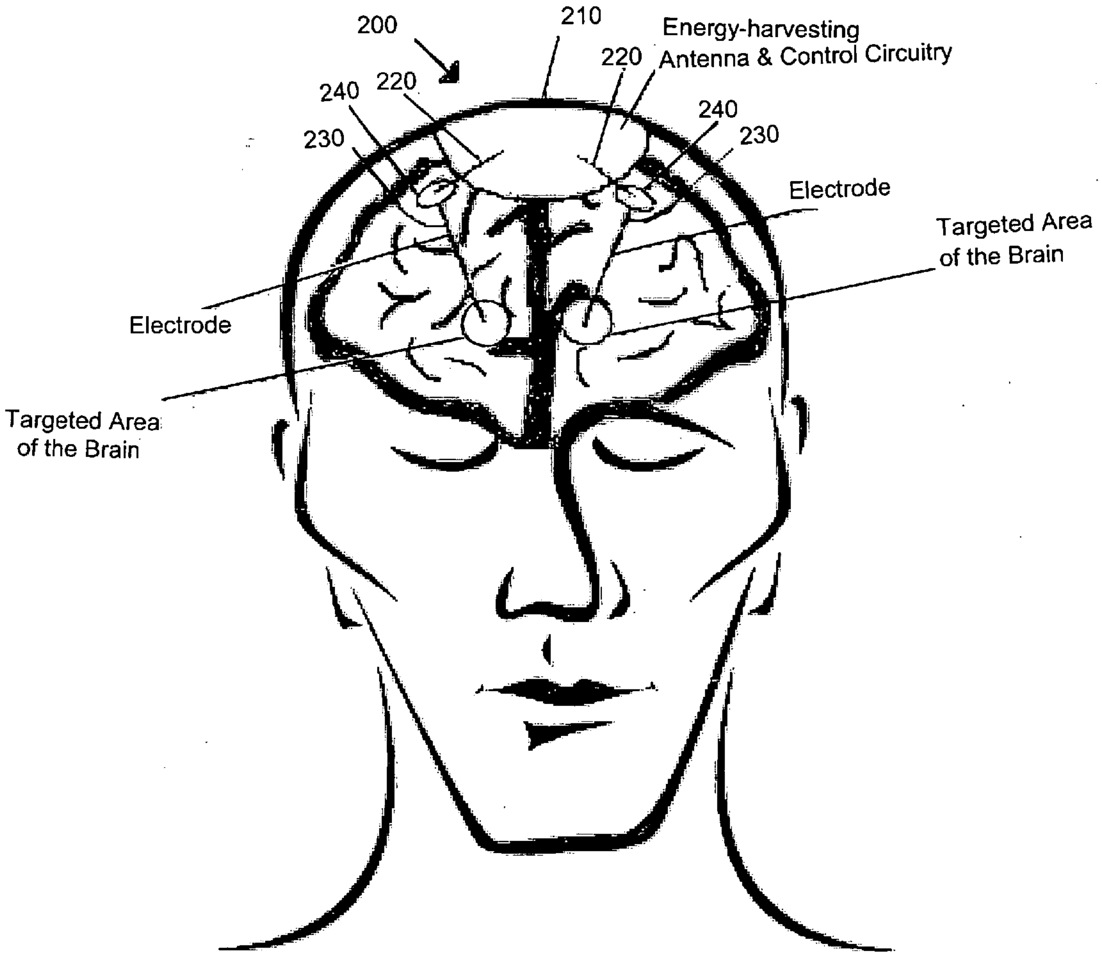
200
210
Energy-harvesting Antenna & Control Circuitry
240
220
230
220
240
230
Electrode
Targeted Area of the Brain
Electrode
Targeted Area of the Brain

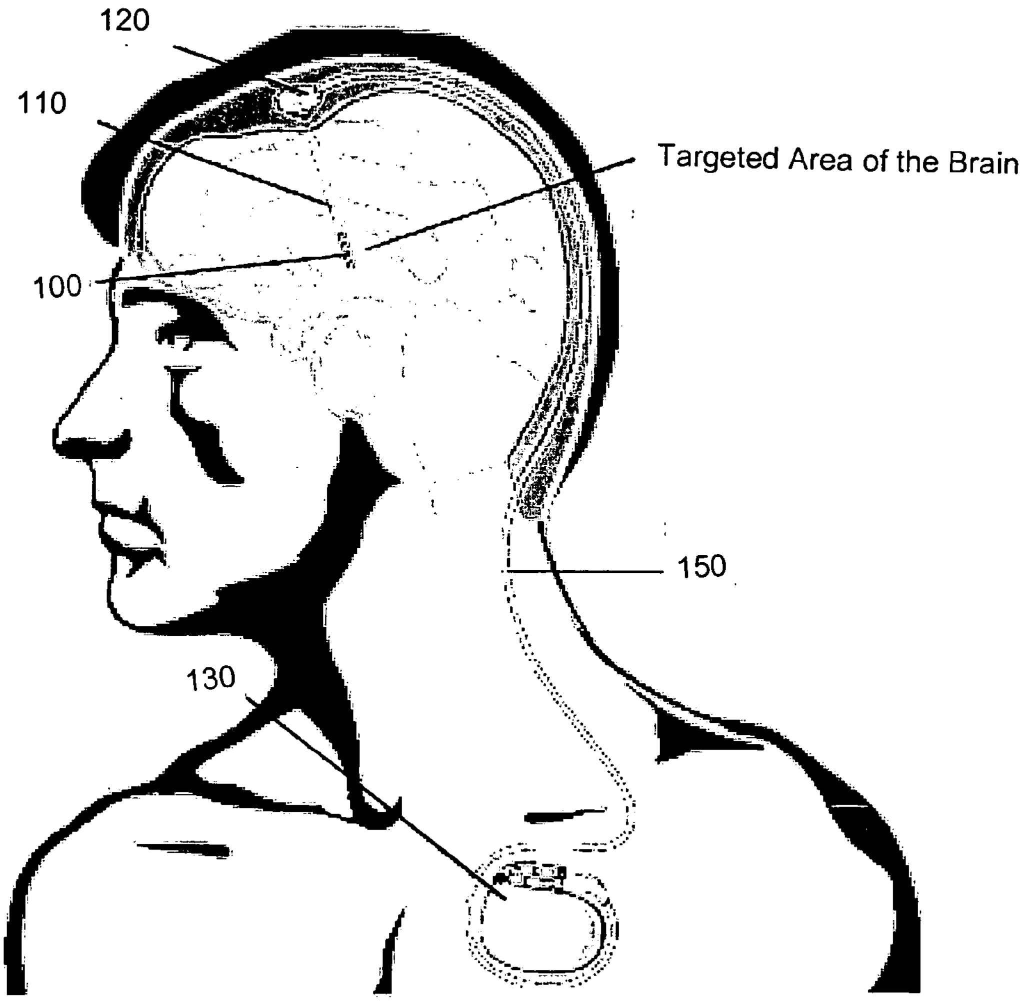
120
110
Targeted Area of the Brain
100
150
130

PRIOR ART

FIG. 1

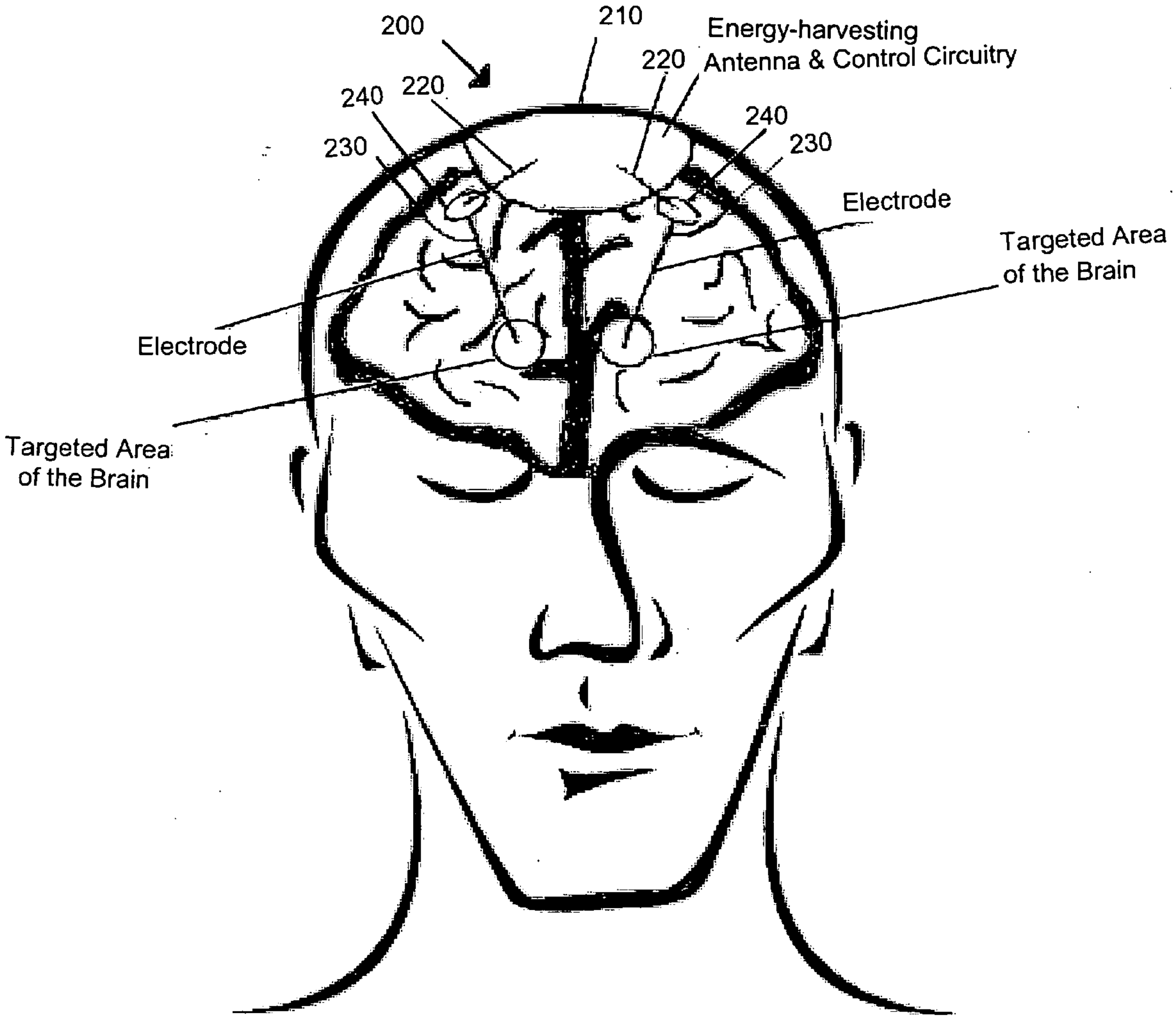
200
210
Energy-harvesting
Antenna & Control Circuitry
220
240
220
230
240
230
Electrode
Targeted Area
of the Brain
Electrode
Targeted Area
of the Brain

FIG. 2A

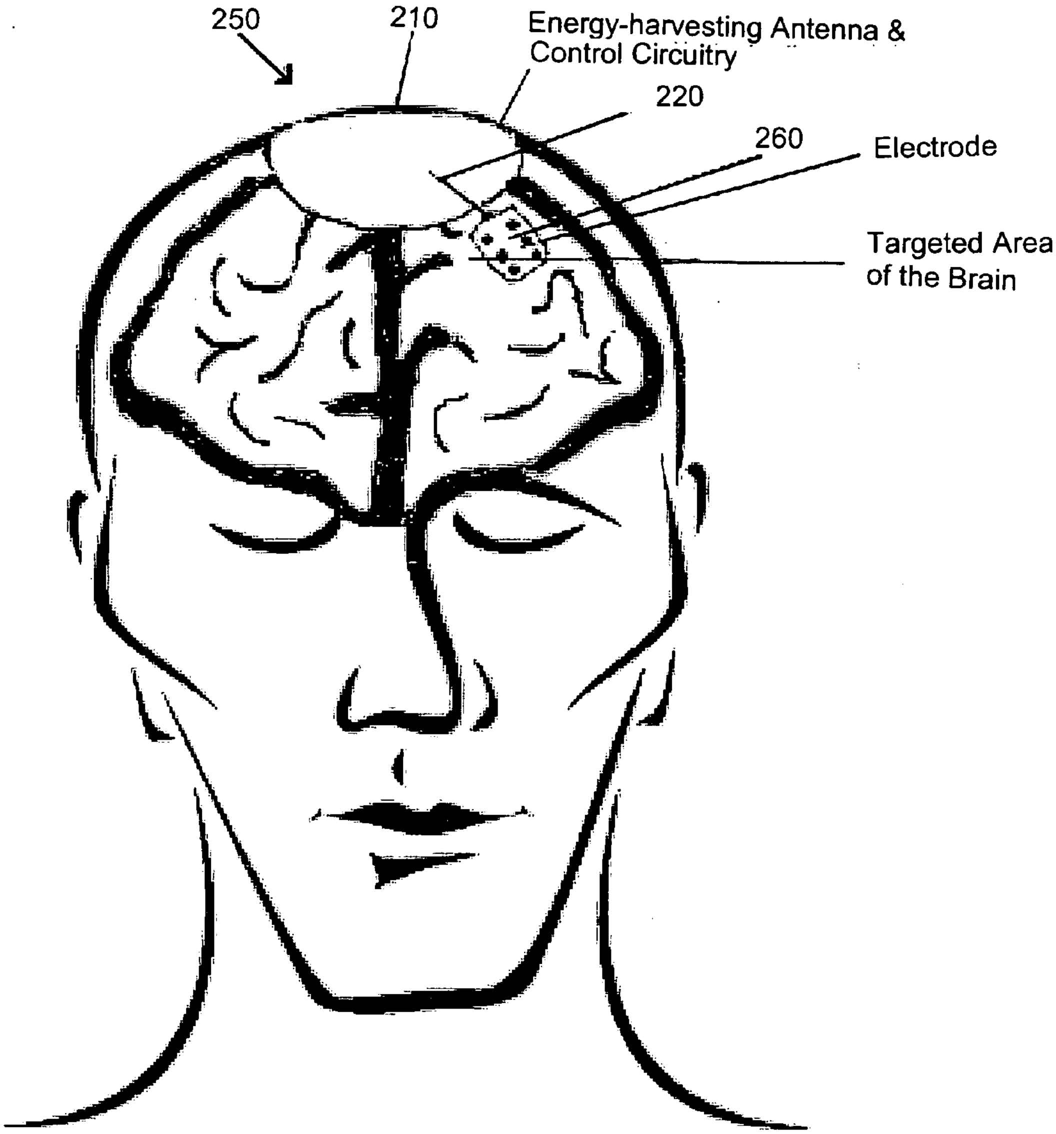
250
210
Energy-harvesting Antenna &
Control Circuitry
220
260
Electrode
Targeted Area
of the Brain

FIG. 2B

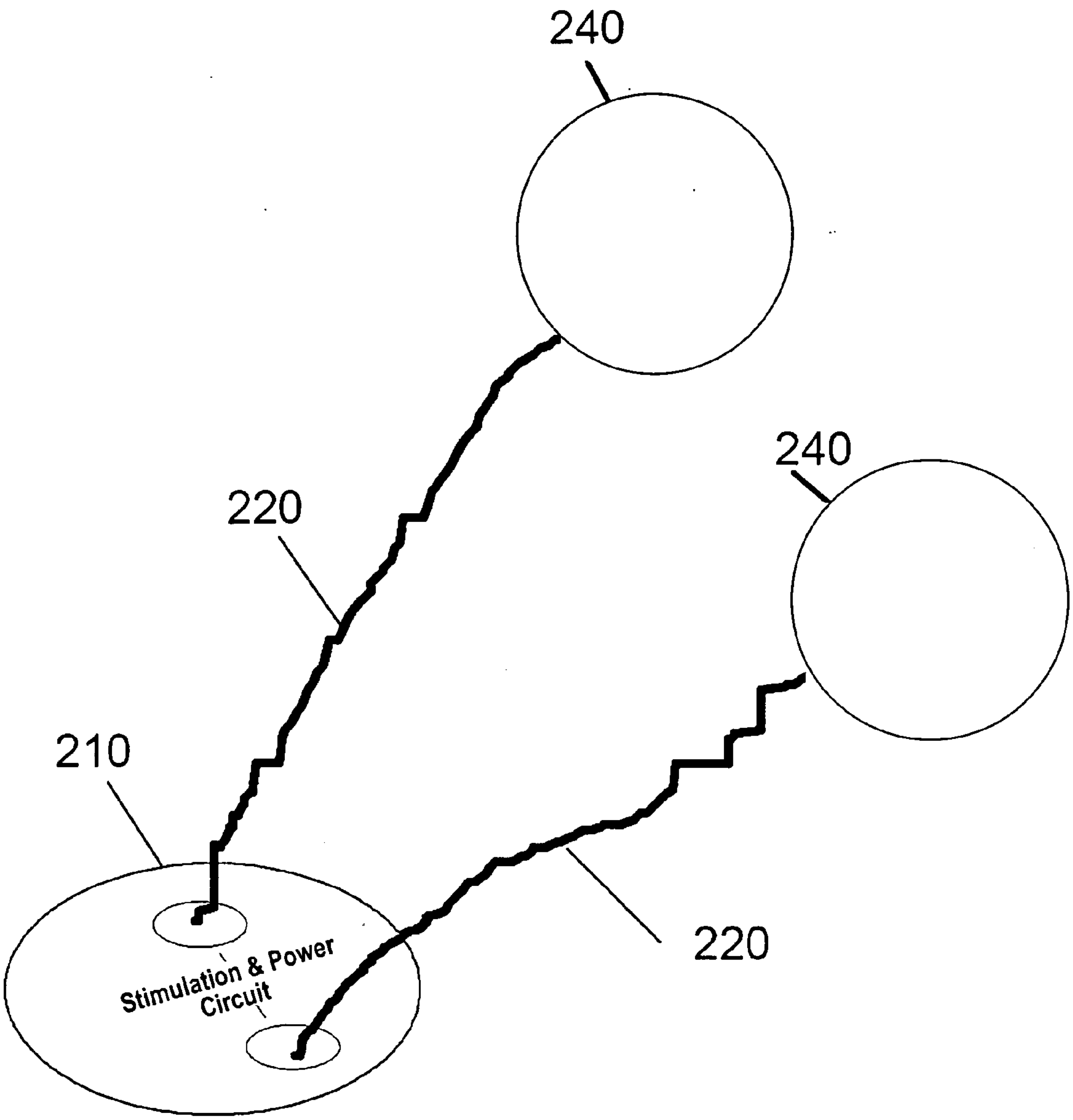
240
240
220
210
220
Stimulation & Power
Circuit

FIG. 2C

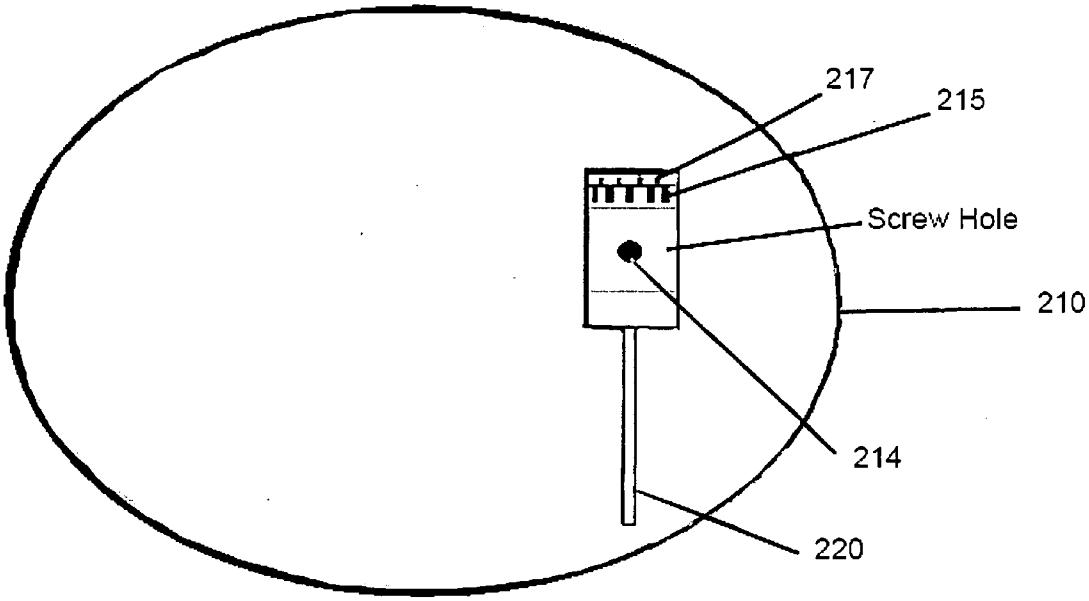
217
215
Screw Hole
210
214
220

FIG. 2D

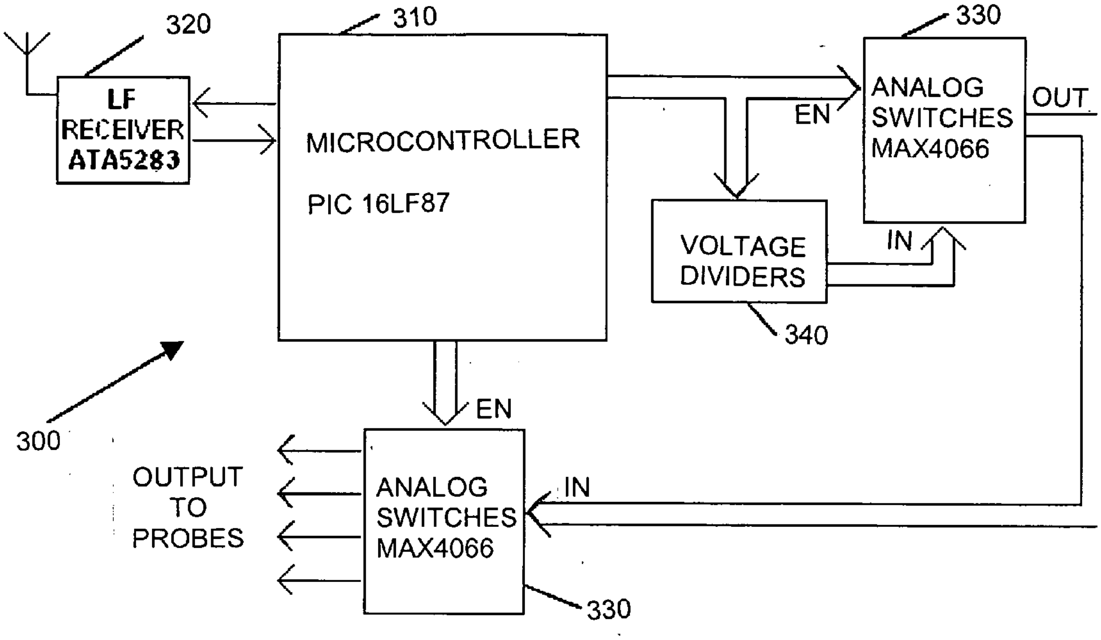
320
LF RECEIVER ATA5283
310
MICROCONTROLLER
PIC 16LF87
330
ANALOG SWITCHES MAX4066
OUT
EN
IN
VOLTAGE DIVIDERS
340
300
EN
OUTPUT TO PROBES
ANALOG SWITCHES MAX4066
IN
330

FIG. 3

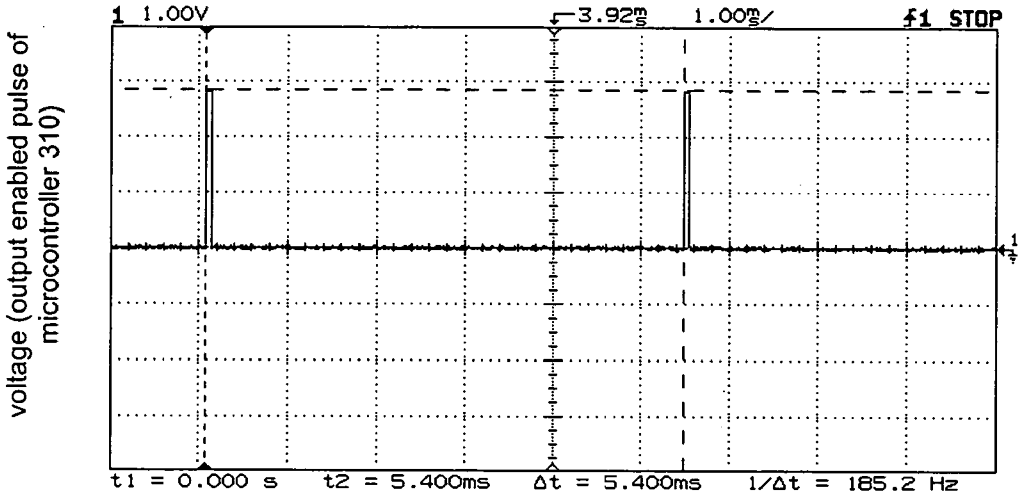
voltage (output enabled pulse of microcontroller 310)
1 1.00V
-3.92ms
1.00ms/
f1 STOP
1
t1 = 0.000 s
t2 = 5.400ms
Δt = 5.400ms
1/Δt = 185.2 Hz

FIG. 4

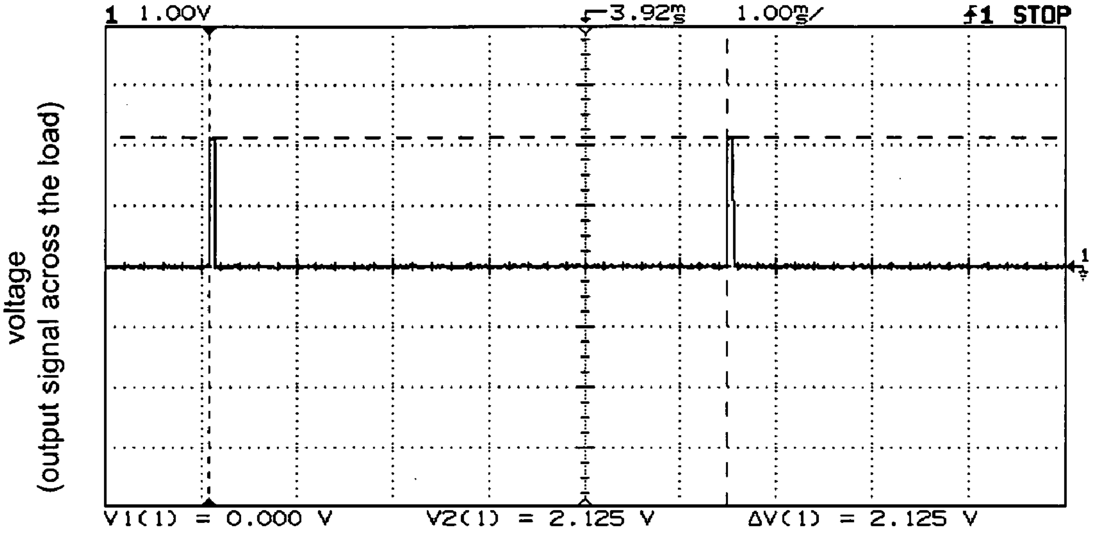
1 1.00V
3.92ms
1.00ms/
f1 STOP
voltage
(output signal across the load)
1
V1(1) = 0.000 V
V2(1) = 2.125 V
ΔV(1) = 2.125 V

FIG. 5

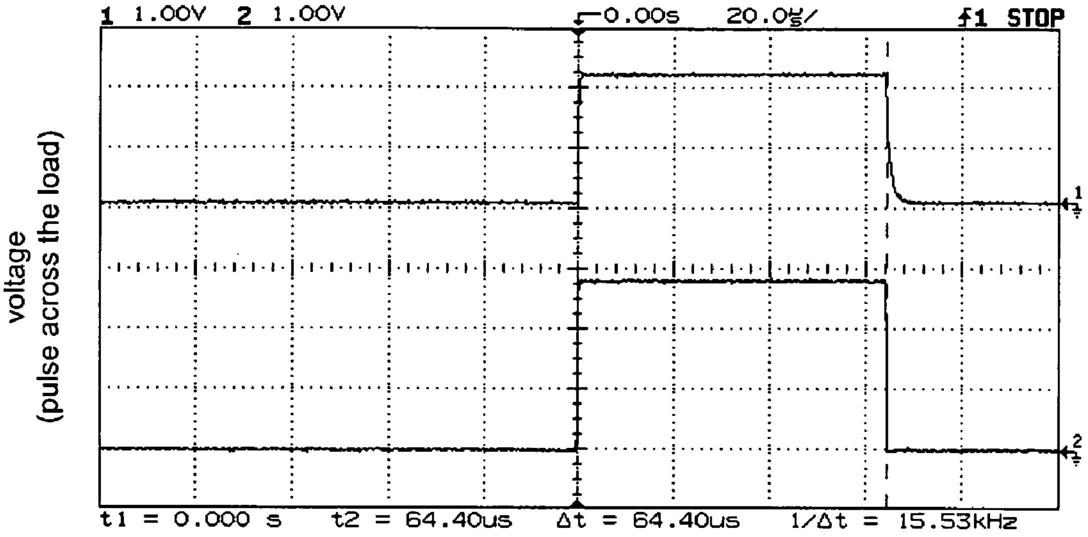
1 1.00V 2 1.00V
0.00s 20.0µs/
f1 STOP
voltage
(pulse across the load)
1
2
t1 = 0.000 s t2 = 64.40us Δt = 64.40us 1/Δt = 15.53kHz

FIG. 6

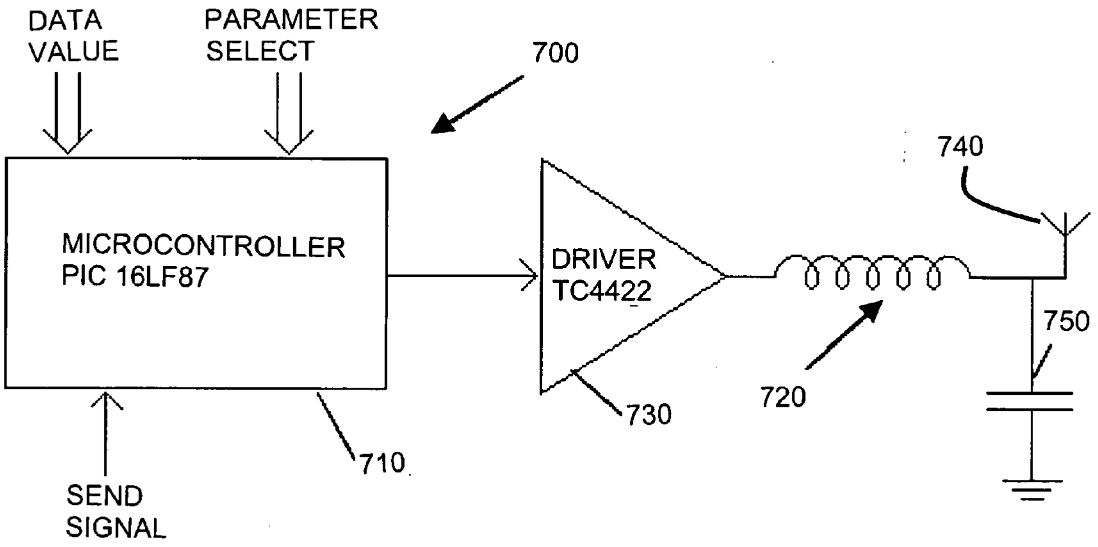
DATA
VALUE
PARAMETER
SELECT
700
740
MICROCONTROLLER
PIC 16LF87
DRIVER
TC4422
750
730
720
710
SEND
SIGNAL

FIG. 7

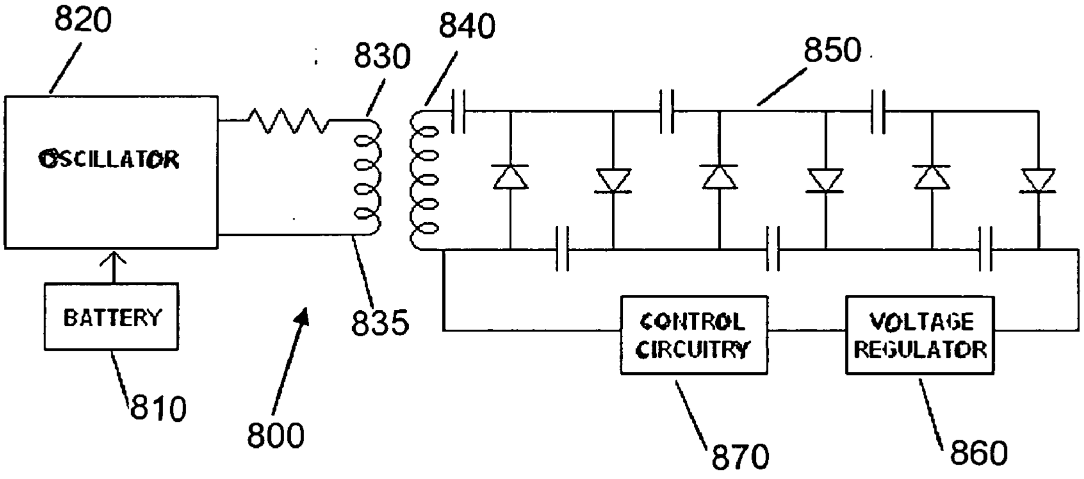
OSCILLATOR
BATTERY
CONTROL CIRCUITRY
VOLTAGE REGULATOR

FIG. 8A

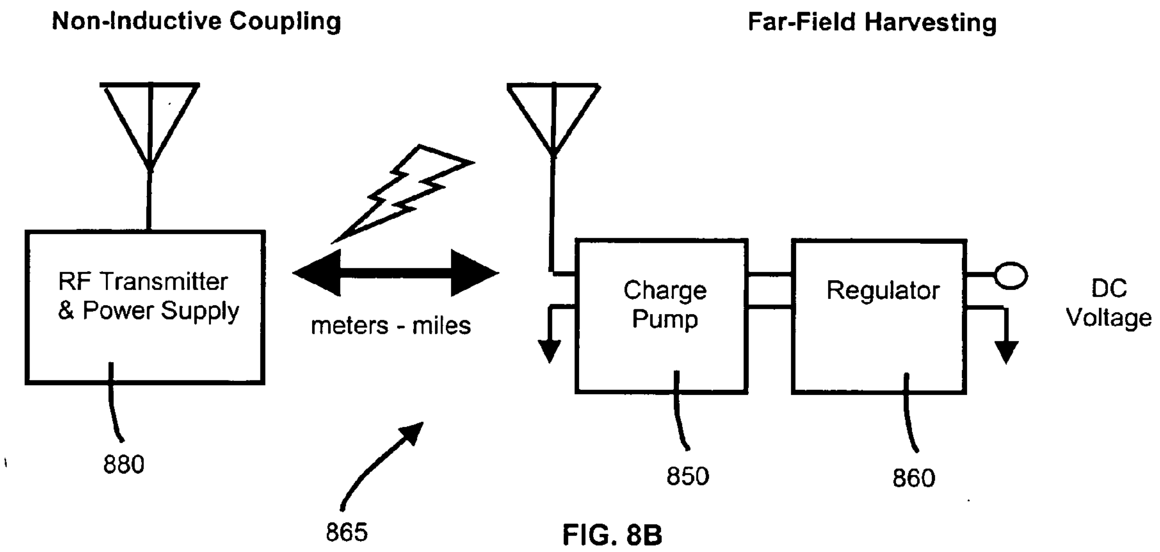
Non-Inductive Coupling
Far-Field Harvesting
RF Transmitter & Power Supply
meters - miles
Charge Pump
Regulator
DC Voltage
880
865
850
860

FIG. 8B

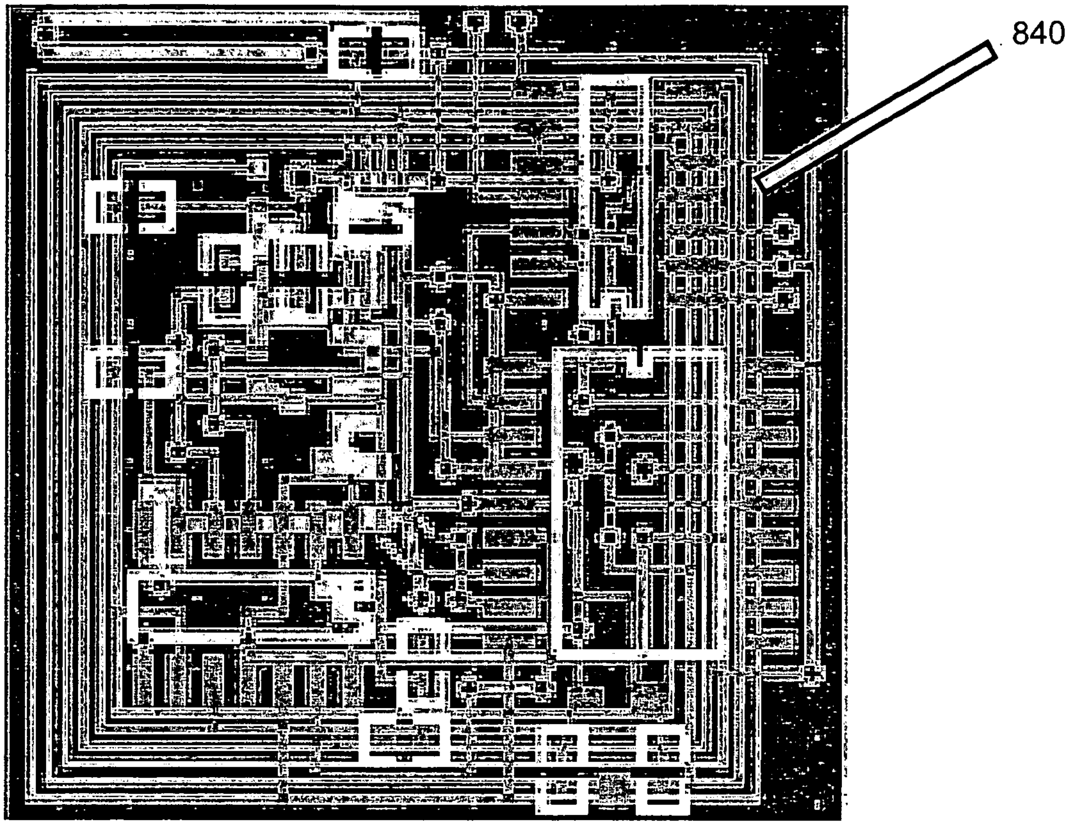

FIG. 9

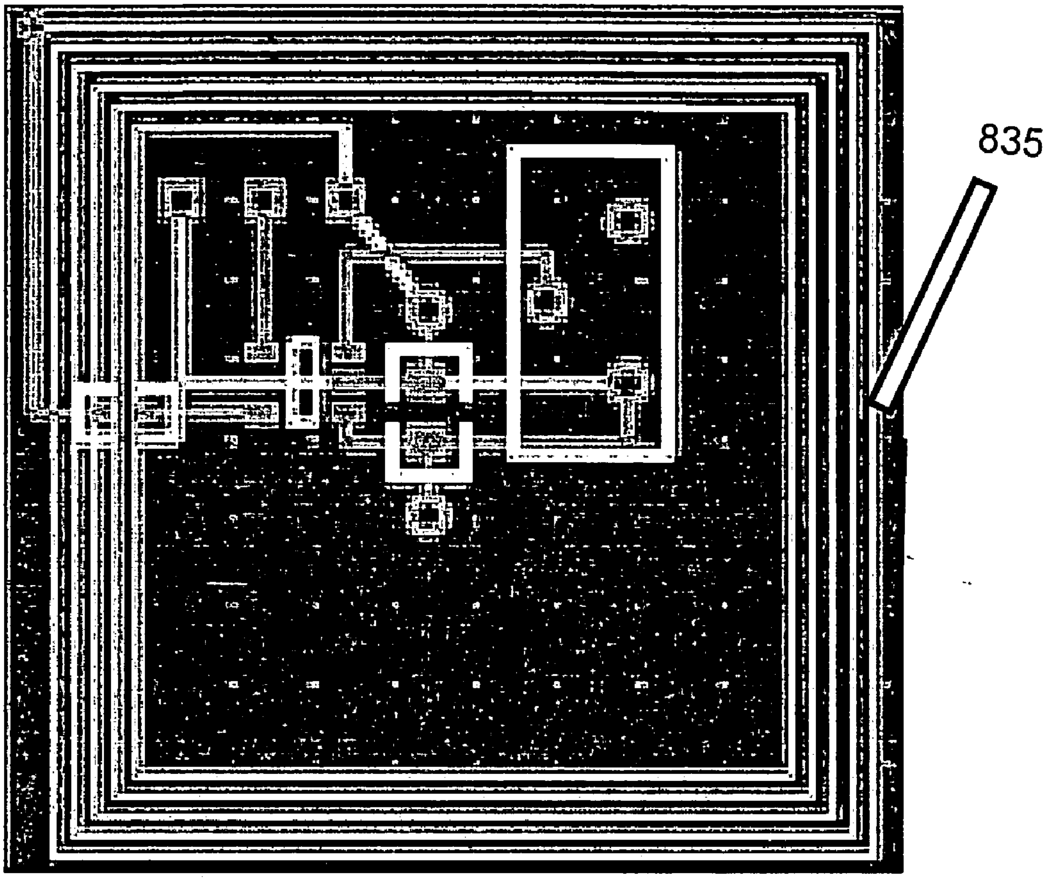

FIG. 10

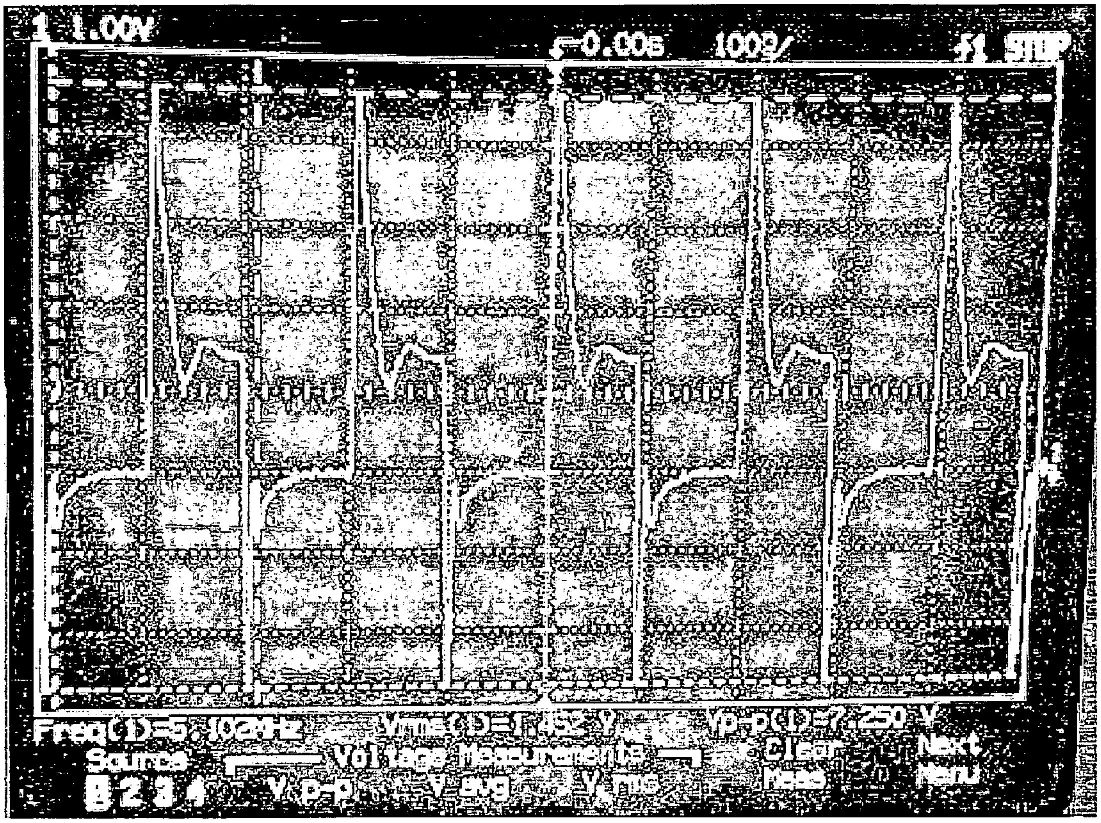
1 1.00V
0.00s
STOP
Freq(1)=5.102Hz
Vrms(1)=1.452 V
Vp-p(1)=7.250 V
Source
Voltage Measurements
Clear
Meas
Next
Menu
1 2 3 4
V p-p
V avg
V rms

FIG. 11

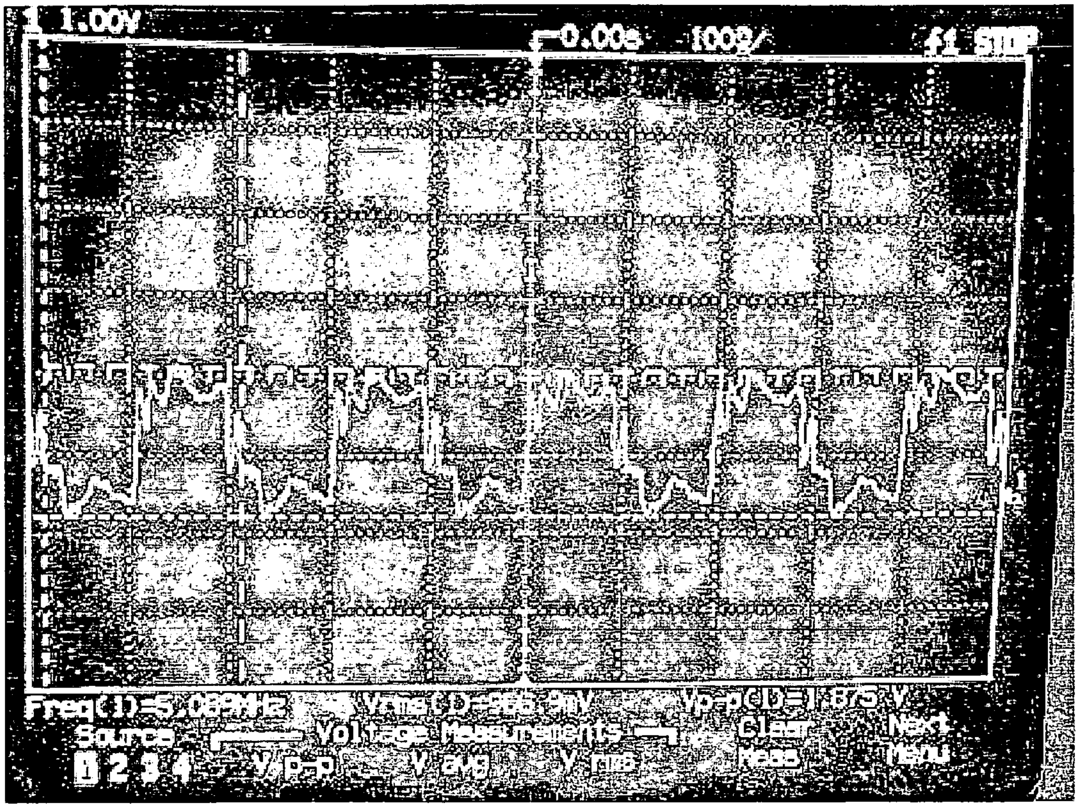
1 1.00V
0.00s
STOP
Source
Voltage Measurements
Clear
Meas
Next
Menu
1 2 3 4
V p-p
V avg
V rms

FIG. 12

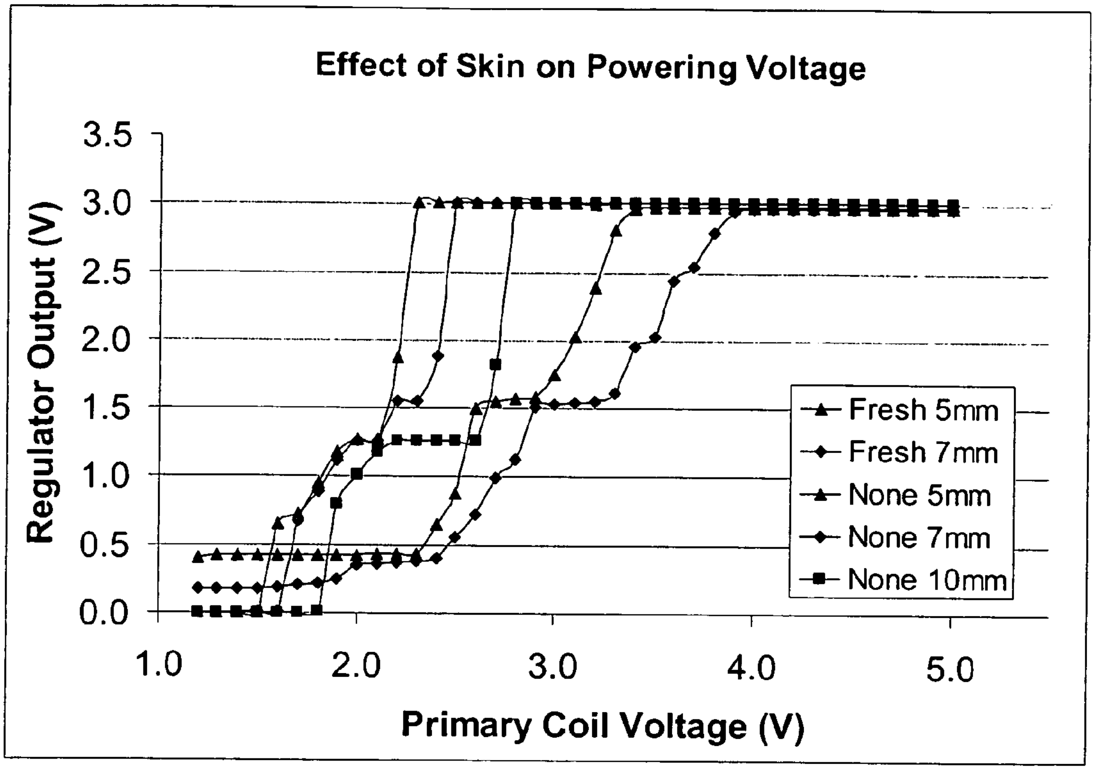
Effect of Skin on Powering Voltage
Regulator Output (V)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1.0
2.0
3.0
4.0
5.0
Primary Coil Voltage (V)
Fresh 5mm
Fresh 7mm
None 5mm
None 7mm
None 10mm

FIG. 13

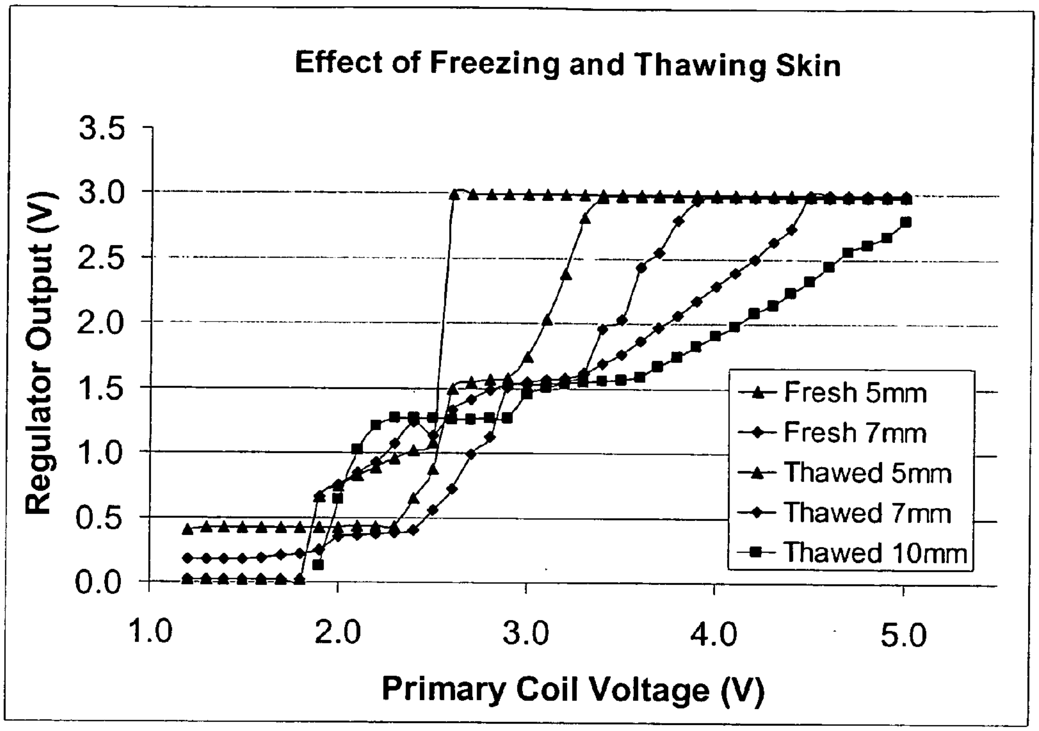
Effect of Freezing and Thawing Skin
Regulator Output (V)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1.0
2.0
3.0
4.0
5.0
Primary Coil Voltage (V)
Fresh 5mm
Fresh 7mm
Thawed 5mm
Thawed 7mm
Thawed 10mm

FIG. 14

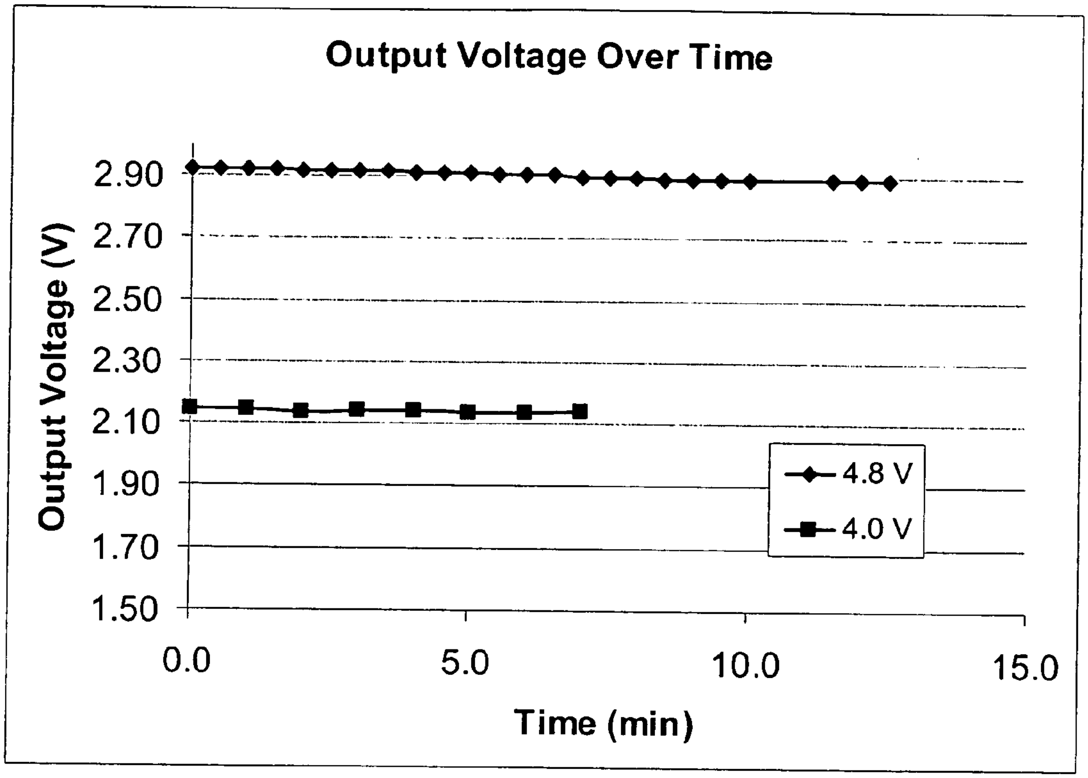
Output Voltage Over Time
Output Voltage (V)
2.90
2.70
2.50
2.30
2.10
1.90
1.70
1.50
0.0
5.0
10.0
15.0
Time (min)
4.8 V
4.0 V

FIG. 15

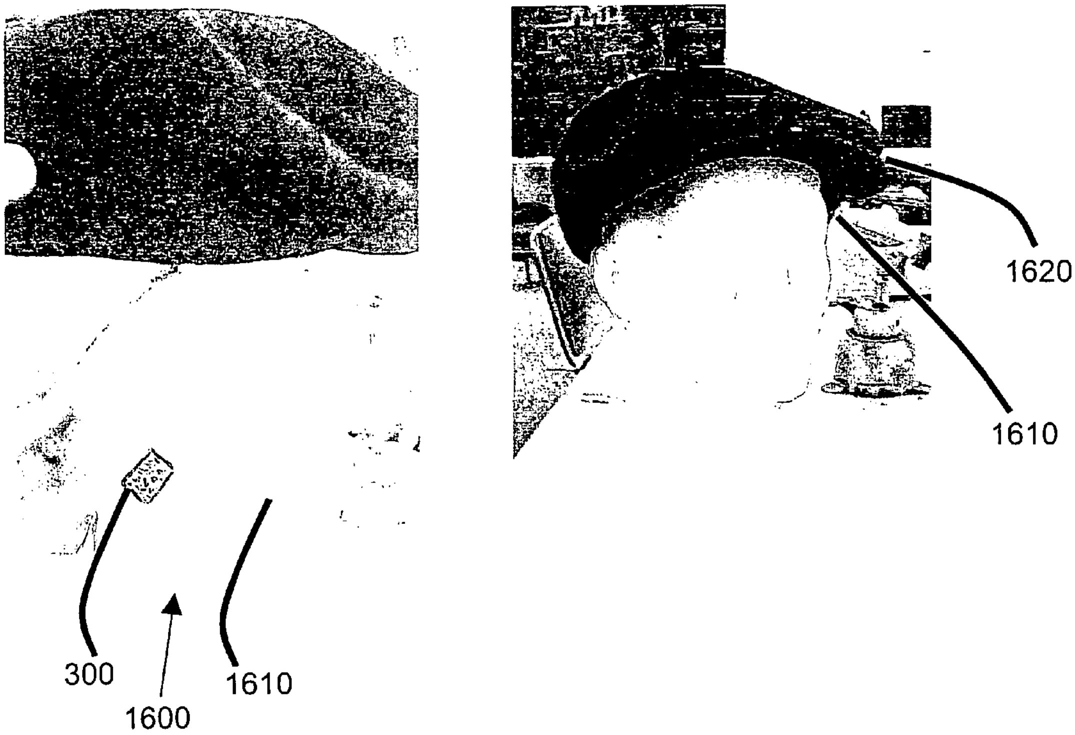
1620
1610
300
1600
1610

FIG. 16

DEVICE FOR BRAIN STIMULATION USING RF ENERGY HARVESTING

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims priority under 35 U.S.C. 119(e) from provisional patent application Ser. No. 60/606,853, entitled "Device For Deep Brain Stimulation (DBS) Using RF Energy Harvesting", filed on Sep. 2, 2004, the disclosure of which is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

[0002] The present invention relates generally to systems and apparatus for providing brain stimulation and more particularly to a device for harvesting radio frequency (RF) energy that can be implanted under a human scalp to produce stimulation in different regions of the brain, including deep brain stimulation (DBS).

BACKGROUND OF THE INVENTION

[0003] DBS is a surgical technique first used in humans over 25 years ago. DBS has been used in a wide variety of brain targets, including the thalamus, globus pallidus and the subthalamic nucleus. Diseases that have been effectively treated with DBS include movement disorders including essential tremor [Lyons K E, Pahwa R. Deep Brain Stimulation and Essential Tremor, J Clin Neurophysiol. 2004 January-February;21(1):2-5], Parkinson's disease [Byrd D L, Marks W J Jr, Starr P A. Deep brain stimulation for advanced Parkinson's disease. AORN J. 2000 September;72(3):387-90, 393-408] and dystonia [Vidailhet M. et al., Bilateral deep-brain stimulation of the globus pallidus in primary generalized dystonia. N Engl J Med. 2005 Feb. 3;352(5):459-67]. Other indications for DBS are being explored, including chronic pain, cluster headache, persistent vegatative state, epilepsy, Alzheimer's, and psychiatric disorders including obsessive-compulsive disorder and intractable depression.

[0004] Parkinson's disease (PD) is an idiopathic neurodegenerative disorder that is characterized by the presence of tremor, rigidity, akinesia or bradykinesia (slowness of movement) and postural instability. It is believed to be caused by the loss of a specific, localized population of neurons in a region of the brain called the substantia nigra. These cells normally produce dopamine, a neurotransmitter that allows brain cells to communicate with each other. These dopaminergic cells in the substantia nigra are part of an elaborate motor circuit in the brain that runs through a series of discrete brain nuclei known as the basal ganglia that control movement. It is believed that the symptoms of PD are caused by an imbalance of motor information flow through the basal ganglia.

[0005] Conventionally, a medication known as levodopa has been the mainstay of treatment for patients with Parkinson's disease. However, long-term therapy with levodopa has several well-known complications that limit the medications effectiveness and tolerability. The first of these is the development of involuntary movements known as dyskinesias. These movements can be violent at times and as or more disabling than the Parkinson's symptoms themselves. The other frequent complication is the development of "on-off" fluctuations, where patients cycle between periods of good function (the "on" period) and periods of poor function (the "off" period). These fluctuations can become very frequent, up to 7 or more cycles per day, and can cause patients to become suddenly and unpredictably "off" to the point where they cannot move.

[0006] Lesioning procedures such as pallidotomy were known to improve the motor symptoms of Parkinson's disease, presumably by disruption of the abnormal neuronal activity in the motor circuitry of the basal ganglia. The discovery that MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) produced a Parkinsonian-like state in non-human primates allowed electrophysiologic study of this phenomenon by numerous investigators. The discovery that high frequency stimulation could mimic the effect of lesioning led to the use of DBS for PD in humans in the early 1990's. DBS was found to improve all of the cardinal symptoms of Parkinson's disease while allowing the patient to decrease or sometimes even eliminate the amount of levodopa medication, therefore decreasing both dyskinesia and "on-off" fluctuations.

[0007] DBS is currently the surgical treatment of choice for medically refractory Parkinson's disease. Two brain targets have been found to provide clinical benefit when chronically stimulated; the subthalamic nucleus (STN) and the internal segment of the globus pallidus (GPi). In a recent prospective, double-blinded cross-over study involving 96 patients with STN DBS and 38 patients with GPi DBS, the STN group reported an improvement in the percentage of time spent during the day with good mobility and without dyskinesia from 27% to 74%. The GPi group also reported a significant improvement, from 28% to 64%.

[0008] Although the mechanism of action is not fully understood it is thought to act by either depolarization blockade, synaptic inhibition, synaptic depression, or stimulation-induced modulation of pathologic network activity [McIntyre C C, Savasta M, Walter B L, Vitek J L. How does deep brain stimulation work? Present understanding and future questions. J Clin Neurophysiol. 2004 January-February;21(1):40-50]. It is believed that DBS acts somehow to suppress the neuronal activity by the stimulation of the region of the brain immediately adjacent to the electrode. This hypothesis seems to be supported by the fact that lesioning a specific structure in the brain has the same clinical effect as stimulating that same structure at high (greater than 100-150 Hz) frequency. In fact, DBS has largely replaced the older lesioning procedures (such as pallidotomy and thalamotomy) that used to be the mainstay of surgical treatment for movement disorders such as Parkinson's disease. The high frequency stimulation may act to hyperpolarize immediately adjacent neurons such that they become incapable of producing normal action potentials. An alternative hypothesis is that DBS may be altering more distant structures or even fibers from far removed nerve cells that are passing through or near the area of stimulation. Whatever the mechanism of action, DBS has a distinct advantage over the older lesioning techniques because it is an adjustable therapy and does not involve destruction of the patient's brain tissue.

[0009] Prior art DBS devices have several limitations that can lead to adverse effects including infection, cutaneous erosion, and lead breaking or disconnection [Temel Y,

Ackermans L, Celik H, Spincemaille G H, Van Der Linden C, Walenkamp G H, Van De Kar T, Visser-Vandewalle V. Management of hardware infections following deep brain stimulation. Acta Neurochir (Wien) 2004;146(4):355-61; Putzke J D, Wharen R E, Jr., Wszolek Z K, Turk M F, Strongosky A J, Uitti R J. Thalamic deep brain stimulation for tremor-predominant Parkinson's disease. Parkinsonism Relat Disord 2003;10(2):81-8.; Umemura A, Jaggi J L, Hurtig H I, Siderowf A D, Colcher A, Stern M B, Baltuch G H. Deep brain stimulation for movement disorders: morbidity and mortality in 109 patients. J Neurosurg 2003;98(4):779-84; Hariz M I. Complications of deep brain stimulation surgery. Mov Disord 2002;17(Suppl 3):S162-**6**]. One study found that 27% of 66 patients with implanted DBS devices had hardware problems [Kondziolka D, Whiting D, Germanwala A, Oh M. Hardware-related complications after placement of thalamic deep brain stimulator systems. Stereotact Funct Neurosurg 2002;79(3-4):228-33. This relatively high incidence of hardware problems is similar to the results of a study where 20 (25.3%) of 79 patients who received 124 permanent DBS electrode implants had 26 hardware-related complications [Oh M Y, Abosch A, Kim S H, Lang A E, Lozano A M. Long-term hardware-related complications of deep brain stimulation. Neurosurgery 2002;50(6):1268-74; discussion 1274-**6**]. In addition, intracranial electrode implantation can induce a hematoma or contusion. Nonetheless, most authors agree that the benefit to risk ratio of DBS is favorable.

[0010] A prior art DBS device is shown in **FIG. 1** and includes an electrode **100** disposed in a targeted area of the brain. The electrode is coupled to a lead **110** held in place at the top of the skull by a securement device **120**. The lead **110** is coupled to a neurostimulator **130** powered by a battery-powered pulse generator **140** by means of a lead **150**. The lead **150**, which averages about 15 inches in length, is implanted under the scalp and traverses the length of the patient's neck to the chest where the neurostimulator **130** and battery **140** are implanted.

[0011] The pulse generator **140** is typically placed underneath the skin just below the collar bone and is capable of stimulating at one or any combination of the four contacts present on the end of the electrode **110** in the brain. The parameters of the stimulating current (voltage, frequency, pulse width) can also be selected by the treating physician or health care worker. The pulse generator **140** is programmed through the skin via a telemetry device that allows the practitioner to select the desired stimulation parameters and also perform diagnostic tests on the device.

[0012] Implantation of the conventional DBS device is costly as for implantation of a single electrode in the brain for treatment of one side of the body the procedure requires three incisions; one on the top of the head, one behind the ear and the third just below the collarbone where the leads are connected. The implantation of the electrode **110** and the implantable pulse generator **140** is sometimes performed on different days. The incisions can be prone to infection in the immediate postoperative period. In some elderly patients with thin skin, the pulse generator **140** or wire can erode through the skin and become exposed to potential contamination. Infection or erosion often results in the need to remove the entire device, as antibiotic treatment alone in this setting rarely will clear the infection adequately. The lead **150** restricts the patient's mobility in the neck region and may break. Furthermore, the battery **140** must be replaced every three to five years. Additional drawbacks of the DBS device include the risk of erosion of the leads or hardware, infection, and magnetic sensitivity.

[0013] A prior art deep brain stimulation system is disclosed in U.S. Pat. No. 6,920,359 entitled "Deep Brian Stimulation System for the Treatment of Parkinson's Disease or Other Disorders". The DBS system includes a small, implantable pulse generator implanted directly in the cranium of the patient, thereby eliminating the long lead wires conventionally used. The disclosed system does not provide for the harvesting of energy to power the pulse generator.

[0014] As noted in Table 1, there are several current and potential indications for deep brain stimulation.

[0015] Known systems for providing electrical stimulus to the motor cortex of the brain, such as the Northstar Stroke Recovery Treatment System available from Northstar Neuroscience, Inc., also include an implantable pulse generator implanted in the pectoral area of a patient. A cortical stimulation lead includes an electrode connected to the implantable pulse generator which is used to deliver stimulation to the cortex. The electrode is placed on top of the dura and coupled to the implantable pulse generator by means of a lead which traverses the length of the patient's neck to the patient's pectoral area.

[0016] Motor cortex stimulation (MCS) is a process involving the application of stimulation signals to the motor cortex in the brain of a patient during physical rehabilitation of the disabled region of the body. The MCS system includes a pulse generator connected to a strip electrode that is surgically implanted over a portion of only the motor cortex (precentral gyrus). Because MCS involves the application of stimulation signals to surface regions of the brain rather than deep neural structures, electrode implantation procedures for MCS are significantly less invasive and time consuming than those for DBS. The current evaluation of MCS is for stroke. Stroke-related disabilities affect more than 200,000 people in the U.S. each year. Good results have been reported in MCS treatment of stroke victims. With a MCS device, a stamp-sized electrode is placed on the surface of the brain. This is attached to a wire that goes through the neck to an implantable pulse generator in the pectoral area.

[0017] Neurostimulation and responsive neurostimulation (RNS) are also being tested for the treatment of medically refractory epilepsy. In treating epilepsy, the RNS system can be designed to detect abnormal electrical activity in the brain and respond by delivering electrical stimulation to normalize brain activity before the patient experiences seizure symptoms. For either neurostimulation or RNS for treatment of epilepsy the electrode or electrodes of the device deliver a short train of electrical pulses to the brain near the patient's seizure focus.

[0018] In order to obviate the need for long leads and batteries, attempts have been made in the prior art to transmit energy through space from a base station to a remote station. One such system is disclosed in U.S. Pat. No. 6,289,237 entitled "Apparatus for Energizing a Remote Station and Related Method". The base station transmits energy which may be RF power, light, acoustic, magnetic or other suitable forms of space transmitted or "radiant" energy to the remote station. Within the remote station, the received

energy is converted into DC power which serves to operate the remote station. The source of power for the remote station is the base station and, therefore, there is no need for the remote station to carry an electrical storage device such as a battery. It is suggested that this facilitates the remote station being encapsulated within a suitable protective material, such as a resinous plastic. Homopolymers, elastomers and silicon dioxide are also suggested as suitable materials for such purposes. Further, it is suggested that this facilitates miniaturization of the remote station and placing the remote station in functionally desirable locations which need not be readily accessible. The remote station, for example, could be implanted in a patient.

[0019] The use of a wireless communication link between a base station and transponders in a radio frequency identification system employing modulated back-scattered waves is also known. See Rao, An Overview of Bulk Scattered Radio Frequency Identification System (RFID) IEEE (1999). It has also been suggested to employ a silicon chip in a transponder having a change pump on voltage doubler current. Hornby, RFID Solutions for the Express Parcel and Airline Baggage Industry, Texas Instruments, Limited (Oct. 7, 1999).

[0020] For use in miniaturized electronic chip systems, an electronic article containing a microchip having at least one antenna structured to communicate with an antenna remotely disposed with respect to the microchip is disclosed in U.S. Pat. No. 6,615,074 entitled "Apparatus for Energizing a Remote Station and Related Method". Power enhancement is achieved using a voltage doubler. The antenna of the disclosed apparatus is comparable in volume to a Smart Dust device. Smart Dust is a combination MEMS/Electronic device on the order of 1 mm×1 mm×1 mm.

[0021] What is needed therefore is a brain stimulation device that overcomes the disadvantages of the prior art brain stimulation devices. What is needed is a brain stimulation device that requires a single implantation site and surgery. What is also needed is a brain stimulation device that uses RF energy as a power source. What is further needed is a brain stimulation device that converts RF energy and stores the converted RF energy. What is also needed is a brain stimulation device that is flexible and implantable under the scalp. What is needed is a brain stimulation device that does not require leads or a pulse generator to be placed outside of the head area that are subject to disconnection or breakage. What is also needed is a brain stimulation device for electrical stimulation in the brain that is smaller and more self-contained and that does not require a pulse generator to be implanted elsewhere in the body. What is further needed is a device that is less susceptible to hardware problems or complications. What is needed is a device that has less potential for erosion through the skin. What is also needed is a device that is has a power source that does not need to be replaced.

SUMMARY OF INVENTION

[0022] The device for brain stimulation using RF energy harvesting of the present invention overcomes the disadvantages of the prior art, fulfills the needs in the prior art, and accomplishes its various purposes by providing a brain stimulation device that harvests radio frequency energy and is implantable under the scalp. The brain stimulation device of the invention may include an electrode that penetrates into the brain to provide neurostimulation to the brain. The brain stimulation device may also include an electrode that is used to provide stimulation to the brain cortex.

[0023] In accordance with one aspect of the invention, a device for brain stimulation using radio frequency harvesting includes a circuit implantable under a scalp of a patient, the circuit comprising a radio frequency harvesting power circuit and a stimulation circuit, and a plurality of electrodes coupled to the circuit, the plurality of electrodes providing brain stimulation to targeted areas of the brain. An advantage of this system is that it may use "trickle charging" wherein the device is charged by the harvesting power circuit. Moreover, anther advantage of this invention is the power transmitter which sends power to the device can be used both to send power and to send information.

[0024] There has been outlined, rather broadly, the more important features of the invention in order that the detailed description thereof that follows may be better understood, and in order that the present contribution to the art may be better appreciated. There are, of course, additional features of the invention that will be described below and which will form the subject matter of the claims appended herein.

[0025] In this respect, before explaining at least one embodiment of the invention in detail, it is to be understood that the invention is not limited in its application to the details of design and to the arrangement of components set forth in the following description or illustrated in the drawings. The invention is capable of other embodiments and of being practiced and carried out in various ways. Also, it is to be understood that the phraseology and terminology employed herein, as well as the abstract, are for the purpose of description and should not be regarded as limiting.

[0026] As such, those skilled in the art will appreciate that the conception upon which this disclosure is based may readily be utilized as a basis for the designing of other methods and systems for carrying out the several purposes of the present invention. It is important, therefore, that the claims be regarded as including such equivalent methods and systems insofar as they do not depart from the spirit and scope of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0027] The accompanying drawings, which are incorporated in and form a part of the specification, illustrate the embodiments of the present invention and together with the description, serve to explain the principles of the invention. In the drawings:

[0028] **FIG. 1** is a schematic representation of a prior art DBS device;

[0029] **FIG. 2A** is a schematic representation of a device for deep brain stimulation using RF energy harvesting in accordance with the invention;

[0030] **FIG. 2B** is a schematic representation of a device for cortical stimulation using RF energy harvesting in accordance with the invention;

[0031] **FIG. 2C** is a schematic representation of the device of **FIG. 2A** illustrating lead securement devices;

[0032] **FIG. 2D** is a schematic representation of the device of **FIG. 2A** illustrating an attachment means for connecting a lead wire to a circuit of the device;

[0033] **FIG. 3** is a schematic representation of a stimulation circuit in accordance with the invention;

[0034] **FIG. 4** is a graph showing an output enable pulse from a microcontroller of the stimulation circuit shown in **FIG. 3** in accordance with the invention;

[0035] **FIG. 5** is a graph showing an output signal from the microcontroller of the stimulation circuit shown in **FIG. 3** applied across a resistive load in accordance with the invention.

[0036] **FIG. 6** is a graph showing pulses across the resistive load;

[0037] **FIG. 7** is a schematic representation of an external programming circuit in accordance with the invention;

[0038] **FIG. 8A** is a schematic representation of an external power circuit inductively coupled to a power circuit in accordance with the invention;

[0039] **FIG. 8B** is a schematic representation of an alternative embodiment of the external power circuit non-inductively coupled to the power circuit in accordance with the invention;

[0040] **FIG. 9** is an illustration of a PCB layout of the stimulating circuit in accordance with the invention;

[0041] **FIG. 10** is an illustration of the external power circuit in accordance with the invention/

[0042] **FIG. 11** is an oscilloscope screen showing the output voltage from an oscillator of the external power circuit in accordance with the invention;

[0043] **FIG. 12** is an oscilloscope screen showing a voltage across a primary coil series resistance of the external power circuit in accordance with the invention;

[0044] **FIG. 13** is a graph showing the effect of skin disposed between the primary coil and a secondary coil of the stimulating circuit in accordance with the invention;

[0045] **FIG. 14** is a graph showing the effect of freezing the thawing the skin in accordance with the invention;

[0046] **FIG. 15** is a graph of the output voltage over time in accordance with the invention; and

[0047] **FIG. 16** is a pictoral representation of a model having the stimulation circuit implanted in a scalp and the external powering circuit disposed in a hat in accordance with the invention.

DETAILED DESCRIPTION OF THE INVENTION

[0048] A device for deep brain stimulation using RF energy harvesting **200** of he invention is shown implanted under a human scalp in **FIG. 2A**. A flexible, implantable disc-shaped portion **210** having a diameter of about 6 cm and a thickness of between 3 and 4 mm may be formed of a biocompatible material and include circuitry as further described herein. Lead wires **220** may lead from the circuitry and be coupled to electrodes **230** disposed in targeted areas of the brain. Electrodes **230** may include conventional electrodes used for DBS. Neurostimulation lead securement devices **240** including burr hole caps may serve to secure the lead wires **220** to the electrodes. The circuitry may be operable to harvest and store RF energy, control the operation of the device **200** and provide neurostimulation pulses and signals to the targeted areas of the brain.

[0049] A device for cortical brain stimulation using RF energy harvesting **250** of the invention is shown in **FIG. 2B**. The flexible, implantable disc-shaped portion **210** is shown implanted under the scalp. Lead wires **270** may lead from the circuitry of the disc-shaped portion **260** and be coupled to electrodes **280** disposed on the cortical dura.

[0050] With reference to **FIGS. 2C and 2D**, lead securement devices **240** are shown. Lead securement devices **240** may include StimLoc devices available from ign. Lead securement devices **240** may minimize dislodgment of lead wires **220**.

[0051] Lead wires **220** may be coupled to the circuitry of the disc-shaped portion **210** by means of connectors **215**. Connectors **215** may include a plurality of male contacts **217** for providing electrical contact to corresponding female contacts of the circuitry (not shown). A screw hole **219** may be formed in the connector **215** for securing the connectors **215** to the disc-shaped portion **210** and for securing the disc-shaped portion **210** to the skull of the patient.

[0052] The circuitry may include a stimulation circuit **300** as shown in **FIG. 3** and a portion of the power circuit as shown in **FIG. 8A**. The stimulation circuit **300** may include a circuit printed onto the disc-shaped portion **210**. For purposes of illustration, the stimulation circuit **300** may be modeled using discreet components. The stimulation circuit **300** may include a PIC microcontroller **310** such as the PIC16LF87. The microcontroller **310** may manage the internal stimulation circuitry. A low frequency receiver chip **320** such as the ATA5283 may be coupled to the microcontroller **310** and may convert RF communications into programming commands which the microcontroller **310** interprets. An array of analog switches **330** such as the MAX4066 may be coupled to the microcontroller **310** and connect to voltage dividers **340** to output stimulation locations. Analog switches **330** may be coupled to electrodes **230** (**FIG. 2A**) and **260** (**FIG. 2B**).

[0053] According to internal parameters which can be modified via an external RF programming signal, the microcontroller **310** may control analog switch states to determine a voltage applied to any combination of four output locations including four output locations on electrodes **230** and **260**. The maximum possible voltage is determined by the supply voltage to the circuit **300**. A pulsing frequency, nominally 185 Hz, can be adjusted slightly as well as whether a stimulation pulse is applied or not. In order to conserve energy, the microcontroller **310** may enter a standby mode for 4 ms between pulses, greatly reducing power consumption. The microcontroller **310** may be operated with an internal clock frequency of 125 KHz, giving an efficient tradeoff between power conservation and proper functionality. This clock frequency allows pulse durations in increments of 32 micro-seconds. The output pulse duration can be adjusted between ~60 and ~180 microseconds. With reference to **FIG. 4**, **FIG. 5** and **FIG. 6**, the frequency output, varied voltage output and pulse duration of the microcontroller **310** are shown respectively.

[0054] Every pulsing cycle, the programming input from the low frequency receiver chip **320** may be checked. If a programming signal is present, an input code may be read

sequentially and the specified parameter adjusted to a new value, after which the program continues its pulsing routine.

[**0055**] The low frequency receiver chip **320** used for receiving external programming commands uses an amplitude shift keying (ASK) protocol. The state of a 125 KHz signal being received determines the output voltage of the low frequency receiver chip **320**: on-high, off-low. While waiting for a signal, the low frequency receiver chip **320** may remain in standby mode, conserving power. Upon the presence of a programming signal, the low frequency receiver chip **320** may wake up and send the coded data to the microcontroller **310**, after which the microcontroller **310** may tell the low frequency receiver chip **320** to enter standby mode again. The programming signal may include a preliminary "on" time to wake up the low frequency receiver chip **320**, a 4-bit header, a 3-bit parameter identifier, and a 4-bit data value. Each bit time is 2 milliseconds, allowing enough time for the microcontroller **310** to process the bit reception before the next bit arrives. An antenna attached to a coil input of the low frequency receiver chip **320** may be a short wire having a strong programming signal.

[**0056**] Eight analog switches **330** may be used to control the output pulsing. Four switches **330** may determine a path of the selected voltage to the four possible output locations. Each of these may be controlled by one of the microcontroller outputs, which are in turn enabled or disabled depending on the internal variable for output locations. The inputs of the four switches **330** may be attached to the outputs of the other four switches **330**. The inputs of these four switches **330** may all be attached to different voltage dividers **340**, providing four different voltage levels, ranging from three quarters of the supply voltage to the supply voltage maximum of 3V. Each switch **330** may be controlled by an individual microcontroller signal, which also drives the voltage divider **340** for its particular switch **330**. For every pulse, only one of these four microcontroller signals is active, enabling the voltage from its divider **340** to be sent to the output switches **330** and ultimately to the electrodes **230** and **260**. The use of static voltage dividers **340** to provide output voltage scaling may minimize power consumption. In an alternative embodiment of the invention, a custom digital-analog converter could be used to allow for a higher range of stimulation voltages.

[**0057**] When tested for power consumption, a ~1 Ω resistor was put in series with a powering circuitry described herein. The voltage measured across the resistor while in operation was approximately 17 μV, implying that the DC current required is ~17 μA. At a supply voltage of 3 V, this equates to a power consumption of 51 μW. If operated for 24 hours, the implant would consume a little over 4.4 Joules/day. Typical parameters of a stimulation signal provided for Parkinson's disease are a series of pulses of 120 microsecond duration, 2.5 volts in strength at a repetition rate of 185 pulses per second. Assuming these typical parameters, there are:

185 pulses/second*60 seconds/minute*60 minutes/hour*24 hours/day=1.5984*10^7 pulses/day.

With pulse duration of 120 micro seconds, this gives a total energy application duration of 1.5984*10^7*120*$10^{6=1918.08}$ seconds. With 2.5 volts and 50 micro amps=120 micro watts, the total energy required for stimulation is 120*10^6 watts*1918.08 seconds=0.2302 joules per day. As disclosed herein, energy harvesting by the power circuit is on the order of 12-15 joules per day and the stimulation energy required is more than adequately provided by the power circuit.

[**0058**] An external programmer circuit **700** may include a microcontroller **710** including a PIC16LF87, an inductor/capacitor (LC) oscillating circuit **720** (125 KHz), and an intermediate MOSFET driver **730** including a TC4422 as shown in **FIG. 7**. The MOSFET driver **730** may supply enough energy for driving the LC circuit **720**. When a programming signal is to be sent, a button (not shown) may be pressed, telling the microcontroller **710** to read its inputs and stimulate the MOSFET driver **730** to oscillate the LC circuit **720** according to a communication protocol. Input voltages may be controlled by simple switches. Four switches may dictate the value to be sent, while five switches may dictate which parameter is to be changed. Only one of these switches should be on at one time. A Phidget RFID antenna **740** designed for 125 KHz may be attached to the high voltage side of a capacitor **750** of the LC circuit **720** for sending the programming signal. The circuit **700** may be powered via a 12-Volt wall supply. The 12 V drives the MOSFET driver **730** and is regulated to 5 V for the switches and microcontroller **710**.

[**0059**] An external powering circuit **800** may include a battery **810** for powering an oscillator **820** which drives a transformer-like setup **830** as shown in **FIG. 8A**. The coils **835** on one side of the transformer **830** may be disposed in a cap worn on the head of a patient, a headband worn on the head of the patient, or on a headboard of a bed in which the patient lies. The coils **840** on the other side of the transformer **830** may be coupled to the stimulation circuit **300** and may be disposed proximate the coils **835**. An AC signal coming from coil **840** may be amplified and rectified through a charge pump **850** having three stages, after which a voltage may be clamped with a regulator **860** to prevent spiking. A control circuit **870** may control operation of the voltage regulator **860**.

[**0060**] The oscillator **820** may include an LTC6900. This oscillator **820** produces a 50% duty cycle square wave to drive the primary coil **835** of the transformer **830** and requires only a potentiometer for adjusting the frequency. The charge pump **850** may be a Cockroft Walton voltage multiplier, utilizing a ladder of diodes and capacitors to rectify and amplify the signal. The amplification depends on the number of stages used. Three stages have been found to be enough for a substantial voltage multiplication across a load of 200 KΩ. The capacitors may be 0.1-μF each and the diodes may include BAT54SW surface mount diodes with a forward voltage drop of ~0.24-V. The regulator **860** may include an LT1521-3, which clamps a higher input voltage to 3 V.

[**0061**] Previous empirical testing showed square coils (both primary **835** and secondary **840**), 1 in.×1 in., with 5 turns each are effective for transferring enough energy to power the stimulation circuit **300**. Coils **840** are shown in PCB layout in **FIG. 9** and coils **835** are shown in PCB layout in **FIG. 10**.

[**0062**] The optimal frequency depends on the dielectric and distance between coils **835** and **840**. Frequencies in the range of 2 MHz to 15 MHz may be used. The oscillator **820** can be powered with 3 AAA batteries (4.5 V). In examining

the actual signal through the primary coil **835**, the voltage waveforms in **FIG. 11** and **FIG. 12** were obtained. **FIG. 11** shows the output voltage from the oscillator **820**. **FIG. 12** shows the voltage across the primary coil series resistance, from which the RMS current is calculated to be 29.36 mA_{RMS}.

[**0063**] The embodiment described above provides for near field harvesting and includes inductive coupling between coils **835** and **840**. With reference to **FIG. 8B** and in an alternative embodiment of the invention, the power circuit **865** for powering the stimulation circuit **300** may be non-inductively coupled to an external source of RF energy **880**. In this far field embodiment, the power circuit **865** may be disposed in a wrist band worn by the patient, in a room transmitter or in a transmitter disposed in a building occupied by the patient. In yet another alternative embodiment of the invention, the power circuit **865** may harvest ambient RF energy such as energy transmitted in space by using an inherently tuned antenna as described in U.S. Pat. No. 6,856,291, the description of which is incorporated by reference in its entirety herein. Furthermore, a rechargeable battery or other storage device (not shown) may be employed to store harvested energy. "Non-inductive" as described herein being directed RF.

[**0064**] To demonstrate the effectiveness of the powering and programming schemes through tissue, the device **200** was tested through swine skin. Clear tape was used to cover the conductive surfaces on the primary coil **835** and the secondary coil **840** to prevent interaction with the moisture on the skin. This tape had negligible effect on the inductive coupling.

[**0065**] Three different tests were performed, each following the same procedure. At a certain separation, the voltage powering the primary coil oscillator **820** was adjusted and the maximum output voltage from the secondary coil voltage regulator **860** was measured. The primary coil oscillator voltage started at 5 V and was decreased in increments of 0.1 V until the maximum regulator output voltage had reached a steady minimum.

[**0066**] The first test was performed with no skin between the transformer coils **835** and **840**. Data was acquired at separations of 5 mm, 7 mm, and 10 mm, values chosen based on the common range of human scalp thickness. The second test used fresh swine skin of thicknesses 5 mm and 7 mm between the transformer coils **835** and **840**. The test was interrupted, preventing the testing of 10 mm thick skin. The third test used the same pieces of swine skin, 5, 7, and 10 mm thick, after they had been frozen and thawed.

[**0067**] **FIG. 13** shows the results from the first two tests for comparison purposes. The presence of the skin reduced the inductive coupling between the coils, and hence the possible maximum output voltage in the range of 1.5-3.8 V. At 4.0 V and above, the maximum output voltage of ~3 V is obtainable even with the presence of the skin.

[**0068**] The same pieces of skin were tested a second time due to interruption of the first test. However, they had all been frozen and thawed in the interim, affecting the results slightly. **FIG. 14** shows the effect that the freezing and thawing of the skin had on the energy transfer of the transformer coils. Both the 7- and 10-mm thick pieces of skin reduced the inductive coupling, but the 5-mm skin actually improved in performance. This may be due to the presence of a layer of fat in both the 7- and 10-mm pieces that is absent in the 5-mm piece.

[**0069**] Another test was performed to find the effect of the skin over time. The stimulus for this test was the degrading performance of the 10-mm thick skin over time during the interrupted test mentioned above. For this test, the 7-mm thick piece of skin was used between the primary and secondary coil. The frequency was adjusted to produce a maximum output voltage, which was measured successively over a period of time. The results shown in **FIG. 15** support the fact that performance does not degrade over time. The slight drop in output voltage is likely due to the mechanical nature of the frequency-tuning potentiometer. Notice that the output voltage reaches a steady value and remains constant after that point.

[**0070**] In order to demonstrate the concept and functionality of the device **200**, a model **1600** was created as shown in **FIG. 16**. The stimulation circuit **300** was put on top of a Styrofoam head **1610** with wires running down through the bottom for power- and pulse-monitoring purposes. An ABS Plastic cap (not shown) was made to simulate the head's scalp, covering the stimulation circuit **300**. The primary powering coil **835** with the batteries **810** was secured in a hat **1620** over the position of the stimulation circuit **300** to provide for near field inductive coupling.

[**0071**] The device **200** was tested on a cadaver head to show that a signal may be generated through the scalp and to demonstrate the programmability of the device **200** during stimulation. First, an incision was made in the scalp of the cadaver head. The secondary coil **840** was inserted and placed on the skull and the incision was sewn, leaving the lead wires **220** of the circuit exposed. Six wires were used on the implanted circuitry, four representing the electrodes **230**, which were connected to an oscilloscope and two wires for power and ground. The primary coil **835** was then taped on the scalp directly on top of the implanted circuitry and connected to a power supply.

[**0072**] The experiment began by demonstrating the programmability of the stimulation circuit **300**. Four parameters were varied and displayed on the oscilloscope; pulse width (60, 120 and 180 micro seconds), amplitude (2.34V, 2.75V, 2.94V, 3.13 V), frequency (191 and 194 Hz) and the shifting from one stimulating probe to another (i.e. probe **1** to probe **2** or probe **1** to all four probes). A 10K OHM resistor was used to represent the brain resistance, however this resistance is higher than the resistance for the brain (900 to 1100 Ohms), but a 10 k Ohm resistor was used to ensure there was enough power.

[**0073**] Next, several voltages were tested to determine the output source voltage of the power circuit **865**. Initially, the power supply connected to the primary coil **835** was set at 5 V and was decremented by 0.1 V to 1.2 V. The voltage on the secondary coil **840** was clamped so as not to exceed 3V. As the voltage decreased on the power supply, the output voltage on the secondary coil **840** was steady at 3 V until it declined around 2.2 V. Once the voltage decreased from 3V, a potentiometer was adjusted to obtain the maximum voltage. The data obtained shows the when the voltage drops, the amplitude voltage and frequency drop off as well.

[**0074**] The device for brain stimulation using RF harvesting of the present invention provides a brain stimulation device that requires a single implantation site and surgery to thereby reduce both the cost and trauma to the patient of the implantation procedure. The brain stimulation device further uses RF energy as a power source to eliminate the need for a battery implanted in the pectoral area of the patient. The brain stimulation device further converts RF energy and

stores the converted RF energy for use in stimulation of targeted brain areas. The brain stimulation device is flexible and implantable under the scalp to minimize discomfort for the patient.

[0075] The foregoing description of the embodiments of the invention has been presented for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed. Many modifications and variations are possible in light of the above teaching. It is intended that the scope of the invention be limited not by this detailed description, but rather by the claims appended hereto.

TABLE 1

Indication	Prevalence	Reference for Prevalence
	FDA Approved Indications	
Essential tremor (1997)	1.5–3% population	Am J Med 115:134–42, 2003
Parkinson's (2002)	1% of population > 50	
Dystonia (2004)	~150,000 in US	Mov Disord 3:188–94, 1988
	Current Clinical Trials	
Obsessive-compulsive disorder	2–3% of population	J Clin Psych 53 Suppl: 4–10, 1992
Tourette's syndrome	~1–2% children	J Psychosom Res. 55:3–6, 2003; Can J Neurol Sci Suppl 1:S64–71, 2003
Intractable epilepsy	~100,000 in US	Neurology 56: 1445–52, 2001; Rev Neurol (Paris); 160 Spec No 1:5S31–5, 2004
Intractable depression	1,000,000	Psychiatr Clin North Am 19:179–200, 1996; http://www.mhsource.com/depconsult/june2004.jhtml?__requestid = 605984

We claim:

1. A system for brain stimulation using radio frequency harvesting comprising:

a device implantable under a scalp of a patient, the device comprising a radio frequency harvesting power circuit and a stimulation circuit; and

at least one electrode coupled to the stimulation circuit, the at least one electrode providing brain stimulation to targeted areas of the brain.

2. The system of claim 1, wherein the device is fabricated from a biocompatible substrate.

3. The system of claim 1, wherein the device is flexible and conformable to a shape of the scalp.

4. The system of claim 1, wherein the power circuit comprises a charge pump inductively coupled to a primary coil of an external power circuit.

5. The system of claim 1, wherein the power circuit comprises an inherently tuned antenna for harvesting energy transmitted in space.

6. The system of claim 1, wherein the at least one electrode provides deep brain stimulation.

7. The system of claim 1, wherein the at least one electrode provides cortical stimulation.

8. The system of claim 1, further comprising a programming circuit operable to control the stimulation circuit.

9. The system of claim 8, wherein the programming circuit is operable to control a stimulation circuit voltage output.

10. The system of claim 8, wherein the programming circuit is operable to control a stimulation circuit output pulse width.

11. The system of claim 8, wherein the programming circuit is operable to control a stimulation circuit output frequency.

12. The system of claim 1, further comprising an energy storage device coupled to the device.

13. A system for brain stimulation comprising:

a device implantable under a scalp of a patient, the device comprising a coupled power circuit and a stimulation circuit; and

at least one electrode coupled to the stimulation circuit, the at least one electrode providing brain stimulation to targeted areas of the brain.

14. The system of claim 13, wherein the coupled power circuit is inductively coupled.

15. The system of claim 13, wherein the coupled power circuit is non-inductively coupled.

16. A system for brain stimulation comprising:

a device implantable under a scalp of a patient, the device comprising a power circuit powered by ambient radio frequency energy and a stimulation circuit; and

at least one electrode coupled to the stimulation circuit, the at least one electrode providing brain stimulation to targeted areas of the brain.

17. The system of claim 1, wherein the at least one electrode provides brain stimulation to targeted areas of the brain in response to signals received from the stimulation circuit.

18. A method of providing brain stimulation, comprising:

harvesting power in a power harvesting circuit in a device implantable under the scalp of a patient; and

providing brain stimulation to targeted areas of the brain with at least one electrode connected to the power harvesting circuit.

19. The method of claim 18, wherein the power harvesting circuit harvests radio frequency energy.

20. The method of claim 18, wherein the power harvesting circuit harvests energy by an inductively coupled power circuit in the device.

21. The method of claim 18, wherein the power harvesting circuit harvests energy by a non-inductively coupled power circuit in the device.

22. The method of claim 18, wherein the brain stimulation is used to treat Parkinson's disease.

23. The method of claim 18, wherein the brain stimulation is used to treat stroke patients.

* * * * *

US 20060184209A1

(19) **United States**
(12) **Patent Application Publication**
John et al.
(10) **Pub. No.: US 2006/0184209 A1**
(43) **Pub. Date: Aug. 17, 2006**

(54) DEVICE FOR BRAIN STIMULATION USING RF ENERGY HARVESTING

(76) Inventors: **Constance M. John**, San Francisco, CA (US); **Varghese John**, San Francisco, CA (US); **Martin H. Mickle**, Pittsburgh, PA (US)

Correspondence Address:
BUCHANAN INGERSOLL PC
(INCLUDING BURNS, DOANE, SWECKER & MATHIS)
POST OFFICE BOX 1404
ALEXANDRIA, VA 22313-1404 (US)

(21) Appl. No.: **11/219,404**

(22) Filed: **Sep. 2, 2005**

Related U.S. Application Data

(60) Provisional application No. 60/606,853, filed on Sep. 2, 2004.

Publication Classification

(51) **Int. Cl.**
A61N 1/34 (2006.01)

(52) **U.S. Cl.** **607/45**

(57) **ABSTRACT**

A device for brain stimulation using radio frequency harvesting is disclosed. The device includes a circuit implantable under a scalp of a patient, the circuit comprising a radio frequency harvesting power circuit and a stimulation circuit, and a plurality of electrodes coupled to the circuit, the plurality of electrodes providing brain stimulation to targeted areas of the brain. The electrodes may provide stimulation to targeted areas of the brain including deep brain stimulation for the treatment of Parkinson's disease and cortical stimulation for the treatment of stroke victims.

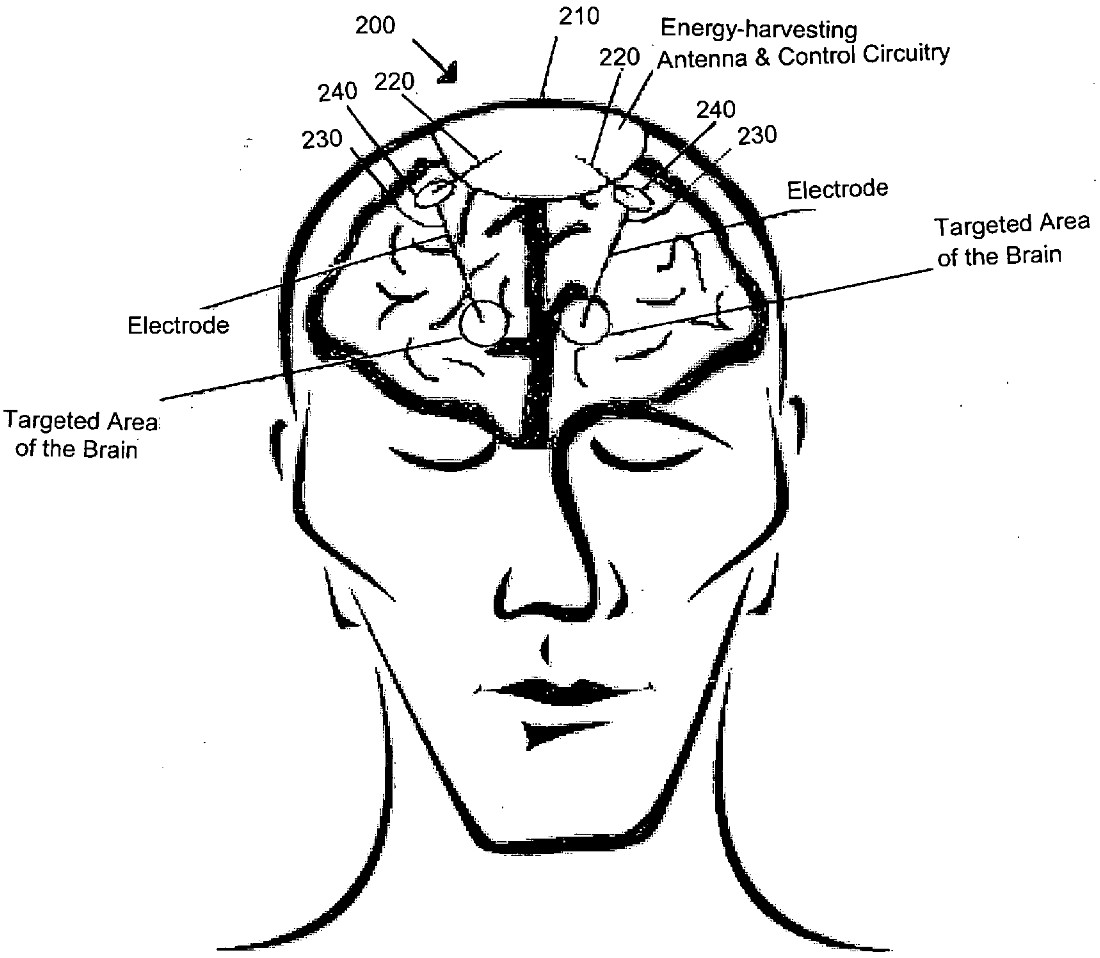
200
210
Energy-harvesting Antenna & Control Circuitry
240
220
220
240
230
230
Electrode
Targeted Area of the Brain
Electrode
Targeted Area of the Brain

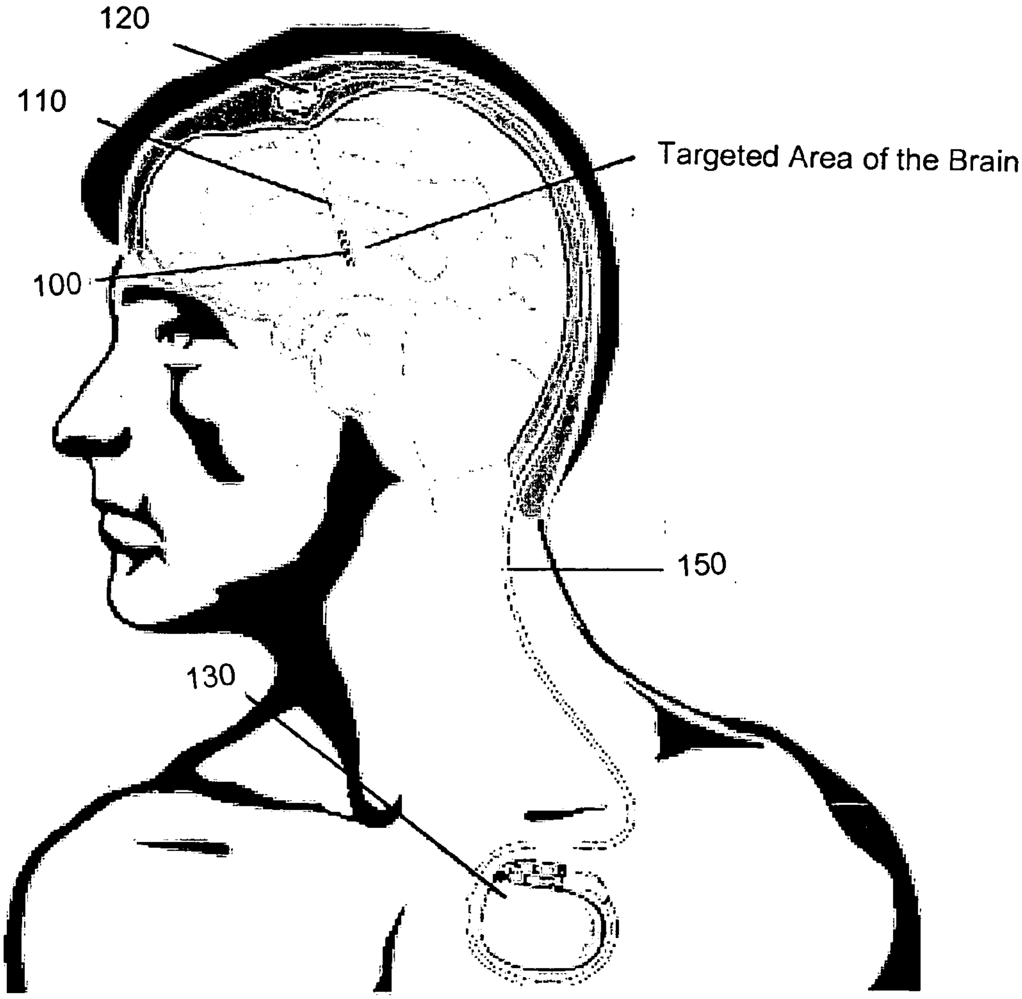
120
110
Targeted Area of the Brain
100
150
130

PRIOR ART

FIG. 1

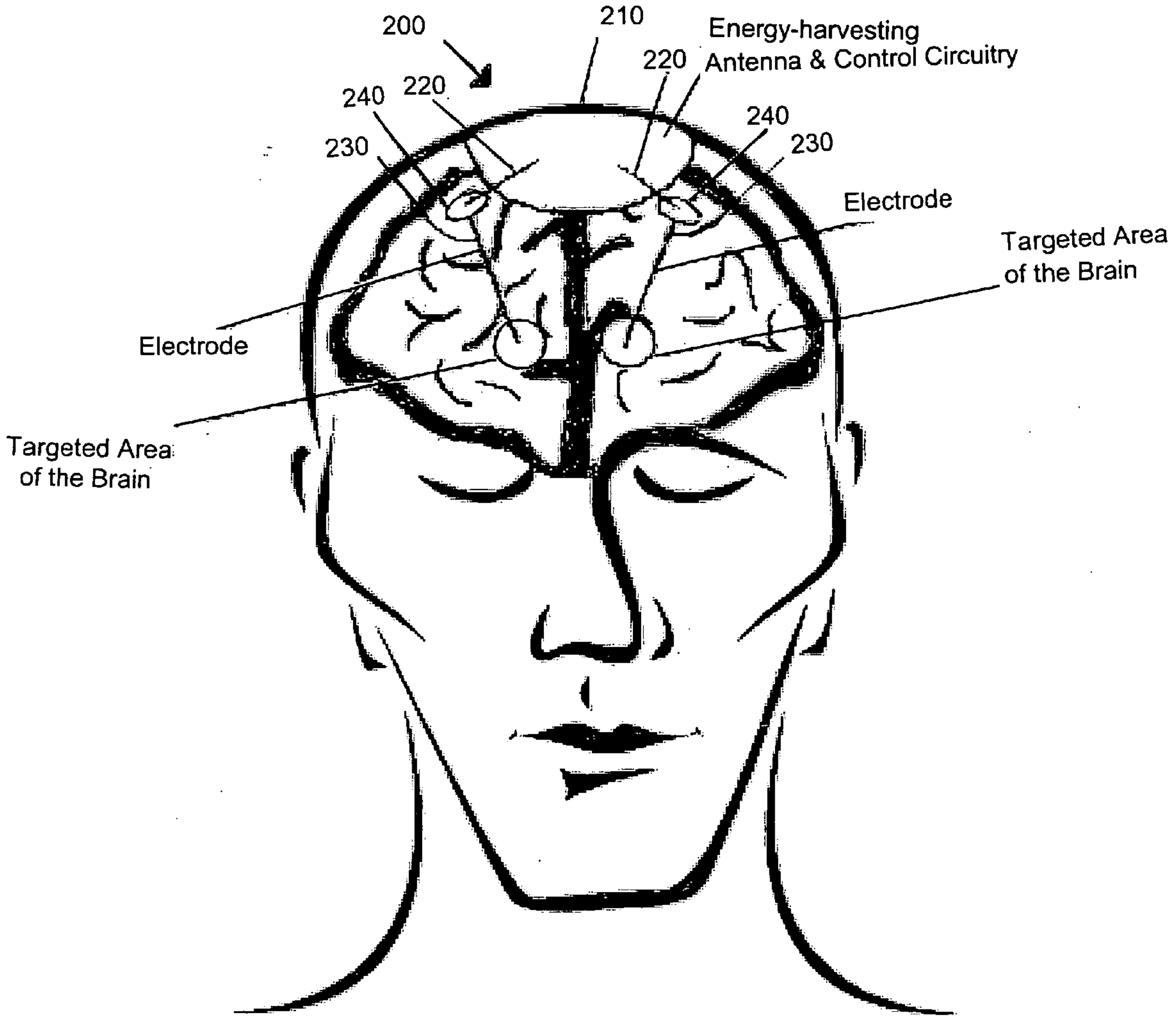
Energy-harvesting Antenna & Control Circuitry
Electrode
Targeted Area of the Brain
Electrode
Targeted Area of the Brain

FIG. 2A

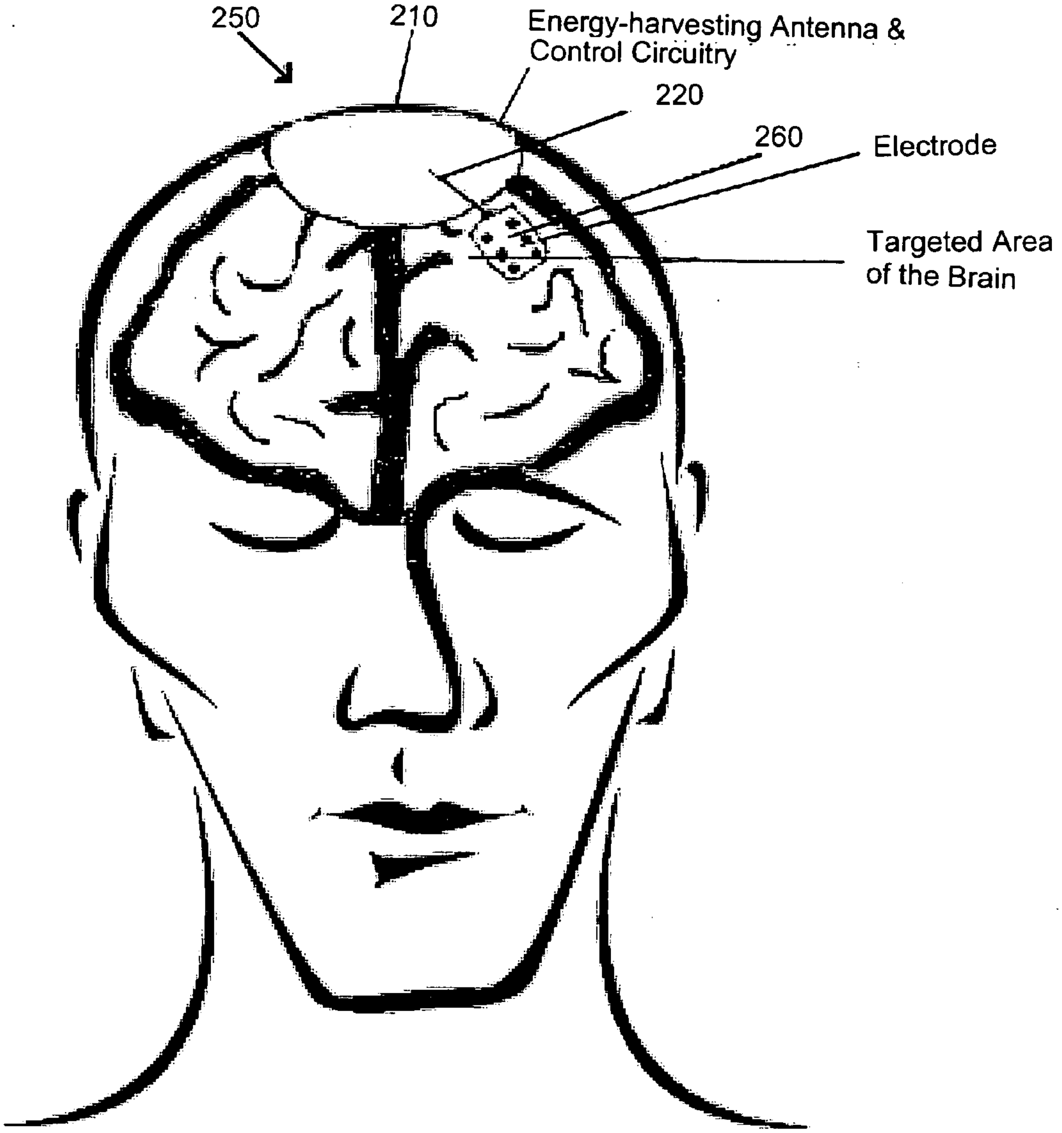
250
210
Energy-harvesting Antenna & Control Circuitry
220
260
Electrode
Targeted Area of the Brain

FIG. 2B

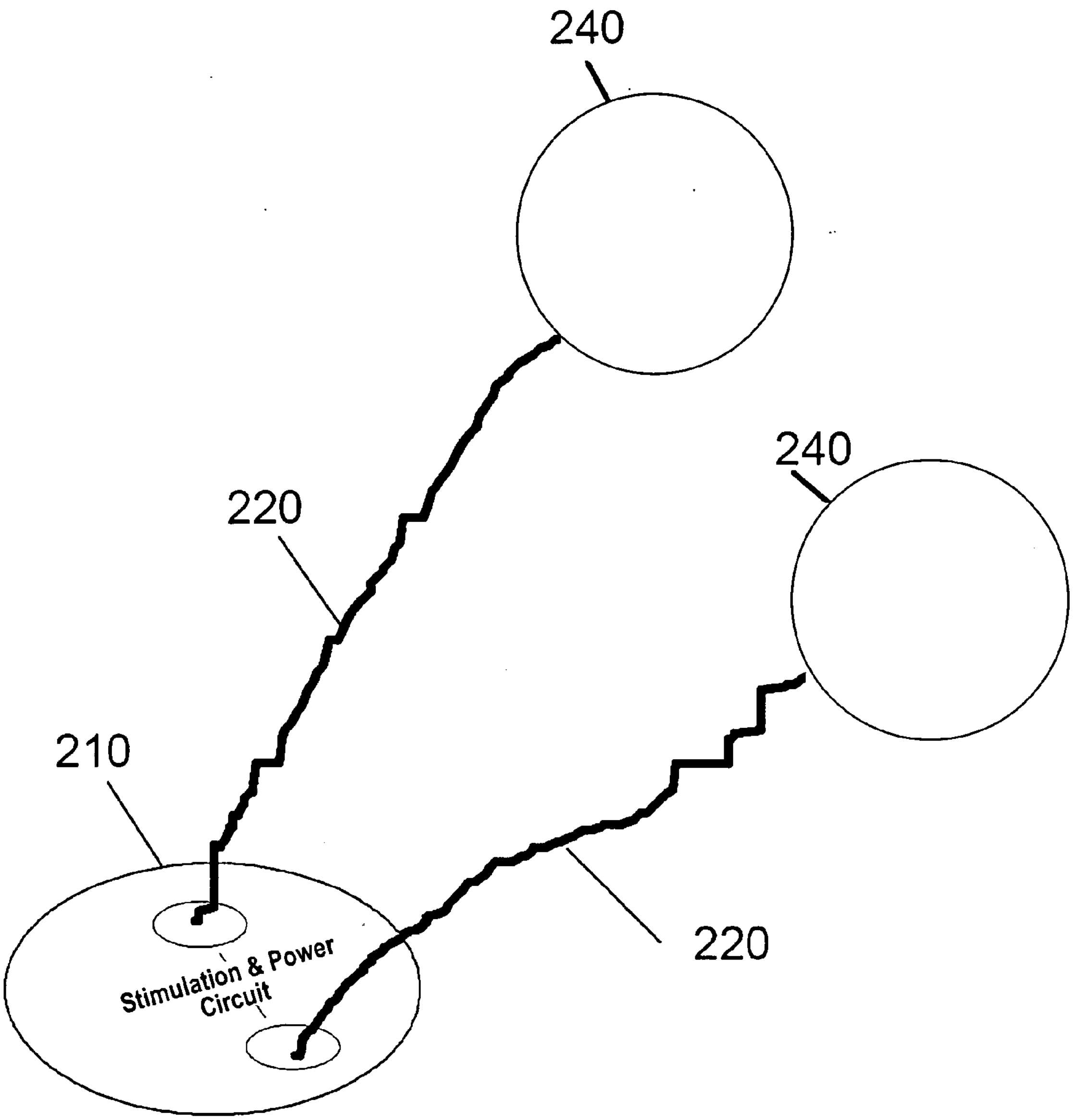
240
240
220
220
210
Stimulation & Power Circuit

FIG. 2C

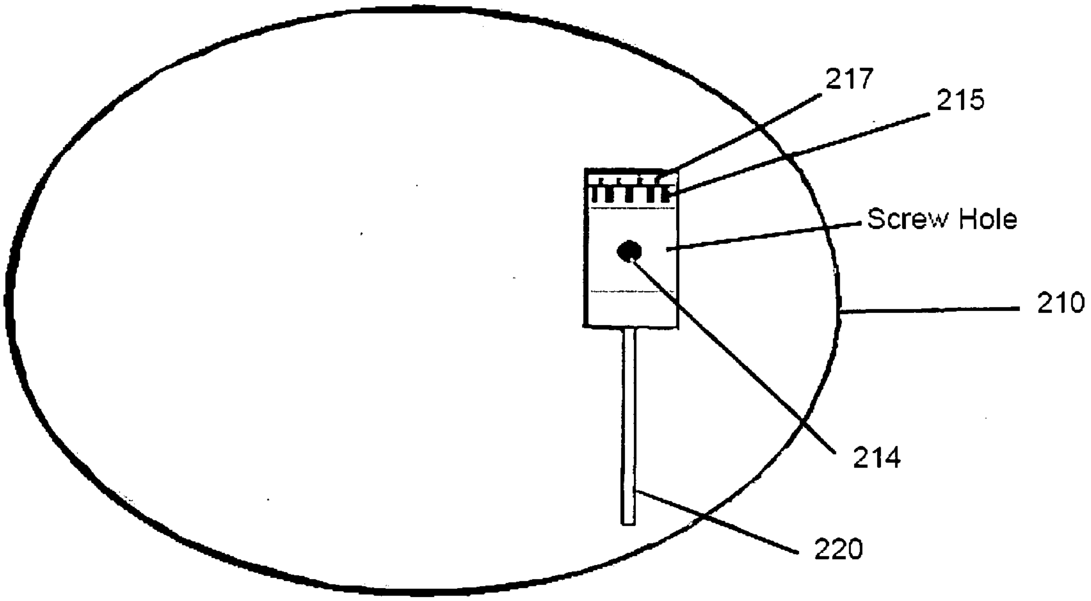
217
215
Screw Hole
210
214
220

FIG. 2D

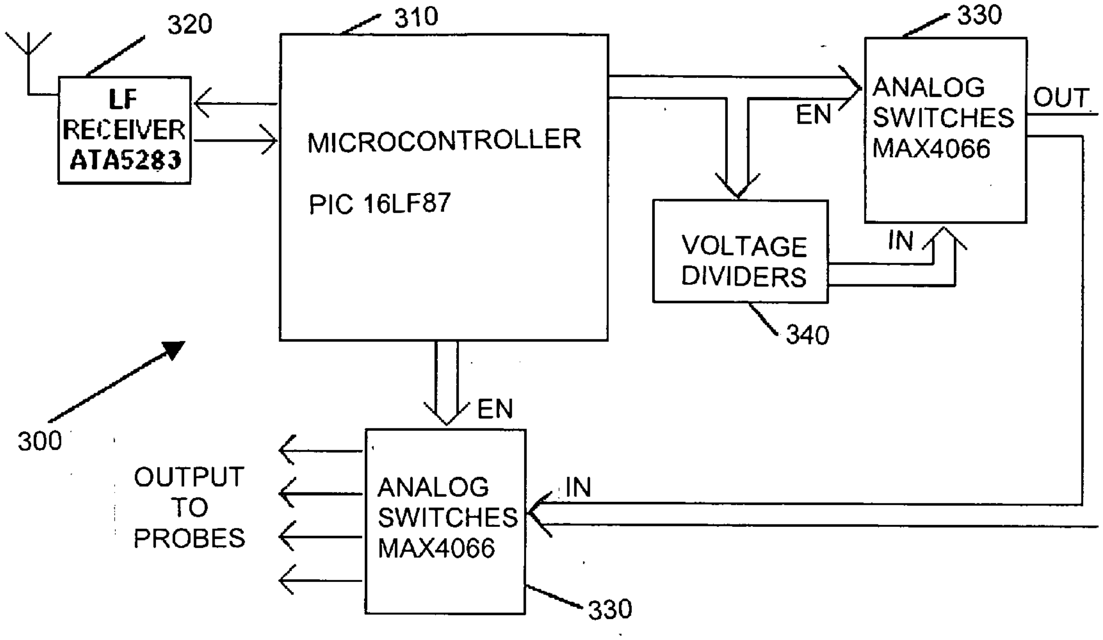
320
LF RECEIVER ATA5283
310
MICROCONTROLLER
PIC 16LF87
330
ANALOG SWITCHES MAX4066
OUT
EN
VOLTAGE DIVIDERS
IN
340
300
EN
OUTPUT TO PROBES
ANALOG SWITCHES MAX4066
IN
330

FIG. 3

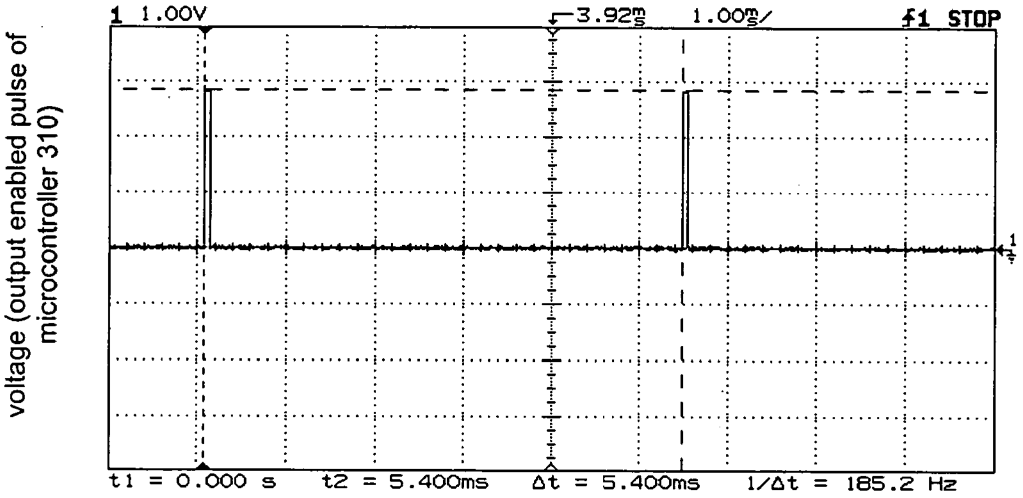
voltage (output enabled pulse of microcontroller 310)
1 1.00V
−3.92ms
1.00ms/
f1 STOP
t1 = 0.000 s
t2 = 5.400ms
Δt = 5.400ms
1/Δt = 185.2 Hz

FIG. 4

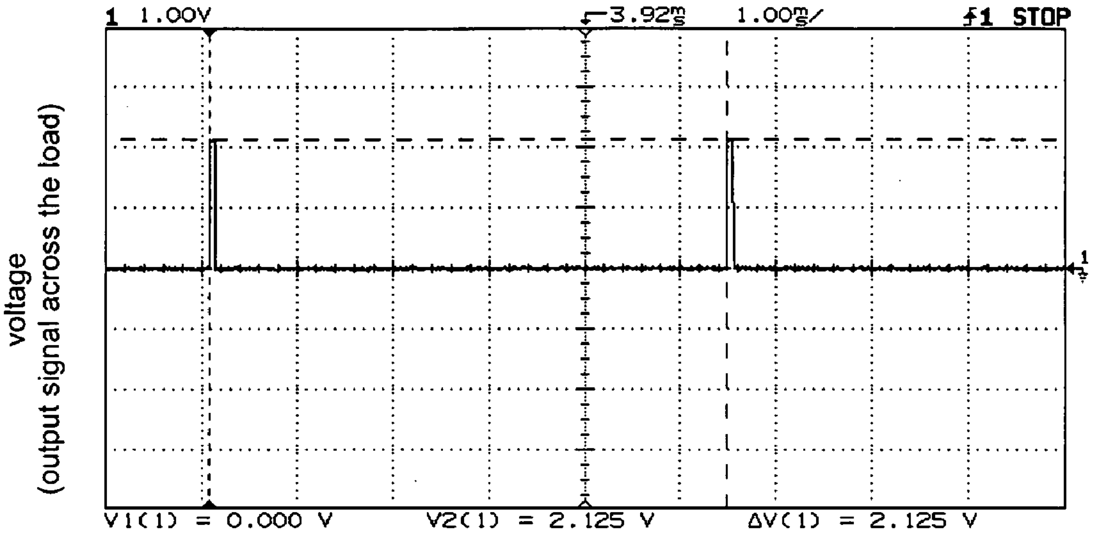
1 1.00V
3.92ms
1.00ms/
f1 STOP
voltage
(output signal across the load)
1
V1(1) = 0.000 V
V2(1) = 2.125 V
ΔV(1) = 2.125 V

FIG. 5

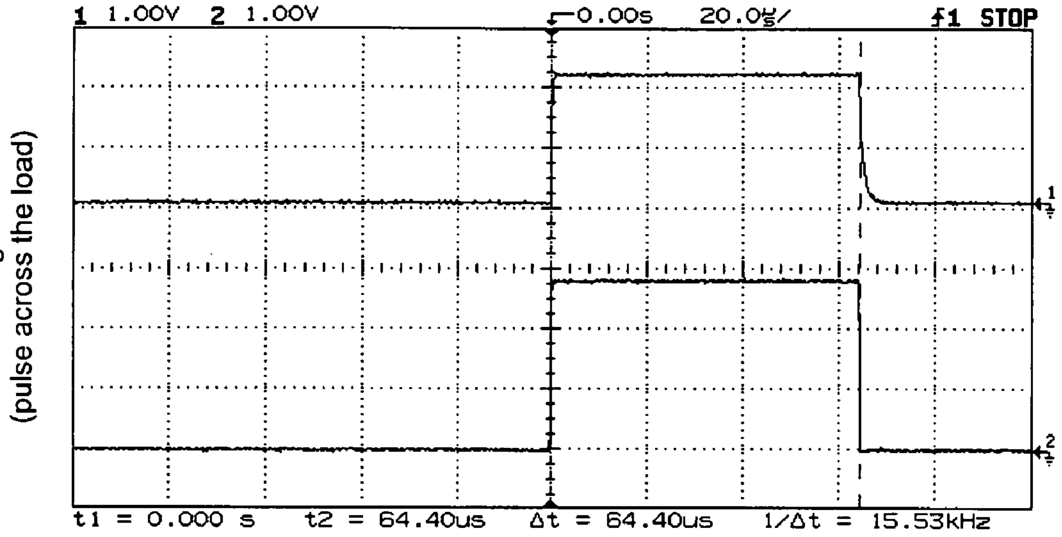
1 1.00V 2 1.00V
0.00s
20.0us/
f1 STOP
1
2
t1 = 0.000 s
t2 = 64.40us
Δt = 64.40us
1/Δt = 15.53kHz
voltage
(pulse across the load)

FIG. 6

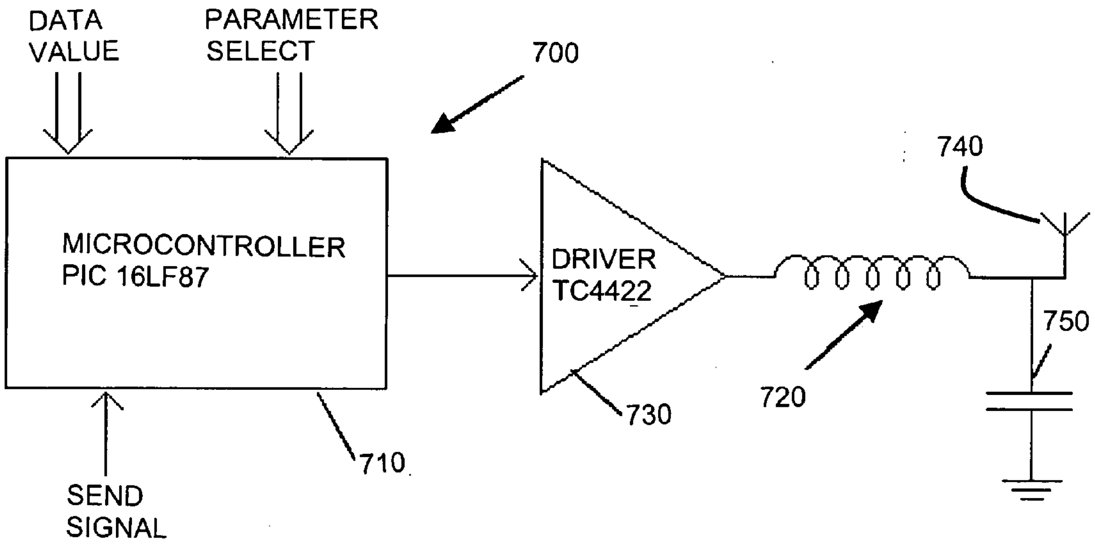
DATA VALUE
PARAMETER SELECT
700
740
MICROCONTROLLER PIC 16LF87
DRIVER TC4422
750
730
720
710
SEND SIGNAL

FIG. 7

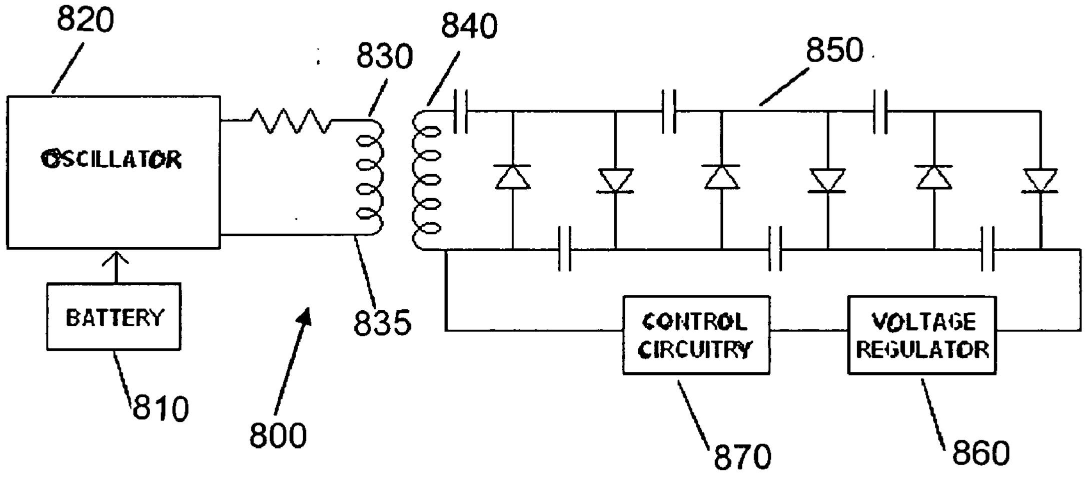
OSCILLATOR
BATTERY
CONTROL CIRCUITRY
VOLTAGE REGULATOR

FIG. 8A

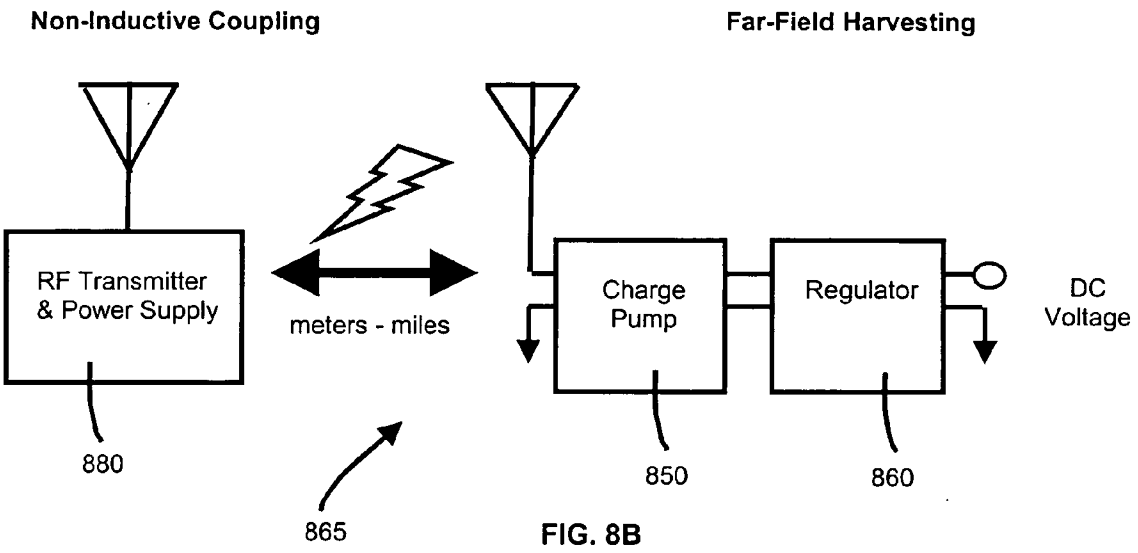
Non-Inductive Coupling
Far-Field Harvesting
RF Transmitter & Power Supply
meters - miles
Charge Pump
Regulator
DC Voltage
880
865
850
860

FIG. 8B

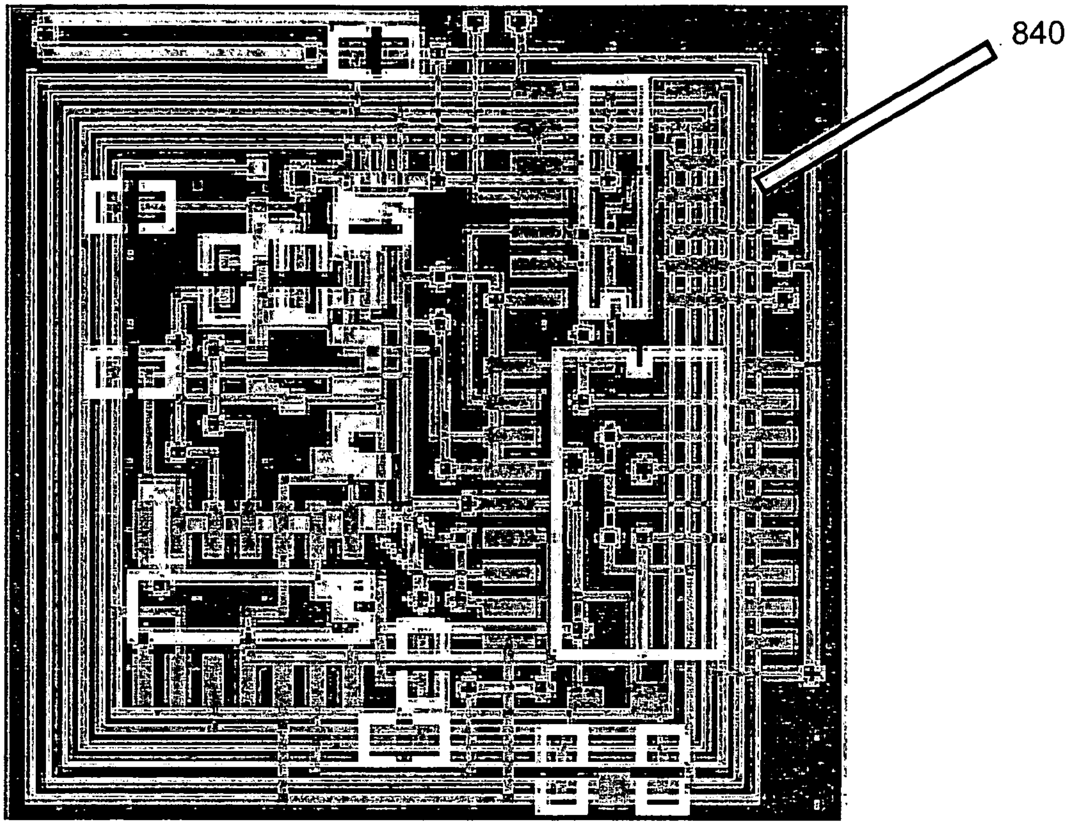

FIG. 9

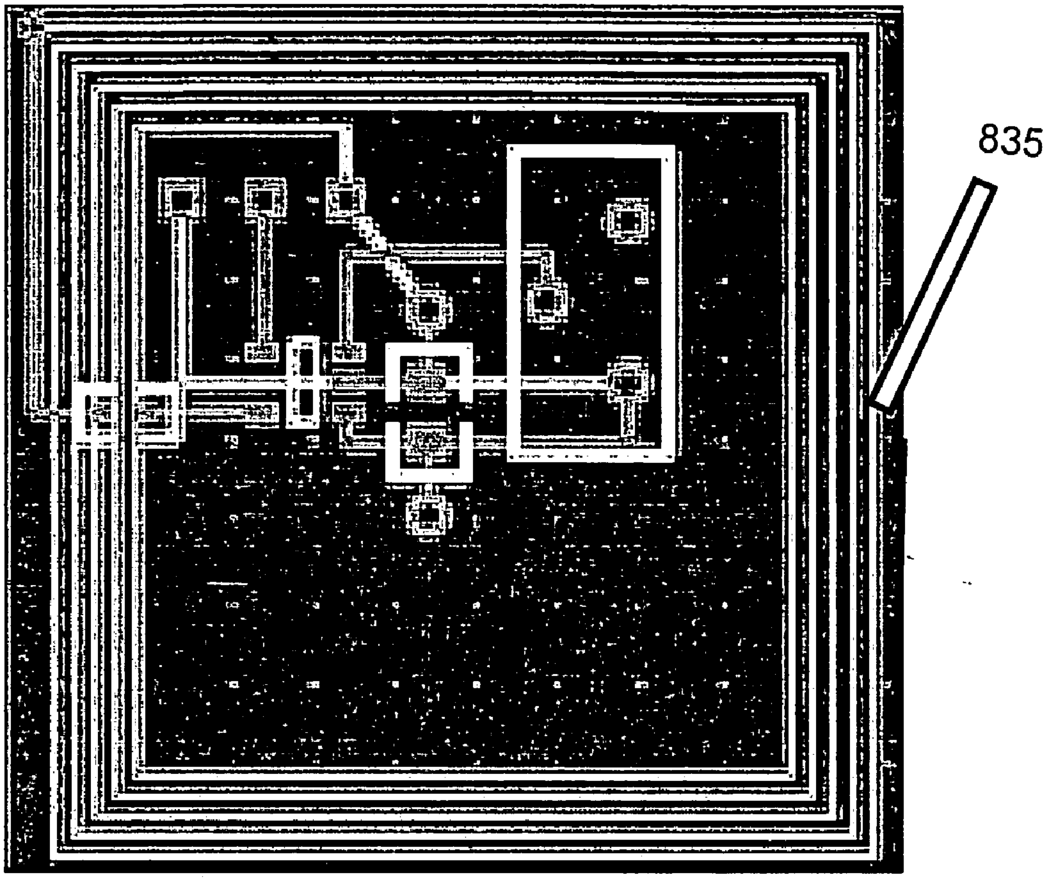

FIG. 10

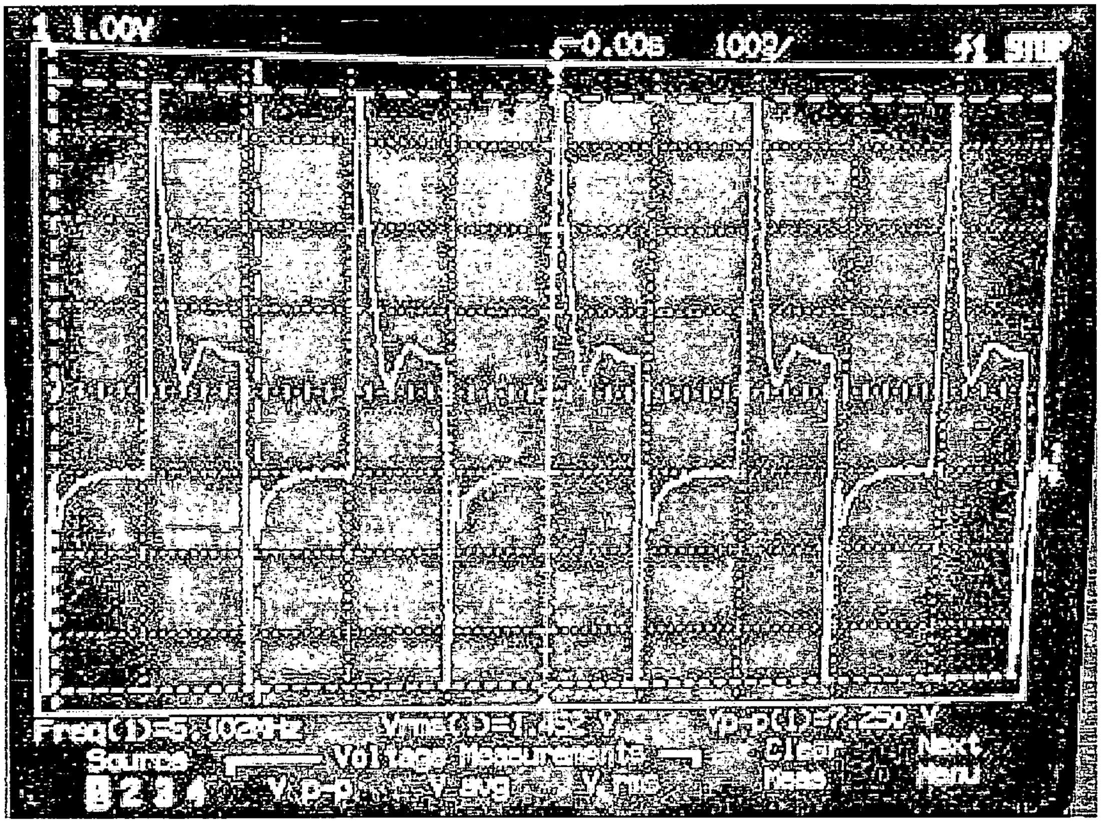

Freq(1)=5.102MHz Vrms(1)=1.452 V Vp-p(1)=7.250 V

FIG. 11

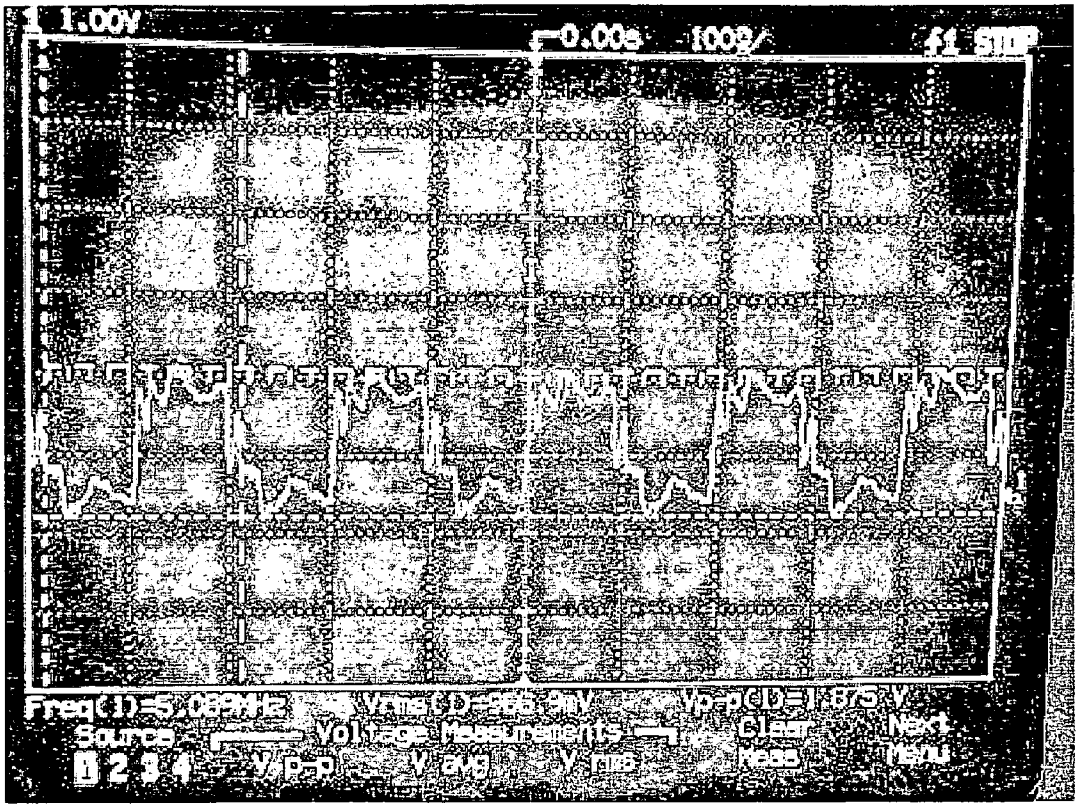
1 1.00V
0.00s 100ns/
f1 STOP
Freq(1)=5.08MHz
Vrms(1)=865.9mV
Vp-p(1)=1.875V
Source
Voltage Measurements
Clear Meas
Next Menu
1 2 3 4
V p-p
V avg
V rms

FIG. 12

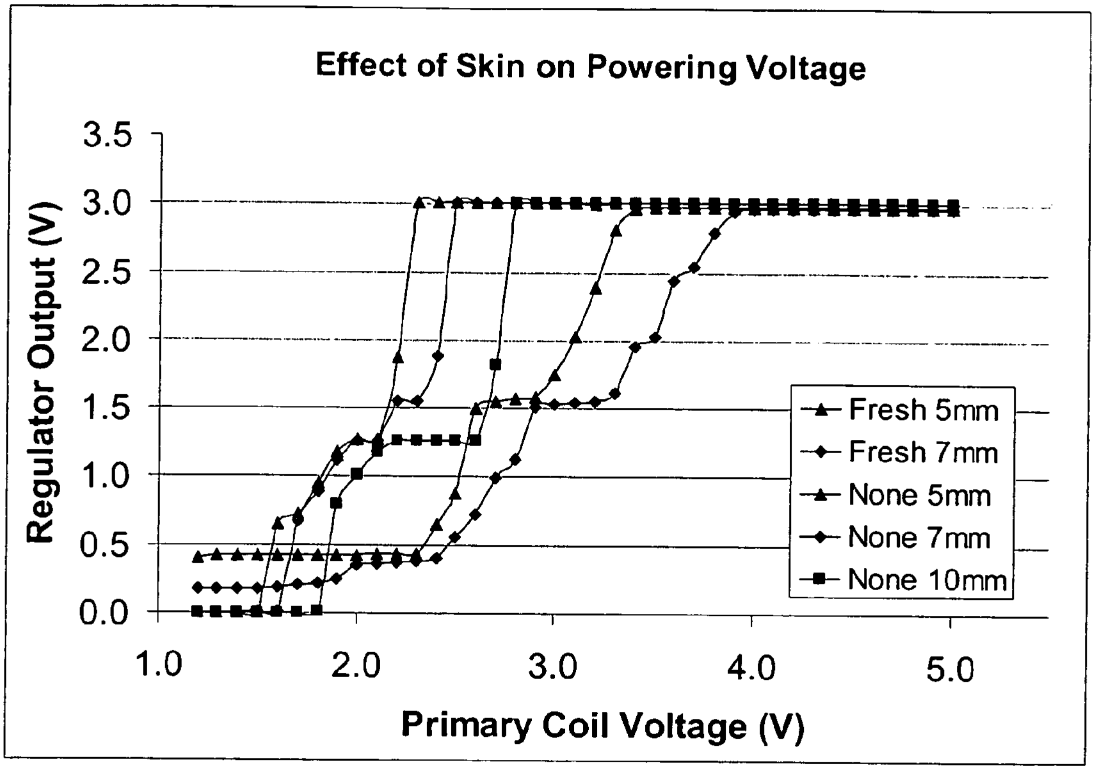
Effect of Skin on Powering Voltage
Regulator Output (V)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1.0
2.0
3.0
4.0
5.0
Primary Coil Voltage (V)
Fresh 5mm
Fresh 7mm
None 5mm
None 7mm
None 10mm

FIG. 13

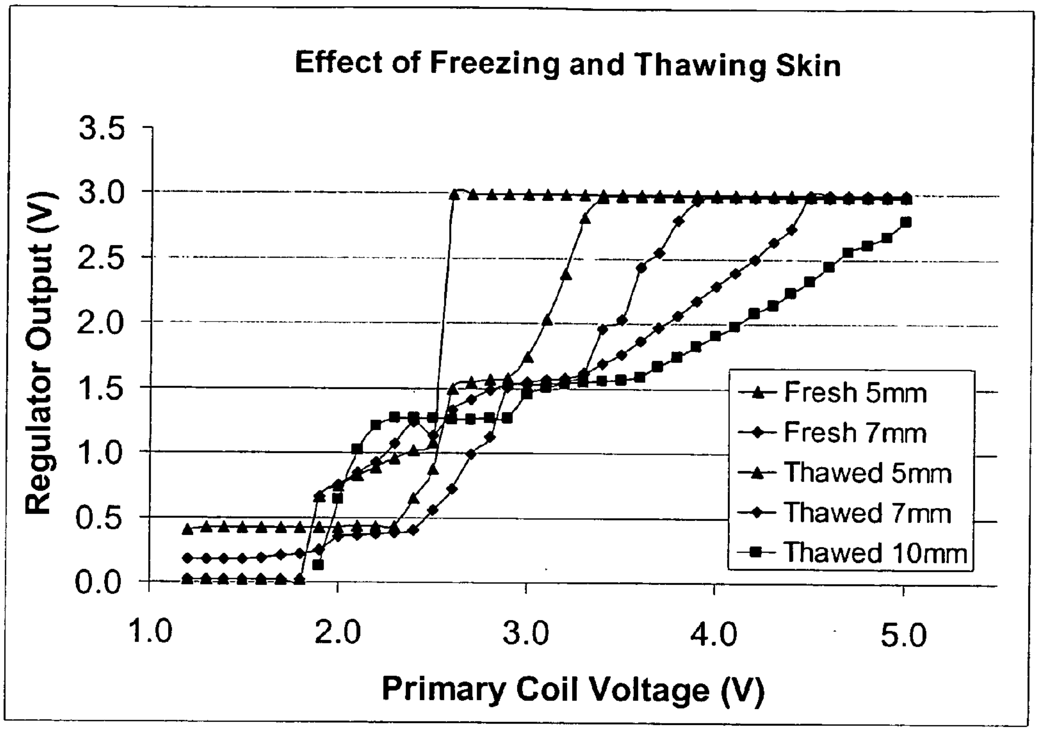
Effect of Freezing and Thawing Skin
Regulator Output (V)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1.0
2.0
3.0
4.0
5.0
Primary Coil Voltage (V)
Fresh 5mm
Fresh 7mm
Thawed 5mm
Thawed 7mm
Thawed 10mm

FIG. 14

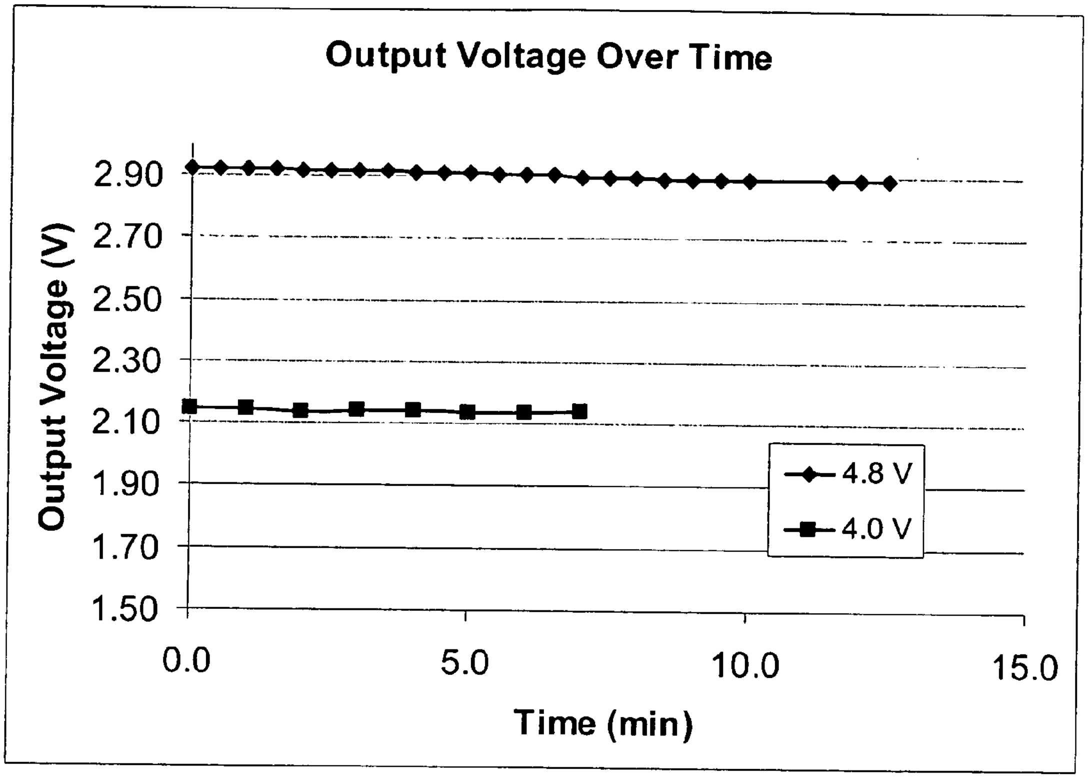
Output Voltage Over Time
Output Voltage (V)
2.90
2.70
2.50
2.30
2.10
1.90
1.70
1.50
0.0
5.0
10.0
15.0
Time (min)
4.8 V
4.0 V

FIG. 15

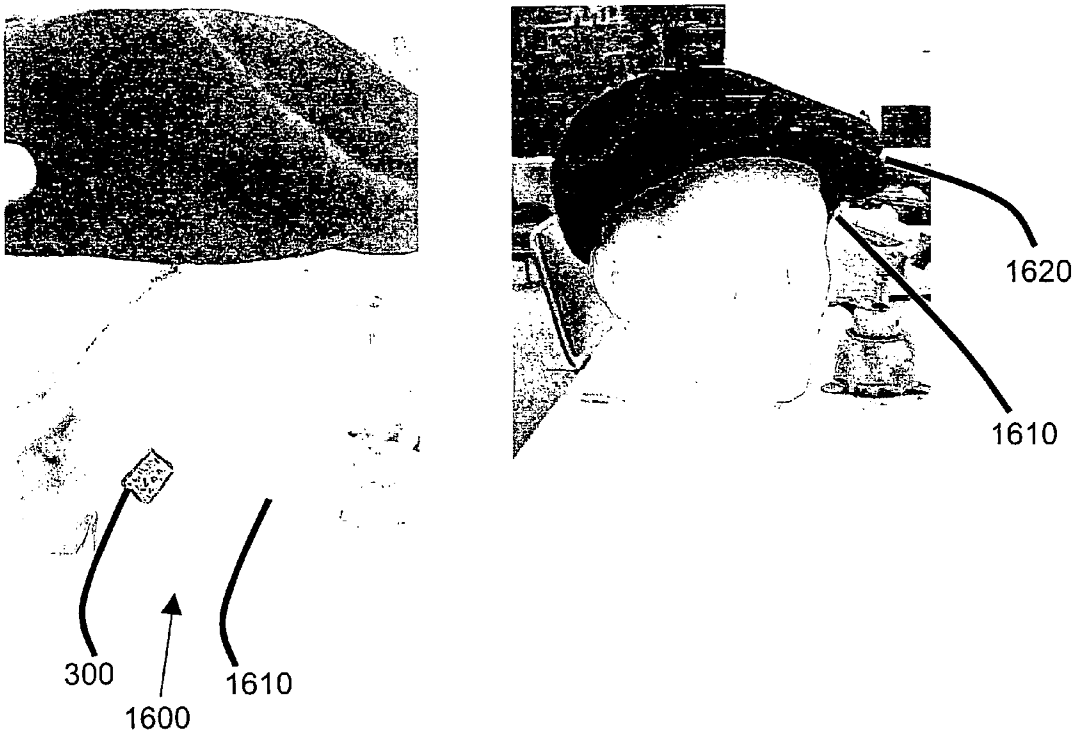
1620
1610
300
1600
1610

FIG. 16

DEVICE FOR BRAIN STIMULATION USING RF ENERGY HARVESTING

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims priority under 35 U.S.C. 119(e) from provisional patent application Ser. No. 60/606,853, entitled "Device For Deep Brain Stimulation (DBS) Using RF Energy Harvesting", filed on Sep. 2, 2004, the disclosure of which is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

[0002] The present invention relates generally to systems and apparatus for providing brain stimulation and more particularly to a device for harvesting radio frequency (RF) energy that can be implanted under a human scalp to produce stimulation in different regions of the brain, including deep brain stimulation (DBS).

BACKGROUND OF THE INVENTION

[0003] DBS is a surgical technique first used in humans over 25 years ago. DBS has been used in a wide variety of brain targets, including the thalamus, globus pallidus and the subthalamic nucleus. Diseases that have been effectively treated with DBS include movement disorders including essential tremor [Lyons K E, Pahwa R. Deep Brain Stimulation and Essential Tremor, J Clin Neurophysiol. 2004 January-February;21(1):2-5], Parkinson's disease [Byrd D L, Marks W J Jr, Starr P A. Deep brain stimulation for advanced Parkinson's disease. AORN J. 2000 September;72(3):387-90, 393-408] and dystonia [Vidailhet M. et al., Bilateral deep-brain stimulation of the globus pallidus in primary generalized dystonia. N Engl J Med. 2005 Feb. 3;352(5):459-67]. Other indications for DBS are being explored, including chronic pain, cluster headache, persistent vegatative state, epilepsy, Alzheimer's, and psychiatric disorders including obsessive-compulsive disorder and intractable depression.

[0004] Parkinson's disease (PD) is an idiopathic neurodegenerative disorder that is characterized by the presence of tremor, rigidity, akinesia or bradykinesia (slowness of movement) and postural instability. It is believed to be caused by the loss of a specific, localized population of neurons in a region of the brain called the substantia nigra. These cells normally produce dopamine, a neurotransmitter that allows brain cells to communicate with each other. These dopaminergic cells in the substantia nigra are part of an elaborate motor circuit in the brain that runs through a series of discrete brain nuclei known as the basal ganglia that control movement. It is believed that the symptoms of PD are caused by an imbalance of motor information flow through the basal ganglia.

[0005] Conventionally, a medication known as levodopa has been the mainstay of treatment for patients with Parkinson's disease. However, long-term therapy with levodopa has several well-known complications that limit the medications effectiveness and tolerability. The first of these is the development of involuntary movements known as dyskinesias. These movements can be violent at times and as or more disabling than the Parkinson's symptoms themselves. The other frequent complication is the development of "on-off" fluctuations, where patients cycle between periods of good function (the "on" period) and periods of poor function (the "off" period). These fluctuations can become very frequent, up to 7 or more cycles per day, and can cause patients to become suddenly and unpredictably "off" to the point where they cannot move.

[0006] Lesioning procedures such as pallidotomy were known to improve the motor symptoms of Parkinson's disease, presumably by disruption of the abnormal neuronal activity in the motor circuitry of the basal ganglia. The discovery that MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) produced a Parkinsonian-like state in non-human primates allowed electrophysiologic study of this phenomenon by numerous investigators. The discovery that high frequency stimulation could mimic the effect of lesioning led to the use of DBS for PD in humans in the early 1990's. DBS was found to improve all of the cardinal symptoms of Parkinson's disease while allowing the patient to decrease or sometimes even eliminate the amount of levodopa medication, therefore decreasing both dyskinesia and "on-off" fluctuations.

[0007] DBS is currently the surgical treatment of choice for medically refractory Parkinson's disease. Two brain targets have been found to provide clinical benefit when chronically stimulated; the subthalamic nucleus (STN) and the internal segment of the globus pallidus (GPi). In a recent prospective, double-blinded cross-over study involving 96 patients with STN DBS and 38 patients with GPi DBS, the STN group reported an improvement in the percentage of time spent during the day with good mobility and without dyskinesia from 27% to 74%. The GPi group also reported a significant improvement, from 28% to 64%.

[0008] Although the mechanism of action is not fully understood it is thought to act by either depolarization blockade, synaptic inhibition, synaptic depression, or stimulation-induced modulation of pathologic network activity [McIntyre C C, Savasta M, Walter B L, Vitek J L. How does deep brain stimulation work? Present understanding and future questions. J Clin Neurophysiol. 2004 January-February;21(1):40-50]. It is believed that DBS acts somehow to suppress the neuronal activity by the stimulation of the region of the brain immediately adjacent to the electrode. This hypothesis seems to be supported by the fact that lesioning a specific structure in the brain has the same clinical effect as stimulating that same structure at high (greater than 100-150 Hz) frequency. In fact, DBS has largely replaced the older lesioning procedures (such as pallidotomy and thalamotomy) that used to be the mainstay of surgical treatment for movement disorders such as Parkinson's disease. The high frequency stimulation may act to hyperpolarize immediately adjacent neurons such that they become incapable of producing normal action potentials. An alternative hypothesis is that DBS may be altering more distant structures or even fibers from far removed nerve cells that are passing through or near the area of stimulation. Whatever the mechanism of action, DBS has a distinct advantage over the older lesioning techniques because it is an adjustable therapy and does not involve destruction of the patient's brain tissue.

[0009] Prior art DBS devices have several limitations that can lead to adverse effects including infection, cutaneous erosion, and lead breaking or disconnection [Temel Y,

Ackermans L, Celik H, Spincemaille G H, Van Der Linden C, Walenkamp G H, Van De Kar T, Visser-Vandewalle V. Management of hardware infections following deep brain stimulation. Acta Neurochir (Wien) 2004;146(4):355-61; Putzke J D, Wharen R E, Jr., Wszolek Z K, Turk M F, Strongosky A J, Uitti R J. Thalamic deep brain stimulation for tremor-predominant Parkinson's disease. Parkinsonism Relat Disord 2003;10(2):81-8.; Umemura A, Jaggi J L, Hurtig H I, Siderowf A D, Colcher A, Stern M B, Baltuch G H. Deep brain stimulation for movement disorders: morbidity and mortality in 109 patients. J Neurosurg 2003;98(4):779-84; Hariz M I. Complications of deep brain stimulation surgery. Mov Disord 2002;17(Suppl 3):S162-**6**]. One study found that 27% of 66 patients with implanted DBS devices had hardware problems [Kondziolka D, Whiting D, Germanwala A, Oh M. Hardware-related complications after placement of thalamic deep brain stimulator systems. Stereotact Funct Neurosurg 2002;79(3-4):228-33. This relatively high incidence of hardware problems is similar to the results of a study where 20 (25.3%) of 79 patients who received 124 permanent DBS electrode implants had 26 hardware-related complications [Oh M Y, Abosch A, Kim S H, Lang A E, Lozano A M. Long-term hardware-related complications of deep brain stimulation. Neurosurgery 2002;50(6):1268-74; discussion 1274-**6**]. In addition, intracranial electrode implantation can induce a hematoma or contusion. Nonetheless, most authors agree that the benefit to risk ratio of DBS is favorable.

[0010] A prior art DBS device is shown in **FIG. 1** and includes an electrode **100** disposed in a targeted area of the brain. The electrode is coupled to a lead **110** held in place at the top of the skull by a securement device **120**. The lead **110** is coupled to a neurostimulator **130** powered by a battery-powered pulse generator **140** by means of a lead **150**. The lead **150**, which averages about 15 inches in length, is implanted under the scalp and traverses the length of the patient's neck to the chest where the neurostimulator **130** and battery **140** are implanted.

[0011] The pulse generator **140** is typically placed underneath the skin just below the collar bone and is capable of stimulating at one or any combination of the four contacts present on the end of the electrode **110** in the brain. The parameters of the stimulating current (voltage, frequency, pulse width) can also be selected by the treating physician or health care worker. The pulse generator **140** is programmed through the skin via a telemetry device that allows the practitioner to select the desired stimulation parameters and also perform diagnostic tests on the device.

[0012] Implantation of the conventional DBS device is costly as for implantation of a single electrode in the brain for treatment of one side of the body the procedure requires three incisions; one on the top of the head, one behind the ear and the third just below the collarbone where the leads are connected. The implantation of the electrode **110** and the implantable pulse generator **140** is sometimes performed on different days. The incisions can be prone to infection in the immediate postoperative period. In some elderly patients with thin skin, the pulse generator **140** or wire can erode through the skin and become exposed to potential contamination. Infection or erosion often results in the need to remove the entire device, as antibiotic treatment alone in this setting rarely will clear the infection adequately. The lead **150** restricts the patient's mobility in the neck region and may break. Furthermore, the battery **140** must be replaced every three to five years. Additional drawbacks of the DBS device include the risk of erosion of the leads or hardware, infection, and magnetic sensitivity.

[0013] A prior art deep brain stimulation system is disclosed in U.S. Pat. No. 6,920,359 entitled "Deep Brian Stimulation System for the Treatment of Parkinson's Disease or Other Disorders". The DBS system includes a small, implantable pulse generator implanted directly in the cranium of the patient, thereby eliminating the long lead wires conventionally used. The disclosed system does not provide for the harvesting of energy to power the pulse generator.

[0014] As noted in Table 1, there are several current and potential indications for deep brain stimulation.

[0015] Known systems for providing electrical stimulus to the motor cortex of the brain, such as the Northstar Stroke Recovery Treatment System available from Northstar Neuroscience, Inc., also include an implantable pulse generator implanted in the pectoral area of a patient. A cortical stimulation lead includes an electrode connected to the implantable pulse generator which is used to deliver stimulation to the cortex. The electrode is placed on top of the dura and coupled to the implantable pulse generator by means of a lead which traverses the length of the patient's neck to the patient's pectoral area.

[0016] Motor cortex stimulation (MCS) is a process involving the application of stimulation signals to the motor cortex in the brain of a patient during physical rehabilitation of the disabled region of the body. The MCS system includes a pulse generator connected to a strip electrode that is surgically implanted over a portion of only the motor cortex (precentral gyrus). Because MCS involves the application of stimulation signals to surface regions of the brain rather than deep neural structures, electrode implantation procedures for MCS are significantly less invasive and time consuming than those for DBS. The current evaluation of MCS is for stroke. Stroke-related disabilities affect more than 200,000 people in the U.S. each year. Good results have been reported in MCS treatment of stroke victims. With a MCS device, a stamp-sized electrode is placed on the surface of the brain. This is attached to a wire that goes through the neck to an implantable pulse generator in the pectoral area.

[0017] Neurostimulation and responsive neurostimulation (RNS) are also being tested for the treatment of medically refractory epilepsy. In treating epilepsy, the RNS system can be designed to detect abnormal electrical activity in the brain and respond by delivering electrical stimulation to normalize brain activity before the patient experiences seizure symptoms. For either neurostimulation or RNS for treatment of epilepsy the electrode or electrodes of the device deliver a short train of electrical pulses to the brain near the patient's seizure focus.

[0018] In order to obviate the need for long leads and batteries, attempts have been made in the prior art to transmit energy through space from a base station to a remote station. One such system is disclosed in U.S. Pat. No. 6,289,237 entitled "Apparatus for Energizing a Remote Station and Related Method". The base station transmits energy which may be RF power, light, acoustic, magnetic or other suitable forms of space transmitted or "radiant" energy to the remote station. Within the remote station, the received

energy is converted into DC power which serves to operate the remote station. The source of power for the remote station is the base station and, therefore, there is no need for the remote station to carry an electrical storage device such as a battery. It is suggested that this facilitates the remote station being encapsulated within a suitable protective material, such as a resinous plastic. Homopolymers, elastomers and silicon dioxide are also suggested as suitable materials for such purposes. Further, it is suggested that this facilitates miniaturization of the remote station and placing the remote station in functionally desirable locations which need not be readily accessible. The remote station, for example, could be implanted in a patient.

[0019] The use of a wireless communication link between a base station and transponders in a radio frequency identification system employing modulated back-scattered waves is also known. See Rao, An Overview of Bulk Scattered Radio Frequency Identification System (RFID) IEEE (1999). It has also been suggested to employ a silicon chip in a transponder having a change pump on voltage doubler current. Hornby, RFID Solutions for the Express Parcel and Airline Baggage Industry, Texas Instruments, Limited (Oct. 7, 1999).

[0020] For use in miniaturized electronic chip systems, an electronic article containing a microchip having at least one antenna structured to communicate with an antenna remotely disposed with respect to the microchip is disclosed in U.S. Pat. No. 6,615,074 entitled "Apparatus for Energizing a Remote Station and Related Method". Power enhancement is achieved using a voltage doubler. The antenna of the disclosed apparatus is comparable in volume to a Smart Dust device. Smart Dust is a combination MEMS/Electronic device on the order of 1 mm×1 mm×1 mm.

[0021] What is needed therefore is a brain stimulation device that overcomes the disadvantages of the prior art brain stimulation devices. What is needed is a brain stimulation device that requires a single implantation site and surgery. What is also needed is a brain stimulation device that uses RF energy as a power source. What is further needed is a brain stimulation device that converts RF energy and stores the converted RF energy. What is also needed is a brain stimulation device that is flexible and implantable under the scalp. What is needed is a brain stimulation device that does not require leads or a pulse generator to be placed outside of the head area that are subject to disconnection or breakage. What is also needed is a brain stimulation device for electrical stimulation in the brain that is smaller and more self-contained and that does not require a pulse generator to be implanted elsewhere in the body. What is further needed is a device that is less susceptible to hardware problems or complications. What is needed is a device that has less potential for erosion through the skin. What is also needed is a device that is has a power source that does not need to be replaced.

SUMMARY OF INVENTION

[0022] The device for brain stimulation using RF energy harvesting of the present invention overcomes the disadvantages of the prior art, fulfills the needs in the prior art, and accomplishes its various purposes by providing a brain stimulation device that harvests radio frequency energy and is implantable under the scalp. The brain stimulation device of the invention may include an electrode that penetrates into the brain to provide neurostimulation to the brain. The brain stimulation device may also include an electrode that is used to provide stimulation to the brain cortex.

[0023] In accordance with one aspect of the invention, a device for brain stimulation using radio frequency harvesting includes a circuit implantable under a scalp of a patient, the circuit comprising a radio frequency harvesting power circuit and a stimulation circuit, and a plurality of electrodes coupled to the circuit, the plurality of electrodes providing brain stimulation to targeted areas of the brain. An advantage of this system is that it may use "trickle charging" wherein the device is charged by the harvesting power circuit. Moreover, anther advantage of this invention is the power transmitter which sends power to the device can be used both to send power and to send information.

[0024] There has been outlined, rather broadly, the more important features of the invention in order that the detailed description thereof that follows may be better understood, and in order that the present contribution to the art may be better appreciated. There are, of course, additional features of the invention that will be described below and which will form the subject matter of the claims appended herein.

[0025] In this respect, before explaining at least one embodiment of the invention in detail, it is to be understood that the invention is not limited in its application to the details of design and to the arrangement of components set forth in the following description or illustrated in the drawings. The invention is capable of other embodiments and of being practiced and carried out in various ways. Also, it is to be understood that the phraseology and terminology employed herein, as well as the abstract, are for the purpose of description and should not be regarded as limiting.

[0026] As such, those skilled in the art will appreciate that the conception upon which this disclosure is based may readily be utilized as a basis for the designing of other methods and systems for carrying out the several purposes of the present invention. It is important, therefore, that the claims be regarded as including such equivalent methods and systems insofar as they do not depart from the spirit and scope of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0027] The accompanying drawings, which are incorporated in and form a part of the specification, illustrate the embodiments of the present invention and together with the description, serve to explain the principles of the invention. In the drawings:

[0028] **FIG. 1** is a schematic representation of a prior art DBS device;

[0029] **FIG. 2A** is a schematic representation of a device for deep brain stimulation using RF energy harvesting in accordance with the invention;

[0030] **FIG. 2B** is a schematic representation of a device for cortical stimulation using RF energy harvesting in accordance with the invention;

[0031] **FIG. 2C** is a schematic representation of the device of **FIG. 2A** illustrating lead securement devices;

[0032] **FIG. 2D** is a schematic representation of the device of **FIG. 2A** illustrating an attachment means for connecting a lead wire to a circuit of the device;

[0033] **FIG. 3** is a schematic representation of a stimulation circuit in accordance with the invention;

[0034] **FIG. 4** is a graph showing an output enable pulse from a microcontroller of the stimulation circuit shown in **FIG. 3** in accordance with the invention;

[0035] **FIG. 5** is a graph showing an output signal from the microcontroller of the stimulation circuit shown in **FIG. 3** applied across a resistive load in accordance with the invention.

[0036] **FIG. 6** is a graph showing pulses across the resistive load;

[0037] **FIG. 7** is a schematic representation of an external programming circuit in accordance with the invention;

[0038] **FIG. 8A** is a schematic representation of an external power circuit inductively coupled to a power circuit in accordance with the invention;

[0039] **FIG. 8B** is a schematic representation of an alternative embodiment of the external power circuit non-inductively coupled to the power circuit in accordance with the invention;

[0040] **FIG. 9** is an illustration of a PCB layout of the stimulating circuit in accordance with the invention;

[0041] **FIG. 10** is an illustration of the external power circuit in accordance with the invention/

[0042] **FIG. 11** is an oscilloscope screen showing the output voltage from an oscillator of the external power circuit in accordance with the invention;

[0043] **FIG. 12** is an oscilloscope screen showing a voltage across a primary coil series resistance of the external power circuit in accordance with the invention;

[0044] **FIG. 13** is a graph showing the effect of skin disposed between the primary coil and a secondary coil of the stimulating circuit in accordance with the invention;

[0045] **FIG. 14** is a graph showing the effect of freezing the thawing the skin in accordance with the invention;

[0046] **FIG. 15** is a graph of the output voltage over time in accordance with the invention; and

[0047] **FIG. 16** is a pictoral representation of a model having the stimulation circuit implanted in a scalp and the external powering circuit disposed in a hat in accordance with the invention.

DETAILED DESCRIPTION OF THE INVENTION

[0048] A device for deep brain stimulation using RF energy harvesting **200** of he invention is shown implanted under a human scalp in **FIG. 2A**. A flexible, implantable disc-shaped portion **210** having a diameter of about 6 cm and a thickness of between 3 and 4 mm may be formed of a biocompatible material and include circuitry as further described herein. Lead wires **220** may lead from the circuitry and be coupled to electrodes **230** disposed in targeted areas of the brain. Electrodes **230** may include conventional electrodes used for DBS. Neurostimulation lead securement devices **240** including burr hole caps may serve to secure the lead wires **220** to the electrodes. The circuitry may be operable to harvest and store RF energy, control the operation of the device **200** and provide neurostimulation pulses and signals to the targeted areas of the brain.

[0049] A device for cortical brain stimulation using RF energy harvesting **250** of the invention is shown in **FIG. 2B**. The flexible, implantable disc-shaped portion **210** is shown implanted under the scalp. Lead wires **270** may lead from the circuitry of the disc-shaped portion **260** and be coupled to electrodes **280** disposed on the cortical dura.

[0050] With reference to **FIGS. 2C and 2D**, lead securement devices **240** are shown. Lead securement devices **240** may include StimLoc devices available from ign. Lead securement devices **240** may minimize dislodgment of lead wires **220**.

[0051] Lead wires **220** may be coupled to the circuitry of the disc-shaped portion **210** by means of connectors **215**. Connectors **215** may include a plurality of male contacts **217** for providing electrical contact to corresponding female contacts of the circuitry (not shown). A screw hole **219** may be formed in the connector **215** for securing the connectors **215** to the disc-shaped portion **210** and for securing the disc-shaped portion **210** to the skull of the patient.

[0052] The circuitry may include a stimulation circuit **300** as shown in **FIG. 3** and a portion of the power circuit as shown in **FIG. 8A**. The stimulation circuit **300** may include a circuit printed onto the disc-shaped portion **210**. For purposes of illustration, the stimulation circuit **300** may be modeled using discreet components. The stimulation circuit **300** may include a PIC microcontroller **310** such as the PIC16LF87. The microcontroller **310** may manage the internal stimulation circuitry. A low frequency receiver chip **320** such as the ATA5283 may be coupled to the microcontroller **310** and may convert RF communications into programming commands which the microcontroller **310** interprets. An array of analog switches **330** such as the MAX4066 may be coupled to the microcontroller **310** and connect to voltage dividers **340** to output stimulation locations. Analog switches **330** may be coupled to electrodes **230** (**FIG. 2A**) and **260** (**FIG. 2B**).

[0053] According to internal parameters which can be modified via an external RF programming signal, the microcontroller **310** may control analog switch states to determine a voltage applied to any combination of four output locations including four output locations on electrodes **230** and **260**. The maximum possible voltage is determined by the supply voltage to the circuit **300**. A pulsing frequency, nominally 185 Hz, can be adjusted slightly as well as whether a stimulation pulse is applied or not. In order to conserve energy, the microcontroller **310** may enter a standby mode for 4 ms between pulses, greatly reducing power consumption. The microcontroller **310** may be operated with an internal clock frequency of 125 KHz, giving an efficient tradeoff between power conservation and proper functionality. This clock frequency allows pulse durations in increments of 32 micro-seconds. The output pulse duration can be adjusted between ~**60** and ~180 microseconds. With reference to **FIG. 4**, **FIG. 5** and **FIG. 6**, the frequency output, varied voltage output and pulse duration of the microcontroller **310** are shown respectively.

[0054] Every pulsing cycle, the programming input from the low frequency receiver chip **320** may be checked. If a programming signal is present, an input code may be read

sequentially and the specified parameter adjusted to a new value, after which the program continues its pulsing routine.

[0055] The low frequency receiver chip **320** used for receiving external programming commands uses an amplitude shift keying (ASK) protocol. The state of a 125 KHz signal being received determines the output voltage of the low frequency receiver chip **320**: on-high, off-low. While waiting for a signal, the low frequency receiver chip **320** may remain in standby mode, conserving power. Upon the presence of a programming signal, the low frequency receiver chip **320** may wake up and send the coded data to the microcontroller **310**, after which the microcontroller **310** may tell the low frequency receiver chip **320** to enter standby mode again. The programming signal may include a preliminary "on" time to wake up the low frequency receiver chip **320**, a 4-bit header, a 3-bit parameter identifier, and a 4-bit data value. Each bit time is 2 milliseconds, allowing enough time for the microcontroller **310** to process the bit reception before the next bit arrives. An antenna attached to a coil input of the low frequency receiver chip **320** may be a short wire having a strong programming signal.

[0056] Eight analog switches **330** may be used to control the output pulsing. Four switches **330** may determine a path of the selected voltage to the four possible output locations. Each of these may be controlled by one of the microcontroller outputs, which are in turn enabled or disabled depending on the internal variable for output locations. The inputs of the four switches **330** may be attached to the outputs of the other four switches **330**. The inputs of these four switches **330** may all be attached to different voltage dividers **340**, providing four different voltage levels, ranging from three quarters of the supply voltage to the supply voltage maximum of 3V. Each switch **330** may be controlled by an individual microcontroller signal, which also drives the voltage divider **340** for its particular switch **330**. For every pulse, only one of these four microcontroller signals is active, enabling the voltage from its divider **340** to be sent to the output switches **330** and ultimately to the electrodes **230** and **260**. The use of static voltage dividers **340** to provide output voltage scaling may minimize power consumption. In an alternative embodiment of the invention, a custom digital-analog converter could be used to allow for a higher range of stimulation voltages.

[0057] When tested for power consumption, a ~1 Ω resistor was put in series with a powering circuitry described herein. The voltage measured across the resistor while in operation was approximately 17 μV, implying that the DC current required is ~17 μA. At a supply voltage of 3 V, this equates to a power consumption of 51 μW. If operated for 24 hours, the implant would consume a little over 4.4 Joules/day. Typical parameters of a stimulation signal provided for Parkinson's disease are a series of pulses of 120 microsecond duration, 2.5 volts in strength at a repetition rate of 185 pulses per second. Assuming these typical parameters, there are:

185 pulses/second*60 seconds/minute*60 minutes/hour*24 hours/day=$1.5984*10^7$ pulses/day.

With pulse duration of 120 micro seconds, this gives a total energy application duration of $1.5984*10^7*120*10^{6=1918.08}$ seconds. With 2.5 volts and 50 micro amps=120 micro watts, the total energy required for stimulation is $120*10^6$ watts*1918.08 seconds=0.2302 joules per day. As disclosed herein, energy harvesting by the power circuit is on the order of 12-15 joules per day and the stimulation energy required is more than adequately provided by the power circuit.

[0058] An external programmer circuit **700** may include a microcontroller **710** including a PIC16LF87, an inductor/capacitor (LC) oscillating circuit **720** (125 KHz), and an intermediate MOSFET driver **730** including a TC4422 as shown in **FIG. 7**. The MOSFET driver **730** may supply enough energy for driving the LC circuit **720**. When a programming signal is to be sent, a button (not shown) may be pressed, telling the microcontroller **710** to read its inputs and stimulate the MOSFET driver **730** to oscillate the LC circuit **720** according to a communication protocol. Input voltages may be controlled by simple switches. Four switches may dictate the value to be sent, while five switches may dictate which parameter is to be changed. Only one of these switches should be on at one time. A Phidget RFID antenna **740** designed for 125 KHz may be attached to the high voltage side of a capacitor **750** of the LC circuit **720** for sending the programming signal. The circuit **700** may be powered via a 12-Volt wall supply. The 12 V drives the MOSFET driver **730** and is regulated to 5 V for the switches and microcontroller **710**.

[0059] An external powering circuit **800** may include a battery **810** for powering an oscillator **820** which drives a transformer-like setup **830** as shown in **FIG. 8A**. The coils **835** on one side of the transformer **830** may be disposed in a cap worn on the head of a patient, a headband worn on the head of the patient, or on a headboard of a bed in which the patient lies. The coils **840** on the other side of the transformer **830** may be coupled to the stimulation circuit **300** and may be disposed proximate the coils **835**. An AC signal coming from coil **840** may be amplified and rectified through a charge pump **850** having three stages, after which a voltage may be clamped with a regulator **860** to prevent spiking. A control circuit **870** may control operation of the voltage regulator **860**.

[0060] The oscillator **820** may include an LTC6900. This oscillator **820** produces a 50% duty cycle square wave to drive the primary coil **835** of the transformer **830** and requires only a potentiometer for adjusting the frequency. The charge pump **850** may be a Cockroft Walton voltage multiplier, utilizing a ladder of diodes and capacitors to rectify and amplify the signal. The amplification depends on the number of stages used. Three stages have been found to be enough for a substantial voltage multiplication across a load of 200 KΩ. The capacitors may be 0.1-μF each and the diodes may include BAT54SW surface mount diodes with a forward voltage drop of ~0.24-V. The regulator **860** may include an LT1521-3, which clamps a higher input voltage to 3 V.

[0061] Previous empirical testing showed square coils (both primary **835** and secondary **840**), 1 in.×1 in., with 5 turns each are effective for transferring enough energy to power the stimulation circuit **300**. Coils **840** are shown in PCB layout in **FIG. 9** and coils **835** are shown in PCB layout in **FIG. 10**.

[0062] The optimal frequency depends on the dielectric and distance between coils **835** and **840**. Frequencies in the range of 2 MHz to 15 MHz may be used. The oscillator **820** can be powered with 3 AAA batteries (4.5 V). In examining

the actual signal through the primary coil **835**, the voltage waveforms in **FIG. 11** and **FIG. 12** were obtained. **FIG. 11** shows the output voltage from the oscillator **820**. **FIG. 12** shows the voltage across the primary coil series resistance, from which the RMS current is calculated to be 29.36 mA_{RMS}.

[**0063**] The embodiment described above provides for near field harvesting and includes inductive coupling between coils **835** and **840**. With reference to **FIG. 8B** and in an alternative embodiment of the invention, the power circuit **865** for powering the stimulation circuit **300** may be non-inductively coupled to an external source of RF energy **880**. In this far field embodiment, the power circuit **865** may be disposed in a wrist band worn by the patient, in a room transmitter or in a transmitter disposed in a building occupied by the patient. In yet another alternative embodiment of the invention, the power circuit **865** may harvest ambient RF energy such as energy transmitted in space by using an inherently tuned antenna as described in U.S. Pat. No. 6,856,291, the description of which is incorporated by reference in its entirety herein. Furthermore, a rechargeable battery or other storage device (not shown) may be employed to store harvested energy. “Non-inductive” as described herein being directed RF.

[**0064**] To demonstrate the effectiveness of the powering and programming schemes through tissue, the device **200** was tested through swine skin. Clear tape was used to cover the conductive surfaces on the primary coil **835** and the secondary coil **840** to prevent interaction with the moisture on the skin. This tape had negligible effect on the inductive coupling.

[**0065**] Three different tests were performed, each following the same procedure. At a certain separation, the voltage powering the primary coil oscillator **820** was adjusted and the maximum output voltage from the secondary coil voltage regulator **860** was measured. The primary coil oscillator voltage started at 5 V and was decreased in increments of 0.1 V until the maximum regulator output voltage had reached a steady minimum.

[**0066**] The first test was performed with no skin between the transformer coils **835** and **840**. Data was acquired at separations of 5 mm, 7 mm, and 10 mm, values chosen based on the common range of human scalp thickness. The second test used fresh swine skin of thicknesses 5 mm and 7 mm between the transformer coils **835** and **840**. The test was interrupted, preventing the testing of 10 mm thick skin. The third test used the same pieces of swine skin, 5, 7, and 10 mm thick, after they had been frozen and thawed.

[**0067**] **FIG. 13** shows the results from the first two tests for comparison purposes. The presence of the skin reduced the inductive coupling between the coils, and hence the possible maximum output voltage in the range of 1.5-3.8 V. At 4.0 V and above, the maximum output voltage of ~3 V is obtainable even with the presence of the skin.

[**0068**] The same pieces of skin were tested a second time due to interruption of the first test. However, they had all been frozen and thawed in the interim, affecting the results slightly. **FIG. 14** shows the effect that the freezing and thawing of the skin had on the energy transfer of the transformer coils. Both the 7- and 10-mm thick pieces of skin reduced the inductive coupling, but the 5-mm skin actually improved in performance. This may be due to the presence of a layer of fat in both the 7- and 10-mm pieces that is absent in the 5-mm piece.

[**0069**] Another test was performed to find the effect of the skin over time. The stimulus for this test was the degrading performance of the 10-mm thick skin over time during the interrupted test mentioned above. For this test, the 7-mm thick piece of skin was used between the primary and secondary coil. The frequency was adjusted to produce a maximum output voltage, which was measured successively over a period of time. The results shown in **FIG. 15** support the fact that performance does not degrade over time. The slight drop in output voltage is likely due to the mechanical nature of the frequency-tuning potentiometer. Notice that the output voltage reaches a steady value and remains constant after that point.

[**0070**] In order to demonstrate the concept and functionality of the device **200**, a model **1600** was created as shown in **FIG. 16**. The stimulation circuit **300** was put on top of a Styrofoam head **1610** with wires running down through the bottom for power- and pulse-monitoring purposes. An ABS Plastic cap (not shown) was made to simulate the head’s scalp, covering the stimulation circuit **300**. The primary powering coil **835** with the batteries **810** was secured in a hat **1620** over the position of the stimulation circuit **300** to provide for near field inductive coupling.

[**0071**] The device **200** was tested on a cadaver head to show that a signal may be generated through the scalp and to demonstrate the programmability of the device **200** during stimulation. First, an incision was made in the scalp of the cadaver head. The secondary coil **840** was inserted and placed on the skull and the incision was sewn, leaving the lead wires **220** of the circuit exposed. Six wires were used on the implanted circuitry, four representing the electrodes **230**, which were connected to an oscilloscope and two wires for power and ground. The primary coil **835** was then taped on the scalp directly on top of the implanted circuitry and connected to a power supply.

[**0072**] The experiment began by demonstrating the programmability of the stimulation circuit **300**. Four parameters were varied and displayed on the oscilloscope; pulse width (60, 120 and 180 micro seconds), amplitude (2.34V, 2.75V, 2.94V, 3.13 V), frequency (191 and 194 Hz) and the shifting from one stimulating probe to another (i.e. probe **1** to probe **2** or probe **1** to all four probes). A 10K OHM resistor was used to represent the brain resistance, however this resistance is higher than the resistance for the brain (900 to 1100 Ohms), but a 10 k Ohm resistor was used to ensure there was enough power.

[**0073**] Next, several voltages were tested to determine the output source voltage of the power circuit **865**. Initially, the power supply connected to the primary coil **835** was set at 5 V and was decremented by 0.1 V to 1.2 V. The voltage on the secondary coil **840** was clamped so as not to exceed 3V. As the voltage decreased on the power supply, the output voltage on the secondary coil **840** was steady at 3 V until it declined around 2.2 V. Once the voltage decreased from 3V, a potentiometer was adjusted to obtain the maximum voltage. The data obtained shows the when the voltage drops, the amplitude voltage and frequency drop off as well.

[**0074**] The device for brain stimulation using RF harvesting of the present invention provides a brain stimulation device that requires a single implantation site and surgery to thereby reduce both the cost and trauma to the patient of the implantation procedure. The brain stimulation device further uses RF energy as a power source to eliminate the need for a battery implanted in the pectoral area of the patient. The brain stimulation device further converts RF energy and

stores the converted RF energy for use in stimulation of targeted brain areas. The brain stimulation device is flexible and implantable under the scalp to minimize discomfort for the patient.

[0075] The foregoing description of the embodiments of the invention has been presented for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed. Many modifications and variations are possible in light of the above teaching. It is intended that the scope of the invention be limited not by this detailed description, but rather by the claims appended hereto.

TABLE 1

Indication	Prevalence	Reference for Prevalence
	FDA Approved Indications	
Essential tremor (1997)	1.5–3% population	Am J Med 115:134–42, 2003
Parkinson's (2002)	1% of population > 50	
Dystonia (2004)	~150,000 in US	Mov Disord 3:188–94, 1988
	Current Clinical Trials	
Obsessive-compulsive disorder	2–3% of population	J Clin Psych 53 Suppl: 4–10, 1992
Tourette's syndrome	~1–2% children	J Psychosom Res. 55:3–6, 2003; Can J Neurol Sci Suppl 1:S64–71, 2003
Intractable epilepsy	~100,000 in US	Neurology 56: 1445–52, 2001; Rev Neurol (Paris); 160 Spec No 1:5S31–5, 2004
Intractable depression	1,000,000	Psychiatr Clin North Am 19:179–200, 1996; http://www.mhsource.com/depconsult/june 2004.jhtml?_requestid = 605984

We claim:

1. A system for brain stimulation using radio frequency harvesting comprising:

a device implantable under a scalp of a patient, the device comprising a radio frequency harvesting power circuit and a stimulation circuit; and

at least one electrode coupled to the stimulation circuit, the at least one electrode providing brain stimulation to targeted areas of the brain.

2. The system of claim 1, wherein the device is fabricated from a biocompatible substrate.

3. The system of claim 1, wherein the device is flexible and conformable to a shape of the scalp.

4. The system of claim 1, wherein the power circuit comprises a charge pump inductively coupled to a primary coil of an external power circuit.

5. The system of claim 1, wherein the power circuit comprises an inherently tuned antenna for harvesting energy transmitted in space.

6. The system of claim 1, wherein the at least one electrode provides deep brain stimulation.

7. The system of claim 1, wherein the at least one electrode provides cortical stimulation.

8. The system of claim 1, further comprising a programming circuit operable to control the stimulation circuit.

9. The system of claim 8, wherein the programming circuit is operable to control a stimulation circuit voltage output.

10. The system of claim 8, wherein the programming circuit is operable to control a stimulation circuit output pulse width.

11. The system of claim 8, wherein the programming circuit is operable to control a stimulation circuit output frequency.

12. The system of claim 1, further comprising an energy storage device coupled to the device.

13. A system for brain stimulation comprising:

a device implantable under a scalp of a patient, the device comprising a coupled power circuit and a stimulation circuit; and

at least one electrode coupled to the stimulation circuit, the at least one electrode providing brain stimulation to targeted areas of the brain.

14. The system of claim 13, wherein the coupled power circuit is inductively coupled.

15. The system of claim 13, wherein the coupled power circuit is non-inductively coupled.

16. A system for brain stimulation comprising:

a device implantable under a scalp of a patient, the device comprising a power circuit powered by ambient radio frequency energy and a stimulation circuit; and

at least one electrode coupled to the stimulation circuit, the at least one electrode providing brain stimulation to targeted areas of the brain.

17. The system of claim 1, wherein the at least one electrode provides brain stimulation to targeted areas of the brain in response to signals received from the stimulation circuit.

18. A method of providing brain stimulation, comprising:

harvesting power in a power harvesting circuit in a device implantable under the scalp of a patient; and

providing brain stimulation to targeted areas of the brain with at least one electrode connected to the power harvesting circuit.

19. The method of claim 18, wherein the power harvesting circuit harvests radio frequency energy.

20. The method of claim 18, wherein the power harvesting circuit harvests energy by an inductively coupled power circuit in the device.

21. The method of claim 18, wherein the power harvesting circuit harvests energy by a non-inductively coupled power circuit in the device.

22. The method of claim 18, wherein the brain stimulation is used to treat Parkinson's disease.

23. The method of claim 18, wherein the brain stimulation is used to treat stroke patients.

* * * * *

We use cookies to make sure our websites work effectively and to improve your user experience. If you continue to use this site we will assume that you are happy with this. However, you can change your cookie settings at any time.

More Info/Change Settings

Continue

The Open University | Study with The Open University

OpenLearn | Free Learning from The Open University

The Open University

Home | Latest | Free courses | Subjects | Skills | TV & Radio |

Create account / Sign in

Subjects ▸ Science, Maths & Technology ▸ All content ▸ **Building a crystal radio set**

Science, Maths & Technology | Featured content | Free courses | All content

Building a crystal radio set

Updated Tuesday 3rd July 2007

For many of the original teenagers, home-built crystal radios gave them their first contact with the world beyond their bedrooms. Here, Fraser Robertson describes how to make your own.

Share this article

Author

A crystal set is the most basic type of radio receiver. It is capable of receiving AM (amplitude modulated) transmissions, which are generally found on the long, medium and short wave bands.

The magic of the crystal set is that no power source is required; a tiny fraction of the transmitted signal provides sufficient power for the signal to be heard. What follows, while listing the parts you'll need, isn't a step-by-step how to guide, as part of the fun in building a crystal set is experimenting and optimising performance.

Aerial

Unless the crystal set is very close to the transmitter, an aerial, (or antenna) will be required in order to receive a big enough signal for the crystal set to work. Generally the aerial will be a length of wire, which should be as long and as high as possible.

A wire strung from an upstairs window to a distant tree will work well; say 30 metres long and seven metres high. The type of wire is not critical, but PVC covered stranded copper wire of about 2 milimetres diameter will work well.

If the wire is too thin it may break in the wind, and/or birds may fly into it; too thick and it will sag under its own weight.

The aerial should be insulated at each end from its support, a short length of washing line or other non-conductive cord can be used. An effective aerial will produce about one tenth of a volt radio frequency (RF), which will result in good reception.

Fraser Robertson (Other Open University departments)

Fraser Robertson is a project officer at The Open University, supporting various research projects.

View author profile >

Related content

Crystal Radio

Transmitters

Receivers

More +

More like this

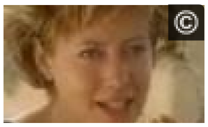

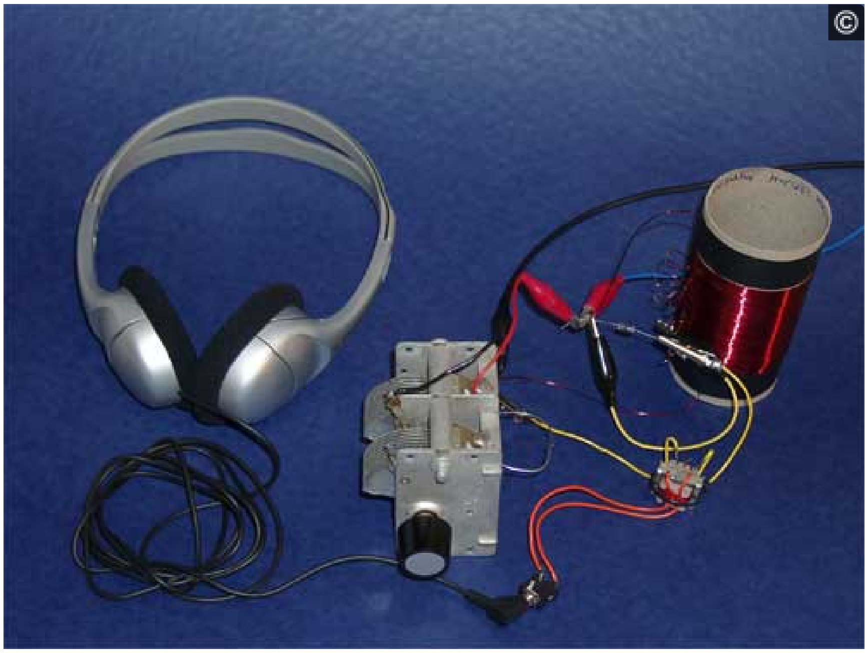

Earth

An earth connection is required to complete the aerial circuit. This can be a length of copper pipe driven into the ground, or a length of buried un-insulated wire.

As with the aerial, a better earth connection will improve reception. It is possible to connect to a metal heating or water pipe, but **this is not recommended on safety grounds** should there be an electrical fault. For the same reason, **the mains earth** should not be used.

When the crystal set is not in use the aerial should be connected directly to the earth connection. This will prevent any possible static build up on the aerial.

Tuned circuit

The aerial and earth are connected across a parallel tuned circuit. The purpose of this is to select the wanted station.

A simple crystal set will not have good selectivity so if two strong signals are received on nearby frequencies it will not be possible to separate them.

The tuned circuit comprises a coil (inductor) and a capacitor, which is usually variable for tuning. Pre-wound ferrite rod coils can be bought, or you can wind your own: for example, for medium wave, 100 turns of 0.5 milimetre enameled copper wire wound on a toilet roll or similar sized non-metallic former, twisted/tapped every ten turns.

Kathy's Carriacou diary: Arrival

Bang Goes The Theory: Show me the bunny

Sir Bernard Lovell: An appreciation

The science of making musical instruments

The greater number of turns, the longer the wavelength (lower frequency), so a coil for long wave will require more turns. If using a pre-wound ferrite rod, ignore the small link windings.

An air spaced variable capacitor (with a value around 500 picofarad) removed from an old radio is ideal, or alternatively a small polyvaricon capacitor as used in transistor radios.

Variable capacitors often have two or more capacitors in one (known as **ganged**), in which case these can be connected together to increase the maximum capacitance (lower the tuning frequency).

Some variable capacitors also have FM (VHF) tuning sections included, which can be ignored. If a variable capacitor is not available, a fixed capacitor can be used (e.g. 150 picofarad - 470 picofarad), with the tuning adjusted by changing taps on the coil. In which case, taps every five turns would be useful.

Detector

The purpose of the detector is to recover the audio signal (waveform) from the radio frequency signal. Early crystal sets used an adjustable ‘cats whisker’ detector, but a germanium or Schottky diode is now used, examples are OA90, OA91, BAT81, BAT85, 1N34, 1N5711. Any one of these can be used.

Silicon diodes, (such as 1N4148) do not work as well due to their higher forward voltage drop. Modern zero gate threshold voltage FETs can be used for better sensitivity, although these are not readily available in small quantities. The detector diode is polarized (directional), but in the crystal set circuit it does not matter which way round it is connected.

Headphones

Traditional crystal sets used sensitive high impedance headphones, typically several thousand Ohms, which are hard to obtain now. A crystal earpiece can be used, or a sensitive pair of walkman style headphones (usually 32+32 Ohms) connected through a transformer (e.g. type LT700) to raise the impedance. Low impedance headphones will ‘load’ the tuned circuit excessively.

Look for headphones with a high sensitivity figure, for example 115 decibels - a higher number indicates higher sensitivity; CPC AV03626 106 decibel headphones work well. Be aware that a higher quoted decibel specification does not always guarantee higher sensitivity in practice.

Note that if a crystal earpiece is used, a resistor (such as a 47k Ohms) may need to be connected across it to provide a DC path for the detector diode.

Construction

The fine detail here is left to the builder. Component leads can be soldered, or twisted together, or ‘crocodile clips’ can be used to make connections.

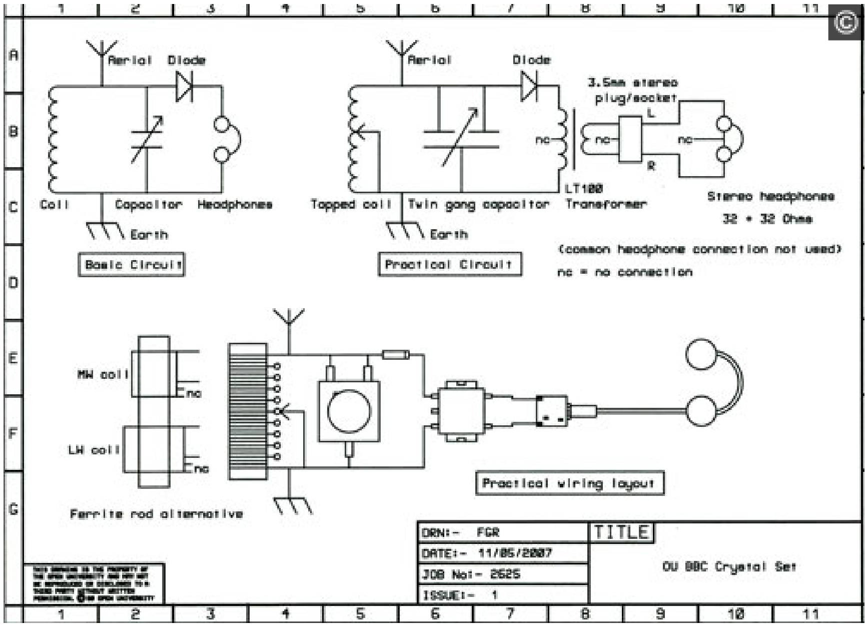
1 2 3 4 5 6 7 8 9 10 11
A B C D E F G
Aerial
Diode
Coil
Capacitor
Headphones
Earth
Basic Circuit
Aerial
Diode
3.5mm stereo plug/socket
L
R
nc
nc
nc
Tapped coil
Twin gang capacitor
LT100 Transformer
Earth
Stereo headphones
32 + 32 Ohms
Practical Circuit
(common headphone connection not used)
nc = no connection
MW coil
nc
LW coil
nc
Ferrite rod alternative
Practical wiring layout
DRN:- FGR
DATE:- 11/05/2007
JOB No:- 2625
ISSUE:- 1
TITLE
OU BBC Crystal Set

The diagram above, and the photograph at the top of the page, offers some guidance. **You can download a larger pdf of the diagram by clicking here**.

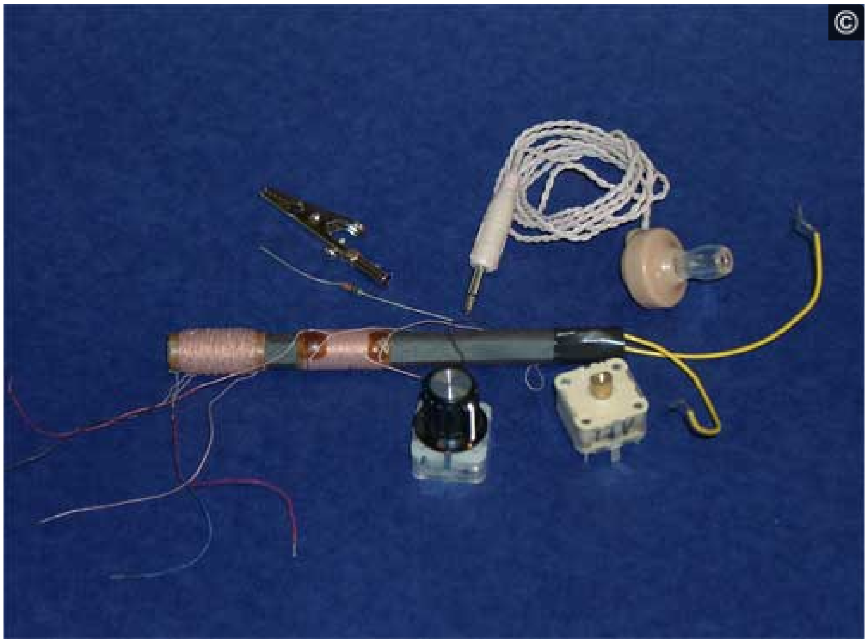

This photograph shows examples of alternative suggested parts. The components can be lashed together initially for experimentation, then connected more permanently and housed in any suitable enclosure once the crystal set is working.

Typical performance and troubleshooting

In the UK it will normally be possible to pick up two or three stations on medium wave, and/or BBC Radio 4 from Droitwich on long wave. If nothing at all is heard check:

- that all connections are sound, and ensure the enamel insulation on coil wire has been removed by being scraped off where connections are made
- that the correct connections for each component have been used
- that the tuning capacitor vanes are not bent and touching together.

If you get weak reception it may be that you need to improve the aerial / earth connections, or try a different pair of headphones.

Weblink

Watch a video showing the process of making a crystal set from **the OU's Interfaculty Electronics and Computing department**.

Copyright information

+

Ratings ★★★★☆ Your rating None. Average rating 3.8 out of 5, based on 30 ratings

Be the first to post a comment

Leave a comment

Sign in or create your **OpenLearn account** to join the discussion.

We invite you to discuss this subject, but remember this is a public forum.

Please be polite, and avoid your passions turning into contempt for others. We may delete posts that are rude or aggressive; or edit posts containing contact details or links to other websites.

For further information, take a look at our frequently asked questions which may give you the support you need.

Have a question?

Other content you may like

Science, Maths & Technology

DIY: The Saucepan Radio

A step-by-step guide on how to make a radio, part of the BBC/OU's programme website for Rough Science 1

▶ Video

10 mins

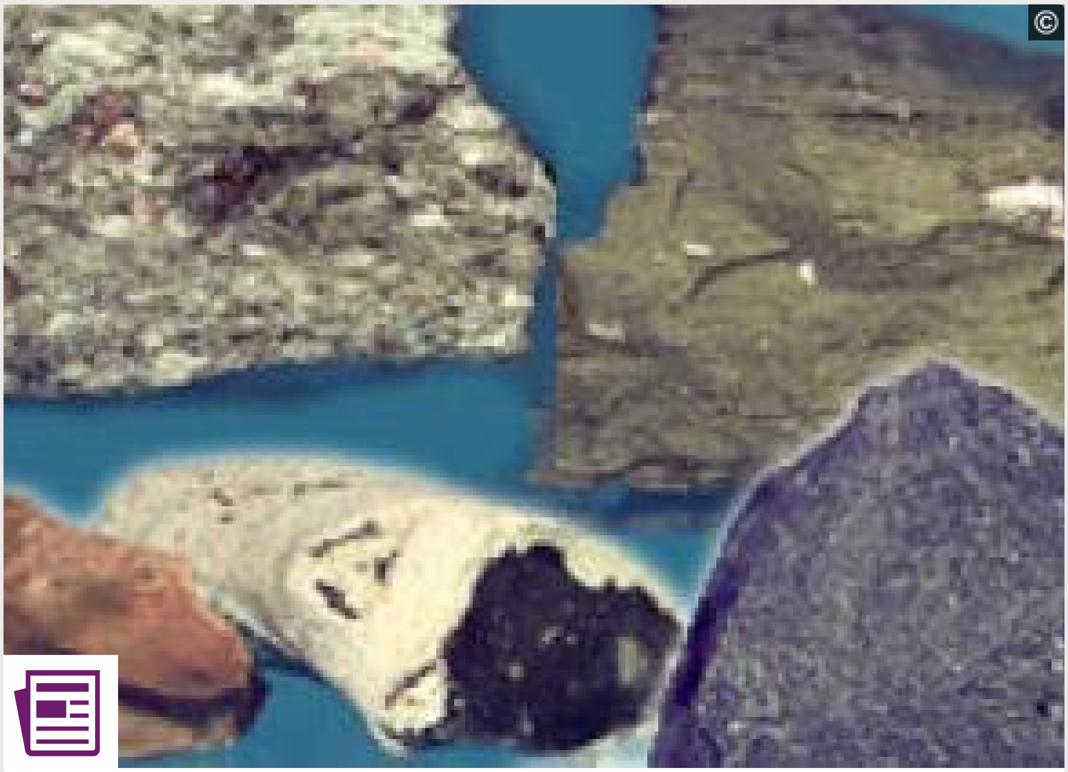

Science, Maths & Technology

You have chosen crystals

What is your crystaline rock?

Article

TV, Radio & Events

Rough Science 5 Zanzibar: Jonathan...

Jonathan Hare's diary about the challenge for the Call of the Wild programme, part of the fifth BBC/OU TV series Rough Science, based in...

Article

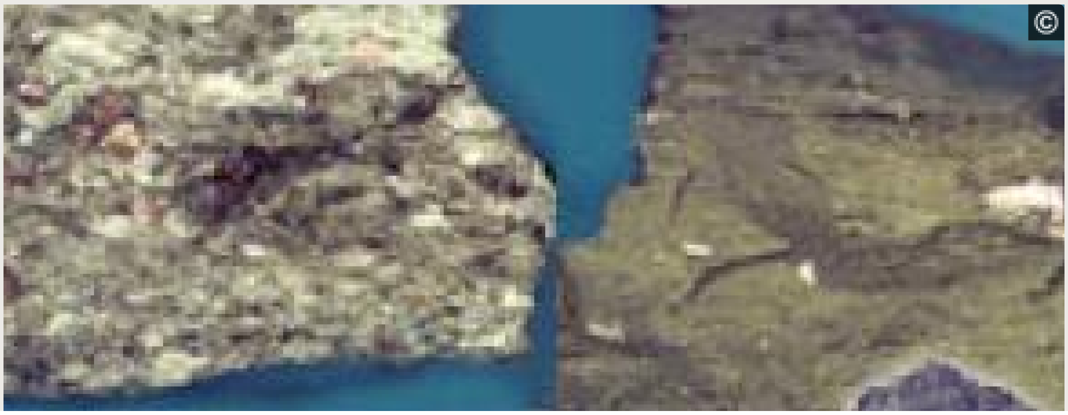

Science, Maths & Technology

You have chosen random

A step closer to identiifying your rock

Article

Science, Maths & Technology

Gneiss

A brief description of the nature of gneiss

Article

History & The Arts

Speaking Shakespeare: How was Shakespeare...

How do we know what actors sounded like when they performed premieres of Shakespeare plays? The clues, as ever, are in the lines themselves...

Video 10 mins

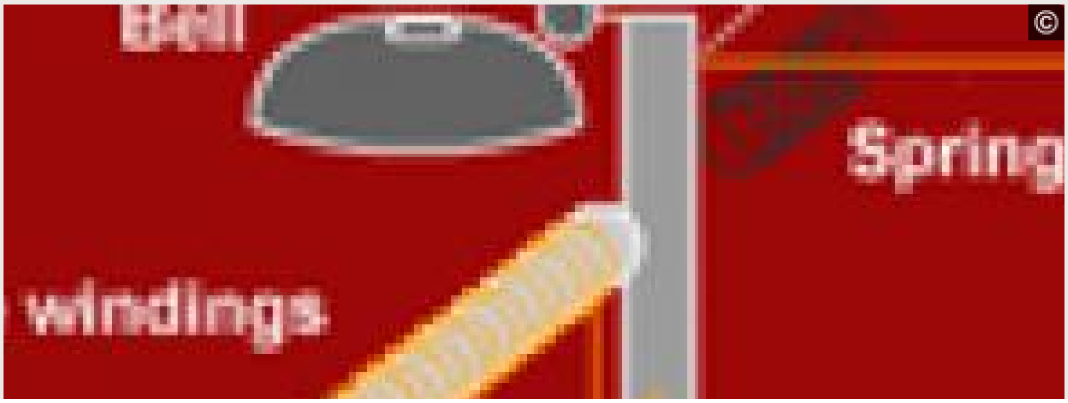

Science, Maths & Technology

Challenge: Transmit and Receive a Signal

The science behind transmitting and receiving radio signals, part of the BBC/OU's programme website for Rough Science 2

Article

Science, Maths & Technology

Microgabbro/ dolerite

A brief description of the nature of microgabbro or dolerite

Article

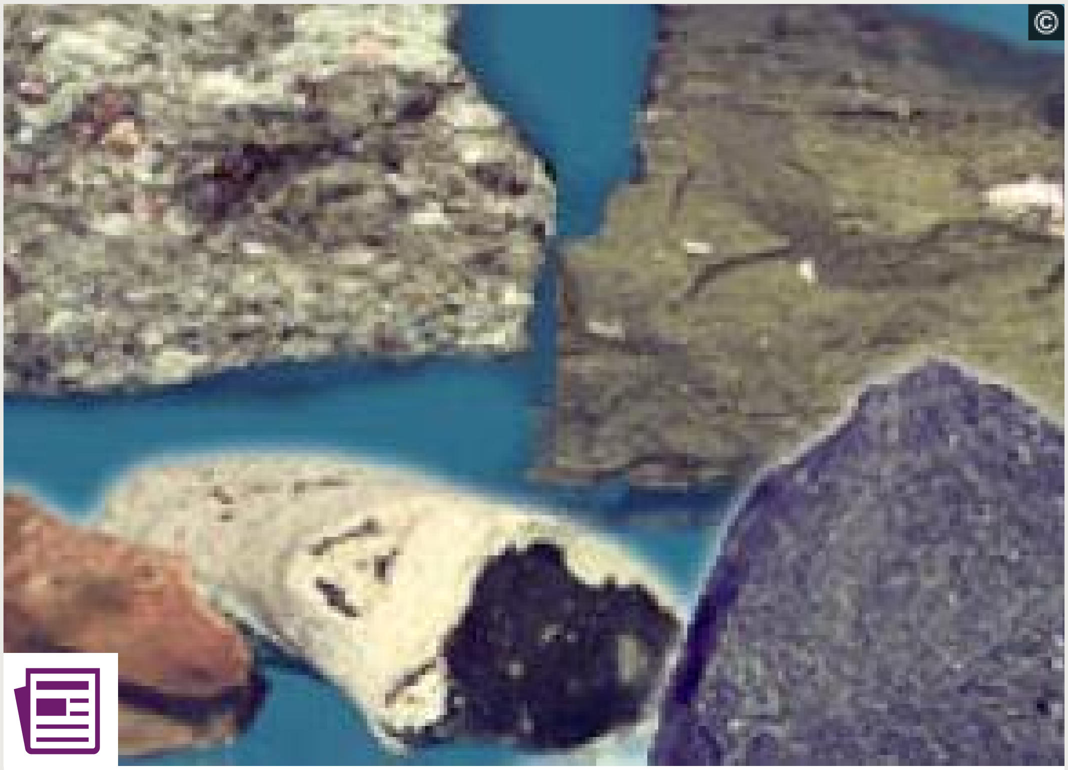

Science, Maths & Technology

Rock analyser: Just the facts

Our step by step rock analyser will help you identify your sample...

Article

Back to top

Search for free courses, interactives, videos and more!

About OpenLearn

About us

Frequently asked questions

Study with The Open University

Subscribe to our newsletter

Contact OpenLearn

OpenLearn Create

OpenLearn Cymru

Explore subjects

Money & Business

Education & Development

Health, Sports & Psychology

History & The Arts

Languages

Nature & Environment

Science, Maths & Technology

Society, Politics & Law

Explore resources

Latest from OpenLearn

Free courses

Skills for study

Skills for work

OU on TV & radio

Order our free print items

My OpenLearn profile

Our partners

OpenLearn works with other organisations by providing free courses and resources that support our mission of opening up educational opportunities to more people in more places.

Support us

©2017. All rights reserved. The Open University is incorporated by Royal Charter (RC 000391), an exempt charity in England & Wales and a charity registered in Scotland (SC 038302). The Open University is authorised and regulated by the Financial Conduct Authority in relation to its secondary activity of credit broking.

Conditions of use Privacy and cookies
Modern Slavery Act Copyright

Twitter Facebook YouTube

US 20070117596A1

(19) **United States**

(12) **Patent Application Publication**

Greene et al.

(10) **Pub. No.: US 2007/0117596 A1**

(43) **Pub. Date: May 24, 2007**

(54) **RADIO-FREQUENCY (RF) POWER PORTAL**

(75) Inventors: **Charles E. Greene**, Pittsburgh, PA (US); **Daniel W. Harrist**, Carnegie, PA (US); **John G. Shearer**, Ligonier, PA (US)

Correspondence Address:
Ansel M. Schwartz
Suite 304
201 N. Craig Street
Pittsburgh, PA 15213 (US)

(73) Assignee: **Powercast, LLC**

(21) Appl. No.: **11/601,142**

(22) Filed: **Nov. 17, 2006**

Related U.S. Application Data

(60) Provisional application No. 60/738,523, filed on Nov. 21, 2005.

Publication Classification

(51) **Int. Cl.**

H03D 7/16	(2006.01)
H04B 1/10	(2006.01)
H04B 1/16	(2006.01)
H04B 1/38	(2006.01)

(52) **U.S. Cl.** **455/572**; 455/299; 455/131; 455/343.1

(57) **ABSTRACT**

A system for powering a mobile device having an energy harvester which receives wireless energy and converts the energy into current includes a first portal in which wireless energy is transmitted for the energy harvester of the mobile device to receive and convert it into current when the device is in the first portal. The system comprises a second portal separate and apart from the first portal with a gap between the first portal and the second portal, the second portal in which wireless energy is transmitted for the energy harvester of the mobile device to receive and convert it into current after the device has passed through the first portal and the gap and is in the second portal. A method for powering a mobile device having an energy harvester which receives wireless energy and converts the energy into current is also disclosed.

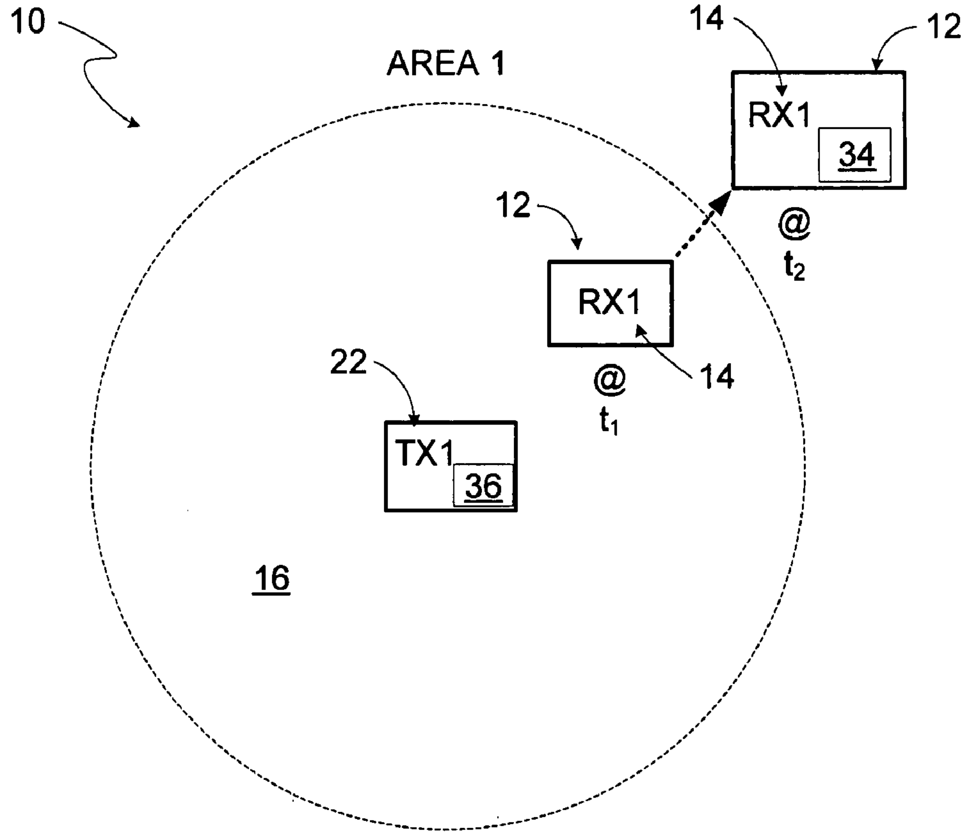

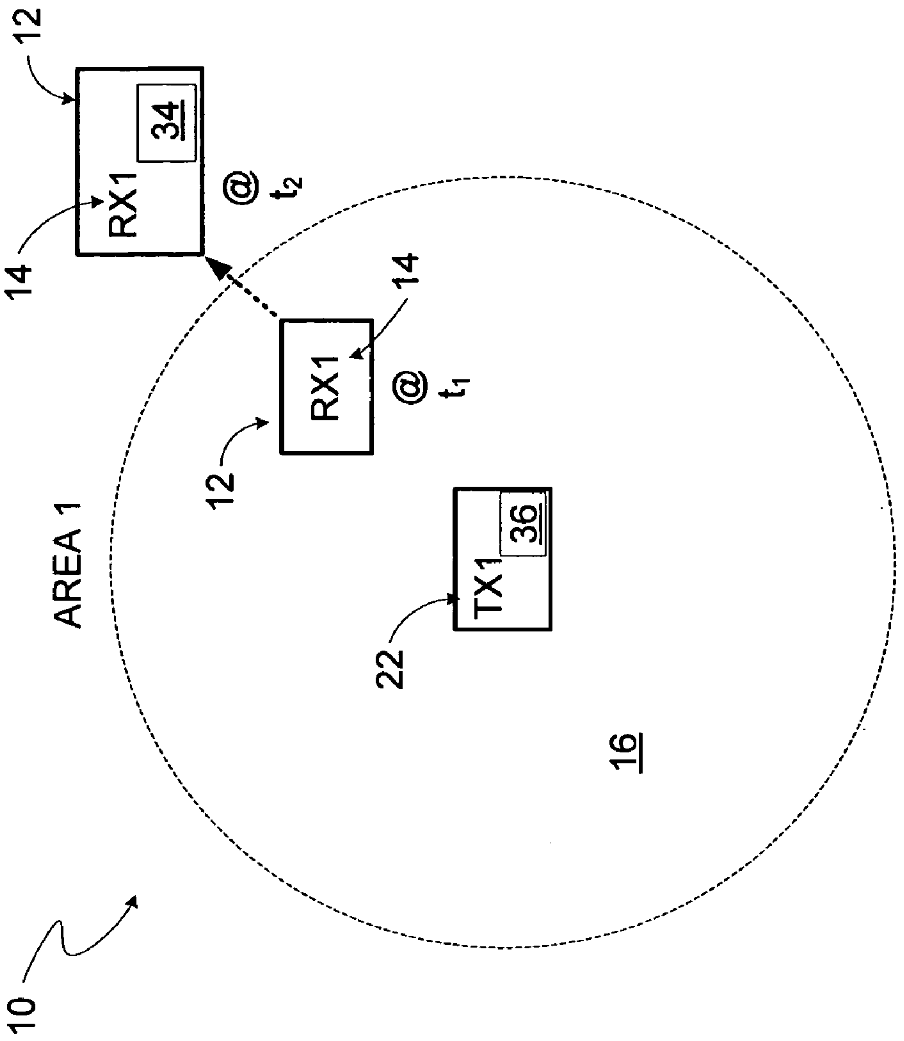
12
34
RX1
14
@ t_2
RX1
14
@ t_1
12
AREA 1
TX1
36
22
16
10

Fig. 1

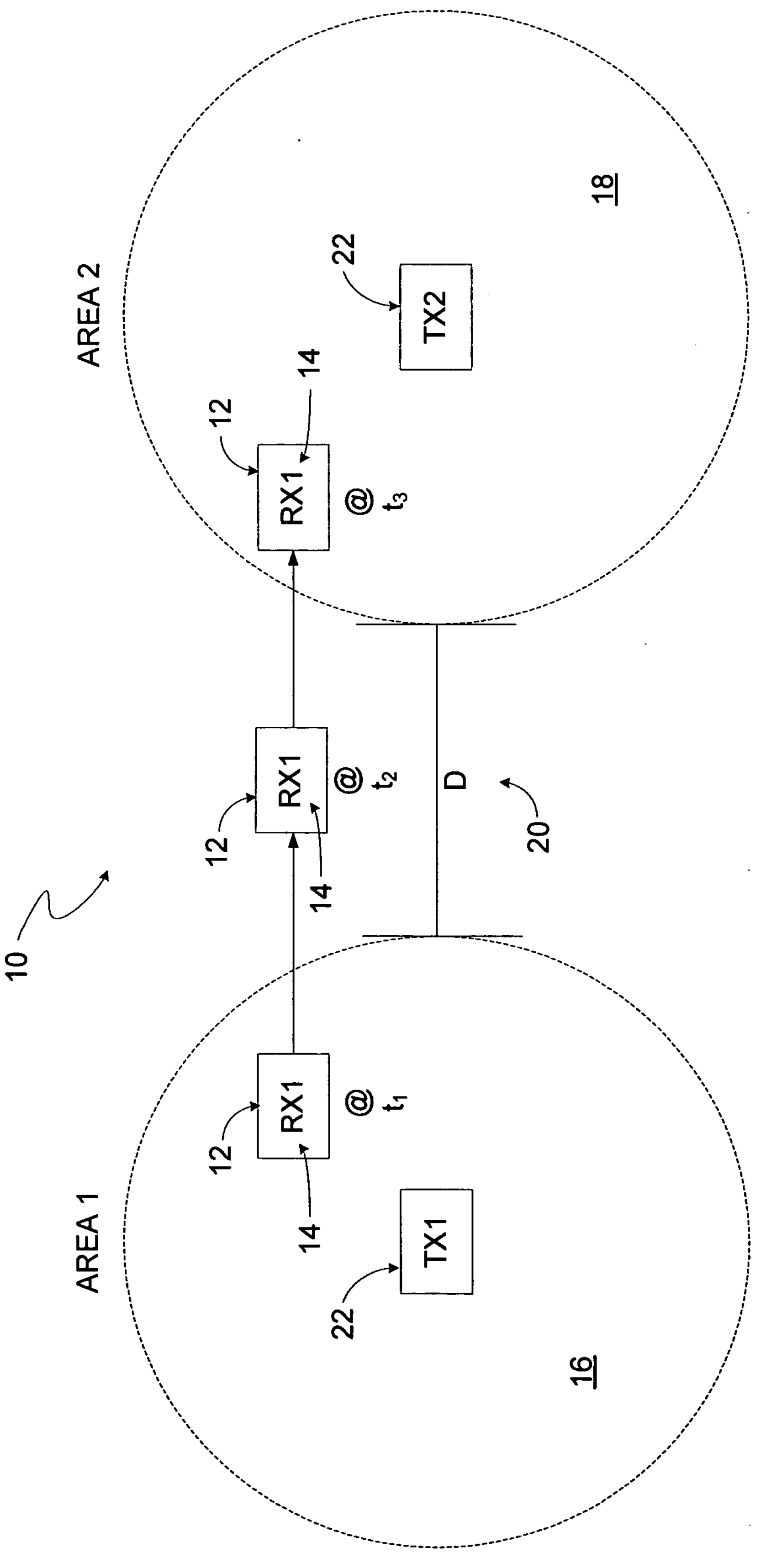
10
AREA 1
16
TX1
22
RX1
12
14
@ t_1
RX1
12
14
@ t_2
D
20
AREA 2
18
RX1
12
14
@ t_3
TX2
22

Fig. 2

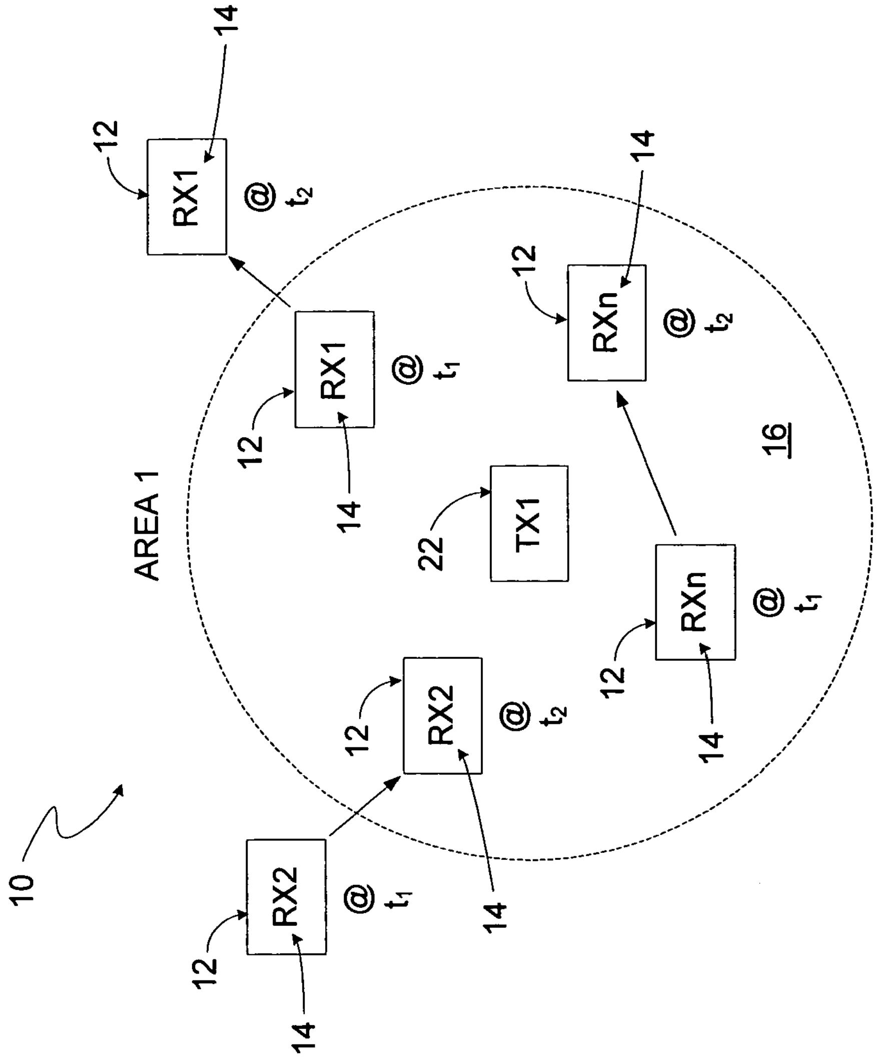
10
AREA 1
RX1
@ t_2
12
14
RX1
@ t_1
12
14
TX1
22
RXn
@ t_2
12
14
RXn
@ t_1
12
14
16
RX2
@ t_2
12
14
RX2
@ t_1
12
14

Fig. 3

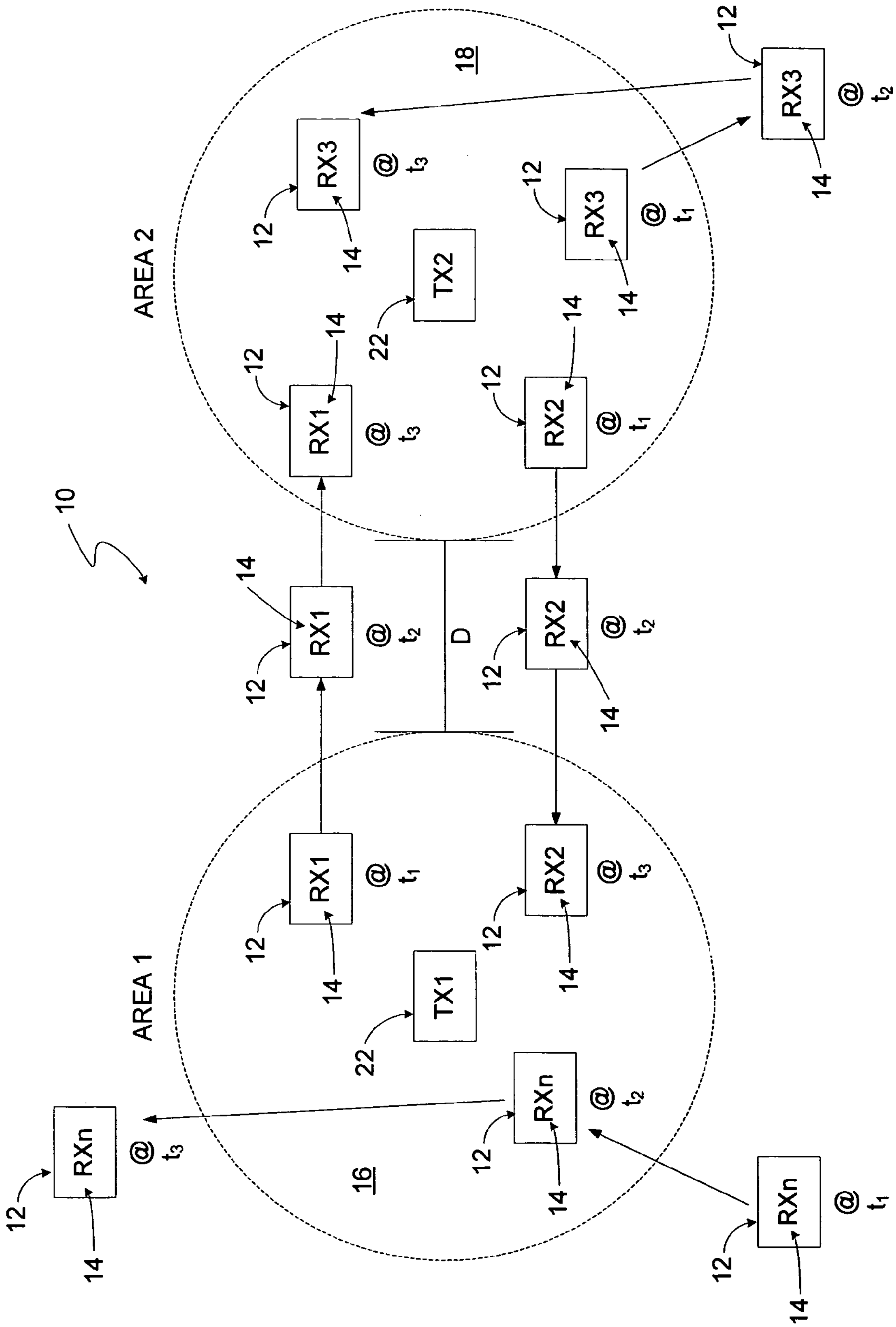
10
AREA 1
16
AREA 2
18
D
TX1
22
TX2
22
RX1 @ t_1
RX1 @ t_2
RX1 @ t_3
RX2 @ t_1
RX2 @ t_2
RX2 @ t_3
RX3 @ t_1
RX3 @ t_2
RX3 @ t_3
RXn @ t_1
RXn @ t_2
RXn @ t_3
12
14

Fig. 4

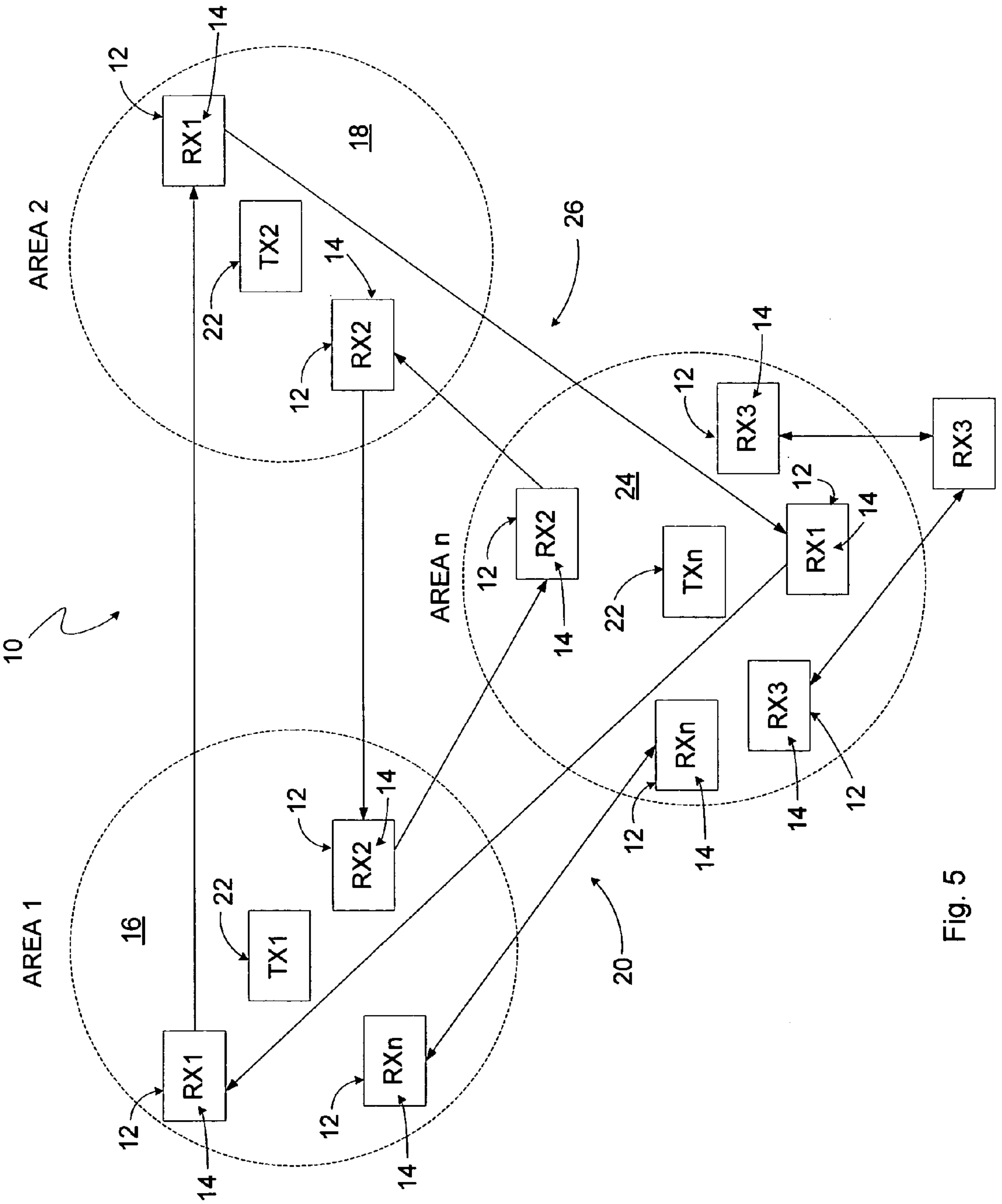
10
AREA 1
16
RX1
TX1
RX2
RXn
22
12
14
AREA 2
18
RX1
TX2
RX2
AREA n
24
RX2
TXn
RX1
RX3
RX3
RX3
RXn
20
26

Fig. 5

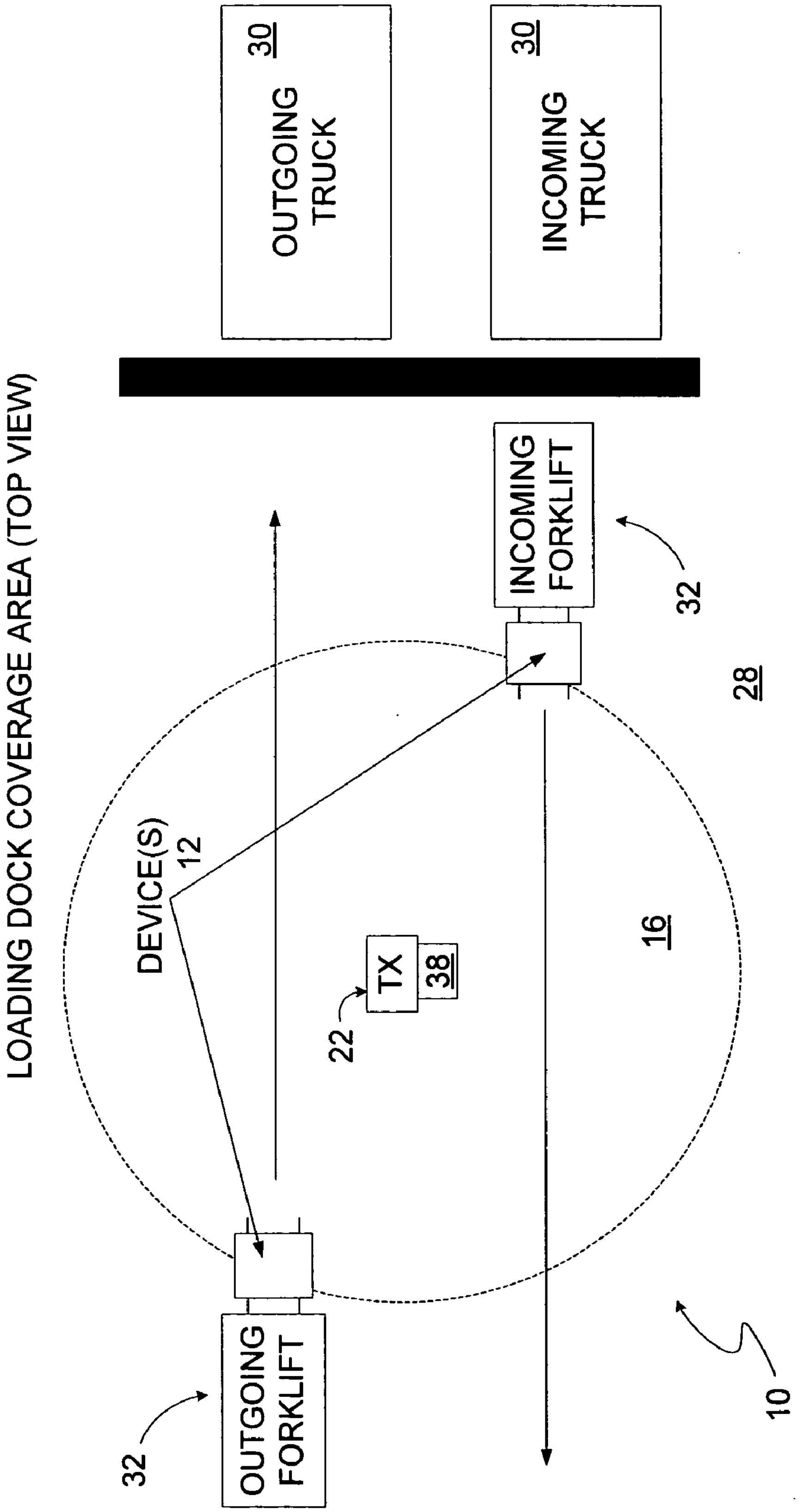
LOADING DOCK COVERAGE AREA (TOP VIEW)
OUTGOING TRUCK
30
INCOMING TRUCK
30
INCOMING FORKLIFT
32
DEVICE(S) 12
TX
38
22
16
28
OUTGOING FORKLIFT
32
10

Fig. 6

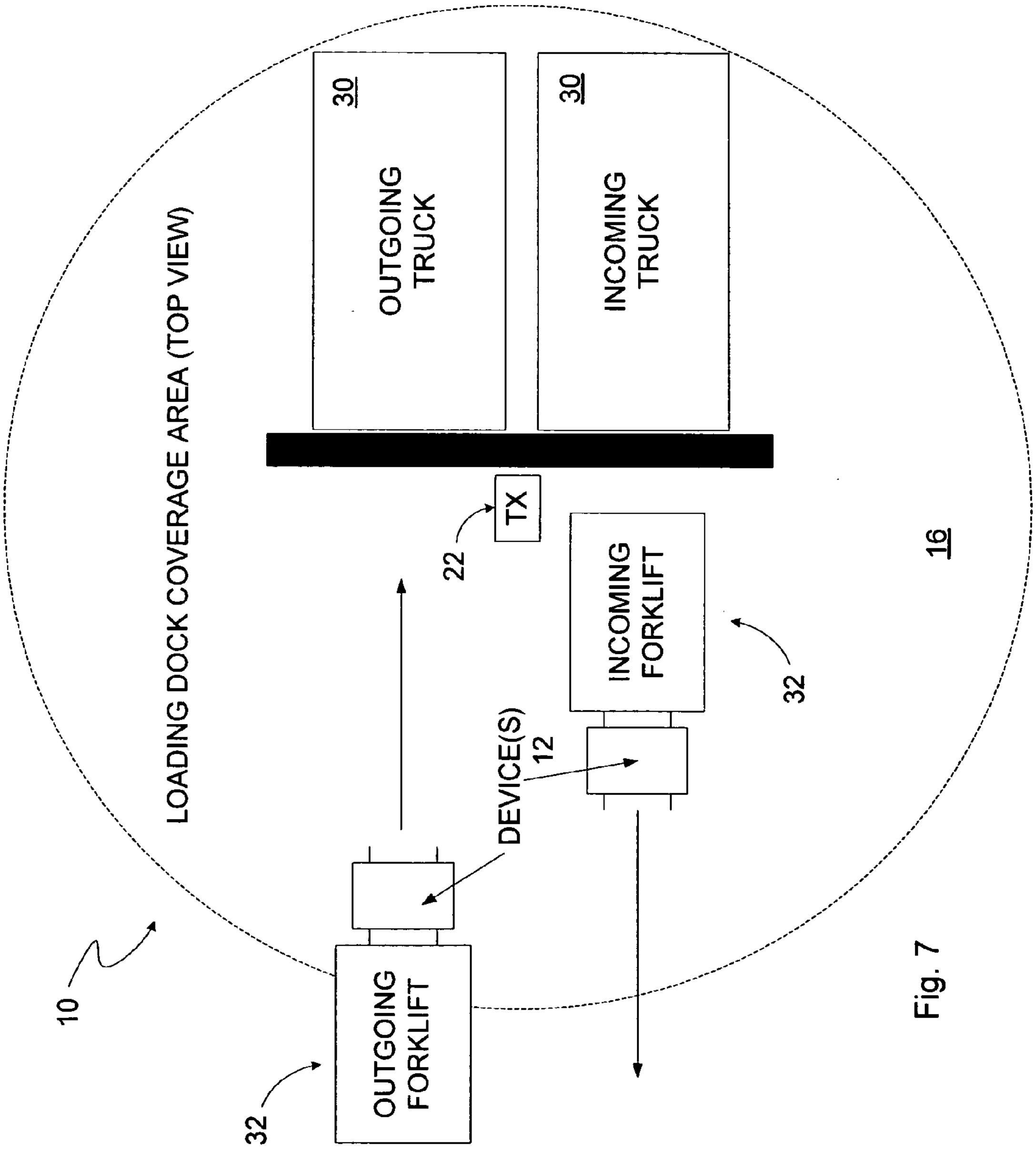
10
LOADING DOCK COVERAGE AREA (TOP VIEW)
30
OUTGOING TRUCK
30
INCOMING TRUCK
TX
22
INCOMING FORKLIFT
32
16
DEVICE(S) 12
OUTGOING FORKLIFT
32

Fig. 7

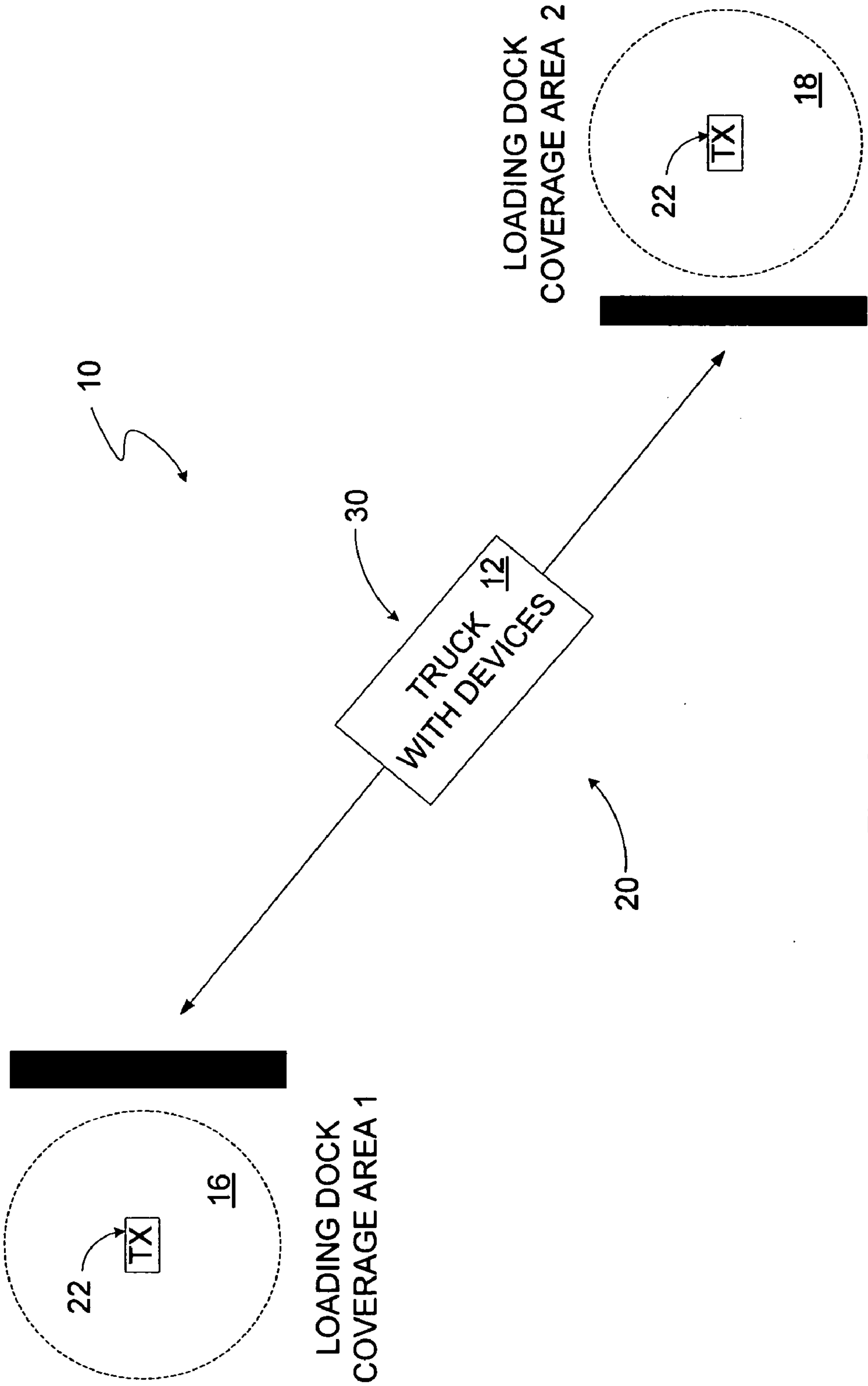
LOADING DOCK
COVERAGE AREA 2
18
TX
22
10
30
TRUCK
WITH DEVICES
12
20
LOADING DOCK
COVERAGE AREA 1
16
TX
22

Fig. 8

RADIO-FREQUENCY (RF) POWER PORTAL

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Patent Application Ser. No. 60/738,523, filed Nov. 21, 2005.

BACKGROUND OF THE INVENTION

[0002] 1. Field of the Invention

[0003] The present invention relates to a system for powering a mobile device having an energy harvester which receives wireless energy and converts the energy into current. More specifically, the present invention relates to a system for powering a mobile device having an energy harvester which receives wireless energy and converts the energy into current when it is in a portal in which wireless energy is transmitted for the energy harvester of the mobile device.

[0004] 2. Description of Related Art

[0005] As processor capabilities have expanded and power requirements have decreased, there has been an ongoing explosion of devices that operate completely independent of wires or power cords. These "untethered" devices range from cell phones and wireless keyboards to building sensors and active Radio Frequency Identification (RFID) tags.

[0006] Engineers and designers of these untethered devices continue to have to deal with the limitations of portable power sources, primarily using batteries as the key design parameter. While the performance of processors and portable devices has been doubling every 18-24 months driven by Moore's law, battery technology in terms of capacity has only been growing at a measly 6% per year. Even with power conscious designs and the latest in battery technology, many devices do not meet the lifetime cost and maintenance requirements for applications that require a large number of untethered devices such as logistics and building automation. Today's devices that need two-way communication require scheduled maintenance every three to **18** months to replace or recharge the device's power source (typically a battery). One-way devices that simply broadcast their status without receiving any signals, such as automated utility meter readers, have a better battery life typically requiring replacement within 10 years. For both device types, scheduled power-source maintenance is costly and can be disruptive to the entire system that a device is intended to monitor and/or control. Unscheduled maintenance trips are even more costly and disruptive. On a macro level, the relatively high cost associated with the internal battery also reduces the practical, or economically viable, number of devices that can be deployed.

[0007] The ideal solution to the power problem for untethered devices is a device or system that can collect and harness sufficient energy from the environment. The harnessed energy would then either directly power an untethered device or augment a power supply. However, this ideal solution may not always be practical to implement due to low energy in the environment, and site restrictions may limit the ability to use a dedicated energy supply. The proposed invention takes these factors into account and provides a solution for both the ideal situation and also for more restrictive circumstances.

[0008] U.S. Pat. No. 6,127,799 discloses a method and apparatus for wireless powering and recharging by use of an electromagnetic field with an antenna. The antenna may be one or more dipole antennas that are combined to form at least two subsets of dipole antenna element array, wherein one subset may be oriented at an acute or a right angle with respect to at least one other subset. The technology disclosed in U.S. Pat. No. 6,127,799 is limited by the fact that the power conversion efficiency of the antenna array is dependent on the number of dipoles. Additionally, U.S. Pat. No. 6,127,799 discloses the use of dipole antennas, which may not be practical for most battery-operated devices due to the size of the antenna and antenna array. This prior disclosure, states that the dipole antennas are used to cover more than one side of a battery that has a width of 12.5 cm which makes in impractical for certain devices and battery sizes. There are also limitations on the battery that can be used with the prior disclosure due to metal pieces inside the battery, which tend to absorb RF energy. The invention disclosed herein does not suffer from the described limitations of the antenna and battery and is applicable with any antenna and battery type.

[0009] U.S. Pat. No. 5,994,871 discloses a system for charging a secondary cell using transmitted microwave energy, however, it relies solely on the use of an existing microwave oven as the charging means.

[0010] U.S. Pat. No. 5,982,139 discloses a remote charging system for a vehicle capable of transmitting electrical energy to the electrical storage device via a wireless energy beam. The referenced patent relies on the beam being able to track the vehicle or vehicles, which limits the number of devices (vehicles) by the number of beams available at a power transmitting location. The invention disclosed herein does not suffer from this limitation and is capable of transmitting power to many devices using a single beam.

[0011] U.S. Pat. Nos. 5,411,537, 5,733,313, and 5,769,877 describe systems for providing power to implantable biomedical devices by use of electromagnetic induction techniques using power transmitting and receiving coils. The devices in these patents are limited by the fact that they can only receive power in the inductive or near-field region. Additionally, the implanted devices have a dedicated transmitter with the purpose of supplying power solely to a single implanted device.

BRIEF SUMMARY OF THE INVENTION

[0012] The present invention pertains to a system for powering a mobile device having an energy harvester which receives wireless energy and converts the energy into current. The system comprises a first portal in which wireless energy is transmitted for the energy harvester of the mobile device to receive and convert it into current when the device is in the first portal. The system comprises a second portal separate and apart from the first portal with a gap between the first portal and the second portal, the second portal in which wireless energy is transmitted for the energy harvester of the mobile device to receive and convert it into current after the device has passed through the first portal and the gap and is in the second portal.

[0013] The present invention pertains to a method for powering a mobile device having an energy harvester which receives wireless energy and converts the energy into current. The method comprises the steps of moving the device into a first portal in which wireless energy is transmitted for the energy harvester of the mobile device to receive and convert it into current when the device is in the first portal. There is the step of converting the wireless energy transmitted in the first portal by the energy harvester of the mobile device into current when the device is in the first portal. There is the step of moving the device from the first portal to a second portal separate and apart from the first portal through a gap between the first portal and the second portal in which wireless energy is transmitted for the energy harvester of the mobile device to receive and convert it into current when the device is in the second portal. There is the step of converting the wireless energy transmitted in the second portal by the energy harvester of the mobile device into current when the device is in the second portal.

[0014] The present invention pertains to a method for powering a mobile device having an energy harvester which receives wireless energy and converts the energy into current. The method comprises the steps of moving the device into a portal in which wireless energy is transmitted for the energy harvester of the mobile device to receive and convert it into current when the device is in the portal. There is the step of converting the wireless energy transmitted in the portal by the energy harvester of the mobile device into current when the device is in the portal. There is the step of moving the device out of the portal. There is the step of moving the device back into the portal for the energy harvester of the mobile device to receive and convert the energy into current when the device is in the portal. There is the step of converting the wireless energy transmitted in the portal by the energy harvester of the mobile device into current when the device is back in the portal.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[0015] FIG. **1** shows a single RF power portal with an associated coverage area.

[0016] FIG. **2** shows multiple RF power portals with associated coverage areas and separated by a region without coverage.

[0017] FIG. **3** shows a single RF power portal with an associated coverage area with multiple devices.

[0018] FIG. **4** shows multiple RF power portals with associated coverage areas with multiple devices and separated by a region without coverage.

[0019] FIG. **5** shows multiple RF power portals with associated coverage areas with multiple devices and each separated by regions without coverage.

[0020] FIG. **6** shows an RF power portal implemented at a loading dock.

[0021] FIG. **7** shows an expanded RF power portal implemented in a loading dock.

[0022] FIG. **8** shows the implementation of a multiple load dock example.

DETAILED DESCRIPTION OF THE INVENTION

[0023] A complete understanding of the invention will be obtained from the following description when taken in connection with the accompanying drawing figures wherein like reference characters identify like parts throughout.

[0024] For purposes of the description hereinafter, the terms “upper”, “lower”, “right”, “left”, “vertical”, “horizontal”, “top”, “bottom”, and derivatives thereof shall relate to the invention as it is oriented in the drawing figures. However, it is to be understood that the invention may assume various alternative variations and step sequences, except where expressly specified to the contrary. It is also to be understood that the specific devices and processes illustrated in the attached drawings, and described in the following specification, are simply exemplary embodiments of the invention. Hence, specific dimensions and other physical characteristics related to the embodiments disclosed herein are not to be considered as limiting.

[0025] Referring now to the drawings wherein like reference numerals refer to similar or identical parts throughout the several views, and more specifically to FIGS. **1** and **2** thereof, there is shown a system **10** for powering a mobile device **12** having an energy harvester **14** which receives wireless energy and converts the energy into current. The system **10** comprises a first portal **16** in which wireless energy is transmitted for the energy harvester **14** of the mobile device **12** to receive and convert it into current when the device **12** is in the first portal **16**. The system **10** comprises a second portal **18** separate and apart from the first portal **16** with a gap **20** between the first portal **16** and the second portal **18**, the second portal **18** in which wireless energy is transmitted for the energy harvester **14** of the mobile device **12** to receive and convert it into current after the device **12** has passed through the first portal **16** and the gap **20** and is in the second portal **18**.

[0026] A method and apparatus for high efficiency rectification for various loads (energy harvester and energy harvesting), which are suitable for receiving the RF power distributed by the present invention, have been discussed in detail in U.S. Provisional Patent Application No. 60/729,792 and U.S. patent application Ser. No. 11/584,983, which are incorporated herein by reference.

[0027] Preferably, the portal is defined by a minimum field strength. The portal preferably has at least one wireless energy transmitter **22**, and could have a plurality of wireless transmitters. Preferably, the wireless energy transmitter **22** produces pulses of energy. The system **10** preferably includes a third portal **24** separate and apart from the first portal **16** and the second portal **18** with a second gap **26** between the third portal **24** and the second portal **18**. The third portal **24** in which wireless energy is transmitted for the energy harvester **14** of the mobile device **12** to receive and convert it into current after the device **12** has passed through the second portal **18** and the second gap **26** and is in the third portal **24**. Preferably, each portal has a wireless energy transmitter **22**. The device can include a charge storage element **34** in the device in communication with the energy harvester to receive the current from the harvester to charge the storage element **34**.

[0028] The present invention pertains to a method for powering a mobile device **12** having an energy harvester **14** which receives wireless energy and converts the energy into current. The method comprises the steps of moving the device **12** into a first portal **16** in which wireless energy is transmitted for the energy harvester **14** of the mobile device

12 to receive and convert it into current when the device **12** is in the first portal **16**. There is the step of converting the wireless energy transmitted in the first portal **16** by the energy harvester **14** of the mobile device **12** into current when the device **12** is in the first portal **16**. There is the step of moving the device **12** from the first portal **16** to a second portal **18** separate and apart from the first portal **16** through a gap **20** between the first portal **16** and the second portal **18** in which wireless energy is transmitted for the energy harvester **14** of the mobile device **12** to receive and convert it into current when the device **12** is in the second portal **18**. There is the step of converting the wireless energy transmitted in the second portal **18** by the energy harvester **14** of the mobile device **12** into current when the device **12** is in the second portal **18**.

[**0029**] There can be the step of communicating with the device **12** with a communication transceiver **36**. The communication transceiver **36** can be co-located with a wireless energy transmitter **22**. There can be the step of charging a charge storage element **34** in the device.

[**0030**] Preferably, there is the step of moving the device **12** from the second portal **18** to a third portal **24** separate and apart from the first portal **16** and the second portal **18** through a second gap **26** between the second portal **18** and the third portal **24** in which wireless energy is transmitted for the energy harvester **14** of the mobile device **12** to receive and convert it into current when the device **12** is in the third portal **24**. The moving the device **12** into the first portal **16** step preferably includes the step of moving the device **12** into the first portal **16** which is at a loading dock **28**.

[**0031**] The present invention pertains to a method for powering a mobile device **12** having an energy harvester **14** which receives wireless energy and converts the energy into current. The method comprises the steps of moving the device **12** into a portal in which wireless energy is transmitted for the energy harvester **14** of the mobile device **12** to receive and convert it into current when the device **12** is in the portal. There is the step of converting the wireless energy transmitted in the portal by the energy harvester **14** of the mobile device **12** into current when the device **12** is in the portal. There is the step of moving the device **12** out of the portal. There is the step of moving the device **12** back into the portal for the energy harvester **14** of the mobile device **12** to receive and convert the energy into current when the device **12** is in the portal. There is the step of converting the wireless energy transmitted in the portal by the energy harvester **14** of the mobile device **12** into current when the device **12** is back in the portal.

[**0032**] Preferably, the portal is defined by a minimum field strength. The portal preferably has a wireless energy transmitter **22**. There can be the step of communicating with the device **12** with a communication transceiver **36**. The communication transceiver **36** can be co-located with the energy transmitter **22**.

[**0033**] Preferably, the wireless energy transmitter **22** produces pulses of energy. The moving the device **12** into the portal step preferably includes the step of moving the device **12** into the portal which is at a loading dock **28**. Alternatively, the device **12** can be an RFID tag, and there can be the step of communicating data from the RFID tag to a data receiver **38**.

[**0034**] The moving the device **12** into the portal step includes the step of moving the device **12** into the portal which is at a recharging cabinet. There can be the step of eliminating interference with other equipment and protecting humans from the energy with shielding in the recharging cabinet.

[**0035**] Untethered devices **12** can be classified into two categories, stationary and mobile. A stationary device **12** is defined as, but not limited to, a device **12** that is designed not to move. A mobile device **12** is defined as, but not limited to, a device **12** that is designed with the ability to move or be moved. Mobile devices pose a problem when trying to implement a powering solution due to the constant change in the device **12** location. One solution for mobile devices was proposed in U.S. provisional patent application 60/683,991, "Power Transmission Network"; in U.S patent application Ser. No. 11/438,508, "Power Transmission Network"; and in U.S. provisional patent application 60/763,582, "Power Transmission Network and Method", all of which are incorporated by reference herein. The application describes how a Radio Frequency (RF) power-harvesting network could be setup to provide power to devices within a specific area. However, mobile devices may not always be confined to the coverage area. In this situation, once the device **12** leaves the area, it will no longer receive power, making the device **12** rely solely on its power supply, if one exists. The proposed invention addresses this issue by creating an RF power portal. The purpose of the power portal is to provide a device **12** with enough operating power so the device **12** can function while it is outside the coverage area until it needs to return to a power portal to obtain additional power. This method is analogous to recharging a cell phone. Once the battery charger has charged the battery, the cell phone can be removed from the charger. The cell phone is then free to move away from the charger and operate until it uses the charge stored in the battery. At this point, the cell phone must return to the battery charger to obtain more power. In this example, the RF power portal is analogous to the cell phone battery charger.

[**0036**] In the previous example, the cell phone must be directly connected (wired) to the charger in order for charge to be transferred. The proposed invention eliminates the need for a wired connection in order to transfer charge. The charge is transferred to the device **12** from the RF power portal preferably in the form of electromagnetic waves or RF energy. The device **12** contains an energy harvester **14**, which preferably is used to convert the electromagnetic waves captured by the antenna into usable current such as Direct Current (DC). This invention should not be confused with power transfer by inductive coupling, which requires the device **12** to be relatively close to the power transmission source. The RFID Handbook by the author Klaus Finkenzeller defines the inductive coupling region as distance between the transmitter and receiver of less than 0.16 times lambda where lambda is the wavelength of the RF wave. The proposed invention can obtain power in the near-field (sometimes referred to as inductive) region as well as the far-field region. The far-field region is distances greater than 0.16 times lambda. This means the RF power portal can cover a large area, such as entire room, or be localized to a portion of a room, such as around an office desk. One method of transmitting RF energy was described in U.S. provisional patent application 60/656,165, "Method, Apparatus and System for Power Transmission"; in U.S patent application Ser. No. 11/356,892, "Method, Apparatus and System for Power Transmission"; and in U.S. provisional patent application 60/758,018, “Pulse Transmission Method”, all of which are incorporated by reference herein, which describes the use of RF energy pulses to increase the efficiency of the energy harvester **14** in the receiving device **12**. The pulsing method can also be used with the invention proposed herein to provide a larger coverage area than those available for Continuous Wave (CW) implementations. The coverage area can be defined by an instantaneous, peak, or average field strength or power density with respect the power transmitter. As an example, the energy harvester **14** may require an electric field strength of one volt per meter in order to convert the electromagnetic energy to DC at an acceptable efficiency. For a CW power transmitter, this field strength may correspond to a distance from the power transmitter of six feet. However, the device **12** may be located at ten feet meaning it is outside the CW coverage area. If pulsing is used, a larger amount of power can be transmitter at a particular duty cycle in order to average to the same amount of power as the CW case, however, the peak electric field will be greater than the CW case at six feet although the average is the same. The electric field strength at ten feet will then have a peak value greater than one volt per meter, which will enable the energy harvester **14** to capture and convert energy.

[**0037**] The RF power portal can be configured in numerous ways. Several of these configurations are presented and should not be considered limiting but rather examples of the invention. The first embodiment of the RF power portal is a single transmitter having a single coverage area, AREA **1**. A coverage area (or volume) is defined by a minimum electric, magnetic, and/or electromagnetic field strength and is represented as a dashed line in all figures. It should be noted that the figures illustrate a circular coverage area, however, the shape of the coverage area may take on other shapes for different antenna types and environments. As an example, Area **1** in FIG. **1** may be defined as the area in which the electric field strength generated by Transmitter **1** (TX**1**) is greater than two volts per meter (2 V/m). It should be noted that TX**1** in FIG. **1** preferably contains an RF transmitter and an antenna. Subsequent figures may use the same transmitter block or may separate the transmitter and antenna. The transmitter will be represented in the figures by a TX block with or without a trailing number such as TX, TX**1**, TX**2**, etc.

[**0038**] For the configuration in FIG. **1**, the device **12**, RX**1**, is receiving operational and recharging power from the RF field produced by the RF power transmitter, TX**1**, while it is in the coverage area, AREA**1**. The device **12**, RX**1**, at time t**1** in FIG. **1**, is receiving not only operational power but also additional charge that may be stored in a charge storage element such as, but not limited to, a battery and/or capacitor, or other charge storage element, for use if the device **12** leaves the coverage area, AREA**1**. As an example, at time t**2** in FIG. **1**, the device **12** is operating from power captured while in the coverage area. When the charge captured at time t**1** has been depleted the device **12** is operating from charge stored previously in the charge storage element. As can be seen by examining FIG. **1**, the device **12** can continuously operate if the device **12** periodically returns to the coverage area to obtain additional charge. As an example, the device **12** may require 1 joule of energy to operate for a twenty-four hour period. If the device **12** then returns to the coverage area after the twenty-four hour period, it may be able to capture an amount of power equivalent to that used in the twenty-four hour period, 1 joule, by remaining in the coverage area for one hour. In this example, the device **12** can capture all charge needed for operation from the RF power transmitter, TX**1**. There may, however, be applications where the device **12** is away from the coverage area for longer periods of time than twenty-four hours, which can be combated by increasing the time the device **12** must remain in the coverage area. There may also be cases where the lifetime of the device **12** is a known finite value and therefore, the device **12** may not need to capture the total amount of charge used in the time away from the coverage area. Rather, the device **12** needs to capture an amount less than the charge used but enough to insure operation for the lifetime of the device’s battery or other storage element. There may also be instances where the device **12** may leave the coverage area for an unknown extended period of time. In which case, the RF power transmitter must fully charge the device **12** to the peak capacity of the charge storage element to insure the device **12** can function until it is able to return to the RF power portal.

[**0039**] Another embodiment of the RF power portal is shown in FIG. **2**. FIG. **2** shows multiple coverage areas separated by a distance, D. In this case, the device **12** may obtain power in the first coverage area, AREA**1**, at time t**1** from the RF power transmitter, TX**1**, for the purpose of having enough power to travel the distance, D, to the second coverage area, AREA**2**, where at time t**3** the device **12** would be able to capture charge from RF power transmitter, TX**2**, to replace the charge used during time t**2** outside the coverage areas, AREA**1** and AREA**2**. As an example, the device **12** may take twenty-four hours to travel the distance, D, between the first coverage area and the second coverage area and may use 10 joules of energy while outside the coverage areas. The first RF power transmitter, TX**1**, must therefore supply the device **12** with 10 joules of energy in order to travel the distance, D, between the two coverage areas. Once the device **12** reaches the second coverage area, AREA**2**, it can capture an amount of power equivalent to that used in the twenty-four hour period, 10 joules, by remaining in the coverage area for a given period of time.

[**0040**] It should be noted that the inventions in FIG. **1** and FIG. **2** may be used with more than one device **12** which is shown in FIG. **3** and FIG. **4**, respectively.

[**0041**] It should be noted that the invention may be extended to include multiple RF power transmitters each with an associated coverage area with each area having any number of devices which can move in and out of the coverage areas provided by the RF power transmitters. An expanded version of the invention is shown in FIG. **5**.

[**0042**] If found to be advantageous, the device **12** may communicate with a communication transceiver **36** located inside or outside the coverage area(s) while the device **12** is inside or outside the coverage area(s). A communication transceiver **36** is defined as a transmitter and/or receiver using wireless communication techniques to transfer data. A practical implementation of the invention with a communication transceiver **36** would be to have the RF power transmitter collocated with the communication transceiver **36** and have the device **12** communicate while it is inside the coverage area. Communication inside the coverage area allows the device **12** to use additional power to transmit/receive data because the RF power transmitter is supplying the device **12** with additional power to recharge its storage element, which could then be used to supply the additional power to turn on the communication transceiver **36**. The device **12** may transmit multiple data readings collected while it was outside the range of the communication transceiver **36**.

[**0043**] Specifically, the RF power portal is a defined area (or volume) where a device **12** can receive and store operational power for current and later use inside or outside of the coverage area. More specifically, the RF power portal may contain at least one RF power transmitter with at least one associated antenna that defines a coverage area where mobile devices with energy harvesters **14** may be located. The portal is designed so the device **12** will be able to capture enough power from the transmitter for operation during its time away from the RF power portal. To give farther insight into the described invention, several examples of implementation will be given. The portal can have a plurality of wireless energy transmitters, all of which are the same, or some of which are different in regard to the amplitude of the energy they output and/or the polarization, frequency or pulse of the output.

[**0044**] One implementation of an RF power portal is a battery recharging cabinet. The cabinet may be a new or existing structure where one or more RF transmitters are located in order to provide power to one or more devices. The cabinet or structure may or may not be shielded to minimize RF interference depending on its environment. A specific example of the battery recharging cabinet was designed for a hospital environment. The cabinet was shielded to eliminate interference with hospital equipment and was designed to recharge the battery or batteries in a patient-tracking device **12**. The shielding also allows the power to be elevated without risk of exposing humans to unsafe levels of RF energy as regulated by the Federal Communications Commission (FCC) and other standards bodies. The recharging cabinet was designed to eliminate the need for a hardwired recharging device **12** and to allow the nurse or other staff member to simply place the tracking device **12** inside the cabinet in any direction rather than in a recharging cradle. After a predetermined charging period, the device **12** can be removed from the cabinet for use in the hospital. The door of the cabinet was connected to a switch to control the operation of the transmitter, but this does not have to be the case. When the door is open, the transmitter power was turned off to eliminate any interference with hospital equipment. The power could also be turned down instead of off. The device **12** was required to operate for several days before recharging was required. The output power of the transmitter inside the cabinet was designed to provide the battery with enough charge to supply the required power to the device **12** for the average stay of a patient. The average patient stay was chosen because the device **12** can operate longer if needed. As an example, the average patient stay may be two days meaning the battery recharging cabinet will supply the battery with two days of charge. The patient, however, may stay in the hospital for five days. This does not mean the device **12** will fail to operate after the second day. After the second day, the device **12** will run off of the reserve charge stored in the battery, and the battery will begin to drain below a reference level. When the device **12** is returned to the cabinet after the fifth day, it will receive enough charge to operate for an additional two days. However, the next patient may only have a stay of one day. As can be seen by using the average, the power stored in the battery will average to the reference level allowing the device **12** to operate without problems.

[**0045**] It may also be advantageous to implement the system **10** described above in an existing structure within the hospital. One existing structure that was explored was the sterilization equipment. The patient-tracking device **12** needs to be sterilized after each use. The sterilization may be performed in an autoclave, by ethylene oxide, or by another sterilization system. In any case, the sterilization system **10** is a predefined location where devices are placed for a specific reason. The proposed idea was to use the sterilization system **10** to sterilize and recharge the patient-tracking devices. The method eliminated the step of placing the devices in the cabinet after sterilization. Rather, the devices could be recharged while they were being sterilized.

[**0046**] Another implementation of the RF power portal can be constructed in a loading dock **28**. The RF power portal can be used to provide charge to devices as they are exiting or entering the dock to or from delivery trucks **30**. In this case, interference with other equipment may not be a problem, so the RF power portal may not need shielding meaning the RF power portal could be open to allow devices **12** to pass through and not impede normal loading dock **28** functions. It should be noted that the invention has the ability of being transparent to the user by allowing mobile devices **12** to remain mobile rather than requiring them to be stationary for a given period of time in order to be recharged. As an example, the device **12** may be attached to a pallet that is being moved by a forklift **32**. The forklift **32** could simply drive through the RF power portal, which would then supply the device **12** with the required power as shown in FIG. **6**. The device **12** may be but is not limited to an active RFID tag. An active RFID tag is defined as a device **12** with power storage, which uses backscatter, transformer impedance reflection, or a standard antenna to antenna link to communicate with a data receiver **38**. The active RFID tag may be used to measure temperature of a product or may be connected to a cargo container to monitor the contents and other environmental factors such as tampering.

[**0047**] To better illustrate the operation of the RF power portal, consider a typical temperature sensor, which requires one joule of energy to transmit the temperature at a sampling rate of once per hour. If the device **12** will be away from an RF power portal for 12 hours, the device **12** will need to obtain 0.5 joules from the RF power portal for the 12-hour trip. If it is assumed that the device **12** is in the portal area for 60 seconds, the amount of power that must be provided by the portal is 0.5 joules divided by 60 seconds or 8.3 milli-joules per second or 8.3 milli-watts of continuous power. It may not be advantageous for the device **12** to be in the portal area for 60 seconds and therefore the portal area may be expanded to include the trucks **30**, which will be in the portal area for a time far greater than 60 seconds for loading purposes. The expansion of the RF power portal allows the devices **12** to charge not only in the loading dock **28** area but also inside the trucks **30**. The expanded RF power portal can be seen in FIG. **7**.

[**0048**] In a similar example shown in FIG. **8**, an active RFID tag or other sensing device **12** could obtain enough charge at a first loading dock **28** to ensure its operation for the duration of the travel to a second loading dock **28** where it would again receive power. In this application, charging the device **12** at discrete locations ensures that the device **12** can function without wired battery recharging or battery replacement throughout the lifetime or duration of the device **12**. It should be noted that the device **12** may have a power storage device **12** other than a battery such as, but not limited to, a capacitor.

[**0049**] The concept of an RF power portal can be expanded without limit to include greater than two areas. The increased number of RF power portals will decrease the demands on each individual RF power portal by decreasing the distance between portals thereby decreasing the amount of power that each area must supply to the device **12** for its travel between RF power portals.

[**0050**] For another example, a plurality of RF power portals may be located in different environments, for example, at a desk, in a car, at home on a kitchen counter, and/or at a coffee shop. A cell phone containing a device **12** may receive power from each or any of the RF power portals as the cell phone is within each power portal's coverage area. For example, a user may have the cell phone at work. The cell phone receives power from the power portal at the desk when in the coverage area for the transmitter of the desk power portal. When the user then enters his car, the cell phone would receive power from the power portal in the car. When the user places his cell phone on the kitchen counter, the cell phone would receive power from the power portal on the kitchen counter. When the user goes to the coffee shop, the cell phone would receive power from the power portal at the coffee shop. The cell phone would collect power from each power portal which would provide power to the cell phone during travel between the power portals.

[**0051**] Although the invention has been described in detail in the foregoing embodiments for the purpose of illustration, it is to be understood that such detail is solely for that purpose and that variations can be made therein by those skilled in the art without departing from the spirit and scope of the invention except as it may be described by the following claims.

The invention claimed is:

1. A system for powering a mobile device having an energy harvester which receives wireless energy and converts the energy into current comprising:

a first portal in which wireless energy is transmitted for the energy harvester of the mobile device to receive and convert it into current when the device is in the first portal; and

a second portal separate and apart from the first portal with a gap between the first portal and the second portal, the second portal in which wireless energy is transmitted for the energy harvester of the mobile device to receive and convert it into current after the device has passed through the first portal and the gap and is in the second portal.

2. A system as described in claim 1 wherein each portal is defined by a minimum field strength.

3. A system as described in claim 1 wherein each portal has a wireless energy transmitter.

4. A system as described in claim 3 wherein the wireless energy transmitters produce pulses of energy.

5. A system as described in claim 1 including a third portal separate and apart from the first portal and the second portal with a second gap between the third portal and the second portal, the third portal in which wireless energy is transmitted for the energy harvester of the mobile device to receive and convert it into current after the device has passed through the second portal and the second gap and is in the third portal.

6. A system as described in claim 5 wherein each portal has a wireless energy transmitter.

7. A method for powering a mobile device having an energy harvester which receives wireless energy and converts the energy into current comprising the steps of:

moving the device into a first portal in which wireless energy is transmitted for the energy harvester of the mobile device to receive and convert it into current when the device is in the first portal;

converting the wireless energy transmitted in the first portal by the energy harvester of the mobile device into current when the device is in the first portal;

moving the device from the first portal to a second portal separate and apart from the first portal through a gap between the first portal and the second portal in which wireless energy is transmitted for the energy harvester of the mobile device to receive and convert it into current when the device is in the second portal; and

converting the wireless energy transmitted in the second portal by the energy harvester of the mobile device into current when the device is in the second portal.

8. A method as described in claim 7 including the step of moving the device from the second portal to a third portal separate and apart from the first portal and the second portal through a second gap between the second portal and the third portal in which wireless energy is transmitted for the energy harvester of the mobile device to receive and convert it into current when the device is in the third portal.

9. A method as described in claim 7 wherein the moving the device into the first portal includes the step of moving the device into the first portal which is at a loading dock.

10. A method for powering a mobile device having an energy harvester which receives wireless energy and converts the energy into current comprising the steps of:

moving the device into a portal in which wireless energy is transmitted for the energy harvester of the mobile device to receive and convert it into current when the device is in the portal;

converting the wireless energy transmitted in the portal by the energy harvester of the mobile device into current when the device is in the portal;

moving the device out of the portal;

moving the device back into the portal for the energy harvester of the mobile device to receive and convert the energy into current when the device is in the portal; and

converting the wireless energy transmitted in the portal by the energy harvester of the mobile device into current when the device is back in the portal.

11. A method as described in claim 10 wherein the portal is defined by a minimum field strength.

12. A method as described in claim 10 wherein the portal has a wireless energy transmitter.

13. A method as described in claim 12 wherein the wireless energy transmitter produces pulses of energy.

14. A method as described in claim 13 wherein the moving the device into the portal includes the step of moving the device into the portal which is at a loading dock.

15. A method as described in claim 9 wherein the device is an RFID tag.

16. A method as described in claim 15 including the step of communicating data from the RFID tag to a data receiver.

17. A method as described in claim 7 including the step of communicating with the device with a communication transceiver.

18. A method as described in claim 17 wherein the communication transceiver is co-located with a wireless energy transmitter.

19. A method as described in claim 10 wherein the moving the device into the portal step includes the step of moving the device into the portal which is at a recharging cabinet.

20. A method as described in claim 19 including the step of eliminating interference with other equipment and protecting humans from the energy with shielding in the recharging cabinet.

21. A method as described in claim 10 including the step of communicating with the device with a communication transceiver.

22. A method as described in claim 21 wherein the communication transceiver is co-located with the energy transmitter.

23. A method as described in claim 12 wherein the portal has a plurality of wireless energy transmitters.

24. An apparatus as described in claim 6 wherein the portal has a plurality of wireless energy transmitters.

25. A method as described in claim 6 including the step of charging a charge storage element in the device.

26. A method as described in claim 10 including the step of charging a charge storage element in the device.

27. A system as described in claim 1 wherein the device includes a charge storage element in the device in communication with the energy harvester to receive the current from the harvester to charge the storage element.

* * * * *

Crazy Crystal Set

copyright © 2006 by Larry J Solomon

Experimenting with various circuits, and using the Mystery set as a starting base, I wanted to create a receiver with a double-ganged (oder *döppelganger*) capacitor. This turned out to be not as simple as it first seemed. Simply replacing the two variables with a döppelganger made the set untunable, because the two rotors are connected together and are inseparable. So, I began tinkering with the circuitry, trying this and that. The circuit that worked best seemed impossible. Indeed, some who see this will probably proclaim it to be impossible. To the contrary, it is not only possible, but a fine performer.

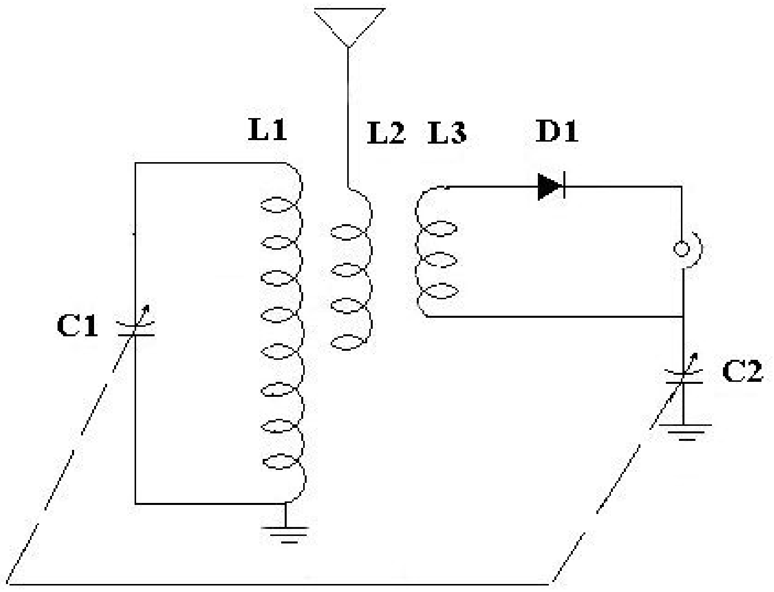

L1 - 45 turns 24 gauge magnet wire
L2 - 15 turns 26 gauge magnet wire wound on top of L1 and L3 separated by a layer of varnish
L3 - 25 turns 30 gauge magnet wire wound parallel to L1 after 12th turn of L1
C1-C2 500uF double-ganged air variable capacitor
D1 - 1N34 diode or rock crystal
Coil form 2.5 inch diameter x 2.5 inch long

I

I call this weird circuit the **Crazy Crystal Set**. Notice that there is not a single resistor or fixed capacitor. The astonishing thing is the dead end coil, L2, connected to an antenna but going nowhere! I have never seen such a circuit in print before. However, it works beautifully! Here are some of my photos.

<>

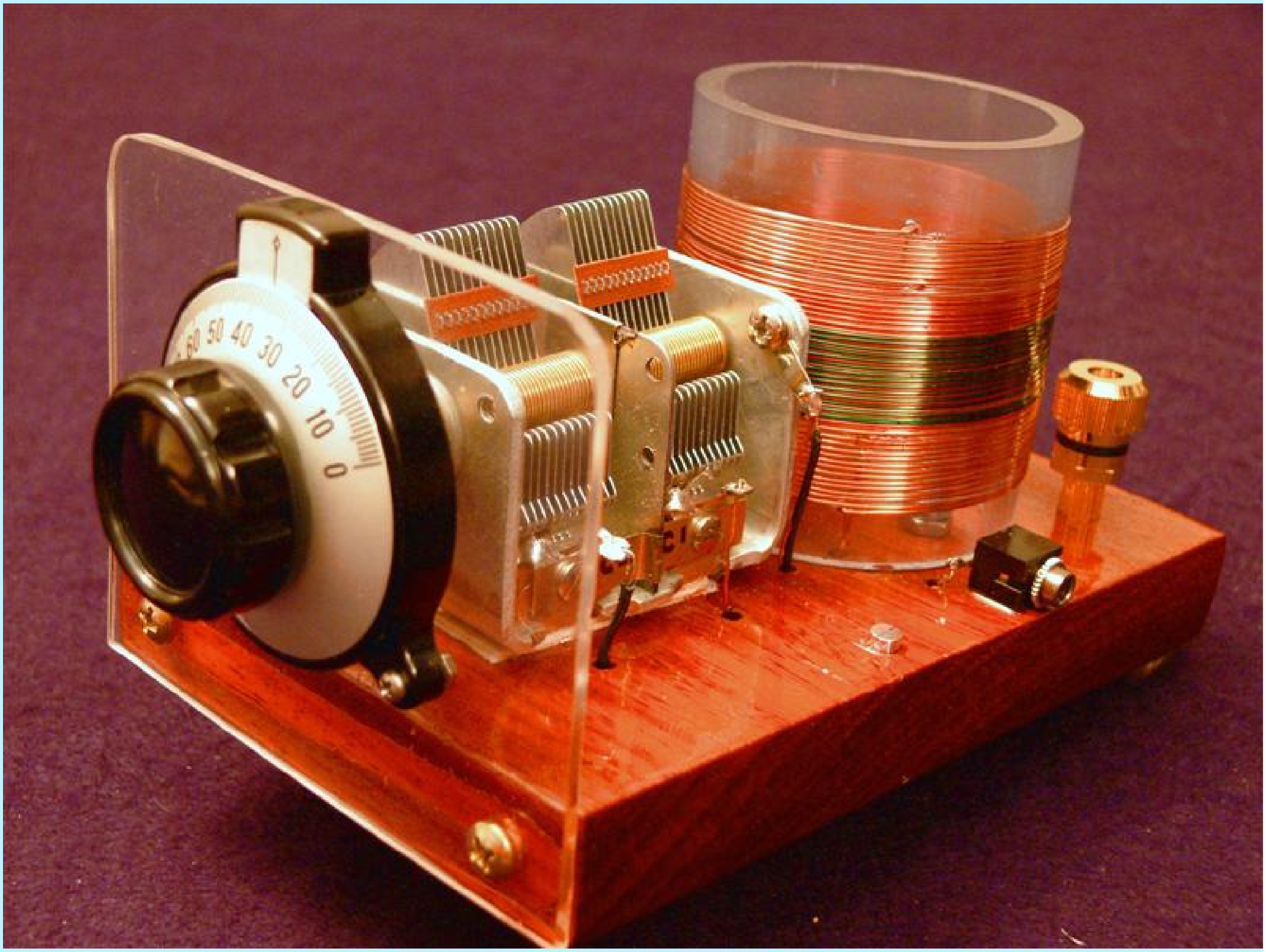

This shot shows the simplicity of the set, the vernier dial, coil, and capacitor.Most of the wiring is under the chassis. The phone jack can be used for high impedance phones or for audio output to magnetic phono (low level) input on a stereo amplifier. Sensitivity and selectivity are both good.

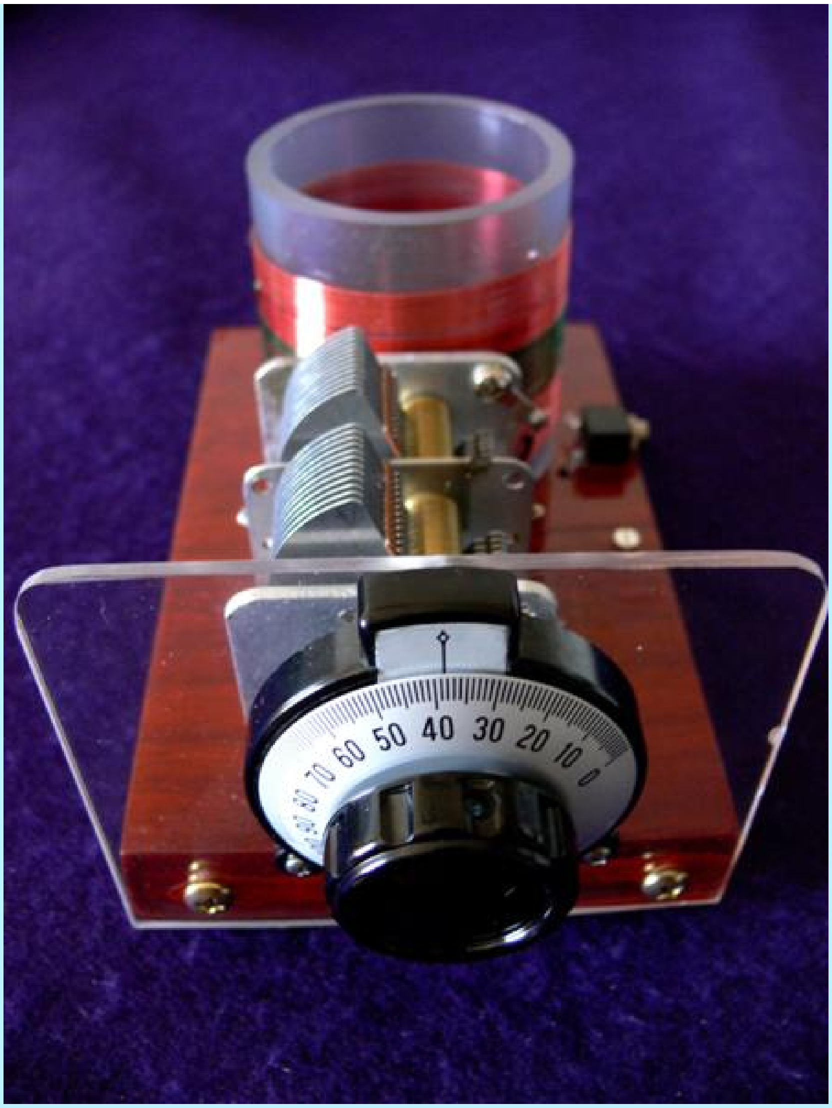
90
80
70
60
50
40
30
20
10
0

The whole view from the front showing vernier.

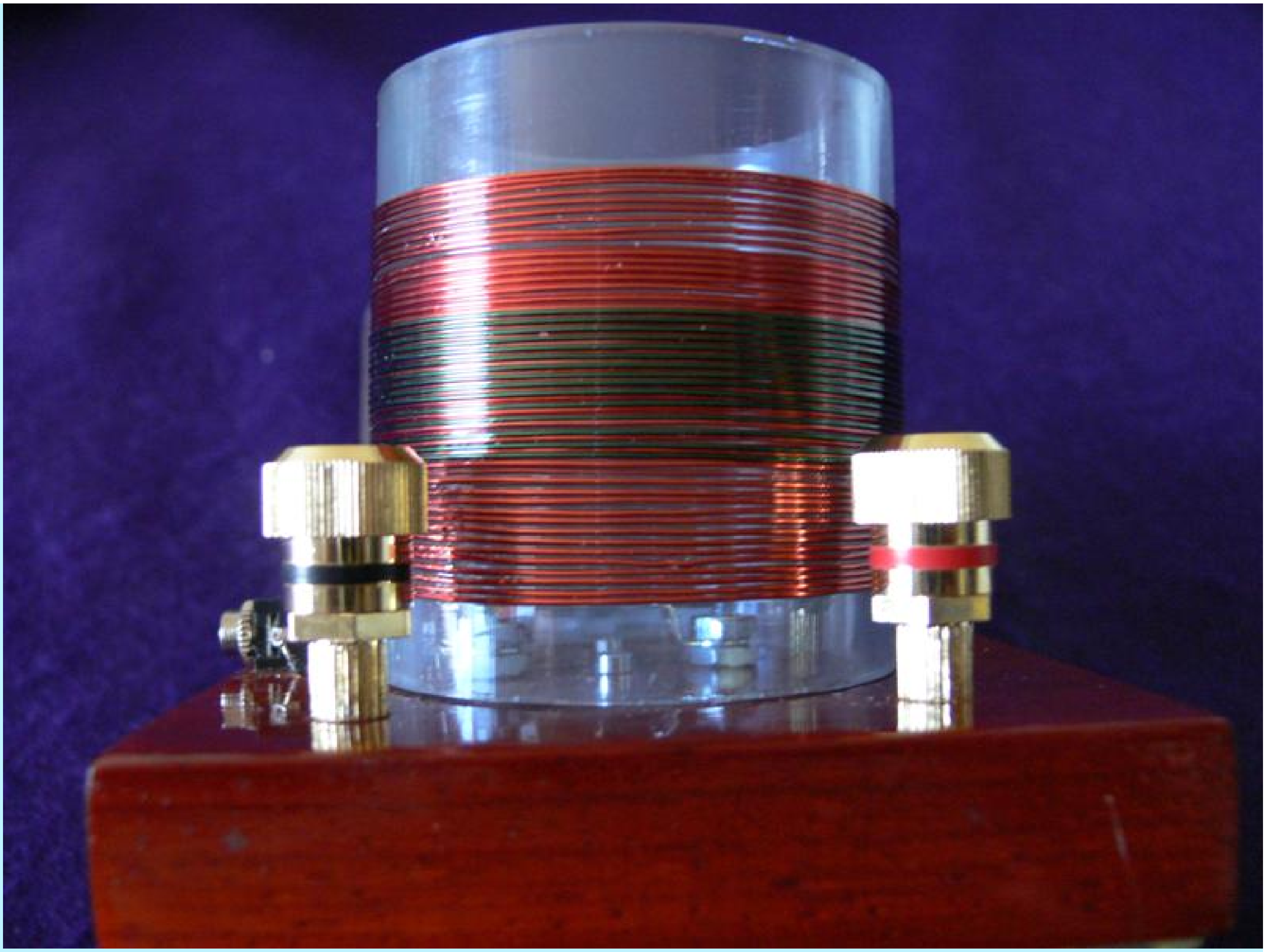

This close-up photo shows the view from the back with the space-wound coil. The crystal diode mount can barely be seen inside the coil. The antenna post is red, and the ground is black The green coil is L2..

Another back view all the way to the front.

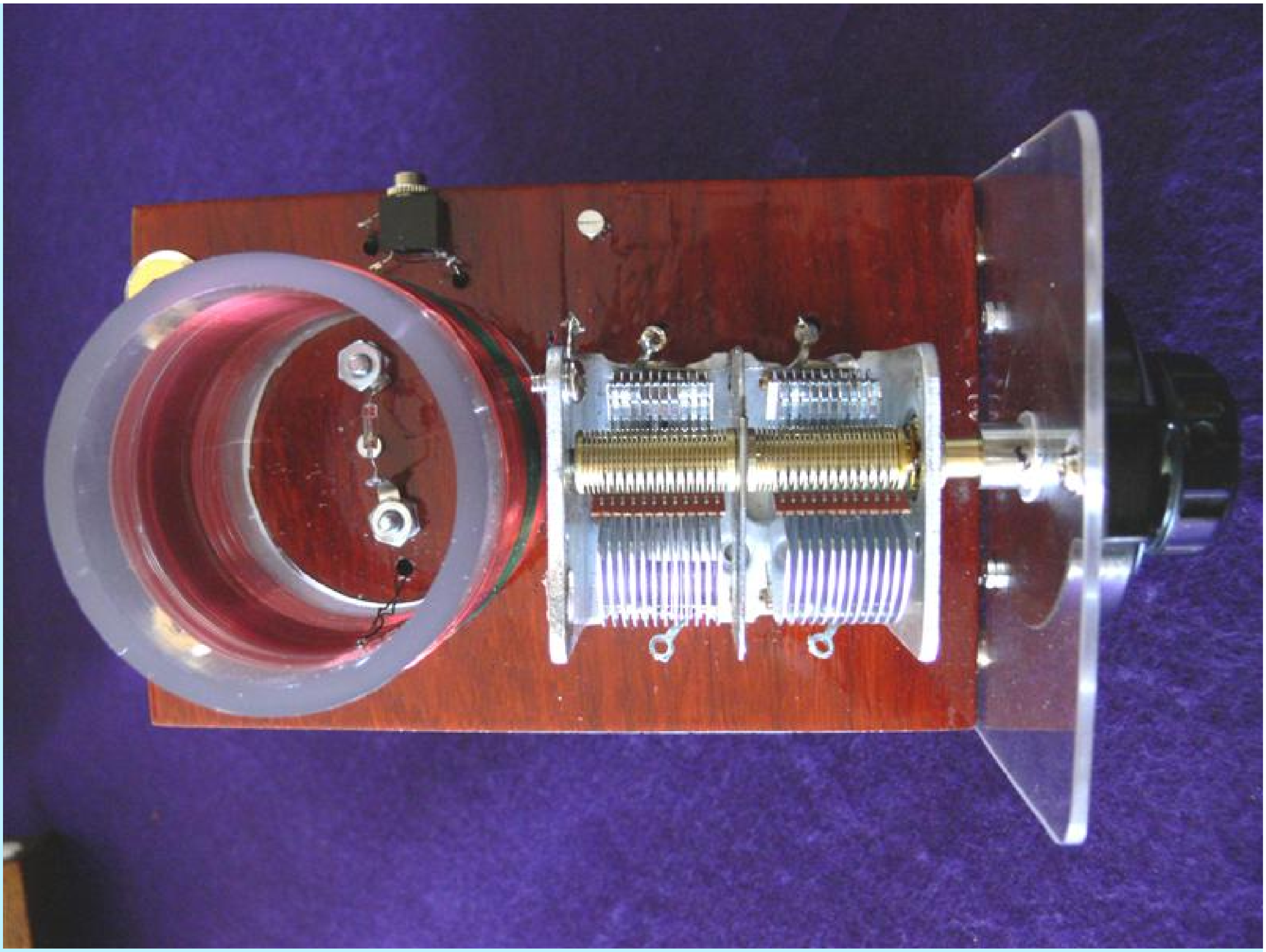

This is a view from above. Note the crystal inside the coil.

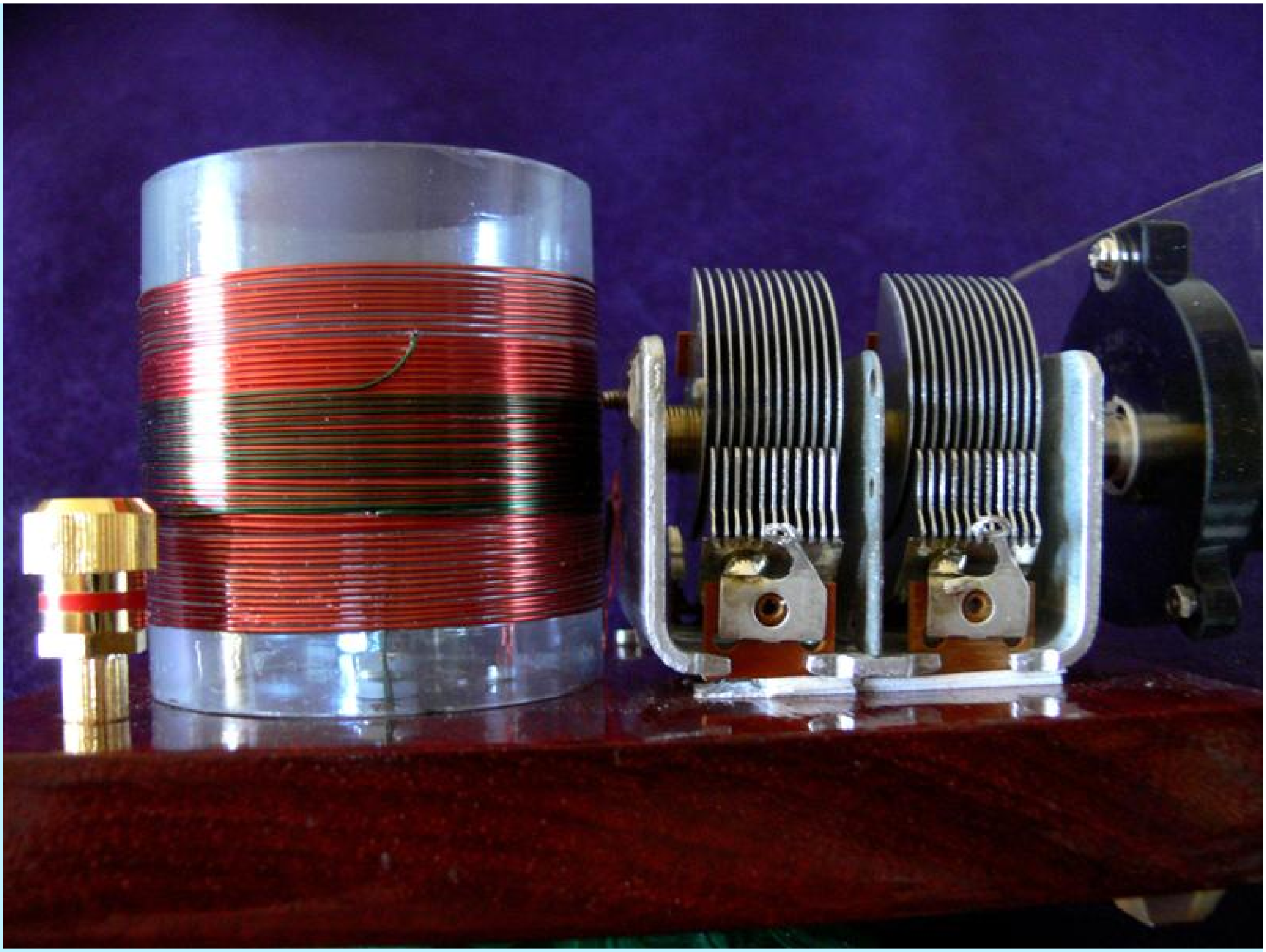

This view is from the left side.

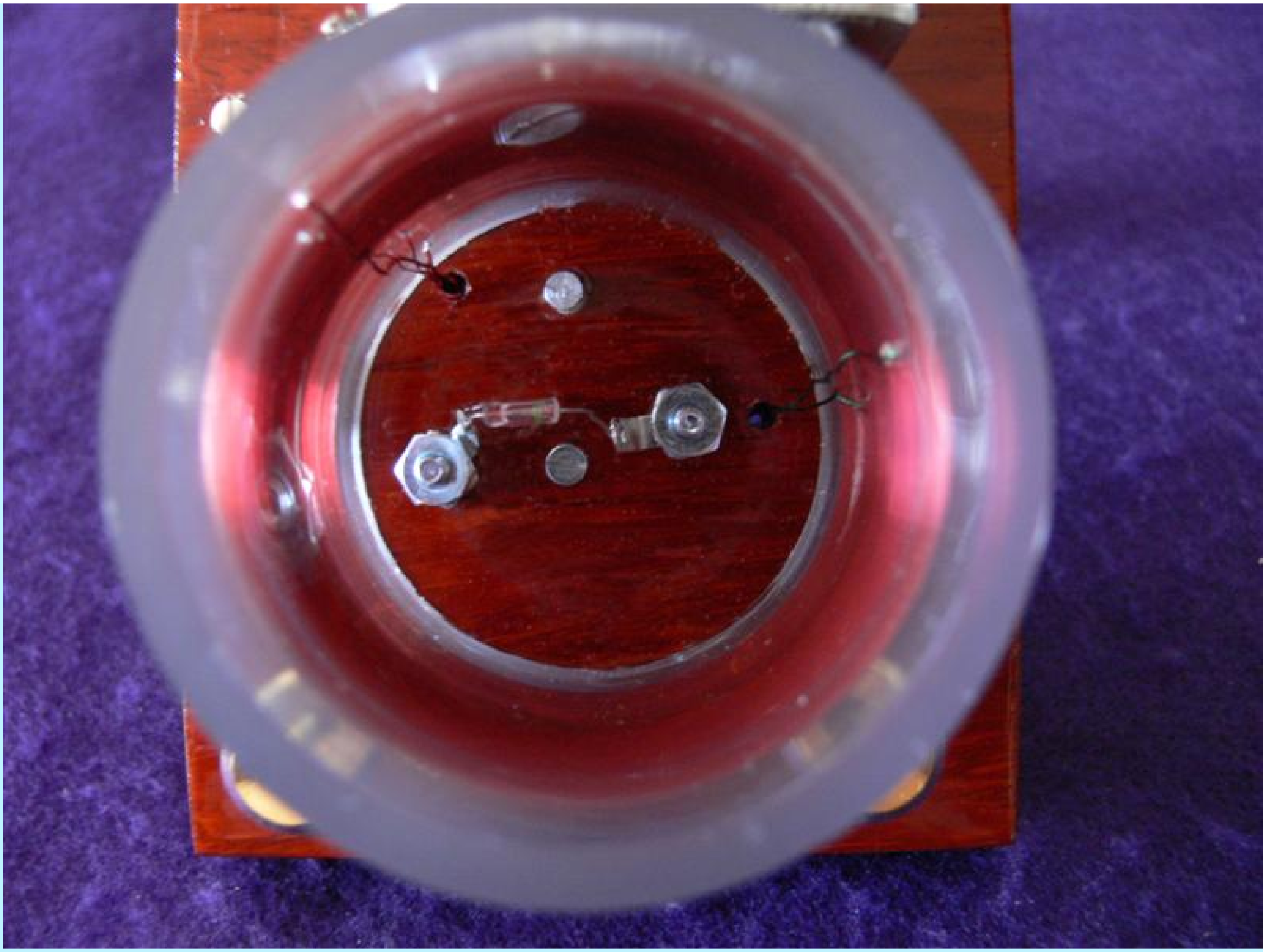

This is a closeup looking down into the coil, showing how the crystal is mounted.

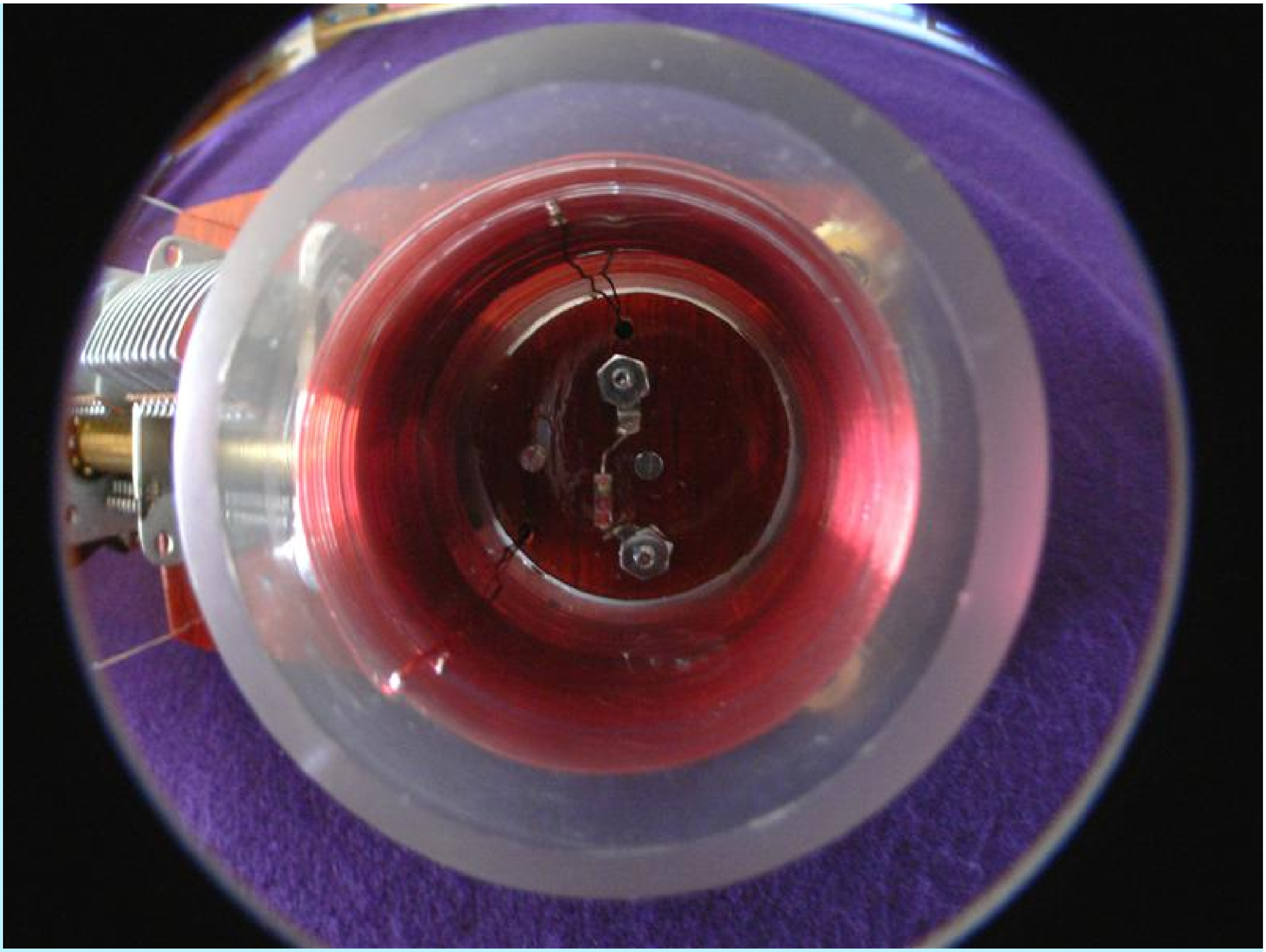

A fisheye view of the same.

Home

BIRMINGHAM, ALABAMA

CRYSTAL RADIO GROUP

crystalradio.us

- Home
- E-Mail
- B-Ham Sets
  - 80M Receiver
  - Brian's
  - Dee's Radio
  - Gil's Superhet
  - 80M Homebrew
  - Hop's Radio
  - Jack's 1AD
  - BC-221 Regen
  - JB's XtalSet
  - Lem's Regen
  - Lem's Rook Set
  - Jack's Bandpass
  - MosFet Set
  - Progressive
  - Dave's Regen
  - Steve's 221
  - Tom's 1 Tube
- Crystal
  - 2004-1
  - 2004-2
  - 2004-3
  - 2005-1
  - 2005-2
  - 2006-1
  - 2006-2
  - 2006-3
  - 2007-1
  - 2007-2
  - 2008-1
  - 2008-2
  - 2008-3
  - 2008-4
  - 2009-1
  - 2009-2
  - 2009-3
  - 2009-4
  - 2009-5
  - 2009-6
  - 2010-1
  - 2010-2
  - 2010-3
  - 2010-4
- 1AD Radios
  - 2005-1
  - 2006-1
  - 2006-2
  - 2007-1
  - 2007-2
  - 2008-1
  - 2009-1
  - 2009-2
  - 2009-3
- Contest Results

- 2004
- 2005
- 2006
- 2007
- 2008
- 2009
- 2010
- 2006 Sprint
- 2008 Sprint
- 2009 Sprint

Matching The Antenna

Many crystal sets use a variable capacitor in series with the antenna and the top of the antenna tank circuit (parallel coil/ capacitor) or between the ground and the antenna tank circuit. However, that is not the only way!

In the following pages we will look at some other approaches.

Lem's Antenna Matching Unit

Front of the unit. Lem used a black plastic clipboard from WalMart for the front panel.

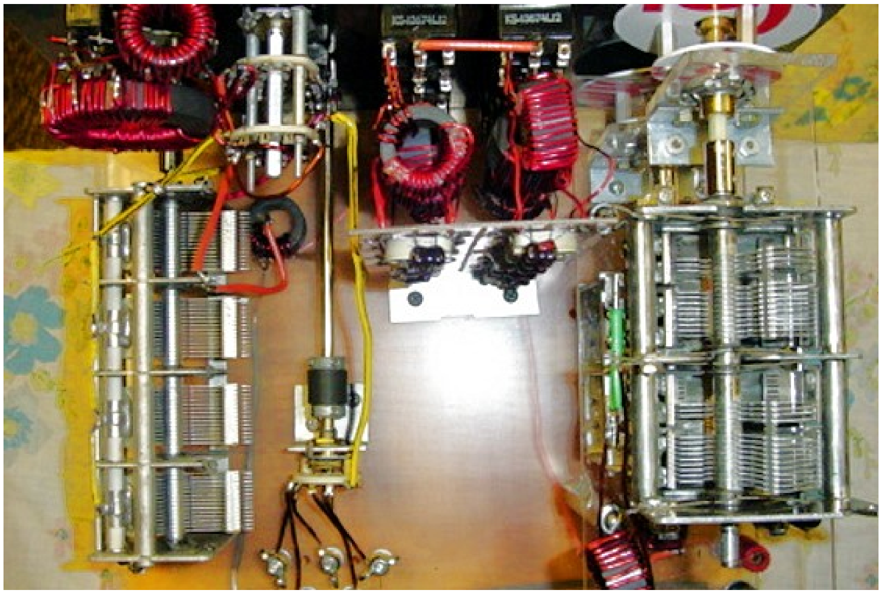

Top view showing lots of FT114A-61 cores.

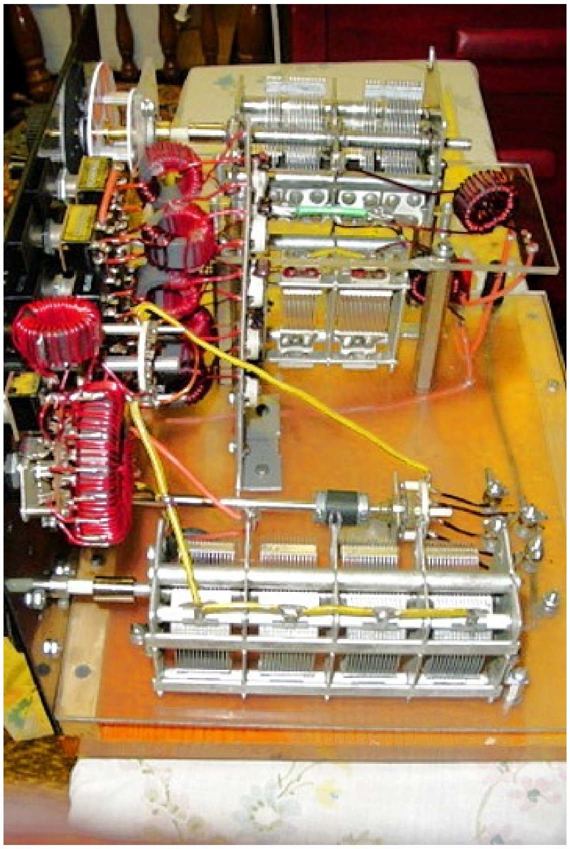

Right view showing the variable traps and series connected 8:1 verniers for an effective ratio of 64:1.

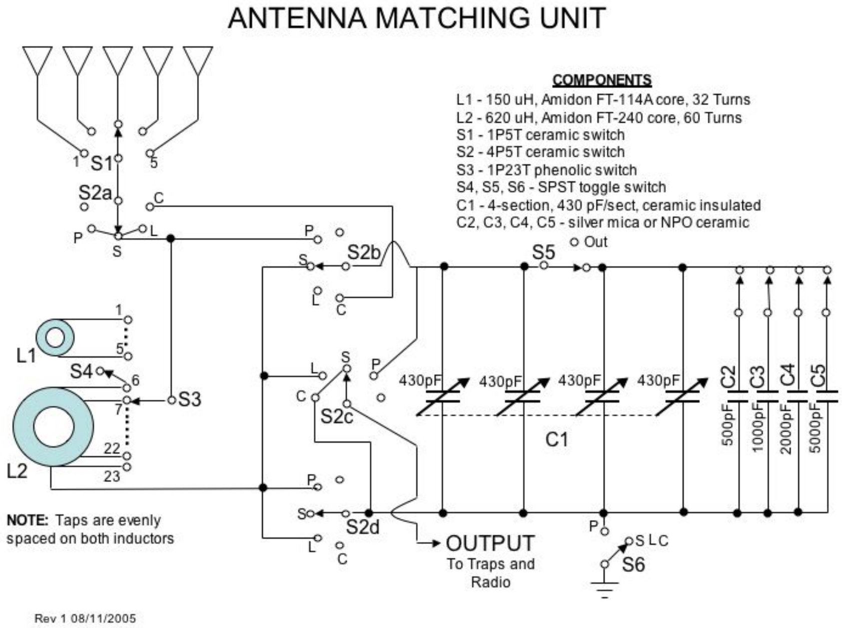

ANTENNA MATCHING UNIT

COMPONENTS
L1 - 150 uH, Amidon FT-114A core, 32 Turns
L2 - 620 uH, Amidon FT-240 core, 60 Turns
S1 - 1P5T ceramic switch
S2 - 4P5T ceramic switch
S3 - 1P23T phenolic switch
S4, S5, S6 - SPST toggle switch
C1 - 4-section, 430 pF/sect, ceramic insulated
C2, C3, C4, C5 - silver mica or NPO ceramic

NOTE: Taps are evenly spaced on both inductors

Rev 1 08/11/2005

1
S1
5
S2a
C
P
L
S
P
S
S2b
L
C
S5
Out
1
5
L1
S4
6
S3
7
22
L2
23
L
S
P
C
S2c
430pF
430pF
430pF
430pF
C1
C2
C3
C4
C5
500pF
1000pF
2000pF
5000pF
P
S
S2d
L
C
OUTPUT
To Traps and Radio
P
S L C
S6

[Click for large schematic view.](http://www.crystalradio.us/antennamatching/index.htm)

---

1 EMC Test Antennas - up to 40Ghz & up to 45dBi gain
High-end Broadband EMC, EMI, RF & Microwave Test Antennas,1Hz to 40GHz aaronia.com/antenna

2 Hafnium Wire and Rod
Hf wire and rod at factory price. High purity(99.95%)-Youtian Metal. njytmetal.com

Select Menu

- Home
- About Us
- Contest Results
- 1AD Contests
- 1AD Entries
- Antenna Matching
- Antennas
- BHAM Radios
- Crystal Contests
- Crystal Radio Entries
- Sprint Contest
- Headphones
- Projects, Mods
- Favorite Links
- Contact Jack

Airtable

a view for every work style

Grid

Kanban

Gallery

Calendar

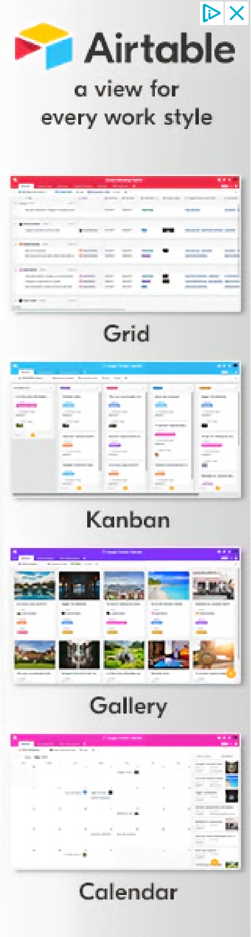
Airtable
a view for
every work style
Grid
Kanban
Gallery
Calendar

© Copyright 2000 – 2012, Jack Bryant, [crystalradio.us](). All rights reserved.

Website by [Dave Schmarder]()

CRYSTAL SETS
4
A Crystal Set By
Kenneth Rankin

Home | Contact | Site Map | Radio Stations & Memorabilia | Amateur Radio

Crystal Sets Introduction | Resistor & Capacitor Conversion Tables

Crystal Sets (Part1) | Build Your Own Crystal Set (Part 2)

Spider's Web Crystal Set (Part 3) | Experimental Crystal Sets (Part 5) | Crystal Radio Links

CRYSTAL SETS 4: A CRYSTAL SET BUILT BY KENNETH RANKIN

MDS975.co.uk reader Roddie Rankin kindly sent in the details of a crystal set built by his son Kenneth for a school project which won him great praise and merit. The crystal set is so nicely made and the project so well written that I thought it deserved a special page of its own.

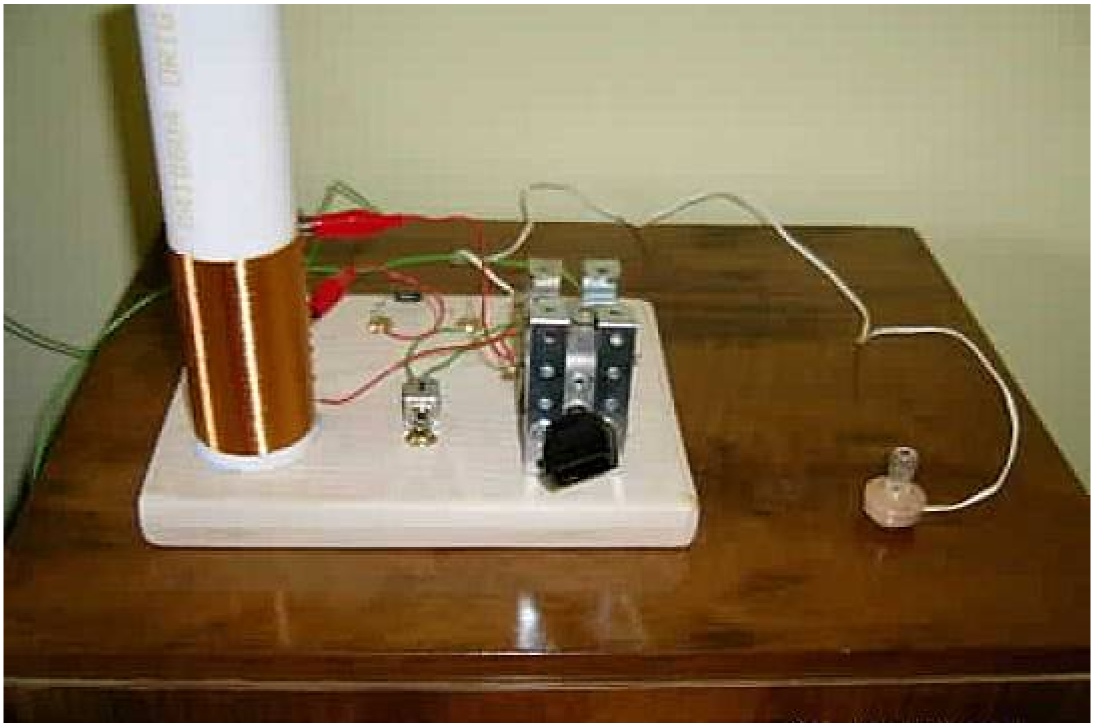

Photograph Of Kenneth Rankin's Rather Super Crystal Set

CONTENTS

1 - What Is A Crystal set?
2 - How A Crystal Set Works
3 - History Of Crystal Sets
4 - How I Built My Crystal Set
5 - The Results I Obtained
6 - Where I Found The Information
7 - Quiz

1 - WHAT IS A CRYSTAL SET?

Perhaps you think, like I did, that a crystal set is just a collection of lovely stones. In fact a crystal set is a very early, simple domestic radio. The photograph above shows the crystal set I built.

The crystal set does not have a battery. It runs completely from the radio waves it picks up in the air. It is absolutely free to run!

Please read on to find out more about this exciting invention!

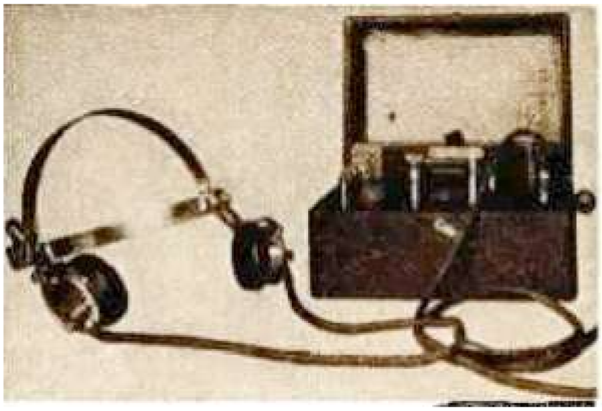

The birth of domestic radio: A deluxe crystal set produced by Marconi for the reception of programmes from the first of the BBC's radio stations, 2LO in London.
[from Pageant Of The Century - Odhams Press]

2 - HOW A CRYSTAL SET WORKS

The diagram below shows the circuit diagram of the crystal set I have built. It consists of a coil (L1), a tuning capacitor (C2), a diode (X1), an aerial, a crystal earpiece and an earth. (Capacitors C1 and C3 are not essential.)

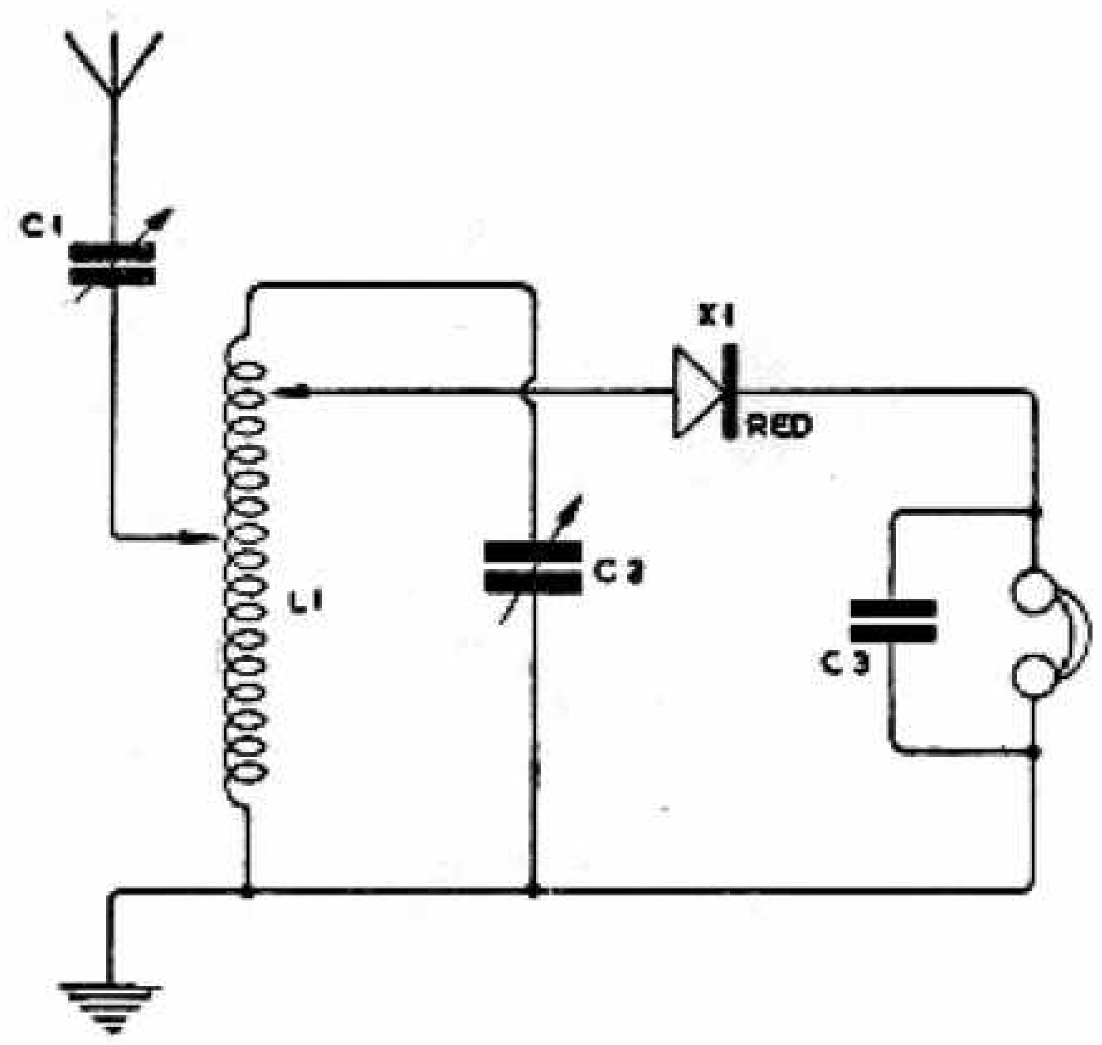

Circuit Diagram Of The Crystal Set

First the aerial picks up radio waves from radio stations all over the world. Then the required station needs to be selected, this is called tuning, and this is achieved with a Tuned Circuit. The Tuned Circuit in the crystal set consists of two items; a component called a variable capacitor and a coil of wire.

By adjusting the value of the capacitor and coil tuning is achieved. The coil taking unwanted signals of low frequency straight through to the earth, which gets rid of them, while the capacitor takes unwanted signals of high frequency straight through to the earth. The remaining signal, in between the high frequency and low frequency signals, is the signal of the required station which is not passed to earth by either the coil or the capacitor. This is the frequency to which the set is now tuned in.

You normally change the frequency to which a set is tuned in to by turning the knob on the tuning capacitor, thus receiving different stations by varying the capacitance of the variable capacitor. The number of turns on the coil can also be changed [called varying the inductance of the coil] and this will give different tuning ranges. A larger number of turns provided longer wavelengths [lower frequencies], while a smaller number of turns of wire provides shorter wavelengths [higher frequencies].

[It is also worth noting that the inductance of a coil can be increased by inserting a component called a Ferrite Rod into the centre of the coil.]

Once tuned into the required station this remaining signal goes to earth by travelling through the diode and earpiece. The diode takes the sound signal out of the radio signal [called demodulation], and passes it to the earpiece. The sound signal travels through the crystal in the earpiece which vibrates the air in the earpiece giving us the sound of the radio station to which you are tuned.

3 - THE HISTORY OF THE CRYSTAL SET

The history of crystal sets begins with the first broadcast stations being opened. This happened in England in 1922. The simplest form of receiver for these broadcasts was the crystal set.

In those early sets they had "cat’s whiskers" instead of diodes. A cat’s whisker was a fine piece of wire which was adjusted to make a suitable contact with a crystal. The crystal would be made of a substance such as coke or galena. An example from 1923 is shown in the picture below. Notice how well made these instruments were.

There were several limitations to the crystal set: it needed a big aerial, an earth connection, the clumsy cat’s whisker, and the weak signal could only be listened to by one person at a time with headphones. Very quickly the crystal set began to be replaced by valve radios with loudspeakers, powered by battery (see the picture from 1923).

In World War II, crystal sets were used by prisoners of war in prison camps to listen to news from home. As they were often badly treated it was hard to get all the right parts to build the set.

By this time crystal sets were no longer used in the home. Instead radios like the attractive 1937 model in the picture were being used.

Nowadays people only use crystal sets as a hobby. There are various different kits available on the market (see websites in the references below).

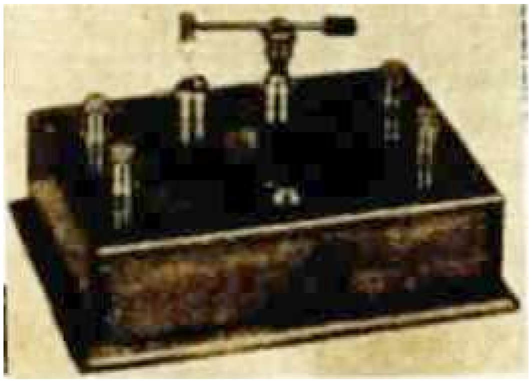

Photograph Of A Crystal Set From 1923

[from Pageant Of The Century - Odhams Press]

Crystal sets were primitive instruments used for 'listening in', and were fitted with the now famous cats whisker. Listening was done though a pair of headphones which meant that only one person could listen at a time.

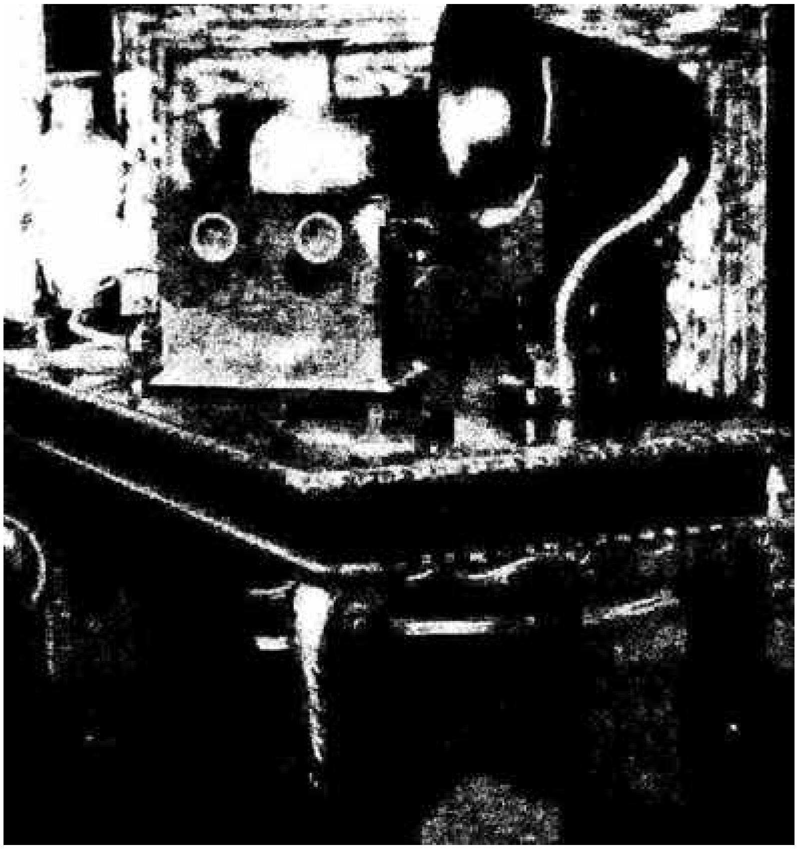

Photograph Of a 1923 Radio

[from Pageant Of The Century - Odhams Press]

Later valve amplification could be added to radios which allowed them to power a loudspeaker so that everyone in a room could listen in.

4 - HOW I BUILT MY CRYSTAL SET

The crystal set is a fun and easy thing to build.

The first thing I had to do was put up an aerial and earth. For the aerial I drilled a hole in the ceiling and fed wire through the hole. Then I went into the attic and took the wire to the other end of the attic, fastening it to rafters as I went. When I got to the other side I fed the wire out of an air vent to the outside of the house. I then went outside tied wire to a nearby tree and joined it up to the wire coming out of the air vent. This gave me an aerial of about 25 metres.

Next the earth. For this I drilled another hole in the ceiling and fed another wire through and out of an air vent at the other side of the house (see previous paragraph). I attached this wire to the metal frame of my swing. This is a good earth because its poles stick deep into the ground.

The coil used enamelled copper wire wound on to a piece of plastic pipe with a diameter of 5 cm. There were 120 turns of wire. Every 10 turns I made a loop on which a crocodile clip could be fixed. See the picture below:

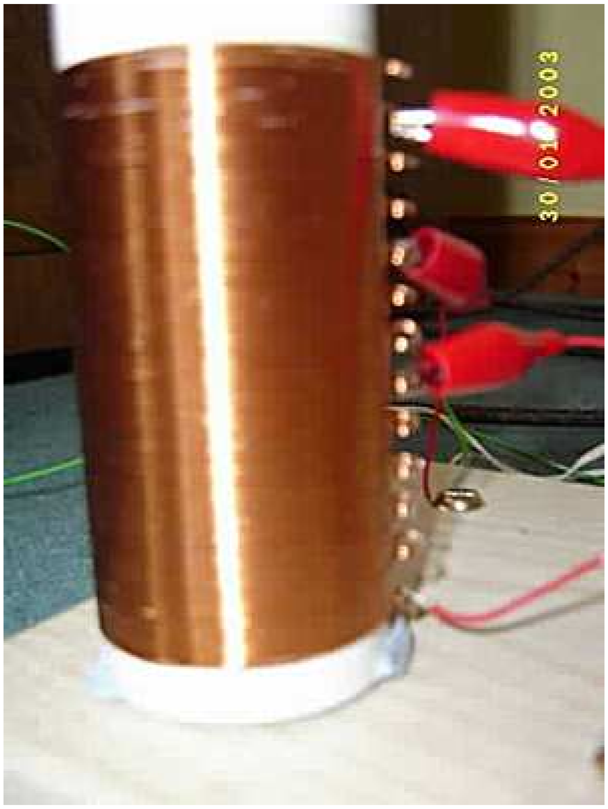

30/01/2003

Photograph Of The Coil Showing The Tapping Points & Crocodile Clips

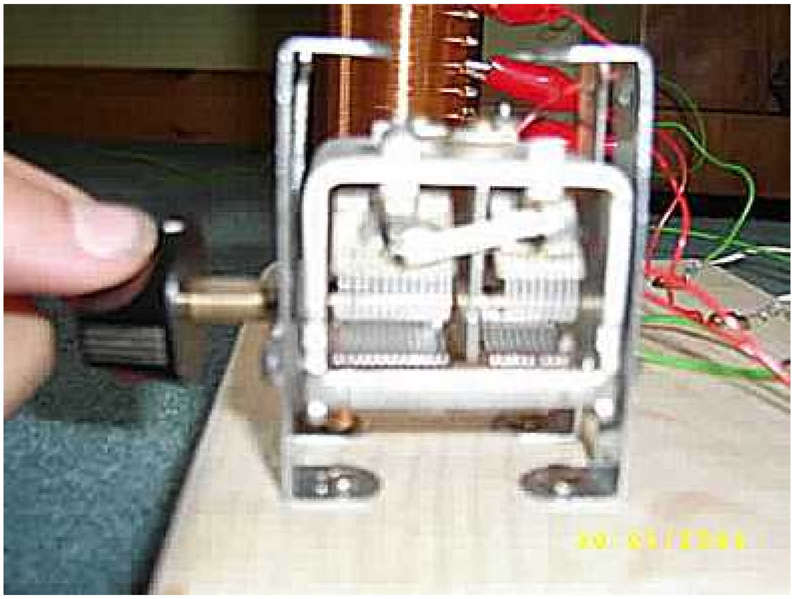

The Tuning Capacitor - Being Carefully Adjusted

For the crystal set itself, I first had to plan what I would be using and where the pieces would be placed (see picture and the plan drawing). I was using the circuit shown above. I built the crystal set on a wooden board about 20cm by 15cm. I drilled 4 holes 4cm apart. I put screws with cup washers in each hole and fastened the wires of the parts of the crystal set and the aerial and earth wires under the cup washers. I fastened the capacitor to the board using Meccano brackets and screws. The coil was fixed to the board using glue. The earpiece plugged into a socket which was fastened to the board using a thunderbolt shaped piece of Meccano.

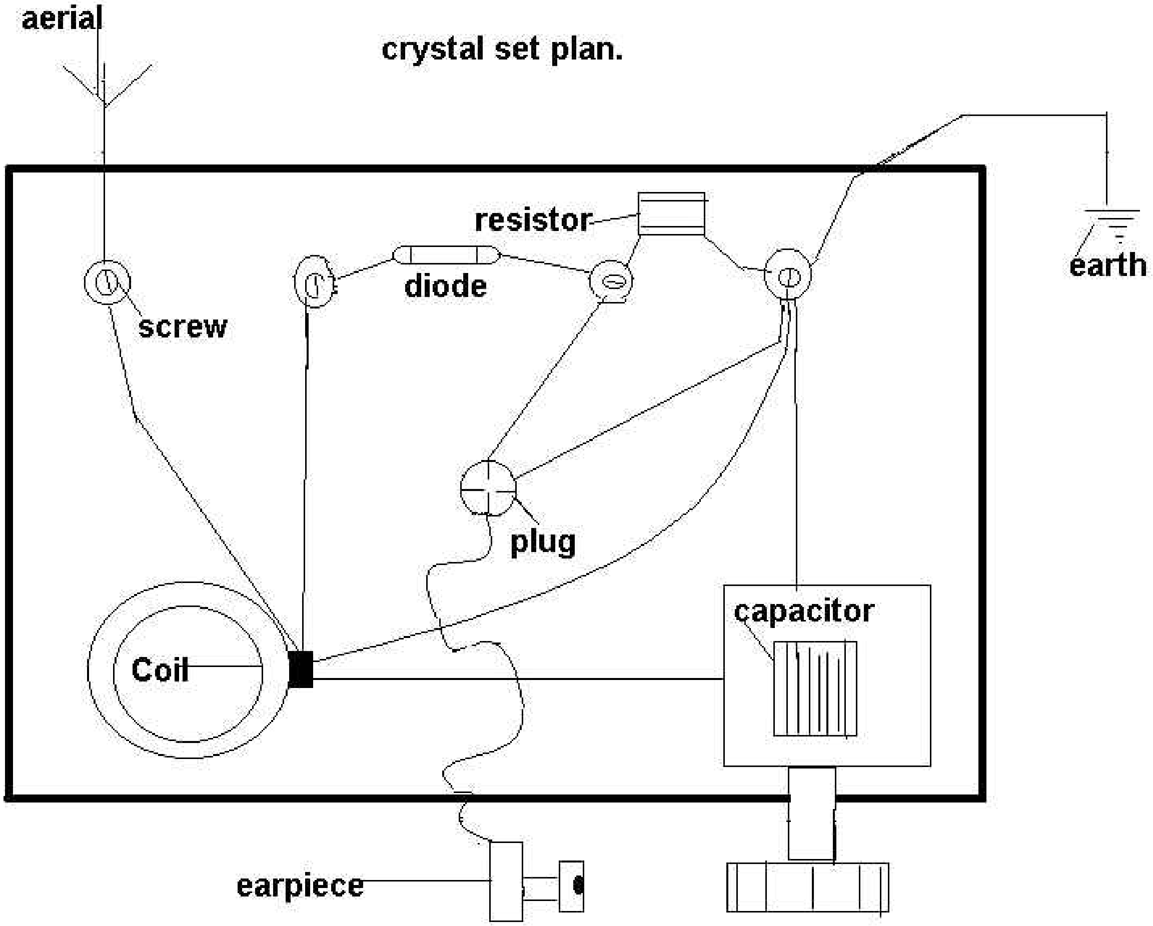
aerial
crystal set plan.
resistor
diode
screw
earth
plug
capacitor
Coil
earpiece

Drawing Showing Layout Of Components

5 - THE RESULTS I OBTAINED

At first all I could hear was a foreign station. But then, I tried late at night and could sometimes hear a station from Florida in English! By tuning the capacitor and moving the crocodile clips up and down the coil I managed to receive a few foreign stations on Short Wave. What I really wanted was to tune in to Medium Wave. I wanted to receive English stations like Radio Scotland and music stations. I tried everything including a longer aerial and using different parts but to no avail. I came to the conclusion that I could not get a strong enough Medium Wave signal in Northerly area of Scotland where we live.

Here is a picture of me operating my crystal set:

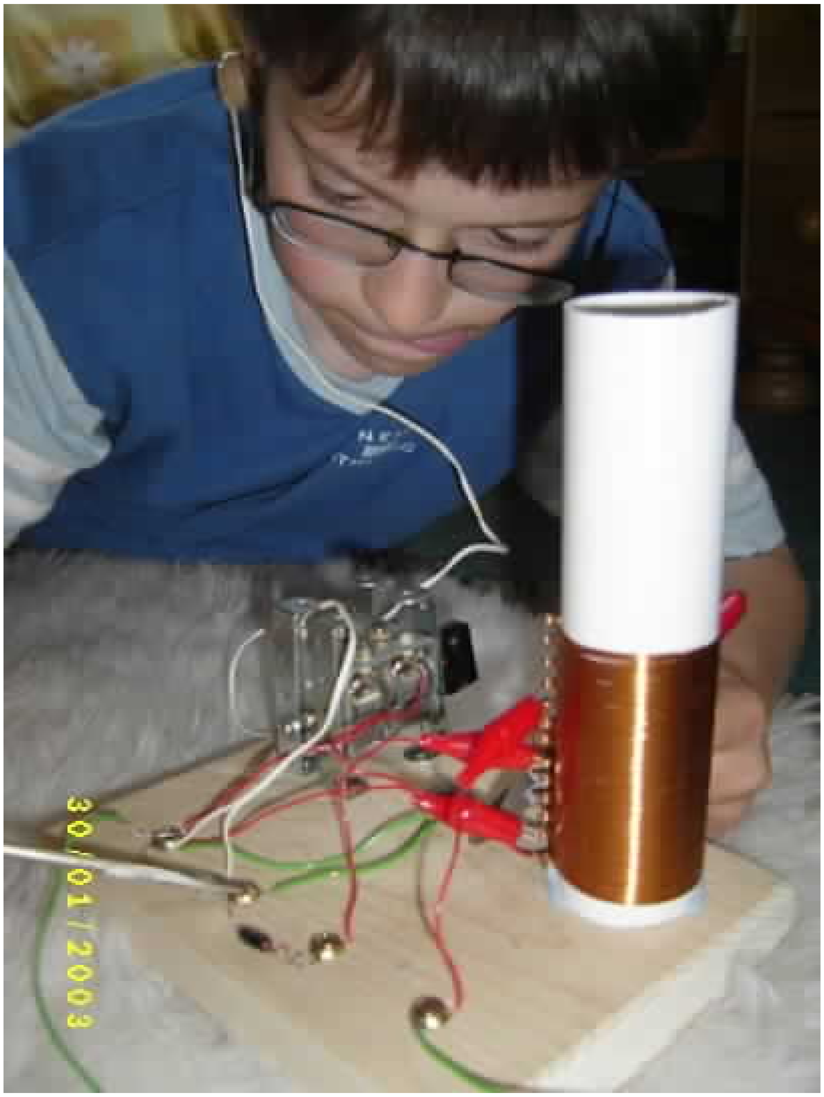

Photograph Showing The Crystal Set In Operation

6 - WHERE I FOUND THE INFORMATION

The History Of Crystal Sets

Book: The Pageant of the Century, Odhams Press, 1933.

www.oldradioworld.de/gollum/dhistor.htm

How I Built My Crystal Set:

Web page describing how to build a crystal set:

http://www.mds975.co.uk/Content/crystalsets2.html (THIS SITE)

Book:

"Making A Transistor Radio" written by Rev. G.C. Dobbs, Ladybird, 1972

Kits Are Available From:

www.midnightscience.com/kits.html

www.oldradioworld.de/gollum/xkits.htm

1N34A.com - crystal radio parts, including litz wire is http://1n34a.com

Other Hobby Websites:

Dave Schmarder's Crystal Set Pages http://www.makearadio.com/crystal/index.php http://www.makearadio.com

"Heart of England Crystal Radio Club" based in Birmingham, West Midlands - http://crystalradioclub.co.uk

More Crystal Set Links here >>

7 - QUIZ QUESTIONS

CRYSTAL SET QUIZ QUESTIONS:

1. What is a crystal set?
2. When were the first crystal sets sold?
3. Name 3 essential parts for a crystal set.
4. What does the earth do?
5. Does the coil pass signals of low frequency or high frequency?
6. How far apart were the screws on my crystal set?
7. What happened when I made the aerial longer?
8. What is this an anagram of: i a c c p r o t a?

(ANSWERS ARE FURTHER DOWN THIS PAGE)

PARTS THAT ARE REQUIRED

1	Wooden Base Panel
1	Coil Former 5cm dia. UPVC Pipe
1	Tuning Capacitor approx 500pF Air-Spaced or Solid Dielectric
1	Reel Of Enamelled Copper Wire for coil
1	Germanium Diode eg OA47 and IN34 which are especially suitable or other alternatives such as OA81, OA90 OA91, 1N94 etc
1	Crystal Earphone
1	47 k Ohm Resistor across earphone
6	½ inch No. 6 Brass Screws and Screw Cups
1	3.5mm mono Jack Socket
	Mounting Hardware (Meccano) for coil, capacitor and jack socket
	PVC covered wire for aerial and earth

A Superb RGD Valve Radio from 1937 - carefully restored in 1994 by Roddie Rankin

With thanks to Kenneth and Roddie Rankin for sending in this excellent article and photographs!

Crystal Sets Part 5 >

QUIZ ANSWERS

1. The first kind of radio
2. 1922.
3. Coil, diode, capacitor, earpiece, aerial, earth.
4. Gets rid of signals you don’t want.
5. Low.
6. 4 cm.
7. It did not make much difference.
8. Capacitor.

No AM radio stations or transmitters in your locality or country?

Spitfire
AM Transmitter

Has your local medium wave broadcast station closed or been moved to VHF/FM or Digital? Don't worry. You can still build and experiment with crystal sets and TRF radios by also buying or even building a simple low power AM transmitter. So, not only can you use your crystal sets but you can also run your own radio station that can be heard in and around your home - playing the music or programmes that you want to hear!

SSTRAN AMT3000 : Superb high fidelity medium wave AM transmitter kits from SSTRAN. Versions available for 10kHz spacing in the Americas (AMT3000 or AMT3000-SM) and 9kHz spacing in Europe and other areas (AMT3000-9 and AMT3000-9SM). Superb audio quality and a great and well designed little kit to build: http://www.sstran.com/pages/products.html

POWER
GAIN
MODULATION
COMPRESSION
SSTRAN
AMT3000

http://www.sstran.com/

Other AM transmitters available:

Spitfire & Metzo: Complete, high quality ready built medium wave AM Transmitters from Vintage Components:
http://www.vcomp.co.uk/index.htm Vintage Components offer a choice of the high quality Spitfire and Metzo transmitters:

SPITFIRE AM Medium Wave Transmitter with 100 milliwatt RF output power:

http://www.vcomp.co.uk/spitfire/spitfire.htm

Spitfire
AM Transmitter

METZO AM Medium Wave Transmitter with built in compressor:

Gain
Comp
Mod
Metzo

http://www.vcomp.co.uk/metzo/metzo.htm

AM88 LP A basic AM transmitter kit from North County Radio.
http://www.northcountryradio.com/Kitpages/am88.htm

LINKS:

BOWOOD ELECTRONICS - A friendly, helpful and very speedy source for many of your electronic components at prices that won't frighten your wallet!

THE FOXHOLE and P.O.W RADIOS - Simple crystal set receivers used by soldiers during the war and by prisoners of war (P.O.W.'s).

VINTAGE COMPONENTS - A great resource for crystal sets, components, valve radio kits and medium wave AM transmitters!

6V6 - Electronic Nostalgia and Vintage Components

MORE CRYSTAL RADIO LINKS HERE >>

Crystal Sets (Part1) | Build Your Own Crystal Set (Part 2)

Spider's Web Crystal Set (Part 3) | Experimental Crystal Sets (Part 5) | Crystal Radio Links

^Top Of Page

Amateur Radio

M0MTJ Visit my amateur radio pages M0MTJ

Visit my Amateur Radio Pages >

Home | Contact | Site Map | Radio Stations & Memorabilia | Amateur Radio

Crystal Sets Introduction | Resistor & Capacitor Conversion Tables

www.mds975.co.uk © 2005 - 2014

Jeffrey La Favre August 1, 2015

Understanding and Operating Your Crystal Radio

Your crystal radio receiver is designed to tune stations on the AM broadcast band (radio frequencies of 540 to 1610 kHz). You will be limited in the number of stations you can hear when operating the radio without amplification. Keep in mind that a crystal radio operates ONLY on the power received from the radio station transmitter. Therefore, your ability to hear a station will depend to a great extent on the distance from the radio transmitter to your house and the amount of power the station uses to broadcast their signal. Another important factor will be the antenna you use with your radio.

Most of our club members live in Geauga County. As far as I know, there is no radio station broadcasting on the AM band in our county (at least not a powerful station). From my home in Chesterland, the strongest station tuned on my crystal radio is WTAM. This station broadcasts with 50,000 watts of power on a frequency of 1100 kHz. The radio transmitter is located at 8200 Snowville Road, Brecksville, Ohio. The distance from the transmitter to Chesterland is about 23 miles (21 miles to Auburn Twp and 31 miles to Chardon). Figure 1 is an aerial view of the WTAM transmitting antenna, which is 480 feet tall.

Figure 1 WTAM transmitting antenna

As radio waves travel away from the transmitter, the power density of the radio energy varies inversely with the square of the distance. What does that mean? We can write this as a proportion:

$$power\ density \propto \frac{1}{distance^2}$$

Let me give some examples. The power density of a radio wave at a distance of 10 miles from the transmitter is:

$$power\ density \propto \frac{1}{10 \times 10} = \frac{1}{100} = 0.01$$

The power density at 20 miles is:

$$power\ density \propto \frac{1}{20 \times 20} = \frac{1}{400} = 0.0025$$

Therefore, the power density at 20 miles is 1/4th the power density at a distance of 10 miles. At a distance of 40 miles, the power density will be 1/16th the power density at 10 miles. All other things being equal, if your crystal radio is located 40 miles from the transmitter, the signal strength delivered to the radio will be 1/16th what it would be if located 10 miles from the transmitter. This is why you will have a better chance of hearing a station, the closer you are to its transmitter. Of course the amount of power used in transmitting makes a difference. If two stations are the same distance away, but one transmits with 1,000 watts of power and another transmits with 50,000 watts, the 50,000 watt station will be considerably louder than the one with 1,000 watts.

Antenna for the Crystal Radio

The antenna you connect to your crystal radio will determine to a great extent the number of stations you can hear. It is doubtful that an indoor antenna will be sufficient unless you live within a few miles of an AM radio transmitter. Provided that you have permission from your parents, you can install an outdoor antenna. You should ask your parents for assistance in erecting the antenna. This can be a temporary antenna, put up only when you are using the radio, or a more permanent installation. Again, you should ask your parents which kind of installation they will allow. Some people think that antennas are not attractive and you must keep this in mind.

The antenna that I use for my crystal radio is of the long wire type (14 gauge, stranded and insulated). It is permanently installed. I drilled a hole through the window frame to allow the antenna wire to exit from the GEAR workshop to the outside. I installed a hook on the edge of the roof and with an insulator, attached the wire antenna. The wire length from window to the edge of the roof is 11 feet. Then the wire runs to a tree 50 feet from the roof. The attachment to the tree is 18 feet above the ground. The wire then runs to another tree 61 feet from the first tree and also 18 feet above the ground. At the second tree the wire goes through a pulley and then down to the ground where it is attached to a weight. The total length of the antenna wire, from the window to the end is 140 feet (you will also need additional length inside the house to reach the radio). The pulley is used as part of the antenna tensioning system. When the wind moves the tree branches, the wire moves through the pulley to prevent breakage of the wire. The weight at the end of the wire provides the tensioning of the antenna so that it does not dip excessively toward the ground between its supports. This antenna works well and I can hear four or five stations during the day.

When you work 18 feet above the ground, there is danger of a serious or fatal injury from a fall. If you want to erect an antenna like the one I have described, **DO NOT ATTEMPT THIS ON YOUR OWN. PLEASE HAVE YOUR PARENTS HELP YOU**. I used a ladder to attach the antenna wire to the trees. This kind of work can be very dangerous and you must use proper safety procedures. The job takes a minimum of two people, one to hold the ladder while the other person climbs the ladder. The ladder should be secured with ropes so that it will not slip off the tree. It is also a good idea to use a safety harness that connects the climber to the ladder in case they slip on a ladder rung.

A safer procedure, although possibly more frustrating, is to throw a light line over a tree branch from the ground. Attach a weight to the end of a string and attempt to throw the weight over a branch. If that does not work you might try using a slingshot. It would be a good idea to wear a hard hat or bicycle helmet while doing this in case the thrown weight comes back down on your head! If you have a very long pole, you might use that as well to hoist the wire over a branch. These methods are preferred over the ladder since they are safer. After you are successful in getting a string line over a tree branch, connect the string to your antenna wire and pull the string to hoist the antenna wire in place.

Ground Connection for the Crystal Radio

In addition to the antenna, a connection to ground is required. The ground connection is actually part of the antenna system. For the antenna I have described, I used a grounding rod, which you can find in the electrical department at a hardware store. The rod is 8 feet long and all but a few inches are driven into the ground with a sledge hammer. Make sure there are no pipes or underground electrical wires in the area where you install the grounding rod. It is best to install the rod near where the antenna wire exits the house. The rod is copper-clad steel. A special bronze clamp is used to attach a ground wire to the rod. The other end of this ground wire is connected to the ground connection (black wire) of the antenna coil of the radio. The grounding rod must be in moist soil to work well. Of course it is difficult to know how moist your soil is 8 feet below the surface. But if your antenna does not seem to be working very well, you may need to water the soil around the grounding rod.

If you are not sure you will use your radio very much, you may decide to erect a shorter outdoor antenna. I erected a temporary long wire antenna, 11 feet from window to edge of roof, 50 feet to a tree branch that was 7 feet off the ground, then ran the wire to the ground and tied to a weight (total of 68 feet of wire from the window to the end of the wire). For this antenna I did not use a grounding rod. Instead, I connected the ground wire of the antenna coil to the ground system of the house electrical wiring. This can be done by connecting a wire to the screw of a faceplate for an electrical receptacle – see Figure 2 (**DO NOT PUSH THE WIRE INTO ONE OF THE RECEPTICLE SLOTS AS THIS COULD RESULT IN A FATAL ELECTRIC SHOCK**).

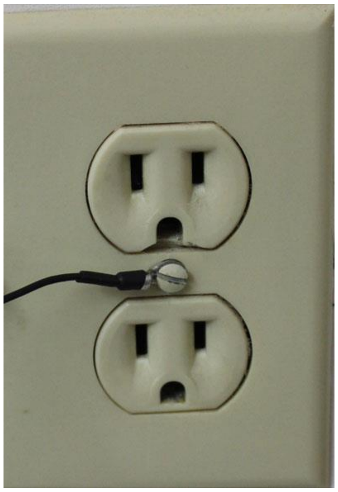

Figure 2 ground wire connected to faceplate

Using the shorter antenna and ground to faceplate, I could only hear station WTAM 1100. After you have completed the audio amplifier for your crystal radio, you will be able to hear more stations with the shorter antenna. However, if you have an interest in hearing distant stations, the longer antenna with grounding rod will be better.

When there is a chance of an electrical storm in your area, it is advisable to disconnect your antenna at a location outside your house. While lighting strikes on an antenna like yours are rare, it is better to be safe and prevent the very high current and voltage from entering your house. It is best if your exterior antenna is connected to ground outside the house when not in use.

Calibrating the Tuning Dial

The knob connected to the variable capacitor of the radio has a white pointing line to indicate the position of the variable capacitor. The position of the variable capacitor determines the radio frequency tuned (the station). However, the front face of the radio does not have any markings to indicate the frequencies for various positions of the knob. Mr. La Favre will help you mark the face of the radio at intervals of frequency. To do this we will use a signal generator. The signal generator will be connected to a coil that will substitute for the antenna coil. The signal generator can be adjusted to any radio frequency desired. In addition, the radio frequency can be modulated (AM - amplitude modulation) with an audio frequency, which you can hear as a tone in the radio earphone. We will start by adjusting the signal generator to a radio frequency of 600 kHz, modulated with an audio frequency of 400 Hz (400 Hz is a tone just a bit lower than the musical note A above middle C – that A is 440 Hz). The coil attached to the signal generator is placed next to the tuning coil of the radio. Then the tuning knob of the radio is adjusted until the 400 Hz tone is heard clearly in the radio earphone. The knob is adjusted to the position where the tone is heard with the highest volume. At that point a felt-tipped pen is used to place a mark on the front plate of the radio to match the position of the white line of the tuning knob. Then that line is labeled with the number 600 (600 kHz). We will continue by adjusting the signal generator in increments of 100 kHz, marking each knob position for each frequency. When finished, you will have marks on the face of the radio for 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500 and 1600 kHz. These marks will be helpful in adjusting the variable capacitor to tune a specific station.

Functions of the Various Components of the Crystal Radio

The most important reason for completing the Crystal Radio project is to learn how a radio works. Of course I also hope you will enjoy using your radio. To begin, let us take a look at the schematic for the radio provided below.

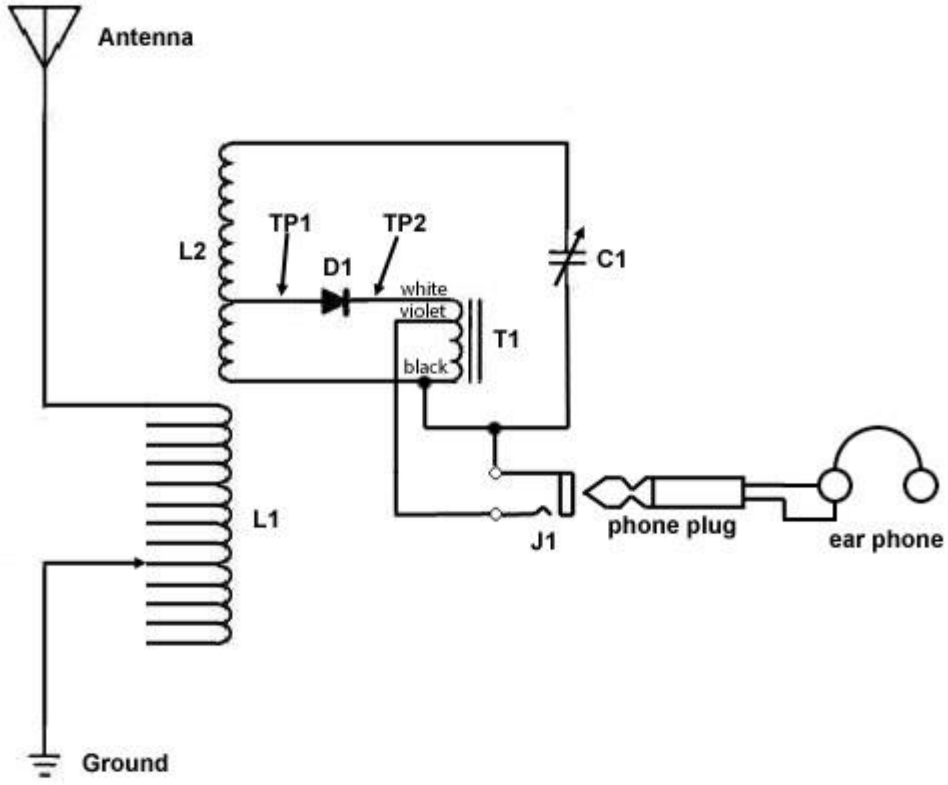

Figure 3 schematic of the Crystal Radio

L1 is the antenna coil and L2 is the tuning coil. The antenna coil has 12 taps. The ground wire is terminated with an alligator clip, which is used to connect to one of the taps. After tuning a specific station with the variable capacitor, C1, the ground wire is connected to the tap of coil L1 which gives the loudest signal as heard in the earphone. By selecting the best tap on coil L1, you are actually tuning the antenna to *resonate* at the tuned frequency. An antenna that is tuned to a specific radio frequency will deliver the strongest signal to the radio. When you adjust the variable capacitor to tune another station, you may need to adjust the tap connection on L1, especially if the new station tuned is of a significantly different frequency than the previously tuned station. Higher frequency stations will require connection to a tap closer to the antenna end of the coil. As you tune stations of lower frequency, you will need to adjust the tap connection to a tap further from the antenna end of the coil. In other words, a station of lower frequency will require the signal to travel through more coil turns than one of higher frequency. The exact tap to use will also depend on the length of your antenna.

Therefore, it is not possible for me to tell you which tap to use for a specific station. You will need to establish this by trial and error. If you are not able to find a tap on the coil that provides a stronger signal than other taps, then it is likely that your antenna is too short to tune to resonance. Nevertheless, this will not prevent you from hearing stronger stations.

Figure 4 below is a photo of the assembled radio. The green wire labeled A is the antenna wire and the black wire labeled B is the ground wire which connects to the grounding rod just outside the window of the GEAR workshop. D is the antenna coil (L1 in schematic). C is the alligator clip, attached to one of the taps of the coil. The red wire of the antenna coil connects to the antenna wire and the black wire connects to the ground wire. After making these connections, and placing the antenna coil (D) near the tuning coil (E), your radio is ready for use.

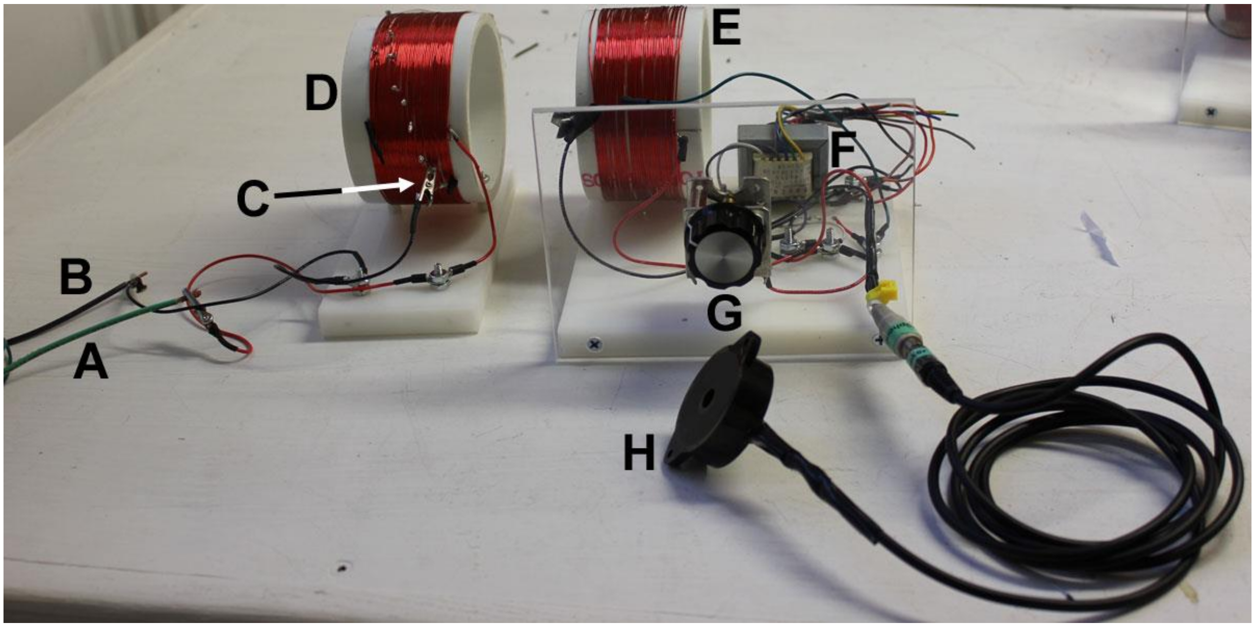

Figure 4 assembled crystal radio

F is the audio transformer, which is labeled as T1 in the schematic of Figure 3. The audio transformer is used to match the impedance of the earphone (H). G is the knob attached to the variable capacitor.

The radio signal present in the antenna coil (D) is transferred to the tuning coil (E) by the magnetic field generated by the antenna coil. When current flows through a coil, a magnetic field develops around the coil. If another coil is placed near that coil, then the second coil is also immersed in the magnetic field created by the first coil. This is the means for transferring the radio signal from the antenna coil to the tuning coil.

The transfer of the signal between the coils can only occur if the current flowing in the antenna coil is alternating current. As it happens, the current established in the antenna coil is alternating current. When an antenna captures radio waves, the radio waves are converted to an alternating current in the antenna circuit, which includes the antenna coil. As the current flows back and forth through the antenna coil, the magnetic field builds and then collapses around the antenna coil. It is the changing magnetic field which is required to transfer the electrical energy from the antenna coil to the tuning coil.

If the current flowing in the antenna coil was a constant direct current, there would be no transfer of electrical energy to the tuning coil.

The antenna coil and tuning coil together constitute a type of transformer. Transformers play important roles in electrical systems. The usual function of a transformer is conversion of voltage from one level to another. There is a transformer mounted to the electric pole near our house. It converts the high voltage (several thousand volts) of the electric line down to 240/120 volts for service to our house. In the case of our radio, we are not so much interested in converting the voltage of the signal in our antenna circuit as just transferring that signal to the tuning coil. We could just use one coil and connect the antenna, ground and variable capacitor to that coil. However, better results are obtained by using separate coils to tune the antenna and then transfer that tuned signal to the tuning coil via a magnetic field.

The antenna coil and tuning coil are mounted on separate bases so that you can adjust the distance between the coils. There will be a specific distance between the coils that results in the strongest signal received in the tuning coil and this may vary with the radio frequency tuned. The optimum spacing is not necessarily the closest spacing. As part of the process of tuning a station, you should adjust the space between the coils for maximum loudness heard in the earphone. In any case, the distance between the coils for best transfer of signal will be similar to that seen in Figure 4. For example, if you place the coils 12 inches apart, there will be little signal transfer.

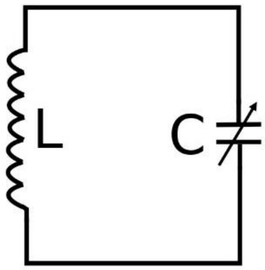

Figure 5 simple LC circuit with coil and variable capacitor

The variable capacitor C1 works together with the tuning coil (L2 schematic or E in photo) to tune a specific radio frequency. The tuning coil has a fixed amount of *inductance* and it is the adjustment in *capacitance* of C1 that allows tuning of different frequencies. Figure 5 shows a simplified version of the LC circuit of the crystal radio. The actual circuit for our radio contains a tap on the coil for diode D1 as well as other connections. But let us concentrate only on the coil L and capacitor C. In a certain respect the circuit of Figure 5 is an open circuit because current cannot pass through capacitor C. To be more precise, the circuit is open as far as direct current is concerned. However, we can consider the circuit closed in respect of alternating current. Capacitors behave in a fashion that allows current to flow temporarily in a circuit as the capacitor charges. With alternating current, the current flows for a short period only in one direction, especially at radio frequency rates of alternating current. As the current charges the capacitor, current is allowed to flow in the circuit. Once the capacitor is fully charged, current will no longer flow. What we must keep in mind is that at radio frequencies, current flows in one direction for a very short period of time. Let us take a frequency of 1000 kHz as an example, a frequency in the middle of the AM broadcast band. That frequency is equivalent to 1,000,000 cycles per second. Each cycle results in current flow in one direction followed by current flow in the opposite direction. The duration of flow in one direction is only 1/2,000,000th of a second or 0.5 microseconds. A microsecond is one millionth of a second. That is a very short amount of time. Even so, the electric current in the circuit in Figure 5 could travel around the circuit in much less time than 0.5 microseconds IF we did not include the capacitor and inductor (coil).

In order to tune the circuit for resonance at a specific radio frequency, we want the capacitor to charge and discharge at a rate that equals the frequency. If we removed the inductor (L) from the circuit, the capacitor would charge and discharge much faster than the rate required for tuning the circuit to resonance. By including the inductor, we slow down the rate of current flow, which in turn results in a slower charge and discharge of the capacitor. It is the interesting behavior of an inductor (coil) that we have employed to establish resonance. When voltage is first applied across an inductor, very little current is allowed to flow through the coil. Instead, a magnetic field starts to build around the coil. As time wears on (here we are talking about very little time), more current is allowed to flow through the inductor. The degree of hindrance in current flow by an inductor depends on the amount of inductance of the coil. We measure this in units named Henries, just as capacitance of a capacitor in measured in units of the Farad. An LC circuit containing a coil of a specific value of inductance and a capacitor of a specific value of capacitance will be resonant at a specific radio frequency.

Resonance can be a difficult thing to understand. Let us take the analogy of a child on a swing. As the child swings back and forth, the parent gives a push just as the backward movement stops. The end result of these timed pushes is that the child continues to swing. The swinging of the child is in resonance with the pushes given by the parent and the child will continue to swing indefinitely. Now suppose that the parent delivered the pushes while the child was still moving backward. In this case, with each cycle of the swinging back and forth, the amount of the swing will diminish until it ceases. In this case there is not resonance of the swinging and pushes.

The LC circuit is much the same. At a specific radio frequency, the frequency of resonance, the current flows back and forth in the circuit, timed exactly to match the rate of capacitor charging and discharging (the capacitor is somewhat like the parent pushing). At frequencies other than the resonance frequency, the current flowing back and forth in the circuit does not match the charging and discharging rate of the capacitor. Thus, the electrical energies of the non-resonant frequencies are absorbed in the LC circuit and not passed on to the remaining circuits of the radio. It is by this method that a signal of a specific frequency is allowed to pass through the radio while other frequency are not. If this did not happen, then we would be hearing many radio stations at one time.

Two important characteristics of a radio are **sensitivity** and **selectivity**. Sensitivity is the ability of a radio to deliver a sound that you can hear, derived from a weak radio signal received. Radios other than the crystal type include some type of signal amplification so that stations with weak signals can be heard. Sensitivity can be measured as the minimum signal voltage received that results in audible sound. Crystal radios have poor sensitivity because they rely only on the power received from the tuned radio waves. The efficiency of the radio antenna in capturing energy from the passing radio waves affects the signal strength received by the radio. Since crystal radios do not amplify the signal, it is especially important to use efficient antennas with crystal radios.

Crystal radios are also known in general for poor selectivity, although better designs have better selectivity. It is not uncommon for more than one station to be heard at a particular setting of the variable capacitor due to poor selectivity in the crystal radio. Nevertheless, you will be able to find some stations that you can hear clearly without interference from another station. It just depends on the separation in frequencies of the strong radio stations received at your location. More advanced radios, of the kind you are familiar with, have much better selectivity and usually tune only one station at a time, even when strong stations have frequencies very close to each other. I will not detail the reasons for this any further at this point as the improvements in selectivity require a discussion of circuits not found in crystal radios.

Figure 6 is a portion of the radio schematic showing part of the tuning coil, the detector diode D1 and the audio transformer T1. Two test points are marked on the schematic. We will connect an oscilloscope to these test points so that we can visualize the radio signal at these points in the radio. This will help in understanding the radio.

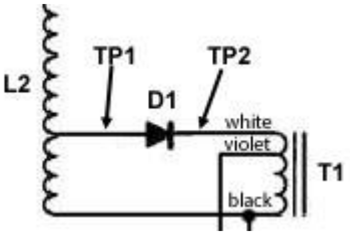

Figure 6 detector schematic of radio

An oscilloscope is an instrument that can be used to visualize rapidly changing electric signals. Figure 7 is the oscilloscope traces at test points one and two (see Figure 6 for locations of test points). A signal generator is connected to the radio providing a radio frequency of 1100 kHz, modulated by an audio frequency of 400 Hz. The blue trace displays the radio frequency at test point one modulated by the audio frequency. In the trace you can only see the waves of the audio signal. The waves of the radio frequency are so numerous that they appear as a solid blue color. It would be necessary to stretch the display to a great extent in the horizontal direction before we could see the individual radio frequency waves in the blue colored trace.

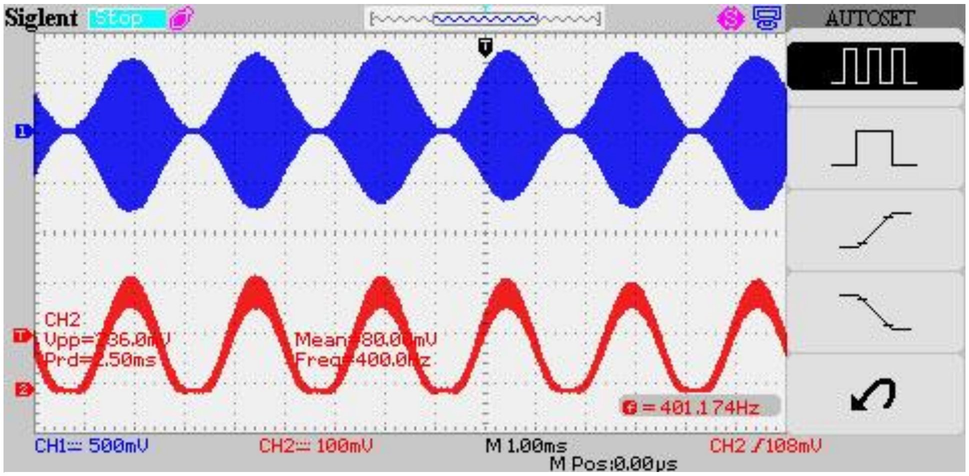

CH2
Vpp=136.0mV
Prd=2.50ms
Mean=80.00mV
Freq=400.0Hz
= 401.174Hz
CH1⎓ 500mV
CH2⎓ 100mV
M 1.00ms
M Pos:0.00μs
CH2 ╱108mV

Figure 7 oscilloscope traces at test point 1 (blue) and test point 2 (red)

The height of the blue radio waves is *modulated* by the audio signal (amplitude modulation or AM). The audio signal is much lower in frequency than the radio frequency (the number of audio modulations are much fewer than the number of radio frequency waves). Notice the blue tab labeled “1” on the left side of the blue trace. That tab marks the position of 0 volts. Any time the wave is above the tab, the voltage will be greater than 0. Any time the wave is below the tab, the voltage will be below 0 (a negative voltage). The radio frequency must have alternating positive and negative voltage in order to force periodic reversal in the direction of current flow (like quickly switching the connections to a battery). Otherwise, the current would not change directions and it would not be alternating current.

You should notice that the red trace, representing the signal at test point two, looks very different. Test point two is on the opposite side of the detector diode from test point one. Also note that for this test the diode connection to the transformer T1 was disconnected. Since the diode blocks current flow in one direction, we should not expect to see positions of negative and positive voltage on the wave trace here. In fact, notice where the 0 volt red tab with the label “2” is located. The bottoms of the red waves are at 0 volts. All positions of the waves are at 0 volts or higher because the diode has rectified the radio signal. We call this *detection*. The diode has converted the alternating current radio signal into a pulsating direct current audio signal. It is the audio signal that we need to apply to the earphone in order to hear the sound. If we were to just apply the radio frequency itself to the earphone we could not hear any sound. Humans are not capable of hearing sound at a frequency higher than 20,000 Hz.

After passing through the detector diode D1, the extracted audio signal passes through the audio transformer. The transformer transfers the audio signal to the earphone where the electric signal is converted into sound waves that you can hear. The transformer performs an important function in matching the *impedance* requirement of the tuning circuit to the impedance requirement of the device used to convert the electric signal to sound. The transformer has many taps that represent different impedances. The tuning circuit is attached to the white and black wires of the transformer, which provide about 40,000 ohms of impedance. The earphone is attached to the violet and black wires of the transformer, which provides 10,000 ohms of impedance. This is the best impedance matching I have found by listening to the signal in the earphone and selecting the loudest sound produced when testing all possible connections on the transformer. If you happen to have another high impedance earphone or headset you would like to use with your radio, you may need to use another tap on the audio transformer to connect to the device. The black wire should always be used as one connection. The gray wire provides 20,000 ohms, the blue wire 5,000 ohms, the green wire 2,500 ohms, the yellow wire 1,200 ohms, the orange wire 600 ohms and the red wire 300 ohms. I happen to have a very old pair of high impedance headphones from about 1920 that are designed for crystal radios. It works well when connected to the yellow wire of 1,200 ohms impedance. Modern headsets and earbuds are mostly low impedance devices that will not work with the crystal radio. However, if you pass the signal through the amplifier you will build (which has a high impedance input and low impedance output), then you can use your modern headphones to listen to your crystal radio. The amplifier works well using the 10,000 ohm violet wire connection to the transformer, the same as the earphone supplied with the radio kit.

Impedance is a measure of the resistance to alternating current that a device presents to a connected signal. It is not the exact same thing as a pure resistance, which we can call ohmic resistance. Impedance also accounts for something called *reactance*. Inductors (coils) have reactance, a kind of resistance to current flow as already mentioned. Impedance is a measure of resistance that accounts for the combined effects of ohmic resistance (like a resistor) and reactance (like a coil). In fact, the speakers that are used in radios that you typically use employ a voice coil attached to a diaphragm to convert an electrical signal to sound. The long length of wire used in the coil presents some ohmic resistance to current flow. Since the wire is wound into a coil, it also presents some reactance to current flow. Both of these are combined into a measure of resistance known as impedance. Typical speakers have impedances of 4 or 8 ohms and require much too much power to work with a crystal radio.

Keeping a Log for Your Radio Listening

Most people listen to radio for the programming content. However, if you are serious about radio technology, you may wish to maintain a log of your listening. It can be fun to see how many stations you can hear with your radio. The conditions for propagation of radio waves varies over time and it is probable that you will hear some stations at certain times and not at other times. Therefore, you should listen at different times of the day and night. By keeping a log, recording the stations you can hear, and the date and time, you will create a log of your listening. From this log you may discover some interesting trends. For example, some stations can be heard at night but not during the day.

Even though your radio is calibrated, you cannot know for sure the exact frequency for a station until you hear an announcement of the frequency by the radio announcer. Some stations announce this many times each hour and some do not announce their broadcast frequency very often. So it may be an exercise in patience on your part. The strong station in Cleveland, WTAM, broadcasts on 1100 kHz and announces their broadcast frequency several times per hour. The announcement may sound something like this: "This is WTAM eleven hundred." You may need to listen carefully to catch the announcement because it is usually done fairly fast. Don't expect to hear something like this: "This is station WTAM broadcasting on a frequency of eleven hundred kilohertz." You might have heard something like that 50 years ago but not today!

There are some amateur radio operators (Ham Radio) that have become big fans of the crystal radio. They have studied the construction of crystal radios and compete with each other to design radios capable of tuning stations from distant locations. In fact, one Ham Radio operator located in Hawaii was able to tune a station in Cuba with his special crystal radio. In the case of the radio you have constructed, we cannot expect that level of performance. In fact, if you hope to hear distant stations with your crystal radio, you will probably need to use an amplifier. After you finish your crystal radio you will build an amplifier to use with your radio. Then you will be able to tune many more local AM stations and also distant stations at night if you have a good antenna.

There are different modes of radio wave propagation that determine which radio stations you can hear and at what time of day or night. This also varies with the frequency of the radio signal. For long distance reception, frequencies above those of the AM broadcast band are more useful. These frequencies, between 2 to 30 MHz, are sometimes referred to as shortwave radio or high frequency (HF) radio. Some of the Ham Radio bands are included in this range. International radio stations also broadcast in this frequency range.

Radio wave propagation can be divided into two broad types called ground waves and sky waves. Ground waves travel close to the surface of the Earth and sky waves bounce off the Earth's upper atmosphere (ionosphere), which allows propagation to more distant locations. The bouncing of radio waves off the ionosphere depends on the amount of ionization. The ionization is caused by high energy radiation (UV light) from the Sun. Therefore, there is more ionization in the upper atmosphere during the day than at night. That is why radio propagation varies between day and night. This subject is considerably more complex than my short description might suggest, but we will leave it at this point for now.

During the day radio wave propagation in the AM band is generally confined to the ground wave type. Ground waves don't travel much farther than the horizon. The distance to the horizon depends on the height of the radio transmitting antenna and its elevation above sea level. Nevertheless, propagation by ground wave is usually limited to 100 miles or less.

During the night the Sun does not shine on Earth's upper atmosphere and the degree of ionization of air molecules and atoms decreases. With these changes, radio waves of the frequencies used in the AM band can bounce off the ionosphere. When this happens it becomes possible to tune in AM stations hundreds of miles from your location. Therefore, if you want to try tuning in distant stations, you will need to try this at night. The signal strength of these stations will be weak and you will have much better luck hearing these stations if you connect an amplifier to your crystal radio. In addition, it is especially important to use a good outdoor antenna if you are trying to tune in distant stations.

As you build a log of listening, you may wish to gather more technical information about the distant stations you have been able to hear. A good web site to consult is the FCC (Federal Communications Commission) site for AM stations: https://www.fcc.gov/encyclopedia/am-query-broadcast-station-search . The information at this site includes the location of the station's transmitter (given in longitude and latitude) and the amount of power used by the station. Some AM stations are required to broadcast with lower power at night in order to avoid interfering with other radio stations operating on the same frequency (due to longer transmission distances at night). The FCC licenses a limited number of "clear channel" stations in various locations in the United States. These stations are required to transmit with 50,000 watts of power day and night and these clear channel stations are the most likely ones you will be able to hear from distant locations. There is a list of clear channels at this web site: https://en.wikipedia.org/wiki/Clear-channel_station . Our strong AM station in Cleveland, WTAM 1100, is a clear channel.

RF Diagnostics, LLC

WELCOME TO RF DIAGNOSTICS

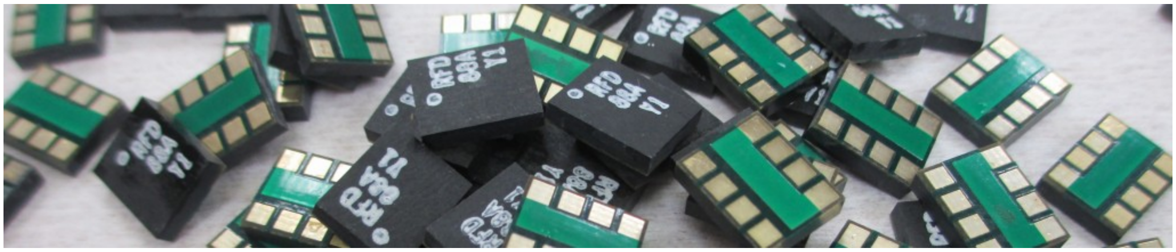

Energy Harvesting Modules

WIRELESS ENERGY HARVESTING PRODUCTS RF Diagnostics, LLC designs and sells surface mount RF to DC converters and other building block modules for energy harvesting systems. With our products you can harvest ambient AC/AM/FM/Wireless Energy to power your IoT battery free system. We also provide custom designs for RF/microwave and general electrical engineering products and design contracting services to clients in the commercial wireless and defense electronics businesses. We strive to be a high quality and cost effective design resource for your design team.

OUR PRODUCTS

Microwave Energy Detector
This detector receives wireless energy at 2.5GHz and converts the signal to a DC voltage which powers a light emitting diode (LED). Place the detector on your microwave window or near your wireless LAN router and the LED will brighten and fade as the wireless field changes. No batteries needed!
RFD102A RF-DC Converter
2.5GHz Dipole Antenna
Light Emitting Diode

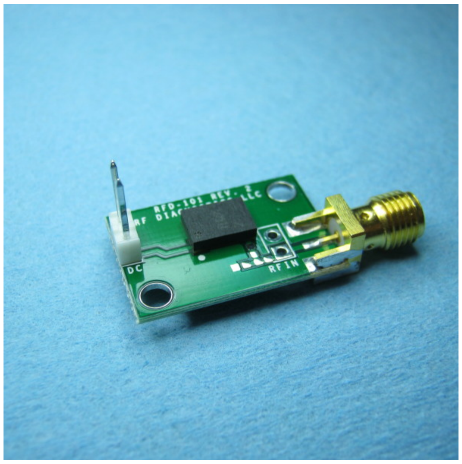
RFD-101 REV. 2
RF DIAGNOSTICS LLC
RF
DC
RF IN

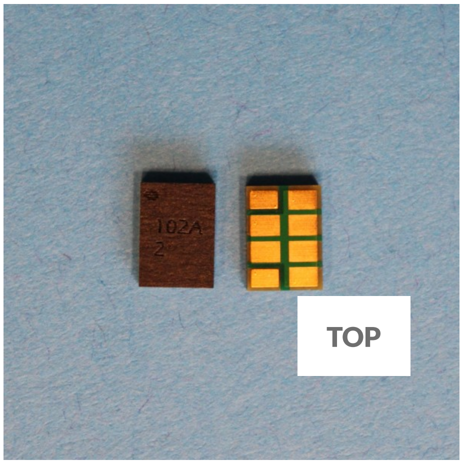

TOP

RFD102A-DET: Microwave Energy Detector	RFD102A-TB: 60Hz…6GHz Energy Harvesting Test Board	RFD102A: Wireless Energy Harvesting Module
$18.00	$33.00	$12.00
Add to cart	Add to cart	Add to cart

OUR DESIGN SERVICES

Energy Harvesting Module & System Design

Any of our standard products can be customized for your application. We can support your application with matching, testing, antenna design, assembly and many other tasks. Have a new idea for a product? We can help you realize it. We are developing new products and are always interested in your feedback to guide our product development.

MMIC Design

We use Agilent's Advanced Design System for MMIC design and PCB design. Previous designs completed are quad-band GSM power amplifier modules, 2.4-2.5GHz/5-6GHz wireless LAN power amplifiers, broadband power amplifiers, switch-filter designs, RFID detectors, RF/DC converters and filter designs.

Front End Module Design

We use Agilent's Advanced Design System and Orcad PCB Editor for Front-End Module designs. Previous design examples are high-volume switch filter modules, dual-band 802.11a/b/g front end modules, 76GHz car radar module, RFID card design.

RFID Detector System

On an NSF Phase I & II grant to one of our clients, operating as a subcontractor, the RF Diagnostics design team created a complete RFID MEMS resonator detection system. Both hardware and software were developed under this contract. The work was published at an SPIE conference in July 2010.

LEARN MORE

NISKAYUNA, NY 12309 USA
518-810-6560
INFO@RFDIAGNOSTICS.COM

TOP

RF DIAGNOSTICS, LLC COPYRIGHT © 2017.

WEBSITE BY ROCK THE DEADLINE

Build a Homebrew Radio Telescope

Explore the basics of radio astronomy with this easy to construct telescope.

Mark Spencer, WA8SME

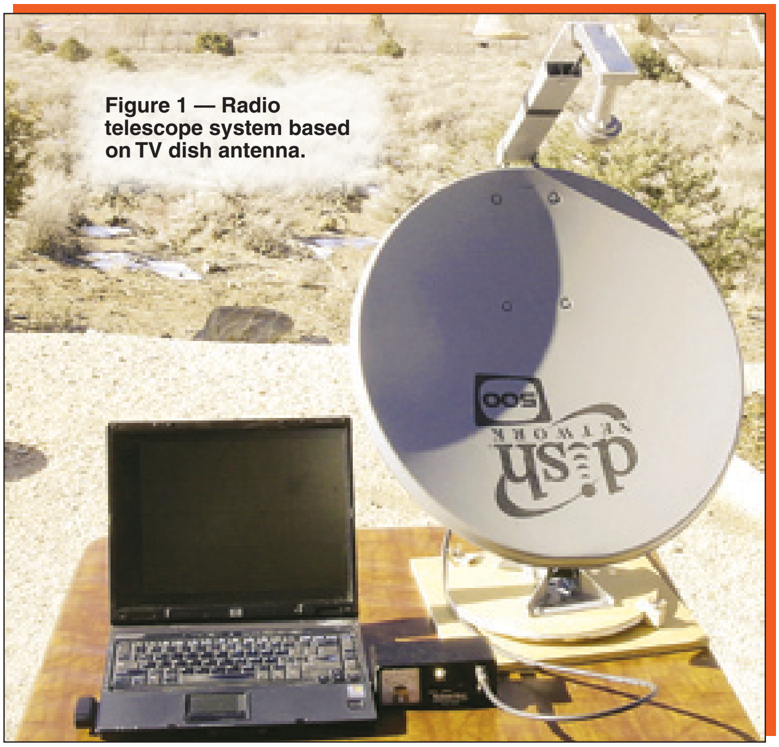
Figure 1 — Radio telescope system based on TV dish antenna.

There are many ham radio related activities that provide a rich opportunity to explore and learn more about the science of radio. One of those opportunities is radio astronomy.

All matter emits radio frequency (RF) energy dependent on the temperature and makeup of the matter, including the matter in space. The foundation of radio astronomy is to study the heavens by collecting and analyzing the RF energy that is emitted by bodies in space, very much as optical astronomers use light energy collected by telescopes. It sounds complicated. While professionals use very sophisticated and expensive equipment, you can, with some simple equipment and a little investment, build a radio telescope that will allow you to learn and explore the fundamentals of radio astronomy.

A Homemade Radio Telescope

In this article, I will build on an existing design of a radio telescope made from one of those ubiquitous TV dish antennas that you see around your neighborhood. The radio telescope (RT) project described here can easily be reproduced. Although this is not a fully capable RT, it can provide a wonderful learning opportunity for you, or perhaps students in your local school.

Figure 1 shows the radio telescope set up. The major components include a modified TV dish antenna mounted on a wooden support structure to allow pointing the antenna, a commercial satellite signal strength detector that displays the signal strength of signals collected by the dish on a meter and an interface that converts the signal strength into a amplitude modulated tone. The tone is fed into a computer sound card and finally a computer and software graphically displays the signal strength as a function of time.

The TV dish modifications are structural, and any available TV dish system can be used. The signal strength detector costs between $40 and $65 and is widely available from Web retailers. The interface circuit, which will be described shortly, is easily duplicated and costs approximately $20. Finally, the display software is free.

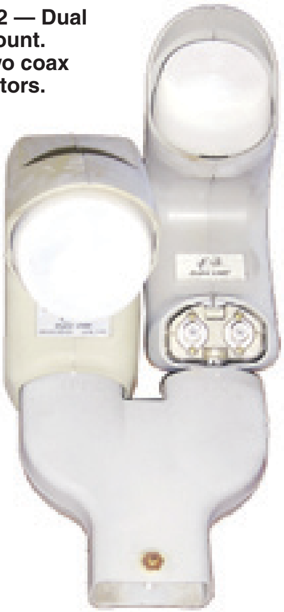
Figure 2 — Dual LNB mount. Note two coax connectors.

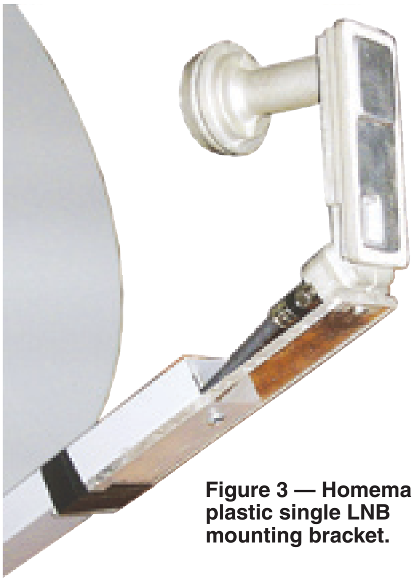
Figure 3 — Homemade plastic single LNB mounting bracket.

What it Can Do

The following is just a sample of what you can do with this simple RT:

- Use the sun to study and determine the beamwidth of the dish and verify the mathematic formula that is used to predict dish antenna performance.
- Measure the radiation intensity of the Sun and perhaps detect changes in solar activity.
- Measure the relative changes in the surface temperature of the moon.
- Learn about and explore a common radio astronomy collection technique called the *drift scan*.

From June 2009 QST © ARRL

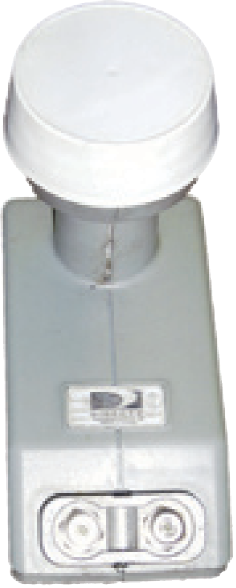

Figure 4 — Dual coax connector configured LNB. Terminate one connector with a dummy load.

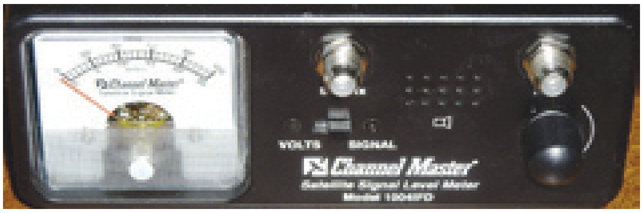
Channel Master
VOLTS
SIGNAL
Channel Master
Satellite Signal Level Meter
Model 1004IFD

Figure 5 — CM satellite signal strength meter.

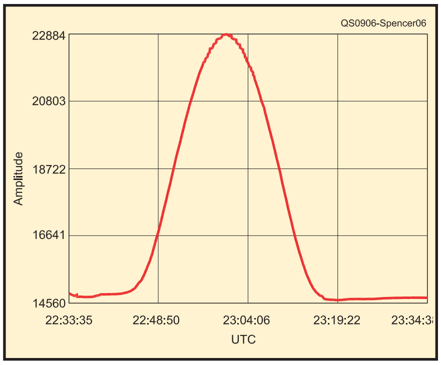

Figure 6 — *SkyPipe* screen showing antenna response.

- Explore the fundamental principle of energy emission as a function of temperature by detecting the relative differences between the temperatures of emitting bodies.
- Detect satellites parked along the Clarke Belt in geosynchronous orbit and illustrate how crowded space has become.
- Detect the Earth's rotation around the Sun and the Earth's spin on its axis by comparing daily drift scans of the horizon.

Antenna Subsystem

The basic RT system is based on the "Itty-Bitty" design that is described in two Web pages.[1,2] The TV dish is an offset 18 inch dish that has down converter(s) mounted at the focal point of the dish. The down converter is called a low noise block (LNB). The LNB is a preamplifier/down converter that converts the satellite signals from around 12 GHz down to around 2.4 GHz. Most modern dishes have two or more LNBs to access more than one TV satellite at a time without changing the pointing of the dish (Figure 2). The LNBs are mounted to share the focal point of the dish. Since only one LNB is required for the RT, I made a minor adjustment to the published Itty-Bitty design to position the single LNB at the dish focal point. Mounting the single LNB at the focal point really helps in pointing the antenna.

I used the existing LNB housing and mounting bracket as a template to determine the distance between the edge of the mounting arm to the mounting hole of the LNB. I then used a piece of plastic to fabricate a new mounting bracket for the LNB as shown

---

[1] Notes appear on page 45.

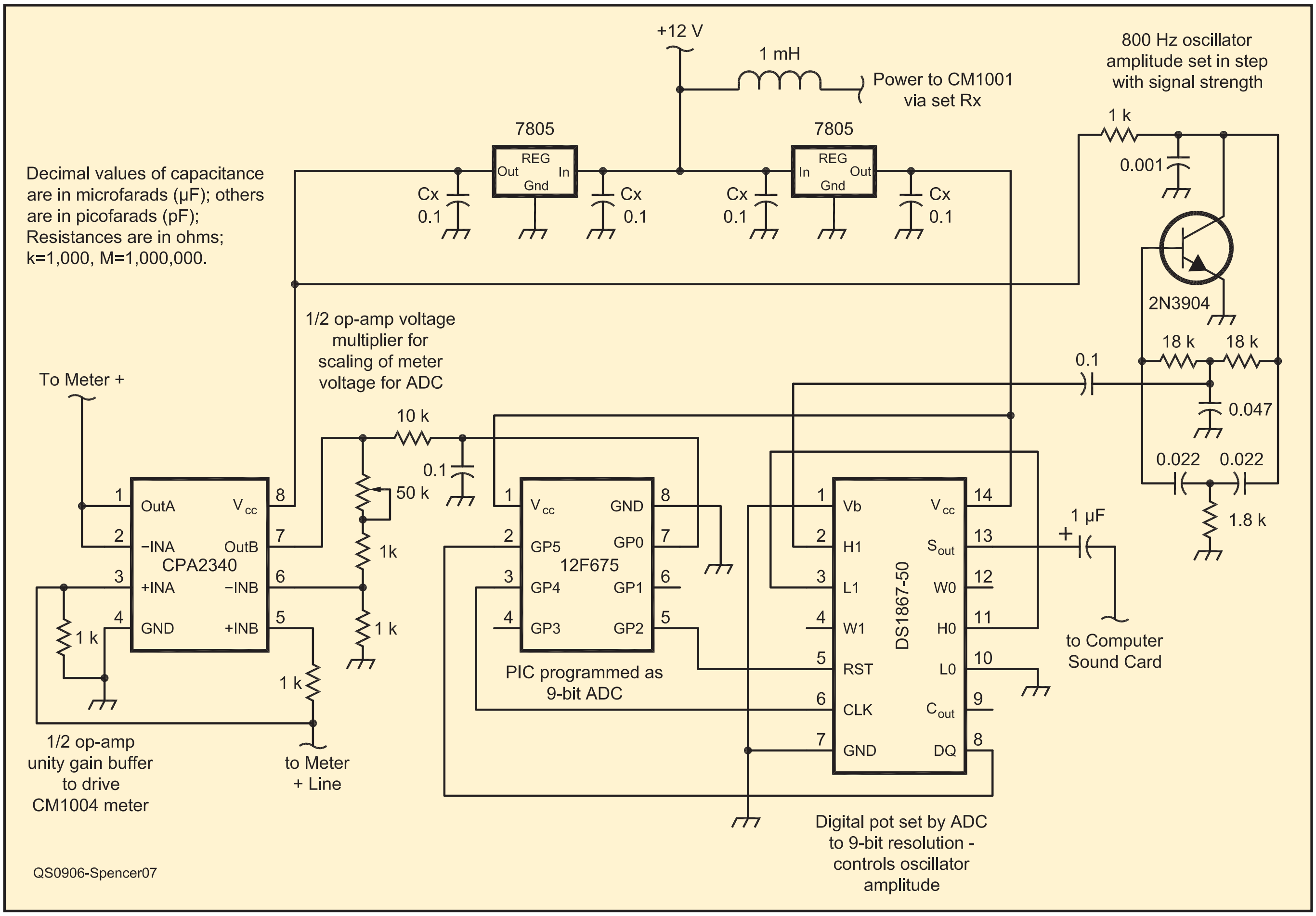
+12 V
1 mH
Power to CM1001 via set Rx
800 Hz oscillator amplitude set in step with signal strength
7805
REG
Out
In
Gnd
7805
REG
In
Out
Gnd
Cx 0.1
Cx 0.1
Cx 0.1
Cx 0.1
1 k
0.001
2N3904
18 k
18 k
0.1
0.047
0.022
0.022
1.8 k
Decimal values of capacitance are in microfarads (µF); others are in picofarads (pF); Resistances are in ohms; k=1,000, M=1,000,000.
1/2 op-amp voltage multiplier for scaling of meter voltage for ADC
To Meter +
10 k
0.1
50 k
1k
1 k
1 k
1 k
OutA
Vcc
−INA
OutB
CPA2340
+INA
−INB
GND
+INB
Vcc
GND
GP5
GP0
12F675
GP4
GP1
GP3
GP2
PIC programmed as 9-bit ADC
Vb
Vcc
H1
Sout
L1
W0
W1
H0
RST
L0
CLK
Cout
GND
DQ
DS1867-50
1 µF
to Computer Sound Card
1/2 op-amp unity gain buffer to drive CM1004 meter
to Meter + Line
Digital pot set by ADC to 9-bit resolution - controls oscillator amplitude
QS0906-Spencer07

Figure 7 — RT Interface circuit diagram.

From June 2009 QST © ARRL

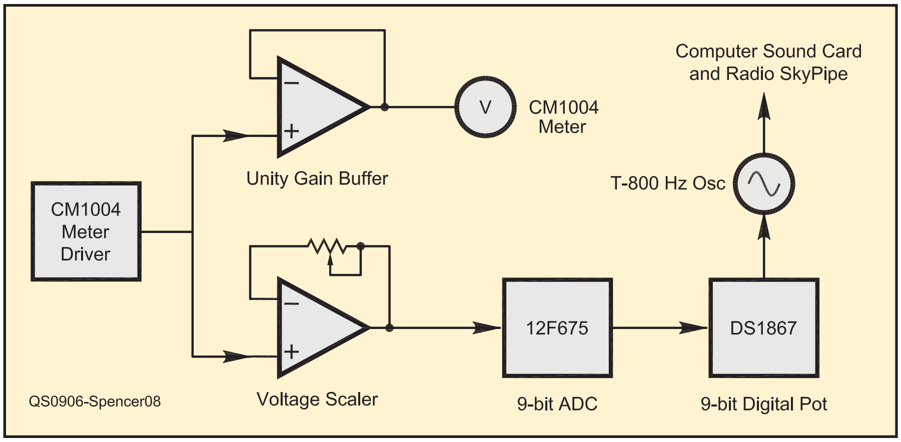

Figure 8 — RT Interface block diagram.

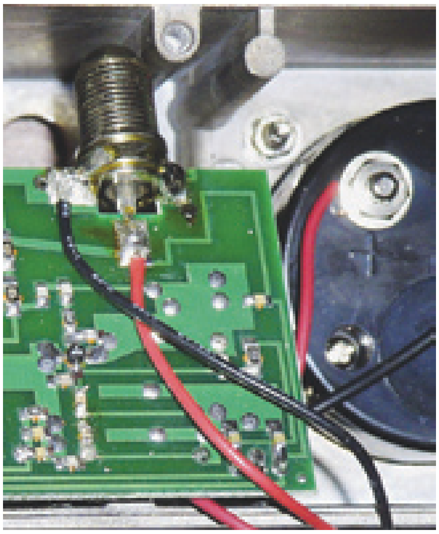

Figure 9 — Power and ground connection to CM board.

in Figure 3. The dimensions are not super critical, but careful placement certainly will improve the RT performance.

Some LNBs have two coax connectors. Only one will be used in the RT (Figure 4). It is a good idea to terminate the extra coax connector with a 75 Ω dummy load plug to balance the load on the LNB. The dummy loads for F type TV coax connectors are readily available from electronic parts retailers.

Note that the dish is mounted upside down. Though this orientation is not ideal for receiving satellite signals, this arrangement helps with pointing the dish in its radio telescope role.

Satellite Detector

The detector used in this project is the Channel Master (CM) satellite signal level meter model 1004IFD (Figure 5).[3] The CM is connected to the LNB. Power is supplied to the LNB through the coax connection from the CM. The CM detects the signal coming from the LNB and gives a meter indication of the signal strength and also varies the frequency of an audio tone to help technicians point the dish at the desired satellite. As you move the dish through the beam coming from the satellite, the meter indication will increase and then decrease coincident with the pitch of the audio tone.

The Itty-Bitty plans detail how to connect power to the CM and in turn connect power to the LNB (this power connection is handled by the interface in this project). Though somewhat effective, the CM meter and variable frequency tone indications provide limited utility in detecting changes in signal strengths required for radio astronomy.

Display

To really study the signals received by the RT, you will need to see them displayed graphically on a strip chart. There is an excellent software package called *Radio-SkyPipe* that is posted on radio astronomy Web sites.[4] The free version of this software is a good place to start. *SkyPipe* uses the computer sound card to measure the incoming signal strength and graphically displays the signal strength as a function of time. Figure 6 is illustrative of a signals detected by the RT. *SkyPipe* is very easy to use but some study of the HELP files will make it easier for you to fully tap into the capabilities of this software.

SkyPipe requires audio signals to be fed into the sound card MICROPHONE jack. The output of the CM detector is either an analog meter reading or a frequency modulated (constant amplitude) tone that is not really compatible with *SkyPipe*. An interface is required.

Interface

What is required to make the CM output work with *SkyPipe* and a sound card is to convert the signal level into an amplitude varying audio tone. The interface designed to do this is shown in Figure 7 and as a block diagram in Figure 8. Refer to the block diagram during the description of the interface function.

The unity-gain op-amp is used as a buffer between the CM meter driver circuit and the analog meter. The other op-amp is used as a voltage multiplier to scale the CM meter driver output voltage to match the 5 V reference voltage of the following analog to digital converter (ADC). The variable resistor in this voltage multiplier circuit is used to calibrate the CM to *SkyPipe*. The voltage from the multiplier is fed to a programmable interface controller (PIC) that is programmed as a 9-bit ADC to covert the analog voltage that is a function of received signal strength to a 9-bit digital word that is used to control a digitally controlled variable resistor. The interface includes a simple Twin-T audio oscillator circuit that provides a tone of approximately 800 Hz that is fed to the computer sound card. The amplitude of this audio oscillator is varied by the digital pot that is being controlled by the PIC. The result is the audio amplitude being varied in step with the signal strength detected by the CM.

The circuit provides power to the CM and the LNB. A 12 V source in the CM is tapped through an RF choke and this is connected to the LNB coax connector inside the CM (Figure 9). The 12 V is also regulated to 5 V to provide power to the interface. Though probably not required, there are two 5 V sources, one for the digital components of the interface, and the other for the analog components with one common ground point. This arrangement is used to isolate potential digital and analog noise sources within the circuit.

The interface is built on a circuit board and mounted right inside the CM box

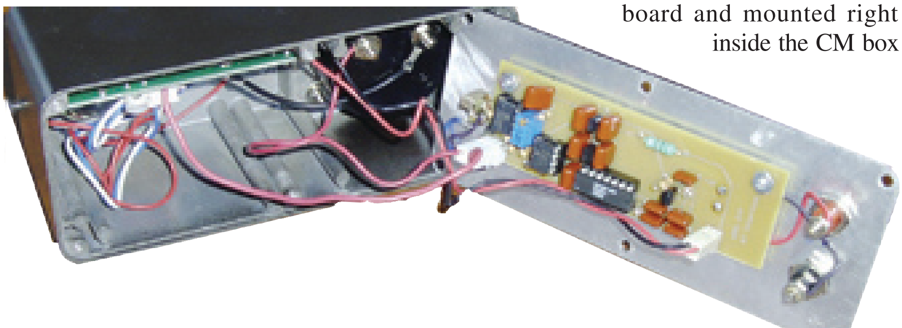

Figure 10 — CM with interface board.

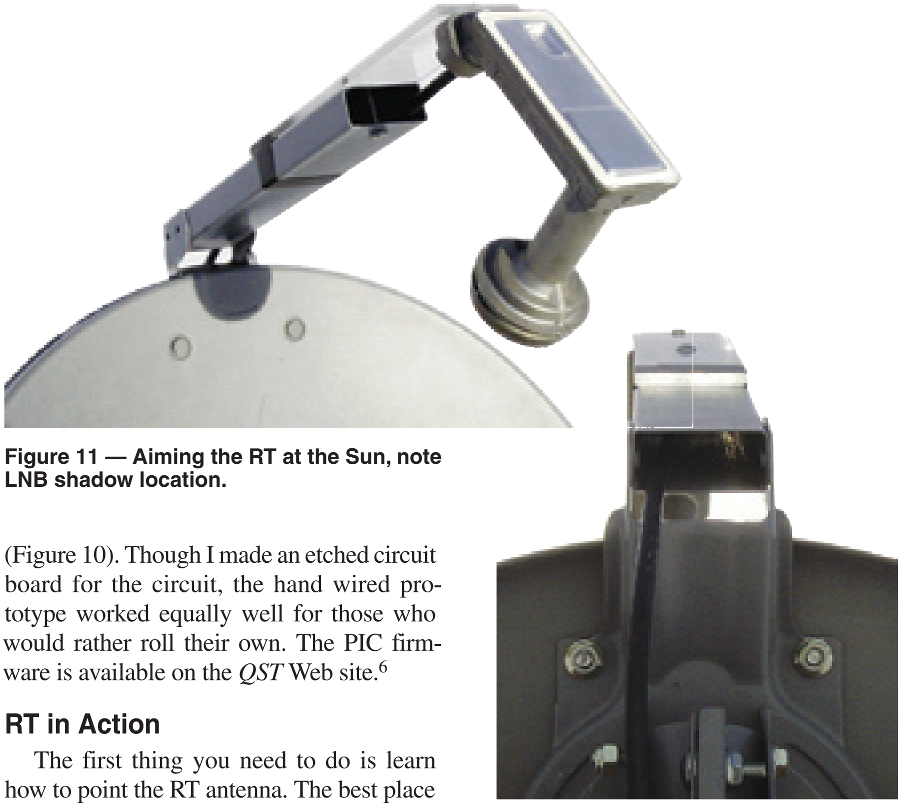

Figure 11 — Aiming the RT at the Sun, note LNB shadow location.

(Figure 10). Though I made an etched circuit board for the circuit, the hand wired prototype worked equally well for those who would rather roll their own. The PIC firmware is available on the *QST* Web site.[6]

RT in Action

The first thing you need to do is learn how to point the RT antenna. The best place to start is to connect the CM to the antenna and point the antenna at the Sun. *Caution*: Do not look into the Sun as you do this, or at any time. Adjust the pointing angle and elevation until you get peak signal strength as indicated on the CM meter or hear the highest pitch audio tone. With the antenna pointed directly at the Sun, take note of the position of the shadow of the LNB on the surface of the dish (left in Figure 11). If you look from behind the dish, along the LNB

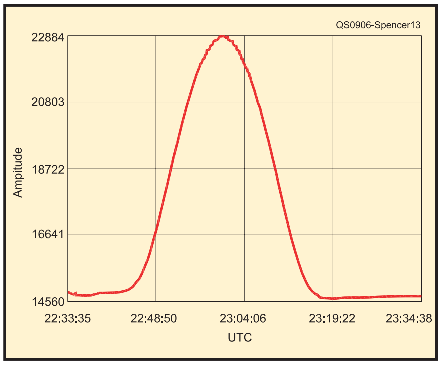

Figure 13 — Drift scan of the Sun indicating antenna's azimuth pattern.

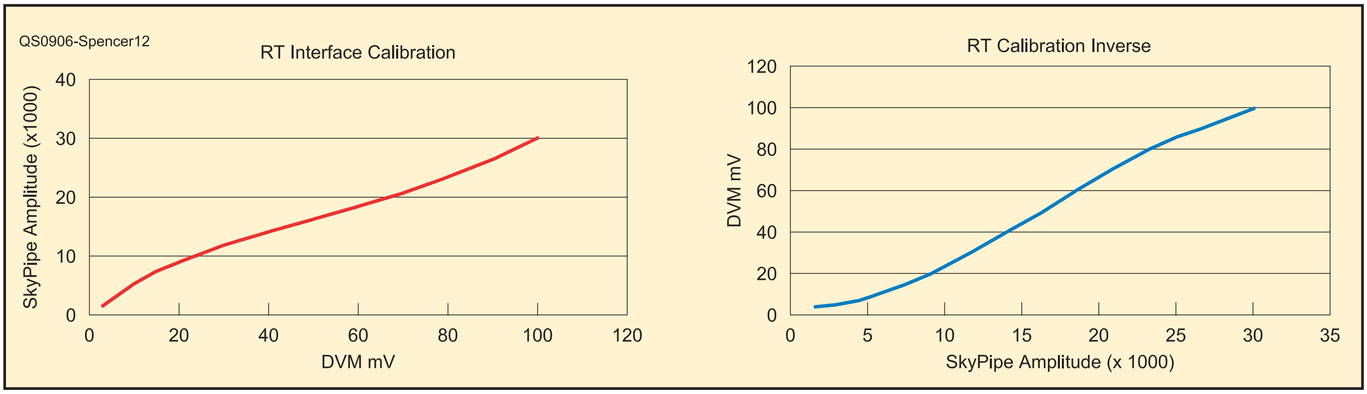

Figure 12 — Example calibration curves.

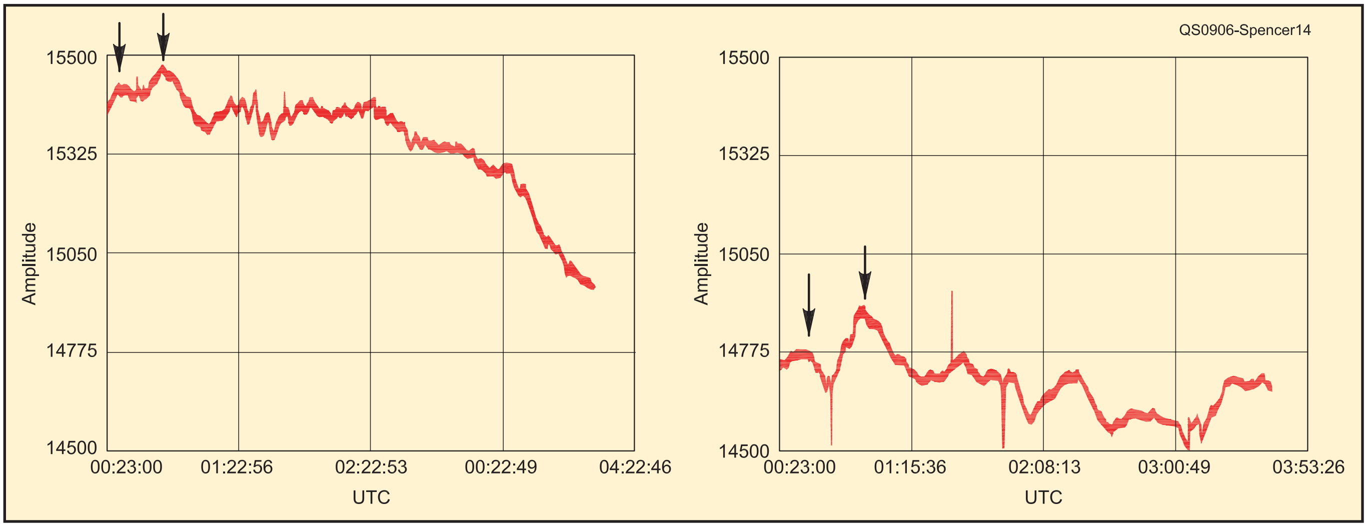

Figure 14 — Sequential drift scans. Note the time offsets between the peaks.

From June 2009 QST © ARRL

supporting arm (between the arm and the rim of the dish), you will see the Sun being blocked by the LNB.

Once you have the RT set up, it needs to be calibrated to match the output of the CM to *SkyPipe*. I have developed an *Excel* spreadsheet template to help with the calibration and a few of the other activities that you can accomplish with the RT (also available from the *QST* Web site). Turn the RT to a signal source, the Sun, or the side of a building would work. Turn the gain control of the CM to set the meter to maximum. Run *SkyPipe* and adjust the variable resistor on the interface board until you get a reading on the *SkyPipe* graph vertical (y) axis of approximately 32,000. With the maximum value set, adjust the CM gain control through the voltage range (0 to 100 mV) in 10 mV steps and record the corresponding y axis value on *SkyPipe*. This data is entered into the *Excel* spreadsheet to compute the calibration curve between voltage and y axis value. Both voltage and y axis values are used in analyzing recorded signal strength data (Figure 12).

A good first activity is to do a drift scan of the Sun. A drift scan means that you set the antenna azimuth (AZ) and elevation (EL) to some fixed pointing angle and allow the Earth to serve as the rotator to drag the antenna across the sky. To do a drift scan of the Sun, first set the elevation and azimuth to point directly at the Sun (maximum signal) and then move the azimuth toward the west (leave the elevation set) until you are off the peak signal. Now start *SkyPipe*. In about 15 minutes, the Sun will pass through the antenna pattern beam width and the result will be as illustrated in Figure 13. You can also use this collection technique to explore the antenna performance parameters.

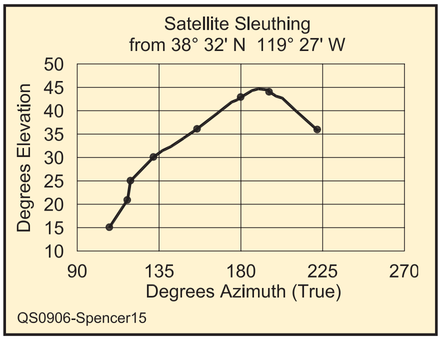

Figure 15 — Clarke Belt plot — tracking down satellites.

A good second activity is to do two drift scans of the night sky on two consecutive nights (beginning the scans at the same time each night) using the same fixed antenna azimuth (AZ) and elevation (EL). Figure 14 shows two such drift scans. Although at first glance they may not seem similar, there are some interesting features that are pointed to by arrows. If you compare the time that these two peaks occurred, the time difference is about 4.5 minutes. This shift is the result of the distance the Earth had traveled during the 24 hours between collections. This illustrates that the Earth's rotation as well as its travel in orbit needs to be considered when comparing drift scans. Enough to make your head spin (pun intended)?

A final good starting activity is to point the antenna toward the Clarke Belt and find all the satellites in geosynchronous orbit transmitting on 12 GHz. If you record signal strength peaks and AZ and EL for each peak, you will develop a graph of the Clarke belt as illustrated in Figure 15.

I have only scratched the surface, and the sky is the limit of this little project. The RT project can certainly broaden your horizons and expand your understanding of our universe. If you would like more detail than can be presented here, please contact the author.

Notes

[1] **www.setileague.org/articles/lbt.pdf**.
[2] **www.aoc.nrao.edu/epo/teachers/ittybitty/procedure.html**.
[3] **www.pctinternational.com/channelmaster/0612/satellite.html**.
[4] **radiosky.com/skypipeishere.html**.
[5] **en.wikipedia.org/wiki/Geostationary**.
[6] **www.arrl.org/files/qst-binaries/**.

Mark Spencer, WA8SME, is ARRL Education and Technology Program Coordinator. He is an ARRL member and holds an Amateur Extra class license. You can reach Mark at 774 Eastside Rd, Coleville, CA 96107 or at **wa8sme@arrl.org**.

Did you enjoy this article?
Cast your vote at:
VOTE
www.arrl.org/members-only/qstvote.html

From June 2009 QST © ARRL

Double-ganged Mystery Crystal Set

copyright © 2006 by Larry J Solomon

Here's another great performer based on the Mystery set. This one uses a 500 pF döppelganger.

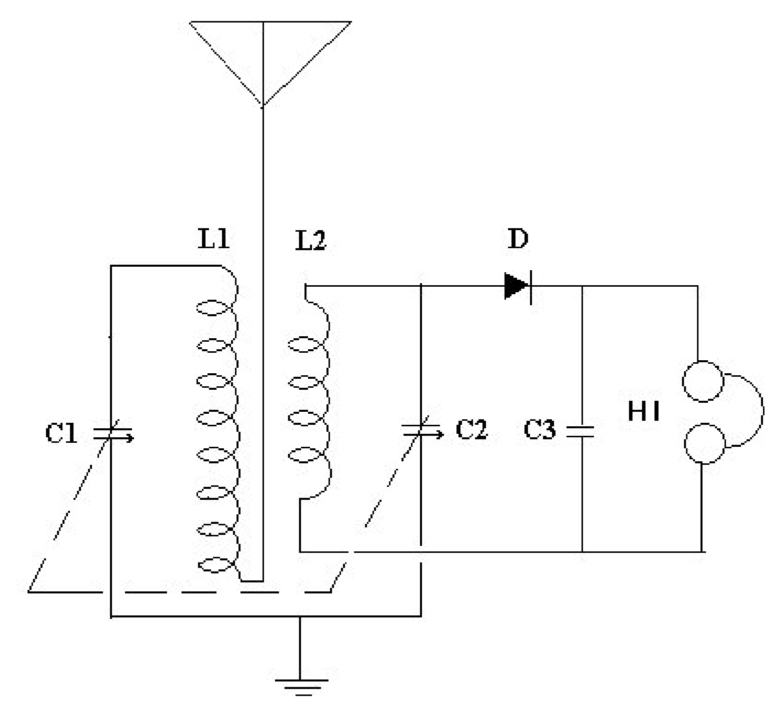

C1, C2 - 500 pF double-ganged air-variable capacitor
C3 - .001 uF capacitor
D - 1N34A diode or rock crystal
L1 - 50 turns #24 awg
L2 - 25 turns #24 awg wound in tandem with middle of L1

The Coil is air-spaced on each end of L1. L2 is close-wound with L1 after 12 turns of L1. Coils are wound in same direction. Coil form is 2.625 inches high by 2.375 inches in diameter. Although a high impedance headphone is shown, the set can be directly connected to the magnetic phono input of an amplifier instead of the headphones.

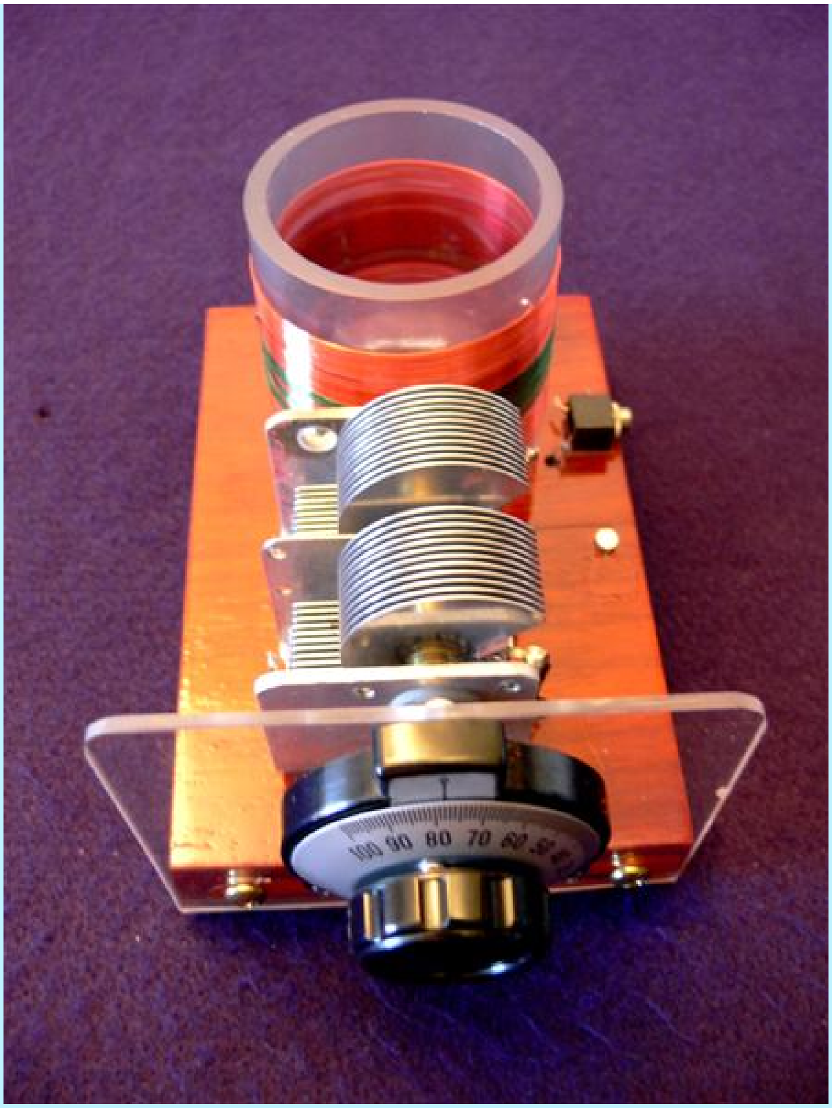
100 90 80 70 60

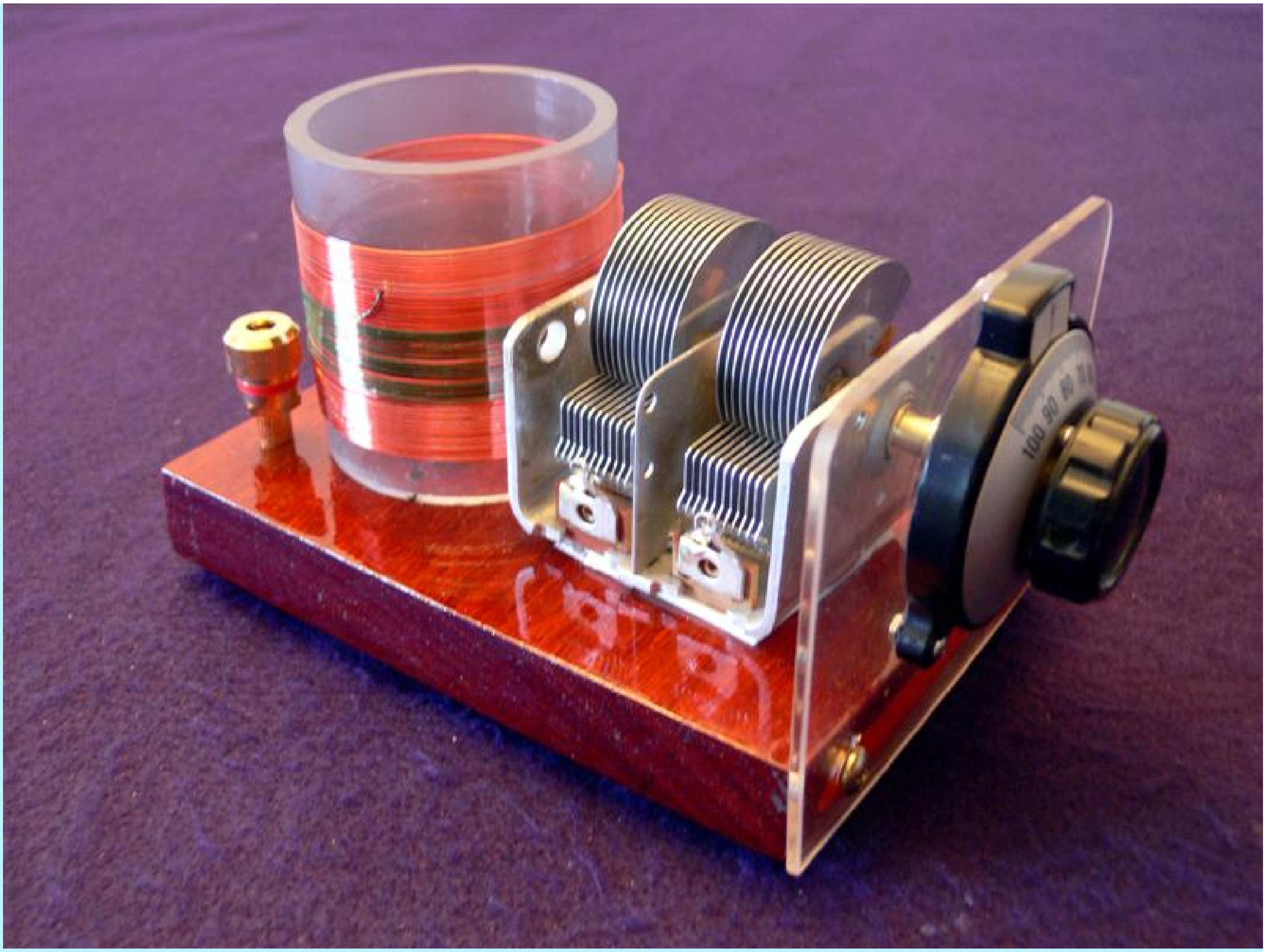
100 90 80 70

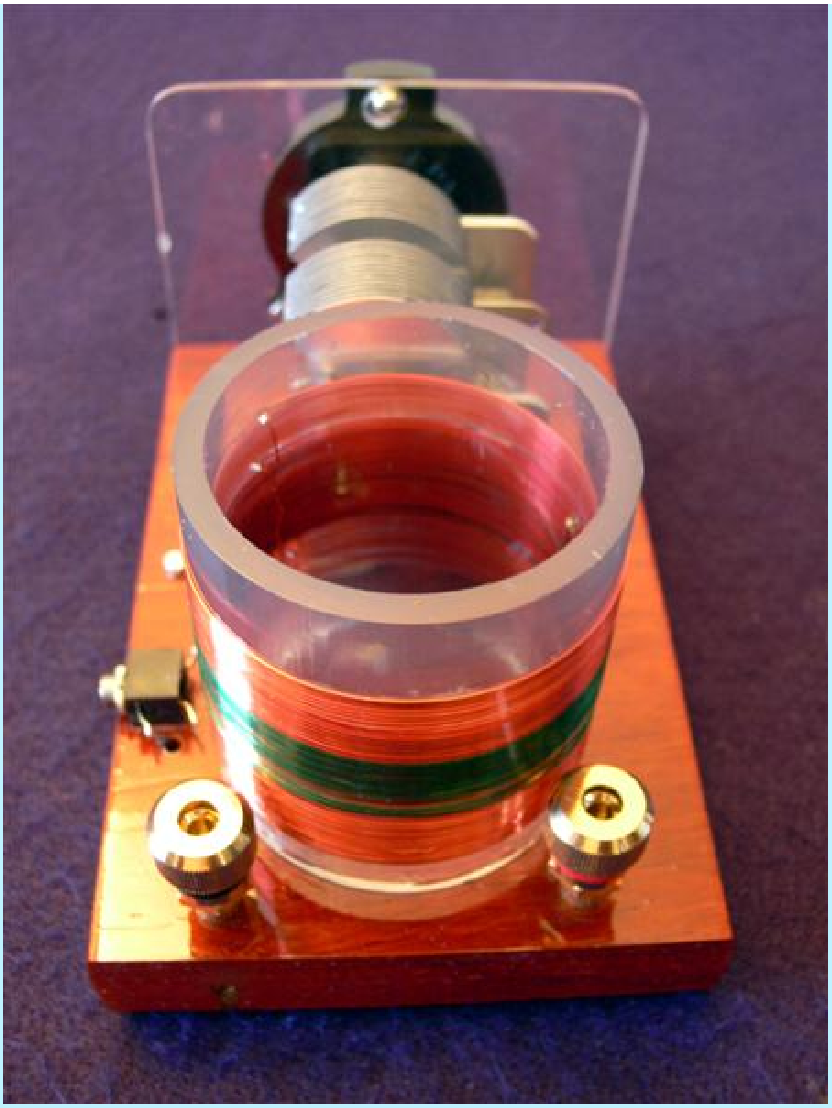

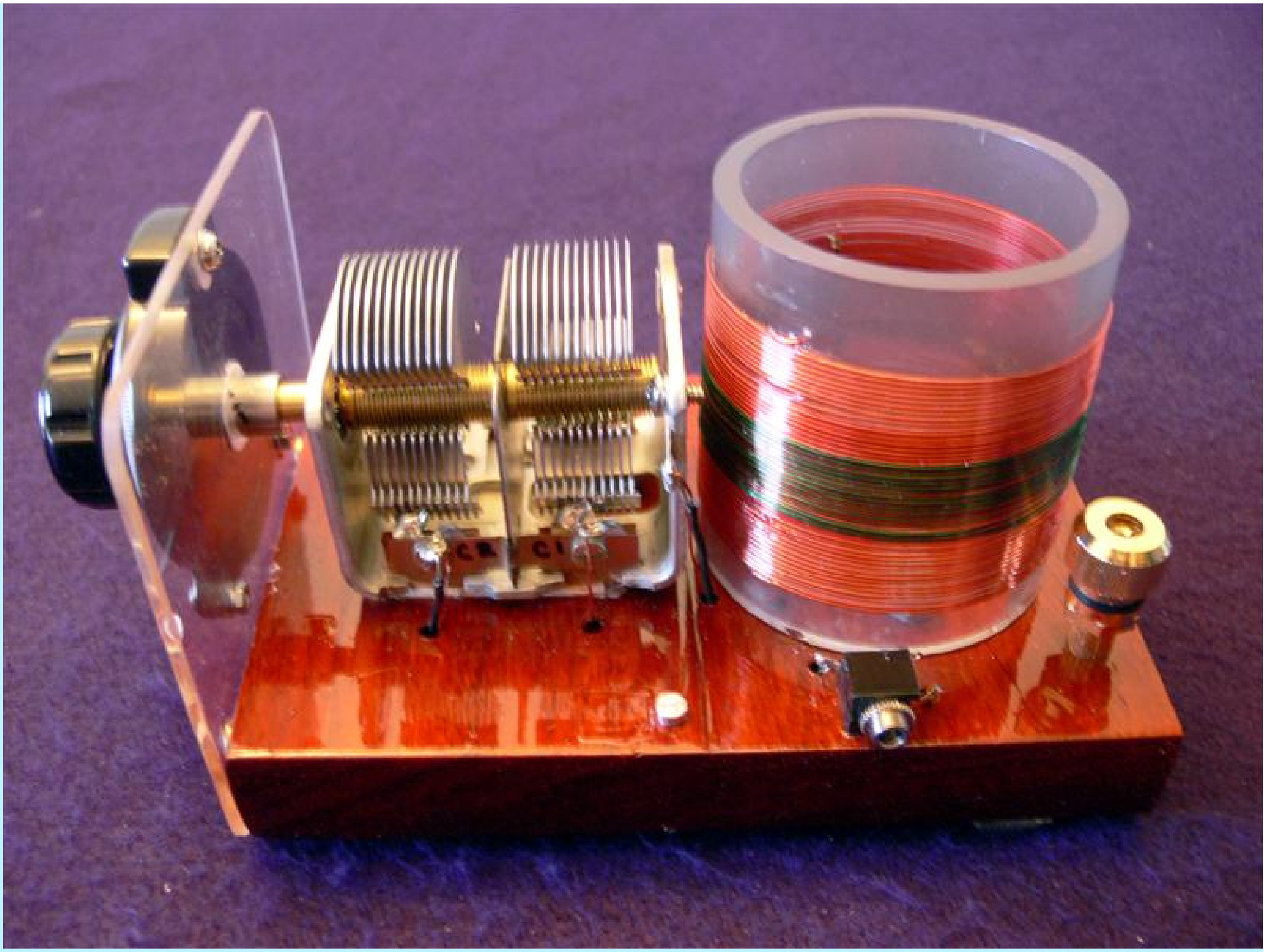

Home

My DIY radio wave energy harvesting circuit

proteus-h (61) (/@proteus-h) in **technology (/trending/technology)** • 2 months ago

Here's a writeup of a simple radio-frequency energy harvester I built earlier this year. Now that I'm back at my apartment with my equipment I should be able to post more DIY projects. This device is crude, unoptimized, and imperfect ... *but it works*.

This super simple circuit converts radio waves collected into a crude wire antenna into electrical energy which is stored onboard two electrolytic capacitors. This is not a "free energy" device because it actually works: All of the energy brought into the capacitors comes directly from the energy of electromagnetic waves. You can build your very own harvester circuit with just 9 components and a piece of perfboard.

The actual amount of energy collected is absolutely tiny if you are only using ambient RF, so like most of my projects this is just a proof of concept. It's not going to charge a powerbank off of your cell phone emissions or anything like that ...

In case you don't want to get into the details, here's a picture of the completed crude harvester:

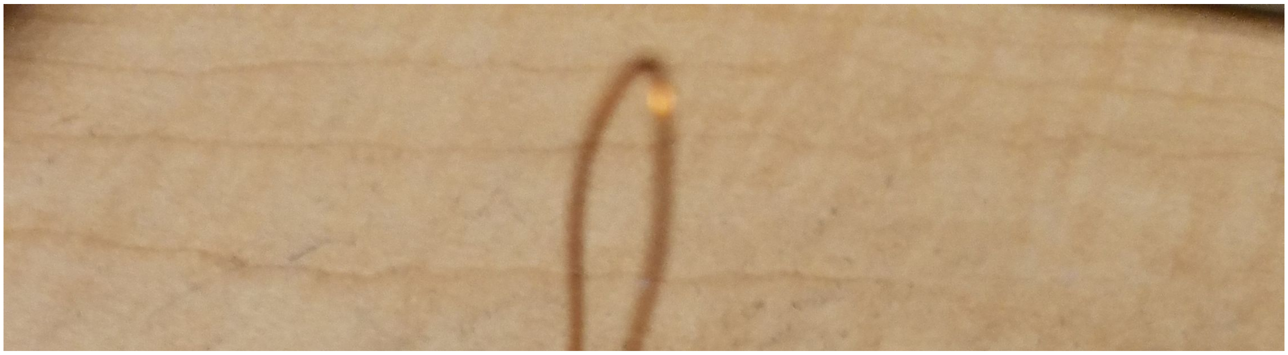

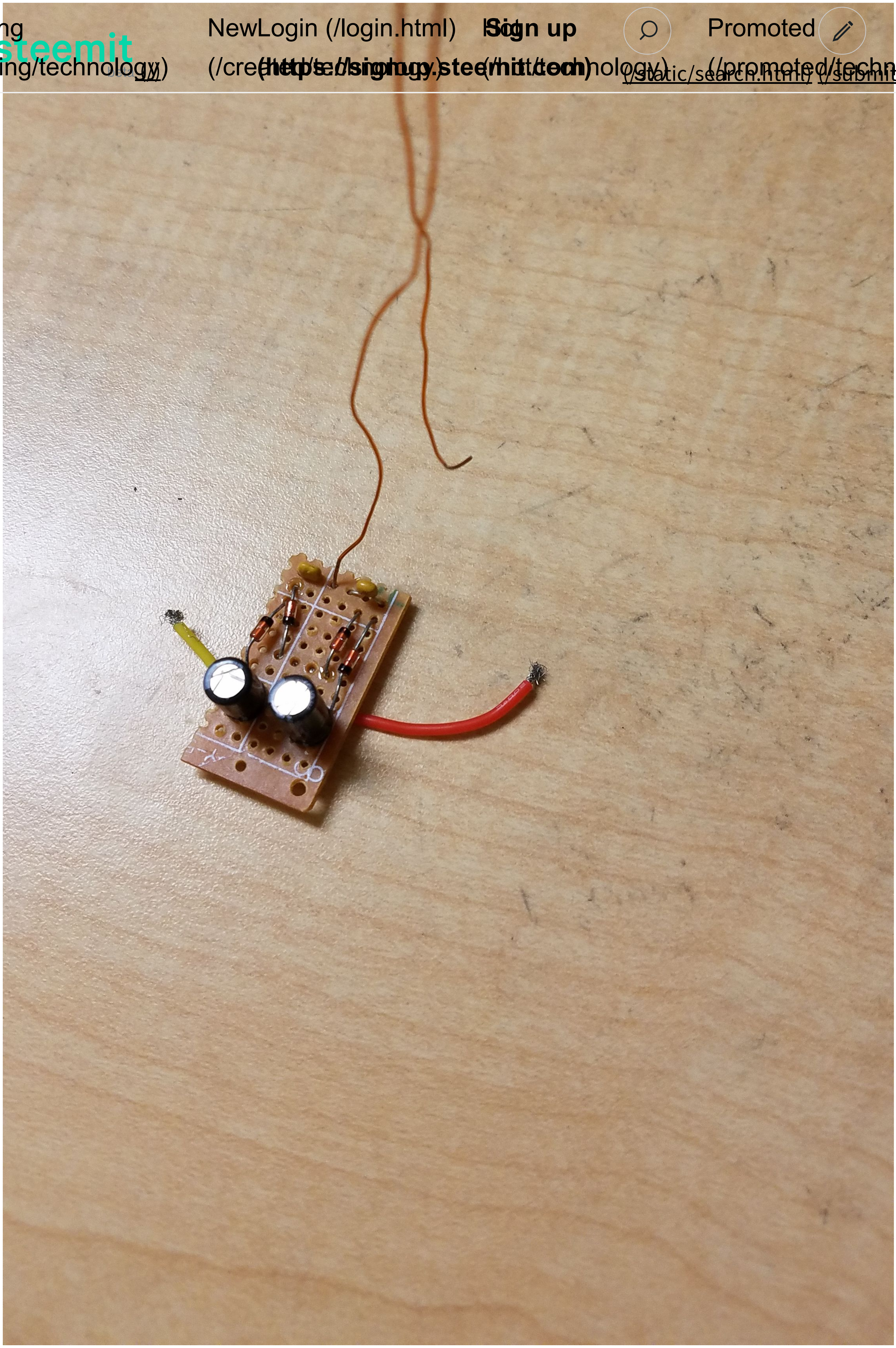

Operating Principles

Ambient electromagnetic radiation is all around you, harmlessly emitted from most electronics. Computers transmit 2.4 GHz microwaves for wifi/bluetooth and car FM transmitters transmit around 100 MHz to send audio to your car's radio receiver. Light itself is also, of course, electromagnetic radiation, but this device only works for much lower frequencies. I had the best luck around 100-200 MHz radiation. Electromagnetic radiation at these low frequencies (radio and microwave) is not ionizing and as such harmlessly passes through or is absorbed by your body.

EM radiation can be collected via antennas. Any radio receiver does this but only uses the captured radio wave energy to produce a detectable signal so that you can hear the broadcast. But these waves contain *energy*, and we can capture this energy and use it for other purposes, small as it may be.

Crystal radios (https://en.wikipedia.org/wiki/Crystal_radio), common starter DIY electronics projects, do just this. A crystal radio is a simple receiver that can be built from mostly scrap parts, and lets you listen to radio broadcasts *without a battery*. The crystal radio produces energy out of its piezoelectric speaker: Sound energy, that is. This energy *has to come from somewhere*, and in the crystal radio's case, it is harvested directly from the incoming electromagnetic radio waves.

A crystal radio receiver.
Image credit (https://en.wikipedia.org/wiki/Crystal_radio)

Other devices, such as passive RFID tags, also work by extracting energy from electromagnetic fields.

This device works on the same principle: Extract useful energy from electromagnetic waves. Here's the schematic for the circuit (found by me online - I didn't design this, although it is pretty simple)

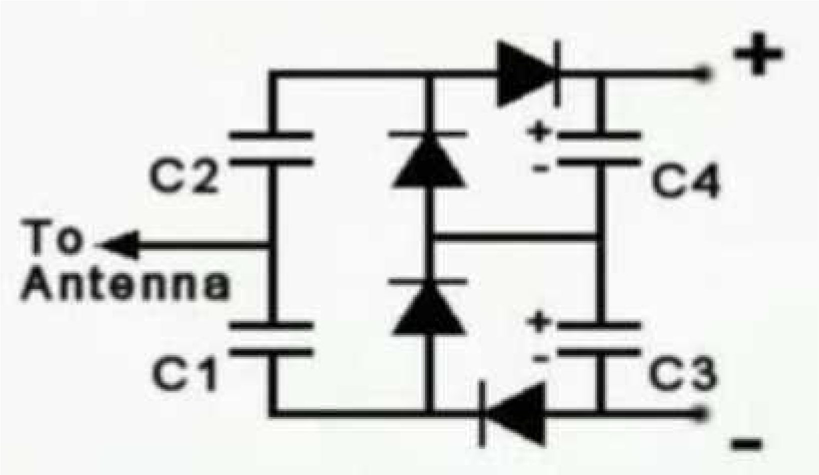

Image credit (https://electronics.stackexchange.com/questions/248525/circuit-for-free-energy-from-radio-waves)

Capacitors C1 and C2 serve to isolate the antenna. The antenna in my case is made from a piece of copper wire soldered to the two capacitors. The wire is probably 20 centimeters long and was not optimized with any particular frequency in mind (this was a quick project). The four diodes form a full wave rectifier that converts the incoming alternating current produced by the inbound radio wave into direct current. Capacitors C3 and C4 then serve to store the harvested energy in an electric field - think of them as tiny batteries. For C3 and C4, I used 100 microfarad capacitors rated to 35 Volts. These are very small capacitors but will store enough energy to easily measure on my cheap multimeter.

Assembly

It's important to note that you can't just use any diodes for this and expect it to work. The diodes used need to be able to handle sufficiently high frequency and have a low voltage drop. To handle this task, I used the **1N34A Germanium diode** (I got some on Ebay for a reasonable price). You can peruse the datasheet here. (http://www.nteinc.com/specs/original/1N34A.pdf) In particular, these diodes are used for *FM radio detection* - this is essentially the same thing the energy harvester circuit is doing, since it is rectifying an input signal close to the typical FM band in frequency. Note that although the datasheet states that the forward voltage drop is 1 volt, this drop is much lower at the low currents we are dealing with here.

1N34A Germanium Diode
Image credit (https://www.radioshack.com/products/nte-1n34a-germanium-diode)

For assembly, I used a broken piece of perfboard I had lying around and a soldering iron to quickly connect all of the components. It is critical to get the polarity on the electrolytic capacitors correct. For the two capacitors connecting the antenna to the circuit, I used ceramic capacitors that don't need to be hooked up with a certain polarity. Once again, I used a badly optimized piece of wire as an "antenna" just to see if this would work. I unfortunately don't have the knowledge of RF theory yet to go much further, but the wire ended up working out just fine.

Here is a close-up shot of the final circuit. Yeah, I know it's ugly.

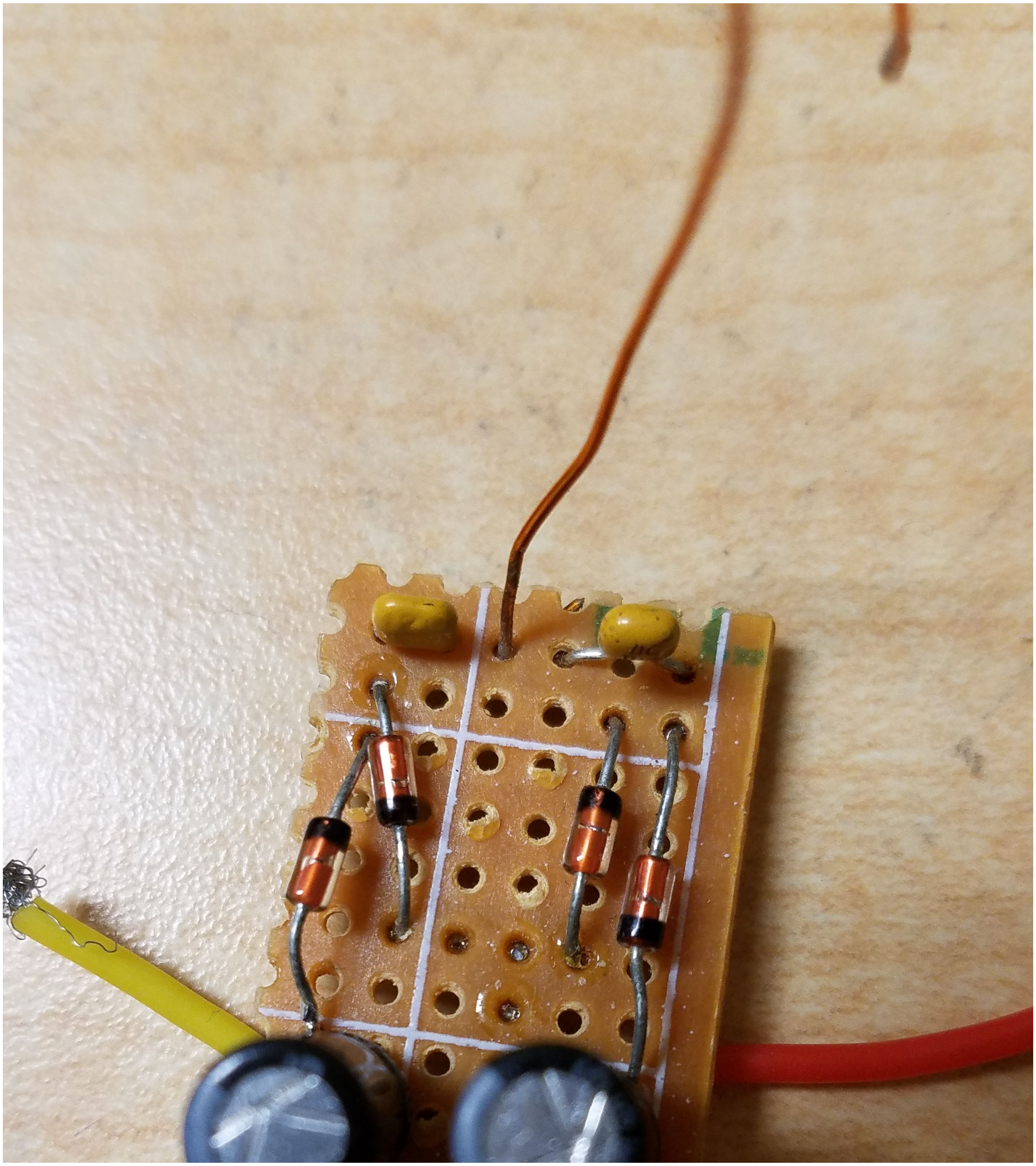

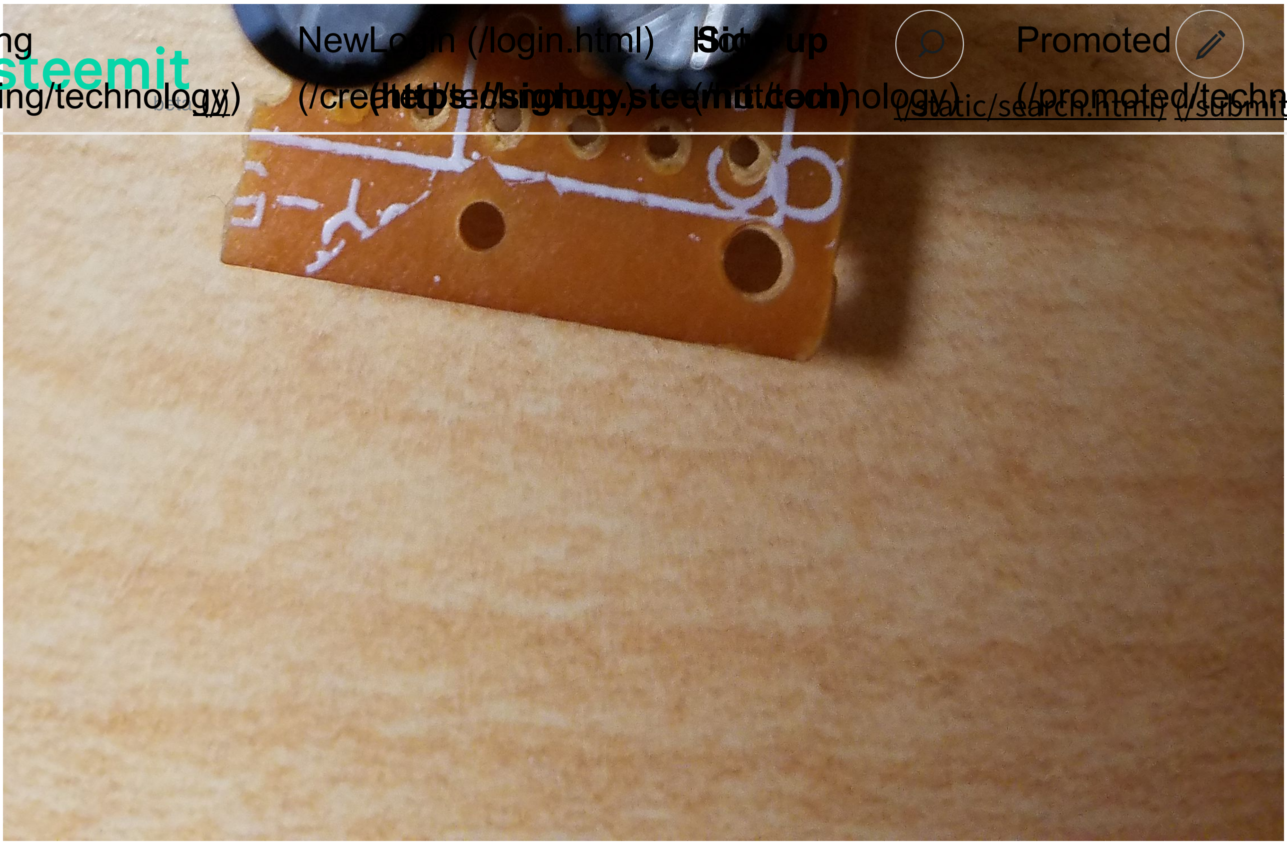

The red and yellow wires are just to make it easy for me to read the voltage on the storage capacitors - they are completely unnecessary.

I originally expected to get tiny (but detectable) energy generation, picked up on C3 and C4, just by leaving the circuit out in the room I built it in. When this didn't work, I thought it was broken. Turns out it wasn't: A ham radio transmitter confirmed that.

Using a Baofeng UV-5R handheld radio transmitter on the VHF band (~140 MHz) near the harvester causes the storage capacitor voltage to spike - so much so that I would be worried about them failing if the transmitter was left on for more than a few seconds. In a test I just did a couple minutes ago, quickly transmitting around 142 MHz (4W transmit power, nobody was using that frequency at the time) from a few inches away from the harvester caused the storage capacitors to quickly spike up to over 16 volts - not insignificant at all. This voltage was measured across the two electrolytic capacitors in series. In fact, this is easily enough to light an LED for a brief moment (or even break it, since LEDs are typically not rated for 16 volts).

VFO/MR
A/B
BAOFENG
BAND
MENU
EXIT
1 STEP
2 TXP
3 SAVE
* SCAN
4 VOX
5 WN
6 ABR
0 SQL
7 TDR
8 BEEP
9 TOT
#
UV-5R

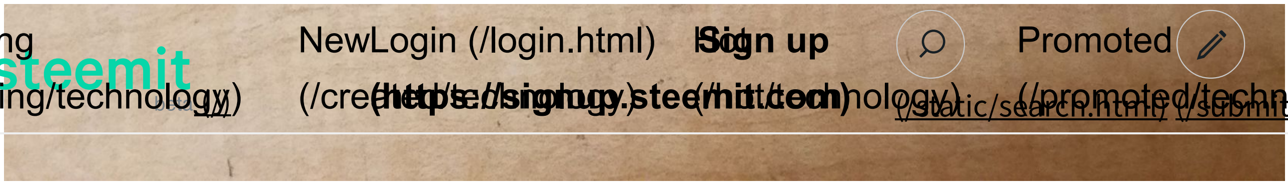

My UV-5R transceiver. I used this handheld to test the harvester circuit above. If you're looking to get into ham radio, I highly recommend this device because it's very cheap (~$25), is somewhat sturdy, and offers UHF and VHF transmission/reception.

The UV-5R transmitter proves conclusively that the harvester is actually working ... somewhat. While you can harvest pretty decent amounts of energy an inch away from a four watt radio transmitter, this isn't useful outside of an experiment since ambient RF is nowhere near that energetic. Detecting ambient RF with this thing would be nice ...

Well, I was able to do that too with this circuit. I just conducted another test where I shorted the storage capacitor and measured the storage capacitor voltage after 1 minute. The result: 18 millivolts. Here's the multimeter screen:

ETEKCITY
MEASURE UP
018

OLD
200M
OFF
200m
2000 m
20M
20
V⎓
200k
200
Ω
20k
50
000
50
V
200
200

Various other informal tests returned similar results. For some reason, the circuit does actually harvest tiny amounts of energy from radio waves in my apartment but not in the room where I originally built the circuit. Not sure why there is more RF here, but whatever. The capacitors hooked up to the harvester charge themselves without external power source ... other than the radio waves transferring energy to the circuit, of course, to conserve energy.

Of course, 18 millivolts (that's 0.018 volts ... you should be laughing here) is pretty damn useless. But once again, this isn't exactly intended to be a useful product. There really isn't a lot of energy in ambient RF waves floating around to begin with. This just showed that the circuit works on both low and high levels of incident radio waves.

Conclusion

I don't personally see a practical use for this device. Ambient energy collection produces too low a capacitor voltage to be useful (although I may try some lower capacitance capacitors to see if I can get a higher voltage more quickly, which could then be potentially useful like my nuclear battery was). While you could technically produce okay amounts of energy by strapping this to the back of your ham radio and collecting whenever you talk, you are technically hindering your transmission range in the direction of the harvester by doing this - again, the energy has to come from somewhere, and if you are taking energy of out of the radio waves leaving the transmitter, those absorbed waves don't get to make it to the receiver of whoever you are calling. For this reason, putting such a harvester on, say, your cell phone, would just make it more difficult to receive calls from cell towers in the direction of the harvester, since the harvester is essentially providing partial shielding by absorbing some of the RF emissions.

If there was a way to produce practical voltages from ambient RF using this, then I could see this being useful (even if just as a gimmick). This is of course totally possible, and electronics manufacturers have already built circuits that can do this. It's just a lot harder with DIY tools, big components, and only minimal understanding of RF antenna theory. But I'll probably keep trying and keep you posted if I have any luck.

I hope you learned something from this project! If you have any questions, comments, corrections, or rants, feel free to let me know. I hope to eventually improve this device a bit, so let me know if you have any suggestions for improvement!

Thanks for reading!

Images not credited are my own. You are welcome to use them in your own posts with credit.

technology (/trending/technology) diy (/trending/diy) steemstem (/trending/steemstem) science (/trending/science) education (/trending/education)

2 months ago by **proteus-h** (61) (/@proteus-h) Reply 3 (/technology/@proteus-h/my-diy-radio-wave-energy-harvesting-circuit) 152

$54.96 27 votes

Authors get paid when people like you upvote their post.
If you enjoyed what you read here, create your account today and start earning FREE STEEM!

Sign up. Get STEEM

Sort Order: **Trending**

steemiteducation (63) (/@steemiteducation) · 2 months ago (/technology/@proteus-h/my-diy-radio-wave-energy-harvesting-circuit#@steemiteducation/re-proteus-h-my-diy-radio-wave-energy-harvesting-circuit-20180110t105559787z) [-]

Your post has been selected by
STEEMIT EDUCATION
Please follow & support us by upvoting, resteeming & following
our blog...
@steemiteducation

$0.07 1 vote Reply

drmake (53) (/@drmake) · 2 months ago (/technology/@proteus-h/my-diy-radio-wave-energy-harvesting-circuit#@drmake/re-proteus-h-my-diy-radio-wave-energy-harvesting-circuit-20180108t201511765z) [-]

I recently read about a company that is transferring power via waves between 1-5m. There must be some wave range where this works quite well then. Dont remeber where it was.

$0.00 Reply

proteus-h (61) (/@proteus-h) · 2 months ago (/technology/@proteus-h/my-diy-radio-wave-energy-harvesting-circuit#@proteus-h/re-drmake-re-proteus-h-my-diy-radio-wave-energy-harvesting-circuit-20180108t223403628z) [-]

Wireless power definitely works, but it's usually pretty inefficient. After writing this I had the idea of using this harvester as an RF detector by adding an LED that lights up when it gets close to something transmitting by using energy in the storage capacitors. This would let you detect stuff like FM transmitters without seeing them. I'm going to try to throw it together and make a post in a few days about it if you're interested.

$0.69 2 votes Reply

Enhanced Mystery Crystal Set Also Pulls in FM

copyright © 2006 by Larry J Solomon

The Australian "Mystery Crystal Radio" has been praised since its original publication in 1932, but I don't think anyone has ever mentioned its capability for FM reception. To clarify somewhat, it is the Mystery Plus set, rather than the original circuit to which I am here referring. The original Mystery set had two coils wound in tandem, whereas the Plus set has three coils -- an extra one for antenna tuning. The only other substantive change I have made is an additional variable capacitor on the ground circuit, an outboard rock detector jack, and an audio output jack for playing through a sound system (optional). The capacitors cannot be ganged. Here is the circuit.

Enhanced Mystery Crystal Radio

with Optional Audio Output and FM Detection

© 2006 Larry Solomon

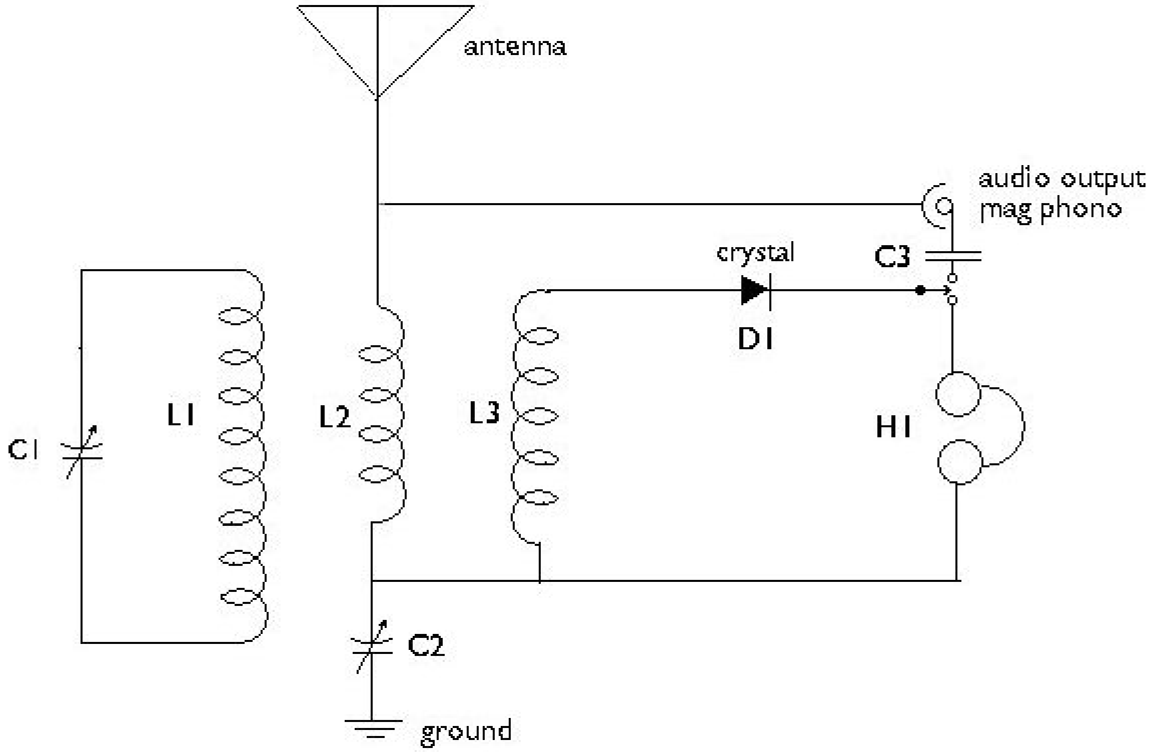

Parts List

C1 350 pf air variable capacitor

C2 350 pf air variable capacitor

C3 2750 pf capacitor

L1 45 turns 22 gauge magnet wire

L2 15 turns 22 gauge magnet wire wound on top of L1+L3 and separated from them by a layer of duct tape

L3 25 turns 30 gauge magnet wire wound parallel to L1 after the twelfth turn of L1

H1 high-impedance headphones

D1 Detector: 1N34 germanium diode or crystal rock detector

Coil form 2.75" diameter and 3" long (PVC pipe)

*original Mystery circuit: The Sunday Mail - Brisbane, Australia, April 16, 1933

This set is an amazing performer. It has crystal clear reception, good sensitivity, and excellent selectivity. To my surprise, in addition to the expected panoply of AM stations, I heard a faint signal that I could not

tune out. At first I could not identify it. To my surprise I heard the announcement "KiiM FM, 99.5"! This is a country music station here in Tucson. It was all over the dial, untunable, but the much louder AM signals masked it when they were tuned in.

I set myself the task of trying to improve the FM signal. I tried some simple circuit modifications, which did not improve anything. Then I connected a dipole antenna instead of the AM antenna I normally use. Suddenly, the FM signal was much clearer, although still weak. By using the audio output and sound system amplifier, I was even more amazed that four different FM stations came in loud and clear. I found that changing the telescoping antenna length and position I could tune the stations in and out. They were KRQ, KLPX, KiiM, and KHYT all local FM stations with transmitters nearby. Their reception was also affected by the length and position of the audio output cable.

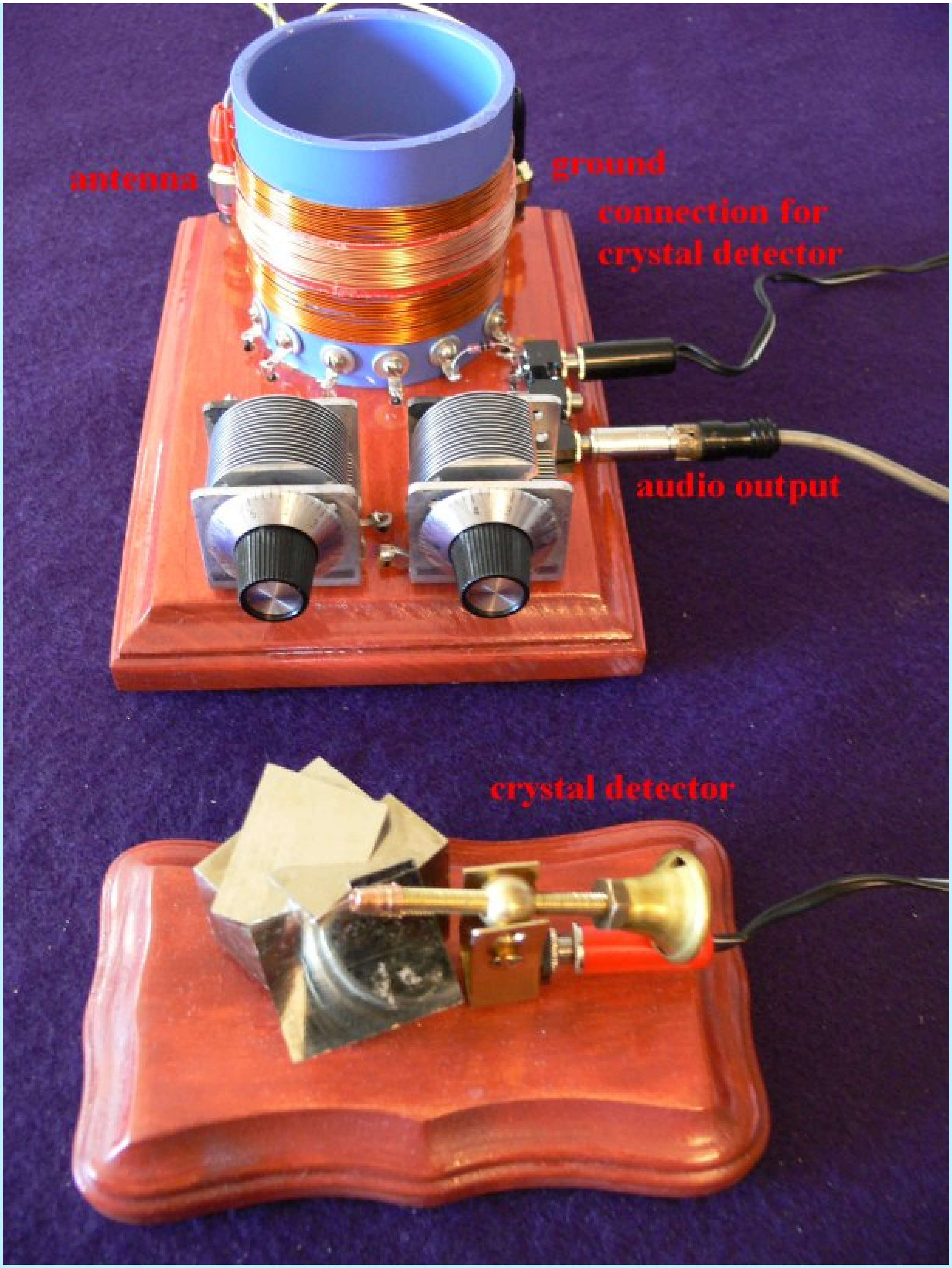
antenna
ground
connection for
crystal detector
audio output
crystal detector

The "rock" is a beautiful "twinned" pyrite from Spain. I have found that many metallic rock crystals will work as detectors, including galena, molybdenite, various copper ores such as "Peacock" copper, hematite, and zincite. Watch for a future webpage on crystal detectors.

layout

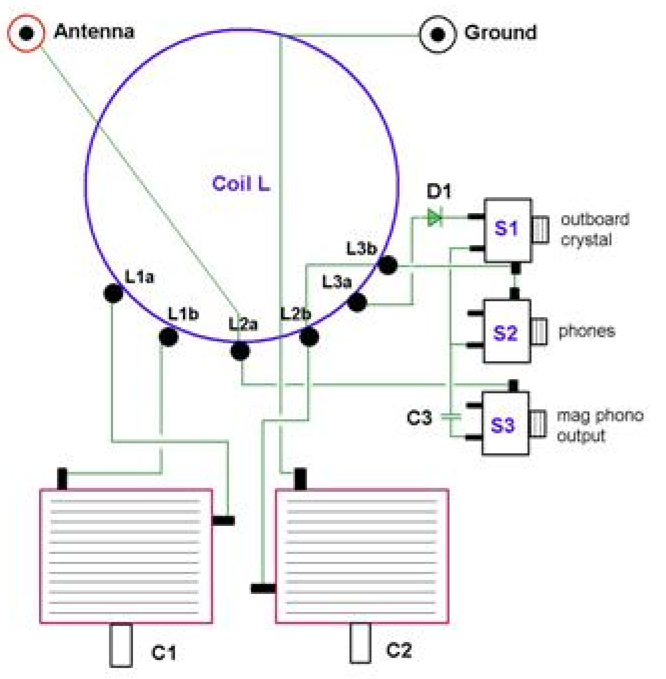
Antenna
Ground
Coil L
D1
S1
outboard crystal
L3b
L1a
L3a
L1b
L2a
L2b
S2
phones
C3
S3
mag phono output
C1
C2

Parts List

C1 350 pf air variable capacitor
C2 350 pf air variable capacitor
C3 2750 pf capacitor
L1 45 turns 22 gauge magnet wire
L2 15 turns 22 gauge magnet wire wound on top of L1+L3 and separated from them by a layer of duct tape
L3 25 turns 30 gauge magenet wire wound parallel to L1 after the 12th turn of L1
D1 Detector crystal (1N34 diode)
Coil L 3" dia x 3" long PVC pipe
S1, S2, S3 miniature phone jacks (mono)

Home

Electronics Repair

Make Electronics Repair
Great Again

DON'T MISS LED TV T-Con Board Problem

AdChoices RF Strength Fix

Home > Guest Post > Building Your Own RF Field Strength Meter

Building Your Own RF Field Strength Meter

By *Jestine Yong* on *June 1, 2013*

29 SHARES SHARE TWEET SHARE SHARE 34 COMMENTS

A simple and easy project for a rainy Sunday!

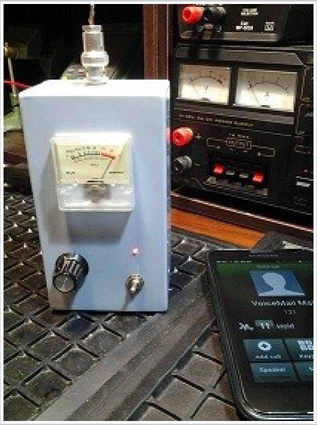

Why would you need one? Sometimes we build electronic projects for fun only, or just for satisfying our technical curiosity; this is what I do sometimes. This project, however, started with a need.

The need originated when a friend called for help to solve an intermittent problem with the UHF transceiver on his boat. The receiving ends reported frequent cuts in his transmission.

I quickly improvised a very simple Field Strength Meter composed of a small antenna, a detector diode, a resistor, capacitor and a Micro Ammeter. Assuming that this simple circuit would be sensitive enough to detect the HF field, providing we keep it close enough to the transmitting antenna. This worked quite well, and the intermittent problem was easily diagnosed and fixed; it was caused by a broken antenna cable, making intermittent contacts. With the meter we were able to see a clear drop of the field strength when moving the cable.

From there, the technical curiosity took over; what if we could improve the circuit, making it more sensitive and, how high in frequency it would still work? Would it work for much higher frequencies such as 2.4 GHz, to check RF fields from Wi-Fi routers and mobile phones?

I gave up (at least temporarily!) to the idea of building a wide range RF amplifier to amplify the antenna signal over that wide range of frequencies. This might be another project for another rainy Sunday (or many more rainy Sundays!). Instead, I preferred trying to amplifying the DC signal after the RF detection and see how far we can go.

And here we go; we have a very simple instrument that can detect the activity of your Wi-Fi router, your mobile phone and about any RF transmitter through a wide range of frequencies.

Interview Report
How To Be A Successful
Electronics Repairer
Jestine Yong

Jestine's Free Report

Enter your name and email below to get Free Instant Access to the Free Report "How To Be A Successful Electronics Repairer".

Over 60,000 people have enjoyed my Free Report and received exclusive Electronics Repair tips.

Name:

Email:

SUBMIT

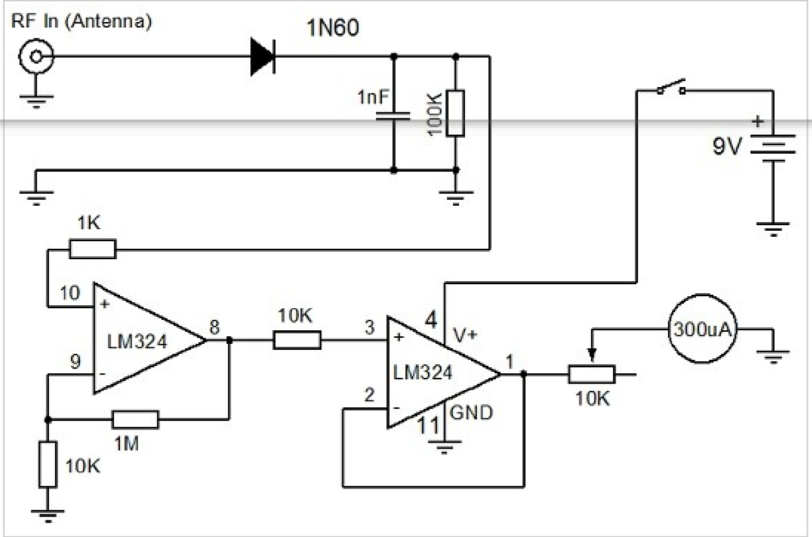

Fig. 1 Schematic diagram

It is powered by a 9 V battery. The diode is a critical component; it should have a low threshold voltage. The diode used in this circuit is a germanium diode 1N60. A ghost from the past, but it is still available on the market (Escol for example).

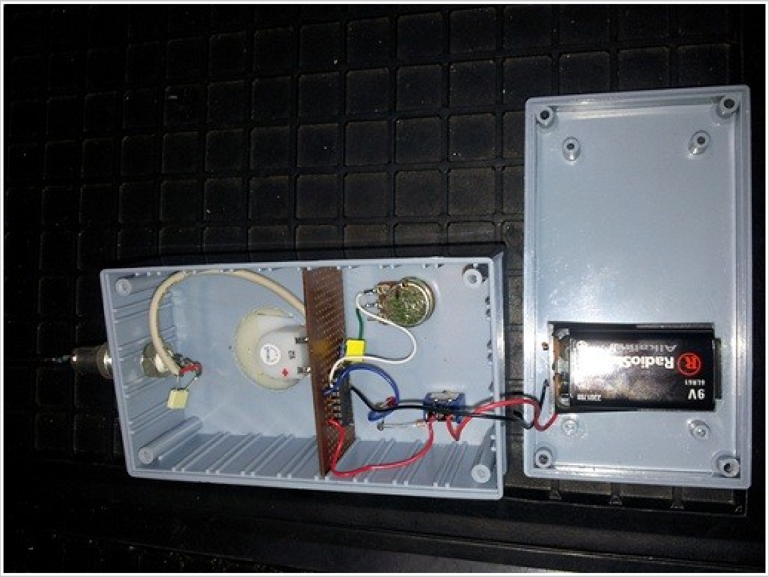

Fig 2. Inside the box. Plenty of space for further developments!

The diode and the 1 nF capacitor are soldered directly at the base of the antenna connector. In the next version of this circuit I would put the 100 K resistor there as well.

Another critical point is the antenna. It should be cut to ¼ wavelengths for the frequency range you

1 PLC Calibration ISO 17025
Over 60 years combined experience in Electronics and RF
plccalibration.com

2 Radiation Monitors
Measure radiation doserate and environmental contaminants.
tracerco.com

3 #1 VPN Service
Fastest Speeds. 100% Anonymous. No Logs. Money Back Guarantee.
IPVanish VPN

4 PCB Topological Autorouting - with hand routing results.
Finally an autorouter that doesn't require cleanup.
simplifyda.com

5 Better Than Bitcoin
Here are the top 5 cryptocurrencies that will overtake bitcoin. Free report.
Early Investing

ELECTRONICS REPAIR CATEGORIES
Electronics Repair Categories
Select Category ▼

REPAIR TIPS POPULAR COMMENTS

Lada 111 1.6L 16V 2005 Year – Immo Off
...
January 9, 2018 7

Servicing Technics RS-M85-MK2 Tape Deck
...
January 9, 2018 4

HOME ABOUT JESTINE YONG REPAIRING E-STORE FREE SUBSCRIPTION TECHNICAL TRAINING
MUST HAVE-BLUE METERS EXPERTS KENT'S EBOOKS DAMON'S EBOOKS

21 Inch Hotpoint CRT Television (Model 21JI8US) Wide Horizontal Lines Repaired
...
January 3, 2018 26

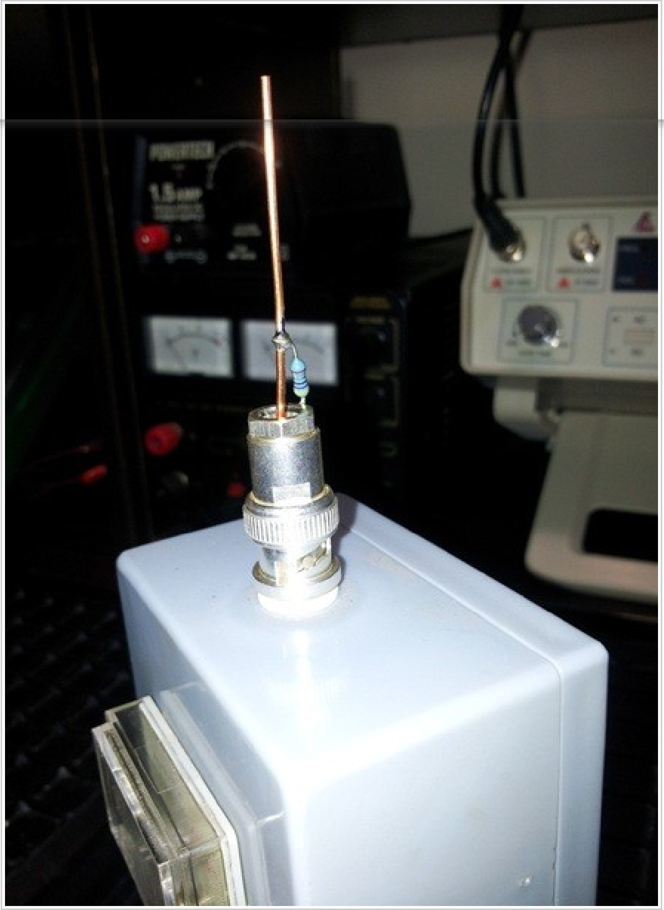

Fig 3. The antenna for 2.4 GHz

This simple RF Field Strength Meter proved to be much more effective than expected. It will react to your Wi-Fi router and mobile phone transmitting activities. You can also check the leaks of your microwave hoven, but be careful, put the sensitivity button to a very low position as the signal proved to be very strong! I guess this is normal…

Desoldering With The Help Of Medical Needles (Hypodermic Needle)

…

December 26, 2017 22

Servicing Coby LCD TV Monitor

…

December 26, 2017 6

Noahtech Electronics Training
The Best Hands On In Technical Training

LIKE US ON FACEBOOK

Noahtech Electronics …
1,574 likes
Like Page
Share

Be the first of your friends to like this

Album **Advanced Electronics Repair Course**

Noahtech Electronics Training
about a month ago

Just completed the Advanced Electronics Repair Course for two participants from Malaysia

NOAHTEC TRAINING CENTER
Advanced Electronics Repair

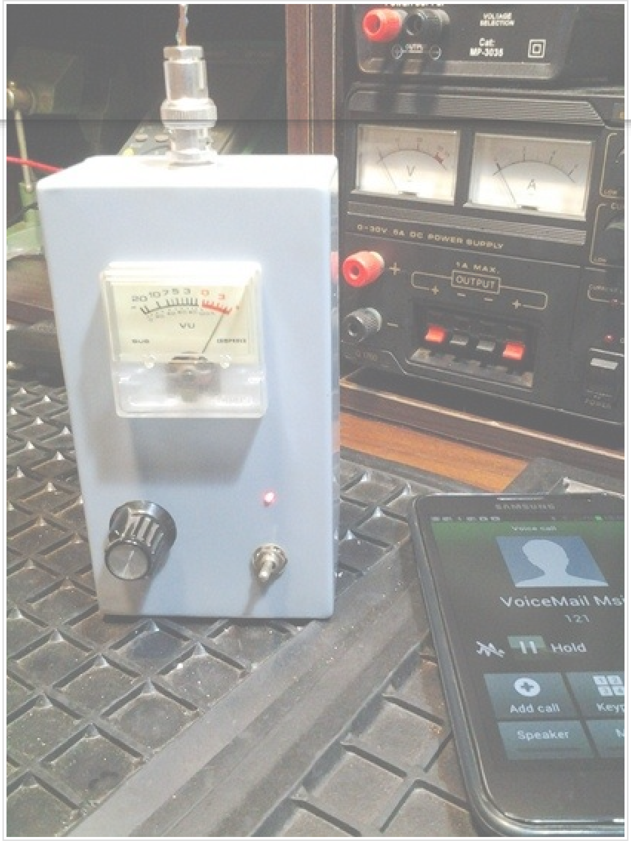

Fig 4. Reacting to the mobile phone transmission.

Note that the sensitivity button is at a quite low position, showing that the device it quite sensitive at those frequencies.

This project can be built with components we have in our drawers. Who want to go out buying components on a rainy Sunday? I used a Vu-Meter from a scrapped amplifier for the indicator. The scale is not important as the measurements are comparative anyway. Those Vu-Meters are quite sensitive. The OP Amp can be any one of the common type as we amplify DC only. I used one fond in a drawer…

Possible improvements:

– A power jack so we can power the device with an external power supply when it is convenient, saving the precious battery life.

– A better antenna with a better gain. This device can be used to compare the performances of various antennas.

– A wide range RF amplifier… maybe later.

This article was contributed by Gerald Musy from Penang, Malaysia

Note: If you have any interesting repair article that you want to share to others please do email me HERE

Please give a support by clicking on the social buttons below. Your feedback on the post is welcome. Please leave it in the comments-thanks.

Likes (88) Dislikes (0)

Kent Liew's LCD/LED/Plasma TV Repair Ebooks

Damon Morrow's LCD/LED/Plasma TV Repair Ebooks

NEW E-BOOK SMPS Modification Click Here Now To Be An Expert

Click Here to be an EXPERT In Switch Mode Power Supply Repair

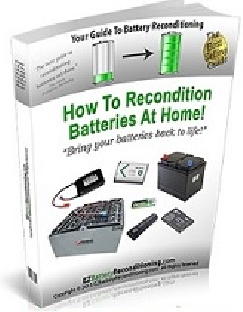

Click Here Now

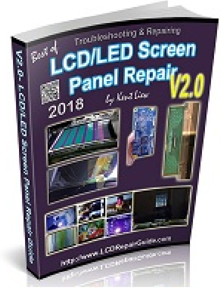

Click Here Now

FM Crystal Set

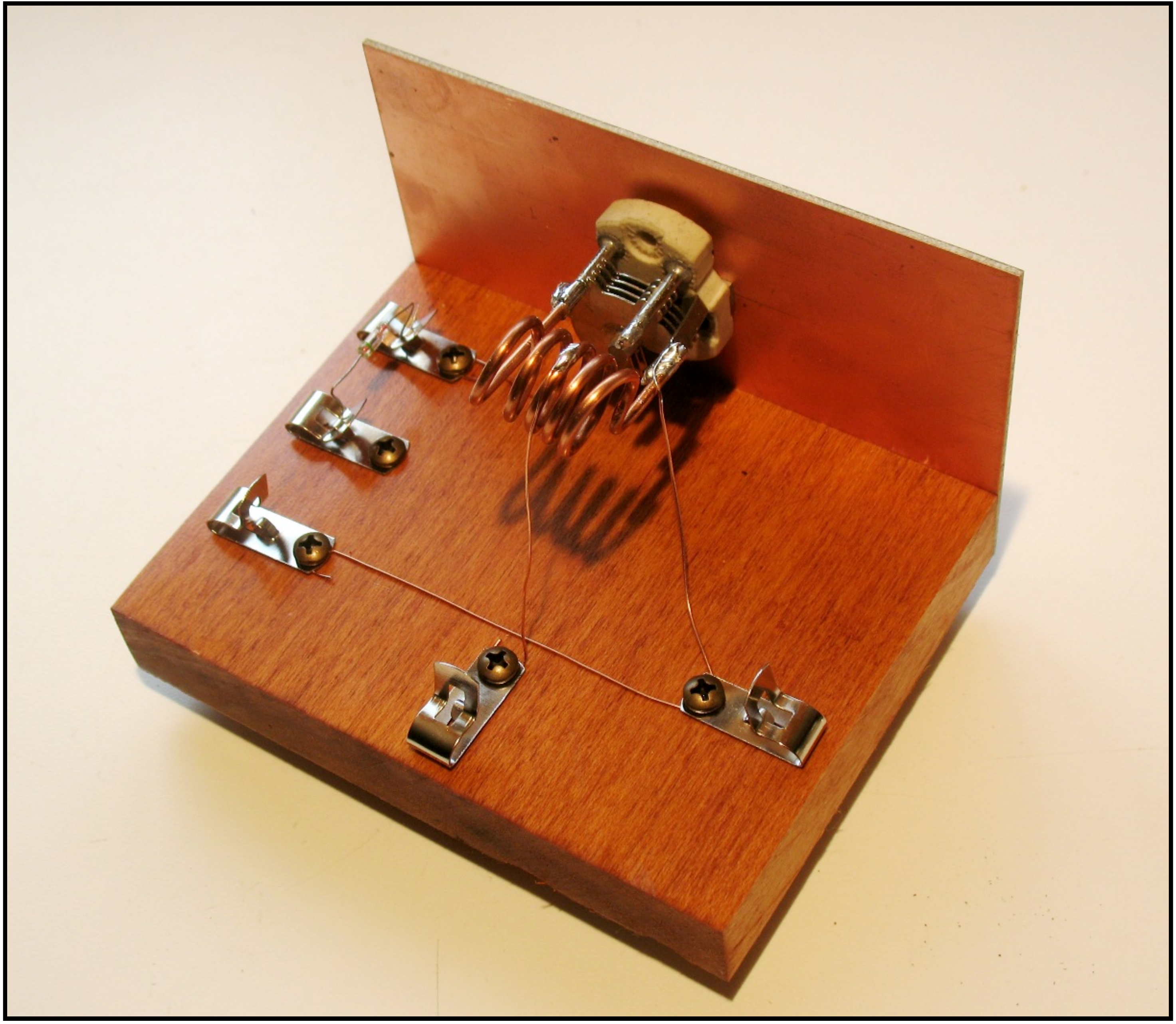

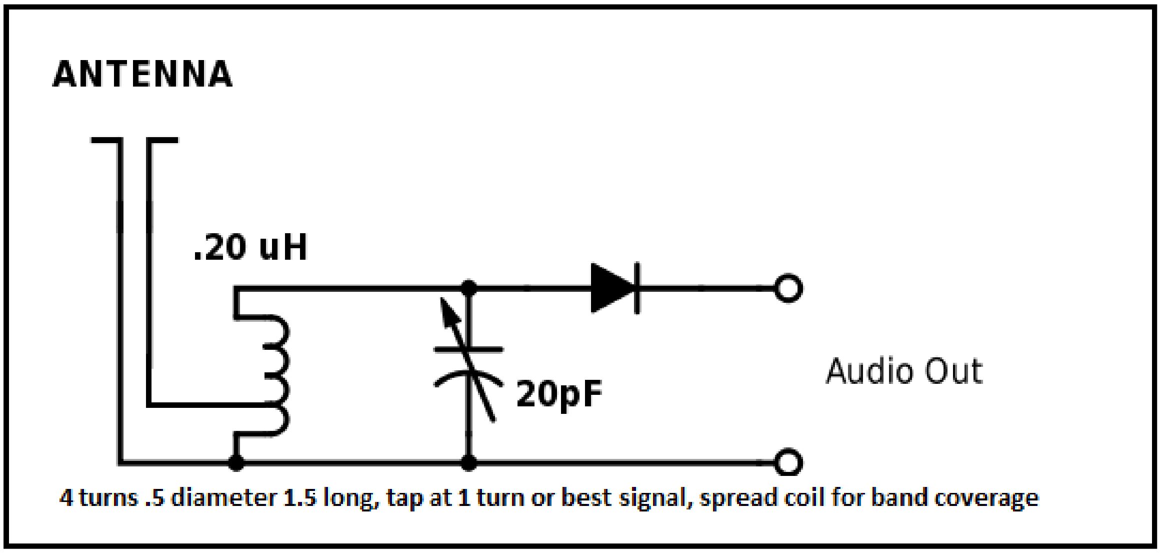
ANTENNA
.20 uH
20pF
Audio Out
4 turns .5 diameter 1.5 long, tap at 1 turn or best signal, spread coil for band coverage

By slope detection you can get FM. If you live close to the stations you should do well with Hi Z phones. I use an audio amp, I am far from the stations. I get three.

Fighting the pollution by public radio broadcast antennas

Un texte en français :	La pollution radio par les émetteurs FM
Een tekst in het Nederlands:	De radiovervuiling door FM-stralen
A petition:	www.petition-electrosmog.be

The electromagnetic pollution has now reached such a level that not only hypersensitive persons are being harmed. The debate most often focuses on cell phones, their relay antennas, Wi-Fi routers and similar technologies. It is often pointed out that domestic appliances like power saving lamps and cheap electronic devices, are harmful too. Recently, I came to the conclusion that I too had problems with electromagnetic radiations. Yet after a few months investigating, I was surprised to find out that far out the main cause of my problems was nothing I had heard about. A 100 MHz FM public radio broadcast antenna is located atop a local hospital and burns the whole city with an astounding force of 2 V/m (pike value).

I found other persons that are harmed. For example people who develop severe headaches or tiredness after a few hours working in a given place. They have no problems in other places, that are less exposed to the FM radio waves. It seems to me that people who live in strongly exposed habitations often develop psychiatric problems. My guess is that probably several thousand persons in the city are severely affected. Maybe tens of thousands are affected but without getting obviously sick.

Most puzzling is the fact that this hammering by radio waves is completely useless. It wouldn't change anything to the quality of the reception by the listeners in the city, if the force of the radio waves was divided by 100 or preferably 1,000. The purpose of those antennas is to emit far away in the countryside. That's why they are located atop hills or tall buildings and they have a huge power. But, due to their rudimentary construction, quite much of that power is wasted and hits the nearby city.

It is legal to hit the ground with such force. Whether I measure in rooms or in the open, it is at worst a little below the norms. Hence the problem is with the norms. They keep being too high, despite the efforts of many scholars and health organizations. The situation is catastrophic. If the people did understand the harm that is done to their children, revolts would break open.

Below is a picture of a cheap calculator that I transformed. It contains no more battery nor solar power cell. Instead it draws its energy from a 1 meter long antenna. Anyplace in the city center, in direct view of the emitter, the power of the radio waves is strong enough to feed the device. I made this because nobody was frightened by the figures I was reporting, except some friends that are into electronics. Everybody knows that a car can kill even at 30 km/h speed but almost nobody has any knowledge about radio waves figures. When I showed the calculator, everybody understood and some whitened. (The calculator also works in some classrooms where I tried it out. The emitter was visible through the windows... Should I warn the parents of the students that failed their exams?)

Powered by 000webhost

This didn't help anyway, so I invented another device: the "snake". The picture below shows six of them, an early version.

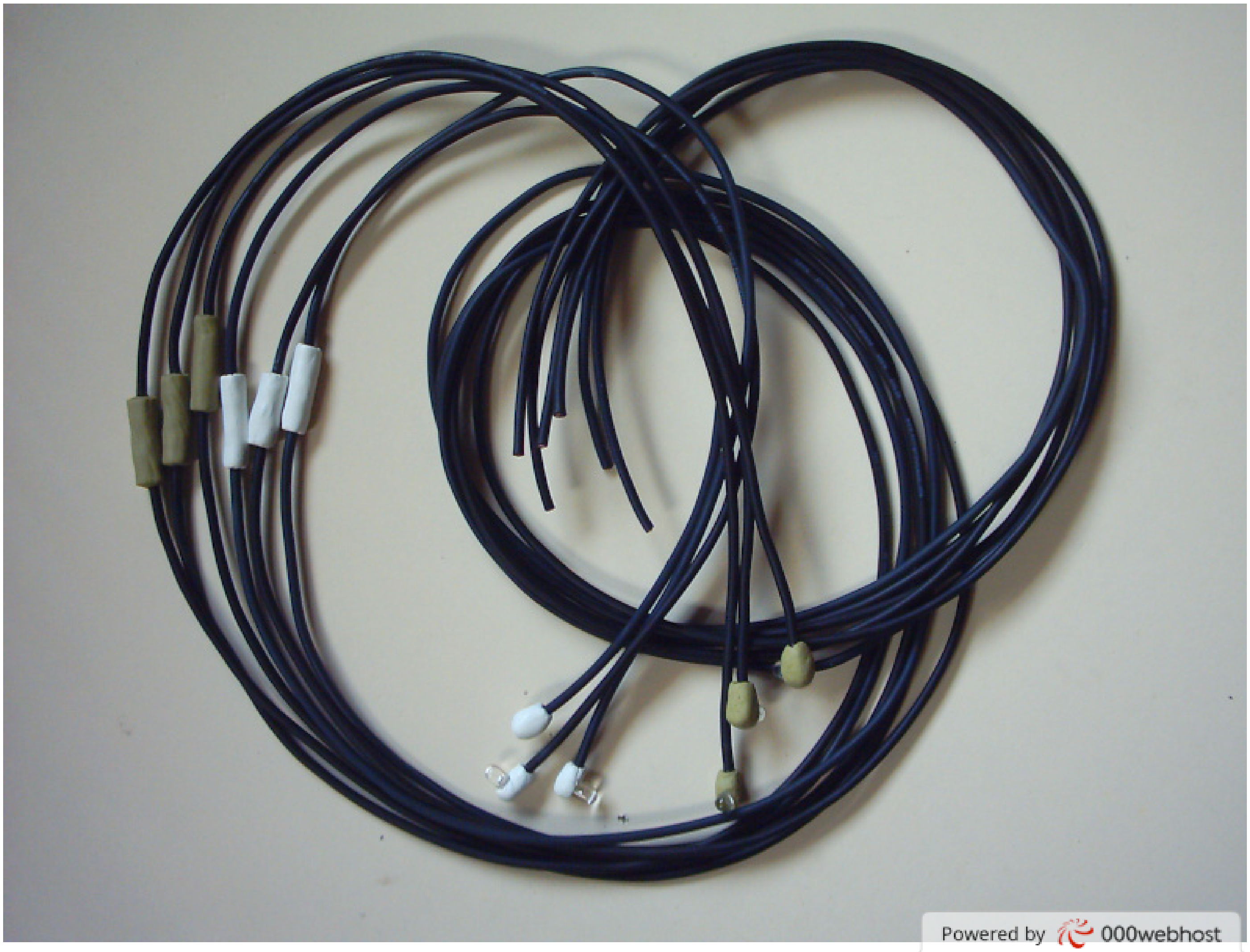

Powered by 000webhost

You just hang the snake in the flux of radio waves and the LED lamp at one end will bright up, clearly visible in daytime and enough to light and read a text at night.

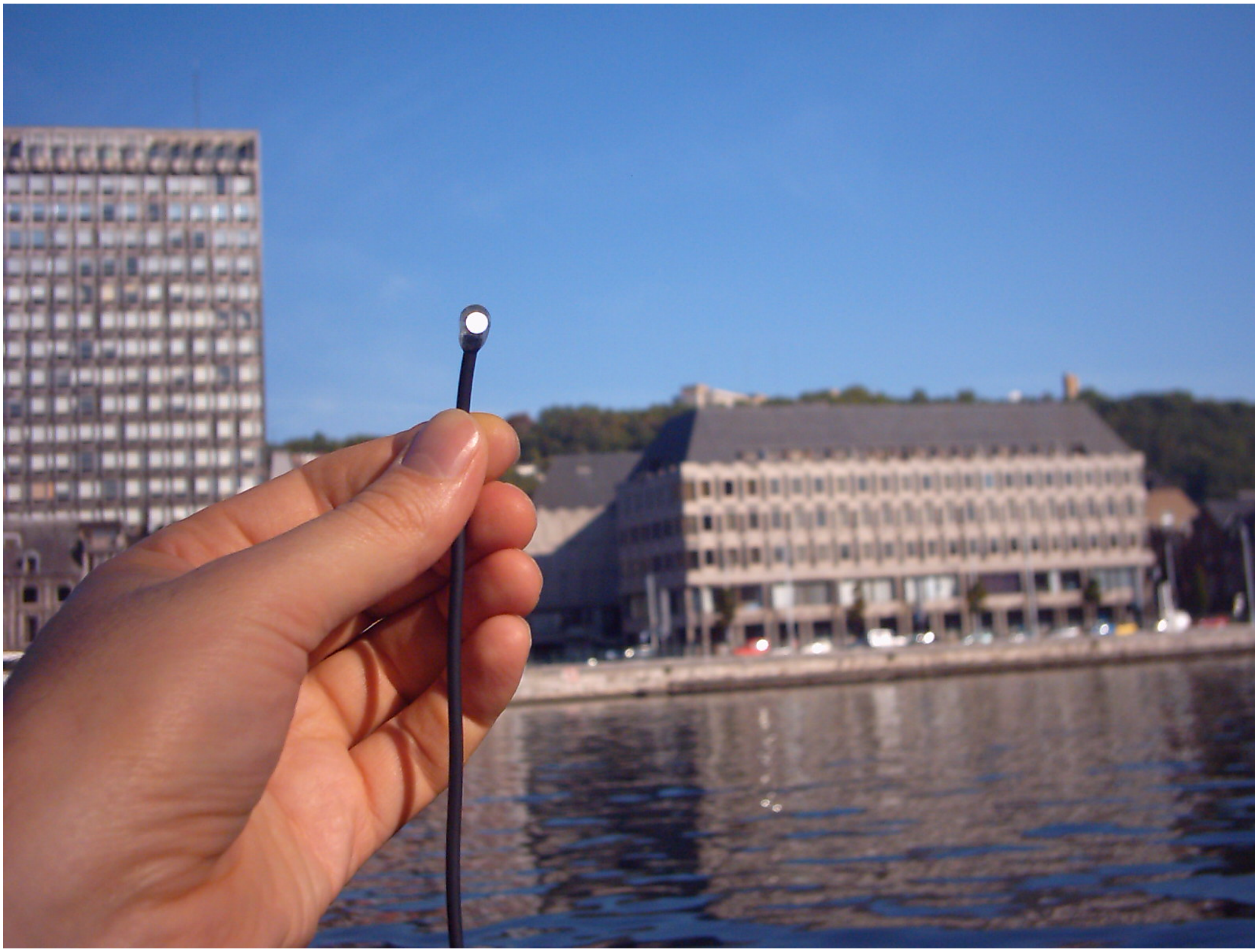

The snake is less impressive than the calculator but it is much cheaper and easier to make and it can be sent in an envelope. I made tens of them and sent them to politicians, newspapers, universities... I gave some to local people, together with a user guide on a sheet of paper. Just like the calculator, the snake contains no battery. The lamp lights up solely on the power of the radio waves. The picture below shows a later version, that can be rolled up in an envelope that fits the conditions to be sent with only one stamp.

Powered by 000webhost

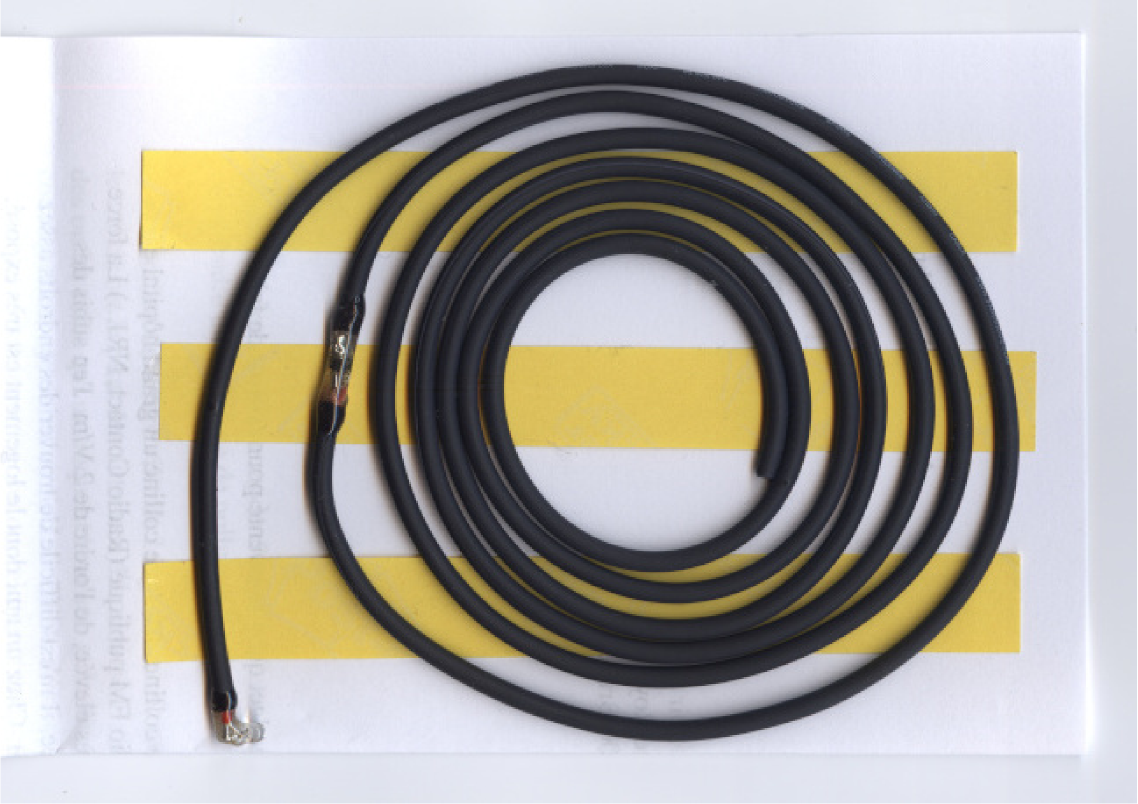

However mad it is to send such power on a city, nothing has been done yet against this. It's like an 19th century chemical plant sending its pollutants all around. Some complain but others argue that closing the factory would mean people laid off. Some then reply that being employed to feed children that die from ugly sicknesses may be nonsense... And so on, except for the fact that the chemical factory indeed had a purpose while the strength of the radio waves we're talking about, has no serious purpose at all.

Something common to all heavy sources of electromagnetic pollution is that it wouldn't cost much to make them harmless. For example, high tension power lines can be made harmless by feeding them DC current instead of AC current. This requires some more heavy electronics at the input and the output but the price of it is a detail in the whole. I studied the schematics of power saving lamps and simply noticed that they were conceived by people who do not understand the propagation of radio waves. Adding the necessary components to make the lamps radio-silent would almost not increase their prices. Enforcing proper norms will not change our way of life, it will just stop to favor the sociopathics amongst the industrials and lobbyists.

So, what can I do to help people understand the situation? They need to appropriate the scientific knowledge involved. Most assume that the necessary studies will be made by labs and then the governments will adapt the laws accordingly... That would be OK if it wasn't for the trillions of money involved. A minister in Belgium tried to prevent sugar dispensers in schools, she immediately got a made-up scandal on her shoulders. All she ultimately managed to get was to have some health advice notices glued on the sacred dispensers...

Flows of observations are now available, showing the effects of radio waves that match the norms: brains of rats destroyed, people unable to work, eggs that don't hatch, quarreling birds, increased risk of cancer, cows aging abnormally... I myself had seeds germinate and saw them die quicker when closer to a power saving lamp. Their strategy to cope with this is simple: "OK, there does exist scientific evidence of some danger... but there also does exist evidence that there is no danger at all! So let's balance the whole..." The norms are lowered but kept above what cell phone operators need for the current systems. The fact in itself of making a balance between scientific evidence is nonsense. Either a study proves a given danger and it can be verified, or bust. But anyway, what are those studies that tend to prove that there would be no danger? Lots of them are simply payed for by the cell phone industry... Yet some are perfectly serious, for example those concluding that radiations within the norms will not heat the brain or the body abnormally. This is perfectly true. There is no problem of scientific fairness involved with these studies. But, is a bullet harmless because it only very slightly heats your bones? A more sophisticated approach is to pick out studies that have borderline conclusions: "there may be a problem but it is not clear..." The study will be assumed to conclude that there is no problem at all, while it just was an inconclusive experiment like many are in science.

Powered by 000webhost

(When I'm exposed, I tend to compulsively perform simple tasks. What if a study examines if being exposed hampers the ability to perform that kind of tasks?)

Then, there is the problem that I'm going to make people sick. Indeed, electromagnetic hypersensitivity has a psychological component. Once people understand that they are victim to the radio waves and what symptoms the radio waves cause, they can overreact and become very sick when they start feeling the symptoms or just when they know there is a strong radiation force around. The radio waves induce a physiological stress, that hampers the proper functioning of the brain. When you add a psychological stress, due to the awareness of the problem, the summation of the two multiplies... Therapies for hypersensitive persons imply to dislearn them to overreact. Some friends of mine just refuse to start talking about electromagnetic sensitivity because they don't want to embark in the sickness. They do so with everything, meanwhile they eat very good food and have lots of pleasure in life. It works... with people with a strong health and a tad of selfishness.

When you are in a cold wind and you feel a pain in the neck, you know it is due to the wind and you take cover. Almost nobody can do this with the symptoms of an exposition to strong radio waves. You feel tired, some brain fog, like if your head was in a clamp, a headache, or simply you make stupid errors... You will blame yourself or the location, possibly the noisy people around... while in fact you are mainly undergoing electromagnetic sensitivity. The home of a friend is strongly exposed. After about twenty minutes I can nor more find my words and speak correctly, unless I wrap my head in aluminum foil. There is a fair chance that a while ago, I would have blamed my friend for being annoying (he is) to the point of making me speechless. There also is a fair chance that my friend is annoying because he hangs in that electromagnetic pollution everyday. Since he lives there, he became an alcoholic, lost his job, manages to keep no girlfriend... (Everything I mention here can have other causes. Quite often, such causes just add up. For example, the person blaming the noisy people around, would maybe have felt nothing if the noisy people weren't there. Like you don't feel your bad knee when you carry no bag.)

The impact of the electromagnetic pollution is severe. It is a far too heavy, and pointless, contribution to our burdens and losses. But how do you establish, scientifically, be it as an amateur, that there indeed is a problem? My initial approach can be an illustration of this. As my home is uninhabitable, I spend most of the day working in local libraries. I noticed that I could work quite well at some places and not at all at other places. Then I saw an article taped on a board, about the effects of cell phone and Wi-Fi radiations. Okay... I made a spreadsheet, with a row for each location and a note from 0 to 10 according to my ability to work there. Although I knew it was a weak approach, I then filled in each row the strength of the closest Wi-Fi router, as listed by my laptop. There was no correlation... but this proved not at all that the Wi-Fi was innocent. Weeks later I made a nonsensical but working little assembly that made the variability of the radio waves hearable. This time I got a correlation. Mild, but obvious. The ability to work is horizontally while the intensity of the radio signals is vertically:

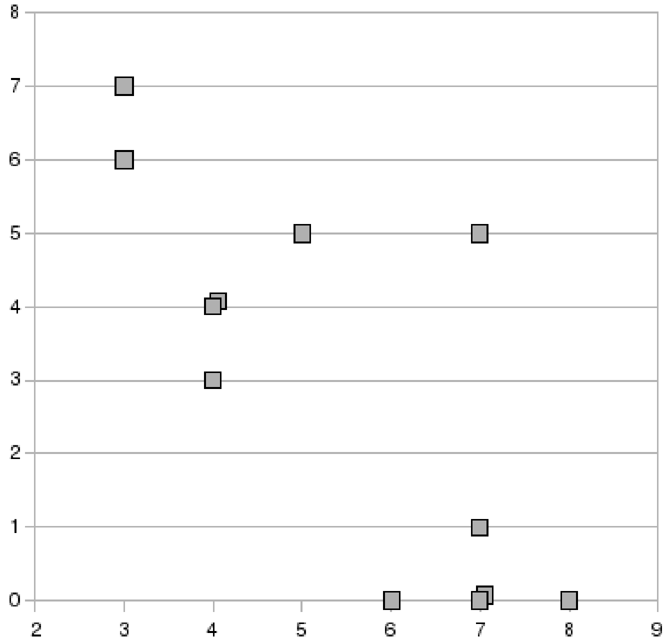

It quickly became obvious that I was measuring no strong signal in the GHz bands. I had no doubt that there were cell phone and Wi-Fi emissions coming from all directions and echoing all around but their strength was obviously

Powered by 000webhost

weak. The problem was with lower frequencies. I accused the fluorescent lamps for a while, till I finally understood that there was only one strong source, hitting me everywhere; those 100 MHz public radio broadcasts. The picture below shows the final correlation. Only one dot disagrees with the theory:

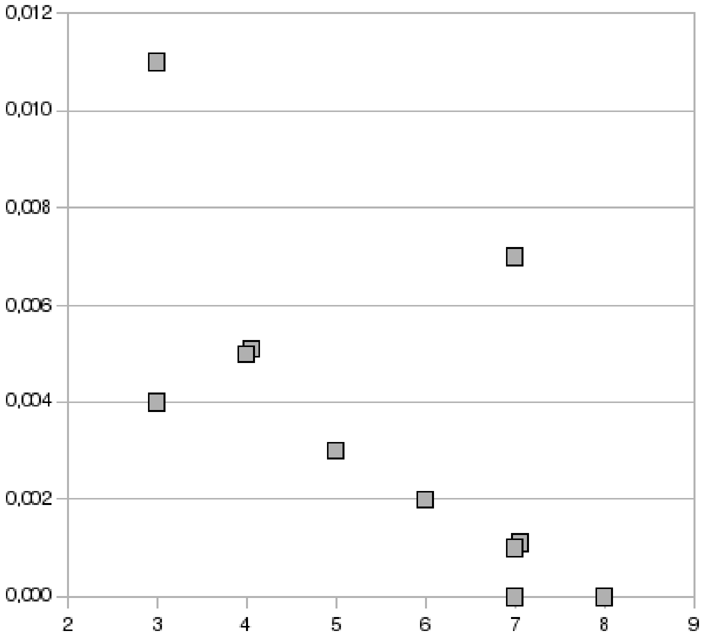

I was heavily disturbed by the radio waves, simply unable to work, long before I started puzzling on the subject. This is a strong proof that the matter is not just psychological. Reciprocally, once I understood the psychological component, I tried to use it positively, to counterbalance the effects of the electrosmog. I did get some results but I was never able to work correctly in the most exposed places. It is by no means "just psychological". It is useless "to be willing to make a small effort". I did efforts that frightened other people. One thing that does clearly help is eating a lot of sugar (Organic jam on rice bread...) (by the way, is that why sugar dispensers are placed everywhere kids are supposed to think and learn, to counterbalance the effects of the nearby cell phone relay antennas?) Alcohol does also help and I'm afraid this is why people seem to become alcoholics in exposed habitations.

Also, it does sometimes happen that I feel like undergoing strong radio waves but when I pop out my measuring device, I measure nothing serious. I have a natural tendency to be knocked-out... The radio waves are far out my heaviest problem but they are not always responsible.

I started studying other sources of radio pollution. I had an obvious benefit of removing every power saving lamp in my home. I put ferrite cores on most power chords. On one power chord that I sometimes use to feed my laptop inside my shielded tent, I had to put ten ferrites till I felt no more direct effect. Something very interesting is that till then I had no problem when using my cell phone. You guessed correctly: I started having problems, to the point that I never more hold it against my ear. I always use an earplug linked to the cell phone by a wire with two ferrite cores. I suppose that since I started avoiding the exposure to strong radio waves (which significantly improved my life and ability to work), my brains are no more constantly knocked out. So they get knocked out if I hold my cell phone against my head. And I feel it. This is one more synergy between the public broadcasts emitters and the cell phone industry: one hides the other away. 'Want to prove that inhabitants close to a relay antenna are no worse than those further away? Do the experiment in a city that is strongly exposed to public broadcasts...

So: buy a measuring tool or have a friend assemble one for you. And start comparing your problems and those of the people around you, with the level of electromagnetic pollution. If you can, make large-scale experiments with tens of persons, under advice of somebody with knowledge in statistics (to build an expressive statistic yet completely false, is *easy*). We need many such experiments, till it becomes unbearable for local interlocutors to negate the problem and not to dim the emissions. Step by step, we need to conquer our world back. Some enterprises and some countries have already set their norms at responsible levels.

One very important parameter is the rate at which the problems build up and then disappear, when exposed. Some people claim that they instantly feel when their cell phone is going to ring. I never could verify this. A friend claims that his cat reacts when he is going to get a text message on his cell phone. I tried several times to see the cat change

Powered by 000webhost

behavior while making my cell phone emit close to him but I got nothing. So I suppose that the cat reacts to the tone of the cell phone and my friend brain-confuses the timing. (I don't pretend nobody can feel a text message arriving, I'm just saying that I never could verify.) What I'm sure about is that my problems need minutes to build up, most often some ten or twenty minutes, the problem possibly becoming really painful after an hour. Once the problem built up, it needs much more time to diminish, maybe four times more. I often had to sleep a whole night on it. One day I "burned" my brains by holding a very polluting power saving lamp about 3 decimeters away from my head. I only noticed the problem after half an hour. But then it was horrible, like if a flame had been sweeping inside my head. Really just like the ache of a burn wound. I had a strong nausea and headache, that calmed down only hours later. I identified a few persons whose problems are obviously due to the electromagnetic pollution, in all cases the same applies: it needs time to build up. This means you cannot simply walk in and out of an exposed zone and claim you feel no difference. Also, you cannot get into a safe zone and claim you feel no better. If you were already poisoned, you will stay that way a long while. The only sure approach is to stay really long in a safe zone, preferably spend the night there, then go working in a polluted area. Try to be aware of what happens to you. The next day go working in a safe zone and try to see the difference. (This is a goldmine to build inconclusive studies.)

Everybody is sensitive to electromagnetic waves. The question is about the level of the sensibility. That's why people who get seriously sick with the current levels of exposition are called "hypersensitive". They are more sensitive than the average, sometimes with very painful and debilitating consequences.

There is a whole ladder of symptoms. This is roughly the ladder for me, exposed to the 100 MHz FM broadcasts:

0.01	V/m	No problem
0.1	V/m	No more able to concentrate and work
1	V/m	head in a clamp, maybe difficulties to talk
10	V/m	Tried it once for 2 hours, had to stay the whole next day in bed with an awful headache.

The serious problems begin for me at 0.1 V/m. Yet I know some people that wouldn't be called hypersensitive and that get problems similar to mine at about 0.3 V/m. A friend spent with me the two hours under the 10 V/m radiation and claimed he endured no problem... but he made surprising errors on the way back. It is very difficult for me, in my city, to find locations where the force is around 0.01 V/m. Values between 0.1 and 0.3 are common. It is easy to find building with rooms reaching 1 V/m. The force is quite predictable, according to the location and the orientation of the building, how much it emerges out of other buildings... So, because many people are being affected and because the radiation is reasonably easy to measure or to estimate, making statistics linking common health problems with the level of exposition is quite doable.

Hypersensitive persons would be 1% of the population. The need to protect this minority is enough to stop those powerful broadcasts from reaching the ground. I don't understand why I have to add that everybody is being harmed, be it to a lesser degree. I met official people whose job is to help hypersensitive persons. They've seen their pain and anguish by themselves... but they feel clueless as to how to stop the emissions. It's a very strange situation, when even the people close to the steer and convinced of a problem, can't do anything to stop it.

It has been claimed that the hypersensitivity of some persons is caused by an accumulation of toxic metals or organic pollutants. In such cases, antioxidants can be of immediate help and detoxifications would greatly help. (Note that common medical analysis like looking for mercury and the like in samples of blood, urine or hair, will reveal nothing even under severe intoxication, because blood and hair renew constantly. Only a sudden intoxication can be revealed by such analysis. The toxics slowly accumulated in standing body cells like the brain... Best method seem to be to take a well-tuned quantity of chelating molecules and then perform an analysis of the blood or urine.)

Cell phone and Wi-Fi radiations are different, for several reasons: they don't spread the same way in the body, they are pulsated and their frequency make they will not target the same macromolecules in the body. It would seem that you need 3 times more force of FM broadcast waves than cell phone or Wi-Fi waves, to have a same global impact...

I was frightened when I understood that under those 2 V/m force of radiation, an electric current in the order of 1 mA oscillates through the body. Such a current at a high frequency of 100 MHz will interact with the body quite differently than a DC or low frequency AC current. You cannot make direct comparisons... Anyway, 100 mA of DC current can be enough to kill a person by electrocution. 1 mA of low frequency current is enough to feel the current as a tickle. What are the consequences of such a current on the chemical reactions in the liver, on the production of

Powered by 000webhost

hormones by glands everywhere in the body? What if the person being exposed is a pregnant woman, with the baby just at the worst place in the flux of current? The length of an "antenna" matters much to allow it to build up current. Teenagers have the "optimal" length for 100 MHz radio waves... One thing I'm sure about is that if the government decided that from now on the school kids will be latched electrodes on the body and a current of 1 mA send through several hours a day, whatever the frequency, the parents will not accept. But this *is* happening, through radio waves, and I could feel the neuroleptic-like effect of it. (By the way, this implies that those experiments that show brains of rats destroyed by cell phones waves within the norms, apply only to a lesser degree to human brains, because of the difference in size. Rats are close to the optimal length to be harmed by cell phone waves... The optimal wavelength to dammage teenagers is that of those 100 MHz public broadcasts...)

The radiations harm me more inside buildings than outside. I may get more sick under 0.2 V/m in a room than under 2 V/m directly exposed to an antenna. I don't know exactly why. Here are four possibilities:

- The roofs atop buildings act like prisms and send the waves towards the inhabitants beneath. Radio waves coming from above would be more harmful than when coming from aside.
- The radio waves resonate and echo inside the building and build up in metallic structures. This creates halos of proximity that will inject more power inside nearby living bodies. For example, while simply wrapping aluminum foil around my head, it is obvious that some designs of shielding will make me even more sick than using no aluminum at all. The same way, when some parts of the shielding of my tent wear off and stop to be conducting, I can get serious problems till I replace or complement the faulty area.
- Buildings contain their own sources of electromagnetic pollution. They add their effects to those of the pollution from the outside...
- Close to metallic structures, the electromagnetic field can become rotating. This because waves with equivalent strength come from different directions with different phase shifts.

Again, the four possibilities mentioned above are potentially very suitable to build experiments that demonstrate there is no problem. Perform the experiment in an anechoic place with the waves coming from aside and no other radiation than the one from the experiment... Just the opposite from real world situations but you can claim to be serious at avoiding parasitic variables...

If you want to find the FM emitters somewhere, fmscan.org is the reference I used till now.

This is the schematic of the "snake":

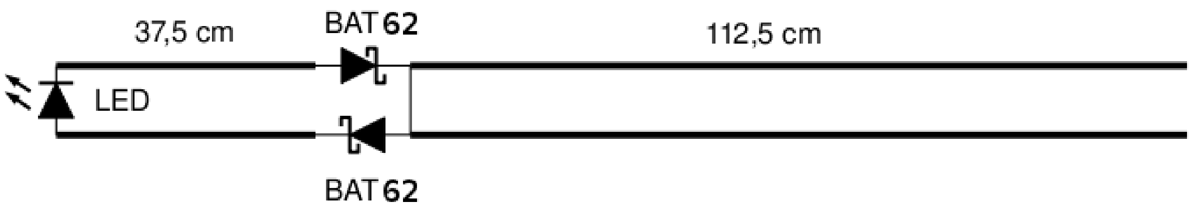

BAT62 detection diodes are no more produced. **BAT15** diodes work fine but they wear out; after a few months they stop functioning. **SMS7630** diodes are great but very little and mechanically fragile. **MMSD701T1G** diodes are sturdy and powerful; an excellent choice for a beginner. Such SMD diodes do also work for cell phone frequencies, which allows to test out a snake with a calling cell phone pushed against it. But any detection diodes that can manage 100 MHz will do.

The LED I'm currently using is the **L-7113SEC-H** . It lights up with a low tension and a very low current (the bluer a LED, the more tension it needs). Its color is red yet close to orange hence it is easily seen by the human eye (the eye is most sensitive to green, yellow and orange). The beam is quite narrow so when the LED is directed towards somebody's eyes it will appear quite bright.

For the lengths of 37.5 and 112.5 centimeters, any electric wire with two copper conductors will do. Audio signal wire is a practical solution. Use the shielding as one of the two conductors. The lengths of the two segments must not be precise. What matters is that the total length of the snake be 1.5 meters. Do not hesitate to try out if a little

Powered by 000webhost

longer or shorter snake gives better results.

A schematic of my current probes, that I connect to a standard multimeter, measuring Volts DC. The measure displayed by the multimeter must be multiplied by 10. When using a 200.0 mV scale, just read while forgetting the dot:

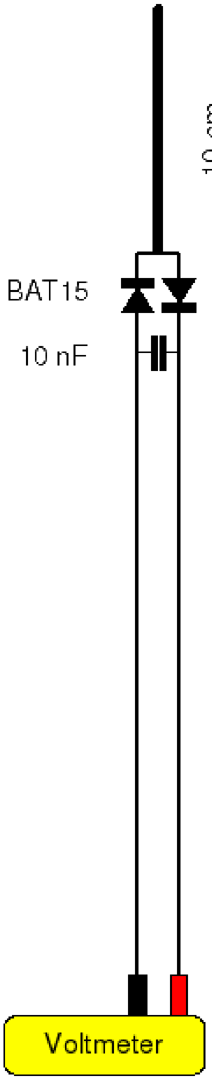

This is a picture of a device that tries to pack the whole in a neat gadget. The electronic circuit is a €7 digital Volts display with a scale of 200.0 mV. I soldered away the jumper that makes the dot be displayed. The antenna is only 7 cm long, so an RMS value is displayed (for roughly steady radio waves). The whole cost about €12. Hold the antenna perpendicular to the direction the waves come from and try to hold the device away from conducting objects like your body or metallic objects. You cannot closely trust such simple device yet it is very handy and adequate to compare the radiation level in different places:

Powered by 000webhost

Powered by 000webhost

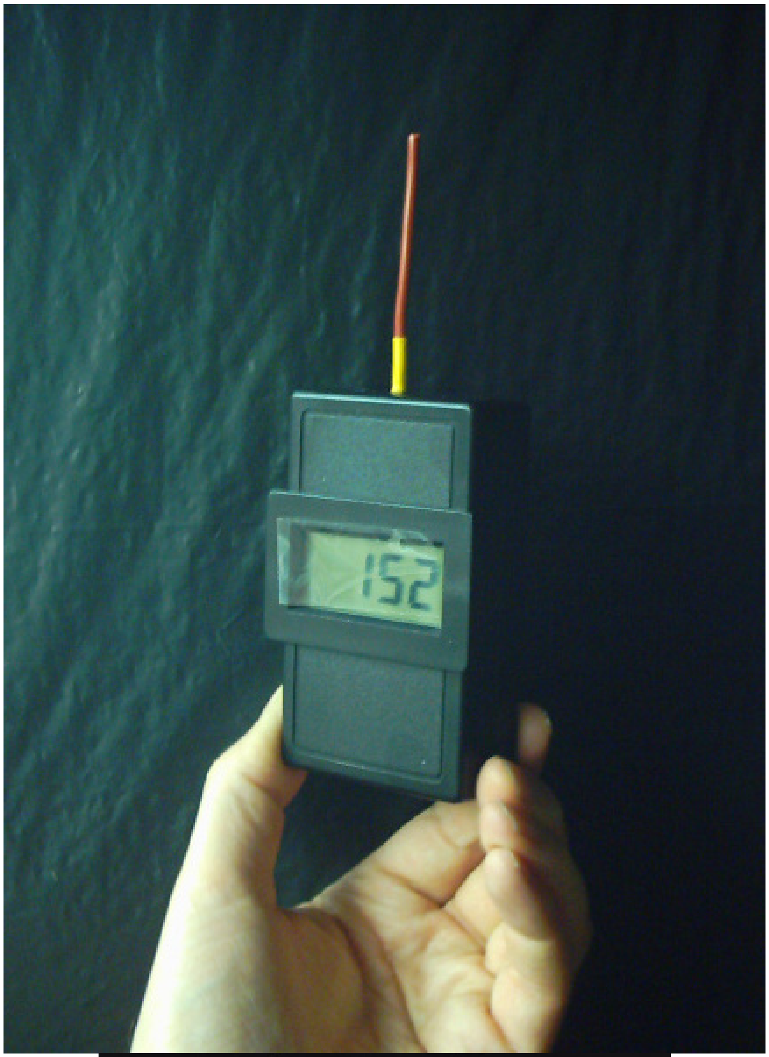

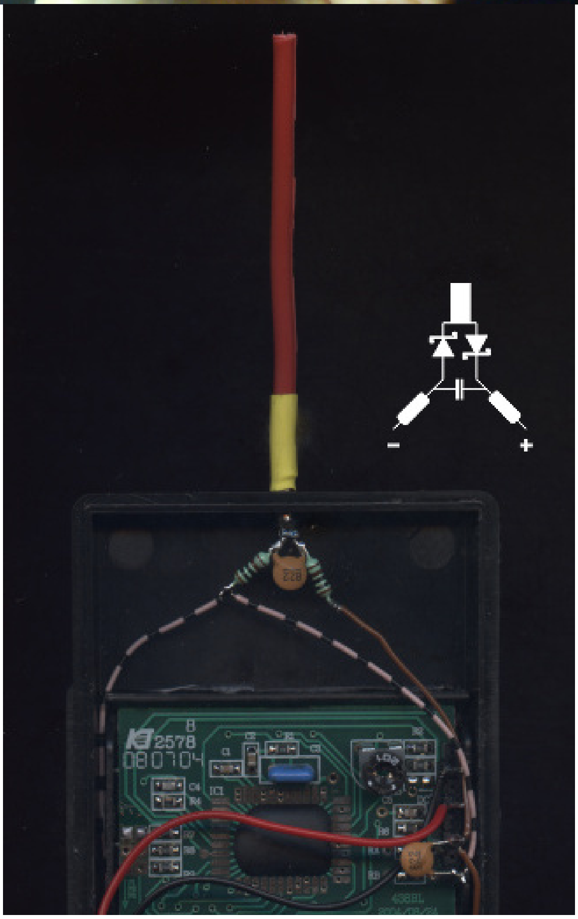

Powered by 000webhost

Those probes and device, quite easily allow to know the direction some 100 MHz radiation come from. Hold the thing at some decimeters from your body, roughly at the height of the belly. Your body and the device form a directional antenna... When the device is towards the emitter, it will display up to twice the value, while when it is hidden from the emitter by your body, the measure will be very low.

If you measure something strong, yet by turning in different directions the device always shows approximately the same value, then you are probably undergoing low-frequency radio waves like short wave broadcasts.

If the emitter is close, like a power saving lamp or a wrongly build power supply, converge towards it simply using the fact that the closer you get, the stronger the measure.

Such measuring tools are handy to find out power chords and appliances that pollute, yet then the measure displayed doesn't mean much. Once you are really close to an emitter, the measuring device and the emitter interact and the power flowing through the measuring device can be tremendous even though the device is only weakly polluting. You are in the "halo of proximity" of the device... Best example is most quality laptops, which emit almost no radio waves. I never could notice a problem while using my laptop when it is plugged to nothing... Yet if I hold a measuring tool against the screen I get frightening data... A few decimeters further, where my head can be, the measuring tool tells there is nothing... So, the high figures close to the laptop indeed tell that strong currents flow through it, but competent people made things such that those currents harm nobody...

If you want a global measure, away from the effects of your body, hold the device above your head. Or place it on some plastic or cardboard box and walk some distance away.

The windows are a common entry point of radio waves and electric wires can be awfull ducts, yet in some cases I noticed that metallic structures unrelated with electricity, like the copper tubes of water heating radiators, were the main ducts of radio pollution towards a room.

Below is a device I assembled to focus on 1.8 GHz cell phone radio waves. Close to the nearby cell phone relay antenna, in direct view of it, I could not measure much more than 0.1 V/m (pike value). While this is still too much, it is by no means comparable to the frightening 2 V/m intensity of the FM broadcasts. I also went measuring in places where people complain about problems similar to mine. They thought the reason was the nearby cell phone relay antennas. But I could measure no significant intensity of cell phone radio waves at all, while the force of the FM broadcasts was above 0.2 V/m...

Powered by 000webhost

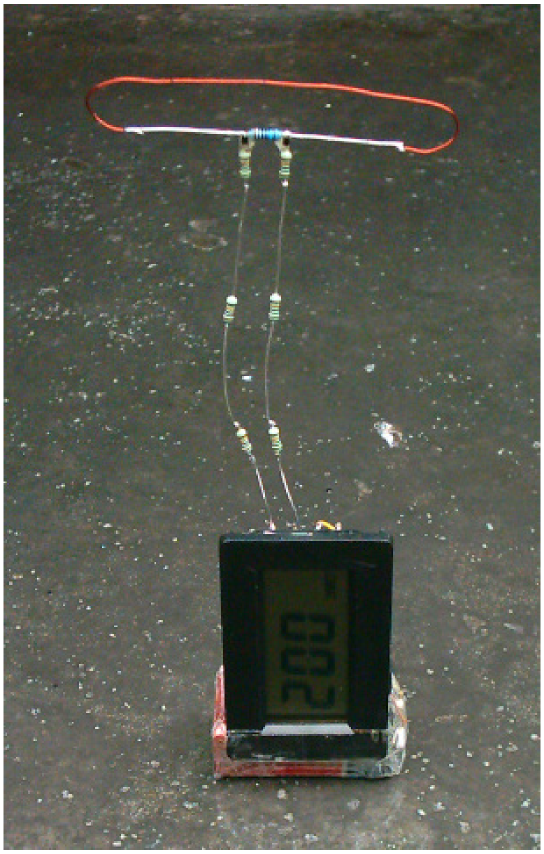

After about a year experimenting, the most simple protection I could come up with against 100 MHz broadcasts is a headband of aluminum, much the shape and size of a sweatband. I hide it inside a bonnet. I have no idea as to how and why it helps, aside from the general fact that any conducting structure will change something to the propagation of high frequency currents. Don't make it too high, a 3 to 4 centimeters height is OK. To try this out, you can just fold in a circle a 1 meter long wrap of aluminum and compress it to a flattened ring that fits your head. If you intent to re-use it, then you must strengthen it with lengths of tape before (inside) and after (outside) the ring is constituted. Make sure the tape does not hamper a perfect electric contact all around the ring. This protection is not perfect but it really helps.
When the field strength is above 0.1 V/m, I need a complementary vertical band in order to be able to work.
This also is a cheap way to make statistics. Compare days with the headband to days without the headband. In order to be scientific, you must ask somebody else to put the headband inside your bonnet every morning, at random. Every morning, the other person writes down if there was a conducting headband inside the bonnet or not. Every evening, you write a comment down to how your day was (headache? brain fog? errors? doomed day?) Then after a while you compare the two. At worst, if you have to do it alone, buy two identical bonnets, push a plastic headband inside one and an aluminum headband inside the other, and spend each day without knowing which one you picked. At the end of the day, write your appreciation down before checking and writing down which bonnet you picked.

Powered by 000webhost

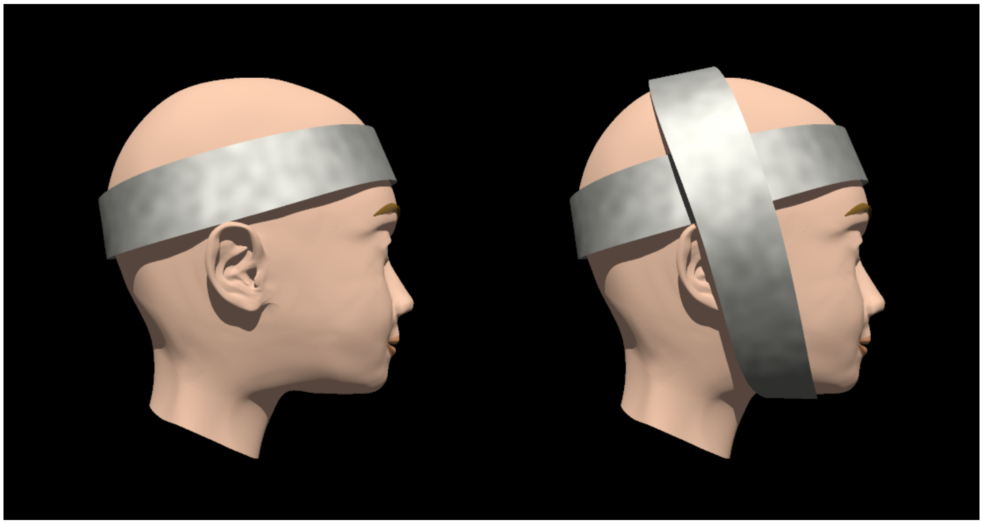

A friend who is a physician points out that in many administrations, be they public or private, the employees are being persecuted. (That often happens when the work done by the administration is mostly useless and only serves to justify the social position and the salaries of a few people.) A statistic about the health impact of radio waves in such administrations could be of few significance, because the people have all kinds of health problems due to the harassment. The addition to this by the radio waves may be comparatively negligible. (Yet another way to build a statistic that shows few or no impact of the radio waves: perform it in a place where people bow under humilations...)

I was quite frightened to discover that the symptoms caused by radio waves, as I see them around me, are very close to those of "radiation poisoning". This is part of nuclear medicine. When a victim is exposed to a short and strong burst of radioactivity or any other ionizing radiation, its body will quickly react like if it had been poisoned. A "low" poisoning by radiations yields headaches, tiredness, difficulties to think... while stronger poisoning can lead to the person die in a few days like if it was heavily burn. The exposure to ionizing radiations is measured in "Gray". 1 Gy leads to a low poisoning. So I computed out how much energy a person gets from the local FM emitter, like the person I know that get sick at their desk. The result is 0.01 Gy for 8 hours of exposition, which is quite close.

Radio waves can induce all the typical problems caused by that ionisinig radiations but the energy levels involved can be quite different. For example, ionizing radiations cause burns without increasing the temperature of the body, while radio waves can only burn if they are powerful enough to heat the body. Radio waves can induce cancer but again the energy required is ways higher, like spending ten years using a cell phone against the head, several hours a day.

Body cells can be attacked by all kinds of means. Ionizing radiations lead to direct damage of the cell membrane. Radio waves disturb the molecular cell gates operate. The cell will always react the same way, as for poisoning: it will close the cell gates, to try to shun itself from the outside world and from the harm. This has many consequences. The cell is less able to fulfill its purpose inside the body. It uses resources to try to protect its internal parts. It will accumulate waste. That's why people get tired, less efficient or even sick. When the aggression made to the cell is deemed too strong, the cell will auto-destruct. That's the way radiations and chemicals can be used against cancer, making the cancer cell suicide.

The cells can be helped to better recover from the state of shock. Also they can be made to go less likely into the state of shock. Medication and health advice can be helpful for this.

A link has been proposed between autism and radio pollution. The growth of the human brain is very complex, with cells traveling from one side of the brain towards the other to find their place. This involves a tremendous amount of communication and coordination between the cells. What if those cells are too often put into a state of shock?

These are two scientific publications about the link between autism and radio waves :

http://www.ncbi.nlm.nih.gov/pubmed/16530334

http://www.ncbi.nlm.nih.gov/pubmed/14962625

Powered by 000webhost

Publications about how radio waves disturb cells :

http://www.ncbi.nlm.nih.gov/pubmed/10860806

http://www.ncbi.nlm.nih.gov/pubmed/12379225

Links

http://www.teslabel.be
http://www.clag.be
http://www.001.be.cx
http://www.criirem.org
http://www.next-up.org
http://www.robindestoits.org
http://www.beperkdestraling.org
http://www.stopumts.nl

A petition: www.petition-electrosmog.be

Eric Brasseur - October 3 2009 till October 16 2015

Powered by 000webhost

A Novel RF Energy Harvesting Module Integrated on a Single Substrate

Monika Mathur[1], Ankit Agarawal[2], Ghanshyam Singh[3], S. K. Bhatnagar[4]

Department of Electronics and Communication
[1,2,4] Swami Keshvanand Institute of Technology , Management and Gramothan, Jaipur, Rajasthan, INDIA.
[1,3]Malaviya National Institute of Technology Jaipur, Rajasthan, INDIA
Email id: [1]monikamathur16@ gmail.com, [2]ankitsagarwal@ gmail.com, [3]gschoudhary75@ gmail.com, [4]bhatnagar_skb@yahoo.de

ABSTRACT

This paper presents the RF energy harvesting module (RECTENNA). The working range of this module includes multiple bands i.e. GSM, ISM, WLAN, and UWB band. To enhance the capturing RF power capability an array arrangement of coplanar monopole antenna has been proposed. Wilkinson power combiner has also been implemented to combine the powers of this antenna array. The RF-DC converter circuit having seven stages has also been integrated with this structure. This module produces the DC voltage of 1.8 V with respect to + 40 dB RF input. It is the unique module because it has no need of port connectors. The impedance matching of antenna and converter has been fulfilled by incorporating the passive component at the combiner's branch. The value of this passive component is kept equal to the existing value of impedance at input port of converter circuit.

Key words: *Rectenna, coplanar monopole antenna, Wilkinson power combiner, antenna array, impedance matching, RF-DC converter circuit.*

I. INTRODUCTION

The primary focus of the Rectenna circuit is to design a suitable RF receiving antenna. For this purpose, an antenna should be designed to resonate on single, dual and multiband frequencies according to the need of the applications. The main aim of proposed work is to design a RF energy harvesting module for a wide range of frequencies which covers most of functional bands, e.g. GSM, Radio, ISM and UWB. It is well known that the RF spectrum widely employed for 900 MHz – 2 MHz (television and radio applications, GSM), 2.1GHz - 2.6 GHz (ISM band for various applications), and 3.1GHz -10.6 GHz (ultra wideband for satellite applications) [1]. Narrow band systems such as WLAN (3.1GHz - 4.4GHz), HIPERLAN (5.1GHz - 5.3GHz), C-BAND (4.4GHz - 5GHz) are also the useful bands. Keeping this in mind a compact coplanar monopole antenna has been designed and its 2 X 2 arrangement was reported [2]. For preparing the array arrangements of the antenna the literature [3-5] were profoundly studied.

This antenna works for the wide range of frequencies of the above specified bands. As reported in the structure [2] the Wilkinson power combiners were preferred because it accomplishes two essential tasks of the structure. Firstly it combines the powers of two or four antennas serially and increases the input power of RF-DC converter circuit of energy harvesting module. Secondly it considered as the element of afford, for the impedance matching of antenna port and converter circuit. The maximum power transfer concept has been chosen for this impedance matching.

The multi band operation of the RF energy harvesting module developed on a single substrate is the innovation of this paper.

II. ANTENNA DESIGN

In RF energy harvesting module antenna is the major functionary part. It receives the power from the electromagnetic spectrum. The received RF power is converted into the DC output voltage from the module. So the antenna should be capable enough to receive the maximum energy/power from the spectrum. As the working range of our design is preferred from 900 MHz -10 GHz, the modification in the monopole antenna structure for multi frequencies resonance has investigated. For increasing the area of receiving as well as gain of the structure the 2 x 2 array of the coplanar monopole antenna has been designed and presented [2]. In this paper [2] only the simulated results were presented. The simulated antenna structure was as shown in figure 1[2]. The power of the individual structure has been combined by the Wilkinson power combiner. The reason for selecting this combiner is that we may externally adjoin the surface mount components at the combining branch of it for impedance matching purpose.

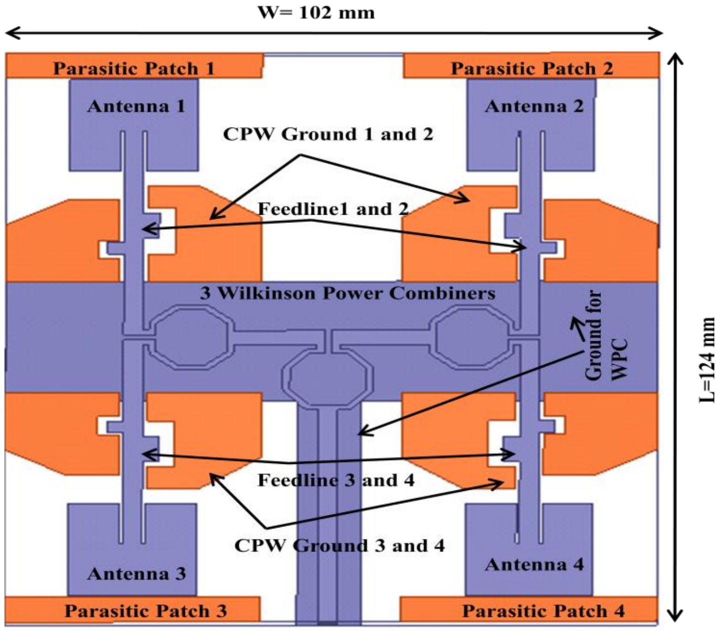
W= 102 mm
Parasitic Patch 1
Parasitic Patch 2
Antenna 1
Antenna 2
CPW Ground 1 and 2
Feedline1 and 2
3 Wilkinson Power Combiners
Ground for WPC
L=124 mm
Feedline 3 and 4
CPW Ground 3 and 4
Antenna 3
Antenna 4
Parasitic Patch 3
Parasitic Patch 4

Figure 1. Simulated structure of 2 x 2 array arrangement of coplanar monopole antenna [2].

The simulated structure from licensed software HFSS version16 has been exposed resonant on the multi frequencies of different allocated bands as 900 MHz and 1.75 GHz (GSM band) 2.7 GHz (ISM), 3.5 GHz, 4 GHz and 5.2 GHz (WLAN), 6.5 GHz, 7.4 GHz, 8.5 GHz and 10 GHz (UWB).

Further in this paper all measured results have been reported for the validation of proposed antenna structure. For the measurement purpose proposed structure has been fabricated as shown in figure 2. The measured result of return loss from VNA (Keysight Technologies) for fabricated structure is shown in figure 3. The resonant frequencies of the fabricated structure has exposed on 900 MHz (GSM), 1.29 GHz (ISM), 4.1 GHz and 5.6 GHz (WLAN), 6.8 GHz and 9 GHz (UWB), that is the almost same as simulated result of the structure reported earlier [2].

.

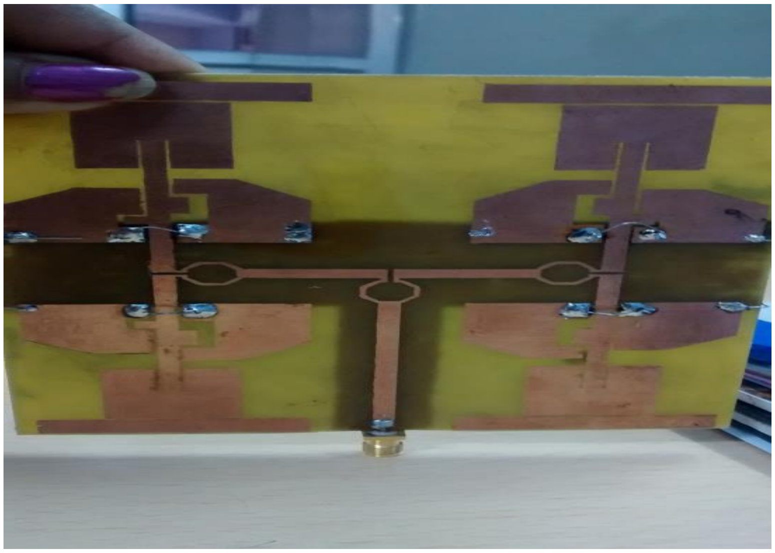

Figure 2. Fabricated structure of the 2 x 2 array of coplanar monopole antenna.

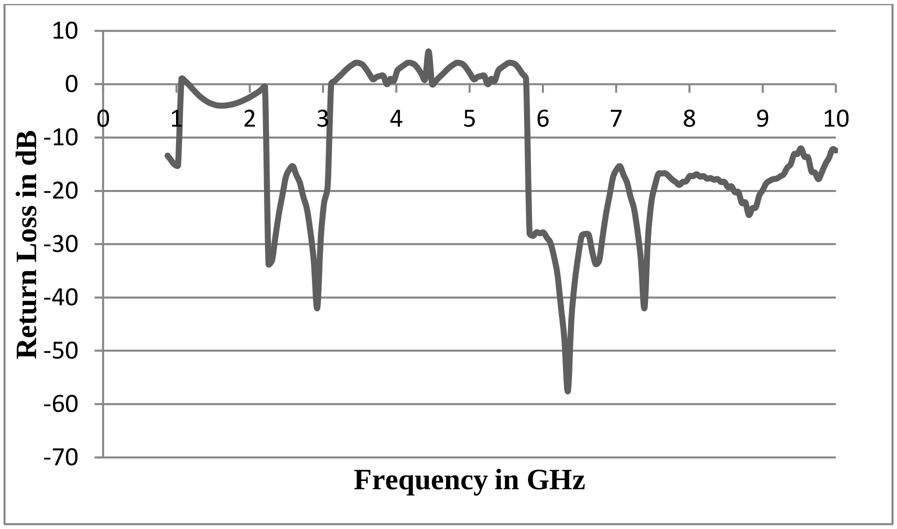

Figure 3. Measured return loss (S_{11}) of the 2 x 2 array of coplanar monopole antenna.

The measured H- field patterns for these frequencies are shown in figure 4(a) and E-field patterns are shown in figure 4 (b). These patterns show that the antenna supposed to be worked as omnidirectional.

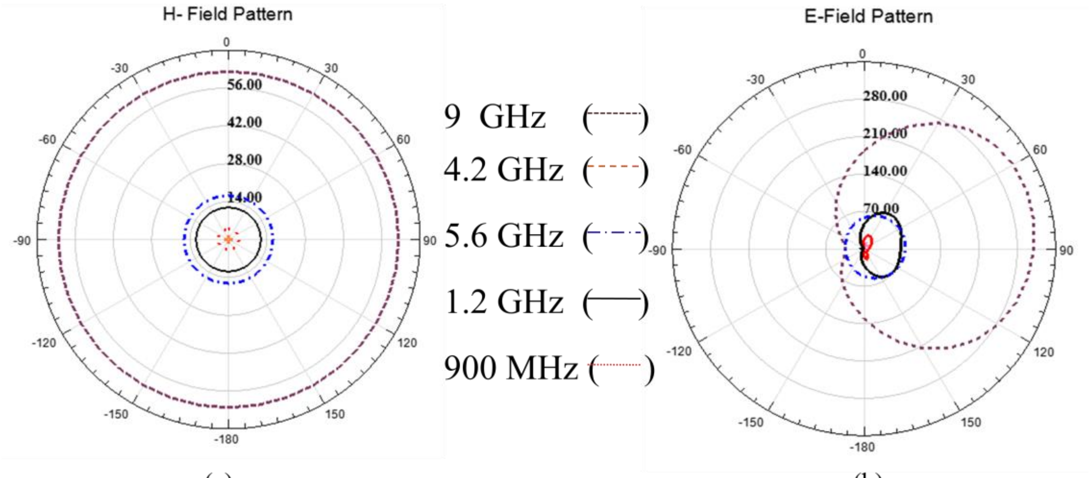

(a) (b)

Figure 4. (a) H- field pattern (b) E- field Pattern

III. RF-DC CONVERTER

The RF-DC converter circuit is also the important part of energy harvesting module. The output of harvester circuit is the DC voltage. The input of this module is the magnitude of AC voltage at the antenna port. This AC voltage is proportional to the captured power on the antenna from the RF spectrum. In this paper for energy conversion purpose a voltage doubler circuit has been used. The number of stages of the converter circuit may be increased for escalating the DC output voltage value. The relation of number of stages (n) and output voltage ($V_{o/p\ volt}$) of converter circuit is considered as

$$V_{o/p\ volt} = \frac{nV_{o/p\ volt}}{nR_0 + R_L} R_L$$

In this paper conversion design contains seven numbers of stages of Villard voltage doubler circuit. The main function of this conversion circuit is to convert the RF harvested energy from the ambient into the direct current (DC) voltage. This seven stage Schottky diode voltage doubler circuit is first designed, modelled and simulated on ANSYS RF circuit designer and then fabricated and tested. After that, this circuit has integrated with the same substrate of antenna. The converter circuit in this design uses zero bias Schottky diode (SMS7621-001-SOT23) from Skyworks. The simulated design of seven stage RF-DC converter circuit has been presented for this purpose [6]. The simulated DC output voltage at the first stage and seventh stage of RF-DC converter circuit from the RF circuit simulator was as 48.2 mV and 289.6 mV respectively. The fabricated circuit of seven stage RF-DC converter and its analysis has already been presented [6]. So in this paper we implemented that designed RF-DC circuit (having seven stages) [6] with the same single substrate of designed antenna array.

IV. RF ENERGY HARVESTING MODULE

This part of paper presented a RF energy harvesting module on a single substrate working on multi bands of the frequency range 900 MHz to 10 GHz. The DC output from the multiband harvesting circuit has been presented it produces 3.4 V DC output [8]. The motive of this part is to design whole module on a single substrate for a wide range of operation and should be able to produce the DC output voltage of the value more than 1.5 V from the ambient RF spectrum. This DC voltage may be boosted by adding the booster circuit. In addition to the structure on a single substrate of RF module, it is also proposed that any value of impedance measured at the converter circuit input impedance may be matched to antenna port by using SMD (Surface mount device) resistor components. This is possible only due to the use of Wilkinson power combiner implemented at the antenna structure. No need of matching of connectors impedances. So it may be called connector-less module. This is the uniqueness of the structure.

The novel idea of impedance matching is that; for any wilkinson power combiner a given number of output ports N, an input impedance R_S and output load impedance R_L, and the characteristic impedance of each quarter-wave section, $Z_{\lambda/4}$, is given by

$$Z_{\lambda/4} = \sqrt{N\, R_L R_S}$$

Additionally

$$R = R_L$$

the resistor (R) that connect between the common junction and the output ports are simply equal to the load impedances (R_L) at each output. Now considering the input impedance of RF-DC converter circuit as the value of load resistance (R_L) of the output of the Wilkinson power combiner. This impedance has been measured by vector network analyzer. The SMD resistor of this measured value (R) has been mounted in between the combining leg of Wilkinson power combiner. In this way we can easily match the impedance of antenna and converter circuit. This satisfies the condition of maximum power transfer.

The fabricated RF energy module on a single substrate with impedance matching arrangement is shown in figure 5. The seven stage RF-DC converter has been implemented by the combination of seven Schottky diodes and seven capacitors. The load resistance and source resistance has also been added into the circuit. The overall size of the module is 124 x 102 mm^2.

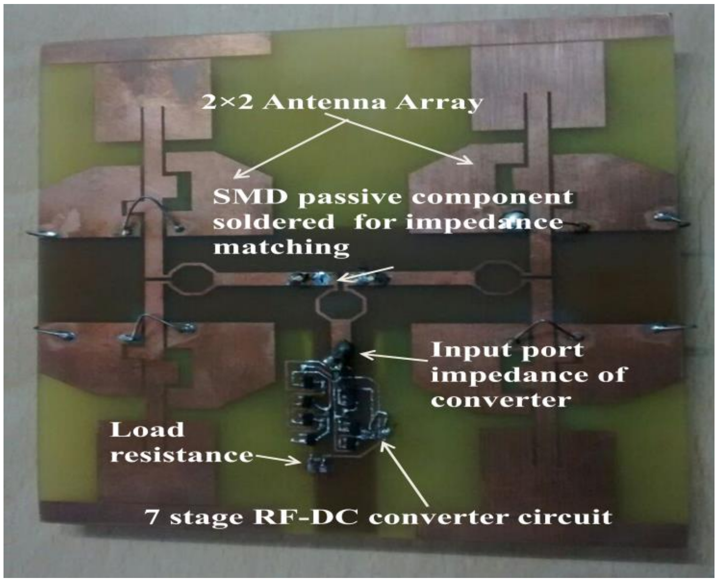
2×2 Antenna Array
SMD passive component soldered for impedance matching
Input port impedance of converter
Load resistance
7 stage RF-DC converter circuit

Figure 5. Fabricated RF energy harvesting module integrated on a single substrate.

The measurement with module in RF lab has been shown in figure 6. The measurement has taken at ICRS (International center for Radio Science) lab, Jodhpur. Rajasthan, INDIA. The RF source is of 40 GHz. First of all with respect to the source on the resonant frequencies of the module; power received at the antenna port has been measured on the power-meter.

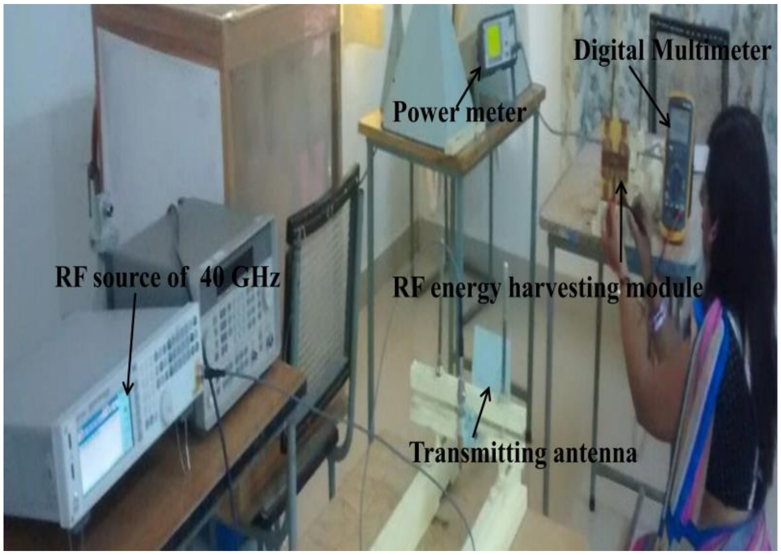
Digital Multimeter
Power meter
RF source of 40 GHz
RF energy harvesting module
Transmitting antenna

Figure 6. Measurement of output DC voltage from the module with respect to RF input voltage from RF source.

The DC output voltage at the module output with respect to RF input power varying from - 40 dB m to + 40 dB m has been plotted for five resonant frequencies (900 MHz, 1.2 GHz, 4.1 GHz, 5.6 GHz and 9 GHz). The plot of the data is as shown in figure 7. The maximum DC output is 1.823 V.

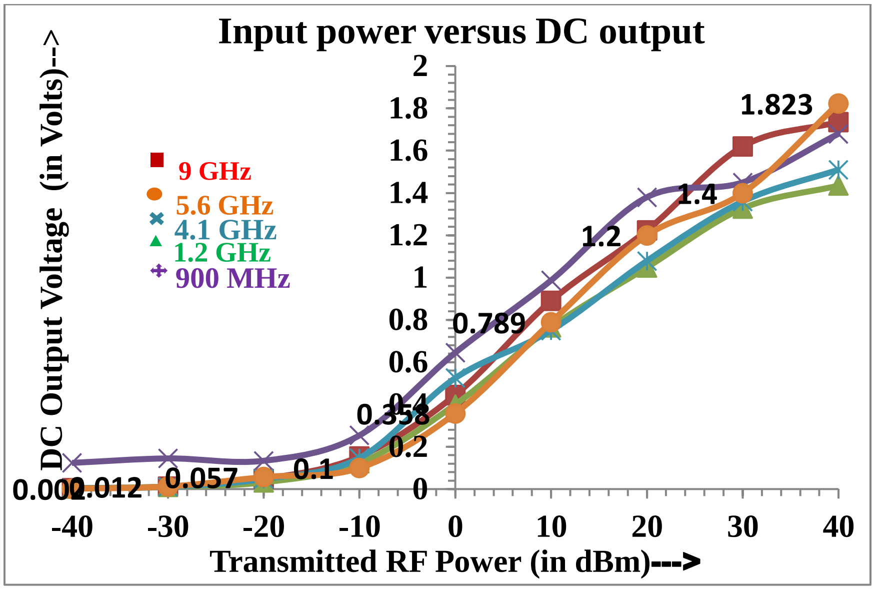

Figure 7. The transmitted RF input power versus DC output at the module graph.

The conversion efficiency (in %)of the module is calculated by the formula

$$\eta = \frac{V_{DC\ output}}{V_{input}} \times 100$$

The $V_{DC\ output}$ is the output measured by the digital multimeter and V_{input} is the magnitude of the input voltage at RF power source. The efficiency for the 9 GHz frequency is shown in the figure 8. As shown in figure the efficiency increases as the RF power increases.

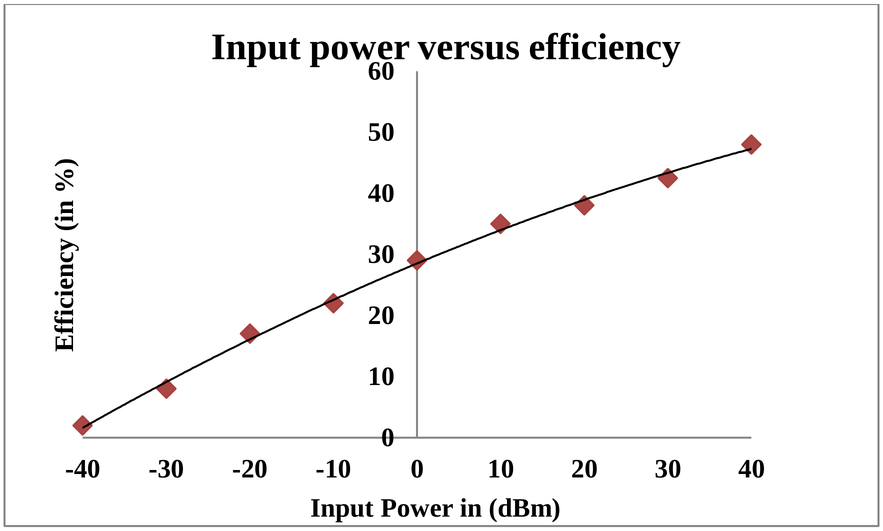

Figure 8. Input RF power versus efficiency graph.

The efficiency increases by the change in load resistance. For the input RF power +10 dBm the change in efficiency with respect to the change in load resistance is shown in figure 9. The maximum DC output voltage (1.823 V) for +40 dBm has been noted on the load resistance 22 k-ohms.

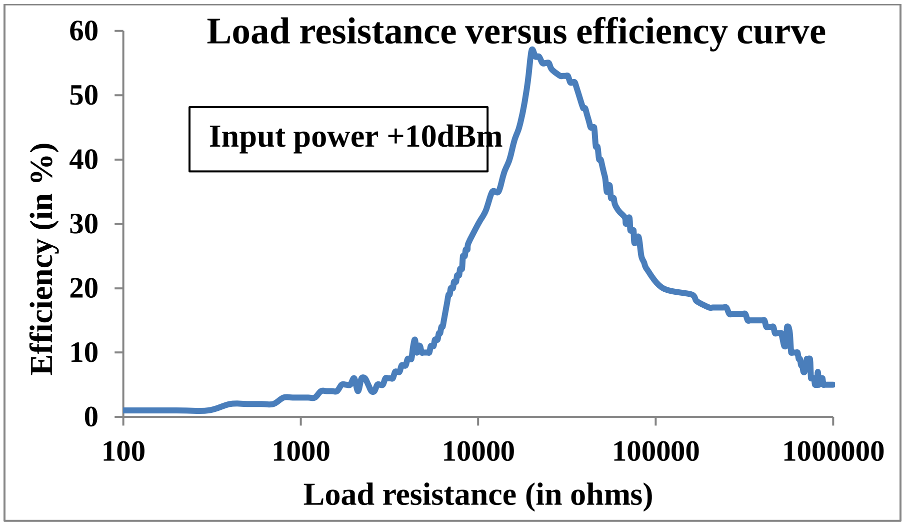

Figure 9. The load resistance versus efficiency curve.

V. CONCLUSION

A novel RF energy harvesting module has been presented in this paper. The antenna for this module has the structure of 2×2 array of the coplanar monopole antenna. The novelty which is proposed lies in the fact that a seven stage RF-DC converter circuit has been integrated on the same substrate of antenna. This makes the module connector-less. Another uniqueness of this module is that the impedance between the antenna and the converter circuit is matched by using the maximum power transfer concept. The Wilkinson power combiners have been specifically selected for this purpose as they combine the powers of antenna arrays. For impedance matching between the antenna output port and RF-DC converter port, the existing input impedance at the input port of the converter circuit is measured. And the component of that measured value has been mounted at the Wilkinson combining branch. This arrangement reported the matching between the antenna and circuit ports. The DC output of the module has been measured. The maximum DC output is 1.823 V with respect to +40 dB m input RF power, with the load resistance 22 k ohm for the frequency 9 GHz. This voltage may be boosted by integrating a booster circuit. The voltage may be increased by using the super capacitor at the output of module.

VI. ACKNOWLEGEMENT

The authors are thankful to our Institutes Swami Keshvanand Institute of Technology, Management and Gramothan, Jaipur and Malaviya National Institute of Technology Jaipur, Rajasthan, India for providing the licensed software and fabrication facilities. The authors are also thankful to Dr. O. P. N Calla, Director- ICRS (International Center for Radio Science) lab, Jodhpur, Rajasthan, India for providing the measurement facilities.

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

REFERENCES

[1] Federal Communications commission, First report and Order, (Revision of part 15 of commission's rule regarding UWB transmission system FCC, Washington, DC, pp. 02-48, 2002.

[2] Monika Mathur, Ankit Agrawal, Ghanhyam Singh, S. K. Bhatnagar, " The array structure of 2 x 2 coplanar monopole antenna with wilkinson power combiner for RF energy harvesting applications", Proc. of IEEE Xplore 2016 , pp. 1-4, ISBN 978-1-5090-2807-8, 2016.
DOI: 10.1109/ICRAI.2016.7939563

[3] U. Olgun, C. C. Chen, J. L. Volakis, " Invesigation of rectenna array configurations for enhanced RF power harvesting" IEEE Antenna and Wirel. Prop. Letters, vol. 10, pp. 262-265, 2011.

[4] N. hasan and S. K. giri, "Design of low power RF to DC generator for energy harvesting applications", Int. Journal of Applied Scien. and Engg. Research, vol.1, Issue 4, pp. 562-569, 2012.

[5] T. Sakamoto, Y. Ushijima, E. Nishiyama, M. Aikawa and I. Toyoda, " 5.8 GHz series/parallel connected rectenna array using expandable differential rectenna units" , IEEE Trans. on Antenna and Prop. , vol. 61, no. 9, 2013.

[6] Monika Mathur, Ankit Agrawal, Ghanshyam Singh, S. K. Bhatnagar, "A Novel Design Module of RF Energy harvesting for Powering the Low Power Electronics Devices", ACM digital library- Proc. of AICTC , article no. 68, pp 1-4, 2016.
DOI: 10.1145/2979779.2979847.

UNITED STATES PATENT OFFICE.

NIKOLA TESLA, OF NEW YORK, N. Y.

APPARATUS FOR TRANSMITTING ELECTRICAL ENERGY.

1,119,732. Specification of Letters Patent. **Patented Dec. 1, 1914.**

Application filed January 18, 1902, Serial No. 90,245. Renewed May 4, 1907. Serial No. 371,817.

To all whom it may concern:

Be it known that I, Nikola Tesla, a citizen of the United States, residing in the borough of Manhattan, in the city, county, and State of New York, have invented certain new and useful Improvements in Apparatus for Transmitting Electrical Energy, of which the following is a specification, reference being had to the drawing accompanying and forming a part of the same.

In endeavoring to adapt currents or discharges of very high tension to various valuable uses, as the distribution of energy through wires from central plants to distant places of consumption, or the transmission of powerful disturbances to great distances, through the natural or non-artificial media, I have encountered difficulties in confining considerable amounts of electricity to the conductors and preventing its leakage over their supports, or its escape into the ambient air, which always takes place when the electric surface density reaches a certain value.

The intensity of the effect of a transmitting circuit with a free or elevated terminal is proportionate to the quantity of electricity displaced, which is determined by the product of the capacity of the circuit, the pressure, and the frequency of the currents employed. To produce an electrical movement of the required magnitude it is desirable to charge the terminal as highly as possible, for while a great quantity of electricity may also be displaced by a large capacity charged to low pressure, there are disadvantages met with in many cases when the former is made too large. The chief of these are due to the fact that an increase of the capacity entails a lowering of the frequency of the impulses or discharges and a diminution of the energy of vibration. This will be understood when it is borne in mind, that a circuit with a large capacity behaves as a slackspring, whereas one with a small capacity acts like a stiff spring, vibrating more vigorously. Therefore, in order to attain the highest possible frequency, which for certain purposes is advantageous and, apart from that, to develop the greatest energy in such a transmitting circuit, I employ a terminal of relatively small capacity, which I charge to as high a pressure as practicable. To accomplish this result I have found it imperative to so construct the elevated conductor, that its outer surface, on which the electrical charge chiefly accumulates, has itself a large radius of curvature, or is composed of separate elements which, irrespective of their own radius of curvature, are arranged in close proximity to each other and so, that the outside ideal surface enveloping them is of a large radius. Evidently, the smaller the radius of curvature the greater, for a given electric displacement, will be the surface-density and, consequently, the lower the limiting pressure to which the terminal may be charged without electricity escaping into the air. Such a terminal I secure to an insulating support entering more or less into its interior, and I likewise connect the circuit to it inside or, generally, at points where the electric density is small. This plan of constructing and supporting a highly charged conductor I have found to be of great practical importance, and it may be usefully applied in many ways.

Referring to the accompanying drawing, the figure is a view in elevation and part section of an improved free terminal and circuit of large surface with supporting structure and generating apparatus.

The terminal D consists of a suitably shaped metallic frame, in this case a ring of nearly circular cross section, which is covered with half spherical metal plates P P, thus constituting a very large conducting surface, smooth on all places where the electric charge principally accumulates. The frame is carried by a strong platform expressly provided for safety appliances, instruments of observation, etc., which in turn rests on insulating supports F F. These should penetrate far into the hollow space formed by the terminal, and if the electric density at the points where they are bolted to the frame is still considerable, they may be specially protected by conducting hoods as H.

A part of the improvements which form the subject of this specification, the transmitting circuit, in its general features, is identical with that described and claimed in my original Patents Nos. 645,576 and 649,621. The circuit comprises a coil A which is in close inductive relation with a primary C, and one end of which is connected to a ground-plate E, while its other end is led through a separate self-induction coil B and a metallic cylinder B′ to the terminal D.

The connection to the latter should always be made at, or near the center, in order to secure a symmetrical distribution of the current, as otherwise, when the frequency is very high and the flow of large volume, the performance of the apparatus might be impaired. The primary C may be excited in any desired manner, from a suitable source of currents G, which may be an alternator or condenser, the important requirement being that the resonant condition is established, that is to say, that the terminal D is charged to the maximum pressure developed in the circuit, as I have specified in my original patents before referred to. The adjustments should be made with particular care when the transmitter is one of great power, not only on account of economy, but also in order to avoid danger. I have shown that it is practicable to produce in a resonating circuit as E A B B′ D immense electrical activities, measured by tens and even hundreds of thousands of horse-power, and in such a case, if the points of maximum pressure should be shifted below the terminal D, along coil B, a ball of fire might break out and destroy the support F or anything else in the way. For the better appreciation of the nature of this danger it should be stated, that the destructive action may take place with inconceivable violence. This will cease to be surprising when it is borne in mind, that the entire energy accumulated in the excited circuit, instead of requiring, as under normal working conditions, one quarter of the period or more for its transformation from static to kinetic form, may spend itself in an incomparably smaller interval of time, at a rate of many millions of horse power. The accident is apt to occur when, the transmitting circuit being strongly excited, the impressed oscillations upon it are caused, in any manner more or less sudden, to be more rapid than the free oscillations. It is therefore advisable to begin the adjustments with feeble and somewhat slower impressed oscillations, strengthening and quickening them gradually, until the apparatus has been brought under perfect control. To increase the safety, I provide on a convenient place, preferably on terminal D, one or more elements or plates either of somewhat smaller radius of curvature or protruding more or less beyond the others (in which case they may be of larger radius of curvature) so that, should the pressure rise to a value, beyond which it is not desired to go, the powerful discharge may dart out there and lose itself harmlessly in the air. Such a plate, performing a function similar to that of a safety valve on a high pressure reservoir, is indicated at V.

Still further extending the principles underlying my invention, special reference is made to coil B and conductor B′. The latter is in the form of a cylinder with smooth or polished surface of a radius much larger than that of the half spherical elements P P, and widens out at the bottom into a hood H, which should be slotted to avoid loss by eddy currents and the purpose of which will be clear from the foregoing. The coil B is wound on a frame or drum D¹ of insulating material, with its turns close together. I have discovered that when so wound the effect of the small radius of curvature of the wire itself is overcome and the coil behaves as a conductor of large radius of curvature, corresponding to that of the drum. This feature is of considerable practical importance and is applicable not only in this special instance, but generally. For example, such plates at P P of terminal D, though preferably of large radius of curvature, need not be necessarily so, for provided only that the individual plates or elements of a high potential conductor or terminal are arranged in proximity to each other and with their outer boundaries along an ideal symmetrical enveloping surface of a large radius of curvature, the advantages of the invention will be more or less fully realized. The lower end of the coil B—which, if desired, may be extended up to the terminal D—should be somewhat below the uppermost turn of coil A. This, I find, lessens the tendency of the charge to break out from the wire connecting both and to pass along the support F.

Having described my invention, I claim:

1. As a means for producing great electrical activities a resonant circuit having its outer conducting boundaries, which are charged to a high potential, arranged in surfaces of large radii of curvature so as to prevent leakage of the oscillating charge, substantially as set forth.

2. In apparatus for the transmission of electrical energy a circuit connected to ground and to an elevated terminal and having its outer conducting boundaries, which are subject to high tension, arranged in surfaces of large radii of curvature substantially as, and for the purpose described.

3. In a plant for the transmission of electrical energy without wires, in combination with a primary or exciting circuit a secondary connected to ground and to an elevated terminal and having its outer conducting boundaries, which are charged to a high potential, arranged in surfaces of large radii of curvature for the purpose of preventing leakage and loss of energy, substantially as set forth.

4. As a means for transmitting electrical energy to a distance through the natural media a grounded resonant circuit, comprising a part upon which oscillations are impressed and another for raising the ten-

sion, having its outer conducting boundaries on which a high tension charge accumulates arranged in surfaces of large radii of curvature, substantially as described.

5. The means for producing excessive electric potentials consisting of a primary exciting circuit and a resonant secondary having its outer conducting elements which are subject to high tension arranged in proximity to each other and in surfaces of large radii of curvature so as to prevent leakage of the charge and attendant lowering of potential, substantially as described.

6. A circuit comprising a part upon which oscillations are impressed and another part for raising the tension by resonance, the latter part being supported on places of low electric density and having its outermost conducting boundaries arranged in surfaces of large radii of curvature, as set forth.

7. In apparatus for the transmission of electrical energy without wires a grounded circuit the outer conducting elements of which have a great aggregate area and are arranged in surfaces of large radii of curvature so as to permit the storing of a high charge at a small electric density and prevent loss through leakage, substantially as described.

8. A wireless transmitter comprising in combination a source of oscillations as a condenser, a primary exciting circuit and a secondary grounded and elevated conductor the outer conducting boundaries of which are in proximity to each other and arranged in surfaces of large radii of curvature, substantially as described.

9. In apparatus for the transmission of electrical energy without wires an elevated conductor or antenna having its outer high potential conducting or capacity elements arranged in proximity to each other and in surfaces of large radii of curvature so as to overcome the effect of the small radius of curvature of the individual elements and leakage of the charge, as set forth.

10. A grounded resonant transmitting circuit having its outer conducting boundaries arranged in surfaces of large radii of curvature in combination with an elevated terminal of great surface supported at points of low electric density, substantially as described.

NIKOLA TESLA.

Witnesses:
M. LAMSON DYER,
RICHARD DONOVAN.

N. TESLA.

APPARATUS FOR TRANSMITTING ELECTRICAL ENERGY.

APPLICATION FILED JAN. 18, 1902. RENEWED MAY 4, 1907.

1,119,732.

Patented Dec. 1, 1914.

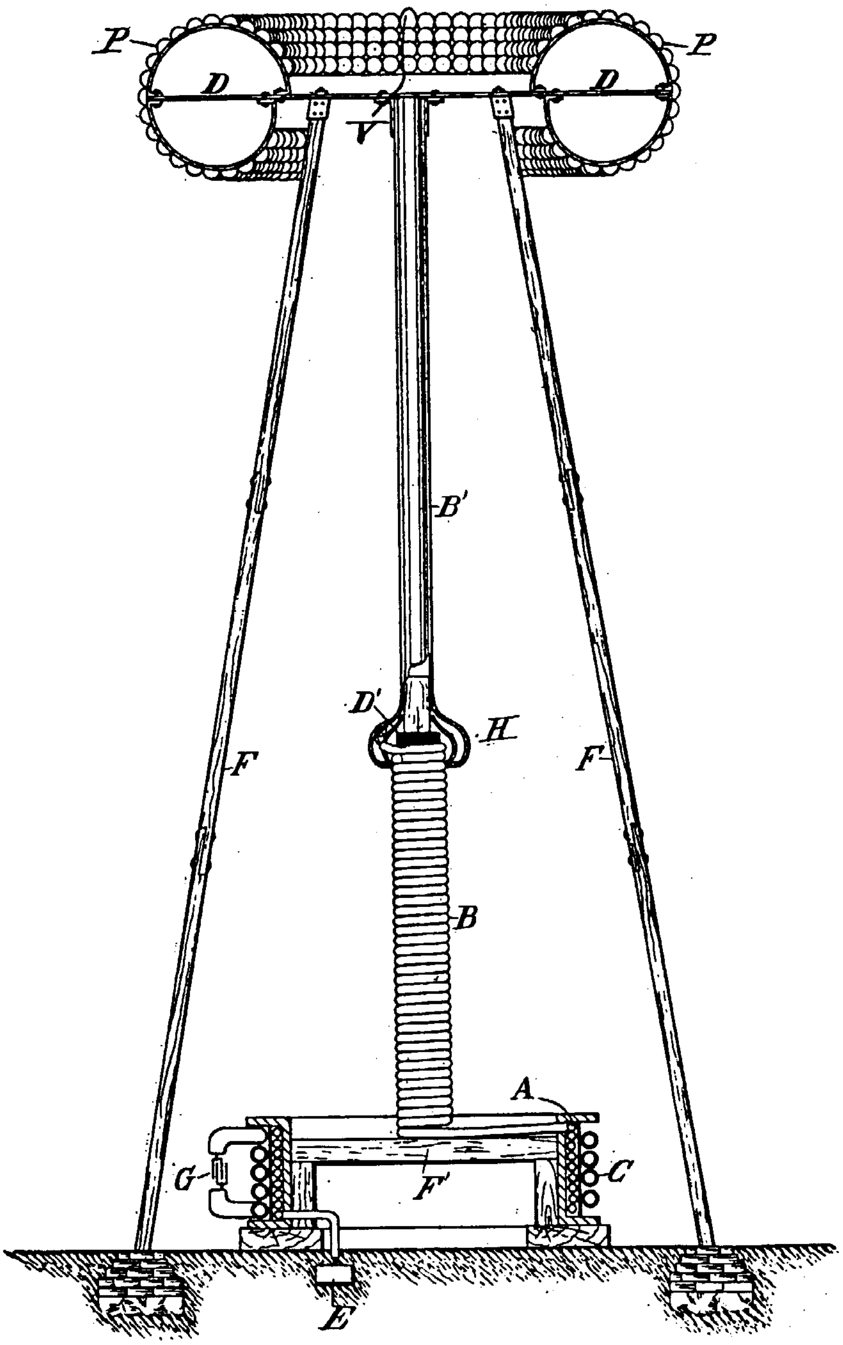

WITNESSES:

M. Lawson Dyer

Benjamin Miller.

Nikola Tesla, INVENTOR,

BY Kerr, Page & Cooper,

his ATTORNEYS.

TPC-8

TESLA AGAINST MARCONI

The Dispute for the Radio Patent Paternity

Paul Brenner, *M.Sc., Senior Member, IEEE, Member WREN, Israeli Representative in the World Renewable Energy Council*

Abstract: **The goal of this paper is to present the multilateral personality of the greatest inventor in history, Nikola Tesla and the claims against Guglielmo Marconi for the radio patent paternity.**

Index Terms: **Tesla (T.), Marconi (M.), coherer, magnifying transformer, LW (long wave), SW (short wave), MW (medium wave).**

I. INTRODUCTION:

A genius is born - Nikola Tesla [1]

Fig. 1.Nikola Tesla at the age of 23. *(Courtesy of Wikipedia)*

Nikola Tesla (Fig.1) saw the daylight for the first time in his life in Smiljan, a small village in Croatia, in the Lika region, on July 10, 1856. His father, Rev. Milutin Tesla was a priest in the Serbian Orthodox Church Metropolitanate of. Sremski Karlovci and his mother, Ana Kalinić, a very talented home craft tools producer, never learned to read, although she was able to recall from memory entire epopees. After the school in Gospić, where he finished a 4 year term in only 3, T. started in 1875 to learn electrical engineering at the Austrian Polytechnic in Graz. Despite the affirmations of some sources, there are no precise indications if he received his Baccalaureate degrees from the university at Graz, since this university declared that he did not receive any title and did not continue his study after the first semester of the third year [1]. In December 1878 he left Graz, breaking also all relationships with his family. Some friends thought that he had drowned in Mura. What is known is that T.was employed for a year as assistant engineer in Maribor, Slovenia, where he suffered a nervous breakdown. Persuaded by his father to continue his studies, he attended the Charles-Ferdinand University in Prague in 1880, where he received the strong influence of Ernst Mach. However, he did not finish his study and left the university after the death of his father, completing a term of only 1 year. But he was genial, even without an academic title. Tesla was an autodidact. He started to read many works, memorizing whole books. Specialists supposed that T. had a photographic memory. In his autobiography he tells that many times he experienced detailed moments of inspiration. Since his childhood, T. was stricked by halucinations accompanied frequently by blinding flashes of light. Much of the effects of this peculiar affliction were related to a word or an idea; the simple hearing of the name of an item was able to induce its detailed envisioning in Tesla's mind. Most of his inventions would have been *apriori visualized* in detail in his mind.(*picture thinking*). This perfect photographic memory was perhaps a hereditary inheritance from his mother, possessing, as said, a natural gift in remembering entire epic poems - who knows?

Hungary and France

After moving to Budapest in 1881 he started to work in Tivadar Puskás's Hungarian National Telephone Company. There, he worked with Nebojša Petrović, a young inventor from Austria, on a project using twin turbines to create continual power. On the opening of the telephone exchange in Budapest, 1881, Tesla became the chief electrician of the company, and later, the engineer for the country's first telephone system. He also deviced a telephone repeater or amplifier, but according to others, the first loudspeaker. In 1882 he moved to Paris, attending the *Continental Edison Company*, where he was charged with the designing of improvements for electric equipments. In the same year, Tesla built the induction motor and improved a range of applications using rotating magnetic fields for which he received patents in 1888. After the death of his mother, in April 1882. he fell ill. Recuperation followed in Gospić and Tomingaj, his mother's birthplace.

United States

Tesla arrived in New York on 06.06.1884 with a recommendation letter from Charles Batchelor, his former manager, to Thomas Alva Edison, in which was stipulated textually: "*I know two great men: you are one of them; the other is this young man*". Tesla got the task to redisign Edison company's DC motors

and generators. He claimed later that he was offered US$50,000 (~ US$1.1 million in 2007, adjusted for inflation) for this task. Tesla gave the Edison company several efficient new patents, but as he asked in 1885 about the payment, Edison replied: "Tesla, you don't understand our American humor" and refused to pay. T. resigned after a new refuse of Edison to raise his weekly salary from 18 USD to 25. After a period of digging ditches for the Edison company, T. formed his own company, *Tesla Electric Light & Manufacturing*. The initial investors disagreed with T's plan for alternating current and he was forced again to work as a common worker a year to feed himself and to subventionate his next project. In 1887, he invented the brushless alternating current motor, presented to the AIEE (now IEEE) in 1888. In the same year, he started to develop his Tesla coil and began working with George Westinghouse at Westinghouse Electric & Manufacturing Company. In April of 1887, Tesla began investigating the X-rays, with his own single node vacuum tubes, a device without target electrode. Tesla was aware about the damages of these rays on the skin, but considered that the ozone and the nitrous acide are merely responsible and not the radiation itself. He made photos of his hand bones before Roentgen and sent them later to Roentgen, without commenting his discovery. T. was sure that these waves are longitudinal waves, produced in force-free magnetic fields, like in plasma, a fact confirmed also by others.
. This concept was demonstrated in 1891 (see fig.2). Unfortunately, much of these early researches of T. were lost in the 5th Avenue laboratory fire in March 1895. T.proposed later, in 1904, a "world system" in which energy could be transmitted wireless through various media, starting from a high-power ultraviolet beam forming a vertical ionized channel directly above the transmitter-receiver stations. The same concept was applied in further inventions of T., like the virtual lighting rods, the electrolaser electroshock weapon, being proposed for disabling war vehicles.

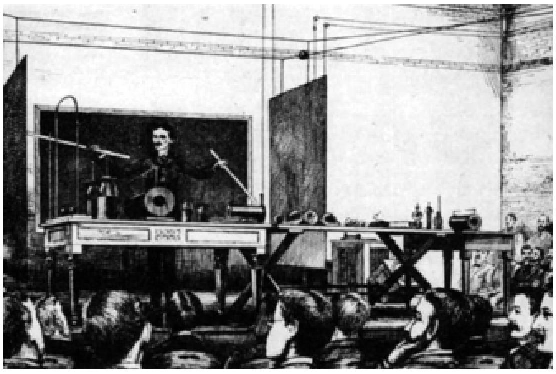

Fig.2. Tesla demonstrating his theory about the wireless transmission of energy waves in 1891[*Courtesy of Wikipedia*]

II. WIRELESS ENERGY TRANSMISSION – THE FIRST 100 YEARS' LANDMARKS [2]

1864 – James Clerk Maxwell – the mathematical model of electromagnetic radiation

1888 – Heinrich Hertz – the *apparatus for generating electromagnetic waves* – by many considered as the first radio transmitter.

1894 – Jagdish Chandra Bose ignited gunpower and rang a bell with electromagnetic waves.

1895 – Jagdish Chandra Bose succeeded to transmit electromagnetic signals at a distance of approx. one mile.

1897 – Nikola Tesla (the real inventor of the radio, inventor of microwaves and alternating current) patented his first inventions in the field of wireless energy transmission using for experiments his Colorado Springs tower.

1900 – Gulielmo Marconi failed to patent his radio invention . The patent office remarks: "*Marconi's pretended ignorance of the nature of a "Tesla oscillator" being little short of absurd...*"

1901 – Gulielmo Marconi sends and receives signals over the Atlantic. Replying to one of his collaborators, engineer Otis Pond, Tesla answers: "*Marconi is a good fellow. Let him continue. He is using **seventeen of my patents**.*"

1904 – at the St. Louis World Fair, a prize was offered for operating an aeromodel having a motor of 0.1 HP from a distance of 100 ft.

1926 – Shintaro Uda and Hidetsugu Yagi published their first article concerning the future Yagi antenna.

1964 – CBS News presented a reality program in which William C. Brown (former technical director of *JPL Raytheon*) and Walter Cronkite demonstrated the operation of a microwave-powered model helicopter, supplied from a microwave beam.

Starting with the 1975 Goldstone Deep Space Communications Experiments a large range of researches, using powerful microwave devices and lasers are now currently performed.

III. TESLA IN COLORADO SPRINGS [1, 2, 4, 5]

About a year before his lecture on wireless energy transmission (1891) T. arrived to the conviction that rarefied air can be a suitable conductor and therefore he went to Colorado Springs to build a transmission station in this part of the Rockies. Leonard E. Curtis, a friend,, found land and provided power for the research from the El Paso Power Company of Colorado Springs and Colonel John Jacob Astor contributed with 30,000 US$ to the construction. Nobody knew exactly what Tesla intended to perform in this laboratory (fig.2) erected on the highest mountain of the Southern Rockies– the Pikes Peak – he generally informed the press that he wants to transmit a signal from the Pikes Peak to Paris.

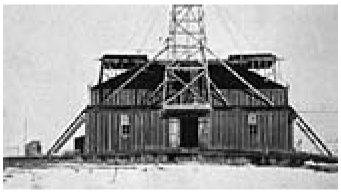

Fig.3. Tesla's Pikes Peak experimental station near Colorado Springs *[Courtesy of PBS]*[

What was experimented in this experimental station is still not clear. In Colorado, with its strong storms, T. performed measurements and soon came to the idea that the earth is a relatively good conductor. He emitted the hypothesis that man could transmit strong impulses to any location on Earth practically without losses. To confirm his theory he constructed a new laboratory in Colorado Springs that would be able to produce the strong impulses that will be conducted to the earth. This new laboratory has a rolled back roof to prevent it from sparks and a wooden tower 80 ft high. On its top was placed a 142 ft metal mast that supported a large copper sphere. Inside the wooden tower was placed a huge Tesla coil. In an experiment of one second, the secondary coil produced wonderful electric sparks and a blue corolla. T. ordered to repeat the experiment without a time limit. Huge 100 ft long man-made lightning bolts were produced from the metal mast to the earth. The play came to an end with the burning of the power plant's generators. Since this event, T. received no more free electricity for his experiments from the El Paso Power Plant. The experiments in Colorado Spring lasted 9 months. Though T. wrote all the experiments in his diary, a question remains still not elucidated: did T. transmit powerful energy waves at Pikes Peak? Some reports conclude that several vacuum tubes planted in the soil in a certain distance from his laboratory were alighted, but this fact can be attributed to the specific high conductivity of this soil [1].

Another approach of T. was on the resonance frequency of the atmosphere between the earth's surface and the ionosphere. T. calculated that this may be about 8 Hz, a result which was verified in the 50's by other researchers that reached the same result.

A third approach of T. was in the field of the ionosphere. T. anticipated correctly that this high stratus of the atmosphere may be highly ionized and therefore highly conductive. He started to think about a device able to transmit electrical energy from the Earth's surface to this high altitude. During this study, he received a repeating signal through his transmitter. He considered – and despite his ridiculization at his time for this announcement - today we may think too that Tesla was perhaps *the first person on Earth receiving a cosmic radio signal.*

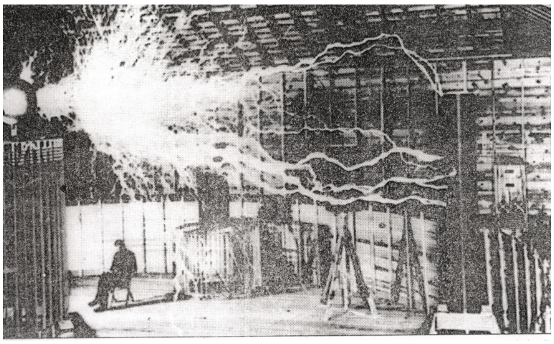

Fig. 4. The magnifying transformer at work in Colorado Springs*[Courtesy of Wikipedia]*[1]

IV. THE LECTURES OF TESLA AND HIS PREDICTIONS [1, 2, 4, 5, 6]

Hertz had demonstrated the existence of electromagnetic waves in 1887 using as source of high frequency oscillations the sparks produced by discharges of condensers [6]. In his experiments, Tesla needed high frequency power on a vast scale and at very high voltages. In 1891 Tesla developed an AC generator with 384 poles and an output frequency of 10,000 Hz and soon he invented other generators even at higher frequencies, up to 25,000 Hz. In his study of increasing frequency and voltages, he started with the invention of such giant rotating machines, but finally he came to the conclusion that they are unable to produce the necessary effects and switched to other alternatives. Tesla was all the time convinced that undamped current generation is very important, a method which at his time was still inapplicable and will be introduced only after 1910. His early models utilizing the induction principle with stationary coils formed the equipment used for the large commercial machines built afterwards for radio communications. In his experiments Tesla used tuned circuits, reduction coils and oscillatory spark circuits, which, combined with an oscillation transformer create spectacular arc discharges. He exposed his theory in some famous lectures. The first of them was in 1892, before the American Institute of Electrical Engineers in New York, the second, also in 1892, before the Institution for electrical Engineers and the Royal Institution in London, the Societe Internationale Francaise des Electriciens and the Societe Francaise de Physique in Paris and the third in 1893 before the Franklin Institute in Philadelphia and the National Electric Light Association in St. Louis. In these lectures, Tesla included also references on the structure of matter. In 1893 he also had an extensive exhibit at the Chicago World's Fair. A couple of years before, he explained how metals and dielectric materials could be heated using high frequency coils. He also described the effect of high frequency on the human body (diathermy). He produced vacuum tubes and introduced them in high frequency fields. In this way, he became the father of the neon tube and of the fluorescent lamp. He predicted the use of cheap synchronous electrical clocks. Such clocks were presented at the World's Fair in 1893. Tesla claimed that the future of aviation – than still inexistent, except the balloons - depended on the use of aluminum. Aluminum was a scarce and expensive material at this time and only the development of electricity contributed essentially to its price reduction. Tesla described in his early lectures and patents devices that became important many years later. For example, rotary sparks gaps and series spark gaps with small spacings predicted by T. in 1892-93 became a main issue in the wireless telegraphy from 1909 to 1920. In February 1892, Sir William Crookes published a prediction concerning the possibility of electromagnetic waves to be used for communications. Tesla has the idea to disturb the electrostatic field of the Earth using standing waves by excitation of its entire surface with high frequency power and than taking off power anywhere. Tesla described in his lecture in 1893 large antennas connected to the ground by wires for both purposes – transmission and reception and emphasizes the importance of applying the *electrical resonance principle to these schemes, using even variable tuning*. In 1896 Tesla succeeded to transmit electromagnetic signals to a boat on the Hudson River. About 1900, talking about the radio, he declared: "*I have no doubt that it will prove very efficient in enlightening the masses, particularly in still uncivilized countries and less accessible regions, and that it will add materially to general safety comfort and convenience, and maintenance of peaceful plants, all of which are capable of transmitting individualized signals to the uttermost confines of the earth. Each of them will be preferably located near some important center of civilization and the news that it receives through any channel will be flashed to all points of the globe. A cheap and simple device, which might be carried in one's pocket, may then be set up somewhere on sea or land, and will record the world's news or such special messages as may be intended for it.*" Since 1893, Tesla put as basic objective for himself a worldwide range of grandiose electrical effects, many of his priory invented patents were left for others to make from them less ambitious, but more practical applications. A few other researchers, inspired by his experiments and revelations, like Marconi, used his published patents for their own profit. [1].

V. TESLA – FATHER OF THE AC SYSTEM[1,16]

We remember that as a young scientist Tesla had to compete with the then universally recognized "Father of the DC electricity", Edison. The AC system invented by Tesla was much more effective than the DC system proposed by Edison. The facility to transport electrical energy at long distances was the main reason which determined the abandon of DC networks in favor of Tesla's AC network. The Tesla system was adopted for the first time at the Niagara Falls power plant, which was put in service in 1895. The first recorded assets for economical use of the Niagara Falls waters are dated 1759 (the Daniel Joncairs sawmill's channel). The Niagara Hydraulic Power & Manufacturing Company was founded in 1853. In 1875, the water turbines in the new erected powerhouse began to turn, but electrical power was produced commercially here only starting from 1879, since Schoellkopf introduced electrical technology to his powerhouse turbines. This was the start of one of the first power stations in the world. Illuminating the spectacular falls was a desiderate of many sponsors. In 1860, this occurred using expensive calcium flares, but in 1881, Charles Brush installed electric carbon arc lights, supplied by a small DC generator. In 1904 a new power plant, featuring 34,000 HP was built. In 1918, Schoellkopf's Hydraulic Power Company merged with the Edward Dean Adams' Niagara Falls Power Company. In 1886, Thomas Evershed proposed a project of 200,000 HP electricity generation at Niagara Falls, but this plan was demised due to the impossibility to distribute DC power at long distances. The Niagara Falls Power Company offered a 100,000 US$ award for anyone who could invent a system able to transmit electricity long distance. No one responded to this offer, because nobody was able to oppose to Edison, except Tesla. The greatest scientific minds of the world met in London to discuss the solutions. Against the advice of Thomas Edison and William Kelvin, Tesla's alternating current system was selected as the standard to be utilized. George Westinghouse started to develop Tesla's system. In 1883, Westinghouse replaced the illumination system of the Falls with AC technology. Ten years later, on May 6th, 1893, the Cataract Construction Company, a subsidiary founded in 1889 of the Niagara Falls Power Company, headed by the world's greatest financiers (J.P. Morgan, John Astor, William Vanderbilt and Edward Dean Adams), decided to use in its works in the Niagara Falls region alternating current (AC) for power generation and transmission. A new era in the history of electricity began. The AC system – today extended in the entire world – is one of the major inventions in electroenergetics, and all this - thanks to Tesla.

VI. FROM COLORADO SPRINGS BACK TO NEW YORK [1] .

Back to New York, Tesla wrote an enthousiastic paper for the "Century Magazine", where he describes in a futuristic view means to capture the Sun's radiation energy, to control the weather using electricity, predicts machines that would make wars impossible, and exposes his theory on the worldwide wireless energy transmission, as a global network that would link power stations, telecommunication, radio and even transmit images at remote locations. One of the most powerful magnats of this time, J.P. Morgan, offered to Tesla a capital of 150,000 US$ to build a transmission tower and a power plant, but the sum was too small to cover all the expenses. An offer of 1,000,000 US$ would be more reasonable for this goal. T. accepted this sum and started the construction works. Unfortunately, the history was cruel with this project. After tangled construction phases, due to lack of funds, in 1917 the Marines dynamited it by order of the USA Government, to avoid German spies' use.

VII. TESLA – FATHER OF ROBOTICS [1].

Arrived in New York from Colorado Springs, Tesla tried to shock the assistance in his presentations. For example, a short time after his arrival, in 1898, he presented in the Madison Square Garden arena in New York, in an electrical exhibition, *the world's first radio controlled vessel model.* This is the first attempt to build *automatic robots.* Tesla Patent Nr. 613,809 describes the first device having a wireless remote control. The working model, the *teleautomaton* was controlled with radio signals and powered with an internal battery. On the device were mounted some small lamps by means of which – by switching them wireless on and off from the remote controller, Tesla responded to some easy mathematical questions of the assistance. Being questionned by a reporter about the possibility to use wireless controlled robotics in war missions, like explosive charges' bearing machines, T.replyed: "*You do not see there a wireless torpedo, you see there the first of a race of robots, mechanical men which will do the laborious work of the human race.*"

VIII. TESLA – FATHER OF WIRELESS [1,2,3,5,6]

In a paper entitled "*The True Wireless*"[3], published in the *Electrical Experimenter* in May 1919, Tesla explains his patent concerning wireless transmission of energy using earth's electric resonance in a more popular language: *The arrangement, as I described it in detail, is shown in Fig. 8 [Fig.5 – P.B.]*[3]. *In this case an alternator energizes the primary of a transformer, the high tension secondary of which is connected to the ground and an elevated capacity and tuned to the imprest oscillations. The receiving circuit consists of an inductance connected to the ground and to an elevated terminal without break and is resonantly responsive to the transmitted oscillations.*

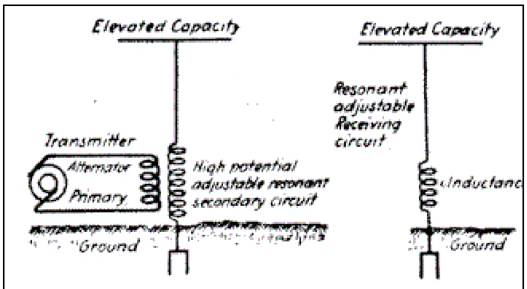
Elevated Capacity
Elevated Capacity
Resonant adjustable Receiving circuit
Transmitter
Alternator
Primary
High potential adjustable resonant secondary circuit
Inductanc
Ground
Ground

Fig. 5. Tesla's System of Wireless Transmission Thru the Earth as Actually Exposed In His Lectures Before the Franklin Institute and Electric Light Association in February and March, 1893 [3]

I gave to the world a wireless system of potentialities far beyond anything before conceived. *I made explicit and repeated statements that I contemplated transmission, absolutely unlimited as to terrestrial distance and amount of energy. But, altho I have overcome all obstacles which seemed in the beginning unsurmountable and found elegant solutions of all the problems which confronted me, yet, even at this very day, the majority of experts are still blind to the possibilities which are within easy attainment. My confidence that a signal could be easily flashed around the globe was strengthened thru the discovery of the "rotating brush," a wonderful phenomenon which I have fully described in my address before the Institution of Electrical Engineers, London, in 1892… This is undoubtedly the most delicate wireless detector known, but for a long time it was hard to produce and to maintain in the sensitive state. These difficulties do not exist now and I am looking to valuable applications of this device, particularly in connection with the high-speed photographic method, which I suggested, in wireless, as well as in wire, transmission"*[3].

IX.NOBEL PRIZE CONTROVERSIES [1].

Thomas Edison and Tesla were proposed as potential laureates to share the Nobel Prize of 1915 in a press dispatch, but finally due to their reciprocal animosity, the prize was not given to them., despite their enormous scientific contributions. In the following events after the rumors, neither Tesla nor

Edison won the prize (although Edison did receive one of 38 possible bids in 1915, and Tesla did receive one bid out of 38 in 1937). Earlier, rumors circulated that Tesla was have been nominated for the Nobel Prize of 1912, for his experiments with tuned circuits using high-voltage high-frequency resonant transformers. In 1915, Tesla filed an unsuccessful lawsuit against Marconi.

X. TESLA'S AWARDS [1].

For his multilateral groundwork in theoretical electrotechnics and for his prodigious inventor's and teacher's activity Tesla was awarded with some of the highest titles and prizes a scientist and an engineer can dream about during his life – except the Nobel Prize, that he virtually might have received instead of Marconi if in 1909 the Nobel Commission would had been correct. Tesla received the Elliott Cresson Gold Medal in 1893, the Edison Medal in 1916 and the John Scott Medal in 1934.

XI. THE RIVAL APPEARS [10, 12, 13, 14, 15]

Fig.6. Guglielmo Marconi, [*Courtesy of INVENTORS*] [10]

Guglielmo Marconi (fig.6) was born on April 25,.1874 in Bologna, Italy. His father was a rich Italian landowner and his mother, a granddaughter of the founder of the Jameson whiskey distillery. Marconi learned in Bologna in the lab of Augusto Righi, in Florence at the Istituto Cavallero, and, finally in Livorno. First experiments were performed at home, in December 1894 . After sending LW signals at a distance of 1 mile in 1895, he went to England and founded the Marconi Company at Chelmsford, Essex in 1897. His first patent in wireless telegraphy, Nr. 12039, was received on June 2, 1896, followed in July by experiments from the General Post Office building in London in a distance of 1.5 miles. On September 2, 1896 the span had increased to 8 miles with tests at Salisbury plain and then, on 11.05.1897, across the Bristol Channel (3.5 miles). Daily Express, in July 1898 was the first newspaper using radiotelegraph. Another device was installed at the board of Queen Victoria's Royal yacht. The extension of the new invention was really rapid. In December 1898 wireless telegraphy was installed between East Goodwin light ship and the South Foreland lighthouse and on March 3, 1899, wireless communication was employed to save life. In summer 1899 the Marconi's wireless signals crossed the English Channel and on November 15, 1899 the first ocean wireless transmissions linked ships. On April 26, 1900, M. received his patent Nr. 7777 for *"Tuned Syntonic Telegraphy"*, an invention that used tuned circuits, enabling a wireless set to be tuned to specific emitting station (like a radio receiver today). On January 23, 1901 the first long distance (196 miles) wireless transmission was performed from the Isle of Wight to The Lizzard in Cornwall. The next task was to transmit over the Atlantic and this was realized on Thursday the 12th of December 1901.The letter "S" transmitted from Poldhu, Cornwall, was – according to Marconi and his assistant, Georges Piget - received at Signal Hill, St. Johns, Newfoundland. In a span of a few decades, radio transmissions changed the world and brought nations closer to each other. In 1909, Marconi shared the Nobel Prize with Karl Ferdinand Braun, from which he "borrowed" some parts of his work (like Braun's British patent on tuning), as he will recognize later. Guglielmo Marconi used also the patents of others, as we will see in the next chapter.. Marconi's 1901 tests were contested by scientists (Edison and others), motivating that M.receives only parasite signals from the atmosphere. Dr. Jack Belrose estimated that the Poldhu transmitter has a frequency of 850 Hz and what Marconi received were only SW, emitted by the ionosphere. Fearing to be challenged by concurrents, M. conducted new tests over the ocean, this time using a ship. These tests gave coherer-tape reception up to 2,496 km (1,551 mi), and audio reception up to 3,378 km (2,099 mi) during the night and half of these results during daytime, i.e. that Marconi's mid-day Atlantic wireless first crossing in 1901 with LW and MW was seemingly not real. An 1902 unsuccessful test from Glacier Bay confirmed this supposition. In 1903, President Theodore Roosevelt greeted wireless King Edward VII to HM birthday [I still must look for the way - directly or by intermediary ship-stations (?)]. The two Marconi's company radio operators on the "Titanic" rescued survivors by their SOS signals.. In 1915 the Chelmsford Marconi factory, using a vacuum tube transmitter, organized the first entertainment radio broadcast and in 1922 such transmissions became regular (from the Marconi Research Centre at Writtle). Marconi. reached the Italian Senate in 1914 and was made a marchese in 1924 by King Victor Emmanuel III. In 1923 he had adhered to the fascist party of Benito Mussolini, which made him in 1930 President of the Accademia d'Italia. Later he was appointed a Member of the Fascist Grand Council. Marconi died in Rome in 1937 at the age of 63, after a series of heart attacks. All the radio stations in the world kept silence for two minutes. After state funerals his remains were deposed in his house, Villa Griffone, Sasso Marconi, Emilia-Romagna.

XI. A LATE RECOGNITION[2, 9, 12, 13, 17]

Marconi's work is based on copies of patents of many other inventors without their permission. His so-called original "two-circuit" device, a spark-gap transmitter plus a coherer-receiver, was similar to those used by Oliver Lodge in a series of worldwide reported tests in 1894. Tesla disputed that Marconi was able to signal for greater distances than anyone else when using the spark-gap and coherer combination. In 1900 Alexander Stepanovich Popov declared to the Congress of Russian Electrical Engineers: "[...] *the emission and reception of signals by Marconi by means of electric oscillations* [was] *nothing new. In America, the famous engineer Nikola Tesla carried the same experiments in 1893*." The Fascist regime in Italy credited Marconi as inventor of the radio. The doubt is if his own contribution was sufficient to constitute an independent patent, or if his component devices were too similar to the original ones (if not copies) developed years before by Hertz, Popov, Branly, Tesla, and Lodge – and in this case, they are copies. Marconi's "four-circuit" design, having two tuned-circuits at the transmitting and receiver antennas, **British Patent Nr.7777** from 26 April 1900 is a compound of the earlier work of Tesla and other inventors. 5of 9 judges of the Supreme Court stated that Marconi's work had been anticipated by **John Stone Stone** (patent No.714,756) and **Oliver Lodge** (patent No. 609,154). The Supreme Court also studied **Tesla's patent No. 645,576** and noted that **Tesla had used four tuned circuits before Marconi**. In order to defend himself, in 1911 Marconi purchased through his company the Lodge-Muirhead Syndicate, whose main performance was Oliver Lodge's tuning patent, invented in 1897. Thus, the 7777 Patent was vividly discussed at this time from validation to complete rejection. In 1943, the US Supreme Court stated that: "*The Tesla patent No. 645,576, applied for September 2, 1897 and allowed March 20, 1900, disclosed a four-circuit system, having two circuits each at transmitter and receiver, and recommended that all four circuits be tuned to the same frequency. [... He] recognized that his apparatus could, without change, be used for wireless communication, which is dependent upon the transmission of electrical energy.*" In its decision, the court noted: "*Marconi's reputation as the man who first achieved successful radio transmission rests on his original patent, which became reissue No. 11,913, and which is not here [320 U.S. 1, 38] in question. That reputation, however well-deserved, does not entitle him to a patent for every later improvement which he claims in the radio field. Patent cases, like others, must be decided not by weighing the reputations of the litigants, but by careful study of the merits of their respective contentions and proofs*". The court also stated that: "*It is well established that as between two inventors priority of invention will be awarded to the one who by satisfying proof can show that he first conceived of the invention.*" Following to this decision, most of Marconi's patents were overturned by the US Supreme Court. A decision that came unfortunately 9 months after Tesla's death.and

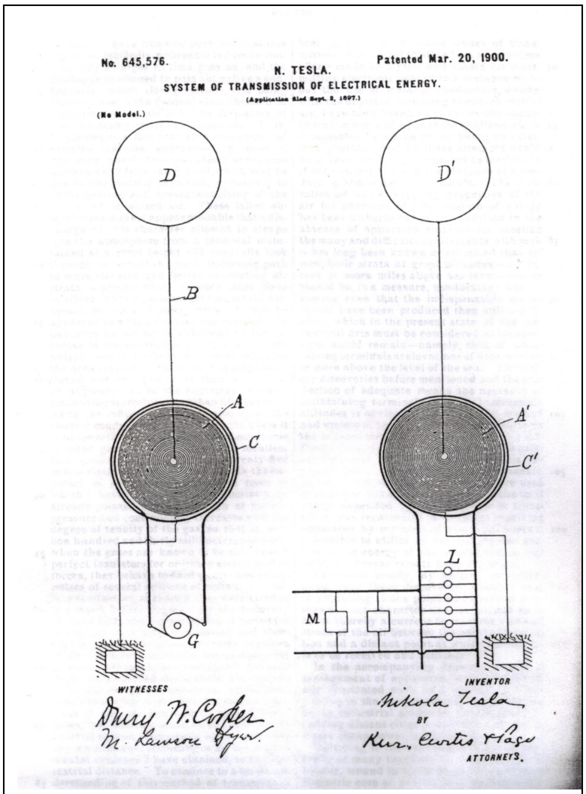

Fig. 7. Tesla's System of transmission of electrical energy, Patent No. 645,576, patented Mar. 20, 1900 [5]

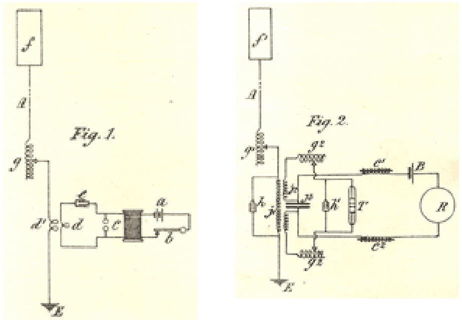

a) b)

Fig. 8. Marconi's British Patent No.12039 from 1900 – the so called "4 circuits patent": a) - emitter diagram, b) – receiver diagram [*Courtesy of Sparkmuseum*][9]

which was pronounced not mainly in order to do simply justice. At the time, the US Army and Marconi's company were involved in a trial on the radio patent rights, and in order to anullate Tesla's former claims for compensation, the Court nullified Marconi's other inventions. The priority of Tesla's US Patent No.645, 576 on the Marconi's US Patent No.763,772 is more evident if we use the following suggestive comparative timeline diagram, published in the Internet (Fig.9) [1]

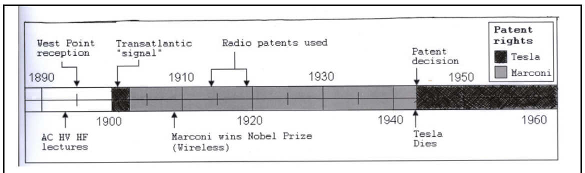

Fig. 9. Comparative timeline diagram illustrating the priority of Tesla's Patent No.645,576 on Marconi's radio patent [*Courtesy of Wikipedia*][1]

Tesla was not alone in claiming against Marconi. Oliver Lodge claimed the usage of his unpatented ideas in Marconi's first radio patent. Don Kimberlin comments: *"…What did happen by April 1900 was that Marconi uniquely had combined Professor Righi's transmitter, Branly's coherer, Tesla's elevated antenna and Lodge's tuned circuits to realize what ultimately became first wireless telegraphy…"*[18]. In order to compare the patents, please see the figures (Fig.5, 7, 8a and b). From these drawings results that the use of resonant circuits was clearly primordially demonstrated (1893) and filed (1898) by Tesla, the Marconi patent was only afterwards "adjusted". Marconi's invention is only an application of Tesla's (and of others') prior researches and studies. The differences consist in the used media – in Tesla's patent it was the Earth, in Marconi's patent it is the air. In Tesla's patent, a high frequency polyphasic alternator provides the HF initial impulses, whereas in Marconi's feature, the impulses are produced by switching on-off manually the telegraph's switch. The differences are not in the principle, but in the application's purpose. Tesla – as discussed – wanted to create his planned worldwide wireless system, whereas Marconi – using the wireless transmission knowledge and technology previously set by Tesla and others – performed an apparatus for common use– the radiotelegraph - and fructified the results.

XII. CONCLUSION

Tesla was the first to invent the radio. Marconi did not really invent the radio, he only combined, copied and adapted patents of Tesla and others and commercialized them as his own patent. He died in a plenty of glory and richess, whereas Tesla died as a poor man. No comments…

XIII. REFERENCE

[1] Wikipedia – *Nikola Tesla*
[2] Wikipedia – *Wireless energy transfer*
[3] Nikola Tesla – *The True Wireless, Electrical Experimenter*, May 1919, Wikipedia
[4] Nikola Tesla – *Colorado Springs Notes*, Nolit, Beograd, 1978
[5] Leland Anderson - *Work with Alternating Currents and TheirAplication to Wireless Telegraphy, Telephony and Transmission of Power,* Apendix B, United States Patents Office - Nikola Tesla, System of transmission of electrical energy, No. 645,576, patented 20.03.1900, Sun Publishing, Denver, 1992.
[6] Charles A. Jost – *The Tesla Experiment, Lightning and Earth electrical resonance – A Commentary on 1983 Research and Test Results*, Millbrae, CA, 1983.
[7] *Catalogue of Tesla's patents*, Beograd, 1987.
[8] Margaret Cheney - *Tesla: Man Out of Time*, New York: Dorset Press, 1981
[9] Sparkmuseum, Internet.
[10] Inventors, Rockradio, UK, Internet.
[11] PBS, Internet.
[12] Joe Craig, VO1NA - *Marconi's First Transatlantic Wireless Experiment***,** Marconi Radio Club of Newfoundland
[13] A. David Wunsch - *Misreading the Supreme Court: A Puzzling Chapter in the History of Radio*, Internet, Antenna, Newsletter of the Mercurians, in the Society for the History of Technology
[14] Barry Mishkind - *This is the Marconi Archive Section of The Broadcast Archive*, Internet.
[15] Helen Briggs - *Profile: Marconi, the wireless pioneer*, BBC Sci-Tech, Internet
[16] *Niagara Falls – History of Power*, Internet
[17] Leland Anderson – *Priority in the invention of the radio*, *Tesla vs. Marconi*
[18] Don Kimberlin - *Jurassic Telecommunications, Part II, - The Global Influences Leading Up To Marconi's "Lucky 7's" Patent*, The Jurassic Radio Section of the Broadcasting Archive, maintained by Barry Mishkind, Internet.

Paul Brenner , M.Sc.,finished his studies at the Bucharest Polytechnic Institute in 1970 and worked in the field of energy balances' achievement (and, as well, also in other projects), all of them for large industrial platforms and for the coal mining industry, until his emigration to Israel in 1982. In parallel with his actual daily activity in the Israel Electric Corporation in Haifa, Israel, he is a permanent member of the World Renewable Energy Network and Council and participates regularly in WREC/ WREN congresses. Author of more than 30 papers, presentations and a course of energy balances' achievement, he sustained together with **Christopher Dunn**, an USA laser engineer, the known author of the bestseller ***The Giza Power Plant - Technologies from Ancient Egypt*** a workshop in Aberdeen, UK, in May 2005, in the pre-congress days of WREC 2005 Europe, entitled ***Hi-Tech Antiquity***. *E-mail addresses: utc51@iec.co.il and pbrenner@bezeqint.net*

RF Harvesting – Experiment 1 (PASS)

by w5fcx | Feb 24, 2017 | Antennas, electronics, Software-defined Antenna | 0 comments

I'm happy to report that the first experiment on RF harvesting for DC power for the BatBeam project was a big success. I recorded the event, with the help of my wife Sandy, who did an excellent job taking the video, which you will find below. The significance of this experiment may not be apparent at first. What it means is that the first step in establishing the feasibility of a fully-wireless J-Box has been taken successfully. This is a critical step to have passed, as we must have DC power at each J-Box in order to power the RF switching relay and J-Box controller circuitry.

Here's the video. The schematic of the simple circuit along with some additional photos are below.

Amateur Radio RF Harvesting Experiment 1

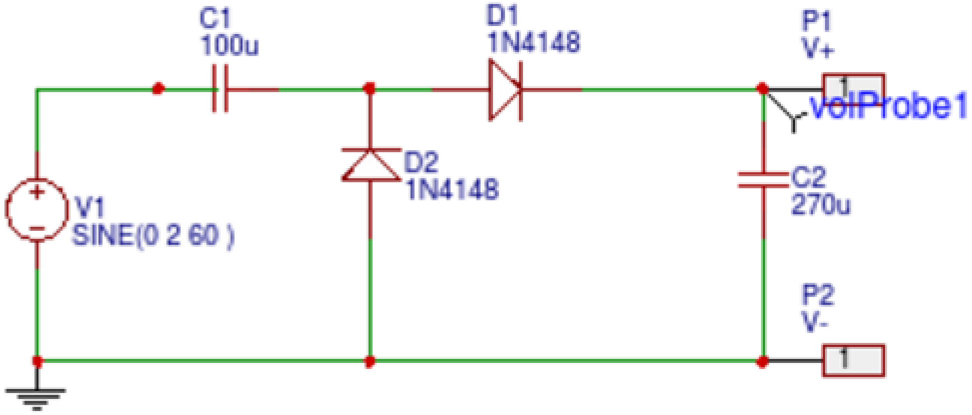
C1
100u
D1
1N4148
P1
V+
1
volProbe1
D2
1N4148
C2
270u
V1
SINE(0 2 60)
P2
V-
1

In the above schematic, the signal generator is replaced by the RF harvesting inductor, 10 turns of wire wrapped about the end of one G5RV antenna segment (a 20 meter G5RV). The RF harvesting inductor is shown below.

Below you will see the extremely simple proof of concept rectifier circuit, along with the RF inductor feedline.

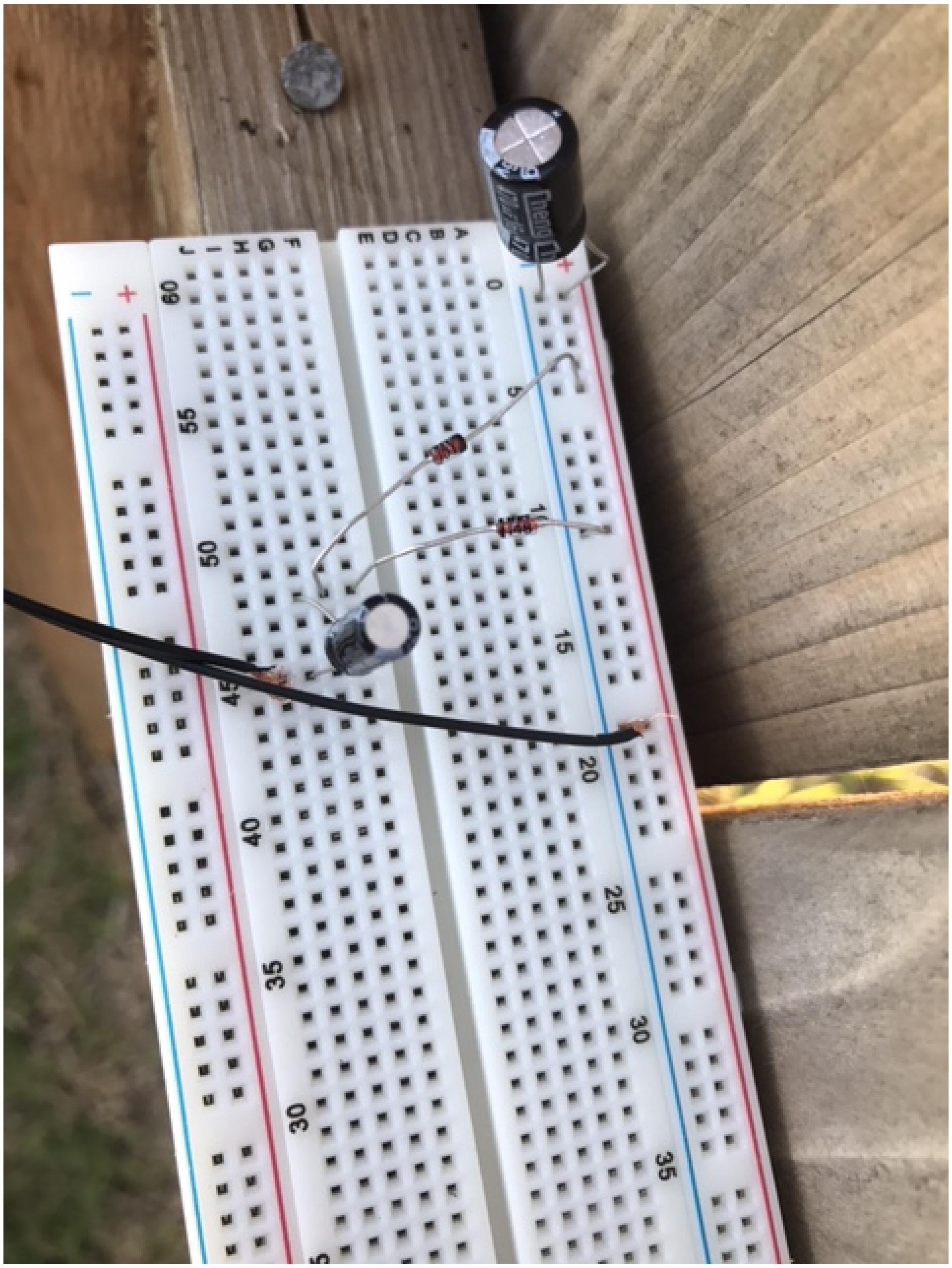

This concludes RF harvest experiment 1, which PASSED.

PDF

Trackbacks/Pingbacks

1. Prescaling Mhz into Khz using PICDIV 1000 | W5FCX - […] the frequency counter and RF detector. As RF is harvested (see earlier posts on RF Harvesting 1 2 3…

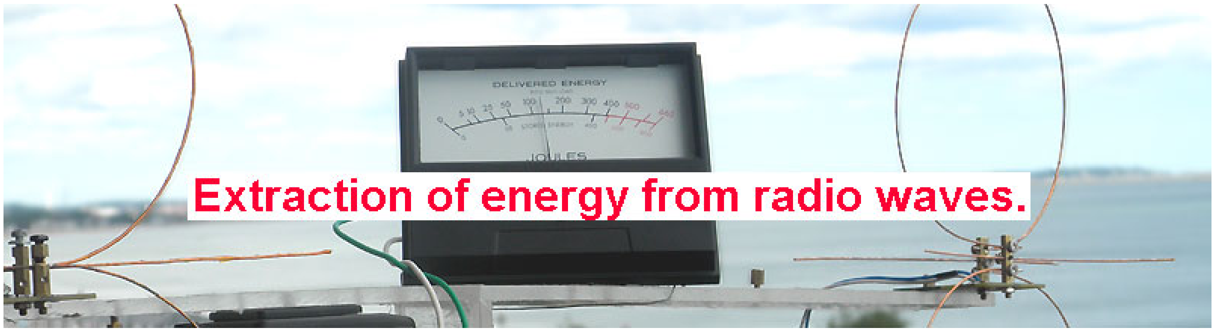
DELIVERED ENERGY
JOULES
Extraction of energy from radio waves.

Simply, I like to check things.

Harvesting power from radio waves.

Assuming that most people came here to just find out how much energy can be obtained from radio waves, we will be short.

On the banner, two '**D**ual-**Loop**' directional antennas, designed specifically to collect power at 750MHz. One is connected to 100uA (700Ohms) head, and the average current you can see on it. Both antennas have 'on-board' rectifiers.

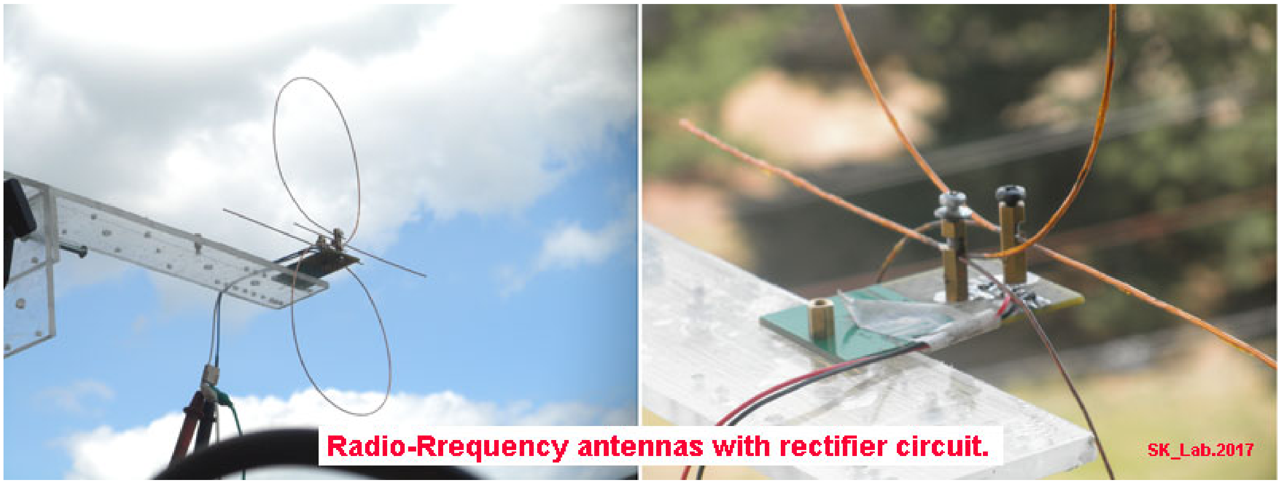
Radio-Rrequency antennas with rectifier circuit.
SK_Lab.2017

Reality is not as good as we would like. One antenna produces about 1 volt (at 10MOhms), and a current of about 0,4 mA (at 700Ohms).

The circuit is simple as a crystal receiver. Diodes with capacitors work as a voltage doubler (rectifier). The remaining inductors and capacitors isolate the output from high frequency.

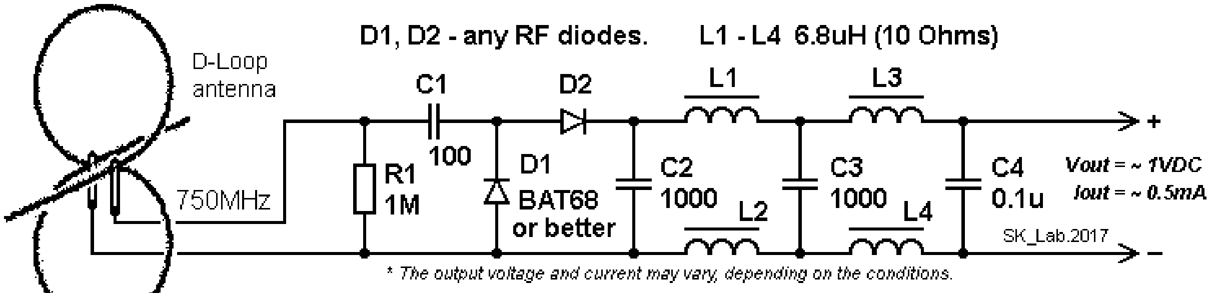
D1, D2 - any RF diodes.
L1 - L4 6.8uH (10 Ohms)
D-Loop antenna
750MHz
C1
100
D2
R1
1M
D1
BAT68 or better
C2
1000
L1
L2
C3
1000
L3
L4
C4
0.1u
+
–
Vout = ~ 1VDC
Iout = ~ 0.5mA
SK_Lab.2017
* The output voltage and current may vary, depending on the conditions.

Harvesting RF power. Absorbs energy from cell towers at 750MHz.

Any RF diodes with junction capacitance less then 1.5pF will be suitable for this application. See the end of this page for a suitable choice.

It's hard to believe, but the old style panel meter is the best indicator of the power that was obtained.

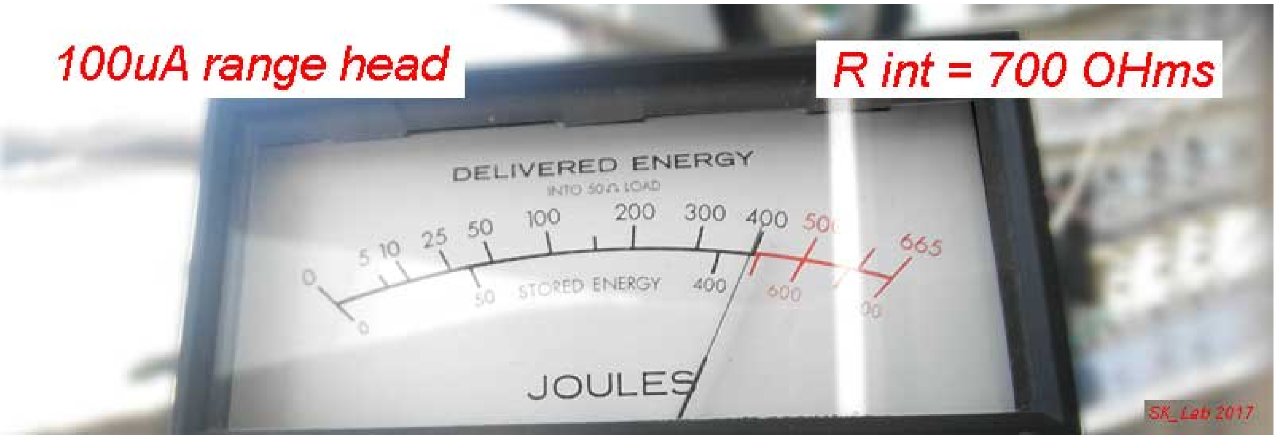

After simple calculations, the total power, received at the moment, is about 0.5mW.

It was a good, top result. But interesting to know what average power can be expected. A simple installation allows me to get more reliable results. We measure the open voltage (at 5 MOhms) and short current (at 4 Ohms) produced by one antenna. The measurement time for obtaining the average is about half an hour. Both multimeters work together to eliminate an accidental error.

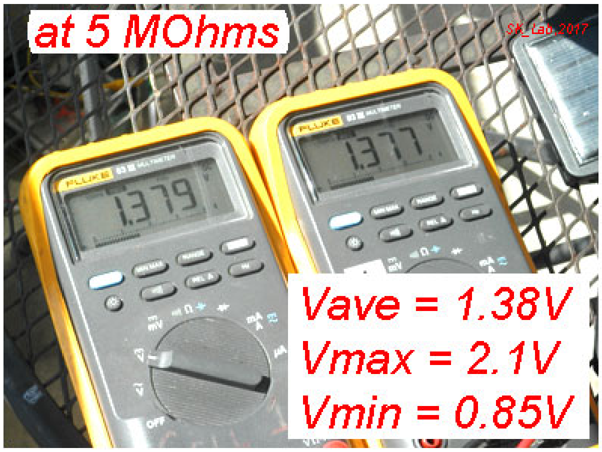

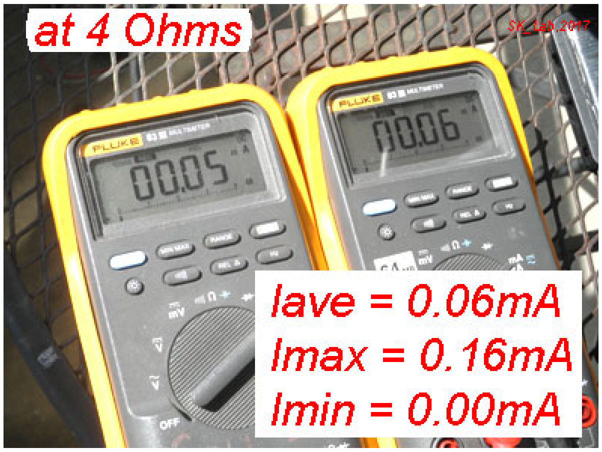

And here we return to reality. During the measurement time, the average power taken from the air by one antenna was ~0.02mW. Even the fact that the maximum power jumped to ~0.8 mW did not change anything.

For most people, these digits do not say anything. So let's move on to the practical side.

Suppose you want to light the LED. 0.02mW is not enough to light the LED even very, very dimly. One way to do this is to charge the capacitor and when enough energy is stored, discharge it through the LED.

Let's say you want to light a white LED (3V), with a current of 0.01A, for 0.1 second. It requires 3**V***0.01**A***0.1**sec**.= 0.003**J**oule. The antenna produces, an average, 0.00002W*1**sec**.=0.00002**J** each second. This means that to get the necessary energy (0.003J), you need to wait 150 seconds or 2.5 minutes.

But it is not all that bad. During the 24 hours, this simple setup with one antenna, can collect about 1.7**J** of totally free energy. This is enough that a data logger would work, and once a day transmit data via wireless or so. For example, a device (or LED) consuming 10mA at 3.6V can run at least 45 seconds per day on the stored energy.

In order to make this idea run, in your particular case, you need to:
a) find a 'hot spot' using any RF detector;
b) with any SDR Receiver find RF frequency with max. power;
c) make a simple antenna (Dual-Loop or so), add RF rectifier and enjoy 'free' energy :)

Yes. It's all. It's simple like that.

* A good article, about radio emission in London UK,(pdf): 'Ambient RF Energy Harvesting in Urban and Semi-Urban'.

In case of questions - address is below.

Marwyn and John's home pages

... where you will find details of some of the things that we want to make available to friends and family.

Main menu

- Home
- Holidays
- Sailing
- Music
- Family History
- Computers
- Astronomy
- Amateur Radio
- Contacts

Links

- Lothians Radio Society

Navigation: Home > Amateur Radio > Homebrew > Directional coupler

John's homebrew pages

Microwave directional coupler - and an RF detector

More deviations from the main project.

I was concerned that, as I manage to generate a bit more power (hopefully!) at microwave frequencies, a simple RF probe would not be sufficient for power measurement, and the diode could be over-voltaged. A directional coupler allows a small amound of the forward signal to be sampled and measured. A very simple design is to use microstrip, as shown in the photo. A BAT85 Schottky diode is used as the detector, with 100pF and 10nF decoupling. The other end of the coupling loop is terminated with two 100 ohm resistors in parallel.

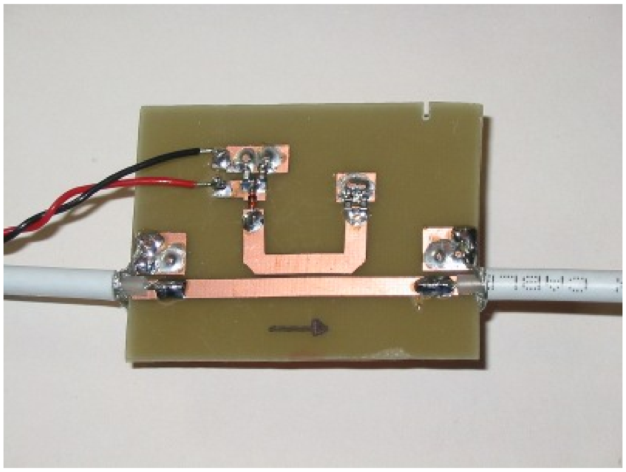

Here it is in use, with the signal being dissipated in the dummy load. A 100 microamp meter is used as the detector (100-0-100 microamps here since that's what I had available). It seems to work well, and is very directional indeed - very little signal is measured if it's connected the other way round. I don't know what the coupling is with any accuracy yet, but it will be fine for relative measurements for now.

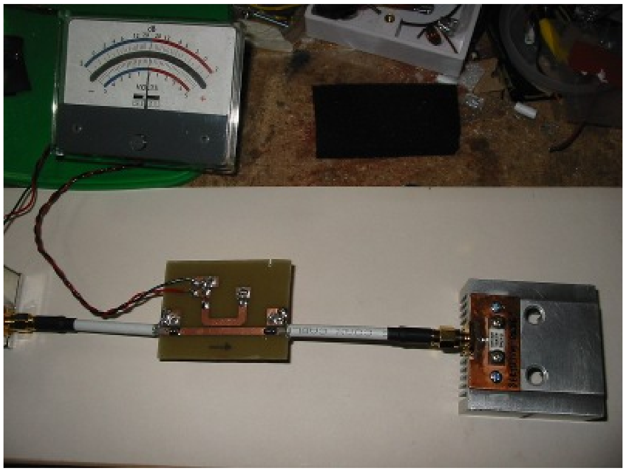

Having proved its success, I will probably make one with forward and reverse detectors that can be put in a box and used as a "through line" power meter for forward and reverse measurements.

I also wanted something a bit more reliable than my hand held RF probe for making measurements of power in test rigs. With that, since it really needs to be mounted in a screened tube, the signal can vary a bit depending on how the probe is set up. I thought that if I could put the RF detector into a BNC plug, then I might get rather more consistent measurements. Here are the components before assembly into a BNC compression plug. The BAT85 diode fits right down near the probe; two SMD 1nF capacitors are soldered to a brass washer, and to the other end of the diode, to give the decoupling; and for good measure, the output is taken through a 2mm ferrite bead. When I get round to scanning some circuit diagrams I'll try to remember to add a sketch of the arrangement here.

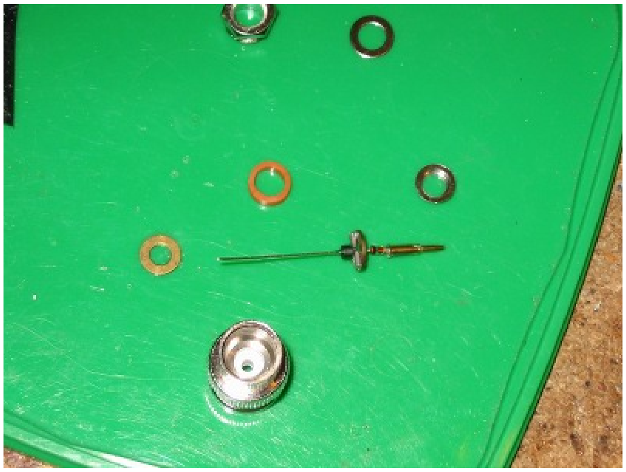

This new detector seems to be much more consistent than the RF probe, though I'm not yet happy that I can be sure of power measurements calculated from the detected DC voltage!

© Copyright Marwyn and John Cooke 2000-2015

The Basics of Patch Antennas, Updated

By D. Orban and G.J.K. Moernaut,
Orban Microwave Products
www.orbanmicrowave.com

Introduction

This article introduces the basic concepts of patch antennas. We use a simple rectangular, half wave long, probe-fed patch operating in its fundamental mode as an example.

Topics include principles of operation, impedance matching, radiation patterns, circular polarization, bandwidth, efficiency, alternative feed types, stacked patches and higher mode behavior.

This article was originally published in September 2005. After it was published, we received substantial feedback from those who read it. The article has been revised based upon the feedback we received.

Properties Of A Basic Microstrip Patch

A microstrip or patch antenna is a low-profile antenna that has a number of advantages over other antennas: it is lightweight, inexpensive, and electronics like LNA's and SSPA's can be integrated with these antennas quite easily. While the antenna can be a 3-D structure (wrapped around a cylinder, for example), it is usually flat and that is why patch antennas are sometimes referred to as planar antennas.

Figure 1 shows a patch antenna in its basic form: a flat plate over a ground plane. This antenna is often built of printed circuit board material and the substrate makes up the patch antenna's dielectric. The distance between the patch and the ground plane – the substrate or dielectric height h – determines the bandwidth.

A thicker substrate increases the gain to some extent, but may lead to undesired effects like surface wave excitation: surface waves decrease efficiency and perturb the radiation pattern.

The ground plane should extend beyond the edges of the patch by at least 2 to 3 times the board thickness for proper operation. A ground plane that is too small will result in a reduced front to back ratio.

Making the ground plane larger also increases the gain, but as the ground plane size increases, diffraction near the edges plays less of a role and increasing the size of an already "large" ground plane has very little effect on gain.

In the antenna in *Figure 1*, the center conductor of a coaxial line serves as the feed probe to couple electromagnetic energy in and/or out of the patch. A thicker substrate leads to a longer feed probe, a larger feed probe inductance and a degradation of impedance matching. This can be compensated by using a different feed type and we'll look at alternative feed methods further down.

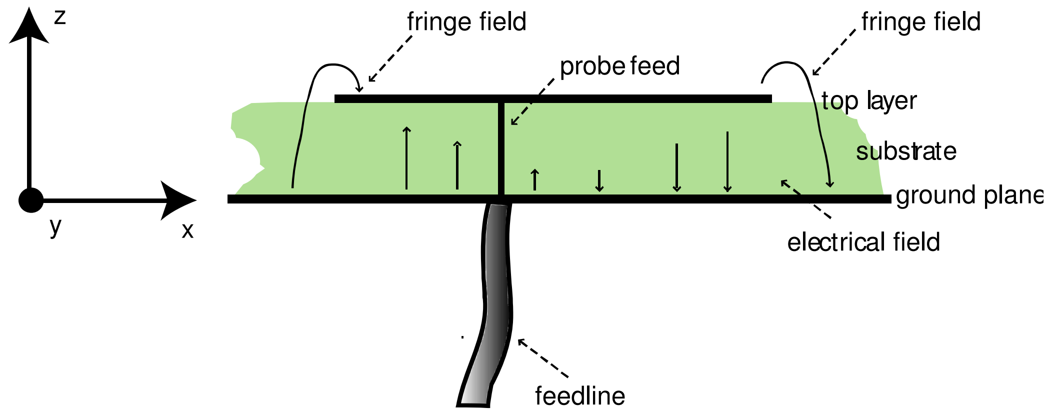

Figure 1: *Cross section of a patch antenna in its basic form*

A half wave long patch operates in what we call the fundamental mode: the electric field is zero at the center of the patch, maximum (positive) on one side, and minimum (negative) on the opposite side. These minima and maxima continuously change side like the phase of the RF signal.

The electric field does not stop abruptly near the patch's edges like it would in a cavity: the field extends beyond the outer periphery. These field extensions are known as fringing fields and cause the patch to radiate. Some popular analytic modeling techniques for patch antennas are based on this leaky-cavity concept and the fundamental mode of a rectangular patch is often denoted using cavity theory like the TM10 mode.

This TM notation often leads to confusion and here is an attempt to explain that: *Figure 1* uses a Cartesian coordinate system, where the x and y axes are parallel with the ground-plane and the z-axis is perpendicular to it.

TM stands for a magnetic field distribution –between patch and ground– that is transverse to the z-axis of the antenna shown in *Figure 1*. This 'transverse' with respect to the z-axis is usually dropped because the magnetic fields in patch antennas are always transverse to their z-axis.

So, we can simplify things and only consider three field components instead of six (magnetic and electric fields in each x, y and z axis): the electric field in the z direction, and the magnetic field components in x and y directions.

In general, modes are designated as TMnmp. The 'p' value is mostly omitted because the electric field variation is considered negligible in the z-axis since only a phase variation exists in the z axis. So, TMnm represents the field variations in the x and y directions. The field variation in the y direction (impedance width direction) is negligible and m is 0. The field has one minimum-to-maximum variation in the x direction (resonance length direction and a half wave long), n is 1 in this case and we say that this patch operates in the TM10 mode.

Dimensions

The resonant length (the x axis in *Figure 2*) determines the resonant frequency and is about $\lambda_d/2$ for a rectangular patch excited in its fundamental mode where λ_d is the wavelength in the PCB material. The patch is actually a bit larger electrically than its physical dimensions due to the fringing fields and the difference between electrical and physical size is mainly dependent on the PC board thickness and dielectric constant of the substrate.

A good approximation for the resonant length is:

$$L \approx 0.49\ \lambda_d = 0.49\ \frac{\lambda_0}{\sqrt{\varepsilon_r}}.$$

This formula includes a first order correction for the edge extension due to the fringing fields, with:

- L = resonant length
- λ_d = wavelength in PC board
- λ_0 = wavelength in free space
- ε_r = dielectric constant of the printed circuit board material

Other parameters that have less influence on the resonant frequency include:

- Ground plane size
- Metal (copper) and dielectric thickness
- Patch (impedance) width

Impedance Matching

The feed position of a patch antenna excited in its fundamental mode is typically located in the center of the patch width direction (y axis) and somewhere along the patch resonant length direction (x axis). The exact position along the resonant length is determined by the electromagnetic field distribution in the patch. Looking at the current (magnetic field) and voltage (electric field) variation along the patch, the current has a *maximum* at the center and a *minimum* near the left and right edges, while the electric field is *zero* in the center and *maximum* near the left and *minimum* near the right edges. Keep in mind that the field distribution constantly changes in amplitude and sign. *Figures 2 and 3* below clarify this:

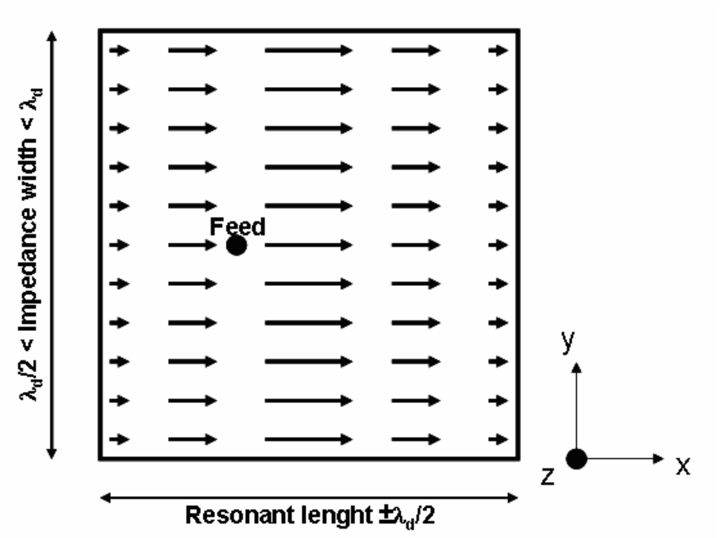
$\lambda_d/2$ < Impedance width < λ_d
Feed
y
z
x
Resonant lenght $\pm\lambda_d/2$

Figure 2: *Current distribution on the patch surface*

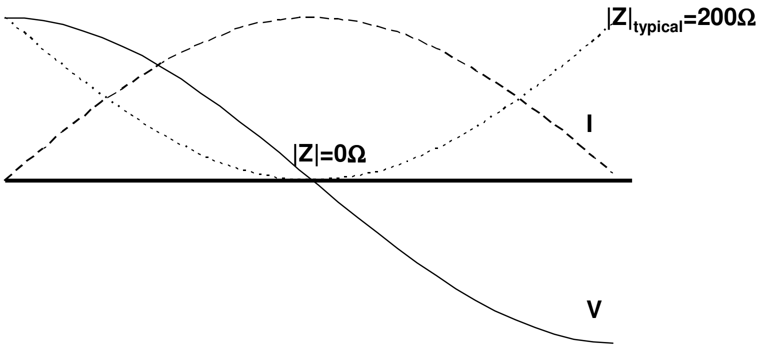

Figure 3: *Voltage (V), current (I) and impedance (|Z|) distribution along the patch's resonant length*

From the magnitude of the current and the voltage, we can determine that the impedance is minimum (theoretically zero Ω) in the center of the patch and maximum (typically a couple hundred Ω) near the edges. This means that there are two points where the impedance is 50 Ω somewhere along the resonant length (x) axis of the element and this is where you would typically connect to the antenna.

The possibility to connect to the patch at other impedance points is quite useful and impedances up to 200 Ω are common. For example, a two element array can be fed with a simple parallel feed by matching the individual patch elements to 100 Ω and connecting them in parallel results in 50 Ω end impedance without the need for impedance transformers. The same woks for a 4 element array with the elements connected at their 200 Ω point.

If you wanted to connect to the edge of the patch and were looking for a specific impedance, you could modify the width of the patch to yield the impedance your are looking for. Increasing the width decreases the impedance.

Fundamental Specifications Of Patch Antennas

Radiation Pattern

A patch antenna radiates power in certain directions and we say that the antenna has directivity (usually expressed in dBi). If the antenna had a 100% radiation efficiency, all directivity would be converted to gain. Typical half wave patches have efficiencies well above 90%.

The directivity of a patch can be estimated quite easily:

The radiating edges of a patch can be seen as two radiating slots placed above a ground-plane and, assuming all radiation occurs in one half of the hemisphere (on the patch side of the ground), we get a 3 dB directivity increase. This would be an antenna with a perfect front-to-back ratio where all radiation occurs towards the front and no radiation towards the back. This front-to-back ratio is highly dependent on ground-plane size and shape in real life.

Another 3 dB can be added because there are 2 slots. The length of these slots typically equals the impedance width (length in the y-axis) of the patch and the width of these slots equals the substrate height. These slots typically have a directivity of 2 to 3 dB compared to an isotropic radiator and behave like a dipole.

All of this results in a total maximum directivity of 8 to 9 dBi.

The rectangular patch excited in its fundamental mode has a maximum directivity in the direction perpendicular to the patch (z-axis or broadside). The directivity decreases when moving away from broadside towards lower elevations. The 3 dB beamwidth is the width at which the gain of the beam decreases by 3 dB relative to the gain in broadside to either side of the main beam.

Figure 4 shows a typical radiation pattern for a square, half wave patch.

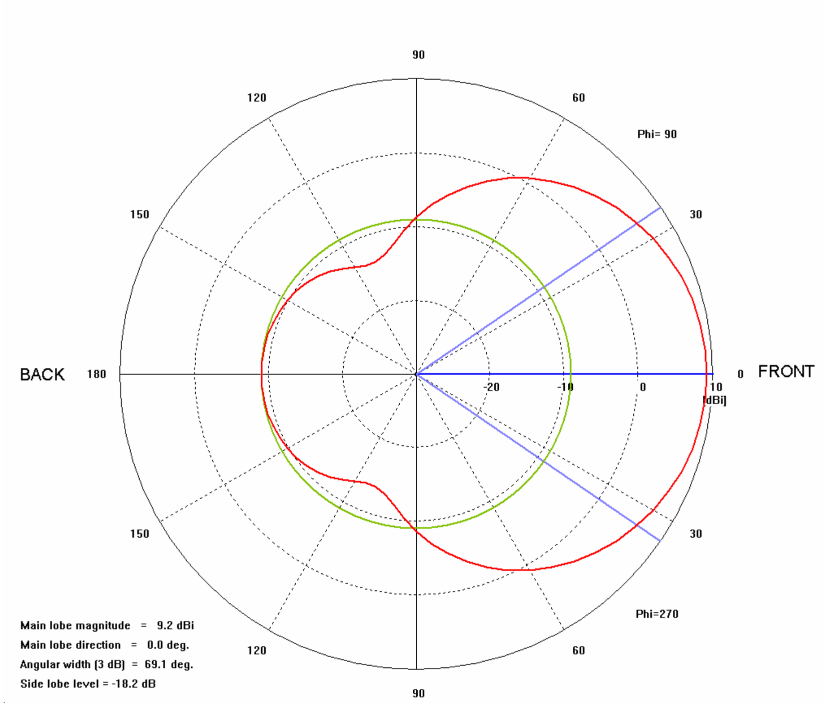

Figure 4: *Typical radiation pattern of a simple square patch*

So far, the directivity has been defined relative to an isotropic radiator and we use dB*i*. An isotropic radiator emits an equal amount of power in all directions and it has no directivity. Antenna directivity can also be specified relative to that of a dipole. A dipole has 2.15 dB*i* of directivity over an isotropic radiator. When we specify the directivity of an antenna relative to a dipole, we use dB*d*.

No antenna losses have been included so far and the integrated average of the directivity pattern over an entire sphere has to be 0 dBi. This implies that creating directivity in a certain direction reduces directivity in other directions.

Antenna Gain

Antennas do not have gain because they are passive structures. Antenna gain is defined as antenna directivity times a factor representing the radiation efficiency. Radiation efficiency is always lower than 100% so the antenna gain is always lower than antenna directivity. This efficiency quantifies the losses in the antenna and is defined as the ratio of radiated power (Pr) to input power (Pi). The input power is transformed into radiated power, surface wave power and a small portion is dissipated due to conductor and dielectric losses. Surface waves are guided waves captured within the substrate and partially radiated and reflected back at the substrate edges. Surface waves are more easily excited when materials with higher dielectric constants and/or thicker materials are used. Surface waves are not excited when air dielectric is used. Several techniques to prevent surface wave excitation exist, but this is beyond the scope of this article.

Antenna gain can also be specified using the total efficiency rather than just the radiation efficiency. This total efficiency is a combination of the radiation efficiency and efficiency linked to the impedance matching of the antenna.

Polarization

The plane in which the electric field varies is also known as the polarization plane. The basic patch covered so far is linearly polarized since the electric field varies in only one direction. This polarization can be anything between vertical and horizontal depending on the orientation of the patch. The polarization plane is the xz-plane in *Figure 1*. For optimum system performance, transmit and receive antennas must have the same polarization. The patch described above yields horizontal polarization and when rotated by 90°, the current flows in the vertical plane and the antenna is now vertically polarized.

A large number of applications like satellite communications, do not work well with linear polarization because the relative orientation of the antennas is unknown and because of Faraday rotation. In these applications, circular polarization is useful since it is not sensitive to antenna orientation. In a circularly polarized antenna, the electric field varies in two orthogonal planes (x and y direction) with the same magnitude and a 90° phase difference. The result is the simultaneous excitation of two modes, i.e. the TM10 mode (x direction) and the TM01 mode (y direction). One of the modes is excited with a 90° phase delay with respect to the other mode. A circularly polarized antenna can either be right-hand circular polarized (RHCP) or left-hand circular polarized (LHCP). The antenna is RHCP when the phases are 0° and -90° for the antenna in *Figure 5* when it radiates towards the reader, and it is LHCP when the phases are 0° and +90°.

To excite circular polarization in a patch we need to do three things:

- Split the signal in two equal parts.
- Feed one signal to the horizontal radiator (x axis) and the other to the vertical radiator (y axis). Each radiator behaves like a pair of radiating slots in the patch antenna as shown in *Figure 5*.
- Change the phase of one of the signals by 90°.

Splitting the signal in half can be done with a Wilkinson power divider or other splitter. If a square patch is fed with two feed points like in *Figure 5* and a 90° delay is added to one of the signal lines, a circularly polarized antenna is built.

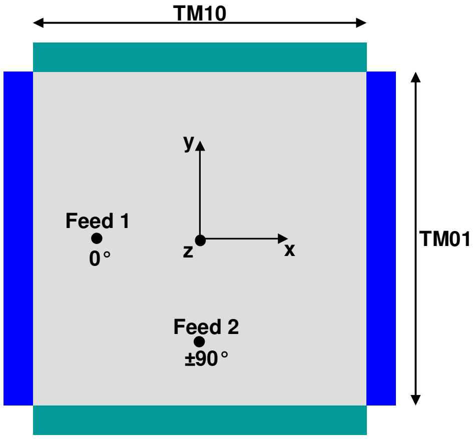

Figure 5: *The nearly square mechanism for circular polarization*

While this works well, the splitter and delay line take up valuable board space, introduce losses, tend to radiate and may degrade the radiation pattern.

A different approach is based on the fact that each patch mode (TM10, TM01) behaves like a parallel RLC resonant circuit. If we make the x and y dimensions of the patch slightly different, there will be two resonant frequencies, fa and fb and we will see two different RLC circuits. This creates a phase shift change versus frequency as shown in *Figure 6*:

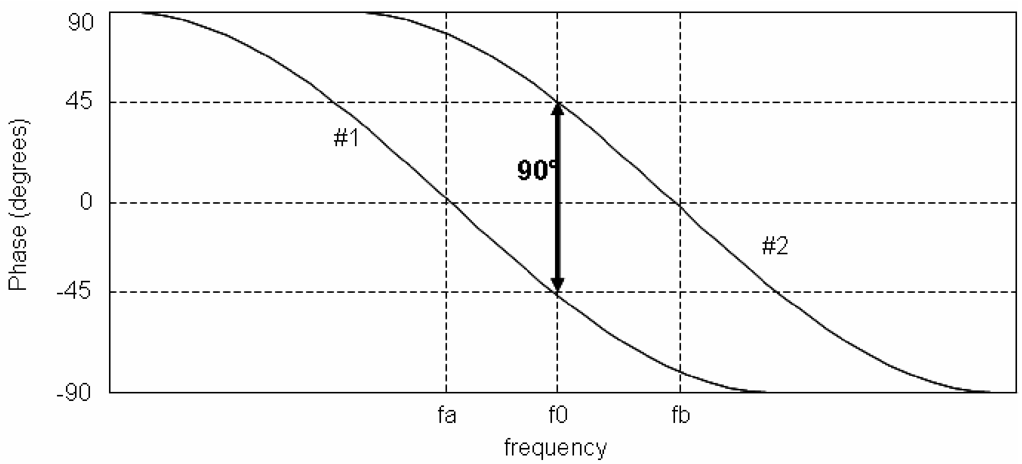

Figure 6: *The dual resonant circuit analogy of the nearly square mechanism*

If we pick these two resonance frequencies right, there will be a small frequency band where the phase difference between the two RLC circuits is 90°.

So, circular polarization can be achieved by building a patch with two resonance frequencies in the orthogonal directions and using the antenna right in between the two resonances at f0. The two modes must be excited with equal power and with a 90° phase difference.

One way to implement this is shown in *Figure 7*: this nearly square patch has slightly different lengths in the x and y axis. This causes the excitation of two orthogonal modes with the required 90° phase shift. Feeding the antenna close to the diagonal splits the power equally between the two orthogonal modes and all three requirements to generate circular polarization have been met.

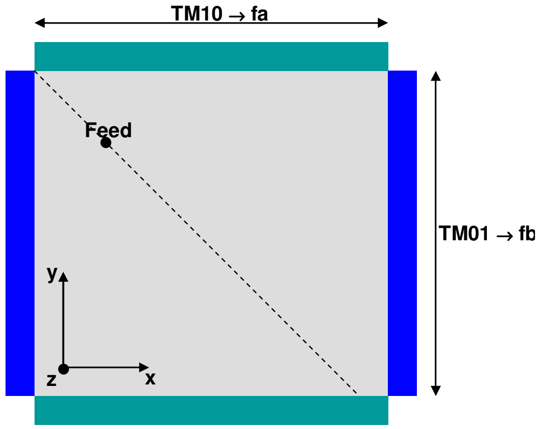

Figure 7: *Nearly square patch fed on the diagonal*

The corners truncated patch (often used in GPS L1 antennas), is another implementation of the same idea. This is shown in *Figure 8*.

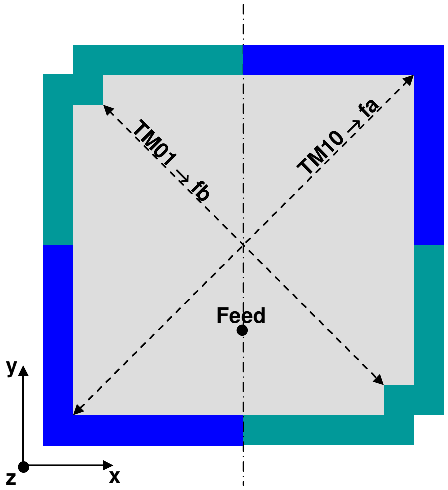

Figure 8: *The corners truncated patch antenna*

The nearly square and corners truncated patches have a smaller (1 to 3%) circular polarization bandwidth than the double fed patch (at least 3%): the polarization bandwidth of the antenna is mainly determined by the bandwidth of the splitter-phase shifter.

The quality of the circular polarization is commonly quantified as the axial ratio (AR) and is expressed in dB. A 3 dB axial ratio is considered sufficient for most applications. Since the phase difference between modes varies with frequency, it is clear that the axial ratio also varies with frequency: it has a theoretical optimum of 0 dB right in between the resonance frequencies of the two orthogonal modes. Not shown so far is that the axial ratio varies with elevation as well: the AR is optimal in broadside (z-axis), degrades towards lower elevations (the x-y plane) and the amount of degradation is highly dependent on the antenna geometry.

The quality of the circular polarization can also be expressed by showing the co- and cross-polar radiation patterns. The co-polar radiation pattern is the radiation pattern of the wanted polarization, and the cross-polar radiation pattern is the radiation pattern of the unwanted and opposite polarization. The co and cross polar behavior is often quantified using the co/cross polar ratio (20log(co/cross)) and expressed in dB.

The relationship between the axial ratio and the co/cross polar ratio is shown in *Figure 9*.

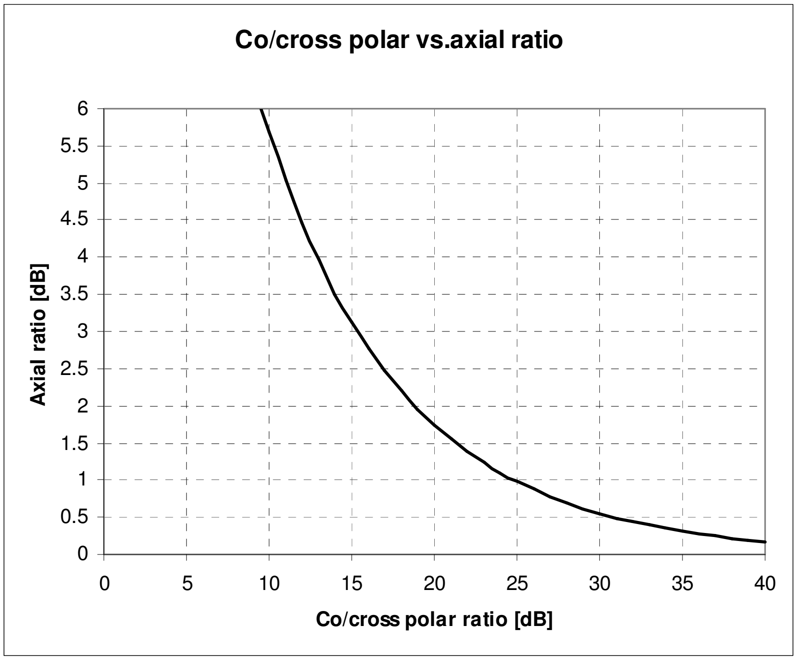

Figure 9: *Co/cross polar ratio versus axial ratio*

Bandwidth

Another important parameter of an antenna is the bandwidth it covers. Most of the time only impedance or return loss bandwidth is specified. However, it is important to realize that several other definitions of bandwidth exist: directivity bandwidth, polarization bandwidth and efficiency bandwidth. Directivity and efficiency are often combined to gain bandwidth.

Impedance/return loss bandwidth:

This is the frequency range over which the structure has a good impedance matching with respect to a certain reference impedance.

The impedance bandwidth depends on a large number of parameters related to the patch antenna element itself like quality factor Q and the type of feed used. The plot in *Figure 10* shows the return loss bandwidth of a patch antenna. The impedance bandwidth of a square, half wave patch antenna is typically limited to 1 to 3% and this is a major disadvantage of this type of patch antenna.

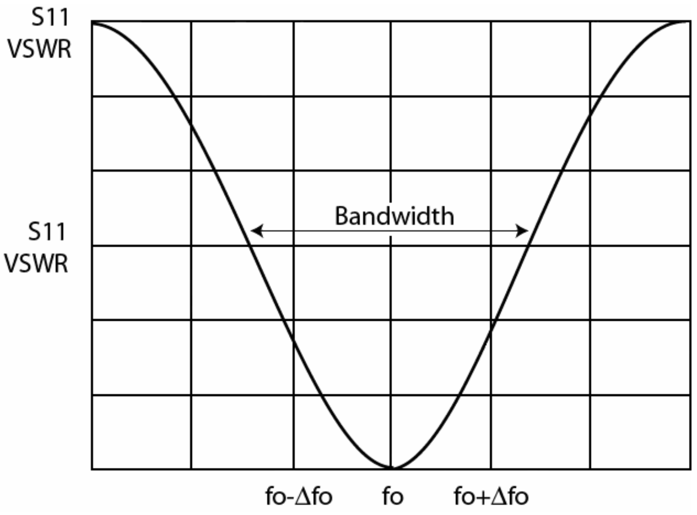

Figure 10: *Impedance bandwidth definition*

Keep in mind that impedance bandwidth can be defined in a couple of ways and that these do not forcibly mean the same thing: VSWR = 2:1, S11 <–10 dB or the maximum real impedance divided by the square root of two [Z(Re)/√2] bandwidth.

Directivity/gain bandwidth:

This is the frequency range where the antenna meets a certain directivity/gain requirement (like a 1 dB gain flatness).

Efficiency bandwidth:

This is the frequency range where the antenna has a defined efficiency.

Polarization bandwidth:
This is the frequency range where the antenna maintains a suitable co/cross ratio.

Axial ratio bandwidth:
This bandwidth is related to the polarization bandwidth and this number expresses the quality of the circular polarization of an antenna.

Alternative Feed Types

A classic way to feed a patch antenna is by using a coaxial probe or connector mounted at the appropriate impedance point and we will look at a couple of alternate ways to connect to a patch. Since in most patch antennas, the impedance bandwidth is much smaller than the radiation bandwidth, we will also look at some techniques to increase the bandwidth.

While the probe feed discussed so far has a bandwidth of only a couple %, it is still very useful for a lot of applications like GPS and WLAN.

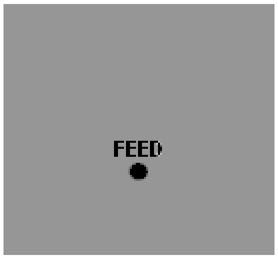

Figure 11: *Probe fed patch*

Another way to connect to the patch is a microstrip line at the edge of the element as shown in *Figure 12*.

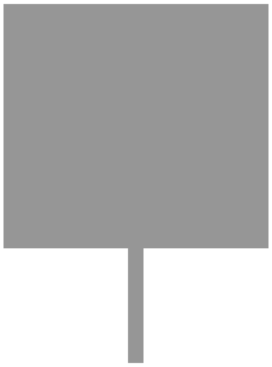

Figure 12: *Microstrip line fed patch antenna*

With this type of feed, impedance transformation to a useful value is needed as the impedance near the edge of the patch is quite high (a couple hundred Ω typical).

The advantage of this structure is the ability to place circuitry near the elements on the same substrate and avoid a multi-layer board. Also, elements can be connected in parallel in the same plane to form an array.

The impedance transformer itself tends to radiate and may distort the radiation pattern of the antenna.

A nice solution is a combination of the antennas from *Figures 11 and 12* and shown in *Figure 13*.

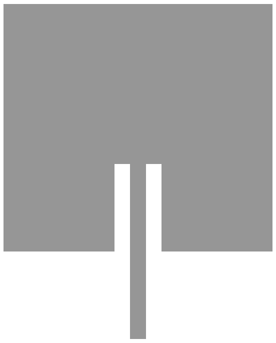

Figure 13: *Inset microstrip line fed patch antenna*

We keep the microstrip feed from *Figure 12* and create an inset to find the same impedance point from *Figure 11*. The gaps distort the radiation pattern of the element, but it is well worth the advantage of being able to connect a standard impedance circuitry straight to the element.

As discussed before, the impedance bandwidth of a basic patch is a couple % at best. The radiation bandwidth is usually much larger and can be up to 50%. Aperture coupling to a patch is a classic impedance bandwidth enhancement technique. The aperture coupled patch has a slot cut in the ground plane under the radiating element as shown in *Figures 14 and 15*. The slot is excited with a microstrip or stripline transmission line over the slot on the opposite side of the ground plane.

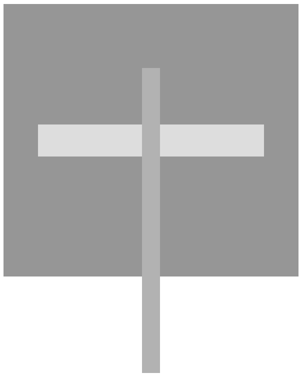

Figure 14: *Aperture coupled patch antenna*

The microstrip fed aperture coupled patch may suffer from a reduced front to back ratio. A stripline feed avoids this: the feed line is 'captured' between two ground layers. A cross section of a stripline fed, aperture coupled patch is shown in *Figure 15*. Impedance bandwidths of 10% can be readily achieved but this structure tends to be a bit more challenging to design.

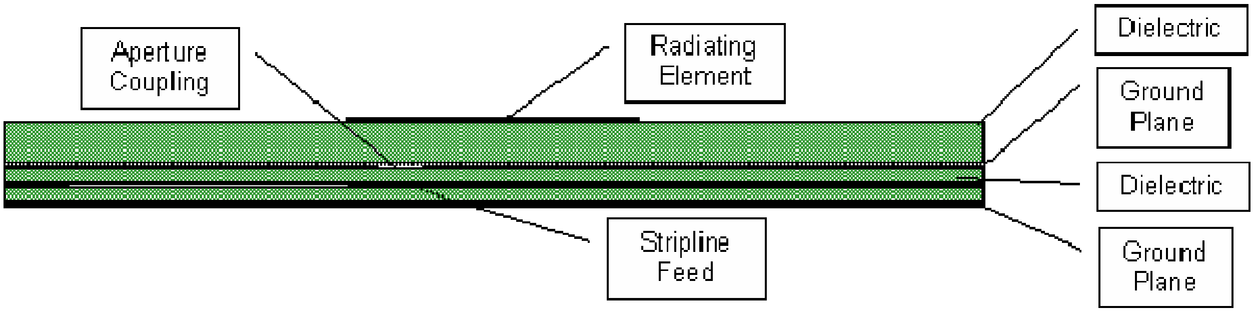

Figure 15: *Cross section of the aperture coupled patch antenna*

The Stacked Patch

Two patches can be vertically stacked and built like a multi layer printed circuit board. Vertically stacking two patches creates a dual band or a wide band patch antenna.

There are three ways to connect to this antenna:

- Connecting to both patches individually (dual band)
- Connecting to the upper patch only: series fed (dual band)
- Connecting to the lower patch only: shunt fed (wide band)

Connecting to both patches is a useful technique to create a dual band patch with separate feeds for each band. The required volume is considerable lower than when using two separate patch antennas for each band. The antenna itself acts as a diplexer, eliminating the need for an additional discrete diplexer. These features effectively reduce the overall system cost.

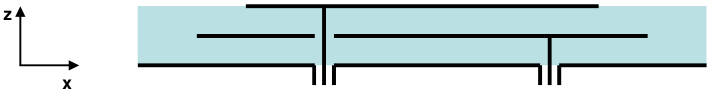

Figure 16: *Stacked patches with dual feed*

Feeding the upper patch only (series feed): this is a classic technique to create a dual band antenna with a single feed. Typically, the lower frequency band uses the lower (and larger) patch while the upper frequency band uses the upper (and smaller) patch. At the lower frequency, the leaky cavity between the lower patch and the ground plane resonates while the leaky cavity between upper patch and lower patch is electrically short circuited because it is off-resonance (the off-resonance impedance of patch antennas is effectively zero). At the higher frequency, the leaky cavity between the upper patch and lower patch resonates while the leaky cavity between lower patch and the ground plane is electrically short circuited because it is off-resonance. This is illustrated in the impedance versus frequency plot in *Figure 18* and the input reflection versus frequency plot in *Figure 19*. It is important that the upper patch is sufficiently smaller than the lower patch in order to avoid fringing field coupling between the two patches. This implies that the frequency bands should be separated by at least 5 to 10%.

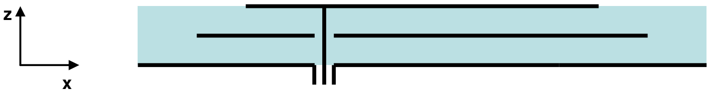

Figure 17: *Series fed stacked patches*

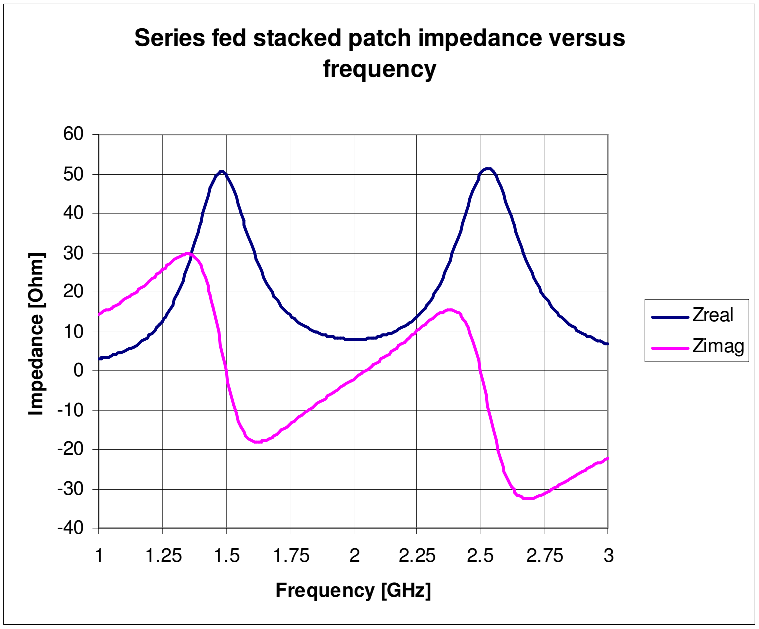

Figure 18: *Series fed stacked patch impedance behavior versus frequency*

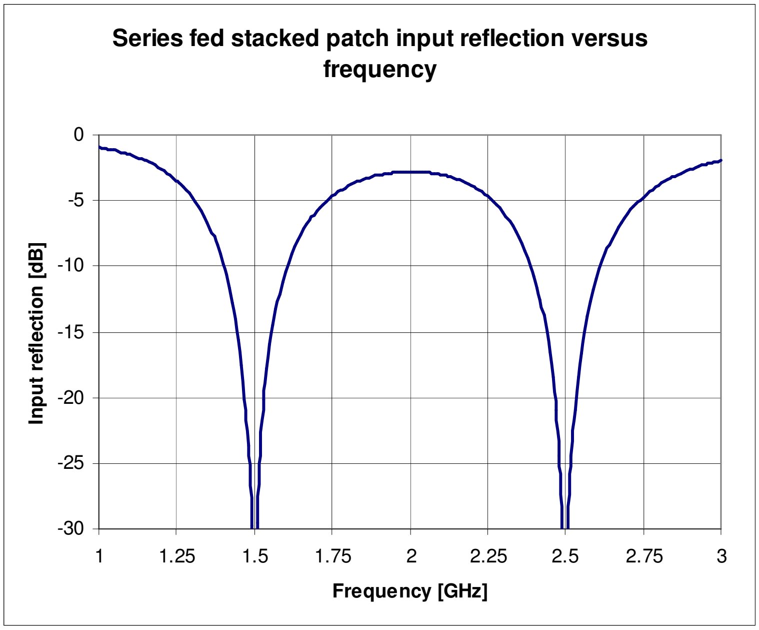

Figure 19: *Series fed stacked patch input reflection versus frequency behavior*

Feeding the lower patch only (shunt feed): this is a classic technique to create a wide band (3 to 10%) patch antenna where each of the two vertically stacked patches covers a sub-band. Dual frequency operation does not work very well with this configuration because the upper patch is electromagnetically coupled with the lower patch. This puts a constraint on the dimensions of the patches making this a wide band rather than dual band structure.

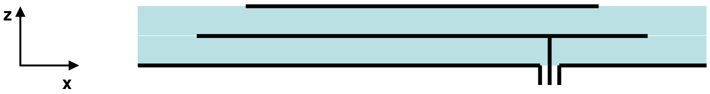

Figure 20: *Shunt fed stacked patches*

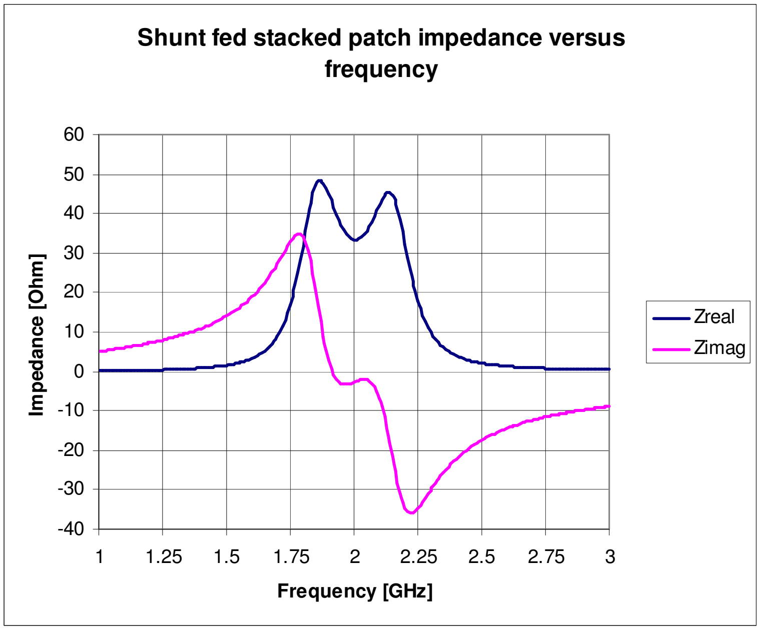

Figure 21: *Shunt fed stacked patch impedance behavior versus frequency*

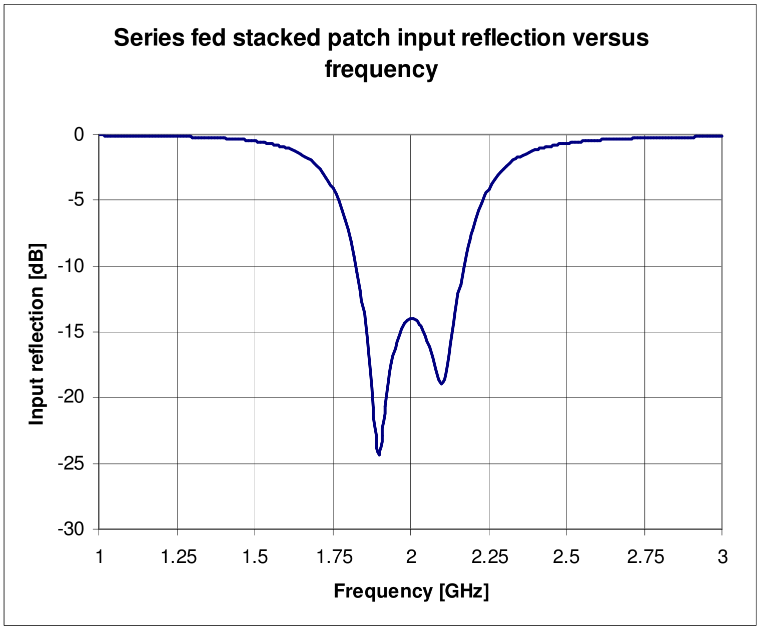

Figure 22: *Shunt fed stacked patch input reflection versus frequency behavior*

Series fed patches work very well (and only then) when the two bands are well separated and the antennas present an effective short at each other's bands of operation. Shunt fed patches work well (and only then) when both resonances are within a couple % of each other.

Higher Modes

So far, we've been talking about the fundamental TM10 mode. But, patch antennas are resonators and inherently exhibit higher modes.

Only a few TM modes are potentially useful and the other useful ones in a rectangular patch are TM20 and TM30. Other modes exist but result in exotic radiation patterns and have little -if any- practical use.

The radiation pattern cross sections of antennas operating in TM10, TM20 and TM30 modes have been plotted in *Figure 23*.

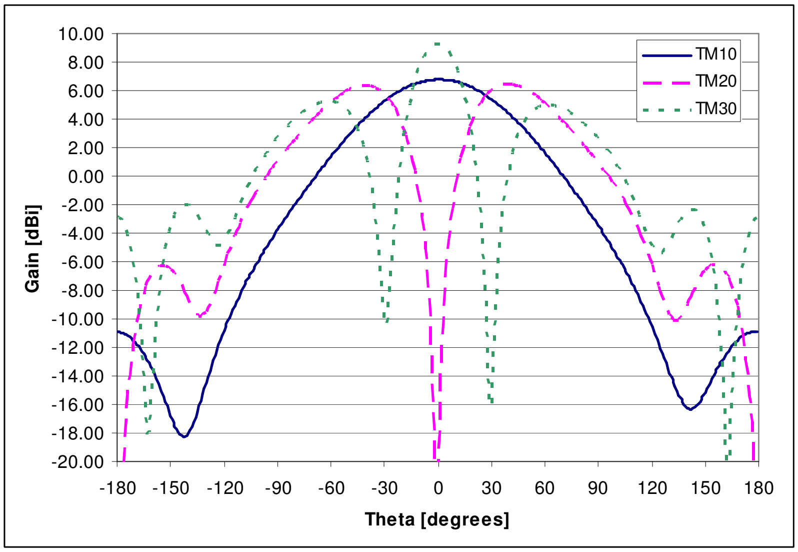

Figure 23: *Radiation patterns of higher rectangular patch modes*

This plot shows that the fundamental mode has one main lobe in broadside, the TM20 mode has two lobes symmetrically around broadside with a radiation null in broadside and the TM30 mode has three lobes: one in broadside and two symmetrically next to the broadside lobe.

These are for a rectangular patch only and differently shaped patch antennas exhibit other mode behavior and associated radiation patterns. Another excellent example is the round patch antenna shown in *Figure 24*:

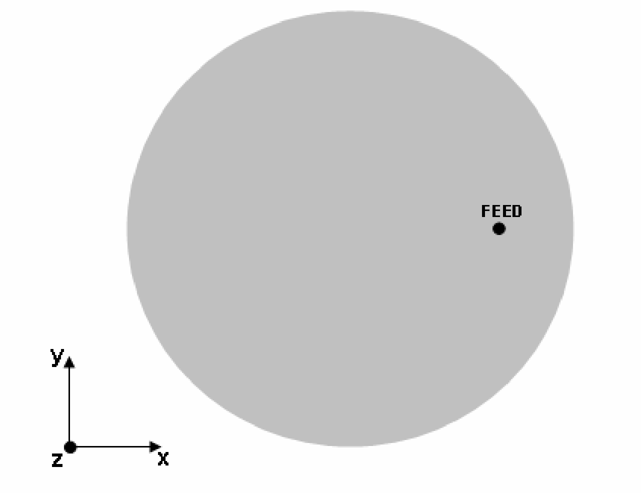

Figure 24: *Probe fed circular disk patch antenna*

The useful three modes of a round patch are TM11, TM21 and TM31. The radiation pattern cross sections are shown in *Figure 25*.

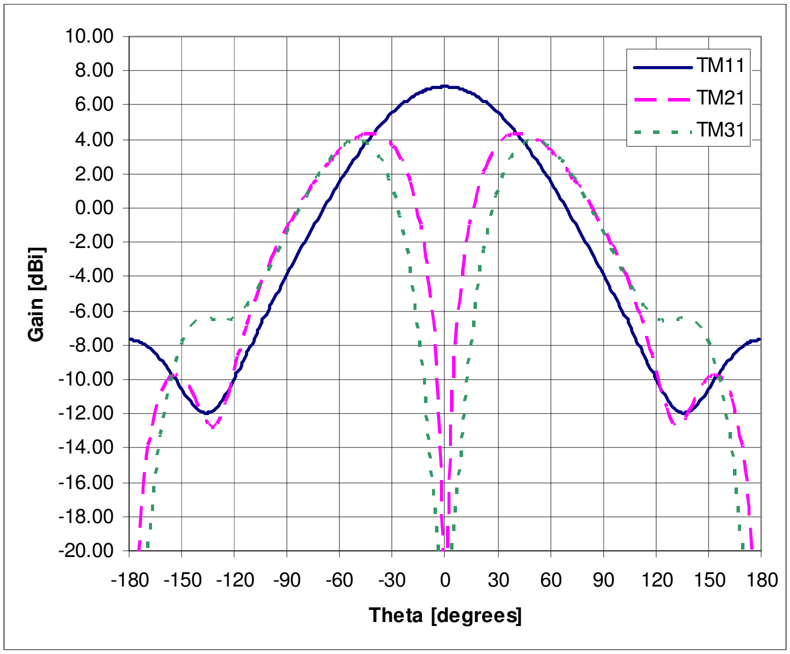

Figure 25: *Radiation patterns of higher circular disk patch modes*

The fundamental mode, TM11, radiates like a square patch operating in TM10. The second mode, TM21, radiates like a square patch operating in TM20. The third mode, TM31, radiates different from the rectangular patch: there is no broadside lobe.

Tuning higher modes tends to be a bit of a challenge relative to fundamental mode tuning because parasitic effects such as cross polar behavior and feed reactance are more problematic. The higher the mode, the more complex tuning is going to be.

Summary

In this article, some of the basic properties of linear and circular polarized patch antennas have been discussed. A basic set of specifications has been defined allowing the reader to understand and write a set of requirements for a specific application. A couple of patch antenna feeding techniques have been described. The stacked patch configuration has been introduced and the higher mode behavior of patch antennas has been explained.

Besides the patch antennas covered here, many more design options and different implementations are possible. Coverage of these alternatives is beyond the scope of this article, but they should be considered during the specification and development phases of the antenna.

If you have any questions, you can contact the authors at info@orbanmicrowave.com

June 19, 1945. **G. O. STRIKER** **2,378,663**

RADIO RECEIVER

Filed Jan. 16, 1942 5 Sheets-Sheet 1

Fig.1

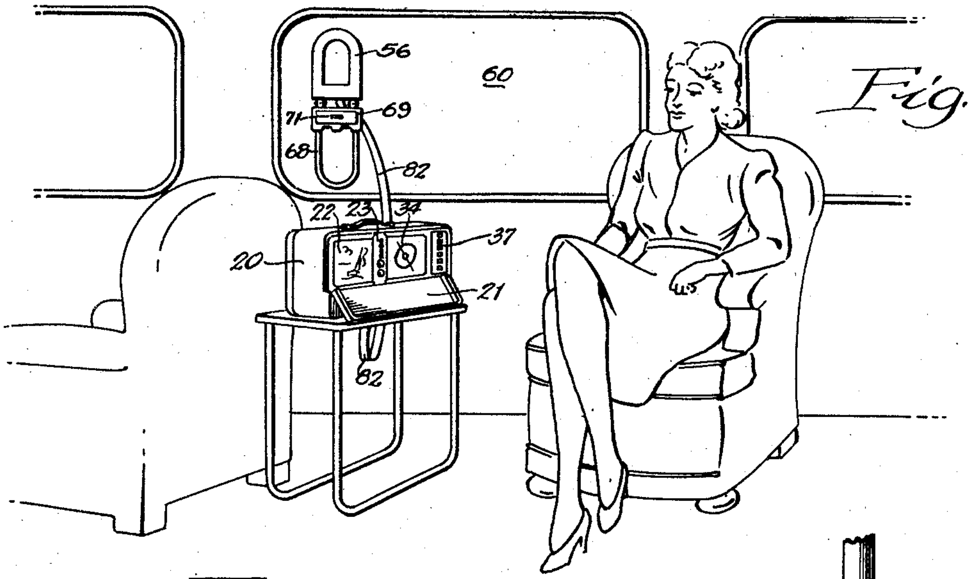

56
60
71
69
68
82
22
23
34
37
20
21
82

Fig.2

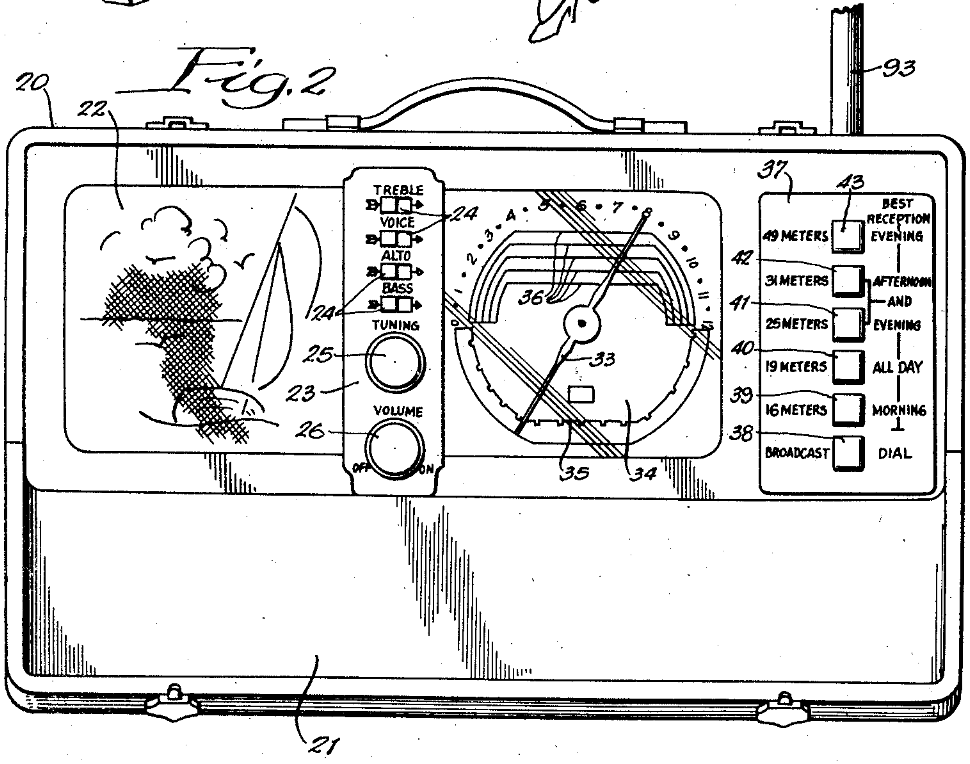

20
22
93
37
43
BEST RECEPTION
49 METERS
EVENING
TREBLE
VOICE
ALTO
BASS
24
24
24
TUNING
25
23
VOLUME
26
OFF
ON
0
1
2
3
4
5
6
7
8
9
10
11
12
36
33
42
31 METERS
AFTERNOON
AND
EVENING
41
25 METERS
40
19 METERS
ALL DAY
39
16 METERS
MORNING
38
BROADCAST
DIAL
35
34
21

Inventor:
George O. Striker,
By McCaleb, Wendt & Dickinson
Attorneys.

June 19, 1945. G. O. STRIKER **2,378,663**

RADIO RECEIVER

Filed Jan. 16, 1942 5 Sheets-Sheet 2

Fig. 3

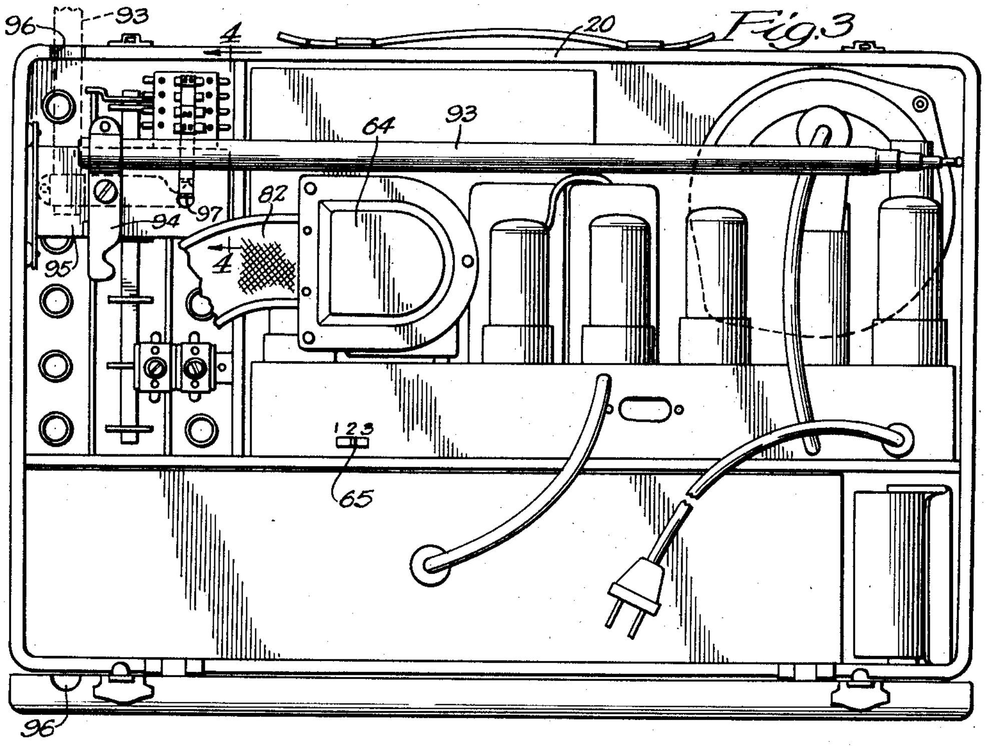
96
93
4
20
64
93
82
97
94
95
4
1 2 3
65
96

Fig. 4

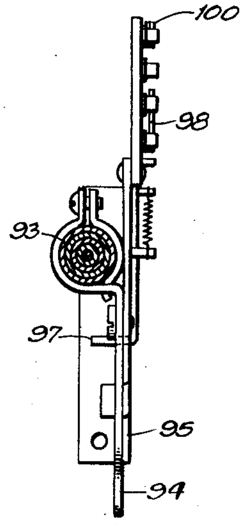
100
98
93
97
95
94

Fig. 5

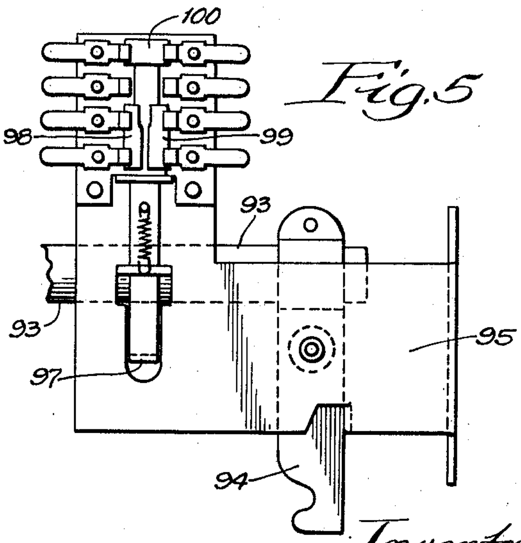
100
98
99
93
93
97
95
94

Inventor:
George O. Striker,
By McCaleb, Wendt & Dickinson
Attorneys.

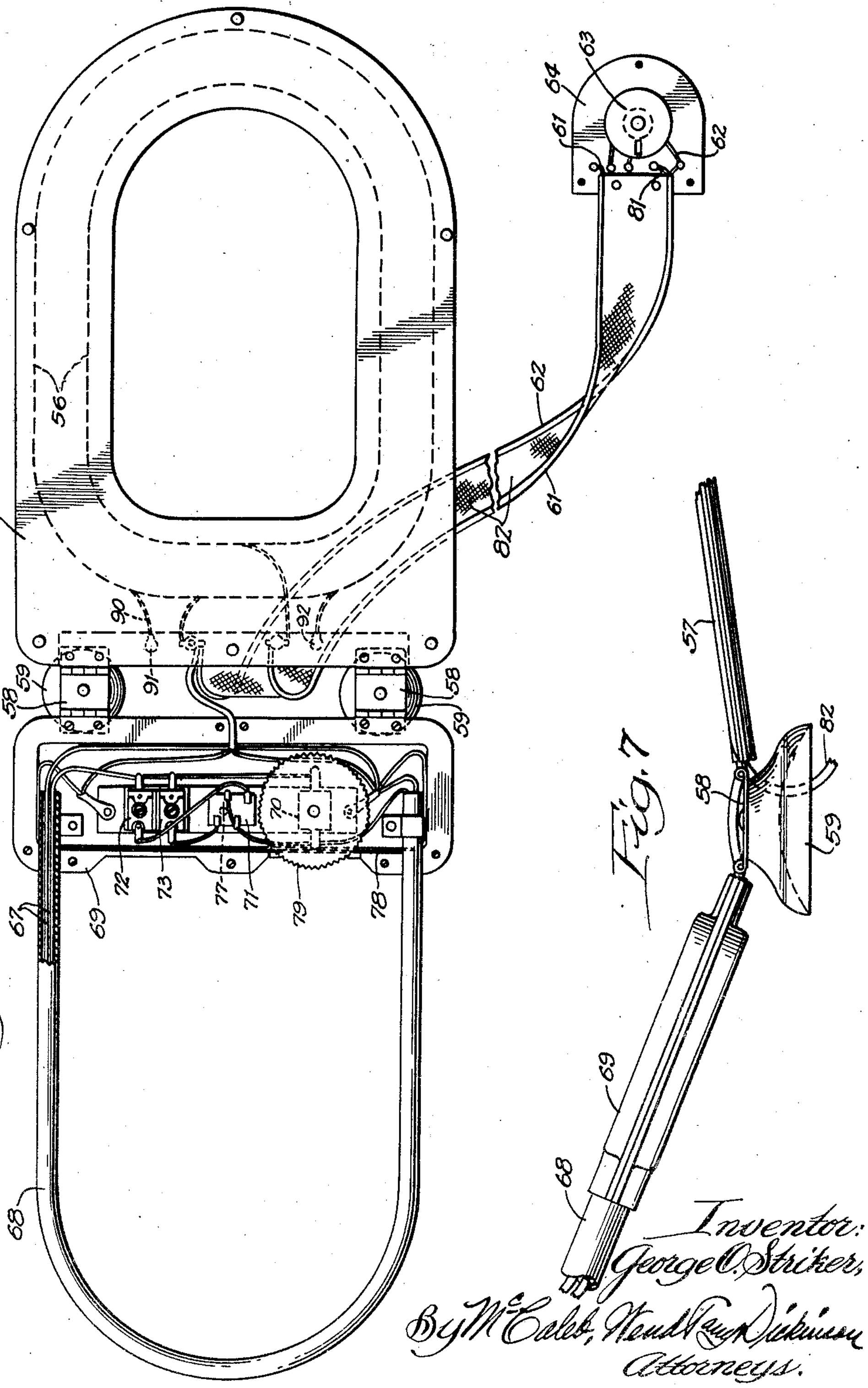
63
64
61
62
81
62
61
82
56
57
90
92
91
59
58
58
59
70
72
73
77
71
79
78
69
67
68

Fig. 6

57
82
58
59
69
68

Fig. 7

Inventor:
George O. Striker,
By McCaleb, Wendt Tenney Dickinson
Attorneys.

June 19, 1945. G. O. STRIKER 2,378,663

RADIO RECEIVER

Filed Jan. 16, 1942 5 Sheets-Sheet 4

Fig. 8

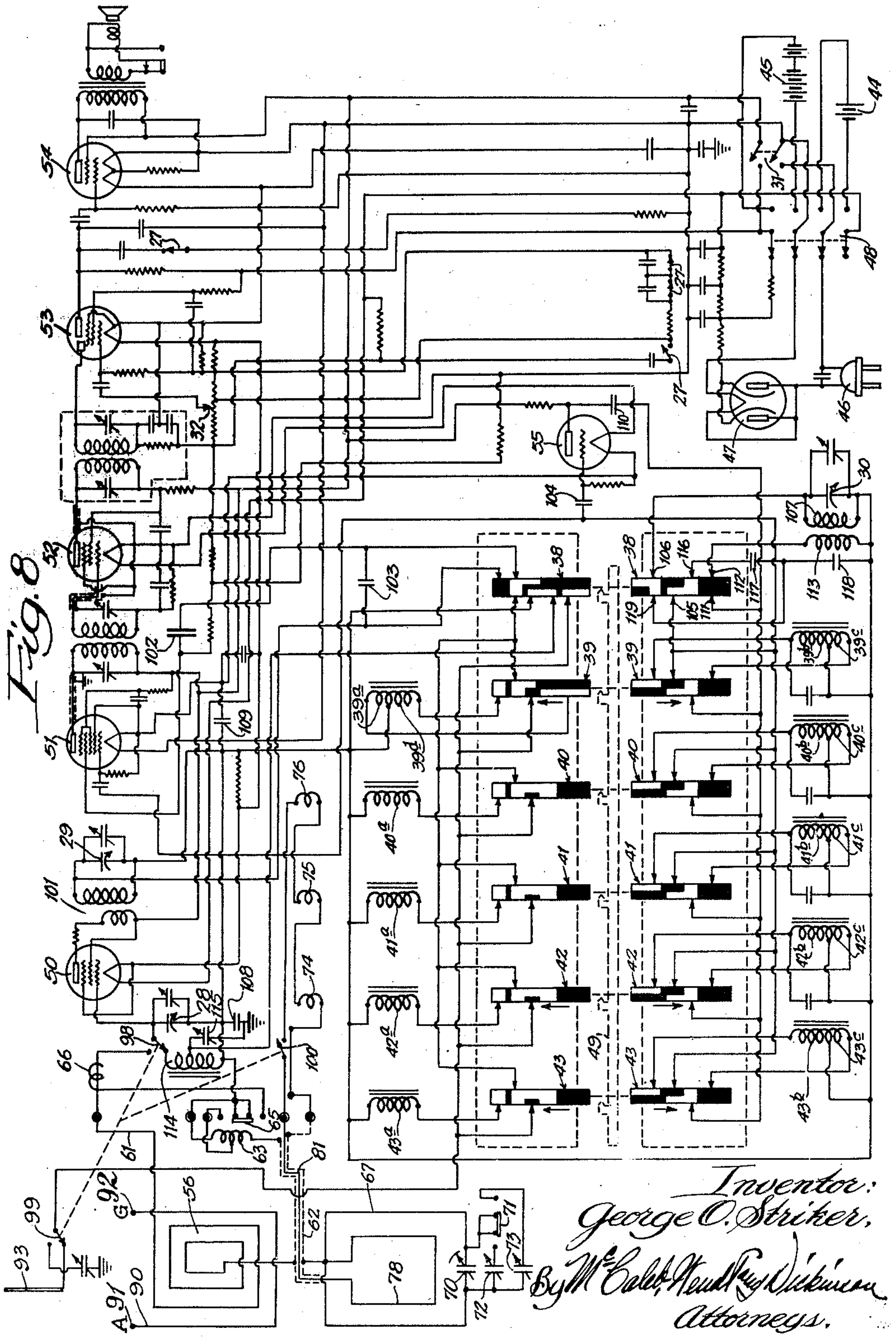
54
53
52
51
50
27
27
27
32
102
109
29
101
28
115
108
98
66
114
65
63
61
99
92
G
93
91
A
90
56
62
81
67
78
70
72
73
71
100
74
75
76
39a
39d
40a
41a
42a
43a
103
38
39
40
41
42
43
49
38
39
40
41
42
43
55
104
110
45
44
31
48
46
47
30
107
106
116
112
117
113
118
119
105
111
39b
39c
40b
40c
41b
41c
42b
42c
43b
43c

Inventor:
George O. Striker,
By McCaleb, Hendley & Dickinson
Attorneys.

June 19, 1945.　　G. O. STRIKER　　2,378,663

RADIO RECEIVER

Filed Jan. 16, 1942　　5 Sheets-Sheet 5

Fig. 11

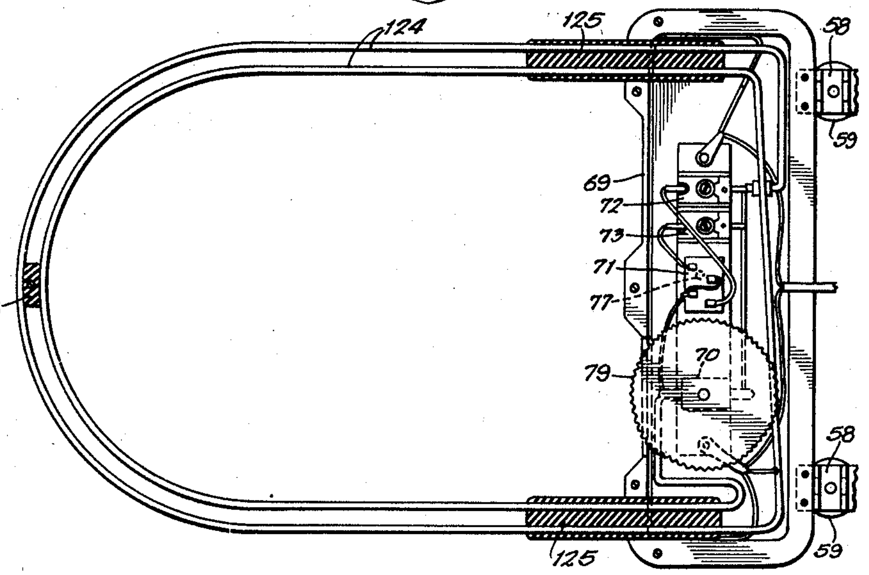
124
125
58
59
69
72
73
71
77
126
79
70
58
59
125

Fig. 9

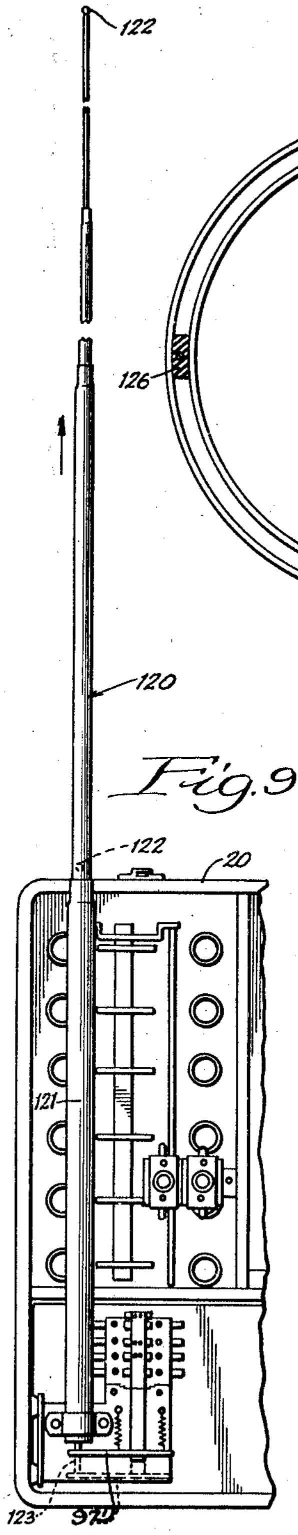
122
120
122
20
121
123
97

Fig. 10

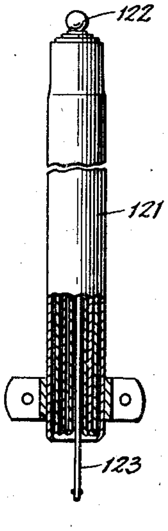

Inventor:
George O. Striker,
By McCaleb, Wendt & Dickinson
Attorneys.

UNITED STATES PATENT OFFICE

2,378,663

RADIO RECEIVER

George O. Striker, Chicago, Ill., assignor to Zenith Radio Corporation, Chicago, Ill., a corporation of Illinois

Application January 16, 1942, Serial No. 426,991

43 Claims. (Cl. 250—20)

This invention relates to radio receivers and particularly to receivers adapted to operate on a plurality of bands.

One of the objects of the invention is to provide a portable radio receiver which is provided with a plurality of antennae suitable for different wave bands and for use under different circumstances with relation to different locations and different conditions as to man-made static and other forms of interference.

A further object of the invention is to provide a radio receiver with a short wave loop antenna which may be located at positions relatively remote from the receiver.

A further object of the invention is to provide means for tuning the loop to adapt it for the reception of signals in different bands.

A further object of the invention is to provide a radio receiver with a plurality of antennae with means whereby the preparation of one antenna for use renders another antenna inoperative.

A further object of the invention is to provide a radio receiver having a fish-pole antenna and switch means controlled thereby connecting and disconnecting said antenna to the input of the receiver and for connecting and disconnecting another antenna from the input of the receiver.

Other objects, advantages and capabilities of the invention will appear from the following description of a preferred embodiment thereof, taken in conjunction with the accompanying drawings, in which:

Figure 1 is a perspective view showing the manner in which the portable radio receiver may be used by a railroad passenger;

Fig. 2 is a front elevational view of the receiver, the front cover being opened;

Fig. 3 is a rear elevational view, the rear cover being removed to show internal parts;

Fig. 4 is a sectional detail taken on the line 4—4 of Fig. 3 showing the telescoped fish-pole antenna in cross-section and associated elements in elevation;

Fig. 5 is a front view of the mechanism shown in Fig. 4;

Fig. 6 is an elevational view of the two-loop antenna structure, a cover of the short wave loop being removed and an associated part thereof being shown in section;

Fig. 7 is a fragmentary plan view thereof on a somewhat larger scale;

Fig. 8 is a wiring diagram of the receiver;

Fig. 9 is a fragmentary rear view of the receiver with the rear cover removed, showing a modified form of fish-pole antenna;

Fig. 10 is an elevational view partly in section of such fish-pole antenna in collapsed form; and

Fig. 11 is a view similar to the left-hand side of Fig. 6 showing a modified form of short wave antenna.

Referring to the drawings, the radio receiver as a whole is contained within a case **20** which is provided with a hinged front cover **21**. When the front cover **21** is opened, music and other broadcast material may be emitted from the loud speaker **22**.

Adjacent the loud speaker **22** is a panel **23** which serves as a mounting for tone controls **24**, a tuning control **25**, and a volume control **26**. The tone controls **24** open and close the four switches **27** in Fig. 8 to effect control of the tone. The tuning control **25** controls a gang of condensers **28**, **29** and **30** for tuning the RF input circuit, mixer input circuit, and oscillator circuit, respectively (Fig. 8).

The volume control **26** controls the double-pole single-throw switch **31** (Fig. 8), and the volume control **32** also shown in that figure. The tuning control actuates not only the gang of condensers but also the pointer **33** of the dial **34**.

The dial **34** comprises indicia **35** corresponding to the broadcast band and five sets of indicia **36** which correspond to five short wave bands, for example, the 16-meter, the 19-meter, the 25-meter, the 31-meter, and the 49-meter bands.

Adjacent the dial **34** is an escutcheon plate **37** which carries a series of projecting switch buttons **38**, **39**, **40**, **41**, **42** and **43**, corresponding to the broadcast bands just referred to.

As best seen in Fig. 2, the escutcheon plate **37** bears opposite these buttons the name of the corresponding band and also indicia adjacent the buttons indicating when the best reception is received with that band. Thus, the best reception of the 49-meter band is in the evening, with the 31- and 25-meter bands in the afternoon and evening, the 19-meter band all day, and the 16-meter band in the morning.

The double-pole single-throw switch **31** controls the supply of electrical energy to the tubes of the receiver. This energy may be supplied by A and B batteries **44** and **45** (Fig. 8) carried in the receiver case or it may be supplied by the plug **46** which may be inserted in any available wall receptacle, the power pack associated therewith being arranged for A. C. and D. C. operation. This power pack includes a rectifier tube **47** and the filter instrumentalities illustrated in Fig. 8.

The receiver may be changed over from battery to power line operation by means of a four-pole double-throw switch **48** shown in power line position in Fig. 8.

The band switch buttons **38** to **43** constitute the outer ends of axially movable switch bars bearing the same numbers and shown in Fig. 8. In this figure two views of each switch bar are shown in relatively inverted relation. Thus, if one of the switch bars were actuated in Fig. 8, then the upper showing of that bar would be displaced upwardly while the lower showing of that bar would be displaced downwardly. The black parts of the switch bars represent insulation and the white parts conducting material.

The bars **38** to **43** are associated with a locking bar **49** of known type shown diagrammatically in Fig. 8. This locking bar is so arranged that it holds any bar which is displaced inwardly in its inward switching position. The bar is so arranged that it is actuated to releasing position by the inward movement of another bar. Consequently, only one of the band switches **38** to **43** is operative at any time.

The receiver may suitably be a superheterodyne receiver, tube **50** being a radio frequency amplifier, tube **51**, the mixer, tube **52**, an intermediate frequency amplifier, tube **53**, a detector and audio amplifier, tube **54**, a power amplifier, and tube **55**, the oscillator tube.

The receiver is provided with a plurality of antenna structures which will now be described. For broadcast pickup, a loop antenna **56** of the type described and claimed in Patents Nos. 2,164,251, 2,200,674, 2,237,260, and 2,250,387, issued to Eugene F. McDonald, Jr., may be employed.

The loop **56** is enclosed within a suitable nonconductive enclosure **57** which is hinged to plate members **58**. These plate members **58** carry suction cups **59** whereby the loop may be mounted on the glass window **60** of a conveyance. This conveyance may, for example, be a steel railroad car, a bus, an airplane, or an automobile. Of course, the invention is not intended to be limited to a conveyance since its antenna may be used to great advantage in a steel or metallized house, or it may be used irrespective of any enclosure whatsoever. Normally, the suction cups **59** do provide a very simple quickly attachable and detachable means for supporting the loop **56** on or near a dielectric wall such as a pane of glass in an enclosure which is constituted for the most part by walls of conducting material. It is within the scope of my invention to provide one or more suction cups to support loop **56**.

The ends of loop **56** are connected to two leads **61** and **62**, lead **62** appearing in Fig. 8 as the outer conductor of a coaxial cable having a center conductor **81**. For the purposes of the broadcast loop **56**, it is important that the distributed capacity between the leads **61** and **62** should be maintained low and should be maintained substantially constant. For this reason the conductor **81** and its shield **62** are located along one edge of a stout ribbon **82** and the conductor **61** is located along the opposite edge of this ribbon. This ribbon is relatively broad and stiff so that the spacing between the conductors at the opposite sides is maintained substantially constant. The latter lead is connected to a small loading coil **63** on a five-prong plug **64**. The five prongs are shown as ringed dots in the upper left-hand corner of Fig. 8. The conductor **61** is connected to one prong. The coil **63** is connected to another. A tap on this coil is connected to still another prong.

On the receiver is mounted a three-position switch **65** which in one position shorts the upper part of the coil **63** and in another position introduces a small capacity **66** across the leads **61** and **62**. In its intermediate position the switch **65** has no effect on the inductance or capacity of the circuit including the loading coil **63** and the loop **56**. The purpose of the switch **65** is to introduce a slight correction for improving the tuning of the loop **56** which may be impaired by mounting the loop too near the edge of a steel window, for example, or from some other cause. The capacity **66** may suitably be constituted by a few turns of wire around the conductor leading to conductor **61** as is illustarted in Fig. 8.

The housing **57** also contains a coil **90** of a few turns in close proximity to the loop **56**, the ends of the coil **90** being connected to terminals **91** and **92**. In the case of a semi-permanent installation, for example in the case of a sojourn at a country home, it may be desirable to operate with an outside aerial which may be connected to the terminal **91**, the terminal **92** being grounded. It will thus be seen that the coil **90** is coupled to the loop **56** and that the operation for broadcast reception with an outside aerial is attained in a remarkably simpler manner.

For short wave reception the loop antenna **56** is no longer suitable. An antenna having a low inductance and capacity is required. In accordance with the present invention, I provide a short wave loop antenna **67** which consists essentially of a few turns of wire extending around a U-shaped dielectric tube **68**. The tube **68** has its ends mounted in a housing **69** which is hinged to the plates **58** as best seen in Fig. 6.

If the loop **67** were not tuned separately from the receiver and were located at a point remote from the receiver, the inductance and capacitance of the connections between the loop and receiver, especially at high frequencies, would seriously mistune the loop. In fact, such a loop with its connections to the receiver would probably be resonant at a relatively low frequency. By providing a small degree of coupling between loop **67** and the receiver input circuit it is possible to eliminate the mistuning effect of the connections sufficiently that the loop **67** may be tuned substantially independently of the receiver even at high frequencies. In accordance with this invention, adjustable tuning means is provided at the loop to tune the loop to any desired frequency. Such an arrangement is shown in Fig. 8. Opposite ends of the loop **67** are connected across a condenser **70** which is arranged to tune the loop to a desired frequency within one of the short wave bands. A three-position switch **71** is arranged to cut in a condenser **72** or a condenser **73** into parallel connection with the condenser **70** so as to tune the loop for two other short wave bands of longer wave length.

When using a short wave loop such as loop **67** it is even more desirable than with the broadcast loop **56** to provide a relatively long transmission line for connecting the loop remotely to the receiver input so that the loop may be placed in a position removed from the receiver. This is so because at high frequency the conductors of the receiver itself if too near the loop, shield or damp the loop even more so than at lower frequencies and render it less effective to transmit a signal to the receiver input. At the same time, the use of such a line becomes more difficult at high frequencies because of radiation loss from and radiation pickup on the transmission line itself. Also, the problems of shunt and stray capacity of the transmission line, discussed in connection with the broadcast loop 56, become more troublesome. For the purpose of transmitting signals intercepted by the short wave loop 67 and transferring the same to the receiver located remotely from the antenna, and more specifically for transferring intercepted signals to a control grid of the mixer tube 51, there is provided a link circuit extending in space from the loop 67 to the receiver and comprising a first coupling coil 78 coupled to the loop 67, a pair of conductors 62 and 81, a switch 100, and three coupling coils 74, 75 and 76, all connected in series relation. The coil 78 suitably comprises a single turn of wire positioned closely to the loop antenna 67 and preferably has a common point therewith. The loop antenna 67 and the coil 78 thus form a step down transformer.

The switch 100 is closed when using the receiver with the short wave antenna 67 and is actuated by the fishpole antenna 93 as will be explained hereinafter. The coils 74, 75 and 76 are located at the receiver and coupled to coils 42*a*, 41*a* and 40*a*, respectively, which can be connected in circuit with the tank circuit comprising the secondary winding of transformer 101 and variable condenser 29 for the purpose of applying the signal intercepted by the antenna 67 to a control grid of the mixer tube 51. The variable condenser 29 tunes the tuned circuit 101, 29 to the same signal frequency as that to which the loop 67 is tuned. Tuning of the receiver input as well as the loop increases the selectivity of the set in an amount proportional to the product of the quality factors of the two tuned circuits. This is desirable for best operation of a superheterodyne receiver.

Since the coils 74, 75 and 76 each have a smaller number of turns than the coils 42*a*, 41*a* and 40*a*, respectively coupled thereto, there is an upwards transformation of voltage therebetween. The voltage transformations, down at the antenna and up at the receiver at the opposite ends of the link circuit results in good energy transfer and yet permit the transmission line to operate at a lower voltage thereby reducing the effect of the shunt and stray capacity of the line. Furthermore, this transformation produces a better impedance match between the high impedance tuned antenna circuit and the untuned transmission line, and between the line and the high impedance tuned input circuit of the receiver. In the specific embodiment of the invention shown, the loop antenna 67 is arranged to operate on the 31-meter, the 25-meter, and the 19-meter bands, these being the bands controlled by the switch bars 40, 41 and 42, respectively. It will, however, be understood that any desired number of condensers may be provided so as to be brought into parallel relation with the condenser 70 and the windings 74, 75 and 76 should be likewise increased in number.

Thus, I may provide, instead of the switch 71, a five-position switch, and provide two condensers in addition to 72 and 73, arranged to be connected in parallel with the condenser 70 by the five-position switch. In addition, two windings in series with the windings 74, 75 and 76, would be provided in the same relation with the coils associated with the bars 39 and 43 as the coils 74, 75 and 76 bear to the coils associated with the bars 42, 41, and 40, respectively.

The switch 71 is mounted on the housing 69 and is provided with a projecting button 77 (Fig. 6). It will, of course, be understood that suitable indicia (not shown) is associated with the button 77 so as to indicate the particular short wave band for which the loop 67 is tuned by the switch 71.

The condensers 72 and 73 are adjustable trimmer condensers of appropriate value. The condenser 70 is preferably variable by means of a manually operable member 79. The manually operable member may consist, for example, of a small rough edged wheel of insulating material, the center of which is operatively connected to the condenser 70.

The edge of the wheel 79 projects out of the housing 69 so that it can be actuated manually to provide a slight adjustment of the condenser 70. This slight adjustment corrects for variable capacity effects introduced by variable mounting of the loop relative to conductors such as the body of a steel car.

The connection of the loop 67 and the coil 78 of the link circuit to the receiver will now be described in detail. The connections comprise a common conductor 62 which is a shield. This shield extends to the plug 64 where it is connected, as previously indicated, to one end of the loading coil 63 and to the fifth prong of the plug. Within this shield is a conductor 81 which is connected to the other end of the coil 78. This conductor is connected to the fourth prong of the plug 64. The shield 62 and conductor 81 form a coaxial line. Of course two separate wires may be used, if desired, but the shielded line provides one of the simplest means of achieving the desired characteristics of the transmission line, namely, low signal radiation loss and stray radiation pick up, and relatively low shunt capacity. Of importance, for obtaining suitable signal pickup in a shielded structure, is the tuned loop antenna winding 67 which is coupled through an untuned link circuit to the input circuit of a receiver tuned to the same frequency as is the loop antenna.

The composite link circuit includes the first coil 78 on one end of the conductors 62, 81 and on the other end the serially connected coils 74, 75 and 76 which, for purposes of analysis, may be grouped and considered as a second coil serially connected with said first coil 78 through the elongated conductors 62, 81 comprising, preferably, the co-axial cable, as shown in Figure 6.

Coil 78 is mounted near to and coupled to antenna 67 directly through transformer action, the winding of antenna 67 serving as one element of the transformer, and coil 78 serving as the other element of the transformer. This type of coupling is preferred, inasmuch as a minimum number of circuit elements is necessary and efficiency of energy conversion is high.

Of importance in short-wave reception, is that tuning of the antenna 67 and tuning of the input circuit of the receiver results in high selectivity and high image ratio. This is especially desirable at high frequencies.

Furthermore, the particular arrangement of the link circuit with the tuned loop antenna and tuned receiver input circuit results in a decrease in both the inductance at the loop antenna and at the input circuit of the receiver, in accordance with the principles set forth in "Radio Engineer's Hand Book," by Frederick E. Terman, 1943, published by McGraw-Hill Book Company.

Furthermore, inasmuch as the over-all co-efficient of coupling between the tuned loop antenna 67 and the tuned input circuit of the receiver is quite small, the tuned loop antenna 67 is coupled in small degree to the tuned input circuit, as is evidenced by the constructions shown in

Figs. 6, 8 and 11. It is remembered that the co-efficient of coupling between two circuits is a ratio and is defined as the ratio of the mutual inductance actually present between two windings to the maximum possible value that can occur. In accordance with this accepted definition of the co-efficient of coupling, the magnitude of the coupling between loop antenna **67** and adjacent link circuit coil **78** in Figs. 6 and 11 is in the range of .15 to .25.

The co-efficient of coupling between the tuned input circuit of the receiver and the adjacent link circuit coil, comprising, in fact, three coils **74**, **75** and **76**, has as its maximum possible theoretical magnitude one-third or .33, inasmuch as the flux from only one of such coils **74**, **75** and **76** threads the magnetically coupled input circuit of the receiver. As shown by Fig. 8, the secondary coils **40A**, **41A**, and **42A** are, respectively, close to their associated iron cores, and primary coils **76**, **75** and **74** are spaced a substantial distance from the corresponding secondary coils and cores. Therefore, the co-efficient of coupling between each of the coils **76**, **75** and **74** and its associated secondary coil is substantially less than unity, with the result that the co-efficient of coupling between the link circuit coil comprising serially connected coils **74**, **75** and **76** is substantially less than the possible theoretical maximum magnitude of .33.

In accordance with a commonly accepted definition of "loose" and "tight" coupling, as set forth in the above-mentioned "Radio Engineer's Hand Book," these recited magnitudes of the various coupling co-efficients of .15 to .25 and .33 neither correspond to "tight" coupling nor to "loose" coupling, inasmuch as such publication states that closely coupled coils usually have a coupling co-efficient of .5 or greater, and loosely coupled coils have a coupling co-efficient of .01 or less. Thus, these two magnitudes of .15 to .25 and .33 fall within a range which, for purposes of this specification, corresponds to a state of medium coupling.

The short-wave loop antenna shown herein is thus not tightly coupled to the input circuit of the receiver upon consideration that the over-all coupling is the product of the individual coupling co-efficients and which has a maximum product of .25×.33 or .0825. Of course, this actual magnitude of .0825 for the over-all coupling co-efficient is the maximum theoretical magnitude and, for the reasons explained, is smaller in actual physical receivers embodying apparatus having the precise arrangement of elements shown herein.

This small degree of coupling, which gives efficient energy transfer consistent with high selectivity and good image ratio, is desirable not only for the reasons stated above including making it possible to tune the loop **67** substantially independently of the receiver but also because it makes possible the use of a critical amount of coupling which by definition is the condition for maximum energy transfer. In accordance with the present invention, I also provide a fish-pole or extensible rod antenna **93** which is normally housed in telescoped and unerected position, as shown in Fig. 3, the fish-pole antenna being then located in the rear of the housing. To render the fish-pole antenna operative, the rear door of the receiver housing is opened and the telescoped fish-pole antenna **93** is swung into vertical position, that is, the position in which it is shown in dotted lines in Fig. 3. For this purpose the lowermost section of the fish-pole antenna **93** is rigidly carried on a lever **94**.

The lever **94** is pivotally mounted on a bracket **95** secured to one side of the case **20** of the receiver. When the fish-pole antenna **93** is erected, its lowermost section is brought into registry with two semi-circular openings **96** in the upper wall of the case **20** and on the upper edge of the rear door thereof so that this door may now be closed, leaving the fish-pole antenna projecting upwardly from the case in the vertical direction. It will, of course, be understood that the fish-pole antenna may be pulled out or extended in normal manner, illustrated in Fig. 9, with respect to a modified arrangement of fish-pole antenna.

Referring to Figs. 4 and 5, it will be readily understood that when the lever **94** is swung by the movement of the fish-pole antenna **93** into erected position, one end of the lever **94** engages a projection **97** and moves it upwardly, thus displacing contacts **98**, **99** and **100** upwardly. These contacts are elements of the three switches **98**, **99** and **100** shown in Fig. 8. These switches are shown in Fig. 8 in the condition in which they are after the fish-pole antenna has been erected. It will be understood that when the fish-pole antenna is again unerected or turned into horizontal position, the switches **98** and **99** assume their different positions and the switch **100** becomes closed.

In the embodiment of the invention shown in Figs. 9 and 10, a different fish-pole antenna **120** is employed. This fish-pole antenna has a lower section **121** which is fixedly mounted within the case **20** of the receiver. As shown in Figs. 9 and 10, the fish-pole antenna comprises a plurality of sections which can be telescoped together. When fully telescoped, the knob **122** of the fish-pole antenna projects from the upper side of the case **20** as shown in dotted lines in Fig. 9. The fish-pole antenna may, of course, be extended fully or to a desirable degree by merely grasping the knob **122** and pulling it upwardly.

The innermost section of the fish-pole antenna consists of a rod **123** which is somewhat longer than the other sections of the fish-pole antenna. Consequently, when the fish-pole antenna is fully telescoped back into the case **20**, the rod **123** projects beyond and below the lowermost section **121**. In this embodiment of the invention the switch elements **98**, **99** and **100** are adapted to be actuated by the movement of a projection **97'**, this projection having a portion extending below the fish-pole antenna so as to be actuated by the rod **123**. The projection **97'** is biased upwardly by suitable springs.

It will readily be understood that when the knob **122** is moved upwardly from its normal position and when the knob **122** is returned to its normal position, the same switching operation effected by the erection and reclining of the fish-pole antenna **93** will take place.

The loop antenna shown in Fig. 11 differs from the loop antenna **67** primarily in that it consists of two turns **124** of relatively stiff wire. This wire is sufficiently strong so that it does not need to be enclosed in a dielectric tube as shown in the previously described embodiment.

The two turns **124** are supported where they enter the housing **69** by means of two bushings **125** of suitable dielectric material. The two loops **124** are further secured together preferably at a position remote from the housing **69** by means of a suitable block or spacer **126** of insulating material. The remainder of the elements associated with the loops 124 may be precisely the same as in the previously described embodiment.

The operation is as follows: To prepare the receiver for battery operation, the four-pole switch 48 is thrown to the right, as viewed in Fig. 8, and the double-pole single-throw switch 31 is closed, supplying electrical energy to the tubes. On the other hand, if it is desired to operate from power lines with A. C. or D. C. current, the plug 46 is plugged into a suitable outlet, the switch 48 is thrown to the left, as viewed in Fig. 8, and the switch 31 is used as an "on and off" switch as it was when the receiver was battery operated.

For operation in the broadcast band, the antenna employed is selected in accordance with the location. At the beach or at a football game, it is preferred to employ the fish-pole antenna 93 for this puropse, whereas on a train or in a metal airplane, or indeed in any partially shielded structure or in any location where there is a large amount of man-made static, better results will be obtained with the loop 56.

Operation with the loop 56 will first be described. The loop 56, together with the loop 67, is mounted by means of the section cups 59 at a suitable location on a train window as shown in Fig. 1. Considerable improvement in reception can usually be obtained by slight changes of the mounting position on the window and changes of orientation of the loop relative to the window.

The broadcast bar 38 is pushed inwardly and tuning is effected by means of control 25, the volume being controlled by the button 26 which actuates the switch 31. The tuning is indicated on the scale and indicia 35, this being the broadcast scale.

Reception is improved by orientating the loop 56 relative to the plates 58 whereby the loop is brought into the best plane for reception. Reception may be improved further by operating the switch 65 into the best of its three positions.

Since the fish-pole antenna 93 is in its unerected position, the switches 98 and 99 are in their other positions from those shown in Fig. 8, switch 100 being closed.

Consequently, conductor 61 is connected directly to the control grid of the RF amplifier tube 50. The loop is tuned by the condenser 28 of the gang of condensers, the other side of the condenser 28 being connected through loading coil 63 to conductor 62.

Since the broadcast bar 38 has been actuated, its upper portion in Fig. 8 is displaced upwardly and its lower portion in that figure is displaced downwardly. As a result of the upward displacement of the bar 38, the high side of the secondary of the radio frequency transformer 101 is connected through condenser 102 to a grid of the mixer tube 51. Unless the bar 38 is pushed inwardly, this connection is not made and consequently no broadcast signal can pass from the radio frequency tube 500 to the mixer tube 51.

It may be noted that the condenser 102 is relatively large being, for example, a .02 microfarad condenser. On the other hand, the condenser 103 which connects the high side of the secondary of the transformer 101 to the condenser 102 is very small, having a capacity of around 15 micro-microfarads. Thus at broadcast frequencies the condenser 103 is not large enough to couple the secondary of transformer 101 to the mixer tube 51. Condenser 103 is used in short wave reception as will be described. The actuation of the bar 38 shorts the condenser 103. The actuation of the bar 38 tunes the oscillator tube 55 to an appropriate frequency band. The grid of this tube is connected through condenser 104, contacts 105 and 106 to the tank circuit 107, the other side of which is connected to radio frequency ground. Radio frequency ground is determined by condenser 108 which may suitably be a .1 microfarad condenser. The cathode of the tube 55 is connected thereto through condenser 109 which may suitably be a .07 microfarad condenser. The plate of the tube 55 is connected through condenser 110 and contacts 111 and 112 to a coil 113 coupled to the tank circuit 107.

The same circuits are established when an outdoor aerial is employed with the aid of the coil 90.

Employing the fish-pole antenna 93, the erection of the fish-pole or its extension in the case of a fish-pole of the second described variety, moves the switches 98, 99 and 100 into the positions in which they are shown in Fig. 8. The circuit of loop 56 is now broken and signal passes from the fishpole antenna 93 through switch 99, through contacts on the actuated bar 38 up to the tap on coil 114 to radio frequency ground.

Condenser 115 is a trimmer to compensate for variations in fish-pole circuit capacities. The signal energizes the circuit including coil 114 and tuning condenser 28, the high side of which circuit is connected to a grid of the radio frequency tube 50. The oscillator circuit is as previously described.

The fish-pole 93 can also be employed in the case of all the short wave bands. In this connection it is pointed out that when operating on any of the short wave bands, the broadcast bar 38 is in its outer position, so that no broadcast signal from the loop 56 can pass to the mixer as has previously been described. At short wave frequencies the loop 56 is substantially shorted so that substantially no short wave signal passes from tube 50 to tube 51 even though they are coupled at these frequencies through condenser 103.

If the button 43 is actuated, the fish-pole signal passes to coil 43*a*, the low side of which is connected to radio frequency ground. The high side of the coil 43*a* is connected through the unactuated bar 38 and through coupling condenser 102 to a grid of the mixer tube 51. At the same time the high side of the coil 43*a* is connected through the relatively small condenser 103 to the high side of the tank circuit including the secondary of the transformer 101 and the tuning condenser 29. The condenser 29 tunes the signal input circuit of tube 51 in coordination with the tuning of the oscillator circuit. The coil 43*a* thus forms a part of a circuit tuned to the signal frequency.

The actuation of the bar 43 (lower portion in Fig. 8) connects the circuit of the grid of oscillator tube 55 to the upper end of coil 43*b*, the lower end of which (actually a tap) is connected to radio frequency ground.

At the same time the plate of the oscillator tube 55 is connected through condenser 110 and the bar 43 to the coil 43*c* coupled with coil 43*b*, the other end of the coil 43*c* being connected to radio frequency ground (actually the same tap).

The grid of the oscillator tube 55 is connected through condenser 104 and contacts 105 and 116 on the unactuated bar 38 to a condenser 117. The condenser 117 is connected to radio frequency ground by a condenser 118 and it is also connected by contacts 119 and 106 on the unactuated bar 38 to the tank circuit 107. This permits the use of condenser 30 as a band spread tuner. Since this condenser is in series with the small condenser 117 and is shunted by fixed condenser 118, the effect in change of capacity when condenser 30 is varied is, therefore, very small. The oscillator frequency is thus tuned by the variable condenser 30 to give an appropriate oscillator frequency for this band.

Similar results can be obtained with the bars 42, 41 and 40, the associated coils being indicated 42*a*, 42*b*, and 42*c*, etc.

The bar 39 controls an oscillator circuit in precisely the same way as the bars 40 to 43, coils corresponding to 40*b*, 40*c* being designated 39*b* and 39*c*, respectively. The fish-pole signal, however, goes through the coil 39*a* to radio frequency ground. This coil is inductively coupled to a winding 39*d*, thus decreasing the coupling of the fish-pole to the input circuit. The winding 39*d* is connected through the two uppermost contacts on bar 39 (Fig. 8) through condenser 102 to the grid of the mixer tube 51 and through condenser 103, to the tank circuit including the secondary of transformer 101 and the tuning condenser 29.

When the button 38 is not in actuated condition and when the fish-pole antenna 93 is in its unerected position or the corresponding fish-pole antenna 120 is in fully telescoped condition, both the loop 56 and the fish-pole are rendered inoperative. In the first case, the high side of the secondary of the transformer 101 is effectively disconnected and in the second case the switch 99 is opened.

The receiver is now in condition for operation by the loop 67 (or the equivalent loop 124). It is to be noted that the switch 100 is closed when the fish-pole is stowed away. This switch may be omitted if desired since when the fish-pole antenna is in operation, the two loops 56 and 67 are stowed away in the back of the case 20. At the high frequencies for which the loop 67 is suited, this loop is substantially shielded by the other nearby conductors in the case and is not in the least effective. However, in some cases it is preferred to use the switch 100 and if it is used, it must be closed when short wave loop reception is desired.

The loop 67 being mounted on window 60 of a railroad car, for example, it is orientated manually to give best reception and the button 40, 41 or 42 is pushed in accordance with the band which it is desired to receive.

The switch 71 is moved into the corresponding position to select the same band, this switch being located on the housing 69 of the loop 67 as has been previously described.

The receiver is, of course, turned on by the switch 31 on the volume control 26 and tuning is effected by means of the button 25 to bring in the desired station. Control wheel 79 may then be adjusted to tune the loop 67 accurately within the selected band to the station manually selected by button 25. Slight correction for variation of capacity of the loop 67 by reason of varying position on the window 60 is also readily taken care of by actuating the control wheel 79 of the condenser 70.

The signal from the loop 67 is conveyed through the link circuit which includes the coil 78, the leads 82 and 81 and the coils 74, 75 and 76. The signal is transferred from the appropriate one of these coils to the coil 42*a*, 41*a*, or 40*a*, depending upon which of the three buttons 42, 41 or 40 is in the actuated position.

As has previously been explained, one end of each of the coils 42*a*, 41*a* and 40*a* is connected to radio frequency ground and the high sides of these coils are connected, when their respective buttons 42, 41 and 40 are pushed, to the grid of the mixer tube 51 and to the tuning condenser 29 in the manner previously described in connection with short wave operation with the fish-pole antenna.

Although the invention has been described in connection with specific details of preferred embodiments thereof, it must be understood that such details are not intended to be limitative of the invention except in so far as set forth in the accompanying claims.

Having thus described my invention, I declare that what I claim is:

1. In combination, a radio receiver, said receiver having an input circuit, a loop antenna winding spaced a substantial distance from said input circuit, a low impedance untuned link circuit extending between said input circuit and said antenna winding and arranged to transfer radio frequency energy therebetween, said link circuit including a coil having less turns than said antenna winding, means mounting said antenna winding and coil in close proximity to one another such that said antenna winding and coil comprise elements of a transformer, and means for tuning said input circuit to a frequency equal to the frequency at which said antenna winding is tuned whereby radio energy is efficiently transferred between said antenna winding and link circuit and variations of stray capacity of said link circuit produce no substantial effect on transfer of such radio energy.

2. In combination, a radio receiver, said receiver having a plurality of input circuits tunable to the frequency of signals in any one of a plurality of bands in a relatively large frequency spectrum, means for tuning said input circuits, each of said input circuits including a coil, a loop antenna arranged to be mounted remotely from said receiver, means proximate to said loop for tuning it to a predetermined resonant frequency corresponding to a frequency in any one of said plurality of bands, a low impedance untuned link circuit arranged to pass signals in all of said bands and coupling said loop antenna to said coil of each of said input circuits, and means for selectively connecting one of said tunable input circuits to said receiver for the transfer of signals from said link circuit to said receiver.

3. In combination, a portable radio receiver unit having an input circuit, a coil connected in said input circuit for the transfer of energy to said unit, means for tuning said coil to a frequency desirable for operation of said radio unit, a short wave loop antenna arranged to be mounted remotely relative to said unit, means proximate to said loop for tuning it to a predetermined resonant frequency equal to the first mentioned frequency, an untuned link circuit coupling said loop antenna to said coil in said input circuit, said link circuit comprising: a two-conductor cable extending from said loop antenna to said unit and having its conductors connected to opposite terminals of a first coil coupled to said loop antenna for voltage stepdown from the loop to said first coil and having its conductors connected to a second coil coupled to said coil in the input circuit, for the transfer of signals received in said loop antenna to said input circuit.

4. In combination, a radio receiving unit, a plurality of input circuits tunable to the frequency of signals in any one of a plurality of bands, each of said input circuits including a coil, means for tuning said coil to a frequency desirable for operation of said radio unit, a loop antenna arranged to be mounted remotely from said unit, means proximate to said loop for tuning it to a predetermined resonant frequency equal to the first mentioned frequency and corresponding to a frequency in any one of said plurality of bands, a low impedance untuned link circuit coupling said loop antenna to said coil of each of said input circuits, said link circuit comprising: a two-conductor cable extending from said loop antenna to said unit and having its conductors connected to opposite terminals of a first coil near said loop antenna for voltage stepdown between said loop antenna and said first coil and having its terminals connected to opposite terminals of a second coil coupled to said coil in each of said input circuits, for the transfer of energy from said loop antenna to said input circuit, and means for selectively connecting one of said tunable input circuits to said unit for the transfer of signals from said loop over said link circuit to said unit.

5. In combination, a radio receiver having an input circuit, a short wave loop antenna arranged to be mounted remotely from said unit, a base structure carrying said loop, variable capacitance means proximate to said loop for tuning it to any selected one of a group of predetermined resonant frequencies, an untuned coil mounted on said base and coupled to said loop for substantial voltage stepdown from said loop to said untuned coil, means comprising a low impedance network for transferring a voltage induced in said coil to the input circuit of said receiver, and means for tuning said input circuit to the said predetermined resonant frequencies, said coil having such a number of turns so that it serves to produce such substantial voltage stepdown and serves to match the impedance of the resulting tuned loop antenna with the impedance of such transferring means.

6. In combination, a receiver having an input circuit, a short wave loop antenna comprising a loop, a base structure carrying said loop, mounting means for said base whereby the loop may be located in desired positions, hinge means connecting said base to said mounting means whereby said loop may be oriented for best reception, variable capacitance means proximate to said loop whereby it may be tuned to a predetermined resonant frequency, a coil coupled to said loop and mounted on said base, means for transferring a voltage induced in said coil from said loop to the input circuit of said receiver, and means for tuning said input circuit to the same frequency as said predetermined resonant frequency.

7. In combination, a radio receiver having an input circuit, a short wave antenna comprising a loop, a base structure arranged to be placed remotely from said receiver and carrying said loop, a plurality of capacitors carried by said base, switching means for cutting individual capacitors into circuit with said loop whereby the same is tuned to any one of a plurality of predetermined resonant frequencies, a low impedance coil coupled to said loop for substantial voltage stepdown from said loop to said coil, said coil being mounted on said base, means for transferring a voltage induced in said coil from said loop to the input circuit of said receiver, and means for tuning said input circuit to the same frequency as said predetermined frequencies.

8. In combination, a portable radio receiver unit having an input circuit, means for tuning said input circuit to a frequency desirable for operation of said radio unit, a short wave loop antenna arranged to be mounted remotely relative to said unit, variable capacitance means arranged to tune said loop to a predetermined resonant frequency in a band of frequencies, and an untuned low impedance link circuit coupled to said loop and coupled to said input circuit of the unit for the transfer of radio energy selectively received in said loop to said unit.

9. In combination, a radio receiver having an input circuit, means for tuning said input circuit to a frequency desirable for operation of said receiver, a loop antenna arranged to be mounted in remote relation thereto, an untuned low impedance link circuit coupled to said antenna for substantial voltage stepdown from said loop to said link circuit, said link circuit being coupled to said input circuit and extending between said loop and radio receiver to communicate signals from said antenna to said receiver, and means proximate to said loop antenna for changing its resonant frequency to the first mentioned frequency.

10. In combination, a radio receiver having an input circuit, a pair of loop antennas, mounting means whereby said antennas may be mounted in alignment with a suitable dielectric surface, means pivotally connecting said loop antennas to said mounting means whereby said loop antennas may be oriented in the most advantageous direction for signal reception, tuning means proximate to one of said loop antennas whereby the same is adapted for the reception of a plurality of bands, a first coil coupled to said one loop antenna for substantial voltage stepdown from said last one loop antenna to said first coil, a second coil coupled to said input circuit, a pair of conductors extending between and connecting said first and second coils, said first coil and second coil and conductors comprising a link circuit, conductors connecting the other loop antenna to the input circuit of the receiver, cable means including all of said conductors, and switching means on the receiver for rendering each of said loop antennas alternatively effective to transfer signals to said receiver.

11. In combination, a radio receiver having an input circuit, a pair of loop antennas, mounting means whereby said antennas may be mounted in alignment with a suitable dielectric surface, means pivotally connecting said loop antennas to said mounting means whereby said loop antennas may be oriented in the most advantageous direction for signal reception, tuning means proximate to one of said loop antennas whereby the same is adapted for the reception of a plurality of broadcast bands, a first coil coupled to said one loop antenna for substantial voltage stepdown from said one antenna to said coil, a second coil coupled to said input circuit, a pair of conductors extending between and connecting said first and second coils in a series circuit, said series circuit comprising an untuned low impedance link circuit, conductors between the other loop antenna and input circuit of the receiver, cable means including said first and last mentioned conductors, switching means on the receiver for rendering each of said loop antennas alternatively effective

to transfer signal energy to said receiver, said switching means including band selecting means for adapting the receiver for the reception of signals in any of a plurality of said bands.

12. In combination, a radio receiver having an input circuit and a housing, a pair of loop antennas, mounting means whereby said antennas may be mounted in alignment with a suitable dielectric surface, means pivotally connecting said loop antennas to said mounting means whereby said loop antennas may be oriented in the most advantageous direction for signal reception, tuning means proximate to one of said loop antennas for tuning it at a predetermined resonant frequency for the reception of signals in a plurality of bands, a first coil coupled to said one loop antenna for substantial voltage stepdown, a second coil coupled to said input circuit, conductors extending between and connecting said first and second coils in a series circuit, said first coil, second coil and conductors forming an untuned low impedance link circuit, at least one additional conductor extending between and connecting the other loop antenna to the input circuit of the receiver, cable means including all of said conductors and extending from said loop antennas to the receiver, a telescopic antenna, said telescopic antenna being of such size with respect to said housing and arranged so that said antenna may be moved from a position substantially within said housing to a position where said antenna extends substantially entirely outside of said housing, switching means for alternatively connecting said other loop antenna and said telescopic antenna to said input circuit, an operating member for said switching means positioned in the path of movement of said telescopic antenna such that movement of said telescopic antenna between said position substantially within said housing and said position substantially outside of said housing results in alternatively connecting said other loop and said telescopic antenna to said input circuit, and switching means on the receiver for rendering each of said loop antennas alternatively effective to transfer signals therefrom to said input circuit.

13. In combination with an antenna for the reception of radiated energy, a receiver arranged to be energized by such energy and located at a substantial distance from said antenna, a tuned input circuit for the receiver, said antenna being tuned to the same frequency as said input circuit a low impedance link circuit extending between said antenna and receiver and including means at its ends for coupling said antenna to said link circuit and for coupling said link circuit to said input circuit of the receiver, said coupling means comprising means for reducing the voltage of said energy passing from said antenna to said link circuit and for increasing the voltage of said energy passing from said link circuit to the input circuit of said receiver.

14. In combination, radio apparatus, a housing for the radio apparatus, said radio apparatus including metal parts, a first antenna, a second antenna, means within said housing for movably mounting said first antenna with respect to said housing, said first antenna being of such size with respect to said housing and arranged so that said first antenna may be moved from a position substantially within said housing wherein said metal parts impair the effectiveness of said first antenna to a position substantially entirely outside of said housing wherein said metal parts influence in small degree the effectiveness of said first antenna, and switching means for rendering said first antenna and said second antenna alternatively effective for the passage of signals between said antennas and said apparatus, said switching means including a member arranged to move with said first antenna during at least a portion of the movement of said first antenna between said position within said housing and said position outside of said housing, said member being arranged to disconnect said second antenna from said apparatus and to connect said first antenna to said apparatus in response to movement of said first antenna to said outside position, the resulting connections being reversed upon movement of said first antenna to said position within said housing.

15. In combination, radio apparatus including an energy transferring circuit, a houisng for said apparatus, said radio apparatus including metal parts, an antenna movable from a position within said housing wherein said metal parts impair the effectiveness of said antenna to a position substantially outside of said housing wherein said metal parts influence in small degree the effectiveness of said antenna, and means including an operating member positioned in the path of movement of said antenna operated in response to movement of said antenna from said position within said housing to a position outside of said housing for electrically connecting said antenna to said circuit for the transfer of radio energy between said radio apparatus and antenna.

16. In combination, a radio apparatus, a housing for said radio apparatus, said radio apparatus including metal parts, a first antenna, a second antenna, means for movably mounting said first antenna with respect to said housing, said first antenna being of such size with respect to said housing and arranged so that said first antenna may be moved from a first position substantially within said housing wherein said metal parts impair the effectiveness of said first antenna to a second position substantially entirely outside of said housing wherein said metal parts influence in small degree the effectiveness of said first antenna, and means including an operating member positioned in the path of movement of said first antenna operated in response to movement of said first antenna between said position within said housing to said position substantially entirely outside of said housing for alternatively connecting electrically said antennas to said apparatus for the passage of signals therebetween.

17. In combination, radio apparatus including a radio energy transferring circuit, said apparatus including metal parts, a first antenna movable from a position relatively near said apparatus wherein the effectiveness of said antenna is impaired by said metal parts to a position more remote from said apparatus wherein said metal parts influence in small degree the effectiveness of said antenna, a second antenna, and means responsive to movement of said first antenna from said position relatively near said apparatus to said more remote position for electrically disconnecting said second antenna from said energy transferring circuit and for electrically connecting said first antenna to said energy transferring circuit for the transfer of radio energy between said radio apparatus and first antenna, said disconnecting and connecting means including a member arranged to move with said first an-

tenna during at least a portion of the movement of said first antenna between said positions, said member being arranged to move to open a circuit between said second antenna and said energy transferring circuit in response to said movement and to close a circuit between said first antenna and said energy transferring circuit in response to said movement.

18. In combination, a radio receiver having a housing, said receiver including metal parts, a telescopic rod antenna mounted on said receiver for transferring signals thereto and arranged to extend substantially within said housing when the antenna is telescoped wherein said metal parts impair the effectiveness of said antenna and to extend substantially outside of said housing when the antenna is extended in which position said metal parts influence in small degree the effectiveness of said antenna, a second antenna for transferring signals to the receiver, switching means including an operating member positioned in the path of telescoping movement of the first mentioned antenna operated in response to movement of said first mentioned antenna from its telescoped position within said housing to its extended position outside of said housing for rendering said first mentioned and second antennas alternatively effective for the passage of signals between said antennas and receiver.

19. In combination, a radio receiver, said receiver including metal parts, a telescopic antenna mounted on said receiver for transferring signals thereto and arranged to be positioned in close proximity to said receiver when the antenna is telescoped in which position said metal parts impair the effectiveness of said antenna and to extend a substantial distance away from the receiver when the antenna is extended in which position said metal parts influence in small degree the effectiveness of said antenna, a loop antenna for transferring signals to the receiver, and switching means for rendering said first and second antennas alternatively effective for the passage of signals between the antennas and receiver, said switching means including a member arranged to move with said first mentioned antenna during at least a portion of the movement of the first mentioned antenna from its position close to said receiver to its position a substantial distance from said receiver, said member being arranged to move to disconnect said loop antenna from said receiver and to connect said first mentioned antenna to said receiver in response to said movement.

20. In combination, radio apparatus, a housing for said apparatus, said apparatus including metal parts, an antenna, means within said housing for movably mounting said antenna with respect to said housing, said antenna being of such size with respect to said housing and arranged so that said antenna may be moved from a position substantially within said housing wherein said metal parts impair the effectiveness of said antenna to a position where said antenna extends substantially entirely outside of said housing wherein said metal parts influence in small degree the effectiveness of said antenna, and means including an operating member positioned in the path of movement of said antenna responsive to movement of said antenna from said position substantially within said housing to said position entirely outside of said housing for operatively connecting electrically said antenna to said apparatus for the passage of signals therebetween.

21. In combination, radio apparatus, a housing for said radio apparatus, said apparatus including metal parts, an antenna movable from a position within said housing wherein said metal parts impair the effectiveness of said antenna to a position outside of said housing wherein said metal parts influence in small degree the effectiveness of said antenna, and means including an operating member positioned in the path of movement of said antenna responsive to movement of said antenna from said position within said housing to said position outside of said housing for electrically connecting said antenna to the apparatus for the passage of signals therebetween.

22. In combination, radio apparatus, a housing for said apparatus, a telescopic rod antenna for the reception of signals and movable from a telescoped position within said housing in close proximity to said apparatus to an extended position outside of said housing removed from said apparatus, said apparatus having metal parts which materially influence a function of the antenna when said antenna is positioned within said housing and in proximity to said apparatus, and means for connecting electrically said antenna to the apparatus for the passage of signals therebetween, said connecting means including a member arranged to move with said antenna during at least a portion of the movement of said antenna between its telescoped and extended positions, said member being arranged to move to close a circuit between said antenna and said apparatus in response to said movement.

23. In combination, radio apparatus normally disposed near the earth's surface, a rod antenna attached mechanically to said apparatus and movable from a vertical position with respect to the earth's surface where said antenna is highly effective to pass signals to a position where said rod antenna extends substantially parallel to the earth's surface, and means for operatively connecting electrically said antenna to said apparatus for the passage of signals therebetween, said connecting means including a member arranged to move with said antenna during at least a portion of the movement of said antenna from said horizontal position to said vertical position, said member being arranged to move to close a circuit between said antenna and said apparatus in response to said movement.

24. In combination, radio apparatus including a radio energy transferring circuit, a first antenna movable with respect to said apparatus, a second antenna, and switching means for alternately connecting said first antenna and second antenna electrically with said radio energy transferring circuit for the alternate transfer of radio energy between said antennas and said apparatus, said switching means including a member arranged to move with said first antenna during at least a portion of the movement of said first antenna with respect to said apparatus, said member being arranged to move to connect said circuit to one of said antenna and disconnect said circuit from the other antenna in response to said movement.

25. In combination, means for receiving radio frequency energy, said means having a high impedance input circuit, a loop antenna winding spaced a substantial distance from said input circuit, a low impedance untuned link circuit extending between said input circuit and antenna winding and arranged to transfer radio frequency energy therebetween, said link circuit including a

coil, means mounting said antenna winding and link circuit coil in close proximity to each other such that said antenna winding and link circuit coil comprise elements of a transformer, means for tuning said antenna winding to a frequency of signals desirable for operation of said radio receiving means, means for tuning said input circuit to the same frequency at which said antenna winding is tuned, and means for matching the impedance of the resulting tuned antenna winding to the low impedance of said link circuit, and means for matching the high impedance of said input circuit in said radio receiving means to the low impedance of said link circuit whereby radio energy is efficiently transferred between said antenna winding and link circuit and input circuit and variations of stray capacity of such link circuit produce no substantial effect on the transfer of such radio energy.

26. In combination, a radio receiver having an input circuit, a pair of loop antennas for the reception of radio signals, said antennas being arranged to be mounted remotely relative to the receiver, tuning means proximate to one of said antennas for tuning it to a predetermined resonant frequency, means for tuning said input circuit to a frequency equal to said predetermined frequency, means including conductors for coupling said one antenna to the input circuit of said receiver, means including at least one other conductor for coupling the other antenna to the input circuit of said receiver, flexible cable means including said conductors extending between said antennas and receiver, switching means on the receiver for rendering alternatively effective said one and the other antenna to transfer signals therefrom to said input circuit, said switching means including band selecting means for adapting the receiver for the reception of signals in any of a plurality of broadcast bands.

27. In combination, means for translating radio frequency energy, said means having an energy transferring circuit, a loop antenna spaced a substantial distance from said energy transferring circuit, means for tuning said antenna to a frequency desirable for the transmission of energy to said translating means, a low impedance untuned link circuit extending between said energy transferring circuit and said antenna and arranged to transfer radio frequency energy therebetween, means for tuning said transferring circuit to said desirable frequency, whereby radio energy is efficiently transferred between said antenna and link circuit and variations of stray capacity of said link circuit produce no substantial effect on transfer of such radio energy, a low capacity antenna arranged to be connected to said energy transferring circuit, and switching means for alternatively connecting said antennas to said energy transferring circuit for the transfer of energy therebetween.

28. In combination, means for translating radio frequency energy, said means having an energy transferring circuit, a loop antenna spaced a substantial distance from said energy transferring circuit, means proximate to said antenna for tuning it to predetermined frequencies, a low impedance untuned link circuit extending between said energy transferring circuit and said antenna and arranged to transfer radio frequency energy therebetween, means for tuning said transferring circuit to frequencies equal to said predetermined frequencies, whereby radio energy is effectively transferred between said antenna and link circuit and variations of stray capacity of said link circuit produce no substantial effect on transfer of such radio energy, a movable low capacity antenna arranged to be connected to said energy transferring circuit, and means responsive to movement of said last antenna for alternatively connecting said antennas to said energy transferring circuit for the transfer of energy therebetween.

29. In combination, a radio receiver having an input circuit, a pair of antennas operable for the reception of radio signals in different frequency bands and being arranged to be mounted remotely relative to the receiver, means proximate to one of said antennas for tuning it to predetermined frequencies, means for tuning said input circuit to frequencies equal to said predetermined frequencies, means for coupling said antennas to the input circuit for the transfer of radio frequency energy in different frequency bands to said input circuit, and means on the receiver for rendering each of said antennas alternatively effective to transfer radio frequency energy in different frequency bands to said receiver.

30. In combination, a radio receiver having an input circuit, a first and a second antenna operable for the reception of signals in different frequency bands, means proximate to said first antenna for tuning said first antenna to predetermined frequencies, said input circuit including a plurality of tuned circuits tunable to frequencies equal to said predetermined frequencies and arranged to be selectively coupled to said first antenna for the selective reception of signals in a plurality of frequency bands, switching means for selectively coupling one of said tuned circuits to said first antenna, means for coupling said second antenna to the input circuit of the receiver, and means responsive to operation of said switching means for uncoupling said first antenna from said input circuit.

31. In combination, a radio receiver having an input circuit, a first antenna and a second antenna arranged for the reception of radio signals in different frequency bands, means for coupling said first antenna to said input circuit for the transfer of radio energy therebetween, means for coupling said second antenna to said input circuit for the transfer of energy therebetween, said last mentioned means including one of a plurality of tuned circuits tuned to predetermined frequencies for the selective transfer of signals between said second antenna and input circuit in any one of a plurality of frequency bands, means for tuning said second antenna to frequencies equal to said predetermined frequencies, means for selectively connecting one of said plurality of tuned circuits in operative relationship for the selective transfer of signals between said second antenna and input circuit, and means responsive to actuation of said connecting means for uncoupling said first antenna from said input circuit.

32. In combination, an antenna, means proximate to said antenna for tuning the same to a single frequency within a range of frequencies, radio receiving apparatus, an input circuit for said radio receiving apparatus located remotely relative to said antenna, means for tuning said input circuit to a single frequency equal to the first mentioned frequency, and a low impedance circuit coupling the resulting tuned antenna to the resulting tuned input circuit, the overall coupling between the antenna and input circuit being of small degree such that efficient energy

transfer consistent with high selectivity and image ratio results.

33. In combination, a radio receiver, said receiver having a plurality of input circuits each tunable to the frequency of signals in a different range, means for connecting any selected one of said tunable input circuits to said receiver, means for tuning the selected circuit to any desired signal frequency within its range, a loop antenna arranged to be mounted remotely from said receiver, antenna tuning means proximate to the loop comprising discontinuous means for tuning the loop to any selected one of a plurality of frequencies each in a different one of said ranges and continuous means for tuning the loop to any signal frequency within the range for which the discontinuous means has been set, and means for coupling said antenna to the selected input circuit, said coupling means comprising a line having a coil at one end coupled to said loop antenna for substantial voltage stepdown from antenna to coil, means to match the impedance of said loop antenna to the impedance of the line, and means to match the line impedance to the impedance of the selected input circuit, said coupling means having a low characteristic impedance, a low shunt capacity, and providing a small degree of coupling between said antenna and input circuit.

34. In combination, radio receiving apparatus having an input circuit, means for adjustably tuning said input circuit, a loop antenna, means proximate to said antenna for adjustably tuning said antenna to a frequency corresponding to the frequency to which the input circuit is tuned, means for coupling said antenna to said input circuit, said coupling means comprising a long flexible line having a coil at one end coupled to said loop and arranged for substantial voltage stepdown from said loop to said coil, the overall coupling between said loop and said input circuit being of small degree, and means for quickly attachably and detachably supporting said antenna at a point removed from said receiving apparatus.

35. In combination, a radio receiver having an input circuit, means for adjustably tuning said input circuit to a frequency desirable for operation of said receiver, a loop antenna arranged to be mounted in remote relation thereto, a line having a low characteristic impedance and a low shunt capacity, means for coupling in small degree said antenna to said line for substantial voltage stepdown from said antenna to said line, and means for coupling in small degree said line to said input circuit, said line extending between said antenna and input circuit to communicate signals from said antenna to said input circuit, and means proximate to said loop antenna for tuning it to the first mentioned frequency.

36. In combination, a radio receiver, said receiver having a plurality of input circuits tunable to the frequency of signals in any one of a plurality of bands in a relatively large frequency spectrum, means for connecting any selected one of said tunable circuits to said receiver, means for tuning the selected circuit to any signal frequency within the selected band, a loop antenna arranged to be mounted remotely from said receiver, antenna tuning means proximate to this loop comprising discontinuous means for tuning the loop to any selected one of a plurality of frequencies each in a different one of said bands, and continuous means for tuning the loop to any signal frequency within the band corresponding to the frequency to which said input circuit is tuned, and means for coupling said antenna to the selected input circuit, said coupling means comprising a line having a coil at one end coupled to said loop antenna for substantial voltage stepdown from antenna to coil, said coupling means having a low characteristic impedance, a low shunt capacity, and a low overall coefficient of coupling.

37. In combination, a radio receiver, said receiver having a plurality of input circuits tunable to the frequency of signals in any one of a plurality of bands in a relatively large frequency spectrum, means for connecting any selected one of said tunable input circuits to said receiver, means for tuning the selected circuit to any signal frequency within the selected band, a loop antenna arranged to be mounted remotely from said receiver, antenna tuning means proximate to the loop comprising discontinuous means for tuning the loop to any selected one of a plurality of frequencies each in a different one of said bands and continuous means for tuning the loop to any signal frequency within the band corresponding to the frequency to which said input circuit is tuned, and means for coupling said antenna to the selected input circuit, said coupling means having a low characteristic impedance and a low shunt capacity and a low overall coefficient of coupling and comprising a line having a coil at one end coupled to said loop antenna, said coupling means being arranged to maintain the voltage therein substantially lower than the voltage in either said antenna or the selected input circuit, means to improve the impedance match between said antenna and said line and between said line and said input circuit, and means for quickly attachably and detachably supporting said antenna at a point removed from said receiver.

38. In combination, a radio receiver having an input circuit, means for tuning said input circuit to any desired frequency within a range of frequencies, a first antenna, means proximate to said antenna for tuning it to the same frequency as said input circuit, said first antenna being movable to a position removed from said input circuit, means for coupling said antenna to said input circuit, said coupling means having a low characteristic impedance and being arranged to provide overall coupling between said antenna and said input circuit of such small degree that efficient energy transfer consistent with high selectivity results, a second antenna, and means for connecting said second antenna with said input circuit and disconnecting said coupling means therefrom and alternatively for disconnecting said second antenna from said input circuit and connecting said coupling means thereto.

39. In combination, radio receiving apparatus having an input circuit, means for adjustably tuning said input circuit to any desired frequency within the short wave range, a loop antenna, means proximate to said antenna for adjustably tuning said antenna to a frequency corresponding to said desired frequency, and long flexible means for coupling said antenna to said input circuit, the overall coefficient of coupling between said antenna and said input circuit lying in a range of medium to loose coupling, said coupling means having a low characteristic impedance and a low shunt capacity.

40. In combination, a radio receiver having an input circuit, means for tuning said input circuit to any desired frequency within a range of frequencies, a first antenna, means proximate to

said antenna for tuning it to the same frequency as said input circuit, said first antenna being movable to a position removed from said input circuit, means for coupling said antenna to said input circuit, said coupling means having a low characteristic impedance and being arranged to provide overall coupling between said antenna and said input circuit of such small degree that efficient energy transfer consistent with high selectivity results, a second antenna, and means responsive to movement of one of said antennas with respect to said receiver for connecting said second antenna with said input circuit and disconnecting said coupling means therefrom and alternatively for disconnecting said second antenna from said input circuit and connecting said coupling means thereto.

41. In combination, an antenna, means proximate to said antenna for adjustably tuning the antenna to any desired frequency, radio receiving apparatus, an input circuit for said radio receiving apparatus, means for adjustably tuning said input circuit to a frequency equal to said desired frequency, and a flexible line having a low characteristic impedance and a low shunt capacity for coupling said antenna to said input circuit, the overall coupling between said antenna and input circuit being of small degree, means for supporting said antenna and said line on said receiving apparatus, said line being of sufficient length that when coupled between said antenna and input circuit, the antenna may be located at a remote point with respect to the input circuit, and means for quickly attachably and detachably supporting said antenna at such a remote point.

42. In combination, means for receiving radio frequency energy, said means having a high impedance input circuit, a loop antenna winding arranged to be placed a substantial distance from said energy receiving means, a line having a low characteristic impedance and a low shunt capacity extending between said energy receiving means and antenna winding and arranged to transfer energy therebetween, said line including a coil, means mounting said antenna winding and coil in close proximity to each other such that said antenna winding and coil comprise elements of a transformer, means variable to tune said antenna winding to a frequency of signals desirable for operation of said receiving means, means variable to tune said energy receiving means to the same frequency as the antenna winding is tuned, means for matching the high impedance of the tuned antenna winding to the low impedance of said line, and means for matching the high impedance of said energy receiving means to the low impedance of said line, said line being coupled to said antenna and said receiving means in small degree.

43. In combination, a radio receiver, said receiver having a plurality of input circuits tunable to the frequency of signals in any one of a plurality of bands in a relatively large frequency spectrum, means for connecting any selected one of said tunable input circuits to said receiver, means for tuning the selected circuit to any signal frequency within the selected band, a loop antenna arranged to be mounted remotely from said receiver, antenna tuning means proximate to the loop comprising discontinuous means for tuning the loop to any selected one of a plurality of frequencies each in a different one of said bands and continuous means for tuning the loop to any signal frequency within the band corresponding to the frequency to which said input circuit is tuned, and means for coupling said antenna to the selected input circuit, said coupling means having a low characteristic impedance, a low shunt capacity, and a low overall coefficient of coupling.

GEORGE O. STRIKER.

Nov. 12, 1957

L. R. CRUMP

2,813,242

POWERING ELECTRICAL DEVICES WITH ENERGY
ABSTRACTED FROM THE ATMOSPHERE
Filed March 12, 1954

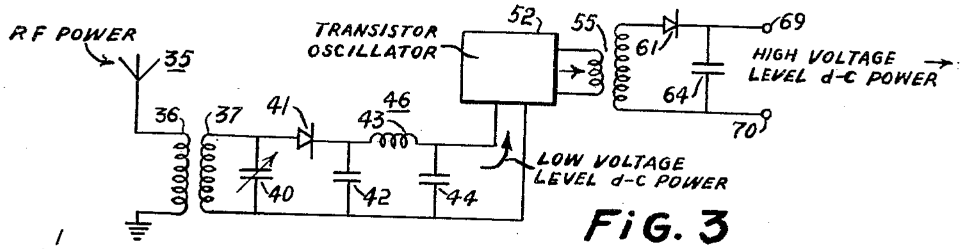
RF POWER
35
36
37
41
46
43
40
42
44
TRANSISTOR OSCILLATOR
52
55
61
64
69
70
HIGH VOLTAGE LEVEL d-c POWER
LOW VOLTAGE LEVEL d-c POWER

FiG. 3

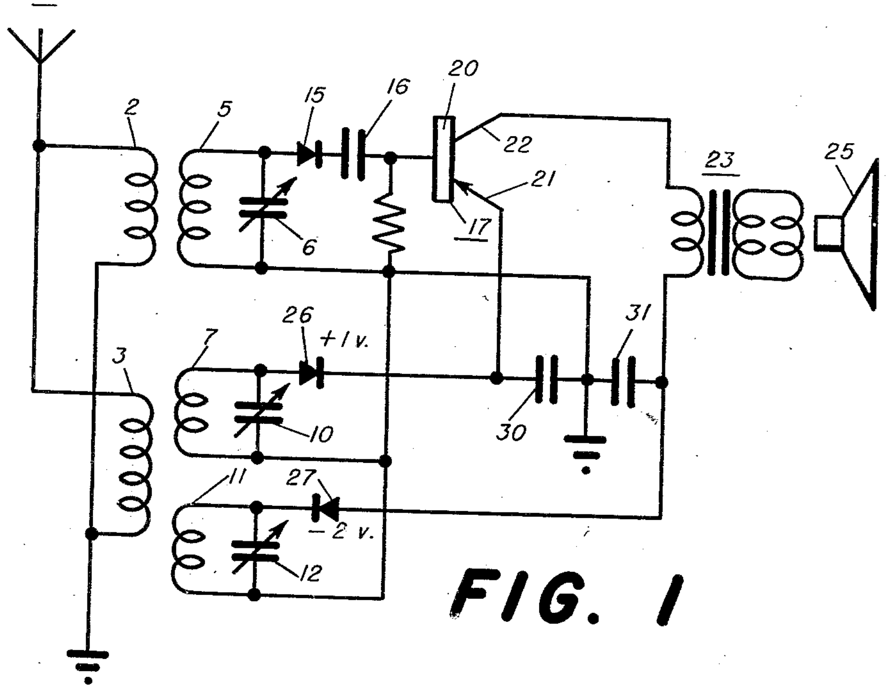
1
2
3
5
6
7
10
11
12
15
16
17
20
21
22
23
25
26
+1 v.
27
−2 v.
30
31

FIG. 1

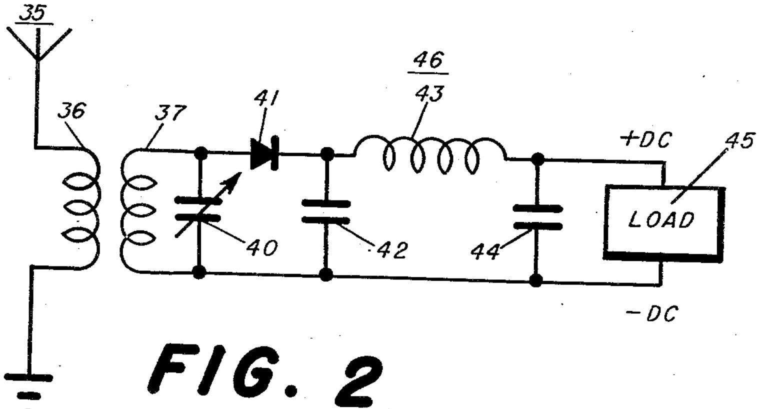
35
36
37
41
46
43
40
42
44
+DC
45
LOAD
−DC

FIG. 2

INVENTOR
Lloyd R. Crump

BY W. E. Thibodeau + A. W. Dew

ATTORNEYS

United States Patent Office

2,813,242

Patented Nov. 12, 1957

2,813,242

POWERING ELECTRICAL DEVICES WITH ENERGY ABSTRACTED FROM THE ATMOSPHERE

Lloyd R. Crump, Silver Spring, Md.

Application March 12, 1954, Serial No. 415,986

1 Claim. (Cl. 321—2)

(Granted under Title 35, U. S. Code (1952), sec. 266)

The invention described herein may be manufactured and used by or for the Government for governmental purposes without payment to me of any royalty thereon.

This invention relates to the convenient and economical provision of power for the operation of electronic circuits and devices using transistors, and of other electrical devices having modest power requirements.

A great advantage of transistors, and a major reason for their enthusiastic reception since their introduction a few years ago, is the fact that they will operate satisfactorily with very low supply voltages and currents. One milliwatt or even less is sufficient to power a transistor in many applications. Various batteries have been developed to provide, in a minimum of space, the relatively minute amounts of power needed by transistors.

My invention provides methods and means that permit transistor circuits, and also other low-powered electrical devices, to be economically and conveniently operated without any batteries whatever, and indeed without any power supply whatever as power supplies are ordinarily conceived.

The invention centers around my discovery that it is practicable to construct operative transistor circuits that are able to abstract from the atmosphere sufficient electromagnetic energy to provide all necessary supply voltages and currents for their own operation. Circuits and devices powered according to my invention will operate indefinitely without any local power source whatever.

I have successfully constructed and demonstrated such circuits. For example, I have constructed a batteryless transistor radio receiver on which I have listened to either nearby or distant broadcast stations as desired, using either headphones or a loudspeaker; this receiver has been powered entirely by electromagnetic energy abstracted from the atmosphere.

From the successful operation of this receiver, and from other experimental work, it becomes clear that, by the methods and means of the invention, a great variety of practical and useful transistor circuits can be powered entirely by energy abstracted from the atmosphere.

Furthermore, as will become apparent below, my invention is applicable to the powering of other electrical devices requiring relatively small amounts of power.

An object of the present invention is to provide methods and means for powering transistor circuits entirely from radiofrequency energy abstracted from the atmosphere.

Another object is to provide methods and means for powering remote radio receivers, low-powered radio transmitters, and other low-powered electrical devices, with energy received by radio from a master station, so that no local power supplies are needed by the devices and so that the powering or non-powering of the remote device is under the control of the master station.

A further object is to provide methods and means for powering transistor circuits and other low-powered electrical devices with radiofrequency energy received from one or more remote radio transmitters.

Other objects, aspects, uses, and advantages of the invention will become apparent from the following description and from the drawing.

Figure 1 is a schematic diagram of a transistor radio receiver in which all necessary power is supplied by energy abstracted from the atmosphere in accordance with the invention.

Figure 2 is a schematic diagram showing a general application of the invention to provide direct-current power to a load.

Figure 3 is a schematic diagram of a system for obtaining a high energy D.-C. source at a high voltage level using energy abstracted from the atmosphere.

Referring to Figure 1, a receiving antenna **1** is connected to antenna coupling coils **2** and **3**, the other ends of which are connected to ground. A parallel resonant circuit consisting of coil **5** and capacitor **6** is coupled to coil **2**. A second parallel resonant circuit consisting of coil **7** and capacitor **10** is coupled to coil **3**. A third parallel resonant circuit consisting of coil **11** and capacitor **12** is also coupled to coil **3**.

Coil **5** and capacitor **6** are tuned to the frequency of a radio transmitter from which it is desired to receive information—for instance, an amplitude-modulated standard broadcast station. The signal received from this transmitter need not be strong. The signal is detected by diode **15** to obtain an audio-frequency information signal. This audio signal is coupled through a capacitor **16** and is amplified by a circuit that includes a transistor **17** having a base **20**, an emitter **21**, and a collector **22**. The amplified audio output of the transistor is coupled through an audio transformer **23** to an electroacoustical transducer, preferably a permanent-magnet dynamic loudspeaker **25** as shown.

The novelty of the invention lies largely in the method and means by which the necessary direct-current power is supplied to the emitter and collector circuits of transistor **17**. This method and means will now be described.

Coil **7** and capacitor **10**, and also coil **11** and capacitor **12**, are tuned to receive radio signals of relatively high strength. It does not matter whether these signals contain information. These power signals are rectified by diodes **26** and **27** to provide direct-current power that is filtered by capacitors **30** and **31**. The D.-C. power thus obtained is utilized to power the transistor **17**.

In the circuit shown, two tuned circuits (coil **7** and capacitor **10**, and coil **11** and capacitor **12**) are tuned to power signals and the D.-C. voltages obtained from each are connected in series. The tuned power circuits may be tuned to the same or different power signals. Under certain circumstances it may be desirable to use more than two tuned power circuits and to tune them to more than two power signals; in this way power can be obtained from several signals and combined. On the other hand, if a strong power signal is available, a single tuned power circuit may suffice to give the needed D.-C. power.

Even weak information signals can be received successfully. A plurality of transistor amplifier stages can be used if desired, or other circuits such as superheterodyne circuits can be used. It is merely necessary that a sufficiently strong power signal or signals be available to provide the small amount of D.-C. needed to power the transistors.

If the information signal happens to be strong, it can be used as the power signal; all of the tuned circuits (coil **5** and capacitor **6**, coil **7** and capacitor **10**, coil **11** and capacitor **12**) are tuned to the information signal.

Engineers who have observed my invention in operation have been surprised at the unexpectedly good results obtained, even with readily available power signals of quite moderate strength. For instance, sufficient power for satisfactory operation of a loudspeaker at low volumes is

readily obtained from a 5-kilowatt standard broadcast station 5 miles away, using only an indoor antenna to pick up the power signal as well as information signals. In typical operation under these conditions a D.-C. voltage of about 2.5 to 3 volts is obtained between the emitter and the collector, at a current of about 250 microamperes; D.-C. power input to the transistor is thus of the order of 0.5 to 1 milliwatt. So far as I am aware, no one has ever before discovered and demonstrated the practicability of this method of powering a radio receiver.

Because existing broadcast stations within a radius of a number of miles provide adequate power signals, the invention is readily practicable with existing power signals in almost any location in or near any city in the United States.

Although I have described a transistor radio receiver powered by my invention, it will be readily apparent that the invention is applicable to the powering of any transistor circuit using one or a number of transistors, and to the powering of other devices requiring relatively small amounts of power. For instance, sensitive electromechanical, electrochemical, or electrothermal devices can be operated by the method of the invention.

Referring to Figure 2, which shows a more general embodiment of my invention, an antenna **35** picks up radiofrequency energy from the atmosphere. This energy flows through coil **36**, which is coupled to a tuned circuit consisting of coil **37** and capacitor **40**. The radiofrequency voltage across capacitor **40** is rectified by diode **41** and filtered by a low-pass filter **46** consisting of capacitors **42** and **44** and choke coil **43**. The resulting D.-C. voltage is applied to a load **45**.

In the practice of my invention, larger amounts of power can be obtained for short periods of time by storing received energy in a suitable energy storage device. Stored energy may then be withdrawn at intervals at a more rapid rate than that at which it was received and put into the storage device. In this way the invention can be used to provide short pulses of relatively very high electrical energy. This result can be readily obtained by charging a relatively large capacitor with direct current and then discharging the capacitor rapidly into a load when desired. This rapid discharge can be initiated automatically when the voltage across the capacitor reaches a certain level, or it can be initiated when a transistor radio receiver receives a certain information signal.

Higher voltages can be obtained with the invention by means of well known devices for raising D.-C. voltages as shown in Figure 3. The D.-C. voltage output from the capacitor **44** can be used to power a low frequency transistor oscillator **52** whose A.-C. output is raised to a higher voltage level by the transformer **55**. This relatively high A.-C. voltage can then be rectified by a diode **61** and fed to a capacitor **64** to provide a high energy D.-C. source at a relatively high voltage level at the terminals **69** and **70**. If desired, energy can now be withdrawn from the capacitor **64** at intervals in short pulses of high energy at a high voltage level. Pulsed radio transmission is one of the possible uses for this form of the invention. Other uses would be to provide a single relatively powerful pulse needed to actuate an electrothermal or electromechanical device.

As has been indicated above, in many locations and particularly anywhere in or near most American cities, power signals normally present in the atmosphere are readily available for the easy and convenient practice of the invention. However, the invention also has important applications in systems in which the necessary power signal is generated and transmitted specifically for the operation of the particular system. Such systems can, for example, comprise a master station transmitting all the power that is needed for hundreds or thousands of fixed or mobile transistor receivers or other remote devices over a range of many miles. This eliminates the need for hundreds or thousands, as the case may be, of local power supplies. At the same time, such a system has the advantage that all of the remote devices can be simultaneously activated or deactivated at the will of the master station, simply by starting or stopping the transmission of the power signal. In such systems it will often be advantageous to use power signals of frequencies sufficiently high to permit the use of resonant receiving antennas of small physical dimensions for signal pickup at the remote devices. In addition to the power signal, the master station may transmit an information signal on the same or a different carrier.

Certain types of devices powered entirely by received radio waves are of course well known. The well-known "crystal set" of the early days of radio, which used a diode rectifier to demodulate an amplitude-modulated radiofrequency signal, is an outstanding example of such a device. My invention is readily distinguishable from such prior devices, however. In typical prior devices a modulated radiofrequency signal is applied to a diode to obtain unidirectional half-wave pulses whose amplitudes vary with modulation. These pulses are integrated by means of a capacitor to obtain a unidirectional signal the amplitude of which follows the audiofrequency modulation envelope. If the radiofrequency signal is received with sufficient strength the audio signal may have sufficient power to operate headphones or similar utilization device without power amplification; but the signal is utilized for its information content, rather than to supply non-information-containing power.

My invention, on the other hand, entails the utilization of received radiofrequency energy to supply power to at least one pair of circuit points (across capacitor **31** in Fig. 1, for example), such circuit points requiring power solely for its power content and not for any information or modulation it may contain. In other words, my invention entails the utilization of radiofrequency energy to supply power that would otherwise have to be supplied by batteries, generator, or other local power source.

It will be apparent that the embodiments shown are only exemplary and that various modifications can be made in construction and arrangement within the scope of the invention as defined in the appended claim.

I claim:

An electrical device for obtaining a high energy D.-C. source at a high voltage level using energy abstracted from the atmosphere, said device comprising in combination: resonant means for receiving radio waves, first rectifier means for converting said radio waves into first direct current energy, first capacitor means for storing said first direct current energy, an oscillator powered by said direct current energy, said oscillator producing an A.-C. output, transformer means for raising said A.-C. output to an increased voltage level, second rectifier means for converting the A.-C. output of increased voltage level from said transformer into second direct current energy, and second capacitor means for storing said second direct current energy.

References Cited in the file of this patent

UNITED STATES PATENTS

1,540,998	Plauson	June 9, 1925
2,056,824	Cawley	Oct. 6, 1936
2,553,734	Adler	May 22, 1951
2,569,687	Pan	Oct. 2, 1951
2,647,957	Mallinckrodt	Aug. 4, 1953

OTHER REFERENCES

Turner: "A crystal receiver with transistor amplifier," Radio and Television News, January 1950.

March 25, 1969 W. C. BROWN ET AL **3,434,678**

MICROWAVE TO DC CONVERTER

Filed May 5, 1965 Sheet 1 of 3

FIG. 1

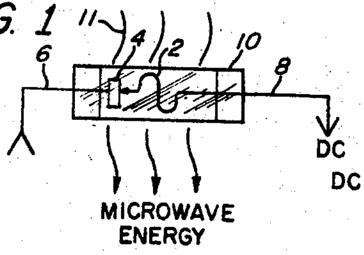
11
4
2
10
6
8
DC
MICROWAVE
ENERGY

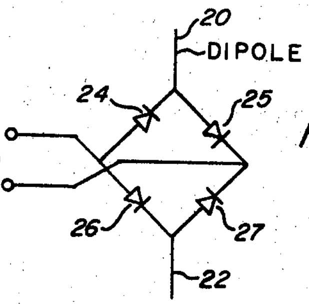
20
DIPOLE
24
25
DC LOAD
26
27
22

FIG. 2

FIG. 3

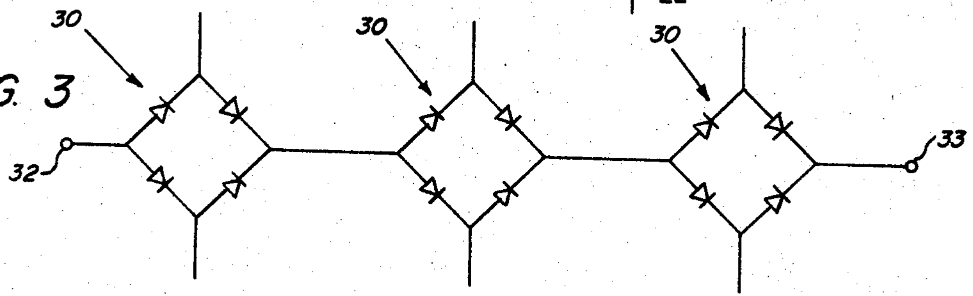
30
30
30
32
33

FIG. 4

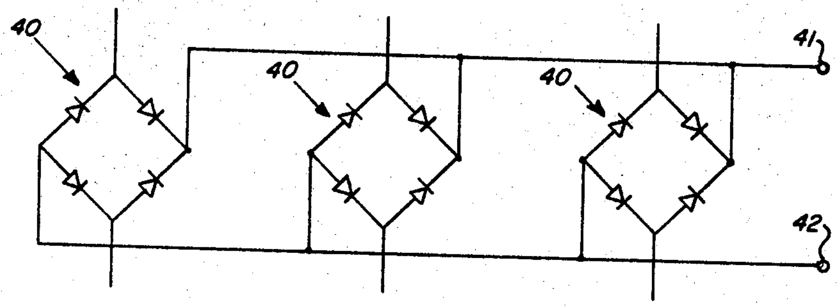
40
40
40
41
42

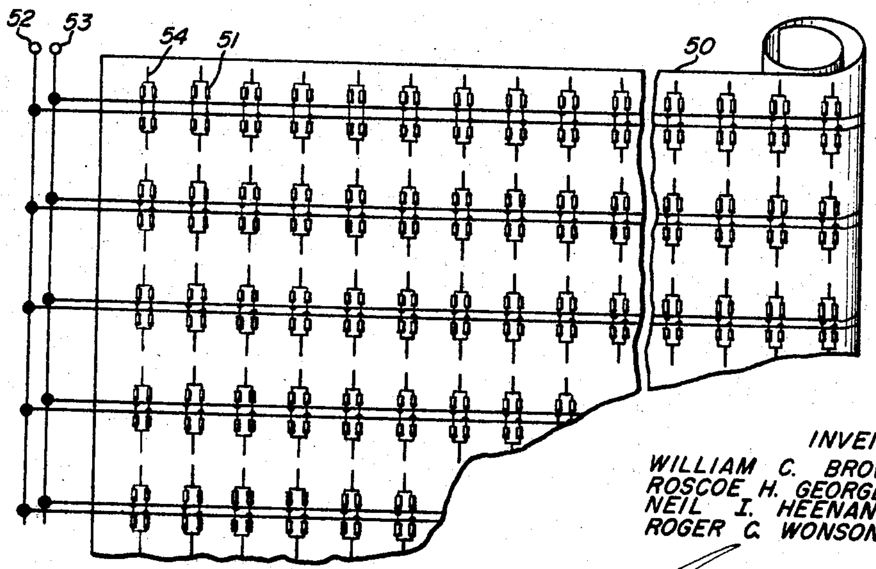
52
53
54
51
50

FIG. 5

INVENTORS
WILLIAM C. BROWN
ROSCOE H. GEORGE
NEIL I. HEENAN
ROGER C. WONSON

BY Edgar O. Rost
ATTORNEY

March 25, 1969 W. C. BROWN ET AL 3,434,678

MICROWAVE TO DC CONVERTER

Filed May 5, 1965 Sheet 2 of 3

FIG. 7

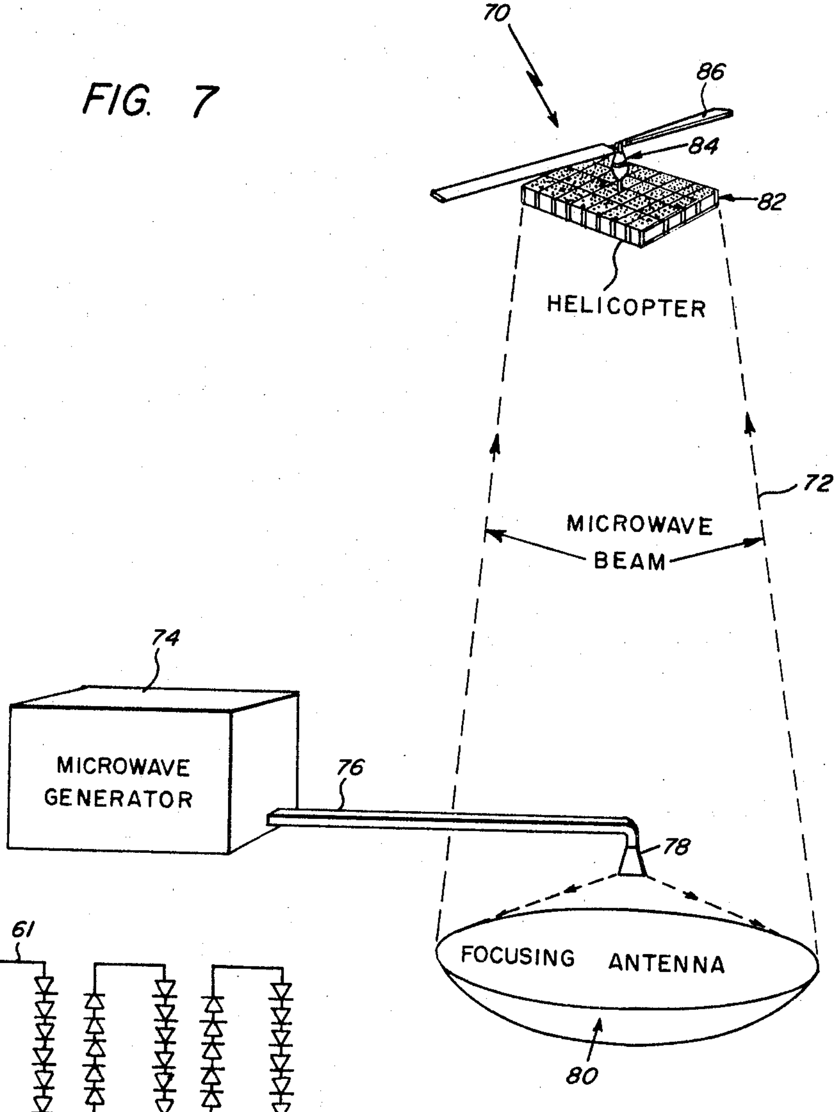
70
86
84
82
HELICOPTER
72
MICROWAVE
BEAM
74
MICROWAVE
GENERATOR
76
78
FOCUSING ANTENNA
80

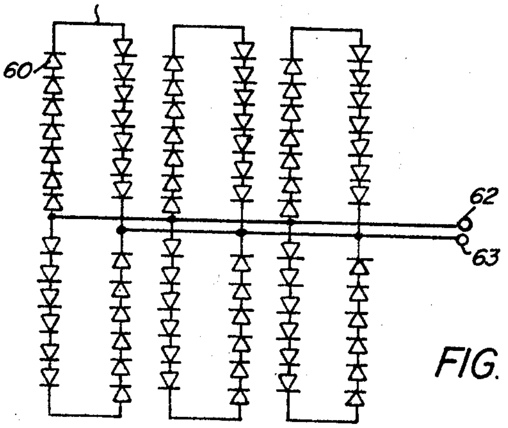

FIG. 6

INVENTORS
WILLIAM C. BROWN
ROSCOE H. GEORGE
NEIL I. HEENAN
ROGER C. WONSON

BY Edgar O. Rost
ATTORNEY

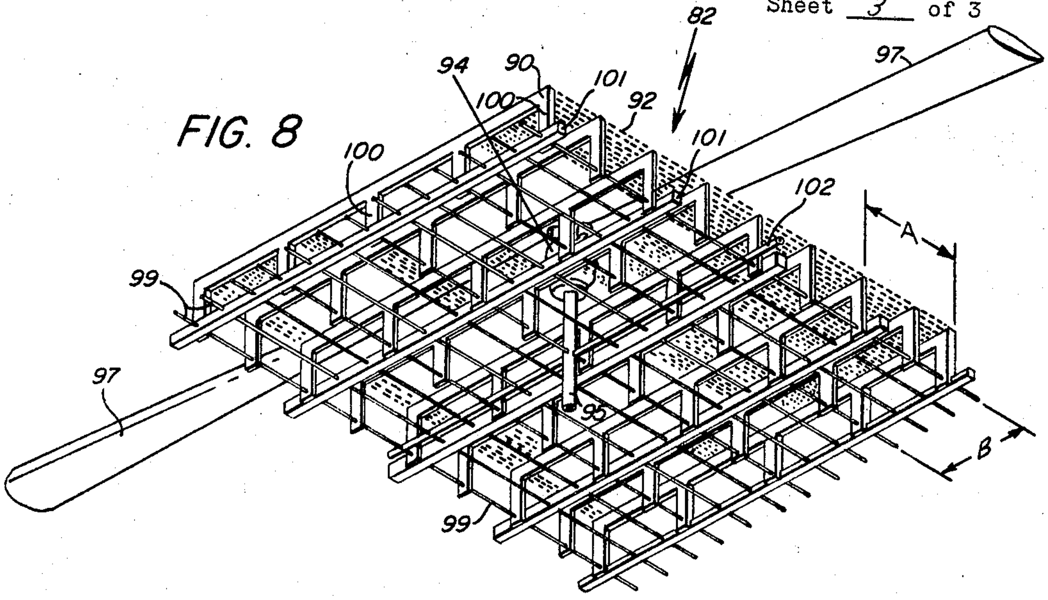
82
94
90
101
92
100
100
101
97
102
A
99
97
95
99
B

FIG. 8

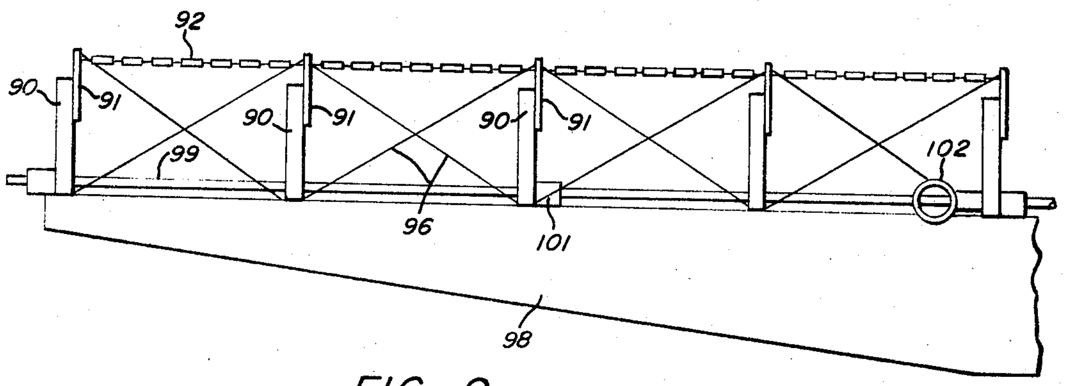
92
90
91
90
91
90
91
99
102
96
101
98

FIG. 9

INVENTORS
WILLIAM C. BROWN
ROSCOE H. GEORGE
NEIL I. HEENAN
ROGER C. WONSON

BY

ATTORNEY

3,434,678
Patented Mar. 25, 1969

3,434,678

MICROWAVE TO DC CONVERTER

William C. Brown, Weston, Mass., Roscoe H. George, West Lafayette, Ind., and Neil I. Heenan, Needham, and Roger C. Wonson, Beverly, Mass., assignors to Raytheon Company, Lexington, Mass., a corporation of Delaware

Filed May 5, 1965, Ser. No. 453,415
Int. Cl. B64g *1/00*
U.S. Cl. 244—1 3 Claims

ABSTRACT OF THE DISCLOSURE

A combined antenna and conversion mechanism for reception of beamed high frequency electromagnetic energy in space including a large array of unidirectional current semiconductor rectifier devices. A self-supporting space vehicle utilizing the rectified DC electrical energy for propulsion is disclosed in an illustrative embodiment.

The present invention relates in general to the transfer of energy by means of an electromagnetic wave beam and more particularly to interception and rectification of such energy into low frequency electrical DC energy with a high degree of efficiency.

Improved technology in the field of microwave energy generation at superpower levels has resulted in the realization of electrical energy transmission over considerable distances for remote energization of devices or vehicles without the aid of wires. The transmission of microwave electromagnetic energy into space has been commonly employed in the radar pulse echo systems for the detection and orientation of desired objects within a predetermined scanning range of a transmitting antenna. Beams of a similar nature may now be employed for other useful purposes and the advantages attendant the utilization of electromagnetic energy in the microwave region in contrast with other wavelengths may now be enumerated.

Microwaves have been generally defined as high frequency radio waves whose wavelength is less than 30 centimeters, with a lower wavelength limit on the order of 1 millimeter sometimes being applied to what is commonly referred to as the "microwave region." The superiority of high frequency microwaves is due in part to the fact that it is generally desirable to focus the transmitted energy so as to achieve a high power density at a remote point or area with respect to a given power source. In accordance with the laws of optics, the sharpness of the microwave beam produced by a transmitting antenna varies as the ratio of antenna dimensions to the wavelength of the transmitted energy. Therefore, for a given or desired power density or beam sharpness, a decrease in the wavelength of the transmitted energy permits a corresponding decrease in the dimensions of the antenna. From the standpoint of mechanical considerations, it is desirable to employ small antennas and other components, and it is therefore advantageous to employ high frequency energy of very short wavelength. In addition, the difficulties encountered in long wave transmission as a result of natural and man-made interference or noise do not occur with any appreciable significance at microwave frequencies. Further, in aerospace applications with considerable distances separating the transmitter at an earth or mother planet location and the employment of shorter wavelength beamed energy is preferred since longer wave signals will generallly be reflected at certain altitudes by reflecting layers in the atmosphere.

In view of certain losses due to absorption which may occur in the atmosphere, microwaves in the region having the approximate bounds of 2 and 30 centimeters are readily adaptable to the convenient radiation of power to remote points without the utilization of wires. The preferred wavelengths are of the order of 5 or 10 centimeters to provide efficient focusing with existing transmitting antenna systems which may be maintained at a reasonable size. An illustrative device of the superpower high frequency microwave generators operative in the desired band is the so-called Amplitron which is an amplifier having a broad bandwidth and excellent performance characteristics for the focusing of the beam. Such devices are capable of producing 15 or 20 kilowatts of average continuous wave power in the neighborhood of 10 centimeters in wavelength with capabilities expected in the region of 500 kilowatts or more average power with 50 megawatts peak power. A complete description of such devices may be had by referring to Patent No. 2,933,723 issued Apr. 19, 1960 to William C. Brown and assigned to the assignee of the present invention.

With microwave energy capable of being generated and directed over longer distances conversion of such high frequency electromagnetic energy is of paramount concern. One conversion mechanism in the prior art involves direct conversion of such energy into heat which may then be utilized directly or indirectly for propulsion or generation of flight-producing forces. Examples of such devices for heat energy exchange as well as space vehicles utilizing such energy may be noted in Patent No. 3,174,-705, issued Mar. 23, 1965, to D. Schiff et al., as well as U.S. Letters Patent No. 3,083,528, issued Apr. 2, 1963 and No. 3,114,517, issued Dec. 17, 1963, to William C. Brown. The heat exchanger method of conversion of electromagnetic energy into useful power is limited by the overall efficiencies of approximately 25 percent in the conversion of heat into mechanical or electrical work. Desirable, therefore, would be the direct rectification of the high frequency electromagnetic energy into low frequency electrical energy for the operation of many useful aerospace devices as well as systems.

The present invention has for its primary object the conversion of high frequency electromagnetic energy in the microwave region directly into low frequency electrical energy.

A further object of the present invention is the provision of a combined nondirectional receiving antenna and microwave electromagnetic energy to low frequency electrical energy conversion means in a unitary structure.

A still further object of the present invention is a provision of a new and novel combined nondirectional receiving antenna and microwave to DC energy converter for aerospace applications.

Another object of the present invention is the provision of a new and novel nondirectional receiving antenna and microwave to DC energy converter having a high degree of efficiency.

Still another object of the present invention is the provision of a new and novel aerospace vehicle with nondirectional receiving antenna and microwave to DC energy converter means with said vehicle being capable of being supported by its own energy generation means at a distance spaced apart from the power generation means.

In accordance with the teachings of the present invention, the above and other objects are achieved by the employment of efficient unidirectional microwave power rectifiers and dipole antenna means. Such rectifying devices, while being individually limited in power-handling capabilities, normally in the order of fractions of watts, have been found to be highly efficient means for the rectification of microwave power when assembled in large numbers in various arrays. It is interesting to note that the observed collective efficiency was on the order of 40 to 70 percent. In an illustrative embodiment, point-

contact semiconductor diodes were arranged in four arm bridge connected networks with the networks interconnected in various configurations such as series, parallel and series-parallel.

In discussing aerospace applications, an additional problem is encountered in the beaming of microwave energy to a remote point and the interception and utilization of such electrical energy. In such applications the advantages of a vehicle which may be maintained in space for indeterminate periods of time without employing a local fuel source are readily apparent. Such devices could readily provide communication networks, surveillance functions using radar techniques along with numerous other functions. The acapture of the beamed high frequency electromagnetic energy raises the need for an efficient antenna means capable of intersecting the beam at high altitudes. Conventional techniques employed in microwave radar usage such as receiving antenna horns are capable of intersecting only a small portion of the beam energy and add considerable weight in applications involving heavier-than-air vehicles. In an exemplary embodiment of the invention a space vehicle, namely a helicopter, is disclosed for either moving flight or a stationary location with self-supporting electrically operative propulsion means. The semi-conductor diode rectifier arrays have been demonstrated to fulfill the receiving antenna functions as well as the electrical energy rectification means in a highly efficient manner. Such combined antenna and rectifier means has also assisted in reduction of the weight problem in airborne devices. Further, it has provided a nondirectional means for the interception of the microwave energy to thereby reduce the problems of focusing inherent in prior art directional horn type receiving antennas.

With the above features, advantages and objects in mind the invention will now be described by reference to the following detailed description together with the accompanying drawings in which:

FIG. 1 is a perspective view of an illustrative diode rectifier;

FIG. 2 is a schematic circuit diagram of a bridge connected diode network with dipole antenna means;

FIG. 3 is a schematic circuit diagram of a plurality of bridge connected networks arranged in series;

FIG. 4 is a schematic circuit diagram of a parallel bridge connected network array;

FIG. 5 is a perspective view of an illustrative embodiment of a combined antenna and rectifier array in a folded or rolled up configuration;

FIG. 6 is a schematic circuit diagram illustrating the bridge connected diode array incorporated in the aerospace vehicle shown in FIG. 7;

FIG. 7 is a schematic representation in elevation illustrative of a heavier-than-air aerospace vehicle incorparating the structure of the present invention;

FIG. 8 is a perspective view of the aerospace vehicle embodiment as viewed from the under portion thereof; and

FIG. 9 is an enlarged partial view in elevation of a portion of the illustrative embodiment shown in FIG. 8.

FIG. 1 illustrates a point-contact semiconductor diode rectifier of the type employed in radar microwave receiver apparatus to rectify returned radar pulses. Any of the high burnout semiconductor diodes having high rectification characteristics are preferred and are commercially available, such as the 1N82 or 1N830. The rectifying junction is formed by whisker element **2** contacting the semiconductor element **4** respectively connected to leads **6** and **8**. Silicon is preferred over germanium for element **4** because of its ability to operate at higher temperatures and thereby handle higher powers. Envelope **10** houses the rectifying elements and may be of a hermetically sealed dielectric material or combination metal and ceramic composition. The inherent characteristic of such diode rectifiers is that the microwave energy is intercepted and rectified in a unidirectional manner and the line **11** indicate pictorially the rays of the beamed electromagnetic microwave energy in a plane normal to the envelope. In FIG. 2 a full-wave bridge connected diode network is illustrated with the forward direction of the rectified DC electric current indicated by the direction of the arrow symbols. The network shown consists of half-wave dipoles **20** and **22** each terminated with a diode rectifier element **24** to **27** in an arm of the bridge connected network. The dipole elements **20** and **22** are of the half-wave configuration and may be spaced apart from each other a one-half wavelength at the frequency of the beamed electromagnetic energy.

Referring now to FIG. 3, an array of bridge connected diode networks each with the half-wave dipoles are shown connected in series. Each network is referred to by the numeral **30** and is similar in the bridge connections to the single element network shown in FIG. 2. The DC output of the collective rectified energy is coupled by means of terminals **32** and **33**. In FIG. 4, a similar number of individual bridge connected diode-dipole networks are shown connected in a parallel array. Each network is indicated by the numeral **40**, and the output terminals are indicated as **41** and **42**.

Any number of diode-dipole networks may be provided and in FIG. 5 such a multi-element array is illustrated by mounting on a flexible material **50** which may be rolled or folded into any desired package or enclosed within a capsule to be launched and released at a predetermined point in space. Any flexible material which is pervious to electromagnetic energy is preferred. The total power desired would be the determining factor in a number of individual diode-dipole elements required. In this embodiment, the bridge connected networks **51** are connected in parallel to the output load indicated by terminals **52** and **53**, and representative measurements of electrical characteristics have shown that approximately five watts of DC electrical energy is realizable for each square foot of area of the combined antenna-rectifier. While the dipole elements **54** have been indicated in a particular array, it is within the scope of the invention to stagger the placement of such dipoles to increase the over-all efficiency of the antenna-rectifier.

To further increase the DC powder output, the full-wave bridge connected networks are preferably arranged with a plurality of diodes in series in each arm of the bridge. An illustrative schematic circuit diagram of such a configuration is shown in FIG. 6 wherein seven diodes **60** are shown in each arm of the bridge circuit and are connected in series for a total of twenty-eight diodes in each bridge network. The dipole members will then be the substantially U-shaped end portions **61** at the ends of each brace of seven diodes. In the illustration three such twenty-eight diode bridge networks are shown connected in parallel to terminals **62** and **63**. This closer spacing and compact arrangement has been shown to be a source of improved power output and is capable of a high degree of reliability through the redundant nature of the parallel series connections within each bridge network. If one of the diode rectifiers fails to function the over-all voltage drop across this element would be divided among the six remaining diode rectifiers. If any of the connecting wires between the diode elements should break, the adjacent arms of the other bridge assemblies would take the additional load due to the close proximity of the respective arms to each other. In addition, it is possible to have a number of open connections or inoperative diodes dispersed throughout the array without any serious impairment in performance.

In relation to the array concept to be hereinafter described it may be stated that within a six inch square area ten such individual bridge networks each containing twenty-eight diode rectifiers for a total of 280 diode rectifiers may be deployed in such a manner as to provide

maximum exposed area for each diode as well as the connecting leads. Such an arrangement will be hereinafter referred to as a "module" and a DC output in excess of fourteen watts has been measured for such a module. Any number of such modules could be connected provided for a desired power yield and this module concept readily lends itself to use in certain aerospace applications now to be described.

In FIG. 7 a propelled type of space vehicle **70** is shown wholly supported by means of the transfer and rectification of continuous wave electromagnetic energy via a microwave beam **72**. The source of the microwave energy which may be of the Amplitron type device as described in the aforementioned issued Patent No. 2,933,723 is indicated as **74**. This energy is fed by waveguide means **76** to a transmitting horn **78** to illuminate an ellipsoidal beam forming focusing antenna **80** for the transmission of the microwave beam **72**. It will be appreciated by those skilled in the art that the representations of the microwave generation and transmitting antenna means are pictorial representations to illustrate the usage of the invention in diagrammatic form and the present invention is not limited to any particular source of microwave energy or transmitting antenna assembly. It may be stated the reflector of the antenna assembly is considerably larger than most of the reflectors of the prior art in order to focus a large amount of the microwave power at high altitudes for use in the transfer of energy to space vehicles. Such antenna assemblies may be partially supported in a large hollowed area on the earth's surface or other convenient means of support.

The space vehicle or helicopter **70** can be described as a main body member supporting antenna-rectifier means **82** including a large number of the so-called modules connected together and rigidly supported in a planar parallel array. A motor **84** is supported by the combined body member and the receiving antenna-rectifier means and actuates the rotor **86** of conventional design employed in such self-propelled hovering vehicles. The disclosed vehicle provides for the illumination of the planar array of the semiconductor diode dipole elements by the microwave beam and the direct conversion of the microwave power transmitted by the beam into usable electrical energy for the self-propulsion of the device without any local fuel supply being required.

FIGS. 8 and 9 illustrate a space vehicle **82** comprising a plurality of the combined receiving antenna-rectifier module means for interception and rectification of the electromagnetic microwave energy beam emanating from an earth or mother planet source. A planar array of the antenna-rectifier modules is mechanically supported by means of structural members **90** of any lightweight wood or metal. Insulators **91** positioned coextensive with the members **90** support the diode rectifier array and avoid interference with the receiving and electrical performance characteristics by the structural support members. Carrying forward the module concept of 280 diode rectifiers to provide an approximate power output of 14 watts, it was noted that any number of such modules may be coupled together since the individual module outputs are relatively insensitive to a wide range of load resistances connected to the common output terminals. To achieve the desired electrical output of approximately 120 volts and 250 watts of power, subgroups of four modules each were assembled and parallel connected with an approximate 30 volts available for each subgroup. Four such subgroups were series-connected to result in a total of 4,480 diode rectifiers or 16 modules assembled in a two foot square self-supporting planar parallel array structure. The individual diode rectifiers connected in each arm of the bridge network are indicated by the numeral **92**. An exemplary module configuration would extend within the area delineated by the dotted lines and reference letter A on one side and similar dotted lines and reference letter B on the other side.

A motor **94** is connected to the DC side of the overall array and may be additionally supported by tubular member **95**. A shaft and propulsion means consisting of rotor blades **97** provide for the upward lift of the overall vehicle for the self-supporting of same in space applications. Additional structural support such as interlaced rigging **96** of a high tensile strength material such as nylon or steel wire, as well as bracing member **98**, may be employed for strengthening of the body means to withstand the vibrational forces and downwash from the propulsion means.

In accordance with the well known technology of microwave transmission the combined array of diode rectifiers and propulsion means presents a specific load impedance which must be suitably matched to the transmitted microwave energy beam to result in maximum efficiency. In aerospace applications a mismatch of approximately ten to one may be evident. Matching of the load impedance to a value of approximately 377 ohms as the free space value will be provided by a plurality of coplanar parallel metallic rod members **99** disposed in a grating array in front of the diode rectifiers a predetermined distance. Rod members **99** are linearly disposed and extend in a similar direction as the assembled diode rectifiers. A selected frontal spacing of one-quarter of the wavelength of the microwave frequency being transmitted has been experimentally determined to be suitable for impedance matching purposes. An approximate spacing of two inches between the respective members was preferred for a selected microwave frequency of 2,450 megacycles. Each of the members **90** are provided with lateral sections **100** to support the elongated bar members **101** which in turn maintain the rod members **99** in the desired position. A tubular member **102** of a lightweight metal may also be provided to combine with the motor support member **95** for structural support.

The combined antenna-rectifier array provides a source of electrical energy to render any space vehicle self-supporting. The diode rectifier elements when assembled in the antenna array have been found to be nondirectional with respect to interception of the beamed microwave energy. This represents a large step forward in the utilization of high power microwave energy over the prior art horn-type receiving antennas which must be accurately focused and pointed in a particular direction for the reception of any energy. The connections between the respective members of the diode rectifier array and deployment in the parallel configuration serves to provide maximum exposed area. Such connections and in particular the end loop portions adjacent the terminus of each arm of the bridge networks serves as an efficient dipole for the interception of the microwave energy.

Although it is not intended as a full explanation of the high degree of efficiency attained with the disclosed antenna-rectifier array, it is believed that the whisker elements within the semiconductor diodes themselves are a contributing factor and may function as additional dipole elements. The disclosed embodiment functioned efficiently when illuminated by microwave energy generating a vertically polarized beam. Hence, an efficient and light weight energy conversion apparatus is disclosed which may be self-supporting without the requirement of a large local fuel supply payload.

It may be within the purview of the invention to use the available rectified electrical energy for performing many functions in addition to the actuation of the propulsion means. Hence, communications' payloads may be maintained at predetermined positions in space in a hovering attitude utilizing a portion of the electrical energy available. Relay signals to other such vehicles or return signals to ground stations would then be within the realm of possibility. Such available energy may also be employed for servomechanisms, stabilizing and countertorque systems for the navigation of such vehicles.

The electrical efficiencies realized with the combined

receiving antenna and rectifier means have also provided certain weight advantages over other energy converters in aerospace applications. Examples of such converters would be heat exchangers or solar cells. In comparison to the present invention where five to eight pounds per kilowatt of energy realized is a normal characteristic, other energy conversion means weigh in the vicinity of 150 pounds per kilowatt of realizable energy. The inherent advantages of the present invention are therefore apparent. While the technology in the diode rectifier art is being continually advanced, new diode power rectifiers as well as integrated circuit techniques are readily available to future configurations of the present invention. The so-called Schottky barrier diodes could be employed to produce combined antenna-rectifier means weighing even less than two pounds per kilowatt of available energy.

Although the foregoing detailed description has referred to DC power rectification it will be evident that with suitable circuit components low frequency AC energy may also be made available. In addition, other propulsion means may be readily substituted using electrical energy. The embodiments disclosed herein are illustrative only and other modifications or alterations will be apparent to those skilled in the art which do not depart from the scope of the broadest aspects of the present invention as defined in the appended claims.

What is claimed is:

1. A space vehicle comprising:

body means;

said body means including spaced structural support members;

combined antenna and DC electrical energy rectification means for the interception and rectification of incident high frequency electromagnetic microwave energy carried by said support members in a planar parallel array;

said rectification means comprising a plurality of four arm full-wave bridge connected rectifier circuit networks each having a plurality of unidirectional semiconductors in each arm;

said networks being electrically interconnected to common output terminals;

electrically operable propulsion means comprising a motor and rotor members carried by said body means and connected to said terminals for the utilization of said rectified DC energy; and

means for matching the load impedance of said combined antenna and electrical energy rectification means to the incident microwave energy.

2. A space vehicle according to claim **1** wherein said load impedance matching means are arranged in a coplanar array coextensive with said antenna and energy rectification means array, and spaced therefrom a distance of approximately one-quarter of a wavelength at the frequency of the microwave energy.

3. A space vehicle according to claim **2** wherein said load impedance matching means comprise a plurality of parallel disposed elongated metallic members.

References Cited

UNITED STATES PATENTS

1,217,149	2/1917	Caldwell	321—27 X
2,165,055	7/1939	Kafka	321—27
2,444,458	7/1948	Master	321—8
2,927,321	3/1960	Harris	343—6.8 X
3,098,971	7/1963	Richardson	325—592 X
3,174,705	3/1965	Schiff et al.	244—1

RODNEY D. BENNETT, *Primary Examiner.*

MALCOMB F. HUBLER, *Assistant Examiner.*

U.S. Cl. X.R.

307—151; 318—16; 321—27; 325—494; 343—100

March 25, 1969 W. C. BROWN ET AL 3,434,678

MICROWAVE TO DC CONVERTER

Filed May 5, 1965

Sheet 1 of 3

FIG. 1

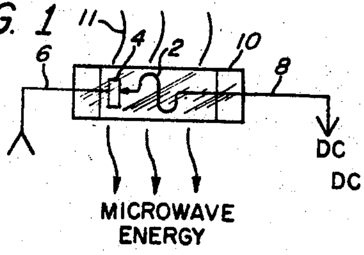
11
4
2
10
6
8
DC
MICROWAVE ENERGY

FIG. 2

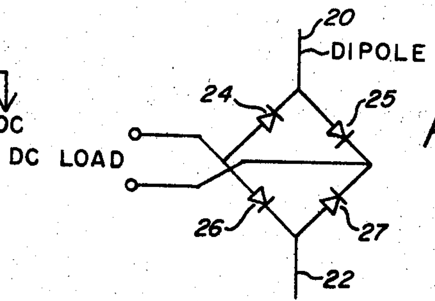
20
DIPOLE
24
25
DC LOAD
26
27
22

FIG. 3

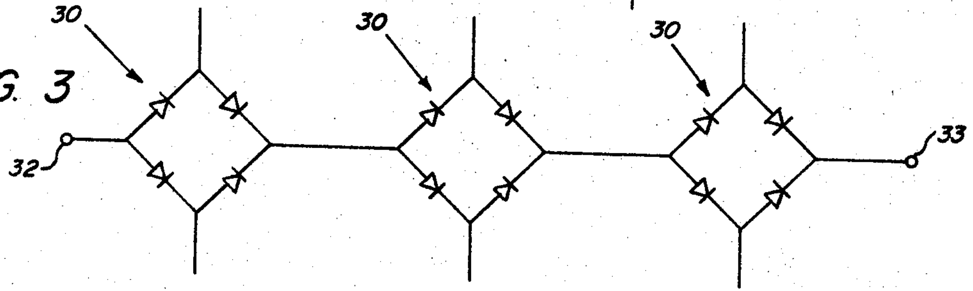
30
30
30
32
33

FIG. 4

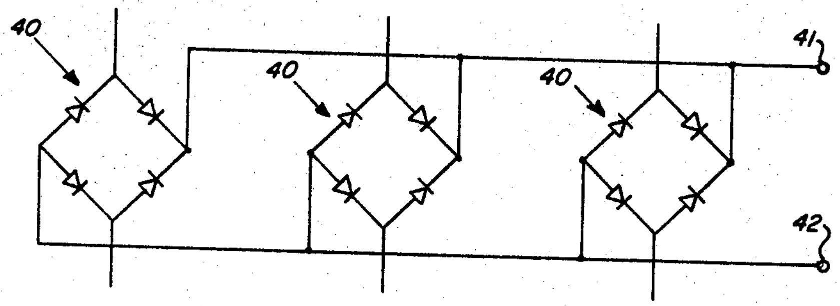
40
40
40
41
42

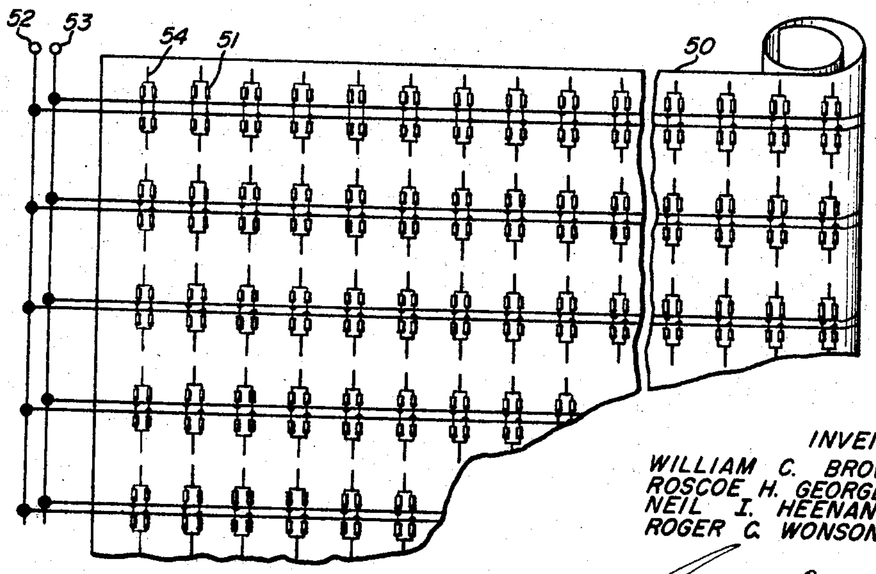
52
53
54
51
50

FIG. 5

INVENTORS
WILLIAM C. BROWN
ROSCOE H. GEORGE
NEIL I. HEENAN
ROGER C. WONSON

BY Edgar O. Rost
ATTORNEY

March 25, 1969 W. C. BROWN ET AL 3,434,678

MICROWAVE TO DC CONVERTER

Filed May 5, 1965 Sheet 2 of 3

FIG. 7

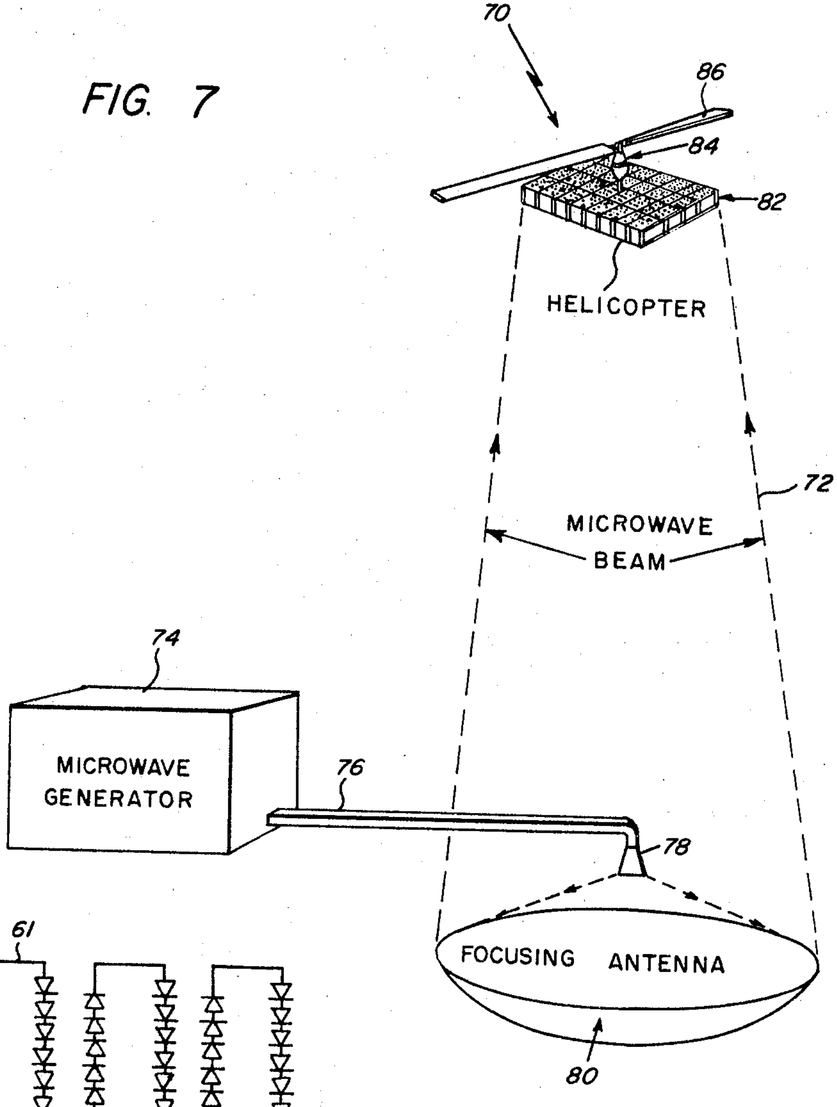
70
86
84
82
HELICOPTER
72
MICROWAVE
BEAM
74
MICROWAVE
GENERATOR
76
78
FOCUSING ANTENNA
80

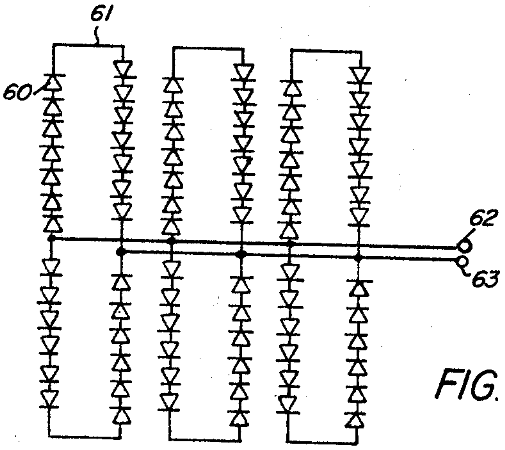

FIG. 6

INVENTORS
WILLIAM C. BROWN
ROSCOE H. GEORGE
NEIL I. HEENAN
ROGER C. WONSON

BY Edgar O. Rost

ATTORNEY

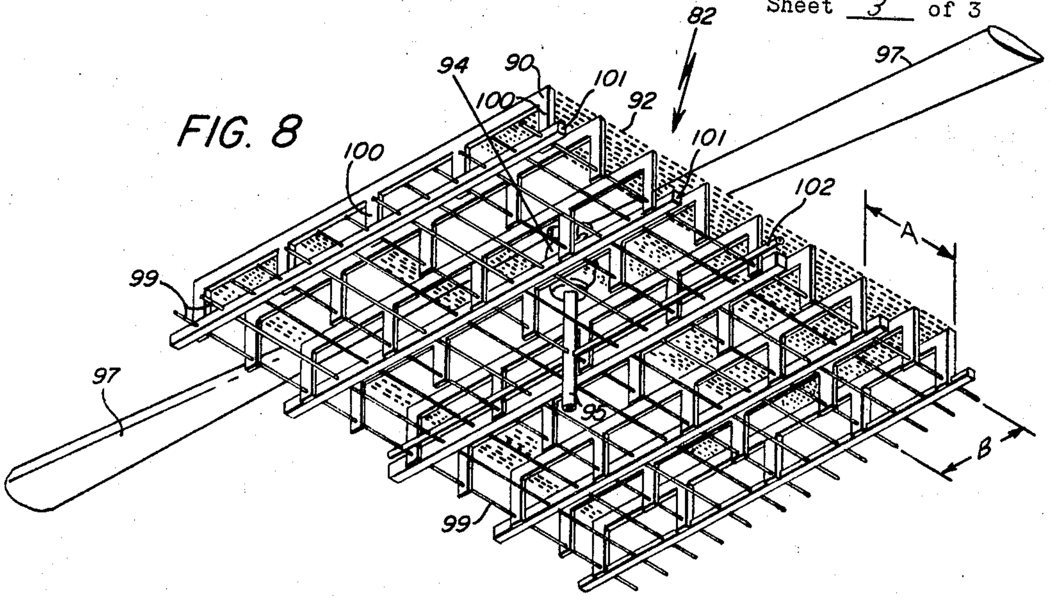
82
94
90
101
100
92
101
100
97
102
A
99
97
95
B
99

FIG. 8

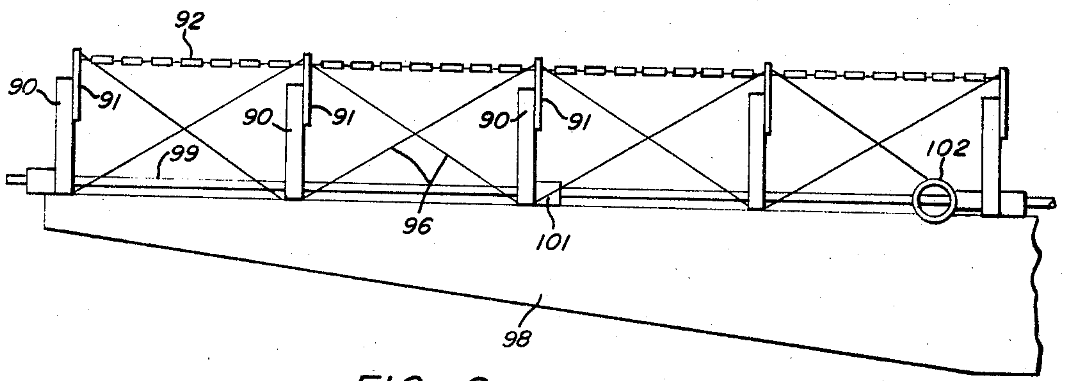
92
90
91
90
91
90
91
99
102
96
101
98

FIG. 9

INVENTORS
WILLIAM C. BROWN
ROSCOE H. GEORGE
NEIL I. HEENAN
ROGER C. WONSON

BY Edgar O. Rost
ATTORNEY

3,434,678
Patented Mar. 25, 1969

3,434,678

MICROWAVE TO DC CONVERTER

William C. Brown, Weston, Mass., Roscoe H. George, West Lafayette, Ind., and Neil I. Heenan, Needham, and Roger C. Wonson, Beverly, Mass., assignors to Raytheon Company, Lexington, Mass., a corporation of Delaware

Filed May 5, 1965, Ser. No. 453,415
Int. Cl. B64g *1/00*
U.S. Cl. 244—1
3 Claims

ABSTRACT OF THE DISCLOSURE

A combined antenna and conversion mechanism for reception of beamed high frequency electromagnetic energy in space including a large array of unidirectional current semiconductor rectifier devices. A self-supporting space vehicle utilizing the rectified DC electrical energy for propulsion is disclosed in an illustrative embodiment.

The present invention relates in general to the transfer of energy by means of an electromagnetic wave beam and more particularly to interception and rectification of such energy into low frequency electrical DC energy with a high degree of efficiency.

Improved technology in the field of microwave energy generation at superpower levels has resulted in the realization of electrical energy transmission over considerable distances for remote energization of devices or vehicles without the aid of wires. The transmission of microwave electromagnetic energy into space has been commonly employed in the radar pulse echo systems for the detection and orientation of desired objects within a predetermined scanning range of a transmitting antenna. Beams of a similar nature may now be employed for other useful purposes and the advantages attendant the utilization of electromagnetic energy in the microwave region in contrast with other wavelengths may now be enumerated.

Microwaves have been generally defined as high frequency radio waves whose wavelength is less than 30 centimeters, with a lower wavelength limit on the order of 1 millimeter sometimes being applied to what is commonly referred to as the "microwave region." The superiority of high frequency microwaves is due in part to the fact that it is generally desirable to focus the transmitted energy so as to achieve a high power density at a remote point or area with respect to a given power source. In accordance with the laws of optics, the sharpness of the microwave beam produced by a transmitting antenna varies as the ratio of antenna dimensions to the wavelength of the transmitted energy. Therefore, for a given or desired power density or beam sharpness, a decrease in the wavelength of the transmitted energy permits a corresponding decrease in the dimensions of the antenna. From the standpoint of mechanical considerations, it is desirable to employ small antennas and other components, and it is therefore advantageous to employ high frequency energy of very short wavelength. In addition, the difficulties encountered in long wave transmission as a result of natural and man-made interference or noise do not occur with any appreciable significance at microwave frequencies. Further, in aerospace applications with considerable distances separating the transmitter at an earth or mother planet location and the employment of shorter wavelength beamed energy is preferred since longer wave signals will generallly be reflected at certain altitudes by reflecting layers in the atmosphere.

In view of certain losses due to absorption which may occur in the atmosphere, microwaves in the region having the approximate bounds of 2 and 30 centimeters are readily adaptable to the convenient radiation of power to remote points without the utilization of wires. The preferred wavelengths are of the order of 5 or 10 centimeters to provide efficient focusing with existing transmitting antenna systems which may be maintained at a reasonable size. An illustrative device of the superpower high frequency microwave generators operative in the desired band is the so-called Amplitron which is an amplifier having a broad bandwidth and excellent performance characteristics for the focusing of the beam. Such devices are capable of producing 15 or 20 kilowatts of average continuous wave power in the neighborhood of 10 centimeters in wavelength with capabilities expected in the region of 500 kilowatts or more average power with 50 megawatts peak power. A complete description of such devices may be had by referring to Patent No. 2,933,723 issued Apr. 19, 1960 to William C. Brown and assigned to the assignee of the present invention.

With microwave energy capable of being generated and directed over longer distances conversion of such high frequency electromagnetic energy is of paramount concern. One conversion mechanism in the prior art involves direct conversion of such energy into heat which may then be utilized directly or indirectly for propulsion or generation of flight-producing forces. Examples of such devices for heat energy exchange as well as space vehicles utilizing such energy may be noted in Patent No. 3,174,705, issued Mar. 23, 1965, to D. Schiff et al., as well as U.S. Letters Patent No. 3,083,528, issued Apr. 2, 1963 and No. 3,114,517, issued Dec. 17, 1963, to William C. Brown. The heat exchanger method of conversion of electromagnetic energy into useful power is limited by the overall efficiencies of approximately 25 percent in the conversion of heat into mechanical or electrical work. Desirable, therefore, would be the direct rectification of the high frequency electromagnetic energy into low frequency electrical energy for the operation of many useful aerospace devices as well as systems.

The present invention has for its primary object the conversion of high frequency electromagnetic energy in the microwave region directly into low frequency electrical energy.

A further object of the present invention is the provision of a combined nondirectional receiving antenna and microwave electromagnetic energy to low frequency electrical energy conversion means in a unitary structure.

A still further object of the present invention is a provision of a new and novel combined nondirectional receiving antenna and microwave to DC energy converter for aerospace applications.

Another object of the present invention is the provision of a new and novel nondirectional receiving antenna and microwave to DC energy converter having a high degree of efficiency.

Still another object of the present invention is the provision of a new and novel aerospace vehicle with nondirectional receiving antenna and microwave to DC energy converter means with said vehicle being capable of being supported by its own energy generation means at a distance spaced apart from the power generation means.

In accordance with the teachings of the present invention, the above and other objects are achieved by the employment of efficient unidirectional microwave power rectifiers and dipole antenna means. Such rectifying devices, while being individually limited in power-handling capabilities, normally in the order of fractions of watts, have been found to be highly efficient means for the rectification of microwave power when assembled in large numbers in various arrays. It is interesting to note that the observed collective efficiency was on the order of 40 to 70 percent. In an illustrative embodiment, point-

contact semiconductor diodes were arranged in four arm bridge connected networks with the networks interconnected in various configurations such as series, parallel and series-parallel.

In discussing aerospace applications, an additional problem is encountered in the beaming of microwave energy to a remote point and the interception and utilization of such electrical energy. In such applications the advantages of a vehicle which may be maintained in space for indeterminate periods of time without employing a local fuel source are readily apparent. Such devices could readily provide communication networks, surveillance functions using radar techniques along with numerous other functions. The acapture of the beamed high frequency electromagnetic energy raises the need for an efficient antenna means capable of intersecting the beam at high altitudes. Conventional techniques employed in microwave radar usage such as receiving antenna horns are capable of intersecting only a small portion of the beam energy and add considerable weight in applications involving heavier-than-air vehicles. In an exemplary embodiment of the invention a space vehicle, namely a helicopter, is disclosed for either moving flight or a stationary location with self-supporting electrically operative propulsion means. The semi-conductor diode rectifier arrays have been demonstrated to fulfill the receiving antenna functions as well as the electrical energy rectification means in a highly efficient manner. Such combined antenna and rectifier means has also assisted in reduction of the weight problem in airborne devices. Further, it has provided a nondirectional means for the interception of the microwave energy to thereby reduce the problems of focusing inherent in prior art directional horn type receiving antennas.

With the above features, advantages and objects in mind the invention will now be described by reference to the following detailed description together with the accompanying drawings in which:

FIG. 1 is a perspective view of an illustrative diode rectifier;

FIG. 2 is a schematic circuit diagram of a bridge connected diode network with dipole antenna means;

FIG. 3 is a schematic circuit diagram of a plurality of bridge connected networks arranged in series;

FIG. 4 is a schematic circuit diagram of a parallel bridge connected network array;

FIG. 5 is a perspective view of an illustrative embodiment of a combined antenna and rectifier array in a folded or rolled up configuration;

FIG. 6 is a schematic circuit diagram illustrating the bridge connected diode array incorporated in the aerospace vehicle shown in FIG. 7;

FIG. 7 is a schematic representation in elevation illustrative of a heavier-than-air aerospace vehicle incorparating the structure of the present invention;

FIG. 8 is a perspective view of the aerospace vehicle embodiment as viewed from the under portion thereof; and

FIG. 9 is an enlarged partial view in elevation of a portion of the illustrative embodiment shown in FIG. 8.

FIG. 1 illustrates a point-contact semiconductor diode rectifier of the type employed in radar microwave receiver apparatus to rectify returned radar pulses. Any of the high burnout semiconductor diodes having high rectification characteristics are preferred and are commercially available, such as the 1N82 or 1N830. The rectifying junction is formed by whisker element **2** contacting the semiconductor element **4** respectively connected to leads **6** and **8**. Silicon is preferred over germanium for element **4** because of its ability to operate at higher temperatures and thereby handle higher powers. Envelope **10** houses the rectifying elements and may be of a hermetically sealed dielectric material or combination metal and ceramic composition. The inherent characteristic of such diode rectifiers is that the microwave energy is intercepted and rectified in a unidirectional manner and the line **11** indicate pictorially the rays of the beamed electromagnetic microwave energy in a plane normal to the envelope. In FIG. 2 a full-wave bridge connected diode network is illustrated with the forward direction of the rectified DC electric current indicated by the direction of the arrow symbols. The network shown consists of half-wave dipoles **20** and **22** each terminated with a diode rectifier element **24** to **27** in an arm of the bridge connected network. The dipole elements **20** and **22** are of the half-wave configuration and may be spaced apart from each other a one-half wavelength at the frequency of the beamed electromagnetic energy.

Referring now to FIG. 3, an array of bridge connected diode networks each with the half-wave dipoles are shown connected in series. Each network is referred to by the numeral **30** and is similar in the bridge connections to the single element network shown in FIG. 2. The DC output of the collective rectified energy is coupled by means of terminals **32** and **33**. In FIG. 4, a similar number of individual bridge connected diode-dipole networks are shown connected in a parallel array. Each network is indicated by the numeral **40**, and the output terminals are indicated as **41** and **42**.

Any number of diode-dipole networks may be provided and in FIG. 5 such a multi-element array is illustrated by mounting on a flexible material **50** which may be rolled or folded into any desired package or enclosed within a capsule to be launched and released at a predetermined point in space. Any flexible material which is pervious to electromagnetic energy is preferred. The total power desired would be the determining factor in a number of individual diode-dipole elements required. In this embodiment, the bridge connected networks **51** are connected in parallel to the output load indicated by terminals **52** and **53**, and representative measurements of electrical characteristics have shown that approximately five watts of DC electrical energy is realizable for each square foot of area of the combined antenna-rectifier. While the dipole elements **54** have been indicated in a particular array, it is within the scope of the invention to stagger the placement of such dipoles to increase the overall efficiency of the antenna-rectifier.

To further increase the DC powder output, the full-wave bridge connected networks are preferably arranged with a plurality of diodes in series in each arm of the bridge. An illustrative schematic circuit diagram of such a configuration is shown in FIG. 6 wherein seven diodes **60** are shown in each arm of the bridge circuit and are connected in series for a total of twenty-eight diodes in each bridge network. The dipole members will then be the substantially U-shaped end portions **61** at the ends of each brace of seven diodes. In the illustration three such twenty-eight diode bridge networks are shown connected in parallel to terminals **62** and **63**. This closer spacing and compact arrangement has been shown to be a source of improved power output and is capable of a high degree of reliability through the redundant nature of the parallel series connections within each bridge network. If one of the diode rectifiers fails to function the over-all voltage drop across this element would be divided among the six remaining diode rectifiers. If any of the connecting wires between the diode elements should break, the adjacent arms of the other bridge assemblies would take the additional load due to the close proximity of the respective arms to each other. In addition, it is possible to have a number of open connections or inoperative diodes dispersed throughout the array without any serious impairment in performance.

In relation to the array concept to be hereinafter described it may be stated that within a six inch square area ten such individual bridge networks each containing twenty-eight diode rectifiers for a total of 280 diode rectifiers may be deployed in such a manner as to provide

maximum exposed area for each diode as well as the connecting leads. Such an arrangement will be hereinafter referred to as a "module" and a DC output in excess of fourteen watts has been measured for such a module. Any number of such modules could be connected provided for a desired power yield and this module concept readily lends itself to use in certain aerospace applications now to be described.

In FIG. 7 a propelled type of space vehicle **70** is shown wholly supported by means of the transfer and rectification of continuous wave electromagnetic energy via a microwave beam **72**. The source of the microwave energy which may be of the Amplitron type device as described in the aforementioned issued Patent No. 2,933,723 is indicated as **74**. This energy is fed by waveguide means **76** to a transmitting horn **78** to illuminate an ellipsoidal beam forming focusing antenna **80** for the transmission of the microwave beam **72**. It will be appreciated by those skilled in the art that the representations of the microwave generation and transmitting antenna means are pictorial representations to illustrate the usage of the invention in diagrammatic form and the present invention is not limited to any particular source of microwave energy or transmitting antenna assembly. It may be stated the reflector of the antenna assembly is considerably larger than most of the reflectors of the prior art in order to focus a large amount of the microwave power at high altitudes for use in the transfer of energy to space vehicles. Such antenna assemblies may be partially supported in a large hollowed area on the earth's surface or other convenient means of support.

The space vehicle or helicopter **70** can be described as a main body member supporting antenna-rectifier means **82** including a large number of the so-called modules connected together and rigidly supported in a planar parallel array. A motor **84** is supported by the combined body member and the receiving antenna-rectifier means and actuates the rotor **86** of conventional design employed in such self-propelled hovering vehicles. The disclosed vehicle provides for the illumination of the planar array of the semiconductor diode dipole elements by the microwave beam and the direct conversion of the microwave power transmitted by the beam into usable electrical energy for the self-propulsion of the device without any local fuel supply being required.

FIGS. 8 and 9 illustrate a space vehicle **82** comprising a plurality of the combined receiving antenna-rectifier module means for interception and rectification of the electromagnetic microwave energy beam emanating from an earth or mother planet source. A planar array of the antenna-rectifier modules is mechanically supported by means of structural members **90** of any lightweight wood or metal. Insulators **91** positioned coextensive with the members **90** support the diode rectifier array and avoid interference with the receiving and electrical performance characteristics by the structural support members. Carrying forward the module concept of 280 diode rectifiers to provide an approximate power output of 14 watts, it was noted that any number of such modules may be coupled together since the individual module outputs are relatively insensitive to a wide range of load resistances connected to the common output terminals. To achieve the desired electrical output of approximately 120 volts and 250 watts of power, subgroups of four modules each were assembled and parallel connected with an approximate 30 volts available for each subgroup. Four such subgroups were series-connected to result in a total of 4,480 diode rectifiers or 16 modules assembled in a two foot square self-supporting planar parallel array structure. The individual diode rectifiers connected in each arm of the bridge network are indicated by the numeral **92**. An exemplary module configuration would extend within the area delineated by the dotted lines and reference letter A on one side and similar dotted lines and reference letter B on the other side.

A motor **94** is connected to the DC side of the overall array and may be additionally supported by tubular member **95**. A shaft and propulsion means consisting of rotor blades **97** provide for the upward lift of the overall vehicle for the self-supporting of same in space applications. Additional structural support such as interlaced rigging **96** of a high tensile strength material such as nylon or steel wire, as well as bracing member **98**, may be employed for strengthening of the body means to withstand the vibrational forces and downwash from the propulsion means.

In accordance with the well known technology of microwave transmission the combined array of diode rectifiers and propulsion means presents a specific load impedance which must be suitably matched to the transmitted microwave energy beam to result in maximum efficiency. In aerospace applications a mismatch of approximately ten to one may be evident. Matching of the load impedance to a value of approximately 377 ohms as the free space value will be provided by a plurality of coplanar parallel metallic rod members **99** disposed in a grating array in front of the diode rectifiers a predetermined distance. Rod members **99** are linearly disposed and extend in a similar direction as the assembled diode rectifiers. A selected frontal spacing of one-quarter of the wavelength of the microwave frequency being transmitted has been experimentally determined to be suitable for impedance matching purposes. An approximate spacing of two inches between the respective members was preferred for a selected microwave frequency of 2,450 megacycles. Each of the members **90** are provided with lateral sections **100** to support the elongated bar members **101** which in turn maintain the rod members **99** in the desired position. A tubular member **102** of a lightweight metal may also be provided to combine with the motor support member **95** for structural support.

The combined antenna-rectifier array provides a source of electrical energy to render any space vehicle self-supporting. The diode rectifier elements when assembled in the antenna array have been found to be nondirectional with respect to interception of the beamed microwave energy. This represents a large step forward in the utilization of high power microwave energy over the prior art horn-type receiving antennas which must be accurately focused and pointed in a particular direction for the reception of any energy. The connections between the respective members of the diode rectifier array and deployment in the parallel configuration serves to provide maximum exposed area. Such connections and in particular the end loop portions adjacent the terminus of each arm of the bridge networks serves as an efficient dipole for the interception of the microwave energy.

Although it is not intended as a full explanation of the high degree of efficiency attained with the disclosed antenna-rectifier array, it is believed that the whisker elements within the semiconductor diodes themselves are a contributing factor and may function as additional dipole elements. The disclosed embodiment functioned efficiently when illuminated by microwave energy generating a vertically polarized beam. Hence, an efficient and light weight energy conversion apparatus is disclosed which may be self-supporting without the requirement of a large local fuel supply payload.

It may be within the purview of the invention to use the available rectified electrical energy for performing many functions in addition to the actuation of the propulsion means. Hence, communications' payloads may be maintained at predetermined positions in space in a hovering attitude utilizing a portion of the electrical energy available. Relay signals to other such vehicles or return signals to ground stations would then be within the realm of possibility. Such available energy may also be employed for servomechanisms, stabilizing and countertorque systems for the navigation of such vehicles.

The electrical efficiencies realized with the combined

receiving antenna and rectifier means have also provided certain weight advantages over other energy converters in aerospace applications. Examples of such converters would be heat exchangers or solar cells. In comparison to the present invention where five to eight pounds per kilowatt of energy realized is a normal characteristic, other energy conversion means weigh in the vicinity of 150 pounds per kilowatt of realizable energy. The inherent advantages of the present invention are therefore apparent. While the technology in the diode rectifier art is being continually advanced, new diode power rectifiers as well as integrated circuit techniques are readily available to future configurations of the present invention. The so-called Schottky barrier diodes could be employed to produce combined antenna-rectifier means weighing even less than two pounds per kilowatt of available energy.

Although the foregoing detailed description has referred to DC power rectification it will be evident that with suitable circuit components low frequency AC energy may also be made available. In addition, other propulsion means may be readily substituted using electrical energy. The embodiments disclosed herein are illustrative only and other modifications or alterations will be apparent to those skilled in the art which do not depart from the scope of the broadest aspects of the present invention as defined in the appended claims.

What is claimed is:

1. A space vehicle comprising:

body means;

said body means including spaced structural support members;

combined antenna and DC electrical energy rectification means for the interception and rectification of incident high frequency electromagnetic microwave energy carried by said support members in a planar parallel array;

said rectification means comprising a plurality of four

arm full-wave bridge connected rectifier circuit networks each having a plurality of unidirectional semiconductors in each arm;

said networks being electrically interconnected to common output terminals;

electrically operable propulsion means comprising a motor and rotor members carried by said body means and connected to said terminals for the utilization of said rectified DC energy; and

means for matching the load impedance of said combined antenna and electrical energy rectification means to the incident microwave energy.

2. A space vehicle according to claim **1** wherein said load impedance matching means are arranged in a coplanar array coextensive with said antenna and energy rectification means array, and spaced therefrom a distance of approximately one-quarter of a wavelength at the frequency of the microwave energy.

3. A space vehicle according to claim **2** wherein said load impedance matching means comprise a plurality of parallel disposed elongated metallic members.

References Cited

UNITED STATES PATENTS

1,217,149	2/1917	Caldwell	321—27 X
2,165,055	7/1939	Kafka	321—27
2,444,458	7/1948	Master	321—8
2,927,321	3/1960	Harris	343—6.8 X
3,098,971	7/1963	Richardson	325—592 X
3,174,705	3/1965	Schiff et al.	244—1

RODNEY D. BENNETT, *Primary Examiner.*

MALCOMB F. HUBLER, *Assistant Examiner.*

U.S. Cl. X.R.

307—151; 318—16; 321—27; 325—494; 343—100

United States Patent [19]

Fletcher et al.

[11] **3,760,257**

[45] **Sept. 18, 1973**

[54] **ELECTROMAGNETIC WAVE ENERGY CONVERTER**

[76] Inventors: **James C. Fletcher**, Administrator of the National Aeronautics and Space Administration with respect to an invention of; **Robert L. Bailey**, Gainesville, Fla.

[22] Filed: **Sept. 27, 1972**

[21] Appl. No.: **292,698**

[52] **U.S. Cl.** **321/1.5,** 136/89, 250/212

[51] **Int. Cl.** **H02m**

[58] **Field of Search** 136/89; 250/212; 333/21; 321/1.5

[56] **References Cited**

UNITED STATES PATENTS

3,026,439	3/1962	Geer	250/212 X
3,121,648	2/1964	Jensen	136/89
3,232,795	2/1966	Gillette et al.	250/212 X
3,350,234	10/1967	Ule	250/212 X
3,427,200	2/1969	Lapin et al	136/89
3,448,273	6/1969	Webb	250/212 X
3,615,853	10/1971	Paine	136/89

Primary Examiner—William M. Shoop, Jr.
Attorney—R. F. Kempf et al.

[57] **ABSTRACT**

Electromagnetic wave energy is converted into electric power with an array of mutually insulated electromagnetic wave absorber elements each responsive to an electric field component of the wave as it impinges thereon. Each element includes a portion tapered in the direction of wave propagation to provide a relatively wideband response spectrum. Each element includes an output for deriving a voltage replica of the electric field variations intercepted by it. Adjacent elements are positioned relative to each other so that an electric field subsists between adjacent elements in response to the impinging wave. The electric field results in a voltage difference between adjacent elements that is fed to a rectifier to derive d.c. output power. The element pairs may be arranged in a two-dimensional array to provide power conversion of randomly polarized electromagnetic waves, such as sunlight.

19 Claims, 3 Drawing Figures

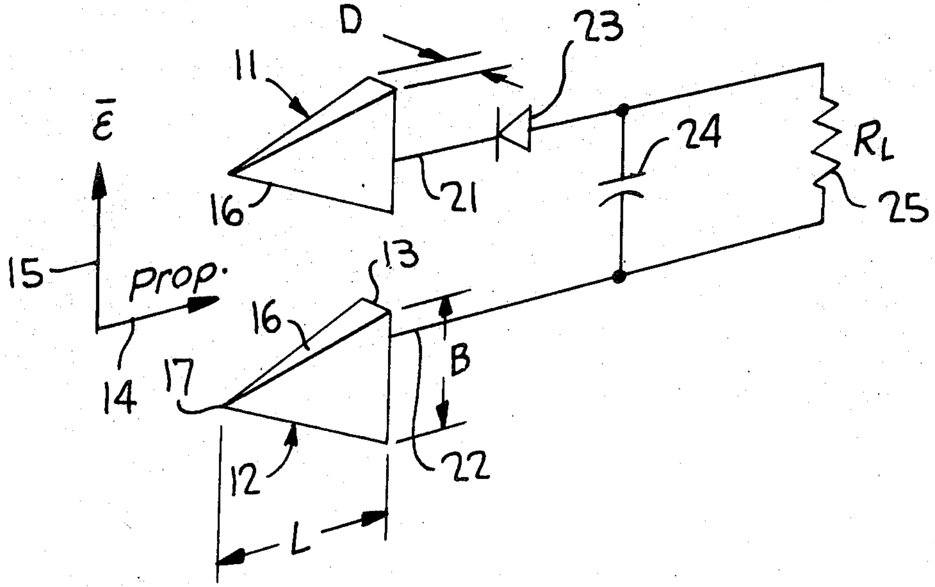

FIG. 1

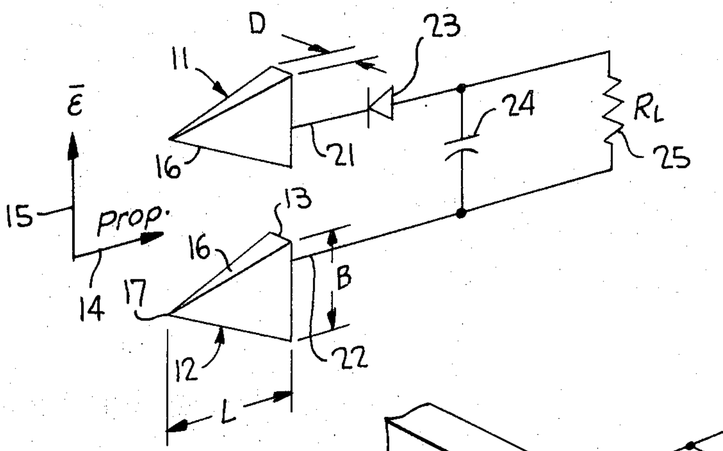
D
11
23
$\bar{\varepsilon}$
24
R_L
16
21
25
15
prop.
13
16
B
14
17
12
22
L

FIG. 2

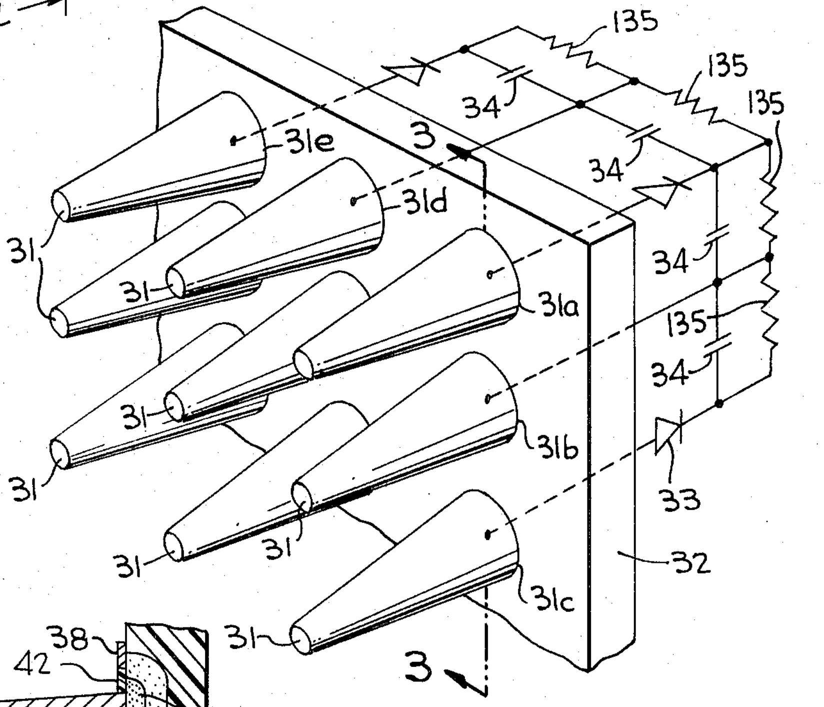
135
135
34
135
31e
3
34
31d
31
31
34
31a
31
135
31
34
31
31b
33
31
31
32
31c
31
3

FIG. 3

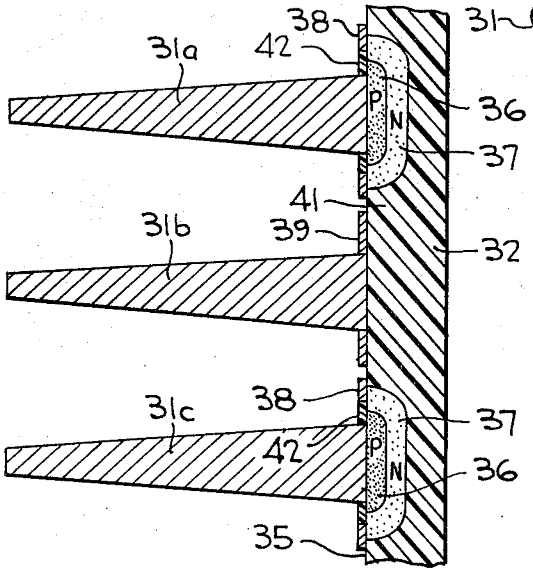
38
31a
42
P
36
N
37
41
31b
39
32
38
31c
37
42
P
N
36
35

ELECTROMAGNETIC WAVE ENERGY CONVERTER

ORIGIN OF THE INVENTION

The invention described herein was made in the performance of work under a NASA contract and is subject to the provisions of Section 305 of the National Aeronautics and Space Act of 1958, Public Law 85-568 (72 Stat. 435; 42 U.S.C. 2457).

FIELD OF INVENTION

The present invention relates generally to devices for converting electromagnetic wave energy into electric power and, more particularly, to a device including a number of relatively closely spaced electromagnetic wave absorber elements having tapered portions responsive to wide band electromagnetic wave radiation.

BACKGROUND OF THE INVENTION

Devices for converting radiant energy into usable power have been considered extensively in the past. Presently, the most commonly utilized devices for converting radiant energy into usable power are solar cells which are adapted to convert solar energy directly into electricity for power generating purposes. Solar cells generally include oppositely doped semiconductor junctions which generate current in response to solar energy impinging thereon. Solar cells depend for their operation on quantum properties of light energy to provide charge separation and hence current flow in the junction region. To date, the maximum efficiency of solar cells in converting solar energy into electric energy is approximately 13 percent. In addition to the relatively low conversion efficiency of solar energy into electric power, solar cells are very expensive, as well as fragile, and usually must be mounted on a rigid, preferably flat substrate. The requirement for a rigid, flat substrate frequently substantially increases the combined cost of the solar cells and their supporting structure; if the supporting structure is a spacecraft panel, problems of folding the panel and maintaining it in a rigid condition in use are encountered.

BRIEF DESCRIPTION OF THE INVENTION

In accordance with the present invention, the wave properties of electromagnetic energy are utilized to convert wide band electromagnetic wave energy into usable electric power. The wide band energy impinges on an array of tapered absorber elements mutually insulated from each other. The length, base area, and separation of the elements are such that relatively wide band energy can be absorbed. To achieve maximum conversion of wide band energy, each element has a length, in the direction of wave propagation, equal to several wavelengths of the longest wavelength energy to be converted. Adjacent elements have spaced, sloping converging sides to enable maximum conversion of energy wavelengths in the spectrum to be converted. A voltage replica of the electric field variations absorbed by each element is derived at an output of each element. The difference between the voltage variations derived from adjacent elements is derived and supplied to a rectifier to provide the power conversion. To maximize the voltage difference, bases of adjacent absorber elements are spaced from each other by less than one wave length and preferably less than ½ wavelength, of the received energy. The close spacing, relatively long length and tapered sides also enable the absorber elements to function effectively as absorbers and prevent substantial re-radiation and reflection of the energy impinging thereon because the wave radiation is trapped in the converging interstices between the abosrbing elements.

In a preferred embodiment, the absorbers are formed of metallic elements, such as copper or other suitable material; however, absorbers may consist of dielectric elements which function to direct the electromagnetic wave energy onto an electromagnetic wave energy-to-electric voltage converter that derives a voltage replica of the wave energy.

The invention may be utilized in conjunction with converting wave energy from the microwave region through the visible light spectrum. If the invention is utilized for wave lengths in the solar spectrum, it is ideally suited for use as a solar energy-to-electric power converter. In such a configuration, the absorber elements and electric components utilized for converting the voltage wave replicas into electric power might be formed utilizing integrated circuit type manufacturing processes.

If the device is utilized for absorbing plane polarized electromagnetic waves, adjacent elements in the direction of the wave electric field are aligned with each other in pairs such that maximum electric field variations are derived between them but adjacent elements at right angles to the electric field are packed as close as possible to each other without electric contact, to maximize the active absorber area. If the electromagnetic wave radiation is circularly or randomly polarized, as in the case of solar energy, a two-dimensional array of elements is provided. The elements of the two-dimensional array are positioned in mutually orthogonal directions having aligned columns and rows, and the elements preferably have symmetrical bases, such as a square or circle. Voltage differences are derived between adjacent pairs in both orthogonal directions to provide maximum conversion efficiency of the circularly or randomly polarized wave energy.

Because the conversion process is in response to the wave properties of the impinging electromagnetic wave energy, rather than the quantum properties of such waves, and because of the small number of loss mechanisms individually and collectively optimal, it is believed possible to achieve conversion efficiencies considerably greater than existing solar cell type devices. A major advantage of the present invention is the separation of the wave absorption means and the conversion means permitting each to be individually optimized for the incident wave electromagnetic power spectrum. Another major advantage is that by suitable choice of geometry for the absorbing elements the device can be made to match the incident radiation spectrum. There is no known means of achieving this desirable result with present art solar cells. Another advantage of the present invention is that it does not utilize temperature-sensitive active semiconductor elements, for the basic absorption process; the only semiconductor components are passive diodes. A further advantage of the present invention is that the substrate on which the elements are mounted can be mechanically flexible, to substantially eliminate many of the problems inherent with prior art fragile solar cells mounted on rigid substrates.

It is, accordingly, an object of the present invention to provide a new and improved device for converting electromagnetic wave energy into electric power.

Another object of the invention is to provide a new and improved device for converting wide band electromagnetic wave energy into electric power with relatively high efficiency, utilizing the wave properties of the energy.

An additional object of the invention is to provide a new and improved device for converting circularly or randomly polarized electromagnetic wave energy into electric power in response to electric field gradients.

A further objet of the invention is to provide a relatively inexpensive device for converting electromagnetic wave energy into electric power, which device is relatively temperature-insensitive and can be mounted on a flexible substrate.

An additional object is to provide a new and improved integrated circuit type device for converting solar energy into electric power.

An additional object is to provide a new and improved means of matching the response of the device to the incident electromagnetic wave spectrum.

The above and still further objects, features and advantages of the present invention will become apparent upon consideration of the following detailed description of several specific embodiments thereof, especially when taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWING

FIG. 1 is a perspective view schematically illustrating the principles of the present invention;

FIG. 2 is a perspective diagram illustrating an embodiment of the invention particularly adapted for converting solar energy into electric power; and

FIG. 3 is a side sectional view of the embodiment illustrated in FIG. 2, along the lines 3—3.

DETAILED DESCRIPTION OF THE DRAWING

Reference is now made to FIG. 1 of the drawing wherein is illustrated a pair of mutually insulated, substantially identical, pyramidal or conical, aligned and relatively closely packed, preferably metallic, as copper or other suitable material, electromagnetic wave absorber elements **11** and **12** which are responsive to plane polarized wide band electromagnetic wave energy that is propagating in a direction indicated by arrow **14** and has an electric field with a gradient in a direction indicated by arrow **15.** Elements **11** and **12** are constructed so that the shapes of every cross section parallel to their bases **13** are similar; the cross-sectional areas decrease for increasing distances from the bases **13** of the elements. Each of elements **11** and **12** includes a co-planar base **13**, which may have any desired configuration, such as a rectangle, square or circle, that lies in a plane at right angles to the direction of wave propagation and is parallel to the electric field gradient **15.** Elements **11** and **12** have a substantial height, L, or length in the direction of wave energy propagation, between their base **13** and apex **17.** The height is generally several (up to approximately seventy five) wave lengths of the longest wavelength of the incident electromagnetic energy. Elements **11** and **12** include adjacent, tapered side surfaces **16** that converge toward base **13** and are positioned relative to each other to intercept the electric field **15** and absorb energy over a wide band of the electromagnetic wave. The apex of absorber elements **11** and **12** may be truncated in which case L is the length of the truncated absorber elements.

Voltage variations derived in elements **11** and **12** in response to the electric field variations of the electromagnetic wave energy impinging on surfaces **16** are transduced or converted into an electric voltage at terminals **21** and **22** of the elements. The voltages at terminals **21** and **22** are replicas of the electric field variations impinging on absorber elements **11** and **12,** whereby the voltage difference between terminals **21** and **22** is proportional to the electric field gradient established by the plane polarized wave between facing surfaces **16** of elements **11** and **12.**

The centerline spacing between adjacent elements **11** and **12** is such that for the incident electromagnetic wave spectrum, maximum voltage between **21** and **22** occurs. It is of the order of a wavelength or less of the incident electromagnetic energy. The present invention responds to this voltage difference as derived over the desired spectrum of the incident wave energy, to convert the wave energy into electric power.

The voltage difference between terminals **21** and **22** is converted into useful d.c. electric power by a rectifier **23** and a filter capacitor **24,** the voltage across which is supplied to a suitable external load, such as resistor **25.** The cathode is connected between the anode of the rectifier diode and terminal **22.** It is to be understood that a full wave rectifier can be employed and that the d.c. voltage developed across external load **25** can be supplied to any suitable device, such as a d.c.-to-a.c. power frequency device.

The breadth,B, and depth,D, of base **13** as well as the length, L, and spacing between the elements **11** and **12** enable the band width to which the elements **11** and **12** are responsive to be adjusted to provide optimum matching to the spectrum of the electromagnetic wave energy impinging on absorber elements **11** and **12.** Close spacing between adjacent elements **11** and **12,** in addition to enabling the electric field gradient to be coupled with the greatest voltage difference to terminals **21** and **22,** enables the electric field to be effectively trapped in the converging interstices between facing, adjacent elements **11** and **12.** Trapping of the electric field variations occurs because elements **11** and **12** are absorptive to the electric electromagnetic wave energy, whereby the energy is not reflected and re-radiated from the elements.

Because the device of FIG. **1** is responsive to a plane polarized electromagnetic wave, it is preferable for the depth, D, of each of the elements **11** and **12** to be relatively narrow, less than ¼ wave length, so that adjacent elements (not shown) can be packed as closely as possible to either side of elements **11** and **12** in the horizontal direction. Close packing of the elements enables maximum conversion of the electromagnetic wave energy into electric power over the entire area of an array which may be fabricated out of a multiplicity of element pairs as illustrated in FIG. **1.** It is to be noted that the device of FIG. **1** is responsive most efficiently to electromagnetic waves polarized so that the E field is in a plane parallel or coplanar with the long dimension of bases **13.** If the electromagnetic wave polarization direction were rotated 90°, there would be no substantial electric field variation between the adjacent elements **11** and **12** and there would be substantially zero

voltage developed between terminals **21** and **22** with a resulting zero conversion of electromagnetic wave energy into electric power.

The power of the electromagnetic wave incident on absorber elements **11** and **12** is the incident power density times the effective area of the two elements and can thereby be stated approximately as: $2(\bar{E}\times\bar{H})BD$, where $\bar{E}$ and $\bar{H}$ are respectively the electric and magnetic fields of the wave energy. The power indicated by the equation is available so that it can be converted into electric power supplied to a load, except for losses in elements **11** and **12**, rectification, and stray losses. The losses of elements **11** and **12** are principally due to skin effect or di-electric losses, while the rectifier losses are the series resistance of the diode and stray losses are of the capacitor, and supporting substrate for elements **11** and **12**. Stray losses can be minimized by locating elements **11** and **12** on a low loss dielectric substrate to minimize loss currents between the abosrber terminals.

The maximum wavelength restriction (about 75λ of the longest wavelength of the received spectrum) is provided to enable the shortest wavelength of the desired spectrum to be trapped between adjacent elements and to provide a relatively large surface area for extracting power from it. If the maximum wavelength restriction, which corresponds approximately with the length of a cone in an eyeball of a mammal (about 75λ of the longest wavelength of the received spectrum), is exceeded by the incident electromagnetic energy spectrum, then the voltage output between terminals **21–22** of the device decreases.

If the device of FIG. **1** is utilized to convert plane polarized microwave electromagnetic wave energy into electric power, the rectifier circuit can be connected to the terminals by a coaxial cable having a center conductor connected directly to terminal **21** and a shielded outer conductor, possibly with a conventional balun, connected to terminal **22**. In the alternative, the rectifier circuit can be connected to elements **11** and **12** by a wave guide excited by the voltage difference between terminals **21** and **22** so that propagation in the wave guide is in one of the transverse electric modes, such as TE_{01}. In such a configuration, upper and lower conducting surfaces of a rectangular wave guide, between which the electric field is developed, are respectively connected to terminals **21** and **22**.

One microwave absorber actually constructed and tested in accordance with the present invention includes a pair of sheet copper pyramidal elements **11** and **12** having a length (L) of 13.3 centimeters, a breadth (B) of 6.3 centimeters, and a depth (D) of 2.0 centimeters. This absorber pair had a center frequency of approximately 475 Mhz and a pass band between approximately 200 and 700 Mhz. Increasing the length of the abosrber elements, without any changes in the base dimensions of the pyramid, lowered the pass band center frequency, as well as the upper and lower cut-off frequencies.

The basic converter structure of FIG. **1** can be modified, as illustrated in FIG. **2**, to respond to circularly polarized electromagnetic waves or randomly polarized electromagnetic waves, as subsist in solar energy. In FIG. **2**, a multiplicity of absorber elements **31** are arranged in a two-dimensional array of aligned orthogonally directed columns and rows. Elements **31** are mutually insulated from each other and mounted on a dielectric substrate **32**, which is preferably flexible but may be rigid. Each of elements **31** preferably has a symmetrical base and cross section parallel to the base, which may be either circular or square, to enable equal voltages to be derived between adjacent ones of elements **31** in the two orthogonal directions in response to orthogonal electric field components at the same wavelength. Thereby, the total array power output of FIG. **2** is insensitive to the polarization direction of the energy impinging thereon and the array can respond to circularly polarized, as well as randomly polarized electromagnetic wave energy.

To convert the electromagnetic wave energy impinging on elements **31** into electric power, adjacent elements are interconnected with each other by a half wave rectifying network similar to that illustrated in FIG. **1**. In particular, aligned vertical elements **31*a***, **31*b***, and **31*c*** are interconnected so that the anodes of rectifier diodes **33**, connected to elements **31*a*** and **31*c*** (which are separated from each other by element **31*b***) are connected directly to the elements, while the cathodes of the rectifier diodes are connected to one electrode of different capacitors **34**, the other electrodes which have a common connection to element **31*b***. Similarly, horizontally aligned elements **31*a***, **31*d***, and **31*e*** are connected to rectifier circuits such that the center element of the triad has a common connection to a pair of capacitors **34** which are connected to the cathodes of diodes **33**, having anodes connected to be responsive to the voltage replicas respectively derived by absorber element pairs **31*a*–31*d*** and **31*e*–31*d***. D. C. voltages developed across capacitors **34** are supplied to a matrix of load resistors **135**, one of which is connected across each of the capacitors.

Elements **31** of FIG. **2** are illustrated as frustoconical structures having upper bases with considerably smaller areas than the bases of the elements that are secured to substrate **32**. The frusto-conical configuration is preferred in certain instances because sharp points, as illustrated in the pyramidal elements **11** and **12**, FIG. **1**, may have a tendency to fracture.

To enable the structure illustrated in FIG. **2** to function as a device for converting solar electromagnetic wave energy, which has the majority of its power in the spectrum from 0.3 to 1.1 microns, into d.c. power, the element dimensions and inter-element spacing must be on the order of a micron, whereby integrated circuit techniques are preferably employed in fabrication. To this end, a solar energy converter in accordance with the present invention may take the form illustrated by the cross-sectional view illustrated in FIG. **3** but not limited to this embodiment. In FIG. **3**, electrically conductive cones **31*a***, **31*b***, and **31*c*** project from surface **35** of substrate **32**. Alternate ones of cones **31**, such as cones **31*a*** and **31*c***, are ohmically connected to a P-doped regions **36** of substrate **32**, which are formed on surface **35** of substrate **32** in super-position with the bases of cones **31*a*** and **31*c***. A junction is formed between P-doped regions **36** and N-doped regions **37** which are formed in substrate **32** beneath P-doped regions **36**. Regions **36** and **37** respectively correspond with the anode and cathode of diodes **33** connected to cones **31*a*** and **31*c***, FIG. **2**. Annular, dielectric oxide films **42** are formed on face **35** over the intersections between the otherwise exposed portions of the junctions between regions **36** and **37**. Oxide films **42** cover the peripheries of P-doped regions **36** and the inner circumference of N-doped regions **37**. On the portion of

P-doped regions **36** not covered by oxide films **42** and the inner radial portion of the films, metal elements **31***a* and **31***c* are placed. The bases of elements **31***a* and **31***c* are ohmically connected to P-doped regions **36.** Annular, metal films **38** are formed on face **35** in superposition with the otherwise exposed regions **37** to form ohmic contacts with the N-doped regions. Metal films **38** are electrically insulated from elements **31***a* and **31***c* by dielectric films **42** which are disposed between metal films and elements. Metal film **39** is formed on face **35** at the base of element **31***b* and is electrically connected to the base. The areas **41** of substrate **32** between adjacent metal films **38** and **39** form dielectrics for capacitors **34,** having electrodes comprised of the metal film portions **38** and **39.** A d.c. load is connected between metal film portions **38** and **39** so that the film portions form load terminals.

The structure of FIG. 3 is formed by diffusing P and N regions **36** and **37** onto surface **35** of substrate **32,** utilizing conventional integrated circuit techniques. Regions **36** and **37** are diffused only onto the areas beneath and slightly to the sides of where cones **31***a* and **31***d* are to be located. Thereafter, annular oxide regions **42** are formed over the exterior portion of P-doped region **36** and over a slight segment of the interior of N-doped region **37** on surface **35.** After oxide regions **42** have been formed, thin metal film contacts **38** are vacuum vapor deposited on the exposed portions of N regions **37.** Simultaneously with the formation of contacts **38** thin, metal film, annular contact **39** is vacuum vapor deposited on surface **35** in the region outside of cone **31***b*. After metal films **38** and **39** have been deposited, cones **31***a*, **31***b*, and **31***d* are suitably attached.

While metal is believed preferable in most instances for the absorber elements, it is to be understood that dielectric elements can also be utilized as elements for directing electromagnetic wave radiation to an optical to d.c. converter diode. The dielectric elements would preferably have the same configuration and dimensions as indicated supra while the optical to d.c. converter diode may be of a type described by Javin in the IEEE Spectrum, October, 1971, page 91.

While there have been described and illustrated several specific embodiments of the invention, it will be clear that variations in the details of the embodiments specifically illustrated and described may be made without departing from the true spirit and scope of the invention as defined in the appended claims.

I claim:

1. A device for converting electromagnetic wave energy having an electric field into electric power comprising a plurality of mutually insulated electromagnetic wave absorber elements each responsive to the electric field, adjacent ones of said elements converging toward each other in the direction of the wave propagation, each element including means for deriving a voltage replica of variations of the electric field intercepted by the element, adjacent ones of said elements being positioned relative to each other so that an electric field subsists between said adjacent elements, and means responsive to the voltage difference between the derived voltage replicas between two adjacent elements.

2. The device of claim **1** wherein each of the absorber elements is metallic.

3. The device of claim **1** where each of the absorber elements comprises a metallic surface and has a length in the direction of wave propagation several wavelengths of the impinging electromagnetic wave.

4. The device of claim **3** wherein the adjacent elements have substantially coplanar bases, and the spacing between the adjacent elements at the bases thereof is of the order of no more than a wavelength of the electromagnetic energy.

5. The device of claim **1** wherein the adjacent elements have substantially coplanar bases, and the spacing between the adjacent elements at the bases thereof is of the order of no more than a wavelength of the electromagnetic energy.

6. The device of claim **1** wherein the conversion means responsive to the voltage difference includes means for rectifying the voltage difference to derive a d.c. voltage.

7. A device for converting circularly polarized or randomly polarized electromagnetic wave energy having an electric field into electric power comprising a two-dimensional array of mutually insulated electromagnetic wave absorber elements each responsive to a component of the electric field, each elements converging toward each other in the direction of propagation of the wave and including means for deriving a voltage replica of variations of the electric field intercepted by the element, adjacent ones of said elements being positioned relative to each other so that an electric field subsists between said adjacent elements, and means responsive to the voltage difference between the voltage replicas derived between two adjacent elements, said pairs being spaced at right angles to each other in the array.

8. The device of claim **7** wherein each of the absorber elements is metallic.

9. The device of claim **7** where each of the absorber elements comprises a metallic surface and has a length in the direction of wave propagation several wavelengths of the electromagnetic wave.

10. The device of claim **9** wherein the adjacent elements have substantially co-planar bases and the spacing between the adjacent elements at the bases thereof of the order of one wavelength or less of the electromagnetic energy.

11. The device of claim **7** wherein the adjacent elements have substantially coplanar bases and the spacing between the adjacent elements at the bases thereof is of the order of one wavelength or less of the electromagnetic energy.

12. A device for converting solar, randomly polarized electromagnetic wave energy in the wave length band from approximately 0.3 to 1.1 microns into electric power comprising a dielectric substrate having a face substantially at right angles to the direction of propagation of the solar electromagnetic wave energy, a plurality of mutually insulated electromagnetic wave, metallic absorber elements mounted on said face, each of said elements including tapered portion extending away from the face in the direction of propagation of the wave energy, each of said elements having a decreasing cross-sectional area as the distance from the face increases and a length in the direction of propagation of the wave several wavelengths of the shortest wavelength of the spectrum, an integrated circuit diode in said substrate connected to be responsive to wave energy intercepted by alternately spaced ones of said

elements, a metallic film on said substrate ohmically connected to the remaining metallic elements, whereby a capacitor is formed in the substrate dielectric between adjacent ones of said metallic films.

13. The device of claim **12** wherein each of the elements includes a base with a symmetrical geometry in contact with said face, each element having a cross section parallel to the base similar to the base.

14. The device of claim **13** wherein said face has a square cross section.

15. The device of claim **13** wherein said face has a circular cross section.

16. The device of claim **13** wherein each of said elements is a frustum.

17. The device of claim **12** wherein the spacing between adjacent ones of said elements is of the order of one wavelength or less of the shortest wavelength of the spectrum.

18. The device of claim **12** wherein the several wavelengths are between 1.25 and 75.

19. A device for converting solar, randomly polarized electromagnetic wave energy in the wave length band from approximately 0.3 to 1.1 microns into electric power comprising a dielectric substrate having a face substantially at right angles to the direction of propagation of the solar electromagnetic wave energy, a plurality of mutually insulated electromagnetic wave, metallic absorber elements mounted on said face, each of said elements including a tapered portion extending away from the face in the direction of propagation of the wave energy, each of said elements having a decreasing cross-sectional area as the distance from the face increases, adjacent ones of said elements being spaced from each other by a distance on the order of a wavelength or less of said energy, an integrated circuit diode in said substrate connected to be responsive to wave energy intercepted by alternately spaced ones of said elements, a first metallic film on said substrate ohmically connected to one electrode of each of said diodes, a second metallic film on said substrate ohmically connected to the remaining metallic elements, said first and second films including load terminals whereby a capacitor is formed in the substrate dielectric between adjacent ones of said metallic films, and said load terminals.

* * * * *

US007400253B2

(12) **United States Patent**
Cohen

(10) **Patent No.: US 7,400,253 B2**
(45) **Date of Patent: Jul. 15, 2008**

(54) HARVESTING AMBIENT RADIO FREQUENCY ELECTROMAGNETIC ENERGY FOR POWERING WIRELESS ELECTRONIC DEVICES, SENSORS AND SENSOR NETWORKS AND APPLICATIONS THEREOF

(75) Inventor: **Marc H. Cohen**, Silver Spring, MD (US)

(73) Assignee: **MHCMOS, LLC**, Silver Spring, MD (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 239 days.

(21) Appl. No.: **11/459,459**

(22) Filed: **Jul. 24, 2006**

(65) **Prior Publication Data**

US 2007/0109121 A1 May 17, 2007

Related U.S. Application Data

(60) Provisional application No. 60/705,636, filed on Aug. 4, 2005.

(51) **Int. Cl.**
G08B 13/14 (2006.01)

(52) **U.S. Cl.** **340/572.1**; 340/572.3; 340/572.4; 340/539.1; 340/825.69; 340/825.72; 340/10.1; 235/383; 235/385; 320/108; 320/109

(58) **Field of Classification Search** 340/572.1, 340/572.3, 572.4, 539.1, 635, 657, 825.69, 340/825.72, 10.1; 235/383, 385; 320/108, 320/109; 455/422.1
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,127,799	A	10/2000	Krishnan	
6,798,716	B1	9/2004	Charych	
6,972,682	B2 *	12/2005	Lareau et al.	340/568.1
7,084,605	B2	8/2006	Mickle et al.	
7,348,875	B2 *	3/2008	Hughes et al.	340/10.4
2003/0199778	A1	10/2003	Mickle et al.	
2004/0113790	A1 *	6/2004	Hamel et al.	340/572.1
2004/0259604	A1	12/2004	Mickle et al.	

* cited by examiner

Primary Examiner—Hung T. Nguyen
(74) *Attorney, Agent, or Firm*—Mohammad S. Rahman, Esq.; Gibb & Rahman, LLC

(57) ABSTRACT

A system and device for harvesting various frequencies and polarizations of ambient radio frequency (RF) electromagnetic (EM) energy for making a passive sensor (tag) into an autonomous passive sensor (tag) adapted to collect and store data with time-stamping and some primitive computation when necessary even when an interrogating radio frequency identification (RFID) reader is not present (not transmitting). A specific source of ambient RF EM energy may include wireless fidelity (WiFi) and/or cellular telephone base stations. The system and device may also allow for the recharging of energy storage units in active and battery assisted passive (BAP) devices. The system could be a "smart building" that uses passive sensors with RF EM energy harvesting capability to sense environmental variables, security breaches, as well as information from "smart appliances" that can be used for a variety of controls and can be accessed locally or remotely over the Internet or cellular networks.

35 Claims, 15 Drawing Sheets

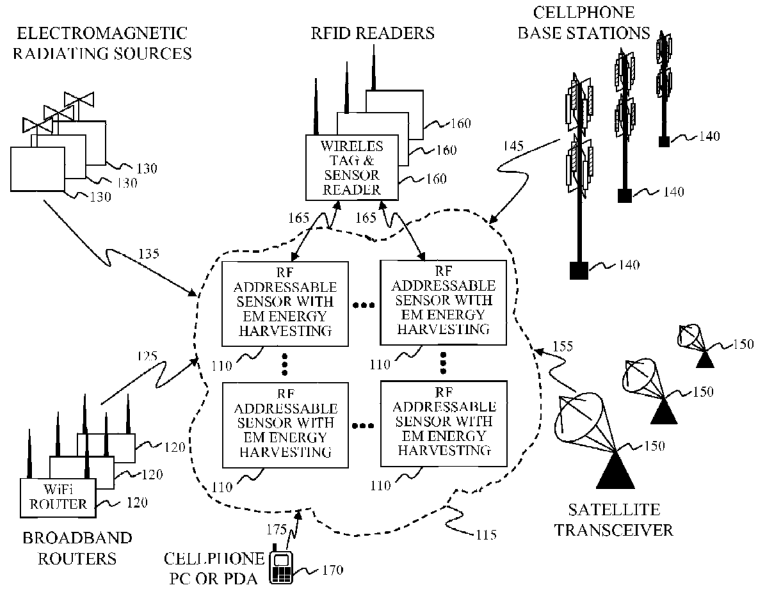
ELECTROMAGNETIC RADIATING SOURCES
RFID READERS
CELLPHONE BASE STATIONS
130
130
130
135
160
160
160
WIRELESS TAG & SENSOR READER
145
140
140
140
140
165
165
RF ADDRESSABLE SENSOR WITH EM ENERGY HARVESTING
RF ADDRESSABLE SENSOR WITH EM ENERGY HARVESTING
RF ADDRESSABLE SENSOR WITH EM ENERGY HARVESTING
RF ADDRESSABLE SENSOR WITH EM ENERGY HARVESTING
110
110
110
110
125
155
150
150
150
120
120
120
WiFi ROUTER
BROADBAND ROUTERS
175
115
CELLPHONE PC OR PDA
170
SATELLITE TRANSCEIVER

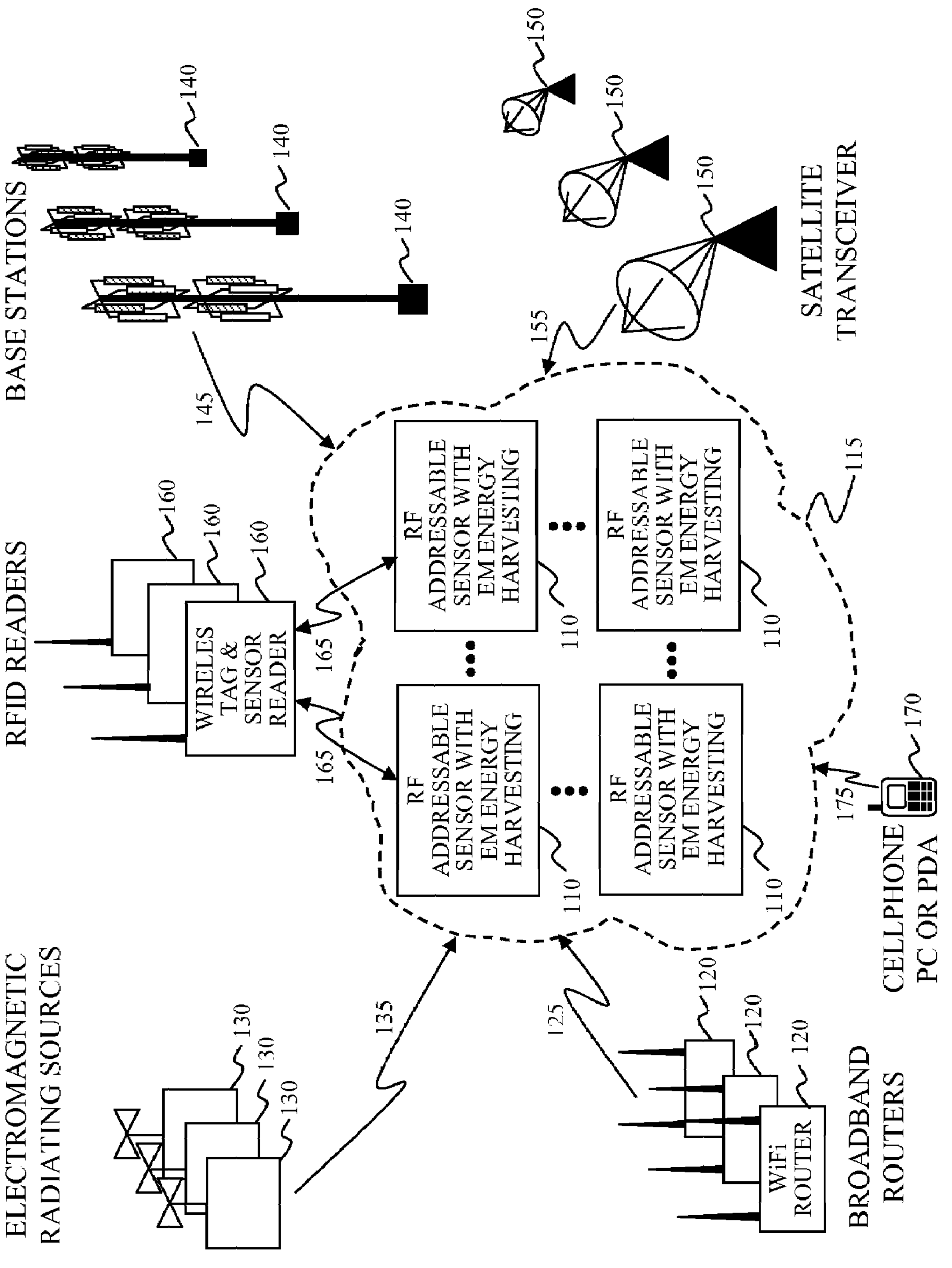
ELECTROMAGNETIC RADIATING SOURCES
RFID READERS
WIRELES TAG & SENSOR READER
CELLPHONE BASE STATIONS
RF ADDRESSABLE SENSOR WITH EM ENERGY HARVESTING
SATELLITE TRANSCEIVER
WiFi ROUTER
BROADBAND ROUTERS
CELLPHONE PC OR PDA

FIG. 1

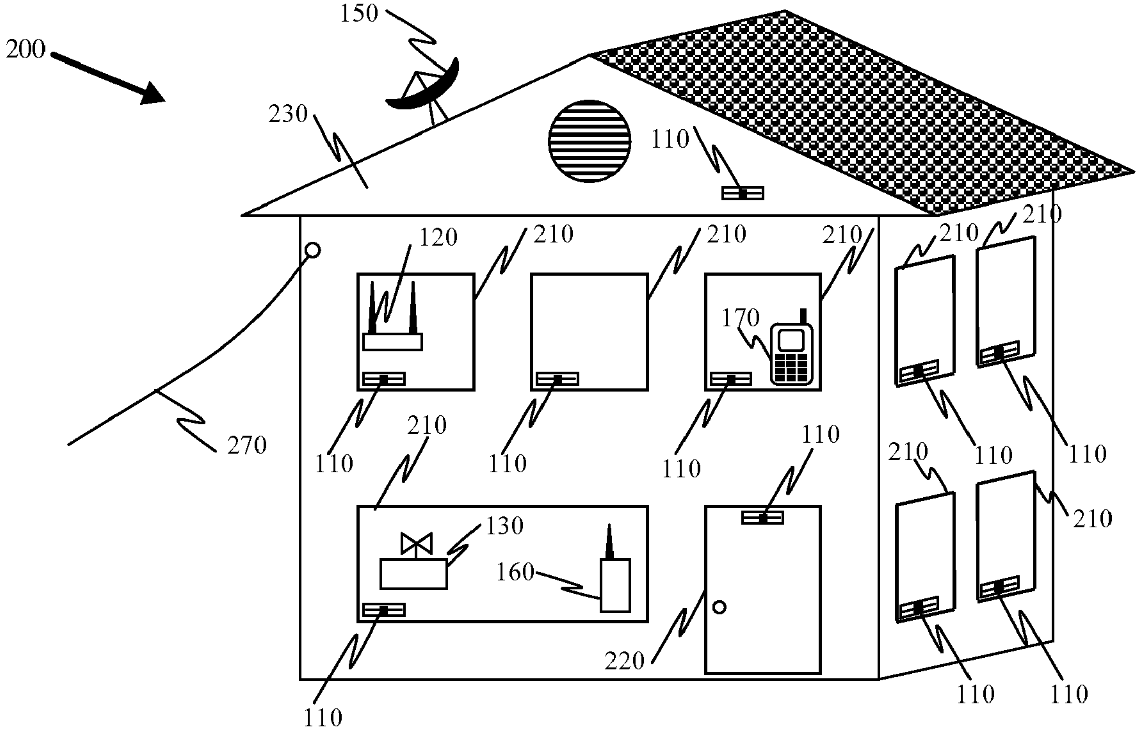

FIG. 2

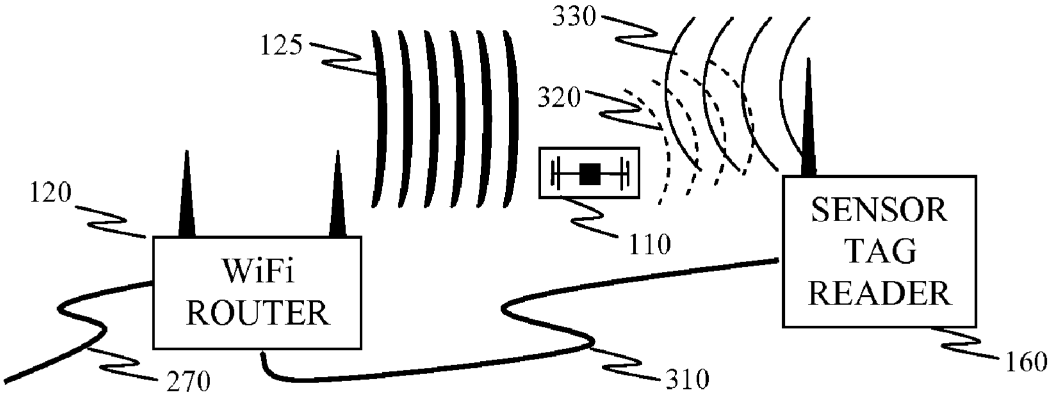
330
125
320
120
110
SENSOR
TAG
READER
WiFi
ROUTER
160
270
310

FIG. 3A

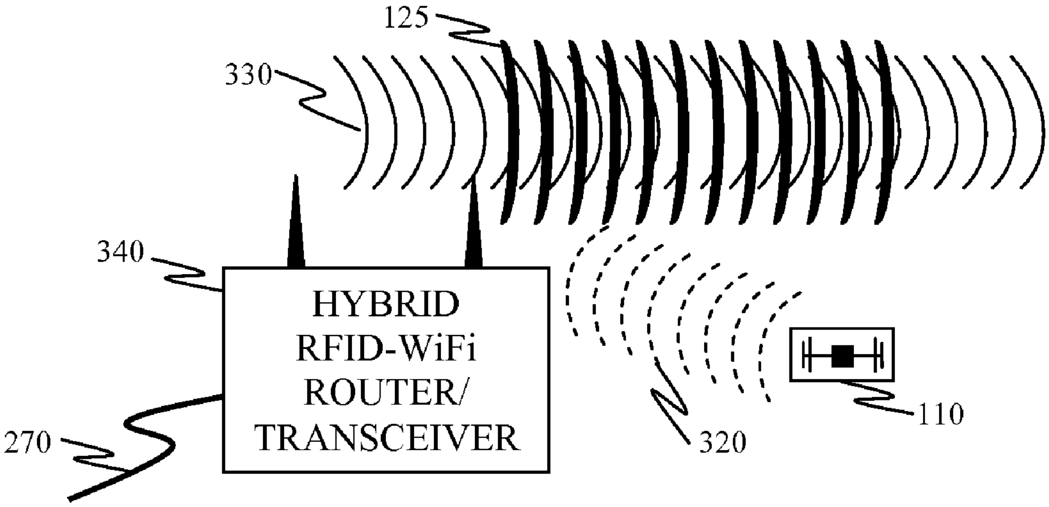
125
330
340
HYBRID RFID-WiFi ROUTER/ TRANSCEIVER
270
320
110

FIG. 3B

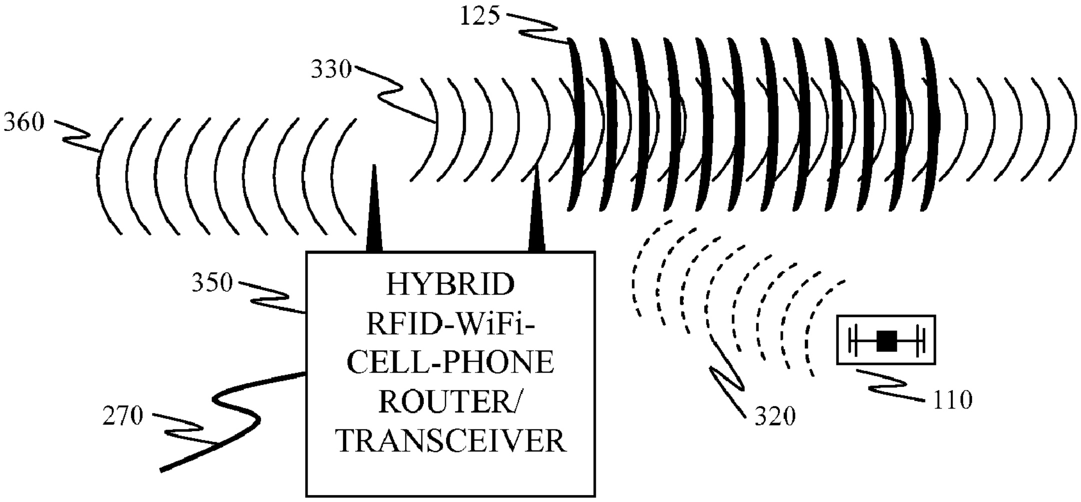
125
330
360
350
HYBRID
RFID-WiFi-
CELL-PHONE
ROUTER/
TRANSCEIVER
270
320
110

FIG. 3C

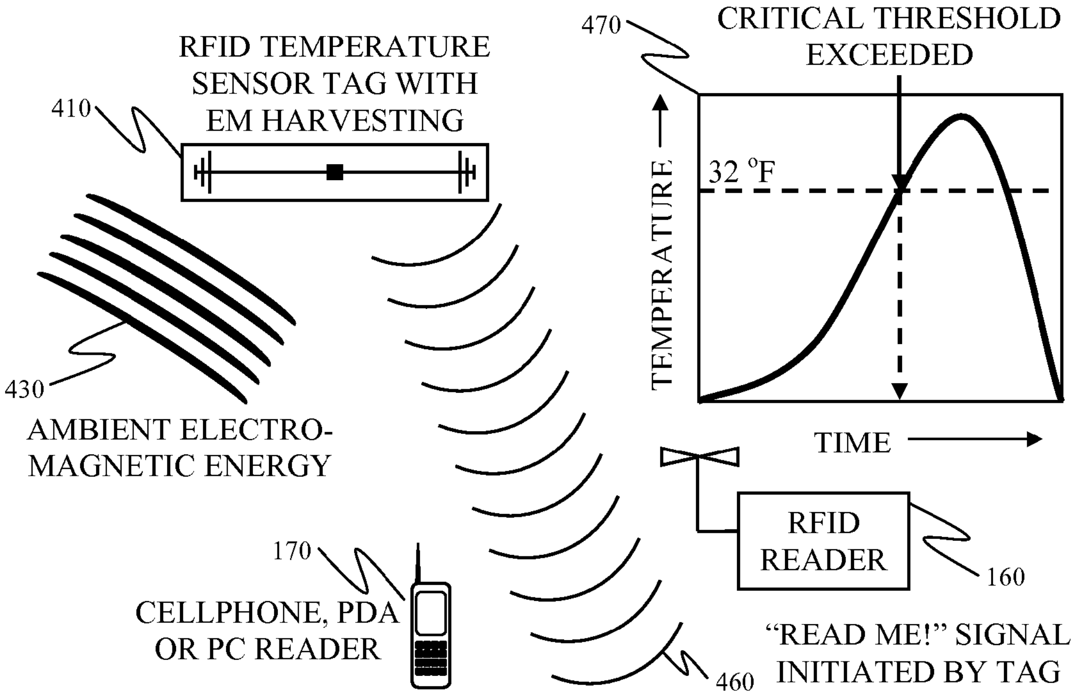
RFID TEMPERATURE SENSOR TAG WITH EM HARVESTING
410
470
CRITICAL THRESHOLD EXCEEDED
TEMPERATURE
32 °F
TIME
430
AMBIENT ELECTRO-MAGNETIC ENERGY
RFID READER
160
170
CELLPHONE, PDA OR PC READER
"READ ME!" SIGNAL INITIATED BY TAG
460

FIG. 4A

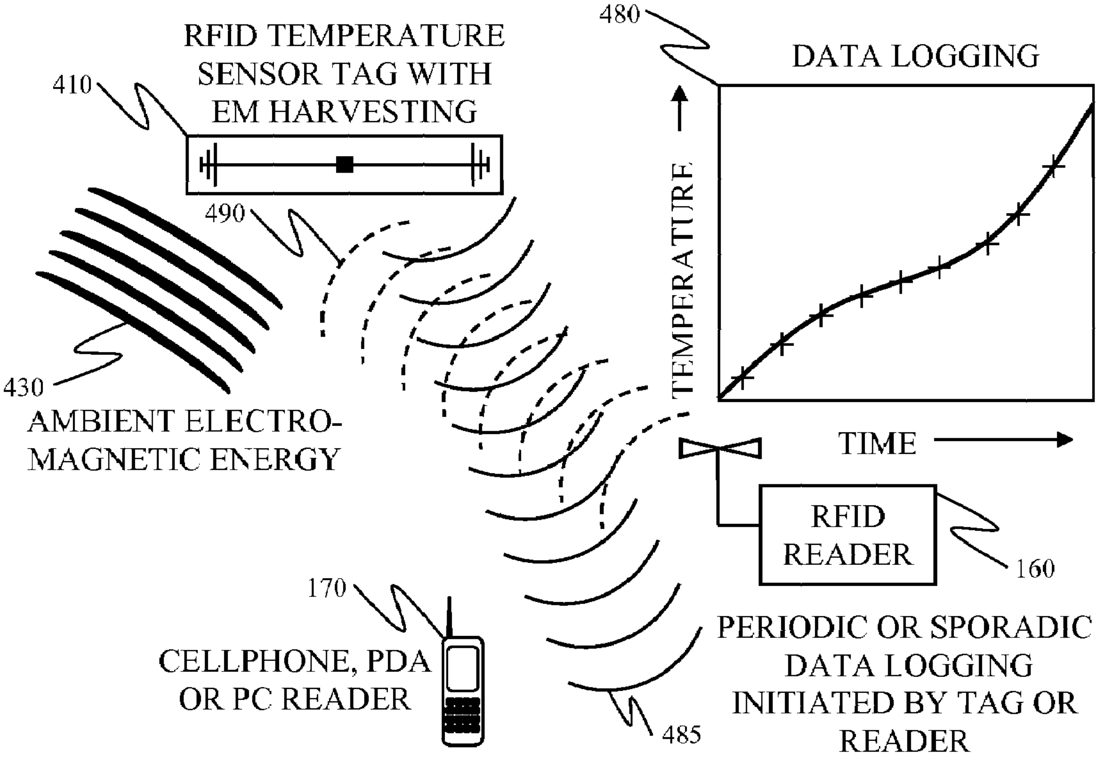
RFID TEMPERATURE SENSOR TAG WITH EM HARVESTING
410
490
430
AMBIENT ELECTRO-MAGNETIC ENERGY
480
DATA LOGGING
TEMPERATURE
TIME
RFID READER
160
170
CELLPHONE, PDA OR PC READER
485
PERIODIC OR SPORADIC DATA LOGGING INITIATED BY TAG OR READER

FIG. 4B

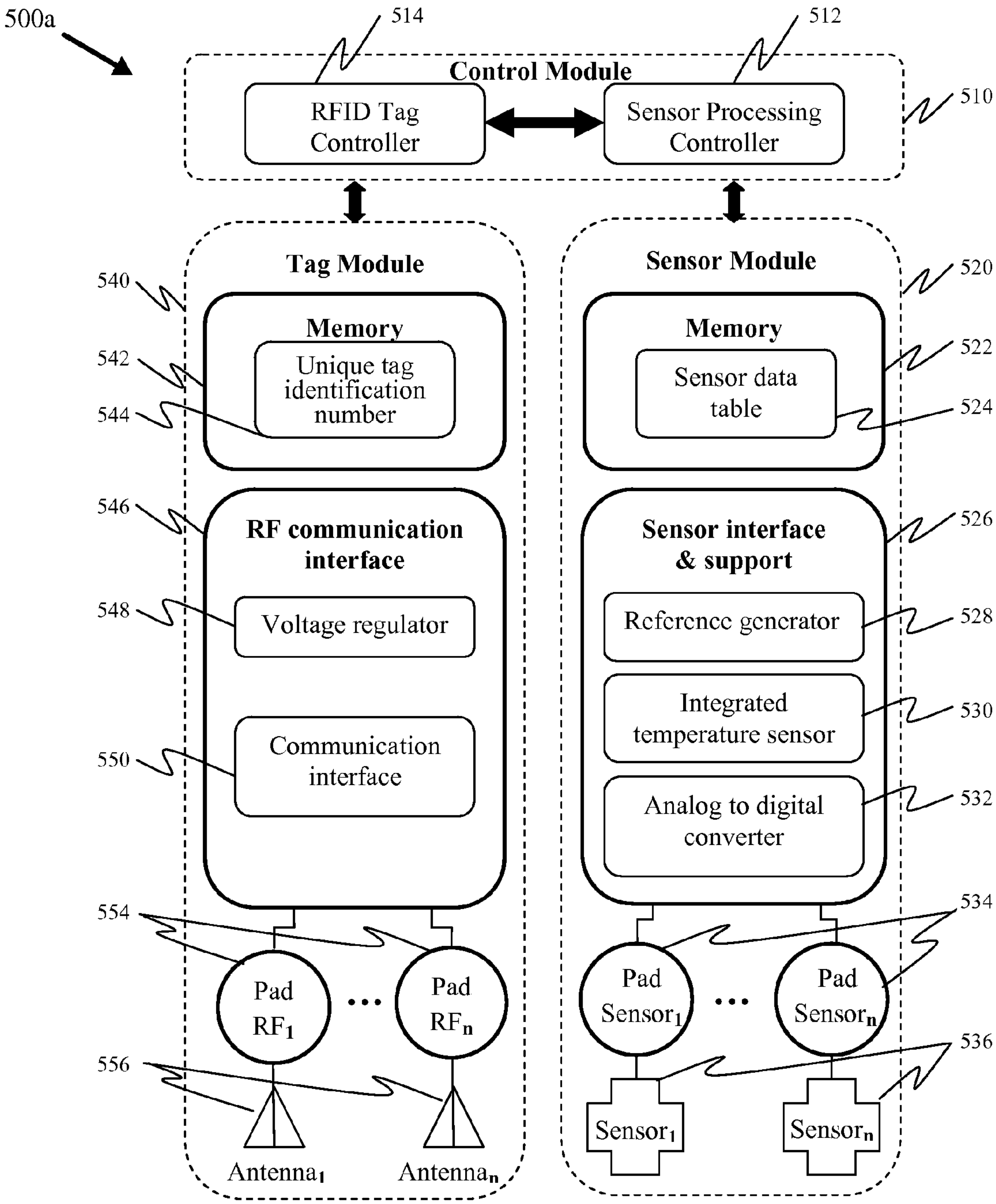
500a
514
512
Control Module
RFID Tag Controller
Sensor Processing Controller
510
Tag Module
540
Memory
542
Unique tag identification number
544
RF communication interface
546
Voltage regulator
548
Communication interface
550
554
Pad RF$_1$
Pad RF$_n$
556
Antenna$_1$
Antenna$_n$
Sensor Module
520
Memory
522
Sensor data table
524
Sensor interface & support
526
Reference generator
528
Integrated temperature sensor
530
Analog to digital converter
532
534
Pad Sensor$_1$
Pad Sensor$_n$
536
Sensor$_1$
Sensor$_n$

FIG. 5A

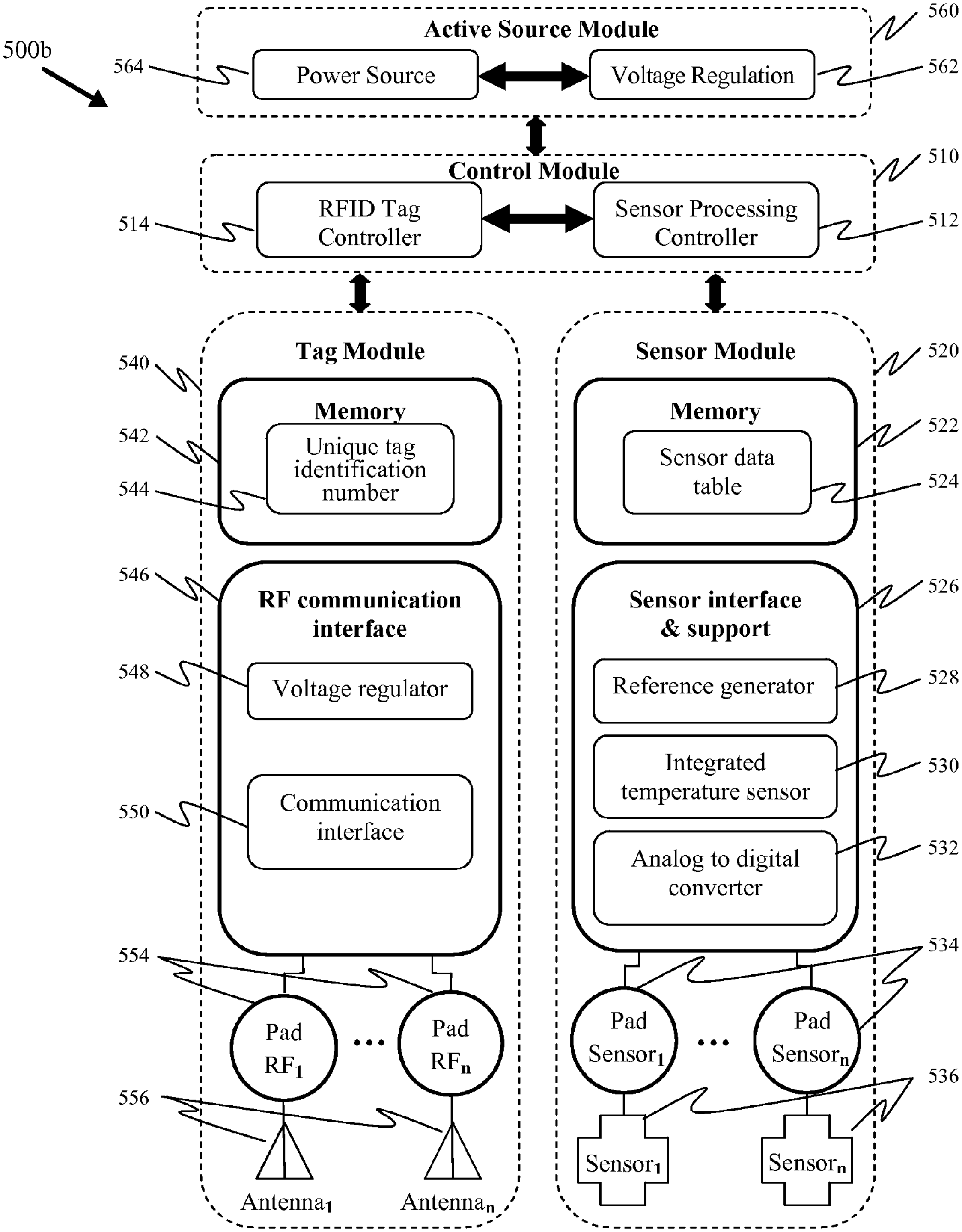
500b
Active Source Module
560
564
Power Source
Voltage Regulation
562
Control Module
510
514
RFID Tag Controller
Sensor Processing Controller
512
Tag Module
540
Sensor Module
520
Memory
542
Unique tag identification number
544
Memory
522
Sensor data table
524
546
RF communication interface
548
Voltage regulator
550
Communication interface
Sensor interface & support
526
Reference generator
528
Integrated temperature sensor
530
Analog to digital converter
532
554
Pad RF$_1$
Pad RF$_n$
534
Pad Sensor$_1$
Pad Sensor$_n$
556
536
Sensor$_1$
Sensor$_n$
Antenna$_1$
Antenna$_n$

FIG. 5B

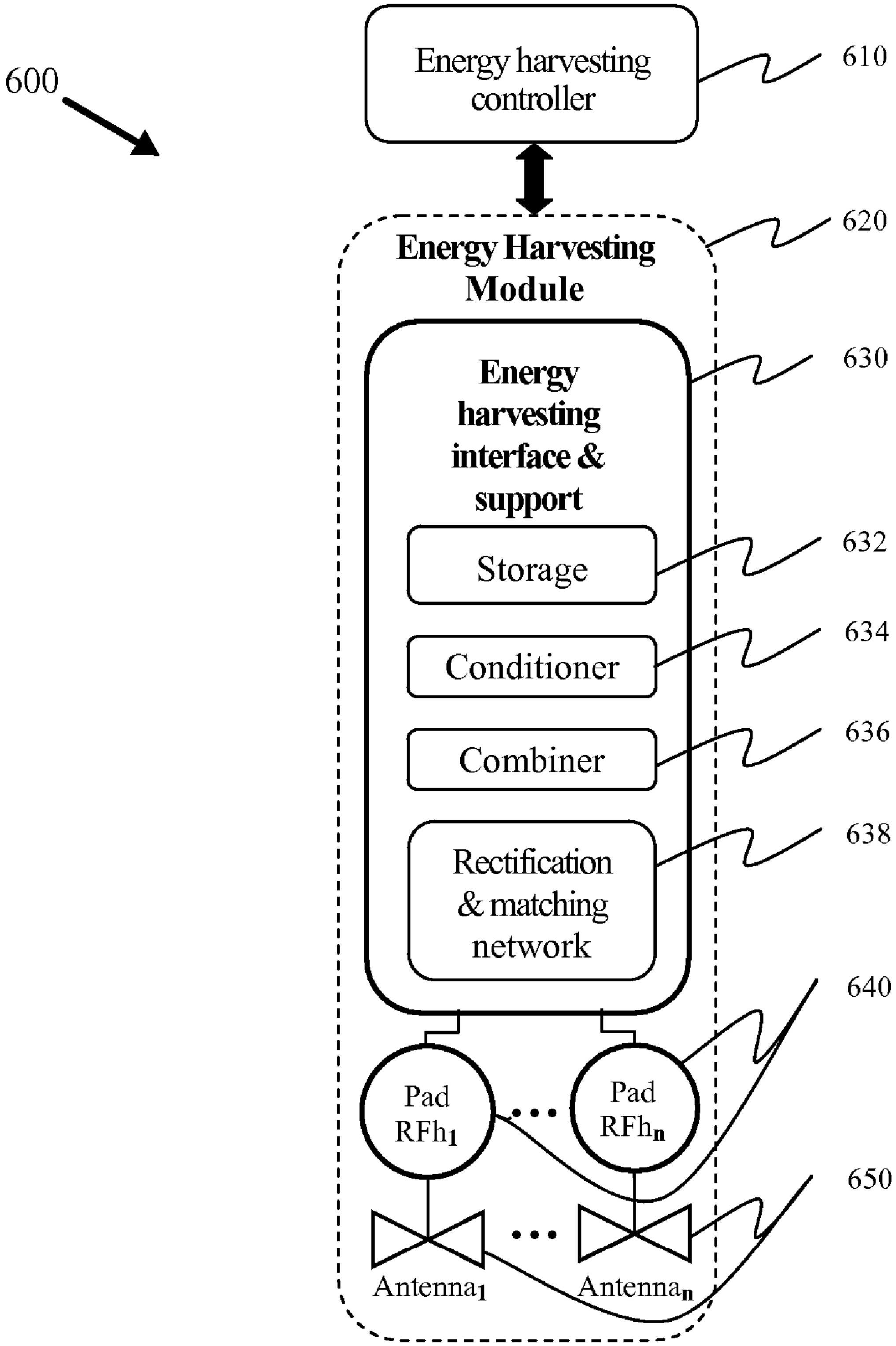
600
Energy harvesting controller
610
Energy Harvesting Module
620
Energy harvesting interface & support
630
Storage
632
Conditioner
634
Combiner
636
Rectification & matching network
638
Pad RFh$_1$
Pad RFh$_n$
640
Antenna$_1$
Antenna$_n$
650

FIG. 6

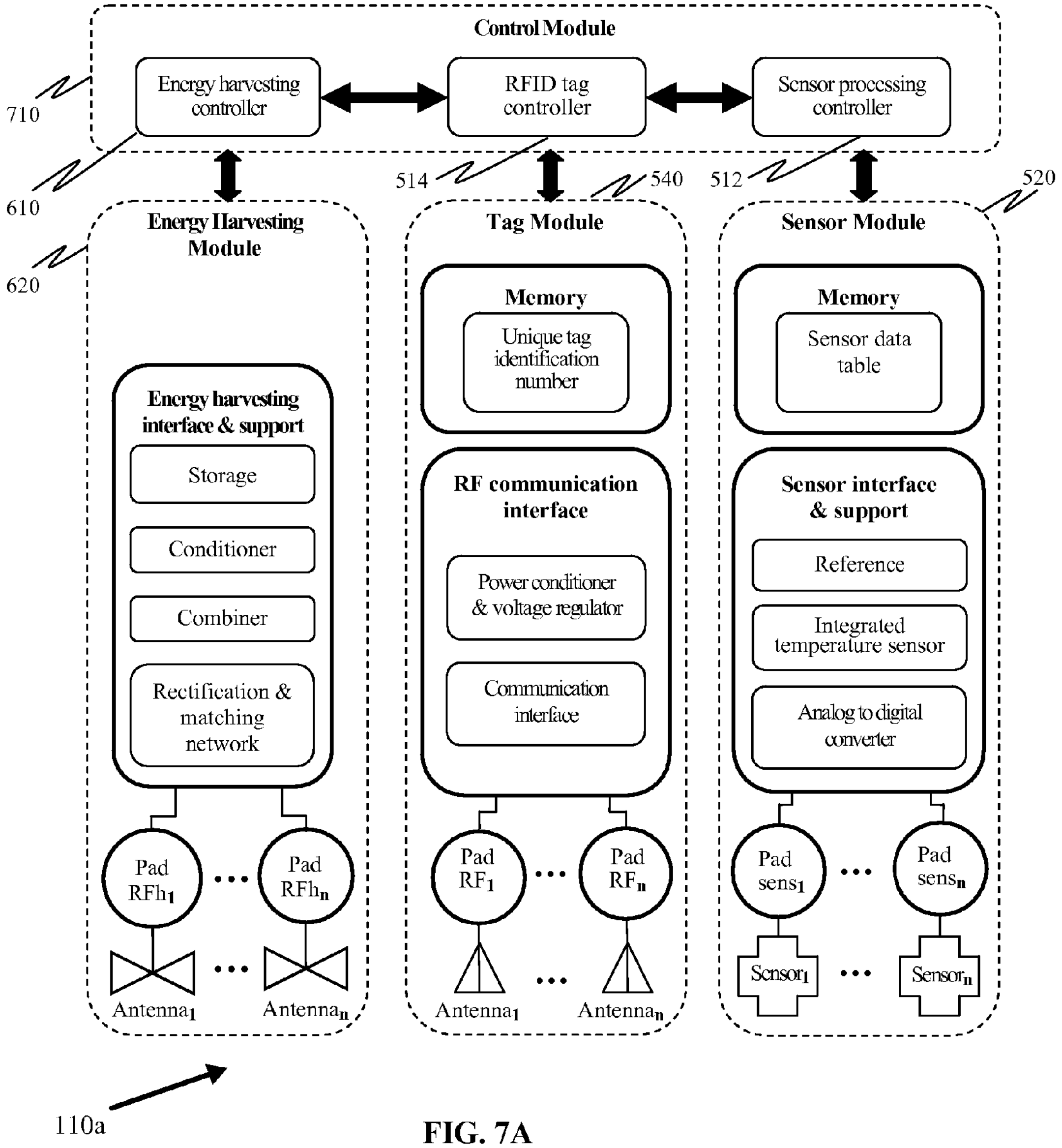
Control Module
Energy harvesting controller
RFID tag controller
Sensor processing controller
710
610
514
540
512
520
620
Energy Harvesting Module
Tag Module
Sensor Module
Energy harvesting interface & support
Storage
Conditioner
Combiner
Rectification & matching network
Memory
Unique tag identification number
RF communication interface
Power conditioner & voltage regulator
Communication interface
Memory
Sensor data table
Sensor interface & support
Reference
Integrated temperature sensor
Analog to digital converter
Pad RFh$_1$
Pad RFh$_n$
Pad RF$_1$
Pad RF$_n$
Pad sens$_1$
Pad sens$_n$
Antenna$_1$
Antenna$_n$
Antenna$_1$
Antenna$_n$
Sensor$_1$
Sensor$_n$
110a

FIG. 7A

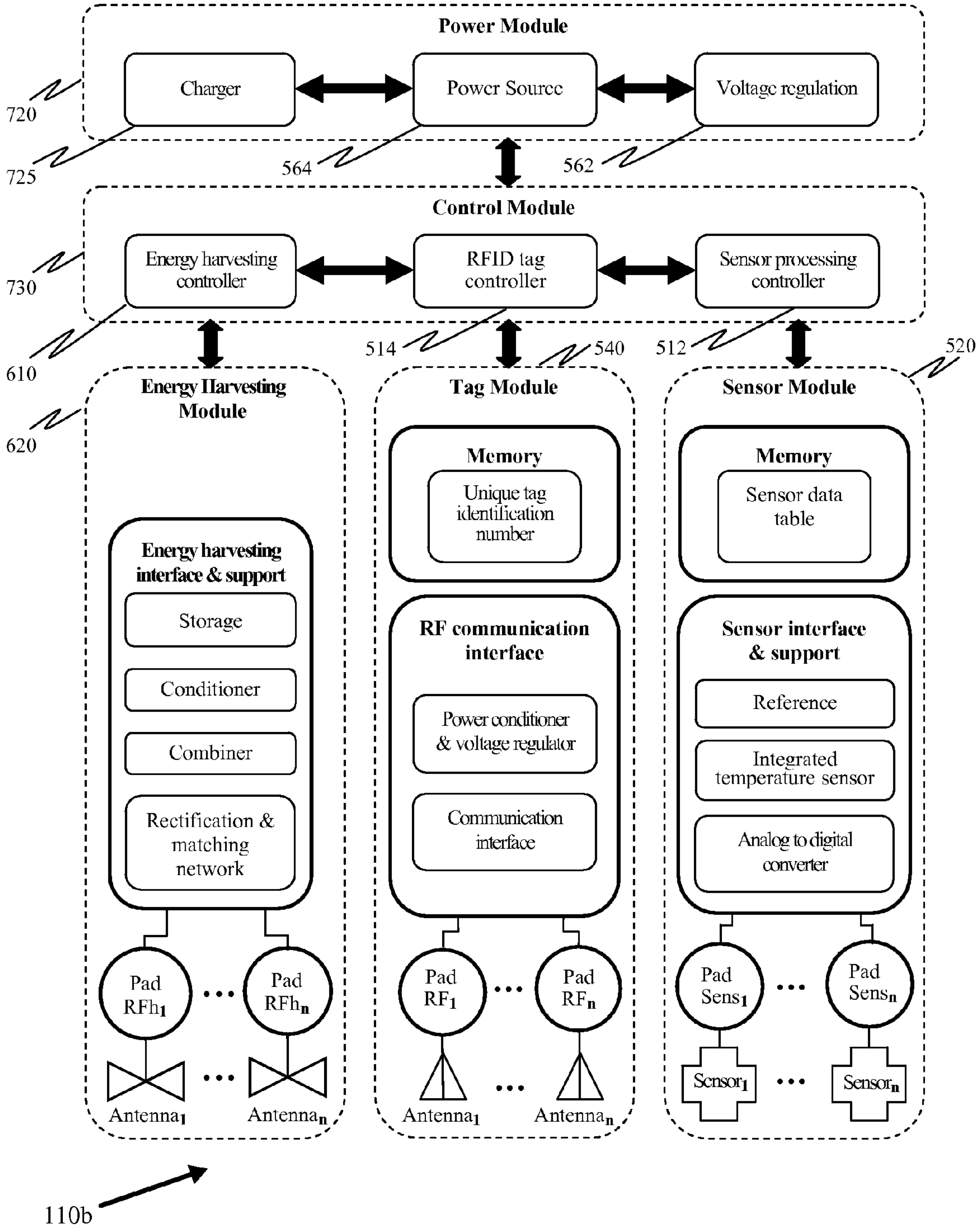
Power Module
Charger
Power Source
Voltage regulation
720
725
564
562
Control Module
Energy harvesting controller
RFID tag controller
Sensor processing controller
730
610
514
540
512
520
620
Energy Harvesting Module
Energy harvesting interface & support
Storage
Conditioner
Combiner
Rectification & matching network
Pad RFh1
Pad RFhn
Antenna1
Antennan
Tag Module
Memory
Unique tag identification number
RF communication interface
Power conditioner & voltage regulator
Communication interface
Pad RF1
Pad RFn
Antenna1
Antennan
Sensor Module
Memory
Sensor data table
Sensor interface & support
Reference
Integrated temperature sensor
Analog to digital converter
Pad Sens1
Pad Sensn
Sensor1
Sensorn
110b

FIG. 7B

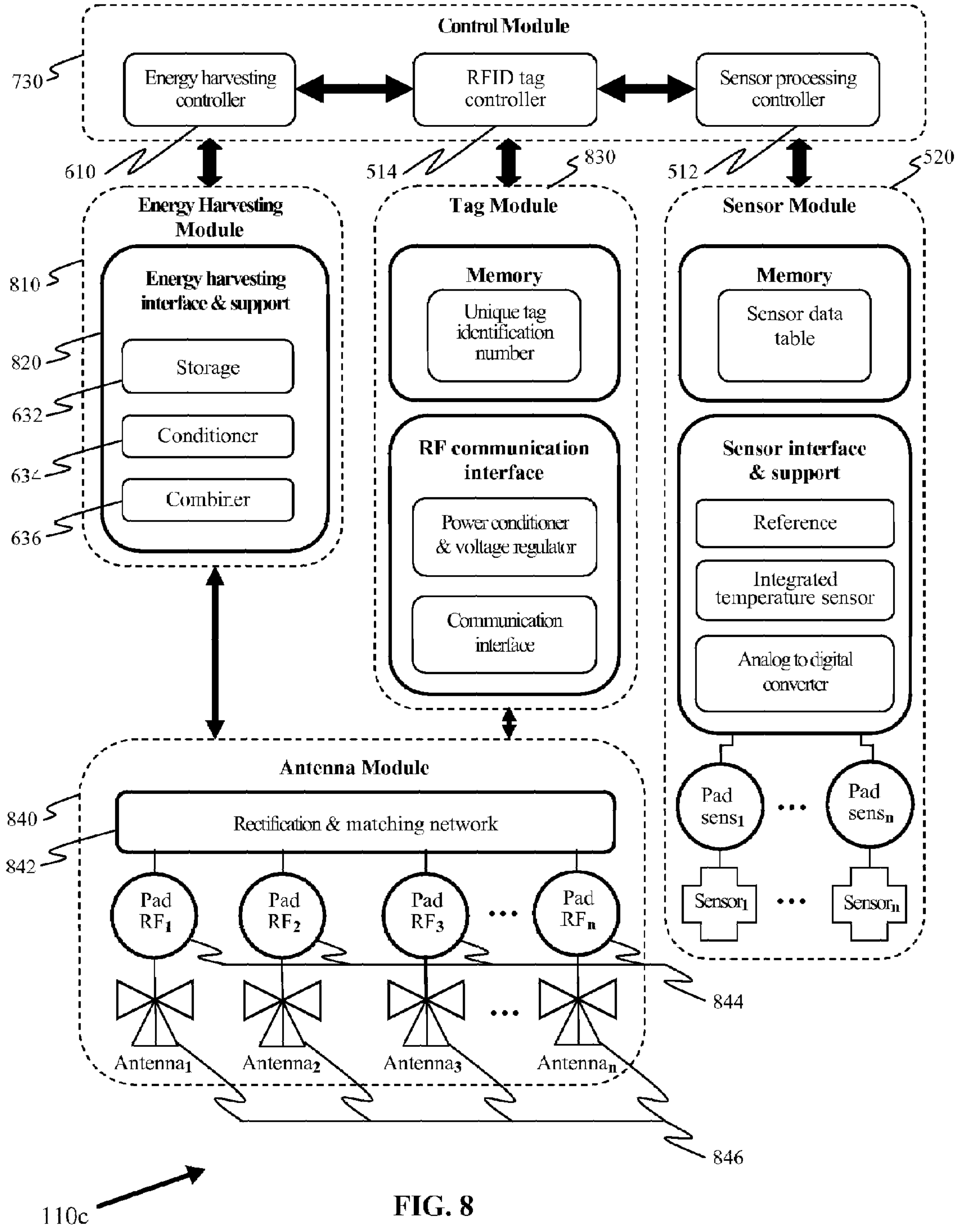
Control Module
730
Energy harvesting controller
RFID tag controller
Sensor processing controller
610
514
830
512
520
Energy Harvesting Module
Tag Module
Sensor Module
810
Energy harvesting interface & support
820
Storage
632
Conditioner
634
Combiner
636
Memory
Unique tag identification number
RF communication interface
Power conditioner & voltage regulator
Communication interface
Memory
Sensor data table
Sensor interface & support
Reference
Integrated temperature sensor
Analog to digital converter
Pad sens$_1$
Pad sens$_n$
Sensor$_1$
Sensor$_n$
Antenna Module
840
Rectification & matching network
842
Pad RF$_1$
Pad RF$_2$
Pad RF$_3$
Pad RF$_n$
844
Antenna$_1$
Antenna$_2$
Antenna$_3$
Antenna$_n$
846
110c

FIG. 8

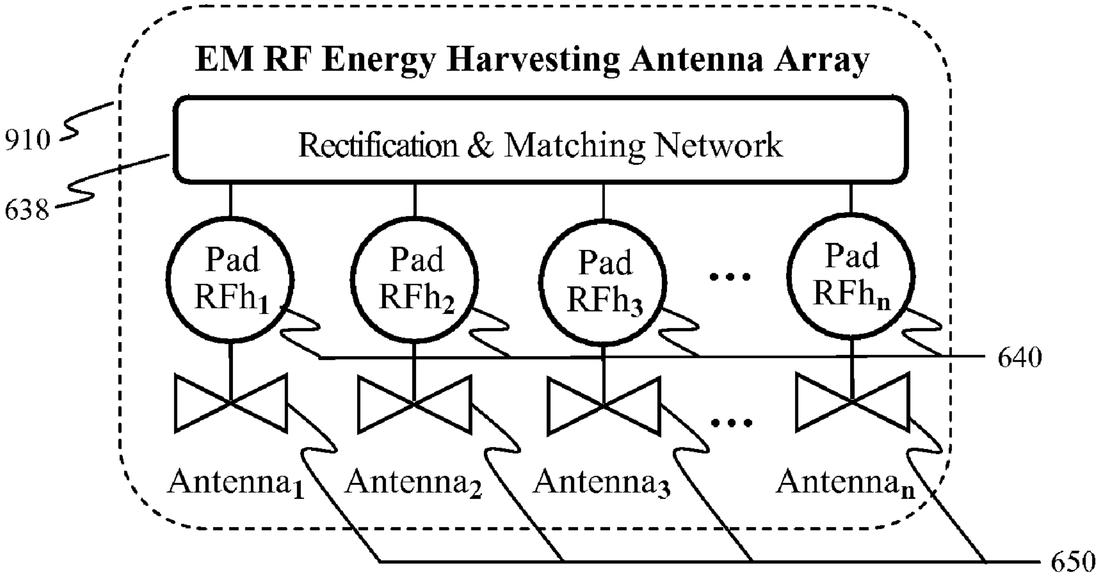
EM RF Energy Harvesting Antenna Array
910
638
Rectification & Matching Network
Pad RFh$_1$
Pad RFh$_2$
Pad RFh$_3$
...
Pad RFh$_n$
640
Antenna$_1$
Antenna$_2$
Antenna$_3$
...
Antenna$_n$
650

FIG. 9A

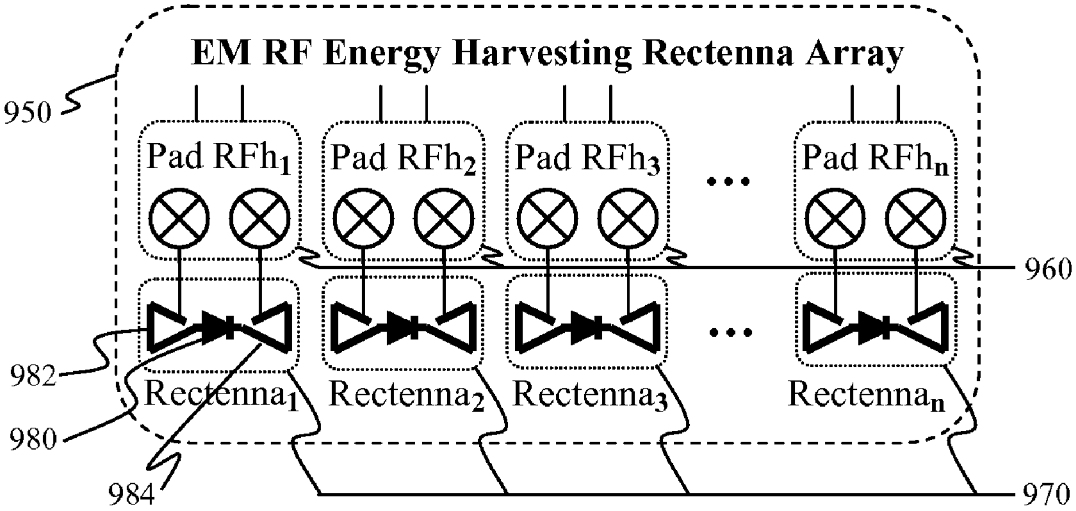
EM RF Energy Harvesting Rectenna Array
950
Pad RFh1
Pad RFh2
Pad RFh3
Pad RFhn
960
982
980
984
Rectenna1
Rectenna2
Rectenna3
Rectennan
970

FIG. 9B

HARVESTING AMBIENT RADIO FREQUENCY ELECTROMAGNETIC ENERGY FOR POWERING WIRELESS ELECTRONIC DEVICES, SENSORS AND SENSOR NETWORKS AND APPLICATIONS THEREOF

CROSS REFERENCE TO RELATED APPLICATION

This application claims the benefit of U.S. Provisional Patent Application No. 60/705,636 filed on Aug. 4, 2005, the contents of which in its entirety is herein incorporated by reference.

BACKGROUND

1. Technical Field

The embodiments herein generally relate to wireless technology, and, more particularly, to harvesting ambient radio frequency (RF) electromagnetic (EM) energy for supplying, storing, recharging or supplementing the energy necessary to power active or passive wireless electronic devices, sensors and sensor networks.

2. Description of the Related Art

The ability of actively powered, battery assisted passive (BAP) and passive wireless sensors to remotely acquire, store and/or transmit data may open up a wide variety of medical, diagnostic, quality control, safety, supply chain, logistic and security applications. For an individual, the ability to determine, prior to consumption, whether a food item contains harmful bacteria or ingredients to which the individual is allergic is highly desirable. From a population's perspective, recent national and international events have increased the need for personal or distributed systems that can monitor and detect continuously in real-time or at periodic or irregular intervals in time, chemical agents, biological agents, radiological agents and other hazards locally or over wide geographical areas.

Examples of personal handheld wireless devices that can address these issues are cellular telephones (cell phones), personal data assistants (PDAs) and portable computers or notebooks (PCs). These are pervasive electronic devices that are networked via the Internet or cellular telephone networks. Generally, the abovementioned consumer electronic devices all require local energy storage (batteries) to provide mobility and portability, and do so until their local source of energy is depleted.

A new class of sensors such as Electronic Product Code (EPC) and Radio Frequency Identification (RFID) tags that include sensing capabilities are emerging as a generally inexpensive and effective means of addressing many wireless sensor applications such as, but not exclusively, cold-supply chain, food safety, quality control, environmental safety, medical, diagnostic, electro-immunoassays, consumer goods as well as homeland security, property and personal security applications.

Purely passive sensors, when actively interrogated by an RF transceiver (reader), use this received source of EM energy to power themselves up, to acquire readings from their sensors and then rebroadcast or reflect their specific identification code and sensor readings back to the interrogator. BAP tags use their batteries to acquire and store or log data from one or more sensors, but typically do not use their batteries to enhance their RF communication abilities. Active sensors have their own built in power source that can be used to acquire and store sensor readings at any time as well as enhance the reception and transmission of RF communications. Generally, passive EPC and RFID tags equipped with one or more sensors will require a source of energy to measure and store their acquired information at times other than during active interrogation by a reader.

Because of the cost of the conventional sensors and sensor readers in particular, broad deployment of a sensor network over a large geographical area or widespread use by individuals is currently not particularly feasible. Generally, a considerable problem with large geographical deployment is that if the sensors are actively powered, their power sources need to be replaced when depleted, again adding to the cost. In addition, wear on sensors or sensor surfaces, generally requires sensors to be replaced on a regular basis further increasing the cost. In addition, imprecision in sensors generally requires cross validation to eliminate false positives, increasing the number of sensors that must be deployed for each application and thus also increasing the cost.

Accordingly, in view of the foregoing, there remains a need for low cost wireless sensors that can harvest ambient EM energy to power themselves up, acquire sensor data or recharge their power sources (batteries in the case of active sensors and batteries in BAP sensors).

SUMMARY

In view of the foregoing, the embodiments herein provide wireless RF addressable sensor network architectures for individuals, homes, industries, and homeland security where a wireless reader can supply RF power to and communicate with RF addressable tags that include sensors. Each active, BAP or passive tag can be equipped to harvest ambient EM energy from its environment allowing it to be autonomous: the tag can acquire and store data from its sensors independently of any particular reader event. The harvested ambient EM energy could, in addition, be used to transmit an alarm or "read me!" signal when a vital "sensor event" occurs, or to charge a battery on or near the tag. For example, the reader could be an RFID reader, a broadband wireless fidelity (WiFi) device, a cell phone, PDA, PC or a hybrid thereof comprising of a device adapted to read an RFID tag's unique identification and sensor information and bi-directionally communicate this information over the Internet or cellular telephone networks. The embodiments herein are also directed towards RF addressable tags with sensors that may be produced at costs lower than most types of active or passive tags conventionally available and furthers the state of the art by including ambient EM energy harvesting capabilities.

The RF addressable sensor tag combines RFID tag functionality with sensor functionality. The RF addressable sensor includes one or more antennas for communicating with the tag reader, one or more sensor elements, an RF power and communications interface, an RFID control module, and a sensor interface. The RFID control module includes logic to control RFID tag communications with an RFID tag reader. The RFID control module may also include logic and memory necessary to store and/or process the acquired sensor data.

The embodiments herein additionally furthers the current state of the art by adding an ambient EM energy harvesting sub-system to passive, BAP, or active wireless sensor tags. The harvested energy is used to provide power to the passive sensor or to provide power to the BAP or active sensor or for recharging an energy source on the active sensor tag or BAP sensor tag. This harvested energy is used by the sensor tag for: the facilitation of autonomous data collection and/or storage; the processing of acquired data; and the transmission of an alarm or "read-me!" signal when a vital sensor variable is above or below a critical threshold or vital sensor variables are above or below critical thresholds.

The embodiments herein are also adapted to identify WiFi and cellular base stations as sources of readily available, known frequency, ambient RF EM energy that can be harvested to provide energy to the sensor tag. The embodiments herein are also capable of identifying microwave band frequencies consistent with IEEE 521-1984 designations (Bands L, S, C, X, Ku, K, Ka, V, and W) as ambient RF EM energy that can be harvested to provide energy to the sensor tag.

In an instantiation of the embodiments herein, an RFID reader is used to communicate with nearby sensor tags that are being energized by harvested ambient RF EM energy supplied by a nearby WiFi router/transceiver or cellular base station. The WiFi router/transceiver or cellular base station not only supplies the ambient energy for the sensor tags, but also bi-directionally communicates and distributes the collected sensor tag data over the Internet or cellular telephone networks.

In an aspect of the embodiments herein, the RFID reader and WiFi router/transceiver are combined into a hybrid RFID-WiFi reader whereby the RFID reader sub-system communicates with nearby sensor tags that are being energized by harvested ambient RF EM energy supplied by the WiFi router/transceiver part of the hybrid RFID-WiFi reader. The WiFi router/transceiver part of the hybrid RFID-WiFi reader also facilitates the local and/or remote distribution of the collected sensor tag data over the Internet.

In another aspect of the embodiments herein, the RFID, WiFi router/transceiver, and cell phone technology are combined into a hybrid RFID-WiFi-cell phone reader whereby the RFID and/or WiFi reader sub-system communicates with nearby RFID sensor tags and the WiFi and cell phone subsystems facilitate local and remote access to the collected sensor tag data over the Internet or via the cellular telephone networks. The WiFi router/transceiver part of the hybrid RFID-WiFi-Cell phone reader is also the source of ambient energy available for harvesting by the sensor tags.

In another aspect of the embodiments herein, a source that radiates EM energy at a known frequency may be placed in the vicinity of the passive, BAP or active sensor tags for the sole purpose of supplying ambient RF EM energy at frequencies that these wireless sensors can harvest.

The energy harvesting system may include an antenna or an array of antennas to collect ambient EM energy having a single wavelength or a multitude of wavelengths of single polarization or arbitrary polarizations. In another aspect of the embodiments herein the energy harvesting antenna or antennas may be used to supplement the wireless tag's communication antenna or antennas. In another instantiation of the embodiments herein, the energy harvesting antenna or array of antennas connects to an impedance matching network and a rectifying and combining module before being fed to the power converting and/or conditioning module.

In yet another instantiation of the embodiments herein, the energy harvesting antenna or array of antennas may be designed to directly rectify the harvested energy (a rectenna or array of rectennas) and feed the rectified energy to an energy combining module that in turn feeds a power converting and/or conditioning module.

In another aspect of the embodiments herein, the energy harvesting power converting and/or conditioning module can supplement the wireless tag's RF transceiver power and thereby its communication capabilities. Furthermore, the embodiments herein preferably do not interfere with the communications between RFID readers and wireless RF addressable tags.

Generally, the embodiments provide a system comprising an ambient RF EM energy source; a sensor reader operatively connected to the ambient RF EM energy source; and a sensor in communication with the sensor reader, wherein the sensor comprises a control module comprising an energy harvesting controller; a tag module controlled by the control module; a sensor module controlled by the control module; an energy harvesting module controlled by the energy harvesting controller and adapted to wirelessly collect the ambient RF EM energy; and a power module adapted to provide a power supply for the control module, the power supply being generated by the collected ambient RF EM energy, wherein the energy harvesting controller is adapted to (i) assess power requirements of each of the tag and sensor modules, (ii) transfer power from the energy harvesting module to any of the tag and sensor modules, and (iii) charge a local energy storage device.

In one embodiment, the sensor further comprises an antenna module shared by the energy harvesting module and the tag module, wherein the antenna module comprises a rectification and impedance matching network; at least one RF pad operatively connected to the rectification and impedance matching network; and at least one antenna operatively connected to each of the at least one RF pad. In another embodiment, the tag module comprises a memory component; a RF communication interface operatively connected to the memory component; at least one RF pad operatively connected to the RF communication interface; and at least one antenna operatively connected to each of the at least one RF pad. The sensor module may comprise a memory component; a sensor interface and support sub-system operatively connected to the memory component; at least one sensor pad operatively connected to the sensor interface and support sub-system; and at least one antenna operatively connected to each of the at least one sensor pad. Alternatively, the sensor module comprises a memory component; a sensor interface and support sub-system operatively connected to the memory component; and at least one antenna operatively connected to sensor interface and support sub-system. Preferably, the control module further comprises a RFID tag controller; and a sensor processing controller operatively connected to the RFID tag controller.

In one embodiment, the energy harvesting module comprises an energy harvesting interface and support sub-system; at least one RF harvesting pad operatively connected to an energy harvesting interface and support sub-system; and at least one antenna operatively connected to each of the at least one RF harvesting pad. Alternatively, the energy harvesting module comprises an energy harvesting interface and support sub-system; at least one pair of RF harvesting pads operatively connected to the energy harvesting interface and support sub-system; and at least one rectenna operatively connected to each of the at least one pair RF harvesting pads. The energy harvesting interface and support sub-system may comprise a storage component; a power conditioner component operatively connected to the storage component; and a combiner operatively connected to the power conditioner, wherein the energy harvesting interface and support subsystem comprises a rectification and impedance matching network operatively connected to the combiner. Preferably, the power module comprises a power charger; a power source operatively connected to the power charger; and a voltage regulator operatively connected to the power source.

Additionally, the collected ambient RF EM energy enables the electronic device to be a fully autonomous data collector, alarm generator, event generator, data logger, and data processor independent of any other device. Moreover, the sensor reader may comprise any of a cell phone, a PDA, a PC reader, a RFID reader, a broadband WiFi device, and a combination thereof. Preferably, the power supply is generated only by the collected ambient RF EM energy. Furthermore, the sensor may be adapted to transmit a wireless signal to the sensor reader, and wherein the sensor reader is adapted to receive the transmitted wireless signal from the sensor. Also, the collected ambient RF EM energy facilitates transmission of the wireless signal when a selected sensor variable meets a predetermined threshold. In one embodiment, the ambient RF EM energy source comprises a WiFi router adapted to allow the sensor reader to communicate over the Internet.

Alternatively, the ambient RF EM energy source comprises a hybrid WiFi router transceiver adapted to allow the sensor reader to communicate over the Internet. In another embodiment, the sensor is a RF ambient EM energy harvesting sensor and is arranged with a plurality of other the sensors each adapted to collect data for climate control, security alarms, environmental alarms, and information gathering from other devices located within a communicable distance from the sensor, wherein all of the sensors are adapted to be any of locally and remotely monitored via any of the Internet and a cell phone network.

These and other aspects of the embodiments herein will be better appreciated and understood when considered in conjunction with the following description and the accompanying drawings. It should be understood, however, that the following descriptions, while indicating preferred embodiments and numerous specific details thereof, are given by way of illustration and not of limitation. Many changes and modifications may be made within the scope of the embodiments herein without departing from the spirit thereof, and the embodiments herein include all such modifications.

BRIEF DESCRIPTION OF THE DRAWINGS

The embodiments herein will be better understood from the following detailed description with reference to the drawings, in which:

FIG. **1** is a schematic diagram illustrating an example of ambient RF EM energy harvesting for gathering, accumulating, and storing energy and powering passive wireless sensors or sensor tags and/or recharging BAP and/or active wireless sensors or sensor tags all of which are equipped with ambient EM energy harvesting technology according to an embodiment herein;.

FIG. **2** is a schematic diagram illustrating the deployment of wireless addressable sensor tags with on-board EM energy harvesting technology in a home, business, office, warehouse, or building according to an embodiment herein;

FIG. **3**A is a schematic diagram illustrating an example of a WiFi wireless router/transceiver, connected to the Internet, bathing passive and/or BAP and/or active wireless sensor tags with energy at a known frequency or known frequencies according to an embodiment herein;

FIG. **3**B is a schematic diagram illustrating an example of a hybrid RFID-WiFi wireless router comprising an RFID tag reader module combined with a WiFi router/transceiver connected to the Internet according to an embodiment herein;

FIG. **3**C is a schematic diagram illustrating an example of a hybrid RFID-WiFi-cell phone wireless router comprising an RFID tag reader module, a WiFi router/transceiver module that facilitates connection to the Internet, and a cell phone module that facilitates connection to cellular telephone networks according to an embodiment herein;

FIG. **4**A is a schematic diagram illustrating an example of an RFID sensor tag equipped with an energy harvesting subsystem that initiates an alarm or "read me!" signal when a measured variable or combination of measured variables exceeds a predefined critical threshold according to an embodiment herein;

FIG. **4**B is a schematic diagram illustrating an example of an RFID sensor tag equipped with the energy harvesting sub-system logging (writing to memory) the sensed data according to an embodiment herein;

FIG. **5**A is a block diagram illustrating a passive wireless RF addressable tag module with a sensor module;

FIG. **5**B is a block diagram illustrating an active wireless RF addressable tag module with a sensor module;

FIG. **6** is a block diagram illustrating an EM energy harvesting sub-system architecture according to an embodiment herein;

FIG. **7**A is a block diagram illustrating a passive wireless RF addressable tag with a sensor module combined with an EM energy harvesting sub-system architecture according to an embodiment herein;

FIG. **7**B is a block diagram illustrating an active wireless RF addressable tag architecture with a sensor module combined with an illustrative EM energy harvesting sub-system architecture according to an embodiment herein;

FIG. **8** is a block diagram illustrating a passive wireless RF addressable tag architecture with a sensor module, an EM energy harvesting sub-system, and an antenna module that could be used for RF communication with a reader as well as to harvest ambient EM energy according to an embodiment herein;

FIG. **9**A is a block diagram illustrating an example of an EM energy harvesting architecture comprising an antenna or an array of antennas with associated rectification and matching networks according to an embodiment herein; and

FIG. **9**B is a block diagram illustrating an example of an EM energy harvesting architecture comprising a rectenna (rectifying antenna) or an array of rectennas with rectification and matching built into each rectenna according to an embodiment herein.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

The embodiments herein and the various features and advantageous details thereof are explained more fully with reference to the non-limiting embodiments that are illustrated in the accompanying drawings and detailed in the following description. Descriptions of well-known components and processing techniques are omitted so as to not unnecessarily obscure the embodiments herein. The examples used herein are intended merely to facilitate an understanding of ways in which the embodiments herein may be practiced and to further enable those of skill in the art to practice the embodiments herein. Accordingly, the examples should not be construed as limiting the scope of the embodiments herein.

As mentioned, there remains a need for low cost wireless sensors that can harvest ambient EM energy to power themselves up, acquire sensor data or recharge their power sources (batteries in the case of active sensors and batteries in BAP sensors). The embodiments herein achieve this by providing a technique for harvesting ambient RF EM energy for supplying, storing, recharging or supplementing the energy necessary to power active, BAP or passive wireless electronic devices, sensors and sensor networks, allowing them to become autonomous "data loggers" and/or empowering them to transmit their unique identification and sensor information when a sensed variable or a combination of sensed variables exceeds a threshold and triggers a vital "sensor event".

More specifically, the embodiments provide a system and device for harvesting various frequencies and polarizations of ambient RF EM energy for making a passive sensor (tag) into an autonomous passive sensor (tag) adapted to collect and store data with time-stamping and some primitive computation when necessary even when an interrogating RFID reader is not present (not transmitting). A specific source of ambient RF EM energy may include WiFi. The system and device may also allow for the recharging of energy storage units in active and BAP devices. The system could be a "smart building" that uses passive sensors with RF EM energy harvesting capability to sense environmental variables, security breaches, as well as information from "smart appliances" that can be used for a variety of controls and can be accessed locally or remotely over the Internet or cellular telephone networks. Referring now to the drawings, and more particularly to FIGS. **1** through **9**B, where similar reference characters denote corresponding features consistently throughout the figures, there are shown preferred embodiments.

The embodiments herein relate to ambient EM energy harvesting for powering passive wireless sensors and for recharging energy storage devices on active or BAP wireless sensors. The RF frequency band from 3 KHz to 300 GHz, and a subset of this band, the Microwave Frequency Band from 1 GHz to 110 GHz (as designated in IEEE 521-1984) are identified as the most ubiquitous sources of ambient energy available for energy harvesting.

The embodiments herein furthers the current state of the art by focusing on RFID sensor tag technologies that may be purely passive (i.e., no on-board power supply but may have on-board energy storage devices) and/or active (i.e., having on-board energy storage as the power supply) and/or BAP (i.e., also having on-board energy storage as the power supply). The distinction between active and BAP wireless sensors is made here: an active tag uses its local power source to power up its sensors and boost its RF receive and transmit communications power while a BAP tag uses its local power source to power up and collect data from its sensors alone.

In particular IEEE standards 802.11a, 802.11b, and 802.11g and WiFi routers/transceivers used in homes, businesses, warehouses, stores and in towns or cities transmit a known wavelength (or frequency) in the S-band (2 GHz to 4 GHz) and the C-band (4 GHz to 8 GHz). This source of energy can be optimally harvested with a combination of appropriately designed antenna or antennas and rectenna or rectennas, impedance matching circuits, voltage regulation circuits and energy storage devices and can be used for supplying energy to wireless sensors, be they passive, BAP or active.

EM energy transmitted by cell phone towers (cellular base stations) is also designated as a ubiquitous source of energy of known wavelength (or frequencies in the 800 MHz to 1.9 GHz), and may likewise be optimally harvested and processed for supplying energy to wireless sensors, be they passive, BAP or active.

In addition a further improvement to the state of the art and provided by the embodiments herein is a hybrid RFID-WiFi transceiver that combines an RFID sensor tag reader and a WiFi router/transceiver into a device that can provide ambient EM energy to RFID sensor tags, read their unique identity and data from the tag and transmit this data over the Internet. The hybrid RFID-WiFi transceiver enables remote query of RFID sensor tags' identity and data via the Internet.

An additional improvement to the state of the art provided by the embodiments herein is a hybrid RFID-WiFi-cell phone transceiver that combines an RFID sensor tag reader, a WiFi router/transceiver, and a cell phone into a device that can provide ambient EM energy to RFID sensor tags, read their unique identity and data from the tag and transmit this data over the Internet and/or cellular telephone networks. The hybrid RFID-WiFi-cell phone transceiver enables remote query of RFID sensor tags' identity and data over the Internet as well as via cellular telephone networks.

The embodiments herein further improves upon the state of the art by making it possible for purely passive sensors to harvest available EM energy from their environment, and locally accumulate and store this energy so that passive tags become autonomous; i.e., they can log data and they can transmit an alarm or "read me!" signal to a nearby reader when a critical sensor event occurs or when a critical combination of sensor events occur or when a sensor level or combination of sensor levels has crossed a threshold.

FIG. **1** is an illustrative scenario of some of the forms of ambient EM energy that may be available for harvesting by passive RF addressable sensor tags **110** with energy harvesting capability. The ambient EM energy bathing the environment **115** in which the tags **110** are placed can come from a number of different sources including broadband WiFi routers/transceivers 120 (including but not limited to IEEE 802.1x, IEEE 802.15.4 "Zigbee®" electronic devices, Bluetooth® telecommunication equipment, HomeRF® computer hardware/software, HiperLAN/1 and HiperLAN/2), an EM transmitter **130** continuously or periodically radiating a known frequency (or frequencies), Microwave IEEE 521-1984 or cell phone base stations **140**, satellite transceivers **150**, wireless tag sensor readers **160**, and cell phones, PDAs or PCs **170**.

The energy that each of these sources contribute to the total amount of ambient EM energy available for harvesting by the antenna (not shown) of the tag **110**, $P_{RF}(t)$ **125**, **135**, **145**, **155**, **165**, **175** respectively, is a function of time, position/distance, frequency, solid angle and polarization state:

$$P_{RF}=\frac{1}{f_2-f_1}\int_{f_1}^{f_2}\int_0^{4\pi}\Phi(\Omega,f,t)A_{eff}(\Omega,f)d\Omega df$$

where Ω is the solid angle in steradians, $\Phi(\Omega f,t)$ is the time-varying frequency and angle-dependent incident power density and A_{eff} is the angle-, frequency-, and polarization-dependent effective area of the antenna.

The available power $P_{RF}(f_i)$ at any frequency f_i is an alternating current (ac) signal that must be converted to a direct current (dc) signal by a nonlinear device or rectifier (not shown) before being summed together with all other energy sources of different frequencies, $f_{1\ldots n}$, which have been similarly rectified. The dc power, P_{dc} at a single frequency, f_i is given by:

$$P_{dc}(f_i)=P_{RF}(t,f_i)n[P_{RF}(t,f_i),\beta]$$

where $\eta=P_{dc}/P_{RF}$ is the total received rectified and summed power conversion efficiency (including the loss due to reflected power), and $\beta(P_{dc},f_i,)$ represents the extent to which the rectifying device (not shown) and the antenna are impedance matched at the received dc power level, P_{dc} and frequency f_i.

The anticipated ambient power densities that can be harvested and converted to dc power at useful voltage levels is approximately in the range 10^{-1} mW/cm^2 (the FCC exposure limit for 880 MHz and 1990 MHz) to 10^{-6} mW/cm^2. One skilled in the art can optimize the harvested dc power, P_{dc} by appropriately adjusting the parameters A_{eff} and/or the parameters η and, β.

FIG. **2** is an illustrative example of a "smart home" 200 where RF addressable sensor tags **110** with energy harvesting capability are mounted on windows **210**, a door **220**, and in the attic **230**. Wired **270** or satellite transceiver **150** broadband Internet access is connected to a WiFi router **120** which bathes the smart house **200** in RF energy. Alternatively, an RF EM source **130** of known frequency continuously or periodically radiates energy specifically for the purpose of providing energy for the RF addressable sensor tags **110** with energy harvesting capability. An RFID tag reader **160** or a cell phone, PDA or PC **170** can read the unique identity and sensor data from each passive sensor tag **110**. The sensor tags **110** can be used to monitor environmental conditions such as temperature, humidity, carbon monoxide levels, radiation levels or smoke inside the house and the data read from the sensor tags **110** may be used to control HVAC systems or sound alarms when sensor thresholds are exceeded. The same sensor tags **110** can be used to detect unwanted intruders via vibration detectors, breakage detectors and open door detectors. Accordingly, the data from such sensors **110** can be used in a home security monitoring device. Similarly, addressable sensor tags **110** with energy harvesting capability can be placed on home appliances such as refrigerators, stoves, hot water heaters and gas/oil furnaces to monitor their operating state. Since the smart house **200** is connected to the Internet **270**, **150**, and/or via a cell phone **170**, all sensor data can be accessed locally and/or remotely over existing Internet and/or cellular telephone networks. Again, these sensors **110** harvest ambient energy from any or all of the available modes of RF EM energy bathing the house **200**. This energy could be used to power the passive sensor tags **110** or recharge batteries or other power sources associated with active or BAP tags **110**.

FIG. **3**A illustrates an example of a WiFi router/transceiver **120** connected to the Internet **270**, bathing passive, BAP or active wireless sensor tags **110** with energy at a known frequency **125**. This particular source of ambient energy supplies the passive sensor tag **110** with power or facilitates recharging batteries or other on-board or nearby energy storage devices. A single tag reader **160** (or a plurality of tag readers) is adapted to interrogate **330** each uniquely identifiable sensor tag **110** for information about its particular sensor tag reading(s) **320** and pass this information to the WiFi router/transceiver **120** via a direct connection **310** to make it locally or remotely accessible via the Internet and/or cell phone networks. This particular source of ambient energy can supply the passive sensor tag **110** equipped with energy harvesting capability with power for autonomous data collection and storage, or facilitate recharging batteries or other energy storage devices.

FIG. **3**B illustrates a hybrid RFID-WiFi wireless router **340** comprising an RFID tag reader module combined with a WiFi router/transceiver connected to the Internet **270**. This hybrid RFID-WiFi reader/module **340** provides a source of ambient energy **125** which the sensor tag **110** can harvest to supply itself with power to become an autonomous sensor **110** able to collect and save data as well as send out an alarm if its reading(s) has/have crossed a predefined threshold or a combination of predefined thresholds. The harvested energy can also facilitate recharging batteries or other energy storage devices associated with each tag **110**. The hybrid RFID-WiFi router/transceiver **340** periodically or sporadically interrogates **330** each uniquely identifiable sensor tag **110** for its data, and each uniquely identifiable sensor tag **110** rebroadcasts **320** its particular sensor(s) reading(s) back to the RFID-WiFi router/transceiver **340** which makes that data accessible over the Internet.

FIG. **3**C illustrates a hybrid RFID-WiFi-Cell phone router/transceiver **350** comprising an RFID tag reader module, a WiFi router/transceiver module and a cell phone module. This hybrid RFID-WiFi-Cell phone reader module **350** provides a source of ambient energy **125** which the sensor tag **110** can harvest to supply itself with power to become an autonomous sensor **110** able to collect and store data as well as send out an alarm if its reading(s) has/have crossed a predefined threshold or a combination of predefined thresholds. The harvested energy can also facilitate recharging batteries or other energy storage devices associated with each tag **110**. The hybrid RFID-WiFi-Cell phone router/transceiver **350** periodically or sporadically interrogates **330** each uniquely identifiable sensor tag **110** for its data, and each uniquely identifiable sensor tag **110** rebroadcasts **320** its particular sensor(s) reading(s) back to the hybrid RFID-WiFi-Cell phone router/transceiver **350** which is connected to the Internet **270** and makes that data accessible over the Internet as well as over cellular telephone networks.

FIG. **4**A illustrates an RFID temperature sensor tag **410** equipped with energy harvesting capability. Ambient EM energy **430** is harvested by the sensor tag **410**. This harvested energy is used by the tag **410** to continuously, periodically, or sporadically measure its sensor's value and compare that value with a pre-programmed threshold **470**. If this threshold is crossed, the tag **410** initiates and broadcasts an alarm or "read me!" signal **460** which is read by an RFID reader **160** or cell phone, PDA, or PC reader **170** along with its unique identification number. In this particular illustrative example the tag **410** senses temperature, but in general, the sensor or sensors of the tag **410** could measure a variety of environmental or physical variables. The unique identification and sensor data of the tag **410** can then be read by a standalone RFID reader **160** or by a cell phone, PDA, or PC reader **170**.

FIG. **4**B illustrates an RFID temperature sensor tag **410** equipped with energy harvesting capability that uses ambient EM energy **430** to provide a continuous, periodic, or sporadic source of power to the RFID temperature sensor tag **410**. The sensor tag **410** logs data to local memory **480** and can periodically or sporadically initiate and broadcast a signal **485**, which can be read by the RFID reader **160**. The RFID reader **160** can also periodically or sporadically initiate a read **490** of the unique identification and data of the tag **410**. In this particular illustrative example the tag **410** senses temperature, but in general, the sensor or sensors of the tag **410** could measure a variety of environmental or physical variables. The unique identification and sensor data of the tag **410** can then be read periodically or sporadically either when a sensor event initiates a "read me!" signal **485**, or when a standalone RFID reader **160** or a cell phone, PDA, or PC reader **170** initiates a read cycle **490**.

FIG. **5**A illustrates a passive wireless RF addressable sensor tag **500***a*, which may be used as a basis for designing the embodiments herein. The passive RFID tag **500***a* includes three main modules: the control module **510**, the tag module **540**, and the sensor module **520**. The control module **510** has an RFID tag controller **514** and a sensor processing controller **512** that can communicate with each other. The RFID tag controller **514** communicates with and controls the tag module **540**. The tag module **540** has on-board memory **542** and an RF communication interface **546**. The memory **542** holds the unique identification number and additional information **544** associated with that particular sensor tag **500***a*. The RF communication interface **546** has a voltage regulator **548** and a communication interface **550**. The RF communication interface **546** connects to a single antenna or a multitude of antennas **556** through an RF pad or a multitude of RF pads **554**. The sensor module **520** has its own memory **522** and a sensor interface and support sub-system **526**. The memory **522** stores the sensor data table **524** that include sensor calibration parameters, lookup tables as well as data collected from the sensor or sensors of the sensor tag **500***a*. The sensor interface and support sub-system **526** has a voltage and/or current reference generator **528**, an integrated temperature sensor **530** and an analog to digital converter **532**. An external sensor or a multitude of external sensors **536** are connected to the sensor interface and support sub-system **526** via a sensor pad or a multitude of sensor pads **534**.

FIG. **5**B illustrates an active or BAP wireless RF addressable sensor tag **500***b*, which may be used as a basis for designing the embodiments herein. In addition to all of the components described with FIG. **5**A above, there is also an active source module **560** which communicates with the control module **510**. The active source module **560** includes a power source **564** which could be replaceable and/or rechargeable, and voltage regulation circuitry **562** which regulates the voltage and/or current necessary to recharge the power source **564** and supply power to the remainder of the sensor tag **500***b*.

FIG. **6** is an illustrative example of a block diagram of the embodiments herein as it relates to the harvesting of ambient EM energy for supplying, storing, recharging, or supplementing the energy necessary to power active, BAP or passive wireless electronic devices, sensors and sensor networks. The EM energy harvesting sub-system architecture **600** comprises a single antenna **650**, rectifier and matching network **638** or an array of antennas **650**, rectifiers and matching networks **638**, a combiner **636**, a power conditioner **634**, and an energy storage device **632**. The energy harvesting controller **610** communicates bi-directionally with the energy harvesting module **620** and RFID tag controller **514**, wherein both the energy harvesting controller **610** and energy harvesting module **620** may or may not be integrated into an RFID sensor chip (not shown), depending on the particular application. The energy harvesting controller **610** sporadically, periodically or continuously controls and monitors the state of the energy harvesting module **620**. The energy harvesting module **620** includes an energy harvesting interface and support sub-system **630** that connects to an external energy harvesting antenna or multitude of external energy harvesting antennas **650** through an RFh (RF harvesting) pad or a multitude of RFh pads **640**.

The antenna or multitude of antennas **650** may be designed to collect ambient EM energy at either a single frequency f_1, a range of frequencies f_i, i=1, 2, 3 . . . N, or a specific set of frequencies, for example $f_{\{i,j,k\}l}$={13.56 MHz, 438 MHz, 928 MHz, 2.54 GHz}. Moreover, the polarization state of the energy within the aperture of a single energy harvesting antenna or a multitude of energy harvesting antennas **650** can be arbitrary. Furthermore, the aperture or effective area A_{eff} of an energy harvesting antenna or multitude of energy harvesting antennas **650** can be optimized for a wavelength, and/or a range or set of wavelengths of the received EM energy, and can be readily accomplished by those skilled in the art.

The EM energy harvested by an energy harvesting antenna or a multitude of energy harvesting antennas **650** presents an ac signal or a multitude of ac signals to the rectification and matching network **638**. The matching network **638** is preferably designed to optimally match the frequency, polarization state and signal strength (power) dependence $\beta(P_{dc}, f_i)$ of the output impedance of the harvesting antenna(s) **650** and the input impedance of the rectification element or elements **638** to ensure maximum power transfer or harvesting efficiency. The rectifier or rectifiers **638** convert the ac signal(s) into a dc or voltage signal(s).

The combiner **636** takes its input from the output of the rectification and matching network **638** and combines the currents or voltages in any combination of series or parallel connections in order to provide the necessary quantity of dc power for passive, BAP, or active sensor tag operation. The conditioner **634** takes the dc power from the combiner **636** and converts this dc power into a specified voltage level and/or current level with efficiency η by using circuit techniques known by those skilled in the art. The voltage and/or current level may be determined by the requirements of the local energy storage device **632** and/or the active source module **560**. The storage device **632** acts as local storage and may be implemented using any or a combination of capacitors, supercapacitors, plastic batteries, polymer batteries, Li-Ion batteries, or any rechargeable energy storing technology that can be integrated within the energy harvesting module **620**.

FIG. **7**A is an illustrative block diagram of an instantiation of a passive sensor tag **110***a* that includes a control module **710**, a tag module **540**, a sensor module **520**, and an energy harvesting module **620**. The energy harvesting controller **610** is shown here as being part of the control module **710** and communicates with the RFID tag controller **514** and the sensor processing controller **512**. The energy harvesting controller **610** is responsible for assessing the power needs of the tag module **540** and the sensor module **520** and can divert power from its storage element **632** to the tag module **540** and the sensor module **520** as necessary.

FIG. **7**B is an illustrative block diagram of an instantiation of a BAP or active sensor tag **110***b* that includes the energy harvesting module **620**. The difference between this sensor tag **110***b* of FIG. **7**B and sensor tag **110***a* of FIG. **7**A is the addition of the power module **720** whose purpose is to control and distribute power to the control module **730**, the tag module **540**, and the sensor module **520** in a BAP or active sensor tag **110***b*. The power module **720** includes, but is not limited to, a charger **725**, a power source **564**, and voltage regulation **562**.

The charger **725** controls the continuous, periodic, or sporadic charging of energy storage devices, in particular the power source **564** which may be a capacitor, supercapacitor, plastic battery, polymer battery, Li-Ion batteries, or any rechargeable energy storage technology. The voltage regulation sub-system **562** regulates the voltage and/or current that is needed to charge/recharge the power source **564** or divert power to the tag module **540** and/or the sensor module **520** as determined by the control module **730**.

FIG. **8** is a block diagram of another illustrative example of a sensor tag **110***c*with EM energy harvesting capability. As with the sensor tag **110***a* of FIG. **7**A and sensor tag **110***b* of FIG. **7**B, the sensor tag **110***c* of FIG. **8** comprises a control module **730**, a sensor module **520**, a tag module **830**, an energy harvesting module **810**, and an antenna module **840**. The tag module **830** differs from the tag module **540** of FIGS. **5**A, **5**B, **7**A, and **7**B in that there are no RF pads **554** and no antennas **556**. The remaining functionality of the tag module **830** is the same as described above.

Likewise, the energy harvesting module **810** differs from the energy harvesting module **620** of FIGS. **6**, **7**A, and **7**B in that there are no RF pads **554** and no RF antenna **556**, nor are there RFh pads **640** or energy harvesting antennas **650** associated with this module **810**. The remaining components, storage **632**, conditioner **634**, and combiner **636** have the same functionality as described above.

The antenna module **840** serves both the tag module **830** and the energy harvesting module **810**. An antenna or a multitude of antennas **846** are used for both RF EM energy harvesting and for bidirectional RF communications with an RFID, hybrid RFID-WiFi or hybrid RFID-WiFi-Cell phone reader. The antenna or antennas are designed to optimize RFID read range as well as to optimize A_{eff}. Each antenna **846** is connected via an RF pad **844** to the rectification and matching network **842**.

The rectification and matching network **842** is used to match the impedance of the antenna(s) **846** to that of the rectifier(s) as described above. In the context of FIG. **8**, the rectification and matching network **842** includes the rectifier(s). Moreover, those skilled in the art would readily understand that diodes, transmission lines, microstrip elements, and surface mount components could be further included in the rectification and matching network **842** in order to provide the rectifier(s). Since this network **842** also allows RF communications to occur, frequencies corresponding to those used by the RFID, RFID-WiFi, or RFID-WiFi-Cell phone reader (not shown in FIG. **8**) are preferably isolated and directed to the communication interface **550** on the tag module **830**. The harvested EM energy, once rectified, is sent first to the combiner **636**, then to the conditioner **634**, and finally to storage **632** as described above.

FIG. **9**A is an illustrative example of a block diagram of one of many possible designs for an EM RF energy harvesting antenna array module **910** in accordance with the embodiments herein, with the functionality of the various components included therein as previously described above. FIG. **9**B is an illustrative example of a block diagram of an alternate method for harvesting EM RF energy. In this aspect of the embodiments herein, an EM RF energy harvesting rectenna array module **950** is provided as a possibly more efficient way to harvest ambient EM RF energy.

In FIG. **9**B, an array of rectennas **970**, $\text{Rectenna}_1, \ldots, \text{Rectenna}_n$, each comprise a pair of antennas **982**, **984**, optimized to collect a broad range of RF frequencies having time-varying polarization states, directly integrated with an impedance matched rectifier **980**. Each rectenna **970** could be made up of two interdigitated antennas, one with a clockwise winding **982**, the other with a counterclockwise winding **984**, with the rectifier **980** connected between the two inner ports of the antennas. The rectified signals are collected from the two outermost ports of the rectenna **970** and connected to pads $\text{RHh}_1, \ldots, \text{RFh}_n$, **960**.

The techniques provided by the embodiments herein may be implemented on an integrated circuit (IC) chip or using printable electronic technologies (not shown). The chip or printable electronic circuit design is created in a graphical computer programming language, and stored in a computer storage medium (such as a disk, tape, physical hard drive, or virtual hard drive such as in a storage access network). If the designer does not fabricate chips or printable electronic circuits or the photolithographic masks used to fabricate chips or printable electronic circuits, the designer transmits the resulting design by physical means (e.g., by providing a copy of the storage medium storing the design) or electronically (e.g., through the Internet) to such entities, directly or indirectly. The stored design is then converted into the appropriate format (e.g., GDSII or CIF) for the fabrication of photolithographic masks, which typically include multiple copies of the chip design in question that are to be formed on a wafer or printed on a suitable substrate. The photolithographic masks are utilized to define areas of the wafer or printable electronic circuits (and/or the layers thereon) to be etched or otherwise processed or printed.

The resulting integrated circuit chips or printable electronic circuits can be distributed by the fabricator in raw wafer form (that is, as a single wafer that has multiple unpackaged chips), as a bare die, or in a packaged form or as individual printed circuits or in a sheet or roll of printed circuits. In the latter case the chip is mounted in a single chip package (such as a plastic carrier, with leads that are affixed to a motherboard or other higher level carrier) or in a multichip package (such as a ceramic carrier that has either or both surface interconnections or buried interconnections). In any case the chip might then be integrated with other chips, discrete circuit elements, and/or other signal processing devices as part of either (a) an intermediate product, such as a mother or daughter-board, or (b) an end product. The end product can be any product that includes integrated circuit chip or chips and/or printed circuits, ranging from toys and other low-end applications to advanced computer products having a display, a keyboard or other input device, and a central processor.

Harvesting ambient EM energy would transform a purely passive sensor into an autonomous sensor **110***a*, **110***b*, **111***c* capable of periodic or sporadic sensor data acquisition, data logging and storage at times other than when being actively interrogated by a reader **160** as well as providing the necessary energy to either autonomously transmit a notable "sensor event", broadcast an alarm or "read me!" signal **460**, **485**, **490** when a threshold of the sensor **110***a*, **110***b*, **110***c* has been exceeded or when a combination of various thresholds of the sensor **110***a*, **110***b*, **110***c* have been exceeded.

The combination of passive and/or BAP and/or active wireless sensors **110***a*, **110***b*, **110***c* having ambient EM energy harvesting capabilities and wireless communication capabilities as provided by the embodiments herein provide a new low cost means for remotely sensing the environment, monitoring the state of a home, buildings, structures, goods and devices as well as remote and local personal/patient monitoring, medical diagnostic and medical care solutions.

Cell phones, PDAs and PCs **170** are pervasive electronic devices that are already networked via the Internet. A combination or integration of RFID readers, WiFi technologies and/or cell phone technologies into a hybrid reader that transparently communicates with RFID sensor tags, the Internet and cell phone networks would facilitate the widespread deployment and usage of passive sensor tags equipped with EM energy harvesting technology.

Broadband wireless communications networks such as WiFi (including WiMax™ telecommunication equipment and WiBro™ telecommunication equipment and any combination of IEEE 802.1x (where x is used to convey any or all of the following: 1a, 1b, 1g, 5 or 6), IEEE 802.15.4 "Zigbee®" electronic devices, Bluetooth® telecommunication equipment, HomeRF® computer hardware/software, HiperLAN/1 and HiperLAN/2) are being widely deployed in offices, factories, warehouses, stores, homes, communities, and cities. With long read ranges (up to several hundred feet), RFID tags and RFID readers can benefit from aspects of the embodiments herein by: using RFID and/or WiFi for short range communication between a tag reader and wireless sensor tags; using the WiFi and/or the cellular network infrastructure to communicate the acquired sensor tag data over the Internet or cellular telephone networks (bi-directionally); and using WiFi routers/transceivers or cellular base stations as a source of ambient EM energy, with known frequency, that can be optimally harvested to provide energy to inexpensive passive (BAP and/or active tags with rechargeable energy sources) sensor tags.

Generally, the embodiments herein provide a technique for harvesting ambient RF EM energy for supplying, storing, recharging or supplementing the energy necessary to power active, BAP or passive wireless electronic devices, sensors and sensor networks, allowing them to become autonomous "data loggers" and/or empowering them to transmit their unique identification and sensor information when a sensed variable exceeds a threshold and triggers a vital "sensor event". The data gathered by these autonomous data loggers can also be transparently read at any time by a variety of reader technologies such as RFID readers, cell phones, personal data assistants, personal computers, WiFi, WiMax™telecommunication equipment and WiBro™ telecommunication equipment, including IEEE 802.1x (where x is used to convey any or all of the following: 1a, 1b, 1g, 5, and 6), "Zigbee®" electronic devices, Bluetooth® telecommunication equipment, HomeRF® computer hardware/software, HiperLAN/1and HiperLAN/2 and any hybrid combinations thereof. The embodiments herein also provide a hybrid RFID-WiFi reader/transceiver **340** and a hybrid RFID-WiFi-cell phone reader/transceiver **350**. The RFID reader module communicates with RFID tags, the WiFi module supplies ambient energy which the passive sensor tags are equipped to harvest and provides connectivity to the Internet and may also read WiFi sensor tags, and the cell phone module provides connectivity to cellular telephone networks.

The foregoing description of the specific embodiments will so fully reveal the general nature of the embodiments herein that others can, by applying current knowledge, readily modify and/or adapt for various applications such specific embodiments without departing from the generic concept, and, therefore, such adaptations and modifications should and are intended to be comprehended within the meaning and range of equivalents of the disclosed embodiments. It is to be understood that the phraseology or terminology employed herein is for the purpose of description and not of limitation. Therefore, while the embodiments herein have been described in terms of preferred embodiments, those skilled in the art will recognize that the embodiments herein can be practiced with modification within the spirit and scope of the appended claims.

What is claimed is:

1. An electronic device comprising:

a control module comprising an energy harvesting controller;

a tag module controlled by said control module;

a sensor module controlled by said control module;

an energy harvesting module controlled by said energy harvesting controller and adapted to wirelessly collect ambient radio frequency (RF) electromagnetic (EM) energy; and

a power module adapted to provide a power supply for said control module, said power supply being generated by the collected ambient RF EM energy,

wherein said energy harvesting controller is adapted to (i) assess power requirements of each of the tag and sensor modules, (ii) transfer power from said energy harvesting module to any of said tag and sensor modules, and (iii) charge a local energy storage device.

2. The electronic device of claim **1**, further comprising an antenna module shared by said energy harvesting module and said tag module.

3. The electronic device of claim **2**, wherein said antenna module comprises:

a rectification and impedance matching network;

at least one RF pad operatively connected to said rectification and impedance matching network; and

at least one antenna operatively connected to each of said at least one RF pad.

4. The electronic device of claim **1**, wherein said tag module comprises:

a memory component;

a RF communication interface operatively connected to said memory component;

at least one RF pad operatively connected to said RF communication interface; and

at least one antenna operatively connected to each of said at least one RF pad.

5. The electronic device of claim **1**, wherein said sensor module comprises:

a memory component;

a sensor interface and support sub-system operatively connected to said memory component;

at least one sensor pad operatively connected to said sensor interface and support sub-system; and

at least one antenna operatively connected to each of said at least one sensor pad.

6. The electronic device of claim **1**, wherein said sensor module comprises:

a memory component;

a sensor interface and support sub-system operatively connected to said memory component; and

at least one antenna operatively connected to sensor interface and support sub-system.

7. The electronic device of claim **1**, wherein said control module further comprises:

a radio frequency identification (RFID) tag controller; and

a sensor processing controller operatively connected to said RFID tag controller.

8. The electronic device of claim **1**, wherein said energy harvesting module comprises:

an energy harvesting interface and support sub-system;

at least one RF harvesting pad operatively connected to an energy harvesting interface and support sub-system; and

at least one antenna operatively connected to each of said at least one RF harvesting pad.

9. The electronic device of claim **1**, wherein said energy harvesting module comprises:

an energy harvesting interface and support sub-system;

at least one pair of RF harvesting pads operatively connected to said energy harvesting interface and support sub-system; and

at least one rectenna operatively connected to each of said at least one pair RF harvesting pads.

10. The electronic device of claim **9**, wherein said energy harvesting interface and support sub-system comprises:

a storage component;

a power conditioner component operatively connected to said storage component; and

a combiner operatively connected to said power conditioner.

11. The electronic device of claim **10**, wherein said energy harvesting interface and support sub-system comprises a rectification and impedance matching network operatively connected to said combiner.

12. The electronic device of claim **1**, wherein said power module comprises:

a power charger;

a power source operatively connected to said power charger; and

a voltage regulator operatively connected to said power source.

13. The electronic device of claim **1**, wherein said collected ambient RF EM energy enables said electronic device to be a fully autonomous data collector, alarm generator, event generator, data logger, and data processor independent of any other device.

14. The electronic device of claim **1**, wherein said power supply is generated only by said collected ambient RF EM energy.

15. A system comprising:

an ambient radio frequency (RF) electromagnetic (EM) energy source;

a sensor reader operatively connected to said ambient RF EM energy source; and

a sensor in communication with said sensor reader, said sensor comprising:

a control module comprising an energy harvesting controller;

a tag module controlled by said control module;

a sensor module controlled by said control module;

an energy harvesting module controlled by said energy harvesting controller and adapted to wirelessly collect said ambient RF EM energy; and

a power module adapted to provide a power supply for said control module, said power supply being generated by the collected ambient RF EM energy,

wherein said energy harvesting controller is adapted to (i) assess power requirements of each of the tag and sensor modules, (ii) transfer power from said energy harvesting module to any of said tag and sensor modules, and (iii) charge a local energy storage device.

16. The system of claim **15**, wherein said sensor further comprises an antenna module shared by said energy harvesting module and said tag module.

17. The system of claim **16**, wherein said antenna module comprises:

a rectification and impedance matching network;

at least one RF pad operatively connected to said rectification and impedance matching network; and

at least one antenna operatively connected to each of said at least one RF pad.

18. The system of claim **15**, wherein said tag module comprises:

a memory component;

a RF communication interface operatively connected to said memory component;

at least one RF pad operatively connected to said RF communication interface; and

at least one antenna operatively connected to each of said at least one RF pad.

19. The system of claim **15**, wherein said sensor module comprises:

a memory component;

a sensor interface and support sub-system operatively connected to said memory component;

at least one sensor pad operatively connected to said sensor interface and support sub-system; and

at least one antenna operatively connected to each of said at least one sensor pad.

20. The system of claim **15**, wherein said sensor module comprises:

a memory component;

a sensor interface and support sub-system operatively connected to said memory component; and

at least one antenna operatively connected to sensor interface and support sub-system.

21. The system of claim **15**, wherein said control module further comprises:

a radio frequency identification (RFID) tag controller; and

a sensor processing controller operatively connected to said RFID tag controller.

22. The system of claim **15**, wherein said energy harvesting module comprises:

an energy harvesting interface and support sub-system;

at least one RF harvesting pad operatively connected to an energy harvesting interface and support sub-system; and

at least one antenna operatively connected to each of said at least one RF harvesting pad.

23. The system of claim **15**, wherein said energy harvesting module comprises:

an energy harvesting interface and support sub-system;

at least one pair of RF harvesting pads operatively connected to said energy harvesting interface and support sub-system; and

at least one rectenna operatively connected to each of said at least one pair RF harvesting pads.

24. The system of claim **15**, wherein said energy harvesting interface and support sub-system comprises:

a storage component;

a power conditioner component operatively connected to said storage component; and

a combiner operatively connected to said power conditioner.

25. The system of claim **24**, wherein said energy harvesting interface and support sub-system comprises a rectification and impedance matching network operatively connected to said combiner.

26. The system of claim **15**, wherein said power module comprises:

a power charger;

a power source operatively connected to said power charger; and

a voltage regulator operatively connected to said power source.

27. The system of claim **15**, wherein said collected ambient RF EM energy enables said electronic device to be a fully autonomous data collector, alarm generator, event generator, data logger, and data processor independent of any other device.

28. The system of claim **15**, wherein said sensor reader comprises any of a cell phone, a personal display assistant (PDA), a personal computer (PC) reader, a radio frequency identification (RFID) reader, a broadband wireless fidelity (WiFi) device, and a combination thereof.

29. The system of claim **15**, wherein said power supply is generated only by said collected ambient RF EM energy.

30. The system of claim **15**, wherein said sensor is adapted to transmit a wireless signal to said sensor reader, and wherein said sensor reader is adapted to receive the transmitted wireless signal from said sensor.

31. The system of claim **15**, wherein said collected ambient RF EM energy facilitates transmission of said wireless signal when a selected sensor variable meets a predetermined threshold.

32. The system of claim **15**, wherein said ambient RF EM energy source comprises a wireless fidelity (WiFi) router adapted to allow said sensor reader to communicate over the Internet.

33. The system of claim **15**, wherein said ambient RF EM energy source comprises a hybrid wireless fidelity (WiFi) router transceiver adapted to allow said sensor reader to communicate over the Internet.

34. The system of claim **15**, wherein said sensor is a RF ambient EM energy harvesting sensor and is arranged with a plurality of other said sensors each adapted to collect data for climate control, security alarms, environmental alarms, and information gathering from other devices located within a communicable distance from said sensor, wherein all of said sensors are adapted to be any of locally and remotely monitored via any of the Internet and a cell phone network.

35. A device comprising:

a control module comprising:

an energy harvesting controller;

a radio frequency identification (RFID) tag controller operatively connected to said energy harvesting controller; and

a sensor processing controller operatively connected to said RFID tag controller;

a tag module controlled by said control module;

a sensor module controlled by said control module; and

an energy harvesting module directly connected to said control module and controlled by said energy harvesting controller, wherein said energy harvesting module is adapted to wirelessly collect, process, and store ambient radio frequency (RF) electromagnetic (EM) energy,

wherein said energy harvesting controller is adapted to (i) assess power requirements of each of the tag and sensor modules, (ii) transfer power from said energy harvesting module to any of said tag and sensor modules, and (iii) charge a local energy storage device, and

wherein power for operation of said control module, said tag module, said sensor module, and said energy harvesting module is generated only by the collected ambient RF EM energy.

* * * * *

US008552597B2

(12) United States Patent
Song et al.

(10) **Patent No.: US 8,552,597 B2**
(45) **Date of Patent: Oct. 8, 2013**

(54) PASSIVE RF ENERGY HARVESTING SCHEME FOR WIRELESS SENSOR

(75) Inventors: **Zhen Song**, Plainsboro, NJ (US); **Chellury R. Sastry**, South Brunswick, NJ (US)

(73) Assignee: **Siemens Corporation**, Iselin, NJ (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 1542 days.

(21) Appl. No.: **11/691,545**

(22) Filed: **Mar. 27, 2007**

(65) **Prior Publication Data**

US 2012/0256492 A1 Oct. 11, 2012

Related U.S. Application Data

(60) Provisional application No. 60/744,028, filed on Mar. 31, 2006.

(51) **Int. Cl.**
H02J 17/00 (2006.01)

(52) **U.S. Cl.**
USPC **307/149**; 307/104

(58) **Field of Classification Search**
USPC 307/104, 149
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2,462,102	A	2/1949	Istvan	
2,554,270	A	5/1951	Rosenberg	
2,805,332	A	9/1957	Bell	
3,327,222	A	6/1967	King, Jr	
3,987,448	A *	10/1976	Scheppman	343/702
3,996,522	A	12/1976	Koch et al.	
4,229,742	A	10/1980	Rotunda	
5,075,670	A	12/1991	Bower et al.	
5,151,848	A	9/1992	Finello	
5,585,999	A	12/1996	De Long et al.	
6,037,743	A	3/2000	White et al.	
6,259,372	B1	7/2001	Taranowski et al.	
6,417,653	B1	7/2002	Massie et al.	
6,691,236	B1 *	2/2004	Atkinson	713/320
6,825,754	B1	11/2004	Rolin	
6,856,291	B2 *	2/2005	Mickle et al.	343/701
6,882,128	B1	4/2005	Rahmel et al.	

(Continued)

FOREIGN PATENT DOCUMENTS

WO	WO 01 37497	A1	5/2001

OTHER PUBLICATIONS

Pervasive Computing, Published by the IEEE CS and IEEE ComSoc, 1536-1268/05, copyright 2005 IEEE. Energy Scavenging for Mobile and Wireless Electronics, by Joseph A. Paradiso and Thad Starner. pp. 18-27.*

Primary Examiner — Fritz M Fleming

(57) ABSTRACT

A radio frequency energy harvester harvests RF energy from the environment. The energy harvester may include a local power source. The energy harvester may have an associated crystal radio antenna that harvests AM radio waves ubiquitously. The radio frequency energy harvested may be stored via a super capacitor. The stored energy may power a self-contained processing unit directly when the level of the energy harvested remains above a threshold level. However, if the energy harvested drops below the threshold level, the processing unit, which may include a radio, may switch to the local power source and/or a low power mode of operation. The low power mode of operation may reduce the amount of processing and/or transmitting by the radio. The processing unit may be part of a network of radios. The energy harvester may supply energy for a variety of applications, including building automation, industrial automation, power generation, and healthcare.

8 Claims, 9 Drawing Sheets

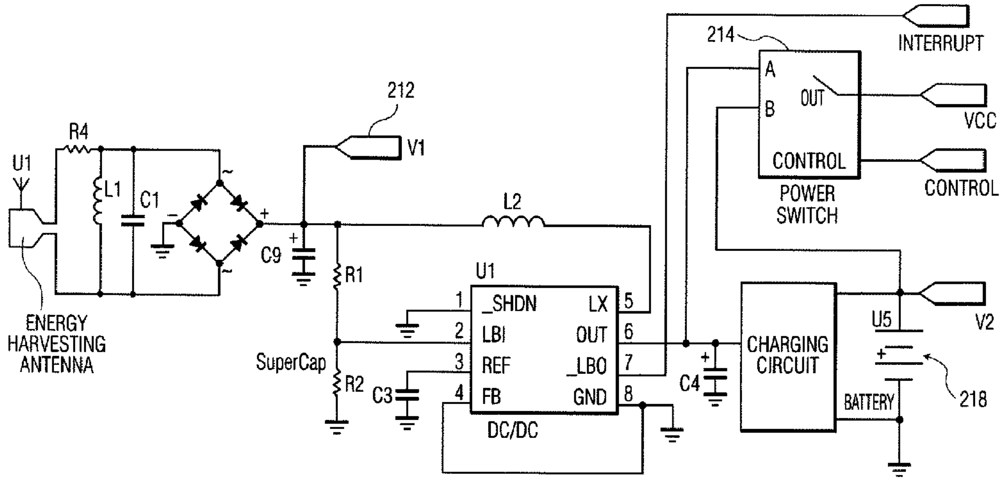

(56) References Cited

U.S. PATENT DOCUMENTS

6,924,772	B2	8/2005	Kiernan, Jr. et al.	
7,038,630	B1	5/2006	Bally et al.	
7,089,089	B2	8/2006	Cumming et al.	
7,132,757	B2 *	11/2006	Steigerwald et al.	290/1 R
7,139,239	B2	11/2006	McFarland	
2002/0116460	A1	8/2002	Treister et al.	
2005/0115600	A1 *	6/2005	DeSteese et al.	136/205

* cited by examiner

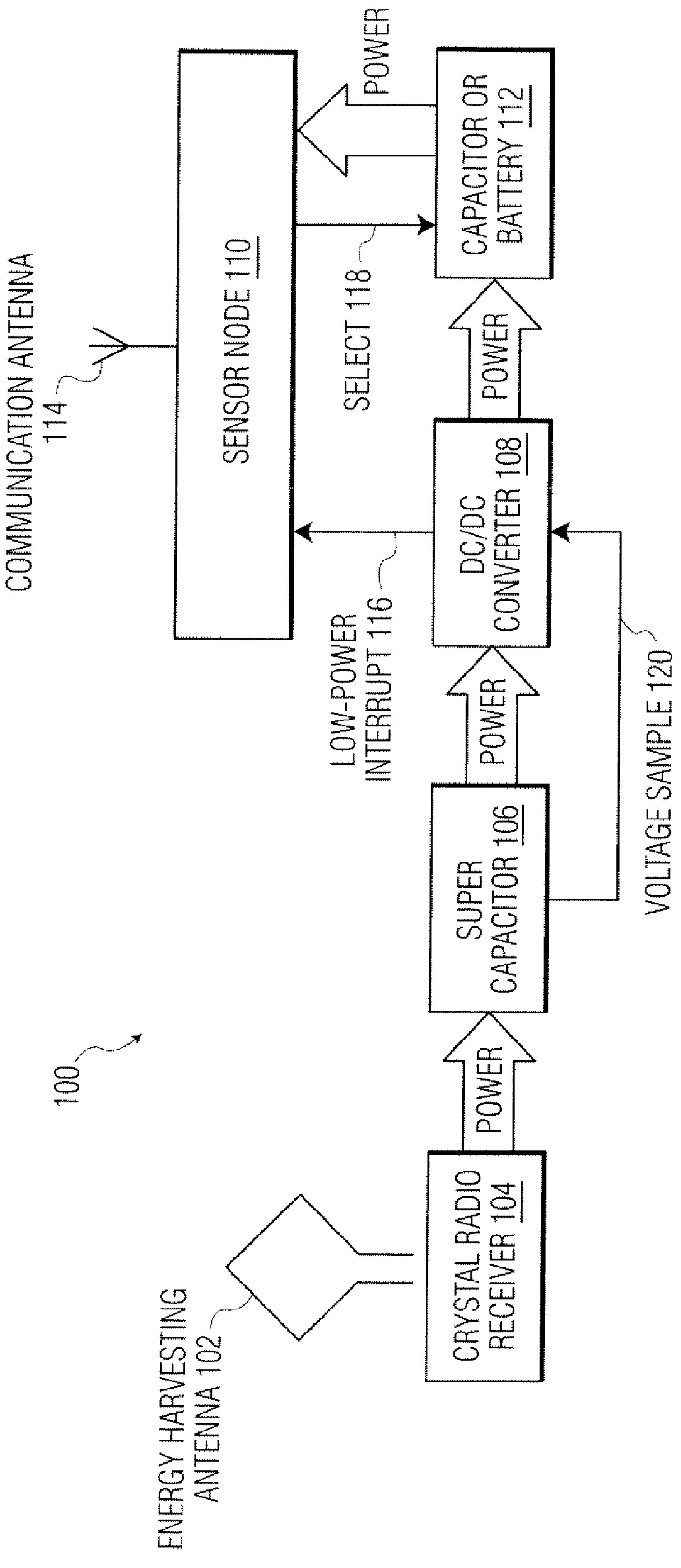
100
ENERGY HARVESTING ANTENNA 102
CRYSTAL RADIO RECEIVER 104
POWER
SUPER CAPACITOR 106
POWER
DC/DC CONVERTER 108
POWER
CAPACITOR OR BATTERY 112
POWER
SENSOR NODE 110
COMMUNICATION ANTENNA 114
LOW-POWER INTERRUPT 116
SELECT 118
VOLTAGE SAMPLE 120

FIG. 1

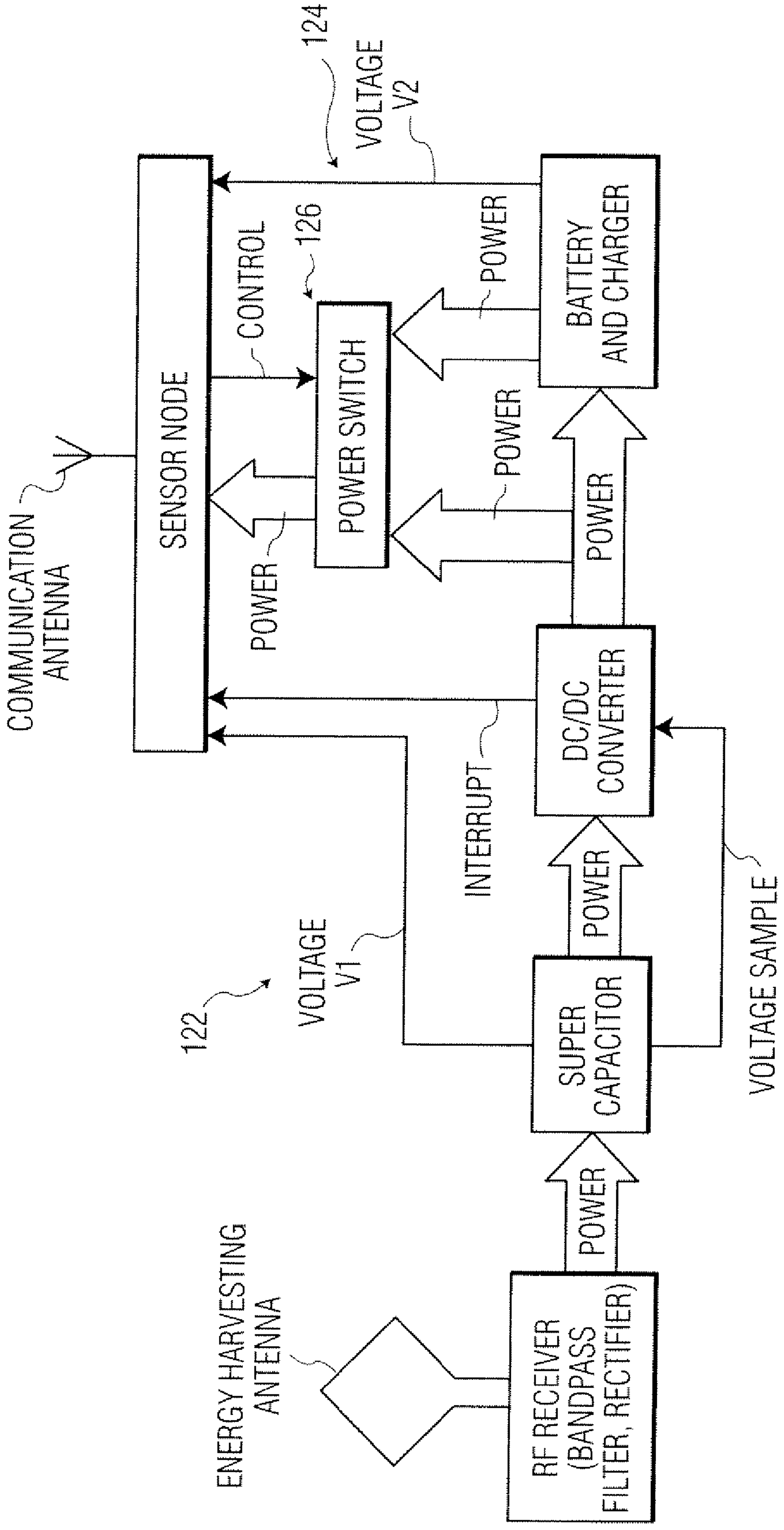
ENERGY HARVESTING ANTENNA
RF RECEIVER (BANDPASS FILTER, RECTIFIER)
POWER
SUPER CAPACITOR
122
VOLTAGE V1
VOLTAGE SAMPLE
POWER
DC/DC CONVERTER
INTERRUPT
POWER
POWER
COMMUNICATION ANTENNA
SENSOR NODE
POWER
POWER SWITCH
CONTROL
126
POWER
BATTERY AND CHARGER
VOLTAGE V2
124

FIG. 1a

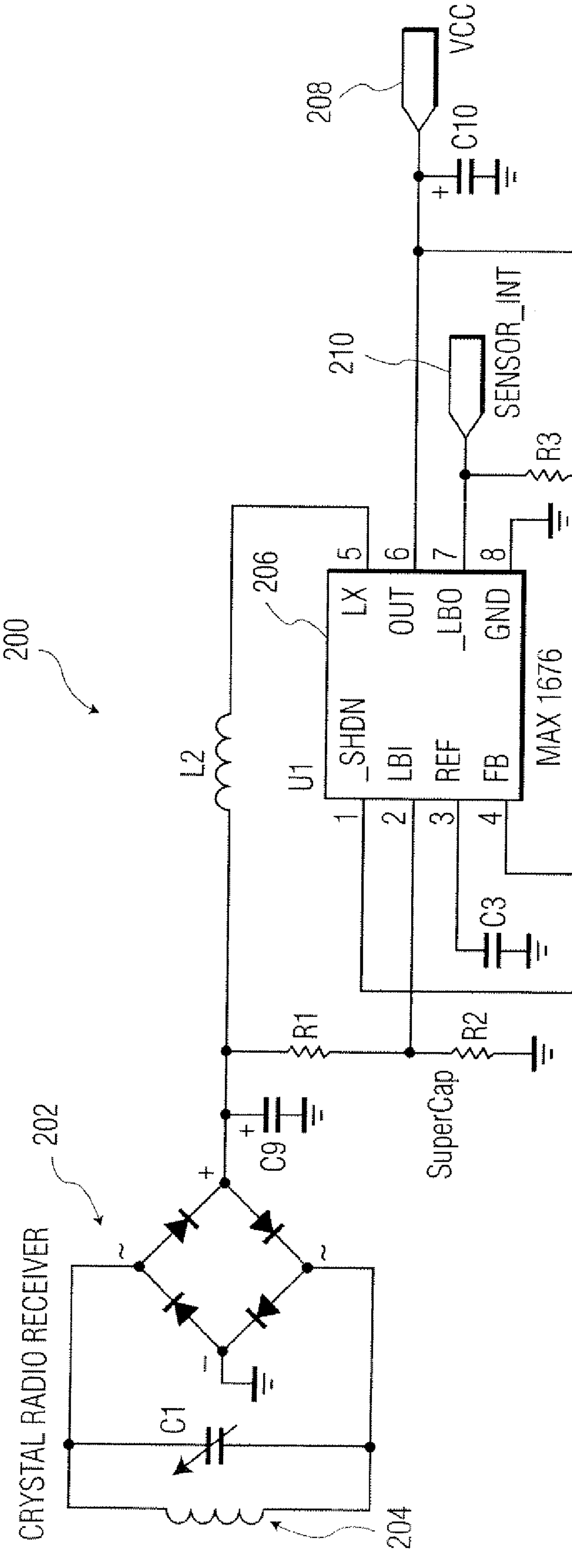
CRYSTAL RADIO RECEIVER
204
C1
202
~
~
+
-
C9
+
SuperCap
R1
R2
L2
U1
206
200
_SHDN
LBI
REF
FB
LX
OUT
_LBO
GND
MAX 1676
1
2
3
4
5
6
7
8
C3
R3
210
SENSOR_INT
C10
+
208
VCC

FIG. 2

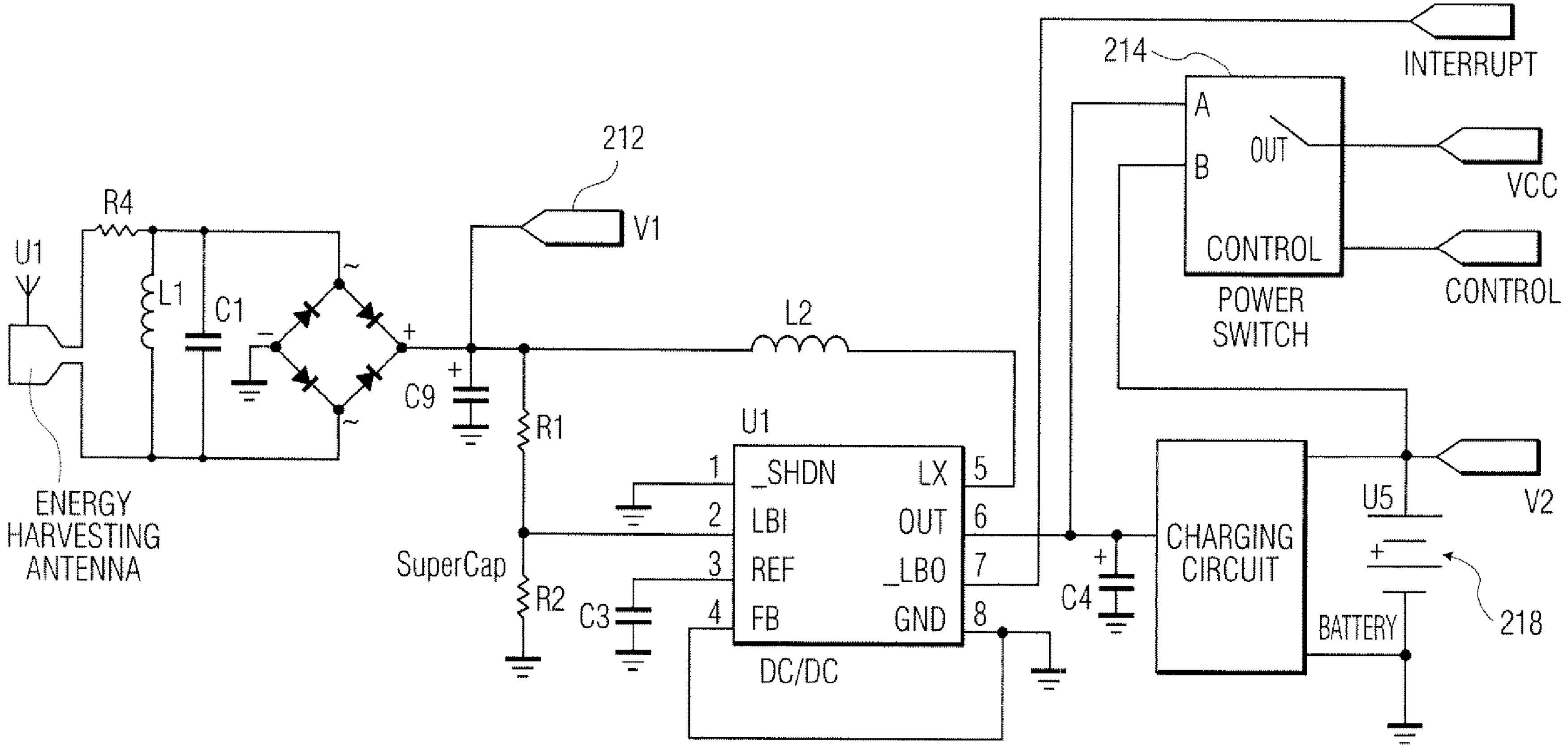
U1
R4
L1
C1
ENERGY HARVESTING ANTENNA
−
~
+
~
C9
+
212
V1
R1
SuperCap
R2
L2
U1
C3
1
2
3
4
_SHDN
LBI
REF
FB
LX
OUT
_LBO
GND
5
6
7
8
DC/DC
C4
+
214
A
B
OUT
CONTROL
POWER SWITCH
INTERRUPT
VCC
CONTROL
CHARGING CIRCUIT
U5
+
BATTERY
V2
218

FIG. 2a

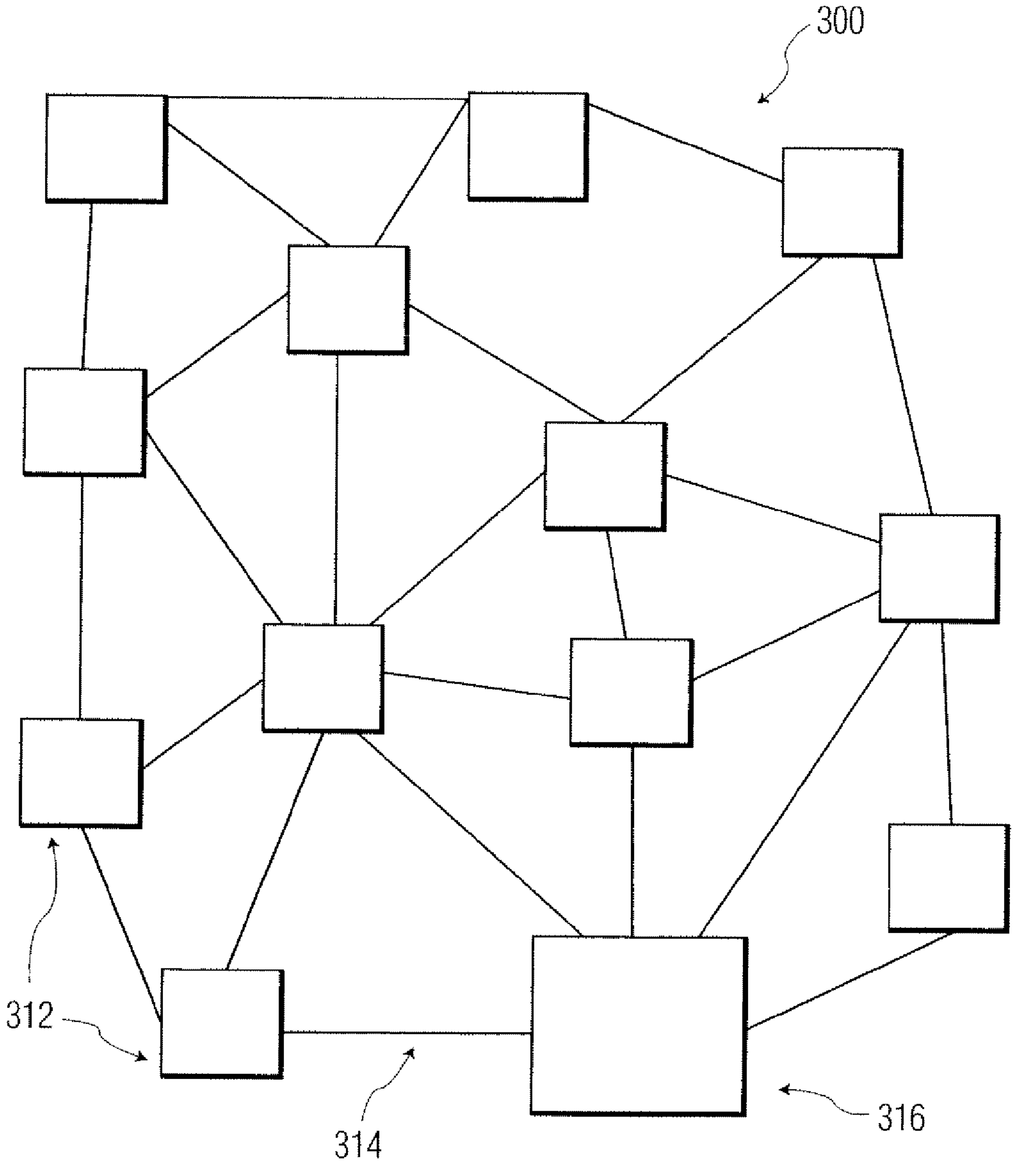
300
312
314
316

FIG. 3

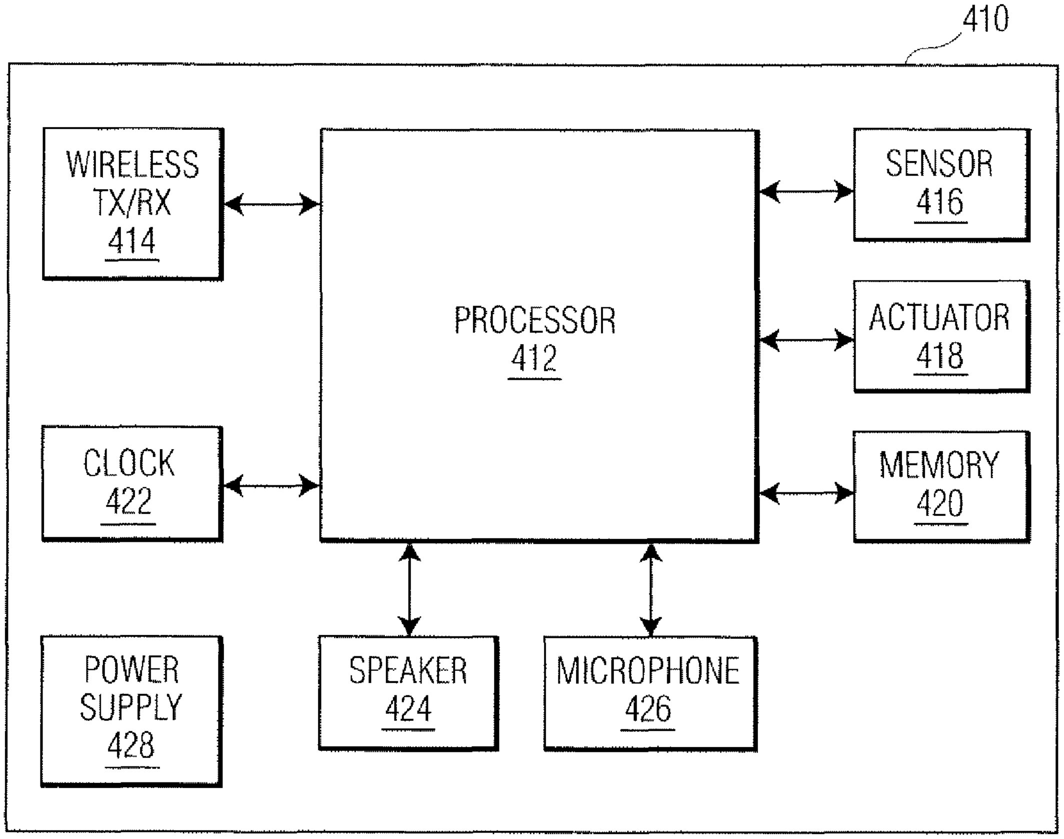
410
WIRELESS TX/RX 414
PROCESSOR 412
SENSOR 416
ACTUATOR 418
CLOCK 422
MEMORY 420
POWER SUPPLY 428
SPEAKER 424
MICROPHONE 426

FIG. 4

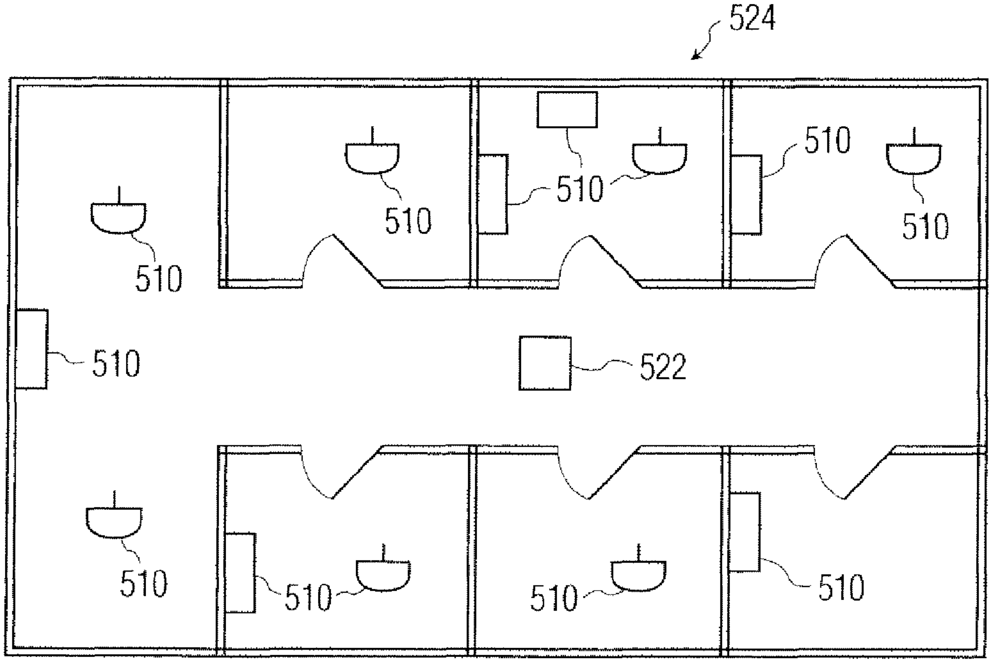

FIG. 5

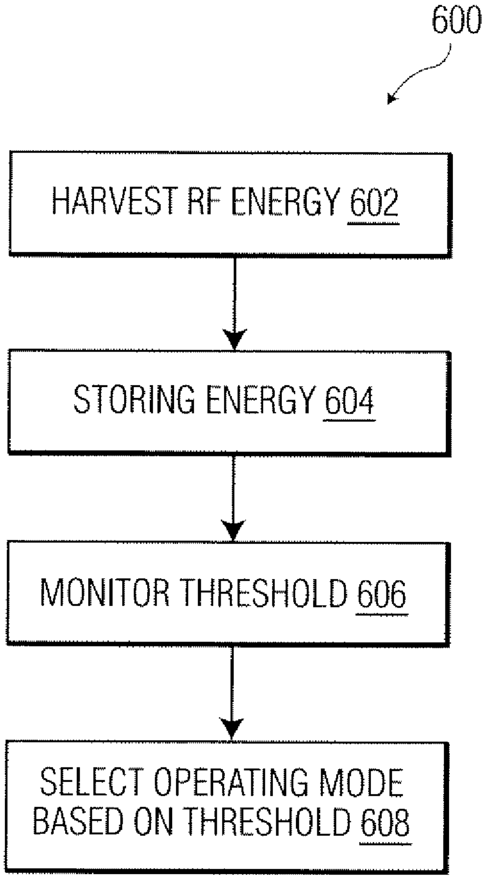
600
HARVEST RF ENERGY 602
STORING ENERGY 604
MONITOR THRESHOLD 606
SELECT OPERATING MODE BASED ON THRESHOLD 608

FIG. 6

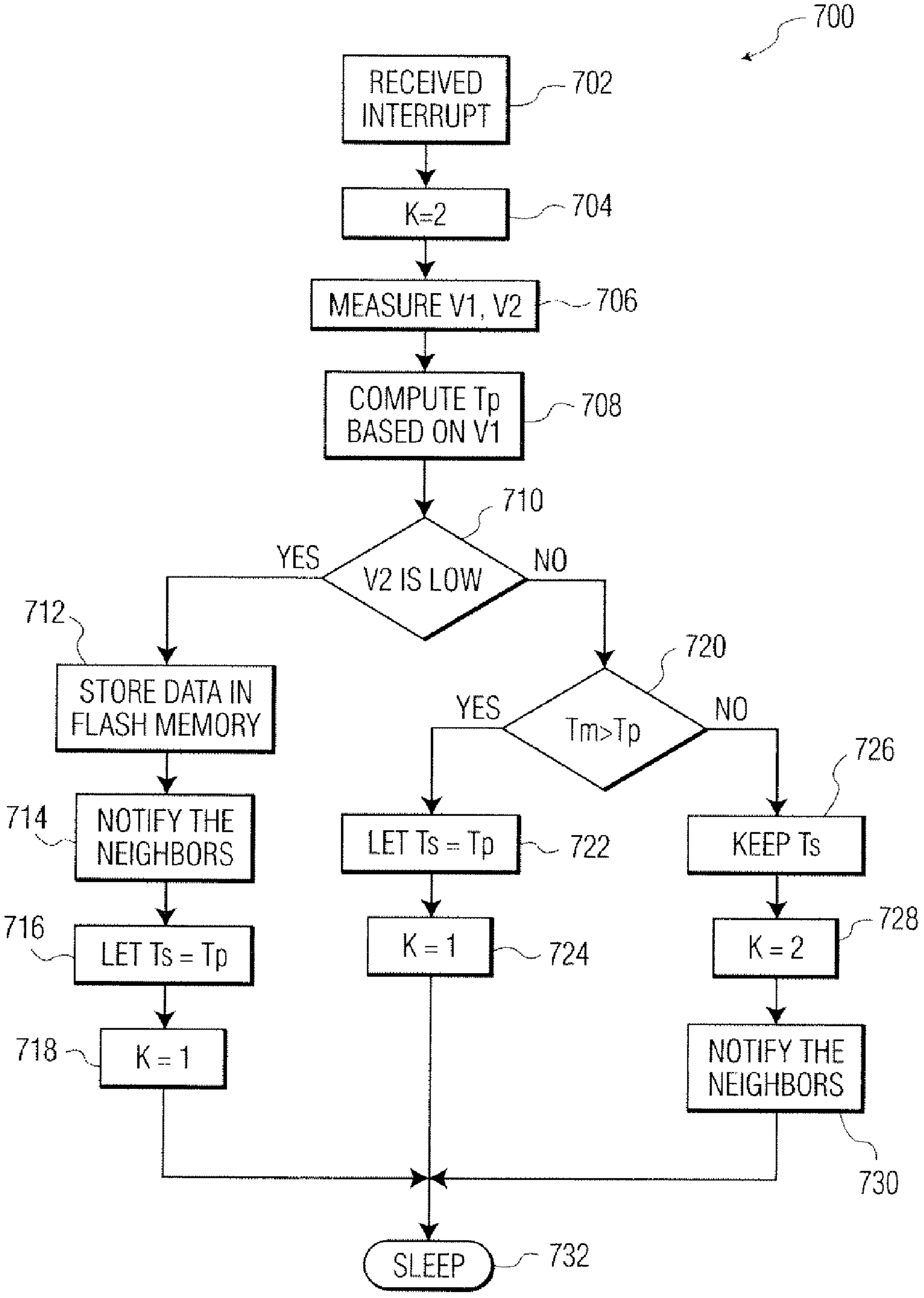
700
RECEIVED INTERRUPT
702
K=2
704
MEASURE V1, V2
706
COMPUTE Tp BASED ON V1
708
710
YES
V2 IS LOW
NO
712
STORE DATA IN FLASH MEMORY
720
YES
Tm>Tp
NO
726
714
NOTIFY THE NEIGHBORS
LET Ts = Tp
722
KEEP Ts
716
LET Ts = Tp
K = 1
724
K = 2
728
718
K = 1
NOTIFY THE NEIGHBORS
730
SLEEP
732

FIG. 7

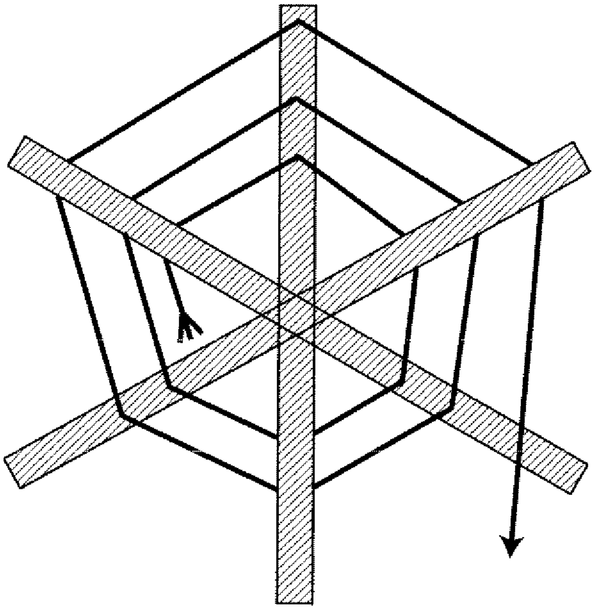

FIG. 8

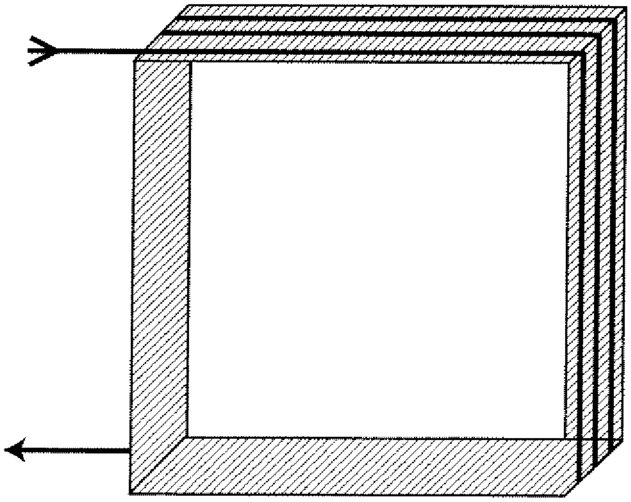

FIG. 9

PASSIVE RF ENERGY HARVESTING SCHEME FOR WIRELESS SENSOR

PRIORITY AND CROSS-REFERENCE TO RELATED APPLICATION

This application claims priority under 35 U.S.C. §119(e) to provisional application Ser. No. 60/744,028, filed on Mar. 31, 2006, which is incorporated by reference in its entirety herein.

BACKGROUND

The present embodiments relate generally to harvesting environmental energy. More particularly, the present embodiments relate to harvesting energy from radio frequency (RF) signals to power a sensor node and/or wireless radio.

Wireless sensor networks (WSNs) of so-called "smart" sensors can be used in remote monitoring and control applications. However, in almost all applications that involve WSNs, there is a trade off between the need to sense and communicate useful information and the requirement to do so in an energy efficient manner. A WSN, once deployed, may be intended to work for long periods of time with only minimal human intervention. For instance, remote monitoring applications may include individual sensor nodes, numbering into the hundreds or even thousands, that each run on a limited supply of energy, such as a battery. As a result, with conventional sensor nodes, excessive on-board processing and, perhaps more importantly, frequent invocation of the sensor node's radio may rapidly deplete the on-board energy source. Therefore, there is a need to develop sensor network protocols and algorithms with energy-efficient and self-organizing capabilities.

Additionally, replacing depleted batteries may be inefficient or impractical. A network may include a large number of individual sensor nodes that may be difficult to access due to being implanted within a wall, a ceiling, a larger device, or a patient. Accordingly, there is also a need for novel energy harvesting schemes operable to power wireless sensor nodes, as well as automatically replenish on-board energy storage units.

BRIEF SUMMARY

By way of introduction, the embodiments described below include methods, processes, apparatuses, instructions, or systems for harvesting energy from freely available radio frequency (RF) signals. The energy harvested may power processing units, such as self-contained processing units or sensor nodes. A "passive RF energy harvesting system," or "harvesting system," may be connected to a processing unit having a processor and delivers power supply to the processing unit. Each harvesting system may be associated with a crystal radio receiver and/or an all pass or band pass filter operable to harvest energy from all radio frequency bands simultaneously or ubiquitously. As a result, more of the available RF energy may be harvested from the existing electromagnetic signals. The harvesting is ubiquitous since no tuning is required if the harvester is moved from place to place. Most of the RF energy scavenged may be harvested from amplitude modulated (AM) radio signals. The harvesting system may have a super capacitor that temporarily stores the RF energy captured. If the level of RF energy being harvested falls below a threshold level, an interrupt signal may be generated that directs the processing unit to conserve energy. For

instance, the processing unit may limit on-board processing or reduce the number and/or strength of signals transmitted via a wireless radio, or both.

In a first aspect, a method harvests passive radio frequency energy. The method includes harvesting radio frequency energy using a crystal receiver. The method also includes powering a processing unit directly from the radio frequency energy being harvested when the radio frequency energy being harvested or an associated signal is approximately equal to or above a threshold level, and alternatively powering the processing unit from an on-board power source when the radio frequency energy being harvested or the associated signal is below the threshold level.

In a second aspect, a method harvests passive radio frequency energy. The method includes harvesting radio frequency energy via a crystal receiver and storing the radio frequency energy harvested on a super capacitor for subsequent use.

In a third aspect, a system harvests passive radio frequency energy. The system includes a network of radios within a building operable to direct the operation of building equipment to control a building environment of a building and at least one radio within the network being a sensor node operable to harvest radio frequency energy via a crystal receiver, the radio frequency energy harvested being used to power the sensor node.

In a fourth aspect, a computer-readable medium having instructions executable on a computer stored thereon is described. The instructions include monitoring for an interrupt signal that is triggered when the radio frequency energy being harvested via a crystal receiver falls below certain threshold, the radio frequency energy being used to at least partially power a sensor node or charge the battery on the harvesting system. The instructions also include automatically reducing the amount of power being expended by the sensor node if the interrupt signal is detected.

The present invention is defined by the claims herein. Nothing in this section should be taken as a limitation on those claims. Further aspects and advantages of the invention are discussed below in conjunction with the preferred embodiments and may be later claimed independently or in combination.

BRIEF DESCRIPTION OF THE DRAWINGS

The components in the figures are not necessarily to scale, emphasis instead being placed upon illustrating the principles of the invention.

FIGS. **1** and **1***a* are system block diagrams of an exemplary passive RF energy harvesting systems;

FIGS. **2** and **2***a* are schematics of additional exemplary passive RF energy harvesting systems;

FIG. **3** is a schematic of an exemplary network of sensor nodes;

FIG. **4** is a block diagram of an exemplary sensor node;

FIG. **5** is a top plan view of an exemplary network of sensor nodes within a building;

FIG. **6** illustrates an exemplary method of harvesting radio frequency energy;

FIG. **7** illustrates an exemplary interrupt handling method; and

FIGS. **8** and **9** illustrate exemplary antennas.

DETAILED DESCRIPTION OF THE DRAWINGS AND PRESENTLY PREFERRED EMBODIMENTS

The present embodiments are directed toward methods, processes, apparatuses, instructions, or systems for harvest-

ing freely available radio frequency (RF) energy from the environment, such as signals generated from amplitude modulated (AM) and frequency modulated (FM) radio stations. The RF energy harvested is used to power a "sensor node." The sensor node may include a processor, a sensor, an actuator, a wireless radio, and/or other components. The sensor node may be part of a network of radios in direct or indirect communication with one another, such as a wireless sensor network (WSN).

The sensor node may have an associated crystal radio receiver. The crystal radio receiver may operate as an all pass filter to harvest energy from all radio frequency bands simultaneously or ubiquitously. As such, more of the available RF energy may be captured for the given electromagnetic signals being received by the sensor node at a specific time.

The energy harvesting system may have an associated super capacitor that temporarily stores the RF energy received or a related signal, such as a voltage, a current, or other electrical parameter or signal associated with the RF energy harvested. In one embodiment, if a voltage associated with the RF energy being harvested falls below a given threshold, an interrupt signal may be generated that directs the sensor node to switch to a "low-power" mode to conserve energy. In low-power mode, the sensor node may reduce the amount of processing that an on-board processor performs. Alternatively or additionally, in low-power mode, the operation of an on-board wireless radio or transmitter may be altered and/or limited. For example, the wake-up period of the sensor nodes may be increased for energy saving purposes. Other manners of reduced sensor node operation may be employed to conserve energy being expended by the sensor node.

I. Crystal Radio Receivers

Generally, the embodiments discussed herein provide energy harvesting or scavenging for sensor nodes and/or wireless sensor networks (WSNs). In particular, AM RF signals may be harvested via crystal radio receivers to provide the primary, or even sole, source of power for the sensor nodes. Other crystal receivers may be used.

Crystal radios were popular by the 1950's, but are now almost obsolete. One of the key features of crystal radios is that they do not require batteries. The earplugs and/or speakers of a crystal radio may be powered solely by amplitude modulated (AM) radio signals. The crystal radio may employ a piezo-electric crystal, dielectric disk, oscillating crystal, or another type of crystal. The crystal may oscillate or vibrate due to the radio frequency signals being received. Exemplary crystal radio receivers are disclosed by U.S. Pat. Nos. 2,554,270, 2,805,332, 3,327,222, and 3,996,522, which are all incorporated herein by reference in their entireties. Other crystal radio receivers may be used.

The crystal radio technology may be used in new applications due to the advancement of semi-conductors and reduced power requirements of sensor nodes. In other words, the available energy that may be harvested from radio frequency signals via a crystal radio receiver, may contain sufficient energy to power milliwatt level instruments and micro-devices. Additionally, the size of a crystal radio antenna may be configured to be relatively small, ranging from a few to tens of centimeters. An antenna of this size may be used with a number of today's wireless sensor node applications.

Harvesting small amounts of energy from radio signals has been studied before, although not in the context of powering nodes in a WSN by AM RF signals. Previous studies have concluded that RF energy is of little utility unless a receiver is very close to a transmitter, such as a cell phone station. An alternate approach would be to introduce an RF signal transmitter for the explicit purpose of transmitting RF energy to a receiver that then harvests the RF energy to perform certain tasks. In other embodiments, such an approach may not be practical in the context of WSNs.

The energy available from RF signals in the unlicensed band, for e.g., the 2.4 GHz spectrum, may be typically too limited to be of any practical use. Such RF power harvesting may be impractical due to (1) very limited power being available (for an antenna size of 100 cm^2, the received power may be less than 0.1 milliwatt), and (2) very limited range (not only may a transmitter be required, but the sensor also may have to be placed very close to the transmitter). On the other hand, the principles on which crystal radios operate may finally permit the successful harvesting of energy from freely available RF signals.

As AM radio signals may be strong enough to power an earplug or a speaker, the energy available from RF signals may be more than past estimates indicated. A typical sensor node may require about 20 milliamps in active mode and 5 microamps in sleep mode, and the operating voltage required may be about 3 volts. Thus, the required power for a typical sensor node is comparable to the power that may be harvested via RF signals. Additionally, low-duty low-energy sensors may not require a transmitter, further reducing the amount of energy needed to power the sensor node. Therefore, RF signals may be a viable "free" and persistent (available 24 hours a day) energy source for a considerable amount of a given radio broadcast receiving area.

II. Estimation of Received Radio Energy

Table 1 below illustrates an estimation of the available RF energy in Princeton, N.J. from AM and FM RF signals transmitted by area radio stations. Through this rough estimation analysis, the "free" RF energy in the Princeton area is shown to be within the 10 milliwatt level. The estimation is based upon the Friis propagation model,

$$P_r = \frac{p_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L},$$

TABLE 1

Radio Stations that can be Received in Princeton, N.J.

Call Sign	Frequency	Dist./Signal	City	Format	Estimated Received Energy (w)
WNJT	88.1 FM	7.1 mi.	Trenton, NJ	Public Radio	$2.8 * 10^{-6}$
WWFM	89.1 FM	6.7 mi.	Trenton, NJ	Classical	$3.1 * 10^{-6}$
WPST	94.5 RM	17.7 mi.	Trenton, NJ	Top-40	$0.40 * 10^{-6}$
WTHK	97.5 FM	12.5 mi.	Burlington, NJ	Classic Rock	$0.75 * 10^{-6}$
WMGQ	98.3 FM	10.7 mi.	New Brunswick, NJ	Adult Contemporary	$1.0 * 10^{-6}$

TABLE 1-continued

Radio Stations that can be Received in Princeton, N.J.

Call Sign	Frequency	Dist./Signal	City	Format	Estimated Received Energy (w)
WAWZ	99.1 FM	19.3 mi.	Zarephath, NJ	Christian Contemporary	$0.30 * 10^{-6}$
WKXW	101.5 FM	7.1 mi.	Trenton, NJ	News/Talk	$2.1 * 10^{-6}$
WPRB	103.3 FM	7.2 mi.	Princeton, NJ	College	$2.0 * 10^{-6}$
WWJZ	640 AM	24.5 mi.	Mount Holly, NJ	Children's	$9.1 * 10^{-3}$
WFAN	660 AM	55.0 mi.	New York, NY	Sports	$1.7 * 10^{-3}$
WOR	710 AM	40.4 mi.	New York, NY	Talk	$2.7 * 10^{-3}$
WABC	770 AM	46.1 mi.	New York, NY	News/Talk	$1.8 * 10^{-3}$
WCBS	880 AM	55.0 mi.	New York, NY	News/Talk	$0.95 * 10^{-3}$
WPHT	1210 AM	32.9 mi.	Philadelphia, PA	Talk	$1.4 * 10^{-3}$
WHWH	1350 AM	9.5 mi.	Princeton, NJ	Business News	$13.5 * 10^{-3}$

Here, p_t, p_r are the power of the transmitter and the receiver, respectively; G_t, G_r are the antenna gains of the transmitter and the receiver, respectively; λ is the wavelength of the RF signal in meters; d is the distance between the transmitter and the receiver; and L is the system losses (L>1). The estimation is based upon various assumptions, such as (1) radio station information (frequency and distance), (2) the transmitting power of AM radio stations being 10 kilowatt and FM stations being 5 kilowatt, (3) the antenna gains, G, are 1 (for a properly designed antenna, G>1), and (4) the system losses, L, are 1.

The results of the estimation indicate that (1) the total power received was 31.120 milliwatts, (2) the AM power received was 31.108 milliwatts, or 99.96% of the total power, and (3) the FM power received was 0.0126 milliwatt, or 0.04% of the total power. The estimation is based on ideal environments. In practice, L is more than 1 and the received power is less than the ideal case. Observing the Friis propagation model reveals that the effect of wavelength on the received power is more than that of the transmission power. Accordingly, signals with longer wavelength, such as AM signals, are much more suitable for RF energy harvesting.

A typical sensor node may require about 41 milliwatts in active mode and 1.5 microwatts in sleep mode. This power requirement is of the same order of magnitude as the power that is available from RF signals. With reasonable conditions, such as the sensor nodes being located near one or more AM radio stations and/or using relatively efficient antennas, the sensor nodes may be powered solely by the AM signal(s) and be kept running uninhibited all of the time. Alternatively, the sensor nodes may be powered primarily by the AM signal(s), if backed up by an on-board energy source.

III. Estimation of Available Energy from Other RF Signal Sources

As noted above, AM signals may include much more energy available for harvesting than FM signals. Energy from other RF signal sources also may be very small when compared with AM signals. For example, the RF signals from cell phones quickly attenuate as a function of distance. As a result, a cell phone signal may be used as a source of energy to power a sensor node only if the distance from the cell phone to the sensor node is about one meter or less.

Energy also may be received from IEEE 802.11.x protocol-based devices and television sets. As IEEE 802.11.x devices operate in the 2.4 GHz band, the energy available for harvesting may be even less than what is available from cell phones. Similarly, due to their high frequency, television signals may contribute very little towards harvestable energy. Therefore, AM signals may provide the best freely available RF source of power for wireless sensor nodes.

IV. Hybrid Scheme

A number of harvesting schemes may be utilized to capture energy from RF signals. As an example, harvesting energy from AM signals may not only be useful for indoor sensor nodes, but could also prove fruitful for outdoor sensor deployments.

In one aspect, the RF harvesting schemes discussed herein may combine two concepts. A sensor node may include crystal radio technology that is operable in conjunction with an energy storage unit, such as a battery and/or a capacitor. Once the output of the crystal radio receiver is connected to act as a source of energy and/or a "charger" to the energy storage unit, the system may be ready for use.

The sensor node may include a RF energy harvesting device that includes a so-called "energy harvesting antenna" operable to capture RF energy. The antenna may harvest AM, as well as FM, or RF signals. The antenna may be a spider antenna, braided antenna, omni-directional antenna, double-braided polyester rope, loop antenna, air coil loop antenna, sparrow antenna, Yagi-Uda antenna, or other type of antenna.

For example, FIG. **8** is an exemplary spider antenna. As shown, the antenna may have a number of legs of differing size. FIG. **9** is another exemplary antenna. Other shapes may be used, such as circular or disk shapes.

The antenna may be a laminated antenna having a number of wires etched onto a semi-conductor or other micro-device. The antenna wires may have either a flat or rounded shape, or be disk shaped. The antenna may have a number of spokes or legs, such as the spider antenna disclosed by U.S. Pat. No. 4,229,742, or the antenna may be a grid antenna similar to the one disclosed by U.S. Pat. No. 7,038,630, both of which are incorporated herein by reference in their entireties. As the antenna characteristics may be determined based upon the bandwidth and gain properties desired for a specific application, alternative antennas may be used.

In one embodiment, for both convenience and cost reduction purposes, a 1950's type of coil or wire antenna connected to the crystal radio receiver may be replaced by modern flexible printed circuit board. Printed circuit board similar to that used in conventional radio frequency identification (RFID) devices may be used. An exemplary RFID antenna is disclosed by U.S. Pat. No. 6,825,754, which is incorporated herein by reference in its entirety. Other types of printed circuit board material and shapes may be used as the antenna.

Once the RF energy is scavenged, the sensor node may regulate the energy within the crystal radio circuit and convert

the RF signals received to a DC signal, such as via a rectifier, without any frequency filtering. As an example, the RF signals harvested may not be filtered via low, high, or band pass filters. After the RF energy harvested passes through the rectifier, the energy may be stored on a super capacitor. However, since the voltage on the two sides of the super capacitor may not satisfy the requirements of the sensor node, a DC/DC converter and/or another capacitor may be employed (such as shown in FIG. **2**, discussed below).

Clean DC power may be available on both sides of the second capacitor. In addition, the DC/DC converter may sample the voltage on the super capacitor, and issue an interrupt to the sensor node once the voltage, which is an indication of the energy being stored and/or harvested, is lower than a certain threshold. The sensor node may then switch to the on-board power source of the harvesting system, such as a battery, a capacitor, or other energy storage device. If the energy being harvested provides a voltage level greater than that of an on-board energy storage device, the energy being harvested may power the sensor node directly, as well as provide a trickle charge to the on-board energy storage device. On the other hand, if the energy being harvested provides a voltage level lower than of the on-board energy storage device, the on-board energy storage device may provide power to the sensor node via a trickle discharge. The harvested RF energy is used to charge the battery under this case.

Therefore, the wireless and self-contained sensor nodes may employ a “hybrid” RF energy harvesting scheme. Each wireless and self-contained sensor node may be equipped with (1) a RE-based energy harvester and (2) an on-board power source. As long as there is enough available power from the RF energy harvester, the sensor node may operate using the RF power. Intermittently, however, the power from the harvester may not be at the threshold required to operate the sensor node. As a result, an interrupt may be generated that switches the sensor node to the on-board power source. The hybrid scheme ensures that the sensor node may be continuously operational while simultaneously greatly lengthening the life-time of the on-board power source.

There are some distinctions between RF energy harvesting as discussed herein and conventional harvesting schemes. With respect to solar cell techniques, the proposed RF harvesting schemes are not affected by the weather or time of day (darkness). The proposed RF harvesting schemes may provide a continuous power supply as most radio stations broadcast around the clock.

Additionally, the proposed RF harvesting schemes may provide more power than solar cells, such as with indoor environments, and at less cost, as typical solar cells may be more expensive than an RF antenna or coil. The costs associated with the proposed schemes may be further reduced if the antenna is implemented using technology similar to that of a RFID tag.

As compared to biomechanics-based harvesting schemes, the proposed RF harvesting schemes may be more suitable for sensor node applications. Conventional biomechanics harvesting methods tend to provide energy on an inconsistent basis and inconveniently toggle sensors periodically to recharge the power source. On the other hand, the proposed RF harvesting schemes may provide a consistent milliwatt-level power source. Toggling may be used with RF harvesting.

With respect to vibration or thermoelectric methods, it may be more difficult to find a vibration or heat source than it is to locate an AM radio station, especially in populated regions of the United States. When a free source is not available, the vibration or thermoelectric based schemes may require an artificial source. In many environments, such as an office or building, installation of such artificial energy sources may not be feasible.

In sum, new RF energy harvesting schemes are described that provide energy to self-contained sensor nodes. The schemes provide an improved crystal radio receiver and harvest AM RF signals as a source of power. The RF harvesting schemes are suitable for indoor applications that involve wireless sensor networks, such as building automation applications.

The RF harvesting schemes may be especially useful in areas having AM stations within about 20 kilometers to satisfy typical sensor node power requirements. However, the effective distance of the harvesting schemes may be extended. The antennas employed may be configured to harvest more energy and provide enhanced gain. Sensor nodes may be configured to be more energy efficient. For example, the sensor nodes may be programmed to operate more efficiently, such as spending more time in various reduced power modes of operation. Additionally, the sensor nodes equipped with RF energy harvesters may be supplemented with other types of energy harvesters, such as solar, piezo-electric, MEMS (micro-electro-mechanical systems), biomechanical, thermoelectric, and other harvesters. The RF harvesting schemes, except for requiring relatively nearby AM radio stations, may have no other substantial apparent limitations, such as the weather, periodic toggling, or other limits.

V. Exemplary RF Energy Harvesting Systems

FIG. **1** illustrates an exemplary passive RF energy harvesting system **100**. The system **100** may include an energy harvesting antenna **102**, a crystal radio receiver **104**, a super capacitor **106**, a DC/DC converter **108**, a sensor node **110**, a local power source **112**, and/or a communication antenna **114**. The system may include additional, fewer, or alternate components.

The crystal radio receiver **104** or crystal set may employ an antenna **102**. The crystal radio receiver **104** may not require a battery or power source except the power received from radio waves via the antenna **102**. The crystal radio receiver **104** may include an antenna wire, a tuning coil of copper wire, a rectifier, and other components. The crystal radio receiver **104** may be a passive radio receiver, not requiring an actively powered amplifier associated with conventional radios. The crystal radio receiver **104** may preserve electrical power from RF signals received via the antenna **102** and convert the RF signals to electronic power. The crystal radio receiver **104** may include various types of antennas and various “low-tech” or traditional electrical circuitry components, such as capacitors, resistors, inductors, and diodes. As such the crystal radio receiver **104** may be simpler and less expensive to manufacture than current “smart” radios, such as cellular phones.

The antenna **102** may be a conventional crystal radio antenna and include copper or silver wire coils, or a ferrite coil. The antenna **102** may be a Yagi, log-periodic, fiberglass, cable, repeater, or other type of antenna, including those discussed elsewhere herein. The antenna **102** may be manufactured from printed circuit board, aluminum, copper, silver, other metals, or other materials. The antenna may omni-directionally collect all of the available AM radio signals currently being received. The shape, size, circumference, and composition of the antenna **102** may be altered for a specific application and/or to enhance the gain associated with the antenna **102**.

The antenna **102** and crystal radio receiver **104** may act independently or in conjunction to harvest RF energy ubiquitously. The crystal radio receiver **104** may require no tuning and resonant with all of the available frequencies. As a result,

energy may be collected simultaneously from a number of radio stations, which may increase the average level of RF energy being harvested. This may be in contrast with typical receivers that may employ RC resonant circuits, which eliminate some RF signals from being harvested. In one embodiment, the antenna may have a physical length of between 10 and 20 cm.

The low power mode may include a number of algorithms and/or software instructions stored in the sensor node that direct the sensor node or an associated processor to conserve or reduce the amount of energy that the sensor node is using. A processing unit on the sensor node may perform reduced or limited processing in low power mode. The radio may change the message routing being employed based upon the amount of RF energy being sensed or harvested. A transmitter may transmit less often in low power mode, or not at all until the level of the RF energy being harvested returns to above a threshold or self-sustaining level. The transmitter may alter the type of signal being transmitted in low power mode. For example, the sensor node may not relay all the received communication packets, and/or the transmitter on the sensor node may transmit at a reduced power level. The transmission at a reduced power level may be intended to only reach the next or closest available sensor node or radio within the network.

The low power mode may include buffering or storing data, instead of constant or even intermittent transmission of information to other sensor nodes. With typical sensor nodes, the operation of the radio and/or transmitting information may require a substantial amount of energy. Therefore, reducing radio transmissions in low power mode may enhance the effectiveness of the RF harvesting schemes discussed herein.

The super capacitor **106** may store the RF energy being harvested. The capacitance of an electrode associated with the super capacitor **106** may be several hundred times the capacitance of a conventional double-layer capacitor. The super capacitor **106** may have one or more electrodes manufactured from titanium, palladium, aluminum, silicon, or other metals. Alternatively, the electrodes may be doped or have titanium, palladium, aluminum, silicon, or other metal coatings. The super capacitor **106** may be similar to those disclosed by U.S. Pat. Nos. 5,585,999 and 5,151,848, which are incorporated herein by reference in their entireties. Alternate types of super or other capacitors may be used.

The DC-to-DC converter **108** may operate to produce an output voltage signal and/or an interrupt signal **116** based upon a signal received from the super capacitor **106** or an electrical parameter of the super capacitor **106** monitored, such as a voltage, current, or other signal. The DC-to-DC converter **108** may be similar to the DC-to-DC converter disclosed by U.S. Pat. No. 6,417,653, which is incorporated herein by reference in its entirety. Alternate DC-to-DC converters may be used.

In one embodiment, the DC-to-DC converter **108** samples the voltage **120** on the super capacitor **106**. The voltage sampled may be a reflection the current level of RF energy being harvested crystal radio receiver **104** and/or stored by the super capacitor **106**. If the voltage sampled falls below a threshold level, such as a level required by the sensor node **110** to operate normally (i.e., with no power operating restrictions), then the DC-to-DC converter may generate a low-power interrupt **116** signal. The sensor node **110** may switch to a low or reduced power operating mode upon the receipt of the low-power interrupt **116**.

The on-board energy storage unit may be a capacitor or battery **112**. The on-board energy storage unit **112** may be on a trickle charge when the voltage sampled **120** from the super capacitor **106** remains above the threshold level (such as the actual battery or capacitor voltage) with the carryover RF energy harvested directly powering the sensor node **110**. Alternatively, the RF energy harvested may be sent directly from the DC-to-DC converter **108** to the sensor node **110** when the level of RF energy being harvested remains at or above the threshold level.

On the other hand, if the low-power interrupt **116** is sent to the sensor node **110**, the sensor node **110** may select **118** or switch to the on-board storage unit **112** as a direct source of power. Alternatively, when the voltage from the DC-to-DC converter falls below the threshold level (such as the actual capacitor or battery voltage), the sensor node **110** may automatically start to be powered by the on-board power storage unit **112** as the storage unit voltage would be higher than the output voltage of the DC-to-DC converter, as can be envisioned by FIG. **1**. Other manners of powering of the sensor node from the super capacitor, either solely or in conjunction with the storage unit, may be used.

The sensor node **110** may be mounted on a device that includes the other components shown in FIG. **1** or be associated with the other components in another fashion. The sensor node **110** may include a communication antenna **114** that provides for wireless communication with other sensor nodes. The sensor node **110** may include the components shown in FIG. **4** and discussed below, in addition to the components illustrated by FIG. **1**. The sensor node may include additional, fewer, or alternate components.

FIG. **1***a* illustrates another exemplary passive RF energy harvesting system **100**. In one embodiment, the system may be targeted to facilitate the powering of sensor nodes in an ad-hoc wireless sensor network. As each sensor node may have communication functionality through an on-board RF module, a message communication subsystem may not be required to be a part of the RF energy harvesting system **100**. Also, the RF energy harvesting system **100** may be purely passive, and thus, no registration with the FCC is required.

As shown in FIG. **1***a*, the system **100** may be connected to the sensor node and its communication antenna. The system **100** may be a band pass and/or all pass filter to match all AM frequencies in order to provide ubiquitous energy harvesting capability and also to harness the maximum available harvestable energy.

As shown in FIG. **1***a*, a super capacitor is introduced as an energy buffer to the rechargeable batteries. Super capacitors have large number of recharge cycles, but limited energy capacity when compared to rechargeable batteries. In one embodiment, since there is bound to be some variation of RF signal strength from time to time and place to place, the harvested RF energy may not be consistent. Thus, as can be seen from FIGS. **1** and **1***a*, harvested energy can be used to power the sensor node or charge the on-hoard power source (such as batteries) on each sensor node. Thus, the demands on recharge cycles of the batteries are relaxed and the battery lifetime may be improved.

FIG. **1***a* also illustrates the interrupt signal. If the harvested RF energy is not enough and falls below a certain threshold, the DC/DC converter (for example, a MAX1674/MAX1675) may trigger the sensor node via an interrupt signal as shown. The processor on the sensor node, upon receiving this interrupt, may measure the voltages of the super capacitor, V1 **122**, and that of the battery, V2 **124**. Depending upon what these measured values are, the sensor node may make intelligent decisions. In one embodiment, the sensor node controls a power switch **126** that is operable to switch between the super capacitor and the batteries as a source of power. An exemplary

flow chart of an interrupt handling process that may be implemented by software on the sensor node is illustrated by FIG. **7**.

FIG. **2** illustrates the circuit exemplary passive RF energy harvesting system **200**. The system **200** may include a crystal radio receiver **202**, an energy harvesting antenna **204**, a DC/DC converter **206**, a voltage output **208**, and an interrupt output **210**. The system may include additional, fewer, or alternate components.

As shown in FIG. **2**, the crystal radio receiver **202** may include a capacitor bank and a rectifier that converts the RF energy being received into a DC signal. The DC signal harvested may be stored upon the super capacitor C**9**. The DC/DC converter **206** may sense the voltage on the super capacitor C**9** and provide an output voltage VCC **210** to the sensor node, as well as the sensor interrupt **210** when the output voltage approximately reaches or falls below a threshold level. The sensor nodes may use RF energy harvesting systems having additional, fewer, or alternate components.

FIG. **2***a* illustrates another embodiment of the circuit of the exemplary passive RF energy harvesting system **200**. With respect to FIG. **1***a*, FIG. **2***a* illustrates in more detail from where V**1 212** and V**2 218** may be sensed. V**1 212** may be sensed from the super capacitor C**2**. V**2 218** may be sensed from the battery. FIG. **2***a* also shows the power switch **214**. The system may include additional, fewer, or alternate components.

VI. Exemplary Network

FIG. **3** illustrates an exemplary network **300** of wireless radios or sensor nodes **312**. The network **300** may utilize a dynamic routing algorithm that permits data transmitted to travel the shortest distance or link **314** between sensor nodes **312** to a destination, which decreases the required transmission time for a given message, as well as the required power level of that transmission. The destination may be another sensor node **312** or a control radio **316**. Each wireless and self-contained sensor node **312** and control radio **316** may have a dedicated processor, a receiver, and a transmitter. The network **300** may include additional, fewer, or alternate components.

Each sensor node **312** may communicate its associated routing information to every nearby or adjacent sensor node **312** or control radio **316**. After a sensor node **312** receives a data transmission, a processor of the sensor node **312** may determine what to do with that data, including whether to retransmit the data to an adjacent or nearby sensor node **312** or control radio **316**. The base station **316** may function as a network controller that directs the overall operation of the network **300**.

The network **300** may provide continuous communication with otherwise unavailable sensor nodes **312**. For instance, some sensor nodes **312** may become obstructed by obstacles, such as equipment, containers, furniture, or other items, or may fail. However, the network **300** may reconfigure itself around blocked paths by redirecting transmissions from one sensor node to the next until communication with a lost sensor node is re-established. The network **300** also may provide enhanced communication reliability between sensor nodes **312** as a single sensor node **312** may be in direct communication with a number of other sensor nodes **312**, as shown in FIG. **3**.

The network **300** may implement IEEE 802.15.4 protocols. Other protocol standards may be used. The network **300** may operate as a mesh network, as described in more detail below. Alternate control or routing algorithms/instructions may be used.

VII. Control of Building Equipment

One application of the sensor nodes described herein relates to a network that controls building equipment. The network may include multiple sensor nodes, each having a wireless radio, as well as one or more control radios or base stations that direct the network. Each sensor node may be a so-called “smart” sensor node that includes a receiver, a transmitter, a processor, memory, one or more sensors, an actuator, and/or other components. Each sensor node may transmit messages to a control radio acting as network controller. Alternatively, the network controller may be a dedicated processor. The network may have one or more network controllers and/or control radios. The term network herein may include the entire network, a sub-set of a network, a number of sensor nodes, one or more network controllers, one or more control radios, or a combination of sensor nodes with one or more network controllers or control radios.

A network controller may assimilate and analyze a number of messages received from a plurality of sensor nodes. In response to each of the messages received, the network controller may determine that a change in the currently operating building equipment, or the operating modes thereof, is in order. Subsequently, the network controller may transmit a message to one or more sensor nodes that direct the operation of building equipment. Upon receiving the message, a sensor node may alter the operation of building equipment.

The sensors associated with the sensor nodes/wireless radios may monitor specific parameters pertaining to building environmental conditions or specific operating equipment. The actuators associated with the sensor nodes may control the operation of certain building equipment. A sensor node may transmit the value of a parameter sensed by a sensor to the network. In response to the values of the parameters received, the network or another sensor node may automatically alter the operation of building equipment, such as by sending messages that operate the actuators that control the building equipment.

For example, the sensors may be temperature sensors that sense the temperature in an area of a building. Each temperature sensor may be connected with a sensor node having a wireless radio, the sensor nodes being dispersed throughout a building. Each sensor node having a temperature sensor may transmit a message to the network regarding the temperature sensed in the building area in which the sensor node is located. In response to the temperature information received, the network may direct that cooling, heating, ventilation, HVAC, emergency, or other building equipment be operated to alter the building environment of the building area in which the sensor node is located.

The network may employ multiple sensor nodes in each building area to monitor temperature. Conventional wall mourned temperature sensors and/or thermostats may be single point sources of information. However, the average value of individual temperature parameters received from a plurality of temperature sensors dispersed in a given building area may better reflect the actual temperature in the building area. Accordingly, the building environmental equipment may be directed to maintain the temperature of a building area closer to the desired temperature based upon the more accurate temperature information received.

The sensors also may be motion sensors that sense motion in a building area. Each motion sensor may be connected with a sensor node having a wireless radio, the sensor nodes being dispersed throughout a building. Each sensor node having a motion sensor may transmit a message to the network regarding the motion sensed in a building area. In response to the motion information received, the network may direct the operation of building equipment.

The motion detected may alert the network that a building area has recently become occupied or unoccupied. In response, the network may ensure that lighting equipment provides adequate light in or near the building area in which motion was sensed. The network may direct that building environmental equipment, such as cooling, heating, ventilation, HVAC, or other equipment, be operated to alter the building environment of the building area. The motion information received also may be used by the network to determine that a security breach has occurred. Accordingly, the network may trigger an alarm, secure passageways, and operate other security equipment in response to the security breach.

A sensor node having a wireless radio may be connected with an identification device located on an individual. After the sensor node located on the identification device transmits a message to the network, the network may determine the identification and/or location of the associated individual. In response, the network may transmit instructions to building environmental equipment to automatically alter the environmental conditions of the specific building area in which the individual is currently located based upon stored or transmitted environmental preferences associated with that individual.

In a building having numerous pieces of operating equipment, it may be desirable to automatically monitor various parameters associated with various pieces of equipment. For instance, in a power plant, refinery, factory, or other plant, it may be advantageous to monitor temperatures, pressures, alarms, tank levels, bilge levels, hydraulic levels, atmospheric conditions, operating pumps or fans, and other parameters. The change in various temperatures, pressures, levels, or equipment operating temperatures may indicate problematic conditions.

The network may automatically identify problematic conditions associated with operating building equipment. The various parameters monitored each may be sensed by a sensor located on a sensor node. The sensor node having a wireless radio may transmit the value of the parameter to the network, either periodically or upon being queried by the network or sensing an out of specification value. The sensor node may determine whether a parameter is within specification, i.e., a predetermined satisfactory range.

If a parameter is not within specification, the network may take corrective action to restore the parameter and/or building conditions to specification. For example, the running speed of a problematic piece of equipment may be shifted, increased, or decreased. The problematic piece of equipment also may be secured and an alternate piece of equipment may be started or placed on line to replace it. Additional, fewer, or alternate courses of action may be taken to correct problematic or out of specification parameters. Sensor nodes may be connected with or part of the equipment being controlled for wireless control of the equipment.

VIII. Mesh Network

In one embodiment, the network may include a number of sensor nodes, each having a wireless radio, arranged as a mesh network that also may be used to locate movable assets and/or operate building environmental equipment. The mesh network provides the capability of routing data and instructions between and among the network of radios. The mesh network permits data to be efficiently transmitted from one radio in the network to the next until the data reaches a desired destination.

The mesh network may be implemented over a wireless network or partially wireless network. Each sensor node having a radio within the network may function as a repeater that transmits data received from adjacent sensor nodes having radios to other nearby sensor nodes that are within range. The coverage area of the mesh network may be increased by adding additional radios. As a result, a network may be established that may cover an area of desired size, such as a floor of a building or an entire building.

Each sensor node having a radio within the mesh network is typically only required to transmit data as far as the next sensor node having a radio within the network. Hence, if a sensor node having a wireless radio has a limited power supply, the reduction in the distance that each sensor node is required to transmit permits lower power level transmissions, which may extend the operating life of the on-board power supply.

A number of protocols may be used to implement the mesh network. The sensor nodes may implement a protocol that uses low data rates and low power consumption. As noted above, the mesh network may employ devices that use very small amounts of power to facilitate significantly increased battery or power supply life. In some situations, power supply life may be extended by minimizing the time that the radio device of the sensor node is "awake" or in normal power using mode, as well as reducing the power at which a signal is transmitted.

Alternatively, the sensor nodes having wireless radios may implement a protocol that uses moderate or high data rates and power consumption. For instance, the sensor nodes may implement IEEE 802.11 protocols. An IEEE 802.11 LAN may be based on a cellular architecture where the system is subdivided into cells, where each cell is controlled by a base station. Other protocols may be implemented.

Additionally, by reducing the distance between sensor nodes, each sensor node having a wireless radio may be able to transmit signals at a reduced power level, which may extend the life of a power supply while the signals transmitted remain strong enough to reach an adjacent sensor node having a radio. The sensor nodes within the network may be synchronized such that each sensor node talks or listens at a particular time. Alternatively, one or more control radios may be generally active, while the remaining radios remain predominantly passive. The control radios may be hardwired directly to a power supply such that they are not confined by a limited power supply.

The mesh network may utilize the Zigbee protocol or other IEEE 802.15.4 Low-Rate Wireless Personal Area Network (WPAN) standards for wireless personal area networking. Zigbee is a published specification set of high level communication protocols designed for use with small, low power digital radios based upon the IEEE 802.15.4 standard. Other IEEE 802.15 standards also may be implemented, including those using Bluetooth or other WPAN or WLAN protocols or any other protocol.

The mesh network of sensor nodes may employ a dynamic routing algorithm. As a result, the mesh network may be self configuring and self mending. Each sensor node within the network may be able to identify neighboring sensor nodes. After receiving a message, a receiving sensor node may determine that it is not the sensor node closest to the destination and/or that it should not relay the message to another sensor node based upon the currently known configuration of operating sensor nodes. The receiving sensor node may wait a predetermined period and listen for another sensor node with a radio to relay the message. If after a predetermined time, the sensor node determines that the message has not been relayed as expected, the receiving sensor node may transmit or relay the message to a nearby sensor node.

By transmitting messages to only reach nearby or adjacent sensor nodes in the network, the messages within the network

may be transmitted at lower power. The low power transmission requires less energy from the on-board power supply of each sensor node. Additionally, the low power transmissions by the sensor nodes prevent one message from occupying the entire network and permits messages to be simultaneously transmitted from different sensor nodes and travel throughout the network of sensor nodes in parallel.

IX. Exemplary Sensor Nodes

FIG. **4** illustrates an exemplary sensor node or wireless radio **410**. The sensor node **410** includes a processor **412**, a wireless radio frequency transmitter and/or receiver **414**, a sensor **416**, an actuator **418**, a memory **420**, a clock **422**, a speaker **424**, a microphone **426**, and an on-board power supply **428**. The sensor node **410** may include additional, different, or fewer components.

For example, the sensor node **410** may be free of the sensor **416**, actuator **418**, memory **420**, clock **422**, speaker **424**, the microphone **426**, and/or power supply **428**. The sensor node **410** may include the processor **412** and the wireless transmitter and/or receiver **414**. For instance, the sensor node **410** may comprise the processor **412**, the wireless radio frequency transmitter and/or receiver **414**, the sensor **416**, the actuator **418**, and the power supply **428**. Or the sensor node **410** may comprise the processor **412**, the wireless radio frequency transmitter and/or receiver **414**, the sensor **416**, and the power supply **428**. The sensor node **410** may include other combinations employing additional, different, or fewer components.

In one embodiment, the sensor node **410** may be operable to automatically control building equipment and locate movable items within a building. In another embodiment, the sensor node **410** may be portable, such as in the case of being mounted upon a movable item, or affixed at a specific location or to an immovable item. The sensor node **410** may be a controller, actuator, sensor, locator or other device in a security, fire, environment control, HVAC, lighting, or other building automation system. The sensor node **410** may determine its present location, sense conditions within a building, report conditions within a building, generate a signal representative of a building condition, and/or respond to an interrogator. The sensor node **410** also or alternatively may actuate building control components. As a controller, the sensor node **410** may be free of the sensor **416** and/or the actuator **418**. In another embodiment, the sensor node **410** is a wireless device free of wired connections to other devices making the sensor node **410** self-contained and/or portable.

The sensor **416** may be a single sensor or include multiple sensors. The sensor **416** may be a temperature, pressure, humidity, fire, smoke, occupancy, air quality, flow, velocity, vibration, rotation, enthalpy, power, voltage, current, light, gas, CO_2, CO, N_2, O_2, chemical, radiation, fluid level, tank level, motion, Global Positioning System (GPS), infrared, or other sensor or combination thereof. The sensor **416** also may be a limit or proximity switch. Alternate sensors may be used.

The sensor **416** may be a motion sensor that detects when a portable sensor node **410** is moving. If it is sensed that the sensor node **410** is moving, the processor **412** may wake the sensor node **410** up from a sleep mode that draws less energy from the power supply **428**. Upon waking up, the sensor node **410** may transmit via the wireless transmitter **414** to the network a message indicating that sensor node **410** is moving.

The sensor **416** may be motion sensor that detects when there is movement within a predetermined distance. For example, the sensor **416** may be wall mounted to detect when an individual has entered a specific building area. If the building area was previously unoccupied, the sensor node **410** on which the sensor **416** is mounted may transmit a message to the network that the building area is no longer unoccupied. As a result, the network may direct that the environmental conditions be altered accordingly, such as increase the temperature during cold weather, decrease the temperature during hot weather, turn on one or additional lights, or adjust the room to the individual's personal preferences.

The sensor **416** may be a GPS unit capable of receiving GPS signals and determining the location of the sensor node **410**. The GPS unit may be able to determine the latitudinal and longitudinal coordinates, as well as the elevation, of the sensor node **410**. The location of the sensor node **410** determined by the GPS unit may be subsequently transmitted to the network via the wireless transmitter **414**.

The actuator **418** may be a single actuator or include multiple actuators. The actuator **418** may be a valve, relay, solenoid, speaker, bell, switch, motor, motor starter, turbine generator, motor generator, diesel generator, pneumatic device, damper, or pump actuating device or combinations thereof. For example, the actuator **412** may be a valve for controlling flow of fluid, gas, or steam in a pipe, or a damper controlling or redirecting air within an air duct. As another example, the actuator **412** may be a relay or other electrical control for opening and closing doors, releasing locks, actuating lights, or starting, stopping, and shifting motors and pumps. As a further example, the actuator **412** may be a solenoid that opens or closes valves, dampers, or doors, such as for altering the flow of fluid or air within piping or ducting. Alternate actuating devices also may be used.

The sensor node **410** may function as a controller or base station. The controller may be positioned at either a known or an unknown location. As a controller, the sensor node **410** interacts with other sensor nodes **410** for control or reporting functions.

The processor **412** is capable of processing data and/or controlling operation of the sensor node **410**. The processor **412** may be a general processor, digital signal processor, application-specific integrated circuit (ASIC), field programmable gate array, analog circuit, digital circuit, network of processors, programmable logic controller, or other processing device. The processor **412** may have an internal memory.

The sensor node **410** also may have a memory unit **420** external to the processor **412**. The memory unit **420** may store data and instructions for the operation and control of the sensor node **410**, such as discussed herein. Additional or alternate types of data also may be stored in the memory unit **420**.

A program may reside on the internal memory or the memory unit **420** and include one or more sequences of executable code or coded instructions that are executed by the processor **412**. The instructions may include monitoring a signal associated with radio frequency energy being harvested via a crystal radio receiver, the radio frequency energy being used to at least partially power a sensor node. The instructions also may include automatically reducing the amount of power being expended by the sensor node if the magnitude of the signal monitored approximately reaches or falls below a threshold level. The instructions may include directions to accomplish other functionality described herein.

The program may be loaded into the internal memory or memory unit **420** from a storage device. The processor **412** may execute one or more sequences of instructions of the program to process data. Data may be input to the data processor **412** with a data input device and/or received from a network. The program and other data may be stored on or read from machine-readable medium, including secondary storage devices such as hard disks, floppy disks, CD-ROMS, and

DVDs, or other forms of machine readable medium, either currently known or later developed.

The processor **412** is capable of directing the transmission or reception of data by the wireless transmitter or receiver **414**, the speaker **424** or the microphone **426**. For example, the processor **412** may direct the acoustic speaker **424** to transmit an ultrasound signal. The processor **412** may also direct the microphone **426** to receive an ultrasound signal and determine a distance from another device as a function of the received signal. Alternatively or additionally, the processor **412** may direct the wireless transmitter or receiver **414** to transmit data for determining the distance. Additionally or alternatively, the wireless transmitter **414** transmits a determined distance or distances as well as data regarding the processes and operation of the sensor **416** and/or the actuator **418**.

The wireless transmitter and receiver **414** or the speaker **424** may be alternate wireless transmitters capable of transmitting a signal for distance determination. Similarly, the wireless receiver **414** and microphone **426** may be alternative wireless receivers capable of transmitting a signal for distance determination.

FIG. **5** illustrates a floor layout for a network of sensor nodes having wireless radios **510** operating with one or more control radios **522** within a building **524**. The sensor nodes **510** may be dispersed throughout the building **524**. One or more of the sensor nodes **510** may be located in each room or other building area. Alternate dispersed arrangements of the sensor nodes **510** may be provided. While one control radio **522** is shown, a plurality of control radios **522** may be provided in other embodiments. Additional, different or fewer sensor nodes **510** and control radios **522** may be provided. While shown as a single floor of a building **524**, the network of sensor nodes **510** and control radios **522** may be distributed over multiple floors, a portion of the floor, a single room, a house, a structure, or any other building **524** or portion thereof.

The various sensor nodes **510** may be of the same configuration or a different configuration than each other. For example, some of the sensor nodes **510** may correspond to sensor arrangements, while other sensor nodes **510** may correspond to actuator arrangements. The same or different communication device, such as a wireless radio frequency transmitter and/or receiver, may be provided for each of the sensor nodes **510**. Alternatively, different communications mechanisms and/or protocols are provided for different groups of the sensor nodes **510**. The sensor nodes **510** may operate in an integrated manner for implementing one or multiple types of building automation control. Alternatively, different networks may be provided for different types of building automation, such as security, HVAC, heating, ventilation, and fire systems.

The sensor nodes powered by RF energy as discussed herein may be useful for a wide variety of applications, in addition to the building control functionality described above. For instance, the applications may include environmental and human body monitoring to military surveillance, building automation, and industrial monitoring. Each sensor node may facilitate one or more control, homeland security, reconnaissance, or other tasks. In one embodiment, the sensor node may be implanted internally within a patient, and sense and transmit data pertaining to a biological characteristic of the patient. The sensor nodes may be used for other applications.

X. Exemplary Method

FIG. **6** illustrates an exemplary RF energy harvesting method. The method may include harvesting RF energy **602**, storing the RF energy harvested **604**, monitoring a signal associated with the RF energy harvested **606**, and selecting an operating mode based upon the level of the signal monitored **608**. The method may include additional, fewer, or alternate actions.

The method may include harvesting RF energy **602**. RF energy may be harvested via a crystal radio receiver as discussed herein. Most of the energy harvested may be scavenged from AM radio signals. The RF energy may be harvested ubiquitously. The RF energy may be used to primarily or solely power a self-contained sensor node.

The method may include storing the RF energy harvested **604**. The RF energy or an associated signal may be measured from a super capacitor or other energy storage means. In one aspect, the RF energy harvestor may be used in unison with another type of energy harvester, such as a solar, biomechanical, thermoelectric, or vibration driven device.

The method may include monitoring a signal associated with the RF energy **606**. A voltage on the super capacitor may be monitored. Other signals related to the RF energy harvested also may be monitored.

The method may include selecting an operating mode based upon the level of the signal monitored **608**. The operating mode selected may relate to power constraints imposed upon the sensor node. If the level of the signal being monitored, such as a voltage signal or other signal, falls below a threshold level, the sensor node may be directed to operate in a low power mode, as discussed herein, to conserve energy until the amount of RF energy being harvested returns to a self-sustaining level with respect to the operating power requirements of the sensor node.

XI. Exemplary Method

FIG. **7** illustrates an exemplary interrupt handling method **700**. The method **700** may be implemented by software on the sensor node. As shown, the method **700** may use a state variable K having 2 states. When K=1, the power switch in FIG. **2***a* may connect pin A to pin Out, which in turn is connected to the pin labeled VCC. The voltage of pin VCC, which is connected to the sensor node, is equal to the output voltage of the DC/DC converter. Thus, in this case, the sensor node is powered through the energy harvester via the super capacitor and DC/DC converter. On the other hand, when K=2, the power switch connects pin B to pin Out and in this case, the sensor node is powered by the battery.

Ts in FIG. **7** is the wake-up period (which is configurable through software depending upon the application). The sensor node wakes up repetitively with time period of Ts. Each time when the sensor node wakes up, it may sense the environment, process the data, and transmit the date to the sink, if necessary. Then the sensor node may enter sleep mode, which costs almost no energy.

Tm in FIG. **7** is the maximum wake-up period allowed by the system. If Ts>Tm, the wake-up period is longer than the upper limit required by the system. Thus, the sensor node cannot provide spontaneous samples to satisfy the quality of service (QoS) requirements of the network.

Tp in FIG. **7** denotes a continuous block of time during which harvested energy is enough to support the continuous operation of the sensor node to execute for a time Te, which is the minimum time required by a sensor node to complete sensing tasks as required by an application. In other words, each time the sensor node wakes up, it may have to execute for at least a period of Te. And the sensor node may have to make a decision as to whether it powers itself from the energy harvester or through the battery for this amount of time. The

following brief analysis is presented to describe the novel sensor node operation that may be facilitated through the energy harvesting circuit.

Let the capacity of the super capacitor be represented by C. Then, the instantaneous energy E stored in the super capacitor is given by:

$$E=½CV_1^2.$$

If power rating of the sensor node is P, then the following equation holds

$$E+D'T_p=PT_e,$$

where E' is the derivative of E with respect to time Thus,

$$\begin{aligned} T_p &= \frac{PT_e - E}{E'} \\ &= \frac{PT_e - \frac{1}{2}CV_1^2}{CV_1V_1'}, \end{aligned}$$

where V_1' is the derivative of V**1** and V_1' can be estimated in practice by taking the difference of the V**1** values at consecutive time intervals.

The flow chart shown in FIG. **7** may be described as follows. After receiving the interrupt **702**, the sensor node sends a control signal to the power switch such that the sensor node is powered through the battery **704**, since the energy on the capacitor may be depleted soon. The voltages V**1** and V**2** are measured by the sensor node **706**. Tp is also computed based on V**1** **708**, as shown in the above equation.

V**2** is checked to see if it is low **710**. If it is, then all of the energy is going to be depleted soon. Recall that when the interrupt occurred, V**1** is already low. Thus, in this case, neither the energy harvester nor the battery can power the sensor node. The sensor node must take emergency actions, such as backup data **712** and notify its neighbors **714**. Finally, the sensor node sets Ts to be equal to Tp **716**, as well as sets K to 1 **718**. The rationale being that since the energy in the battery is going to be depleted soon, the node should wait in sleep mode **732** until enough energy has been harvested by the RF energy harvester.

When V**2** is not low, the decision is based on V**1**. If V**1** is low such that Tm≤Tp **720**, which means that the energy harvester cannot satisfy the sensor node energy requirements, then the sensor node once again has no choice but to use power from the battery source. Since there is enough battery energy (recall that V**2** is not low), the sensor node can be powered by the battery **728** through the current wake-up period Ts **726**. The neighbors also may be notified **730**. However, note that the energy from the harvester can still be used to charge the battery through the DC/DC converter. Thus, the system still benefits from the energy harvester even though the node is not powered directly by the energy harvester.

If V**1** is high enough to support the minimal energy requirements of the sensor during its wake-up period, i.e., Tm>Tp, then the sensor node can increase its wake-up period. In this case, Ts is set to be Tp **722** and the sensor node is powered by the energy harvester through the DC/DC converter **724**.

To summarize, the embodiments discussed herein facilitate the following complex sensor node behaviors: (1) switching the operating mode on the sensor node, such as standby or sleep modes; (2) sending a notification message to a sink via other sensor nodes; (3) saving important data to the flash memory on the sensor node, in case it is lost; (4) changing a sampling rate and transmission rate, and reducing the rates according to the voltages V**1** and V**2**, (5) adjusting an energy weighting factor in the network routing metric (if the voltages of the sensor node are not high enough, the use of this node as a router to route other wireless message packets is minimized or avoided), and (6) switching the input power source from the RF energy harvester to a battery.

The methods and systems discussed herein may use one, two, or more antennas. If two antennas are used, one antenna may be dedicated to the sensor node or communications unit, and the other antenna may be dedicated to the RF energy harvesting system, such as shown in the Figures. As both the sensor node and the energy harvester each may have their own dedicated antenna, switching a single antenna between the sensor node and the energy harvester may be alleviated.

The operational mode of the harvesting system and/or the sensor node may be switched by the sensor node itself. Additionally or alternatively, the RF energy harvester discussed herein may trigger the sensor node and provide voltage measurements. For energy conservation purposes, the harvesting system discussed herein may not switch operation modes based on. RF signals.

Therefore, the RF energy harvesters described herein may be directly connected to sensor nodes, purely passive and not needing to pass FCC requirements, and/or based on a bandpass or all pass filter for ubiquitous and wide spectrum RF energy harvesting. The systems discussed herein may facilitate advanced sensor behaviors. The system modes (active, standby and sleep) may be triggered by an interrupt signal and controlled by the sensor node, as compared to altering the operating mode according to received RF signals. The designs of the energy harvesting antenna may be different from known antennas. A super capacitor and/or a DC/DC converter may be employed by the energy harvester. Based on efficiency, size, and costs, the DC/DC converter may provide benefits as compared to transformers. A buffered energy storage system may be constructed by the super capacitor and one or more rechargeable batteries. The harvesting system may improve battery life time.

While the invention has been described above by reference to various embodiments, it should be understood that many changes and modifications can be made without departing from the scope of the invention. The description and illustrations are by way of example only. Many more embodiments and implementations are possible within the scope of this invention and will be apparent to those of ordinary skill in the art. The various embodiments are not limited to the described environments, and have a wide variety of applications including integrated building control systems, environmental control, security detection, communications, industrial control, power distribution, and hazard reporting.

It is intended in the appended claims to cover all such changes and modifications which fall within the true spirit and scope of the invention. Therefore, the invention is not limited to the specific details, representative embodiments, and illustrated examples in this description. Accordingly, the invention is not to be restricted except in light as necessitated by the accompanying claims and their equivalents.

We claim:

1. A method of passive radio frequency energy harvesting, the method comprising:

harvesting radio frequency energy using a crystal receiver;

powering a processing unit only from the radio frequency energy being harvested when the radio frequency energy being harvested, or an indication of the amount of radio frequency energy being harvested, is above a threshold level;

generating an interrupt signal to switch the processing unit to a low power mode when the radio frequency energy

being harvested, or the indication of the amount of radio frequency energy being harvested, falls below the threshold level; and

alternatively powering the processing unit only from an on-board power source when the radio frequency energy being harvested, or the indication of the amount of radio frequency energy being harvested, is below the threshold level.

2. The method of claim **1**, wherein the crystal receiver employs an antenna that is fabricated from printed circuit board.

3. The method of claim **1**, wherein the radio frequency energy harvested is increased by a super capacitor before being used to power the processing unit.

4. The method of claim **1**, wherein the on-board power source is a battery or a capacitor.

5. The method of claim **1**, wherein the crystal receiver employs a ferrite coil antenna.

6. The method of claim **1**, wherein the radio frequency energy is harvested ubiquitously.

7. A method of passive radio frequency energy harvesting, the method comprising:

harvesting radio frequency energy using a crystal receiver;

powering a processing unit only from the radio frequency energy being harvested when the radio frequency energy being harvested, or an indication of the amount of radio frequency energy being harvested, is above a threshold level, wherein the processor unit is a self-contained sensor node operable to direct the operation of building equipment to control an environment of a building;

generating an interrupt signal to switch the processing unit to a low power mode when the radio frequency energy being harvested, or the indication of the amount of radio frequency energy being harvested, falls below the threshold level; and

alternatively powering the processing unit only from an on-board power source when the radio frequency energy being harvested, or the indication of the amount of radio frequency energy being harvested, is below the threshold level.

8. The method of claim **7**, wherein the processor unit has power requirements of greater than 10 milliwatts.

* * * * *

US008922347B1

(12) **United States Patent**
de Rochemont

(10) **Patent No.: US 8,922,347 B1**
(45) **Date of Patent: Dec. 30, 2014**

(54) **R.F. ENERGY COLLECTION CIRCUIT FOR WIRELESS DEVICES**

(76) Inventor: **L. Pierre de Rochemont**, Austin, TX (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 233 days.

(21) Appl. No.: **12/818,025**

(22) Filed: **Jun. 17, 2010**

Related U.S. Application Data

(60) Provisional application No. 61/187,687, filed on Jun. 17, 2009.

(51) **Int. Cl.**
H04Q 5/22 (2006.01)
H01Q 9/26 (2006.01)

(52) **U.S. Cl.**
USPC **340/10.4**; 343/803

(58) **Field of Classification Search**
CPC G06K 19/0723; G06K 19/0707; G06K 19/0775; G06K 19/07767; H01L 2223/6677; H01L 21/4807; H02J 17/00
USPC 340/10.1–10.4; 343/803
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2,283,925	A	5/1942	Harvey
2,886,529	A	5/1959	Louis
3,574,114	A	4/1971	Monforte
3,614,554	A	10/1971	Shield et al.
3,983,077	A	9/1976	Fuller et al.
4,400,683	A	8/1983	Eda et al.
4,455,545	A	6/1984	Shelly
4,523,170	A	6/1985	Huth, III
4,646,038	A	2/1987	Wanat
4,759,120	A	7/1988	Bernstein
4,859,492	A	8/1989	Rogers, Jr. et al.
4,880,770	A	11/1989	Mir et al.
4,967,201	A	10/1990	Rich, III
5,084,749	A	1/1992	Losee et al.
5,130,675	A	7/1992	Sugawara
5,139,999	A	8/1992	Gordan et al.
5,154,973	A	10/1992	Imagawa et al.
5,198,824	A	3/1993	Poradish
5,217,754	A	6/1993	Santiago-Aviles et al.
5,219,377	A	6/1993	Poradish

(Continued)

FOREIGN PATENT DOCUMENTS

EP	0026056	A1	4/1981
EP	0939451	A1	1/1999
EP	1376759	A2	2/2004
GB	1125897	A	9/1968

OTHER PUBLICATIONS

GigaCircuits, "The Wireless Revolution Continues . . . ", Jul. 19, 2006, http://www.gigacircuits.com.*

(Continued)

Primary Examiner — Benjamin C Lee
Assistant Examiner — Adam Carlson
(74) *Attorney, Agent, or Firm* — Burns & Levinson LLP; Jacob N. Erlich; David W. Gomes

(57) **ABSTRACT**

An R.F. energy collection circuit, provides a low-loss R.F. front end tuned to minimize lost power in R.F. bands commonly used for communications for passing substantially coherent R.F. signals received therefrom, a rectifier circuit coupled to receive coherent R.F. energy from the front end, and an energy storage circuit coupled to receive energy from the rectifier circuit.

10 Claims, 6 Drawing Sheets

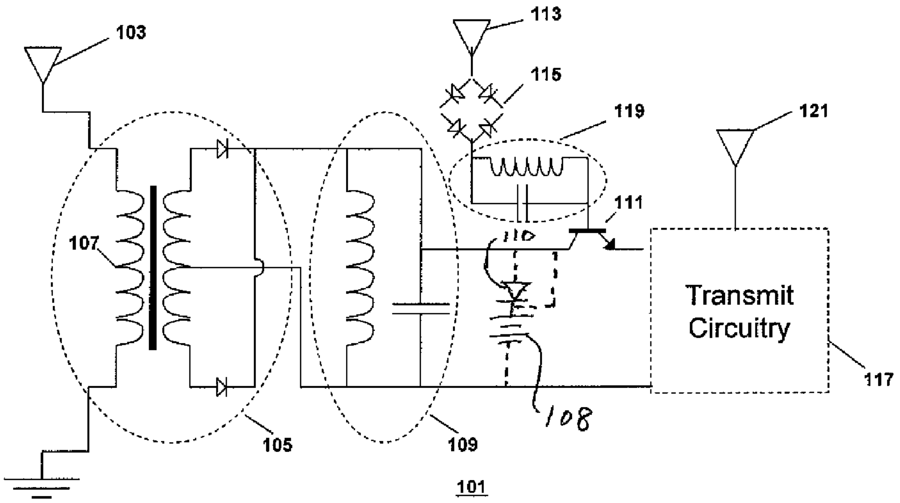

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,263,198	A	11/1993	Geddes et al.	
5,272,485	A	12/1993	Mason et al.	
5,403,797	A	4/1995	Ohtani et al.	
5,427,988	A	6/1995	Sengupta et al.	
5,456,945	A	10/1995	McMillan et al.	
5,478,610	A	12/1995	Desu et al.	
5,513,382	A	4/1996	Agahi-Kesheh et al.	
5,535,445	A	7/1996	Gunton	
5,540,772	A	7/1996	McMillan et al.	
5,543,773	A	8/1996	Evans et al.	
5,584,053	A	12/1996	Kommrusch et al.	
5,590,387	A	12/1996	Schmidt et al.	
5,614,252	A	3/1997	McMillan et al.	
5,625,365	A	4/1997	Tom et al.	
5,635,433	A	6/1997	Sengupta	
5,707,459	A	1/1998	Itoyama et al.	
5,707,715	A	1/1998	deRochemont et al.	
5,747,870	A	5/1998	Pedder	
5,759,923	A	6/1998	McMillan et al.	
5,764,189	A	6/1998	Lohninger	
5,771,567	A	6/1998	Pierce et al.	
5,854,608	A	12/1998	Leisten	
5,859,621	A	1/1999	Leisten	
5,888,583	A	3/1999	McMillan et al.	
5,889,459	A	3/1999	Hattori et al.	
5,892,489	A	4/1999	Kanba et al.	
5,903,421	A	5/1999	Furutani et al.	
5,933,121	A	8/1999	Rainhart et al.	
5,945,963	A	8/1999	Leisten	
6,023,251	A	2/2000	Koo et al.	
6,027,826	A	2/2000	deRochemont	
6,028,568	A	2/2000	Asakura et al.	
6,031,445	A	2/2000	Marty et al.	
6,040,805	A	3/2000	Huynh et al.	
6,046,707	A	4/2000	Gaughan et al.	
6,052,040	A	4/2000	Hino	
6,111,544	A	8/2000	Dakeya et al.	
6,143,432	A	11/2000	deRochemont et al.	
6,154,176	A	11/2000	Fathy et al.	
6,157,321	A *	12/2000	Ricci	340/902
6,176,004	B1	1/2001	Rainhart et al.	
6,181,297	B1	1/2001	Leisten	
6,188,368	B1	2/2001	Koriyama et al.	
6,195,049	B1	2/2001	Kim et al.	
6,204,203	B1	3/2001	Narwankar et al.	
6,208,843	B1	3/2001	Huang et al.	
6,222,489	B1	4/2001	Tsuru et al.	
6,266,020	B1	7/2001	Chang	
6,271,803	B1	8/2001	Watanabe et al.	
6,300,894	B1	10/2001	Lynch et al.	
6,320,547	B1	11/2001	Fathy et al.	
6,323,549	B1	11/2001	deRochemont et al.	
6,492,949	B1	12/2002	Breglia et al.	
6,496,149	B1	12/2002	Birnbaum et al.	
6,501,415	B1	12/2002	Viana et al.	
6,541,820	B1	4/2003	Bol	
6,552,693	B1	4/2003	Leisten	
6,559,735	B1	5/2003	Hoang et al.	
6,583,699	B2	6/2003	Yokoyama	
6,605,151	B1	8/2003	Wessels et al.	
6,611,419	B1	8/2003	Chakravorty	
6,620,750	B2	9/2003	Kim et al.	
6,639,556	B2	10/2003	Baba	
6,642,908	B2	11/2003	Pleva et al.	
6,650,303	B2	11/2003	Kim et al.	
6,670,497	B2	12/2003	Tashino et al.	
6,680,700	B2	1/2004	Hilgers	
6,683,576	B2	1/2004	Achim	
6,686,406	B2	2/2004	Tomomatsu et al.	
6,690,336	B1	2/2004	Leisten et al.	
6,697,605	B1	2/2004	Atokawa et al.	
6,742,249	B2	6/2004	deRochemont et al.	
6,743,744	B1	6/2004	Kim et al.	
6,762,237	B2	7/2004	Glatkowski et al.	
6,791,496	B1	9/2004	Killen et al.	
6,864,848	B2	3/2005	Sievenpiper	
6,906,674	B2	6/2005	McKinzie, III et al.	
6,919,119	B2	7/2005	Kalkan et al.	
7,047,637	B2	5/2006	deRochemont et al.	
7,405,698	B2	7/2008	deRochemont	
7,564,887	B2	7/2009	Wang et al.	
7,763,917	B2	7/2010	deRochemont	
7,840,305	B2 *	11/2010	Behr et al.	700/174
8,115,448	B2 *	2/2012	John	320/108
8,237,561	B2 *	8/2012	Beigel et al.	340/572.1
2001/0023779	A1	9/2001	Sugaya et al.	
2001/0027119	A1	10/2001	Furutani et al.	
2001/0048969	A1	12/2001	Constantino et al.	
2002/0039667	A1	4/2002	Takaya et al.	
2002/0047768	A1	4/2002	Duffy	
2002/0070983	A1	6/2002	Kozub et al.	
2002/0092472	A1	7/2002	Hayashi et al.	
2002/0110004	A1	8/2002	Parks	
2002/0190818	A1	12/2002	Endou et al.	
2003/0002045	A1	1/2003	Nemat-Nasser et al.	
2003/0034124	A1	2/2003	Sugaya et al.	
2003/0073565	A1	4/2003	Ellis et al.	
2003/0080325	A1	5/2003	Uchiyama et al.	
2003/0107455	A1	6/2003	Imanaka et al.	
2003/0111714	A1	6/2003	Bates et al.	
2003/0122647	A1	7/2003	Ou	
2003/0148024	A1	8/2003	Kodas et al.	
2003/0161959	A1	8/2003	Kodas et al.	
2003/0170436	A1	9/2003	Sumi et al.	
2003/0221621	A1	12/2003	Pokharna et al.	
2004/0000964	A1	1/2004	Killen et al.	
2004/0000966	A1	1/2004	Killen et al.	
2004/0000970	A1	1/2004	Killen et al.	
2004/0000972	A1	1/2004	Killen et al.	
2004/0000975	A1	1/2004	Killen et al.	
2004/0000976	A1	1/2004	Killen et al.	
2004/0001024	A1	1/2004	Killen et al.	
2004/0001026	A1	1/2004	Killen et al.	
2004/0001027	A1	1/2004	Killen et al.	
2004/0001028	A1	1/2004	Killen et al.	
2004/0012081	A1	1/2004	Kwon	
2004/0033654	A1	2/2004	Yamagata	
2004/0070915	A1	4/2004	Nagai et al.	
2004/0084080	A1	5/2004	Sager et al.	
2004/0092236	A1	5/2004	Irie et al.	
2004/0113738	A1	6/2004	Ahn et al.	
2004/0113790	A1 *	6/2004	Hamel et al.	340/572.1
2004/0118448	A1	6/2004	Scher et al.	
2004/0189528	A1	9/2004	Killen et al.	
2005/0036269	A1	2/2005	Ma et al.	
2005/0051870	A1	3/2005	Yamazaki et al.	
2005/0104553	A1 *	5/2005	Mickle et al.	320/101
2006/0086994	A1	4/2006	Viefers et al.	
2006/0092079	A1 *	5/2006	de Rochemont	343/700 MS
2006/0094425	A1 *	5/2006	Mickle et al.	455/434
2006/0125681	A1	6/2006	Smith et al.	
2006/0134491	A1	6/2006	Hilchenko et al.	
2007/0003781	A1	1/2007	deRochemont	
2007/0137257	A1	6/2007	Desantolo et al.	
2007/0139976	A1	6/2007	deRochemont	
2007/0166453	A1	7/2007	Van Duren et al.	
2007/0170535	A1	7/2007	deRochemont	
2007/0259768	A1	11/2007	Kear et al.	
2008/0136395	A1	6/2008	Bennett	
2008/0186245	A1 *	8/2008	Hilgers	343/803
2008/0231421	A1 *	9/2008	Tuttle	340/10.1
2009/0004370	A1	1/2009	Zurcher et al.	
2009/0011922	A1	1/2009	deRochemont	
2009/0015314	A1	1/2009	Kirchmeier et al.	
2011/0021007	A1	1/2011	deRochemont	
2011/0049394	A1	3/2011	deRochemont	
2011/0065224	A1	3/2011	Bollman et al.	
2011/0248900	A1	10/2011	deRochemont	

OTHER PUBLICATIONS

McKinzie III, W.E. et al. Mitigation of Multipath Through the Use of an Artificial Magnetic Conductor for Precision CPS Surveying Antennas. © 2002 IEEE pp. 640-643.

(56) **References Cited**

OTHER PUBLICATIONS

McKinzie, W. et al. A Multi-Band Artificial Magnetic Conductor Comprised of Multiple FSS Layers. © 2003 IEEE pp. 423-426.

Monorchio, A. et al. Synthesis of Artificial Magnetic Conductors by Using Multilayered Frequency Selective Surfaces. IEEE Antennas and Wireless Propagation Letters vol. 1 2002 pp. 196-199.

Mosallaei, H. et al. Periodic Bandgap and Effective Dielectric Materials in Electromagnetics: Characterization and Applications in Nanocavities and Waveguides. IEEE Transactions on Antennas and Propagation 51(3). Mar. 2003. pp. 549-563.

Pontes, F.M. et al. Study of the dielectric and ferroelectric properties of chemically processed BaxSr1-xTiO3 thin films. Thin Solid Films, vol. 386, No. 1 (2001) 91-98.

Rogers, S. et al. AMCs Comprised of Interdigital Capacitor FSS Layers Enable Lower Cost Applications. © 2003 IEEE pp. 411-414.

Rogers, S. et al. An AMC-Based 802.11a/b Antenna for Laptop Computers. © 2003 IEEE pp. 10-13.

Sievenpiper, D.F. et al. Two-Dimensional Beam Steering Using an Electrically Tunable Impedance Surface. IEEE Transactions on Antennas and Propagation 51(10). Oct. 2003. pp. 2713-2722.

Sun, J. et al. Efficiency of Various Photonic Bandgap (PBG) Structures. 3rd Int'l. Conf. on Microwave and Millimeter Wave Technology Proceedings 2002. pp. 1055-1058.

Tsunemine, Y. et al. Pt/BaxSr(1-x)TiO3/Pt Capacitor Technology for 0.15 micron Embedded Dynamic Random Access Memory. Jap. J. Appl. Phys. vol. 43 No. 5A (2004) 2457-2461.

Vest, R.W. Metallo-organic decomposition (MOD) processin of ferroelectric and electro-optic films: A review. Ferroelectrics, 102:1, 53-68.

Viviani, M. et al. Positive Temperature Coefficient of Electrical Resistivity below 150k of Barium Strontium Titanate. J. Amer. Ceram. Soc. vol. 87 (2004) 756-758.

Weily, A.R. et al. Antennas Based on 2-D and 3-D Electromagnetic Bandgap Materials. © 2003 IEEE pp. 847-850.

Yang, H-Y. D. et al. Surface Waves of Printed Antennas on Planar Artificial Periodic Dielectric Structures. IEEE Transactions on Antennas and Propagation 49(3). Mar. 2001. pp. 444-450.

Andrenko, A.S. et al. EM Analysis of PBG Substrate Microstrip Circuits for Integrated Transmitter Front End. MMET 2000 Proceedings pp. 295-297.

Bardi, I. et al. Plane Wave Scattering From Frequency-Selective Surfaces by the Finite-Element Method. IEEE Transactions on Magnetics 38(2) Mar. 2002. pp. 641-644.

Chappell, W. et al. Composite Metamaterial Systems for Two-Dimensional Periodic Structures. © 2002 IEEE pp. 384-387.

Cheng, Y.L. et al. Preparation and Characterization of (Ba.Sr)TiO3 thin films using interdigital electrodes. Microelectronic Engineering vol. 66 (2003) 872-879.

Clavijo, S. et al. Design Methodology for Sievenpiper High-Impedance Surfaces: An Artificial Magnetic Conductor for Positive Gain Electrically Small Antennas. IEEE Transactions on Antennas and Propagation 51(10). Oct. 2003. pp. 2678-2690.

Diaz, R. et al. Magnetic Loading of Artificial Magnetic Conductors for Bandwidth Enhancement. © 2003 IEEE pp. 431-434.

Hansen, R.C. Effects of a High-Impedance Screen on a Dipole Antenna. IEEE Antennas and Wireless Propagation Letters vol. 1 2002 pp. 46-49.

Joshi , P.C. et al. Processing and Characterization of Pure and Doped Ba0.6Sr 0.4TiO3 thin films for tunable microwave applications. Mat. Res. Soc. Symp. Proc. vol. 656E (2001) DD4.9.1-DD4.9.6.

Kern, D.J. et al. Active Negative Impedance Loaded EBG Structures for the Realization of Ultra-Wideband Artificial Magnetic Conductors. © 2003 IEEE pp. 427-430.

Kern, D.J. et al. The Synthesis of Metamaterial Ferrites for RF Applications Using Electromagnetic Bandgap Structures. © 2003 IEEE pp. 497-500.

Kern, D.J. et al. Ultra-thin Electromagnetic Bandgap Absorbers Synthesized via Genetic Algorithms. © 2003 IEEE pp. 1119-1122.

Khun, R. et al., Characterization of Novel Mono- and Bifacially Active Semi-Transparent Crystalline Silicon Solar Cells, IEEE Transactions on Electron Devices, 46(10), Oct. 1999, p. 2013-2017.

Kretly, L.C. et al. The Influence of the Height Variation on the Frequency Bandgap in an AMC Artificial Magnetic Conductor for Wireless Applications: an EM Experimental Design Approach. Proceedings SBMO/IEEE MTT-S IMOC 2003 pp. 219-223.

Lee, Y. et al. Investigation of Electromagnetic Bandgap (EBG) Structures for Antenna Pattern Control. © 2003 IEEE, pp. 1115-1118.

Zhang, Y. et al. Planar Artificial Magnetic Conductors and Patch Antennas. IEEE Transactions on Antennas and Propagation 51(10). Oct. 2003. pp. 2704-2712.

Ziroff, A. et al. A Novel Approach for LTCC Packaging Using a PBG Structure for Shielding and Package Mode Suppression. 33rd European Microwave Conference—Munich 2003 pp. 419-422.

* cited by examiner

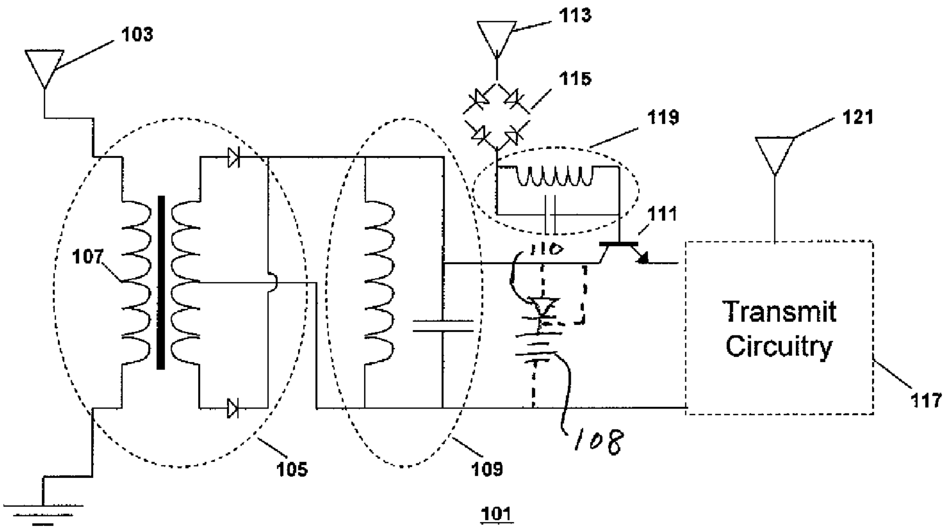
Transmit Circuitry

FIG. 1

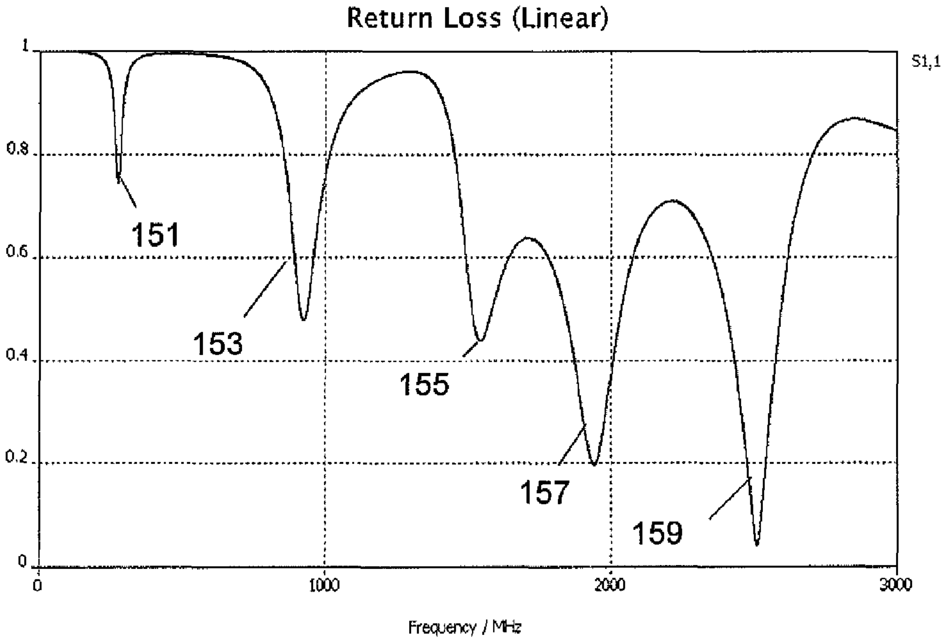
Return Loss (Linear)
1
0.8
0.6
0.4
0.2
0
0
1000
2000
3000
Frequency / MHz
S1,1
151
153
155
157
159

FIG. 2

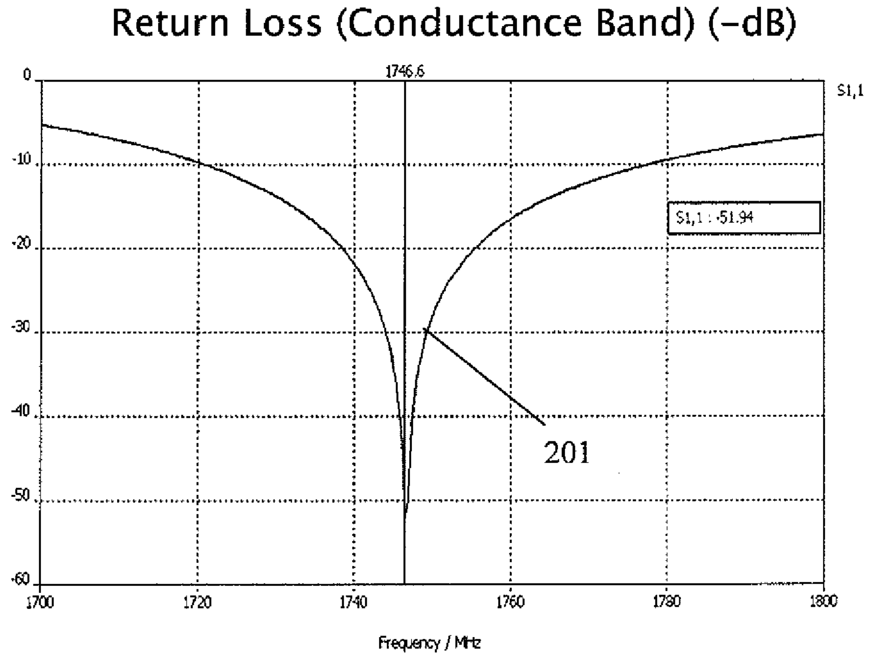
Return Loss (Conductance Band) (-dB)
1746.6
S1,1
S1,1 : -51.94
201
0
-10
-20
-30
-40
-50
-60
1700
1720
1740
1760
1780
1800
Frequency / MHz

FIG 3A.

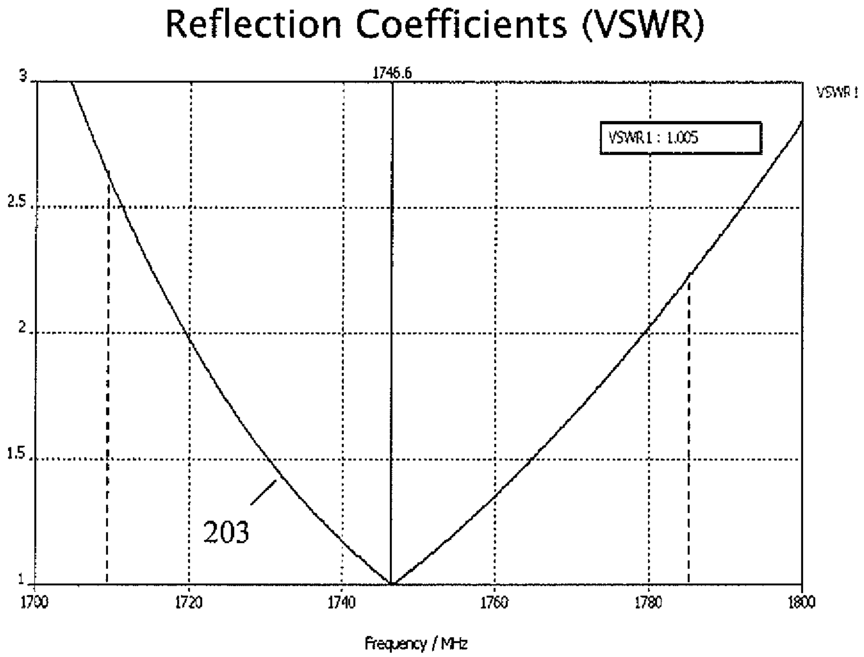
Reflection Coefficients (VSWR)
1746.6
VSWR1
VSWR1 : 1.005
3
2.5
2
1.5
1
203
1700
1720
1740
1760
1780
1800
Frequency / MHz

FIG. 3B.

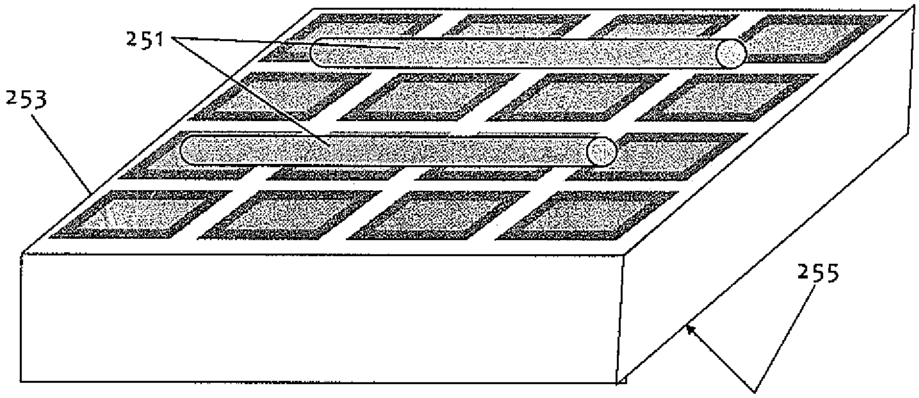

FIG. 4

	Conventional CDMA				Narrow Conductance Band Antenna			
Component	RF Input Power	Power Lost	DC Input Power	Wasted Power	RF Input Power	Power Lost	DC Input Power	Wasted Power
Secondary Band Filter	1 mW	1 mW		1 mW	1 mW	—		—
Power Amplifier (PA)	1 mW	-506 mW	1267 mW	761 mW	1 mW	-250 mW	629 mW	379 mW
PA/Duplexer matching	507 mW	23 mW		23 mW		—		—
SAW Duplexer	484 mW	212 mW		212 mW		—		—
Coupler	272 mW	6 mW		6 mW		—		—
Band Select Switch	266 mW	15 mW		15 mW		—		—
Power to Antenna	251 mW				251 mW	—		—
				1018 mW				379 mW

TABLE 1.

FIG. 5

R.F. ENERGY COLLECTION CIRCUIT FOR WIRELESS DEVICES

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims priority from U.S. Provisional Patent Application Ser. No. 61/187,687, filed Jun. 17, 2009, entitled EXTENDED RANGE PASSIVE RFID SYSTEM and incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

The present invention relates generally to the application of ideal materials sets and frequency-selective antennas to collect R.F. energy and to greatly extend the range of passive radio-frequency identification (RFID) systems.

BACKGROUND OF THE INVENTION

U.S. Pat. No. 6,027,826 to de Rochemont, et al. discloses articles and methods to form oxide ceramic on metal substrates to form laminate, filament and wire metal-ceramic composite structures using liquid aerosol spray techniques. U.S. Pat. Nos. 6,323,549 and 6,742,249 to de Rochemont, et al. disclose articles that comprise, and methods to construct, an interconnect structure that electrically contacts a semiconductor chip to a larger system using at least one discrete wire that is embedded in amorphous silica ceramic, as well as methods to embed passive components within said interconnect structure. U.S. Pat. Nos. 5,707,715 and 6,143,432 to de Rochemont, et al. (the '715 and '432 patents) disclose articles and methods to relieve thermally-induced mechanical stress in metal-ceramic circuit boards and metal-ceramic and ceramic-ceramic composite structures. U.S. Pat. No. 6,027,826 to de Rochemont, et al. disclose articles and methods to form oxide ceramic on metal substrates to form laminate, filament and wire metal-ceramic composite structures using liquid aerosol spray techniques. U.S. Pat. Nos. 6,323,549 and 6,742,249 to de Rochemont, et al. disclose articles that comprise, and methods to construct, an interconnect structure that electrically connects a semiconductor chip to a larger system using at least one discrete wire that is embedded in silica ceramic, as well as methods to embed passive components within said interconnect structure. U.S. patent application Ser. No. 11/479,159 (U.S. Pub. No. 2007/0003781) (hereinafter "de Rochemont '159") discloses articles and methods that laminate passive components (resistors, capacitors, inductors) on an exposed or a buried microelectronic surface using liquid aerosol sprays or inkjet deposition systems. U.S. patent application Ser. No. 11/660,042 (Pub. No. 2007/0139976) (hereinafter "de Rochemont '042) discloses articles and methods that laminate passive components on an exposed surface layer or a buried microelectronic layer to form a power management device using laminated resistors, capacitors, inductors or transformers. U.S. Pat. No. 7,405,698 is used to miniaturize artificial magnetic conducting ground planes and form meta-material dielectrics that, in combination with de Rochemont '159, enable electrically small antennas. The contents of each of these references are incorporated herein by reference as if laid out in their entirety.

RFID systems allow an object to be remotely identified by a radio frequency "interrogator" signal, which activates a "chirp" or "read" signal from the RFID tag. Passive RFID systems do not have any internal power sources (such as battery elements) within the tag to avoid maintenance needs and, consequently, will have read ranges—(the distance from which the tag can be remotely identified)—limited to 65 to 100 feet. This is the case because the interrogator signal must activate and power the "read signal" from the RFID transponder. Therefore, the interrogator signal must be powerful to simultaneously overcome signal path losses between the read head and the RFID tag, to activate the "read" signal, and still be detectable by the remote read head. State-of-the-art passive RFID systems that incorporate surface-acoustic wave systems will require power levels of −15 dB to −18 dB simply to activate the chirped read signal.

RFID systems are used to manage assets in inventory and during transport. Recently, piracy and theft during logistical transport has developed a need for passive RFID sensors that can not only report asset location, but additional information that might indicate whether or not the asset has been compromised or tampered to a surveillance vehicle located several miles away from the asset, such as an unmanned aerial vehicle.

Therefore, methods that enable passive RFID sensors to remotely report the location or tampering of an asset at distances up to and exceeding a distance of 10 miles is desirable. It is also desirable that such an RFID system have a planar structure so that it may be concealed in the wall of a container to protect it from being detected, damaged or destroyed prior to or during the time an asset is tampered.

SUMMARY OF THE INVENTION

One embodiment of the present invention provides an R.F. energy collection circuit, comprising a low-loss R.F. front end tuned to minimize lost power in R.F. bands commonly used for communications for passing substantially coherent R.F. signals received therefrom, a rectifier circuit coupled to receive coherent R.F. energy from the front end, and an energy storage circuit coupled to receive energy from the rectifier circuit.

The storage circuit may include a tank circuit. The storage circuit may include a rechargeable battery. The rectifier circuit may include a center-tapped transformer.

The R.F. front end may include a folded dipole antenna formed in composite material having an effect permittivity which varies $\leq 5\times10^{-2\circ}\,\mathrm{C.}^{-1}$. The effective permittivity of the composite material may be made to vary $\leq 9\times10^{-3\circ}\,\mathrm{C.}^{-1}$.

The circuit may further comprise an RFID circuit coupled to receive response transmission energy from the energy storage circuit. The RFID circuit may include an interrogation reception circuit and an ID transmission circuit, and further wherein the interrogation reception circuit is coupled to connect energy from the energy storage circuit to power the ID transmission circuit in response to an interrogation signal being received by the interrogation reception circuit.

The circuit may further comprise a wireless appliance coupled to receive power from the energy storage circuit.

The present invention describes methods and embodiments that extend the read range of easily concealable passive RFID sensor systems by minimizing activation energies and harvesting power from the RF environment.

An RFID circuit generally includes a stray RF energy collection circuit, an storage device for the collected stray RF energy, an ID transmission circuit and an interrogation reception circuit adapted for coupling stored energy to the ID transmission circuit in response to an interrogation signal sent to the RFID circuit. The energy collection circuit may have resonant antenna sections tuned to frequencies used for global communications and navigation. Narrow conductance band frequency-selective antenna tuning may be used for the interrogator-receive and tag transmit circuits, and also for the read-head antenna system that interrogates the RFID circuit. Meta-material dielectrics may be used to apply the aforementioned antenna turnings to electrically small antennas embedded within the meta-material dielectric. The ID transmit circuitry may optionally include an oscillator, a power amplifier, and modulation circuitry that encodes specific information about the status of an asset associated with the RFID circuit for transmission back to the remote read head.

BRIEF DESCRIPTION OF THE DRAWINGS

For a better understanding of the present invention, together with other and further aspects thereof, reference is made to the following description taken in conjunction with the accompanying figures of the drawing, wherein:

FIG. **1** is a schematic representation of an extended range passive RFID architecture that harvests energy from the RF environment to power the signal sent to the remote interrogator;

FIG. **2** depicts the return loss of a broadband antenna element that is tuned to selectively absorb energy from signals commonly found in the modern RF environment;

FIG. **3** depicts the return loss of narrow conductance band antenna that eliminates the need for additional RF filtering stages in the RFID tag or the read head;

FIG. **4** depicts a physical design of a planar RFID device; and

FIG. **5**. shows TABLE 1, which compares the power losses in a conventional RF front-end used in code-division multiple access (CDMA) to the power losses in an RF front-end that contains a narrow conductance band antenna.

DETAILED DESCRIPTION OF THE INVENTION

The read range of an RFID device, passive or otherwise, is determined by the residual power detectable in dBWatts (dBW) at the remote read head, P_r, which is determined by:

$$P_r = P_t - L_p + G_t + G_r - L_t - L_r, \tag{1}$$

where P_t is the radio frequency (RF) power transmitted by the RFID tag, L_p is the signal path loss, G_t and G_r are the antenna gain of the RFID tag (t) and the remote read head (r) respectively, L_t and L_r represent the transmission line losses between the transmitter circuit and the tag antenna and the read antenna and the read head receiver circuitry, with all values reported in dBW except for the antenna gains which are reported in dBi. The free space signal path loss is determined by:

$$L_p = 36.6 + 20\log(f[\text{MHz}]) + 20\log(d[\text{miles}]), \tag{2}$$

where f is the signal frequency in megahertz units (MHz) and d is the distance in units of miles. Therefore, at a distance of 10 miles the free space path loss for a 900 MHz signal is significant (~116 dBW). Signal loss due to scattering and attenuation caused by features in the landscape and buildings can easily drive path loss values up to 140-160 dBW, which makes the task of remotely detecting the RFID tag quite challenging.

In order to extend the range of remote RFID devices it therefore is essential to increase the transmitted power at the RIFD tag (P_t), increase the gain of the tag and read head antennas (G_t) and (G_r), respectively, and minimize the transmission line losses in the interconnects placing the transmission circuitry in electrical communication with the tag antenna (L_t), and the receiver circuitry in electrical communication with the receive antenna (L_r). As described in the background, the present invention provides a means to minimize transmission line losses (L_t and L_r) by embedding low resistivity alloys or noble metals (copper, silver or gold) in amorphous silica glass ceramic, which is used in optical fiber applications for its ultra-low dielectric loss ($\tan \delta \approx 2\times10^{-5}$).

A limitation to state-of-the-art passive RFID tags is that the interrogator signal is used to activate and power the chirp signal. At a distance of 10 miles, the power of a 900 MHz interrogator signal is already reduced by −116 dBW (in free space). If it requires −15 dBW to −18 dBW to activate the transmitted read signal, the total loss (assuming no transmission line losses i.e., L_t and L_r=0) for the round trip is −247 dBW to −250 dBW. This level of loss makes it extremely difficult to operate a passive RFID system, particularly if the interrogator signal powers the chirp signal. This limitation is overcome by the current invention.

Reference is now made to FIGS. **1** & **2** which depict features of the present invention. The extended range passive RFID **101**, shown in FIG. **1**, uses a selective-frequency broadband antenna system **103** to harvest ambient power from the RF environment. The antenna system is a dipole antenna folded in such a way that current vectors along the dipole arms form capacitive loads (when aligned in anti-parallel directions) and inductive loads (when aligned in parallel directions). By properly folding the arms one can construct one or more RF filtering stages by inserting capacitive and inductive load elements in series or in parallel to create frequency selective pass bands. Further details of this method and technology are described in U.S. Provisional Patent Application Ser. No. 61/355, 755, filed Jun. 17, 2010 and entitled FREQUENCY SELECTIVE DIPOLE ANTENNAS, the contents of which are hereby incorporated by reference herein.

FIG. **2** shows the selective frequency response of a folded dipole that is designed to be resonant at the primary frequencies for civilian communications and navigation, with resonant frequencies at the 450 MHz cellular band **151**, the 900 MHz bands used for the American Mobile Phone System (AMPS) bands **153**, the 1600 MHz global positioning satellite (GPS) bands **155**, the 1800 MHz and 1900 MHz global system mobile (GSM) bands **157**, and the 2.5 GHz (2500 MHz) WiFi bands **159**. This selective-frequency broadband antenna system **103** allows the passive RFID system **101** to harvest energy from the RF environment that is readily available from the civilian wireless communications infrastructure. The selective-frequency broadband antenna need not be tuned to these frequencies, and can be tuned to an arbitrary set of frequencies that could readily harvest power from any given RF environment. These bands are cited here because they are widely used in the global civilian communications infrastructure and can readily supply RF power to the passive tag in most locations around the world.

The RF power harvested by selective-frequency broadband antenna element **103** is then fed through a rectification circuit **105**, which may either be a diode bridge or a center-feed transformer **107** as shown in FIG. **1**, which stores the energy in a lumped circuit (LC) storage tank **109**. The LC storage tank **109** may comprise one or more inductor coils, capacitor and/or resistor elements, or may comprise any device that is capable of storing the harvested RF power for later use. This energy harvested from the RF environment is now available to augment the power of the transmitted signal P_t from the passive RFID tag, which is needed to extend its read range.

Alternatively to the RF ID circuit shown in fig one, the energy collection and storage circuitry described thus far in components **103**, **105**, **107** and **109** may be used for providing power to any sort of wireless device or appliance. This includes, but is not limited to cell phones, computing devices, GPS devices, pagers, radios, etc. For this purpose, it may be desirable to include a rechargeable battery **108** and diode **110**, connected in phantom, to store power either from storage tank **109** or directly from rectifier **105**.

As referenced above and in the background, conventional passive RFID systems rely on the interrogator signal to both activate and supply the power for the transmitted chirp signal. At long distances, such as 10 miles or more, the interrogator power is already diminished by signal path losses, so a −15 dBW to −18 dBW loss associated with activating the chirp signal is highly undesirable. It is therefore a desirable embodiment of the present invention to utilize a weak interrogator signal to activate a switch **111** that dumps the harvested and stored RF power into the transmit circuitry of the extended range passive RFID system **101**. Reference is now made to FIGS. **1**, **3**A, **3**B and Table 1 of FIG. **5** to illustrate preferred embodiments of the invention to maximize performance in this regard. The tag's interrogator-receive antenna **113** registers the interrogator signal, which it then passes through a rectifier circuit **115**. Rectifier circuit **115** is representational in FIG. **1** and may be a center-feed transformer or a rectifying diode bridge. The interrogator-receive circuitry may additionally contain command circuitry (not shown) that may receive digitally encoded instructions to alter the modulation encoding (also not shown) in the RFID system's **101** transmit circuitry **117** to communicate information about the asset, such as maximum exposed temperatures, applied g-forces, or evidence of tampering. The weak interrogator is then used as a base current for a power transistor that functions as the switch **111**. The LC storage tank **109** supplies both the collector voltage V_c and the collector I_c for the switch **111**, while the rectified weak interrogator signal provides base current for the switch **111**. Since gain for the switch **111** is a function of the base current derived from the weak interrogator signal, it may be desirable for the interrogator-receive circuit to also contain a power management or collection system **119** that allows the base current supplied to the switch **111** to obtain a threshold value that maximizes the transmitted power P_t from the extended range passive RFID system. The emitter current and voltage emerging from the switch **111** is then used to power the transmit circuitry **117** that may optionally include an oscillator, a power amplifier, and modulation circuitry that encodes specific information about the asset's status for transmission back to the remote read head.

As alternatively referred to above, the representational transmit circuitry **117** may also be embodied by any form of wireless device or appliance such as cell phones, computing devices, GPS devices, pagers, radios, etc. while components **103**, **105**, **107** and **109** work as a recharging and power supply.

As noted above, it is desirable to minimize transmission line losses in the tag and the read head. This may also include RF front-end losses from filters that are used to improve overall signal integrity in the system. To this end, it is a preferred embodiment of this invention to utilize antenna elements that have been tuned to have narrow conductance band frequency selectivity as shown in FIGS. **3**A, **3**B using the current vector alignment methods described above. In this instance, the RF filtering stages built in to the folded dipole antenna limit the resonant band of the interrogator-receive antenna **113** and the transmit antenna **121** to a narrow communications band **201** of signal frequencies. Since this antenna tuning provides a voltage standing wave ratios (VSWR) frequency **203** bandwidth that is much wider than the conductance bandwidth **201**, this antenna tuning both filters spurious noise from the interrogator-receive antenna **113** and eliminates undesirable emissions from the transmit antenna **121**.

Table 1 provides an analysis of the power that is wasted in an RF front-end designed for code division multiple access (CDMA) communications using conventional filters to an RF front-end that uses a narrow conductance band antenna. Not only is there a 60% reduction in the power wasted through the front-end circuit, the resultant power savings additionally allow a power amplifier (PA) contained in the transmit circuitry **117** to operated at lower back-off voltage, which allows it to operate at much higher efficiency. It is therefore a preferred embodiment of the present invention to use narrow conductance band frequency-selective antenna tuning not just for the interrogator-receive and tag transmit circuits, but also for the read-head antenna system.

Signal path loss can be reduced by using a lower signaling frequency. For instance, 10-30 MHz signaling frequencies reduce the free space path loss from the tag to the receive head from −116 dBW to 74-80 dBW. Lower signaling frequencies consequently require antennas that are much larger in size to achieve higher efficiencies and high gain. This proportionally makes the device much larger in overall size. For the purpose of concealment, it is therefore advantageous to have a smaller device that operates at lower frequency and also has high gain and excellent power efficiency characteristics. It is therefore a preferred embodiment of the invention to use meta-material dielectrics that apply the aforementioned antenna turnings to electrically small antennas embedded within the meta-material dielectric.

Finally, it is desirable that the present invention be easily concealable within a planar surface of the asset or its container. Reference is now made to FIG. **4** to describe a preferred embodiment of the invention in which the antenna systems **251** are configured on a thermally stable miniaturized artificial magnetic conducting (AMC) ground plane **253** that comprises high-energy ceramic meta-material dielectric. To improve system planarity it is also desirable to have all the additional circuit elements, such as the power transistors, diodes, passive components (transformers, inductors, capacitors, resistors)—(not shown in FIG. **4**)—laminated on the backside **255** of said AMC ground plane **253** adjacent to any active semiconductor chip components that may be needed to complete the circuit.

All of the electronic components shown and represented in FIG. **1** may be optimally constructed using the techniques described in the patents and patent applications incorporated herein by reference.

Although the invention has been described with respect to various embodiments, it should be realized this invention is also capable of a wide variety of further and other embodiments within the spirit and scope of the present invention.

What is claimed is:

1. An extended range passive RFID transponder with an R.F. energy collection circuit, comprising:

a first narrow conductance band antenna forming an interrogator receive antenna and in electrical communication with the gate electrode of a transistor switch,

wherein the first narrow conductance band antenna comprises a folded dipole that is resonant at a frequency or frequencies that match the frequency or frequencies of an interrogator signal for the RFID transponder, providing −30 dB or better signal isolation from RF frequencies that do not match the frequency or frequencies of the interrogator signal,

an ID transmission circuit;

an R.F. energy collection circuit, comprising:

a low-loss R.F. front end including a folded dipole antenna tuned to minimize lost power in two or more R.F. bands commonly used for communications for passing substantially coherent R.F. signals received therefrom;

wherein the low-loss R.F. front end comprises low-resistivity alloys or noble metals embedded in a composite material having an effective permittivity that varies $5\times10^{-2\circ}$ C.$^{-1}$;

a rectifier circuit coupled to receive coherent R.F. energy from the front end; and

an energy storage circuit coupled to receive energy from the rectifier circuit and coupled to provide energy to the ID transmission circuit in response to an interrogation signal received by the first narrow conductance band antenna.

2. The transponder of claim **1**, wherein the storage circuit includes a tank circuit.

3. The transponder of claim **1**, wherein the storage circuit includes a rechargeable battery.

4. The transponder of claim **1**, wherein rectifier circuit includes a center-tapped transformer.

5. The transponder of claim **1**, wherein the effective permittivity of the composite material varies $<9\times10^{-3\circ}$ C.$^{-1}$.

6. The transponder of claim **1**, wherein the ID transmission circuit is coupled to the R.F. energy collection circuit through the transistor having the gate electrode coupled to the first narrow conductance hand antenna.

7. The transponder of claim **6**, further comprising a resonant tank circuit connected between the first narrow conductance band antenna and the transistor gate electrode.

8. The transponder of claim **1**, further comprising a wireless appliance coupled to receive power from the energy storage circuit.

9. The transponder of claim **1**, wherein the folded dipole antenna is resonant in two or more separate R.F. bands.

10. The transponder of claim **1**, wherein the energy storage circuit and the passive RFID system is functional without a battery.

* * * * *

US009438063B2

(12) **United States Patent**
Lin et al.

(10) **Patent No.: US 9,438,063 B2**
(45) **Date of Patent: Sep. 6, 2016**

(54) **CHARGE APPARATUS**

(71) Applicant: **Industrial Technology Research Institute**, Hsinchu (TW)

(72) Inventors: **Jui-Ying Lin**, Hsinchu County (TW); **Mu-Tao Chu**, Hsinchu (TW); **Yu-Chen Yu**, Hsinchu County (TW); **Wen-Yih Liao**, Taichung (TW); **Ming-Hsien Wu**, Tainan (TW)

(73) Assignee: **Industrial Technology Research Institute**, Hsinchu (TW)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 261 days.

(21) Appl. No.: **14/197,229**

(22) Filed: **Mar. 5, 2014**

(65) **Prior Publication Data**

US 2014/0217955 A1 Aug. 7, 2014

Related U.S. Application Data

(63) Continuation-in-part of application No. 12/833,218, filed on Jul. 9, 2010, now Pat. No. 8,692,505.

(51) **Int. Cl.**

H02J 7/00	(2006.01)
H02J 7/02	(2016.01)

(Continued)

(52) **U.S. Cl.**
CPC ***H02J 7/025*** (2013.01); ***H02J 7/34*** (2013.01); ***H02J 7/35*** (2013.01); ***H02J 17/00*** (2013.01); *H02J 5/005* (2013.01); *Y02E 10/766* (2013.01)

(58) **Field of Classification Search**
CPC H02J 7/35
USPC 320/101
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,982,139 A	11/1999	Parise
6,567,002 B2	5/2003	Manneschi et al.

(Continued)

FOREIGN PATENT DOCUMENTS

TW	533646	5/2003
TW	M255370	1/2005

(Continued)

OTHER PUBLICATIONS

Takanori Narita, et al., "The Development of Space Solar Power System Technologies," Mitsubishi Heavy Industries Technical Review vol. 48 No. 4, Dec. 2011, pp. 17-22.

(Continued)

Primary Examiner — Suresh Memula
(74) *Attorney, Agent, or Firm* — Jianq Chyun IP Office

(57) **ABSTRACT**

A charge apparatus including a natural energy conversion module, an energy converter, an energy transmitter, an energy receiver, and an electricity storage device is provided. The natural energy conversion module receives the natural energy and converts the natural energy into a first electric energy. The energy converter is electrically connected to the natural energy conversion module and converts the first electric energy into a wireless energy. The energy transmitter is electrically connected to the energy converter and transmits the wireless energy. The energy receiver receives the wireless energy and converts the wireless energy into a second electric energy to charge an external apparatus. The energy transmitter monitors the current charge status of the external apparatus so as to adjust transmitting the wireless energy. The electricity storage device is electrically connected to the natural energy conversion module to store the first electric energy.

17 Claims, 11 Drawing Sheets

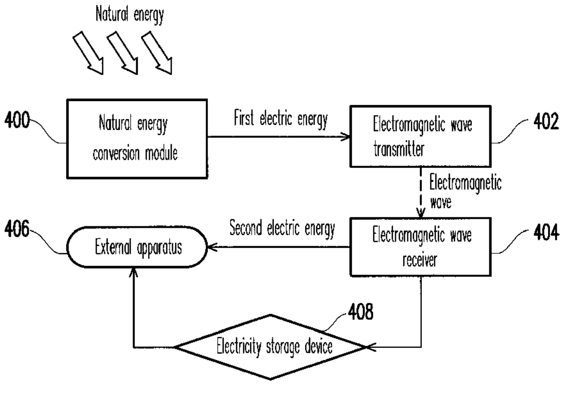
Natural energy
400
Natural energy conversion module
First electric energy
Electromagnetic wave transmitter
402
Electromagnetic wave
406
External apparatus
Second electric energy
Electromagnetic wave receiver
404
408
Electricity storage device

(51) **Int. Cl.**

H02J 7/35	(2006.01)
H02J 7/34	(2006.01)
H02J 17/00	(2006.01)
H02J 5/00	(2016.01)

(56) References Cited

U.S. PATENT DOCUMENTS

6,882,128	B1 *	4/2005	Rahmel	H02J 17/00 320/101
6,912,904	B2	7/2005	Storm, Jr. et al.	
6,949,909	B2	9/2005	Chuang	
6,967,462	B1	11/2005	Landis	
6,994,448	B1	2/2006	Gorrell	
7,471,062	B2	12/2008	Bruning	
8,487,478	B2	7/2013	Kirby et al.	
8,547,057	B2	10/2013	Dunworth et al.	
8,611,815	B2	12/2013	Mohammadian et al.	
8,629,650	B2	1/2014	Mohammadian et al.	
2003/0098057	A1	5/2003	Mizuno et al.	
2004/0142733	A1	7/2004	Parise	
2005/0062454	A1	3/2005	Raghunath et al.	
2006/0238365	A1	10/2006	Vecchione et al.	
2006/0266917	A1	11/2006	Baldis et al.	
2007/0019693	A1	1/2007	Graham	
2007/0114967	A1	5/2007	Peng	
2008/0236648	A1	10/2008	Klein et al.	
2009/0284082	A1	11/2009	Mohammadian	
2009/0284218	A1	11/2009	Mohammadian et al.	
2009/0284220	A1	11/2009	Toncich et al.	
2009/0284227	A1	11/2009	Mohammadian et al.	
2009/0284245	A1	11/2009	Kirby et al.	
2009/0284369	A1	11/2009	Toncich et al.	
2009/0286470	A1	11/2009	Mohammadian et al.	
2009/0286475	A1	11/2009	Toncich et al.	
2009/0286476	A1	11/2009	Toncich et al.	
2010/0194335	A1	8/2010	Kirby et al.	
2010/0201189	A1	8/2010	Kirby et al.	
2010/0201201	A1	8/2010	Mobarhan et al.	
2010/0201202	A1	8/2010	Kirby et al.	
2010/0201311	A1	8/2010	Kirby et al.	
2010/0201312	A1	8/2010	Kirby et al.	
2010/0201314	A1	8/2010	Toncich et al.	
2010/0201533	A1	8/2010	Kirby et al.	
2010/0207572	A1	8/2010	Kirby et al.	
2010/0213895	A1	8/2010	Keating et al.	
2010/0225270	A1	9/2010	Jacobs et al.	
2010/0225272	A1	9/2010	Kirby et al.	
2010/0253281	A1	10/2010	Li	
2010/0289341	A1	11/2010	Ozaki et al.	
2010/0323616	A1	12/2010	Von Novak et al.	
2011/0030757	A1	2/2011	Lin et al.	
2011/0115431	A1	5/2011	Dunworth et al.	
2011/0115432	A1	5/2011	El-Maleh et al.	
2011/0119135	A1	5/2011	Grilli et al.	
2011/0119144	A1	5/2011	Grilli et al.	
2012/0019057	A9	1/2012	Kirby et al.	
2012/0153731	A9	6/2012	Kirby et al.	
2013/0134922	A1 *	5/2013	Takagi	H01M 10/44 320/101
2013/0147428	A1	6/2013	Kirby et al.	
2013/0147429	A1	6/2013	Kirby et al.	
2013/0249481	A1	9/2013	Jacobs et al.	
2013/0300358	A1	11/2013	Kirby et al.	

FOREIGN PATENT DOCUMENTS

TW	M317697		8/2007
TW	200931763		7/2009
WO	2007008646		1/2007
WO	2009140216	A2	11/2009
WO	2009140217	A2	11/2009
WO	2009140218	A1	11/2009
WO	2009140219	A1	11/2009
WO	2009140220	A1	11/2009
WO	2009140221	A2	11/2009
WO	2009140222	A1	11/2009
WO	2009140223	A1	11/2009
WO	2009140216	A3	1/2010
WO	2009140217	A3	1/2010
WO	2009140221	A3	1/2010
WO	2009140221	A8	4/2010
WO	2010047850	A1	4/2010
WO	2010057224	A1	5/2010
WO	2010093719	A1	8/2010
WO	2010093721	A1	8/2010
WO	2010093723	A1	8/2010
WO	2010093724	A1	8/2010
WO	2010093728	A2	8/2010
WO	2010093729	A1	8/2010
WO	2010093730	A2	8/2010
WO	2010093964	A2	8/2010
WO	2010093967	A2	8/2010
WO	2010093973	A2	8/2010
WO	2010118161	A2	10/2010
WO	2010093973	A3	11/2010
WO	2009140221	A9	12/2010
WO	2010093728	A3	12/2010
WO	2010093729	A2	12/2010
WO	2010093730	A3	12/2010
WO	2010144885	A1	12/2010
WO	2010118161	A3	4/2011
WO	2010093964	A3	5/2011
WO	2011063053	A2	5/2011
WO	2011063054	A1	5/2011
WO	2011063057	A2	5/2011
WO	2011063057	A3	5/2011
WO	2011063058	A2	5/2011
WO	2010093967	A3	7/2011
WO	2011063053	A3	7/2011
WO	2011063058	A3	7/2011
WO	2013052178		4/2013

OTHER PUBLICATIONS

Brown, "The History of Power Transmission by Radio Waves", IEEE Transactions on Microwave Theory and Techniques, vol. 32, Issue:9, Sep. 1984, pp. 1230-1242.

Baarman, "Understanding Wireless Power", Dec. 2009, Fulton Innovation LLC, pp. 1-13.

Shinohara, "Wireless Power Transmission for Solar Power Satellite (SPS)", Space Solar Power Workshop, Georgia Institute of Technology, 2006, pp. 1-43.

* cited by examiner

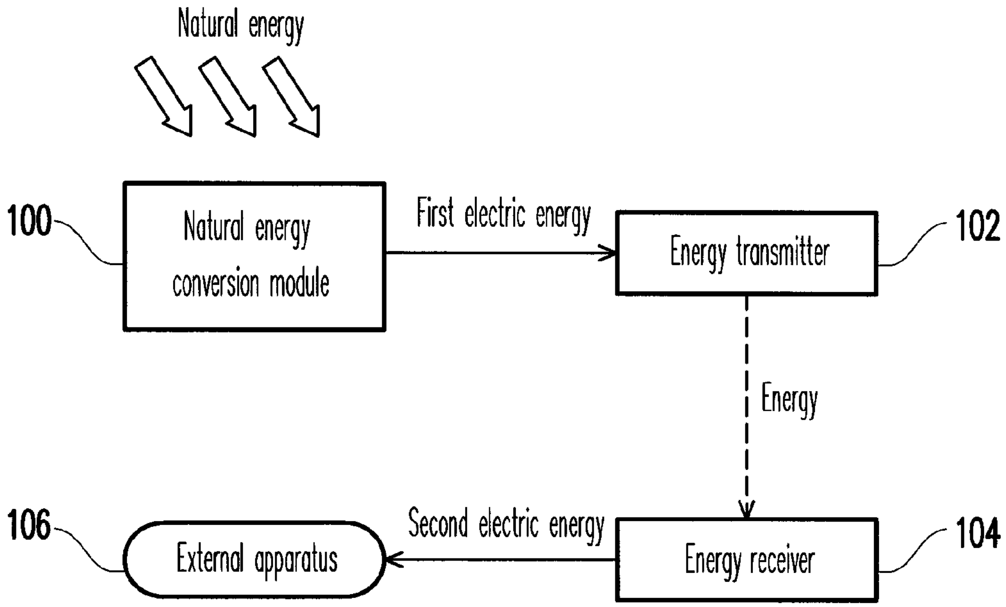
Natural energy
100
Natural energy conversion module
First electric energy
Energy transmitter
102
Energy
106
External apparatus
Second electric energy
Energy receiver
104

FIG. 1

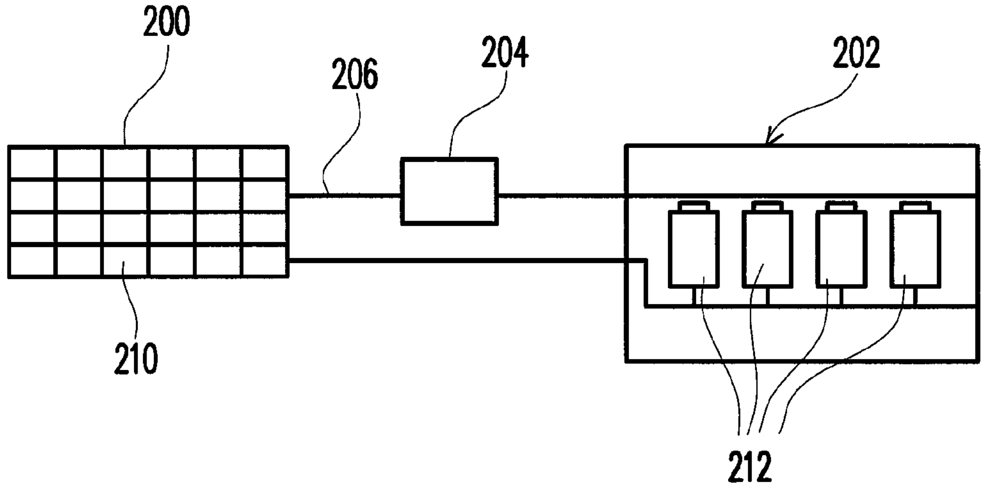

FIG. 2

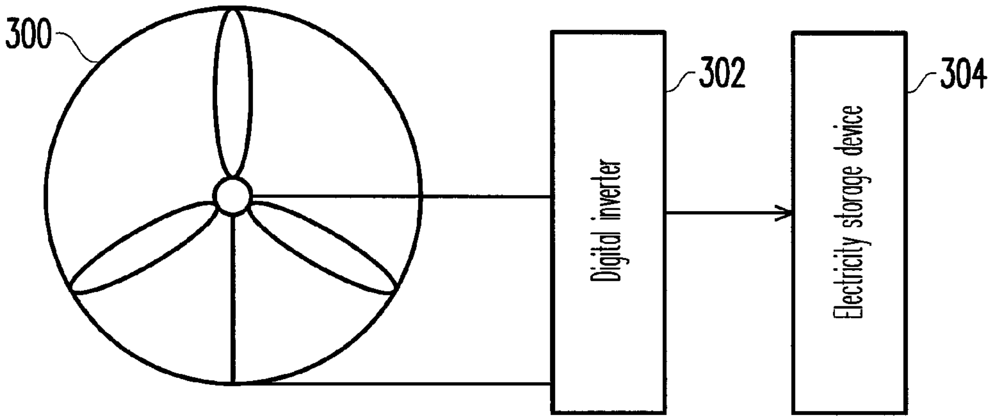
Digital inverter
Electricity storage device

FIG. 3

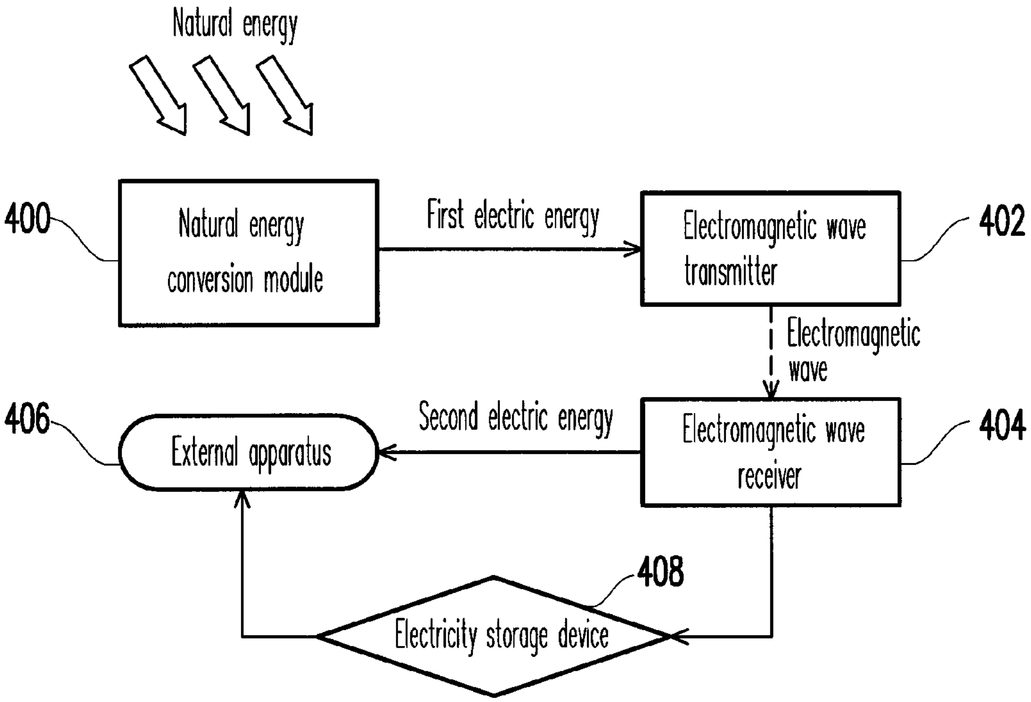
Natural energy
400
Natural energy conversion module
First electric energy
Electromagnetic wave transmitter
402
Electromagnetic wave
406
External apparatus
Second electric energy
Electromagnetic wave receiver
404
408
Electricity storage device

FIG. 4

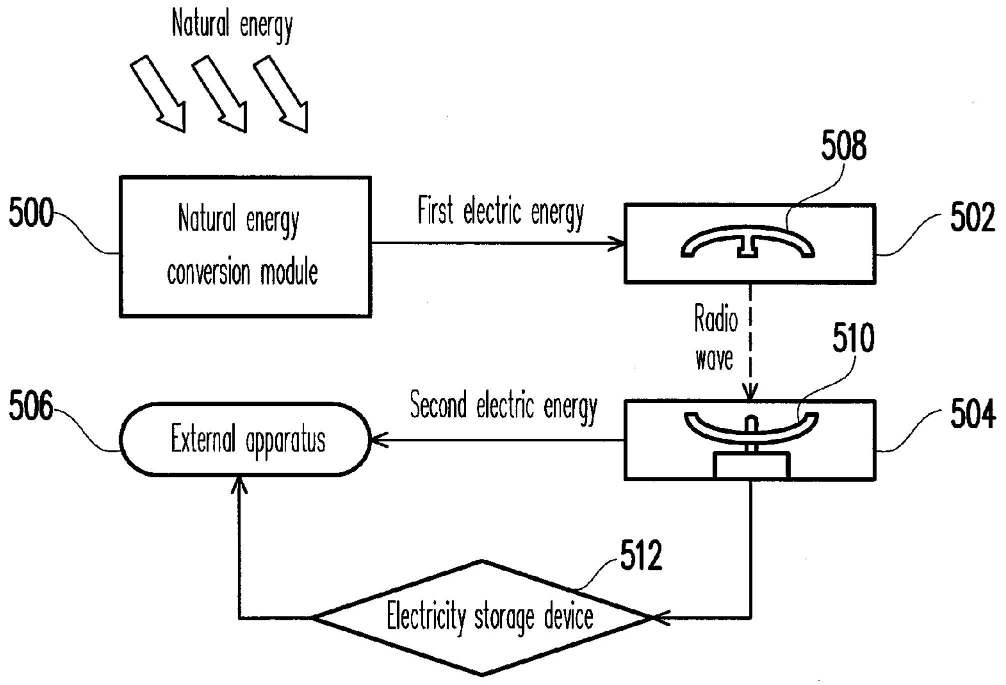
Natural energy
500
Natural energy conversion module
First electric energy
508
502
Radio wave
510
504
506
External apparatus
Second electric energy
512
Electricity storage device

FIG. 5

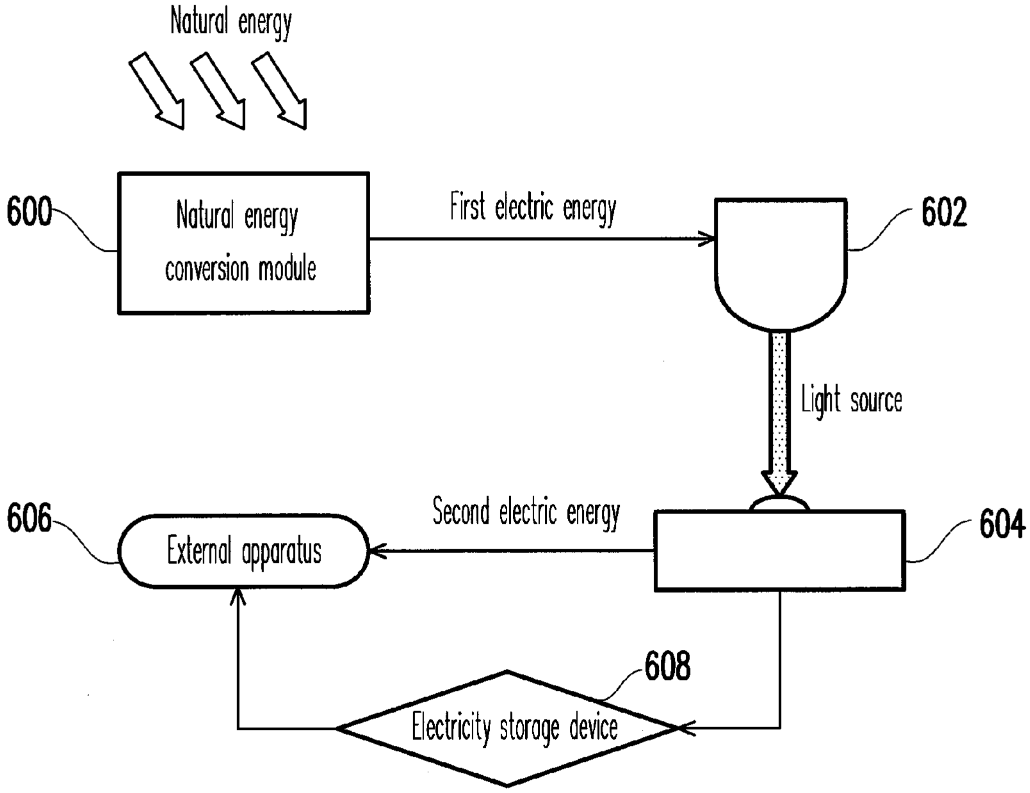
Natural energy
600
Natural energy conversion module
First electric energy
602
Light source
606
External apparatus
Second electric energy
604
608
Electricity storage device

FIG. 6

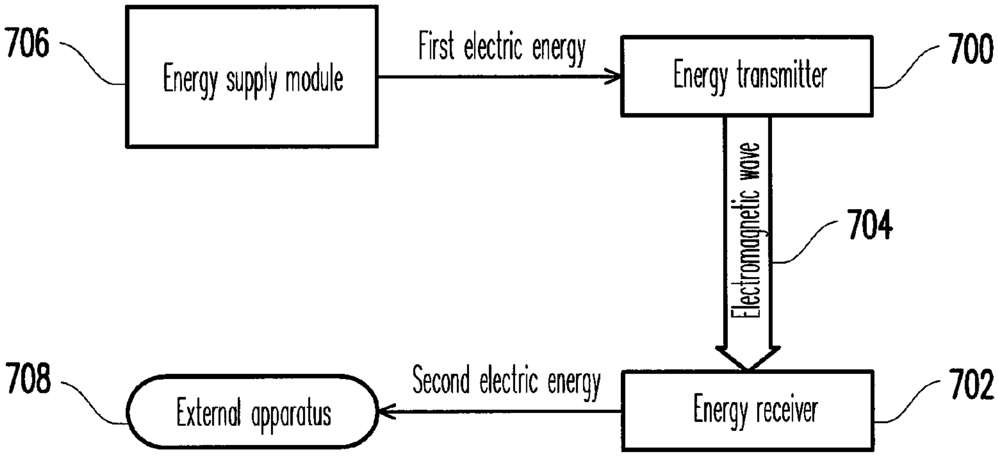
706
Energy supply module
First electric energy
Energy transmitter
700
Electromagnetic wave
704
708
External apparatus
Second electric energy
Energy receiver
702

FIG. 7

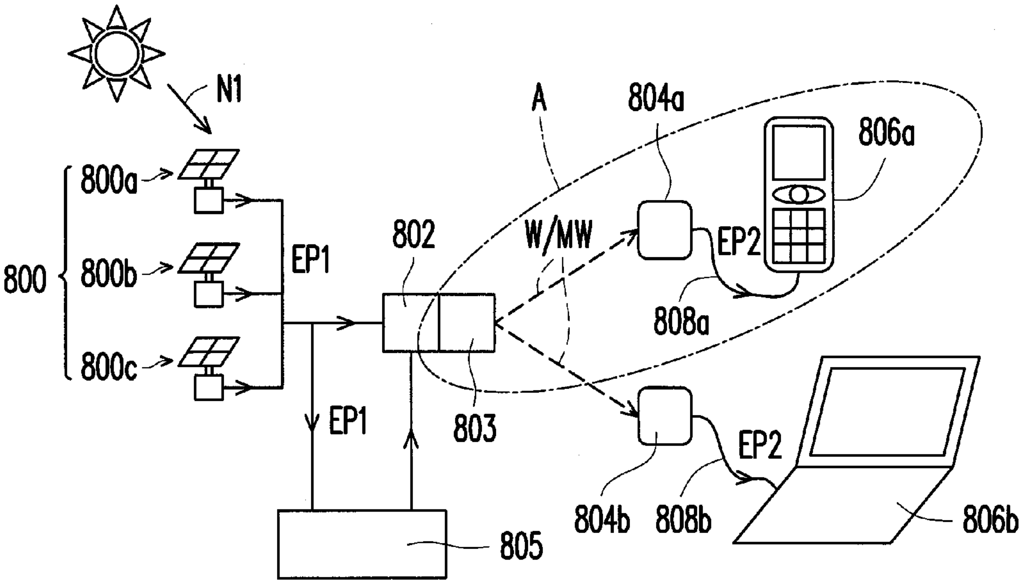
N1
EP1
EP1
A
W/MW
EP2
EP2

FIG. 8A

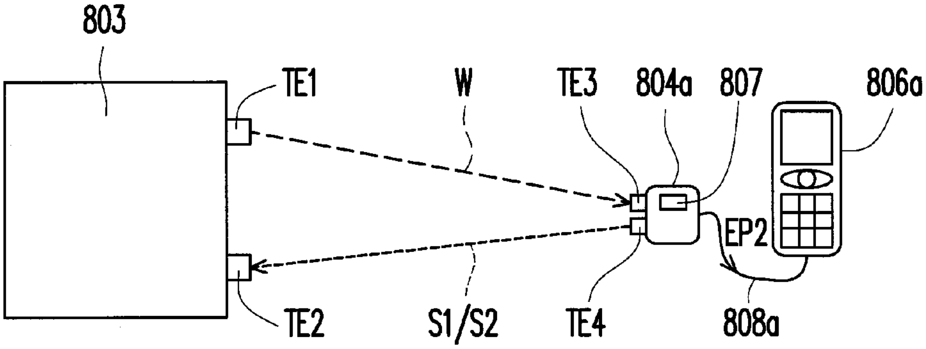
803
TE1
W
TE3
804a
807
806a
EP2
TE2
S1/S2
TE4
808a

FIG. 8B

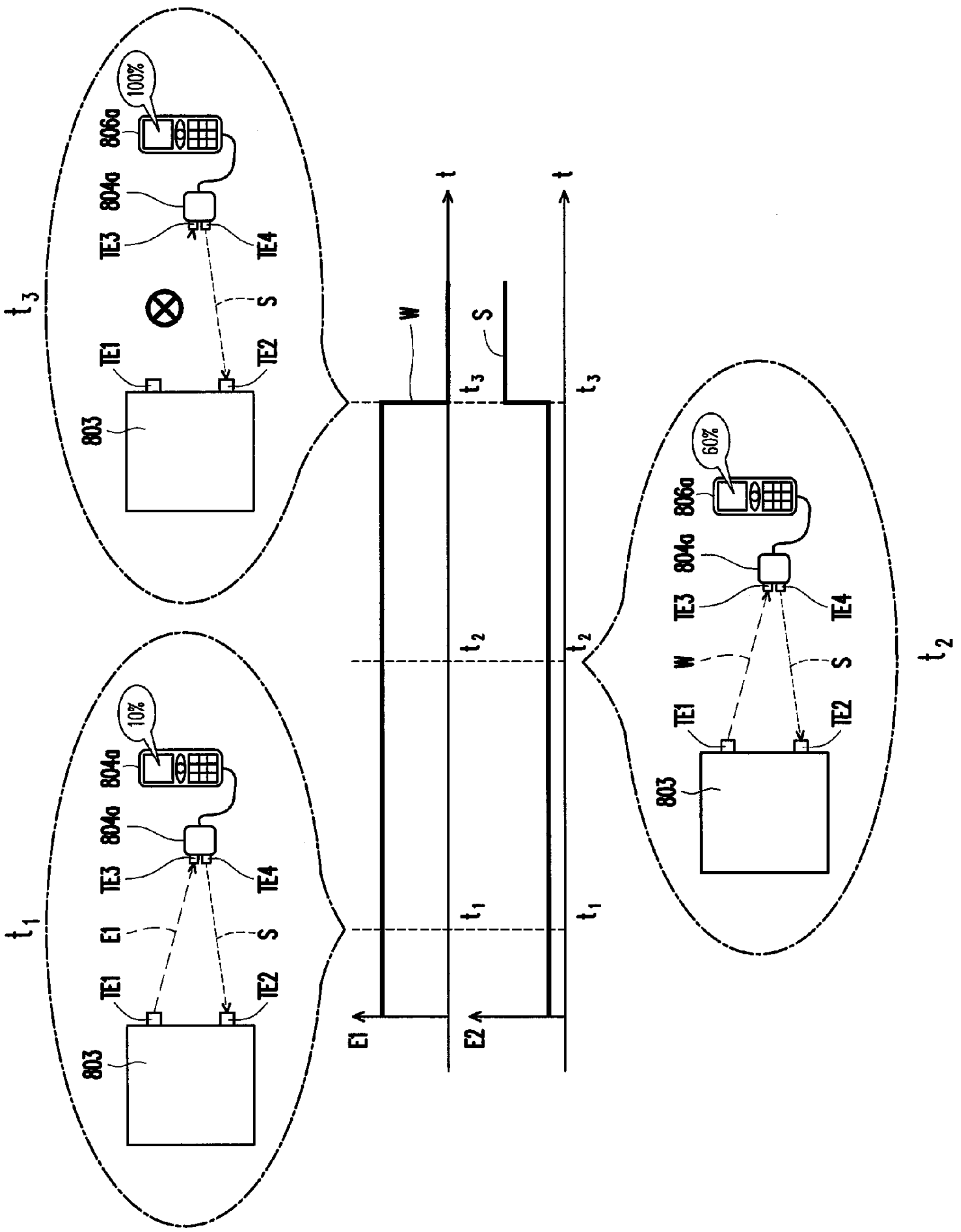
t1
803
TE1
E1
TE3
804a
804a
10%
S
TE2
TE4
t3
803
TE1
TE3
804a
806a
100%
S
TE2
TE4
E1
E2
W
S
t
t
t1
t2
t3
t1
t2
t3
803
TE1
W
TE3
804a
806a
60%
S
TE2
TE4
t2

FIG. 9

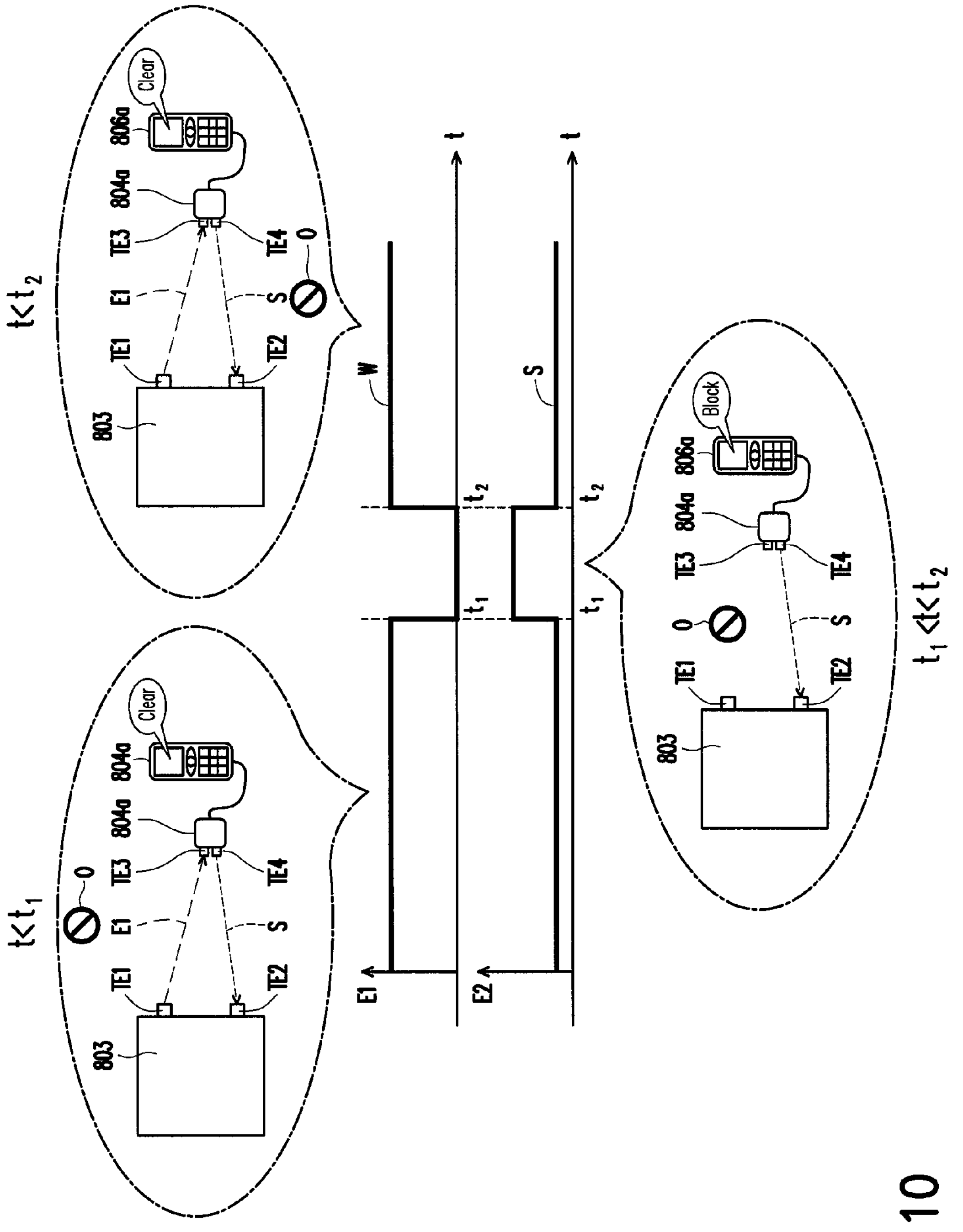
t<t1
803
TE1
E1
S
TE2
O
TE3
TE4
804a
804a
Clear
t<t2
803
TE1
E1
S
TE2
O
TE3
TE4
804a
806a
Clear
E1
W
t1
t2
t
E2
S
t1
t2
t
t1<t<t2
803
TE1
TE2
O
S
TE3
TE4
804a
806a
Block

FIG. 10

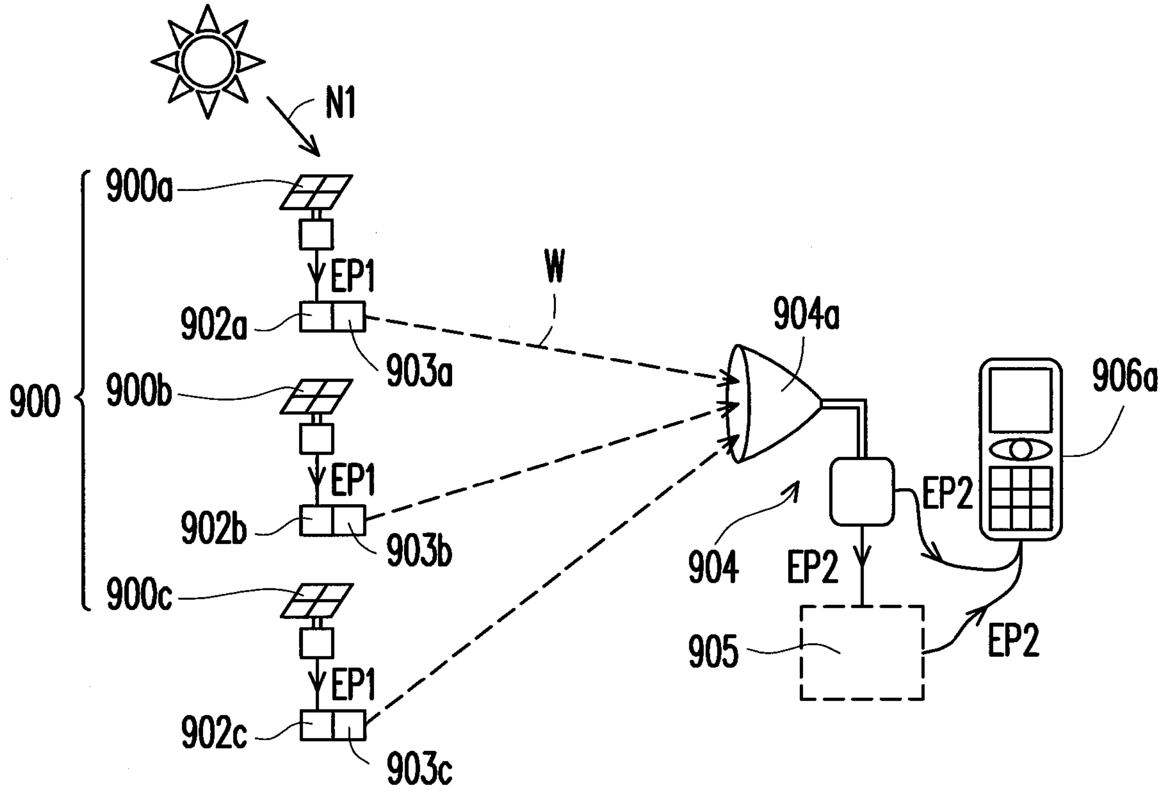
N1
EP1
EP1
EP1
W
EP2
EP2
EP2

FIG. 11

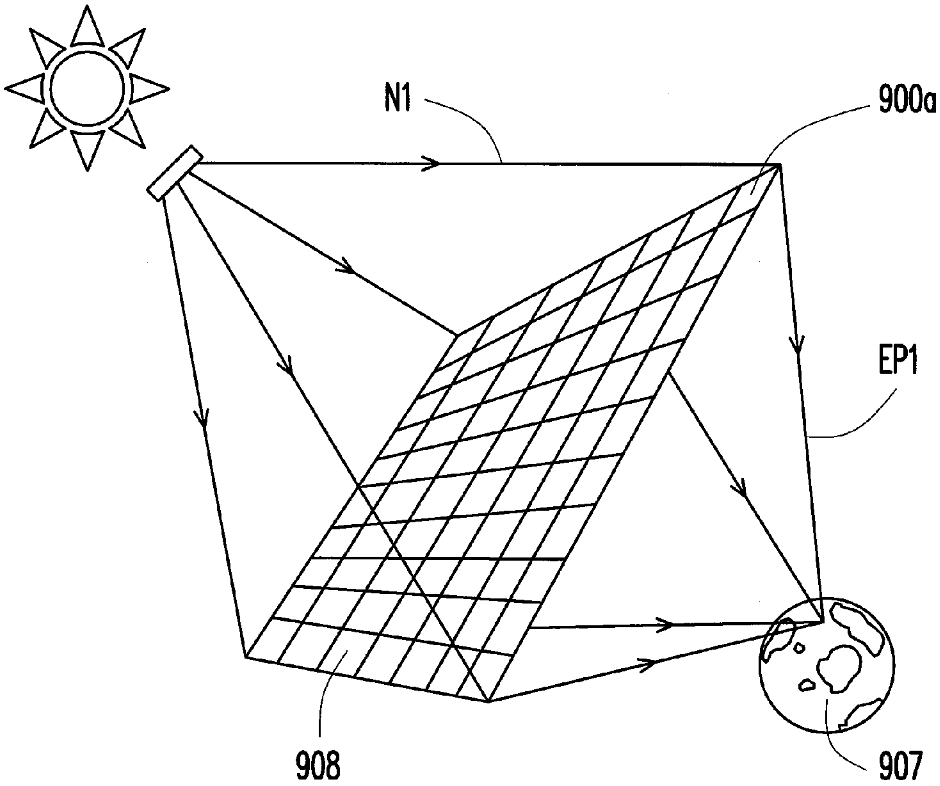
N1
EP1

FIG. 12

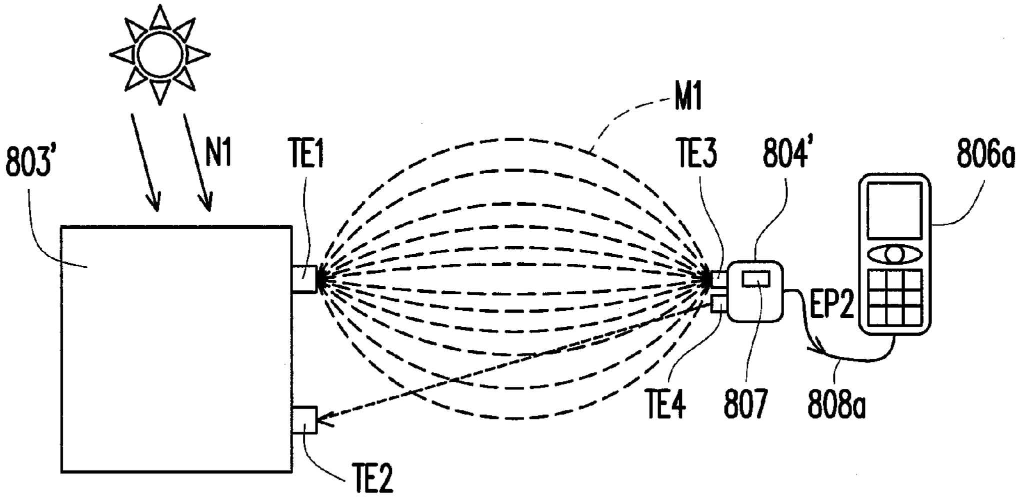
M1
803'
N1
TE1
TE3
804'
806a
EP2
TE4
807
808a
TE2

FIG. 13

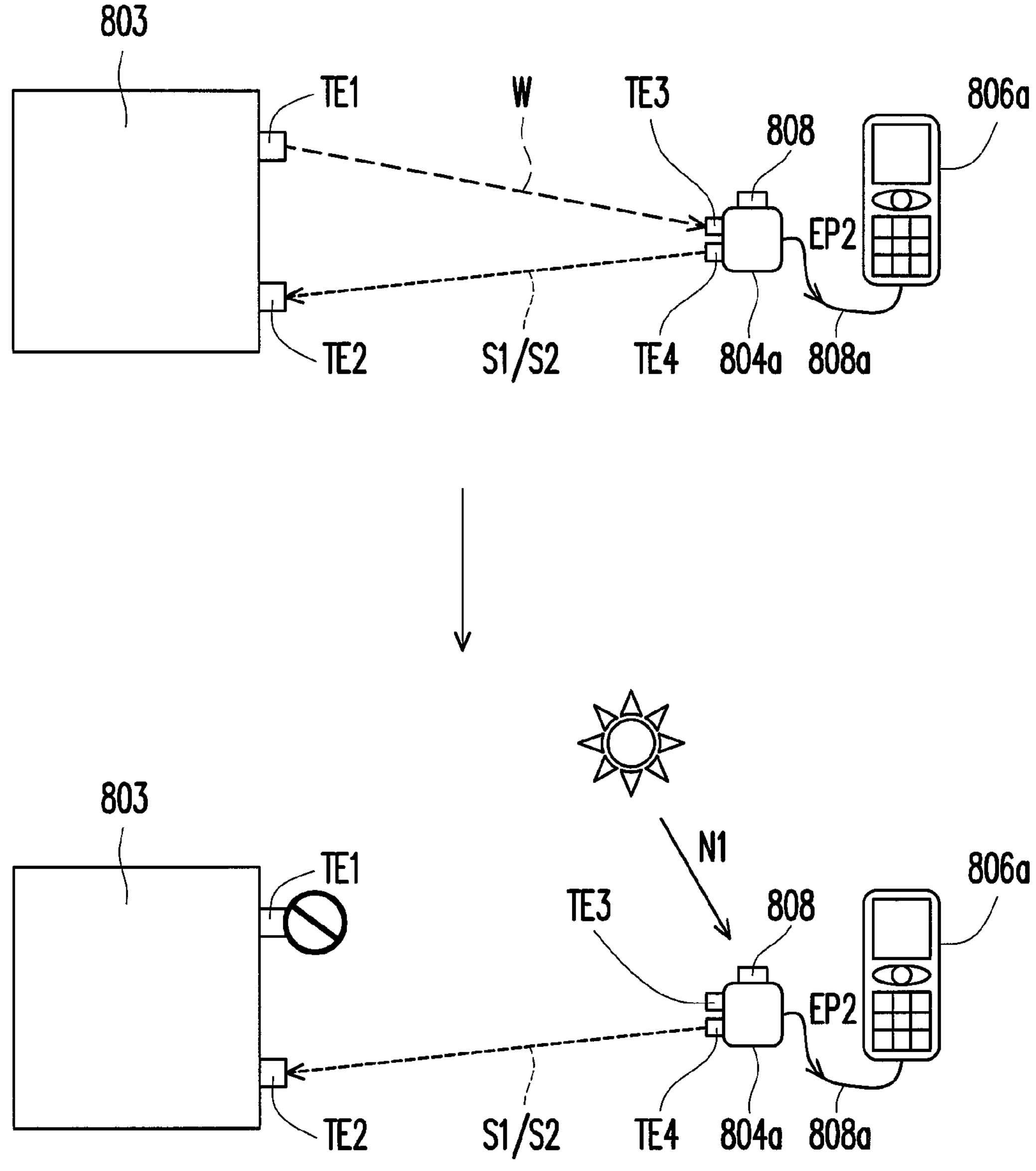
803
TE1
W
TE3
808
806a
EP2
TE2
S1/S2
TE4
804a
808a
803
TE1
N1
TE3
808
806a
EP2
TE2
S1/S2
TE4
804a
808a

FIG. 14

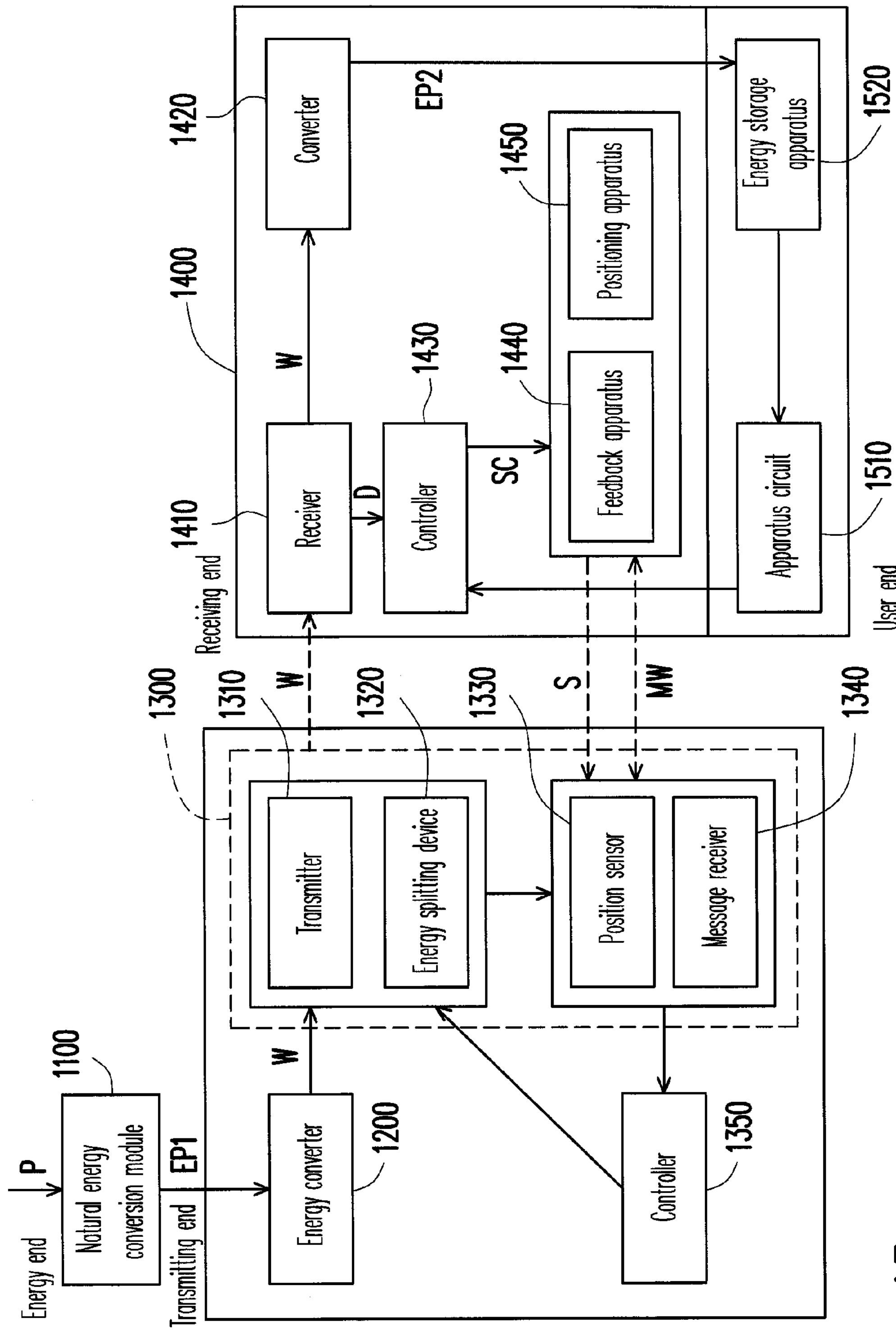
Energy end
P
Natural energy conversion module
1100
EP1
Transmitting end
Energy converter
1200
W
1300
Transmitter
1310
Energy splitting device
1320
Position sensor
1330
Message receiver
1340
Controller
1350
W
S
MW
Receiving end
Receiver
1410
D
Controller
1430
SC
Feedback apparatus
1440
Positioning apparatus
1450
W
Converter
1420
1400
EP2
User end
Apparatus circuit
1510
Energy storage apparatus
1520

FIG. 15

CHARGE APPARATUS

CROSS REFERENCE TO RELATED APPLICATION

This application is a continuation-in-part of prior U.S. patent application Ser. No. 12/833,218 filed on Jul. 9, 2010. The entirety of the above-mentioned patent application is hereby incorporated by reference herein and made a part of specification.

TECHNICAL FIELD

The disclosure relates to a charge apparatus, and more particularly, to an energy-converting charge apparatus.

BACKGROUND

Natural energy resources offer low pollution, high regeneration, and easy collection. Wind and solar energy are two ideal natural energy resources compared to coal, petroleum, and gas for thermal energy generation (which have limited quantities and are highly pollutive) and water for hydraulic energy generation (which is restricted by the terrain). Particularly, a stable and continuous energy supply can be ensured if different natural energy resources are combined.

In recent years, application of solar energy in energy generation has been increasing significantly. Solar panels are broadly deployed by various manufacturers to reduce the cost in power consumption and set up an environment-friendly manufacturing site. How to effectively convert a natural energy into electricity and store it has become one of the major subjects in today's energy and resources research. A light emitting device which uses solar energy as its power supply is disclosed in U.S. Pat. No. 6,994,448.

However, electric energy is the essential element for driving electronic devices and appliances. Conventionally, electric energy is generated by using thermal energy, hydraulic energy, or nuclear energy and then conducted to electronic devices and appliances to charge the same through transformers, power cables, and receptacles. Since the conventional technique for driving electronic devices and appliances depends on the contact between different elements, the charging effect may not be satisfactory due to oxidized contact points and sparks, electric shocks, power failures may be caused. In addition, different electronic devices and appliances require different power cables and charge apparatuses. Thus, the cables may be tangled together if different electronic products are used, and the working area of these electronic products may be limited by the position of the power source and the lengths of the cables.

Some wireless charging techniques have been developed in recent years. For example, an optical charge apparatus including a light source and a sensor connected to the light source is disclosed in U.S. Patent Publication 2007/0114967. When the sensor detects that a portable electronic apparatus having photovoltaic cells as its receiving device is getting close to the light source, the sensor controls the light source to emit light. The light emitted by the light source reaches the photovoltaic cells in the portable electronic apparatus to charge the portable electronic apparatus. However, because the photovoltaic cells are used as the receiving end of the energy transmission, this charge apparatus disclosed in the U.S. Patent Publication 2007/0114967 has a low photoelectric conversion rate therefore is not so practical.

A non-radiative energy wireless transmission technique is disclosed in PCT Publication WO 2007/008646 A2, wherein the energy is transmitted through resonance coupling between two resonance structure apparatuses.

Additionally, a wireless energy transmission system is disclosed in U.S. Patent Publication 2006/0266917, wherein an energy signal is issued and the position of an electronic product to be charged is determined at the receiving end through a reflection mechanism so as to charge the electronic product. A wireless charging system adaptable to a vehicle is disclosed in U.S. Pat. No. 5,982,139. When the wireless charging system receives a positioning signal of the vehicle, the wireless charging system charges the vehicle. The receiving end of the wireless charging system includes a logic circuit for detecting whether the electric volume of the vehicle reaches a charging standard. The wireless charging system can charge only one vehicle at one time.

In U.S. Patent Publication 2007/00019693, the position of an electronic product that is to be charged is determined through an image captured by a camera, and the electric volume of the electronic product is detected to determine whether to turn on a wireless energy transmission function. Furthermore, an optical receiver is disclosed in U.S. Pat. No. 6,972,904, wherein a diffractive optical element (DOE) is adopted for diffracting a light beam to a reflection plane, and the reflection plane is suitable for focusing the light beam into a specific direction.

SUMMARY

A charge apparatus is provided herein.

According to one of the exemplary embodiments of the charge apparatus, the charge apparatus includes a natural energy conversion module, an energy converter, an energy transmitter, an energy receiver, and an electricity storage device. The natural energy conversion module receives the natural energy and converts the natural energy into a first electric energy. The energy converter is electrically connected to the natural energy conversion module and converts the first electric energy into a wireless energy. The energy transmitter is electrically connected to the energy converter and transmits the wireless energy. The energy receiver receives the wireless energy and converts the wireless energy into a second electric energy to charge an external apparatus. The energy transmitter monitors the current charge status of the external apparatus so as to adjust transmitting the wireless energy. The electricity storage device is electrically connected to the natural energy conversion module to store the first electric energy.

According to one of the exemplary embodiments of the charge apparatus, the charge apparatus includes a natural energy conversion module, an energy converter, a transmitting antenna, a receiving antenna, and an electricity storage device. The natural energy conversion module receives the natural energy and converts the natural energy into a first electric energy. The energy converter is electrically connected to the natural energy conversion module and converts the first electric energy into an electromagnetic wave. The transmitting antenna is electrically connected to the energy converter and transmits the electromagnetic wave in a wireless manner. The receiving antenna receives the electromagnetic wave in a wireless manner and converts the electromagnetic wave into a second electric energy to charge an external apparatus. The transmitting antenna monitors the current charge status of the external apparatus so as to adjust transmitting the electromagnetic wave. The electricity storage device is connected to the receiving antenna to store the second electric energy.

According to one of the exemplary embodiments of the charge apparatus driven by a natural energy, the charge apparatus includes an energy receiver and an external apparatus. The energy receiver receives a wireless energy and converts the wireless energy into a second electric energy. The external apparatus is charged by the energy receiver wherein a charge status of the external apparatus is monitored so as to adjust the wireless energy transmitting.

Several exemplary embodiments accompanied with figures are described in detail below to further describe the disclosure in details.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings are included to provide further understanding, and are incorporated in and constitute a part of this specification. The drawings illustrate exemplary embodiments and, together with the description, serve to explain the principles of the disclosure.

FIG. **1** is a block diagram illustrating a charge apparatus driven by a natural energy according to a first exemplary embodiment of the present disclosure.

FIG. **2** is a block diagram illustrating a solar electricity generating apparatus applied to the charge apparatus in the first exemplary embodiment.

FIG. **3** is a block diagram illustrating a wind electricity generating apparatus applied to the charge apparatus in the first exemplary embodiment.

FIG. **4** is a block diagram illustrating a charge apparatus driven by a natural energy according to a second exemplary embodiment of the present disclosure.

FIG. **5** is a block diagram illustrating a charge apparatus driven by a natural energy according to a third exemplary embodiment of the present disclosure.

FIG. **6** is a block diagram illustrating a charge apparatus driven by a natural energy according to a fourth exemplary embodiment of the present disclosure.

FIG. **7** is a block diagram illustrating an electromagnetic wave charge apparatus according to a fifth exemplary embodiment of the present disclosure.

FIG. **8**A is a diagram illustrating a charge apparatus driven by a natural energy according to a sixth exemplary embodiment of the present disclosure.

FIG. **8**B is an enlarged view of area A in FIG. **8**A.

FIG. **9** is a diagram illustrating how an energy transmitter monitors the charge status of an external apparatus.

FIG. **10** is a diagram illustrating how an energy transmitter determines whether there is any obstacle in a transmission path.

FIG. **11** is a diagram illustrating a charge apparatus driven by a natural energy according to a seventh exemplary embodiment of the present disclosure.

FIG. **12** is a diagram illustrating a charge apparatus driven by a natural energy according to another exemplary embodiment of the present disclosure.

FIG. **13** is a diagram illustrating how an energy transmitter transmits a wireless energy to an energy receiver according to another exemplary embodiment of the present disclosure.

FIG. **14** is a diagram illustrating how an energy transmitter transmits a wireless energy to an energy receiver according to another exemplary embodiment of the present disclosure.

FIG. **15** is a block diagram illustrating a charge apparatus driven by a natural energy according to another exemplary embodiment of the present disclosure.

DETAILED DESCRIPTION OF DISCLOSED EMBODIMENTS

FIG. **1** is a block diagram illustrating a charge apparatus driven by a natural energy according to a first exemplary embodiment of the present disclosure.

Referring to FIG. **1**, the natural-energy-driven charge apparatus in the present embodiment includes a natural energy conversion module **100**, an energy transmitter **102**, and an energy receiver **104**. The natural energy conversion module **100** receives a natural energy and converts it into a first electric energy. The energy transmitter **102** receives the first electric energy and converts it into an energy. The energy transmitter **102** may be a light emitting device or a radio wave transmitter. For example, the light emitting device may be a light emitting diode (LED), an incandescent, a laser diode, or other collimated light sources, and the radio wave transmitter may be a transmitting antenna that can convert the first electric energy into an electromagnetic wave and radiate it into the space. In addition, a sensor (not shown) may be disposed on the energy transmitter **102** for detecting the position of the energy receiver **104**. When it is detected that the energy receiver **104** is getting close, the sensor drives the energy transmitter **102** to start working. The energy receiver **104** receives the energy and converts it into a second electric energy to be consumed by an external apparatus **106**. The energy receiver **104** may be a photo detector or a radio wave receiver. For example, the photo detector may be a photovoltaic semiconductor, a photodiode, a phototransistor, or a photovoltaic cell. If the energy receiver **104** is a radio wave receiver, the energy receiver **104** is composed of a receiving antenna that receives the electromagnetic wave and converts it into the second electric energy. In addition, the natural-energy-driven charge apparatus may further include an electricity storage device (not shown) for storing the second electric energy. The energy transmitter **102** and the energy receiver **104** may be induction coils. Energy is transmitted between the energy transmitter **102** and the energy receiver **104** (as indicated by the dotted arrow in FIG. **1**) without any conventional power cable. For example, if the energy transmitter **102** is a light emitting device, the light it generates can be transmitted to the energy receiver **104** through an optical fiber, and if the energy transmitter **102** is a radio wave transmitter, the electromagnetic wave it generates can be transmitted to the energy receiver **104** through waveguide (when the electromagnetic wave is a radio wave) or the air.

Referring to FIG. **1** again, the natural energy conversion module **100** may be a solar electricity generating apparatus or a wind electricity generating apparatus. When the natural energy conversion module **100** is a solar electricity generating apparatus, it converts solar energy into the first electric energy. For example, as shown in FIG. **2**, the solar electricity generating apparatus includes a photovoltaic cell array **200**, a charge battery array **202**, a unidirectional current device **204**, and an outer lead **206**. The photovoltaic cell array **200** includes a plurality of solar chips **210** for receiving solar energy and converting it into the first electric energy. The charge battery array **202** includes a plurality of charge batteries **212** for storing the first electric energy. The unidirectional current device **204** is located between the photovoltaic cell array **200** and the charge battery array **202**. The unidirectional current device **204** may be a blocking diode, and this blocking diode can prevent power consumption caused by countercurrent. The outer lead **206** connects the photovoltaic cell array **200**, the unidirectional current device **204**, and the charge battery array **202**. On the other hand, when the natural energy conversion module **100** is a wind electricity generating apparatus, it converts wind energy into the first electric energy. For example, as shown in FIG. **3**, the wind electricity generating apparatus includes a wind driven generator **300**, a digital inverter **302**, and an electricity storage device **304**. The wind driven generator **300** converts wind energy into the first electric energy. The digital inverter **302** converts the voltage of the first electric energy. The electricity storage device **304** stores the first electric energy.

FIG. **4** is a block diagram illustrating a charge apparatus driven by a natural energy according to a second exemplary embodiment of the present disclosure.

Referring to FIG. **4**, the natural-energy-driven charge apparatus in the present embodiment includes a natural energy conversion module **400**, an electromagnetic wave transmitter **402**, and an electromagnetic wave receiver **404**. The natural energy conversion module **400** receives a natural energy and converts it into a first electric energy. The electromagnetic wave transmitter **402** receives the first electric energy and converts it into an electromagnetic wave (for example, a laser wave or a radio wave). The electromagnetic wave transmitter **402** may be a laser diode (LD) or any other device that can emit a directional light energy. After the electromagnetic wave transmitter **402** generates the laser wave, the light energy is transmitted to the electromagnetic wave receiver **404** through optical fiber or through peer-to-peer transmission with the air as the transmission medium. In addition, a sensor (not shown) may be disposed on the electromagnetic wave transmitter **402** for detecting the position of the electromagnetic wave receiver **404**. When it is detected that the electromagnetic wave receiver **404** is getting close, the sensor drives the electromagnetic wave transmitter **402** to emit the electromagnetic wave. The laser wave may be implemented by using a GaAs visible or invisible laser (for example, a laser having its wavelength between 600 nm and 1600 nm), or a GaN visible or deep ultraviolet laser (for example, a laser having its wavelength between 200 nm and 600 nm). Besides, the laser wave may also be an InP laser (for example, a laser having its wavelength between 1100 nm and 1800 nm). The laser wave may further be a solid-state laser, a gas laser, or a LD.

Referring to FIG. **4** again, the electromagnetic wave receiver **404** in the natural-energy-driven charge apparatus receives the aforementioned laser wave and converts it into a second electric energy to be consumed by an external apparatus **406**. For example, the electromagnetic wave receiver **404** in the present embodiment includes a photo detector. The photo detector may be a photovoltaic semiconductor, a photodiode, a phototransistor, or a photovoltaic cell. The photo detector is applicable to the detection of both visible and invisible lights. The electric energy converted by the photo detector is usually a current source. The current source can be converted into a voltage source. Alternatively, the second electric energy is first stored in an electricity storage device (for example, a battery) **408** and then used for driving the backend external apparatus **406**, e.g. a light emitting device, a mobile device, or a display device, etc. For example, a GaAs laser with wavelength 980 nm emits a 100 mW laser beam and transmits the laser beam to the photo detector through optical fiber to generate a current of 40 mA. Herein this current of 40 mA is enough for driving a small bulb. In the second embodiment, the transmission medium between the electromagnetic wave transmitter **402** and the electromagnetic wave receiver **404** may be air or optical fiber. The natural energy conversion module **400** can be referred to the natural energy conversion module **100** in the first embodiment.

FIG. **5** is a block diagram illustrating a charge apparatus driven by a natural energy according to a third exemplary embodiment of the present disclosure.

Referring to FIG. **5**, the natural-energy-driven charge apparatus in the present embodiment includes a natural energy conversion module **500**, a radio wave transmitter **502**, and a radio wave receiver **504**. The natural energy conversion module **500** receives a natural energy and converts it into a first electric energy. The natural energy conversion module **500** can be referred to the natural energy conversion module **100** in the first embodiment. The radio wave transmitter **502** is connected to the natural energy conversion module **500**, receives the first electric energy and converts the first electric energy into a radio wave. In addition, a sensor (not shown) may be disposed on the radio wave transmitter **502** for detecting the position of the radio wave receiver **504**. When it is detected that the radio wave receiver **504** is getting close, the sensor drives the radio wave transmitter **502** to emit the radio wave. The radio wave receiver **504** is disposed opposite to the radio wave transmitter **502**, receives the radio wave and converts the radio wave into a second electric energy to be consumed by an external apparatus **506**. The radio wave transmitter **502** includes a transmitting antenna **508**, and the transmitting antenna **508** converts the first electric energy into an electromagnetic wave and radiates it into the space. The radio wave receiver **504** also includes a receiving antenna **510** for receiving the radio wave and converting it into the second electric energy. The natural-energy-driven charge apparatus may further include an electricity storage device **512**. The electricity storage device **512** is connected to the radio wave receiver **504** and stores the second electric energy to be consumed by the external apparatus **506**. In the present embodiment, the radio wave transmitter **502** and the radio wave receiver **504** may be induction coils, and the transmission medium between the radio wave transmitter **502** and the radio wave receiver **504** may be waveguide or air.

FIG. **6** is a block diagram illustrating a charge apparatus driven by a natural energy according to a fourth exemplary embodiment of the present disclosure.

Referring to FIG. **6**, the natural-energy-driven charge apparatus in the present embodiment includes a natural energy conversion module **600**, a light emitting device **602**, and a photo detector **604**. The natural energy conversion module **600** receives a natural energy and converts it into a first electric energy. The natural energy conversion module **600** can be referred to the natural energy conversion module **100** in the first embodiment. The light emitting device **602** is connected to the natural energy conversion module **600**, and which receives the first electric energy and converts it into a light source, wherein the light source may be an electromagnetic light (for example, a :laser wave) or a non-electromagnetic light (for example, the light emitted by a LED or a fluorescent bulb). In the present embodiment, the light emitting device **602** may be a LED or a fluorescent bulb. In addition, a sensor (not shown) may be disposed on the light emitting device **602** for detecting the position of the photo detector **604**. When it is detected that the photo detector **604** is getting close, the sensor drives the light emitting device **602** to emit light. Because the photo detector **604** in the present embodiment can detect light of related wavelength, a high photoelectric conversion rate over 60% is achieved. Thus, the energy conversion rate is increased considerably compared to that in a photovoltaic cell. Besides, the wavelengths of light absorbed by a photovoltaic cell range from 0.4 μm to 1.1 μm. In the present disclosure, the selection of wavelength range is made more flexible by adopting a photo detector corresponding to a light source.

Referring to FIG. **6** again, the photo detector **604** is disposed opposite to the light emitting device **602**, receives the aforementioned light source and converts the light source into a second electric energy to be consumed by an external apparatus **606**. The photo detector **604** may be a photovoltaic semiconductor, a photodiode, a phototransistor, or a photovoltaic cell, wherein the phototransistor is usually composed of a photodiode and a transimpedance amplifier. The transimpedance amplifier converts the current supplied by the photodiode into a voltage so that a voltage source is output. Besides, in the present embodiment, the natural-energy-driven charge apparatus further includes a first electricity storage device **608**. The first electricity storage device **608** is connected to the photo detector **604** for storing the second electric energy.

FIG. **7** is a block diagram illustrating an electromagnetic wave charge apparatus according to a fifth exemplary embodiment of the present disclosure.

Referring to FIG. **7**, the electromagnetic wave charge apparatus in the present embodiment includes an energy transmitter **700** and an energy receiver **702**. The energy transmitter **700** receives a first electric energy and converts it into an electromagnetic wave (for example, a radio wave or a laser wave). The energy receiver **702** receives the electromagnetic wave from the energy transmitter **700** through a transmission medium **704** and converts the electromagnetic wave into a second electric energy. In addition, a sensor (not shown) may be disposed on the energy transmitter **700** for detecting the position of the energy receiver **702**. When it is detected that the energy receiver **702** is getting close, the sensor drives the energy transmitter **700** to start working. The transmission medium **704** may be air, optical fiber, or waveguide, wherein the optical fiber may be quartz optical fiber or plastic optical fiber. Moreover, the electromagnetic wave charge apparatus in the present embodiment further includes an energy supply module **706** for supplying the first electric energy to the energy transmitter **700**. The energy supply module **706** may be a general power supply or a natural energy conversion module. For example, the general power supply may be a battery or mains electricity, and the natural energy may be solar energy or wind energy. Finally, the second electric energy converted by the energy receiver **702** is supplied to and consumed by an external apparatus **708**.

FIG. **8**A is a diagram illustrating a charge apparatus driven by a natural energy according to a sixth exemplary embodiment of the present disclosure.

Referring to FIG. **8**A, the natural-energy-driven charge apparatus in the present embodiment includes a natural energy conversion module **800**, an energy converter **802**, an energy transmitter **803**, and a plurality of energy receivers **804***a* and **804***b* (only two of which are illustrated in FIG. **8** as examples).

The natural energy conversion module **800** receives a natural energy **N1** (for example, a solar energy) and converts it into a first electric energy EP**1**. In the present embodiment, the natural energy conversion module **800** includes conversion units **800***a*, **800***b*, and **800***c*. The conversion units **800***a*, **800***b*, and **800***c* may be photovoltaic cells, and these photovoltaic cells are electrically connected with each other and converts the solar energy into the first electric energy EP**1**. In another embodiment, the natural energy **N1** includes at least one of solar energy, water energy, geothermal energy, wave energy, tidal energy, wind energy, and biomass energy.

Referring to FIG. **8**A again, the energy converter **802** converts the first electric energy EP**1** into a wireless energy W. The wireless energy W may be a magnetic energy or an electromagnetic wave, wherein the electromagnetic wave may be a microwave, a light wave, a radio wave, or a laser wave.

The energy transmitter **803** is electrically connected to the energy converter **802**. The energy transmitter **803** receives the wireless energy W and transmits it in a wireless manner. To be specific, when the energy converter **802** converts the first electric energy EP**1** into the wireless energy W which can be conveniently transmitted in a wireless manner, an energy splitting device (not shown) in the energy transmitter **803** splits the wireless energy W into a plurality of energy portions and transmits these energy portions to the energy receivers **804***a* and **804***b* of a plurality of external apparatuses in a wireless manner. The energy splitting device may be disposed on the energy transmitter **803** to simplify the apparatus.

Then, the energy receivers **804***a* and **804***b* receive the wireless energy W in a wireless manner and convert the received wireless energy W into a second electric energy EP**2** to be consumed by external apparatuses **806***a* and **806***b*. Herein the external apparatuses **806***a* and **806***b* may be cell phones or notebook computers. In addition, the wireless energy W in the present embodiment may be laser. Since laser is highly collimated, easy to integrate, and free of radio frequency (RF) interference, the highly intensive wireless energy W to the energy receivers **804***a* and **804***b* can be exactly transmitted.

As shown in FIG. **8**A, the energy receivers **804***a* and **804***b* may be connected to the external apparatuses **806***a* and **806***b* by, for example, the circuits **808***a* and **808***b*, and the circuits **808***a* and **808***b* transmit the converted second electric energy EP**2** to the external apparatuses **806***a* and **806***b* to charge the external apparatuses **806***a* and **806***b*.

Additionally, the energy transmitter **803** in the present embodiment may also transmit a microwave signal MW to locate the energy receiver **804***a* or **804***b*. In other words, in the present embodiment, the position of the external apparatus **806***a* or **806***b* is determined by using microwave. Compared to the conventional positioning technique in which a regular light beam is used, positioning within a large area can be carried out by using microwave due to the high directivity of microwave. Besides, because the transmission of microwave is not affected by the weather (unlike a regular light beam), after the energy transmitter **803** determines the position of the energy receiver **804***a* or **804***b*, it can transmit energy to the external apparatus **806***a* or **806***b* in the specific direction so that the energy transmission efficiency can be improved.

Referring to FIG. **8**A again, the natural-energy-driven charge apparatus in the present embodiment further includes an electricity storage device **805**. The electricity storage device **805** is electrically connected to the natural energy conversion module **800**, and the first electric energy EP**1** is stored in the electricity storage device **805** through circuit transmission. After that, the energy converter **802** converts the first electric energy EP**1** into the wireless energy W, and the energy transmitter **803** transmits the wireless energy W to the energy receivers **804***a* and **804***b* in a wireless manner.

To be specific, in the present embodiment, a converter (not shown) may be disposed on the energy receivers **804***a* and **804***b* to convert the wireless energy W into the second electric energy EP**2**, so that the second electric energy EP**2** can be consumed by the external apparatuses **806***a* and **806***b*. In another embodiment, the energy transmitted between the energy receivers **804***a* and **804***b* and the external apparatuses **806***a* and **806***b* may be through waveguide instead of the circuits **808***a* and **808***b*.

For example, in another embodiment, the energy transmitter **803** may be a transmitting antenna, such as a dipole antenna, a panel antenna, a microstrip antenna, a multiple element dipole antenna, a Yagi antenna, a parabolic dish, an omniantenna, a grid antenna, or a disc antenna. The energy receivers **804***a* and **804***b* may be an receiving antenna, such as a dipole antenna, a panel antenna, a microstrip antenna, a multiple element dipole antenna, a Yagi antenna, a parabolic dish, an omniantenna, a grid antenna, or a disc antenna.

The transmitting frequency of the wireless energy W maybe within the range of 110~150 kHz. The transmitting frequency of the wireless energy W maybe within the range of 350~375 kHz. The transmitting frequency of the wireless energy W maybe within the range of 10~15 MHz. The transmitting frequency of the wireless energy W maybe within the range of 470~860 MHz. The transmitting frequency of the wireless energy W maybe within the range of 850~950 MHz. The transmitting frequency of the wireless energy W maybe 2.45 GHz.

A dipole antenna would include two identical conductive elements such as metal wires or rods, which are usually bilaterally symmetrical Several different variations of the dipole are also used, such as the folded dipole, short dipole, cage dipole, bow-tie, and batwing antenna. Dipoles may be used as standalone antennas themselves, but they are also employed as feed antennas (driven elements) in many more complex antenna types, such as the Yagi antenna, parabolic antenna, reflective array, turnstile antenna, log periodic antenna, and phased array. Yagi antenna is consisting of a driven element and additional parasitic elements. The parasitic element is slightly longer (typically 5% longer) than the driven dipole. a parabolic dish antenna would include a metal parabolic reflector with a small feed antenna suspended in front of the reflector at its focus, pointed back toward the reflector. The reflector is a metallic surface formed into a paraboloid of revolution and usually truncated in a circular rim that forms the diameter of the antenna. Parabolic antennas is nonresonant, so it can function over a wide range of frequencies, that is a wide bandwidth. To change the frequency of operation would involve replacing the feed antenna with one that works at the new frequency. Some parabolic antennas receive at multiple frequencies by having several feed antennas mounted at the focal point, close together.

FIG. **8**B is an enlarged view of area A in FIG. **8**A. As shown in FIG. **8**B, the energy transmitter **803** includes a transmitting end TE**1** and a receiving end TE**2**. The transmitting end TE**1** transmits the wireless energy W to the energy receiver **804***a* in a wireless manner. The receiving end TE**2** receives an instruction signal S1 from the energy receiver **804***a*, wherein the instruction signal S1 provides an energy transmission information to the energy transmitter **803**. In the present embodiment, the energy transmission information may be a current charge status of the external apparatus **806***a*.

On the other hand, the energy receiver **804***a* in the present embodiment may further include a receiving end TE**3** and a transmitting end TE**4**. The receiving end TE**3** receives the wireless energy W from the energy receiver **803**, and the transmitting end TE**4** transmits the instruction signal S1 to the energy receiver **803**. To be specific, when the receiving end TE**3** of the energy receiver **804***a* receives the wireless energy W from the transmitting end TE**1** of the energy transmitter **803**, the transmitting end TE**4** of the energy receiver **804***a* transmits the instruction signal S**1** to the receiving end TE**2** of the energy receiver **803**, wherein the instruction signal S**1** represents the energy transmission information (i.e., the current charge status of the external apparatus **806***a*).

FIG. **9** is a diagram illustrating how the energy transmitter **803** monitors the charge status of the external apparatus **806***a*.

Referring to FIG. **9**, when the external apparatus **806***a* is not fully charged yet (for example, $t=t_1$ and $t=t_2$), the energy transmitter **803** continues to supply the wireless energy W. Contrarily, after the external apparatus **806***a* is fully charged (for example, $t=t_3$), the intensity E**2** corresponding to the instruction signal S changes and is transmitted back to the energy transmitter **803** so that the energy transmitter **803** stops supplying the wireless energy W. As described above, the energy transmitter **803** can monitor the charge status of the external apparatus **806***a* according to the instruction signal S**1**, so as to determine whether to continues supplying the wireless energy W. Thereby, the power consumption can be reduced.

Additionally, as shown in FIG. **8**B, in the present embodiment, the energy receiver **804***a* may further include a controller **807**. The controller **807** executes a logic operation. The energy receiver **804***a* transmits a displacement signal S2 to the energy transmitter **803** according to the operation result of the controller **807** such that the energy transmitter **803** can adjust the position or the angle of the transmitting end TE1.

To be specific, when the energy receiver **804***a* starts to transmit the wireless energy W, the controller **807** performs the position adjustment to find the maximum value of the wireless energy W within the area in which the energy receiver **804***a* is located. Meanwhile, the energy receiver **804***a* keeps receiving the wireless energy W. When the intensity E**1** of the received wireless energy W is lower than a predetermined threshold, the transmitting end TE**4** of the energy receiver **804***a* outputs the displacement signal S**2** to notify the energy transmitter **803** to adjust the position or the angle of the transmitting end TE**1**. Namely, the energy transmitter **803** adjusts its transmitting position or the transmitting angle when the received energy (for example, an electromagnetic wave energy or a magnetic energy) of the energy receiver **804***a* is lower than a predetermined threshold. So that an optimal working point can be maintained between the transmitting end TE1 of the energy transmitter **803** and the receiving end TE3 of the energy receiver **804***a*.

It should be noted that in the present embodiment, even though the wireless energy W is transmitted in a single direction, other signals may be transmitted in a single direction or in two different directions. As described above, the energy transmitter **803** transmits the microwave signal MW to locate the energy receiver **804***a* or **804***b*. After that, the energy receiver **804***a* transmits the displacement signal S**2**, and the energy transmitter **803** determines whether to adjust the position or direction of the transmitting end TE**1** according to the displacement signal S**2**. Namely, the signal transmission between the transmitting end TE**4** and the receiving end TE**2** is not limited to the single direction from the transmitting end TE**4** to the receiving end TE**2**. Instead, it could be bidirectional.

Moreover, as shown in FIG. **10**, besides monitoring the charge status, the transmitting end TE**4** and the receiving end TE**2** may also simultaneously provide an information indicating whether an obstacle O blocking the energy transmission exists in the transmission path. Namely, the energy transmission information may be a transmission path information. Referring to FIG. **10**, when the obstacle O does not exist in the transmission path between the receiving end TE3 and the transmitting end TE**1** (for example, t<t1), the instruction signal S instructs the energy transmitter **803** to supply the wireless energy W continuously. Contrarily, when the obstacle O exists in the transmission path (for example, t**1**<t<t**2**), the energy E**2** of the instruction signal S changes to notify the energy transmitter **803** to stop supplying the wireless energy W. After the obstacle O leaves the transmission path (for example, t>t2), the intensity E2 of the instruction signal S changes again to notify the energy transmitter **803** to supply the wireless energy W again.

As described above, through the obstacle monitoring function, energy loss caused by invalid transmission is prevented, and meanwhile, any damage caused by the transmitted wireless energy to a living creature that passes through the transmission path is also prevented. In other words, the charge apparatus in the present embodiment offers a life protection function. In addition, the functions of the transmitting end TE**4** and the receiving end TE**2** are not limited to those described in the present embodiment, and other possible states (for example, any malfunction or the charge status of the energy transmitter **803**) of the energy transmitter **803** and the energy receiver **804***a* may be shared between the transmitting end TE**4** and the receiving end TE**2**.

Referring to FIG. **8**B, in another embodiment, the transmitting end TE**1** of the energy transmitter **803** may also be a fiber optic connector. In other words, the energy transmitter **803** may be considered a charging station with a fiber optic connector. Thus, the energy transmitter **803** is connected to the energy receiver **804***a* through optical fiber to transmit light energy. After that, the energy receiver **804***a* converts the light energy into the second electric energy EP2 to be consumed by the external apparatus **806***a*.

FIG. **11** is a diagram illustrating a charge apparatus driven by a natural energy according to a seventh exemplary embodiment of the present disclosure.

Referring to FIG. **11**, the natural-energy-driven charge apparatus in the present embodiment includes a natural energy conversion module **900**, a plurality of energy converters **902**, a plurality of energy transmitters **903**, and an energy receiver **904***a*. The natural energy conversion module **900** includes a plurality of conversion units **900***a*, **900***b*, and **900***c*.

In the present embodiment, the conversion units **900***a*, **900***b*, and **900***c* convert a natural energy N**1** into a first electric energy EP**1**, and the conversion units **900***a*, **900***b*, and **900***c* may be photovoltaic cells. The charge apparatus in the present embodiment is similar to the charge apparatus in the sixth embodiment, while the main difference between the two is that each of the conversion units **900***a*, **900***b*, and **900***c* has a corresponding energy converters **902***a*, **902***b*, or **902***c* and a corresponding energy transmitters **903***a*, **903***b*, or **903***c*. The energy converters **902***a*, **902***b*, and **902***c* convert the first electric energy EP**1** into a wireless energy W. The energy transmitters **903***a*, **903***b*, and **903***c* receive the wireless energy W and transmit it to the energy receiver **904** in a wireless manner.

To be specific, when the energy converters **902***a*, **902***b*, and **902***c* convert the first electric energy EP**1** into the wireless energy W (for example, an electromagnetic wave energy), the energy transmitters **903***a*, **903***b*, and **903***c* transmit the wireless energy W to the energy receiver **904** in a wireless manner. In the present embodiment, to increase the wireless transmission efficiency, the energy receiver **904** includes an energy receiving head **904***a*. In the present embodiment, a receiving surface of the energy receiving head **904***a* is in a cone shape. However, in other embodiments, the receiving surface of the energy receiving head **904***a* may also be in a parabolic shape, an elliptical shape, a hemispherical shape, a funneled shape, a bowl shape, a disc shape, or any other shape that can increase the receiving efficiency. For example, in another embodiment, the energy receiver **904** may be an antenna, such as a dipole antenna, a panel antenna, a microstrip antenna, a multiple element dipole antenna, a Yagi antenna, a parabolic dish, an omniantenna, a grid antenna, or a disc antenna.

As shown in FIG. **11**, because the energy receiver **904** in the present embodiment has the energy receiving head **904***a*, the wireless energy W transmitted by the energy transmitters **903***a*, **903***b*, and **903***c* from different directions can be simultaneously received. In addition, an electricity storage device **905** may be connected to the energy receiver **904** externally. After the energy receiver **904** converts the wireless energy W into the second electric energy EP**2**, the second electric energy EP**2** may be stored into the electricity storage device **905** and then a determination for determining whether to transmit the second electric energy EP**2** to an external apparatus **906***a* to charge the same according to the actual requirement is performed.

In another embodiment, the energy receiver **904** may be implemented as a rotary energy receiver such that the receiving angle can be adjusted according to the positions of the energy transmitters **903***a*, **903***b*, and **903***c*. For example, the energy receiving head **904***a* can rotate to sequentially receive the wireless energy W emitted by the energy transmitters **903***a*, **903***b*, and **903***c* from different directions. In other embodiments, the energy transmitters **903***a*, **903***b*, and **903***c* may be disposed in different power plants, and the energy receiver **904** may be disposed on the top of a building along with the electricity storage device **905**, so that the energy receiver **904** can receive the wireless energy W from the power plants, convert the wireless energy W into the second electric energy EP**2**, and store the second electric energy EP**2** into the electricity storage device **905** or supply the second electric energy EP**2** to the residents in the building.

Additionally, in the present embodiment, the conversion units **900***a*, **900***b*, and **900***c* are all solar panels. However, in another embodiment, the conversion units **900***a*, **900***b*, and **900***c* may also be wind power plants, geothermal power plants, or hydraulic power plants.

Moreover, as shown in FIG. **12**, in another embodiment, the conversion unit **900***a* may also be disposed in the space to directly collect solar energy (i.e., the natural energy N**1**). For example, a solar panel **908** that has a surface area of up to several square kilometres can be deployed on a synchronous orbit outside the aerosphere of the Earth **907** for collecting the more intensive solar energy N**1** compared to that collected on the ground. Then, the solar energy N**1** is transmitted back to the energy receiver **904** (illustrated in FIG. **11**) on the Earth **907** through laser or microwave, and the energy receiver **904***a* may be a giant disc antenna located within a prohibited area on the sea or a reservoir.

FIG. **13** is a diagram illustrating how an energy transmitter transmits a wireless energy to an energy receiver according to another exemplary embodiment of the present disclosure.

As shown in FIG. **13**, in the present embodiment, the energy transmitter **803'** transmits the wireless energy W through a magnetic field M**1**. To be specific, the energy transmitter **803'** converts the first electric energy EP**1** received from the energy converter **802** (illustrated in FIG.

8A) into a magnetic energy **M1** and radiates it into the space. The external apparatus **806***a* receives the magnetic energy **M1** through an energy receiver **804'** and converts the magnetic energy **M1** into the second electric energy EP**2** to be consumed by the external apparatus **806***a*. In addition, the energy receiver **804'** may also be a multipolar receiver having multiple magnetic poles. Namely, the energy receiver **804'** is a receiver with multiple N/S polarity combinations such that the magnetic energy **M1** in the space can be effectively collected and converted into the second electric energy EP**2** to be consumed by the external apparatus **806***a*.

FIG. **14** is a diagram illustrating how an energy transmitter transmits a wireless energy to an energy receiver according to another exemplary embodiment of the present disclosure.

As shown in FIG. **14**, besides that the external apparatus **806***a* receives the wireless energy W from the energy transmitter **803** and the energy receiver **804***a* converts the wireless energy W into the second electric energy EP**2**, as illustrated at the lower part of the FIG. **14**, when the energy transmitter **803** stops transmitting energy, if the energy receiver **804***a* is located within the transmission environment of other wireless energy (for example, light energy, radio wave, or solar energy), the energy receiver **804***a* directly converts the natural energy N**1** (for example, the solar energy) into the second electric energy EP**2** to be consumed by the external apparatus **806***a* by using an auxiliary natural energy conversion module **808**. The auxiliary natural energy conversion module **808** may be a solar panel disposed on the energy receiver **804***a*.

Moreover, the energy receiver **804***a* may have a general receptacle, and when the supply of the natural energy N**1** stops, energy can still be supplied to the external apparatus **806***a* through the receptacle or a battery. As described above, the energy receiver **804***a* in the present embodiment is a multifunctional apparatus.

FIG. **15** is a block diagram illustrating a charge apparatus driven by a natural energy according to another exemplary embodiment of the present disclosure. The natural-energy-driven charge apparatus in the present embodiment includes an energy end, a transmitting end, a receiving end, and a user end. The energy end may be an energy conversion module **1100**, and the energy conversion module **1100** receives an energy P, converts the energy P into a first electric energy EP**1**, and transmits the first electric energy EP**1** to the transmitting end, wherein the energy P may be a natural energy, the mains electricity, or other energies. The transmitting end includes an energy converter **1200** and an energy transmitter **1300**. The energy converter **1200** converts the first electric energy EP**1** into a wireless energy W and transmits the wireless energy W to the energy transmitter **1300**.

In the present embodiment, the energy transmitter **1300** includes a transmitter **1310**, an energy splitting device **1320**, a position sensor **1330**, and a message receiver **1340**. The transmitter **1310** transmits the wireless energy W to the receiving end in a wireless manner. The wireless transmission may be achieved through electromagnetic induction, microwave, or light energy, wherein electromagnetic induction has shorter transmission distance but is easier to implement and safer to human health.

In addition, the energy splitting device **1320** allows the energy transmitter **1300** to transmit different portions of the wireless energy W to a plurality of receiving ends. As shown in FIG. **15**, the receiving end may be an energy receiver **1400**, and the energy receiver **1400** includes a receiver **1410** and a converter **1420**. The receiver **1410** receives the wireless energy W and transmits it to the converter **1420** in a wireless manner. The converter **1420** converts the wireless energy W into a second electric energy EP**2** to be consumed by an external apparatus. Herein the external apparatus may be an apparatus circuit **1510** or an energy storage apparatus **1520** at the user end, wherein the energy storage apparatus **1520** may be an electricity storage device deployed on the top of a building to be used by residents in the building.

The energy receiver **1400** further includes a controller **1430**, a feedback apparatus **1440**, and a positioning apparatus **1450**. The positioning apparatus **1450** sends a displacement signal S to the energy transmitter **1300** to allow the energy transmitter **1300** to adjust the transmission direction of the wireless energy W. The controller **1430** controls the feedback apparatus **1440** and the positioning apparatus **1450** according to an information D of the receiver **1410**, wherein the information D may be the status of the wireless energy W currently received by the receiver **1410**. When the controller **1430** detects that the intensity of the wireless energy W received by the receiver **1410** is lower than a predetermined threshold, the feedback apparatus **1440** and the positioning apparatus **1450** that receive a control signal CS**2** send an instruction signal S to the position sensor **1330** and the message receiver **1340** at the transmitting end. Then the controller **1350** at the transmitting end adjusts the energy distribution quantity of the energy splitting device **1320** or the transmission position or direction of the energy transmitter **1300**, so as to allow the receiver **1410** at the receiving end to receive a more intensive wireless energy W. On the other hand, a position sensor **1330** and the positioning apparatus **1450** may communicate with each other through microwave in a single direction or in two different directions between the transmitting end and the receiving end to determine the exact position of the receiving end. Moreover, similar to that in the embodiment described above, the natural-energy-driven charge apparatus in the present embodiment also offers the charge status detecting and obstacle monitoring functions such that the power consumption can be reduced and human safety can be protected. The present embodiment can be referred to related description of foregoing embodiments therefore will not be described herein.

In summary, a natural-energy-driven charge apparatus provided by the present disclosure can convert a natural energy into electric energy, and therefore is very economical and environment-friendly. Besides, the natural-energy-driven charge apparatus provided by the present disclosure can control the charge status of a charged apparatus. In addition, in a natural-energy-driven charge apparatus provided by the present disclosure, the electric energy is converted into an electromagnetic wave by an electrically-driven energy transmitter, and the electromagnetic wave is received and stored as electric energy by an energy receiver, so that the effect of conventional energy charging through power cable is achieved. Thus, the problems of oxidized contact point, power failure, and electromagnetic interference in the existing contact charging technique are prevented, and the problem of tangled cables when different electronic products are used can be avoided. Moreover, in the natural-energy-driven charge apparatus provided by the present disclosure, no additional circuit is required for connecting the energy transmitter and the energy receiver so that the application of the natural-energy-driven charge apparatus is not limited by its working environment.

It will be apparent to those skilled in the art that various modifications and variations can be made to the structure of the disclosed embodiments without departing from the scope or spirit of the disclosure. In view of the foregoing, it is intended that the disclosure cover modifications and variations of this disclosure provided they fall within the scope of the following claims and their equivalents.

What is claimed is:

1. A charge apparatus driven by a natural energy, comprising:
   - a natural energy conversion module receiving the natural energy and converting the natural energy into a first electric energy;
   - an energy converter, electrically connected to the natural energy conversion module to convert the first electric energy into a wireless energy;
   - an energy transmitter, electrically connected to the energy converter to transmit the wireless energy;
   - an energy receiver, receiving the wireless energy and converting the wireless energy into a second electric energy to charge at least one external apparatus, wherein the energy transmitter monitors the current charge status of the external apparatus so as to adjust transmitting the wireless energy; and
   - an electricity storage device is electrically connected to the natural energy conversion module to store the first electric energy.

2. The charge apparatus according to claim **1**, wherein the energy transmitter stops supplying the wireless energy when the external apparatus is fully charged.

3. The charge apparatus according to claim **1**, wherein a receiving angle of the energy receiver is adjusted according to the position of the energy transmitter.

4. The charge apparatus according to claim **1**, wherein the energy receiver has an energy receiving head to rotate to receive the wireless energy.

5. The charge apparatus according to claim **1**, wherein the wireless energy comprises at least one of a magnetic energy and an electromagnetic wave.

6. The charge apparatus according to claim **5**, wherein the electromagnetic wave is a microwave, a light wave, a radio wave, or a laser.

7. The charge apparatus according to claim **1**, wherein the energy transmitter comprises a position sensor for detecting a position of the energy receiver.

8. A charge apparatus driven by a natural energy, comprising:
   - a natural energy conversion module, receiving the natural energy and to convert the natural energy into a first electric energy;
   - an energy converter, electrically connected to the natural energy conversion module to convert the first electric energy into an electromagnetic wave;
   - a transmitting antenna, electrically connected to the energy converter to transmit the electromagnetic wave in a wireless manner;
   - a receiving antenna, receiving the electromagnetic wave in a wireless manner to convert the electromagnetic wave into a second electric energy to charge an external apparatus, wherein the transmitting antenna monitors the current charge status of the external apparatus so as to adjust transmitting the electromagnetic wave; and
   - an electricity storage device is connected to the receiving antenna to store the second electric energy.

9. The charge apparatus according to claim **8**, wherein the charge apparatus further comprises a position sensor.

10. The charge apparatus according to claim **8**, wherein the transmitting antenna stops supplying the electromagnetic wave when the external apparatus is fully charged.

11. The charge apparatus according to claim **8**, wherein a receiving angle of the receiving antenna is adjusted according to the position of the transmitting antenna.

12. The charge apparatus according to claim **8**, wherein the receiving antenna has a receiving head to rotate to receive the electromagnetic wave.

13. A charge apparatus driven by a natural energy, comprising:
   - an energy receiver receiving a wireless energy and converting the wireless energy into a second electric energy; and
   - an external apparatus is charged by the energy receiver wherein a charge status of the external apparatus is monitored so as to adjust the wireless energy transmitting.

14. The charge apparatus according to claim **13**, wherein the energy receiver has an energy receiving head which rotates to receive the wireless energy.

15. The charge apparatus according to claim **13**, wherein the charge apparatus, further comprising:
   - a natural energy conversion module receiving the natural energy and converting the natural energy into a first electric energy;
   - an energy converter, electrically connected to the natural energy conversion module to convert the first electric energy into a wireless energy;
   - an energy transmitter, electrically connected to the energy converter to transmit the wireless energy; and
   - an electricity storage device is electrically connected to the natural energy conversion module to store the first electric energy.

16. The charge apparatus according to claim **15**, wherein the energy transmitter stops supplying the wireless energy when the external apparatus is fully charged.

17. The charge apparatus according to claim **15**, wherein a receiving angle of the energy receiver is adjusted according to the position of the energy transmitter.

* * * * *

US009680327B2

(12) **United States Patent**
Scheb

(10) **Patent No.: US 9,680,327 B2**
(45) **Date of Patent: Jun. 13, 2017**

(54) **RF ENERGY HARVESTING BY A NETWORK NODE**

(71) Applicant: **Landis+Gyr Innovations, Inc.**, Alpharetta, GA (US)

(72) Inventor: **Jeffrey E. Scheb**, Roswell, GA (US)

(73) Assignee: **Landis+Gyr Innovations, Inc.**, Alpharetta, GA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 313 days.

(21) Appl. No.: **14/318,970**

(22) Filed: **Jun. 30, 2014**

(65) **Prior Publication Data**

US 2015/0380973 A1 Dec. 31, 2015

(51) **Int. Cl.**

H01F 27/42	(2006.01)
H01F 37/00	(2006.01)
H01F 38/00	(2006.01)
H02J 7/02	(2016.01)
H02J 5/00	(2016.01)

(Continued)

(52) **U.S. Cl.**
CPC ***H02J 7/025*** (2013.01); ***H02J 5/00*** (2013.01); ***H02J 17/00*** (2013.01); ***H02J 50/20*** (2016.02); ***H04B 1/1607*** (2013.01); ***H04W 52/0251*** (2013.01); ***H04W 52/0296*** (2013.01); *H04W 52/0216* (2013.01); *H04W 52/0248* (2013.01)

(58) **Field of Classification Search**
CPC .. H02J 5/005; H02J 7/025; H02J 17/00; H02J 50/12; H02J 50/10; H02J 50/90; H02J 50/80; H02J 7/00; H02J 7/0042; H02J 2007/0096; H02J 50/20; H02J 50/60; H02J 7/04; H02J 50/05; H02J 50/27
USPC 307/104
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,490,257 B1 * 12/2002 Watanabe H04B 7/0608 370/318
2006/0104291 A1 * 5/2006 Rodriguez H04B 3/54 370/401

(Continued)

FOREIGN PATENT DOCUMENTS

WO	2007079490	7/2007
WO	2016003552	1/2016

OTHER PUBLICATIONS

Ostaffe, "RF-based Wireless Charging and Energy Harvesting Enables New Applications and Improves Product Design", mouser.com, Applications & Technologies, at least as early as Mar. 28, 2014, 2 pages.

(Continued)

Primary Examiner — Jared Fureman
Assistant Examiner — Esayas Yeshaw
(74) *Attorney, Agent, or Firm* — Kilpatrick Townsend & Stockton LLP

(57) **ABSTRACT**

Disclosed are various embodiments for energy harvesting by circuitry of a node in a network. The node operates in a communication mode that includes communicating, via an RF transceiver, data with another node in the network. The node includes an energy storage device that provides power for the node. The node switches, via an RF switch, from said communication mode to an energy harvest mode. The energy harvest mode includes receiving RF emissions accessible to the node via an antenna. The RF emissions are converted to DC electrical energy, and the DC energy is stored in the energy storage device.

19 Claims, 5 Drawing Sheets

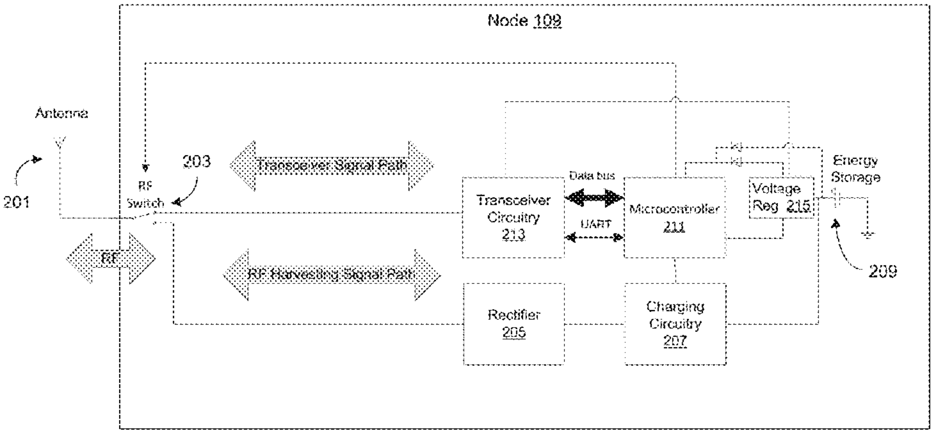

(51) **Int. Cl.**

H02J 17/00	(2006.01)
H04B 1/16	(2006.01)
H02J 50/20	(2016.01)
H04W 52/02	(2009.01)

(56) **References Cited**

U.S. PATENT DOCUMENTS

2009/0174361	A1*	7/2009	Duron	H02J 17/00 320/101
2010/0134257	A1*	6/2010	Puleston	G06K 7/0008 340/10.4
2010/0142509	A1	6/2010	Zhu et al.	
2012/0161721	A1*	6/2012	Neethimanickam	H02J 17/00 320/167
2013/0137455	A1*	5/2013	Xia	H02J 17/00 455/456.1
2013/0234536	A1*	9/2013	Chemishkian	H02J 17/00 307/149
2013/0257360	A1*	10/2013	Singh	H01F 38/14 320/108
2014/0004912	A1	1/2014	Rajakarunanayake	
2014/0111032	A1	4/2014	Shearer et al.	
2014/0159647	A1*	6/2014	Nishioka	H02J 7/025 320/107
2014/0370830	A1*	12/2014	Steer	H04W 52/18 455/127.1
2015/0369870	A1*	12/2015	Okada	H01M 10/44 324/430

OTHER PUBLICATIONS

Publitek Marketing Communication, “Tune in, Charge up: RF Energy Harvesting Shows Its Potential”, Digi-Key Electronics, May 29, 2013, 5 pages.

PCT/US2015/032367, “International Search Report and Written Opinion”, Nov. 2, 2015, 16 pages.

International Application No. PCT/US2015/032367, Invitation to Pay Additional Fees and Partial Search Report mailed on Aug. 17, 2015, 6 pages.

* cited by examiner

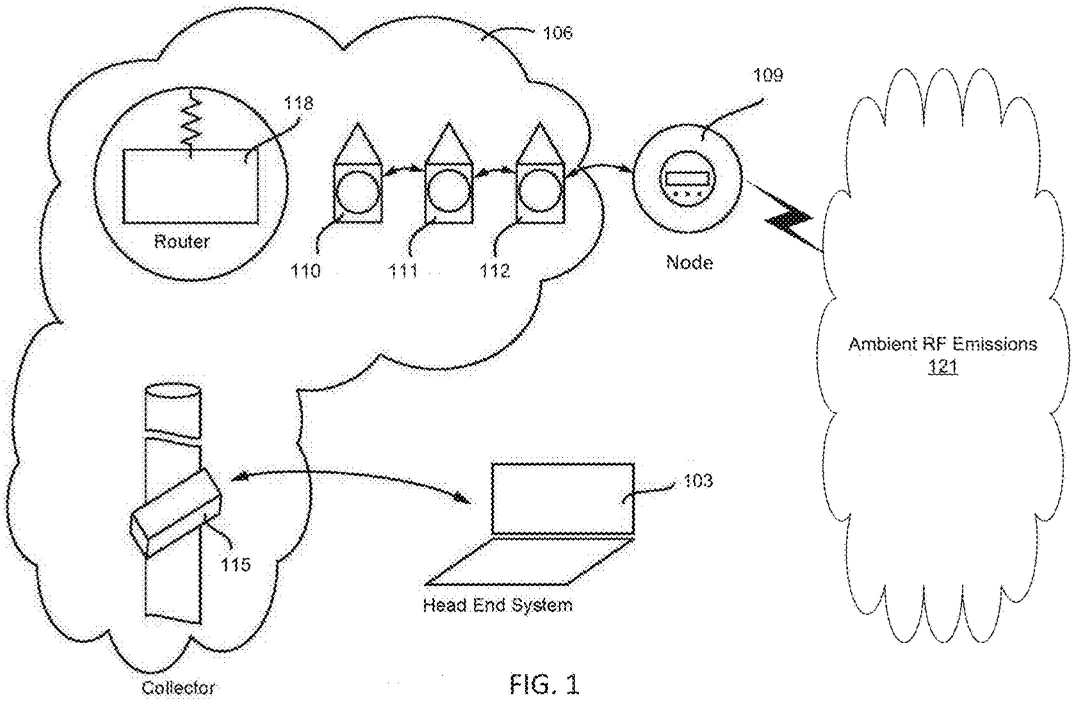
Router
Node
Ambient RF Emissions
Collector
Head End System

FIG. 1

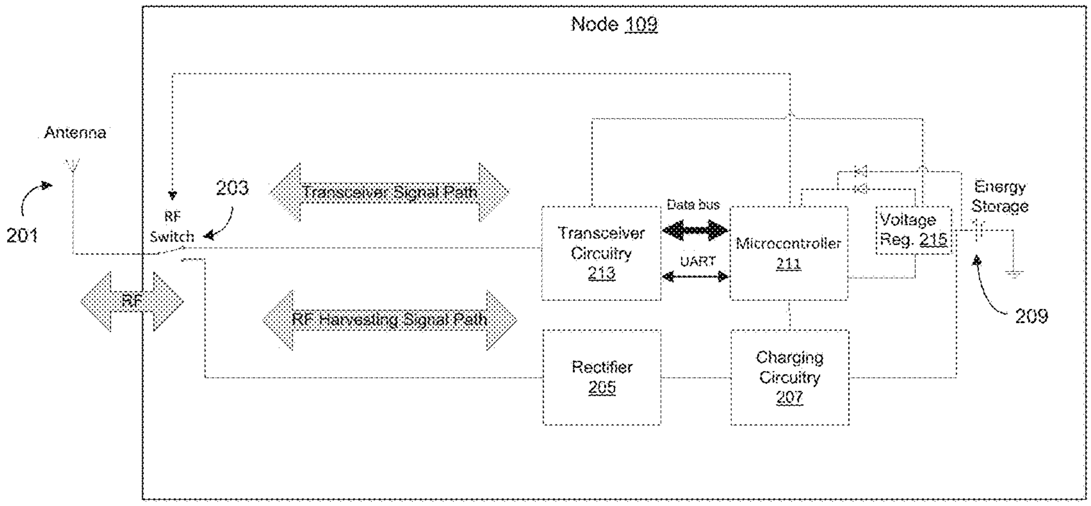
Node 109
Antenna
201
203
RF
Switch
Transceiver Signal Path
RF
RF Harvesting Signal Path
Data bus
UART
Transceiver
Circuitry
213
Microcontroller
211
Voltage
Reg. 215
Energy
Storage
209
Rectifier
205
Charging
Circuitry
207

FIG. 2

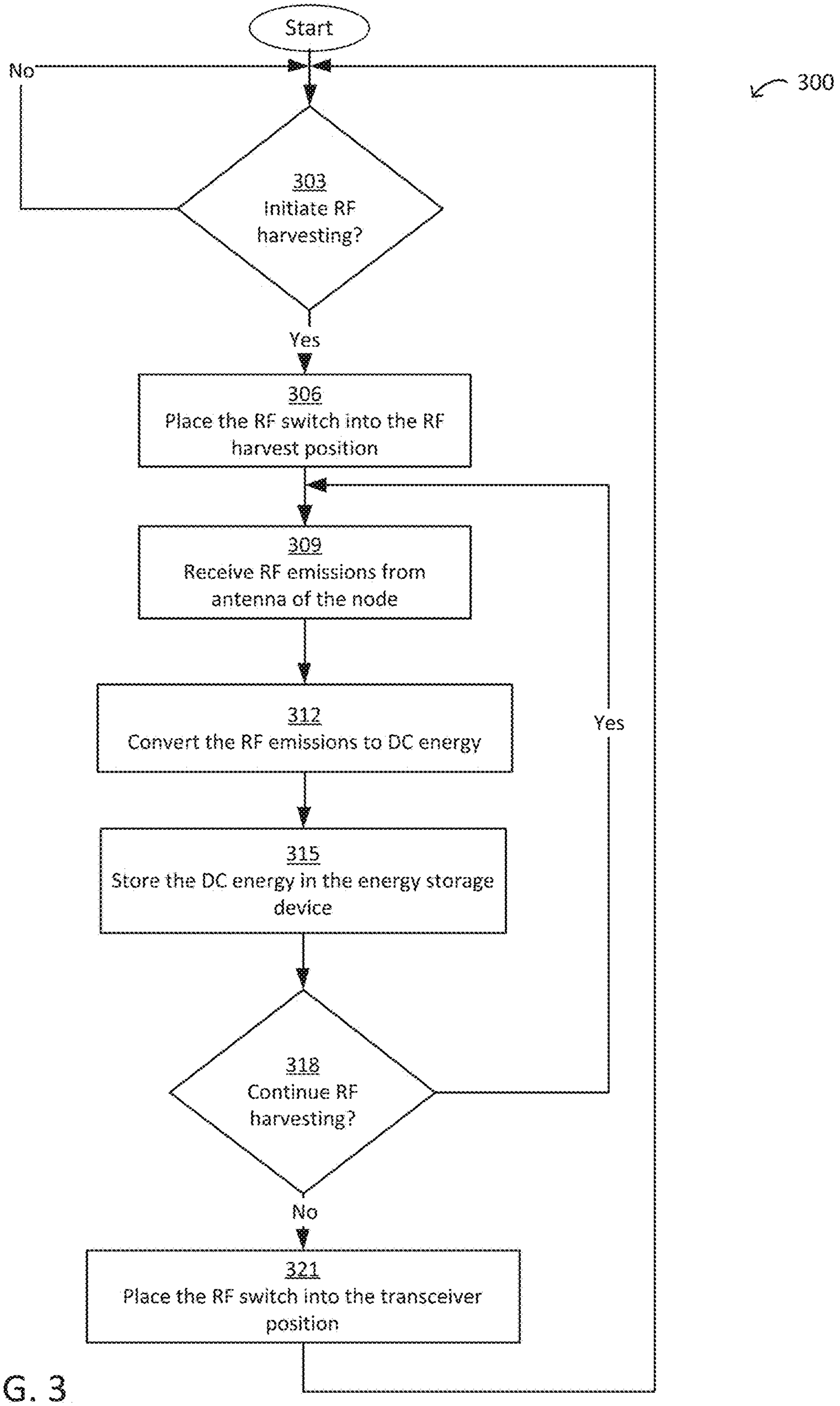
Start
No
300
303
Initiate RF harvesting?
Yes
306
Place the RF switch into the RF harvest position
309
Receive RF emissions from antenna of the node
312
Convert the RF emissions to DC energy
Yes
315
Store the DC energy in the energy storage device
318
Continue RF harvesting?
No
321
Place the RF switch into the transceiver position

FIG. 3

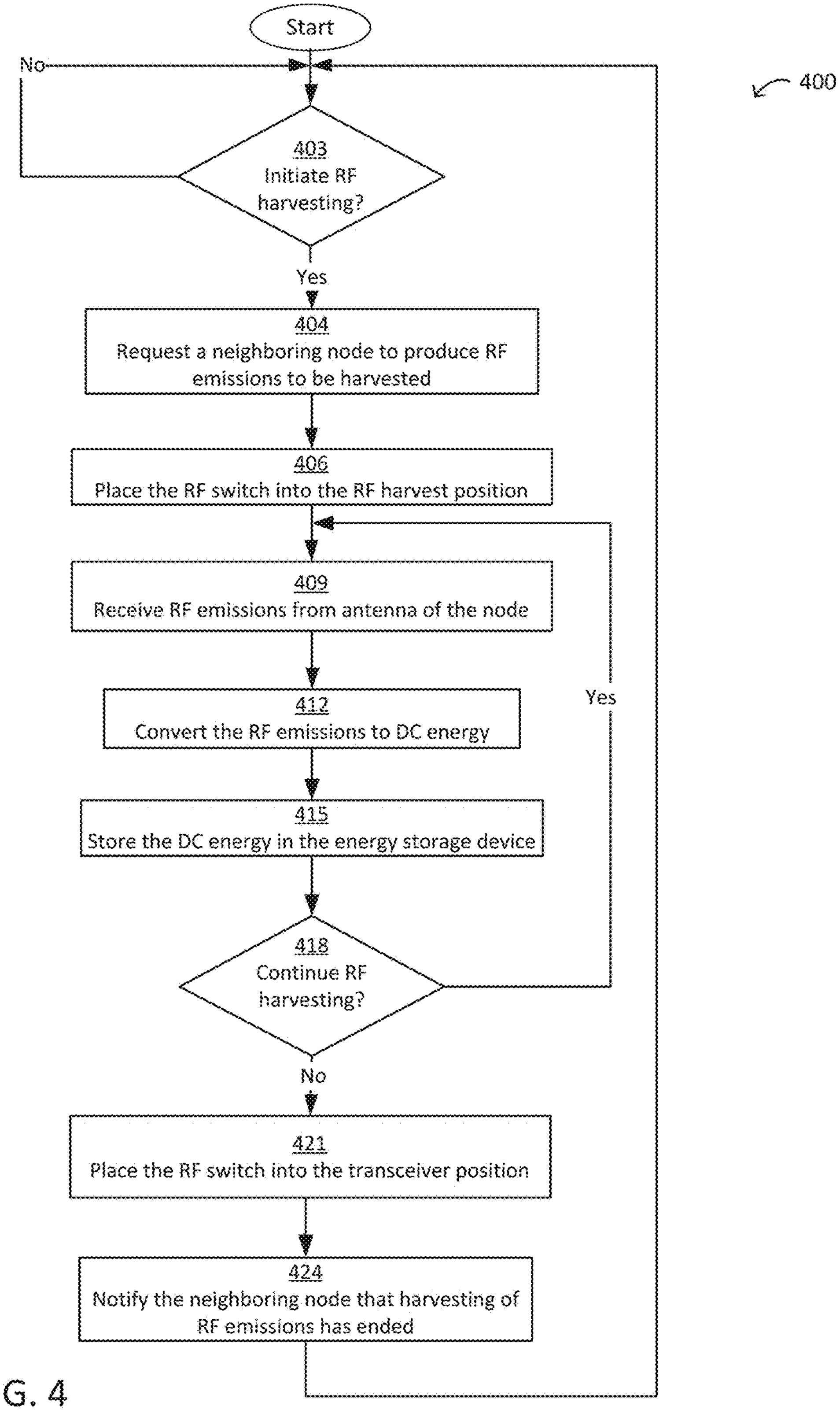
Start
No
400
403
Initiate RF harvesting?
Yes
404
Request a neighboring node to produce RF emissions to be harvested
406
Place the RF switch into the RF harvest position
409
Receive RF emissions from antenna of the node
412
Convert the RF emissions to DC energy
415
Store the DC energy in the energy storage device
Yes
418
Continue RF harvesting?
No
421
Place the RF switch into the transceiver position
424
Notify the neighboring node that harvesting of RF emissions has ended

FIG. 4

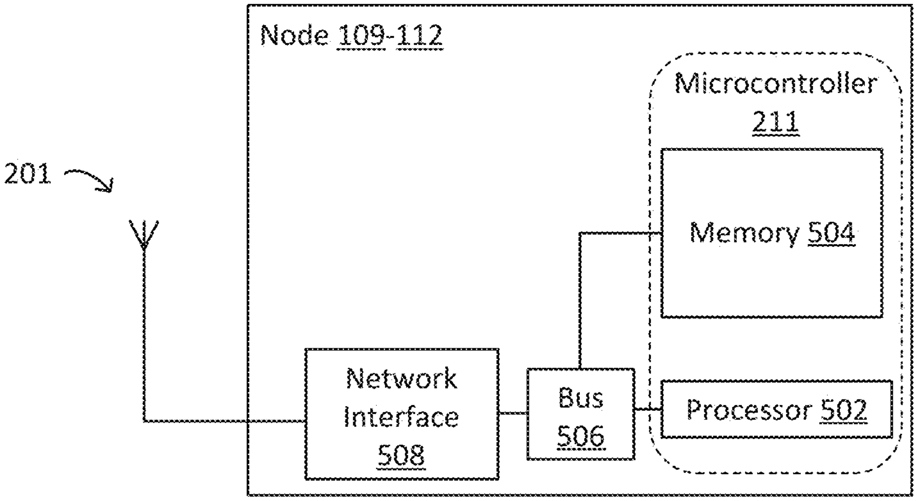
Node 109-112
Microcontroller 211
201
Memory 504
Network Interface 508
Bus 506
Processor 502

FIG. 5

RF ENERGY HARVESTING BY A NETWORK NODE

BACKGROUND

A communications module for a gas or water meter (or any other battery-powered utility device) must be able to support a battery life of 15-20 years. This is primarily achieved through the use of efficient energy-storage batteries as well as by minimizing the power consumption of the communications module (radio). In order to minimize power consumption, the radio may spend most of its time in a low-power "sleep" mode, waking up on some periodic basis to communicate any data with its network and then returning to sleep.

Replacing exhausted batteries in numerous such devices deployed in operation is a time-consuming and difficult task. Increasing the usable life-span of the batteries in these metering devices can produce significant savings in time and money for utility operators.

SUMMARY

Various aspects of the present invention relate to capturing ambient radio frequency (RF) emissions in order to supplement the energy stored by a node, such as a utility meter that may be deployed as part of an advanced metering infrastructure (AMI) system. In one implementation, the node operates in a communication mode that includes communicating, via an RF transceiver, data with another node in the network. The node includes an energy storage device, such as a battery, that provides power for the node. In various embodiments, the node may not be connected to an external source of power.

The node uses an RF switch in order to switch from the communication mode to an energy harvest mode. The mode switch may occur as a result of the node also entering a low-energy sleep mode and/or having no additional data to be communicated via the network. The energy harvest mode includes receiving RF emissions accessible to the node via an antenna and does not attempt to decode the RF emissions. The node converts the RF emissions to direct current (DC) energy and stores the DC energy in the energy storage device. The RF emissions may be converted to DC energy using a rectifier. The node may switch back to the communication mode based in part upon having data to be transmitted, data scheduled to be received, and/or other possible circumstances.

BRIEF DESCRIPTION OF THE DRAWINGS

Many aspects of the present disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, with emphasis instead being placed upon clearly illustrating the principles of the disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. **1** is a drawing of a networked environment according to various embodiments of the present disclosure.

FIG. **2** is a block diagram of an illustrative example of a node in the networked environment of FIG. **1** according to various embodiments of the present disclosure.

FIG. **3** is a flowchart illustrating one example of RF harvesting functionality implemented in a node in the networked environment of FIG. **1** according to various embodiments of the present disclosure.

FIG. **4** is a flowchart illustrating another example of RF harvesting functionality implemented in a node in the networked environment of FIG. **1** according to various embodiments of the present disclosure.

FIG. **5** is a block diagram that provides another example illustration of a node employed in the networked environment of FIG. **1** according to various embodiments of the present disclosure.

DETAILED DESCRIPTION

The techniques disclosed herein are directed to a node capturing ambient RF emissions in order to supplement the energy provided to the node by an energy storage device. The node may use an antenna to capture the energy from RF emissions from various commercially used RF bands, such as cellular, broadcast television, and unlicensed industrial, scientific and medical (ISM) bands (900 MHz and 2.4 GHz). For example, research studies have shown an ambient RF energy density of between −60 and −14 dB/m^2 in typical urban environments, all within the 685 MHz to 3.5 GHz frequency range. In these studies, the most energy was found to be in the 1.8-1.9 GHz band.

A node, such as used in a meter for utility service (e.g. electricity, water or gas) as part of an AMI (advanced metering infrastructure) network, can incorporate one or more wideband antennas tuned for all or a portion of these frequency bands. The wideband antenna would collect or "harvest" the RF energy, which could then be converted to a DC electrical energy by, for example, an RF/microwave rectifier circuit. The DC output of the rectifier would then be stored in an energy storage device of the node, such as a battery or supercapacitor, via a charging circuit. The energy stored in the energy storage device would then provide DC power for the node during its operation. While the amount of energy harvested may not be equivalent to the energy consumed by the node during all stages of operation, the energy may be used to supplement the energy stored in the energy storage device.

As defined herein, a "node" includes an intelligent device capable of performing functions related to distributing messages in a mesh network. In one system, a node can be a meter located at a facility, such as a house or apartment, that measures the consumption of a utility such as gas, water, or electric power. Such a meter can be part of an RF network used for AMI. Other examples of nodes include a router, collector or collection point, host computer, hub, or other electronic device that is attached to a network and is capable of sending, receiving, or forwarding information over a communications channel.

A node can contain several components that enable it to function within implementations of the present invention. For example, a node can include a radio that can enable it to communicate with like nodes and/or other devices in the mesh network. The radio of each node may have a programmable logic controller (PLC)-like device that can enable the radio to function like a computer, carrying out computer and command functions to provide implementations of the present invention described herein. A node may also include a storage medium for storing information related to communication with other nodes. Such storage mediums can include a memory, a floppy disk, CD-ROM, DVD, or other storage devices located internal to the node or accessible by the node via a network, for example. A node may also include a crystal oscillator (i.e. a clock) to provide timekeeping and an energy storage device (e.g. a battery) to provide back-up power. Some nodes may be powered only by an energy storage device.

With reference to FIG. **1**, shown is an exemplary illustration of a mesh network that may be employed as part of an AMI or other metering network. The head-end system **103** controls the metering network by communicating through the network **106**. The network **106** that the head-end system **103** may utilize to communicate to a node **109** may include devices such as other nodes, **110-112**, collectors **115**, routers **118**, and possibly other devices. Communication may proceed utilizing any appropriate protocol and any appropriate network configuration. Protocols used by the network **106** include, but are not limited to the IEEE 802.15.4, PRIME, G3, and/or TCP/IP protocols. Several nodes **109-112** may transmit data to a router **118**. The router **118**, in turn, may route data to a collector **115** in the network **106**. A collector may receive data from multiple routers. The collector **115** communicates with the head-end system **103**. The head-end system may receive and send information to multiple collectors.

The nodes, such as the node **109**, may be meters that are usually in geographically dispersed locations such as homes or businesses. The meters are used to monitor a utility resource such as electricity, water, or natural gas and to measure the usage of the resource. Some meters may be smart meters that support a variety of service commands. These service commands may allow utilities to disconnect, or limit service remotely or manually at the meter. In addition, some meters may store an event log that contains entries of functions the meter has performed. Service commands may originate from the head-end system **103** and are sent via the network **106** to nodes **109-112**.

To support functionality and continuous operation of a meter, the node **109** may include an energy storage device either as a supplemental/backup power source or as a sole power source. In order to reduce the electrical power consumed during operation the node **109**, may use an antenna to capture the energy from ambient RF emissions **121** from various commercially used RF bands, such as cellular, broadcast television, and ISM bands. While harvesting RF emissions in this manner has the benefit of reducing electrical power consumed from other sources, it can be especially critical when the node is reliant upon an energy storage device as the only power source for an extended period.

In some embodiments, the node **109** may begin harvesting energy from the ambient RF emissions **121** during periods in which no communications are expected to occur. During such periods, a portion of the components of the node **109** may enter a low-energy "sleep" mode in order to conserve electrical power. In these embodiments, an RF switch may be used to switch the one or more antennas of the node from the transceiver circuitry to the circuitry used for RF harvesting.

In other embodiments, a node may harvest energy from the ambient RF emissions **121** independent of communications of node through the use of antenna(s) and other circuitry that is separate from the communications circuitry. In these embodiments, the RF harvesting may occur continuously or when a threshold is met associated with the energy storage device of the node (e.g. the energy level is less than 50% of capacity).

In addition, a node may communicate with other devices of the network **106** in order to request these devices to begin transmitting RF emissions in order to augment the ambient RF emissions **121** received by the node. Such a scenario would typically occur with, for example, a battery-powered node requesting that another device powered from an external source provide additional RF emissions. A request for additional RF emissions may occur, for example, when the energy storage device of the node reaches a threshold, when the ambient RF emissions **121** are insufficient, and/or based upon other criteria.

Referring next to FIG. **2**, shown is a block diagram of an illustrative embodiment of the node **109** which implements the RF harvesting techniques. The node **109** includes one or more antennas, represented by the antenna **201**, which may be used to transmit and receive communications for the node **109**, as well as capture RF emissions to be stored as energy. The design of the antenna **201** may be affected by the RF band(s) at which the node communicates in a network **106** and the RF band(s) at which the node will capture RF emissions. For example, the antenna **201** may include an antenna for the ISM 900 MHz bands, in addition to another antenna used for the 1.8-1.9 GHz RF bands.

In this illustrative embodiment of FIG. **2**, the antenna **201** is connected to an RF switch **203** that connects the antenna **201** to either the transceiver signal path or the RF harvesting signal path. As described previously, other arrangements allowing, for example, simultaneous RF harvesting and data communication may also be possible. In FIG. **2**, the transceiver signal path includes the circuitry used for communicating data to and from the node **109**, while the RF harvesting signal path includes the circuitry used to charge the energy storage device using the captured RF emissions. A node may be said to operate in a "communication mode" while the transceiver signal path is selected via the RF switch, and in an "RF harvest mode" while the RF harvesting signal path is selected.

The RF harvesting signal path comprises a rectifier **205**, charging circuitry **207**, an energy storage device **209**, and/or other components not described in detail herein. The rectifier **205** is an RF/microwave rectifier that converts the energy from the captured RF emissions as input to a DC electrical energy output. The rectifier may be implemented using, for example, diodes.

The charging circuitry **207** may adjust the voltage or other components of the DC output from the rectifier **205** to a form suitable for storing in the energy storage device **209**. The energy storage device **209** represents one or more energy storage devices such as a battery, capacitor, supercapacitor, and/or other electrical energy storage devices. In the embodiment shown in FIG. **2**, the energy storage device **209** is the only source of power for the node **109**, though other configurations that use an external power source in conjunction with a local energy storage device are possible.

A microcontroller **211** is responsible for managing the operation and data communications of the node **109**. To this end, the microcontroller **211** manages the RF switch **203** and the transceiver circuitry **213** that facilitates transmission and reception of data along the transceiver signal path. In some embodiments, the microcontroller **211** may also monitor charging of the energy storage device **209** via the charging circuitry **207**. The various components of the node **109** may receive electrical power from the energy storage device **209**. The electrical power from the energy storage device **209** may be adjusted by the voltage regulator **215** in order to accommodate the electrical needs of disparate electrical components. For example, the voltage regulator **215** may adjust the 3.0 V output from the energy storage device to the 3.45 V required by the microcontroller **211**.

Referring next to FIG. **3**, shown is a flowchart that provides an exemplary implementation of the RF harvesting operations for a method **300** of a node, such as the node **109**, according to various embodiments. It is understood that the flowchart of FIG. **3** provides merely an example of the many different types of functional arrangements that may be employed to implement the RF harvesting operations of the method **300** as described herein. The operations depicted in the flowchart of FIG. **3** illustrate the transition of the node from operating in a communication mode to operating in a harvest mode. As described previously, variations of this implementation may also be possible allowing a node to communicate via a network, while concurrently capturing energy from ambient RF emissions.

Beginning with block **303**, the node determines whether to begin the RF harvest mode or to remain in the communication mode. The determination may be made based upon the current or expected data communications with other devices in a network, a threshold in the energy storage device of the node, and/or other criteria as can be appreciated. For example, if the node had no current data communication taking place, nor any that were expected, some components (e.g. the transceiver circuitry) of the node might enter a low-power "sleep mode," at which time the node may also enter a harvest mode to capture RF emissions. Similarly, if the node detected that the energy level of the energy storage device was, for example, below a threshold of 50%, the node may begin RF harvesting instead of taking place in non-essential communications.

If the node determines that RF harvesting is not ready to begin, then execution of the method **300** returns to block **303**. Alternatively, if RF harvesting should be initiated, in block **306**, the node may place the RF switch in a position to provide received RF emission to the RF harvesting signal path. Thereafter, in block **309**, the node may receive RF emissions via the one or more antennas of the node. The node may use the antenna(s) to capture the energy from RF emissions from various commercially used RF bands, such as cellular, broadcast television, ISM bands (900 MHz and 2.4 GHz), and/or other possibilities. For example, research studies have shown an ambient RF energy density of between −60 and −14 dB/m2 in typical urban environments, all within the 685 MHz to 3.5 GHz frequency range. In these studies, the most energy was found to be in the 1.8-1.9 GHz band.

Next, in block **312**, the rectifier of the node converts the energy from the various received RF emissions to DC electrical energy. Then, in block **315**, the DC energy output from the rectifier may be stored in the energy storage device of the node through use of charging circuitry. The charging circuitry may be used to monitor and adjust the characteristics of the DC energy used to charge the energy storage device, monitor the current capacity of the energy storage device, and/or other possible re-charging functions. In some embodiments, some or all of the DC energy output from the rectifier may be used to directly power, at least in part, one or more other components of the node such as, for example, the microcontroller. During periods of low power usage, such as during the sleep mode, the energy collected from RF emissions may be sufficient to provide a substantial portion of the power consumed by various components (e.g. the microcontroller).

Subsequently, in block **318**, the node determines whether to continue the RF harvesting (i.e. harvest mode). The determination may be made based upon the current or expected data communications with other devices in a network, a threshold in the energy storage device of the node, and/or other criteria as can be appreciated. For example, if the node detected that the energy level of the energy storage device was at a threshold of 100%, the node may no longer need to continue to harvest the energy since no more capacity exists to store it. Similarly, if the node is scheduled to soon report a utility meter reading to a head-end system, the node may end the harvest mode in order to deliver the scheduled communication data.

In the event that the RF harvesting is determined to continue, execution of the method **300** returns to block **309**. Alternatively, if it is determined that the current period of RF harvesting should end, in block **321**, the node places the RF switch in a position that connects the antenna to the transceiver signal path in order to enable the node to communicate with other devices in the network. Thereafter, execution of the method **300** may return to block **303**.

Moving on to FIG. **4**, shown is a flowchart that provides another exemplary implementation of the RF harvesting operations for a method **400** of a node, such as the node **109**, according to various embodiments. It is understood that the flowchart of FIG. **4** provides merely an example of the many different types of functional arrangements that may be employed to implement the RF harvesting operations of the method **400** as described herein. The operations depicted in the flowchart of FIG. **4** illustrate the transition of the node from operating in a communication mode to operating in a harvest mode, where the node requests a neighboring node to supplement the ambient RF emissions by generating additional RF emissions to be harvested by the node. As described previously, variations of this implementation may also be possible allowing a node to communicate via network, while concurrently capturing energy from ambient RF emissions.

Beginning with block **403**, the node determines whether to begin the RF harvest mode or to remain in the communication mode. The determination may be made based upon the current or expected data communications with other devices in a network, a threshold in the energy storage device of the node, and/or other criteria as can be appreciated.

If the node determines that RF harvesting is not ready to begin, then execution of the method **400** returns to block **403**. Alternatively, if RF harvesting should be initiated, in block **404**, the node may communicate with one or more neighboring nodes in the network to request that the node(s) produce additional RF emissions in order to supplement the energy received during the harvest mode. Typically, a node would make a request to neighboring nodes known to have an external power source (i.e. power supplied through the electric utility grid) or that otherwise have a substantial reserve of energy available.

Then, in block **406**, the node may place the RF switch in a position to provide received RF emission to the RF harvesting signal path. Thereafter, in block **409**, the node may receive RF emissions via the one or more antennas of the node. The neighboring node(s) responding to the request for addition RF emissions may produce a continuous waveform (CW) of RF energy that may be received by the requesting node, as well as other nodes in the area. The node may use the antenna(s) to capture the energy from RF emissions from various commercially used RF bands, such as cellular, broadcast television, ISM bands (900 MHz and 2.4 GHz), and/or other possibilities.

Next, in block **412**, the rectifier of the node converts the energy from the various received RF emissions to DC electrical energy. Then, in block **415**, the DC energy output from the rectifier may be stored in the energy storage device of the node through use of charging circuitry. The charging circuitry may be used to monitor and adjust the characteristics of the DC energy used to charge the energy storage device, monitor the current capacity of the energy storage device, and/or other possible re-charging functions.

Subsequently, in block **418**, the node determines whether to continue the RF harvesting (i.e. harvest mode). The determination may be made based upon the current or expected data communications with other devices in a network, a threshold in the energy storage device of the node, and/or other criteria as can be appreciated. For example, if the node detected that the energy level of the energy storage device was at a threshold of 100%, the node may no longer need to continue to harvest the energy since no more capacity exists to store it. Similarly, if the node is scheduled to soon report a utility meter reading to a headend system, the node may end the harvest mode in order to deliver the scheduled communication data.

In the event that the RF harvesting is determined to continue, execution of the method **400** returns to block **409**. Alternatively, if it is determined that the current period of RF harvesting should end, in block **421**, the node places the RF switch in a position that connects the antenna to the transceiver signal path in order to enable the node to communicate with other devices in the network. As an optional step, in block **424**, once the node has returned to the communication mode, the node may transmit a notification to the neighboring nodes indicating that the node is no longer in harvest mode and that the transmission of additional RF emission is no longer needed by the node. In some implementations, the neighboring nodes producing additional RF emissions as requested in block **404** may do so for a predefined period of time. Thereafter, execution of the method **400** may return to block **403**.

Next, in FIG. **5**, shown is another block diagram depicting an example of a node **109-112** used for implementing the techniques disclosed herein. The node **109-112** can include a processing device **502**. Non-limiting examples of the processing device **502** include a microprocessor, an application-specific integrated circuit ("ASIC"), a state machine, or other suitable processing device. The processing device **502** can include any number of processing devices, including one. The processing device **502** can be communicatively coupled to computer-readable media, such as memory device **504**. The processing device **502** can execute computer-executable program instructions and/or access information respectively stored in the memory device **504**. In some implementations, the processor **502** and memory **504** may be incorporated as components of a microcontroller, such as the microcontroller **211**.

The memory device **504** can store instructions that, when executed by the processing device **502**, cause the processing device **502** to perform operations described herein. The memory device **504** may be a computer-readable medium such as (but not limited to) an electronic, optical, magnetic, or other storage device capable of providing a processor with computer-readable instructions. Non-limiting examples of such optical, magnetic, or other storage devices include read-only ("ROM") device(s), random-access memory ("RAM") device(s), magnetic disk(s), magnetic tape(s) or other magnetic storage, memory chip(s), an ASIC, configured processor(s), optical storage device(s), or any other medium from which a computer processor can read instructions. The instructions may comprise processor-specific instructions generated by a compiler and/or an interpreter from code written in any suitable computer-programming language. Non-limiting examples of suitable computer-programming languages include C, C++, C#, Visual Basic, Java, Python, Perl, JavaScript, and the like.

The nodes **109-112** can include a bus **506** that can communicatively couple one or more components of the node **109-112**. Although the processor **502**, the memory **504**, and the bus **506** are depicted in FIG. **5** as separate components in communication with one another, other implementations are possible. For example, the processor **502**, the memory **504**, and the bus **506** can be components of printed circuit boards or other suitable devices that can be disposed in a node **109-112** to store and execute programming code.

The nodes **109-112** can also include network interface device **508**, such as may be implemented by the transceiver circuitry **213**. The network interface device **508** can be a transceiving device configured to establish one or more of the wireless communication links via an antenna **201**. A non-limiting example of the network interface device **508** is an RF transceiver and can include one or more components for establishing a communication links to other nodes **109-112** in the mesh network **106**.

Numerous specific details are set forth herein to provide a thorough understanding of the claimed subject matter. However, those skilled in the art will understand that the claimed subject matter may be practiced without these specific details. In other instances, methods, apparatuses, or systems that would be known by one of ordinary skill have not been described in detail so as not to obscure claimed subject matter.

Some portions are presented in terms of algorithms or symbolic representations of operations on data bits or binary digital signals stored within a computing system memory, such as a computer memory. These algorithmic descriptions or representations are examples of techniques used by those of ordinary skill in the data processing arts to convey the substance of their work to others skilled in the art. An algorithm is a self-consistent sequence of operations or similar processing leading to a desired result. In this context, operations or processing involves physical manipulation of physical quantities. Typically, although not necessarily, such quantities may take the form of electrical or magnetic signals capable of being stored, transferred, combined, compared or otherwise manipulated. It has proven convenient at times, principally for reasons of common usage, to refer to such signals as bits, data, values, elements, symbols, characters, terms, numbers, numerals, or the like. It should be understood, however, that all of these and similar terms are to be associated with appropriate physical quantities and are merely convenient labels. Unless specifically stated otherwise, it is appreciated that throughout this specification discussions utilizing terms such as "processing," "computing," "calculating," "determining," and "identifying" or the like refer to actions or processes of a computing device, such as one or more computers or a similar electronic computing device or devices, that manipulate or transform data represented as physical electronic or magnetic quantities within memories, registers, or other storage devices, transmission devices, or display devices of the computing platform.

The system or systems discussed herein are not limited to any particular hardware architecture or configuration. A computing device can include any suitable arrangement of components that provide a result conditioned on one or more function calls. Suitable computing devices include multipurpose microprocessor-based computer systems accessing stored software that programs or configures the computing system from a general-purpose computing apparatus to a specialized computing apparatus implementing one or more aspects of the present subject matter. Any suitable programming, scripting, or other type of language or combinations of languages may be used to implement the teachings contained herein in software to be used in programming or configuring a computing device.

Aspects of the methods disclosed herein may be performed in the operation of such computing devices. The order of the blocks presented in the examples above can be varied—for example, blocks can be re-ordered, combined, and/or broken into sub-blocks. Certain blocks or processes can be performed in parallel.

The use of “adapted to” or “configured to” herein is meant as open and inclusive language that does not foreclose devices adapted to or configured to perform additional tasks or steps. Additionally, the use of “based on” is meant to be open and inclusive, in that a process, step, calculation, or other action “based on” one or more recited conditions or values may, in practice, be based on additional conditions or values beyond those recited. Headings, lists, and numbering included herein are for ease of explanation only and are not meant to be limiting.

While the present subject matter has been described in detail with respect to specific aspects thereof, it will be appreciated that those skilled in the art, upon attaining an understanding of the foregoing, may readily produce alterations to, variations of, and equivalents to such aspects. Accordingly, it should be understood that the present disclosure has been presented for purposes of example rather than limitation, and does not preclude inclusion of such modifications, variations, and/or additions to the present subject matter as would be readily apparent to one of ordinary skill in the art.

Therefore, the following is claimed:

1. A method for energy harvesting by circuitry of a node in a network, comprising:

operating in a communication mode comprising:

communicating, via a radio frequency (RF) transceiver, data with another node in the network, wherein the node comprises an energy storage device that provides power for the node; and

in response to occurrence of a time period during which no further data communication is scheduled in the node, switching, via an RF switch in the node, from said communication mode to an energy harvest mode comprising:

receiving RF emissions accessible to the node via an antenna;

providing the received RF emissions to a rectifier;

converting the RF emissions to direct current (DC) energy, wherein said converting is implemented using the rectifier; and

storing the DC energy in the energy storage device.

2. The method of claim **1**, further comprising switching back to the communication mode when the node determines additional data needs to be transmitted or received by the node, the communication mode further comprising:

receiving additional RF emissions accessible to the node via the antenna; and

decoding the additional data from the additional RF emissions.

3. The method of claim **1**, wherein the node is a utility meter deployed as part of an advanced metering infrastructure (AMI) system.

4. The method of claim **1**, wherein the energy storage device is the only source of electrical power for the node during operation.

5. The method of claim **1**, wherein said switching to the energy harvesting mode is based at least in part upon the node determining that a sleep mode can begin.

6. The method of claim **1**, wherein said receiving RF emissions comprises receiving a plurality of RF signals in one or more RF bands.

7. A node, comprising:

a processor;

an antenna;

a radio frequency (RF) switch for switching connectivity to the antenna;

an energy storage device that provides power for the node;

a rectifier;

an RF transceiver for communicating on a network accessible to the node; and

a memory configured by a management application executed in the node, the management application comprising logic causing the node to:

operate in an energy harvest mode comprising:

receiving RF emissions accessible to the node via the antenna;

converting, via the rectifier, the RF emissions to direct current (DC) energy; and

storing the DC energy in the energy storage device; and

in response to occurrence of a time period during which data communication is scheduled in the node, switch, via the RF switch, from said energy harvest mode to a communication mode comprising:

communicating, via the RF transceiver, data with another node in the network, wherein said communicating uses power from the energy storage device.

8. The node of claim **7**, wherein said switch to the communication mode occurs when the node determines additional data needs to be transmitted or received by the node.

9. The node of claim **8**, wherein the additional data needs to be transmitted or received by the node on a periodic basis.

10. The node of claim **7**, wherein the management application further comprises logic to request the other node in the network to transmit additional RF emissions to the node.

11. The node of claim **7**, wherein the network is a wireless mesh network of a plurality nodes that includes the node.

12. The node of claim **7**, wherein the RF transceiver operates in one or more of the industrial, scientific and medical (ISM) RF bands.

13. The node of claim **12**, wherein said receiving RF emissions comprises receiving a plurality of RF signals in one or more RF bands, a portion of the one or more RF bands being outside the ISM RF bands.

14. A non-transitory computer-readable medium embodying a program executable by a processor of a node, comprising:

code for operating the node in a communication mode comprising:

communicating, via a radio frequency (RF) transceiver, data with another node in a network, wherein the node comprises an energy storage device that provides power for the node;

code for operating the node in an energy harvest mode comprising:

receiving RF emissions accessible to the node via an antenna;

providing the received RF emissions to a rectifier;

converting the RF emissions to direct current (DC) energy, wherein said converting is implemented using the rectifier; and

storing the DC energy in the energy storage device; and code for switching, via an RF switch in the node, between the modes based upon a determination made in the node, wherein the node switches to the energy harvesting mode in response to occurrence of a time period during which no further data communication is scheduled in the node.

15. The non-transitory computer-readable medium of claim **14**, wherein the node receives electrical power only from the energy storage device in the node.

16. The non-transitory computer-readable medium of claim **15**, wherein the program further comprises code for requesting the other node in the network to transmit additional RF emissions to the node, wherein the other node receives electrical power from an external source.

17. The non-transitory computer-readable medium of claim **14**, wherein the node is a utility meter deployed as part of an advanced metering infrastructure (AMI) system.

18. The non-transitory computer-readable medium of claim **14**, wherein the determination comprises switching to the energy harvest mode based upon a power level of the energy storage device reaching a threshold.

19. The non-transitory computer-readable medium of claim **14**, wherein the determination comprises switching to the energy harvest mode based upon a completion of a scheduled data communication.

* * * * *

US 20090152954A1

(19) **United States**

(12) **Patent Application Publication**

Le et al.

(10) **Pub. No.: US 2009/0152954 A1**

(43) **Pub. Date: Jun. 18, 2009**

(54) RF ENERGY HARVESTING CIRCUIT

(76) Inventors: **Triet Tu Le**, Portland, OR (US); **Terri S. Flez**, Corvallis, OR (US); **Kartikeya Mayaram**, Corvallis, OR (US)

Correspondence Address:
LUMEN PATENT FIRM
2345 YALE STREET, SECOND FLOOR
PALO ALTO, CA 94306 (US)

(21) Appl. No.: **12/218,761**

(22) Filed: **Jul. 17, 2008**

Related U.S. Application Data

(60) Provisional application No. 60/961,009, filed on Jul. 17, 2007, provisional application No. 60/993,260, filed on Sep. 10, 2007.

Publication Classification

(51) **Int. Cl.**
H02M 3/18 (2006.01)
H02J 17/00 (2006.01)

(52) **U.S. Cl.** **307/110**; 307/149

(57) ABSTRACT

A voltage doubler rectifier RF power harvesting system is provided having at least one power harvesting module with a voltage doubler rectifier that includes a DC voltage input and output, signal input, first and second floating gate transistors that are connected in series between the DC voltage input and output, and a gate control of the floating gate is connected to a drain of the transistor. The voltage double rectifier includes a first capacitor disposed between the input signal and the drain of the first floating gate transistor, and a second capacitor disposed between a ground and the drain of the second floating gate transistor. The system includes an input antenna and a transmitter with a transmitter antenna, where a full-wave peak-to-peak voltage of an incoming RF signal is rectified. The system also includes a powered device connected to the DC voltage output to utilize the rectified signal.

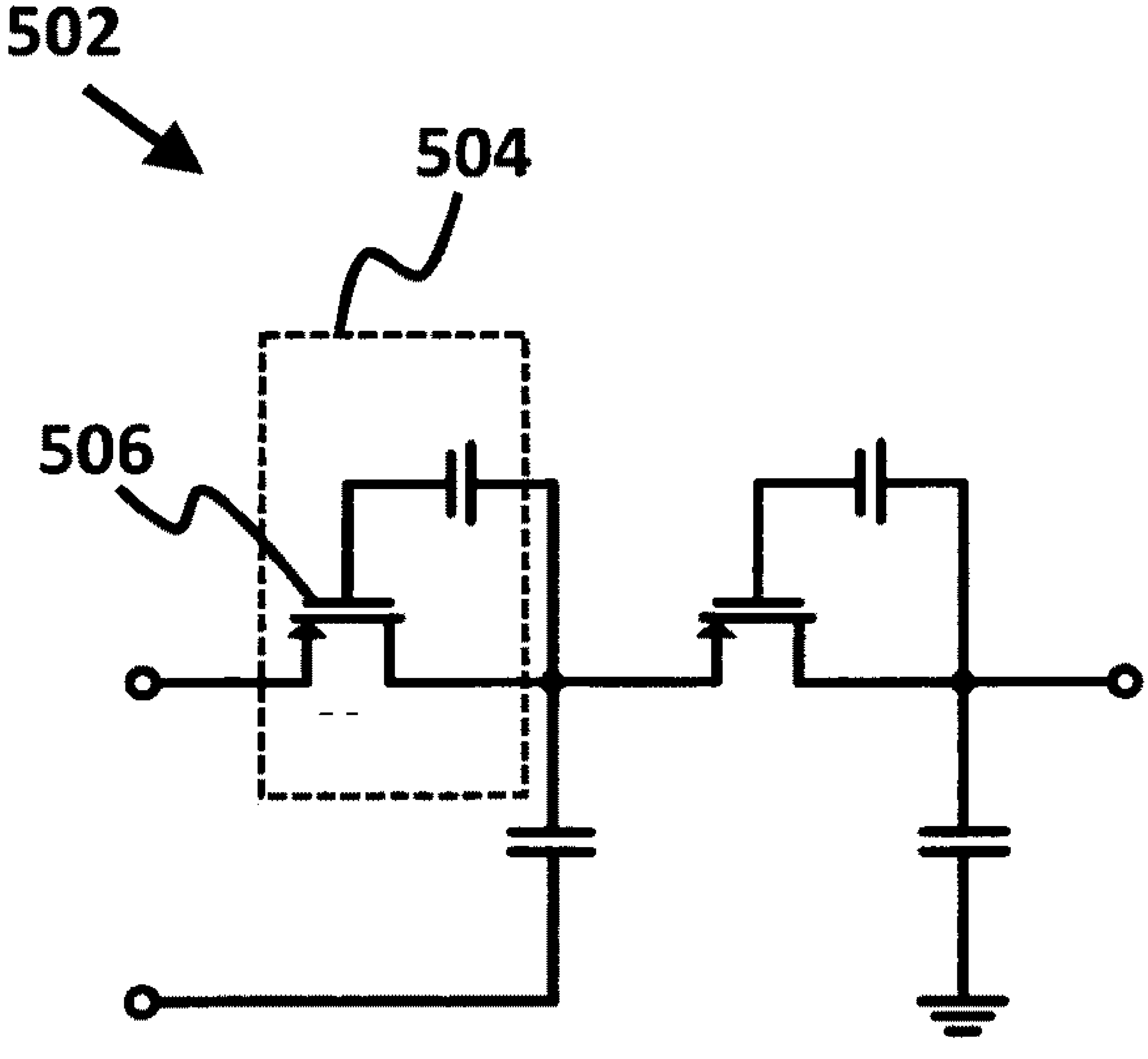

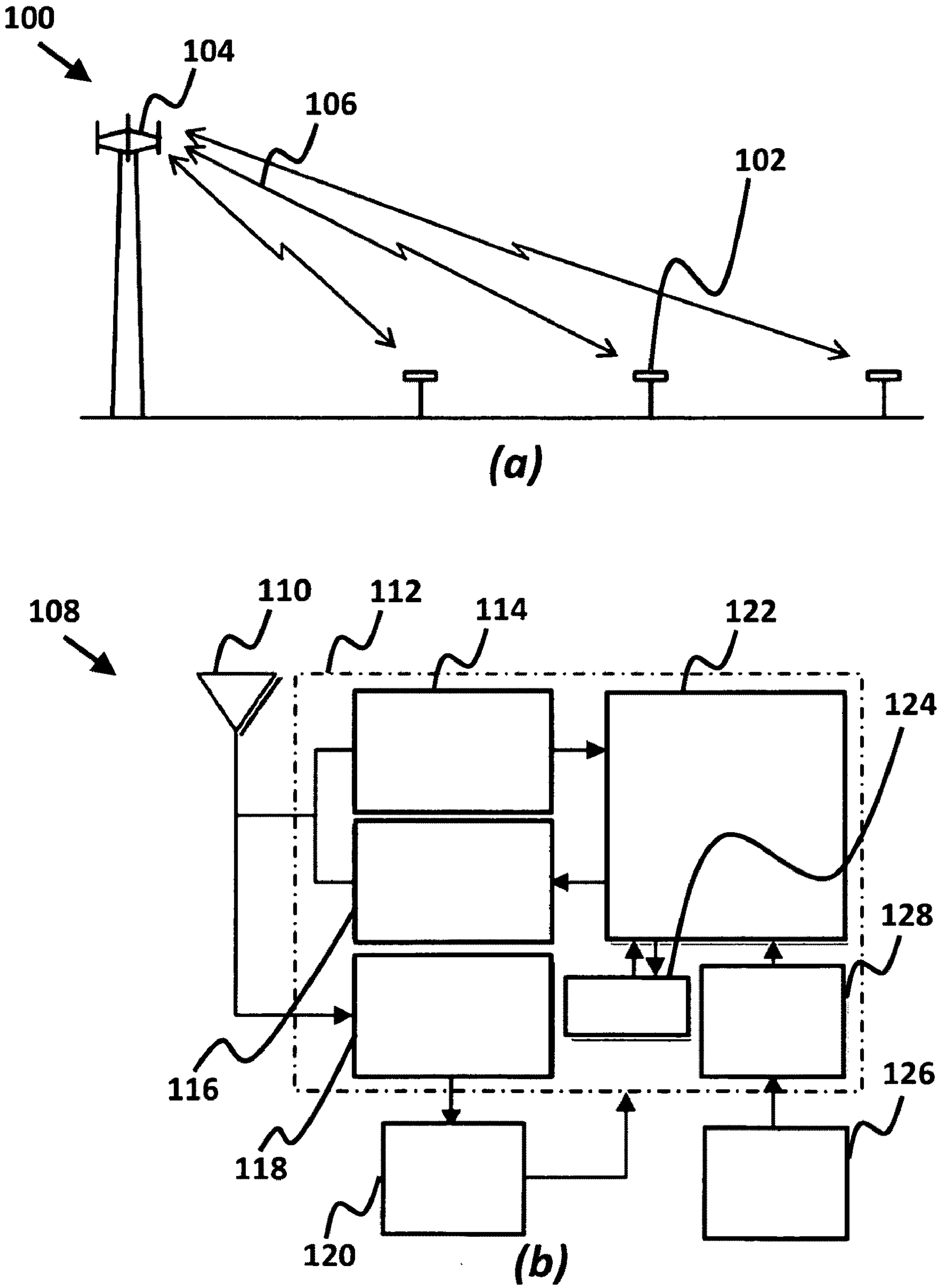
100
104
106
102
108
110
112
114
122
124
128
126
116
118
120

(a)

(b)

FIG. 1

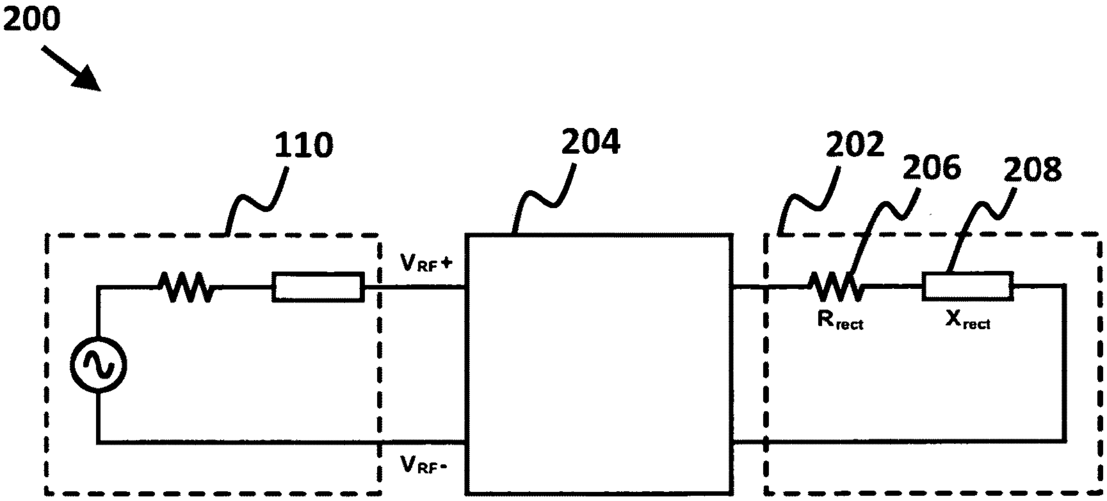
200
110
204
202
206
208
$V_{RF}+$
$V_{RF}-$
R_{rect}
X_{rect}

FIG. 2

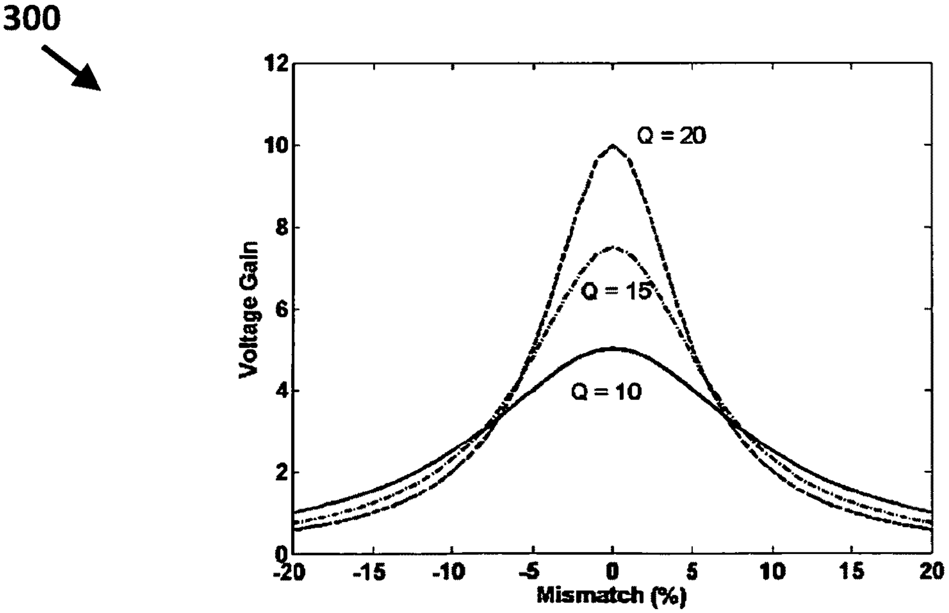
300
Q = 20
Q = 15
Q = 10
Voltage Gain
Mismatch (%)
0
2
4
6
8
10
12
-20
-15
-10
-5
0
5
10
15
20

FIG. 3

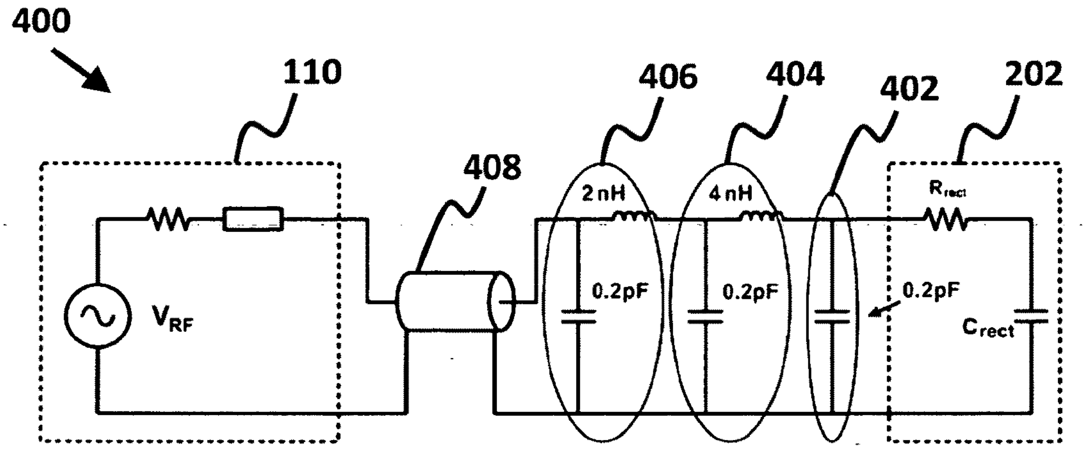
400
110
408
406
404
402
202
V_{RF}
2 nH
4 nH
0.2pF
0.2pF
0.2pF
R_{rect}
C_{rect}

FIG. 4

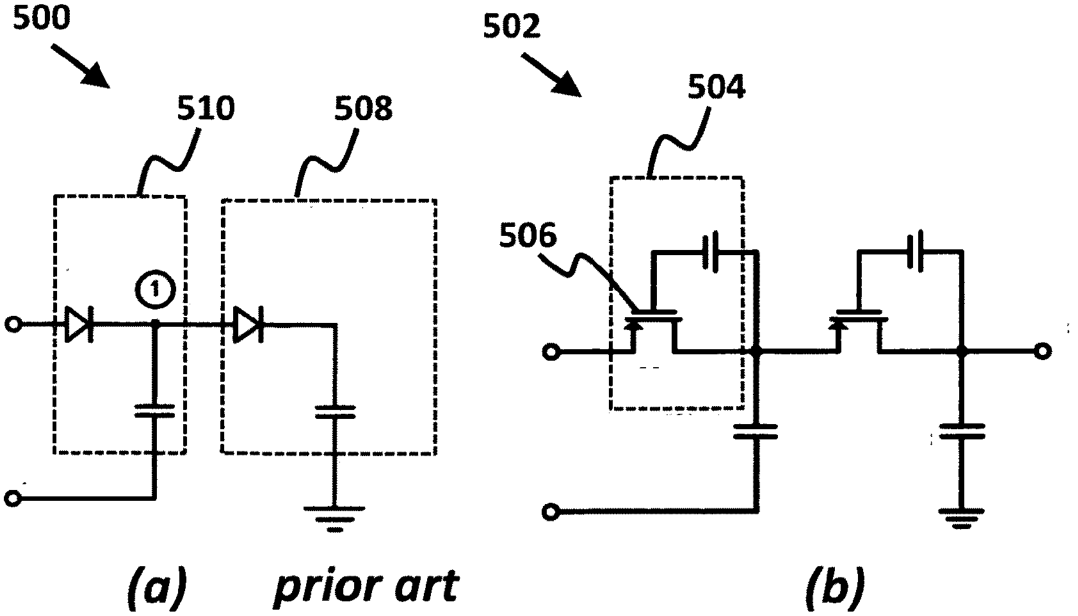
500
510
508
①
502
504
506

(a) prior art (b)

FIG. 5

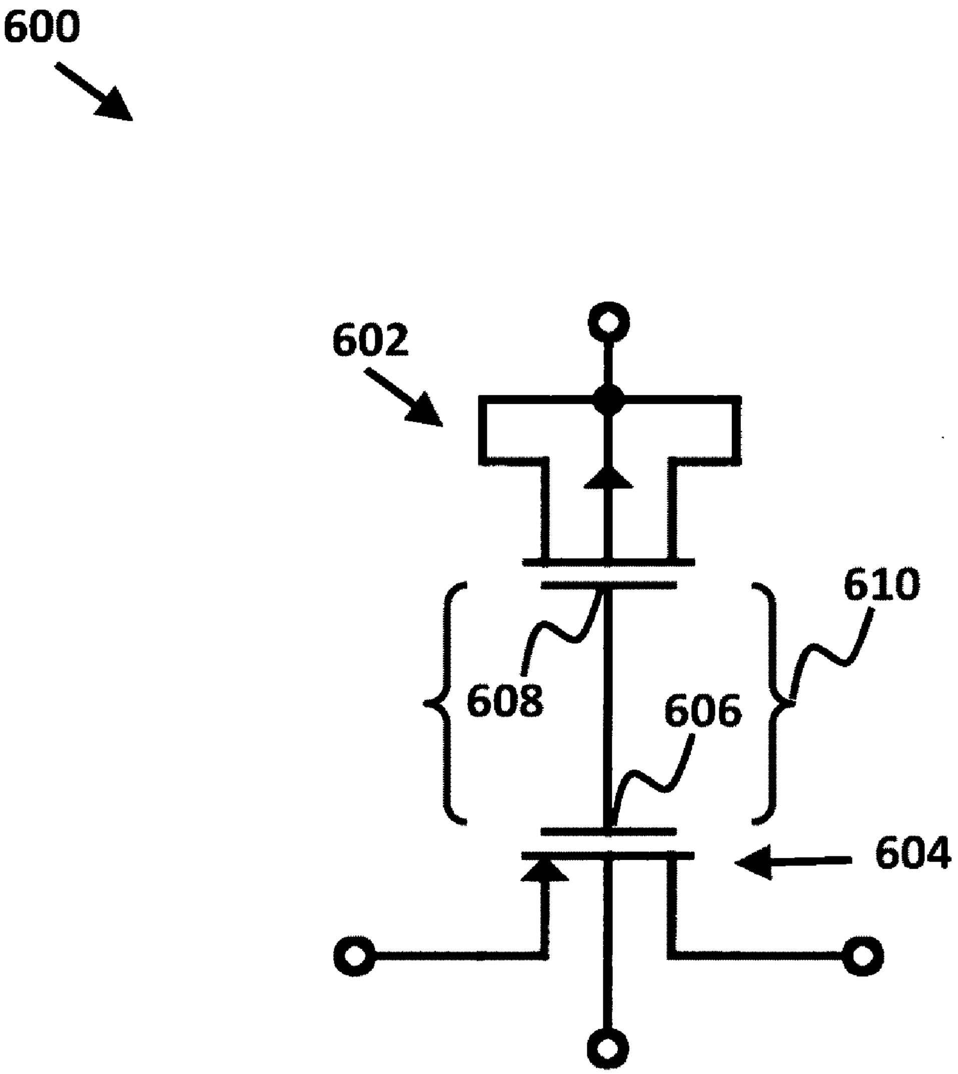
600
602
610
608
606
604

FIG. 6

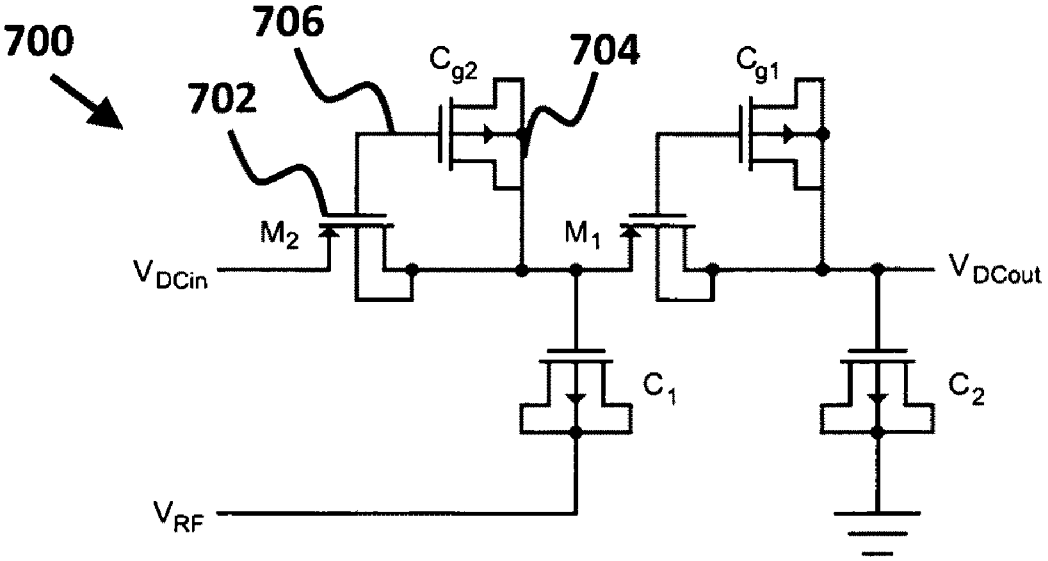
700
706
702
704
C_{g2}
C_{g1}
M_2
M_1
V_{DCin}
V_{DCout}
C_1
C_2
V_{RF}

FIG. 7

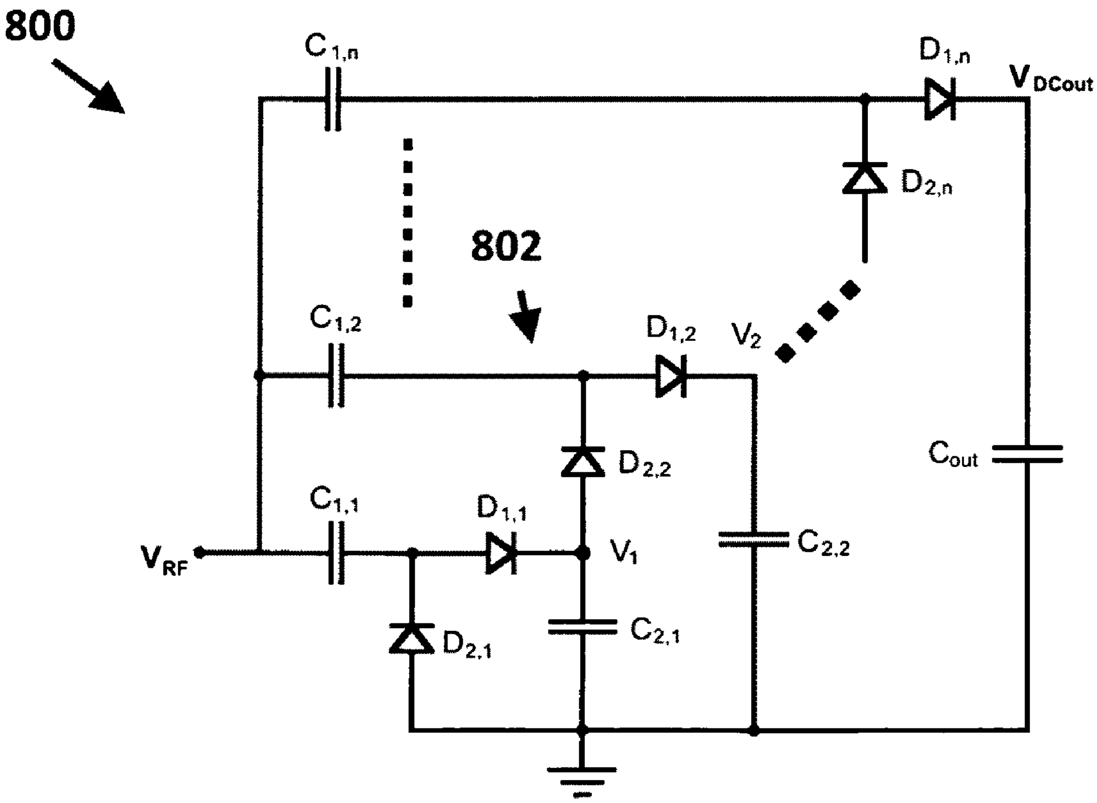
800
$C_{1,n}$
$D_{1,n}$
V_{DCout}
$D_{2,n}$
802
$C_{1,2}$
$D_{1,2}$
V_2
$D_{2,2}$
C_{out}
$C_{1,1}$
$D_{1,1}$
V_1
$C_{2,2}$
V_{RF}
$D_{2,1}$
$C_{2,1}$

FIG. 8

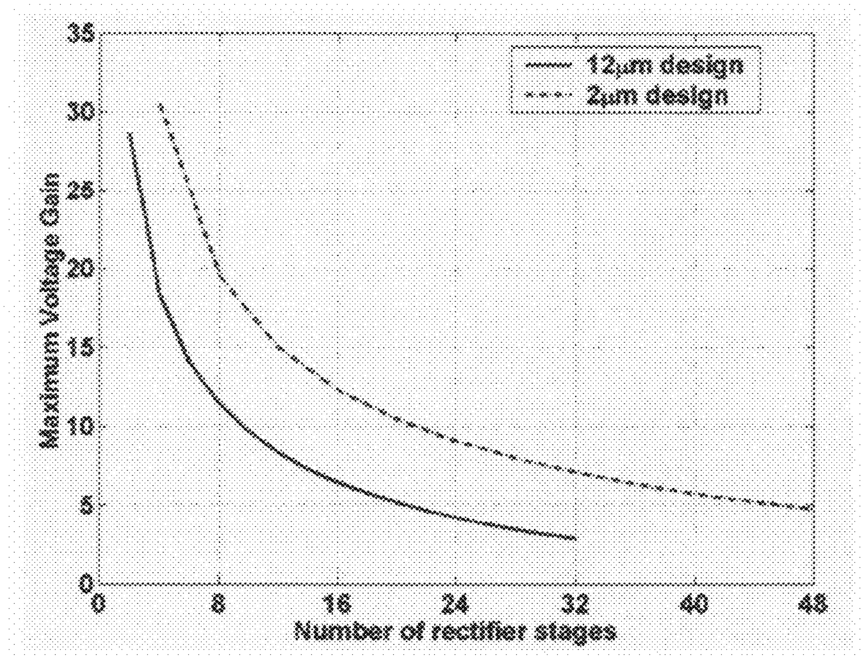
35
30
25
20
15
10
5
0
0
8
16
24
32
40
48
Maximum Voltage Gain
Number of rectifier stages
12μm design
2μm design

(a)

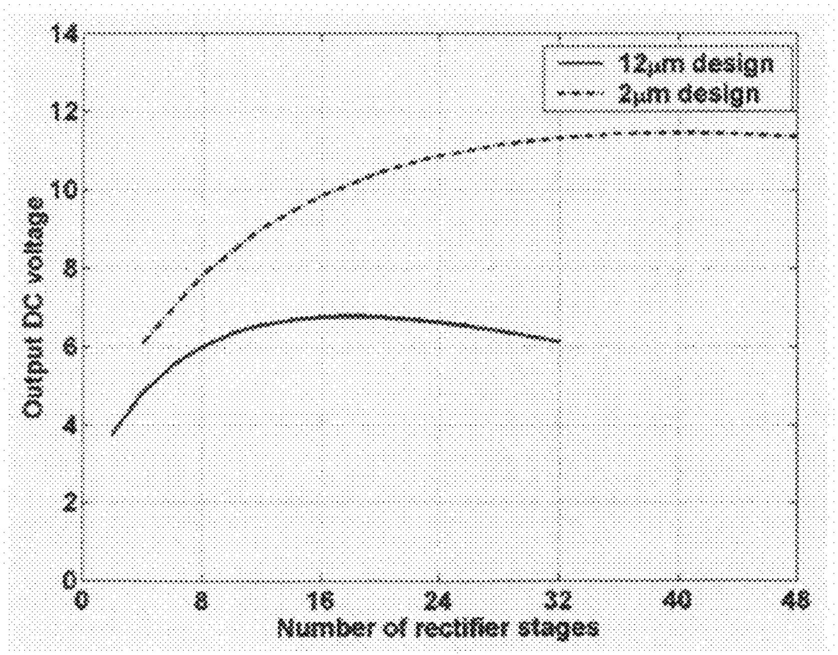
14
12
10
8
6
4
2
0
0
8
16
24
32
40
48
Output DC voltage
Number of rectifier stages
12μm design
2μm design

(b)

FIG. 9

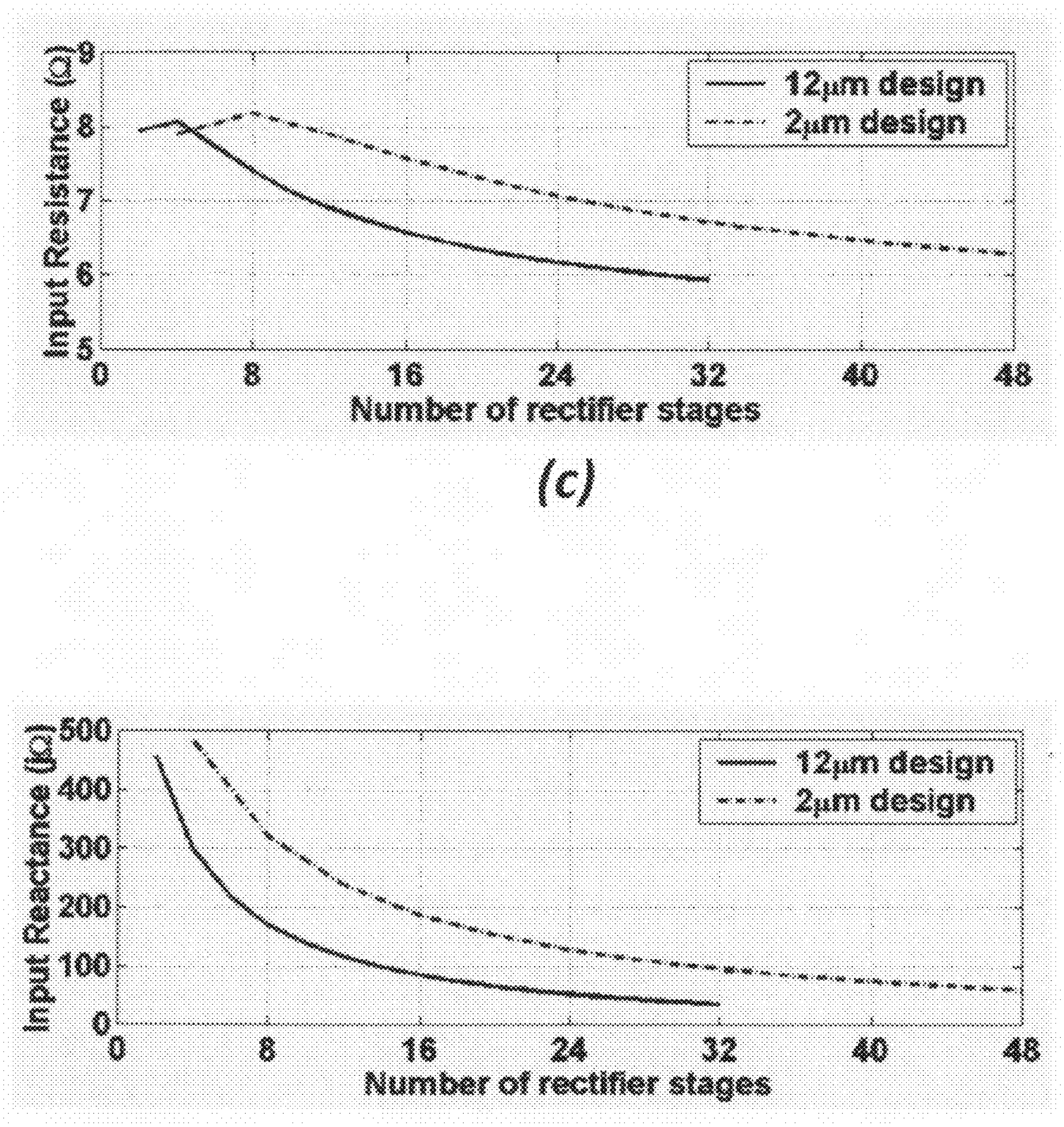
Input Resistance (Ω)
9
8
7
6
5
12μm design
2μm design
0
8
16
24
32
40
48
Number of rectifier stages
(c)
Input Reactance (jΩ)
500
400
300
200
100
0
12μm design
2μm design
0
8
16
24
32
40
48
Number of rectifier stages
(d)

FIG. 9 (cont.)

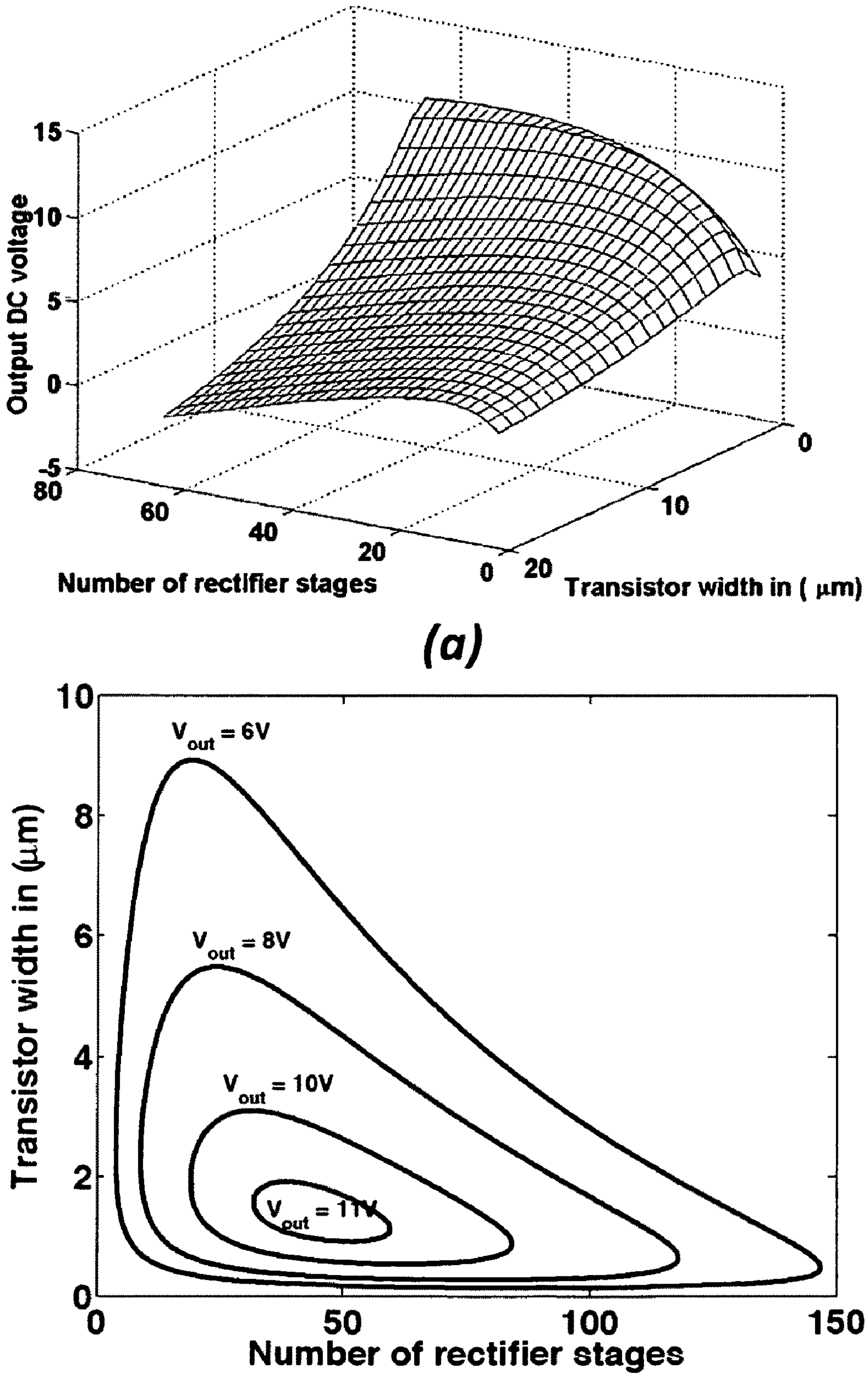
Output DC voltage
15
10
5
0
-5
80
60
40
20
0
Number of rectifier stages
20
10
0
Transistor width in (μm)
Transistor width in (μm)
10
8
6
4
2
0
V_{out} = 6V
V_{out} = 8V
V_{out} = 10V
V_{out} = 11V
0
50
100
150
Number of rectifier stages

(a)

(b)

FIG. 10

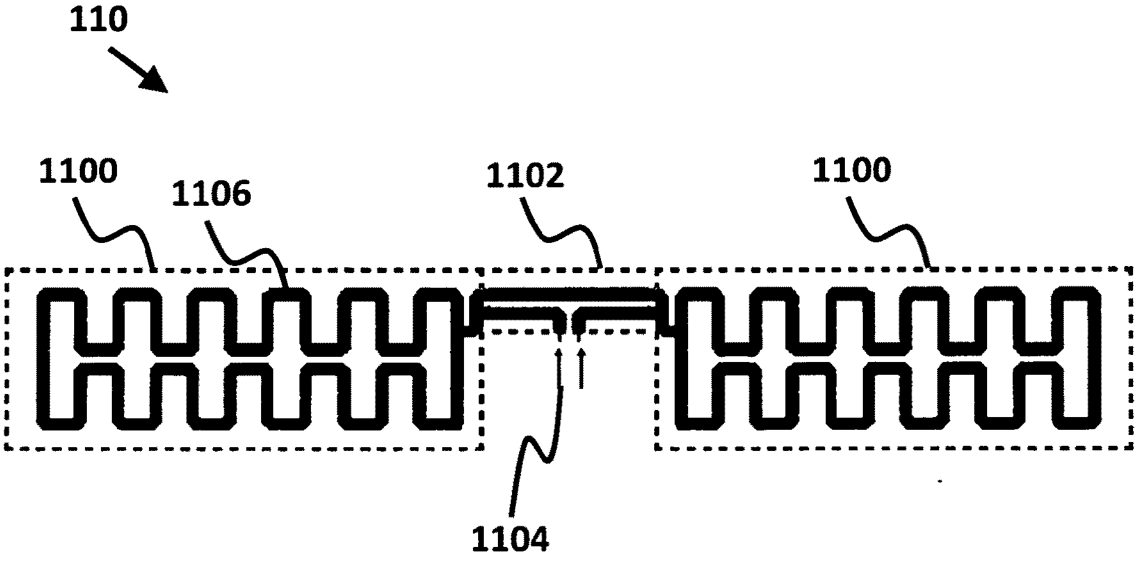
110
1100
1106
1102
1100
1104

FIG. 11

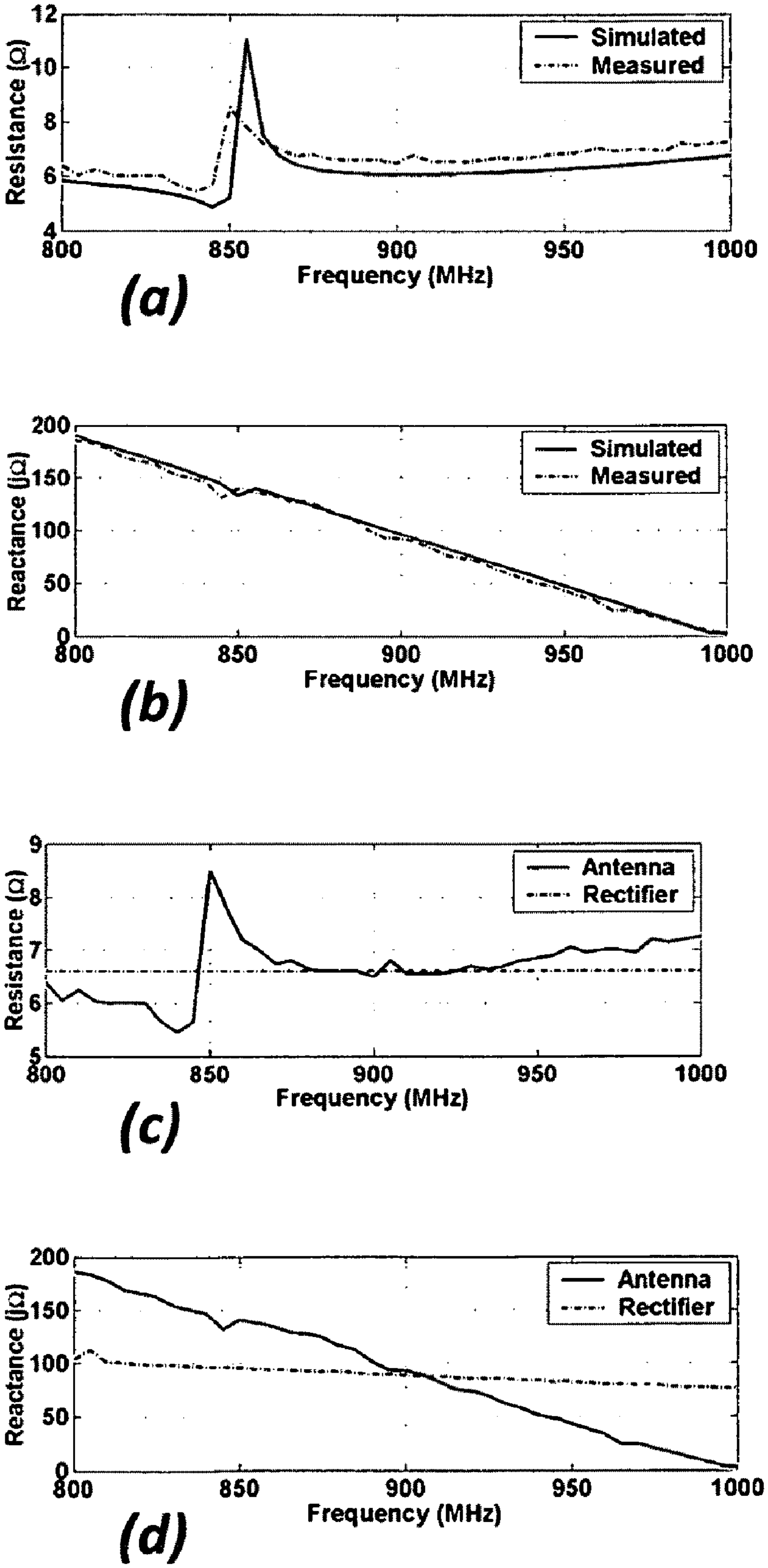
Simulated
Measured
Resistance (Ω)
12
10
8
6
4
800
850
900
950
1000
Frequency (MHz)
(a)
Simulated
Measured
Reactance (jΩ)
200
150
100
50
0
800
850
900
950
1000
Frequency (MHz)
(b)
Antenna
Rectifier
Resistance (Ω)
9
8
7
6
5
800
850
900
950
1000
Frequency (MHz)
(c)
Antenna
Rectifier
Reactance (jΩ)
200
150
100
50
0
800
850
900
950
1000
Frequency (MHz)
(d)

FIG. 12

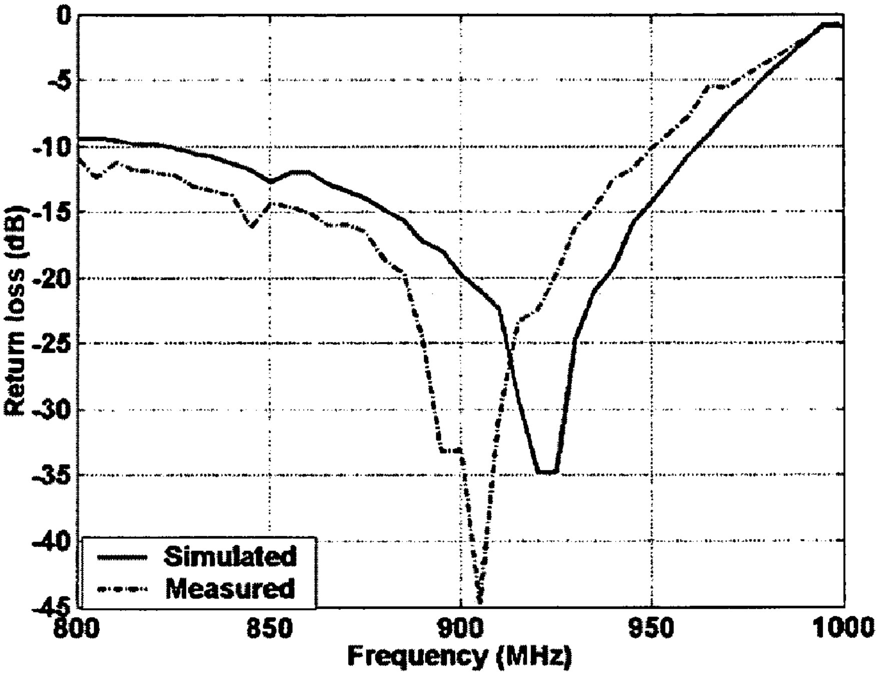
0
-5
-10
-15
-20
-25
-30
-35
-40
-45
Return loss (dB)
Simulated
Measured
800
850
900
950
1000
Frequency (MHz)

(a)

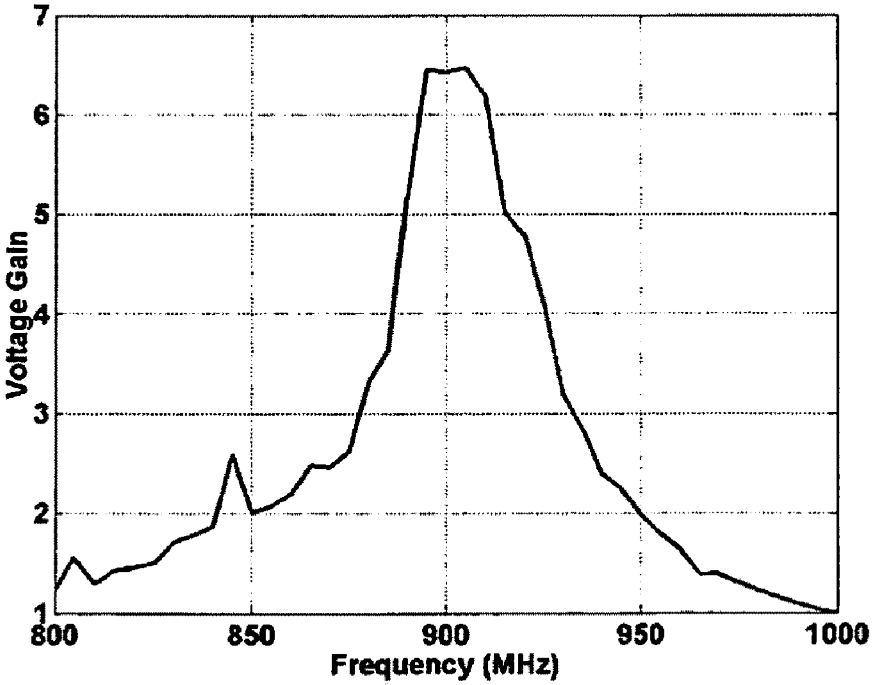
7
6
5
4
3
2
1
Voltage Gain
800
850
900
950
1000
Frequency (MHz)

(b)

FIG. 13

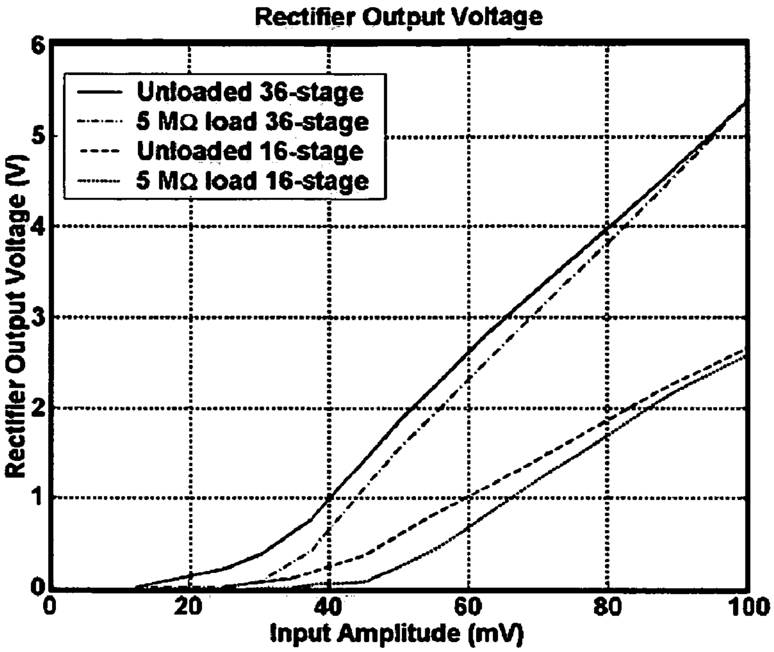
Rectifier Output Voltage
Unloaded 36-stage
5 MΩ load 36-stage
Unloaded 16-stage
5 MΩ load 16-stage
Rectifier Output Voltage (V)
0
1
2
3
4
5
6
0
20
40
60
80
100
Input Amplitude (mV)

(a)

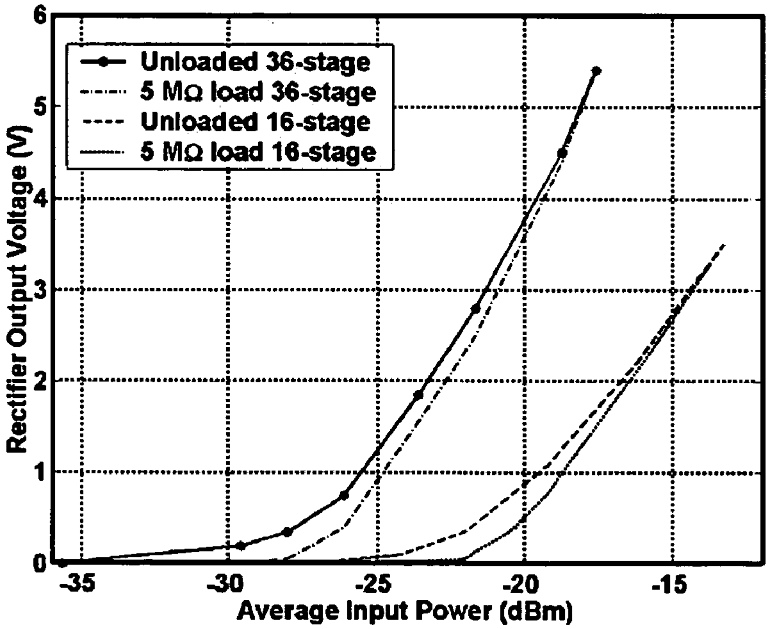
Unloaded 36-stage
5 MΩ load 36-stage
Unloaded 16-stage
5 MΩ load 16-stage
Rectifier Output Voltage (V)
0
1
2
3
4
5
6
-35
-30
-25
-20
-15
Average Input Power (dBm)

(b)

FIG. 14

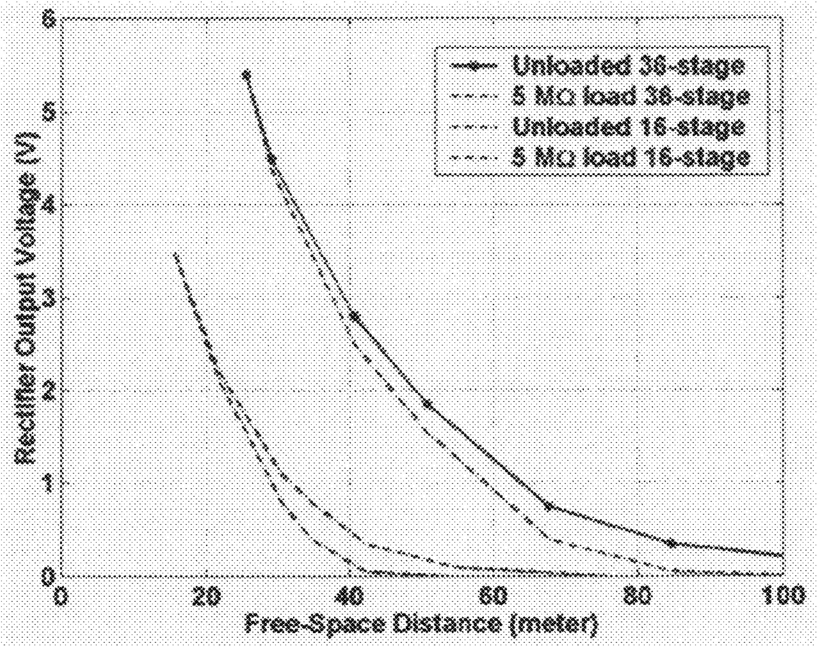

(c)

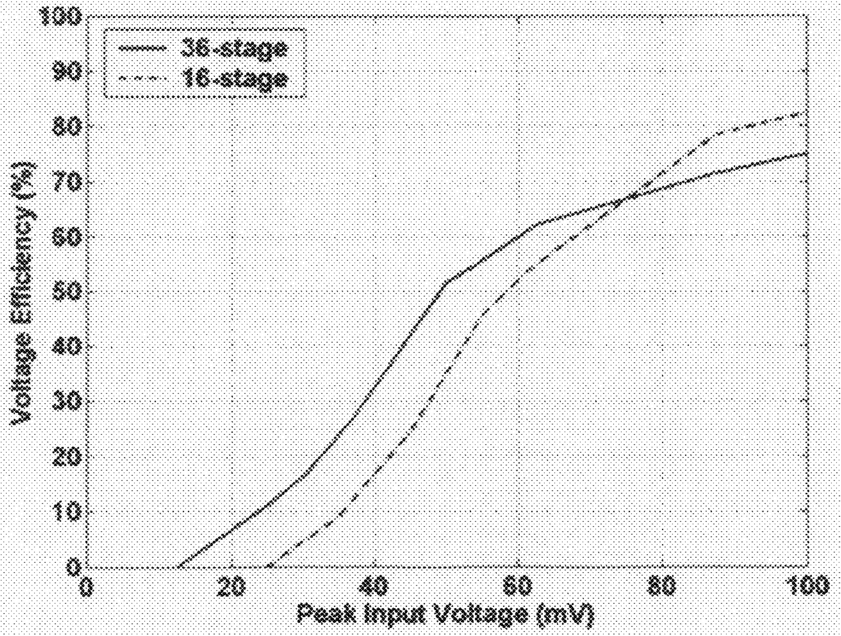

(d)

FIG. 14 (cont.)

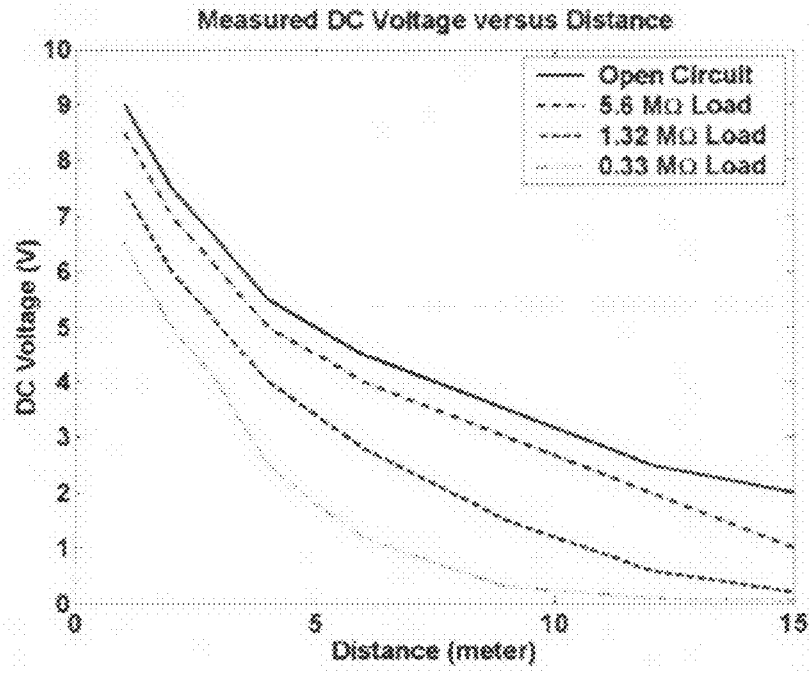
Measured DC Voltage versus Distance
Open Circuit
5.6 MΩ Load
1.32 MΩ Load
0.33 MΩ Load
DC Voltage (V)
10
9
8
7
6
5
4
3
2
1
0
0
5
10
15
Distance (meter)

(a)

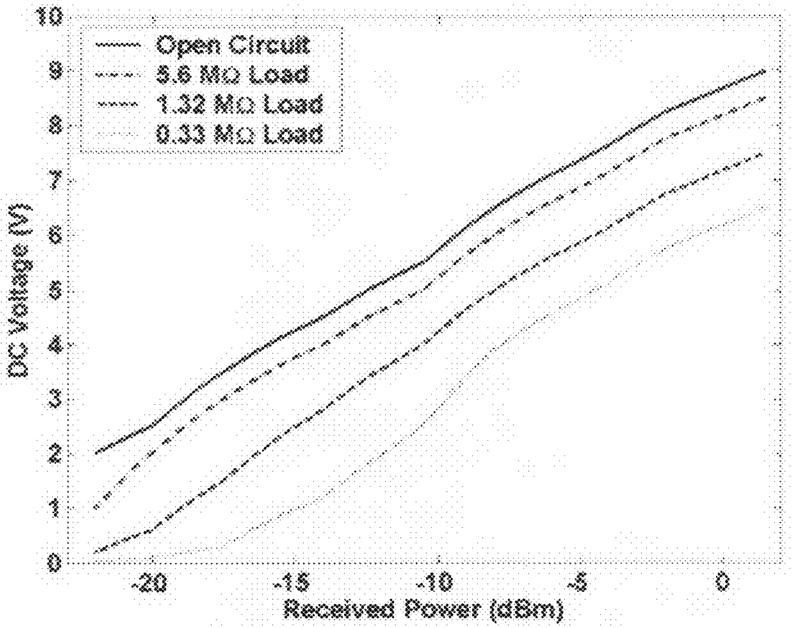
Open Circuit
5.6 MΩ Load
1.32 MΩ Load
0.33 MΩ Load
DC Voltage (V)
10
9
8
7
6
5
4
3
2
1
0
-20
-15
-10
-5
0
Received Power (dBm)

(b)

FIG. 15

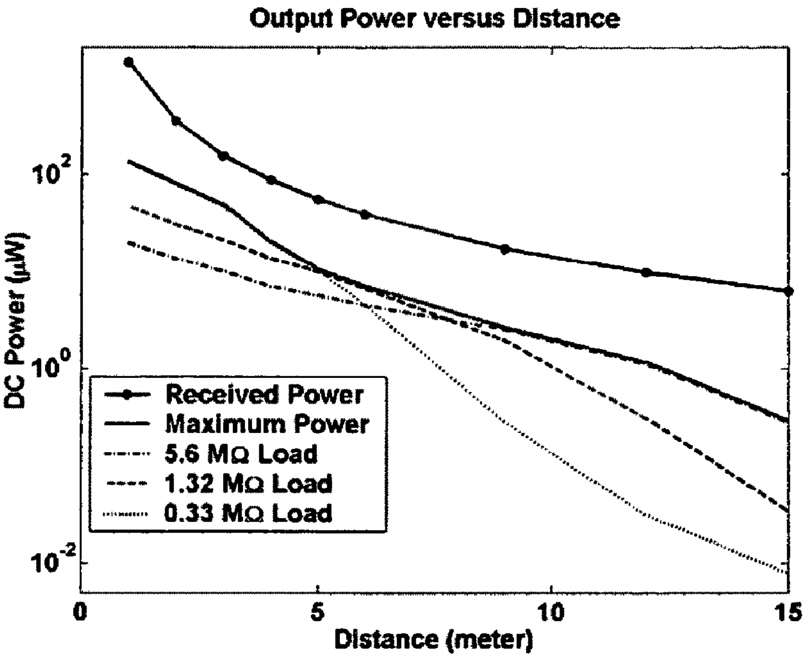

(c)

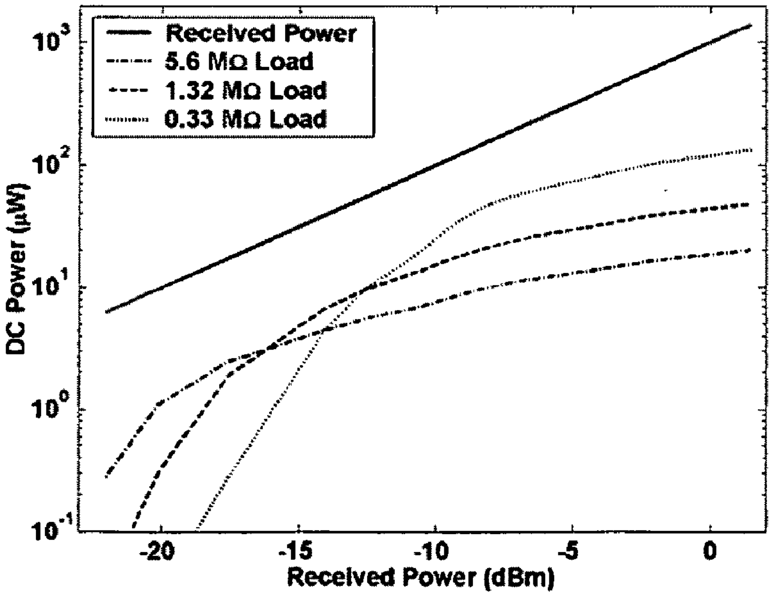

(d)

FIG. 15 (cont.)

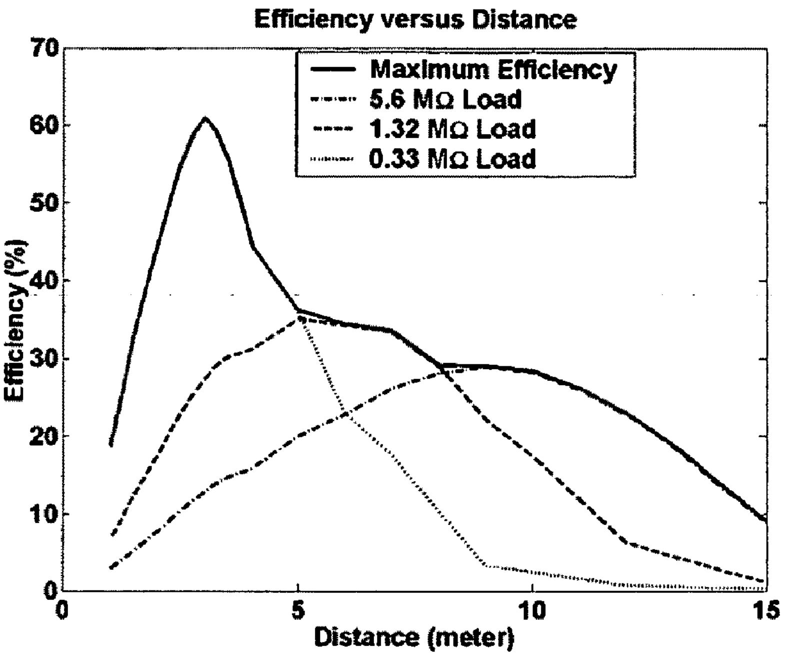

(e)

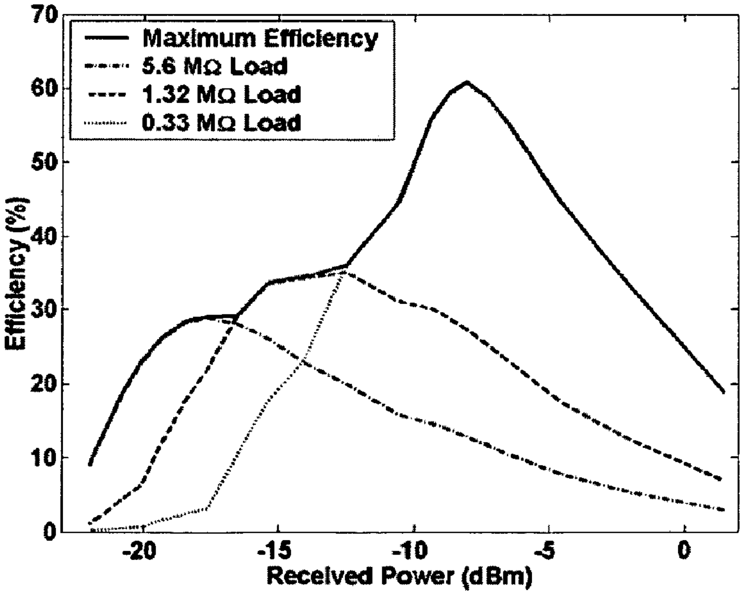

(f)

FIG. 15 (cont.)

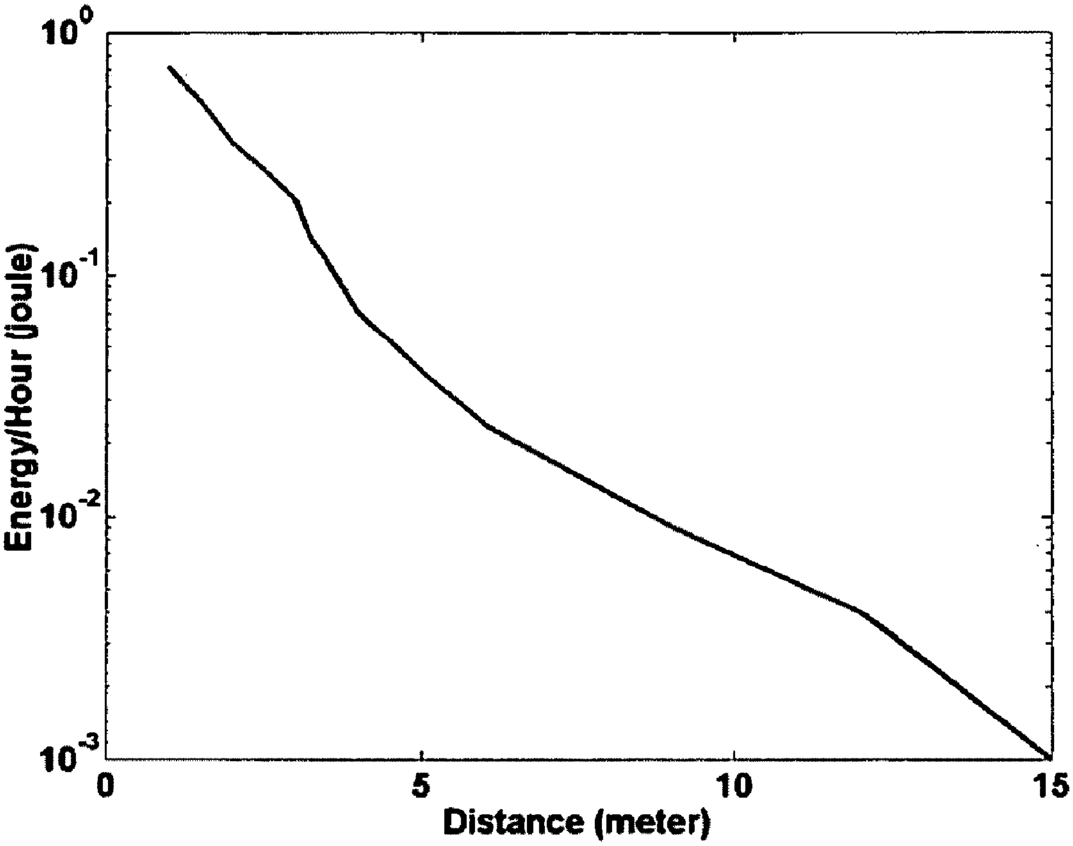

FIG. 16

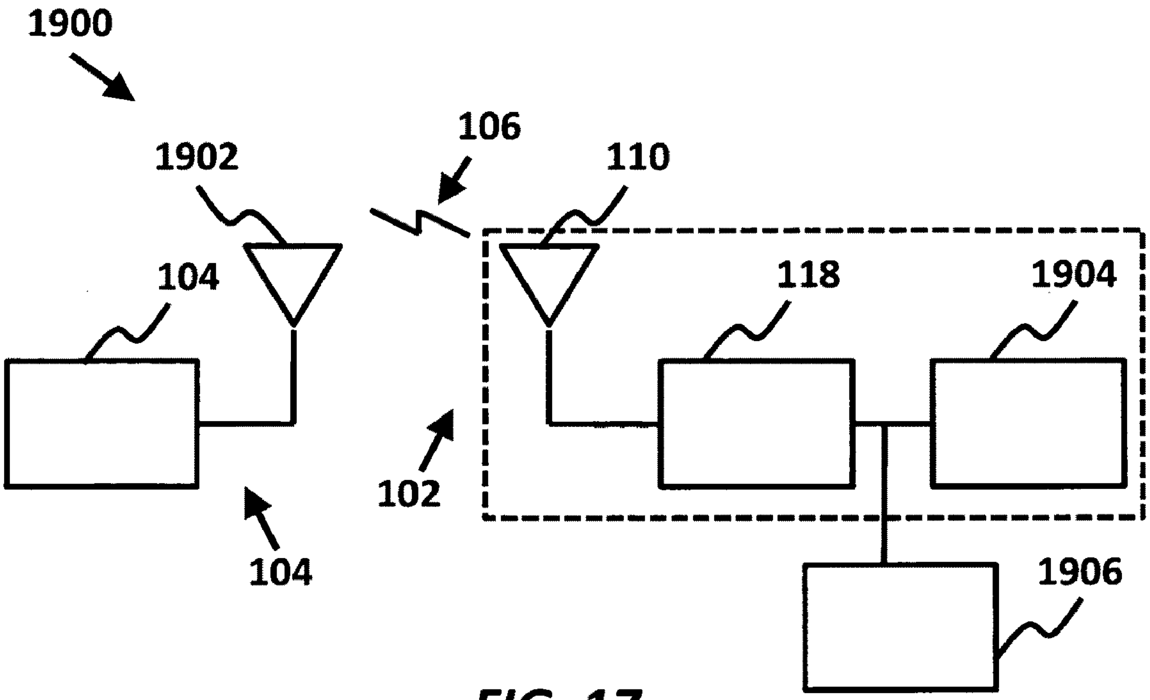
1900
1902
1904
1906

FIG. 17

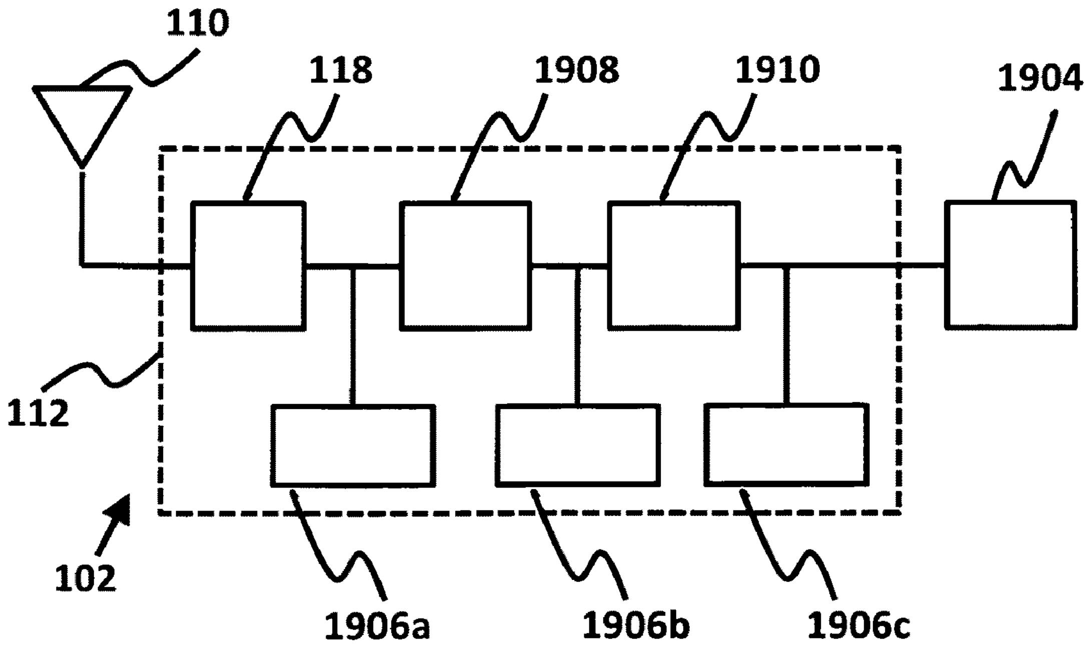
110
118
1908
1910
1904
112
102
1906a
1906b
1906c

FIG. 18

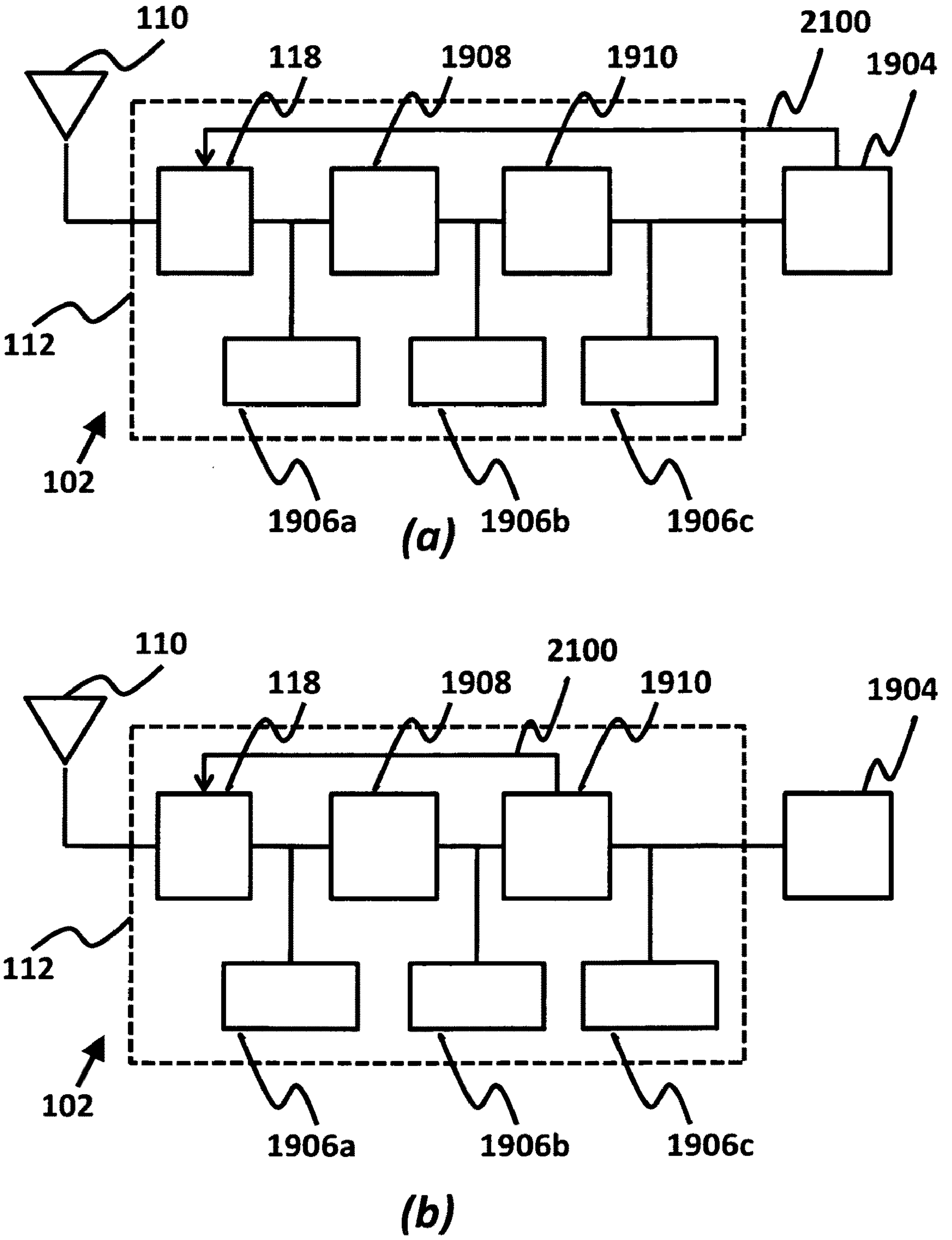
110
118
1908
1910
2100
1904
112
102
1906a
1906b
1906c
110
118
1908
2100
1910
1904
112
102
1906a
1906b
1906c

(a)

(b)

FIG. 19

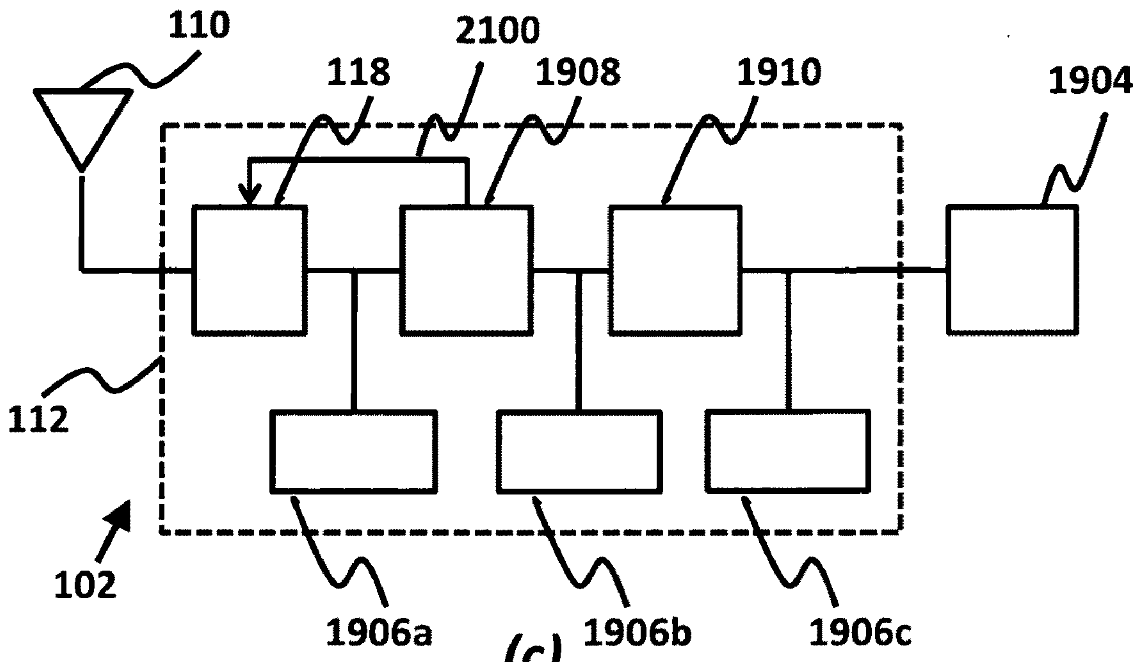
110
2100
118
1908
1910
1904
112
102
1906a
1906b
1906c

(c)

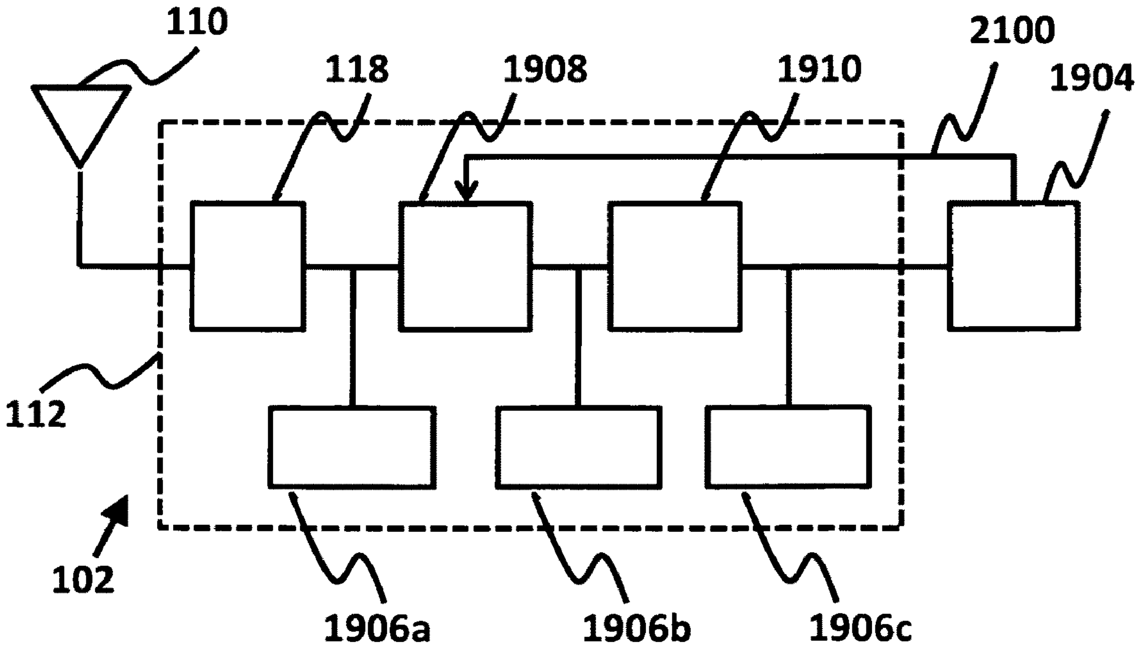
110
118
1908
1910
2100
1904
112
102
1906a
1906b
1906c

(d)

FIG. 19 (cont.)

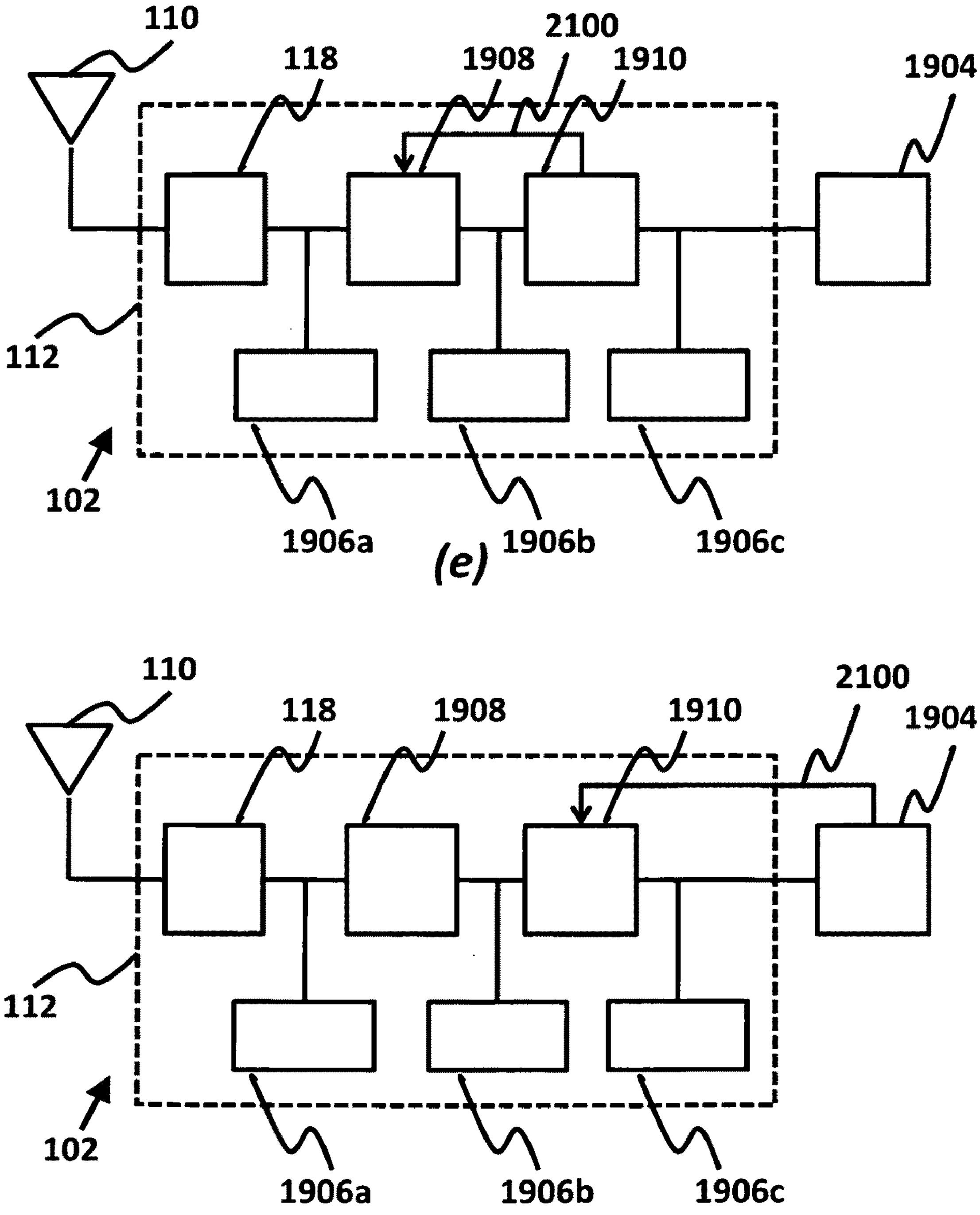
110
118
1908
2100
1910
1904
112
102
1906a
1906b
1906c
110
118
1908
1910
2100
1904
112
102
1906a
1906b
1906c

(e)

(f)

FIG. 19 (cont.)

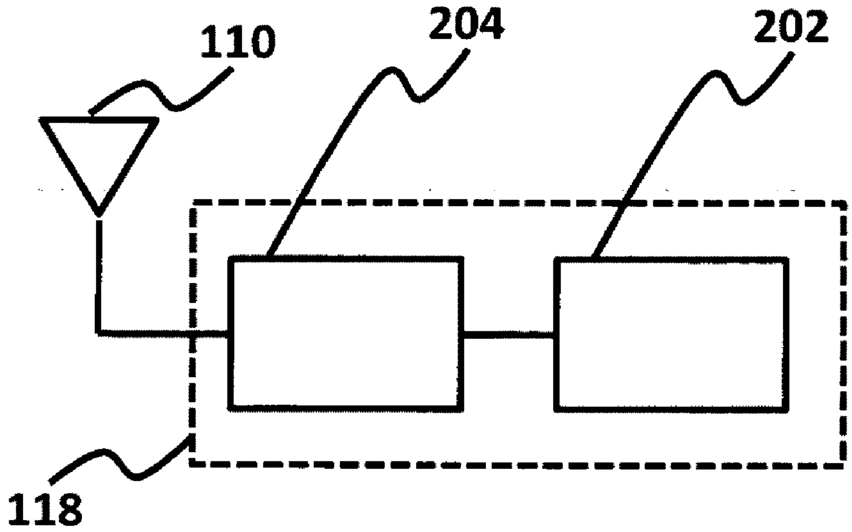
110
204
202
118

FIG. 20

RF ENERGY HARVESTING CIRCUIT

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is cross-referenced to and claims the benefit from U.S. Provisional Patent Application 60/961,009 filed Jul. 17, 2007, and from U.S. Provisional Patent Application 60/993,260 filed Sep. 10, 2007, which are hereby incorporated by reference.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] The present invention was supported in part by grant number DBI-0529223 from the National Science Foundation (NSF). The U.S. Government has certain rights in the invention.

FIELD OF THE INVENTION

[0003] The invention relates generally to energy harvesting. More particularly, the invention relates to the conversion of radio frequency electromagnetic power to electrical power.

BACKGROUND

[0004] Power extraction is an increasingly important technology, especially in applications relating to harvesting power from propagating radio frequency (RF) signals. RF powered devices are often used in applications where battery replacement is impossible, such as structural monitoring where the RF powered devices are embedded into a structure. Other applications for RF powered devices are in telemetry systems to remotely measure and report data back to a central processing unit, and in passive radio frequency identification (RFID) or passive RF tags to replace the bar code as a new form of data collection, where passive RFID tags are typically used in the range of 1-3 meters. Further, many modem biomedical implants are passively powered with radio waves to prolong the lifetime of the implanted device, and to reduce the chances of infection and chemical instability from the use of batteries. These biomedical implants generally operate within close proximity of the base station (typically 1-50 cm) and require robust designed since there is little tolerance for error in implanted devices. RF powered devices are also used in ultra-low power sensor networks in remote areas to eliminate the need for batteries in the sensor system to keep the sensor network free of maintenance, however they may have a backup battery in case the power provided by the RF radiation is insufficient. The applications for these sensor networks normally require an operating distance of 3 to 100 meters. Other applications for RF powered devices include access control, equipment monitoring and even personal identification.

[0005] RF powered devices require a power conversion circuit that can extract enough DC power from the incident electromagnetic waves for the passive device to operate. Far field RF powered devices are known to generally operate from distances of less than 10 meters from the RF source due to the high power loss from RF wave propagation at UHF frequencies. Other known devices achieve sufficient power but provide low output voltage with higher load current making them inadequate for use in passively powered wireless sensor networks.

[0006] For far-field RF energy harvesting applications, the RF energy is extracted from the air at a very low power density, since the propagation energy drops off rapidly as distance from the source is increased. More specifically, in free space, the electric field and power density drop off at the rate of $1/d^2$, where d is the distance from the radiating source. Here, the available power to a receiver in a far-field RF harvesting device decreases by 6 dB for every doubling of distance from the transmitter. Further, with multi-path fading, the power density drops off at a much faster rate than $1/d^2$. It is therefore critical that the power conversion circuit operate at very low receive power to achieve longer operating distance.

[0007] One of the major challenges to achieving this goal is the relatively high voltage requirement of rectifying circuits currently employed. When the available RF power to the receiver is under 100 μW, the available voltage for rectification in the RF to DC conversion system falls below 0.3V, which is much too low to overcome the threshold voltage (V_{th}) of conventional rectifier circuits. Currently, far-field RF energy harvesting circuits known in the art require a start-up circuit.

[0008] Accordingly, there is a need to develop an efficient rectifier circuit as well as improved system level design for RF to DC power conversion. What is needed is a highly efficient passive power conversion circuit for long distance passive sensing in distributed sensor networks. Further, new rectification circuits for far-field RF energy harvesting devices are needed to improve on the minimum power-threshold requirements for the system to operate. To overcome this power-threshold, significantly more efficient circuit and system level design is required over what is known in the art, especially in reducing or eliminating the need for a start-up circuit. Solutions must be found to circumvent or diminish the “dead-zone” in voltage rectification and otherwise reduce the effective threshold voltage in standard CMOS rectifier designs.

SUMMARY OF THE INVENTION

[0009] To address the current needs for a robust and efficient energy harvesting system, a voltage doubler rectifier RF power harvesting system is provided. The system includes at least one power harvesting module, where the power harvesting module has a voltage doubler rectifier structure circuit that includes a DC voltage input, a DC voltage output, a signal input, a first floating gate transistor and a second floating gate transistor, where the floating gate transistors are connected in series between the DC voltage input and the DC voltage output, and a gate control of the floating gate is connected to a drain of the transistor. The voltage double rectifier structure further includes a first capacitor disposed between the input signal and the drain of the first floating gate transistor, and a second capacitor disposed between a ground and the drain of the second floating gate transistor. The voltage double rectifier RF power harvesting system further includes a harvester antenna, where the harvester antenna is disposed to provide a signal to the signal input. Additionally included are a transmitter and a transmitter antenna, where the transmitter provides the signal for output from the transmitter antenna, and a full-wave peak-to-peak voltage of an incoming RF signal from the antenna to the signal input is rectified. The system also includes a powered device, connected to the DC voltage output to utilize the rectified signal.

[0010] In one embodiment of the current invention, the at least one power harvester further includes at least one charge storage device, where the charge storage device can include capacitors, rechargeable batteries, or non-rechargeable batteries, and the charge storage device is disposed between the voltage doubler rectifier structure and the powered device.

[0011] According to another embodiment of the invention, the at least one power harvester further includes a voltage limiting/regulation device, where the voltage limiting/regulating device is disposed between the voltage doubler rectifier structure and the powered device. Here, the embodiment can further include feedback loops where the feedback loops are feedback connections that can include the powered device to the voltage doubler rectifier structure, the powered device to the voltage limiting/regulation device, or the voltage/regulation device to the voltage doubler rectifier structure. This embodiment can also include at least one charge storage device disposed between the powered device and the voltage doubler rectifier.

[0012] According to another embodiment of the invention, the at least one power harvester further includes a power management device, where the power management device is disposed between the voltage doubler rectifier structure and the powered device. Here, the embodiment can further include feedback loops, where the feedback loops are feedback connections that can include the powered device to the voltage doubler rectifier structure, the powered device to the voltage limiting/regulation device, the powered device to the power management device, the power management device to the voltage limiting/regulation device, the power management device to the voltage doubler rectifier structure, or the voltage/regulation device to the voltage doubler rectifier structure. This embodiment can also include at least one charge storage device disposed between the powered device and the voltage doubler rectifier.

[0013] According to another embodiment, the at least one power harvester further can include an impedance matching device disposed between the harvester antenna and the voltage doubler rectifier. In this embodiment the impedance matching is an adjustable impedance matching, where the adjustment is programmable. Further, the adjustable impedance matching can be programmable.

[0014] In a further embodiment of the current invention, the at least one power harvester further includes passive components disposed between the signal input and the harvesting antenna, where a wider band match is provided along a cascade of the power harvesters.

[0015] In yet another embodiment, the floating gate transistor can be a PMOS-PMOScap, a PMOS-NMPScap, an NMOS-NMOScap, an NMOS-PMOScap, an NMOS-MiMcap, or a PMPS-MiMcap.

[0016] In a further embodiment, the harvesting antenna is a meanderline antenna, where the meanderline antenna includes an inner loop disposed between a pair of approximately matching outer loops, where the inner loop has a pair of antenna ports, and an input impedance of the meanderline antenna matches an in impedance of the voltage doubler rectifier. Here, the antenna outer loop can have a perimeter of 35 centimeters. Further, the antenna inner loop can have a perimeter to match the input impedance of the rectifier. Additionally, the antenna matching outer loops are disposed to divide the antenna into two identical parts as a virtual ground, where a fully differential signal is provided at the antenna ports.

[0017] In a further embodiment, the voltage doubler rectifier structure can be disposed in at least a two-stage rectifier cascade. According to one embodiment, the cascade is a 36-stage cascade. According to another embodiment the cascade is a 16 stage cascade.

BRIEF DESCRIPTION OF THE FIGURES

[0018] The objectives and advantages of the present invention will be understood by reading the following detailed description in conjunction with the drawing, in which:

[0019] FIG. **1**(*a*) shows a drawing of communication links between base station (hub) and passively powered sensors in a passively powered sensor network according to the present invention.

[0020] FIG. **1**(*b*) shows a diagram of an RF-DC power conversion system in a passively powered sensor according to the present invention.

[0021] FIG. **2** shows a block diagram of a passive RF-DC conversion circuit showing the equivalent circuit representation for the antenna and rectifier according to the present invention.

[0022] FIG. **3** shows a graph of the effect of impedance mismatch in high Q resonators.

[0023] FIG. **4** shows a diagram of parasitic components that affect performance of the RF-DC power conversion circuit.

[0024] FIG. **5**(*a*) shows a prior art conventional voltage doubler rectifier.

[0025] FIG. **5**(*b*) shows a diagram of a PMOS floating-gate rectifier according to the present invention.

[0026] FIG. **6** shows a diagram of a PMOS implementation of a floating gate transistor according to the present invention.

[0027] FIG. **7** shows a transistor level schematic of PMOS floating-gate rectifier according to the present invention.

[0028] FIG. **8** shows a diagram of a voltage doubler rectifier with N-stages in cascade according to the present invention.

[0029] FIGS. **9**(*a*)-(*d*) show the effect of a number of rectifier stages on (a) maximum voltage gain, (b) output DC voltage, and (c and d) input impedance of rectifier, according to the present invention.

[0030] FIGS. **10**(*a*)-(*b*) show (a) output voltage curves as function of the number of rectifier stages and the transistor width. (b) contour plot of constant output voltage according to the present invention.

[0031] FIG. **11** shows an antenna for an RF-DC converter according to the present invention.

[0032] FIGS. **12**(*a*)-(*d*) show (a) and (b) simulated and measured components of the antenna input impedance, and (c) and (d) measured input impedance of antenna and rectifier circuit according to the present invention.

[0033] FIGS. **13**(*a*)-(*b*) show (a) measured return loss, and (b) voltage gain for 36-stage rectifier design according to the present invention.

[0034] FIGS. **14**(*a*)-(*b*) show die images of (a) 16-stage rectifier circuit, and (b) 36-stage rectifier circuit fabricated in a 0.25 mm CMOS process according to the present invention.

[0035] FIGS. **15**(*a*)-(*d*) show measured output DC voltage (a) as a function of the sinusoidal input, (b) as function of input power, (c) as function of distance, and (d) measured voltage efficiency according to the present invention.

[0036] FIG. **16** shows PCB to establish performance of the RF-DC power converter at the system level according to the present invention.

[0037] FIGS. **17**(*a*)-(*f*) show measured DC output voltage (a) as function of distance and load for the 36-stage rectifier, and (b) as function of received power for the 36-stage rectifier. (c) Measured output power performance as function of distance and load for the 36-stage rectifier. (d) Measured output power as function of received power for the 36-stage rectifier. (e) Measured power conversion efficiency curves as function of distance and load for the 36-stage rectifier. (f) Measured power conversion efficiency curves as function of received power for the 36-stage rectifier according to the present invention.

[0038] FIG. **18** shows a graph of the measured energy extracted per hour as function of distance for the 36-stage rectifier according to the present invention.

[0039] FIG. **19** shows a shows a block diagram of one embodiment of the energy harvesting system according to the present invention.

[0040] FIG. **20** shows a diagram of one embodiment of a power harvester with voltage limiting/regulation, charge storage devices and power management according to the present invention.

[0041] FIGS. **21**(*a*)-(*i*) show diagrams of various feedback loops with a power harvester with voltage limiting/regulation, charge storage devices and power management according to the present invention.

[0042] FIG. **22** shows a block diagram of one embodiment of power harvester according to the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0043] Although the following detailed description contains many specifics for the purposes of illustration, anyone of ordinary skill in the art will readily appreciate that many variations and alterations to the following exemplary details are within the scope of the invention. Accordingly, the following preferred embodiment of the invention is set forth without any loss of generality to, and without imposing limitations upon, the claimed invention.

[0044] The present invention is directed to systems, circuitry and techniques for obtaining, recovering, acquiring and/or harvesting electrical energy from an environment having/including radio frequency (RF) signals (for example, signals in the is a frequency or rate of oscillation within the range of about 3 Hz and about 30 GHz). The RF signals may be periodic or non-periodic. The environment is generally described as being local to the system (for example, within a radius of 40 meters, and more preferably, within a radius of 3 meters). In one embodiment, the system consists of RF power harvester circuitry to obtain, recover, acquire and/or harvest (hereinafter collectively "harvest" or the like (for example, "harvests", "harvested", "harvesting")) the RF electromagnetic waves from the environment. The RF waves may come in the form of a power transmitter or ambience RF wave. In operation, an RF signal is received by the harvester circuitry, which converts the RF signal into a voltage to be used by devices (for example, a substantially DC voltage). In one embodiment, the power harvester circuitry employs floating-gate transistors to provide improved power conversion efficiency of the system.

[0045] An RF-DC power conversion system is provided, which efficiently converts far-field RF energy to DC voltages at very low received power and voltages. Passive rectifier circuits are provided in a 0.25 μm CMOS technology using floating gate transistors as rectifying diodes. A 36-stage rectifier disclosed, according to one embodiment, can rectify input voltages as low as 50 mV with a voltage gain of 6.4 and operate with received power as low as 5.5 μW (−22.6 dBm). Implemented for far field, the circuit operates at a distance of 44 meters from a 4 W EIRP source. The high voltage range achievable at low load current, makes it practical for use in passively powered sensor networks.

[0046] The current invention provides a highly sensitive and efficient rectifier circuit and system level design for RF to DC power conversion. Fully passive rectifier circuits are provided in a 0.25 μm CMOS technology disposed to operate at very low received power. A receive antenna is provided in a 4-layer FR4 board to maximize power transfer in the system.

[0047] The harvested electrical energy can then be used to power one or more electrical devices. After RF signals are converted to DC signals, the invention can power an electrical device directly or it can incorporate other components, such as charge storage devices (e.g. batteries and capacitors), voltage limiters/regulators, power management devices, or any combination thereof. In another embodiment, the system can also incorporate feedbacks between any two or more of the components including the electrical device to be powered, the RF to DC converter, the voltage regulator, and the power management device.

[0048] When RF powered devices harvest their power from RF wave radiation, a radiating source or base station is required to transmit a high intensity RF signal wirelessly through the air. The RF signals can be within the range of about 3 Hz and about 30 GHz and may be periodic or non-periodic. The environment is generally described as being local to the system within a radius range of 3 to 40 meters.

[0049] Referring now to the figures, FIG. **1**(*a*) shows one embodiment of how the system operates as a network **100** where multiple energy harvesting systems **102** receive their energy from the same power source (shown as a hub or base station) **104**, however the RF waves **106** may also come in the form of ambient RF waves **106**. FIG. **1**(*b*) shows a diagram of an RF-DC power conversion system **108** in a passively powered sensor **102** according to one embodiment of the current invention, where the RF-DC power conversion system **108** includes an antenna **110**, connected to an integrated circuit **112**. Here, the integrated circuit **112** includes a receiver **114** and a transmitter **116**, which are connected to the antenna **110** and, optionally, to each other. The integrated circuit further includes an RF-DC converter **118** connected to the antenna **110**. According to this embodiment, the integrated circuit **112** has a holding capacitor **120** connected to the RF-DC converter **116** for receiving and holding the converted power, and providing power back to the integrated circuit **112**. Further shown is a baseband processing block **122** connected to receive a signal from the receiver **114** and connected to provide a signal to the transmitter **116**. Further, the baseband processing block **122** is connected to receive and provide signals with an electrically erasable programmable read-only memory (EEPROM) **124**. Further shown in this embodiment is an ancillary sensor **126** connected to the integrated circuit **112** through an analog to digital converter interface **128**, which may provide an additional signal for further power conversion or other sensing capabilities.

[0050] FIG. **2** shows a diagram of a passive RF-DC conversion circuit **200** showing the equivalent circuit representation for the antenna **110**, rectifier **202** and impedance matching **204**. The RF to DC power conversion system of the current invention operates in UHF frequencies in the industrial, scientific and medical bands (ISM band) of 902-928 MHz. In this frequency range, RF power is transmitted more efficiently for longer distances and experiences lower propagational losses than higher frequency bands (i.e., 2.4 GHz). The system is optimized to operate at distances above 10 meters with the capacity to store charge for long periods of time. Given that the power density drops off at the rate of $1/d^2$ in free space, the propagational RF signal loss through the air at 915 MHz can be calculated to be 51.6 dB with the Friis equation for freespace loss. The maximum transmit power allowed by the FCC in the 902-928 MHz band is 36 dBm, equivalent isotropically radiated power (EIRP) (30 dBm maximum transmit power with 6 dB antenna gain), thus the received power for distances greater than 10 meters is below 27 μW, this translates to less than 75 mV in a 50Ω matched system. The power conversion circuit must be highly sensitive for long-range operation, thus the threshold voltage of the system must be greatly reduced to improve power conversion efficiency at distances greater than 10 meters. It is also essential to come up with design techniques at the system level that can increase the voltage available for rectification to further increase the power conversion efficiency.

[0051] An RF-DC power conversion system is provided to passively amplify the voltage available for rectification by forming a high-Q resonator. The system maximizes the voltage coming into the RF-DC power conversion system **108** so that it can provide a stable DC output voltage at ultra-low receive power. The block diagram **200** of the system **108**, shown in FIG. **2**, consists of an antenna **110** to pick up the power radiated by the RF waves **106**, an impedance matching network **204** to ensure maximum power transfer in the system **108**, and a rectifier circuit **202** to convert the RF signal **106** to a DC voltage. The passive amplification of voltage is done by matching the impedances between the receive antenna **110** and the rectifier circuit **202**. Due to the high-Q nature of the voltage rectification circuit **202** (see circuit schematic in FIG. **5***b*) the impedance matching creates a high-quality factor (Q factor) resonator between the receive antenna **110** and the rectifier circuit **202**.

[0052] A key aspect to improve the efficiency of the RF-DC power conversion at the system level, according to the current invention, is to maximize the input voltage to the rectifier **202**. This is done by forming a resonator (see circuit schematic in FIG. **5***b*) with high-Q factor between the impedances of the receive antenna **110** and the rectifier circuit **202**. By doing this, it passively amplifies the incoming RF signal **106**. To ensure the maximum possible power is transferred to the rectifier circuit **202**, the impedance of the receive antenna **110** is matched to the input impedance of the rectifier circuit **202**. Due to the fact that systems with high-Q resonate with greater amplitude at the resonant frequency than systems with low-Q, the high-Q resonator acts as a passive voltage-amplifier to increase the peak voltage coming into the input of the rectifier without dissipating additional power. The passive voltage gain from the high-Q resonator is directly proportional to its Q. With an increase in the amplitude of the voltage coming to the input of the rectifier **202**, the output voltage of the rectifier **202** also increases and, therefore, increases the overall power conversion efficiency of the system **108**. One drawback of the high-Q resonator is it can reduce the operating bandwidth of the RF-DC power conversion system **108** since

$$Q = \omega \cdot \left(\frac{\text{Energy Stored}}{\text{Average Power Dissipated}} \right) \text{ or } Q = \frac{f_c}{\Delta f} \tag{1}$$

where ω is the resonant frequency in radians/second, f_c is the center frequency of operation, and Δf is the bandwidth of the system. For the far-field RF-DC power conversion system **108** operating in the band 902-928 MHz, the maximum Q that can be attained without sacrificing bandwidth is 35. This limitation on the system Q does not cause much concern since on-chip components rarely have Q of more than 10, and the parasitic resistance from these components damp out the resonator to prevent the Q of the system from limiting the bandwidth. In the case of a series connected matched LC resonator, the reactive components are complex conjugates of each other and the resistive components are matched. The Q of the resonator is therefore

$$Q = \frac{1}{2}\frac{\omega L}{R} = \frac{X_L}{2R} \text{ or } Q = \frac{1}{2}\frac{1}{\omega RC} = \frac{X_C}{2R} \tag{2}$$

where X_L and X_C are the reactive components, and R is the resistive component of the LC resonator. The Q of the resonator is half the Q of the rectifier and antenna since the resistance in the series connected matched resonator is doubled. To achieve a high system Q, it is therefore desirable to increase the reactive components of the rectifier **202** and antenna **110** while reducing resistive components.

[0053] A matching network **204** between the receive antenna **110** and rectifier **202** is necessary to fine tune the impedance match between the antenna **110** and the rectifier **202** to further reduce transmission loss and increase the voltage gain. Coarse impedance matching is done through circuit and antenna design but fine impedance matching must also be done on the PCB for more accurate matching. FIG. **3** shows a graph of the simulated effect **300** of impedance mismatch for a typical high Q resonator. The maximum voltage gain that can be attained is equal to the Q of the resonator, or half that of the rectifier Q. With impedance mismatch greater than 7%, the passive voltage gain from the resonator is reduced to below 3 regardless of the resonator Q. With impedance mismatch greater than 15%, the high-Q matching network provides no voltage gain and even yields a voltage attenuation that becomes greater with increasing Q. The impedance mismatch between the antenna and rectifier must be minimized to obtain a high voltage gain.

[0054] From the top-level diagram of FIG. **2**, the rectifier circuit is modeled by impedance with a real part R_{rect} **206** and a reactive part X_{rect} **208**. From the system point of view, the rectifier circuit **202** must be designed to reduce threshold voltage loss (V_{th}) as much as possible to improve the efficiency of the RF-DC power conversion system **108**. The rectifier circuit (see FIG. **5**(*b*)) must also be provided so that the output voltage can be scaled by cascading multiple rectifier stages in series (see FIG. **8**). Improving the Q of the rectifier input impedance is essential to increasing the power conversion efficiency of the overall system **108**. The rectifier input impedance Q should also be kept as high as possible and parasitic components in the rectifier must be kept as low as possible to maintain a high overall system Q and thus power conversion efficiency.

[0055] The number of cascaded rectifier stages (see FIG. **8**) in the RF-DC conversion system **108** also has a significant effect on the rectifier input impedance. In a conventional voltage rectification circuit design (see FIG. **5**(*a*)) in CMOS technology, the rectifier impedance as seen from the input is capacitive and resistive due to the gate capacitance and the channel resistance (r_{ds}) of the MOS transistor. In general, cascading multiple rectifier stages in series causes capacitive components to increase linearly with the number of stages, while providing parallel paths causes the resistive component to decrease. With a large number of stages, the resistive component from the rectifier is so low that it is dominated by other sources of parasitic resistance (i.e. drain and source connection resistors) and hence, the Q of the system is reduced due to the linearly increased parasitic capacitance.

[0056] Conversely, if there are too few rectifier stages in cascade, the output voltage of the rectifier may not be high enough to operate the sensor node. As the number of rectifier stages increases, the DC output voltage increases until the number of rectifier stages reaches an optimal point. Adding more stages beyond the optimal point reduces the system Q and causes a reduction in the DC output voltage. Thus, it is critical for the number of rectifier stages to be selected through thorough circuit simulation so that the output DC voltage is maximized while maintaining a high system Q to achieve maximum power conversion efficiency.

[0057] The current invention provides a reduction of the unwanted parasitic components that affect system performance. Since the system is to be designed with a high-Q resonator, any additional parasitic components between the antenna and rectifier will greatly diminish the performance of the power conversion system. It is therefore critical to specify an accurate parasitic model for simulating the effect of all parasitic components that affect power conversion efficiency. In the design of the RF-DC power conversion system, the traces connected to inputs of the rectifier are most sensitive to parasitics since a high-Q resonance is required at these inputs. For example, FIG. **4** shows an equivalent circuit model **400** for the parasitic components that affect the power conversion efficiency of the rectifier circuit.

[0058] These parasitic components cannot be avoided altogether, but the value of each parasitic component can be reduced through careful layout and package selection. The bond pad **402** can be designed to be minimum size and only consisting of the top two metal layers to reduce bond pad capacitance to the substrate. The package for the integrated circuit is selected so that pin-parasitics are minimized and critical input pins **404** are placed in locations where the length of bond wires **406** is minimized. PCB traces **408** are made as short as possible and they are impedance controlled to reduce the parasitic capacitance and inductance.

[0059] The antenna **110** structure is critical in the RF-DC power conversion system since it must extract the power radiated by the RF waves **106**. The antenna **110** performs best when it is impedance matched to the rectifier circuit **202** at the operating frequency to reduce transmission loss from PCB traces. Also, the antenna **110** must be small in area and must have a bandwidth large enough to cover the frequency band from 902-928 MHz.

[0060] FIG. **5**(*a*) shows the conventional voltage doubling rectification circuit **500**, and FIG. **5**(*b*) shows a floating-gate rectification circuit **502** according to the current invention. For the floating-gate rectification circuit **502**, floating-gate devices **504** are used to create a gate-source bias to reduce the threshold voltage loss of the MOS transistor **506**.

[0061] The voltage doubler rectifier structure is provided for the RF-DC power conversion system because it rectifies the full-wave peak-to-peak voltage of the incoming RF signal and it can be arranged in cascade to increase the output voltage. The voltage conventional doubler rectifier **500** in FIG. **5**(*a*) consists of a peak rectifier **508** formed by D_1 and C_2 and a voltage clamp **510** formed by D_2 and C_1. The voltage clamp **510** and the peak rectifier **508** are arranged in cascade configuration to provide a passive level shift in voltage before rectification.

[0062] In the negative phase of the input, current flows through diode D_2 while D_1 is cutoff. The voltage across diode D_2 stays constant around its threshold voltage and the voltage at node (1) is charged to $-V_{th2}$. At the negative peak, the voltage across capacitor C_1 is $V_{amp}-V_{th2}$, where V_{amp} is the amplitude of the input signal. In the positive phase of the input, current flows through diode D_1 while D_2 is in cutoff. The voltage across capacitor C_1 remains the same as the previous phase because it has no way to discharge. At the positive peak, the voltage across D_2 is $2V_{amp}-V_{th2}$. Since D_1 is conducting current to charge C_2, the voltage at the output is a threshold voltage below that across D_2, i.e., the voltage at the output V_{out} is $2V_{amp}-V_{th2}-V_{th1}$.

[0063] For the floating-gate rectifier circuit **502** of the current invention, the diodes D_1 and D_2 of the conventional rectifier circuit **500** are replaced by diode-tied floating gate transistors **504**. The floating gate devices **504** are provided to passively reduce the threshold voltage of the rectifier circuit **502**. In a floating gate device **504**, when charge is injected into the floating gate of the transistor **506**, it remains in the gate oxide because of the high impedance provided by the oxide layer. The gate oxide is a very good insulator, which keeps the charge from leaking off in the floating gate **504**.

[0064] To provide a floating gate device **600** in a standard CMOS process, a MOS capacitor **602** is placed in series with the gate **606** of the diode-tied transistor **604** as shown in FIG. **6**. The gate of the diode-tied transistor **606** and the MOS capacitor gate **608** are connected together to form a high impedance node to trap charges in the floating gate **610**. The charge in the floating gate **610** is therefore fixed which results in a fixed voltage bias across the MOS capacitor **602**. The charges that are trapped inside the floating gate device **600** act as a gate-source bias to passively reduce the effective threshold voltage of the transistor **604**.

[0065] The charge on the floating gate **610** can be injected via Fowler-Norheim (F-N) tunneling when the rectifier is not operating, or it can be charged by injecting a relatively large sinusoidal signal to the input of the rectifier at any time. The Fowler-Norheim tunneling technique charges the floating gate **610** to the desired voltage much faster, but the amount of charge is harder to control and also, additional circuitry is needed to inject or remove charge from the floating gate **610**. Referring again to FIG. **5**(*b*), the rectifier circuit **502** of the current invention is provided to have charge injected to the floating gate **504** with a relatively large sinusoidal input with amplitude larger than the threshold voltage of the transistor **506** used for rectification. With the large sinusoidal voltage injection, charge is injected to the floating gate **504** via the parasitic capacitance between the gate-source and gate-drain junction of the transistor **506**. With the floating-gate device **504**, the threshold voltages of the diode-tied transistors M_1 and M_2 are reduced by creating a gate-source bias. The gates of transistors M_1 and M_2 in FIG. **5**(*b*) are high impedance nodes so any charge trapped in the floating gates **504** can be retained for a long time. Retaining charge in floating-gate devices **504** is critical to the useful lifetime for the power conversion circuit in the current invention. With the 70-angstrom oxide thickness in the 0.25 μm CMOS process, the device retains charge in the floating gate **504** in excess of 10 years for normal operation at room temperature. However, the performance of the rectifier circuit **502** may reduce slightly as charge is leaked from the floating gate **504**. During fabrication, the residual charge trapped in the floating gate **504** may also affect the threshold voltage of the rectifier circuit **502**, hence the floating gate **504** must be programmed to account for these residual charges.

[0066] The transistor level schematic of the 36-stage floating-gate rectifier **700** is shown in FIG. **7**. V_{RF} represents the input signal extracted from the RF wave **106** (not shown), V_{DCin} is the input DC voltage coming from the previous rectifier stage (not shown) and V_{DCout} is the output DC voltage of the rectifier **700**. The 36-stage embodiment **700** uses diode-tied PMOS transistors **702** as rectifying devices and a MOS capacitor (MOSCAP) **704** is used to create the gate-source bias for each individual diode-tied transistor **702** in the rectifier **700**. The MOSCAPs **704** (C_{g1} and C_{g2}) generally operate in the depletion region and their capacitance is a function of the applied voltage. To create the gate-source bias, a large sinusoidal signal is applied at the input of the rectifier **700** and charge is injected over time through the parasitic capacitance of the diode-tied transistors **702**. When the gate-source bias reaches a potential close to V_{th}, the capacitance reaches a flat point and remains constant. The amount of charge stored in the floating gate **708** reaches equilibrium and the MOSCAP **704** operates in the inversion region, the source-gate bias voltage is, therefore, approximately the threshold voltage of the diode-tied transistors **702** (M_1 and M_2).

[0067] The individual stages of the floating-gate voltage doubler rectifier circuit can be arranged in cascade to increase the output voltage of the rectifier. FIG. **8** shows N stages of a voltage doubler rectifier in cascade **800**. When the rectifier stages **802** are cascaded, each rectifier stage **802** acts as a passive voltage level shifter in addition to the voltage shift in voltage clamp and peak rectification. The number of rectifier stages **602** used in the design is important since too few rectifier stages yields insufficient output voltage and too many rectifier stages damps out the effect of the high-Q resonator.

[0068] One of the important tradeoffs in the voltage doubler rectifier is the size of the transistor versus parasitic capacitance. The smaller the transistor size, the less parasitic capacitance it has, however, rectification efficiency is lowered by the smaller transistor size since smaller transistor can deliver less current to the load. The transistor sizes can be reduced to a few times the minimum width to reduce parasitic capacitance as seen from the input of the rectifier, however, the reduction in channel width may cause a decrease in the performance of the rectifier due to the increase in the channel resistance of the diode-tied transistors. As an example, two different sample designs are provided to compare the tradeoffs between a reduction in the parasitics and the rectifier performance.

[0069] The first sample design uses a relatively large device size of 12 mm/0.24 μm (NMOS) and the second uses 2 μm/0.24 μm PMOS devices. FIG. **9**(*a*) shows the maximum voltage gain that can be achieved in the rectifier with different numbers of floating-gate rectifier stages arranged in cascade. As the number of rectifier stages increase, the resonator Q between the antenna and rectifier impedance is reduced. In fact, the resonator Q is inversely proportional to the number of rectifier stages.

[0070] The output voltage of the rectifier for both designs with an input voltage with 300 mV amplitude is shown in FIG. **9**(*b*). The output voltage initially increases as more rectifier stages are added until an optimal point then it reduces as the resonator Q is decreased. The impact of the number of rectifier stages on the input impedance is shown in FIGS. **9**(*c*) and **9**(*d*). As the number of rectifier stages increase, the capacitive component in the rectifier input impedance increases thus reducing the reactive component of the rectifier input impedance. The reactive component of the rectifier input impedance is found to be inversely proportional to the number of rectifier stages. With the decrease in the reactive component, the maximum voltage gain that can be achieved at the input is also decreased at the same rate since the resistive component in the input impedance stays fairly constant. In this example, the optimized number of cascaded rectifier stages for the 12 μm (NMOS) and 2 μm (PMOS) designs is 16 and 36 stages, respectively. The simulated input impedance for the two samples is 6.6-j85Ω and 6.6-j84Ω, almost identical to each other.

[0071] To establish the tradeoff between the transistor size used for rectification versus the number of rectifier stages, a model is extracted by curve fitting of the output voltage data for the two different designs. FIG. **10**(*a*) shows a 3-dimensional surface plot for the rectifier output voltage as a function of the number of rectifier stages and the transistor width. It can be observed that a width of 2 μm is the best design choice for the highest output voltage when the number of rectifier stages is below 40. Any increase in device size from 2 μm results in a reduction in the output voltage. Reducing the device size below 2 μm could lead to a slight improvement in the output voltage. However, the number of rectifier stages required is significantly increased which results in a larger die size. FIG. **10**(*b*) shows the constant output voltage contour plot as function of the transistor width and the number of rectifier stages. The 36-stage design with a 2 μm device size is very close to the highest voltage contour, thus, this design is optimal for the output voltage.

[0072] The antenna **110** for the RF-DC conversion circuit **108** is designed with meander lines on a printed circuit board to reduce the area of the antenna **110** and to provide the desired antenna input impedance to the impedance matching network. FIG. **11** shows one embodiment of the antenna with two outer loops **1100** on each side connected to an inner loop **1102** in the center. The antenna ports **1104** are in the middle and connect to the antenna **110** through the inner loop **1102** to transform the impedance as seen from the outer loops **1100** of the antenna **110**. To achieve resonance at the desired frequency, the antenna **110** is provided to match with the input impedance of the floating-gate rectifier circuit **502**, **700**, **800**). The outer loop **1100** of the antenna **110** is designed to tune the antenna **110** to the correct operating frequency and the perimeter of the loop **1100** is 35 centimeters, roughly equal to one wavelength at 916 MHz, according to this embodiment. The perimeter of the inner loop **1102** varies depending on the value of the input impedance of the rectifier (**502**, **700**, **800**). The line of symmetry between the two sides of the antenna **110** splits the antenna **110** into two identical parts and this acts as a virtual ground to create the fully differential signal at the antenna ports **1104**. The dimension of the drawn area of the antenna is 15 cm×2 cm (6×0.8 inches). The antenna **110** is tuned for maximum received power at 916 MHz and the antenna is designed on the top layer of a 4-layer FR4 board with carefully controlled impedance to ensure the impedance of PCB traces does not vary from the intended values. According to one embodiment, the traces **1106** in the antenna **110** are exposed so that they can be tapped to fine tune the operating frequency. The impedance matching network **204**

(see FIG. **2**) can be placed anywhere along the antenna **110** to fine-tune the matching between the antenna **110** and the rectifier circuit (**502**, **700**, **800**). The antenna **110** has a bandwidth of about 40 MHz to maintain a high antenna Q. Due to the fact that the antenna **110** must be impedance matched to the rectifier (**502**, **700**, **800**), the loop configuration provides little antenna gain in the antenna **110** throughout the frequency bandwidth it operates.

[0073] FIG. **12**(*a*) shows the simulated and FIG. **12**(*b*) shows measured input impedance of the antenna **110** around the operating frequency band. The measured resistive component shows less than 1Ω variation from the simulated value and the reactive component also correlates well with the simulated results. The simulated impedance shown in FIG. **12**(*c*) and the measured input impedance shown in FIG. **12**(*d*) corresponds to the components of the receive antenna **110** for the 36-stage rectifier configuration (see FIGS. **7** and **8**). From the measurement data, the measured resistive components of the antenna **110** and rectifier (**502**, **700**, **800**) match well throughout the intended bandwidth for both embodiments. The reactive component of the 36-stage design matches at 910 MHz.

[0074] FIG. **13**(*a*) shows the plot of the return loss (S_{11}) and FIG. **13**(*b*) shows the voltage gain for the 36-stage configuration. The antenna **110** matches so that less than 1% of the transmitted signal is reflected back to the antenna (i.e. S_{11} below −20 dB) in the frequency band of operation 902-928 MHz. The configuration achieves voltage gain above 5 in a 30 MHz bandwidth, however the peak voltage gain is shifted down by approximately 10 MHz from the intended 915 MHz center frequency.

[0075] As an example of the efficacy of the current invention, rectifier devices with 16 NMOS stages and 36 PMOS stages were fabricated in a 0.25 µm 5 metal single-poly CMOS process as shown in FIGS. **14**(*a*) and (*b*), respectively. The active die areas are 350 µm×600 µm and 400 µm×1000 µm, respectively. The majority of the active area is consumed by the MOSCAPs for floating gate devices and the holding capacitors. The 16-stage and 36-stage rectifier layouts are arranged in two rows of 8 stages each, and six rows of 6 stages each, respectively. Both dies have a pair of 1-stage rectifiers included so that the functionality of the NMOS and PMOS diode-tied floating-gate devices can be individually characterized. Each rectifier is packaged in a 0.5 mm pin pitch QFN-32 package to reduce bond wire parasitics and trace lengths on the PCB.

[0076] In order to test the rectifier circuit, a small-amplitude 1 MHz sinusoidal signal is applied directly to the input. The low frequency signal minimizes parasitic effects introduced by the board as well as voltage losses in the PCB traces. FIG. **15**(*a*) shows the measured output DC voltage unloaded and with a 5 MW load for both rectifier designs as a function of the input sinusoidal amplitude, FIG. **15**(*b*). From a straight-line extrapolation of the output voltage curves, the threshold voltage for the 16-stage rectifier is 33 mV compared to 30 mV for the 36-stage rectifier. Since the difference in threshold voltages is very small between the two rectifier designs, the improvement in the output voltage of the 36-stage rectifier is primarily due to the higher number of cascaded rectifier stages used in the design.

[0077] The performance of the rectifier circuit can be specified by the sensitivity of the rectifier or the minimum input power required for voltage rectification. The average input power can be expressed as:

$$P = \frac{QV_{rms}^2}{R} = \frac{QV_{amp}^2}{2R} \tag{3}$$

where Q is the passive voltage gain from the antenna to the input of the rectifier, VMS is the input rms voltage to the rectifier, V_{amp} is the rectifier input amplitude and R is the input impedance of the rectifier circuit. The voltage gain of the 16-stage and 36-stage rectifiers, measured from the return loss of the matching between the rectifier and antenna is 4.8 and 6.4, respectively. Assuming a rectifier input impedance of 7Ω, the equivalent input power is derived and shown in FIG. **15**(*b*). In the case of the 16-stage rectifier, 2V is achieved for 80 mV input, which is equivalent to a power 20.95 µW (−16.8 dBm). Using the Friis formula for free-space wave propagation, this corresponds to an operating distance 23 meters (FIG. **15**(*c*)) assuming a 4 W radiating source. This is the maximum theoretical distance that can be achieved with this design while keeping radiated power and antenna gain under the FCC emission limit. For the 36-stage rectifier circuit, the rectifier performs better with low voltages due to the increased number of voltage doubler rectifier stages and the higher passive voltage gain. The 36-stage rectifier circuit achieves 2V for an input voltage amplitude as low as 50 mV. Using (3), these values are equivalent to 4.7 µW of power (−23.3 dBm) at 2.0 V output. In free-space, this corresponds to 49 meters in distance.

[0078] Defining the voltage efficiency as the percentage of the theoretical voltage achieved, the theoretical output for the voltage doubler architecture is the number of cascaded rectifier stages times the peak-to-peak voltage into the rectifier. In the case of the 16-stage rectifier, the theoretical output voltage is 16 times the input peak-to-peak voltage. For input voltages above 100 mV, the rectifier efficiency is over 80% as shown in FIG. **15**(*d*) and, hence, it is more efficient at higher input voltages. The voltage efficiency for the 36-stage design is higher as compared to the 16-stage rectifier design for low voltages.

[0079] To test the overall performance of the RF-DC rectifier in actual applications, a custom printed circuit board was fabricated that includes the antenna with exposed metal traces for impedance and frequency tuning as shown in FIG. **16**. The chip is directly soldered to the board and physically abuts to the antenna to minimize any stray inductance. The output DC voltage is observed via the SMA connector with an oscilloscope or a digital multimeter (DMM).

[0080] The performance of the RF-DC power conversion system can be measured wirelessly with propagating electromagnetic wave radiation. The output of an HP/Agilent 8665A signal generator is amplified by a Mini Circuits LZY-2 power amplifier to generate signals with 6 W of power in the frequency range of 902-928 MHz. The signal at the output of the power amplifier is then fed to an antenna through several series connected SMA cables. The losses in the cable combined with the return loss between the power amplifier and the transmit antenna are measured to be around 1.7 dB. Thus, the maximum radiated power is about 36 dBm (4 Watts).

[0081] Only the 36-stage rectifier design is wirelessly measured since it has superior performance in terms of received power sensitivity. Before measuring the overall performance, it is necessary to tune the frequency of the resonator circuit. The resonant frequency can be tuned to the 902-928 MHz range by placing a shorted stub tuning on the bare metal trace at the end of the antenna on the PCB. The tuning of the resonant frequency does affect the output voltage of the RF-DC conversion circuit, since the shorted stub tuning slightly modifies the physical shape of the antenna. The decrease in the output voltage is greater when the original resonant frequency is further from the desired (i.e., longer length stub tuning) band and the decrease is smaller for a resonant frequency closer to the desired band. Although the frequency tuning does affect the output voltage, this change in output voltage is much smaller compared to the change in output voltage caused by impedance mismatch. For the 36-stage rectifier design, the resonant frequency is originally 860 MHz so the size of the loop is reduced only slightly to increase the resonant frequency to 906 MHz. For the wireless measurement of the RF-DC power conversion circuit, the center frequency is set at 906 MHz for the 36-stage rectifier design.

[0082] FIGS. **17**(*a*) and (*b*) show the measured output voltage of the 36-stage rectifier as a function of distance and input power for various resistive loads. The 36-stage conversion circuit is capable of outputting DC signal levels of 2V at distances up to 15 meters with a 36 dBm radiating source when it is not loaded as shown in FIG. **17**(*a*). The measured output DC voltage decreases exponentially as distance increases and also decreases as the load resistance is decreased. With a 0.33MΩ load, the conversion circuit operates within 7 meters from the radiating source while providing 1 volt DC. From observation, the operating distance of the rectifier reduces quickly as the load current is increased while at closer distances, the dependence on load current is much less. FIG. **17**(*b*) shows the output voltage curves of the 36-stage design as function of received power calculated from the Friis equation for free-space propagation loss. The slope of the output voltage curves are relatively constant over the full range since the output voltage is inversely proportional to the log of the operating distance. The 36-stage design operates well when the received power is higher than −22 dBm.

[0083] FIG. **17**(*c*) shows the measured output power as a function of distance. The top curve represents the theoretical received power in free-space as calculated by the Friis formula for free-space propagation loss given a 36 dBm radiating source. With a 0.33MΩ resistive load, the current requirement at the load is higher so more current is drawn from the output of the rectifier to drive the load hence the curve shows a steeper slope. The slope of the roll off is directly proportional to the load current at the output of the rectifier. With a 1.32 MΩ load, 1 μW at 10 meters is possible and up to 12.3 meters with the 5.6 MΩ load. FIG. **17**(*d*) shows the output power as function of received power. From this plot, the point where the power curve starts to roll off from the straight (open circuit) line can be seen more clearly. For the 36-stage rectifier design, the output power starts to roll off at about −8 dBm for the 0.33 MΩ load, −13 dBm for the 1.32 MΩ load, and −18 dBm for the 5.6 MΩ load.

[0084] The measured power conversion efficiency versus distance curves are shown in FIGS. **17**(*e*) and (*f*). The power conversion efficiency curve for each individual load is shown as well as the maximum efficiency that can be achieved at a particular distance. For a perfectly matched resonator network, the power available for rectification is half of the received power in the antenna since only half of the power is dissipated in the rectifier. The power conversion efficiency of the RF-DC power conversion system is defined as the ratio between the DC output power and available power for rectification (half of the received power) from the antenna. The maximum efficiency measured is 60% with the 0.33 Mb load. For lower load currents, the output power is higher at longer distances so the power conversion efficiency is also sufficient at longer distances from the RF source. The configuration is most efficient at 3 meters and the power conversion efficiency curves rolls off as the inverse of distance from the RF source. The peak efficiency with the 5.6 MΩ and 1.32 MΩ load is 29% at about 9 meters and 35% at 5 meters, respectively.

[0085] FIG. **18** shows the maximum energy that can be extracted from the RF signal in one hour at various distances from the radiating source. From the logarithmic plot of extracted energy against distance, the extracted energy curve is approximately a straight line. The maximum extractable energy in an hour is 0.8 J at 1 meter, 40 mJ at 5 meters, 7 mJ at 10 meters and about 1 mJ at 15 meters from the source. With one-hour charge time, the RF-DC power conversion circuit acquires and replenishes enough energy to operate a circuit that dissipates 1 mW for one second or 100 mW for 10 ms at distances up to 15 meter. For an increase of every meter in distance, the amount of energy that can be extracted from the RF signal is reduced by a factor of 2.6. This establishes the trade off between power dissipation and operating distance of the integrated circuit powered by the RF-DC power conversion circuit.

[0086] According to one aspect of the present invention, acquiring and/or harvesting electrical energy is done from an environment having RF frequency, or rate of oscillation, within the range of about 3 Hz and about 30 GHz. The RF signals may be periodic or non-periodic. FIG. **19** shows a block diagram of one embodiment of the energy harvesting system **1900**, where the diagram shows how the RF waves **106** may come from a power transmitter **104** having a transmitter antenna **1902** however it is within the scope of the invention that the RF waves may be ambient RF waves, and be received by the energy harvester system **102**. In operation, an RF signal **106** is received by the harvester circuitry **112** (see FIG. **2**), which converts the RF signal **106** into a voltage to be used by devices **1904**, for example, a substantially DC voltage. FIG. **19** is a diagram of an RF power harvesting system **102** that harvests RF signal **104** from ambient RF energy via one or more RF signal sources **104**. The RF signal source **104** provides one or more RF signals **106**, which may be continuous or non-continuous, and/or periodic or non-periodic, and/or wideband/multi-band or narrowband, where the power transmitter **104** can be powered by a wall outlet, a battery or other power sources. The harvesting antenna **110** collects the RF signal(s) **106** and the power harvester **102** converts the RF signal(s) **106** to a DC supply. The DC voltage may be used by various devices **1902**, or stored on charge storage devices **1906**, where the charge storage device may be external or internal to the system **102**. The charge storage device **1906** stores the charge output by the power harvester **102** and, as needed, supplies the charge to the device **1904** for example, when its power dissipation exceed the power harvested by the power harvester **102**. The charge storage **1906** device may also extend the operational usage time of the device by providing extra charge.

[0087] According to one embodiment of the invention, the power harvester **102** can communicate back to the transmitter **104** via radio link or via backscattering of the incident RF wave **106**. The charge storage device may be considered external to the power harvester.

[0088] FIG. **20** is a diagram of an exemplary embodiment of a power harvester **102** with voltage limiting/regulation **1908**, charge storage devices **1906** and power management **1910**. The antenna **110** collects the RF signal **106** (not shown), the RF-DC rectifies **118** the RF signal **106** to a DC voltage that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage device **1906***a* and its output charges a different charge storage device **1906***b*. Power management **1910**, in this embodiment, controls the conduction path between the output of voltage regulator **1908** and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed or instructed. Charge storage **1906***c* holds the excess charge from the power harvester **102** when device is in operation and gives off charge when device is drawing more current than the power harvester circuit **112** can provide. Notably, the power management **1910** may include a switch or circuitry, which controls, monitors and/or analyzes:

[0089] (i) the power harvesting operation of the power harvester and/or components thereof, for example, the power conversion circuitry;

[0090] (ii) the operating characteristics of the power harvester and/or components thereof;

[0091] (iii) the characteristics of the output power of the power harvester, for example current; voltage and temporal characteristics thereof;

[0092] (iv) the storage operation of one or more of the charge storage devices and/or charge supplied thereto (via, for example, the power harvester); and/or

[0093] (v) the characteristics of the output power of one or more of the charge storage devices, for example, current, voltage and temporal characteristics thereof.

[0094] The voltage limiting/regulation **1908** limits the voltage that can be charged to circuits and charge storage devices **1906** to prevent it from breaking down. The voltage regulation suppresses the voltage supply voltage variation to a small range, preferably under 100 mV of variation, over wide range of temperature, supply voltage and process variations.

[0095] The power management circuitry **1910** controls the charge transfer from the power harvester **102** to the device for safe operation. The power management **1910** may turn on the conduction path for charge transfer when power at the device **1904** is needed and turns off the conduction path when power is no longer needed. The power management **1910** may connect the power harvester **102** to the external device **1904** continuously, at a regular interval, or only once, depending on the operational power requirement of the device **1904**.

[0096] FIGS. **21**(*a*)-(*f*) are diagrams of exemplary embodiments of a power harvester **102** with voltage limiting/regulation **1908**, charge storage devices **1906** and power management **1910**. In the embodiment shown in FIG. **21**(*a*), a feedback **2100** between the device **1904** and the RF-DC power conversion **118** gives the device **1904** the control of the power harvester **102**. The feedback loop **2100** can be used to partially or fully enabling and disabling the RF-DC power conversion **118**. It can be used for tuning the RF-DC power conversion **118** for desirable performance and it can also be used to select between different circuit structures in the RF-DC power conversion **118**. Feedback **2100** is preferred when the RF-DC power conversion **118** requires tuning for enhanced and/or optimized matching, or when communication to the transmitter is required. The device **1904** can trigger the impedance matching network in the RF-DC **118** from matched to total mismatch. For example, this exemplary embodiment of the power harvester **102** may enable a communication link to the transmitter via backscattering of the incident RF signal. On a system where the antenna is shared, the feedback can disconnect and isolate the RFDC block from the antenna.

[0097] In the embodiment shown in FIG. **21**(*b*), the feedback **2100** between the power management **1910** and the RF-DC power conversion **118** allows the power management **1910** control of power harvester **102**. The feed back loop **2100** can be used to partially or fully enabling and disabling the RF-DC power conversion **118**. It can be used for tuning the RF-DC power conversion **118** for desirable performance and it can also be used to select between different circuit structures in the RF-DC power conversion **118**. This feedback configuration may be preferred when the RF-DC power conversion **118** requires tuning for enhanced and/or optimized matching that must be controlled by the power management **1910**. This configuration may be used to adjust the output voltage level of the RF-DC power conversion circuit **112**, which in turn varies the voltage at regulator output to flexibly accommodate for the voltage requirement of the device **1904**. On a system where the antenna is shared, the feedback may be employed to disconnect and isolate the RF-DC block from the antenna so it can be used for other purposes. The feedback loop **2100** can also be used for the purpose of triggering the charging the floating gate devices in certain rectifier circuits to enhance power conversion efficiency or minimize or prevent degradation in the performance of the power harvester.

[0098] FIG. **21**(*c*) shows a diagram of a power harvester **102** with voltage limiting/regulation **1908**, charge storage devices **1906** and power management **1910**. The antenna **110** collects the RF signal **104** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage device **1906***a* and its output charges a different charge storage device **1906***b*. Power management **1910** controls the conduction path between the output of voltage regulator **1908** and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed. Charge storage **1906***c* holds the excess charge from the power harvester **102** when device **1904** is in operation and gives off charge when device **1904** is drawing more current than the power harvester circuit **112** (not shown) can provide. The feedback **2100** between the voltage limiting/regulation **1908** and the RF-DC power conversion **118** allows the voltage limiting/regulation **1908** control of the power harvester **102**. The feed back loop **2100** can be used to partially or fully enabling and disabling the RF-DC power conversion **118**. It can be used for tuning the RF-DC power conversion **118** for desirable performance and it can also be used to select between different circuit structures in the RF-DC power conversion **118**. This feedback configuration is preferred when the voltage limiter/regulation **1908** needs to directly adjust the output voltage level of the RF-DC power conversion **118** to prevent the voltage regulator circuit from getting over saturated in voltage or breaking down. The configuration can be use to aid in the tuning of the impedance matching networks in the RF-DC power conversion **118**. It can also be used for the purpose of triggering the charging the floating gate devices in the rectifier circuits to enhance power conversion efficiency or prevent degradation in the performance of the power harvester.

[0099] FIG. **21**(*d*) shows a diagram another embodiment of the power harvester **102** with voltage limiting/regulation **1908**, charge storage devices **1906** and power management **1910**. The antenna **110** collects the RF signal **106** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage device **1906***a* and its output charges a different charge storage device **1906***b*. Power management **1910** controls the conduction path between the output of voltage regulator **1908** and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed. Charge storage **1906***c* holds the excess charge from the power harvester when device is in operation and gives off charge when device **1904** is drawing more current than the power harvester circuit **112** (not shown) can provide. The feedback **2100** between the device **1904** and the voltage limiting/regulation **1908** allows the device **1904** to control the regulated output voltage level of the power harvester **102**. The feed back loop **2100** can be used to partially or fully enabling and disabling the voltage limiter/regulation **1908**. It can be to select between different devices **1904** or reference voltages in the voltage limiter/regulation **1908** to robustly vary the voltage level as seen from the device **1908**. This configuration is preferred when the device does not operate on a wide range power supply and requires the voltage regulation to be very accurate, preferably to within less than 50 mV of the desire operating voltage.

[0100] FIG. **21**(*e*) shows a diagram of a power harvester with voltage limiting/regulation **1908**, charge storage devices **1906** and power management **1910**. The antenna **110** collects the RF signal **106** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage **1906***a* device and its output charges a different charge storage device **1906***b*. Power management **1910** controls the conduction path between the output of voltage regulator **1908** and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed. Charge storage **1906***c* holds the excess charge from the power harvester **102** when device **1904** is in operation and gives off charge when device **1904** is drawing more current than the power harvester circuit **112** (not shown) can provide. The feedback **2100** between the power management **1910** and the voltage limiting/regulation **1908** allows power management circuits to control the regulated output voltage level of the power harvester **102**. The feed back loop **2100** can be used to partially or fully enabling and disabling the voltage limiter/regulation **1908**. It can be to select between different devices or reference voltages in the voltage limiter/regulation block to robustly vary the voltage level as seen from the device.

[0101] FIG. **21**(*f*) shows a diagram of a power harvester with voltage limiting/regulation **1908**, charge storage devices **1906** and power management **1910**. The antenna **110** collects the RF signal **106** (not shown), the RF-DC rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage device **1906***a* and its output charges a different charge storage device **1906***b*. Power management **1910** controls the conduction path between the output of voltage regulator and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed. Charge storage **1906***c* holds the excess charge from the power harvester **102** when device **1904** is in operation and gives off charge when device **1904** is drawing more current than the power harvester circuit can provide. The feedback **2100** between the device **1904** and the power management **1910** allows device **1904** some control of the power management circuits. The feed back loop **2100** can be used to partially or fully enabling and disabling the power management **1910**. This is desirable when the device **1904** operates frequently such that, power can be saved by turning off the power management circuit. The feedback **2100** also allows the control of how much charge is transferred from the power harvester **201** to the device **1904**.

[0102] It should be obvious that the above embodiments can be provided without the power management block **1910**, and without charge storage device **1906***c*, where the antenna **110** collects the RF signal **106** (not shown), the RF-DC rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage **1906***a* device and its output charges a different charge storage device **1906***b*. Without the power management circuit, the device **1904** is directly powered by the power from the voltage limiting/regulation **1908** and the charge stored on charge storage **1906***b*.

[0103] It should be obvious that the feedback **2100** provided in FIGS. **21**(*a*)-(*f*) can be applied in a similar manner without the power management block **1910**, and without charge storage device **1906***c*. The antenna **110** collects the RF signal **106** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage device **1906***a* and its output charges a different charge storage device **1906***b*. Without the power management **1910**, the device **1904** is directly powered by the power from the voltage limiting/regulation **1908** and the charge stored on charge storage **1906***b*. The feedback **2100** between the device **1904** and the RF-DC power conversion **118** gives the device **1904** the control the power harvester **102**. The feed back loop **2100** can be used to partially or fully enabling and disabling the RF-DC power conversion **118**. It can be used for tuning the RF-DC power conversion **118** for desirable performance and it can also be used to select between different circuit structures in the RF-DC power conversion **118**. This feedback **2100** configuration is preferred when the RF-DC power conversion **118** requires tuning for optimized matching or communication to the transmitter **104** (not shown) is required. The device **1904** can trigger the impedance matching network **204** (not shown) in the RF-DC **118** from matched to total mismatch. This enables a communication link to the transmitter **104** via backscattering of the incident RF signal. On a system where the antenna is shared, the feedback can disconnect and isolate the RF-DC block from the antenna.

[0104] In another embodiment provided without the power management block **1910**, and without charge storage device **1906***c* and having feedback **2100**, the antenna **110** collects the RF signal **106** (not shown), the RF-DC **118** rectifies the RF signal **106** to a DC voltage, the voltage limiter/regulator **1908** regulates the DC voltage and its output is available to an external device **1904**. The feedback **2100** between the voltage limiting/regulation **1908** and the RF-DC power conversion **118** allows the voltage limiting/regulation **1908** control of power harvester **102**. The feedback loop **2100** can be used to partially or fully enabling and disabling the RF-DC power conversion **1910**. It can be used for tuning the RF-DC power conversion **118** for desirable performance and it can also be used to select between different circuit structures in the RF-DC power conversion **118**. This feedback **2100** configuration may be preferred when the voltage limiter/regulation **1908** needs to directly adjust the output voltage level of the RF-DC power conversion **118** to prevent the voltage regulator **1908** from getting over saturated in voltage or breaking down. The configuration can be use to aid in the tuning of the impedance matching networks in the RF-DC power conversion **118**. It can also be used for the purpose of triggering the charging of the floating gate devices in the rectifier circuits to enhance power conversion efficiency or prevent degradation in the performance of the power harvester.

[0105] In another embodiment provided without the power management block **1910**, and without charge storage device **1906***c* and having feedback **2100**, the antenna **110** collects the RF signal **106** (not shown), the RF-DC **118** rectifies the RF signal **106** to a DC voltage, that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage device **1906***a* and its output charges a different charge storage device **1906***b*. Without the power management circuit, the device is directly powered by the power from the voltage limiting/regulation **1908** and the charge stored on charge storage **1906***b*. The feedback between the voltage limiting/regulation **1908** and the RF-DC power conversion **118** allows the voltage limiting/regulation **1908** control of power harvester **102**. The feed back loop **2100** can be used to partially or fully enabling and disabling the RF-DC power conversion **118**. It can be used for tuning the RF-DC power conversion **118** for desirable performance and it can also be used to select between different circuit structures in the RF-DC power conversion **118**. This feedback **2100** configuration is preferred when the voltage limiter/regulation **1908** needs to directly adjust the output voltage level of the RF-DC power conversion **118** to prevent the voltage regulator circuit **1908** from getting over saturated in voltage or breaking down. The configuration can be use to aid in the tuning of the impedance matching networks in the RF-DC power conversion **118**. It can also be used for the purpose of triggering the charging of the floating gate devices in the rectifier circuits to enhance power conversion efficiency or prevent degradation in the performance of the power harvester.

[0106] In another embodiment provided without the power management block **1910**, and without charge storage device **1906***c* and having feedback **2100**, the antenna **110** collects the RF signal **106** (not shown), the RF-DC **118** rectifies the RF signal **106** to a DC voltage, that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage device **1906***a* and its output charges a different charge storage device **1906***b*. Without the power management circuit, the device **1904** is directly powered by the power from the voltage limiting/regulation circuit **1908** and the charge stored on charge storage **1906***b*. The feedback **2100** between the device **1904** and the voltage limiting/regulation **1908** allows the device **1904** to control the regulated output voltage level of the power harvester **102**. The feed back loop **2100** can be used to partially or fully enabling and disabling the voltage limiter/regulation **1908**. It can select between different devices **1904** or reference voltages in the voltage limiter/regulation **1908** to robustly vary the voltage level as seen from the device **1904**. This configuration is preferred when the device **1904** does not operate on a wide range power supply and requires the voltage regulation to be very accurate, preferably to within less than 50 mV of the desire operating voltage.

[0107] It should be obvious that the above embodiments can be provided without the power without charge storage device **1906***c*. Charge storage after the power management block **1910** is not necessary when the conduction path between the power harvester output (not shown) and device **1904** is strongly on, and sufficient power can be delivered through this conduction path. The antenna **110** collects the RF signal **104** (not shown), the RF-DC **188** rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage device **1906***a* and its output charges a different charge storage device **1906***b*. Power management **1910** controls the conduction path between the output of voltage regulator **1908** and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed.

[0108] Charge storing **1906***b* after the voltage limiting/regulation **1908** is not necessary when the power management **1910** is turned on often, thus charging charge **1906***b*. The antenna **110** collects the RF signal, the RF-DC rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*, the voltage limiter/regulator **1908** regulates the DC voltage from the charge storage device **1906***a* and its output is feeding to the power management **1910**. Power management **1910** controls the conduction path between the output of voltage regulator **1908** and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed. Charge storage **1906***c* holds the excess charge from the power harvester **102** when device **1904** is in operation and gives off charge when device **1904** is drawing more current than the power harvester **102** can provide.

[0109] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks and without charge storage **1096***b* after regulator **1908** are possible.

[0110] Charge storage **1906***a* is not needed after the RF-DC power conversion **118** when the power received by the RF-DC power conversion **118** is stable or when the power requirement of the voltage limiting/regulation **1908** does not exceed the power harvested up to RF-DC conversion. The antenna **110** collects the RF signal **104** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage and is fed directly to the voltage limiter/regulation block **1908**, the voltage limiter/regulator **1908** regulates the DC voltage from the output of RF-DC **118** and its output charges a charge storage device **1908***b*. Power management controls **1910** the conduction path between the output of voltage regulator **1908** and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed. Charge storage **1906***c* holds the excess charge from the power harvester when device **1904** is in operation and gives off charge when device **1904** is drawing more current than the power harvester circuit **102** can provide.

[0111] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks and without charge storage **1096***a* after regulator **1908** are possible.

[0112] In cases where the device **1904** operates at peak power that does not exceed harvested power and/or the regulated power, charge storage device **1906***b* after the voltage limiter/regulator **1908** and charge storage device **1906***c* after regulation and power management **1910** are not require as part of the power harvester **102**. The antenna **110** collects the RF signal **104** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*, the voltage regulator **1908** regulates the DC voltage on charge storage **1906***a*. Power management **1910** controls the conduction path between the output of voltage regulator to device and turn on the conduction switch when power is needed at the device **1904**.

[0113] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks and without charge storage **1096***b* after regulator **1908** and **1906***c* after power management **1910** are possible.

[0114] In another embodiment, the charge storage devices **1906***a* and **1906***c* are not implemented. The antenna **110** collects the RF signal **104** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage that is then applied to the voltage limiting/regulation **1908**. The voltage limiter/regulator **1908** regulates the DC voltage from the RF-DC power conversion **118** and its output charges charge storage device **1906***b*. Power management **1910** controls the conduction path between the output of voltage regulator **1908** and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed.

[0115] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks and without charge storage **1096***a* after RF converter **118** and **1906***c* after power management **1910** are possible.

[0116] In another embodiment, the charge storage devices **1906***a* and **1906***b* are not implemented. The antenna **110** collects the RF signal **104** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage that is then applied to the input of the voltage limiter/regulation **1908**. The voltage limiter/regulator **1908** regulates the DC voltage from the output of the RF-DC power conversion circuit **118** and its output is fed to the power management **1910**. Power management **1910** controls the conduction path between the output of voltage regulator **1908** and the device **1904**. Power management **1910** supplies power to the device **1904** when it is needed. Charge storage **1906***c* holds the excess charge from the power harvester **210** when device **1904** is in operation and gives off charge when device **1904** is drawing more current than the power harvester circuit **102** can provide.

[0117] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks and without charge storage **1096***a* after RF converter **118** and **1906***b* after voltage limiter/regulator **1908** are possible.

[0118] In another embodiment, the charge storage devices **1906***a*, **1906***b* and **1906***c* are not implemented. In cases that device **1904** operates once within a time interval and/or at power level lower than or equal to the harvested power from the RF environment, charge storage devices **1906***a* **1906***b* and **1906***c* are not require as part of the power harvester **102**. The antenna **101** collects the RF signal **104** (not shown), the RF-DC rectifies the RF signal to a DC voltage, the voltage regulator regulates the DC voltage from the output of RF-DC. Power management **1910** controls the conduction path between the output of voltage regulator **1908** to device **1904** and turn on the conduction switch when power is needed at the device **1904**.

[0119] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks and without charge storage **1096***a* after RF converter **118**, **1906***b* after voltage limiter/regulator **1908**, and **1906***c* after power manager **1910** are possible.

[0120] In another embodiment, the charge storage devices **1906***b* and **1906***c* are not implemented, and power management **1910** is not implemented. In cases where the device **1904** operates at peak power that does not exceed harvested power and/or the regulated power, power management **19019** and charge storage device **1906***c* and charge storage device **1906***b* after voltage regulation **1908** is not require as part of the power harvester **102**. The antenna **110** collects the RF signal **104** (not shown), the RF-DC rectifies the RF signal to a DC voltage that is stored on a charge storage device (charge storage **1**), the voltage regulator regulates the DC voltage on **1906***a*. The voltage regulator supplies **1908** the power directly to the device **1904**.

[0121] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks **2100** and without the charge storage devices **1906***b* and **1906***c*, and power management **1910** are possible.

[0122] In another embodiment, the charge storage devices **1906***a* and **1906***c* are not implemented, and power management **1910** is not implemented. In cases that device **1904** operates once within a time interval and/or at power level lower than or equal to the harvested power from the RF environment, power management **1910** and charge storage **1906***c* after power management **1910**, and charge storage **1906***a* after RF-DC **118** are not require as part of the power harvester **102**. The antenna **110** collects the RF signal **104** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage, the voltage regulator **1908** regulates the DC voltage from the output of RF-DC **118**, that is provided to device **1904**.

[0123] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks **2100** and without the charge storage devices **1906***a* and **1906***c*, and power management **1910** are possible.

[0124] In another embodiment, the charge storage devices **1906***a*, **1906***b* and **1906***c* are not implemented, and power management **1910** is not implemented. In cases that device **1904** operates once within a time interval and/or at power level lower than or equal to the harvested power from the RF environment, charge storage devices **1906***a*, **1906***b* and **1906***c* and power management **1910** is not require as part of the power harvester **102**. The antenna **110** collects the RF signal **104** (not shown), the RF-DC rectifies the RF signal to a DC voltage, the voltage regulator regulates the DC voltage from the output of RF-DC and provides it to device **1904** and turns on a conduction switch (not shown) when power is needed at the device **1904**.

[0125] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks **2100** and without the charge storage devices **1906***a*, **1906***b* and **1906***c*, and power management **1910** are possible.

[0126] In another embodiment, the charge storage device **1906***b* and voltage limiter/regulator **1908** are not implemented. In cases that device **1904** operates once within a time interval and/or at power level lower than or equal to the harvested power from the RF environment, charge storage device **1906***b* is not require as part of the power harvester. The antenna collects the RF signal, the RF-DC rectifies the RF signal to a DC voltage, the voltage is provided to power manager **1910** which can store needed power in storage device **1906***c* for powering device **1904**. The power management **1910** also supplies the power directly to the device **1904**.

[0127] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks **2100** and without the charge storage devices **1906***b* after voltage limiter/regulator **1908** and voltage limiter/regulator **1908** are possible.

[0128] In another embodiment, the charge storage devices **1906***b* and **1906***c*, and voltage limiter/regulator **1908** are not implemented. In cases where the device **1904** is able to tolerate wide power supply ranges, charge storage devices **1906***b* and **1906***c*, and voltage limiter/regulator **1908** are not require as part of the power harvester **102**. The antenna **110** collects the RF signal **104** (not shown), the RF-DC rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*. Power management **1910** controls the conduction path between the output of RF-DC **118** to device **1904** and turn on a conduction switch (not shown) when power is needed at the device **1904**.

[0129] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks **2100** and without the charge storage devices **1906***b* after voltage limiter/regulator **1908** and charge storage devices **1906***b* after power manager **1910**, and voltage limiter/regulator **1908** are possible.

[0130] In another embodiment, the charge storage devices **1906***a* and **1906***b*, and voltage limiter/regulator **1908** are not implemented. In cases where the device **1904** is able to tolerate wide power supply ranges, charge storage devices **1906***a* and **1906***b*, and voltage regulator **1908** are not require as part of the power harvester **102**. The antenna **110** collects the RF signal **104** (not shown), the RF-DC **118** rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***c*. Power management **1910** controls the conduction path between the output of RF-DC **118** to device **1908** and turn on a conduction switch (not shown) when power is needed at the device **1904**.

[0131] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks **2100** and without the charge storage devices **1906***a* and **1906***b*, and voltage limiter/regulator **1908** are possible.

[0132] In another embodiment, the charge storage devices **1906***a*, **1906***b* and **1906***c*, and voltage limiter/regulator **1908** are not implemented. In cases where the device **1904** is able to tolerate wide power supply ranges, with peak power less than the harvestable energy, the charge storage devices **1906***a*, **1906***b* and **1906***c*, and voltage limiter/regulator **1908** are not require as part of the power harvester **102**. The antenna **110** collects the RF signal **104** (not shown), the RF-DC rectifies the RF signal to a DC. In this embodiment, power management **1910** may control the conduction path between the output of the RF-DC **118** to external device **1904**. The power management **1910** may responsively "turn on" the conduction when power is needed at the device **1904** or as instructed (for example, by the external device and/or other device).

[0133] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks **2100** and without charge storage devices **1906***a*, **1906***b* and **1906***c*, and voltage limiter/regulator **1908** are possible.

[0134] In another embodiment, the charge storage devices **1906***b* and **1906***c*, and voltage limiter/regulator **1908** and power management **1910** are not implemented. In cases where the device **1904** is able to tolerate wide power supply range, the charge storage devices **1906***b* and **1906***c*, and voltage limiter/regulator **1908** and power management **1910** are not require as part of the power harvester **102**. The antenna **110** collects the RF signal **104** (not shown), the RF-DC rectifies the RF signal to a DC voltage that is stored on a charge storage device **1906***a*. The device **1904** is powered by the charge stored on charge storage **1906***a*.

[0135] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks **2100** and without the charge storage devices **1906***b* and **1906***c*, and voltage limiter/regulator **1908** and power management **1910** are possible.

[0136] In another embodiment, the charge storage devices **1906***a*, **1906***b* and **1906***c*, and voltage limiter/regulator **1908** and power management **1910** are not implemented. This embodiment can be advantageous where the external device **1904** is able to tolerate wide power supply range a, with peak power less than the harvestable energy, the charge storage devices **1906***a*, **1906***b* and **1906***c*, and voltage limiter/regulator **1908** and power management **1910** are not require as part of the power harvester **102**. In this embodiment, the antenna **110** collects the RF signal **104** (not shown), the RF-DC rectifies the RF signal to a DC. The device **1904** draws power directly from the output of RF-DC power conversion circuitry **118**.

[0137] It should be obvious that all different possible feedback **2100** configurations for power harvester **102** with feedbacks **2100** and without the charge storage devices **1906***a*, **1906***b* and **1906***c*, and voltage limiter/regulator **1908** and power management **1910** are possible.

[0138] FIG. **22** shows a block diagram of one embodiment of power harvester, according to the current invention, where shown is the antenna **110** connected to the impedance match **204**, which is connected to the rectifier **202** of the RF-DC converter **118**. In cases where the device **1904** (not shown) is able to tolerate wide power supply range, with peak power less than the harvestable energy, the charge storage devices **1906***a*, **1906***b* and **1906***c*, and voltage limiter/regulator **1908** and power management **1910** not require as part of the power harvester. The RF signal **104** (not shown) is collected at the antenna **110** and passes through impedance matching networks **204** before reaching the rectifier **202**. In this embodiment, the impedance matching circuitry **204** may provide an enhanced and/or predetermined impedance match between the output of the antenna **110** and the input of the RF-DC converter **118**. For example, the impedance matching circuitry may be high-Q to passively increase the voltage amplitude to the rectifier or it can be low-Q to increase the bandwidth of harvestable RF signal. In certain embodiment, it may be preferred that the impedance be matched to within 5% of the ideal impedance value to achieve higher passive voltage gain from high-Q matching and higher power from low-Q matching.

[0139] The present invention has now been described in accordance with several exemplary embodiments, which are intended to be illustrative in all aspects, rather than restrictive. Thus, the present invention is capable of many variations in detailed implementation, which may be derived from the description contained herein by a person of ordinary skill in the art.

[0140] All such variations are considered to be within the scope and spirit of the present invention as defined by the following claims and their legal equivalents.

What is claimed:

1. A voltage doubler rectifier RF power harvesting system comprising:

a. at least one power harvesting module, wherein said power harvesting module comprises:
  i. a voltage doubler rectifier structure circuit, wherein said voltage doubler rectifier structure comprises:
    1. a DC voltage input;
    2. a DC voltage output;
    3. a signal input;
    4. a first floating gate transistor and a second floating gate transistor, wherein said floating gate transistors are connected in series between said DC voltage input and said DC voltage output, wherein a gate control of said floating gate is connected to a drain of said transistor;
    5. a first capacitor disposed between said input signal and said drain of said first floating gate transistor; and
    6. a second capacitor disposed between a ground and said drain of said second floating gate transistor;

b. a harvester antenna, wherein said harvester antenna is disposed to provide a signal to said signal input;

c. a transmitter;

d. a transmitter antenna, wherein said transmitter provides said signal for output from said transmitter antenna, wherein a full-wave peak-to-peak voltage of an incoming RF signal from said transmitter antenna to said signal input is rectified; and

e. a powered device, wherein said powered device is connected to said DC voltage output to utilize said rectified signal.

2. The RF power harvesting system of claim **1**, wherein said at least one power harvesting module further comprises at least one charge storage device, wherein said charge storage device is selected from a group consisting of capacitors, rechargeable batteries, and non-rechargeable batteries.

3. The RF power harvesting system of claim **1**, wherein said at least one power harvester further comprises a voltage limiting/regulation device, wherein said voltage limiting/regulation device is disposed between said voltage double rectifier structure and said powered device.

4. The RF power harvesting system of claim **3**, wherein said at least one power harvesting module further comprises at least one charge storage device, wherein said charge storage device is selected from a group consisting of capacitors, rechargeable batteries, and non-rechargeable batteries.

5. The RF power harvesting system of claim **3**, further comprises feedback loops wherein said feedback loops are feedback connections selected from a group consisting of said powered device to said voltage doubler rectifier structure, said powered device to said voltage limiting/regulation device, and said voltage doubler rectifier structure to said voltage limiting/regulation device.

6. The RF power harvesting system of claim **5**, wherein said at least one power harvesting module further comprises at least one charge storage device, wherein said charge storage device is selected from a group consisting of capacitors, rechargeable batteries, and non-rechargeable batteries.

7. The RF power harvesting system of claim **1**, wherein said at least one power harvesting module further comprises a power management device, wherein said power management device is disposed between said voltage doubler rectifier structure and said powered device.

8. The RF power harvesting system of claim **7**, wherein said at least one power harvesting module further comprises at least one charge storage device, wherein said charge storage device is selected from a group consisting of capacitors, rechargeable batteries, and non-rechargeable batteries.

9. The RF power harvesting system of claim **7**, wherein said at least one power harvesting module further comprises feedback loops wherein said feedback loops are feedback connections selected from a group consisting of said powered device to said voltage doubler rectifier structure, said powered device to said power management device, and said power management device to said voltage doubler rectifier structure.

10. The RF power harvesting system of claim **7**, wherein said at least one power harvesting module further comprises at least one charge storage device, wherein said charge storage device is selected from a group consisting of capacitors, rechargeable batteries, and non-rechargeable batteries.

11. The RF power harvesting system of claim **1**, wherein said at least one power harvesting module further comprises a power management device and a voltage limiting/regulation device, wherein said power management device and said voltage limiting/regulation device are disposed between said voltage doubler rectifier structure and said powered device.

12. The RF power harvesting system of claim **11**, wherein said at least one power harvesting module further comprises at least one charge storage device, wherein said charge storage device is selected from a group consisting of capacitors, rechargeable batteries, and non-rechargeable batteries.

13. The RF power harvesting system of claim **10** further comprises feedback loops wherein said feedback loops are feedback connections selected from a group consisting of said powered device to said voltage doubler rectifier structure, said powered device to said voltage limiting/regulation device, said powered device to said power management device, said power management device to said voltage limiting/regulation device, said power management device to said voltage doubler rectifier structure, and said voltage/regulation device to said voltage doubler rectifier structure.

14. The RF power harvesting system of claim **11**, wherein said at least one power harvesting module further comprises at least one charge storage device, wherein said charge storage device is selected from a group consisting of capacitors, rechargeable batteries, and non-rechargeable batteries.

15. The RF power harvesting system of claim **1**, wherein said at least one power harvester further comprises an impedance matching device disposed between said harvester antenna and said voltage doubler rectifier.

16. The RF power harvesting system of claim **15**, wherein said impedance matching is an adjustable impedance matching, wherein said adjustment is programmable.

17. The RF power harvesting system of claim **16**, wherein said adjustable impedance matching is programmable.

18. The RF power harvesting system of claim **1**, wherein said at least one power harvester further comprises passive components disposed between said signal input and said harvesting antenna, whereby a wider band match is provided along a cascade of said power harvesters.

19. The RF power harvesting system of claim **1**, wherein said floating gate transistor is selected from a group consisting of a PMOS-PMOScap, a PMOS-NMPScap, an NMOS-NMOScap, an NMOS-PMOScap, an NMOS-MiMcap, and a PMPS-MiMcap.

20. The RF power harvesting system of claim **1**, wherein said harvesting antenna is a meanderline antenna, wherein said meanderline antenna comprises an inner loop disposed between a pair of approximately matching outer loops, wherein said inner loop comprises a pair of antenna ports, wherein an input impedance of said meanderline antenna matches an in impedance of said voltage doubler rectifier.

21. The RF power harvesting system of claim **20** wherein said antenna outer loop comprises a perimeter of 35 centimeters.

22. The RF power harvesting system of claim **22** wherein said antenna inner loop comprises a perimeter to match said input impedance of said rectifier.

23. The RF power harvesting system of claim **23** wherein said antenna matching outer loops are disposed to divide said antenna into two identical parts as a virtual ground, wherein a fully differential signal is provided at said antenna ports.

24. The RF power harvesting system of claim **1**, wherein said voltage doubler rectifier structure is disposed in at least a two-stage rectifier cascade.

25. The RF power harvesting system of claim **24**, wherein said cascade is a 36-stage cascade.

26. The RF power harvesting system of claim **24**, wherein said cascade is a 16-stage cascade.

* * * * *

US 20120106103A1

(19) **United States**
(12) **Patent Application Publication**
Nohra
(10) **Pub. No.: US 2012/0106103 A1**
(43) **Pub. Date: May 3, 2012**

(54) **RADIO FREQUENCY ENERGY HARVESTING ENCLOSURE FOR RADIO FREQUENCY CONNECTED DEVICES**

(76) Inventor: **Tanios Nohra**, Broadview Heights, OH (US)

(21) Appl. No.: **13/163,707**

(22) Filed: **Jun. 19, 2011**

Related U.S. Application Data

(60) Provisional application No. 61/357,590, filed on Jun. 23, 2010.

Publication Classification

(51) **Int. Cl.**
H05K 7/00 (2006.01)

(52) **U.S. Cl.** **361/752**

(57) **ABSTRACT**

A Radio frequency energy harvester enclosure and protective enclosure for Radio Frequency connected devices that harvests and recycles some of the transmitted Radio Frequency energy for the activation of an application circuit, or to additionally charge a storage component.

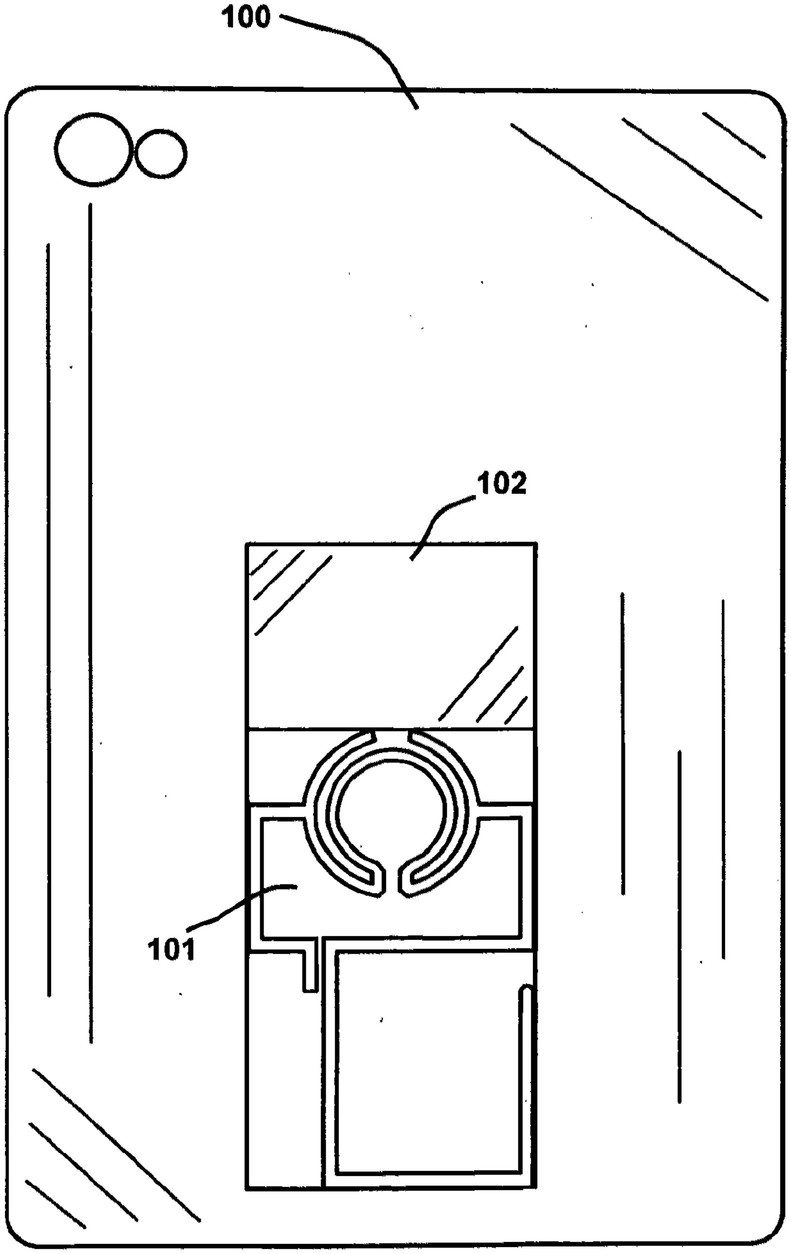

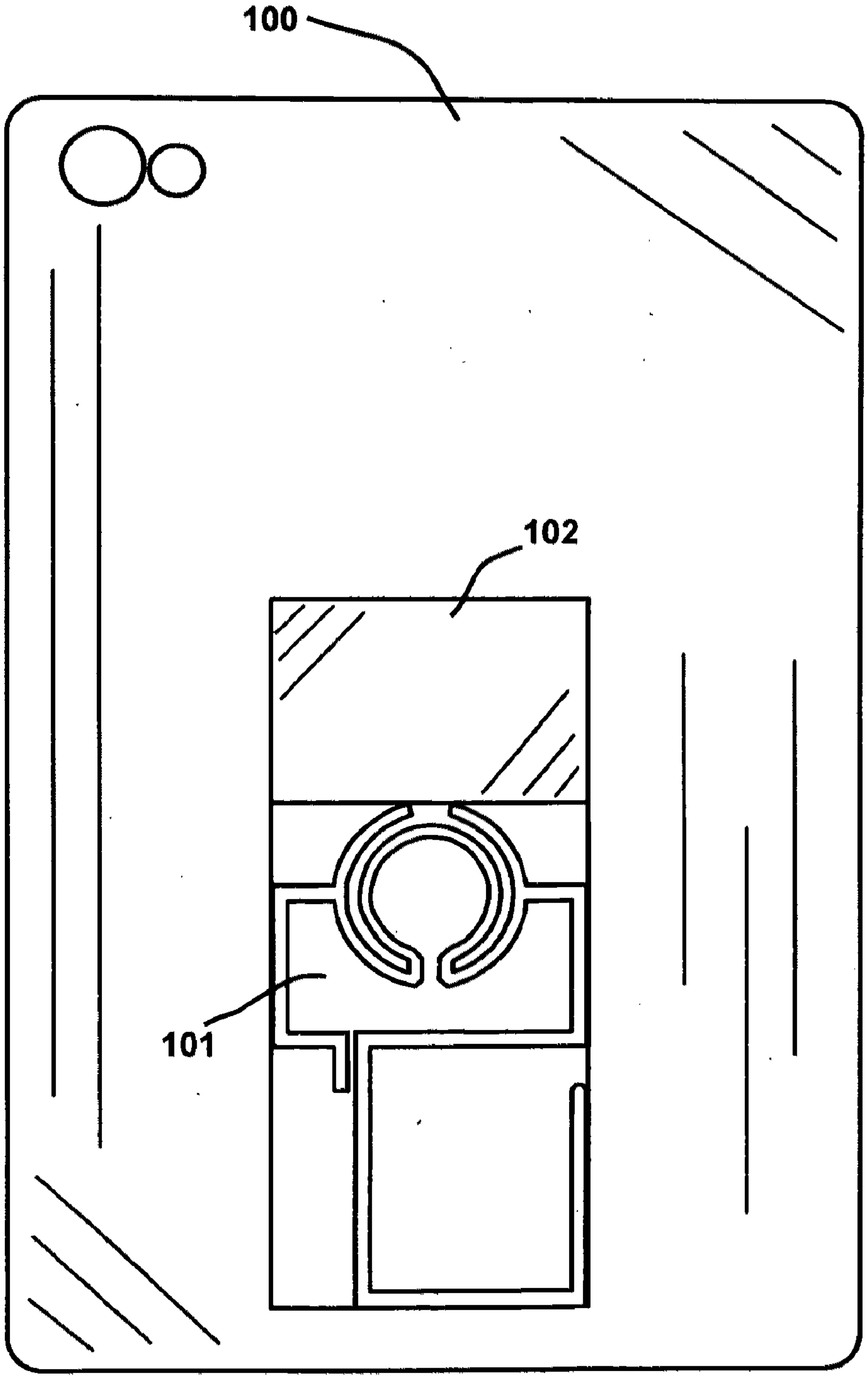

FIG. 1A

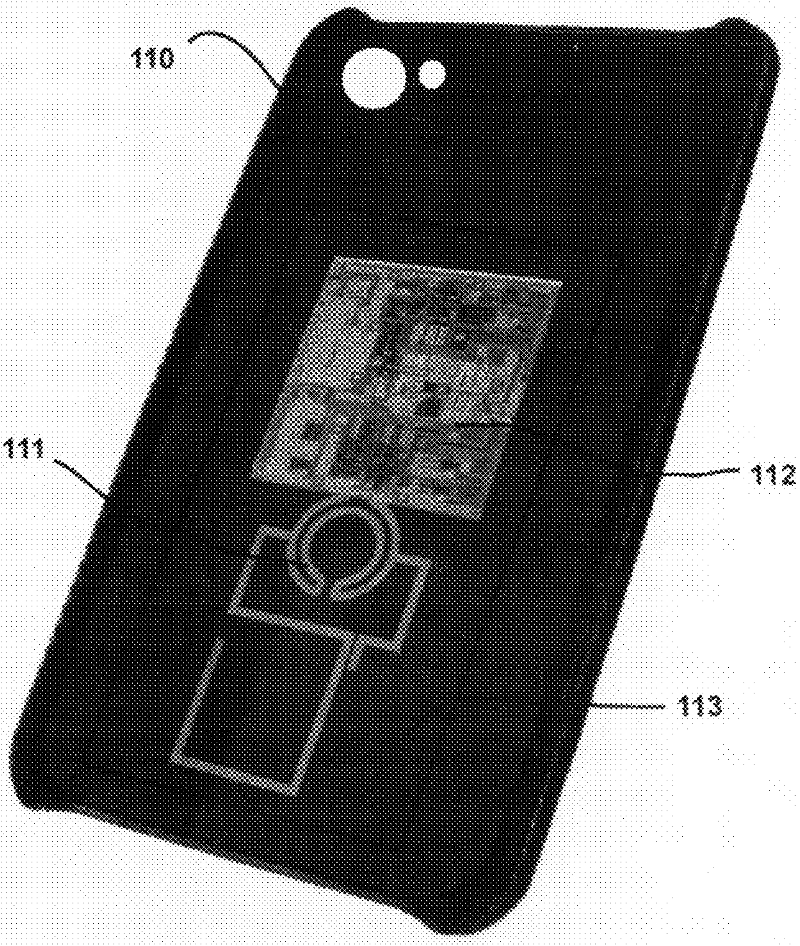
110
111
112
113

FIG. 1B

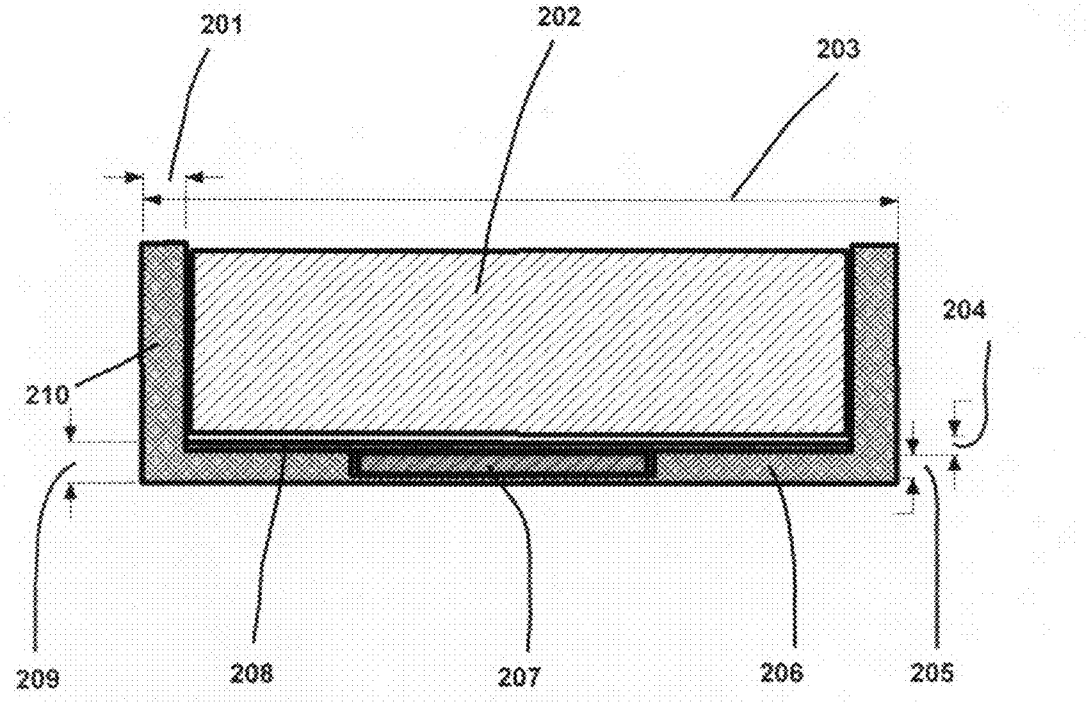
201
202
203
204
210
209
208
207
206
205

FIG. 2

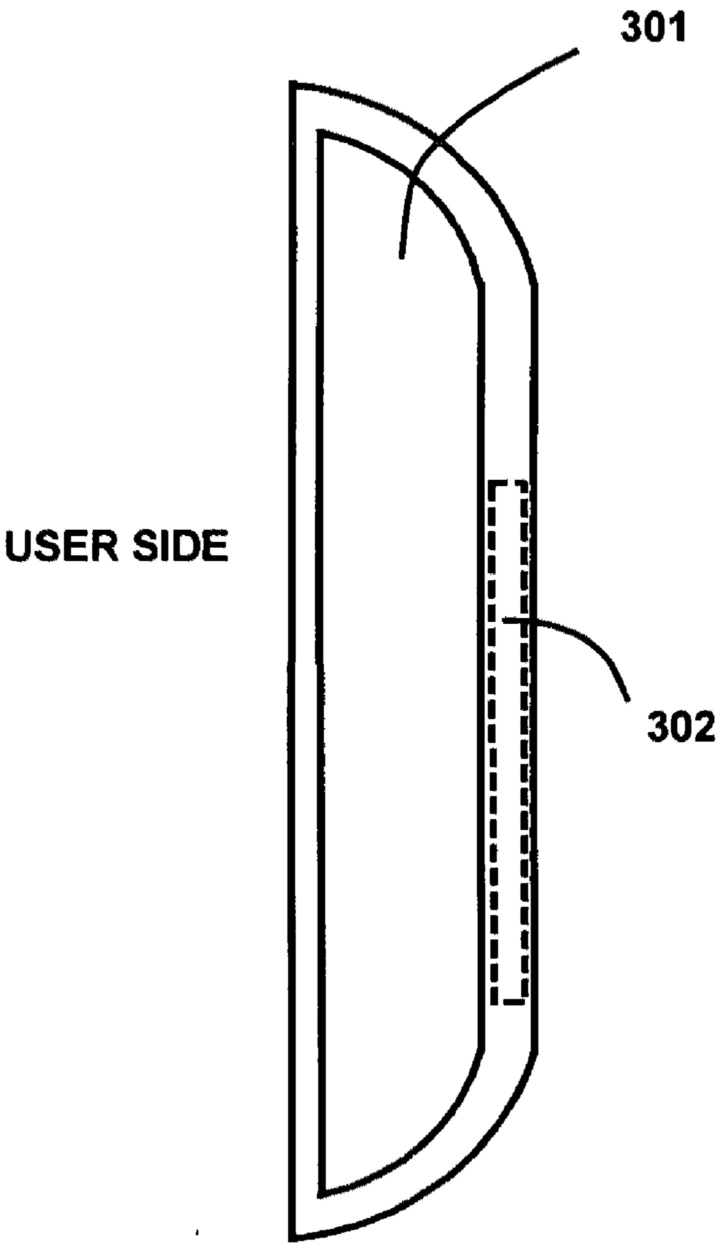
301
USER SIDE
302

FIG. 3

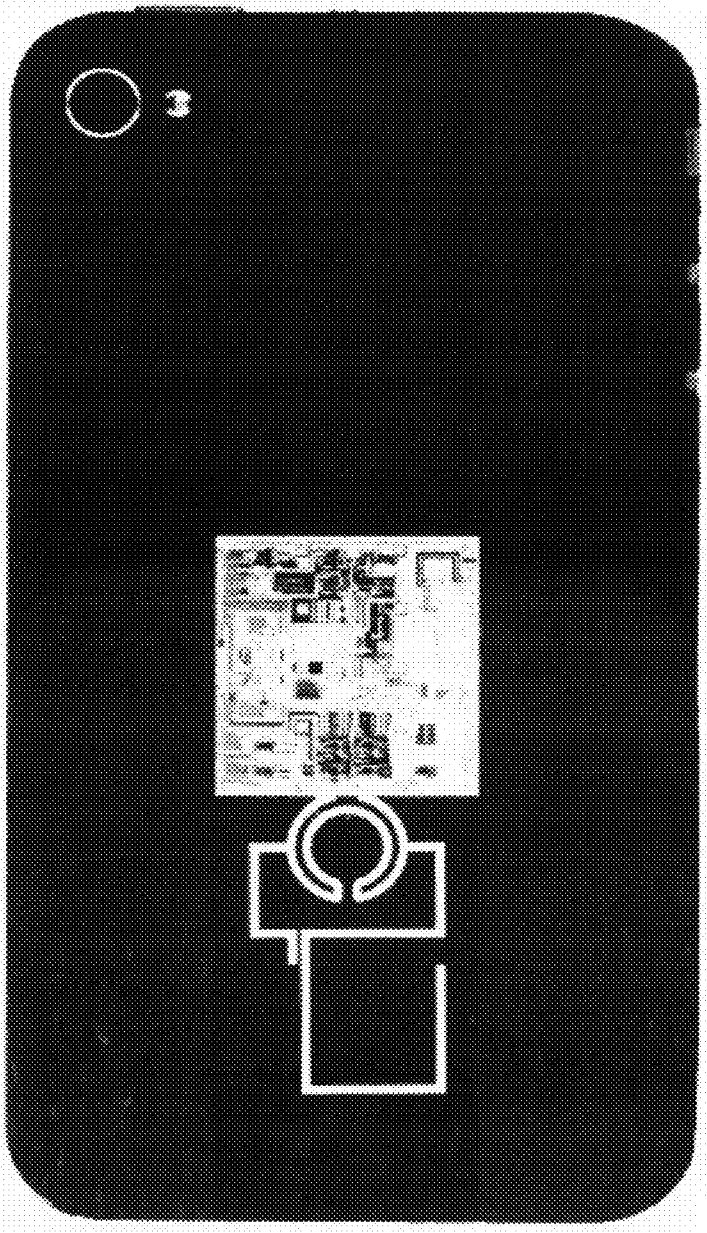

FIG. 4

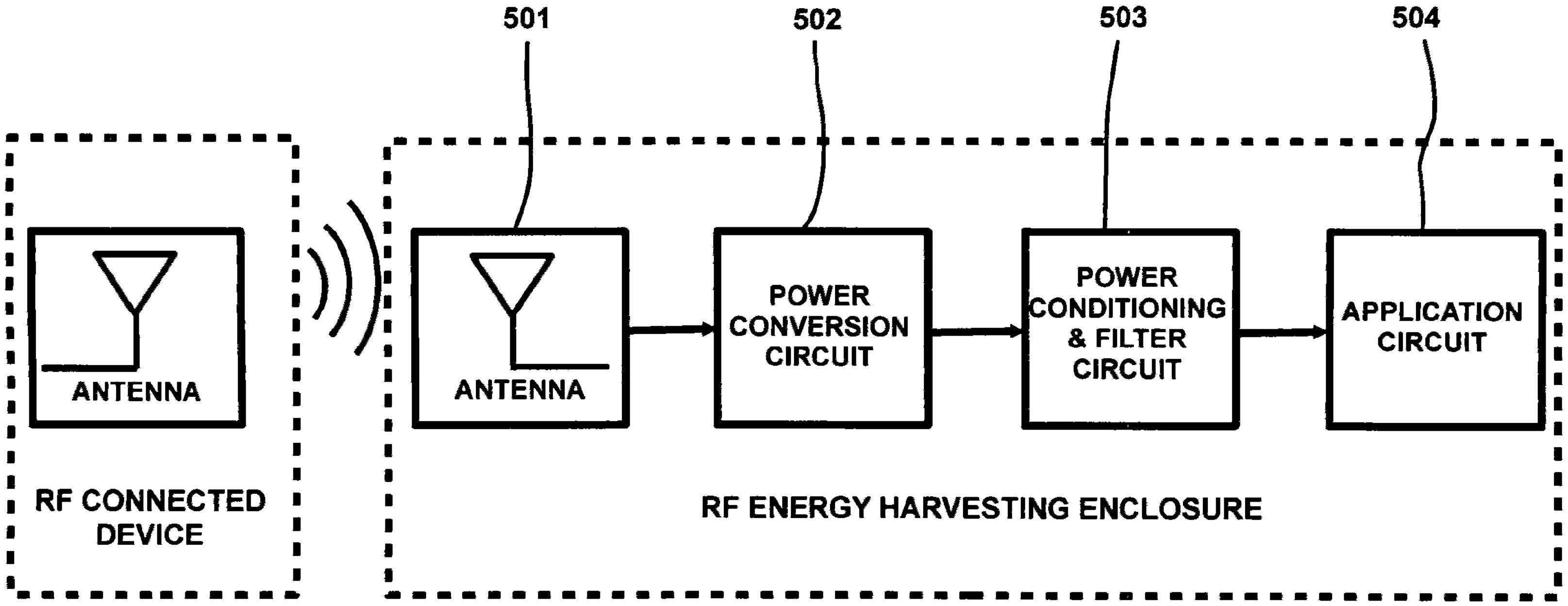
501
502
503
504
ANTENNA
RF CONNECTED DEVICE
ANTENNA
POWER CONVERSION CIRCUIT
POWER CONDITIONING & FILTER CIRCUIT
APPLICATION CIRCUIT
RF ENERGY HARVESTING ENCLOSURE

FIG. 5

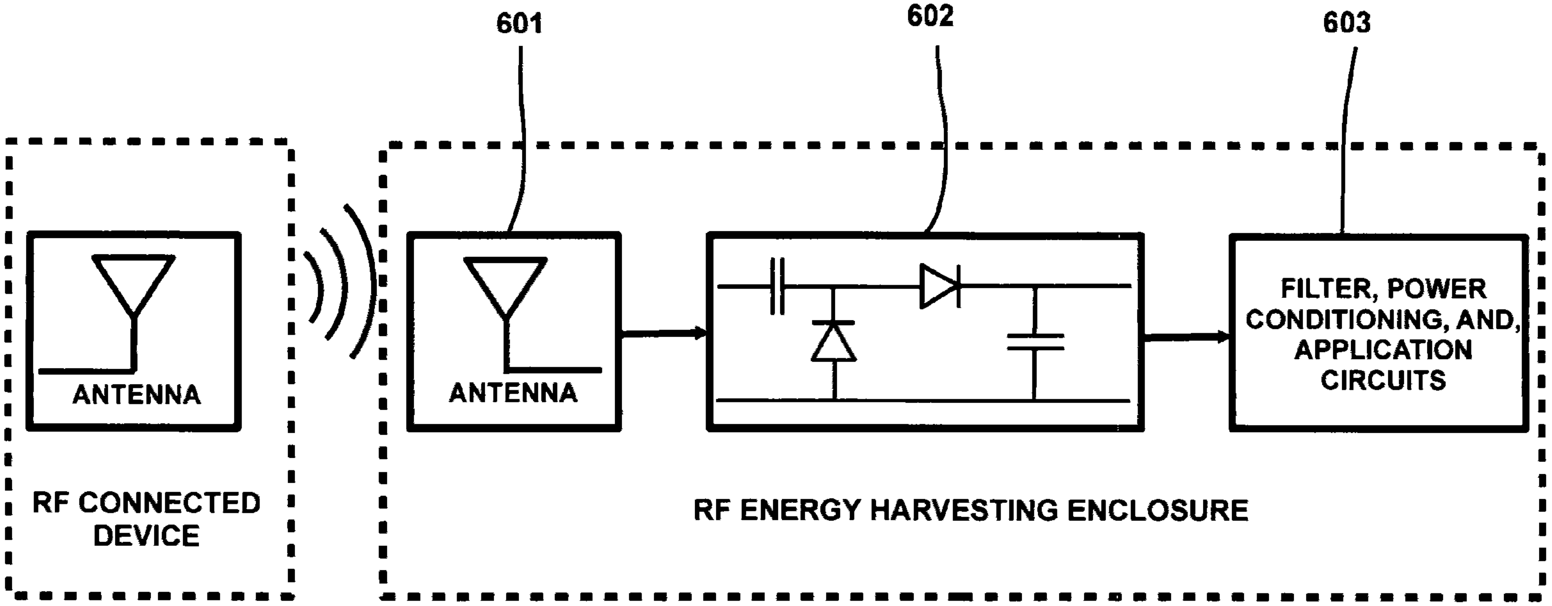
601
602
603
ANTENNA
RF CONNECTED DEVICE
ANTENNA
FILTER, POWER CONDITIONING, AND, APPLICATION CIRCUITS
RF ENERGY HARVESTING ENCLOSURE

FIG. 6

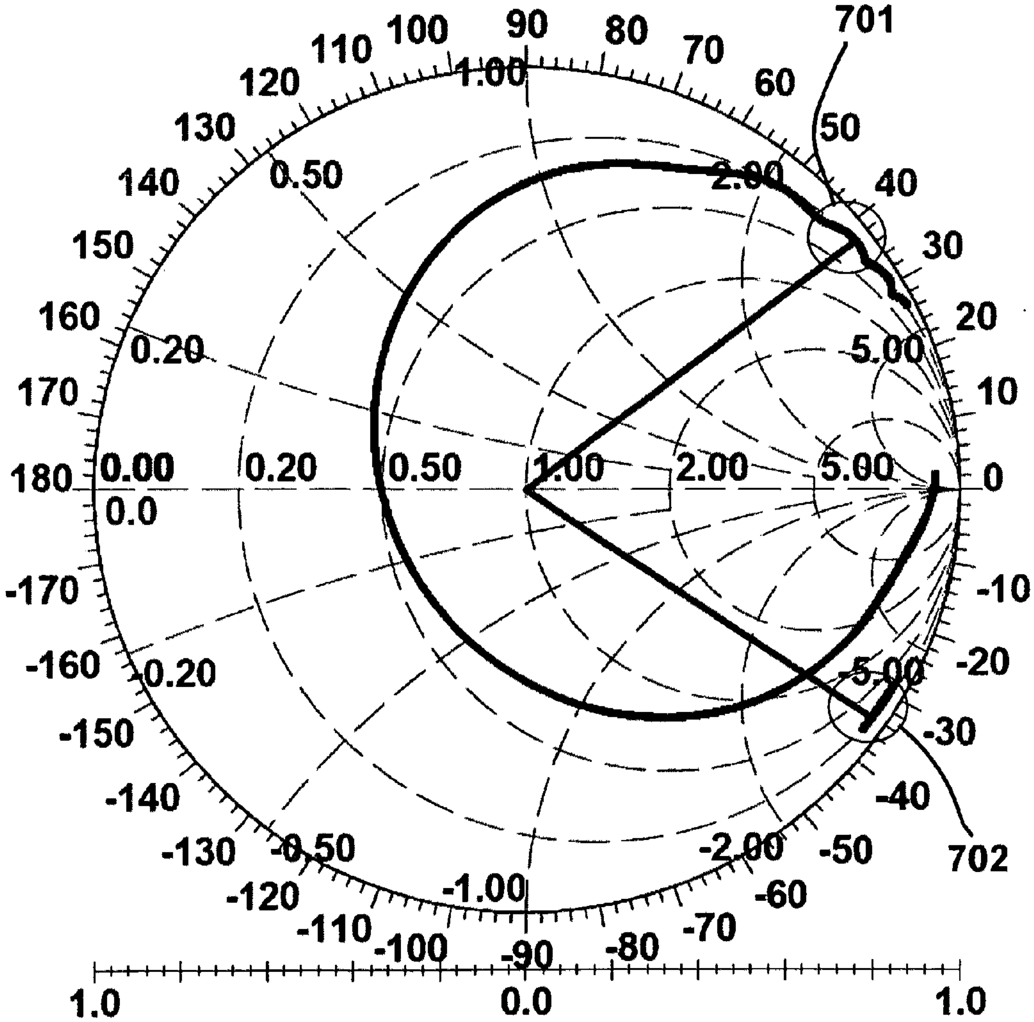
701
702
90
80
70
60
50
40
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-100
-110
-120
-130
-140
-150
-160
-170
180
170
160
150
140
130
120
110
100
1.00
0.50
2.00
0.20
5.00
0.00
0.20
0.50
1.00
2.00
5.00
0.0
-0.20
-5.00
-0.50
-2.00
-1.00
1.0
0.0
1.0

FIG. 7A

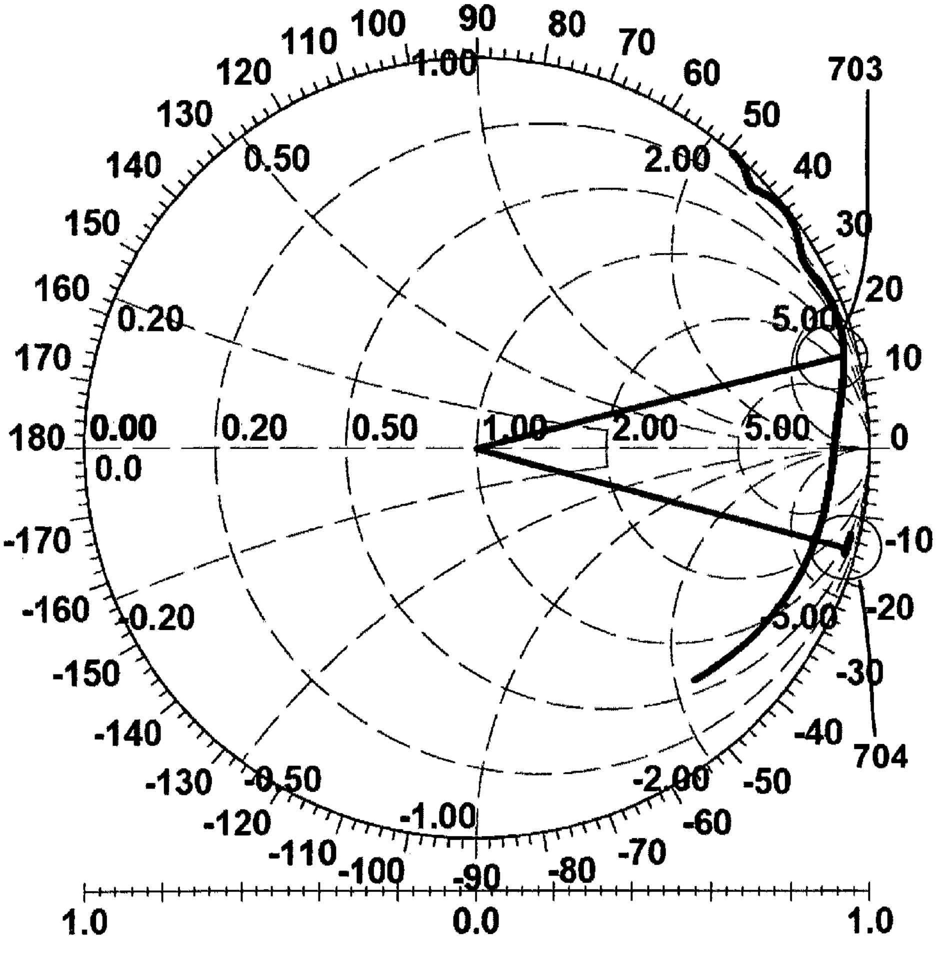
90
80
70
60
50
40
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-100
-110
-120
-130
-140
-150
-160
-170
180
170
160
150
140
130
120
110
100
1.00
0.50
2.00
5.00
0.20
0.00
0.0
0.20
0.50
1.00
2.00
5.00
-5.00
-0.20
-0.50
-2.00
-1.00
703
704
1.0
0.0
1.0

FIG. 7B

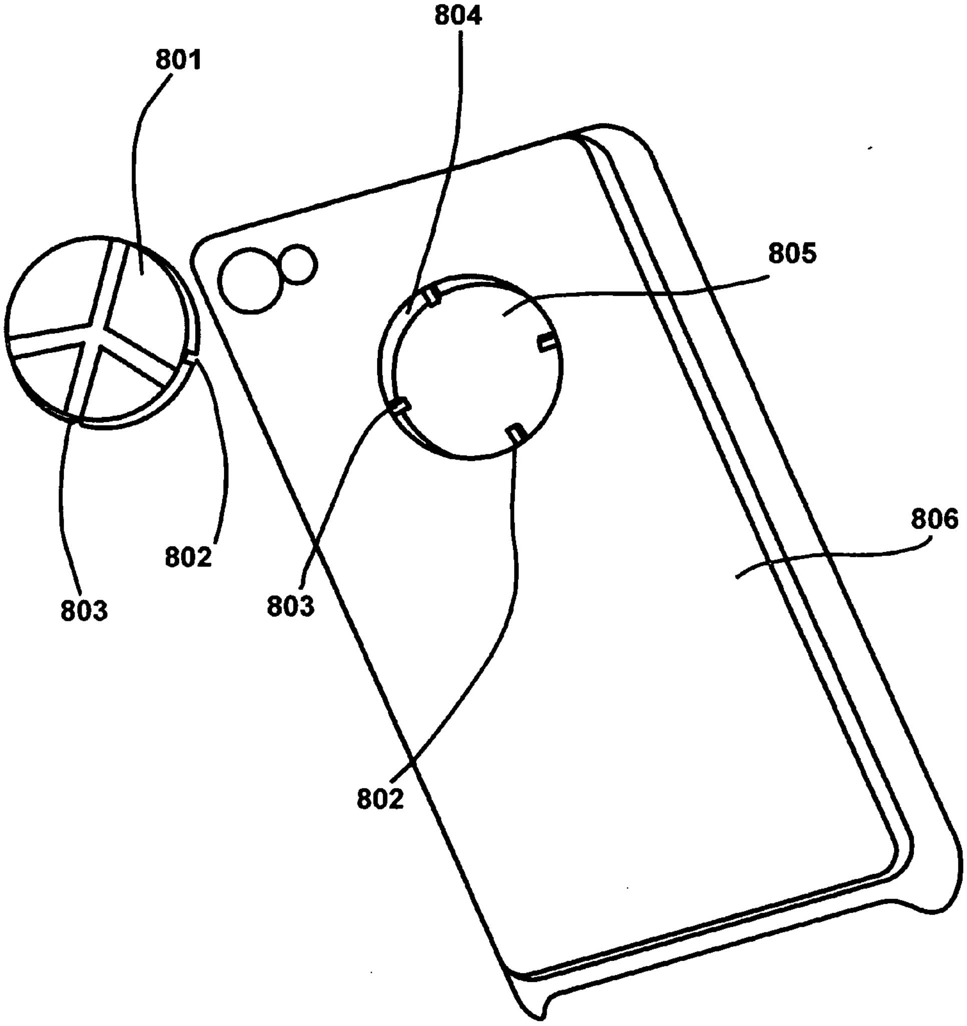
801
804
805
806
802
803
803
802
802

FIG. 8

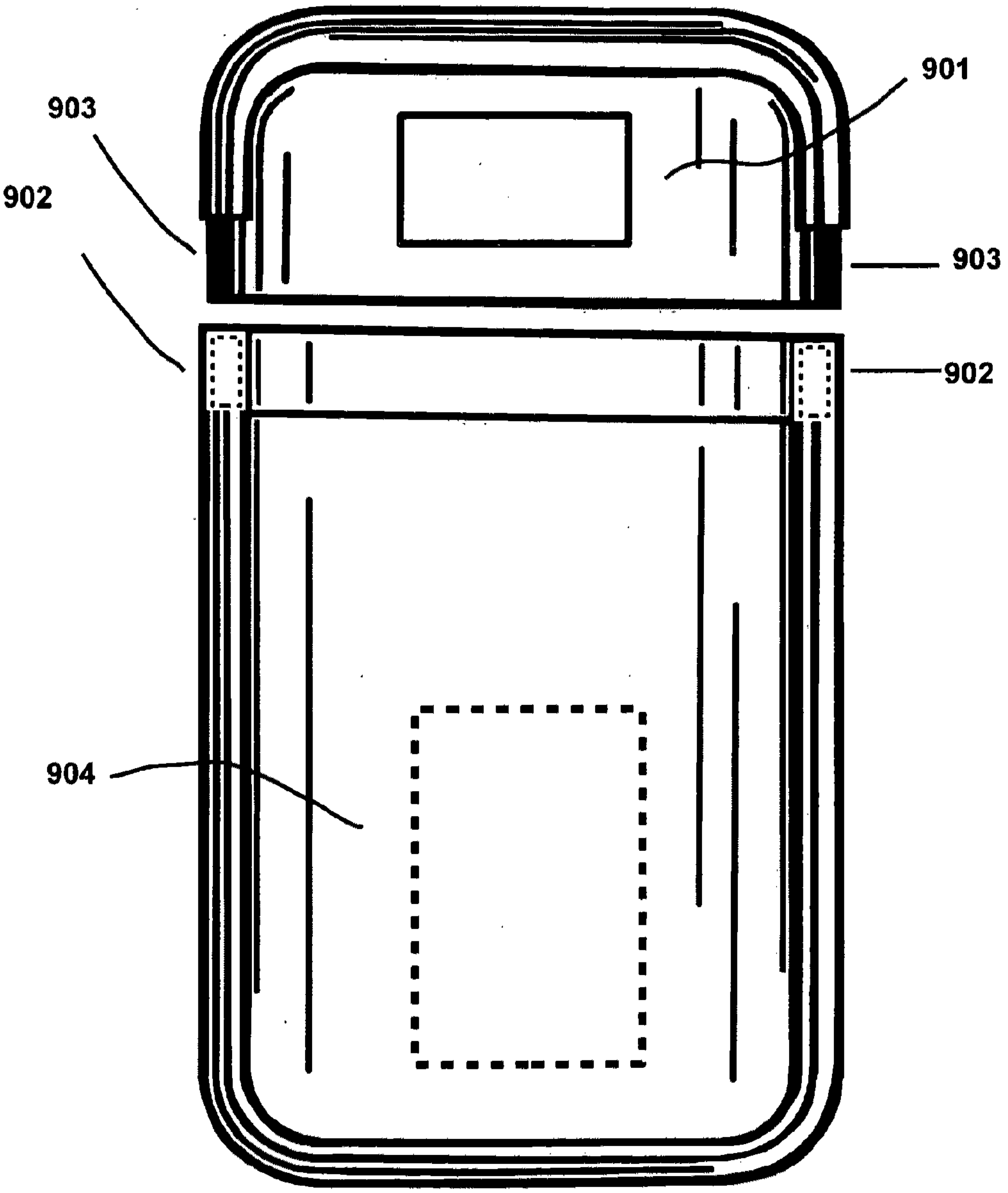
901
903
902
903
902
904

FIG. 9

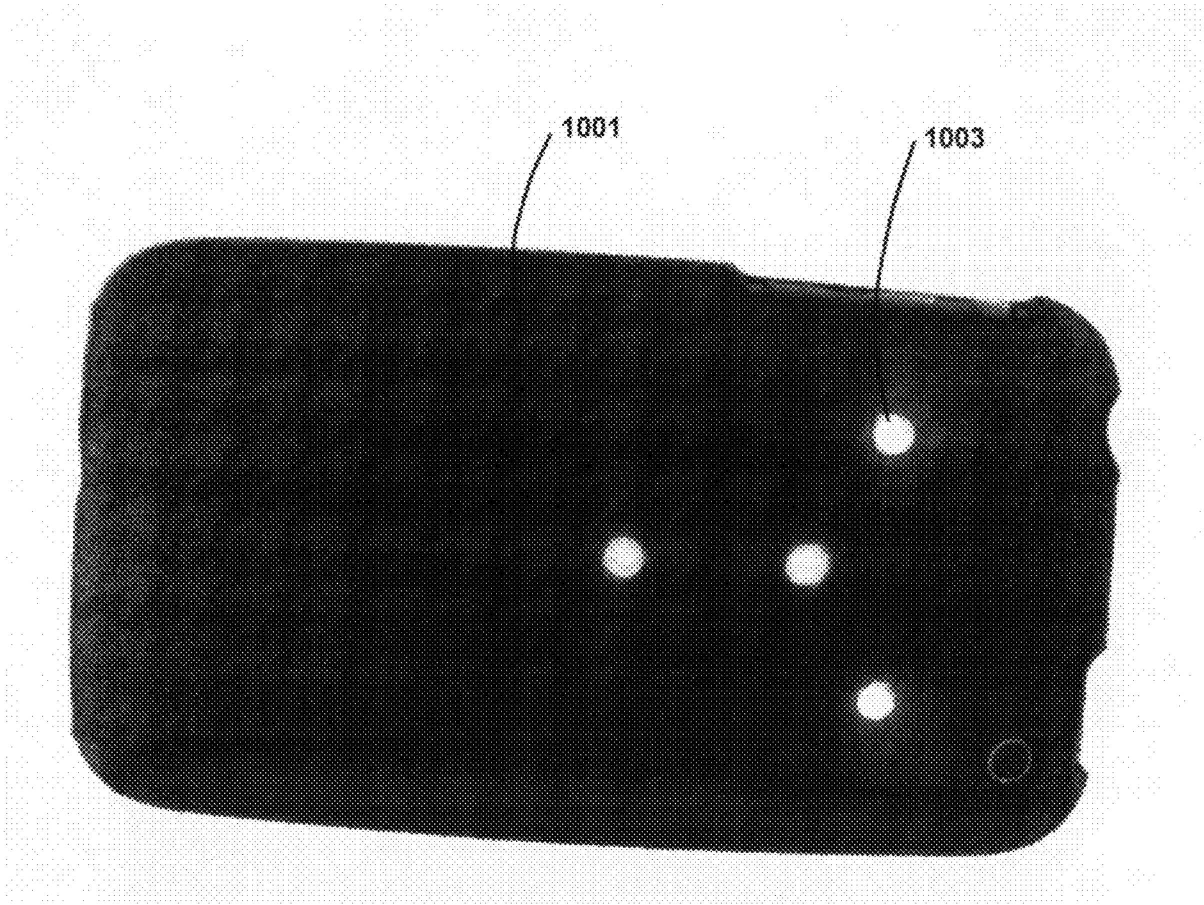

FIG. 10

RADIO FREQUENCY ENERGY HARVESTING ENCLOSURE FOR RADIO FREQUENCY CONNECTED DEVICES

RELATED U.S. APPLICATION DATA

[0001] Provisional patent application No. 61/357,590 filed Jun. 23, 2010.

FIELD OF THE INVENTION

[0002] This invention relates to an enclosure for radio frequency (hereinafter referred to as "RF") connected devices, such as cellular phones, personal electronic tablet computers, and similar electronic devices, that contains a RF energy harvester. More particularly, this invention pertains to harvesting RF energy and converting it to direct current (hereinafter referred to as "DC") electrical energy for later use by an application circuitry located either within or on the enclosure of the RF connected device, such as displaying light-emitting images.

BACKGROUND

[0003] Radio Frequency (RF) connected devices such as cellular phones, personal electronic tablet computers, and similar devices use electromagnetic waves to communicate with base-stations that exchange information through RF energy. During such communication period, these devices send and receive RF energy to establish a communication link. This power of RF energy during communication ranges approximately from 32 μW to 1 W peak power.

[0004] It is recognized in the art that the RF energy is sometimes a health risk, and attempts have been made to limit the exposure of the user's head to RF energy. U.S. Patent Application Publication 2010/0113111, to Alfred Wong, et al., describes an apparatus to redirect RF energy from the device away from the user's head. The present invention attempts to avoid the health risks associated with RF energy by harvesting some of the RF energy and keeping it contained.

[0005] It is also recognized in the art that RF energy can be harvested and converted into DC electricity. Applications of such systems are known in biomedical implants, in order to avoid health risks associated with batteries. U.S. Patent Application Publication 2009/0152954, to Triet Tu Le, et al., describes a RF harvesting circuit with a voltage doubler rectifier circuit. The present invention obtains RF harvesting results for mobile RF connected devices.

[0006] Likewise, stylized aftermarket cellular phone cases and cases for other RF devices, such as iPod Touch® or iPad®, are popular for users to personalize their devices. The present invention supplements the options that are available to users to customize the appearance of their devices by displaying lighted images or designs. Such images or designs can also have intermittent lighting patterns according to the transmitted RF energy activity that is occurring with the device.

[0007] A primary feature of the present invention is an energy harvesting enclosure, such as a cellular phone protective case, having the ability to capture some of the transmitted RF energy, and recycle it to DC electric power, without affecting the manufacturers' intended functionality of the device operation. The harvested power and resulting voltage is derived from capturing RF energy within the existing transmitted RF field created during mobile device to base-station communication activity. Examples of base-station communication activities include, but are not limited to, incoming and outgoing calls, text messaging, and interne browsing.

[0008] Accordingly, the proposed invention harvests some of the transmitted RF energy during transmission periods while an RF connected device is in communication with a base station, and recycles it into DC power. This is accomplished in order to activate an attached or detached application circuitry. Examples of application circuitry include, but are not limited to, light-emitting devices such as light-emitting diodes, one-pixel LCD, organic LEDs, etc. Likewise, the harvested and recycled RF energy can be used to charge a storage component such as rechargeable battery or super-capacitor for a later use of the stored power.

SUMMARY OF THE INVENTION

[0009] To address the current trends of customizing and personalizing RF connected devices such as cell phones and personal electronic tablet computers, a system for harvesting RF energy and converting it into DC power to activate a light-emitting device, is provided. The system includes an enclosure for a portable RF connected device, with the enclosure comprising a first end at a first longitudinal edge of the device opposite a second end at a second longitudinal edge of the device. As well as a body section having a first side on the back of the device that is opposite the second side of the device where a screen is located. The enclosure also comprises an interior surface of the first side that is facing inward on the device toward the battery, and an exterior surface of the first side that is facing outward toward the device's outer environment. Furthermore, a RF energy harvesting antenna is located on the interior surface of the first side of the device enclosure, as well as a RF to DC power conversion circuit located on the interior surface of the first side of the device enclosure. Additionally, a conditioning circuit located on the interior surface of the first side of the device enclosure. Finally, an electric application load is located on a surface of the first side of the device enclosure. This application load is a light-emitting device. The light emitting device could be designed in such a way to depict a visual image. In addition to a light-emitting device, the enclosure of a RF connected device of is further comprising a charge storage component.

[0010] According to another embodiment of the invention, the enclosure of a RF connected device of comprises a first piece and a second piece, that mechanically fasten, with an electrical interface between the first piece and second piece. In this embodiment, the RF harvesting antenna, RF to DC conversion circuit, and conditioning circuit are located on the first piece, whereas the electric application load, in the form of a light-emitting device, is located on the second piece. The electrical interface between the first and second piece allow the RF harvesting antenna, RF to DC conversion circuit, and conditioning circuit to provide power to the application load on the second piece.

[0011] In each of the embodiments of the present invention, the RF energy harvesting multiband antenna has a an output impedance that is substantially equal to the complex conjugate input impedance of the RF to DC power conversion circuit at the center frequency of the bands of operation for the RF connected device.

[0012] Also, the RF energy harvesting multiband antenna has a surface placed at a critical distance to the metallic surface on the interior of the RF connected device. The RF energy harvesting multiband antenna is on the same plane as the device antenna, which is parallel to the broad-side, back surface of the device. If the RF energy harvesting multiband antenna is not on the same plane as the device antenna, then it is located on a plane that is within a range of 1 mm to 5 mm from the plane of the device antenna.

[0013] In any embodiment of the present invention, the RF energy harvesting multiband antenna is made of either a flexible substrate or a rigid substrate. This application is specific to the RF device design and materials.

[0014] In yet another embodiment of the present invention, the enclosure of the RF connected device comprises an intermittent flashing light-emitting device that flashes in reaction to the transmitted power level communication activities of the RF connected device.

[0015] According to a final embodiment of the present invention, there is a method for displaying light-emitting images on an enclosure for RF connected devices comprising a RF energy harvesting antenna located on an interior surface of the device enclosure, a RF to DC power conversion circuit located on the interior surface of the device enclosure, a conditioning circuit located on the interior surface of the device enclosure, and an electric application load in the form of a light-emitting device located on the exterior surface of the device enclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

[0016] FIG. **1**A shows a protective cellular phone case with a RF energy harvester antenna and application circuit located on the interior surface of the enclosure.

[0017] FIG. **1**B shows a perspective view of an exposed RF energy harvester mounted on the interior surface of the protective enclosure.

[0018] FIG. **2** shows a cross section of the of a protective cellular phone enclosure design including an embedded RF energy harvester mounted on the interior surface.

[0019] FIG. **3** shows a side view of the protective enclosure and the embedded RF energy harvester and application circuit.

[0020] FIG. **4** shows an illustration of the radio frequency energy harvester and application circuit embedded in the protective enclosure mounted on a cellular phone.

[0021] FIG. **5** shows a block diagram of the RF energy harvester and the application circuit, such as a light-emitting device.

[0022] FIG. **6** shows the electrical block diagram including voltage doubler circuit.

[0023] FIGS. **7**A & **7**B show the Smith chart plot of the multiband harvesting antenna output impedance and the RF to DC power conversion circuit input impedance on the same chart, at the PCS (1850-1910 MHz) and the Cellular (824-849 MHz) frequency bands.

[0024] FIG. **8** shows an exemplary protective enclosure comprising an interchangeable first piece that has an electrical and mechanical interface to a second piece, whereas the first piece contains a light emitting logo, or symbol on the exterior back surface of the enclosure.

[0025] FIG. **9** shows an exemplary protective enclosure for a cell phone, comprising two pieces slide-lock interface and user customizable section contain a light emitting logo, or symbol on the exterior back surface of the customizable section.

[0026] FIG. **10** shows a cell phone protective enclosure having four LEDs being activated by the RF energy harvester.

DETAILED DESCRIPTION OF THE INVENTION

[0027] The present invention is described herein with reference to a variety of RF connected device enclosures, such as those for cell phones and personal electronic tablet computers. The detailed description set forth below, in association with the drawings and diagrams, is intended as a description of exemplary embodiments of the present invention and does not represent the only embodiments in which the present invention can be practiced. The detailed description includes specific details for the purpose of providing an understanding of the present invention, and is not limiting in any way. Those skilled in the art will recognize that the invention can be practiced in various applications while deviating from the specific details described herein.

[0028] The following defined terms are of exemplary terms used within this disclosure. Except where noted otherwise, variants of all terms, including singular or plural forms, fall within each exemplary term meaning:

[0029] I. "Enclosure" as used herein means, but is not limited to the protective plastic case that is the hardware of a cellular phone, gaming device, music player, personal electronic tablet computer, etc.

[0030] II. "Radio Frequency connected" or "RF connected" as used herein, includes but is not limited to, a communicative relationship between devices or circuits via electromagnetic waves.

[0031] III. "RF to DC power conversion circuit," as used herein means a diode rectifier circuit, voltage doubler circuit, full-wave rectifier circuit, or similar solid state rectifier components and integrated circuit.

[0032] IV. "RF energy harvester," as used herein means a RF harvesting antenna and RF to DC power conversion circuit.

[0033] FIG. **1**A illustrates the interior of a RF connected device enclosure **100** with a RF harvesting antenna **101** and application circuit **102**. FIG. **1**B illustrates a RF connected device enclosure **110** made of a plastic such as ABS or similar material. RF energy is captured by means of RF harvesting antenna **111** and recycled by means of RF to DC power conversion circuit **112**. The RF energy harvester **111** and the application circuits **112** are populated on printed circuit board **113** using a thin flexible substrate or on a thin rigid printed circuit board substrate. The enclosure **110** is similar in appearance to a cellular phone external protective case.

[0034] FIG. **5** illustrates the block diagram for a RF energy harvester system. RF energy is captured by the antenna structure **501** and recycled by RF to DC power conversion circuit **502** and a power conditioning and filter circuit **503**. The combination of these blocks supply the required power to an application circuitry **504**, such as a light-emitting device. The RF to DC power conversion circuit **502** is made of small package surface mount zero bias Schottky detector diodes such as the HSMS-286x series or similar rectifier devices or integrated circuit and devices, and a filter circuit.

[0035] The application circuit **504** can be made from active devices, passive devices, or a combination of active and passive devices. Such active and passive devices include, but are not limited to, light-emitting diodes, super capacitors, solid state batteries, rechargeable batteries, and electronic circuits.

The application circuit **504** maybe embedded in the enclosure or it can be an auxiliary apparatus that is either attached or detached from the enclosure.

[0036] The presented RF energy harvester **501** does not require an impedance matching network. However, an additional impedance matching network may be utilized to broaden the impedance match of the antenna structure **501** to the RF to DC power conversion circuit **502**. The various components and structures, which are integral to the enclosure, convert mainly some of the transmitted RF energy to a DC power and voltage for activation of a single or multiple application circuits.

[0037] Several types of harvesting antenna structures **501** may be employed. A key aspect of the harvesting antenna is its impedance value. The harvesting antenna **501** impedance must be equal or near the complex conjugate impedance of that of the voltage doubler **502**, **602** input circuit, Z (Antenna) =Z*(voltage doubler), as illustrated in FIGS. 7A and 7B. The harvesting antenna **501** combined with the RF to DC power conversion circuit **502** form a resonant network.

[0038] In addition to specific antenna design practice, the embedded harvesting antenna structure is configured utilizing the enclosure material, such as ABS plastic or other enclosure material, thickness and electrical properties such as Dielectric Constant to achieve the required impedance value **701** and **703**. The embedded harvester antenna design also accounts for the distance **204** between the harvesting antenna structure **207** and the internal hardware **202** of the RF connected device. Additionally, the harvester antenna design accounts for the thickness of the enclosure material **206** that surrounds harvesting antenna structure **207**, in order to tune the harvesting antenna for optimum RF energy capturing and improve the system efficiency.

[0039] FIGS. 7A and 7B show simulated data points on a Smith Plot, of the antenna structure impedance **701**, **703** and the RF to DC power conversion circuit impedance **702**, **704** at various frequencies. In consideration of the range of frequency bands utilized for cellular phones, the harvesting antenna output impedance **701**, **703** is substantially near the complex conjugate of the RF to DC power conversion circuit input impedance **702**, **704** at the average frequency of each band.

[0040] The RF harvester employs and makes use of the RF to DC power conversion circuit **502** and the rectifier diodes junctions capacitance **602** to form the resonant circuit and impedance match **701**, **702**, such that the energy transfer to the RF to DC power conversion circuitry **502** and application circuitry **504** is optimized without the use of any matching network circuit.

[0041] Prior art teaches that an additional impedance matching circuit is required to maximize power transfer. According to the detailed description above, the present invention does not require a matching circuit to maximize power transfer, and therefore reduces the overall cost of the RF harvester. A matching network maybe used to further improve the power transfer between the harvester antenna structure and the RF to DC power conversion circuit, but is not necessary for satisfactory operation.

[0042] As illustrated in FIG. **6**, the RF to DC power conversion circuit **602** is made of a voltage doubler circuit or full wave rectifier. The output voltage can be increased by cascading more than one voltage doubler stage. The signal conditioning block **603** may include a voltage regulator and a filter circuit.

[0043] In an embodiment that employs more than one voltage doubler stage, the harvesting antenna output impedance must be tuned to the complex conjugate of the cascaded voltage doublers input impedance.

[0044] To maximize the voltage amplitude at the input of the RF to DC power conversion circuit, a virtual L-C resonant circuit is formed by the antenna inductive impedance and the RF to DC power conversion circuit diodes junctions capacitance.

[0045] The number of voltage doubler stages used has significant impact on the harvester efficiency. It will change the RF to DC power conversion circuit equivalent input capacitance. If more voltage doubler stages are used, the RF to DC power conversion circuit junction and parasitic capacitance components vary, and therefore effecting the impedance match and the resonance frequency. The presented patent utilizes but not limited to one voltage doubler stage as the RF to DC power conversion circuit FIG. **6**.

[0046] The application circuit and load impedance and resistance have an effect on the overall harvester efficiency. The presented enclosure is designed to accommodate the integrated harvesting antenna, RF to DC power conversion circuit, filter and power conditioning circuit, and the application circuitry. The presented methodology is an illustration of an exemplary enclosure and is general. One with skill in the art will recognize that different methodologies, combinations of elements, and materials maybe used.

[0047] The harvesting antenna structure placement, orientation, height, and position with respect to the transmitting antenna of the RF connected device, such as a cellular phone, is controlled during the design stage to maximize the energy harvested and minimize the effect on the Radio Frequency connected device mode of operation.

[0048] Shown in FIG. **2** is a cross-section view of the presented plastic enclosure **210**. The enclosure width **203** varies and could be different for each device and is designed to fit on the RF connected device just like in any typical protective case. In an exemplary embodiment, the enclosure wall thickness **201** ranges from about 1.5 mm to 2.5 mm. The enclosure's broad side thickness **209** ranges from about 2.5 mm to 4 mm including the cover layer **208**. The enclosure's broad side thickness **209** is designed to accommodate the integrated harvester and application circuit.

[0049] The presented RF to DC power conversion circuit and topology has a direct effect on the system efficiency and impedance matching. The RF to DC power conversion circuit input capacitance is used in conjunction with the antenna output inductance to form the high Q resonant circuit. The embedded harvester **207** and application circuit thickness **205** is about 1.5 mm. The exemplary distance **204** between the Radio Frequency connected device back **202** and the harvester, range from 1.5 mm to 5 mm and it's critical to the harvester antenna operation and impedance. It affects the harvester efficiency.

[0050] As illustrated in FIG. **8** and FIG. **10**, an embodiment of the presented enclosure including RF harvester pertains to a personalized and customizable protective cover for a person's RF connected device, such as a cellular phone. The enclosure **806** has an interface to lock, hold, and supply electrical power to a logo or icon **801**. The logo or icon section **801** may be of any shape, such as square, circle, or octagon. The logo or icon section **801** may incorporate any design such as a drawing, image, or symbol. The customizable protective cover **806** comprises an opening **805** to house an interchangeable logo or icon **801**. FIG. **8** shows a twist or push-to-lock design interface that incorporates a mechanical lock **802** and electrical contacts **803**. The thickness **804** of the presented enclosure varies to suit the logo or icon mechanical and electrical requirements.

[0051] FIG. **9** shows another embodiment of a two pieces slide-lock interface. The user customizable section **901** of the enclosure is inter-changeable and contains a logo, icon, or any application circuit, such as a light-emitting device. The larger enclosure section **904** contains the RF energy harvester and the power conditioning circuit. The two separate sections **901**, **904** of the enclosure join together by means of an electrical interface **903** and a mechanical interface **902**. A primary feature of the user customizable section **901** of the enclosure is the ability of customizing a single enclosure using multiple logo, icon, or design options. The RF connected device user is able to have one enclosure that is quickly modifiable and customizable. This is accomplished by replacing only the smaller section **801**, **901** containing a logo or icon with another that incorporates a different design or function. Interchangeability of the logo or icon section **801**, **901** is possible by incorporating the same mechanical interface **802**, **902** and electrical interface **803**, **903**, as the larger enclosure section **806**, **904**. Therefore, the user creates a new customized look and a new enclosure design or function without acquiring a new enclosure altogether.

[0052] The logo or icon section **801**, **901** may or may not incorporates electrical components. In one embodiment the logo or icon section **801**, **901** only incorporates the graphic design. The light-emitting device, such as an LED, is incorporated and embedded in the larger enclosure section **806**, **904**. The light-emitting device is placed in the larger enclosure section cavity **805**, and mounted between the internal hardware of the RF device and the logo or icon section **801**, **901** such that emitted light shines through the logo or icon to create an image for visual stimulation.

[0053] Likewise, it can be recognized and appreciated in an embodiment where a logo or icon is directly incorporated on a singular and unitary RF device enclosure, as apparent in FIG. **10**. In such an embodiment, the logo or icon is not interchangeable from the enclosure as a whole.

[0054] In FIG. **10** a cellular phone enclosure **1001** is embedded with light-emitting diodes (LEDs) **1003**. The LED excitation and illumination is facilitated through the harvested energy, without the use of a battery or any form of plugged or wired power source. The enclosure **1001** may include single or multiple antenna structures, RF to DC power conversion circuits, storage devices such as super-capacitor, electronics devices, and integral LED(s) **1003** as visual signaling devices. The light-emitting diodes **1003** are suitable for visual signaling of an incoming communication, such as phone calls or text messages, when the cellular phone is screen side down and the ringer is turned off. Likewise, the illuminated LEDs **1003** provide visual stimulation of their own as in FIG. **10** or through illuminating art such as an icon or logo **801**, as in FIG. **8**. The illumination can be constant light, or flashing light that is synchronized to the cellular phone base-station communication activities of the attached RF connected device.

[0055] Where the RF energy harvester is used to directly power the LED(s) **1003** to decorate a protective case logo or icon **801**, the RF connected device transmitted power levels and operation modes affect the brightness level of the LED(s) **1003** and is an indication of the transmitted power level. Light-emitting diode **1003** brightness is directly proportional to the peak power level of the radio frequency wave being transmitted by the Radio Frequency connected device. In another embodiment where conditioning circuit and charge storage device is used, the brightness level of the LED(s) is controlled.

[0056] In any every embodiment of the present invention, all or some of the electronic circuitry, including the logo or icon embedded circuitry, will be powered by the RF energy harvester which is incorporated and integrated within the physical structure of the RF device protective enclosure.

[0057] The logo or icon section **801**, **901** may includes batteries to power additional electronic circuitry that is triggered and controlled by the enclosure energy harvester signal and circuit. Furthermore, audible and tactile notification mechanisms, such as the ringer and vibration, can be turned of and the user will be alerted of incoming communications utilizing only the RF harvested energy to activate its integral light-emitting device on the enclosure. An obvious benefit of such method of operation results in an extension of the internal battery cycle life for the RF connected device, while continuously indicating operation of the attached device.

What is claimed is:

1. An enclosure for a portable RF connected device, the enclosure comprising:

a first end at a first longitudinal edge of the device opposite a second end at a second longitudinal edge of the device;

a body section having a first side on the back of the device that is opposite the second side of the device where a screen is located;

an interior surface of the first side that is facing inward on the device toward the battery;

an exterior surface of the first side that is facing outward toward the device's outer environment;

a RF energy harvesting antenna located on the interior surface of the first side of the device enclosure;

a RF to DC power conversion circuit located on the interior surface of the first side of the device enclosure;

a conditioning circuit located on the interior surface of the first side of the device enclosure; and

an electric application load located on a surface of the first side of the device enclosure.

2. The enclosure of a RF connected device of claim **1**, further comprising a light-emitting device on the exterior surface.

3. The enclosure of a RF connected device of claim **2**, further comprising a visual image on the exterior surface of the first side of the device enclosure.

4. The enclosure of a RF connected device of claim **1**, further comprising a charge storage component.

5. The enclosure of a RF connected device of claim **1**, further comprising:

a first piece and a second piece, that mechanically fasten;

an electrical interface between the first piece and second piece;

the RF harvesting antenna, RF to DC power conversion circuit, and conditioning circuit located on the first piece;

the electric application load located on the second piece;

the electrical interface between the first and second piece allow the RF harvesting antenna, RF to DC conversion circuit, and conditioning circuit to provide power to the application load on the second piece.

6. The enclosure of a RF connected device of claim **1**, further comprising an embedded RF energy harvesting multiband antenna having an output impedance that is substantially equal to the complex conjugate input impedance of the RF to DC power conversion circuit at the center frequencies of the bands of operation for the RF connected device.

7. The enclosure of a RF connected device of claim **5**, further comprising a RF energy harvesting multiband antenna located on the interior of the RF connected device that has a surface placed at a critical distance to the metallic surface of the internal hardware of the RF connected device, said distance in the range of 1.5 mm to 5 mm.

8. The enclosure of a RF connected device of claim **1**, further comprising a RF energy harvesting multiband antenna made of a flexible substrate.

9. The enclosure of a RF connected device of claim **1**, further comprising a RF energy harvesting multiband antenna made of a rigid substrate.

10. The enclosure of a RF connected device of claim **1**, further comprising an intermittent flashing light-emitting device synchronized by the transmitted power level of communication activities.

11. A method for displaying light emitting images on an enclosure for RF connected devices comprising:

a RF energy harvesting antenna located on an interior surface of a first side of the device enclosure;

a RF to DC power conversion circuit located on the interior surface of the first side of the device enclosure;

a conditioning circuit located on the interior surface of the first side of the device enclosure; and

an electric application load located on the exterior surface of the first side of the device enclosure.

* * * * *

US 20140111032A1

(19) **United States**

(12) **Patent Application Publication**

Shearer et al.

(10) **Pub. No.: US 2014/0111032 A1**

(43) **Pub. Date: Apr. 24, 2014**

(54) **POWERING DEVICES USING RF ENERGY HARVESTING**

(71) Applicant: **POWERCAST CORPORATION**, Ligonier, PA (US)

(72) Inventors: **John G. Shearer**, Ligonier, PA (US); **Charles E. Greene**, Pittsburgh, PA (US); **Daniel W. Harrist**, Carnegie, PA (US)

(73) Assignee: **POWERCAST CORPORATION**, Ligonier, PA (US)

(21) Appl. No.: **14/143,334**

(22) Filed: **Dec. 30, 2013**

Related U.S. Application Data

(60) Continuation of application No. 12/499,618, filed on Jul. 8, 2009, now Pat. No. 8,621,245, which is a division of application No. 11/447,412, filed on Jun. 6, 2006, now abandoned.

(60) Provisional application No. 60/688,587, filed on Jun. 8, 2005.

Publication Classification

(51) **Int. Cl.**
H02J 17/00 (2006.01)

(52) **U.S. Cl.**
CPC *H02J 17/00* (2013.01)
USPC **307/149**

(57) **ABSTRACT**

Disclosed is an apparatus for an application including a core device for the application. The apparatus includes a power (preferably RF energy) harvester connected to the core device to power the core device. Also disclosed is a method for an application. The method includes the steps of converting RF energy into usable energy. There is the step of powering the core device with the usable energy.

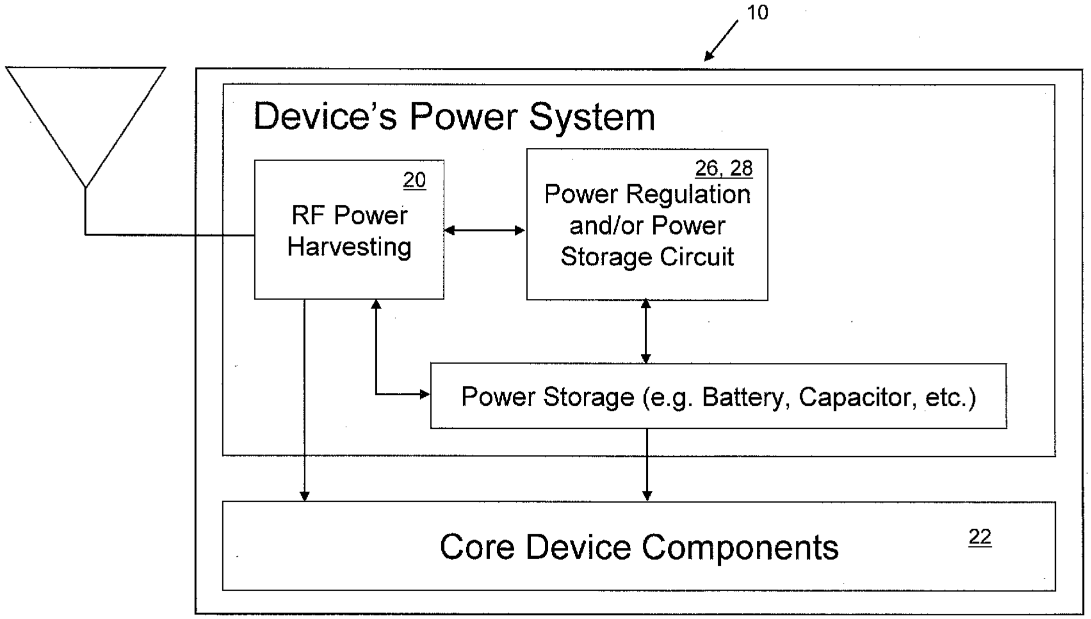
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

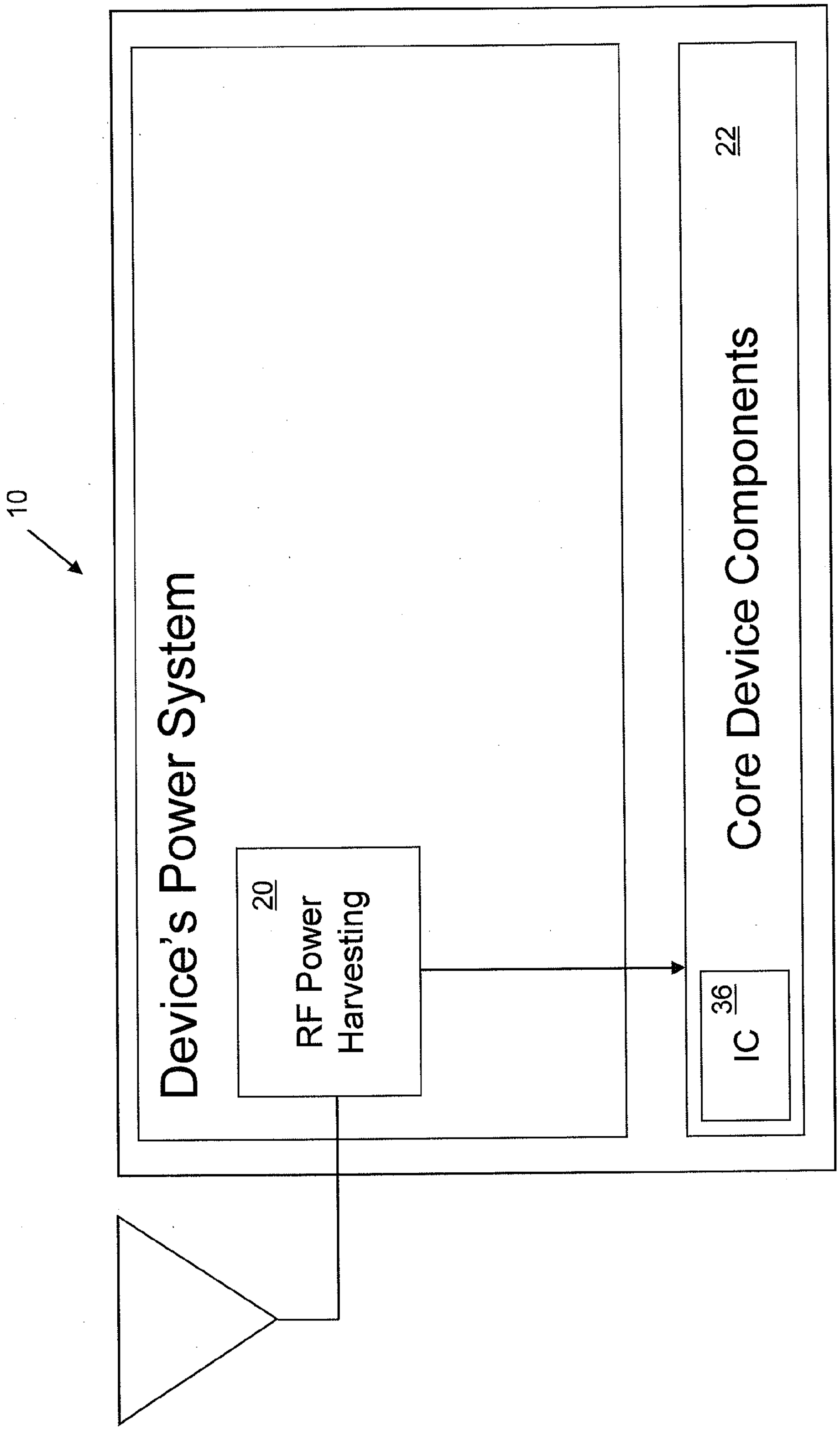
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
IC
36

Fig. 1

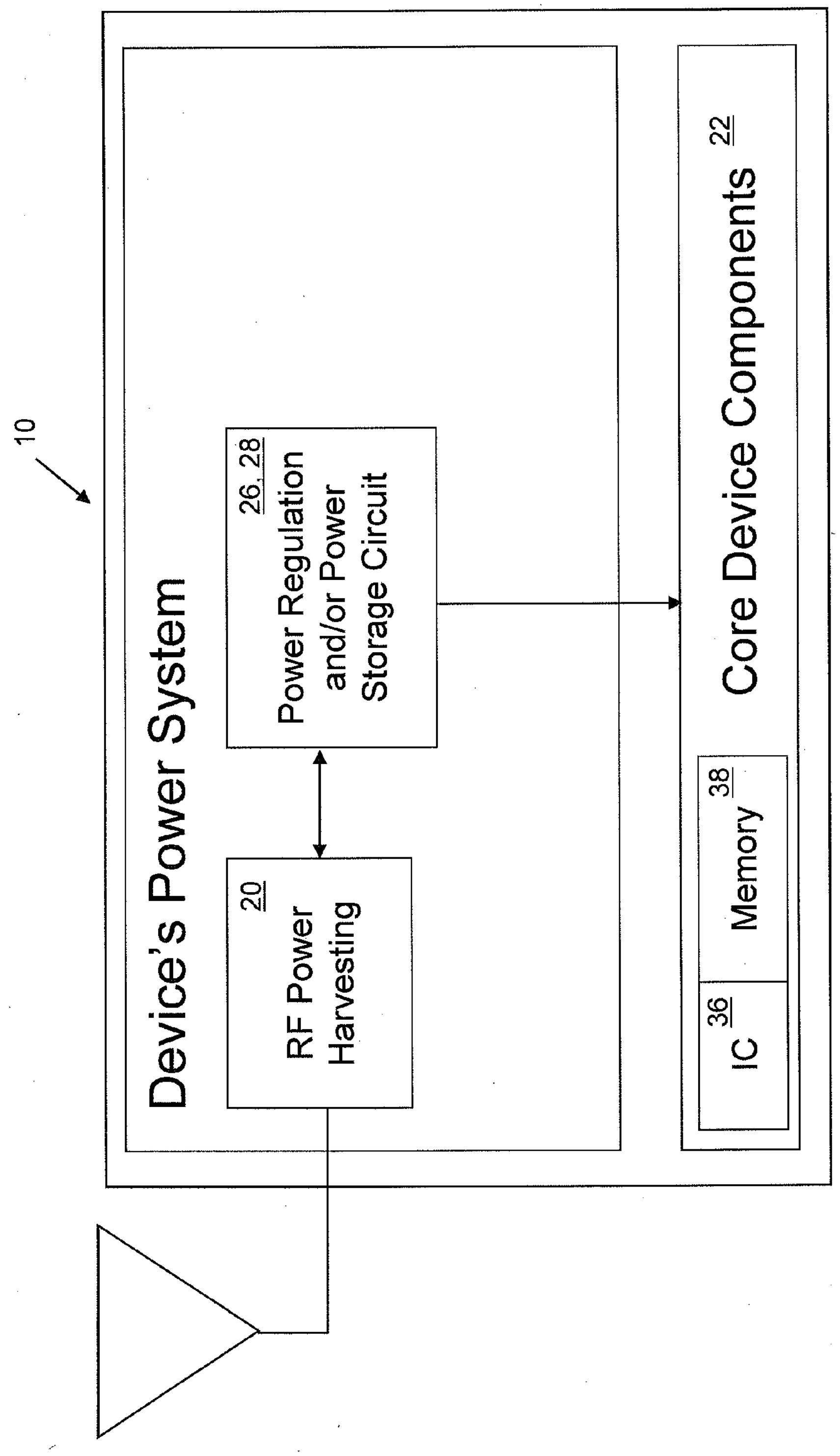
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Core Device Components
22
IC
36
Memory
38

Fig. 2

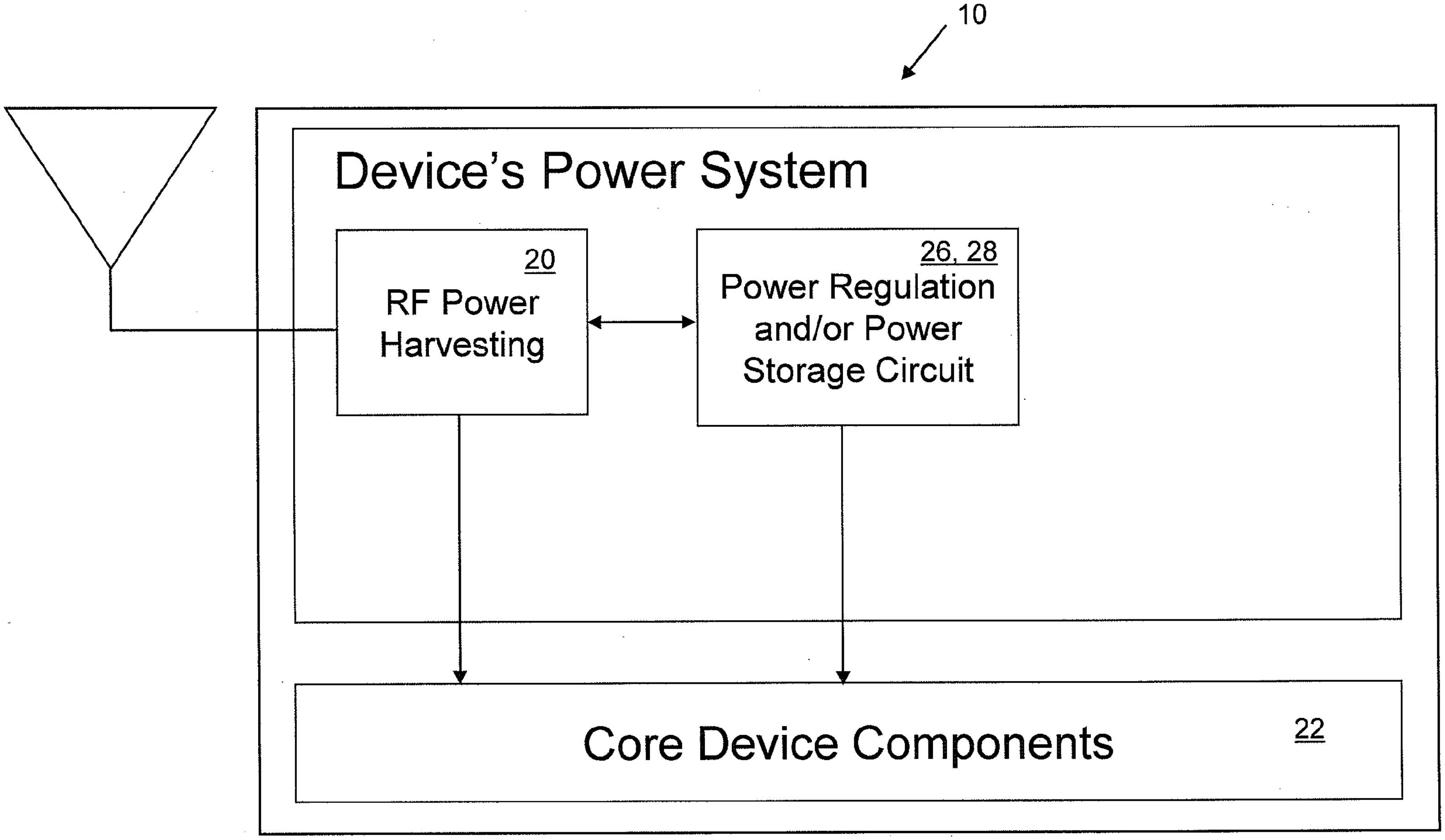
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Core Device Components
22

Fig. 3

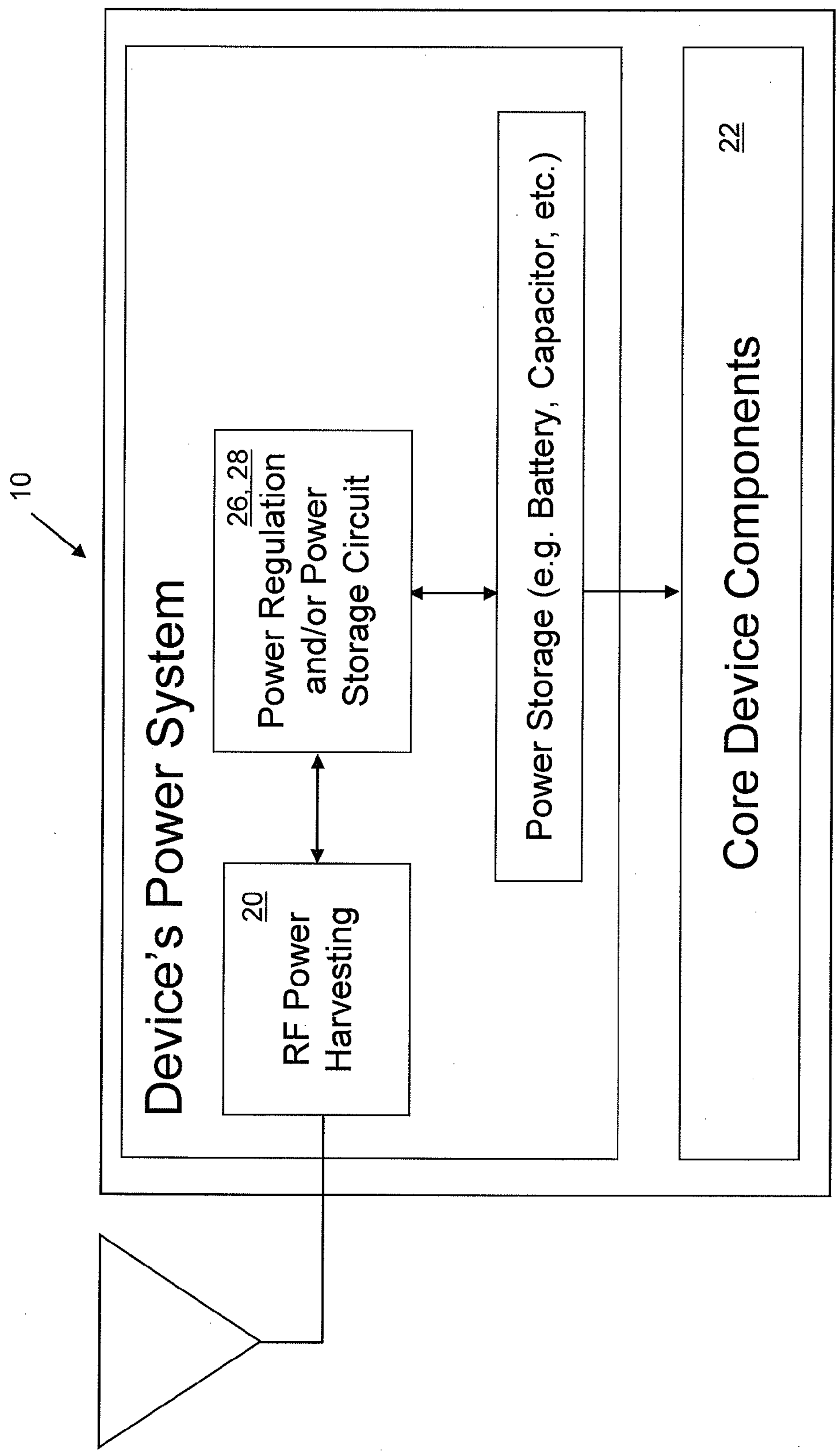
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 4

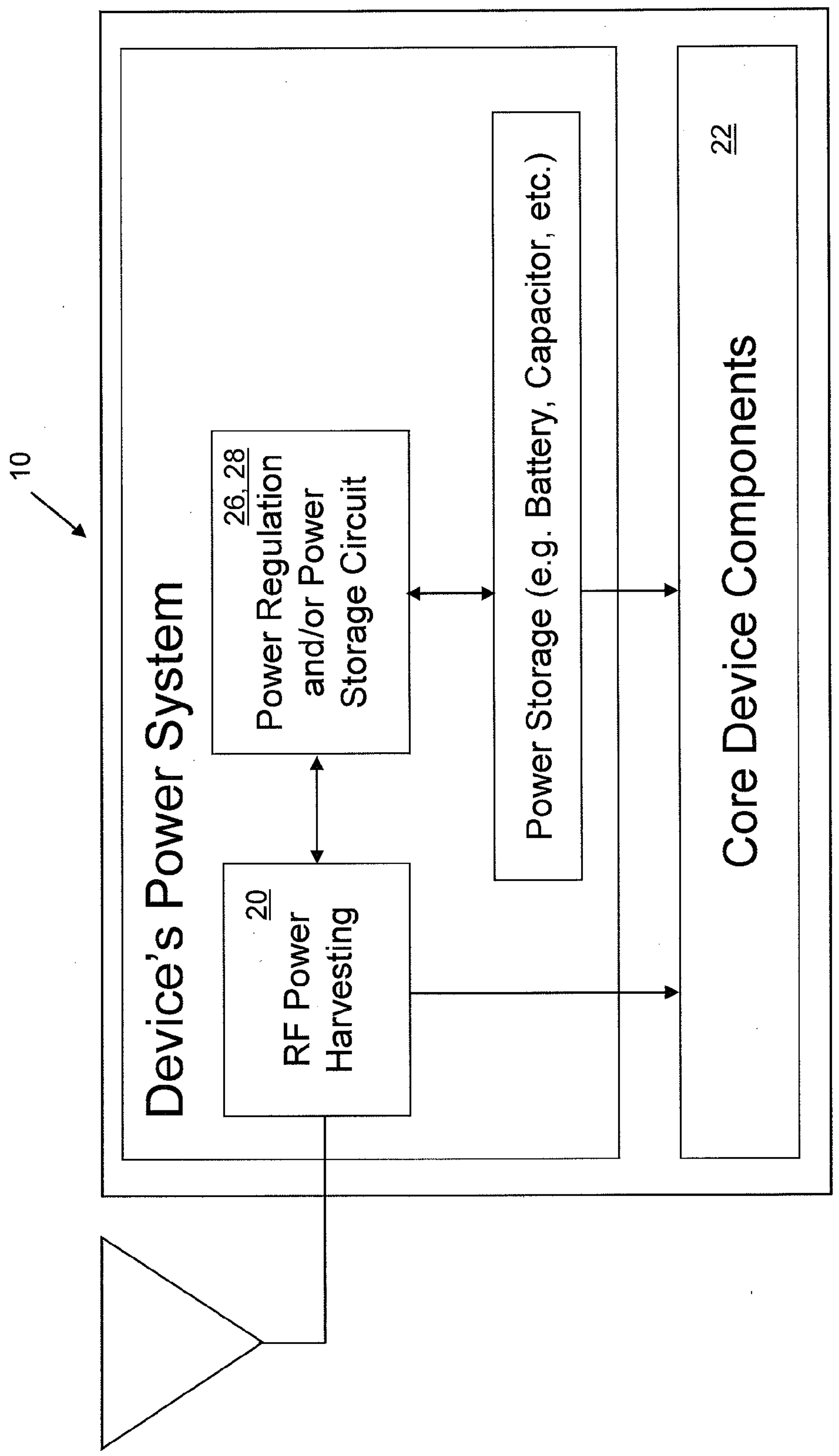
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 5

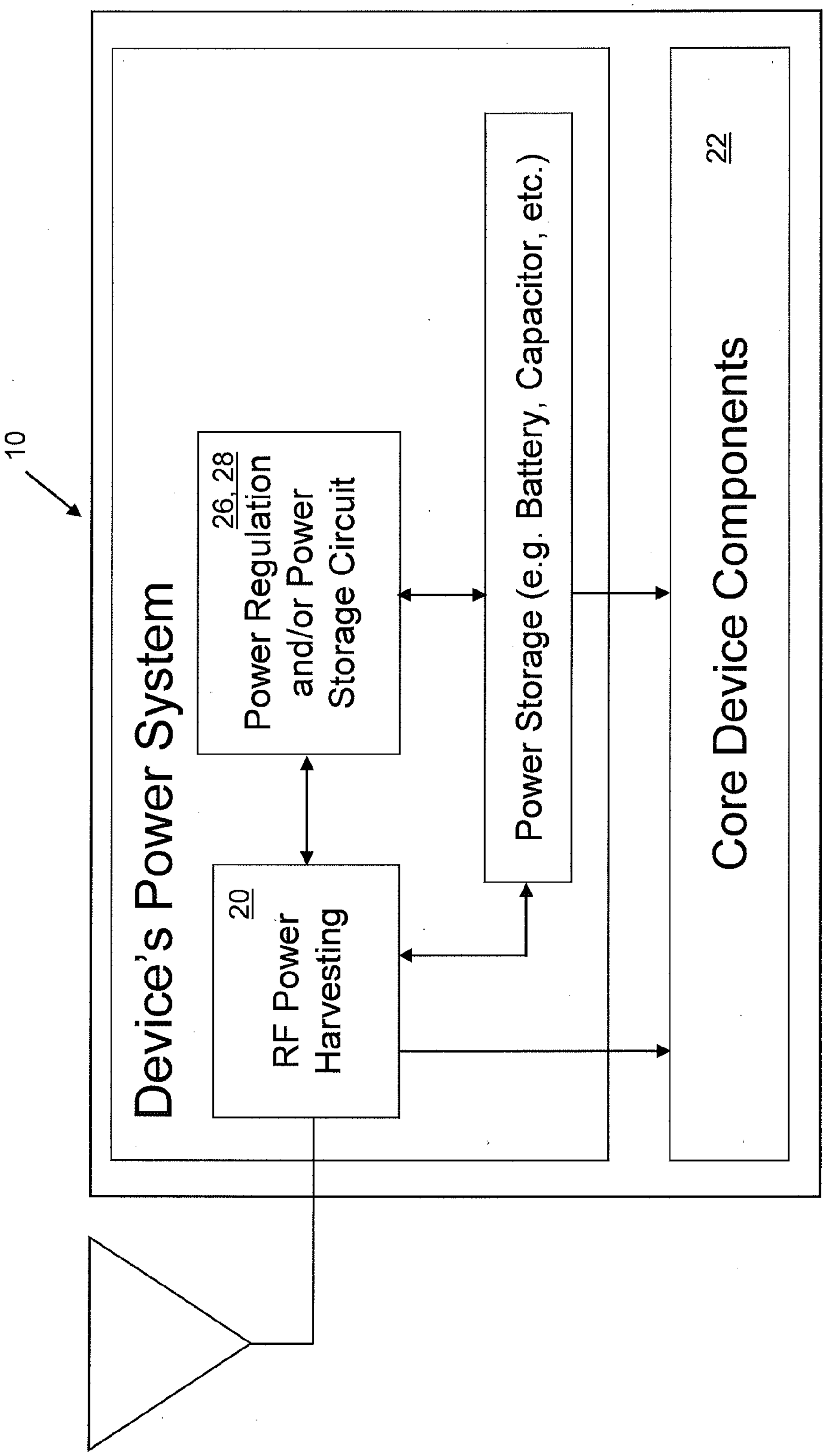
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 6

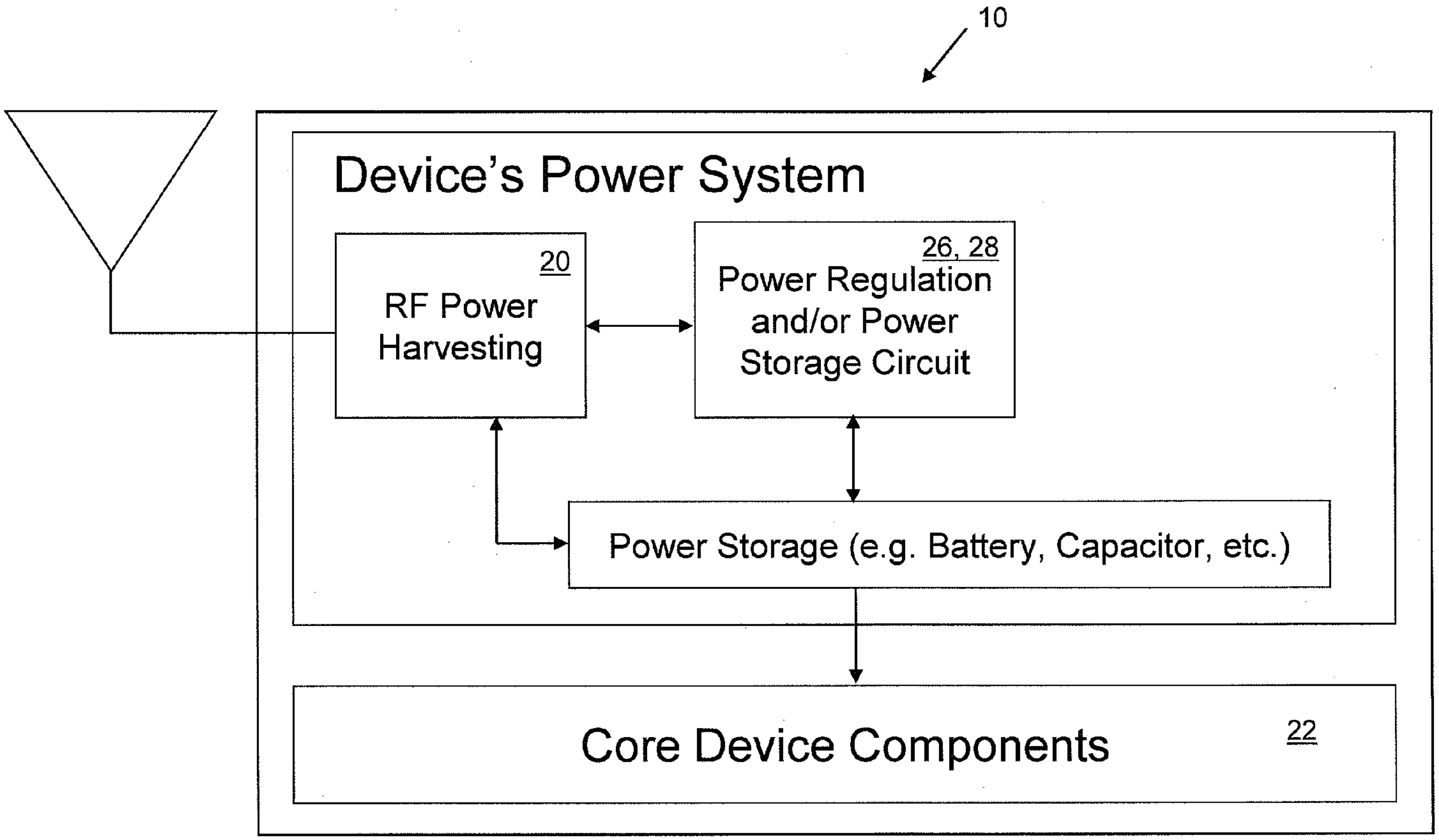
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 7

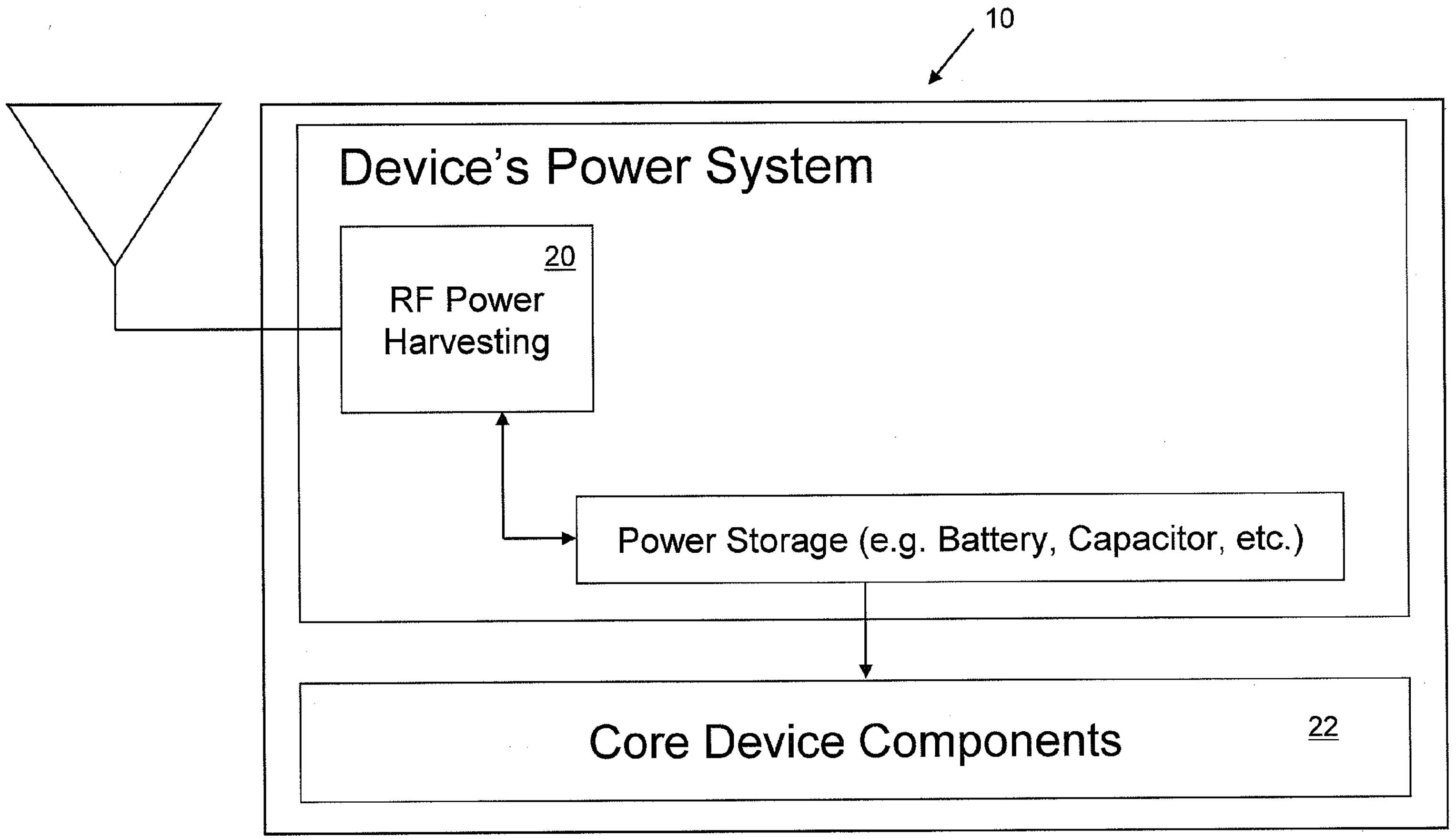
10
Device's Power System
20
RF Power
Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 8

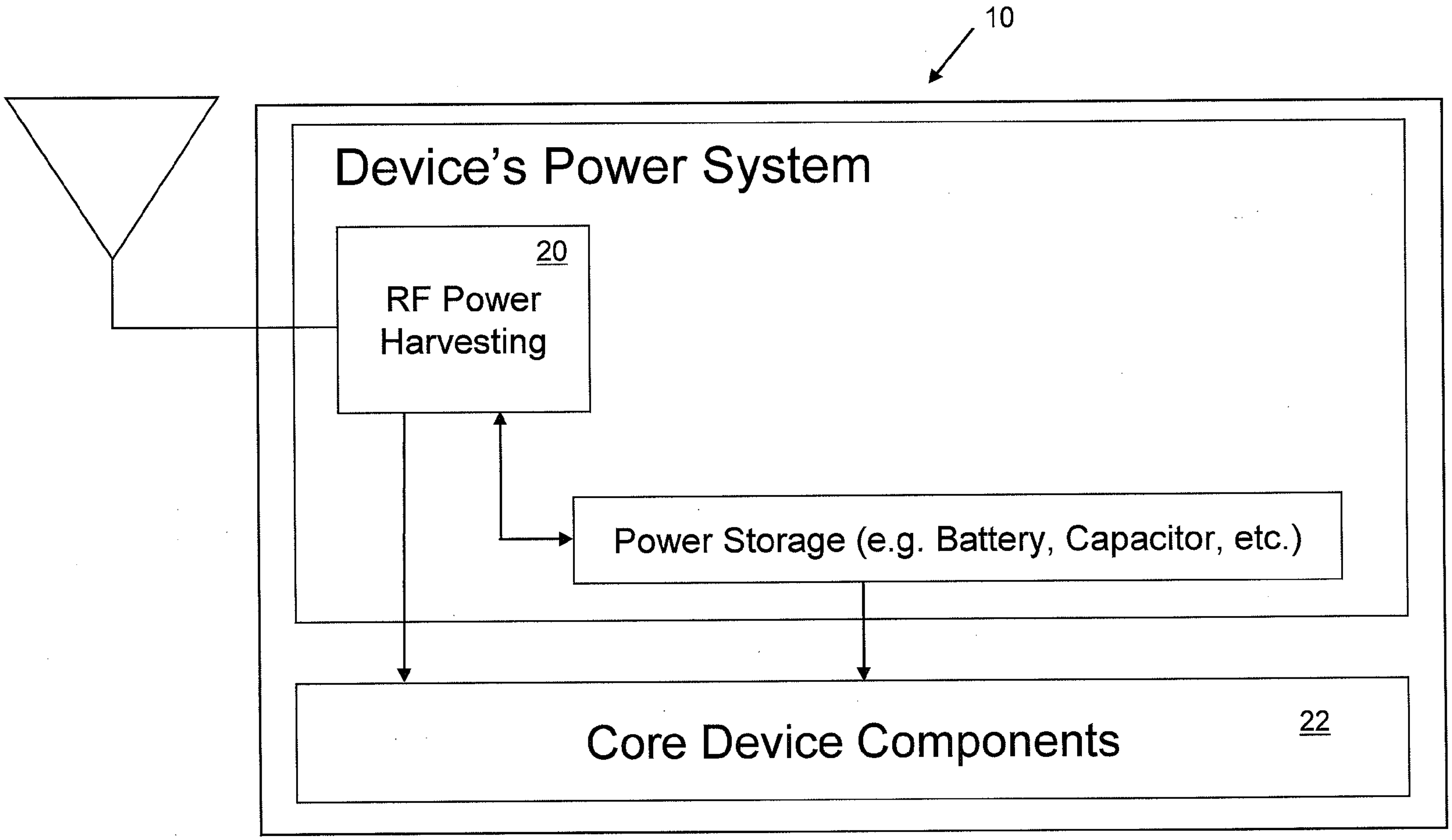
10
Device's Power System
20
RF Power Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 9

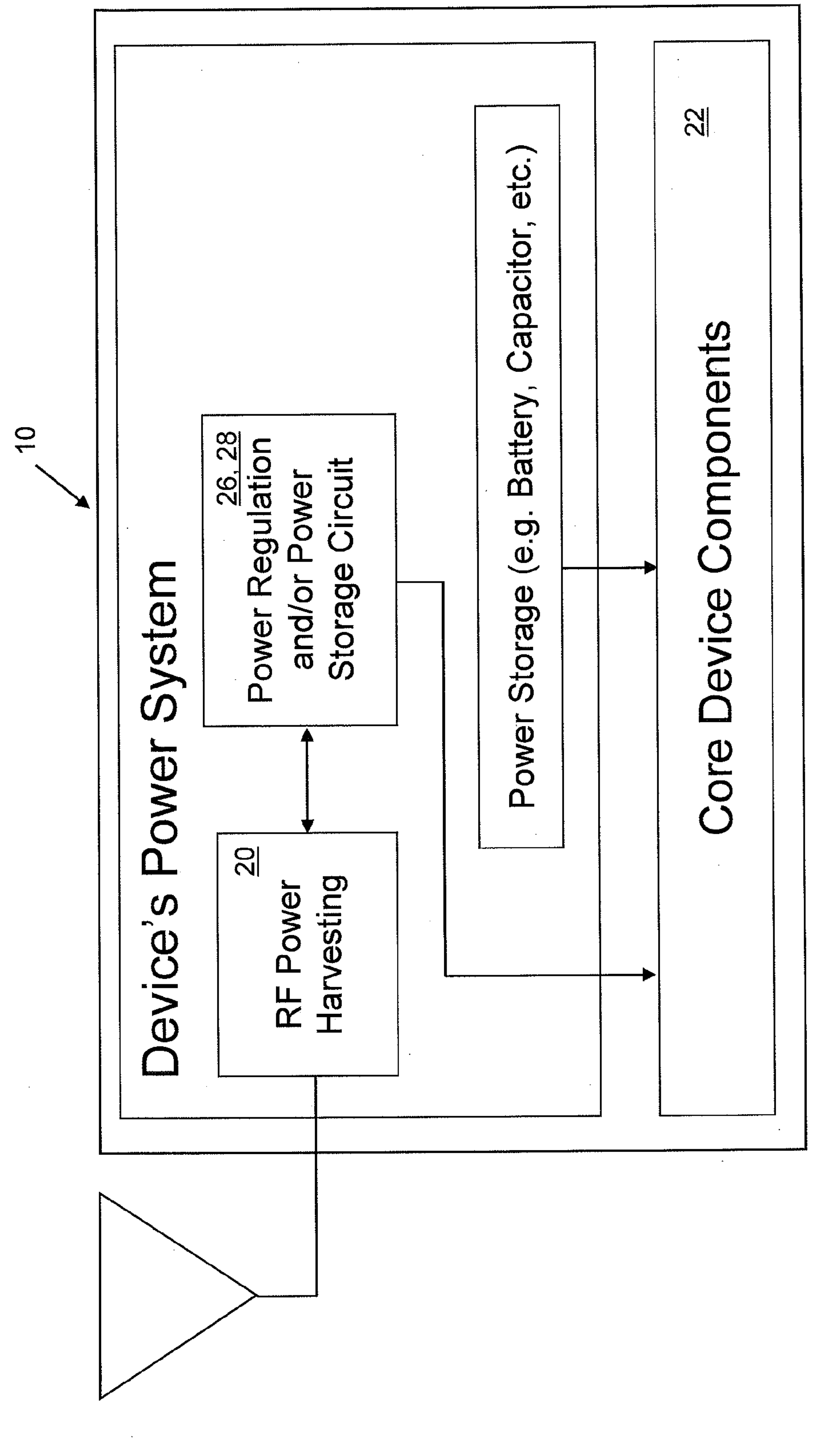
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 10

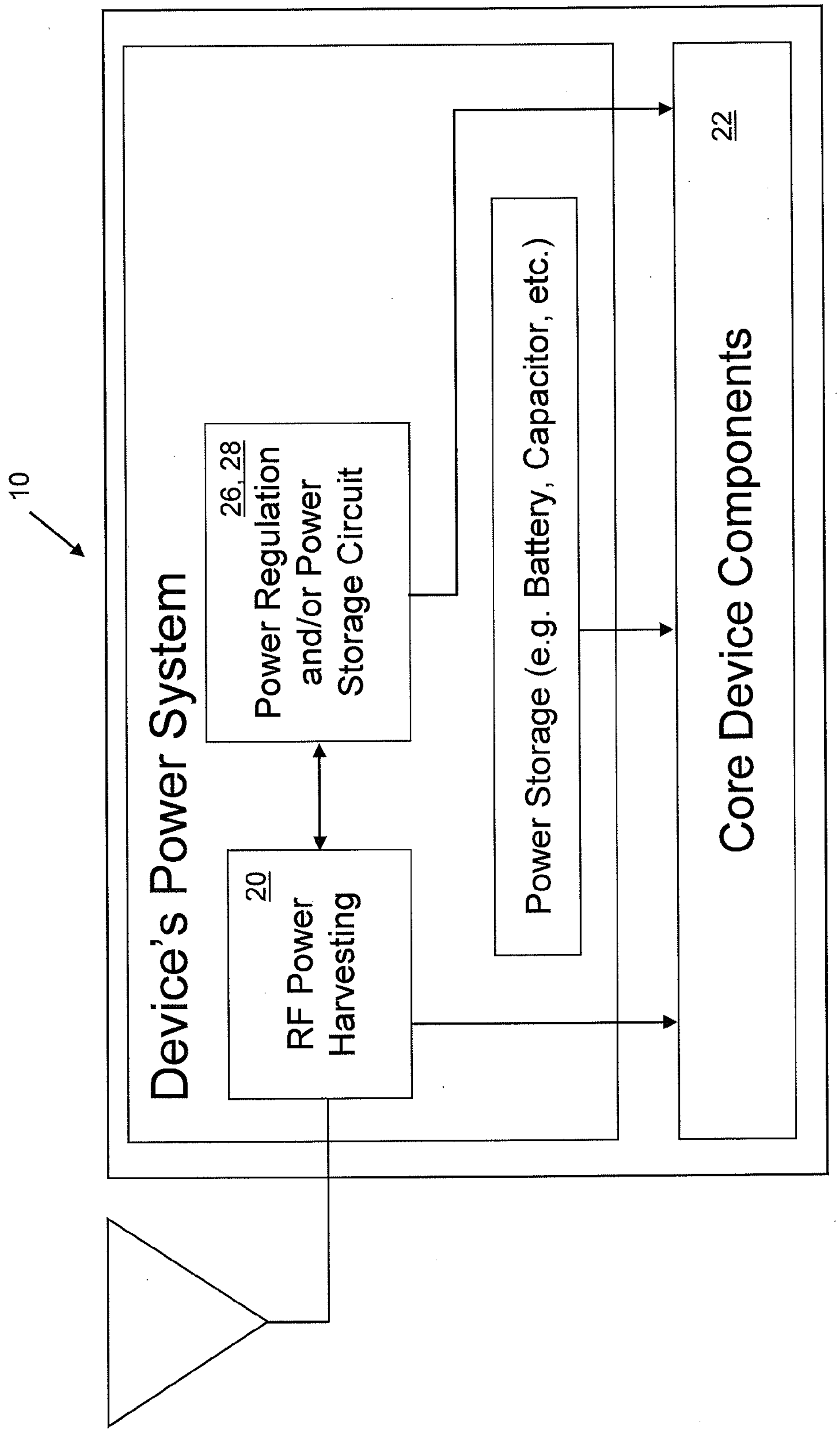
10
Device's Power System
26, 28
Power Regulation and/or Power Storage Circuit
20
RF Power Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 11

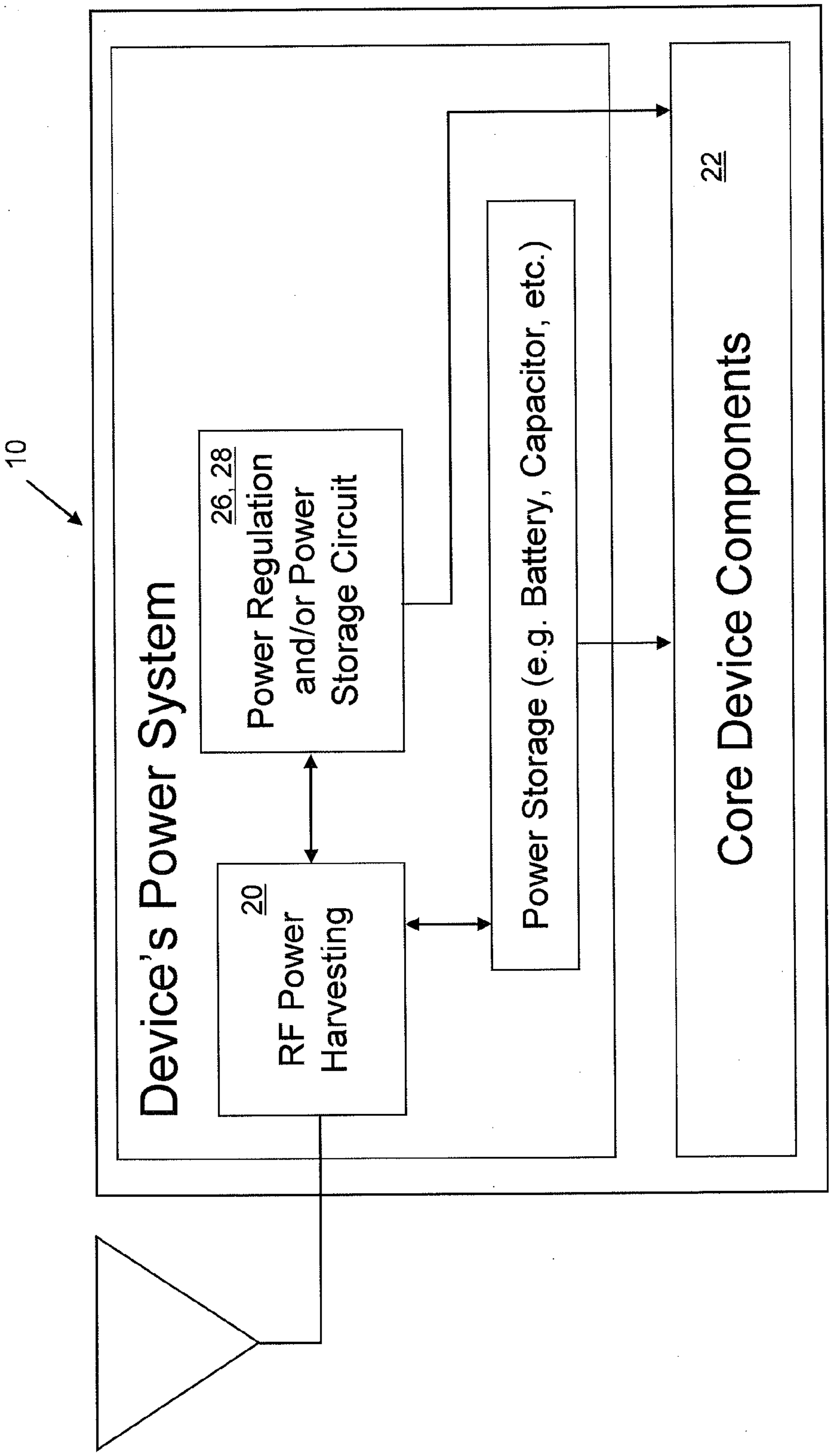
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 12

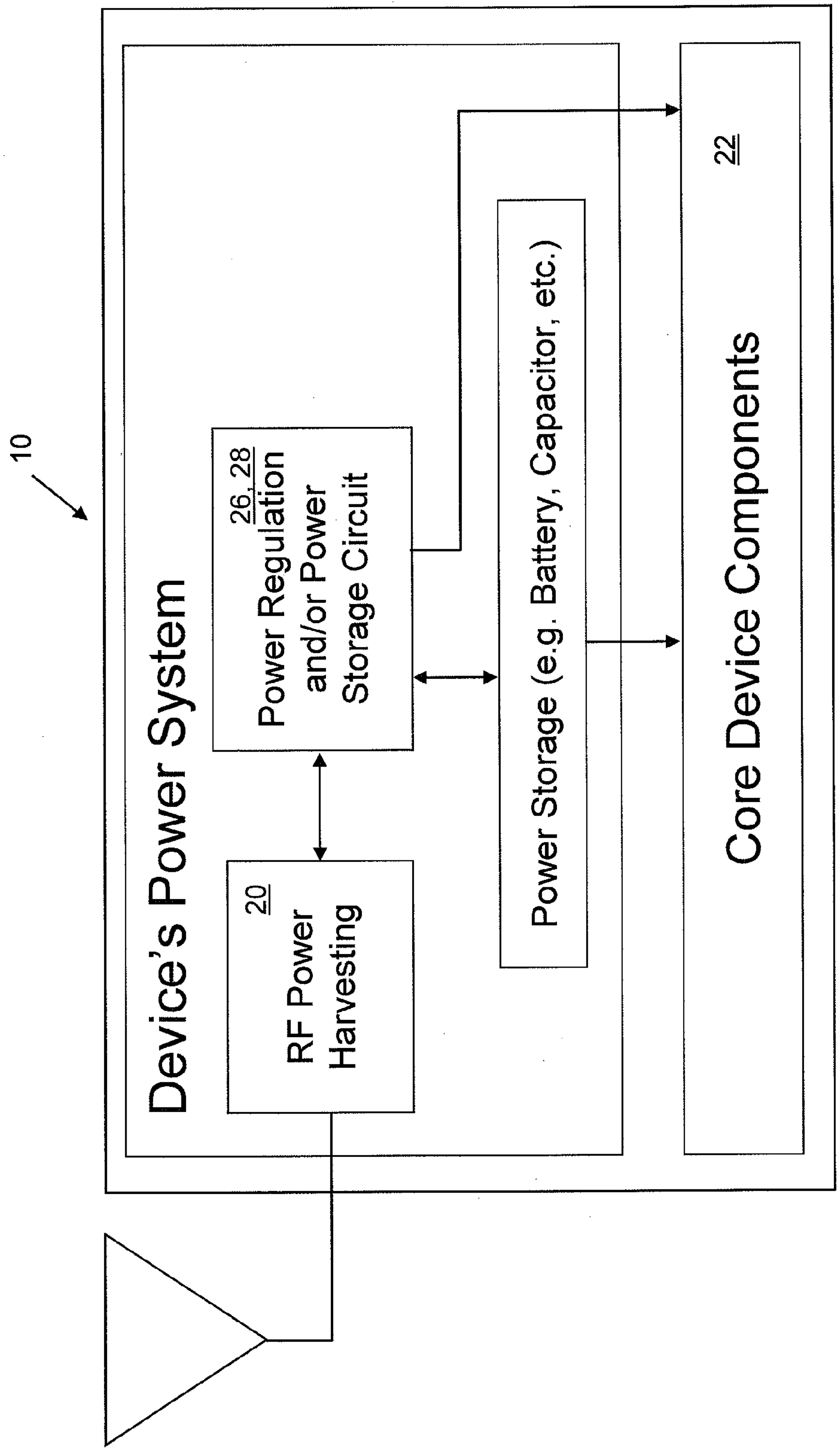
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 13

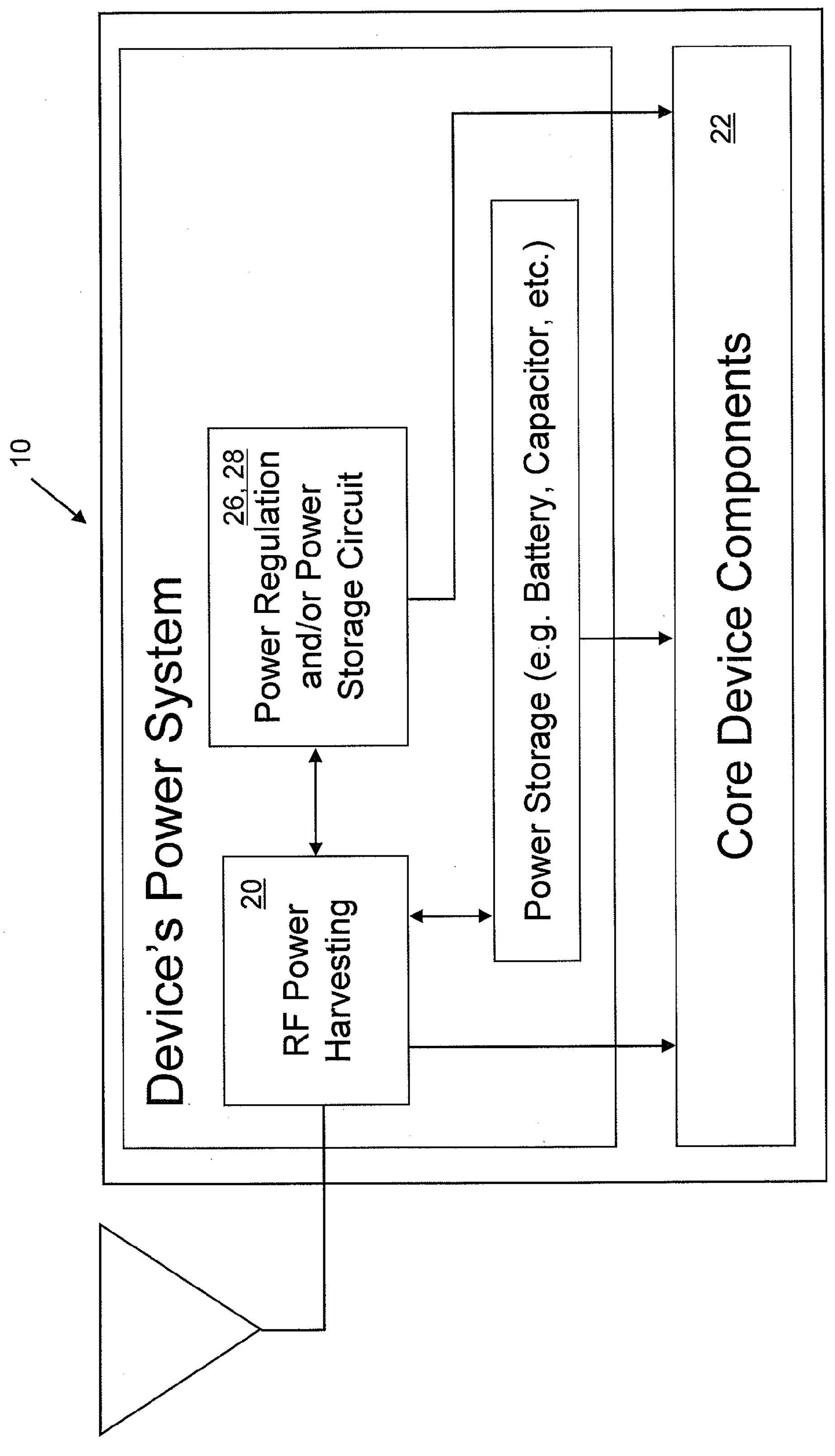
10
Device's Power System
26, 28
Power Regulation and/or Power Storage Circuit
20
RF Power Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 14

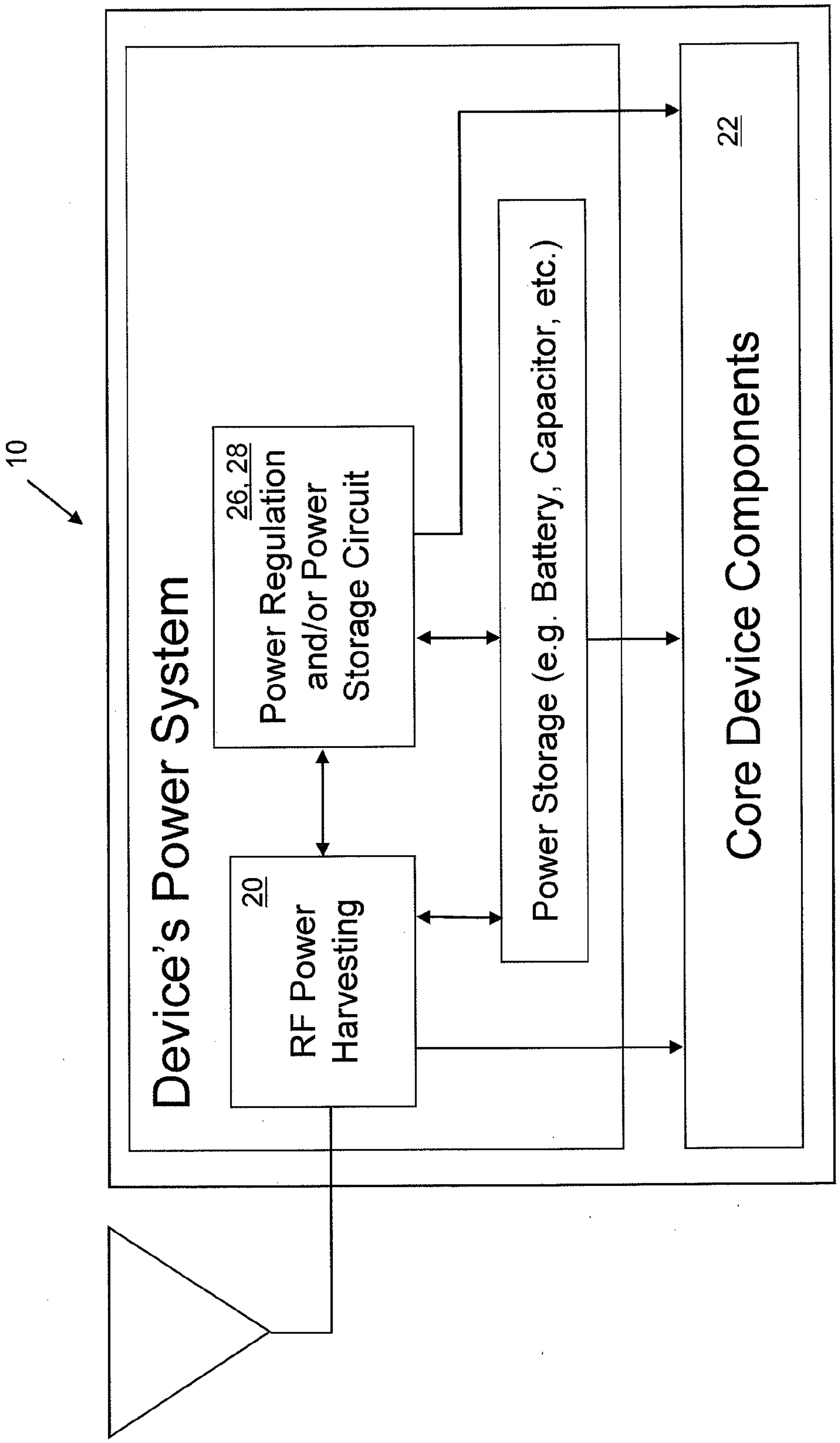
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 15

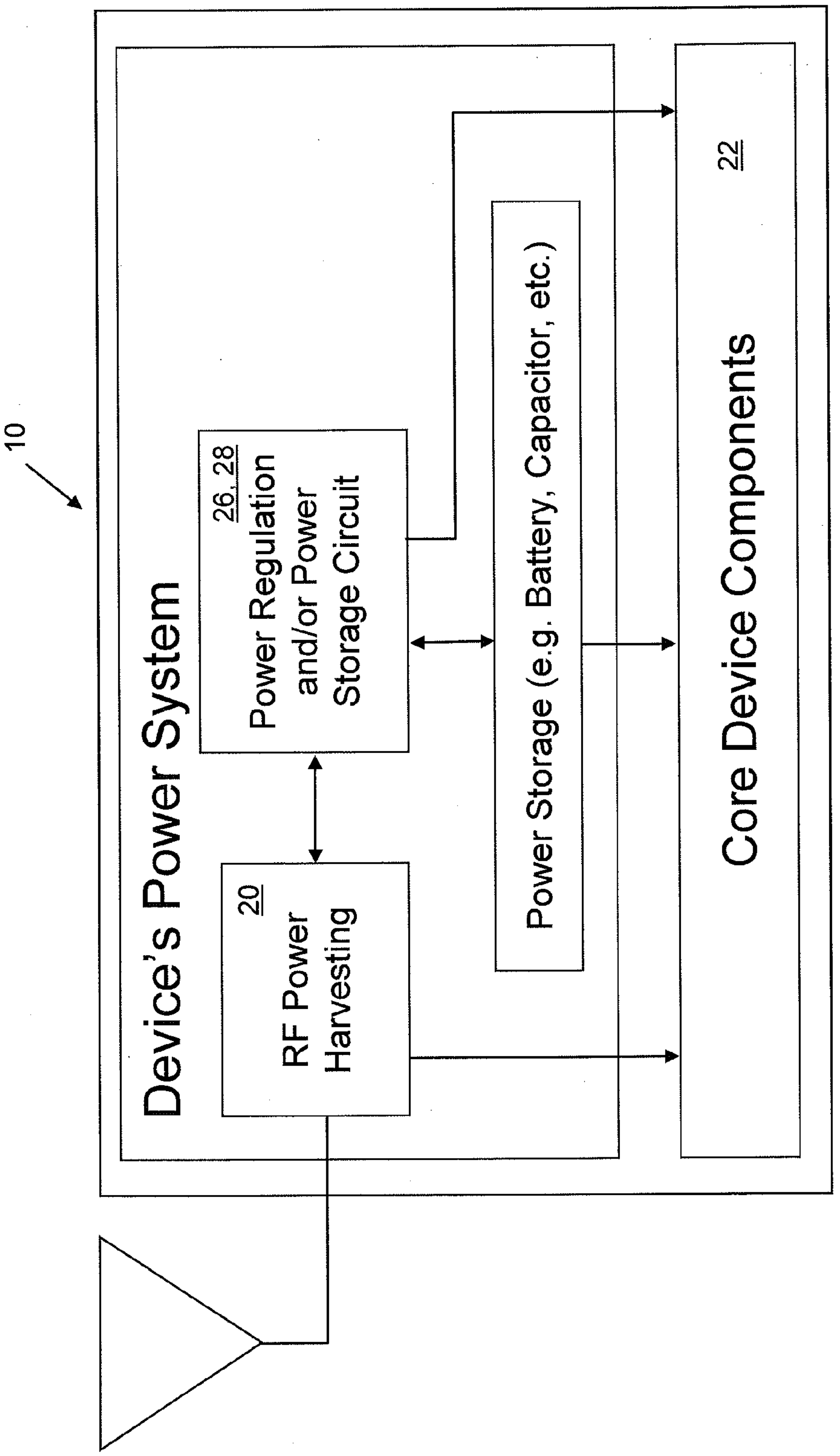
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 16

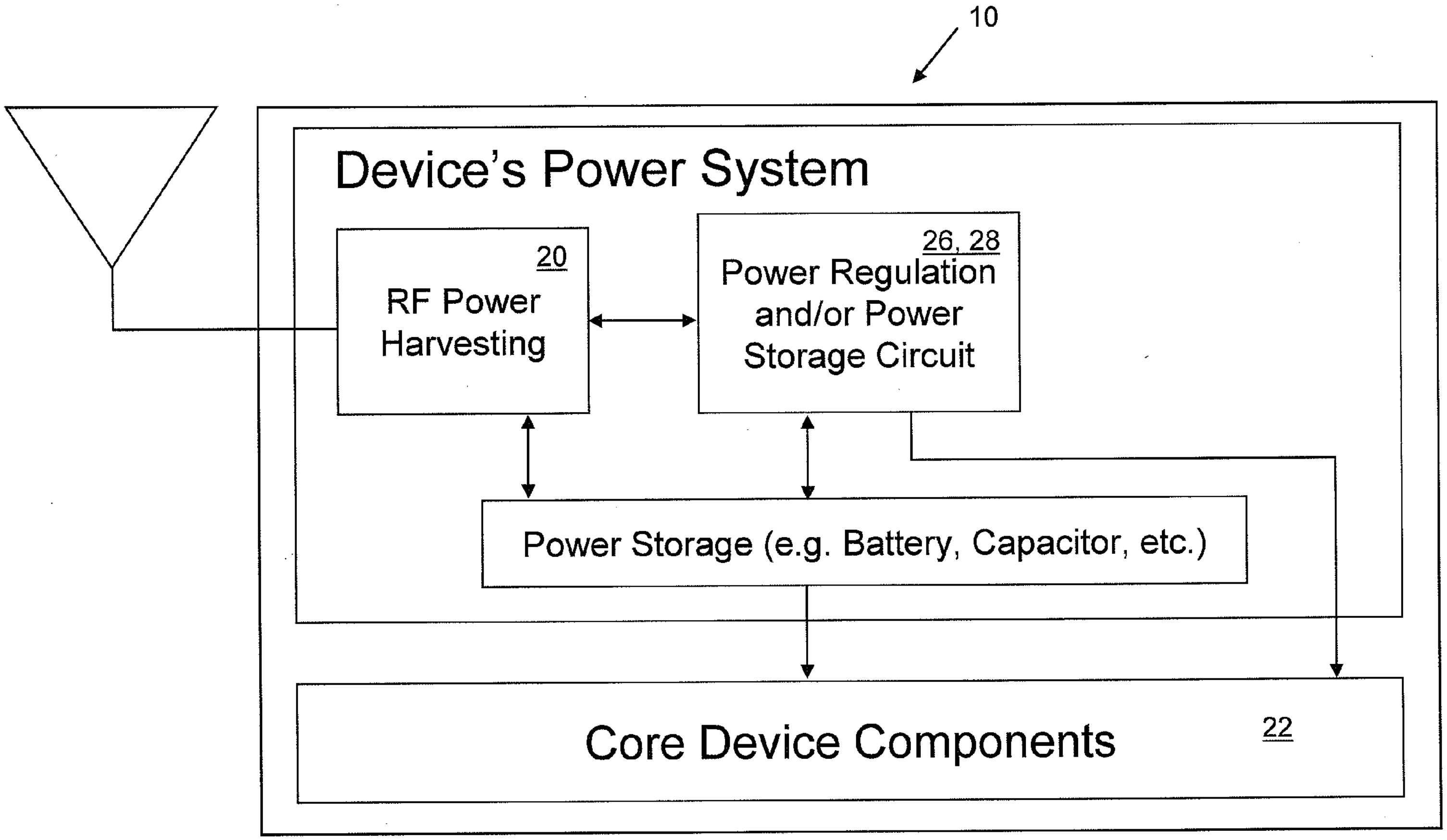
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 17

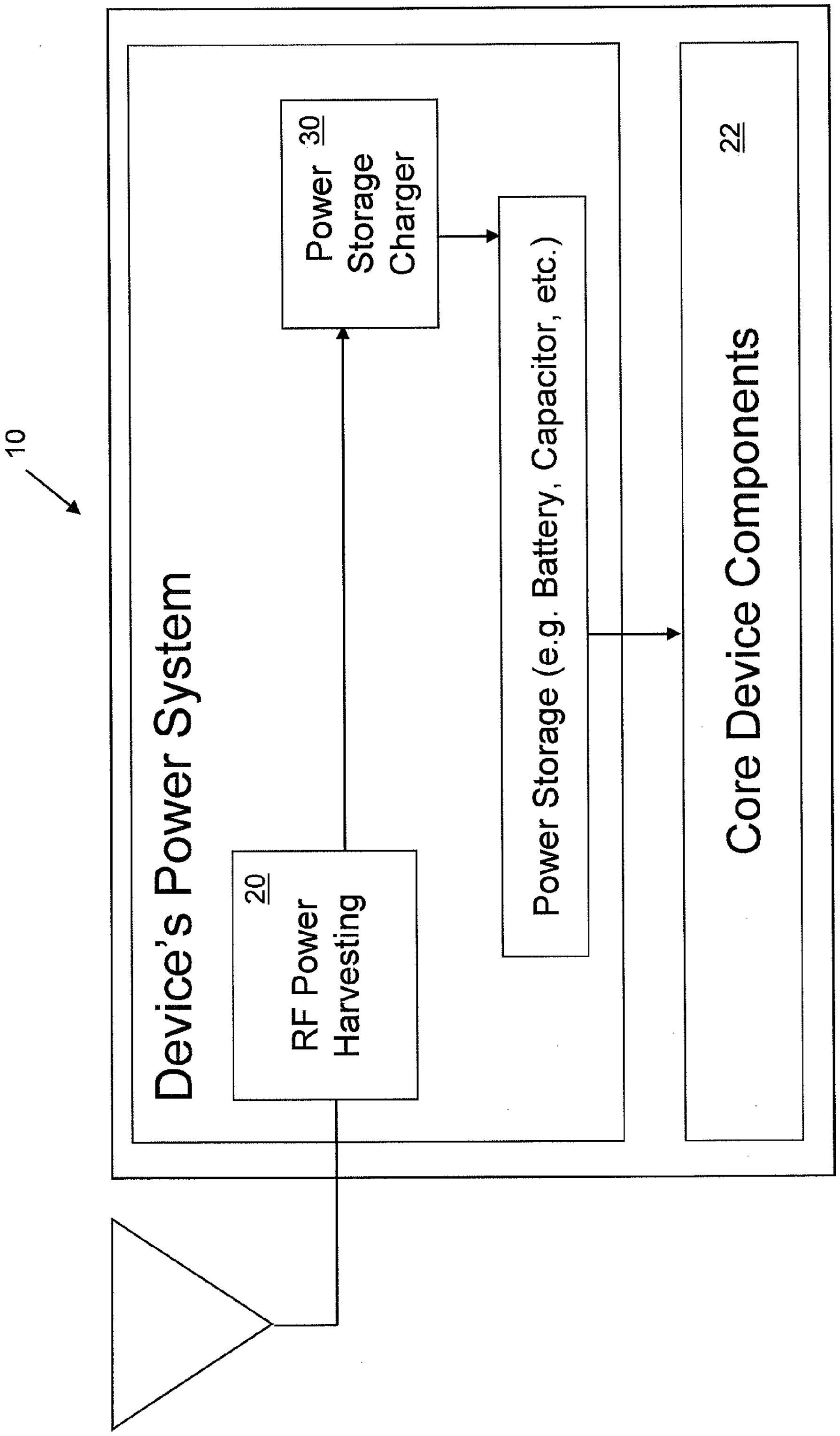
10
Device's Power System
RF Power Harvesting 20
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 18

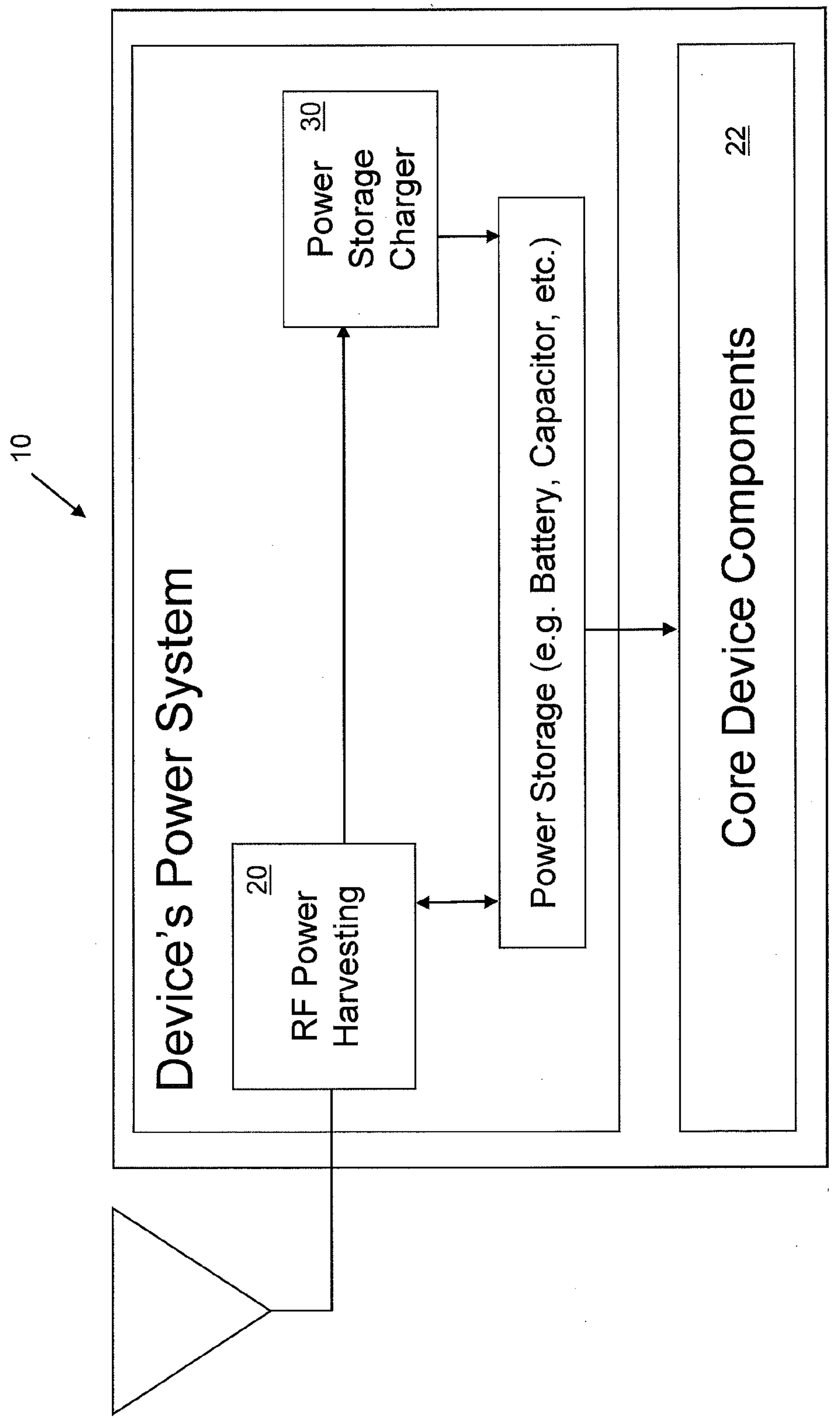
10
Device's Power System
RF Power Harvesting
20
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 19

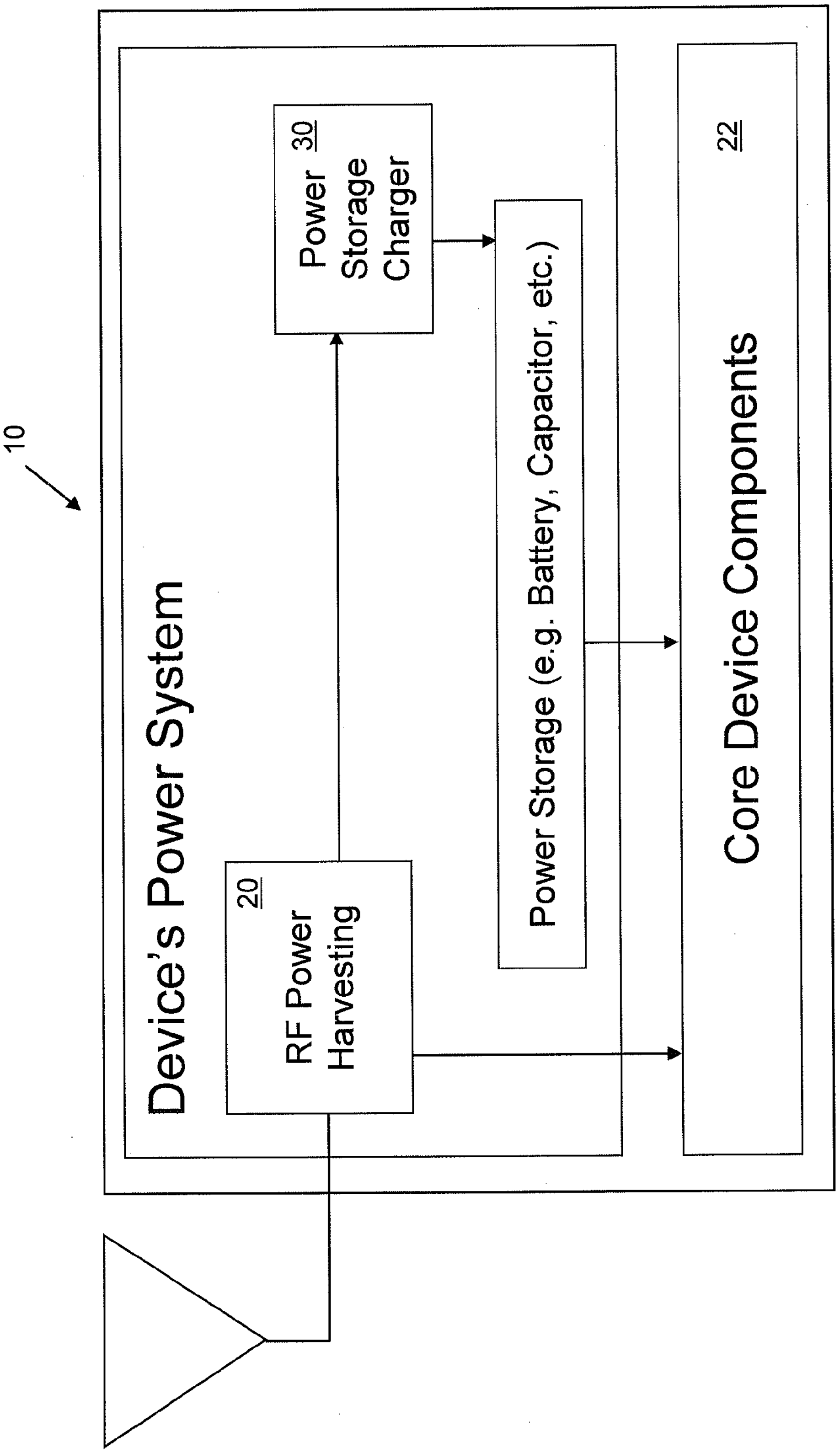
10
Device's Power System
20
RF Power Harvesting
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 20

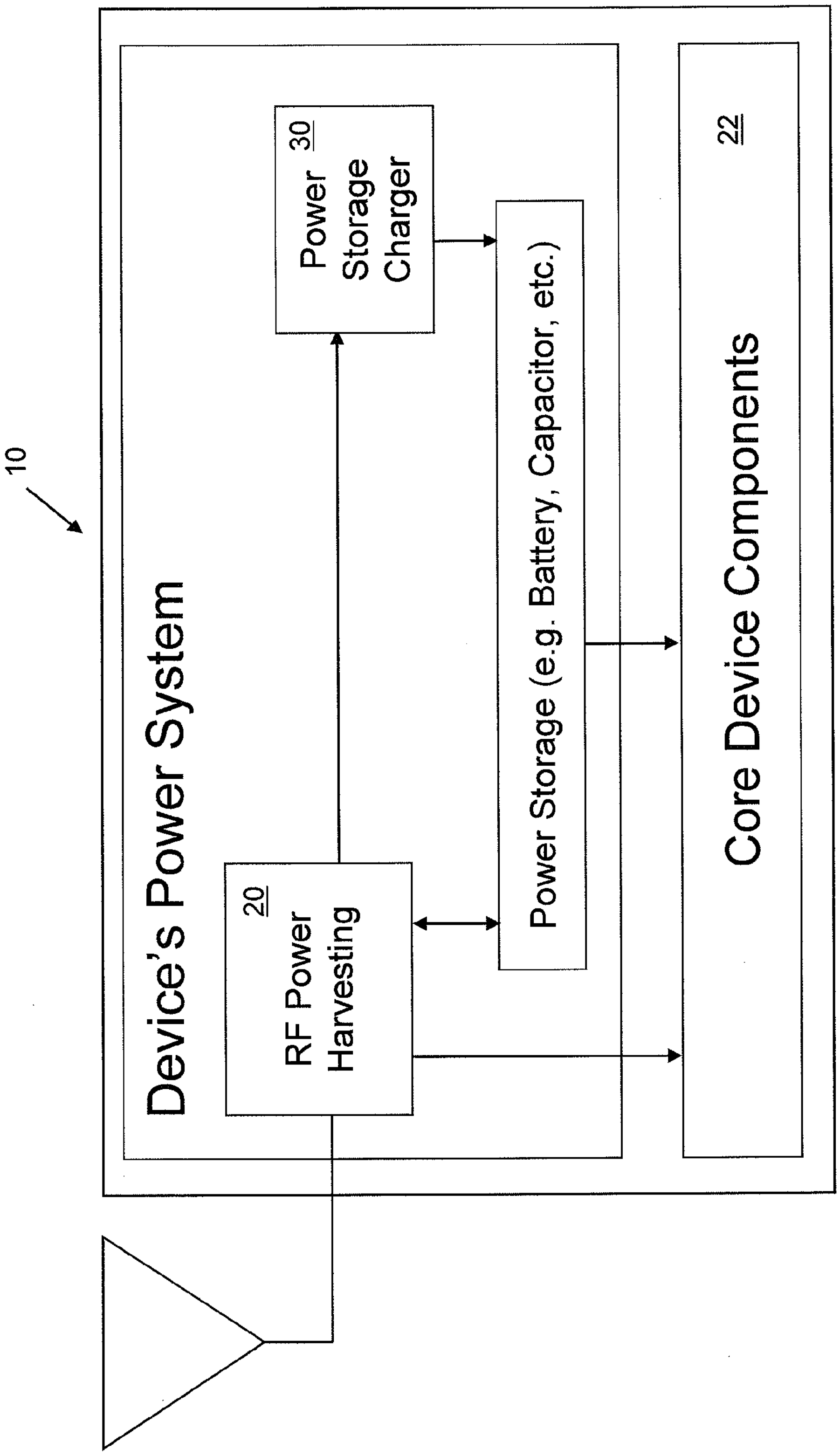
10
Device's Power System
20
RF Power Harvesting
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 21

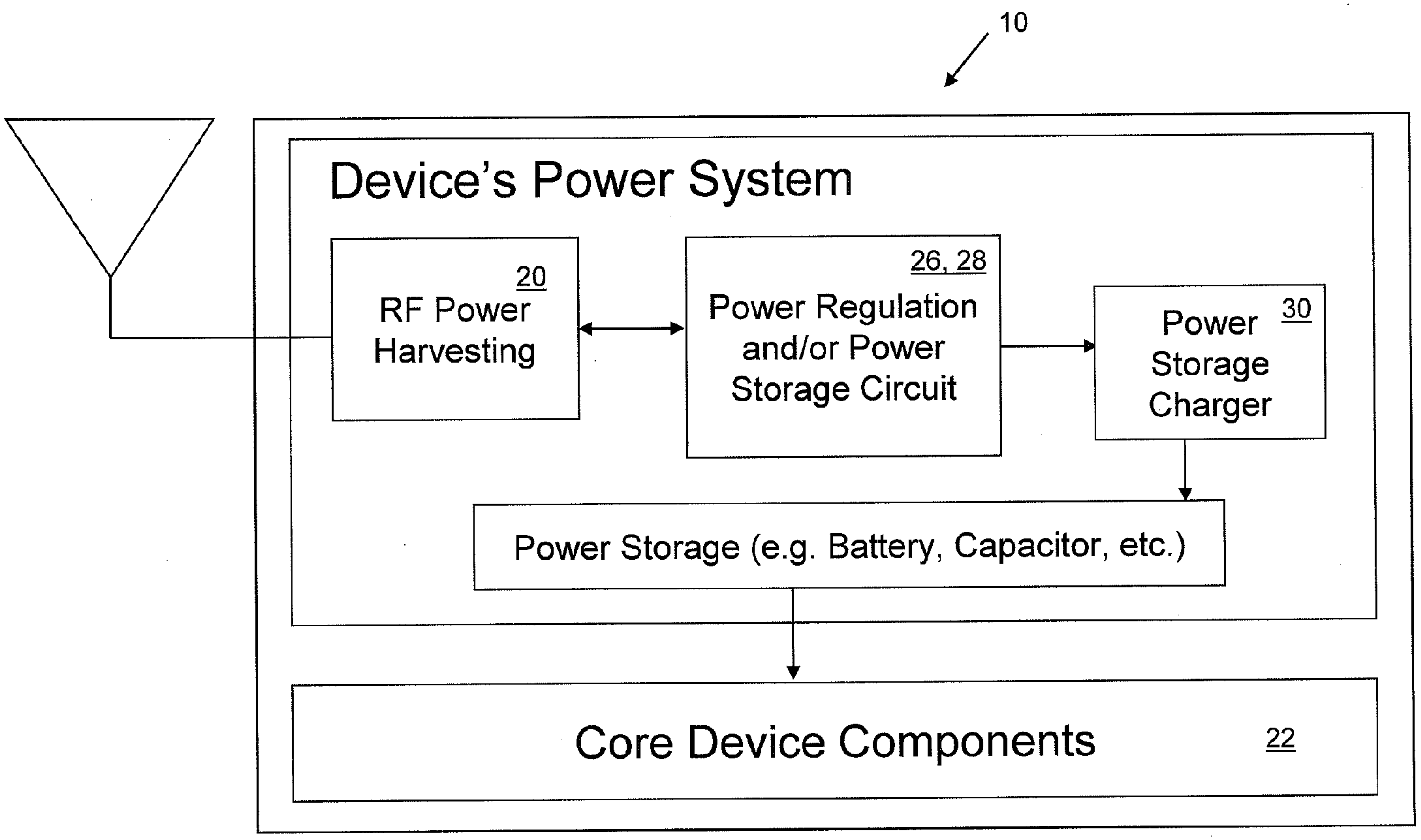
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 22

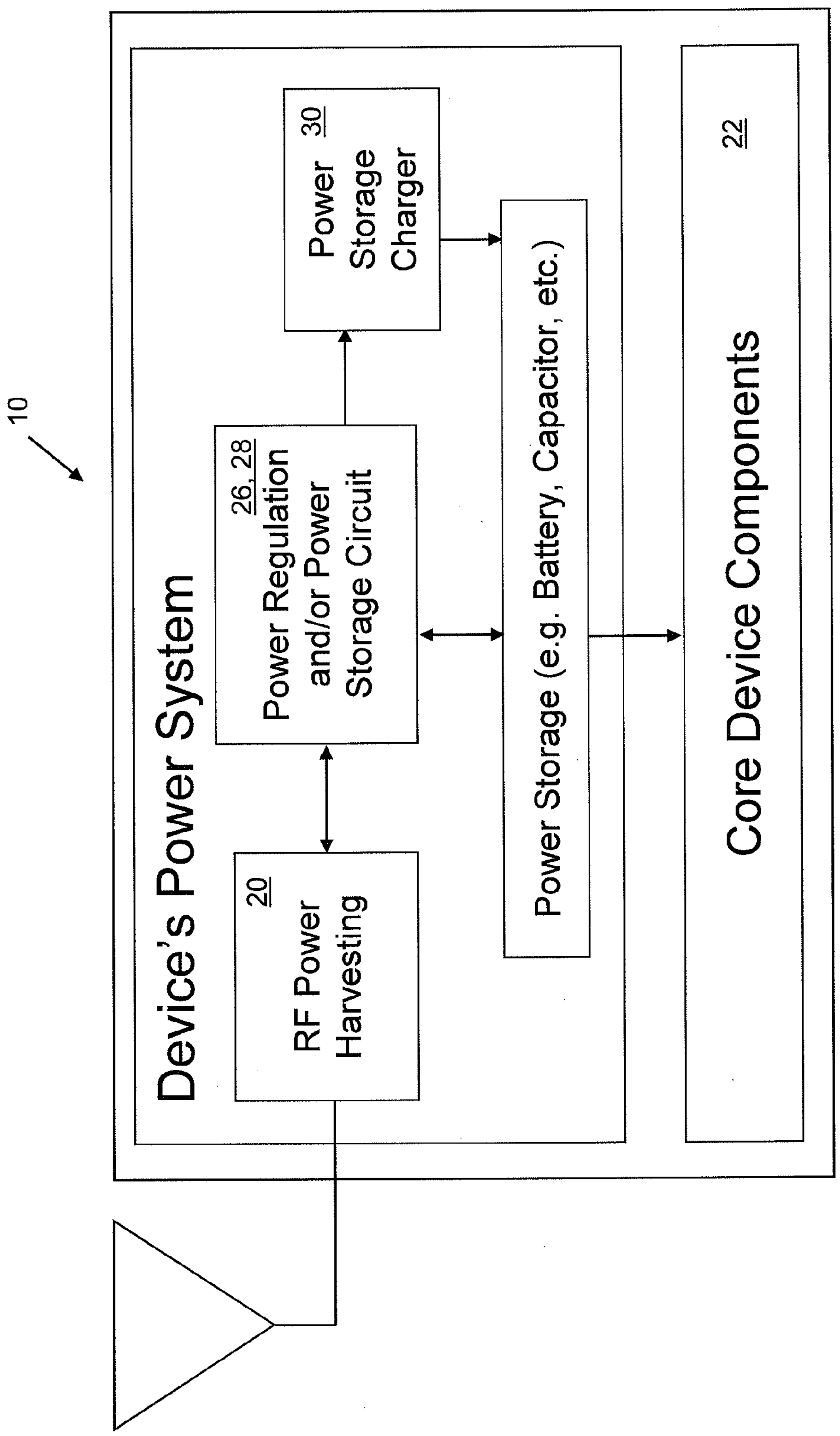
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 23

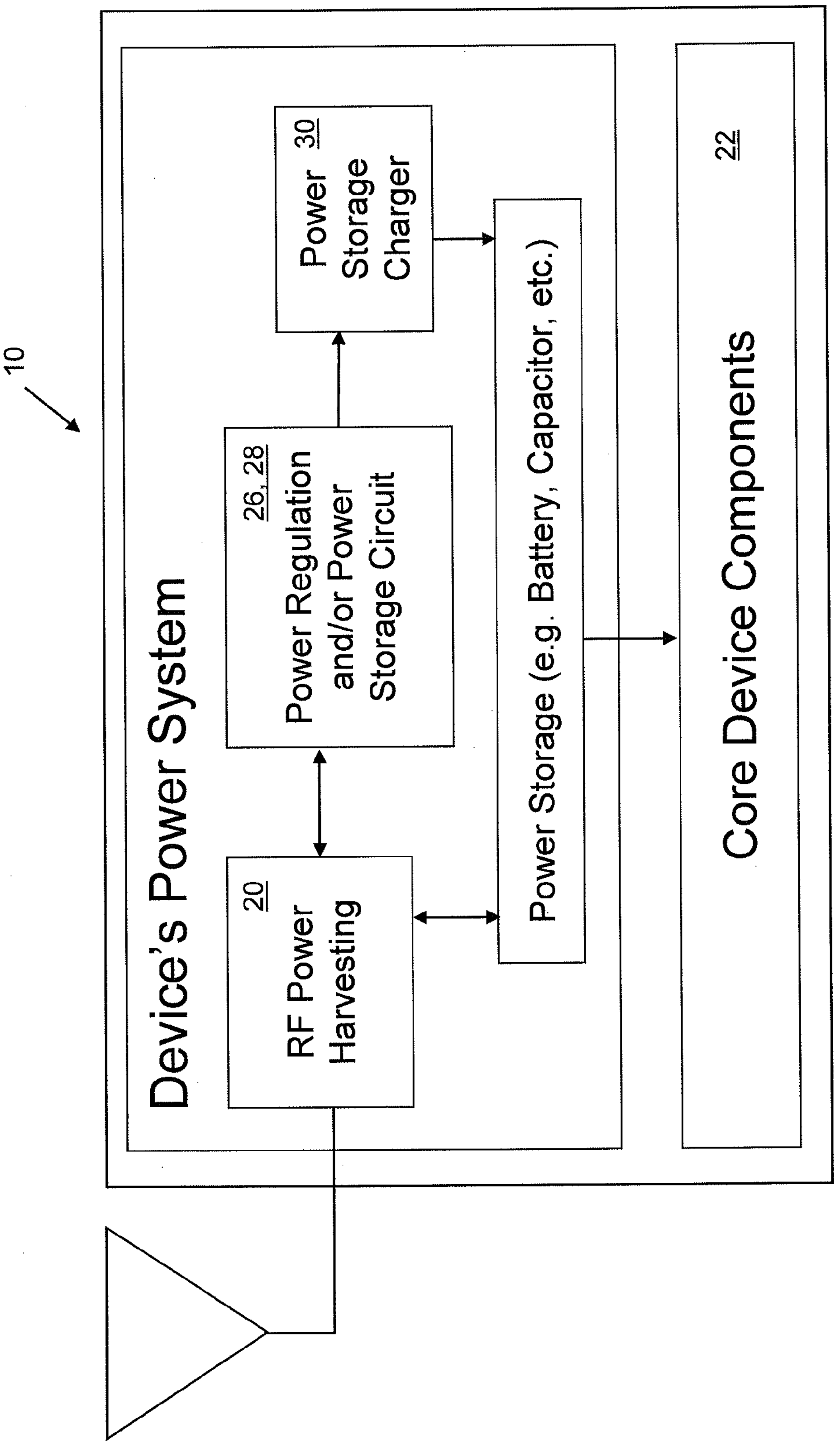
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 24

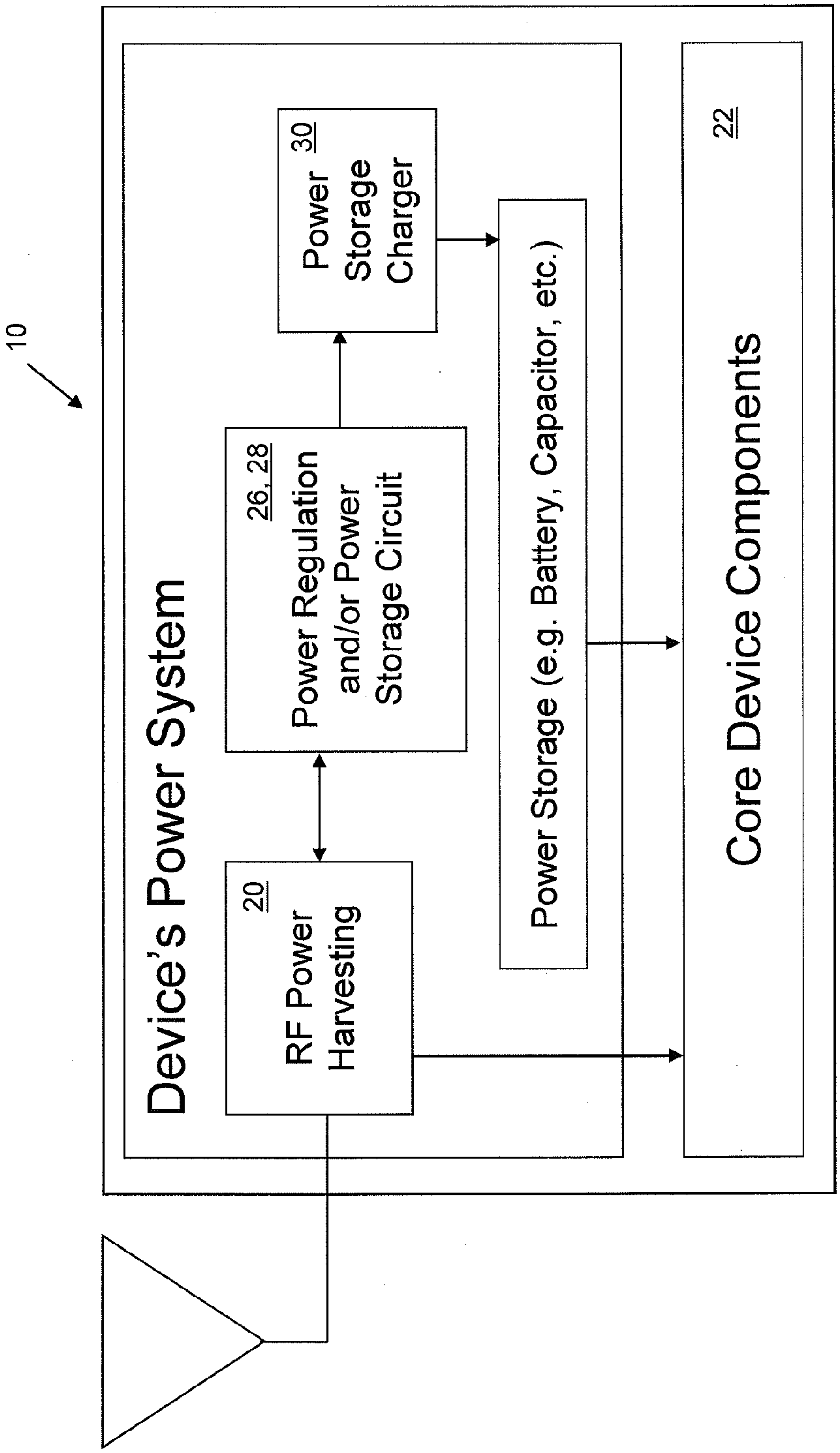
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 25

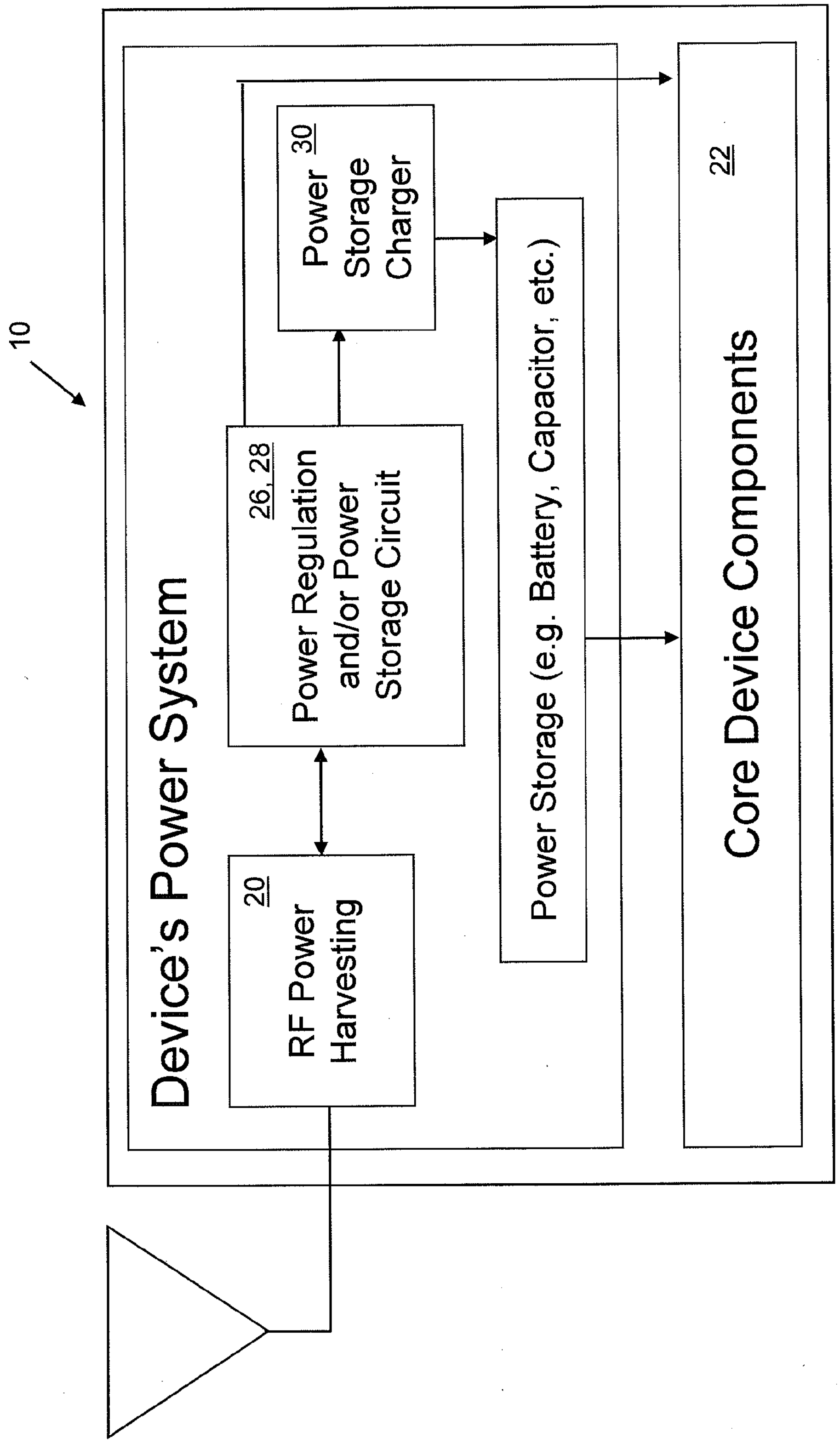
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 26

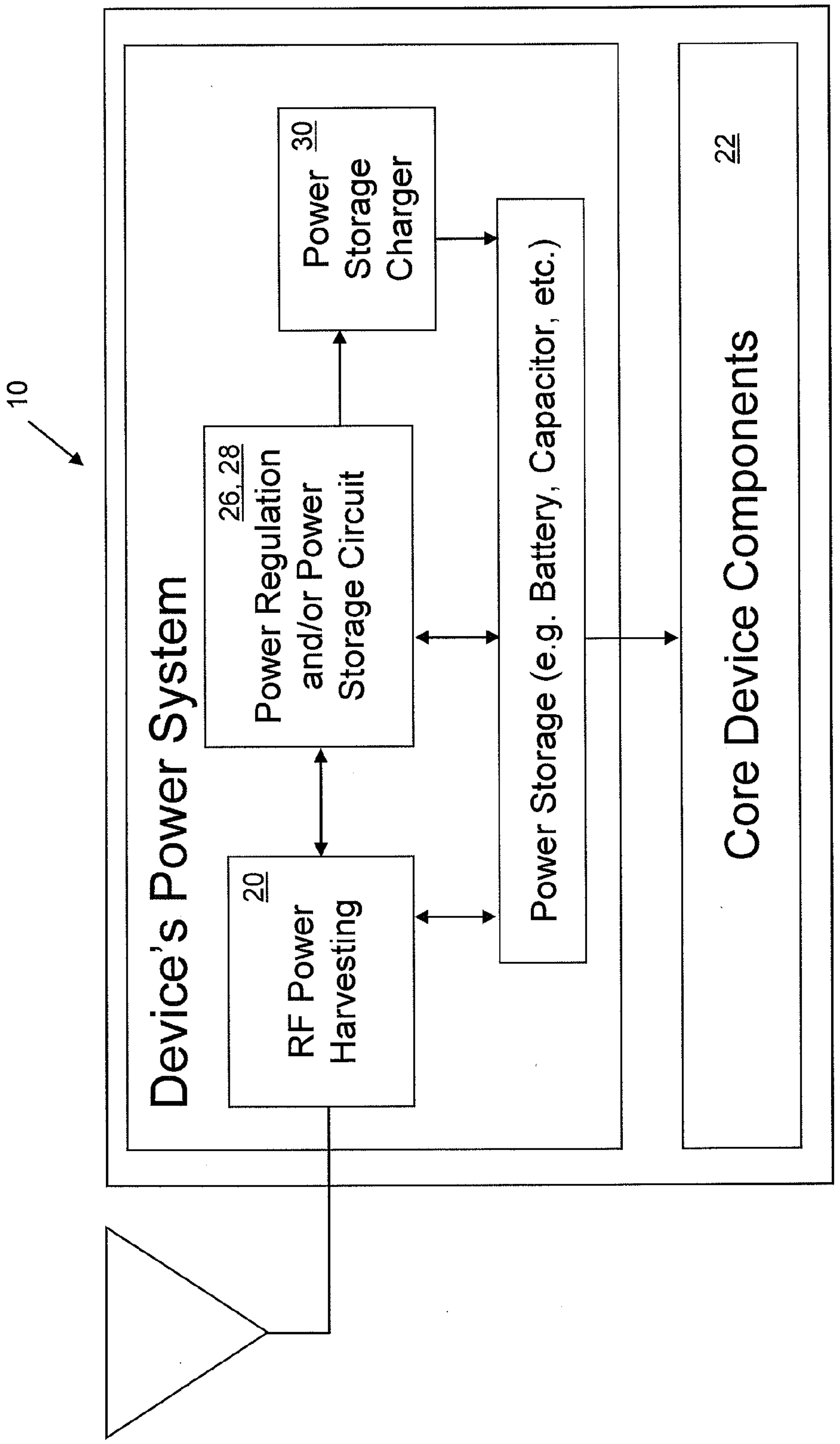
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 27

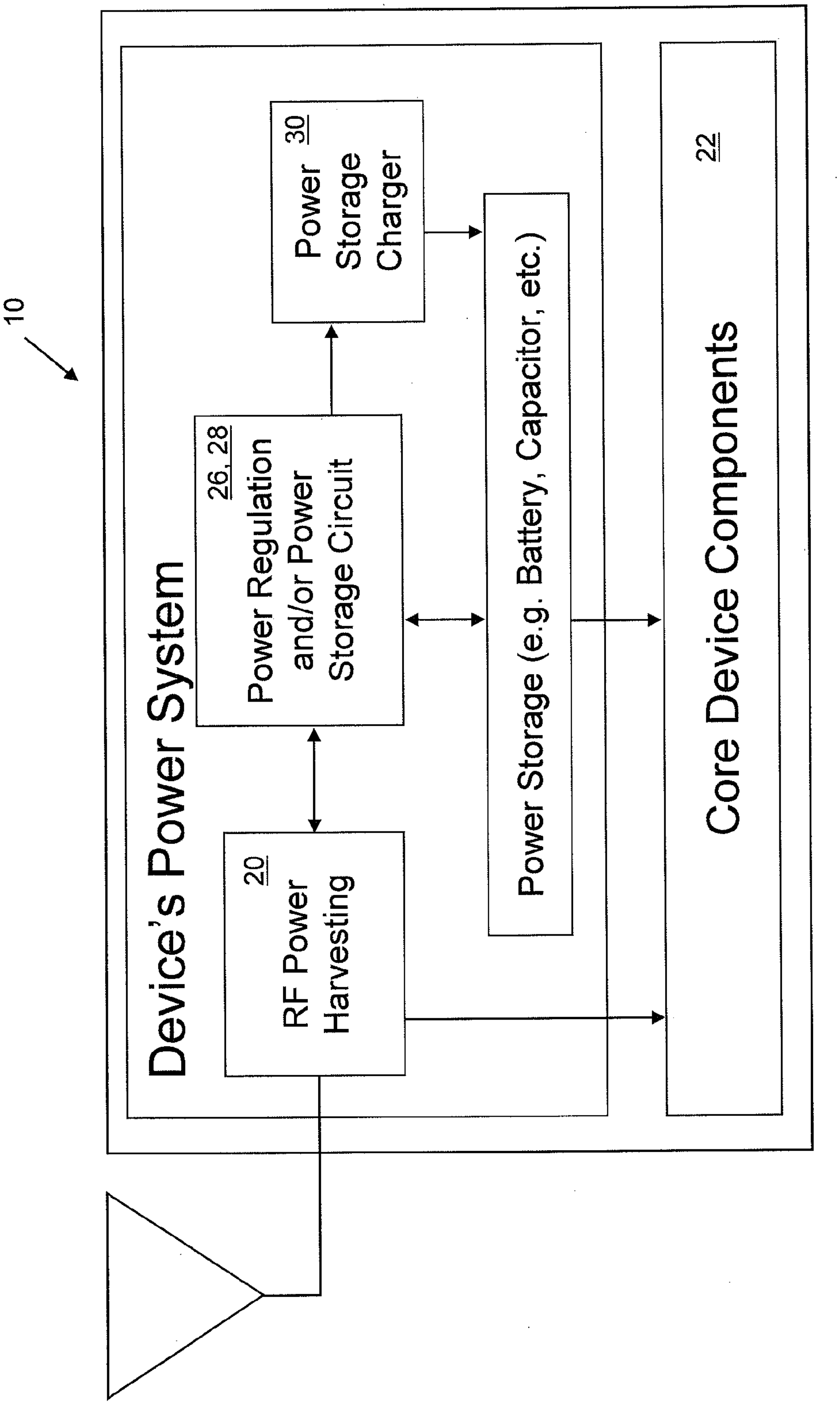
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 28

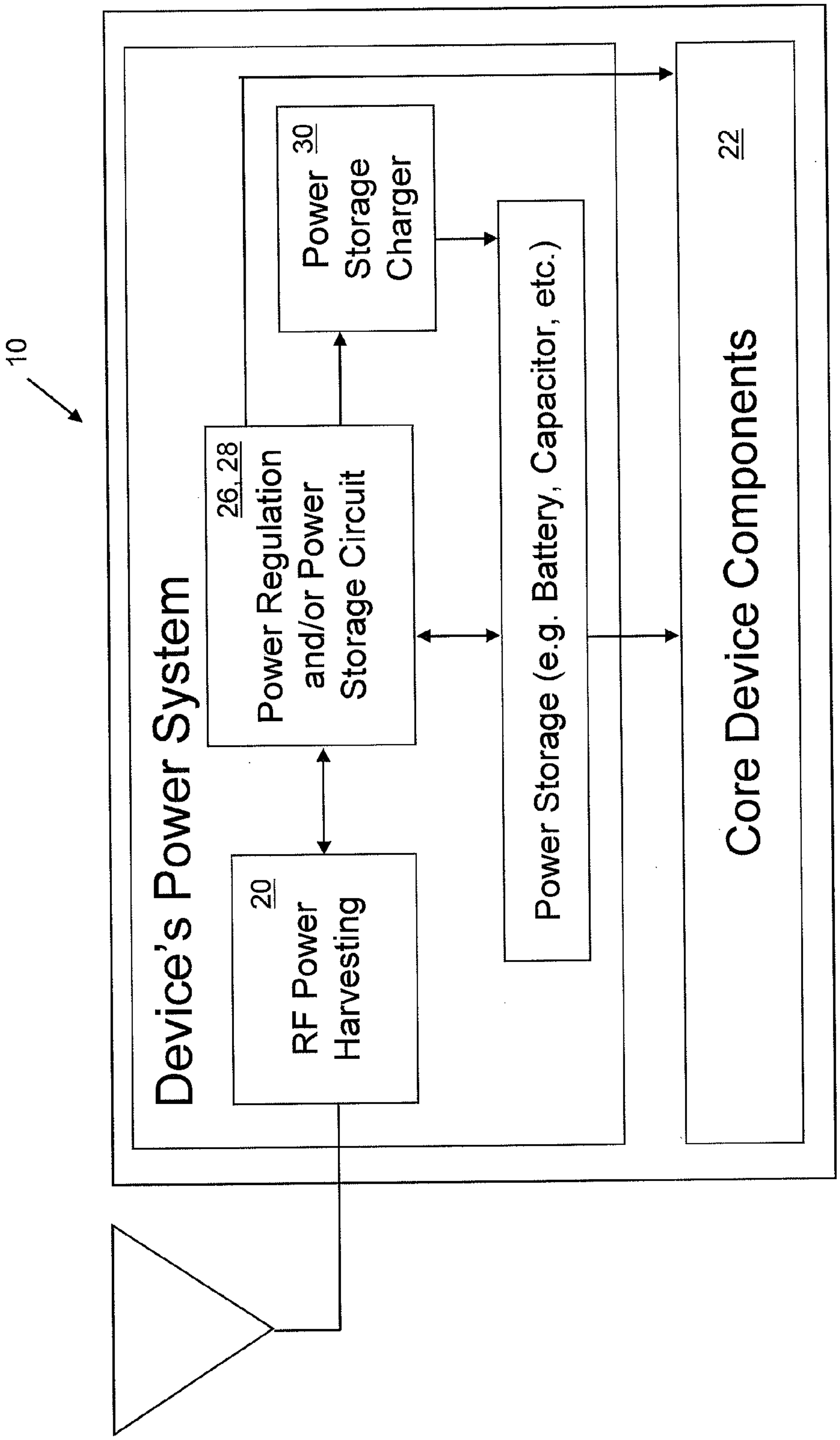
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 29

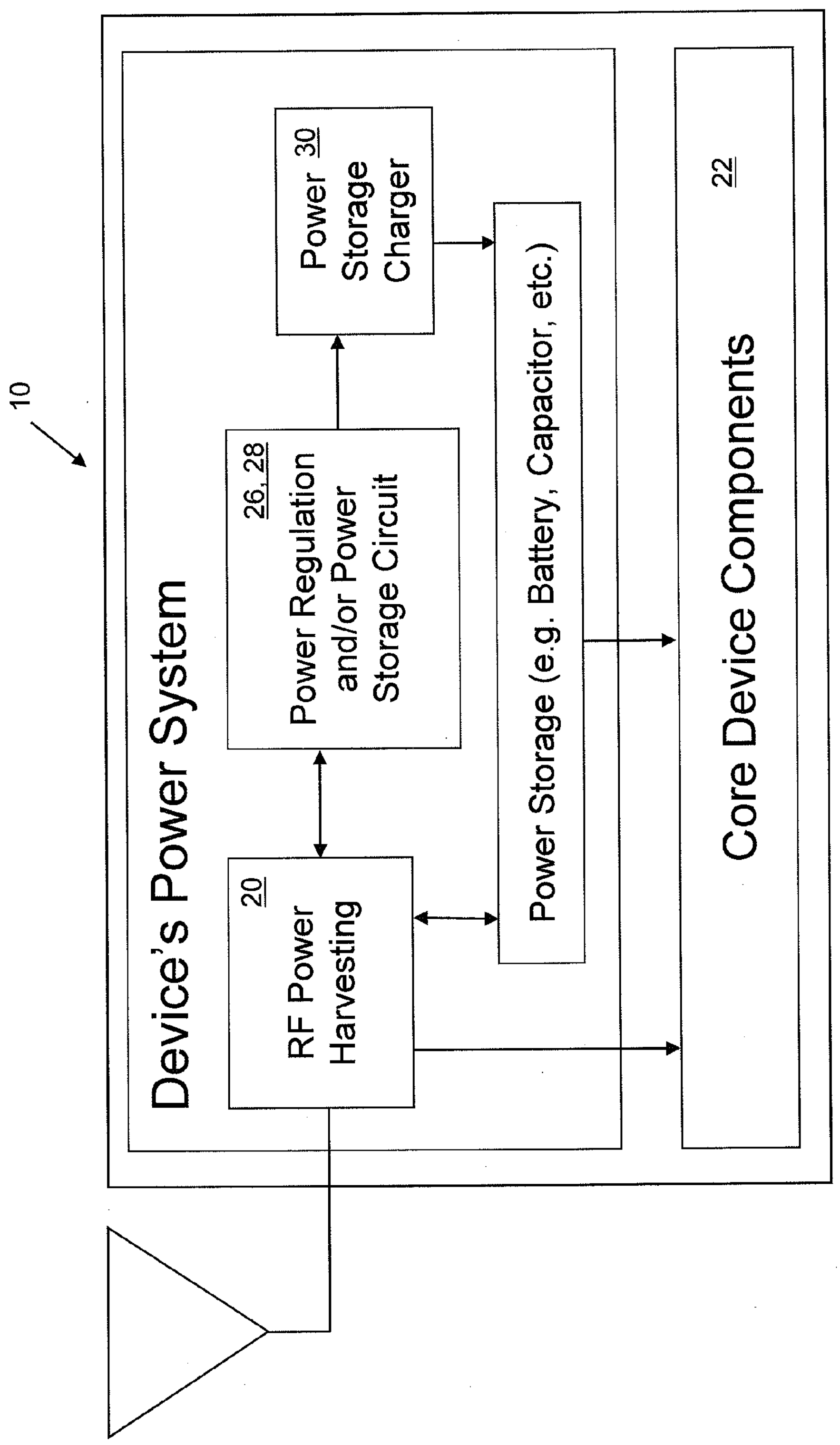
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 30

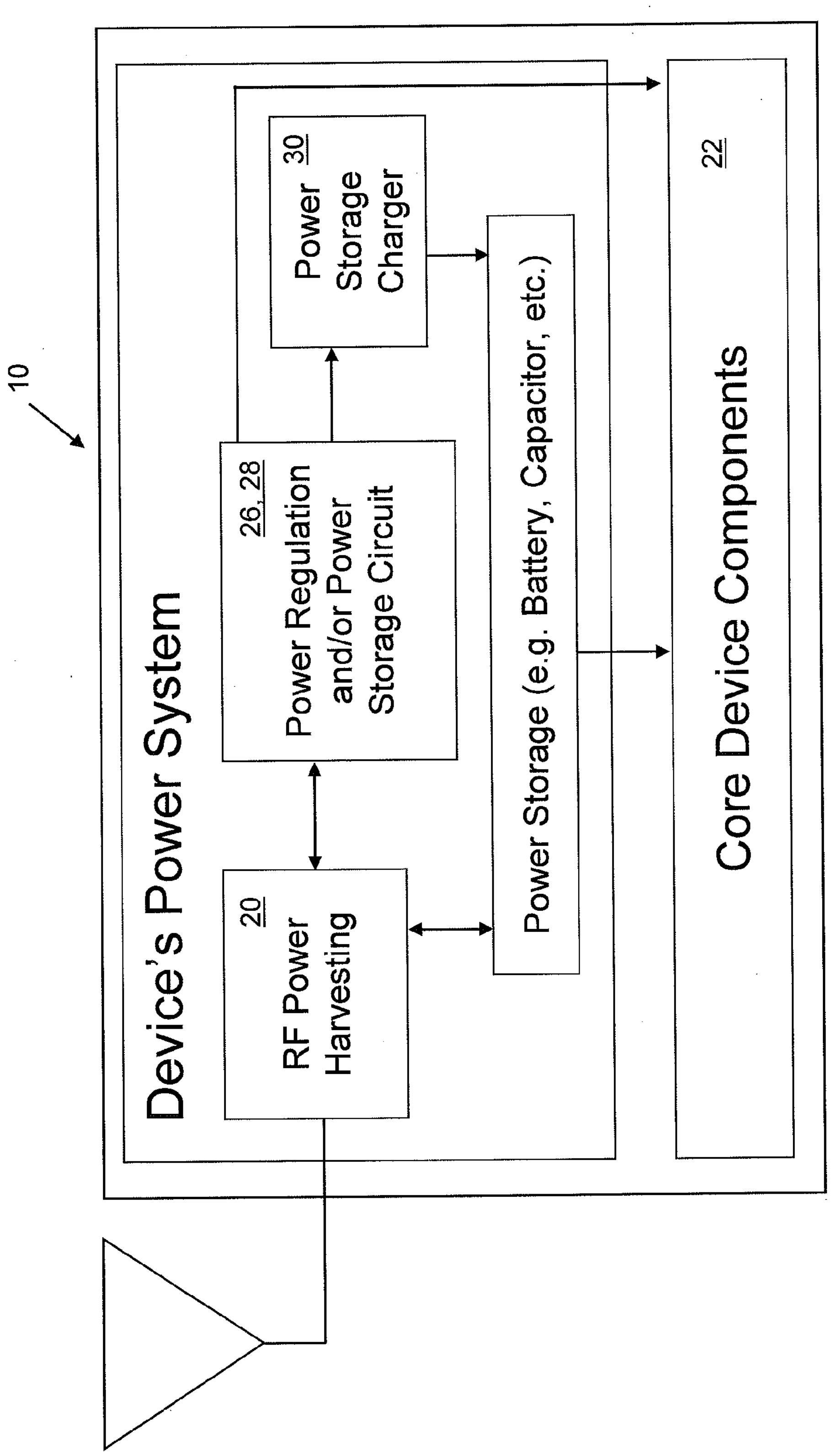
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 31

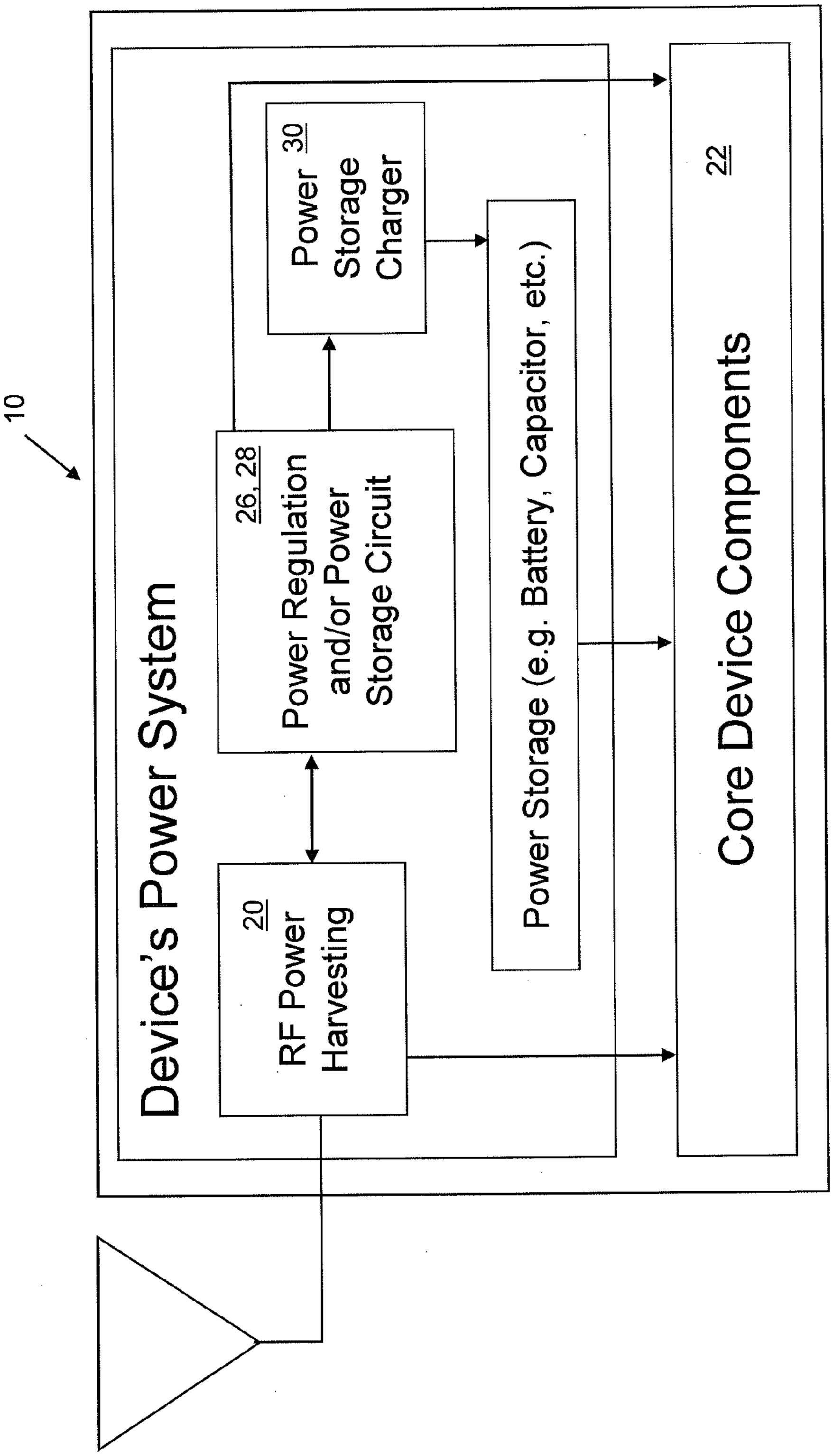
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 32

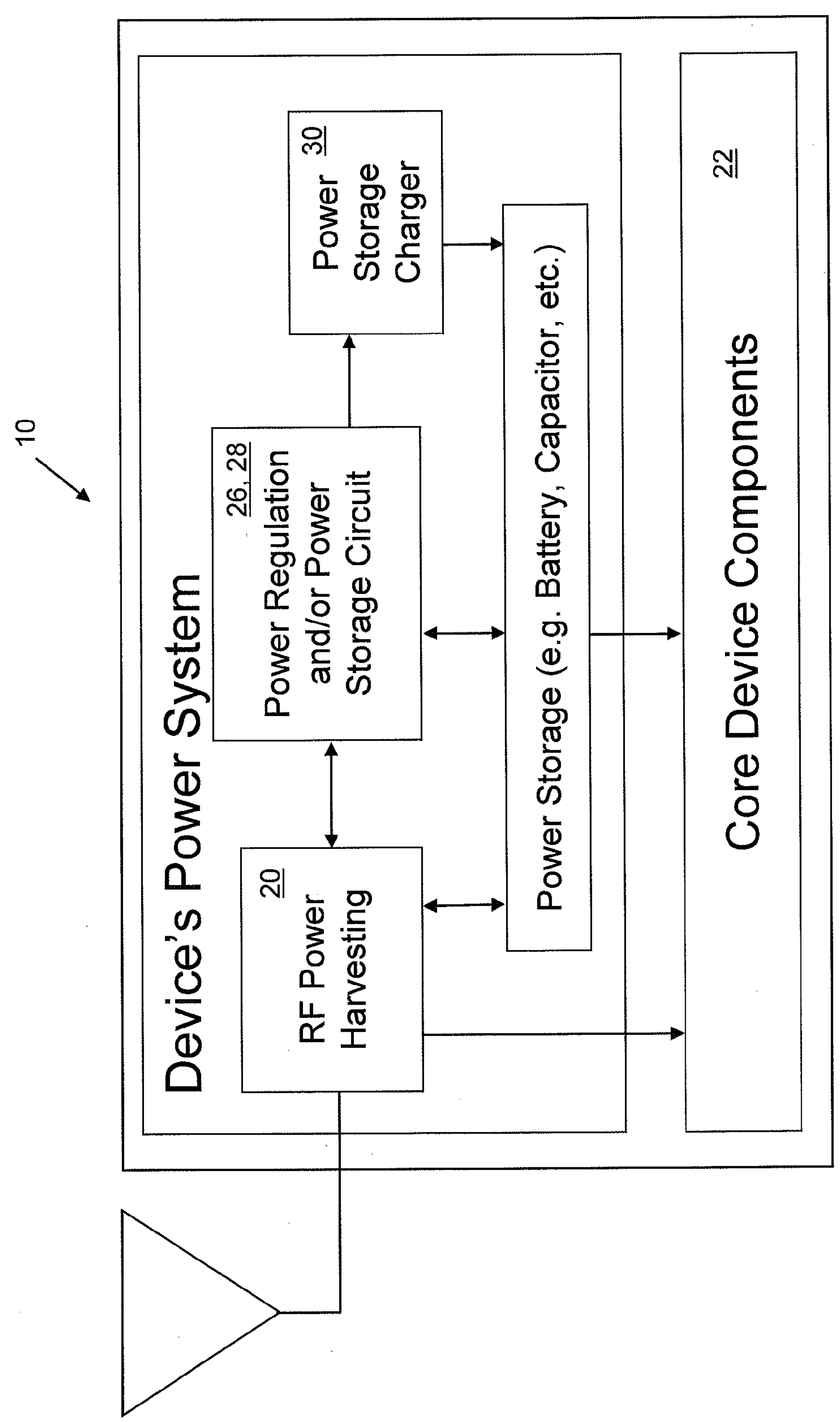
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 33

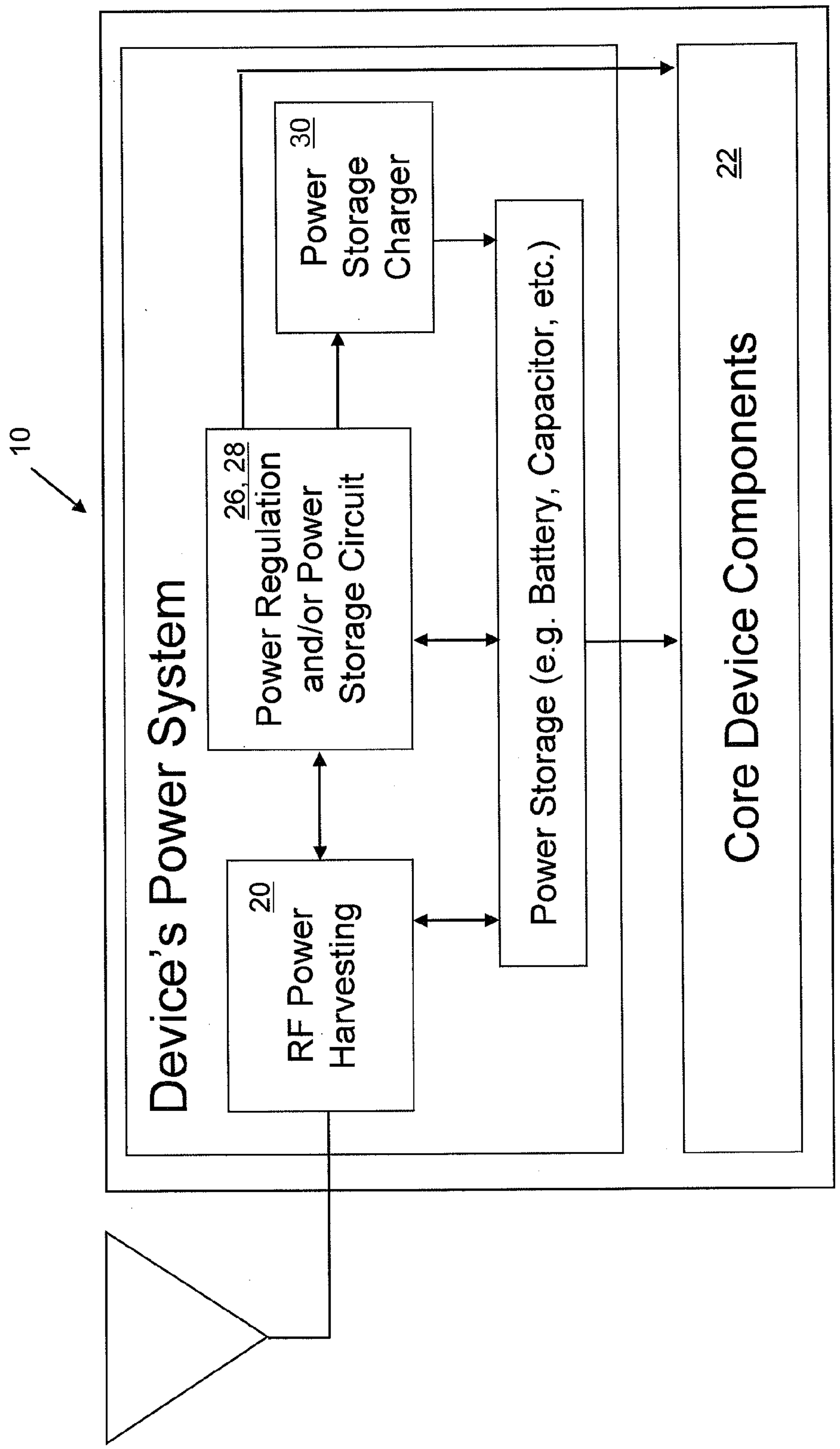
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 34

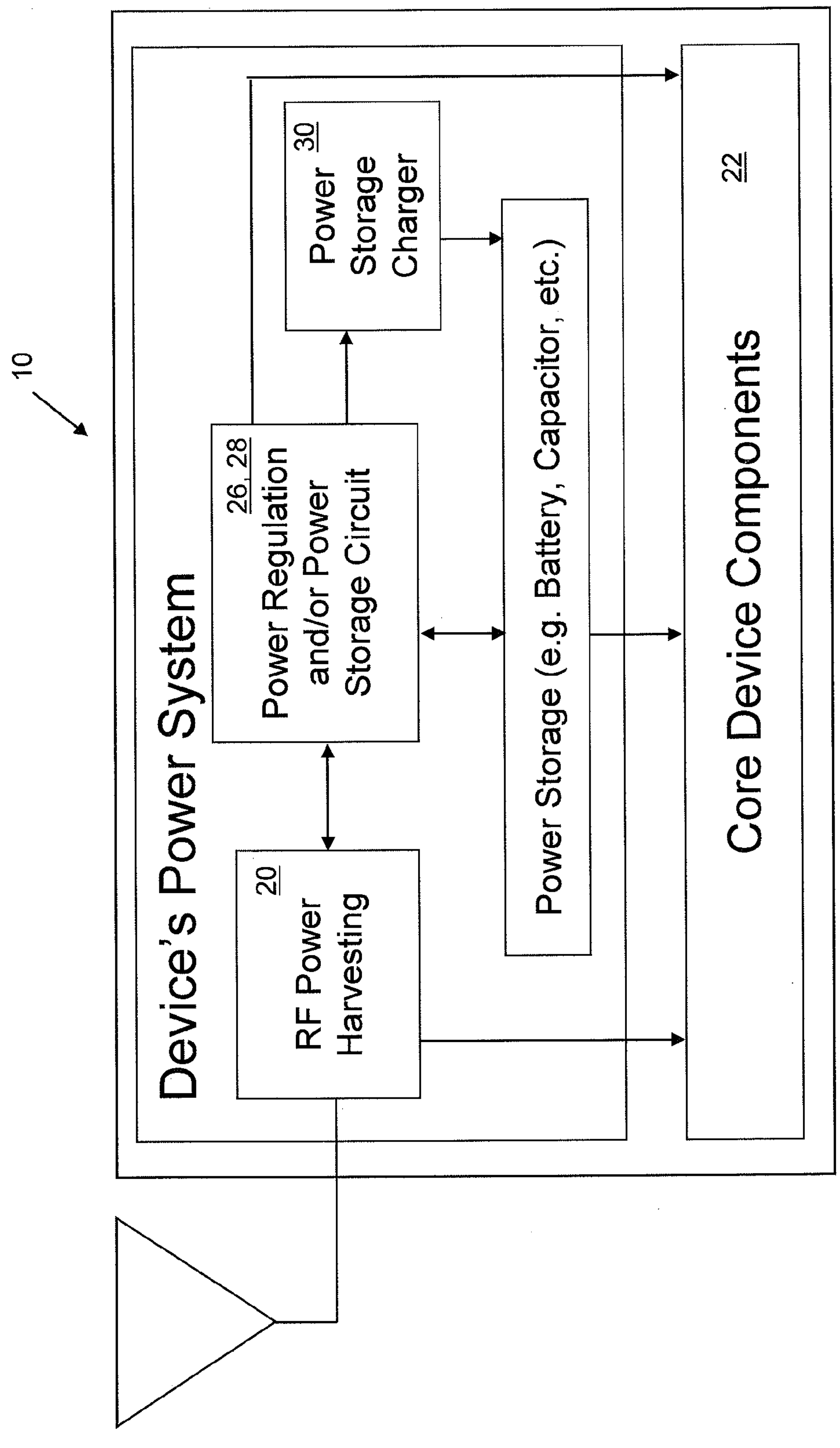
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 35

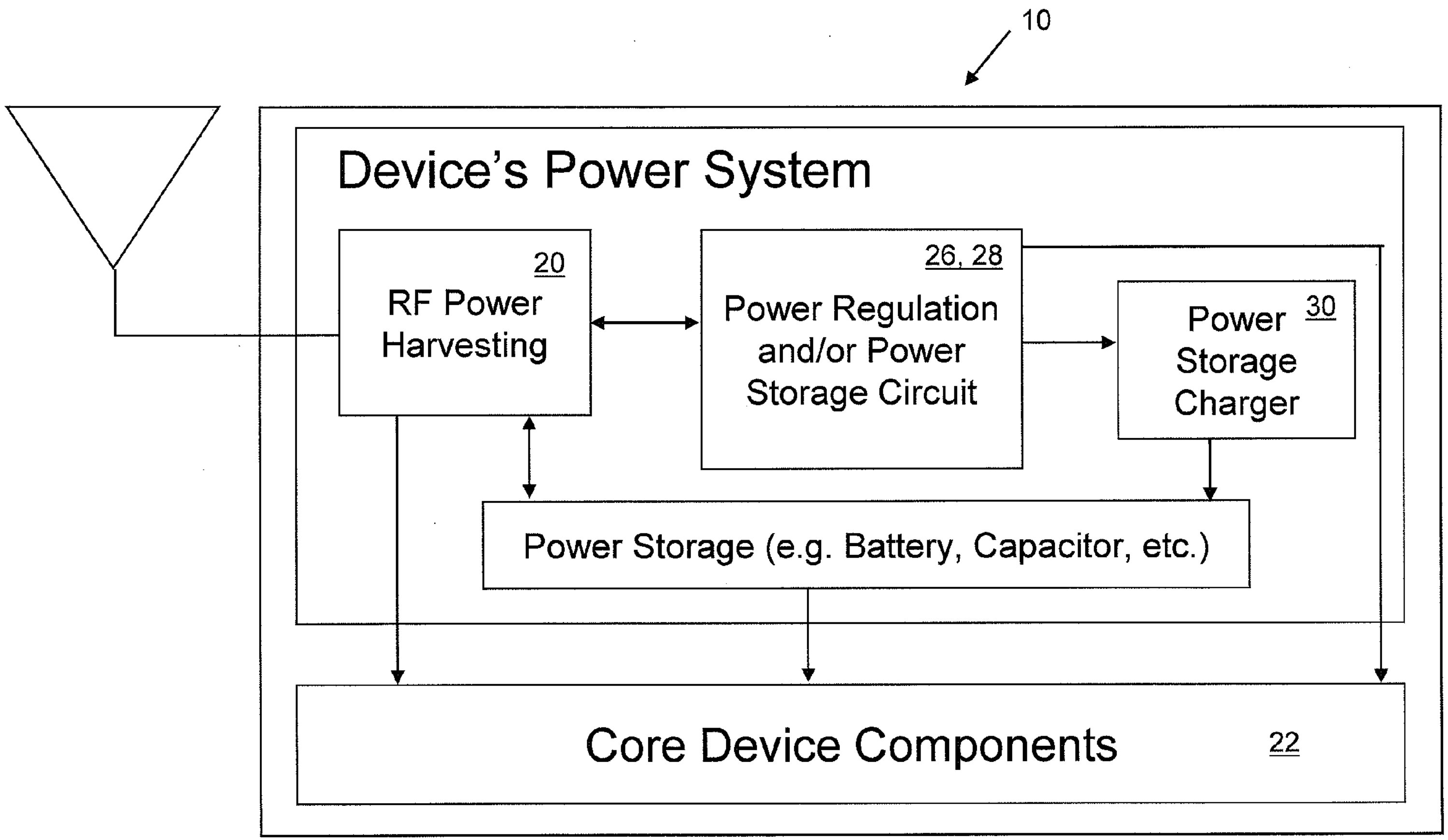
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 36

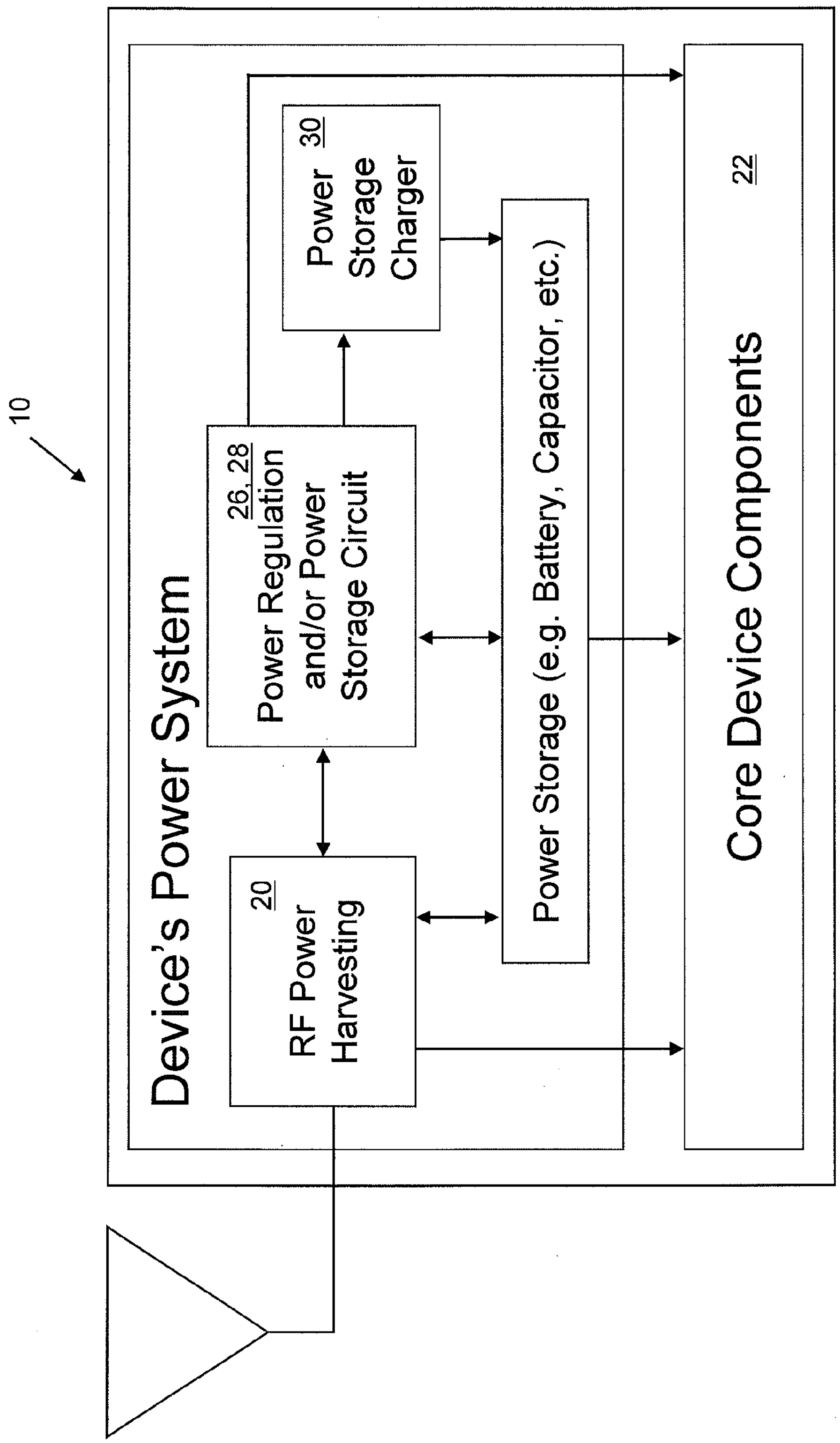
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 37

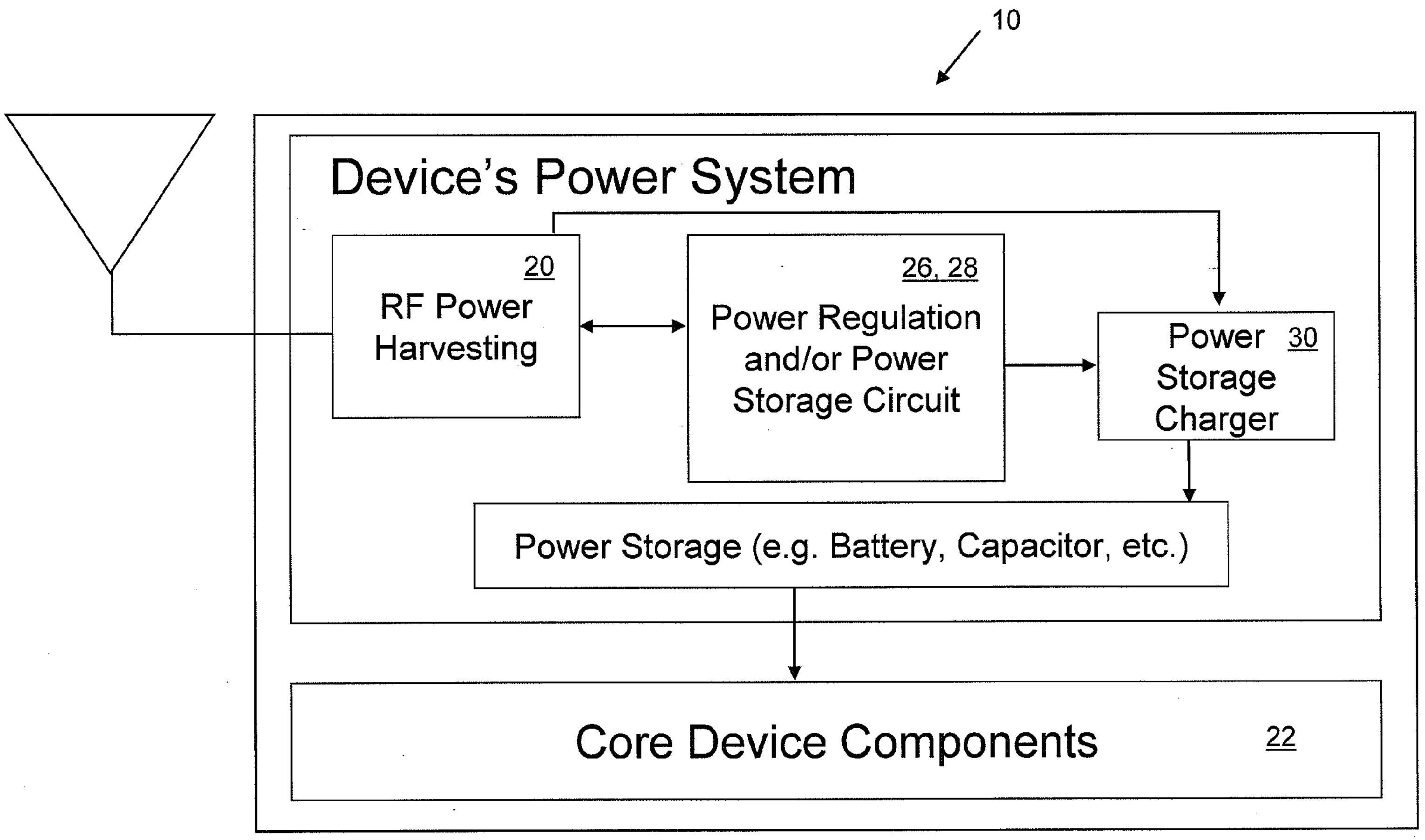
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 38

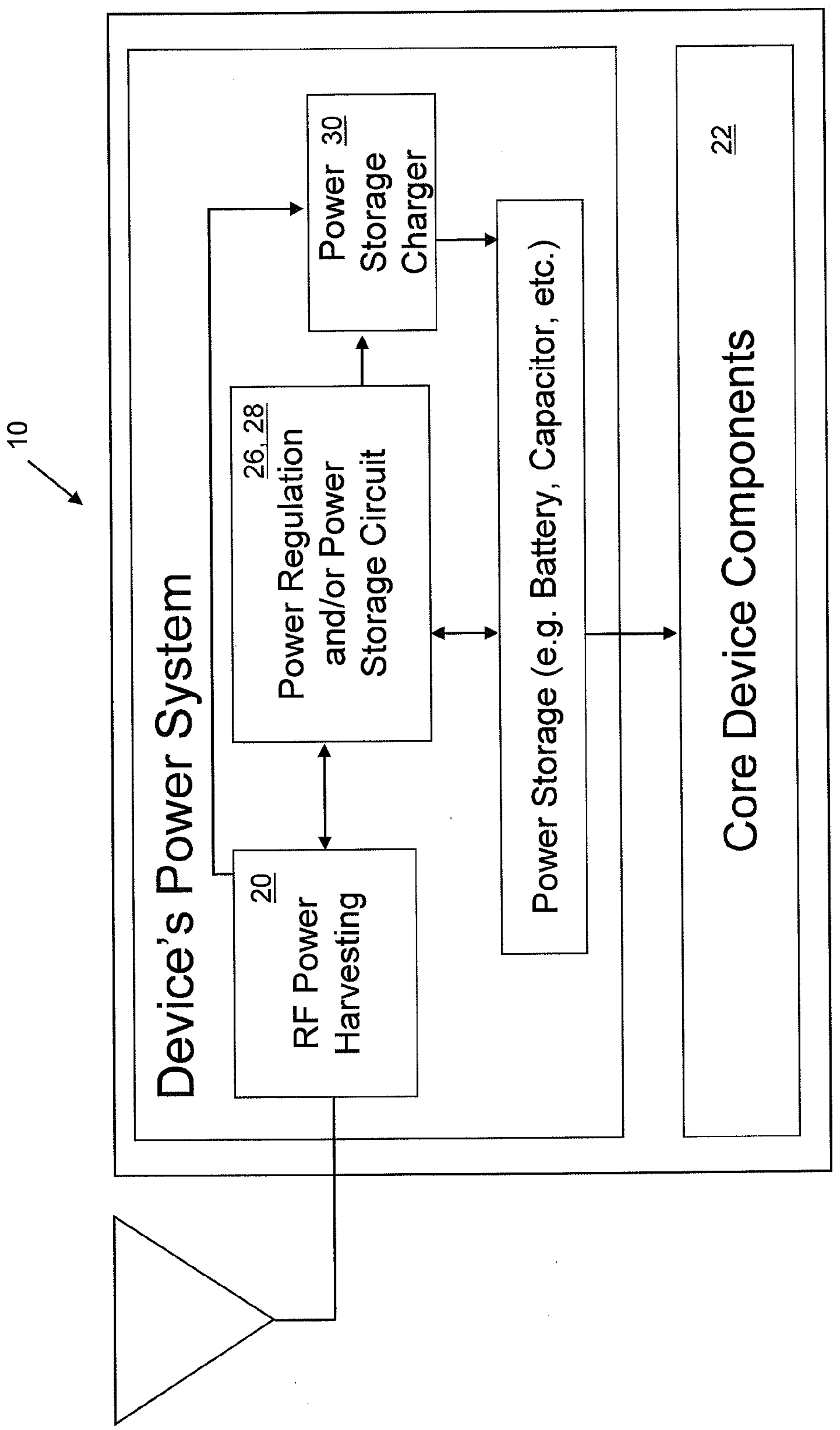
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30 Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 39

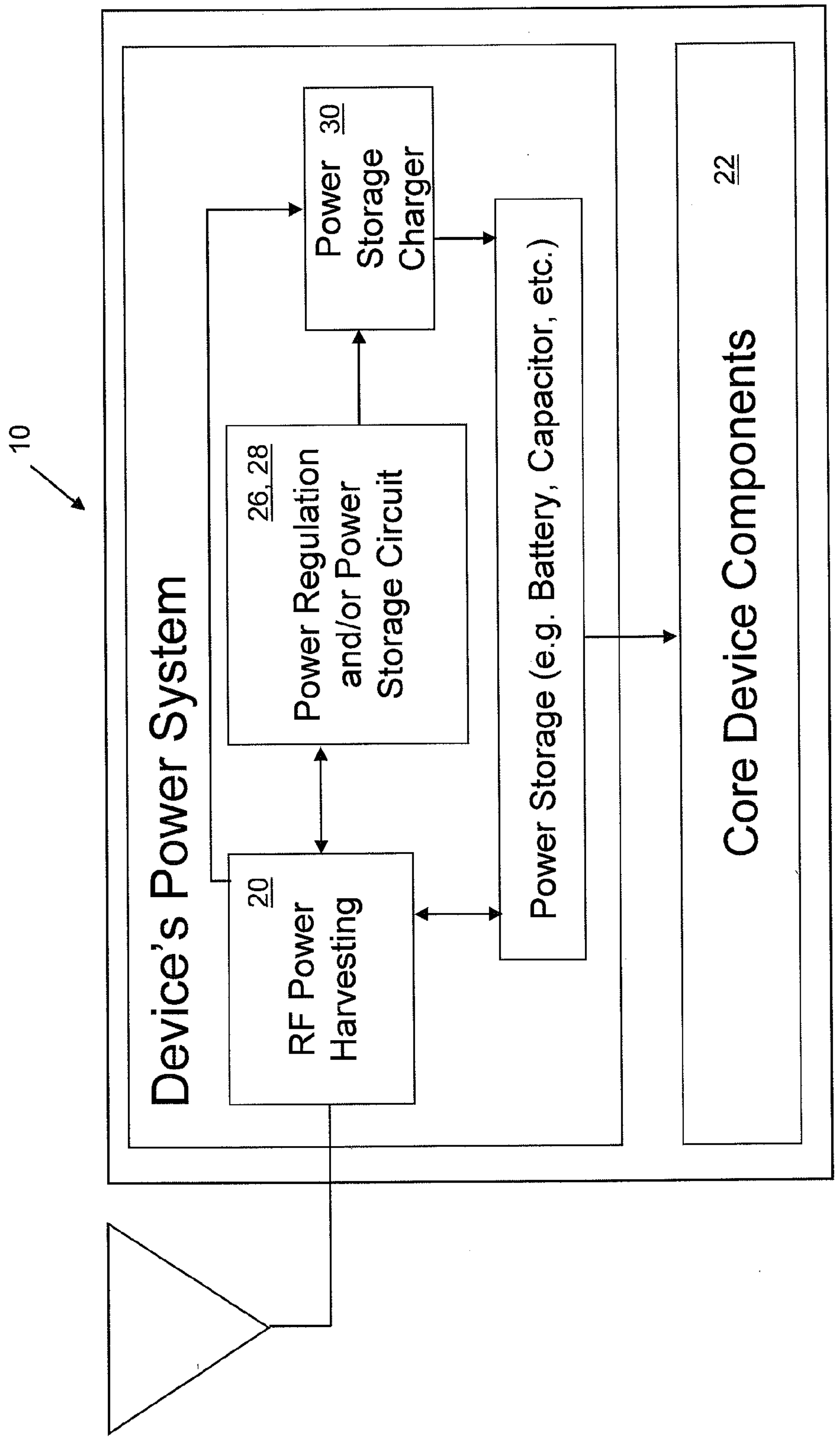
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 40

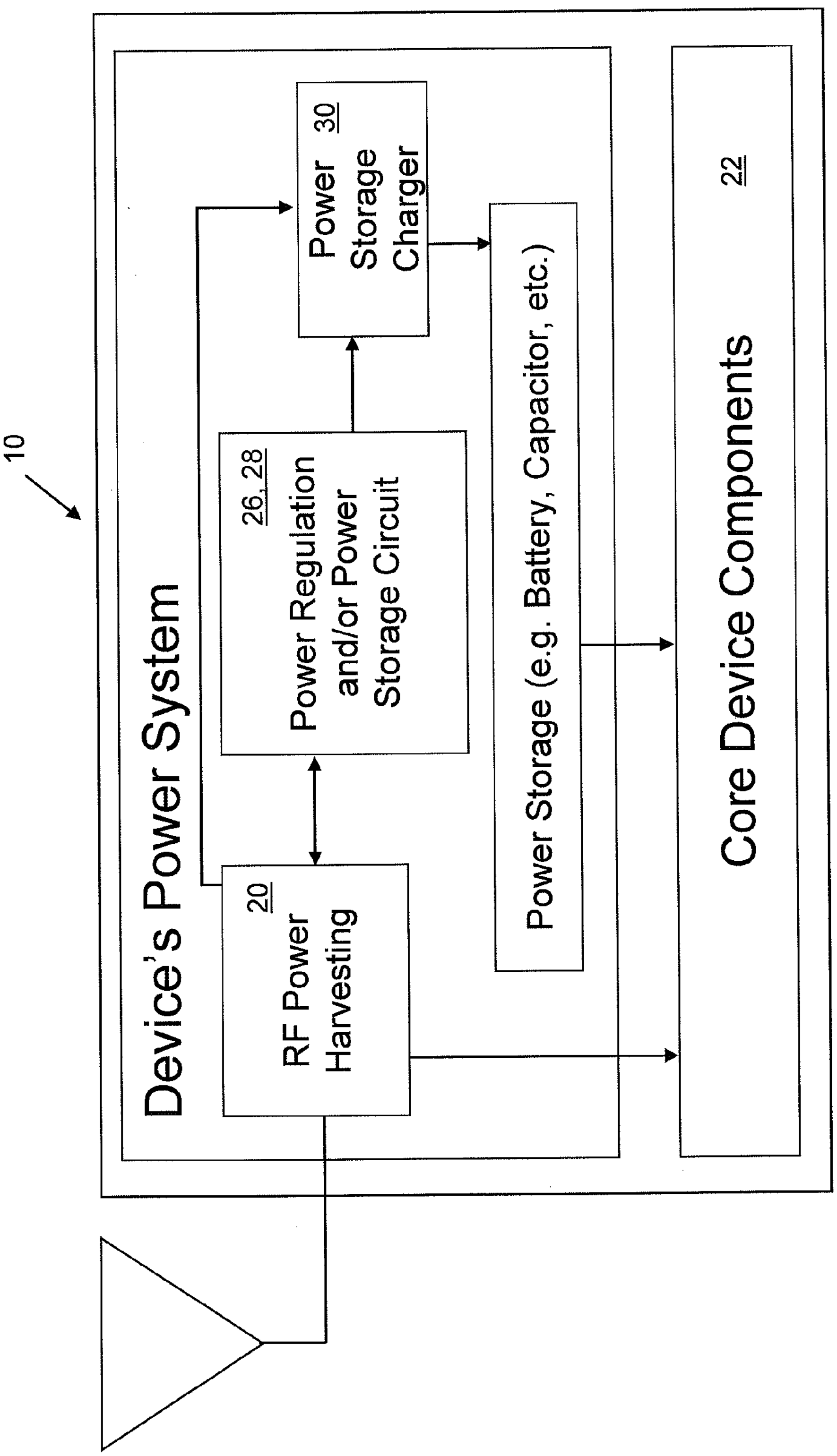
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 41

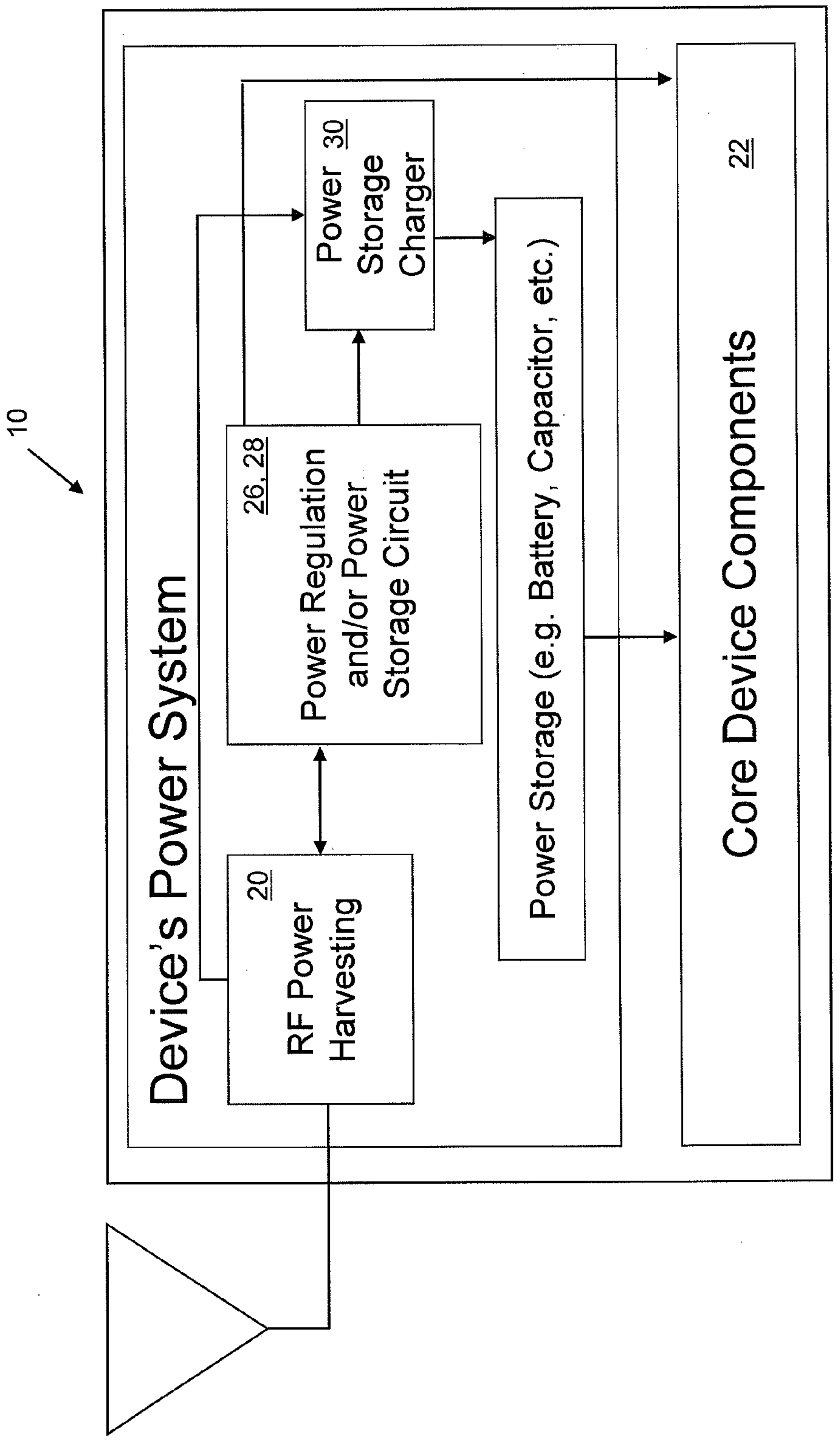
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 42

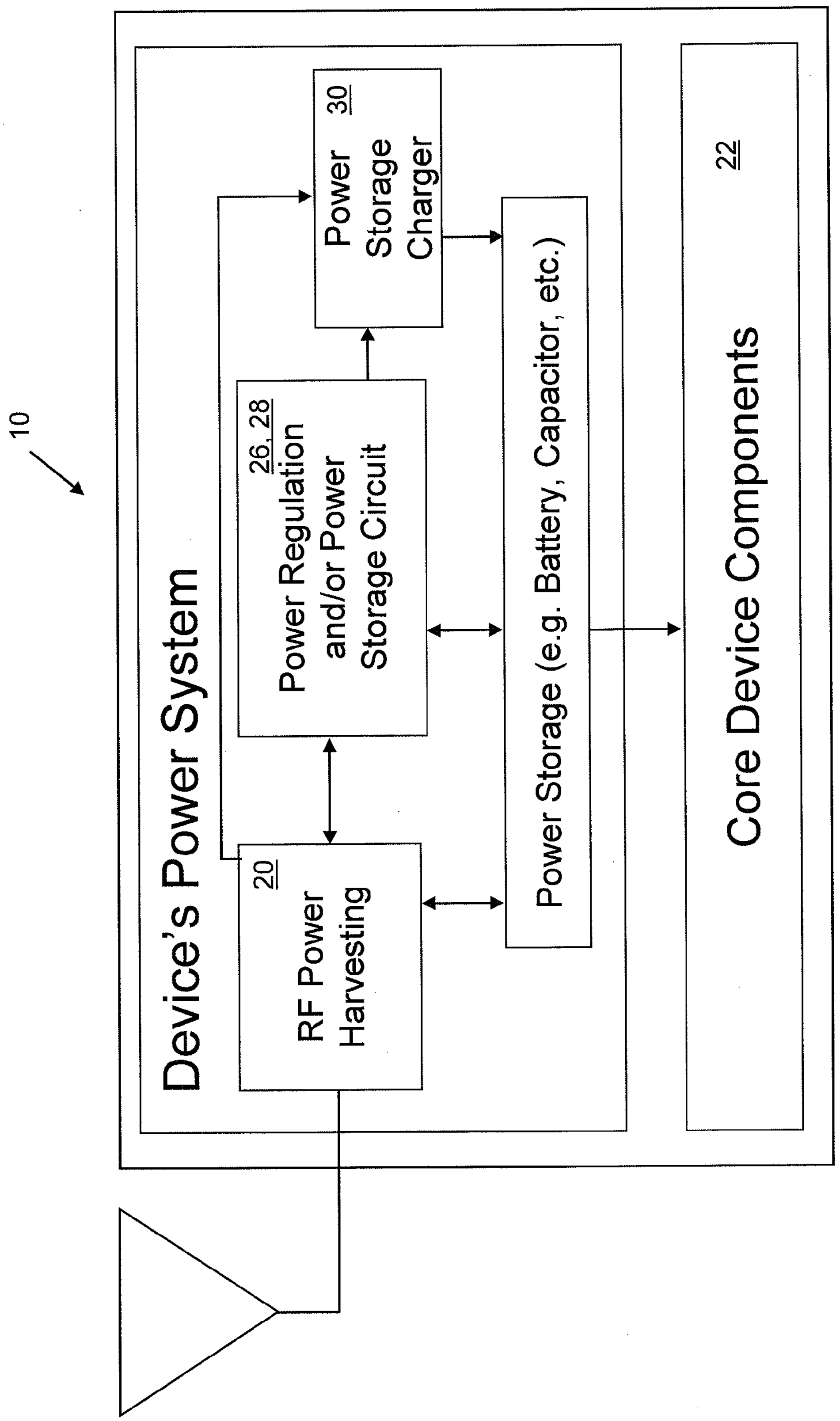
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 43

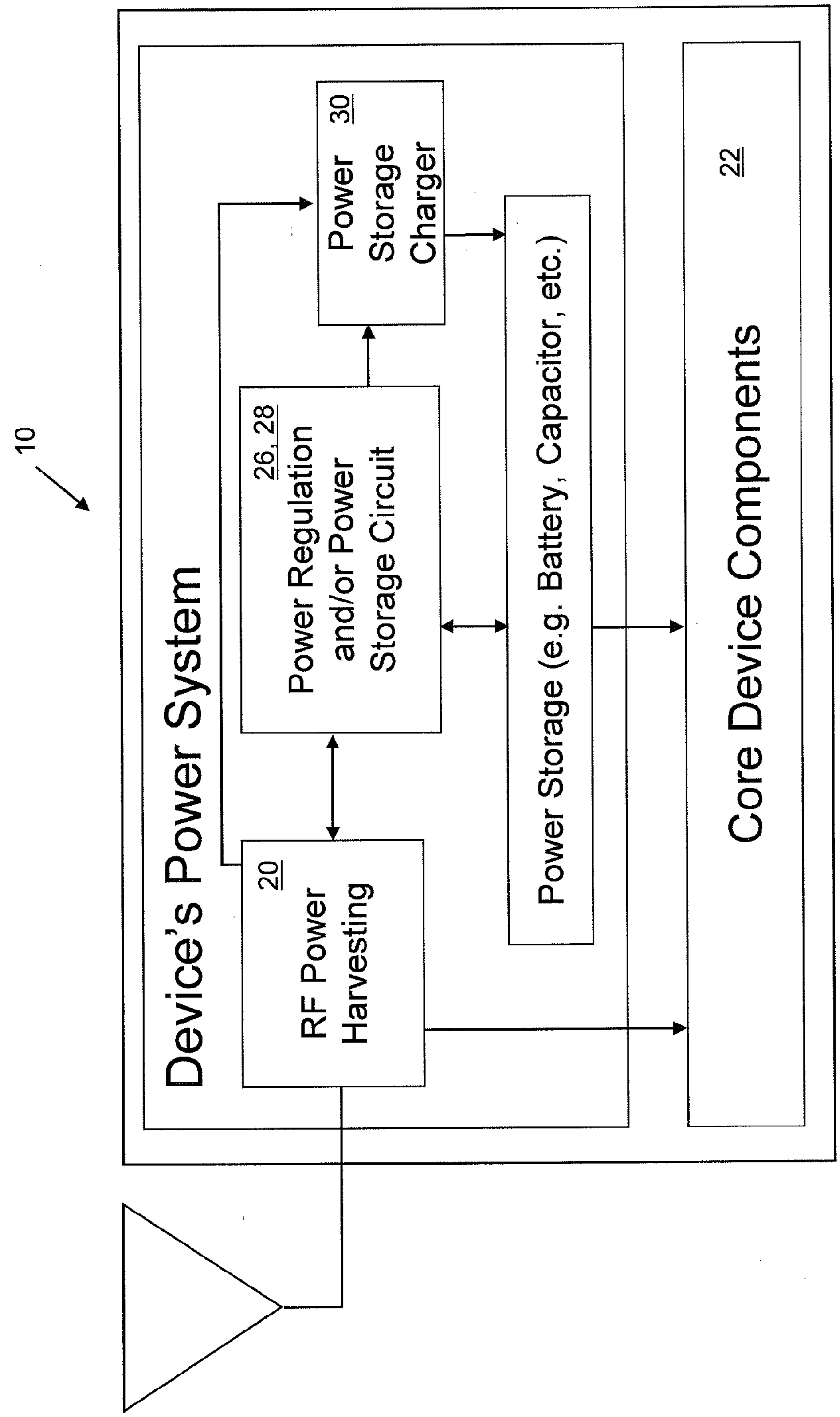
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 44

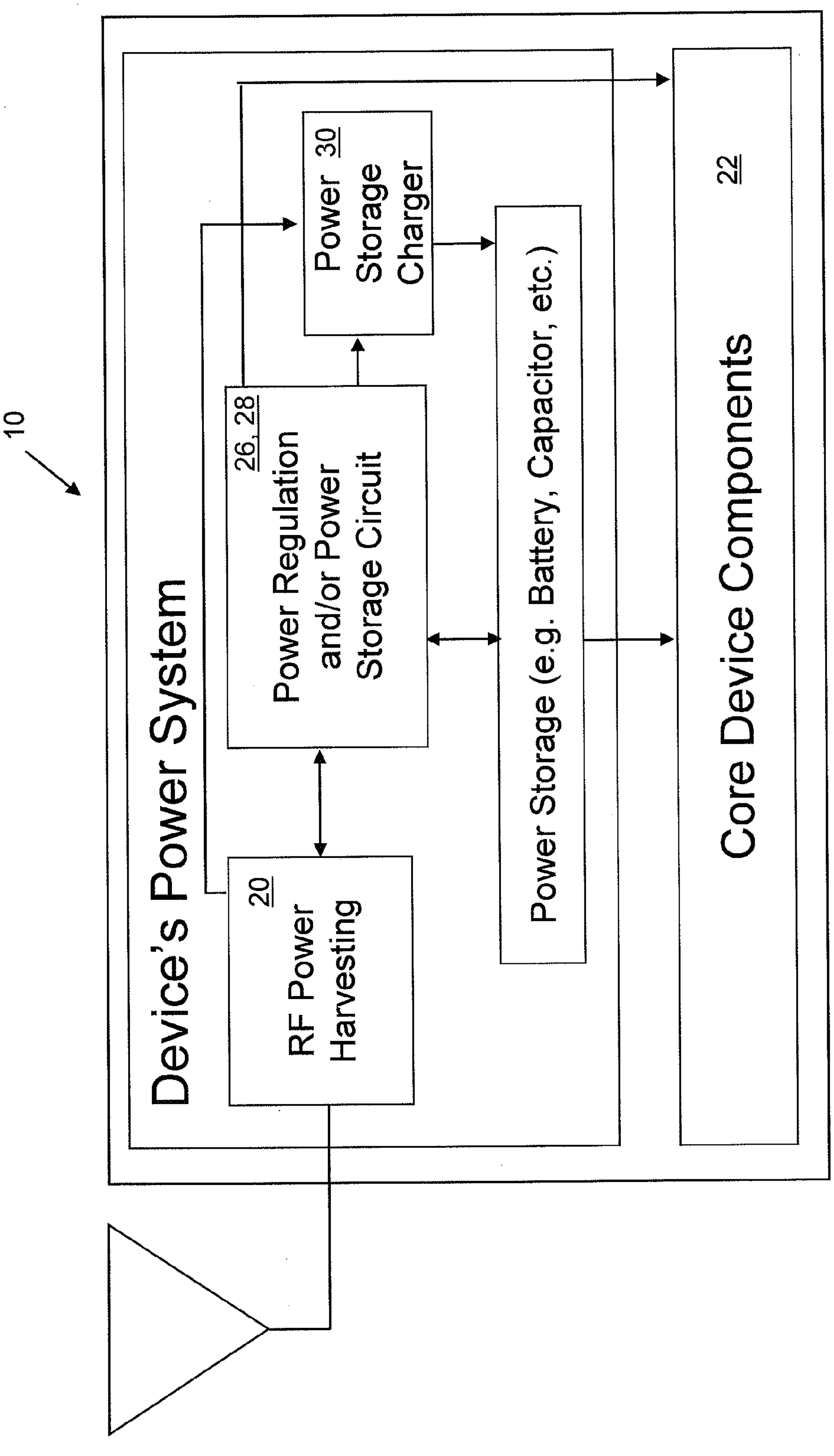
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30 Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 45

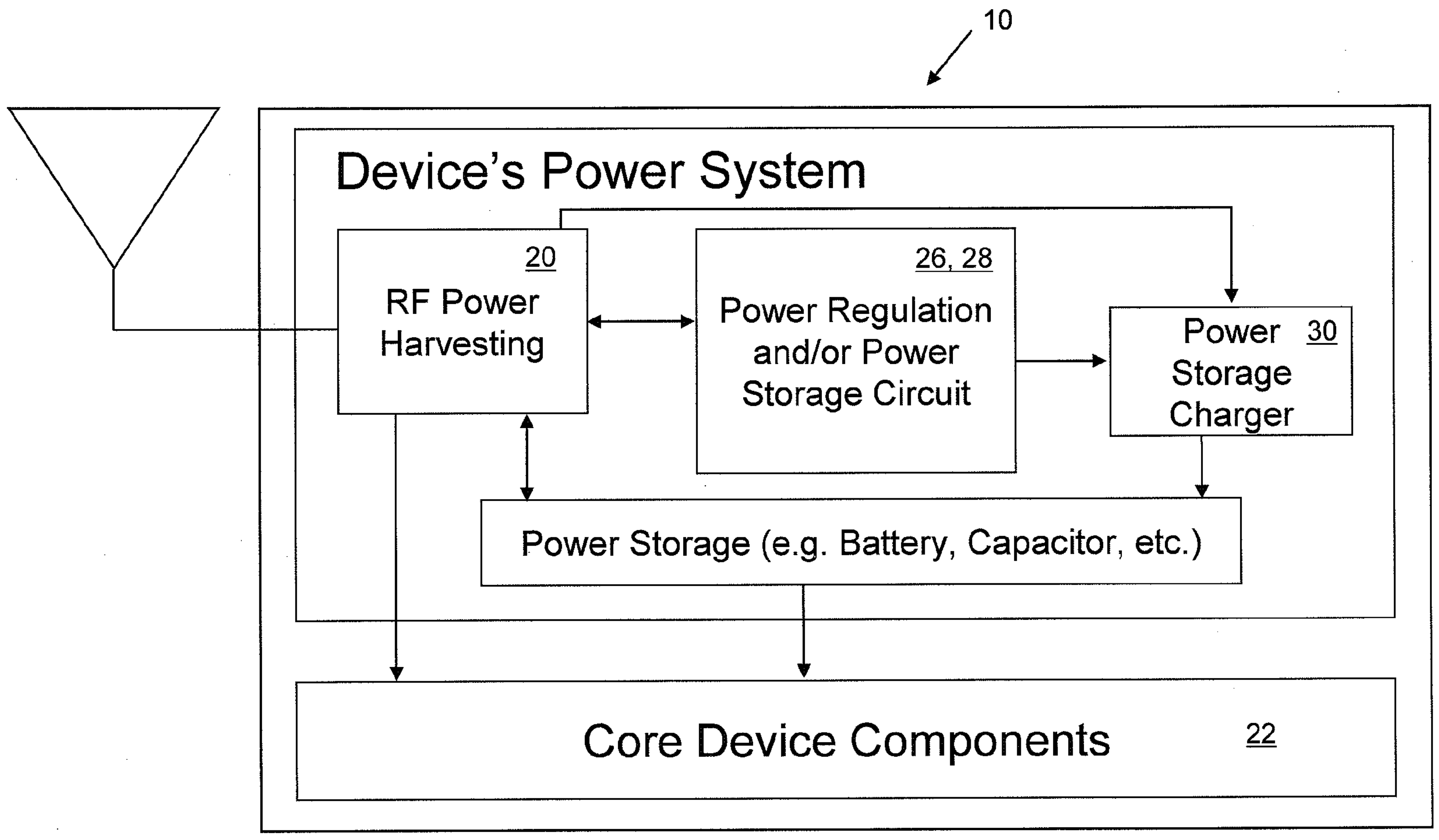
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 46

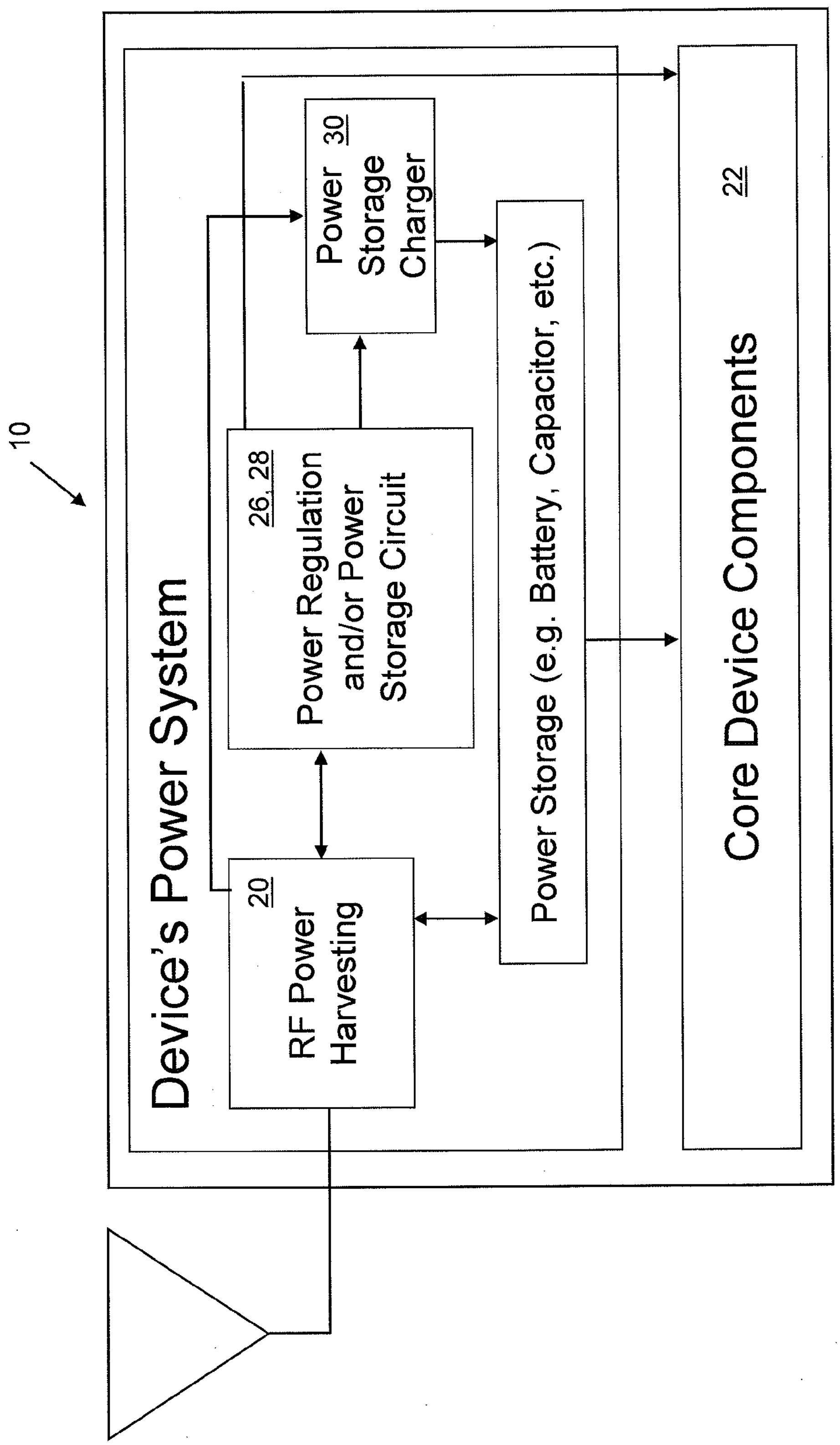
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 47

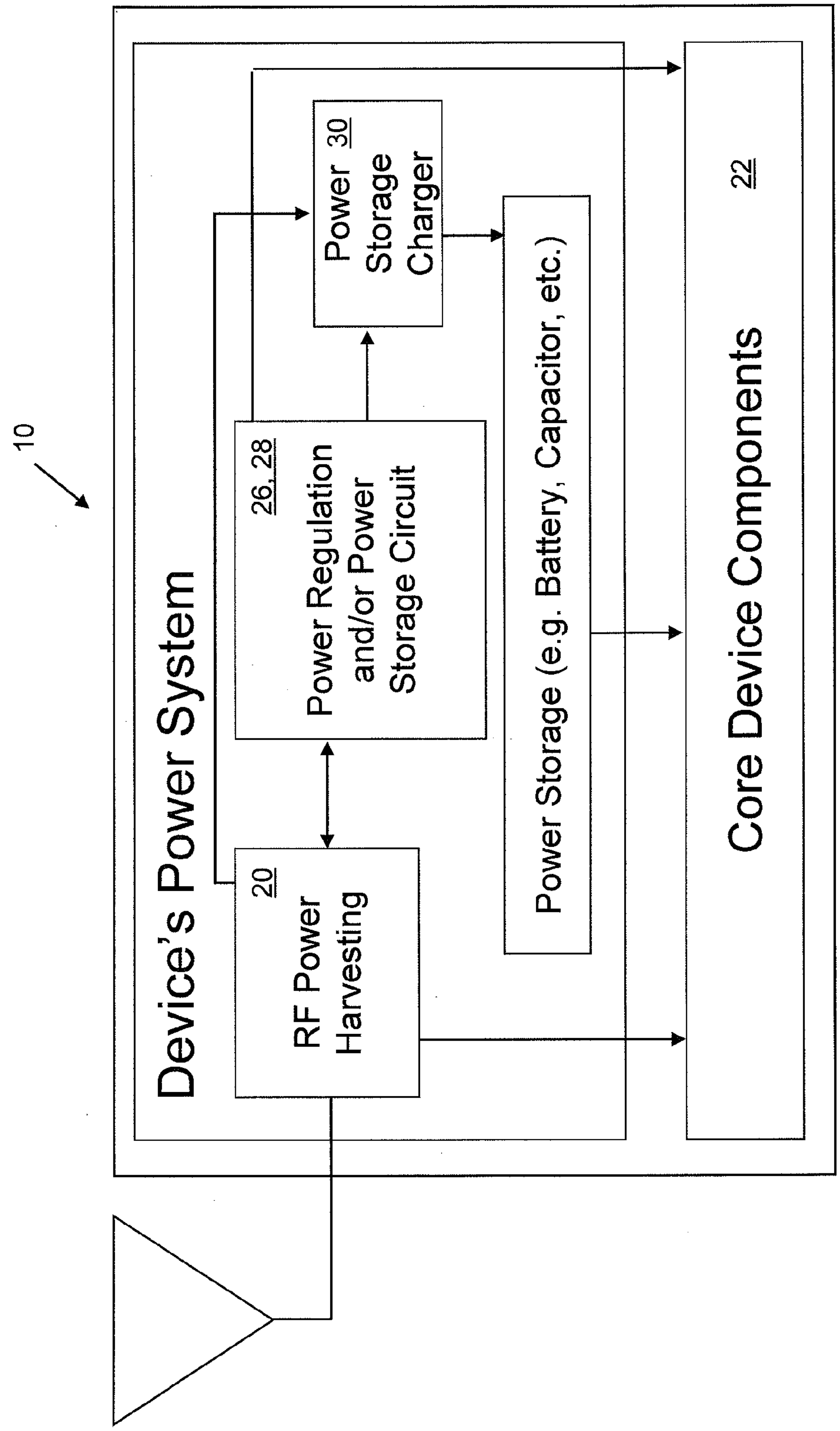
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 48

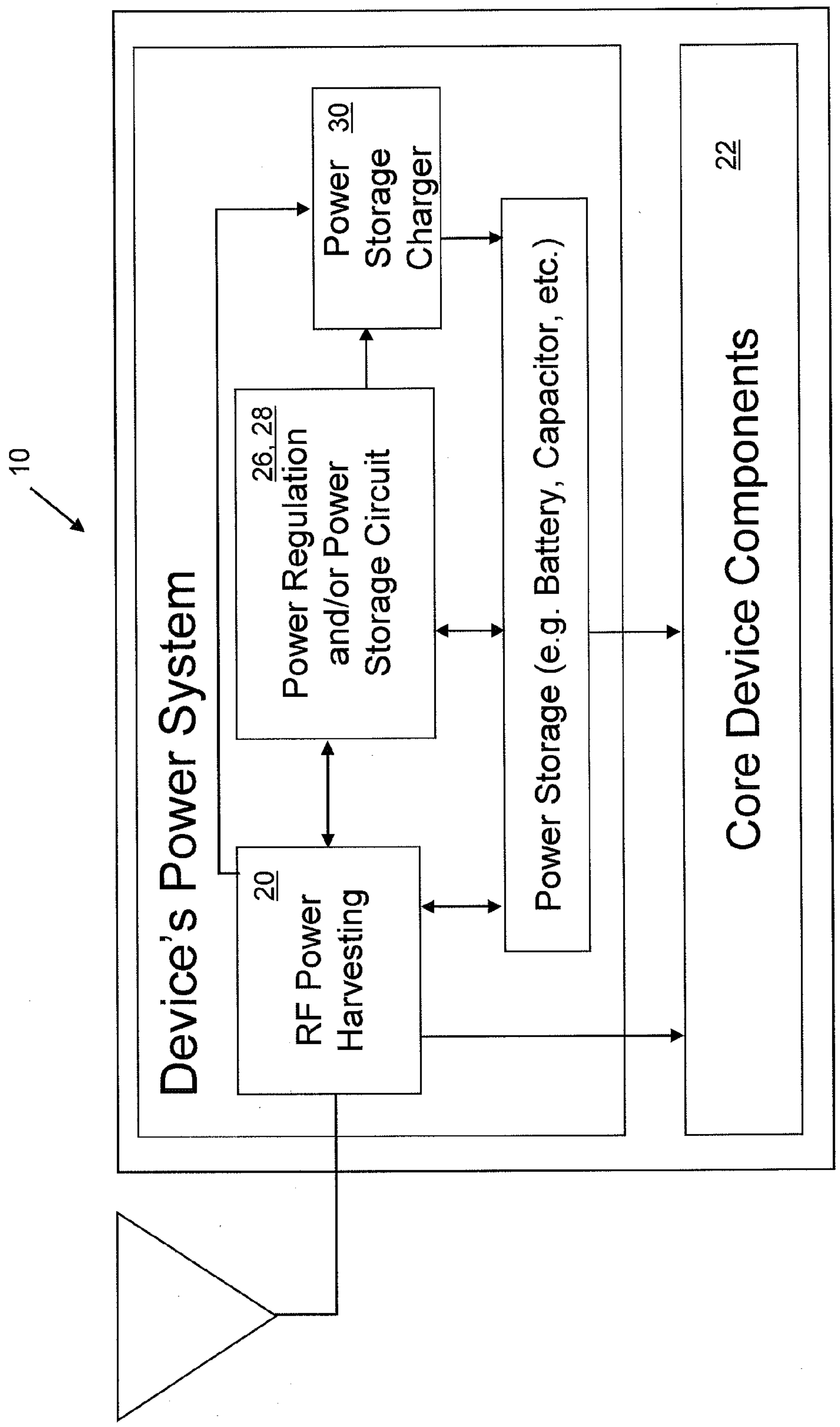
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 49

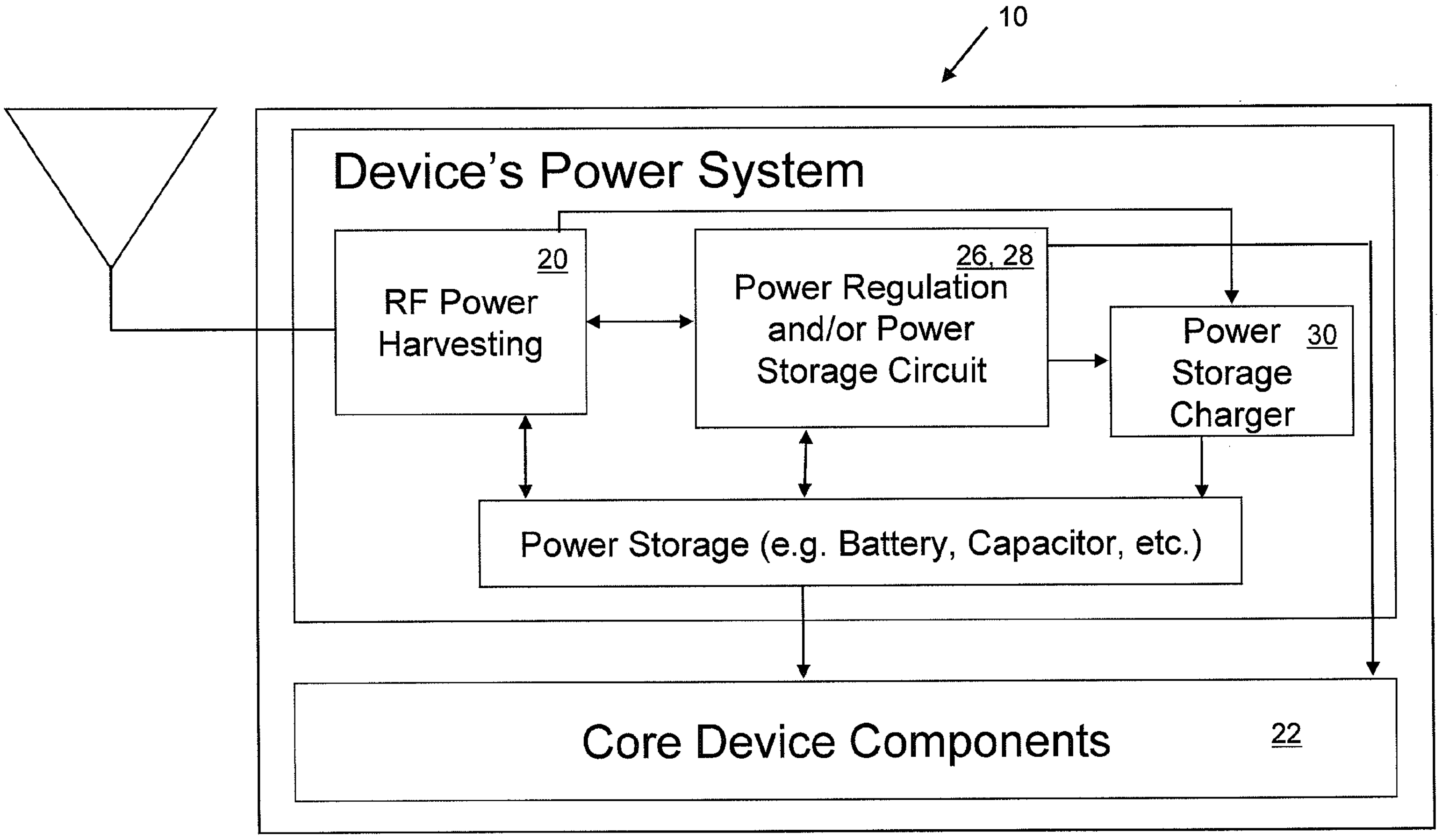
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 50

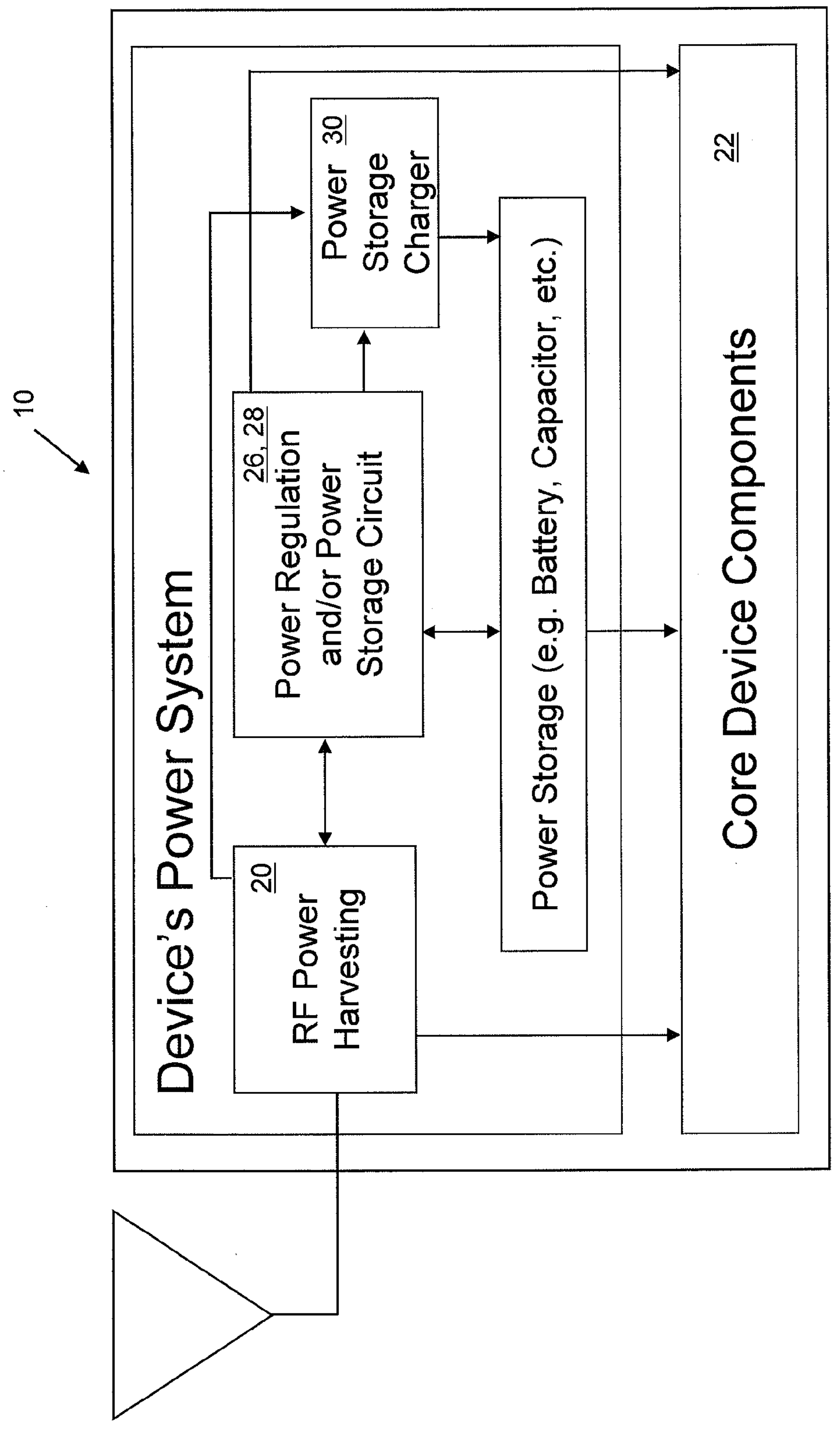
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 51

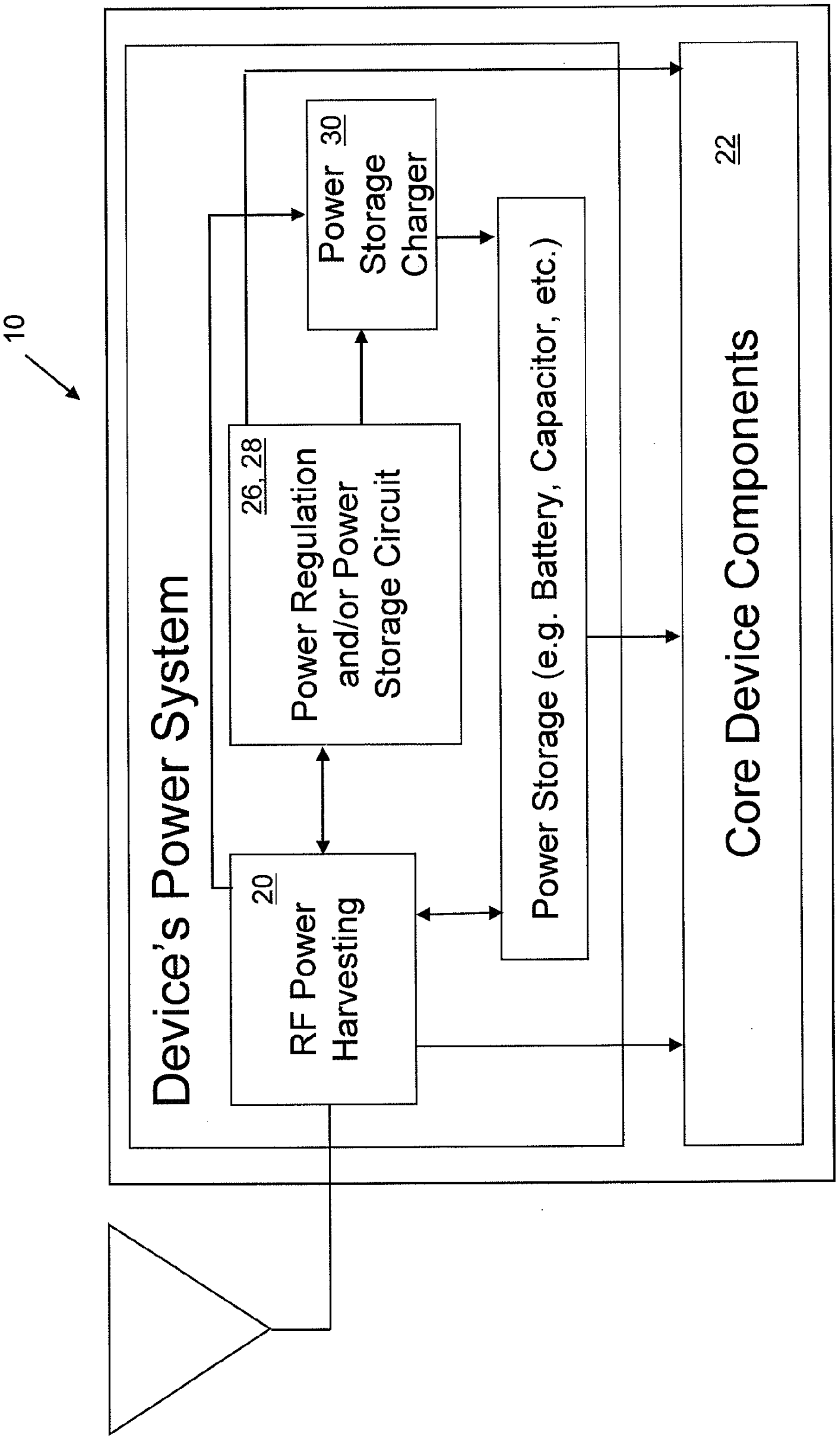
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 52

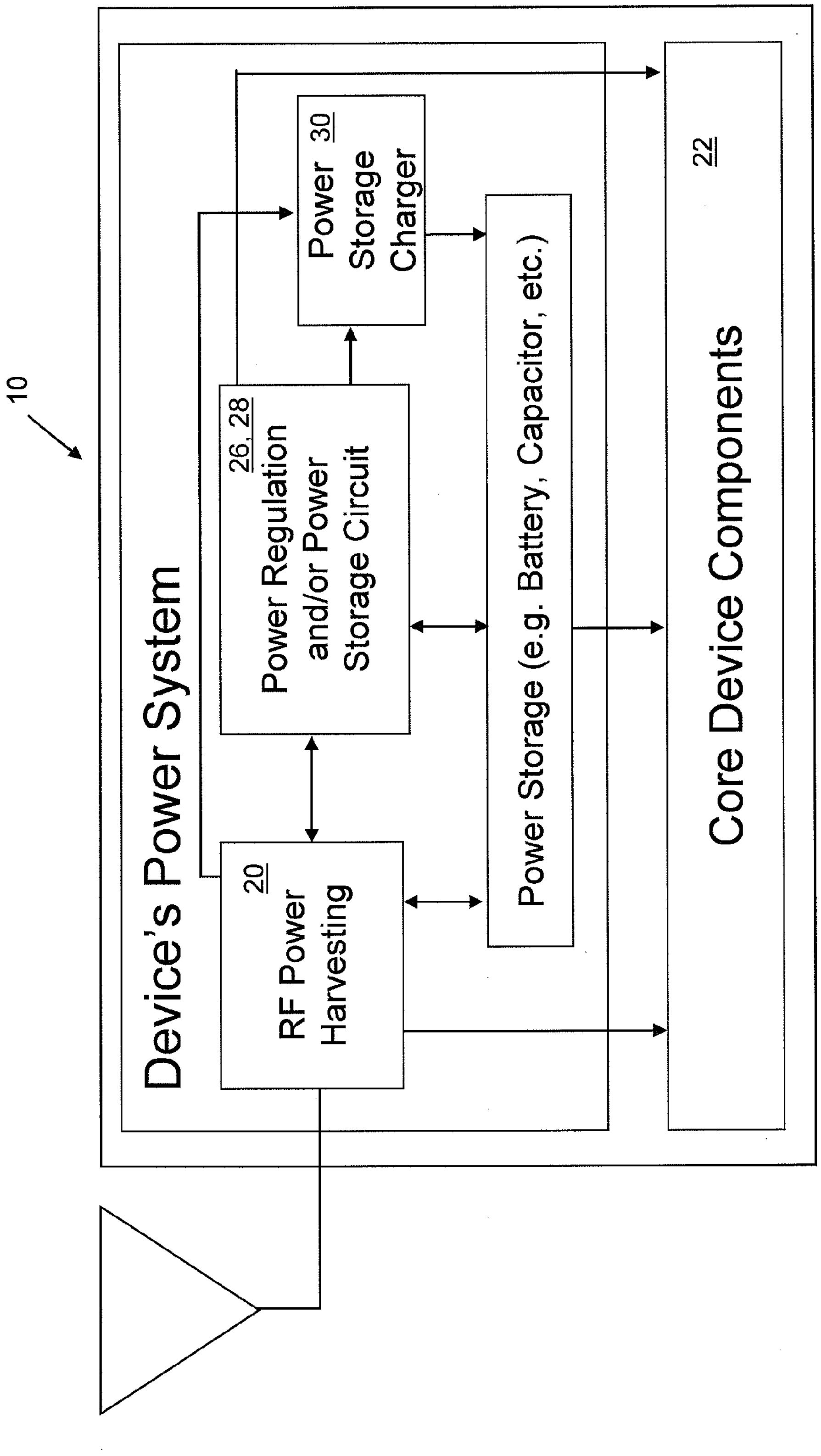
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30 Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 53

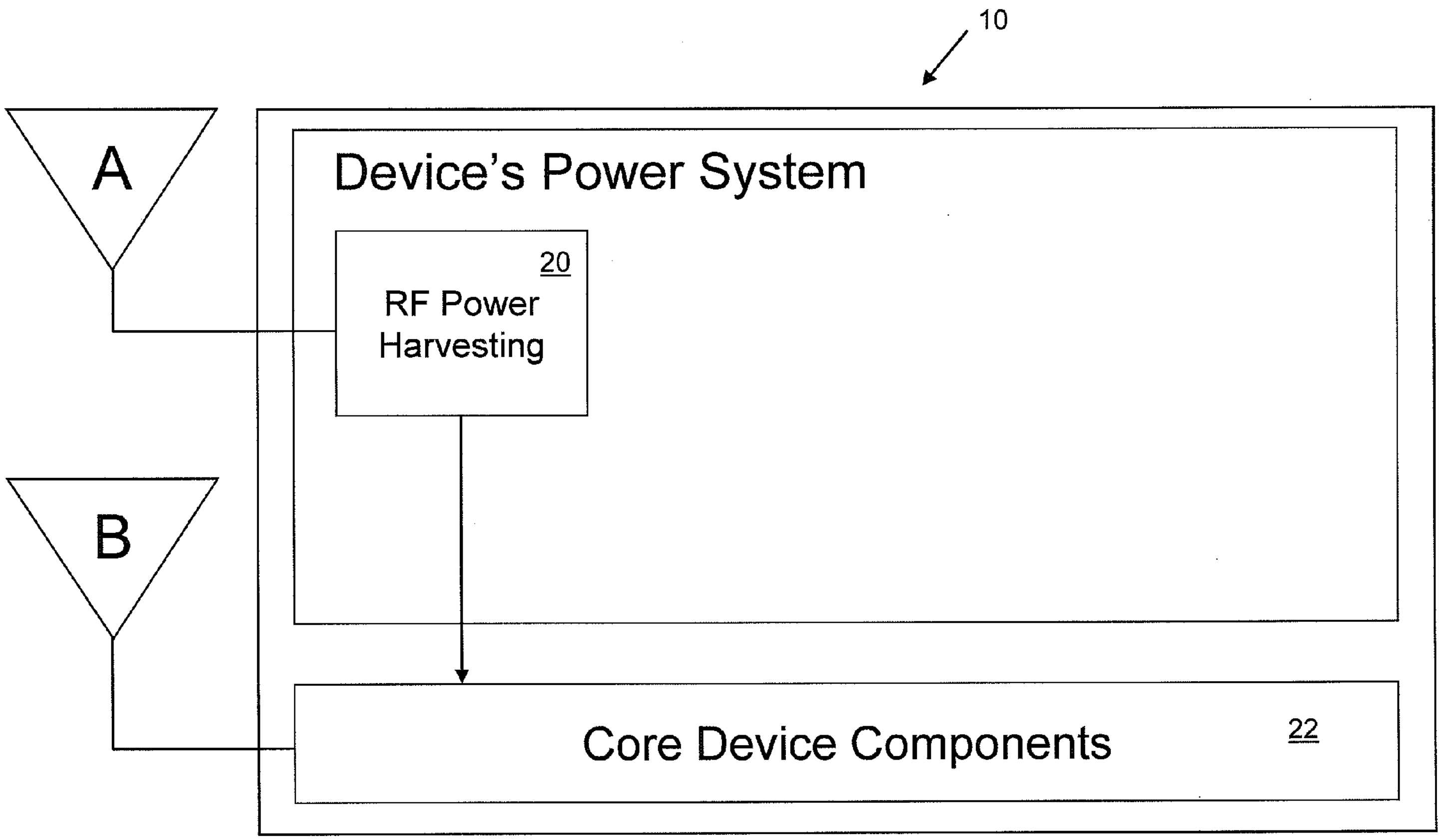
10
A
B
Device's Power System
20
RF Power Harvesting
Core Device Components
22

Fig. 54

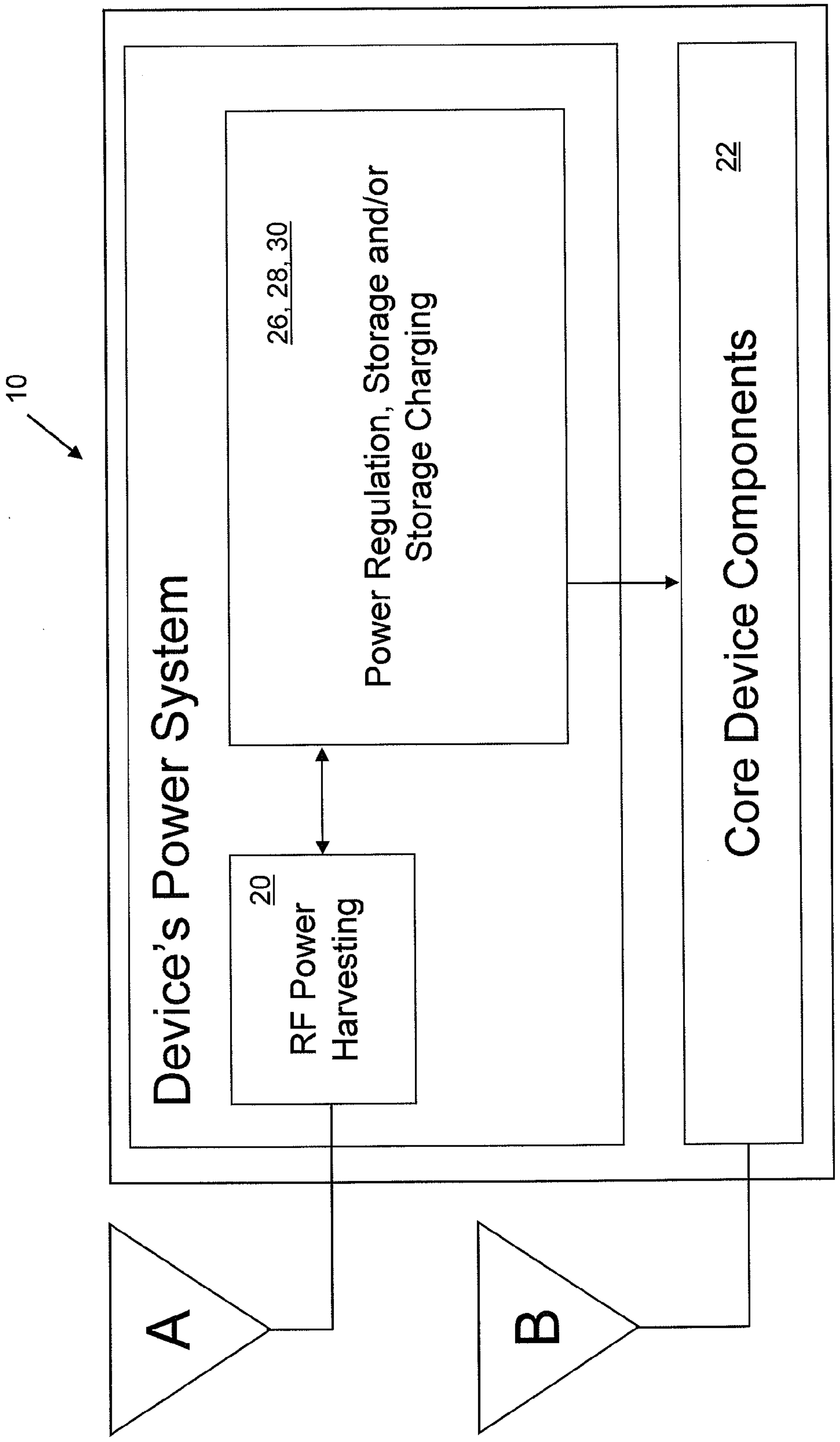
10
Device's Power System
RF Power Harvesting
20
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
A
B

Fig. 55

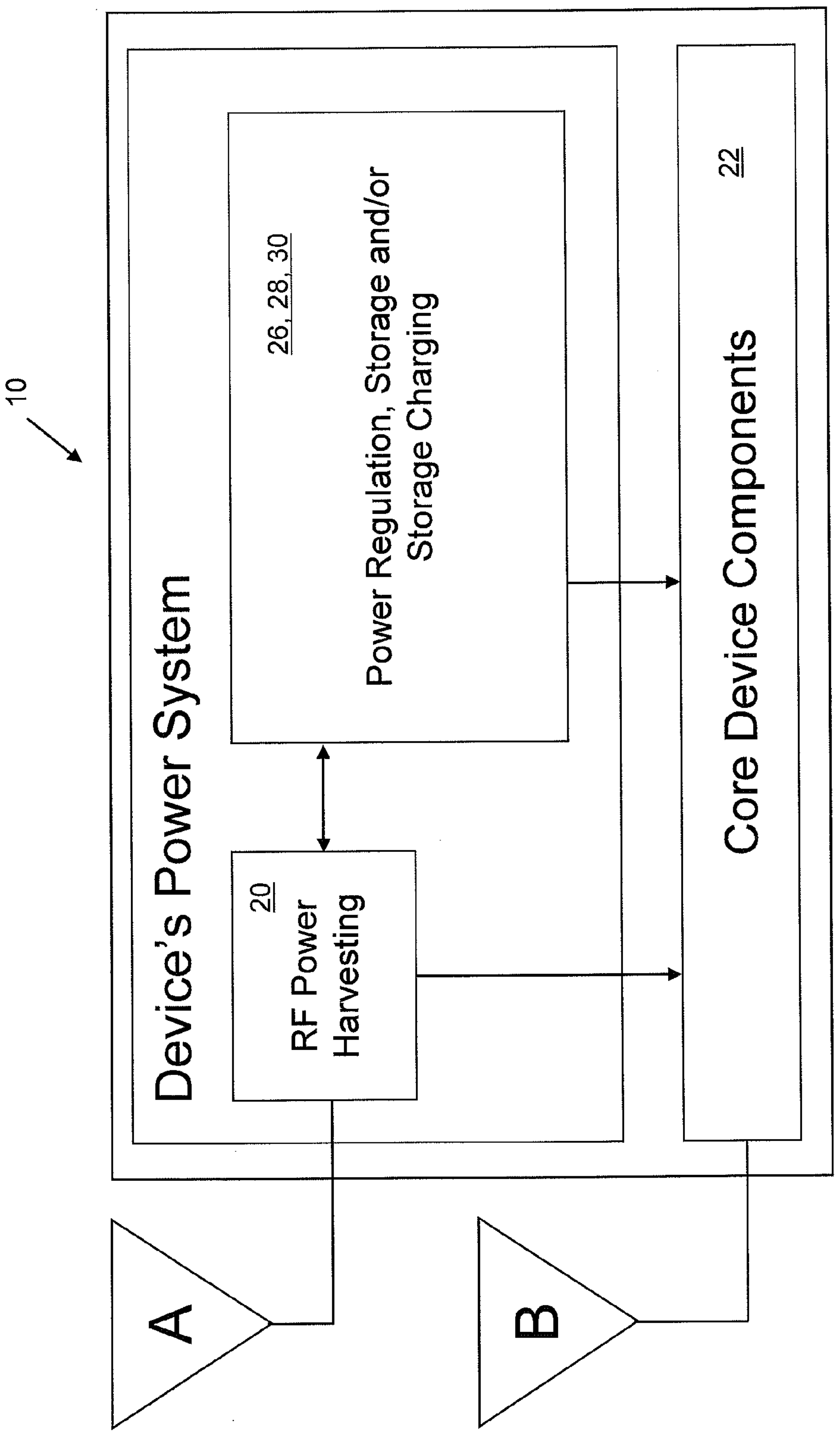
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
A
B

Fig. 56

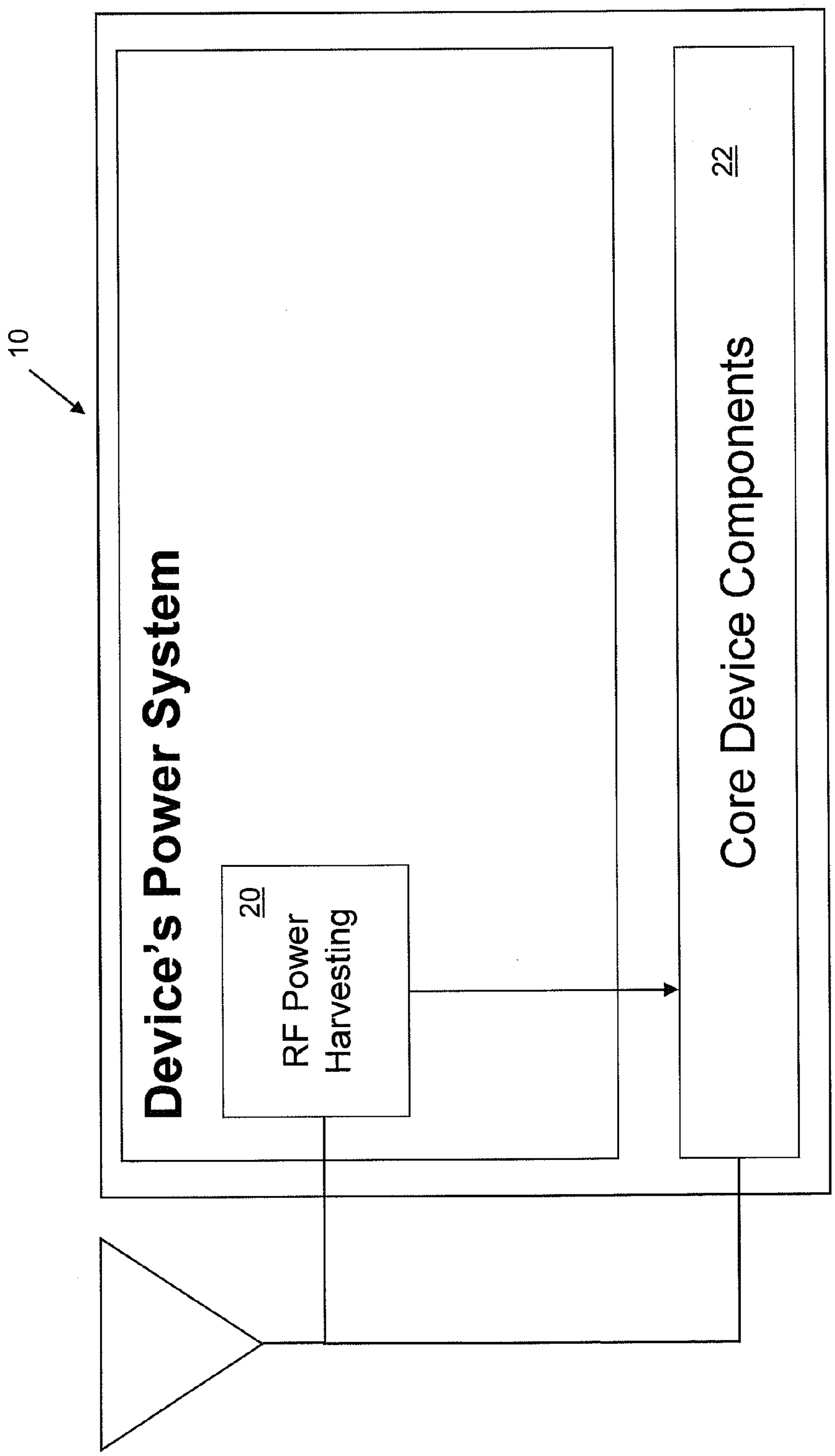
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22

Fig. 57

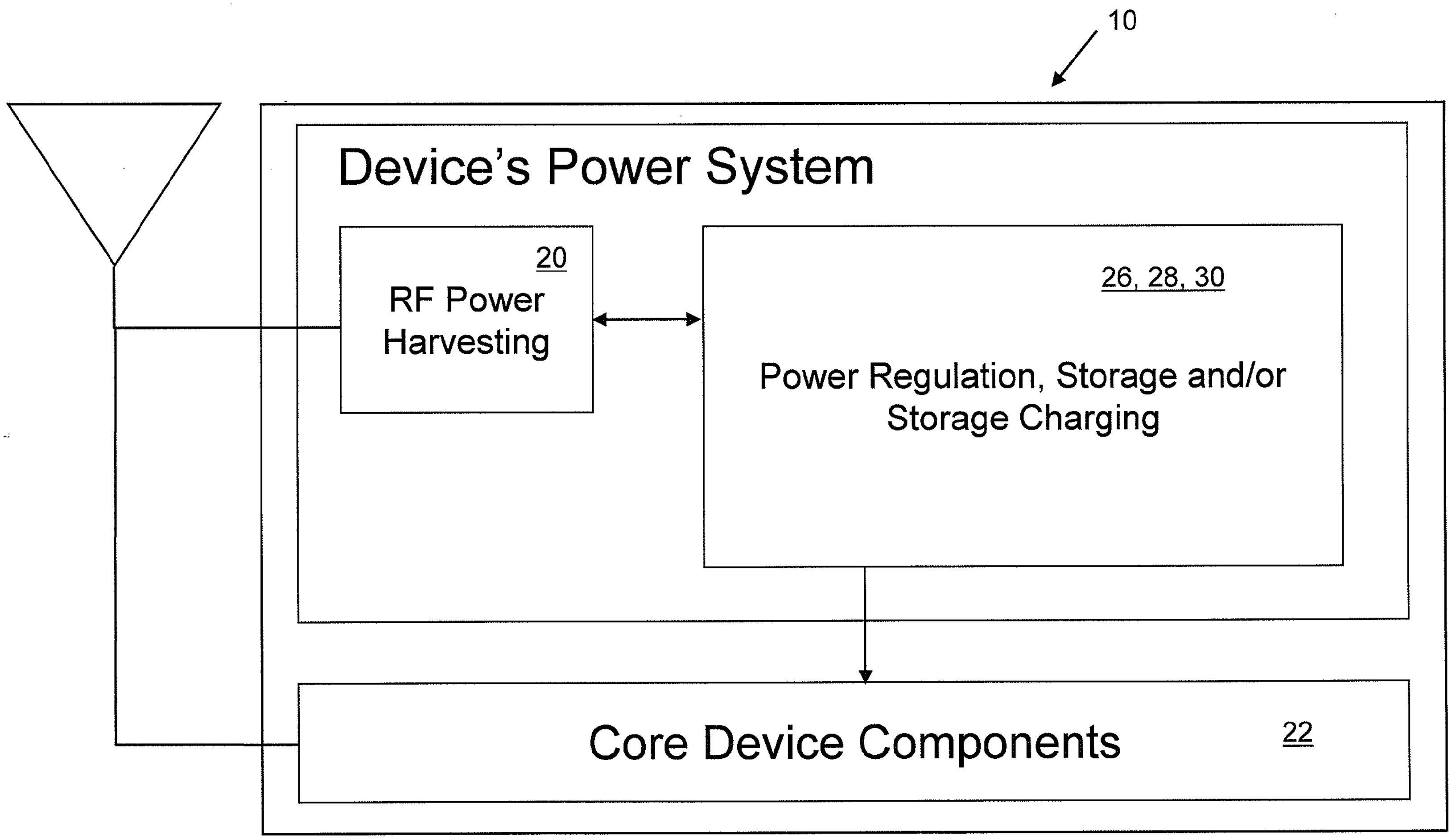
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22

Fig. 58

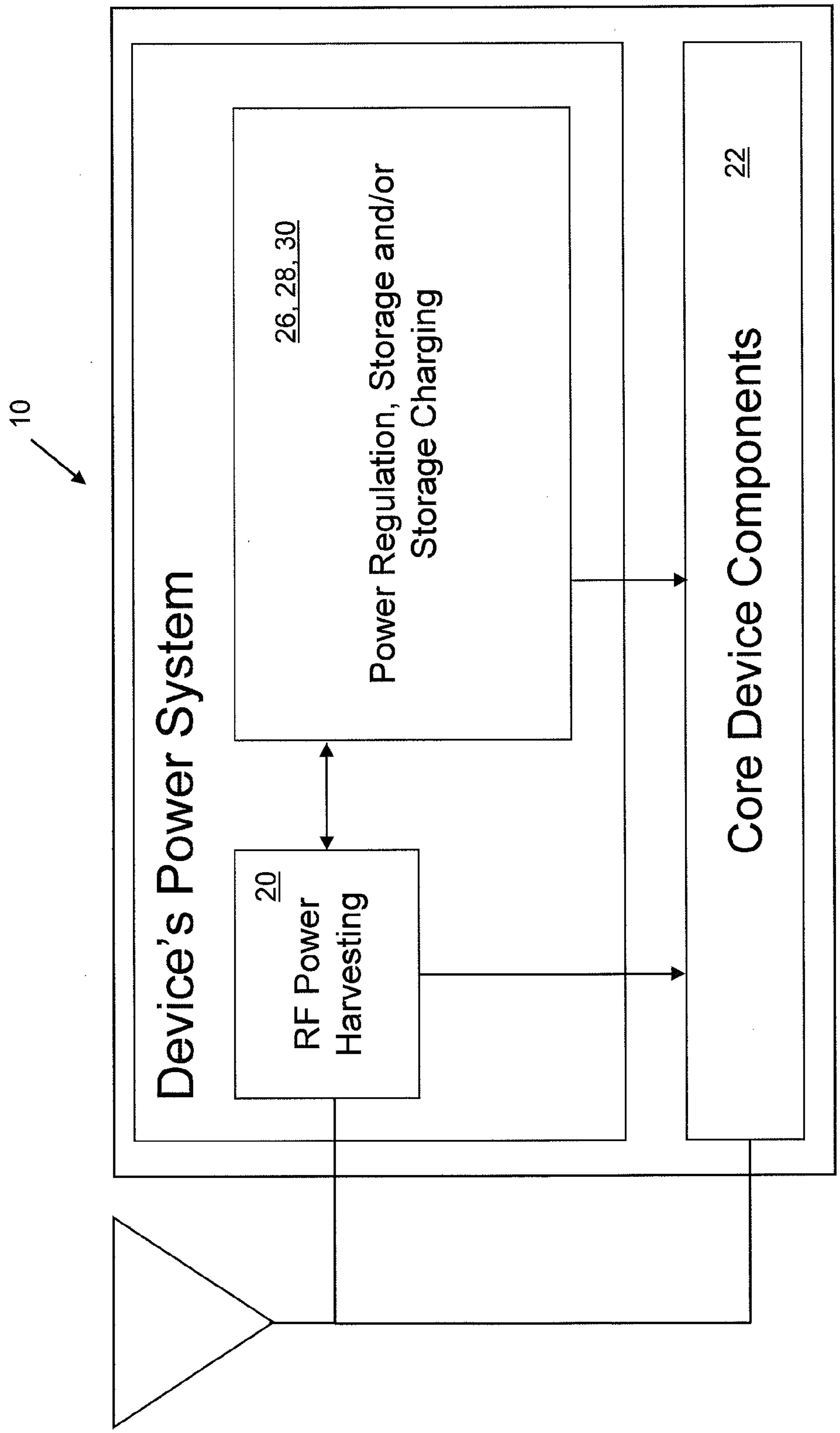
10
Device's Power System
RF Power Harvesting
20
Power Regulation, Storage and/or Storage Charging
26, 28, 30
Core Device Components
22

Fig. 59

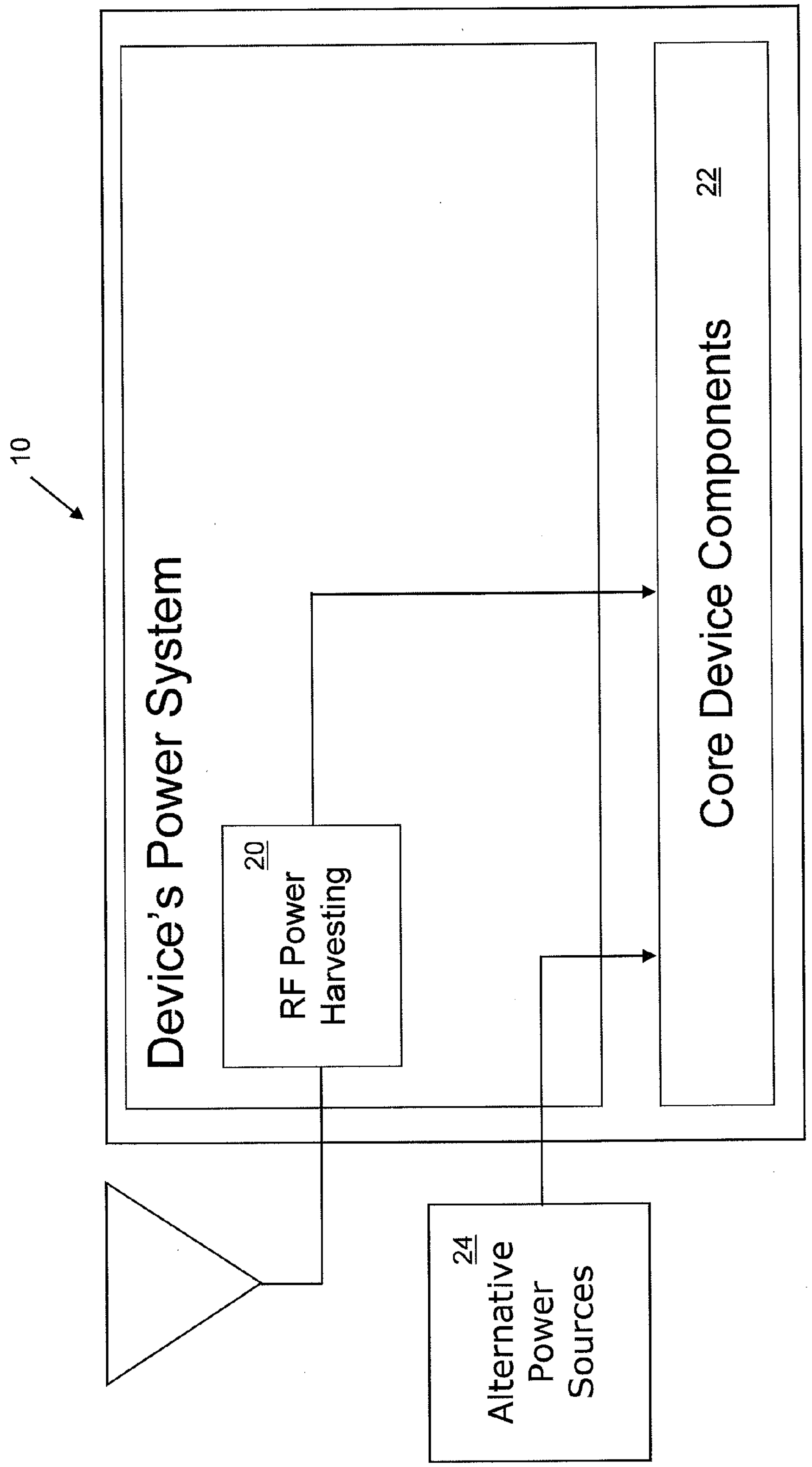
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 60

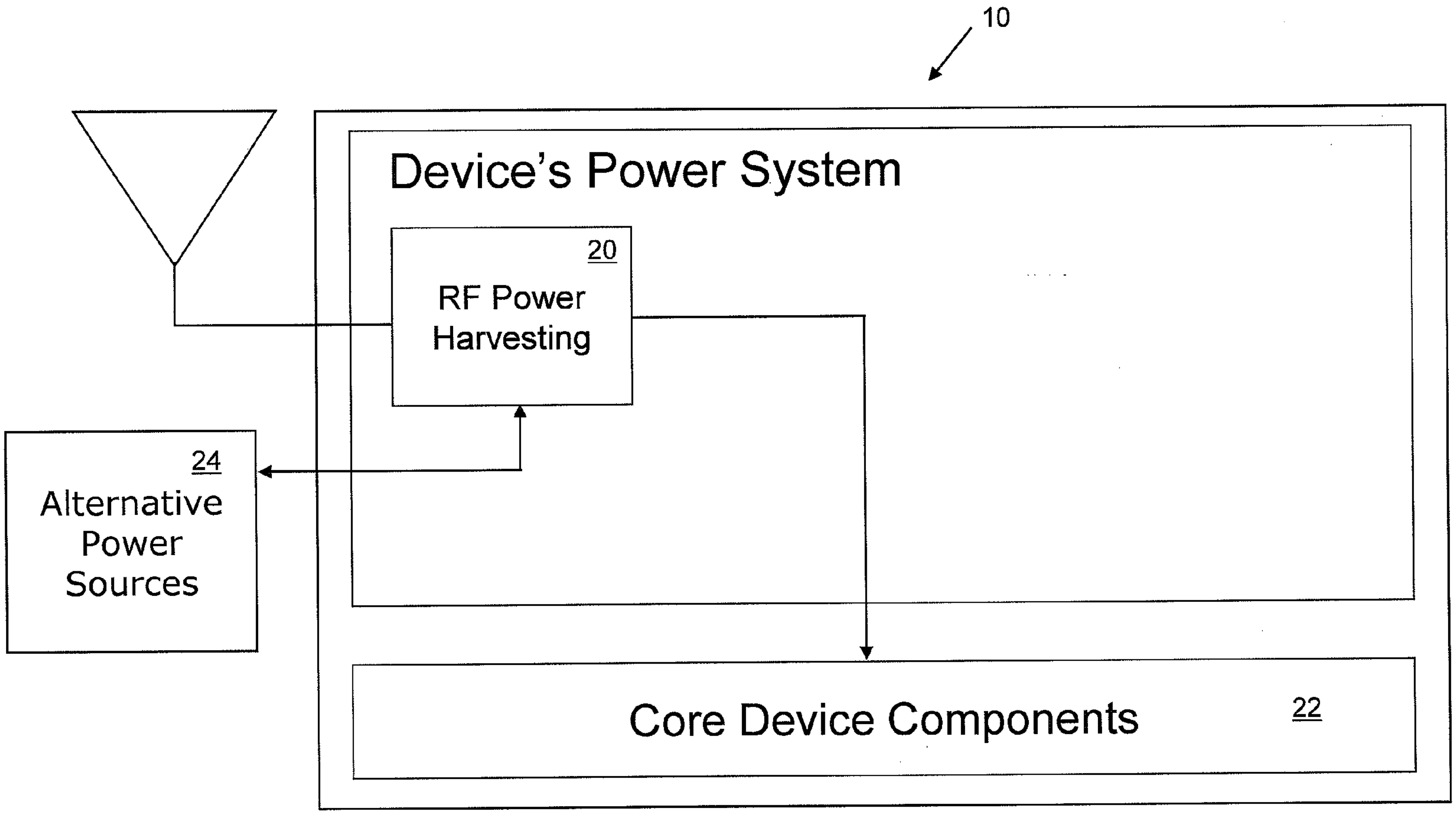
10
Device's Power System
20
RF Power Harvesting
24
Alternative Power Sources
Core Device Components
22

Fig. 61

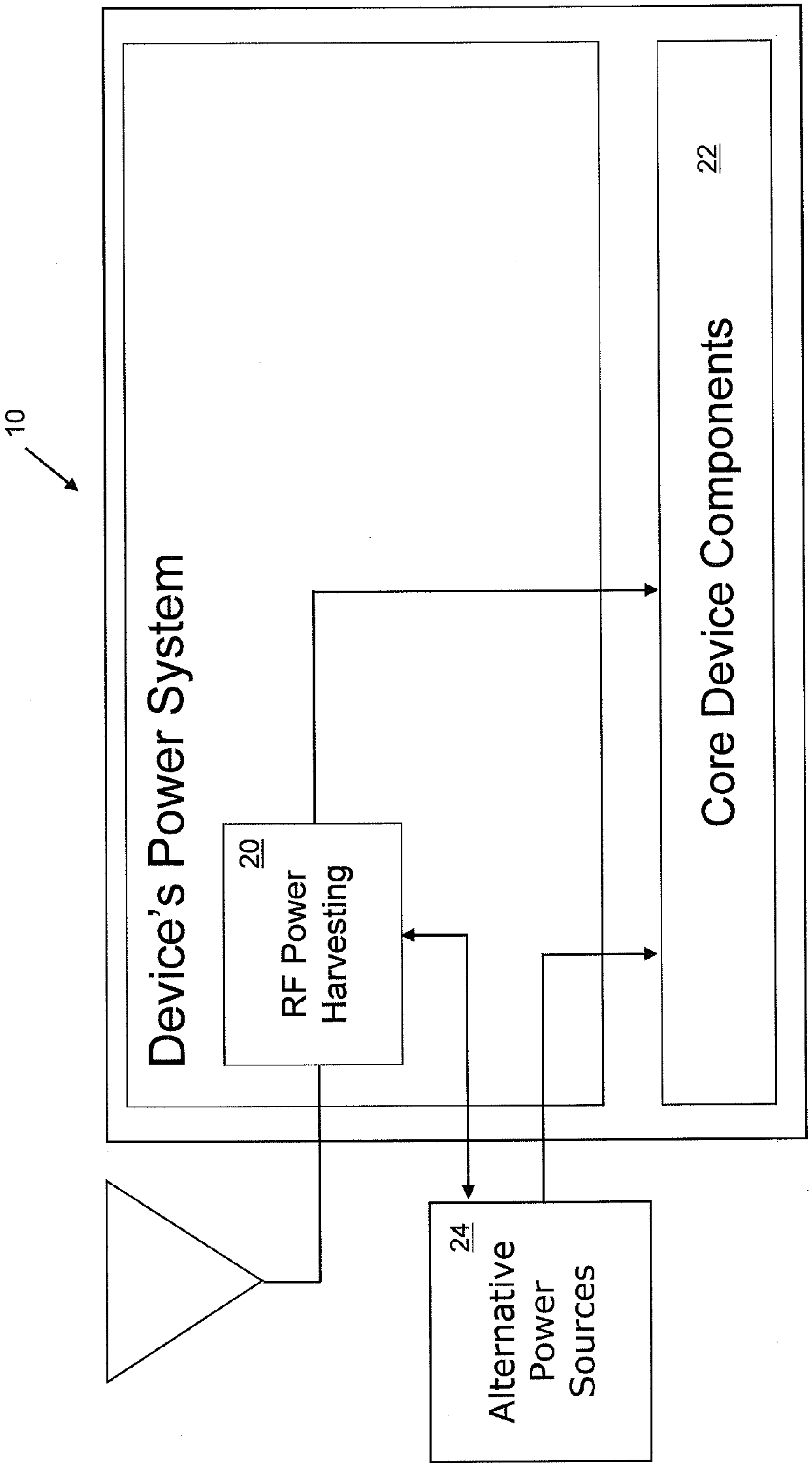
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 62

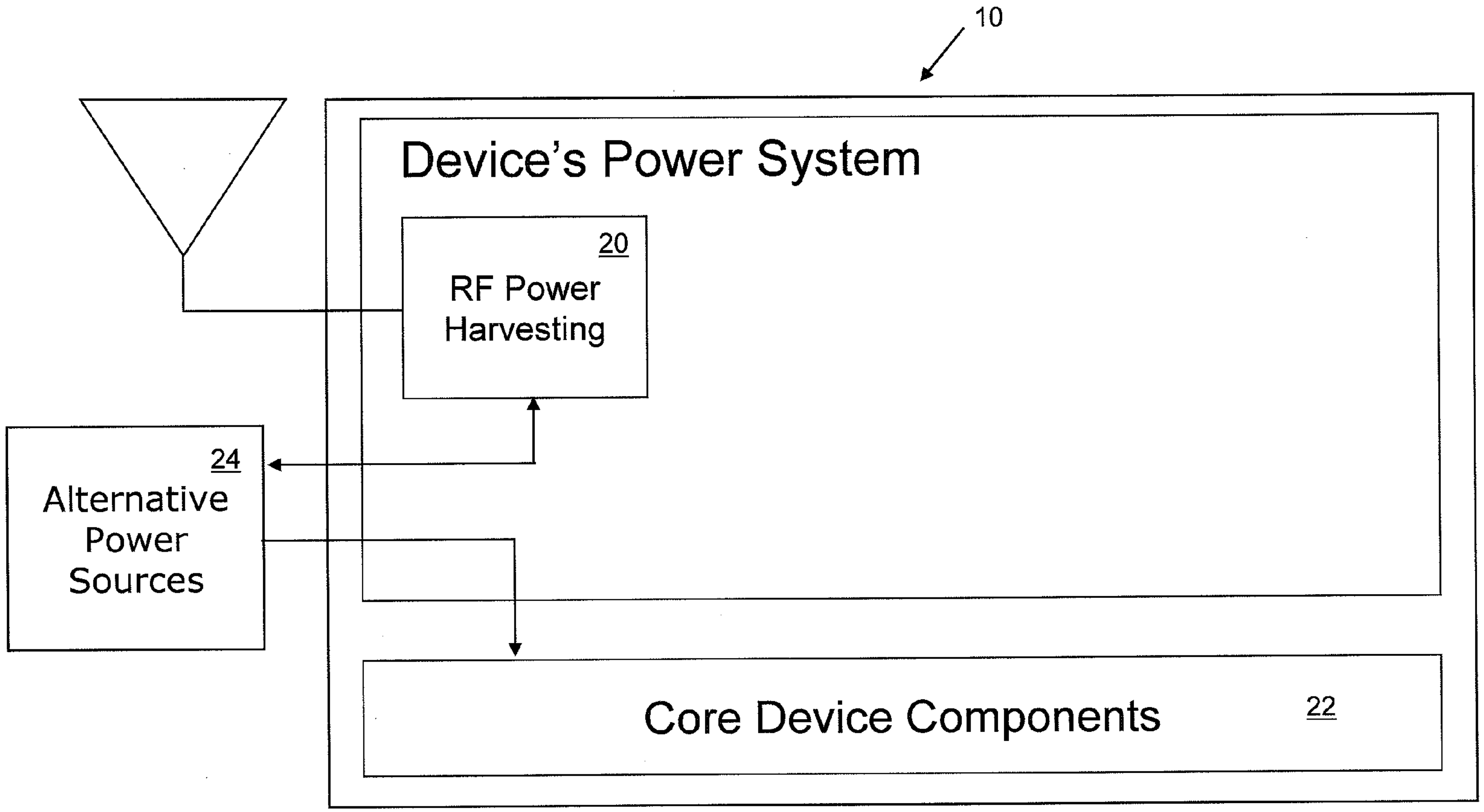
10
Device's Power System
20
RF Power Harvesting
24
Alternative Power Sources
Core Device Components
22

Fig. 63

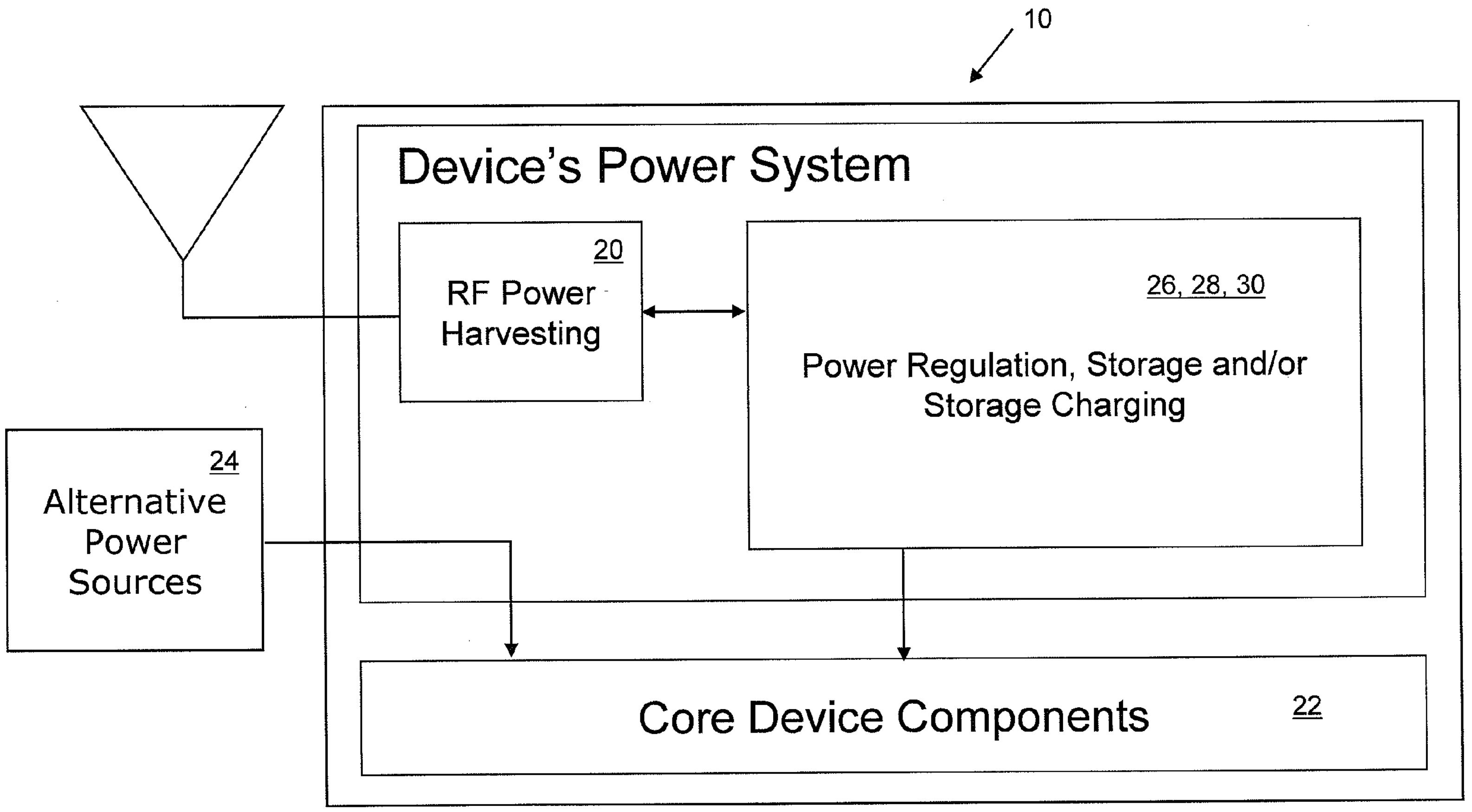
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 64

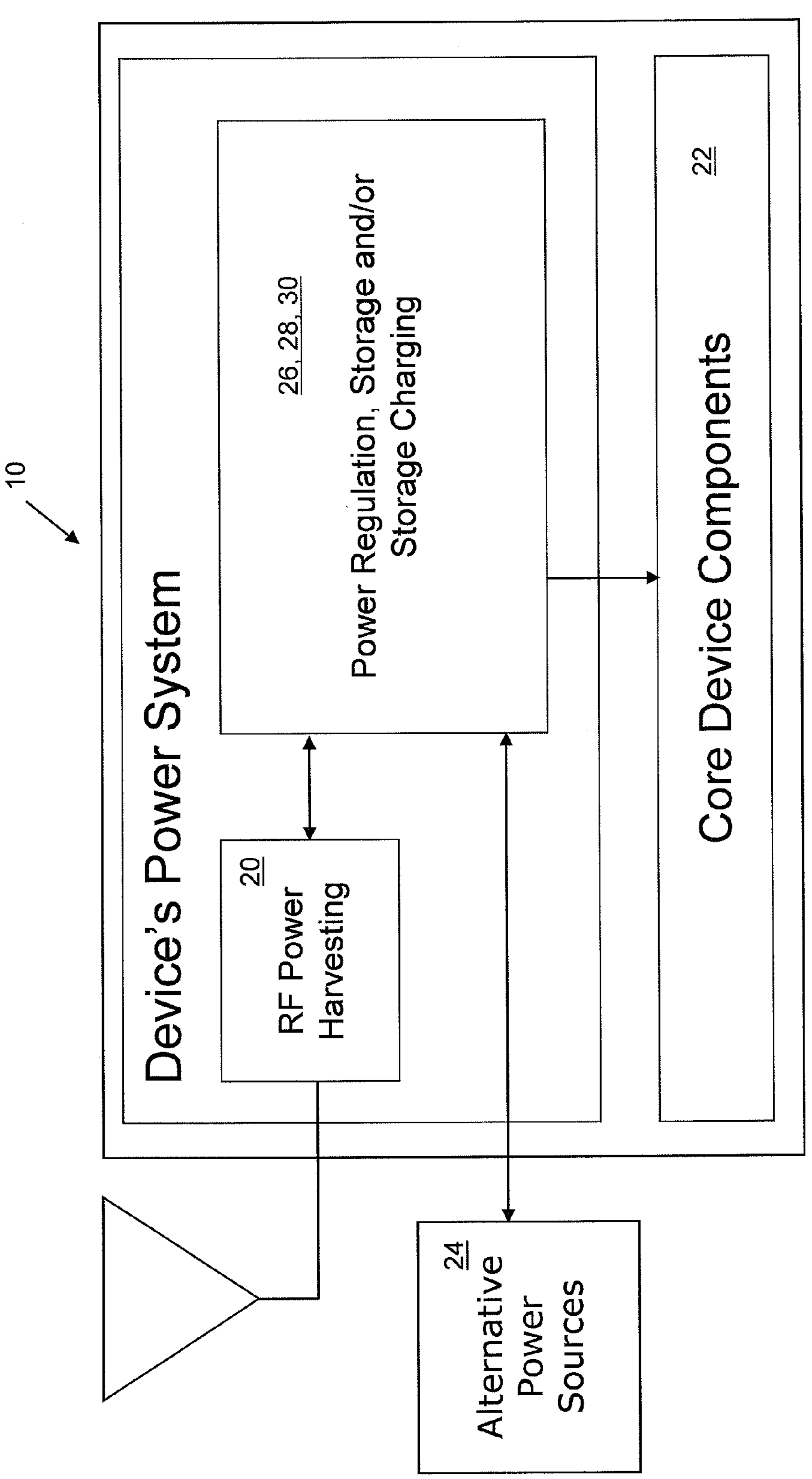
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 65

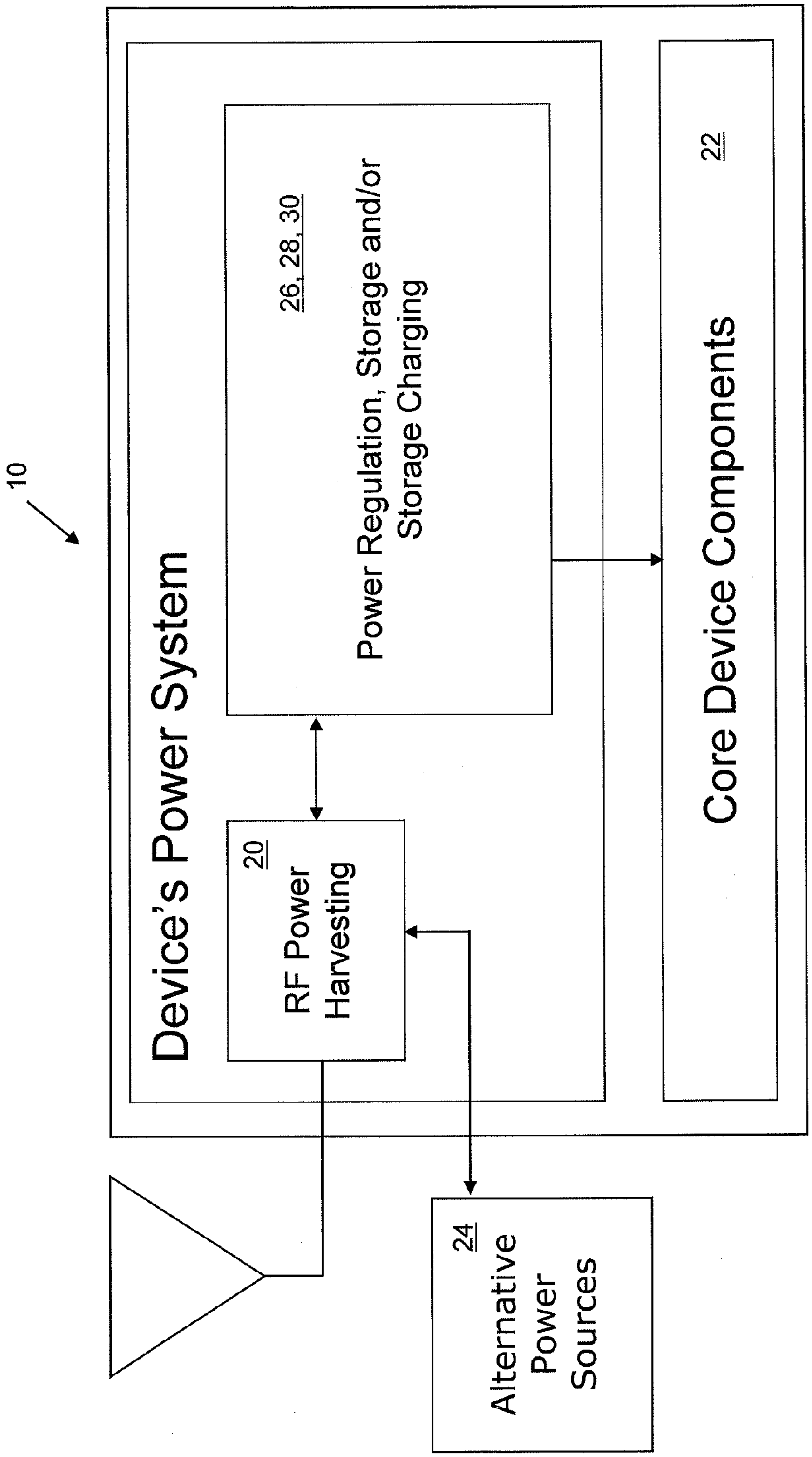
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 66

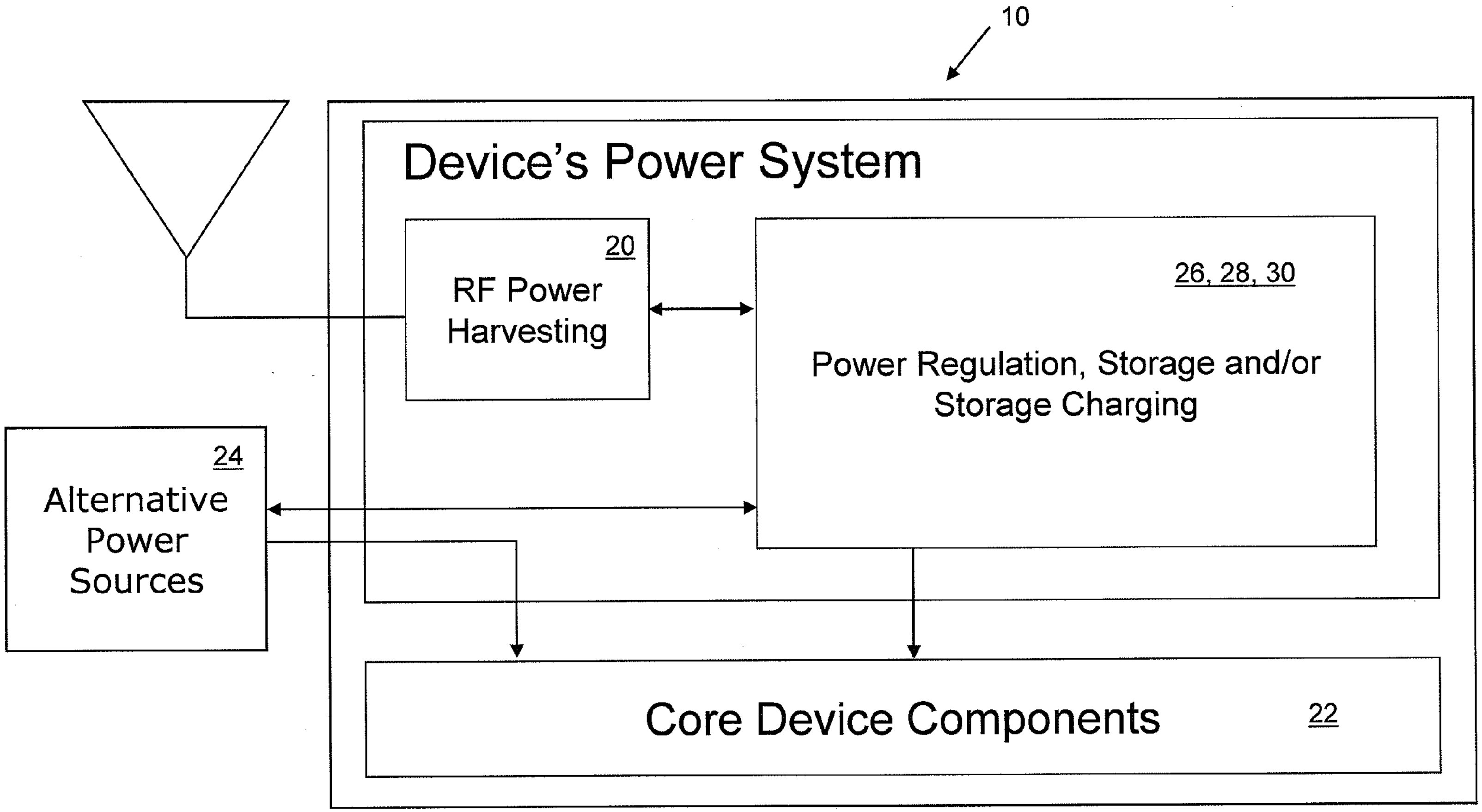
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 67

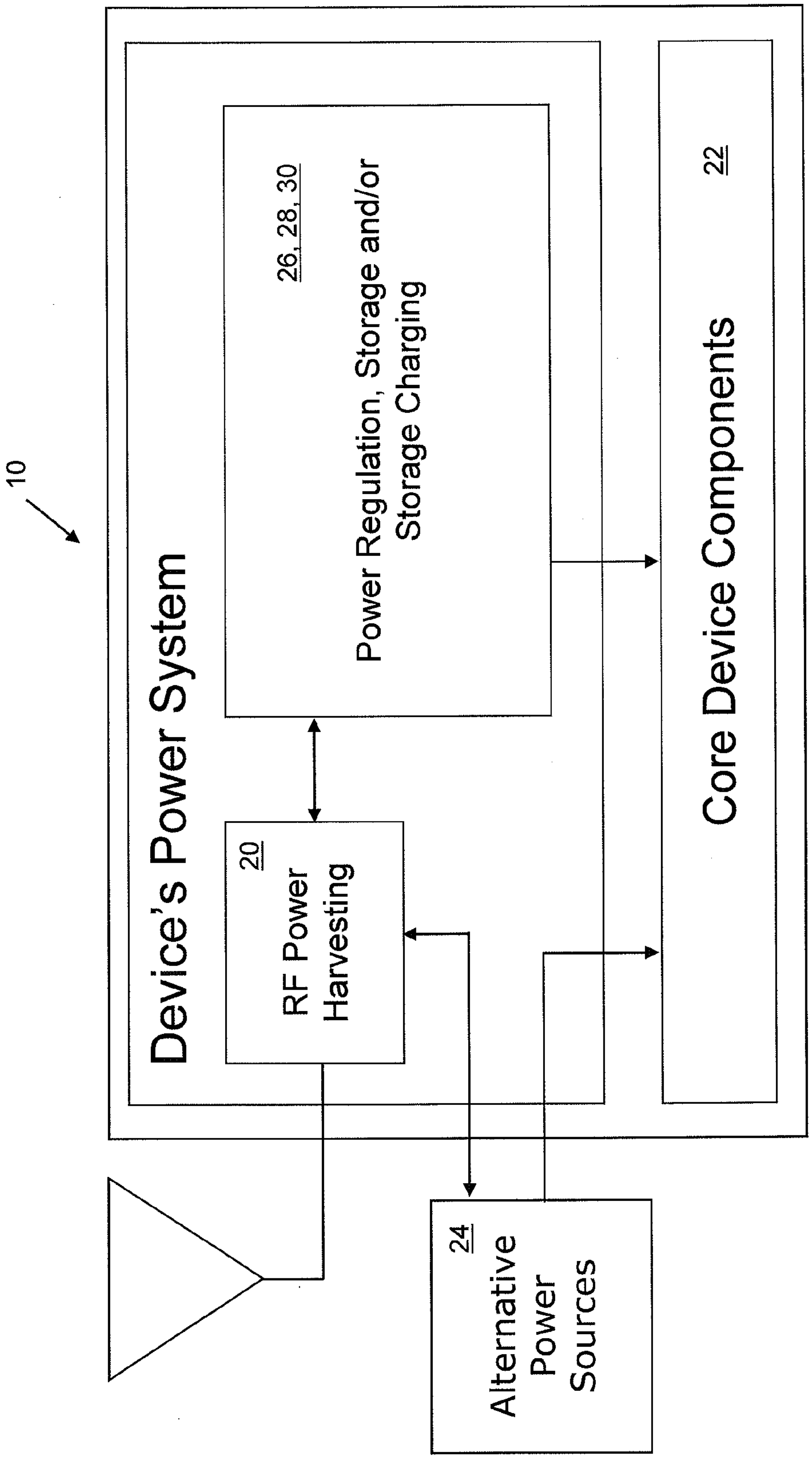
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 68

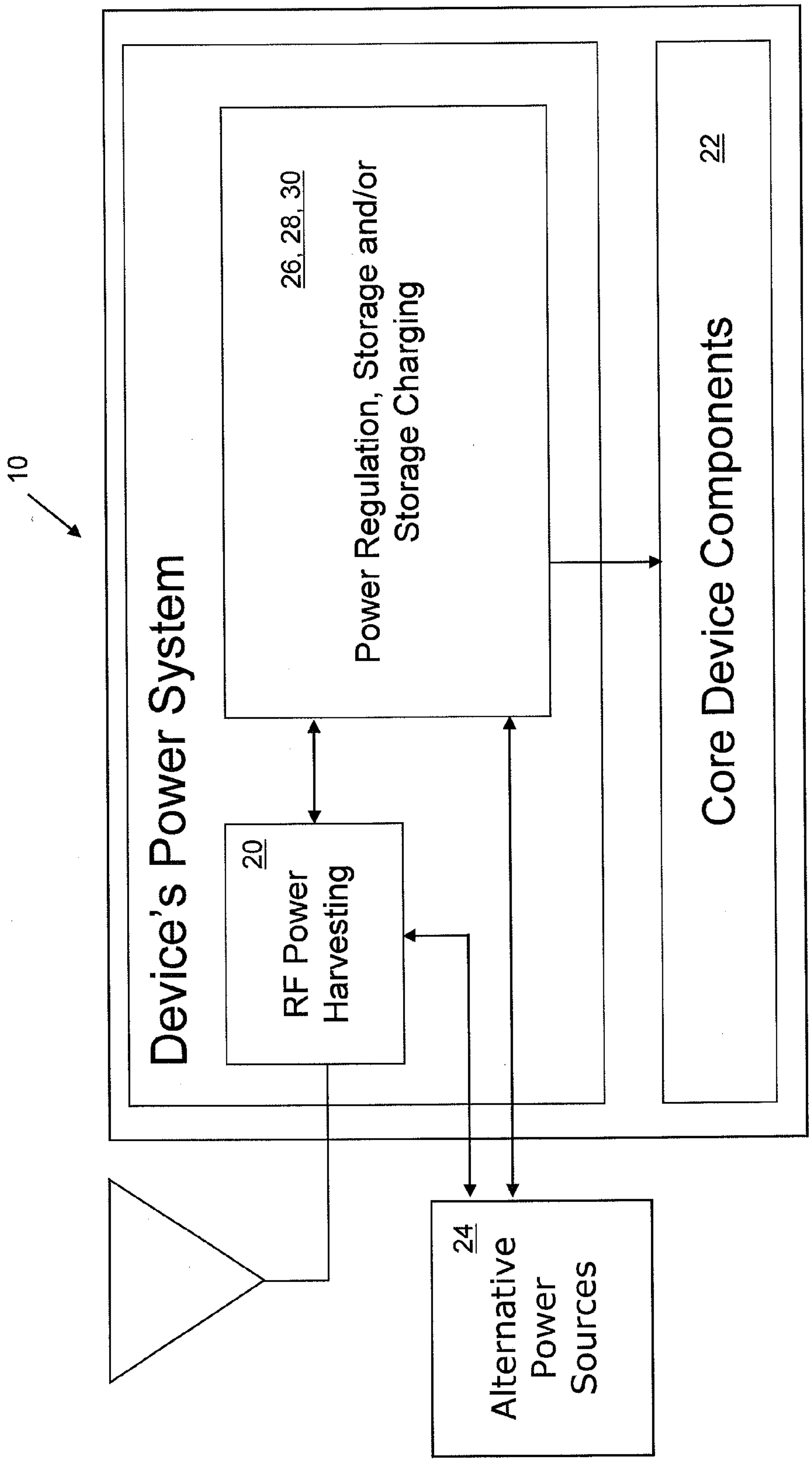
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
22
Core Device Components
24
Alternative Power Sources

Fig. 69

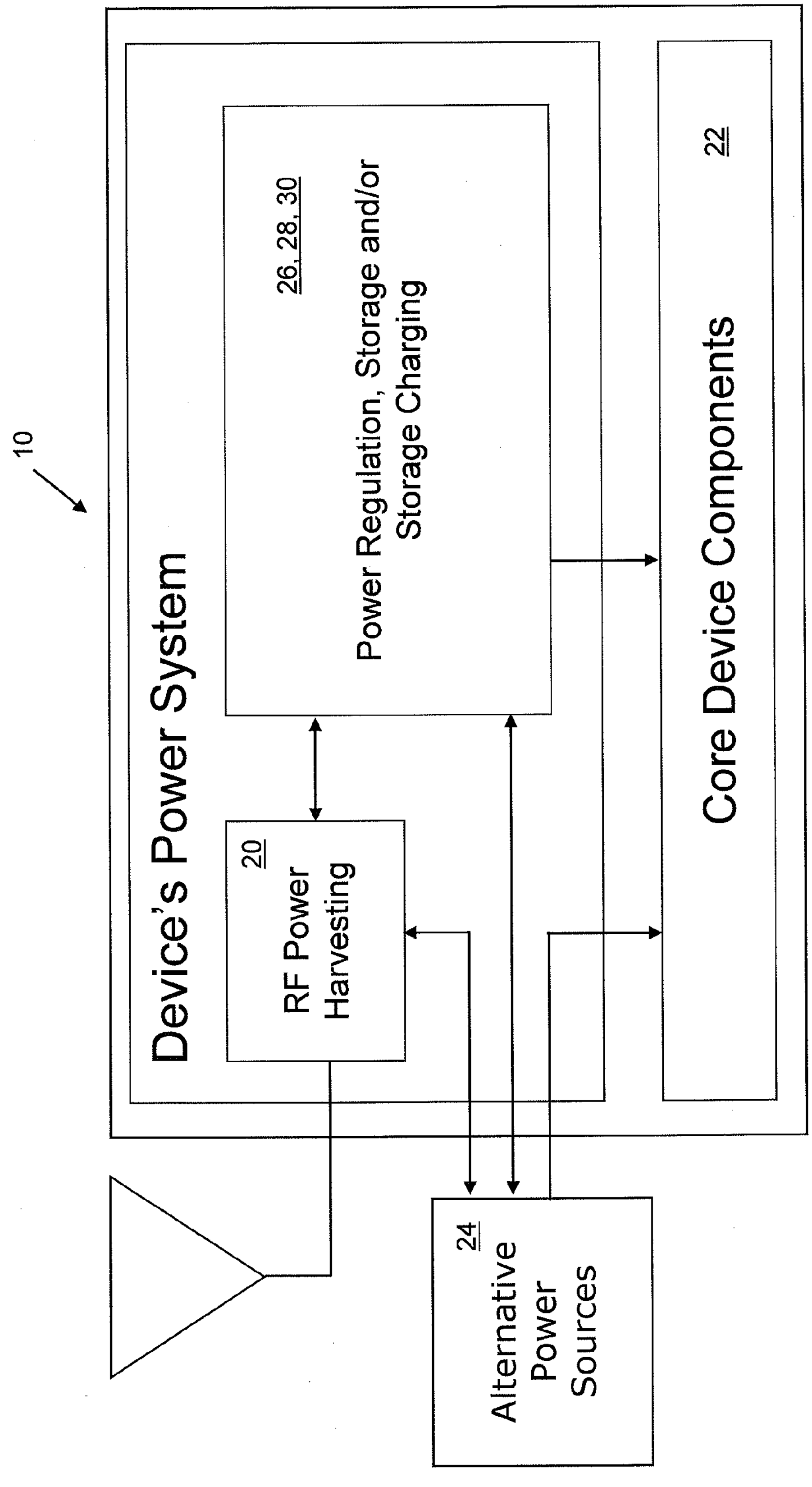
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
22
Core Device Components
24
Alternative Power Sources

Fig. 70

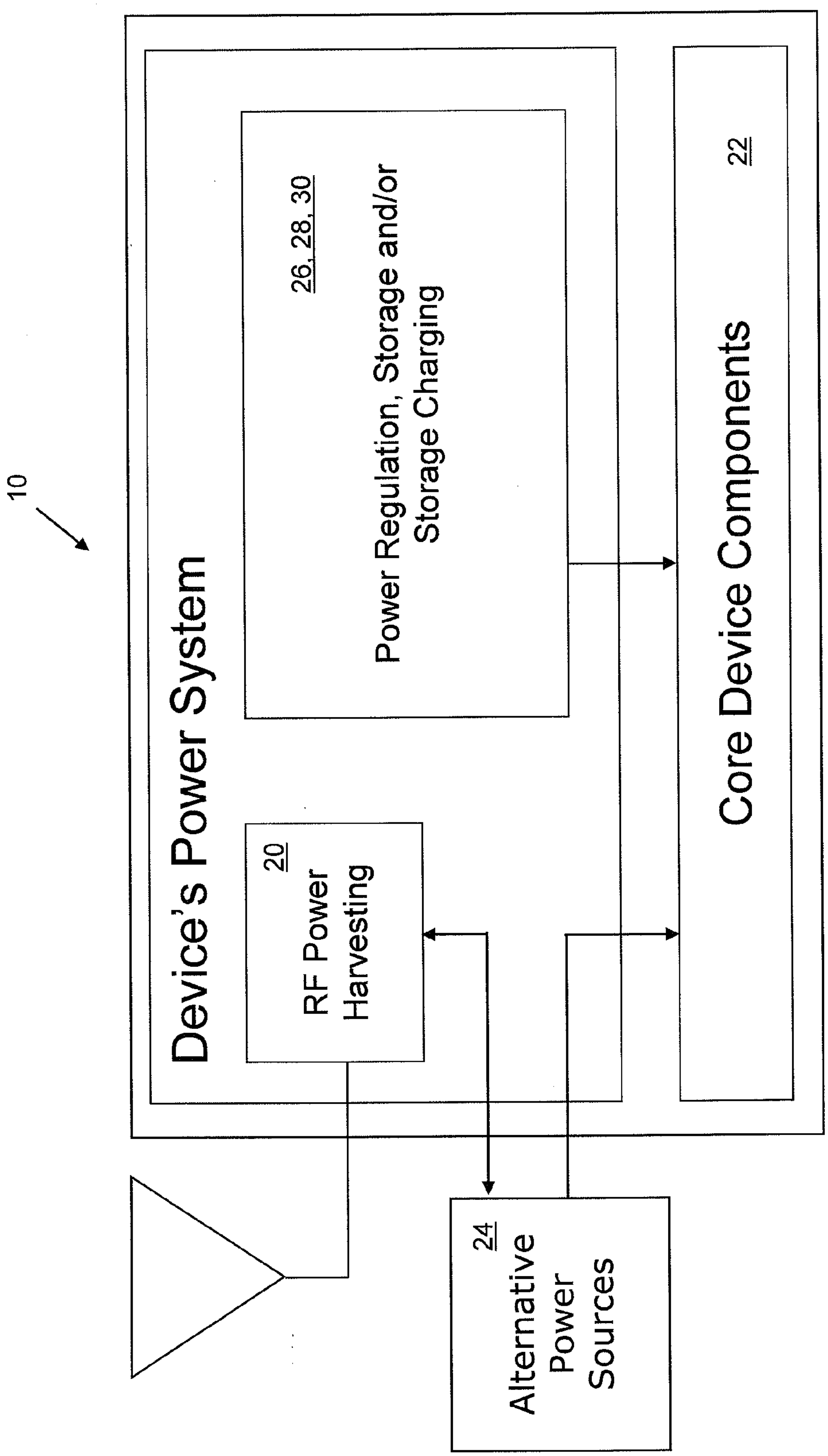
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 71

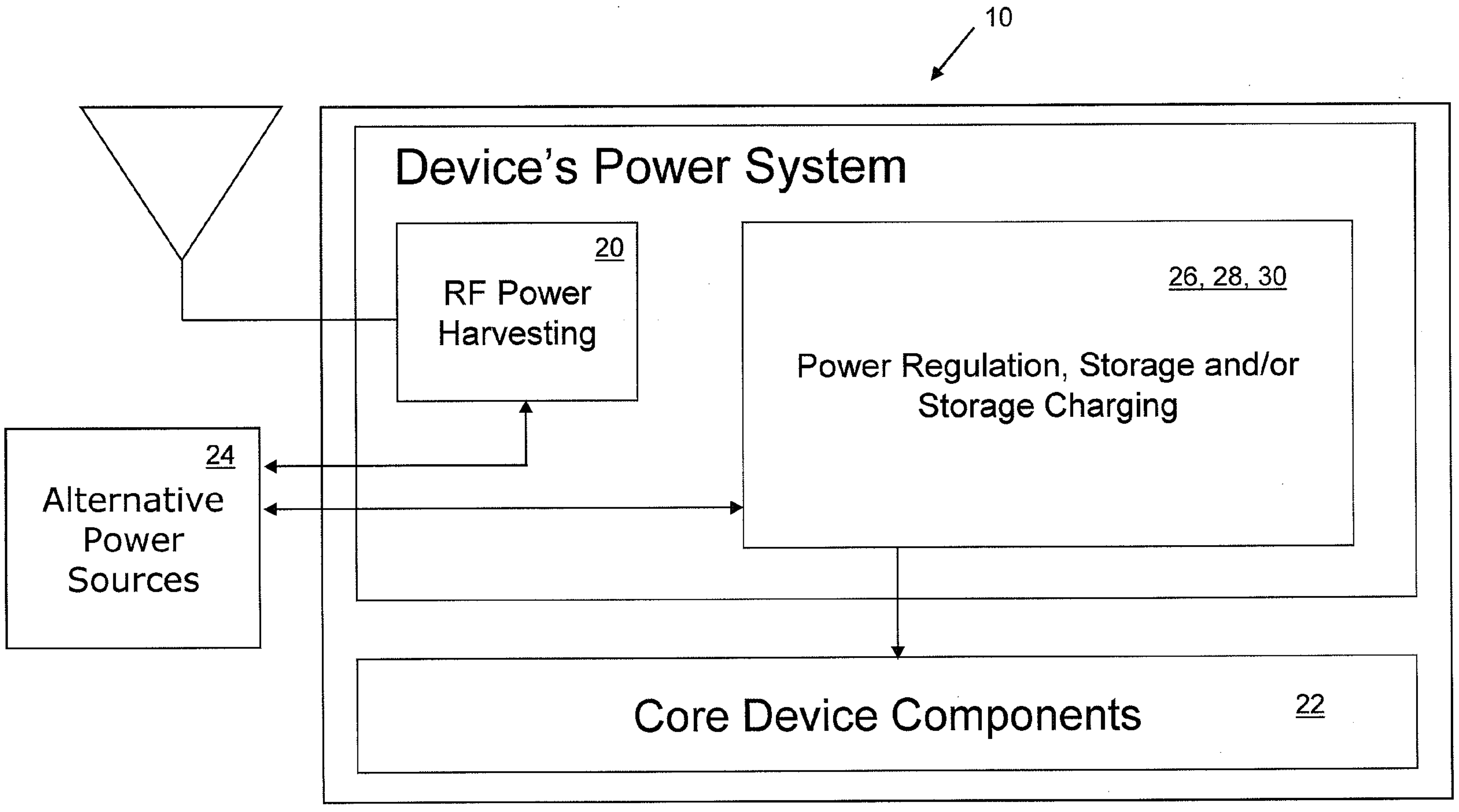
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 72

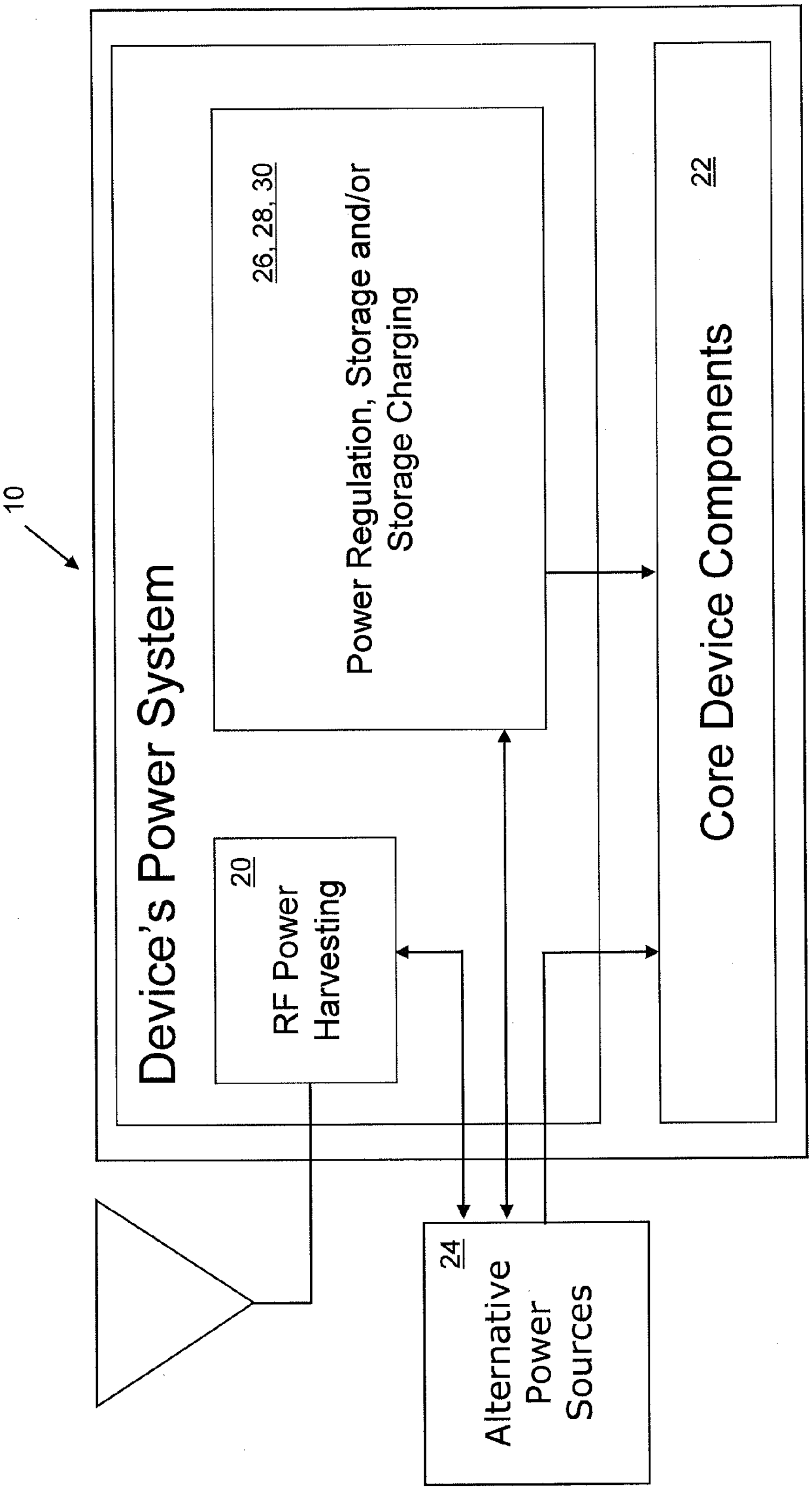
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 73

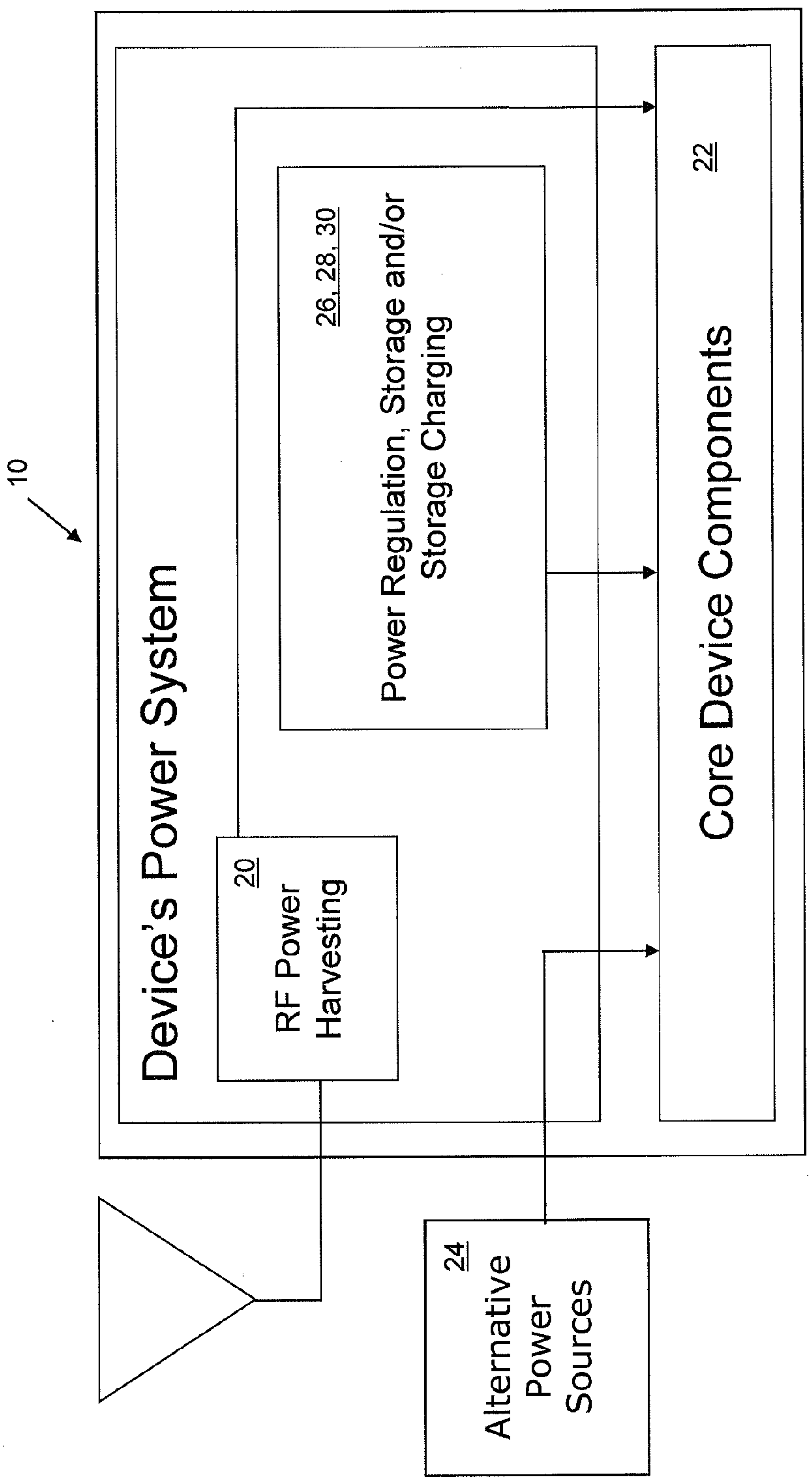
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 74

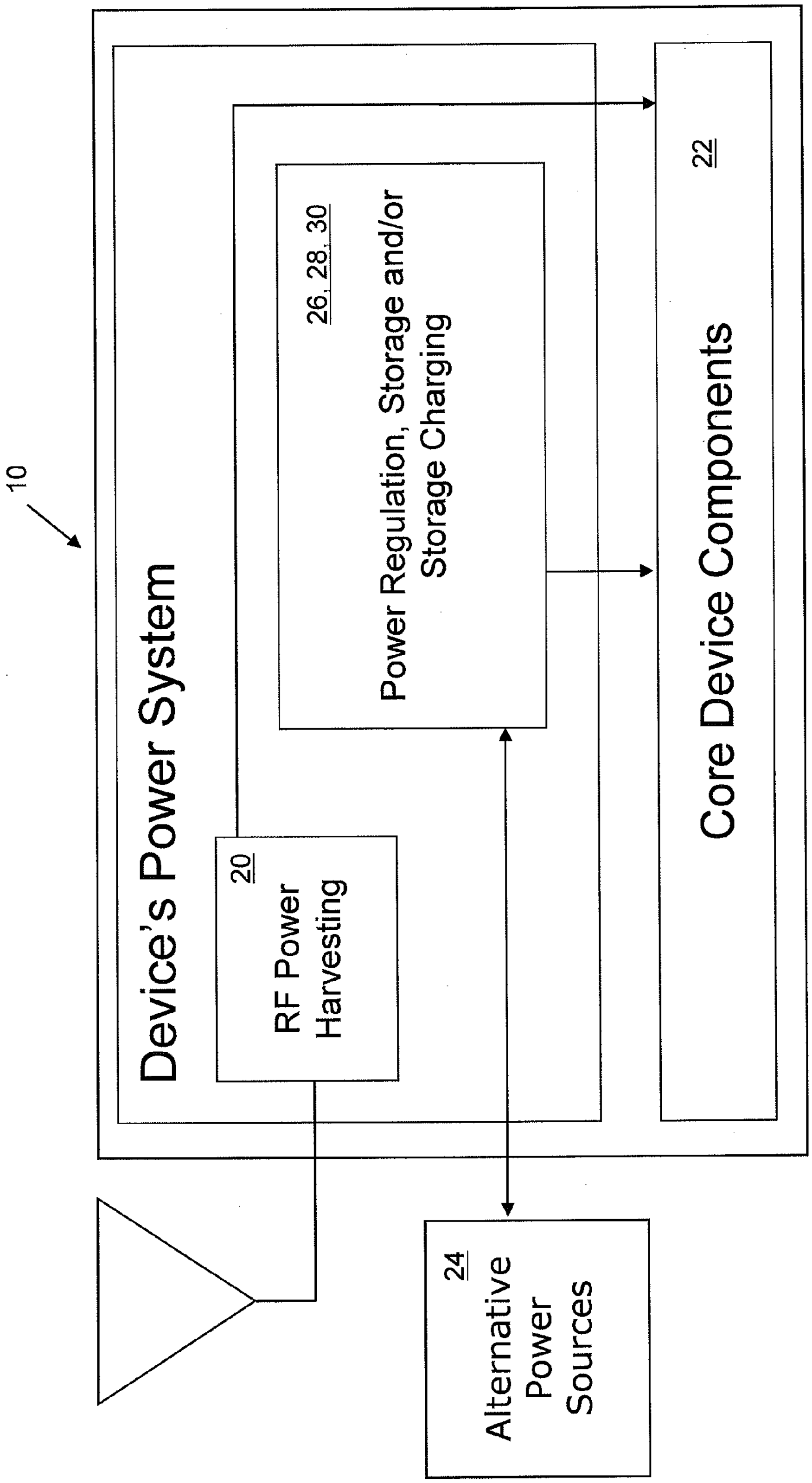
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 75

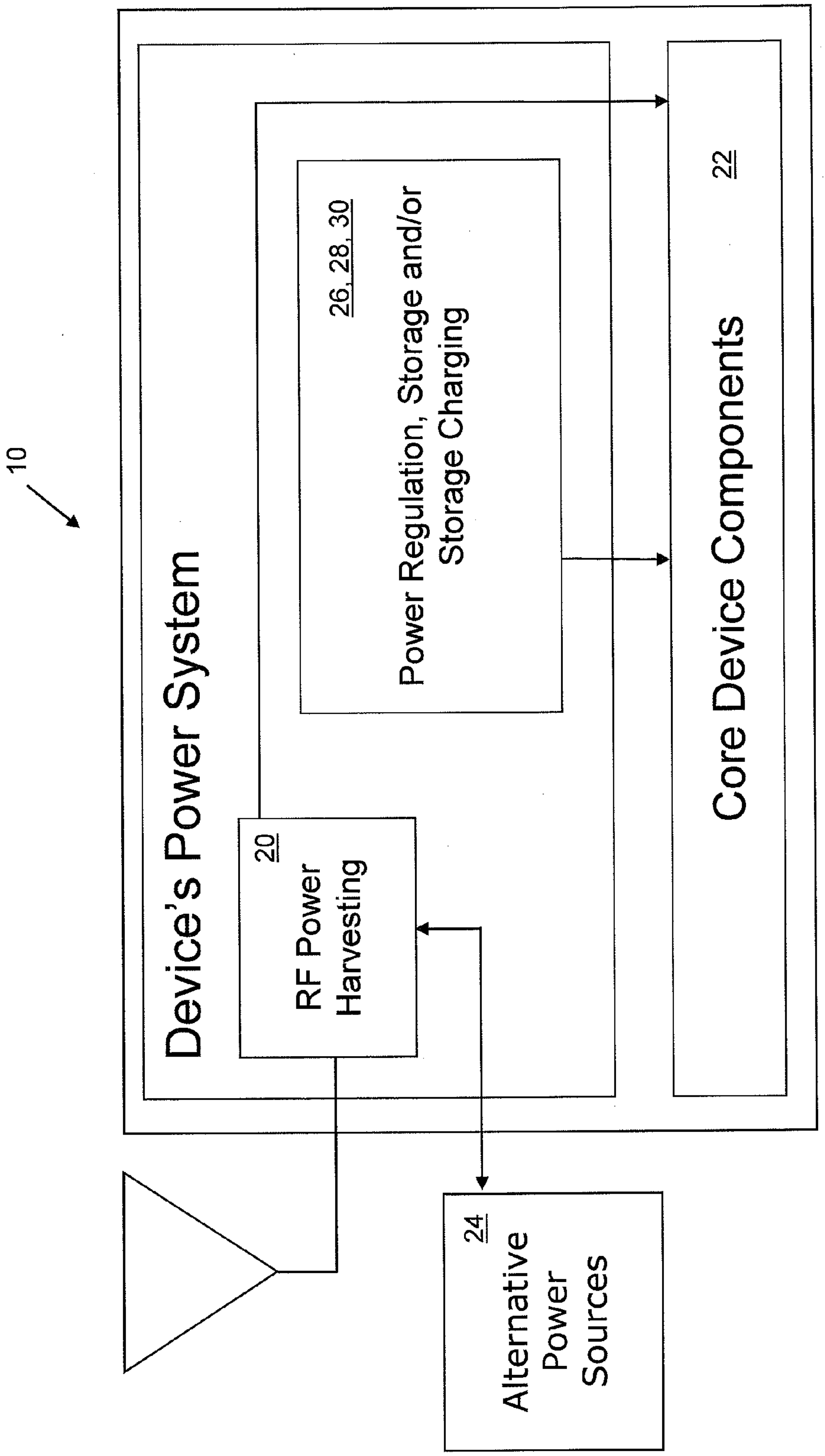
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 76

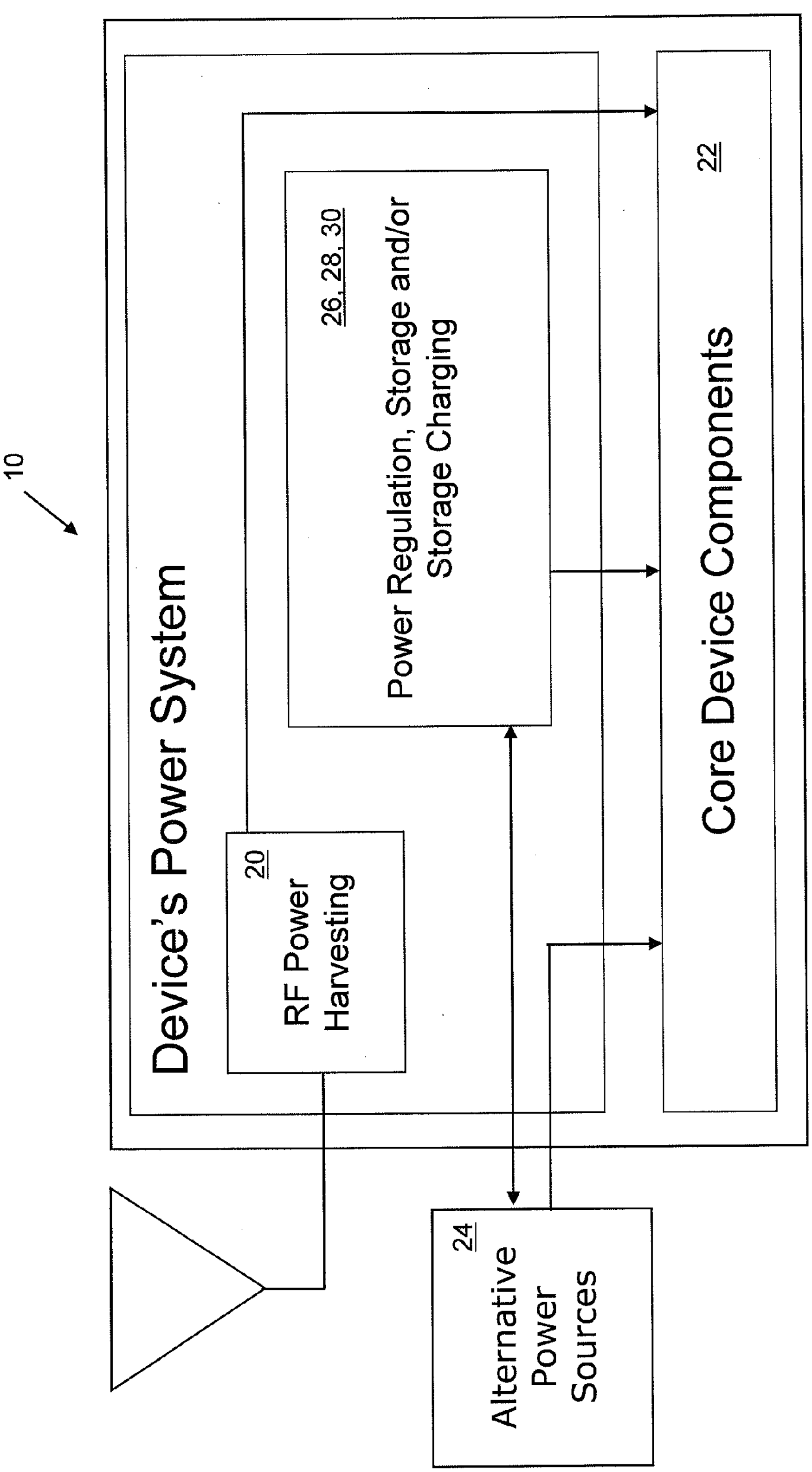
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 77

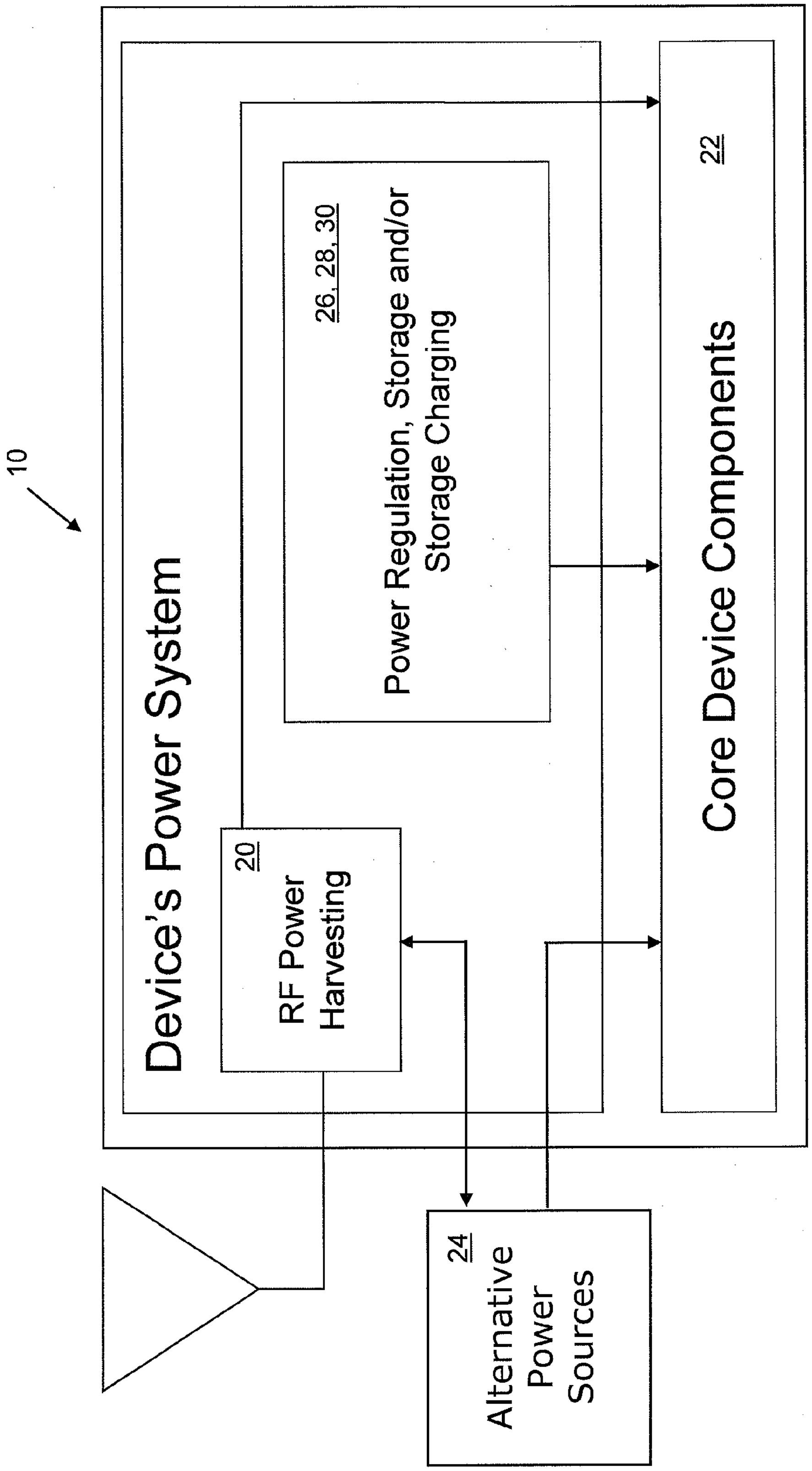
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 78

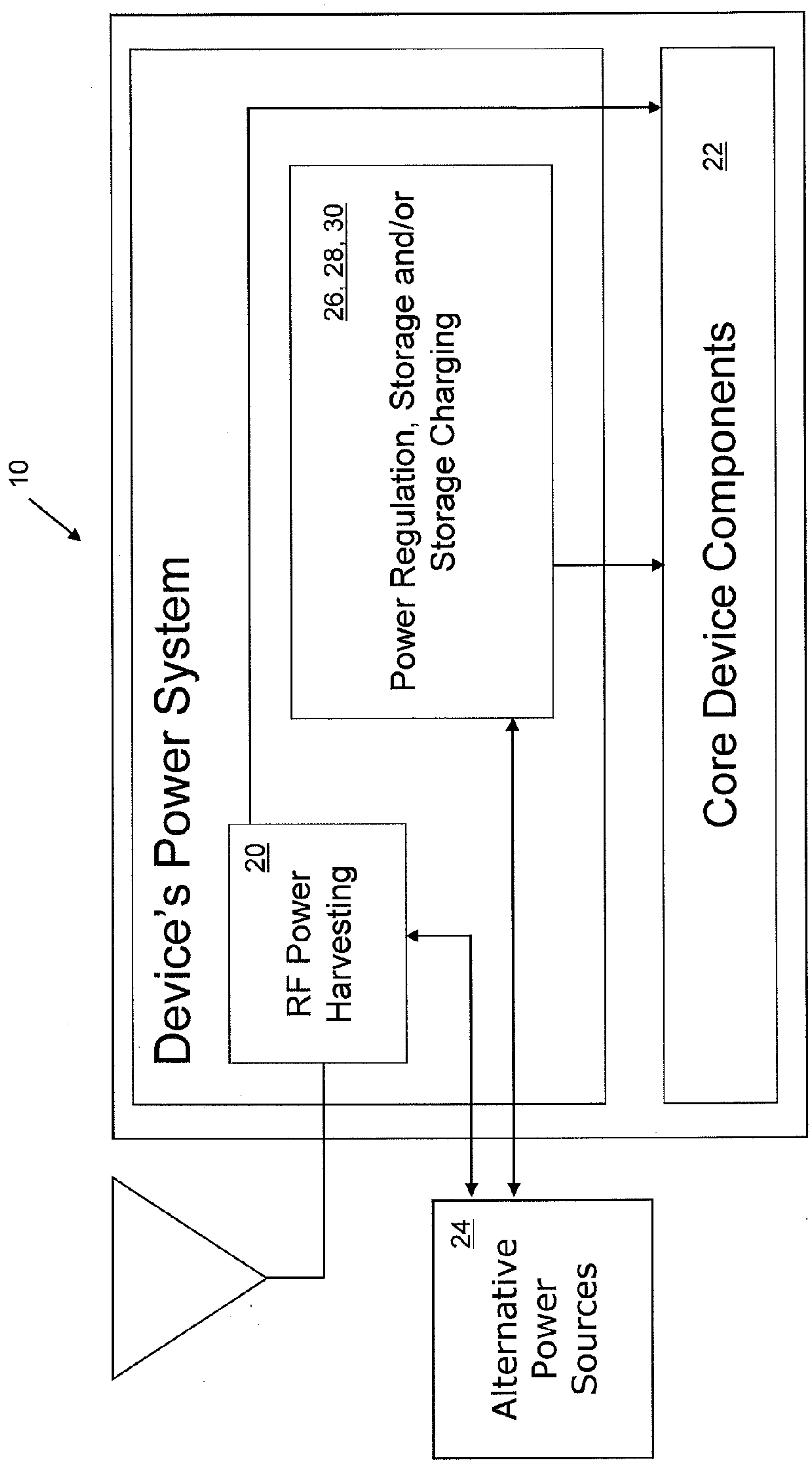
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 79

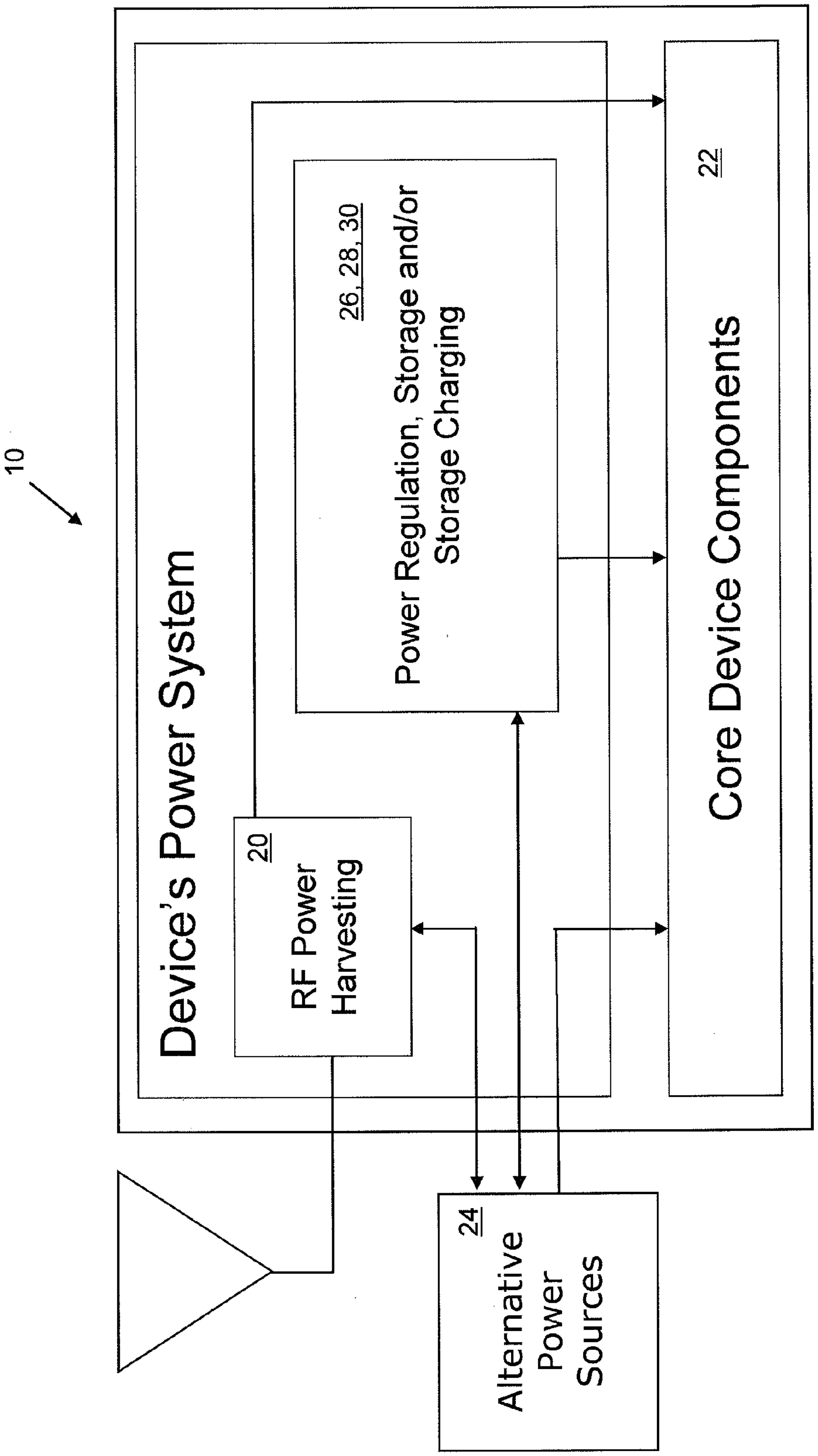
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 80

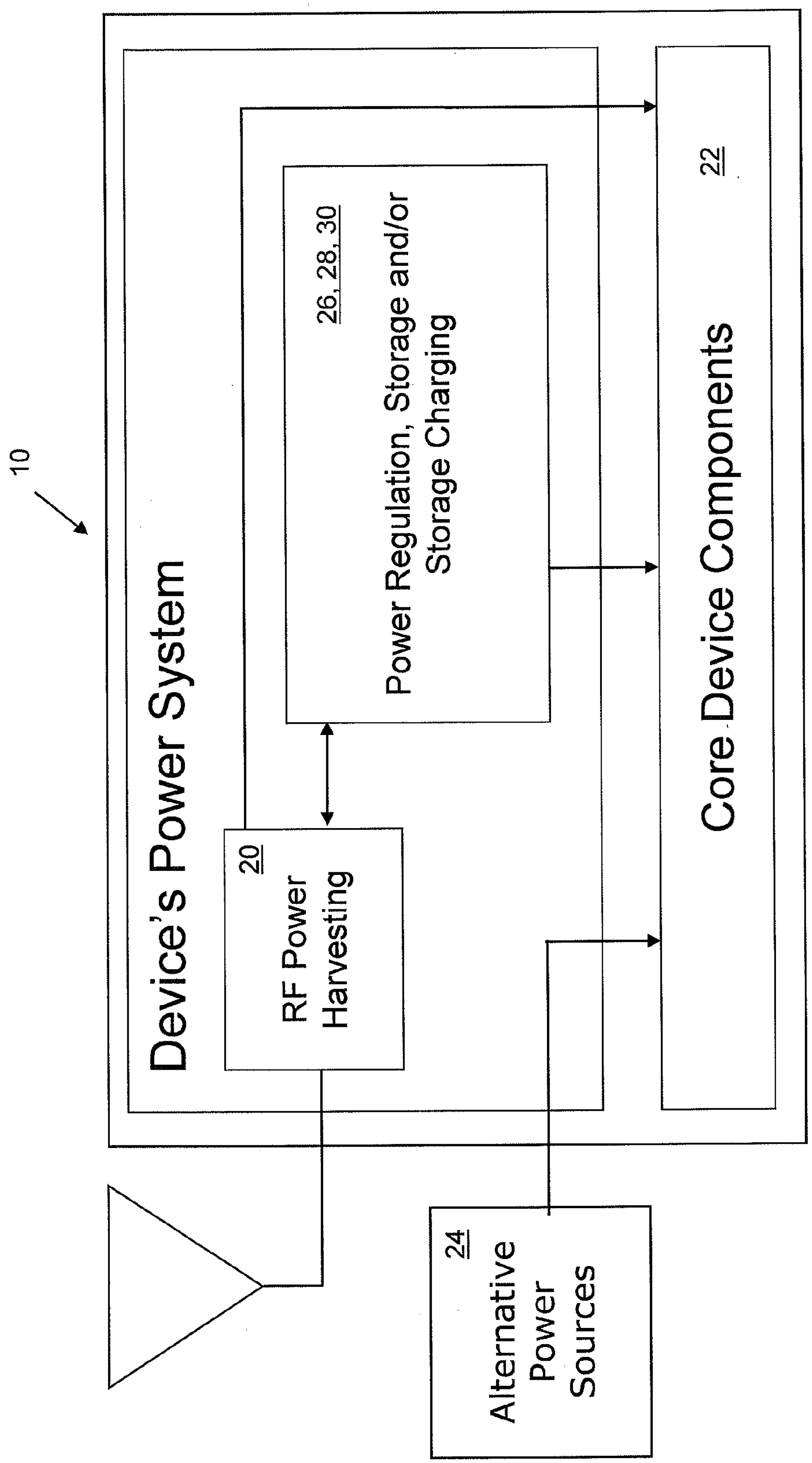
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 81

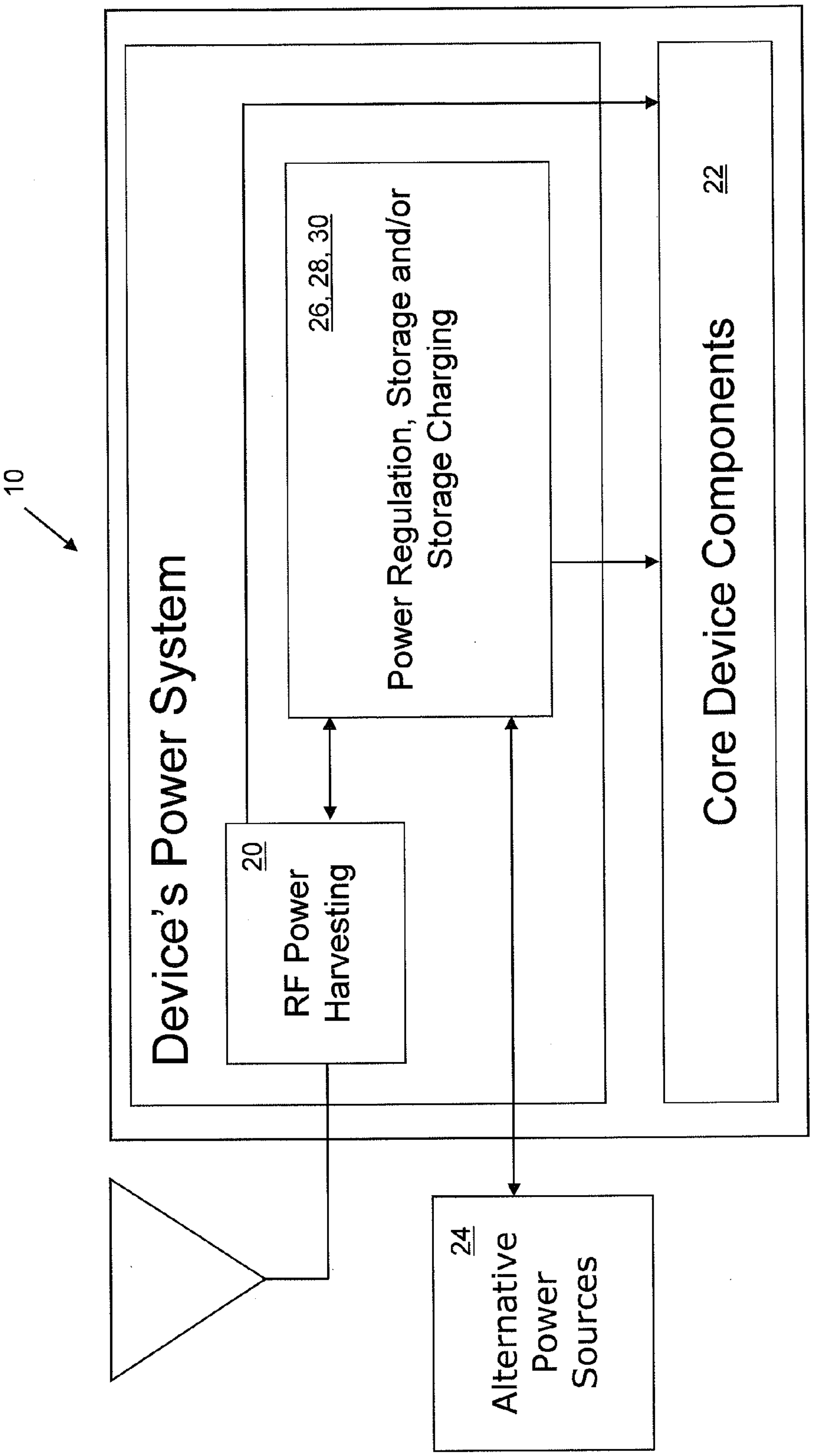
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 82

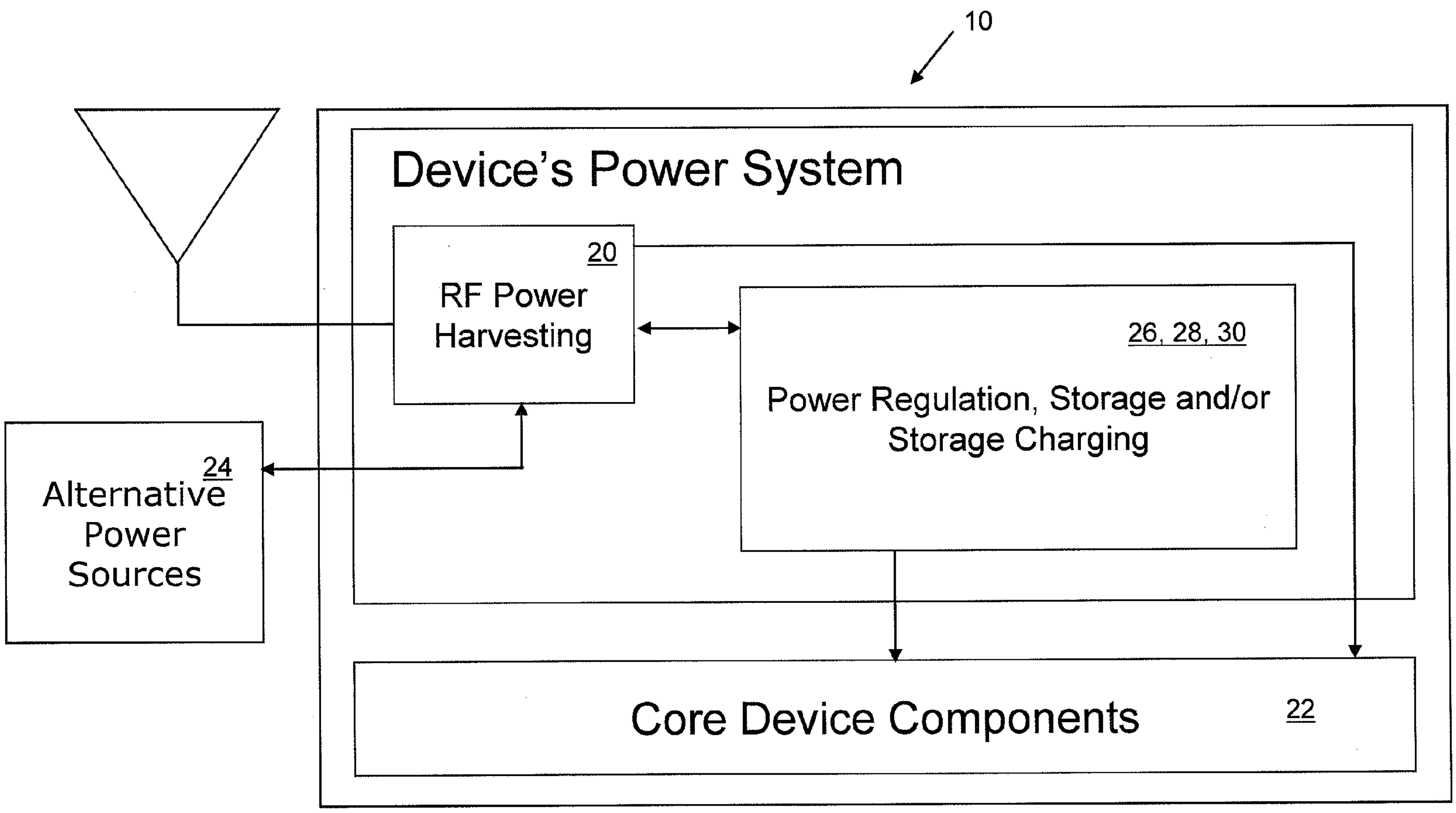
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 83

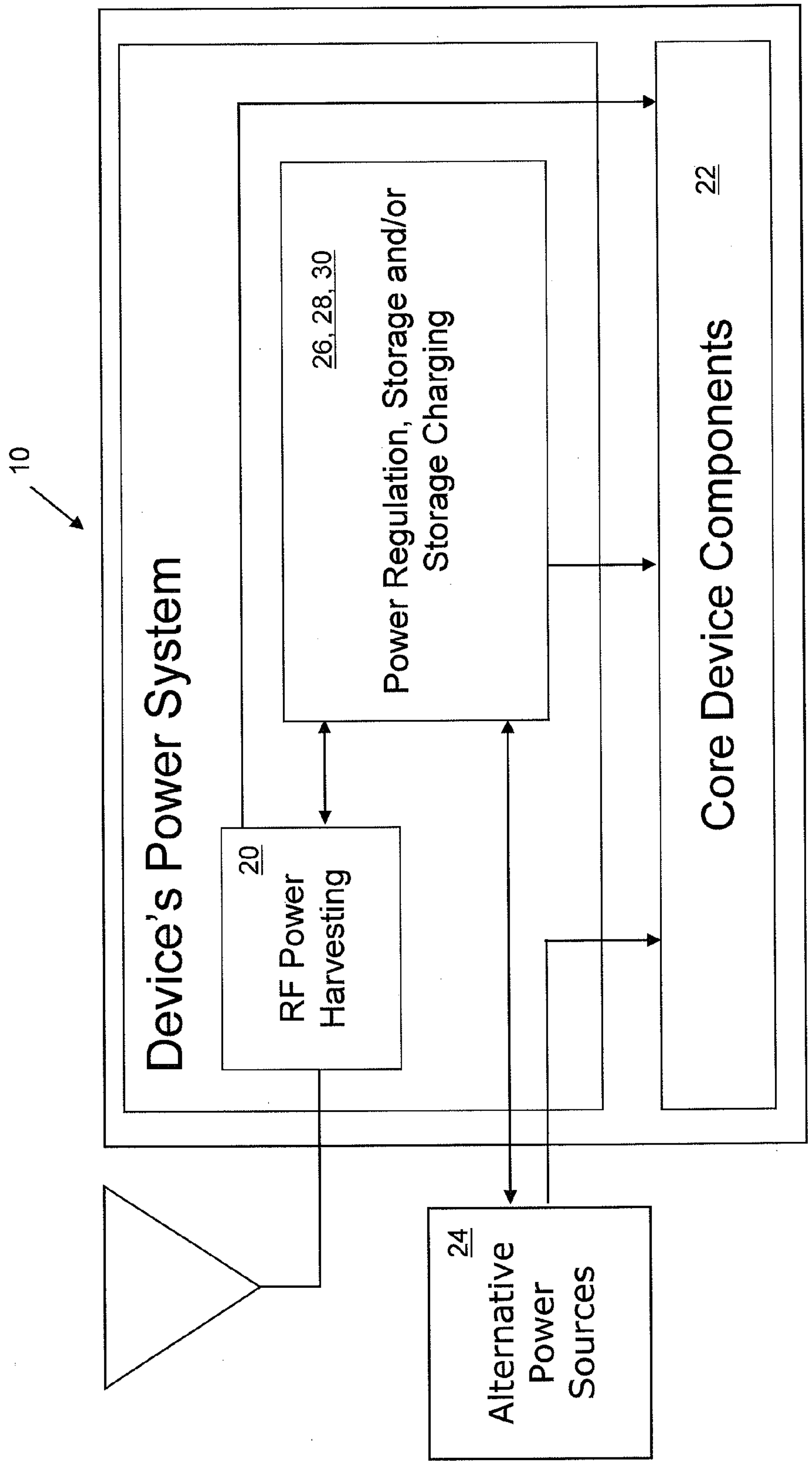
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 84

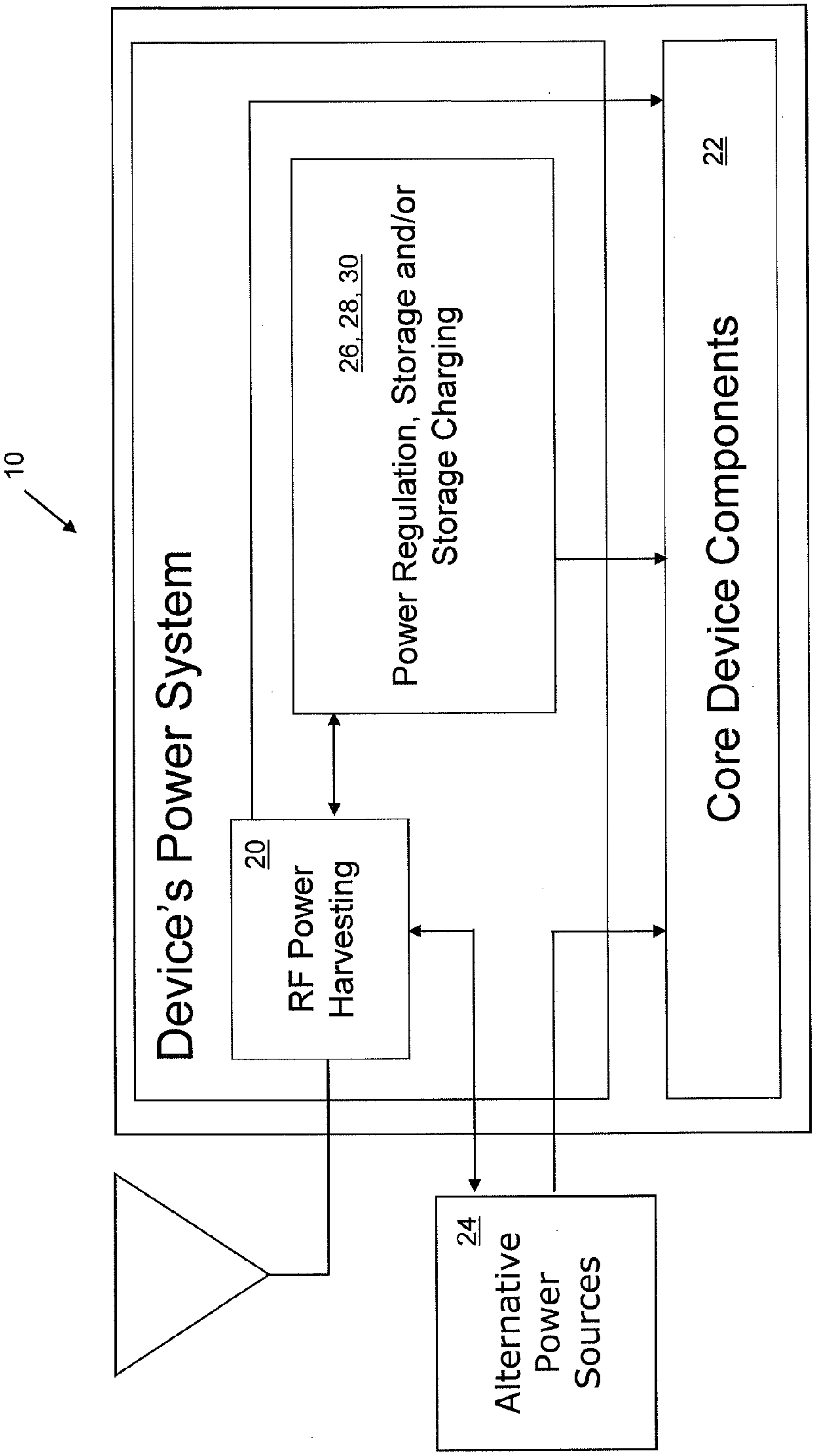
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 85

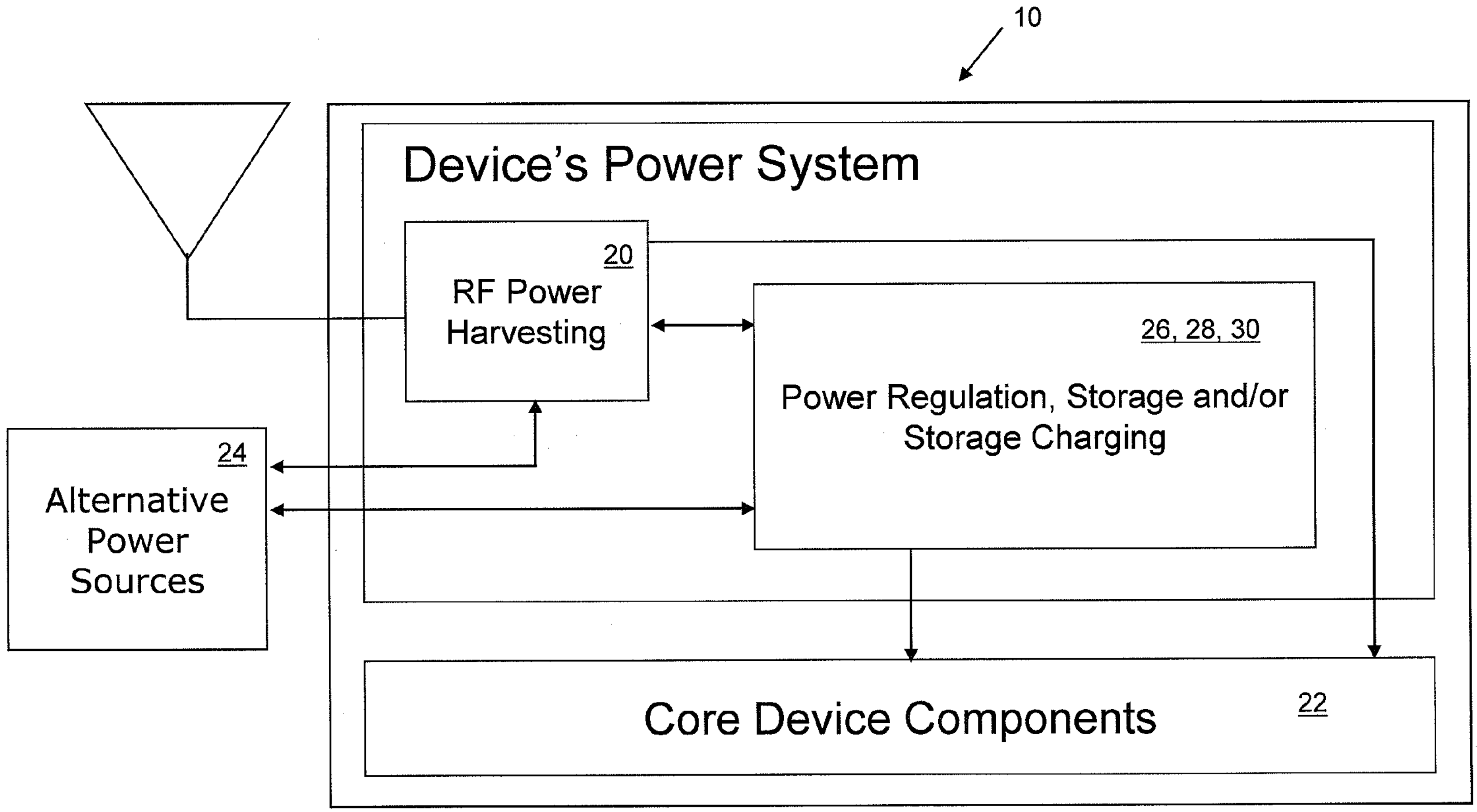
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 86

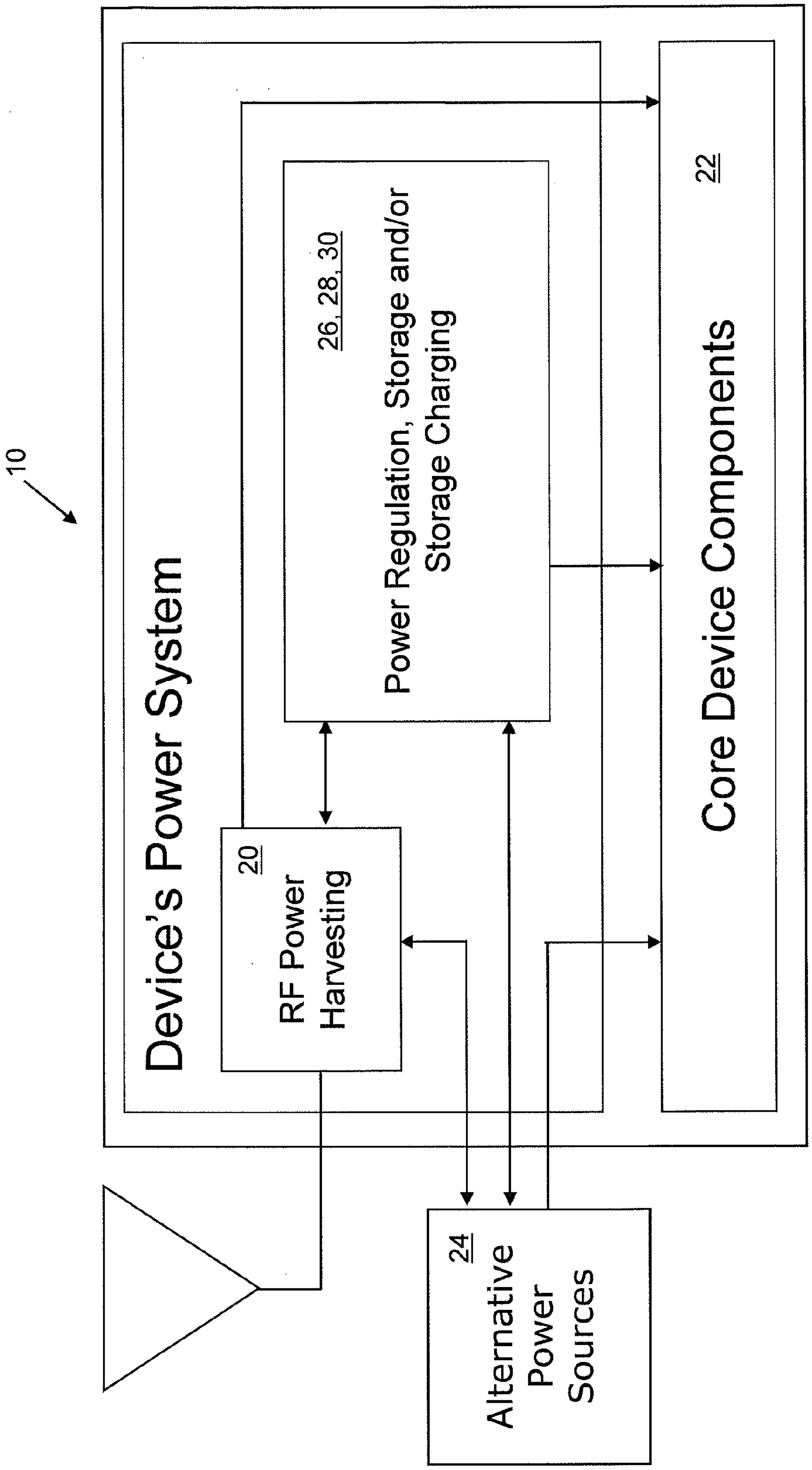
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 87

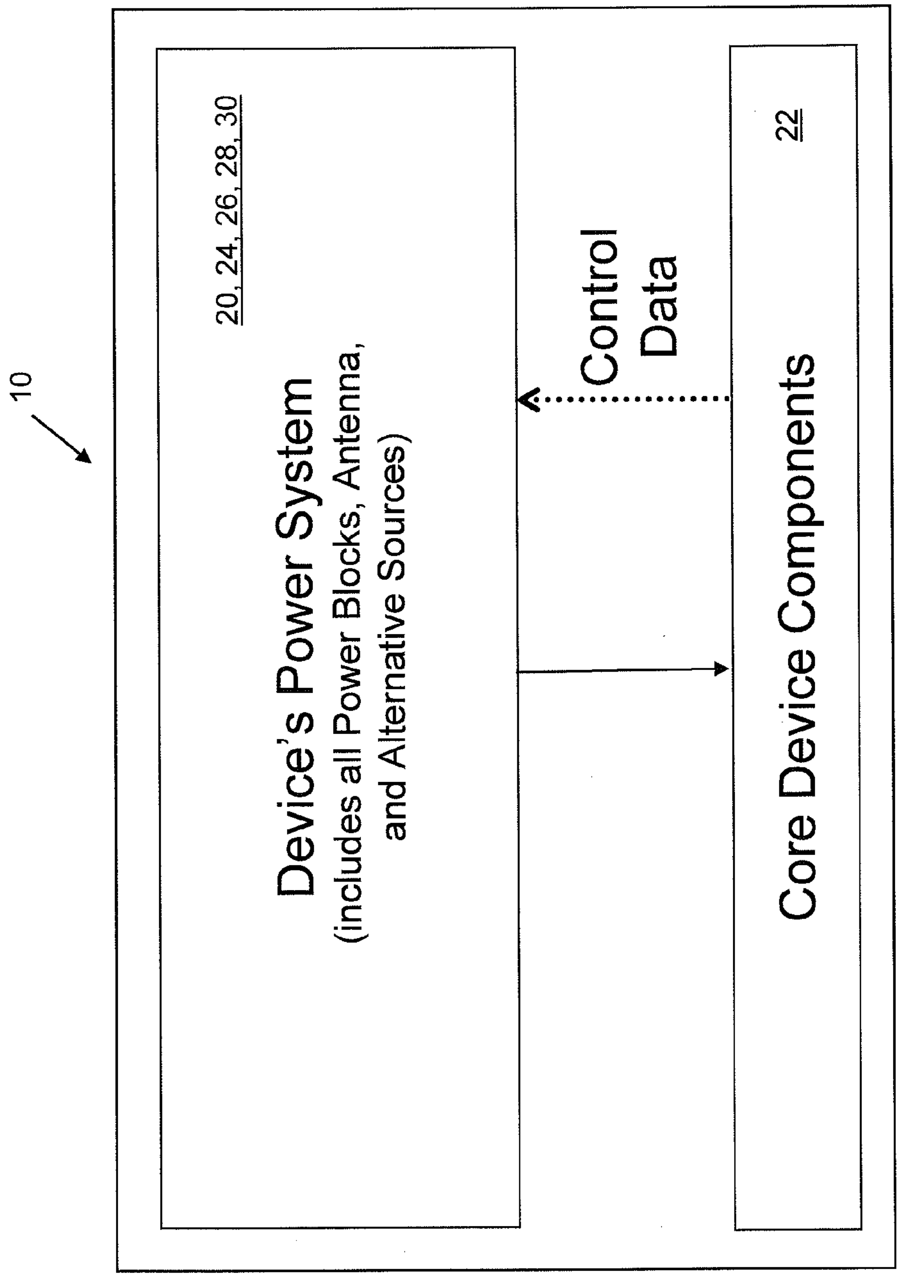
10
Device's Power System
(includes all Power Blocks, Antenna, and Alternative Sources)
20, 24, 26, 28, 30
Control Data
Core Device Components
22

Fig. 88

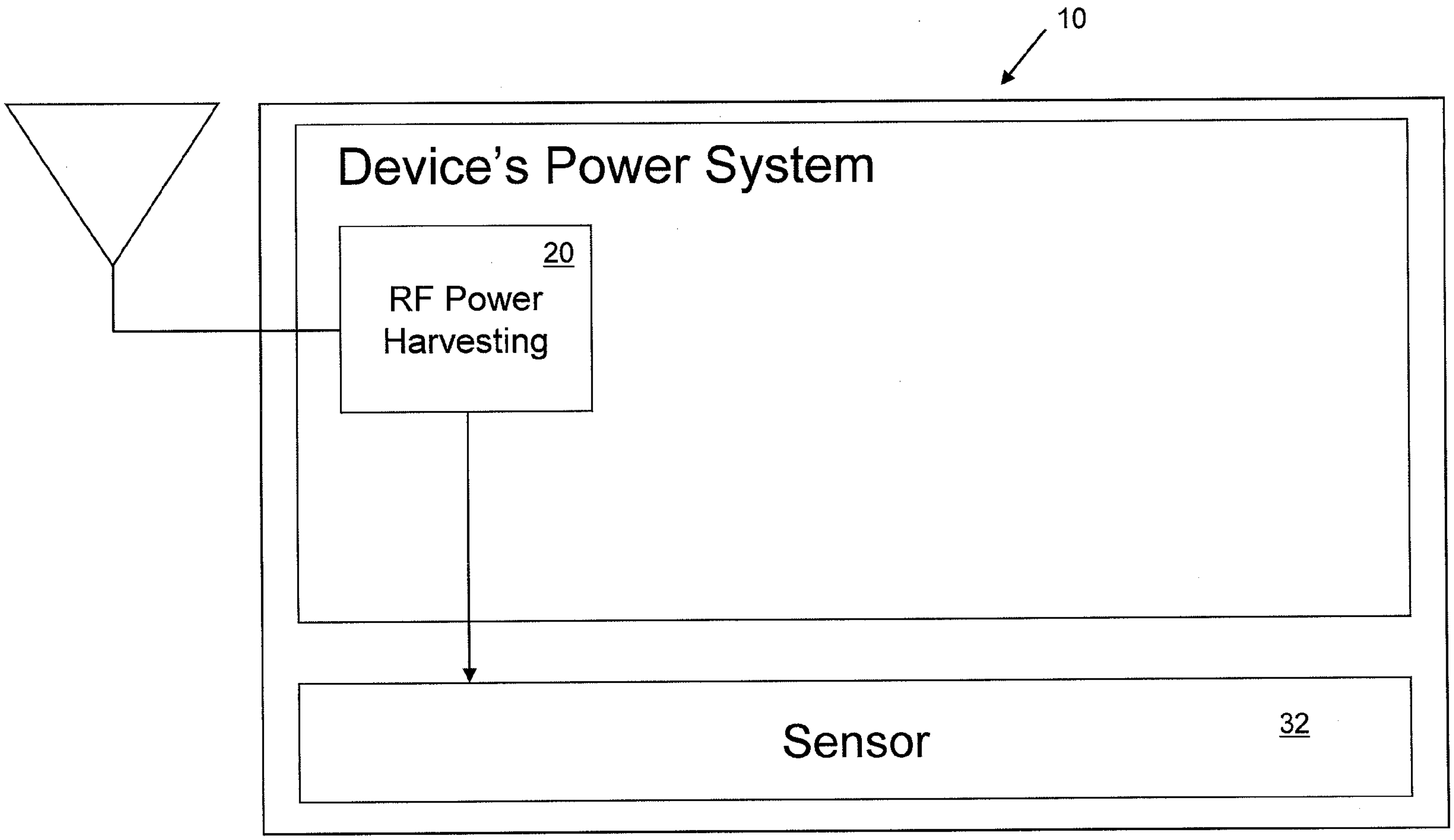
10
Device's Power System
20
RF Power
Harvesting
Sensor
32

Fig. 89

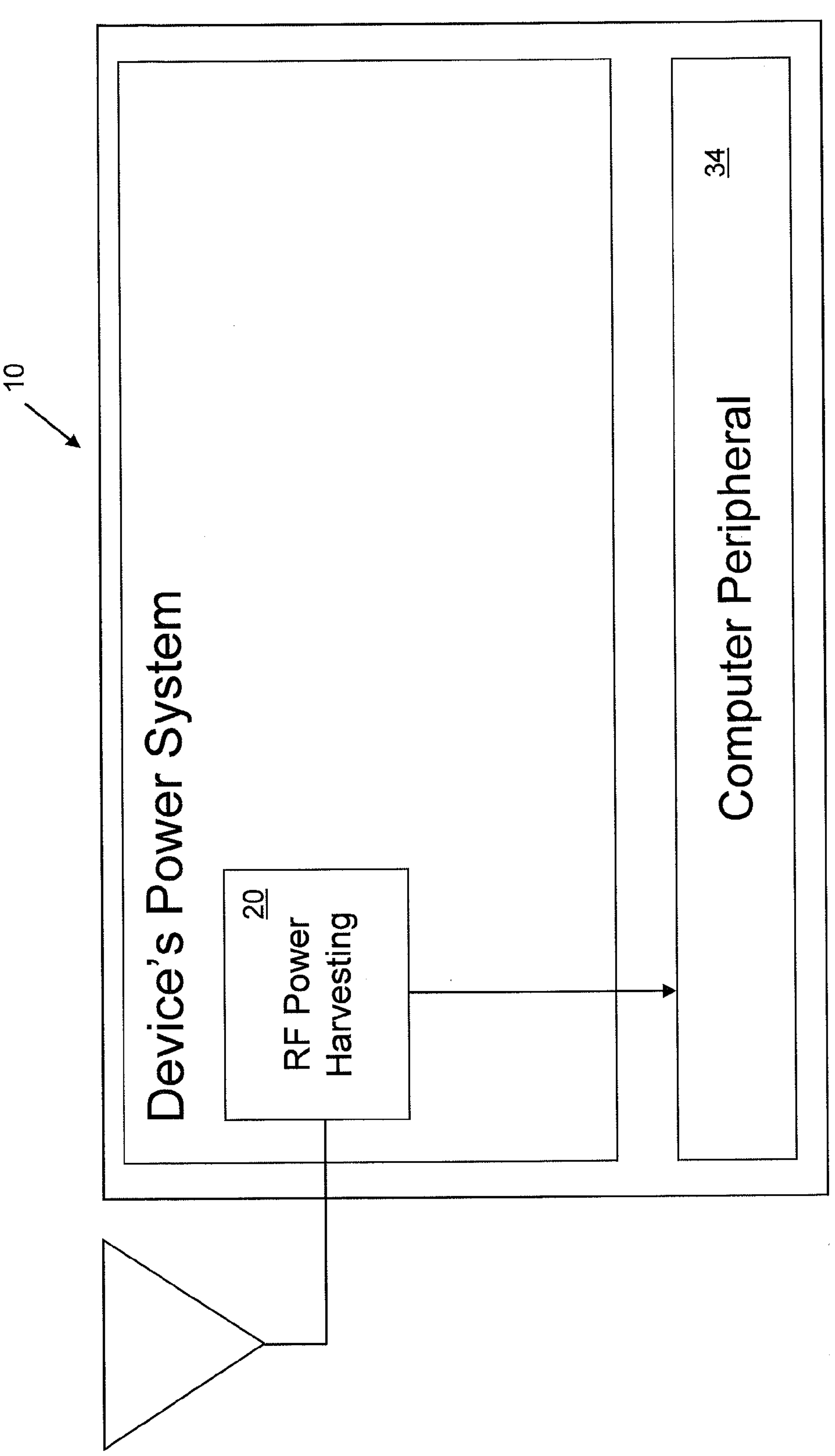
10
Device's Power System
20
RF Power Harvesting
Computer Peripheral
34

Fig. 90

POWERING DEVICES USING RF ENERGY HARVESTING

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to and is a continuation application of U.S. patent application Ser. No. 12/499,618, filed Jul. 8, 2009, which is a divisional application of U.S. patent application Ser. No. 11/447,412, filed Jun. 6, 2006, which claims priority to U.S. Provisional Patent Application No. 60/688,587, filed on Jun. 8, 2005; each of which is incorporated herein by reference in its entirety.

BACKGROUND

[0002] The present invention is related to the wireless powering of devices. More specifically, the present invention is related to the wireless powering of devices with a power harvester.

[0003] As processor capabilities have expanded and power requirements have decreased there has been an ongoing explosion of devices that operate completely independent of wires or power cords. These "untethered" devices range from cell phones, and wireless keyboards to building sensors and active RFID tags.

[0004] Engineers and designers of these untethered devices continue to have to deal with the limitations of portable power sources, primarily batteries as the key design parameter. While performance of processors and portable devices have been doubling every 18-24 months driven by Moore's law, battery technology in terms of capacity has only been growing at measly 6% per year. Even with power conscious designs and the latest in battery technology, many devices do not provide the lifetime cost and maintenance requirements for applications that require a large number of untethered devices such as logistics, and building automation. Today's devices that need two-way communication require scheduled maintenance every three to 18 months to replace or recharge the device's power source (typically a battery). One-way devices simply broadcasting their status (one-way) such as automated utility meter readers have a better battery life, typically requiring replacement within 10 years. For both device types, scheduled power-source maintenance is costly and disruptive to the entire system that a device is intended to monitor and/or control. Unscheduled maintenance trips are even more costly and disruptive. On a macro level, the relatively high cost associated with the internal battery also reduces the practical, or economically viable, number of devices that can be deployed.

[0005] The ideal solution to the power problem for untethered devices is a device or system that can collect and harness sufficient energy from the external environment. The harnessed energy would then either directly power an untethered device or augment a battery or other storage component. Directly powering an untethered device enables the device to be constructed without the need for a battery. Augmenting a storage component could be along two lines: 1) increasing the overall life of the device or 2) by providing more power to the device to increase the functionality of the device. The other parameters for an ideal solution is that the harnessing device could be used in a wide range of environments including harsh and sealed environments (e.g. nuclear reactors), would be inexpensive to produce, would be safe for humans, and would have a minimal effect on the basic size, weight and other physical characteristics of the untethered device.

SUMMARY

[0006] The present invention pertains to an apparatus for an application. The apparatus comprises a core device preferably having an integrated circuit for the application. The apparatus comprises a power harvester (preferably a radio frequency (RF) energy harvester) connected to the core device to power the core device.

[0007] The present invention pertains to an apparatus for an application. The apparatus comprises a core device preferably having an integrated circuit for the application. The apparatus comprises means for receiving energy wirelessly and providing power from the energy to the core device to power the integrated circuit of the core device. The receiving means is connected to the core device.

[0008] The present invention pertains to a method for an application. The method comprises the steps of converting RF energy into usable energy. There is the step of preferably powering an integrated circuit of the core device with the usable energy.

[0009] This invention pertains to a technique that uses radio frequency (RF) energy as a source of energy to directly power a device or augment a power storage component in a device. The present invention meets the requirements described in the previous "Background of the Invention" section.

[0010] Traditional RF receiving devices have focused on maximizing selectivity of the frequency to isolate and to be coherent without interference from other sources. In contrast, while the present invention operates at a specific frequency or range of frequencies, the device accepts any interference to supplement the output power of the device. Also, the research related to power harvesting that uses RF energy as a source has primarily focused on devices in close proximity of the source (inductive or near-field energy). In most cases, prior research assumed a dedicated or directed source of RF to power the device.

[0011] The invention should not be confused with power transfer by inductive coupling, which requires the device to be relatively close to the power transmission source. The RFID Handbook by the author Klaus Finkenzeller defines the inductive coupling region as distance between the transmitter and receiver of less than 0.16 times lambda where lambda is the wavelength of the RF wave. The invention can be implemented in the near field (sometimes referred to as inductive) region as well as the far-field region. The far-field region is distances greater than 0.16 times lambda.

[0012] It is an object of this invention to provide a method and apparatus to

[0013] 1. remotely energize an untethered device without using direct wiring

[0014] 2. power or augment the life of the power storage component so it matches the life of the device and, ultimately, powers the off-grid device with or without the use of batteries

[0015] 3. allow untethered devices to be virtually maintenance free

[0016] 4. provide augmentation for other energy harvesting technologies (solar, piezoelectric, etc.)

[0017] 5. provide backup power to tethered devices

[0018] It is a further object of this invention to directly power a device or augment a power storage component in a device in conjunction with other power harvesting technologies and storage elements.

[0019] With this method and apparatus a device's power storage components do not require replacement, thus enabling the device to be permanently placed off-grid, where it may be physically impractical, costly, or dangerous (due to a harsh environment) to provide maintenance.

[0020] For devices on-grid (tethered) or with reliable power sources, RF power harvesting can be used as a backup in case the primary power source is lost.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] In the accompanying drawings, the preferred embodiment of the invention and preferred methods of practicing the invention are illustrated in which:

[0022] FIG. **1** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

[0023] FIG. **2** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0024] FIG. **3** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0025] FIG. **4** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0026] FIG. **5** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0027] FIG. **6** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0028] FIG. **7** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0029] FIG. **8** is a block diagram of an RF Power Harvesting block in communication with a Power Storage block.

[0030] FIG. **9** is a block diagram of an RF Power Harvesting block in communication with a Power Storage block and used to supply power to Core Device Components.

[0031] FIG. **10** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0032] FIG. **11** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0033] FIG. **12** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block.

[0034] FIG. **13** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0035] FIG. **14** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

[0036] FIG. **15** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

[0037] FIG. **16** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0038] FIG. **17** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block.

[0039] FIG. **18** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger.

[0040] FIG. **19** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and the RF Power Harvesting block in communication with a Power Storage block.

[0041] FIG. **20** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and Core Device Components.

[0042] FIG. **21** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and Core Device Components and the RF Power Harvesting block in communication with a Power Storage block.

[0043] FIG. **22** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0044] FIG. **23** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0045] FIG. **24** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0046] FIG. **25** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0047] FIG. **26** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0048] FIG. **27** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0049] FIG. **28** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the RF Power Harvesting block supplying power to Core Device Components.

[0050] FIG. **29** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0051] FIG. **30** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

[0052] FIG. **31** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0053] FIG. **32** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power Core Device Components.

[0054] FIG. **33** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or

Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0055] FIG. **34** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

[0056] FIG. **35** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

[0057] FIG. **36** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0058] FIG. **37** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

[0059] FIG. **38** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0060] FIG. **39** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0061] FIG. **40** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to the Power Storage Charger.

[0062] FIG. **41** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0063] FIG. **42** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0064] FIG. **43** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

[0065] FIG. **44** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the RF Power Harvesting block supplying power to Core Device Components and used to supply power to a Power Storage Charger.

[0066] FIG. **45** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

[0067] FIG. **46** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0068] FIG. **47** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

[0069] FIG. **48** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power Core Device Components and used to supply power to a Power Storage Charger.

[0070] FIG. **49** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0071] FIG. **50** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

[0072] FIG. **51** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0073] FIG. **52** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0074] FIG. **53** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

[0075] FIG. **54** is a block diagram of an RF Power Harvesting block using Antenna A to directly supply power to Core Device Components.

[0076] FIG. **55** is a block diagram of an RF Power Harvesting block using Antenna A to supply power to a Power Regulation, Storage and/or Storage Charging block.

[0077] FIG. **56** is a block diagram of an RF Power Harvesting block using Antenna A to supply power to a Power Regulation, Storage and/or Storage Charging block and used to supply power to Core Device Components.

[0078] FIG. **57** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

[0079] FIG. **58** is a block diagram of an RF Power Harvesting block used to supply power to a Power Regulation, Storage and/or Storage Charging block.

[0080] FIG. **59** is a block diagram of an RF Power Harvesting block used to supply power to a Power Regulation, Storage and/or Storage Charging block and used to supply power to Core Device Components.

[0081] FIG. **60** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

[0082] FIG. **61** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0083] FIG. **62** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and used to directly supply power to Core Device Components.

[0084] FIG. **63** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0085] FIG. **64** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

[0086] FIG. **65** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

[0087] FIG. **66** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0088] FIG. **67** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

[0089] FIG. **68** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0090] FIG. **69** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0091] FIG. **70** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0092] FIG. **71** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0093] FIG. **72** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0094] FIG. **73** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

[0095] FIG. **74** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

[0096] FIG. **75** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

[0097] FIG. **76** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0098] FIG. **77** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

[0099] FIG. **78** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0100] FIG. **79** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

[0101] FIG. **80** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources.

[0102] FIG. **81** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

[0103] FIG. **82** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

[0104] FIG. **83** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0105] FIG. **84** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

[0106] FIG. **85** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

[0107] FIG. **86** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0108] FIG. **87** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

[0109] FIG. **88** is a block diagram of an entire power system for the device.

[0110] FIG. **89** is a block diagram of a power harvesting block used to supply power to a core device having a sensor.

[0111] FIG. **90** is a block diagram of a power harvesting block used to supply power to a core device having a computer peripheral.

DETAILED DESCRIPTION

[0112] A complete understanding of the invention will be obtained from the following description when taken in connection with the accompanying drawing figures wherein like reference characters identify like parts throughout.

[0113] There is shown an apparatus **10** for an application. The apparatus **10** comprises a core device **22** preferably having an integrated circuit **36** for the application. The apparatus **10** comprises a power harvester **20** connected to the core device **22** to power the core device **22**.

[0114] The apparatus **10** preferably includes an alternative power source **24** connected to the core device **22** to power the core device **22** in conjunction with the power harvester **20**. Preferably, the apparatus **10** includes a power regulator **26** and/or power storage circuit **28** connected to the power harvester **20**. The apparatus **10** preferably includes a power storage charger **30** connected to the power harvester **20**. Preferably, the apparatus **10** includes a power storage connected to the power harvester **20**.

[0115] Preferably, the core device **22** includes a memory **38** connected to the integrated circuit **36** and to the power harvester **20** to power the memory **36**.

[0116] The core device **22** can include a sensor **32**, as shown in FIG. **89**. The sensor **32** can include a proximity sensor, an intrusion sensor, an environmental sensor, a chemical sensor, a biological sensor, a sensor in contact with an automobile, an occupancy sensor, a motion sensor, a position sensor, a metal detector, or a sensor **32** in contact with an aircraft. The sensor **32** can include an alarm connected to the power harvester **20** to power the alarm, a display connected to the power harvester **20** to power the display, a sensor **32** disposed in a building, an industrial automation sensor, a sensor **32** in contact with an elevator, a temperature sensor, a fire sensor, or an accelerometer.

[0117] The sensor **32** can include a level sensor, a gas level sensor, a fluid level sensor, a light level sensor. The sensor **32** can include a flow sensor, a gas flow sensor, a fluid flow sensor, a light flow sensor, or a plasma flow sensor.

[0118] The sensor **32** can include a pressure sensor, a gas pressure sensor, a fluid pressure sensor, a light sensor, an infrared light sensor, an ultraviolet light sensor, an x-ray sensor, a cosmic ray sensor, a visible light sensor, a gamma ray sensor, a stress sensor, a strain sensor, a depth sensor, or an electrical characteristic sensor.

[0119] The sensor **32** includes a voltage sensor, a current sensor, a viscosity sensor, an acoustical sensor, a sound sensor, a listening sensor, a thickness sensor, a density sensor, a surface quality sensor, a volume sensor, a physical sensor, a mass sensor, a weight sensor, a conductivity sensor, a distance sensor, an orientation sensor, or a vibration sensor.

[0120] The sensor **32** can include a radioactivity sensor, a field strength sensor, an electric field sensor or a magnetic field sensor, a smoke detector, a carbon monoxide detector, a radon detector, an air quality sensor, a humidity sensor, a glass breakage sensor, or a break beam detector. The sensor can include a thermal energy sensor, an electromagnetic sensor, a mechanical sensor, an optical sensor, a radiation sensor, a sensor in contact with a vehicle, or a sensor **32** in contact with a water craft.

[0121] The present invention pertains to an apparatus **10** for an application. The apparatus **10** comprises a core device **22** having an integrated circuit **36** for the application. The apparatus **10** comprises means for receiving energy wirelessly and providing power from the energy to the core device **22** to power the integrated circuit **36** of the core device **22**. The receiving means is connected to the core device **22**. Preferably, the core device **22** includes means for sensing.

[0122] The present invention pertains to a method for an application. The method comprises the steps of converting RF energy into usable energy. There is the step of preferably powering an integrated circuit **36** of the core device **22** with the usable energy.

[0123] Preferably, there is the step of regulating the usable energy provided to the core device **22**. There is preferably the step of storing the usable energy. Preferably, there is the step of providing power to the core device **22** from an alternative power source **24** in conjunction with the usable energy.

[0124] Alternatively, the core device **22** can include a computer peripheral **34**, as shown in FIG. **90**. The computer peripheral **34** can include a handheld game, a gaming system, a game controller, a controller, a keyboard, a mouse, a computer terminal, computer storage, or computer equipment.

[0125] The present invention can be implemented in numerous ways. A number of these ways are depicted in FIGS. **1-90**. These figures contain multiple blocks that are configured in multiple ways.

[0126] In the figures, an arrow represents the flow of power unless otherwise stated. Single-headed (one-way) arrows represent that the power is flowing from one block to another. The single-headed arrow may represent multiple wires that provide power from one block to multiple parts in the other block. Two-headed (two-way) arrows represent a single wire that can have power flow in either direction or multiple wires each having power flow in a single direction.

[0127] As an example, a two-headed arrow between an RF power harvesting **20** block and a power regulation **26** and/or power storage circuit **28** block can represent a single wire that allows harvested power to flow into a storage device such as a capacitor. The same block diagram can also represent two wires between the two blocks with the first wire allowing harvested power to flow into a voltage regulator **26**. The second wire can allow the regulated voltage to feedback to the RF power harvesting **20** block to provide power to internal components such as transistors to increase the performance of the RF power harvesting **20** block.

[0128] Each block is described in detail below. Each block represents the functionality described below associated with it. For instance, the RF power harvesting **20** block describes a power harvester **20**, and the power regulation **26** and/or power storage circuit **28** block describes a power regulator **26** and a power storage circuit **28**.

Device's Power System Block

[0129] The device's power system block includes all components/circuitry required to power the device. This block may include an RF power harvesting **20** block, a power regulation **26** and/or power storage circuit **28** block, a power storage block, and/or a power storage charger **30** block (all discussed below).

RF Power Harvesting **20** Block

[0130] The RF power harvesting **20** block is connected to an antenna. The antenna may or may not be used as the communications antenna for the core device **22** components. The RF power harvesting **20** block is used to convert the energy captured by the antenna into usable power, such as direct current (DC) voltage.

[0131] This block may include antenna matching, rectifying circuitry, voltage transforming circuitry, and/or other performance optimizing circuitry. The rectifying circuitry may include a diode(s), a transistor(s), or some other rectifying device or combination. Examples of the rectifying circuitry include, but are not limited to, half-wave, full-wave, and voltage doubling circuits.

[0132] The output of the RF power harvesting **20** block is a DC voltage or current. The RF power harvesting **20** block may accept feedback (or input) from other circuitry or blocks, which may be used to control the harvesting circuitry to improve the performance or vary the output. This feedback may include, but is not limited to, a DC voltage or a clock from the core device **22** components.

[0133] U.S. Pat. No. 6,615,074 (FIGS. **8**, **9**, **12***a*, **12***b*, **13**, **14**), incorporated by reference, herein, shows numerous examples of RF power harvesting circuits that can be used to implement the block and function described.

Power Regulation **26** and/or Power Storage Circuit **28** Block

[0134] It may be necessary to regulate the converted power (i.e., hold the power at a constant level) for specific devices. The devices that would need this block require a fairly constant voltage or current. Deviations from the required values may cause the device to not perform within its specifications.

[0135] The regulation can be implemented in many different ways. The block can be as simple as using a Zener diode, or as complicated as using an integrated circuit such as a linear voltage regulator **26** or switching regulator **26** to hold the voltage at a constant level.

[0136] Certain devices have a more tolerable power requirement. For these devices, the regulation stage may be excluded.

[0137] This block may also include, with or without the regulation, a storage device such as a capacitor, a battery, or some other device able to store charge. The output from the power regulation **26** and/or power storage circuit **28** block may be used as feedback to other blocks within the device's power system block or to the alternative power sources **24** block (described below), if they require a regulated supply voltage or stored power.

[0138] U.S. Pat. No. 6,894,467 (FIGS. **1**, **3**), Linear Voltage Regulator, incorporated by reference, herein, is an example of a practical application of implementing the regulation described in the block. U.S. Pat. No. 6,297,618 (FIGS. **1-4**), Power storage device and method of measuring voltage of storage battery, incorporated by reference, herein, is an example of a practical application of implementing the storage described in the block.

Power Storage Charger **30** Block

[0139] The power storage charger **30** block may be needed if a power storage component of the device requires a special charging mechanism, such as pulse charging or trickle charging. This block controls how the captured and converted power is supplied to the power storage component.

[0140] U.S. Pat. No. 6,836,095 (FIGS. **1-3**), Battery Charging Method and Apparatus, incorporated by reference herein, is an example of a practical application of implementing the special charging mechanism described in the block.

Power Storage Block

[0141] If a device has intermittent power requirements, it may be necessary to store the captured power for use at a later time. The power can be stored in the power storage block, which could include a battery, a capacitor, and/or another type of power storage component. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other power storage elements.

[0142] The output from the power storage block may be used as feedback to other blocks within the device's power system or to the alternative power sources **24** block, if they require a dedicated and predictable supply voltage.

[0143] U.S. Pat. No. 6,297,618 (FIGS. **1-4**), Power Storage Device and Method of Measuring Voltage of Storage Battery, incorporated by reference herein, is an example of a practical application of implementing the storage described in the block. U.S. Pat. No. 6,835,501, Alkaline Rechargeable Battery, incorporated by reference herein, is also an example of a practical application of implementing the storage described in the block.

Core Device **22** Components Block

[0144] The core device **22** components block is the portion of the device that is receiving power from the device's power system. This block may be, but is not limited to, the devices listed in the subsequent pages of this document. It may be advantageous for the core device **22** components block to communicate with any of the blocks that are supplying power to it. This communication can include, but is not limited to, a feedback control signal such as a clock or an ON/OFF command. As an example, the device may want to turn off the alternative power sources **24** block if it is receiving sufficient power from the RF power harvesting **20** block.

Alternative Power Sources **24** Block

[0145] RF energy harvesting also has the ability to be augmented by other types of power harvesting, storage components, or dedicated sources (e.g. power line). The alternative power sources **24** block shows how this type of system could be implemented. The augmenting power harvesting technologies include, but are not limited to, solar, light (visible and non-visible), piezoelectric, vibration, acoustic, thermal, microgenerators, wind, and other environmental elements. This block can work independently or have communication with other blocks.

[0146] U.S. Pat. No. 6,784,358, Solar Cell Structure Utilizing an Amorphous Silicon Discrete By-Pass Diode, incorporated by reference herein, is an example of a practical application of implementing an alternative power source **24** described by the block. U.S. Pat. No. 6,858,970, Multi-Frequency Piezoelectric Energy Harvester, incorporated by reference herein, is also an example of a practical application of implementing an alternative power source **24** described by the block.

Power Regulation **26**, Storage **28** and/or Storage Charging **30** Block

[0147] The power regulation **26**, storage **28** and/or storage charging **30** block contains all the combinations of the power regulation **26** and/or power storage circuit **28** block, power storage charger **30** block, and power storage block. This block is used in the later figures to reduce the number of figures needed to show how the blocks can interconnect.

[0148] The disclosed invention is an apparatus and method for an application for retrieving radio frequency (RF) energy by an antenna, converting that energy into direct current (DC) power, regulating that energy using an optimized circuit, storing that energy in an optimized component, and/or supplying the power for a specific device.

Retrieval of RF Energy

[0149] The RF energy is retrieved from the environment by the use of the antenna. The antenna can be shared or standalone with respect to an antenna used for the device's wireless communication. FIGS. **54-56** show a device that has an antenna A for use by the RF power harvester **20** and an antenna B used for wireless one-way or two-way communication. FIGS. **57-59** show a device where the antenna is shared by both a device's communication module (within the core device **22** components) and the RF power harvester **20**. In terms of form factor, the antenna used by the apparatus **10** can be a separate component or integrated directly into the form factor of the device.

[0150] The antenna is able to capture two types of available RF energy. The first type of energy exists as ambient RF energy. This type of RF surrounds us in our day-to-day lives and is usually generated to carry one way or two-way combinations of voice, video and data communications. The sources that the antenna can harvest from include, but are not limited to, medium-frequency AM radio broadcast, very-high-frequency (VHF) FM radio broadcast and television broadcast, ultra-high-frequency (UHF) broadcast, cellular base stations, wireless data access points, super-high-frequency (SHF) frequencies, and the industrial, scientific, and medical (ISM) bands. These sources cover transmission frequencies from 300 kHz to 30 GHz.

[0151] The second type of energy available is directed RF energy. This type of RF energy is directed from a transmitter specifically designed to deliver RF energy for harvesting by the antenna. The transmitter can be configured as a standalone device or integrated into an existing device.

Conversion of the Energy into DC

[0152] The RF energy captured by the antenna must be converted into a useful form of energy for the specific device. This conversion is shown in block form in all FIGS. (**1-90**) as the RF power harvesting **20** block. The most common form of useable energy is DC energy. To perform this conversion, the block includes circuitry to rectify the captured alternating current (AC) energy to create DC energy. The rectification in this block can be done with a diode(s), a transistor(s), or some other rectifying device or combination.

Regulation of the Energy

[0153] It may be necessary to regulate the converted power (hold the power at a constant level) for specific devices. FIGS. **2-7**, **10-17**, and **22-53** show how this regulation can be implemented using a power regulation **26** and/or power storage circuit **28** block. The devices that would need this block require a fairly constant voltage or current. Deviations from the required values may cause the device to not perform within its specifications. The regulation can be implemented in many different ways. The block can be as simple as using a Zener diode, or as complicated as using an integrated circuit such as a linear voltage regulator **26** or a switching regulator **26** to hold the voltage at a constant level. Certain devices have a more tolerable power requirement. For these devices, the regulation stage may be excluded.

Storage of the Energy

[0154] If a device has intermittent power requirements, such as the devices exampled by FIGS. **2-53**, it may be necessary to store the captured power for use at a later time. The power can be stored in the power storage block or the power regulation **26** and/or power storage circuit **28** block. Storage devices can include, but are not limited to, a battery, a capacitor, or another type of power storage component. In certain applications, it may be necessary to include additional circuitry that controls how the power is transferred to the storage device. The power storage charger **30** block is shown in FIGS. **18-53**. This may be needed if the power storage component requires a special charging mechanism such as pulse charging or trickle charging. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other storage elements. There are devices that will not require storage. These devices can run directly off of the converted power. These devices also may or may not require regulation of the captured power.

Supplying the Power

[0155] The captured DC power, which may or may not be regulated and/or stored, is supplied to the device, which is represented by the core device **22** components block in the figures. This may be a single connection or it may supply multiple parts of the device with power.

[0156] RF energy harvesting also has the ability to be augmented by other types of power harvesting or storage components. Other power harvesting technologies include, but are not limited to, solar, light (visible and non-visible), piezoelectric, vibration, acoustic, thermal, microgenerators, wind, and other environmental elements. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other storage elements. FIGS. **60-88** show how the alternative power sources **24** block can be connected to an RF energy harvesting system. These figure show how the RF energy harvesting **20** block and the alternative power sources **24** block can work independently or have communication with each other. The antenna configurations shown in FIGS. **54-59** are still applicable with the addition of an alternative power sources **24** block. These antenna configurations can be applied to FIGS. **60-88**.

[0157] RF energy harvesting also has the ability to provide a backup to devices on-grid (tethered) or with reliable power sources, which can be used in case the primary power source is lost. As an example, it may be mandated by regulations that a sensor has auxiliary power in case the primary supply is lost. It could be possible to use a rechargeable battery that obtains its charge from the primary supply when in operation. However, if the primary supply is lost for a time greater than the life of the rechargeable battery, the specification of uninterrupted power is not met. RF energy could be used to supply power to the described device while the primary supply is not available. The primary supply could include, but is not limited to, an on-grid connection, a generator, a battery, or other reliable power supply.

[0158] RF energy harvesting with or without alternative source augmentation is applicable to provide electric power directly or indirectly to a range of electronic components contained in any specific electrical or electronic device and includes, but is not limited to:

[0159] Passive electronic components, active electronic components;

[0160] Resistors, fixed resistors, variable resistors, thermistors, thyristor, thermocouple;

[0161] Capacitors, Electrolytic Capacitors, Tantalum Capacitors, Ceramic Capacitors, Multilayer Ceramic Capacitors, Polystyrene Film Capacitors, Electric Double Layer Capacitors (Super Capacitors), Polyester Film Capacitors, Polypropylene Capacitors, Mica Capacitors, Metallized Polyester Film Capacitors, Variable Capacitors;

[0162] Diodes, Voltage regulation diodes, light-emitting diodes, organic light-emitting diodes, Variable capacitance diodes, Rectification diodes, Switching diodes, Regulation Diodes, Diode bridges, Schottky barrier diodes, tunnel diodes, PIN diodes, Zener diodes, Avalanche diodes, TVSs;

[0163] Integrated circuits, microcontroller unit (MCU), microprocessor unit (MPU), logic circuits, memory, printed circuits, circuit boards, printed wiring boards;

[0164] Transistors, MOSFETs, FETs, BJTs, JFETs, IGBTs, Relays, Antennas, semiconductors, conductors, inductors, relays, diacs, triacs, SCRs, MOVs;

[0165] Fuses, circuit breakers;

[0166] Batteries, Non-rechargeable batteries, rechargeable batteries, coin cell batteries, button cell batteries, alkaline batteries, lithium batteries, lithium ion batteries, lithium polymer batteries, NIMH batteries, NICAD batteries, Lead acid batteries, Zinc air batteries, Manganese Lithium batteries, Niobium Titanium Lithium batteries, Vanadium Pentoxide Lithium batteries, Carbon

Zinc batteries, Zinc Chloride batteries, Lithium Thionyl Chloride batteries, Manganese Dioxide batteries, Lithium Poly-Carbonmonofluoride batteries, Lithium Manganese Dioxide batteries, Lithium Chloride batteries, Lead Acid Calcium batteries, Lead Acid Tin batteries, Oxy Nickel batteries, Silver Oxide batteries, Magnesium batteries;

[0167] Inductors, Coils, High Frequency Coils, Toroidal Coils, Transformers, switches, chokes;

[0168] Motors, DC motors, stepper motor, AC motors, Fans;

[0169] Crystals, Oscillators, Clocks, Timers; and

[0170] Displays, LCDs, LED displays.

[0171] RF energy harvesting with or without alternative source augmentation is applicable across a range of markets and specific devices and includes but is not limited to:

[0172] Consumer electronics:

[0173] Electronic equipment, wired devices, battery powered devices, wireless communication devices, cell phones, telephones, phones, cordless phones, portable phones, Bluetooth devices, Bluetooth headsets, hands-free headsets, headsets, headphones, Wireless headsets, radios, AM/FM radios, shortwave radios, weather radios, Two-way radios, portable radio, lights, lanterns, portable lights, flashlights, nightlights, spotlights, search lights, calculators, graphing calculators, desk calculators, clocks, alarm clocks, wall clocks, desk clocks, travel clocks, watches, wristwatches, pocket watches, stop watches, timers, voice recorders, Dictaphones, laser pointers, power tools, cordless power tools, electronic razors, electric razors, handheld games, gaming systems, game controllers, wireless game controllers, remote controls, battery chargers, computers, portable computers, keyless entry, toys, toy guns, toy laser guns, games, microphones, musical instruments, musical effects processors, musical instrument tuners, metronomes, electronic chord charts, door openers, garage door openers, PDA, Cameras, Video recorders, Multimeter, electronic test equipment, hand-held electronics, portable electronics, wireless pens, sound generators, noise generators, language translators, electric toothbrushes, portable televisions, pagers, transceivers, toy vehicles, remote control vehicles, toy planes, remote control planes, pet containment systems, invisible fence pet sensors, memory backup, base station battery backups, appliance battery backups, uninterrupted power supplies, GPS devices, memory retention power supplies, metal detectors, stud finders, metal stud finders, stun guns, tazers, wearable devices, baby monitors, intercoms, doorbells, wireless doorbells, electronic office supplies, electronic staplers, radar jammers, radar detectors, digital scales, microfilm cassettes, video head testers, compasses, noise canceling headphones, air samplers, depth finders, barometers, weather measurement instruments, data transfer devices, automatic distress signaling unit, Wireless audio speakers, Satellite radios, Police scanners, Car navigation systems (GPS devices), Decorative lights, Christmas lights, garden lights, lawn lights, ornamental lights, porch lights;

[0174] Multi-media players: MP3, DVD, analog music players, CD players, tape players, digital music players, digital video players, minidisc; and

[0175] Computer: keyboards, mice, peripherals, computer equipment, electronic computers, computer storage, computer terminals;

[0176] Building/Industrial Automation:

[0177] Sensors: Position, elevator, temperature, fire, accelerometers, level, gas level, fluid level, light level, flow, gas flow, fluid flow, light flow, plasma flow, pressure, gas pressure, fluid pressure, motion, light, infrared light, ultraviolet light, X-rays, cosmic rays, visible light, gamma rays, chemical, stress, strain, depth, electrical characteristics, voltage, current, viscosity, acoustical, sound, listening, thickness, density, surface quality, volume, physical, mass, weight, force, conductivity, distance, orientation, vibration, radioactivity, field strength, electric field strength, magnetic field strength, occupancy, smoke detector, carbon monoxide detector, radon detector, air quality, humidity, glass breakage, break beam detector;

[0178] Controls: Position, elevator, temperature, fire, accelerometers, level, gas level, fluid level, light level, flow, gas flow, fluid flow, light flow, plasma flow, pressure, gas pressure, fluid pressure, motion, light, infrared light, ultraviolet light, X-rays, cosmic rays, visible light, gamma rays, chemical, stress, strain, depth, electrical characteristics, voltage, current, viscosity, acoustical, sound, listening, thickness, density, surface quality, volume, physical, mass, weight, force, conductivity, distance, orientation, vibration, radioactivity, field strength, electric field strength, magnetic field strength, occupancy, smoke detector, carbon monoxide detector, radon detector, air quality, humidity; and

[0179] Devices: Thermostats, light switches, door locks, smart-card door locks, lighting, emergency lighting, motion lighting, safety lighting, highway lighting, construction lighting, sign lighting, roadway sign lighting, construction sign lighting, automatic flushing units, automatic soap dispenser, automatic paper towel dispenser, automatic faucets, automatic door sensors, identification reader, fingerprint reader, credit card readers, card readers, valve actuators, gauges, analog gauges, digital gauges, fire extinguishers, wireless switches, remotely operated inspection equipment, gas/oil pipeline monitoring systems, robotic pipeline inspection gauges, "auto-reclosers" for electric power lines, sonar buoys, telemetry systems, electronic record tracking systems, robbery tracking devices, interrogators, programmers, emergency exit alarms, alarms, flood alarms, gas alarms, electronic entry systems, security keypads, silo transducers, data recorders, signal tracers, antistatic strap testers, radiosonde weather balloons, utilities load controllers, profilometers, noise cancellation equipment, infrared beacons;

[0180] Military/Government:

[0181] Tracking tags: Weapons, vehicle, soldier, gear/assets, staff, general population, security access badges;

[0182] Sensors: Proximity, intrusion, environmental, chemical/biological; and

[0183] Equipment: Battery charger, surveillance, card readers, identification reader, fingerprint reader, retinal scanners, satellites, rockets, space vehicles, search and rescue transponders (SARTs), emergency position-indicating rescue beacons (EPIRBs), emergency locator transmitters (ELTs), military radios, electronic toll collection systems, postal tracking systems, communica-

tions, thermal imaging, night vision, training targets, field medical equipment, house arrest monitors, laser tags, electronic parking meters, multiple integrated laser engagement system, munitions and mines, ship sensors;

[0184] Utility:

[0185] Gas consumption meters, water consumption meters, and electric consumption meters;

[0186] Logistics & Supply Chain Management:

[0187] Radio-frequency identification devices (RFID), RFID readers;

[0188] Tracking: Asset tags, cargo container location beacons, transponders, transceivers; and

[0189] Devices: Smart price tags, smart shelving, hand-held barcode scanner, barcode scanners, credit card readers, card readers, retail signage, hotel door locks;

[0190] Homeland Security:

[0191] Sensors: Occupancy, proximity, environmental, chemical/biological, motion, position; and

[0192] Metal detector wand;

[0193] Medical:

[0194] Implantable: cochlear implants, neural stimulators, pace makers, medication administration, defibrillator;

[0195] Body function monitors: pressure, temperature, respiration, blood oxygenation, insulin, hearing aid, pulse, EKG, heart, Holter;

[0196] Tracking tags: Patient, baby identification, assets, supplies, staff, medication, instruments;

[0197] Devices: Home healthcare equipment, ambulatory infusion pumps, blood analyzers, biofeedback systems, bone growth stimulators, thermometers, digital thermometers, stimulators, galvanic stimulators, muscle stimulators, pediatric scales; and

[0198] Agriculture—livestock tracking and asset tracking:

[0199] Tracking: livestock, asset, wildlife tracking devices; and

[0200] Equipment: cattle prods;

[0201] Automotive:

[0202] Automotive antennas, Automotive Audio Systems, Automotive Lighting, Automotive Video Systems, Computers, Processors, Controls, Switches, Electric Motors, Actuators, Ignition Systems, Starter Systems, Injection Systems, Powertrain Electronics, Radar Detectors, Proximity Detectors, Safety Systems, Security Systems, Sensors, Regulators, Distributors, Vehicle Control, Wiper Systems, Washer Systems, Radio, Video Systems, Entertainment Systems, Navigation Systems, GPS systems, Power Mirror Systems, Emission control systems;

[0203] Appliances:

[0204] Monitoring systems and control systems for major and small appliances including washing machines, dryers, refrigerators, freezers, coolers, air conditioners, humidifiers, dehumidifiers, air purifiers, air filters, fans, furnaces, water heaters, boilers, space heaters, sowing machines, ice makers, microwave ovens, convection ovens, ovens, toaster ovens, ranges, range hoods, cooktops, stoves, stovetops, crock pots, hot plates, dishwashers, garbage disposals, can openers, vacuum cleaners, blenders, mixers, food processors, irons, coffee makers, toasters, grills, hair dryers, electric tooth brushes, electric razors, electric drills, electric screwdrivers, chainsaws, lawnmowers, push mowers, riding mowers, trimmers, brush cutters, pruners, edgers, vending machines;

[0205] Ventilation, Heating, Air-Conditioning, and Commercial Refrigeration Equipment:

[0206] Monitoring systems, control systems;

[0207] Engine, Turbine, and Power Transmission Equipment;

[0208] Monitoring systems, control systems;

[0209] Other General Purpose Machinery Manufacturing:

[0210] Monitoring systems, control systems;

[0211] Telecommunications:

[0212] Monitoring systems, control systems;

[0213] Portable;

[0214] Aircraft:

[0215] Monitoring systems, control systems, actuator systems, sensors.

[0216] It should be noted that devices within a specific category may be applicable across multiple areas even if they are not specifically listed. (e.g., temperature sensors apply to Industrial and Building Automation and Appliances).

[0217] To retrofit or redesign the devices listed, it is possible to implement the described systems in numerous ways. It may be advantageous to leave the device design as-is including the existing power supply. As an example, a device may use non-rechargeable batteries to operate. The device will most likely have a protection circuit to prevent damage if the batteries are installed incorrectly. The protection mechanism is commonly a diode inline with the positive terminal of the battery. In this case, the RF power harvester **20** with or without the alternative power source **24** could be inserted, with an antenna, into the device. The power generated by the RF power harvester **20** (and the alternative power source **24**, if applicable) could be connected to the device after the protection mechanism described to avoid potential charging of a non-rechargeable battery.

[0218] Another way to configure the device's power system is to replace the non-rechargeable batteries with rechargeable batteries. In this instance, the output from the RF power harvester **20** (and alternative power source **24**) could be connected to either side of the protections device. If the connection is before the protection mechanism, the RF power harvester **20** (and the alternative power source **24**, if applicable) will recharge the battery and supply power to the device. If the connection is after the protection mechanism, the RF power harvester **20** (and the alternative power source **24**, if applicable) will supply power to the device and the battery will supply any extra power needed that could not be supplied by the RF power harvester **20** (and the alternative power source **24**, if applicable). It should be noted that the protection device in this case is not needed for proper operation. Its only function would be to protect the batteries from being installed incorrectly. An antenna could be contained inside or placed on the outside of the device.

[0219] Another configuration of the device's power system is to remove the existing batteries and install the RF power harvester **20** (and the alternative power source **24**) in the enclosure provided for the batteries. An antenna could be contained inside or placed on the outside of the device.

[0220] Yet another method of configuring the device's power system would be to reduce the number of batteries and replace them with the RF power harvester **20** (and the alternative power source **24**). In this case, the output from the system would be connected to the batteries in series or par-

allel depending on the original battery configuration. An antenna could be contained inside or placed on the outside of the device.

[0221] An additional option, would be to completely redesign the product and integrate the required circuitry and storage components into the device. This method is probably the most advantageous because it can fully take advantage of the benefits offered by the RF power harvester **20** (and the alternative power source **24**). An antenna could be contained inside or placed on the outside of the device.

[0222] If the RF power harvester **20** (and the alternative power source **24**) is used as a backup to the primary power supply, a switch could be implemented into the system in order to switch the RF power harvester **20** (and the alternative power source **24**) on when the primary source is lost. In this case, an antenna could be contained inside or placed on the outside of the device.

[0223] To show the flexibility of RF energy harvesting, several products were retrofitted to include an RF power harvester **20** (energy harvesting circuitry). These products include a wireless keyboard, a wall clock, and a desk calculator.

[0224] The wireless keyboard is an example of recharging and augmenting a battery to supply power to a device. This system is shown in FIG. **13**. The output from the regulation circuitry recharges the battery and supplies power to the keyboard. The battery is also used to supply power to the keyboard. The keyboard also includes a separate antenna for receiving power and for data communications. The antenna configuration can be seen in FIG. **55**.

[0225] The wall clock is an example of a direct powering system. The wall clock was retrofitted to include an RF power harvester **20** and the internal AA battery was removed. This system is shown in FIG. **2**. The wall clock did not need regulation, but did require a capacitor for storage to supply the pulse of power to move the second hand.

[0226] The calculator is an example of using RF energy harvesting with another energy harvesting technology. The calculator had an internal 1.5V coin cell battery and a small solar panel. The internal battery was removed, however, the solar panel was left intact. This system is shown in FIG. **60**. In this system, the calculator can receive power from both the solar panel and the RF power harvester **20** to eliminate the need for a battery.

[0227] As an additional example, an RF power harvester **20** (energy harvesting circuit) similar to the ones shown in U.S. Pat. No. 6,615,074 (FIGS. **8**, **9**, **12***a*, **12***b*, **13**, **14**), was connected in series with a 0.5V solar cell. Individually, the solar cell was able to provide 0.480V to a 10 kilo-ohm resistor, which was being used to simulate the core device **22** components. This corresponds to 23 microwatts. The RF power harvesting circuit by itself was able to provide 2.093V across the 10 kilo-ohm resistor when being supplied by 1 milliwatt of RF power. This corresponds to 438 microwatts. The two circuit outputs were then combined in series by connecting the output from the RF energy harvesting circuit to the ground of the solar cell. The output of the solar cell was then connected to the resistor. The other end of the resistor was connected to the ground of the RF energy harvesting circuit. The voltage across the resistor with the circuits connected as shown in FIG. **63** was 2.445V. This corresponds to 598 microwatts.

[0228] As can be seen, the combination of the two technologies produces a result higher than the addition of the individual powers. From this, it can be determined that the two technologies can cooperate in a way that produces favorable results.

[0229] In the example given, the solar cell produces current to supply the load and helps to bias the RF rectifying diodes, which allows the RF energy harvesting circuit to operate a higher efficiency. The solar cell also changes the impedance seen by the RF energy harvesting circuit, which produces a beneficial result. To be more specific, when examining the power output of the individual circuit (solar and RF power harvesting), the sum of the power captured by the individual circuits was 23 μW+438 μW=461 μW. However, when the two circuits are combined and are allowed to work in conjunction with one another, the output power becomes 598 μW. This result shows that combining the two power-harvesting technologies produces a 30 percent increase in the output power for this example. This same technique can be applied to multiple energy harvesting technologies to produce even greater output power. The equations for this example are shown below.

[0230] Individual Circuits

$$P_I=P_1+P_2+\ldots+P_N$$

[0231] Combined Circuits

$$P_C>P_I=P_1+P_2+\ldots+P_N$$

[0232] where P_I is the sum of the individual output powers

[0233] P_C is the output of the combined circuit

[0234] P_1 is the output power from the first power harvesting technology

[0235] P_2 is the output power from the second power harvesting technology

[0236] P_N is the output power from the N^{th} power harvesting technology

[0237] N is the number of power harvesting technologies or circuits.

[0238] As can be seen by the previous examples, RF energy harvesting can be used alone or in conjunction with alternative power sources to power a wide range of devices. The addition of RF energy harvesting technology to the device allows for increased battery life, increased functionality, or the removal of the primary battery.

[0239] Although the invention has been described in detail in the foregoing embodiments for the purpose of illustration, it is to be understood that such detail is solely for that purpose and that variations can be made therein by those skilled in the art without departing from the spirit and scope of the invention except as it may be described by the following claims.

What is claimed is:

1. An apparatus, comprising:

a first antenna configured to wirelessly receive energy from a source, the source being remote from the apparatus;

a second antenna different from and electrically isolated from the first antenna, the second antenna configured to wirelessly communicate with the source;

a power harvester electrically coupled to the first antenna and configured to convert the energy received at the first antenna to a DC voltage;

a radio-frequency identification (RFID) device electrically coupled to the power harvester and to the second antenna, the RFID device configured to wirelessly communicate with the source via the second antenna;

an integrated circuit configured to operate based on the DC voltage; and a sensor electrically coupled to the integrated circuit, the sensor configured to operate based on the DC voltage and to detect at least one of an environmental condition associated with a location of the apparatus or a physical condition associated with an item coupled to the apparatus.

2. The apparatus of claim **1**, further comprising:

a regulator coupled to the power harvester and configured to regulate the DC voltage to produce a regulated DC voltage, the sensor configured to operate based on the regulated DC voltage received from the regulator.

3. The apparatus of claim **1**, wherein the integrated circuit is configured to perform an application associated with the at least one of the environmental condition or the physical condition detected by the sensor.

4. The apparatus of claim **1**, further comprising:

an electronic display configured to operate based on the DC voltage.

5. The apparatus of claim **1**, wherein the sensor includes at least one of a temperature sensor, a motion sensor, a vibration sensor, a humidity sensor, an orientation sensor, or a position sensor.

6. The apparatus of claim **1**, wherein the RFID device does not operate based on a battery, and the sensor does not operate based on a battery.

7. The apparatus of claim **1**, wherein the apparatus is a RFID tag.

8. An apparatus, comprising:

an antenna configured to wirelessly receive energy from a source, the source being remote from the apparatus;

a power harvester electrically coupled to the antenna and configured to convert the energy received at the antenna to a DC voltage;

a radio-frequency identification (RFID) device electrically coupled to the antenna and configured to wirelessly communicate with the source via the antenna, the RFID device operatively coupled to the power harvester and including an integrated circuit configured to operate based on the DC voltage; and

a sensor electrically coupled to the RFID device and the power harvester, the sensor configured to operate based on the DC voltage and to detect at least one of an environmental condition associated with a location of the apparatus or a physical condition associated with an item coupled to the apparatus,

the integrated circuit configured to perform an application associated with the at least one of the environmental condition or the physical condition detected by the sensor.

9. The apparatus of claim **8**, wherein the power harvester is configured to wirelessly communicate with the source via the antenna.

10. The apparatus of claim **8**, wherein the sensor is electrically coupled to the antenna and is configured to wirelessly communicate with the source via the antenna.

11. The apparatus of claim **8**, further comprising:

a regulator coupled to the power harvester and configured to regulate the DC voltage to produce a regulated DC voltage, the integrated circuit configured to operate based on the regulated DC voltage received from the regulator.

12. The apparatus of claim **8**, further comprising:

an electronic display configured to operate based on the DC voltage.

13. The apparatus of claim **8**, wherein the sensor includes at least one of a temperature sensor, a motion sensor, a vibration sensor, a humidity sensor, or a position sensor.

14. The apparatus of claim **8**, wherein the RFID device does not operate based on a battery, and the sensor does not operate based on a battery.

15. An apparatus, comprising:

an antenna configured to wirelessly receive energy from a source, the source being remote from the apparatus;

a power harvester electrically coupled to the antenna and configured to convert the energy received at the antenna to a DC voltage to produce a power associated with the DC voltage;

a regulator coupled to the power harvester and configured to regulate the power associated with the DC voltage to produce a regulated power;

a radio-frequency identification (RFID) device electrically coupled to the antenna and configured to wirelessly communicate with the source via the antenna, the RFID device electrically coupled to the power harvester and the regulator, the RFID device including an integrated circuit configured to operate based on at least one of the power received from the power harvester or the regulated power received from the regulator; and

an electronic display electrically coupled to the RFID device and configured to operate, at least in part, based on the regulated power received from the regulator.

16. The apparatus of claim **15**, wherein the electronic display is electrically coupled to the regulator, the electronic display is configured to operate based on the regulated power received from the regulator and the power received from the power harvester.

17. The apparatus of claim **15**, wherein the electronic display is one of a liquid-crystal display (LCD) or a light-emitting diode (LED) display.

18. The apparatus of claim **15**, further comprising:

a sensor configured to operate, at least in part, based on the regulated power received from the regulator, the sensor including at least one of a temperature sensor, a motion sensor, a vibration sensor, a humidity sensor, or a position sensor.

19. The apparatus of claim **15**, wherein the power harvester is configured to wirelessly communicate with the source via the antenna.

20. The apparatus of claim **15**, wherein the RFID device does not operate based on a battery, and the electronic display does not operate based on a battery.

* * * * *

US 20150381080A1

(19) **United States**
(12) **Patent Application Publication**
Virella
(10) **Pub. No.: US 2015/0381080 A1**
(43) **Pub. Date: Dec. 31, 2015**

(54) **ELECTROMAGNETIC POWER GENERATOR**

(71) Applicant: **Eddie Oquendo Virella**, Norwich, CT (US)

(72) Inventor: **Eddie Oquendo Virella**, Norwich, CT (US)

(21) Appl. No.: **14/753,755**

(22) Filed: **Jun. 29, 2015**

Related U.S. Application Data

(60) Provisional application No. 62/018,826, filed on Jun. 30, 2014.

Publication Classification

(51) **Int. Cl.**

H02N 11/00	(2006.01)
H02J 7/00	(2006.01)
H02K 9/00	(2006.01)
H02J 3/38	(2006.01)

(52) **U.S. Cl.**
CPC ***H02N 11/002*** (2013.01); ***H02J 3/382*** (2013.01); ***H02J 7/0068*** (2013.01); ***H02K 9/005*** (2013.01)

(57) **ABSTRACT**

An electromagnetic generator comprising an antenna that receives radiofrequency energy. The antenna is connected to a rectifier circuit, which is used to charge a first battery using the radiofrequency energy received by the antenna. The first battery supplies power to a control unit, which powers at least one electromagnet to generate a magnetic field. A flywheel having at least one magnet is configured to rotate when the magnetic field is generated. An alternator, connected to the flywheel, charges a second battery based on the rotation of the flywheel. The second battery is then used to supply power to an external device.

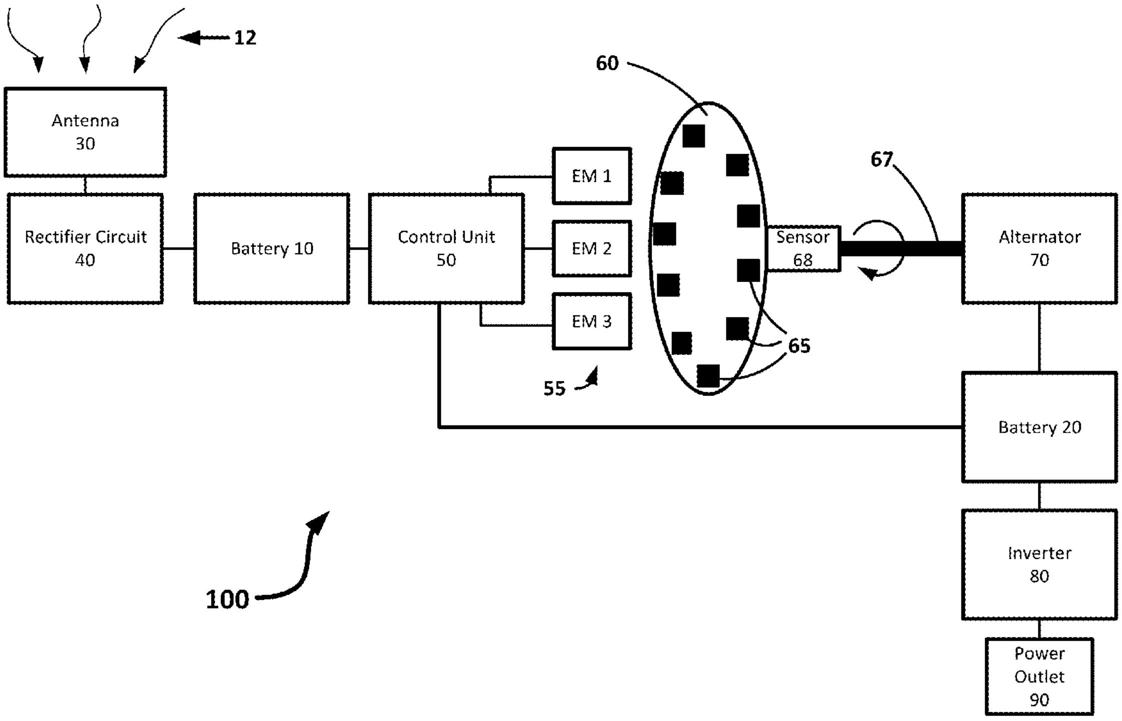
12
Antenna
30
60
67
Rectifier Circuit
40
Battery 10
Control Unit
50
EM 1
EM 2
EM 3
Sensor
68
Alternator
70
65
55
Battery 20
Inverter
80
Power
Outlet
90
100

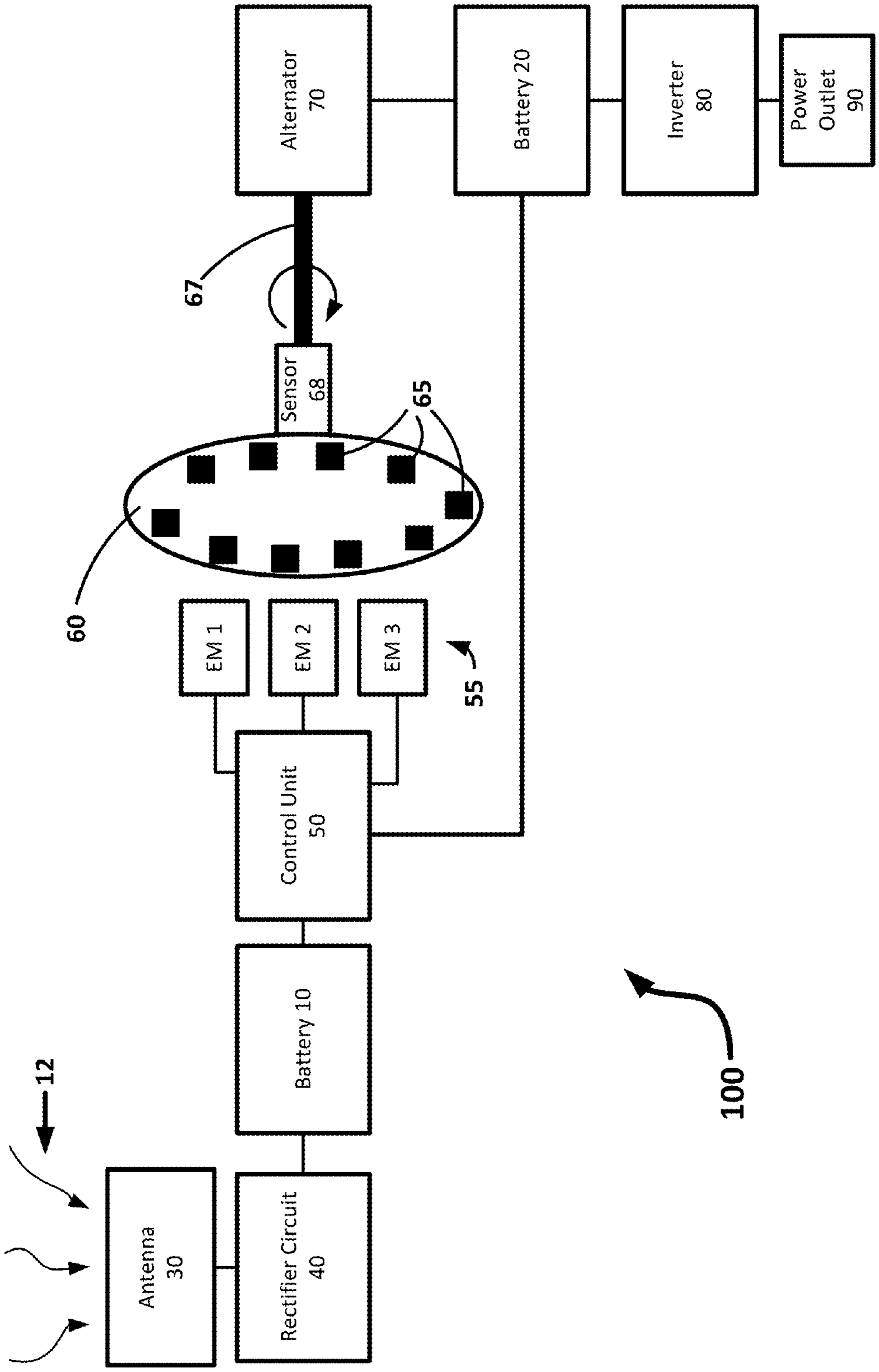
12
Antenna
30
Rectifier Circuit
40
Battery 10
Control Unit
50
EM 1
EM 2
EM 3
55
60
65
Sensor
68
67
Alternator
70
Battery 20
Inverter
80
Power
Outlet
90
100

FIG. 1

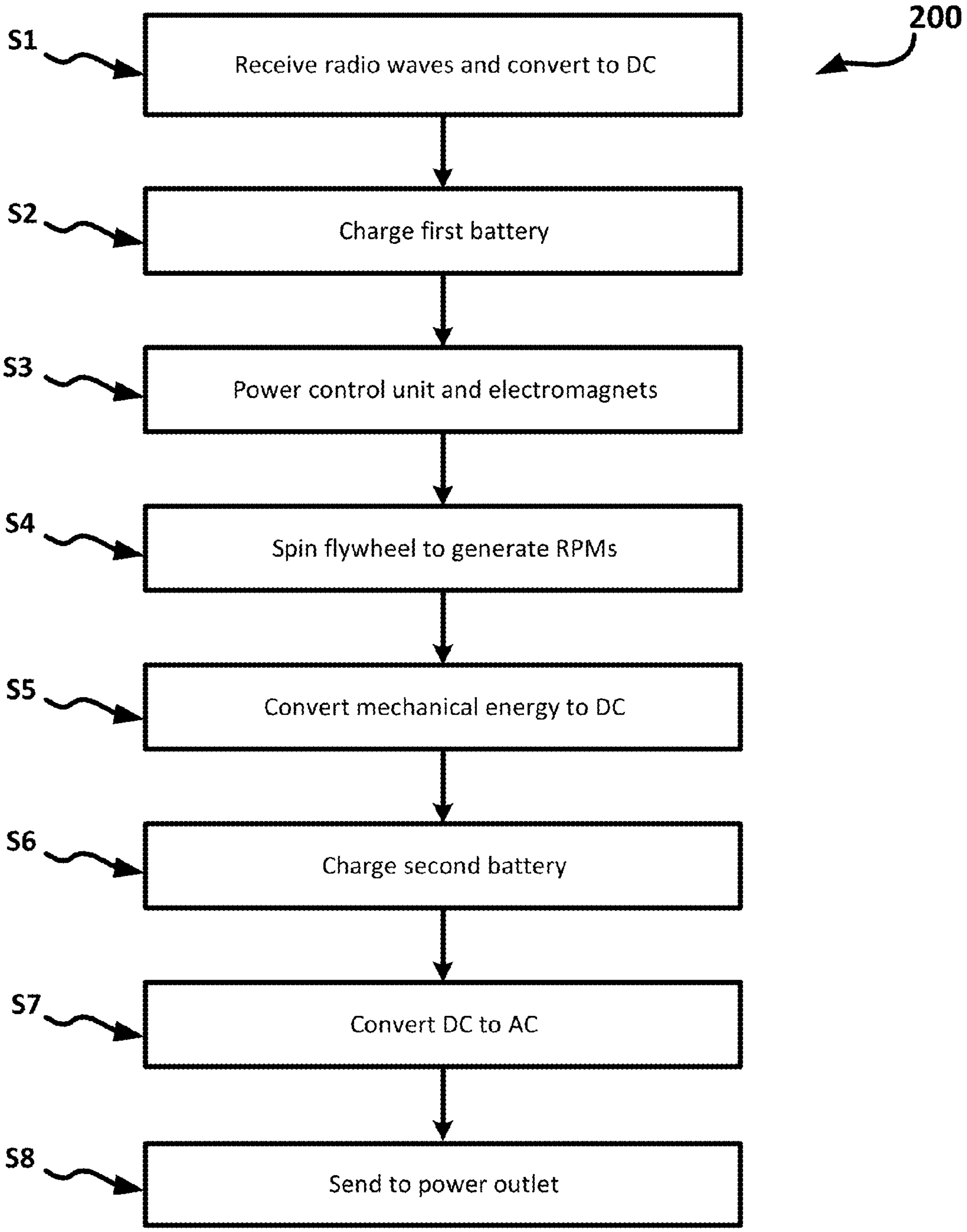
200
S1
Receive radio waves and convert to DC
S2
Charge first battery
S3
Power control unit and electromagnets
S4
Spin flywheel to generate RPMs
S5
Convert mechanical energy to DC
S6
Charge second battery
S7
Convert DC to AC
S8
Send to power outlet

FIG. 2

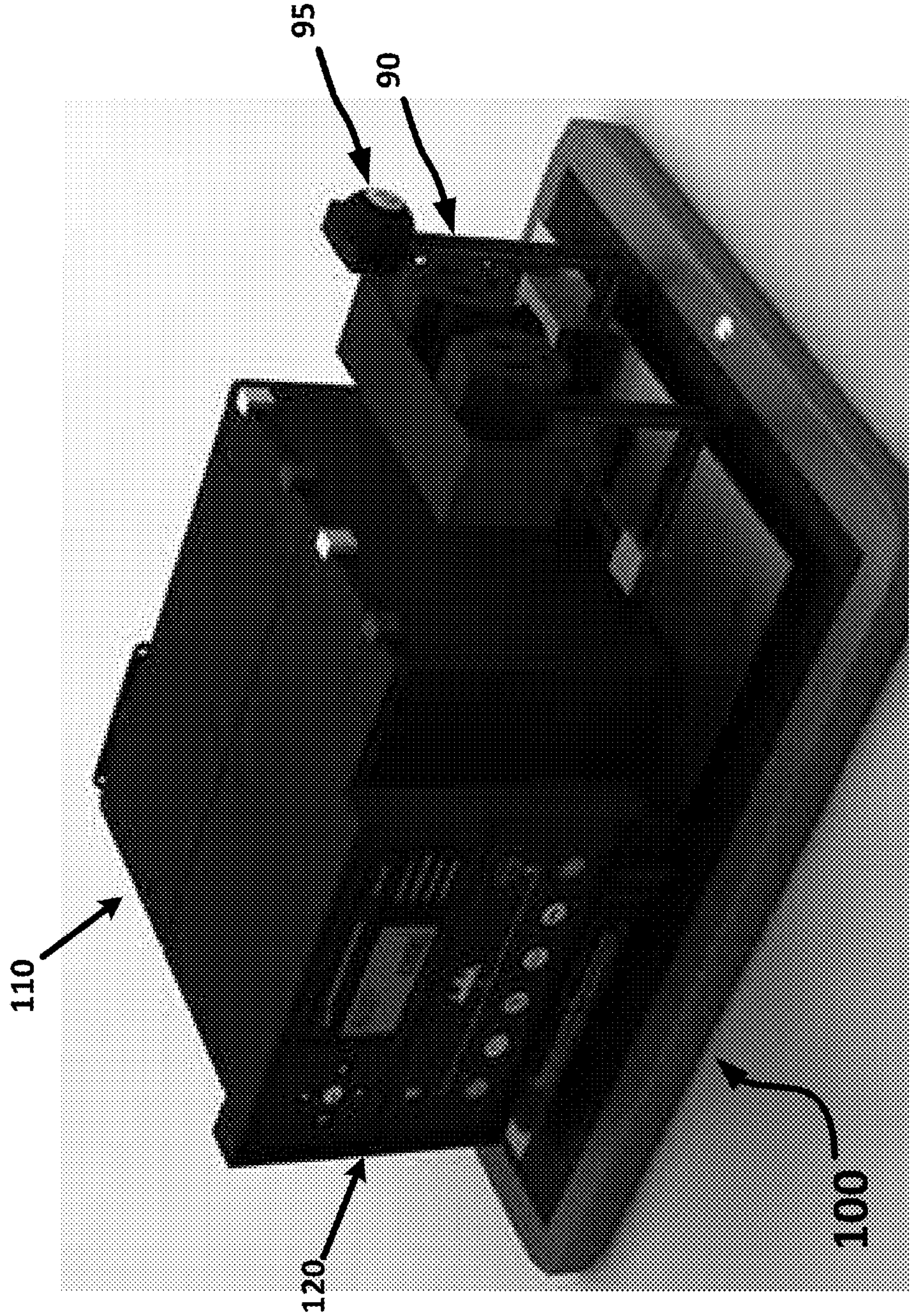
95
90
110
120
100

FIG. 3

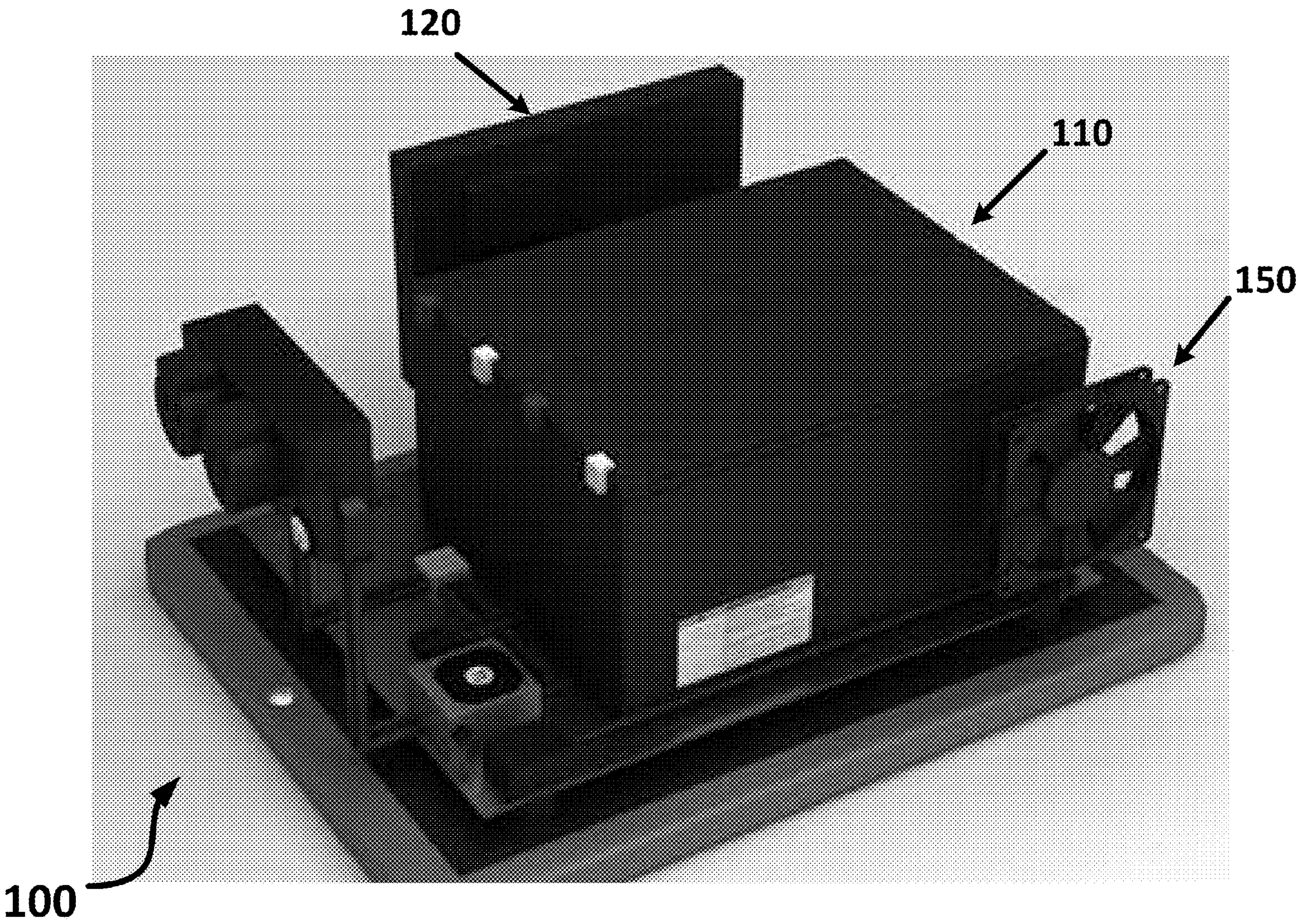
120
110
150
100

FIG. 4

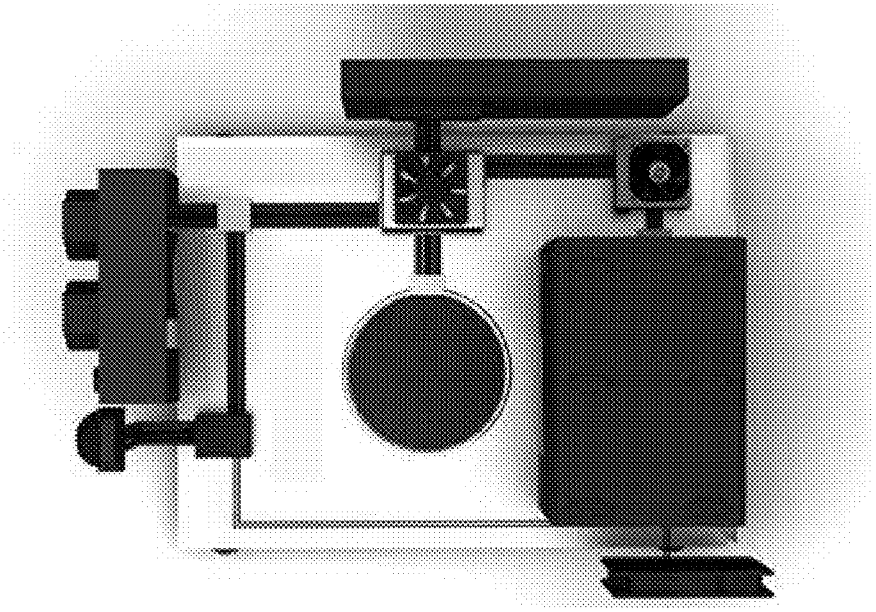

FIG. 5A

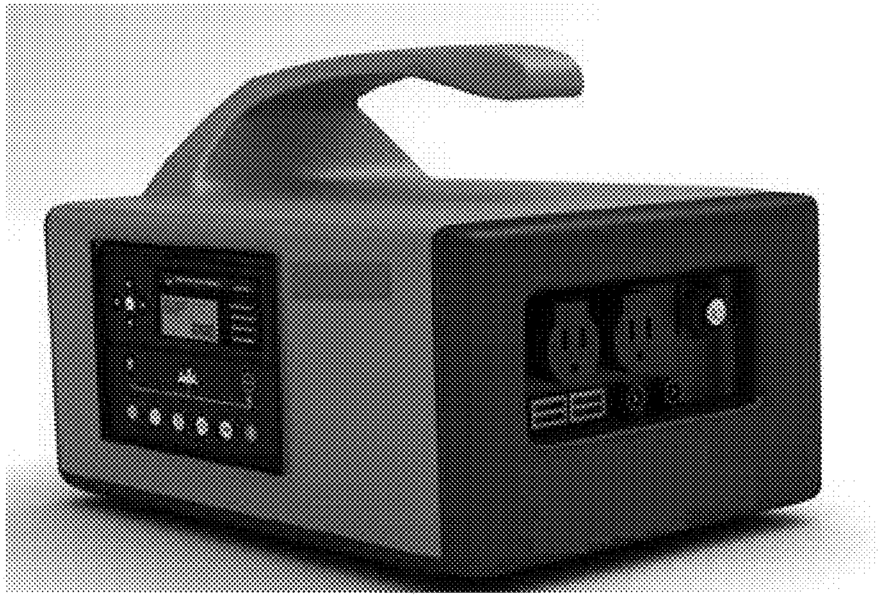

FIG. 5B

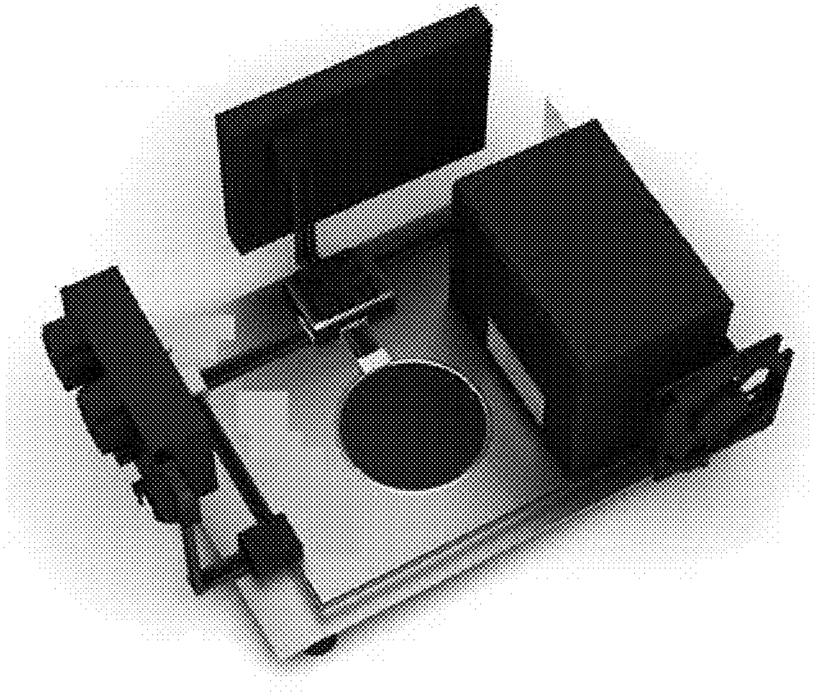

FIG. 5C

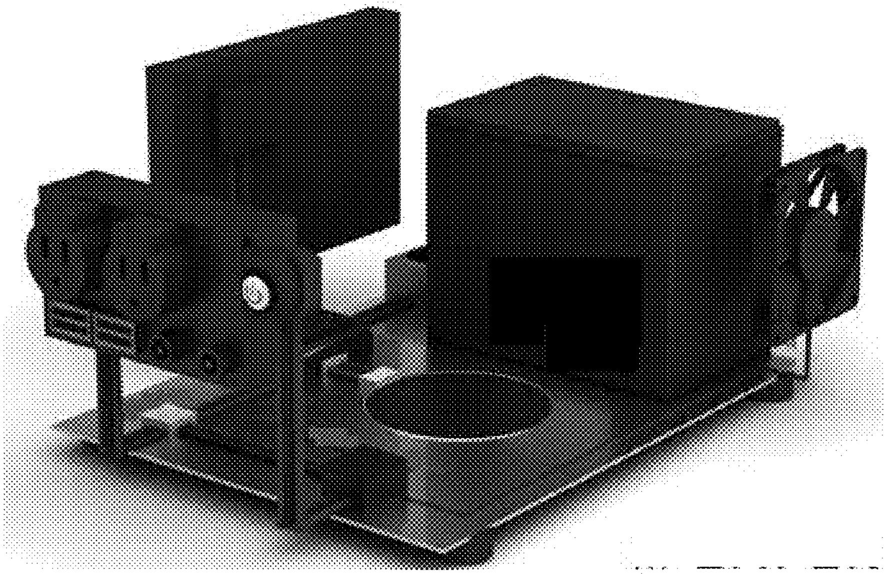

FIG. 5D

ELECTROMAGNETIC POWER GENERATOR

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of and priority from U.S. Provisional Application No. 62/018,826, filed Jun. 30, 2014, which is herein incorporated by reference in its entirety and for all purposes.

FIELD

[0002] The disclosure relates generally to electromagnetic generators, in particular, electromagnetic generators that utilize radiofrequency signals as an energy input.

BACKGROUND

[0003] Consumers rely on electricity to power various devices and appliances used on a daily basis. For example, electricity is used to charge cell phones, power televisions, heat water, propel cars, cook food, and the like. Electricity is a critical component to everyday life. Consumers typically receive electricity in one of two ways: (1) from a utility company or (2) from an electric generator.

[0004] Often, utility companies rely on large-scale electric generators to create electricity in a similar manner to smaller-scale electric generators used by individuals. Electric generators are devices that convert one form of energy (e.g., mechanical energy) into electrical energy (i.e., electricity), which can be used to power external devices or can be stored in storage devices, such as batteries and capacitors, for later use. For example, some electric generators use mechanical energy input to turn or rotate permanent magnets. The rotating permanent magnets create changing magnetic fields, which react with conductors to drive an electric current. Accordingly, at least a portion of the input mechanical energy is converted into usable electric energy (i.e., electricity). The input mechanical energy may be provided manually (e.g., via a hand crank), provided from the output of an internal combustion engine, provided from a steam driven turbine (e.g., as done in a nuclear or fossil fuel power plant), or the like.

[0005] Large amounts of input mechanical energy are typically required to generate a meaningful amount of electricity. This often requires paying a utility company for a portion of the electricity generated by the utility company or the burning of fossil fuels in an internal combustion engine driven personal generator. Electricity generated by utility companies can be expensive. Additionally, the burning of fossil fuels can be expensive and produces potentially harmful byproducts, such as carbon dioxide, carbon monoxide, sulfur, and the like. Accordingly, improved electricity sources are desired.

SUMMARY

[0006] One embodiment of the present invention relates to an electromagnetic generator comprising an antenna configured to receive radiofrequency energy and a first circuit connected to the antenna and configured to charge a first battery using the radiofrequency energy received by the antenna. A control unit is configured to receive power from the first battery and is further configured to power at least one electromagnet to generate a magnetic field. A rotatable body having at least one magnet is configured to rotate when the magnetic field is generated. A second circuit is connected to the rotatable body and configured to supply power to an external device.

[0007] Another embodiment relates to a method for supplying power using radiofrequency energy. The method includes receiving radio waves through an antenna and converting energy from the radio waves to a first current. The first current is used charge a first battery. At least one electromagnet is powered to produce a magnetic field using energy supplied from the first battery. The method further includes rotating a rotatable body due to the magnetic field produced by the at least one electromagnet and convertin energy produced by the rotation of the rotatable body to a second current. The second current is used to power an external device.

[0008] A further embodiment relates to an electromagnetic generator. The generator includes an antenna configured to receive radiofrequency energy. The generator further includes a rectifier circuit connected to the antenna and configured convert the radiofrequency energy to direct current energy used to charge a first battery. The generator includes a control unit configured to receive power from the first battery and further configured to power at least one electromagnet to generate a magnetic field. The generator further includes a flywheel having at least one magnet and configured to rotate when the magnetic field is generated. The generator includes an alternator mechanically connected to the flywheel and configured to charge a second battery based on the rotation of the flywheel. The second battery is configured to supply power to an external device.

BRIEF DESCRIPTION OF THE FIGURES

[0009] Features, aspects, and advantages of the present invention will become apparent from the following description and the accompanying exemplary embodiments shown in the drawings, which are briefly described below.

[0010] FIG. **1** shows a schematic diagram of an exemplary embodiment of an electromagnetic generator of the present invention.

[0011] FIG. **2** shows a flow chart illustrating a process in which radiofrequency energy is utilized in order to charge a battery to power devices.

[0012] FIG. **3** shows a perspective front view of the electromagnetic generator.

[0013] FIG. **4** shows a perspective back view of the electromagnetic generator.

[0014] FIGS. **5**A-**5**D shows a perspective views of additional embodiments of the electromagnetic generator.

DETAILED DESCRIPTION

[0015] Referring to the figures generally, systems and methods for generating usable quantities of electricity from radio waves via an electromagnetic electricity generator are described. Radio waves are electromagnetic waves that are emitted and received by many different types of devices. Each radio wave contains electromagnetic energy that can be converted into usable electricity through the described electromagnetic electricity generator. The electromagnetic electricity generator utilizes an antenna to receive the radio wave and a rectifier circuit to convert the electromagnetic electricity generator into an electric current. The electric current is used to drive a magnetic flywheel, which generates mechanical energy. The mechanical energy is then converted into electricity, which can be used to charge batteries and/or power devices. The radio waves received by the antenna can be generated from a remote power company or may be scavenged from the ambient environment (e.g., from television broadcasts, radio broadcasts, cellular phone base stations, wireless networking devices, etc.).

[0016] FIG. **1** schematically shows an electromagnetic generator **100** according to an exemplary embodiment. The electromagnetic generator **100** includes an antenna **30** connected to a rectifier circuit **40**. The antenna **30** is configured to receive radio waves **12** from the ambient environment. In some arrangements, the received radio waves have a frequency of approximately 945 MHz. The radio waves **12** generate an alternating current (AC) in the antenna **30**. The radio waves **12** may originate from a dedicated radio wave generation plant (e.g., a utility company) or may be scavenged from the ambient environment (e.g., from television broadcast transmitters, radio broadcast transmitters, cellular phone base stations, wireless networking devices, etc.).

[0017] The generated AC power is then provided from the antenna **30** to the rectifier circuit **40**, which converts the AC power into direct current (DC). This DC power is used to charge a first battery **10**. The first battery **10** is a rechargeable battery. Accordingly, the first battery **10** may be any of a nickel cadmium (NiCd) battery, a nickel metal hydride (NiMH) battery, a lithium ion (LiIon) battery, a sealed lead acid (SLA) battery, or the like. The first battery **10** may include a plurality of individual battery cells. In some arrangements, the first battery **10** includes a charging controller that controls the input of DC power into the first battery **10** to prevent damage to the first battery **10** that may result from overcharging and/or overheating. Because the generator **100** continuously receives radio waves **12** through the antenna **30** and continuously converts the electromagnetic energy of the received radio waves **12** into DC power, the first battery **10** can be continuously charged over time (e.g., trickle charged). The first battery **10** is not limited to a single battery. Alternatively, the first battery **10** may comprise a plurality of batteries. In such arrangements, once one battery of the plurality of batteries is sufficiently charged, other batteries may be charged by the received DC power from the rectifier circuit **40**. This allows for additional storage of electric energy from the radio waves **12**.

[0018] Once the first battery **10** is sufficiently charged, the battery **10** is then used to supply electric power to a control unit **50**. The control unit **50** controls a plurality of electromagnets **55**. Although three electromagnets **55** are shown in FIG. **1**, it should be understood that any number of electromagnets **55** may be utilized by the generator **100**. The control unit **50** includes a power input that receives electric power from the first battery **10**. The control unit **50** includes a processor and memory. The memory stores programming modules that, when executed by the processor, control the operation of the control unit **50** as described in further detail below. The electromagnets **55** are mounted adjacent to a flywheel **60** that has a plurality of permanent magnets **65** affixed thereon. The permanent magnets **65** are positioned about a circumference of the flywheel. The electromagnets **55** are aligned such that the electromagnets **55** project a magnetic field over at least a portion of the flywheel **60** when the electromagnets are powered. In arrangements where a plurality of electromagnetics **55** are utilized by the generator **100**, each of the electromagnets can have a different position with respect to the flywheel **60**. For example, the electromagnets **55** may be positioned around the circumference of the flywheel **60**. Moreover, each of the electromagnets **55** may be further configured to produce different magnetic field strengths with respect to each one another. The flywheel **60** is a cylinder that has a diameter that is significantly larger than its axial length. In some arrangements, the diameter of the flywheel **60** is at least ten times the axial length of the flywheel **60**.

[0019] The flywheel **60** is rotationally coupled to an axle **67**. The axle **67** extends along a central axis of the flywheel **60**. Accordingly, when the flywheel **60** rotates, the axle **67** rotates with the same rotational velocity as the flywheel **60**. The axle **67** is connected to an alternator **70**. When the flywheel **60** rotates, the axle **67** rotates and provides rotational mechanical energy into the alternator **70**. The alternator **70** is an electromechanical device that converts rotational mechanical energy into electricity.

[0020] During operation, the control unit **50** selectively and independently powers each of the electromagnets **55** to generate magnetic fields. The magnetic fields interact with the permanent magnets **65** positioned on the flywheel **60**, which causes the flywheel **60** to spin. In some operating conditions, the control unit **50** will cause the magnetic field generated by a given electromagnet to be of the same polarity as the permanent magnets **65** to repel the permanent magnets **65**. In other operating conditions, the control unit **50** will cause the magnetic field generated by a given electromagnet to be of the opposite polarity as the permanent magnets **65** to attract the permanent magnets **65**. The control unit **50** determines when to power each of the electromagnets **55**, the amount of power to provide to each electromagnet **55** when powered on, the polarity of the magnetic field to be generated, the strength of the magnetic field to be generated, and when to turn off each electromagnet **55** based on at least in part on feedback from a position and speed sensor **68**. The position and speed sensor **68** measures the position and rotational speed of either the flywheel **60** or the axle **67**. By selectively and independently powering each of the electromagnets **55**, the control unit **50** can accelerate the flywheel **60** to a high rotational speed (e.g., above 1000 RPM) and maintain the high rotational speed of the flywheel **60**. The flywheel **60** may be mounted to a low friction bearing such that little energy input is needed to maintain the rotational speed of the flywheel **60**.

[0021] The spinning of the flywheel **60** produces rotational mechanical energy. The rotational mechanical energy is transferred from the flywheel **60** to the alternator **70** via the axle **67**. The alternator **70** converts the mechanical energy into DC power. The DC power output by the alternator **70** is then used to charge a second battery **20**. The second battery **20** is a rechargeable battery. Accordingly, the second battery **20** may be any of a NiCd battery, a NiMH battery, a LiIon battery, a SLA battery, or the like. The second battery **20** may include a plurality of individual battery cells. In some arrangements, the second battery **20** includes a charging controller that controls the input of DC power into the second battery **20** to prevent damage to the second battery **20** that may result from overcharging and/or overheating. The charging controller may provide feedback to the control unit **50** to indicate a charge of the second battery **20**. For example, the charging controller may provide a voltage of the second battery **20** to the control unit **50**. Based on the feedback provided to the control unit **50**, can stop powering the flywheel **60** or accelerate the flywheel **60**. The second battery **20** stores the power until it is needed. As described above with the first battery **10**, the second battery **20** may also comprise a plurality of batteries. These batteries may also be continuously and/or alternately charged and discharged to store and provide power, allowing for an increase in capacity of the generator **100**.

[0022] The power stored by the second battery **20** is then sent to a power inverter **80**, which converts the DC power supplied by the second battery into AC power. The AC power is output from the generator. In some arrangements, the AC power is supplied to a power outlet **90** to supply power to external devices when needed. In other arrangements, the AC power is provided to a power system (e.g., for a building) or to a power grid (e.g., a power grid that supplies a plurality of buildings and devices). In an alternative arrangement, the power inverter **80** is bypassed and DC power is output from the generator **100** (e.g., to power a device via a USB port).

[0023] FIG. **2** illustrates a flow chart of a method **200** of converting radiofrequency energy into electricity according to an exemplary embodiment. Method **200** is performed by the generator **100** as described above with respect to FIG. **1**. In step S**1**, radio waves **12** are received by the antenna **30** and converted to DC power by the rectifier circuit **40**. The DC power is then used to charge the first battery **10** in step S**2**. At step S**3**, the energy stored on the first battery **10** is used to power the control unit **50** and the electromagnets **55**. As described above, the control unit **50** selectively and independently activates and deactivates each electromagnet to generate magnetic fields. The magnetic fields interact with the permanent magnets **65** mounted on the flywheel **60**, which cause the flywheel **60** to rotate at step S**4**. At step S**5**, the rotational mechanical energy of the flywheel **60** is transferred to the axle **67**, and is provided to the alternator **70**, which converts the received mechanical energy into DC power. The DC power is then output to the second battery **20** at step S**6**. The second battery **20** stores the electric energy until needed. When power is needed, it is first converted from DC to AC by the inverter **80** at step S**7**. The AC power is then outputted from the generator **100** at step S**8**. In some arrangements, the AC power is supplied to a power outlet **90** to supply power to external devices when needed. In other arrangements, the AC power is provided to a power system (e.g., for a building) or to a power grid (e.g., a power grid that supplies a plurality of buildings and devices). In an alternative arrangement, step s**7** is bypassed (e.g., the power inverter **80** is bypassed) and DC power is output from the generator **100** (e.g., to power a device via a USB port).

[0024] FIGS. **3** and **4** illustrate perspective front and back views of one embodiment of the electromagnetic generator **100**. The components described above are preferably housed in a case **110** of the generator **100**. As shown in FIGS. **3** and **4**, the generator **100** may include a control panel **120**. The control panel **120** may comprise indicator lights that may indicate to the user the current state of the generator **100** and/or the storage levels of the batteries. The control panel **120** may also indicate to the user other measurements of the system, such as current input or output levels. As shown in FIG. **3**, the generator **100** may also include a power switch **95** to send energy stored in the second battery **20** to the power outlet **90** when required. As shown in FIG. **4**, the generator **100** may further include a cooling system **150**, such as a fan, to prevent overheating of the system. FIGS. **5**A-**5**D illustrate perspective views of an additional embodiment of the electromagnetic generator **100**, with an alternative case and component design.

[0025] The above-described generator **100** may include additional features. For example, the generator **100** may include an integrated cooling system that cools the various components (e.g., the first battery **10**, the second battery **20**, the control unit **50**, etc.) to prevent overheating. The cooling system may be a liquid cooling system or an air cooling system. The generator **100** may include a passive voltage control system and an electrical and a fuse system to prevent any overload problem. In some arrangements, a gearbox is coupled between the rotating axle **67** and the alternator **70** to modify the torque and rotational speed input into the alternator **70**.

[0026] In some arrangements, the above-described generator **100** is distributed by a power delivery company. In such arrangements, the power delivery company may distribute generators to its subscribers. The power delivery company generates wireless signals (e.g., from a transmitter tower). Each generator may have an identifier associated with each subscriber. The identifier can be communicated wirelessly (e.g., through radio-frequency identification, Bluetooth, WiFi, etc.). Alternatively, the identifier may be associated with the entry of a specific user ID into the control panel **120** (e.g., user ID and password, user PIN, etc.). Each generator also includes a data transceiver (e.g., a cellular data transceiver) that provides real-time feedback to the power delivery company relating to an amount of electricity consumed by the associated subscriber. The power delivery company can then accurately bill each of its subscribers for actual electricity used.

[0027] The above-described generator **100** may be adapted to power specific devices. For example, the generator **100** may be integrated into and provide operational power to tools, such as powered gardening equipment (e.g., lawn mowers, weed whackers, electric tree trimmers, etc.), tools (e.g., air compressors, power washers, saws, drills, etc.), hospital equipment to provide off-grid power backup, vehicles or other propulsion systems (e.g., electric motor boats, electric aircraft, electric automobiles, electric motorbikes, etc.), appliances and HVAC equipment (e.g., refrigeration systems, heating systems, clothes washers, clothes dryers, etc.), commercial construction equipment, and the like. The generator **100** may be used as a replacement for traditional fossil fuel based electricity generators, such as diesel or gasoline generators. Furthermore, the generator **100** is not limited in its use of radiofrequency signals from transmission towers. The generator **100** may alternatively be used in space systems, utilizing radiofrequency energy present in space communications.

[0028] The above-described generator **100** may also be used in combination with other power-transmitting devices. For example, the generator **100** may be used together with a traditional gas engine, electric motor, or a hybrid engine. The generator **100** may also be used in combination with other forms of alternative energy, such as solar power, wind power, or hydropower.

[0029] In the foregoing description, certain terms have been used for brevity, clearness, and understanding. No unnecessary limitations are to be implied therefrom beyond the requirement of the prior art because such terms are used for descriptive purposes and are intended to be broadly construed. The different configurations, systems and method steps described herein may be used alone or in combination with other configurations, systems and method steps. It is to be expected that various equivalents, alternatives and modifications are possible.

[0030] It should be noted that any use of the term "exemplary" herein to describe various embodiments is intended to indicate that such embodiments are possible examples, representations, and/or illustrations of possible embodiments (and such term is not intended to connote that such embodiments are necessarily extraordinary or superlative examples).

[0031] The use of the term "approximately" in relation to numbers, values, and ranges thereof refers to plus or minus five percent of the stated of numbers, values, and ranges thereof.

[0032] The terms "coupled" and the like as used herein mean the joining of two members directly or indirectly to one another. Such joining may be stationary (e.g., permanent) or moveable (e.g., removable or releasable). Such joining may be achieved with the two members or the two members and any additional intermediate members being integrally formed as a single unitary body with one another or with the two members or the two members and any additional intermediate members being attached to one another.

[0033] References herein to the positions of elements (e.g., "top," "bottom," "above," "below," etc.) are merely used to describe the orientation of various elements in the FIGURES. It should be noted that the orientation of various elements may differ according to other exemplary embodiments, and that such variations are intended to be encompassed by the present disclosure.

[0034] It is important to note that the construction and arrangement of the various exemplary embodiments are illustrative only. Although only a few embodiments have been described in detail in this disclosure, those skilled in the art who review this disclosure will readily appreciate that many modifications are possible (e.g., variations in sizes, dimensions, structures, shapes and proportions of the various elements, values of parameters, mounting arrangements, use of materials, colors, orientations, etc.) without materially departing from the novel teachings and advantages of the subject matter described herein. For example, elements shown as integrally formed may be constructed of multiple parts or elements, the position of elements may be reversed or otherwise varied, and the nature or number of discrete elements or positions may be altered or varied. The order or sequence of any process or method steps may be varied or re-sequenced according to alternative embodiments, and elements from different embodiments may be combined in a manner understood to one of ordinary skill in the art. Other substitutions, modifications, changes and omissions may also be made in the design, operating conditions and arrangement of the various exemplary embodiments without departing from the scope of the present invention.

What is claimed is:

1. An electromagnetic generator comprising:

an antenna configured to receive radiofrequency energy;

a first circuit connected to the antenna and configured to charge a first battery using the radiofrequency energy received by the antenna;

a control unit configured to receive power from the first battery and further configured to power at least one electromagnet to generate a magnetic field;

a rotatable body having at least one magnet and configured to rotate when the magnetic field is generated; and

a second circuit connected to the rotatable body and configured to supply power to an external device based on the rotation of the rotatable body.

2. The electromagnetic generator of claim **1**, further comprising a second battery, wherein the second circuit is further configured to charge the second battery based on the rotation of the rotatable body.

3. The electromagnetic generator of claim **2**, wherein the second battery is configured to supply power to the external device.

4. The electromagnetic generator of claim **1**, wherein the rotatable body is a flywheel.

5. The electromagnetic generator of claim **1**, wherein the first circuit is a rectifier.

6. The electromagnetic generator of claim **1**, wherein the second circuit is an alternator.

7. The electromagnetic generator of claim **1**, wherein the first battery comprises at least two batteries.

8. The electromagnetic generator of claim **2**, wherein the second battery comprises at least two batteries.

9. The electromagnetic generator of claim **1**, comprising three electromagnets.

10. The electromagnetic generator of claim **9**, wherein the three electromagnets are each configured to have different magnetic field strengths.

11. The electromagnetic generator of claim **1**, further comprising a cooling system.

12. The electromagnetic generator of claim **1**, further comprising a control panel configured to indicate to a user charging levels of the first battery and the second battery.

13. A method for supplying power using radiofrequency energy, comprising:

receiving radio waves through an antenna;

converting energy from the radio waves to a first current;

using the first current to charge a first battery;

powering at least one electromagnet to produce a magnetic field using energy supplied from the first battery;

rotating a rotatable body due to the magnetic field produced by the at least one electromagnet;

converting energy produced by the rotation of the rotatable body to a second current; and

using the second current to power an external device.

14. The method of claim **13**, further comprising the steps of using the second current to charge a second battery and powering an external device using energy supplied from the second battery.

15. The method of claim **13**, wherein the first battery is a plurality of batteries.

16. The method of claim **13**, wherein the rotatable body is a flywheel.

17. The method of claim **13**, wherein the first current is a direct current.

18. The method of claim **13**, wherein the second current is an alternating current.

19. The method of claim **14**, wherein the second current is a direct current.

20. An electromagnetic generator comprising:

an antenna configured to receive radiofrequency energy;

a rectifier circuit connected to the antenna and configured convert the radiofrequency energy to direct current energy used to charge a first battery;

a control unit configured to receive power from the first battery and further configured to power at least one electromagnet to generate a magnetic field;

a flywheel having at least one magnet and configured to rotate when the magnetic field is generated; and an alternator mechanically connected to the flywheel and configured to charge a second battery based on the rotation of the flywheel;

wherein the second battery is configured to supply power to an external device.

* * * * *

US 20160149441A1

(19) **United States**
(12) **Patent Application Publication**
Nayak
(10) **Pub. No.: US 2016/0149441 A1**
(43) **Pub. Date: May 26, 2016**

(54) **RADIO FREQUENCY HARVESTER ASSEMBLY**

(71) Applicant: **UTC FIRE AND SECURITY AMERICAS CORPORATION, INC.**, Bradenton, FL (US)

(72) Inventor: **Jagannath Nayak**, Aska (IN)

(21) Appl. No.: **14/903,322**

(22) PCT Filed: **Jul. 7, 2014**

(86) PCT No.: **PCT/US2014/045572**
§ 371 (c)(1),
(2) Date: **Jan. 7, 2016**

Related U.S. Application Data

(60) Provisional application No. 61/843,667, filed on Jul. 8, 2013.

Publication Classification

(51) **Int. Cl.**

H02J 50/27	(2006.01)
H01Q 21/06	(2006.01)
H01Q 1/38	(2006.01)

(52) **U.S. Cl.**
CPC ***H02J 50/27*** (2016.02); ***H01Q 1/38*** (2013.01); ***H01Q 21/06*** (2013.01)

(57) **ABSTRACT**

A device for harvesting radio frequency (RF) energy to generate power includes a first spiral antenna having a first spiral direction, the first spiral antenna having a spiral cross-section shape along a first flat plane. The device includes a second spiral antenna having a second spiral direction opposite the first spiral direction, the second spiral antenna having a spiral cross-section shape along a second plane parallel to the first plane. The device further includes a ground layer between the first and second spiral antennae, the first spiral antenna electrically connected to the second spiral antenna at a center of the first and second spiral antennae by a conductive element passing through the ground layer without contacting the ground layer.

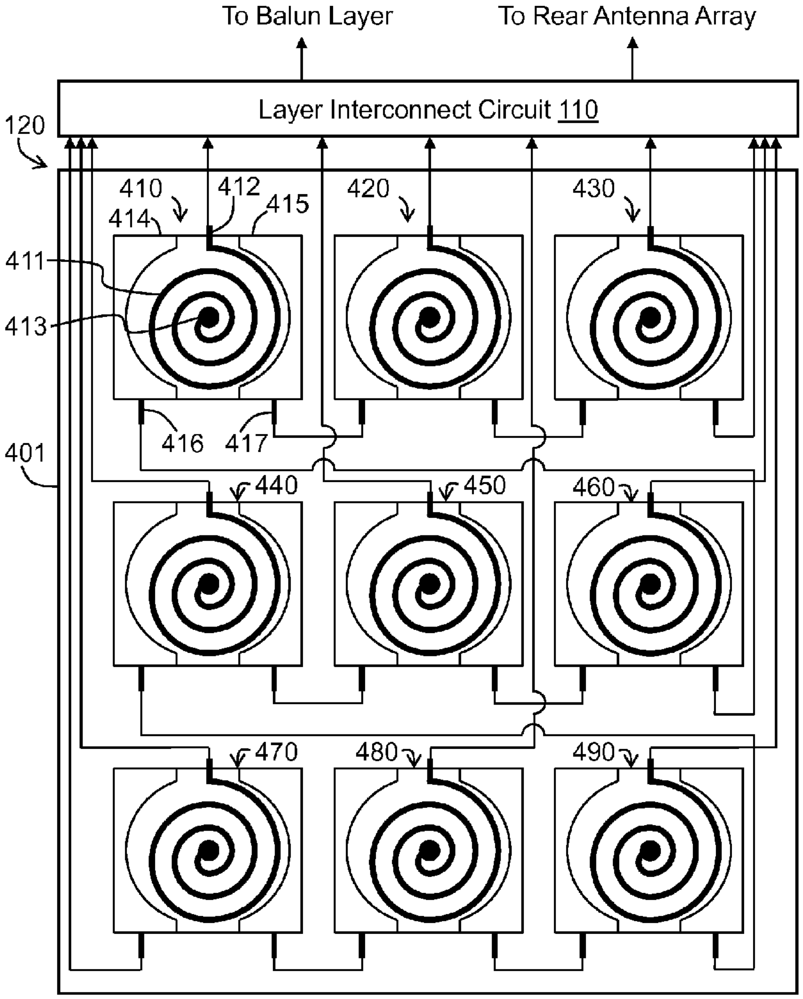

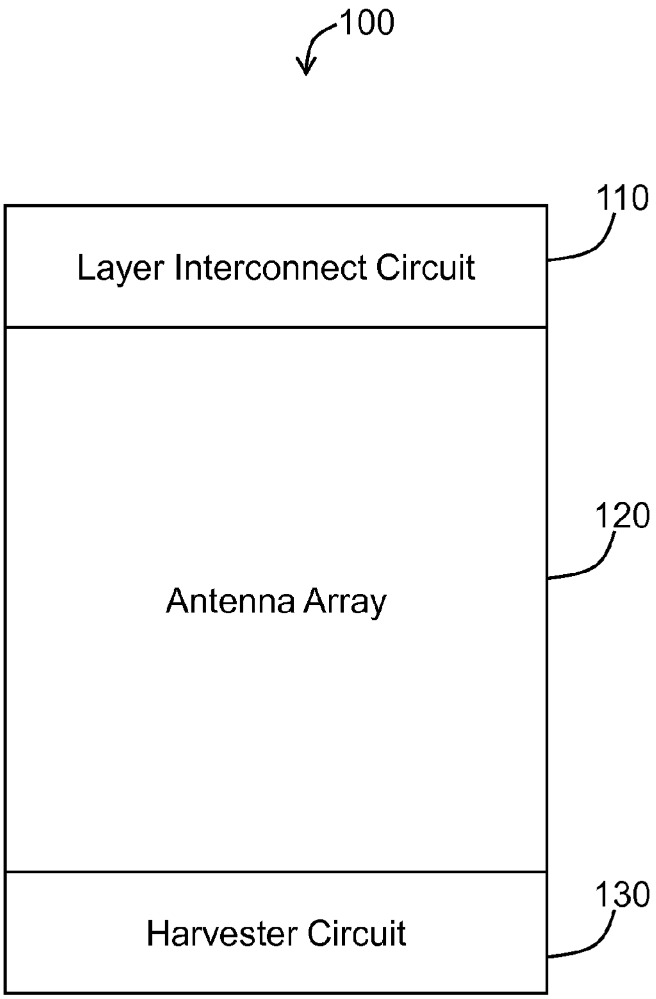
100
110
Layer Interconnect Circuit
120
Antenna Array
130
Harvester Circuit

FIG. 1

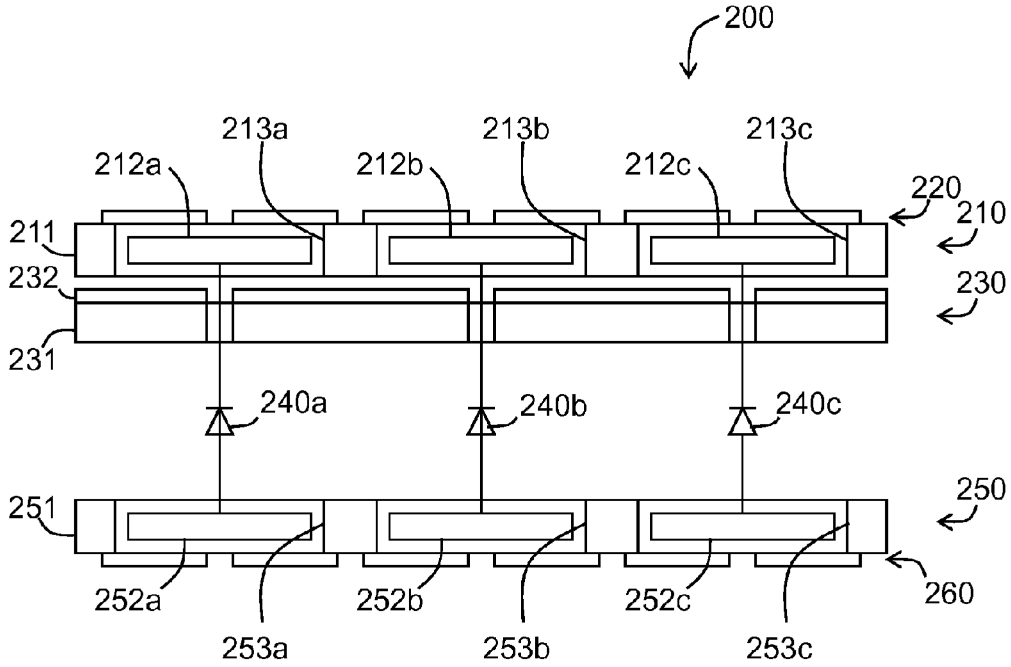

FIG. 2

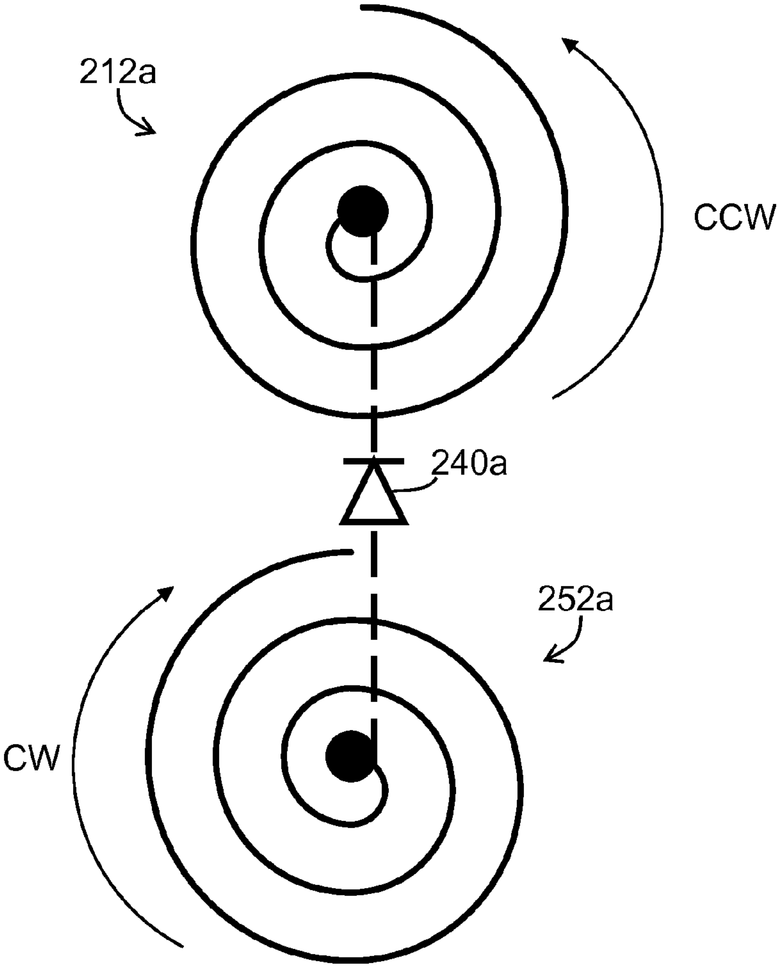
212a
CCW
240a
252a
CW

FIG. 3

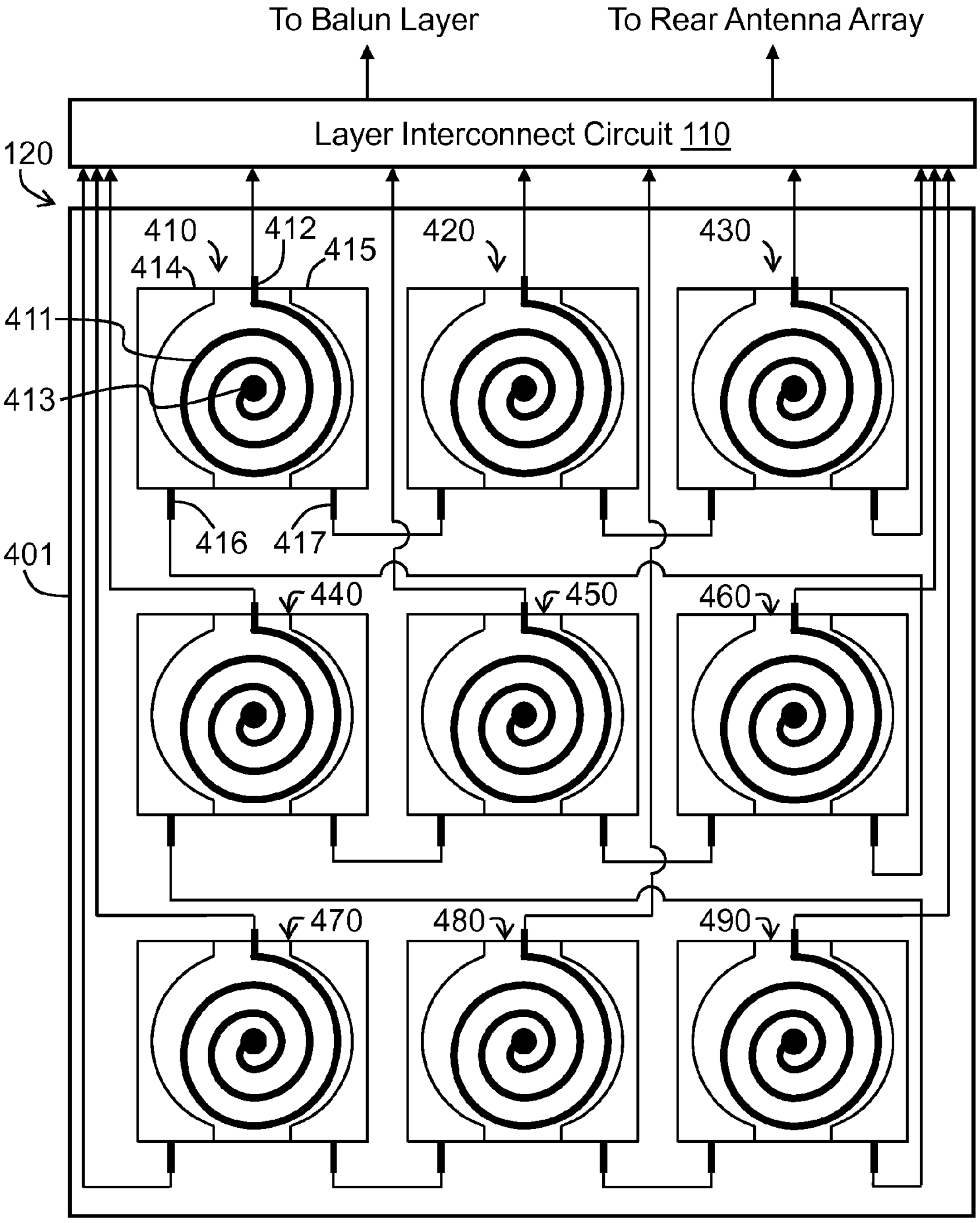
To Balun Layer
To Rear Antenna Array
Layer Interconnect Circuit 110

FIG. 4

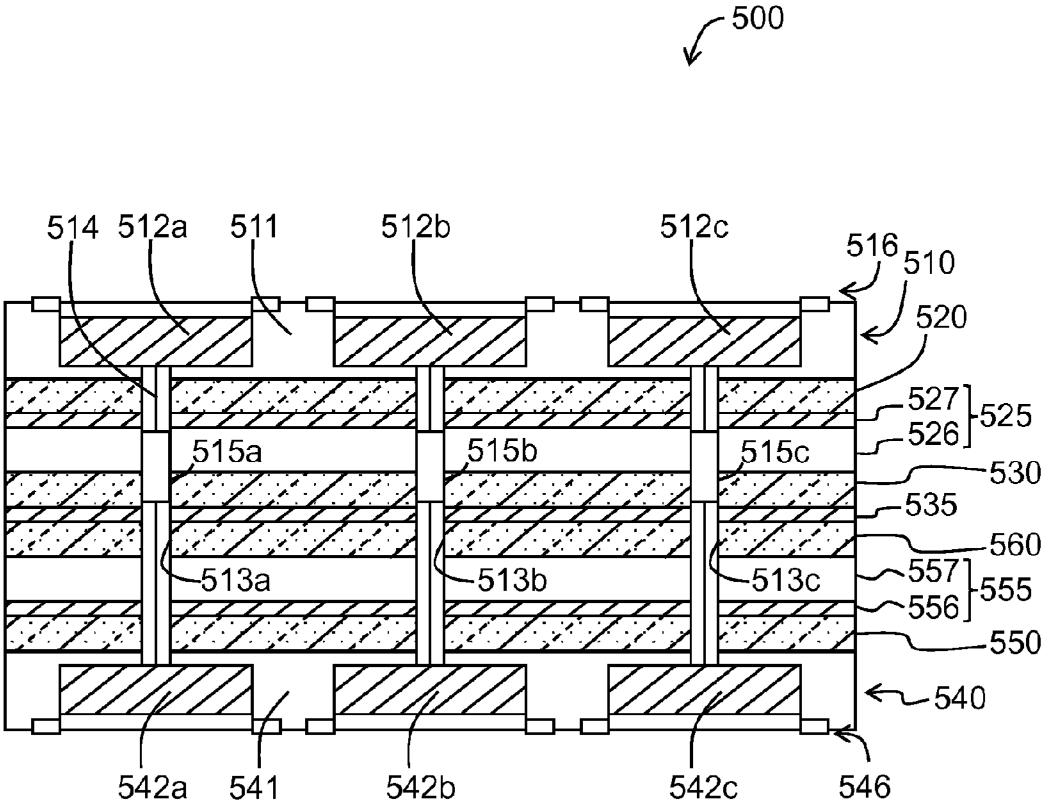

FIG. 5

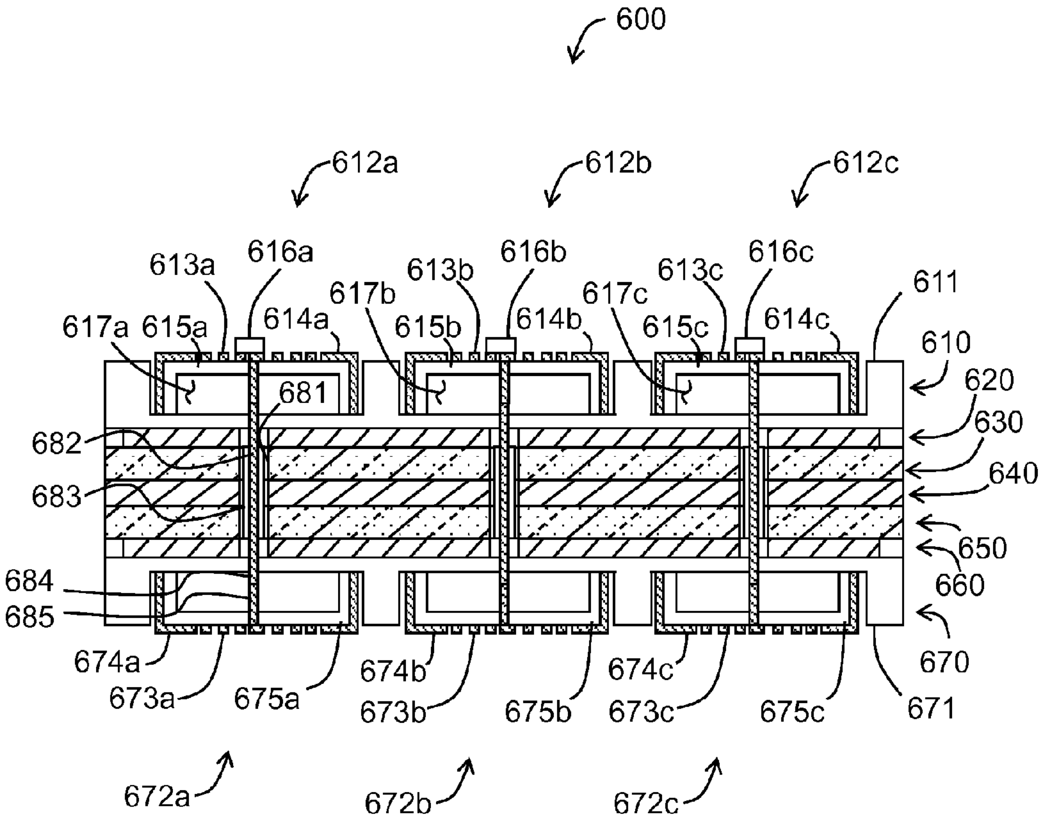

FIG. 6

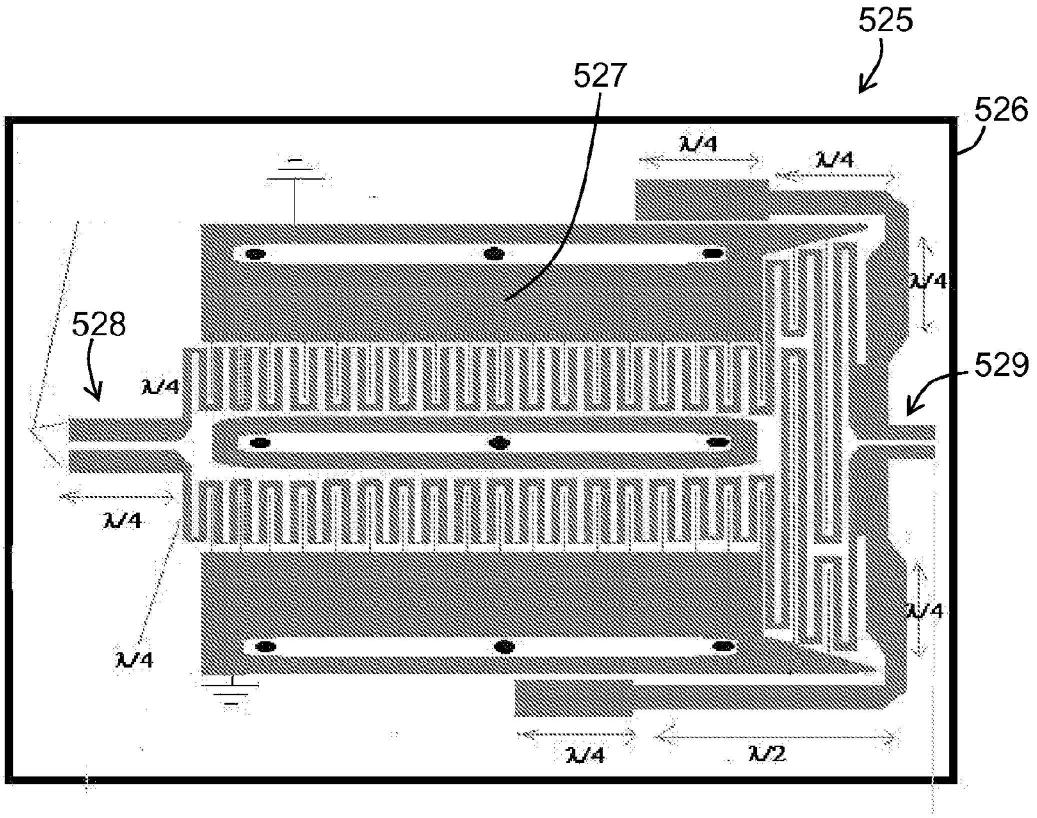
525
527
526
λ/4
λ/4
λ/4
528
λ/4
529
λ/4
λ/4
λ/4
λ/4
λ/2

FIG. 7

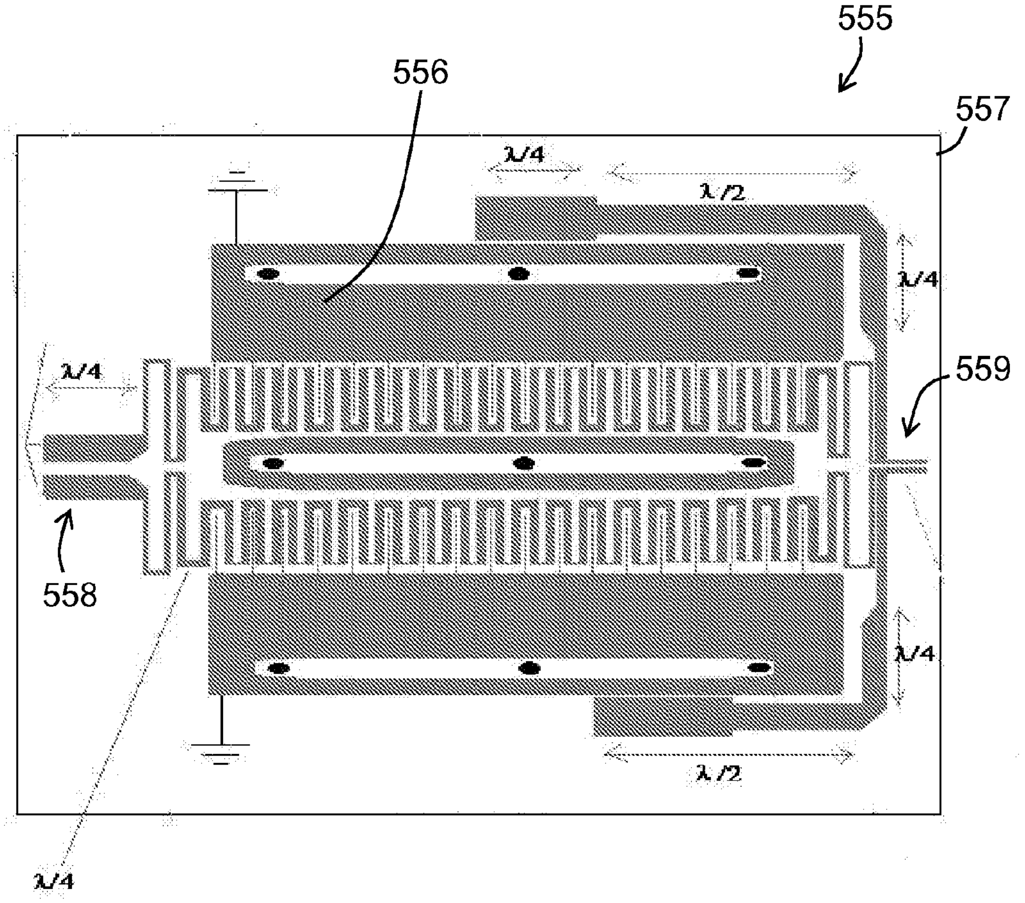
555
556
557
λ/4
λ/2
λ/4
λ/4
559
558
λ/4
λ/2
λ/4

FIG. 8

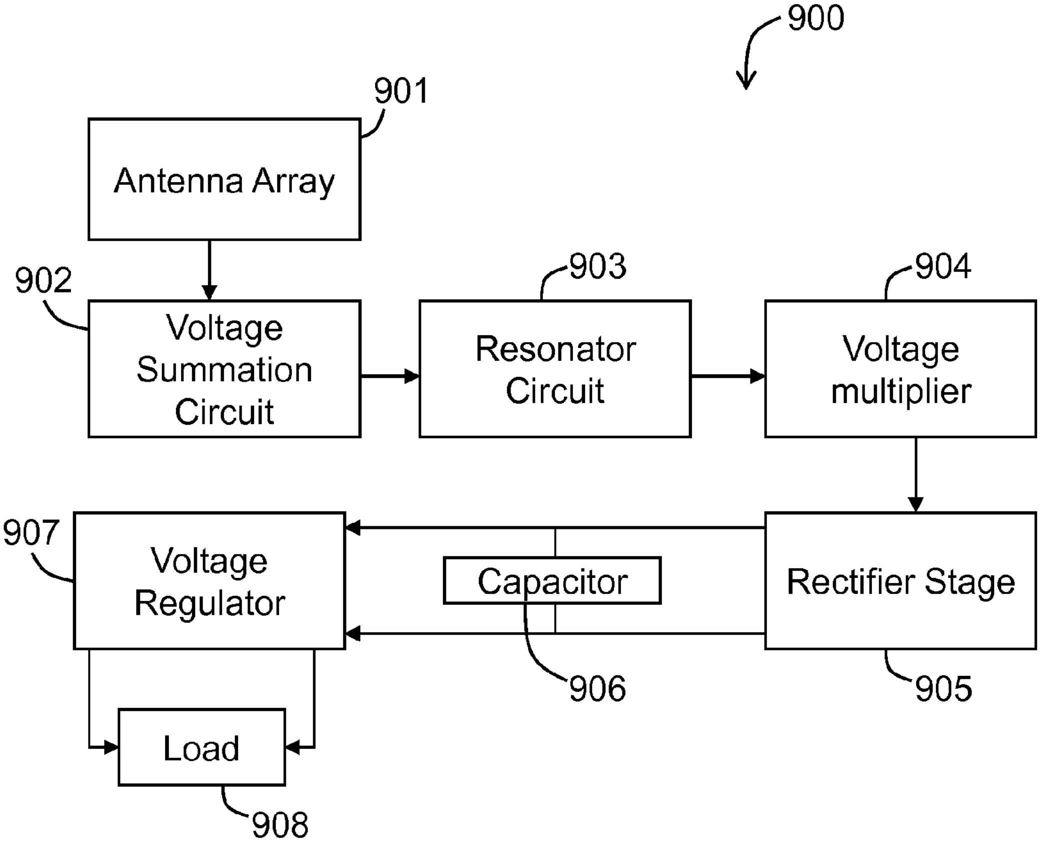
900
901
Antenna Array
902
Voltage Summation Circuit
903
Resonator Circuit
904
Voltage multiplier
907
Voltage Regulator
Capacitor
Rectifier Stage
906
905
Load
908

FIG. 9

RADIO FREQUENCY HARVESTER ASSEMBLY

BACKGROUND OF THE INVENTION

[0001] Embodiments of the invention relate to radio frequency harvesters, and in particular to spiral antennae for harvesting RF signals to generate power.

[0002] Numerous methods are being developed and used to harvest non-conventional or ambient energy available in nature to power small electronic devices. Examples of non-conventional power sources include solar energy, vibration energy, radio frequency (RF) energy, heat energy, etc. Some of these energy sources are ubiquitous, while others are available only at particular times or locations.

[0003] The increasing use of wireless communication and wireless networks makes RF energy everywhere in the world. Most of the energy generated to facilitate RF communication is dissipated in the environment (e.g. air molecules, structures, organisms, etc.) Efforts are underway to capture some of this otherwise-lost RF energy to generate power.

BRIEF DESCRIPTION OF THE INVENTION

[0004] Embodiments of the invention include a device for harvesting radio frequency (RF) energy to generate power. The device includes a first spiral antenna having a first spiral direction, the first spiral antenna having a spiral cross-section shape along a first flat plane. The device includes a second spiral antenna having a second spiral direction opposite the first spiral direction, the second spiral antenna having a spiral cross-section shape along a second plane parallel to the first plane. The device further includes a ground layer between the first and second spiral antennae, the first spiral antenna electrically connected to the second spiral antenna at a center of the first and second spiral antennae by a conductive element passing through the ground layer without contacting the ground layer.

[0005] Additional embodiments include a device for harvesting radio frequency (RF) energy to generate power. The device includes a first array of spiral antennae arranged along a first flat plane and a second array of spiral antennae arranged along a second plane parallel to the first plane. The device also includes a ground layer arranged between the first array and the second array and a harvester circuit configured to receive an RF signal based on RF energy gathered by the first and second arrays of spiral antennae and to generate direct current (DC) power based on the RF signal.

BRIEF DESCRIPTION OF THE DRAWINGS

[0006] The subject matter, which is regarded as the invention, is particularly pointed out and distinctly claimed in the claims at the conclusion of the specification. The foregoing and other features, and advantages of the invention are apparent from the following detailed description taken in conjunction with the accompanying drawings in which:

[0007] FIG. **1** illustrates an antenna assembly according to an embodiment of the invention;

[0008] FIG. **2** illustrates a side view of a part of an antenna assembly according to an embodiment of the invention;

[0009] FIG. **3** illustrates interconnected spiral antennae according to an embodiment of the invention;

[0010] FIG. **4** illustrates an antenna array according to an embodiment of the invention;

[0011] FIG. **5** illustrates layers of an antenna assembly according to an embodiment of the invention;

[0012] FIG. **6** illustrates layers of an antenna assembly according to another embodiment of the invention;

[0013] FIG. **7** illustrates a first balun layer according to an embodiment of the invention;

[0014] FIG. **8** illustrates a second balun layer according to an embodiment of the invention; and

[0015] FIG. **9** illustrates a block diagram of harvester assembly according to an embodiment of the invention.

[0016] The detailed description explains embodiments of the invention, together with advantages and features, by way of example with reference to the drawings.

DETAILED DESCRIPTION OF THE INVENTION

[0017] Radio frequency (RF) energy is generated to communicate between RF devices, but the majority of RF energy is dissipated in the environment. Embodiments of the invention relate to harvesting RF energy with spiral antenna arrays.

[0018] FIG. **1** illustrates an antenna assembly **100** according to an embodiment of the invention. The antenna assembly **100** includes a layer interconnect circuit **110**, an antenna array **120** and a harvester circuit **130**. The antenna array **120** is a multi-layer antenna array, and the layer interconnect circuit **110** connects the multiple layers of the antenna array **120**. The harvester circuit **130** collects the electrical energy generated by the antenna array **120** and proves an electrical outlet to provide the electrical energy to an electrical device, such as a battery or other device that runs on electricity or stores an electrical charge.

[0019] FIG. **2** illustrates a side view of an antenna assembly **200** according to an embodiment of the invention. The antenna assembly **200** includes a first spiral antenna layer **210**, a first flat antenna layer **220**, a ground layer **230**, a second spiral antenna layer **250**, and a second flat antenna layer **260**. The first spiral antenna layer **210** includes three spiral antennae **212***a*, **212***b* and **212***c* located in openings **213***a*, **213***b* and **213***c* in a first substrate **211**. The second spiral antenna layer **250** includes three spiral antennae **252***a*, **252***b* and **252***c* located in openings **253***a*, **253***b* and **253***c* in a second substrate **251**. The three spiral antennae **212***a*, **212***b* and **212***c* are connected to the three spiral antennae **252***a*, **252***b* and **252***c* by electrical components including Schottky diodes **240***a*, **240***b* and **240***c*.

[0020] The first flat plate antenna layer **220** is formed on the first substrate **211**. The first flat plate antenna layer **220** comprises a conductive material, and a separate flat plate antenna of the flat plate antenna layer **220** may be positioned above each spiral antenna **212***a*, **212***b* and **212***c*. Similarly, the second flat plate antenna layer **260** is formed on the second substrate **251**. The flat plate antenna layers **220** and **260** will be described in further detail below with respect to FIG. **4**.

[0021] The ground layer **230** includes a substrate **231** and ground plate **232**. The ground plate **232** may be a conductive plate having holes cut into the plate to pass conductive lines connecting the first spiral antennae **212***a*, **212***b* and **212***c* and the second conductive antennae **252***a*, **252***b* and **252***c*. Since the first and second spiral antenna layers **210** and **250** are separated by a ground plate **232**, the radiation fields of the spiral antennae of the respective layers **210** and **250** are at least partially separated, and polarization loss and electromagnetic interaction loss are reduced or eliminated.

[0022] FIG. **3** illustrates connected spiral antennae **212***a* and **252***a* according to an embodiment of the invention. The spiral antennae **212***a* and **252***a* are arranged parallel to each other, and spiral in opposite directions. As illustrated in FIG. **3**, the first spiral antenna **212***a* spirals in a counter-clockwise direction (CCW) and the second spiral antenna **252***a* spirals in a clockwise direction (CW) from its center outward. In embodiments of the invention, the spiral antennae **212***a* and **252***a* connected by the Schottky diode **240***a* form a rectanna, which is an antenna that detects or receives microwaves and generates a direct current (DC) output based on the detected or received microwaves. The Schottky diode **240***a* rectifies the signals generated by the spiral antennae **212***a* and **252***a* and a load (not shown) may be connected across the Schottky diode **240***a*.

[0023] In embodiments of the invention, the spirals of the spiral antennae **212***a* and **252***a* may be Archimedean-style, single-armed spirals. Technical effects of such spiral antennae include higher spectral efficiency than other planner antennae, and the spiral antennae are frequency independent. In addition, in some embodiments the bandwidths of the spiral antennae reach up to 40:1 for both the input impedance and radiation patterns.

[0024] In addition, as illustrated in FIGS. **2** and **3**, the connected spiral antennae, such as spiral antennae **212***a* and **252***a,* result in dual polarization, which is achieved via dual planes of Archimedean spiral antennae, such as antennae **212***a* and **252***a*. In addition, as illustrated in FIG. **2**, the ground plate **232** or **272** between the connected spiral antennae electromagnetic interaction due to parallelism of spirals having opposite polarities. Further, while FIG. **2** illustrates different antennae, plates and layers of varying thicknesses, the illustrated thicknesses and sizes are for purposes of illustration only, and embodiments of the invention are not limited to sizes or proportions illustrated in FIG. **2**.

[0025] FIG. **4** illustrates an antenna array **120** according to an embodiment of the invention. The antenna array **120** includes antenna modules **410**, **420**, **430**, **440**, **450**, **460**, **470**, **480** and **490**. The antenna module **410** will be described in further detail, although the characteristics of the antenna module **410** are the same as the characteristics of the antenna modules **420**, **430**, **440**, **450**, **460**, **470**, **480** and **490**. In addition, the antenna modules **410**, **420** and **430** may correspond to the spiral antennae **212***a,* **212***b* and **212***c* of FIG. **2**.

[0026] The antenna module **410** includes a spiral antenna **411** and flat plate antennae **414** and **415**. The flat plate antennae **414** and **415** make up a portion of the flat plate antenna layer **220** of FIG. **2**. The flat plate antenna **415** includes an output connected to an adjacent flat plate antenna of the antenna module **420**, and the flat plate antenna **414** includes an output connected to a different flat plate antenna of the antenna module **460**.The spiral antenna **411** includes a first end connector **412** connected to a layer interconnect circuit **110** and a center connector **413** connected to a spiral antenna located on an opposite side of a spiral antenna assembly from the spiral antenna **411**, as illustrated in FIG. **2**.

[0027] The flat plate antennae **414** and **415**, also referred to as patch antennae **414** and **415**, capture leaky microwaves, or microwaves that may be emitted by the spiral antenna **411**. The flat plate antennae **414** and **415** have angular shapes that substantially surround the spiral antenna **411** along a circumference of the spiral antenna **411**, or in a radial direction with respect to the center of the spiral antenna **411**. FIG. **2** illustrates the flat plate antenna layer **220** located on top of the substrate **211** for purposes of description. However, embodiments encompass flat plate antennae that are inserted into grooves in a substrate **211** to have an outer surface that is substantially flush with the outer surface of the substrate **211**. The flat plate antennae **414** and **415** capture leaky waves and improve spiral antenna **411** broadband characteristics by improving a frequency vs. impedance curve of the spiral antenna **411**.

[0028] The outputs from the antenna modules **410** to **490**, including the spiral antennae **411** and the patch antennae **414** and **415** of each antenna module **410** to **490**, are output to the layer interconnect circuit **110**, which is a printed circuit in electrical communication with the array **120**. In one embodiment, the layer interconnect circuit **110** is printed on the same substrate in which the antennae modules **410** to **490** are located. The substrate **401** may be, for example, a printed wiring board.

[0029] FIG. **5** illustrates a cross-section view of layers of an antenna assembly **500** according to an embodiment of the invention. The antenna assembly includes a first spiral antenna layer **510** including a substrate **511**, spiral antennae **512***a,* **512***b* and **512***c* and patch antennae **516**. The substrate **511** may be made of any electrically insulating material, such as a printed circuit board (PCB), resin, semiconductor material, or any other electrically insulating material. The antennae **512***a,* **512***b* and **512***c* and the patch antennae **516** may be made of any conductive material, such as copper, aluminum, gold, silver, or any other conductive material or alloy. The antenna **512***a,* **512***b* and **512***c* and the patch antennae **516** may be pre-formed and deposited on the substrate **511**, may be formed by photolithography or chemical deposition on the substrate **511**, or may be formed or deposited by any other process.

[0030] Below the first spiral antenna layer **510** is a first thin dielectric layer **520**. Below the first thin dielectric layer **520** is a first balun layer **525**. The first balun layer **525** includes balun wiring **527** on a substrate **526**. Below the first balun layer **525** is a second thin dielectric layer **530**. Below the thin dielectric layer **530** is a ground layer **535**, including a conductive ground plate or wiring.

[0031] The antenna assembly **500** further includes a second spiral antenna layer **540** including a substrate **541**, spiral antennae **542***a,* **542***b* and **542***c* and patch antennae **546**. Above the second spiral antenna layer **540** is a third thin dielectric layer **550**. Above the third thin dielectric layer **550** is a second balun layer **555**. The second balun layer **555** includes balun wiring **557** on a substrate **556**. Above the second balun layer **555** is a fourth thin dielectric layer **530**. Above the fourth thin dielectric layer **560** is the ground layer **535**.

[0032] As illustrated in FIG. **5**, all of the layers **510**, **520**, **525**, **530**, **535**, **540**, **550**, **555**, and **560** have holes in them to form channels **513***a,* **513***b* and **513***c* extending between the spiral antennae **512***a* to **512***c* and the spiral antennae **542***a* to **542***c.* A rectifier **515***a,* **515***b* or **515***c* is located along wires **514***a,* **514***b* and **514***c* extending between the spiral antennae **512***a* to **512***c* and the spiral antennae **542***a* to **542***c.* In one embodiment, the rectifiers **515***a* to **515***c* are Schottky diodes. Although the rectifiers **515***a* to **515***c* are illustrated as being within the channels **513***a* to **513***c,* embodiments of the invention encompass any configuration in which the rectifiers **515***a* to **515***c* are located electrically between the spiral antennae **512***a* to **512***c* and the spiral antennae **542***a* to **542***c.* For example, the rectifiers **515***a* to **515***c* may be circuit elements mounted onto, or within, one of the dielectric layers **520**, **530**, **550**, or **560**, or mounted on the substrates **526** or **557**. In addition, while FIG. **5** illustrates a wire or line **514** extending through the channel **513***a*, embodiments of the invention encompass any method of electrical connection, such as providing a conductive filling material to fill the channels **513***a* to **513***c*.

[0033] While the connections between the spiral antennae **512***a* to **512***c* and the spiral antennae **542***a* to **542***c* are illustrated in FIG. **5**, it is understood that the antenna assembly **500** includes additional electrical and physical connections between various layers, including between the spiral antennae **512***a* to **512***c* and the first balun circuit **527**, or balun wiring layer, and between the first balun circuit **527** and the first ground plate **537**.

[0034] While the terms "above" and "below" have been used to describe the layers in view of FIG. **5**, it is understood that embodiments of the invention encompass devices having layers arranged as in FIG. **5**, but having any orientations, such as at an angle greater than zero with respect to FIG. **5**. In addition, embodiments of the invention encompass antenna assemblies having additional layers or layers omitted, relative to FIG. **5**. In one embodiment, the total thickness of the antenna assembly, such as the assembly **500** of FIG. **5**, satisfies the following: $0.006\lambda o<h<=0.05\lambda o$, where λo is the wavelength of a radio frequency (RF) wave in free space.

[0035] FIG. **6** illustrates a cross-section view of layers of an antenna assembly **600** according to an embodiment of the invention. The antenna assembly includes a first spiral antenna layer **610** including a substrate **611** and spiral antennae modules **612***a*, **612***b* and **612***c*. The substrate **611** may be made of any electrically insulating material, such as a printed circuit board (PCB), resin, semiconductor material, or any other electrically insulating material. The antennae modules **612***a*, **612***b* and **612***c* each include a dielectric insulating layer **615***a*, **615***b* or **615***c* which, together with the substrate **611**, forms a cavity **617***a*, **617***b* or **617***c*. A spiral antenna **613***a*, **613***b* or **613***c* is formed on the dielectric insulating layer **615***a*, **615***b* or **615***c*. Patch antennae **614***a*, **614***b* or **614***c* are formed to surround the spiral antenna **613***a*, **613***b* or **613***c* on the surface of the dielectric insulating layer **615***a*, **615***b* or **615***c*. The patch antennae **614***a*, **614***b* and **614***c* are also configured to surround lateral sides of the dielectric insulating layers **615***a*, **615***b* or **615***c*. In other words, a portion of the patch antennae **614***a*, **614***b* or **614***c* is located on the same surface of the dielectric insulating layers **615***a*, **615***b* or **615***c* as the spiral antennae **613***a*, **613***b* or **613***c*, and a portion of the patch antennae **614***a*, **614***b* or **614***c* is located on lateral sides of the dielectric insulating layers **615***a*, **615***b* or **615***c*.

[0036] Below the first spiral antenna layer **610** is a first thin dielectric layer **620**. Below the first thin dielectric layer **620** is a first balun layer **630**. Below the first balun layer **620** is a second thin dielectric layer **630**. Below the second thin dielectric layer **630** is a ground layer **640**, which may be a conductive plate. The antenna assembly **600** further includes a second spiral antenna layer **670** including a substrate **671** and spiral antennae modules **672***a*, **672***b* and **672***c*. The substrate **671** may be made of any electrically insulating material, such as a printed circuit board (PCB), resin, semiconductor material, or any other electrically insulating material. The antennae modules **672***a*, **672***b* and **672***c* each include a dielectric insulating layer **675***a*, **675***b* or **675***c*. A spiral antenna **673***a*, **673***b* or **673***c* is formed on the dielectric insulating layer **675***a*, **675***b* or **675***c*. Patch antennae **674***a*, **674***b* or **674***c* are formed to surround the spiral antenna **673***a*, **673***b* or **673***c* on the surface of the dielectric insulating layer **675***a*, **675***b* or **675***c*. The patch antennae **674***a*, **674***b* and **674***c* are also configured to surround lateral sides of the dielectric insulating layers **675***a*, **675***b* or **675***c*.

[0037] As illustrated in FIG. **6**, all of the layers **610**, **620**, **630**, **640**, **650**, **660** and **670** have holes in them to form channels **681**. For purposes of clarity, only the channel **681** extending between the spiral antennae **613***a* and **673***a* is labeled in FIG. **6**, but each opposing pair of antennae **613***a* and **673***a*, **613***b* and **673***b*, and **613***c* and **673***c* is connected by a conductive line **682** extending through channels **681** in the layers **610**, **620**, **630**, **640**, **650**, **660** and **670**. In one embodiment, an insulator **683** surrounds the conductive line **682** within the channel **681** at locations corresponding to the ground layer **640** and balun layers **620** and **650** to prevent electrical contact between the conductive line **682** and the layers **640**, **620** and **650**.

[0038] In one embodiment, a rectifier **616***a*, **616***b* or **616***c*, such as a Schottky diode, is electrically connected to a center of the spiral antennae **613***a*, **613***b* or **613***c*. For example, the Schottky diode may be surface mounted to the center of the spiral antennae **613***a*, **613***b* or **613***c*. The rectifier **616***a*, **616***b* or **616***c* is located along the electrical path between the spiral arms of the spiral antennae **613***a* to **613***c* and the spiral antennae **673***a* to **673***c*.

[0039] In one embodiment, the spiral antenna modules **612***a* to **612***c* and **672***a* to **672***c* are readily connectable and disconnectable from the conductive line **682**. As illustrated in FIG. **6**, the conductive line **682** has a connector **684** that is located to connect to a connector **685** of each spiral antenna module **612***a* to **612***c* and **672***a* to **672***c*. The connectors **684** and **685** may be of any connection type, such as a male-female connector in which one connector is inserted into another. In alternative embodiments, the spiral antenna modules **612***a* to **612***c* and **672***a* to **672***c* may be connected to the substrates **711** and **771** or the conductive line **682** by adhesives, soldering or any other permanent connection process.

[0040] While the various layers and other features have been illustrated in FIG. **6** with different thicknesses, it is understood that the thicknesses and lengths of materials and features of FIG. **6** are provided by way of illustration only, and embodiments of the invention are not limited to the specific widths, lengths, thicknesses or relative dimensions illustrated in FIG. **6**.

[0041] FIG. **7** illustrates a first balun layer **525** according to an embodiment of the invention. The balun layer illustrated in FIG. **7** may also correspond to the first balun layer **620** illustrated in FIG. **6**. The first balun layer **525** includes the substrate **526** and the balun wiring **527**. The balun wiring **527** may be any conductive material, such as copper, aluminum, etc. The balun wiring **527** may be printed on the substrate **526** via photolithography, etching, deposition, or any other method. The balun wiring **527** includes an input **528** from the first spiral antenna layer **510**, and in particular from the spiral antennae **512***a*, **512***b* and **512***c*. The balun wiring **527** also includes outputs **529** to a harvester circuit (not shown in FIG. **7**). The balun wiring **556** is also connected to ground, such as the ground plate **537**. The balun wiring **527** is configured to match the impedance of the spiral antennae **512***a* to **512***c* with the harvester circuit.

[0042] FIG. **8** illustrates a second balun layer **555** according to embodiments of the invention. The balun layer illustrated in FIG. **8** may also correspond to the second balun layer **660** of FIG. **6**. The second balun layer **555** includes a substrate **557** and balun wiring **556**. The balun wiring **556** includes an input

558 from the second spiral antenna layer **540**, and in particular from the spiral antennae **542***a,* **542***b* and **542***c*. The balun wiring **556** also includes outputs **559** to a harvester circuit (not shown in FIG. **8**). The balun wiring **556** is also connected to ground, such as the ground plate **566**. The balun wiring **556** is configured to match the impedance of the spiral antennae **542***a* to **542***c* with the harvester circuit.

[0043] In embodiments of the invention, the balun wirings **527** and **556** comprise an "unbalanced" port converted into two "balanced" ports. The unbalanced port can either be the input or the output as the device's balanced ports conversely can be either the output or the input. The balun wirings **527** and **556** are designed to have a precise 180-degree phase shift, with minimum loss and equal balanced impedances. The balun wirings **527** and **556** provide technical benefits of improving isolation between the antenna arrays of the antenna layers **510** and **540**, improving the impedance match between the antenna arrays and a harvesting circuit, improving balance/unbalance transformation characteristics, and decreasing noise interaction between the ports of the balun wirings **527** and **556**. The baluns **527** and **556** are configured to transfer the input power from an unbalanced line (i.e. from the spiral antennae **512***a* to **512***c* and **542***a* to **542***c*) to a balanced transmission line at the output. The balun wirings **527** and **556** comprise a quarter-wavelength transformer, two branches or delay lines and output lines.

[0044] The balun wirings **527** and **556** use electro-magnetic coupling to provide an equal level of power at both the output ports but with a phase difference of 180 degrees over a wide frequency range. The balun wirings **527** and **556** include $\lambda/4$ and $\lambda/2$ elements (where λ is the center frequency wavelength) with alternate interdigital capacitor-like structures formed by micro strip line projections from ground (such as the ground plates **537** and **566**) coupling with alternate loops of microstrip lines. For improved bandwidths multiple sections of half wavelength lines are interconnected by quarter wavelengths. The equivalent electrical model would be series inductors with parallel capacitors whose one end is connected ground. As a whole they form coupled line couplers. The coupled line couplers include of two quarter wavelength lines with capacitive coupling between them. The coupling between the two lines is the result of interaction of electromagnetic fields of each line. Using the planar parallel and series inductor and capacitor circuits to realize the effective open and short circuits respectively.

[0045] As illustrated in FIGS. **7** and **8**, the input lines **528** and **558** of the balun wirings **527** and **556** form two sets of loop elements. One set of loop elements has a constant $\lambda/4$ length, resulting in a balanced structure. The other set of loop elements has 180 degree out-of-phase structure, resulting in an unbalanced output along with pi-structure formation with unbalanced $\lambda/4$ and $\lambda/2$ elements. The resulting balun wirings **527** and **556** each have four ports, two input ports and two output ports.

[0046] Referring to FIGS. **4**, **5**, **7** and **8** the layer interconnect circuit **110** of FIG. **4** connects the outputs from the spiral antennae modules **410** to **490**, including the spiral antennae **411** and the patch antennae **414** and **415** to each other and to the balun layers **525** and **555**. In one embodiment, the outputs **412** of all of the spiral antennae **411** of each planar array **120**, are interconnected. For example, the array **120** illustrated in FIG. **4** is one array corresponding to the first spiral antenna layer **510** of FIG. **5** and the second spiral antenna layer **540** includes a similar array having spiral antenna with spirals of an opposite direction. In such an embodiment, all of the spiral antennae of the first spiral antenna layer **510** are connected to each other, and all of the spiral antennae of the second spiral antenna layer **540** are connected to each other. The outputs from the spiral antennae of the first spiral antenna layer **510** comprise one input **528** of the first balun wiring **525** and one input **558** of the second balun wiring **555**. Similarly, the outputs from the spiral antennae of the second spiral antenna layer **540** comprise one input **528** of the first balun wiring **525** and one input **558** of the second balun wiring **555**.

[0047] In one embodiment, the wiring of the interconnect circuit **110** connects the spiral antennae of the first spiral antenna layer **510** and the spiral antennae of the second spiral antenna layer **540** together in series, such that all of the spiral antennae of the antenna assembly **500** are connected in series, and ends of the spiral antennae are connected to the balun wirings **527** and **556**. Similarly, in one embodiment, all of the patch antennae **516** and **546** are connected in series, and ends of the series of patch antennae are connected to the balun wiring layers **527** and **556**.

[0048] FIG. **9** illustrates a block diagram of a harvester assembly **900** according to an embodiment of the invention. The assembly **900** includes an antenna array **901** connected to a voltage summation circuit **902**. The voltage summation circuit **902** may include an operational amplifier (op-amp). The antenna array **901** gathers energy from microwaves each antenna of the antenna array **901** transmits the gathered energy to the voltage summation circuit **902** to combine the gathered voltage into one power signal. The antenna array **901** may correspond to the antenna array **120** of FIGS. **1** and **4**, for example. The gathered energy is harvested via a resonator circuit **903** and voltage multiplier **904**, such as a two-stage Dickson Charge pump. The resonator circuit **903**, also referred to as a resonator tank, ensures that sufficient voltage swing is supplied in the power signal from the voltage summation circuit **902** to drive the rectifier stage **905**.

[0049] The resulting signal is input to a rectifier stage **905**. In one embodiment, the rectifier stage **905** includes a three or four stage rectifier circuit. Since some power loss occurs in the rectifier circuit, due to the energy required to drive the diode junction barrier, a capacitor **906** may be provided between the rectifier **905** and a voltage regulator **907**. The capacitor **906** may be selected such that the charging time of the capacitor is not more than one to three seconds, and the charge level is high. In one embodiment, multiple capacitors, or multiple super-capacitors, are connected in series. The rectifier **905** converts the RF power from the antenna array **901** to DC power. The voltage regulator **907** receives the rectified voltage from the rectifier stage **905** and outputs a regulated voltage at a predetermined level. The power from the voltage regulator **907** is output to a load, such as a battery.

[0050] In one embodiment, the harvester assembly **900** includes one or more microcontrollers configured to calculate a power input level and to auto-tune a matching circuit for available RF signals at a particular frequency.

[0051] Embodiments of the invention encompass rectanna antennae, and in particular, arrays of spiral antennae that are parallel to each other, that spiral in opposite directions and that are connected to each other. A ground plane is provided between the arrays of antennae to isolate the electric fields of the antennae. A dielectric layer is provided between the ground plane and the arrays of antennae to increase structural strength and to increase a gain vs. frequency of the assembly.

[0052] A rectifier, such as a Schottky diode, connects the centers of the spiral antennae of different arrays via a hole that extends through the antenna array assembly, including ground layers, dielectric layers and balun layers. In some embodiments, the Schottky diode itself is used for the matching and filtering of the output signal. However, in some embodiments of the invention, a transmission-line matching circuit is provided to match the optimal diode impedance for all frequencies, according to the maximum power transfer theorem.

[0053] The dimensions of the spiral antennae may be adjusted according to desired design characteristics. In one embodiment, the diameter of the rectenna elements (or spiral antennae) satisfies the following: $0.3333\ \lambda o<L<0.5\ \lambda o$), where L is the diameter of each rectenna element and λo is the free space wavelength of a radio frequency (RF) wave. To obtain a tilt beams, the outermost periphery of the rectenna array must be more than $2\lambda^8$. The tilted beam is obtained by superimposing a radiation field between a first active region, such as a spiral antenna having a circumference of $1\lambda^g$ and a second active region, such as a spiral antenna having a circumference of $2\lambda^g$. Here, λ^g=a wave length in a dielectric medium of an antenna assembly, such as the antenna assembly illustrated in FIG. **5**. In embodiments of the invention, the dielectric constant, ϵr, of the substrate in which the antennae are housed, such as the substrate **511** of FIG. **5**, satisfies the following: $0.2<\epsilon r<12$. In addition, the dielectric thickness, h, satisfies the following: $0.003\lambda o<h<0.05\lambda o$.

[0054] The number of turns of the spiral antennae depends on an antenna impedance calculation that matches with the optimal diode impedance. A spiral antenna having a larger number of turns with optimal spacing has more spectral efficiency and bandwidth. The circumference of the radiation zone determines the radiation frequency. In embodiments of the invention, the circumference of the radiation zone is greater than 2×λo, where λo is the wavelength of the RF wave in free space.

[0055] There are two possible excitations for spiral antennae: out of phase excitation (balanced) and in-phase excitation. The out of phase excitation creates the so-called normal or the first spiral mode and the in-phase excitation creates the second mode. Since the normal mode has a single lobe broadside pattern, it requires a balanced input. This can be achieved by using a broadband balun, such as the balun layers **525** and **555** of FIG. **5**.

[0056] While the invention has been described in detail in connection with only a limited number of embodiments, it should be readily understood that the invention is not limited to such disclosed embodiments. Rather, the invention can be modified to incorporate any number of variations, alterations, substitutions or equivalent arrangements not heretofore described, but which are commensurate with the spirit and scope of the invention. Additionally, while various embodiments of the invention have been described, it is to be understood that aspects of the invention may include only some of the described embodiments. Accordingly, the invention is not to be seen as limited by the foregoing description, but is only limited by the scope of the appended claims.

1. A device for harvesting radio frequency (RF) energy to generate power, comprising:

a first spiral antenna having a first spiral direction, the first spiral antenna having a spiral cross-section shape along a first flat plane;

a second spiral antenna having a second spiral direction opposite the first spiral direction, the second spiral antenna having a spiral cross-section shape along a second plane parallel to the first plane; and

a ground layer between the first and second spiral antennae, the first spiral antenna electrically connected to the second spiral antenna at a center of the first and second spiral antennae by a conductive element passing through the ground layer without contacting a ground plate of the ground layer.

2. The device of claim **1**, wherein the conductive element includes a Schottky diode.

3. The device of claim **1**, wherein the ground layer defines a third plane parallel to the first flat plane and the second plane.

4. The device of claim **1**, further comprising:

a first dielectric layer surrounding the first spiral antenna along the first flat plane; and

a first flat antenna mounted to the first dielectric layer, the first flat antenna defining a fourth plane parallel to the first flat plane.

5. The device of claim **4**, wherein the first dielectric layer has a thickness greater than a thickness of the first spiral antenna, such that the fourth plane does not intersect the first flat plane.

6. The device of claim **1**, further comprising:

a first balun layer between the first spiral antenna and the ground layer, the first balun layer electrically connected to the first spiral antenna and the ground layer; and

a second balun layer between the second spiral antenna and the ground layer, the second balun layer electrically connected to the second spiral antenna.

7. The device of claim **6**, further comprising:

a harvester circuit connected to the outputs of the first and second balun layers, the harvester circuit configured to receive a radio frequency (RF) signal gathered by the first and second spiral antennae and to generate a direct current (DC) signal based on the RF signal.

8. The device of claim **1**, wherein the first spiral antenna includes a first array of multiple first spiral antennae arranged along the same first flat plane, and

the second spiral antenna includes a second array of multiple second spiral antennae arranged along the same second plane.

9. The device of claim **8**, wherein the multiple first spiral antennae and the multiple second spiral antennae are arranged in series electrically.

10. A device for harvesting radio frequency (RF) energy to generate power, comprising:

a first array of spiral antennae arranged along a first flat plane;

a second array of spiral antennae arranged along a second plane parallel to the first plane;

a ground layer arranged between the first array and the second array; and

a harvester circuit configured to receive an RF signal based on RF energy gathered by the first and second arrays of spiral antennae and to generate direct current (DC) power based on the RF signal.

11. The device of claim **10**, wherein each spiral antenna of the first array has a first spiral direction and each spiral antenna of the second array has a second spiral direction opposite the first spiral direction.

12. The device of claim **10**, wherein the each spiral antenna of the first array is connected to a respective spiral antenna of the second array via a conductive element extending through a hole in the ground plane.

13. The device of claim **12**, wherein the conductive element includes a Schottky diode.

14. The device of claim **10**, wherein the first array includes a first dielectric layer surrounding the first spiral antennae of the first array along the first flat plane; and

an array of flat antenna mounted to the first dielectric layer, the array of first flat antennae defining a fourth plane parallel to the first flat plane.

15. The device of claim **14**, wherein the first dielectric layer has a thickness greater than a thickness of the spiral antennae of the first array, such that the fourth plane does not intersect the first flat plane.

16. The device of claim **10**, further comprising:

a first balun layer between the first array and the ground layer, the first balun layer electrically connected to the spiral antennae of the first array and the ground layer; and

a second balun layer between the second array and the ground layer, the second balun layer electrically connected to the spiral antennae of the second array.

17. The device of claim **16**, further comprising:

a first dielectric layer between the first array and the first balun layer;

a second dielectric layer between the first balun layer and the ground layer;

a third dielectric layer between the second array and the second balun layer; and

a fourth dielectric layer between the second balun layer and the ground layer.

18. The device of claim **17**, wherein the each spiral antenna of the first array is connected to a respective spiral antenna of the second array via a conductive element extending through holes in each of the ground plane, the first and second balun layers, and the first, second, third and fourth dielectric layers.

* * * * *

Sensors **2012**, *12*, 13636-13663; doi:10.3390/s121013636

OPEN ACCESS

sensors

ISSN 1424-8220

www.mdpi.com/journal/sensors

Article

Optimization of Passive Low Power Wireless Electromagnetic Energy Harvesters

Antwi Nimo *, Dario Grgić and Leonhard M. Reindl

Department of Microsystems Engineering, Laboratory for Electrical Instrumentation,
University of Freiburg, IMTEK, Georges-Köhler-Allee 106, 79110 Freiburg, Germany;
E-Mails: grgic@imtek.de (D.G.); leonhard.reindl@imtek.uni-freiburg.de (L.M.R.)

* Author to whom correspondence should be addressed; E-Mail: antwi.nimo@imtek.de;
Tel.: +49-761-203-7223; Fax: +49-761-203-7222.

Received: 13 August 2012; in revised form: 28 September 2012 / Accepted: 28 September 2012 / Published: 11 October 2012

Abstract: This work presents the optimization of antenna captured low power radio frequency (RF) to direct current (DC) power converters using Schottky diodes for powering remote wireless sensors. Linearized models using scattering parameters show that an antenna and a matched diode rectifier can be described as a form of coupled resonator with different individual resonator properties. The analytical models show that the maximum voltage gain of the coupled resonators is mainly related to the antenna, diode and load (*remote sensor*) resistances at matched conditions or resonance. The analytical models were verified with experimental results. Different passive wireless RF power harvesters offering high selectivity, broadband response and high voltage sensitivity are presented. Measured results show that with an optimal resistance of antenna and diode, it is possible to achieve high RF to DC voltage sensitivity of 0.5 V and efficiency of 20% at −30 dBm antenna input power. Additionally, a wireless harvester (*rectenna*) is built and tested for receiving range performance.

Keywords: RF energy harvesting; wireless power transmission; coupled resonators; Schottky diode; RF to DC power converter; impedance matching; PI-matching; L-matching; rectenna

---

1. Introduction

For autonomous powering of sensor nodes in remote or inaccessible areas, wireless power transfer provides the only viable option to power them from an energy source. Due to the low power density of ambient RF at far-field from transmitters, there is a need to optimize each aspect of a wireless RF energy harvester for possible realistic applications. Today remote autonomous sensors are mostly powered by batteries, which have limited lifespan. Renewable powering has the potential to power autonomous sensors perpetually. Due to the expansion of telecommunications technology ambient electromagnetic (EM) power is among the most common sources of ambient energy. There are power transmitters/receivers scattered in practically any society, ranging from television transmission stations to cell phone transmitters and even wireless routers in our homes/offices or mobile phones. These transmitters in our environment and others which are on special dedicated frequencies produce ambient RF power (on the order of microwatts) which can be used as a source for powering remote microwatt budget sensors through wireless energy harvesting. This work presents different matching techniques based on different application requirements using Schottky diode-based RF to DC power converting circuits for wireless remote EM energy harvesting around 434 MHz and 13.6 MHz. Generalized analytical models and limitations of the matched RF to DC power converters are discussed. A wireless RF energy harvester consisting of an antenna and a matched diode rectifier is then realized and its performance tested. Passive wireless energy harvesting also finds applications in near field communications (NFC) [1], RFID tags [2–5], implantable electronics [6,7], and environmental monitoring [8], among others.

1.1. State of the Art

Hertz was the first to demonstrate the propagation of EM waves in free space and to demonstrate other properties of EM waves such as reflection using parabolic reflectors [9]. Wireless power transmission was then investigated and demonstrated for possible wireless remote powering by Tesla. Electromagnetic power beaming for far field wireless power transfer using collimated EM waves was proposed in the 1950s [9]. Recent advances in ultralow power sensors means ambient omni-directional EM power can be used as a source for powering remote sensors without the need to collimate the EM power through the wireless space. Mickle [10] and McSpadden [11] have presented earlier work on wireless energy harvesting systems using Schottky diodes and rectennas where the usability of ambient RF power into DC power was shown. Sample [12] presented a wireless harvester which can harvest EM power from TV and radio base stations transmitting 960 kW of effective radiated power; 60 μW was harvested at a range of about 4 km. Umeda [13] and Le [14] have presented more integrated wireless energy harvesters based on CMOS RF to DC rectifying circuits. CMOS-based rectifying power converters provide full compatibility with standard CMOS technologies and have advantages in batch processes for mass production. The drawback of CMOS-based diode connected transistors is the need to bias the gate of the transistors for the rectifying circuits to effectively function. This gate bias is provided externally, which makes the system not passive. Without the injection of external charges or a biasing of the transistor gate, the circuit has low efficiency, especially when the amplitude of the input voltage is low [15]. Shameli [2] presented a passive CMOS RF to DC power converter with a voltage sensitivity of 1 V at −14.1 dBm input, but the circuit efficiency was only 5 %. Zbitou [16] presented an RF to DC converter based on Schottky diodes and achieved 68 % efficiency at 20 dBm RF input power. Ungan [17,18] presented antennas and high quality factor RF to DC power converters at 24 MHz and 300 MHz for RF wireless energy harvesting at -30 dBm input power. The power converter used high quality factor resonators for impedance matching the EM source and the diodes and achieved high open circuit voltage sensitivity of 1 V/μW. Boquete [19] presented a risk assessment system for calculating insurance premiums by monitoring mobile phone usage while driving. This was done by harvesting EM power from detected mobile phone usage during driving for risk assessment. Heikkinen [20] presented rectennas on different substrates at 2.4 GHz using transmisson lines to match the antennas output resistance (*at resonance*) to the rectifying diodes. Akkermans [21] presented a rectenna design by complex conjugating impedance provided by a microstrip structure to a diode so that resonance may be achieved for a working frequency. This design approach may need sophisticated tools to realize and the dominant resonance frequency of the rectenna can be unpredictable in practice. Hagerty [22] presented rectenna arrays for broadband ambient EM harvesting and characterized the harvesters from 2 GHz to 18 GHz; rectennas combine impedance matching the RF rectifying circuit and the antenna into one compact device, but an array of rectennas may increase the overall size of an EM harvester. Herb [23] and Vullers [24] have provided a comprehensive state of the art for micro energy harvesting and have explored the various techniques used for harvesting ambient renewable energy.

2. RF to DC Power Converter

2.1. Diode Rectifier

A junction diode equivalent circuit and simple Schottky diode rectifier are shown in Figure 1. R_{DS} is the diode resultant series resistance, C_{DS} is the diode resultant series capacitance, R_{DP} is the diode resultant parallel resistance, C_{DP} is the diode resultant parallel capacitance, V_S is the sinusoidal source voltage and V_C is the voltage across the capacitor.

Figure 1. (**a**) Diode series equivalent model, (**b**) Diode parallel equivalent model, (**c**) Simple diode detector.

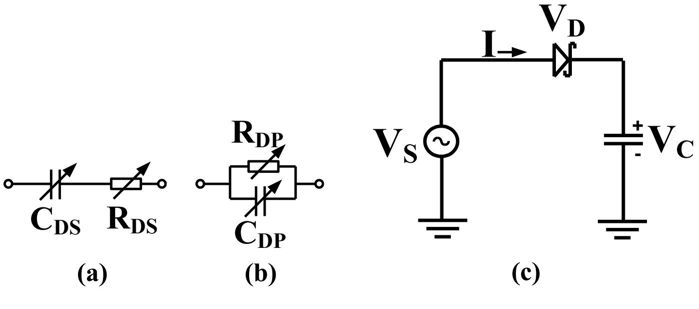

The diode capacitive impedance is mainly due to the junction capacitances provided by the metal, its passivation and the semiconductor forming the diode. AC power incident on a forward biased diode input is converted to DC power at the output. The current-voltage behavior of a single metal/semiconductor diode is described by the Richardson equation [25] as in Equation (1):

$$I = I_S\left(e^{\left(qV_D/nKT\right)} - 1\right) \tag{1}$$

where I is the current through the diode, I_S is the saturation current, q is the charge of an electron, V_D is the voltage across the diode, T is the temperature in degrees Kelvin and K is Boltzmann constant. The voltage equation around the loop can be derived from Figure 1(c) and is given in Equation (2):

$$V_D = V_S - V_C \tag{2}$$

Since the same current flows through the diode and the capacitor, one can find the average current through the circuit by integrating Equation (1) over a time period. By substituting Equation (2) into Equation (1), V_C can be expressed in terms of V_S by averaging the diode current to zero. This is given in Equation (3) [26]:

$$V_C = \frac{KT}{q}\ln\left[\vartheta_0\left(\frac{qV_S}{KT}\right)\right], \tag{3}$$

where ϑ_0 is the series expansion of the sinusoidal source voltage. Equation (3) can further be simplified for very small amplitude V_S as Equation (4):

$$V_C \approx \frac{qV_S^2}{4KT} \tag{4}$$

Equation (4) shows that for a small voltage source, the circuit output voltage is proportional to the square of the input sinusoidal voltage; hence it's so-called square law operation. Extensions of this model for voltage multipliers and other input signals are presented in [27] and [28]. Equation (4) further confirms that for low input voltage (power ≤ 10 dBm), an impedance matching network between the source and the diode is necessary to improve the detected output voltage and efficiency.

2.2. Impedance Matching

The maximum power transfer theorem states that the highest power is transferred to the load when the source resistance is the same as the load resistance. For systems with both resistive and reactive impedances from source and load, the source and the load impedance should be adjusted in a way that they are the complex conjugate of each other through impedance matching. For the purposes of this work, a 50 Ω resistive source is chosen as reference for load impedance matching. The antenna which captures the ambient RF signals is tuned to provide this source resistance at resonance for the rectifying circuit in a complete EM wireless remote harvester. The load is the resistance of the Schottky diodes and the actual connected resistance (*remote sensor*). The specific type of matching network which can be used for complex conjugation depends on the nature of load or source impedance, the desired RF to DC converter functionality and other factors like circuit size, cost, *etc.* The response of a matched RF to DC power converter depends on the matching network used as well as the source or load component quality factors and impedances.

2.3. Diode Impedance

Schottky diodes HSMS-285C and HSMS-286C from Avago [29,30] are used to build the RF to DC power converters. The HSMS-285× or 286× series diodes can be operated as zero biased with relatively low forward junction potential. This allows for the realization of completely passive RF to DC power converters for wireless energy harvesting. The HSMS-285C or 286C is a pair of series connected Schottky diodes in a SOT-323 package. The impedance of the HSMS-285C and HSMS-286C diodes was first measured so it can be matched to the resistance (50 Ω) of the antenna source. This is done by connecting the input of the diodes to a network analyzer and measuring the scattering parameters. These scattering parameters are then converted to the corresponding impedances. The input impedance of a diode depends mainly on the resistive and capacitive impedance provided by the junction of the diode and its connected load. For a couple of diodes arranged in a package such as the HSMS-285C or 286C, the input impedance is the vector sum of the impedances provided by each diode in the package arrangement, the extra impedance associated with the packaging and the connected load. The diode measuring board is as shown in Figure 2. The diodes were measured at room temperature for an input power of -30 dBm at a diode connected load of 1 MΩ with a 100 pF filter capacitor. For the sake of this work, the input impedance of the diodes will always be referred to at these connected load conditions.

Figure 2. (**left**) Reference circuit layout for measuring diodes input impedance, (**right**) measuring printed circuit board (PCB) for diodes input impedance on 1 mm FR4 substrate.

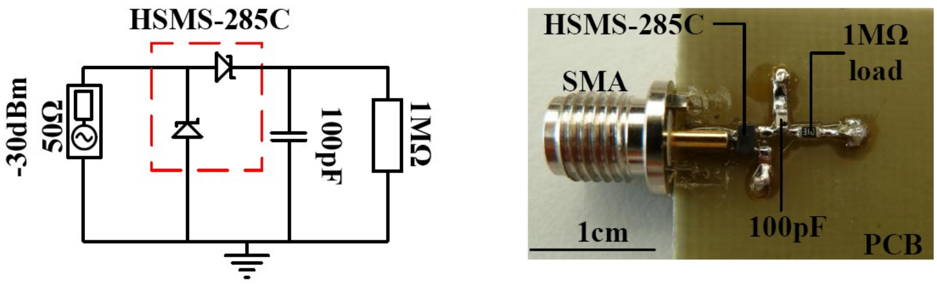

Figure 3. Measured input impedance (Δ resistive, □ capacitive) of HSMS-285C (**left**) and HSMS-286C (**right**) diodes at −30 dBm input with 1 MΩ load and 100 pF filter.

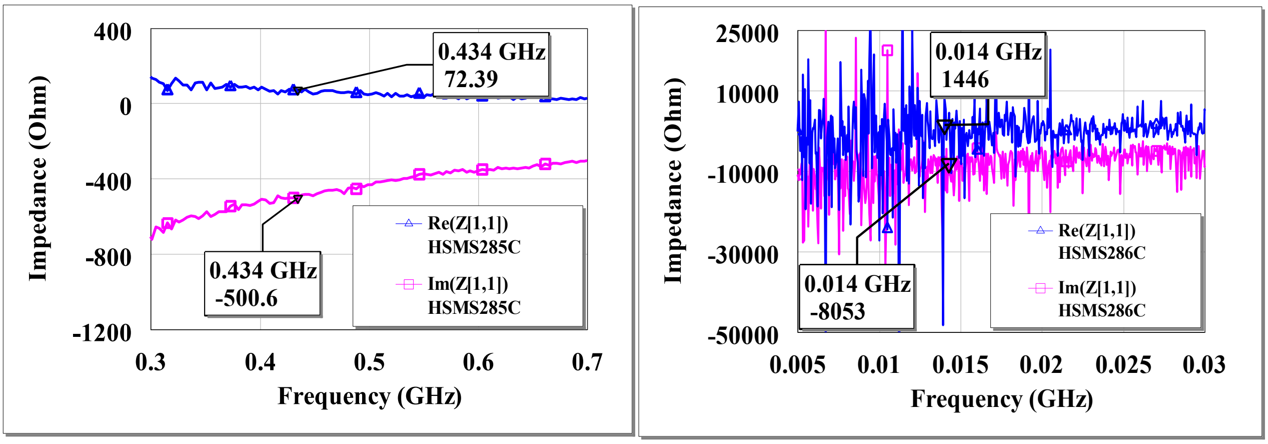

The board is fabricated such that components are soldered directly one into another to prevent additional impedances introduced by copper route. The PCB backside had the ground layer. An example of measured input impedance for HSMS-285C and HSMS-286C is shown in Figure 3.

The diodes quality factor is given by $X_{DS}R_{DS}^{-1}$, where X_{DS} is the resultant series capacitive impedance of the diodes. At an input power of −30 dBm, the measured input impedance of the HSMS-285C diodes is 72–j501 Ω at 434 MHz and 587–j1239 Ω at 13.6 MHz. For HSMS-286C diodes, it is 10–j503 Ω at 434 MHz and ~1.5–j8.1 kΩ at 13.6 MHz for −30 dBm input. The measured impedance of the HSMS-286C diodes at low frequencies (< 60 MHz) shows pronounced fluctuations. The low-frequency excess flicker noise and the shot noise observed in the HSMS-286C have been studied by several authors [31–33]. The pronounced presence of trap states in the depletion region of the semiconductor, mobility fluctuations in carriers, edge effects among other reasons is reported to cause deviations from the ideal Schottky diode behavior and hence generation-recombination noise for some diodes such as the HSMS-286C [34]. When a diode rectifier is matched at a reference operating condition, the matching network may function less effectively at other input power levels, connected load and other operating frequencies. This is due to possible changes in the diode input impedance. Throughout this work the imperfections of the matching circuit at other operating conditions away from the matched reference conditions are accepted without changes to the matching network.

2.4. Voltage Doubler

The Delon voltage doubler and Greinacher doubler are both used to realize the RF to DC power converters presented in this work. The Delon voltage doubler and Greinacher doubler are shown in Figure 4. The diodes output voltage (V_{out}) is doubled what is detected by a simple detector circuit shown in Figure 1. Both doublers produce the same output performance, the only difference is that the Delon doubler has an instantaneous input ground which is not shared with the output.

Figure 4. Circuit diagram of voltage doubler, (**a**) Delon doubler and (**b**) Greinacher doubler.

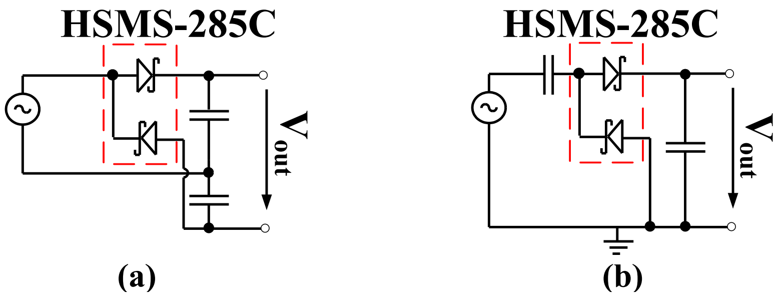

(a) **(b)**

2.5. Matching Techniques for Antenna Source and RF to DC Power Converter

2.5.1. L-match RF to DC Power Converter

An L-match network converts a source series impedance to its equivalent load parallel impedance or *vice-versa* and tunes out by subtracting or adding any surplus reactance from the load or source with the counter impedance. Series impedance is converted to its parallel equivalent impedance using Equations (5–7):

$$Q_S = \frac{X_S}{R_S} \tag{5}$$

$$Q_P = \frac{R_P}{X_P} \tag{6}$$

where X_S is the total series reactive impedance, R_S is the total series resistance, R_P is the total parallel resistance, X_P is the total parallel reactive impedance, Q_S and Q_P are the series and parallel quality factors respectively:

$$R_S + jX_S = \frac{R_P \times jX_P}{R_P + jX_P} \tag{7}$$

Equation (7) is the equation of a series sum of impedances and a parallel sum of impedances. It is interesting to note that Q_S and Q_P from an L-matched network may be different from the individual component quality factors as a result of the inherent resistive and reactive impedances in that component. By virtue of Equation (7), Q_S and Q_P must be equal in an L-matched network. Using Equations (5,6) and (7), the ratio of the parallel resistance (*or reactance*) to the series resistance (*or reactance*) can be derived in terms of the quality factors Q_P or Q_S [35]. Since at match conditions, only the resistive impedances dissipate power, the loaded quality factor Q, of the L-matched network can be expressed as in Equation (8):

$$R_P = (Q^2 + 1)R_S \tag{8}$$

Using Equations (5,6) and (8), series impedance can be converted to its parallel equivalent for a fixed frequency and power level. As an example; a series impedance 72–j501 Ω (HSMS-285C at 434 MHz for −30 dBm input power) is easily converted to −*j*510(3519)/(−*j*510 + 3519) Ω as its parallel equivalent with a component quality factor of 6.96. The source resistance is taken as part of the parallel matching network in an L-match circuit if the source series equivalent resistance is greater than the load series equivalent resistance. On the other hand, the load resistance is taken as part of the parallel matching network if the load series equivalent resistance is greater than the source series equivalent resistance. For the purpose of this work, inductors were only used for series impedance matching and capacitors as shunts. This prevents power seeping through any shunt inductor used for impedance matching due the short circuit provided by a shunt inductor to ground and resulting in less output efficiency. Resistors were not used for impedance matching.

2.5.2. L-match RF to DC Converter Generalized Analytical Model

The classical matching technique using Equations (5,6) and (8) is first used to L-match the 50 Ω resistance of the antenna to the resistance of the HSMS-286C diodes (and load) at 434 MHz for −30 dBm input and then the generalized model is discussed. The antenna source resistance was L-matched to the resistance of the diodes (and load). The 50 Ω resistance of the antenna is taken as the parallel matching component and the diodes 10 Ω resistance is the series matching component. The loaded Q is found as 2 between the 50 Ω antenna source resistance and the 10 Ω diode series resistance using Equation (8). From this loaded Q, a shunt capacitive impedance of 25 Ω (14.6 pF at 434 MHz) using Equation (6) and a series inductive impedance of 20 Ω (7.3 nH at 434 MHz) using Equation (5) will match the 50 Ω source to the 10 Ω HSMS-286C diodes (and load) series resistance at −30 dBm input. Since the HSMS-286C diodes inherently provides 503 Ω series capacitive impedance at −30 dBm, a resultant series inductive impedance of 523 Ω (192 nH at 434 MHz) is needed to tune the 50 Ω resistive source to the complete HSMS-286C diodes impedance at 434 MHz for −30 dBm input. The L-matched HSMS-286C diodes rectifier is as shown in Figure 5(a).

Figure 5. (**a**) L-match RF to DC harvester using the HSMS-286C diodes at 434 MHz for −30 dBm input. (**b**) Small signal impedance model of a generalized L-matched RF to DC power converter as capacitive coupled series RLC resonator with different resonator elements.

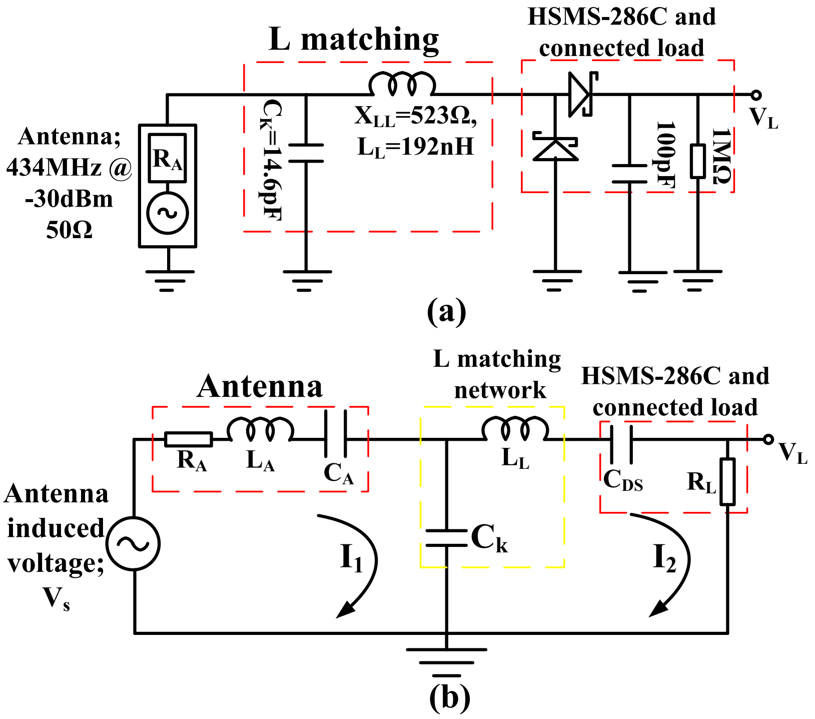
L matching
HSMS-286C and connected load
Antenna; 434MHz @ -30dBm 50Ω
R_A
C_K=14.6pF
X_{LL}=523Ω, L_L=192nH
100pF
1MΩ
V_L
Antenna
L matching network
HSMS-286C and connected load
Antenna induced voltage; V_s
R_A
L_A
C_A
L_L
C_k
C_{DS}
R_L
I_1
I_2
V_L

(a)

(b)

C_K is the tuning capacitance, L_L is the tuning inductance, X_{LL} is the tuning inductive impedance, C_{DS} is the diodes series capacitance, X_{DS} is the diodes series capacitive impedance, V_S is the antenna captured ambient EM voltage, R_A is the resistance of antenna, L_A is the inductance of antenna, C_A is the capacitance of antenna, R_L is the resultant series resistance from the diodes and the connected load resistance, V_L is the resistive load voltage. From Figure 5(a) the power dissipated in the resistance of the diodes (and connected load); P_L is given by Equation (9), where R_L is the series resistance of the diodes and load:

$$P_L = \frac{V_L^2}{R_L} \tag{9}$$

The source power; P_S is given by Equation (10), where V_S* is the root mean squared (RMS) antenna captured source voltage. Half of the source power is transferred to the resistance of the diodes (and connected load) at match conditions as described by the maximum power transfer theorem:

$$P_S = \frac{V_S^2}{R_A} \quad \text{or} \quad P_S = \frac{V_{S*}^2}{2R_A} \tag{10}$$

Equating P_L and half RMS P_S gives a condition of maximum voltage gain for the matched RF to DC power converter shown in Figure 5(a):

$$\frac{V_L}{V_{S*}} = \frac{1}{2}\sqrt{\frac{R_L}{R_A}} \tag{11}$$

From Equation (8), substituting the series and parallel resistance ratio into Equation (11) the voltage gain can be expressed in terms of the loaded quality factor as in Equations (12) and (13), where Q is the loaded quality factor of the RF to DC power converter:

$$\frac{V_L}{V_{S*}} = \frac{1}{2}\sqrt{\frac{1}{1 + Q^2}} \tag{12}$$

Equation (12) is the voltage gain in-terms of the loaded Q if the resistance of the diodes (and connected load) is part of the series matching network and the resistance of the antenna source is part of the parallel matching network as in Figure 5(a). If the resistance of the diodes is part of the parallel matching network, then Equation (13) may be written as the voltage gain in-terms of the loaded Q in an L-matched circuit:

$$\frac{V_L}{V_{S*}} = \frac{1}{2}\sqrt{1 + Q^2} \tag{13}$$

Equations (12) and (13) shows that the maximum voltage gain is directly related to the relative differences between the diodes (and connected load) resistance and source resistance at matched conditions or the circuit loaded quality factor. It is interesting to note that the circuit shown in Figure 5(a) has a loaded Q of 2, but an HSMS-286C unloaded quality factor of 50 (at 434 MHz for −30 dBm).

Figure 5(a) is generally modeled as capacitive coupling of two series RLC resonators with a voltage source. This linearized model can be made at any defined frequency and power level. The model however neglects the metal/semiconductor physics of the diode’s junction potentials which results in a Schottky barrier. The first series RLC resonator is modeled as impedance from the antenna with or without some passive matching components. The voltage source V_S, is the antenna captured electromagnetic voltage. The second series RLC resonator is the impedance from the diodes (at a defined condition), connected resistance and some passive matching components. Ck is modeled as the coupling element between the two series RLC resonators. Figure 5(b) gives a more general look at the special scenario shown in Figure 5(a). The voltage equations in the two loops are given by Equations (14,15) according to Kirchhoff’s voltage loop laws, where ω is the angular frequency and I_1, I_2 are the currents in the first loop and second loop, respectively:

$$V_S = I_1\left[R_A + j\omega L_A - \frac{j}{\omega C_A} - \frac{j}{\omega C_K}\right] + \frac{jI_2}{\omega C_K} \tag{14}$$

$$0 = \frac{jI_1}{\omega C_K} + I_2\left[R_L + j\omega L_L - \frac{j}{\omega C_{DS}} - \frac{j}{\omega C_K}\right] \tag{15}$$

Using Cramers rule, I_2 can be expressed as:

$$I_2 = \frac{\dfrac{-jV_S}{\omega C_K}}{\left[R_A + j\omega L_A - \dfrac{j}{\omega C_A} - \dfrac{j}{\omega C_K}\right]\left[R_L + j\omega L_L - \dfrac{j}{\omega C_{DS}} - \dfrac{j}{\omega C_K}\right] + \dfrac{1}{\omega^2 C_K^2}}. \tag{16}$$

The voltage across R_L is V_L; given by I_2R_L:

$$V_L = \frac{\dfrac{-jV_S}{\omega C_K} R_L}{\left[R_A + j\omega L_A - \dfrac{j}{\omega C_A} - \dfrac{j}{\omega C_K}\right]\left[R_L + j\omega L_L - \dfrac{j}{\omega C_{DS}} - \dfrac{j}{\omega C_K}\right] + \dfrac{1}{\omega^2 C_K^2}} \tag{17}$$

The voltage gain of the coupled resonator can be expressed as in Equation (18):

$$\frac{V_L}{V_S} = \frac{\dfrac{-jR_L}{\omega C_K}}{\left[R_A + j\omega L_A - \dfrac{j}{\omega C_A} - \dfrac{j}{\omega C_K}\right]\left[R_L + j\omega L_L - \dfrac{j}{\omega C_{DS}} - \dfrac{j}{\omega C_K}\right] + \dfrac{1}{\omega^2 C_K^2}} \tag{18}$$

At resonance, there is no resultant reactance in the RLC resonators or the capacitive and inductive impedances become equal; hence Equation (19) can be written:

$$\omega L_A - \frac{1}{\omega}\left\{\frac{1}{C_A} + \frac{1}{C_K}\right\} = 0 \text{ and } \omega L_L - \frac{1}{\omega}\left\{\frac{1}{C_{DS}} + \frac{1}{C_K}\right\} = 0 \tag{19}$$

Equations in Equation (19) can be used to find the resonant frequencies of the series coupled resonator. The voltage gain of the coupled resonator at resonance can then be expressed as in Equation (20):

$$\frac{V_L}{V_S} = V_{gain} = \frac{\dfrac{-jR_L}{\omega C_K}}{R_A R_L + \dfrac{1}{\omega^2 C_K^2}} \tag{20}$$

where V_{gain} is the voltage gain. V_{gain} at resonance is a function of the resistance of the source and load and the coupling element. The maximum of Equation (20) is obtained when:

$$\frac{dV_{gain}}{dC_K} = 0. \tag{21}$$

This gives the results as in Equation (22):

$$\frac{dV_{gain}}{dC_K} = \frac{j2R_L}{\omega^3 C_K^4} - j\left\{R_A R_L + \frac{1}{\omega^2 C_K^2}\right\}\frac{R_L}{\omega C_K^2} = 0 \text{ or } R_A R_L^2 = \frac{R_L}{\omega^2 C_K^2} \tag{22}$$

Equation (22) can be simplified to find $C_{K(max)}$:

$$C_{K(\max)} = \pm\frac{1}{\omega}\sqrt{\frac{1}{R_A R_L}} \tag{23}$$

where C_{Kmax} is the value of the coupling element where maximum power transfer from the first resonator to the second resonator occurs. Using Equations (19) and (23) the unknown optimal matching impedances can be found from the known impedances just like the classical L-matched procedure using Equations (5,6) and (8). By substituting $C_{K(max)}$ into Equation (20) and taking the magnitude of V_{gain}, gives the maximum voltage gain of the coupled series resonator at resonance:

$$\left|\frac{V_L}{V_S}\right| = \frac{1}{2}\sqrt{\frac{R_L}{R_A}} \text{ or simply } \frac{V_L}{V_{S*}} = \frac{1}{2}\sqrt{\frac{R_L}{R_A}} \tag{24}$$

For wireless harvesters consisting of an antenna and a diode rectifying circuit, the diode resistive impedance at any condition is dependent on the diode realized parameters, signal frequency, connected load and the input power level. The source impedance is determined by the impedance of the antenna. For maximum efficiency, the ratio of the source resistance to the load resistance must tend to zero at matched conditions. The efficiency η of the circuit is given by Equation (25):

$$\eta = \frac{P_L}{P_S}\,;\ \eta \rightarrow 1 \ \ when \ \frac{R_A}{R_L} \rightarrow 0 \tag{25}$$

2.5.3. L-Match RF to DC Converter Experimental Results and Discussion

The presented circuit was L-matched between the 50 Ω resistance of the antenna source and the resistance of the HSMS-285C diodes (and load) at 434 MHz for −30 dBm input as shown in Figure 6. Since the series equivalent resistance of the HSMS-285C diodes and load (72 Ω) is greater than the 50 Ω series resistive antenna source, the diode is taken as parallel matching network with a parallel equivalent impedance of $-j510(3519)/(-j510 + 3519)$ Ω. The analysis follows the same procedure as in Section 2.5.2 after this step. Figure 6(b) shows the resultant L-matched RF to DC power converter. C_{DP}* is the resultant shunt matching capacitance.

Figure 6. (**a**) L-matched impedance circuit for matching the HSMS-285C diodes at 434 MHz for −30 dBm input. (**b**) Resultant network, (**c**) PCB layout of the L-matched Delon doubler with adjusted values on FR4 substrate (**d**) Fabricated PCB of the L-network matched Delon voltage doubler.

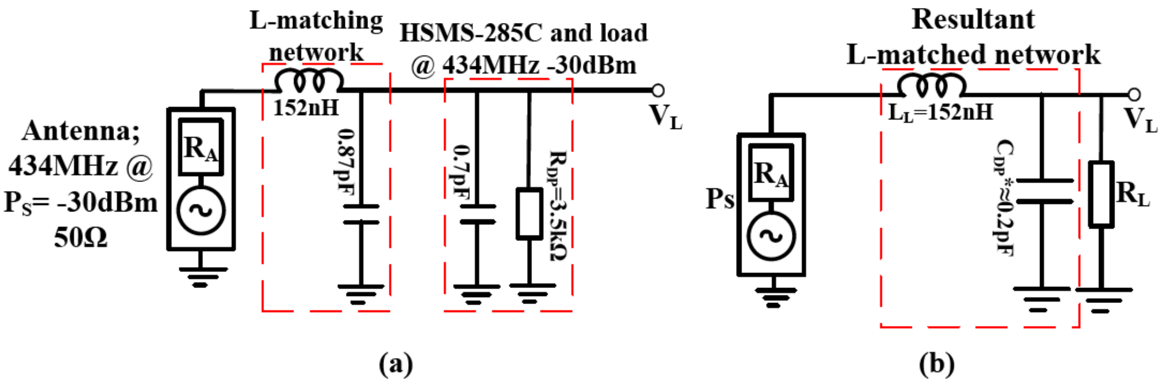

Figure 6. *Cont.*

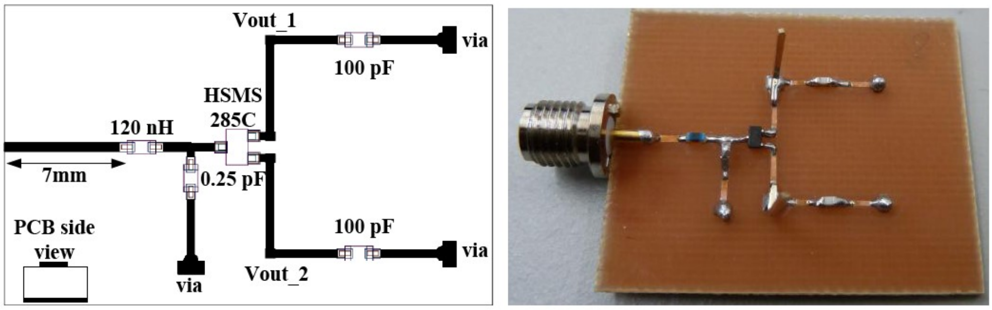

(c) (d)

Figure 6(a,b) assume perfect characteristic impedance between the various components in the matched circuit. When a copper route is introduced between components and on a material substrate, it must be accounted for in the total impedance as seen by the source or load. This PCB impedance compensation is carried out in Advance Design Systems (ADS) from Agilent [36]. ADS has extensive models for microstrip substrates to account for its impedances. The optimized layout using ADS microstrip models and its compensated values in the passive tuning components for a Delon doubler is shown in Figure 6(c).

The circuit reflection coefficient (S_{11}) and input impedance at open circuit are shown in Figure 7. There is high return loss and resonance around 434 MHz. The circuit input impedance at open circuit conditions is ~38 Ω at resonance for −40 dBm and ~17 Ω at −10 dBm input.

The measured L-matched circuit efficiency and voltage sensitivity is shown in Figure 8. The maximum measured L-matched efficiency at −30 dBm is 22% at ~20 kΩ load and an open circuit voltage of 124 mV. At −10 dBm, the maximum efficiency and open circuit voltage is 47% and 2 V respectively. At the optimal load of ~20 kΩ, the detected voltage is 58 mV and 1 V at −30 dBm and −10 dBm respectively.

Figure 7. Measured open circuit S_{11} of the L-matched Delon circuit at different input power levels from a 50 Ω source (**left**), measured open circuit input impedance at −10 dBm and −40 dBm of the L-matched circuit (**right**).

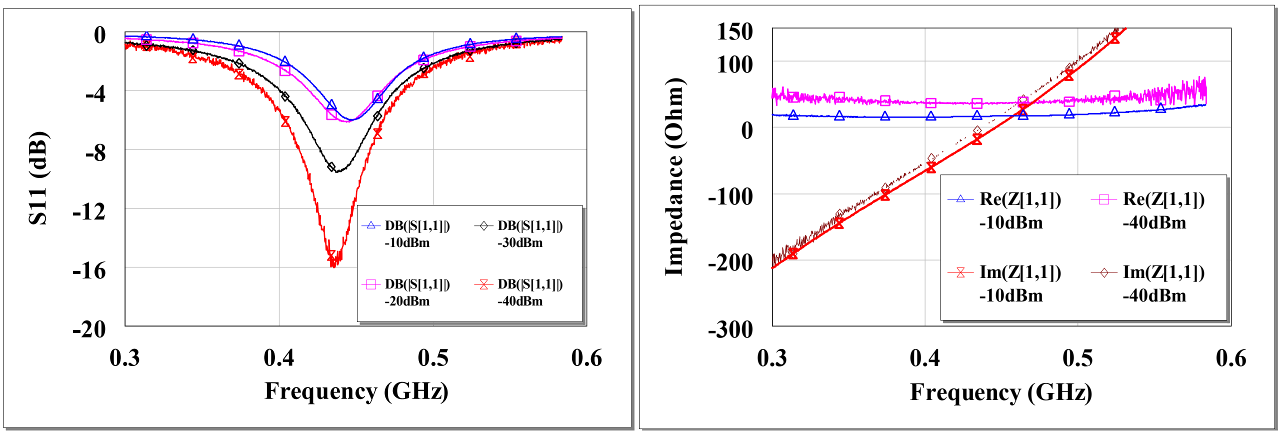

The open circuit voltage gain is 25 at −30 dBm and 40 at −10 dBm. The maximum measured efficiency at −35 dBm is 27%. This is higher than that of −30 dBm due to the better matched circuit impedance at −35 dBm (35 Ω) than at −30 dBm (27 Ω) input. The L-matched RF to DC power converter has a loaded *Q*, sensitivity and efficiency determined mainly by the diodes resistance, diodes junction potential, connected resistance and antenna source resistance at matched conditions.

Figure 8. Measured L-matched circuit efficiency *versus* resistive load at various input power levels at 434 MHz (**left**), measured open circuit voltage and at 17 kΩ load *versus* input power at 434 MHz (**right**).

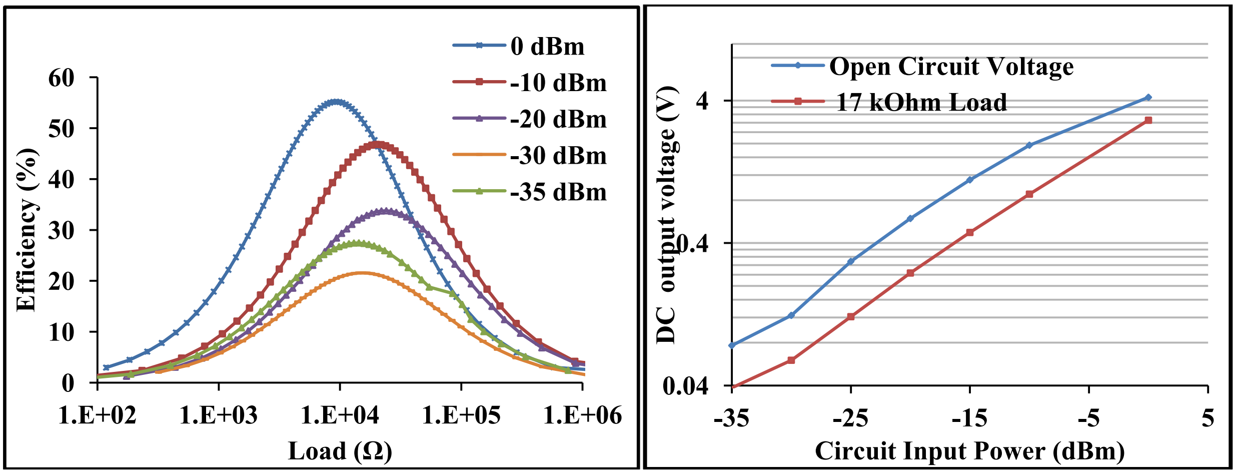

2.5.4. PI-match RF to DC Power Converter

A highly selective or small frequency bandwidth RF power converter is realized with a PI-network in-between the source impedance from the antenna and the diode rectifier. A PI-network is a ‘back to back’ L-network that are both configured to match the load and source impedance to an invisible resistance located at the junction between the two L-networks [37]. The quality factor of the L-network with the parallel resistance is given by Equation (26):

$$Q_P^* = \sqrt{\frac{R_P}{R^*} - 1}, \tag{26}$$

where R_P is the parallel resistance, R^* is a virtual resistance and Q_P^* is the quality factor of the L-network with the parallel resistance. The quality factor of the L-network with the series resistance is given by Equation (27):

$$Q_S^* = \sqrt{\frac{R_S}{R^*} - 1}, \tag{27}$$

where Q_S^* is the quality factor of the L-network with the series resistance. The unloaded quality factor; Q_S^* or Q_P^* is set higher than what is normally achieved with a single L-network [37] to realize the small frequency bandwidth circuit. The resistance of the load is assigned the parallel network in a PI-matched circuit if its series equivalent resistance is higher than the source series equivalent resistance; the opposite is true if the source is higher than the load. Equation (26) and Equation (27) are synonymous to Equation (8), except the lowest resistive impedance in Equation (8) is substituted with the virtual resistance which is dependent on the newly desired circuit selectivity. From Equations (26) and (27) the loaded quality factor of the PI-matched circuit can be written as Equation (34) in terms of Q_S^* and Q_P^*:

$$Q^2 = \left[\left(\frac{Q_P^{*2} + 1}{Q_S^{*2} + 1}\right) - 1\right], \tag{28}$$

where Q is the loaded quality factor of the PI-network. Q_S* or Q_P* are the unloaded quality factors of the PI-matched network. The larger value among the unloaded quality factors result in small frequency bandwidth response which is desired when matching a source and load impedance with a PI-network. Some authors approximate the highest value of Q_S* or Q_P* or their algebraic sum as the loaded quality factor of the PI-network as in [35] and [37], but Equation (28) gives the exact loaded Q of the PI-matched circuit in terms Q_S* and Q_P*. This allows for the correct estimation of the maximum voltage gain from the loaded quality factor.

2.5.5. Selectivity RF to DC Converter Generalized Analytical Model

An example of a PI-matched RF to DC converter using the HSMS-285C diodes operating at 434 MHz for −30 dBm input is presented first and then the generalized model is discussed. The circuit is matched for Q_P* of 60 between the antenna and the resistance of the diodes as shown in Figure 9.

Figure 9. Impedance diagram of PI-matched RF power converter; (**a**) Impedance diagram of 50 Ω source and the HSMS-285C diodes at 434 MHz, (**b**) Resultant PI matched network between the antenna source and load resistance.

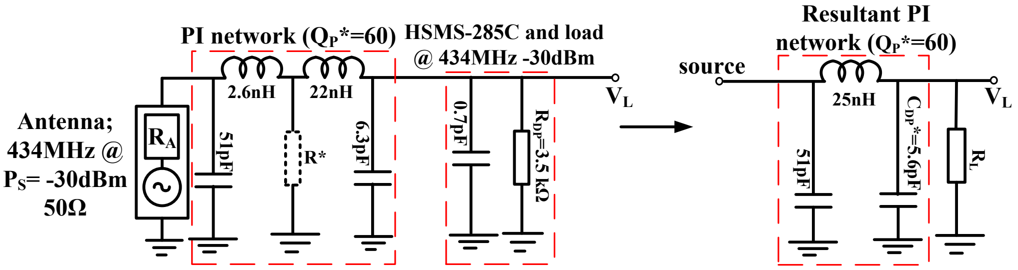
PI network (Q_P*=60)
HSMS-285C and load @ 434MHz -30dBm
Resultant PI network (Q_P*=60)
source
Antenna; 434MHz @ P_S= -30dBm 50Ω
R_A
2.6nH
22nH
51pF
R*
6.3pF
0.7pF
R_{DP}=3.5 kΩ
V_L
25nH
51pF
C_{DP}*=5.6pF
R_L
V_L

(a) **(b)**

Figure 9(b) can also be modeled as an inductive coupling of two parallel RC circuits. A more general look at such a circuit is shown in Figure 10, as an inductive coupling of two parallel RLC resonators with a current source.

Figure 10. Inductive coupled parallel RLC small signal model of a generalized PI-matched antenna and diode rectifier.

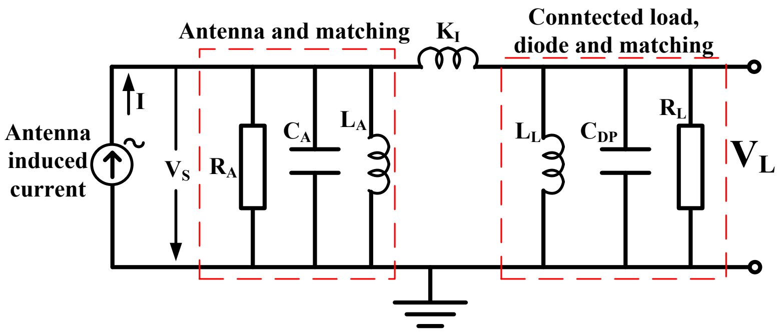
Antenna and matching
K_I
Conntected load, diode and matching
I
Antenna induced current
V_S
R_A
C_A
L_A
L_L
C_{DP}
R_L
V_L

The first parallel RLC resonator is modeled as impedance from the antenna and some passive matching components. The second parallel RLC resonator is modeled as impedance from the linearized diodes, its connected load and some passive matching components. I is the antenna induced current, V_S this time is the voltage across the parallel R_A and K_1 is the coupling element between the two parallel resonators. Using Kirchoff's current laws, the node equations can be expressed as Equations (29) and (30):

$$I = V_S\left[\frac{1}{R_A} + j\omega C_A - \frac{j}{\omega L_A} - \frac{j}{\omega K_1}\right] + \frac{jV_L}{\omega K_1} \tag{29}$$

$$0 = \frac{jV_S}{\omega K_1} + V_L\left[\frac{1}{R_L} + j\omega C_{DP} - \frac{j}{\omega L_L} - \frac{j}{\omega K_1}\right] \tag{30}$$

Load voltage (V_L) and the source voltage (V_S) at resonance are given by the equations in Equation (31). The resonance frequencies are given by Equation (32):

$$V_L = \frac{\frac{-jI}{\omega K_1}}{\left[\frac{1}{R_A R_L} + \frac{1}{\omega^2 K_1^2}\right]} \quad \text{and} \quad V_S = \frac{\frac{I}{R_L}}{\left[\frac{1}{R_A R_L} + \frac{1}{\omega^2 K_1^2}\right]} \tag{31}$$

$$\omega C_A - \frac{1}{\omega}\left\{\frac{1}{L_A} + \frac{1}{K_1}\right\} = 0 \text{ and } \omega C_{DP} - \frac{1}{\omega}\left\{\frac{1}{L_L} + \frac{1}{K_1}\right\} = 0 \tag{32}$$

From V_L and Vs in Equation (31), the voltage gain at resonance can be expressed as:

$$\frac{V_L}{V_S} = \frac{R_L}{j\omega K_1} \tag{33}$$

The maximum of Equation (33) is obtained when:

$$j\omega K_1 \to 0 \quad or \quad R_L \to \infty \tag{34}$$

Since $j\omega K_1$ is restricted by the conditions in Equation (32) to attain resonance, one cannot manipulate $j\omega K_1$ alone without changing the resonance conditions. What can drive the voltage gain is if R_L is very large at resonance conditions. If the input impedance (V_S/I) of the coupled resonator is maximum at resonance, conditions in Equation (35) hold:

$$\left(\frac{Vs}{I}\right) \to maximum \;\; when \;\; \frac{R_L}{\omega^2 K_1^2} \to 0 \tag{35}$$

Equation (36) may be assumed when $\frac{R_L}{\omega^2 K_1^2} \to 0$:

$$\frac{V_S}{I} = R_A \tag{36}$$

Under these conditions and an optimal coupling coefficient K_{1max}, the maximum voltage gain of the parallel coupled resonator can be written as in Equation (37), where K_{1max} is given by Equation (38):

$$\left|\frac{V_L}{V_S}\right| = \left|V_{gain}\right| = \left|\frac{1}{2}\sqrt{\frac{R_L}{R_A}}\right| \tag{37}$$

$$K_{1(\max)} = \mp\frac{1}{\omega}\sqrt{R_A R_L} \tag{38}$$

The analysis of Section 2.5.2 and parallel coupled RLC resonators show that any antenna and matched rectifying diode can be described as an equivalent circuit of a coupled resonator at a defined operating point. This general model can be applied to optimize other harvesters with complex output impedance such as piezo-harvesters or vibration harvesters for maximum transfer of power or voltage to its connected load. The model can also be applied to near field magnetically coupled antennas/coils for optimization.

2.5.6. Broadband RF to DC power converter

A broadband network is preferred when an RF to DC power converter is to be operated for a wide range of frequencies. A broadband converter is realized by connecting successive L-networks together in a multi-network between the antenna source and the rectifying diodes. The result is broadband or multiband RF power converter around certain frequencies. This can be deduced from the general model of a coupled resonators that by choosing certain passive components between a source and the load, it is possible to have more frequencies (ω) fulfilling Equation (32) and hence a result of multiple resonant frequencies or broader bandwidth at match conditions. For a two stage L-connected match, the quality factor of the L-network with the parallel resistance is given by Equation (39):

$$Q_P^{\ *} = \sqrt{\frac{R_P}{R^*} - 1} \tag{39}$$

The quality factor of the L-network with the series resistance is given by Equation (40):

$$Q_S^{\ *} = \sqrt{\frac{R_*}{R_S} - 1} \tag{40}$$

From Equations (39) and (40) the loaded quality factor of the two stage L-connected broadband network may be written as Equation (41) in terms of the unloaded quality factors; Q_S* and Q_P*:

$$Q^2 = \{(Q_P^{*2} + 1)(Q_S^{*2} + 1)\} - 1 \tag{41}$$

$R*$ in this case may be chosen if it is larger than R_S and lower than the R_P. The highest possible bandwidth between a resistive source and resistive load is found for a virtual resistance ($R*$) given in Equation (42) [37]:

$$R^* = \sqrt{R_S R_P} \tag{42}$$

For complex loads such as rectifying diodes or transistors, the largest achievable bandwidth prescribed by Equation (42) is limited by the load or source component quality factor, since Equation (42) does not take into account reactive impedance associated with the source or load.

2.5.7. Broadband-Match RF to DC Converter Results and Discussion

The antenna source resistance was broadband matched to the HSMS-285C diodes (and load) resistance at −30 dBm input around 434 MHz. For a desired Q_P* and Q_S* of 2.7 there is ~0.4 pF inherent diode capacitance which is un-tuned using a two stage L-matching network [Figure 11(b)]. This is because the HSMS-285C diodes provides an inherent component quality factor of 6.96 at 434 MHz for −30 dBm input, hence a broadband circuit with Q_P* lower than this inherent component quality factor of the diodes (and load) is difficult to achieve without trade-offs. However, connected L-networks with Q_P* as high as the diode component quality factor may perform worse than a single L-matched network with similar loaded quality factor. This is due to redundant components of the connected L-networks which have inherent losses.

Figure 11. Impedance diagram of broadband RF power converter; (**a**) Broadband match around 434 MHz with loaded Q of 2.7, (**b**) Resultant impedance matching network with un-turned capacitance of 0.4 pF.

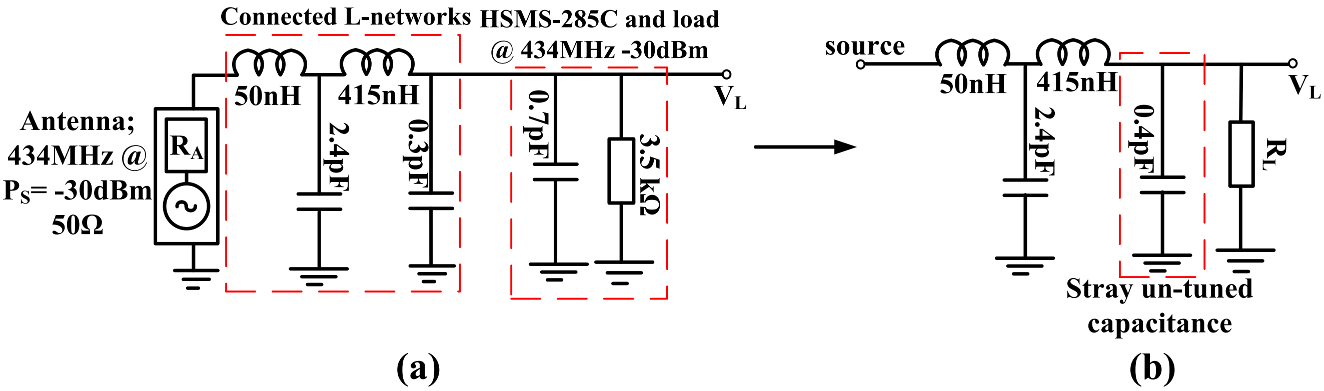
Connected L-networks
HSMS-285C and load
@ 434MHz -30dBm
Antenna;
434MHz @
P_S= -30dBm
50Ω
R_A
50nH
415nH
2.4pF
0.3pF
0.7pF
3.5 kΩ
V_L
(a)
source
50nH
415nH
2.4pF
0.4pF
R_L
V_L
Stray un-tuned
capacitance
(b)

Figure 12. Measured open circuit S_{11} of the broadband circuit around 434 MHz at different input power levels from a 50 Ω source (**left**), measured open circuit input impedance at −10 dBm and −40 dBm of the broadband circuit (**right**).

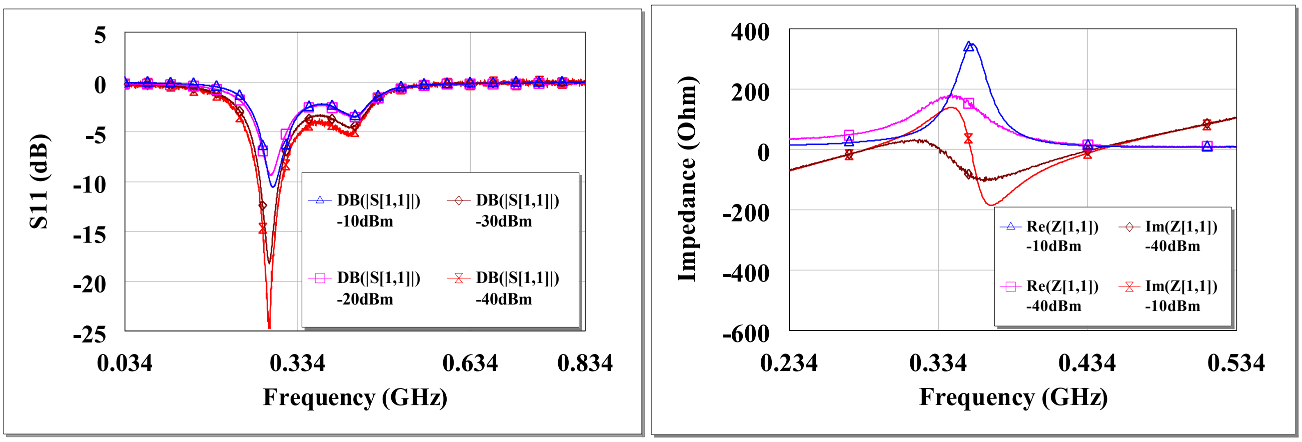

Therefore the broadband circuit is matched for Q_P* of 2.7, notwithstanding the un-tuned shunt capacitance as can be seen in Figure 11(b). Figure 12 shows the circuit S_{11} at various input power levels and input impedance at open circuit conditions. From Figure 12 (left) there is ~−5 dB return loss from 200 MHz to 500 MHz providing an operating band of ~300 MHz. The impedance of the circuit shows resonances at ~290 MHz and ~450 MHz as shown in Figure 12(right). A third resonance occurs around 356 MHz at −10 dBm as the frequency of highest harvester input resistance (~350 Ω) and where the reactive impedances approach their extremes. Figure 12 show that a wireless EM harvester can exhibit different resonance scenarios depending on the dominant instantaneous conditions. The efficiency and voltage sensitivity of the broadband matched wireless EM harvester are shown in Figure 13. The average open circuit voltage is 47 mV and 1.1 V at −30 dBm and −10 dBm, respectively, when operating from 200 MHz to 500 MHz.

Figure 13. Measured open circuit voltage *versus* frequency sweep from 200 MHz to 500 MHz for −10 dBm and −30 dBm (**left**), measured efficiency at 17 kΩ load *versus* frequency sweep for −10 dBm and −30 dBm (**right**).

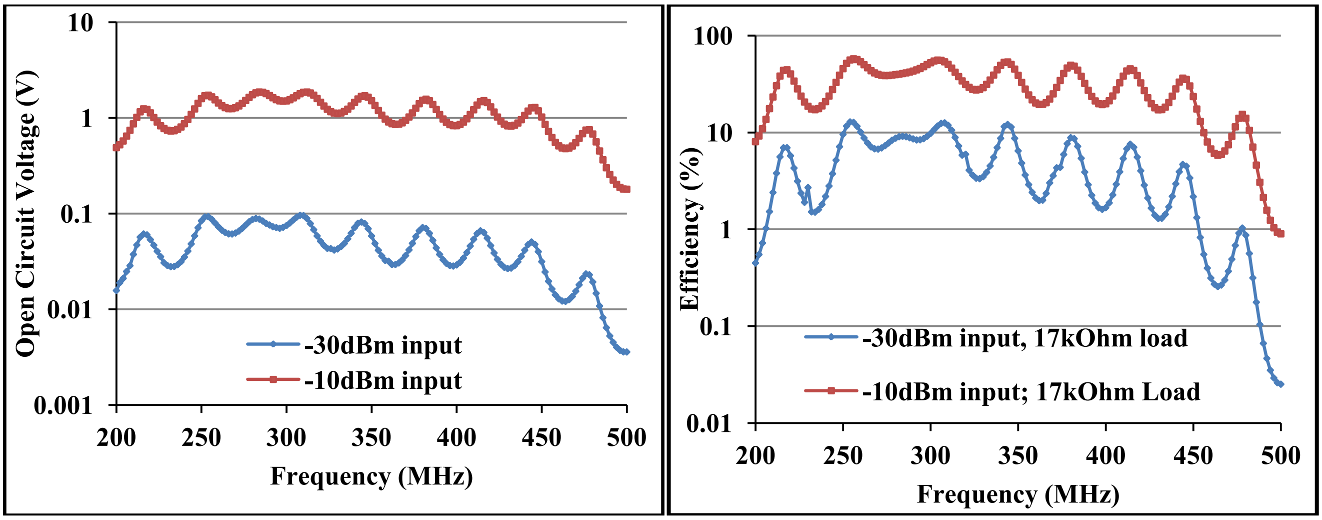

The broadband circuit achieves average efficiency of 5% at 17 kΩ load for −30 dBm and 30% at 17 kΩ load for −10 dBm input power from 200 MHz to 500 MHz. Figure 13 further confirm a direct link between frequency response and the unloaded quality factors. For Q_S* and Q_P* of ~2.7, the circuit response is broadband around 434 MHz.

2.6. High Voltage Sensitive RF to DC Converter

The current state of the art low power remote sensors would require a DC voltage supply of about 1 V and DC current of about 30 µA for operation. Therefore, the issue is not only how efficient a wireless EM harvester is in converting RF to DC power, but also what the output DC voltage and current of the EM harvester are at the RF input power level [38]. Equations (11,24) and (33) show that the maximum voltage sensitivity of a coupled resonator system or an RF to DC power converter is mostly related to the load and the source resistances at resonance. Therefore high voltage sensitive wireless EM harvester can be achieved with a diode voltage doubler with a very high input resistance relative to the antenna source without the need to cascade the diodes as in voltage multipliers. If the diodes been used for the RF to DC power conversion cannot provide high resistive impedance at the working frequency relative to the antenna source, then a DC-DC converter can be applied after the EM harvester as presented in [39] or the diodes may be cascaded by way of multipliers as presented in our earlier work [40] and by several other authors [3,5,14]. In case of multipliers, the input voltage ought to be high enough to overcome the junction potential of the several diodes in the multiplier network. If frequency is not a constraint, then a frequency sweep *versus* impedance for the diodes can be made and the frequency where the diodes exhibits high resistive impedance can be used to realize high voltage sensitive wireless RF harvester. For Schottky diodes, high resistive impedance occurs mostly at lower frequencies (see Figure 3). The measured voltage gain of a high resistive diode pair (voltage doubler) is presented in the next results.

2.6.1. High Voltage Sensitive RF to DC Converter Results and Discussion

The presented result was L-matched using 50 Ω resistance of the antenna source and the resistance of the HSMS-286C diodes (and load). The HSMS-286C diodes do provide high resistive impedance at low frequencies; notwithstanding the flicker noise which causes its resistive (and reactive) impedance to fluctuate. The HSMS-286C has low forward junction potential (~350 mV at 1 mA) per diode and series impedance of ~1.5–j8.1 kΩ or parallel impedance of ~$-j8.3(46.3)/(-j8.3 + 46.3)$ kΩ at 13.6 MHz for −30 dBm input. Even though the HSMS-286C diodes unloaded component quality factor at 13.6 MHz is similar to that of the HSMS-285C diodes at 434 MHz, the elevated resistive impedance at 13.6 MHz fulfills the condition for high voltage sensitivity relative to a 50 Ω antenna source at resonance conditions.

Figure 14. (**a**) L-matched impedance diagram for matching the HSMS-286C diodes at 13.6 MHz at -30 dBm input. (**b**) Resultant network, (**c**) PCB layout of the L-matched Greinacher doubler with adjusted values due to impedances provided by copper route on FR4 substrate with thickness of 1 mm. (**d**) Fabricated PCB of the L-matched RF to DC power converter.

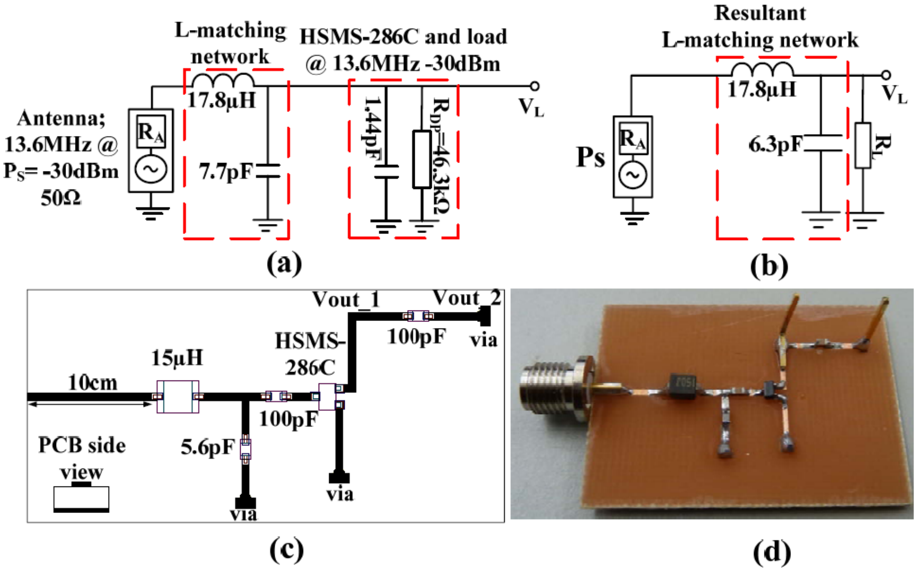
L-matching network
HSMS-286C and load @ 13.6MHz -30dBm
Antenna; 13.6MHz @ P_S= -30dBm 50Ω
R_A
17.8µH
7.7pF
1.44pF
R_{DP}=46.3kΩ
V_L
(a)
Resultant L-matching network
Ps
R_A
17.8µH
6.3pF
R_L
V_L
(b)
Vout_1
Vout_2
HSMS-286C
100pF
via
15µH
10cm
100pF
PCB side view
5.6pF
via
via
(c)
(d)

The high voltage sensitive EM harvester operating at 13.6 MHz is as shown in Figure 14. On the realized PCB is a Greinacher doubler. An inductance of 15 µH and a shunt capacitance of 5.6 pF were the adjusted values after the microstrip contributions.

The measured S_{11} and input impedance at open circuit are shown in Figure 15. There is high return loss and resonance around 13.6 MHz. The circuit input impedance at open circuit conditions is 58 Ω at resonance for both −40 dBm and −10 dBm.

Figure 15. Measured open circuit S_{11} of the L-matched HSMS-286C diodes at 13.6 MHz for different input power levels from a 50 Ω source (**left**), measured open circuit input impedance at −10 dBm and −40 dBm of the L-matched HSMS-286C diode at 13.6 MHz (**right**).

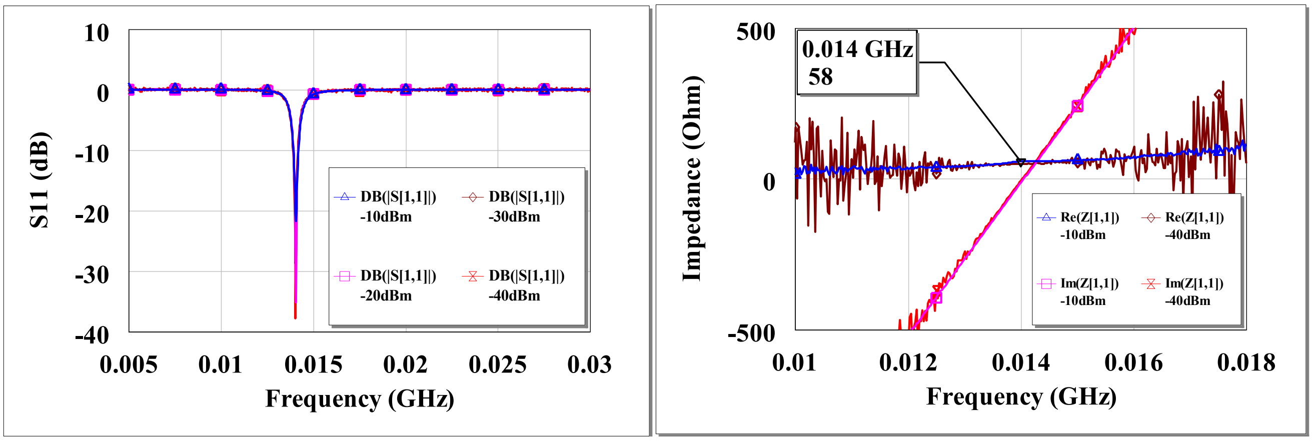

The efficiency and voltage sensitivity of the high voltage sensitive wireless EM harvester are shown in Figure 16.

Figure 16. Measured circuit efficiency *versus* load at various input power levels at 13.6 MHz (**left**), measured open circuit voltage and at 200 kΩ load *versus* input power at 13.6 MHz (**right**).

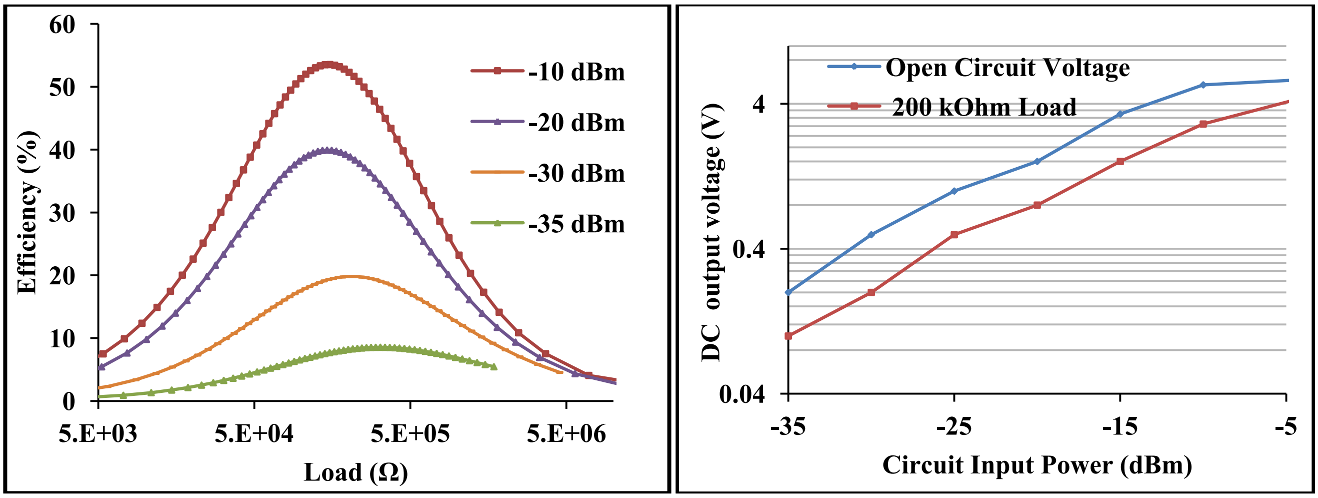

The maximum measured efficiency at −30 dBm is 20% for ~200 kΩ load and an open circuit voltage of 0.5 V. At −10 dBm, the maximum efficiency and open circuit voltage are 54% and 5.4 V respectively. At the optimal load of ~200 kΩ, the detected voltage is 0.2 V and 2.9 V at −30 dBm and −10 dBm respectively. The open circuit voltage gain is 100 at −30 dBm and 108 at −10 dBm.

Even though the RF to DC converter presented in Section 2.5.3 is the same as the L-match circuit realized with the HMSM-286C diodes at 13.6 MHz, the voltage gain is increased by a factor of 4 due to the large difference between the diodes (and load) resistance and source resistance so that at matched conditions high voltage gain occurs. The loaded *Q* of the L-matched circuit is 30 which results in small frequency bandwidth just like a PI-matched diode rectifier presented in our earlier work [40]. From this result and the results from our earlier presented PI-matched EM harvester, it can be inferred that all high loaded *Q* RF to DC circuits have high selectivity but not all highly selective RF to DC circuits have high loaded *Q*. The voltage sensitivity of the matched HSMS-286C diode at 13.6 MHz can be improved if its resistive impedance is not lowered by the flicker noise.

3. Wireless EM Power Harvester

A wireless EM harvester, consisting of a rectifying antenna (*rectenna*) was designed to find a compromise between size and performance of its antenna. The rectenna is shown in Figure 17.

Figure 17. Rectenna realized on a Duroid 5880, 1.57 mm substrate. (**a**) Backside of the rectenna, (**b**) cross-section of antenna output coplanar stripline dimensions (**c**) frontside of the rectenna.

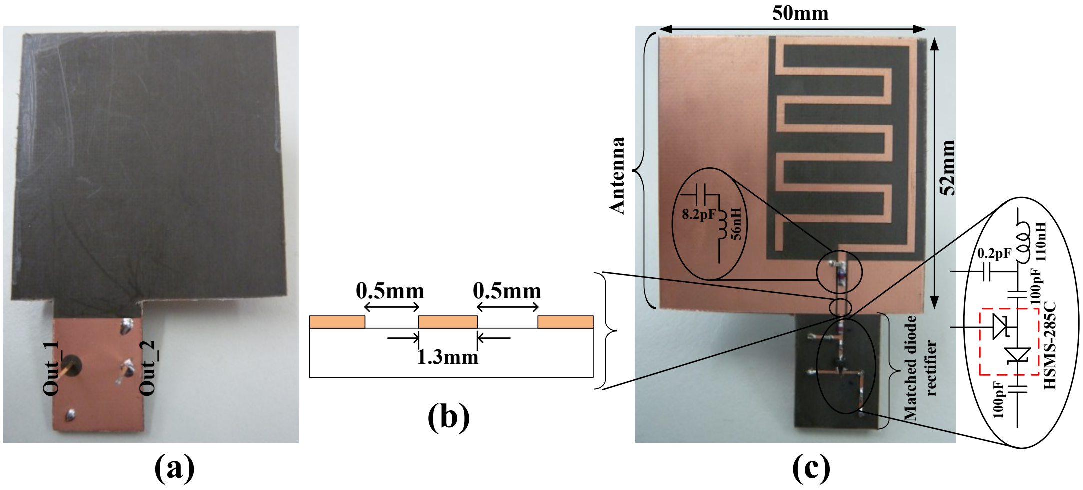

The antenna (planar) part of the rectenna is based on our earlier work [41]. In contrast to the earlier presented antenna, this rectenna is realized on a Duroid [42] substrate of thickness 1.57 mm. Duroid 5880 has lower loss tangent of 0.0004 at 1 MHz compared to 0.025 at 1 MHz for FR4. This means there is less loss in the transmission of signals on a Duroid PCB at this frequency range. The antenna part is fabricated to resonate around 434 MHz; hence its dimensions of 5 × 5.2 cm make it electrically small. The antenna is tuned with a chip inductor and a capacitor to achieve the resonance frequency around 434 MHz [Figure 17(c)]. This is done at a cost of reduced antenna radiation efficiency. An antenna is one of the few components the size of which is related to the operating frequency. Thus, if the size of an antenna is fixed, resonance frequency reduction of the antenna can only be achieved with penalty factors [10]. The antenna's output impedance is tuned with the dimensions of the coplanar stripline as shown in Figure 17(b).

HFSS [43] was used to simulate the presented antenna and to find the correct capacitive and inductive components for frequency tuning before the optimized design was fabricated. The simulated antenna resonances occur at 438 MHz and 445 MHz. At these frequencies, the radiation efficiency is 20% and a peak gain of −6 dBi. The rectifying part of the rectenna consists of L-matched HSMS-285C diodes (Figure 17(c)). The L-matched HSMS-285C part of the rectenna can be engineered to be as small as possible if required. The separate parts of the rectenna were characterized by terminating their ends and measuring the individual reflection coefficients just like the power converters presented in Section 2. Figure 18 shows the measured antenna and matched rectifier individual S_{11} and impedance. Figure 18 (left) also show the HFSS simulated S_{11} results. From Figure 18 (right), the measured antenna resonance where the input impedance is at maximum is ~434 MHz. At ~434 MHz, the antenna input resistance is 376 Ω and the reactive impedances approach their extreme (*so called anti-resonance*). The other resonance occurs when the input resistance is 'finite' and the reactive impedance is zero; at ~441 MHz. The input resistance is 57 Ω at ~441 MHz. The rectifier circuit is matched for the antenna's resistance at ~441 MHz.

Figure 18. Antenna HFSS simulated, antenna measured, and measured L-matched diode rectifier S_{11} on a Duroid 5880 PCB for −30 dBm input (**left**), Measured open circuit input impedance of antenna and rectifier at −30 dBm input (**right**).

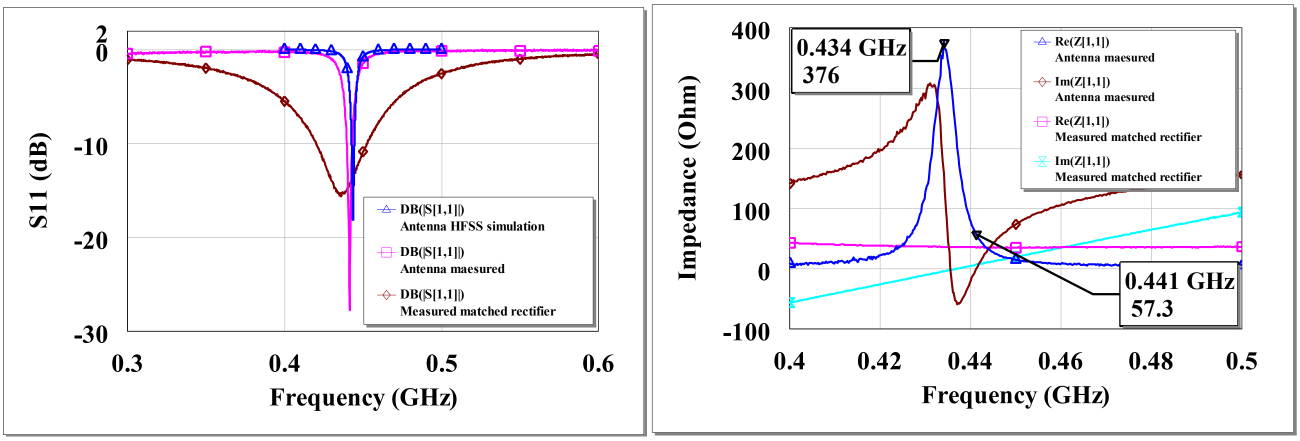

3.1. EM Range Results and Discussion

At far field between wireless EM transmitting and receiving antenna, the coupling mechanism between the transmitting and receiving antenna is neither capacitive nor inductive as is the case for the RF to DC converters. The coupling is radiative which can be described by the Friis equation of transmission on the assumption that the transmitting and receiving antenna are in free space [44]. A modified Friis equation for a transmitting and receiving antenna at far-field (R >> λ and R >> transmitting antenna largest dimension) to each other at a specified direction is given by Equation (43) [45]. Equation (43) assumes real world open space conditions:

$$\frac{P_r}{P_t} = F_{envt} G_t G_r \left(\frac{\lambda}{4\pi R} \right)^2 , \tag{43}$$

where *Pr* is the power at the receiving antenna port, *Pt* is the power supplied at the transmitting antenna port, F_{envt} is a factor accounting for environmental effects as such ground reflections among others, *Gt and Gr* are the transmitting and receiving antenna gain (*at specified direction*) respectively. *R* is the distance between the transmitting and receiving antenna and λ is the wavelength of the transmitting EM wave. The rectenna receiving range measurements were carried out in an open space (*hall*) with the antennas 2 m above ground level. The transmitting and receiving antennas were arranged in the direction of their peak gain. The rectenna range performance is shown in Figure 19. According to Equation (43), the efficiency of RF power transferred between a sending and receiving antenna depends on controllable factors like the gain of the antennas in the arranged direction and the radiation efficiency of the antennas. Since the receiving/transmitting antenna's incorporated in remote harvesters for sensor powering are normally small in relation to their operating frequencies, they tend to be less efficient.

Figure 19. Rectenna receiving range performance by sending 17 dBm (50 mW) at a gain of −6 dBi at 437 MHz. Output DC voltage *versus* receiving distance for different loads (**left**), loads output power *versus* receiving distance (**right**).

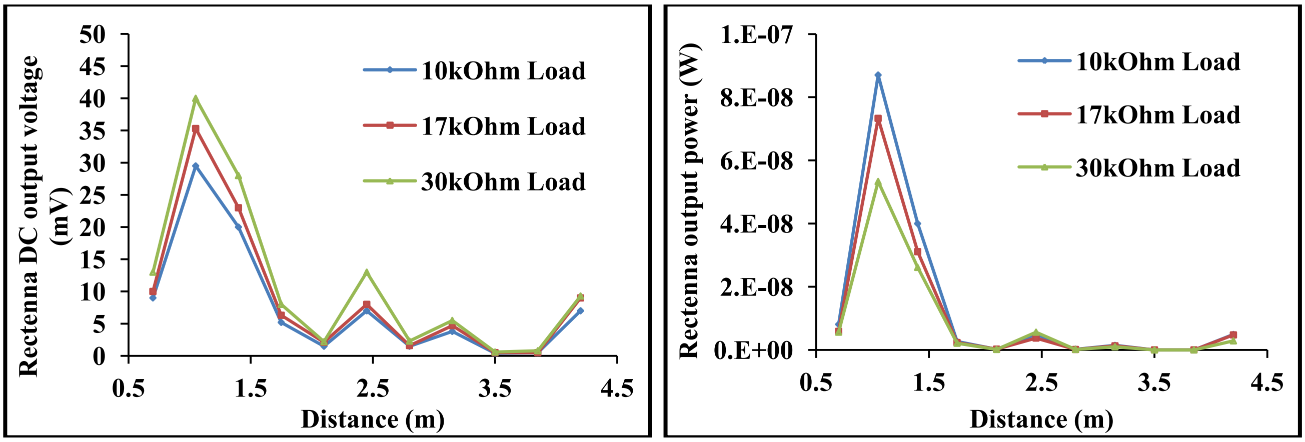

The efficiency of the rectenna's antenna is ~20% at resonance. A 'perfectly' matched RF to DC power converter operating in its square law region has efficiencies in the region of 20% as depicted in Section 2. The transmitting antenna was the same as the antenna incorporated in the rectenna. By transmitting the EM power with a small antenna (5 cm × 5.2 cm) at 437 MHz with efficiency of ~20% and at a gain of −6 dBi, the power delivered by the rectenna is generally low at far-field from the transmitter as can be seen in Figure 19. A mediocre transmitting antenna was used to transmit the EM waves due to limitations in the European Union about transmitting EM power at certain frequencies; so the goal in the rectenna range experiment is to show the lowest limit functionality of such a harvester. At 4.2 m from the electrically small transmitting antenna transmitting at 17 dBm, the rectenna harvested DC voltage and power are 9 mV and 5 nW respectively for 10 kΩ load. It can be seen from Figure 19 that the harvested voltage/power generally degrades as an inverse square of distance from transmitter as described by Friis equation. The measured received power however alternate along this R^{-2} fit as shown in Figure 19. This anomaly is accounted for by F_{envt} [Equation (43)] as influence of ground reflections and polarization in real world open field measurements [45]. For any particular distance *R*, the signals reflected from ground can be constructive with the direct signal to the rectenna, in which case the measured power may be higher than that predicted by the original Friss equation as in [44]. The ground effect can also be destructive, in which case the measured power will be lower than what is predicted by the original Friis equation.

4. Conclusions

Optimization of Schottky diode-based RF to DC power converters using different matching techniques for wireless EM energy harvesting applications is presented. Using scattering parameters for small signal modeling, it is shown that wireless EM harvesters can be generally described as coupled resonators with efficiencies and maximum voltage sensitivity depending mostly on the source and load resistances under matched conditions. The analytical models allow systematic control in the design of passive wireless EM harvesters. Based on these analyses, a rectenna is built and tested for lower limit functionality from harvesting ambient EM waves. The analysis presented in this work may also be applied to optimize derivatives of wireless EM harvesters like RFID tags, NFC, wireless chargers *etc.*, for efficient powering of their sensors or integrated circuits. Generally, most energy harvesters and their matched loads can be described as coupled resonators and thus may be optimized with the methods presented in this work.

Acknowledgments

This work is part of the graduate program GRK 1322 Micro Energy Harvesting at IMTEK, University of Freiburg, funded by the German Research Foundation (DFG). Special thanks to Daniela Ohnemus for the PCB preparations and Uwe Burzlaff for antenna range measurements.

References

1. Fischer, J. NFC in cell phones: The new paradigm for an interactive world [Near-Field Communications]. *IEEE Commun. Mag.* **2009**, *47*, 22–28.
2. Shameli, A.; Safarian, A.; Rofougaran, A.; Rofougaran, M.; de Flaviis, F. Power harvester design for passive UHF RFID tag using a voltage boosting technique. *IEEE Trans. Microw. Theor. Tech.* **2007**, *55*, 1089–1097.
3. Karthaus, U.; Fischer, M. Fully integrated passive UHF RFID transponder IC with 16.7-µW minimum RF input power. *IEEE J. Solid-State Circuits* **2003**, *38*, 1602–1608.
4. Lee, S.H.; Jin, I.S. Interoperation of an UHF RFID reader and a TCP/IP device via wired and wireless links. *Sensors* **2011**, *11*, 10664–10674.
5. Mandal, S.; Sarpeshkar, R. Low-power CMOS rectifier design for RFID applications. *IEEE Trans. Circuits Syst. I: Regul. Papers* **2007**, *54*, 1177–1188.
6. Yu, H.; Bashirullah, R. A Low Power ASK Clock and Data Recovery Circuit for Wireless Implantable Electronics. In *Proceedings of IEEE Custom Integrated Circuits Conference, CICC '06*, San Jose, CA, USA, September 2006; pp. 249–252.
7. Hannan, M.A.; Abbas, S.M.; Samad, S.A.; Hussain, A. Modulation techniques for biomedical implanted devices and their challenges. *Sensors* **2011**, *11*, 297–319.

8. Cho, N.; Song, S.-J.; Lee, J.Y.; Kim, S.; Kim, S.; Yoo, H.-J. A 8-μW, 0.3-mm2 RF-powered Transponder with Temperature Sensor for Wireless Environmental Monitoring. In *Proceedings of IEEE International Symposium on Circuits and Systems, ISCAS*, Kobe, Japan, May 2005; Volume 5, pp. 4763–4766.
9. Brown, W.C. The history of power transmission by radio waves. *IEEE Trans. Microw. Theor. Tech.* **1984**, *32*, 1230–1242.
10. Mickle, M.H.; Mi, M.; Mats, L.; Capelli, C.; Swift, H. Powering autonomous cubic-millimeter devices. *IEEE Antennas Propag. Mag.* **2006**, *48*, 11–21.
11. McSpadden, J.O.; Chang, K. A Dual Polarized Circular Patch Rectifying Antenna at 2.45 GHz for Microwave Power Conversion and Detection. In *Proceedings of IEEE MTT-S International Microwave Symposium Digest*, San Diego, CA, USA, May 1994; Volume 3, pp. 1749–1752.
12. Sample, A.; Smith, J.R. Experimental Results with Two Wireless Power Transfer Systems. In *Proceedings of the 4th International Conference on Radio and Wireless Symposium*, San Diego, CA, USA, January 2009; pp. 16–18.
13. Umeda, T.; Yoshida, H.; Sekine, S.; Fujita, Y.; Suzuki, T.; Otaka, S. A 950-MHz rectifier circuit for sensor network tags with 10-m distance. *IEEE J. Solid-State Circuits* **2006**, *41*, 35–41.
14. Le, T.; Mayaram, K.; Fiez, T. Efficient far-field radio frequency energy harvesting for passively powered sensor networks. *IEEE J. Solid-State Circuits* **2008**, *43*, 1287–1302.
15. Yuan, F. *CMOS Circuits for Passive Wireless Microsystems*; Springer: New York, NY, USA, 2011.
16. Zbitou, J.; Latrach, M.; Toutain, S. Hybrid rectenna and monolithic integrated zero-bias microwave rectifier. *IEEE Trans. Microw. Theor. Tech.* **2006**, *54*, 147–152.
17. Ungan, T.; Le Polozec, X.; Walker, W.; Reindl, L. RF Energy Harvesting Design Using High Q Resonators. In *Proceedings of IEEE MTT-S International Microwave Workshop on Wireless Sensing, Local Positioning and RFID, IMWS 2009*, Cavtat, Croatia, September 2009; pp. 1–4.
18. Ungan, T.; Freunek, M.; Muller, M.; Walker, W.D.; Reindl, L.M. Wireless Energy Transmission Using Electrically Small Antennas. In *Proceedings of IEEE Radio and Wireless Symposium, RWS '09*, San Diego, CA, USA, January 2009; pp. 526–529.
19. Boquete, L.; Rodríguez-Ascariz, J.M.; Barea, R.; Cantos, J.; Miguel-Jiménez, J.M.; Ortega, S.; Data acquisition, analysis and transmission platform for a pay-as-you-drive system. *Sensors* **2010**, *10*, 5395–5408.
20. Heikkinen, J.; Salonen, P.; Kivikoski, M. Planar Rectennas for 2.45 GHz Wireless Power Transfer. In *Proceedings of Radio and Wireless Conference, RAWCON 2000*, Denver, CO, USA, 10–13 September 2000; pp. 63–66.
21. Akkermans, J.A.G.; van Beurden, M.C.; Doodeman, G.J.N.; Visser, H.J. Analytical models for low-power rectenna design. *IEEE Antenn. Wireless Propag. Lett.* **2005**, *4*, 187–190.
22. Hagerty, J.A.; Helmbrecht, F.B.; McCalpin, W.H.; Zane, R.; Popovic, Z.B. Recycling ambient microwave energy with broad-band rectenna arrays. *IEEE Trans. Microw. Theor. Tech.* **2004**, *52*, 1014–1024.
23. Harb, A. Energy harvesting: State-of-the-art. *Renew. Energ.* **2011**, *36*, 2641–2654.
24. Vullers, R.J.M.; van Schaijk, R.; Doms, I.; van Hoof, C.; Mertens, R. Micropower energy harvesting. *Solid-State Electron.* **2009**, *53*, 684–693.

25. Sah, C.-T. *Fundamentals of Solid-State Electronics*; World Scientific: Singapore, Singapore, 1991.
26. Wetenkamp, S. Comparison of Single Diode *vs.* Dual Diode Detectors for Microwave Power Detection. In *Proceedings of IEEE MTT-S International Microwave Symposium Digest*, Boston, MA, USA, June 1983; pp. 361–363.
27. Cardoso, A.J.; Schneider, M.C.; Montoro, C.G. Design of Very Low Voltage CMOS Rectifier Circuits. In *Proceedings of Circuits and Systems for Medical and Environmental Applications Workshop (CASME)*, Merida, Mexico, December 2010; pp. 1–4.
28. Cardoso, A.J.; de Carli, L.G.; Galup-Montoro, C.; Schneider, M.C. Analysis of the Rectifier Circuit Valid Down to Its Low-Voltage Limit. *IEEE Trans. Circuits Syst. I: Regul. Papers* **2012**, *59*, 106–112.
29. Data sheet HSMS-285x. Avago Technologies. Available online: http://www.avagotech.com (accessed on 28 September 2012).
30. Data sheet HSMS-286x. Avago Technologies. Available online: http://www.avagotech.com (accessed on 28 September 2012).
31. Hsu, T.S. Low-frequency excess noise in metal—Silicon Schottky barrier diodes. *IEEE Trans. Electron Devices* **1970**, *17*, 496–506.
32. Hastas, N.A.; Dimitriadis, C.A.; Dozsa, L.; Gombia, E.; Amighetti, S.; Frigeri, P. Low frequency noise of GaAs Schottky diodes with embedded InAs quantum layer and self-assembled quantum dots. *J. Appl. Phys.* **2003**, *93*, 3990.
33. Gomila, G.; Reggiani, L.; Rubı́, J.M. Shot-noise suppression in Schottky barrier diodes. *J. Appl. Phys.* **2000**, *88*, 3079.
34. Kleinpenning, T.G.M. Low-frequency noise in Schottky barrier diodes. *Solid-State Electron.* **1979**, *22*, 121–128.
35. Lee, T.H. *The Design of CMOS Radio-Frequency Integrated Circuits*; Cambridge University Press: Cambridge, UK, 2004.
36. Advanced Design System. Agilent Technologies: Santa Clara, CA, USA.
37. Bowick, C.; Blyler, J.; Ajluni, C.J. *RF Circuit Design*; Newnes/Elsevier: Amsterdam, The Netherlands, 2008.
38. Joe, J.; Chia, M.; Marath, A.; Ang, C. Zero Bias Schottky Diode Model for Low Power, Moderate Current Rectenna. In *Proceedings of Defence Electronics Technology Seminar (DETS 97)*, Singapore, Singapore, November 1997; pp. 1–5.
39. Visser, H.J.; Vullers, R.J.M.; Veld, B.O.h.; Pop, V. Remote RF Battery Charging. In *Proceedings of Power MEMS 2010*, Leuven, Belgium, December 2010; pp. 37–40.
40. Nimo, A.; Grgic, D.; Reindl, L.M. Impedance optimization of wireless electromagnetic energy harvester for maximum output efficiency at μW input power. *Proc. SPIE* **2012**, *8341*, 83410W:1–83410W:14.
41. Nimo, A.; Grgic, D.; Reindl, L.M. Electrically small Planner Antenna for Compact Electromagnetic (EM) Wireless Energy Harvester. In *Proceedings of Power MEMS 2011*, Seoul, Korea, November 2011; pp. 310–313.
42. RT-duroid-5870–5880-Data-Sheet; Rogers Corporation: Rogers, CT, USA, 2011.

43. *HFSS*; Ansys Corporation. Ansys, Inc: Canonsburg, PA, USA, 2005.
44. Friis, H.T. A Note on a Simple Transmission Formula. *Proc. IRE* **1946**, *34*, 254–256.
45. Kvaksrud, T.I. *Range Measurements in an Open Field Environment*, *Design Note DN018*; Texas Instrum. Incorporated: Dallas, TX, USA, 2008.
46. *LabView*; National Instruments: Austin, TX, USA, 2009.

Appendix A: Measuring Setup for RF Rectifier Efficiency and Voltage Sensitivity

The measuring setup is as shown in Figure A1.

Figure A1. RF to DC Power converter characterization setup.

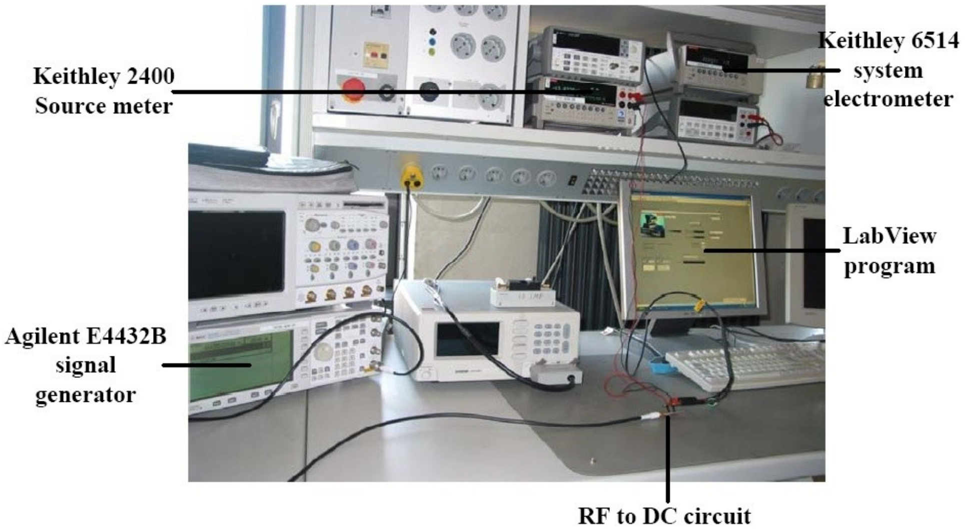

The RF to DC circuit efficiency and voltage sensitivity measurements were made with a Keithley 2400 source meter and Keithley 6514 system electrometer with an Agilent E4432B signal generator providing 50 Ω RF signal into the circuit board.

The closed circuit current drawn by the RF to DC power converter (*without load*) from the generator is first determined by the Keithley 2400 source meter; then starting from this current, the value of the current is decreased at set intervals to creates virtual load resistances to the circuit for up to a lowest current of 0.1 μA. The 6514 system electrometer is used to measure the output voltage. The number of data point is set through LabView [46] as well as the measurements. Additionally open circuit voltage or at specific loads and frequency sweep can be made through the LabView program. At −40 dBm input power and below, the detected voltages and currents were difficult to measure accurately with the measuring setup; hence measurements were made up to a minimum of -35 dBm input power. The circuit layout for the efficiency and voltage sensitivity measurements is schematically shown in Figure A2.

Figure A2. RF to DC power converter characterization circuit.

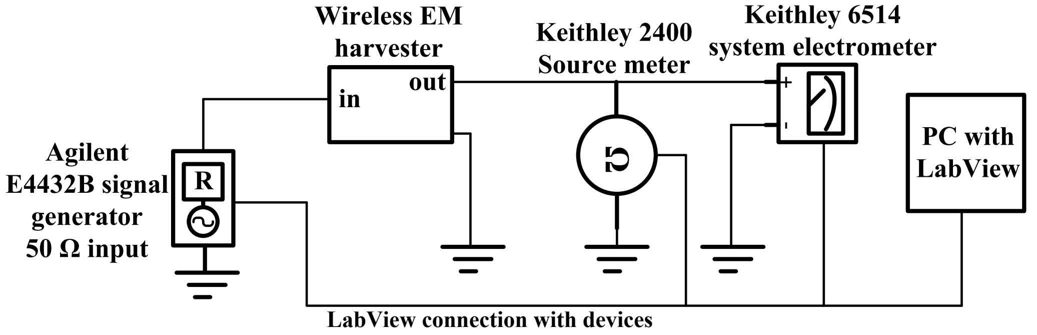
Wireless EM harvester
in
out
Keithley 2400 Source meter
Keithley 6514 system electrometer
+
-
PC with LabView
Agilent E4432B signal generator 50 Ω input
R
LabView connection with devices

© 2012 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).

This website uses cookies to ensure you get the best experience on our website. More info

New patented fabrication technique key to new solar power technology

February 27, 2013 by Colin Poitras

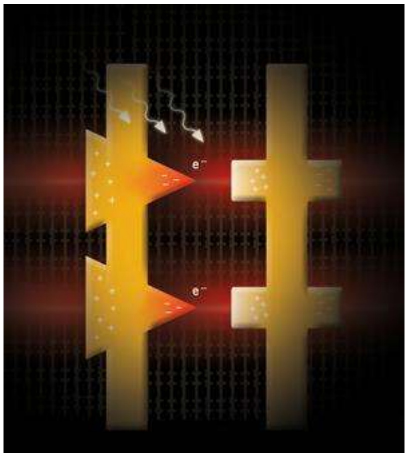

This is an illustration of a working optical rectifying antenna, or rectenna. Credit: (Illustration by Justine Braisted/UConn)

A novel fabrication technique developed by a University of Connecticut engineering professor could provide the breakthrough technology scientists have been looking for to vastly improve the efficiency of today's solar energy systems.

Capital One

Capital One® & Android Pay™
Pay Quickly and Securely

Android Pay and Capital One Help You Pay Quickly and Securely—Learn More!
capitalone.com/MobileFeatures/AndroidPay

For years, scientists have studied the potential benefits of a new branch of solar energy technology that relies on nanosized antenna arrays theoretically capable of harvesting more than 70 percent of the sun's electromagnetic radiation and simultaneously converting it into usable electric power.

But while nanosized antennas that also serve as rectifiers have shown promise in theory, scientists have lacked the technology required to construct and test them. The fabrication process is immensely challenging. The nano-antennas – known as "rectennas" because of their ability to both absorb and rectify solar energy from alternating current to direct current – must be capable of operating at the speed of visible light and be built in such a way that their core pair of electrodes is a mere 1 or 2 nanometers apart, a distance of approximately one millionth of a millimeter, or 30,000 times smaller than the diameter of human hair.

The potential breakthrough lies in a novel fabrication process called selective area atomic layer deposition (ALD) that was developed by Brian Willis, an associate professor of chemical and biomolecular engineering at the University of Connecticut and the former director of UConn's Chemical Engineering Program.

It is through atomic layer deposition that scientists believe they can finally fabricate a working rectenna device. In a rectenna device, one of the two interior electrodes must have a sharp tip, similar to the point of a triangle. The secret is getting the tip of that electrode within one or two nanometers of the opposite electrode, something similar to holding the point of a needle to the

Capital One

Capital One® & Android Pay™
Pay Quickly and Securely

Android Pay and Capital One Help You Pay Quickly and Securely—Learn More!
capitalone.com/MobileFeatures/AndroidPay

Featured Last comments **Popular**

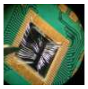
New silicon structure opens the gate to quantum computers Dec 12, 2017 3

Ancient penguin was as big as a (human) Pittsburgh Penguin Dec 12, 2017 1

Industrial Revolution left a damaging psychological 'imprint' on today's populations Dec 10, 2017 24

Officials: Right whales, after deadly year, could become extinct Dec 10, 2017 1

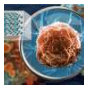
Hyperlens crystal capable of viewing living cells in unprecedented detail Dec 11, 2017 4

more »

Phys.org
G+ Follow

Phys.org on facebook
Follow 1.3M people are following Phys.org. Sign Up to see who your friends are following.

Email newsletter
email Subscribe

plane of a wall. Before the advent of ALD, existing lithographic fabrication techniques had been unable to create such a small space within a working electrical diode. Using sophisticated electronic equipment such as electron guns, the closest scientists could get was about 10 times the required separation. Through atomic layer deposition, Willis has shown he is able to precisely coat the tip of the rectenna with layers of individual copper atoms until a gap of about 1.5 nanometers is achieved. The process is self-limiting and stops at 1.5 nanometer separation.

Relevant PhysicsForums posts

B -> s μμ decays: Current status 20 minutes ago

CPT asymmetry question, in regards to recent experiment@CERN 2 hours ago

Where I can find the article: Higgs-Thomson-Fibonacci generation 3 hours ago

About Quarks 3 hours ago

About Noether Theorem 8 hours ago

What does gauge theory explain? 8 hours ago

More from High Energy, Nuclear, Particle Physics

1 | Install Solar for $0 Down Free Savings Calculation. Check Rebates & Incentives. Easy to Qualify. energybillcruncher.com

2 | Invention Idea-Start Here We Help Turn Ideas into Products. No Cost Consultation - Start Today. davison.com/Invention

The size of the gap is critical because it creates an ultra-fast tunnel junction between the rectenna's two electrodes, allowing a maximum transfer of electricity. The nanosized gap gives energized electrons on the rectenna just enough time to tunnel to the opposite electrode before their electrical current reverses and they try to go back. The triangular tip of the rectenna makes it hard for the electrons to reverse direction, thus capturing the energy and rectifying it to a unidirectional current.

Brian Willis holds a rectenna device. Credit: Sean Flynn/UConn Photo

Impressively, the rectennas, because of their extremely small and fast tunnel diodes, are capable of converting solar radiation in the infrared region through the extremely fast and short wavelengths of visible light – something that has never been accomplished before. Silicon solar panels, by comparison, have a single band gap which, loosely speaking, allows the panel to convert electromagnetic radiation efficiently at only one small portion of the solar spectrum. The rectenna devices don't rely on a band gap and may be tuned to harvest light over the whole solar spectrum, creating maximum efficiency.

Willis and a team of scientists from Penn State Altoona along with SciTech Associates Holdings Inc., a private research and development company based in State College, Pa., recently received a $650,000, three-year grant from the National Science Foundation to fabricate rectennas and search for ways to maximize their performance.

"This new technology could get us over the hump and make solar energy cost-competitive with fossil fuels," says Willis. "This is brand new technology, a whole new train of thought."

The Penn State Altoona research team – which has been exploring the theoretical side of rectennas for more than a decade – is led by physics professor Darin Zimmerman, with fellow physics professors Gary Weisel and Brock Weiss serving as co-investigators. The collaboration also includes Penn State emeritus physics professors Paul Cutler and Nicholas Miskovsky, who are principal members of Scitech Associates.

"The solar power conversion device under development by this collaboration of two universities and an industry subcontractor has the potential to revolutionize green solar power technology by increasing efficiencies, reducing costs, and providing new economic opportunities," Zimmerman says.

"Until the advent of selective atomic layer deposition (ALD), it has not been possible to fabricate practical and reproducible rectenna arrays that can harness solar energy from the infrared through the visible," says Zimmerman. "ALD is a vitally important processing step, making the creation of these devices possible. Ultimately, the fabrication, characterization, and modeling of the proposed rectenna arrays will lead to increased understanding of the physical processes underlying these devices, with the promise of greatly increasing the efficiency of solar power conversion technology."

The atomic layer deposition process is favored by science and industry because it is simple, easily reproducible, and scalable for mass production. Willis says the chemical process is particularly applicable for precise, homogenous coatings for nanostructures, nanowires, nanotubes, and for use in the next generation of high-performing semi-conductors and transistors.

The method being used to fabricate rectennas also can be applied to other areas, including enhancing current photovoltaics (the conversion of photo energy to electrical energy), thermoelectrics, infrared sensing and imaging, and chemical sensors.

Over the next year, Willis and his collaborators in Pennsylvania plan to build prototype rectennas and begin testing their efficiency.

"To capture the visible light frequencies, the rectenna have to get smaller than anything we've ever made before, so we're really pushing the limits of what we can do," says Willis. "And the tunnel junctions have to operate at the speed of visible light, so we're pushing down to these really high speeds to the point where the question becomes 'Can these devices really function at this level?' Theoretically we know it is possible, but we won't know for sure until we make and test this device."

Explore further: **Patented technique key to new solar power technology**

Provided by: University of Connecticut

0 shares

feedback to editors

1 Still Paying For Power? - WHY?
GO SOLAR AND LET THE SUN PAY FOR YOUR ELECTRICITY.
kuriosenergy.com/Solar

2 Faster, Cheaper, More Accurate - Izon's TRPS
More Reliable & Accurate Than NTA. Concentration, Size Analysis & More
izon.com/TRPS

Related Stories

Patented technique key to new solar power technology February 5, 2013

(Phys.org)—A novel fabrication technique developed by UConn engineering professor Brian Willis could provide the breakthrough technology scientists have been looking for to vastly improve today's solar energy systems.

New energy technologies promise brighter future

November 19, 2012

In three studies published in the current issue of Technology and Innovation – Proceedings of the National Academy of Inventors, innovators unveil creative technologies that could change our sources of energy, change our ...

Recommended for you

Engineers create plants that glow December 13, 2017

Imagine that instead of switching on a lamp when it gets dark, you could read by the light of a glowing plant on your desk.

Accelerating the self-assembly of nanoscale patterns for next-generation materials December 13, 2017

The ability to quickly generate ultra-small, well-ordered nanopatterns over large areas on material surfaces is critical to the fabrication of

Plasmonic device converts light into electricity • November 9, 2011

(PhysOrg.com) -- While the most common device for converting light

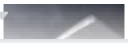

Nanotexturing creates bacteria-killing spikes on stainless steel surfaces

3 comments

2.5 Adjust slider to filter visible comments by rank

Display comments: **newest first**

EyeNStein 2.3 / 5 (6) Feb 27, 2013

Hi Guys: You already did this one 3 weeks ago:-
http://phys.org/n...wer.html

grondilu 3.7 / 5 (3) Feb 27, 2013

As if we didn't have enough of these news about some amazing ground-breaking new techniques in solar energy every day already, you have to post some of them twice. Jeez.

pico 1 / 5 (1) Mar 05, 2013

glad U did I missed it the 1st time. It sounds very promising! The scale I think its' will work but still it will be possible to reach even pico size and truely make this marketable to all at reasonable cost.

Commenting is closed for this article.

Trending News

Powered by Ideal Media

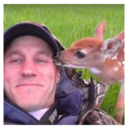

Baby Deer Refused To Leave Side Of Man Who Rescued It

An outdoorsman rehabilitated an injured fawn with plans to release it to ...

Teen Frees Scared, Entangled Owl

Blake Milliken went to the school's soccer field to kick a ball around with ...

Uber Charged A Passenger $14,000 For A 21-Minute Ride

The company has since issued a full refund

Watch Lonely Rescued Cockatoo Find Happiness

It's impossible not to smile when watching this fun-loving cockatoo get the ...

Must-Have Travel Gadgets

Modern technology is making it easier and safer to go where we want to go ...

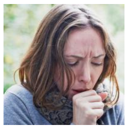

10 Signs You May Have Cancer

As a rule, cancer is hard to detect in its early stages

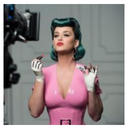

Katy Perry Asks Warren Buffett For His Thoughts On Cryptocurrency

Popstar Katy Perry set crypto twitter aflame when she posted a picture of ...

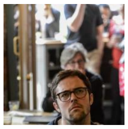

10 People You Should Not Be Tipping This Holiday Season

Holiday tipping can cost you more than feeding your family for an entire ...

Top
Home
Search
Mobile version
Help
FAQ
About
Contact
Science X Account
Sponsored Account
Newsletter
RSS feeds
Feature Stories
Latest news
Week's top
Archive
Android app
iOS app
Amazon Kindle
Push notification
Connect

© Phys.org 2003 - 2017, Science X network

Privacy Policy Terms of Use

Skip to main content

Login to your account

Publisher Main Menu Search
Search Search SpringerOpen articles Search

Publisher main menu

- Get published
- Explore Journals
- About
- Books

Take part in our short survey to give us your opinion on Springer Nature and its family of journals

Micro and Nano Systems Letters

- About
- Articles
- Submission Guidelines

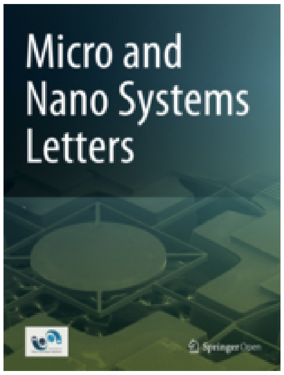

Review
Open Access

RF power harvesting: a review on designing methodologies and applications

- Le-Giang Tran[1],
- Hyouk-Kyu Cha[2] and
- Woo-Tae Park[1, 3]Email authorView ORCID ID profile

Micro and Nano Systems Letters2017**5**:14

https://doi.org/10.1186/s40486-017-0051-0

© The Author(s) 2017

Received: 8 November 2016

Accepted: 17 February 2017

Published: 28 February 2017

Abstract

Wireless power transmission was conceptualized nearly a century ago. Certain achievements made to date have made power harvesting a reality, capable of providing alternative sources of energy. This review provides a summ ary of radio frequency (RF) power harvesting technologies in order to serve as a guide for the design of RF energy harvesting units. Since energy harvesting circuits are designed to operate with relatively small voltages and currents, they rely on state-of-the-art electrical technology for obtaining high efficiency. Thus, comprehensive analysis and discussions of various designs and their tradeoffs are included. Finally, recent applications of RF power harvesting are outlined.

Keywords

Wireless power harvestingRadio frequencyImpedance matching networkVoltage multiplierAntennaWireless sensor networks

Background

Motivation for wireless energy harvesting

An early definition of a wireless power transmission system portrays a unit that emits electrical power from one place and captures it at another place in the Earth's atmosphere without the use of wires or any other supporting medium [1]. The history of RF power scavenging in free space originated in the late 1950s with a microwave-powered helicopter system [2]. Later, the concept of power harvesting or energy scavenging was explained as a technique for reaping energy from the external environment using different methods including thermoelectric conversion, vibrational excitation, solar energy conversion, and pressure gradients. This technique promises tremendous scope for the replacement of small batteries in low power electrical devices and systems. Figure 1 introduces the structure of an RF energy harvesting system and factors that contribute to the performance of the whole system.

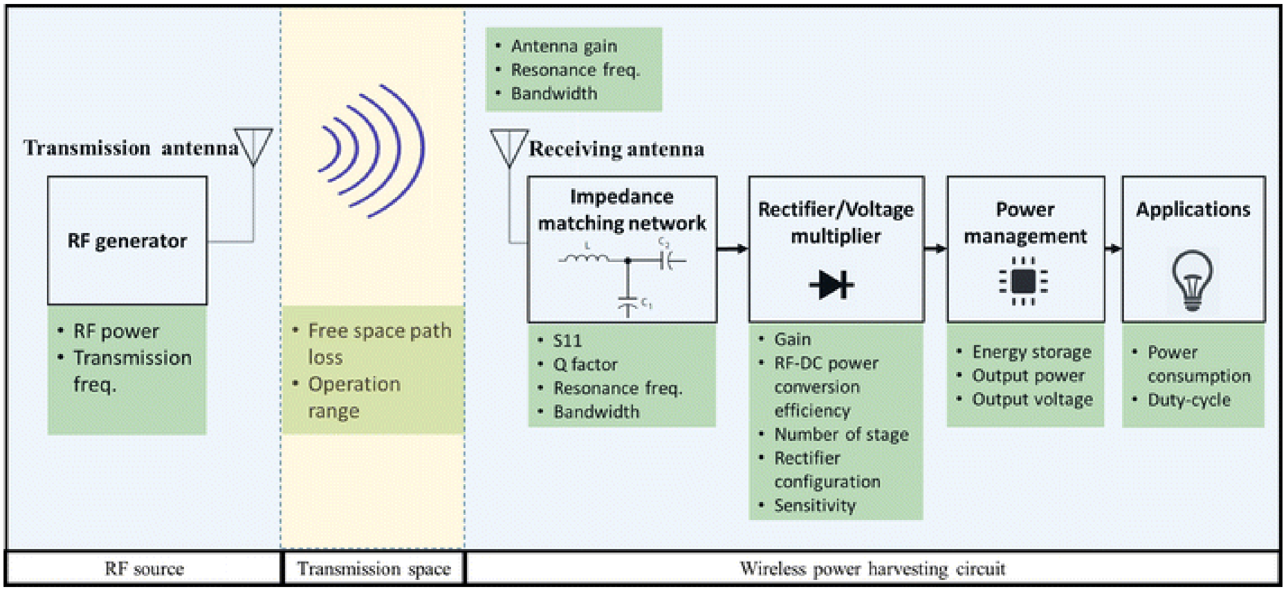
Antenna gain
Resonance freq.
Bandwidth
Transmission antenna
Receiving antenna
RF generator
Impedance matching network
L
C2
C1
Rectifier/Voltage multiplier
Power management
Applications
RF power
Transmission freq.
Free space path loss
Operation range
S11
Q factor
Resonance freq.
Bandwidth
Gain
RF-DC power conversion efficiency
Number of stage
Rectifier configuration
Sensitivity
Energy storage
Output power
Output voltage
Power consumption
Duty-cycle
RF source
Transmission space
Wireless power harvesting circuit

Fig. 1

Conceptual block diagram of an RF power harvesting system

RF wireless power harvesting (WPH) holds vast potential for replacing batteries or increasing their lifespans. Currently, batteries power a majority of low-power remote sensor devices and embedded equipment. In fact, batteries have finite lifespans and require periodic replacements. By applying power harvesting technologies, devices and equipment can become self-sustaining with respect to the energy required for operation, thereby obtaining an unlimited operating lifespan. Thus, the demand for power maintenance will become negligible.

The sources for WPH are available in many forms, such as solar power [3, 4, 5], wind energy [6, 7, 8], thermal energy [9, 10, 11], electromagnetic energy [12, 13, 14], kinetic energy [15, 16, 17], etc. Among them, electromagnetic energy is abundant in space and can be retrieved without limit. Electromagnetic waves come from a variety of sources such as satellite stations, wireless internet, radio stations, and digital multimedia broadcasting. A radio frequency power harvesting system can capture and convert electromagnetic energy into a usable direct current (DC) voltage. The key units of an RF power harvesting system are the antenna and rectifier circuit that allows the RF power or alternating current (AC) to be converted into DC energy.

The processing of battery wastes is a critical problem. A majority of batteries end up in landfills, leading to the pollution of the land and water underneath. The most effective solution for reducing battery wastes is to avoid using them. Applying WPH technology will help to reduce the dependency on batteries, which will ultimately have a positive impact on the environment. Moreover, the process of harnessing electromagnetic energy will not generate wastes as it is a clean energy source.

A summary of the present sources of energy available for power harvesting is shown in Table 1. The data in Table 1 was collected from references [18, 19]. In comparison with thermal or kinetic energy, electromagnetic energy is not limited by space or time. The RF wave is available both indoors and outdoors, in rural and urban areas, throughout the day. Despite its low power density in the environment, an intentional source can be added for more efficient power transmission and a boosting circuit can be built to suit the requirements of the load application. This feature promotes research to realize RF WPH technology through applications such as wireless sensor networks (WSNs) and Internet of Things (IoT).

Table 1

Overview of alternative sources of energy to replace batteries

Source	**Power density**	**Harvesting tech.**	**Advantages**	**Disadvantages**
Solar	Indoor: 10 μW/cm^2 Outdoor: 10 mW/cm^2	Photovoltaic	High power density Mature	Not always available Required exposure to light (not implantable) Expensive
Vibration	Human: 4 μW/cm^2 Industrial: 100 μW/cm^2	Piezoelectric Electrostatic Electromagnetic	Implantable High efficiency	Not always available Material physical limitation
Thermal	Human: 30 μW/cm^2 Industrial: 1–10 mW/cm^2	Thermoelectric Pyroelectric	High power density Implantable	Not always available Excess heat
RF	GSM: 0.1 μW/cm^2 WI-FI: 1 mW/cm^2	Antenna	Always available Implantable	Low density Efficiency inversely proportional to distance

Fundamentals of RF transmission

Understanding electromagnetic waves is important to design an RF WPH system. The behavior of electromagnetic waves varies according to the distance, frequency, and conducting environment. Depending on the requirements of the application, the designer needs to select suitable parameters for electromagnetic waves to obtain best results.

The loss of power in space can be characterized by free space path loss (FSPL), which is the loss of signal power during propagation in free space. Calculating FSPL requires information about the antenna gain, frequency of transmitting wave, and distance between the transmitter and receiver.

The behavior of electromagnetic waves depends upon the distance from the transmitting antenna. These characteristics are categorized into two segments: far-field and near-field [20]. While the electromagnetic wave pattern at the far-field is relatively uniform, in the near-field, the electric and magnetic components are very strong and independent such that one component can dominate the other. The near-field region is considered as the space that lies within the Fraunhofer's distance and the far-field region lies outside the Fraunhofer's distance. The distribution of the near and far field regions is illustrated in Fig. 2. The Fraunhofer's distance is defined as

$$d_f = \frac{2D^2}{\lambda},$$

(1)

where d_f is the Fraunhofer's distance, D is the maximum dimension of the radiator (or diameter of the antenna), and λ is the wavelength of the electromagnetic wave. Although the Fraunhofer creates a boundary region, the actual transition between regions is not distinct. Inside the near-field, the space from the antenna to a distance of $\sqrt[3]{\frac{D}{2\lambda}}$ is a non-radiative/reactive near-field region, where E and H fields are not in phase, creating energy distortion. As we propagate forward in this region towards the far-field, we encounter a radiative near-field or Fresnel region, where the reactive behavior of electromagnetic waves is not dominant but the phases of E and H fields still vary with distance.

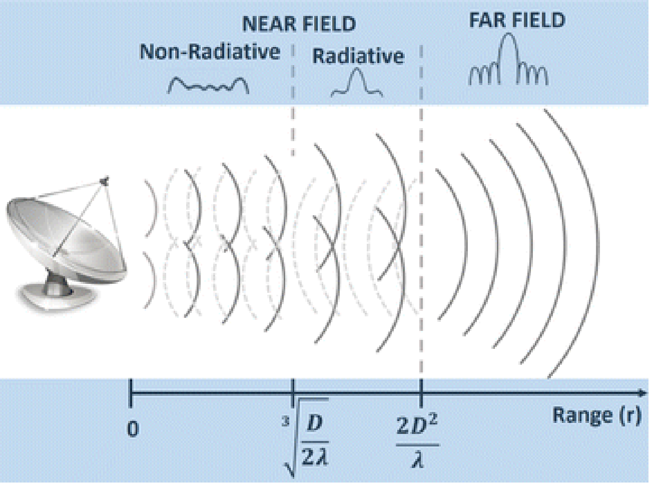

Fig. 2

Distribution of near field and far field regions in space

For a transmitter–receiver antenna in the far-field free space, the power propagation at the receiver antenna can be expressed as

$$P_R = \frac{P_T G_T G_R \lambda^2}{(4\pi R)^2},$$

(2)

where P_R is power at the receiver antenna; G_R is receiver antenna gain relative to the isotropic source (dBi); λ is the wavelength of the electromagnetic signal, which is equal to the speed of light in vacuum divided by the signal frequency, $\lambda = \frac{c}{f}$; and $k = 2\pi/\lambda$ is the wave number. From the above formula, the FSPL, P_L for far-field can be inferred as

$$P_L = \frac{P_T}{P_R} = \frac{(4\pi R)^2}{G_T G_R \lambda^2} = \frac{(4\pi f R)^2}{G_T G_R c^2} = \frac{4}{G_T G_R}(kR)^2$$

(3)

or

$$P_L\,(dB) = 20\log_{10}(f) + 20\log_{10}(R) + 20\log_{10}\left(\frac{4\pi}{c}\right) - G_T - G_R$$

(4)

In case f is measured in MHz, distance R is measured in km, and gain G_T and G_R are measured in dBi, the above function becomes

$$P_L\,(dB) = 20\log_{10}(f) + 20\log_{10}(R) + 32.44 - G_T - G_R$$
(5)

By using the path loss equation, it is possible to indicate the signal power at the far field region. However, this equation does not present all the factors that affect the propagation process such as reflection, diffraction, absorption, etc.

In contrast to the uniform far-field wave, the electric and magnetic fields behave unusually in the near-field. The relationship between electric and magnetic waves in the near-field varies unpredictably with time and space and is too complicated to predict. All these factors make estimating the power density at this range problematic.

The FSPL estimation process is critical to design the energy-harvesting unit. Being informed about the amount of power that the system is required to handle helps the designer in choosing the right technology and method.

Since the electromagnetic field (EMF) is widely used for a variety of purposes, there must be a standard to ensure the safe operation of EMF devices. For this purpose, the International Electrotechnical Commission [21] established international standards and frequency limits for EMFs. For safety measures, there must be a separate emission and immunity standard for each class of application including broadcast, communication, military, medical, and household appliances.

On the other hand, radiation exposure must be considered carefully, especially in biomedical applications. EMFs at different spectrums affect the human body differently. Between 1 MHz to 10 GHz frequency, EMF possibly penetrates through tissues and produce heat because of energy absorption. EMFs over 10 GHz are strongly blocked by the skin, and then the heat created causes damage such as eye cataracts or skin burns if the field density is above 1000 W/m^2. Hoping to prevent negative health effects when utilizing EMF applications, the International Commission on Non-Ionizing Radiation Protection (ICNIRP) developed general guidelines for EMF exposure limits recognized under the World Health Organization (WHO). To date, there have been no consistent reports corroborating that lasting exposure to EMF shortens one's life span or induces adverse health conditions. However, for safety issues, RF emission power and frequency must be considered carefully to avoid unwanted damages.

Wireless power harvesting evaluation metrics

There are various parameters that need to be evaluated, which decide the performance of a WPH design. Evaluation merits change depending on different applications. Nevertheless, critical values such as efficiency, sensitivity, operation distance, output power are defined as standards to make comparisons. However, tradeoffs exist between these values such as operation distance and overall efficiency. Moreover, besides this merit, other manufacturing auxiliary factors like low cost, maturity of fabrication process, and bulk manufacturing availability are also paramount.

Operation range

Operation distance is mostly related to operation frequency. In fact, transmission at high frequencies is attenuated by atmospheric conditions more than at low frequencies while low frequency penetrates deeper through matter. Therefore, if the application of WPH is for implantable devices, then the transmitting frequency should not exceed the megahertz range.

RF-DC power conversion efficiency (PCE)

This is the ratio between the amount of power applied to the load and that retrieved by the antenna. Generally, the RF-DC PCE covers the efficiency of the rectifier, voltage multiplier, and storage elements. PCE can be simply calculated as the ratio of power delivered to the load to the retrieved power. RF transmission loss in space is not considered in this term.

$$\eta_{PCE} = \frac{P_{load}}{P_{retrived}},$$
(6)

where P_{load} is power delivered to the load and $P_{retrieved}$ is harvested power at the antenna. Factors that determine the value of PCE include parasitic effects, leakage in the circuits, design topologies, and nonlinear thresholds of electrical components.

Resonator Q factor

The Q factor is generally defined as a dimensionless value that describes how strong the resonance is and the bandwidth of resonance [22]. In electric circuits, the Q factor represents how much the peak voltage increases when the system resonates at resonant frequency. From this concept, the equation of Q factor can be express as the formula below [23]:

$$Q = 2\pi \frac{Energy\, stored}{Energy\, dissipated\, per\, cycle}\, or\; Q = \frac{f_c}{\Delta f}$$

(7)

From the equation above, it is inferred that a high Q factor comes with a narrow resonant bandwidth, but high voltage gain is obtained at resonance. Further, this equation also indicates that the Q factor is inversely proportional to the energy dissipated per cycle. The specific Q factor for capacitor C and inductor L at frequency ω is given by

$$\text{Capacitor Q factor:}\, Q_C = \frac{1}{\omega R_c C} = \frac{X_c}{R_c},$$

(8)

$$\text{Inductor Q factor:}\, Q_L = \frac{\omega L}{R_L} = \frac{X_L}{R_L},$$

(9)

where R_C is the series resistance of capacitor and R_L is the series resistance of inductor. From the equations of Q_C and Q_L, we note that resistive component is the cause of power dissipation. Additionally, in an electrical circuit, energy loss can be reduced by adding reactive components such as capacitors or inductors and minimizing resistive components. Since obtaining a high Q factor is a usual consideration in designing WPH systems, this feature is typically included in the impedance matching network design.

Sensitivity

Sensitivity of a WPH system is defined as the minimum limit of incident power that is required to trigger the operation of the system. Sensitivity is quantified as follows:

$$Sensitivity\ (dBm) = 10 \log_{10}\left(\frac{P}{1mW}\right),$$

(10)

where P is the minimum power the system requires to perform a task.

The threshold voltage of CMOS technology affects sensitivity. A lower threshold CMOS is more sensitive but comes with more leakage current, which reduces the overall efficiency.

Output power

Usually, the outcome of a WPH system is DC power, which is characterized by load voltage V_{DD} and current I_{DD}. Measuring open-load voltage demonstrates the performance of WPH in general since V_{DD} and I_{DD} depend on load impedance. If the load is a sensor, V_{DD} is more important than I_{DD} while in applications like electrolysis or LED, current is the dominant parameter.

Recent advances in wireless power harvesting

Antenna/rectenna design

Gain, resonance frequency, bandwidth are parameters that characterize the performance of an antenna. Assuming an unobstructed space and isotropic transmitting source, the spreading of the waves in all directions is uniform. Thus, the power per unit area at a distance from the source is inversely proportional to the square of the distance:

$$S_{isotropic} = \frac{P_T}{4\pi R^2},$$

(11)

where R is the distance from the source, S is power per unit area at the distance R, and P_T is the transmitted power. However, it is important to note that antennas do not always transmit power spherically (isotropic antenna), they also transmit energy in some specific directions according to their designs. The ratio between the maximum power density of an antenna at a given distance to the power density of an optimal omnidirectional or isotropic antenna at the same distance, radiating the same power, is known as antenna gain (G). This parameter represents the directivity of an antenna. Since power density (S) is non-directional, the definition of S was developed as a function of direction i.e. S (θ, φ). Consequently, the gain of the antenna is a directional function as well. It is defined as the ratio of the antenna's power density in a given direction to the power density of an isotropic antenna as formula below

$$G\left(\theta,\varphi\right)=\frac{\mathrm{S}\left(\theta,\varphi\right)}{S_{isotropic}}=\frac{4\pi R^2 S\left(\theta,\varphi\right)}{\mathrm{P}_t},$$

(12)

Thus, the power density at a distance R from an antenna in general is given by

$$S=\frac{P_T G_T}{4\pi R^2},$$

(13)

where P_T is the transmitted power by the antenna and G_T is the transmitting antenna gain. It is known that an ideal isotropic antenna has $G_T = 0$ dBi. The aforementioned formula is also applicable to receiving antennas. Specifically, for power harvesting applications, a receiving antenna, which constitutes a rectifier, is called a rectenna.

Preference for a high gain antenna depends on the application requirements. In case of RF transmission, if the positions of source and receiving antennas are known, then a high gain rectenna is advantageous. On the other hand, if the positions of the source and receiving antenna are relatively uncertain, a low gain antenna is preferable in order to collect signals from various directions simultaneously.

Every antenna has its own optimal operation frequency known as resonance frequency (Fig. 3). The resonance frequency is determined by the capacitance and inductance characteristics of the antenna. As frequency increases, inductive behavior becomes dominant and capacitance decreases. The frequency at which the inductance and capacitance nullify each other, minimizing the impedance of the antenna, is called the resonance frequency.

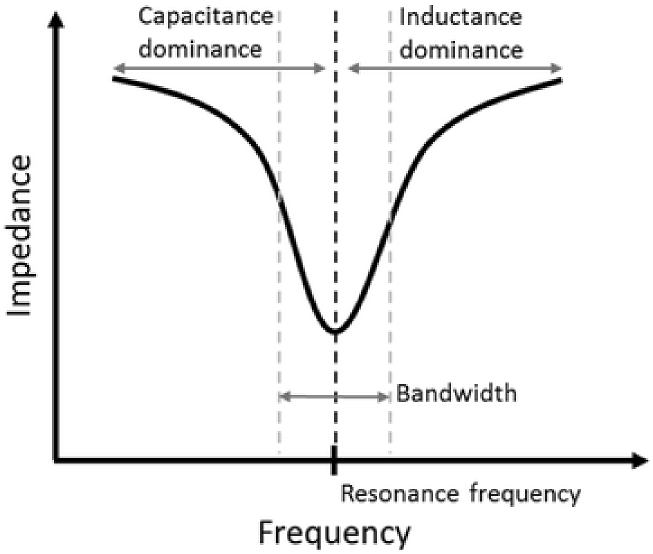

Fig. 3

Correlation between impedance and resonance frequency, and bandwidth of an antenna

The capacitance and inductance are functions of the frequency and physical size of the antenna. The larger the dimensions of the antenna, the lower the resonance frequency. Therefore, transmitting and receiving low frequency waves requires a large aperture, which is not suitable for small device applications. The antenna designed in [24] has a sensitivity as low as −35 dBm (0.32 µW), but the tradeoff is a large size aperture measuring up to 64.68 cm^3 (7 × 7 × 1.32 cm).

The bandwidth of an antenna is the range of frequencies in which the antenna can operate efficiently. A wide bandwidth antenna can collect signals from a wider range of frequencies than a narrow bandwidth antenna. Hence, a wide bandwidth antenna is advantageous in retrieving incident energy but presents a high risk of interference by noise through unwanted frequencies (Table 2).

Table 2

Comparison of published work regarding power-harvesting antennas

Ref.	Type	Gain (dBi)	Freq. (GHz)	Dimension (mm)	Sensitivity	RF-DC PCE
[25]	Air-substrate patch	7	2.45	261 × 5	–	30% @ 1 μW/cm^2
[26]	Polarization patch	–	5.1–5.8 5.8–6.1	90 × 160	–	23.8%–31.9% 22.7%–24.5%
[27]	Patch	–	2.45	100 × 70	–	73.9% @ 207 μW/cm^2
[28]	Dual-linearly polarized patch	7.45–7.63	2.45	70 × 47.5	–	78% @ 295.3 μW/cm^2
[29]	Microstrip	5.3	160	–	–	8.5% @ −2.46 dBm
[30]	Patch	–	1.9–3.2	40 × 40 × 0.8	20 μW/cm^2	70% @ 50 μW/cm^2
[31]	Patch	−20.20	0.915	Π × 6 × 1.584		~ 60%
[32]	Patch, metal strip	4.33 6.64	4.9 5.9	68 × 34	–	65.2% 64.8%
[33]	Stack differential	–	5.8	120 × 40	–	44.1% @ 0.041 W/m^2
[24]	Cross dipole	–	1.8–2.5	70 × 70 × 13.2	−35 dBm	55% @ −10 dBm
[34]	Dual-polarized patch	–	2.45	100 × 100 × 3.8	–	82.3% @ 22 dBm
[35]	1 × 4 quasi-Yagi	10.9 13.3	1.8 2.2	300 × 300 × 1.6	–	40% @ 455 μW/cm^2
[36]	Dipole	–	0.915 2.45	60 × 60 × 60	1 μW/cm^2	48% @ 0 dBm 39% @ 0 dBm
[37]	Microstrip	8.6	2.45	–		83%

Ref.	Type	Gain (dBi)	Freq. (GHz)	Dimension (mm)	Sensitivity	RF-DC PCE
[38]	Patch	4	2.45	–	–	70%
[39]	Patch	2.19 3.6	2.45 5.8	40 × 43	–	65% @ 10 mW/cm^2 46% @ 10 mW/cm^2

Communication antennas have been studied for decades. However, power-harvesting antennas are currently in the developmental stage. At first, antenna classification was based on design characteristics and applications. It originally included wire antennas, aperture antennas, printed planar antennas, and reflector antennas. An illustration of some examples of antennas is shown in Fig. 4. To date, the growth in technology has paved the way for a variety of antenna design and fabrication methods for making it more compact and mature.

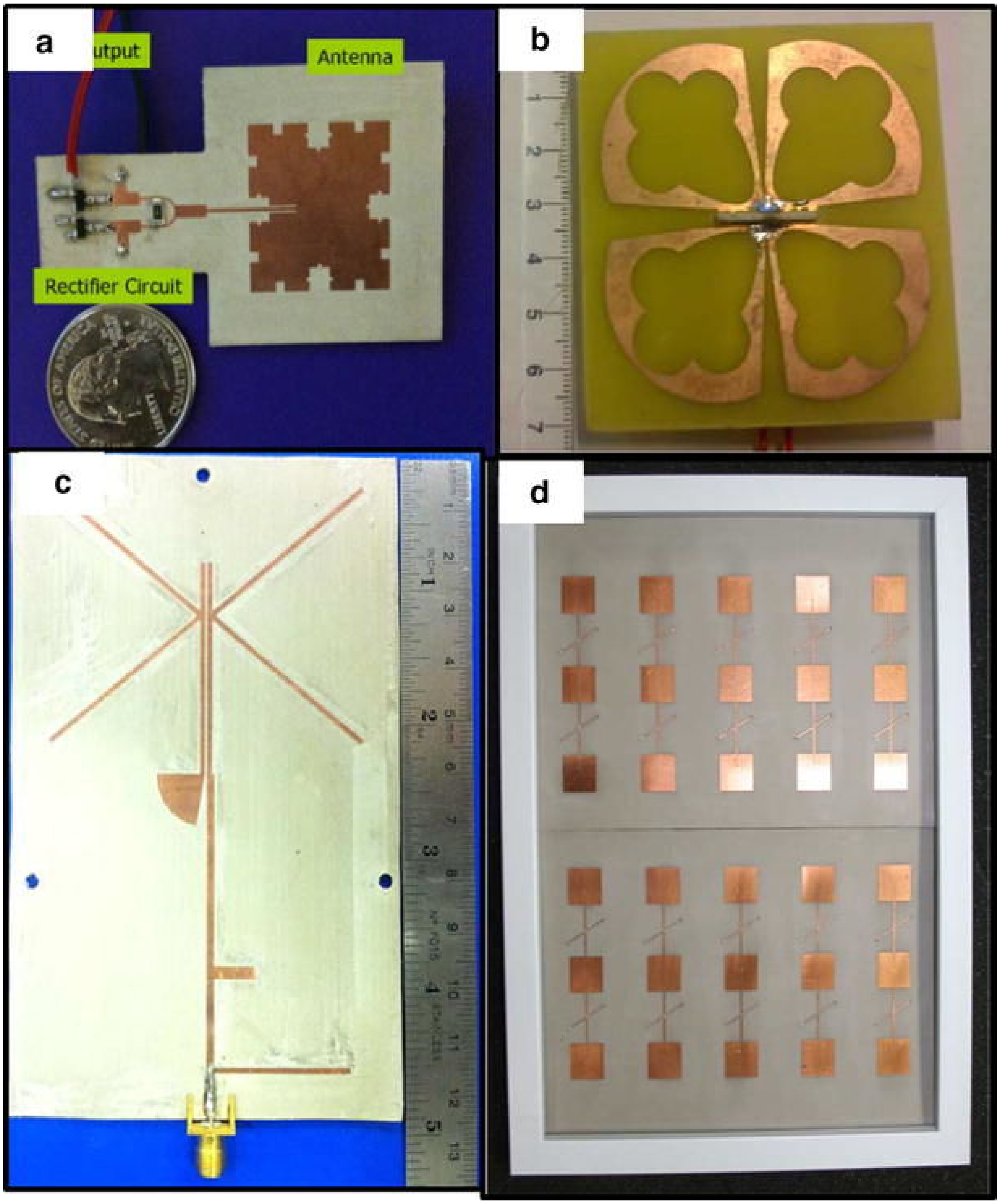

Fig. 4

Types of antennas **a** 2.45 GHz patch antenna with rectifier [38]. ©2010 IEEE. All rights reserved; **b** Planar dual-polarized antenna [24]. ©2015 IEEE. All rights reserved; **c** Microstrip antenna [37]. ©2012 IEEE. All rights reserved; **d** Array of stacked differential patch antenna [33]. ©2015 IEEE. All rights reserved

The plate antennas are popular and have many applications [27, 34, 38]; on-chip antennas are preferred for small and compact applications. Recently, many publications addressed wide-band and multi-band antennas. It has been proven that narrow-band antennas offer high energy conversion efficiencies but can only retrieve a limited amount of energy. On the other hand, wide-band or multi-band frequency antennas can retrieve more RF energy in space. However, the tradeoffs are low overall efficiency and large aperture. In [32], antennas with a resonance frequency of 4.9 and 5.9 GHz were designed with PCEs of 65.2 and 64.8%, respectively. Further work by Lu et al. [26] on polarization antennas supports the assertion that expanding the bandwidth of an antenna leads to increasing the amount of power harvested. In this work, the demonstration of broadband polarization antennas with three separate modes allows the antenna to operate in a wider range of frequencies. One common mechanism in the aforementioned works is the control of the antenna configuration by switching the diodes on and off, thus altering its resonant frequencies. However, since it uses separate modes for different frequencies, this antenna is not able to simultaneously resonate at two frequencies. On the other hand, the antenna presented in [39] is capable of operating at 2.45 and 5.8 GHz, simultaneously, providing 2.6 V output with a PCE of 65% and power density of 10 mW/cm^2.

The main purpose of aligning antennas in arrays is to enhance the antenna gain and obtain high voltage/current. Array antennas are preferred over large aperture antennas because they do not require large breakdown voltage diodes to operate. Antenna arrays can be connected before or after rectification. The first configuration enhances the retrieved power at the main beam while the second configuration expands the ability to retrieve power from various angles away from the main beam [40]. In case the RF waves are combined before rectification, the rectifier requires a large breakdown diode. If RF waves are combined after rectification, combining DC current becomes problematic. Antenna arrays can be connected in series or parallel to obtain high voltage or large current. Nonetheless, expanding the arrays yields better output but this might cause a deduction in conversion efficiency [33].

For demonstrating antenna arrays, Sun et al. [35] invented a T-junction to connect four quasi-Yagi antennas together. The advancement in this work was that the T-junction was flexible in changing from 1 × 4 array to 2 × 2 array topologies. Consequently, the system was able to operate at an ambient power level as low as 455 µW/cm^2 while obtaining 40% PCE.

Impedance matching network

In low-power consumption electrical systems, power leakage during transmission may lead to energy insufficiency. In these circumstances, adding an impedance matching network (IMN) ensures that the maximum power transfers between the RF source and load. For WPH applications, the receiving antenna is considered as the source while the rectifier/voltage multiplier is considered as the load. It is acknowledged that in DC, power transfer is optimum when the resistances of the source and load are indistinguishable. In an RF circuit, the impedance is referred to instead of resistance. An impedance mismatch between the source and load creates reflected power flow in the circuit that lowers the efficiency of the system. As its name indicates, the IMN ensures that the impedance of the source and load are identical by adding reactive components in between.

There are three basics matching configurations i.e. L, T and π matching networks (Fig. 5). The L matching is commonly used since it typically has two components, which simplifies the designing and controlling process. Additionally, the L matching networks do not alter the quality factor (Q) of the circuit.

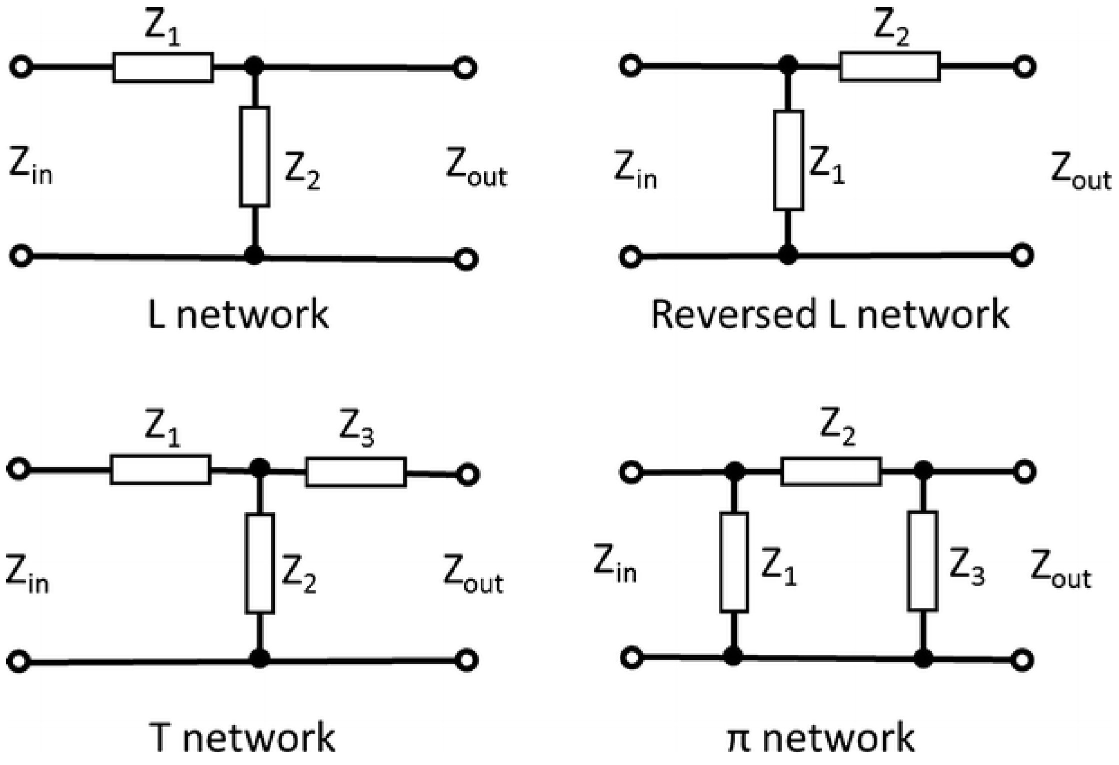

Fig. 5

Configuration of common impedance matching networks

The T and π matching configurations are more complex than the L network. Furthermore, organizing the T and π configurations into multiple stages will retain the final matching results but will change the Q factor. This strategy is useful in improving voltage boost.

There are tradeoffs between the attributes of an IMN, which include frequency, bandwidth, adjustability, and complexity. For instant, in [41], the suboptimal impedance matching and multiport ladder matching methods were introduced to enhance the matching performance and harvested power of antennas, respectively. However, the tradeoff was that implementation of these configurations required more components than traditional matching networks, thus, escalating the circuit's complexity.

Etor et al. [42] designed an IMN for THz frequencies applications using transmission lines and self-designed metal–insulator-metal diodes instead of lumped components. Moreover, fixed IMN and tunable IMN [43, 44, 45] were introduced as a technique for better matching with wide-band and multi-band antennas.

Rectifier/voltage multiplier

RF energy extracted from free space usually possesses low power density since the electric field power density decreases at the rate of $1/d^2$, where d is the distance from the RF source [46]. Therefore, a power amplifier circuit is required that yields enough DC energy from the electromagnetic waves to drive the loads. This gives rise to two possibilities, if the power consumption of the load is lower than the average power harvesting, the electronic devices at the load may work continuously; otherwise, if the load consumes more energy than the power harvesting circuit can generate, the devices cannot work continuously [47].

Rectifying is the most popular application of diodes, which refers to the conversion of AC current to DC current. In terms of power harvesting application, the RF signal retrieved in the antenna has a sinusoidal waveform. The signal after transformation through IMN would be rectified and boosted to meet the power requirements of the application.

The most fundamental topology of the rectifier is the half-wave rectifier that comprises of a single diode D1 (Fig. 6a). When AC voltage transfers through D1, only the positive cycle remains and the negative cycle is cutoff; thus, it diminishes half of the AC power. Moreover, the output V_{out} is discontinuous since the negative cycle is cutoff. Despite its simplicity, a half-wave rectifier is usually inadequate for common applications. Hence, a full-wave rectifier is more preferable. The circuit design of the full-wave rectifier is shown in Fig. 6b. During the first negative cycle of AC input, diode D1 is conductive and capacitor C1 is charged to the corresponding energy level of V_{peak} of the input. Then, at the next positive cycle, diode D1 is blocked, diode D2 is conductive so that capacitor C2 is also charged. In consequence, the output V_{out} would see two capacitors in series (each one is storing a voltage of $V_{peak)}$. Thus, V_{out} is twice V_{peak}. Therefore, this topology is more stable and efficient than the half-wave rectifier. There is also a bridge rectifier that rectifies both positive and negative cycles of the AC input but retains $V_{out} = V_{peak}$ by alternatively blocking pairs of diodes D1, D4 and D2, D3 (Fig. 6c).

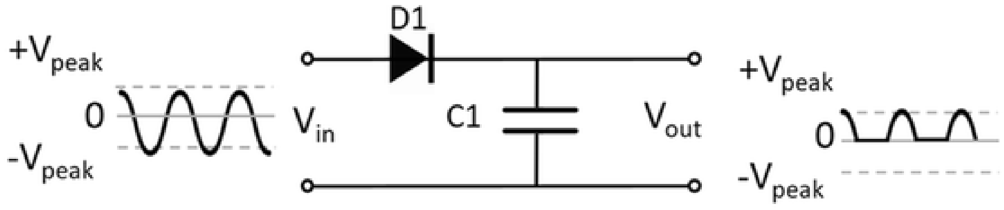

a Half-wave

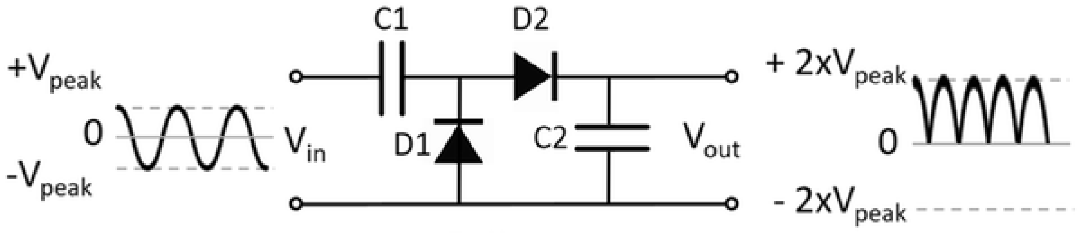

b Full-wave

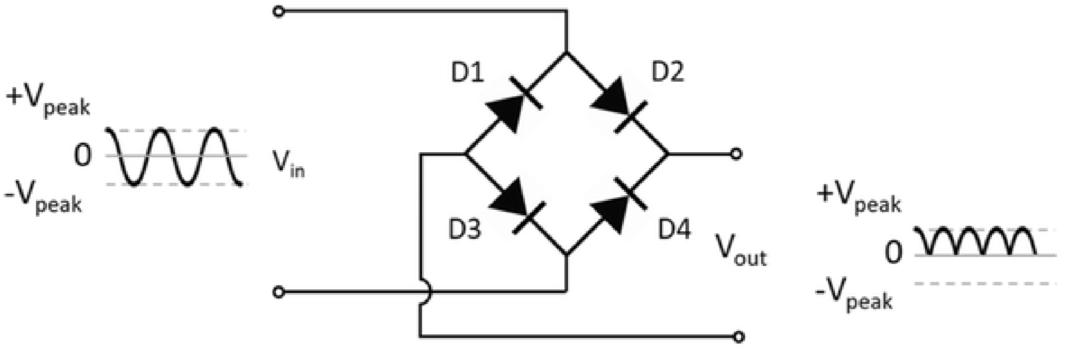

c Bridge

Fig. 6

Some common topologies of a rectifier

Voltage multiplier is a special type of rectifier circuit that converts and boosts AC input to DC output. In some case where the rectified power is inadequate for the application, there is a need for boosting the output DC by stacking single rectifiers into series, forming the voltage multiplier [48]. Several configurations of the voltage multiplier are shown in Fig. 7. The most fundamental configuration is the Cockcroft–Walton voltage multiplier (Fig. 7a). This circuit's operational principle is similar to the full-wave rectifier (Fig. 6b) but has more stages for higher voltage gain. The Dickson multiplier in Fig. 7b is a modification of Cockcroft–Walton's configuration with stage capacitors being shunted to reduce parasitic effects. Thus, the Dickson multiplier is preferable for small voltage applications. However, it is challenging to obtain high PCE due to the high threshold voltage among diodes creating leakage current, thus reducing the overall efficiency. Additionally, for high resistance loads, output voltage drops drastically leading to low current supply to the load. A summary of recent works related to voltage multiplier is shown in Table 3.

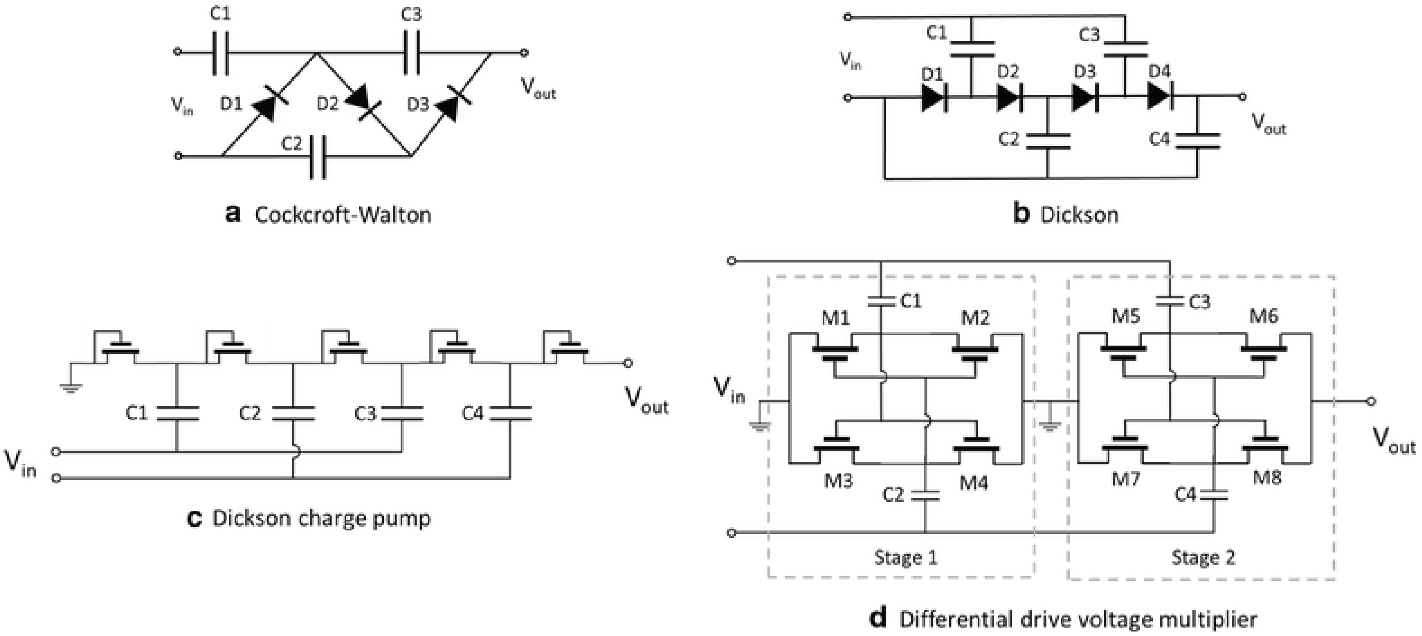
C1
C3
Vin
D1
D2
D3
Vout
C2
a Cockcroft-Walton
C1
C3
Vin
D1
D2
D3
D4
Vout
C2
C4
b Dickson
C1
C2
C3
C4
Vout
Vin
c Dickson charge pump
C1
M1
M2
C3
M5
M6
Vin
Vout
M3
M4
C2
M7
M8
C4
Stage 1
Stage 2
d Differential drive voltage multiplier

Fig. 7

Common voltage multiplier configurations: **a** Three stages Cockcroft–Walton voltage multiplier, **b** Four stages Dickson voltage multiplier, **c** Four stage Dickson voltage multiplier using CMOS technology, **d** Two stages voltage multiplier comprised of differential drive unit

Table 3

Published work comparison regarding voltage multiplier

Ref.	Rectifier topology	No. of stage	Freq. (MHz)	Tech.	Range (dBm)	Maximum PCE	Size
[51]	Differential	2	433	0.18 µm CMOS	0–20	74% @ 2 dBm	–
[52]	PMOS transistors	7	900	40 nm CMOS	–	44% @ > 390 mV	0.04 mm^2
[53]	Half-wave	4	900	0.18 µm CMOS	–	37.42% @ 390 mV	–
[54]	Comparator-based/active-diode	3	13.56	0.18 µm CMOS	8–15	67.9% @ 12.8 dBm	–
[55]	Dickson	3	13.56	250 nm CMOS	–	72%	0.13 mm^2
[56]	–	–	915	0.13 µm CMOS	−21.6	22.6% @ −16.8 dBm	0.186 mm^2

Ref.	Rectifier topology	No. of stage	Freq. (MHz)	Tech.	Range (dBm)	Maximum PCE	Size
[57]	–	–	900, 2400	130 nm CMOS	–	41% @ −20.6 dBm	–

The MOSFET (metal–oxide–semiconductor field effect transistor) technology is overcoming the limitations of diodes and becoming an alternate solution for rectifying and boosting. Owing to MOSFET technology, Dickson multiplier can be integrated together in integrate circuits (IC) by replacing diodes with NMOS as shown in Fig. 7c. Relatively low threshold voltages and high PCEs are features of this design. Moreover, differential drive voltage multiplier (Fig. 7d) is widely used because of its low leakage current and potential for further modification for specific applications. A detail explanation and analysis is presented in [49, 50].

There is a strong relationship between the number of stages of a voltage multiplier and its sensitivity and efficiency. As the number of stages increase, there is more loss across each added stage. However, the tradeoff is higher voltage multiplication and small threshold voltage at the first stage. On the other hand, a voltage multiplier with a few stages has less voltage drop between its stages but requires higher threshold voltage for all stages to work simultaneously. For this reason, a voltage multiplier becomes more sensitive when a large number of stages are present and becomes more efficient at fewer stages. This tradeoff feature was analyzed in several researches [53, 58]. Therefore, the optimal number of stages should be considered depending upon the application targets.

Pros and cons of Schottky diode and CMOS technology

Rectifying elements such as diodes or transistors are indispensable components of the rectifier/voltage multiplier block, which determine operation frequency and power-conversion efficiency. Traditionally, a Schottky diode was used because of its low threshold voltage. Moreover, there are also Esaki (tunnel) diodes, metal–insulator–metal (MIM) diodes, and spin diodes with recent technology enhancements that make it more mature. Particularly, the Esaki diode can operate at very high frequency with fast response because of its low parasitic elements. Moreover, the MIM diodes technology makes integration with CMOS process possible, which is a limitation of Schottky diodes. Spin diodes offer lower threshold voltages than Schottky diodes.

There are several parameters that need to be considered when choosing a diode for designing a voltage multiplier. These include RF to DC power conversion efficiency, parasitic efficiency, and matching efficiency. When it comes to microwatt applications, Schottky diodes face some limitation in RF-DC conversion efficiency because of their high zero-bias junction resistance [59]. However, owing to recent developments in technology, if the retrieved power is less than −30 dBm (1 µW), then the low-barrier SMS7630 and VDI W-Band ZBD diodes are recommended due to their low parasitic power losses and low power dissipation in the matching network [60].

Recently, many publications have moved toward CMOS processes, since they not only assist in custom designing in electronics but are also more sensitive to low operation voltages than traditional Schottky diodes. Many researches focus on modifying present rectifier and voltage multiplier topologies to achieve higher gain, sensitivity, efficiency [52, 55, 56, 57, 61]. For instance, the voltage multiplier achieved maximum 11% PCE at −24 dBm (4 µW) in [62] and 41% PCE at −20.6 dBm (8.7 µW) input power in [57]. While the majority of publications utilized 0.18 µm CMOS technology in their work, [52] applied commercial 40 nm CMOS process to build a low-voltage operation voltage multiplier that reached 44% PCE at 390 mV input. Hwang et al. [54] applied reducing reverse loss to minimize the reverse leakage loss in the rectifier, thus increasing the overall efficiency.

Similar to Schottky diodes, the number of stages directly dictates the performance efficiency in the task of designing a custom CMOS voltage multiplier. A basic assumption when using transistors is that the retrieved RF signal from the antenna must be high enough to trigger the transistors on and off. Different voltage multiplier topologies possess different features and characteristics. However, as long as transistors are used in the design, there is considerable amount of voltage drop among transistors. It has been observed that the major cause of rectified voltage drop is the threshold loss. If the number of stages increases, the threshold voltage at the transistors also adds up, requiring the higher amplitude RF signal to trigger the transistors [62]. Therefore, the number of stages in the majority of publications usually does not exceed 10.

Designing and optimizing the WPH system

The efficiencies of individual modules and the integration of all modules together constitute the total efficiency of a WPH system. As a consequence, the only way to optimize the total efficiency is to maximize efficiencies of each module and consolidate them in harmony. Commonly, several tests and adjustments need to be done to obtain an optimized WPH system. The work flow of designing a WPH system is described in Fig. 8.

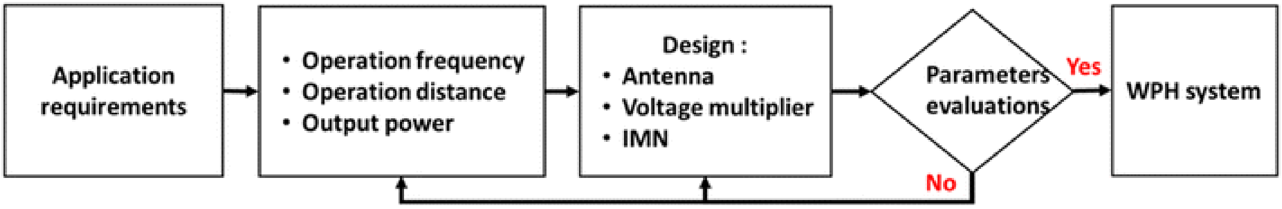

Fig. 8

WPH system's designing protocol

Firstly, choosing the appropriate operation frequency to the application's requirement critically dictates the efficiency of the WPH system. Secondly, the operation range also need to be specified. Long range (far-field) power harvesting or transmission requires high frequency such as 2.45, 5.24 GHz, whereas short range (near-field) applications requires electromagnetic waves at frequencies in the megahertz range. Additionally, for power harvesting in dense environment different from air, very low frequency (~kHz) is preferable. Besides the operation frequency and distance, the required output power and voltage determines the suitable topology of the voltage multiplier.

Based on the operation and application, the antenna's design has to match the gain, frequency and size. Choosing the right rectifying element is also critical. For rectifier and voltage multiplier modules in general, the dependence of RF-DC PCE on frequency and output load was discussed in [50]. In addition, it is obligated to have a correct-tuned IMN for maximum power transfer inside the WPH circuit.

Detailed features of individual modules of the WPH system are explained and discussed in the previous sections. The problem of optimizing the WPH system becomes the game of defining the goals and integrating individual modules in harmony to archive the goal.

Applications

RF power harvesting in medical and healthcare

In order to deploy WPH in the real world, power consumption rate of the device should be less than the harvested power. Since retrieved power is unstable and difficult to predict, an energy storage module is highly recommended to enhance its consistency. If the WPH is integrated to the device as a single system on chip, the total size of the system would be reduced significantly.

The first wireless battery-free bio-signal processing system on chip [63] was introduced by [64]. This system was able to monitor various bio-signals via electrocardiogram (ECG), electromyogram (EMG), and electroencephalogram (EEG). The total size of the chip was 8.25 mm^2. This chip only consumed 19 µW to measure heart rate. The power module of this system comprised an RF WPH that was supported by a thermoelectric generator, which together supplied a voltage of 1.35 V. Table 4 presents the performance of several health monitoring system on chips (SoCs). The majority of chips have the advantages of small size, low voltage supply and low power consumption which is applicable to be integrated with WPH technology.

Table 4

Power consumption of published SoC devices

Ref.	Sensors	Operation voltage (V)	Chip power (µW)	Size (mm)
[21]	Intraocular pressure (IOP)	–	0.0053	0.5 × 1.5 × 2.5
[65]	ECG, Bio-impedance	1.2	345–374	7 × 7

Ref.	Sensors	Operation voltage (V)	Chip power (µW)	Size (mm)
[66]	ECG	1.2	32	5 × 4.7
[64]	ECG, EMG, EEG	1.2	19	2.5 × 3.3
[67]	ECG, thoracic impedance variance (TIV)	1.2	2400	5 × 5
[68]	EEG	1	77.1	2.5 × 2.5
[69]	Temperature, pressure	0.4/0.5	7.7	2.2 × 2.7
[70]	Neural, ECG, EMG, EEG	1	500	2.5 × 1

It is a fact that a considerable amount of TV broadcast signals never reach the TV. Fortunately, WPH technology can harness this wasted energy without causing any effect on the TV broadcasting quality. In this context, Nishimoto et al. [71] have explored this potential source of energy to remove batteries from temperature monitoring WSNs. Since the power density of TV broadcasts is relatively weak, an array antenna was designed to collect multi-frequency waves. Compared to the antenna array designed in [72], the antenna rectifying efficiency obtained was higher by 50% in the frequencies between 15 and 800 MHz. This design is reported to successfully measure and transmit data every 5 s.

In some research experiments where free movement is unavoidable, stimulation signal and response data has to be delivered and collected wirelessly. In these circumstances, RF WPH is a superior choice in comparison with inductive coupling transmission technology to power the wireless system. Cobo et al. designed an implantable micro-pump that is powered by inductive coupling coils. Even though this work has obtained high efficiency power transmission, the operation range is limited to a few centimeters from the source. On the other hand, in [73], the RF electromagnetic waves were deployed to power optogenetic microneedles to be wirelessly injected into mice. The design of WPH in this work was able to retrieve 4.08 mW power at a distance of 1 m from the 7.9 W source at a frequency of 910 MHz. Another rectenna was built for mounting on the mice’ heads for deep brain stimulation (DBS) devices [31]. In this system, the power-harvesting module retrieves 0.254 mW DC power at a maximum operation distance of 20 cm. With this available power, the DBS device generates 200 µA pulses at 130 Hz. The above results show that RF WPH technology offers a better operation range than traditional inductive coupling power transmission technology. It is now possible to harvest sufficient power to run low-power consumption devices in the range of meters. The noninvasive bio-telemetry system for monitoring and investigating various health indexes is a promising field for WPH technology. Cheng et al. [74] introduced a power harvesting design that included a loop antenna to be attached to the pig eye with 31% PCE and output power up to 2.10 mW. Additionally, in vivo experiments, using WPH technology to supply medical devices for implantation in rapid eyes was demonstrated without recorded side effects [63, 75].

Wireless power harvesting network (IoT/WSN)

Along with the development of the micro-electro-mechanical systems (MEMS) technology, wireless sensor networks have gained widespread popularity in recent years. This field has achieved several milestones and is still developing further. WSNs applications widespread from smart house, healthcare to industry and military. In addition to the issues of sensing reliability, communication protocol, and network services, energy is a critical concern in WSNs [76]. Sensor nodes of WSNs are low power devices, and minimization of energy consumption and application of WPH to replenish power storage elements are two concurrent schemes to maximize the network lifespan. Recently, several researches were done to find optimized approaches to power WSNs including RF power harvesting [77, 78, 79, 80, 81].

Energy is the greatest challenge in deploying WSNs [82]. Applying RF energy to recharge the batteries is one approach to enhance the lifespan of WSNs [83, 84]. Lee et al. [85] is an example of an RF charger utilizing WPH technology. The system was able to work with a minimum input power of −10 dBm (0.1 mW) in the frequency range from 1.96 to 1.98 GHz to charge a 5 V super-capacitor. Maximum RF to DC power conversion efficiency is 81% at 6 dBm (3.98 mW). Gudan et al. [86] also designed a WPH system to charge a NiMH battery using either ambient Wi-Fi or Bluetooth signals. With the sensitivity down to −20 dBm, this system was capable of charging a battery to 5.8 μJ after 1 h.

Table 5 lists some low-power consumption sensor applications that can be integrated with a WPH module. The majority of these sensors were powered by battery with limited lifetime. Once WPH technology is implemented with these applications, it could be used in as a wireless sensor node. In particular, the work presented in [81] utilized RF power harvesting into WSNs for building's structure monitoring. In this work, far-field RF powered wireless sensors were developed with novel data transmission approach to detect humidity, temperature and light inside a building. Results show at a distance of 1 m from 3 W source, the sensor node received 3.14 mW power through air, 1.53 mW of power through 2 inches of brick, 2.88 mW through wood and 0.7 mW through 2 inches of steel. This amount of power is sufficient for the sensors presented in [88, 89, 90, 91]. This result indicates the applicability of RF WPH toward WSNs.

Table 5

State-of-the-art published sensors for WSNs

Ref.	**Sensors**	**Total power consumption (mW)**	**Supplied voltage (V)**	**Power source**	**Applications**
[87]	Pressure	12.69		Battery	Tire pressure monitoring
[88]	Pressure	0.03		Battery	Blood pulses monitoring, small vibration recording
[89]	Temperature	0.0068	1.5-2		RFID, implantable devices
[90]	Temperature	0.065	2.5–5.5		Room temperature monitoring
[91]	Temperature	0.000071	1.2	Battery	Environment monitoring
[92]	Gas	24		DC power source	C_2H_5OH detection
[93]	Gas	2	5		Smart phone gas sensor, air monitor

Furthermore, placement of power transmitters and sensor nodes in WSNs also play a critical role in the efficiency of the system [94, 95, 96]. To enhance directivity of RF transmissions, Shao et al. [97] proposed a beam-steered antenna array to direct its main beam toward energy harvesting units. By controlling the length of transmission line to each antenna unit in an array, the phase radiating difference between units created shift in the main beam. By this way, the harvesting unit remains fed with energy even it was misaligned with the source by 35°. Using the harvesting unit presented in [98] as a demonstration, within 1 min, this system was able to harvest 68.7 μJ at 0° and 13.6 μJ at 35° at a distance of 2 m from the beam-steered source antenna array.

Conclusion

This paper summarized the state-of-the-art of RF power harvesting technology in recent years. This technology will play a key role in replacing batteries in the near future. Some applications of RF power harvesting have been practically realized. A basic RF power-harvesting unit includes three main modules: the antenna, IMN, and voltage multiplier. The total efficiency of the system is dictated by the harmony in the integration of all the modules. The designs and principles of each module were also discussed in the paper.

The RF electromagnetic waves are harmless, abundant in space, and is able to penetrate through soft tissues. Those are properties that make RF electromagnetic waves an alternative source of energy to replace batteries in many applications. Particularly, RF power harvesting supports low-power medical and healthcare devices and facilitates the development of WSNs and IoT by providing mobility of use. Additionally, the progress in integrating RF power harvesting circuits into CMOS technology creates a completely wireless SoC.

Besides progressive accomplishment in recent years, there are still a variety of rooms to further optimize the RF power harvesting technology such as increasing operation range, reducing transmission loss, optimizing PCE, and minimizing system dimensions are the targets of RF power harvesting research. Furthermore, research focusing on harsh working environments for RF WPH such as implantation conditions or underwater zones is attracting much attention to extend the capabilities of this technology. A fine manufacturing and packing process with cost-effective production is necessary to make this technology mature.

In short, the RF WPH technology is gradually becoming a reality. Although this technology still faces many issues, overcoming these challenges can lead the power industry into a new era of clean and sustainable energy.

Declarations

Authors' contributions

LGT carried out literature search and wrote the manuscript. HKC and WTP reviewed/edited the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Grant No. 10060065 from the Industrial Source Technology Development Programs of the MOTIE (Ministry Of Trade, Industry and Energy), Korea. This work was also supported by the Radiation Technology R&D program through the National Research Foundation of Korea funded by the Ministry of Science, ICT & Future Planning (NRF-2013M2A2A9043274).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Authors' Affiliations

(1)
Convergence Institute of Biomedical Engineering and Biomaterials, Seoul National University of Science and Technology
(2)
Department of Electrical and Information Engineering, Seoul National University of Science and Technology
(3)
Department of Mechanical and Automotive Engineering, Seoul National University of Science and Technology

References

1. *Brown WC (1996) The history of wireless power transmission. Solar Energy 56:3–21*

2. *Brown WC (1969) Experiments involving a microwave beam to power and position a helicopter. IEEE Trans Aerosp Electron Syst AES-5:692–702*
3. *Raghunathan V, Kansal A, Hsu J, Friedman J, Srivastava M (2005) Design considerations for solar energy harvesting wireless embedded systems. In: Proceedings of the 4th international symposium on Information processing in sensor networks, p 64*
4. *Brunelli D, Benini L, Moser C, Thiele L (2008) An efficient solar energy harvester for wireless sensor nodes. In: 2008 design, automation and test in europe, pp 104–109*
5. *Abdin Z, Alim MA, Saidur R, Islam MR, Rashmi W, Mekhilef S et al (2013) Solar energy harvesting with the application of nanotechnology. Renew Sustain Energy Rev 26:837–852*
6. *Ackermann T, Söder L (2000) Wind energy technology and current status: a review. Renew Sustain Energy Rev 4:315–374*
7. *GM Joselin Herbert, S. Iniyan, E. Sreevalsan, and S. Rajapandian, "A review of wind energy technologies," Renewable and Sustainable Energy Reviews, vol. 11, pp. 1117-1145, 8//2007*
8. *Şahin AD (2004) Progress and recent trends in wind energy. Prog Energy Combust Sci 30:501–543*
9. *Xin L, Shuang-Hua Y (2010) Thermal energy harvesting for WSNs. In: 2010 IEEE international conference on systems man and cybernetics (SMC), pp 3045–3052*
10. *Dalola S, Ferrari V, Marioli D (2010) Pyroelectric effect in PZT thick films for thermal energy harvesting in low-power sensors. Procedia Eng 5:685–688*
11. *Cuadras A, Gasulla M, Ferrari V (2010) Thermal energy harvesting through pyroelectricity. Sens Actuators A Phys 158:132–139*
12. *Cao X, Chiang WJ, King YC, Lee YK (2007) Electromagnetic energy harvesting circuit with feedforward and feedback DC–DC PWM boost converter for vibration power generator system. IEEE Trans Power Electron 22:679–685*
13. *Beeby SP, Torah RN, Tudor MJ, Glynne-Jones P, Donnell TO, Saha CR et al (2007) A micro electromagnetic generator for vibration energy harvesting. J Micromech Microeng 17:1257*
14. *Yang B, Lee C, Xiang W, Xie J, He JH, Kotlanka RK, Low SP, Feng H (2009) Electromagnetic energy harvesting from vibrations of multiple frequencies. J Micromech Microeng 19:035001*
15. *Beeby SP, Tudor MJ, White NM (2006) Energy harvesting vibration sources for microsystems applications. Meas Sci Technol 17:R175*
16. *Challa VR, Prasad M, Shi Y, Fisher FT (2008) A vibration energy harvesting device with bidirectional resonance frequency tunability. Smart Mater Struct 17:015035*
17. *Khaligh A, Zeng P, Zheng C (2010) Kinetic energy harvesting using piezoelectric and electromagnetic technologies—state of the art. IEEE Trans Ind Electron 57:850–860*
18. *Vullers RJM, van Schaijk R, Doms I, Van Hoof C, Mertens R (2009) Micropower energy harvesting. Solid-State Electron 53:684–693*
19. *Akhtar F, Rehmani MH (2015) Energy replenishment using renewable and traditional energy resources for sustainable wireless sensor networks: a review. Renew Sustain Energy Rev 45:769–784*
20. *Yaghjian A (1986) An overview of near-field antenna measurements. IEEE Trans Antennas Propag 34:30–45*
21. *Chen G, Ghaed H, Haque RU, Wieckowski M, Kim Y, Kim G et al (2011) A cubic-millimeter energy-autonomous wireless intraocular pressure monitor. In: 2011 IEEE international solid-state circuits conference, 2011, pp 310–312*
22. *Harlow JH (2004) Electric power transformer engineering. CRC Press, Boca Raton*
23. *Lee TH (2004) The design of cmos radio-frequency integrated circuits. Commun Eng 2:47*
24. *Song C, Huang Y, Zhou J, Zhang J, Yuan S, Carter P (2015) A high-efficiency broadband rectenna for ambient wireless energy harvesting. IEEE Trans Antennas Propag 63:3486–3495*
25. *Momenroodaki P, Fernandes RD, Popovi Z (2016) Air-substrate compact high gain rectennas for low RF power harvesting. In: 2016 10th European conference on antennas and propagation (EuCAP), pp 1–4*
26. *Lu P, Yang XS, Li JL, Wang BZ (2016) Polarization reconfigurable broadband rectenna with tunable matching network for microwave power transmission. IEEE Trans Antennas Propag 64:1136–1141*
27. *Sun H (2016) An enhanced rectenna using differentially-fed rectifier for wireless power transmission. IEEE Antennas Wirel Propag Lett 15:32–35*
28. *Sun H, Geyi W (2016) A new rectenna with all-polarization-receiving capability for wireless power transmission. IEEE Antennas Wirel Propag Lett 15:814–817*
29. *Zhu P, Ma Z, Vandenbosch GAE, Gielen G (2015) 160 GHz harmonic-rejecting antenna with CMOS rectifier for millimeter-wave wireless power transmission. In: 2015 9th European conference on antennas and propagation (EuCAP), pp 1–5*
30. *Zhang J, Wu ZP, Liu CG, Zhang BH, Zhang B (2015) A double-sided rectenna design for RF energy harvesting. In: 2015 IEEE international wireless symposium (IWS), pp 1–4*
31. *Hosain MK, Kouzani AZ, Samad MF, Tye SJ (2015) A miniature energy harvesting rectenna for operating a head-mountable deep brain stimulation device. IEEE Access 3:223–234*
32. *Lu P, Yang XS, Li JL, Wang BZ (2015) A compact frequency reconfigurable rectenna for 5.2- and 5.8-GHz wireless power transmission. IEEE Trans Power Electron 30:6006–6010*

33. *Matsunaga T, Nishiyama E, Toyoda I (2015) 5.8-GHz stacked differential rectenna suitable for large-scale rectenna arrays with DC connection. IEEE Trans Antennas Propag 63:5944–5949*
34. *Chou JH, Lin DB, Weng KL, Li HJ (2014) All polarization receiving rectenna with harmonic rejection property for wireless power transmission. IEEE Trans Antennas Propag 62:5242–5249*
35. *Sun H, Guo Y, He M, Zhong Z (2013) A dual-band rectenna using broadband yagi antenna array for ambient RF power harvesting. IEEE Antennas Wirel Propag Lett 12:918–921*
36. *Niotaki K, Kim S, Jeong S, Collado A, Georgiadis A, Tentzeris MM (2013) A compact dual-band rectenna using slot-loaded dual band folded dipole antenna. IEEE Antennas Wirel Propag Lett 12:1634–1637*
37. *Hucheng S, Yong-Xin G, Miao H, Zheng Z (2012) Design of a high-efficiency 2.45-GHz rectenna for low-input-power energy harvesting. IEEE Antennas Wirel Propag Lett 11:929–932*
38. *Olgun U, Chen CC, Volakis JL (2010) Wireless power harvesting with planar rectennas for 2.45 GHz RFIDs. In: 2010 URSI international symposium on electromagnetic theory (EMTS), pp 329–331*
39. *Ren YJ, Farooqui MF, Chang K (2007) A compact dual-frequency rectifying antenna with high-orders harmonic-rejection. IEEE Trans Antennas Propag 55:2110–2113*
40. *Olgun U, Chen CC, Volakis JL (2011) Investigation of rectenna array configurations for enhanced RF power harvesting. IEEE Antennas Wirel Propag Lett 10:262–265*
41. *Shen S, Murch RD (2016) Impedance matching for compact multiple antenna systems in random RF fields. IEEE Trans Antennas Propag 64:820–825*
42. *Etor D, Dodd LE, Wood D, Balocco C (2015) Impedance matching at THz frequencies: optimizing power transfer in rectennas. In: 2015 40th international conference on infrared, millimeter, and terahertz waves (IRMMW-THz), pp 1–2*
43. *Hoarau C, Corrao N, Arnould JD, Ferrari P, Xavier P (2008) Complete design and measurement methodology for a tunable RF impedance-matching network. IEEE Trans Microw Theory Tech 56:2620–2627*
44. *Marrocco G (2008) The art of UHF RFID antenna design: impedance-matching and size-reduction techniques. IEEE Antennas Propag Mag 50:66–79*
45. *Mingo JD, Valdovinos A, Crespo A, Navarro D, Garcia P (2004) An RF electronically controlled impedance tuning network design and its application to an antenna input impedance automatic matching system. IEEE Trans Microw Theory Tech 52:489–497*
46. *Hatay M (1980) Empirical formula for propagation loss in land mobile radio services. IEEE Trans Veh Technol 29:317–325*
47. *Radiom S, Vandenbosch G, Gielen G (2008) Impact of antenna type and scaling on scavenged voltage in passive RFID tags. In: International workshop on antenna technology: small antennas and novel metamaterials, 2008. iWAT 2008, pp 442–445*
48. *Gosset G, Flandre D (2011) Fully-automated and portable design methodology for optimal sizing of energy-efficient CMOS voltage rectifiers. IEEE J Emerg Sel Top Circuits Syst 1:141–149*
49. *Facen A, Boni A (2007) CMOS power retriever for UHF RFID tags. Electron Lett 43:1424*
50. *Kotani K, Sasaki A, Ito T (2009) High-efficiency differential-drive CMOS rectifier for UHF RFIDs. IEEE J Solid-State Circuits 44:3011–3018*
51. *Chouhan SS, Nurmi M, Halonen K (2016) Efficiency enhanced voltage multiplier circuit for RF energy harvesting. Microelectron J 48:95–102*
52. *Wang W, Xiangjie C, Wong H (2015) Analysis and design of CMOS full-wave rectifying charge pump for RF energy harvesting applications. In: 2015 IEEE Region 10 conference TENCON 2015, pp 1–4*
53. *Rodriguez AN, Cruz FRG, Ramos RZ (2015) Design of 900 Mhz AC to DC converter using native Cmos device of TSMC 0.18 micron technology for RF energy harvest application. Univers J Electr Electron Eng 3:7*
54. *Hwang YS, Lei CC, Yang YW, Chen JJ, Yu CC (2014) A 13.56-MHz low-voltage and low-control-loss RF-DC rectifier utilizing a reducing reverse loss technique. IEEE Trans Power Electron 29:6544–6554*
55. *Haddad PA, Gosset G, Raskin JP, Flandre D (2014) Efficient ultra low power rectification at 13.56 MHz for a 10 μA load current. In: 2014 SOI-3D-subthreshold microelectronics technology unified conference (S3S), pp 1–2*
56. *Hameed Z, Moez K (2014) Hybird forward and backward threshold-compensated RF-DC power converter for RF energy harvesting. IEEE J Eng Sel Top Circuits Syst 4:9*
57. *Karolak D, Taris T, Deval Y, Béguéret JB et al (2012) Design comparison of low-power rectifiers dedicated to RF energy harvesting. In: 2012 19th IEEE international conference on electronics, circuits and systems (ICECS), pp 524–527*
58. *Kadupitiya JCS, Abeythunga TN, Ranathunga PDMT, De Silva DS (2015) Optimizing RF energy harvester design for low power applications by integrating multi stage voltage doubler on patch antenna. In: 2015 8th international conference on Ubi-Media computing (UMEDIA), pp 335–338*
59. *Hemour S, Zhao Y, Lorenz CHP, Houssameddine D, Gui Y, Hu CM et al (2014) Towards low-power high-efficiency RF and microwave energy harvesting. IEEE Trans Microw Theory Tech 62:965–976*
60. *Lorenz CHP, Hemour S, Wu K (2016) Physical mechanism and theoretical foundation of ambient RF power harvesting using zero-bias diodes. IEEE Trans Microw Theory Tech 64:2146–2158*
61. *Sun H, Xu G (2015) A differentially-driven rectifier for enhanced RF power harvesting. In: 2015 IEEE MTT-S international microwave workshop series on advanced materials and processes for RF and THz applications (IMWS-*

AMP), pp 1–3

62. *Papotto G, Carrara F, Palmisano G (2011) A 90-nm CMOS threshold-compensated RF energy harvester. IEEE J Solid-State Circuits 46:1985–1997*
63. *Lingley AR, Ali M, Liao Y, Mirjalili R, Klonner M, Sopanen M et al (2011) A single-pixel wireless contact lens display. J Micromech Microeng 21:125014*
64. *Zhang Y, Zhang F, Shakhsheer Y, Silver JD, Klinefelter A, Nagaraju M et al (2013) A batteryless 19 W MICS/ISM-band energy harvesting body sensor node SoC for ExG applications. IEEE J Solid-State Circuits 48:199–213*
65. *Helleputte NV, Konijnenburg M, Pettine J, Jee DW, Kim H, Morgado A et al (2015) A 345 uW multi-sensor biomedical SoC with bio-impedance, 3-channel ECG, motion artifact reduction, and integrated DSP. IEEE J Solid-State Circuits 50:230–244*
66. *Kim H, Kim S, Helleputte NV, Artes A, Konijnenburg M, Huisken J et al (2014) A configurable and low-power mixed signal SoC for portable ECG monitoring applications. IEEE Trans Biomed Circuits Syst 8:257–267*
67. *Yan L, Bae J, Lee S, Roh T, Song K, Yoo HJ (2011) A 3.9 mW 25-electrode reconfigured sensor for wearable cardiac monitoring system. IEEE J Solid-State Circuits 46:353–364*
68. *Verma N, Shoeb A, Bohorquez J, Dawson J, Guttag J, Chandrakasan AP (2010) A micro-power EEG acquisition SoC with integrated feature extraction processor for a chronic seizure detection system. IEEE J Solid-State Circuits 45(4):804–816*
69. *Chen G, Fojtik M, Kim D, Fick D, Park J, Seok M et al. (2010) Millimeter-scale nearly perpetual sensor system with stacked battery and solar cells. In: 2010 IEEE international solid-state circuits conference—(ISSCC), pp 288–289*
70. *Rai S, Holleman J, Pandey JN, Zhang F, Otis B (2009) A 500 µW neural tag with 2 µVrms AFE and frequency-multiplying MICS/ISM FSK transmitter. In: 2009 IEEE international solid-state circuits conference—digest of technical papers, pp 212–213*
71. *Nishimoto H, Kawahara Y, Asami T (2010) Prototype implementation of ambient RF energy harvesting wireless sensor networks. In: Sensors, 2010 IEEE, pp 1282–1287*
72. *Shinohara N, Kawasaki S (2009) Recent wireless power transmission technologies in Japan for space solar power station/satellite. In: 2009 IEEE radio and wireless symposium, pp 13–15*
73. *Kim T-I, McCall JG, Jung YH, Huang X, Siuda ER, Li Y et al (2013) Injectable, cellular-scale optoelectronics with applications for wireless optogenetics. Science 340:211–216*
74. *Cheng HW, Yu TC, Huang HY, Ting SH, Huang TH, Chiou JC et al (2014) Design of miniaturized antenna and power harvester circuit on the enucleated porcine eyes. IEEE Antennas Wirel Propag Lett 13:1156–1159*
75. *Chow EY, Yang CL, Ouyang Y, Chlebowski AL, Irazoqui PP, Chappell WJ (2011) Wireless powering and the study of RF propagation through ocular tissue for development of implantable sensors. IEEE Trans Antennas Propag 59:2379–2387*
76. *Yick J, Mukherjee B, Ghosal D (2008) Wireless sensor network survey. Comput Netw 52:2292–2330*
77. *Correia R, Carvalho NB, Kawasaki S (2016) Continuously power delivering for passive backscatter wireless sensor networks. IEEE Trans Microw Theory Tech 64:3723–3731*
78. *Praveen MP, Mehta NB (2016) Trade-offs in analog sensing and communication in RF energy harvesting wireless sensor networks. In: 2016 IEEE international conference on communications (ICC), pp 1–6*
79. *Collado A, Georgiadis A (2014) Optimal waveforms for efficient wireless power transmission. IEEE Microwave Wirel Compon Lett 24:354–356*
80. *Zhao Y, Chen B, Zhang R (2013) Optimal power allocation for an energy harvesting estimation system. In: 2013 IEEE international conference on acoustics, speech and signal processing, pp 4549–4553*
81. *Ruisi G, Hong P, Zhibin L, Na G, Jinhui W, Xiaowei C (2016) RF-powered battery-less wireless sensor network in structural monitoring. In: 2016 IEEE international conference on electro information technology (EIT), pp 0547–0552*
82. *Seah WKG, Eu ZA, Tan HP (2009) Wireless sensor networks powered by ambient energy harvesting (WSN-HEAP)—survey and challenges. In: 2009 1st international conference on wireless communication, vehicular technology, information theory and aerospace and electronic systems technology, pp 1–5*
83. *Jabbar H, Song YS, Jeong TT (2010) RF energy harvesting system and circuits for charging of mobile devices. IEEE Trans Consum Electron 56:247–253*
84. *Che W, Chen W, Meng D, Wang X, Tan X, Yan N et al (2010) Power management unit for battery assisted passive RFID tag. Electron Lett 46:589–590*
85. *Lee JH, Jung WJ, Jung JW, Jang JE, Park JS (2015) A matched RF charger for wireless RF power harvesting system. Microw Opt Technol Lett 57:1622–1625*
86. *Gudan K, Chemishkian S, Hull JJ, Thomas SJ, Ensworth J, Reynolds MS (2014) A 2.4 GHz ambient RF energy harvesting system with −20 dBm minimum input power and NiMH battery storage. In: RFID technology and applications conference (RFID-TA), 2014 IEEE, pp 7–12*
87. *Nagaraju MB, Lingley AR, Sridharan S, Gu J, Ruby R, Otis BP (2015) A 0.8 mm3 ± 0.68 psi single-chip wireless pressure sensor for TPMS applications. In: 2015 IEEE international solid-state circuits conference—(ISSCC) digest of technical papers, pp 1–3*
88. *Gong S, Schwalb W, Wang Y, Chen Y, Tang Y, Si J et al (2014) A wearable and highly sensitive pressure sensor with ultrathin gold nanowires. Nat Commun 5:3132*

89. *Souri K, Chae Y, Makinwa KAA (2013) A CMOS temperature sensor with a voltage-calibrated inaccuracy of ±0.15 °C (3σ) from −55 °C to 125 °C. IEEE J Solid-State Circuits 48:292–301*
90. *Aita AL, Pertijs MAP, Makinwa KAA, Huijsing JH, Meijer GCM (2013) Low-power CMOS smart temperature sensor with a batch-calibrated inaccuracy of ±0.25 °C (±3σ) from −70 °C to 130 °C. IEEE Sens J 13:1840–1848*
91. *Jeong S, Foo Z, Lee Y, Sim JY, Blaauw D, Sylvester D (2014) A fully-integrated 71 nW CMOS temperature sensor for low power wireless sensor nodes. IEEE J Solid-State Circuits 49:1682–1693*
92. *Moon SE, Lee HK, Choi NJ, Kang HT, Lee J, Ahn SD et al (2015) Low power consumption micro C2H5OH gas sensor based on micro-heater and ink jetting technique. Sens Actuators B Chem 217:146–150*
93. *Zhou Q, Sussman A, Chang J, Dong J, Zettl A, Mickelson W (2015) Fast response integrated MEMS microheaters for ultra low power gas detection. Sens Actuators A Phys 223:67–75*
94. *Erol-Kantarci M, Mouftah HT (2012) Mission-aware placement of RF-based power transmitters in wireless sensor networks. In: 2012 IEEE symposium on computers and communications (ISCC), pp 000012–000017*
95. *He S, Chen J, Jiang F, Yau DKY, Xing G, Sun Y (2013) Energy provisioning in wireless rechargeable sensor networks. IEEE Trans Mob Comput 12:1931–1942*
96. *Li Y, Fu L, Chen M, Chi K, Zhu YH (2015) RF-based charger placement for duty cycle guarantee in battery-free sensor networks. IEEE Commun Lett 19:1802–1805*
97. *Shao S, Gudan K, Hull JJ (2016) A mechanically beam-steered phased array antenna for power-harvesting applications [Antenna Applications Corner]. IEEE Antennas Propag Mag 58:58–64*
98. *Gudan K, Shao S, Hull JJ, Ensworth J, Reynolds MS (2015) Ultra-low power 2.4 GHz RF energy harvesting and storage system with −25 dBm sensitivity. In: 2015 IEEE international conference on RFID (RFID), pp 40–46*

Copyright

© The Author(s) 2017

Download PDF

Export citations

Citations & References

- Papers, Zotero, Reference Manager, RefWorks (.RIS)
- EndNote (.ENW)
- Mendeley, JabRef (.BIB)

Article citation

- Papers, Zotero, Reference Manager, RefWorks (.RIS)
- EndNote (.ENW)
- Mendeley, JabRef (.BIB)

References

- Papers, Zotero, Reference Manager, RefWorks (.RIS)
- EndNote (.ENW)
- Mendeley, JabRef (.BIB)

Table of Contents

- Abstract
- Background
- Fundamentals of RF transmission
- Wireless power harvesting evaluation metrics
- Recent advances in wireless power harvesting
- Impedance matching network
- Rectifier/voltage multiplier
- Pros and cons of Schottky diode and CMOS technology
- Designing and optimizing the WPH system

- Applications
- Wireless power harvesting network (IoT/WSN)
- Conclusion
- Declarations
- References
- Comments

Metrics

- Article accesses: 6116
- Citations: 1 more information

Share this article

- Share on Twitter
- Share on Facebook
- Share on LinkedIn
- Share on Weibo
- Share on Google Plus
- Share on Reddit

See updates

Check for updates

Other Actions

- Order reprint

Follow

- Follow us on Twitter
- Follow us on Facebook

Publisher secondary menu

- Contact us
- Jobs
- Language editing for authors
- Scientific editing for authors
- Leave feedback

- Terms and conditions
- Privacy statement
- Accessibility
- Cookies

Follow Springer Open

- Twitter
- Facebook
- Google Plus

By continuing to use this website, you agree to our Terms and Conditions, Privacy statement and Cookies policy.

SPRINGER NATURE

© 2017 BioMed Central Ltd unless otherwise stated. Part of Springer Nature.

We use cookies to improve your experience with our site. More information about our cookie policy

Close

US008830710B2

(12) **United States Patent**
Perreault et al.

(10) **Patent No.: US 8,830,710 B2**
(45) **Date of Patent: Sep. 9, 2014**

(54) **RF ENERGY RECOVERY SYSTEM**

(71) Applicant: **Eta Devices, Inc.**, Cambridge, MA (US)

(72) Inventors: **David J. Perreault**, Brookline, MA (US); **Joel L. Dawson**, Roslindale, MA (US)

(73) Assignee: **Eta Devices, Inc.**, Cambridge, MA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **13/755,335**

(22) Filed: **Jan. 31, 2013**

(65) **Prior Publication Data**

US 2013/0343106 A1 Dec. 26, 2013

Related U.S. Application Data

(60) Provisional application No. 61/663,930, filed on Jun. 25, 2012.

(51) **Int. Cl.**

H02M 7/10	(2006.01)
H03F 3/00	(2006.01)
H02M 7/06	(2006.01)
H02M 7/217	(2006.01)
H03H 7/38	(2006.01)

(52) **U.S. Cl.**

CPC . ***H03H 7/38*** (2013.01); ***H03F 3/00*** (2013.01); ***H02M 7/06*** (2013.01); ***H02M 7/217*** (2013.01)
USPC **363/67**; 363/69

(58) **Field of Classification Search**

CPC H02M 7/217; H03H 7/38
USPC 363/65, 67, 69
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,319,359	A *	3/1982	Wolf	375/238
5,568,088	A	10/1996	Dent et al.	
5,638,024	A	6/1997	Dent et al.	
5,708,573	A	1/1998	Lusher et al.	

(Continued)

FOREIGN PATENT DOCUMENTS

GB	1506633	4/1978
WO	WO 0067373	11/2000

(Continued)

OTHER PUBLICATIONS

Beltran, et al.; "An Outphasing Transmitter Using Class-E PAs and Asymmetric Combining: Part 1;" High Frequency Electronics: Summit Technical Media, LLC; Apr. 2011; pp. 18-26.

(Continued)

Primary Examiner — Matthew Nguyen
(74) *Attorney, Agent, or Firm* — Daly, Crowley, Mofford & Durkee, LLP

(57) **ABSTRACT**

An energy recovery system is provided that can be used in place of a resistive termination. The energy recovery system can simulate the behavior and performance of a resistive termination, while recovering at least some of the energy that would otherwise have been dissipated within the termination. In one embodiment, the energy recovery system includes a rectification system having a plurality of tuned rectifier circuits and an RF input network to provide input power distribution to the plurality of tuned rectifier circuits and input impedance shaping to shape the collective input impedance of the plurality of tuned rectifier circuits as seen at the RF input port of the RF input network.

22 Claims, 16 Drawing Sheets

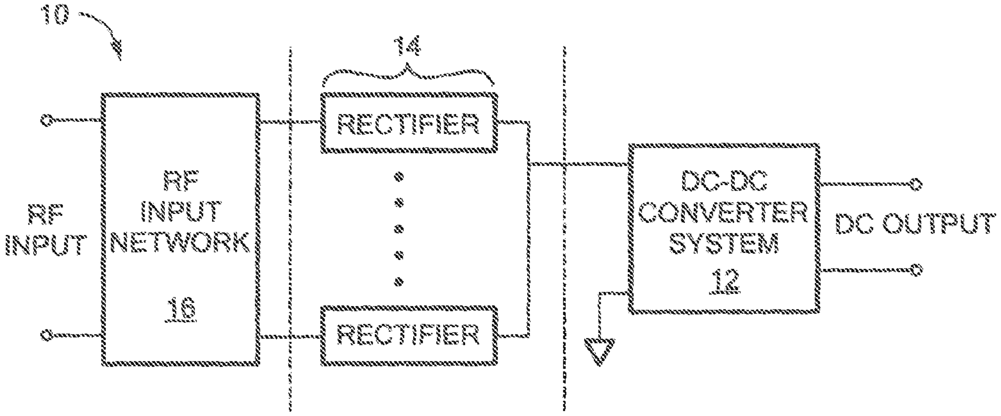

(56) References Cited

U.S. PATENT DOCUMENTS

5,736,877	A	4/1998	Tihanyi	
5,892,395	A	4/1999	Stengel et al.	
6,600,384	B2	7/2003	Mohwinkel et al.	
6,887,339	B1	5/2005	Goodman et al.	
7,279,971	B2	10/2007	Hellberg et al.	
7,508,262	B1 *	3/2009	Wagner et al.	330/136
7,535,133	B2	5/2009	Perreault et al.	
7,589,605	B2	9/2009	Perreault et al.	
7,889,519	B2	2/2011	Perreault et al.	
7,956,572	B2	6/2011	Zane et al.	
8,026,763	B2	9/2011	Dawson et al.	
8,164,384	B2	4/2012	Dawson et al.	
8,212,541	B2	7/2012	Perreault et al.	
8,451,053	B2	5/2013	Perreault et al.	
8,638,156	B2	1/2014	Shibata et al.	
2004/0027209	A1	2/2004	Chen et al.	
2006/0193095	A1	8/2006	Hunter et al.	
2009/0200985	A1	8/2009	Zane et al.	
2010/0120384	A1	5/2010	Pennec	
2011/0187437	A1	8/2011	Perreault et al.	
2011/0215866	A1	9/2011	Dawson et al.	
2012/0176195	A1	7/2012	Dawson et al.	
2012/0313602	A1	12/2012	Perreault et al.	
2012/0326684	A1	12/2012	Perreault et al.	
2013/0241625	A1	9/2013	Perreault et al.	
2013/0343107	A1	12/2013	Perreault	
2014/0118063	A1	5/2014	Briffa et al.	
2014/0118065	A1	5/2014	Briffa et al.	
2014/0118072	A1	5/2014	Briffa et al.	
2014/0120854	A1	5/2014	Briffa et al.	
2014/0125412	A1	5/2014	Dawson et al.	
2014/0132354	A1	5/2014	Briffa et al.	

FOREIGN PATENT DOCUMENTS

WO	WO 01/10013	A1	2/2001
WO	WO 2005/106613	A1	11/2005
WO	WO 2006/119362	A2	11/2006
WO	WO 2007/082090	A2	7/2007
WO	WO 2007/094921	A1	8/2007
WO	WO 2010/056646	A1	5/2010
WO	WO 2011/097387	A1	8/2011
WO	WO 2013/109719	A1	7/2013
WO	WO 2013/109743	A2	7/2013
WO	WO 2013/109797	A1	7/2013
WO	WO 2013/134573	A1	9/2013
WO	WO 2013/191757	A1	12/2013
WO	WO 2013/048747	A2	1/2014
WO	WO 20141/028441	A2	2/2014
WO	WO 2014/070474	A1	5/2014
WO	WO 2014/070475	A1	5/2014
WO	WO 2014/070998	A1	5/2014
WO	WO 2014/085097	A1	6/2014

OTHER PUBLICATIONS

Beltran, et al.; "An Outphasing Transmitter Using Class-E PAs and Asymmetric Combining: Part 2;" High Frequency Electronics; Summit Technical Media, LLC; May 2011; pp. 34-46.

Chang, et al.; "A Power-Recycling Technique for Improving Power Amplifier Efficiency Under Load Mismatch;" IEEE Microwave and Wireless Components Letters; vol. 21; No. 10; Oct. 2011; pp. 571-573.

Eun, et al.; "A High Linearity Chireix Outphasing Power Amplifier Using Composite Right/Left-Handed Transmission Lines;" Proceedings of the 4th European Radar Conference; Oct. 2007; pp. 343-346.

Gerhard, et al.; "Novel Transmission Line Combiner for High Efficient Outphasing RF Power Amplifiers;" Proceedings of the 37th European Microwave Conference; Oct. 2007; pp. 1433-1436.

Gody, et al.; "Outphasing Energy Recovery Amplifier with Resistance Compression for Improved Efficiency;" IEEE Transactions on Microwave Theory and Techniques; vol. 57; No. 12; Dec. 2009; pp. 2895-2906.

Gutmann, et al.; "Power Combining in an Array of Microwave Power Rectifiers;" IEEE Microwave Symposium Digest; Apr. 30, 1979-May 2, 1979; pp. 453-455.

Han, et al.; "Resistance Compression Networks for Radio-Frequency Power Conversion;" IEEE Transactions on Power Electronics; vol. 22; No. 1; Jan. 2007; pp. 41-53.

Langridge, et al.; "A Power Re-Use Technique for Improved Efficiency of Outphasing Microwave Power Amplifiers;" IEEE Transactions on Microwave Theory and Techniques; vol. 47; No. 8; Aug. 1999; pp. 1467-1470.

Nitz, et al.; "A New Family of Resonant Rectifier Circits for High Frequency DC-DC Converter Applications;" Applied Power Electronics Conference and Exposition; 1988; APEC 88; Conference Proceedings 1988; Feb. 1-5, 1988; pp. 12-22.

Paing, et al., "Resistor Emulation Approach to Low-Power RF Energy Harvesting;" IEEE Transactions on Power Electronics; vol. 23; No. 3; May 2008; pp. 1494-1501.

Perreault; "A New Power Combining and Outphasing Modulation System for High-Efficiency Power Amplification," IEEE Transactions on Circuits and Systems-I: Regular Papers; vol. 58; No. 8; Aug. 2011; pp. 1713-1726.

Rivas, et al.; "New Architecture for Radio Frequency dc/dc/ Power Conversion;" 2004 35th Annual EIII Power Electronics Specialist Conference; Aachen, German; vol. 15; Jun. 2004; pp. 4074-4084.

International Search Report and Written Opinion of the ISA for PCT/US2006/016981 dated Dec. 2006.

U.S. Pat. No. 7,535,133 issued on May 19, 2009.

PCT Search Report of the ISA for PCT/US2013/048747 dated Jan. 31, 2014.

PCT Written Opinion of the ISA for PCT/US2013/046747 dated Jan. 31, 2014.

Office Action dated Mar. 28, 2014 from U.S. Appl. No. 13/800,221, filed Mar. 13, 2013.

Response to Office Action Mar. 28, 2014 Office Action as filed on Apr. 10, 2014 from U.S. Appl. No. 13/800,221, filed Mar. 13, 2013.

Notice of Allowance for U.S. Appl. No. 13/800,221, filed Mar. 13, 2013.

* cited by examiner

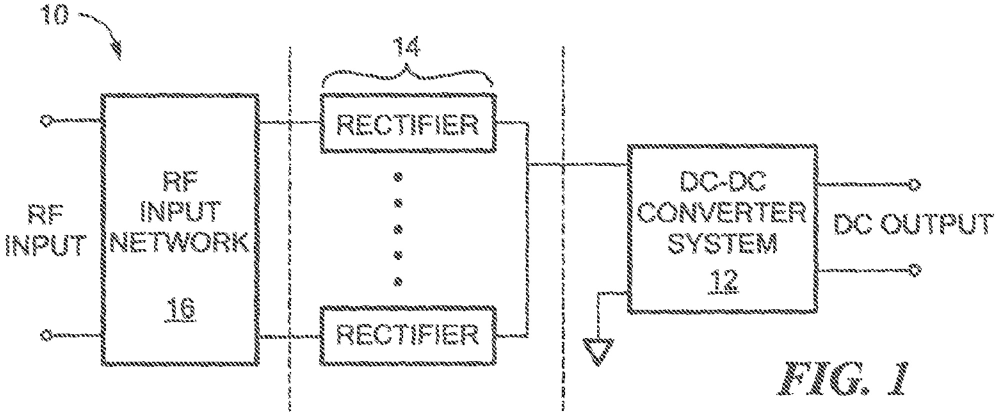
10
14
RF INPUT
RF INPUT NETWORK
16
RECTIFIER
RECTIFIER
DC-DC CONVERTER SYSTEM
12
DC OUTPUT

FIG. 1

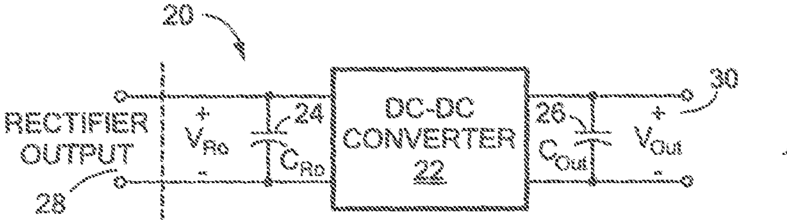
20
RECTIFIER OUTPUT
28
+
V_{Ro}
-
24
C_{Ro}
DC-DC CONVERTER
22
26
C_{Out}
+
V_{Out}
-
30

FIG. 2A

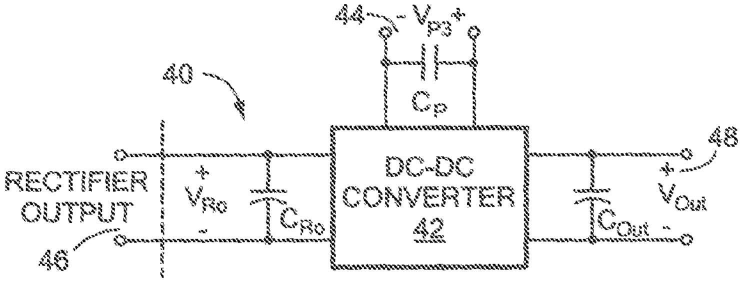
44
- V_{P3} +
C_p
40
RECTIFIER OUTPUT
46
+
V_{Ro}
-
C_{Ro}
DC-DC CONVERTER
42
C_{Out}
+
V_{Out}
-
48

FIG. 2B

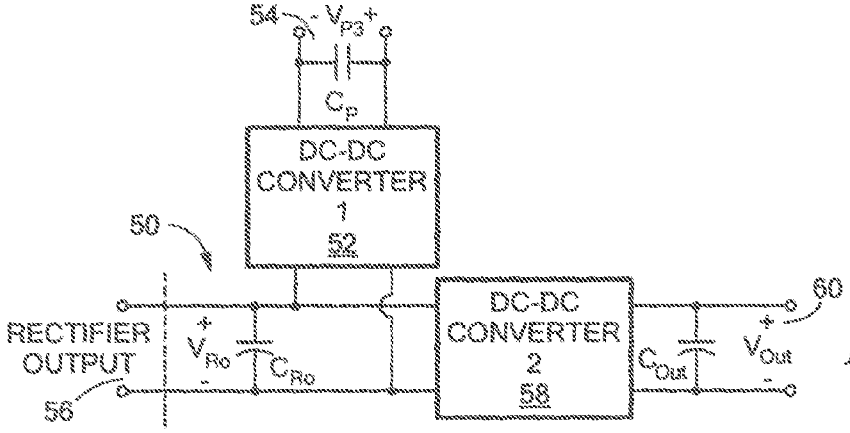
54
- V_{P3} +
C_p
DC-DC CONVERTER 1
52
50
RECTIFIER OUTPUT
56
+
V_{Ro}
-
C_{Ro}
DC-DC CONVERTER 2
58
C_{Out}
+
V_{Out}
-
60

FIG. 2C

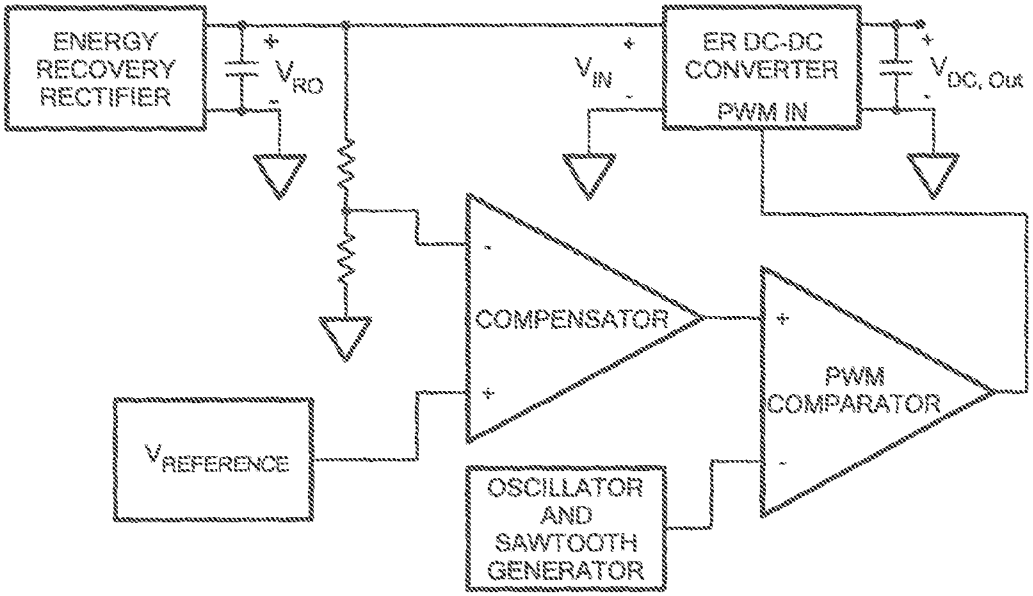
ENERGY RECOVERY RECTIFIER
+
V_{RO}
-
V_{IN}
+
-
ER DC-DC CONVERTER
PWM IN
+
$V_{DC, Out}$
-
-
COMPENSATOR
+
$V_{REFERENCE}$
+
PWM COMPARATOR
-
OSCILLATOR AND SAWTOOTH GENERATOR

FIG. 2D

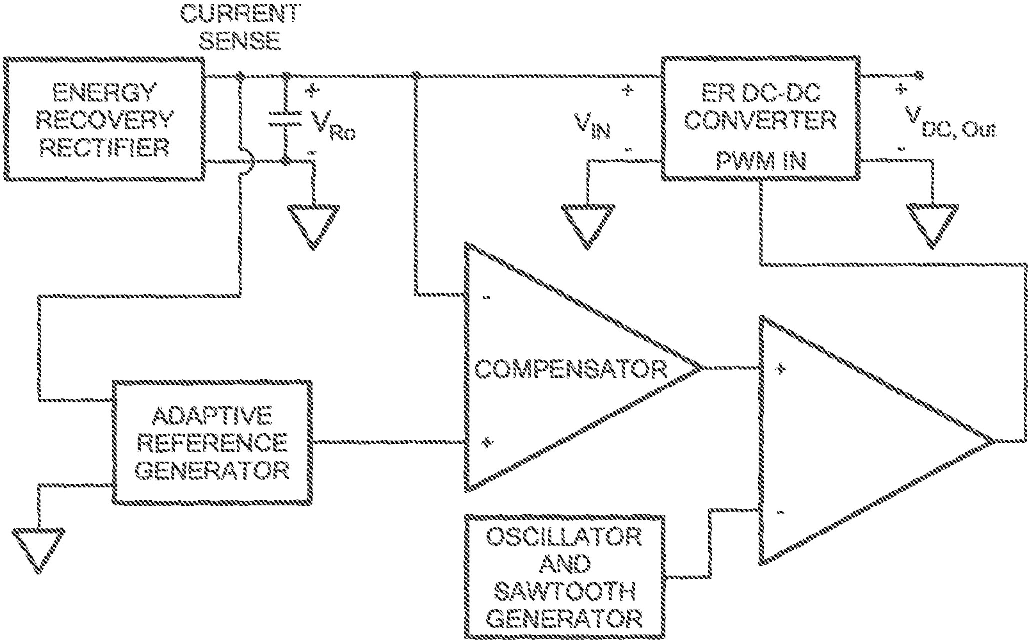
CURRENT SENSE
ENERGY RECOVERY RECTIFIER
+
V_{Ro}
-
V_{IN}
+
-
ER DC-DC CONVERTER
PWM IN
+
$V_{DC, Out}$
-
-
COMPENSATOR
+
ADAPTIVE REFERENCE GENERATOR
+
-
OSCILLATOR AND SAWTOOTH GENERATOR

FIG. 2E

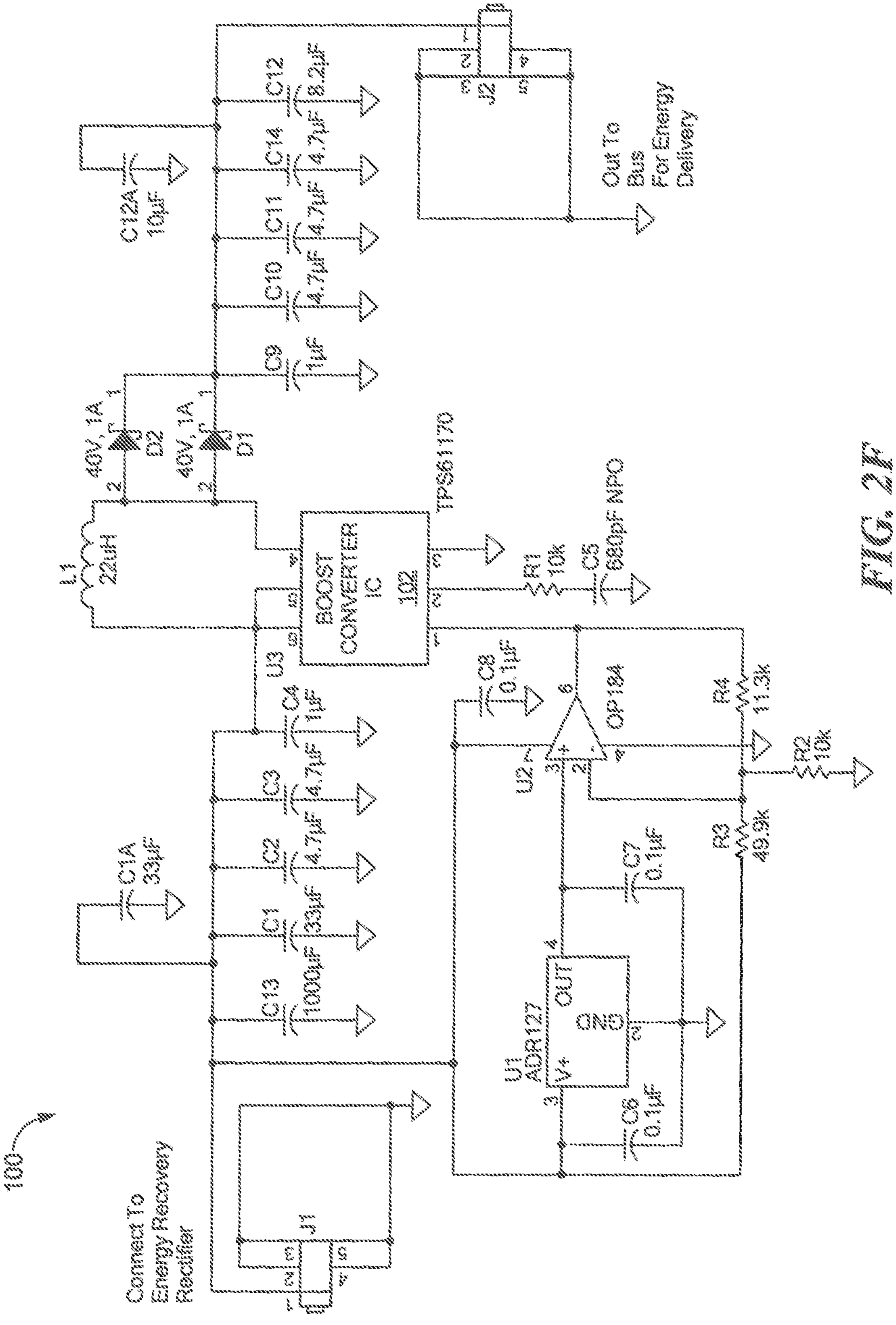
100
Connect To Energy Recovery Rectifier
J1
1
2
3
4
C13
1000μF
C1
33μF
C1A
33μF
C2
4.7μF
C3
4.7μF
C4
1μF
U3
L1
22uH
BOOST CONVERTER IC
102
TPS61170
R1
10k
C5
680pF NPO
D1
40V, 1A
D2
40V, 1A
C9
1μF
C10
4.7μF
C11
4.7μF
C14
4.7μF
C12
8.2μF
C12A
10μF
J2
Out To Bus For Energy Delivery
U1
ADR127
V+
OUT
GND
C6
0.1μF
C7
0.1μF
U2
OP184
C8
0.1μF
R4
11.3k
R2
10k
R3
49.9k

FIG. 2F

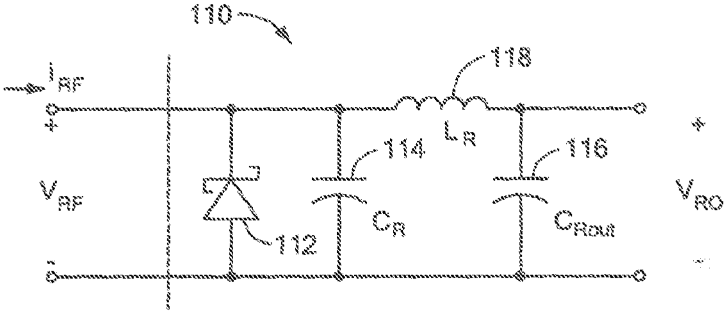
110
i_{RF}
118
+
L_R
114
116
+
V_{RF}
V_{RO}
C_R
C_{Rout}
112
-
-

FIG. 3A

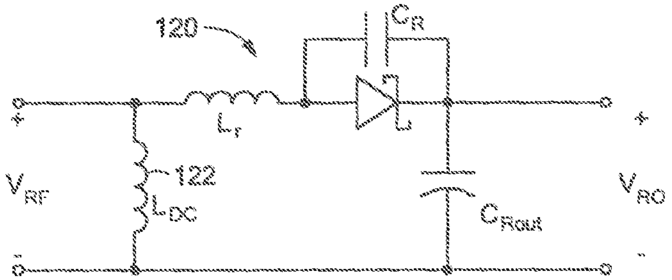
120
C_R
+
L_r
+
122
V_{RF}
L_{DC}
V_{RO}
C_{Rout}
-
-

FIG. 3B

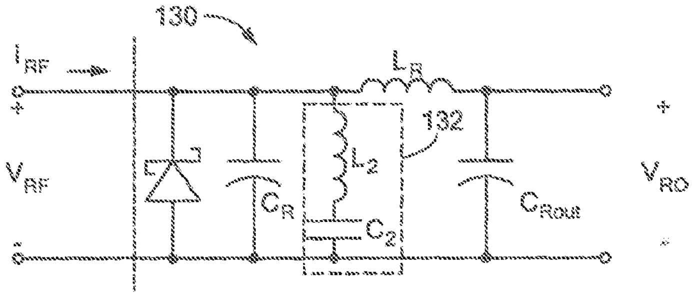
130
i_{RF}
L_R
+
132
+
V_{RF}
L_2
V_{RO}
C_R
C_{Rout}
C_2
-
-

FIG. 3C

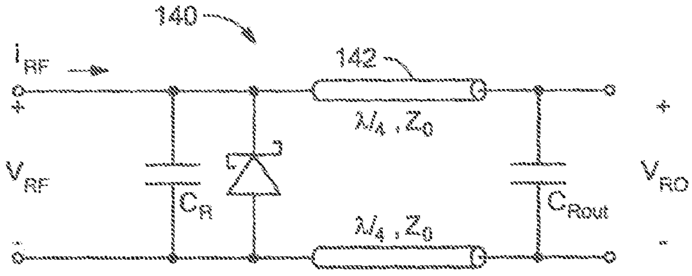
140
i_{RF}
142
+
$\lambda/_4, Z_0$
+
V_{RF}
V_{RO}
C_R
C_{Rout}
$\lambda/_4, Z_0$
-
-

FIG. 3D

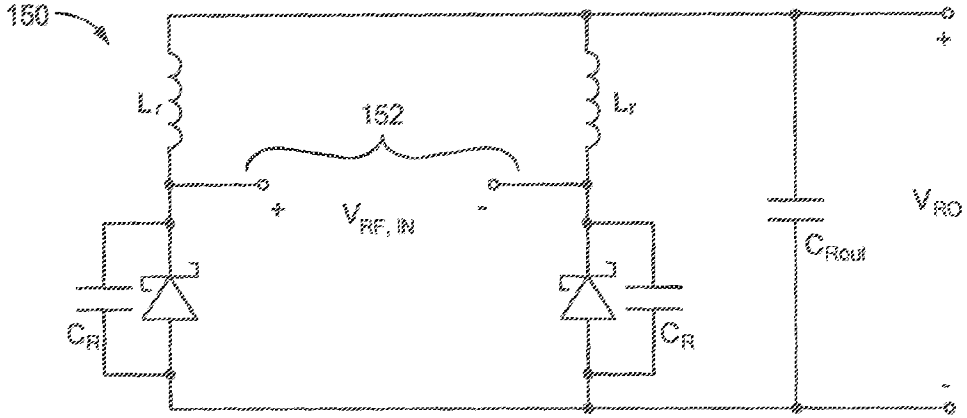
150
L_r
152
L_r
+
$V_{RF,IN}$
-
+
C_R
C_R
C_{Rout}
V_{RO}
-

FIG. 3E

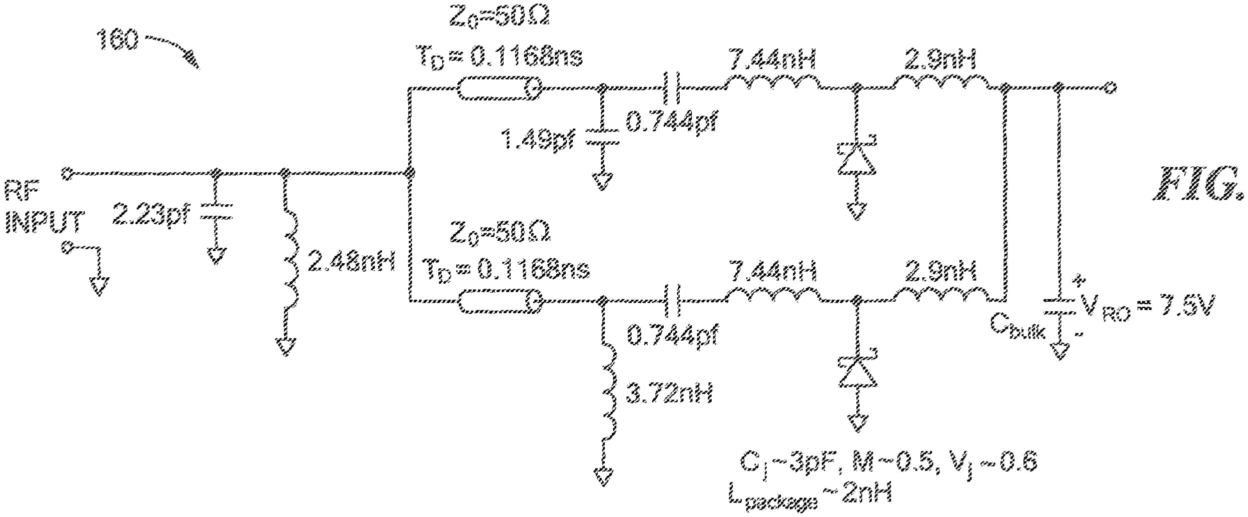
160
Z_0=50Ω
T_D = 0.1168ns
7.44nH
2.9nH
1.49pf
0.744pf
RF
INPUT
2.23pf
2.48nH
Z_0=50Ω
T_D = 0.1168ns
7.44nH
2.9nH
0.744pf
C_{bulk}
+
V_{RO} = 7.5V
-
3.72nH
C_j~3pF, M~0.5, V_j~0.6
$L_{package}$~2nH

FIG. 3F

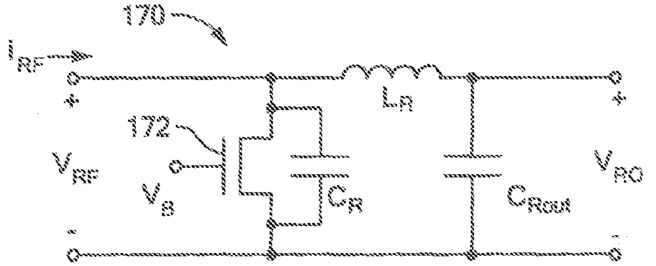
170
i_{RF}
+
V_{RF}
-
172
V_B
L_R
C_R
C_{Rout}
+
V_{RO}
-

FIG. 4A

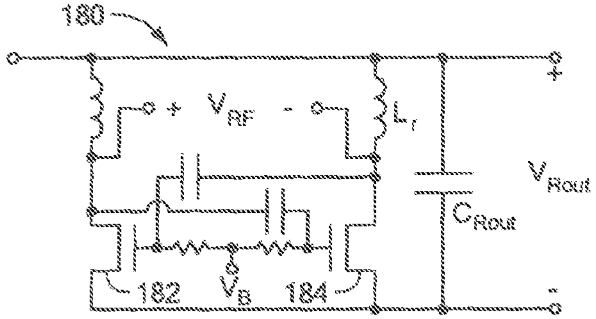
180
+ V_{RF} -
L_r
+
V_{Rout}
C_{Rout}
-
182
V_B
184

FIG. 4B

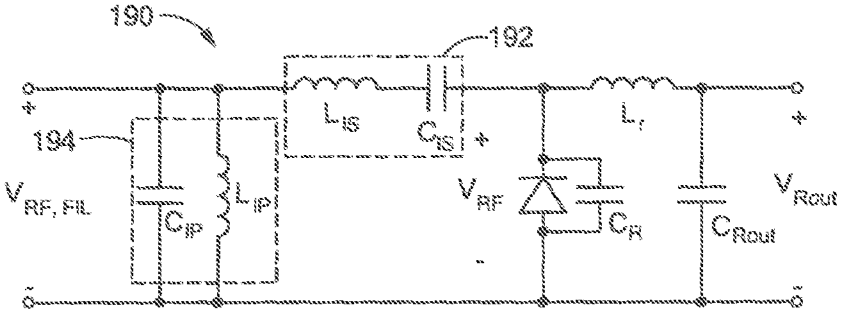
190
192
194
+
$V_{RF, FIL}$
-
C_{IP}
L_{IP}
L_{IS}
C_{IS}
+
V_{RF}
-
C_R
L_r
C_{Rout}
+
V_{Rout}
-

FIG. 5

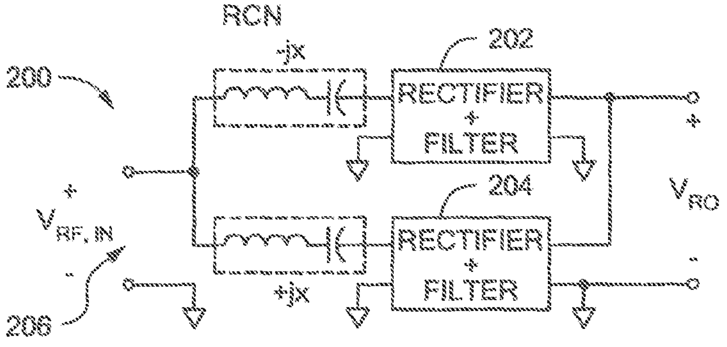
RCN
-jx
200
202
RECTIFIER + FILTER
+
V_{RO}
-
+
$V_{RF, IN}$
-
204
RECTIFIER + FILTER
+jx
206

FIG. 6A

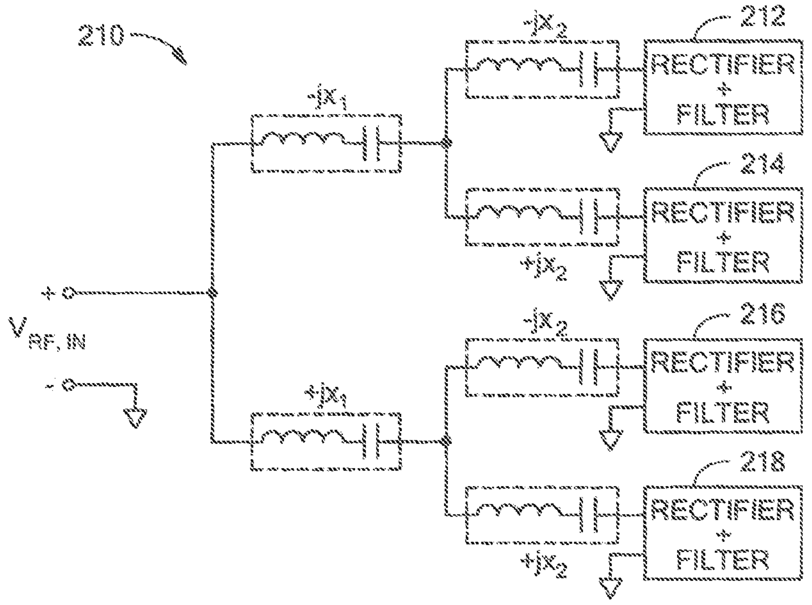
210
212
$-jX_2$
RECTIFIER
+
FILTER
$-jX_1$
214
RECTIFIER
+
FILTER
$+jX_2$
+
$V_{RF,IN}$
-
216
$-jX_2$
RECTIFIER
+
FILTER
$+jX_1$
218
RECTIFIER
+
FILTER
$+jX_2$

FIG.6B

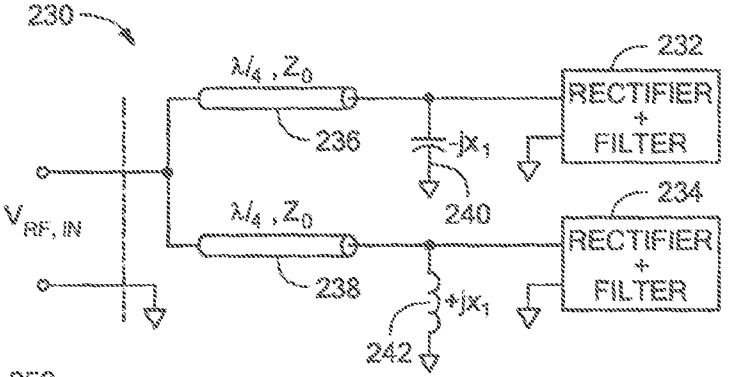
230
$\lambda/_4$, Z_0
232
RECTIFIER
+
FILTER
236
$-jX_1$
240
$V_{RF,IN}$
$\lambda/_4$, Z_0
234
RECTIFIER
+
FILTER
238
$+jX_1$
242

FIG.6C

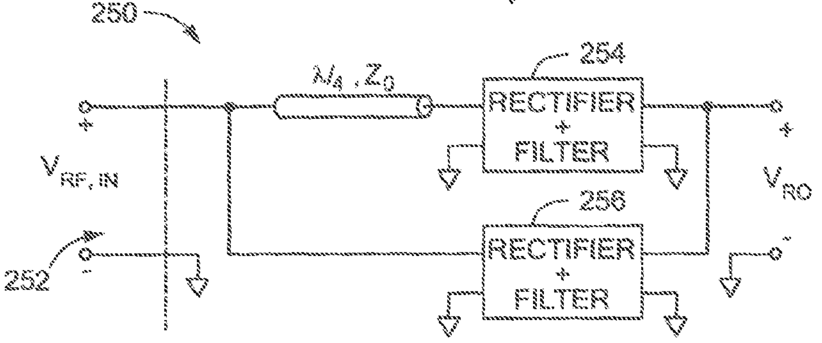
250
254
$\lambda/_4$, Z_0
RECTIFIER
+
FILTER
+
$V_{RF,IN}$
+
V_{RO}
256
RECTIFIER
+
FILTER
252
-
-

FIG.6D

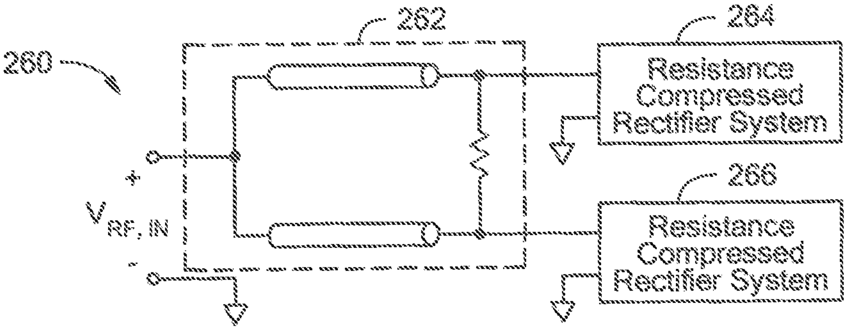
260
262
264
Resistance Compressed Rectifier System
266
Resistance Compressed Rectifier System
+
$V_{RF,\ IN}$
-

FIG. 6E

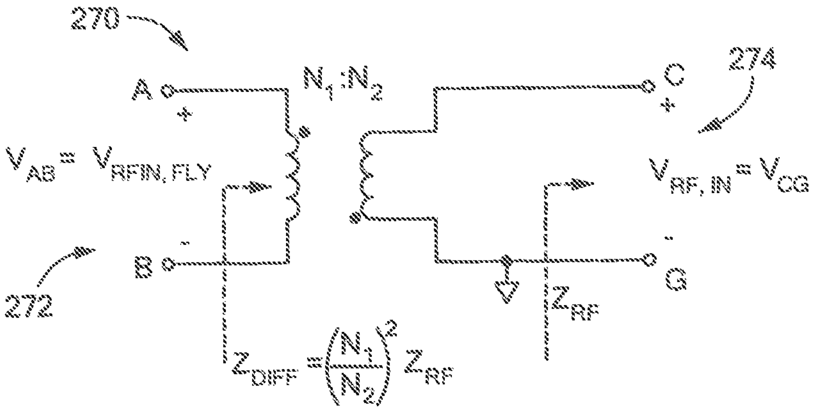
270
A
+
$N_1:N_2$
C
+
274
$V_{AB} = V_{RFIN,\ FLY}$
$V_{RF,\ IN} = V_{CG}$
-
B
272
-
G
Z_{RF}
$Z_{DIFF} = \left(\frac{N_1}{N_2}\right)^2 Z_{RF}$

FIG. 7A

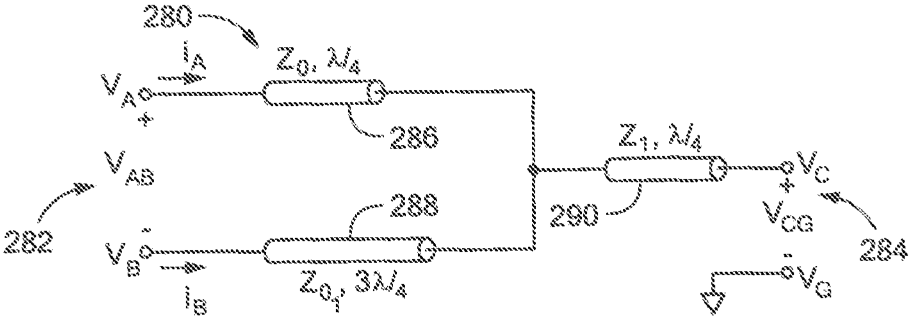
280
i_A
$Z_0,\ \lambda/_4$
V_A
+
286
V_{AB}
$Z_1,\ \lambda/_4$
V_C
290
+
V_{CG}
284
288
282
-
V_B
i_B
$Z_{0_1},\ 3\lambda/_4$
-
V_G

FIG. 7B

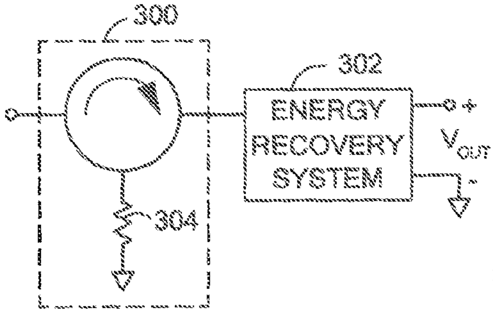
300
302
ENERGY RECOVERY SYSTEM
+
V_{OUT}
-
304

FIG.8A

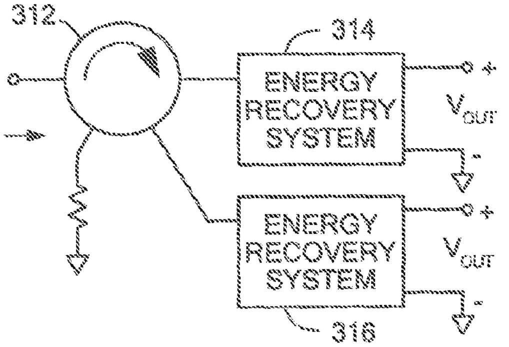
312
314
ENERGY RECOVERY SYSTEM
+
V_{OUT}
-
310
ENERGY RECOVERY SYSTEM
+
V_{OUT}
-
316

FIG.8B

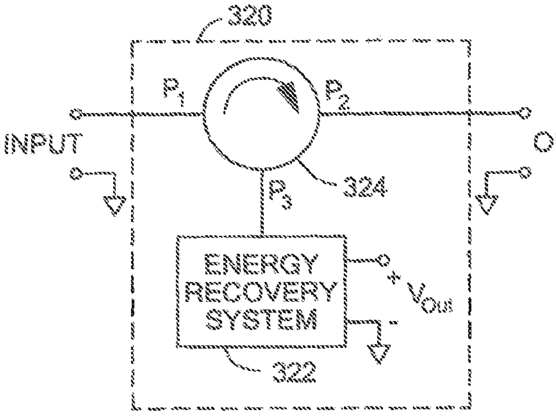
320
P_1
P_2
INPUT
OUTPUT
P_3
324
ENERGY RECOVERY SYSTEM
+
V_{Out}
-
322

FIG.9A

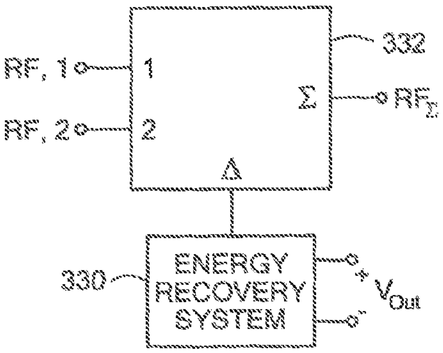
332
RF, 1
1
Σ
RF_{Σ}
RF, 2
2
Δ
330
ENERGY RECOVERY SYSTEM
+
V_{Out}
-

FIG.9B

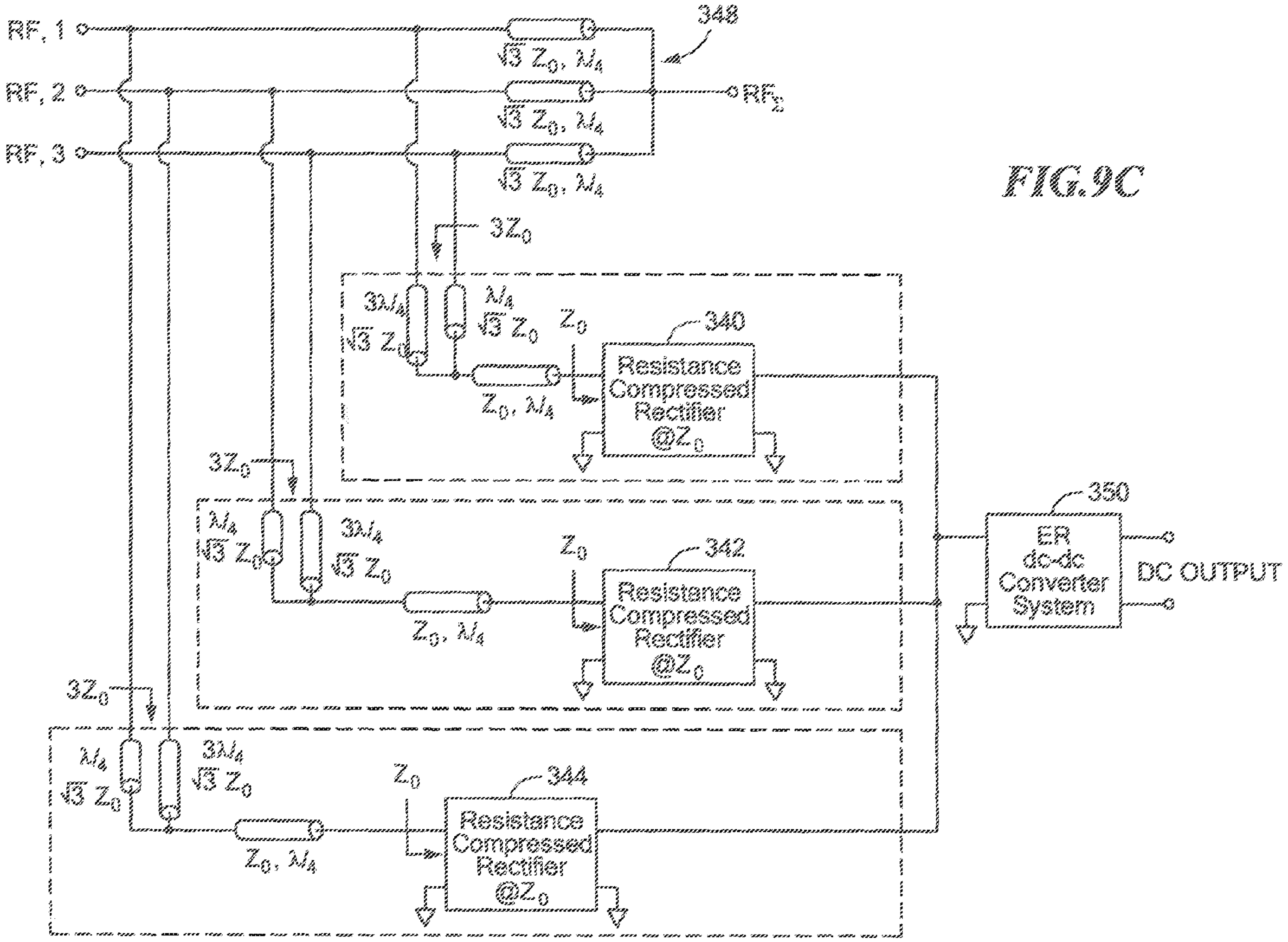
RF, 1
RF, 2
RF, 3
$\sqrt{3}\ Z_0,\ \lambda/_4$
$\sqrt{3}\ Z_0,\ \lambda/_4$
$\sqrt{3}\ Z_0,\ \lambda/_4$
348
RF_Σ
$3Z_0$
$3\lambda/_4$ $\sqrt{3}\ Z_0$
$\lambda/_4$ $\sqrt{3}\ Z_0$
Z_0
340
Resistance Compressed Rectifier @Z_0
$Z_0,\ \lambda/_4$
$3Z_0$
$\lambda/_4$ $\sqrt{3}\ Z_0$
$3\lambda/_4$ $\sqrt{3}\ Z_0$
Z_0
342
Resistance Compressed Rectifier @Z_0
$Z_0,\ \lambda/_4$
350
ER dc-dc Converter System
DC OUTPUT
$3Z_0$
$\lambda/_4$ $\sqrt{3}\ Z_0$
$3\lambda/_4$ $\sqrt{3}\ Z_0$
Z_0
344
Resistance Compressed Rectifier @Z_0
$Z_0,\ \lambda/_4$

FIG. 9C

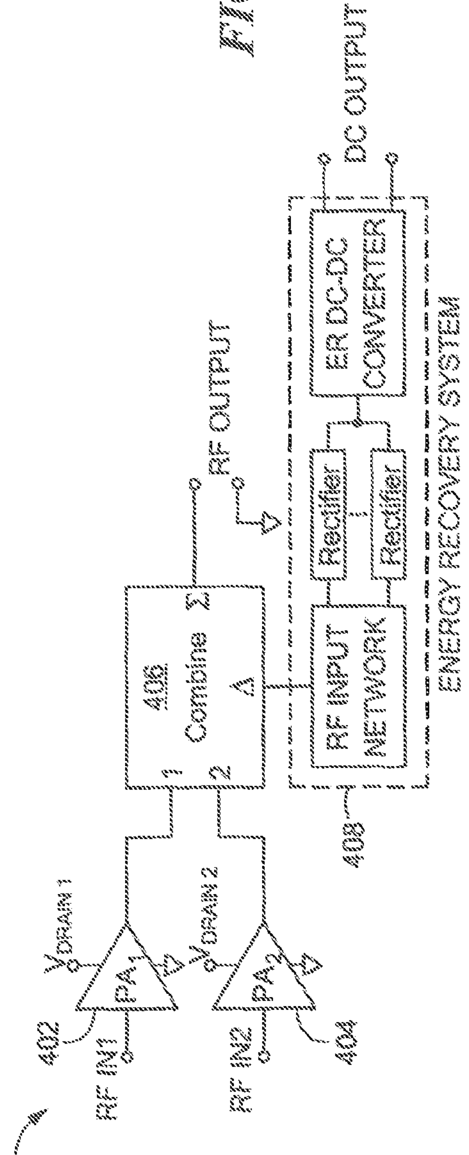
400
RF IN1
402
$V_{DRAIN\ 1}$
PA_1
$V_{DRAIN\ 2}$
PA_2
RF IN2
404
1
2
406
Combine
Δ
Σ
RF OUTPUT
RF INPUT NETWORK
Rectifier
Rectifier
ER DC-DC CONVERTER
DC OUTPUT
ENERGY RECOVERY SYSTEM
408

FIG. 9D

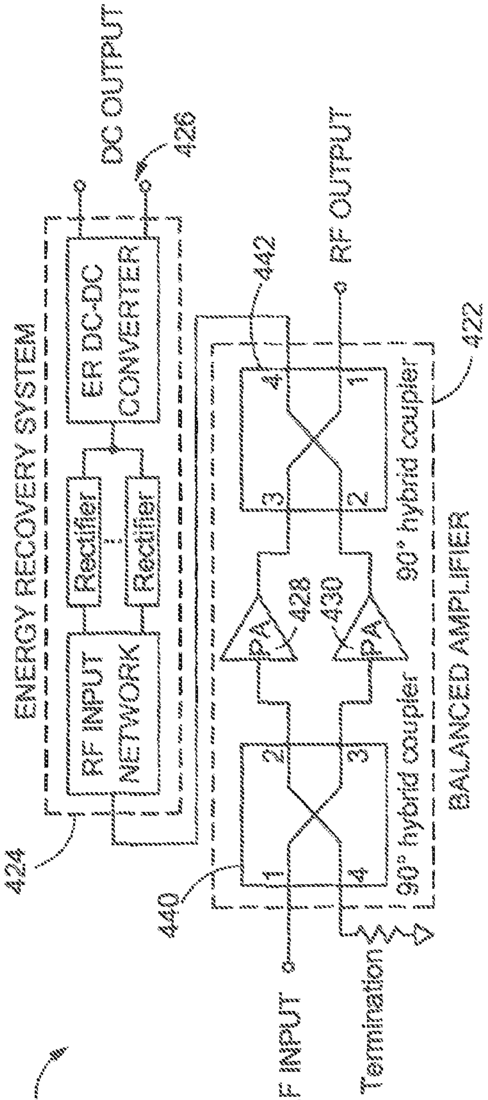
420
RF INPUT
Termination
440
1
4
2
3
90° hybrid coupler
PA
428
PA
430
3
2
4
1
90° hybrid coupler
442
RF OUTPUT
BALANCED AMPLIFIER
422
ENERGY RECOVERY SYSTEM
424
RF INPUT NETWORK
Rectifier
Rectifier
ER DC-DC CONVERTER
DC OUTPUT
426

FIG. 9E

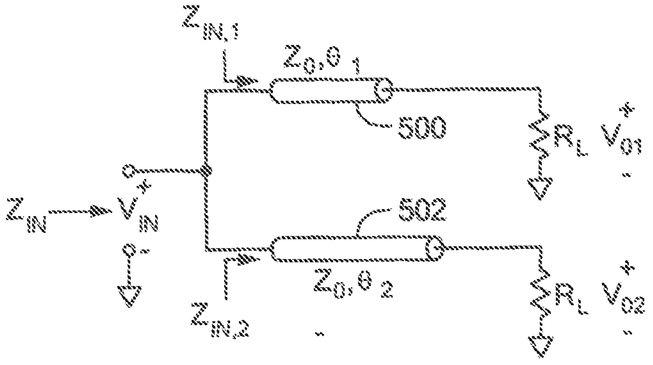
$Z_{IN,1}$
Z_0, θ_1
500
R_L
V_{01}
+
-
Z_{IN}
V_{IN}
+
-
502
Z_0, θ_2
$Z_{IN,2}$
R_L
V_{02}
+
-

FIG.10A

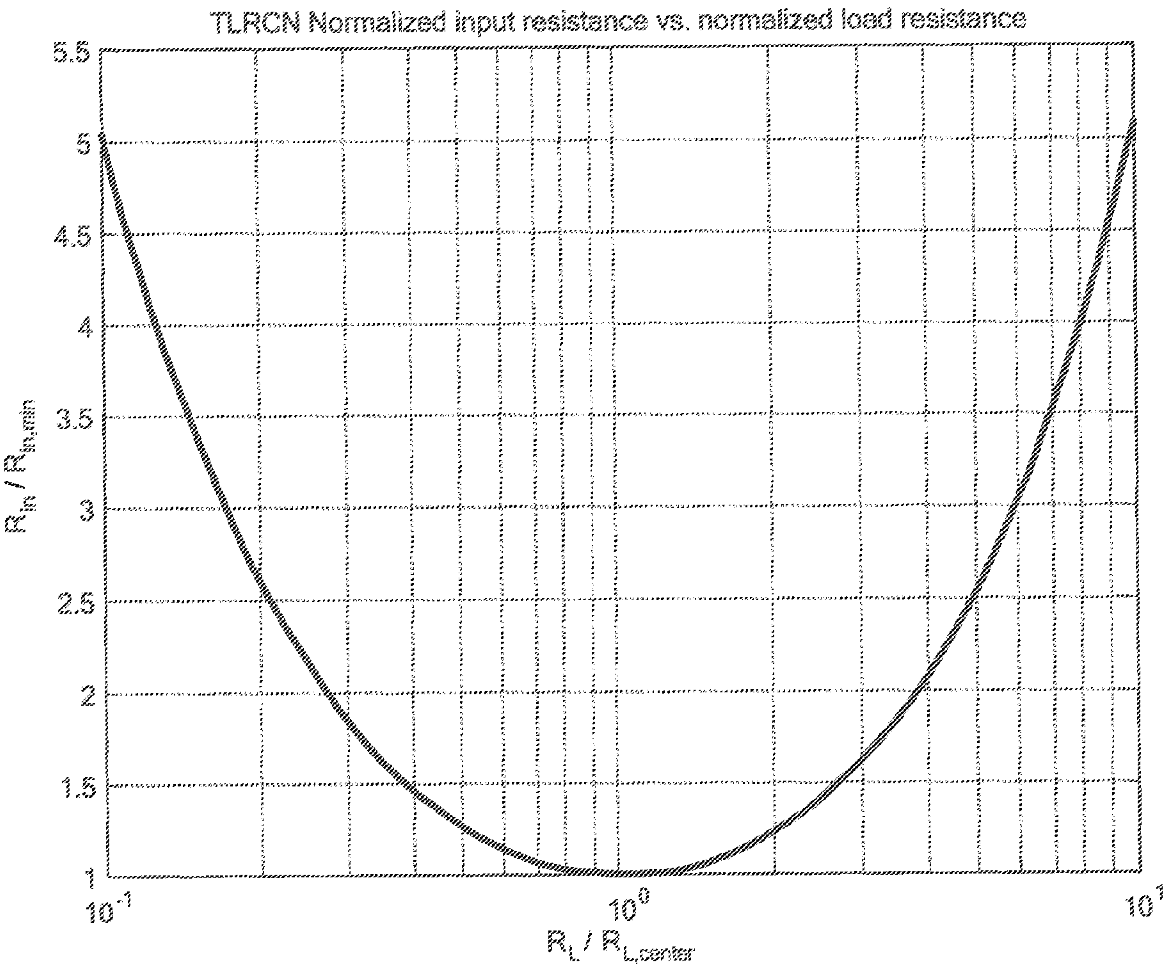
TLRCN Normalized input resistance vs. normalized load resistance
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
$R_{in} / R_{in,min}$
10^{-1}
10^{0}
10^{1}
$R_L / R_{L,center}$

FIG.10B

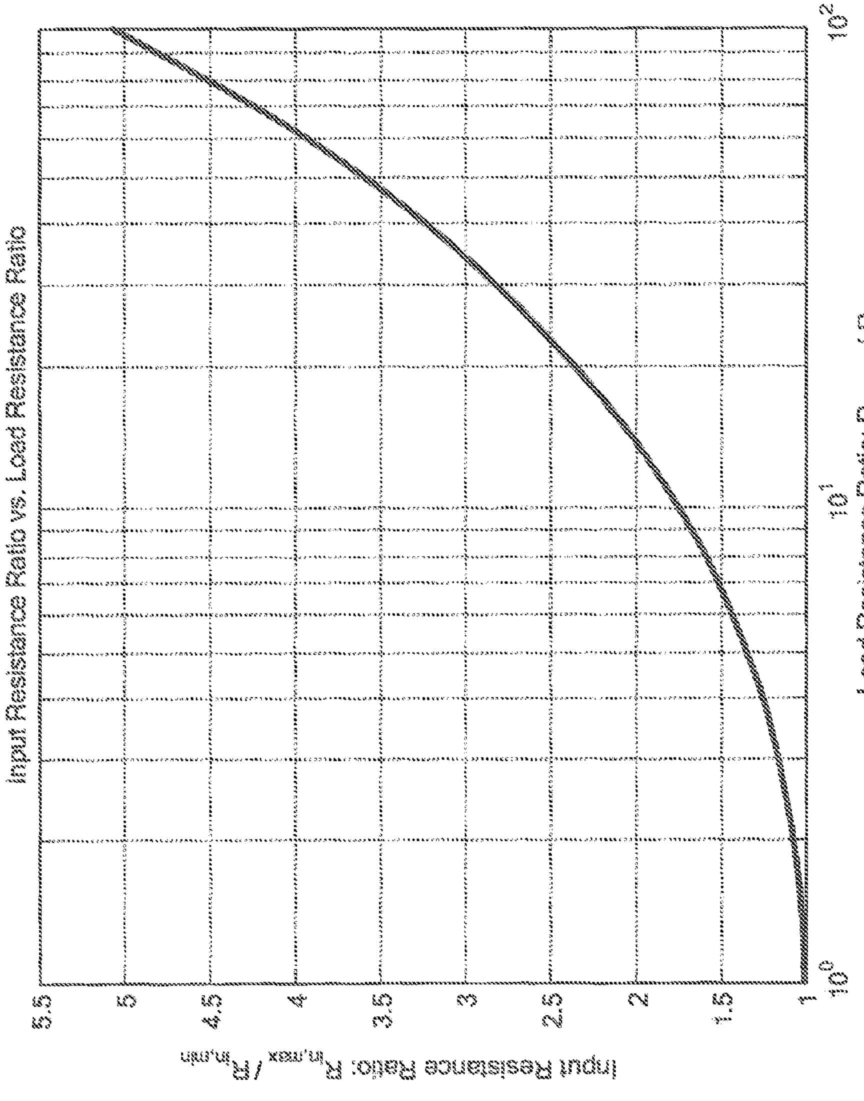
Input Resistance Ratio vs. Load Resistance Ratio
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
Input Resistance Ratio: $R_{in,max}$ / $R_{in,min}$
10^0
10^1
10^2
Load Resistance Ratio: $R_{L,max}$ / $R_{L,min}$

FIG.10C

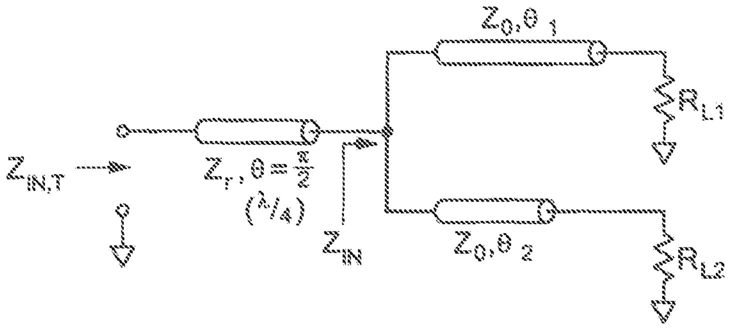
Z_0, θ_1
R_{L1}
$Z_{IN,T}$
$Z_T, \theta = \frac{\pi}{2}$
$(\lambda/4)$
Z_{IN}
Z_0, θ_2
R_{L2}

FIG. 10D

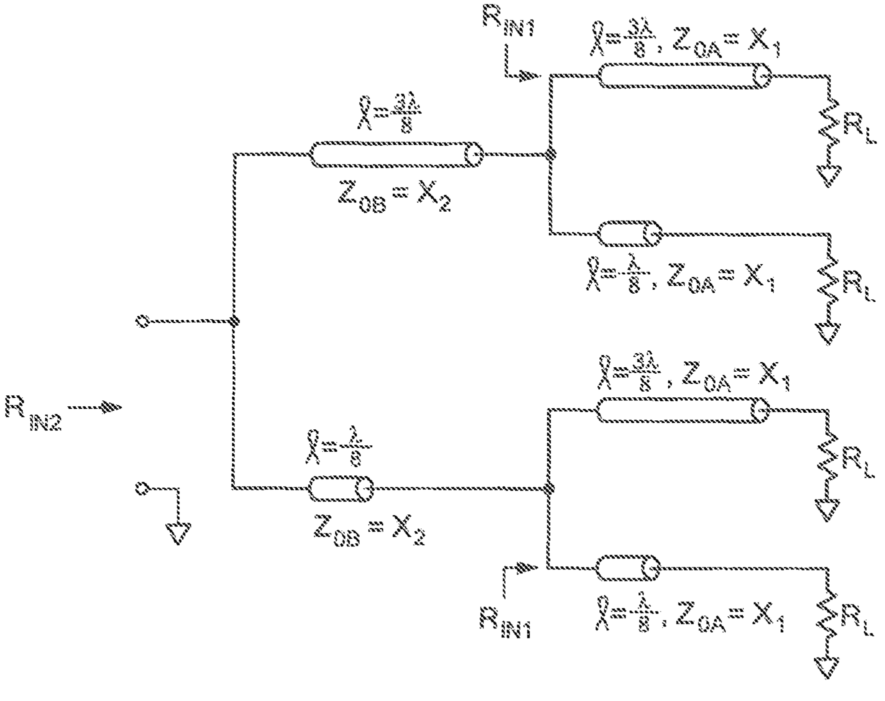
R_{IN1}
$\ell = \frac{3\lambda}{8}, Z_{0A} = X_1$
$\ell = \frac{3\lambda}{8}$
$Z_{0B} = X_2$
R_L
$\ell = \frac{\lambda}{8}, Z_{0A} = X_1$
R_L
R_{IN2}
$\ell = \frac{3\lambda}{8}, Z_{0A} = X_1$
R_L
$\ell = \frac{\lambda}{8}$
$Z_{0B} = X_2$
R_{IN1}
$\ell = \frac{\lambda}{8}, Z_{0A} = X_1$
R_L

FIG. 10E

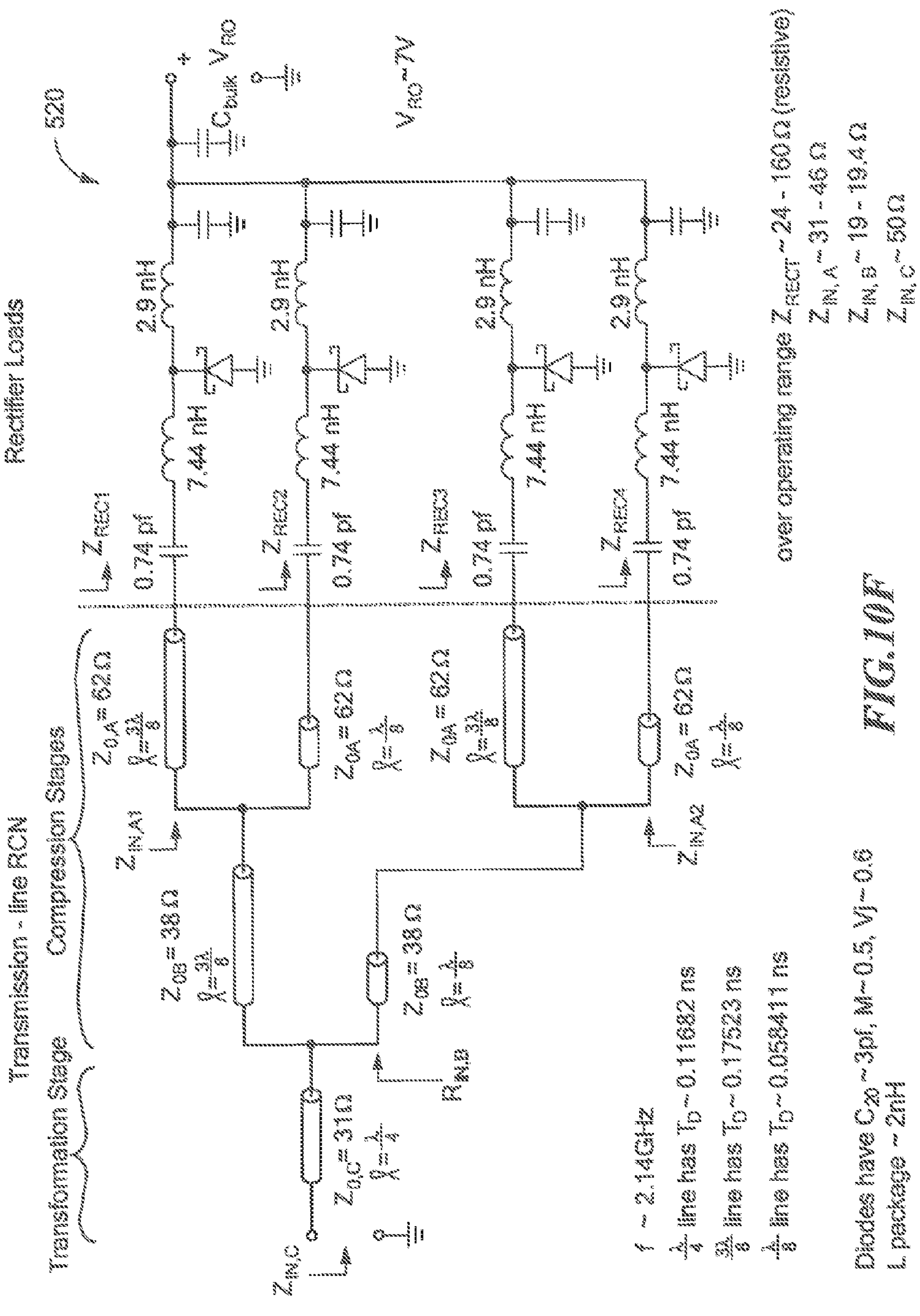
520
Rectifier Loads
Transmission - line RCN
Transformation Stage
Compression Stages
$Z_{IN,C}$
$Z_{0,C} = 31\Omega$
$\ell = \frac{\lambda}{4}$
$Z_{0B} = 38\Omega$
$\ell = \frac{3\lambda}{8}$
$Z_{0B} = 38\Omega$
$\ell = \frac{\lambda}{8}$
$R_{IN,B}$
$Z_{IN,A1}$
$Z_{IN,A2}$
$Z_{0A} = 62\Omega$
$\ell = \frac{3\lambda}{8}$
$Z_{0A} = 62\Omega$
$\ell = \frac{\lambda}{8}$
$Z_{0A} = 62\Omega$
$\ell = \frac{3\lambda}{8}$
$Z_{0A} = 62\Omega$
$\ell = \frac{\lambda}{8}$
Z_{REC1}
Z_{REC2}
Z_{REC3}
Z_{REC4}
0.74 pf
0.74 pf
0.74 pf
0.74 pf
7.44 nH
7.44 nH
7.44 nH
7.44 nH
2.9 nH
2.9 nH
2.9 nH
2.9 nH
C_{bulk}
+
V_{RO}
V_{RO}~7V
over operating range Z_{RECT} ~ 24 - 160 Ω (resistive)
$Z_{IN,A}$ ~ 31 - 46 Ω
$Z_{IN,B}$ ~ 19 - 19.4 Ω
$Z_{IN,C}$ ~ 50 Ω
f ~ 2.14GHz
$\frac{\lambda}{4}$ line has T_D ~ 0.11682 ns
$\frac{3\lambda}{8}$ line has T_D ~ 0.17523 ns
$\frac{\lambda}{8}$ line has T_D ~ 0.058411 ns
Diodes have C_{J0} ~ 3pf, M~0.5, Vj~0.6
L package ~ 2nH

FIG. 10F

RF ENERGY RECOVERY SYSTEM

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 61/663,930 filed Jun. 25, 2012 under 35 U.S.C. §119(e) which application is hereby incorporated herein by reference in its entirety.

FIELD

Subject matter disclosed herein relates generally to radio frequency (RF) systems and, more particularly, to techniques for enhancing the power efficiency of RF systems.

BACKGROUND

Many radio-frequency circuits incorporate power resistors that dissipate energy during circuit operation. Such circuits include radio-frequency (RF) hybrids, power combiners and dividers, isolators, duplexers, outphasing RF power amplifiers, composite RF amplifier systems, termination networks and filters, among other devices. In such applications, resistors are often specified as isolation resistors, terminating resistors, dump resistors, etc. While the power dissipation associated with these resistors is undesirable, their terminal characteristics (and absorption of RF electrical power) are typically necessary for proper circuit function. For example, a terminating resistor of a specified resistance value is necessary in an isolating power combiner (such as, for example, a Wilkinson combiner) in order to preserve isolation between the combiner input ports. Likewise, a circulator having a terminating resistor coupled to a port thereof may be used as an RF isolator that passes signals flowing in one direction while terminating signals flowing in an opposite direction. The energy absorption of the resistor coupled to the circulator is required for the isolator to function as intended.

Because efficiency and energy draw are important in many RF applications, the loss associated with such resistors is undesirable. Moreover, the heat generated by such resistors is undesirable and can limit miniaturization of these items and place burdens on supporting devices (e.g., increasing cooling requirements, requiring fans, etc.). It would thus be desirable to realize the electrical function of these power resistors (e.g., providing a nearly constant equivalent resistive input impedance at a specified operating frequency and absorbing RF electrical power), while reducing the dissipation of the absorbed power.

SUMMARY

Systems and techniques are described herein for providing desired operational characteristics of a resistive termination in an RF system, while also allowing power that would normally be dissipated in a resistive termination to be recovered and converted to a useable form.

In accordance with one aspect of the concepts, systems, circuits, and techniques described herein, an energy recovery system is provided that is capable of simulating the behavior of a resistive termination by providing an approximately constant, resistive input impedance to a radio frequency (RF) circuit coupled to the energy recovery system, while recovering energy that would otherwise have been dissipated had a resistive termination been used. The energy recovery system may comprise: a rectification system having a plurality of tuned rectifier circuits and a plurality of rectifier input ports corresponding to the plurality of tuned rectifier circuits, the rectification system to rectify RF energy received at the plurality of rectifier input ports to generate a direct current (dc) signal at one or more rectifier output ports, wherein the tuned rectifier circuits are each tuned to provide an approximately resistive input impedance across a finite input power range; an RF input network having an RF input port and multiple RF output ports, the RF input port forming an input of the energy recovery system and the multiple RF output ports being coupled to the plurality of rectifier input ports of the rectification system, the RF input network to provide input power distribution to the plurality of tuned rectifier circuits and input impedance shaping to shape the combined input impedance of the plurality of tuned rectifier circuits as seen at the RF input port of the RF input network; and a dc-dc converter system having an energy recovery input port and a dc output port, the energy recovery input port of the dc-dc converter system coupled to the output port of at least one rectifier to convert rectified power to a useable dc output voltage, wherein the voltage at the energy recovery input port of the do-dc converter system is regulated.

In accordance with a further aspect of the concepts, systems, circuits, and techniques described herein, a machine implemented method for use in capturing energy output by a radio frequency (RF) circuit comprises: receiving an RF signal from the RF circuit; distributing the RF signal among a plurality of tuned rectifier circuits, the plurality of tuned rectifier circuits to each generate a direct current (dc) signal at an output thereof in response to the RF input signal, wherein the plurality of tuned rectifier circuits are each tuned to provide an approximately resistive input impedance across a finite input power range; providing impedance shaping for the approximately resistive input impedances of the plurality of tuned rectifier circuits to provide an approximately constant, resistive input impedance to the RF circuit; and regulating an output voltage of the plurality of tuned rectifier circuits.

BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing features may be more fully understood from the following description of the drawings in which:

FIG. **1** is a block diagram illustrating an exemplary energy recovery system in accordance with an embodiment;

FIGS. **2**A-**2**F are schematic diagrams illustrating energy recovery dc-dc converter systems in accordance with various embodiments;

FIGS. **3**A-**3**F are schematic diagrams illustrating tuned rectifier circuits that use diodes to perform rectification in accordance with various embodiments;

FIGS. **4**A-**4**B are schematic diagrams illustrating tuned rectifier circuits that use transistors to perform rectification in accordance with various embodiments;

FIG. **5** is a schematic diagram illustrating a rectifier circuit having a series tank and a parallel tank coupled to an input thereof in accordance with an embodiment;

FIG. **6**A is a schematic diagram illustrating an example resistance compression network (RCN) that may be used to provide power dividing and impedance shaping at the input of an energy recovery system in accordance with an embodiment;

FIG. **6**B is a schematic diagram illustrating a higher order RCN that may be used to provide power dividing and impedance shaping at the input of an energy recovery system in accordance with an embodiment;

FIG. **6**C is a schematic diagram illustrating an RCN that utilizes transmission-line sections and reactances to provide power dividing and impedance shaping at the input of an energy recovery system in accordance with an embodiment;

FIG. **6**D is a schematic diagram illustrating a network that provides un-equal power sharing and impedance shaping at the RF input of an energy recovery system in accordance with an embodiment;

FIG. **6**E is a schematic diagram illustrating a network that may be used to provide additional power division using an isolating power splitter to divide power among multiple resistance-compressed rectifier systems in accordance with an embodiment;

FIGS. **7**A and **7**B are schematic diagrams illustrating conversion networks that may be used to transfer energy from a floating input voltage port to a common-referenced output port in accordance with embodiments;

FIG. **8**A is a schematic diagram illustrating an energy recovery system using an isolator as part of the energy recovery system input network in accordance with an embodiment;

FIG. **8**B is a schematic diagram illustrating an energy recovery system using a circulator and providing “multilevel” isolation and energy recovery in accordance with an implementation;

FIG. **9**A is a block diagram illustrating an isolator system using energy recovery at a termination port thereof in accordance with an embodiment;

FIG. **9**B is a block diagram illustrating a combiner/splitter using energy recovery in accordance with an embodiment;

FIG. **9**C is a block diagram illustrating a combiner/splitter using energy recovery that utilizes differential to single-ended conversion in the energy recovery system in accordance with an embodiment;

FIGS. **9**D and **9**E are block diagrams illustrating example power amplification systems using energy recovery in accordance with various embodiments;

FIG. **10**A is a schematic diagram illustrating a transmission-line resistance compression network (TLRCN) providing power dividing and impedance shaping and having an input port and two output ports, with the output ports loaded with a pair of resistive loads;

FIG. **10**B is a plot of the normalized input resistance versus the normalized load resistance for a single-level transmission-line resistance compression network, with the normalized load resistance plotted on a log scale;

FIG. **10**C is a plot showing the ratio of maximum input resistance to minimum input resistance plotted versus the ratio of maximum load resistance to minimum load resistance for a single-level transmission-line resistance compression network, assuming balanced compression;

FIG. **10**D is a schematic diagram illustrating a transmission-line resistance compression network (TLRCN) providing power dividing and impedance shaping and having an input port and two output ports, with the output ports loaded with a pair of resistive loads;

FIG. **10**E is a schematic diagram of an exemplary multi-level transmission-line resistance compression network (TL-RCN) providing power dividing and impedance shaping and having an input port and four output ports, with the output ports loaded with resistive loads; and

FIG. **10**F is a schematic diagram of an exemplary resistance-compressed rectifier system incorporating a multi-level transmission-line resistance compression network (TLRCN) and a set of resonant rectifiers.

DETAILED DESCRIPTION

Systems and techniques are described herein for providing desired operational characteristics of a resistive termination in a radio frequency (RF) system, while also allowing power that would normally be dissipated in a resistive termination to be recovered and converted to a useable form. Likewise, a termination to a radio-frequency power receiver (such as an antenna or coil) that maintains desirable operating characteristics (e.g., providing a constant matched resistive termination) as operating conditions change can capture RF power that would normally be reflected or lost. To recover RF power and convert it to direct current (dc) signals, the power must be rectified. However, rectification presents many challenges in practice, especially at microwave frequencies (e.g., frequencies above 300 MHz).

First, while one may desire a constant resistive input impedance (e.g., 50 Ohms) to realize a desired circuit function, rectifiers at microwave frequencies typically present effective input impedances that are difficult to make resistive and which vary with RF power level and dc output voltage. In systems having a high peak-to-average power ratio, the problem is particularly challenging. Second, harmonics in the RF input or associated with the rectifier system itself can affect rectifier behavior, and harmonic voltages or currents generated at the energy recovery system input by rectifier operation can be problematic to overall system operation. Third, devices capable of rectifier operation at microwave frequencies are often small (and hence cannot individually handle needed power levels), and rectifier circuits employing them are constrained to operate at low alternating current (ac) input voltages and at dc output voltages that may not be useful for utilizing recovered energy. Moreover, at microwave frequencies, the parasitics associated with rectifier devices (such as, for example, device capacitance and package inductance) can have substantial influence on circuit operation, and make it difficult to provide the desired RF input characteristics. Lastly, the RF input port where the resistor is being replaced is often floating with respect to system ground, while the recovered energy must often be provided at an output port that is referenced to system ground. In various aspects described herein, techniques and circuits are provided that make energy recovery at microwave frequencies practical and effective.

FIG. **1** is a block diagram illustrating an exemplary energy recovery system **10** in accordance with an embodiment. As illustrated, the energy recovery system **10** includes: an energy recovery dc-dc converter system **12**, a plurality of tuned rectifier circuits **14**, and an RF input network **16** providing functions such as input impedance shaping, filtering, RF power distribution to the plurality of rectifier circuits, and/or level shifting. As will be described in greater detail, in some alternative embodiments, multiple sets of rectifiers and RF input networks may be coupled to the same energy recovery dc-dc converter. In some other embodiments, an individual energy recovery dc-dc converter system may be provided for each rectifier. Other architectures are also possible.

As mentioned above, one challenge to realizing a practical RF energy recovery system is the fact that effective rectifier input impedance (i.e., the complex ratio of the fundamental component of RF voltage to fundamental RF current at the rectifier input port) depends upon the RF input power level and the rectifier output voltage. Moreover, the output voltage that a practical rectifier device or circuit can operate into is often not at a useful level for utilizing recovered energy. The energy recovery dc-dc converter system **12** is provided to address this challenge.

FIG. **2**A is a schematic diagram illustrating an exemplary energy recovery dc dc converter system **20** in accordance with an embodiment. As illustrated, the energy recovery dc-dc converter system **20** includes a single dc-dc converter **22**. In some implementations (e.g. designs operated to include burst-mode operation), input and output capacitances C_{RO} **24** and C_{OUT} **26** may be provided at the input and output of converter **22** to limit voltage ripple under bursting at frequencies below the converter switching frequency. The energy recovery dc-dc converter system **20** of FIG. **2**A may be used in an energy recovery system that uses, for example, a single do-dc converter (e.g., the energy recovery system of FIG. **1**). As illustrated, converter system **20** includes an input port **28** to be coupled to the outputs of the plurality of rectifiers and an output port **30** to be coupled to a load network to which recovered energy is to be delivered. In order to provide desired impedance characteristics at the RF energy recovery system input, the energy recovery dc-dc converter system **20** may be controlled to regulate the voltage at input port **28** while absorbing the power delivered from the rectifiers and delivering the absorbed power to output port **30**. This can be achieved using, for example, a dc-dc converter such as a buck converter, a boost converter, a flyback converter, a buck-boost converter, or other topology, with control designed to regulate the input voltage of the dc-dc converter. Output power from the converter can be provided to a regulated or unregulated output.

The energy recovery system may often be required to operate at high peak-to-average RF power ratios. Consequently, the instantaneous power delivered to the energy recovery dc-dc converter input port **28** may vary over a wide range. Therefore, in some embodiments, converter **22** may be optimized for operation over a wide range of power levels, including the use of burst mode, cycle skipping, gate-width switching, phase shedding, and related techniques to maintain high efficiency over a wide power range.

In one implementation, the dc-dc converter **22** of FIG. **2**A may be rated for the peak RF power to be recovered, thereby enabling continuous operation at peak RF power conditions. In some applications, however, it is known that such peak power conditions only occur for a limited duration. In these applications, the energy recovery converter input capacitor **24** (or other energy storage element) may be made sufficiently large such that the converter only needs to process a fraction of the peak power rating of the recovery system (but at least the average power rating), while still enabling the voltage at the rectifier output (and the energy recovery converter input **28**) to be regulated to within a desired voltage range.

In many applications, it may be desirable to regulate the energy recovery converter input **28** to a nearly fixed voltage independent of rectifier input power, to provide desired operating characteristics (and effective input impedance values) of the rectifiers and, in turn, desired input impedance characteristics of the energy recovery system. In one exemplary embodiment, this is achieved by comparing the energy recovery power converter input voltage to a reference voltage to generate an error signal, and controlling the converter input current, switch current, burst rate, switching frequency, duty ratio or other control variable as a function of the error signal. That is, in this approach, the converter input voltage is regulated to a value at or near the reference voltage by feedback control of the converter. FIG. **2**D is a block diagram illustrating example circuitry for performing this voltage regulation in accordance with an exemplary embodiment.

Alternatively, the energy recovery converter input voltage V_{RO} may be regulated to a voltage that is a function of the current or power delivered to or from the rectifiers. This can be used to extend the operating power range over which the energy recovery system has desirable effective input impedance characteristics (e.g., 50 Ohms, etc.). To understand this, it should be recognized that the effective input resistance of a rectifier is a function of both the power level at which the rectifier is operating and of the rectifier output voltage. For example, as is known in the art, some rectifiers ideally provide an input resistance at the fundamental operating frequency of the form $R_{eff}=k(V_{RO})^2/P$, where R_{eff} is the effective resistance at the rectifier input, V_{RO} is the rectifier dc output voltage, P is the rectifier output power, and k is a constant that depends on the rectifier topology. For an ideal full-bridge diode rectifier topology, for example, $k=8/\pi^2$. For an ideal voltage-doubler rectifier topology, $k=2/\pi^2$ and so on. By adjusting the energy recovery converter input voltage (rectifier output voltage) to be smaller for low power (or rectifier (or rectifier current) levels, the effective resistance provided by the rectifier can be more approximately constant (vary over a smaller range) than if a constant voltage V_{RO} is used. This can be achieved by adjusting the above-mentioned reference voltage as a function of measured, estimated, or expected power, current or another related variable. It should be noted that to achieve this dynamic reference alternative, it is desirable to size the energy recovery converter input capacitance C_{RO} (in conjunction with any rectifier output capacitance) large enough such that the rectifier output voltage ripple at the RF frequency and harmonics is small, but small enough that the rectifier output voltage can adjust on a time scale associated with the power variations to be compensated. FIG. **2**E is a block diagram illustrating example circuitry for performing voltage regulation with a dynamic reference in accordance with an exemplary embodiment.

FIG. **2**B is a schematic diagram illustrating an energy recovery dc-dc converter system **40** in accordance with another embodiment. In this embodiment, the energy recovery converter **42** is provided with a third port **44** from which energy may be drawn and/or delivered. With a third port, the energy recovery converter **42** can be configured to regulate the voltage at both its input port (V_{RO}) **46** and its output port (V_{OUT}) **48**. For example, in one approach, the converter **42** can be configured as a two-input dc-dc converter with a single output. In another approach, the converter **42** can be configured to transfer power in any direction among the ports **44**, **46**, **48**. Power needed to regulate the output port voltage that is not sourced from the input port (and rectifiers) can instead be sourced from the third port **44**. Typically, the third port **44** will provide an average source of power, but in some cases it may be constructed as a large energy buffer only, such that long-term average power is always sourced from the input port and only temporarily stored at the third port **44**. As described above, V_{RO} may be regulated to a fixed voltage or to a variable voltage.

FIG. **2**C is a schematic diagram illustrating another exemplary energy recovery dc-de converter system **50** in accordance with an embodiment. In this embodiment, a first converter **52** supplied from a third port **54** (V_{P3}) regulates the voltage (V_{RO}) at the energy recovery converter system input port **56** (i.e., the rectifier output), while a second converter **58** is supplied from this voltage and regulates voltage (V_{OUT}) at an output port **60**. While the first and second converters **52**, **58** may both be provided as switched-mode converters, in some applications the second converter **58** (regulating V_{OUT}) may be realized as a linear regulator to reduce noise at the output port **60**. In some embodiments, additional converters powered from V_{RO} may be utilized to regulate additional outputs. The additional converters can each be powered directly from V_{RO} or they can be cascaded so that the output of one converter powers the input of another, and so on.

FIG. **2**F is a schematic diagram illustrating an energy recovery dc-dc converter circuit **100** that may be used within an energy recovery system in accordance with an embodiment. The dc-dc converter circuit **100** uses an off the shelf, integrated, high voltage boost converter IC **102** (e.g., Texas Instruments TPS61170) that is coupled to control circuitry for supporting operation in an energy recovery system. The high voltage boost converter IC **102** has a feedback port that was designed for use in controlling the output voltage of the converter. As shown in FIG. **2**F, the energy recovery dc-dc converter circuit **100** includes logic coupled to the feedback port of the boost converter IC **102** that causes the chip to regulate the input voltage of the converter rather than the output voltage. It should be appreciated that the energy recovery dc-dc converter circuit **100** of FIG. **2**F is an example of one possible converter design that may be used within an energy recovery system in an embodiment. Other converter designs may be used in other implementations.

As described previously, one of the functions of the dc-dc converter system **12** of FIG. **1** is to convert the dc output signal of the rectifier(s) to a useable dc output voltage. As used herein, the phrase "useable dc output voltage" and similar terms refer to voltages that are useable to perform some desired function. In various implementations, the dc output voltage of the converter may be used to, for example, power other circuitry or devices, charge a battery, provide power to a power amplifier and/or to control circuitry, or perform some other function. In each case, the dc-dc converter system may be used to generate a dc output voltage that is useable to support the corresponding function.

As described above, one component in the energy recovery system **10** of FIG. **1** is the rectifier circuit that converts RF to dc. As shown in FIG. **1**, in some implementations, a plurality of rectifiers **14** may be used to provide the overall conversion from RF to dc. A plurality of rectifiers may be used for multiple reasons. First, devices capable of efficient rectification at microwave frequencies are often small and, therefore, may not be able to handle the needed power levels individually. Second, it is often easier to take advantage of circuit parasitics (or mitigate parasitics) for smaller rectifier devices than for larger ones. In addition, as will be described in greater detail, use of a plurality of rectifiers permits a rectifier system to be constructed that provides lower effective input impedance variation than is provided by a single rectifier.

In the discussion that follows, the design of individual tuned rectifier circuits will be described. Techniques for integrating multiple rectifier circuits into a rectification system will be described later. It is desired to have a rectifier that provides approximately resistive effective input impedance (i.e., having the fundamental component of RF voltage substantially in phase with fundamental RF current at the rectifier input port) with as little variation in effective impedance (ratio of fundamental voltage to current) as possible across operating conditions. As is known in the art, the effective impedance presented by a rectifier is a function of power and rectifier output voltage. Moreover, with conventional rectifiers (e.g., implemented with diodes), even the small parasitics associated with device capacitance, package capacitance and package inductance can cause substantial deviations from resistive input impedance. To address these challenges, in at least one embodiment, rectifiers incorporating tuned networks are used, where the tuned networks incorporate device and/or package parasitics to achieve the desired operating characteristics. Such tuned or resonant rectifiers can provide nearly resistive input resistance characteristics across a wide power range (e.g., impedance phase magnitude of less than 20 degrees over more than a 5:1 power range, etc.).

FIG. **3**A is a schematic diagram of an exemplary tuned rectifier circuit **110** in accordance with an embodiment. The tuned rectifier circuit **110** is shown using a Schottky diode **112**, but other device types may also be used (e.g., lateral field-effect rectifiers, etc.). Capacitor **114** represents a capacitance (C_R) including one or more of diode capacitance, diode package capacitance, and added circuit capacitance, and forms part of the tuning network. Capacitor **116** represents the rectifier output capacitance (C_{ROUT}). It provides a low impedance to RF, yielding approximately constant voltage at the rectifier output. Inductor **118** represents the rectifier inductance (L_R) which also forms part of the tuned network. The value of the rectifier inductance (L_R) may include diode package inductance.

In one approach, inductor **118** may be tuned near resonance with capacitor **114** (e.g., L_R on the order of $1/(4\pi^2 f^2 C_R)$). For a given value of C_R, the exact value of L_R can be selected using simulation software, such as SPICE. The network is loaded with a specified fixed output voltage V_{RO} and driven with a sinusoidal input current I_{RF} (at the design frequency), with L_R selected such that the fundamental of V_{RF} is nearly in phase with I_{RF} for a range of current amplitudes. One may adjust the value of C_R (by changing discrete capacitance, diode size, etc.) and subsequently retune L_R in order to achieve resistive operation of the rectifier at a desired range of power levels (and to best utilize the rectifier device capabilities). In some cases, to achieve the best resistive input impedance across power, a small reactance may be added (inductive or capacitive, not shown) in series with the positive RF input terminal of the rectifier **110**. This optional reactance, which would typically be much smaller than the reactance of L_R or C_R, can be used to offset any residual reactance from the tuning process.

FIG. **3**B is a schematic diagram illustrating an exemplary tuned rectifier circuit **120** in accordance with an embodiment. In tuned rectifier **120**, C_{ROUT} is again the rectifier output capacitance, which provides a low impedance at the operating frequency. C_R is the resonant capacitance and incorporates the diode capacitance, and L_R is the resonant inductance and may incorporate diode package inductance. In general, there needs to be a dc current path at the rectifier input. In the embodiment of FIG. **3**B, this path is provided by an explicit inductor (i.e., optional choke inductance **122** (L_{DC})) across the input terminals. In other embodiments, elements external to the rectifier **120** may be used to provide the dc current path. Tuning of this rectifier **120** may be carried out in a similar manner to the rectifier of FIG. **3**A. However, instead of driving the rectifier circuit **120** with a sinusoidal current during tuning, the circuit may be driven with a sinusoidal voltage while observing the fundamental of the rectifier input current.

In some implementations, rectifier circuits may be provided that also include tuning for higher-order harmonics. FIG. **3**C is a schematic diagram illustrating such a tuned rectifier circuit **130** in accordance with an embodiment. As illustrated, the rectifier circuit **130** of FIG. **3**C includes a series tank **132** having an inductance L**2** and a capacitance C**2** tuned at the second harmonic to reduce device voltage stress. Rectifier circuits may also be provided that use transmission-line sections to perform some or all of the tuning function. For example, FIG. **3**D is a schematic diagram illustrating a tuned rectifier circuit **140** using transmission line segments in accordance with an embodiment. As illustrated, the rectifier circuit **140** of FIG. **3**D includes a quarter-wavelength line segment **142** to provide reduced voltage stress. In some implementations, the line length of this segment may be adjusted slightly longer or shorter than a quarter wavelength to compensate for rectifier device capacitance and/or inductance. FIG. **3**E is a schematic diagram illustrating a class E/F rectifier **150** having a differential input port **152** in accordance with an embodiment. This rectifier circuit **150** can be driven differentially, or its input can be converted to a single-ended port using a splitter/combiner. It will be appreciated that other types of tuned rectifier circuit may be used, including those with other harmonic tunings, and those incorporating multiple diodes. In general, the tuning of a rectifier may involve selecting the inductances and capacitances such that the ac current (or voltage) at the rectifier input is substantially in-phase with the ac voltage (or current) at the rectifier input for the frequency of interest.

FIG. **3**F is a schematic diagram illustrating a resistance compressed rectifier system **160** having a pair of tuned rectifiers preceded by a resistance compression network having a pair of 50 ohm transmission line segments in accordance with an embodiment. The resistance compression network acts as an RF input network for the tuned rectifiers (although additional RF input network circuitry may be added in some implementations). The resistance compressed rectifier system **160** of FIG. **3**F is operative at 2.14 GHz with a dc output voltage of 7 V and an RF input impedance near 50 Ohms.

While the tuned rectifier circuits of FIGS. **3**A-**3**F are shown with Schottky diodes, it should be appreciated that other diode types or device types may be employed in other implementations to perform the rectification function. In some implementations, for example, transistors may be used to provide rectification (e.g., MOSFETs, JFETs, HEMTs, HBTs, etc.). FIG. **4**A is a schematic diagram illustrating a tuned rectifier circuit **170** that uses a transistor **172** to perform rectification in accordance with an embodiment. As illustrated in FIG. **4**A, to perform rectification, the gate of the transistor **172** may be biased to an appropriate dc potential (e.g., V_B). Alternatively, the gate terminal may be actively driven at RF to provide synchronous rectification (e.g., V_B in FIG. **4**A may include a dc bias plus an RF signal with appropriate phase relative to V_{RF}). FIG. **4**B is a schematic diagram illustrating a tuned rectifier circuit **180** that performs synchronous rectification using transistors **182**, **184** that are “cross-driven” from the RF input signal itself. In general, synchronous rectification involves turning a transistor on and off with timing such that it carries an average dc current, thus providing conversion from an ac input to a dc output. In one approach, each transistor gate may be driven such that the transistor conducts current of substantially one polarity, and is turned off otherwise. That is, the transistor may be driven to conduct in a manner similar to a diode, but having a lower device drop than a diode.

Another component in the energy recovery system of FIG. **1** is the RF input network **16** that addresses challenges associated with capturing RF energy. These challenges include the generation of undesired frequencies by the rectifiers, the need to operate with small individual rectifier devices, rectifier input impedance variations, operation at high peak-to-average power ratios, and energy capture at circuit ports that are not referenced to a common potential. As will be described below in greater detail, each of these challenges may be addressed by specific elements of the RF input network.

A byproduct of the rectification process is often the presence of dc, switching harmonics, and other undesired frequency components at the rectifier inputs. These undesired frequency components can cause deleterious effects in the energy recovery system. For example, the unwanted frequency content can pollute the system that the energy recovery system is connected to, can cause undesired interactions among the multiple rectifier circuits, as well as other negative effects. To suppress this content, filtering may be provided at the rectifier inputs and possibly at points closer to the energy recovery system input. In some embodiments, these filters may include tuned series and/or parallel tanks in series with the rectifier inputs (and/or interconnections) or in parallel with the rectifier inputs. For example, FIG. **5** is a schematic diagram illustrating a rectifier circuit **190** having a series tank **192** and a parallel tank **194** coupled to an input thereof, where the series tank **192** is tuned to the input frequency to pass through the desired fundamental and the parallel tank **194** is tuned to the input frequency to suppress unwanted content. Other bandpass and/or bandstop filtering (e.g., tuned to harmonics) may also be employed, as well as low-pass and high-pass filters to suppress undesired content. These filters may be implemented with discrete or integrated passive components, with transmission-line sections, or as a combination of these elements.

As described previously, better rectifier performance can sometimes be achieved at RF frequencies using rectifier devices that are small in power rating and physical size, making it advantageous to construct an energy recovery system from multiple rectifiers of reduced power rating. Moreover, utilizing a plurality of rectifiers, one can construct a rectifier system providing lower effective input impedance variation than is provided by a single rectifier, as described in detail below. To implement a multi-rectifier system, in at least one embodiment, the RF input network may include a power dividing and impedance shaping network that splits power among the various rectifiers and provides an input impedance that varies over a smaller range (e.g., a ratio of input resistances or input impedance magnitudes) than the individual rectifiers. Desirable attributes of such a power dividing network include relatively even power splitting among the individual rectifiers and lossless or very low loss operation.

In some embodiments, a resistance compression network (RCN) may be used to provide the power dividing and impedance shaping network of the RF input network. FIG. **6**A is a schematic diagram illustrating an exemplary resistance compression network **200** that may be used in an embodiment. This network is ideally lossless and splits input power among two rectifiers **202**, **204** approximately equally, while providing a resistance at its input port **206** that varies over a smaller range than the individual rectifiers **202**, **204**. This thus forms a “resistance-compressed rectifier system.” The compression branch impedance magnitude X is selected based on the input impedance range provided by the rectifiers **202**, **204**. In constructing RCNs of the type in FIG. **6**A for operation at microwave frequencies, care should be taken to minimize or compensate (e.g., “resonate out”) shunt parasitics at the intermediate nodes (e.g., between the inductors and capacitors).

The resistance compression network **200** of FIG. **6**A, constructed with discrete components, has the benefits of being compact and providing good operational bandwidth. Higher-order resistance compression networks (or “multilevel” compression networks) can also be used, such as the “second order” (or two-level) RCN **210** of FIG. **6**B that can split power among four rectifier circuits **212**, **214**, **216**, **218**. An RCN of order M, or an M-level RCN, can split power among N=2M rectifier circuits. Other RCN variants may alternatively be used, including RCNs that are achieved by topological duality or star-delta transformations of the network elements, RCNs that use different component types to realize the needed reactances, and RCNs that use circuit board or microstrip elements.

Power dividing and resistance compression can also be achieved with circuits incorporating transmission-line sections (e.g., section of microstrip, etc.). FIG. **60** is a schematic diagram illustrating a resistance compression network **230** that uses quarter wavelength transmission line sections **236**, **238** and reactances **240**, **242** to perform resistance compression for first and second rectifiers **232**, **234**. In resistance compression network **230**, the reactive components **240**, **242** may be realized as discrete components, components implemented in the circuit board, or as transmission-line stubs. In this embodiment, in addition to working with the shunt elements to realize resistance compression, the characteristic impedance of the quarter-wave lines **236**, **238** can be selected to provide a desired degree of impedance transformation, FIG. **3**F described previously illustrates the details of an example implementation of a resistance-compressed rectifier system **160** that uses the resistance compression network architecture illustrated in FIG. **60**.

The power dividing and impedance shaping network of the RF input network may also be implemented using other network types and topologies. For example, FIG. **6**D is a schematic diagram illustrating a network **250** that serves to provide a degree of impedance shaping such that the impedance at the RF input **252** varies over a smaller range (e.g., ratio of input resistances or input impedance magnitudes) than the individual rectifiers **254**, **256**, though it does not cause power to be shared equally among the rectifiers **254**, **256** across operating conditions. The network **250** can be extended for use with additional rectifiers by paralleling additional rectifier-loaded quarter-wave-line branches having different characteristic impedances and appropriately designed rectifiers.

Additional power dividing can be obtained using networks that are not ideally lossless and/or which do not provide reduction in impedance range. For example, power may be split further using an isolating power splitter to divide up power to multiple resistance-compressed rectifier systems. FIG. **6**E illustrates a network **260** that uses a two way splitter **262** (e.g., a Wilkinson splitter, a Gysel splitter, a hybrid splitter, etc.) to split power between first and second resistance-compressed rectifier systems **264**, **266**. Many-way splitters can be used when there are more than two resistance-compressed rectifier systems. While providing power dividing in this fashion does not provide additional reduction in impedance ranges, it can be made highly efficient, and the isolation helps soften the variations in input impedance. Moreover, it enables systems of different energy recovery power capacities to be simply constructed.

In some embodiments, the RF input port of an energy recovery system is referenced to a common potential (e.g., ground, etc.). However, in some cases, the input port may be "differential" or floating (or "flying") with respect to system ground, while the recovered energy must be provided at an output port that is referenced to system ground. For example, the isolation port of a Wilkinson combiner is not ground referenced, but represents a "differential" input. One technique to recover energy at a floating or "flying" input port involves the use of an RF transformer or balun to transfer energy from the flying input voltage to a common-referenced port. FIG. **7**A is a schematic diagram illustrating an example conversion network **270** for doing this in accordance with an embodiment. If such an approach is used, the energy may be recovered at the common-referenced port as described above. In the embodiment of FIG. **7**A, energy at port AB **272** is delivered to output port CG **274** with $V_{CG}=(N_2/N_1)\cdot V_{BA}$. If the effective turns ratio is not unity, network **270** also provides an impedance transformation such that $Z_{diff}=(N_1/N_2)^2\cdot Z_{RF}$. Depending on dc levels, one or both of the ports **272**, **274** may require a series blocking capacitor, and one may also need to tune out transformer parasitics (such as leakage inductances and/or magnetizing inductance) at the RF frequency of interest (e.g., using capacitors).

Another technique to transfer energy from a "flying" input port to a common-referenced port for energy recovery purposes involves the use of transmission-line sections. FIG. **7**B is a schematic diagram illustrating an example conversion network **280** that uses transmission line sections **286**, **288**, **290** to transfer energy between ports in accordance with an embodiment. At the design frequency, conversion network **280** transfers energy from a floating port **282** to a common-referenced port **284** according to the relationship $V_{CG}=(Z_1/Z_0)\cdot V_{BA}$. Network **280** can provide both a differential to single-ended conversion and an impedance transformation, such that $Z_{diff}=(Z_0/Z_1)^2\cdot Z_{RF}$. In some cases, the Z_1 transmission line section can be omitted. For example, in cases where the load provided by the energy recovery system is at or near impedance Z_1, the transmission section can be omitted. Similarly, in cases where the impedance inversion provided by the quarter-wave section **290** is not a problem, this transmission line section can be omitted. Also, it will be recognized that blocking capacitors may be used in series with one or more of the terminals of the differential conversion network in some implementations. Note that the network can also be implemented with lumped transmission-line approximations, with other immittance converting circuits taking the function of each quarter-wave section, or combinations of lumped and distributed structures serving the same or similar functions.

Other alternative techniques for converting "flying" or differential inputs to common-referenced outputs for energy recovery can be utilized in other embodiments. Combined with the other portions of the energy recovery system, these techniques enable RF energy to be recovered from "flying" ports.

It is recognized that even with filtering, some applications may be sensitive to any reflected power from the energy recovery system (including at the operating frequency, harmonics, or other frequencies). In such cases, an isolator may be used as part of the energy recovery input network to terminate any reflected power generated by the energy recovery system. As shown in FIG. **8**A, an isolator **300** may be placed at the input of an energy recovery block **302** so that reflections from the input are directed to termination **304**. It should be appreciated that other isolator locations are also possible.

FIG. **8**B is a schematic diagram illustrating an energy recovery system **310** that uses an isolator **312** to provide "multilevel" isolation and energy recovery in accordance with an implementation. As shown, a number of energy recovery blocks **314**, **316** are coupled to ports of isolator **312**. Each energy recovery block **314**, **316** incorporates one or more energy recovery rectifiers and possibly other subsystems, as described above. Multilevel isolation provides the opportunity to capture RF power at different power levels and contents, and may be implemented with a 4 or more port circulator and a terminating resistor, or a cascade of 3-port circulators and a terminating resistor.

The availability of an energy recovery system also enables additional functions and features to be realized. One such function is power monitoring. For example, an additional circuit can be provided that monitors information about the energy recovery system, such as the input and/or output power of the system. In some implementations, the system output power is monitored using the sensing and/or control signals of the energy recovery dc-dc converter, in some other implementations, the output power may be monitored using additional low-frequency sensors or other structures. In at least one embodiment, the input power may be monitored based on the operating point of the rectifier circuits. Alternatively, the input power may be monitored based on the operation of the energy recovery converter (e.g., estimating input power based on the known characteristics of the system and the converter input or output current, voltage, or power or rectifier input or output voltage, current, or power). Additional monitoring circuitry may also be provided in some embodiments. This may include, for example, circuitry for monitoring system temperature, operating status, and/or other critical operating parameters.

Many RF circuits incorporate power resistors that dissipate energy during circuit operation. Such circuits include RF hybrids, power combiners and dividers, isolators, RF power amplifier systems, duplexers, filters and termination networks, among other devices. Each of these circuit types can benefit by incorporating the energy recovery techniques described herein. Likewise, many RF circuits seek to absorb energy from an antenna, transformer secondary, coil, or other means of receiving RF energy, including in rectenna systems, wireless power transfer systems, inductive power coupling systems, radio-frequency power converter systems including RF dc-dc converters, and microwave power transmission systems. It is desirable to be able to capture RF energy in these systems and convert it to useable form, while minimizing reflected or dissipated RF power. Each of these circuit types can likewise benefit by incorporating the energy recovery techniques described herein.

One type of circuit that can benefit from the described techniques is an isolator. FIG. **9**A is a block diagram illustrating the use of an energy recovery system **322** (e.g., the energy recovery system of FIG. **1**, etc.) as a termination for an isolator **320**. As illustrated, the isolator **320** includes a circulator **324** having three ports, with the energy recovery system **322** coupled to one of the three ports (i.e., instead of a conventional resistive termination). A first port of the circulator serves as the input port of the isolator **320** and a second port serves as the output port. RF power delivered to the isolator input port is transferred to the isolator output port. Any power entering the isolator output port (e.g., power reflected back into the isolator output port from a mismatched load and/or any other power) is transferred to the energy recovery system **322** at the third port of the circulator **324**. In this manner, the energy may be at least partially recovered for other uses instead of being lost to dissipation as would occur with a conventional resistive termination.

Another application that can benefit from energy recovery is power combining and/or splitting. FIG. **9**B is a block diagram illustrating the use of an energy recovery system **330** to provide energy recovery for a combiner/splitter **332** in accordance with an embodiment. The combiner/splitter **332** may include, for example, a rat-race hybrid, a branchline hybrid, a 90 or 180 degree hybrid, or others.

In the illustrated embodiment, the combiner/splitter **332** includes four ports: port 1, port 2, a sum port, and a difference (delta) port. The energy recovery system **330** may be coupled to either the sum port or the difference port. The energy recovery system **330** may provide, for example, a proper termination for the corresponding port (e.g., the difference port in FIG. **9**B) while the other port (e.g., the sum port in FIG. **9**B) is used as a "combined" port. In this manner, combining and/or splitting may be achieved between the combined port and ports 1 and 2, while energy is recovered for isolation at the energy recovery output port.

In some systems, three (or more) way combining/splitting may be used. FIG. **9**C is a block diagram illustrating an implementation that utilizes differential to single-ended conversion in an energy recovery system. The combiner/splitter **348** is implemented as a 3-way Wilkinson combiner/splitter with the ports for isolation (with energy recovery) connected in delta, where each differential input port has an apparent input resistance of approximately $3Z_o$. This system uses three energy recovery rectifier systems **340**, **342**, **344**, each including differential-to-single-ended conversion. The system may also include an RF input network comprising impedance shaping, filtering, distribution, and a plurality of rectifiers, along with a single energy recovery dc-dc converter system **350** to provide recovered energy at the output. In an alternative approach, this system can be reconfigured such that there is a separate energy recovery dc-dc converter for each of the energy recovery systems. This technique may be used for splitting or combining any number of signals.

Another application that can benefit from the energy recovery techniques described herein is power amplification. More specifically, systems where power from multiple power amplifiers is combined to create a single higher power signal. FIG. **9**D is a block diagram illustrating an example power amplification system **400** using energy recovery in accordance with an embodiment. As shown, output power from two power amplifiers **402**, **404** is combined using an isolating combiner **406** with energy recovery (e.g., as illustrated in FIG. **9**B). Depending on the relative magnitude and phase of the outputs of the two power amplifiers **402**, **404**, power will be delivered to one or both of first and second output ports of the combiner **406**. The first output port is illustrated as the sum port and the second output port is illustrated as the difference port in FIG. **9**D. An energy recovery system **408** may be connected to the second output port. This provides the opportunity to load the power amplifiers **402**, **404** in an isolating fashion, and at the same time control power (and signal amplitude and phase) delivered to the output port by adjusting the magnitudes and phases of the outputs of the individual power amplifiers **402**, **404**. Power not delivered to the output port is recovered through the energy recovery system **408** (minus any losses). The magnitudes and phases of the outputs of the individual PAs **402**, **404** may be controlled by, for example, any one or more of: (1) individually adjusting one or more of the pulse widths, timing, drive amplitudes, and phases of the power amplifier input signals (e.g., including amplitude modulation, phase-shift or outphasing modulation, PWM modulation, and on/off modulation), (2) individually adjusting the power supply bias voltages of the power amplifiers (e.g., via discrete or continuous drain modulation), (3) load modulation of the individual power amplifiers (e.g., by implementing each power amplifier as a Doherty amplifier or an outphasing amplifier with lossless combining), and/or other techniques. It should be recognized that such a system can also be realized with a larger number of power amplifiers, such as by using the combining and energy recovery system illustrated in FIG. **9**C. Such a system enables high efficiency linear control of output power by adjusting the relative operation of the individual power amplifiers.

A power amplifier system with multiple power amplifiers can also be implemented as a balanced power amplifier system that realizes improved efficiency under load mismatch (e.g., from a non-ideal load impedance) at its output port. FIG. **9**E is a block diagram illustrating an exemplary system **420** of this type. In system **420**, reflected power from an output port of a balanced amplifier **422** is captured by an energy recovery system **424**. The recovered energy is provided at a dc output port **426** and may be used to supply part of the power to the power amplifiers **428**, **430** of balanced amplifier **422**, for powering computation or control circuitry, powering pre-amplifiers, and/or for other uses. The system **420** of FIG. **9**E includes a first 90-degree hybrid **440** to provide appropriate phases at the two power amplifier inputs, and a second 90-degree hybrid **442** to combine power from the two power amplifiers **428**, **430** for delivery to an RF output port. The energy recovery system **424** is connected to the isolation port of the second hybrid **442**. Reflected power entering the output port (e.g., owing to load mismatch) can be substantially captured by the energy recovery system **424**, thus providing improved operation under load mismatch.

Resistance compression networks (RCNs) absorb energy from a source and deliver it (ideally losslessly) to a plurality of loads (e.g., such a set of rectifiers), providing an input resistance that varies over a narrow range as the resistance of the loads vary together over a wide range. Resistance compression networks can also serve to reduce the phase of the input impedance as compared to the phase of the load impedances (phase compression). Ideally, the RCN splits input power substantially equally among the loads. Conventional resistance compression networks use reactances to accomplish this, and may additionally include transmission-line sections, as illustrated in FIGS. **3**F and **6**C. The reactances themselves may be implemented as transmission-line sections or stubs or otherwise realized in printed-circuit form.

In some embodiments, resistance compression networks may be implemented using transmission-line sections having asymmetric lengths, wherein the transmission-line sections are provided as two-port structures interconnecting among the source and loads. Such Transmission-Line Resistance Compression Network (TLRCN) implementations may provide several benefits. For example, they can (a) have low loss, (b) enable repeatable, low-cost implementation using printed circuit techniques, (c) provide filtering, and (d) with correct length selections, they can provide resistance compression at one or more harmonic frequencies. Moreover, at UHF frequencies and above, the discrete reactances often used in conventional RCN designs represent an increasing challenge. That is, the transmission-line effects associated with their physical size can substantially influence system behavior, and their numerical values can become extremely small, making them difficult to implement accurately and making the system susceptible to parasitic effects. TLRCN implementations, on the other hand, avoid these issues by directly realizing the RCN as a transmission-line structure interconnecting the source and loads.

A description will now be made of the theory and operating characteristics of Transmission-Line Resistance Compression Networks. With a TLRCN network, substantially balanced splitting of power to multiple loads and smaller variation in driving point resistance as compared to load resistance variation may be achieved using only transmission-line sections used as two-port devices to connect among the sources and loads. FIG. **10**A shows a basic TLRCN structure, having a single input port with voltage v_{IN} and two output ports with voltages v_{O1} and v_{O2}, with the output ports loaded with identical resistive loads R_L. In each transmission-line section, or branch, in FIG. **10**A the second terminal at each port of each transmission-line section may be treated as connected to a common potential at that end of the transmission-line section, as naturally occurs in a microstrip or other printed-circuit implementation, and is thus not shown explicitly. The first transmission-line branch **500** coupled to the input port has characteristic impedance Z_0 and a length l_1 corresponding to an electrical angular delay θ_1 at an operating frequency of interest. The second transmission-line branch **502** coupled to the input port likewise has characteristic impedance Z_0 and a different length l_2 corresponding to an electrical angular delay θ_2. (Implementation of systems in which the two branches have different characteristic impedances and which operate best with different loading impedances is also possible.) We can express the lengths of the two branches as a base length plus and minus a delta length (or a base angle+/−a delta angle) as follows:

$$l_1 = l_{Base} + \Delta l$$

$$l_2 = l_{Base} - \Delta l \tag{1}$$

$$\theta_1 = \theta_{Base} + \Delta\theta$$

$$\theta_2 = \theta_{Base} - \Delta\theta \tag{2}$$

We obtain desired operating characteristics at a frequency of interest by proper selection of Z_0, θ_{Base}, and $\Delta\theta$.

While there are multiple possibilities for base lengths, we first consider a base length l_{Base} of $\lambda/4$ (a quarter wavelength at a frequency of interest), corresponding to $\theta_{Base} = \pi/2$ radians. (Any base length with an additional multiple of $\lambda/2$ in length will provide similar results.) Considering the first and second branches **500**, **502** in FIG. **10**A individually, we find branch input admittances at the frequency of interest:

$$\begin{aligned} Y_{in,1} &= \frac{1}{Z_{in,1}} = \frac{1}{Z_0} \cdot \frac{Z_0 - jR_L\cot(\Delta\theta)}{R_L - jZ_0\cot(\Delta\theta)} \\ Y_{in,2} &= \frac{1}{Z_{in,1}} = \frac{1}{Z_0} \cdot \frac{Z_0 + jR_L\cot(\Delta\theta)}{R_L + jZ_0\cot(\Delta\theta)} \end{aligned} \tag{3}$$

Since these admittances are complex conjugates, the network will divide power entering the input port equally to both loads (for identical load resistances). The impedance seen at the input port at the frequency of interest is resistive in this case, and can be shown to be:

$$Z_{in} = R_{in}\frac{|\cot(\Delta\theta)|}{2[1+\cot^2(\Delta\theta)]} \cdot \left[\left\{\frac{R_L}{|\cot(\Delta\theta)|}\right\} + \frac{Z_0^2}{\left\{\frac{R_L}{|\cot(\Delta\theta)|}\right\}}\right] \tag{4}$$

Such a characteristic clearly realizes resistance compression: the input impedance is resistive and varies only over a small range as the load resistances vary together over a wide range. In fact, this input resistance characteristic bears close relation to the input resistance of a type of resistance compression network that uses reactances in series with the load networks (e.g., see FIG. **6**A). With a design selection of $\Delta\theta = \pi/4$ radians ($\Delta l = \lambda/8$), the circuit provides identical characteristics (with the transmission-line characteristic impedance Z_0 taking the place of reactance X in the expression for the input resistance).

Using the above-described network with a base angle of $\theta_{Base} = \pi/2$, "balanced" compression can be achieved for a range of load resistances having a geometric mean of $R_{L,center}$:

$$R_{L,center} = Z_0 \cdot |\cot(\Delta\theta)| \tag{5}$$

At this load resistance value ($R_{L,center}$), the input resistance takes on a minimum value of $R_{in,min}$, with larger input resistance for other load resistances. The value of $R_{in,min}$ may be calculated as follows:

$$R_{in,min} = Z_0 \cdot \frac{|\cot(\Delta\theta)|}{[1+\cot^2(\Delta\theta)]} \tag{6}$$

FIG. **10**B shows a plot of the input resistance (normalized to $R_{in,min}$) as a function of the load resistance (normalized to $R_{L,center}$), with the normalized load resistance on a log scale.

FIG. **10**C shows the detail of the ratio of the maximum to minimum input resistance plotted as a function of the maximum to minimum load resistance, assuming balanced compression (i.e., with $R_{L,center}$ selected as the geometric mean of $R_{L,max}$ and $R_{L,min}$). For this case, we find:

$$\frac{R_{in,max}}{R_{in,min}} = \frac{1}{2}\left[\sqrt{\frac{R_{L,max}}{R_{L,min}}} + \sqrt{\frac{R_{L,min}}{R_{L,max}}}\right] \tag{7}$$

Which, for large values of $R_{L,max}/R_{L,min}$ approaches:

$$\frac{R_{in,max}}{R_{in,min}} \approx \frac{1}{2}\sqrt{\frac{R_{L,max}}{R_{L,min}}} \tag{8}$$

The high degree of resistance compression provided by this system can be seen in these expressions and in FIG. **10**C, with a 10:1 ratio of load resistance compressed to a 1.74:1 ratio in input resistance and a 100:1 ratio in load resistance compressed to only a 5.05:1 ratio in input resistance. As with conventional resistance compression networks, one can also use the TLRCN to provide phase compression in which for identical non-resistive loads, the input impedance is more nearly resistive than the load impedances themselves.

It should be noted that other base lengths besides $l_{Base}=\lambda/4$ ($\theta_{Base}=\pi/2$ radians) can be used. For example, a base length of $l_{Base}=\lambda/2$ ($\theta_{Base}=\pi$ radians) likewise results in equal power transfer to the two loads and a compressed resistive input impedance. The characteristics associated with this base length will be the same as that in Equations (3)-(6), except with $\cot(\Delta\theta)$ replaced with $-\tan(\Delta\theta)$ in each expression. Likewise, operation will be the same with any additional multiple of $\lambda/2$ added to the base length. Because of this, with appropriate selections of base length at a desired fundamental frequency, the system can be designed to also provide resistance compression at one or more harmonic frequencies (and at subharmonic frequencies), provided that the loads have appropriate frequency characteristics.

There are multiple possible design approaches for TLRCN circuits. Considering a base length $l_{Base}=\lambda/4$ ($\theta_{Base}=\pi/2$ radians), one may a priori select a value of $\Delta\theta=\pi/4$ radians ($\Delta l=\lambda/8$), such that $\theta_1=3\pi/4$ and $\theta_2=\pi/4$. For balanced compression, one may then select Z_0 as $R_{L,center}$ (the geometric mean of the maximum and minimum load resistances $R_{L,max}$ and $R_{L,min}$). In this case, the load resistances loading each transmission-line section vary (geometrically) about the characteristic impedance of the lines, which helps reduce required transmission-line reflection and loss. This design choice provides a compressed input resistance having a range of values determined by the load resistances. In cases where this range of input resistance values is not what is desired for the system, an additional impedance transformation stage may be placed at the input of the TLRCN. One highly effective means to do this is to add an additional quarter-wave line at the input to the compression stage, as illustrated in FIG. **10**D.

The whole transmission-line structure of FIG. **10**D is likewise itself a resistance compression network having both a compression stage and a transformation stage. The quarter wave line is selected to have a characteristic impedance Z_T that provides an impedance transformation between the resistance obtained at the input of the compression stage (e.g., $Z_{in,median}$, the median of the range of the input resistance Z_{in} or a similar value related to Z_{in}) and that desired for the input of the system ($Z_{in,T}$). In this case, to obtain an overall input resistance near a desired value $Z_{in,T,desired}$, we may select:

$$Z_T=\sqrt{Z_{in,T,desired}\cdot Z_{in,median}}. \tag{9}$$

Another design approach takes advantage of the freedom to choose both the differential length and the characteristic impedance of the transmission lines. One can also choose the base length. In the example that follows, a base length of $\theta_{Base}=\pi/2$ radians at the operating frequency will be used. By properly selecting the differential length and the characteristic impedance, within certain bounds, one can directly realize both resistance compression and a desired specified input resistance directly with the structure of FIG. **10**A. In this design approach, one may start with a center resistance $R_{L,center}$ about which load resistances are compressed. For balanced compression, $R_{L,center}$ is chosen as the geometric mean of the maximum and minimum load resistances, $R_{L,max}$ and $R_{L,min}$. A desired minimum input resistance $R_{in,min}$ of any value below $R_{L,center}$ may then be chosen. The input resistance will take on values at and above $R_{in,min}$, with the range of R_{in} values depending on the range of the R_L values as per Equation (7). One option is to select $R_{in,min}$ to be the desired resistance $R_{in,desired}$ seen at the input of the TLRCN (recognizing that the actual input resistance will be at or above the desired value). Another option is to set the median value of the input resistance that occurs over the range of load resistances to match the desired input resistance. To achieve this for a given maximum to minimum load resistance ratio, one may select $R_{in,min}$ as follows:

$$R_{in,min} = \frac{2\cdot R_{in,desired}}{1+\frac{1}{2}\sqrt{\frac{R_{L,max}}{R_{L,min}}}+\frac{1}{2}\sqrt{\frac{R_{L,min}}{R_{L,max}}}} \tag{10}$$

One could likewise select $R_{in,min}$ to be some other value close to $R_{in,desired}$ to achieve good results.

Based on selecting $R_{L,center}$ and $R_{in,min}$ as described above (with resistance $R_{in,min}$ necessarily selected below $R_{L,center}$), one can directly choose Z_0 and $\Delta\theta$ of the network of FIG. **10**A to provide both the needed compression and impedance transformation, as follows:

$$\cot(\Delta\theta) = \sqrt{\frac{R_{L,center}}{R_{in,min}} - 1} \tag{11}$$

$$Z_0 = \frac{R_{L,center}}{\sqrt{\frac{R_{L,center}}{R_{in,min}} - 1}} \tag{10}$$

This design choice has the advantage of providing both resistance compression and impedance transformation (where needed) in a compact structure. However, the practicality of such an implementation will depend on the desired values. For example, as $R_{in,min}$ approaches $R_{L,center}$, the characteristic impedance of the transmission line can grow unreasonably large. It should also be appreciated that one can use the selection of the transmission-line impedance and differential length to provide a first-degree of impedance transformation, and add another quarter-wave transformer at the input (and/or a set of quarter-wave transformers between the base compression stage and the loads) to provide additional impedance transformation.

As with conventional resistance compression networks, one may construct multi-stage or multi-level compression networks to provide greater degrees of resistance compression than obtained with a single-level design. As with conventional resistance compression networks, this may be done by cascading single-level resistance compression stages in a tree structure, though other structural implementations of multi-level TLRCNs are possible. FIG. **10**E shows an exemplary two-stage TLRCN. This example utilizes $l_{Base}=\lambda/4$ ($\theta_{Base}=\pi/2$ radians), and differential lengths $\Delta l=\lambda/8$ ($\Delta\theta=\pi/4$) radians ($\Delta l=\lambda/8$), such that $\theta_1=3\pi/4$ and $\theta_2=\pi/4$ in each stage. With this selection of differential lengths, a multi-stage TLRCN can be optimized (in terms of minimizing the peak deviation from a median input resistance) just as with conventional reactance-based multi-level RCNs (see, e.g., "A New Power Combining and Outphasing Modulation System for High-Efficiency Power Amplification," by Perreault et al., *IEEE Transactions on Circuits and Systems—I*, Vol. 58, No. 8, pp. 1713-1726, August 2011; and U.S. Patent Publication No. 2011/0187437 to Perreault et al., entitled "Radio-Frequency (RF) Amplifier Circuits and Related Techniques," which is hereby incorporated by reference in its entirety), with the transmission-line characteristic impedances of the lines in the two stages taking the place of the reactance amplitudes of the two stages. Alternatively, both the characteristic impedances and the base and differential lengths of each stage can be optimized to provide an overall desired input resistance level and transformation, with the opportunity to achieve additional goals through the parameter selection, such as minimizing RCN loss.

Combining a transmission-line resistance compression network with a set of rectifiers forms a resistance compressed rectifier system. An exemplary resistance-compressed rectifier system **520** incorporating a TLRCN is shown in FIG. **10**F. This system is for 2.14 GHz operation, and has four tuned rectifiers as loads for the TLRCN that act as approximately resistive loads, each providing ac input impedance of –24-160 t over an input power range of 0.125-1.07 W (and a dc rectifier output power of 0.08-0.85 W) at a rectifier output voltage of 7 V. The rectifier also includes a series-resonant input filter, and is tuned for resistive operation by appropriate resonance between the device capacitance and output inductance. Note that these filtering and tuning elements can be provided as discrete components or constructed in a printed circuit board (optionally including elements such as transmission-line sections, gap capacitors, etc.). The TLRCN has two compression stages and a transformation stage. The first compression-stage "A" compresses the resistances presented by the rectifiers into a range of 31-46Ω (for the first-stage input impedances $Z_{in,A1}$ and $Z_{in,A2}$, which act as the load for the second stage of compression). The second stage further compresses this to an input impedance $Z_{in,B}$ of approximately 9-19.4Ω resistive over the rectifier load resistance range. The transformation stage, comprising a quarter-wave line, transforms this impedance into the overall input impedance of approximately 50Ω.

It will be appreciated that the inventive energy recovery system can be employed in numerous other applications in which energy is conventionally delivered to a resistive termination. By replacing the lossy termination with the energy recovery system, the power that would otherwise be lost can be captured and converted to a useful form, while maintaining a desirable loading characteristic at the termination port.

Having described exemplary embodiments of the invention, it will now become apparent to one of ordinary skill in the art that other embodiments incorporating their concepts may also be used. The embodiments contained herein should not be limited to disclosed embodiments but rather should be limited only by the spirit and scope of the appended claims. All publications and references cited herein are expressly incorporated herein by reference in their entirety.

What is claimed is:

1. An energy recovery system to simulate behavior of a resistive termination for a radio frequency (RF) circuit coupled to the energy recovery system, the energy recovery system comprising:

a rectification system having a plurality of tuned rectifier circuits and a plurality of rectifier input ports corresponding to the plurality of tuned rectifier circuits, the rectification system to rectify RF energy received at the plurality of rectifier input ports to generate a direct current (dc) signal at one or more rectifier output ports;

an RF input network having an RF input port and multiple RF output ports, the RF input port forming an input of the energy recovery system and the multiple RF output ports being coupled to the plurality of rectifier input ports of the rectification system, the RF input network to provide input power distribution to the plurality of tuned rectifier circuits and input impedance shaping to shape the combined input impedance of the plurality of tuned rectifier circuits as seen at the RF input port of the RF input network; and

a dc-dc converter system having an energy recovery input port and a dc output port, the energy recovery input port of the dc-dc converter system coupled to the output port of at least one rectifier to convert rectified power to a useable dc output voltage, wherein the voltage at the energy recovery input port of the dc-dc converter system is regulated.

2. The energy recovery system of claim **1**, wherein:

the rectification system includes a single rectifier output port that carries a combined dc signal of the plurality of tuned rectifier circuits, the single rectifier output port being coupled to the energy recovery input port of the dc-dc converter system.

3. The energy recovery system of claim **2**, wherein:

the dc-dc converter system comprises a dc-dc converter that regulates its input voltage.

4. The energy recovery system of claim **3**, wherein:

the dc-dc converter is configured so that the regulated input voltage of the dc-dc converter changes based, at least in part, on an RF input power level of the rectification system.

5. The energy recovery system of claim **2**, further comprising:

at least one additional RF input network and at least one additional rectification system coupled to the energy recovery input port of the dc-dc converter system.

6. The energy recovery system of claim **2**, wherein:

the dc-dc converter system comprises a three-or-more port dc-dc converter that includes a first port whose voltage is regulated by the dc-dc converter and which receives an output signal of the rectification system, a second port to output recovered energy to a load device, and a third port to provide power to the dc-dc converter for use in regulating the output voltage of the second port.

7. The energy recovery system of claim **6**, wherein:

the third port of the dc-dc output converter is coupled to an energy buffer for use in storing energy to regulate the output voltage of the second port.

8. The energy recovery system of claim **2**, wherein:

the dc-dc converter system comprises a first dc-dc converter having an input port and an output port, the first dc-dc converter input port coupled to the single rectifier output port to convert rectified power to a useable dc output voltage at the first dc-dc converter output port; and

a second dc-dc converter having an output port coupled to an input port of the first dc-dc converter to regulate an input voltage thereof.

9. The energy recovery system of claim **8**, wherein:

the first and second dc-dc converters are both switched-mode converters.

10. The energy recovery system of claim **8**, wherein:

one of the first and second dc-dc converters is a linear regulator and the other of the first and second dc-dc converters is a switched-mode converter.

11. The energy recovery system of claim **1**, wherein:

the rectification system includes multiple rectifier output ports that each correspond to a subgroup of the plurality of tuned rectifier circuits; and

the dc-dc converter system comprises multiple dc-dc converters having regulated input voltages coupled to the multiple rectifier output ports to receive dc signals from the multiple rectifier output ports and covert them to one or more useable dc output voltages.

12. The energy recovery system of claim **1**, wherein:

the RF input network includes a resistance compression network (RCN) to compress the resistance of the combined input impedance of the plurality of tuned rectifier circuits as seen at the RF input port of the RF input network.

13. The energy recovery system of claim **12**, wherein:

the resistance compression network has an order of 2 or greater.

14. The energy recovery system of claim **12**, wherein:

the resistance compression network is a lossless or near lossless network.

15. The energy recovery system of claim **14**, wherein:

the resistance compression network includes a pair of complementary reactances each having a terminal connected to a common potential.

16. The energy recovery system of claim **12**, wherein:

the resistance compression network includes one or more transmission line sections.

17. The energy recovery system of claim **1**, wherein:

the plurality of tuned rectifier circuits includes at least one tuned rectifier circuit that is tuned to provide an approximately resistive input impedance.

18. A machine implemented method for use in capturing energy output by a radio frequency (RF) circuit, the method comprising:

receiving an RF signal from the RF circuit;

distributing the RF signal among a plurality of tuned rectifier circuits, the plurality of tuned rectifier circuits to each generate a direct current (dc) signal at an output thereof in response to the RF input signal;

providing impedance shaping in coupling the input of the RF circuit to the input impedances of the plurality of tuned rectifier circuits to provide an approximately constant, resistive input impedance to the RF circuit; and

regulating an output voltage of the plurality of tuned rectifier circuits.

19. The method of claim **18**, further comprising:

applying one or more dc signals output by the plurality of tuned rectifier circuits to one or more dc-dc output converters, at least one of which has a regulated input voltage, to convert the one or more dc signals to one or more useable dc output voltages.

20. The method of claim **18**, wherein:

providing impedance shaping includes providing resistance compression to compress the input impedances of the plurality of tuned rectifier circuits to provide the approximately constant, resistive input impedance to the RF circuit.

21. The method of claim **18**, further comprising:

combining output signals of the plurality of tuned rectifier circuits to generate a combined output signal;

delivering the combined output signal to a dc-dc output converter to convert the combined output signal to a useable dc output voltage; and

regulating an input voltage of the dc-dc output converter.

22. The method of claim **18**, wherein:

distributing the RF signal among a plurality of tuned rectifier circuits includes delivering a portion of the RF signal to a first tuned rectifier circuit that is tuned to provide an approximately resistive input impedance.

* * * * *

US007268517B2

(12) **United States Patent**
Rahmel et al.

(10) **Patent No.: US 7,268,517 B2**
(45) **Date of Patent: *Sep. 11, 2007**

---

(54) **METHOD AND SYSTEM FOR ENERGY RECLAMATION AND REUSE**

(75) Inventors: **Ronald S. Rahmel**, Encinitas, CA (US); **Gerald C. Gerace**, Chula Vista, CA (US)

(73) Assignee: **Science Applications International Corporation**, San Diego, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **11/107,841**

(22) Filed: **Apr. 18, 2005**

(65) **Prior Publication Data**

US 2005/0186994 A1 Aug. 25, 2005

Related U.S. Application Data

(63) Continuation of application No. 09/670,613, filed on Sep. 27, 2000, now Pat. No. 6,882,128.

(51) **Int. Cl.**
H04B 1/16 (2006.01)

(52) **U.S. Cl.** **320/101**; 455/572; 455/343.1; 455/101

(58) **Field of Classification Search** 455/572-573, 455/343.1, 343.5, 343.6; 320/101-102, 116-123, 320/134
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,434,678 A		3/1969	Brown et al.	244/1
3,852,755 A		12/1974	Works et al.	343/701
3,927,375 A		12/1975	Lanoe et al.	325/363
4,031,449 A		6/1977	Trombly	320/2
4,079,268 A		3/1978	Fletcher et al.	307/151
4,360,741 A		11/1982	Fitzsimmons et al.	307/151
4,628,299 A		12/1986	Tate et al.	340/540
4,685,047 A	*	8/1987	Phillips, Sr.	363/126
5,043,739 A	*	8/1991	Logan et al.	343/701
5,068,669 A	*	11/1991	Koert et al.	343/700 MS
5,416,486 A		5/1995	Koert et al.	342/42
5,440,300 A		8/1995	Spillman, Jr.	340/825.54
5,446,447 A	*	8/1995	Carney et al.	340/572.4
5,495,519 A		2/1996	Chen	379/58
5,659,237 A	*	8/1997	Divan et al.	320/119
5,659,465 A	*	8/1997	Flack et al.	363/71
5,710,988 A		1/1998	Suzuki et al.	455/90
5,777,453 A	*	7/1998	Imanaga	320/160
5,842,140 A	*	11/1998	Dent et al.	455/573
5,842,838 A		12/1998	Berg	417/331

(Continued)

OTHER PUBLICATIONS

Davies, "Operation of Information Satellites in an interference environment," Abstract, *Microwave Journal*, vol. 15, pp. 33-36, 1972. ("Microwave Journal").

(Continued)

Primary Examiner—Simon Nguyen
(74) *Attorney, Agent, or Firm*—Kilpatrick Stockton LLP

(57) **ABSTRACT**

The present invention relates to a system and method for harvesting ambient electromagnetic energy, and more particularly, to the integration of antennas and electronics for harvesting ubiquitous radio frequency (RF) energy, transforming such electromagnetic energy into electrical power, and storing such power for usage with a wide range of electrical/electronic circuits and modules.

44 Claims, 13 Drawing Sheets

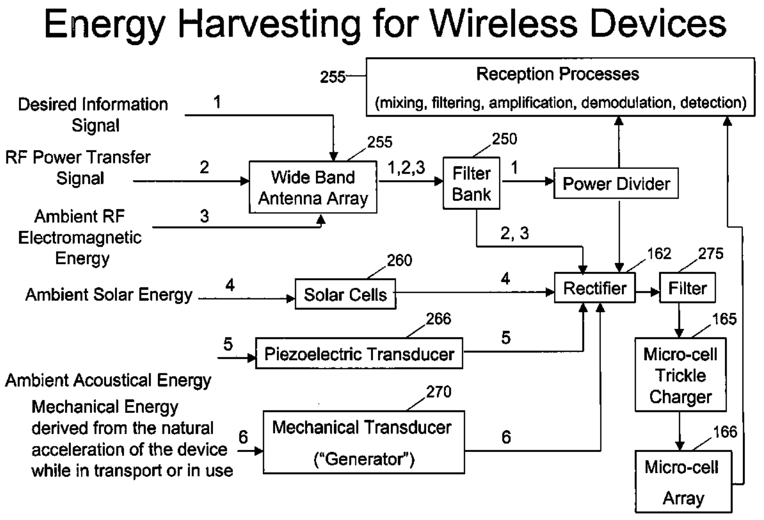
Energy Harvesting for Wireless Devices
255
Reception Processes
(mixing, filtering, amplification, demodulation, detection)
Desired Information Signal
1
RF Power Transfer Signal
2
255
Wide Band Antenna Array
1,2,3
250
Filter Bank
1
Power Divider
Ambient RF Electromagnetic Energy
3
2, 3
162
275
Ambient Solar Energy
4
260
Solar Cells
4
Rectifier
Filter
165
266
5
Piezoelectric Transducer
5
Micro-cell Trickle Charger
Ambient Acoustical Energy
Mechanical Energy derived from the natural acceleration of the device while in transport or in use
6
270
Mechanical Transducer ("Generator")
6
166
Micro-cell Array

U.S. PATENT DOCUMENTS

5,889,383	A	*	3/1999	Teich	320/107
5,898,932	A		4/1999	Zurlo et al.	455/573
5,970,393	A		10/1999	Khorrami et al.	455/66
6,037,743	A	*	3/2000	White et al.	320/103
6,045,339	A		4/2000	Berg	417/332
6,121,752	A	*	9/2000	Kitahara et al.	320/122
6,127,799	A	*	10/2000	Krishnan	320/104
6,140,924	A	*	10/2000	Chia et al.	340/572.5
6,154,664	A		11/2000	Chorey et al.	455/571
6,172,479	B1	*	1/2001	Barton	320/121
6,184,651	B1		2/2001	Fernandez et al.	320/108
6,184,789	B1	*	2/2001	Richley et al.	340/571
6,198,250	B1		3/2001	Gartstein et al.	320/112
6,259,372	B1	*	7/2001	Taranowski et al.	340/683
6,275,010	B1		8/2001	Neuteboom	320/137
6,289,237	B1	*	9/2001	Mickle et al.	600/509
6,329,796	B1	*	12/2001	Popescu	320/134
6,369,759	B1		4/2002	Epp et al.	343/700 MS
6,426,917	B1		7/2002	Tabanou et al.	367/82
6,597,465	B1		7/2003	Jarchow et al.	358/1.12
6,609,013	B1		8/2003	Oh et al.	455/561
6,847,834	B1	*	1/2005	Leem	455/572
6,882,128	B1	*	4/2005	Rahmel et al.	320/101
2002/0001745	A1	*	1/2002	Gartstein et al.	429/61

OTHER PUBLICATIONS

Dickinson, "Satellite Power System (SPS) Microwave Subsystem Impacts and Benefits," Abstract, *Report No. NASA-CR-157951*, 1977. ("Dickinson SPS").

Tonelli et al., "Design and analysis of a 5000-MW GaAlAs satellite power system," Abstract, *Proceedings of the 12th Intersociety Energy Conversion Engineering Conference Part II*, pp. 1412-1420, 1977. ("IECE 1").

Tonelli et al., "Design and evaluation of a 5 GW GaAlAs solar power satellite (SPS)," Abstract, *Proceedings of the 13th Intersociety Energy Conversion Engineering Conference Part I*, pp. 156-161, 1978. ("IECE 2").

Hughes, "The interactions of a solar power satellite transmission with the ionosphere and troposphere," Abstract, *AGARD Conference Proceedings No. 284. Propagation Effects in Space/Earth Paths*, p. 38, 1980. ("AGARD").

Maynard, "Initial MPTS study results: Design considerations and issues," Abstract, *Johnson Space Center Solar Power Satellite Microwave Transmission and Reception*, pp. 28-32, 1980. ("MPTS").

Noreen et al., "The system view," Abstract, *The Deep Space Network: A Radio Communication Instrumentation for Deep Space Exploration*, p. 25, 1983. (Deep Space Network).

Naumenko et al., "Trajectories of relativistic particles moving in free space," Abstract, *Radiotekhnika I Electronika*, vol. 29, No. 9, pp. 144-145, 1984. ("Relativistic Particles").

Yegerov, "The efficiency of energy transmission by an electromagnetic beam and its conversion in a rectenna," translated in *Soviet Journal of Communications Technology & Electronics*, vol. 30, No. 8, pp. 35-41, 1985. ("Energy Transmission").

Bianco, "Measurement Techniques for Evaluating Power System Interference to Instrument Landing Systems (ILS)," Abstract, *Canadian Electrical Association Contract/Grant Nos. CEA100T219 and CEA100T219A*, 1985. ("ILS").

Machina et al., "Power beaming—Energy transmission at 35 GHz and higher frequencies (for terrestrial and space applications with rectennas)," Abstract, *International Communication Satellite Systems Conference and Exhibit, 14th, Technical Papers, Part 3*, pp. 1671-1674, 1992. (Power Beaming).

Waldron, "Rectenna options for a lunar-solar power system," Abstract, *26th International Conference on Environmental Systems*, 1996. ("Rectenna Options").

Keerthi, "Analysis of the multistage CM array for digital communication signals," Abstract, *Conference Record of the Thirtieth Asilomar Conference on Signals, Systems and Computers*, vol. 2, No. 23, p. 1416, 1997. ("Asilomar").

Takeshi et al., "Experimental study of microwave power transmission efficiency," Abstract, *IEIC Technical Report, Institute of Electronics, Information and Communication Engineers*, vol. 98, No. 474, pp. 21-24, 1998. ("IEIC").

Vasudev, "Telecom explosion drives RF ICs," *Electronic Engineering Times, No. 1016*, p. 81, 1998. ("RF ICs").

Takyuki et al., "Effective utilization of electromagnetic energy. Effective utilization of electromagnetic energy in ITS (4). Radiocommunication in turnpike automatic charge collection system," Abstract, *Denji Kankyo Kogaku Joho EMC*, vol. 11, No. 9, pp. 38-45, 1999. ("ITS").

Ohr, "Integrated, one-chip cell phone is still a stretch, panelists concur," *Electronic Engineering Times*, No. 1067, p. 64, 1999. ("One-chip cell phone").

* cited by examiner

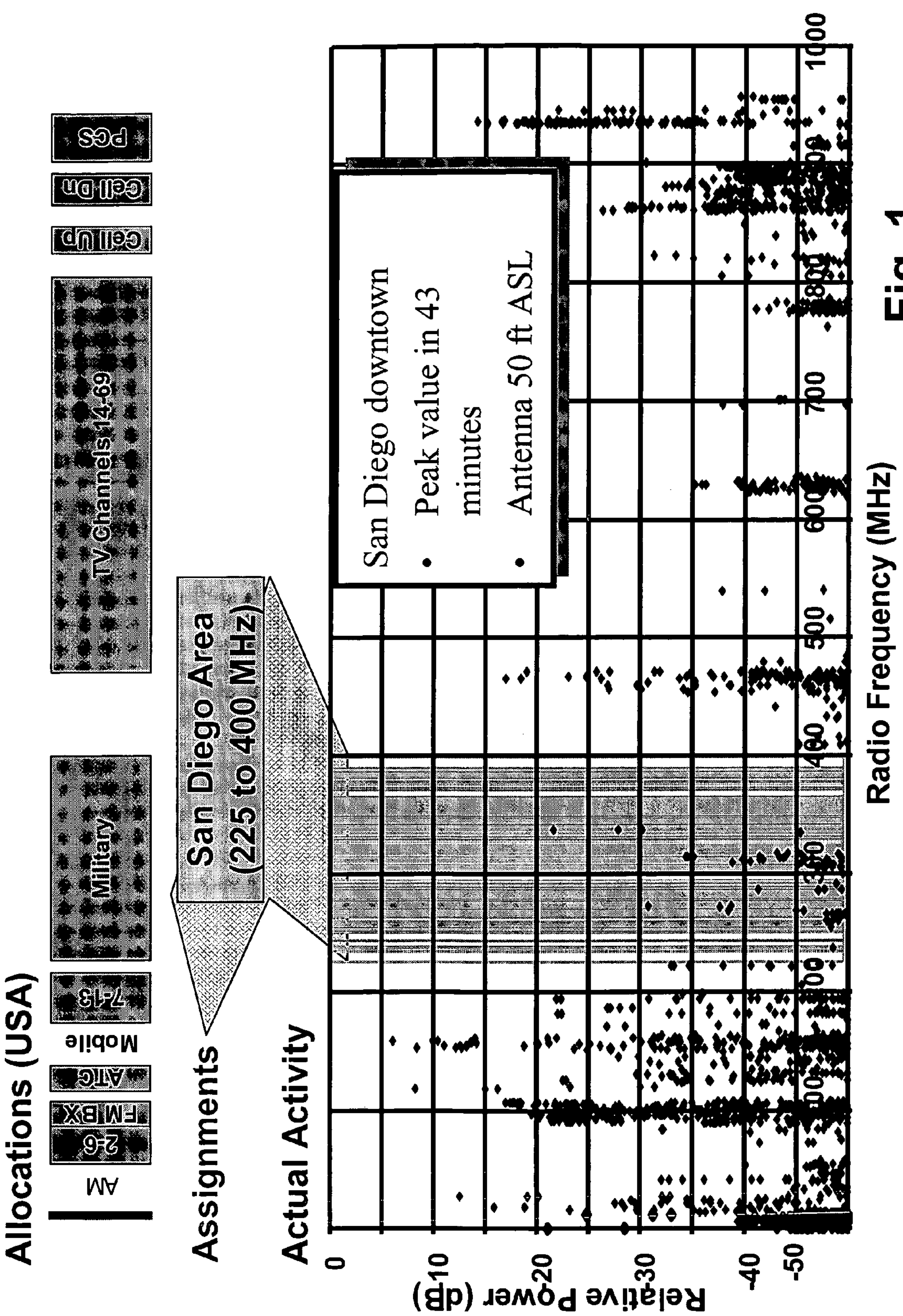
Allocations (USA)
AM
2-6
FM BX
ATC
Mobile
7-13
Military
TV Channels 14-69
Cell Up
Cell Dn
PCS
Assignments
San Diego Area
(225 to 400 MHz)
Actual Activity
San Diego downtown
• Peak value in 43 minutes
• Antenna 50 ft ASL
Relative Power (dB)
0
-10
-20
-30
-40
-50
Radio Frequency (MHz)
100
200
300
400
500
600
700
800
900
1000

Fig. 1

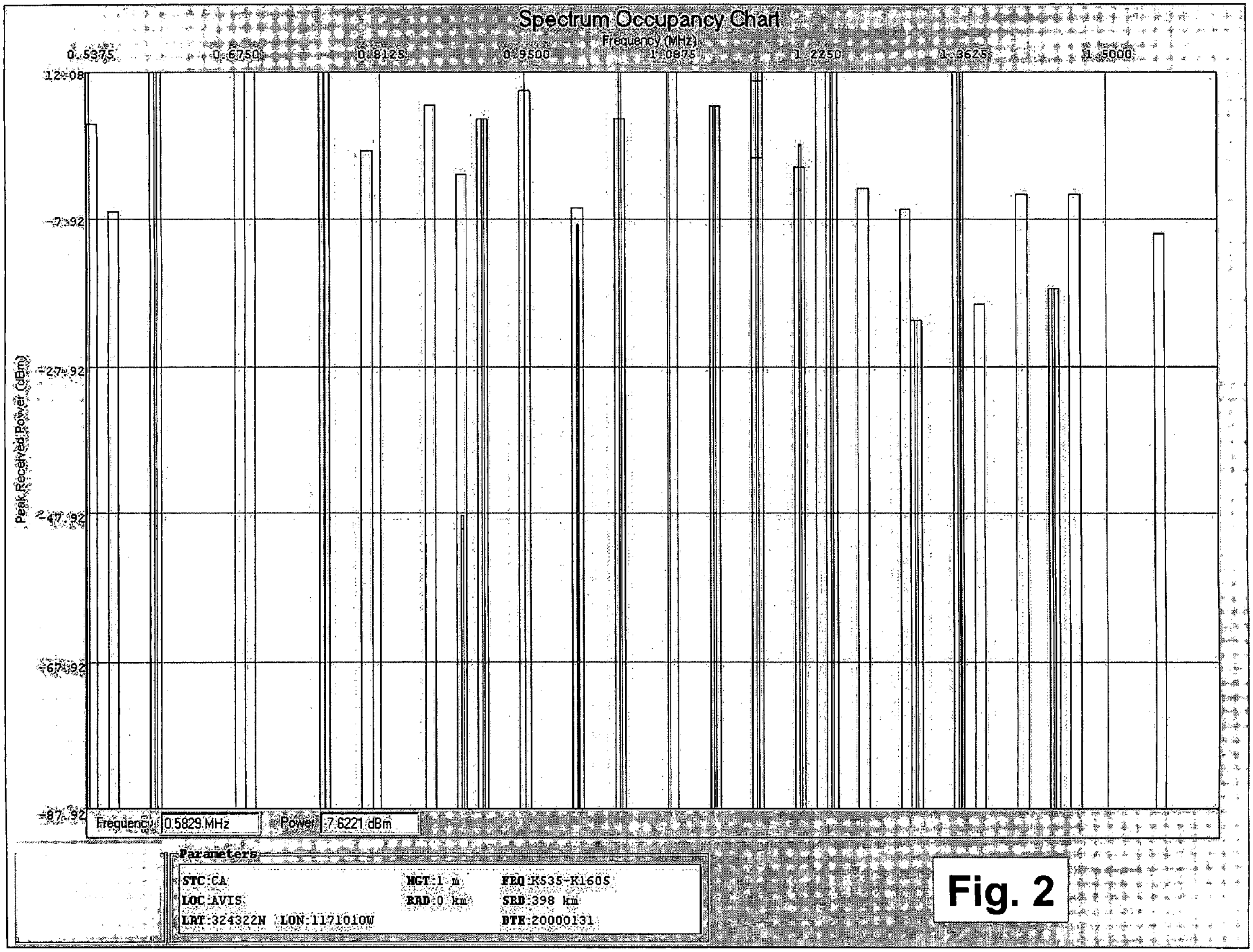
Spectrum Occupancy Chart
Frequency (MHz)
0.5375
0.6750
0.8125
0.9500
1.0875
1.2250
1.3625
1.5000
Peak Received Power (dBm)
12.08
-7.92
-27.92
-47.92
-67.92
-87.92

Frequency 0.5829 MHz Power 7.6221 dBm

Parameters

STC:CA **HGT**:1 m **FRQ**:K535-K1605
LOC:AVIS **RAD**:0 km **SRD**:398 km
LAT:324322N **LON**:1171010W **DTE**:20000131

Fig. 2

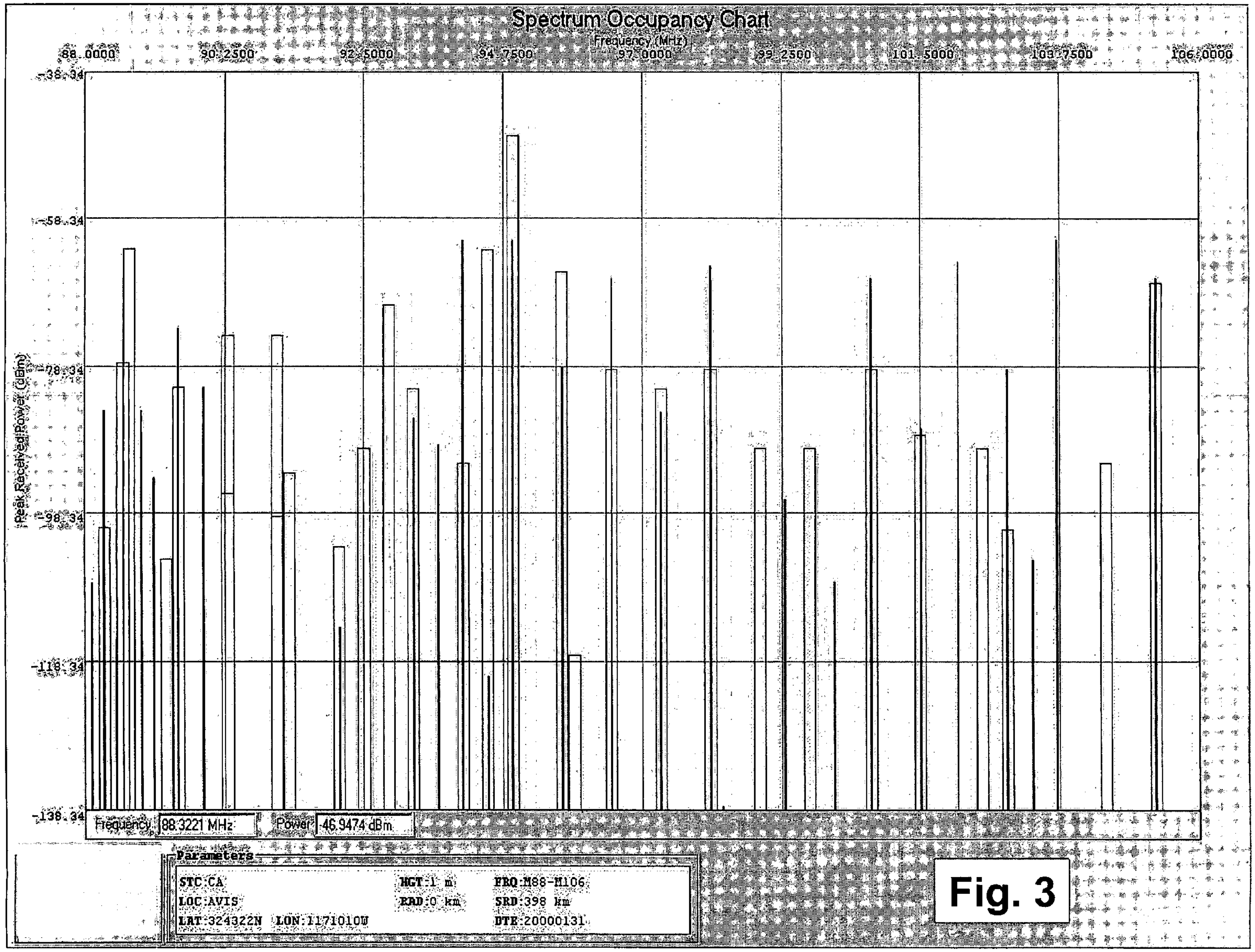
Spectrum Occupancy Chart
Frequency (MHz)
88.0000
90.2500
92.5000
94.7500
97.0000
99.2500
101.5000
103.7500
106.0000
Peak Received Power (dBm)
-38.34
-58.34
-78.34
-98.34
-118.34
-138.34
Parameters

Frequency 88.3221 MHz Power 46.9474 dBm

STC:CA HGT:1 m FRQ:M88-M106
LOC:AVIS RAD:0 km SRD:398 km
LAT:324322N LON:1171010W DTE:20000131

Fig. 3

ERS System Concept

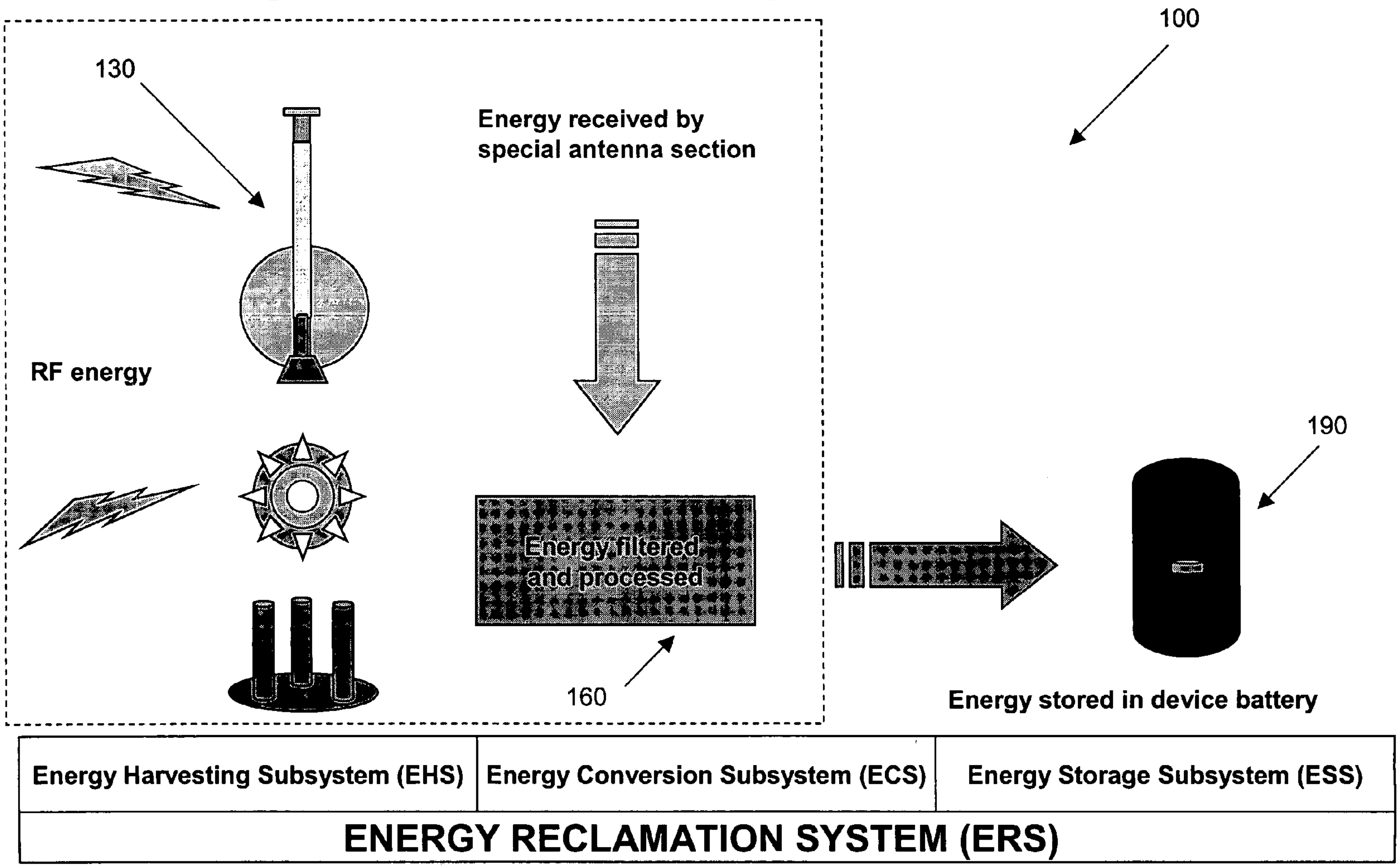
130
Energy received by
special antenna section
RF energy
Energy filtered
and processed
160
100
190
Energy stored in device battery
Energy Harvesting Subsystem (EHS)
Energy Conversion Subsystem (ECS)
Energy Storage Subsystem (ESS)
ENERGY RECLAMATION SYSTEM (ERS)

Fig. 4

Antenna Design

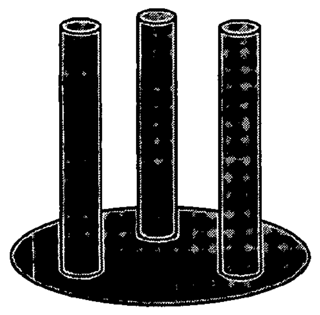

Fig. 5A

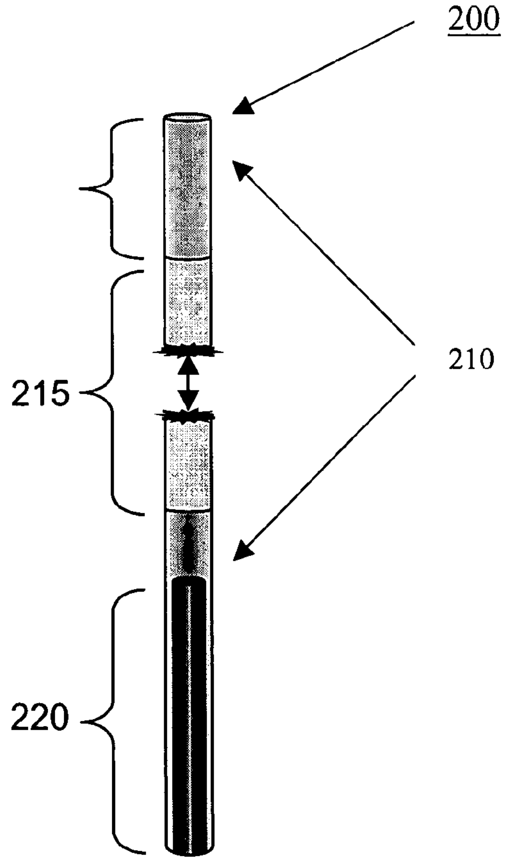
200
210
215
220

Fig. 5B

Alternate Antenna Design

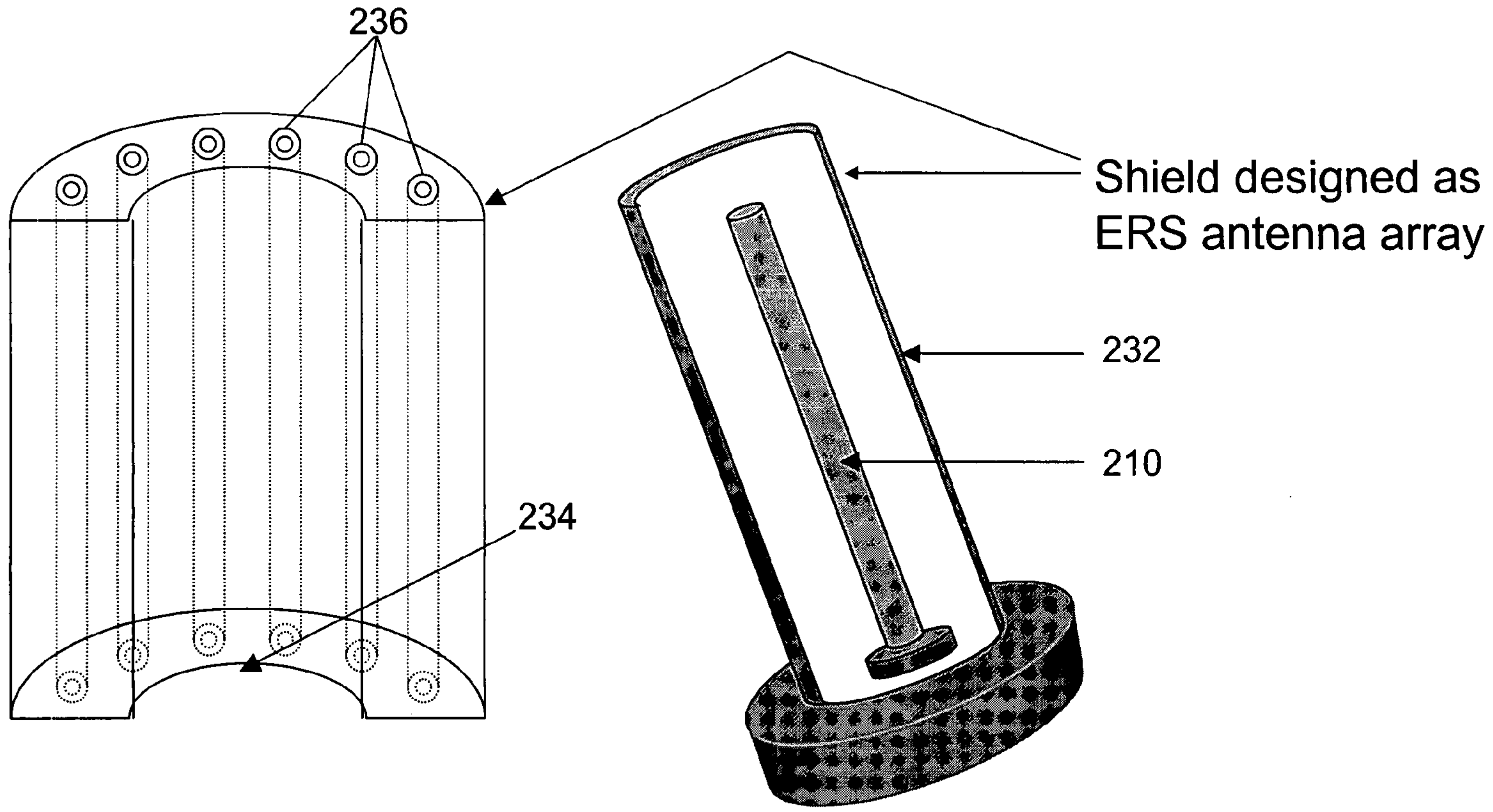
236
Shield designed as
ERS antenna array
232
210
234

Fig. 6

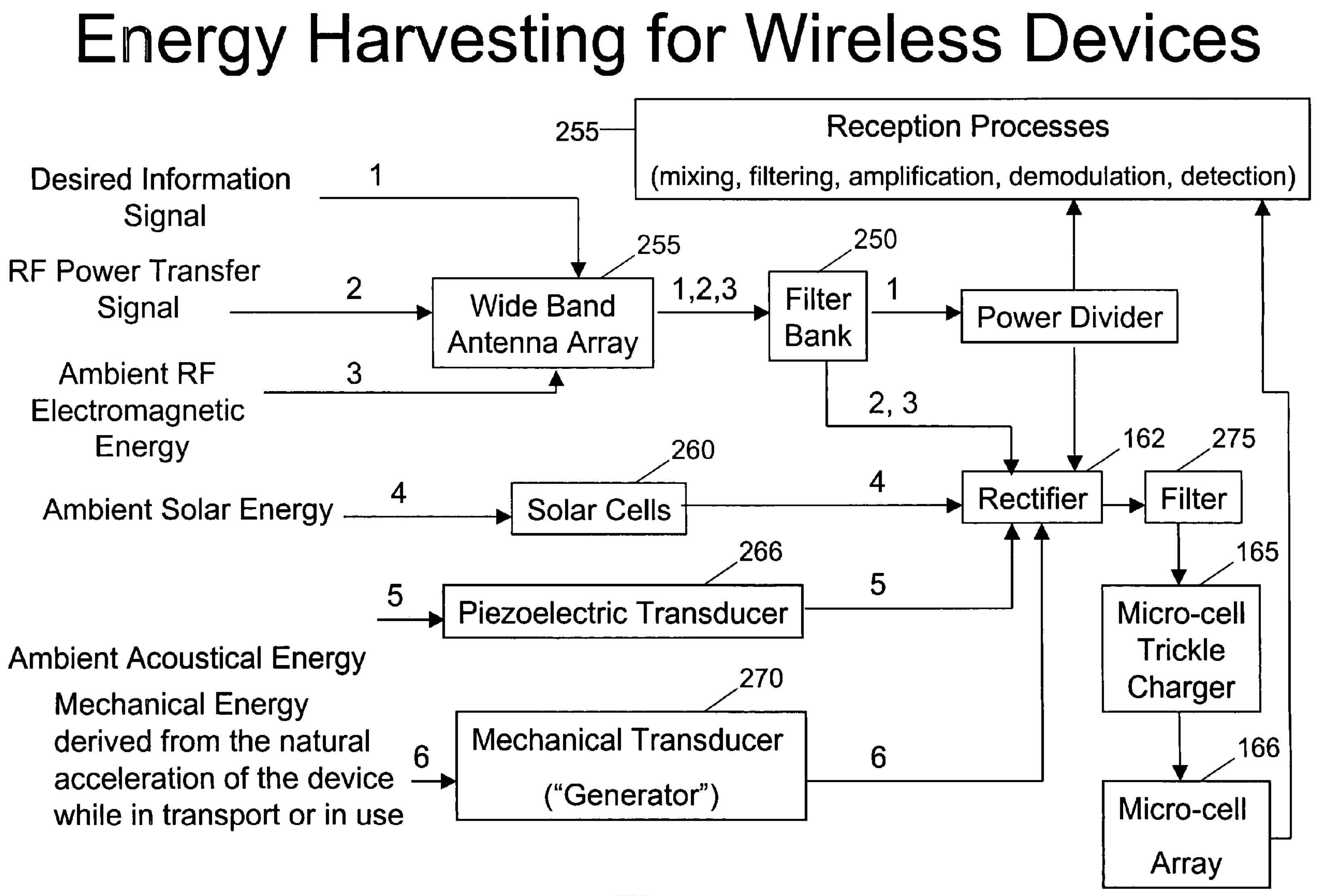
Energy Harvesting for Wireless Devices
255
Reception Processes
(mixing, filtering, amplification, demodulation, detection)
Desired Information Signal
1
RF Power Transfer Signal
2
Ambient RF Electromagnetic Energy
3
255
Wide Band Antenna Array
1,2,3
250
Filter Bank
1
Power Divider
2, 3
162
275
Rectifier
Filter
260
Ambient Solar Energy
4
Solar Cells
4
266
5
Piezoelectric Transducer
5
165
Micro-cell Trickle Charger
Ambient Acoustical Energy
270
Mechanical Energy derived from the natural acceleration of the device while in transport or in use
6
Mechanical Transducer
("Generator")
6
166
Micro-cell Array

Fig. 7

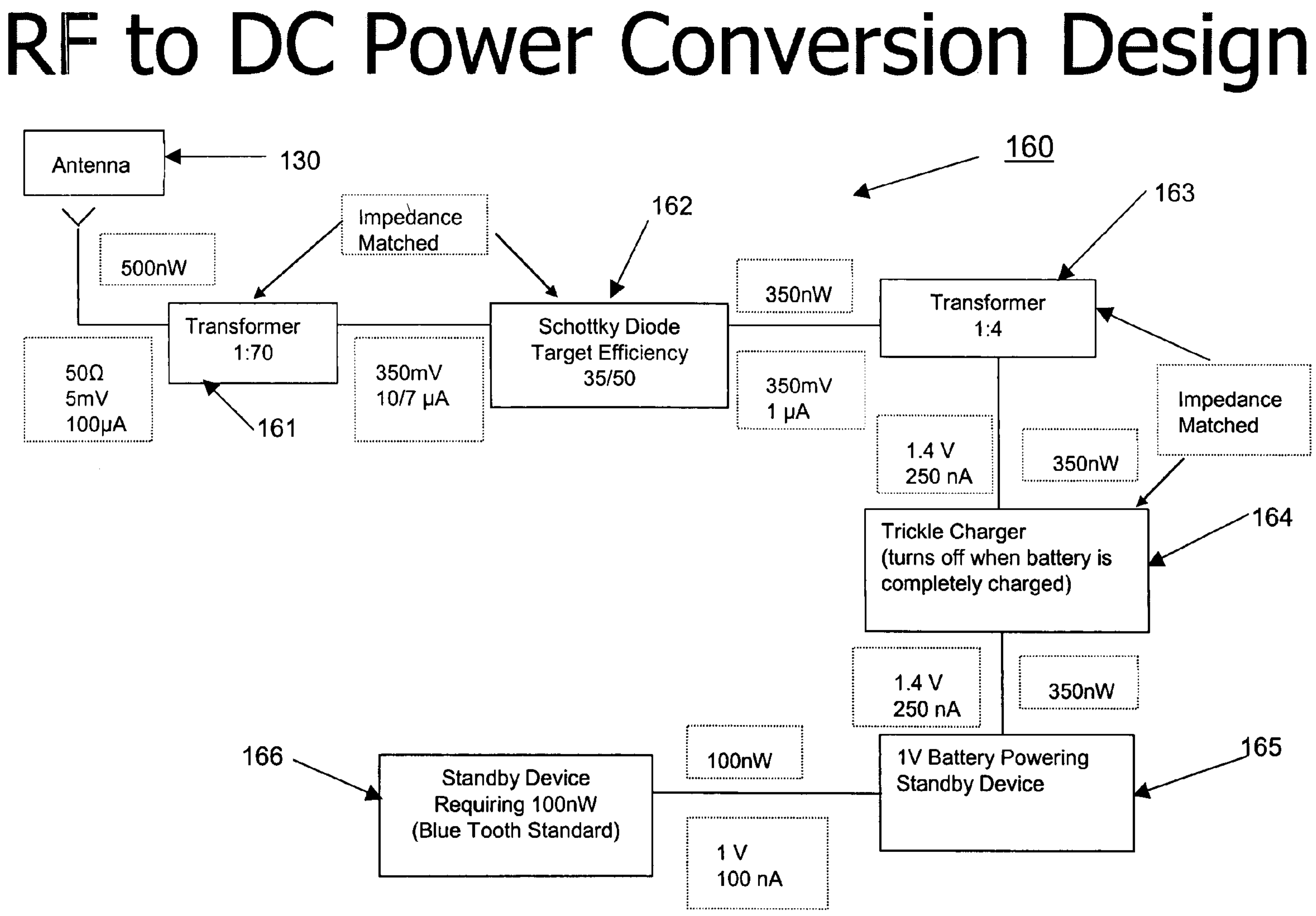
RF to DC Power Conversion Design
160
Antenna
130
500nW
50Ω
5mV
100µA
Impedance
Matched
Transformer
1:70
161
350mV
10/7 µA
162
Schottky Diode
Target Efficiency
35/50
350nW
350mV
1 µA
163
Transformer
1:4
Impedance
Matched
1.4 V
250 nA
350nW
164
Trickle Charger
(turns off when battery is
completely charged)
1.4 V
250 nA
350nW
165
1V Battery Powering
Standby Device
100nW
1 V
100 nA
166
Standby Device
Requiring 100nW
(Blue Tooth Standard)

Fig. 8

RF to DC Power Conversion/Trickle Charge/Power Storage Design Concept

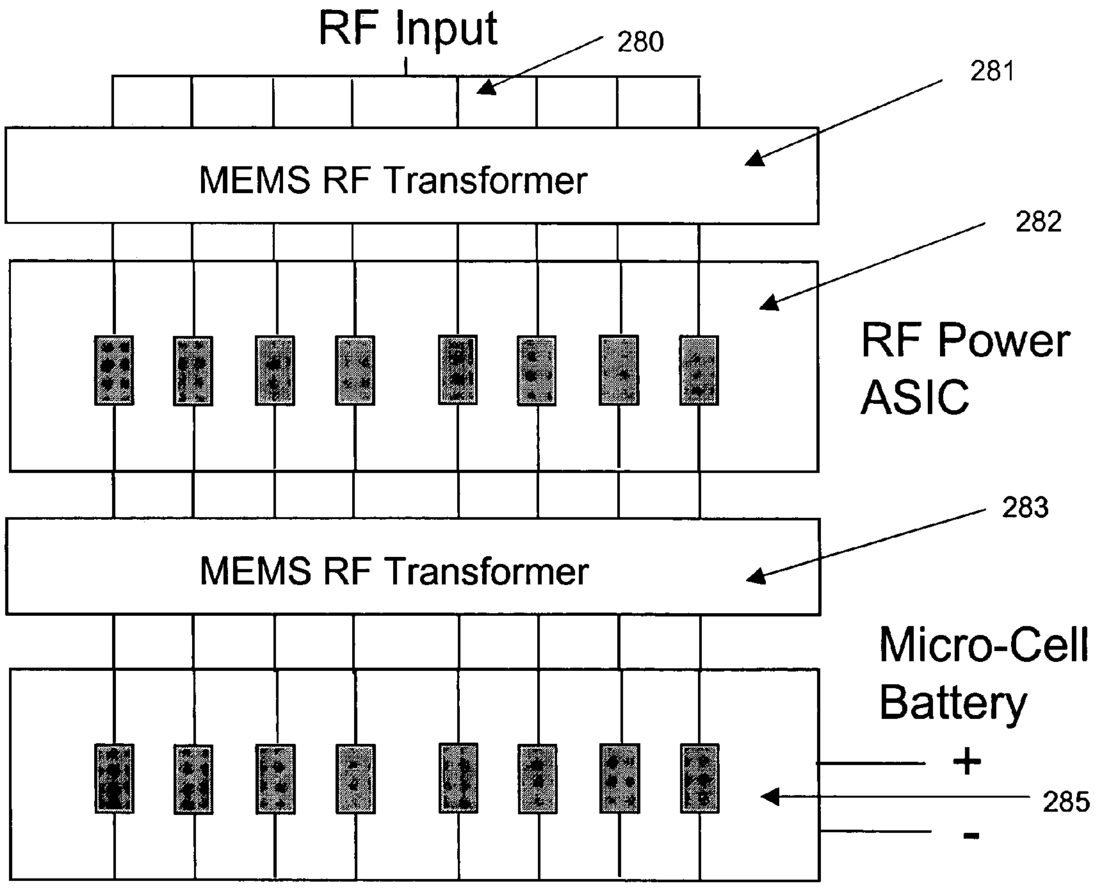
RF Input
280
281
MEMS RF Transformer
282
RF Power ASIC
283
MEMS RF Transformer
Micro-Cell Battery
+
285
-

Fig. 9

Battery Micro-cell Charging (Trickle Charge)

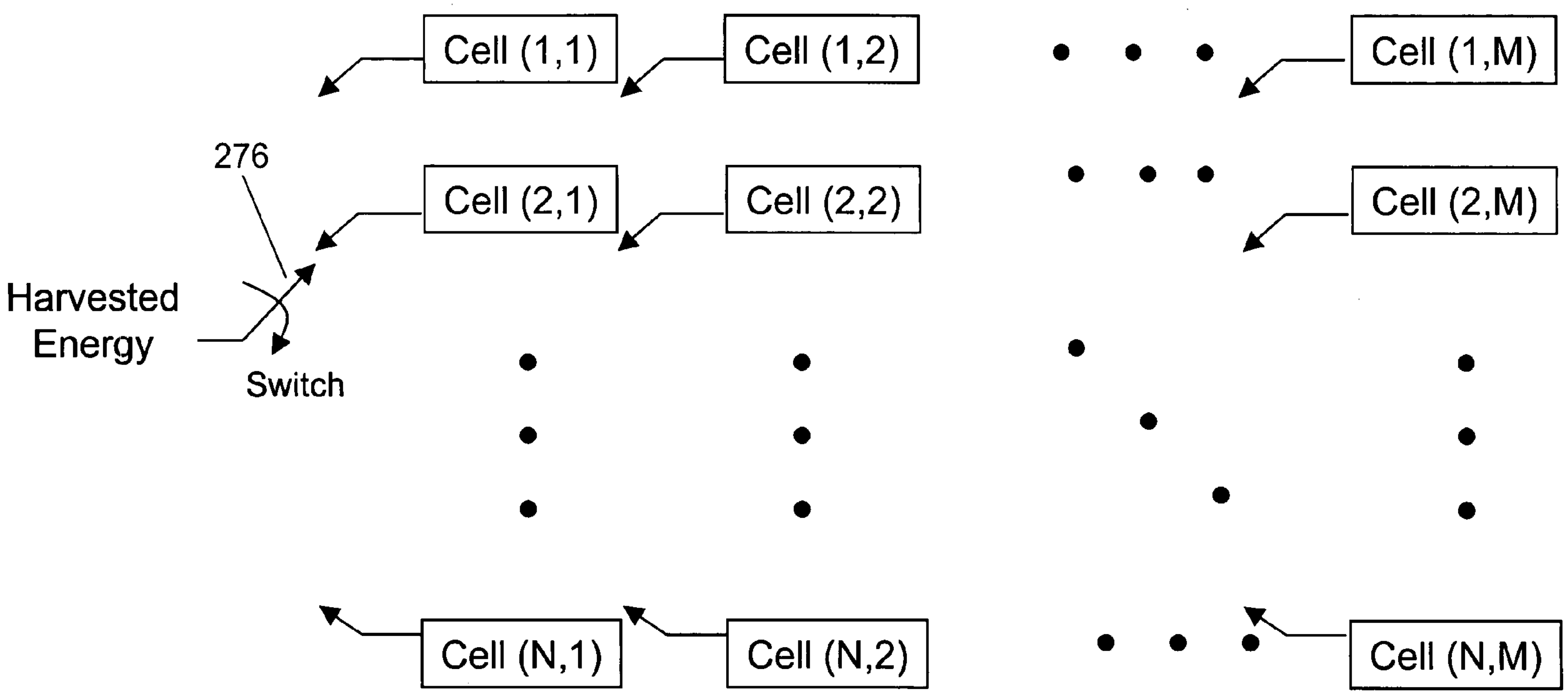
Cell (1,1)
Cell (1,2)
Cell (1,M)
Cell (2,1)
Cell (2,2)
Cell (2,M)
Cell (N,1)
Cell (N,2)
Cell (N,M)
276
Harvested Energy
Switch

Fig. 10

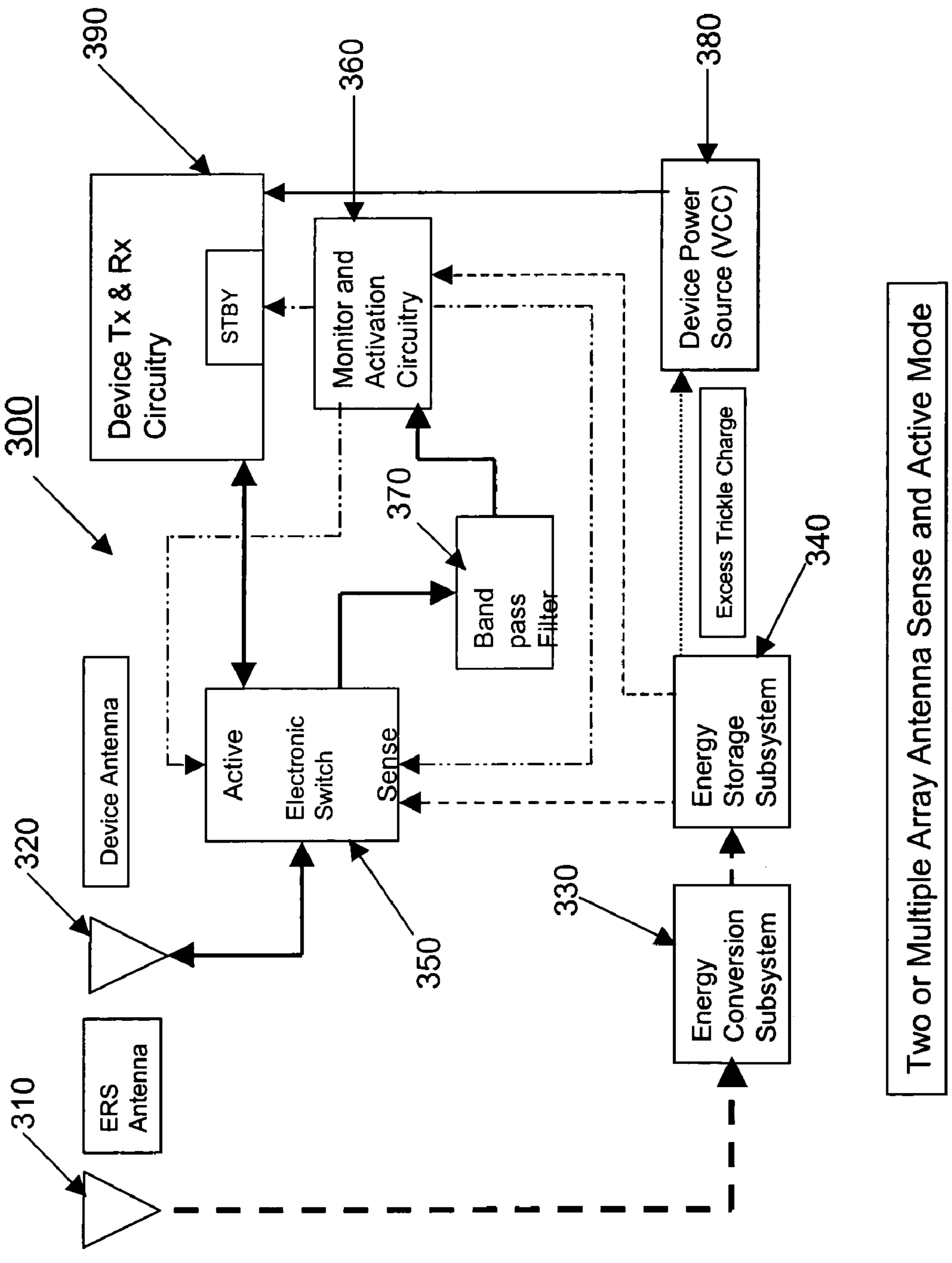
300
390
Device Tx & Rx Circuitry
STBY
360
Monitor and Activation Circuitry
380
Device Power Source (VCC)
370
Band pass Filter
Excess Trickle Charge
340
Device Antenna
Active
Electronic Switch
Sense
Energy Storage Subsystem
320
350
330
Energy Conversion Subsystem
310
ERS Antenna
Two or Multiple Array Antenna Sense and Active Mode

Fig. 11

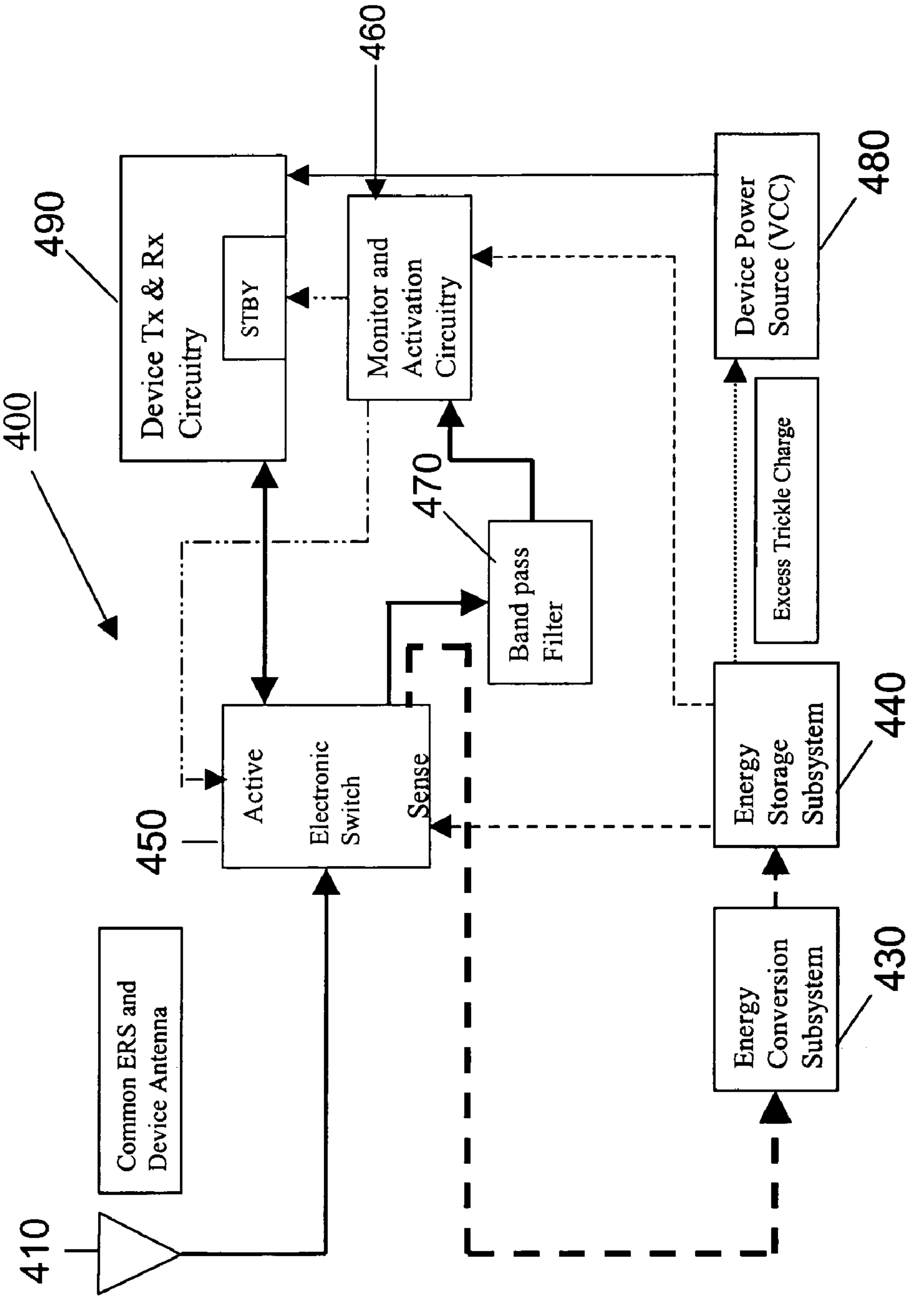
400
490
Device Tx & Rx Circuitry
STBY
460
Monitor and Activation Circuitry
470
Band pass Filter
480
Device Power Source (VCC)
Excess Trickle Charge
450
Active
Electronic Switch
Sense
440
Energy Storage Subsystem
430
Energy Conversion Subsystem
Common ERS and Device Antenna
410

Common Antenna used for Simultaneous Energy Harvesting, Sense and Active Mode

Fig. 12

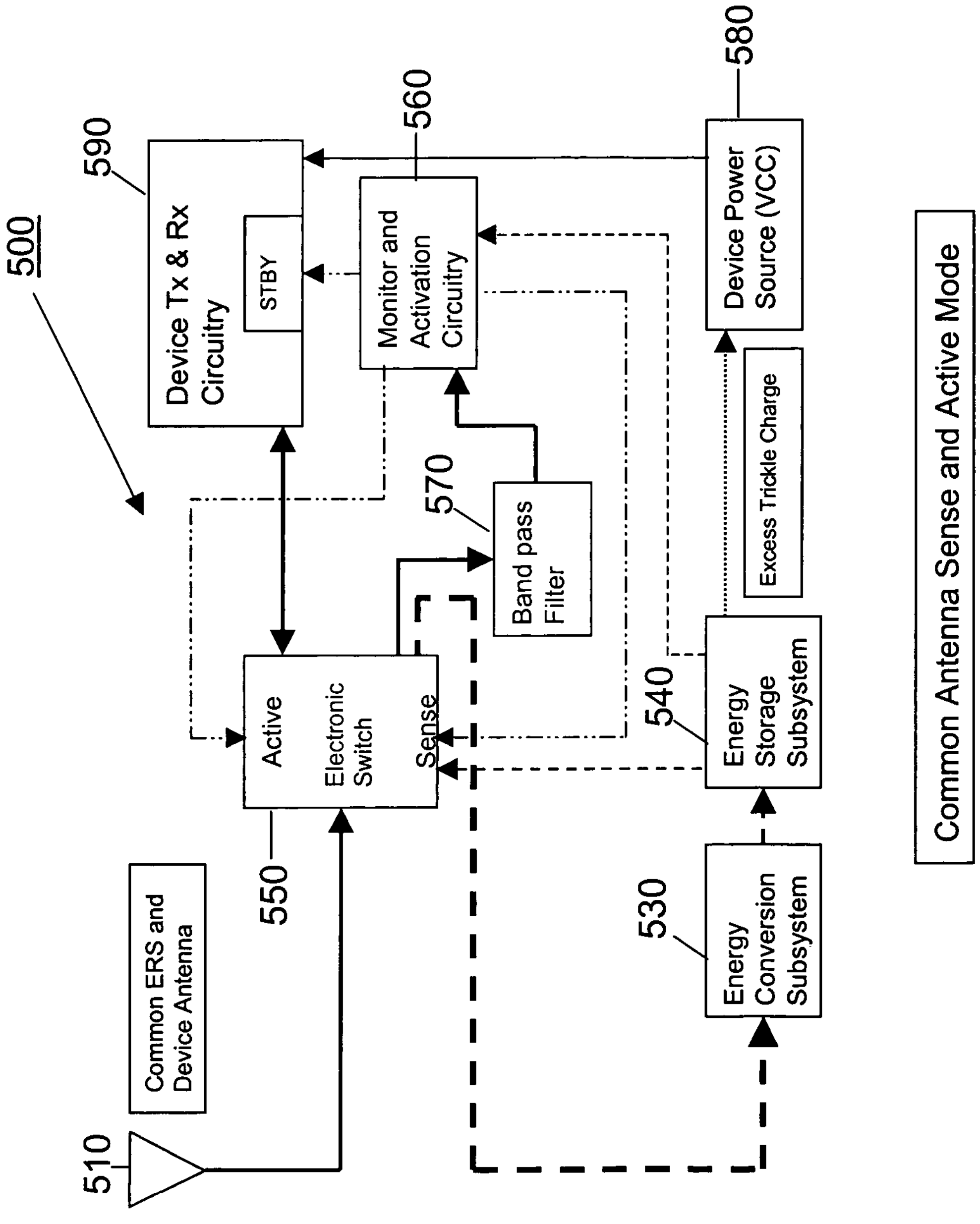
500
590
Device Tx & Rx Circuitry
STBY
560
Monitor and Activation Circuitry
580
Device Power Source (VCC)
570
Band pass Filter
Excess Trickle Charge
540
Energy Storage Subsystem
550
Active
Electronic Switch
Sense
Sense
530
Energy Conversion Subsystem
510
Common ERS and Device Antenna
Common Antenna Sense and Active Mode

Fig. 13

METHOD AND SYSTEM FOR ENERGY RECLAMATION AND REUSE

CROSS-REFERENCE TO RELATED APPLICATIONS

The current application is a continuation application of U.S. application Ser. No. 09/670,613 filed Sep. 27, 2000, which issued as U.S. Pat. No. 6,882,128.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates generally to wireless energy harvesting methods and systems. Specifically, the preferred embodiments of the present invention relate to the integration of antennas and electronics for harvesting ubiquitous electromagnetic energy, transforming that energy into useful power, and storing such power for usage.

2. Description of the Related Art

Typically, energy harvesting techniques and systems are focused on renewable energy such as solar energy, wind energy, and wave action energy from moving water. Solar energy is harvested by arrays of solar cells that convert radiant energy to DC power. Such energy is limited in low-light conditions. Cumbersome wind turbines that convert wind energy to electrical power depend on sufficient wind to spin the blades of the turbines to generate power. Wave action energy is harvested by water turbines or other apparatuses that are capable of converting the energy in the moving water into electrical power. This energy conversion is typically performed at complex hydroelectric power plants or by mechanical/electrical conversion apparatuses, such as those shown in U.S. Pat. Nos. 5,842,838 and 6,045,339 to Berg.

Another approach to energy harvesting is wireless power transmission via focused microwave signals. Unlike the above renewable energy technologies which harvest energy from natural sources, electrical energy of the wireless power transmission technology is remotely harvested and converted from focused microwave signals that are transmitted by a constructed source. For example, U.S. Pat. No. 5,043,739 to Logan et al. discloses a high frequency rectenna device for rectifying electromagnetic energy at microwave frequencies and higher and converting the energy into direct-current (DC) power to conduct wireless power transmission. Applications of this disclosure are focused on high-energy microwave radiation and high temperatures. One such application includes sending power to satellites or high altitude devices from the earth by electromagnetic waves. Another application involves collecting solar power from large space-based arrays of solar cells and transmitting the energy via electromagnetic waves to earth. A third application involves the direct conversion of microwaves generated by fusion reactions into a direct current.

BRIEF SUMMARY OF THE INVENTION

Currently, there is a large amount of renewable electromagnetic energy available in the radio frequency (RF) bands. For instance, AM radio, FM radio, TV, very high frequency (VHF), ultra high frequency (UHF), global system for mobile communications (GSM), digital cellular systems (DCS) and especially the personal communication system (PCS) bands are all lucrative energy sources. These untapped sources of energy grow more and more abundant as a result of the rapid growth in the wireless communication business. Hence, there is a desire to seek new methods and means of collecting the ubiquitous electromagnetic energy (ambient RF noise and signals) opportunistically present in the environment and transforming that energy into useful electrical power.

The preferred embodiments of the present invention can collect ubiquitous ambient radio frequency (RF) electromagnetic energy, convert it to a form useful for powering electronic devices, and store it for reuse. Thus, perpetual self contained subsystems or modules can be realized by using the reclaimed RF energy to power active antenna components, low power circuits, modules and other components such as transceivers, amplifiers, filters, etc. Hence, the preferred embodiments of the present invention can be applicable for any antenna or antenna array, for any antenna physical design (e.g., dipole, monopole, parabolic shape or dish, Yagi, etc.) and any frequency band. In some applications, this technology could be used advantageously and conveniently to transfer power remotely, without a physical connection, using a RF energy source intended for that purpose. In other applications, this technology could be used to maintain devices in a standby sleep mode indefinitely without ancillary power source. Then an RF signal could be used as a triggering mechanism to activate the device. Finally, for low power devices, this technology could provide primary power to the device and eliminate or reduce the traditional battery charge cycle.

Accordingly, the preferred embodiments of the present invention provide an integration of classical antenna physical materials and construction concepts with semiconductor and microprocessor technology to create an antenna structure capable of transforming electromagnetic waves into DC power. This antenna method and system collects and processes RF energy for reuse while simultaneously performing its primary function, such as signal reception, for a specific application. Furthermore, by using either a single antenna element or an array of elements configured to receive a broad range of frequencies in the present invention, energy collection and antenna transmit/receive performance can be maximized and optimized at the same time.

The preferred embodiments of the present invention also provide an Energy Reclamation System (ERS) that collects, converts and stores energy for use from selected frequency bands over the entire electromagnetic spectrum. These selected frequency bands are not required to be contiguous and may based on the operational environment, the application, any physical size limitations, and the technology available to cost effectively integrate an ERS into a device or application. An ERS of the present invention is not restricted to a single energy source because technologies like micro electromechanical systems (MEMS) can be used to integrate the collection of other types of energy, such as electromagnetic energy, solar energy, acoustic energy and/or casual motion energy. Thus, energy diversity enhances an ERS' ability to provide power consistent with the demands of modern devices and applications.

The preferred embodiments of the present invention also provide a method and system for harvesting electromagnetic spectrum energy from unintentional power sources. Thus, the electromagnetic energy may be harvested opportunistically from the abundant ambient RF electromagnetic environment created by the emission of countless commercial RF sources typically present today.

The preferred embodiments of the present invention also provide a method and system for harvesting electromagnetic spectrum energy using efficient wideband omnidirectional antennas for maximizing the collection of RF energy. The antennas can be of any design such as monopole, dipole, Yagi, or contrawound.

The preferred embodiments of the present invention also provide an energy collection and conversion technology that may be used to transmit a control signal to select a device's operating mode (e.g., wake up call in the standby sleep mode). Incorporating this technology into some devices allows them to remain in the sleep mode indefinitely. Thus, the devices may use the continuously reclaimed energy to power the monitoring circuitry and initialize equipment activation as required.

The preferred embodiments of the present invention also provide a method and system for transferring low power levels from an intentional power source advantageously and conveniently. This may be done remotely without a physical connection by using a RF energy source with a focused or blanket radiation pattern and an ERS at the target device. This may be used for numerous applications, including autonomous sensor or communication type applications in any environment (e.g., ground, air, underwater) where the device may use the ERS as its power source backup power source. For example, for underwater sensor locations, energy may come from sonar or other types of energy sources that can propagate with tolerable losses through the water and/or from RF energy delivered to a remote antenna extended above the surface of the water. If remaining operational for extended periods of time requires additional power, the device can be wirelessly and remotely supplied with power for storage and future use.

The preferred embodiments of the present invention further provide a conversion of the energy collected or captured by the antenna or transducer into an electrical current. The electrical current is then converted into a form suitable for input to a selected storage device. For instance, a rectifier and a low pass filter can be used to convert the electrical current into direct current (DC). This power can then be fed into a trickle charger for charging a power storage device, such as batteries, in wireless and other electrical devices. The type of electrical filter and power conditioning used depends on the type of battery or power storage device and the particular process of trickle charging used.

The preferred embodiments of the present invention additionally provide a method and system for harvesting electromagnetic energy that is designed to efficiently collect and convert RF energy using application specific integrated circuit (ASIC) technology. Hence, supporting electronics are integrated with the antenna itself and include, for example, amplifiers, mini-transceivers, trickle chargers for charging micro-cell batteries, filters efficient rectifier design and implementation (with different types of diodes), micro electromechanical systems (MEMS) RF transformers, etc. Thus, embedded circuitries are built into the antenna elements and optimized for a frequency range compatible with any given antenna design.

Additional aspects and novel features of the invention will be set forth in part in the description that follows, and in part will become more apparent to those skilled in the art upon examination of the disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

The preferred embodiments are illustrated by way of example and not limited in the following figures, in which:

FIG. **1** shows data taken from measurements in the RF spectrum between 3 MHz and 1 GHz in downtown San Diego, Calif., USA;

FIG. **2** shows the results of calculations predicting the signal levels in the AM radio band in downtown San Diego, Calif., USA;

FIG. **3** shows the results of calculations predicting the signal levels in the FM radio band in downtown San Diego, Calif., USA;

FIG. **4** shows the energy reclamation system (ERS) concept in accordance to one embodiment of the present invention;

FIGS. **5**A and **5**B show a physical structure of an energy harvesting antenna and an array of such antennas in accordance to an embodiment of the present invention;

FIG. **6** shows a communication antenna having an insulated shield acting as both a safety shield to the user and an energy harvesting antenna array in accordance to an embodiment of the present invention;

FIG. **7** shows a high level schematic diagram of an Energy Harvesting Subsystem (EHS) in accordance to one embodiment of the present invention;

FIG. **8** shows an energy conversion subsystem (ECS) of a perpetual active antenna (PAA) system in accordance to one embodiment of the present invention;

FIG. **9** shows an application specific integrated circuit (ASIC) chip implementation of the ECS circuitry shown in FIG. **8**;

FIG. **10** shows an array of battery micro-cells of an energy storage subsystem in accordance to an embodiment of the present invention;

FIG. **11** shows a high level schematic diagram of the Two or Multiple Array Sense and Active Mode in accordance to an embodiment of the present invention;

FIG. **12** shows a high level schematic diagram of the Common Antenna Sense and Active Mode in accordance to an embodiment of the present invention;

FIG. **13** shows a high level schematic diagram of the Common Antenna for Simultaneous Energy Harvesting, Sense and Active Mode in accordance to an embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The preferred embodiments of the present invention represent a new technology area for the reclamation and reuse of Radio Frequency (RF) power present due to either design or opportunity in certain frequency bands. It was found that RF signals and noise are available in useful power levels in many frequency bands—particularly in the AM, FM, and TV bands as well as the global system for mobile communications (GSM), digital cellular systems (DCS) and communication system (PCS) bands. The abundance of energy available for harvesting is clearly apparent from the measurements of the ambient RF spectrum. For example, FIG. **1** shows one such measurement in downtown San Diego Calif. The measurement indicates the relative power levels of RF signals between 3 MHz and 1 GHz that can be harvested at this location. As can be seen from the figure, there are ample RF signals from which to reclaim energy. FIG. **2** shows the results of calculations predicting the signal levels in just the AM radio band in downtown San Diego. It shows that many of the AM signals approach or are greater than 1 mW (0 dBm). This is because many commercial AM radio transmitters radiate thousands of watts of power. FIG. **3** shows the results of calculations predicting the signal levels in just the FM radio band in downtown San Diego.

These calculations further indicate that there are ample opportunities for harvesting energy from the ambient electromagnetic environment.

Countless candidate devices can benefit from this new energy reclamation technology of the present invention. For instance, the technology can be used for wireless devices that are typically powered by exhaustible or rechargeable batteries. The present invention can be used to provide wireless devices located in certain environments a perpetual life in standby mode and extended normal operational life by using stored reclaimed energy as a power source. The Energy Reclamation System or ERS of the present invention can provide sufficient power (e.g., 0.1-1 mW) to perpetually operate low power wireless devices without external recharging. The ERS of the present invention can also able to provide higher power levels to meet higher demands placed on these wireless devices. According to one preferred embodiment of the present invention, the ERS of the present invention includes a common antenna designed for the energy abundant frequency bands with it's output connected to power conditioning or rectification circuitry appropriately designed for that frequency range. The conditioned or rectified power can then be supplied to a charger (e.g., a trickle charger) or distribution subsystem designed to meet the battery's or storage device's charging specification.

There are innumerable potential applications for the energy reclamation technology of the present invention. For instance, the power harvested from radio frequency (RF) signals can be used for recharging installed power sources, providing power directly to the devices, energizing a device to wake up, and providing emergency backup power for the devices. Countless potential users of this technology also exist. For instance, potential users of the ERS technology may be found in the cellular, personal communications services (PCS) and future generations of wireless communications device industry. Typical consumer wireless electronic products that can be enhanced by ERS technology include, but are not limited to, palm computers, personal digital assistants (PDAs), pocket phones, pagers, cameras, autonomous sensors, garage door openers, remote controls and auto security systems. The energy reclamation technology can also be used in emergency location systems such as GPS position finders, airplane and watercraft emergency beacons and fire alarm systems. The energy reclamation technology can also be applied to wireless personal devices such as video conferencing cell phones, Internet phones and Wireless Personal Area Networks (WPAN). The energy reclamation technology is also flexible enough to be adapted for future electrical and/or electronics applications and devices.

According to a preferred embodiment of the ERS of the present invention, the ERS includes three subsystems: an Energy Harvesting Subsystem (EHS), an Energy Conversion Subsystem (ECS), and an Energy Storage Subsystem (ESS). The Energy Harvesting Subsystem or EHS includes various components used to collect ambient electromagnetic energy via an antenna or antenna array and convert it to electrical current. If acoustic, solar, or motion is the energy source, the antenna or antenna array may be replaced by a transducer or transducer array for the respective energy source. The EHS may also include a combination of antennas and mechanical transducers to collect different types of energy at once or alternately. The Energy Conversion Subsystem or ECS includes circuits, components, and modules integrated together to convert and condition the noise like current out of the antenna or transducer into a form of energy suitable for storage. The Energy Storage Subsystem or ESS includes, for example, batteries, capacitors, and other such components that will store the reclaimed energy for future reuse. The ESS can also include appropriate circuitry (e.g. trickle charger) for charging those devices that the ERS supports.

FIG. **4** shows the ERS concept in accordance to the above preferred embodiment of the present invention. As shown, the ERS **100** includes three parts or subsystems: an Energy Harvesting Subsystem (EHS) **130**, a Power or Energy Conversion Subsystem (ECS) **160**, and a Power or Energy Storage Subsystem (ESS) **190**. The Energy Reclamation System provides for the collection and harvesting of ubiquitous ambient energy over a desired frequency range. In a preferred embodiment, the desired range is 500 KHz-2.4 GHz. For frequencies above 2.4 GHz and for intentionally-radiated high power harvesting applications, the ERS can incorporate the rectenna technology, such as that disclosed in U.S. Pat. No. 5,043,739, into the EHS and the ECS for connection to the ESS. The interconnection of the ESS to the rectenna technology is readily apparent to one skilled in the art.

According to another preferred embodiment of the present invention, the main objective of the ERS **100** is focused on the collection of ambient RF energy present in the ambient environment in the range of 5 watts and below and converting that energy into reusable energy to power numerous electronic devices. The converted energy can be full-wave or half-wave rectified and filtered current. The charging schemes depend on how the ECS **160** is used and its reuse requirements. The output of the ECS **160** can be distributed for reuse to several different types of circuitry. In one embodiment, the converted power may be fed to a trickle charger circuitry that charges a standby energy storage device in the ESS **190** that includes a battery, capacitor, etc. while any excess power is distributed to a charging circuitry for the primary power storage component. Another way to distribute the ECS output power is to feed it directly into circuitry appropriate for affecting indefinite operation of the electronic device in a standby or sleep mode.

The three subsystems of the ERS system are now described in further details.

Energy Harvesting Subsystem (EHS)

The EHS for RF energy collection includes a single antenna of any physical design and shape (e.g., dipole, Yagi, Contrawound Toroidal Helical Antenna, etc.) or an array of such antennas to collect RF energy from any frequency or bands of frequencies. Appropriate baluns (i.e., balanced to unbalanced converters) and impedance matching techniques may be integrated to maximize power transfer into the energy conversion circuitry of the ECS.

For the ERS to reclaim the maximum amount of energy, the energy-harvesting antenna is preferably designed to be a wideband, omnidirectional antenna or antenna array that has maximum efficiency at selected bands of frequencies containing the highest energy levels. If the EHS includes an array of antennas, each antenna in the array can be designed to have maximum efficiency at the same or different bands of frequency from one another. The collected RF energy is then converted into usable DC power using high speed switching semiconductor devices or a diode-type or other suitable rectifier. This power may be used to drive, for example, an amplifier/filter module connected to a second antenna system that is optimized for a particular frequency and application. Thus, one antenna acts as an energy harvester while the other antenna acts as a signal transmitter/receiver. Transforming and/or storing the power garnered by the ERS technology over time makes those devices employing such technology more efficient and extends battery life.

According to a preferred embodiment of the present invention, the antenna design and construction preferably begins with defining a wide band of frequencies over which collection will be required. Next, optimal length or element size is preferably calculated to achieve maximum efficiency over that frequency range. As is apparent to one skilled in the art, antenna efficiency depends on a number of factors related to, for example, the physical design and shape of the antenna. Other factors affecting the amount of energy collected include signal strength at the receiving location from the transmitting source, which in turn depends on the paths from the transmitting source to the ERS of the present invention.

Once the optimal length or element size of the antenna is calculated, it is subsequently reduced to a fraction of a wavelength while integrating semiconductor amplification and impedance matching circuits in the construction of the antenna elements. The antenna circuit elements are preferably constructed using microprocessor wafer manufacturing techniques and shaped for integration or encapsulation in the antenna elements. This preferred design approach reduces the physical size of the antenna while retaining the energy collection efficiencies of a larger antenna. Antenna efficiency of 50% or more is preferred, for higher antenna efficiency resulting in higher energy collection.

FIG. **5**A shows the physical structure of an energy-harvesting antenna and FIG. **5**B shows an array of such antennas in accordance to an embodiment of the present invention. The cylindrical physical design shape was selected for easy explanation of the design concept; however, any physical shape (e.g., toroid, etc.) may evolve that still meets the ERS design concepts and goals. As shown, the antenna **200** includes a standard metal outer sleeve **210**. Under the sleeve **210** is a semiconductor material **215** used to create an apparent physical length optimized for reception to accommodate size limitations. Enclosed by the antenna outer sleeve **210** and semiconductor material **215** is electronic circuitry **220**, which is constructed using microprocessor and ASIC semiconductor technology to develop integrated support modules such as trickle charger, amplifier and filter for the ECS, which will be described later. The outer sleeve **210** may also be used as a heat sink for the electronic circuitry **220**.

According to another preferred embodiment of the present invention, the antenna can also be constructed to comply with recent radiation hazard concerns with wireless and cellular telephone technologies and at the same time increases power collection potential. There are a number of ways to direct the RF energy away from the user or shield the user using cupped shields in the antenna system design. Such design could be optimized to provide an insulated shield and enhance energy collection materials in the antenna system construction. FIG. **6** shows an antenna **210** covered by an insulated shield **232**. The antenna **210** performs its primary function for a specific application, such as receiving RF signals for normal wireless communication. The insulated shield **232** can be used as an electromagnetic shield and also as an ERS antenna or antenna array for energy harvesting. As shown in the figure, the RF shielding material is preferably located on the inside surface **234** of the shield **232**. As is known in the art, the shielding material may be of any material suitable for shielding RF radiation generated by the transmitting antenna **210**. Within the shield **232** is an antenna array **236**, with associated integrated modules of the ECS, for energy harvesting. Alternatively, within the shield **232** is a single antenna for energy harvesting. Each antenna in array **236** or the alternative single antenna is depicted in FIG. **5**A.

FIG. **7** shows a high level schematic diagram of an Energy Harvesting Subsystem (EHS) in accordance to one embodiment of the present invention. The EHS includes a wideband omnidirectional antenna or antenna array **255** for energy harvesting, as described earlier. The antenna **255** may be used to harvest energy from ambient RF signals **3** and/or RF power transfer signal **2** from a remotely-located, intentional power source. It may also be used to concurrently receive the desired communication signal **1** along with the energy-harvesting signals **2** and **3**. The combined signals are then separated by filter bank **250** which will be described in greater detail later.

FIG. **7** also shows that other energy harvesting technologies for collecting ambient solar, acoustical, and mechanical energy can be integrated as part of the EHS to supply electrical energy into the ECS. The electrical energy supplied by these other energy harvesting technologies may interface with the rectifier **162** and/or filter circuitries **250**, which is then fed to filter **275** in the ECS. As apparent to one skilled in the art, the ECS interface depends on the input frequency, quality, quantity and type (constant DC or some form of pulsating DC) of energy supplied.

For ambient solar energy, an array of solar cells or a solar energy conversion device **260** known in the art (such as a MEMS solar/heat electric generator) is preferably used to absorb solar energy from the environment and convert it to electrical energy. This energy is then delivered to a rectifier **162** of the ECS. An array of solar cells may be used to produce sufficient power for a variety of applications.

For ambient acoustical energy, a piezoelectric transducer **266** known in the art can be used to absorb ambient or intentional sound from the environment, which in turn causes a crystal to vibrate and through the piezoelectric effect produce an output voltage. The output energy is then delivered to the rectifier **162** of the ECS. An array of transducers may be used to produce sufficient power for a variety of applications.

For mechanical energy, the energy is preferably converted to electric energy through a MEMS Mechanical Transducer (Generator) **270** known in the art. The MEMS Mechanical Transducer **240** takes mechanical energy derived from the natural acceleration of a thing or person while in transport (e.g., person walking) or in use. These devices can wind a spring, force a piston to move or use some other method to store and convert acceleration energy collected into electrical energy. The output energy is then delivered to the rectifier **162** of the ECS. An array of such transducers may be used to produce sufficient power for a variety of applications.

Energy Conversion Subsystem (ECS)

According to a preferred embodiment of the present invention, the ECS for RF energy collection includes a power charger and other circuitry for performing RF to DC power conversion. Because the energy harvesting in the present invention is focused mainly on recovering small amounts of energy over long periods of time, charging energy storage devices (rechargeable batteries, etc.) may be done over a period of time by trickle charge. As is known in the art, trickle charging is a charging scheme in which a power storage component is charged at a fraction of its capacity rate. FIG. **8** shows an ECS **160** for use in conjunction with the EHS to perform RF to DC power conversion, with specifications in accordance to an embodiment of the present invention.

The ECS **160** preferably includes a transformer **161** connected to an energy harvesting antenna **130** for receiving the RF energy. The transformer **161** is also connected to a diode rectifier circuit **162**. The transformer **161** and the rectifier circuit **162** are designed to match impedance with one another to prevent undesired energy loss between these two elements. After rectification, the converted energy is sent through another transformer **163** and onward to a trickle charger **164**. Again, the second transformer **163** and the trickle charger **164** are designed to match impedance with one another to prevent undesired energy loss between these two elements. The trickle charger **164** is then used to deliver power to a battery **165** for storage and use with an electronic device, such as a standby device **166**. The battery is actually a part of the ESS, which will be described later.

If an array of energy harvesting antennas is used to receive the RF energy, the power collected from all channels (i.e., the power density at all antenna inputs) can be combined, rectified and forwarded to the trickle charger **164**. The power available in each channel is comparable; thus, for example, if 100N nW is required for a load, the power collected in N channels at 100 nW each can be combined to generate the required power.

Although not shown in FIG. **8**, the converted and rectified power may be fed through a filter for signal purification prior to sending it to the transformer **163** and onward to the trickle charger **164**. FIG. **7** shows the use of such filter **275**, with like numbers for like elements shown in FIG. **8**. It should be noted that FIG. **7** depicts a broad overview of the ERS concept according to an embodiment of the present invention; therefore, the figure does not show all detailed components of the ECS as shown in FIG. **8**.

FIG. **9** shows an ASIC chip implementation of the ECS circuitry shown in FIG. **8**. The RF input **280** corresponds to the antenna **130** of FIG. **8** for receiving harvested RF energy; the MEMS RF transformer **281** corresponds to the transformer **161** of FIG. **8**; the RF power ASIC **282** corresponds to the rectifier circuit **162** and the trickle charger **164** of FIG. **8** and any other desired circuitry for the ECS, such as filters, impedance matching circuitry, etc.; and the micro-cell battery **285** corresponds to the battery **165** of FIG. **8**. The battery **165** or **285** is a part of the ESS, which is described next.

Energy Storage Subsystem (ESS)

According to a preferred embodiment of the present invention, the ESS includes a rechargeable battery **165**, which can be a complete battery or an array of micro-cells of battery. If battery **165** includes a complete battery, the trickle charger **164** preferably provides a larger potential difference between terminals and more power for charging during a period of time. If battery **165** includes individual battery cells, the trickle charger **164** preferably provides smaller amounts of power to each individual battery cell, with the charging proceeding on a cell by cell basis. FIG. **10** shows an array of battery micro-cells according to an embodiment of the present invention. The ESS charges one or more micro-cells as available energy allows. As the load requires power, the combined energy is drawn from a charged P×Q subset (sub-array) of the N×M set (array) of cells to power the device on standby or active mode. N, M, P and Q are all natural numbers (i.e., 1, 2, 3, . . .), wherein P is a number less than N, and Q is a number less than M. Charging of the remaining cells continues whenever ambient power is available. As the load depletes cells, switches such as switch **276** are used to replace the depleted cells with charged cells. The rotation of depleted cells and charged cells continues as required. As much energy as possible is banked to ensure adequate power is available on demand. Thus, energy is managed on a micro-cell basis.

According to the present invention, there are numerous ways in which the ERS technology can be configured to harvest RF energy. In one embodiment of the present invention, an electronic device, such as a cellular phone, is preferably designed to have two antennas, one for energy harvesting at a desired frequency band and one for performing RF communication at a different frequency band. An example of this dual-antenna design is previously shown in FIG. **6**. FIG. **11** shows a high level schematic diagram **300** of a dual-antenna design ERS in accordance to one embodiment of the present invention. As explained earlier, the frequency band at which an antenna receives and transmits signals depends on the physical dimensions of the antenna. Thus, the energy-harvesting antenna **310** and the communication antenna **320** are designed with different dimensions to receive signals at their respectively desired frequency bands. The energy-harvesting antenna **310** may comprise one antenna or an array of like antennas. In this embodiment, parallel operations are possible for energy harvesting and communication. In other words, a user can be using the phone for wireless communication in one frequency band, while at the same instance the energy harvesting antenna **310** and associated circuitry within the phone is collecting electromagnetic energy in another frequency band. The collected electromagnetic energy is passed to integrated circuitry that converts the RF energy into electrical energy. This electrical energy is then distributed for storage in batteries of the ESS **340** for use in powering various electrical components of the phone.

As shown in FIG. **11**, communication signals received by the device antenna **320** are sent to communication processing circuitry **390** for transmission and reception. The energy harvesting antenna **310**, on the other hand, is connected to an Energy Conversion Subsystem or ECS **330**, which is designed to match the impedance and frequency band or bands of the energy harvesting antenna **310** for maximizing input energy received. The ECS **330** is as described earlier with respect to FIG. **8**. Its output is conditioned power coming from a trickle charger designed to efficiently charge the power storage devices, such as NiCd, NiMH, or LiIon battery cells, in the energy storage subsystem or ESS **340**. Any excess conditioned power from the trickle charger of the ECS **330** is then distributed to the device's main power source **380**.

The ESS **340** is as described earlier with respect to FIGS. **8** and **10**. It stores the standby power and distributes it to the electronic switching circuitry **350** and the Monitor and Activation Circuitry **360** in both active and standby modes of the device. Alternatively, the ESS **340** may be used to power circuitries **350** and **360** in standby mode of the device, while the device power source **380** powers those circuitries in the device's active mode. The ESS **340** may also be used to complement the device power source **380**.

According to yet another embodiment of the present invention, the harvesting antenna **310** and associated circuitries **330** and **340** may be physically located outside of the cellular phone but yet capable of electrical coupling with the phone. For example, the harvesting antenna **310** and associated circuitries may be built into portable physical structures such as briefcases, suitcases, purses, and wearable apparel (cloths) or fixed physical structures by any known means in the art. Thus, when the phone is desired to be charged, it can be stored in any of the aforementioned physical structures and electrically coupled to the built-in harvesting antenna and circuitry for charging purposes. These separate structures facilitate the use of larger antennas (transducers) or arrays of smaller elements whichever is most efficient for collecting the ambient energy. In any case, the larger surface area intercepts more energy.

According to a further embodiment for harvesting RF energy in the present invention, an electronic device, such as a cellular phone, is designed to have two antennas that respectively harvest RF energy and perform wireless communications in the same desired frequency band. In this embodiment, the energy-harvesting antenna and the communication antenna are designed with the same physical dimensions. The communication antenna is connected to a signal level detection system for detecting the strength of RF signals in the desired frequency band. If the strength of the desired RF signals is sufficiently high to sustain both RF communication and energy harvesting, then the level detection system will cause the activation of the energy-harvesting antenna. Thus harvesting proceeds in parallel with communication and the RF energy is converted into electrical power useful for powering various components of the cellular phone without degrading phone service.

According to another embodiment for harvesting RF energy in the present invention, a single antenna or an array of like antennas in an electronic device, such as a cellular phone, is used for both energy harvesting and communication purposes. FIG. **12** shows a high level schematic diagram **400** of the Common Antenna for Simultaneous Energy Harvesting, Sense and Active Mode in accordance with this embodiment. This mode requires modifications to the device's antenna switching and filter circuitry to allow communication signals in along with collected RF energy for harvesting. The communication signals are filtered and directed to the receiver's communication circuitry while the noise or interference portions of the received RF signals are filtered out and sent to the energy harvesting circuitry within the electronic device. This noise filtering technique is illustrated back in FIG. **7**, where incoming communication signals **1** and the ambient RF energy **3** are filtered by a filter bank **250**, which may reside in the switching circuitry **450** shown in FIG. **12**. The communication signals **1** are then directed to the receiver's communication circuitry **255**, which is shown as communication processing circuitry **490** in FIG. **12**, while the noise or interference RF portions **3** are filtered out and sent to the energy harvesting circuitries, which include the ECS **430** and ESS **440**.

According to still another embodiment for harvesting RF energy in the present invention, a single antenna or an array of like antennas in an electronic device, such as a cellular phone, is again used for both energy harvesting and communication purposes. FIG. **13** shows a high level schematic diagram **500** of the Common Antenna Sense and Active Mode in accordance with this embodiment. Here, the single antenna or antenna array **510** is connected to the energy harvesting circuitries **530** and **540** and communication circuitry **590** by a switch **550**, which is slaved to the various powering modes of the electronic device. For example, when the phone is in the power-on mode, the switch **550** is programmed to connect the antenna **510** with the communication circuitry **590** to receive the RF communication signals and enable wireless communication with the phone. However, when the phone is in the standby mode, the switch is programmed to connect the antenna **510** with the harvesting circuitries, which then converts the electromagnetic energy in the same RF communication signals into electrical power for powering the electronic device in the standby mode.

The normal operation of an electronic device having the ERS technology as depicted in FIGS. **11** and **12** may also include a standby or sleep mode, as described above with regard to FIG. **13**. When the electronic device is set in a standby mode, the electronic switching circuitry (**350** or **450**) of the ERS (**300** or **400**) within the electronic device is placed in a Sense Mode. As shown in both FIGS. **11** and **12**, the monitor and activation circuitry (**360** or **460**) is used to set the electronic switching circuitry (**350** or **450**) in a sense mode. In this mode, signals received by the device's communication antenna (**320** or **410**) are sent to a band pass filter (**370** or **470**) to retrieve an activation signal, which is passed on to the monitor and activation circuitry (**360** or **460**) that is watching for the activation signal. The activation signal can be anything defined by the device's manufacturer, such as a trigger signal or an emergency beacon for activating a desired system. When a turn-on signal brings the electronic device out of standby mode, the monitor and activation circuitry (**360** or **460**) sends an active signal to the electronic switching circuitry (**350** or **450**) to connect the communication antenna (**320** or **410**) to the device's communication circuitry (**390** or **490**). The device's turn-on signal may be manually provided by a user turning on the electronic device, or it may be brought on by the reception of the activation signal by the monitor and activation circuitry (**360** or **460**).

Again, the harvesting circuitries in the above embodiments may be physically located apart from the communication circuitry. For example, the harvesting circuitries may be built into portable structures such as briefcases, suitcases, handbags, and purses by any known means in the art, while the communication circuitries and the switch are built into the phone itself. Thus, when the phone is desired to be charged, it can be stored in any of the aforementioned portable structures and electrically coupled to the harvesting circuitries built into the structure for charging purpose.

Although only a few exemplary embodiments of this invention have been described in detail above, those skilled in the art will readily appreciate that many modifications are possible in the exemplary embodiments without materially departing from the novel teachings and advantages of this invention. Accordingly, all such modifications are intended to be included within the scope of this invention as defined in the following claims. Furthermore, any means-plus-function clauses in the claims (invoked only if expressly recited) are intended to cover the structures described herein as performing the recited function and all equivalents thereto, including, but not limited to, structural equivalents, equivalent structures, and other equivalents.

The invention claimed is:

1. An energy reclamation device, comprising:

a radio frequency (RF) energy collector for receiving ambient RF energy;

a solar energy collector for receiving ambient solar energy;

a vibration energy collector for receiving ambient vibration energy;

at least one conversion circuit for converting the ambient RF energy, the ambient solar energy, and the ambient vibration energy to electrical power; and

a power storage for storing the electrical power;

wherein the RF energy collector comprises one antenna to receive RF signals and wherein the received RF signals are sent to the at least one conversion circuit only when the energy reclamation device is in standby mode.

2. The energy reclamation device of claim **1**, wherein the at least one antenna is designed to receive ambient RF energy in a selected frequency range.

3. The energy reclamation device of claim **1**, wherein the RF energy collector comprises a plurality of antennas.

4. The energy reclamation device of claim **3**, wherein an antenna in the plurality of antennas transmits and receives RF signals used for communications.

5. The energy reclamation device of claim **1**, wherein the ambient RF energy comprises ambient RF signals.

6. The energy reclamation device of claim **1**, wherein the solar energy collector comprises an array of solar cells.

7. The energy reclamation device of claim **1**, wherein the vibration energy collector comprises a piezoelectric transducer.

8. The energy reclamation device of claim **1**, wherein the ambient vibration energy comprises ambient acoustic energy.

9. The energy reclamation device of claim **1**, further comprising a mechanical energy collector for receiving ambient mechanical energy and using the at least one conversion circuit to convert the ambient mechanical energy to electrical power.

10. The energy reclamation device of claim **9**, wherein the mechanical energy collector comprises a mechanical transducer.

11. The energy reclamation device of claim **9**, wherein the ambient mechanical energy is derived from acceleration of the energy reclamation device.

12. The energy reclamation device of claim **1**, wherein at least one conversion circuit comprises a rectifier.

13. The energy reclamation device of claim **1**, wherein the at least one conversion circuit is formed on an application specific integrated circuit (ASIC) chip that is integrated with the RF energy collector, solar energy collector, and the vibration energy collector.

14. The energy reclamation device of claim **1**, wherein the power storage comprises a battery.

15. The energy reclamation device of claim **14**, wherein the battery comprises a plurality of battery micro-cells.

16. The energy reclamation device of claim **15**, further comprising a trickle charger for charging the battery micro-cells on a cell by cell basis.

17. The energy reclamation device of claim **1**, wherein the power storage comprises a plurality of batteries.

18. The energy reclamation device of claim **1**, wherein the electrical power is equal to or less than five watts.

19. The energy reclamation device of claim **1**, wherein the RF antenna to receives RF signals comprising ambient RF signals and RF signals used for communications.

20. The energy reclamation device of claim **19**, further comprising a filter for filtering the received RF signals into an interference portion and a communication portion.

21. The energy reclamation device of claim **19**, wherein the interference portion of the received RF signals is sent to the at least one conversion circuit and the communication portion of the received RF signals is not sent to the at least one conversion circuit.

22. An energy reclamation apparatus, comprising:

a base component;

a radio frequency (RF) energy collecting component for receiving ambient RF energy;

a solar energy collecting component for receiving ambient solar energy;

a vibration energy collecting component for receiving ambient vibration energy;

at least one conversion component for converting the ambient RF energy, the ambient solar energy, and the ambient vibration energy to electrical power used to power the energy reclamation apparatus;

a power storage component for providing additional electrical power to the energy reclamation apparatus if the energy reclamation apparatus requires more electrical power than is converted by the at least one conversion component; and

wherein the RF energy collecting component, the solar energy collecting component, vibration energy collecting component, the at least one conversion component, and the power storage component are mounted onto the base component;

wherein the RF energy collecting component comprises one antenna to receive RF signals and wherein the received RF signals are sent to the at least one conversion component only when the energy reclamation device is in standby mode.

23. The energy reclamation apparatus of claim **22**, further comprising a switching component for receiving an activation signal indicating whether the energy reclamation apparatus is in power-on mode or standby mode.

24. The energy reclamation apparatus of claim **23**, further comprising an activation component for receiving the activation signal from the switching component.

25. The energy reclamation apparatus of claim **22**, wherein the at least one antenna is designed to receive ambient RF energy in a selected frequency range.

26. The energy reclamation apparatus of claim **22**, wherein the RF energy collecting component comprises a plurality of antennas.

27. The energy reclamation apparatus of claim **26**, wherein an antenna in the plurality of antennas transmits and receives RF signals.

28. The energy reclamation apparatus of claim **22**, wherein the ambient RF energy comprises ambient RF signals.

29. The energy reclamation apparatus of claim **22**, wherein the solar energy collecting component comprises an array of solar cells.

30. The energy reclamation apparatus of claim **22**, wherein the vibration energy collecting component comprises a piezoelectric transducer.

31. The energy reclamation apparatus of claim **22**, wherein the ambient vibration energy comprises ambient acoustic energy.

32. The energy reclamation apparatus of claim **22**, further comprising a mechanical energy collecting component for receiving ambient mechanical energy and using the at least one conversion component to convert the ambient mechanical energy to electrical power.

33. The energy reclamation apparatus of claim **32**, wherein the mechanical energy collecting component comprises a mechanical transducer.

34. The energy reclamation apparatus of claim **32**, wherein the ambient mechanical energy is derived from acceleration of the energy reclamation apparatus.

35. The energy reclamation apparatus of claim **22**, wherein at least one conversion component comprises a rectifier.

36. The energy reclamation apparatus of claim **22**, wherein the at least one conversion component is formed on an application specific integrated circuit (ASIC) chip that is mounted onto the base component.

37. The energy reclamation apparatus of claim **22**, wherein the power storage component comprises a battery.

38. The energy reclamation apparatus of claim **37**, wherein the battery comprises a plurality of battery micro-cells.

39. The energy reclamation apparatus of claim **38**, further comprising a trickle charging component for charging the battery micro-cells on a cell by cell basis.

40. The energy reclamation apparatus of claim **22**, wherein the power storage component comprises a plurality of batteries.

41. The energy reclamation apparatus of claim **22**, wherein the electrical power generated by the at least one conversion component is equal to or less than five watts.

42. The energy reclamation apparatus of claim **22**, wherein the RF energy collecting component receives RF signals comprising ambient RF signals and RF signals used for communications.

43. The energy reclamation apparatus of claim **42**, further comprising a filter component for filtering the received RF signals into an interference portion and a communication portion.

44. The energy reclamation apparatus of claim **43**, wherein the interference portion of the received RF signals is sent to the at least one conversion component and the communication portion of the received RF signals is not sent to the at least one conversion component.

* * * * *

US 20070240757A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2007/0240757 A1**

Ren et al. (43) **Pub. Date: Oct. 18, 2007**

(54) **SOLAR CELLS USING ARRAYS OF OPTICAL RECTENNAS**

(75) Inventors: **Zhifeng Ren**, Newton, MA (US); **Krzysztof Kempa**, Billerica, MA (US); **Yang Wang**, Allston, MA (US)

Correspondence Address:
GREENBERG TRAURIG, LLP
ONE INTERNATIONAL PLACE, 20th FL
ATTN: PATENT ADMINISTRATOR
BOSTON, MA 02110 (US)

(73) Assignee: **The Trustees of Boston College**

(21) Appl. No.: **11/250,932**

(22) Filed: **Oct. 14, 2005**

Related U.S. Application Data

(60) Provisional application No. 60/619,262, filed on Oct. 15, 2004.

Publication Classification

(51) **Int. Cl.**
H01L 31/00 (2006.01)

(52) **U.S. Cl.** **136/256**

(57) **ABSTRACT**

The present invention discloses a solar cell comprising a nanostructure array capable of accepting energy and producing electricity. In an embodiment, the solar cell comprises an at least one optical antenna having a geometric morphology capable of accepting energy. In addition, the cell comprises a rectifier having the optical antenna at a first end and engaging a substrate at a second end wherein the rectifier comprises the optical antenna engaged to a rectifying material (such as, a semiconductor). In addition, an embodiment of the solar cell comprises a metal layer wherein the metal layer surrounds a length of the rectifier, wherein the optical antenna accepts energy and converts the energy from AC to DC along the rectifier. Further, the invention provides various methods of efficiently and reliably producing such solar cells.

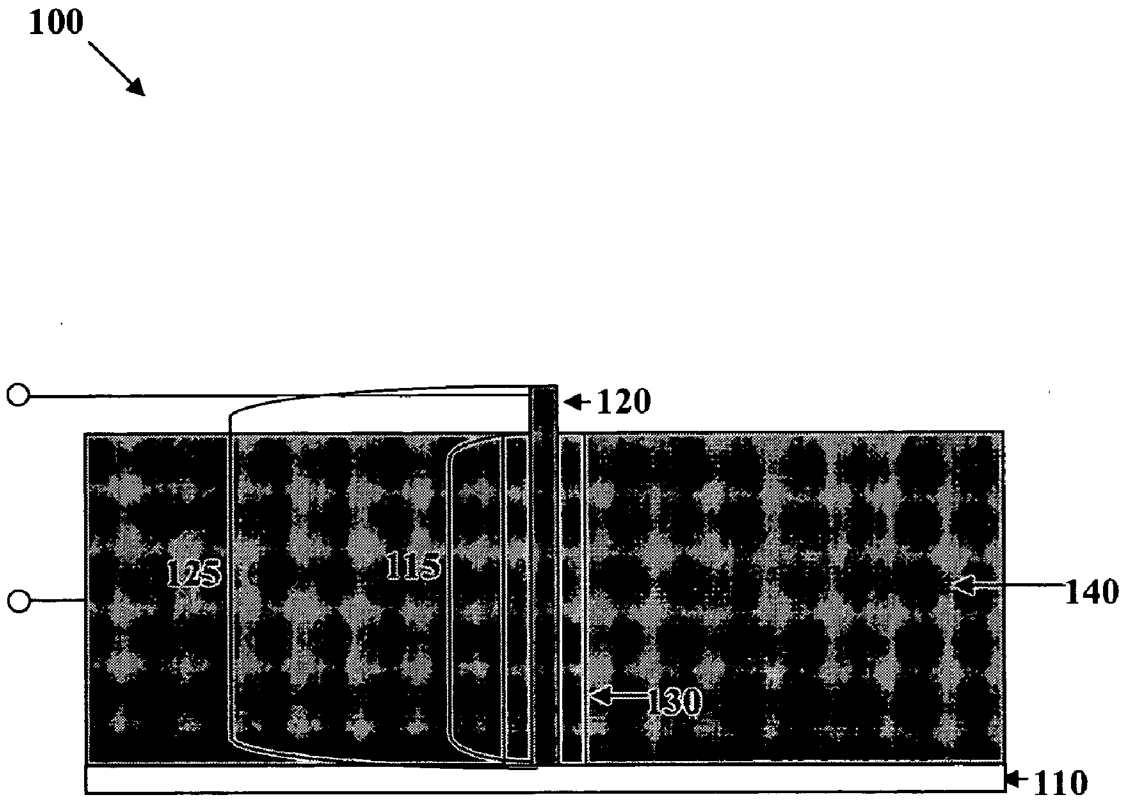

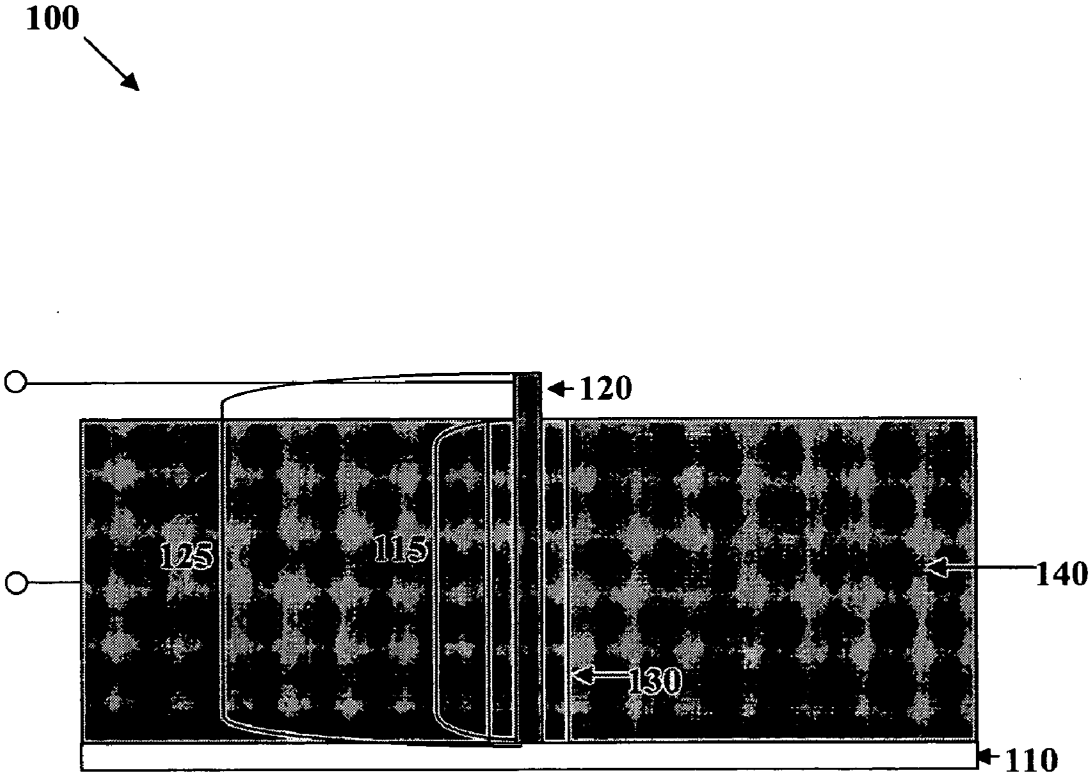
100
120
125
115
140
130
110

FIG. 1

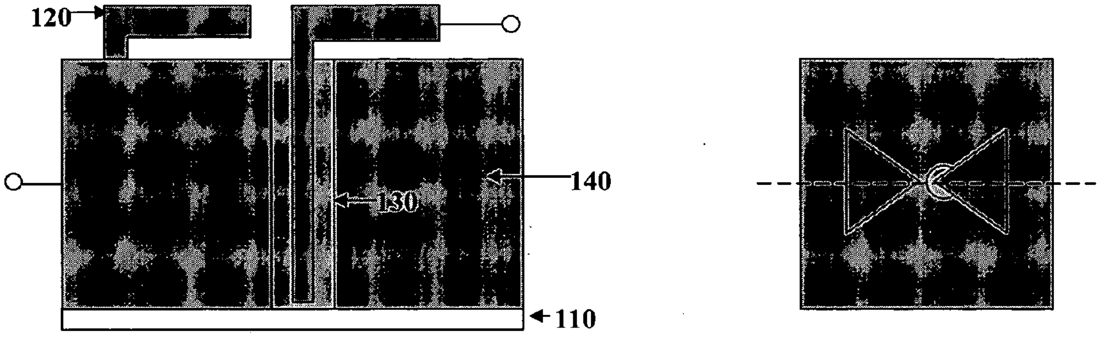
120
130
140
110

FIG. 2

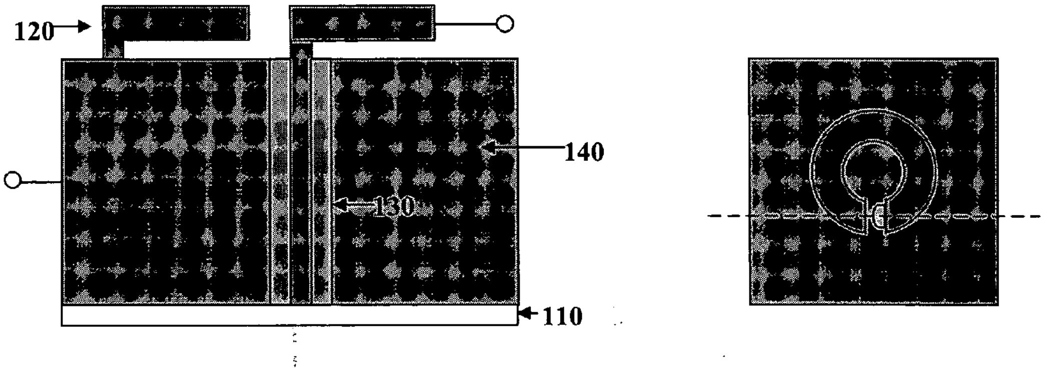
120
140
130
110

FIG. 3

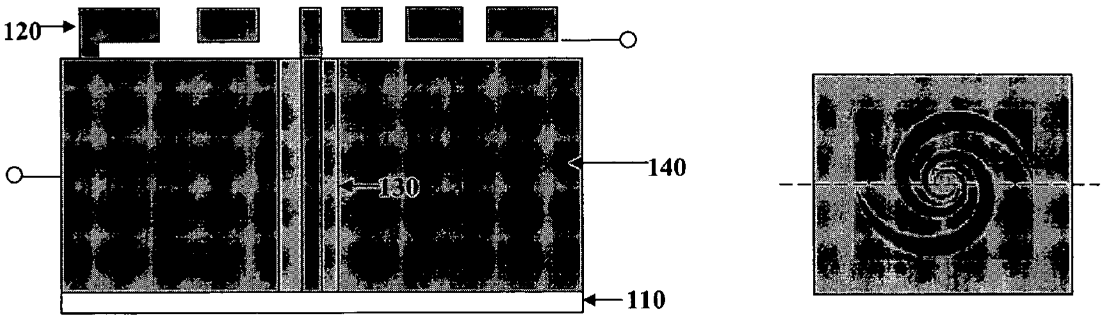
120
140
130
110

FIG. 4

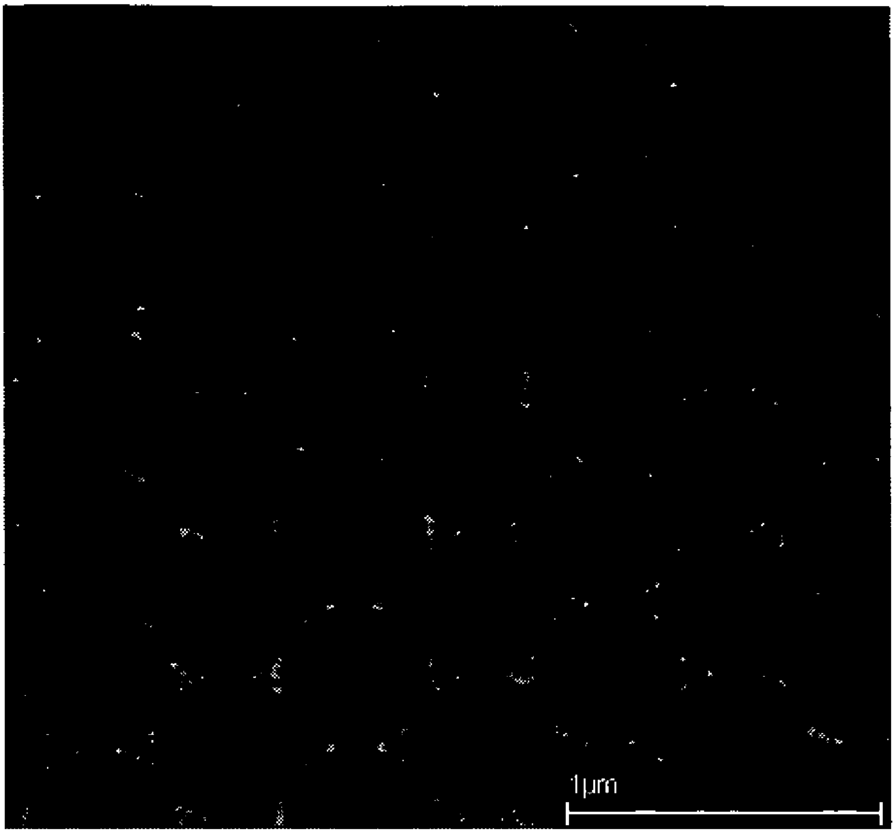

FIG. 5

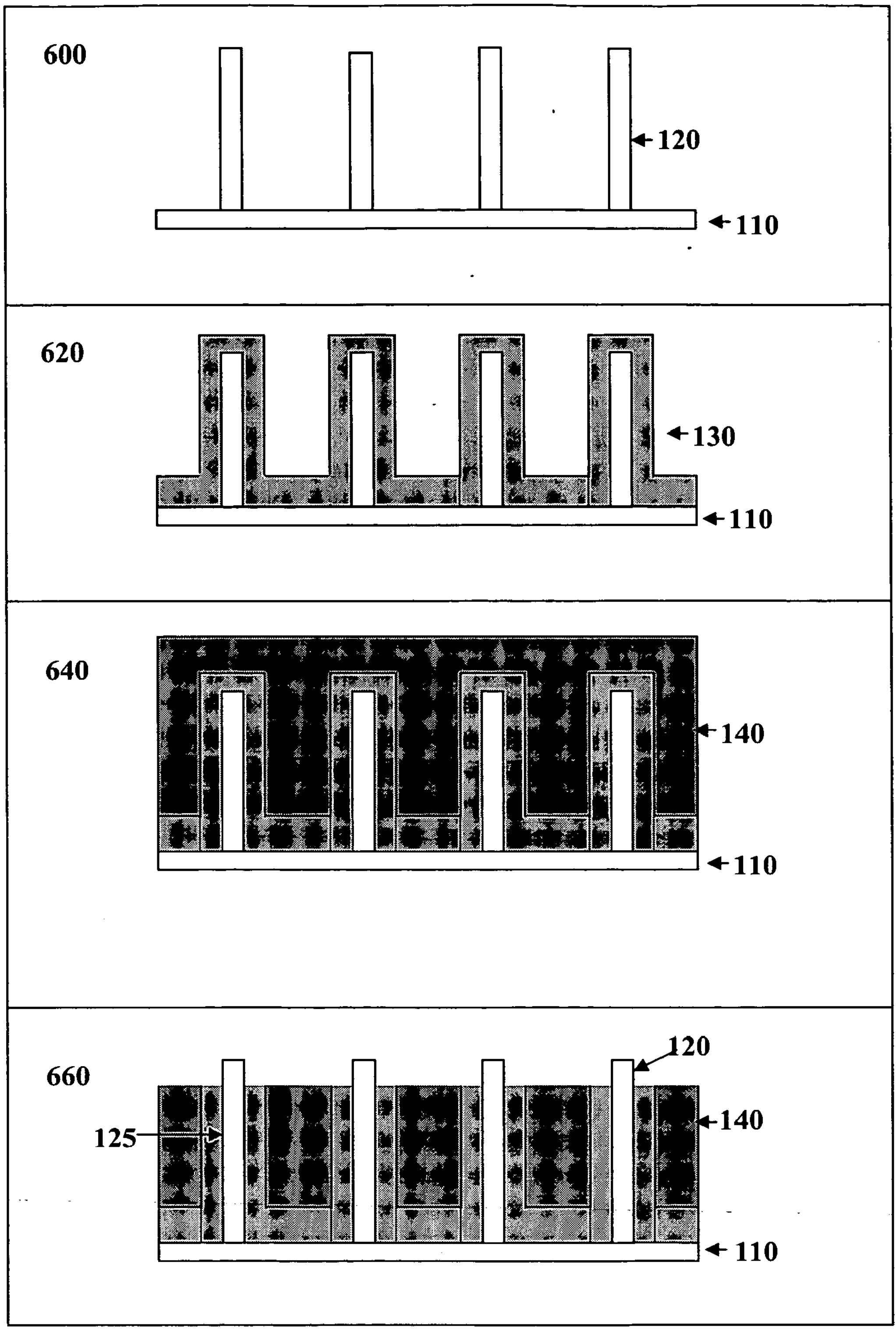
600
120
110
620
130
110
640
140
110
660
120
125
140
110

FIG. 6

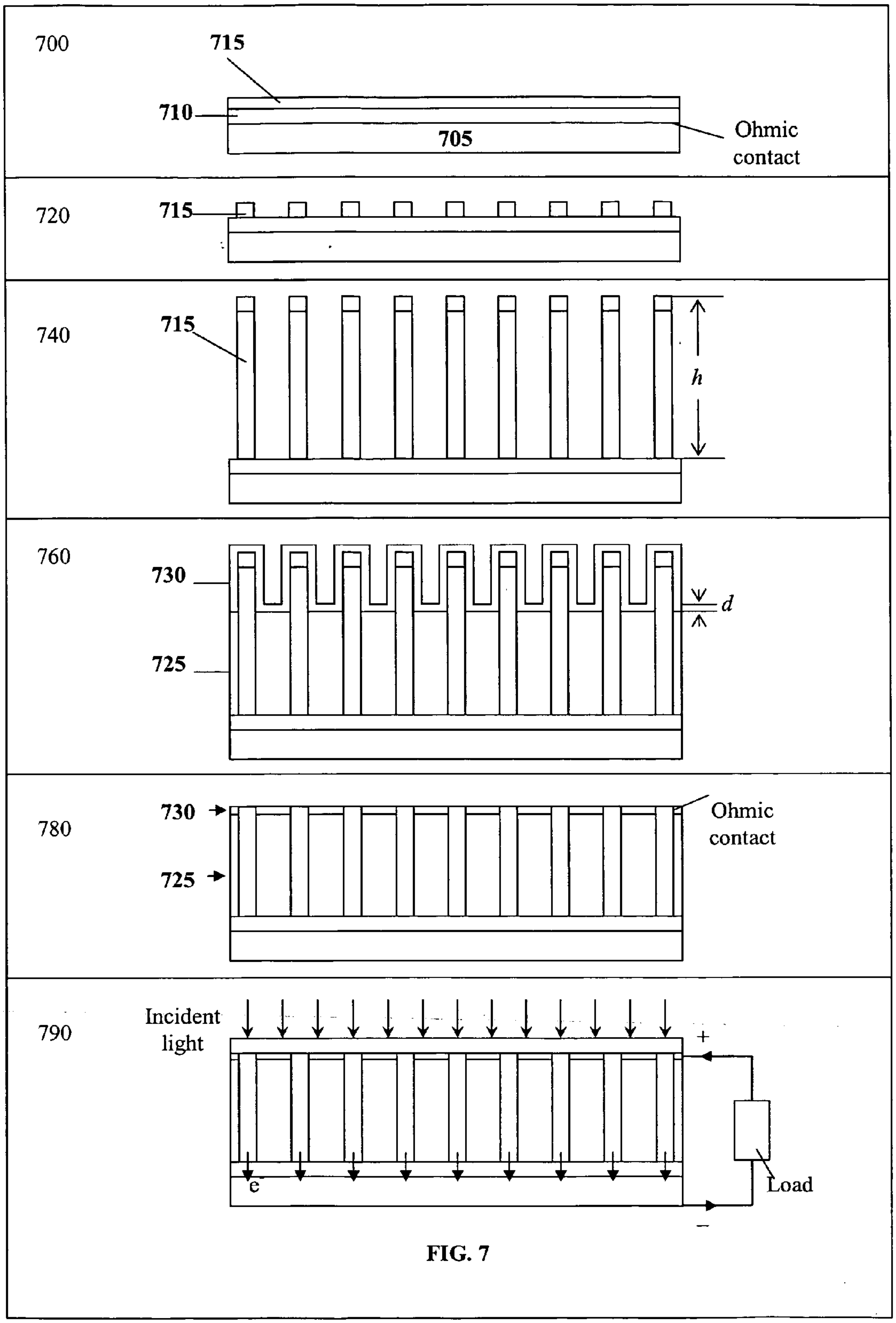
Ohmic contact
h
d
Ohmic contact
Incident light
+
e
Load
−

FIG. 7

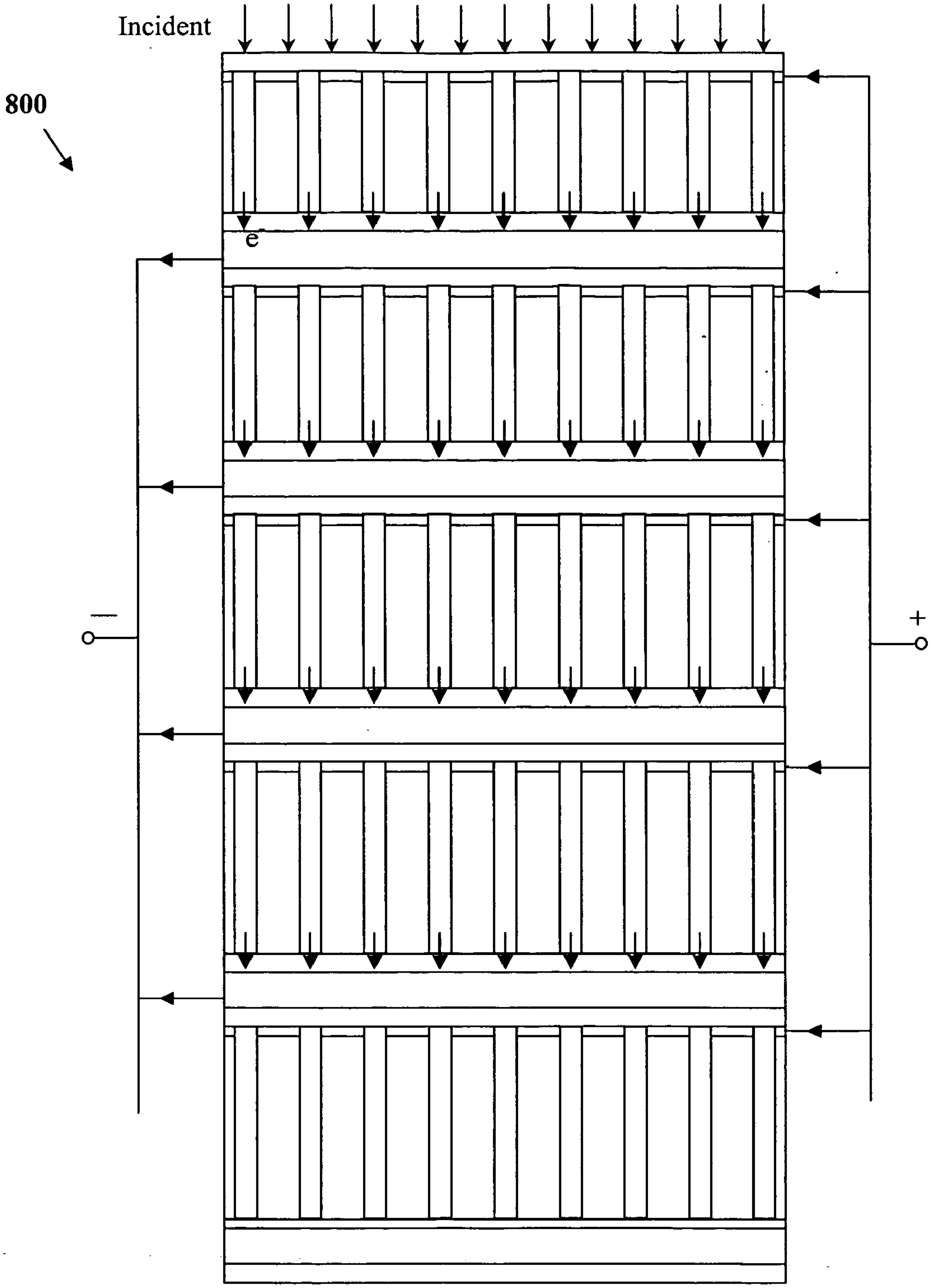
Incident
800
e⁻
—
+

FIG. 8

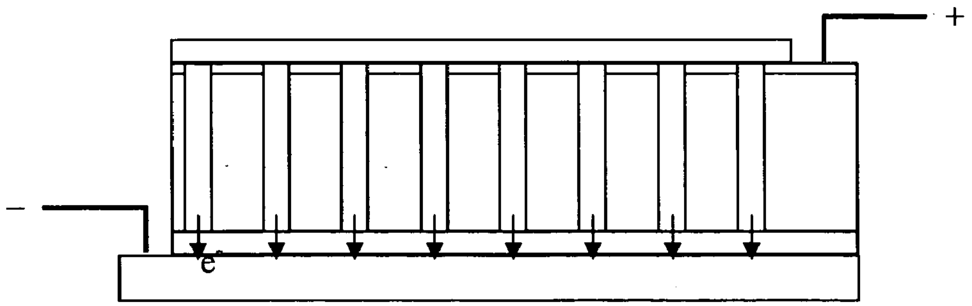

FIG. 9

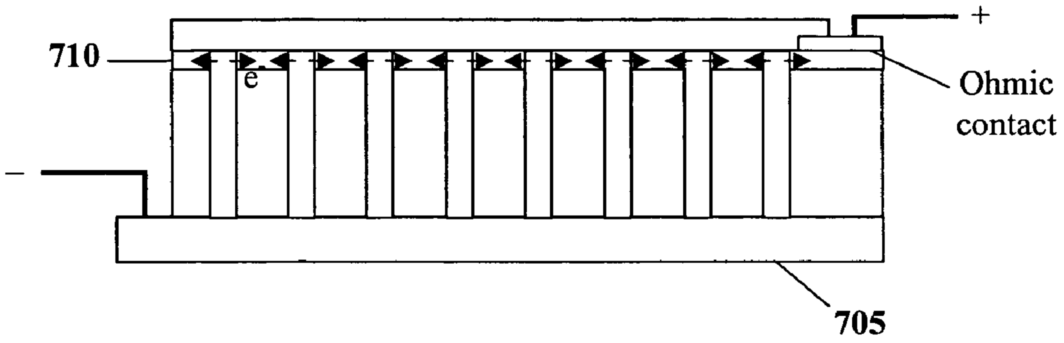
+
710
e
Ohmic contact
−
705

FIG. 10

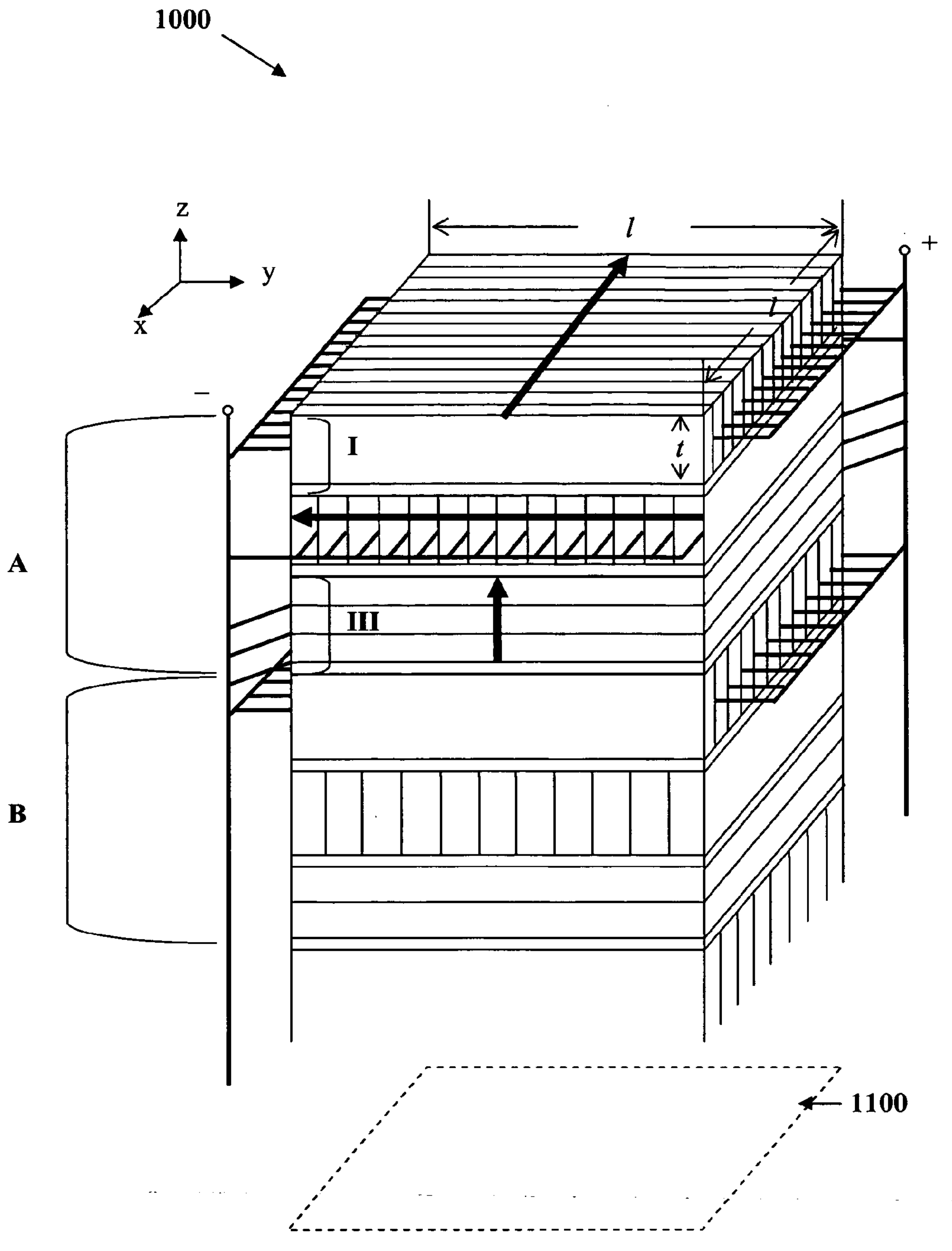
1000
z
y
x
l
l
t
–
+
I
III
A
B
1100

FIG. 11

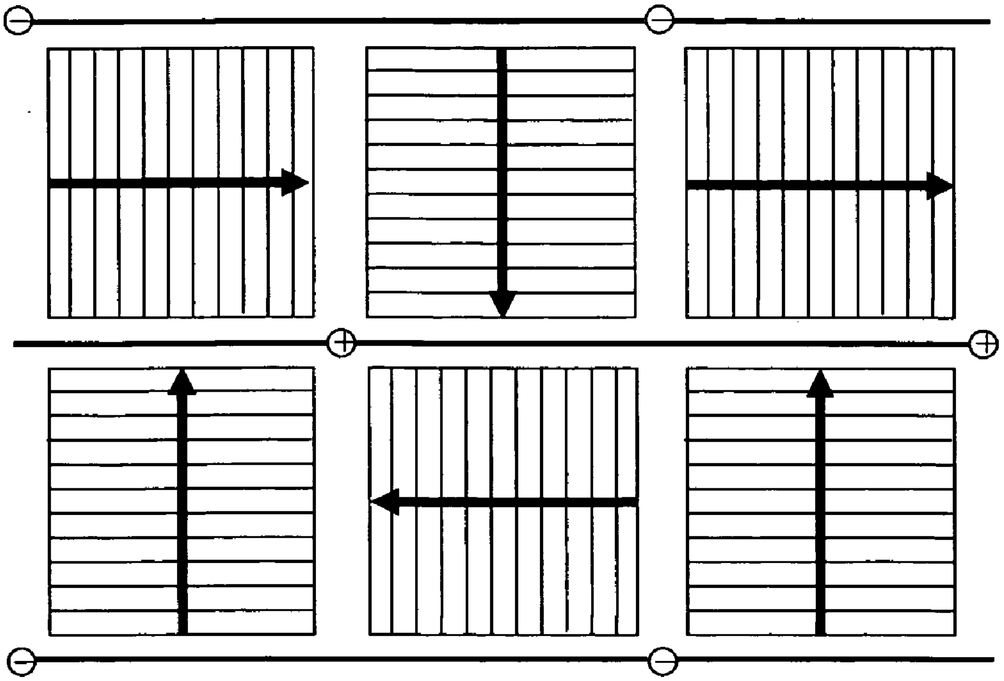

FIG. 12

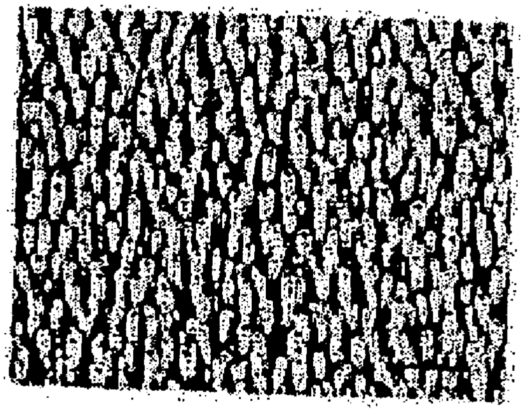

FIG. 13A

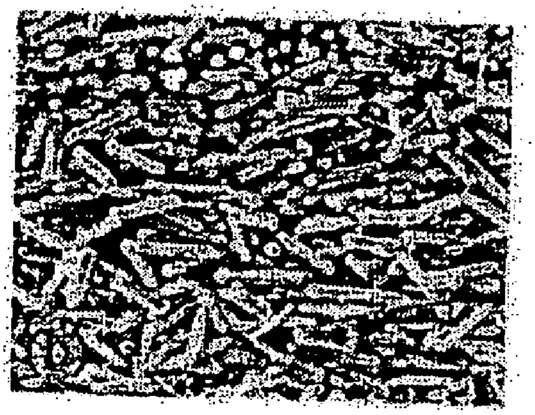

FIG. 13B

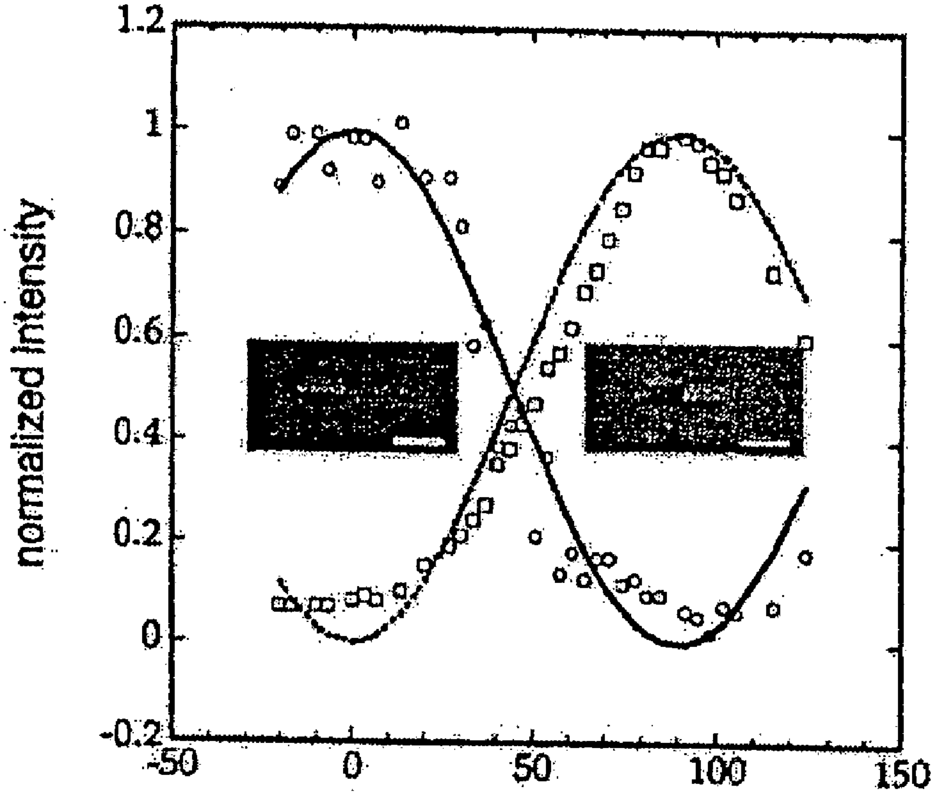

FIG. 14

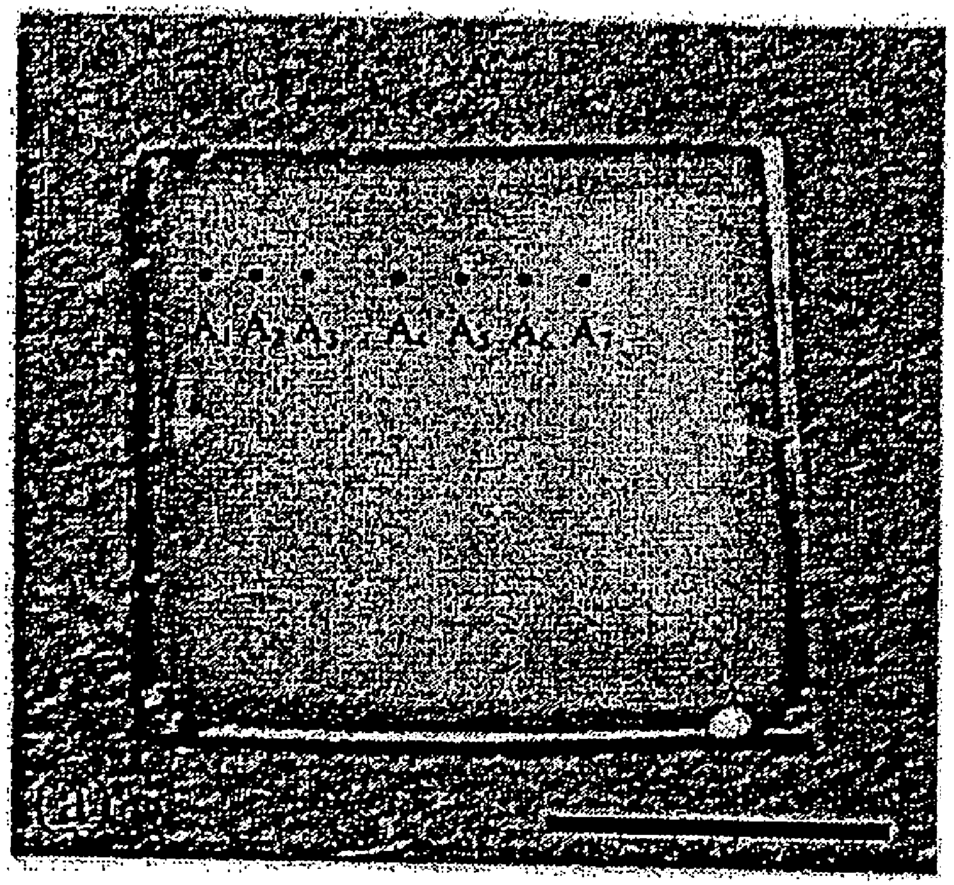
A1
A2
A3
A4
A5
A6
A7

FIG. 15

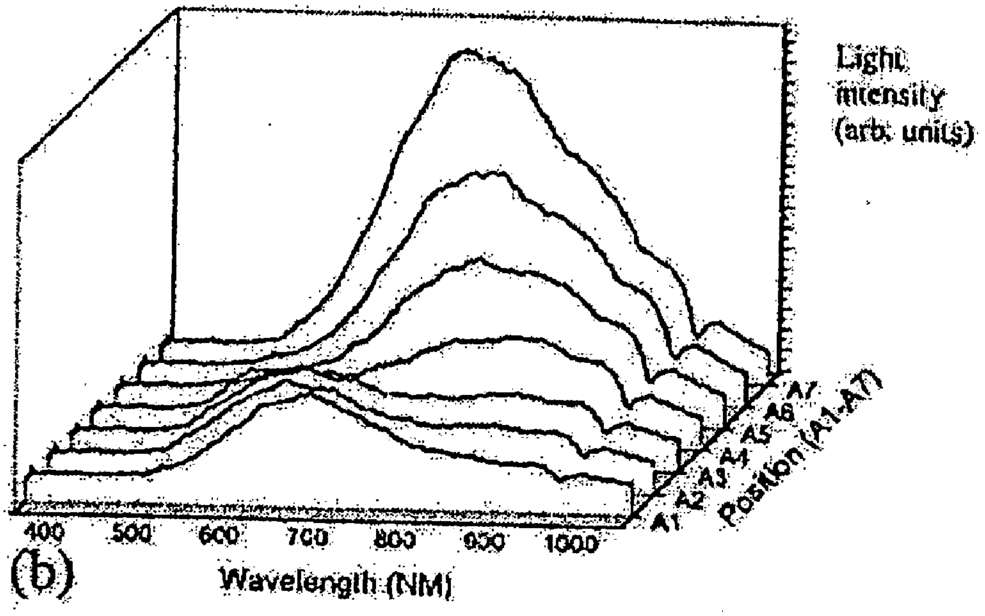
Light intensity (arb. units)
A7
A6
A5
A4
A3
A2
A1
Position (A1-A7)
400
500
600
700
800
900
1000
(b)
Wavelength (NM)

FIG. 16

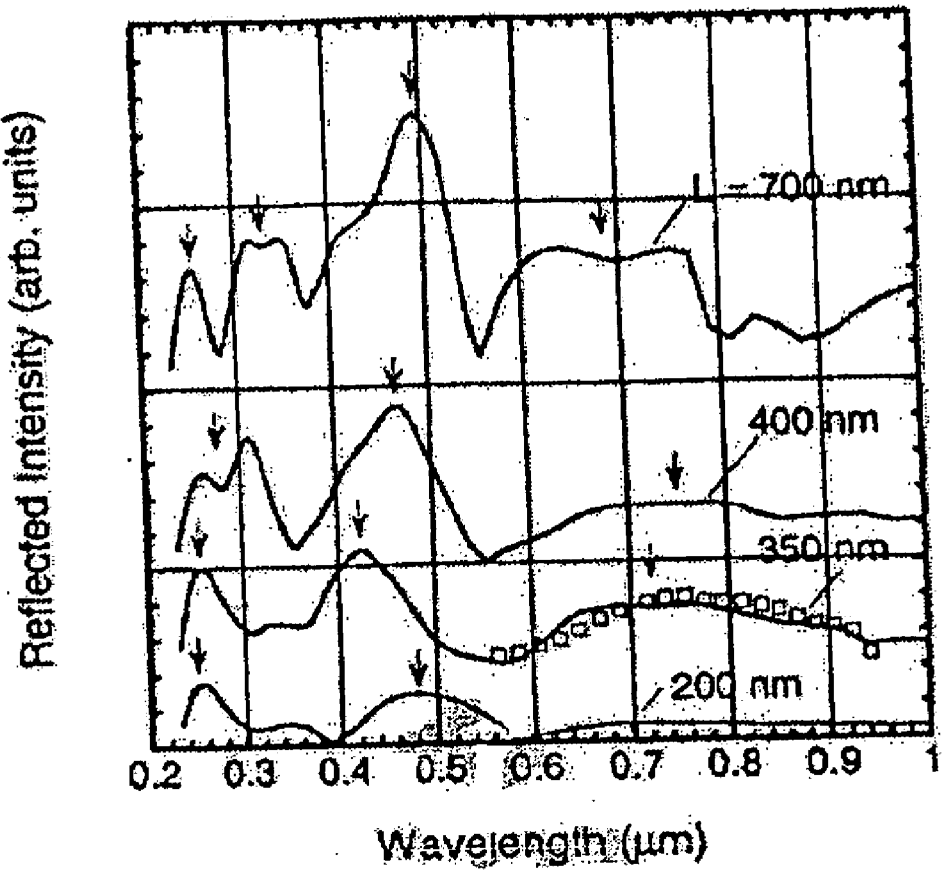
Reflected Intensity (arb. units)
L = 700 nm
400 nm
350 nm
200 nm
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Wavelength (µm)

FIG. 17

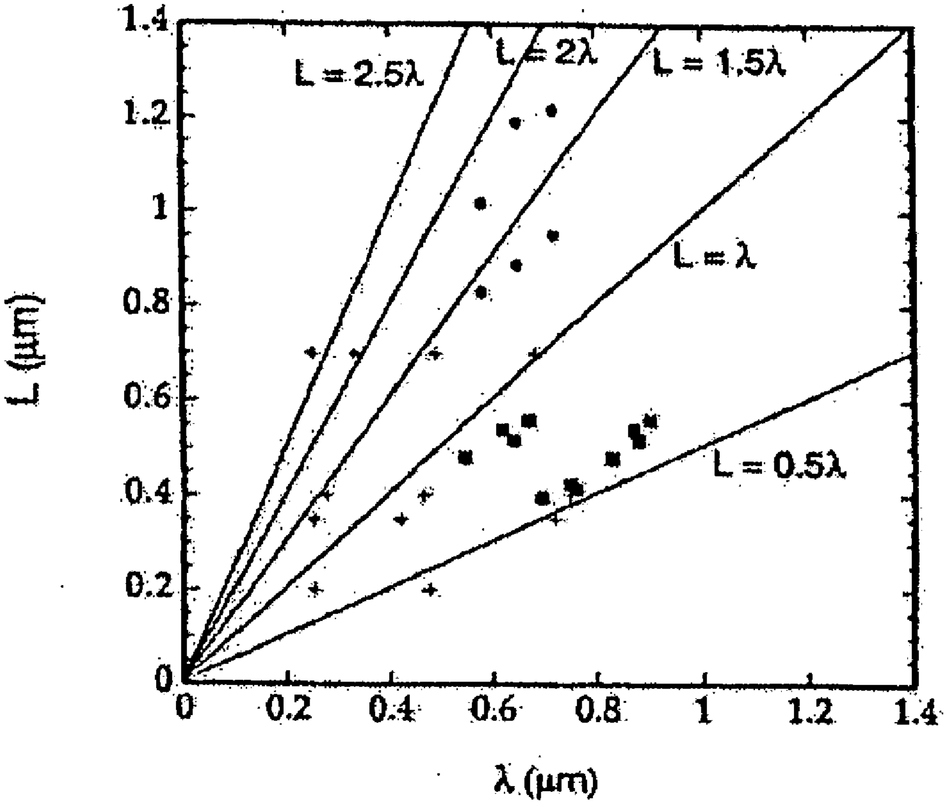
L = 2.5λ
L = 2λ
L = 1.5λ
L = λ
L = 0.5λ
L (μm)
λ (μm)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
0
0.2
0.4
0.6
0.8
1
1.2
1.4

FIG. 18

SOLAR CELLS USING ARRAYS OF OPTICAL RECTENNAS

RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application Ser. No. 60/619,262, filed Oct. 15, 2004, the entirety of which is hereby incorporated herein by reference.

GOVERNMENT SUPPORT

[0002] The present invention was made with partial support from The US Army Natick Soldier Systems Center under Grant Number DAAD16-02-C-0037 and partly by NSF under the grant NIRT 0304506. The United States Government retains certain rights to the invention.

FIELD OF INVENTION

[0003] The embodiments disclosed herein relate to nanoscale energy conversion devices having optical rectennas, and more particularly to high-efficiency solar cells having arrays of optical rectennas capable of receiving and transmitting solar energy and converting the solar energy into direct current electricity.

BACKGROUND OF THE INVENTION

[0004] The concept of using a rectifying antenna (rectenna) to collect solar energy was first proposed by R. L. Bailey in 1972; Since then, different approaches have been taken toward a practical fabrication of solar cells using optical rectennas. To date, however, no substantial progresses in practice have been reported due to major difficulties in achieving large-scale metallic nanostructures at low cost.

[0005] Recently, periodic and random arrays of multi-walled carbon nanotubes (MWCNTs) have been synthesized on various substrates. Each nanotube in the array is a metallic rod of about 10-100 nm in diameter and 200-1000 nm in length. Therefore, one can view interaction of these arrays with the electromagnetic radiation as that of an array of dipole antennas. MWCNTs arrays have been studied in order to determine the antenna-like interactions, since the most efficient antenna interaction occurs when the length of the antennas is of the order of the wavelength of the incoming radiation.

[0006] U.S. Pat. No. 6,038,060, U.S. Pat. No. 6,258,401, and U.S. Pat. No. 6,700,550 disclose various attempts at producing optical antenna arrays. However, there remains a need in the art for high energy conversion devices that employ optical antennas capable of receiving energy and converting AC current into a DC current. In addition, there is a need in the art for an efficient, reproducible method of producing such solar cells.

SUMMARY OF THE INVENTION

[0007] The present invention discloses a solar cell comprising a planar substrate having a top side and a bottom side. The solar cell comprises an at least one optical antenna having a geometric morphology capable of accepting energy. In addition, the cell comprises a rectifier having the optical antenna at a first end and engaging the substrate at a second end wherein the rectifier comprises the optical antenna engaged to a rectifying material. Also, the solar cell comprises a metal layer wherein the metal layer surrounds the rectifier from the top of the substrate to the optical antenna, wherein the optical antenna accepts energy and converts the energy from AC to DC along the rectifier.

[0008] Further, the present invention discloses a solar cell comprising a planar substrate having a conductor layer below a semiconductor layer. In addition, the cell comprises an array of carbon nanotubes engaging the semiconductor layer at a first end and comprising an optical antenna at a second end. In addition, the solar cell comprises a passivation layer wherein the passivation layer surrounds a length of the carbon nanotubes, wherein the optical antenna accepts energy and delivers energy to the solar cell wherein AC is rectified to DC.

[0009] In addition, the present invention discloses methods of producing such solar cells. In an embodiment, a method is disclosed for producing a solar cell which comprises growing a plurality of vertically-aligned nanotubes on a substrate and depositing a layer of a rectifying material onto the nanotubes. In addition, the method comprises depositing a layer of metal to cover a length of the nanotubes.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] The presently disclosed embodiments will be further explained with reference to the attached drawings, wherein like structures are referred to by like numerals throughout the several views. The drawings are not necessarily to scale, the emphasis having instead been generally placed upon illustrating the principles of the presently disclosed embodiments.

[0011] FIG. **1** is a schematic diagram showing an energy conversion device of the disclosed embodiments having a dipole antenna design.

[0012] FIG. **2** is a schematic diagram showing an energy conversion device of the disclosed embodiments having a bow-tie antenna design.

[0013] FIG. **3** is a schematic diagram showing an energy conversion device of the disclosed embodiments having a loop antenna design.

[0014] FIG. **4** is a schematic diagram showing an energy conversion device of the disclosed embodiments having a spiral antenna design.

[0015] FIG. **5** is a scanning electron microscopy image of an energy conversion device of the disclosed embodiments having a bow-tie antenna design.

[0016] FIG. **6** shows method steps for synthesizing an energy conversion device of the disclosed embodiments having a dipole antenna design.

[0017] FIG. **7** shows method steps for synthesizing an energy conversion device of the disclosed embodiments having a dipole antenna design.

[0018] FIG. **8** shows an example of a solar cell created using a number of energy conversion devices of FIG. **7**.

[0019] FIG. **9** shows the electrical connections to the energy conversion device of FIG. **7**.

[0020] FIG. **10** shows an energy conversion device of the disclosed embodiments having a CNT-semiconductor tunnel junction at a distal end of the CNT.

[0021] FIG. **11** shows an example of a solar cell configuration using a number of energy conversion devices.

[0022] FIG. **12** shows how large-scale assemblies of energy conversion devices can be formed where neighbor cells share a common cable to simplify connections.

[0023] FIG. **13**A shows a scanning electron microscope (SEM) image of an array of aligned MWCNTs. FIG. **13**B shows an SEM image of an array of scratched MWCNTs.

[0024] FIG. **14** shows a graph illustrating polarization effect.

[0025] FIG. **15** shows interference colors from the random array of MWCNTs.

[0026] FIG. **16** shows a graph illustrating reflected light intensity radiation wavelength measured in selected points on the sample shown in FIG. **15**.

[0027] FIG. **17** shows calculated reflected light intensity spectra for a model array of random antennas for various nanotube lengths.

[0028] FIG. **18** shows average length of MWCNTs versus wavelength of the incoming radiation at the corresponding maxima of reflected light intensity.

[0029] While the above-identified drawings set forth presently disclosed embodiments, other embodiments are also contemplated, as noted in the discussion. This disclosure presents illustrative embodiments by way of representation and not limitation. Numerous other modifications and embodiments can be devised by those skilled in the art which fall within the scope and spirit of the principles of the presently disclosed embodiments.

DETAILED DESCRIPTION

[0030] The embodiments disclosed herein relate to the field of energy conversion devices and more particularly to a solar cell using random arrays of nanotube optical rectennas. The following definitions are used to describe the various aspects and characteristics of the presently disclosed embodiments.

[0031] As referred to herein, "carbon nanotube", "nanowire", and "nanorod" are used interchangeably.

[0032] As referred to herein, "nanoscale" refers to distances and features below 1000 nanometers (one nanometer equals one billionth of a meter).

[0033] As referred to herein, "single-walled carbon nanotubes" (SWCNTs) consist of one graphene sheet rolled into a cylinder. "Double-walled carbon nanotubes" (DWCNTs) consist of two graphene sheets in parallel, and those with multiple sheets (typically about 3 to about 30) are "multi-walled carbon nanotubes" (MWCNTs).

[0034] As referred to herein, CNTs are "aligned" wherein the longitudinal axis of individual tubules are oriented in a direction substantially parallel to one another.

[0035] As referred to herein, a "tubule" is an individual CNT.

[0036] The term "linear CNTs" as used herein, refers to CNTs that do not contain branches originating from the surface of individual CNT tubules along their linear axes.

[0037] The term "array" as used herein, refers to a plurality of CNT tubules that are attached to a substrate material proximally to one another.

[0038] As referred to herein, a "catalytic transition metal" can be any transition metal, transition metal alloy or mixture thereof. Examples of a catalytic transition metal include, but are not limited to, nickel (Ni), silver (Ag), gold (Au), platinum (Pt), palladium (Pd), iron (Fe), ruthenium (Ru), osmium (Os), cobalt (Co), rhodium (Rh) and iridium (Ir). In an embodiment, the catalytic transition metal comprises nickel (Ni).

[0039] The terms "nanocrystals,""nanoparticles" and "nanostructures," which are employed interchangeably herein, are known in the art. To the extent that any further explanation may be needed, they primarily refer to material structures having sizes, e.g., characterized by their largest dimension, in a range of a few nanometers (nm) to about a few microns. In applications where highly symmetric structures are generated, the sizes (largest dimensions) can be as large as tens of microns.

[0040] The term "CVD" refers to chemical vapor deposition. In CVD, gaseous mixtures of chemicals are dissociated at high temperature (for example, CO_2 into C and O_2). This is the "CV" part of CVD. Some of the liberated molecules may then be deposited on a nearby substrate (the "D" in CVD), with the rest pumped away. Examples of CVD methods include but not limited to, "plasma enhanced chemical vapor deposition" (PECVD), "hot filament chemical vapor deposition" (HFCVD), and "synchrotron radiation chemical vapor deposition" (SRCVD).

[0041] A nanoscale energy conversion device of the presently disclosed embodiments is shown generally at **100** in FIG. **1**. As a brief overview, FIG. **1** shows an embodiment of an optical rectenna **125**. The optical rectenna **125** engages a substrate **110** at a first end and comprises an optical antenna **120** at a second end. The optical antenna **120** receives energy from an outside source and delivers the energy to the solar cell **100**. Further, FIG. **1** shows a rectifier **115** wherein the rectifier **115** comprises a rectifying material **130** engaged to the antenna. Various embodiments of each of these elements is discussed below.

[0042] Referring to FIG. **1**, a nanoscale energy conversion device (solar cell) **100** includes a metal substrate **110** having an array of optical rectennas **125** penetrating the metal substrate **110** and extending beyond the top surface of the metal substrate **110**. Only a single optical rectenna **125** is visible in FIG. **1**, however, an array of optical rectennas **125** exists. The optical rectenna comprises an antenna engaged to a rectifying material **130**. In an embodiment, the rectifying material **130** is an insulator. In an embodiment, the rectifying material is a semiconductor material **130**. In an embodiment, the rectifying material **130** is air. In an embodiment, the rectifying material **130** is a vacuum. The rectifying material **130** may be engaged to the antenna, either before the antennas are grown on the metal substrate **110** or after they are grown on the metal substrate **110**. A thick layer of metal **140** is deposited onto the optical rectennas **125** and portions of the optical rectennas **125** are exposed. The portions of the optical rectennas **125** that are exposed form optical antennas **120** and the portions of the optical rectennas **125** that are embedded in the thick layer of metal form the rectifier (in the form of transmission lines) **115**. In an embodiment, the portion of the optical rectennas **125** that are exposed to the semiconductor material **130** (tunnel junction) results in a rectifying diode. In an embodiment, the semiconductor material **130** is coated onto the optical rectennas **125** after they are grown, such that the optical antenna-semiconductor junction produces a rectifying diode.

[**0043**] The rectifier **115** is capable of rectifying optical frequency alternating current (AC) into direct current (DC) electricity. The optical antennas **120** are connected to a nanowire electrode embedded in the metal substrate **110**, in a vertical configuration (the rectifier section). The array of optical antennas **120** may form various geometric morphologies. In one embodiment, the geometric morphology of the optical antenna is similar to that of conventional microwave antennas. In one embodiment, the geometric morphology is a dipole antenna design. In one embodiment, the geometric morphology is a bow-tie antenna design (non-linear antenna design). In one embodiment, the geometric morphology is a loop antenna design (such as, an antenna forming a loop, parallel to the ground yielding a non-linear antenna design). In one embodiment, the geometric morphology is a spiral antenna design (non-linear antenna design). These designs are, shown respectively in FIGS. **1**, **2**, **3** and **4**. These various configurations allow for the antennas to have various bandwidth response, various directional response patterns with respect to the direction of wave propagation as well as the polarization, and various impedance matching options. Those skilled in the art will recognize that various geometric configurations are within the spirit and scope of the present invention.

[**0044**] In one embodiment, the optical rectennas **125** may be fabricated from a metal nanorod. In one embodiment, the nanorod comprises aluminum. In one embodiment, the nanorod comprises gold. In one embodiment, the optical rectennas comprise carbon nanotubes. In one embodiment, the optical rectennas comprise a dielectric material. Those skilled in the art will recognize that the optical rectennas may comprise various materials and remain within the spirit and scope of the present invention.

[**0045**] Techniques for fabricating the energy conversion device **100** include, but are not limited to, top-down electron beam lithography and bottom-up nanostructure synthesis. In an embodiment, the array of optical antennas **120** forms a dipole antenna design and fabrication of the optical rectennas **125** may be performed using aligned carbon nanotubes grown by a plasma-enhanced chemical vapor deposition method. In an embodiment, the array of optical antennas **120** forms a bow-tie antenna design and fabrication of the optical rectennas **125** may be performed using microsphere lithography where triangular islands with one edge facing each other can be achieved in large scale (as shown in the scanning electron microscopy image of FIG. **5**).

[**0046**] FIG. **6** illustrates a method of fabricating an energy conversion device **100** having a dipole nanoscale optical antenna design. In step **600**, vertically-aligned optical rectennas **125** may be grown, or lithographically created on a metal substrate **110**, that can be etched or dissolved later. In step **620**, a thin intermediate layer, of an insulator **130** (including, but not limited to, SiO_2) or a semiconductor **130** (including, but not limited to, doped or undoped Si, SiC or GaAs), is deposited onto the recntennas **125**. In step **640**, a thick layer of metal **140** (including, but not limited to, Au or Ag) is deposited (by evaporation or sputtering) to cover the rectennas **125**. In step **660**, the surface of the device may be polished, causing the protruding optical antenna **120** tips to be broken up and removed together with the conductive materials coated on the tips, thus exposing the optical antennas **120** cross sections, as shown in step **660**. The optical antennas **120** are in contact with the semiconductor material **130** resulting in a tunnel junction.

[**0047**] The material of the antenna **120** and rectifier **115** sections can be properly chosen to activate plasma resonances, resulting in an enhancement of the antenna **120** response. Nanostructures of gold and silver have plasmonic frequencies in the visible frequency range that may be tuned by changing the antenna **120** geometry. Thus, the electrical field response may be intensified by a factor of several orders of magnitude, both in the case of the antenna **120**, as well as in the rectifier **115**.

[**0048**] The configuration of the embedded rectennas **125** resembles that of a transmission line, impedance matching of which (to the antenna section) may be easily achieved. The energy collected by the antennas **120** will be concentrated in the transmission lines **115** (embedded in the metal substrate **140**) where it is rectified and converted into electricity. The total area of the rectification area, e.g. the transmission line, may be of any size, not limited by the scale of the incident wavelength. The difference of the instantaneous electric field strength on the opposing antennas **120** and metal surfaces **140** causes electrons to tunnel through the intermediate layer (insulating or semiconducting) **130** having an asymmetric barrier height at the two junctions, resulting in net current flow.

[**0049**] When relatively narrow bandwidth antennas **120** are used, stacks of layers of rectenna structures **125** with different working frequencies may be used to respond collectively to a wide solar spectrum (for example, in a dipole design). Alternatively, the same could be achieved by implementing arrays of antennas **120** with random length. If in addition to random length a random orientation of antennas **120** is used, response to an unpolarized light will be maximized.

[**0050**] In an embodiment, a bottom-up procedure is used to fabricate high-efficiency energy conversion devices using random arrays of aligned multi-walled carbon nanotubes (MWCNTs) as the optical rectennas. The MWCNTs are synthesized on substrates by the plasma-enhanced chemical vapor deposition (PECVD) process. The bottom-up fabrication procedure utilizes MWCNTs both as the optical antennas and in the rectifying diodes. A configuration of MWCNT-semiconductor (CNT-Sc) tunnel junction is able to rectify optical frequency AC currents into DC currents. The CNT-Sc configuration features high reproducibility, low series resistance, and low cost.

[**0051**] The bottom-up fabrication procedure takes advantage of nanomaterial synthesis and novel transparent conductive materials and is carried out by a scalable layer-by-layer technique. The use of conductive and semiconducting transparent materials of the presently disclosed embodiments is compatible with large-scale industrial production.

The high-efficiency energy conversion devices disclosed herein are capable of intrinsic energy conversion efficiencies of over about 80%, featuring amplified output current and minimum internal resistance. The characteristics of MWCNTs make the disclosed energy conversion devices useful in a variety of areas such as optoelectronic devices, such as THz and IR detectors and solar cells.

[0052] Aligned MWCNT arrays grown on silicon substrates using PECVD act as optical rectennas, receiving and transmitting light at ultraviolet (UV), visible and infrared (IR) frequencies. Most of the MWCNTs grown by PECVD methods are shown to be truly metallic. In addition, MWCNT-metal junctions have been found to be ohmic and MWCNT-semiconductor junctions have been found to have rectifying behaviors like schottky diodes. The work function of MWCNTs have been measured and found to be close to the work function of graphite which is highly conductive. Recent in situ tunneling electron microscopy studies have shown that the growth of MWCNTs starts off with several graphite layers parallel to the substrate surface at the CNT-substrate interface.

[0053] As is shown in example 1 below, it has been shown that MWCNTs interact with light in the same manner as simple dipole radio antennas. In particular, MWCNTs show both the polarization and the length antenna effect. The first effect is characterized by a suppression of the reflected signal when the electric field of the incoming radiation is polarized perpendicular to the CNT axis. The second, the antenna length effect, maximizes the response when the antenna length is a proper multiple of the half-wavelength of the radiation. The characteristics make the devices disclosed herein useful in a variety of areas such as optoelectronic devices, such as THz or and IR detectors.

[0054] To functionalize MWCNTs as optical rectennas, a femto-second rectifier must be engaged to each MWCNT to change the optical frequency AC current into DC current. An asymmetric metal-insulator-metal (MIM) tunnel junction structure has been disclosed for fabrication of such ultra-fast diodes. However, the methodology requires an unlimited selection range of materials and a very accurate control of the insulating layers thickness at the atomic scale. This greatly restricts the reproducibility and scalability in the practical process. For the case of MWCNTs as an example, the work function is restricted to about 4.9 eV which is prohibitively big compared to the visible frequency photon energy 1.8-3.2 eV The number of available transparent conductive materials is also very limited. Indium Tin Oxide (ITO) is one of such materials that is the most widely used in industry and has a work function in the range of 4.3eV-5.5 eV. Thus, the CNT-insulator-ITO junction will only work in the low-voltage scenario in both forward and reverse biases where the net current density is extremely small ($<10^{-6}$ Acm^{-2} for barrier thickness of 2 nm).

[0055] The disclosed embodiments provide for a CNT-Sc tunnel junction structure at one end of each individual CNT in order to form a rectifying diode. The CNT-Sc tunnel junction can be at either a distal end of the optical antenna (i.e., near the tip) or at a proximal end of the optical rectenna (i.e., at the end where the rectenna and the substrate are formed). The characteristics of the CNT-Sc tunnel junction resemble a conventional metal-semiconductor tunnel junction due to the intrinsic metallic property of the MWCNTs. The CNTs should have an average diameter of less than about 70 nm for significant quantum mechanical tunneling effect to dominate the thermionic emission. The choices of semiconductors are broad, and include, but are not limited to, heavily doped Si, GaAs, SiGe, Sic, and GaN (-type or n-type, doping density $>10^{19}$ cm^{-3} for barrier thickness ~3 nm). For multi-layer fabrications, transparent semiconductors are employed, including, but not limited to, ZnO:Al (AZO, n-type), $SrCu_2O_2$ (SCO, p-type), and $CuAlO_2$ (CAO, p-type), whose doping levels may be well controlled. Ohmic contacts to heavily doped silicon may be achieved by evaporating (sputtering) a catalytic material such as Al, Au, or Ni, onto the silicon and sintering at about 400° C. Ohmic contact to n-ZnO may be achieved by depositing n^+-ZnO. Ohmic contact to p-SCO may be achieved by depositing In_2O_3:SnO_2 (ITO) onto the semiconductors at low temperature respectively. Sputtering targets of these oxide materials are widely available and suitable for large-scale production. The net forward tunneling current density of CNT-(p)Si heterojunctions has been shown to be on the order of 10^{-3} Acm^{-2}-$10^{-2}Acm^{-2}$ under a bias voltage of 1.8V-3.2V. The CNT-Sc tunnel junctions disclosed herein may result in a higher order of magnitude.

[0056] A method of fabricating an energy conversion device **700** using a bottom-up procedure is shown in FIG. **7**. In step **700**, a semiconducting thin film **710** (such as, p^+-Si) is deposited onto a conductive substrate **705** (such as, Al), and standard procedures are carried out to form an ohmic contact. A catalytic material **715**, such as, Ni, Fe, or Co, is deposited onto the semiconducting film **710** using DC magnetron sputtering. In an embodiment, the catalytic material **715** is Ni. The desired length and diameter of the carbon nanotubes (CNTs) are achieved by accurate control of the growth parameters. The deposition thickness of the catalytic material **715** has a direct affect on the average diameter of the aligned CNTs grown. Without being limited to any particular theory, the difference in the average diameter is most likely due to the fact that Ni films of different thicknesses break into catalytic particles of different average sizes by heat treatment during the growth procedure, step **720**. Table 1 lists data showing the average diameters of aligned CNTs grown in a PECVD system resulting from different deposition thickness of Ni as the catalytic material **715**. Clean aligned CNT **715** arrays with Ni particles on top can then be grown using a DC glow discharge plasma in an atmosphere of NH_3 and C_2H_2, as shown in step **740**. Mixing ratios of about 4:1 or about 2:1 can be used. A growth time of about 1-2 min, yields CNTs **715** around or shorter than 1000 nm. The morphology of the CNTs **715** including length, diameters, straightness, etc., can be finely tuned by the other growth parameters such as plasma intensity and etching time, temperature, and total growth time. By modifying the growth parameters, and/or the geometry of the bias voltage electrodes, a nonuniform growth of CNTs **715**, with average length varying across the sample, can also be achieved. In this configuration, at the bottom of each individual CNT **715**, a nano-scale CNT-Sc tunnel junction is formed which rectifies the AC current excited within each antenna into a DC current at optical frequencies.

TABLE 1

Ni film thickness (nm)	Average CNT diameter (nm)
4	30
10	60
16	75
22	100
28	130

[0057] A highly transparent passivation layer **725** is then spin-coated in between the CNTs, as shown in step **760**, up to a height h of $\lambda/4n-d$ or $\lambda/2n-d$ (50 nm-500 nm for visible and near infrared), where λ is the wavelength of incident light in vacuum and n is the refractive index of the passivation material **725**. In an embodiment, the passivation layer **725** is a PMMA/copolymer layer, a silicone elastomer layer or another polymeric material layer. The spin-coating can be performed by varying the viscosity of the polymer solution and the spin rate. After baking (usually <200° C.), a thin film (thickness $d \ll \lambda/4n$ or $\lambda/2n$) of transparent conductive material **730** (such as Indium Tin Oxide (ITO) or n^+-(Zinc Oxide (ZnO)) is deposited on top of the passivation layer **725** and the exposed part of CNTs **715** by e-beam evaporation or sputtering, as shown in step **760**. The CNTs **715** grow sufficiently long ($>\lambda/4n$ or $\lambda/2n$, respectively) so that, by carefully polishing the surface at this stage, the protruding CNT **715** tips will be broken up and removed together with the conductive materials coated on the tips, exposing the CNT **715** cross sections, as shown in step **780**. The cross sections tend to be automatically closed or partially closed through the collapse of the CNT **715** walls near the open end. The CNT-transparent conductive material contacts are ohmic. An additional thin layer of the same passivation material **725** may be again spin-coated on top to provide a uniform dielectric medium surrounding the CNT antennas and protect the CNT antennas from outside attacks. A configuration where all the CNT rectennas **715** as individual current sources are connected in parallel is so achieved and the rectified DC currents will add up to a much higher magnitude accompanied by a substantially reduced total internal resistance of the rectennas **715**. The two conductive layers can be connected across an external load as DC electrodes, as shown in step **790**. The so-established single-wavelength energy conversion device, upon the incident light of wavelength λ polarized in the direction of CNT **715** alignment, will convert the photon energy into DC electricity at an efficiency greater then about 90%.

[0058] As shown in FIG. **8**, a solar cell **800** that catches the whole solar spectrum is constructed by the following arrangement of single-wavelength energy conversion devices **700** disclosed in FIG. **7**. The semiconducting films **710** and the substrates **705** are made of transparent materials. Devices **700** of different wavelength capabilities are cascaded together one below another. The ordering is such that the shorter the wavelength the upper the device **700** (in z direction), since it is easier for longer wavelength to penetrate through the media. All the devices **700** are then wired up in parallel to yield large current and small internal resistance. The electrical connections to each device **700** can be prepared during the original layer-by-layer construction process of the device **700** (see FIG. **9**). The solar cell **800** disclosed herein is capable of collecting full solar spectrum of photons polarized in the CNT direction at an efficiency greater than about 85%.

[0059] As shown in FIG. **10**, the order of the conductive film and the semiconducting film **710** can be interchanged by growing optical rectennas **715** from a conductive substrate **705** and depositing a semiconducting film **710** on the rectennas **715** later. Since the tip, side and bottom of CNTs have different atomic structures, there may be practical merits in this flexibility of changing the junction location. Another advantage of the top CNT-Sc junction configuration is that even smaller contact area can be achieved, if the thickness of the semiconducting film **710** is smaller than a quarter of the CNT diameter. It also means that the CNT diameter can be relatively large, which is easier for growth control and maintenance of CNT straightness during process, while still having a small enough contact area to reduce the parasitic capacitance and increase the switching frequency of the CNT-Sc junction by making the semiconducting film **710** sufficiently thin.

[0060] FIG. **11** shows an embodiment of a solar cell **1000**, where full spectrum unpolarized sunlight may be efficiently converted to DC electricity. The solar cell **1000** is composed of multi-levels (for simplicity, two levels A and B are shown) with shorter wavelength energy conversion devices **1050** closer to the exposed surface and longer ones farther away. Within a single level, there are three sublevels (I, II and III), each consisting of a stack of identical single-wavelength energy conversion devices **1050**. The solid arrows denote the CNT growth directions in the stacks. The three stacks (I, II and III) are so oriented that the CNT array in each aligns with a different orthogonal 3-dimensional axis. The two sublevels of CNTs parallel with the exposed surface (in x-y plane) may be made as wide as possible ($l \gg 10$ μm) but have to be sufficiently thin ($t<10$ μm, ~100 CNT layers) by using devices of strip-shaped substrates for good transmission. Fabrication of substrate strips at this scale may be achieved using contemporary photolithography. The other sublevel in which CNTs are oriented perpendicular to the exposed surface (in z direction) has no limitations in dimensions but the number of devices in a stack, which are larger dimension devices, is compensated so that every sublevel contains similar number of CNTs. The three sublevel configurations (I, II and III) are repeated from one level to another. Although in levels of longer CNTs, an increment in the thickness of the two parallel sublevels seems necessary for maintaining an equivalent CNT quantity to that in shorter CNT levels, the fact that the solar radiation has lower power in longer wavelengths actually predicts a lower demand for the quantity of longer CNTs, thus ensuring no significant change in level thickness along the z direction. If the average bandwidth of each level is about 50 nm (high selectivity) for half-wave antennas according to surface plasmon measurements, the total solar cell thickness should be on the order of about 1 mm. Currently, the best transparent polymeric material has a transmittance of over 95% at this scale of thickness. The solar cell device **1000** is supported on the bottom by a reflecting mirror **1100** facing up for secondary absorption.

[0061] According to the electrical connection pattern shown in FIG. **11**, all the energy conversion devices **1050** are connected in parallel and the currents merge into two major cables located besides the diagonal ridges of the solar cell

1000, resulting in a useful setup when producing large dimension single solar cells is practically prohibitive. The setup allows multiple solar cells to be further integrated into large-scale assemblies where neighbor cells share a common cable to simplify connections (FIG. **12**, top view). All the cables terminate on the surface, ready to be finally collected by a set of parallel wires. The energy conversion efficiency of the solar cell device **1000** is greater then about 80%.

[**0062**] The following provides an example of an embodiment of the current invention. The example in no way is meant to limit any aspect of the current invention.

EXAMPLES

Example 1

[**0063**] With this example, optical measurements of random arrays of aligned carbon nanotubes are disclosed, and show that the response is consistent with conventional radio antenna theory. The example first demonstrates the polarization effect, the suppression of the reflected signal when the electric field of the incoming radiation is polarized perpendicular to the nanotube axis. Next, the example demonstrates the interference colors of the reflected light from an array, and show that they result from the length matching antenna effect. This antenna effect could be used in a variety of optoelectronic devices, including THz and IR detectors.

[**0064**] In recent years, periodic and random arrays of multi-walled carbon nanotubes (MWCNTs) have been synthesized on various substrates, by the plasma-enhanced chemical vapor deposition (PECVD) process. Each nanotube in such arrays is a metallic rod of about 50 nm in diameter and about 200 to about 1000 nm in length. Therefore, one can view interaction of these arrays with the electromagnetic radiation as that of an array of dipole antennas. Since the most efficient antenna interaction occurs when the length of the antennas is of the order of the wavelength of the incoming radiation, the example expects an antenna-like interaction of MWCNT arrays with visible light. There are two major antenna effects. First, the polarization effect suppresses the response of an antenna when the electric field of the incoming radiation is polarized perpendicular to the dipole antenna axis. Second, the antenna length effect maximizes the antenna response when the antenna length is a multiple of half-wavelength of the radiation. The polarization antenna effort has already been observed in the Raman response of single-walled carbon nanotubes. The nanoscopic dipole antenna length effect was recently observed in microbolometer, stripline antenna. This example demonstrates both of these antenna effects in random MWCNT arrays. This example utilizes random nanotube arrays to suppress the intertube diffraction, which obscures the intratube antenna effects that are of interest here.

[**0065**] The MWCNT arrays of this example are fabricated using PECVD. The silicon substrate is coated with a thin film of nickel catalyst (about 20 nm) in a dc magnetron sputtering system, that is then heated to about 550-600 ° C. in a PECVD reaction chamber to break up the nickel film into small catalyst particles. A gas mixture NH_3 and C_2H_2 is introduced into the PECVD chamber at the ratio of 2:1, and a dc glow discharge plasma is then generated and maintained by a bias voltage of about 500-550V. A growth time of about 1-2 minutes yields nanotubes around or shorter than 1000 nm. FIG. **13**A shows the scanning electron microscope (SEM) image of such an array or random MWCNTs. By modifying the growth parameters, and/or the geometry of the bias voltage electrodes, a nonuniform growth of nanotubes, with average length varying across the sample, can also be achieved. This has been utilized to produce the samples used in the present example.

[**0066**] The example first demonstrates the polarization effect. A small piece of silicon wafer (2×1 cm^2) was coated with a thin film of Cr. Subsequently, one-half of the sample was coated with a thin film of Ni catalyst, and processed to grow a random array of MWCNTs. The sample was illuminated with white unpolarized light, and observed in a specular direction through a rotation polarizer. FIG. **14** shows that when the polarizer is oriented parallel to the growth direction of the nanotubes (orientation angle $\theta=0°$), the light reflected from the array is clearly visible, while the exposed metallic half of the sample is dark (not reflecting). With increasing (or decreasing) angle θ, intensity of the light reflected from the nanotube array diminishes, while the intensity reflected from the metallic side increases until at $\theta=90°$, the radiation is observable essentially only from the metallic side.

[**0067**] This behavior follows from the fact that, while in nanotubes currents are excited predominantly along their length, in the metallic film, currents flow in the film plane; that is perpendicular to the nanotubes. Each nanotube acts as an antenna reradiating light with the electric field E, polarized in the plane parallel to the antenna. A polarizer, with its axis of polarization rotated by an angle θ to this plane, transmits radiation with a projected electric field $E'=E\cos\theta$, and therefore the corresponding observed intensity is given by the law of Malus I_{NT} is proportional to $(E'')^2=E^2\cos^2\theta$ (solid line circles in FIG. **14**). For light reflected from the metallic film the situation is exactly "out-of-phase" with that of the array; that is, Imetal is proportional to $(E'')^2=E^2\sin^2\theta$ (dotted line-squares in FIG. **14**).

[**0068**] The second characteristic of an antenna is its resonant response behavior as a function of the radiation wavelength. This results from the condition that the induced current oscillations must "fit" into the antenna length (i.e., satisfy the boundary conditions at the antenna ends). A general equation describing the scattering maxima from a random array of dipole antennas (with vanishing current at each end) is:

$$L=m(\lambda/2)f(\theta,n), \tag{1}$$

where $f(\theta,n)=1$ for a single, simple diode, and $f(\theta,n)=(n^2-\sin^2\theta)^{-1/2}$ in the limit of the very dense array (thin film limit), where the average interantenna distance $D<\lambda$, $f(\theta,n)$ is equal to about 1., and is only weakly dependent on the angle θ. As such, similar behavior is expected for the random array of MWCNTs.

[**0069**] FIG. **15** shows a sample of random array of nanotubes with gradually reduced lengths (from left to right(illuminated by white light. The strong interference colors are due to the antenna length effect. FIG. **16** shows the intensity of the reflected light at the specular direction versus the radiation wavelength of the incoming radiation measured at selected spots (positions **A1**-**A7**) of the sample shown in

FIG. **15**. Experiments were done using the Ocean Optics USB2000 Fiber Optic Spectrometer (FOS). White light emerging from a 50-μm-diameter finer was focused onto the sample surface at about a 30 degree angle of incidence. Incident spot size was of the order of about 0.5 mm. A receiving fiber with a 1 mm diameter is positioned to collect light reflected spectrally from the sample surface. The system was first corrected for dark field and then normalized with respect to the tungsten light source and reflectance from the silicon substrate at the 30° incident angle. Another sample, with longer nanotubes, produced insufficient scattered light intensity for the FOS, and the wavelength was estimated using a high sensitivity CCD camera and optical filters, with accuracy of about 10%. The exact positions at which the data were acquired were permanently marked (scratched with a needle) on each of the samples. The SEM pictures were then used to estimate the average nanotube length at each marked spot. To minimize the parallax error of this estimate, only the collapsed (lying flat on the surface) nanotubes in the scratched area was measured.

[0070] In addition to the experiments described herein, computer simulations of the electromagnetic response from a random dipole antenna array have been performed. The array was modeled as a set of 10 parallel, equal-length antennas, randomly distributed on, and perpendicular to, a flat substrate. Antenna dimensions and the average interantenna distance represent the actual nanotube array. The dielectric constant of the substrate is assumed to be real and equal to 10. The resulting reflection curves for various antenna lengths are shown in FIG. **17**.

[0071] FIG. **18** combines the experimental and theoretical results to demonstrate the antenna length effect. In FIG. **18**, positions of the various reflected intensity maxima are plotted versus the corresponding average antenna length L. Solid lines represent the ideal dipole antenna condition [Eq. (1), with f=1] for different m. The measured results are represented as solid circles and squares. Crosses mark positions of the distinct maxima of the theoretical curves. Arrows indicate those maximas in FIG. **17**. It is clear that both experimental and theoretical results follow closely the ideal dipole antenna condition, and thus demonstrate that MWCNTs can act as light antennas.

[0072] This example also estimates the quality of the nanotube antennas. FIG. **17** shows a comparison between one of the experimental curves of FIG. **16** (for position A_5) shown as squares, and the corresponding calculated response. This comparison shows that the calculation, which assumed infinite conductivity of the metallic antenna, reproduces well the measured line width of the peak. Therefore, the peak width is primarily due to the radiation losses of currents induced inside the nanotubes. As such, this example concludes that the actual scattering rate of the conducting electrons (y) in the nanotube antenna must be much less that the width of the peak in FIG. **17**. After replotting the peak versus frequency (rather than the wavelength), the example finds that the width is about 10^{15s-1}, and therefore y must be of the order $1014s^{-1}$ or less. This is a very low scattering rate of the order of (or better than) that for good metals such as copper.

[0073] The fact that MWCNTs act as high quality light antenna suggests various applications based on the radio analogy. For example, a THz demodulator could be built, if a sufficiently fast diode is attached to (or built into) each antenna in the array mounted on a THz stripline. The modulating THz signal could then be seamlessly introduced into the stripline by shining modulated light onto the array. This scheme could be used in a new generation of THz and possibly IR detectors. The antenna length effect can be tuned by controlling the nanotube length, and to some extent the array density during the growth process, making the devices frequency selective. In principle, the antenna effects should be also detectable in, and the same applications possible with, arrays of aligned single-walled nanotubes. However, at this moment, no scheme for making such arrays of all metallic single-walled nanotubes exists, and there is no reason to believe that such a system would have any advantage over those based on MWCNTs.

[0074] In conclusion, this example demonstrates that MWCNTs interact with light in the same manner as simple diode radio antennas. In particular, they show both the polarization and the length antenna effect. The first effect is characterized by a suppression of the reflected signal when the electric field of the incoming radiation is polarized perpendicular to the nanotube axis. The second, the antenna effect, maximizes the response when the antenna length is a proper multiple of the half-wavelength of the radiation. These effects could be used in a variety of optoelectronic devices, such as THz and/or IR detectors.

[0075] All patents, patent applications, and published references cited herein are hereby incorporated by reference in their entirety. It will be appreciated that various of the above-disclosed and other features and functions, or alternatives thereof, may be desirably combined into many other different systems or applications. Various presently unforeseen or unanticipated alternatives, modifications, variations, or improvements therein may be subsequently made by those skilled in the art which are also intended to be encompassed by the following claims.

What is claimed is:

1. A solar cell comprising:

a planar substrate having a top side and a bottom side;

an at least one optical antenna comprising a geometric morphology capable of accepting energy;

a rectifier having the optical antenna at a first end and engaging the substrate at a second end wherein the rectifier comprises the optical antenna engaged to a rectifying material; and

a metal layer wherein the metal layer surrounds a length of the rectifier,

wherein the optical antenna accepts energy and converts the energy from AC to DC along the rectifier.

2. The cell of claim 1 wherein the geometric morphology of the optical antenna is a bow-tie morphology.

3. The cell of claim 1 wherein the geometric morphology of the optical antenna is a loop morphology.

4. The cell of claim 1 wherein the geometric morphology of the optical antenna is a spiral morphology.

5. The cell of claim 1 wherein the optical antenna comprises carbon nanotubes.

6. The cell of claim 1 wherein the optical antenna comprises an aluminum nanorod.

7. The cell of claim 1 wherein the optical antenna comprises a gold nanorod.

8. The cell of claim 1 wherein the rectifying material is a semiconductor.

9. The cell of claim 9 wherein the semiconductor is selected from the group consisting of doped silicon, undoped silicon, silicon carbide and GaAs.

10. The cell of claim 1 further comprising a plurality of optical antennas.

11. The cell of claim 10 wherein the plurality of optical antennas are of random lengths.

12. The cell of claim 10 wherein the plurality of optical antennas are of random orientation.

13. A solar cell comprising:

a planar substrate having a conductor layer below a semiconductor layer;

an array of carbon nanotubes engaging the semiconductor layer at a first end and comprising an optical antenna at a second end; and

a passivation layer wherein the passivation layer surrounds a length of the carbon nanotubes,

wherein the optical antenna accepts energy and delivers energy to the solar cell wherein AC is rectified to DC.

14. The cell of claim 13 wherein the passivation layer comprises a polymeric material.

15. The cell of claim 13 further comprising a transparent conductive layer above the passivation layer.

16. The cell of claim 15 further comprising a second passivation layer above the transparent conductive layer.

17. A method for producing a solar cell, comprising:

growing a plurality of vertically-aligned nanotubes on a substrate;

depositing a layer of a rectifying material onto the nanotubes; and

depositing a layer of metal to cover a length of the nanotubes.

18. The method of claim 17 wherein the nanotubes are carbon nanotubes.

19. The method of claim 17 wherein the rectifying material is a semiconductor.

20. The method of claim 17 wherein the rectifying material is selected from the group consisting of air, a vacuum, and an insulator.

* * * * *

View ScienceDirect over a secure connection: switch to HTTPS

ScienceDirect

Outline

Purchase PDFPurchase

Export

Search ScienceDirect

Advanced

Outline

Abstract
Keywords
1. Introduction
2. Split ring resonator
3. Antenna configuration
4. Results and discussion
5. Conclusion
References
Vitae

Figures (15)

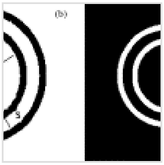
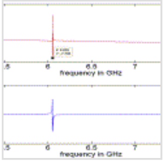
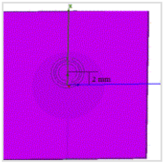
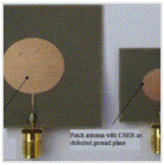
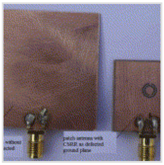
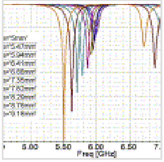
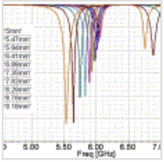
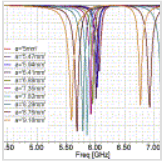
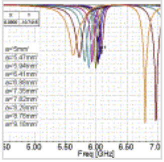
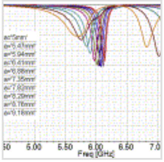
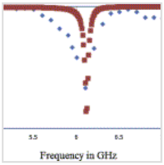

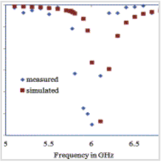

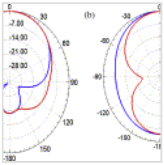
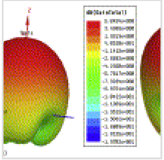
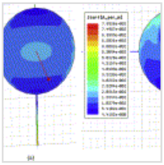

Hide figures

Tables (1)

Table 1

ELSEVIER

AEU - International Journal of Electronics and Communications

Volume 69, Issue 8, August 2015, Pages 1126-1133

AEU
International Journal of Electronics and Communications
8/2015

Reduced size patch antenna using complementary split ring resonator as defected ground plane

Falguni Raval [a] 👤 ✉, Y.P. Kosta [b], Harshita Joshi [a]

⊞ **Show more**

https://doi.org/10.1016/j.aeue.2015.04.013 Get rights and content

Abstract

In this paper, the authors have presented a reduced size metamaterial loaded circular microstrip patch antenna. In the present work, complementary split ring resonator is loaded in the ground plane of circular microstrip patch antenna is. The unloaded circular microstrip patch antenna resonates at 6.11 GHz, whereas after loading it with complementary split ring resonator, the same antenna resonates near 6.11 GHz with reduced size. The effective footprint of the antenna is reduced by nearly 64% compared to the conventional patch antenna. The gain of the metamaterial loaded antenna structure is 5.04 dB.

< **Previous** article **Next** article >

Keywords

Circular microstrip patch antenna; Defected ground plane; Metamaterial; Complementary split ring resonator

Choose an option to locate/access this article:

Check if you have access through your login credentials or your institution.

Check Access

or

Purchase

or

› Check for this article elsewhere

Recommended articles Citing articles (3)

Vitae

Prof. Falguni Raval is pursiung her phd in metamaterial inspired antennas. She did her B.E. in Electronics & communication Engg, M.E. in Communication System Engineering. Presently she is working with the V. T. Patel Dept of E & C Engineering in Chandubhai S. Patel Institute of Technology, Changa – Gujarat as Assistant Professor.

Dr Y. P. Kosta is a senior member of IEEE. He has guided number of Ph. D students and M. Tech students. He has published number of papers in international journals and conferences. He is presently director of Marwadi Education Foundation, Rajkot, india.

Harshita Joshi has completed her bachelor of technology in E & C at V. T. Patel Dept of E & C Engineering in Chandubhai S. Patel Institute of Technology. Presently she is pursuing her master of technology from Nirma university.

Copyright © 2015 Elsevier GmbH. All rights reserved.

Recommended articles

Investigation of gain effect of multi-band patch antenna based on composite rectangular SRRs
Optik - International Journal for Light and Electron Optics, Volume 125, Issue 3, 2014, pp. 930-933
Download PDF View details

Analysis of U-slot loaded truncated corner rectangular microstrip patch antenna for broadband operation
AEU - International Journal of Electronics and Communications, Volume 69, Issue 10, 2015, pp. 1483-1488
Download PDF View details

Analysis and design of dual band compact stacked Microstrip patch antenna with defected ground structure for …
AEU - International Journal of Electronics and Communications, Volume 69, Issue 1, 2015, pp. 39-47
Download PDF View details

View more articles ›

Citing articles (3)

Article Metrics

Citations

Citation Indexes:	3

PLUMX

View details ›

ELSEVIER
About ScienceDirect Remote access Shopping cart Contact and support
Terms and conditions Privacy policy

Cookies are used by this site. For more information, visit the cookies page.

Copyright © 2017 Elsevier B.V. or its licensors or contributors. ScienceDirect ® is a registered trademark of Elsevier B.V.

RELX Group™

rexresearch.com

Atmospheric & RF Electricity Collectors

Tesla Generator? - Build Your Own Generator
Eliminate Your Power Bill Easy Do It Yourself. Great Discount Now! Only $49 tesla-energy.org/Generator

Tesla Generator? - Build Your Own Generator
Eliminate Your Power Bill Easy Do It Yourself. Great Discount Now! Only $49 tesla-energy.org/Generator

WIN A $597 Survival Savage Black Out Bag
ENTER NOW

GB 157263

Process and Apparatus for Converting Static Atmospheric Electrical Energy into Dynamic Electrical Energy of any Suitable High Periodicity

Abstract -- Atmospheric electricity, utilizing; influence machines.-Static atmospheric charges are converted into alternating-currents of any desired frequency by rotary machines, the stator and rotor elements of which form condensers having a capacity which varies during the revolution as described in Specincation 157,262. In the simplest form, Fig. 2, the curved stator plates 2, 1 are connected to earth E2 and the aerial A which collects the atmospheric charges and is earthed through a safety spark gap F. The rotor is motor-driven and consists of concentric plates 3, 4 joined through slip-,rings (not shown) to an inductance 9 shunted by a condenser 5. With a given aerial polarity, rotation of the plates 3, 4 causes current reversals on the circuit A, 1, 9, 2, E@, whereby oscillations are set upon the circuit 9, 5, these being tapped by leads 11, 12. The condenser 5 may consist of concentric tubes rotating with the rotor or of two juxtaposed wound spirals having capacity. The stator and rotor condenser plates may extend over nearly a semicircle in place of the quadrants shown in Fig. 2 and the rotor may comprise two consecutive cylindrical plates or coil condensers each divided into halves, the inner half of each being connected to the outer half of the other. Fig. 7 shows a form in which rings 1, 2, mounted in a casing comprising upper and lower parts insulated from each other, are connected to the stator condenser plates at intervals. The rotor plates 3, in equal number, are connected to slip-rings 5, 6 from which alternating-current is taken by brushes when the rotor is motor-driven. The rings 1, 2 are respectively connected to the aerial collector and earth E2. Several methods of grouping the stator-rotor condenser pairs and the components of the oscillating circuit are described. The oscillatory current in the rotor circuit may react through a transformer on the main charging current and two pairs of stator and rotor plates 1, 2, 3, 4, Fig. 13, may be employed arranged on diameters at right-angles, the oscillatory circuit in this case consisting of a mutual inductance 5 having capacity and a portion of an adjustable inductance 9 in the stator earth connector E2. By suitably proportioning the stator and rotor plates, 'resonance may be established, whereby undamped waves may be produced. In the form shown in Fig. 15, for high-frequency oscillations, the aerial A and earth wire are tapped at points between the stator plates 1, 2 and condensers 16, 18 which are joined through an inductance 9. The rotor current is conveyed to the power circuit 11, 12 through transformers 10, 10<a>. The members composing the stator and rotor plates may be slotted, as shown in Fig. 8, with holes 1 for attaching them to the frame, or they may have spiral slots in their surfaces. Alternatively, they may be formed by embedding spiral wires 3 in a vulcanite mass 1 so as to give a smooth surface as shown in Fig. 11.

Fig.1

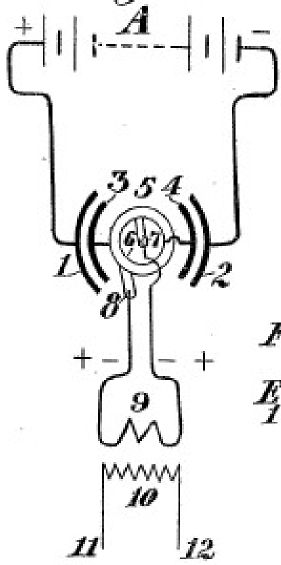
A
+
−
3 5 4
6 6 7
1
2
8
+ −
− +
9
10
11
12

Fig.3

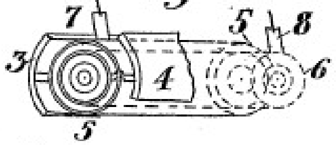

Fig.2

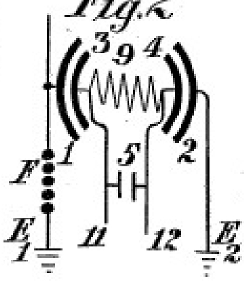
3 9 4
1
5
2
F
E 1
11
12
E 2

Fig 4

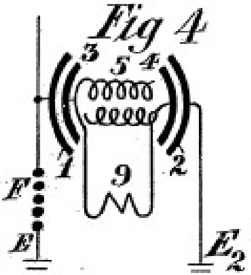
3 5 4
1
9
2
F
E
E 2

Fig.5.

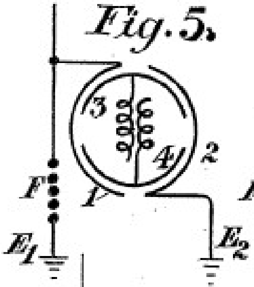

Fig.6.

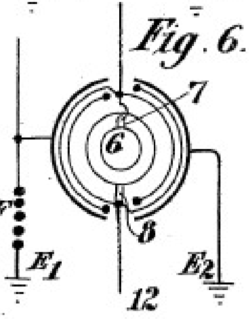
7
6
8
F
E 1
12
E 2

Fig.7.

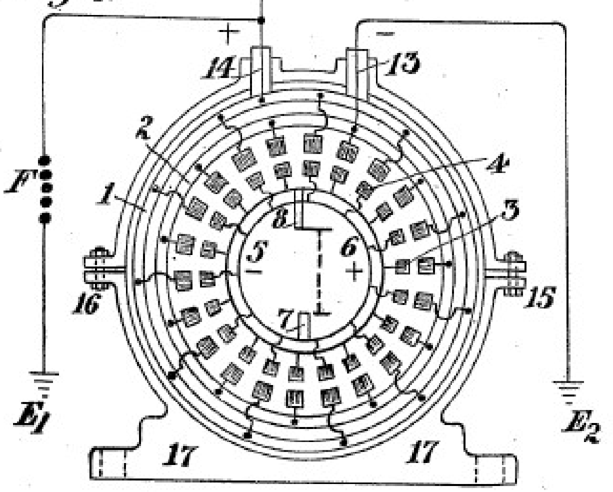
+
−
14
13
2
4
1
3
8
5
−
6
+
7
16
15
F
E 1
E 2
17
17

Fig. 8

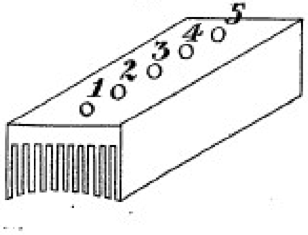

Fig. 10

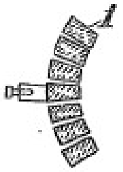

Fig. 9.

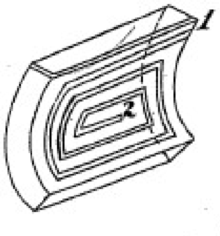

Fig. 11

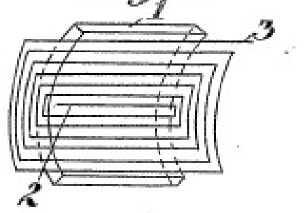

Fig. 12

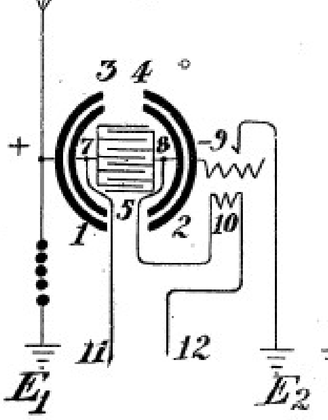
3 4
+
7
8
9
1
5
2
10
11
12
E_1
E_2

Fig. 13.

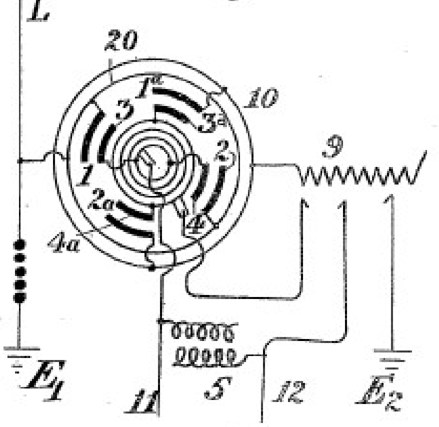
L
20
1^a
10
3
3^a
9
2
1
2a
4
4a
5
11
12
E_1
E_2

Fig. 14

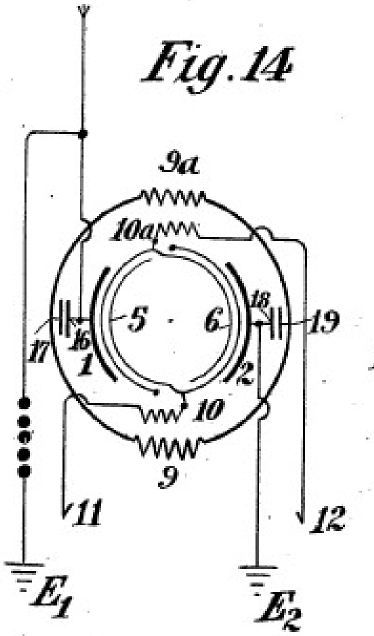
9a
10a
5
6
18
19
17
16
1
2
10
9
11
12
E_1
E_2

Fig. 16

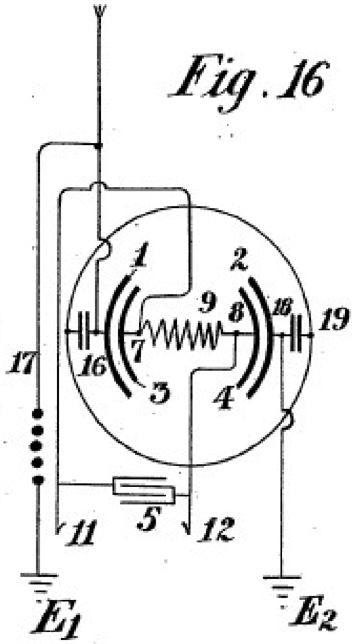
1
2
9
8
18
19
17
16
7
3
4
5
11
12
E_1
E_2

Fig. 15

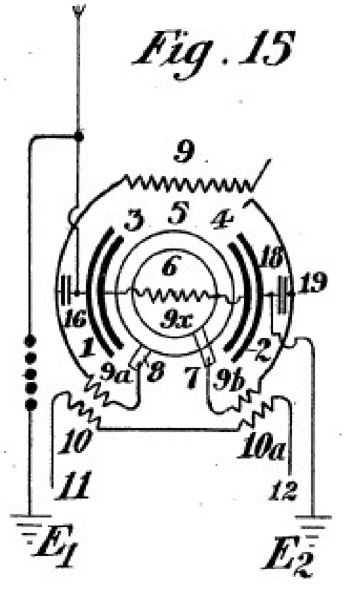
9
3
5
4
6
18
19
16
9x
1
2
9a
8
7
9b
10
10a
11
12
E1
E2

Fig. 17

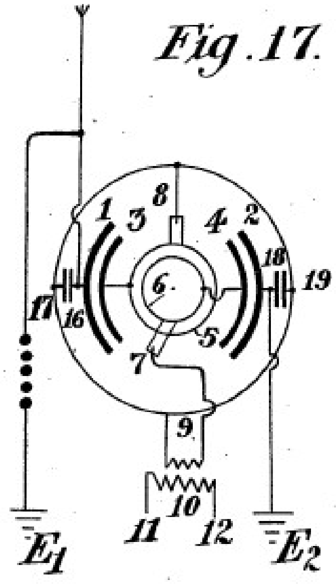
8
1
3
4
2
18
19
17
16
6
5
7
9
10
11
12
E1
E2

French Patent # 551,882

Apparatus for Capture of Electric Currents in the Atmosphere

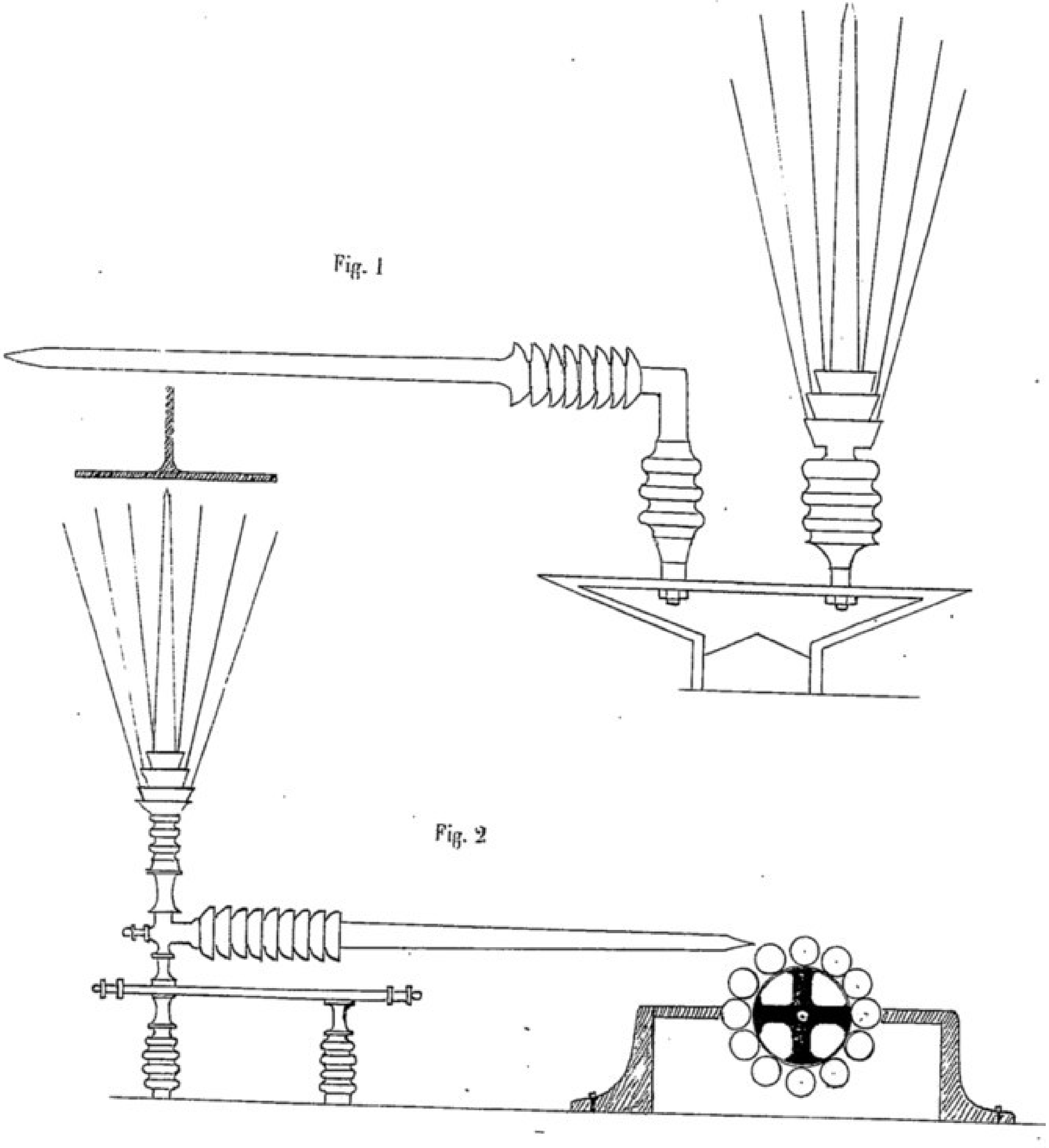

Fig. 1

Fig. 2

French Patent # 565,395

Combined Apparatus for Capture of Atmospheric Electric Currents with Immediate Implementation

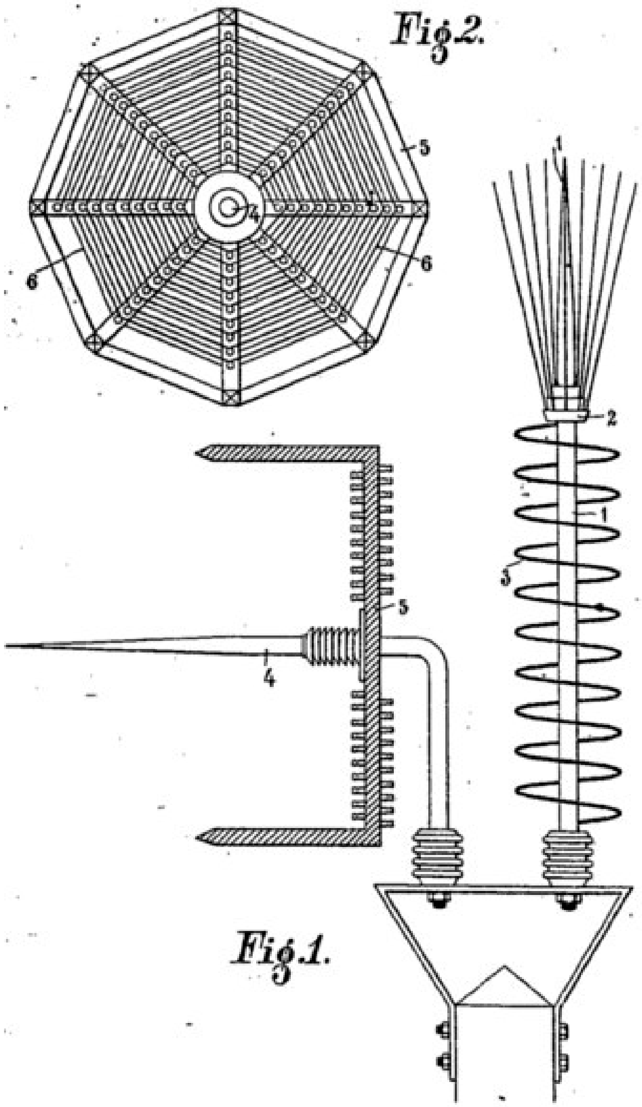
1
5
4
6
6
1
2
1
3
5
4

Fig. 2.

Fig. 1.

Fig. 3.

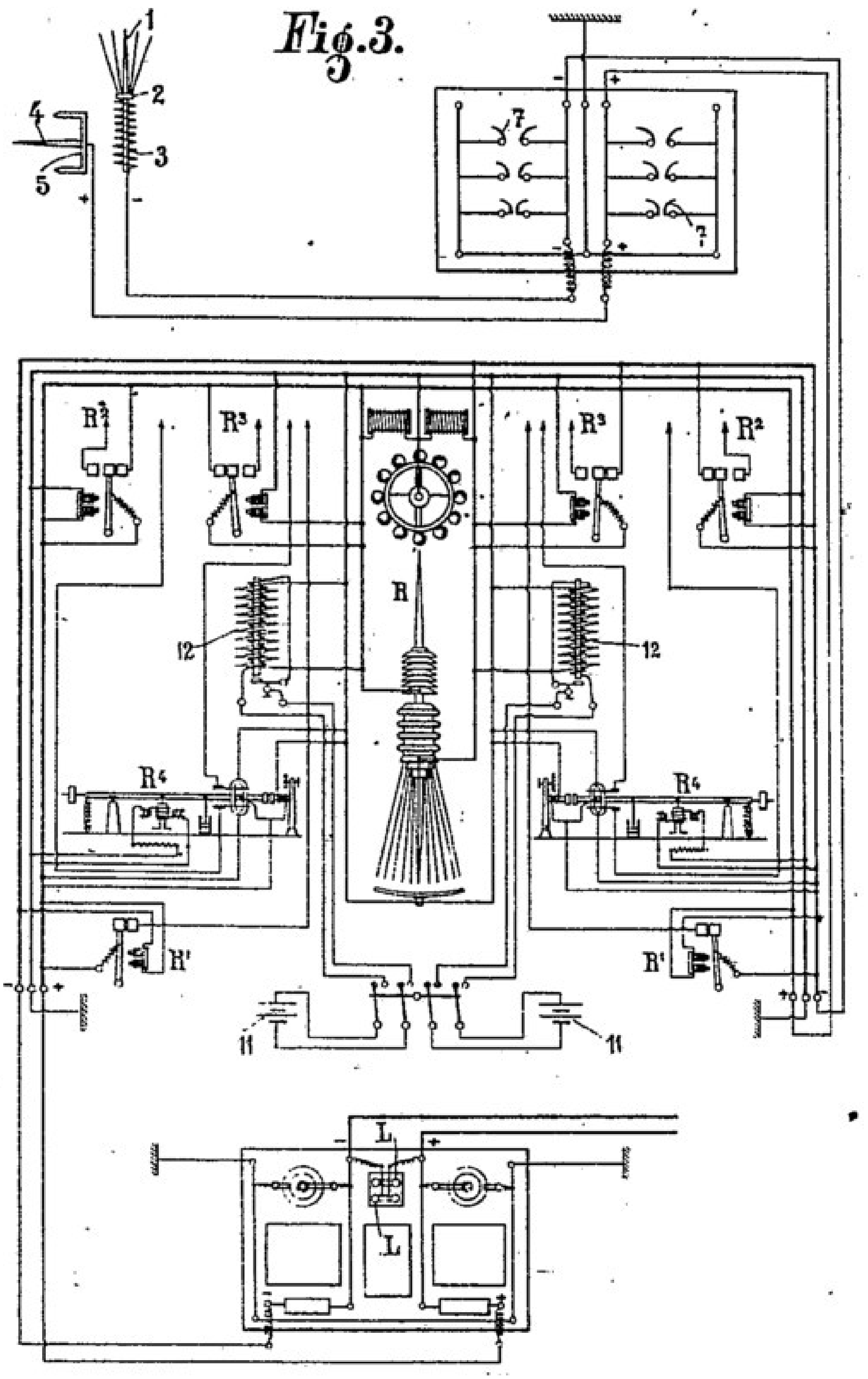
1
2
3
4
5
+
-
7
7
R²
R³
R³
R²
R
12
12
R⁴
R⁴
R¹
R¹
11
11
L
L

Fig. 4.

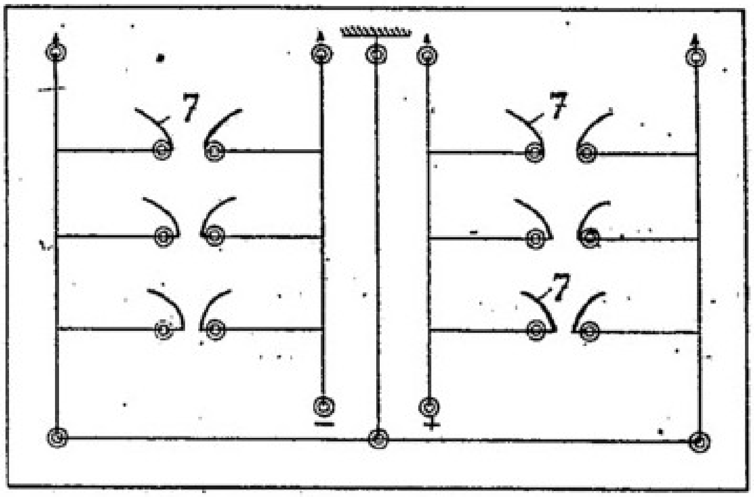

Fig. 5.

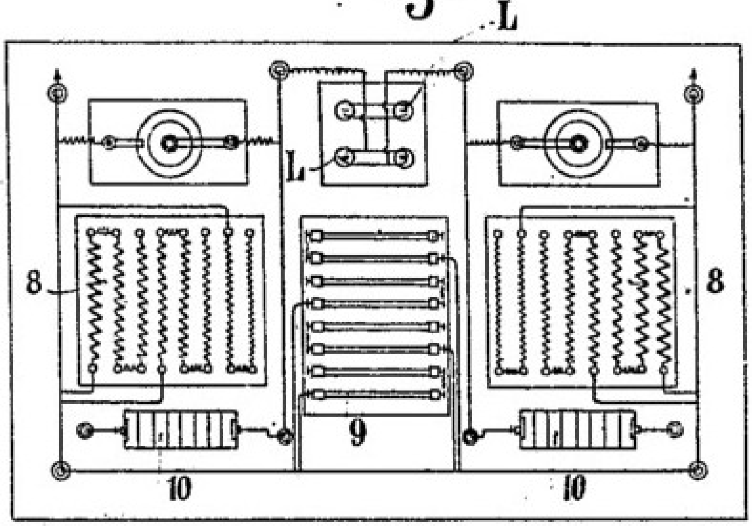

US 28793
Electric Apparatus

Abstract -- An aerostat for obtaining atmpspheric and terrestrial electricity...

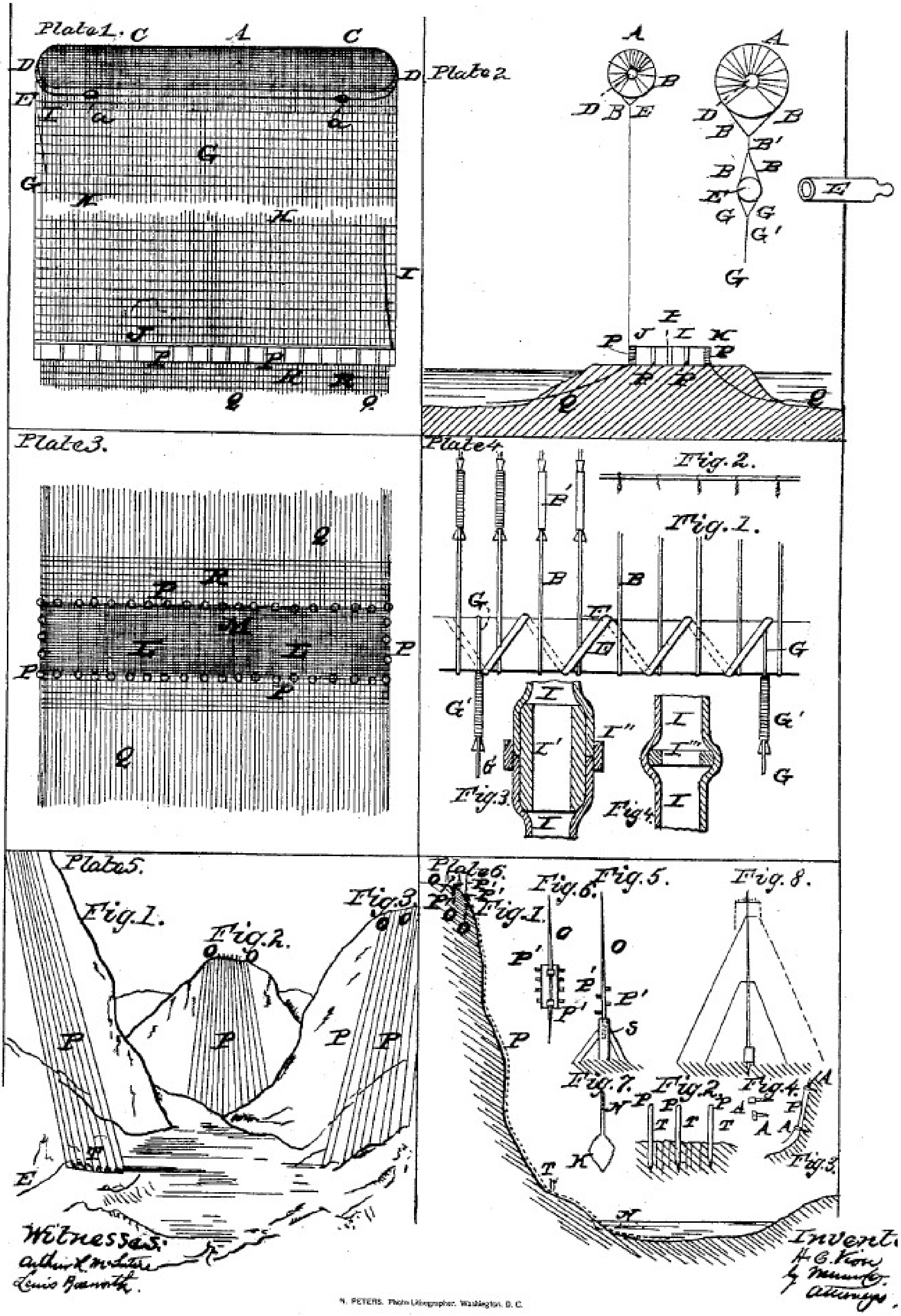
Plate 1.
A
C
D
E
a
G
H
I
J
P
R
Q
Plate 2.
B
D
E
G
G'
B'
J
L
K
Plate 3.
M
Plate 4.
Fig. 1.
Fig. 2.
Fig. 3.
Fig. 4.
E'
G'
I'
I''
Plate 5.
Fig. 1.
Fig. 2.
Fig. 3.
O
T
Plate 6.
Fig. 1.
Fig. 2.
Fig. 3.
Fig. 4.
Fig. 5.
Fig. 6.
Fig. 7.
Fig. 8.
P'
S
N
K
Witnesses:
Invent

N. PETERS, Photo-Lithographer, Washington, D. C.

US 126356
Improvement in Collecting Energy for Telegraph

Abstract -- A tower for the purpose of receiving and imparting natural electricity, so as to be in constant contact with the upper stratum of electricity which surrounds the earth, by tapping which a never-failing supply is formed when brought into contact with the earth.

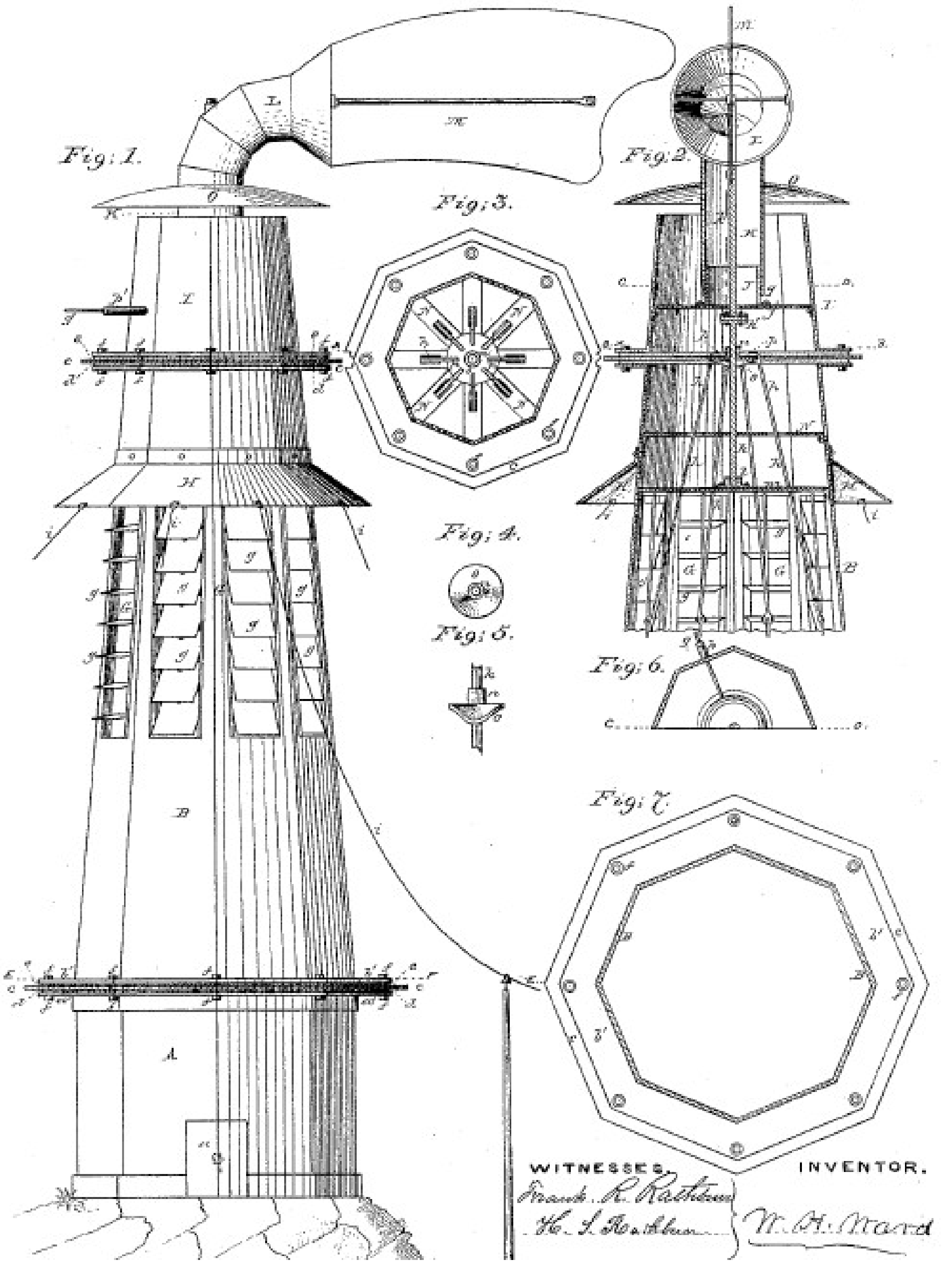
Fig: 1.
Fig: 2.
Fig: 3.
Fig: 4.
Fig: 5.
Fig: 6.
Fig: 7.
WITNESSES.
Frank R. Rathbun
H. L. Rathbun
INVENTOR.
W. H. Ward

US 129971
Telegraph Power

Abstract -- Utilizing natural electricity and establishing an electrical current without the aid of wires, batteries, or cables...

US 674427
Collecting Atmospheric Electricity

Abstract -- An appartus for colecting and driving or conducting atmospheric electricity which renders it practicable to obtain material quantities of teh same in serviceable form.

Fig:1.

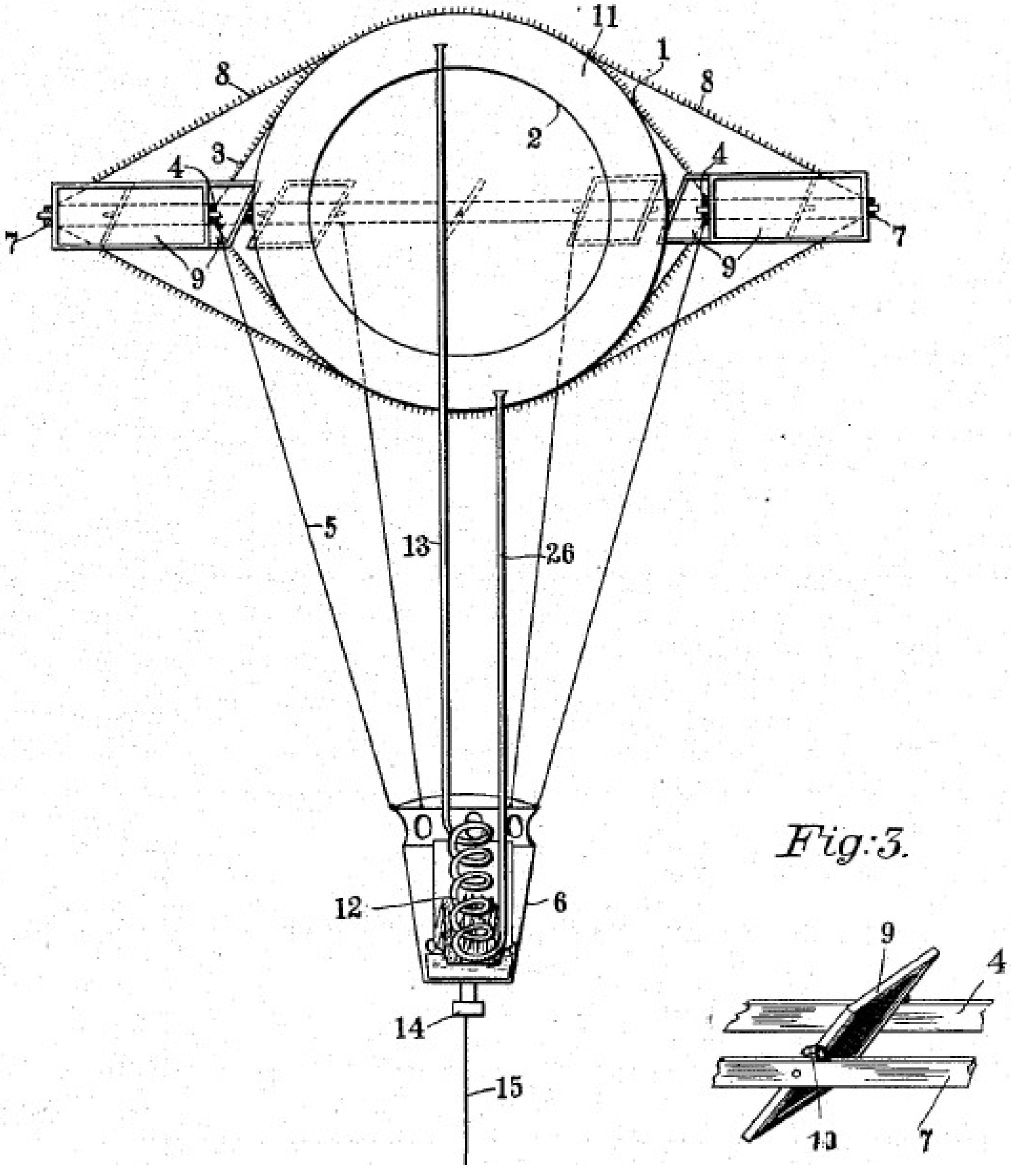

Fig:3.

Fig: 2.

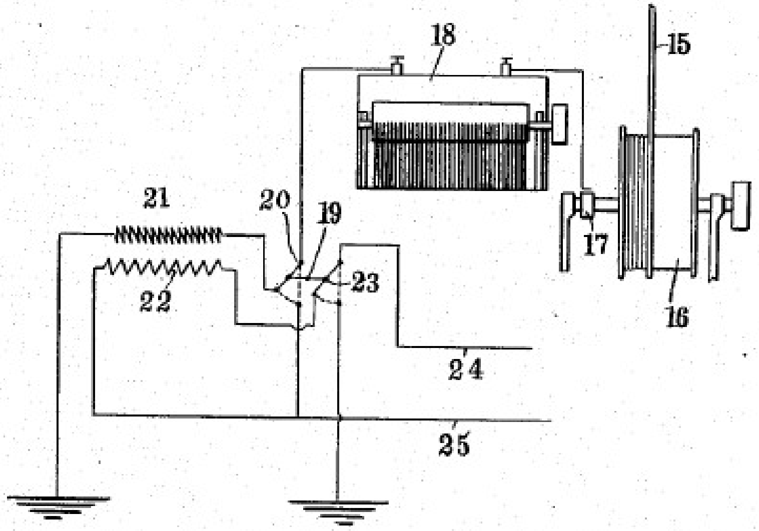

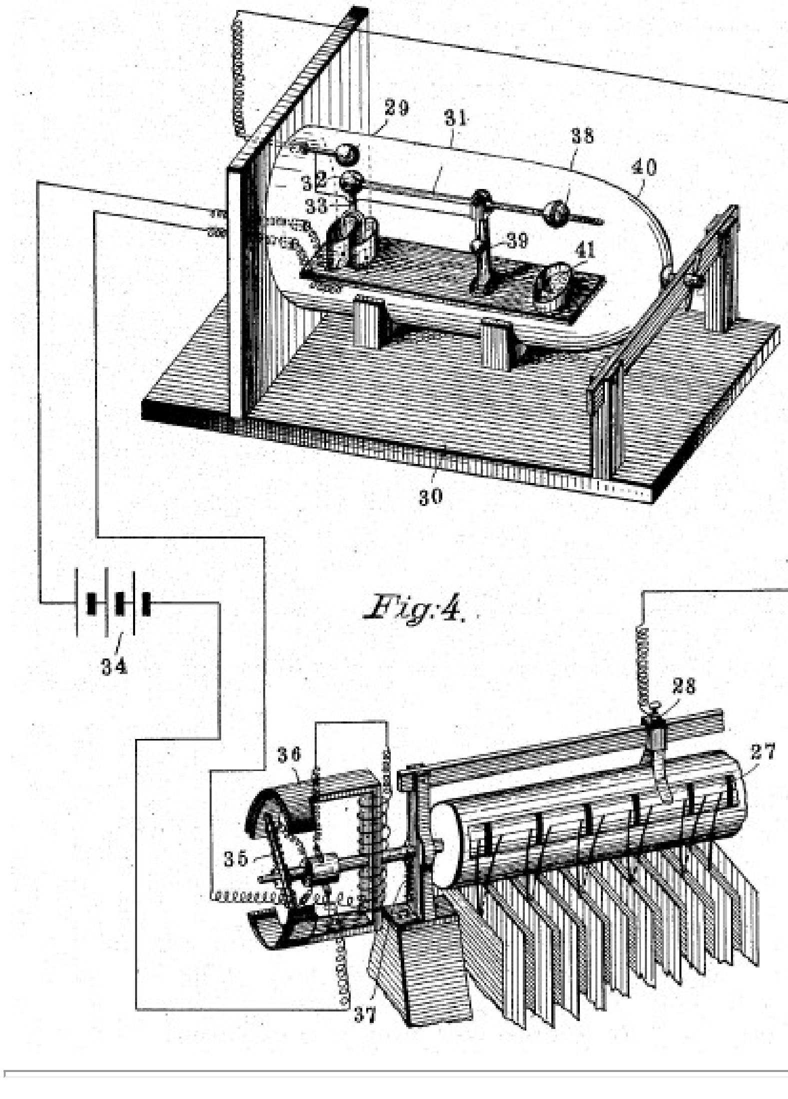

Fig: 4.

US 911260
Collecting Atmospheric Electricity

Abstract -- A method of colecting electricity from a strata laden with electricity at high altitudes in the atmoshere, through the medium of a wire cable suspended from one or more balloons ...

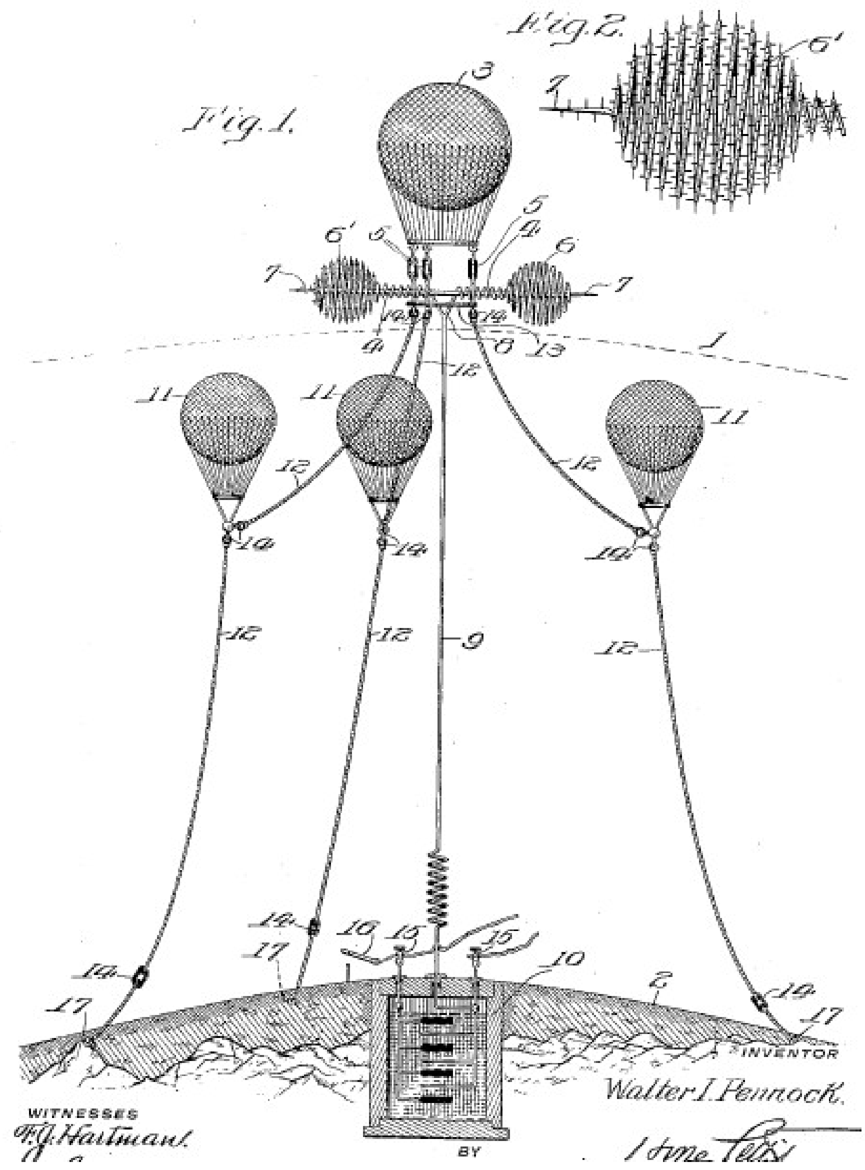
Fig. 2.
6'
7
3
Fig. 1.
5
6'
5
4
6
7
7
14
14
4
8
13
1
12
11
11
11
12
12
14
14
14
12
12
9
12
14
17
16
15
15
10
14
2
14
17
17
INVENTOR
Walter I. Pennock.
WITNESSES
F.G. Hartman
BY

US 1014719
Collecting Atmospheric Electricity

Abstract -- A method of colecting electricity from a strata laden with electricity at high altitudes in the atmoshere, through the medium of a wire cable suspended from one or more balloons ...

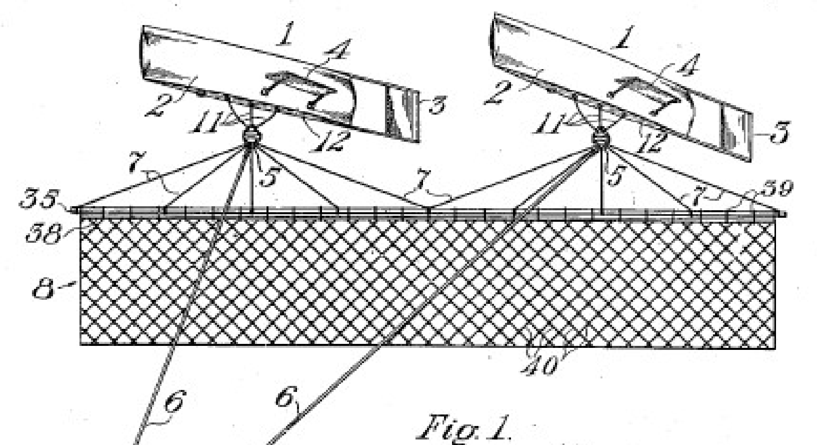
1
2
3
4
11
12
5
7
35
38
39
8
40
6

Fig. 1.

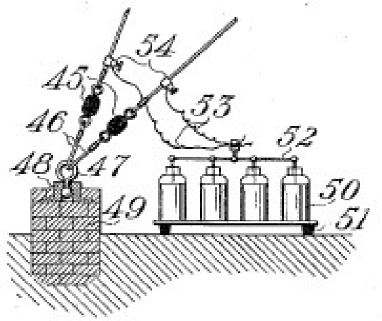
54
45
33
46
52
48
47
50
49
51

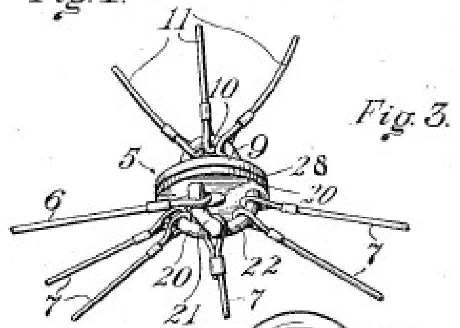
11
10
9
5
28
20
6
7
20
22
21

Fig. 3.

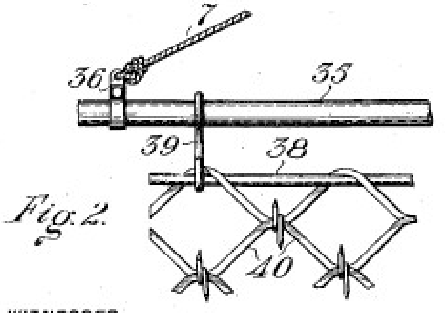
7
36
35
39
38
40

Fig. 2.

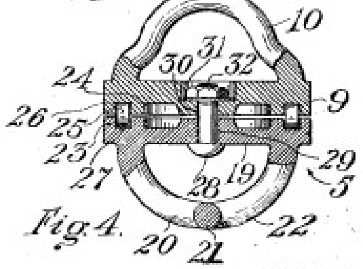
10
24
30
31
32
9
26
25
23
27
28
19
29
5
20
21
22

Fig. 4.

INVENTOR
Walter I. Pennock.

Fig. 5.

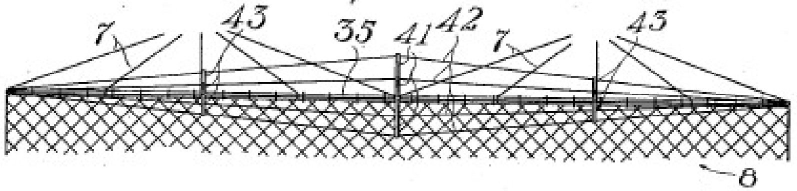
7
43
35
41
42
7
43
8

Fig. 6.

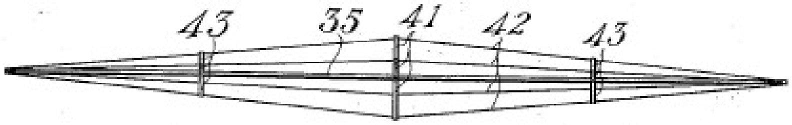
43
35
41
42
43

Fig. 7.

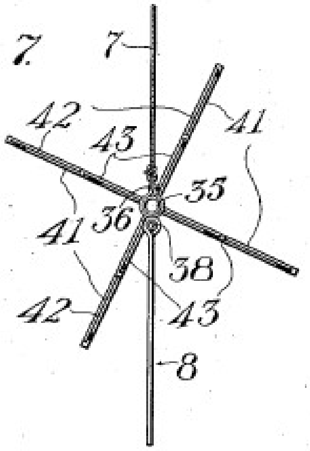
7
42
41
43
35
41
36
38
42
43
8
INVENTOR
Walter I. Pennock.

US 1540998
Conversion of Atmospheric Electric Energy

Abstract -- Method of obtaining atmospheric electricity by means of metallic ballon collectors...

Filed Jan. 13, 1921 12 Sheets-She

Fig. 1.

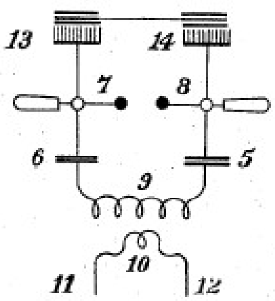
13
14
7
8
6
5
9
10
11
12

Fig. 2.

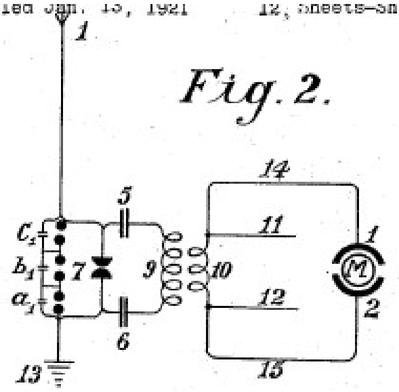
1
14
5
C_1
11
1
b_1
7
9
10
M
a_1
12
2
6
13
15

Fig. 3.

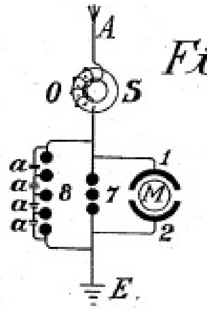
A
O
S
a
a
a
a
8
7
1
M
2
E.

Fig. 4.

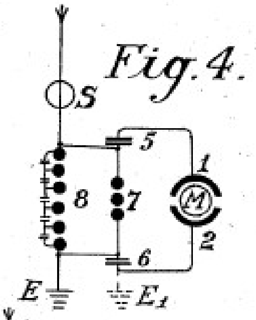
S
5
8
7
1
M
2
6
E
E_1

Fig. 5.

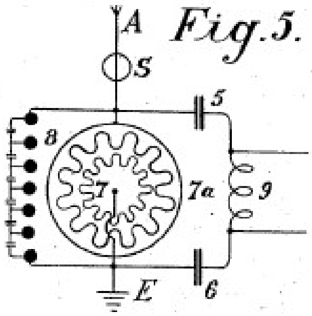
A
S
5
8
7
7a
9
E
6

Fig. 6

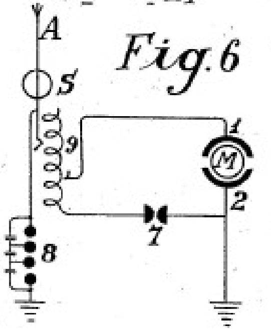

Filed Jan. 13, 1921 12 Sheets-Sheet 2

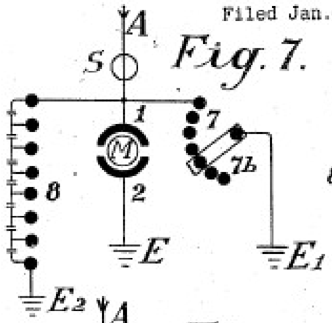
A
S
1
M
2
7
7b
8
E
E1
E2

Fig. 7.

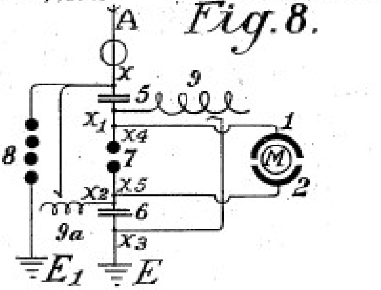
A
x
5
9
x1
x4
1
M
7
8
x5
2
x2
6
9a
x3
E1
E

Fig. 8.

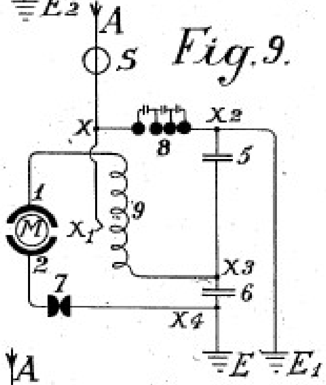
A
S
X
8
X2
5
1
M
X1
9
2
X3
7
6
X4
E
E1

Fig. 9.

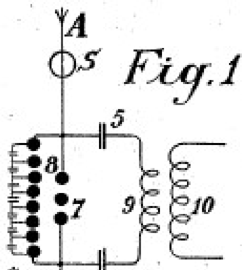

Fig. 10.

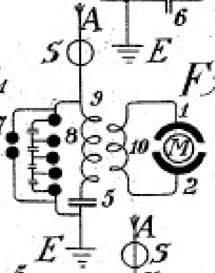
A
S
9
1
7
8
10
M
2
5
E

Fig. 11.

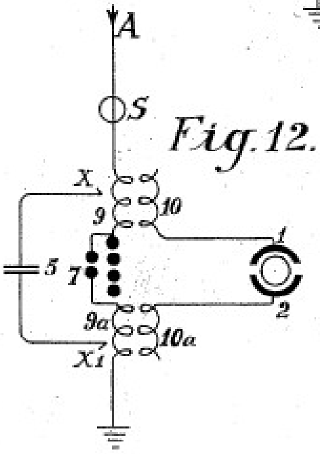
A
S
X
9
10
1
5
7
2
9a
10a
X1

Fig. 12.

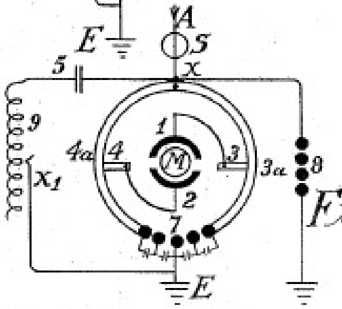
A
S
5
X
9
1
4a
4
M
3
3a
8
X1
2
7
E

Fig. 13.

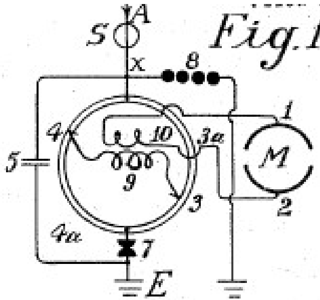
S
A
x
8
1
4
10
3a
M
5
9
3
2
4a
7
E

Fig. 14.

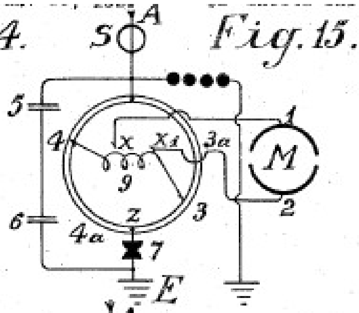
S
A
5
1
4
x
x1
3a
M
9
3
2
6
z
4a
7
E

Fig. 15.

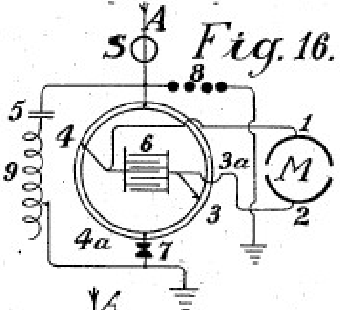
S
A
8
5
1
4
6
3a
M
9
3
2
4a
7

Fig. 16.

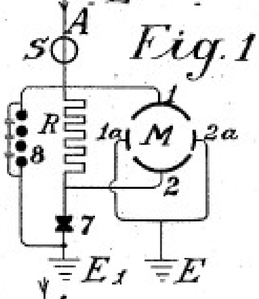
S
A
R
1a
M
2a
8
2
7
E1
E

Fig. 17.

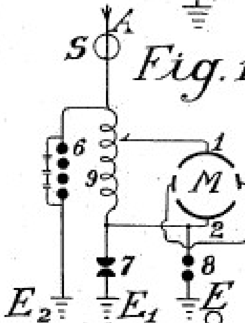
S
A
6
1
M
9
2
7
8
E2
E1
E

Fig. 18.

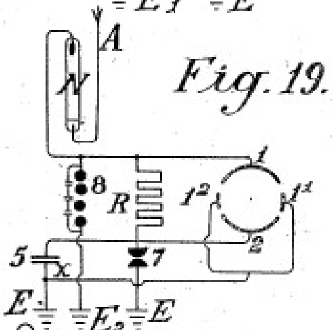
N
A
1
8
R
1^2
1^1
2
5
x
7
E
E2
E

Fig. 19.

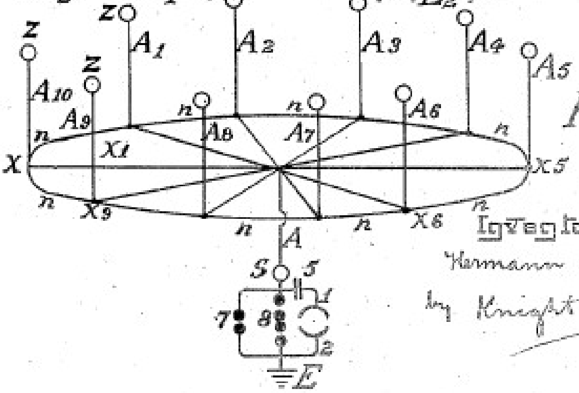
z
z
z
A1
A2
A3
A4
A5
A10
A9
n
A8
A7
A6
x
x1
x5
x9
x6
A
S
5
7
8
1
2
E

Fig. 20.

Inventor
Hermann Plauson
by Knight Bros
Attorneys

Fig. 21.

Fig. 22.

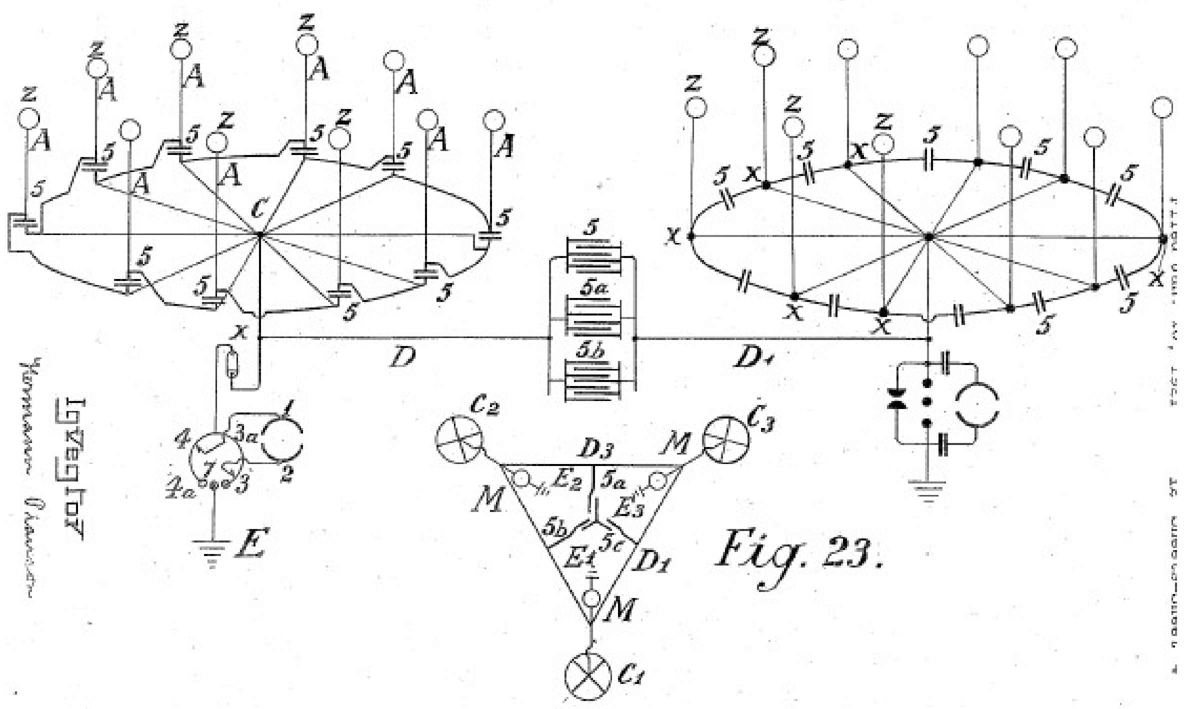
Z
A
5
C
x
D
D1
5
5a
5b
1
2
3
3a
4
4a
7
E
X
C2
C3
C1
D3
M
E2
E3
E1
5a
5b
5c
D1
Inventor
Hermann Plauson

Fig. 23.

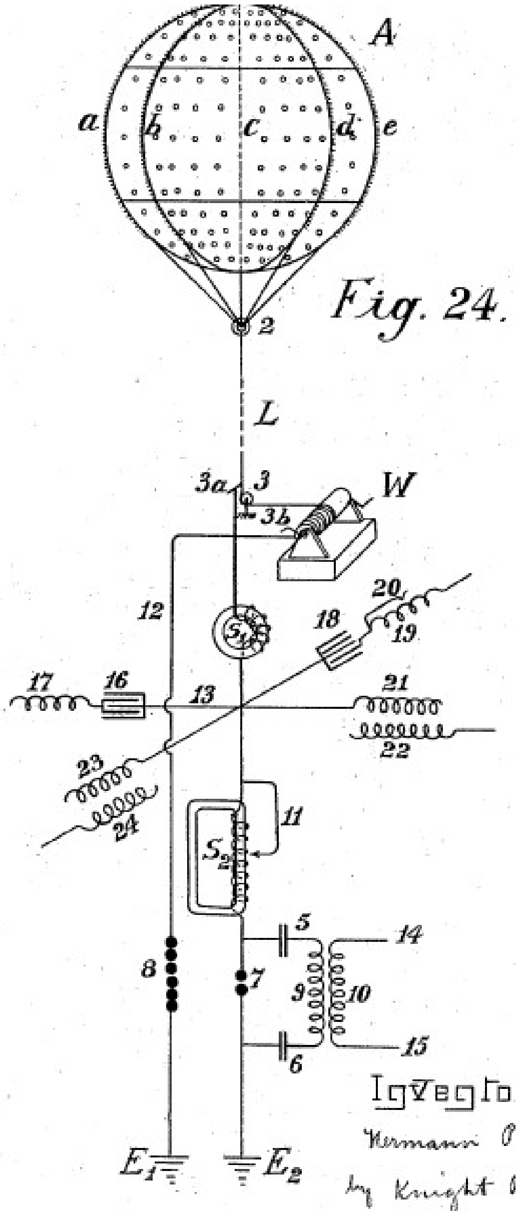
A
a
b
c
d
e
2
L
3a
3
3b
W
12
S1
20
18
19
17
16
13
21
22
23
24
11
S2
5
14
8
7
9
10
6
15
E1
E2
Invento
Hermann
by Knight

Fig. 24.

Fig. 25.

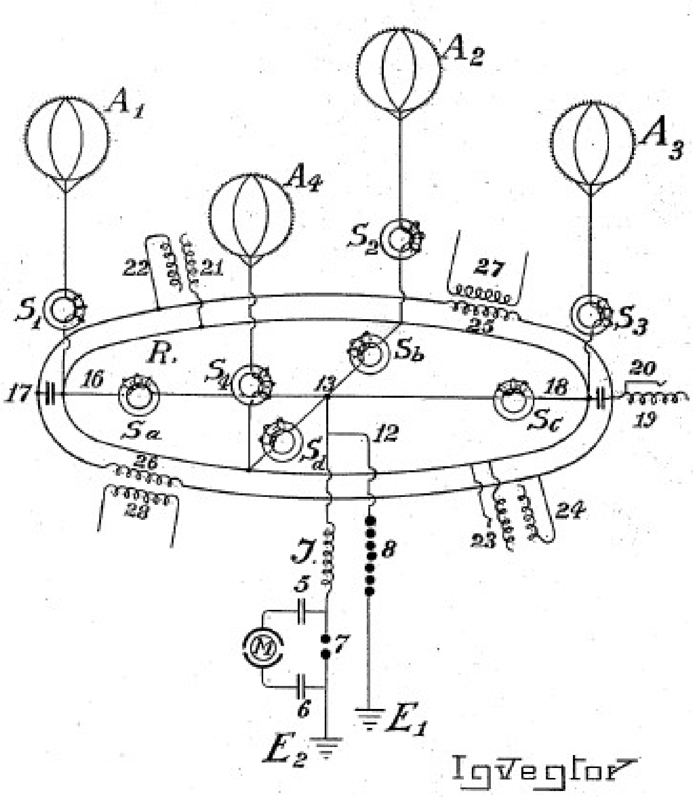
A_1
A_2
A_3
A_4
S_1
S_2
S_3
S_4
S_a
S_b
S_c
S_d
R.
22
21
27
25
17
16
13
18
20
19
12
26
28
23
24
J.
8
5
M
7
6
E_1
E_2
Inventor

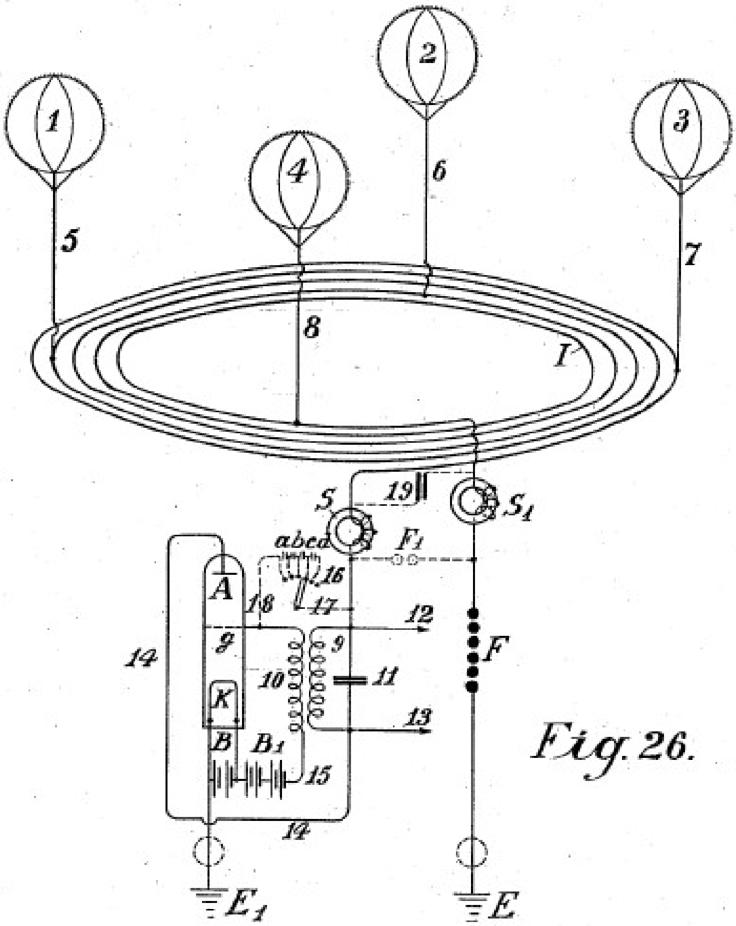
I
S
S_1
abcd
F_1
A
g
K
B
B_1
F
E_1
E

Fig. 26.

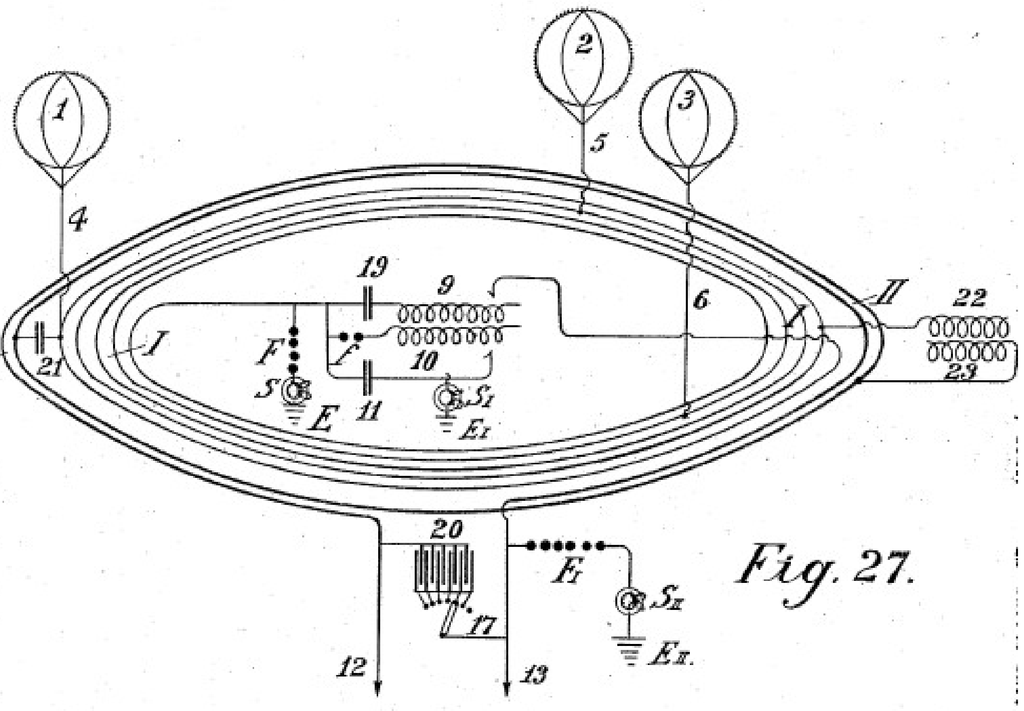
1
2
3
4
5
6
19
9
22
II
II
21
I
I
F
f
S
E
10
11
S_I
E_I
23
20
F_I
S_II
17
E_II.
12
13

Fig. 27.

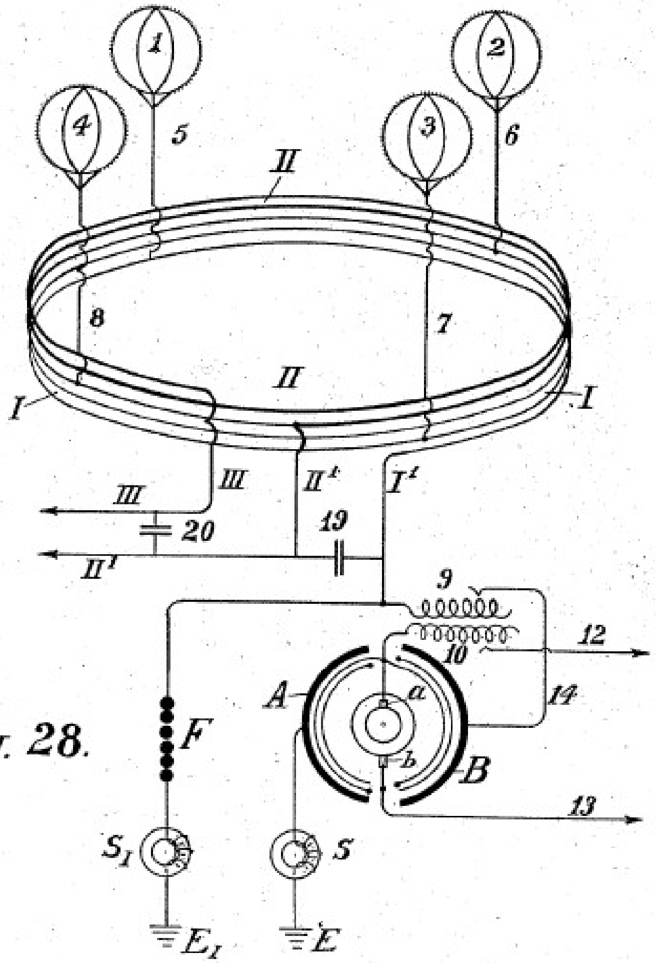
1
2
4
5
3
6
II
8
7
II
I
I
III
III
II^1
II^1
I^1
20
19
9
12
10
A
a
14
B
b
13
F
S_I
S
E_I
E

Fig. 28.

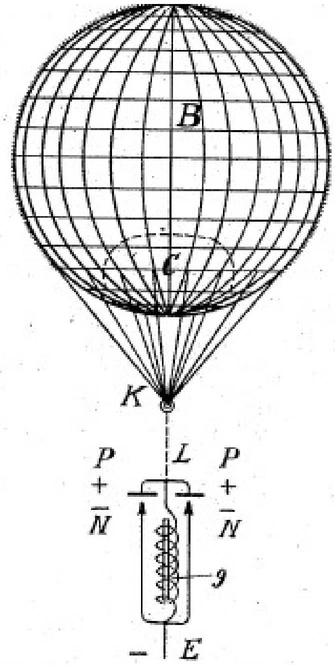
B
C
K
P
L
P
+
+
N
N
9
−
E

Fig. 29.

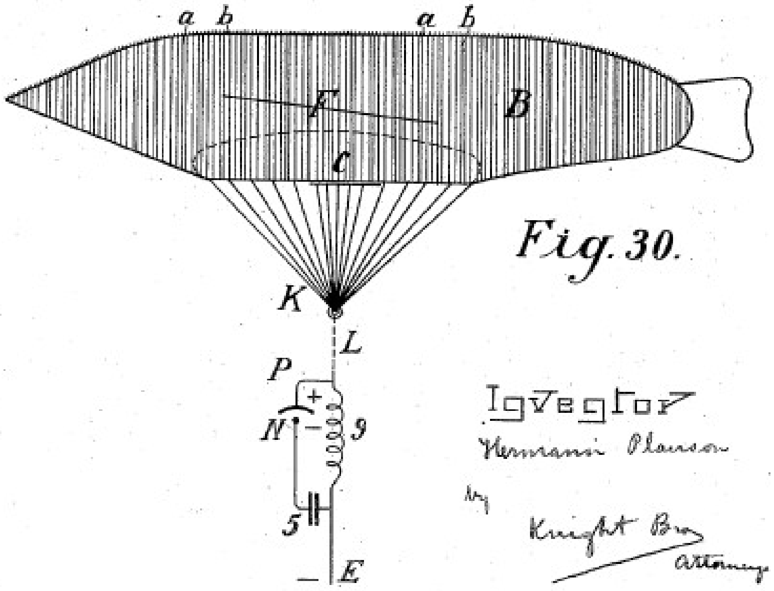
a
b
a
b
F
B
C
K
L
P
+
N
−
9
5
−
E

Fig. 30.

Inventor
Hermann Plauson
by
Knight Bros
Attorneys

Filed Jan. 13, 1921 12 Sheets-Sheet 11

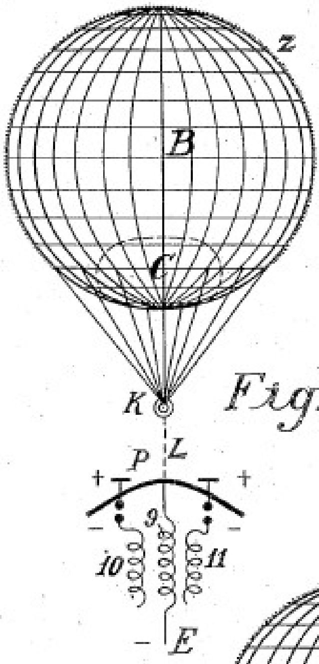
z
B
C
K
P
L
+
+
−
−
9
10
11
−
E

Fig. 31.

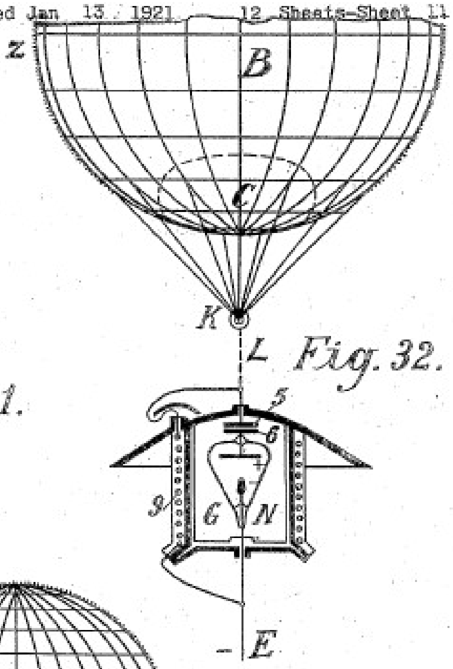
z
B
C
K
L
5
6
9
6
N
−
E

Fig. 32.

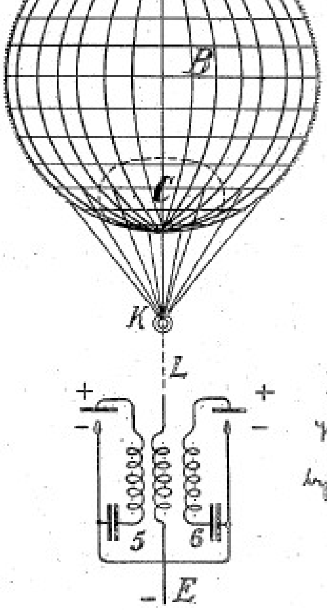

Fig. 33.

Inventor
Hermann Plauson
by Knight Bros
Attorneys

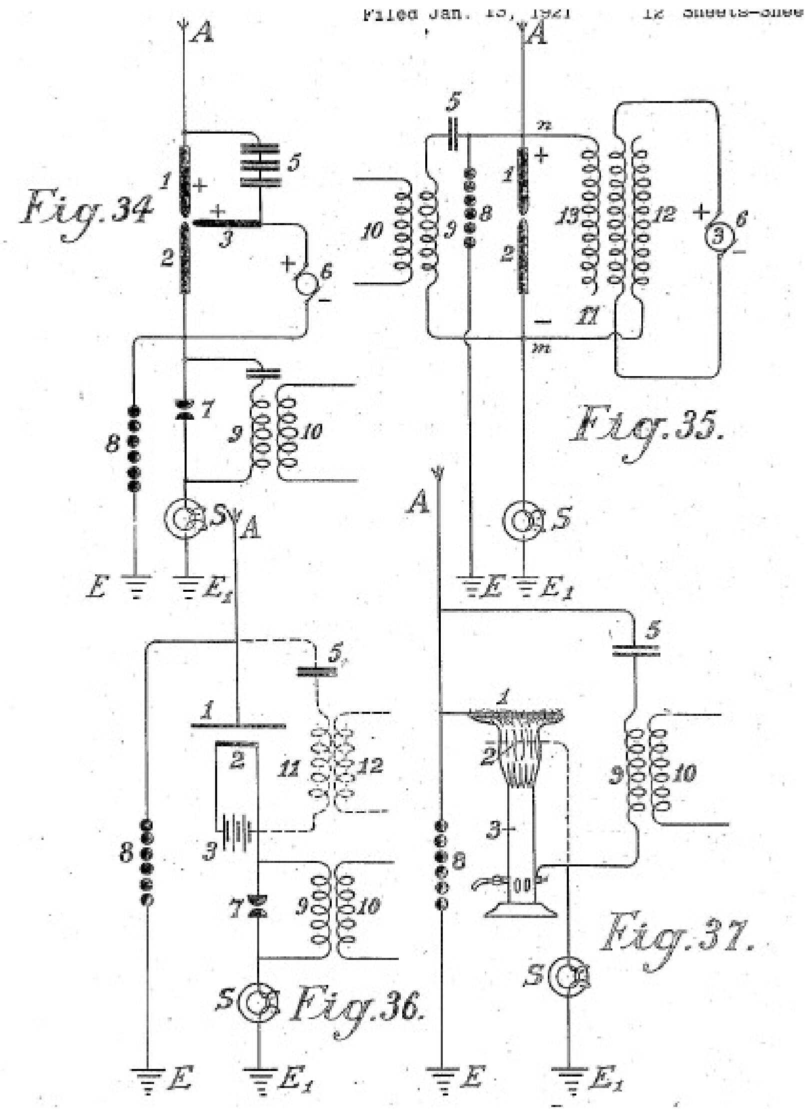
Filed Jan. 15, 1921 12 Sheets-Shee
Fig. 34
Fig. 35.
Fig. 36.
Fig. 37.
A
1
2
3
5
6
7
8
9
10
11
12
13
n
m
S
E
E₁
+
−

GB 149,917
Improvements in & Relating to Radio Communication Systems

Abstract -- Radio transmission and reception through the use of living vegetable organisms such as trees, plants, and the like.

Fig 1

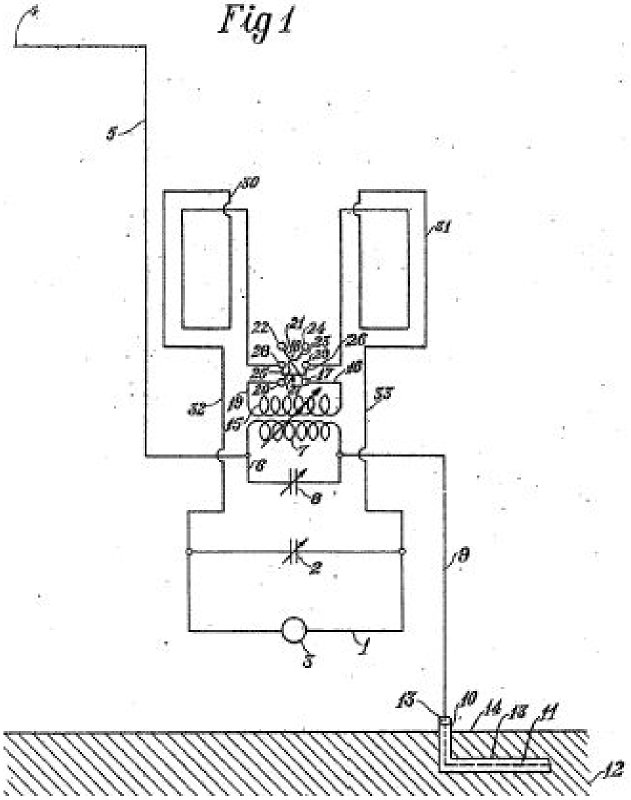
5
30
31
22
21
24
20
25
28
26
39
27
17
16
29
32
19
15
33
7
6
8
2
9
3
1
13
10
14
13
11
12

US 1826727
Radio Apparatus

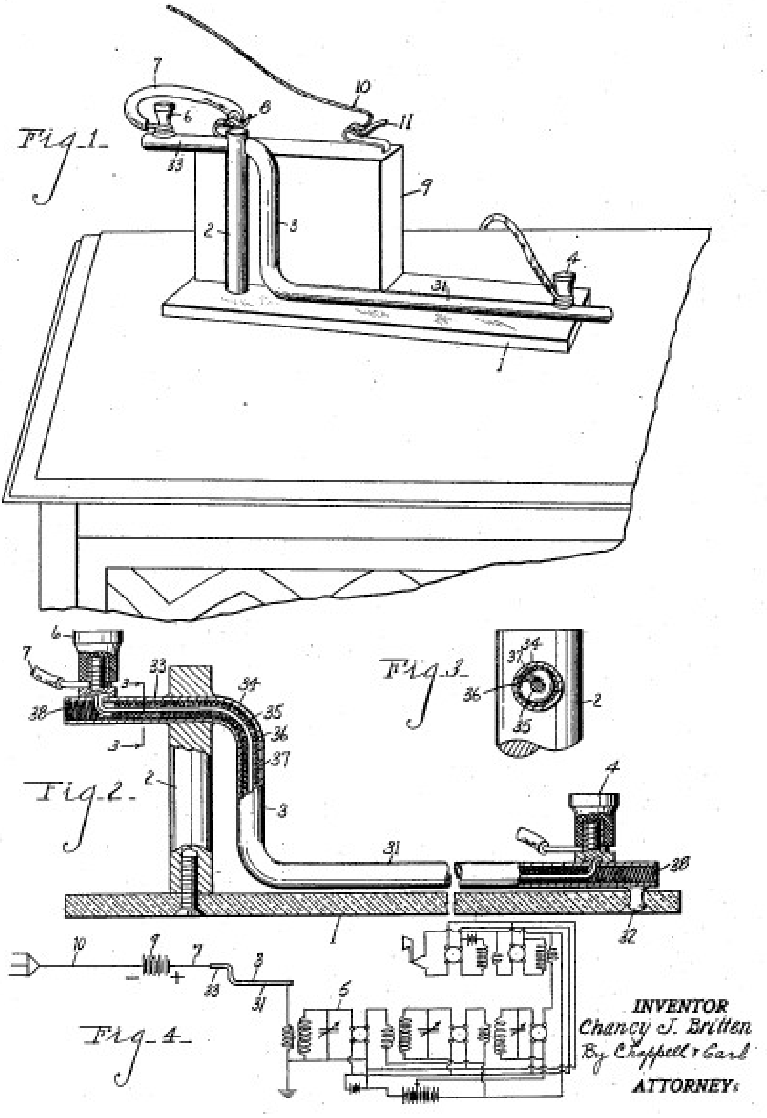
Fig.1.
Fig.2.
Fig.3.
Fig.4.
1
2
3
4
5
6
7
9
10
11
31
32
33
34
35
36
37
38
INVENTOR
Chancy J. Britten
By Chappell & Earl
ATTORNEYS

US 2143437
Radiant Energy System

Abstract -- An apparatus for receiving and transmitting radiant energy within a wide range of frequencies without necessity of a source of local energy.

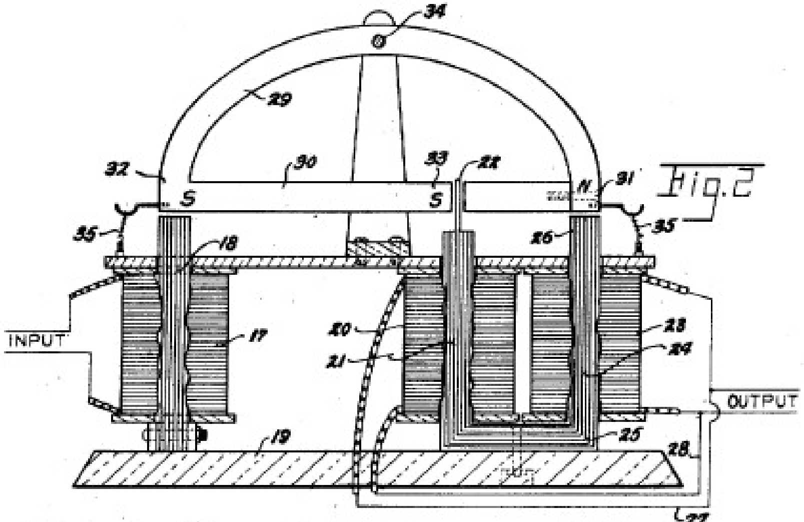
34
29
32
30
33
22
31
S
S
N
35
35
26
18
INPUT
17
20
21
23
24
OUTPUT
19
25
28
27

Fig. 2

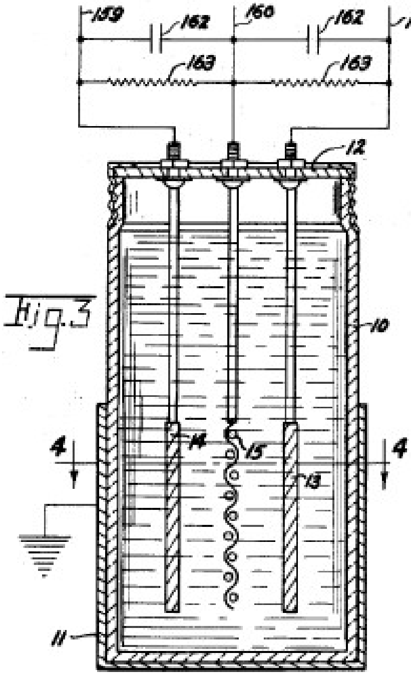
159
162
160
162
161
163
163
12
10
4
14
15
13
4
11

Fig. 3

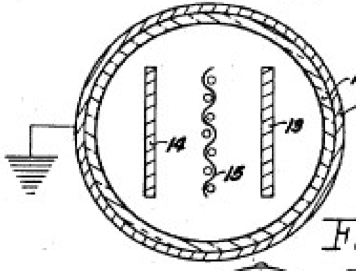
10
11
14
15
13

Fig. 4

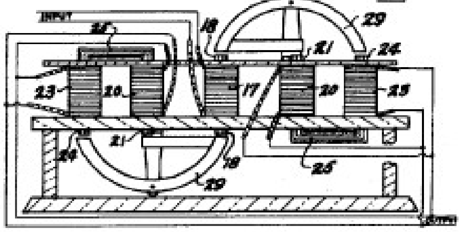
INPUT
22
18
29
21
24
23
20
17
20
23
21
18
24
25
29

Fig. 9

Inventor

James Marion Fisher

By Murray & Zugelter

Attorney

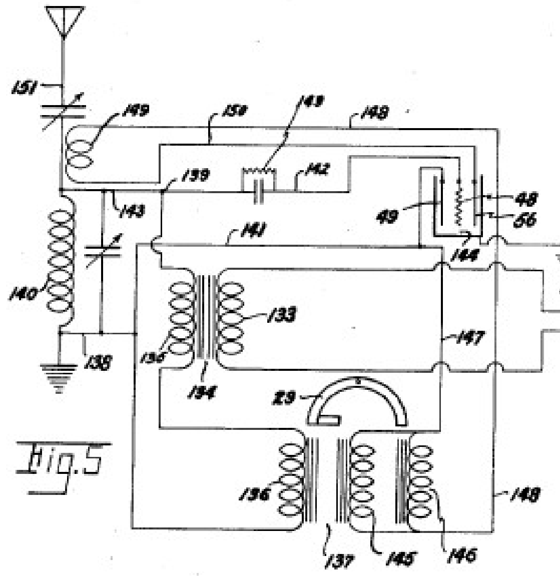
151
149
150
143
148
139
142
143
49
48
56
144
141
140
133
135
134
138
147
152
131
132
153
29
136
137
145
146
148

Fig.5

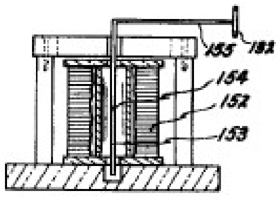
155
152
154
152
153

Fig.7

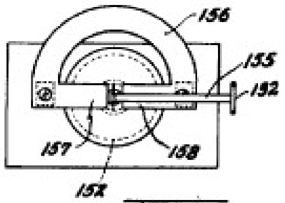
156
155
152
157
158
152

Fig.8

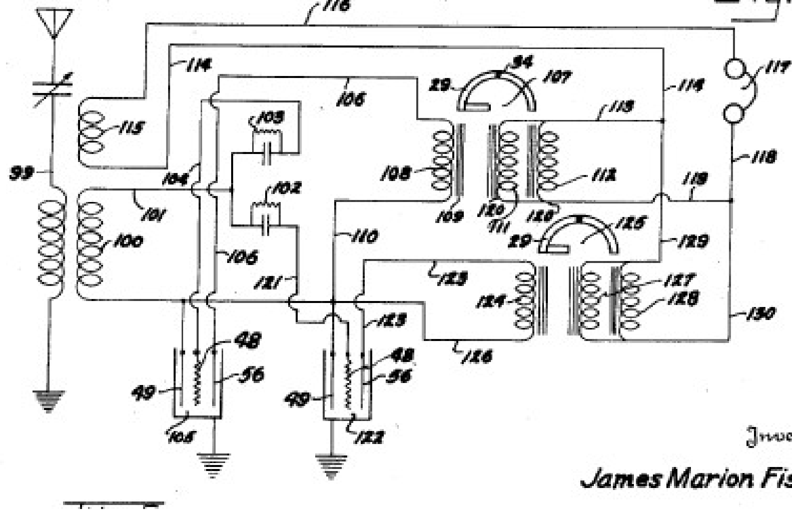
116
114
117
115
106
29
34
107
114
113
103
99
104
108
112
118
101
102
109
120
111
126
119
100
106
110
29
125
129
121
123
124
127
128
125
130
48
56
49
105
48
56
49
122
126

Inventor

James Marion Fisher

US 2813242

Powering Electrical Devices with Energy Abstracted from the Atmosphere

Abstract -- A convenient and economical provision of power for the operation of electronic circuits and devices using transistors, and of other electrical devices hving modest power requirements.

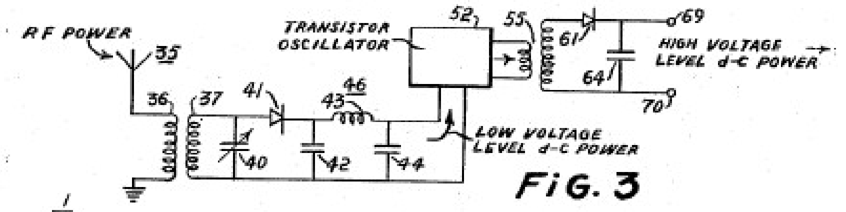
RF POWER
35
36
37
41
43
46
40
42
44
TRANSISTOR OSCILLATOR
52
55
61
64
69
70
HIGH VOLTAGE LEVEL d-C POWER
LOW VOLTAGE LEVEL d-C POWER

FiG. 3

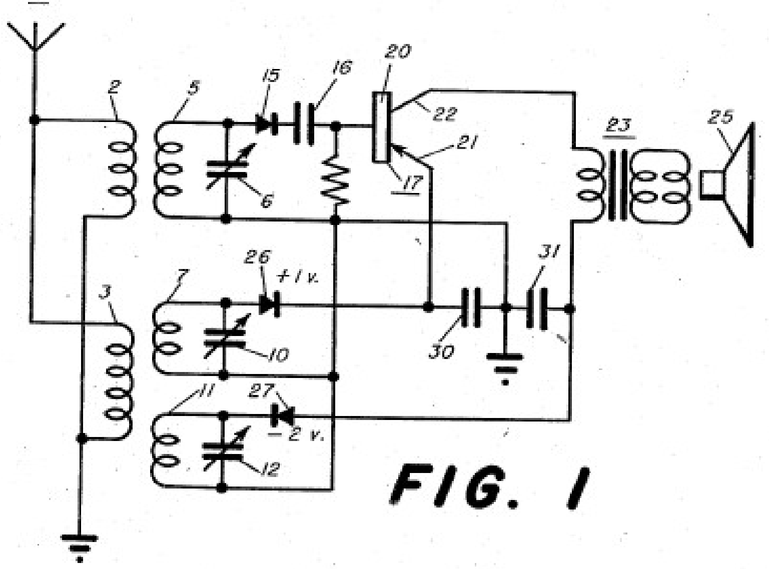
1
2
3
5
6
7
10
11
12
15
16
17
20
21
22
23
25
26
+1 v.
27
-2 v.
30
31

FIG. 1

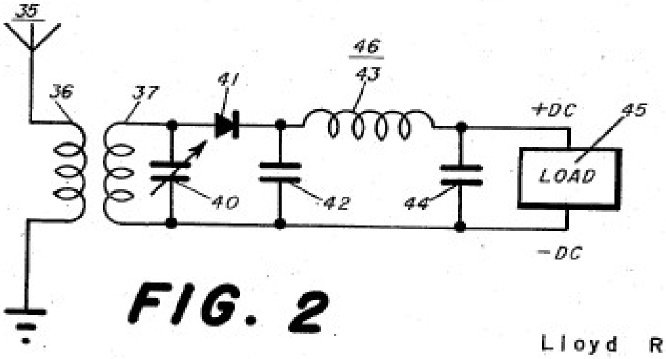
35
36
37
41
46
43
40
42
44
+DC
45
LOAD
-DC

FIG. 2

INVENTOR
Lloyd R. Crump

US 3205381
Ionospheric Battery

Abstract -- A method and apparatus for tapping the enormous reservoir of energy existing in the ionosphere...

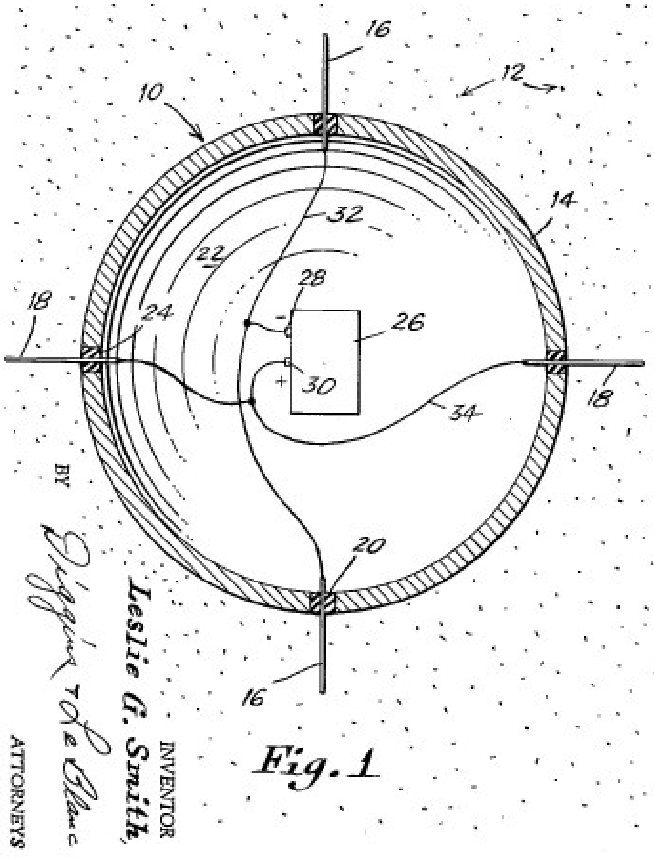
16
12
10
14
32
22
28
26
24
18
30
34
18
20
16
INVENTOR
Leslie G. Smith,
BY
Diggins & LeBlanc
ATTORNEYS

Fig. 1

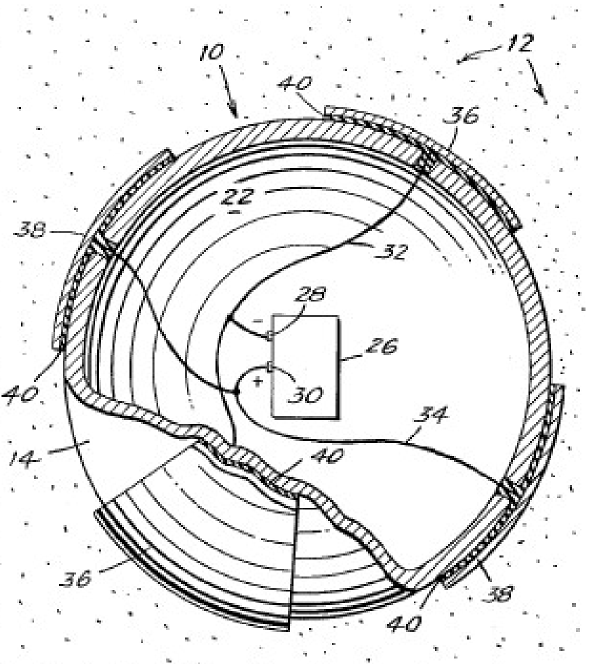
10
40
12
36
22
38
32
28
26
30
40
34
14
40
36
38
40

Fig. 2

US 3273066

Apparatus for Detecting Changes in the Atmospheric Electric Field

Abstract -- A cloud detector and or an apparatus for sensing electrostatic conditions in the earth's atmosphere...

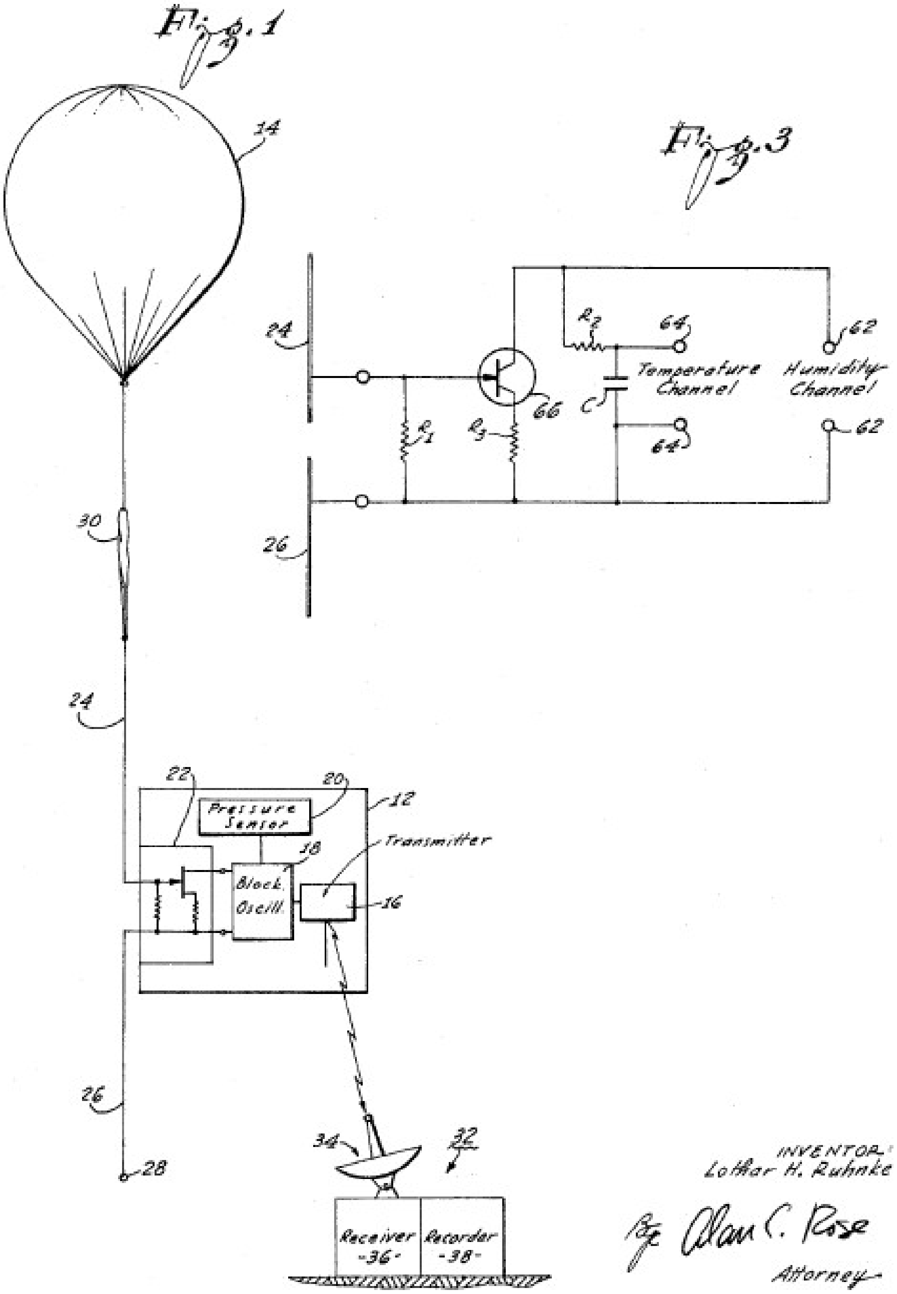
Fig. 1
14
Fig. 3
R2
64
62
24
Temperature Channel
Humidity Channel
C
66
R1
R3
64
62
26
30
24
22
20
12
Pressure Sensor
Transmitter
18
Block. Oscill.
16
26
34
32
28
Receiver -36-
Recorder -38-
INVENTOR
Lothar H. Ruhnke
By Alan C. Rose
Attorney

US 3760257
EM Wave Energy Converter

Abstract -- Electromagnetic wave energy is converted into electric power with an array of mutually insulated electromagnetic wave absorber elements each responsive to an electric field component of the wave as it impinges thereon. Each element includes a portion tapered in the direction of wave propagation to provide a relatively wideband response spectrum. Each element includes an output for deriving a voltage replica of the electric field variations intercepted by it. Adjacent elements are positioned relative to each other so that an electric field subsists between adjacent elements in response to the impinging wave. The electric field results in a voltage difference between adjacent elements that is fed to a rectifier to derive d.c. output power. The element pairs may be arranged in a two-dimensional array to provide power conversion of randomly polarized electromagnetic waves, such as sunlight.

FIG. 1

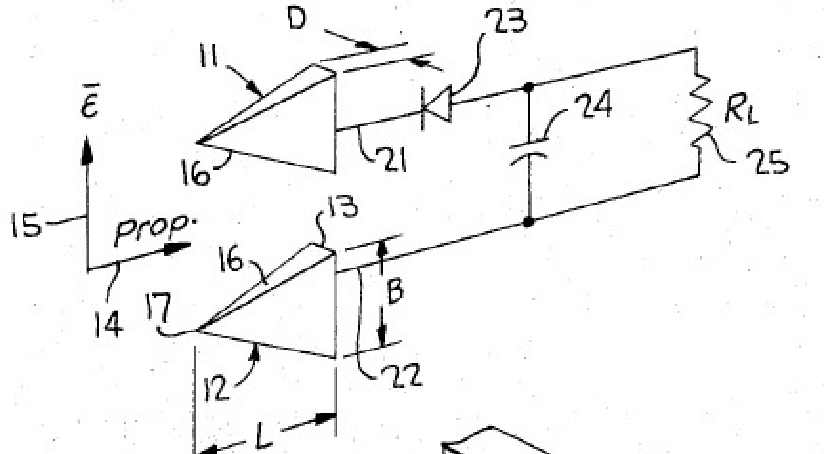
D
23
11
Ē
24
R_L
16
21
25
15
Prop.
13
16
B
14
17
12
22
L

FIG. 2

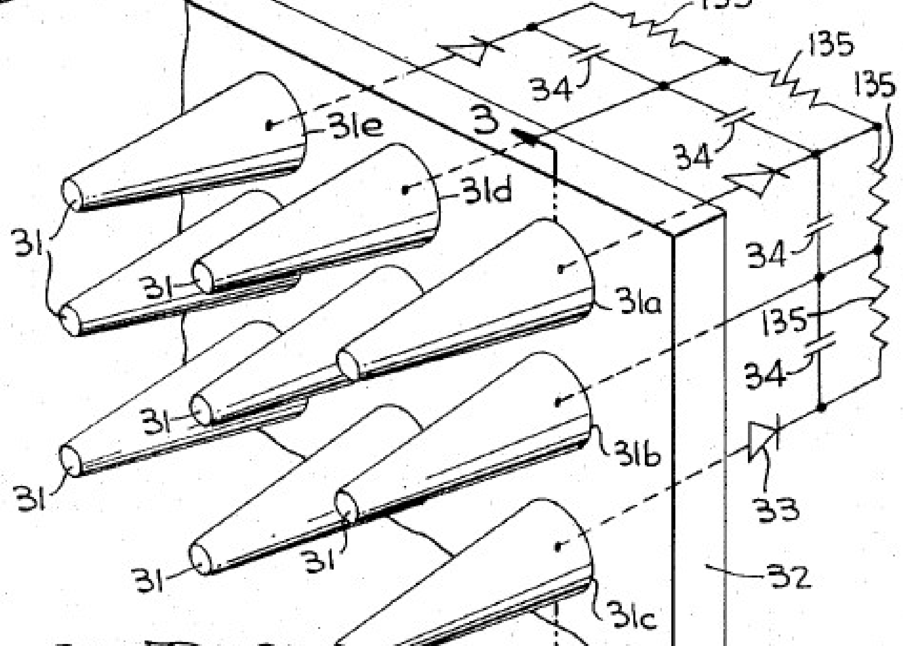
135
135
135
34
3
34
31e
31d
31
31
31a
34
135
34
31
31
31b
33
31
31
32
31c
31
3

FIG. 3

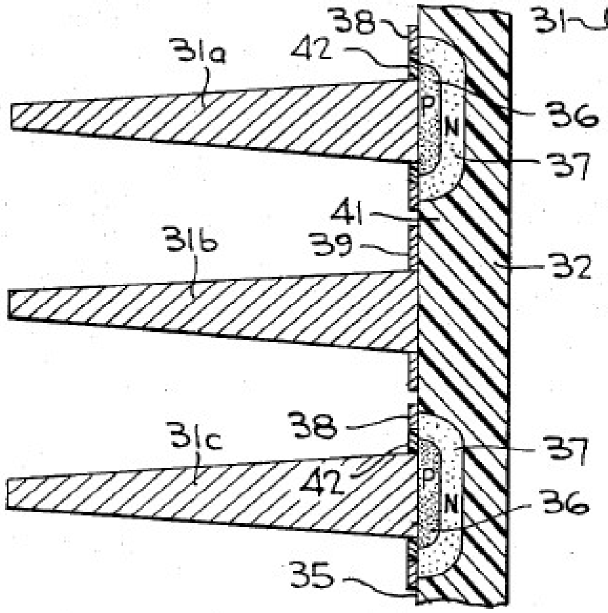
38
31a
42
P
N
36
37
41
31b
39
32
38
31c
37
42
P
N
36
35

US 3931577
Radio Receiver Protection Arrangement

Abstract -- A receiver protection arrangement in which a voltage dependent on aerial voltage and a voltage dependent on aerial current are aggregated and applied to means responsive to a predetermined aggregate level to effectively short circuit the aerial lead. Additional means may be provided to dissipate energy in the input stage of a protected receiver to cause a reduced impedance to be reflected across the output of the protective arrangement.

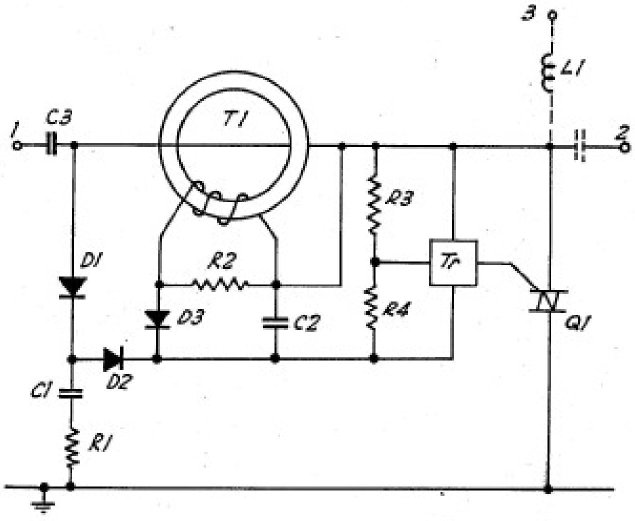
3
L1
1
C3
T1
2
R3
D1
R2
Tr
D3
C2
R4
Q1
D2
C1
R1

Fig. 1.

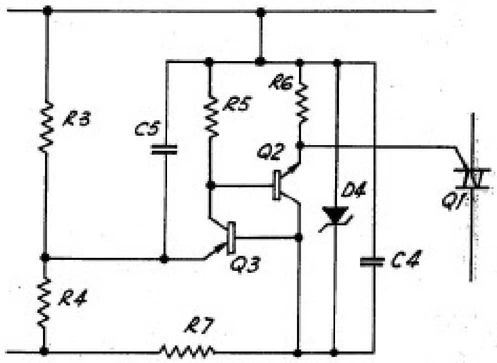
R6
R3
C5
R5
Q2
D4
Q1
Q3
C4
R4
R7

Fig. 2.

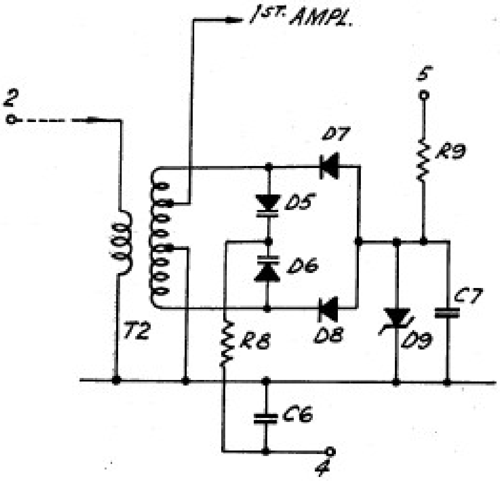
1ST. AMPL.
5
2
D7
R9
D5
D6
C7
T2
R8
D8
D9
C6
4

Fig. 3.

US 4,628,299
Seismic Warning System Using RF Energy Monitor

Abstract -- The ambient broadband radio frequency field strength from broadcast stations is monitored (Figure 4) by periodic sampling (50, 52). A warning indication is provided if the field strength drops significantly. Drops in such field strength have been correlated empirically with the occurrence of seismic activity, usually several days later. Thus the indication serves as an early warning of an impending earthquake. In one preferred embodiment, a broadband, horizontal, very long monopole antenna (40) was connected to a rectifying and smoothing circuit (Figure 3) to provide a dc output proportional to the ambient rf field...

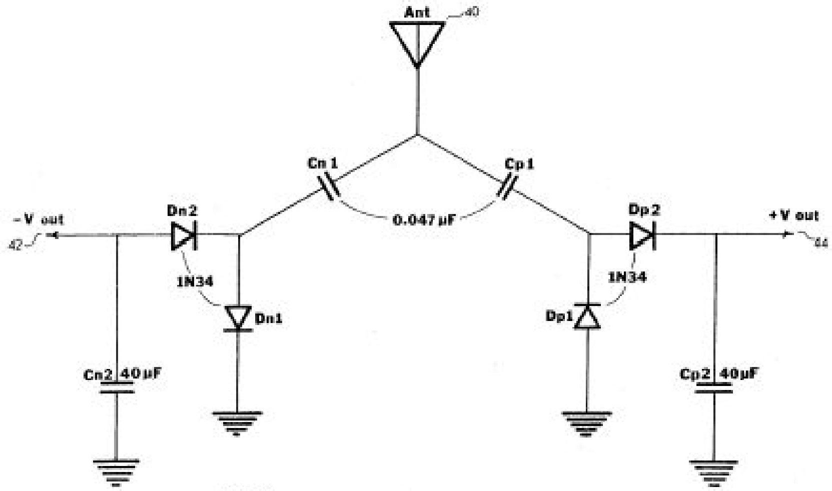

FIG 3

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

---

rexresearch.com

Louis MAICHE

Wireless Telluric Telephone

AS LOW AS $84/MO. WHEN YOU USE YOUR SWEETWATER CARD. CLICK FOR DETAILS
WARM AUDIO
WARM AUDIO WA76 & WA21
Sweetwater
SHOP NOW

Sure, it looks silly now -- but wait until an EMP destroys the modern grid -- then you'll want/need this proven simple, inexpensive alt-tech earth-radio -- directional, low energy, & untappable -- and anyone can build it...

Jerry Vassilatos : Vril Compendium

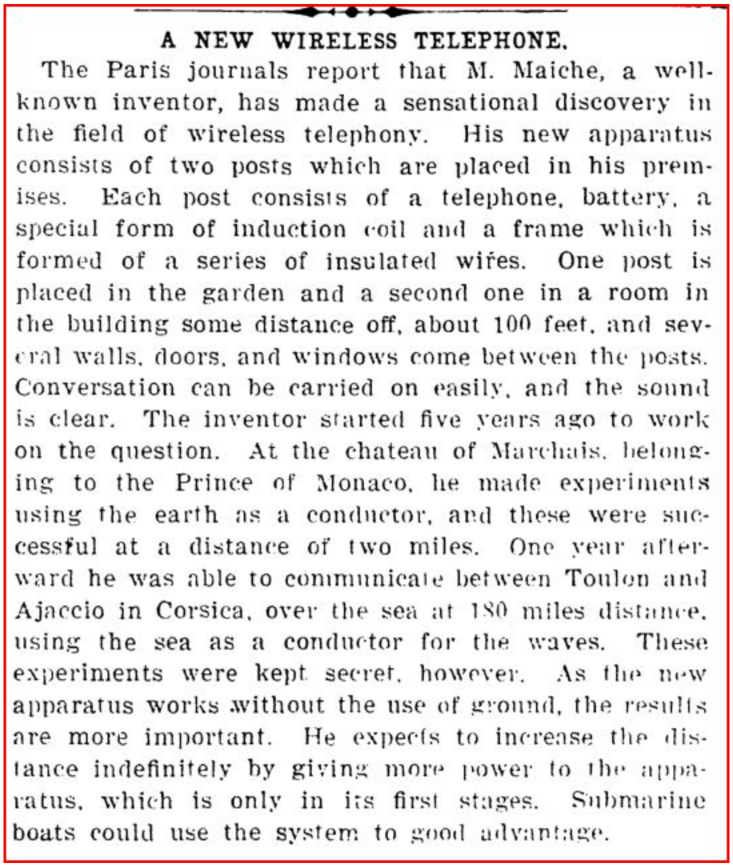

A NEW WIRELESS TELEPHONE.

The Paris journals report that M. Maiche, a well-known inventor, has made a sensational discovery in the field of wireless telephony. His new apparatus consists of two posts which are placed in his premises. Each post consists of a telephone, battery, a special form of induction coil and a frame which is formed of a series of insulated wires. One post is placed in the garden and a second one in a room in the building some distance off, about 100 feet, and several walls, doors, and windows come between the posts. Conversation can be carried on easily, and the sound is clear. The inventor started five years ago to work on the question. At the chateau of Marchais, belonging to the Prince of Monaco, he made experiments using the earth as a conductor, and these were successful at a distance of two miles. One year afterward he was able to communicate between Toulon and Ajaccio in Corsica, over the sea at 180 miles distance, using the sea as a conductor for the waves. These experiments were kept secret, however. As the new apparatus works without the use of ground, the results are more important. He expects to increase the distance indefinitely by giving more power to the apparatus, which is only in its first stages. Submarine boats could use the system to good advantage.

---

https://books.google.com/books?id=de5OAAAAYAAJ&pg=PA477&lpg=PA477&dq=Louis+MAICHE&source=bl&ots=0MwfiZciry&sig=ePs25QJg3djFbnx_ln0ivOIyIWg&hl=en&sa=X&v

The Christian Work and the Evangelist, Volume 82

Wireless Telephone of Louis Maiche, the French Scientist

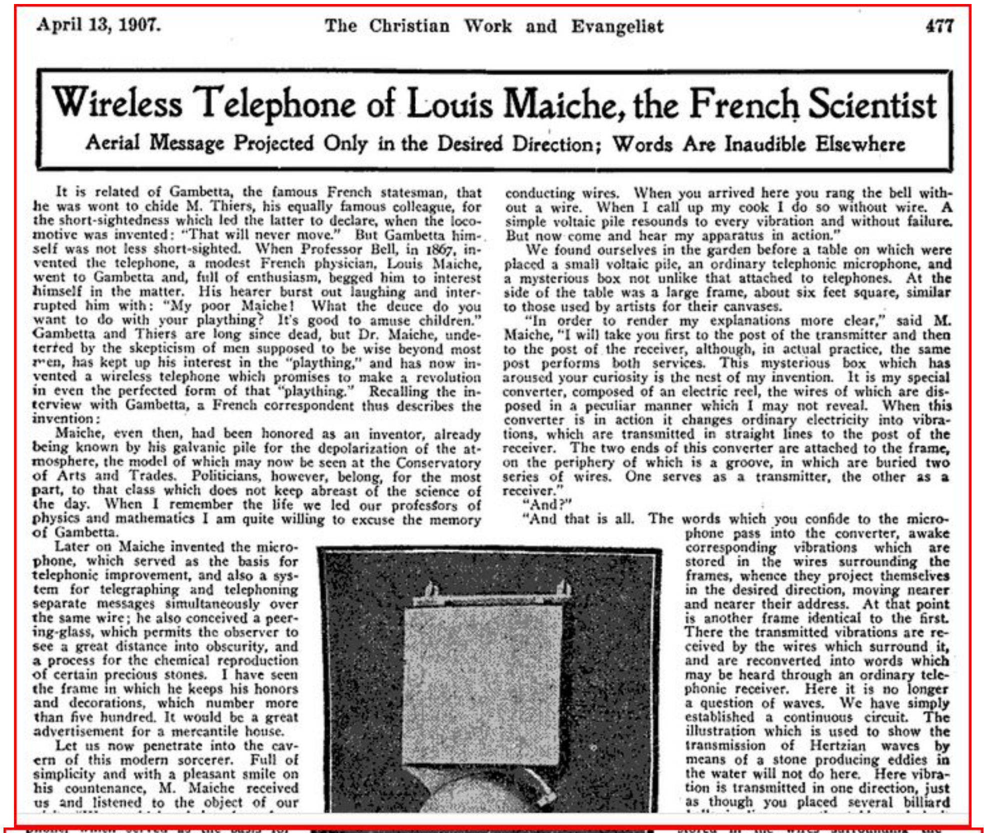

Wireless Telephone of Louis Maiche, the French Scientist

Aerial Message Projected Only in the Desired Direction; Words Are Inaudible Elsewhere

It is related of Gambetta, the famous French statesman, that he was wont to chide M. Thiers, his equally famous colleague, for the short-sightedness which led the latter to declare, when the locomotive was invented: "That will never move." But Gambetta himself was not less short-sighted. When Professor Bell, in 1867, invented the telephone, a modest French physician, Louis Maiche, went to Gambetta and, full of enthusiasm, begged him to interest himself in the matter. His hearer burst out laughing and interrupted him with: "My poor Maiche! What the deuce do you want to do with your plaything? It's good to amuse children." Gambetta and Thiers are long since dead, but Dr. Maiche, undeterred by the skepticism of men supposed to be wise beyond most men, has kept up his interest in the "plaything," and has now invented a wireless telephone which promises to make a revolution in even the perfected form of that "plaything." Recalling the interview with Gambetta, a French correspondent thus describes the invention:

Maiche, even then, had been honored as an inventor, already being known by his galvanic pile for the depolarization of the atmosphere, the model of which may now be seen at the Conservatory of Arts and Trades. Politicians, however, belong, for the most part, to that class which does not keep abreast of the science of the day. When I remember the life we led our professors of physics and mathematics I am quite willing to excuse the memory of Gambetta.

Later on Maiche invented the microphone, which served as the basis for telephonic improvement, and also a system for telegraphing and telephoning separate messages simultaneously over the same wire; he also conceived a peering-glass, which permits the observer to see a great distance into obscurity, and a process for the chemical reproduction of certain precious stones. I have seen the frame in which he keeps his honors and decorations, which number more than five hundred. It would be a great advertisement for a mercantile house.

Let us now penetrate into the cavern of this modern sorcerer. Full of simplicity and with a pleasant smile on his countenance, M. Maiche received us and listened to the object of our

conducting wires. When you arrived here you rang the bell without a wire. When I call up my cook I do so without wire. A simple voltaic pile resounds to every vibration and without failure. But now come and hear my apparatus in action."

We found ourselves in the garden before a table on which were placed a small voltaic pile, an ordinary telephonic microphone, and a mysterious box not unlike that attached to telephones. At the side of the table was a large frame, about six feet square, similar to those used by artists for their canvases.

"In order to render my explanations more clear," said M. Maiche, "I will take you first to the post of the transmitter and then to the post of the receiver, although, in actual practice, the same post performs both services. This mysterious box which has aroused your curiosity is the nest of my invention. It is my special converter, composed of an electric reel, the wires of which are disposed in a peculiar manner which I may not reveal. When this converter is in action it changes ordinary electricity into vibrations, which are transmitted in straight lines to the post of the receiver. The two ends of this converter are attached to the frame, on the periphery of which is a groove, in which are buried two series of wires. One serves as a transmitter, the other as a receiver."

"And?"

"And that is all. The words which you confide to the microphone pass into the converter, awake corresponding vibrations which are stored in the wires surrounding the frames, whence they project themselves in the desired direction, moving nearer and nearer their address. At that point is another frame identical to the first. There the transmitted vibrations are received by the wires which surround it, and are reconverted into words which may be heard through an ordinary telephonic receiver. Here it is no longer a question of waves. We have simply established a continuous circuit. The illustration which is used to show the transmission of Hertzian waves by means of a stone producing eddies in the water will not do here. Here vibration is transmitted in one direction, just as though you placed several billiard

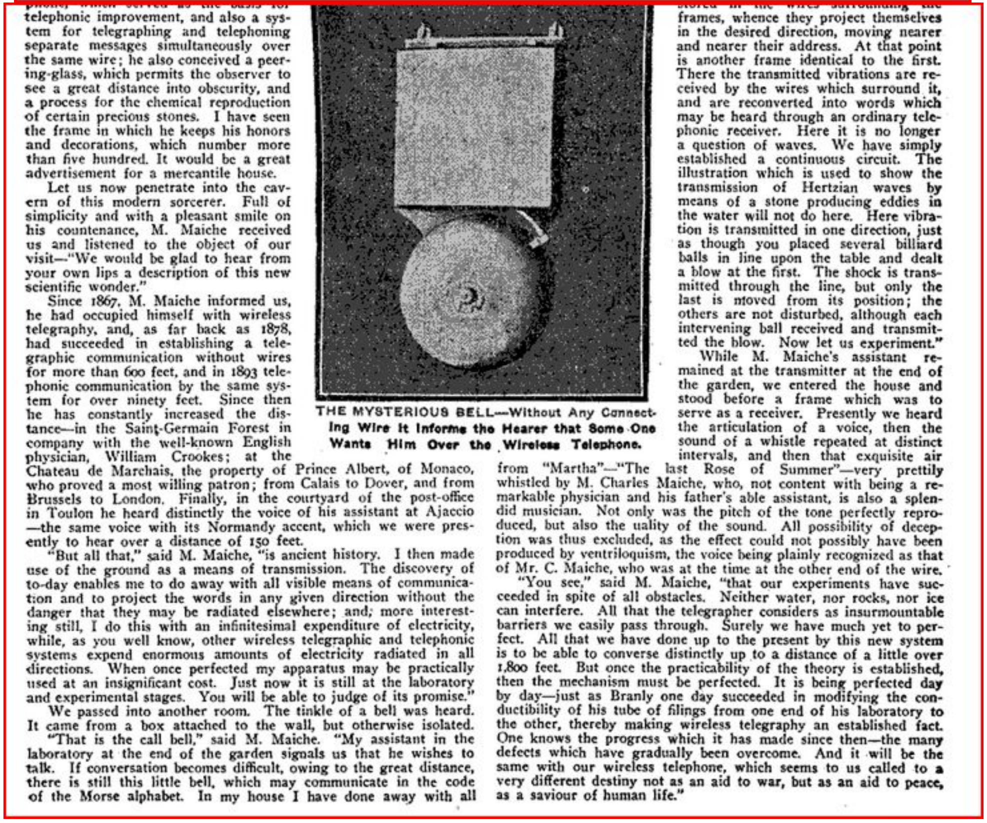

telephonic improvement, and also a system for telegraphing and telephoning separate messages simultaneously over the same wire; he also conceived a peering-glass, which permits the observer to see a great distance into obscurity, and a process for the chemical reproduction of certain precious stones. I have seen the frame in which he keeps his honors and decorations, which number more than five hundred. It would be a great advertisement for a mercantile house.

Let us now penetrate into the cavern of this modern sorcerer. Full of simplicity and with a pleasant smile on his countenance, M. Maiche received us and listened to the object of our

visit—"We would be glad to hear from your own lips a description of this new scientific wonder."

Since 1867, M. Maiche informed us, he had occupied himself with wireless telegraphy, and, as far back as 1878, had succeeded in establishing a telegraphic communication without wires for more than 600 feet, and in 1893 telephonic communication by the same system for over ninety feet. Since then he has constantly increased the distance—in the Saint-Germain Forest in company with the well-known English physician, William Crookes; at the Chateau de Marchais, the property of Prince Albert, of Monaco, who proved a most willing patron; from Calais to Dover, and from Brussels to London. Finally, in the courtyard of the post-office in Toulon he heard distinctly the voice of his assistant at Ajaccio—the same voice with its Normandy accent, which we were presently to hear over a distance of 150 feet.

"But all that," said M. Maiche, "is ancient history. I then made use of the ground as a means of transmission. The discovery of to-day enables me to do away with all visible means of communication and to project the words in any given direction without the danger that they may be radiated elsewhere; and, more interesting still, I do this with an infinitesimal expenditure of electricity, while, as you well know, other wireless telegraphic and telephonic systems expend enormous amounts of electricity radiated in all directions. When once perfected my apparatus may be practically used at an insignificant cost. Just now it is still at the laboratory and experimental stages. You will be able to judge of its promise."

We passed into another room. The tinkle of a bell was heard. It came from a box attached to the wall, but otherwise isolated.

"That is the call bell," said M. Maiche. "My assistant in the laboratory at the end of the garden signals us that he wishes to talk. If conversation becomes difficult, owing to the great distance, there is still this little bell, which may communicate in the code of the Morse alphabet. In my house I have done away with all

THE MYSTERIOUS BELL—Without Any Connecting Wire It Informs the Hearer that Some One Wants Him Over the Wireless Telephone.

frames, whence they project themselves in the desired direction, moving nearer and nearer their address. At that point is another frame identical to the first. There the transmitted vibrations are received by the wires which surround it, and are reconverted into words which may be heard through an ordinary telephonic receiver. Here it is no longer a question of waves. We have simply established a continuous circuit. The illustration which is used to show the transmission of Hertzian waves by means of a stone producing eddies in the water will not do here. Here vibration is transmitted in one direction, just as though you placed several billiard

balls in line upon the table and dealt a blow at the first. The shock is transmitted through the line, but only the last is moved from its position; the others are not disturbed, although each intervening ball received and transmitted the blow. Now let us experiment."

While M. Maiche's assistant remained at the transmitter at the end of the garden, we entered the house and stood before a frame which was to serve as a receiver. Presently we heard the articulation of a voice, then the sound of a whistle repeated at distinct intervals, and then that exquisite air from "Martha"—"The last Rose of Summer"—very prettily whistled by M. Charles Maiche, who, not content with being a remarkable physician and his father's able assistant, is also a splendid musician. Not only was the pitch of the tone perfectly reproduced, but also the uality of the sound. All possibility of deception was thus excluded, as the effect could not possibly have been produced by ventriloquism, the voice being plainly recognized as that of Mr. C. Maiche, who was at the time at the other end of the wire.

"You see," said M. Maiche, "that our experiments have succeeded in spite of all obstacles. Neither water, nor rocks, nor ice can interfere. All that the telegrapher considers as insurmountable barriers we easily pass through. Surely we have much yet to perfect. All that we have done up to the present by this new system is to be able to converse distinctly up to a distance of a little over 1,800 feet. But once the practicability of the theory is established, then the mechanism must be perfected. It is being perfected day by day—just as Branly one day succeeded in modifying the conductibility of his tube of filings from one end of his laboratory to the other, thereby making wireless telegraphy an established fact. One knows the progress which it has made since then—the many defects which have gradually been overcome. And it will be the same with our wireless telephone, which seems to us called to a very different destiny not as an aid to war, but as an aid to peace, as a saviour of human life."

---

https://books.google.com/books?id=euDmAAAAMAAJ&pg=PA79&lpg=PA79&dq=Louis+MAICHE&source=bl&ots=1k4NAB8mnd&sig=NTExYinKMaUfIbhxtExc0kaKA-0&hl=en&sa=X&ved=0ahUKEwjgp8uehoHQAhVj0FQKHeFyDlk4ChDoAQgbMAA#v=onepage&q&f=false

Telephony, Volumes 3-4

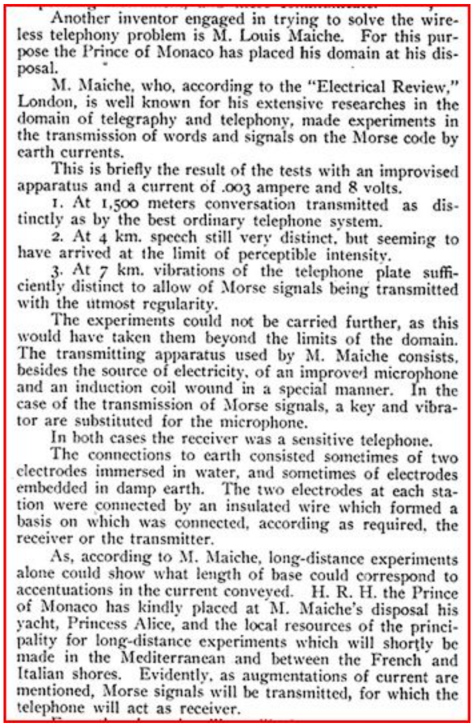

Another inventor engaged in trying to solve the wireless telephony problem is M. Louis Maiche. For this purpose the Prince of Monaco has placed his domain at his disposal.

M. Maiche, who, according to the "Electrical Review," London, is well known for his extensive researches in the domain of telegraphy and telephony, made experiments in the transmission of words and signals on the Morse code by earth currents.

This is briefly the result of the tests with an improvised apparatus and a current of .003 ampere and 8 volts.

1. At 1,500 meters conversation transmitted as distinctly as by the best ordinary telephone system.

2. At 4 km. speech still very distinct, but seeming to have arrived at the limit of perceptible intensity.

3. At 7 km. vibrations of the telephone plate sufficiently distinct to allow of Morse signals being transmitted with the utmost regularity.

The experiments could not be carried further, as this would have taken them beyond the limits of the domain. The transmitting apparatus used by M. Maiche consists, besides the source of electricity, of an improved microphone and an induction coil wound in a special manner. In the case of the transmission of Morse signals, a key and vibrator are substituted for the microphone.

In both cases the receiver was a sensitive telephone.

The connections to earth consisted sometimes of two electrodes immersed in water, and sometimes of electrodes embedded in damp earth. The two electrodes at each station were connected by an insulated wire which formed a basis on which was connected, according as required, the receiver or the transmitter.

As, according to M. Maiche, long-distance experiments alone could show what length of base could correspond to accentuations in the current conveyed. H. R. H. the Prince of Monaco has kindly placed at M. Maiche's disposal his yacht, Princess Alice, and the local resources of the principality for long-distance experiments which will shortly be made in the Mediterranean and between the French and Italian shores. Evidently, as augmentations of current are mentioned, Morse signals will be transmitted, for which the telephone will act as receiver.

---

http://www.telephones-anciens.org/l-maiche-electrophone-a-pupitre/

L.Maiche électrophone à pupitre

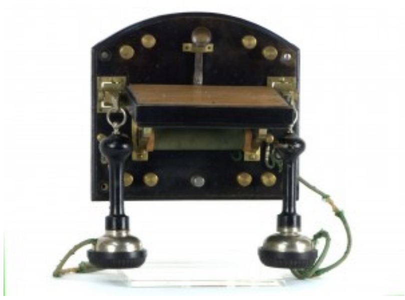

Appareil dit Electrophone de Maiche: nom donné par Louis Maiche a ses appareils téléphoniques. Celui-ci est équipé d'une paire d'écouteurs a manches Lensczewski et d'un microphone type Ader. Le modèle en photo doit être antérieur à 1890

---

http://www.lookandlearn.com/history-images/M240118/Maiches-wireless-telephone-system-1906?img=15874

Maiche's wireless telephone system, 1906

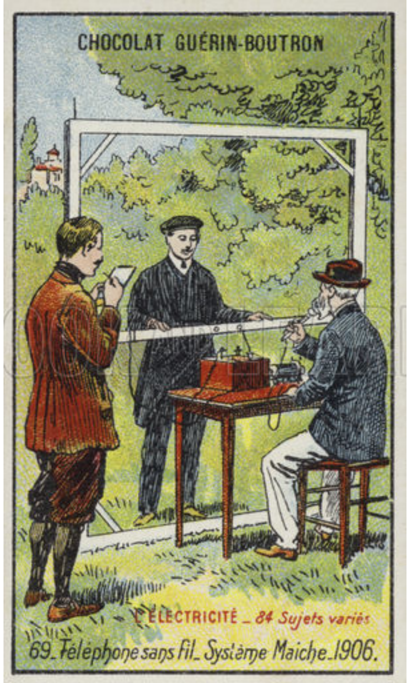
CHOCOLAT GUÉRIN-BOUTRON
L'ÉLECTRICITÉ _ 84 Sujets variés
69. Téléphone sans fil_ Système Maiche_1906.

---

http://thethingsienjoy.blogspot.com/2013_04_21_archive.html

The world's first "mobile" phone (in 1906)

French inventor Louis Maiche and his wireless phone, arguably also the world's first mobile phone.
Image published in the Swedish weekly Hvar 8 Dag in September 1906 (Click to enlarge)

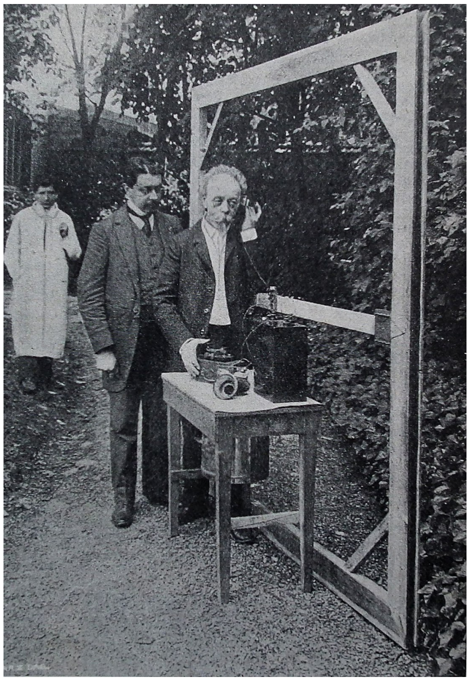

"M. Louis Maiche is the inventor of the wireless telephone and, like many scientists, is not so famous as he ought to be".

That is how an article in the Auckland Star on March 6, 1907, introduced the French inventor Louis Maiche, who also could be called the inventor of a mobile phone.

In the article M. Maiche describes how it all began:

"My first trials took place in 1867, but they were concerned more with wireless telegraphy, and it was not until 1893 that I exchanged a conversation through a wireless telephone at a distance of 30 yards."

Two months earlier, the New York Times had also published an article on M. Maiche and his wireless telephone invention. In the article the inventor refers to his successful experiments with telegraphy and then describes his work with the wireless telephone:

"I then made use of the ground as a means of transmission. The discovery of today enables me to do away with all visible means of communication and to project the words in any given direction without the danger that they may be radiated elsewhere; and, more interesting still, I do this with an infinitesimal expenditure of electricity, while, as you well know, other wireless telegraphic and telephone systems expend enormous amounts of electricity radiated in all directions."

"When once perfected my apparatus may be practically used at an insignificant cost. Just now it is still at the laboratory and experimental stages."

"You see", said M. Maiche,"that our experiments have succeeded in spite of all obstacles. Neither the water, nor rocks, nor ice can interfere. All that the telegrapher considers as insurmountable barriers we easily pass through. Surely we have much to perfect! All that we have done up to the present is to be able to converse distinctly up to a distance of a little over 1,800 feet. But once the practicality of the theory is established, then the mechanism must be perfected."

"It is being perfected day by day - just as Branly one day succeeded in modifying the conductibility of his tube fillings from one end of his laboratory to the other, thereby making wireless telegraphy an established fact. And it will the same with our wireless telephone,which seems to us called to a very different destiny, not as aid to war, but as an aid to peace, as a savior of human life."

In the Auckland Star article M. Maiche gives some examples of the future use of his wireless telephone:

"Imagine two steamers steering through a fog. With my little apparatus the captain can tell the direction the other ship is taking. In case of an accident to a submarine my apparatus would enable the crew in danger to communicate with the convoying tug without fearing the breaking of a line as in the case of a telephone buoy. Miners entombed after a disaster like that at Courrieres could communicate with the rescue party. Two army corps making a night attack could keep in constant touch without risk of interruption. The ordinary citizen could have it in his drawing room, and would no longer be dependent upon the vagaries of the telephone exchange."

Both articles also include some technical descriptions of the use of M. Maiche's new wireless phone. It is possible that the French inventor's theories were not as successful as he himself thought. Otherwise it is difficult to understand why this page (in French) is the only one with information about him that I have been able to locate. Still, his concept of wireless (and mobile) telephony should merit him a place in the hall of fame of modern communication technology .

Patents

Nouveau mode de transmission et de réception sans fil des signaux électriques

FR335990

1904-02-22

The object of the present invention is a new method of transmitting and receiving wireless electrical signals. The essential means on which is based the system is the production transmits position and uses the position receiver electricity manifested in a new form which may be referred to as "electrostatic vibration" and is obtained as will be explained further by modification of a spark inductor. One of distinctive characters of this new manifestation of electricity is that it can be transmitted in space without the aid antennas and ground at all distances, and, in the same form as sound, light and heat, that is to say without poles.

The inductor being communicated c with a power source, a battery, for example, and a special switch which will be described below, each interruption is to be born in the armature current giving a spark, and it is the latter that is transform an "electrostatic vibration." The armature out of the coil wire after the last winding outside, is for this purpose connected to a large metal surface which can be exposed to air. This surface may be constituted either by c a hollow or solid cylinder or by one or more sheets of iron, copper or any other metal usual more or less thick. The surface is preferably rounded at the corners or packed with a border all around to remove spikes, as in the conductive electrostatic machines. It should be about one square meter per watt of the current flowing in the inductor. The other end of the armature is connected to an adjusting screw of which the tip k of platinum or gold is next to a button, also in a gold or platinum, carried on the large surface.

FIGS. 1 and 2 of the accompanying drawings show two provisions which may be 5 h adopted for the preparation of the transmitter station.

FIGS. 3 and k show two provisions and receiver.

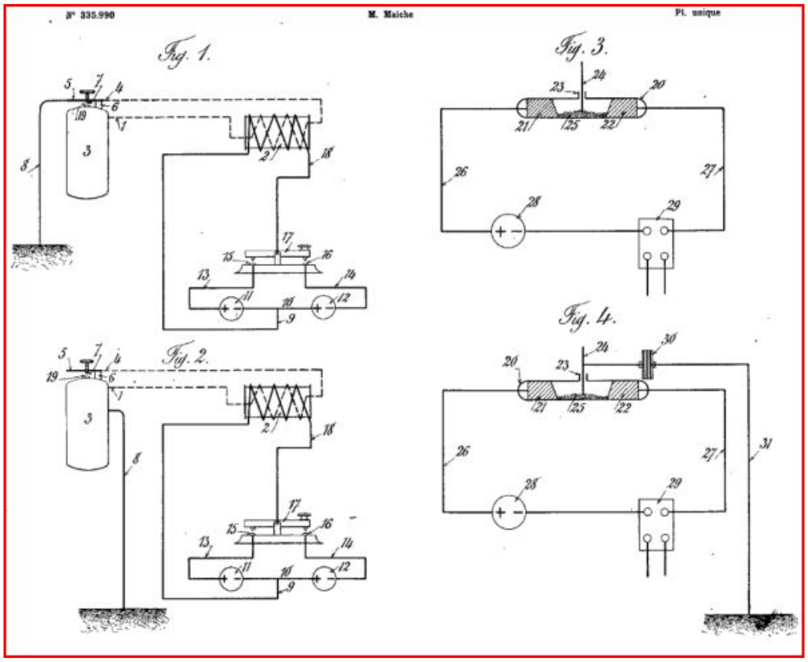

One end of the armature wire of the coil -j is connected to the cylinder 3 rounded at its ends, and the end k of the wire is connected to plate 5 mounted on the insulating support 6 which is fixed to the cylinder 3 . The plate 5 is traversed by a setscrew 55 provided beyond tip 7 made of gold or platinum. The end of the induced wire k is that by which begins rolling on a bobbin. Wire 8 communicates with the earth or the plate 5 (Fig. I) is the cylinder 3 (Fig. A).

Improvements in and relating to Electric Transformers.
GB190727157

1908-10-29

Comprises a large number of windings wound in the form of a cable a on a coreless frame b, and led out to terminals 1 ... 10 and, adapted to be connected in quantity or tension by means of a commutator so that three or more independent circuits may be made.

COMPLETE SPECIFICATION.

The present invention relates to electric transformers of the class described by Specifications Nos. N;821901 (Amberg), z1854 (Swinburne), and 4/I83 (Clerc) and it particularly concerns the arrangement of the windings.

The invention is represented in the accompanying drawings;-

Figs. 1 and 2, show a section and a front view respectively of the improved transformer.

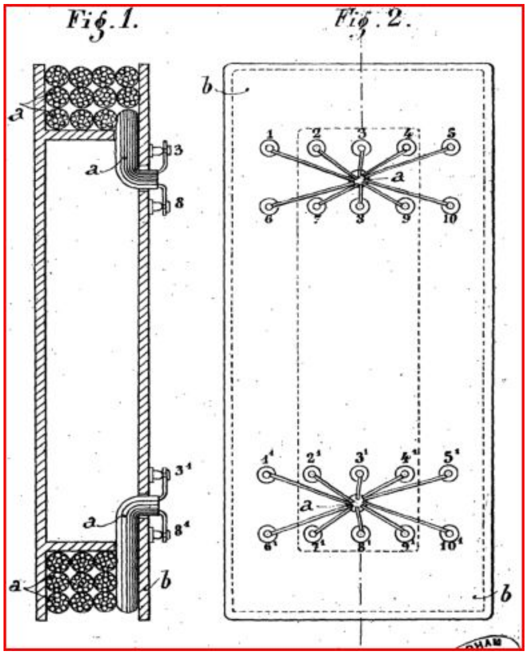

The transformer which has no magnetic core, is merely constituted by a convenient number of insulated wires, 10 for example, which are united to form a cable a which is wound around a frame b. The ends of the cable project through two holes of the frame and the ends of each wire are joined to terminals 1 to 10 and 11 to 101 respectively. In the drawings are shown 20 terminals, the first ten of which correspond to the beginnings of the 10 wires, and the other ten correspond to the ends of the same wires.

Thus, a winding of a certain number of identical wires is obtained which are all wound parallel and simultaneously and which may be connected by means of a commutator so as to form groups to 'be varied according to requirement; so that there are thus constituted three or more circuits which are entirely independent the one of the other and the length of which can be varied at will.

By conveniently joining the free ends of some of the wires, a primary winding of weak resistance is obtained with the wires of same joined for quantity, while one or more of the secondary windings are joined for tension, thus providing any required resistance.

The construction of the transformer by means of a great number of separate wires offers numerous advantages amongst which are the facility of connecting the separate wires at will to constitute the circuits; the arrangement enabling a supplementary circuit to be branched off at any point of the said circuits; and the reciprocal independence of these circuits with large number.

Improvements in and relating to Electric Transformers.
GB190721535

1908-05-28

Transformers for telegraphy and wireless telegraphy are made without a magnetic core, and of little depth but of large diameter. They have the primary winding e enclosed within the secondary winding c, g, but separated therefrom by insulation d, f. The windings may be formed in sections, and coupled up in any order by means of a commutator.

COMPLETE SPECIFICATION.

This invention relates to electric transformers the construction of which is improved so that currents of induction of very short periodicity are obtained.

The currents, the duration of which is extremely short, possess special properties which are particularly useful when applied in telegraphy and in wireless telephony.

Transformers without magnetic core and with the secondary winding wound around the primary, or with the primary winding enclosed within insulating material round wh,ich is wound the secondary winding so that the whole forms a cable, have been provided, and my invention is limited to the form and construction of the transformer as hereinafter set forth.

My improved electric transformer contains its primary wire within a cable of the secondary wires in the form of a single coil of little thickness but of very large dia,meter, said coil or cable being applied to a ring-shaped transformer body.

The accompanying drawing shows my improved electric transformer.

Fig. 1 is a section on line A-B of Fig. 2 and

Fig. 2 is a plan of the ring-shaped transformer.

Fig.1.

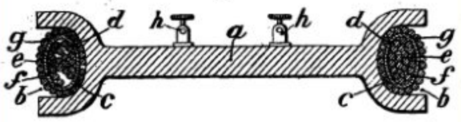

Fig.2.

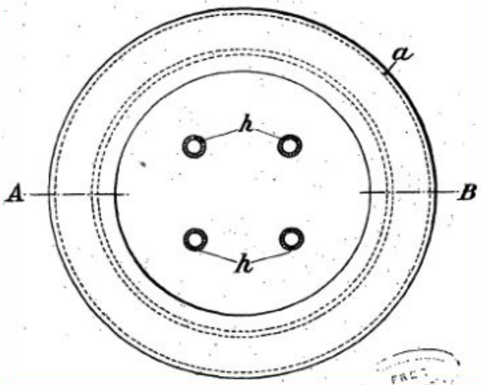

As will be seen from the drawings, the transformer a is of comparatively large diameter which may be several meters. In the drawing the transformer is represented of circular cross-section, it could however be of square or other cross-section. This transformer is very short because all the wires of the winding are located on its periphery in an appropriate hollow or groove b. For manufacturing the trans-former I place around the rim of this kind of wheel or dial: a, a first layer of secondary wire which is covered by a suitable insulator d and upon this insulator the primary wire e is wound which is covered by another layer of insulating material f. The two insulating layers d and f are connected in such a manner that the primary wires e are well insulated, and the coil is then finished by applying an outer winding of secondary wire g, the latter surrounding the primary in such a manner as to form a cable. In this manner a ring of large diameter is obtained which is composed of the secondary wires c, g which enclose a core of suitably insulated and wholly surrounded primary wires e and which ring is applied to a ring-shaped transformer-body having a peripheral groove of suitable cross-section.

The apparatus comprises four terminals h, two of which serve for fixing the ends of the secondary wires, the other two serving for fixing the ends of the primary winding.
There could be arranged several independent cables or rings in each of which the secondary wires entirely surround the primary wires and in such case each of the terminals would be replaced by a multiple commutator permitting to join up the different secondary wires as well as the different primary wires either for tension or for quantity according to the effect to be obtained.

Improvements in connection with Wireless Telegraphy and Telephony.
GB190721252

1907-11-21

The transmitting and receiving circuits at one station for an inductive system of telephony or telegraphy are arranged as shown. A coil a, consisting of a few turns of thick wire, is used alone in transmitting, but is connected in series with a coil b, consisting of a large number of turns of fine wire, for receiving; the circuit of the coil b is broken automatically, during transmission, by means of a Morse key c. The circuit of the coil a includes a battery n and a microphone o for telephony, or an interrupter for telegraphy, and the coil b is connected with a telephone h.

There are known systems for telegraphy and telephony without connecting wires which utilise, instead of Hertzian waves, the phenomena of distant induction by means of electric vibrations of very short duration, but great amplitude. With such systems, one and the same winding is used fur sending and for receiving; for sending it is placed in circuit with a source of electricity (battery, accumulator, etc.) and with a microphone or an interrupter destined to produce the vibrations in question. The same winding is placed in circuit with a .receiving telephone .for receiving the currents which arc induced at a. distance from the sending winding of the other station.

It is obvious that this single coil cannot equally well serve for the sending as for the receiving. To transmit well it ought to be made from thick and short wire, whilst for receiving well it ought to be of long and thin wire.

It would evidently be the simplest to provide each station with two different coils of which one serves as transmitter and the other as receiver, but lucre is another difficulty to be considered; for receiving, the long and thin coil has to be closed by the telephone and in that condition it would absorb nearly all the energy which ought to be sent out by the transmitter.

The present invention rotates to a special method of mounting the parts constituting a station in order to overcome the above mentioned difficulty; besides, the special method of mounting permits to join the two coils one behind the other for receiving whereby the receiving power of the station is considerably increased.

The accompanying drawing represents diagrammatically a station according to the present Invention.

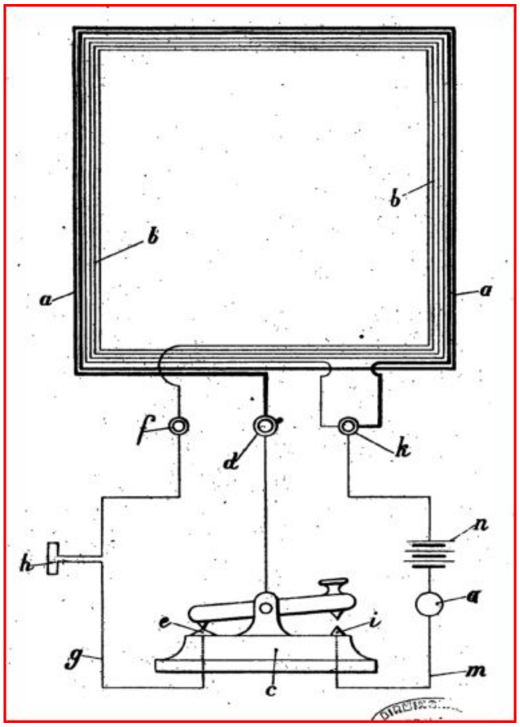

a is the inductor wire which serves fur transmitting; this wire is thick and relatively short. b is the induced winding for receiving the induced vibrations from the inductor winding of the transmitting station; this winding consists of thin wire of considerable length. c is a. Morse-manipulator the lever of which is connected through the contact let with one end of wire a, its connector e is through the contact f connected with one of the ends of wire; the wire g which connects the contact f with wire b comprises the telephonic receiver h ;the other connector i of the Morse apparatus is connected through terminal with the free ends of the two wires a and b which are joined at this point. The connecting wire m which serves for this purpose, comprises the battery n and the microphone o.

It is easily' understood that in the inoperative position of the Morse key (as represented in the drawings) the two coils and b form one circuit closed on the telephone h. This position is the receiving position. For transmitting, the key is depressed and brought in contact with the connector whereby the coil b is opened whilst the coil a is closed on the microphone and the battery.

It is to be understood, that the station described serves for the transmission of sounds. For the transmission of telegraphic signals, the microphone o is replaced by a suitable interruptor-vibrator and the apparatus is operated through the Morse key as usual.

Improvements in and relating to Inductive Telegraphy and Telephony.
GB190712529

1908-03-12

In a system of telegraphy or telephony by magnetic induction, the coils are placed vertically and the same coil is used for transmitting and receiving. The coil b is wound on a frame a of any suitable shape, which may be lined with iron. One terminal of the coil b is connected to a key f, one contact of which is connected through a microphone p and battery o to the other terminal d of the coil. A telephone m is connected across the terminals of the coil through another contact of the key f, and its circuit is normally completed for receiving telephonic or telegraphic messages. Telegraphy is effected by manipulating the kev f to complete the circuit of battery o ; to telephone, the key is depressed and the microphone p is utilized. A high-frequency alternator may be used instead of the battery o, and a screen composed of sheets of paper soaked in paraffin, varnish, or glue may be placed behind the frame.

The present invention relates to improvements in that class of magnetic induction telegraphy and telephony in which transmission is effected between similar coils of wire at the receiving and transmitting stations. The invention is characterized by the fact that senders and receivers.are reversible, the member for transmitting the waves being the, same as the member which receives the waves.

In the drawing, a diagram of a station according to the present system is represented.

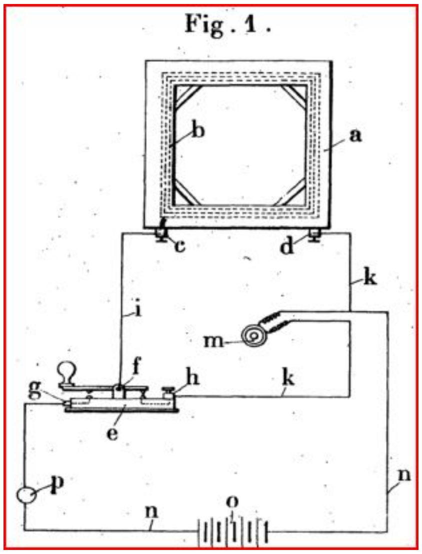

The transmitter and receiver for the electric waves consists of a. wooden frame (a) of fairly large dimensions; for example 4 meters square for a distance of 20 kilometers. In the drawing the frame is shown as being square but it may be of any other geometric shape having as center a.rectangle a lozenge, a circle, etc. The wall of the frame is provided with a groove (b) in which an insulated conducting wire is placed which is connected with the terminals (c) and (d). This wire is wound a. varying number of times around the frame so that it is of sufficient length according to the power of transmission required of the apparatus.

The frame (a) must be of sufficiently large dimensions to prevent the currents generated in opposite directions in the sides of the frame from tending to neutralise each other.

Each station comprises besides a Morse manipulator (e), the commutator of which is connected with one of the terminals (c) of frame (a) through the wire, its terminals (r) and (h) being connected with the other terminal (d) of frame (a).

An ordinary telephonic receiver (m) is inserted in, the circuit (k) which leads from the terminal (h.) of the manipulator to the frame, and in the other circuit (n), connecting the terminal (g) of the manipulator with. the frame, a battery (o) is inserted near the frame and , microphone (p) is inserted near the manipulator.

When the manipulator (c) is at rest (position shown in the drawing) the station is ready to receive the transmissions, the electric circuit (a, c, i, f, m, k, d, a) enclosing the frame and the telephone -being closed. One has only to hold the telephone to the ear to perceive all the currents which a.re induced in the frame (a).

The telephone (m) is put out of circuit when the commutator (f) is pressed down, whilst t.he circuit will be closed upon the frame (a.) and battery (o). The station has thus become 11 sender, the battery sending variable currents into the frame according to the nature of the transmission to be effected. To telegraph according to the Morse alphabet one manipulates the commutator (f) in the usual manner, the telephone (m) of the other station permitting to distinctly perceive the dots and dashes. The operator simply gives one contact for the dot and two contacts following in quick succession for a dash. In this case the microphone (p) is superfluous but it does not disturb. When however one wants to telephone, the contact (g) has to be kept closed and one speaks in front of or into the microphone (p); the telephone (in.) of the other station will reproduce the sounds.

The source of electricity (o) can be a battery or an electro-magnetic generator giving alternating currents of high frequency. The currents which are thus sent into the frame (a) of the sender station have no appreciable effect upon the telegraphic receivers which could be provided at this station; they even do not influence the magnet needle owing to their **illst,nntfmeousness.**

To establish the communication between two stations the two frames (a) have to be placed in parallel planes, perpendicular to the line which would connect their centers, which can be effected through various means. The action (if the one of the frames on the other is effected through any obstacles whatsoever, no earth connection or air connection being necessary.

The reciprocal action of the frames can be increased through lining the bottom of groove (b) for its entire length with soft sheet iron, the lining being continuous or interrupted. The frame will thus become a kind of electro-magnet, the proportion of which are just the reverse of those generally used.

To communicate at very great distances several frames (n,) are joined up, the wires being joined up in series or for tension.

The efficiency can be considerably increased and the direction of transmission localised in avoiding dispersion at the sides or behind the frame, by means of a thick screen composed of superposed sheets of paper soaked in paraffn ; the paper could also be impregnated with varnish, glue or any other equivalent material.

Improvements in and relating to Telegraphic and Telephonic Systems.
GB190126600

1902-11-06

Relates to the arrangements of circuits in telegraph and telephone systems. In Fig. 1, a microphone b is shown in a main line a, and at the receiving-station is connected to the middle of a split battery f, g, two receivers l, m being connected between battery and earth. When induction coils are employed, two separate primaries take the place of the receivers in the Figure, and the receiver employed is placed in a common secondary. In another modification, the transmitter is bridged across two parallel closed circuits, including primaries and batteries, the secondary forming the line circuit.

This invention relates to telegraphic and telephonic systems and has for its object certain improvements hereinafter more particularly referred to and finally pointed out in the claims.

As is well known, in order to prevent over-heating of' the receiving and transmitting apparatus in telephonic or telegraphic installations it is necessary to prevent very strong current passing over them. It may however, in certain cases, be necessary to provide currents of varying energy within rather large limits, and this invention has for its object to provide a simple device which will permit currents of any force being passed through telegraphic and telephonic apparatus without the slightest risk of deterioration to such apparatus.

In order that my invention may be readily understood and carried into effect I will now proceed to describe the same fully with reference to the accompanying drawings wherein are shewn diagrammatically several methods of carrying my. invention into practice.

In the drawings :

Figure 1 shows the principle of my improved arrangement, and

Figures 2. and 3 are modified arrangements.

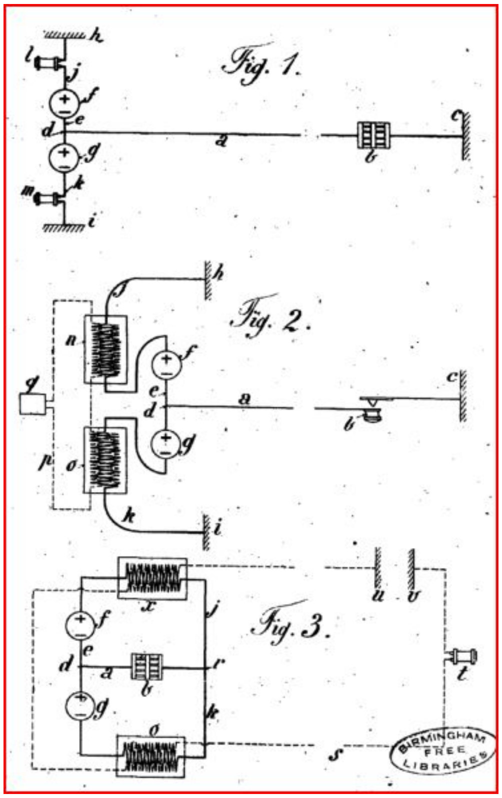

My improved system consists essentially of a line a of any convenient length in which is interpolated a microphone b or other appropriate transmitting apparatus. One end of this line is grounded at c and the other end is connected at d to.a wire e adapted to place into communication the opposite poles of the two sources of electricity of equal power, which for greater clearness sake I have represented as two single batteries f and g. These two batteries are respectively connected to earth at h and i. In the earth wires j and k are interpolated telegraphic or telephonic receivers l, m, of any convenient construction. In the example shown in the drawing the receivers and m are indicated as being telephonic receivers.

Having thus described the general arrangement of the system I will now explain the function of the various apparatus of the system.

Presuming that the wire a is non-existent and that the communication of the battery f and g with the earth has been established, a current passes from the positive pole of the battery f to earth through the wire then back .through the wire k, battery 9 and finally through the wire e to the negative pole of the battery f. Thus a closed circuit is established. Admitting now the wire a to be connected to the wire e at d and the transmitter b operated, the single current generated in the previous case will be replaced, by two currents which circulate in the same direction.

One of these currents starts from the positive pole of the battery f through the wire through earth from h to c through the transmitter b, wire a and wire e to the negative pole of the battery f.

The other current starts from the positive pole of the battery 9 through the 'wires e and a, the transmitter b, through earth from c to i and wire k to the negative pole of the battery g.

It will thus be seen that the wire a and transmitter b are simultaneously traversed by two equal currents in inverse directions.

In Figure 2 I have shown a telegraphic transmitter or key b but the general arrangement of the system is the the same as that previously described with the exception that the wires j and k form the primary conductors of the transformers n and o. The secondary conductor p, which is common to both transformers, leads,to the receiving apparatus q, for example a telegraphic receiving instrument.

Each contact ejected with the key b produces, as before explained a double derivation acting by subtraction of intensity upon the two transformers n and o, produces an induced current which is reproduced in totality in the receiving apparatus q. In the case of telephonic transmission the key must be replaced by a microphone as shown in Figure 1, while further the receiver q would also be a telephonic receiver.

By extending the same principle I can, as indicated in Figure 3, replace the two earth wires by two conductors and k joined together at r after having traversed the transformers x and o. The induced current may then be conveyed to the receiving instrument t over two lines, or as in the previous example over a line s and two earth connections u and v, placed more or less apart from each other.

It will be noticed that the microphone b forms part of the neutral conductors.

It acts exactly as in Figure 1. The differences of resistance produced by its vibrations influence the two coils x and o, and it is easy to understand that, whatsoever may be the intensity of the two electric sources f and g, the microphone is never exposed to the risk of becoming overheated for the reason that it is traversed by two equal currents in opposite directions.

It will thus be understood that the transmissions will be rendered perceptible in the receivers through the differences of intensity of the electric current; the differences being due to the differences of resistance of the neutral conductor which, in receiving two equal currents in opposite directions, is not much affected itself.

Improvements relating to the Transmission of Telegraphic and Telephonic Signals.
GB190117711

1901-02-19

The local circuits at each station are doubly earthed at e, f, j, k symmetrically about a straight line joining the stations. The earths may be formed of porous pots containing water in which the ends of the conductors are immersed, or deep ditches filled with gas coke may be employed. Insulating screens may be placed in deep ditches dug behind the earths.

This invention relates to an improved method of transmitting electrical telegraphic and telephonic signals from one point to another without the employment of wires connecting the two points.

In my improved system of transmission, the two poles of the transmitting station are put to earth and the two conductors, upon which the receiving apparatus is mounted are also both put to earth; the signals emitted by the transmitting apparatus are transmitted to the receiving station through the earth alone.

In order that my invention may be readily and clearly understood, I have represented in the accompanying drawing, by way of example only, a diagrammatic view of my device as a whole.

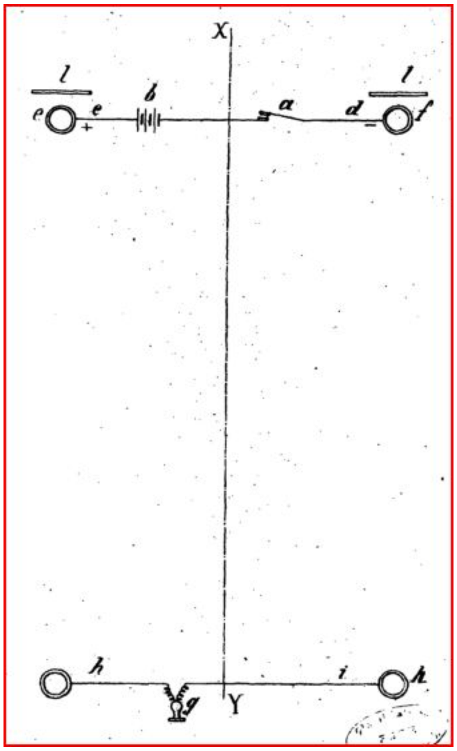

As shewn in the drawing, the transmitting station comprises a transmitter a one of the terminals of which is electrically connected to one of the poles of a source of electricity b whilst the other pole communicates with the earth; the second pole of the source of electricity is also run to earth and with this object the two conductors c d terminate in two porous pots e f embedded in the ground and filled with water.

The receiving station comprises a receiver 9 such as a telephone for example, the conductors h i which are put to earth by the intermediary of two porous pots j k similar to those of the transmitting station.

The porous pots of the receiving station should be arranged upon a line parallel with that which joins the two porous pots of the transmitting station, and they should be equidistant from an axis passing through the middle point of this line and at right angles to said axis.

The porous pots may advantageously be replaced by deep ditches sunk down to the humid earth, filled with gas coal or coke and containing at their centre a plate of carbon which is a good conductor, to which the transmission and receiving wires respectively are attached.

When contact is established by manipulation the transmitter a, the electric circuit passing through the source of electricity b, the conductor c, the porous pot e earth-connected to the other porous pot f the conductor d and the transmitter a, is closed. The current thus produced causes a change of condition in the ground ; all around the porous pots electric radiation is produced, the polarity of which is that of the corresponding conductor; its density decreases in proportion as it is further removed from the point of departure. This radiation is infinite with respect to space if nothing obstructs its passage.

The radiation extends so much the further according as the distance between the two porous pots of the transmitting station is greater.

These radiations, starting respectively from the porous pots e f thus create two zones one of which is positive and the other negative, the intersection of which constitutes % line absolutely neutral. It follows from this that the porous pots of the receiving station should be arranged upon either side of the neutral axis, because if these two earth connections were upon the same side they would both collect fluid of the same polarity, and in these conditions the receiver would not be traversed by any current.

In order to collect the maximum of current at the receiving station, the two porous receiving pots should be as nearly as possible equidistant from the neutral axis, as if one of the two porous pots is nearer than the other, the difference of separation of that which is the furthest removed is not only lost but is prejudical since it causes useless resistance without adding anything to the effect.

Experience has demonstrated that the disposition which furnishes the maximum of reception is that which consists in arranging the porous pots of the receiver at the same interval as separates the porous pots of the transmitting station.

In order to install my method of electrical transmission, it is first of all necessary to draw a straight line a y connecting the two stations and .then to produce upon each of these stations a line proceeding from this line x y at right angles, the said perpendicular lines being of equal length.

The transmitting and receiving appliances may be arranged at any convenient point on the conductor which connects the porous pots.

In order that the whole of the radiation emitted may be utilised at the receiving station, the propagation of this radiation is prevented (at the transmitting station upon the side opposite to that on which the receiving station is situated by burying in the ground an insulating screen l which may be of glass, tarred fabric, ebonite or the like, or by sinking a sufficiently deep ditch with or without an insulating screen.

Each time contact is made at the transmitter a, a sharp blow is audible at 'the receiving telephone g ; a second blow is heard when the circuit is broken.

It is obvious that there may be arranged both at the transmitting and receiving station transmitting and receiving appliances of the greatest possible sensibility, either for serving as relays for telegraphic purposes, or for directly receiving and recording the signals; for example an ink siphon mounted upon a galvanometric frame, or any other suitable device.

My method of transmission also permits of the transmission of speech; it is only necessary to replace the transmitter by an ordinary microphone interposed either directly upon the circuit, or arranged upon the primary of an induction coil, the secondary wire of which terminates at each extremity of the two porous pots.

It will of course be understood that in practice each station comprises both a transmitter and a receiver.

Improvements relating to the Transmission of Electric Currents
GB190108906

1901-04-30

Telegraph and telephone systems are arranged with the line-wire a connected to the center point of a battery c, c, the opposite poles of which are earthed and shunted by a normally closed circuit key g or microphone. The earth connections are made through non-polarizable cells f, f, containing, for example, two plates of copper immersed in copper-sulphate solution. The receiver is bridged between one earth and a switch h by which the transmitter is normally connected to line. The microphone may be inserted in the primary of an induction coil, the two secondaries of which oppose and are connected to the two earths and line. The two earth connections may be brought nearer together by the receiver and switch in the primary of the induction coil, so that this coil acts both to receive and transmit according to the position of the switch h. The microphone is arranged in the primary of a second induction coil.

This invention relates to improvements in the transmission of electric currents and is more particularly adapted for telegraph and telephone lines, whereby condensation in the cable is avoided, while at the same time "tapping" of the , line at any point thereof, for the purpose of ascertaining the purport of the message in transmission is rendered impossible.

My improved system of transmission comprises a single line wire between the two stations and branched at these two stations to the zero point of a circuit, the two extremities of which are connected to earth or the sea ; the current furnished at each station by a local source of electricity being sent through the circuit by means of any convenient transmitting apparatus, (such as a Morse key, microphone or the like) either direct or by induction.

Each station is furnished amongst other apparatus, with a receiving apparatus consisting of a key-instrument connected with the line-wire and permitting the current sent into the line from the transmitting station, to pass into a conductor leading direct to one of the wires running to earth and to which is connected the receiver proper.

In order to avoid the polarization of the earth-wires, each such wire is connected with a conductor leading to a copper or other metallic blade immersed in an impolarizable saline solution in which is disposed a second similar blade connected to earth or sea.

In the accompanying drawings :

Figure 1, is a diagram showing my improved system as applied to telegraphy

Figure 2, is a diagram showing the same applied to telephony.

Figure 3 shows the system of Figure 2 with induced current.'

Figure 4, shows a modified arrangement.

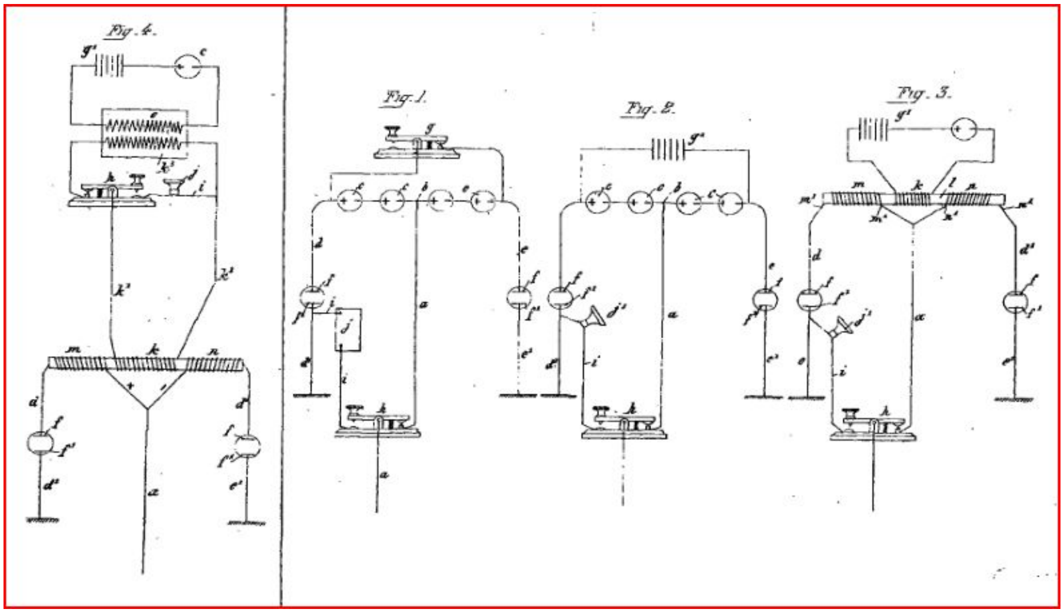

In all the figures the same letters of reference designate the same parts.

Referring to Figure 1 the two stations are connected by a single line-wire a, branched at the zero point b of the battery formed of the different elements c of which the end poles are connected to earth, or to sea, by a conductor d e lead- ing to a copper or other metallic plate f immersed in a liquid of good conductivity, such as for example a solution of a sulphate of copper, in which another copper plate f1 is placed connected to earth, or sea by a conductor d1 .

A transmitting apparatus g of any convenient construction, such as a Morse- key instrument, is connected with the two end poles of the battery in such a manner that in its position of rest it closes the circuit of the latter by establishing a short circuit so that no current passes, (or at most a very weak one) through the wire line a or through the earth-wires.

To send a current to the receiving station, the key 9 is depressed whereupon the short circuit is interrupted and the current is caused to pass simultaneously through the line-wire a and the earth-wires d d1 e e*. The line-wire is in reality traversed by two currents in opposite directions, so that it is in an apparently neutral state, and thus no interception of the messages can be effected for the, reason that if a receiver be interpolated, at any point of. the line, said receiver would receive two currents in opposite directions' which would neutralize each other.

Each station is furnished also with a receiving apparatus consisting of a commutator h connected with the line-wire a and permitting, when depressed, the current from the transmitter to pass into a wire i leading direct to one of the earth-wires (d1 for example) and in which is interpolated the receiving instrument j.

To receive signals sent from the transmitting station, it 'is necessary that the commutator lever h be kept depressed, so as to cause the current from the transmitting station to pass into the wire In connection with the receiving instrument.

'The connections may also be established in such a manner that when the commutator h is in its normal .position, it cuts the wire (and connects the wire with that part of the said wire which leads to the other station ; in this case for transmitting a signal it is necessary to depress the commutator lever h previously so as to re-establish the continuity of the line-wire and the operator then acts on the key-lever g as hereinbefore stated.

My arrangement is equally applicable to telephonic messages as will now be explained with reference to Figure 2, The transmitting instrument g is in this case replaced by a microphone g1 and the receiving instrument j is replaced by a telephone receiver j 1 the operation of the arrangement is exactly, the same as that above described.

The vibrations of the microphone cut the short circuit and the current from the local, source of electricity thereupon passes over the line-wire a and the earth-wires d d1 c el as previously stated.

The telephonic message is received in the same manner as before through the intervention of the commutator h which passes the current, sent from the transmitting station; to the telephonic receiving instrument jl.

The line-wire may also be operated by induction currents a3 is indicated at Figure 3.

The local circuit in which the transmitting apparatus (for instance, the microphone g1) is inserted comprises a primary winding k surrounding a central soft-iron core, or bundle of wires, l which is further provided with two secondary windings m m in opposite directions. The line wire a is connected to each of these secondary windings at the two extremities m1 n1 of opposite polarity and the opposite extremities ms p2 are connected with the earth-wires d d c set in the manner already explained.

When operating the transmitting apparatus or when speaking into the microphone 0 for example, induction currents are generated in the secondary windings m n and they then affect the receiving apparatus jl at the receiving station as already stated.

The primary winding and the two secondary windings may have the three wires wound in the same direction upon the soft iron core and the two extremities of the primary convolutions are connected with the local circuit in which is interpolated the transmitter while two extremities of opposite polarities of the secondary convolutions are connected to line and the two osiers to earth as previously explained.

In my improved transmitting arrangement the two earth wires at 'each station must be the more separated from each other the longer the distance is between the two stations so that the resistance of the earth between the said earth-wires will be sufficient to cause the single line-wire to be traversed by the greatest possible quantity of the current.

In order to reduce as far as possible the distance between the earth-wires the beginning of the single wire a and the wires rl rl' leading to earth are advantageously rm distance from the transmitting instrument and receiving instrument respectively at the telegraphic or telephonic station and thus permit the two earth connections of the two telegraphic or telephonic stations to be placed nearer each other.

The telegraphic or telephonic station is then divided and the transmitting instrument is connected with the magnet with its earth connections and line wire a ' by two conducting wires kl k2 connected with the ends of the convolutions k which generates induction currents in the secondary windings m n (Figure 4).

The current sent into the conducting wires k1 k2 and. passing through the convolutions k may itself be an induced current generated in the coil k3 when a current from the source of electricity c is passing through the. primary coil o through the action of the transmitting microphone gl for example.

In the case when' the distance of the telegraphic or .telephonic transmitting station from its earth-connections is not too excessive the current traversing the conductors k1 1.:2 instead of being an induction current may be furnished direct by the source of electricity of such station, and naturally the, transmitting instrument, (microphone or 'key), and the source of electricity are then placed direct in the circuit;1 k2.

An interrupter A interpositioned in the conductor kl permits of cutting the circuit in leading to the secondary coil k3 and of closing the circuit for the conductor i leading to the conductor k 2 in which is interpolated' the receiving instrument j.

Thus the coil k is either inductor or induction according to whether there is transmission or reception.

It is to be well understood that my system of transmission is entirely independent of the nature of the telegraphic or telephonic apparatus employed therewith, and that the source of electricity at the stations may be of any convenient kind, either derived from a battery, accumulators, dynamo, magneto-electric machine or the like.

COMPLETE SPECIFICATION.

This invention elates to improvements in the transmission of electric currents and is more particularly adapted for telegraph and telephone lines, whereby condensation in the cable is avoided, while at the same time "tapping" of the line at any, point thereof, for the purpose of ascertaining the purpose of the message in transmission is rendered impossible.

My improved system of transmission comprises a single line-wire between the two stations and branched at these two stations to the zero point of a circuit, the two extremities of which are connected to earth or the set; the current furnished at each station by a local source of electricity being sent through the circuit by means of any convenient transmitting apparatus, '(such as a Morse key .microphone or the like) either direct or by induction.

Each station is furnished, amongst other apparatus, with a receiving apparatus consisting of a key-instrument connected, with the line-wire and permitting the current sent into the line from the transmitting station, to pass into a conductor leading direct to one of the wires running to earth and to which is connected the receiver proper.

In order to avoid the polarization of the earth-wires, each such wire is connected with a conductor leading to a copper or other metallic blade immersed in an impolarizable saline solution in which is disposed a second similar blade connected to earth or sea..

In order that my invention may be clearly understood and readily carried into effect I will now proceed to describe the same more fully with reference to the drawings accompanying my Provisional Specification in which:.

Figure 1, is a diagram showing my improved system as applied to telegraphy.

Figure 2, is a diagram showing the same applied to telephony.

Figure 3 shows the system of Figure 2 with induced current.

Figure 4 shows a modified arrangement.

In all the figures the same letters of reference designate the same parts.

Referring to Figure 1 the two stations are connected by a single line-wire a branched at the zero point b of the battery formed of the different elements c of 'which 'the end poles are connected to earth, or to sea, by a conductor d e leading to a copper or other metallic plate f immersed in a liquid of good conductivity, such as for example a solution of sulphate of copper, in which another copper plate f1 is placed connected to earth, or sea by a. conductor d1 e1.

A transmitting apparatus q of any convenient construction, such as a Morse-key instrument, is connected with the two end poles of the battery in such a manner that in its position of rest it closes the circuit of the latter by establishing a short circuit so that no current passes, (or at most a very weak one) through the wire line a or through the earth-wires.

To send a current to the receiving station, the key g is depressed whereupon the short circuit is interrupted and the current is caused to pass simultaneously through the line-wire a and the earth-wires d d1 e el. The line wire is in reality traversed by two currents in opposite directions, so that it is in an apparently neutral state, and thus non interception of the messages can be effected for the reason that if a receiver be .interpolated at any point of the line, said receiver would receive two currents in opposite directions which would neutralize each other

Each station is furnished also with a receiving apparatus consisting of a commutator h connected with the line-wire a and permitting, when depressed. the current from the transmitter to pass into a wire i leading direct to one of the earth-wires (d1 for example) and in which is interpolated the receiving instrument j.

To receive signals sent from the transmitting station, it is necessary that the commutator lever h be kept depressed, so as to cause the current from the transmitting station to pass into the wire i in connection with the receiving instrument.

The connections may also be established in such a manner that when the commutator h is in its normal position, it cuts the wire' a and connects the wire i with that part of the said wire which leads to the other station; in this case for transmitting a signal it is necessary to depress the commutator lever h previously so as to re-establish the continuity of the line-wire and the operator then acts on the key-lever as hereinbefore stated.

My arrangement is equally applicable to telephonic messages as will now be explained with reference to Figure 2. The transmitting instrument g is in this case replaced by a microphone 1 and the receiving instrument j is replaced by a telephone receiver il the operation of the arrangement is exactly the same as that above described.

The vibrations of the microphone cut the short circuit and the current from the local source of electricity thereupon passes over the line-wire a and the earth-wires d d1e e1 as previously stated,

The telephonic message is received in the same manner as before through the intervention of the commutator h which passes the current sent from, the transmitting station to the telephonic receiving instrument j1.

The line wire may also be operated by induction currents as is indicated at Figure 3.

The local circuit in which the transmitting apparatus (for instance, the microphone g1). is inserted comprises a ,primary winding k surrounding a central soft iron core,, or bundle of wires, 1 which is further provided with two secondary windings m n in opposite directions. The line wire a is connected to each of these secondary windings at the two extremities in n1 of opposite polarity and the opposite extremities m2 n2 are connected with the earth wires clt c ex in the manner already explained.

When operating the transmitting apparatus or when speaking into the microphone for example, induction currents are generated in the secondary windings m n and they then affect the receiving apparatus l at the receiving station as already stated. The primary winding and the two secondary windings may have the three wires wound in the same direction upon the soft iron core and the two extremities of the primary convolutions are connected with the local circuit in which is interpolated the transmitter while two extremities of opposite polarities of the secondary convolutions are connected to line and the two others to earth as previously explained.

In my improved transmitting, arrangement the. two earth wires at each. station must be the more separated from each other the longer the distance is between the two stations so that the resistance of the earth between the said earth-wires will be sufficient to cause the single line-wire to be traversed by the greatest possible quantity of the current.

In order to reduce as far as possible the distance between the earth-wires the beginning of the single line wire a and the wires d d1 leading to earth are advantageously removed at a distance front the transmitting instrument and receiving instrument respectively at the telegraphic or telephonic station and thus permit the two earth connections of the two telegraphic or telephonic stations to be placed nearer each other.

The telegraphic or telephonic station is then divided and the transmitting instrument is connected with the magnet with its earth connections and line wire z by two conducting wires k' k2 connected with the ends of the convolutions which generates Induction currents in the secondary windings m n (Figure 4).

The current sent into the conducting wires k1 k2 .and, passing through the convolutions k may itself be an induced current generated in the coil k3 when a current from the source of electricity c is passing through the primary coil or through the action of the transmitting microphone f/I for example.

In the case when the distance of the telegraphic or telephonic transmitting station from its earth connections is not too excessive the current traversing the conductors k le2 instead of. being an induction current may be furnished direct by the) source, of electricity of such station, and naturally transmitting instrument (microphone or key) and the source of electricity are then placed direct in the circuit leI k-S.

An interrupter It interpolated in the conductor k1 permits of cutting the circuit in leading to the secondary coil k3 and of closing the circuit for in conductor i leading to the conductor 3 in which is interpolated the receiving instrument j1.

Thus the coil k is either inductor or induction according to whether there is transmission or reception.

It is to be well understood that my system of transmission is entirely independent of the nature of the telegraphic or telephonic apparatus employed therewith, and that the source of electricity at the stations may be of any convenient kind,

---

Improvements in Connections of Telephonic and Telegraphic Instruments to Conducting Lines.
GB189706773

1897-03-15

Telegraph and telephone lines are arranged so that any pair of lines bearing instruments may be used as a lead or return for another set of instruments without interference. Thus, as an example, four metallic circuits carry four separate sets of instruments individually but may have circuits A and B, C and D, used as other metallic circuits, or A and B may form a lead and C and D a return for a seventh metallic circuit. The Specification describes telephone instruments as used on the metallic circuits and telegraph or telephone instruments inserted in the earth return circuits. Fig. 1 shows a pair of lines L<1>, L<2> forming a metallic circuit and capable of use as the line with an earth return through L<3>, L<3>, the lines L<2> being connected to the loop at the centre of the secondary coil S', S<2> of the induction coil X. The receiving-instrument is shown in a bridge, such as Z. The positions of the transmitting T and receiving R instruments may be reversed. Each side of a metal loop may also be used as the line with an earth return.

This invention relates to the connecting of telephonic and telegraphic instruments to conducting lines in such a manner that several telephonic and telegraphic messages can be simultaneously transmitted without interfering with one another.

Fig. 1 of the accompanying drawings shews diagrammatically connections for a telephone and also for another telephone or telegraph to a double conducting line.

Fig. 2 shews a modification.

Fig. 3 shews how seven sets of the connections may be applied to four double conducting lines.

Fig. 4 shews the connections for a telephone and telegraph to a single line with earth return.

Fig..5 shews connections for telephoning over two telegraph lines one of them forming a metal return.

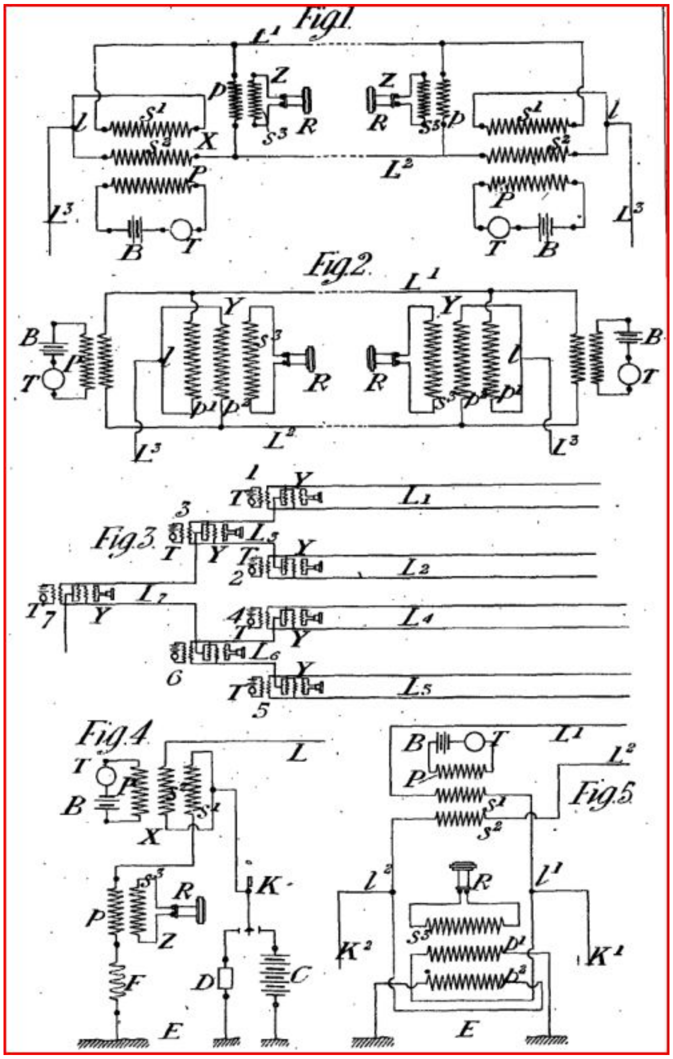

Referring first to Fig. 1 the two lines Ll L2 are at each end connected as shewn at X through two equal secondary coils 8 1 and s 2 wound with a primary coil P which is in circuit with a telephone transmitter T and local battery B.

The lines L1 L2 are so connected to the secondary coils 8 1 s 2 that the currents induced in the lines by the telephone transmitter are always in opposite directions.

The telephone receiver R, as shewn at Z, is in circuit with a secondary coil 8 3 wound with a primary coil p connecting the two lines L1 L2.

At a point l intermediate to the coils s1 s2 a third line L3 is connected, this line being employed for telephone or telegraph transmission. When either of the telephone transmitters T is employed, the induced currents pass through the primary coil p thus causing the receivers R to act. But should any current flow by L3 it must flow in opposite directions through 8 1 and s2, and therefore in the same direction along the lines L1 L2 so that it cannot influence the receivers R nor the transmitters T. At each end a line L3 may be connected as shewn by Y, Fig. 2, at l a point intermediate to two equal primary coils p1 p2 which connect the lines Ll LO and are wound with the secondary S3 of each telephone receiver R.

With connections thus arranged any current flowing in the same direction through the coils pl p 2 actuates the telephone receivers R, whereas a current from L3 flows in opposite directions through the coils pl p 2 and therefore is ineffective on the receivers R. By combining several ordinary transmitters T and several of the receiver arrangements marked Y in Fig. 2 with a number of double lines as shewn in Fig. 3, a number of messages can be sent over these lines without interfering with each other.

Thus referring to Fig. 3 the transmitter 1 works over the pair of lines L1 in the - usual way, one of these lines serving as metal return. In like manner the transmitter 2 works over the pair L2, transmitter 4 over the pair L 4 and transmitter 5 over the pair L 5. Again the transmitter 3 works through the pair of lines L 3 and the two pairs L 1 and L 2, one of these pairs serving as metal return.

In like manner the transmitter 6 works through the pair of lines L 6 and through the two pairs L 4 L 5. Finally the transmitter 7 works through the pairs of lines L 7 and through the two pairs L 3 and L 6 and the. four pairs L 1 L 2 L 4 and L 5, two of these pairs such as L 4 and L 5 serving as metal return.

Obviously this system may be duplicated, and reduplicated so that one transmitter can be made to work over 8 pairs or 16 pairs of lines or other multiple of 4.

As shewn in Fig. 4 a telephone and a telegraph are connected to one line L having earth return. In this case X and Z are the telephone arrangements so marked in Fig. 1, K is the key, C the battery and D the receiver for telegraphing in the connection of Z to earth there is an adjustable resistance F to balance that of the line. As shewn in Fig. 5 two sets of the telegraphic connections shewn in Fig. 4 and two lines Ll L2 which may be both to one place or each to a different place with earth return are combined with one telephone arrangement.

The transmitter T by its primary P induces currents in the secondaries $ 1 89 both in the same direction and therefore in opposite directions in the lines Ll L2. The currents being in the same direction in p1 and p2 act on the secondaries 1 3 of the receivers 1,. Currents from the telegraph gi dividing at 11 pass in opposite directions through sl and pl and therefore do not affect the telephonic receivers. And in like manner the currents from the telegraph K2 are ineffective on the receivers.

For the sake of clearness the primary and secondary coils have been shewn in the form of many lines side by side.

It is to be understood however that they are wound in each case all on one bobbin.

In Figs. 3,4 and 5 the arrangements are shewn at only one end of the conducting lines.

It is to be understood however that they are symmetrically repeated at the other end as shewn in Figs. 1 and 2.

COMPLETE SPECIFICATION.

This invention relates to the connecting of telephonic and telegraphic instruments to conducting lines in such a manner that several telephonic and telegraphic messages can be simultaneously transmitted without interfering with one other.

Figure 1 of the drawings accompanying my Provisional Specification shews diagrammatically connections for a telephone and also for another telephone or .telegraph to a double conducting line. Figure 2 shews a modification. Figure 3 shews how seven sets of the connections may be applied to four double conducting . lines: Figure 4 shews the connections for a telephone and telegraph to a single line with earth return. Figure 5 shews connections for telephoning over two telegraph lines one of them forming a metal return.

1 . Referring first to Figure 1 the two lines Ll L2 are at each end connected as shewn at X through two equal secondary coils s1 and 8 2 wound with a primary coil P which is in circuit with a telephone transmitter T and local battery B.

The lines Ll L 2 are so connected to the secondary coils 8 1 s'" that the currents induced in the lines by the telephone transmitter arc always in opposite directions.

The telephone receiver R, as shewn at Z, is in circuit with a secondary coil s3 wound with a primary coil p of high resistance connecting the two lines L1 L2.

At the point intermediate to the coils sr s 2 a third line L3 is connected, this line being employed for telephone or telegraph transmission. 'When either of tho telephone transmitters T is employed, the induced currents pass through tho primary coil p thus causing the receivers R to act. But should any current flow by L3 it must flow in opposite directions through sl and s2, and therefore in the same direction along the lines Ll L2 so that it cannot influence the receivers It nor the transmitters T. At each end a line L3 may be, connected as shewn by Y, Figure 2, at l a point intermediate to two equal primary coils p1 p2 which connect the lines L1 L2 and are wound with the secondary 8 3 of each telephone receiver R. With connections thus arranged any current flowing in the same direction through the coils yx actuates the telephone receivers R, whereas a current from L3 flows in opposite directions through the coils p1 p2 and therefore is ineffective on the receivers R. By combining several ordinary transmitters T and several of the receiver arrangements marked Y in Figure 2 with a number of double lines as shewn in Figure 3, a number of messages can be sent over these lines without interfering with each other.

Thus referring to Figure 3 the transmitter 1 works over the pair of lines L 1 in the usual way, one of these lines serving as metal return.. In like manner tho transmitter 2 works over the pair L 2, transmitter 4 over the pair L 4 and transmitter 5 over the pair L 5. Again the transmitter 3 works through the pair of lines L 3 and the two pairs L 1 and L 2, one of these pairs serving as metal return.

In like manner the transmitter ti works through the pair of lines JLA b and through the two pairs L 4 L 5. Finally the transmitter 7 works through the pairs of lines L 7 and through the two pairs L 3 and L 6 and the four pairs L 1 L 2 L 4 and L 5, two of these pairs such as L 4 and L 5 serving as metal return.

Obviously this system may be duplicated, and reduplicated so that) one transmitter can be made to work over 8 pairs or 16 pairs of lines or other multiple of 4.

As shewn in Figure 4 a telephone and a telegraph are connected to one line L having earth return. In this case X and Z are the telephone arrangements so marked in Figure 1, K is the key, C the battery and D the receiver for telegraphing.

In the connection of p to earth there is an adjustable resistance F to balance that of the line. As shewn in Figure 5 two sets of the telegraphic connections shewn in Figure 4 and two lines Ll L2 which may be both to one place or each to a different place with earth return are combined with one telephone arrangement.

The transmitter T by its primary P induces currents in the secondaries sl S2 both in the same direction and therefore in opposite directions in the lines L1 L2. The currents being in the same direction in pI and pS act on the secondaries i1 of the receivers It.

Currents from the telegraph K1 dividing at 1 1 pass in opposite directions through sl and I > I and therefore do not affect the telephonic receivers. And in like manner the currents from the telegraph K2 are ineffective on the receivers.

The connections of p1 and p2 to earth may have suitable resistances as shewn at F in Figure 4.

For the sake of clearness the primary and secondary coils have been shewn in the form of many lines side by side.

It is to be understood however that they are wound in each case all on one bobbin.

In Figures 3,4 and 5 the arrangements are shewn at only one end of the conducting lines.

It is to be understood however that they are symmetrically repeated at the other end as shewn in Figures 1 and 2.,

---

Perfectionnements aux bobines d'induction
FR382386

1908-02-05

The present invention relates to an improved induction coil in order to obtain induction currents in very short periods. These currents, whose durations are extremely short 5, enjoy special properties which find application mainly in telegraphy and wireless telephony.

This induction coil is characterized by the absence of the magnetic core and the winding mode of the inductor and the armature on a short length coil, but a very large diameter, the inductor coil being virtually drowned in May 1 the armature winding.

The appended drawing schematic represents for example an induction coil according to the present invention.

FIG. 1 is a section along line A-20B of FIG. 2 ;

FIG. 2 is a plan.

Fig.1

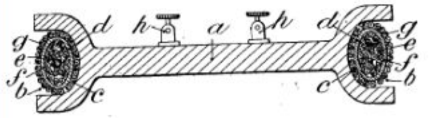

Fig.2

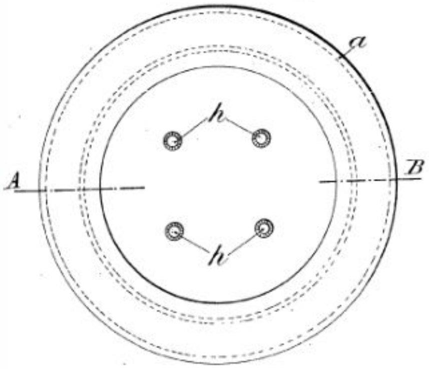

As shown in these drawings, this coil has a very large diameter which can be up to several meters; 26 is drawing the round, but it could equally be square or affect any other polygonal shape. This reel is very long because all winding son are housed on its outskirts in a suitable groove b; it will be understood by 3o begins winding on the rim of this kind of pulley a first layer of wire induced E and then covered with a suitable insulating and d are wound over this insulating layer the inductor wire E which is covered another 35 isolated layer; we **réuuïi uêUi** the insulating layers d and f, so as to isolate the inductors e and ends the coil by an external winding armature wire g. ko

Is carried out in this way a large-diameter ring compound induced son c, g, containing a central core of an adequately insulated son inductors.

The machine has four terminals 5 h which two are used to secure the ends of the armature winding and the other two ends of the inductor winding.

We can arrange several induced son and son more inductors; in this case, each terminal 5o h will be replaced by a multiple switch that will bring together the different armatures on, and the different inductors on, either voltage or the contrary in quantity, following the effects that one wishes to obtain.

Dispositif de transmission téléphonique
FR377785

1907-09-14

Telephone transmission system of mounting each station transmitter and or serial receivers and to bring the two stations by three son line, including 5o first two respectively together the poles of the two stations and the third is mounted derivation at a point of each station caught between the transmitters and the receiver or receivers, one or even two of these three son may be replaced by equivalent landings.

The present invention provides a way to increase the distance at which it is possible to call or to increase the intensity of sound transmitted at a given distance 5.

When installing a regular telephone line, it is customary to bring the two stations using two wires, one for the second round for the return; at-10 there is no fear induction from neighboring lines, one connecting the two stations by a thread alone and return through the ground at each station.

FIG. 1 of the accompanying drawing shows the 15 most simple provision of a regular line with two son; there is one at each station magnetic phone, a1 met by both son line c and d; when speaking in the phone, the phone is 30 a1 receiver and vice versa; this provision is entirely satisfactory from the standpoint of the actual transmission, but it has the obvious disadvantage that the two parties can not break interfere, each necessarily listen to while the other speaks .

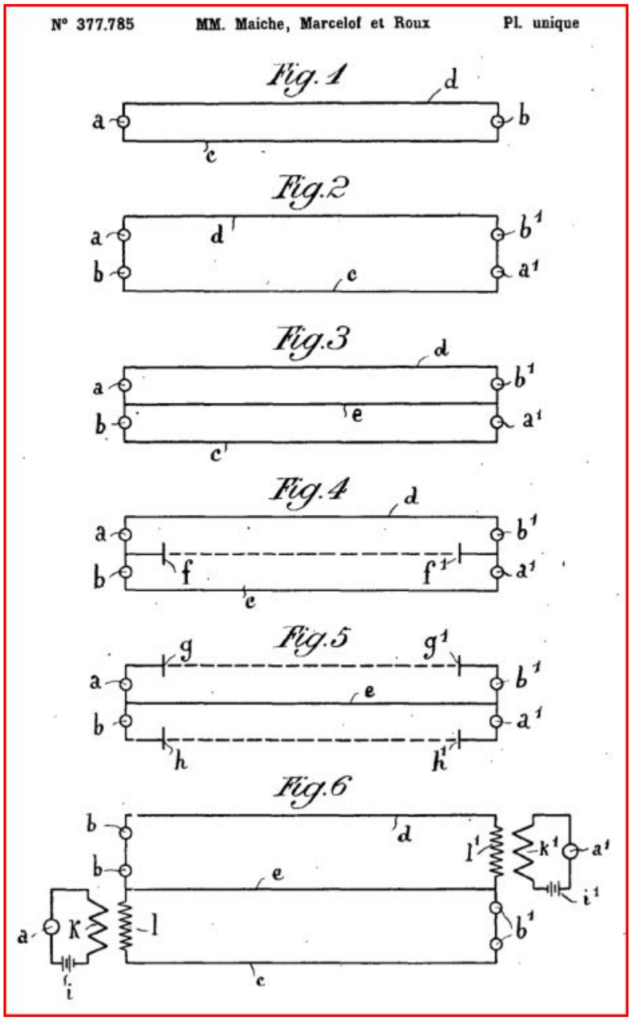

To overcome this disadvantage, it immediately offered to have each transmitter station a phone or a 1 and a receiver or telephone b bl, allowing 3 to the two interlocutors to talk while listening. This solution suffers a disadvantage from the standpoint of the transmission; in fact, as shown in Fig. 2, the four phone a, b, a}, bx 3 are mounted in series on the line cd, so that the sound transmitted by a transmitter has or a1 necessarily travels in addition interesting b1 or b receptor both other phones of the circuit. The telephone circuit thus unnecessarily h has a high resistance which reacts disadvantageously either on the possible length of the line cd, either the intensity of the sound transmitted to a given line length.

The present invention relates to a device easy positive h to remedy the drawbacks both of the two transmission systems according to Fig. 1 and 2 and make a new telephone transmission system providing both the benefits of the 5 two known systems.

As shown in Fig. 3, this invention is simply to dispose between the two stations a third e line wire terminating at each station between the telephone transmitter 5 tor a or a1 and the receiver phone b or b1. In these conditions the wall sent by "will be fully received by b1 without crossing unnecessarily b and a1 resistance and vice versa speech transmitted by a1 will be fully received by b.

The need for three son line c, d, e 5 is required when the line is around other transmissions; but when isolated, it can simplify the transmission as shown in Fig. k or 5.

According to FIG. h, the wire line drifted o-tion e is deleted and replaced by a ground f and f1 taken at each station between the transmitter and receiver.

In special cases, it is even possible to reduce the line to a single wire e in '5 with each station two landings g, h and # 1, h1), the receiver and transmitter are each disposed station in series between the two landings; under these conditions, the two conductors c and d are formed by o the ground and there is no longer a single line wire e.

In the above, to simplify the description and drawings, we assumed the exclusive use of magnetic phones; 5 Fig. 6 shows the circuit diagram in the case of induction phones. In this case each position i includes a battery, a transmitter microphone and primary k of an induction coil, mounted in series o on a closed circuit; the line circuit c, d comprises series headphones b, b and b1, b1 and side and ll of the coil. In this

If the invention is to attach the derivative e line wire between the two receivers b, b or i1, A1 on the one hand and the secondary or ll 35 on the other hand, as shown in FIG. 6.

The result of these provisions is summed up by the possibility of quadrupling of the intensity of the reception for a given length of line, or the length of the line Zio maintaining the same intensity.

Nouveau mode de transmission sans fils des courants télégraphiques et téléphoniques en employant comme conducteurs la terre et l'eau
FR320267

1902-12-05

To understand our invention, it is good to remember that in an ordinary telegraph line a thread, more or less isolated, from a point to another, communicates with earth at each end. A power source; and a transmitter unit are sandwiched between the starting line and earth. Upon arrival, a receiver also communicates with the line and the earth.

We know that each of the posts supporting the line passes a small amount of power that is lost in the ground. Large amount of current is now lost all **plus 1 large 5** that the distance between stations is itself larger and, consequently, the number of' poles is higher. Loss can become considerable, without preventing transmissions. It may even happen that the ao thread without being broken, drags on earth there for a certain length of son'parcours and continues to allow the current of a shift to another, albeit very weakly,

The experience proved us uninsulated conductor could even, for certain distances, be simply placed on the ground and that transmissions would nevertheless be possible.

We are able to continue, transmit signals is greatly increased if, at each end, the uninsulated wire touching the ground, met in a well-insulated wire that extends the line to a ground to the point where so are transmitters and receivers.

It is understood that the ground at the end of a secluded part forces the return current from sinking into the depths of the earth ho and that, consequently, there is an infinitely small closure point where the uninsulated wire touches the ground. In other words, the loss suffered in ground contact, the uninsulated wire, is even less than each end of uninsulated wire so is an insulated wire from the ground at a greater distance.

Hence, if the lead wire, instead of iron, consisting of a much less conductive material, coal, for example, everything would be the same way. We also recognized that even when iron, coal or other similar material, are replaced by placed water, salt if possible, and, consequently, by a stretch of sea or any other part of sea whose geographic provision lends itself to this kind of installation.

When it comes to cross an arm of the sea, the isolated line extension is placed preferably in the direction of the straight line through the two points to be connected.

In case one transmission on a role, isolated leagues will be perpendicular to the straight line joining the two stations.

Isolated lines will not be too long, and therefore, will provide not only coast to coast, but between two points inland. The thickness of the salt water layer being at even greater depths, only about the thickness of the land mass, the proportions remain between water and adequate the earth so that water plays exactly the same role that the uninsulated conductor touching the ground.

To make our evplications as clear as possible, we have shown diagram-matically, in the accompanying drawing, an installation in which our invention is applied.

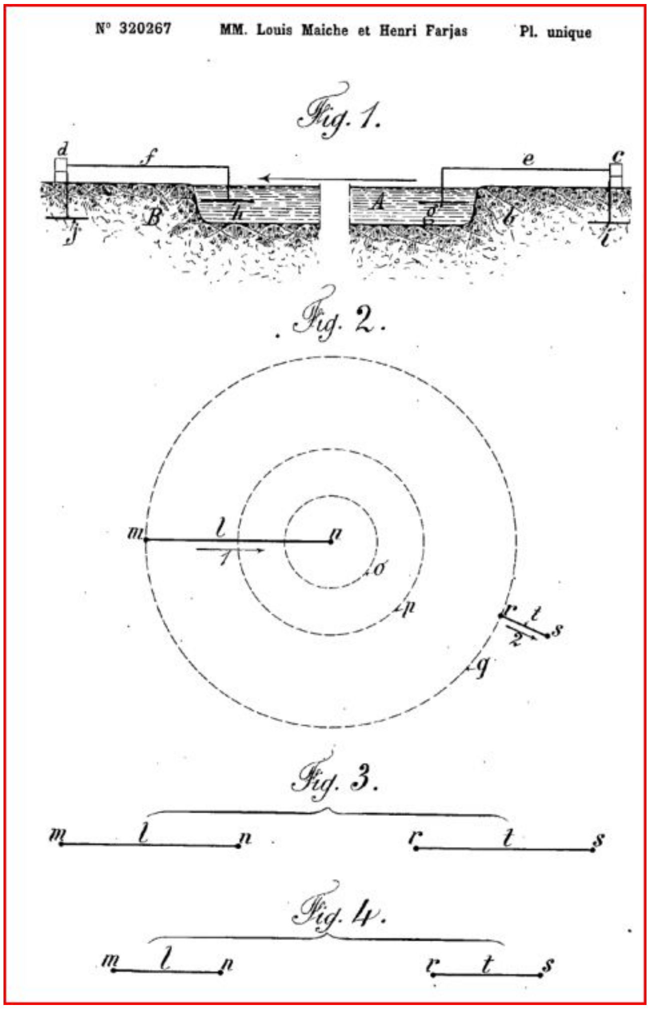

A is the sea (Fig. I) and 13 b are both shores, c and d both stations, e and f insulated, and the plates located in the water, i and it updates earth.
We see that the sea is an extension isolated son e and f, and it plays the role of an uninsulated conductor based on floor.

To make it clear how 3 o operates the propagation of electric currents, we assume an isolated line (Fig. A) within two earthed m and n. If the current is moving in the direction of the **llèche**, electricity arriving ashore in n flourishes spheres 3.) concentric o, p, q, etc., all points of each sphere simultaneously receiving the same amount of electricity if the propagation medium is uniform. Be grounded in r therefore receive the current along the earth m, r and if this land is connected by a line by t, s land to another, even very close to her, the current finding less resistance in the isolated line in the middle (Jun separates the two nets land, will follow this line, as shown in llèche a. It will be thus possible to establish a wireless communication between any point of the line / on which find transmitting devices, and any point on line T which will be **appareds** receptors.

To establish simultaneously on each of the leagues and t, a position transmitter and a receiver will require that the line is the same length as the line 1, J J as shown in Fig. 3.

If the medium is found to be twice as conductor between h and r between m and n, on the one hand, and between r and s on the other hand, the lines l and t may have as half the length of the distance between the points n and r (fig.û).

Of course, transmissions based on the principle that we have indicated, may be made by taking as con- bo different producers, not only the earth and the sea water, but the earth and air, land in this case, acting as a better conductor brought by the air; the insulated wire from the earth lead to the air at each station 7. tion to a more or less remote distance according to the distance that would take. It is obvious that the latter provision, the sea and fresh water advantageously replace the earth. 7°

In summary, we claim as our exclusive property, the new method, described above, transmission without telegraph and telephone currents, said transmission mode of making pass in a conductive medium the current emitted by the base station and to receive on arrival in this current an insulated wire of suitable length, which are sandwiched the receivers and resulting in a second medium whose specific conductivity is less than that of the first medium, but passes the total quantity of electricity because of its mass which is greater than that first medium, an insulated wire, on which the transmitting devices is also leading into the second medium, so that the first medium is an extension of two isolated lines, and the second medium acts as the return wire, the two isolated lines to be used both for shipping and for the reception.

Un nouveau mode de transmission et de réception des courants télégraphiques et téléphoniques sans autre conducteur que la terre ou l'eau

FR318007

1902-10-03

In an electrical circuit consisting of a power source and a metallic conductor whose ends are grounded, which closes the circuit, the earth acts as 5 driver not only a straight line between the two landings, but on a almost infinite extent; However, with a power density decreasing in proportion to the distance.

If parts of the ground are better conductors than others (because of the moisture, the saltiness of the water or any other cause) the current preferably passes through these parts describing curves or bends as does a river following the slopes of the terrain.

It follows that if, on part of the land used as return conductor, it artificially creates a better conductor line as the earth, it will be traversed by the current in preference to any other lower conductive part.

The effect is much more complete and can collect the current and note its existence when the best conductive line is insulated from earth and communicates with it by its ends. The current that would have crossed the land finding a wise not easier by the wire preferably follows and restores much faster electrical neutrality of the ground. It is isolated by cutting the line at any point that can be inserted in the course of the current instruments (telegraphic or telephonic receivers) that can detect its existence.

It is understood that if the isolated artificial line is a very good conductor, if its ends are connected to metal surfaces large enough in contact with the ground, the closing of the circuit will cause the ground to be in a state of electrical neutrality absolute.

Under these conditions, if a current electric is sent intermittently in the metal line, charge the earth, for each issue, a certain amount of electricity instantaneously discharge into the isolated artificial line.

Experience has shown that these phenomena can occur even when there is almost no gap between the two landings transmitter station or between two plates that collect the color rant crossing the line isolated the receiving station, to the condition that at each station the plates which lead to the two lands are carefully insulated from each other by means of a sufficiently thick insulating separation that no induction effect or condensation on one another can occur, and fairly large compared to the dimensions of the plates to prevent current from closing on itself too close to the to start. The current is so obliged to flourish "right and left of the landings, and the point where the positive and negative electric waves begin to reach is considerably related. I can further increase the desired effect by minimizing the current closure on itself through a number of insulating walls arranged below the plates to prevent **farm-i o ture** of the circuit by the depth of the earth, on the side opposite that in which the transmission must be received and, if necessary, above the plates.

In general, the plates will be placed vertically on either side of an insulating wall, their smaller dimensions being turned to the opposite position.

I can also use a wide insulating tube divided in two by a thick wall normal to isolated generators. The plates will be set back to back on either side of this partition in the interior of the tube and placed in communication with the outer conductor through insulated **son**.

The tube may be straight, or consists of two parts forming between them an angle properly calculated so that the circuit is closed by the earth is done by almost entirely in the **3 o** where the current is sent.

It is obvious that the installation is completed by transmitters and telegraph or telephone receivers that may be any, as the source of electricity.

I claim as my exclusive new property above described transmission and reception of telegraph and telephone currents without other driver as land or water, said system of separating the two landings each position by established insulating material to prevent the production of induction phenomena of condensation and the circuit closing on or by the depth of the earth, and force the air to flourish right and left, so that they can begin to get to and distant from the transmitter station, the two earthed of each position can thus without drawbacks, are moved towards one another until they fit on either side of the same insulating wall.

Nouvelle disposition de bobine d'induction
FR647 (E)
1903-02-18

Système de communication télégraphique et téléphonique

FR376100
1907-07-30

Mode de transmission de signaux télégraphiques
FR361733
1906-10-23

Nouveau moyen de transmission et de réception sans fil des courants électriques
FR321454
1903-01-10

Récepteur télégraphique pour toutes transmissions en général, avec ou sans fils
FR320467
1902-12-11

Nouveau récepteur des courants électriques
FR319195
1902-11-06

Nouveau moyen destiné à augmenter la sensibilité des microphones
FR319155
1902-11-06

Nouveau système de télégraphie et de téléphonie sans fil
FR318507
1902-10-17

Système de communication télégraphique et téléphonique
FR7464 (E)
1907-08-09

Dynamo - oder magnetelektrische Maschine
DE46075
1889-02-05

Transformateur téléphonique sans noyau magnétique
CH44013
1909-07-01

Wicklung von eisenfreien Transformatoren für Fernsprechzwecke.
AT40498
1910-01-10

Système de télégraphie et de téléphonie sans fil
FR392252
1908-11-21

Dispositif pour transmetteur et récepteur téléphonique
FR390346
1908-10-02

Other Patents

METHOD OF PRODUCING MOTIVE POWER
CA78259
1902-11-18

Appareil cinématographique
FR388334
1908-08-08

Perfectionnements aux papiers photographiques
FR400693
1909-08-04

Réducteur de pression pour tous fluides
FR361404
1906-06-22

Surchauffeur de vapeur pour locomotives
FR352810
1905-08-22

Apparatus for Heating Steam.
GB189801540
1898-01-19

An Improved Method for the Production of Motive Power by Heat and Fluid Pressure
GB190227627

1903-11-12

Improved Means of and Apparatus for Producing Motive Power
GB190226257
1901-11-29

An Improved Method of Producing Motive Power
GB190202489
1903-01-29

épuration bactériologique de l'eau
FR350264
1905-12-18

Dispositif générateur de vapeur
FR350080
1905-10-13

Perfectionnement aux moteurs à explosions
FR349940
1905-07-27

Mode de production de force motrice
FR349939
1905-07-27

Moyen de propulsion et de direction des navires
FR349935
1905-07-27

Nouveau mode d'application de la vapeur surchauffée
FR339220
1905-01-17

Dispositif pour la condensation de la vapeur d'échappement avec récupération de chaleur
FR337543
1904-04-13

Nouveaux verres d'optique
FR322234
1903-01-30

Improvements in Apparatus for Sterilising Water.
GB189705007
1898-02-19

An Improved Process for the Purification of Spirit.
GB189704089
1898-02-12

An Improved Process for Separating Heavy from Light Particles by Centrifugal Action.
GB189704087
1898-02-12

Improvements relating to the Purification of Water, and to Apparatus therefor.
GB189312941
1894-06-23

Improvements in Apparatus for Heating and Cooling Fluids by Fluids.
GB189706157
1898-03-0

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

US008362745B2

(12) **United States Patent**
Tinaphong

(10) **Patent No.: US 8,362,745 B2**
(45) **Date of Patent: Jan. 29, 2013**

(54) **METHOD AND APPARATUS FOR HARVESTING ENERGY**

(75) Inventor: **Prapan Paul Tinaphong**, Westfield, IN (US)

(73) Assignee: **Audiovox Corporation**, Hauppauge, NY (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 27 days.

(21) Appl. No.: **12/930,421**

(22) Filed: **Jan. 6, 2011**

(65) **Prior Publication Data**

US 2011/0175461 A1 Jul. 21, 2011

Related U.S. Application Data

(60) Provisional application No. 61/335,459, filed on Jan. 7, 2010, provisional application No. 61/402,481, filed on Aug. 31, 2010, provisional application No. 61/404,290, filed on Sep. 30, 2010.

(51) **Int. Cl.**
H01M 10/46 (2006.01)
H01M 10/44 (2006.01)
H02J 7/00 (2006.01)

(52) **U.S. Cl.** **320/108**; 320/101

(58) **Field of Classification Search** 320/108
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,127,799	A	10/2000	Krishnan	320/104
6,289,237	B1	9/2001	Mickle et al.	600/509
6,615,074	B2	9/2003	Mickle et al.	600/509
6,856,291	B2	2/2005	Mickle et al.	343/701
6,882,128	B1	4/2005	Rahmel et al.	320/101
7,057,514	B2	6/2006	Mickle et al.	340/572.7
7,084,605	B2	8/2006	Mickle et al.	320/101
7,283,053	B2	10/2007	Mickle et al.	340/572.2
7,373,133	B2	5/2008	Mickle et al.	455/343.1
7,383,064	B2	6/2008	Mickle et al.	455/573
7,400,253	B2	7/2008	Cohen	340/572.1
7,403,803	B2	7/2008	Mickle et al.	455/573
7,440,780	B2	10/2008	Mickle et al.	455/573
7,528,698	B2	5/2009	Mickle et al.	340/10.34
7,565,968	B2 *	7/2009	Lindley	206/223
7,567,824	B2	7/2009	Mickle et al.	455/573
2003/0199778	A1	10/2003	Mickle et al.	600/509
2004/0085247	A1	5/2004	Mickle et al.	343/701
2004/0259604	A1	12/2004	Mickle et al.	455/572
2005/0104553	A1	5/2005	Mickle et al.	320/101
2005/0210340	A1	9/2005	Townsend et al.	714/701
2006/0058076	A1	3/2006	Mickle et al.	455/574
2006/0094425	A1	5/2006	Mickle et al.	455/434
2006/0136007	A1	6/2006	Mickle et al.	607/45
2006/0168758	A1	8/2006	Chiang et al.	16/342
2006/0281435	A1	12/2006	Shearer et al.	455/343.1
2007/0109121	A1	5/2007	Cohen	340/539.26

(Continued)

OTHER PUBLICATIONS

Notification of Transmittal of the International Search Report and the Written Opinion of the International Searching Authority, or the Declaration (in English) dated Mar. 3, 2011, International Search Report (in English) dated Mar. 3, 2011 and Written Opinion of the International Searching Authority (in English) dated Mar. 3, 2011.

Primary Examiner — Samuel Berhanu
(74) *Attorney, Agent, or Firm* — Gerald T. Bodner

(57) **ABSTRACT**

An energy harvesting circuit includes one or more broadband or narrow band antennas to detect WIFI (wireless fidelity) or other RF (radio frequency) signals. The signals are rectified and voltage multiplied, and the resultant DC voltage is provided to a power management circuit. The output of the power management circuit charges a lithium battery or other storage device within the energy harvesting circuit. The energy stored in the battery or storage device is provided through a DC/DC converter circuit to a USB output connector to power or recharge the batteries of an external electronic device connected to the USB output connector of the energy harvesting circuit.

59 Claims, 11 Drawing Sheets

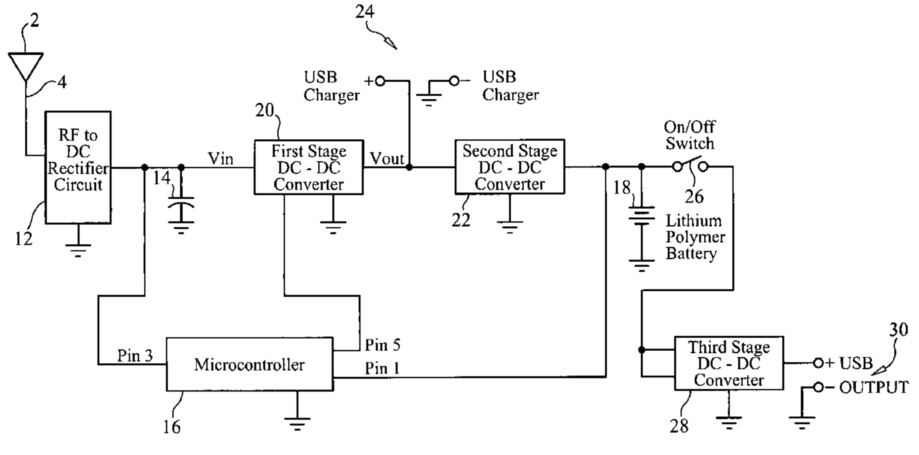

U.S. PATENT DOCUMENTS

2007/0142872	A1	6/2007	Mickle et al.	607/45
2007/0173214	A1	7/2007	Mickle et al.	455/127.1
2007/0191908	A1	8/2007	Jacob et al.	607/48
2008/0174436	A1	7/2008	Landt et al.	340/572.7
2009/0098915	A1	4/2009	Mickle et al.	455/573
2009/0105782	A1	4/2009	Mickle et al.	607/45
2009/0200985	A1	8/2009	Zane et al.	320/108
2009/0207000	A1	8/2009	Mickle et al.	340/10.34
2009/0284245	A1	11/2009	Kirby et al.	323/318
2009/0303076	A1	12/2009	Setiadi et al.	340/870.01
2009/0310393	A1	12/2009	Mickle et al.	363/126
2010/0073188	A1	3/2010	Mickle et al.	340/8.1

* cited by examiner

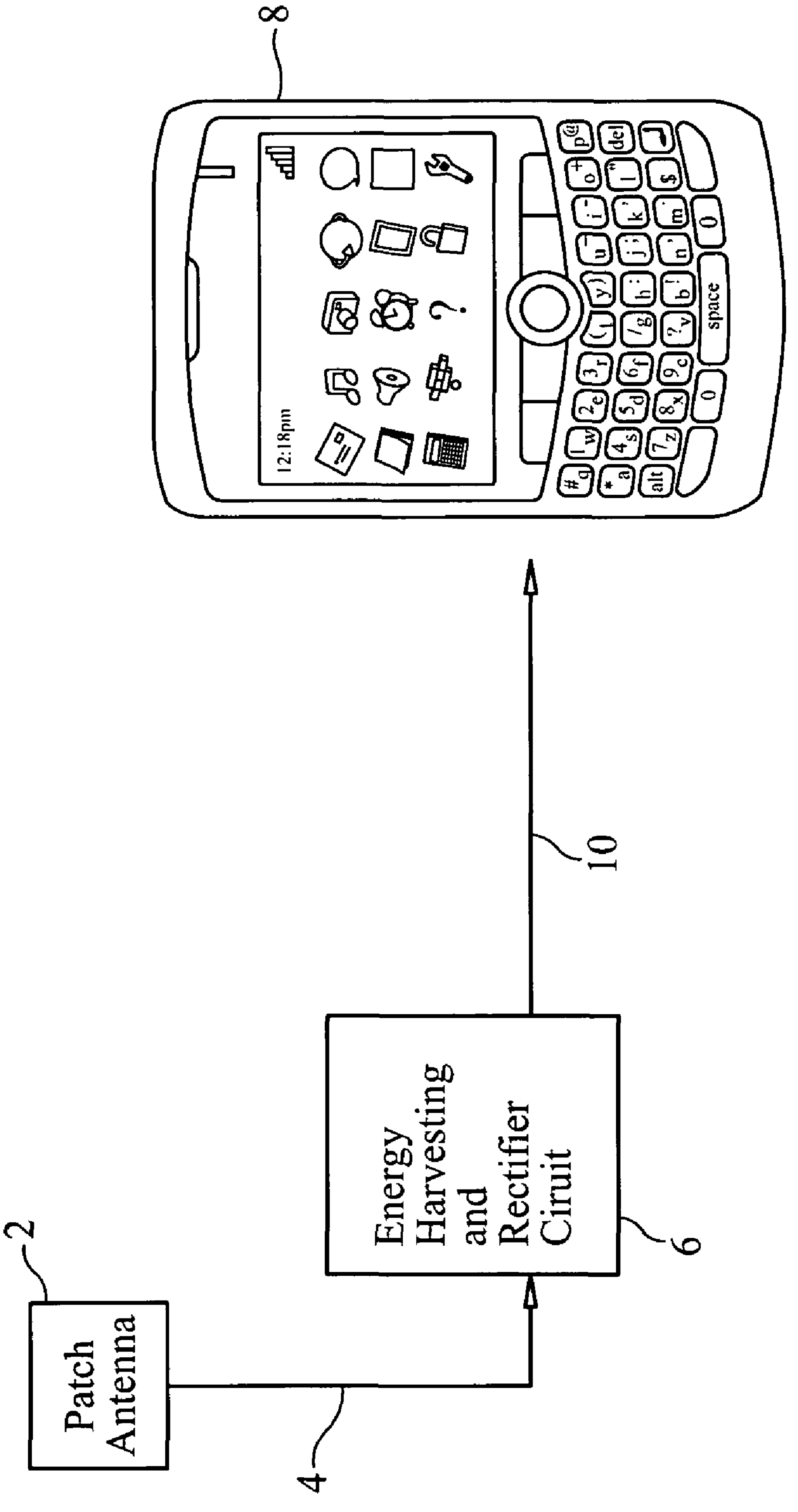
Patch Antenna
2
4
Energy Harvesting and Rectifier Ciruit
6
10
8
12:18pm
space
alt
del

FIG. 1

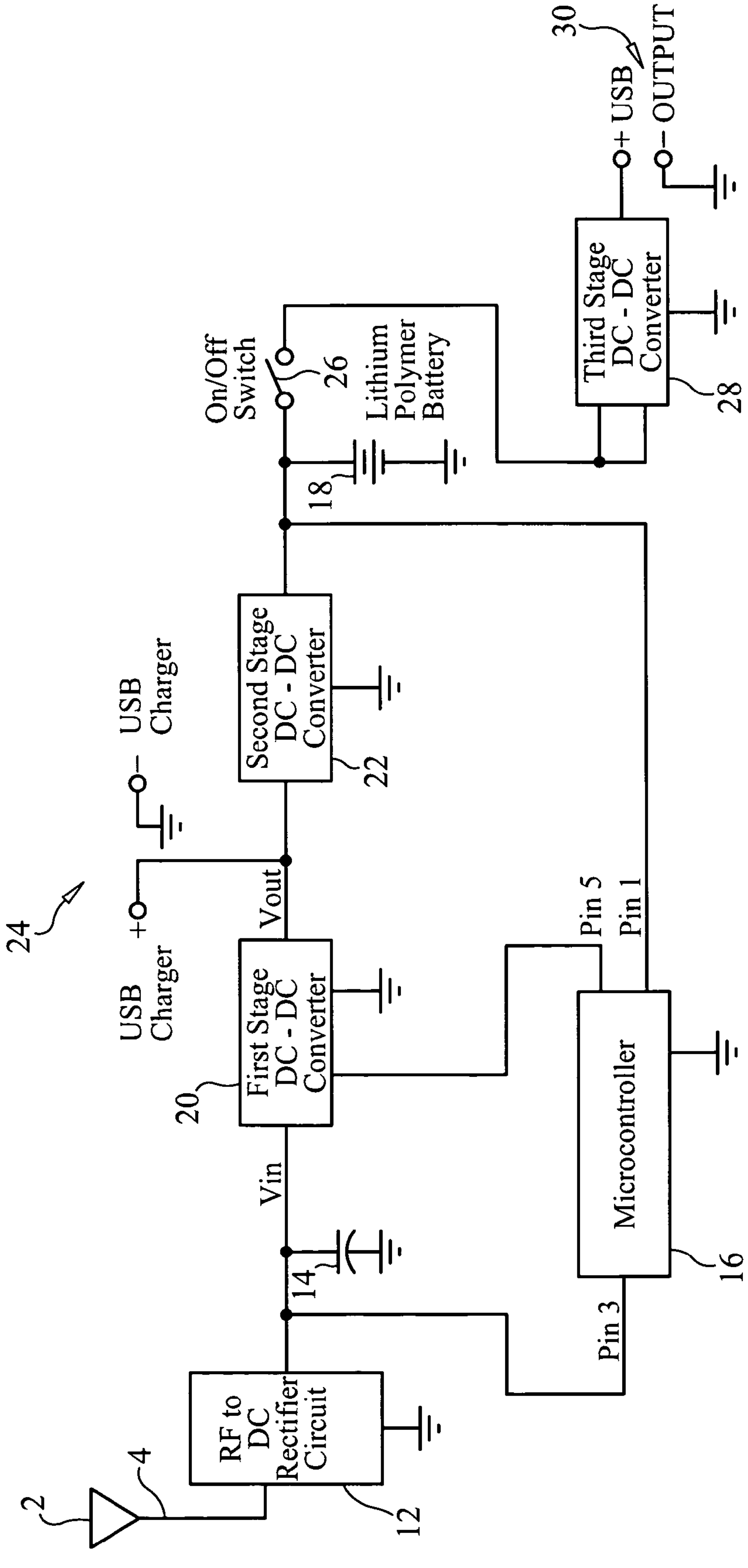
2
4
RF to DC Rectifier Circuit
12
14
Vin
Pin 3
Microcontroller
16
First Stage DC - DC Converter
20
USB Charger
+
Vout
24
Pin 5
Pin 1
USB Charger
−
Second Stage DC - DC Converter
22
18
Lithium Polymer Battery
On/Off Switch
26
Third Stage DC - DC Converter
28
+ USB
− OUTPUT
30

FIG. 2

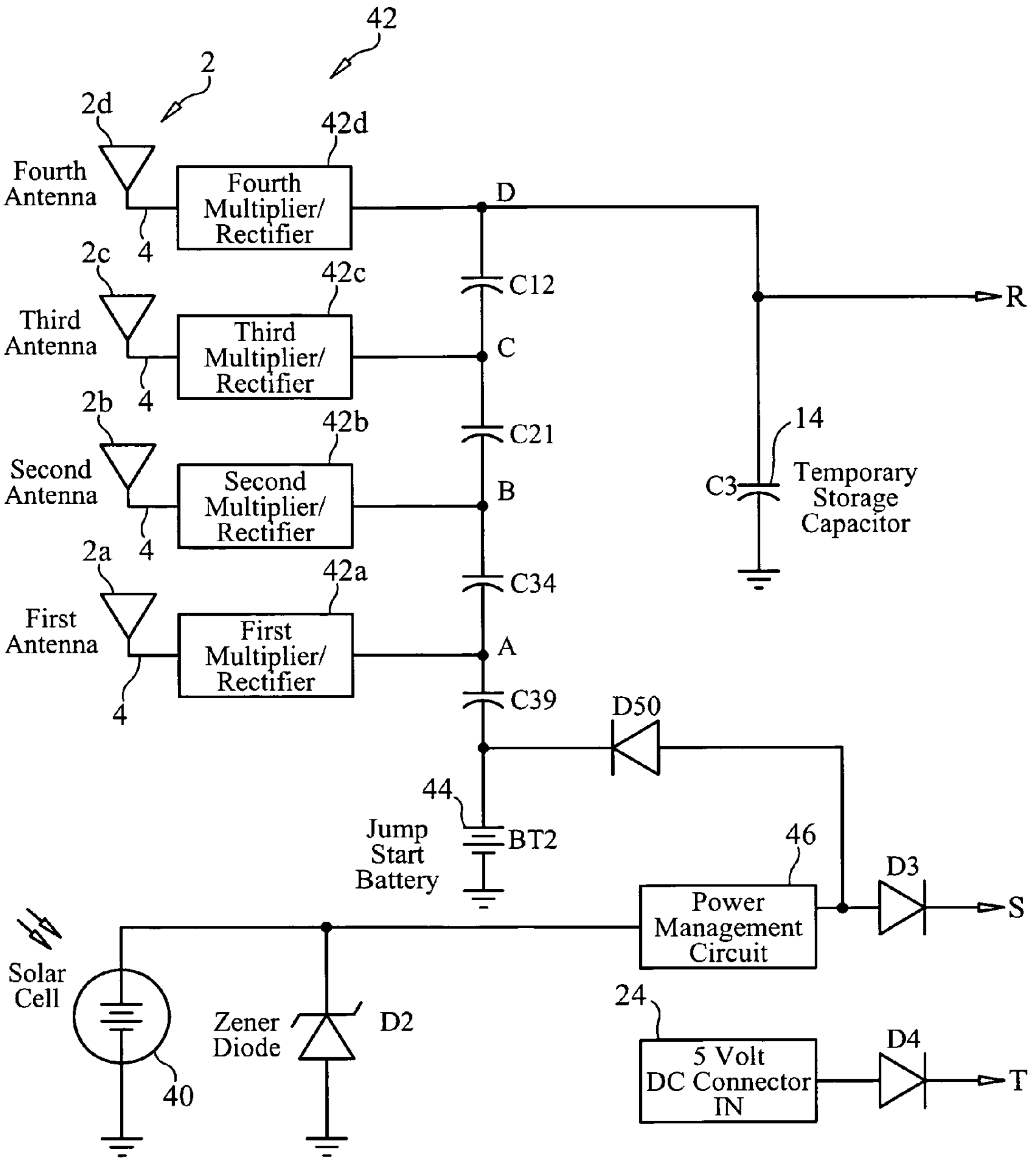
42
2
2d
Fourth Antenna
42d
Fourth Multiplier/ Rectifier
D
4
2c
Third Antenna
42c
Third Multiplier/ Rectifier
C12
C
R
4
2b
C21
14
Second Antenna
42b
Second Multiplier/ Rectifier
B
C3
Temporary Storage Capacitor
4
2a
C34
First Antenna
42a
First Multiplier/ Rectifier
A
4
C39
D50
44
Jump Start Battery
BT2
46
Power Management Circuit
D3
S
Solar Cell
Zener Diode
D2
40
24
5 Volt DC Connector IN
D4
T

FIG. 3A

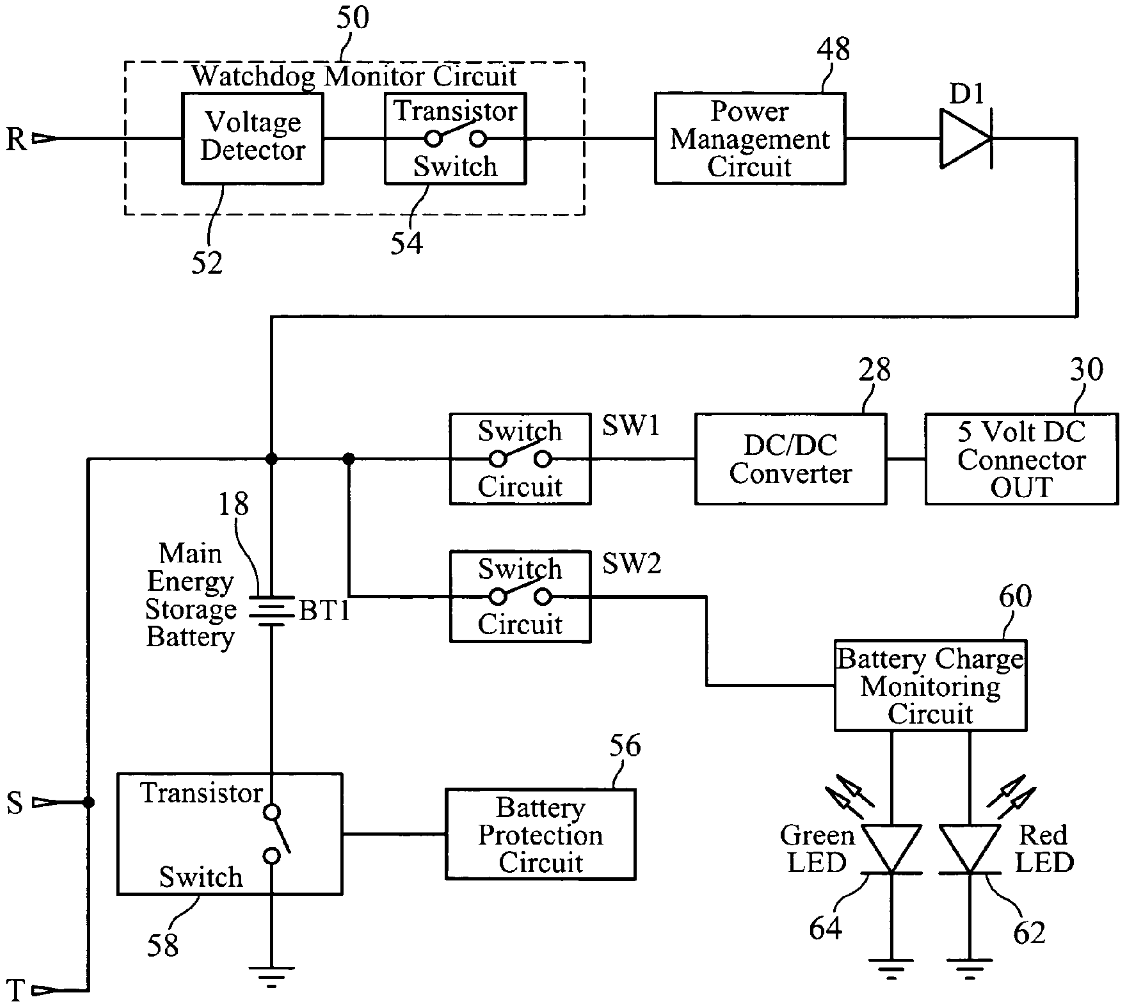
50
Watchdog Monitor Circuit
R
Voltage Detector
Transistor Switch
52
54
48
Power Management Circuit
D1
28
30
Switch Circuit
SW1
DC/DC Converter
5 Volt DC Connector OUT
18
Main Energy Storage Battery
BT1
Switch Circuit
SW2
60
Battery Charge Monitoring Circuit
56
S
Transistor Switch
Battery Protection Circuit
Green LED
Red LED
58
64
62
T

FIG. 3B

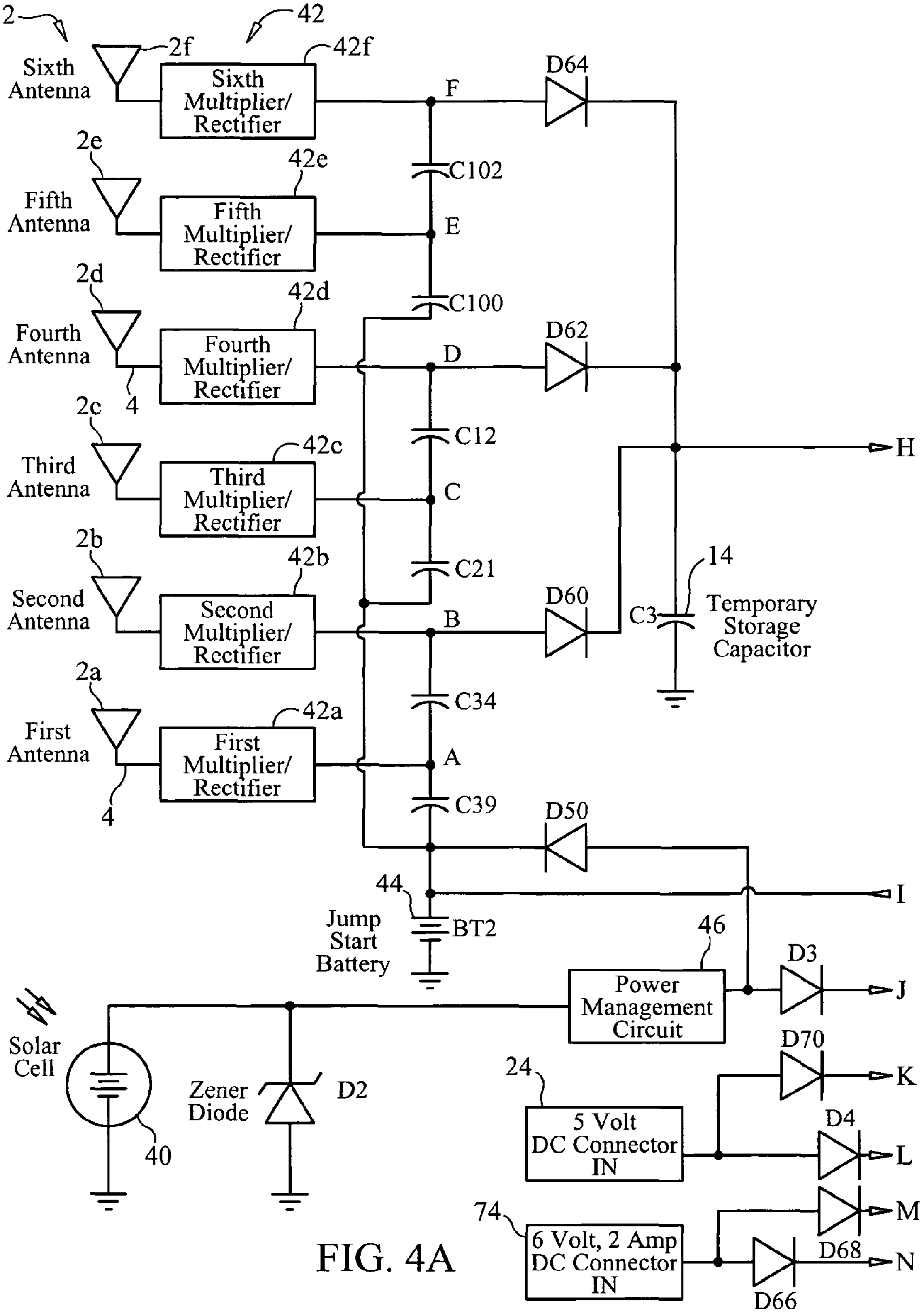
2
42
2f
42f
Sixth Antenna
Sixth Multiplier/ Rectifier
F
D64
2e
42e
C102
Fifth Antenna
Fifth Multiplier/ Rectifier
E
2d
42d
C100
D62
Fourth Antenna
Fourth Multiplier/ Rectifier
D
4
2c
42c
C12
H
Third Antenna
Third Multiplier/ Rectifier
C
2b
42b
C21
14
D60
Second Antenna
Second Multiplier/ Rectifier
B
C3
Temporary Storage Capacitor
2a
42a
C34
First Antenna
First Multiplier/ Rectifier
A
4
C39
D50
44
I
Jump Start Battery
BT2
46
D3
Power Management Circuit
J
Solar Cell
40
Zener Diode
D2
D70
24
5 Volt DC Connector IN
K
D4
L
74
6 Volt, 2 Amp DC Connector IN
M
D68
N
D66

FIG. 4A

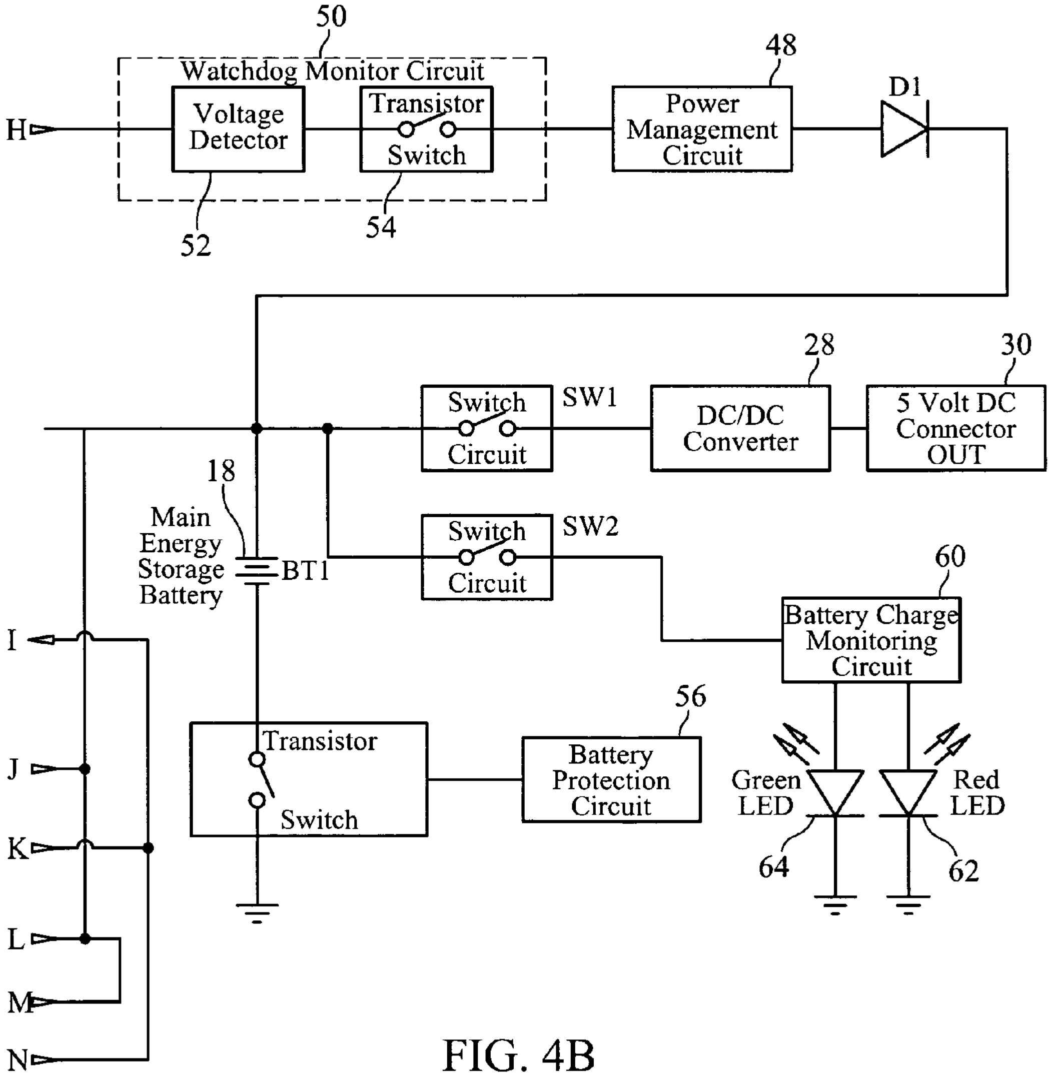
50
Watchdog Monitor Circuit
48
D1
H
Voltage Detector
Transistor Switch
Power Management Circuit
52
54
28
30
Switch Circuit
SW1
DC/DC Converter
5 Volt DC Connector OUT
18
Main Energy Storage Battery
BT1
Switch Circuit
SW2
60
Battery Charge Monitoring Circuit
I
J
K
L
M
N
Transistor Switch
56
Battery Protection Circuit
Green LED
Red LED
64
62

FIG. 4B

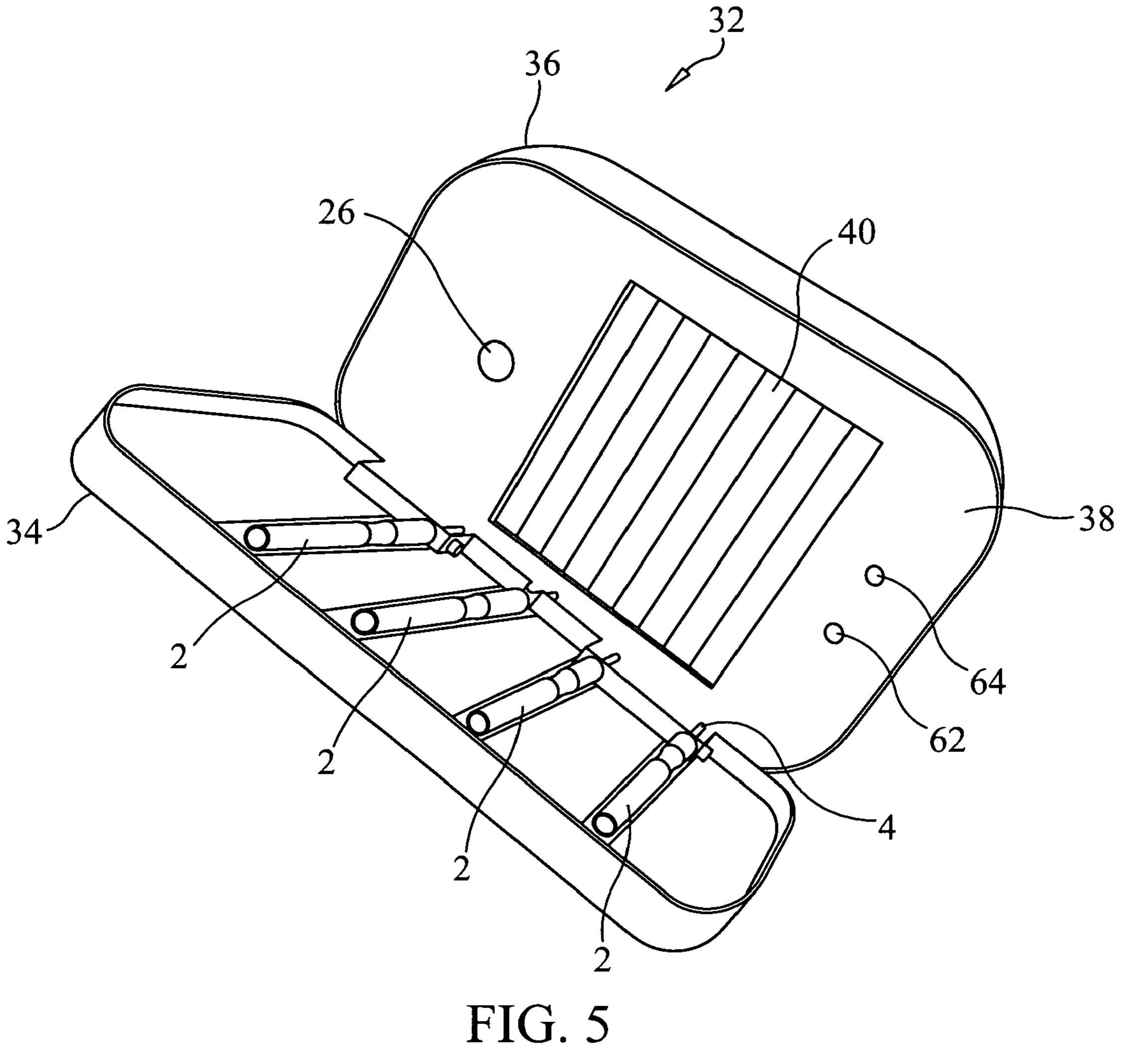
32
36
26
40
38
34
64
62
2
2
2
2
2
4

FIG. 5

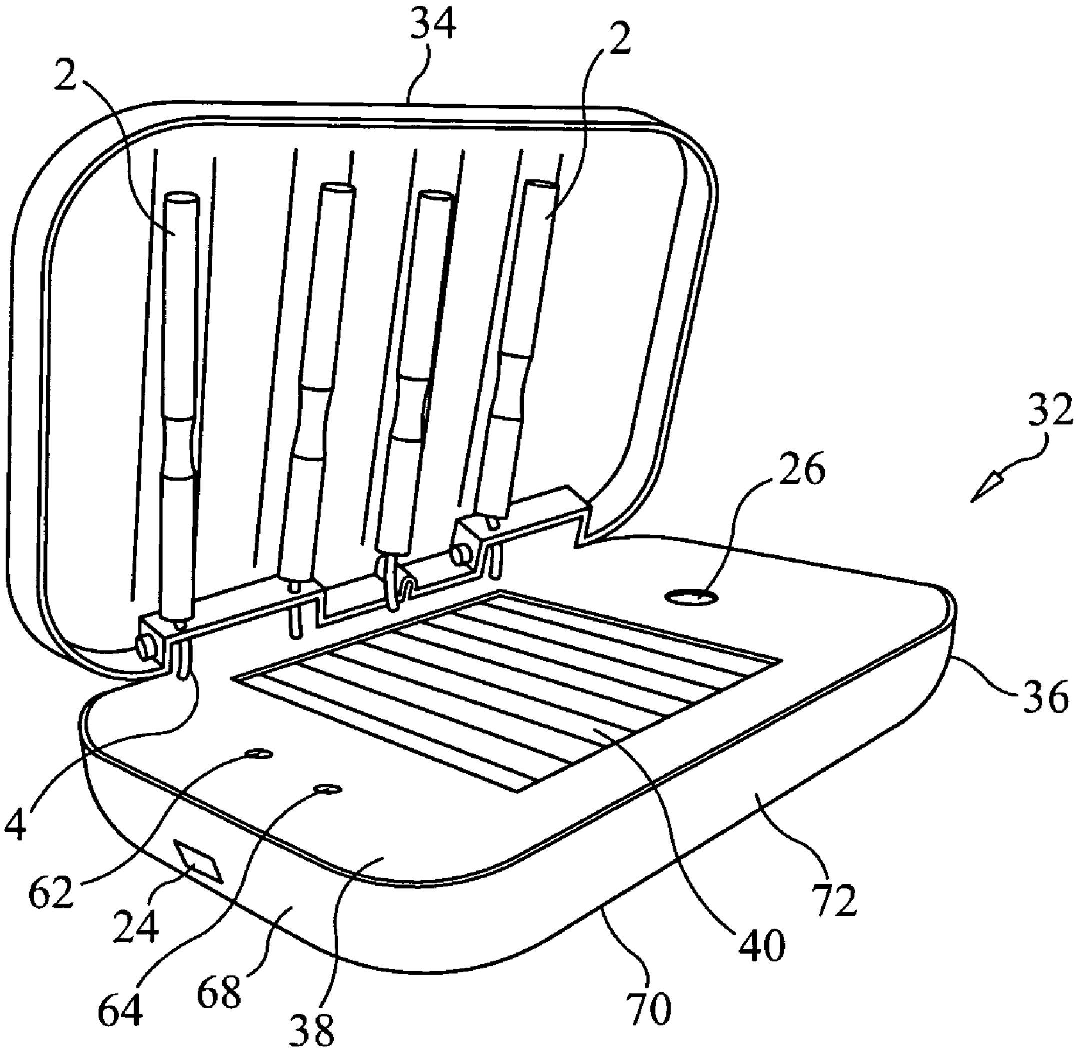
34
2
2
32
26
36
4
62
24
64
68
38
70
40
72

FIG. 6

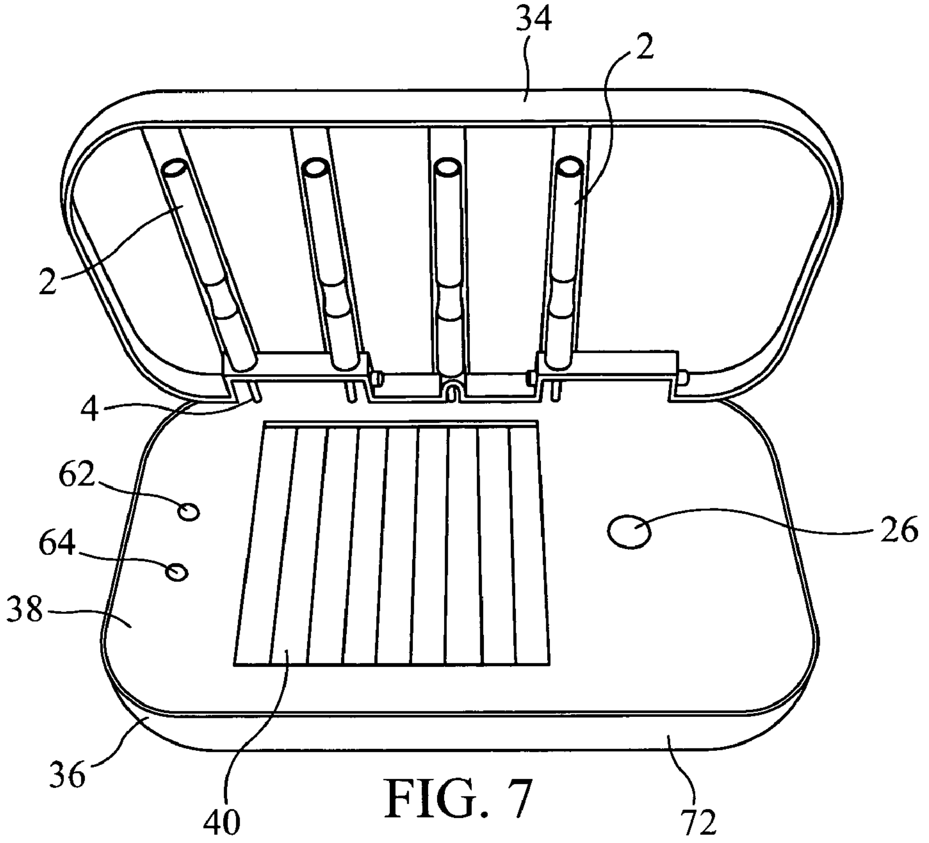
34
2
32
2
4
62
26
64
38
36
40
72

FIG. 7

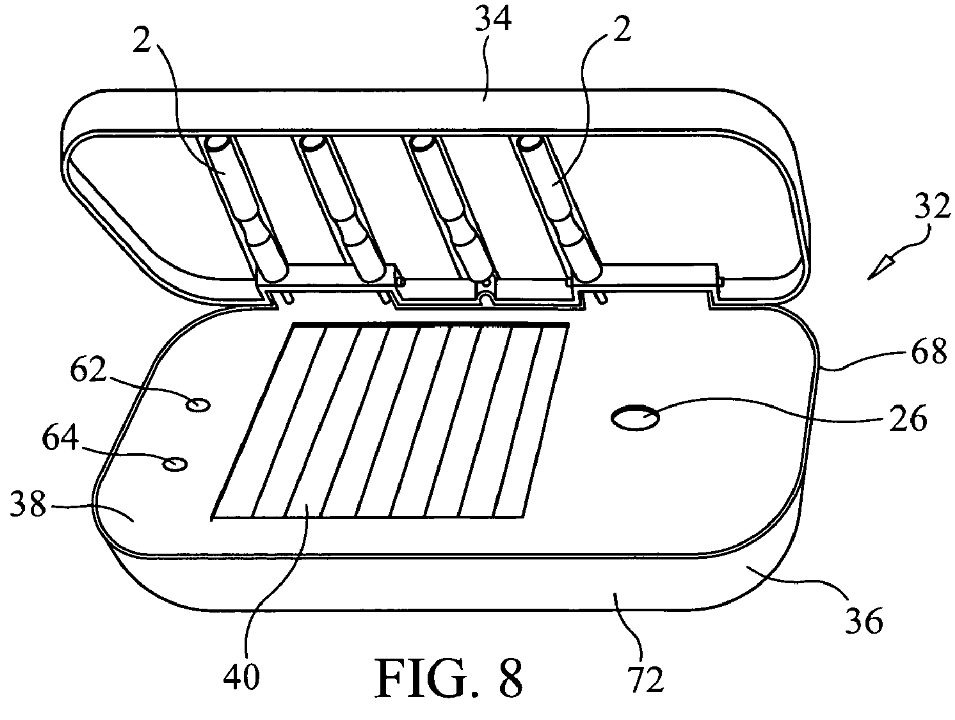
2
34
2
32
68
62
26
64
38
36
40
72

FIG. 8

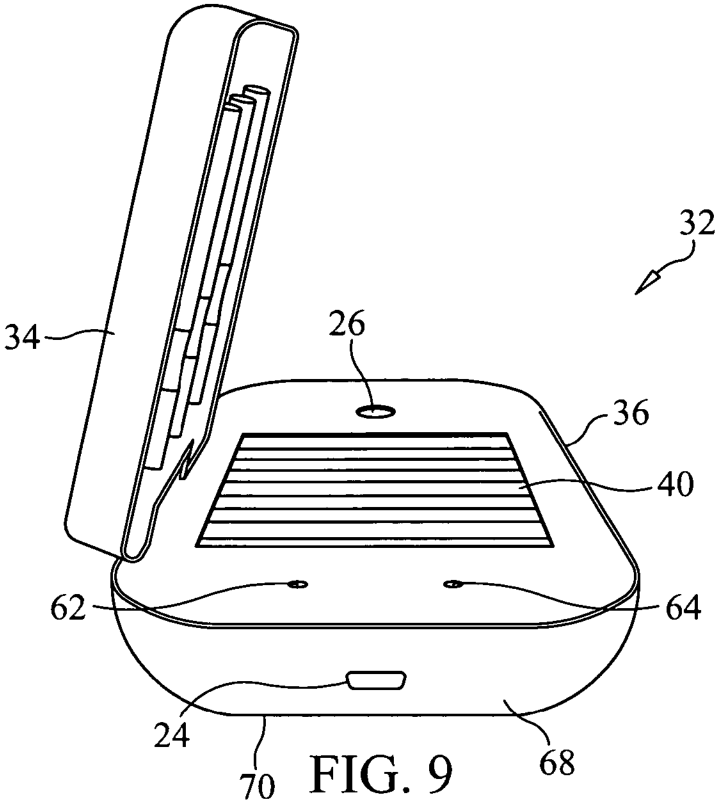
32
34
26
36
40
62
64
24
70
68

FIG. 9

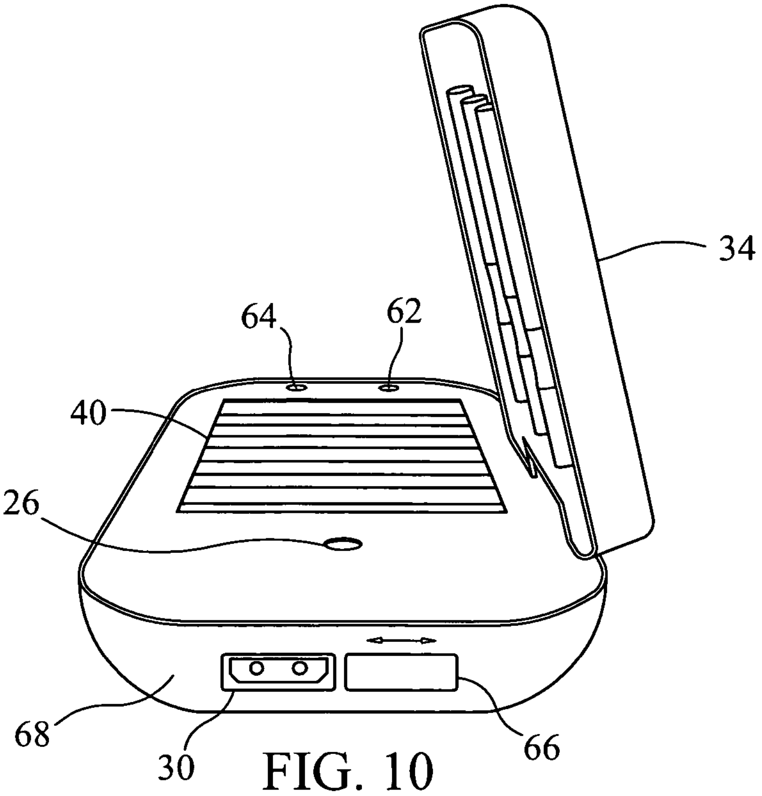
34
64
62
40
26
68
30
66

FIG. 10

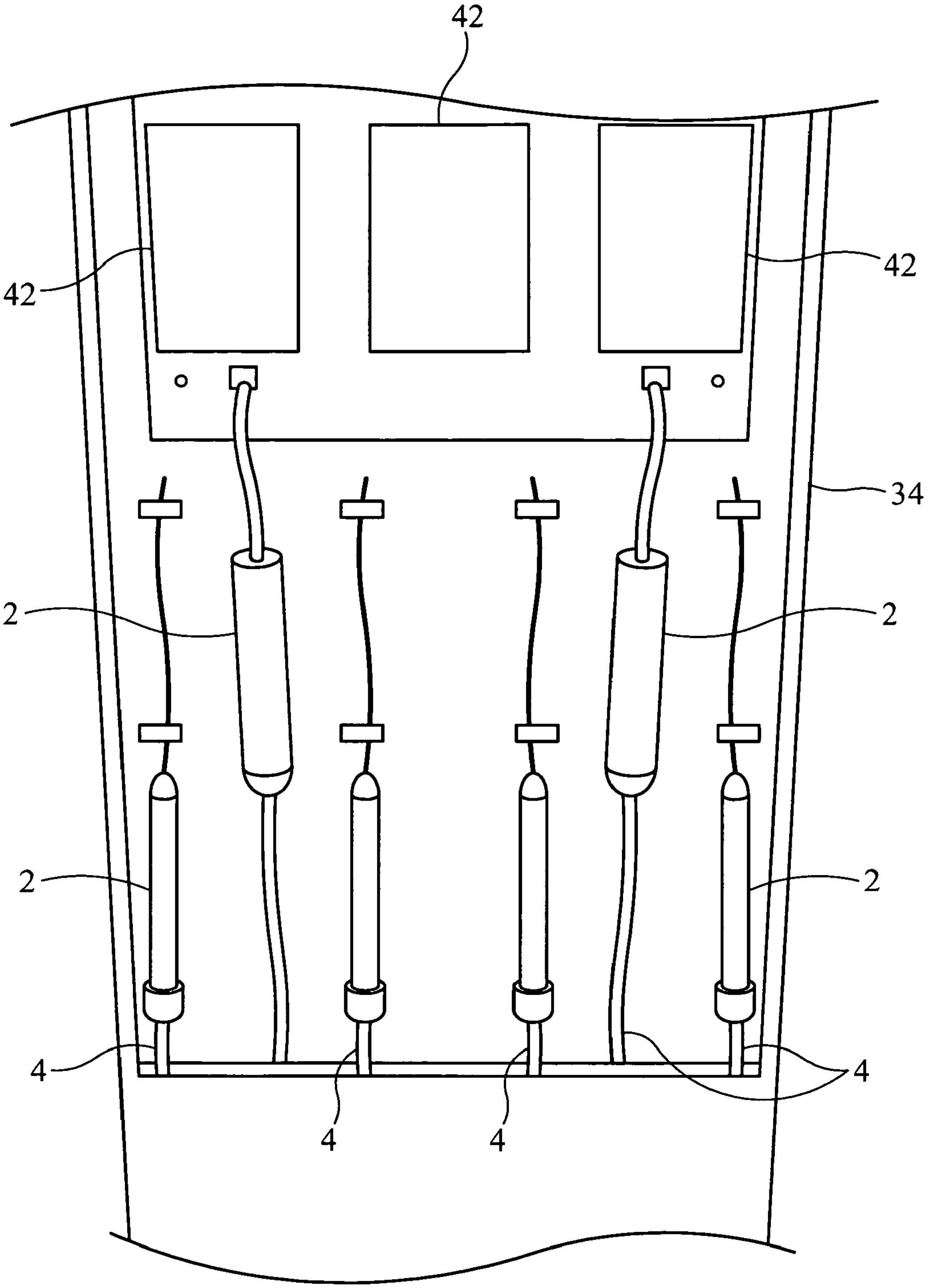
42
42
42
34
2
2
2
2
4
4
4
4

FIG. 11

METHOD AND APPARATUS FOR HARVESTING ENERGY

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is related to U.S. Provisional Application Ser. No. 61/335,459, filed on Jan. 7, 2010, and entitled “Method and Apparatus for Harvesting Energy from WIFI Signals”, U.S. Provisional Application Ser. No. 61/402,481, filed on Aug. 31, 2010, and entitled “Method and Apparatus for Harvesting Energy”, and U.S. Provisional Application Ser. No. 61/404,290, filed on Sep. 30, 2010, and entitled “Method and Apparatus for Harvesting Energy”, the disclosures of which are incorporated herein by reference and on which by priority is hereby claimed.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to energy conservation methods and devices, and more particularly relates to devices which can harvest energy.

2. Description of the Prior Art

Many devices have been proposed or implemented which harvest energy from different types of naturally-occurring energy sources. These energy sources include solar, wind, tide, current and wave power, the energy of which is converted by such devices into electrical power to supplement or replace electrical power provided by a power company. Except for solar cells powering hand held calculators and other small electronic devices, most energy harvesting apparatus are relatively large, expensive and inefficient. To the knowledge of the inventor, there are no RF (radio frequency) commercially available devices which harvest energy from WIFI (wireless fidelity), WiMax, Satellite Radio, Cellular Phone (850 MHz to 900 MHz and 1.8 GHz-1.9 GHz), or any 2.4 GHz and UHF TV signal sources to power or recharge the batteries of small electronic devices, such as cellular phones, personal digital assistants (PDAs), MP3 players or the like.

OBJECTS AND SUMMARY OF THE INVENTION

It is an object of the present invention to provide a device which harvests energy from WIFI, WiMax, Satellite Radio, Cellular Phone (850 MHz to 900 MHz and 1.8 GHz-1.9 GHz), and any 2.4 GHz and UHF TV signals or any signal from outdoor antenna or CATV feed through any “F” type connectors at consumers’ homes.

It is another object of the present invention to provide a method and device for harvesting energy from two or more alternative sources of energy such as solar and vibration from a microphone for powering or recharging the batteries of an electronic device.

It is yet another object of the present invention to provide an energy harvesting device which stores energy for later use to power an electronic device connected thereto.

It is a further object of the present invention to provide a device for harvesting energy from both WIFI or other RF signals and solar power to power an external electronic device connected thereto.

It is yet a further object of the present invention to provide apparatus for harvesting energy from WIFI or other RF signals which converts WIFI or RF signals received thereby into a DC (direct current) voltage, the energy from which is stored in a storage device within the apparatus, and which further has the capability of being connectable to a source of electrical power for storing energy therefrom in the storage device for later consumption by a user to power an external electronic device connected to the energy harvesting apparatus.

In accordance with one form of the present invention, a device for harvesting energy from WIFI or other RF signals includes one or more broadband antennas which are respectively connected to one or more voltage multiplier/rectifier circuits. Each voltage multiplier/rectifier circuit converts the WIFI or RF signals received by the respective antenna coupled thereto into a DC voltage, having an amplitude multiplied a number of times, such as by a factor of 4, 6 or 8. The DC voltage output signals from the voltage multiplier/rectifier circuits are added in series and in the same polarity (i.e., the voltages are “stacked”) or in parallel to provide a combined, detected DC voltage signal, which is provided to a temporary storage device, such as a capacitor.

When the voltage on the capacitor reaches a predetermined level, a watchdog monitor circuit effectively connects the temporary storage capacitor to the input of a power management circuit. The power management circuit includes a charge pump circuit which increases the voltage on the input of the power management circuit to a higher, more useable voltage on its output for charging a storage device, such as a lithium battery (for example, lithium polymer, lithium-ion or any lithium chemical type). Accordingly, the output voltage of the power management circuit is at a predetermined voltage level to charge the lithium battery or other storage device.

In a preferred form of the present invention, the device further includes a solar cell which converts solar energy from light impinging thereon to an electrical voltage, which is also filtered and provided to a second power management circuit. The output of this second power management circuit also provides a voltage at a predetermined level to the lithium battery or other storage device in response to the voltage from the solar cell resulting from the light impinging thereon.

The output of the lithium battery or storage device is connected to the input of a DC/DC boost converter circuit, which provides preferably a 5 volt DC voltage on a USB output connector to which is connectable the external electronic device to be powered thereby or whose battery is to be charged.

In a preferred form of the invention, the energy harvesting device of the present invention may include an input connector for receiving a charging voltage from an external power source, such as another electronic device, or a power transformer for such device which is connected to a cigarette lighter of an automobile, or to an AC/DC adapter for connection to a 110 volt AC wall power receptacle, for example. The power signal from this external electrical power source is provided to the lithium battery or other storage device internal to the energy harvesting circuit so that the lithium battery or storage device may be charged thereby.

These and other objects, features and advantages of the present invention will be apparent from the following detailed description of illustrative embodiments thereof, which is to be read in connection with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a block diagram of a circuit for harvesting energy transmitted from an RF (radio frequency) source to charge and/or power an electronic device, such as a cellular telephone or a PDA or a Blackberry™ device, formed in accordance with the present invention.

FIG. **2** is a schematic diagram of a first preferred form of an electrical circuit constructed in accordance with the present invention for harvesting energy from WIFI signals.

FIG. **3** (i.e., FIGS. **3**A and **3**B) is a block diagram of a second preferred form of the electrical circuit used in the energy harvesting device of the present invention.

FIG. **4** (i.e., FIGS. **4**A and **4**B) is a block diagram of third preferred form of the electrical circuit used in the energy harvesting device of the present invention.

FIG. **5** is a top perspective view of one form of a housing of the energy harvesting device of the present invention, illustrating the cover thereof in an open position.

FIG. **6** is a top perspective view of the housing of the energy harvesting device of the present invention shown in FIG. **5** taken from a different angle and illustrating the cover thereof in an open position.

FIG. **7** is a front perspective view of the housing of the energy harvesting device of the present invention shown in FIGS. **5** and **6**, illustrating the cover thereof in an open position.

FIG. **8** is a front perspective view of the housing of the energy harvesting device of the present invention shown in FIGS. **5-7** taken from a different angle and illustrating the cover thereof in an open position.

FIG. **9** is a first side perspective view of the housing of the energy harvesting device of the present invention shown in FIGS. **5-8** and illustrating the cover thereof in an open position.

FIG. **10** is a second side perspective view of the housing of the energy harvesting device of the present invention shown in FIGS. **5-9** and illustrating the cover thereof in an open position.

FIG. **11** is a perspective view of a portion of the cover of another form of a housing of the energy harvesting device of the present invention, illustrating the cover thereof in an open position and an alternative arrangement of antennas thereon.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is designed to receive radio frequency signals from any WIFI 2.4 to 2.5 GHz source, WiMax, Satellite Radio, Cellular Phone (850 MHz to 900 MHz and 1.8 GHz-1.9 GHz), and UHF TV signals or CATV signals and convert the received radio frequency signal into another form of energy, such as a DC voltage. Even more preferably, the energy harvesting device of the present invention is capable of receiving RF signals from about 500 MHz to about 2.5 GHz, and convert the received signals into electrical energy. The same circuit configuration can also be designed and extended to convert RF signals in VHF, FM and UHF frequencies (such as signals from CATV) into DC electrical energy. This energy is utilized to recharge a battery inside any cell phone through a USB connector. The invention is designed to use WIFI signals, in particular, for energy harvesting because many WIFI signal transmitters are provided and are available in public places, such as airport terminals, coffee shops, fast food restaurants, etc., where an AC power outlet may not be easily accessible for a user to charge or power an electronic device by connecting it to the outlet.

There are many types of energy harvesting devices or systems that are known, such as solar cell, windmill, water pressure to provide movement in power generators, FM radio signals to charge small toys or flying devices (for example, spy monitor devices) inductive chargers, and the like. However, no one heretofore has taken advantage of the many WIFI signals, cell phone signals, Satellite Radio or UHF TV signals that are mostly available in public places, and the energy from those transmitted signals is wasted and not utilized. Such WIFI signals, Satellite Radio, UHF TV signals or cell phone signals are broadcast 24 hours a day, every day.

Most people think that the energy from WIFI or other high frequency radio signals is very minute and cannot be used, but the inventor has found otherwise. Generally, the present invention is directed to harvesting the WIFI, UHF TV, Satellite Radio or cell phone signals and converting the electromagnetic energy in the signals to a DC voltage, and providing a "top charge" to charge the batteries of an external electronic device connected thereto, or to continuously power an internal battery within the energy harvesting device until its storage capacity is full, and that battery energy may be reused when needed to power an electronic device, such as a cellular phone, PDA, MP3 player, camera or other electronic device connected thereto.

The present invention utilizes WIFI or cell phone signals that are available and easy to find in any public place or in a consumer's home almost anywhere in the world. There are many high powered WIFI broadcasting stations whose transmitted signals create a WIFI hotspot, and also WIMAX signals to be implemented at some future time are projected to be rather strong signals that have a 32 mile radial range, and the energy from these signals may also be harvested.

Provided above is a general description of the purpose and implementation of the energy harvesting method and apparatus of the present invention. Now, a detailed description of preferred forms of the present invention is provided.

A block diagram of a circuit for harvesting energy transmitted from an RF (radio frequency) source is shown in FIG. **1** of the drawings. An antenna **2**, such as patch antenna, receives an RF signal, such as a WIFI signal transmitted in public places, such as airport terminals, coffee shops, fast food restaurants and the like, and converts the RF signal to an electrical signal. This electrical signal is provided via a transmission line or signal cable **4** to a circuit **6** which rectifies and harvests the energy of the transmitted RF signal. This circuit **6** converts the RF electrical signal from the antenna **2** to a DC signal of sufficient magnitude to charge and/or power an electronic device **8**, such as a cellular telephone, a PDA or a Blackberry™ device connected to the energy harvesting and rectifier circuit **6** by a USB cable **10** or the like.

FIG. **2** is a schematic diagram of a first preferred form of an electrical circuit constructed in accordance with present invention for harvesting energy from WIFI or other RF signals. The energy harvesting circuit, in accordance with this first embodiment, includes an antenna **2** which receives RF signals, such as WIFI signals, and converts the RF signals to an electrical signal which is provided by a transmission line or cable **4** to the input of an RF-to-DC rectifier circuit **12**. The rectifier circuit **12** converts the RF signal to a DC (direct current) signal on its output. This DC signal is provided to a first storage device **14**, which is preferably a 2F (farad), 2.5 volt supercapacitor. The first storage device **14** stores and accumulates charge thereon from the rectified (DC) signal outputted by the rectifier circuit **12**.

The DC signal outputted by the rectifier circuit **12** is also provided to the input (Pin **3**) of a flash-based CMOS microcontroller **16** preferably having Part No. 12F683-0847 manufactured by Microchip Technology Inc. of Chandler, Ariz. The microcontroller **16** is used as a charge circuit control to prevent a lithium battery used as a second storage device **18** in the energy harvesting circuit from overcharging.

The voltage on the supercapacitor **14** is also provided to the input of a first stage DC-to-DC converter circuit **20**, which acts as a charge pump, or stated another way, a step up voltage regulator and battery management circuit. Preferably, the first stage DC-to-DC converter circuit **20** is Part No. MAX1672 manufactured by Maxim Integrated Products of Sunnyvale, Calif.

When the voltage on the supercapacitor **14** reaches a certain level, such as about 1.5 volts, which is provided to the input of the microcontroller **16**, the microcontroller **16** generates a pulsed signal on its output (Pin **5**), which is provided to the enable input of the first stage DC-to-DC converter circuit **20** to cause the converter circuit **20**, acting as a charge pump, to increase the voltage on its input from about 1.5 volts to about 4.5 volts on its output. If the voltage on the supercapacitor **14** is below a predetermined level, such as 1.5 volts, the microcontroller **16** does not provide a pulsed signal to the enable input of the first stage DC-to-DC converter circuit **20** to effectively cause the converter circuit **20** from operating, so as not to deplete charge on the supercapacitor **14** if the supercapacitor is insufficiently charged.

The output of the first stage DC-to-DC converter circuit **20** is provided to the input of a second stage DC-to-DC converter circuit **22**, preferably being the same integrated circuit as the first stage DC-to-DC converter circuit **20** and having Part No. MAX1672. Again, the preferred voltage on the output of the first stage DC-to-DC converter circuit **20** is about 4.5 volts.

As can be seen from FIG. **2**, the energy harvesting circuit also includes a USB charger port or connector **24** for providing a connection to an external source of voltage (preferably, about 5 volts DC), to charge the internal second storage device **18**, which as will be explained is preferably a lithium polymer battery, using a lap top computer or other device connectable to the USB charger connector **24**. The USB charger connector **24** has its +5 volt input pin connected to the connection between the output of the first stage DC-to-DC converter circuit **20** and the input of the second stage DC-to-DC converter circuit **22**.

The second stage DC-to-DC converter circuit **22** acts as a low drop out, step down voltage converter/regulator to convert the 4.5 volts from the first stage DC-to-DC converter circuit **20** (or the 5 volt DC signal provided on the USB charger connector **24** from an external source) to a 3.7 volt DC signal on the output of the second DC-to-DC converter circuit **22**. This output voltage of the second stage DC-to-DC converter circuit **22** is provided to the positive side of a lithium polymer battery **18** forming part of the energy harvesting circuit, the negative side of the battery being preferably grounded. Preferably, the lithium polymer battery is a 3.7 volt, 1.5 amp battery which is used as a second storage device **18** which is charged by the rectified and DC-converted voltage generated by the energy harvesting circuit from the RF signal received by the antenna **2**, or by an external DC power source connected to the USB charger connector **24**.

The voltage on the lithium polymer battery **18** is monitored so as not to exceed approximately 3.7 volts. The positive side of the lithium polymer battery **18** is connected to a feedback input (Pin **1**) on the microcontroller **16** in order to limit the voltage generated by the energy harvesting circuit provided to the lithium polymer battery **18** and to an external electronic device **8** connected to the energy harvesting circuit. When the voltage on the feedback input of the microcontroller **16** reaches a predetermined threshold voltage, such as 3.7 volts, the microcontroller **16** ceases providing a pulsed signal on its output (Pin **5**) connected to the enable input on the first stage DC-to-DC converter circuit **20** so as to prevent the first stage DC-to-DC converter circuit **20** from operating as a charge pump beyond a predetermined voltage conversion limit. When the voltage on the lithium polymer battery **18** falls to below the predetermined threshold voltage, which is provided to the feedback input of the microcontroller **16**, the microcontroller **16** again provides a pulsed signal on its output to the enable input of the first stage DC-to-DC converter circuit **20** so that the converter circuit **20** resumes operation as a charge pump to increase the input voltage provided thereon to approximately 4.5 volts. However, no pulsed signal will be provided to the enable input of the first stage DC-to-DC converter circuit **20** from the microcontroller **16** if the voltage on the supercapacitor **14** is below a predetermined threshold voltage, such as 1.5 volts.

The positive side of the lithium polymer battery **18** is connected through a user controlled single pole, single throw switch **26** to the input of a third stage DC-to-DC converter circuit **28**, also acting as a charge pump. The third stage DC-to-DC converter circuit **28** is preferably realized using an integrated circuit having Part No. TPS61200DRCT, manufactured by Texas Instruments of Dallas, Tex. The third stage DC-to-DC converter circuit **28** increases the voltage provided by the lithium polymer battery **18** (3.7 volts) to the input of the third stage DC-to-DC converter circuit **28** to approximately 5 volts DC on the output thereof. The output signal of the third stage DC-to-DC converter circuit **28** is provided to a particular pin (+5 volt DC pin) on a USB output connector **30**, having another certain pin grounded. An external electronic or electrical device **8**, such as a cellular phone, PDA, MP3 player, camera or the like, may be connected to the USB output connector **30** of the energy harvesting circuit using a standard USB cable **10** so that the energy harvesting circuit of the present invention may charge the batteries of the external electronic device **8** or power the external electronic device **8** connected to the energy harvesting circuit by the user.

Referring now to FIG. **3** of the drawings, a block diagram of a second preferred form of the electrical circuit of the energy harvesting device of the present invention is shown. Preferably, a plurality of broadband antennas **2**, each preferably the same, is used. In one form of the present invention, four broadband antennas **2***a*, **2***b*, **2***c*, **2***d* are used, and each broadband antenna **2***a*-**2***d* is particularly situated within the housing **32** for the energy harvesting circuit so that each antenna **2***a*-**2***d* will be capable of receiving WIFI or other RF signals which are transmitted no matter what disposition the housing **32** is in. Preferably, each broadband antenna **2***a*-**2***d* is capable of receiving signals from about 500 MHz to about 2.5 GHz.

Even more preferably, and as shown in FIGS. **5**-**10** of the drawings, there are four monopole broadband antennas **2***a*-**2***d* which are mounted on the inside surface of a transparent or translucent cover **34** of the housing **32**. The cover **34** is pivotally joined at one edge thereof to a corresponding lateral or top edge of the main body or base **36** of the housing **32**, so that it may be raised and lowered respectively away from and in proximity to the top surface of the main body or base **36**, thus covering and uncovering a solar cell **40** which is exposed through an opening formed in the top surface **38** of the main body or base **36**. Although the four monopole antennas **2***a*-**2***d* will receive RF energy when the cover **34** is disposed in its lower position, covering the solar cell **40** of the housing **32**, the four monopole antennas will more effectively receive RF energy transmitted from an RF source or WIFI source when the cover **34** is disposed in a raised position, such as shown in FIG. **5** of the drawings. The features of the housing **32** of the energy harvesting device of the present invention will be described in greater detail.

Each antenna **2***a*-**2***d* of the plurality of antennas is coupled to a respective voltage multiplier/full wave rectifier circuit **42**, as shown in FIG. **3**. More specifically, each voltage multiplier/rectifier circuit **42** includes a voltage multiplier so that the WIFI or RF signals received by the respective antenna **2***a*-**2***d* coupled thereto is converted into a DC voltage having an amplitude multiplied a number of times, such as by a factor of 4, 6 or 8, for example.

More specifically, the first antenna **2***a* is coupled to the first multiplier/rectifier circuit **42***a*, which converts the WIFI or RF signals the first antenna **2***a* receives into a DC voltage multiplied by a factor of 4, 6, or 8, for example. The DC output voltage of the first multiplier/rectifier circuit **42***a*, as a result of the detected signal of the first antenna **2***a* coupled thereto, is provided to node A and a first leg of a capacitor C**39** (see FIG. **3**).

The second multiplier/rectifier circuit **42***b* is very similar to the first multiplier/rectifier circuit **42***a*. The second multiplier/rectifier circuit **42***b* converts to a DC voltage the WIFI or RF signals received by the second antenna **2***b* coupled to it, multiplied by a desired factor (e.g., 4, 6 or 8), and this voltage is provided to node B and a first leg of a capacitor C**34**, whose second leg is connected to node A and the first leg of capacitor C**39**. Thus, the output voltage of the second multiplier/rectifier circuit **42***b* is effectively stacked on top of the voltage outputted by the first multiplier/rectifier circuit **42***a*.

The third multiplier/rectifier circuit **42***c* is similar to the second multiplier/rectifier circuit **42***b*, and it is coupled to the third antenna **2***c* in the same manner that the first and second multiplier/rectifier circuits **42***a*, **42***b* are respectively coupled to the first and second antennas **2***a*, **2***b*. Again, the DC voltage outputted by the third multiplier/rectifier circuit **42***c*, which is provided to node C and the first leg of a capacitor C**21**, whose second leg is connected to node B and the first leg of capacitor C**34**, is stacked on top of the DC voltages provided by the first and second multiplier/rectifier circuits **42***a*, **42***b*.

The fourth multiplier/rectifier circuit **42***d* is similar to the third and second multiplier/rectifier circuits **42***c*, **42***b* described previously, and is coupled to the fourth antenna **2***d*. The DC voltage outputted by the fourth multiplier/rectifier circuit **42***d* is provided to node D and the first leg of a capacitor C**12**, whose second leg is connected to node C and the first leg of capacitor C**21**. As a result of RF signals received by the fourth antenna **2***d* coupled to the fourth multiplier/rectifier circuit **42***d*, the DC output voltage of the fourth multiplier/rectifier circuit **42***d* is stacked on top of the DC voltages outputted by the third, second and first multiplier/rectifier circuits **42***c*, **42***b*, **42***a*. Thus, the potential at the first leg of capacitor C**12**, measured with respect to the second leg of capacitor C**39**, could be as much as, or possibly more than, 25 volts DC as a result of the stacked arrangement of output voltages from the first, second, third and fourth multiplier/rectifier circuits **42***a*-**42***d*. As will be explained in greater detail, the second leg of capacitor C**39** is coupled to a "jump start" battery BT**2** (also referred to as a jump start voltage device **44**, and the stacked output voltages from the four multiplier/rectifier circuits **42***a*-**42***d* are raised by the DC voltage provided by battery BT**2**.

The combined stacked voltage generated by the interconnected multiplier/rectifier circuits **42***a*-**42***d* is provided to a capacitor C**3**. Capacitor C**3** acts as a temporary storage device **14** for the voltages detected and rectified by the four multiplier/rectifier circuits **42***a*-**42***d* as a result of RF energy received by the four receiving antennas **2***a*-**2***d*. More specifically, the first leg of capacitor C**3** is connected to the first leg of capacitor C**12** and the second leg of capacitor C**3** is grounded. Preferably, capacitor C**3** is a low leakage tantalum capacitor, although other storage devices and types of capacitors may be used.

As mentioned previously, capacitor C**3** is provided for the temporary storage of the stacked DC voltage outputted by the four multiplier/rectifier circuits **42***a*-**42***d*, as a result of RF or WIFI energy received by the four antennas **2***a*-**2***d* connected thereto, as well as the DC voltage of the jump start battery BT**2** which raises the stacked voltage at node D.

The circuit of the present invention also preferably harvests light energy by using a solar cell **40** and circuit. In this regard, reference should continue to be had to FIG. **3** of the drawings. A solar cell **40** is provided in the energy harvesting device of the present invention and is exposed through an opening in the top surface **38** of the main body or base **36** of the housing **32** of the energy harvesting device. The solar cell **40** can generate between about 3.1 volts with indoor lighting and about 5.5 volts in direct sunlight. The negative terminal of the solar cell **40** is preferably grounded, and the positive terminal of the solar cell **40** is connected to the cathode of a zener diode D**2**, whose anode is connected to ground. Preferably, zener diode D**2** is a 5.6 volt zener diode to limit the output voltage provided by the solar cell **40** so as not damage any circuitry connected thereto.

The output of the solar cell **40** is connected to the VDD input of a power management integrated circuit **46**, such as Part No. CBC3112 manufactured by Cymbet Corporation of Elk River, Minn., or any similar power management circuit made by a number of manufacturers. The output voltage of the power management circuit **46**, which is regulated and selected to be about 4.2 volts, is provided through an isolation diode D**3** to the positive side of the main energy storage battery BT**1** (i.e., second or main storage device **18**) to help maintain the charge on battery BT**1** when no or negligible WIFI or RF signals are present or can be detected and when ambient light conditions permit the charging of battery BT**1** by the solar cell **40**.

To help "jump start" or boost the charge on the temporary storage capacitor C**3**, a 3.7 volt, 60 milliampere-hour or greater, lithium polymer battery BT**2** is used (see FIG. **3**). The negative side of battery BT**2** is grounded, while the positive side is coupled to the second leg of capacitor C**39** in order to raise the stacked voltage generated by the four interconnected multiplier/rectifier circuits **42***a*-**42***d* by 3.7 volts, i.e., the voltage of battery BT**2**. Accordingly, the voltage across capacitor C**3** is the voltage resulting from the combination of the voltage from battery BT**2** and the RF or WIFI energy harvested by the circuit of the present invention. It should be noted that the 25 volt DC signal resulting from the stacked arrangement of the four multiplier/rectifier circuits **42***a*-**42***d* is an unloaded voltage, and when used under load to charge capacitor C**3**, the voltage drops significantly.

The voltage from battery BT**2** is used to boost the voltage temporarily stored on **42***d* capacitor C**3** and to supplement the voltage outputted by the four multiplier/rectifier circuits **42***a*-**42***d* resulting from the WIFI and RF energy which is harvested. Furthermore, the "boost" voltage from battery BT**2** helps insure that a second power management circuit **48** (see FIG. **3**) has a sufficient voltage on its VDD input in order to operate.

It is estimated that a 60 milliampere-hour battery used as battery BT**2** will last for approximately 1500 hours, or about 62 days, when the energy harvesting device of the present invention is stored in a dark room, thus with no voltage being generated by the solar cell **40**, as the circuit draws only about 25 microamperes to about 40 microamperes of current from battery BT**2**.

To maintain the charge on battery BT**2**, the output voltage generated by the power management circuit **46** is provided to the anode of an isolation diode D**50**, whose cathode is connected to the positive side of battery BT**2**. In this way, the power management circuit **46** acts as a buffer between the solar cell and battery BT**2** to prevent battery BT**2** from overcharging, which may result in the battery overheating.

Alternatively, to charge battery BT**2**, the positive lead of the solar cell **40** may be coupled to the positive side of battery BT**2** through a diode (not shown), the anode of the diode being connected to the positive lead of the solar cell **40**, and the cathode of the diode being connected to the positive side of battery BT**2**. Optionally, a PTC (positive temperature coefficient) thermistor (not shown) may be connected to battery BT**2** between the solar cell **40** and the battery BT**2** to provide some protection to the circuit in the remote chance that battery BT**2**, which is preferably a lithium polymer battery, overheats due to overcharging or short circuiting.

Referring again to FIG. **3** of the drawings, it will be seen that a second power management circuit **48** is provided, and is indirectly coupled to the temporary storage capacitor C**3**. Preferably, a watchdog monitor circuit **50** is disposed between capacitor C**3** and the VDD input of the power management circuit **48**, as will be explained in greater detail. Power management circuit **48** may be Part No. CBC3112 manufactured by Cymbet Corporation of Elk River, Minn., or any similar power management circuit made by a number of manufacturers. The VBAT output of the power management circuit **48** is connected to the anode of diode D**1**, as will be explained in greater detail.

Since the power management circuit **48**, like the power management circuit **46**, requires about 2.6 volts on its VDD input in order to operate, the energy harvesting circuit of the present invention preferably includes a watchdog monitor circuit **50** situated between the temporary storage capacitor C**3** and the power management circuit **48** in order to minimize the unnecessary discharge of energy stored in capacitor C**3** when there is insufficient voltage on capacitor C**3** for the power management circuit **48** to operate. The watchdog monitor circuit **50** may include a voltage detector circuit **52** and a transistor switch circuit **54** connected thereto and controlled thereby. The voltage detector circuit **52** may be a constant voltage power monitor with hysteresis and having Part No. S-80825 manufactured by Seiko Instruments, Inc. of Chiba, Japan.

The watchdog monitor circuit **50** comprising the voltage detector circuit **52** and the transistor switch circuit **54** is used to minimize any leakage from the temporary storage capacitor C**3** by preventing a direct connection between capacitor C**3** and the VDD input of the second power management circuit **48** when the voltage on the temporary storage capacitor C**3** is below a predetermined voltage. More specifically, when the voltage is below about 2.6 volts on capacitor C**3**, the voltage detector circuit **52** of the watchdog monitor circuit **50** provides an output signal to the transistor switch circuit **54** to open the transistor switch circuit **54** and prevent the voltage on capacitor C**3** from being connected therethrough to the VDD input of the second power management circuit **48** so that circuit **48** does not drain the energy from the WIFI or RF energy harvesting circuit stored on capacitor C**3**. However, if the voltage on capacitor C**3** rises above about 2.6 volts, the voltage detector circuit **52** of the watchdog monitor circuit **50** outputs a signal to close the transistor switch circuit **54** so that the voltage on capacitor C**3** is provided to the VDD input of the second power management circuit **48**. Typically, the voltage on the temporary storage capacitor C**3** will range between about 2.5 volts and about 2.65 volts, and only when this voltage is equal to or above about 2.6 volts will capacitor C**3** be effectively connected to the input of the second power management circuit **48**. A general form of the watchdog monitor circuit **50** described above may be found in the application notes accompanying the data sheet of the S-808xxC series of voltage detectors disseminated by Seiko Instruments Inc. (see FIGS. **14** and **20** thereof).

The power management circuit **48** has a charge pump which converts the 2.6-2.65 voltage (or, more specifically, any voltage above 2.63 volts) on its input to between about 4.1 volts DC and about 4.2 volts DC on its VBAT output. The VBAT output of the second power management circuit **48** is connected to the anode of a low voltage drop (about 0.1 volts) isolation diode D**1**. The cathode of diode D**1** is connected to the positive side of a 3.7 volt, 1000 milliampere-hour, lithium polymer battery BT**1**, which acts as the main energy storage battery or device **18**.

Preferably, a battery protection circuit **56** is connected to the battery BT**1** to provide overcharge protection and over-discharge protection. The battery protection circuit **56** is preferably Part No. DW01 manufactured by Fortune Semiconductor Corporation of Taipei, Taiwan, which monitors the voltage on the positive side of battery BT**1**, and drives a transistor switch circuit **58** situated between the negative side of the battery and ground to selectively connect and disconnect the negative side of battery BT**1** to and from ground.

When the voltage on the lithium battery BT**1** exceeds a certain predetermined overcharge protection voltage, charging of battery BT**1** is inhibited by the battery protection circuit **56** opening the transistor switch circuit **58** to disconnect the battery BT**1** from ground. On the other hand, when the voltage of the lithium battery BT**1** falls below a predetermined over-discharge protection voltage, discharging of battery BT**1** is inhibited by the battery protection circuit **56** opening the transistor switch circuit **58** to disconnect the battery BT**1** from ground. Essentially, the negative side of battery BT**1** is disconnected from ground when either an overcharge condition or an over-discharge condition exists. When neither condition exists, the battery protection circuit **56** closes the transistor switch circuit **58** to connect the negative side of the battery BT**1** to ground.

The energy harvesting circuit of the present invention also preferably includes a battery charge monitoring circuit **60** to provide an indication to the user of the energy harvesting device whether the lithium battery BT**1** is sufficiently charged to power an external electronic device **8** connected to the energy harvesting circuit or to charge the battery or batteries of the external device. Preferably, the battery charge monitoring circuit **60** includes a voltage divider network (having interconnected resistors) (not shown) operatively coupled to the positive side of the battery BT**1** and used to bias a transistor (not shown) on or off, depending on the voltage of the battery. A single pole, single throw (or equivalent) momentary push button switch SW**2** (or switch circuit) is connected between the positive side of the lithium battery BT**1** and the battery charge monitoring circuit **60**, and is preferably mounted on the top surface **38** of the housing **32** so as to be accessible to the user of the energy harvesting device. Accordingly, so as not to drain the charge from battery BT**1**, the battery charge monitoring circuit **60** will only work when the user presses on momentary push button switch SW**2** to provide power to the battery charge monitoring circuit **60**.

A dual color (red/green) light emitting diode (LED) or separate red and green LEDs **62**, **64** are operatively coupled to and driven by the battery charge monitoring circuit **60**, and are preferably mounted on the top surface **38** of the housing **32**, or elsewhere on the housing, to be viewable by the user of the energy harvesting device.

When the voltage on battery BT**1** is 3.3 volts or greater, the battery charge monitoring circuit **60** will cause the green LED **64** to light when the user presses on momentary push button switch SW**2**. If the voltage on battery BT**1** is 3.0 volts or lower, the battery charge monitoring circuit **60** will cause the red LED **62** to illuminate when the user presses on momentary push button switch SW**2**. The 3.3 volt threshold and the 3.0 volt threshold may be adjusted by changing the values of resistors from which the voltage divider network is formed to turn the transistor on or off at different voltage levels.

As shown in FIG. **3**, a single pole, single throw switch or switch circuit SW**1** has one side connected to the positive side of the lithium polymer battery BT**1** or other storage device. The switch or switch circuit SW**1** is provided to ensure that stored energy is not drawn from the battery until desired by the user. As will be explained in greater detail, the switch SW**1** is coupled to a spring biased, sliding cover **66** on the housing **32** for the energy harvesting circuit (see FIG. **10**) that selectively covers an output USB type A connector **30**, and will be automatically activated to provide electrical conduction therethrough when the user slides the USB connector cover **66** in one direction on the main body or base **36** of the housing **32**.

The other side of the switch or switch circuit SW**1** is connected to the input of a DC/DC boost converter circuit **28**, which may be an integrated circuit having Part No. TPS61200DRCT, manufactured by Texas Instruments of Dallas, Tex. This DC/DC converter circuit **28** is configured with external components to convert the 3.7 volts of the lithium polymer battery BT**1** to a 5 volt DC output.

A 5 volt DC regulated voltage is preferably provided by the DC/DC boost converter circuit **28** to the USB connector **30** of the energy harvesting circuit to which an external electronic device **8** is connectable. Preferably, pin **1** (VCC) of the output connector **30** provides a 5 volt DC output signal to the external electronic device **8**. As shown in FIG. **10** of the drawings, USB output connector or port **30** is preferably positioned on one of the narrower sides **68** of the housing **32** of the energy harvesting device.

Thus, the circuit configuration described above will allow the output of this USB type A connector **28** to be able to provide the correct charge voltage and current for iPhone™, iPod™ and iPad™ devices, which are manufactured by Apple, Inc. Also, this circuit configuration will be compatible and provide correct charge voltage and current to most if not all smart phones, PDAs, MP3 players and Blackberry™ devices.

As mentioned previously, the energy harvesting device of the present invention may include an input connector **24** for receiving a charging voltage from an external power source, such as another electronic device, or a power transformer which may be connected to a cigarette lighter of an automobile, or may be an AC/DC adapter which is connectable to a standard house wall outlet, for example. The power signal from this external electrical power source is provided to the anode of an isolation diode D**4**, whose cathode is in electrical communication with the positive side of the lithium battery BT**1** or other storage device internal to the energy harvesting circuit so that the lithium battery or storage device may be charged thereby.

More specifically, and as shown in FIGS. **3** and **9** of the drawings, a mini USB connector **24** is provided on one side (preferably on one of the narrower sides **68** not occupied by output USB connector **30**) of the housing **32** of the energy harvesting device. Input connector **24** is provided so that the user may connect to connector **24** an AC/DC wall outlet adapter, or another adapter or transformer which is connectable to an auxiliary power outlet or cigarette lighter in an automobile, or to an output port providing a 5 volt DC output signal of an electronic device, such as a laptop computer. Preferably, pin **1** (VCC) of connector **24** provides a 5 volt DC input signal to the circuit.

The 5 volt DC input signal on connector **24** is provided to the anode of isolation diode D**4**. The cathode of diode D**4** is electrically connected directly or through a resistor (not show) to the positive side of the lithium battery BT**1** so that the lithium battery of the energy harvesting circuit may be charged externally by the AC/DC adapter, power transformer, automobile power outlet, laptop computer or the like connected to the mini USB connector **24**.

Also, with an external power source connected to the input USB connector **24**, and with switch SW**1** in a closed position, the external power source will provide power to the output USB connector **30** and to an external electronic device **8** coupled thereto.

Referring now to FIGS. **5**-**10** of the drawings, it will be seen that a preferred form of the housing **32** for the energy harvesting circuit includes a main body or base **36** which is rectangular in shape overall, and a transparent or translucent cover **34** that is pivotally joined to a longer edge (or, alternatively, a narrower edge) of the top surface **38** of the main body **36**. The solar cell **40** is mounted on the top surface **38** of the main body **36** or through a window formed through the thickness of the main body so as to be exposed to ambient light. The transparent cover **34** is pivotable on the top surface **38** between a closed position, where it covers the top surface **38** or at least the solar cell **40** mounted on the top surface, and an open position, where it is raised from the top surface **38** of the main body **36** to a position which is substantially perpendicular to the plane in which the top surface **38** of the main body resides. Thus, if the bottom surface **70**, which is opposite the top surface **38**, of the main body **36** rests on a horizontal supporting surface, such as a table or desk, then the transparent cover **34** may be pivoted to a substantially vertical position.

As mentioned previously, there are preferably four antennas **2***a*-**2***d* which are spaced apart from each other a predetermined distance, which are mounted to the transparent cover **34** on the inside surface thereof. For best WIFI or RF signal reception, it is preferred that these broadband antennas **2***a*-**2***d*, preferably being monopole antennas, are disposed in a vertical position. As such, it is preferred if the user of the energy harvesting device of the present invention raises the transparent cover **34** to a vertical position when he chooses to charge the internal battery BT**1** of the energy harvesting circuit with WIFI or RF signals present in the location where the energy harvesting device is situated.

Preferably, the cover **34** is transparent or translucent, as mentioned previously, so that the solar cell **40** of the energy harvesting circuit will receive light energy through the cover **34** to charge the internal lithium battery BT**1**, even if the cover **34** is in a closed position.

Also, as mentioned previously, so as not to discharge the internal lithium battery BT**1** unnecessarily when no external electronic device **8** is being powered by the energy harvesting circuit, a sliding cover **66** is provided over the output USB connector **30**, and the switch SW**1** is operatively coupled to, and activated by, movement of the USB connector sliding cover **66**.

More specifically, the USB output connector **30** is preferably situated on a smaller side **68** of the main body **36** of the housing **32** opposite that of the input USB connector **24** (although the connectors **30** and **24** could be mounted on opposite wider sides **72** of the housing **32**). There is a small cover **66** for the output USB connector **30** so that a user may slide this cover **66** and connect the output USB connector **30** to an external electronic device **8** in order to power or recharge the batteries of the external electronic device. When the sliding cover **66** is positioned such that it uncovers the output USB connector **30**, it automatically engages and activates the single pole, single throw switch or switch circuit SW**1** so that the battery BT**1** or other storage device is electrically connected to the input of the DC/DC boost converter **28** to provide a 5 volt output voltage on the output USB connector **30**.

When the sliding cover **66** is returned to its closed position, covering the output USB connector **30**, it deactivates the switch or switch circuit SW**1**, breaking the connection between the lithium battery BT**1** or other storage device within the energy harvesting circuit and the DC/DC converter circuit **28**, so that no power is provided to the output USB connector **30**. The sliding cover **66**, in this closed position, covers the output USB connector **30** to prevent access thereto. In a preferred form of the present invention, the sliding cover **30** may be biased by a spring (not shown) so that it automatically returns to a closed position, covering the output connector **30** and deactivating switch SW**1**, when no external electronic device **8** is connected to output connect or connector **30** of the energy harvesting device.

The purpose of having the sliding cover **66**, covering and uncovering the output USB connector **30**, is to minimize any energy drain from the lithium battery BT**1** when there is no need to provide power to an external electronic device **8**. It should be realized that the user controlled switch **26** mounted on the top surface **38** of the base **36** may be omitted if the sliding cover **66** coupled to switch SW**1** to activate switch SW**1** is used. Alternatively, the switch labeled with reference number **26** shown in FIGS. **5-10** may instead be used as switch SW**2**, which may be a momentary push button switch, which the user presses to momentarily activate the battery charge monitoring circuit **60** to illuminate the red LED **62** or the green LED **64** so as to minimize any drain on the lithium battery BT**1**.

In one form of the present invention, the four antennas **2***a*-**2***d* are identical to one another and have the same or similar broadband bandwidth characteristics. In an alternative form of the present invention, the energy harvesting circuit may include several narrow band antennas **2**, such as one UHF antenna to receive signals in the range from between about 500 MHz and about 700 MHz, one cell phone and 900 MHz antenna to receive signals in the range from between about 850 MHz and about 928 MHz, one GSM band antenna to receive signals in the range from between about 1.8 GHz and about 1.9 GHz, and one WiFi/WiMAx/Satellite radio antenna to receive signals in the range from between about 2.3 GHz and about 2.5 GHz, each antenna **2** being connected to a respective multiplier/rectifier circuit **42** through a coaxial cable **4** (see FIG. **9**).

Reference should now be had to FIGS. **4** and **11** of the drawings, which show another (third) preferred embodiment of the energy harvesting device of the present invention. The circuit shown in the schematic diagram of FIG. **4** is similar in many respects to the circuit shown in FIG. **3**. However, in this preferred form, there are six (6) antennas **2***a*-**2***f* provided (also as shown in FIG. **10**) and six (6) multiplier/rectifier circuits **42***a*-**42***f*. Therefore, there are nodes A-F associated with the outputs of the first multiplier/rectifier circuit **42***a* through the sixth multiplier/rectifier circuit **42***f*, respectively.

In this alternative embodiment of the circuit shown in FIG. **4**, the output voltages from the first multiplier/rectifier circuit **42***a* and the second multiplier/rectifier circuit **42***b* are stacked together in series on top of the voltage of the jump start battery BT**2**, and the stacked voltage of the first multiplier/rectifier circuit **42***a* and the second multiplier/rectifier circuit **42***b* is provided at node B and to the temporary storage capacitor C**3** through a diode D**60**, whose anode is connected to node B and whose cathode is connected to temporary storage capacitor C**3**.

Similarly, the output voltages of the third multiplier/rectifier circuit **42***c* and the fourth multiplier/rectifier circuit **42***d* are stacked in series on the voltage of the jump start battery BT**2**. The stacked voltage of the fourth multiplier/rectifier circuit **42***d* and the third multiplier/rectifier circuit **42***c* is provided at node D and to the temporary storage capacitor C**3** through diode D**62**, whose anode is connected to node D and whose cathode is connected to the temporary storage capacitor C**3**.

In the same manner, the output voltage of the fifth multiplier/rectifier circuit **42***e* and the output voltage of the sixth multiplier/rectifier circuit **42***f*, whose inputs are respectively connected to the fifth antenna **2***e* and the sixth antenna **2***f*, are stacked in series on the voltage of the jump start battery BT**2**. The stacked voltage of the sixth multiplier/rectifier circuit **42***f* and the fifth multiplier/rectifier circuit **42***e* is provided at node F (node E is the output voltage at the fifth multiplier/rectifier circuit **42***e*) and is provided to the temporary storage capacitor C**3** through diode D**64**, whose anode is connected to node F and whose cathode is connected to the temporary storage capacitor C**3**.

More specifically, the positive side of the jump start battery BT**2** is connected to a first leg of each of capacitors C**39**, C**21** and C**100**. The second leg of each of capacitors C**39**, C**21** and C**100** is respectively connected to node A and capacitor C**34** (and the output of the first multiplier/rectifier circuit **42***a*); node C and capacitor C**12** (and the output of the third multiplier/rectifier circuit **42***c*); and node E and the first leg of capacitor C**102** (and the output of the fifth multiplier/rectifier circuit **42***e*). The second leg of capacitor C**102** is coupled to node F, the anode of diode D**64** and the output of the sixth multiplier/rectifier circuit **42***f*.

In other words, the outputs of the first and second multiplier/rectifier circuits **42***a*, **42***b* are connected in series, the outputs of the third and fourth multiplier/rectifier circuits **42***c*, **42***d* are connected in series, and the outputs of the fifth and sixth multiplier/rectifier circuits **42***e*, **42***f* are connected in series, to form three groups, and the output voltages of the series-connected groups are effectively connected in parallel with each other and provided to the temporary storage capacitor C**3**. This provides approximately 8-12 volts to charge the temporary storage capacitor C**3**, but at greater current capacity (because of the parallel arrangement of multiplier/rectifier groups) than having the outputs of the individual multiplier/rectifier circuits **42***a*-**42***f* all connected in series in a stack.

It should be realized that a greater or fewer number of parallel-connected groups of series-connected multiplier/rectifier circuits **42***a*-**42***f* may be used to generate the voltage to charge the temporary storage capacitor C**3**. For example, the third group consisting of the fifth and sixth multiplier/rectifier circuits **42***e*, **42***f*, and the corresponding fifth and sixth antennas **2***e*, **2***f* to which the fifth and sixth multiplier/rectifier circuits **42***e*, **42***f* are respectively connected, may be omitted to satisfy space constraints in the arrangement of the antennas **2** on the housing **32** of the energy harvesting device and to provide sufficient separation between adjacent antennas **2** in order to minimize any cross-coupling of the antennas.

This alternative circuit of the energy harvesting device of the present invention also includes a 6 volt DC, 2 amp (or the like) input connector **74**, which allows the main energy storage battery BT**1** to be charged up to approximately 75 percent of its capacity in about 30 minutes. As can be seen from FIG. **4** of the drawings, the output voltage on the quick charge connector **74** is provided to the main energy storage battery BT**1** through diode D**68**, whose anode is operatively connected to the output of connector **74** and whose cathode is operatively connected to the main energy storage battery BT**1**. Also, the output voltage on quick charge connector **74** is provided to the jump start battery BT**2** to charge that battery as well, through diode D**66**, whose anode is operatively connected to the output of the connector **74** and whose cathode is operatively connected to the positive side of jump start battery BT**2**.

Furthermore, the 5 volt DC connector **24** provides a charging voltage to the main energy storage battery BT**1** through diode D**4**, as mentioned previously with the circuit shown in FIG. **3**, but also provides a charging voltage to the jump start battery BT**2** through diode D**70**, whose anode is operatively connected to the output of connector **24** and whose cathode is operatively connected to the positive side of jump start battery BT**2**.

Similarly, the output voltage of the power management circuit **46**, which is derived from the solar cell voltage, is provided through diode D**3** to the main energy storage battery BT**1** and through diode D**50** to the positive side of jump start battery BT**2**, in the same manner as shown in the schematic diagram for the circuit shown in FIG. **3**.

With reference to FIG. **11** of the drawings, it can be seen that the six (6) antennas **2***a*-**2***f* are arranged side-by-side on the cover **34** of the housing **32**, which may be raised from the base **36** of the housing vertically so as to optimize the reception capabilities of the antennas. Although in FIG. **11** the six antennas **2***a*-**2***f* are shown mounted on the cover **34** transversely along the narrower side of the cover, which is hingedly joined to the base **36** of the housing **32**, it is envisioned to be within the scope of the present invention to arrange the six antennas **2***a*-**2***f* along the longer side of the cover **34** and having the longer side of the cover hingedly mounted to the longer side **72** of the base **36** of the housing **32** to provide a greater distance between adjacent antennas **2***a*-**2***f* and to minimize or eliminate any possible coupling between adjacent antennas.

The energy harvesting circuit of the present invention, in its preferred form, detects WIFI signals, WiMax, Satellite Radio, Cellular Phone (850 MHz to 900 MHz and 1.8 GHz-1.9 GHz), any 2.4 GHz and UHF TV signals and light to harvest energy therefrom to charge an internal battery BT**1** or other storage device for use on demand to power or recharge the batteries of an external electronic device **8** connected thereto. When not in use, the energy which is detected and harvested is stored in the battery BT**1** or other storage device so that the battery or other storage device is fully charged and ready to power an external electronic device **8** when connected thereto by the user.

It should be further realized that the energy harvesting device or at least portions of the electrical circuit thereof may be incorporated directly into a cellular phone, PDA, camera, Blackberry™ device or other electronic device **8** to power the device or charge the batteries thereof, without using a stand alone energy harvesting device within a separate housing **32**, such as shown in FIGS. **5-11**. The antenna or antennas **2**, and the solar cell **40**, of the energy harvesting device may be mounted on or within the housing of the electronic device to receive transmitted RF signals and light energy, respectively, in order to power the electronic device **8** or charge the batteries thereof. It is also envisioned to be within the scope of the present invention to incorporate the electrical circuit of the energy harvesting device directly on or in the battery used to power the electronic device **8**.

The various embodiments of the energy harvesting device of the present invention will now further be described in detail.

In one form of the present invention, an energy harvesting device preferably includes an antenna **2** for receiving a transmitted radio frequency (RF) signal, a rectifier circuit **6** operatively coupled to the antenna **2**, the rectifier circuit **6** converting the RF signal received by the antenna **2** to a direct current (DC) voltage, and an output connector **30** operatively coupled to the rectifier circuit **6** and providing the DC voltage thereon, an electronic device **8** being connectable to the output connector **30** for receiving the DC voltage provided on the output connector **30**.

In another form of the present invention, an energy harvesting device preferably includes an antenna **2** for receiving a transmitted radio frequency (RF) signal, the antenna **2** converting the received RF signal to an RF electrical signal, an RF-to-DC rectifier circuit **12**, the RF-to-DC rectifier circuit **12** being operatively connected to the antenna **2** and converting the RF electrical signal to a direct current (DC) voltage, and a first storage device **14**, the first storage device **14** being operatively connected to the RF-to-DC rectifier circuit **12** and being provided with the DC voltage and being charged thereby. The energy harvesting device preferably further includes a microcontroller **16**, the microcontroller **16** having a first signal input operatively connected to the first storage device **14** to monitor the DC voltage on the first storage device **14**, the microcontroller **16** generating a pulsed output signal if the DC voltage on the first storage device **14** is above a first pre-determined voltage level, a first stage DC-to-DC voltage converter circuit **20**, the first stage DC-to-DC voltage converter circuit **20** being operatively connected to the first storage device **14** and generating an output DC voltage which is greater than the DC voltage on the first storage device **14**, and a second stage DC-to-DC voltage converter circuit **22**, the second stage DC-to-DC voltage converter circuit **22** being operatively connected to the first stage DC-to-DC voltage converter circuit **20** and generating an output DC voltage which is less than the output DC voltage generated by the first stage DC-to-DC voltage converter circuit **20**. The energy harvesting device preferably further includes a second storage device **18**, the second storage device **18** being operatively connected to the second stage DC-to-DC voltage converter circuit **22** and being provided with the output DC voltage generated by the second stage DC-to-DC voltage converter circuit **22** and being charged thereby, a third stage DC-to-DC voltage converter circuit **28**, the third stage DC-to-DC voltage converter circuit **28** being operatively connected to the second storage device **18** and generating an output DC voltage which is greater than the DC voltage on the second storage device **18**, and an output connector **30**, the output connector **30** being operatively connected to the third stage DC-to-DC voltage converter circuit **28** and providing the output DC voltage generated by the third stage DC-to-DC voltage converter circuit **28** thereon. An electronic device **8** is connectable to the output connector **30** for receiving thereon the output DC voltage provided to the output connector **30**.

The energy harvesting device described above may further include an input connector **24**. The input connector **24** is operatively connected to the second stage DC-to-DC voltage converter circuit **22**. An external source of DC voltage is connectable to the input connector **24** for providing the DC voltage from the external source to the second stage DC-to-DC voltage converter circuit **22**. The second stage DC-to-DC voltage converter circuit **22** generates the output DC voltage of the second stage DC-to-DC voltage converter circuit **22** in response to the DC voltage from the external source provided on the input connector **24**.

In a preferred form of the invention, the first storage device **14** of the energy harvesting device includes a capacitor. In a preferred form of the invention, the second storage device **18** of the energy harvesting device includes a battery.

Preferably, the microcontroller **16** includes a second signal input operatively connected to the second storage device **18** to monitor the DC voltage on the second storage device **18**. The microcontroller **16** does not generate the pulsed output signal if the DC voltage on the second storage device **18** is above a second predetermined voltage level.

Another preferred form of the energy harvesting device is described as follows. The energy harvesting device includes a plurality of antennas **2** for receiving a transmitted radio frequency (RF) signal. The antennas **2** generate RF electrical signals in response to the transmitted RF signals received thereby. A plurality of multiplier and rectifier circuits **42** is also included. Each multiplier and rectifier circuit **42** is responsive to a respective RF electrical signal generated by an antenna **2** of the plurality of antennas and generates a direct current (DC) voltage in response thereto. At least some of the DC voltages are stacked in series to provide a combined DC voltage, the combined DC voltage being greater in magnitude than each individual DC voltage generated by a respective multiplier and rectifier circuit **42**.

The energy harvesting device of this preferred embodiment further includes a temporary storage device **14**. The temporary storage device **14** is responsive to the combined DC voltage and provides a temporarily stored DC voltage in response thereto. The temporary storage device **14** may be or include a capacitor. A first power management circuit **46** is also preferably included, the first power management circuit **46** being responsive to the temporarily stored DC voltage and generating a DC charging voltage in response thereto.

The energy harvesting device further preferably includes a main energy storage device **18**. The main energy storage device **18** is responsive to the DC charging voltage generated by the first power management circuit **46** and provides a main energy storage DC voltage in response thereto. The main energy storage device **18** may be or include a battery, such as a lithium polymer battery. An output connector **30** is also preferably included in the device, the output connector **30** providing an output connector DC voltage thereon in response to the main energy storage DC voltage. An electronic device **8** is connectable to the output connector **30** for receiving thereon the output connector DC voltage provided on the output connector **30**.

In another preferred form of the present invention, the energy harvesting device may further include a jump start DC voltage device **44**. The jump start DC voltage device **44** provides a jump start DC voltage. The combined DC voltage includes the at least some of the DC voltages generated by the plurality of multiplier and rectifier circuits **42** stacked in series and the jump start DC voltage, if such a jump start DC voltage device **44** is included. The jump start DC voltage device **44** may be or include a battery.

An input connector **24** may be included in the energy harvesting device described above. The input connector **24** is connectable to a source of DC voltage external to the energy harvesting device, the DC voltage of the external source being provided to at least one of the main energy storage device **18**, the jump start DC voltage device **44** and the output connector **30**.

The energy harvesting device described above may also include a DC-to-DC voltage converter circuit **28**. The DC-to-DC voltage converter circuit **28** is responsive to the main energy storage DC voltage and generates an output DC voltage in response thereto. The output connector **30** provides the output connector DC voltage in response to the output DC voltage generated by the DC-to-DC voltage converter circuit **28**, if such is included.

The energy harvesting device also preferably includes a watchdog monitor circuit **50**. The watchdog monitor circuit **50** is responsive to the temporarily stored DC voltage and selectively provides the temporarily stored DC voltage to the first power management circuit **46** if the temporarily stored DC voltage is greater than a predetermined voltage threshold.

In a more preferred form of the present invention, the energy harvesting device includes a solar cell circuit having a solar cell **40**. The solar cell circuit receives light energy and generates a solar cell DC output voltage in response thereto. The solar cell DC output voltage is provided to at least one of the main energy storage device **18**, the jump start DC voltage device **44** and the output connector **30**. Furthermore, a second power management circuit **48** may be included for the solar cell circuit. The second power management circuit **48** is responsive to the solar cell DC output voltage and generates a managed solar cell DC output voltage in response thereto. The managed solar cell DC output voltage is provided to at least one of the main energy storage device **18**, the jump start DC voltage device **44** and the output connector **30**.

It should be noted that at least some of the DC voltages generated by the multiplier and rectifier circuits **42** may be coupled to provide group DC voltages, and the group DC voltages may be combined in parallel to define the combined DC voltage provided to the temporary storage device **14**.

In a more specific form of the present invention, the plurality of antennas **2** of the energy harvesting device includes a first antenna **2***a*, a second antenna **2***b*, a third antenna **2***c* and a fourth antenna **2***d*, the first antenna **2***a* generating a first RF electrical signal, the second antenna **2***b* generating a second RF electrical signal, the third antenna **2***c* generating a third RF electrical signal, and the fourth antenna **2***d* generating a fourth RF electrical signal. Furthermore, the plurality of multiplier and rectifier circuits **42** of the energy harvesting device includes a first multiplier and rectifier circuit **42***a*, a second multiplier and rectifier circuit **42***b*, a third multiplier and rectifier circuit **42***c* and a fourth multiplier and rectifier circuit **42***d*, the first multiplier and rectifier circuit **42***a* being responsive to the first RF electrical signal generated by the first antenna **2***a* and generating a first multiplied and rectified DC voltage in response thereto, the second multiplier and rectifier circuit **42***b* being responsive to the second RF electrical signal generated by the second antenna **2***b* and generating a second multiplied and rectified DC voltage in response thereto, the third multiplier and rectifier circuit **42***c* being responsive to the third RF electrical signal generated by the third antenna **2***c* and generating a third multiplied and rectified DC voltage in response thereto, and the fourth multiplier and rectifier circuit **42***d* being responsive to the fourth RF electrical signal generated by the fourth antenna **2***d* and generating a fourth multiplied and rectified DC voltage in response thereto. The first multiplied and rectified DC voltage and the second multiplied and rectified DC voltage are preferably coupled in series to define a first group DC voltage. Similarly, the third multiplied and rectified DC voltage and the fourth multiplied and rectified DC voltage are preferably coupled in series to define a second group DC voltage. The first group DC voltage and the second group DC voltage are preferably combined in parallel to define the combined DC voltage which is provided to the temporary storage device **14**.

Each antenna **2** of the plurality of antennas may have either a broadband bandwidth characteristic or a narrowband bandwidth characteristic, or the antennas **2** may be a mix of broadband and narrowband bandwidth characteristics.

The energy harvesting device of the present invention preferably includes a housing **32**. The housing **32** has a base **36** defining an internal cavity for containing the electrical circuit of the energy harvesting device, and a cover **34** pivotally mounted on the base **36**. The base **36** has a top surface **38**, and the cover **34** is preferably pivotable between at least a closed position in which the cover **34** is in close proximity to the base **36** to at least partially cover the top surface **38**, and an open position in which the cover **34** is substantially perpendicular to the top surface **38** of the base.

The plurality of antennas **2** is preferably mounted on the pivotable cover **34** of the housing **32** and are positionable in at least a first position in which the plurality of antennas **2** is in close proximity to the top surface **38** of the base **36** when the cover **34** is in the closed position, and a second position in which the plurality of antennas **2** is in a raised position from the top surface **38** of the base **36** when the cover **34** is in the open position.

If a solar cell circuit is included in the energy harvesting device, where the solar cell circuit includes a solar cell **40** for receiving light energy, the solar cell **40** is preferably mounted on the top surface **38** of the base **36** and is selectively covered and uncovered by the cover **34** when the cover **34** is respectively pivoted between the closed position and the open position. The cover **34** is preferably at least partially translucent to at least partially allow light energy to pass therethrough and to be received by the solar cell **40** when the cover **34** is in the closed position.

Alternatively, the energy harvesting device includes a housing **32** having a top surface **38**, whereby the solar cell **40** is mounted on the top surface **38** of the housing. A cover need not be included. The housing **32** may include a base **36** defining an internal cavity for containing the electrical circuit of the energy harvesting device, the top surface **38** of the housing forming a portion of the base **36**. Additionally, a non-pivotable cover **34** may be mounted on the base **36**, the cover **34** at least partially covering the top surface **38** of the base **36** and the solar cell **40** mounted on the top surface **38**. In this embodiment, it is preferred if the cover **34** is at least partially translucent to allow light energy to pass therethrough and impinge on the solar cell **40** situated below the cover **34**.

Now, various forms of a method of harvesting energy in accordance with the present invention will further be described in detail.

In one form of the present invention, a method of harvesting energy includes the steps of receiving a transmitted radio frequency (RF) signal to provide a received RF signal, converting the received RF signal to an RF electrical signal, converting the RF electrical signal to a direct current (DC) voltage, temporarily storing the DC voltage to provide a temporarily stored DC voltage, monitoring the temporarily stored DC voltage and generating a pulsed output signal if the temporarily stored DC voltage is above a first predetermined voltage level, converting the temporarily stored DC voltage to a first converted DC voltage which is greater than the temporarily stored DC voltage, converting the first converted DC voltage to a second converted DC voltage which is less than the first converted DC voltage, storing the second converted DC voltage to provide a stored second converted DC voltage, converting the stored second converted DC voltage to a third converted DC voltage which is greater than the stored second converted DC voltage, and providing the third converted DC voltage to an output connector **30**, an electronic device **8** being connectable to the output connector **30** for receiving thereon the third converted DC voltage provided to the output connector **30**.

The method of harvesting energy described above may further include the step of connecting a source of external DC voltage to an input connector **24**, the input connector **24** receiving the external DC voltage, the first converted DC voltage being at least partially defined by the external DC voltage.

Furthermore, the method of harvesting energy described above may further include the step of monitoring the stored second converted DC voltage and not generating the pulsed output signal if the stored second converted DC voltage is above a second predetermined voltage level.

In another form of the present invention, a method of harvesting energy includes the steps of receiving a transmitted radio frequency (RF) signal to provide a received RF signal, converting the received RF signal to a direct current (DC) voltage, and providing the DC voltage on an output connector **30**, an electronic device **8** being connectable to the output connector **30** for receiving the DC voltage provided on the output connector **30**.

In yet another form of the present invention, a method of harvesting energy includes the steps of receiving by a plurality of antennas **2** a transmitted radio frequency (RF) signal and generating RF electrical signals in response to the transmitted RF signals received by the plurality of antennas **2**, multiplying and rectifying the RF electrical signals generated by the plurality of antennas **2** and generating a plurality of direct current (DC) voltages in response thereto, combining in series at least some of the DC voltages to provide a combined DC voltage, the combined DC voltage being greater in magnitude than each individual DC voltage, temporarily storing the combined DC voltage thereby providing a temporarily stored DC voltage, generating a DC charging voltage in response to the temporarily stored DC voltage, storing the DC charging voltage to provide a stored DC charging voltage, and generating a DC output connector voltage on an output connector **30** in response to the stored DC charging voltage, an electronic device **8** being connectable to the output connector **30** for receiving thereon the output connector DC voltage provided on the output connector **30**.

Furthermore, the method of harvesting energy described above may further include the step of generating a jump start DC voltage, the combined DC voltage including the at least some of the DC voltages generated by the multiplying and rectifying step connected in series and the jump start DC voltage, and may further include the step of connecting an external DC voltage to an input connector **24**, at least one of the stored DC charging voltage, the jump start DC voltage and the output connector DC voltage being defined at least partially by the external DC voltage.

Additionally, the method of harvesting energy described previously may include the step of converting the stored DC charging voltage to the output connector DC voltage provided on the output connector **30**, and may include the further step of monitoring the temporarily stored DC voltage and selectively generating the DC charging voltage if the temporarily stored DC voltage is greater than a predetermined voltage threshold.

If a solar cell circuit is included in the energy harvesting device of the present invention, then the method of harvesting energy described above may include the step of receiving light energy by the solar cell circuit and generating a solar cell DC output voltage in response thereto, the at least one of the stored DC charging voltage, the output connector DC voltage provided on the output connector and the jump start DC voltage being defined at least partially by the solar cell DC output voltage.

The method of harvesting energy may further include the steps of coupling in series at least some of the DC voltages generated by the multiplying and rectifying step to provide group DC voltages, and combining the group DC voltages in parallel to define the combined DC voltage.

The method of harvesting energy may further include the step of adjustably positioning the plurality of antennas **2** between a first position and a second position, wherein the plurality of antennas **2** exhibits better signal reception characteristics to receive a transmitted RF signal when in the first position than when in the second position.

Although illustrative embodiments of the present invention have been described herein with reference to the accompanying drawings, it is to be understood that the invention is not limited to those precise embodiments, and that various other changes and modifications may be effected therein by one skilled in the art without departing from the scope or spirit of the invention.

What is claimed is:

1. An energy harvesting device, which comprises:

at least one antenna for receiving a transmitted radio frequency (RF) signal, the at least one antenna converting the received RF signal to an RF electrical signal;

at least one RF-to-DC rectifier circuit, the at least one RF-to-DC rectifier circuit being operatively connected to the at least one antenna and converting the RF electrical signal to a direct current (DC) voltage;

a first storage device, the first storage device being operatively connected to the at least one RF-to-DC rectifier circuit and being provided with the DC voltage and being charged thereby;

a microcontroller, the microcontroller having a first signal input operatively connected to the first storage device to monitor the DC voltage on the first storage device, the microcontroller generating a pulsed output signal if the DC voltage on the first storage device is above a first pre-determined voltage level;

a first stage DC-to-DC voltage converter circuit, the first stage DC-to-DC voltage converter circuit being operatively connected to the first storage device and generating an output DC voltage which is greater than the DC voltage on the first storage device;

a second stage DC-to-DC voltage converter circuit, the second stage DC-to-DC voltage converter circuit being operatively connected to the first stage DC-to-DC voltage converter circuit and generating an output DC voltage which is less than the output DC voltage generated by the first stage DC-to-DC voltage converter circuit;

a second storage device, the second storage device being operatively connected to the second stage DC-to-DC voltage converter circuit and being provided with the output DC voltage generated by the second stage DC-to-DC voltage converter circuit and being charged thereby;

a third stage DC-to-DC voltage converter circuit, the third stage DC-to-DC voltage converter circuit being operatively connected to the second storage device and generating an output DC voltage which is greater than the DC voltage on the second storage device; and

an output connector, the output connector being operatively connected to the third stage DC-to-DC voltage converter circuit and providing the output DC voltage generated by the third stage DC-to-DC voltage converter circuit thereon, an electronic device being connectable to the output connector for receiving thereon the output DC voltage provided to the output connector.

2. An energy harvesting device as defined by claim **1**, which further comprises:

an input connector, the input connector being operatively connected to the second stage DC-to-DC voltage converter circuit, the input connector being connectable to an external source of DC voltage for providing the DC voltage from the external source to the second stage DC-to-DC voltage converter circuit, the second stage DC-to-DC voltage converter circuit generating the output DC voltage of the second stage DC-to-DC voltage converter circuit in response to the DC voltage from the external source provided on the input connector.

3. An energy harvesting device as defined by claim **1**, wherein the first storage device includes a capacitor.

4. An energy harvesting device as defined by claim **1**, wherein the second storage device includes a battery.

5. An energy harvesting device as defined by claim **1**, wherein the microcontroller includes a second signal input operatively connected to the second storage device to monitor the DC voltage on the second storage device, the microcontroller not generating the pulsed output signal if the DC voltage on the second storage device is above a second predetermined voltage level.

6. An energy harvesting device as defined by claim **1**, which further comprises:

an input connector, the input connector being connectable to a source of DC voltage external to the energy harvesting device, the DC voltage of the external source being provided to at least one of the second storage device and the output connector.

7. An energy harvesting device as defined by claim **1**, which further comprises:

a solar cell circuit, the solar cell circuit receiving light energy and generating a solar cell DC output voltage in response thereto, the solar cell DC output voltage being provided to at least one of the second storage device and the output connector.

8. An energy harvesting device as defined by claim **1**, wherein the at least one antenna includes a plurality of antennas.

9. An energy harvesting device as defined by claim **8**, wherein each antenna of the plurality of antennas has a broadband bandwidth characteristic.

10. An energy harvesting device as defined by claim **8**, wherein each antenna of the plurality of antennas has a narrowband bandwidth characteristic.

11. An energy harvesting device as defined by claim **1**, which further comprises:

a housing, the housing having a base defining an internal cavity, and a cover pivotally mounted on the base, the base having a top surface, the cover being pivotable between at least a closed position in which the cover is in close proximity to the base to at least partially cover the top surface, and an open position in which the cover is substantially perpendicular to the top surface of the base.

12. An energy harvesting device as defined by claim **11**, wherein the at least one antenna is mounted on the pivotable cover of the housing and being positionable in at least a first position in which the at least one antenna is in close proximity to the top surface of the base when the cover is in the closed position, and a second position in which the at least one antenna is in a raised position from the top surface of the base when the cover is in the open position.

13. An energy harvesting device as defined by claim **11**, which further comprises:

a solar cell circuit, the solar cell circuit including a solar cell for receiving light energy, the solar cell circuit generating a solar cell DC output voltage in response to the light energy received by the solar cell, the solar cell being mounted on the top surface of the base and selectively covered and uncovered by the cover when the cover is respectively pivoted between the closed position and the open position.

14. An energy harvesting device as defined by claim **1**, which further comprises:

a solar cell circuit, the solar cell circuit including a solar cell for receiving light energy, the solar cell circuit generating a solar cell DC output voltage in response to the light energy received by the solar cell; and

a housing, the housing having a top surface, the solar cell being mounted on the top surface of the housing.

15. An energy harvesting device as defined by claim **14**, wherein the housing includes a base defining an internal cavity, and a cover mounted on the base, the top surface of the housing forming a portion of the base, the cover at least partially covering the top surface of the base and the solar cell mounted on the top surface.

16. An energy harvesting device as defined by claim **15**, wherein the cover is at least partially translucent to at least partially allow light energy to pass therethrough and to be received by the solar cell.

17. An energy harvesting device as defined by claim **12**, wherein the cover is at least partially translucent to at least partially allow light energy to pass therethrough and to be received by the solar cell when the cover is in the closed position.

18. An energy harvesting device, which comprises:

a plurality of antennas for receiving a transmitted radio frequency (RF) signal, the antennas generating RF electrical signals in response to the transmitted RF signals received thereby;

a plurality of multiplier and rectifier circuits, each multiplier and rectifier circuit being responsive to a respective RF electrical signal generated by an antenna of the plurality of antennas and generating a direct current (DC) voltage in response thereto, at least some of the DC voltages being stacked in series to provide a combined DC voltage, the combined DC voltage being greater in magnitude than each individual DC voltage generated by a respective multiplier and rectifier circuit;

a temporary storage device, the temporary storage device being responsive to the combined DC voltage and providing a temporarily stored DC voltage in response thereto;

a first power management circuit, the first power management circuit being responsive to the temporarily stored DC voltage and generating a DC charging voltage in response thereto;

a main energy storage device, the main energy storage device being responsive to the DC charging voltage generated by the first power management circuit and providing a main energy storage DC voltage in response thereto; and

an output connector, the output connector providing an output connector DC voltage thereon in response to the main energy storage DC voltage, an electronic device being connectable to the output connector for receiving thereon the output connector DC voltage provided on the output connector,

wherein at least some of the DC voltages generated by the multiplier and rectifier circuits are coupled in series to provide group DC voltages, and wherein the group DC voltages are combined in parallel to define the combined DC voltage;

wherein the plurality of antennas includes a first antenna, a second antenna, a third antenna and a fourth antenna, the first antenna generating a first RF electrical signal, the second antenna generating a second RF electrical signal, the third antenna generating a third RF electrical signal, the fourth antenna generating a fourth RF electrical signal;

wherein the plurality of multiplier and rectifier circuits includes a first multiplier and rectifier circuit, a second multiplier and rectifier circuit, a third multiplier and rectifier circuit and a fourth multiplier and rectifier circuit, the first multiplier and rectifier circuit being responsive to the first RF electrical signal generated by the first antenna and generating a first multiplied and rectified DC voltage in response thereto, the second multiplier and rectifier circuit being responsive to the second RF electrical signal generated by the second antenna and generating a second multiplied and rectified DC voltage in response thereto, the third multiplier and rectifier circuit being responsive to the third RF electrical signal generated by the third antenna and generating a third multiplied and rectified DC voltage in response thereto, the fourth multiplier and rectifier circuit being responsive to the fourth RF electrical signal generated by the fourth antenna and generating a fourth multiplied and rectified DC voltage in response thereto;

wherein the first multiplied and rectified DC voltage and the second multiplied and rectified DC voltage are coupled in series to define a first group DC voltage;

wherein the third multiplied and rectified DC voltage and the fourth multiplied and rectified DC voltage are coupled in series to define a second group DC voltage;

and wherein the first group DC voltage and the second group DC voltage are combined in parallel to define the combined DC voltage.

19. An energy harvesting device as defined by claim **18**, which further comprises:

a jump start DC voltage device, the jump start DC voltage device providing a jump start DC voltage, the combined DC voltage including the at least some of the DC voltages generated by the plurality of multiplier and rectifier circuits stacked in series and the jump start DC voltage.

20. An energy harvesting device as defined by claim **18**, which further comprises:

an input connector, the input connector being connectable to a source of DC voltage external to the energy harvesting device, the DC voltage of the external source being provided to at least one of the main energy storage device and the output connector.

21. An energy harvesting device as defined by claim **18**, which further comprises:

a DC-to-DC voltage converter circuit, the DC-to-DC voltage converter circuit being responsive to the main energy storage DC voltage and generating an output DC voltage in response thereto, the output connector providing the output connector DC voltage in response to the output DC voltage generated by the DC-to-DC voltage converter circuit.

22. An energy harvesting device as defined by claim **18**, which further comprises:

a watchdog monitor circuit, the watchdog monitor circuit being responsive to the temporarily stored DC voltage and selectively providing the temporarily stored DC voltage to the first power management circuit if the temporarily stored DC voltage is greater than a predetermined voltage threshold.

23. An energy harvesting device as defined by claim **18**, which further comprises:

a solar cell circuit, the solar cell circuit receiving light energy and generating a solar cell DC output voltage in response thereto, the solar cell DC output voltage being provided to at least one of the main energy storage device and the output connector.

24. An energy harvesting device as defined by claim **19**, which further comprises:

a solar cell circuit, the solar cell circuit receiving light energy and generating a solar cell DC output voltage in response thereto, the solar cell DC output voltage being provided to at least one of the main energy storage device, the output connector and the jump start DC voltage device.

25. An energy harvesting device as defined by claim **18**, which further comprises:

a solar cell circuit, the solar cell circuit receiving light energy and generating a solar cell DC output voltage in response thereto; and

a second power management circuit, the second power management circuit being responsive to the solar cell DC output voltage and generating a managed solar cell DC output voltage in response thereto, the managed solar cell DC output voltage being provided to at least one of the main energy storage device and the output connector.

26. An energy harvesting device as defined by claim **19**, which further comprises:

a solar cell circuit, the solar cell circuit receiving light energy and generating a solar cell DC output voltage in response thereto; and

a second power management circuit, the second power management circuit being responsive to the solar cell DC output voltage and generating a managed solar cell DC output voltage in response thereto, the managed solar cell DC output voltage being provided to at least one of the main energy storage device, the output connector and the jump start DC voltage device.

27. An energy harvesting device as defined by claim **18**, wherein each antenna of the plurality of antennas has a broadband bandwidth characteristic.

28. An energy harvesting device as defined by claim **18**, wherein each antenna of the plurality of antennas has a narrowband bandwidth characteristic.

29. An energy harvesting device as defined by claim **18**, which further comprises:

a housing, the housing having a base defining an internal cavity, and a cover pivotally mounted on the base, the base having a top surface, the cover being pivotable between at least a closed position in which the cover is in close proximity to the base to at least partially cover the top surface, and an open position in which the cover is substantially perpendicular to the top surface of the base.

30. An energy harvesting device as defined by claim **29**, wherein the plurality of antennas is mounted on the pivotable cover of the housing and being positionable in at least a first position in which the plurality of antennas is in close proximity to the top surface of the base when the cover is in the closed position, and a second position in which the plurality of antennas is in a raised position from the top surface of the base when the cover is in the open position.

31. An energy harvesting device as defined by claim **29**, which further comprises:

a solar cell circuit, the solar cell circuit including a solar cell for receiving light energy, the solar cell circuit generating a solar cell DC output voltage in response to the light energy received by the solar cell, the solar cell being mounted on the top surface of the base and selectively covered and uncovered by the cover when the cover is respectively pivoted between the closed position and the open position.

32. An energy harvesting device as defined by claim **18**, which further comprises:

a solar cell circuit, the solar cell circuit including a solar cell for receiving light energy, the solar cell circuit generating a solar cell DC output voltage in response to the light energy received by the solar cell; and

a housing, the housing having a top surface, the solar cell being mounted on the top surface of the housing.

33. An energy harvesting device as defined by claim **32**, wherein the housing includes a base defining an internal cavity, and a cover mounted on the base, the top surface of the housing forming a portion of the base, the cover at least partially covering the top surface of the base and the solar cell mounted on the top surface.

34. An energy harvesting device as defined by claim **33**, wherein the cover is at least partially translucent to at least partially allow light energy to pass therethrough and to be received by the solar cell.

35. An energy harvesting device as defined by claim **30**, wherein the cover is at least partially translucent to at least partially allow light energy to pass therethrough and to be received by the solar cell when the cover is in the closed position.

36. An energy harvesting device as defined by claim **18**, wherein the temporary storage device includes a capacitor.

37. An energy harvesting device as defined by claim **18**, wherein the main energy storage device includes a battery.

38. An energy harvesting device as defined by claim **19**, wherein the jump start DC voltage device includes a battery.

39. A method of harvesting energy, which comprises the steps of:

receiving a transmitted radio frequency (RF) signal to provide a received RF signal;

converting the received RF signal to an RF electrical signal;

converting the RF electrical signal to a direct current (DC) voltage;

temporarily storing the DC voltage to provide a temporarily stored DC voltage;

monitoring the temporarily stored DC voltage and generating a pulsed output signal if the temporarily stored DC voltage is above a first predetermined voltage level;

converting the temporarily stored DC voltage to a first converted DC voltage which is greater than the temporarily stored DC voltage;

converting the first converted DC voltage to a second converted DC voltage which is less than the first converted DC voltage;

storing the second converted DC voltage to provide a stored second converted DC voltage;

converting the stored second converted DC voltage to a third converted DC voltage which is greater than the stored second converted DC voltage; and

providing the third converted DC voltage to an output connector, an electronic device being connectable to the output connector for receiving thereon the third converted DC voltage provided to the output connector.

40. A method of harvesting energy as defined by claim **39**, which further comprises the step of:

connecting a source of external DC voltage to an input connector, the input connector receiving the external DC voltage, the first converted DC voltage being at least partially defined by the external DC voltage.

41. A method of harvesting energy as defined by claim **39**, wherein the step of temporarily storing the DC voltage is at least partially performed by using a capacitor.

42. A method of harvesting energy as defined by claim **39**, wherein the step of storing the second converted DC voltage is at least partially performed by using a battery.

43. A method of harvesting energy as defined by claim **39**, which further comprises the step of:

monitoring the stored second converted DC voltage and not generating the pulsed output signal if the stored second converted DC voltage is above a second predetermined voltage level.

44. A method of harvesting energy, which comprises the steps of:

receiving by a plurality of antennas a transmitted radio frequency (RF) signal and generating RF electrical signals in response to the transmitted RF signals received by the plurality of antennas;

multiplying and rectifying the RF electrical signals generated by the plurality of antennas and generating a plurality of direct current (DC) voltages in response thereto;

combining in series at least some of the DC voltages to provide a combined DC voltage, the combined DC voltage being greater in magnitude than each individual DC voltage;

temporarily storing the combined DC voltage thereby providing a temporarily stored DC voltage;

monitoring the temporarily stored DC voltage and generating a pulsed output signal if the temporarily stored DC voltage is above a first predetermined voltage level;

converting the temporarily stored DC voltage to a first converted DC voltage which is greater than the temporarily stored DC voltage;

converting the first converted DC voltage to a second converted DC voltage which is less than the first converted DC voltage;

storing the second converted DC voltage to provide a stored second converted DC voltage;

converting the stored second converted DC voltage to a third converted DC voltage which is greater than the stored second converted DC voltage; and

providing the third converted DC voltage to an output connector, an electronic device being connectable to the output connector for receiving thereon the third converted DC voltage provided to the output connector.

45. A method of harvesting energy as defined by claim **44**, which further comprises the step of:

generating a jump start DC voltage, the combined DC voltage including the at least some of the DC voltages generated by the multiplying and rectifying step connected in series and the jump start DC voltage.

46. A method of harvesting energy as defined by claim **44**, which further comprises the step of:

connecting an external DC voltage to an input connector, at least one of the stored DC charging voltage and the output connector DC voltage being defined at least partially by the external DC voltage.

47. A method of harvesting energy as defined by claim **44**, which further comprises the step of:

converting the stored DC charging voltage to the output connector DC voltage provided on the output connector.

48. A method of harvesting energy as defined by claim **44**, which further comprises the step of:

monitoring the temporarily stored DC voltage and selectively generating the DC charging voltage if the temporarily stored DC voltage is greater than a predetermined voltage threshold.

49. A method of harvesting energy as defined by claim **44**, which further comprises the step of:

receiving light energy by a solar cell circuit and generating a solar cell DC output voltage in response thereto, at least one of the stored DC charging voltage and the output connector DC voltage provided on the output connector being defined at least partially by the solar cell DC output voltage.

50. A method of harvesting energy as defined by claim **45**, which further comprises the step of:

receiving light energy by a solar cell circuit and generating a solar cell DC output voltage in response thereto, the at least one of the stored DC charging voltage, the output connector DC voltage provided on the output connector and the jump start DC voltage being defined at least partially by the solar cell DC output voltage.

51. A method of harvesting energy as defined by claim **44**, which further comprises the steps of:

coupling in series at least some of the DC voltages generated by the multiplying and rectifying step to provide group DC voltages; and

combining the group DC voltages in parallel to define the combined DC voltage.

52. A method of harvesting energy as defined by claim **44**, wherein each antenna of the plurality of antennas has a broadband bandwidth characteristic.

53. A method of harvesting energy as defined by claim **44**, wherein each antenna of the plurality of antennas has a narrowband bandwidth characteristic.

54. A method of harvesting energy as defined by claim **44**, which further comprises the steps of:

adjustably positioning the plurality of antennas between a first position and a second position, wherein the plurality of antennas exhibits better signal reception characteristics to receive a transmitted RF signal when in the first position than when in the second position.

55. A method of harvesting energy as defined by claim **49**, which further comprises the step of:

mounting the plurality of antennas and at least a portion of the solar cell circuit to a housing, the housing having a base defining an internal cavity, and a cover pivotally mounted on the base, the base having a top surface, the cover being pivotable between at least a closed position in which the cover is in close proximity to the base to at least partially cover the top surface, and an open position in which the cover is substantially perpendicular to the top surface of the base, the plurality of antennas being mounted on the pivotable cover of the housing and being positionable in at least a first position in which the plurality of antennas is in close proximity to the top surface of the base when the cover is in the closed position, and a second position in which the plurality of antennas is in a raised position from the top surface of the base when the cover is in the open position, the at least portion of the solar cell circuit being mounted on the top surface of the base and selectively covered and uncovered by the cover when the cover is respectively pivoted between the closed position and the open position.

56. A method of harvesting energy as defined by claim **55**, wherein the cover is at least partially translucent to at least partially allow light energy to pass therethrough and to be received by the at least portion of the solar cell circuit.

57. A method of harvesting energy as defined by claim **44**, wherein the step of temporarily storing the combined DC voltage is at least partially performed by using a capacitor.

58. A method of harvesting energy as defined by claim **44**, wherein the step of storing the DC charging voltage is at least partially performed by using a battery.

59. A method of harvesting energy as defined by claim **45**, wherein the step of generating a jump start DC voltage is at least partially performed by using a battery.

* * * * *

US006954180B1

(12) **United States Patent**
Braun et al.

(10) **Patent No.: US 6,954,180 B1**
(45) **Date of Patent: *Oct. 11, 2005**

---

(54) **ANTENNA DEVICE FOR TRANSMITTING AND/OR RECEIVING RADIO FREQUENCY WAVES AND METHOD RELATED THERETO**

(75) Inventors: **Christian Braun**, Stockholm (SE); **Olov Edvardsson**, Täby (SE); **Leif Eriksson**, Norrtälje (SE)

(73) Assignee: **AMC Centurion AB**, Akersberga (SE)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **09/712,144**

(22) Filed: **Nov. 15, 2000**

Related U.S. Application Data

(63) Continuation of application No. PCT/SE00/02059, filed on Oct. 24, 2000.

(30) **Foreign Application Priority Data**

Oct. 29, 1999 (SE) 9903945

(51) **Int. Cl.**^{7} **H01Q 1/24**; H04B 7/10

(52) **U.S. Cl.** **343/702**; 343/876; 455/277.2; 455/575.7

(58) **Field of Search** .... 343/702, 700 MS, 343/900, 725, 876, 728; H01Q 1/24; 455/90, 455/575.7, 277.1, 277.2

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,693,088 A		9/1972	Rustako, Jr. et al.	455/506
4,123,759 A		10/1978	Hines et al.	343/876
4,958,382 A		9/1990	Imanishi	455/277.1
5,243,358 A		9/1993	Sanford et al.	343/836
5,294,939 A	*	3/1994	Sanford et al.	343/836
5,463,406 A		10/1995	Vannatta et al.	343/725
5,471,146 A		11/1995	Krayeski et al.	324/637
5,486,836 A	*	1/1996	Kuffner et al.	343/700 MS
5,530,926 A		6/1996	Rozanski	455/277.2
5,541,614 A		7/1996	Lam et al.	343/792.5
5,589,844 A		12/1996	Belcher et al.	343/860
5,617,102 A		4/1997	Prater	342/374
5,680,142 A		10/1997	Smith et al.	342/372
5,689,439 A		11/1997	Weerackody et al.	370/329
5,710,995 A		1/1998	Akaiwa et al.	455/277.2
5,754,104 A		5/1998	Saito	340/539
5,777,581 A		7/1998	Lilly et al.	343/700 MS
5,794,145 A		8/1998	Milam	455/426
5,920,809 A		7/1999	Clarke	455/272
5,933,122 A		8/1999	Sauer et al.	343/876
5,991,643 A		11/1999	Chao-Cheng	455/562
6,021,317 A		2/2000	Irvin	455/78
6,034,638 A		3/2000	Thiel et al.	343/702
6,046,697 A		4/2000	Overbury et al.	342/360

(Continued)

FOREIGN PATENT DOCUMENTS

EP	0546803	6/1993

(Continued)

Primary Examiner—Michael C. Wimer
(74) *Attorney, Agent, or Firm*—Holland & Hart LLP

(57) **ABSTRACT**

An antenna device for transmitting and/or receiving electromagnetic radiation, installable in and connectable to a communication device, inlcudes an antenna structure switchable between a plurality of antenna configuration states. Each antenna configuration state is distinguished by a set of radiation parameters, such as resonance frequency, input impedance, bandwidth, radiation pattern, gain, polarization, and near-field pattern. Each of the plurality of antenna configuration states is adapted for use of the antenna device in the communication device in a respective predefined physical operation environment. A switching device selectively switches the antenna structure between the plurality of antenna configuration states.

25 Claims, 4 Drawing Sheets

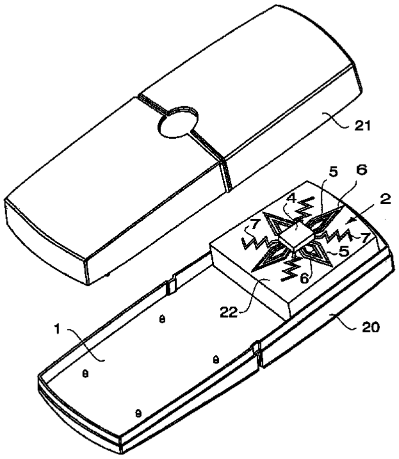

U.S. PATENT DOCUMENTS

6,118,409	A	9/2000	Pietsch et al.	343/703
6,297,780	B1	10/2001	Kirisawa	343/703
6,370,369	B1	4/2002	Kraiem et al.	455/277.1
6,392,610	B1 *	5/2002	Braun et al.	343/876

FOREIGN PATENT DOCUMENTS

EP	0852407	8/1998
EP	0840394	6/1999
GB	2327572	1/1999
GB	2332124	9/1999
JP	10145130	5/1998
JP	10209932	7/1998
WO	WO 94/28598	12/1994
WO	WO 99/44307	9/1999

* cited by examiner

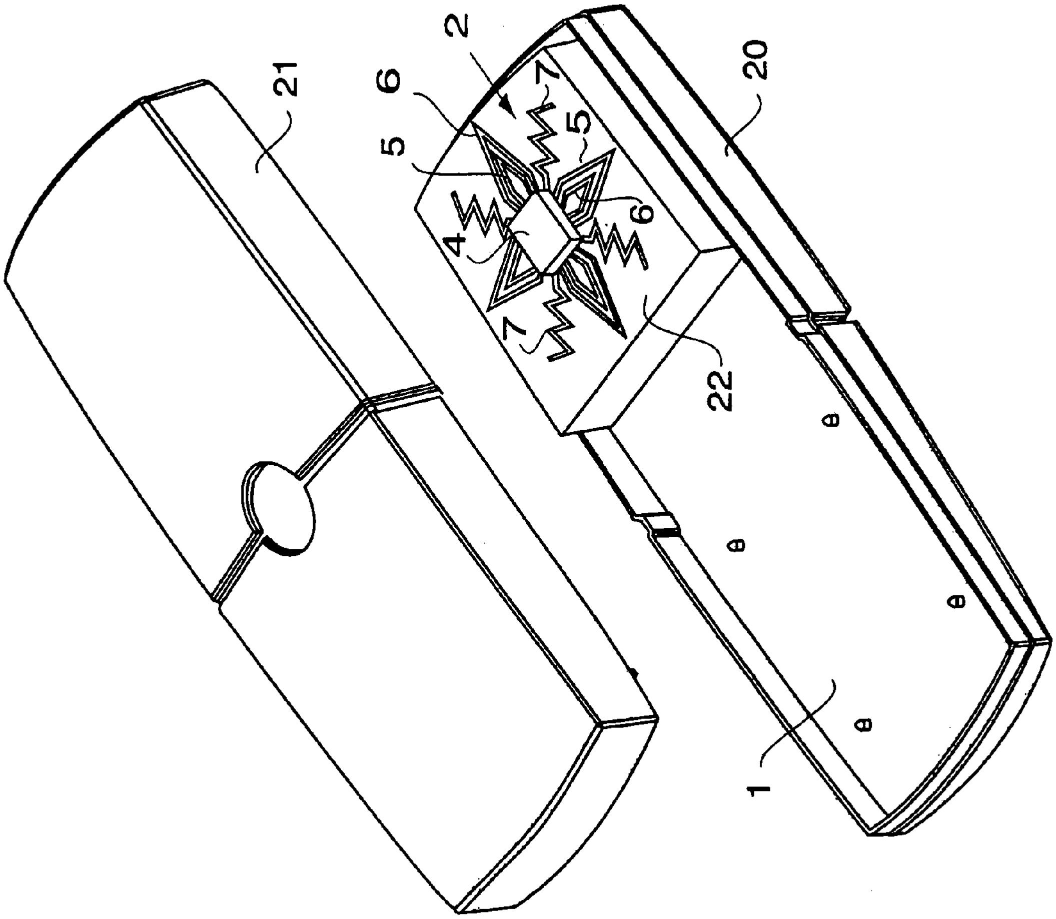
2
20
21
6
7
5
5
6
4
7
22
1

Fig. 1

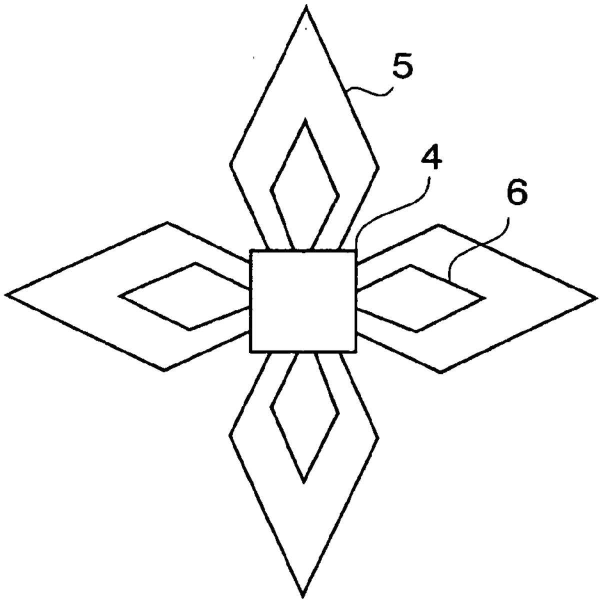

Fig. 2

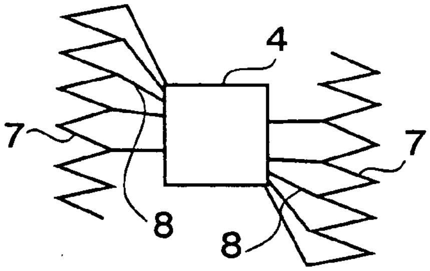

Fig. 3

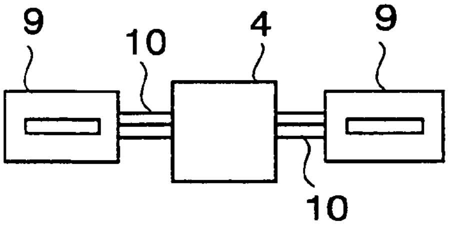

Fig. 4

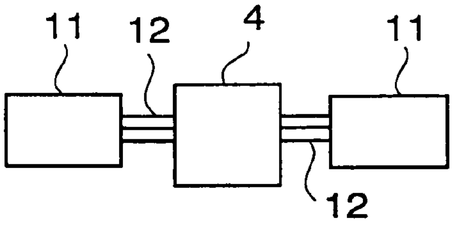

Fig. 5

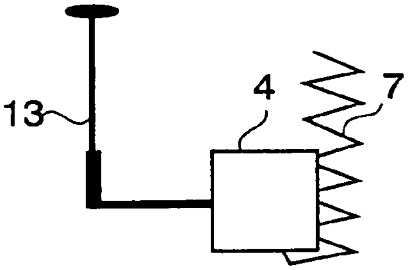

Fig. 6

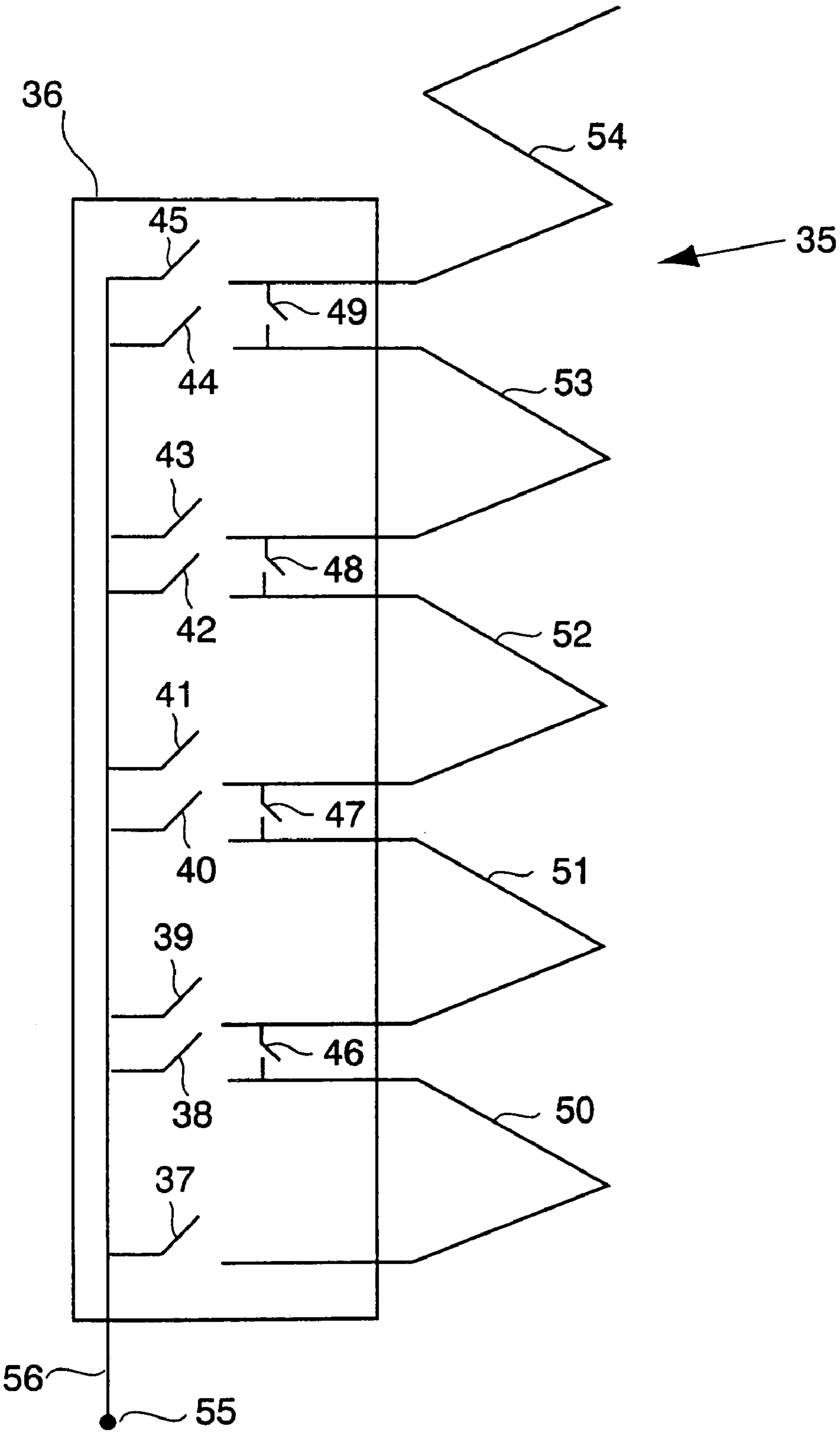
36
54
35
45
49
44
53
43
48
42
52
41
47
40
51
39
46
38
50
37
56
55

Fig. 7

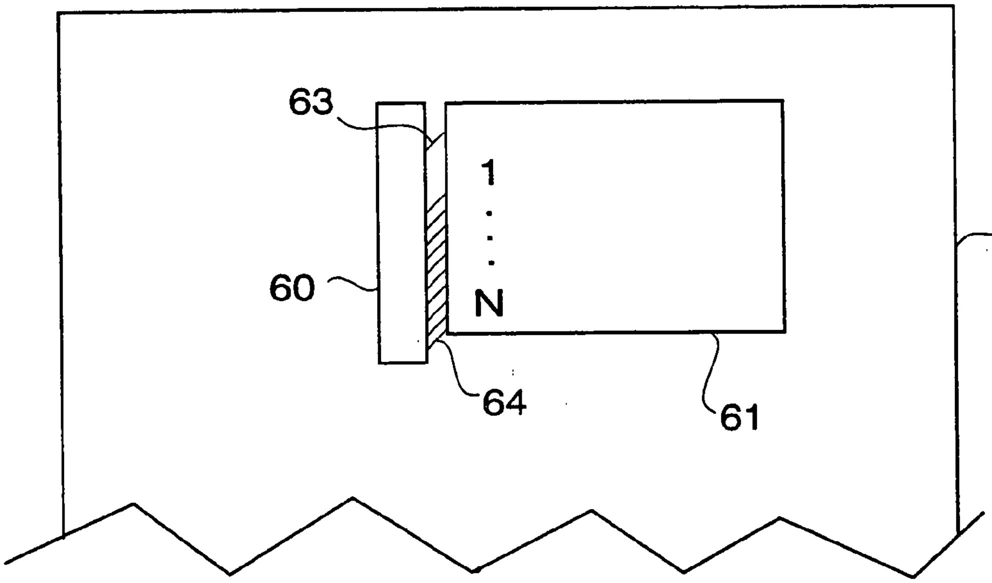
.
.
.
N

Fig. 8

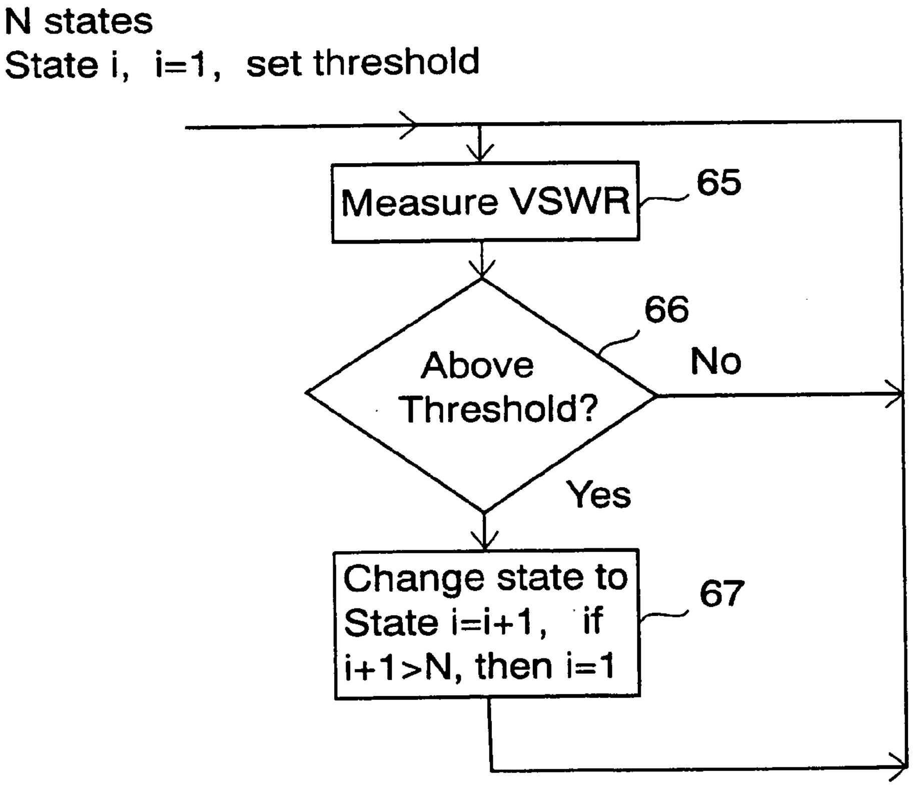
N states
State i, i=1, set threshold
Measure VSWR
65
66
Above Threshold?
No
Yes
Change state to State i=i+1, if i+1>N, then i=1
67

Fig. 9

ANTENNA DEVICE FOR TRANSMITTING AND/OR RECEIVING RADIO FREQUENCY WAVES AND METHOD RELATED THERETO

CROSS REFERENCE TO RELATED APPLICATIONS

The present invention claims priority to commonly assigned Swedish Patent Application Serial No. 9903945-5 filed Oct. 29, 1999 and is a continuation of PCT Patent Application Serial No. PCT/SE00/02059 filed on Oct. 24, 2000, the entire contents of all of which are hereby incorporated by reference in their entirety for all purposes. The present application is also related to commonly assigned, co-pending U.S. patent application Ser. No. 09/712,131 entitled "An antenna device for transmitting and/or receiving RF waves", now U.S. Pat. No. 6,392,610; U.S. patent application Ser. No. 09/712,181 entitled "Antenna device and method for transmitting and receiving radio waves"; and U.S. patent application Ser. No. 09/712,133 entitled "Antenna device and method for transmitting and receiving radio frequency waves", all of which were filed concurrently herewith. These applications are based on the following corresponding PCT applications: PCT/SE00/02058; PCT/SE00/02056; and PCT/SE00/02057, respectively, all filed on Oct. 24, 2000, the entire contents of which are hereby incorporated by reference in their entirety for all purposes.

TECHNICAL FIELD OF THE INVENTION

The present invention relates generally to an antenna device, a radio communication device including the antenna device, and a method for transmitting and receiving electromagnetic waves. More particularly, the present invention is related to an antenna device that is adaptable to a variety of conditions.

BACKGROUND OF THE INVENTION

In modern communication systems, there is an ever-increasing demand for smaller and more versatile portable wireless terminals, e.g., hand-portable telephones. It is well known that the size of an antenna is a critical factor for its performance. Further, the interaction between the antenna, the telephone body and the close-by environment, such as the user, will become more important as the wireless terminals become smaller and smaller. It is thus a formidable task to manufacture such compact and versatile terminals, which exhibit good antenna performance under a variety of conditions.

In current manufacturing of hand-portable telephones, the antenna is commonly adapted to the characteristics of the specific telephone and to be suited for a default use in a default environment. This means that the antenna cannot later on be adapted to any specific condition under which a certain telephone is to be used.

The radiating properties of an antenna device for a portable telephone depends heavily on the shape and size of the support structure such as a printed circuit board (PCB) of the telephone and of the telephone casing. All radiation properties, such as resonance frequency, radiation pattern, polarization, impedance and bandwidth are a product of the antenna device itself and its interaction with the PCB and the telephone casing. Thus, all references to radiation properties made below are intended to be for the whole device in which the antenna is incorporated.

What has been stated above is true also with respect to other communication devices, such as cordless telephones, telemetry systems, wireless data terminals, etc. Thus, the antenna device of the invention is applicable on a broad scale in various communication devices.

SUMMARY OF THE INVENTION

The present invention is therefore directed to an antenna device, a communication device including the antenna device and a method of receiving and transmitting electromagnetic waves that substantially overcomes one or more of the problems due to the limitations and disadvantages noted above.

It is an object of the present invention is to provide a versatile antenna device, adaptable to various conditions, for a communication device. In this respect, it is a particular object of the invention to provide a versatile antenna device, which is adaptable to its close-by environment.

It is yet a further object of the invention to provide an antenna device of which certain characteristics, such as resonance frequency, input impedance bandwidth, radiation pattern, gain, is polarization, and near-field pattern, are easily controllable.

It is still a further object to provide an antenna device that is simple, lightweight, easy to manufacture and inexpensive.

It is yet a further object to provide an antenna device which is efficient, easy to install and reliable, particularly mechanically durable, even after long use.

It is still a further object of the invention to provide an antenna device suited to be used as an integrated part of a communication device.

These and other objects of the present invention will become more readily apparent from the detailed description given hereinafter. However, it should be understood that the detailed description and specific examples, while indicating the preferred embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will become more fully understood from the detailed description of embodiments of the present invention given hereinbelow and the accompanying FIGS. **1–9**, which are given by way of illustration only, and thus are not limitative of the invention.

FIG. **1** is a perspective view of two casing parts of a portable telephone including one embodiment of an antenna device according to the present invention.

FIGS. **2–8** schematically illustrate additional embodiments of an antenna device according to the invention.

FIG. **9** is a flow diagram of an example of a switch-and-stay algorithm for controlling a switching device of an inventive antenna device.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

In the following description, for purposes of explanation and not limitation, specific details are set fourth in order to provide a thorough understanding of the present invention. However, it will be apparent to one skilled in the art that the present invention may be practiced in other embodiments that depart from these specific details. In other instances, detailed descriptions of well-known devices and methods are omitted so as not to obscure the description of the present invention with unnecessary details.

In this disclosure it is to be understood that the antenna system of the invention is operable to transmit or receive electromagnetic signals. Even if a term is used herein that suggests one specific signal direction it is to be appreciated that such a situation can cover that signal direction and/or its reverse. The expression "antenna structure" is intended to include active elements connected to the transmission (feed) line(s) of the communication device circuitry, as well as elements that can be grounded or left disconnected, and hence operate as, e.g., directors, reflectors, impedance matching elements.

According to the present invention, there is provided an antenna device for transmitting and/or receiving RF radiation, which is installable in and connectable to a radio communication device. The antenna device includes an antenna structure, which is switchable between a plurality of antenna configuration states. Each antenna configuration is distinguished by a set of radiation parameters, such as resonance frequency, input impedance, bandwidth, radiation pattern, gain, polarization, and near-field pattern. The antenna device also includes a switching device for selectively switching the antenna structure between the plurality of antenna configuration states.

In the antenna device, each of the antenna configuration states is adapted for use of the antenna device in the radio communication device in a respective predefined physical operation environment. These predefined physical operation environments may include a talk position (TP), a free space position (FS), a waist position (WP) and a pocket position (PP). A predefined physical operation environment as used herein preferably means a close-by environment, which includes objects that affect the above-mentioned parameters of the antenna structure, particularly when being installed in a small-sized radio communication device. Close-by operation environment as used herein preferably means any object at a distance from the radio communication device within which the effect on the antenna parameters is noticeable. This distance may extend ten wavelengths from the device, but optionally it may extend five wavelengths, a few wavelengths, or only roughly about one wavelength from the device. The environment includes, of course, the user of the communication device.

Furthermore, the present invention includes various approaches for sensing the physical operation environment and various procedures for controlling the switching of the antenna device.

The description is hereinbelow divided into five main sections covering various aspects of the present invention. The first section gives an overview of a manifold of different antenna structures and switching devices that may be employed in the present invention. Thereafter, a description of different physical operation environments is given. A discussion about radiation related parameters that may be affected by the different operation environments follows, and which parameter changes may be compensated for by switching to another antenna configuration state. The discussion focuses primarily on the parameters resonance frequency, impedance and radiation pattern and two specific examples are briefly overviewed. Subsequently, some approaches for sensing the physical operation environment are depicted, and, finally some procedures for controlling the switching of the antenna device are outlined.

Antenna Structures and Switching Devices

In FIG. **1** a front part **20** and the back part **21** of the casing of a portable telephone are shown. The main printed circuit board (PCB) of the phone is to be mounted in the space **1** in the front part **20** of the casing. An antenna device **2** of the present invention is printed on a separate supporting device **22** in this embodiment. The support can be a flexible substrate, a Molded Interconnection Device (MID) or a PCB. However, the antenna could have been printed on the main PCB, as well, which can extend along the length of the bottom casing. There are RF feed lines and control lines for the switching device between the phone circuit on the PCB and the antenna device.

The antenna device **2** includes a switching device **4**. The switching device **4** includes a matrix of electrically controllable switching elements. The switching elements can include microelectro-mechanical system switches (MEMS), PIN diode switches, or GaAs field effect transistors (FET).

The switching device **4** is surrounded by an antenna structure including a pattern of antenna elements. Each antenna element is connected to a respective switch in the switching device arranged for connecting and disconnecting the antenna element. In this embodiment, the radiating structure includes four loop-shaped antenna elements **5**. A loop-shaped parasitic element **6** is formed within each of the loop-shaped antenna elements **5**. A meander-shaped antenna element **7** is arranged between each pair of loop-shaped elements **5**, **6**. The antenna elements form a symmetrical pattern around the switching device **4**. However, in certain applications the antenna elements can form an unsymmetrical pattern. Further, the radiation structure can include additional antenna elements not connected to the switching device.

The switching device **4** allows the antenna structure **2** to be selectively switchable between a number of antenna configuration states, each of which is distinguished by a set of radiation parameters, such as resonance frequency, input impedance, bandwidth, radiation pattern, gain, polarization and near-field pattern. The various antenna configuration states may be obtained by connecting loop-shaped antenna elements in parallel or in series with each other, or some elements can be connected in series and some in parallel. Further, one or more elements can be completely disconnected or connected to a RF ground plane. One or more of the meander-shaped antenna elements **7** can be used separately or in any combination with the loop antenna elements. The meander-shaped antenna elements **7** can also be segmented so that only one or more selected portions thereof can be connected if desired.

Although not illustrated in FIG. **1**, other types of antenna elements, such as patch antennas, slot antennas, whip antennas, helical antennas, zigzag antennas and fractal antennas can also be used. In all cases, the switching device **4** may or may not be surrounded by the antenna elements and the antenna elements can also be positioned on one side of the switching device.

All switching of the antenna elements is centralized to the switching device **4**, which can be very small with a controllable interaction with the antenna function. Further, as all switching is centralized to the switching device **4**, switch control signals need only be supplied to the switching device **4**, thereby simplifying the overall antenna structure, as well as providing other advantages.

The connection/disconnection of the antenna elements are easily controllable by the switching device **4**. By appropriate selection of the combination of antenna elements that are connected to the RF feed, i.e., the antenna configuration state, the impedance and/or the resonance frequency of the antenna device can be adjusted without the need for separate connection or disconnection of discrete components. The same effect can be achieved by using parasitic elements, not connected to RF feed, but connected to RF ground or unconnected. The parasitic elements can also be connected to the switching device. If the to use of discrete components in any application is desired, these discrete components can be easily connected or disconnected by use of the same switching device **4** as the other antenna elements.

Further, the radiation pattern of the antenna can be shaped according to demand by appropriate selection of antenna elements. In this way the influence due to objects in the close-by environment of the antenna device, such as the user of a portable phone, can be minimized among other things. It will also be possible to control the tuning, polarization, bandwidth, resonance frequency, gain, input impedance of the antenna device. These above depicted radiation related parameters will be discussed in more detail further below.

Next, a few more antenna configurations will briefly be discussed with reference to FIGS. **2**–**6**.

FIG. **2** is an example of an antenna device including a plurality of loop antenna elements **5**, **6** as in FIG. **1**. The loop antenna elements are arranged so that they start and end at the switching device **4**. The switching device **4** can be used to connect the loop antenna elements to a RF feed line, short-circuited, coupled in series or in parallel with each other. Each loop antenna element can therefore be seen as a portion of the total antenna structure, from now on called "the total antenna", having properties determined by the state of the switching device **4**. That is, the switching device **4** decides how the loop element portions are connected and electrically arranged. At least some of the loop antenna elements **5** can act as an actively radiating element, where the excitation is achieved through direct connection to a RF feed, preferably via a waveguide. Possibly, some of the loop antenna elements **6** can act as parasitic elements, where the excitation of the elements is achieved through parasitic coupling to other antenna elements.

The loop antenna elements can be shaped as three-dimensional structures. Parts or all of the structure can be positioned above the PCB. The pattern can go around, or through the PCB, so that part of the pattern is on the other side of the PCB. Some or all parts of the pattern can extend perpendicular to the PCB. There can be permanent shorting pins and/or components attached to the antenna elements outside of the switching device. The feeding of the antenna elements can also take place outside of the switching device.

The purpose of changing the antenna configuration state can be to match the antenna to a desired impedance. This can be done by switching in/out parasitic elements. The mutual coupling between the elements contributes to the input impedance of the active element, changing the resulting input impedance in a desired manner. Another purpose for altering the antenna configuration is to change the radiation pattern of the total antenna. This can be done by altering the connection of antenna portions so that the radiating currents are altered. This can also be done by switching in/out parasitic elements, thereby directing or reflecting the radiation towards a desired direction.

FIG. **3** shows an example of the antenna device where two meandering antenna elements **7** are connected to the central switching device **4**. The expression "meandering" element is intended also to cover other elements with similar shape and function, such as zigzag shape, snake shape, fractal shape, etc. What has been stated above in connection with the loop antenna elements **5**, **6** in FIG. **2** is applicable also regarding the meander shaped elements of FIG. **3**, as is realized by the person skilled in this art. The only difference between the configurations is the inherent difference in radiation characteristics between these two types of antenna elements as is well known in the art.

In FIG. **3**, connection lines **8**, which provide the RF feed and/or RF ground points of the meander antenna element **7**, can be switched between different positions along the meander antenna element **7**. The aim of this can be to change the input impedance for matching purposes or to change the current flow for radiation pattern control.

FIG. **4** shows an example of an antenna device where slot antenna elements **9** are connected to the central switching device **4**. The slot antenna elements **9** are connected to the switching device **4** via connection lines **10**. The connection lines **10** can be connected directly to a RF feed device, shorted, coupled in series or in parallel with lines to other antenna elements. Each connection line **10** can act as an active feed line and be connected directly to a RF feed device. Parasitic coupling, where there is no direct connection to any RF feed, can also be used.

At least one slot element **9** of the antenna device is fed by at least one connection line **10**, and in various ways tuned by the other lines. For example, the other lines can be shorted or left open so that the slot antenna element **9**, and, in effect, the whole antenna device, is tuned for a desired frequency band. The same technique can be used to change the radiation pattern of the wireless terminal, to which the antenna device is coupled, pattern shaping. Moreover, connecting, disconnecting or tuning other slot elements can provide tuning or pattern shaping.

FIG. **5** shows an example of an antenna device similar to that of FIG. **4**, but where two patch antenna elements **11** are connected to the central switching device **4** via connection lines **12**. The patch antenna elements **11** are placed closed to or in connection to the central switching device. What has been stated above in connection with FIG. **4** is also relevant for the embodiment of FIG. **5**.

FIG. **6** shows an example of an antenna device where a meander element **7** is connected to the central switching device **4** together with a whip antenna element **13**. The whip elements **13** and meander elements **7** can be connected directly to a RF feed device, shorted or coupled in parallel/series. Each element can act as an active radiating element, i.e., be connected directly to a RF feed device, or as a parasitic element, where there is no galvanic connection to a RF feed device. For example, the electrical length of the whip element **13** and/or the meander element **7** can be altered to tune the resonance frequency. There can be other parasitic elements, not shown, close to the whips and/or the meander for tuning and/or for changing the radiation pattern. In this way the radiation pattern can be mainly directed towards a desired direction. The whip element can be replaced by or combined with a helical antenna element.

Of course, the antenna device can include a switching device and any combination of the above described antenna elements forming a symmetrical or an unsymmetrical pattern of radiating elements. The antenna device can be adapted for operation in several frequency bands and for receiving and transmitting radiation of different polarization. In addition, the switching device **4** can be used to connect or disconnect discrete matching components. The invention is not limited to any specific shape of the individual antenna elements as the shapes can be chosen according to the desired function.

Close-by Operation Environments

Next, various physical operation environments that may affect the performance of the antenna device in accordance with the invention will be described. The antenna parameters, such as resonance frequency, input impedance, bandwidth, radiation pattern, gain, polarization, and near-field pattern, of a small-sized wireless communication device are affected by objects in the close-by environment of the device.

A small-sized wireless communication device, such as a mobile telephone, can be used in many different close-by environments. It can be held to the ear as a telephone, it can be put in a pocket (with the front toward the user or the back toward the user), it can be attached to a belt at the waist, or it can be held in the hand or placed on an electrically conductive surface. Many more operation environments may be enumerated. Common for all environments is that there may be objects in the close-by environment of the device, thereby affecting the antenna parameters of the device. Environments with different objects in the proximity of the device have different influence on the antenna parameters.

Two specific operation environments will in the following be specifically discussed.

The free space (FS) operation environment is obtained by locating the radio communication device in empty space, i.e. with no objects in the close-by environment of the device. Air surrounding the device is here considered to be free space. Many operation environments can be approximated by the free space environment. Generally, if the environment has little influence on the antenna parameters, it can be referred to as free space.

The talk position (TP) operation environment is defined as the position, in which the radio communication device is held to the ear by a user. The influence on the antenna parameters varies depending on the person holding the device and on exactly how the device is positioned. Here, the TP environment is considered as a general case, i.e. covering all individual variations mentioned above.

Radiation Related Parameters

Various radiation related parameters that may be affected by the physical operation environment and controlled by an antenna device in accordance with the invention will now be described in more detail.

Antennas for wireless communication devices experience detuning due to the presence of the user. For many antenna types, the resonance frequency drops considerably when the user is present, compared to when the device is positioned in free space. An adaptive tuning between free space and talk position can reduce this problem largely.

A straightforward manner to tune an antenna is to alter its electrical length, and thereby altering the resonance frequency. The longer the electrical length, the lower the resonance frequency. This is also the most straightforward way to create band switching, if the change in electrical length is large enough.

FIG. **7** shows a meander-like antenna structure **35** arranged together with a central switching device **36** including a plurality of switches **37**–**49**. The antenna structure **35** may be seen as a plurality of aligned and individually connectable antenna elements **50**–**54**, which are connectable to a feed point **55** through the switching device **36** and a feed line **56**. Feed point **55** is further connected to a low noise amplifier of a receiver circuitry (not shown) of a communication device, and hence, the antenna structure **35** operates as a receiving antenna. Alternatively, the feed point **55** is connected to a power amplifier of a communication transmitter for receiving an RF power signal, and hence the antenna structure **35** operates as a transmitting antenna. Optionally, the antenna structure **35** may be arranged for both transmission and reception.

A typical example of operation is as follows. Assume that switches **37** and **46**–**49** are closed and remaining switches are opened and that such an antenna configuration state is adapted for optimal performance when the antenna device is arranged in a hand-portable telephone located in free space. When the telephone is moved to talk position, the resonance frequency will be lowered due to the user and thus, in order to compensate for the presence of the user, switch **49** is opened, whereby the electrical length of the connected antenna structure is reduced and accordingly the resonance frequency is increased. This increase, with an appropriate design of antenna structure **35** and switching device **36**, will compensate for the reduction as introduced when the telephone is moved from free space to talk position.

According to the invention, all switching of the antenna elements **50**–**54** required for different purposes is centralized to the switching device **36** which is provided with a single feed line. Instead of tuning a detuned antenna, one can perform adaptive impedance matching, which involves letting the resonance frequency be slightly shifted and compensate this detuning by impedance matching.

An antenna structure can be fed at different locations. Each location has a different ratio between the E and H fields, resulting in different input impedances. This phenomenon can be exploited by switching the feed point, provided that the feed point switching has little influence on the resonance frequency of the antenna. When the antenna experiences detuning due to the presence of the user (or other object), the antenna can be matched to the feed line impedance by altering for example the feed point of the antenna structure. In a similar manner, RF ground points can be altered.

In FIG. **8** is schematically shown an example of such an implementation of an antenna structure **61** that can be selectively RF grounded at a number of different points spaced apart from each other. Antenna structure **61** is in the illustrated case a planar inverted F antenna (PIFA) mounted on a printed circuit board **62** of a radio communication device. Antenna **61** has a feed line **63** and N different spaced RF ground connections **64**. By switching from one RF ground connection to another, the impedance is slightly altered. As before all switching functions are centralized to a central switching device **60**.

Moreover, switching in/out parasitic antenna elements can produce an impedance matching, since the mutual coupling from the parasitic antenna element to the active antenna element produces a mutual impedance, which contributes to the input impedance of the active antenna element. If outer limits for the detuning of the antenna elements can be found, the range of adaptive tuning/matching that needs to be covered by the antenna device can be estimated.

The radiation pattern of a wireless terminal is affected by the presence of a user or other object in its near-field area. Loss-introducing material will not only alter the radiation pattern, but also introduce loss in radiated power due to absorption. This problem can be reduced if the radiation pattern of the terminal is adaptively controlled. The near-field radiation pattern can be directed mainly away from the loss-introducing object, which will reduce the overall losses.

A change in radiation pattern requires the currents producing the electromagnetic radiation to be altered. Generally, for a small device (e.g. a hand-portable telephone), there need to be quite large changes in the antenna structure to produce altered currents, especially for the lower frequency bands. However, this can be done by switching to another antenna type producing different radiation pattern, or to another antenna structure at another position/side of the PCB of the communication device.

Another way may be to switch from an antenna structure that interacts heavily with the PCB of the communication device (e.g., whip or patch antenna) to another antenna not doing so (e.g., loop antenna). This will change the radiating currents dramatically since interaction with the PCB introduces large currents on the PCB (the PCB is used as a main radiating structure).

Sensing of the Physical Operation Environment

According to the present invention, a sensor may be provided for detecting a physical property of a selected close-by environment and a control device may be provided for controlling the switching device, and thus the selective switching of the antenna structure between the various antenna configuration states, in dependence on the detected physical property. The sensor would in the general case not be part of the antenna device, but be located at the surface of the wireless terminal casing. In such an instance, the response of the sensor is received at the antenna module control device.

The sensing of the close-by environment can be performed in several manners. One manner can be to use sensors on different positions at the device. In this manner, objects on different sides of the device may be sensed. The sensors can be of different kinds, e.g., resistive, capacitive or inductive sensors.

Capacitive (negative reactance) or inductive (positive reactance) sensors change their reactance when objects with electrical properties differing from those of free space are close to them. Hence, these sensors may distinguish objects that do not have a large effect on the electric performance of the antenna, e.g. cloth. Capacitive sensors are in general more sensitive to dielectric materials. Capacitive sensors can be found in, e.g., elevator buttons. Inductive sensors, on the other hand, are in general more sensitive to conducting materials. Inductive sensors are often used in the automation industry, for sensing end points of metallic goods.

Another sensor type may be a heat detector for sensing body heat. Optical sensors, e.g., photo detectors, can also be used to detect objects in the close-by environment. Still other sensors that may be employed include pressure, inclination, orientation, or motion sensors, which may detect motion patterns and from them deduce different usage scenarios. Pressure sensors may detect whether the communication device is held by a person and the manner in which it is held.

Also, a measure of the reflection coefficient as measured after the power amplifier of the transmitter can be used to "sense" objects, which cause detuning of the antenna. This is possible since objects with electrical properties, which in the near-field area, i.e., close-by environment, of the antenna differ from those of free space, will influence the antenna-input impedance.

Yet another manner, in which the environment the device is in, may be determined, is the usage state itself, i.e., if the device is used for speech, and no hands-free unit is in use, the antenna is optimized for talk position.

Procedures for Controlling the Switching

The invention will be exemplified below by an algorithm, which uses any suitable sensed parameter such as the reflection coefficient as an optimization parameter. In the following example, the voltage standing wave VSWR is used.

A simple and easily implemented algorithm is a switch-and-stay algorithm, which is shown in the flow diagram of FIG. **9**. Here switching is performed between predefined states i=1, . . . , N (e.g. N=2, one state being optimized for FS and the other state being optimized for TP). A state i=1 is initially chosen, whereafter, in a step **65**, the VSWR is measured. The measured VSWR is then, in a step **66**, compared with a predefined limit (the threshold value). If this threshold is not exceeded the algorithm is returned to step **65**. If this threshold is exceeded, the algorithm proceeds to step **67**, whereby switching to a new state i=i+1 is performed. If i+1 exceeds N, switching is performed to state 1. After this step, the algorithm returns to step **65**.

Using such an algorithm, each state **1**, . . . , N is used until the detected VSWR exceeds the predefined limit. When this occurs, the algorithm steps through the predefined states until a state is reached which has a VSWR below threshold. Both transmitter and receiver antenna structures can be switched at the same time.

Next, the invention will be exemplified by a procedure using a look-up table for determining which antenna configuration state to switch to.

The sensor senses the close-by environment of the radio communication device. Different type of sensors will give different images of the close-by environment. For example, if capacitive or inductive sensors are used at various locations on the device, one may be able to tell towards which direction (as seen from the device) there is least influence from close-by objects. The antenna configuration state is then chosen so as to direct the main radiation towards this direction.

To each set of responses from the sensor(s), there is an associated antenna configuration state that preferably minimizes the influence of the objects, minimizing loss and maximizing radiated power. This can be implemented in the form of a look-up table.

A trial-and-error algorithm works only if an antenna related parameter is measured, for example the reflection coefficient. All sensors not sensing specific antenna parameters could use for example a look-up table type of procedure as described above. Of course, combinations can be used. One example of this is that non-antenna-related parameter sensors are used in order to find the antenna configuration state that works in the best manner possible. A trial-and-error algorithm may optionally fine-tune the antenna configuration state afterwards.

It will be obvious that the invention may be varied in a plurality of ways. Such variations are not to be regarded as a departure from the scope of the invention. All such modifications as would be obvious to one skilled in the art are intended to be included within the scope of the appended claims.

What is claimed is:

1. An antenna device provided for transmitting and/or receiving RF radiation, installable in and connectable to a portable radio communication terminal device, and comprising:

an antenna structure comprising a plurality of antenna elements capable of being connected to and disconnected from each other, said antenna structure being switchable between a plurality of predefined antenna configuration states, in each of which multiple ones of said plurality of antenna elements are connected to each other, and each of which being distinguished by a set of radiation parameters; and

a switching device provided for selectively switching said antenna structure between said plurality of predefined antenna configuration states,

wherein each of said plurality of predefined antenna configuration states is optimized for operation of the antenna device in said portable radio communication terminal device in a respective predefined physical operation environment external to said portable radio communication terminal device,

wherein said plurality of predefined antenna configuration states includes at least three antenna-configuration states, and said switching device is provided for receiving a measure indicating a change from a first one to a second one of said predefined physical operation environments and for switching said antenna structure from a first one to a second one of said plurality of predefined antenna configuration states, in dependence on said received measure, wherein said first predefined antenna configuration state is optimized for use of said antenna device in said portable radio communication terminal device in said first predefined physical operation environment, and said second predetermined antenna configuration state is optimized for use of said antenna device in said portable radio communication terminal device in said second predefined physical operation environment.

2. The antenna device of claim **1**, wherein said plurality of antenna elements includes at least three antenna elements.

3. The antenna device of claim **1**, wherein said plurality of antenna elements includes at least five antenna elements.

4. The antenna device of claim **1**, wherein said plurality of predefined antenna configuration states includes at least five antenna configuration states.

5. The antenna device of claim **1**, wherein at least some of said plurality of predefined antenna configuration states are optimized for operation of the antenna device in said portable radio communication terminal device in different frequency bands.

6. The antenna device of claim **1**, wherein said antenna structure in at least some of said plurality of predefined antenna configuration states have different electrical lengths.

7. The antenna device of claim **1**, wherein said switching device comprises a plurality of switches for connecting and disconnecting said plurality of antenna elements, and wherein said plurality of antenna elements are arranged one after another.

8. The antenna device of claim **1**, wherein said antenna structure in at least some of said plurality of predefined antenna configuration states have different feed positions and/or different ground positions.

9. The antenna device of claim **8**, wherein said at least some of said plurality of predefined antenna configuration states are identical except for the different feed positions and/or different ground positions.

10. The antenna device of claim **9**, wherein said antenna structure comprises a planar or patch antenna and a plurality of feed connectors and/or ground connectors.

11. The antenna device of claim **1**, wherein said antenna structure comprises two separate antennas of different type.

12. The antenna device of claim **11**, wherein said two separate antennas are a whip antenna or a patch antenna, and a loop antenna or a meander antenna.

13. The antenna device as claimed in claim **1**, wherein said switching device comprises a micro electromechanical system (MEMS) switch device.

14. The antenna device of claim **1**, wherein each of said respective predefined physical operation environments is defined by objects affecting RF radiation and located within a distance from said portable radio communication terminal device of less than ten wavelengths of said RF radiation.

15. The antenna device of claim **1**, wherein one of said plurality of predefined antenna configuration states is optimized for use of said antenna device in said portable radio communication terminal device in a position held to an ear of a user as a telephone, and one of said plurality of predefined antenna configuration states is optimized for use of said antenna device in said portable radio communication terminal device in a position held at a waist of the user.

16. The antenna device of claim **1**, wherein said switching device is connectable to a sensor device, which is capable of sensing and identifying each of said respective predefined physical operation environments external to said portable radio communication terminal device, and which is provided for sending said measure indicating a change from a first one to a second one of said predefined physical operation environments to said switching device.

17. The antenna device of claim **16**, wherein said antenna device comprises said sensor device.

18. The antenna device of claim **16**, wherein said sensor device is capable of sensing and identifying each of said respective predefined physical operation environments external to said portable radio communication terminal device by means of sensing resistance, capacitance, inductance, light, temperature, and/or pressure external to said portable radio communication terminal device, or inclination, orientation and/or motion of said portable radio communication terminal device.

19. The antenna device of claim **16**, wherein said sensor device is capable of sensing objects at two opposite sides of said portable radio communication terminal device simultaneously.

20. The antenna device of claim **1**, wherein said antenna device is further provided with an adaptive fine-tuning device capable of controlling switching by said switching device to find an optimum one of a plurality of fine-tuning antenna configuration states, where said plurality of fine-tuning antenna configuration states are variants of said second predefined antenna configuration state.

21. An antenna device provided for transmitting and/or receiving RF radiation, installable in and connectable to a portable radio communication terminal device, and comprising:

an antenna structure comprising a plurality of antenna elements capable of being connected to and disconnected from each other, said antenna structure being switchable between at least three predefined antenna configuration states, each of which being distinguished by a set of radiation parameters, and each of which being associated with a different operation environment external to said portable radio communication terminal device;

a switching device capable of selectively switching said antenna structure between said at least three predefined antenna configuration states; and

a control device provided for receiving a sensed resistance, capacitance, inductance, light, temperature, and/or pressure value external to said portable radio communication terminal device, for determining one of said different operation environments external to said portable radio communication terminal device depending on said sensed resistance, capacitance, inductance, light, temperature, and/or pressure value, and for controlling said switching device to switch said antenna structure to said one of said different operation environments determined depending on said sensed resistance, capacitance, inductance, light, temperature, and/or pressure value.

22. An antenna device provided for transmitting and/or receiving RF radiation, installable in and connectable to a portable radio communication terminal device, and comprising:

an antenna structure comprising a plurality of antenna elements capable of being connected to and disconnected from each other, said antenna structure being switchable between at least three predefined antenna configuration states, each of which being distinguished by a set of radiation parameters, and each of which being associated with a different operation environment external to said portable radio communication terminal device;

a switching device capable of selectively switching said antenna structure between said at least three predefined antenna configuration states; and

a control device provided for receiving a measure of inclination, orientation and/or motion of said portable radio communication terminal device, for determining one of said different operation environments external to said portable radio communication terminal device depending on said measure of inclination, orientation and/or motion of said portable radio communication terminal device, and for controlling said switching device to switch said antenna structure to said one of said different operation environments determined depending on said measure of inclination, orientation and/or motion of said portable radio communication terminal device.

23. In an antenna device installable in and connectable to a portable radio communication terminal device, and comprising an antenna structure comprising a plurality of antenna elements capable of being connected to and disconnected from each other, said antenna structure being switchable between at least three predefined antenna configuration states, each of which being distinguished by a set of radiation parameters, and each of which being associated with a different operation environment external to said portable radio communication terminal device, a method for transmitting and/or receiving RF radiation comprising:

receiving a detected physical property external of said portable radio communication terminal device;

determining one of said different operation environments external to said portable radio communication terminal device depending on said detected physical property;

controlling switching of said antenna structure to said one of said different operation environments determined depending on said detected physical property; and

transmitting and/or receiving RF radiation.

24. In an antenna device installable in and connectable to a portable radio communication terminal device, and comprising an antenna structure comprising a plurality of antenna elements capable of being connected to and disconnected from each other, said antenna structure being switchable between at least three predefined antenna configuration states, each of which being distinguished by a set of radiation parameters, and each of which being associated with a different operation environment external to said portable radio communication terminal device, a method for transmitting and/or receiving RF radiation comprising:

receiving a sensed resistance, capacitance, inductance, light, temperature, and/or pressure value external of said portable radio communication terminal device;

determining one of said different operation environments external to said portable radio communication terminal device depending on said sensed resistance, capacitance, inductance, light, temperature, and/or pressure value;

controlling switching of said antenna structure to said one of said different operation environments determined depending on said sensed resistance, capacitance, inductance, light, temperature, and/or pressure value; and

transmitting and/or receiving RF radiation.

25. In an antenna device installable in and connectable to a portable radio communication terminal device, and comprising an antenna structure comprising a plurality of antenna elements capable of being connected to and disconnected from each other, said antenna structure being switchable between at least three predefined antenna configuration states, each of which being distinguished by a set of radiation parameters, and each of which being associated with a different operation environment external to said portable radio communication terminal device, a method for transmitting and/or receiving RF radiation comprising:

receiving a measure of inclination, orientation and/or motion of said portable radio communication terminal device;

determining one of said different operation environments external of said portable radio communication terminal device depending on said measure of inclination, orientation and/or motion of said portable radio communication terminal device;

controlling switching of said antenna structure to said one of said different operation environments determined depending on said measure of inclination, orientation and/or motion of said portable radio communication terminal device; and

transmitting and/or receiving RF radiation.

* * * * *

US 20120013296A1

(19) **United States**
(12) **Patent Application Publication**
Heydari et al.
(10) **Pub. No.: US 2012/0013296 A1**
(43) **Pub. Date: Jan. 19, 2012**

(54) METHOD AND SYSTEM FOR HARVESTING RF SIGNALS AND WIRELESSLY CHARGING A DEVICE

(76) Inventors: **Soudeh Heydari**, Dubai (AE); **Mohammad Asefi**, Dubai (AE)

(21) Appl. No.: **13/181,448**

(22) Filed: **Jul. 12, 2011**

Related U.S. Application Data

(60) Provisional application No. 61/364,731, filed on Jul. 15, 2010, provisional application No. 61/364,739, filed on Jul. 15, 2010.

Publication Classification

(51) **Int. Cl.**
H02J 7/00 (2006.01)
H02J 17/00 (2006.01)

(52) **U.S. Cl.** **320/108**; 307/104

(57) ABSTRACT

System and method for harvesting radio frequency power from a wireless transmitter includes; an antenna having an output voltage, a matching network coupled to the antenna, a booster circuit coupled to the matching network and having an output voltage, and adapted to amplify the output voltage of the antenna. Matching network is adapted to match a power transfer. A voltage stabilizer circuit is adapted to stabilize the booster circuit output voltage. Method includes, determining at least one operating mode and charging a slave device from a master wireless transmitter located within a predetermined distance from the system. A rechargeable battery in the system is charged from a wireless transmitter, and a slave device is connected to the system. The slave device is charged from the rechargeable battery in the system.

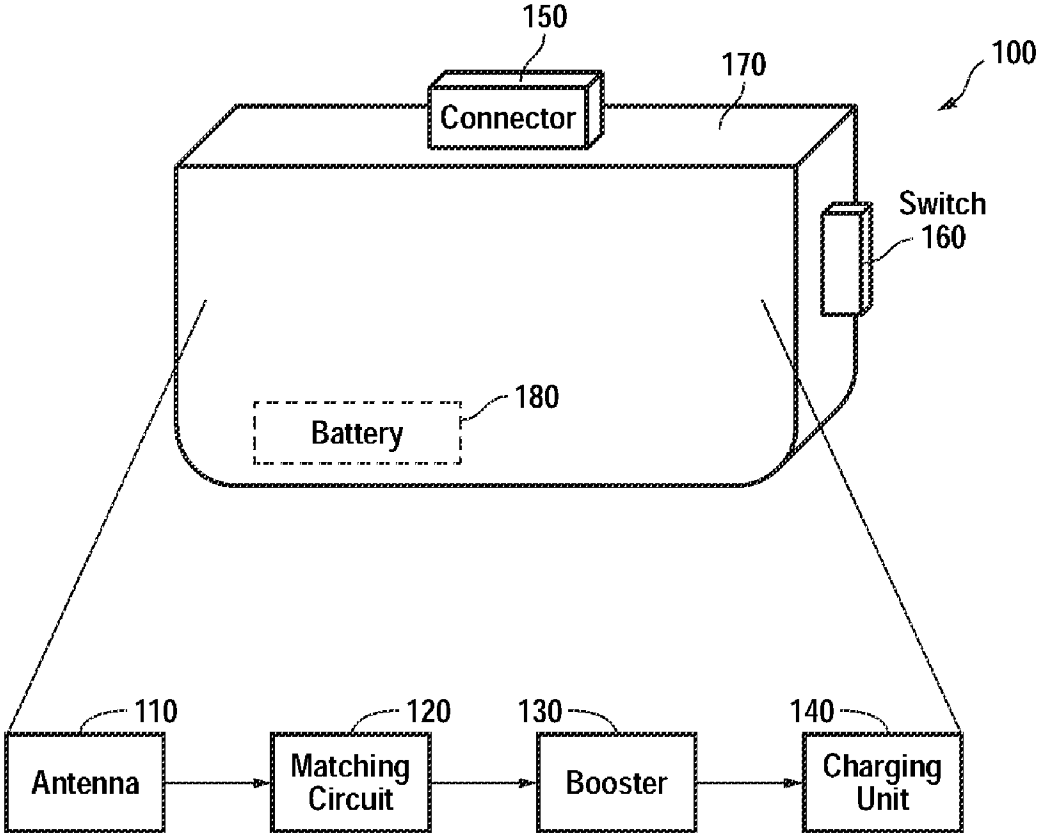

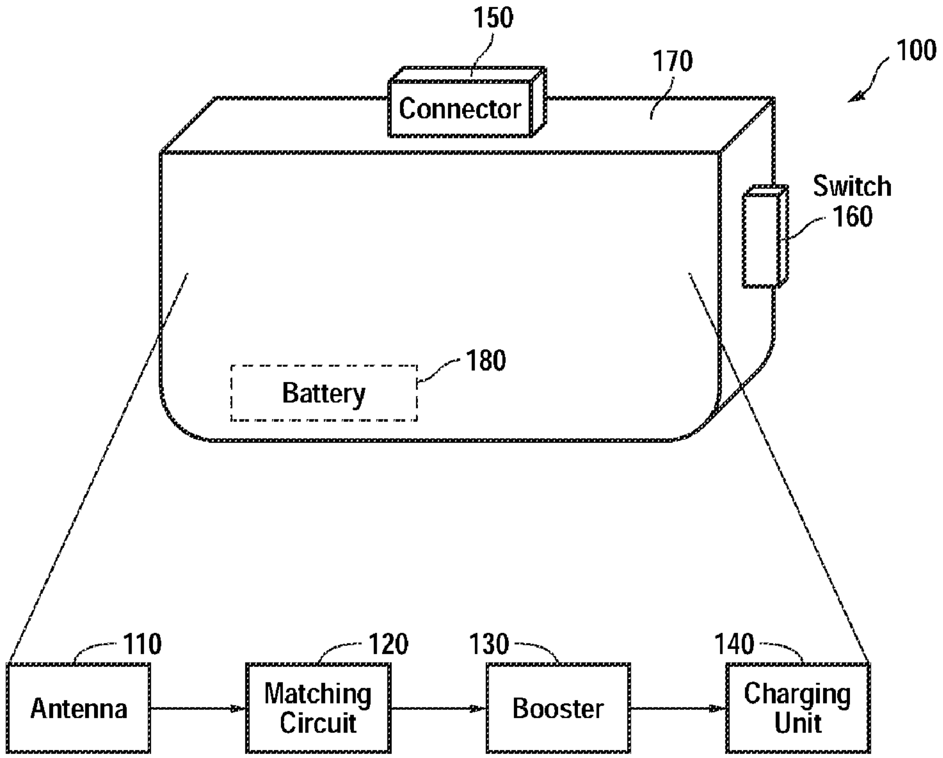
150
Connector
170
100
Switch
160
180
Battery
110
Antenna
120
Matching Circuit
130
Booster
140
Charging Unit

FIG. 1

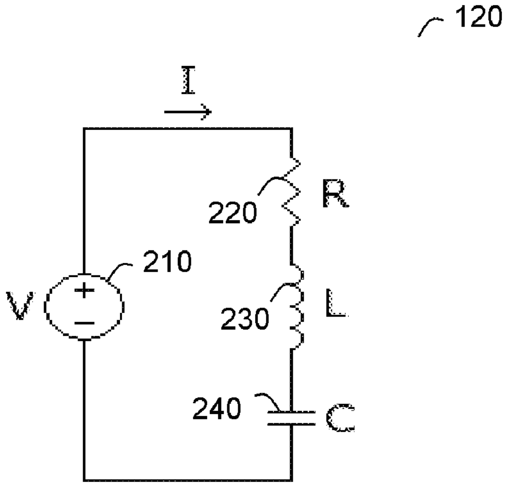
120
I
R
220
210
V
+
−
230
L
240
C

FIG. 2

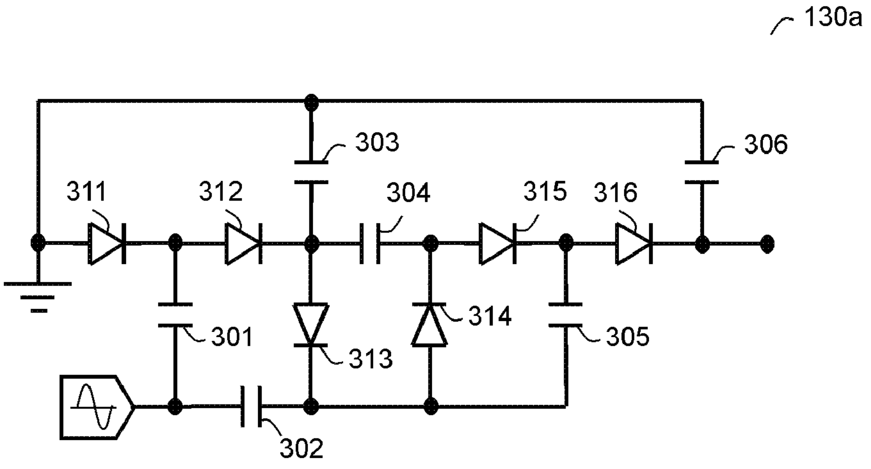
130a
303
306
311
312
304
315
316
314
301
313
305
302

FIG. 3

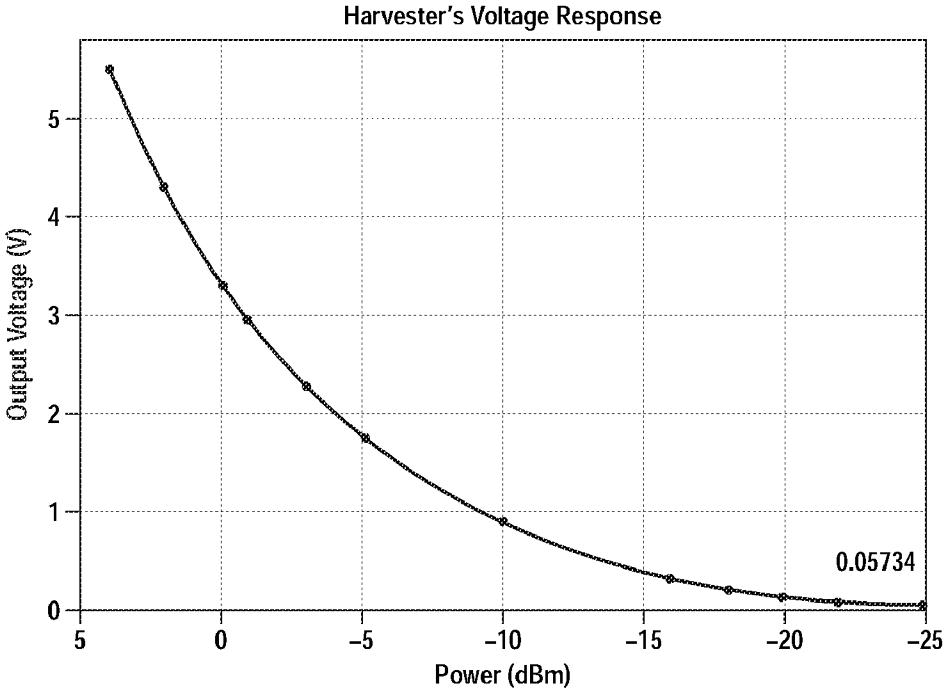
Harvester's Voltage Response
Output Voltage (V)
5
4
3
2
1
0
0.05734
5
0
−5
−10
−15
−20
−25
Power (dBm)

FIG. 4

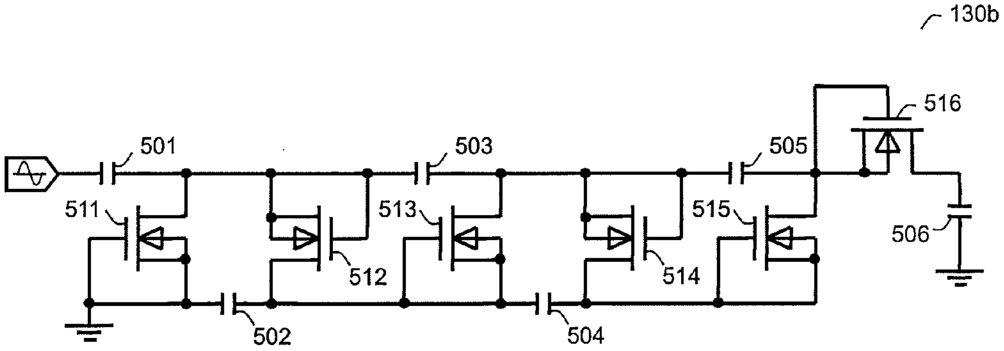
130b
501
503
505
516
511
513
515
506
512
514
502
504

FIG. 5

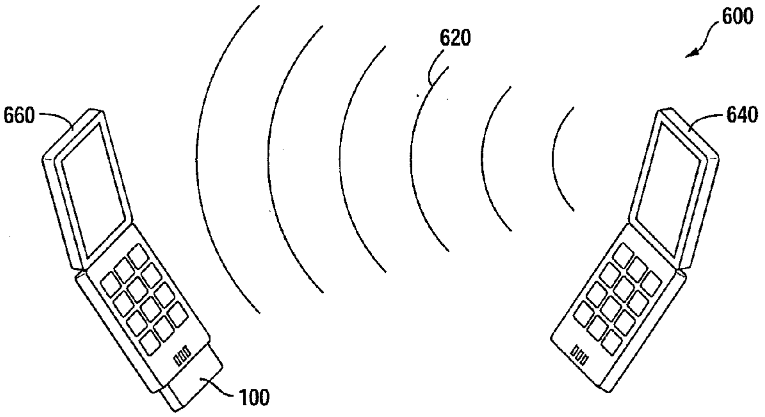
600
620
660
640
100

FIG. 6

METHOD AND SYSTEM FOR HARVESTING RF SIGNALS AND WIRELESSLY CHARGING A DEVICE

CROSS-REFERENCES TO RELATED APPLICATIONS

[0001] The present application claims benefit under 35 USC 119(e) of commonly assigned U.S. provisional application No. 61/364,731, filed Jul. 15, 2010, entitled "Method and System For Harvesting RF Signals and Wirelessly Charging A Device," and U.S. provisional application No. 61/364,739, filed Jul. 15, 2010, entitled "Wireless Battery Charger," the contents of all of which are incorporated herein by reference in their entirety.

BACKGROUND

[0002] The present invention relates generally to supplying power to portable devices. More specifically, some embodiments of the present invention relate to harvesting radio frequency (RF) signals to power portable devices.

[0003] In the world of nanotechnology and wireless communication and networking, power supply of electrical devices still is a matter of concern. Many scientists and researchers are working on various methods of production, transmission and storing energy. Portable devices powered by rechargeable batteries commonly need chargers or base stations that are connected by wire from a power grid. Such portable devices include cell-phones, personal digital assistants, smart-phones, tablets, and remote controls, among others. Some of these devices rely on radio frequency signals for communication purposes.

[0004] The existing wireless chargers have operation distance limitations which prevents them from being used in locations without access to the wired power grid. This means that the device to be charged has to be in locations where a power outlet exists so that the charger can be connected to a power outlet and then the device to be charged must be placed on it (or kept within a short distance from it).

[0005] RF signals in the air have very limited power. Thus, the output power of an RF signal harvester is in the range of micro and milli-watts. Yet, various low power electronic devices can be empowered with this power.

BRIEF SUMMARY

[0006] The present invention relates generally to supplying power to portable devices. More specifically, some embodiments of the present invention relate to harvesting radio frequency (RF) power to power portable devices.

[0007] According to one embodiment of the present invention, a system for harvesting radio frequency power from a wireless transmitter includes; (i) an antenna having an output voltage, (ii) a matching network coupled to the antenna, (iii) a booster circuit coupled to the matching network and having an output voltage, and adapted to amplify the output voltage of the antenna. The matching network is adapted to match a power transfer from the antenna to the booster circuit. The system for harvesting radio frequency power from the wireless transmitter further includes a voltage stabilizer circuit adapted to stabilize the output voltage of the booster circuit.

[0008] According to one specific embodiment, the system further includes a user selectable switch adapted to determine at least one operating mode, and a connector adapted to charge a slave device. At least one operating mode is adapted to charge the slave device from a master wireless transmitter located within a predetermined distance from the system.

[0009] According to another specific embodiment, the system further includes: (i) a user selectable switch adapted to determine at least one operating mode, (ii) a rechargeable battery adapted to be charged by the booster circuit, and (iii) a connector adapted to charge a slave device. The at least one operating mode is adapted to charge the rechargeable battery from a wireless transmitter located within a predetermined distance from the system and the rechargeable battery is adapted to charge the slave device.

[0010] According to another specific embodiment, the antenna is a wideband antenna. According to another specific embodiment, the antenna is a wideband circularly polarized micro-strip patch antenna.

[0011] According to another specific embodiment, the matching network is a resonant circuit tuned at a predetermined frequency. According to another specific embodiment, the booster circuit includes a multi-stage voltage multiplier circuit.

[0012] According to another specific embodiment, the system further includes a software executable on the master wireless transmitter and adapted to provide a user selectable amount of radio frequency power from the master wireless transmitter to the slave device.

[0013] According to another specific embodiment, the master wireless transmitter is adapted to transmit a user selectable file. The duration of the transmission corresponds to the amount of radio frequency power transmitted from the master wireless transmitter to the slave device.

[0014] According to another specific embodiment, the master wireless transmitter is chosen from one of the group of a cell-phone, a laptop, or a personal computer. According to another specific embodiment, the master wireless transmitter is adapted to transmit radio frequency power chosen from of one of the group of a Bluetooth signal, or a Wi-Fi signal.

[0015] According to another specific embodiment, the predetermined distance is about 1 meter. According to another specific embodiment, the predetermined distance is about 500 meters.

[0016] According to another specific embodiment, the wireless transmitter is adapted to transmit radio frequency power for the system. According to another specific embodiment, the wireless transmitter is adapted to transmit radio frequency power for the system in the form of one of the group of a GSM signal, or a Wi-Fi signal. According to another specific embodiment, the system is adapted to receive a signal having a power of at least −25 dBm.

[0017] According to one embodiment of the present invention, a method for harvesting radio frequency power from a wireless transmitter by a system including an antenna having an output voltage, a matching network coupled to the antenna, a booster circuit coupled to the matching network and having an output voltage, and a voltage stabilizer circuit, the method includes, matching a power transfer from the antenna to the booster circuit, amplifying the output voltage of the antenna, and stabilizing the output voltage of the booster circuit.

[0018] According to one specific embodiment, the method further includes determining at least one operating mode. According to another specific embodiment, the method further includes charging a slave device from a master wireless transmitter located within a predetermined distance from the system.

[0019] According to another specific embodiment, the method further includes charging a rechargeable battery in the system from a wireless transmitter located within a predetermined distance from the system, and connecting a slave device to the system. The slave device is charged from the rechargeable battery in the system.

[0020] A better understanding of the nature and advantages of the embodiments of the present invention may be gained with reference to the following detailed description and the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] FIG. **1** is a simplified block diagram showing a system for harvesting radio frequency power from a wireless transmitter in accordance with an embodiment of the present invention;

[0022] FIG. **2** is a simplified schematic diagram showing a matching circuit in accordance with an embodiment of the present invention;

[0023] FIG. **3** is a simplified schematic diagram showing a booster circuit in accordance with a first embodiment of the present invention;

[0024] FIG. **4** is an example of an output voltage response with respect to input RF power for the embodiment shown in FIG. **1**;

[0025] FIG. **5** is a simplified schematic diagram showing a booster circuit in accordance with a second embodiment of the present invention;

[0026] FIG. **6** is a simplified block diagram showing a method for harvesting radio frequency power from a wireless transmitter in accordance with an embodiment of the present invention;

DETAILED DESCRIPTION

[0027] According to the embodiments of the present invention, a system and method for harvesting radio frequency power from a wireless transmitter are described. The system may include a wireless charger that reuses the wasted power of RF signals in the environment. Although the output power of an RF signal harvester is in the range of micro and milliwatts, various low power electronic devices can be empowered with this low power. Some examples of such low power electronic devices may include: calculators, wireless chargers, low power electric devices (sensors, Remote Telemetry Units (RTUs) and the like), low power home appliances (remote controls, watches and the like), biomedical applications (empowering artificial organs such as implanted hearts), and robotics (robotic cells). In such applications, the RF signal harvester can replace batteries. In some applications, a high power transmitter may be placed in the field to provide the RF signals for the RF signal harvester. Using this system, one can charge up his/her portable device enough for a short operation time. For instance, if a cell-phone is to be charged, using the proposed method, the cell-phone will have enough battery power for a few calls.

[0028] Table 1 shows a comparison between an existing RF signal harvester and an RF signal harvester in accordance with embodiments of the present invention. Both of the harvesters were tested under the same conditions (i.e. normal movement, outdoor, and indoor environments). The approximated price of the prototype may be reduced by mass production.

TABLE 1

Harvester	Size (cm)	Output Voltage (mV)	Cost ($)	Special considerations
Commercialized Piezoelectric	4 × 7.5	0.25-0.4	100	Voltage is based on tested measurements
Proposed RF Signal Harvester (prototype)	2 × 3.5	0.15-0.5	11.90	Cost is approximated based on prototype assembly components

[0029] FIG. **1** is a simplified block diagram showing a system **100** for harvesting radio frequency power from a wireless transmitter in accordance with an embodiment of the present invention. System **100** may include an antenna **110**, a matching circuit **120**, a booster circuit **130** a charging unit **140**, a connector **150**, a switch **160**, and a housing **170**. Matching circuit **120** is coupled between antenna **110** and booster circuit **130**. Booster circuit **130** is coupled to charging unit **140**. The antenna **110**, matching circuit **120**, booster circuit **130**, and charging unit **140** may be contained in housing **170**, which may support connector **150** and switch **160**. Alternatively, antenna **110**, matching circuit **120**, booster circuit **130**, and charging unit **140** may be embedded in a low power wireless device. Connector **150** may include either a custom designed connector for each low power electronic device application or may include a general plug and a set of adaptor connectors to couple each low power electronic device application with the general plug. Switch may be user selectable and adapted to determine at least one operating mode to be described below. In one operating mode system **100** may further include an optional rechargeable battery **180**, which is coupled to charging unit **140** in housing **170**. In each operating mode, the switch either couples charging unit **140** to battery **180** and battery **180** to connector **150**, or couples charging unit **140** directly to connector **150**.

[0030] Antenna **110** may be a wideband circularly polarized micro-strip patch antenna. The polarization is selected to be circular so that the antenna can receive signals of all polarizations. Moreover, the antenna is wideband so that a wider range of signals can be harvested. Antenna **110** has an output voltage coupled to matching circuit **120**.

[0031] FIG. **2** is a simplified schematic diagram showing matching circuit **120** in accordance with an embodiment of the present invention. Matching circuit **120** may be tuned at a predetermined frequency, which, depending on the system of wireless communication, can be a Bluetooth frequency of 2.4 GHz or 5 GHz, or a GSM signal of frequencies from 900 MHz to 2 GHz) so that the mismatch losses are reduced and maximum power is delivered to booster circuit **130**. Matching circuit **120** may include a resonant circuit, which can help reduce noise and interference to other operating wireless systems. This noise and interference reduction is achieved by minimization of the reflected back signals due to mismatch. The matching circuit **120** may include a series RLC circuit **220**, **230**, **240** respectively. The design of matching circuit **120** is frequency dependent and may be designed in a way that power transfer losses are minimized.

[0032] Booster circuit **130** may either include diodes and capacitors or MOSFET transistors and capacitors to boost the output voltage of antenna **110** to voltage levels acceptable for charging a low power electronic device or battery **180**. FIG. **3** is a simplified schematic diagram showing a booster circuit **130***a* in accordance with a first embodiment of the present invention, which includes a multi-stage voltage multiplier circuit such as, for example, a six stage voltage multiplier circuit designed with capacitors **301**-**306** and schottky diodes **311**-**316**. Booster circuit **130***a* is designed in a way that changes in the value of capacitors does not affect the output of the booster significantly. In one example, capacitors **301**-**306** are each 1 pF.

[0033] In booster circuit **130***a*, the RF signal wave is rectified at its negative half of the cycle by C**301** and D**311**, and in its positive half by C**303** and D**312**. During the positive half-cycle, the voltage stored in C**301** from the negative half-cycle is transferred to C**303**, which causes the voltage of C**303** to be roughly two times the peak voltage of the RF source and is why this 2-stage portion of booster circuit **130***a* may be called a voltage doubler. The doubled voltage level is minus the turn-on voltage of the diode. Each stage of voltage multiplier contains one diode and one capacitor and operates in the manner described above. The number of stages determines the amount of multiplication (alternatively referred to herein as “amplification”).

[0034] FIG. **4** is an example of an output voltage response with respect to input RF power for the embodiment shown in FIG. **1**. The power of Bluetooth signals may not exceed 0 dBm so the higher powers are provided for lab measurement purposes. As can be observed from FIG. **4**, a voltage of 3 volts can be achieved using a Bluetooth signal to charge a battery or a low power device. As long as a potential deference across the battery is maintained, the current will flow through it, resulting in charging.

[0035] The booster circuit, which is a voltage multiplier, is used to amplify the output voltage of the antenna. Yet, the current gets reduced since the electric power follows the law of conservation of energy. Thus, the booster or the voltage amplifier may be designed in such a way that the power loss through the booster circuit is minimized. To achieve this goal, the booster circuit may include MOSFET transistors. FIG. **5** is a simplified schematic diagram showing a booster circuit **130***b*, in accordance with a second embodiment of the present invention, which includes a multi-stage voltage multiplier circuit such as, for example, a six stage voltage multiplier circuit designed with capacitors **501**-**506** and n-channel MOSFET transistors **511**-**516**. In this example, NMOS transistors are used, however, other types of transistors such as PMOS, and different combinations of PMOS and NMOS transistors may also be used.

[0036] Booster circuit **130***b* works on the same basic principles as booster circuit **130***a*. Booster circuit **130***b* is similar to, but not exactly the same as, booster circuit **130***a*. The difference with booster circuit **130***b* is that the output is more stabilized compared to booster circuit **130***b* since the turn-on voltage of the MOSFETs in booster circuit **130***b* is smaller than the turn-on voltage of the diodes included in booster circuit **130***a*. Thus, each multiplier stage in booster circuit **130***b* “pumps” current with less voltage loss resulting in better circuit power efficiency and the device technology may use standard CMOS technology, without requiring schottky diodes as in booster circuit **130***a*.

[0037] The output signal of booster circuit **130** is a DC signal, which can be used to charge a battery or empower a circuit. However, this DC signal should be stabilized. Thus, a charger unit may include a voltage stabilizer, usually a voltage regulator circuit, and an input-output port to interface between system **100** and the low power device to be charged. Currently, the size of the prototype is 5 cm by 0.4021 cm, which may be further reduced for production.

[0038] System **100** may include two different modes of operation in accordance with an embodiment of the present invention. FIG. **6** is a simplified block diagram showing a method **600** for harvesting a radio frequency power **620** from a master wireless transmitter **640** in accordance with an embodiment of the present invention, by using a first mode to charge a slave device **660** with system **100** coupled to slave device **660**. Master wireless transmitter **640** may include a device with wireless communication capabilities such as a cell-phone, laptop, PC, or any other device with wireless communication capability such as a base-station, router, and the like. The packet may be transmitted either via Wi-Fi or Bluetooth.

[0039] A packet is transmitted from master wireless transmitter **640** to system **100**. The packet may either be any file such as a photo, music, or the like type of file, from master wireless transmitter **640** or a specific file from a software designed for system **100**. The software provides a file for transmission at a predefined frequency of either 980 MHz or 2.4 GHz. The transmission duration of the file, which determines the amount of transferred power, can be selected by the user through the software interface. If an existing file is to be transmitted, its size determines the amount of the transmitted power. Slave device **660** may be a low power device with its own slave battery, such as a receiver phone, which charges by harvesting the energy of the transferred power packet received through system **100**. System **100** may or may not have embedded battery **180**. The longer the time of the connection, the higher the amount of the power that can be harvested by the slave device. The transmitting and receiving devices should be within a range of 1 meter from each other. After a certain time, the slave device charge reaches a point where it cannot be charged any further by system **100**.

[0040] In the second mode of operation, the system **100** gathers energy from the electromagnetic signals existing in its surroundings and charges a rechargeable battery **180** embedded in system **100**. After rechargeable battery **180** is charged, system **100**, including battery **180**, is coupled to the slave device, which then can be charged from rechargeable battery **180** in system **100**. By storing the energy obtained from these signals in a small rechargeable battery embedded in the system, enough power will always be available for emergency cases where no master device is readily available.

[0041] The RF signals existing in the environment can be either from a transmitter designed for system **100** or from the existing GSM, Wi-Fi, or the like signals in the air. As long as the existing signals have a minimum power of −35 dBm, they can be harvested by system **100**. This may sometimes give a user the freedom of being as far as 500 (or even more) meters from a GSM base station and still charge rechargeable battery **180** in system **100**. The limitation of this mode is that rechargeable battery **180** may not hold much power and may require frequent recharging, which may take extended time for recharging.

[0042] Embodiments of the present invention provides a system and method for harvesting radio frequency power from a wireless transmitter. The embodiments of the present invention are not limited by the type of transistor, PMOS, NMOS or otherwise, used to boost the radio frequency power. The embodiments of the present invention are not limited by the number of voltage multiplier stages used to boost the radio frequency power. It will be apparent to those with skill in the art that modifications to the above methods and apparatuses may occur without deviating from the scope of the present invention. Accordingly, the disclosures and descriptions herein are intended to be illustrative, but not limiting, of the scope of the invention, which is set forth in the following claims along with their full scope of equivalents.

What is claimed is:

1. A system for harvesting radio frequency power from a wireless transmitter comprising:

an antenna having an output voltage;

a matching network coupled to the antenna;

a booster circuit coupled to the matching network and having an output voltage, and adapted to amplify the output voltage of the antenna wherein the matching network is adapted to match a power transfer from the antenna to the booster circuit; and

a voltage stabilizer circuit adapted to stabilize the output voltage of the booster circuit.

2. The system of claim **1** further comprising:

a user selectable switch adapted to determine at least one operating mode; and

a connector adapted to charge a slave device wherein the at least one operating mode is adapted to charge the slave device from a master wireless transmitter located within a predetermined distance from the system.

3. The system of claim **1** further comprising:

a user selectable switch adapted to determine at least one operating mode;

a rechargeable battery adapted to be charged by the booster circuit; and

a connector adapted to charge a slave device wherein the at least one operating mode is adapted to charge the rechargeable battery from a wireless transmitter located within a predetermined distance from the system and the rechargeable battery is adapted to charge the slave device.

4. The system of claim **1** wherein the antenna is a wideband antenna.

5. The system of claim **1** wherein the antenna is a wideband circularly polarized micro-strip patch antenna.

6. The system of claim **1** wherein the matching network is a resonant circuit tuned at a predetermined frequency.

7. The system of claim **1** wherein the booster circuit includes a multi-stage voltage multiplier circuit.

8. The system of claim **2** further comprising a software executable on the master wireless transmitter and adapted to provide a user selectable amount of radio frequency power from the master wireless transmitter to the slave device.

9. The system of claim **2** wherein the master wireless transmitter is adapted to transmit a user selectable file wherein the duration of the transmission corresponds to the amount of radio frequency power transmitted from the master wireless transmitter to the slave device.

10. The system of claim **2** wherein the master wireless transmitter is chosen from one of the group of a cell-phone, a laptop, or a personal computer.

11. The system of claim **2** wherein the master wireless transmitter is adapted to transmit radio frequency power chosen from of one of the group of a Bluetooth signal, or a Wi-Fi signal.

12. The system of claim **2** wherein the predetermined distance is about 1 meter.

13. The system of claim **3** wherein the predetermined distance is about 500 meters.

14. The system of claim **3** wherein the wireless transmitter is adapted to transmit radio frequency power for the system.

15. The system of claim **3** wherein the wireless transmitter is adapted to transmit radio frequency power for the system in the form of one of the group of a GSM signal, or a Wi-Fi signal.

16. The system of claim **3** wherein the system is adapted to receive a signal having a power of at least −25 dBm.

17. A method for harvesting radio frequency power from a wireless transmitter by a system including an antenna having an output voltage, a matching network coupled to the antenna, a booster circuit coupled to the matching network and having an output voltage, and a voltage stabilizer circuit, the method comprising;

matching a power transfer from the antenna to the booster circuit;

amplifying the output voltage of the antenna; and

stabilizing the output voltage of the booster circuit.

18. The method of claim **17**, wherein the method further comprises determining at least one operating mode.

19. The method of claim **18**, wherein the method further comprises charging a slave device from a master wireless transmitter located within a predetermined distance from the system.

20. The method of claim **18**, wherein the method further comprises:

charging a rechargeable battery in the system from a wireless transmitter located within a predetermined distance from the system; and

connecting a slave device to the system wherein the slave device is charged from the rechargeable battery in the system.

* * * * *

US007400253B2

(12) United States Patent
Cohen

(10) **Patent No.: US 7,400,253 B2**
(45) **Date of Patent: Jul. 15, 2008**

(54) **HARVESTING AMBIENT RADIO FREQUENCY ELECTROMAGNETIC ENERGY FOR POWERING WIRELESS ELECTRONIC DEVICES, SENSORS AND SENSOR NETWORKS AND APPLICATIONS THEREOF**

(75) Inventor: **Marc H. Cohen**, Silver Spring, MD (US)

(73) Assignee: **MHCMOS, LLC**, Silver Spring, MD (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 239 days.

(21) Appl. No.: **11/459,459**

(22) Filed: **Jul. 24, 2006**

(65) **Prior Publication Data**

US 2007/0109121 A1 May 17, 2007

Related U.S. Application Data

(60) Provisional application No. 60/705,636, filed on Aug. 4, 2005.

(51) **Int. Cl.**
G08B 13/14 (2006.01)

(52) **U.S. Cl.** **340/572.1**; 340/572.3; 340/572.4; 340/539.1; 340/825.69; 340/825.72; 340/10.1; 235/383; 235/385; 320/108; 320/109

(58) **Field of Classification Search** 340/572.1, 340/572.3, 572.4, 539.1, 635, 657, 825.69, 340/825.72, 10.1; 235/383, 385; 320/108, 320/109; 455/422.1
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,127,799	A	10/2000	Krishnan	
6,798,716	B1	9/2004	Charych	
6,972,682	B2 *	12/2005	Lareau et al.	340/568.1
7,084,605	B2	8/2006	Mickle et al.	
7,348,875	B2 *	3/2008	Hughes et al.	340/10.4
2003/0199778	A1	10/2003	Mickle et al.	
2004/0113790	A1 *	6/2004	Hamel et al.	340/572.1
2004/0259604	A1	12/2004	Mickle et al.	

* cited by examiner

Primary Examiner—Hung T. Nguyen
(74) *Attorney, Agent, or Firm*—Mohammad S. Rahman, Esq.; Gibb & Rahman, LLC

(57) **ABSTRACT**

A system and device for harvesting various frequencies and polarizations of ambient radio frequency (RF) electromagnetic (EM) energy for making a passive sensor (tag) into an autonomous passive sensor (tag) adapted to collect and store data with time-stamping and some primitive computation when necessary even when an interrogating radio frequency identification (RFID) reader is not present (not transmitting). A specific source of ambient RF EM energy may include wireless fidelity (WiFi) and/or cellular telephone base stations. The system and device may also allow for the recharging of energy storage units in active and battery assisted passive (BAP) devices. The system could be a "smart building" that uses passive sensors with RF EM energy harvesting capability to sense environmental variables, security breaches, as well as information from "smart appliances" that can be used for a variety of controls and can be accessed locally or remotely over the Internet or cellular networks.

35 Claims, 15 Drawing Sheets

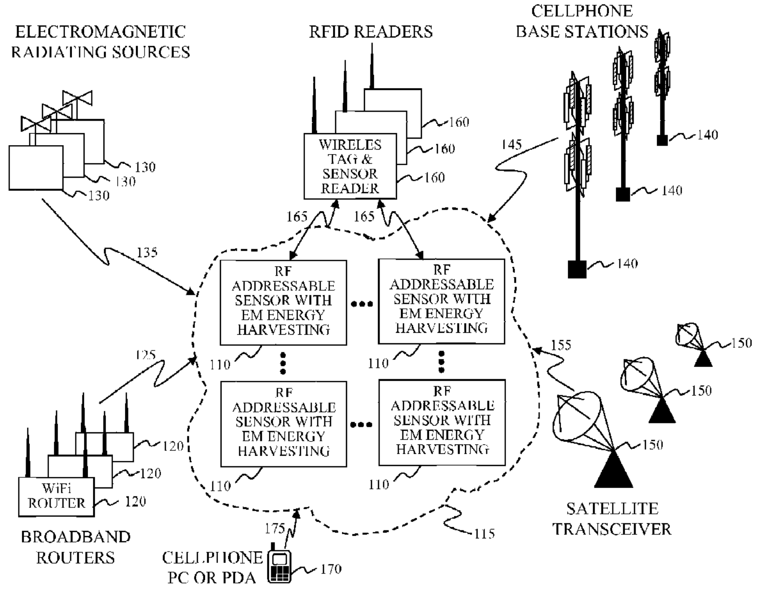
ELECTROMAGNETIC RADIATING SOURCES
RFID READERS
CELLPHONE BASE STATIONS
130
130
130
160
160
160
WIRELES TAG & SENSOR READER
145
140
140
140
135
165
165
RF ADDRESSABLE SENSOR WITH EM ENERGY HARVESTING
RF ADDRESSABLE SENSOR WITH EM ENERGY HARVESTING
125
110
110
155
150
150
150
RF ADDRESSABLE SENSOR WITH EM ENERGY HARVESTING
RF ADDRESSABLE SENSOR WITH EM ENERGY HARVESTING
120
120
120
WiFi ROUTER
110
110
BROADBAND ROUTERS
175
115
SATELLITE TRANSCEIVER
CELLPHONE PC OR PDA
170

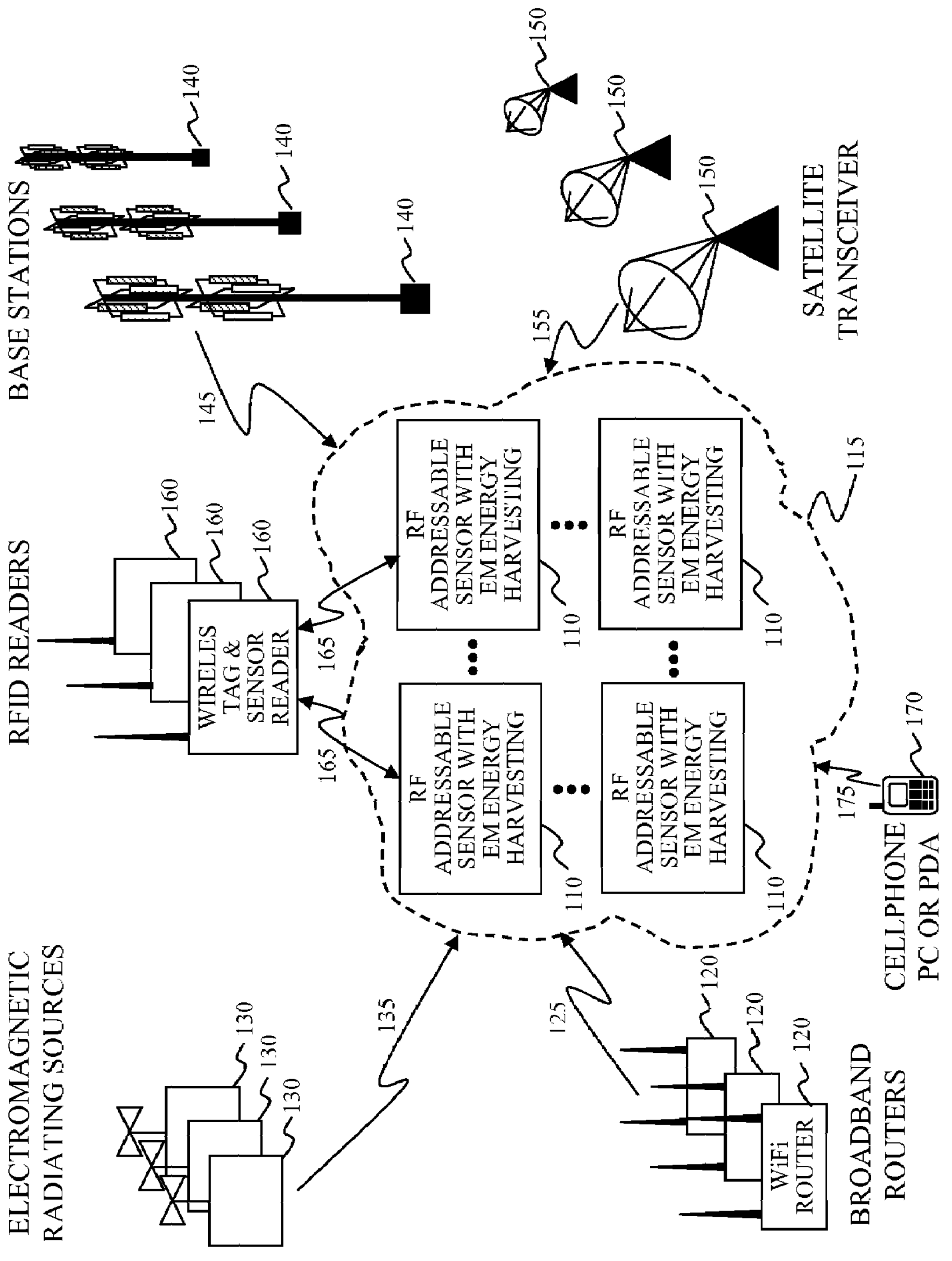
CELLPHONE
BASE STATIONS
140
140
140
150
150
150
SATELLITE
TRANSCEIVER
155
145
RFID READERS
160
160
160
WIRELES
TAG &
SENSOR
READER
165
165
RF
ADDRESSABLE
SENSOR WITH
EM ENERGY
HARVESTING
110
RF
ADDRESSABLE
SENSOR WITH
EM ENERGY
HARVESTING
110
RF
ADDRESSABLE
SENSOR WITH
EM ENERGY
HARVESTING
110
RF
ADDRESSABLE
SENSOR WITH
EM ENERGY
HARVESTING
110
115
170
175
CELLPHONE
PC OR PDA
ELECTROMAGNETIC
RADIATING SOURCES
130
130
130
135
125
120
120
120
WiFi
ROUTER
BROADBAND
ROUTERS

FIG. 1

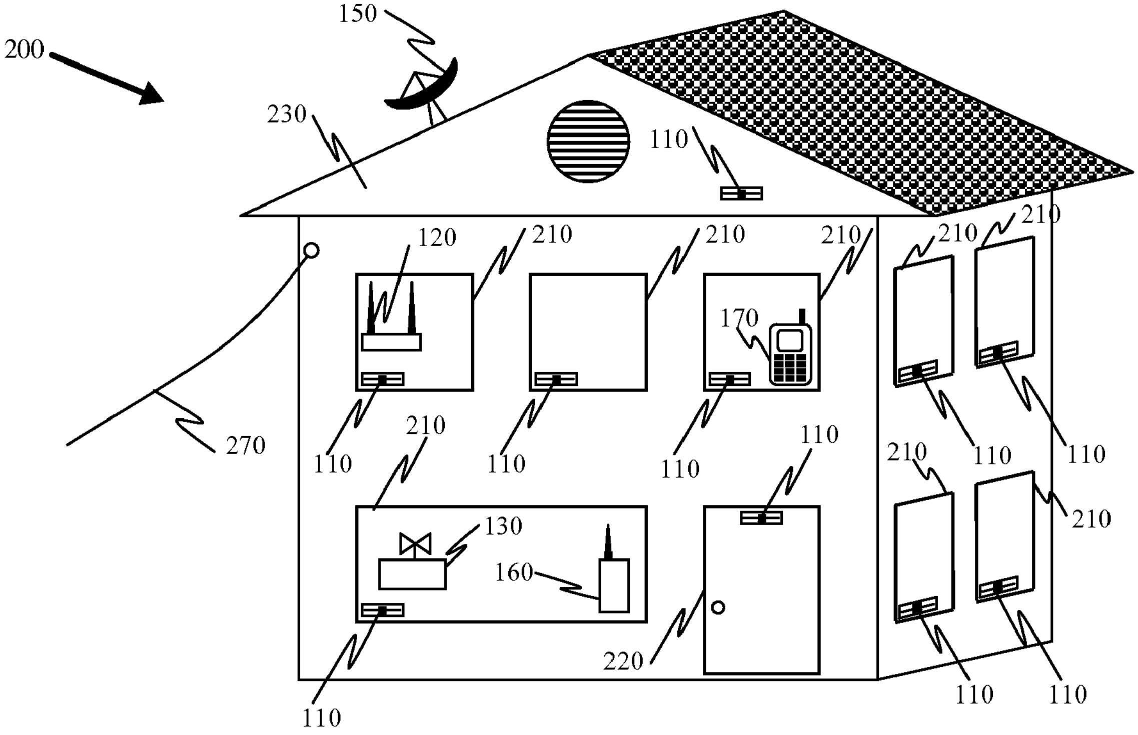

FIG. 2

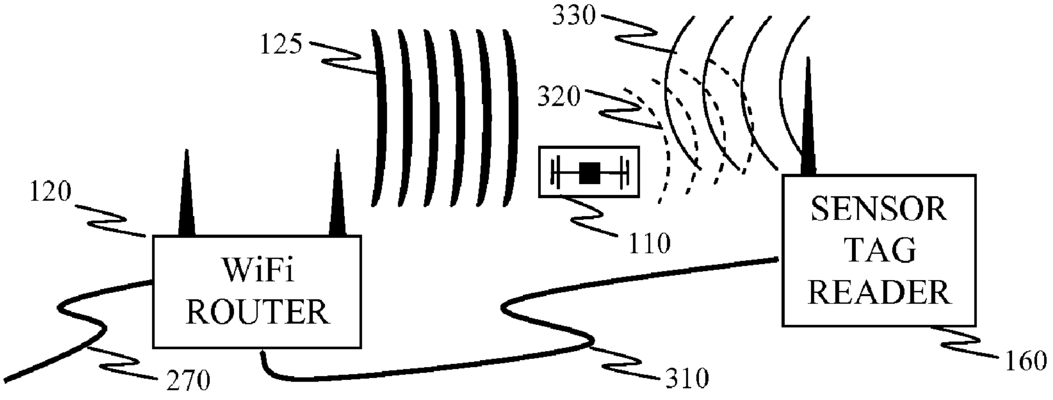
330
125
320
120
110
WiFi
ROUTER
SENSOR
TAG
READER
270
310
160

FIG. 3A

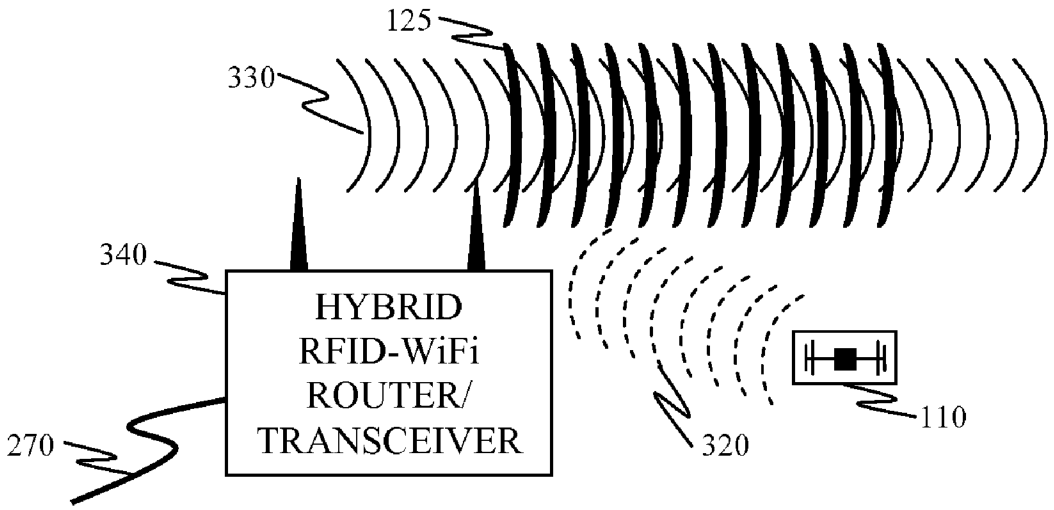
125
330
340
HYBRID
RFID-WiFi
ROUTER/
TRANSCEIVER
270
320
110

FIG. 3B

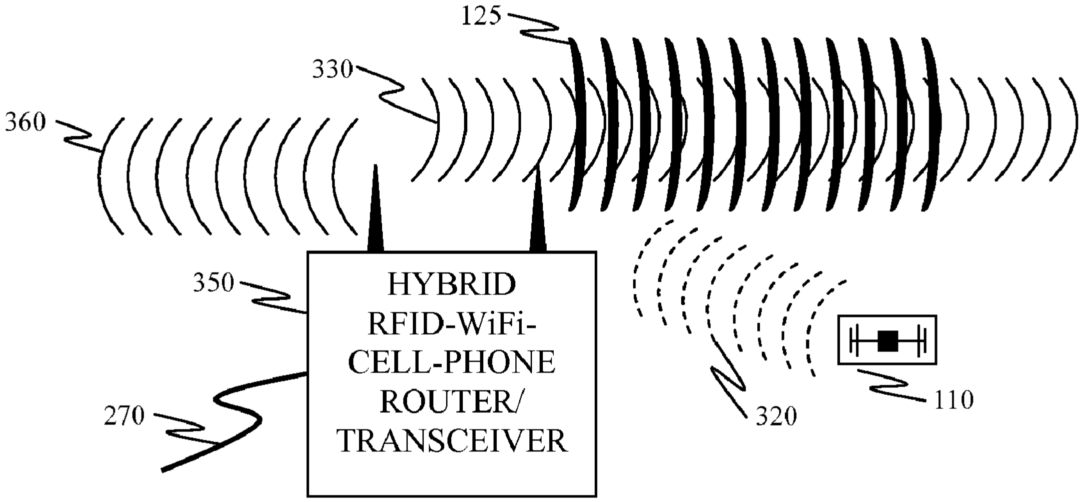
125
330
360
350
HYBRID
RFID-WiFi-
CELL-PHONE
ROUTER/
TRANSCEIVER
270
320
110

FIG. 3C

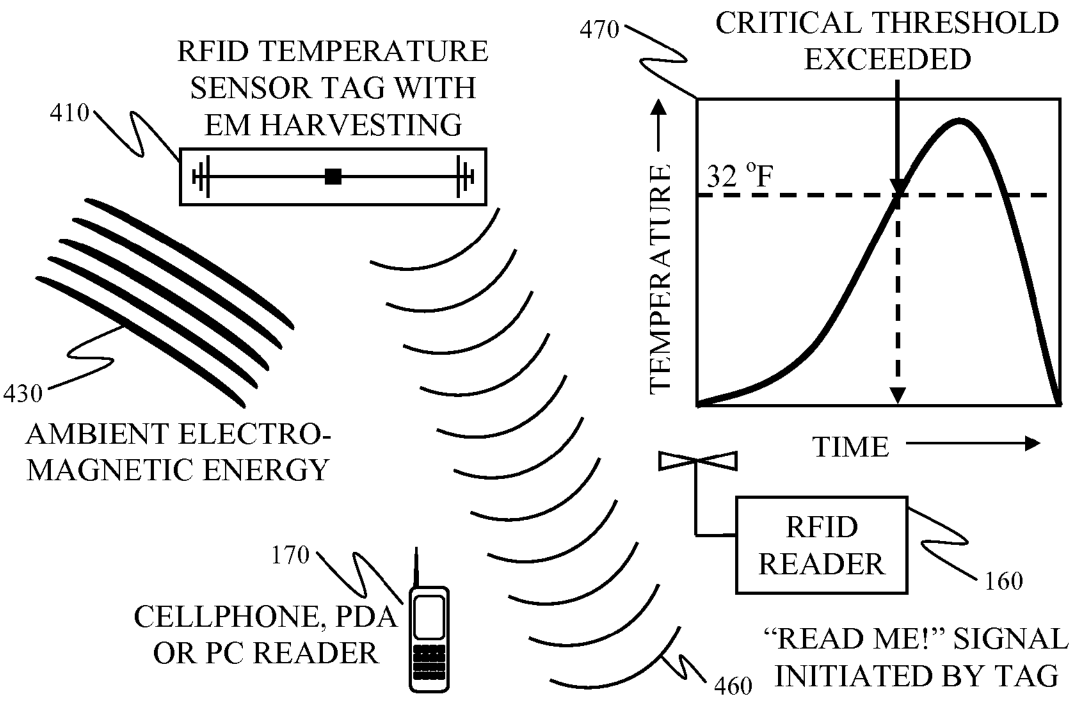
470
CRITICAL THRESHOLD
EXCEEDED
RFID TEMPERATURE
SENSOR TAG WITH
EM HARVESTING
410
TEMPERATURE
32 °F
430
AMBIENT ELECTRO-
MAGNETIC ENERGY
TIME
RFID
READER
160
170
CELLPHONE, PDA
OR PC READER
"READ ME!" SIGNAL
INITIATED BY TAG
460

FIG. 4A

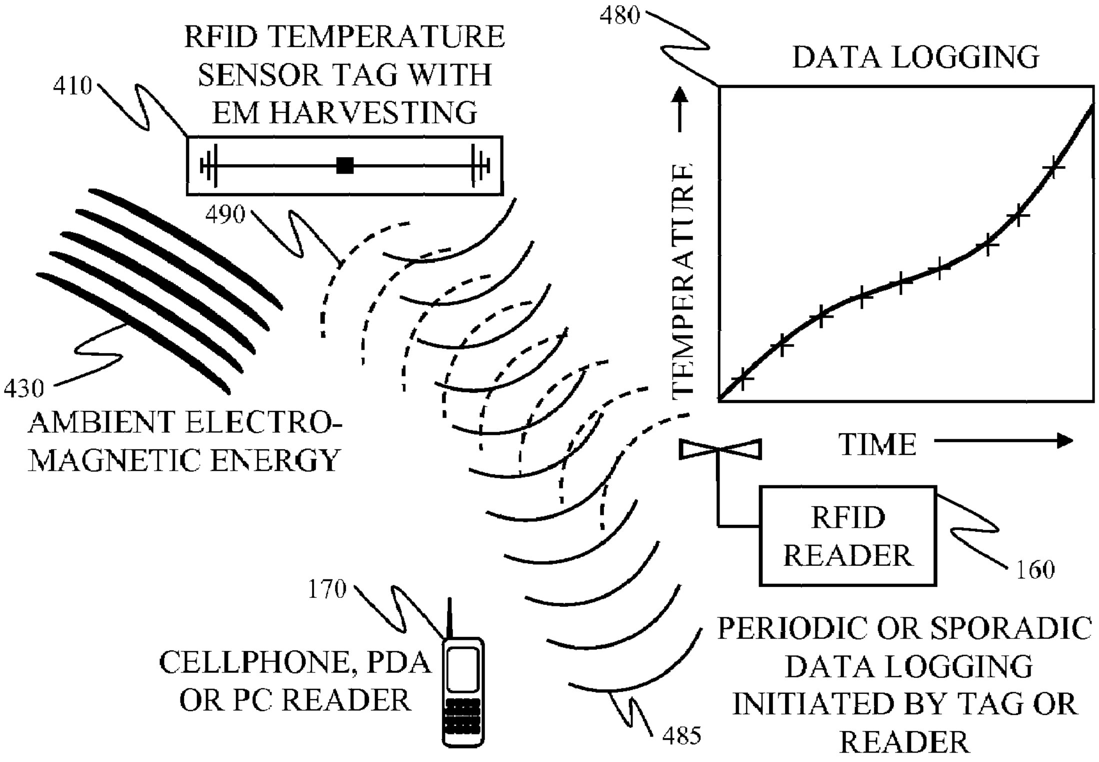
480
DATA LOGGING
RFID TEMPERATURE
SENSOR TAG WITH
EM HARVESTING
410
490
TEMPERATURE
430
AMBIENT ELECTRO-
MAGNETIC ENERGY
TIME
RFID
READER
160
170
CELLPHONE, PDA
OR PC READER
PERIODIC OR SPORADIC
DATA LOGGING
INITIATED BY TAG OR
READER
485

FIG. 4B

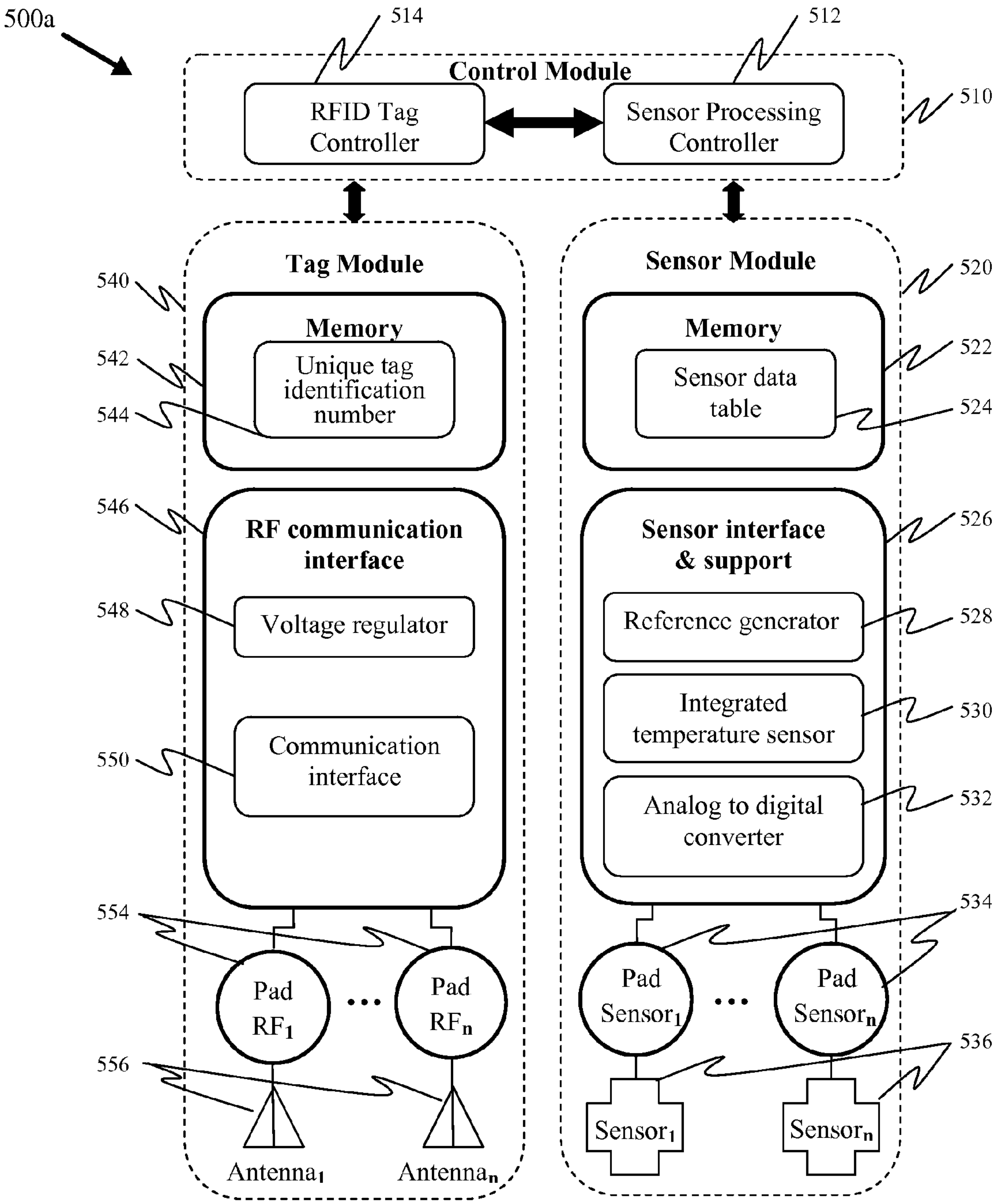
500a
514
512
Control Module
RFID Tag Controller
Sensor Processing Controller
510
Tag Module
Sensor Module
540
520
Memory
Unique tag identification number
542
544
Memory
Sensor data table
522
524
546
RF communication interface
Voltage regulator
548
Communication interface
550
Sensor interface & support
526
Reference generator
528
Integrated temperature sensor
530
Analog to digital converter
532
554
Pad RF$_1$
Pad RF$_n$
534
Pad Sensor$_1$
Pad Sensor$_n$
556
Antenna$_1$
Antenna$_n$
536
Sensor$_1$
Sensor$_n$

FIG. 5A

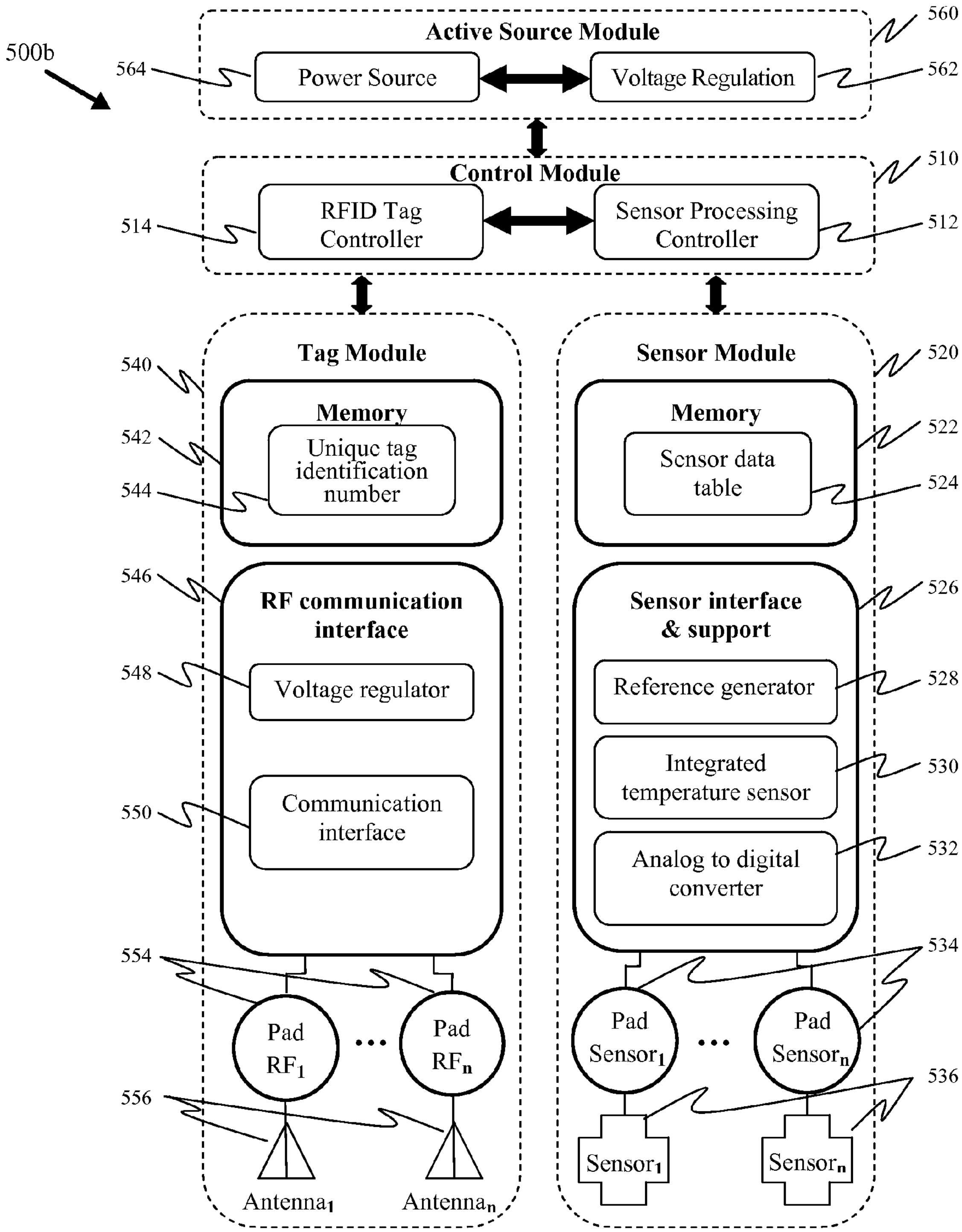
500b
Active Source Module
560
564
Power Source
Voltage Regulation
562
Control Module
510
514
RFID Tag Controller
Sensor Processing Controller
512
Tag Module
Sensor Module
540
520
Memory
Unique tag identification number
542
544
Memory
Sensor data table
522
524
546
RF communication interface
548
Voltage regulator
550
Communication interface
526
Sensor interface & support
Reference generator
528
Integrated temperature sensor
530
Analog to digital converter
532
554
Pad RF_1
Pad RF_n
534
Pad $Sensor_1$
Pad $Sensor_n$
556
536
$Antenna_1$
$Antenna_n$
$Sensor_1$
$Sensor_n$

FIG. 5B

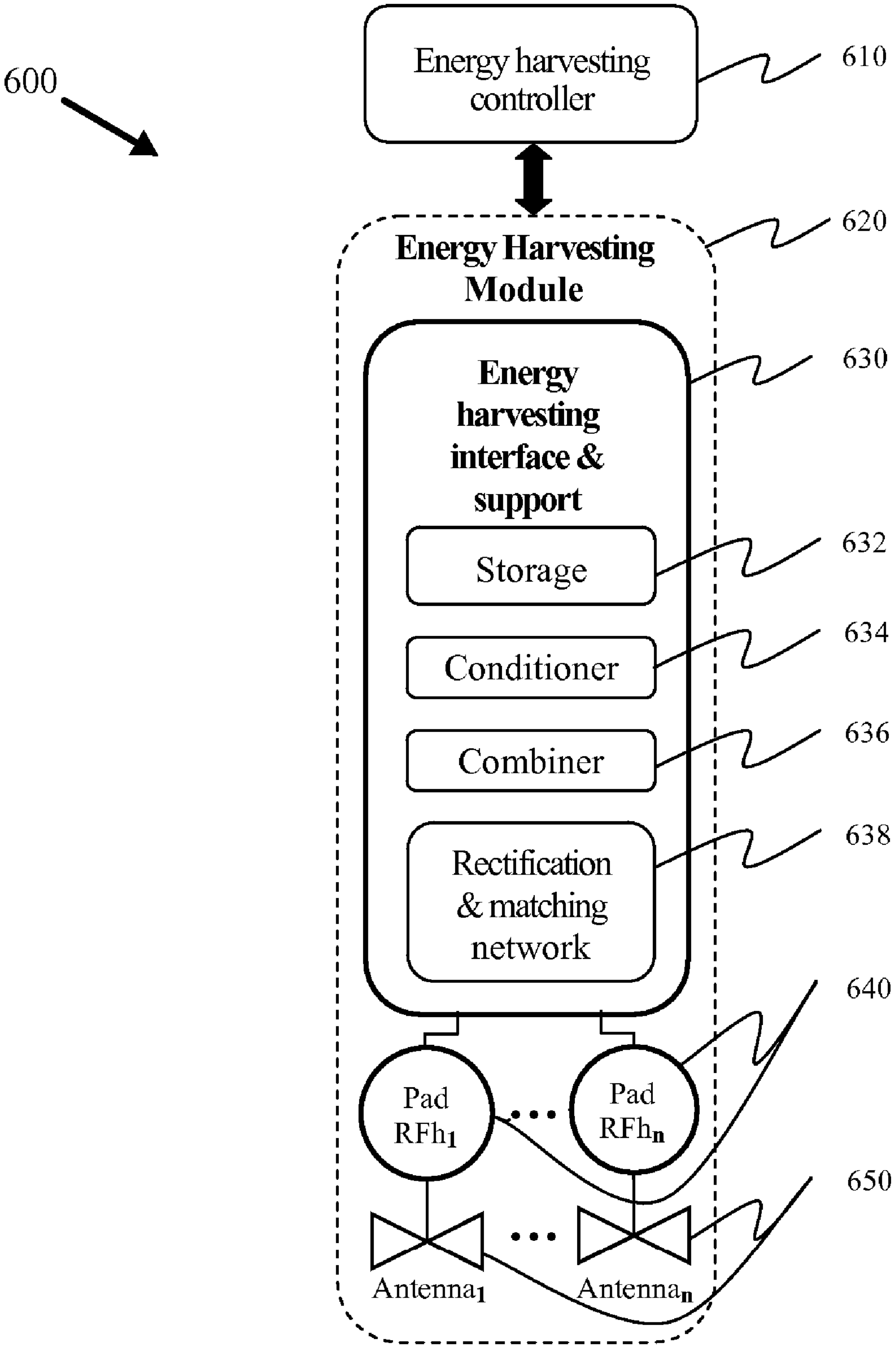
600
Energy harvesting controller
610
620
Energy Harvesting Module
630
Energy harvesting interface & support
632
Storage
634
Conditioner
636
Combiner
638
Rectification & matching network
640
Pad RFh$_1$
Pad RFh$_n$
650
Antenna$_1$
Antenna$_n$

FIG. 6

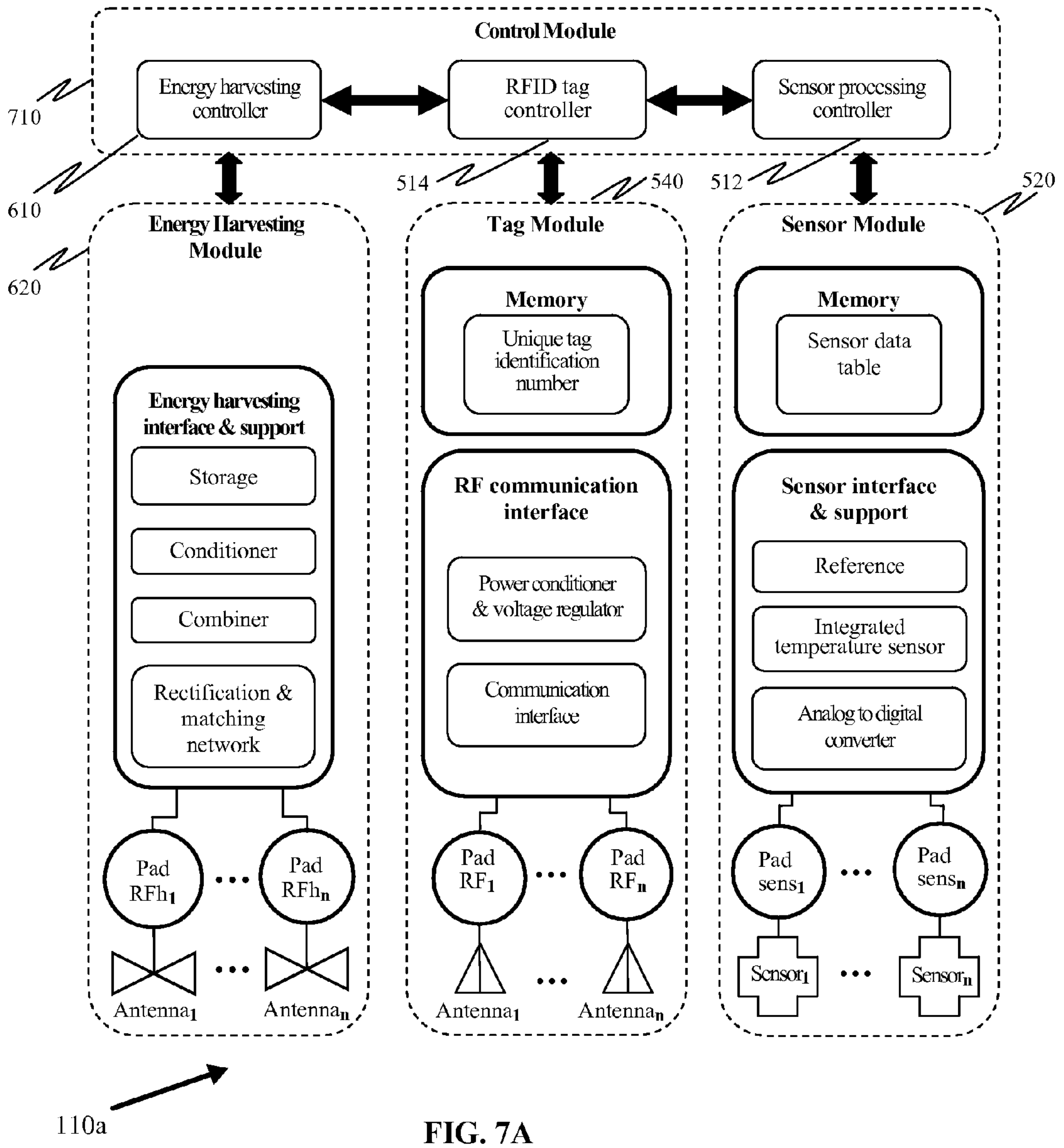
Control Module
Energy harvesting controller
RFID tag controller
Sensor processing controller
710
610
620
514
540
512
520
Energy Harvesting Module
Energy harvesting interface & support
Storage
Conditioner
Combiner
Rectification & matching network
Pad RFh$_1$
Pad RFh$_n$
Antenna$_1$
Antenna$_n$
Tag Module
Memory
Unique tag identification number
RF communication interface
Power conditioner & voltage regulator
Communication interface
Pad RF$_1$
Pad RF$_n$
Antenna$_1$
Antenna$_n$
Sensor Module
Memory
Sensor data table
Sensor interface & support
Reference
Integrated temperature sensor
Analog to digital converter
Pad sens$_1$
Pad sens$_n$
Sensor$_1$
Sensor$_n$
110a

FIG. 7A

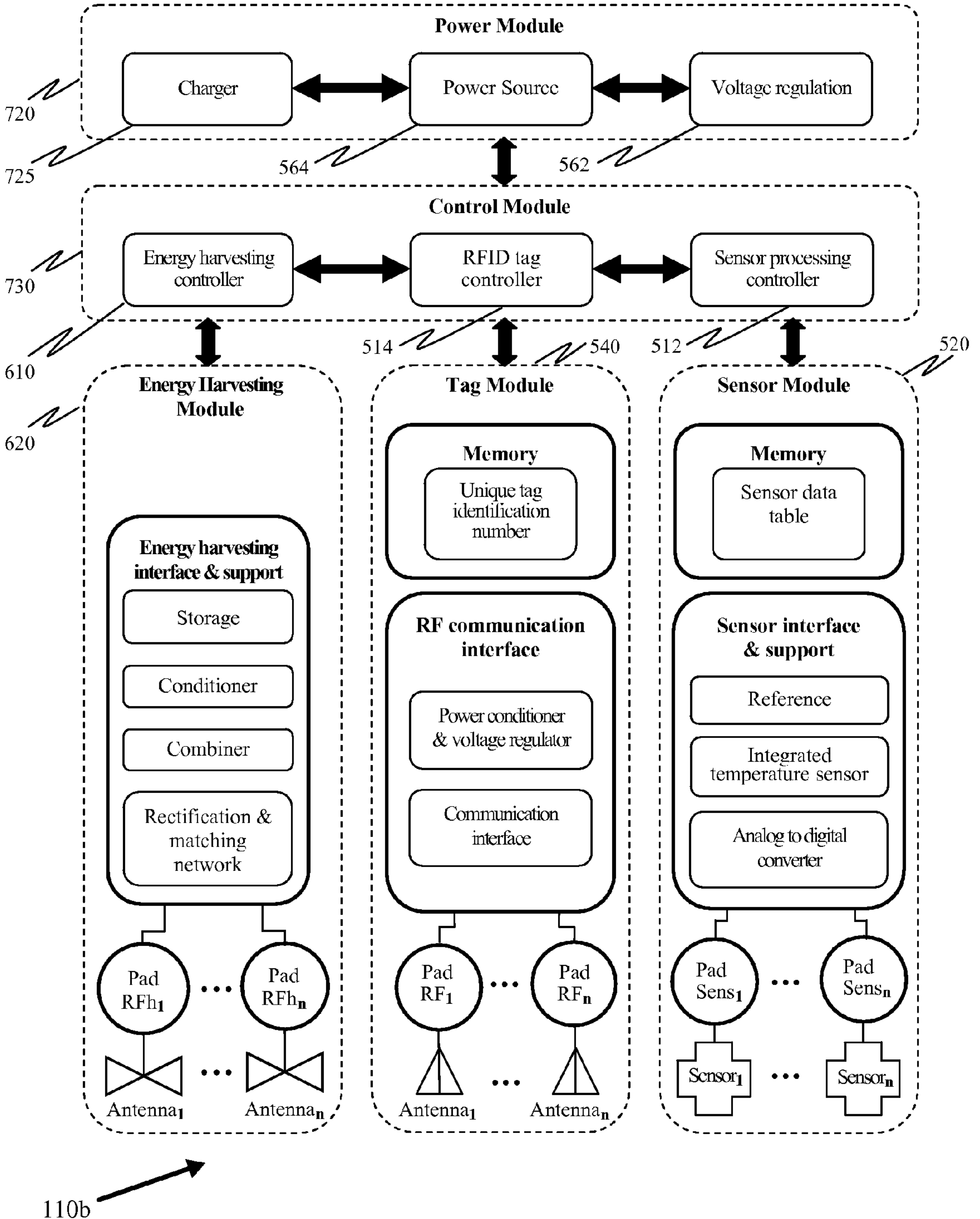
Power Module
Charger
Power Source
Voltage regulation
720
725
564
562
Control Module
Energy harvesting controller
RFID tag controller
Sensor processing controller
730
610
514
540
512
520
620
Energy Harvesting Module
Tag Module
Sensor Module
Memory
Unique tag identification number
Memory
Sensor data table
Energy harvesting interface & support
Storage
Conditioner
Combiner
Rectification & matching network
RF communication interface
Power conditioner & voltage regulator
Communication interface
Sensor interface & support
Reference
Integrated temperature sensor
Analog to digital converter
Pad RFh$_1$
Pad RFh$_n$
Pad RF$_1$
Pad RF$_n$
Pad Sens$_1$
Pad Sens$_n$
Antenna$_1$
Antenna$_n$
Antenna$_1$
Antenna$_n$
Sensor$_1$
Sensor$_n$
110b

FIG. 7B

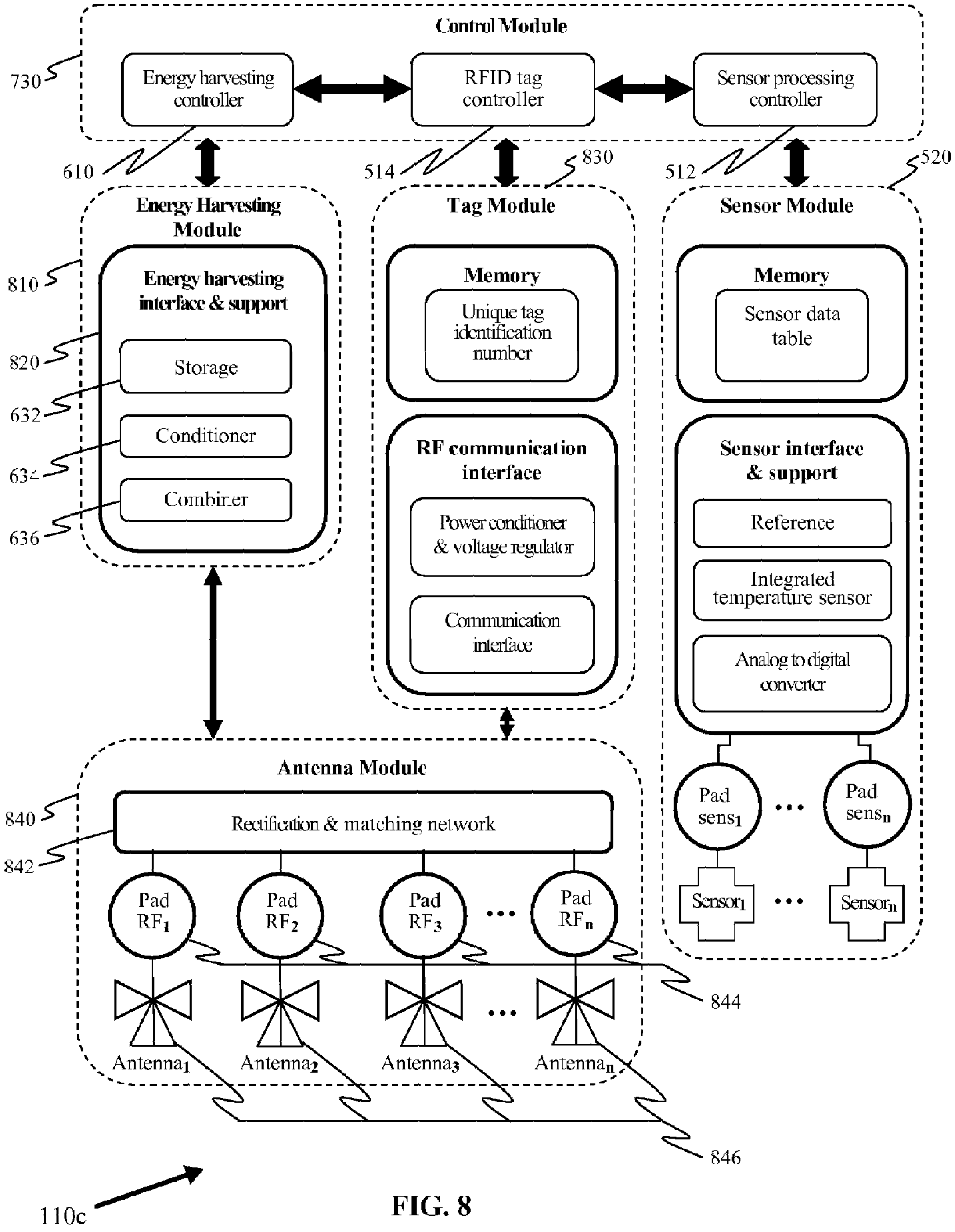
Control Module
Energy harvesting controller
RFID tag controller
Sensor processing controller
730
610
514
830
512
520
Energy Harvesting Module
Tag Module
Sensor Module
810
Energy harvesting interface & support
820
Storage
632
Conditioner
634
Combiner
636
Memory
Unique tag identification number
RF communication interface
Power conditioner & voltage regulator
Communication interface
Memory
Sensor data table
Sensor interface & support
Reference
Integrated temperature sensor
Analog to digital converter
Antenna Module
840
Rectification & matching network
842
Pad RF$_1$
Pad RF$_2$
Pad RF$_3$
Pad RF$_n$
Antenna$_1$
Antenna$_2$
Antenna$_3$
Antenna$_n$
844
846
Pad sens$_1$
Pad sens$_n$
Sensor$_1$
Sensor$_n$
110c

FIG. 8

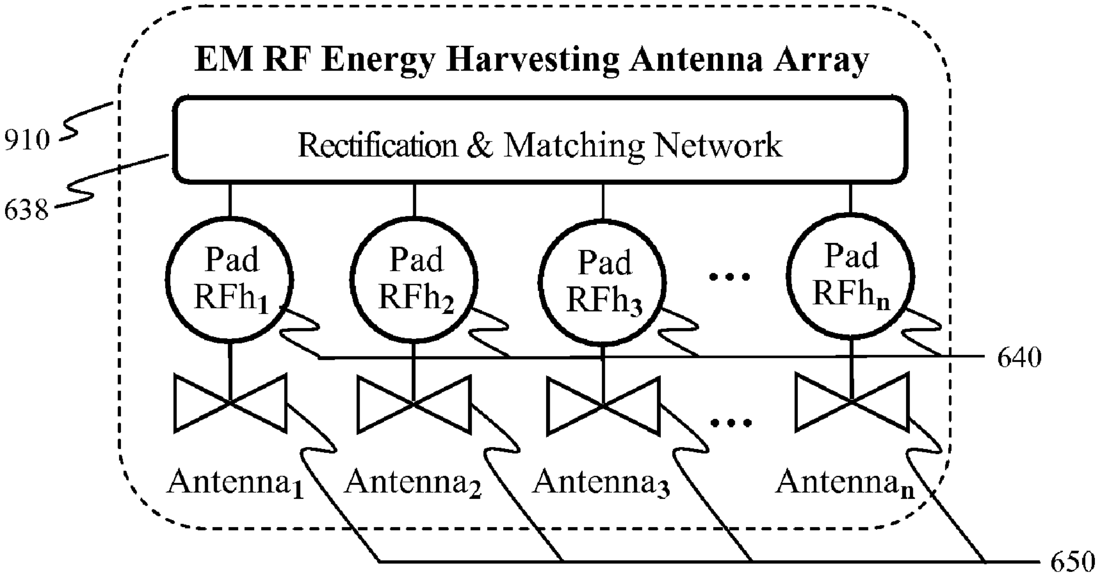
EM RF Energy Harvesting Antenna Array
Rectification & Matching Network
910
638
Pad RFh$_1$
Pad RFh$_2$
Pad RFh$_3$
Pad RFh$_n$
640
Antenna$_1$
Antenna$_2$
Antenna$_3$
Antenna$_n$
650

FIG. 9A

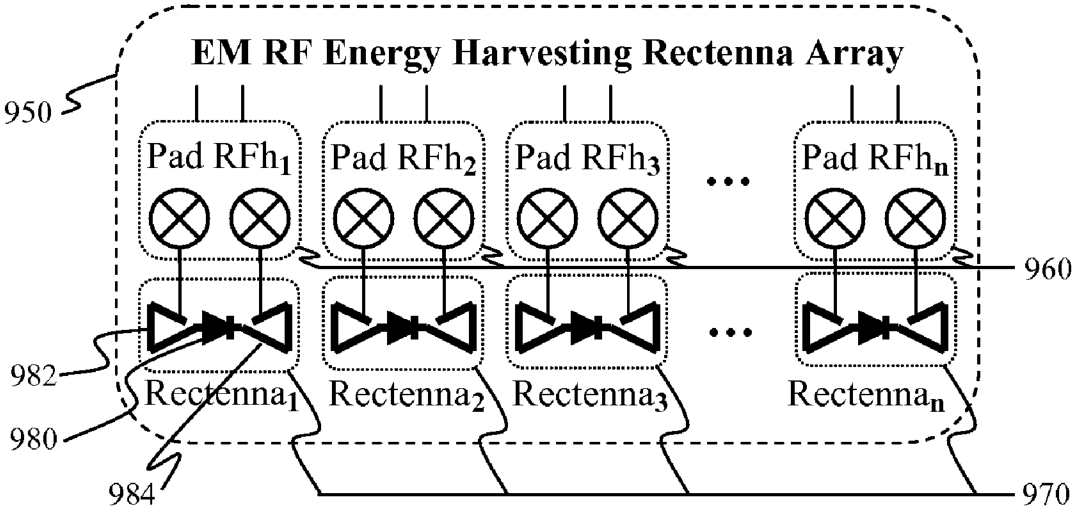
EM RF Energy Harvesting Rectenna Array
950
Pad RFh$_1$
Pad RFh$_2$
Pad RFh$_3$
Pad RFh$_n$
960
982
980
984
Rectenna$_1$
Rectenna$_2$
Rectenna$_3$
Rectenna$_n$
970

FIG. 9B

HARVESTING AMBIENT RADIO FREQUENCY ELECTROMAGNETIC ENERGY FOR POWERING WIRELESS ELECTRONIC DEVICES, SENSORS AND SENSOR NETWORKS AND APPLICATIONS THEREOF

CROSS REFERENCE TO RELATED APPLICATION

This application claims the benefit of U.S. Provisional Patent Application No. 60/705,636 filed on Aug. 4, 2005, the contents of which in its entirety is herein incorporated by reference.

BACKGROUND

1. Technical Field

The embodiments herein generally relate to wireless technology, and, more particularly, to harvesting ambient radio frequency (RF) electromagnetic (EM) energy for supplying, storing, recharging or supplementing the energy necessary to power active or passive wireless electronic devices, sensors and sensor networks.

2. Description of the Related Art

The ability of actively powered, battery assisted passive (BAP) and passive wireless sensors to remotely acquire, store and/or transmit data may open up a wide variety of medical, diagnostic, quality control, safety, supply chain, logistic and security applications. For an individual, the ability to determine, prior to consumption, whether a food item contains harmful bacteria or ingredients to which the individual is allergic is highly desirable. From a population's perspective, recent national and international events have increased the need for personal or distributed systems that can monitor and detect continuously in real-time or at periodic or irregular intervals in time, chemical agents, biological agents, radiological agents and other hazards locally or over wide geographical areas.

Examples of personal handheld wireless devices that can address these issues are cellular telephones (cell phones), personal data assistants (PDAs) and portable computers or notebooks (PCs). These are pervasive electronic devices that are networked via the Internet or cellular telephone networks. Generally, the abovementioned consumer electronic devices all require local energy storage (batteries) to provide mobility and portability, and do so until their local source of energy is depleted.

A new class of sensors such as Electronic Product Code (EPC) and Radio Frequency Identification (RFID) tags that include sensing capabilities are emerging as a generally inexpensive and effective means of addressing many wireless sensor applications such as, but not exclusively, cold-supply chain, food safety, quality control, environmental safety, medical, diagnostic, electro-immunoassays, consumer goods as well as homeland security, property and personal security applications.

Purely passive sensors, when actively interrogated by an RF transceiver (reader), use this received source of EM energy to power themselves up, to acquire readings from their sensors and then rebroadcast or reflect their specific identification code and sensor readings back to the interrogator. BAP tags use their batteries to acquire and store or log data from one or more sensors, but typically do not use their batteries to enhance their RF communication abilities. Active sensors have their own built in power source that can be used to acquire and store sensor readings at any time as well as

enhance the reception and transmission of RF communications. Generally, passive EPC and RFID tags equipped with one or more sensors will require a source of energy to measure and store their acquired information at times other than during active interrogation by a reader.

Because of the cost of the conventional sensors and sensor readers in particular, broad deployment of a sensor network over a large geographical area or widespread use by individuals is currently not particularly feasible. Generally, a considerable problem with large geographical deployment is that if the sensors are actively powered, their power sources need to be replaced when depleted, again adding to the cost. In addition, wear on sensors or sensor surfaces, generally requires sensors to be replaced on a regular basis further increasing the cost. In addition, imprecision in sensors generally requires cross validation to eliminate false positives, increasing the number of sensors that must be deployed for each application and thus also increasing the cost.

Accordingly, in view of the foregoing, there remains a need for low cost wireless sensors that can harvest ambient EM energy to power themselves up, acquire sensor data or recharge their power sources (batteries in the case of active sensors and batteries in BAP sensors).

SUMMARY

In view of the foregoing, the embodiments herein provide wireless RF addressable sensor network architectures for individuals, homes, industries, and homeland security where a wireless reader can supply RF power to and communicate with RF addressable tags that include sensors. Each active, BAP or passive tag can be equipped to harvest ambient EM energy from its environment allowing it to be autonomous: the tag can acquire and store data from its sensors independently of any particular reader event. The harvested ambient EM energy could, in addition, be used to transmit an alarm or "read me!" signal when a vital "sensor event" occurs, or to charge a battery on or near the tag. For example, the reader could be an RFID reader, a broadband wireless fidelity (WiFi) device, a cell phone, PDA, PC or a hybrid thereof comprising of a device adapted to read an RFID tag's unique identification and sensor information and bi-directionally communicate this information over the Internet or cellular telephone networks. The embodiments herein are also directed towards RF addressable tags with sensors that may be produced at costs lower than most types of active or passive tags conventionally available and furthers the state of the art by including ambient EM energy harvesting capabilities.

The RF addressable sensor tag combines RFID tag functionality with sensor functionality. The RF addressable sensor includes one or more antennas for communicating with the tag reader, one or more sensor elements, an RF power and communications interface, an RFID control module, and a sensor interface. The RFID control module includes logic to control RFID tag communications with an RFID tag reader. The RFID control module may also include logic and memory necessary to store and/or process the acquired sensor data.

The embodiments herein additionally furthers the current state of the art by adding an ambient EM energy harvesting sub-system to passive, BAP, or active wireless sensor tags. The harvested energy is used to provide power to the passive sensor or to provide power to the BAP or active sensor or for recharging an energy source on the active sensor tag or BAP sensor tag. This harvested energy is used by the sensor tag for: the facilitation of autonomous data collection and/or storage; the processing of acquired data; and the transmission of an alarm or "read-me!" signal when a vital sensor variable is above or below a critical threshold or vital sensor variables are above or below critical thresholds.

The embodiments herein are also adapted to identify WiFi and cellular base stations as sources of readily available, known frequency, ambient RF EM energy that can be harvested to provide energy to the sensor tag. The embodiments herein are also capable of identifying microwave band frequencies consistent with IEEE 521-1984 designations (Bands L, S, C, X, Ku, K, Ka, V, and W) as ambient RF EM energy that can be harvested to provide energy to the sensor tag.

In an instantiation of the embodiments herein, an RFID reader is used to communicate with nearby sensor tags that are being energized by harvested ambient RF EM energy supplied by a nearby WiFi router/transceiver or cellular base station. The WiFi router/transceiver or cellular base station not only supplies the ambient energy for the sensor tags, but also bi-directionally communicates and distributes the collected sensor tag data over the Internet or cellular telephone networks.

In an aspect of the embodiments herein, the RFID reader and WiFi router/transceiver are combined into a hybrid RFID-WiFi reader whereby the RFID reader sub-system communicates with nearby sensor tags that are being energized by harvested ambient RF EM energy supplied by the WiFi router/transceiver part of the hybrid RFID-WiFi reader. The WiFi router/transceiver part of the hybrid RFID-WiFi reader also facilitates the local and/or remote distribution of the collected sensor tag data over the Internet.

In another aspect of the embodiments herein, the RFID, WiFi router/transceiver, and cell phone technology are combined into a hybrid RFID-WiFi-cell phone reader whereby the RFID and/or WiFi reader sub-system communicates with nearby RFID sensor tags and the WiFi and cell phone subsystems facilitate local and remote access to the collected sensor tag data over the Internet or via the cellular telephone networks. The WiFi router/transceiver part of the hybrid RFID-WiFi-Cell phone reader is also the source of ambient energy available for harvesting by the sensor tags.

In another aspect of the embodiments herein, a source that radiates EM energy at a known frequency may be placed in the vicinity of the passive, BAP or active sensor tags for the sole purpose of supplying ambient RF EM energy at frequencies that these wireless sensors can harvest.

The energy harvesting system may include an antenna or an array of antennas to collect ambient EM energy having a single wavelength or a multitude of wavelengths of single polarization or arbitrary polarizations. In another aspect of the embodiments herein the energy harvesting antenna or antennas may be used to supplement the wireless tag's communication antenna or antennas. In another instantiation of the embodiments herein, the energy harvesting antenna or array of antennas connects to an impedance matching network and a rectifying and combining module before being fed to the power converting and/or conditioning module.

In yet another instantiation of the embodiments herein, the energy harvesting antenna or array of antennas may be designed to directly rectify the harvested energy (a rectenna or array of rectennas) and feed the rectified energy to an energy combining module that in turn feeds a power converting and/or conditioning module.

In another aspect of the embodiments herein, the energy harvesting power converting and/or conditioning module can supplement the wireless tag's RF transceiver power and thereby its communication capabilities. Furthermore, the embodiments herein preferably do not interfere with the communications between RFID readers and wireless RF addressable tags.

Generally, the embodiments provide a system comprising an ambient RF EM energy source; a sensor reader operatively connected to the ambient RF EM energy source; and a sensor in communication with the sensor reader, wherein the sensor comprises a control module comprising an energy harvesting controller; a tag module controlled by the control module; a sensor module controlled by the control module; an energy harvesting module controlled by the energy harvesting controller and adapted to wirelessly collect the ambient RF EM energy; and a power module adapted to provide a power supply for the control module, the power supply being generated by the collected ambient RF EM energy, wherein the energy harvesting controller is adapted to (i) assess power requirements of each of the tag and sensor modules, (ii) transfer power from the energy harvesting module to any of the tag and sensor modules, and (iii) charge a local energy storage device.

In one embodiment, the sensor further comprises an antenna module shared by the energy harvesting module and the tag module, wherein the antenna module comprises a rectification and impedance matching network; at least one RF pad operatively connected to the rectification and impedance matching network; and at least one antenna operatively connected to each of the at least one RF pad. In another embodiment, the tag module comprises a memory component; a RF communication interface operatively connected to the memory component; at least one RF pad operatively connected to the RF communication interface; and at least one antenna operatively connected to each of the at least one RF pad. The sensor module may comprise a memory component; a sensor interface and support sub-system operatively connected to the memory component; at least one sensor pad operatively connected to the sensor interface and support sub-system; and at least one antenna operatively connected to each of the at least one sensor pad. Alternatively, the sensor module comprises a memory component; a sensor interface and support sub-system operatively connected to the memory component; and at least one antenna operatively connected to sensor interface and support sub-system. Preferably, the control module further comprises a RFID tag controller; and a sensor processing controller operatively connected to the RFID tag controller.

In one embodiment, the energy harvesting module comprises an energy harvesting interface and support sub-system; at least one RF harvesting pad operatively connected to an energy harvesting interface and support sub-system; and at least one antenna operatively connected to each of the at least one RF harvesting pad. Alternatively, the energy harvesting module comprises an energy harvesting interface and support sub-system; at least one pair of RF harvesting pads operatively connected to the energy harvesting interface and support sub-system; and at least one rectenna operatively connected to each of the at least one pair RF harvesting pads. The energy harvesting interface and support sub-system may comprise a storage component; a power conditioner component operatively connected to the storage component; and a combiner operatively connected to the power conditioner, wherein the energy harvesting interface and support subsystem comprises a rectification and impedance matching network operatively connected to the combiner. Preferably, the power module comprises a power charger; a power source operatively connected to the power charger; and a voltage regulator operatively connected to the power source.

Additionally, the collected ambient RF EM energy enables the electronic device to be a fully autonomous data collector, alarm generator, event generator, data logger, and data processor independent of any other device. Moreover, the sensor reader may comprise any of a cell phone, a PDA, a PC reader, a RFID reader, a broadband WiFi device, and a combination thereof. Preferably, the power supply is generated only by the collected ambient RF EM energy. Furthermore, the sensor may be adapted to transmit a wireless signal to the sensor reader, and wherein the sensor reader is adapted to receive the transmitted wireless signal from the sensor. Also, the collected ambient RF EM energy facilitates transmission of the wireless signal when a selected sensor variable meets a predetermined threshold. In one embodiment, the ambient RF EM energy source comprises a WiFi router adapted to allow the sensor reader to communicate over the Internet.

Alternatively, the ambient RF EM energy source comprises a hybrid WiFi router transceiver adapted to allow the sensor reader to communicate over the Internet. In another embodiment, the sensor is a RF ambient EM energy harvesting sensor and is arranged with a plurality of other the sensors each adapted to collect data for climate control, security alarms, environmental alarms, and information gathering from other devices located within a communicable distance from the sensor, wherein all of the sensors are adapted to be any of locally and remotely monitored via any of the Internet and a cell phone network.

These and other aspects of the embodiments herein will be better appreciated and understood when considered in conjunction with the following description and the accompanying drawings. It should be understood, however, that the following descriptions, while indicating preferred embodiments and numerous specific details thereof, are given by way of illustration and not of limitation. Many changes and modifications may be made within the scope of the embodiments herein without departing from the spirit thereof, and the embodiments herein include all such modifications.

BRIEF DESCRIPTION OF THE DRAWINGS

The embodiments herein will be better understood from the following detailed description with reference to the drawings, in which:

FIG. **1** is a schematic diagram illustrating an example of ambient RF EM energy harvesting for gathering, accumulating, and storing energy and powering passive wireless sensors or sensor tags and/or recharging BAP and/or active wireless sensors or sensor tags all of which are equipped with ambient EM energy harvesting technology according to an embodiment herein;.

FIG. **2** is a schematic diagram illustrating the deployment of wireless addressable sensor tags with on-board EM energy harvesting technology in a home, business, office, warehouse, or building according to an embodiment herein;

FIG. **3**A is a schematic diagram illustrating an example of a WiFi wireless router/transceiver, connected to the Internet, bathing passive and/or BAP and/or active wireless sensor tags with energy at a known frequency or known frequencies according to an embodiment herein;

FIG. **3**B is a schematic diagram illustrating an example of a hybrid RFID-WiFi wireless router comprising an RFID tag reader module combined with a WiFi router/transceiver connected to the Internet according to an embodiment herein;

FIG. **3**C is a schematic diagram illustrating an example of a hybrid RFID-WiFi-cell phone wireless router comprising an RFID tag reader module, a WiFi router/transceiver module that facilitates connection to the Internet, and a cell phone module that facilitates connection to cellular telephone networks according to an embodiment herein;

FIG. **4**A is a schematic diagram illustrating an example of an RFID sensor tag equipped with an energy harvesting subsystem that initiates an alarm or "read me!" signal when a measured variable or combination of measured variables exceeds a predefined critical threshold according to an embodiment herein;

FIG. **4**B is a schematic diagram illustrating an example of an RFID sensor tag equipped with the energy harvesting sub-system logging (writing to memory) the sensed data according to an embodiment herein;

FIG. **5**A is a block diagram illustrating a passive wireless RF addressable tag module with a sensor module;

FIG. **5**B is a block diagram illustrating an active wireless RF addressable tag module with a sensor module;

FIG. **6** is a block diagram illustrating an EM energy harvesting sub-system architecture according to an embodiment herein;

FIG. **7**A is a block diagram illustrating a passive wireless RF addressable tag with a sensor module combined with an EM energy harvesting sub-system architecture according to an embodiment herein;

FIG. **7**B is a block diagram illustrating an active wireless RF addressable tag architecture with a sensor module combined with an illustrative EM energy harvesting sub-system architecture according to an embodiment herein;

FIG. **8** is a block diagram illustrating a passive wireless RF addressable tag architecture with a sensor module, an EM energy harvesting sub-system, and an antenna module that could be used for RF communication with a reader as well as to harvest ambient EM energy according to an embodiment herein;

FIG. **9**A is a block diagram illustrating an example of an EM energy harvesting architecture comprising an antenna or an array of antennas with associated rectification and matching networks according to an embodiment herein; and

FIG. **9**B is a block diagram illustrating an example of an EM energy harvesting architecture comprising a rectenna (rectifying antenna) or an array of rectennas with rectification and matching built into each rectenna according to an embodiment herein.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

The embodiments herein and the various features and advantageous details thereof are explained more fully with reference to the non-limiting embodiments that are illustrated in the accompanying drawings and detailed in the following description. Descriptions of well-known components and processing techniques are omitted so as to not unnecessarily obscure the embodiments herein. The examples used herein are intended merely to facilitate an understanding of ways in which the embodiments herein may be practiced and to further enable those of skill in the art to practice the embodiments herein. Accordingly, the examples should not be construed as limiting the scope of the embodiments herein.

As mentioned, there remains a need for low cost wireless sensors that can harvest ambient EM energy to power themselves up, acquire sensor data or recharge their power sources (batteries in the case of active sensors and batteries in BAP sensors). The embodiments herein achieve this by providing a technique for harvesting ambient RF EM energy for supplying, storing, recharging or supplementing the energy necessary to power active, BAP or passive wireless electronic devices, sensors and sensor networks, allowing them to become autonomous "data loggers" and/or empowering them to transmit their unique identification and sensor information when a sensed variable or a combination of sensed variables exceeds a threshold and triggers a vital "sensor event".

More specifically, the embodiments provide a system and device for harvesting various frequencies and polarizations of ambient RF EM energy for making a passive sensor (tag) into an autonomous passive sensor (tag) adapted to collect and store data with time-stamping and some primitive computation when necessary even when an interrogating RFID reader is not present (not transmitting). A specific source of ambient RF EM energy may include WiFi. The system and device may also allow for the recharging of energy storage units in active and BAP devices. The system could be a "smart building" that uses passive sensors with RF EM energy harvesting capability to sense environmental variables, security breaches, as well as information from "smart appliances" that can be used for a variety of controls and can be accessed locally or remotely over the Internet or cellular telephone networks. Referring now to the drawings, and more particularly to FIGS. **1** through **9**B, where similar reference characters denote corresponding features consistently throughout the figures, there are shown preferred embodiments.

The embodiments herein relate to ambient EM energy harvesting for powering passive wireless sensors and for recharging energy storage devices on active or BAP wireless sensors. The RF frequency band from 3 KHz to 300 GHz, and a subset of this band, the Microwave Frequency Band from 1 GHz to 110 GHz (as designated in IEEE 521-1984) are identified as the most ubiquitous sources of ambient energy available for energy harvesting.

The embodiments herein furthers the current state of the art by focusing on RFID sensor tag technologies that may be purely passive (i.e., no on-board power supply but may have on-board energy storage devices) and/or active (i.e., having on-board energy storage as the power supply) and/or BAP (i.e., also having on-board energy storage as the power supply). The distinction between active and BAP wireless sensors is made here: an active tag uses its local power source to power up its sensors and boost its RF receive and transmit communications power while a BAP tag uses its local power source to power up and collect data from its sensors alone.

In particular IEEE standards 802.11a, 802.11b, and 802.11g and WiFi routers/transceivers used in homes, businesses, warehouses, stores and in towns or cities transmit a known wavelength (or frequency) in the S-band (2 GHz to 4 GHz) and the C-band (4 GHz to 8 GHz). This source of energy can be optimally harvested with a combination of appropriately designed antenna or antennas and rectenna or rectennas, impedance matching circuits, voltage regulation circuits and energy storage devices and can be used for supplying energy to wireless sensors, be they passive, BAP or active.

EM energy transmitted by cell phone towers (cellular base stations) is also designated as a ubiquitous source of energy of known wavelength (or frequencies in the 800 MHz to 1.9 GHz), and may likewise be optimally harvested and processed for supplying energy to wireless sensors, be they passive, BAP or active.

In addition a further improvement to the state of the art and provided by the embodiments herein is a hybrid RFID-WiFi transceiver that combines an RFID sensor tag reader and a WiFi router/transceiver into a device that can provide ambient EM energy to RFID sensor tags, read their unique identity and data from the tag and transmit this data over the Internet. The hybrid RFID-WiFi transceiver enables remote query of RFID sensor tags' identity and data via the Internet.

An additional improvement to the state of the art provided by the embodiments herein is a hybrid RFID-WiFi-cell phone transceiver that combines an RFID sensor tag reader, a WiFi router/transceiver, and a cell phone into a device that can provide ambient EM energy to RFID sensor tags, read their unique identity and data from the tag and transmit this data over the Internet and/or cellular telephone networks. The hybrid RFID-WiFi-cell phone transceiver enables remote query of RFID sensor tags' identity and data over the Internet as well as via cellular telephone networks.

The embodiments herein further improves upon the state of the art by making it possible for purely passive sensors to harvest available EM energy from their environment, and locally accumulate and store this energy so that passive tags become autonomous; i.e., they can log data and they can transmit an alarm or "read me!" signal to a nearby reader when a critical sensor event occurs or when a critical combination of sensor events occur or when a sensor level or combination of sensor levels has crossed a threshold.

FIG. **1** is an illustrative scenario of some of the forms of ambient EM energy that may be available for harvesting by passive RF addressable sensor tags **110** with energy harvesting capability. The ambient EM energy bathing the environment **115** in which the tags **110** are placed can come from a number of different sources including broadband WiFi routers/transceivers 120 (including but not limited to IEEE 802.1x, IEEE 802.15.4 "Zigbee®" electronic devices, Bluetooth® telecommunication equipment, HomeRF® computer hardware/software, HiperLAN/1 and HiperLAN/2), an EM transmitter **130** continuously or periodically radiating a known frequency (or frequencies), Microwave IEEE 521-1984 or cell phone base stations **140**, satellite transceivers **150**, wireless tag sensor readers **160**, and cell phones, PDAs or PCs **170**.

The energy that each of these sources contribute to the total amount of ambient EM energy available for harvesting by the antenna (not shown) of the tag **110**, $P_{RF}(t)$ **125**, **135**, **145**, **155**, **165**, **175** respectively, is a function of time, position/distance, frequency, solid angle and polarization state:

$$P_{RF} = \frac{1}{f_2 - f_1} \int_{f_1}^{f_2} \int_0^{4\pi} \Phi(\Omega, f, t) A_{eff}(\Omega, f) d\Omega df$$

where $\Omega$is the solid angle in steradians, $\Phi(\Omega f,t)$ is the time-varying frequency and angle-dependent incident power density and $A_{eff}$is the angle-, frequency-, and polarization-dependent effective area of the antenna.

The available power $P_{RF}(f_i)$ at any frequency f_i is an alternating current (ac) signal that must be converted to a direct current (dc) signal by a nonlinear device or rectifier (not shown) before being summed together with all other energy sources of different frequencies, $f_{1 \ldots n}$, which have been similarly rectified. The dc power, P_{dc} at a single frequency, f_i is given by:

$$P_{dc}(f_i) = P_{RF}(t, f_i) n[P_{RF}(t, f_i), \beta]$$

where $\eta = P_{dc}/P_{RF}$ is the total received rectified and summed power conversion efficiency (including the loss due to reflected power), and $\beta(P_{dc}, f_i,)$ represents the extent to which the rectifying device (not shown) and the antenna are impedance matched at the received dc power level, P_{dc} and frequency f_i.

The anticipated ambient power densities that can be harvested and converted to dc power at useful voltage levels is approximately in the range 10^{-1} mW/cm^2 (the FCC exposure limit for 880 MHz and 1990 MHz) to 10^{-6} mW/cm^2. One skilled in the art can optimize the harvested dc power, P_{dc} by appropriately adjusting the parameters A_{eff} and/or the parameters η and, β.

FIG. **2** is an illustrative example of a "smart home" 200 where RF addressable sensor tags **110** with energy harvesting capability are mounted on windows **210**, a door **220**, and in the attic **230**. Wired **270** or satellite transceiver **150** broadband Internet access is connected to a WiFi router **120** which bathes the smart house **200** in RF energy. Alternatively, an RF EM source **130** of known frequency continuously or periodically radiates energy specifically for the purpose of providing energy for the RF addressable sensor tags **110** with energy harvesting capability. An RFID tag reader **160** or a cell phone, PDA or PC **170** can read the unique identity and sensor data from each passive sensor tag **110**. The sensor tags **110** can be used to monitor environmental conditions such as temperature, humidity, carbon monoxide levels, radiation levels or smoke inside the house and the data read from the sensor tags **110** may be used to control HVAC systems or sound alarms when sensor thresholds are exceeded. The same sensor tags **110** can be used to detect unwanted intruders via vibration detectors, breakage detectors and open door detectors. Accordingly, the data from such sensors **110** can be used in a home security monitoring device. Similarly, addressable sensor tags **110** with energy harvesting capability can be placed on home appliances such as refrigerators, stoves, hot water heaters and gas/oil furnaces to monitor their operating state. Since the smart house **200** is connected to the Internet **270**, **150**, and/or via a cell phone **170**, all sensor data can be accessed locally and/or remotely over existing Internet and/or cellular telephone networks. Again, these sensors **110** harvest ambient energy from any or all of the available modes of RF EM energy bathing the house **200**. This energy could be used to power the passive sensor tags **110** or recharge batteries or other power sources associated with active or BAP tags **110**.

FIG. **3**A illustrates an example of a WiFi router/transceiver **120** connected to the Internet **270**, bathing passive, BAP or active wireless sensor tags **110** with energy at a known frequency **125**. This particular source of ambient energy supplies the passive sensor tag **110** with power or facilitates recharging batteries or other on-board or nearby energy storage devices. A single tag reader **160** (or a plurality of tag readers) is adapted to interrogate **330** each uniquely identifiable sensor tag **110** for information about its particular sensor tag reading(s) **320** and pass this information to the WiFi router/transceiver **120** via a direct connection **310** to make it locally or remotely accessible via the Internet and/or cell phone networks. This particular source of ambient energy can supply the passive sensor tag **110** equipped with energy harvesting capability with power for autonomous data collection and storage, or facilitate recharging batteries or other energy storage devices.

FIG. **3**B illustrates a hybrid RFID-WiFi wireless router **340** comprising an RFID tag reader module combined with a WiFi router/transceiver connected to the Internet **270**. This hybrid RFID-WiFi reader/module **340** provides a source of ambient energy **125** which the sensor tag **110** can harvest to supply itself with power to become an autonomous sensor **110** able to collect and save data as well as send out an alarm if its reading(s) has/have crossed a predefined threshold or a combination of predefined thresholds. The harvested energy can also facilitate recharging batteries or other energy storage devices associated with each tag **110**. The hybrid RFID-WiFi router/transceiver **340** periodically or sporadically interrogates **330** each uniquely identifiable sensor tag **110** for its data, and each uniquely identifiable sensor tag **110** rebroadcasts **320** its particular sensor(s) reading(s) back to the RFID-WiFi router/transceiver **340** which makes that data accessible over the Internet.

FIG. **3**C illustrates a hybrid RFID-WiFi-Cell phone router/transceiver **350** comprising an RFID tag reader module, a WiFi router/transceiver module and a cell phone module. This hybrid RFID-WiFi-Cell phone reader module **350** provides a source of ambient energy **125** which the sensor tag **110** can harvest to supply itself with power to become an autonomous sensor **110** able to collect and store data as well as send out an alarm if its reading(s) has/have crossed a predefined threshold or a combination of predefined thresholds. The harvested energy can also facilitate recharging batteries or other energy storage devices associated with each tag **110**. The hybrid RFID-WiFi-Cell phone router/transceiver **350** periodically or sporadically interrogates **330** each uniquely identifiable sensor tag **110** for its data, and each uniquely identifiable sensor tag **110** rebroadcasts **320** its particular sensor(s) reading(s) back to the hybrid RFID-WiFi-Cell phone router/transceiver **350** which is connected to the Internet **270** and makes that data accessible over the Internet as well as over cellular telephone networks.

FIG. **4**A illustrates an RFID temperature sensor tag **410** equipped with energy harvesting capability. Ambient EM energy **430** is harvested by the sensor tag **410**. This harvested energy is used by the tag **410** to continuously, periodically, or sporadically measure its sensor's value and compare that value with a pre-programmed threshold **470**. If this threshold is crossed, the tag **410** initiates and broadcasts an alarm or "read me!" signal **460** which is read by an RFID reader **160** or cell phone, PDA, or PC reader **170** along with its unique identification number. In this particular illustrative example the tag **410** senses temperature, but in general, the sensor or sensors of the tag **410** could measure a variety of environmental or physical variables. The unique identification and sensor data of the tag **410** can then be read by a standalone RFID reader **160** or by a cell phone, PDA, or PC reader **170**.

FIG. **4**B illustrates an RFID temperature sensor tag **410** equipped with energy harvesting capability that uses ambient EM energy **430** to provide a continuous, periodic, or sporadic source of power to the RFID temperature sensor tag **410**. The sensor tag **410** logs data to local memory **480** and can periodically or sporadically initiate and broadcast a signal **485**, which can be read by the RFID reader **160**. The RFID reader **160** can also periodically or sporadically initiate a read **490** of the unique identification and data of the tag **410**. In this particular illustrative example the tag **410** senses temperature, but in general, the sensor or sensors of the tag **410** could measure a variety of environmental or physical variables. The unique identification and sensor data of the tag **410** can then be read periodically or sporadically either when a sensor event initiates a "read me!" signal **485**, or when a standalone RFID reader **160** or a cell phone, PDA, or PC reader **170** initiates a read cycle **490**.

FIG. **5**A illustrates a passive wireless RF addressable sensor tag **500***a*, which may be used as a basis for designing the embodiments herein. The passive RFID tag **500***a* includes three main modules: the control module **510**, the tag module **540**, and the sensor module **520**. The control module **510** has an RFID tag controller **514** and a sensor processing controller **512** that can communicate with each other. The RFID tag controller **514** communicates with and controls the tag module **540**. The tag module **540** has on-board memory **542** and an RF communication interface **546**. The memory **542** holds the unique identification number and additional information **544** associated with that particular sensor tag **500***a*. The RF communication interface **546** has a voltage regulator **548** and a communication interface **550**. The RF communication interface **546** connects to a single antenna or a multitude of antennas **556** through an RF pad or a multitude of RF pads **554**. The sensor module **520** has its own memory **522** and a sensor interface and support sub-system **526**. The memory **522** stores the sensor data table **524** that include sensor calibration parameters, lookup tables as well as data collected from the sensor or sensors of the sensor tag **500***a*. The sensor interface and support sub-system **526** has a voltage and/or current reference generator **528**, an integrated temperature sensor **530** and an analog to digital converter **532**. An external sensor or a multitude of external sensors **536** are connected to the sensor interface and support sub-system **526** via a sensor pad or a multitude of sensor pads **534**.

FIG. **5**B illustrates an active or BAP wireless RF addressable sensor tag **500***b*, which may be used as a basis for designing the embodiments herein. In addition to all of the components described with FIG. **5**A above, there is also an active source module **560** which communicates with the control module **510**. The active source module **560** includes a power source **564** which could be replaceable and/or rechargeable, and voltage regulation circuitry **562** which regulates the voltage and/or current necessary to recharge the power source **564** and supply power to the remainder of the sensor tag **500***b*.

FIG. **6** is an illustrative example of a block diagram of the embodiments herein as it relates to the harvesting of ambient EM energy for supplying, storing, recharging, or supplementing the energy necessary to power active, BAP or passive wireless electronic devices, sensors and sensor networks. The EM energy harvesting sub-system architecture **600** comprises a single antenna **650**, rectifier and matching network **638** or an array of antennas **650**, rectifiers and matching networks **638**, a combiner **636**, a power conditioner **634**, and an energy storage device **632**. The energy harvesting controller **610** communicates bi-directionally with the energy harvesting module **620** and RFID tag controller **514**, wherein both the energy harvesting controller **610** and energy harvesting module **620** may or may not be integrated into an RFID sensor chip (not shown), depending on the particular application. The energy harvesting controller **610** sporadically, periodically or continuously controls and monitors the state of the energy harvesting module **620**. The energy harvesting module **620** includes an energy harvesting interface and support sub-system **630** that connects to an external energy harvesting antenna or multitude of external energy harvesting antennas **650** through an RFh (RF harvesting) pad or a multitude of RFh pads **640**.

The antenna or multitude of antennas **650** may be designed to collect ambient EM energy at either a single frequency f_1, a range of frequencies f_i, i=1, 2, 3 . . . N, or a specific set of frequencies, for example $f_{\{i,j,k\}l}=\{13.56$ MHz, 438 MHz, 928 MHz, 2.54 GHz$\}$. Moreover, the polarization state of the energy within the aperture of a single energy harvesting antenna or a multitude of energy harvesting antennas **650** can be arbitrary. Furthermore, the aperture or effective area A_{eff} of an energy harvesting antenna or multitude of energy harvesting antennas **650** can be optimized for a wavelength, and/or a range or set of wavelengths of the received EM energy, and can be readily accomplished by those skilled in the art.

The EM energy harvested by an energy harvesting antenna or a multitude of energy harvesting antennas **650** presents an ac signal or a multitude of ac signals to the rectification and matching network **638**. The matching network **638** is preferably designed to optimally match the frequency, polarization state and signal strength (power) dependence $\beta(P_{dc},f_i)$ of the output impedance of the harvesting antenna(s) **650** and the input impedance of the rectification element or elements **638** to ensure maximum power transfer or harvesting efficiency. The rectifier or rectifiers **638** convert the ac signal(s) into a dc or voltage signal(s).

The combiner **636** takes its input from the output of the rectification and matching network **638** and combines the currents or voltages in any combination of series or parallel connections in order to provide the necessary quantity of dc power for passive, BAP, or active sensor tag operation. The conditioner **634** takes the dc power from the combiner **636** and converts this dc power into a specified voltage level and/or current level with efficiency η by using circuit techniques known by those skilled in the art. The voltage and/or current level may be determined by the requirements of the local energy storage device **632** and/or the active source module **560**. The storage device **632** acts as local storage and may be implemented using any or a combination of capacitors, supercapacitors, plastic batteries, polymer batteries, Li-Ion batteries, or any rechargeable energy storing technology that can be integrated within the energy harvesting module **620**.

FIG. **7**A is an illustrative block diagram of an instantiation of a passive sensor tag **110***a* that includes a control module **710**, a tag module **540**, a sensor module **520**, and an energy harvesting module **620**. The energy harvesting controller **610** is shown here as being part of the control module **710** and communicates with the RFID tag controller **514** and the sensor processing controller **512**. The energy harvesting controller **610** is responsible for assessing the power needs of the tag module **540** and the sensor module **520** and can divert power from its storage element **632** to the tag module **540** and the sensor module **520** as necessary.

FIG. **7**B is an illustrative block diagram of an instantiation of a BAP or active sensor tag **110***b* that includes the energy harvesting module **620**. The difference between this sensor tag **110***b* of FIG. **7**B and sensor tag **110***a* of FIG. **7**A is the addition of the power module **720** whose purpose is to control and distribute power to the control module **730**, the tag module **540**, and the sensor module **520** in a BAP or active sensor tag **110***b*. The power module **720** includes, but is not limited to, a charger **725**, a power source **564**, and voltage regulation **562**.

The charger **725** controls the continuous, periodic, or sporadic charging of energy storage devices, in particular the power source **564** which may be a capacitor, supercapacitor, plastic battery, polymer battery, Li-Ion batteries, or any rechargeable energy storage technology. The voltage regulation sub-system **562** regulates the voltage and/or current that is needed to charge/recharge the power source **564** or divert power to the tag module **540** and/or the sensor module **520** as determined by the control module **730**.

FIG. **8** is a block diagram of another illustrative example of a sensor tag **110***c*with EM energy harvesting capability. As with the sensor tag **110***a* of FIG. **7**A and sensor tag **110***b* of FIG. **7**B, the sensor tag **110***c* of FIG. **8** comprises a control module **730**, a sensor module **520**, a tag module **830**, an energy harvesting module **810**, and an antenna module **840**. The tag module **830** differs from the tag module **540** of FIGS. **5**A, **5**B, **7**A, and **7**B in that there are no RF pads **554** and no antennas **556**. The remaining functionality of the tag module **830** is the same as described above.

Likewise, the energy harvesting module **810** differs from the energy harvesting module **620** of FIGS. **6**, **7**A, and **7**B in that there are no RF pads **554** and no RF antenna **556**, nor are there RFh pads **640** or energy harvesting antennas **650** associated with this module **810**. The remaining components, storage **632**, conditioner **634**, and combiner **636** have the same functionality as described above.

The antenna module **840** serves both the tag module **830** and the energy harvesting module **810**. An antenna or a multitude of antennas **846** are used for both RF EM energy harvesting and for bidirectional RF communications with an RFID, hybrid RFID-WiFi or hybrid RFID-WiFi-Cell phone reader. The antenna or antennas are designed to optimize RFID read range as well as to optimize A_{eff}. Each antenna **846** is connected via an RF pad **844** to the rectification and matching network **842**.

The rectification and matching network **842** is used to match the impedance of the antenna(s) **846** to that of the rectifier(s) as described above. In the context of FIG. **8**, the rectification and matching network **842** includes the rectifier(s). Moreover, those skilled in the art would readily understand that diodes, transmission lines, microstrip elements, and surface mount components could be further included in the rectification and matching network **842** in order to provide the rectifier(s). Since this network **842** also allows RF communications to occur, frequencies corresponding to those used by the RFID, RFID-WiFi, or RFID-WiFi-Cell phone reader (not shown in FIG. **8**) are preferably isolated and directed to the communication interface **550** on the tag module **830**. The harvested EM energy, once rectified, is sent first to the combiner **636**, then to the conditioner **634**, and finally to storage **632** as described above.

FIG. **9**A is an illustrative example of a block diagram of one of many possible designs for an EM RF energy harvesting antenna array module **910** in accordance with the embodiments herein, with the functionality of the various components included therein as previously described above. FIG. **9**B is an illustrative example of a block diagram of an alternate method for harvesting EM RF energy. In this aspect of the embodiments herein, an EM RF energy harvesting rectenna array module **950** is provided as a possibly more efficient way to harvest ambient EM RF energy.

In FIG. **9**B, an array of rectennas **970**, $\text{Rectenna}_1, \ldots, \text{Rectenna}_n$, each comprise a pair of antennas **982**, **984**, optimized to collect a broad range of RF frequencies having time-varying polarization states, directly integrated with an impedance matched rectifier **980**. Each rectenna **970** could be made up of two interdigitated antennas, one with a clockwise winding **982**, the other with a counterclockwise winding **984**, with the rectifier **980** connected between the two inner ports of the antennas. The rectified signals are collected from the two outermost ports of the rectenna **970** and connected to pads $\text{RHh}_1, \ldots, \text{RFh}_n$, **960**.

The techniques provided by the embodiments herein may be implemented on an integrated circuit (IC) chip or using printable electronic technologies (not shown). The chip or printable electronic circuit design is created in a graphical computer programming language, and stored in a computer storage medium (such as a disk, tape, physical hard drive, or virtual hard drive such as in a storage access network). If the designer does not fabricate chips or printable electronic circuits or the photolithographic masks used to fabricate chips or printable electronic circuits, the designer transmits the resulting design by physical means (e.g., by providing a copy of the storage medium storing the design) or electronically (e.g., through the Internet) to such entities, directly or indirectly. The stored design is then converted into the appropriate format (e.g., GDSII or CIF) for the fabrication of photolithographic masks, which typically include multiple copies of the chip design in question that are to be formed on a wafer or printed on a suitable substrate. The photolithographic masks are utilized to define areas of the wafer or printable electronic circuits (and/or the layers thereon) to be etched or otherwise processed or printed.

The resulting integrated circuit chips or printable electronic circuits can be distributed by the fabricator in raw wafer form (that is, as a single wafer that has multiple unpackaged chips), as a bare die, or in a packaged form or as individual printed circuits or in a sheet or roll of printed circuits. In the latter case the chip is mounted in a single chip package (such as a plastic carrier, with leads that are affixed to a motherboard or other higher level carrier) or in a multichip package (such as a ceramic carrier that has either or both surface interconnections or buried interconnections). In any case the chip might then be integrated with other chips, discrete circuit elements, and/or other signal processing devices as part of either (a) an intermediate product, such as a mother or daughter-board, or (b) an end product. The end product can be any product that includes integrated circuit chip or chips and/or printed circuits, ranging from toys and other low-end applications to advanced computer products having a display, a keyboard or other input device, and a central processor.

Harvesting ambient EM energy would transform a purely passive sensor into an autonomous sensor **110**a, **110**b, **111**c capable of periodic or sporadic sensor data acquisition, data logging and storage at times other than when being actively interrogated by a reader **160** as well as providing the necessary energy to either autonomously transmit a notable "sensor event", broadcast an alarm or "read me!" signal **460**, **485**, **490** when a threshold of the sensor **110**a, **110**b, **110**c has been exceeded or when a combination of various thresholds of the sensor **110**a, **110**b, **110**c have been exceeded.

The combination of passive and/or BAP and/or active wireless sensors **110**a, **110**b, **110**c having ambient EM energy harvesting capabilities and wireless communication capabilities as provided by the embodiments herein provide a new low cost means for remotely sensing the environment, monitoring the state of a home, buildings, structures, goods and devices as well as remote and local personal/patient monitoring, medical diagnostic and medical care solutions.

Cell phones, PDAs and PCs **170** are pervasive electronic devices that are already networked via the Internet. A combination or integration of RFID readers, WiFi technologies and/or cell phone technologies into a hybrid reader that transparently communicates with RFID sensor tags, the Internet and cell phone networks would facilitate the widespread deployment and usage of passive sensor tags equipped with EM energy harvesting technology.

Broadband wireless communications networks such as WiFi (including WiMax™ telecommunication equipment and WiBro™ telecommunication equipment and any combination of IEEE 802.1x (where x is used to convey any or all of the following: 1a, 1b, 1g, 5 or 6), IEEE 802.15.4 "Zigbee®" electronic devices, Bluetooth® telecommunication equipment, HomeRF® computer hardware/software, HiperLAN/1 and HiperLAN/2) are being widely deployed in offices, factories, warehouses, stores, homes, communities, and cities. With long read ranges (up to several hundred feet), RFID tags and RFID readers can benefit from aspects of the embodiments herein by: using RFID and/or WiFi for short range communication between a tag reader and wireless sensor tags; using the WiFi and/or the cellular network infrastructure to communicate the acquired sensor tag data over the Internet or cellular telephone networks (bi-directionally); and using WiFi routers/transceivers or cellular base stations as a source of ambient EM energy, with known frequency, that can be optimally harvested to provide energy to inexpensive passive (BAP and/or active tags with rechargeable energy sources) sensor tags.

Generally, the embodiments herein provide a technique for harvesting ambient RF EM energy for supplying, storing, recharging or supplementing the energy necessary to power active, BAP or passive wireless electronic devices, sensors and sensor networks, allowing them to become autonomous "data loggers" and/or empowering them to transmit their unique identification and sensor information when a sensed variable exceeds a threshold and triggers a vital "sensor event". The data gathered by these autonomous data loggers can also be transparently read at any time by a variety of reader technologies such as RFID readers, cell phones, personal data assistants, personal computers, WiFi, WiMax™telecommunication equipment and WiBro™ telecommunication equipment, including IEEE 802.1x (where x is used to convey any or all of the following: 1a, 1b, 1g, 5, and 6), "Zigbee®" electronic devices, Bluetooth® telecommunication equipment, HomeRF® computer hardware/software, HiperLAN/1and HiperLAN/2 and any hybrid combinations thereof. The embodiments herein also provide a hybrid RFID-WiFi reader/transceiver **340** and a hybrid RFID-WiFi-cell phone reader/transceiver **350**. The RFID reader module communicates with RFID tags, the WiFi module supplies ambient energy which the passive sensor tags are equipped to harvest and provides connectivity to the Internet and may also read WiFi sensor tags, and the cell phone module provides connectivity to cellular telephone networks.

The foregoing description of the specific embodiments will so fully reveal the general nature of the embodiments herein that others can, by applying current knowledge, readily modify and/or adapt for various applications such specific embodiments without departing from the generic concept, and, therefore, such adaptations and modifications should and are intended to be comprehended within the meaning and range of equivalents of the disclosed embodiments. It is to be understood that the phraseology or terminology employed herein is for the purpose of description and not of limitation. Therefore, while the embodiments herein have been described in terms of preferred embodiments, those skilled in the art will recognize that the embodiments herein can be practiced with modification within the spirit and scope of the appended claims.

What is claimed is:

1. An electronic device comprising:

a control module comprising an energy harvesting controller;

a tag module controlled by said control module;

a sensor module controlled by said control module;

an energy harvesting module controlled by said energy harvesting controller and adapted to wirelessly collect ambient radio frequency (RF) electromagnetic (EM) energy; and

a power module adapted to provide a power supply for said control module, said power supply being generated by the collected ambient RF EM energy,

wherein said energy harvesting controller is adapted to (i) assess power requirements of each of the tag and sensor modules, (ii) transfer power from said energy harvesting module to any of said tag and sensor modules, and (iii) charge a local energy storage device.

2. The electronic device of claim **1**, further comprising an antenna module shared by said energy harvesting module and said tag module.

3. The electronic device of claim **2**, wherein said antenna module comprises:

a rectification and impedance matching network;

at least one RF pad operatively connected to said rectification and impedance matching network; and

at least one antenna operatively connected to each of said at least one RF pad.

4. The electronic device of claim **1**, wherein said tag module comprises:

a memory component;

a RF communication interface operatively connected to said memory component;

at least one RF pad operatively connected to said RF communication interface; and

at least one antenna operatively connected to each of said at least one RF pad.

5. The electronic device of claim **1**, wherein said sensor module comprises:

a memory component;

a sensor interface and support sub-system operatively connected to said memory component;

at least one sensor pad operatively connected to said sensor interface and support sub-system; and

at least one antenna operatively connected to each of said at least one sensor pad.

6. The electronic device of claim **1**, wherein said sensor module comprises:

a memory component;

a sensor interface and support sub-system operatively connected to said memory component; and

at least one antenna operatively connected to sensor interface and support sub-system.

7. The electronic device of claim **1**, wherein said control module further comprises:

a radio frequency identification (RFID) tag controller; and

a sensor processing controller operatively connected to said RFID tag controller.

8. The electronic device of claim **1**, wherein said energy harvesting module comprises:

an energy harvesting interface and support sub-system;

at least one RF harvesting pad operatively connected to an energy harvesting interface and support sub-system; and

at least one antenna operatively connected to each of said at least one RF harvesting pad.

9. The electronic device of claim **1**, wherein said energy harvesting module comprises:

an energy harvesting interface and support sub-system;

at least one pair of RF harvesting pads operatively connected to said energy harvesting interface and support sub-system; and

at least one rectenna operatively connected to each of said at least one pair RF harvesting pads.

10. The electronic device of claim **9**, wherein said energy harvesting interface and support sub-system comprises:

a storage component;

a power conditioner component operatively connected to said storage component; and

a combiner operatively connected to said power conditioner.

11. The electronic device of claim **10**, wherein said energy harvesting interface and support sub-system comprises a rectification and impedance matching network operatively connected to said combiner.

12. The electronic device of claim **1**, wherein said power module comprises:

a power charger;

a power source operatively connected to said power charger; and

a voltage regulator operatively connected to said power source.

13. The electronic device of claim **1**, wherein said collected ambient RF EM energy enables said electronic device to be a fully autonomous data collector, alarm generator, event generator, data logger, and data processor independent of any other device.

14. The electronic device of claim **1**, wherein said power supply is generated only by said collected ambient RF EM energy.

15. A system comprising:

an ambient radio frequency (RF) electromagnetic (EM) energy source;

a sensor reader operatively connected to said ambient RF EM energy source; and

a sensor in communication with said sensor reader, said sensor comprising:

a control module comprising an energy harvesting controller;

a tag module controlled by said control module;

a sensor module controlled by said control module;

an energy harvesting module controlled by said energy harvesting controller and adapted to wirelessly collect said ambient RF EM energy; and

a power module adapted to provide a power supply for said control module, said power supply being generated by the collected ambient RF EM energy,

wherein said energy harvesting controller is adapted to (i) assess power requirements of each of the tag and sensor modules, (ii) transfer power from said energy harvesting module to any of said tag and sensor modules, and (iii) charge a local energy storage device.

16. The system of claim **15**, wherein said sensor further comprises an antenna module shared by said energy harvesting module and said tag module.

17. The system of claim **16**, wherein said antenna module comprises:

a rectification and impedance matching network;

at least one RF pad operatively connected to said rectification and impedance matching network; and

at least one antenna operatively connected to each of said at least one RF pad.

18. The system of claim **15**, wherein said tag module comprises:

a memory component;

a RF communication interface operatively connected to said memory component;

at least one RF pad operatively connected to said RF communication interface; and

at least one antenna operatively connected to each of said at least one RF pad.

19. The system of claim **15**, wherein said sensor module comprises:

a memory component;

a sensor interface and support sub-system operatively connected to said memory component;

at least one sensor pad operatively connected to said sensor interface and support sub-system; and

at least one antenna operatively connected to each of said at least one sensor pad.

20. The system of claim **15**, wherein said sensor module comprises:

a memory component;

a sensor interface and support sub-system operatively connected to said memory component; and

at least one antenna operatively connected to sensor interface and support sub-system.

21. The system of claim **15**, wherein said control module further comprises:

a radio frequency identification (RFID) tag controller; and

a sensor processing controller operatively connected to said RFID tag controller.

22. The system of claim **15**, wherein said energy harvesting module comprises:

an energy harvesting interface and support sub-system;

at least one RF harvesting pad operatively connected to an energy harvesting interface and support sub-system; and

at least one antenna operatively connected to each of said at least one RF harvesting pad.

23. The system of claim **15**, wherein said energy harvesting module comprises:

an energy harvesting interface and support sub-system;

at least one pair of RF harvesting pads operatively connected to said energy harvesting interface and support sub-system; and

at least one rectenna operatively connected to each of said at least one pair RF harvesting pads.

24. The system of claim **15**, wherein said energy harvesting interface and support sub-system comprises:

a storage component;

a power conditioner component operatively connected to said storage component; and

a combiner operatively connected to said power conditioner.

25. The system of claim **24**, wherein said energy harvesting interface and support sub-system comprises a rectification and impedance matching network operatively connected to said combiner.

26. The system of claim **15**, wherein said power module comprises:

a power charger;

a power source operatively connected to said power charger; and

a voltage regulator operatively connected to said power source.

27. The system of claim **15**, wherein said collected ambient RF EM energy enables said electronic device to be a fully autonomous data collector, alarm generator, event generator, data logger, and data processor independent of any other device.

28. The system of claim **15**, wherein said sensor reader comprises any of a cell phone, a personal display assistant (PDA), a personal computer (PC) reader, a radio frequency identification (RFID) reader, a broadband wireless fidelity (WiFi) device, and a combination thereof.

29. The system of claim **15**, wherein said power supply is generated only by said collected ambient RF EM energy.

30. The system of claim **15**, wherein said sensor is adapted to transmit a wireless signal to said sensor reader, and wherein said sensor reader is adapted to receive the transmitted wireless signal from said sensor.

31. The system of claim **15**, wherein said collected ambient RF EM energy facilitates transmission of said wireless signal when a selected sensor variable meets a predetermined threshold.

32. The system of claim **15**, wherein said ambient RF EM energy source comprises a wireless fidelity (WiFi) router adapted to allow said sensor reader to communicate over the Internet.

33. The system of claim **15**, wherein said ambient RF EM energy source comprises a hybrid wireless fidelity (WiFi) router transceiver adapted to allow said sensor reader to communicate over the Internet.

34. The system of claim **15**, wherein said sensor is a RF ambient EM energy harvesting sensor and is arranged with a plurality of other said sensors each adapted to collect data for climate control, security alarms, environmental alarms, and information gathering from other devices located within a communicable distance from said sensor, wherein all of said sensors are adapted to be any of locally and remotely monitored via any of the Internet and a cell phone network.

35. A device comprising:

a control module comprising:

an energy harvesting controller;

a radio frequency identification (RFID) tag controller operatively connected to said energy harvesting controller; and

a sensor processing controller operatively connected to said RFID tag controller;

a tag module controlled by said control module;

a sensor module controlled by said control module; and

an energy harvesting module directly connected to said control module and controlled by said energy harvesting controller, wherein said energy harvesting module is adapted to wirelessly collect, process, and store ambient radio frequency (RF) electromagnetic (EM) energy,

wherein said energy harvesting controller is adapted to (i) assess power requirements of each of the tag and sensor modules, (ii) transfer power from said energy harvesting module to any of said tag and sensor modules, and (iii) charge a local energy storage device, and

wherein power for operation of said control module, said tag module, said sensor module, and said energy harvesting module is generated only by the collected ambient RF EM energy.

* * * * *

US 20160301257A1

(19) **United States**

(12) **Patent Application Publication**

PARKS et al.

(10) **Pub. No.: US 2016/0301257 A1**

(43) **Pub. Date: Oct. 13, 2016**

(54) MULTIBAND HARVESTING SYSTEMS AND METHODS INCLUDING SWITCHING NETWORKS

(71) Applicant: **UNIVERSITY OF WASHINGTON**, Seattle, WA (US)

(72) Inventors: **AARON N. PARKS**, Seattle, WA (US); **Joshua R. Smith**, Seattle, WA (US)

(73) Assignee: **UNIVERSITY OF WASHINGTON**, Seattle, WA (US)

(21) Appl. No.: **15/093,459**

(22) Filed: **Apr. 7, 2016**

Related U.S. Application Data

(60) Provisional application No. 62/144,114, filed on Apr. 7, 2015.

Publication Classification

(51) **Int. Cl.**

H02J 50/20 (2006.01)

H04B 1/16 (2006.01)

(52) **U.S. Cl.**

CPC ***H02J 50/20*** (2016.02); ***H04B 1/16*** (2013.01)

(57) **ABSTRACT**

Examples described herein include systems and methods for multiband harvesting. An example system may include a single wideband antenna followed by several narrowband rectifier chains. Each rectifier chain may include a bandpass filter, a tuned impedance matching network, and a rectifier. The outputs of the rectifiers may be combined using a summation network. The summation network, which may be a diode summation network, may in some examples provide good performance even when only a subset o the narrowband harvesters (e.g. a subset of the rectifier chains) is excited.

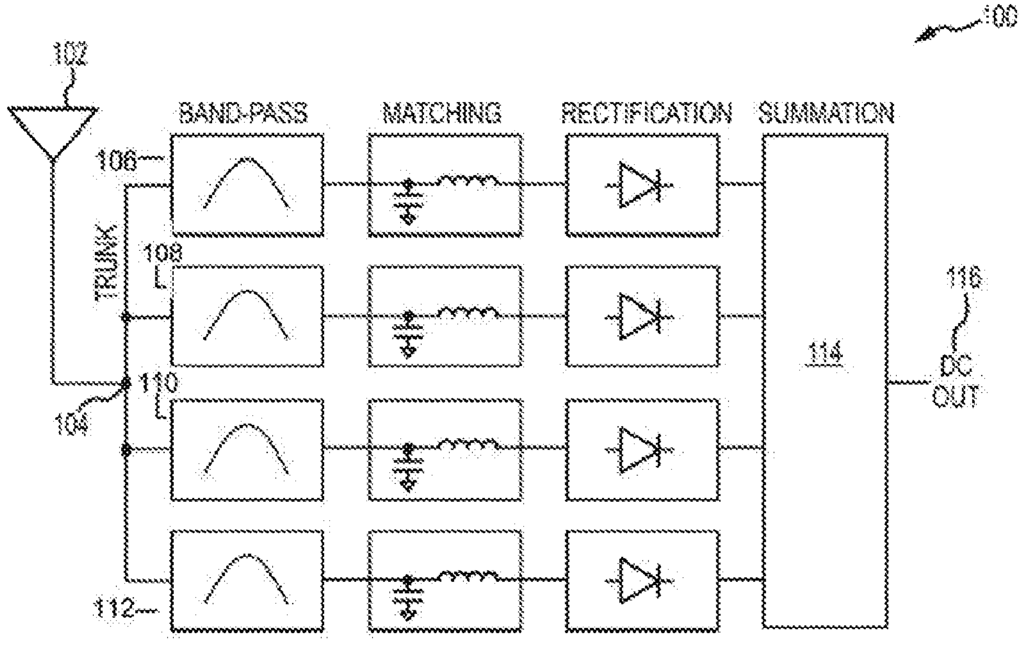

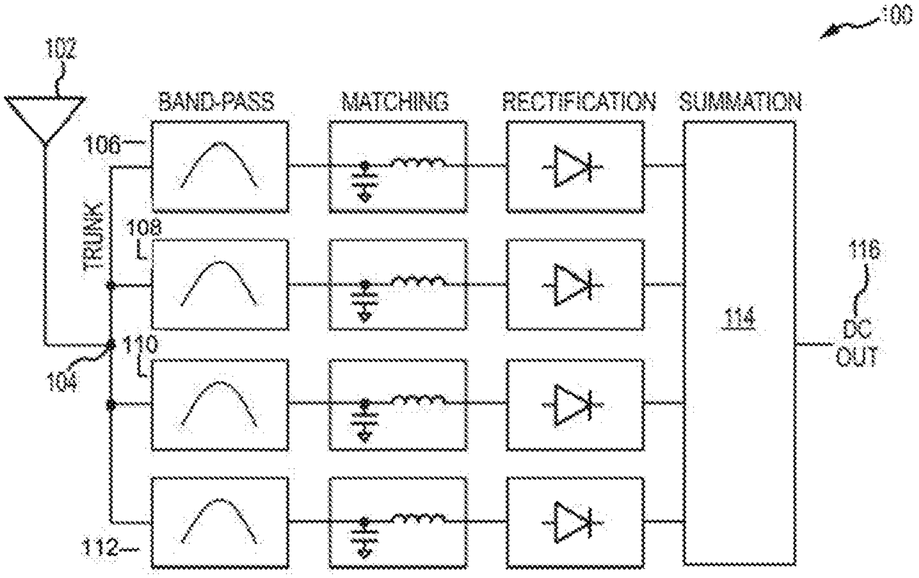
100
102
BAND-PASS
MATCHING
RECTIFICATION
SUMMATION
106
108
TRUNK
110
104
112
114
116
DC
OUT

FIG.1

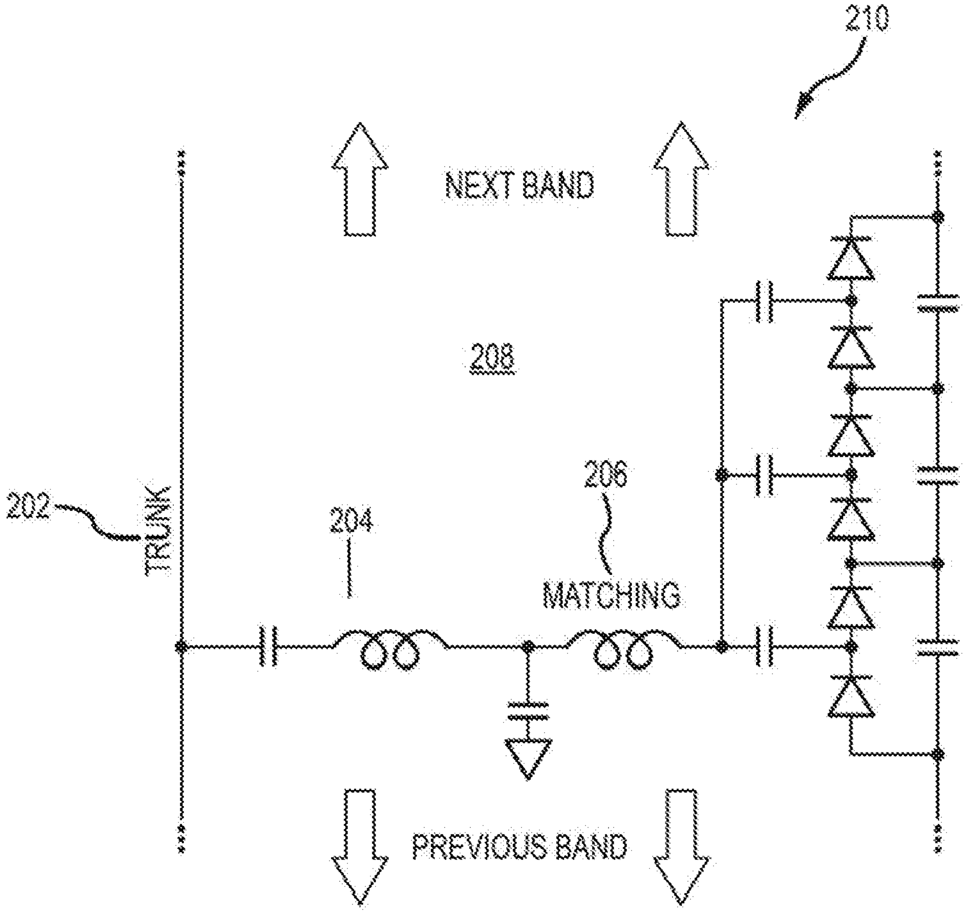
210
NEXT BAND
208
202
TRUNK
204
206
MATCHING
PREVIOUS BAND

FIG.2

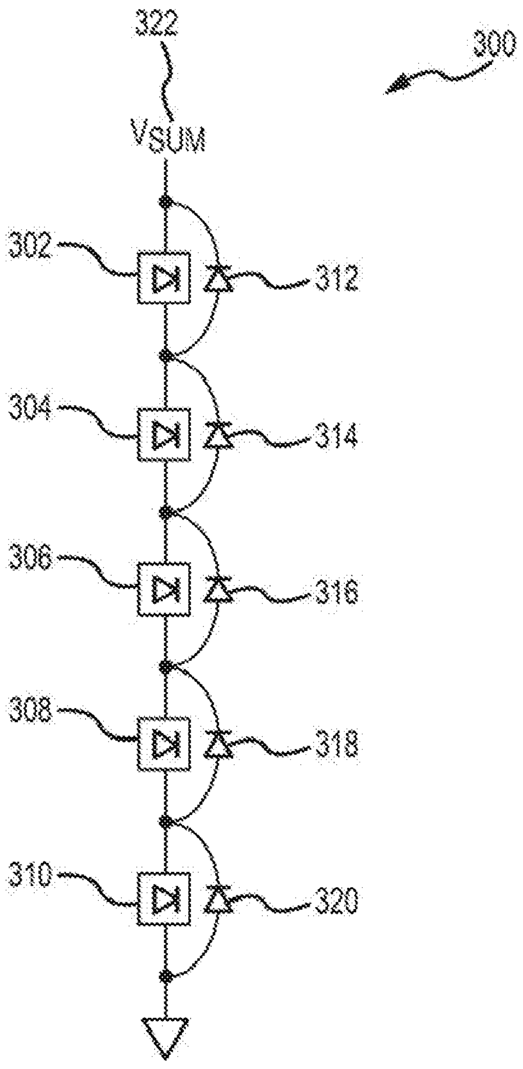
322
300
V_{SUM}
302
312
304
314
306
316
308
318
310
320

FIG.3

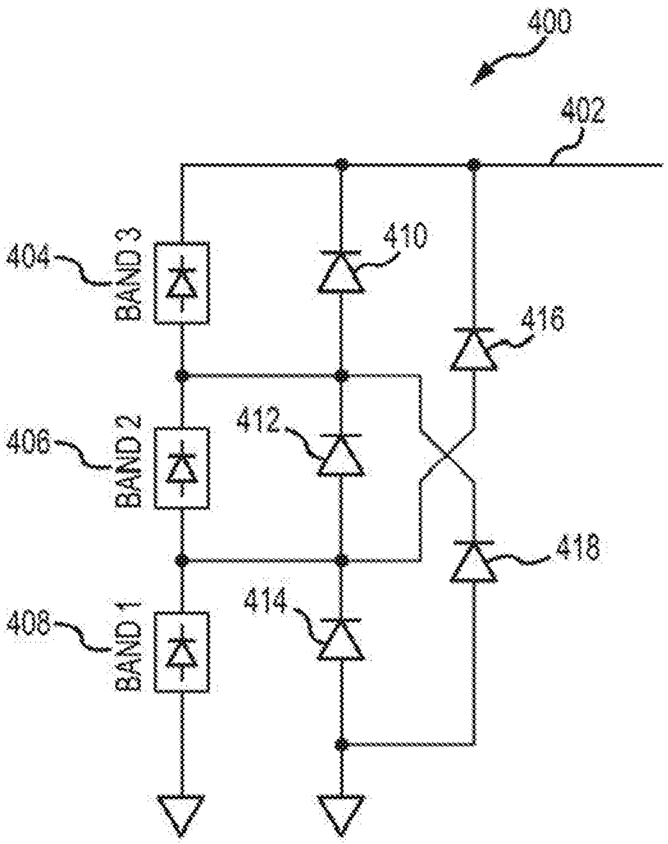
400
402
404
BAND 3
410
416
406
BAND 2
412
418
408
BAND 1
414

FIG.4

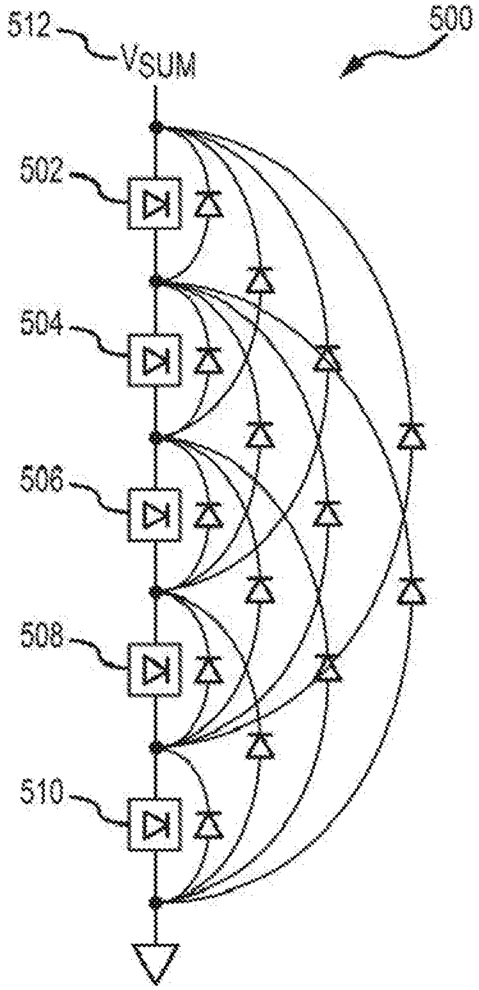
512
VSUM
500
502
504
506
508
510

FIG.5

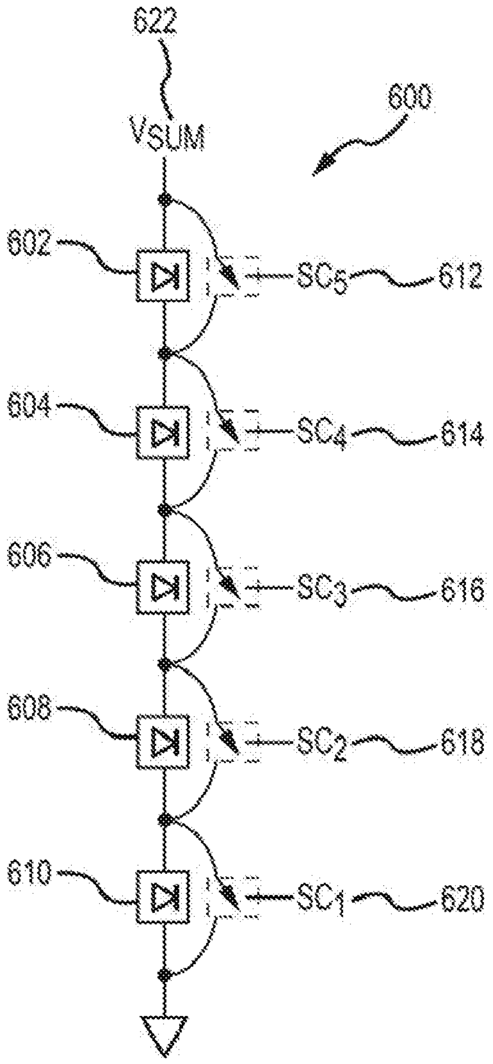
V_{SUM}
SC_5
SC_4
SC_3
SC_2
SC_1

FIG.6

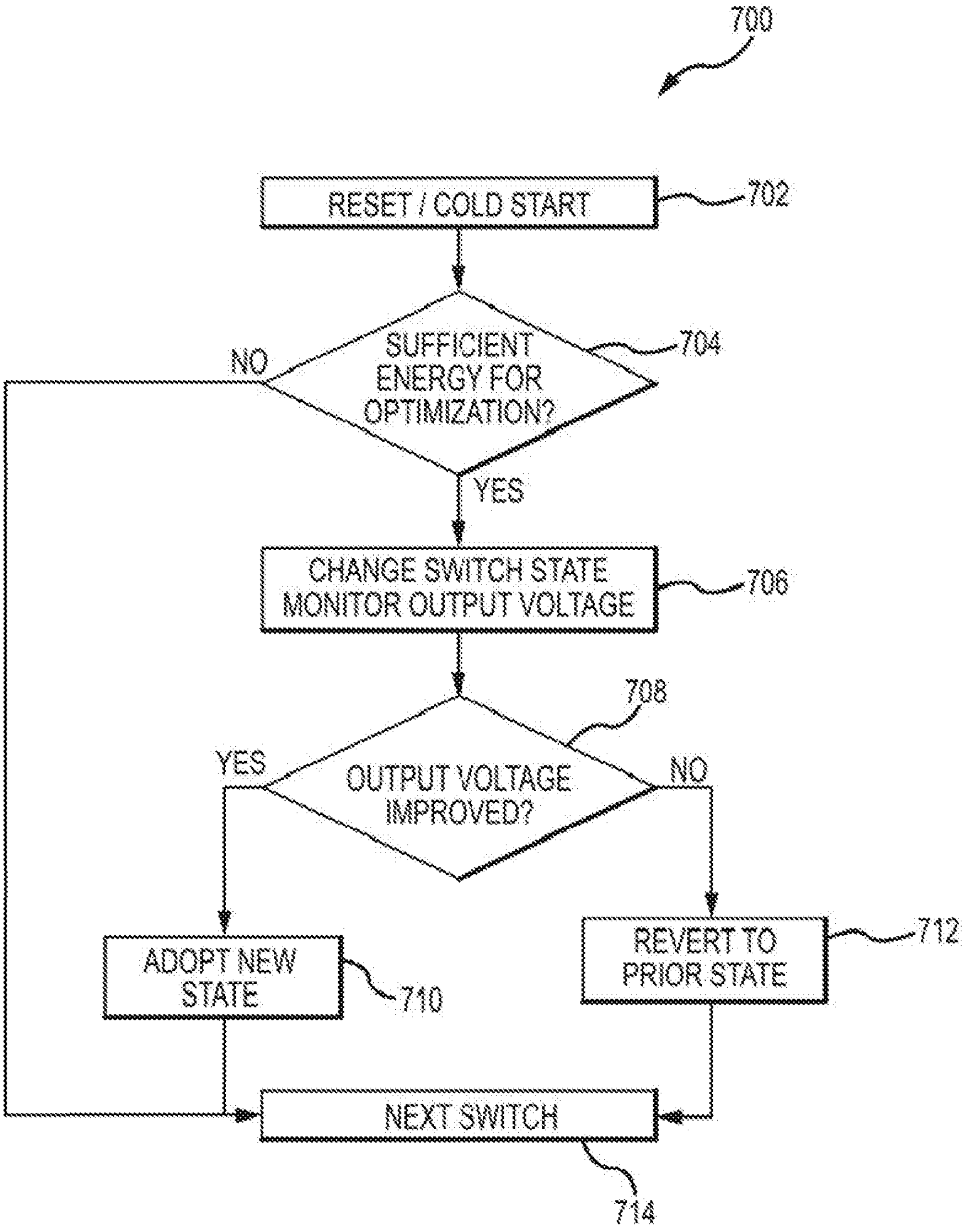
700
RESET / COLD START
702
SUFFICIENT ENERGY FOR OPTIMIZATION?
704
NO
YES
CHANGE SWITCH STATE
MONITOR OUTPUT VOLTAGE
706
708
YES
OUTPUT VOLTAGE IMPROVED?
NO
ADOPT NEW STATE
710
REVERT TO PRIOR STATE
712
NEXT SWITCH
714

FIG.7

MULTIBAND HARVESTING SYSTEMS AND METHODS INCLUDING SWITCHING NETWORKS

CROSS-REFERENCE TO RELATED APPLICATION(S)

[0001] This application claims the benefit under 35 U.S.C. §119 of the earlier filing date of U.S. Provisional Application Ser. No. 62/144,114, filed Apr. 7, 2015, the entire contents of which are hereby incorporated by reference in their entirety for any purpose.

STATEMENT REGARDING RESEARCH & DEVELOPMENT

[0002] This invention was made with government support under CNS-1305072 awarded by the National Science Foundation. The government has certain rights in the invention.

TECHNICAL FIELD

[0003] Examples described herein relate generally to energy harvesting, and examples of wideband harvesters are described.

BACKGROUND

[0004] Harvesting ambient RF power is attractive as a means to operate microelectronics without wires, batteries, or even a dedicated RFID reader. However, previous ambient RF harvesters have been narrowband, making mobile sensing scenarios infeasible: for example, an RF harvester tuned to work in one city will not generally work in another, as the spectral environments tend to differ.

[0005] Ambient radio harvesting is a promising approach to powering battery-free sensing, computing, and communication devices. The vanishingly small amount of power now required by modern microelectronics, along with low power communication techniques, make ambient radio harvesting an increasingly viable power supply option. Compared to solar power, ambient RF has the advantage of being available at night, and is attractive from an industrial design perspective: the antennas already designed in to mobile devices for communication purposes can potentially become a power source, without requiring any changes to the form factor or appearance of the device.

[0006] However, there still exist challenges to widespread adoption of ambient RF harvesting as a power source. Conventional RF harvesting methods are only capable of extracting power from a narrow spectral band. The supply will cease to provide power when its particular source band is not available, either due to geographical fluctuations in spectral occupation, occlusion and shielding (e.g., from the walls of a building), or simply from multipath fading of the ambient signal. When starved of its energy source the device must cease to operate, limiting the application space mostly to a small geographical area with line-of-sight to an ambient radio source of interest.

[0007] Wideband harvesting may capture energy across a large swath of spectrum, but typically results in very low efficiency at any particular source frequency as the quality of the impedance match between the antenna and single rectifier reduces as the bandwidth increases. Tunable harvesting may allow a system to dynamically select a band of interest based on spectral availability, and therefore promises to be able to provide efficient rectification of signals from a single source, regardless of the frequency of that source. However, tunable harvesters in battery free systems may have difficulty bootstrapping to allow the system to coldstart, and also ignore energy outside the band to which they are tuned.

[0008] Existing multiband harvesting generally makes use of multiple antennas, each tuned to a band of interest, and each feeding an independent rectifier through a tuned matching circuit,

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] FIG. **1** is a schematic illustration of a system arranged in accordance with examples described herein.

[0010] FIG. **2** is a schematic illustration of a circuitry branch arranged in accordance with examples described herein.

[0011] FIG. **3** is a schematic illustration of a switching network arranged in accordance with examples described herein.

[0012] FIG. **4** is a schematic illustration of a switching network arranged in accordance with examples described herein.

[0013] FIG. **5** is a schematic illustration of a switching network arranged in accordance with examples described herein.

[0014] FIG. **6** is a schematic illustration of a switching network arranged in accordance with examples described herein.

[0015] FIG. **7** is a flowchart illustrating a method for controlling a switching network with active switches arranged in accordance with examples described herein.

DETAILED DESCRIPTION

[0016] Certain details are set forth below to provide a sufficient understanding of embodiments of the invention. However, it will be clear to one skilled in the art that embodiments of the invention may be practiced without various of these particular details. In some instances, well-known circuits, control signals, timing protocols, signal processing elements or methods, and software operations have not been shown in detail in order to avoid unnecessarily obscuring the described embodiments of the invention.

[0017] Examples described herein include systems and methods for multiband harvesting. An example system may include as single wideband antenna followed by several narrowband rectifier chains. Each rectifier chain may include a bandpass filter, a tuned impedance matching network, and a rectifier. The outputs of the rectifiers may be combined using a summation network. The summation network, which may be a diode summation network, may in some examples provide good performance even when only a subset of the narrowband harvesters (e.g., a subset of the rectifier chains) is excited.

[0018] FIG. **1** is a schematic illustration of a system **100** arranged in accordance with examples described herein. The system **100** includes an antenna **102**, trunk node **104**, circuitry branch **106**, circuitry branch **108**, circuitry branch **110**, circuitry branch **112**, switching network **114**, and output **116**. Each of the circuitry branches includes a respective bandpass filter, respective matching network, and respective rectifier. The antenna **102**, which may be a wideband antenna, is coupled to the trunk node **104**, which in turn is coupled to each of the circuitry branches **106-112**. The circuitry branches **106-112** are coupled to the switching network **114** to sum selected outputs of the circuitry branches and provide a DC voltage at the output **116**,

[0019] The antenna **102** may be a wideband antenna. Generally, the wideband antenna may receive RF signals over a frequency band which is wider than each of the frequency bands filtered by the bandpass filters in the circuitry branches in the system (e.g., circuitry branches **106-112**). Generally, a wideband antenna used to implement the antenna **102** may have approximately the same operating characteristics over the passband of the antenna **102**. The antenna **102** may have a passband that includes frequency bands of interest for energy harvesting (e.g. frequency bands that may be present in an environment where energy harvesting is to be performed) including, but not limited to, cellular frequencies, TV frequencies, and Wi-Fi frequencies. Accordingly, the antenna **102** may receive wireless signals from multiple signal sources, each type of signal source generally configured to transmit wireless signals in a different frequency hand (e,g, short-range transmission. TV broadcast, Wi-Fi transmission). Any signal source may generally be used to harvest energy in accordance with examples described herein, including short-range transmitters (e.g. RFID readers), TV broadcasters, Wi-fi routers, communication devices, and combinations thereof.

[0020] Four circuitry branches are shown in FIG. **1**, however any number of frequency branches may be used in other examples. In some examples, two frequency branches may be used. In some examples, three, four, five, six, seven, eight, nine, or ten frequency branches may be used. Generally, each frequency branch may include a respective bandpass filter configured to pass a particular frequency hand. The frequency hand passed by the respective bandpass filters are each generally narrower than the passband of the antenna **102**. Each (in some examples, at least two) of the respective bandpass filters may he configured to pass a different frequency band. The respective bandpass filters may be orthogonally tuned. The frequency band filtered by the bandpass filter may generally be referred to by specifying its center frequency (e.g. 433 MHz, 915 MHz, 2.45 GHz, in some examples). Examples of frequency bands which may be filtered by the bandpass filters include 433 MHz, 915 MHz, 2.45 GHz in some examples. A range of 400-900 MHz in specified intervals in some examples. Generally, any number and selection of frequencies may be passed, including those in mobile communication. TV, or other bands.

[0021] Each circuitry branch further includes a matching network coupled to the bandpass filter. The matching network may be configured to provide impedance matching for the respective circuitry branch to the respective rectifier in the branch. Each matching network may have one or more resistive elements (e,g, resistors), capacitive elements (e.g. capacitors), and/or inductive elements (e.g. inductors) that provide the impedance.

[0022] Each circuitry branch further includes a respective rectifier. The rectifier is configured to receive the RF signal provided from the matching network and rectify the signal to provide a DC signal. The respective rectifiers may be implemented using respective M stage Dickson charge pumps, which may provide, rectification and voltage amplification in sonic examples,

[0023] Accordingly, each circuitry branch coupled to the antenna **102** (e.g. at the trunk node **104**) may pass a respective frequency band to a matching network and provide the resulting signal to a rectifier. The trunk node **104** may be implemented as a port for the antenna **102**. Multiple rectified signals may be provided, each from a different frequency band. Recall the different frequency bands may correspond to signals received from different environmental frequency sources (e.g., cellular, TV, Wi-Fi).

[0024] The multiple rectified signals may be combined at an output **116**. In some examples, a serial combination (e.g. sum) of the rectified signals is provided, at the output **116**. The output **116** may provide a DC signal (e.g., a DC voltage). The output **116** DC signal may be used to power one or more load circuits (e.g., sensors). Accordingly, the system **100** may be used to harvest energy from many different frequency bands in an environment using a single antenna in some examples.

[0025] Generally, the system **100** may have a number (N) of bandpass filters, and the respective bands may be spaced by frequency deltas (ΔF), and each rectifier may have a respective number of stages.

[0026] Generally, any frequency bands may be selected for the multiband harvester system **100** of FIG. **1**. For example, the system **100** may implement a multi-protocol arrangement targeting several commonly used bands. Each circuitry branch may be designed to filter a particular band. For instance, 433 MHz, 915 MHz, and 2.45 GHz could be filtered by the circuitry branches **106**, **108**, and **110**, respectively. Other frequencies may be used in other examples. Accordingly, knowledge of commonly occupied portions of the spectrum which will be most likely to yield power may be used to select properties of the circuitry branches. This method may achieve high practical efficiency with low complexity, as the number of target bands may be relatively small.

[0027] In some examples, each bandpass tiller may filter a band a set distance from the neighboring bands over a wide bandwidth (e.g. 400-900 MHz in some examples). Accordingly, any number and placement of RF signals may be efficiently harvested in some examples. Accordingly, the energy harvester may capture energy efficiently at nearly any frequency in a larger prescribed bandwidth (e.g., 400-900 MHz in some examples). By allocating circuitry branches for filtering across a complete range, design of the system **100** may not rely on knowledge of commonly occupied bands in some examples, or on current details of regulatory spectral allocation.

[0028] Examples of parallel harvester topologies, such as that shown in FIG. **1**, may provide a good match to one single-port antenna (e.g. Antenna **102**) at multiple bands, at least because a respective matching network may be used for each band. This is in contrast to solutions which may make use of multiple antennas, each covering a particular band of interest, or make use of dual-band antennas with multiple ports and a single rectifier. Therefore, in contrast to existing work, systems described herein may more easily scale in some examples to a large number of frequency bands with no additional antennas or antenna ports required in some examples.

[0029] The bandpass filters and matching networks used in the various circuitry branches may in some examples be tuned for desired performance. Tuning may generally refer to selecting and/or adjusting the parameters (e.g. values) of the various circuit components used m the circuitry branches, including the bandpass filters and the matching networks. In some examples, a cost function is used to optimize the design of the circuitry branches. An example cost function includes minimizing reflected energy in matching the antenna **102** to the trunk node **104** of the circuitry branches. For example, the sum of the return losses (e.g. S11) of the harvester at each frequency band should be minimized. An example cost function may additionally tit instead include minimizing impact of adjacent circuitry branches on impedance of the network at each bandpass frequency of interest. For example, for each center frequency, the impedance of adjacent bands may be maximized,

[0030] In some examples, minimizing a weighted combination of the two above-described cost function factors may be used to select the bandpass filter and match network values for each band of the harvester. For example, overall S11 at each design frequency may be minimized while maximizing adjacent band impedance. For example, considering, a scenario in which only a single continuous wave is presented to the multiband harvester. If overall S11 is simply minimized, the power of this continuous wave may end up distributed over several consecutive rectifiers, resulting in several very poorly excited rectifiers and therefore may be an inefficient system. To also minimize the match quality of adjacent bands generally means that, instead of distributing this continuous wave's power, this power may be focused on one rectifier and thereby maximize or improve the efficiency of the system in harvesting that continuous wave signal. Since most ambient or intentionally radiated signals may be relatively narrow bandwidth compared to the overall scope of the system **100**, most sources may excite only one of the circuitry branches of the harvester and therefore support this optimization strategy. Other optimization strategies may be used in other examples.

[0031] To perform optimization, in some examples, as lumped element model of the harvester may be developed, for example in MATLAB. The model may make use of empirical rectifier impedance measurements taken from implementations of the system **100** in some examples. Each rectifier in the multiband harvester may be characterized at each of the design frequencies (e.g. center frequencies for each of the circuitry branches) in order to achieve a more realistic model of the harvester. A simplifying model of the antenna may be used in which impedance is constant (e.g. 50Ω) across the entire hand of interest, hi some examples, the antenna impedance may be measured across the operating bandwidth and used to better tune the system **100**.

[0032] Approximations for the bandpass filter matching network values for each band may be forward-computed in some examples by assuming each band is independent and selecting the values which minimize S11. The resulting values may be used as the seed values for an optimizing search based on cost functions described herein. The resulting component values would be bandpass and matching network component values which minimize overall S11 at each design frequency while minimizing interaction between adjacent bands.

[0033] Distributed modeling may be used for multiband harvesters with a large band count, which may significantly increase the complexity of modeling, and the difficulty of tuning the harvester.

[0034] In some examples, the system **100** may overcome the Bode-Fano limit, for example, by allowing the match to effectively be distributed across multiple loads. Generally, the Bode-Fano theorem describes the limits of impedance matching between a source and complex load. In conventional single-band RF harvesters or wideband harvesters, the bandwidth over which the antenna can be well matched to a single rectifier will be limited, therefore placing, an upper limit on the site of the hand over which the RF-DC conversion will remain efficient. By allowing the match to be distributed across multiple loads (e.g. multiple circuitry branches), the system **100** may in some examples overcome this limit and achieve improved efficiency over a larger band.

[0035] Note there are at least two impedance interfaces in the system **100** that may be matched to improve efficiency of RF-DC conversion. The first is the interface between the antenna **102** and the trunk node **104** of the multiband harvester. The second is the interface between the trunk node **104** and the rectifier for each of the branches. The Bode-Fano theorem may only place a limit on match bandwidth for complex sources and/or loads. It may model the limitations of the match between the trunk node **104** and the rectifier for each branch, since the rectifier impedance will generally always be complex. However, the antenna-to-trunk interface may be designed such that the impedance looking into each bandpass filter at its resonant frequency may be real and matched to the antenna. Because the adjacent branch bandpass filters may have negligible admittance at any given frequency, they may not contribute significantly to the impedance viewed from the trunk node **104** at adjacent hands. Accordingly, each circuitry branch of the system **100** may be Bode-Fano-constrained, but the antenna-to-trunk match (and therefore the system **100** as a whole) may not be so constrained.

[0036] FIG. **2** is a schematic illustration of a circuitry branch arranged in accordance with examples described herein. The circuitry branch **208** may include trunk node **202**, bandpass filter **204**, matching network **206**, and rectifier **210**. The circuitry branch **208** may be used to implement any of the circuitry branches described herein, including any of those shown in FIG. **1**, for example.

[0037] The circuitry branch **208** includes a trunk node **202**, which may be coupled to an antenna, such as the antenna **102** of FIG. **1**. The circuitry branch **208** may include a bandpass filter **204** coupled to the trunk node **202**. The bandpass filter **204** may filter a particular frequency band, and may include one or more capacitive elements (e.g. capacitors), inductive elements (e.g. inductors), and/or resistive elements (e.g. resistors) selected to filter for the particular frequency band. As shown in FIG. **2**, the bandpass filter **204** includes a capacitor in series with an inductor.

[0038] The bandpass filter **204** is coupled to a matching network **206**. The matching network **206** may include one or more capacitive elements (e.g. capacitors), inductive elements (e.g. inductors), and/or resistive elements (e.g. resistors) selected to match the circuitry branch **208** to the rectifier **210**. For example, the matching network **206** as shown in FIG. **2** includes a capacitor coupled between an output of the bandpass filter **204** and ground, and an inductor coupled between the output of the bandpass filter **204** and an input of the rectifier **210**.

[0039] The rectifier **210** may be implemented using circuitry for rectifying and/or providing voltage amplification. As shown in FIG. **2**, the rectifier **210** may be implemented using an M stage Dickson charge pump.

[0040] As described herein, for example, with reference to FIG. **1**, multiple circuitry branches may be used in systems described herein, and indicated by 'next band' and 'previous band' in FIG. **2**. To sum voltages provided by the rectifiers in each circuitry branch, the outputs of the rectifiers may be connected in series in some examples. In other examples, a parallel combination may be used.

[0041] Accordingly, systems described herein may use a serial connection between independent rectifiers, a topology which, in the context of multiband harvesting, may only provide high efficiency in the case where all bands (e.g. circuitry branches) are excited. If one or more bands in such a serially connected system are not excited, the diode drops in the unexcited bands' charge pumps may need to be overcome by the other bands, resulting in efficiency degradation. For example, unexcited bands' RF-DC conversion diodes may cease to be a source of voltage (e.g. produce an output voltage below a threshold, or no output voltage) at low or zero excitation power (e.g. received power below a threshold) and provide or develop a voltage drop, therefore impeding the flow of current from serially connected excited bands. Summation topologies which improve and/or overcome this effect generally allow current to divert around unexcited bands by providing a path of lesser resistance. Inactive bands accordingly may be shorted using a switching network.

[0042] It is generally difficult to produce semiconductor switches which are normally-closed when no supply voltage is present, a harvester which is capable of cold-starting (starting with all circuit nodes at zero volts) generally may not have normally-closed connections. However, when rectifiers from multiple circuitry branches, as described herein, are connected serially, normally-open switches can be used to short inactive bands and may also behave as open switches during a cold-start. Examples of switching networks described herein may accordingly include normally-open switches to facilitate cold-start of the system.

[0043] The presence of unexcited bands (e.g. circuitry branches) may commonly occur in some examples, for instance where each circuitry branch may be designed to filter and rectify RF signals from a particular type of signal source (e.g. short-range, cellular, TV, Wi-Fi). Those bands may not all always be occupied by signals in the environment. Accordingly, during operation, it may commonly occur that one or more circuitry branches of systems described herein may not be excited at any given time.

[0044] Accordingly, in sonic examples, a switching network may be provided (e.g. the switching network **114** of FIG. **1**), to disconnect certain rectifiers from the output **116**, particularly when the rectifiers are not excited, and would otherwise provide disadvantageous diode drops in a serially-connected output.

[0045] In some examples, the switching network may be implemented using "shortcut" diodes to bypass the circuitry branches which are not active, allowing an automatic optimal configuration to be achieved in summing the voltages produced by each band rectifier. These shortcut diodes my advantageously have both low threshold and low leakage in some examples, attributes which may be achievable as they will be operating at DC.

[0046] In some examples, "dead weight" diodes associated with un-excited circuitry branches may be bypassed with a single, low threshold, low leakage DC diode. Such topology may reduce any extra overhead of unexcited bands, but may still have overhead in some examples if there are many unexcited hands with many bypass diodes for current to traverse.

[0047] FIG. **3** is a schematic illustration of a switching network arranged in accordance with examples described herein. The switching network **300** shown in FIG. **3** is for use with five serially-connected circuitry branches, but any number of circuitry branches may be used. The circuitry branches are represented in FIG. **3** by their rectifiers—rectifier **302**, rectifier **304**, rectifier **306**, rectifier **308**, and rectifier **310**. The switches in the switching network **300** are implemented using diodes. Other switching elements may be used in other examples, including reversible fuses.

[0048] In the example of FIG. **3**, one diode is provided for each circuitry branch to provide a bypass path for current when the rectifier is not producing sufficient voltage. Each diode is accordingly connected across one or more output diodes of the rectifier to provide a bypass path around the output diodes. For example, each diode may be connected form an output of a previous band rectifier to the output of the next band rectifier.

[0049] As shown, the diode **320** is connected between ground and an output of rectifier **310**. Diode **318** is connected between an output of rectifier **310** and an output of rectifier **308**. Diode **316** is connected between an output of rectifier **308** and an output of rectifier **306**. Diode **314** is connected between an output of rectifier **306** and rectifier **304**. Diode **312** is connected between an output of rectifier **304** and output **322**, the output node of the overall system providing voltage V_{sum}.

[0050] When any of the rectifiers **302**-**310** are providing insufficient voltage to maintain the respective diode in an open state, the respective diode will provide a lower resistance current path, bypassing the output diodes of the rectifier, and avoiding presenting the diode drops associated with those output diodes to the output voltage V_{sum}. Other example switching networks may be able to reduce a number of bypass diodes used when multiple circuitry branches are unexcited. In FIG. **3**, generally one bypass diode was used per unexcited circuitry branch.

[0051] A voltage drop associated with bypass diodes in use, may however adversely affect the output voltage V_{sum}.

[0052] In some examples, "shortcut" diodes may be placed between each serially-connected band output and the two output rails (e.g. the output and ground). This may reduce the efficiency hit caused by the bypass diodes i, but may not achieve optimal power combination in all cases. For instance, if the first and second bands are excited the output voltage may be near their sum (minus diode drops), but if the first and third bands are excited the output voltage may be as low as the maximum of the two band outputs, effectively throwing away the benefit of the second excited band.

[0053] FIG. **4** is a schematic illustration of as switching network arranged in accordance with examples described herein. The switching, network **400** is shown for use with three circuitry branches (e.g. bands), however in other examples switching networks may be provided for other numbers of circuitry branches, including the four circuitry branch example of FIG. **1**. In FIG. **4**, each circuitry branch is represented by their respective rectifier—rectifier **404**, rectifier **406**, and rectifier **408**.

[0054] The example of FIG. **4** may allow the output **402** to attain the sum voltage of all excited bands, regardless of whether the excited bands are contiguous or scattered throughout the harvester array. As shown in FIG. **4**, each node in the rectifier stack is connected to every other node with a diode, generally a properly oriented, low threshold and low leakage DC diode.

[0055] For example, the output of rectifier **408** is connected to the output of the rectifier **406** by diode **412**, to output **402** by diode **416**, and to ground by diode **414**. The output of rectifier **406** is connected to the output **402** by diode **410** and to ground by diode **418**.

[0056] If only rectifier **408** is not excited (e.g. insufficient energy has been rectified by the circuitry branch, such that voltage at an output of the rectifier is below a threshold) while rectifiers **406** and **404** are excited, the voltage at the output of the rectifier **408** may be such that diode **414** becomes a short, while the voltages at the outputs of rectifier **406** and rectifier **404** are such that the diodes **412** and **410** are open. In this manner, the output **402** may be a sum of the voltages provided by the rectifier **406** and rectifier **404**, minus a voltage drop provided by the bypass diode **414**. In this manner, the rectifier **408** has been disconnected from the output **402** by diode **414**.

[0057] If rectifier **406** is not excited while rectifier **404** and rectifier **408** are excited, the voltage across diode **412** may cause diode **412** to short, disconnecting the rectifier **406** from the output **402**. However, the voltages at the outputs of the rectifier **408** and the rectifier **404** are such that the diodes **410** and **414** are open. In this manner, the rectifier **406** has been disconnected from the output **402** and the output **402** may provide a voltage equal to the sum of the outputs of rectifier **404** and rectifier **408** less the voltage drop provided by diode **412**.

[0058] If rectifier **404** is not excited while rectifier **406** and rectifier **408** are excited, the voltage across diode **410** may cause diode **410** to short, disconnecting the diode **410** from the output **402**. However, the voltages at the outputs of the rectifier **406** and the rectifier **408** are such that the diodes **412** and **414** are open. In this manner, the rectifier **404** has been disconnected form the output **402** and the output **402** may provide a voltage equal to the sum of the outputs of rectifier **408** and rectifier **406**, less the voltage drop provided by diode **410**.

[0059] If both rectifier **408** and rectifier **406** are not excited while rectifier **404** is excited, the voltage across diode **418** may be such that diode **418** is considered a short, effectively disconnecting the rectifier **408** and the rectifier **406** from the output **402**. The voltage at the output **402** may be the voltage provided by the rectifier **404** less the voltage drop of the diode **418**.

[0060] If both rectifier **404** and rectifier **406** are not excited while rectifier **408** is excited, the voltage across diode **416** may be such that diode **416** is considered a short, effectively disconnecting the rectifier **404** and rectifier **406** from the output **402**. The voltage at the output **402** may be the voltage provided by the rectifier **408** less the voltage drop of the diode **416**.

[0061] If both rectifier **408** and rectifier **404** are not excited while rectifier **406** is excited, the voltages across diode **410** and diode **414** may be such that diode **410** and diode **414** are considered shorts, but the diode **412** is open, disconnecting the rectifier **408** and rectifier **404** from the output **402**. The voltage at the output **402** may be the voltage provided by the rectifier **406** less the voltage drop of the diode **410** and diode **414**.

[0062] Accordingly, the output **402** in the example of FIG. **4** may be a sum total of voltages provided by rectifiers in all excited circuitry branches, minus some number of diode drops which depends on the number and arrangement of bands excited. When multiple bands are not excited, they may be in some cases bypassed with a fewer number of diodes than the number of excited bands (e.g. one diode **418** may be used to bypass rectifier **408** and rectifier **406**). In situations when all hands are excited, the matching network of FIG. **4** may perform equivalently or nearly equivalently to the approach of simple serial combination of bands. When only select bands are excited and others left unexcited, the matching network of FIG. **4** may perform more efficiently than the simple serial approach.

[0063] The leakage current of so many DC shortcut diodes may be considered when implementing a multiband harvester with a large number of bands. However, while the total number of DC shortcut diodes in the system can grow large as the number of bands increases, the leakage may not scale quite as geometrically, as the average of the reverse voltage across a diode in the system will decrease as the number of hands increases (given some number and arrangement of excited bands).

[0064] FIG. **5** is a schematic illustration of a switching network arranged in accordance with examples described herein. The strategy used for the switching network **500** of FIG. **5** is similar to that used in FIG. **4**, however the switching network **500** of FIG. **5** is shown with five circuitry branches—represented by rectifier **502**, rectifier **504**, rectifier **506**, rectifier **508**, and rectifier **510**.

[0065] The output of each rectifier is connected to the output of every other rectifier in the serially chain with a diode. Accordingly, a total of 14 diodes are used in FIG. **5**. A single diode may be used to bypass any number of neighboring rectifiers (e.g. disconnect the rectifiers from the output **512**). For example, a single diode may be used to bypass rectifier **510**, rectifier **508**, rectifier **506**, and rectifier **504** when those corresponding circuitry branches are all unexcited, and only the circuitry branch associated with rectifier **502** is excited.

[0066] FIG. **6** is a schematic illustration of a switching network arranged in accordance with examples described herein. The switching network **600** is shown for use with a system having five circuitry branches, represented in FIG. **6** by their rectifiers—rectifier **602**, rectifier **604**, rectifier **606**, rectifier **608**, and rectifier **610**. In the example of FIG. **6**, active switches are used to bypass un-excited circuitry branches (e.g. disconnect the rectifiers of the unexcited branches from the output).

[0067] The active switch **612** is provided to bypass rectifier **602** and is connected between the output of rectifier **604** and output **622** of the system. The active switch **614** is provided to bypass rectifier **604** and is connected between the output of rectifier **606** and the output of rectifier **604**. The active switch **616** is provided, to bypass rectifier **606** and is connected between the output of rectifier **608** and the output of rectifier **606**. The active switch **618** is provided to bypass rectifier **608** and is connected between the output of rectifier **610** and the output of rectifier **608**. The active switch **620** is provided to bypass rectifier **610** and is connected between ground and an output of rectifier **610**.

[0068] The active switches used in examples described herein may be implemented using a variety of components, including but not limited to transistors and reversible fuses.

[0069] Generally, the active switches each have a control terminal for receipt of a respective control signal, and may close responsive to receipt of an active control signal. In some examples, however, normally closed switches may additionally or instead be used.

[0070] A diode switching network may be present in some examples together with a network of active switches. For example, the switching network **500** of FIG. **5** may be connected in parallel with the switching network **600** of FIG. **6**. In that manner, the diode switching network may be used during cold-start of a system—e.g. when a system is in a startup mode. During cold-start or a start-up mode, it may be undesirable or impractical to control active switches, however passive diodes may supply the bypass functionality described herein. Once startup is achieved and active control is possible (e.g. during normal operation), the bypass functionality may be performed using the active switches.

[0071] FIG. **7** is a flowchart illustrating to method for controlling a switching network with active switches arranged in accordance with examples described herein. The method **700** may be perforated by as controller, logic circuitry, or other computing system provided for control of switching networks including active switches described herein. The method **700** may be implemented in hardware, software, firmware, or combinations thereof.

[0072] The method **700** may begin with a reset or cold start of a system in block **702**. For example, when a system (e.g. the system in any of the Figures shown herein) first receives power or first receives power after a period of time has elapsed without power, it may be in a startup mode (e.g. cold start). In some examples, where a complete or partial passive switching network is present (e,g, using diodes), the diodes may be used to bypass unexcited circuitry branches (e.g. disconnect their rectifiers from an output node).

[0073] In decision block **704**, a determination is made whether the system has sufficient energy for optimization. The controller, logic circuitry, or other computing system implementing the method may monitor available energy, e.g. by monitoring an output voltage, energy stored on a capacitor, or other indicia of available energy. The energy reservoir being monitored may be connected, for example, after voltages from rectifiers in multiple circuitry branches have been summed, for example after a switching network (e.g. after switching network **114** of FIG. **1**). If the available energy is above a threshold for optimization, then the method may proceed. If not, the method may end, or may move to consider a next switch in the system (e.g. Block **714**).

[0074] If there is sufficient energy for optimization, a state of one active switch in the system may be changed in block **706**. Following the change, the outwit voltage of the system may be monitored. Referring to FIG. **6** for example, the active switch **620** may be changed from an open to a closed state, and the output **622** may be monitored. If the output **622** improves (e.g. a greater amount of voltage is available after changing the state of the active switch **620**), the new state may be adopted in block **710**. If the output **622** does not improve (e,g, a same or lesser amount of voltage is available after changing the state of active switch **620**), the controller logic circuitry, or other computing system performing the method **700** may revert the active switch **620** to the prior state (e.g. open).

[0075] In block **714**, the method is repeated for a next switch in the system. For example, an energy check may be performed to confirm sufficient energy for optimization, the switch state may be changed and the output voltage monitored, as described in blocks **704**, **706**, **708**, **710**, and **712**. These blocks may be repeated for each active switch in the switching network (e.g. each active switch shown in FIG. **6**). Once all switches have been considered, in some examples, the system will repeat the process again for all switches. In this manner the state of all switches may be considered continuously during operation of the system. In some examples, once all switches have been considered, the control system may enter a sleep state and perform optimization only from time to time after the first optimization, or not until another reset or cold start in some examples.

[0076] Examples described herein may provide RF energy harvesting which may be suitable for mobile applications (e.g. to power a cell phone, other mobile computing system, and/or sensors). Example energy harvesters described, herein may be used to power ambient RF-powered data logging sensors that may, for example, upload data to RFID readers when in range.

[0077] An integrated multiband system with a single antenna port may be used as a universal RF energy harvesting solution. Systems may be able to be used in a plug-n-play fashion with a wide variety of antennas and across a wide variety of target frequencies.

[0078] Examples of systems (e.g. harvesters) described, herein may provide benefits in some examples. Example benefits are described herein to facilitate understanding. It is to be understood that not all examples may have all, or even any of the described benefits. Some example benefits may include (1) RF source flexibility (2) access to additional power and/or (3) improved sensitivity. RF source flexibility generally refers to the harvester being able to operate using any subset of several potential source frequencies. The benefit of multiband harvesting generally provides access to additional power because the harvester can collect energy from multiple RF sources simultaneously. Another benefit may be that multiband harvesting can provide improved sensitivity in circumstances in which no one source provides enough power to operate the device, but several sources in combination do.

[0079] RF source flexibility may be significant because it may enable mobile RF harvesting: the frequency of the strongest RF source may not generally be the same in each city; examples of multiband harvesters described herein may allow as device to capture power from the strongest sources it encounters in the environment. The summation networks described herein may be advantageous for RF source flexibility. A multiband harvester without the switching networks may require simultaneous excitation in all or most of its input hands. While such a harvester might provide benefits such as increased total power and increased sensitivity, it would not provide RF source flexibility. One can imagine providing RF source flexibility with an actively tunable harvester design, but the active circuitry would typically cost substantial power, resulting, in a system with low net efficiency.

[0080] The efficient multiband harvesting described, herein can enable mobile RF harvesting, increasing the practicality of ambient RF as a power source. Applied in the context of RFID, this can enable a sensor tag that uses ambient RF power to sense and to data, and then downloads the data when interrogated by an RFID reader.

[0081] Example energy harvesting systems described herein may be utilized in a backscattering transponder. Such a device may operate both as a passive RFID tag and as an ambient backscatter transponder, capable of operating from both "planted" energy sourced by an RFID reader, and with "wild" RF energy from existing signals such as those used in Wi-Fi, cellular, and TV transmissions. Such a backscattering transponder could communicate orthogonally at each of the design frequencies of the multiband harvester by performing impedance modulation separately for each band, and may be able to achieve communication diversity and multiplexing gains which could result in higher rates, longer ranges, and improved robustness.

[0082] Example energy harvesting systems described herein may be utilized in stand-alone sensor devices—e.g. Internet-of-things devices. Sensor devices that may utilize energy harvesting systems described herein include, but are not limited to, home sensors where energy harvesting systems described herein may eliminate or reduce a need to change the home sensor batteries.

[0083] Example energy harvesting systems described herein may be utilized in existing devices to provide additional functionality without depleting traditional power sources. For example, energy harvesting systems described herein may be utilized on a cell phone, tablet, laptop, or other device to provide minimum functionality when the battery is depleted (e.g. for use in locating the device and/or providing 'always on' for faster wake-up of the device, and/or maintaining state of the device).

[0084] Example harvesting systems described herein (e.g. circuitry branches and switching networks) may be implemented as a system on a chip or multiple chips in some examples). The chip may have a port for connection to an antenna (e.g. a wideband antenna). Example devices (e.g. cell phones, tablets, appliances, sensors, backscattering transponders) may include such an example chip and use it to harvest energy from the environment and power one or more functionalities of the device.

EXAMPLES

[0085] A SPICE model of a multiband harvester was constructed and subjected to some virtual experimentation. First, a single band model was constructed using Avago HSMS-285C diode models in a 3-stage Dickson charge pump arrangement. The impedance of the single band charge pump was measured at each of the design frequencies for the system. Eight bands in the region of the spectrum were selected: 300 MHz, 356 MHz, 423 MHz, 503 MHz, 597 MHz, 709 MHz, 842 MHz, and 1 GHz. A first order LC bandpass fiber was generated at each design frequency (with a Q of 5) and used to isolate the eight bands. A low-pass L-match network was then constructed for each band, to match the 50Ω source to the load impedance of each band's rectifier at its design frequency. All values were forward-computed based on initial simulated impedance measurements; no iteration was performed once initial component values had been selected,

[0086] A diode summation network was also constructed using the SPICE model for the SDM03U40 Schottky diode manufactured by Diodes. Inc., and used to stun the DC output power from each of the eight bands. The output of the summation network was connected to a 100 kΩ load resistor to simulate a device being powered.

[0087] The RF-DC Version efficiency as a function of frequency tier a single-tone excitation indicated the intended design frequencies aligned well with the conversion efficiency peaks, though there was variation in the conversion efficiency between bands.

[0088] To characterize the effectiveness of the summation network and its impact on the multiband harvester, a series of tests were done on two systems: One with the summation network in place, and one with only a simple serial connection between bands. The two systems were otherwise identical. Eight simulated voltage sources were connected serially to generate up to eight excitation tones. The eight sources were modulated in a binary-weighted fashion, turning on or off each source to ultimately produce every possible combination of excited bands (eight bands yields 256 permutations). The source power per band was set to +3 dBm for this simulation. A transient simulation was performed for each combination of excited bands, and the summed output voltage was measured after a fixed stabilization period had expired.

[0089] Power from multiple bands combined somewhat linearly, with each additional hand contributing approximately 750 μW to the sum output power on average. Note that for each permutation of excited bands, the output power may vary for two reasons: The first is that the RF-DC conversion efficiency for each band may not be equal. The second is more subtle; the diode summation network may exhibit a different operating efficiency depending on the arrangement (not simply the number) of excited bands.

[0090] This arrangement-dependent summation efficiency can be shown by comparing the case in which only bands **1**, **3**, and **5** are excited to the case in which only bands **1**, **2**, and **3** are excited. In the former case, multiple shortcut diode drops must be overcome between the excited bands, while in the latter case the bands are adjacent and therefore sum serially with no diode drop between them.

[0091] On average the summation network provided a significant benefit in terms of delivered power), though there are a number of test eases in which the presence of the summation network reduces overall efficiency. This may be attributed to the reverse leakage current of the summation diodes, which may become a significant factor with the large number of diodes required to implement the network.

[0092] A 2-band and 5-band prototype were implemented with discrete components. The prototypes use 3 stages per band and the described "shortcut" summation topology, and were connected to a wideband log-periodic antenna with a roughly 6 dBi gain when the experiment required. The two-band harvester design frequencies are 539 MHz and 915 MHz, and the fiveband harvester design frequencies are 400 MHz, 480 MHz, 590 MHz, 735 MHz, and 900 MHz. Avago HSMS-285C Schottky; diodes were used in the 3-stage RF-DC conversion charge pumps. The summation network made use of low-threshold and low-leakage SDM03U40 diodes manufactured by Diodes, Inc.

[0093] Tuning of the prototypes were accomplished by first installing the bandpass filters for every band, shorting the L-match network for every band, and measuring RF port impedance at each of the design frequencies with a test power of −10 dBm. The impedance measurements made with the shorted matching network were used to estimate matching component values. Matching was then done empirically for each and in turn, with some iteration due to the impact of each band on the consecutive bands.

[0094] During the following tests to characterize the performance of the harvester. A 100 kΩ resistive load was placed at the output of the summation network.

[0095] A. Single-Tone Response

[0096] The 2-band and 5-band prototypes were first subjected to a single-tone excitation at a power level of −10 dBm (100 μW), and the S11 (reflected power) and RF-DC conversion efficiency were measured.

[0097] During the tuning procedure, the minimization of S11 was the only goal, and this was a potential cause for the decreased efficiency observed in the 5-band harvester: In this system, S11 was not a good predictor for RFDC conversion efficiency. While the minimal S11 and maximal RF-DC efficiency points for the 2-band prototype matched well, the 5-hand harvester did not exhibit the same correlation for every band. For instance, while the fourth band's S11 showed a very good impedance match to the 50Ω source, the RFDC conversion efficiency was extremely poor compared to the other hands. The discrepancy between S11 and RF-DC, efficiency in the 5-band prototype may be due to the large amount of interaction between bands; in that prototype the quality factor of the bandpass filter was low and therefore adjacent hands may have had a strong interaction.

[0098] An experiment was performed to characterize the multi-band excitation response of the 2-band prototype. The goal of this experiment was to quantify how well the harvester can combine power from multiple hands.

[0099] B. Multi-Excitation Response

[0100] The multi-excitation experiment consisted of a two-hand excitation produced by combining the output of two RF function generators. The RF-DC conversion efficiency of the entire system increased as more hands were excited.

[0101] C. In the Wild

[0102] The 5-band harvester was subjected to two scenarios in which "wild" RF energy was present. In the first scenario, it was placed in the field of a UHF RFID reader emitting 30 dBm at 915 MHz into a 6 dBic circularly polarized antenna. At a distance of 5 m from the RFID reader, the harvester was able to produce 2.3 V across a 100 kΩ load, for a delivered power of 53 μW.

[0103] In the second scenario, the 5-hand harvester was placed at a location 4.2 km distant from a TV broadcast tower emitting 1 MW at 539 MHz, and was able to produce 2.5 V across the same 100 kΩ load for a delivered power of 62.5 μW. These "wild" scenarios appeared to indicate that, despite mediocre conversion efficiency when presented with a 50Ω bench source, the antenna-connected harvester appears to work serviceably with single-band excitation from two very different ambient RF sources.

[0104] Switching Networks

[0105] Models of each of the three switching networks shown in FIGS. **3**, **5**, and **6** were constructed. A 5-band harvester model was used as a source in a simulation testbench. The goal of modeling was to characterize the summation benefit of each method over a naive serial combination of DC harvester outputs as a function of the excitation state of the 5-band harvester.

[0106] A SPICE model for the HSMS-282 diode used in serial summation was used as follows:

[0107] D(IS=4.8E−8, CJO=0.649E−12, VJ=5.6, BV=26.7, IBV=10E−4, EG=0.69, N=1.067, RS=7.8, XT1=2, M=0.5)

[0108] A. Test Setup: Multiexcitation Testing

[0109] Two fixed-resistance loads were used in each test scenario, 10 kΩ and 100 kΩ, and represent typical μW-scale loads that may be experienced in an energy harvesting device. A load capacitance of 10 nF was used. Two excitation power levels were used in each test scenario, −10 dBm and −20 dBm. In the multiexcitation tests, each excited band for a particular test was subjected to this excitation power and the other frequency bands were left unexcited. Every possible permutation of excited bands was tested for each data set, For each data point in each set, a transient simulation was allowed to run until a stable output voltage was reached, and the power across the load was then computed.

[0110] Low forward voltage and low reverse saturation current both must be achieved for summation network diodes to have a positive benefit. In all tests the HSMS-282 Schottky barrier diodes were used for summation as they have very low advertised forward voltage while maintaining low reverse saturation current. The parameters of the diode models used in the diode summation network are given in the SPICE model above.

[0111] In the modeled switching network including, active switches, bands left unexcited were bypassed by switches. The switch model used a 50Ω 'on' resistance and a 10 MΩ 'off' resistance, conservatively chosen values for low power analog switches such as the ADC1802 operating at low voltages. Switch power consumption for the ADG802 is under 10 nW per device. To estimate the overhead of the control logic, a TI MSP430G2553 microcontroller power consumption in low power mode (LPM3, with a crystal oscillator active) was assumed. 1. The total power consumption of switches and control logic is just under 1 μW, and all results involving intelligent switching reflect this overhead.

[0112] The diode summation network of FIG. **5** generally provided a benefit excepting the ease in which all five bands were excited. The simple diode summation network of FIG. **3** also reliably provided a benefit, but in general was less beneficial over the complete test space than the full diode summation network.

[0113] The intelligent switching summation network including active switches, however, provided a strong benefit at higher input power but was less advantageous and even became an impediment as excitation power dropped. This decrease in efficacy with very low excitation power is a symptom of the power draw of the logic system for the intelligent switching network, and therefore is dependent on specific implementation.

[0114] From the foregoing it will be appreciated that although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention.

What is claimed is:

1. A system comprising:

a wideband antenna configured to receive wireless signals from at least a first type of signal source configured to transmit in a first frequency band, and at least a second type of signal source configured to transmit in a second frequency band, different from the first frequency band;

a first circuitry branch including:

a first bandpass filter configured to pass at least a portion of the first frequency band;

a first matching network coupled to the first bandpass filter; and

a first rectifier coupled to the first matching network:

a second circuitry branch including:

a second bandpass filter configured to pass at least a portion of the second frequency band:

a second matching network coupled to the first bandpass filter; and

a second rectifier coupled to the first matching network, wherein outputs of the first and second rectifier are configured to couple in series to an output;

a first switch positioned to disconnect the first rectifier from the output when a voltage provided by the first rectifier is below a first threshold; and

as second switch positioned to disconnect the second rectifier from the output when a voltage provided b the second rectifier is below a second threshold.

2. The system of claim **1** wherein the first and second switches comprise diodes.

3. The system of claim **1** wherein the first and second switches comprise active switches.

4. The system of claim **3** wherein the active switches comprise normally-open switches.

5. The system of claim **1** wherein the first and second switches comprise diodes and wherein the system further includes a first active switch positioned to disconnect the first rectifier from the output responsive to a first control signal and a second active switch positioned to disconnect the second rectifier from the output responsive to a second control signal.

6. The system of claim **5** wherein the system is configured to have the first and second diodes provide control during startup of the system, and the first active switch and the second active switch provide control during normal operation of the system.

7. A system comprising:

a wideband antenna configured to receive wireless signals over a first frequency range;

a plurality of circuitry branches, each of the plurality of circuitry branches including:

a respective bandpass filter to pass a different respective frequency range, each different respective frequency range being narrower than the first frequency range; and

a respective rectifier coupled to the respective bandpass filter, the respective rectifier configured to rectify a signal received from the respective bandpass filter;

wherein an output of a plurality of respective rectifiers are configured to be coupled in series to an output; and

a switching network coupled to the output and configured to disconnect selected one or ones of the plurality of respective rectifiers from the output

8. The system of claim **7** further comprising a load circuit coupled to the output, the load circuit configured to receive power from the output.

9. The system of claim **8** wherein the load circuit comprises a backscattering transponder.

10. The system of claim **8** wherein the switching network comprises a diode network.

11. The system of claim **8** wherein the switching network comprises a network of active switches.

12. The system of claim **8** wherein the switching network comprises a diode network and a network of active switches, wherein the diode network is configured to disconnect selected one or ones of the plurality of respective rectifiers from the output during a startup mode, and wherein the network of active switches is configured to disconnect selected one or ones of the plurality of respective rectifiers from the output after the startup mode.

13. The system of claim **7** wherein component values in at least one of the plurality of circuitry branches are selected based, at least in part, on minimizing reflected energy at the wideband antenna at the respective frequency range and maximizing impedance of circuitry branches adjacent the at least one f the plurality of circuitry branches at the respective frequency range.

14. The system of claim **7** wherein an impedance looking into each respective bandpass filter is real and matched to the wideband antenna.

15. A method comprising;

receiving wireless signals at an antenna;

filtering the wireless signals into at least two frequency bands to provide at least two filtered signals;

rectifying each of the at least two filtered signals to provide respective rectified signals; and

combining selected ones of the respective rectified signals to provide an output voltage.

16. The method of claim **15** wherein said combining selected ones of the respective rectified signals comprises selecting the ones of the respective rectified signals having an amplitude greater than a threshold.

17. The method of claim **16** wherein said selecting comprises activating switches to disconnect other ones of the respective rectified signals from an output.

18. The method of claim **15**, wherein the wireless signals include cellular and TV signals, and wherein the at least two frequency bands comprise a first frequency band associated with cellular signals and a second frequency band associated with TV signals,

19. The method of claim **15**, wherein combining selected ones of the respective rectified signals comprises applying the respective rectified signals to a diode network.

20. The method of claim **15**, wherein combining selected ones of the respective rectified signals comprises applying the respective rectified signals to a network of active switches.

21. The method of claim **20**, further comprising selecting a state for each active switch in the network of active switches.

22. The method of claim **21**, wherein selecting the state for each active switch comprises, for each active switch in the network of active switches:

changing as state of the active switch from an initial state to a changed state;

monitoring the output voltage after the state of the active switch changed;

selecting the changed state if the output voltage increased after the state of the active switch changed;

selecting the initial state if the output voltage decreased after the state of the active switch changed.

23. The method of claim **15**, wherein combining selected ones of the respective rectified signals comprises applying the respective rectified signals to a diode network during a startup mode and applying the rectified signals to a network of active switches after the startup mode.

24. The method of claim **15**, further comprising applying the output voltage to a load circuit.

25. The method of claim **24**, wherein the load circuit comprises a backscattering transponder.

* * * * *

US 20070153561A1

(19) **United States**

(12) **Patent Application Publication**

Mickle et al.

(10) **Pub. No.: US 2007/0153561 A1**

(43) **Pub. Date: Jul. 5, 2007**

(54) **MULTIPLE ANTENNA ENERGY HARVESTING**

(75) Inventors: **Marlin H. Mickle**, Pittsburgh, PA (US); **Minhong Mi**, Sewickley, PA (US); **Leonid Mats**, Pittsburgh, PA (US)

Correspondence Address:
ECKERT SEAMANS CHERIN & MELLOTT
600 GRANT STREET
44TH FLOOR
PITTSBURGH, PA 15219

(73) Assignee: **University of Pittsburgh-Of the Commonwealth System of Higher Education**, Pittsburgh, PA

(21) Appl. No.: **11/619,786**

(22) Filed: **Jan. 4, 2007**

Related U.S. Application Data

(60) Provisional application No. 60/756,309, filed on Jan. 5, 2006.

Publication Classification

(51) **Int. Cl.**
H02M 5/06 (2006.01)
H04Q 5/22 (2006.01)

(52) **U.S. Cl.** **363/170**; 340/10.34

(57) **ABSTRACT**

Energy harvesting circuits and associated methods are provided that employ multiple antennas to optimize the amount of energy that is harvested while at the same time making efficient use of tag space. In some embodiments, matching networks are chosen in a manner that optimizes the DC energy that is created from the harvesting process. In other embodiments, phase shifts are introduced into the received signals to allow the signals to be more efficiently combined after they are rectified.

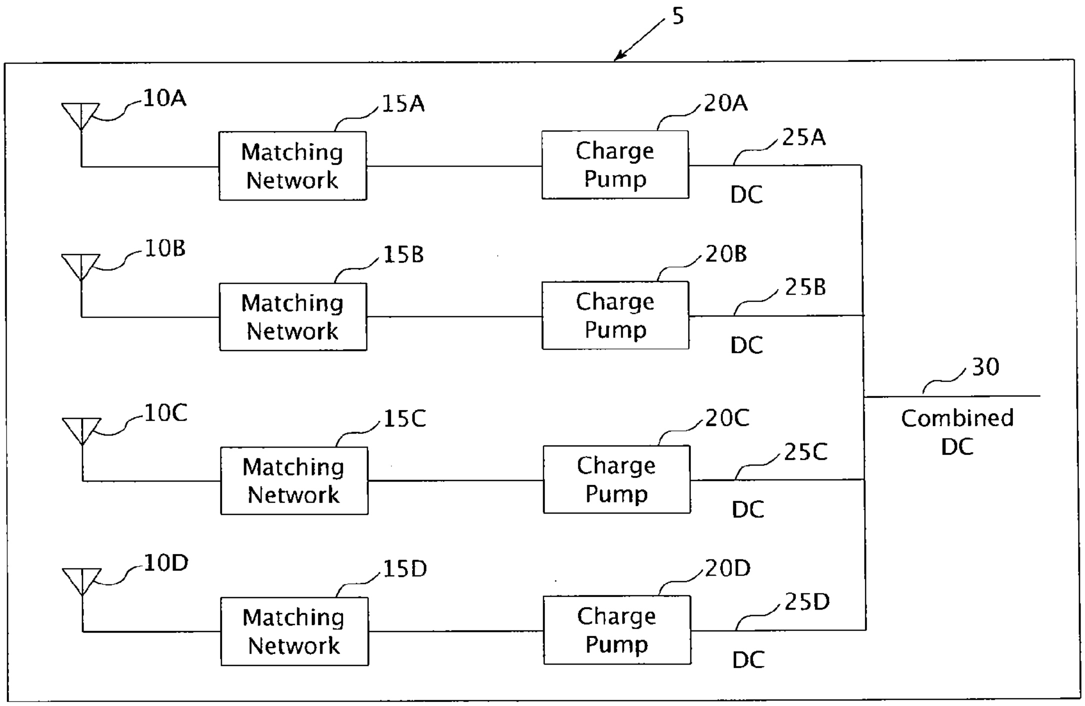
5
10A
15A
Matching Network
20A
Charge Pump
25A
DC
10B
15B
Matching Network
20B
Charge Pump
25B
DC
10C
15C
Matching Network
20C
Charge Pump
25C
DC
10D
15D
Matching Network
20D
Charge Pump
25D
DC
30
Combined DC

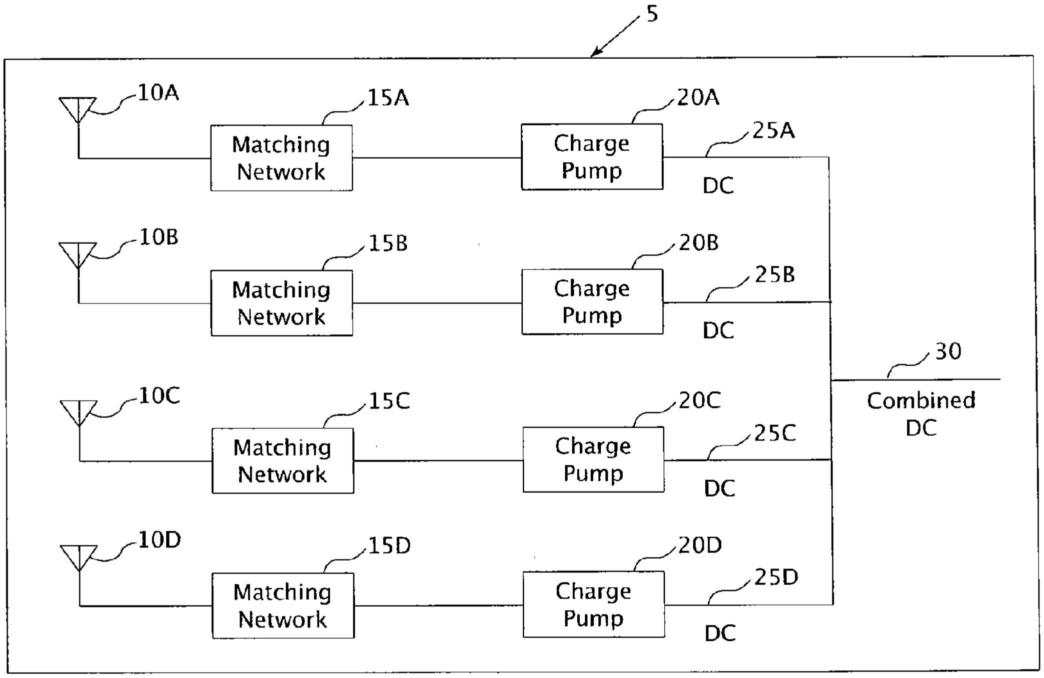
5
10A
15A
Matching Network
20A
Charge Pump
25A
DC
10B
15B
Matching Network
20B
Charge Pump
25B
DC
30
Combined DC
10C
15C
Matching Network
20C
Charge Pump
25C
DC
10D
15D
Matching Network
20D
Charge Pump
25D
DC

FIG. 1

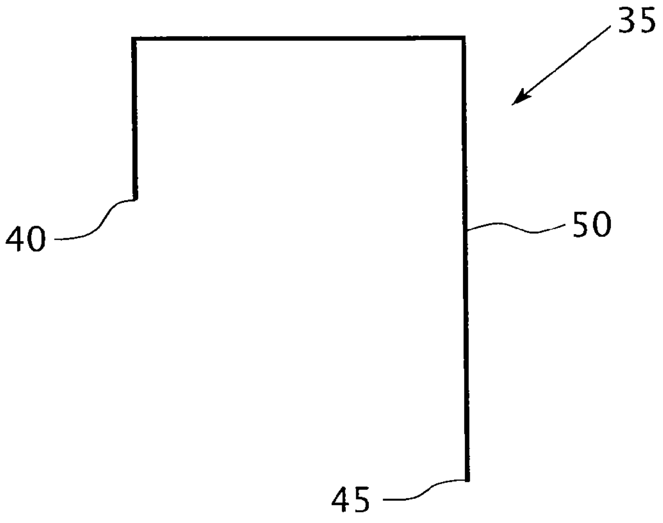

FIG. 2

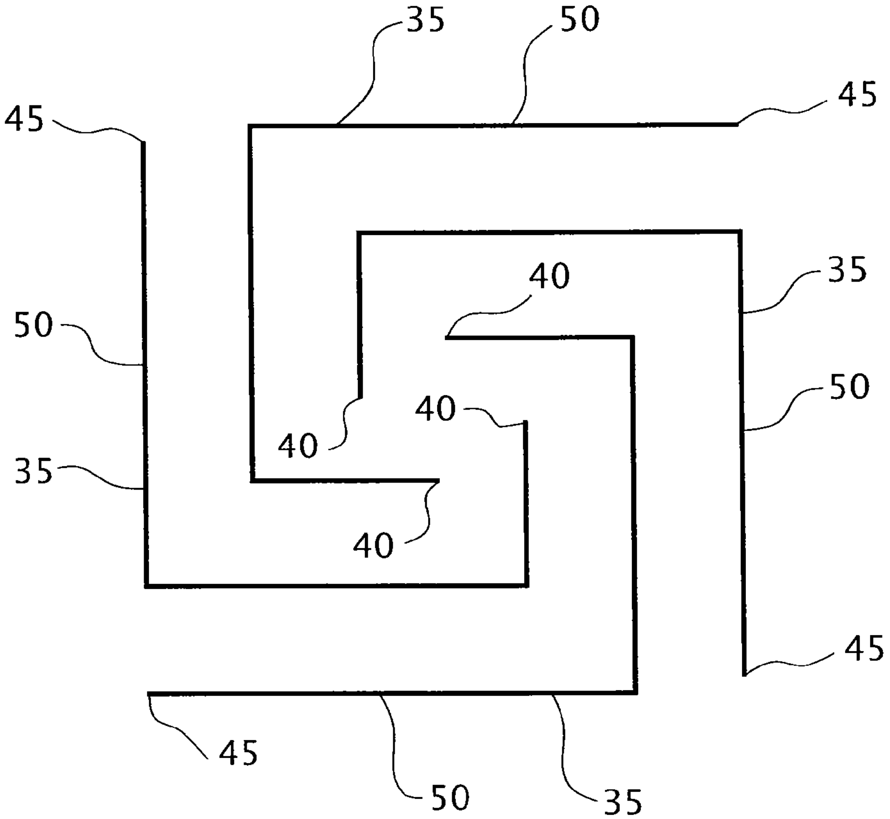
35
50
45
45
40
35
50
40
40
40
50
35
40
45
45
50
35

FIG. 3

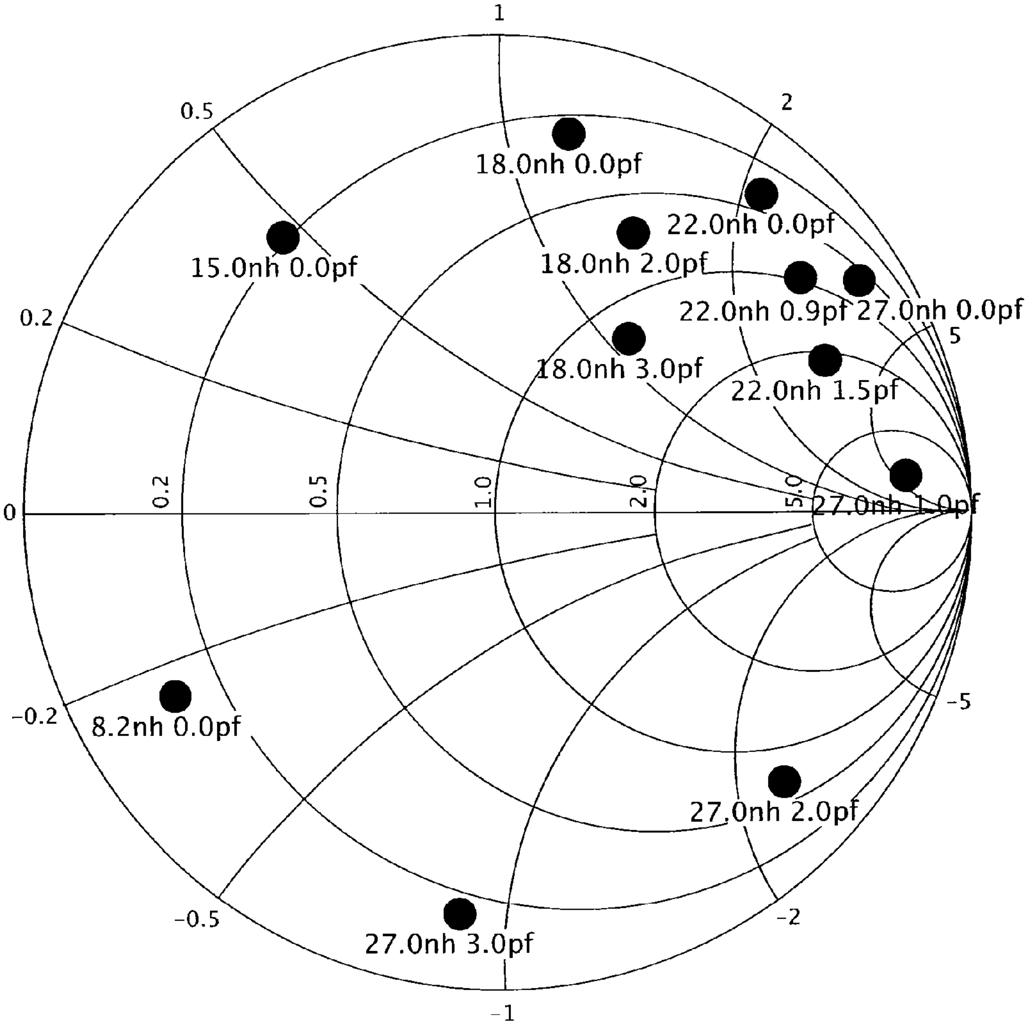
1
0.5
2
0.2
5
0
0.2
0.5
1.0
2.0
5.0
-0.2
-5
-0.5
-2
-1
18.0nh 0.0pf
22.0nh 0.0pf
15.0nh 0.0pf
18.0nh 2.0pf
22.0nh 0.9pf
27.0nh 0.0pf
18.0nh 3.0pf
22.0nh 1.5pf
27.0nh 1.0pf
8.2nh 0.0pf
27.0nh 2.0pf
27.0nh 3.0pf

FIG. 4

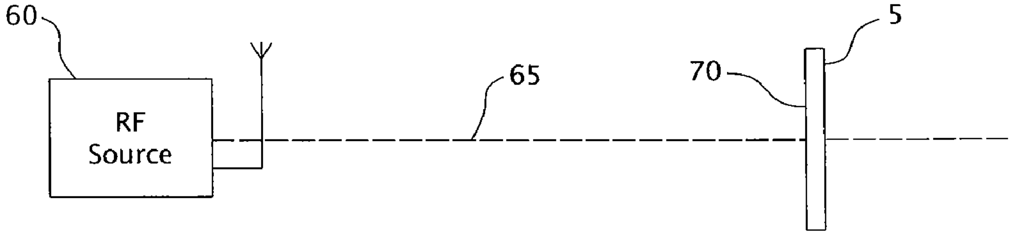
60
RF
Source
65
70
5

FIG. 5

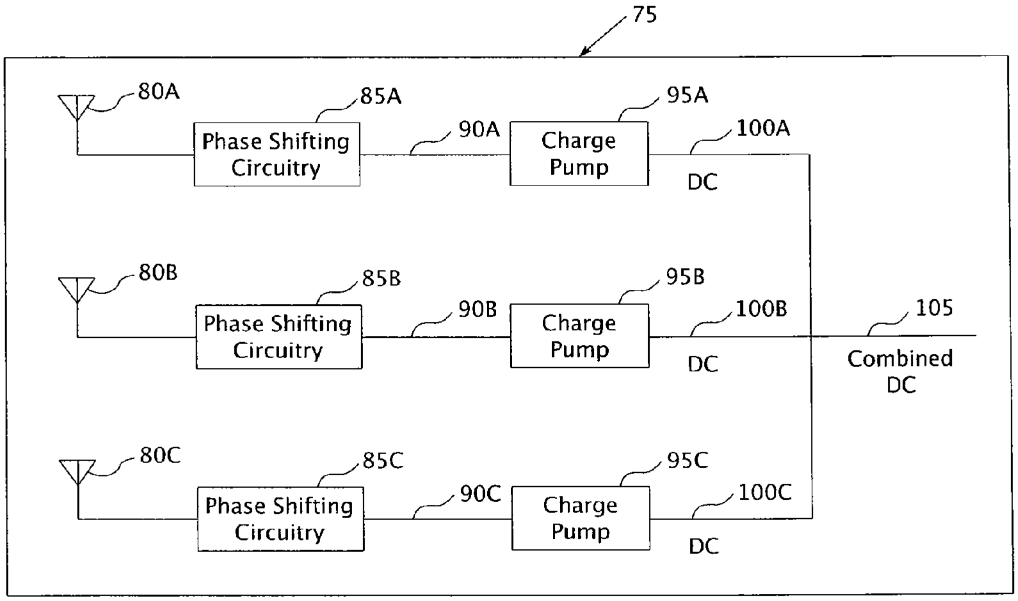
Phase Shifting Circuitry
Charge Pump
DC
Phase Shifting Circuitry
Charge Pump
DC
Combined DC
Phase Shifting Circuitry
Charge Pump
DC

FIG. 6

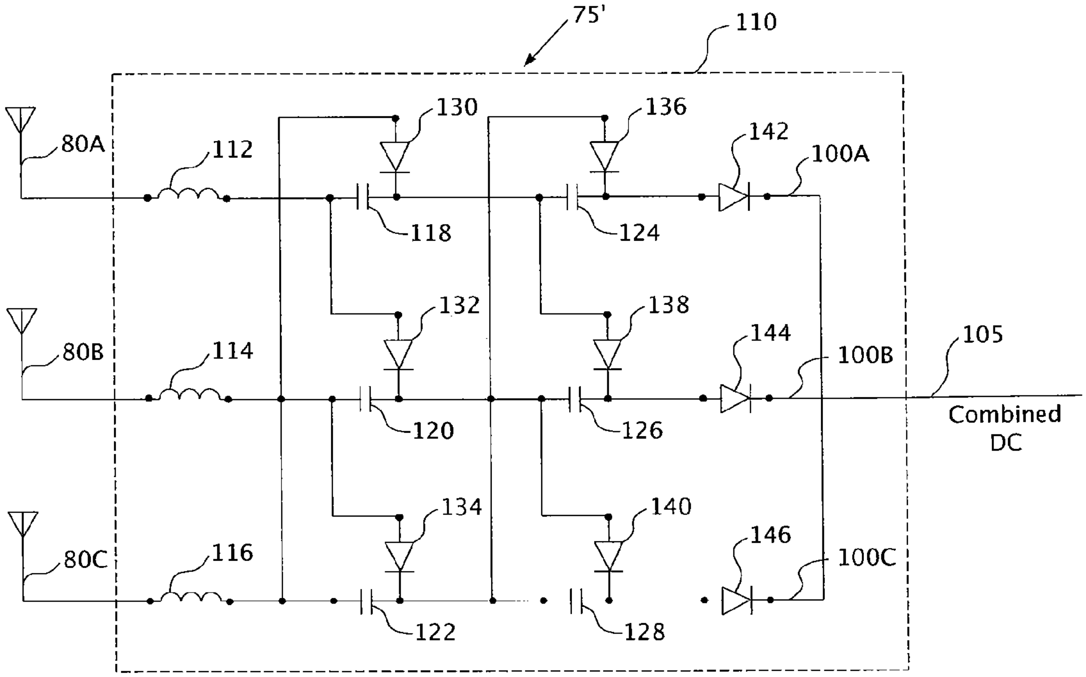
Combined DC

FIG. 7

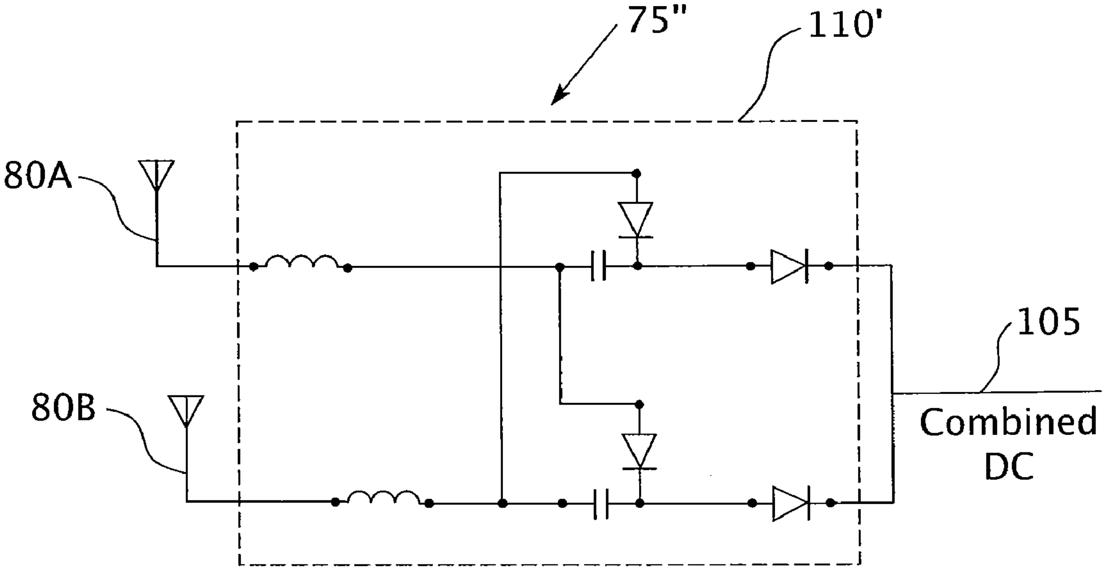
75"
110'
80A
80B
105
Combined DC

FIG. 8

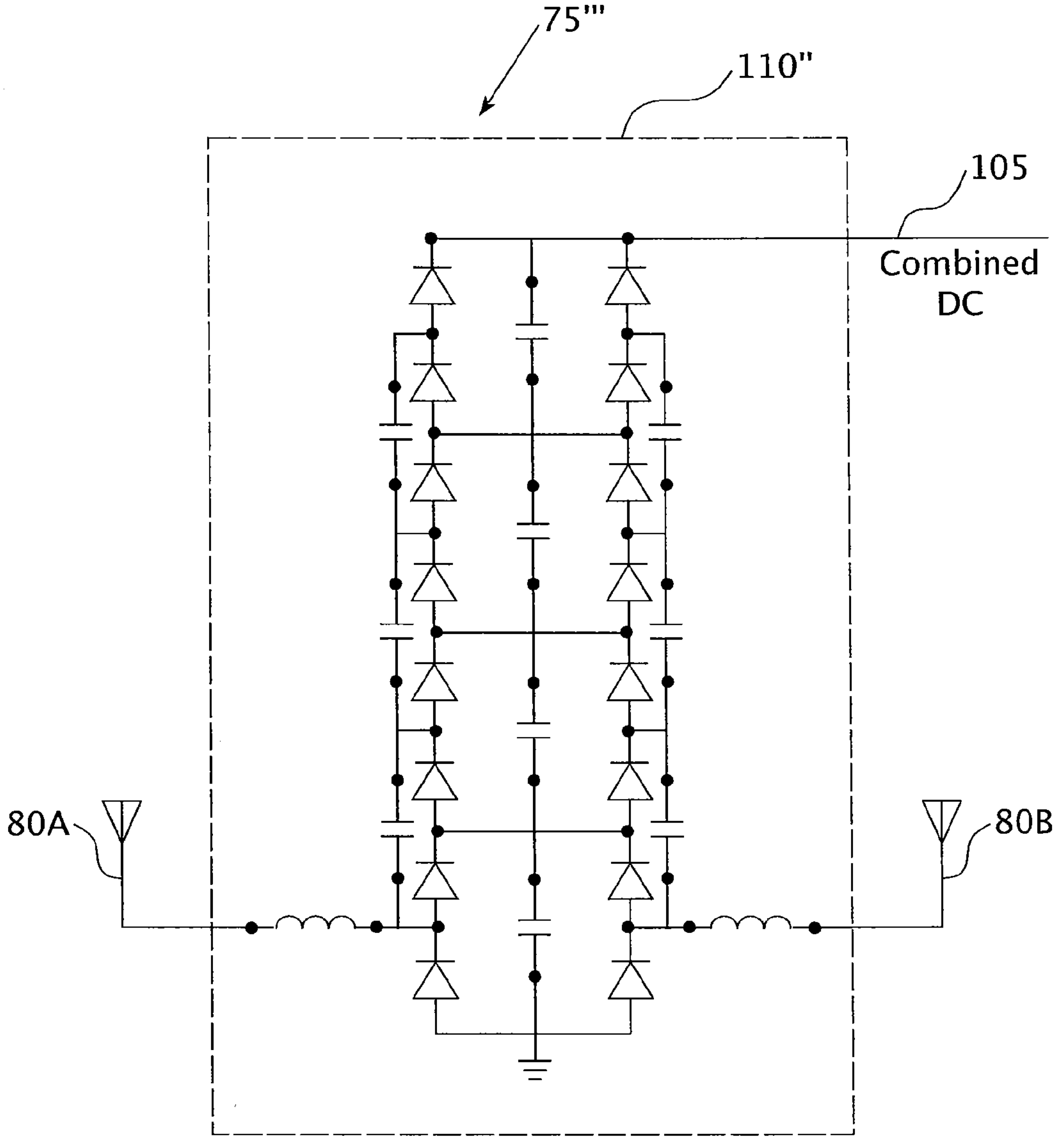
75'''
110"
105
Combined DC
80A
80B

FIG. 9

MULTIPLE ANTENNA ENERGY HARVESTING

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims the benefit of U.S. Provisional Application No. 60/756,309, entitled "Multiple Antenna Energy Harvesting," which was filed on Jan. 5, 2006, the disclosure of which is incorporated herein by reference.

FIELD OF THE INVENTION

[0002] The present invention relates to energy harvesting circuits for harvesting RF energy transmitted through the air and converting it to DC energy for powering electronic devices such as an RFID tag or transponder.

BACKGROUND OF THE INVENTION

[0003] The harvesting of RF energy transmitted through the air for use in powering electronic devices is extremely important in a number of fields, such as radio frequency identification (RFID) systems, security monitoring and remote sensing, among others. For example, RFID systems consist of a number of radio frequency tags or transponders (RFID tags) and one or more radio frequency readers or interrogators (RFID readers). The RFID tags typically include an integrated circuit (IC) chip, such as a complementary metal oxide semiconductor (CMOS) chip, and an antenna connected thereto for allowing the RFID tag to communicate with an RFID reader over an air interface by way of RF signals. In a typical RFID system, one or more RFID readers query the RFID tags for information stored on them, which can be, for example, identification numbers, user written data, or sensed data.

[0004] RFID tags can generally be categorized as either passive tags or active tags. Passive RFID tags do not have an internal power supply. Instead, the electrical current induced in the antenna of a passive RFID tag by the incoming RF signal from the RFID reader provides enough power for the IC chip or chips in the tag to power up and transmit a response. One passive tag technology, known as backscatter technology, generates signals by backscattering the carrier signal sent from the RFID reader. In another technology, described in U.S. Pat. Nos. 6,289,237 and 6,615,074, RF energy from the RFID reader is converted to a DC voltage by an antenna/matching circuit/charge pump combination. The DC voltage is then used to power a processor/transmitter/antenna combination that transmits information to the RFID reader at, for example, a different frequency. In either case, the area of the tag or silicon die is valuable, and therefore it is advantageous to make the most efficient use of the space thereon. In addition, it is known that multiple antennas can be used to generate a DC voltage from an RF signal. For example, U.S. Pat. No. 6,734,797 describes a tag that uses two dipole antennas where the greater of the energies produced from the two antennas is the one that is selected and used. This, however, is not the most efficient use of tag space since the energy for the "loser" antenna is not used. Thus, there is a need for an energy harvesting circuit that is able to employ multiple antennas to optimize the amount of energy that is harvested while at the same time making efficient use of tag space.

SUMMARY OF THE INVENTION

[0005] In one embodiment, an energy harvesting circuit, and associated method, is provided that includes a plurality of antennas, wherein each of the antennas is tuned to the same particular RF frequency range. Each of the antennas is structured to receive an RF signal having the same particular RF frequency and output a respective AC signal. The circuit also includes a plurality of matching networks, wherein each of the matching networks is operatively coupled to a respective one of the antennas and is structured to receive the AC signal output by the respective one of the antennas. The circuit further includes a plurality of voltage boosting and rectifying circuits, wherein each of the voltage boosting and rectifying circuits is operatively coupled to a respective one of the matching networks and is structured to receive the AC signal received by the respective one of the matching networks and output a DC voltage signal by converting the received AC signal into the DC voltage signal. In addition, the DC voltage signals output by the voltage boosting and rectifying circuits are summed together to create a combined DC voltage signal, and the impedance of each of the matching networks is chosen in a manner so as to maximize the voltage level of the DC voltage signal that is output by the associated voltage boosting and rectifying circuit.

[0006] In another embodiment, an energy harvesting circuit, and associated method, is provided that includes a plurality of antennas provided in an antenna plane, wherein each of the antennas is tuned the same particular RF frequency range. Each of the antennas is structured to receive an RF signal having the same particular RF frequency range and output a respective AC signal. The circuit also includes a plurality of matching networks, wherein each of the matching networks is operatively coupled to a respective one of the antennas and is structured to receive the AC signal output by the respective one of the antennas. The circuit further includes a plurality of voltage boosting and rectifying circuits, wherein each of the voltage boosting and rectifying circuits is operatively coupled to a respective one of the matching networks and is structured to receive the AC signal received by the respective one of the matching networks and output a DC voltage signal by converting the received AC signal into the DC voltage signal. In addition, the DC voltage signals output by the voltage boosting and rectifying circuits are summed together to create a combined DC voltage signal, and the impedance of each of the matching networks is chosen so as to cause a voltage level of the combined DC voltage signal to have a minimum deviation as a function of angle of rotation as the antenna plane is rotated about a first axis. Alternatively, the impedance of each of the matching networks may be chosen so as to cause a voltage level of the combined DC voltage signal to have at least a predetermined minimum value as the antenna plane is rotated about a first axis. As still a further alternative, the impedance of each of the matching networks may be chosen by: (i) incrementally rotating the antenna plane through a plurality of angle increments about a first axis, (ii) trying a plurality of different inductance and capacitance value combinations for each LC tank circuit of each of the matching networks at each of the angle increments, (iii) measuring the combined DC voltage signal for each of the inductance and capacitance value combinations at each of the angle increments, (iv) determining which one of the inductance and capacitance value combinations produces a maximum voltage level for the measured combined DC voltage signals at any one of the angle increments; (v) choosing for each the LC tank circuit the one of the inductance and capacitance value combinations that produces the maximum voltage level.

[0007] In yet another embodiment, an energy harvesting circuit, and associated method, is provided that includes a plurality of antennas, wherein each of the antennas is tuned to the same particular RF frequency range, and wherein each of the antennas is structured to receive an RF signal having the particular RF frequency range and output a respective AC signal. The circuit further includes phase shifting and rectifying circuitry operatively coupled to the antennas. The phase shifting and rectifying circuitry is structured to: (i) receive each respective AC signal, (ii) create a plurality of out of phase AC signals by causing each respective AC signal to be out of phase with one another, and (iii) convert each of the out of phase AC signals into a respective DC voltage signal. The DC voltage signals are then summed together to create a combined DC voltage signal.

[0008] Therefore, it should now be apparent that the invention substantially achieves all the above aspects and advantages. Additional aspects and advantages of the invention will be set forth in the description that follows, and in part will be obvious from the description, or may be learned by practice of the invention. Moreover, the aspects and advantages of the invention may be realized and obtained by means of the instrumentalities and combinations particularly pointed out in the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] The accompanying drawings illustrate presently preferred embodiments of the invention, and together with the general description given above and the detailed description given below, serve to explain the principles of the invention. As shown throughout the drawings, like reference numerals designate like or corresponding parts.

[0010] FIG. **1** is a block diagram of one embodiment of an energy harvesting circuit according to the present invention;

[0011] FIG. **2** is a schematic illustration of an embodiment of a square spiral antenna that may be used in the energy harvesting circuits described herein;

[0012] FIG. **3** is a schematic illustration of an embodiment of an antenna layout that may be used in the energy harvesting circuits described herein;

[0013] FIG. **4** is a Smith chart that may be employed to choose the matching networks as described herein;

[0014] FIG. **5** is a schematic illustration of an energy harvesting circuit as described herein positioned within the range of a suitable RF source;

[0015] FIG. **6** is a block diagram of an energy harvesting circuit according to alternative embodiment of the invention in which a phase shift is introduced into each of the received RF signals; and

[0016] FIGS. **7**, **8** and **9** are schematic illustrations of three particular embodiments of the energy harvesting circuit shown in FIG. **6**.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0017] FIG. **1** is a block diagram of one embodiment of an energy harvesting circuit **5** according to the present invention. The energy harvesting circuit **5** may, for example, form part of an RF transponder, such as an RFID tag, for providing power thereto or form part of some other type of remote station for energizing a power storage device thereof or some other object of interest. The energy harvesting circuit **5** includes a plurality of antennas **10**A-**10**D, each of which is tuned to the same particular RF frequency range. As used herein, the term "RF frequency range" shall refer to a single RF frequency or a band of RF frequencies. As is known, each antenna **10**A-**10**D may have a tank circuit (not shown) that includes an inductor and a capacitor, wherein the inductance and capacitance values are chosen to tune the antenna **10**A-**10**D to the desired RF frequency range, i.e., the same particular RF frequency or band of RF frequencies. In this embodiment, each tank circuit is fixed (i.e., fixed inductor and capacitor) so that the antenna **10**A, **10**B, **10**C, **10**D is fixedly tuned to the RF frequency range of interest.

[0018] While four antennas **10**A-**10**D are shown in FIG. **1**, it should be understood that that is meant to be exemplary only, and that the plurality of antennas may include less than or more than four antennas (with each such antenna being operatively coupled to respective accompanying circuitry as described below). Furthermore, each antenna **10**A, **10**B, **10**C, **10**D may be, for example, a square spiral antenna **35** having the form shown in FIG. **2**, with the innermost end **40** being open and the outermost end **45** being connected to a matching network (matching network **15**A, **15**B, **15**C, **15**D shown in FIG. **1**) as described below. In this manner, the square spiral antenna **35** may be viewed as an asymmetric dipole, and thus the square spiral antenna **35** does not need a ground plane. Preferably, the length of the outermost segment **50** of the square spiral antenna **35** is about equal to a quarter of the wavelength of the particular RF frequency or center frequency of the band of RF frequencies to which it is tuned as described herein, and the total length of the square spiral antenna **35** (three segments) is about equal to one half of that wavelength.

[0019] FIG. **3** is a schematic illustration of one particular antenna layout that may be used in the energy harvesting circuit **5** shown in FIG. **1** that includes four antennas **35** positioned and spaced in the manner shown with the outermost end **45** of each antenna **35** being connected to the respective matching network **15**A, **15**B, **15**C, **15**D as described below. It has been observed that the conductor width and spacing of the antennas **35** in FIG. **3** can be varied without seriously degrading the performance measured as the DC output of the respective charge pumps **20**A, **20**B, **20**C, **20**D described below. Therefore, in the layout shown in FIG. **3**, it is possible to put four antennas **35** in an area that is slightly more than two times the area occupied by single antenna **35**.

[0020] Referring again to FIG. **1**, each antenna **10**A, **10**B, **10**C, **10**D is operatively coupled to a respective matching network **15**A, **15**B, **15**C, **15**D and charge pump **20**A, **20**B, **20**C, **20**D as shown in FIG. **1**. In particular, each antenna **10**A, **10**B, **10**C, **10**D is electrically connected to a respective matching network **15**A, **15**B, **15**C, **15**D, which in turn is electrically connected to a respective voltage boosting and rectifying circuit preferably in the form of a one or more stage charge pump **20**A, **20**B, **20**C, **20**D. Charge pumps are well known in the art. Basically, one stage of a charge pump essentially doubles the effective amplitude of an AC input voltage and stores the resulting increased DC voltage on an output capacitor. The voltage could also be stored using a rechargeable battery. Successive stages of a charge pump, if present, will essentially increase the voltage from the previous stage resulting in an increased output voltage. In operation, each antenna **10**A, **10**B, **10**C, **10**D receives energy, such as RF energy, that is transmitted in space by a far-field source, such as an RF source. The RF source may be, for example, an RF interrogator unit which transmits an RF interrogation signal, or a local radio station, in which case the RF energy comprises ambient RF energy in the vicinity of the energy harvesting circuit **5**. The RF energy received by each antenna **10**A, **10**B, **10**C, **10**D is provided, in the form of an AC signal, to the associated charge pump **20**A, **20**B, **20**C, **20**D through the associated matching network **15**A, **15**B, **15**C, **15**D. Each charge pump **20**A, **20**B, **20**C, **20**D rectifies the received AC signal to produce a respective DC signal **25**A, **25**B, **25**C, **25**D, wherein each DC signal **25**A, **25**B, **25**C, **25**D is amplified as compared to what it would have been had a simple rectifier been used. The individual DC signals **25**A, **25**B, **25**C, **25**D are then summed to produce a combined DC signal **30**, which may be used to power a circuit or charge a power storage device as described elsewhere herein. The individual DC signals **25**A, **25**B, **25**C, **25**D may be summed in any suitable manner, such as, without limitation, connecting them in parallel or connecting them in series to produce the combined DC signal **30**.

[**0021**] In one particular embodiment, each matching network **15**A, **15**B, **15**C, **15**D is chosen (i.e., its impedance is chosen) so as to maximize the output of the respective charge pump **20**A, **20**B, **20**C, **20**D in the form of the respective DC signal **25**A, **25**B, **25**C, **25**D (i.e., to maximize the DC signal **25**A, **25**B, **25**C, **25**D). In other words, each matching network **15**A, **15**B, **15**C, **15**D is matching the impedance of the respective antenna **10**A, **10**B, **10**C, **10**D to the respective charge pump **20**A, **20**B, **20**C, **20**D solely on the basis of maximizing the DC output of the respective charge pump **20**A, **20**B, **20**C, **20**D. In the preferred embodiment, each matching network **15**A, **15**B, **15**C, **15**D is an LC circuit of either an L topology (which includes one inductor and one capacitor) or a π topology (which includes one inductor and two capacitors) wherein the inductance of the LC circuit and the capacitance of the LC circuit are chosen so as to maximize the DC output of the respective charge pump **20**A, **20**B, **20**C, **20**D. In one embodiment, an LC tank circuit may be formed by the inherent distributed inductance and inherent distributed capacitance of the conducing elements of each antenna **10**A, **10**B, **10**C, **10**D, in which case the antenna is designed and laid out in a manner that results in the appropriate chosen L and C values.

[**0022**] In the preferred embodiment, each matching network **15**A, **15**B, **15**C, **15**D is chosen so as to maximize the output of the respective charge pump **20**A, **20**B, **20**C, **20**D using a trial and error ("annealing") empirical approach. In particular, various sets of inductor and capacitor values are used as matching elements in each matching network **15**A, **15**B, **15**C, **15**D, and the resulting output of the respective charge pump **20**A, **20**B, **20**C, **20**D is measured for each combination, and the combination that produces the maximum output is chosen. In this process, the input impedance of the charge pump (**20**A, **20**B, **20**C, **20**D) with each matching network (**15**A, **15**B, **15**C, **15**D) may be plotted as a point on a Smith chart with a color coding for the amount of energy harvested as shown in FIG. **4**. After a number of tries, it is easy to see a clustering of the color coded points to selectively choose other points in or around the cluster to achieve a near optimum value. FIG. **4** illustrates a number of points showing the results on a Smith chart. It is important to note that the matching achieved in this manner tends to optimize the asymmetric dipole configuration described elsewhere herein. The matching component values are indicated in FIG. **4** as circles on the Smith chart with grayscale indications of voltage at the charge pump (**20**A, **20**B, **20**C, **20**D) output as compared to the 0 to 5 volt grayscale on the right of FIG. **4**.

[**0023**] In this embodiment, the matching network (**15**A, **15**B, **15**C, **15**D) is chosen individually for each antenna (**10**A, **10**B, **10**C, **10**D) in the energy harvesting circuit **5** in the manner just described. Other embodiments are also possible and are described below.

[**0024**] As will be appreciated by those of skill in the art, each antenna **10**A, **10**B, **10**C, **10**D that is included in the energy harvesting circuit **5**, for example as laid out in FIG. **3**, will have a different polarization with respect to the transmitting antenna of the RF source that is determined by the relative angle of the antenna **10**A, **10**B, **10**C, **10**D in question to the polarization of the transmitting antenna of the RF source. Another embodiment of the invention, described below, chooses each matching network (**15**A, **15**B, **15**C, **15**D) in a manner whereby the relative polarization efficiency of each antenna **10**A, **10**B, **10**C, **10**D will be more efficiently utilized. In particular, in this embodiment, the energy harvesting circuit **5** is set up as shown in FIG. **1** with the antennas **10**A, **10**B, **10**C, **10**D being laid out as desired, for example in the manner shown in FIG. **3**. Then, as shown in FIG. **5**, the energy harvesting circuit **5** is positioned within the range of a suitable RF source **60**, preferably of a type identical or similar to the one with which the energy harvesting circuit **5** will ultimately be used, with the plane **70** in which the antennas are laid out **10**A, **10**B, **10**C, **10**D being substantially perpendicular to a line **65** that is taken through the center of the transmitting antenna of the RF source **60**. While the RF source **60** is transmitting, the antenna is incrementally rotated about the line **65**, preferably through 360 degrees (the rotation will be into and out of the paper in FIG. **5**). At each increment (each angle of rotation θ), the matching elements (the inductance and capacitance values) in each matching network **15**A, **15**B, **15**C, **15**D are varied over a set of different values (preferably predetermined), and with each such variation, the voltage level of the combined DC signal **30** is measured and recorded. The same set of inductance and capacitance values is used at each angle θ. In addition, at each angle θ, the set of values may and preferably will include instances where the values are different for one or more of the matching networks, i.e., each matching network will not have the same values. Once the rotation is complete, the measured and recorded levels of the combined DC signal **30** (as a function of the angle θ) are analyzed and the set of inductance and capacitance values yielding the minimum deviation in the combined DC signal **30** as a function of the angle θ is chosen. The matching networks **15**A, **15**B, **15**C, **15**D are then structured accordingly (i.e., to have the chosen inductance and capacitance values) in the final energy harvesting circuit **5**. As will be appreciated, this will result in the combined DC signal **30** having the minimum deviation in all orientations of the energy harvesting circuit **5** in the plane in which the antennas are laid out. As a result, at least some minimum amount of DC voltage will be able to be harvested in all orientations of the energy harvesting circuit **5**. This may be important if, for example, the circuit or object (or some component thereof) to be powered requires some minimum voltage level to operate. The matching networks **15**A, **15**B, **15**C, **15**D can be chosen as described above to ensure that at least that minimum voltage level is produced in all orientations (i.e., all angles of rotation θ) of the energy harvesting circuit **5**. In addition, once the matching networks **15**A, **15**B, **15**C, **15**D (i.e., the inductance and capacitance values therefor) are chosen in this manner (which may be time consuming) for one energy harvesting circuit **5**, multiple energy harvesting circuits **5** that meet the same parameters can be mass produced without having to go through the same trial and error steps.

[0025] In another embodiment, once the rotation is complete, the measured and recorded levels of the combined DC signal **30** are analyzed and the set of inductance and capacitance values yielding a maximum level for the combined DC signal **30** at any one orientation is chosen. The matching networks are then structured accordingly (i.e., to have the chosen inductance and capacitance values) in the final energy harvesting circuit **5**.

[0026] In the energy harvesting circuit **5** shown in FIG. **1**, it will be appreciated that each of the AC signals provided to the matching networks **15**A, **15**B, **15**C, **15**D will be in phase with one another. As a result, the individual antenna, matching, charge pump, etc. circuits have an RF supply in which the individual RF sinusoidal signals are in phase. This is analogous to the single phase rectifier design which is well known in DC rectifier circuitry. The removal of the minimum values of DC output is strictly on the basis of the angular variation of the antenna circuitry through matching the angles of orientation, and not the angles of the individual sinusoidal sources. Thus, in the energy harvesting circuit **5** shown in FIG. **1** the signals are not combined in the most efficient manner, and a more efficient manner is described below in connection with FIG. **6**.

[0027] FIG. **6** is a block diagram of an energy harvesting circuit **75** according to alternative embodiment of the invention in which a phase shift is introduced into each of the received RF signals. In particular, the energy harvesting circuit **75** includes a plurality of antennas **80**A-**80**C (similar to the antennas **10**A-**10**D), each of which is fixedly tuned to the same particular RF range. While three antennas **80**A-**80**C are shown in FIG. **6**, it should be understood that that is meant to be exemplary only, and that the plurality of antennas may include less than or more than three antennas (with each such antenna being operatively coupled to respective accompanying circuitry as described below). Furthermore, each antenna **80**A, **80**B, **80**C may be, for example, a square spiral antenna **35** having the form shown in FIG. **2**. In the case of four antennas, the layout may be as shown in FIG. **3**.

[0028] As shown in FIG. **6**, each antenna **80**A, **80**B, **80**C is electrically connected to respective phase shifting circuitry **85**A, **85**B, **85**C, which in turn is electrically connected to a respective voltage boosting and rectifying circuit preferably in the form of a one or more stage charge pump **95**A, **95**B, **95**C. In operation, each antenna **80**A, **80**B, **80**C receives energy, such as RF energy, that is transmitted in space by a far-field source, such as an RF source. The RF source may be, for example, an RF interrogator unit which transmits an RF interrogation signal, or a local radio station, in which case the RF energy comprises ambient RF energy in the vicinity of the energy harvesting circuit **75**. The RF energy received by each antenna **80**A, **80**B, **80**C is provided, in the form of an AC signal, to the associated phase shifting circuitry **85**A, **85**B, **85**C. Each phase shifting circuitry **85**A, **85**B, **85**C shifts the phase of the received AC signal in a manner such that the AC signals **90**A, **90**B, **90**C output by the phase shifting circuitry **85**A, **85**B, **85**C are all out of phase with one another. Thus, each phase shifting circuitry **85**A, **85**B, **85**C introduces a different degree of phase shift (which may actually be zero, in which case that particular phase shifting circuitry may be omitted, or, alternatively, be arranged to simply not shift phase or shift phase by 360 degrees). In one embodiment, the phase shift introduced by each phase shifting circuitry **85**A, **85**B, **85**C results in each AC signal **90**A, **90**B, **90**C being out of phase from the immediately adjacent signal AC signal **90**A, **90**B, **90**C by an equal amount. For example, for three AC signals **90**A, **90**B, **90**C, there would be 120 degrees separating immediately adjacent signals (e.g., one with 0 degrees phase shift, one with 120 degrees phase shift and one with 240 degrees phase shift, or, alternatively, one with 20 degrees phase shift, one with 140 degrees phase shift and one with 260 degrees phase shift).

[0029] The AC signal AC signal **90**A, **90**B, **90**C output by each phase shifting circuitry **85**A, **85**B, **85**C is provided to the associated charge pump **95**A, **95**B, **95**C. In an alternative embodiment, the AC signal AC signal **90**A, **90**B, **90**C output by each phase shifting circuitry **85**A, **85**B, **85**C may be provided to the associated charge pump **95**A, **95**B, **95**C through an associated matching network chosen and configured in the manners described elsewhere herein in order to optimize DC output. In either embodiment, each charge pump **95**A, **95**B, **95**C amplifies and rectifies the received AC signal to produce a respective DC signal **100**A, **100**B, **100**C. The individual DC signals **100**A, **100**B, **100**C are then summed (as described elsewhere herein) to produce a combined DC signal **105**, which may be used to power a circuit or charge a power storage device as described elsewhere herein. Because the AC signal **90**A, **90**B, **90**C received by the charge pumps **95**A, **95**B, **95**C are all out of phase with one another, they are able to be converted to DC and summed to produce the combined DC signal **105** in a more efficient manner, thereby leading to an increase on the voltage of the combined DC signal **105** over what would have been the case had the phase shifts described herein not been introduced. This is the case because the phase shifts cause the peaks and valleys of the AC signals **90**A, **90**B, **90**C be offset form one another rather then lined up with one another. It will also be appreciated that the multiphase embodiment provides a second degree of freedom in design for rotational purposes.

[0030] FIG. **7** is schematic illustration of one particular embodiment of the energy harvesting circuit **75**, labeled as **75'**. The energy harvesting circuit **75'** includes antennas **80**A, **80**B, **80**C as described above, each of which is fixedly tuned to the same particular RF frequency range and outputs an AC signal. Each antenna **80**A, **80**B, **80**C is connected to phase shifting and rectifying circuitry **110** that performs the functionality of the phase shifting circuitry **85**A, **85**B, **85**C and the charge pumps **95**A, **95**B, **95**C to produce the respective DC signals **100**A, **100**B, **100**C As described elsewhere herein, the DC signals **100**A, **100**B, **100**C are then summed (as described elsewhere herein) to produce a combined DC signal **105**. The phase shifting and rectifying circuitry **110** includes a well known Cockroft-Walton multiplier and consists of inductors **112**, **114**, and **116**, capacitors **118**, **120**, **122**, **124**, **126**, and **128**, and diodes **130**, **132**, **134**, **1365**, **138**, **140**, **142**, **144** and **146** connected as shown in FIG. **7**. The phase shifting and rectifying circuitry **110** receives each AC signal from the antennas **80**A, **80**B, **80**C, shifts the phases thereof such that each AC signal is out of phase with one another, and boosts and rectifies the resulting signals to produce the DC signals **100**A, **100**B, **100**C. The values of the inductors **112**, **114**, and **116** and the capacitors **118**, **120**, and **122** determines the degree of phase shift applies to each respective AC signal received from the antennas **80**A, **80**B, **80**C. FIG. **8** shows a similar embodiment of an energy harvesting circuit **75**" where two antennas are used with similar phase shifting and rectifying circuitry **110**'. It will be appreciated that four or more antennas may be used in a similar manner. FIG. **9** shows yet another similar embodiment of an energy harvesting circuit **75**"' where two antennas are used with similar phase shifting and rectifying circuitry **110**".

[0031] While preferred embodiments of the invention have been described and illustrated above, it should be understood that these are exemplary of the invention and are not to be considered as limiting. Additions, deletions, substitutions, and other modifications can be made without departing from the spirit or scope of the present invention. Accordingly, the invention is not to be considered as limited by the foregoing description but is only limited by the scope of the appended claims.

What is claimed is:

1. An energy harvesting circuit, comprising:

a plurality of antennas, each of said antennas being tuned to the same particular RF frequency range, each of said antennas being structured to receive an RF signal having said particular RF frequency range and output a respective AC signal;

a plurality of matching networks, each of said matching networks being operatively coupled to a respective one of said antennas and being structured to receive the AC signal output by said respective one of said antennas; and

a plurality of voltage boosting and rectifying circuits, each of said voltage boosting and rectifying circuits being operatively coupled to a respective one of said matching networks and being structured to receive the AC signal received by the respective one of said matching networks and output a DC voltage signal by converting the received AC signal into the DC voltage signal;

wherein the DC voltage signals output by the voltage boosting and rectifying circuits are summed together to create a combined DC voltage signal, and wherein an impedance of each of the matching networks is chosen in manner so as to maximize a voltage level of the DC voltage signal that is output by the associated one of the voltage boosting and rectifying circuits.

2. The energy harvesting circuit according to claim 1, wherein each of said voltage boosting and rectifying circuits is a charge pump.

3. The energy harvesting circuit according to claim 1, wherein each of said matching networks is an LC tank circuit.

4. The energy harvesting circuit according to claim 3, wherein an impedance of each of the matching networks is chosen in manner so as to maximize a voltage level of the DC voltage signal that is output by the associated one of the voltage boosting and rectifying circuits by, for each LC tank circuit of each of the matching networks: (i) trying a plurality of different inductance and capacitance value combinations for the LC tank circuit, (ii) measuring the voltage level of the DC voltage signal that is output by the associated one of the voltage boosting and rectifying circuits for each of the inductance and capacitance value combinations, and (iii) choosing one of the inductance and capacitance value combinations that produces a maximum voltage level of the DC voltage signal that is output by the associated one of the voltage boosting and rectifying circuits.

5. The energy harvesting circuit according to claim 1, wherein each of said antennas is a square spiral antenna having an outermost segment, a middle segment and an innermost segment, wherein a length of each outermost segment of is equal to about a quarter of the wavelength of the particular RF frequency or center frequency of the band of RF frequencies included in said particular RF frequency range, wherein a total length of each square spiral antenna is equal to about one half of said wavelength, and wherein the outermost segment of each of said antennas is operatively coupled to the associated one of said matching networks.

6. The energy harvesting circuit according to claim 5, wherein each square spiral antenna alone occupies a first spatial area, and wherein said antennas are provided in an antenna layout that occupies a second spatial area that is less than two times said first spatial area.

7. An energy harvesting method, comprising:

providing a plurality of antennas, each of said antennas being tuned to the same particular RF frequency range;

receiving a plurality of RF signals using said antennas, each of said RF signals having said particular RF frequency range;

outputting from said antennas a plurality of AC signals in response to receiving said RF signals;

converting each of the AC signals into a respective DC voltage signal after first passing each of the AC signals through a respective matching network having an impedance chosen in manner so as to maximize a voltage level of the respective DC voltage signal; and

summing each of the respective DC voltage signals together to create a combined DC voltage signal.

8. An energy harvesting circuit, comprising:

a plurality of antennas provided in an antenna plane, each of said antennas being tuned to the same particular RF frequency range, each of said antennas being structured to receive an RF signal having said particular RF frequency range and output a respective AC signal;

a plurality of matching networks, each of said matching networks being operatively coupled to a respective one of said antennas and being structured to receive the AC signal output by said respective one of said antennas; and

a plurality of voltage boosting and rectifying circuits, each of said voltage boosting and rectifying circuits being operatively coupled to a respective one of said matching networks and being structured to receive the AC signal received by the respective one of said matching networks and output a DC voltage signal by converting the received AC signal into the DC voltage signal;

wherein the DC voltage signals output by the voltage boosting and rectifying circuits are summed together to create a combined DC voltage signal, and wherein an impedance of each of the matching networks is chosen so as to cause a voltage level of the combined DC voltage signal to have a minimum deviation as a function of angle of rotation as said antenna plane is rotated about a first axis.

9. The energy harvesting circuit according to claim 8, wherein said first axis is a line taken through a center of a transmitting antenna of a source of said RF signal having said particular RF frequency range.

10. The energy harvesting circuit according to claim 8, wherein each of said voltage boosting and rectifying circuits is a charge pump.

11. The energy harvesting circuit according to claim 8, wherein each of said matching networks is an LC tank circuit.

12. The energy harvesting circuit according to claim 11, wherein an impedance of each of the matching networks is chosen so as to cause a voltage level of the combined DC voltage signal to have a minimum deviation as a function of angle of rotation as said antenna plane is rotated about a first axis by: (i) incrementally rotating said antenna plane through a plurality of angle increments, (ii) trying a plurality of different inductance and capacitance value combinations for each LC tank circuit of each of the matching networks at each of said angle increments, (iii) measuring the combined DC voltage signal for each of the inductance and capacitance value combinations at each of said angle increments, (iv) determining which one of the inductance and capacitance value combinations produces a minimum deviation in said measured combined DC voltage signals as a function of each of said angle increments; (v) choosing for each said LC tank circuit the one of the inductance and capacitance value combinations that produces the minimum deviation.

13. The energy harvesting circuit according to claim 8, wherein each of said antennas is a square spiral antenna having an outermost segment, a middle segment and an innermost segment, wherein a length of each outermost segment of is equal to about a quarter of the wavelength of the particular RF frequency or center frequency of the band of RF frequencies included in said particular RF frequency range, wherein a total length of each square spiral antenna is equal to about one half of said wavelength, and wherein the outermost segment of each of said antennas is operatively coupled to the associated one of said matching networks.

14. The energy harvesting circuit according to claim 13, wherein each square spiral antenna alone occupies a first spatial area, and wherein said antennas are provided in an antenna layout that occupies a second spatial area that is less than two times said first spatial area.

15. An energy harvesting method, comprising:

providing a plurality of antennas in an antenna plane, each of said antennas being tuned to the same particular RF frequency range;

receiving a plurality of RF signals using said antennas, each of said RF signals having said particular RF frequency range;

outputting from said antennas a plurality of AC signals in response to receiving said RF signals;

converting each of the AC signals into a respective DC voltage signal; and

summing each of the respective DC voltage signals together to create a combined DC voltage signal;

wherein said converting step is performed after first passing each of the AC signals through a respective matching network having an impedance chosen so as to cause a voltage level of the combined DC voltage signal to have a minimum deviation as a function of angle of rotation as said antenna plane is rotated about a first axis.

16. The energy harvesting method according to claim 15, wherein said first axis is a line taken through a center of a transmitting antenna of a source of each said RF signal having said particular RF frequency range.

17. An energy harvesting circuit, comprising:

a plurality of antennas provided in an antenna plane, each of said antennas being tuned to the same particular RF frequency range, each of said antennas being structured to receive an RF signal having said particular RF frequency range and output a respective AC signal;

a plurality of matching networks, each of said matching networks being operatively coupled to a respective one of said antennas and being structured to receive the AC signal output by said respective one of said antennas; and

a plurality of voltage boosting and rectifying circuits, each of said voltage boosting and rectifying circuits being operatively coupled to a respective one of said matching networks and being structured to receive the AC signal received by the respective one of said matching networks and output a DC voltage signal by converting the received AC signal into the DC voltage signal;

wherein the DC voltage signals output by the voltage boosting and rectifying circuits are summed together to create a combined DC voltage signal, and wherein an impedance of each of the matching networks is chosen so as to cause a voltage level of the combined DC voltage signal to have at least a predetermined minimum value as said antenna plane is rotated about a first axis.

18. The energy harvesting circuit according to claim 17, wherein said first axis is a line taken through a center of a transmitting antenna of a source of said RF signal having said particular RF frequency range.

19. The energy harvesting circuit according to claim 17, wherein each of said voltage boosting and rectifying circuits is a charge pump.

20. The energy harvesting circuit according to claim 17, wherein each of said matching networks is an LC tank circuit.

21. The energy harvesting circuit according to claim 20, wherein an impedance of each of the matching networks is chosen so as to cause a voltage level of the combined DC voltage signal to have at least a predetermined minimum value as said antenna plane is rotated about a first axis by: (i) incrementally rotating said antenna plane through a plurality of angle increments, (ii) trying a plurality of different inductance and capacitance value combinations for each LC tank circuit of each of the matching networks at each of said angle increments, (iii) measuring the combined DC voltage signal for each of the inductance and capacitance value combinations at each of said angle increments, (iv) determining which one of the inductance and capacitance value combinations produces at least said predetermined minimum value at each of said angle increments; (v) choosing for each said LC tank circuit the one of the inductance and capacitance value combinations that produces the at least said predetermined minimum value.

22. The energy harvesting circuit according to claim 17, wherein each of said antennas is a square spiral antenna having an outermost segment, a middle segment and an innermost segment, wherein a length of each outermost segment of is equal to about a quarter of the wavelength of the particular RF frequency or center frequency of the band of RF frequencies included in said particular RF frequency range, wherein a total length of each square spiral antenna is equal to about one half of said wavelength, and wherein the outermost segment of each of said antennas is operatively coupled to the associated one of said matching networks.

23. The energy harvesting circuit according to claim 22, wherein each square spiral antenna alone occupies a first spatial area, and wherein said antennas are provided in an antenna layout that occupies a second spatial area that is less than two times said first spatial area.

24. An energy harvesting method, comprising:

providing a plurality of antennas in an antenna plane, each of said antennas being tuned to the same particular RF frequency range;

receiving a plurality of RF signals using said antennas, each of said RF signals having said particular RF frequency range;

outputting from said antennas a plurality of AC signals in response to receiving said RF signals;

converting each of the AC signals into a respective DC voltage signal; and

summing each of the respective DC voltage signals together to create a combined DC voltage signal;

wherein said converting step is performed after first passing each of the AC signals through a respective matching network having an impedance chosen so as to cause a voltage level of the combined DC voltage signal to have at least a predetermined minimum value as said antenna plane is rotated about a first axis.

25. The energy harvesting method according to claim 24, wherein said first axis is a line taken through a center of a transmitting antenna of a source of each said RF signal having said particular RF frequency range.

26. An energy harvesting circuit, comprising:

a plurality of antennas provided in an antenna plane, each of said antennas being tuned to the same particular RF frequency range, each of said antennas being structured to receive an RF signal having said particular RF frequency range and output a respective AC signal;

a plurality of matching networks, each of said matching networks being operatively coupled to a respective one of said antennas and being structured to receive the AC signal output by said respective one of said antennas; and

a plurality of voltage boosting and rectifying circuits, each of said voltage boosting and rectifying circuits being operatively coupled to a respective one of said matching networks and being structured to receive the AC signal received by the respective one of said matching networks and output a DC voltage signal by converting the received AC signal into the DC voltage signal;

wherein the DC voltage signals output by the voltage boosting and rectifying circuits are summed together to create a combined DC voltage signal, and wherein an impedance of each of the matching networks is chosen by: (i) incrementally rotating said antenna plane through a plurality of angle increments about a first axis, (ii) trying a plurality of different inductance and capacitance value combinations for each LC tank circuit of each of the matching networks at each of said angle increments, (iii) measuring the combined DC voltage signal for each of the inductance and capacitance value combinations at each of said angle increments, (iv) determining which one of the inductance and capacitance value combinations produces a maximum voltage level for said measured combined DC voltage signals at any one of said angle increments; (v) choosing for each said LC tank circuit the one of the inductance and capacitance value combinations that produces the maximum voltage level.

27. An energy harvesting circuit, comprising:

a plurality of antennas, each of said antennas being tuned to the same particular RF frequency range, each of said antennas being structured to receive an RF signal having said particular RF frequency range and output a respective AC signal; and

phase shifting and rectifying circuitry operatively coupled to said antennas, said phase shifting and rectifying circuitry being structured to: (i) receive each said respective AC signal, (ii) create a plurality of out of phase AC signals by causing each said respective AC signal to be out of phase with one another, and (iii) convert each of said out of phase AC signals into a respective DC voltage signal;

wherein the DC voltage signals are summed together to create a combined DC voltage signal.

28. The energy harvesting circuit according to claim 27, wherein said phase shifting and rectifying circuitry comprises: (i) a plurality of phase shifting circuits, each of said phase shifting circuits being operatively coupled to a respective one of said antennas and creating a respective one of said out of phase AC signals, and (ii) a plurality of voltage boosting and rectifying circuits, each of said voltage boosting and rectifying circuits being operatively coupled to a respective one of said phase shifting circuits and converting a respective one of said of said out of phase AC signals into the respective DC voltage signal.

29. The energy harvesting circuit according to claim 28, wherein each of said voltage boosting and rectifying circuits is a charge pump.

30. The energy harvesting circuit according to claim 27, wherein a phase shift between each of said out of phase AC signals and each one of said out of phase AC signals that is adjacent in phase thereto is equal to 360 degrees divided by a total number of said antennas.

31. The energy harvesting circuit according to claim 27, wherein said phase shifting and rectifying circuitry includes a Cockroft-Walton multiplier.

32. An energy harvesting method, comprising:

providing a plurality of antennas, each of said antennas being tuned to the same particular RF frequency range;

receiving a plurality of RF signals using said antennas, each of said RF signals having said particular RF frequency range;

outputting from said antennas a plurality of AC signals in response to receiving said RF signals;

creating a plurality of out of phase AC signals by causing each of said AC signals to be out of phase with one another;

converting each of said out of phase AC signals into a respective DC voltage signal; and

summing each of the respective DC voltage signals together to create a combined DC voltage signal.

* * * * *

US008552597B2

(12) **United States Patent**
Song et al.

(10) **Patent No.: US 8,552,597 B2**
(45) **Date of Patent: Oct. 8, 2013**

(54) **PASSIVE RF ENERGY HARVESTING SCHEME FOR WIRELESS SENSOR**

(75) Inventors: **Zhen Song**, Plainsboro, NJ (US); **Chellury R. Sastry**, South Brunswick, NJ (US)

(73) Assignee: **Siemens Corporation**, Iselin, NJ (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 1542 days.

(21) Appl. No.: **11/691,545**

(22) Filed: **Mar. 27, 2007**

(65) **Prior Publication Data**

US 2012/0256492 A1 Oct. 11, 2012

Related U.S. Application Data

(60) Provisional application No. 60/744,028, filed on Mar. 31, 2006.

(51) **Int. Cl.**
H02J 17/00 (2006.01)

(52) **U.S. Cl.**
USPC **307/149**; 307/104

(58) **Field of Classification Search**
USPC 307/104, 149
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2,462,102	A	2/1949	Istvan	
2,554,270	A	5/1951	Rosenberg	
2,805,332	A	9/1957	Bell	
3,327,222	A	6/1967	King, Jr	
3,987,448	A *	10/1976	Scheppman	343/702
3,996,522	A	12/1976	Koch et al.	
4,229,742	A	10/1980	Rotunda	
5,075,670	A	12/1991	Bower et al.	
5,151,848	A	9/1992	Finello	
5,585,999	A	12/1996	De Long et al.	
6,037,743	A	3/2000	White et al.	
6,259,372	B1	7/2001	Taranowski et al.	
6,417,653	B1	7/2002	Massie et al.	
6,691,236	B1 *	2/2004	Atkinson	713/320
6,825,754	B1	11/2004	Rolin	
6,856,291	B2 *	2/2005	Mickle et al.	343/701
6,882,128	B1	4/2005	Rahmel et al.	

(Continued)

FOREIGN PATENT DOCUMENTS

WO	WO 01 37497 A1	5/2001

OTHER PUBLICATIONS

Pervasive Computing, Published by the IEEE CS and IEEE ComSoc, 1536-1268/05, copyright 2005 IEEE. Energy Scavenging for Mobile and Wireless Electronics, by Joseph A. Paradiso and Thad Starner. pp. 18-27.*

Primary Examiner — Fritz M Fleming

(57) **ABSTRACT**

A radio frequency energy harvester harvests RF energy from the environment. The energy harvester may include a local power source. The energy harvester may have an associated crystal radio antenna that harvests AM radio waves ubiquitously. The radio frequency energy harvested may be stored via a super capacitor. The stored energy may power a self-contained processing unit directly when the level of the energy harvested remains above a threshold level. However, if the energy harvested drops below the threshold level, the processing unit, which may include a radio, may switch to the local power source and/or a low power mode of operation. The low power mode of operation may reduce the amount of processing and/or transmitting by the radio. The processing unit may be part of a network of radios. The energy harvester may supply energy for a variety of applications, including building automation, industrial automation, power generation, and healthcare.

8 Claims, 9 Drawing Sheets

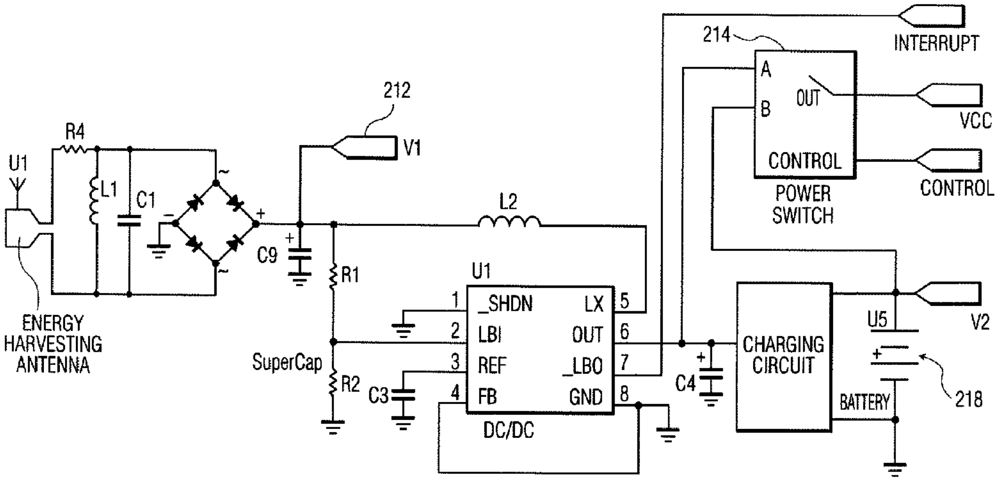

(56) References Cited

U.S. PATENT DOCUMENTS

6,924,772	B2	8/2005	Kiernan, Jr. et al.	
7,038,630	B1	5/2006	Bally et al.	
7,089,089	B2	8/2006	Cumming et al.	
7,132,757	B2 *	11/2006	Steigerwald et al.	290/1 R
7,139,239	B2	11/2006	McFarland	
2002/0116460	A1	8/2002	Treister et al.	
2005/0115600	A1 *	6/2005	DeSteese et al.	136/205

* cited by examiner

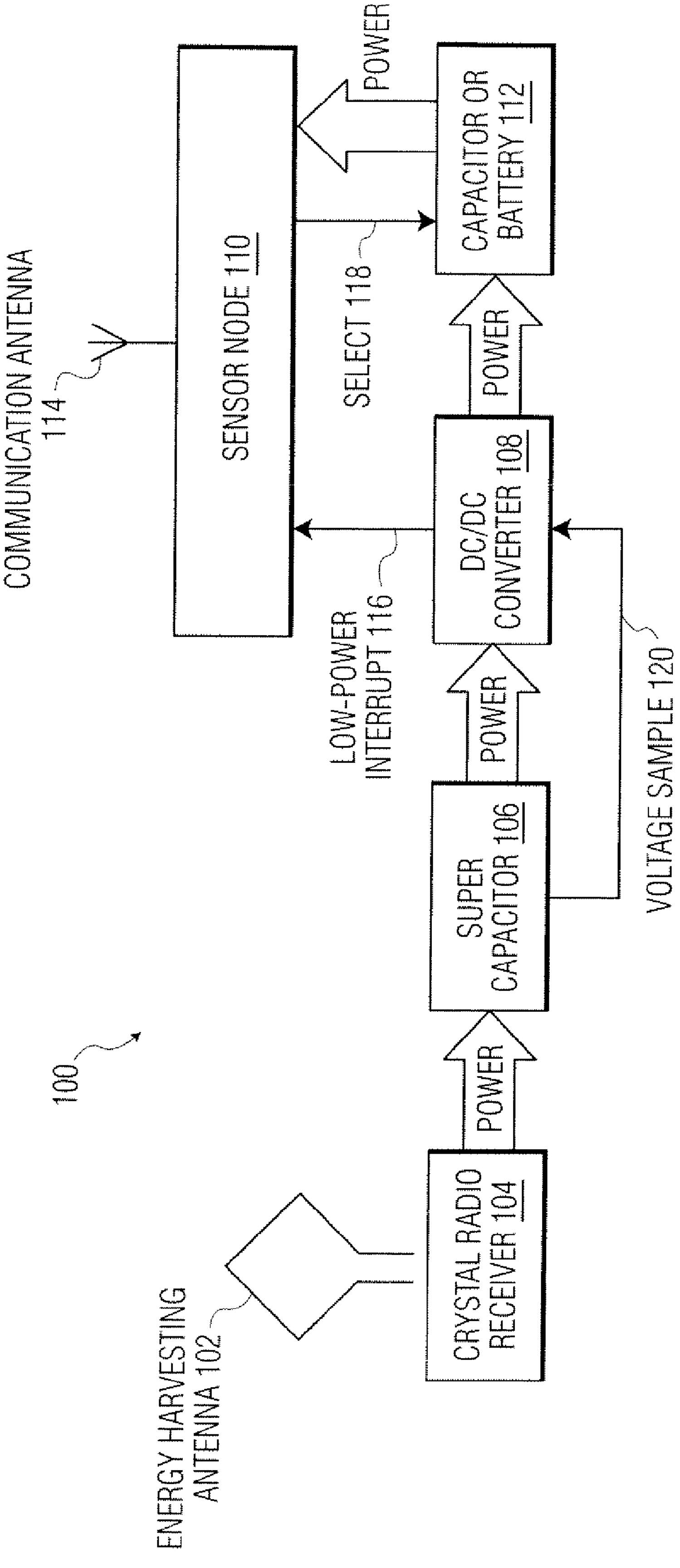
100
ENERGY HARVESTING ANTENNA 102
CRYSTAL RADIO RECEIVER 104
POWER
SUPER CAPACITOR 106
POWER
DC/DC CONVERTER 108
POWER
CAPACITOR OR BATTERY 112
POWER
SENSOR NODE 110
COMMUNICATION ANTENNA 114
LOW-POWER INTERRUPT 116
SELECT 118
VOLTAGE SAMPLE 120

FIG. 1

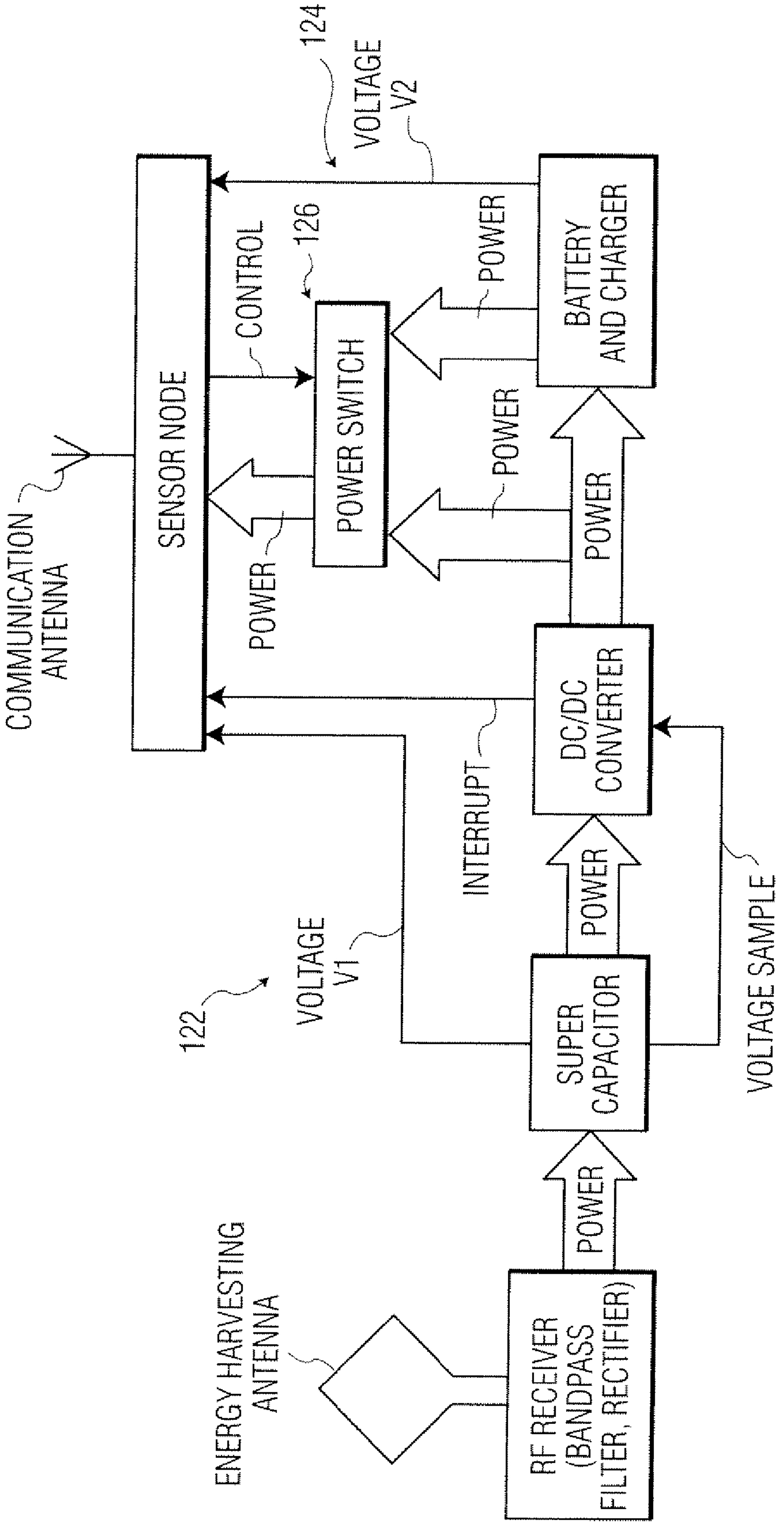
ENERGY HARVESTING ANTENNA
RF RECEIVER (BANDPASS FILTER, RECTIFIER)
POWER
SUPER CAPACITOR
VOLTAGE V1
122
POWER
VOLTAGE SAMPLE
DC/DC CONVERTER
INTERRUPT
POWER
COMMUNICATION ANTENNA
SENSOR NODE
POWER
POWER SWITCH
POWER
CONTROL
126
POWER
BATTERY AND CHARGER
VOLTAGE V2
124

FIG. 1a

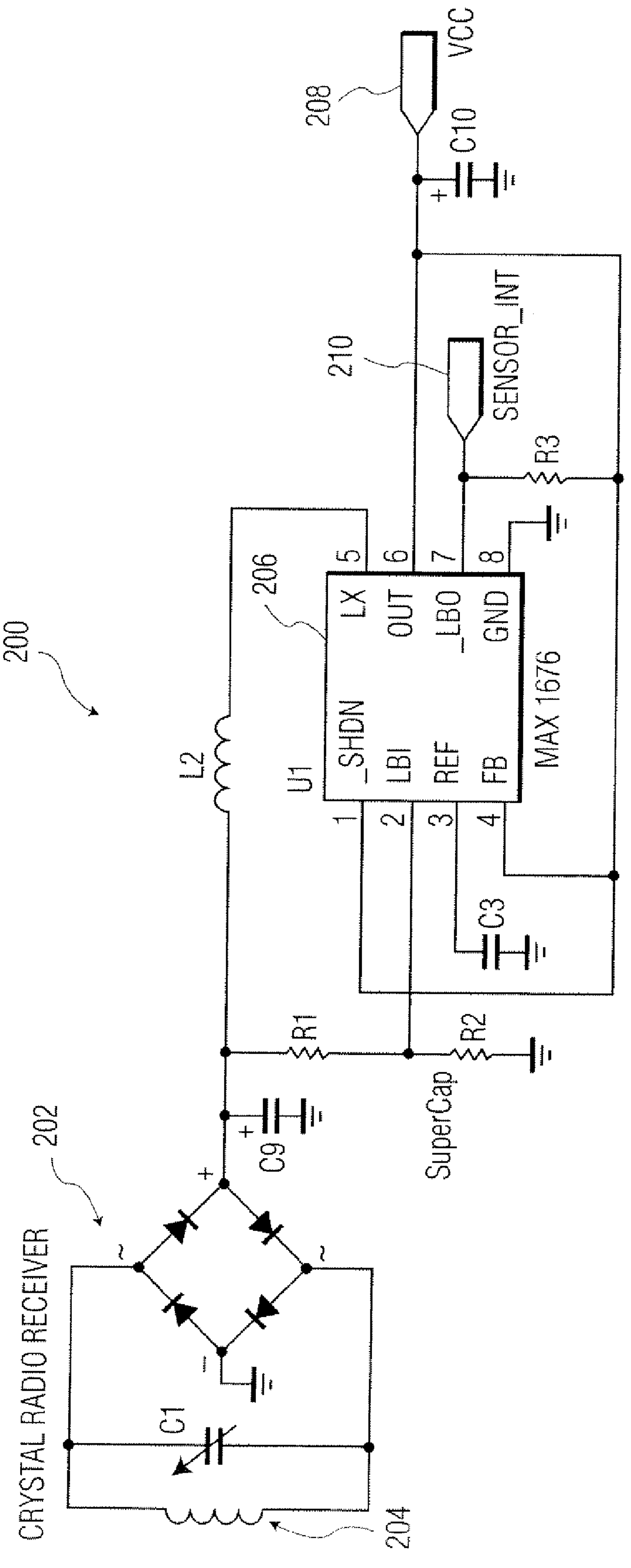
200
CRYSTAL RADIO RECEIVER
202
204
C1
~
~
+
-
C9
+
SuperCap
R1
R2
L2
U1
206
_SHDN
LBI
REF
FB
LX
OUT
_LBO
GND
MAX 1676
1
2
3
4
5
6
7
8
C3
R3
210
SENSOR_INT
208
VCC
C10
+

FIG. 2

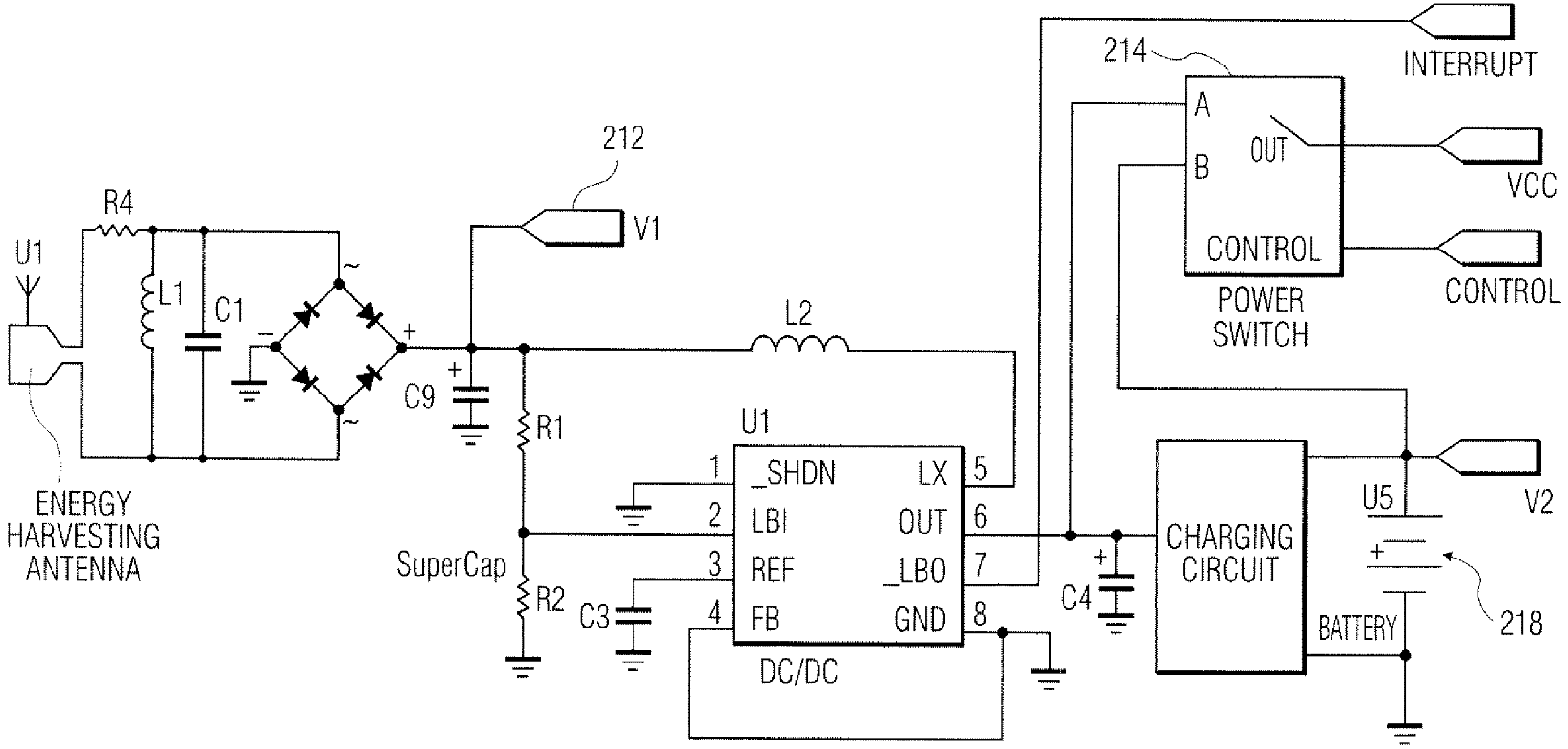
INTERRUPT
214
A
B
OUT
CONTROL
POWER
SWITCH
VCC
CONTROL
212
V1
R4
U1
L1
C1
~
−
+
~
C9
+
R1
L2
U1
1
2
3
4
_SHDN
LBI
REF
FB
LX
OUT
_LBO
GND
5
6
7
8
DC/DC
ENERGY
HARVESTING
ANTENNA
SuperCap
R2
C3
C4
+
CHARGING
CIRCUIT
U5
+
BATTERY
V2
218

FIG. 2a

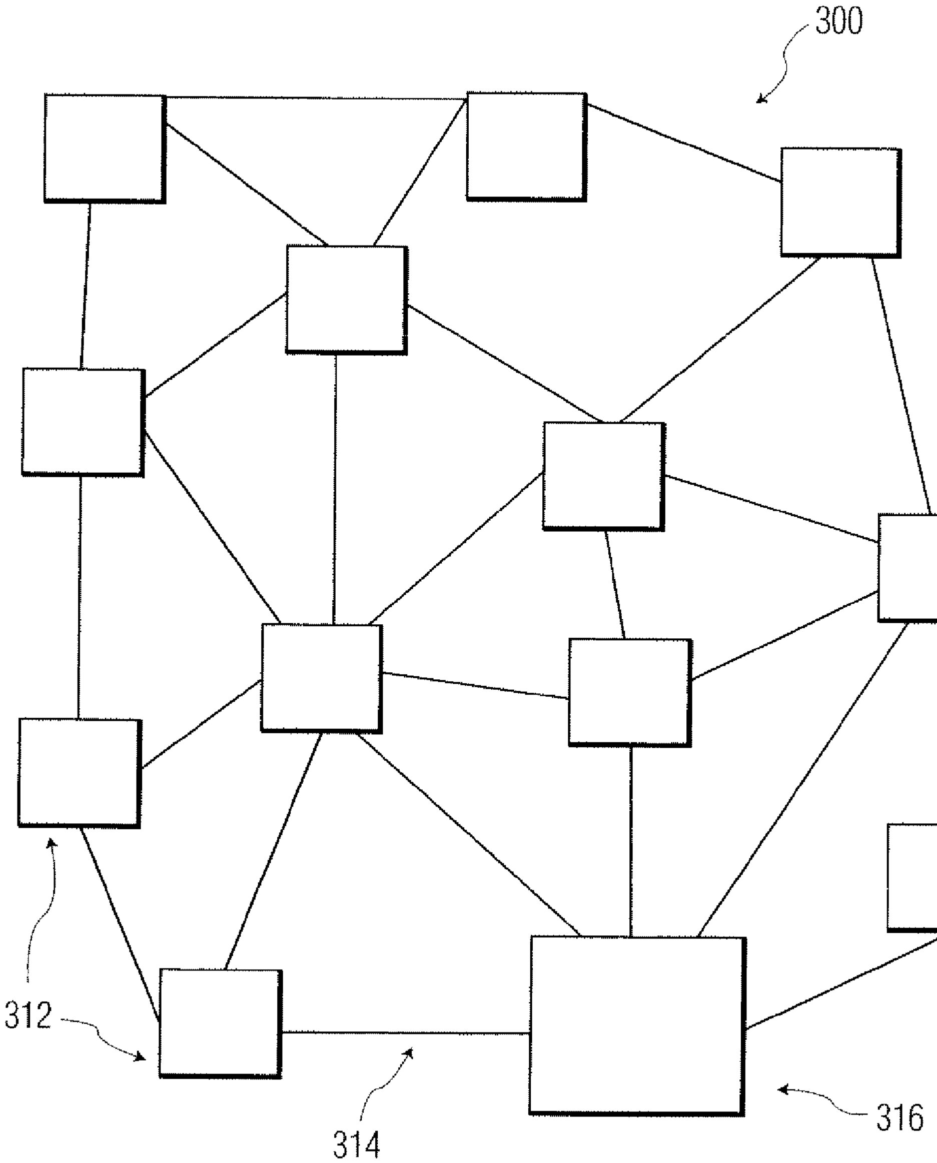
300
312
314
316

FIG. 3

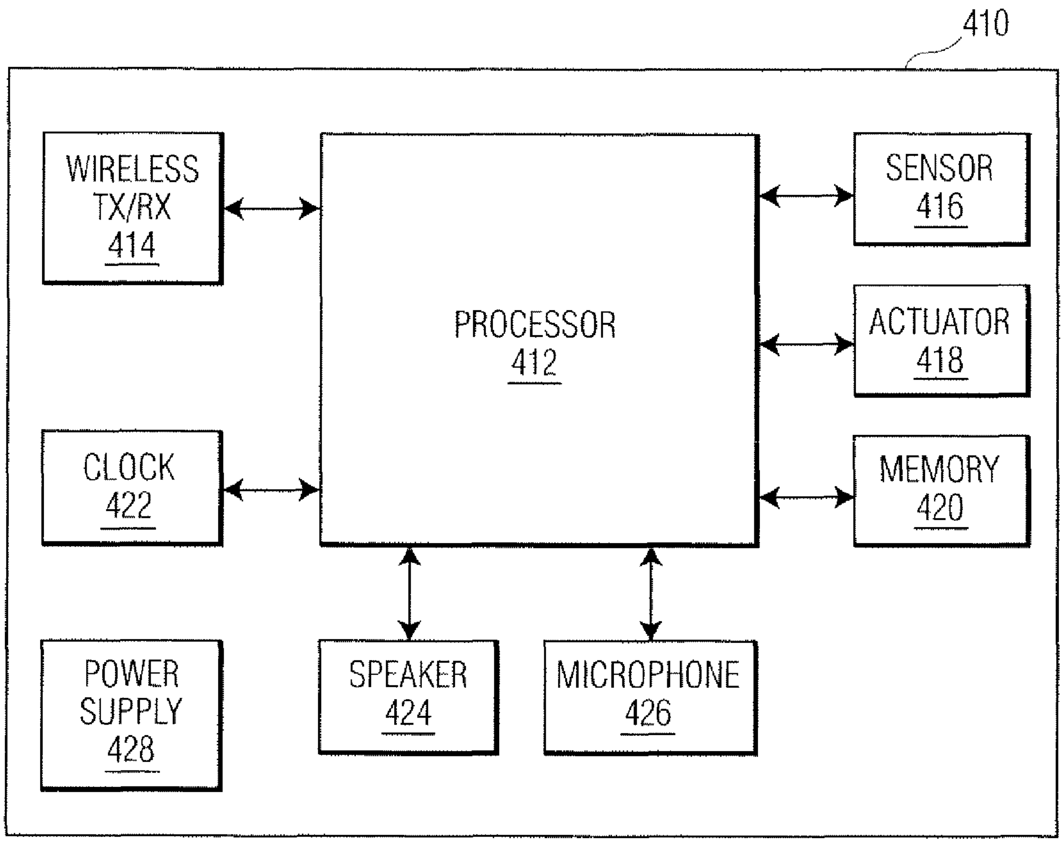
410
WIRELESS TX/RX 414
PROCESSOR 412
SENSOR 416
ACTUATOR 418
CLOCK 422
MEMORY 420
POWER SUPPLY 428
SPEAKER 424
MICROPHONE 426

FIG. 4

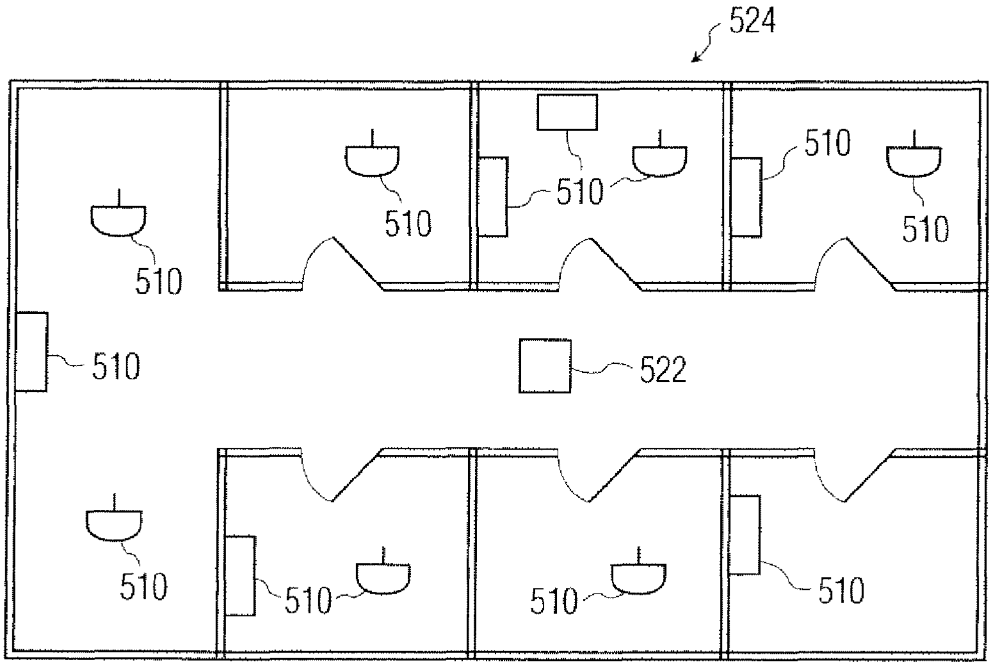

FIG. 5

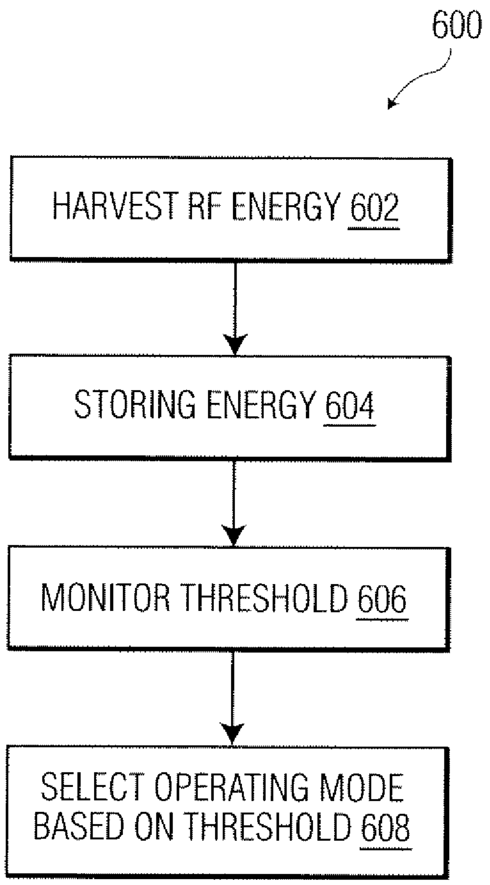
600
HARVEST RF ENERGY 602
STORING ENERGY 604
MONITOR THRESHOLD 606
SELECT OPERATING MODE BASED ON THRESHOLD 608

FIG. 6

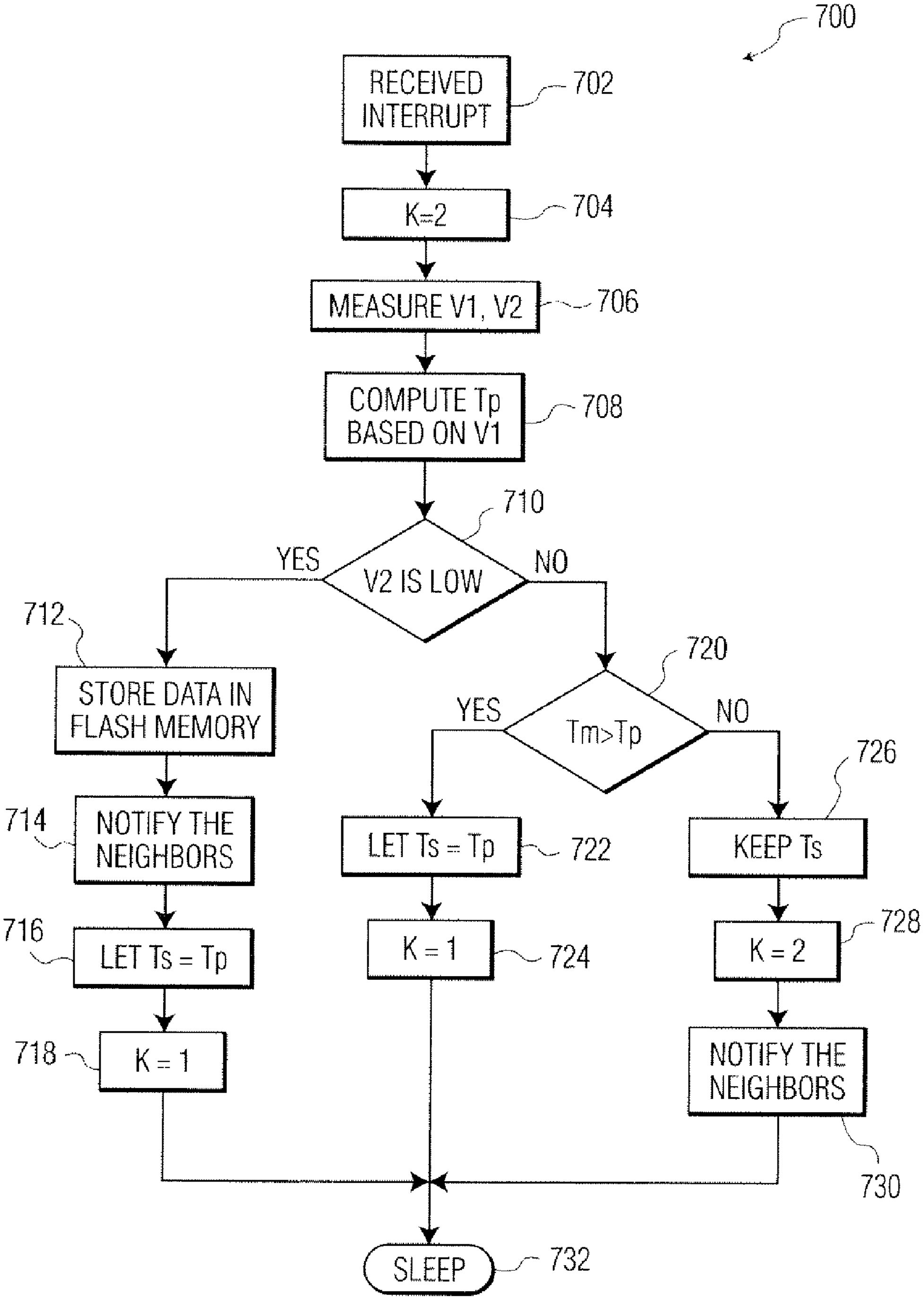
700
RECEIVED INTERRUPT
702
K=2
704
MEASURE V1, V2
706
COMPUTE Tp BASED ON V1
708
710
V2 IS LOW
YES
NO
712
STORE DATA IN FLASH MEMORY
714
NOTIFY THE NEIGHBORS
716
LET Ts = Tp
718
K = 1
720
Tm>Tp
YES
NO
LET Ts = Tp
722
K = 1
724
726
KEEP Ts
K = 2
728
NOTIFY THE NEIGHBORS
730
SLEEP
732

FIG. 7

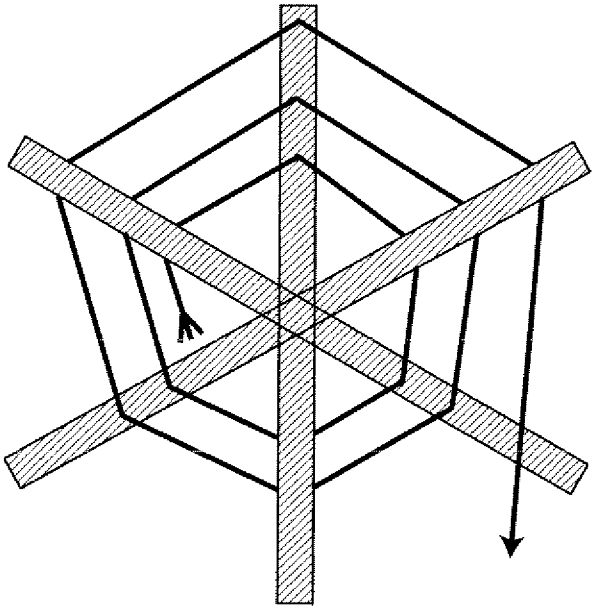

FIG. 8

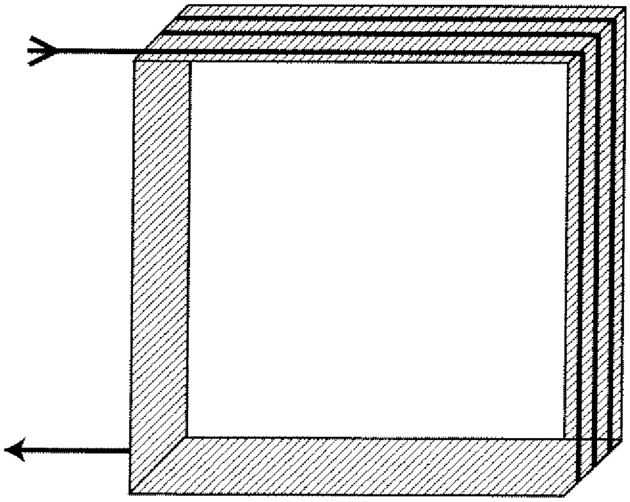

FIG. 9

PASSIVE RF ENERGY HARVESTING SCHEME FOR WIRELESS SENSOR

PRIORITY AND CROSS-REFERENCE TO RELATED APPLICATION

This application claims priority under 35 U.S.C. §119(e) to provisional application Ser. No. 60/744,028, filed on Mar. 31, 2006, which is incorporated by reference in its entirety herein.

BACKGROUND

The present embodiments relate generally to harvesting environmental energy. More particularly, the present embodiments relate to harvesting energy from radio frequency (RF) signals to power a sensor node and/or wireless radio.

Wireless sensor networks (WSNs) of so-called "smart" sensors can be used in remote monitoring and control applications. However, in almost all applications that involve WSNs, there is a trade off between the need to sense and communicate useful information and the requirement to do so in an energy efficient manner. A WSN, once deployed, may be intended to work for long periods of time with only minimal human intervention. For instance, remote monitoring applications may include individual sensor nodes, numbering into the hundreds or even thousands, that each run on a limited supply of energy, such as a battery. As a result, with conventional sensor nodes, excessive on-board processing and, perhaps more importantly, frequent invocation of the sensor node's radio may rapidly deplete the on-board energy source. Therefore, there is a need to develop sensor network protocols and algorithms with energy-efficient and self-organizing capabilities.

Additionally, replacing depleted batteries may be inefficient or impractical. A network may include a large number of individual sensor nodes that may be difficult to access due to being implanted within a wall, a ceiling, a larger device, or a patient. Accordingly, there is also a need for novel energy harvesting schemes operable to power wireless sensor nodes, as well as automatically replenish on-board energy storage units.

BRIEF SUMMARY

By way of introduction, the embodiments described below include methods, processes, apparatuses, instructions, or systems for harvesting energy from freely available radio frequency (RF) signals. The energy harvested may power processing units, such as self-contained processing units or sensor nodes. A "passive RF energy harvesting system," or "harvesting system," may be connected to a processing unit having a processor and delivers power supply to the processing unit. Each harvesting system may be associated with a crystal radio receiver and/or an all pass or band pass filter operable to harvest energy from all radio frequency bands simultaneously or ubiquitously. As a result, more of the available RF energy may be harvested from the existing electromagnetic signals. The harvesting is ubiquitous since no tuning is required if the harvester is moved from place to place. Most of the RF energy scavenged may be harvested from amplitude modulated (AM) radio signals. The harvesting system may have a super capacitor that temporarily stores the RF energy captured. If the level of RF energy being harvested falls below a threshold level, an interrupt signal may be generated that directs the processing unit to conserve energy. For instance, the processing unit may limit on-board processing or reduce the number and/or strength of signals transmitted via a wireless radio, or both.

In a first aspect, a method harvests passive radio frequency energy. The method includes harvesting radio frequency energy using a crystal receiver. The method also includes powering a processing unit directly from the radio frequency energy being harvested when the radio frequency energy being harvested or an associated signal is approximately equal to or above a threshold level, and alternatively powering the processing unit from an on-board power source when the radio frequency energy being harvested or the associated signal is below the threshold level.

In a second aspect, a method harvests passive radio frequency energy. The method includes harvesting radio frequency energy via a crystal receiver and storing the radio frequency energy harvested on a super capacitor for subsequent use.

In a third aspect, a system harvests passive radio frequency energy. The system includes a network of radios within a building operable to direct the operation of building equipment to control a building environment of a building and at least one radio within the network being a sensor node operable to harvest radio frequency energy via a crystal receiver, the radio frequency energy harvested being used to power the sensor node.

In a fourth aspect, a computer-readable medium having instructions executable on a computer stored thereon is described. The instructions include monitoring for an interrupt signal that is triggered when the radio frequency energy being harvested via a crystal receiver falls below certain threshold, the radio frequency energy being used to at least partially power a sensor node or charge the battery on the harvesting system. The instructions also include automatically reducing the amount of power being expended by the sensor node if the interrupt signal is detected.

The present invention is defined by the claims herein. Nothing in this section should be taken as a limitation on those claims. Further aspects and advantages of the invention are discussed below in conjunction with the preferred embodiments and may be later claimed independently or in combination.

BRIEF DESCRIPTION OF THE DRAWINGS

The components in the figures are not necessarily to scale, emphasis instead being placed upon illustrating the principles of the invention.

FIGS. **1** and **1***a* are system block diagrams of an exemplary passive RF energy harvesting systems;

FIGS. **2** and **2***a* are schematics of additional exemplary passive RF energy harvesting systems;

FIG. **3** is a schematic of an exemplary network of sensor nodes;

FIG. **4** is a block diagram of an exemplary sensor node;

FIG. **5** is a top plan view of an exemplary network of sensor nodes within a building;

FIG. **6** illustrates an exemplary method of harvesting radio frequency energy;

FIG. **7** illustrates an exemplary interrupt handling method; and

FIGS. **8** and **9** illustrate exemplary antennas.

DETAILED DESCRIPTION OF THE DRAWINGS AND PRESENTLY PREFERRED EMBODIMENTS

The present embodiments are directed toward methods, processes, apparatuses, instructions, or systems for harvest-

ing freely available radio frequency (RF) energy from the environment, such as signals generated from amplitude modulated (AM) and frequency modulated (FM) radio stations. The RF energy harvested is used to power a "sensor node." The sensor node may include a processor, a sensor, an actuator, a wireless radio, and/or other components. The sensor node may be part of a network of radios in direct or indirect communication with one another, such as a wireless sensor network (WSN).

The sensor node may have an associated crystal radio receiver. The crystal radio receiver may operate as an all pass filter to harvest energy from all radio frequency bands simultaneously or ubiquitously. As such, more of the available RF energy may be captured for the given electromagnetic signals being received by the sensor node at a specific time.

The energy harvesting system may have an associated super capacitor that temporarily stores the RF energy received or a related signal, such as a voltage, a current, or other electrical parameter or signal associated with the RF energy harvested. In one embodiment, if a voltage associated with the RF energy being harvested falls below a given threshold, an interrupt signal may be generated that directs the sensor node to switch to a "low-power" mode to conserve energy. In low-power mode, the sensor node may reduce the amount of processing that an on-board processor performs. Alternatively or additionally, in low-power mode, the operation of an on-board wireless radio or transmitter may be altered and/or limited. For example, the wake-up period of the sensor nodes may be increased for energy saving purposes. Other manners of reduced sensor node operation may be employed to conserve energy being expended by the sensor node.

I. Crystal Radio Receivers

Generally, the embodiments discussed herein provide energy harvesting or scavenging for sensor nodes and/or wireless sensor networks (WSNs). In particular, AM RF signals may be harvested via crystal radio receivers to provide the primary, or even sole, source of power for the sensor nodes. Other crystal receivers may be used.

Crystal radios were popular by the 1950's, but are now almost obsolete. One of the key features of crystal radios is that they do not require batteries. The earplugs and/or speakers of a crystal radio may be powered solely by amplitude modulated (AM) radio signals. The crystal radio may employ a piezo-electric crystal, dielectric disk, oscillating crystal, or another type of crystal. The crystal may oscillate or vibrate due to the radio frequency signals being received. Exemplary crystal radio receivers are disclosed by U.S. Pat. Nos. 2,554,270, 2,805,332, 3,327,222, and 3,996,522, which are all incorporated herein by reference in their entireties. Other crystal radio receivers may be used.

The crystal radio technology may be used in new applications due to the advancement of semi-conductors and reduced power requirements of sensor nodes. In other words, the available energy that may be harvested from radio frequency signals via a crystal radio receiver, may contain sufficient energy to power milliwatt level instruments and micro-devices. Additionally, the size of a crystal radio antenna may be configured to be relatively small, ranging from a few to tens of centimeters. An antenna of this size may be used with a number of today's wireless sensor node applications.

Harvesting small amounts of energy from radio signals has been studied before, although not in the context of powering nodes in a WSN by AM RF signals. Previous studies have concluded that RF energy is of little utility unless a receiver is very close to a transmitter, such as a cell phone station. An alternate approach would be to introduce an RF signal transmitter for the explicit purpose of transmitting RF energy to a receiver that then harvests the RF energy to perform certain tasks. In other embodiments, such an approach may not be practical in the context of WSNs.

The energy available from RF signals in the unlicensed band, for e.g., the 2.4 GHz spectrum, may be typically too limited to be of any practical use. Such RF power harvesting may be impractical due to (1) very limited power being available (for an antenna size of 100 cm^2, the received power may be less than 0.1 milliwatt), and (2) very limited range (not only may a transmitter be required, but the sensor also may have to be placed very close to the transmitter). On the other hand, the principles on which crystal radios operate may finally permit the successful harvesting of energy from freely available RF signals.

As AM radio signals may be strong enough to power an earplug or a speaker, the energy available from RF signals may be more than past estimates indicated. A typical sensor node may require about 20 milliamps in active mode and 5 microamps in sleep mode, and the operating voltage required may be about 3 volts. Thus, the required power for a typical sensor node is comparable to the power that may be harvested via RF signals. Additionally, low-duty low-energy sensors may not require a transmitter, further reducing the amount of energy needed to power the sensor node. Therefore, RF signals may be a viable "free" and persistent (available 24 hours a day) energy source for a considerable amount of a given radio broadcast receiving area.

II. Estimation of Received Radio Energy

Table 1 below illustrates an estimation of the available RF energy in Princeton, N.J. from AM and FM RF signals transmitted by area radio stations. Through this rough estimation analysis, the "free" RF energy in the Princeton area is shown to be within the 10 milliwatt level. The estimation is based upon the Friis propagation model,

$$P_r = \frac{p_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L},$$

TABLE 1

Radio Stations that can be Received in Princeton, N.J.

Call Sign	Frequency	Dist./Signal	City	Format	Estimated Received Energy (w)
WNJT	88.1 FM	7.1 mi.	Trenton, NJ	Public Radio	$2.8 * 10^{-6}$
WWFM	89.1 FM	6.7 mi.	Trenton, NJ	Classical	$3.1 * 10^{-6}$
WPST	94.5 RM	17.7 mi.	Trenton, NJ	Top-40	$0.40 * 10^{-6}$
WTHK	97.5 FM	12.5 mi.	Burlington, NJ	Classic Rock	$0.75 * 10^{-6}$
WMGQ	98.3 FM	10.7 mi.	New Brunswick, NJ	Adult Contemporary	$1.0 * 10^{-6}$

TABLE 1-continued

Radio Stations that can be Received in Princeton, N.J.

Call Sign	Frequency	Dist./Signal	City	Format	Estimated Received Energy (w)
WAWZ	99.1 FM	19.3 mi.	Zarephath, NJ	Christian Contemporary	$0.30 * 10^{-6}$
WKXW	101.5 FM	7.1 mi.	Trenton, NJ	News/Talk	$2.1 * 10^{-6}$
WPRB	103.3 FM	7.2 mi.	Princeton, NJ	College	$2.0 * 10^{-6}$
WWJZ	640 AM	24.5 mi.	Mount Holly, NJ	Children's	$9.1 * 10^{-3}$
WFAN	660 AM	55.0 mi.	New York, NY	Sports	$1.7 * 10^{-3}$
WOR	710 AM	40.4 mi.	New York, NY	Talk	$2.7 * 10^{-3}$
WABC	770 AM	46.1 mi.	New York, NY	News/Talk	$1.8 * 10^{-3}$
WCBS	880 AM	55.0 mi.	New York, NY	News/Talk	$0.95 * 10^{-3}$
WPHT	1210 AM	32.9 mi.	Philadelphia, PA	Talk	$1.4 * 10^{-3}$
WHWH	1350 AM	9.5 mi.	Princeton, NJ	Business News	$13.5 * 10^{-3}$

Here, p_t, p_r are the power of the transmitter and the receiver, respectively; G_t, G_r are the antenna gains of the transmitter and the receiver, respectively; λ is the wavelength of the RF signal in meters; d is the distance between the transmitter and the receiver; and L is the system losses ($L>1$). The estimation is based upon various assumptions, such as (1) radio station information (frequency and distance), (2) the transmitting power of AM radio stations being 10 kilowatt and FM stations being 5 kilowatt, (3) the antenna gains, G, are 1 (for a properly designed antenna, $G>1$), and (4) the system losses, L, are 1.

The results of the estimation indicate that (1) the total power received was 31.120 milliwatts, (2) the AM power received was 31.108 milliwatts, or 99.96% of the total power, and (3) the FM power received was 0.0126 milliwatt, or 0.04% of the total power. The estimation is based on ideal environments. In practice, L is more than 1 and the received power is less than the ideal case. Observing the Friis propagation model reveals that the effect of wavelength on the received power is more than that of the transmission power. Accordingly, signals with longer wavelength, such as AM signals, are much more suitable for RF energy harvesting.

A typical sensor node may require about 41 milliwatts in active mode and 1.5 microwatts in sleep mode. This power requirement is of the same order of magnitude as the power that is available from RF signals. With reasonable conditions, such as the sensor nodes being located near one or more AM radio stations and/or using relatively efficient antennas, the sensor nodes may be powered solely by the AM signal(s) and be kept running uninhibited all of the time. Alternatively, the sensor nodes may be powered primarily by the AM signal(s), if backed up by an on-board energy source.

III. Estimation of Available Energy from Other RF Signal Sources

As noted above, AM signals may include much more energy available for harvesting than FM signals. Energy from other RF signal sources also may be very small when compared with AM signals. For example, the RF signals from cell phones quickly attenuate as a function of distance. As a result, a cell phone signal may be used as a source of energy to power a sensor node only if the distance from the cell phone to the sensor node is about one meter or less.

Energy also may be received from IEEE 802.11.x protocol-based devices and television sets. As IEEE 802.11.x devices operate in the 2.4 GHz band, the energy available for harvesting may be even less than what is available from cell phones. Similarly, due to their high frequency, television signals may contribute very little towards harvestable energy. Therefore, AM signals may provide the best freely available RF source of power for wireless sensor nodes.

IV. Hybrid Scheme

A number of harvesting schemes may be utilized to capture energy from RF signals. As an example, harvesting energy from AM signals may not only be useful for indoor sensor nodes, but could also prove fruitful for outdoor sensor deployments.

In one aspect, the RF harvesting schemes discussed herein may combine two concepts. A sensor node may include crystal radio technology that is operable in conjunction with an energy storage unit, such as a battery and/or a capacitor. Once the output of the crystal radio receiver is connected to act as a source of energy and/or a "charger" to the energy storage unit, the system may be ready for use.

The sensor node may include a RF energy harvesting device that includes a so-called "energy harvesting antenna" operable to capture RF energy. The antenna may harvest AM, as well as FM, or RF signals. The antenna may be a spider antenna, braided antenna, omni-directional antenna, double-braided polyester rope, loop antenna, air coil loop antenna, sparrow antenna, Yagi-Uda antenna, or other type of antenna.

For example, FIG. **8** is an exemplary spider antenna. As shown, the antenna may have a number of legs of differing size. FIG. **9** is another exemplary antenna. Other shapes may be used, such as circular or disk shapes.

The antenna may be a laminated antenna having a number of wires etched onto a semi-conductor or other micro-device. The antenna wires may have either a flat or rounded shape, or be disk shaped. The antenna may have a number of spokes or legs, such as the spider antenna disclosed by U.S. Pat. No. 4,229,742, or the antenna may be a grid antenna similar to the one disclosed by U.S. Pat. No. 7,038,630, both of which are incorporated herein by reference in their entireties. As the antenna characteristics may be determined based upon the bandwidth and gain properties desired for a specific application, alternative antennas may be used.

In one embodiment, for both convenience and cost reduction purposes, a 1950's type of coil or wire antenna connected to the crystal radio receiver may be replaced by modern flexible printed circuit board. Printed circuit board similar to that used in conventional radio frequency identification (RFID) devices may be used. An exemplary RFID antenna is disclosed by U.S. Pat. No. 6,825,754, which is incorporated herein by reference in its entirety. Other types of printed circuit board material and shapes may be used as the antenna.

Once the RF energy is scavenged, the sensor node may regulate the energy within the crystal radio circuit and convert

the RF signals received to a DC signal, such as via a rectifier, without any frequency filtering. As an example, the RF signals harvested may not be filtered via low, high, or band pass filters. After the RF energy harvested passes through the rectifier, the energy may be stored on a super capacitor. However, since the voltage on the two sides of the super capacitor may not satisfy the requirements of the sensor node, a DC/DC converter and/or another capacitor may be employed (such as shown in FIG. **2**, discussed below).

Clean DC power may be available on both sides of the second capacitor. In addition, the DC/DC converter may sample the voltage on the super capacitor, and issue an interrupt to the sensor node once the voltage, which is an indication of the energy being stored and/or harvested, is lower than a certain threshold. The sensor node may then switch to the on-board power source of the harvesting system, such as a battery, a capacitor, or other energy storage device. If the energy being harvested provides a voltage level greater than that of an on-board energy storage device, the energy being harvested may power the sensor node directly, as well as provide a trickle charge to the on-board energy storage device. On the other hand, if the energy being harvested provides a voltage level lower than of the on-board energy storage device, the on-board energy storage device may provide power to the sensor node via a trickle discharge. The harvested RF energy is used to charge the battery under this case.

Therefore, the wireless and self-contained sensor nodes may employ a "hybrid" RF energy harvesting scheme. Each wireless and self-contained sensor node may be equipped with (1) a RE-based energy harvester and (2) an on-board power source. As long as there is enough available power from the RF energy harvester, the sensor node may operate using the RF power. Intermittently, however, the power from the harvester may not be at the threshold required to operate the sensor node. As a result, an interrupt may be generated that switches the sensor node to the on-board power source. The hybrid scheme ensures that the sensor node may be continuously operational while simultaneously greatly lengthening the life-time of the on-board power source.

There are some distinctions between RF energy harvesting as discussed herein and conventional harvesting schemes. With respect to solar cell techniques, the proposed RF harvesting schemes are not affected by the weather or time of day (darkness). The proposed RF harvesting schemes may provide a continuous power supply as most radio stations broadcast around the clock.

Additionally, the proposed RF harvesting schemes may provide more power than solar cells, such as with indoor environments, and at less cost, as typical solar cells may be more expensive than an RF antenna or coil. The costs associated with the proposed schemes may be further reduced if the antenna is implemented using technology similar to that of a RFID tag.

As compared to biomechanics-based harvesting schemes, the proposed RF harvesting schemes may be more suitable for sensor node applications. Conventional biomechanics harvesting methods tend to provide energy on an inconsistent basis and inconveniently toggle sensors periodically to recharge the power source. On the other hand, the proposed RF harvesting schemes may provide a consistent milliwatt-level power source. Toggling may be used with RF harvesting.

With respect to vibration or thermoelectric methods, it may be more difficult to find a vibration or heat source than it is to locate an AM radio station, especially in populated regions of the United States. When a free source is not available, the vibration or thermoelectric based schemes may require an artificial source. In many environments, such as an office or building, installation of such artificial energy sources may not be feasible.

In sum, new RF energy harvesting schemes are described that provide energy to self-contained sensor nodes. The schemes provide an improved crystal radio receiver and harvest AM RF signals as a source of power. The RF harvesting schemes are suitable for indoor applications that involve wireless sensor networks, such as building automation applications.

The RF harvesting schemes may be especially useful in areas having AM stations within about 20 kilometers to satisfy typical sensor node power requirements. However, the effective distance of the harvesting schemes may be extended. The antennas employed may be configured to harvest more energy and provide enhanced gain. Sensor nodes may be configured to be more energy efficient. For example, the sensor nodes may be programmed to operate more efficiently, such as spending more time in various reduced power modes of operation. Additionally, the sensor nodes equipped with RF energy harvesters may be supplemented with other types of energy harvesters, such as solar, piezo-electric, MEMS (micro-electro-mechanical systems), biomechanical, thermoelectric, and other harvesters. The RF harvesting schemes, except for requiring relatively nearby AM radio stations, may have no other substantial apparent limitations, such as the weather, periodic toggling, or other limits.

V. Exemplary RF Energy Harvesting Systems

FIG. **1** illustrates an exemplary passive RF energy harvesting system **100**. The system **100** may include an energy harvesting antenna **102**, a crystal radio receiver **104**, a super capacitor **106**, a DC/DC converter **108**, a sensor node **110**, a local power source **112**, and/or a communication antenna **114**. The system may include additional, fewer, or alternate components.

The crystal radio receiver **104** or crystal set may employ an antenna **102**. The crystal radio receiver **104** may not require a battery or power source except the power received from radio waves via the antenna **102**. The crystal radio receiver **104** may include an antenna wire, a tuning coil of copper wire, a rectifier, and other components. The crystal radio receiver **104** may be a passive radio receiver, not requiring an actively powered amplifier associated with conventional radios. The crystal radio receiver **104** may preserve electrical power from RF signals received via the antenna **102** and convert the RF signals to electronic power. The crystal radio receiver **104** may include various types of antennas and various "low-tech" or traditional electrical circuitry components, such as capacitors, resistors, inductors, and diodes. As such the crystal radio receiver **104** may be simpler and less expensive to manufacture than current "smart" radios, such as cellular phones.

The antenna **102** may be a conventional crystal radio antenna and include copper or silver wire coils, or a ferrite coil. The antenna **102** may be a Yagi, log-periodic, fiberglass, cable, repeater, or other type of antenna, including those discussed elsewhere herein. The antenna **102** may be manufactured from printed circuit board, aluminum, copper, silver, other metals, or other materials. The antenna may omni-directionally collect all of the available AM radio signals currently being received. The shape, size, circumference, and composition of the antenna **102** may be altered for a specific application and/or to enhance the gain associated with the antenna **102**.

The antenna **102** and crystal radio receiver **104** may act independently or in conjunction to harvest RF energy ubiquitously. The crystal radio receiver **104** may require no tuning and resonant with all of the available frequencies. As a result,

energy may be collected simultaneously from a number of radio stations, which may increase the average level of RF energy being harvested. This may be in contrast with typical receivers that may employ RC resonant circuits, which eliminate some RF signals from being harvested. In one embodiment, the antenna may have a physical length of between 10 and 20 cm.

The low power mode may include a number of algorithms and/or software instructions stored in the sensor node that direct the sensor node or an associated processor to conserve or reduce the amount of energy that the sensor node is using. A processing unit on the sensor node may perform reduced or limited processing in low power mode. The radio may change the message routing being employed based upon the amount of RF energy being sensed or harvested. A transmitter may transmit less often in low power mode, or not at all until the level of the RF energy being harvested returns to above a threshold or self-sustaining level. The transmitter may alter the type of signal being transmitted in low power mode. For example, the sensor node may not relay all the received communication packets, and/or the transmitter on the sensor node may transmit at a reduced power level. The transmission at a reduced power level may be intended to only reach the next or closest available sensor node or radio within the network.

The low power mode may include buffering or storing data, instead of constant or even intermittent transmission of information to other sensor nodes. With typical sensor nodes, the operation of the radio and/or transmitting information may require a substantial amount of energy. Therefore, reducing radio transmissions in low power mode may enhance the effectiveness of the RF harvesting schemes discussed herein.

The super capacitor **106** may store the RF energy being harvested. The capacitance of an electrode associated with the super capacitor **106** may be several hundred times the capacitance of a conventional double-layer capacitor. The super capacitor **106** may have one or more electrodes manufactured from titanium, palladium, aluminum, silicon, or other metals. Alternatively, the electrodes may be doped or have titanium, palladium, aluminum, silicon, or other metal coatings. The super capacitor **106** may be similar to those disclosed by U.S. Pat. Nos. 5,585,999 and 5,151,848, which are incorporated herein by reference in their entireties. Alternate types of super or other capacitors may be used.

The DC-to-DC converter **108** may operate to produce an output voltage signal and/or an interrupt signal **116** based upon a signal received from the super capacitor **106** or an electrical parameter of the super capacitor **106** monitored, such as a voltage, current, or other signal. The DC-to-DC converter **108** may be similar to the DC-to-DC converter disclosed by U.S. Pat. No. 6,417,653, which is incorporated herein by reference in its entirety. Alternate DC-to-DC converters may be used.

In one embodiment, the DC-to-DC converter **108** samples the voltage **120** on the super capacitor **106**. The voltage sampled may be a reflection the current level of RF energy being harvested crystal radio receiver **104** and/or stored by the super capacitor **106**. If the voltage sampled falls below a threshold level, such as a level required by the sensor node **110** to operate normally (i.e., with no power operating restrictions), then the DC-to-DC converter may generate a low-power interrupt **116** signal. The sensor node **110** may switch to a low or reduced power operating mode upon the receipt of the low-power interrupt **116**.

The on-board energy storage unit may be a capacitor or battery **112**. The on-board energy storage unit **112** may be on a trickle charge when the voltage sampled **120** from the super capacitor **106** remains above the threshold level (such as the actual battery or capacitor voltage) with the carryover RF energy harvested directly powering the sensor node **110**. Alternatively, the RF energy harvested may be sent directly from the DC-to-DC converter **108** to the sensor node **110** when the level of RF energy being harvested remains at or above the threshold level.

On the other hand, if the low-power interrupt **116** is sent to the sensor node **110**, the sensor node **110** may select **118** or switch to the on-board storage unit **112** as a direct source of power. Alternatively, when the voltage from the DC-to-DC converter falls below the threshold level (such as the actual capacitor or battery voltage), the sensor node **110** may automatically start to be powered by the on-board power storage unit **112** as the storage unit voltage would be higher than the output voltage of the DC-to-DC converter, as can be envisioned by FIG. **1**. Other manners of powering of the sensor node from the super capacitor, either solely or in conjunction with the storage unit, may be used.

The sensor node **110** may be mounted on a device that includes the other components shown in FIG. **1** or be associated with the other components in another fashion. The sensor node **110** may include a communication antenna **114** that provides for wireless communication with other sensor nodes. The sensor node **110** may include the components shown in FIG. **4** and discussed below, in addition to the components illustrated by FIG. **1**. The sensor node may include additional, fewer, or alternate components.

FIG. **1***a* illustrates another exemplary passive RF energy harvesting system **100**. In one embodiment, the system may be targeted to facilitate the powering of sensor nodes in an ad-hoc wireless sensor network. As each sensor node may have communication functionality through an on-board RF module, a message communication subsystem may not be required to be a part of the RF energy harvesting system **100**. Also, the RF energy harvesting system **100** may be purely passive, and thus, no registration with the FCC is required.

As shown in FIG. **1***a*, the system **100** may be connected to the sensor node and its communication antenna. The system **100** may be a band pass and/or all pass filter to match all AM frequencies in order to provide ubiquitous energy harvesting capability and also to harness the maximum available harvestable energy.

As shown in FIG. **1***a*, a super capacitor is introduced as an energy buffer to the rechargeable batteries. Super capacitors have large number of recharge cycles, but limited energy capacity when compared to rechargeable batteries. In one embodiment, since there is bound to be some variation of RF signal strength from time to time and place to place, the harvested RF energy may not be consistent. Thus, as can be seen from FIGS. **1** and **1***a*, harvested energy can be used to power the sensor node or charge the on-hoard power source (such as batteries) on each sensor node. Thus, the demands on recharge cycles of the batteries are relaxed and the battery lifetime may be improved.

FIG. **1***a* also illustrates the interrupt signal. If the harvested RF energy is not enough and falls below a certain threshold, the DC/DC converter (for example, a MAX1674/MAX1675) may trigger the sensor node via an interrupt signal as shown. The processor on the sensor node, upon receiving this interrupt, may measure the voltages of the super capacitor, V**1** **122**, and that of the battery, V**2** **124**. Depending upon what these measured values are, the sensor node may make intelligent decisions. In one embodiment, the sensor node controls a power switch **126** that is operable to switch between the super capacitor and the batteries as a source of power. An exemplary

flow chart of an interrupt handling process that may be implemented by software on the sensor node is illustrated by FIG. **7**.

FIG. **2** illustrates the circuit exemplary passive RF energy harvesting system **200**. The system **200** may include a crystal radio receiver **202**, an energy harvesting antenna **204**, a DC/DC converter **206**, a voltage output **208**, and an interrupt output **210**. The system may include additional, fewer, or alternate components.

As shown in FIG. **2**, the crystal radio receiver **202** may include a capacitor bank and a rectifier that converts the RF energy being received into a DC signal. The DC signal harvested may be stored upon the super capacitor C**9**. The DC/DC converter **206** may sense the voltage on the super capacitor C**9** and provide an output voltage VCC **210** to the sensor node, as well as the sensor interrupt **210** when the output voltage approximately reaches or falls below a threshold level. The sensor nodes may use RF energy harvesting systems having additional, fewer, or alternate components.

FIG. **2***a* illustrates another embodiment of the circuit of the exemplary passive RF energy harvesting system **200**. With respect to FIG. **1***a*, FIG. **2***a* illustrates in more detail from where V**1** **212** and V**2** **218** may be sensed. V**1** **212** may be sensed from the super capacitor C**2**. V**2** **218** may be sensed from the battery. FIG. **2***a* also shows the power switch **214**. The system may include additional, fewer, or alternate components.

VI. Exemplary Network

FIG. **3** illustrates an exemplary network **300** of wireless radios or sensor nodes **312**. The network **300** may utilize a dynamic routing algorithm that permits data transmitted to travel the shortest distance or link **314** between sensor nodes **312** to a destination, which decreases the required transmission time for a given message, as well as the required power level of that transmission. The destination may be another sensor node **312** or a control radio **316**. Each wireless and self-contained sensor node **312** and control radio **316** may have a dedicated processor, a receiver, and a transmitter. The network **300** may include additional, fewer, or alternate components.

Each sensor node **312** may communicate its associated routing information to every nearby or adjacent sensor node **312** or control radio **316**. After a sensor node **312** receives a data transmission, a processor of the sensor node **312** may determine what to do with that data, including whether to retransmit the data to an adjacent or nearby sensor node **312** or control radio **316**. The base station **316** may function as a network controller that directs the overall operation of the network **300**.

The network **300** may provide continuous communication with otherwise unavailable sensor nodes **312**. For instance, some sensor nodes **312** may become obstructed by obstacles, such as equipment, containers, furniture, or other items, or may fail. However, the network **300** may reconfigure itself around blocked paths by redirecting transmissions from one sensor node to the next until communication with a lost sensor node is re-established. The network **300** also may provide enhanced communication reliability between sensor nodes **312** as a single sensor node **312** may be in direct communication with a number of other sensor nodes **312**, as shown in FIG. **3**.

The network **300** may implement IEEE 802.15.4 protocols. Other protocol standards may be used. The network **300** may operate as a mesh network, as described in more detail below. Alternate control or routing algorithms/instructions may be used.

VII. Control of Building Equipment

One application of the sensor nodes described herein relates to a network that controls building equipment. The network may include multiple sensor nodes, each having a wireless radio, as well as one or more control radios or base stations that direct the network. Each sensor node may be a so-called "smart" sensor node that includes a receiver, a transmitter, a processor, memory, one or more sensors, an actuator, and/or other components. Each sensor node may transmit messages to a control radio acting as network controller. Alternatively, the network controller may be a dedicated processor. The network may have one or more network controllers and/or control radios. The term network herein may include the entire network, a sub-set of a network, a number of sensor nodes, one or more network controllers, one or more control radios, or a combination of sensor nodes with one or more network controllers or control radios.

A network controller may assimilate and analyze a number of messages received from a plurality of sensor nodes. In response to each of the messages received, the network controller may determine that a change in the currently operating building equipment, or the operating modes thereof, is in order. Subsequently, the network controller may transmit a message to one or more sensor nodes that direct the operation of building equipment. Upon receiving the message, a sensor node may alter the operation of building equipment.

The sensors associated with the sensor nodes/wireless radios may monitor specific parameters pertaining to building environmental conditions or specific operating equipment. The actuators associated with the sensor nodes may control the operation of certain building equipment. A sensor node may transmit the value of a parameter sensed by a sensor to the network. In response to the values of the parameters received, the network or another sensor node may automatically alter the operation of building equipment, such as by sending messages that operate the actuators that control the building equipment.

For example, the sensors may be temperature sensors that sense the temperature in an area of a building. Each temperature sensor may be connected with a sensor node having a wireless radio, the sensor nodes being dispersed throughout a building. Each sensor node having a temperature sensor may transmit a message to the network regarding the temperature sensed in the building area in which the sensor node is located. In response to the temperature information received, the network may direct that cooling, heating, ventilation, HVAC, emergency, or other building equipment be operated to alter the building environment of the building area in which the sensor node is located.

The network may employ multiple sensor nodes in each building area to monitor temperature. Conventional wall mourned temperature sensors and/or thermostats may be single point sources of information. However, the average value of individual temperature parameters received from a plurality of temperature sensors dispersed in a given building area may better reflect the actual temperature in the building area. Accordingly, the building environmental equipment may be directed to maintain the temperature of a building area closer to the desired temperature based upon the more accurate temperature information received.

The sensors also may be motion sensors that sense motion in a building area. Each motion sensor may be connected with a sensor node having a wireless radio, the sensor nodes being dispersed throughout a building. Each sensor node having a motion sensor may transmit a message to the network regarding the motion sensed in a building area. In response to the motion information received, the network may direct the operation of building equipment.

The motion detected may alert the network that a building area has recently become occupied or unoccupied. In response, the network may ensure that lighting equipment provides adequate light in or near the building area in which motion was sensed. The network may direct that building environmental equipment, such as cooling, heating, ventilation, HVAC, or other equipment, be operated to alter the building environment of the building area. The motion information received also may be used by the network to determine that a security breach has occurred. Accordingly, the network may trigger an alarm, secure passageways, and operate other security equipment in response to the security breach.

A sensor node having a wireless radio may be connected with an identification device located on an individual. After the sensor node located on the identification device transmits a message to the network, the network may determine the identification and/or location of the associated individual. In response, the network may transmit instructions to building environmental equipment to automatically alter the environmental conditions of the specific building area in which the individual is currently located based upon stored or transmitted environmental preferences associated with that individual.

In a building having numerous pieces of operating equipment, it may be desirable to automatically monitor various parameters associated with various pieces of equipment. For instance, in a power plant, refinery, factory, or other plant, it may be advantageous to monitor temperatures, pressures, alarms, tank levels, bilge levels, hydraulic levels, atmospheric conditions, operating pumps or fans, and other parameters. The change in various temperatures, pressures, levels, or equipment operating temperatures may indicate problematic conditions.

The network may automatically identify problematic conditions associated with operating building equipment. The various parameters monitored each may be sensed by a sensor located on a sensor node. The sensor node having a wireless radio may transmit the value of the parameter to the network, either periodically or upon being queried by the network or sensing an out of specification value. The sensor node may determine whether a parameter is within specification, i.e., a predetermined satisfactory range.

If a parameter is not within specification, the network may take corrective action to restore the parameter and/or building conditions to specification. For example, the running speed of a problematic piece of equipment may be shifted, increased, or decreased. The problematic piece of equipment also may be secured and an alternate piece of equipment may be started or placed on line to replace it. Additional, fewer, or alternate courses of action may be taken to correct problematic or out of specification parameters. Sensor nodes may be connected with or part of the equipment being controlled for wireless control of the equipment.

VIII. Mesh Network

In one embodiment, the network may include a number of sensor nodes, each having a wireless radio, arranged as a mesh network that also may be used to locate movable assets and/or operate building environmental equipment. The mesh network provides the capability of routing data and instructions between and among the network of radios. The mesh network permits data to be efficiently transmitted from one radio in the network to the next until the data reaches a desired destination.

The mesh network may be implemented over a wireless network or partially wireless network. Each sensor node having a radio within the network may function as a repeater that transmits data received from adjacent sensor nodes having radios to other nearby sensor nodes that are within range. The coverage area of the mesh network may be increased by adding additional radios. As a result, a network may be established that may cover an area of desired size, such as a floor of a building or an entire building.

Each sensor node having a radio within the mesh network is typically only required to transmit data as far as the next sensor node having a radio within the network. Hence, if a sensor node having a wireless radio has a limited power supply, the reduction in the distance that each sensor node is required to transmit permits lower power level transmissions, which may extend the operating life of the on-board power supply.

A number of protocols may be used to implement the mesh network. The sensor nodes may implement a protocol that uses low data rates and low power consumption. As noted above, the mesh network may employ devices that use very small amounts of power to facilitate significantly increased battery or power supply life. In some situations, power supply life may be extended by minimizing the time that the radio device of the sensor node is "awake" or in normal power using mode, as well as reducing the power at which a signal is transmitted.

Alternatively, the sensor nodes having wireless radios may implement a protocol that uses moderate or high data rates and power consumption. For instance, the sensor nodes may implement IEEE 802.11 protocols. An IEEE 802.11 LAN may be based on a cellular architecture where the system is subdivided into cells, where each cell is controlled by a base station. Other protocols may be implemented.

Additionally, by reducing the distance between sensor nodes, each sensor node having a wireless radio may be able to transmit signals at a reduced power level, which may extend the life of a power supply while the signals transmitted remain strong enough to reach an adjacent sensor node having a radio. The sensor nodes within the network may be synchronized such that each sensor node talks or listens at a particular time. Alternatively, one or more control radios may be generally active, while the remaining radios remain predominantly passive. The control radios may be hardwired directly to a power supply such that they are not confined by a limited power supply.

The mesh network may utilize the Zigbee protocol or other IEEE 802.15.4 Low-Rate Wireless Personal Area Network (WPAN) standards for wireless personal area networking. Zigbee is a published specification set of high level communication protocols designed for use with small, low power digital radios based upon the IEEE 802.15.4 standard. Other IEEE 802.15 standards also may be implemented, including those using Bluetooth or other WPAN or WLAN protocols or any other protocol.

The mesh network of sensor nodes may employ a dynamic routing algorithm. As a result, the mesh network may be self configuring and self mending. Each sensor node within the network may be able to identify neighboring sensor nodes. After receiving a message, a receiving sensor node may determine that it is not the sensor node closest to the destination and/or that it should not relay the message to another sensor node based upon the currently known configuration of operating sensor nodes. The receiving sensor node may wait a predetermined period and listen for another sensor node with a radio to relay the message. If after a predetermined time, the sensor node determines that the message has not been relayed as expected, the receiving sensor node may transmit or relay the message to a nearby sensor node.

By transmitting messages to only reach nearby or adjacent sensor nodes in the network, the messages within the network

may be transmitted at lower power. The low power transmission requires less energy from the on-board power supply of each sensor node. Additionally, the low power transmissions by the sensor nodes prevent one message from occupying the entire network and permits messages to be simultaneously transmitted from different sensor nodes and travel throughout the network of sensor nodes in parallel.

IX. Exemplary Sensor Nodes

FIG. **4** illustrates an exemplary sensor node or wireless radio **410**. The sensor node **410** includes a processor **412**, a wireless radio frequency transmitter and/or receiver **414**, a sensor **416**, an actuator **418**, a memory **420**, a clock **422**, a speaker **424**, a microphone **426**, and an on-board power supply **428**. The sensor node **410** may include additional, different, or fewer components.

For example, the sensor node **410** may be free of the sensor **416**, actuator **418**, memory **420**, clock **422**, speaker **424**, the microphone **426**, and/or power supply **428**. The sensor node **410** may include the processor **412** and the wireless transmitter and/or receiver **414**. For instance, the sensor node **410** may comprise the processor **412**, the wireless radio frequency transmitter and/or receiver **414**, the sensor **416**, the actuator **418**, and the power supply **428**. Or the sensor node **410** may comprise the processor **412**, the wireless radio frequency transmitter and/or receiver **414**, the sensor **416**, and the power supply **428**. The sensor node **410** may include other combinations employing additional, different, or fewer components.

In one embodiment, the sensor node **410** may be operable to automatically control building equipment and locate movable items within a building. In another embodiment, the sensor node **410** may be portable, such as in the case of being mounted upon a movable item, or affixed at a specific location or to an immovable item. The sensor node **410** may be a controller, actuator, sensor, locator or other device in a security, fire, environment control, HVAC, lighting, or other building automation system. The sensor node **410** may determine its present location, sense conditions within a building, report conditions within a building, generate a signal representative of a building condition, and/or respond to an interrogator. The sensor node **410** also or alternatively may actuate building control components. As a controller, the sensor node **410** may be free of the sensor **416** and/or the actuator **418**. In another embodiment, the sensor node **410** is a wireless device free of wired connections to other devices making the sensor node **410** self-contained and/or portable.

The sensor **416** may be a single sensor or include multiple sensors. The sensor **416** may be a temperature, pressure, humidity, fire, smoke, occupancy, air quality, flow, velocity, vibration, rotation, enthalpy, power, voltage, current, light, gas, CO_2, CO, N_2, O_2, chemical, radiation, fluid level, tank level, motion, Global Positioning System (GPS), infrared, or other sensor or combination thereof. The sensor **416** also may be a limit or proximity switch. Alternate sensors may be used.

The sensor **416** may be a motion sensor that detects when a portable sensor node **410** is moving. If it is sensed that the sensor node **410** is moving, the processor **412** may wake the sensor node **410** up from a sleep mode that draws less energy from the power supply **428**. Upon waking up, the sensor node **410** may transmit via the wireless transmitter **414** to the network a message indicating that sensor node **410** is moving.

The sensor **416** may be motion sensor that detects when there is movement within a predetermined distance. For example, the sensor **416** may be wall mounted to detect when an individual has entered a specific building area. If the building area was previously unoccupied, the sensor node **410** on which the sensor **416** is mounted may transmit a message to the network that the building area is no longer unoccupied. As a result, the network may direct that the environmental conditions be altered accordingly, such as increase the temperature during cold weather, decrease the temperature during hot weather, turn on one or additional lights, or adjust the room to the individual's personal preferences.

The sensor **416** may be a GPS unit capable of receiving GPS signals and determining the location of the sensor node **410**. The GPS unit may be able to determine the latitudinal and longitudinal coordinates, as well as the elevation, of the sensor node **410**. The location of the sensor node **410** determined by the GPS unit may be subsequently transmitted to the network via the wireless transmitter **414**.

The actuator **418** may be a single actuator or include multiple actuators. The actuator **418** may be a valve, relay, solenoid, speaker, bell, switch, motor, motor starter, turbine generator, motor generator, diesel generator, pneumatic device, damper, or pump actuating device or combinations thereof. For example, the actuator **412** may be a valve for controlling flow of fluid, gas, or steam in a pipe, or a damper controlling or redirecting air within an air duct. As another example, the actuator **412** may be a relay or other electrical control for opening and closing doors, releasing locks, actuating lights, or starting, stopping, and shifting motors and pumps. As a further example, the actuator **412** may be a solenoid that opens or closes valves, dampers, or doors, such as for altering the flow of fluid or air within piping or ducting. Alternate actuating devices also may be used.

The sensor node **410** may function as a controller or base station. The controller may be positioned at either a known or an unknown location. As a controller, the sensor node **410** interacts with other sensor nodes **410** for control or reporting functions.

The processor **412** is capable of processing data and/or controlling operation of the sensor node **410**. The processor **412** may be a general processor, digital signal processor, application-specific integrated circuit (ASIC), field programmable gate array, analog circuit, digital circuit, network of processors, programmable logic controller, or other processing device. The processor **412** may have an internal memory.

The sensor node **410** also may have a memory unit **420** external to the processor **412**. The memory unit **420** may store data and instructions for the operation and control of the sensor node **410**, such as discussed herein. Additional or alternate types of data also may be stored in the memory unit **420**.

A program may reside on the internal memory or the memory unit **420** and include one or more sequences of executable code or coded instructions that are executed by the processor **412**. The instructions may include monitoring a signal associated with radio frequency energy being harvested via a crystal radio receiver, the radio frequency energy being used to at least partially power a sensor node. The instructions also may include automatically reducing the amount of power being expended by the sensor node if the magnitude of the signal monitored approximately reaches or falls below a threshold level. The instructions may include directions to accomplish other functionality described herein.

The program may be loaded into the internal memory or memory unit **420** from a storage device. The processor **412** may execute one or more sequences of instructions of the program to process data. Data may be input to the data processor **412** with a data input device and/or received from a network. The program and other data may be stored on or read from machine-readable medium, including secondary storage devices such as hard disks, floppy disks, CD-ROMS, and

DVDs, or other forms of machine readable medium, either currently known or later developed.

The processor **412** is capable of directing the transmission or reception of data by the wireless transmitter or receiver **414**, the speaker **424** or the microphone **426**. For example, the processor **412** may direct the acoustic speaker **424** to transmit an ultrasound signal. The processor **412** may also direct the microphone **426** to receive an ultrasound signal and determine a distance from another device as a function of the received signal. Alternatively or additionally, the processor **412** may direct the wireless transmitter or receiver **414** to transmit data for determining the distance. Additionally or alternatively, the wireless transmitter **414** transmits a determined distance or distances as well as data regarding the processes and operation of the sensor **416** and/or the actuator **418**.

The wireless transmitter and receiver **414** or the speaker **424** may be alternate wireless transmitters capable of transmitting a signal for distance determination. Similarly, the wireless receiver **414** and microphone **426** may be alternative wireless receivers capable of transmitting a signal for distance determination.

FIG. **5** illustrates a floor layout for a network of sensor nodes having wireless radios **510** operating with one or more control radios **522** within a building **524**. The sensor nodes **510** may be dispersed throughout the building **524**. One or more of the sensor nodes **510** may be located in each room or other building area. Alternate dispersed arrangements of the sensor nodes **510** may be provided. While one control radio **522** is shown, a plurality of control radios **522** may be provided in other embodiments. Additional, different or fewer sensor nodes **510** and control radios **522** may be provided. While shown as a single floor of a building **524**, the network of sensor nodes **510** and control radios **522** may be distributed over multiple floors, a portion of the floor, a single room, a house, a structure, or any other building **524** or portion thereof.

The various sensor nodes **510** may be of the same configuration or a different configuration than each other. For example, some of the sensor nodes **510** may correspond to sensor arrangements, while other sensor nodes **510** may correspond to actuator arrangements. The same or different communication device, such as a wireless radio frequency transmitter and/or receiver, may be provided for each of the sensor nodes **510**. Alternatively, different communications mechanisms and/or protocols are provided for different groups of the sensor nodes **510**. The sensor nodes **510** may operate in an integrated manner for implementing one or multiple types of building automation control. Alternatively, different networks may be provided for different types of building automation, such as security, HVAC, heating, ventilation, and fire systems.

The sensor nodes powered by RF energy as discussed herein may be useful for a wide variety of applications, in addition to the building control functionality described above. For instance, the applications may include environmental and human body monitoring to military surveillance, building automation, and industrial monitoring. Each sensor node may facilitate one or more control, homeland security, reconnaissance, or other tasks. In one embodiment, the sensor node may be implanted internally within a patient, and sense and transmit data pertaining to a biological characteristic of the patient. The sensor nodes may be used for other applications.

X. Exemplary Method

FIG. **6** illustrates an exemplary RF energy harvesting method. The method may include harvesting RF energy **602**, storing the RF energy harvested **604**, monitoring a signal associated with the RF energy harvested **606**, and selecting an operating mode based upon the level of the signal monitored **608**. The method may include additional, fewer, or alternate actions.

The method may include harvesting RF energy **602**. RF energy may be harvested via a crystal radio receiver as discussed herein. Most of the energy harvested may be scavenged from AM radio signals. The RF energy may be harvested ubiquitously. The RF energy may be used to primarily or solely power a self-contained sensor node.

The method may include storing the RF energy harvested **604**. The RF energy or an associated signal may be measured from a super capacitor or other energy storage means. In one aspect, the RF energy harvestor may be used in unison with another type of energy harvester, such as a solar, biomechanical, thermoelectric, or vibration driven device.

The method may include monitoring a signal associated with the RF energy **606**. A voltage on the super capacitor may be monitored. Other signals related to the RF energy harvested also may be monitored.

The method may include selecting an operating mode based upon the level of the signal monitored **608**. The operating mode selected may relate to power constraints imposed upon the sensor node. If the level of the signal being monitored, such as a voltage signal or other signal, falls below a threshold level, the sensor node may be directed to operate in a low power mode, as discussed herein, to conserve energy until the amount of RF energy being harvested returns to a self-sustaining level with respect to the operating power requirements of the sensor node.

XI. Exemplary Method

FIG. **7** illustrates an exemplary interrupt handling method **700**. The method **700** may be implemented by software on the sensor node. As shown, the method **700** may use a state variable K having 2 states. When K=1, the power switch in FIG. **2***a* may connect pin A to pin Out, which in turn is connected to the pin labeled VCC. The voltage of pin VCC, which is connected to the sensor node, is equal to the output voltage of the DC/DC converter. Thus, in this case, the sensor node is powered through the energy harvester via the super capacitor and DC/DC converter. On the other hand, when K=2, the power switch connects pin B to pin Out and in this case, the sensor node is powered by the battery.

Ts in FIG. **7** is the wake-up period (which is configurable through software depending upon the application). The sensor node wakes up repetitively with time period of Ts. Each time when the sensor node wakes up, it may sense the environment, process the data, and transmit the date to the sink, if necessary. Then the sensor node may enter sleep mode, which costs almost no energy.

Tm in FIG. **7** is the maximum wake-up period allowed by the system. If Ts>Tm, the wake-up period is longer than the upper limit required by the system. Thus, the sensor node cannot provide spontaneous samples to satisfy the quality of service (QoS) requirements of the network.

Tp in FIG. **7** denotes a continuous block of time during which harvested energy is enough to support the continuous operation of the sensor node to execute for a time Te, which is the minimum time required by a sensor node to complete sensing tasks as required by an application. In other words, each time the sensor node wakes up, it may have to execute for at least a period of Te. And the sensor node may have to make a decision as to whether it powers itself from the energy harvester or through the battery for this amount of time. The

following brief analysis is presented to describe the novel sensor node operation that may be facilitated through the energy harvesting circuit.

Let the capacity of the super capacitor be represented by C. Then, the instantaneous energy E stored in the super capacitor is given by:

$$E=\tfrac{1}{2}CV_1^2.$$

If power rating of the sensor node is P, then the following equation holds

$$E+D'T_p=PT_e,$$

where E' is the derivative of E with respect to time Thus,

$$\begin{aligned}T_p &= \frac{PT_e - E}{E'} \\ &= \frac{PT_e - \frac{1}{2}CV_1^2}{CV_1V_1'},\end{aligned}$$

where V_1' is the derivative of **V1** and V_1' can be estimated in practice by taking the difference of the **V1** values at consecutive time intervals.

The flow chart shown in FIG. **7** may be described as follows. After receiving the interrupt **702**, the sensor node sends a control signal to the power switch such that the sensor node is powered through the battery **704**, since the energy on the capacitor may be depleted soon. The voltages **V1** and **V2** are measured by the sensor node **706**. Tp is also computed based on **V1** **708**, as shown in the above equation.

V2 is checked to see if it is low **710**. If it is, then all of the energy is going to be depleted soon. Recall that when the interrupt occurred, **V1** is already low. Thus, in this case, neither the energy harvester nor the battery can power the sensor node. The sensor node must take emergency actions, such as backup data **712** and notify its neighbors **714**. Finally, the sensor node sets Ts to be equal to Tp **716**, as well as sets K to 1 **718**. The rationale being that since the energy in the battery is going to be depleted soon, the node should wait in sleep mode **732** until enough energy has been harvested by the RF energy harvester.

When **V2** is not low, the decision is based on **V1**. If **V1** is low such that Tm≤Tp **720**, which means that the energy harvester cannot satisfy the sensor node energy requirements, then the sensor node once again has no choice but to use power from the battery source. Since there is enough battery energy (recall that **V2** is not low), the sensor node can be powered by the battery **728** through the current wake-up period Ts **726**. The neighbors also may be notified **730**. However, note that the energy from the harvester can still be used to charge the battery through the DC/DC converter. Thus, the system still benefits from the energy harvester even though the node is not powered directly by the energy harvester.

If **V1** is high enough to support the minimal energy requirements of the sensor during its wake-up period, i.e., Tm>Tp, then the sensor node can increase its wake-up period. In this case, Ts is set to be Tp **722** and the sensor node is powered by the energy harvester through the DC/DC converter **724**.

To summarize, the embodiments discussed herein facilitate the following complex sensor node behaviors: (1) switching the operating mode on the sensor node, such as standby or sleep modes; (2) sending a notification message to a sink via other sensor nodes; (3) saving important data to the flash memory on the sensor node, in case it is lost; (4) changing a sampling rate and transmission rate, and reducing the rates according to the voltages **V1** and **V2**, (5) adjusting an energy weighting factor in the network routing metric (if the voltages of the sensor node are not high enough, the use of this node as a router to route other wireless message packets is minimized or avoided), and (6) switching the input power source from the RF energy harvester to a battery.

The methods and systems discussed herein may use one, two, or more antennas. If two antennas are used, one antenna may be dedicated to the sensor node or communications unit, and the other antenna may be dedicated to the RF energy harvesting system, such as shown in the Figures. As both the sensor node and the energy harvester each may have their own dedicated antenna, switching a single antenna between the sensor node and the energy harvester may be alleviated.

The operational mode of the harvesting system and/or the sensor node may be switched by the sensor node itself. Additionally or alternatively, the RF energy harvester discussed herein may trigger the sensor node and provide voltage measurements. For energy conservation purposes, the harvesting system discussed herein may not switch operation modes based on. RF signals.

Therefore, the RF energy harvesters described herein may be directly connected to sensor nodes, purely passive and not needing to pass FCC requirements, and/or based on a bandpass or all pass filter for ubiquitous and wide spectrum RF energy harvesting. The systems discussed herein may facilitate advanced sensor behaviors. The system modes (active, standby and sleep) may be triggered by an interrupt signal and controlled by the sensor node, as compared to altering the operating mode according to received RF signals. The designs of the energy harvesting antenna may be different from known antennas. A super capacitor and/or a DC/DC converter may be employed by the energy harvester. Based on efficiency, size, and costs, the DC/DC converter may be provide benefits as compared to transformers. A buffered energy storage system may be constructed by the super capacitor and one or more rechargeable batteries. The harvesting system may improve battery life time.

While the invention has been described above by reference to various embodiments, it should be understood that many changes and modifications can be made without departing from the scope of the invention. The description and illustrations are by way of example only. Many more embodiments and implementations are possible within the scope of this invention and will be apparent to those of ordinary skill in the art. The various embodiments are not limited to the described environments, and have a wide variety of applications including integrated building control systems, environmental control, security detection, communications, industrial control, power distribution, and hazard reporting.

It is intended in the appended claims to cover all such changes and modifications which fall within the true spirit and scope of the invention. Therefore, the invention is not limited to the specific details, representative embodiments, and illustrated examples in this description. Accordingly, the invention is not to be restricted except in light as necessitated by the accompanying claims and their equivalents.

We claim:

1. A method of passive radio frequency energy harvesting, the method comprising:

harvesting radio frequency energy using a crystal receiver;

powering a processing unit only from the radio frequency energy being harvested when the radio frequency energy being harvested, or an indication of the amount of radio frequency energy being harvested, is above a threshold level;

generating an interrupt signal to switch the processing unit to a low power mode when the radio frequency energy

being harvested, or the indication of the amount of radio frequency energy being harvested, falls below the threshold level; and

alternatively powering the processing unit only from an on-board power source when the radio frequency energy being harvested, or the indication of the amount of radio frequency energy being harvested, is below the threshold level.

2. The method of claim **1**, wherein the crystal receiver employs an antenna that is fabricated from printed circuit board.

3. The method of claim **1**, wherein the radio frequency energy harvested is increased by a super capacitor before being used to power the processing unit.

4. The method of claim **1**, wherein the on-board power source is a battery or a capacitor.

5. The method of claim **1**, wherein the crystal receiver employs a ferrite coil antenna.

6. The method of claim **1**, wherein the radio frequency energy is harvested ubiquitously.

7. A method of passive radio frequency energy harvesting, the method comprising:

harvesting radio frequency energy using a crystal receiver;

powering a processing unit only from the radio frequency energy being harvested when the radio frequency energy being harvested, or an indication of the amount of radio frequency energy being harvested, is above a threshold level, wherein the processor unit is a self-contained sensor node operable to direct the operation of building equipment to control an environment of a building;

generating an interrupt signal to switch the processing unit to a low power mode when the radio frequency energy being harvested, or the indication of the amount of radio frequency energy being harvested, falls below the threshold level; and

alternatively powering the processing unit only from an on-board power source when the radio frequency energy being harvested, or the indication of the amount of radio frequency energy being harvested, is below the threshold level.

8. The method of claim **7**, wherein the processor unit has power requirements of greater than 10 milliwatts.

* * * * *

US 20150365013A1

(19) **United States**
(12) **Patent Application Publication**
Hameed et al.
(10) **Pub. No.: US 2015/0365013 A1**
(43) **Pub. Date: Dec. 17, 2015**

(54) **RF-DC POWER CONVERTER**

(71) Applicant: **THE GOVERNORS OF THE UNIVERSITY OF ALBERTA**, Edmonton (CA)

(72) Inventors: **Zohaib Hameed**, Edmonton (CA); **Kambiz Moez**, Edmonton (CA)

(21) Appl. No.: **14/740,021**

(22) Filed: **Jun. 15, 2015**

Related U.S. Application Data

(60) Provisional application No. 62/011,957, filed on Jun. 13, 2014.

Publication Classification

(51) **Int. Cl.**
H02M 7/217 (2006.01)

(52) **U.S. Cl.**
CPC *H02M 7/217* (2013.01)

(57) **ABSTRACT**

A rectifier comprising a chain of transistors for RF-DC conversion. In order to compensate for the thresholds of the transistors, each transistor can be connected to a junction earlier or later in the chain. By using both p-type and n-type transistors in the same chain, the different types of transistors can be compensated in different directions allowing more transistors to be compensated. Additional transistors connected to the gates of transistors of the main chain can allow the transistors of the main chain to be forward compensated at one part of the input cycle and backward compensated in another part to minimize both the voltage threshold of the rectifier and the leakage current. The line for compensation of the voltage threshold during forward conduction can comprise a solid line or a transistor, and if a transistor is used it may be diode-connected.

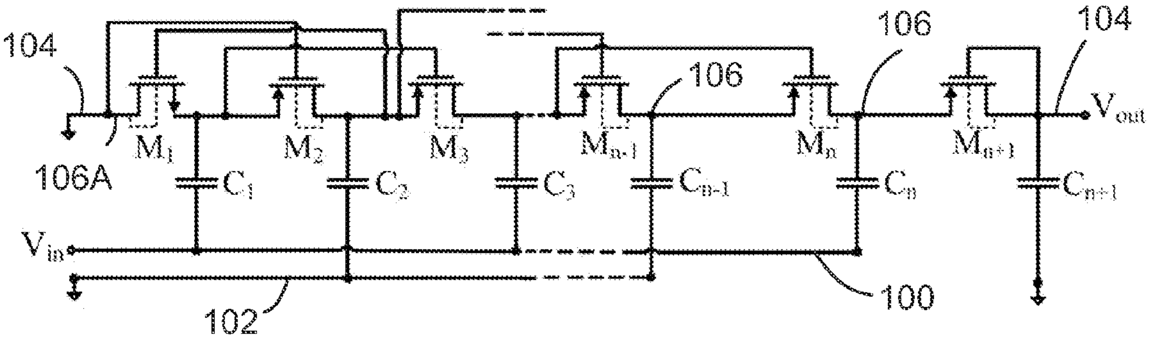

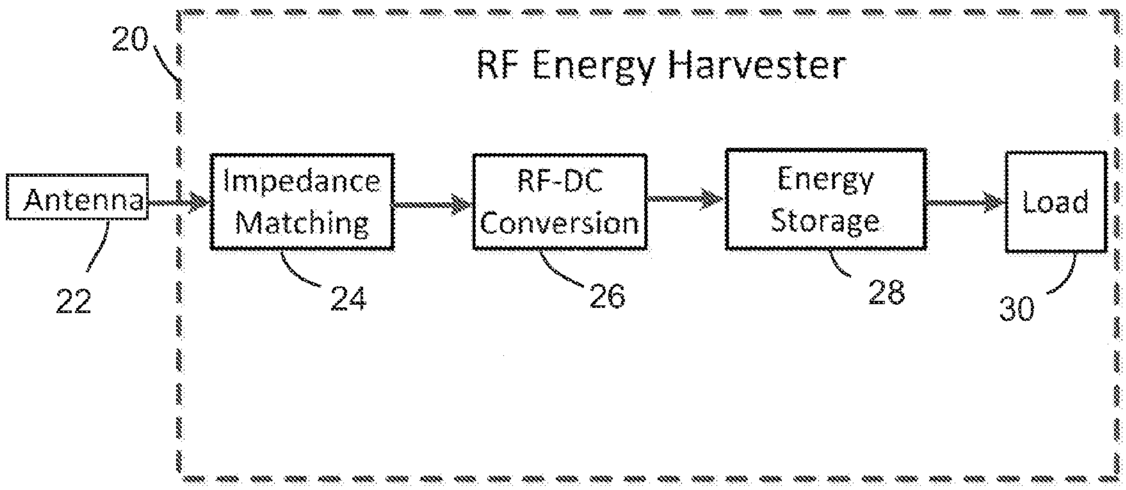
20
RF Energy Harvester
Antenna
Impedance Matching
RF-DC Conversion
Energy Storage
Load
22
24
26
28
30

Fig. 1 (Prior Art)

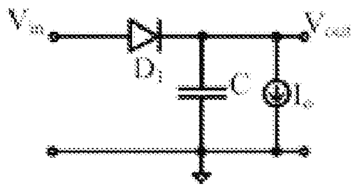
Vin
D1
C
Io
Vout

Fig. 2A (Prior Art)

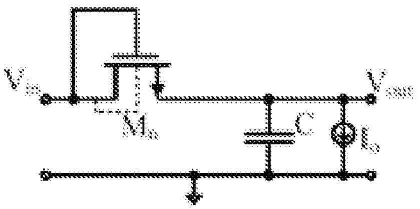
Vin
Mn
C
Io
Vout

Fig. 2B (Prior Art)

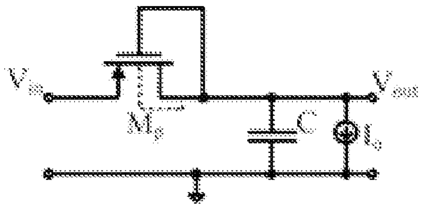
Vin
Mp
C
Io
Vout

Fig. 2C (Prior Art)

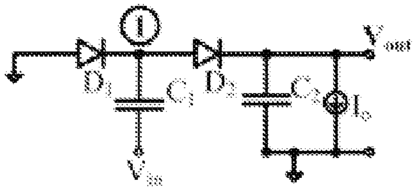
D1
C1
Vin
D2
C2
Io
Vout

Fig. 2D (Prior Art)

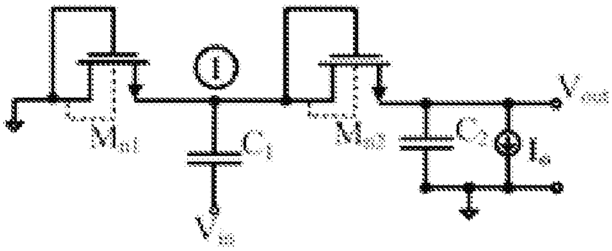
I
V_{out}
M_{n1}
C_1
M_{n2}
C_2
I_o
V_{in}

Fig. 2E (Prior Art)

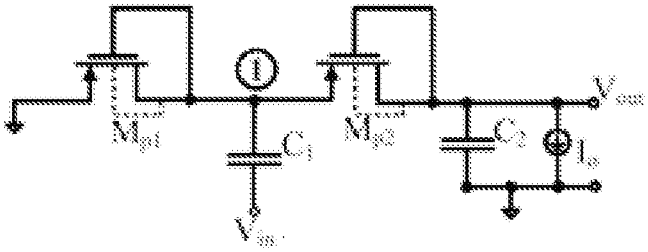
I
V_{out}
M_{p1}
C_1
M_{p2}
C_2
I_o
V_{in}

Fig. 2F (Prior Art)

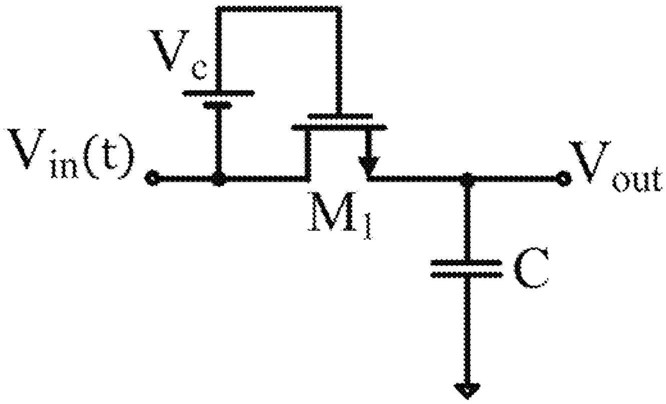
V_c
$V_{in}(t)$
V_{out}
M_1
C

Fig. 2G (Prior Art)

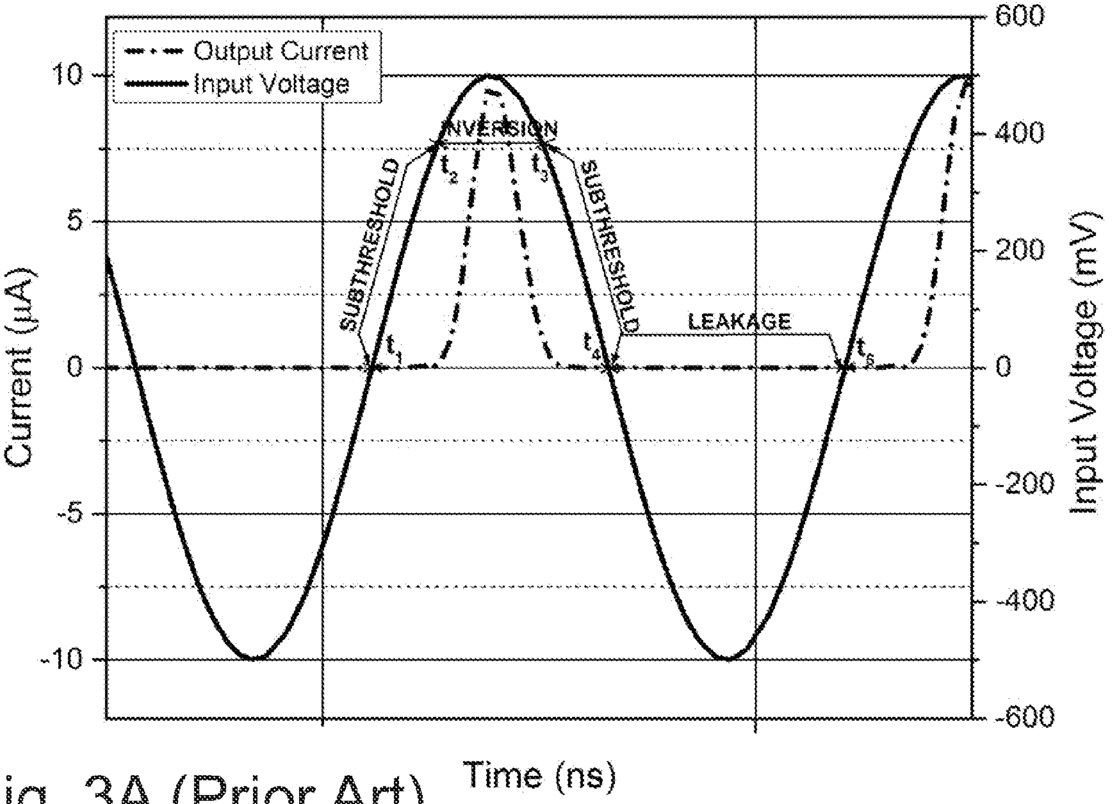

Fig. 3A (Prior Art)

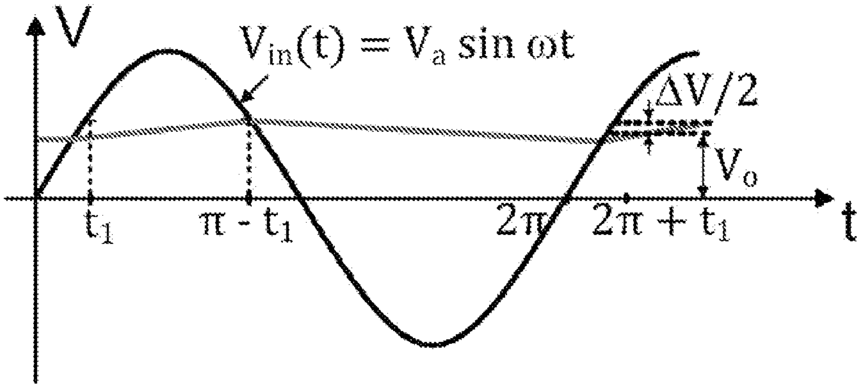
V
$V_{in}(t) = V_a \sin \omega t$
$\Delta V/2$
V_o
t_1
$\pi - t_1$
2π
$2\pi + t_1$
t

Fig. 3B (Prior Art)

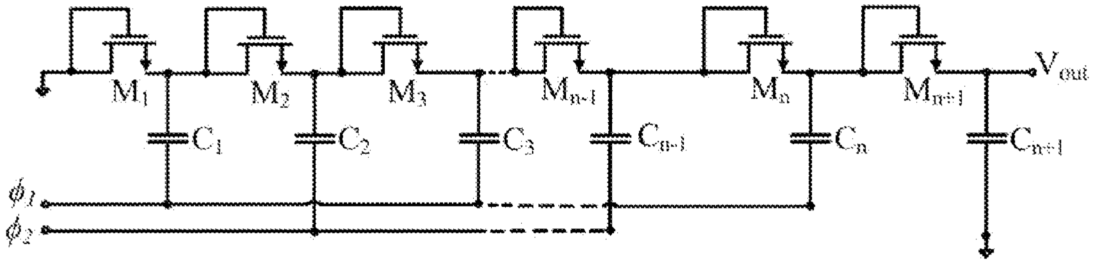
M_1
M_2
M_3
M_{n-1}
M_n
M_{n+1}
C_1
C_2
C_3
C_{n-1}
C_n
C_{n+1}
V_{out}
ϕ_1
ϕ_2

Fig. 4A (Prior Art)

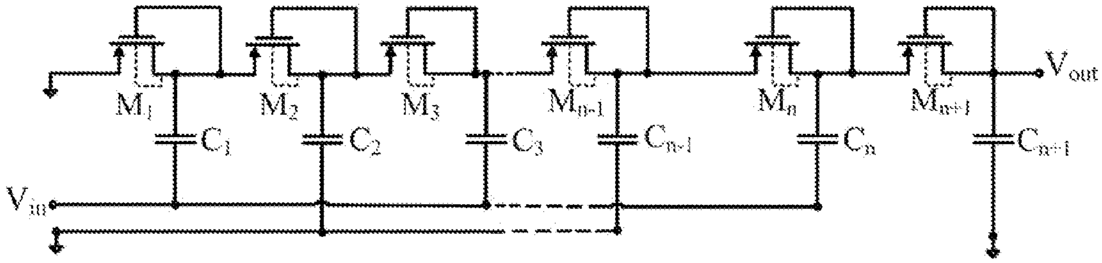
M_1
M_2
M_3
M_{n-1}
M_n
M_{n+1}
C_1
C_2
C_3
C_{n-1}
C_n
C_{n+1}
V_{out}
V_{in}

Fig. 4B (Prior Art)

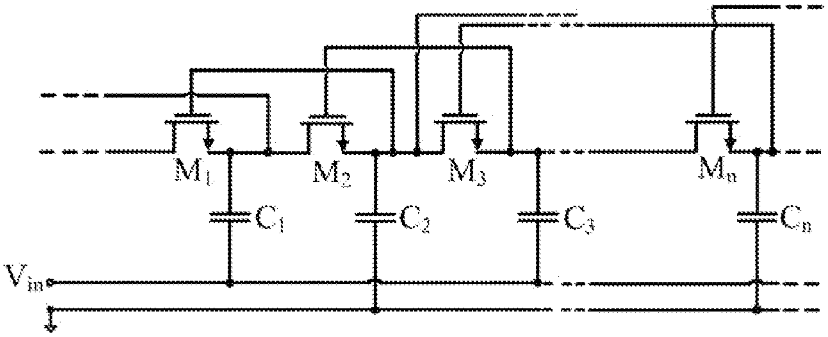
M1
M2
M3
Mn
C1
C2
C3
Cn
Vin

Fig. 5 (Prior Art)

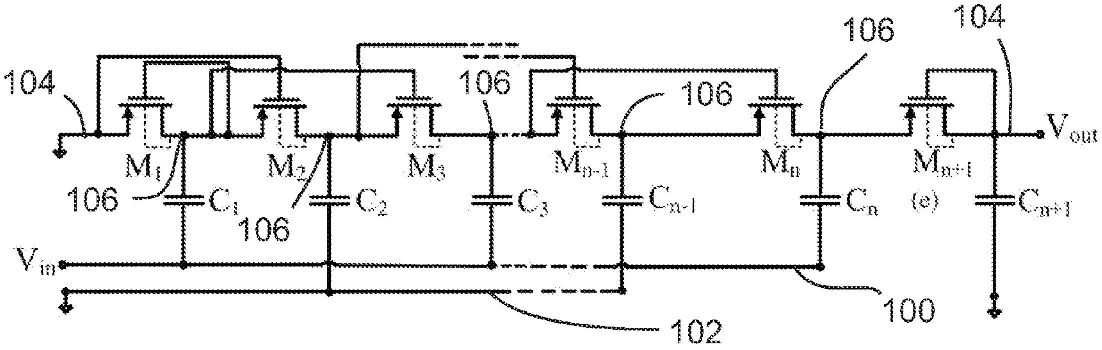
104
106
106
106
106
104
Vout
M1
M2
M3
Mn-1
Mn
Mn+1
106
C1
106
C2
C3
Cn-1
Cn
(e)
Cn+1
Vin
100
102

Fig. 6A

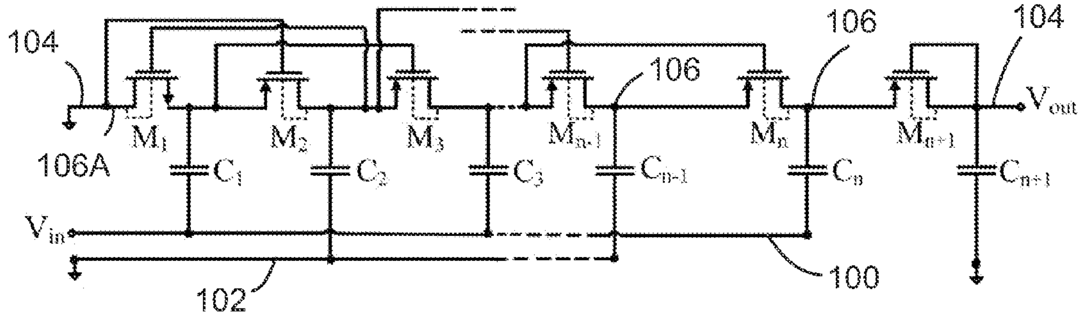
104
106A
M1
M2
M3
Mn-1
Mn
Mn+1
106
106
104
Vout
C1
C2
C3
Cn-1
Cn
Cn+1
Vin
102
100

Fig. 6B

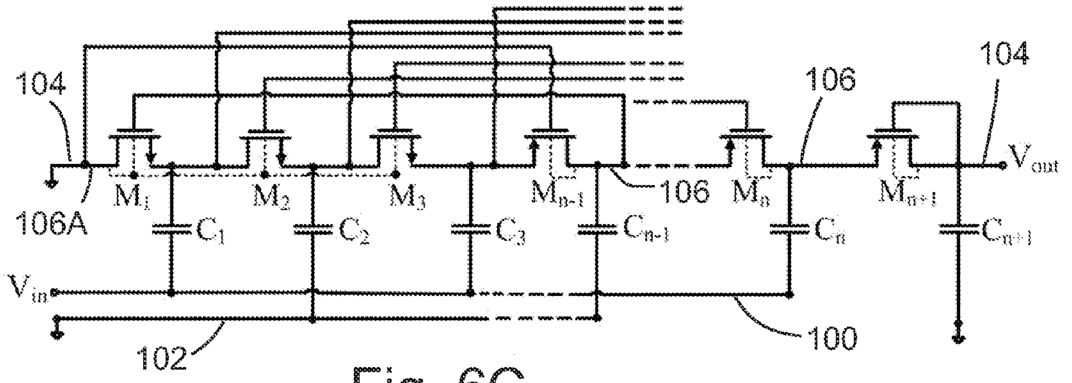
104
106A
M1
M2
M3
Mn-1
106
Mn
Mn+1
106
104
Vout
C1
C2
C3
Cn-1
Cn
Cn+1
Vin
102
100

Fig. 6C

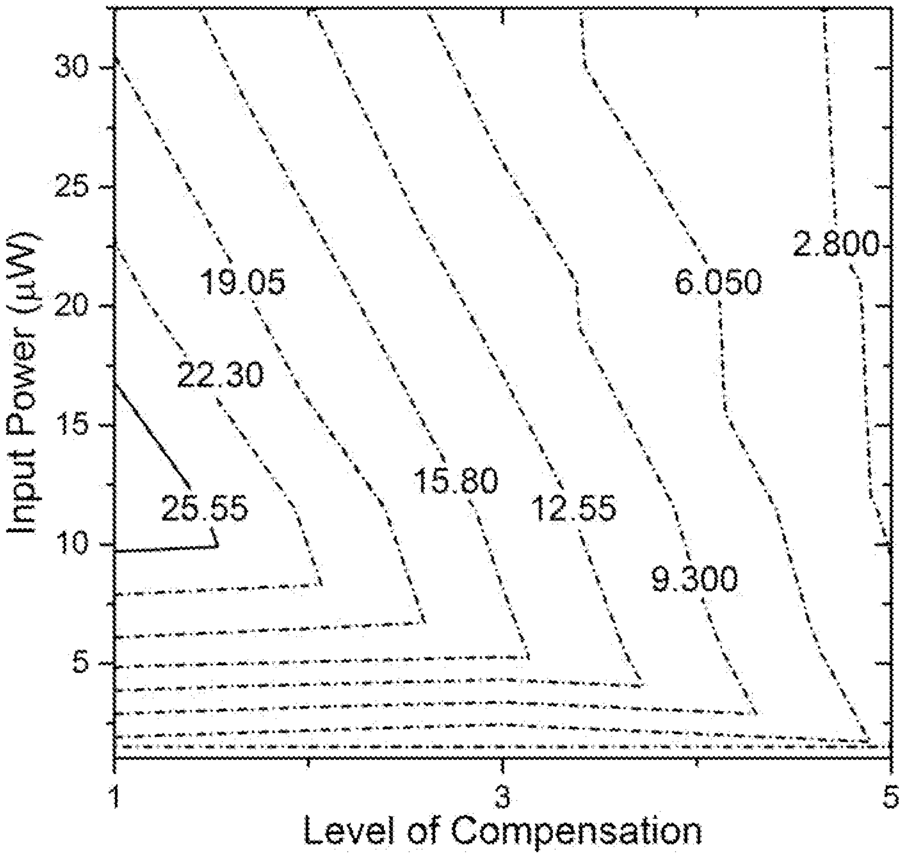
Input Power (μW)
30
25
20
15
10
5
1
3
5
Level of Compensation
19.05
22.30
25.55
15.80
12.55
9.300
6.050
2.800

Fig. 7

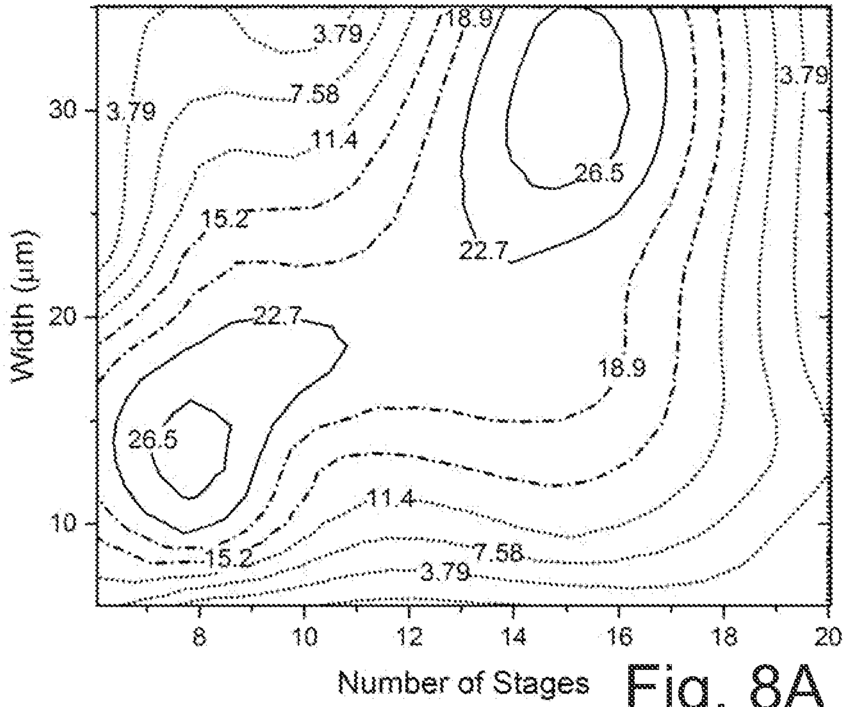
3.79
18.9
3.79
3.79
7.58
11.4
26.5
15.2
22.7
22.7
18.9
26.5
11.4
15.2
7.58
3.79
Width (μm)
30
20
10
8
10
12
14
16
18
20
Number of Stages

Fig. 8A

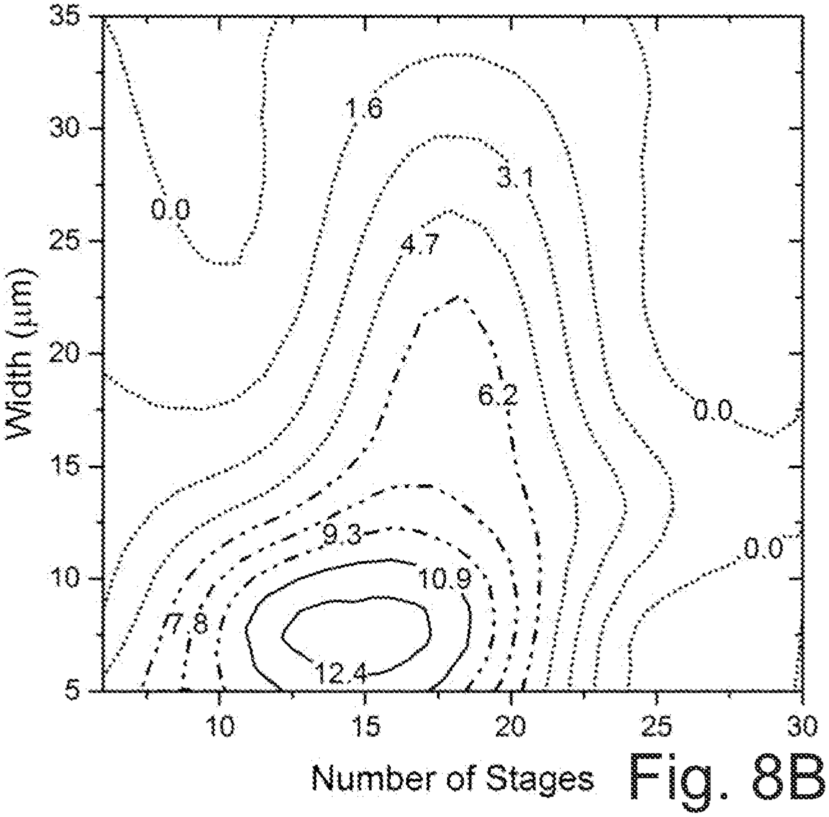
1.6
3.1
0.0
4.7
6.2
0.0
9.3
0.0
10.9
7.8
12.4
Width (μm)
35
30
25
20
15
10
5
10
15
20
25
30
Number of Stages

Fig. 8B

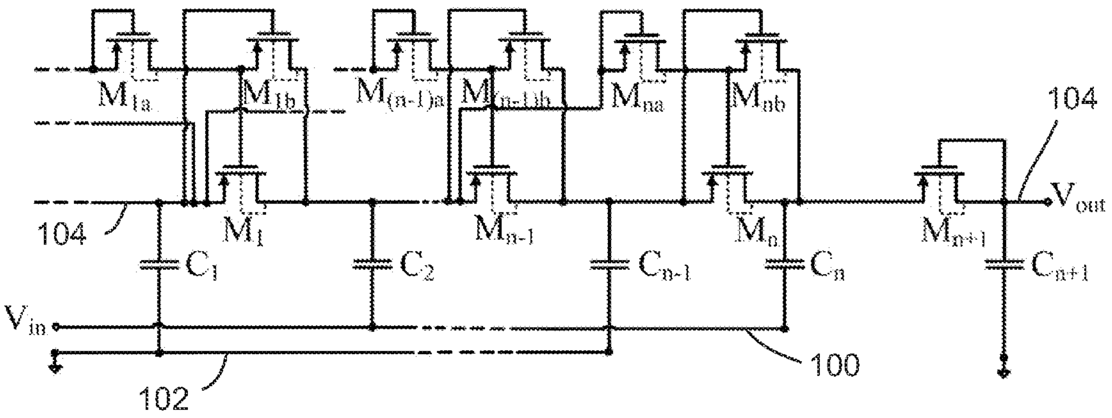
M_{1a}
M_{1b}
$M_{(n-1)a}$
$M_{(n-1)b}$
M_{na}
M_{nb}
104
V_{out}
104
M_1
M_{n-1}
M_n
M_{n+1}
C_1
C_2
C_{n-1}
C_n
C_{n+1}
V_{in}
100
102

Fig. 9A

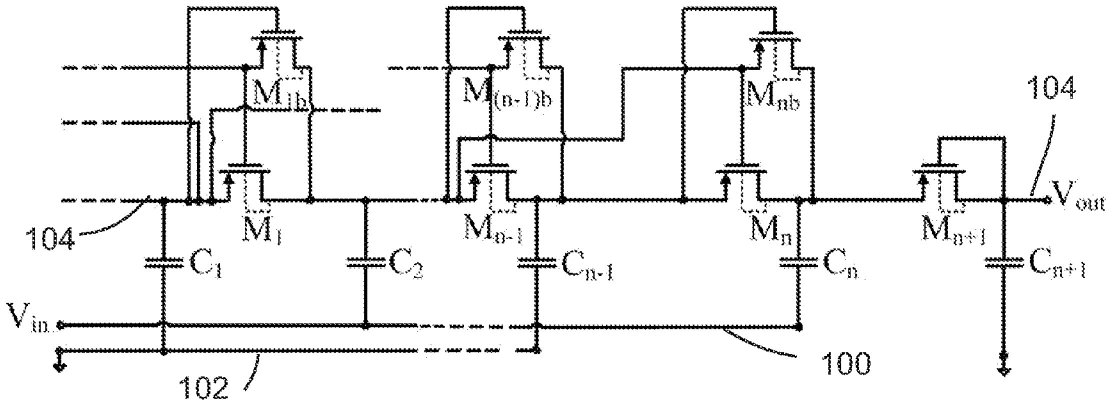
M_{1b}
$M_{(n-1)b}$
M_{nb}
104
V_{out}
104
M_1
M_{n-1}
M_n
M_{n+1}
C_1
C_2
C_{n-1}
C_n
C_{n+1}
V_{in}
100
102

Fig. 9B

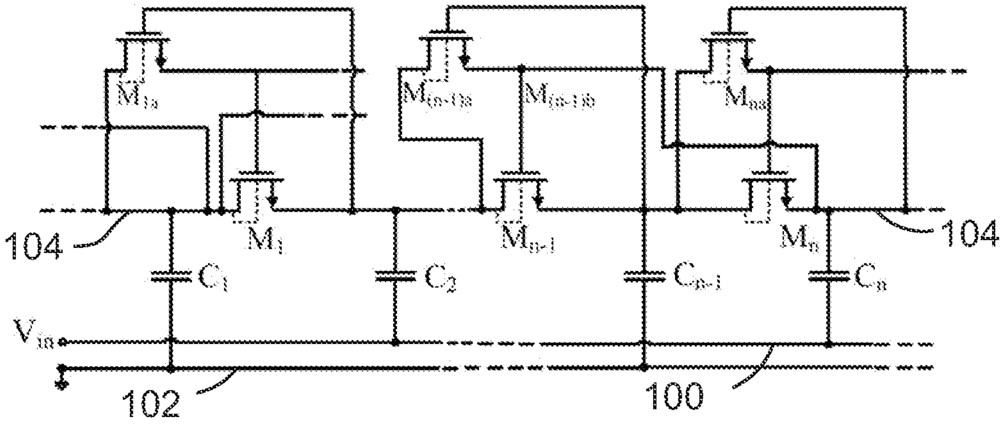
M_{1a}
$M_{(n-1)a}$
$M_{(n-1)b}$
M_{na}
M_1
M_{n-1}
M_n
104
104
C_1
C_2
C_{n-1}
C_n
V_{in}
102
100

Fig. 9C

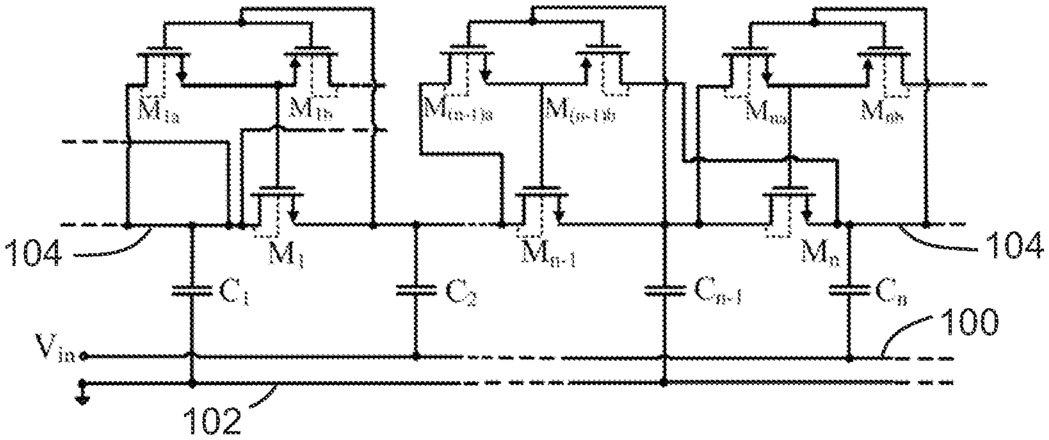
M_{1a}
M_{1b}
$M_{(n-1)a}$
$M_{(n-1)b}$
M_{na}
M_{nb}
M_1
M_{n-1}
M_n
104
104
C_1
C_2
C_{n-1}
C_n
100
V_{in}
102

Fig. 9D

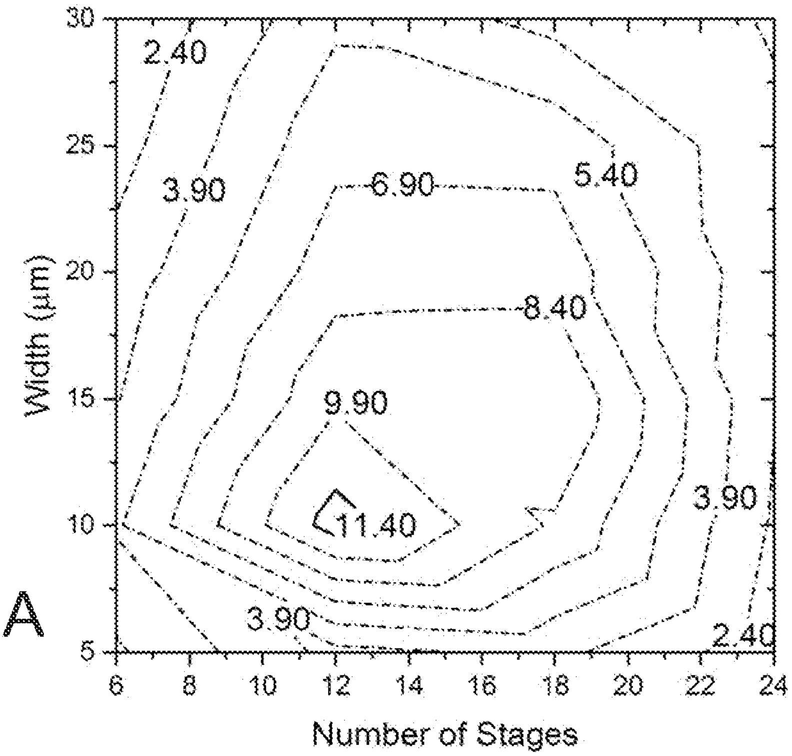

Fig. 10A

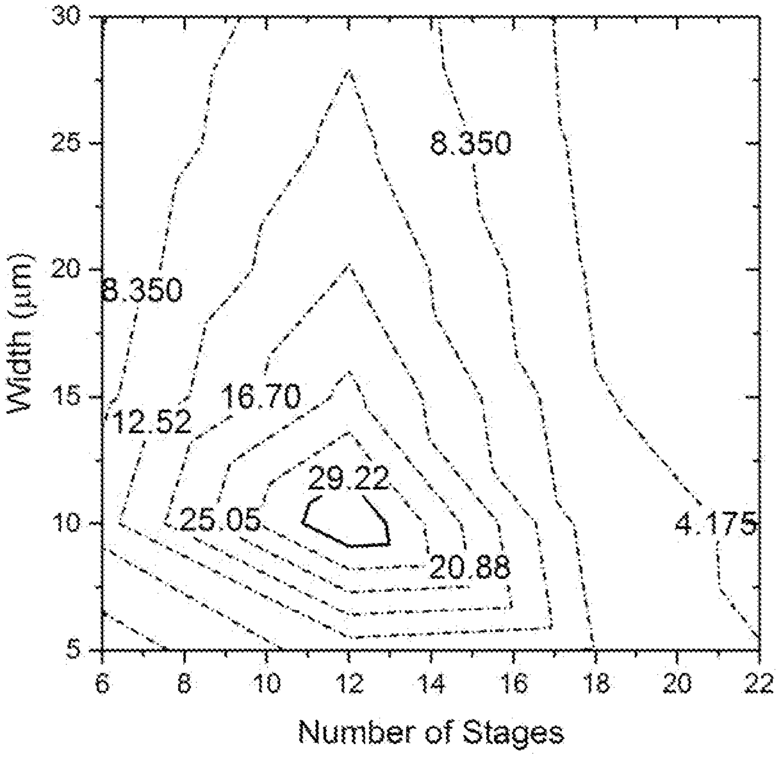

Fig. 10B

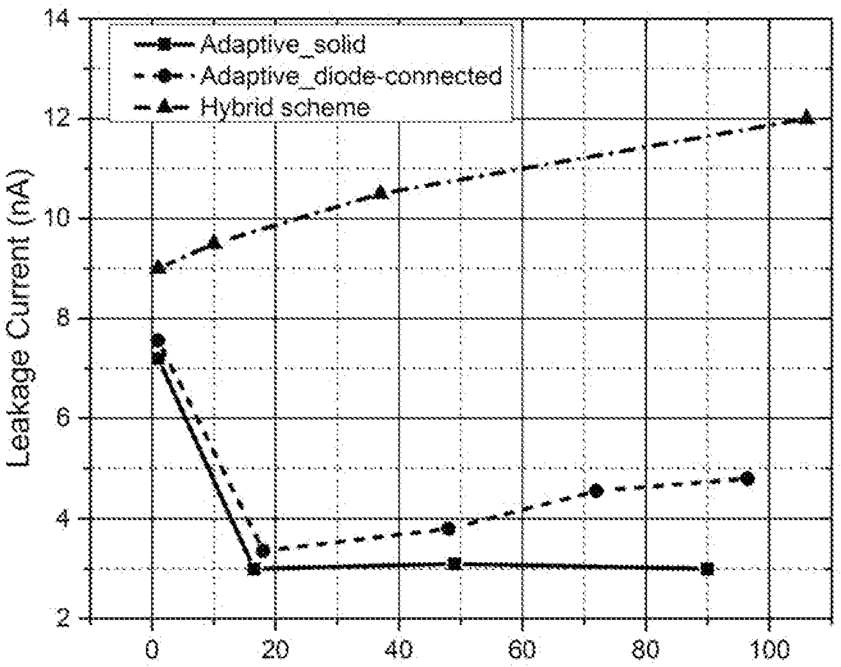

Fig. 11A

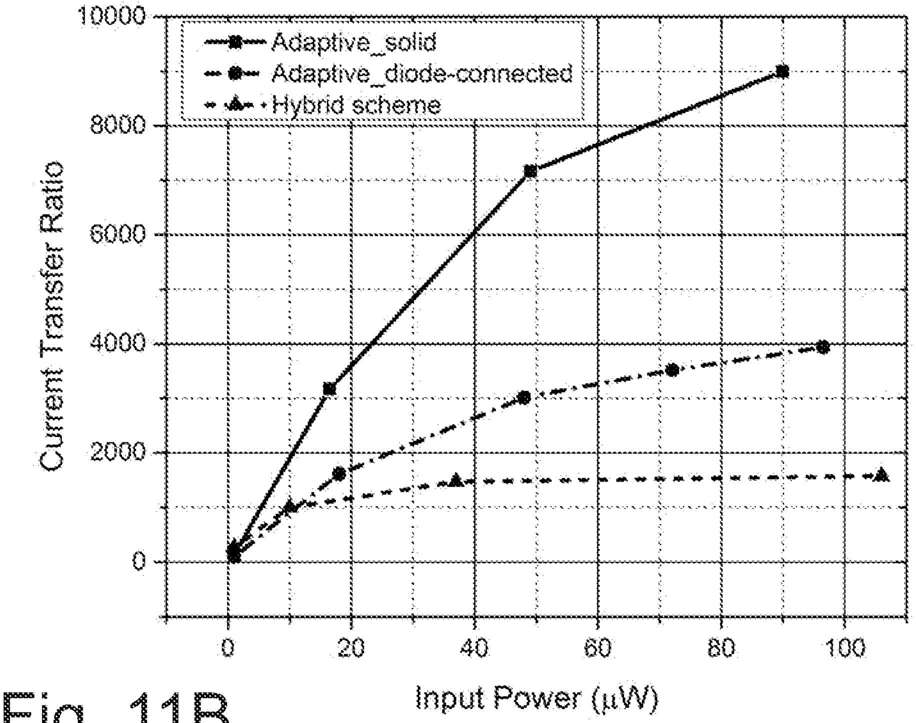

Fig. 11B

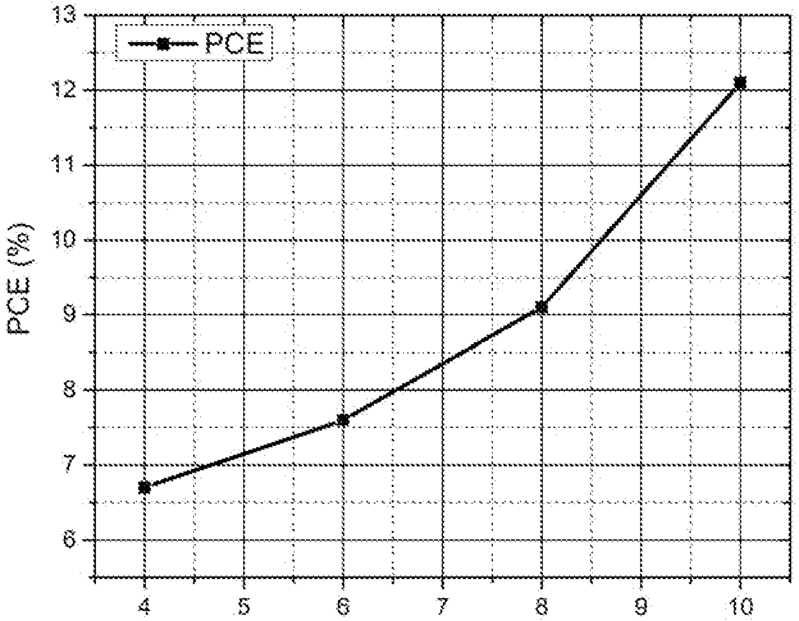

Fig. 12A

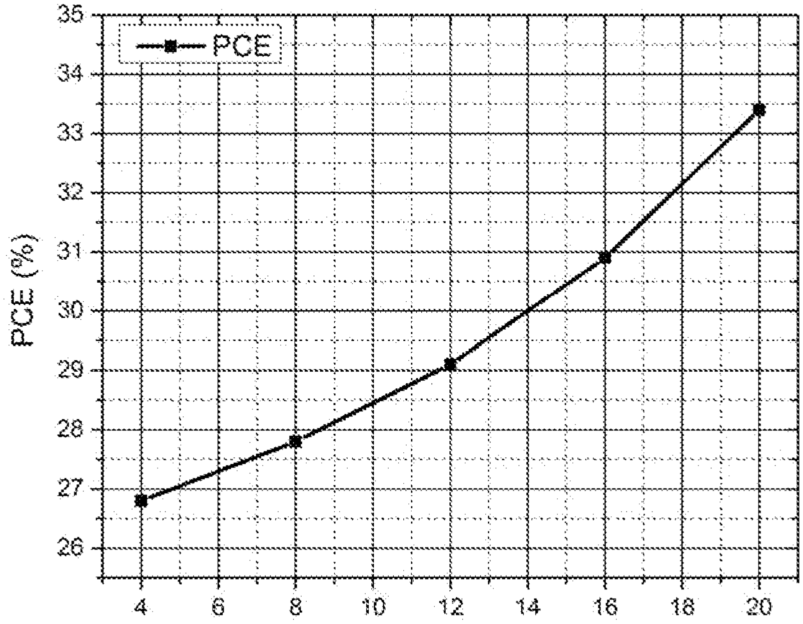

Fig. 12B

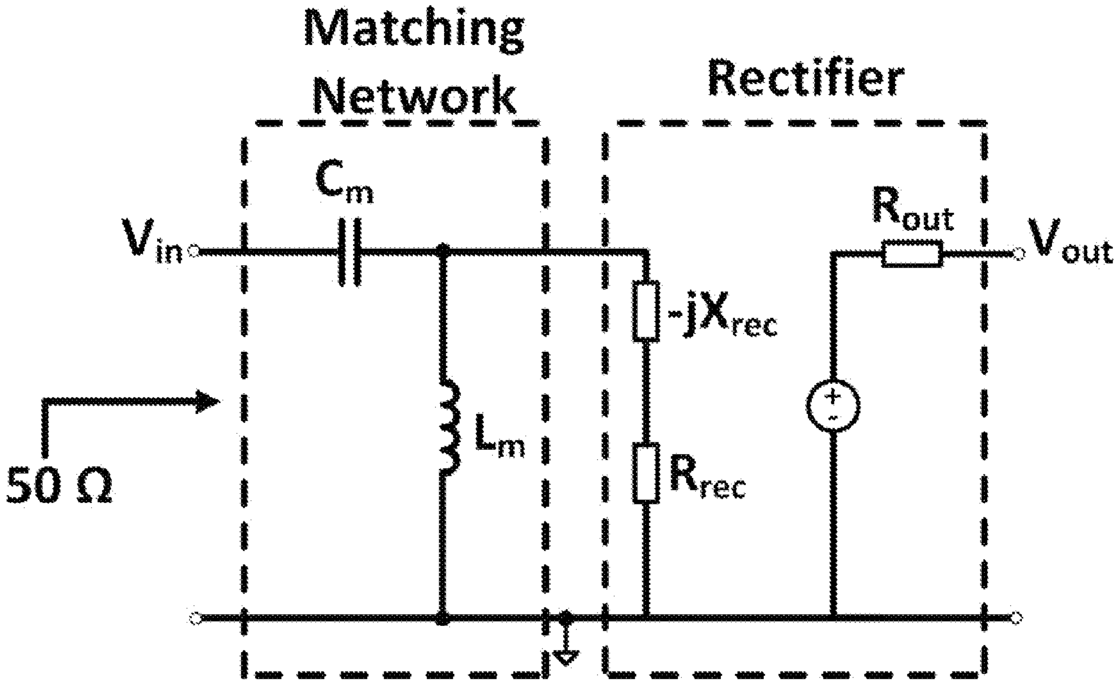
Matching Network
Rectifier
C_m
V_{in}
L_m
50 Ω
$-jX_{rec}$
R_{rec}
R_{out}
V_{out}
+
~

Fig. 13

RF-DC POWER CONVERTER

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit under 35 USC 119(e) of U.S. provisional application Ser. No. 62/011,957 filed Jun. 13, 2014.

TECHNICAL FIELD

[0002] Rectification, particularly for RF energy harvesting.

BACKGROUND

[0003] The vision of realizing Internet of Things (IoT) pervasively connecting large number of sensors and devices requires development of novel solutions that scavenges ambient energy to supply the power required for the operation of the sensors. Relying on batteries as the source of energy for wireless sensors impose several limitations including the need for routine maintenance/charging of batteries, operation interruption and cost involved in replacing batteries specially those employed in harsh environments, and challenges of scaling of battery-powered wireless sensors to millions of nodes. RF energy from ambient sources can be used to power efficiently the sensor networks with or without batteries. In addition, they can also be used to partially/fully charge other portable electronic devices to extend their battery life.

[0004] There is accordingly a growing interest in harvesting ambient energy to partially/fully supply the energy required for the operation of portable electronic devices. Scavenging energy from the ambient electromagnetic wave referred as RF energy harvesting is one of the most popular method for powering low-power wireless sensors. As most of today's integrated circuits are fabricated in CMOS technology, it is highly desirable to integrate RF energy harvesting systems with the rest of the system on a single CMOS chip for reduced cost and small form factor. RF powered devices are widely used in wireless sensor networks operating at power levels of 10 to 100's of μW. RF energy harvesting can be used in applications where the traditional energy sources such as light, vibrations are not available.

[0005] In an RF energy harvesting system **20** as shown in FIG. **1**, an antenna **22** receives the incident RF signal, an impedance matching circuit **24** maximizes the power transfer from the antenna to the power converter **26**, and a RF-DC power conversion circuit **26** converts the incident RF power to DC output power. The output DC voltage is stored in an energy storage component **28** (battery or capacitor) or can directly power the wireless sensors. The major challenge of scavenging RF energy is the limited signal strength of the RF waves and the low efficiency of the harvesting circuit at low input power.

[0006] The radiated signals are received by the RF energy harvesting system **20** and converted into a DC output voltage which is used to power-up the system. The radiated signals have to adhere to the regulation standards set by the federal communication commission (FCC). The unlicensed industrial, scientific and medical (ISM) band with a center frequency of 915 MHz operating at 902-928 MHz with a maximum effective isotropic radiated power (EIRP) of 4 W can be used for RF energy harvesting. The major challenge of scavenging RF energy is the limited signal strength of the RF waves and the low efficiency of the harvesting circuit. The signal strength is limited due to the attenuation in the signal due to the free space loss. Also, the power at the transmitter level cannot be increased without violating the FCC guidelines. Hence modification has to be done at the receiver side. The power consumption for the commercial sensor network nodes vary based on the manufacturers and has been estimated by various authors to be around 10 to 100 μW depending on the sensing application and the radio protocol. A multivolt supply voltage is typically required for the operation of the sensor circuitry.

[0007] The power harvester unit **20** comprising a multistage rectifier **26** is a key component in RF energy harvesting systems. It converts the incoming weak RF signal into a DC voltage. The performance of the rectifier unit **26** can be evaluated based on its power conversion efficiency (PCE) which is the ratio of power delivered to the load to the input power, based on sensitivity i.e. the minimum input power required for production of a DC voltage at the output and finally based on output DC voltage levels. To increase rectifier's PCE, the energy losses such as those introduced by the non-zero ON resistances of the rectifying devices must be reduced. To increase sensitivity and output voltage levels, rectifying devices with lower threshold voltages are required. Hence, these performance parameters of the power harvester are strongly affected by the threshold voltage of the rectifying devices. The power level has to exceed the threshold voltage of the rectifying device to turn on the rectifier unit in the RF-DC power conversion circuit. The minimum power is referred as the power-up threshold of the system. In terms of voltage level, a minimum input voltage is required for the circuit operation and is referred as the dead-zone of the rectifier.

[0008] The performance of an RF energy harvesting system **20** is significantly affected by the threshold voltage of the rectifying device **26**, the voltage that is required to turn on the semiconductor devices used as rectifying devices. A rectifying device **26** with lower threshold voltage enables the operation of RF-DC power converter **26** at low input power levels significantly reducing the power-up threshold of the rectifiers, and increasing the output voltage level for the same input power. The threshold voltage of device can be reduced using different technology-based approaches for the devices including silicon-on-sapphire (SOS), schottky diodes such as silicon-titanium schottky diodes or SMS and the HSMS diodes, special low-threshold-voltage transistors in CMOS process, and floating gate transistors which store a precharged voltage at the gate to lower the threshold voltage. The drawback of using technology-based approaches is additional fabrication steps that increases the production cost and prevents integration of RF energy harvester in mainstream Complementary Metal-Oxide-Silicon Integrated Circuits (CMOS ICs). Active/passive circuit techniques can be alternatively used to reduce the threshold voltage of the device. Active technique requires an external power source or secondary battery and is generally used in active sensors or active RFID. This enables more sophisticated applications at the price of increased cost and maintenance. Passive techniques do not require an additional source of energy but may require additional circuit where an auxiliary rectification chain is used to generate the compensating threshold voltage for the main RF-DC power conversion circuit. The auxiliary chain requires additional power and occupies additional area. An internal threshold voltage cancellation circuit was introduced in Kocer et al. where the compensating voltage was generated passively and stored in a capacitor that is applied at the gate-source terminal of the MOS transistor. This technique requires a large resistance and high capacitance value which leads to a large area on the chip. A self-biasing technique consisting of an off-chip high impedance resistive network was used by Li et al. to provide DC biasing voltages. Another technique can be using floating gate transistors storing a pre-charged voltage at the gate lowering its threshold voltage. The threshold voltage can be varied by changing the body-source potential of the transistor. This technique requires an additional pre-charge phase making it unsuitable for fully battery-less applications. The RF-DC power conversion circuit consisting of NMOS transistors with grounded body terminal leads to an increase in the threshold with the number of stages due to the body effect. This degrades the efficiency of the power conversion circuit. The body terminal of the transistors can be dynamically controlled using additional circuit or floating well devices. The floating well technique generates undesirable substrate current. Also, a triple-well source-body connected device has high parasitic capacitance leading to reduced efficiency. A cross-coupled differential scheme was used in Sciorcioni et al. consisting of triple-well NMOS and standard PMOS transistors to reduce the threshold voltage. A self-compensation scheme based on the Dickson topology was introduced in Curty et al. consisting of triple-well NMOS transistors to provide individual body biasing. The compensating voltage was provided by connecting the gate terminal to later stages. The design of the power converter in the works was focused on only reducing the threshold voltage neglecting the key role played by the reverse leakage current in introducing power losses. The reduction in threshold voltage increases the reverse conduction current causing additional loss.

[0009] To address the trade-off between the reduced threshold voltage and increased leakage current, Lee et al. proposed use of high speed comparators to control the reverse leakage current. The use of comparator increases the power consumption and limits the usefulness of this technique to low-frequency applications. Dynamic CMOS Dickson pump has been designed in previous works to eliminate the V_{th} drop while reducing the reverse leakage current. These circuits are designed for digital application with no emphasis on low-power operation. Differential-drive (**4**T-cell) architecture with the cross coupled bridge configuration and its variant has been used in previous works to reduce the threshold voltage as well as lower the leakage current. Differential circuit requires a PCB balun for the single-ended to differential conversion or differential antenna. Also, the differential circuit requires triple-well NMOS transistors and larger number of rectifying devices for the same number of stages compared with single-ended one.

SUMMARY

[0010] A power conversion circuit is provided having a first input line and a second input line, the first and second input lines configured to receive an alternating voltage differential between the first and second input lines, a multi-stage rectifier comprising transistors arranged in series, each transistor having a gate, a source and a drain, adjacent transistors of the series being connected so that for adjacent p-type transistors the drain of the left p-type transistor is connected to the right adjacent source of the right p-type transistor and source of the right p-type transistor is connected to the left adjacent drain of the left p-type transistor, for adjacent n-type transistors the source of the left n-type transistor is connected to the right adjacent drain of the right n-type transistor and drain of the right n-type transistor is connected to the left adjacent source of the left n-type transistor, for a p-type transistor adjacent to an n-type transistor the source of the right p-type transistor is connected to the source of the left adjacent n-type transistor and drain of the left p-type transistor is connected to the drain of the right adjacent n-type transistor to form a junction, each junction being connected to one of the first input line and the second input line via a capacitor, with adjacent junctions having one junction of the adjacent junctions connected to the first input line and the other junction of the adjacent junctions connected to the second input line, the gate of each transistor of the multi-stage rectifier being connected to a respective junction that is not a junction formed by the connection of that transistor to an adjacent transistor but a junction in the previous stage for the p-type transistor or later stage for the n-type transistor, at least one of the transistor in the multi-stage rectifier being p-type and at least one being n-type.

[0011] In various embodiments, there may be included any one or more of the following features: There may also be an auxiliary chain of p-type transistors, each auxiliary transistor having a gate, a source and a drain, each transistor of the auxiliary chain being connected to a p-type transistor of the multi-stage rectifier in the main chain, so that the gate of the respective auxiliary transistor is connected to the source of the respective transistor in the main chain, the source of the respective auxiliary transistor is connected to the gate of the transistor in the main chain and also connected to the previous stage junction of the main chain which is N transistors away from the respective transistor, and the drain of the respective auxiliary transistor is connected to the drain of the respective transistor or to the junction of the later stage transistor in the main chain. There may also be a diode connected transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of previous stages which is N transistors away from the respective transistor. There may also be an auxiliary chain of n-type transistors, each auxiliary transistor having a gate, a source and a drain, each transistor of the auxiliary chain being connected to an n-type transistor of the multi-stage rectifier in the main chain, so that the gate of the respective auxiliary transistor is connected to the source of the respective transistor in the main chain, the source of the respective auxiliary transistor is connected to the gate of the transistor in the main chain and also connected to the later stage junction of the main chain which is N transistors away from the respective transistor, and the drain of the respective auxiliary transistor is connected to the drain of the respective transistor or to the junction of the previous stage transistor in the main chain. There may also be a diode connected transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of later stages which is N transistors away from the respective transistor.

[0012] There is also provided an auxiliary chain of transistors, each auxiliary transistor having a gate, a source and a drain, each transistor of the auxiliary chain being connected to a respective transistor of a main chain of a multi-stage rectifier in the main chain, so that the gate of the respective auxiliary transistor is connected to the source of the respective transistor in the main chain, the source of the respective auxiliary transistor is connected to the gate of the transistor in the main chain and also connected to the later stage junction of the main chain which is N transistors away from the respective transistor in the case that the respective transistor is n-type, or to the earlier stage junction of the main chain which is N transistors away from the respective transistor in the case that the respective transistor is p-type, and the drain of the respective auxiliary transistor is connected to the drain of the respective transistor or to the junction of a previous stage transistor in the main chain, in the case that the respective transistor is n-type, or to the junction of a later stage transistor in the main chain, in the case that the respective transistor is p-type.

[0013] In various embodiments, there may be included any one or more of the following features: the transistors of the main chain and auxiliary chain may be n-type and there may also be a p-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of later stages, such that the gate of the additional p-type auxiliary transistor is connected to the gate of the n-type auxiliary transistor and the source of the n-type and p-type auxiliary transistors are connected, and the drain of the p-type auxiliary transistors is connected to junction of the main chain which is at 'N' later stages from the respective transistor. The transistors of the main chain and auxiliary chain may be p-type and there may also be an n-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of earlier stages, such that the gate of the additional n-type auxiliary transistor is connected to the gate of the p-type auxiliary transistor and the source of the n-type and p-type auxiliary transistors are connected, and the drain of the n-type auxiliary transistors is connected to junction of the main chain which is at 'N' earlier stages from the respective transistor. There may also be a diode connected p-type or n-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of later stages which is N transistors away from the respective transistor. There may also be a diode connected p-type or n-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of earlier stages which is N transistors away from the respective transistor.

[0014] In any of the above embodiments, there may be included any one or more of the following features: There may also be an additional transistor connected in series with the multistage rectifier, the additional transistor having a gate, a source and a drain, the gate of the additional transistor being connected to the drain of the additional transistor. The gate of each transistor in the main chain may be connected to the junction of previous stages for p-type transistor and later stages for n-type transistor. N of the transistors in the multistage rectifier arranged in series may be n-type and all but N of the transistors in multistage rectifier arranged in series may be p-type. The second input line may be grounded. The body terminal of each p-type transistor may be connected to the respective drain and the body terminal of each n-type transistor may be either grounded or connected to the respective drain terminal.

[0015] These and other aspects of the device and method are set out in the claims, which are incorporated here by reference.

BRIEF DESCRIPTION OF THE FIGURES

[0016] Embodiments will now be described with reference to the figures, in which like reference characters denote like elements, by way of example, and in which:

[0017] FIG. **1** is a block diagram showing energy flow from a receiver through a power harvester to a load;

[0018] FIG. **2**A is a schematic diagram of a prior art half-wave rectifier using a diode;

[0019] FIG. **2**B is a schematic diagram of a prior art half-wave rectifier using diode-connected NMOS;

[0020] FIG. **2**C is a schematic diagram of a prior art half-wave rectifier using diode connected PMOS;

[0021] FIG. **2**D is a schematic diagram of a prior art voltage doubler full wave rectifier using diodes;

[0022] FIG. **2**E is a schematic diagram of a prior art voltage doubler full wave rectifier using diode connected NMOS;

[0023] FIG. **2**F is a schematic diagram of a prior art voltage doubler full wave rectifier using diode connected PMOS;

[0024] FIG. **2**G is a schematic diagram of a prior art threshold-compensated transistor in a single-stage rectifier;

[0025] FIG. **3**A is a graph of output current with respect to input voltage for a prior art PMOS voltage doubler as shown in FIG. **2**F;

[0026] FIG. **3**B is a graph showing input and output voltage over time of the threshold-compensated transistor of FIG. **2**G;

[0027] FIG. **4**A is a schematic diagram of a prior art NMOS Dickson charge multiplier;

[0028] FIG. **4**B is a schematic diagram of a prior art PMOS Dickson charge multiplier;

[0029] FIG. **5** is a schematic diagram of a prior art forward compensated charge multiplier;

[0030] FIG. **6**A is a schematic diagram of a level-1 back compensated PMOS charge multiplier;

[0031] FIG. **6**B is a schematic diagram of a level-1 hybrid forward and back-compensated multiplier using NMOS and PMOS;

[0032] FIG. **6**C is a schematic diagram of a level-3 hybrid forward and back-compensated multiplier using NMOS and PMOS;

[0033] FIG. **7** is a contour plot of constant efficiency for input power versus level of compensation;

[0034] FIG. **8**A is a contour plot of constant efficiency for width versus number of stages for a level-1 compensated multiplier;

[0035] FIG. **8**B is a contour plot of constant efficiency for width versus number of stages for a level-3 compensated multiplier;

[0036] FIG. **9**A is a schematic diagram of a multiplier with an adaptive threshold-voltage compensation scheme using diode-connected transistors;

[0037] FIG. **9**B is a schematic diagram of a multiplier using p-type transistors with an adaptive threshold-voltage compensation scheme using solid-wired connections;

[0038] FIG. **9**C is a schematic diagram of a multiplier using n-type transistors with an adaptive threshold-voltage compensation scheme using solid wired connections;

[0039] FIG. **9**D is a schematic diagram of a multiplier using n-type transistors with an adaptive threshold compensation scheme using p-type and n-type transistors;

[0040] FIG. **10**A is a contour plot of constant efficiency for width versus number of stages for an adaptive diode-connected scheme as shown in FIG. **9**A at an operating frequency of 915 MHz;

[0041] FIG. **10**B is a contour plot of constant efficiency for width versus number of stages for an adaptive solid-wired scheme as shown in FIG. **9**B at an operating frequency of 915 MHz;

[0042] FIG. **11**A is a graph of leakage current versus input power with different schemes at 915 MHz;

[0043] FIG. **11**B is a graph of current transfer ratio versus input power with different schemes at 915 MHz;

[0044] FIG. **12**A is a graph of efficiency versus number of adaptive blocks for the adaptive diode-connected scheme;

[0045] FIG. **12**B is a graph of efficiency versus number of adaptive blocks for the adaptive solid-wired scheme; and

[0046] FIG. **13** is a schematic diagram showing an impedance matching circuit.

DETAILED DESCRIPTION

[0047] There is provided a hybrid forward and backward threshold voltage compensation scheme for RF-to-DC power conversion. In an embodiment, PMOS transistors are used as rectifying devices in all stages except for the first few stages eliminating the need for triple-well NMOS transistors. The compensating voltage is provided by connecting the gate terminal of the PMOS transistors to previous stages, also referred to as back-compensation, and also by connecting the gate terminal of the initial few NMOS transistors to later stages. Using PMOS transistors as a rectifying device in all the stages except for the first few stages eliminate the need for triple-well NMOS transistors. Also, all the MOS transistors in the main rectification chain are compensated in the proposed scheme. In addition to the hybrid scheme, an adaptive method suitable for rectifying devices operating at low power level is proposed to control the ON/OFF operation of the MOS transistors in the main rectification chain. An auxiliary block consisting of PMOS transistor is introduced to control the switching of the MOS transistors in the main chain. The PMOS transistors in the main chain are back-compensated in the conduction phase and forward-compensated when not conducting; increasing the gate-source potential and reducing the reverse leakage. The controlling voltage of the transistors in the auxiliary chain is derived from the local node of the main chain. Moreover, a procedure for designing matching network to provide passive voltage amplification and maximum power transfer is proposed. The dead-zone of the power converter unit is obtained using Y parameter analysis and the matching circuit is designed at the point of maximum power transfer. The RF-to-DC power conversion circuit may be fabricated using conventional methods such as IBM's 0.13 μm CMOS technology.

[0048] Power Harvester Circuit

[0049] FIG. **1** shows a block diagram of the power harvester unit of an RF energy harvesting system **20**. In the RF energy harvesting system shown, the signal is received by an antenna **22**. An impedance matching circuit **24** comprising a high quality factor (Q) resonator is used to ensure that maximum power is transferred from the receiver to the rectifier block **26**. However, the matching network itself introduces some energy losses because of the limited quality factors of on-chip passive inductors and capacitors required to match the input impedance of the rectifier to the output impedance of the antenna. The high Q resonator can also provide passive voltage amplification of the input signal. Depending on the embodiment, the high Q resonator may have a fixed frequency or may be tuneable. The next block is the RF-to-DC power conversion circuit **26** that converts the RF signal to DC voltage. The rectifier preferably operates at low input power levels while providing high power conversion efficiency. Several rectifier units may be cascaded to increase the overall output voltage, referred collectively as a voltage multiplier. There may be a power management block (not shown) to control the flow of power from the energy harvester to the load. The harvested power is then finally used to charge an energy storage device **28** such as a battery or a capacitor, which in turn powers a load **30**. Alternatively, the energy storage device may be omitted and the load powered directly by the power harvester. The power management block may also be omitted with or without the energy storage.

[0050] Rectifier Circuit

[0051] In half-wave rectifiers, the rectifying device is conducting for half of the cycle, passing the input voltage to the output. However, the output voltage is lower than the input by the amount of voltage drop needed for turning ON the rectifying device. A half-wave rectifier can be implemented using diodes as shown in FIG. **2**A. The voltage drop across the device depends on the threshold voltage of the diode. FIGS. **2**B and **2**C show half-wave rectification circuits implemented using diode-connected NMOS and PMOS transistors respectively in CMOS technology. The voltage drop in this case depends on the threshold voltage of the transistor. Both diode-connected PMOS and NMOS transistors can be used as rectifying devices. FIG. **2**D shows the voltage doubler unit which is a cascade of two half-wave rectifiers. The voltage doubler rectifies the AC input in both the positive and negative cycles. Similar to the half wave rectification, the diodes are implemented by connecting the drain and the gate terminal of the MOS transistor together such that the transistor is always in saturation region in forward bias condition, as shown in FIGS. **2**E and **2**F.

[0052] The voltage doubler may be used as a basic unit in a power conversion circuit. FIG. **2**F shows an example voltage doubler. In the voltage doubler, each of the transistors conducts only during one half of the input cycle. M_{P1} conducts during the negative phase while M_{P2} conducts during the positive phase of the input cycle. Each of the transistors goes through the sub-threshold region, inverse region and the leakage region. The maximum voltage at the output of voltage doubler is $2\,Vamp-|VTP1|-|VTP2|$, where V_{amp} is the maximum amplitude of the input signal and V_{TP1}, V_{TP2} are the threshold voltages of the diode-connected transistors M_{P1} and M_{P2} respectively. Thus the maximum possible voltage is twice the RF signal's amplitude only when the threshold voltage of the transistors is equal to zero. To understand the operation of the voltage doubler in detail, we can look at the transient analysis of the doubler circuit of FIG. **2**F, as shown in FIG. **3**A. There are three regions of operation of the circuit as seen from FIG. **3**A. It is described as follows: the sub-threshold operation extends from $V_{in}=0$ to $V_{in}=|V_{TP}|$, where V_{in} is the input voltage. The current in this region is an exponential function of the input voltage. The inversion region extends from $V_{in}=|V_{TP}|$ to $V_{in}=V_{amp}$. The current in the in this region is a square function of the input voltage. In the inversion region, the output current reaches its peak value when $V_{in}=V_{amp}$. Finally, the leakage region extends from $V_{in}=0$ to the next $V_{in}=0$ in the negative half-cycle. The current that flows through transistor M_{P2} during this time interval is referred as reverse leakage current. Thus, the actual charge transfer mechanism is only for a short duration while other parameters such as subthreshold and reverse leakage currents have to be considered for the rest of the cycle.

[0053] Several voltage doublers can be cascaded to increase the overall output voltage. These voltage multipliers are also known as "charge pumps" and can generate a voltage several times larger than their input supply voltage. The Dickson multiplier with diode-connected transistors is commonly used for integrated applications. The Dickson multiplier of

FIG. **4**A is modified for low power energy harvesting applications by grounding the ϕ_2 (out of phase) clock terminal and applying the input signal at the ϕ_1 terminal. FIG. **4**B shows the conventional Dickson multiplier with the PMOS voltage doubler as a rectifier unit. The body terminal of the transistors is connected as shown in FIG. **4**B to further reduce the threshold voltage while the transistors are conducting. For example, in the positive phase of the input cycle, transistors M_2, M_{n-1} and M_{n+1} are conducting. The source of the transistors is at a higher potential compared to the body terminal. This reduces the threshold voltage in the conducting region. At the same time, transistors M_1, M_3 and M_n are in the reverse region (leakage operation). The gate-source and the body-source potential is zero for these transistors. Thus, the threshold voltage of transistors M_2, M_{n-1} and M_{n+1} while they are conducting is lower than the threshold voltage of the transistors M_1, M_3 and M_n which are in the reverse region in the positive input phase. Similarly, in the negative phase of the input cycle, transistors M_1, M_3 and M_n which are conducting have a lower threshold voltage than the transistors M_2, M_{n-1} and M_{n+1}.

[0054] To increase the output voltage of multi-stage rectifiers, the number of rectifier stages must be increased accordingly. However, as the number of stages increases the power conversion efficiency is reduced as larger number of the transistors dissipates more power. Also, the threshold voltage of the standard MOS transistor which lies around 450 mV in 130 nm technology, reduces the sensitivity of the power harvester block. As the number of stages in the rectifier increases, the power losses increases, reducing the overall efficiency. Too few of stages, leads to low output DC voltage even if the PCE is high. To achieve a high DC voltage, large transistors have to be used, leading to high leakage and parasitic losses whereas smaller transistor size affects the charge transfer leading to low DC voltage. The strategy while designing the multi-stage rectifier circuit of FIG. **4**B is as follows: The individual stages of the voltage doubler can be cascaded to increase the output voltage. As the number of stages increases, the output voltage increases. With the increase in the number of stages, the PCE decreases and the output voltage saturates. Hence the transistors as well as the pumping capacitors are scaled while increasing the stages. The scaling is done to maintain the incremental voltage per stage and the PCE relatively constant with the increase in the number of stages. Thus the PCE can be maintained with proper device scaling while increasing the DC voltage for a constant output power. The scaling of width and the pumping capacitance with the stages results in impedance looking into the rectifier to be unchanged so that the input power and thus the PCE remains constant. Thus, when designing a multistage rectifier unit one can optimize of number of stages and width of the transistors to maximize the PCE for a given output voltage.

[0055] The standard Dickson multiplier can be modified for designs involving energy harvesters where the input voltage is low. FIG. **4**A shows an N-stage conventional Dickson multiplier. A threshold self-compensation technique was described by Dickson. The gate of the transistor in this technique is connected to the adjacent source of the transistor instead of the traditional diode-connected structure. Thus, providing bias voltage equivalent to the incremental voltage across each stage. Based on this technique, a forward compensated topology was implemented by Papotto et al. where the bias voltage was increased by extending the gate length connection. FIG. **5** shows the basic forward compensated NMOS topology.

[0056] Threshold voltage reduction techniques requiring additional circuits are not suitable for integrated low power energy harvesting applications as these circuits occupy large areas and cause additional power dissipation. Passive threshold voltage reduction technique such as the self-compensation method of FIG. **5** does not require additional circuit and can reduce the threshold voltage. In the self-compensation method, the threshold voltage of the diode-connected NMOS transistors increases with the stages due to the body effect. The body effect can be dynamically controlled using additional circuits but generates undesirable substrate current. Another alternative is using triple-well process to individually bias the body terminal of NMOS transistors and reduce the threshold voltage variation. Additional parasitic capacitance is introduced due to the well structure increasing the losses at each node and affecting the overall efficiency.

[0057] Hybrid Forward and Back-threshold compensated power converter

[0058] In order to eliminate the need for triple-well processes, we propose a threshold voltage compensation technique scheme as described below. PMOS transistors are used instead of NMOS transistors as rectifying devices. As each PMOS transistor has its own n-well, the body of PMOS transistor can be biased individually without the necessity of a triple-well CMOS process. Unlike an NMOS transistor which needs a higher potential at the gate terminal to offset the threshold voltage, a PMOS transistor requires negative gate-source potential. Therefore, the threshold voltage of PMOS transistors can be reduced by connecting the gate potential to the previous node rather than later node. As shown in FIG. **6**A, alternating input voltage V_{in} is applied on first input line **100** to odd-numbered capacitors $C_1, C_3, \ldots C_n$. Ground voltage is applied on second input line **102** to even numbered capacitors $C_2, \ldots C_{n-1}$. Capacitor C_{n+1}, while also even-numbered, is shown as connected to ground by a separate line. A third line **104** accumulates voltage for the output and is connected to ground at one end and to output at the other. In this and other figures, the end of the line having more negative voltage, which in the embodiments shown is ground, is shown at the left of line **104** and the end of the line having more positive voltage, which in the embodiments shown is the output, is shown at the right of line **104**. The capacitors are each connected to line **104**, and are numbered in the order in which they are connected to line **104** from left to right. Each capacitor maintains a respective voltage differential between the respective input line it is connected to and the part of line **104** it is connected to. This voltage differential varies over time as current flows through line **104** between capacitors. Line **104** comprises transistors $M_1, M_2, M_3, \ldots M_{n-1}, M_n, M_{n+1}$ arranged in series, a transistor being arranged between each pair of adjacent capacitors and between capacitor C_1 and ground. Adjacent transistors define junctions **106**, the capacitors being connected to the junctions. The body terminal of each transistor is connected to the drain terminal of the same transistor, as shown by dotted lines. The gate of the first transistor (for n-level compensation, the first ‘n’ transistors) and the last transistor are connected to their respective drains. The gates of the other transistors are connected to the drain of the transistor n+1 transistors to the left of the respective transistor. The proposed scheme as shown in FIG. **6**A reduces the threshold voltages of all PMOS transistors except the first one leading to an increased output voltage. For an n-level compensation, for the proposed scheme, there will be ‘n’ initial PMOS transistors that will be uncompensated as seen from FIG. **6**A To solve this problem, ‘n’ uncompensated PMOS transistors are replaced by NMOS transistors with grounded body terminals as seen from FIG. **6**B. As shown in FIG. **6**B, the first transistor is an NMOS transistor with a source connected to the source of the PMOS transistor to its right to define a junction and has a gate connected to the junction n stages to the right of the junction defined by its source. That is, the gate is connected to the junction defined by the source of the PMOS transistor **2** (i.e, n+1) transistors to the right. The other transistors are PMOS and, apart from the last, each has a gate connected to a junction n stages to the left of the junction defined by its source. For the purposes of the preceding statement and statements about junctions elsewhere in this document, the grounded end of line **104** can be considered to be a junction **106**A one stage from the junction defined by the connection of M_1 and M_2. Thus the compensating voltage is provided by connecting the gate terminal to later stages for the first few NMOS transistors and connecting the gate terminal to previous stages for PMOS transistors. The scheme shown in FIG. **6**A and FIG. **6**B is level-1 compensation. The compensation level can be increased by connecting the gate terminal of PMOS to the source terminals of the transistor of the following stage rather than the source of its immediate neighbor. The last transistor of the multiplier is left uncompensated to reduce the leakage. FIG. **6**C shows the proposed level-3 hybrid forward- and back-compensated multiplier. As shown in FIG. **6**C, the first **3** (i.e., n) transistors are NMOS transistors with grounded body terminals. Each NMOS (more generally, n-type) transistor has its source on the left and drain on the right, and each PMOS (more generally, p-type) transistor has its drain on the left and source on the right, where “left” and “right” are defined so that that the end of line **104** with more positive voltage is the “right”, as shown in the figures. “Earlier” and “later” can also be used to mean “left” and “right”. While ground is shown in the figures to the left and the output to the right, the ground could also be to the right and the output to the left. See below for a discussion of the definitions of “source” and “drain”. The gates of these first **3** NMOS transistors are connected to the respective junctions **106** three (that is, ‘n’) stages to the right of the junctions defined by their respective sources. The remaining transistors except the last are PMOS transistors with gates connected to the junctions **106** three (that is, ‘n’) stages to the left of the junctions defined by their respective sources. Increasing the level of compensation leads to reduction in the threshold voltage which improves the forward conduction but also leads to increased reverse leakage current degrading the rectifier’s PCE. Only odd level compensation is used as it maximizes the source-gate potential of PMOS transistors due to the alternating voltage phase with successive nodes.

[0059] To find the optimum number of stages and level of compensation, extensive simulations were conducted. FIG. **7** shows the efficiency contour plots with different level of compensation and input power level. Maximum PCE at the lowest input power level is obtained when the level of compensation is one. As the level of compensation increases, the reverse current increases which causes additional power loss and degradation in efficiency. Hence level-1 compensation gives the maximum efficiency while level-3 or higher is advantageous in reducing the threshold voltage of the RF-DC power conversion circuit. FIG. **8**A shows the constant efficiency contour plot as a function of transistor width and the number of stages for level-1 compensation. Eight-stage of doubler design equivalent to 16-stage rectifier with transistor width of 13-μm and fifteen-stage of doubler design equivalent to 30-stage rectifier with transistor width of 28 μm gives the highest efficiency contour. The former one is preferred due to lesser area on the chip. The level-1 and level-3 compensated multi-stage rectifier are referred to as efficiency circuit and voltage circuit respectively Similarly for the voltage circuit constant efficiency contour plot as a function of width and the number of stages is plotted as shown in FIG. **8**B. Twelve-stage of doubler design equivalent to 24-stage rectifier with transistor width of 8 μm is selected based on the plot. For level-3 compensation, more body-grounded NMOS transistors are required compared with the level-1. Also, the reverse leakage loss is higher for level-3 compensation due to the larger compensating voltage compared with the level-1 compensation. Increasing the level of compensation lowers the minimum input voltage requirement whereas increasing the number of stages while lowering the width of the transistors was based on the design strategy discussed earlier.

[0060] During the positive input phase, transistors M_2, M_{n-1} and M_{n+1} in FIGS. **6**A-**6**C are conducting. Since the gate terminals are connected to the previous node V_{SG} in the positive phase (conduction) is relatively high compared to the conventional Dickson multiplier. However, in the negative phase when the transistors M_2, M_{n-1} and M_{n+1} should ideally be not conducting, the V_{SG} is higher compared to the conventional case. Thus the hybrid forward and back-threshold compensated scheme is effective in increasing the forward conduction current but increases the reverse leakage current at the same time. As seen in FIG. **7**, with the increase in the level of compensation the PCE degrades. The simulated maximum PCE of 28% degrades to 11% from level-1 compensation to level-3 compensation respectively. To ameliorate this issue an adaptive scheme is proposed to control the reverse leakage current.

[0061] Especially at extremely low input power (e.g. micro-watts power level), the reverse leakage current has a significant adverse effect on the PCE and the output DC voltage of an RF rectifier as proven by the following analysis. Considering a threshold-compensated transistor of a single-stage rectifier shown in FIG. **2**G driven by an input source $V_{in}=V_a \sin \omega t$ assuming the compensation voltage is modeled by V_C and its input and the output steady-state voltage waveforms shown in FIG. **3**B, the overall PCE is defined as

$$PCE = \frac{P_{out,forward} - P_{leakage}}{P_{input}} \tag{1}$$

[0062] where $P_{out,forward}$ is the output power delivered to the load when transistor is forward-biased, $P_{leakage}$ is the output power lost due to leakage when transistor is reverse-biased, and P_{input} is the input power. The forward region extends from $t=t_1$ to $t=\pi-t_1$ with the current $I_d(t)$ through the MOS transistor M_1 as seen in FIG. **3**B.

$$\frac{P_{out,forward}}{P_{input}} = \tag{2}$$

-continued

$$\frac{\frac{1}{\pi-2t_1}\int_{t_1}^{\pi-t_1} V_{out}\cdot I_d\cdot dt}{\frac{1}{\pi-2t_1}\int_{t_1}^{\pi-t_1} V_{in}\cdot I_d\cdot dt}=\frac{\frac{1}{\pi-2t_1}\int_{t_1}^{\pi-t_1}(V_a-V_{TH}+V_c)\cdot dt}{\frac{1}{\pi-2t_1}\int_{t_1}^{\pi-t_1} V_a\sin\omega t\cdot dt}$$

Assuming the ripple voltage variation A V is much smaller than the average output voltage V_o, the output voltage for the one-stage rectifier can be expressed as $V_o=V_a-V_{TH}+V_C$ where V_{TH} is the threshold voltage of the transistor. Performing integration on (2) gives the ratio of the output power in the forward region to the input power as

$$\frac{P_{out,forward}}{P_{input}}=\frac{\omega\cdot(\pi-2t_1)\cdot(V_a-V_{TH}+V_c)}{2V_a\cos\omega t_1} \tag{3}$$

[0063] The value of t_1 lies between $0<t_1<\pi/2$ based on the value of $V_{TH}-V_C$. The time t_1 indicates the onset of inversion region and is close to zero when the compensating voltage is near the threshold voltage and will be closer to $\pi/2$ when the compensating voltage is near-zero value. As seen in (3), the ratio of the output power in the forward region to the input power increases with increased voltage compensation in the forward region. In the reverse-biased region, the leakage current is expressed as

$$I_{leak}(t)=I_o\cdot(W/L)\cdot e^{(V_{gs}-V_{TH})\eta V_T}\cdot(1-e^{-V_{ds}/\eta V_T}) \tag{4}$$

[0064] Replacing the gate-source bias voltage by V_C, source-drain bias by $V_{in}(t)-V_{out}$ and V_{in} by V_a sin wt, the leakage current as a function of time can be expressed as

$$I_{leak}(t)=I_o\cdot(W/L)\cdot e^{(V_c-V_{TH})/\eta V_T}\cdot(1-e^{(V_a\sin\omega t-V_a+V_{TH}-V_c)/\eta V_T}) \tag{5}$$

[0065] where $(\pi-t_1)<t<(2\pi+t_1)$. With increase in the compensation, the power loss increases due to the higher leakage current. As seen in (3) and (5), the ratio of the output power in the forward region to the input power even though increases with larger threshold-voltage compensation, the losses in the leakage region is higher due to the increased compensation. This indicates the fundamental trade-off between the level of threshold-voltage compensation and the leakage current of the transistors.

[0066] In a field effect transistor, the "source" and "drain" terminals are defined by the direction of flow of charge carriers when the transistor is "on": charge carriers flow from the source to the drain. For an n-type transistor, the charge carriers are electrons and for a p-type transistor the charge carriers are holes. As electrons flow from negative to positive, the source of an n-type transistor is at a lower voltage than the drain and vice versa for a p-type transistor. FETs do not necessarily have any asymmetry between the source and drain terminals beyond this. In the present system, the relative voltages of the source and drain terminals reverse in the normal course of events, as the system is exposed to alternating current in order to rectify it. There are two possible approaches to defining which terminal is the "source" and which terminal is the "drain" in this circumstance. One possible approach is to say that which terminal is the "drain" and which terminal is the "source" varies over time depending on which terminal is at the higher voltage. Thus, by this first definition, the source of an n-type transistor is whichever of the source and drain terminals has the lower voltage at any given moment, and for a p-type transistor whichever has the higher voltage at a given momentA second approach, however, is possible where the desired conduction is in a particular direction only. In the present system, conduction in one direction is desirable and conduction in the other ("leakage") is undesirable, thus the source can be defined according to which terminal is the source when current flows through the transistor in the desired direction; which terminal is considered the "source" and which terminal is considered the "drain" does not change over time in this approach despite the change in voltages. In the discussion here concerning the source-gate voltage (V_{SG}), the first approach is used, i.e. the source-gate voltage is the voltage between the gate and whichever of the source or drain terminals that has the higher or lower voltage at a given time depending on if the transistor is p-type or n-type. In addition, while the sign of a source-gate voltage difference required to operate a transistor depends on whether the transistor is p-type or n-type, V_{SG} is discussed as if the voltage to operate the transistor is always positive for p-type transistor and negative for n-type transistor. For other purposes in this document including the claims, a version of the second approach to defining which terminal is the source and which is the drain is used in which the direction of desired current is based on the desired current direction through the device as a whole. Thus, throughout the figures, p-type transistors are shown as having their sources to the left and n-type transistors are shown as having their sources to the right.

[0067] A diode-connected transistor is a transistor in which one of the source and drain terminals is connected to the gate to cause the transistor to act as a diode. A diode connected FET allows charge carriers to flow from the terminal not connected to the gate to the terminal connected to the gate, if sufficient voltage is applied, and blocks charge carriers from flowing in the opposite direction. The diode connected transistor is diode connected because it is desired that current flow in one direction is allowed and in flow in the other direction blocked. Thus, following a version of the second approach above to defining which terminal is the source and which terminal is the drain based on desired current through the individual transistor, the source or drain terminal of a diode connected transistor that is connected to the gate would be by definition the drain of the diode connected transistor. However, in this document it is the desired current through the device as a whole that is used, thus in e.g. FIG. **9**A transistor M_{na} is diode-connected and the terminal connected to the gate is depicted as the source.

[0068] To investigate the effect of the level of threshold-compensation on the PCE of multi-stage hybrid threshold-compensated rectifiers, level-1 rectifier which is connecting the gate terminal to one previous node for the PMOS transistors (FIG. **6**A) and level-3 rectifier which is connecting the gate terminal to three previous nodes (FIG. **6**C) for the PMOS transistors are simulated to obtain constant maximum efficiency contour plots as shown in FIG. **7**. Increasing the level of compensation leads to reduction in the threshold voltage which improves the forward-conduction but also increases the reverse leakage current. In the multipliers as shown in FIGS. **6**A and **6**C, during the positive input phase, transistor M_n is forward-biased. Since the gate terminals are connected to the previous node, V_{sg} when forward-biased is higher compared to that of the conventional diode-connected case with no threshold-voltage compensation which increases the ratio of output power in the forward region to the input power. However, when the transistor M_n is reverse-biased, the V_{sg} is higher compared to the conventional diode-connected case increasing the reverse leakage current of the reverse-biased transistor. This can also be explained by the efficiency contour plots of the hybrid scheme as seen from FIG. **7**. The contour plots are simulated for constant efficiency with different level of compensation and input power levels. As seen from FIG. **7**, though with the increase in the level of compensation the threshold voltage reduces, the efficiency degrades at the same time which is due to the increased reverse leakage current. The simulated maximum PCE of 28% degrades to 11% from level-1 compensation to level-3 compensation respectively. Accordingly an adaptive scheme is proposed to control the threshold voltage and the reverse leakage current of the rectifying device dynamically to improve the PCE over a wide range of input power. Ideally, the threshold voltage of the rectifying device should be zero when the transistors are forward-biased while the threshold voltage should be high when the transistors are reverse-biased to prevent any leakage or the losses associated with it.

[0069] We propose an adaptive forward and backward threshold-voltage compensation scheme that use minimal additional circuitry to increase the threshold-voltage compensation when transistors are forward-biased and decrease the compensation voltage when they are reverse-biased. The PMOS transistors in the main rectification chain are back-compensated when forward-biased and forward-compensated when reverse-biased; increasing the forward-current and reducing the reverse leakage current dynamically.

[0070] Proposed Adaptive Threshold Voltage Compensation

[0071] The hybrid forward and back-compensated topology described above can be made adaptive using an auxiliary block controlling the gate-source voltage of the MOS transistors in the main rectification chain. The auxiliary structure may be realized using all PMOS transistors to allow for individual body biasing. Preferably the auxiliary blocks are designed using minimum number of PMOS transistors so that the power losses do not increase considerably due to the additional blocks. The controlling voltage of the transistors in the auxiliary chain is derived from the local node of the main rectification chain. Two possible implementation of the proposed adaptive threshold voltage compensation scheme are discussed in the following section, one using diode-connected PMOS transistors and the other solid-wired connection to adaptively adjust the level of the threshold-voltage compensation.

[0072] Using Diode-Connected Transistor for Back-Compensation

[0073] FIG. **9**A shows the hybrid forward and back-threshold compensated rectifier using diode-connected transistors for back-compensation. The back-compensation reduces the threshold voltage when the transistors are forward-biased and forward-compensation reduces the reverse leakage current when reverse-biased with the control signal derived from the local node. In the embodiment shown the last transistor M_{n+1} is left uncompensated to reduce the leakage. During the negative input phase, in which the transistor M_n is forward-biased, the transistor M_n is back-compensated by the diode-connected transistor M_{na} while the V_{SG} terminal voltage for the transistor M_{nb} lies below its threshold voltage resulting in transistor M_{nb} turned OFF. In comparison to the embodiment shown in FIG. **6**B, diode-connected transistor M_{na} lies on the line connecting to the gate of transistor M_n. Transistor M_{nb} lies on a different line not corresponding to a line in FIG. **6**B, providing a forward connection for the gate of transistor M_n when transistor M_{nb} is active. In the embodiment shown in FIG. **9**A p-type transistors are used, but n-type could also be used. Additional transistors of the opposite type may be used at one end as in FIG. **6**B. The back-compensation for the transistor M_n in the main chain enhances the forward current. During the positive input phase, in which the transistor M_n is reverse-biased, the V_{SG} terminal for transistor M_{nb} is enough to turn ON the forward connection thus reducing the V_{SG} bias of the transistor M_n to zero resulting in a reduced leakage current. FIG. **10**A shows the efficiency contour plot for the adaptive diode-connected scheme as a function of width and the number of stages of the rectifier. Generating contour plots are an effective way to optimize the number of stages and the width of the transistors to maximize the PCE while obtaining the required voltage. Twelve-stage of voltage doubler design equivalent to 24-stage rectifier with transistor width of 11 μm gives the highest efficiency contour. The width of the diode-connected transistor in the auxiliary chain should be comparable to the width of the transistors in the main rectification chain so that it provides low forward-resistance when the transistors are conducting. In an embodiment, the width of the diode connected transistor in the auxiliary chain may be 8 μm. The auxiliary transistor used for forward compensation to control the reverse-leakage is selected in an embodiment to be 480 nm, an order of magnitude smaller than the diode-connected transistor to minimize their power consumption and reduce the parasitic at the nodes of the main rectification chain. Larger transistor widths are avoided to reduce the parasitic at the nodes of the main rectification chain.

[0074] Also, for the diode-connected scheme, the auxiliary block is added for every alternate transistor starting from the later stages and adding the blocks towards the initial stages. The maximum PCE increases with the addition of the adaptive auxiliary blocks.

[0075] Using Solid Connection for Back-Compensation

[0076] FIG. **9**B shows the adaptive hybrid topology using solid wired connection instead of the diode-connected transistor for reducing the threshold voltage. When the transistors are in the conduction phase (forward biased) they are back-compensated with solid wired connection instead of the diode-connected transistors which prevents the forward-losses associated with it. In the embodiment of FIG. **9**B, as compared to FIG. **9**A, the transistors M_{na}, $M_{n-1})_a$ etc. are omitted and replaced by a solid wire connection as in FIG. **6**B, but the transistors M_{nb}, $M_{(n-1)b}$ etc. are retained. In the embodiment shown in FIG. **9**A p-type transistors are used, but n-type could also be used. Additional transistors of the opposite type may be used at one end as in FIG. **6**B. When the transistor M_n is reverse-biased (not in the conduction phase), the V_{sg} terminal voltage for the auxiliary transistor M_{nb} is large to turn ON the forward-connection and reduce the source-gate bias of transistor M_n. This decreases the reverse leakage current greatly. FIG. **10**B shows an efficiency contour plot for the solid-wired scheme as a function of width and the number of stages of the rectifier. Twelve stages of doubler design equivalent to 24-stage rectifier with transistor width of 10 μm gives the maximum efficiency contour. The efficiency contour plot for the solid-connection follows similar trend as the diode-connected one. With the increase in the width of the transistors while maintaining the number of stages, the efficiency initially increases and then degrades due to the increased parasitic losses. Similarly, when the width of the transistor is kept constant and the number of stages is increased, the efficiency initially increases and then degrades as more number of stages is added due to the increased power loss with additional stages. Coupling capacitor value of 4 pF is selected as it has a very modest impact on the rectifier's efficiency.

[0077] Further embodiments are shown in FIG. **9**C and FIG. **9**D. The embodiment of FIG. **9**C operates in the same way as the embodiment of FIG. **9**B but uses n-type transistors instead of p-type. Accordingly, the solid wired connections to the gates of the transistors in the main chain are to the right instead of to the left. FIG. **9**D is an embodiment in which the connections to the gates of the transistors of the main chain are each controlled by two transistors. However, unlike in FIG. **9**A, one of these two transistors is p-type and the other is n-type. This allows the gates of these two transistors to be controlled by the same voltage, here using a line connecting to the junction to the right of the corresponding main transistor, and have at most one of the transistors ON at any given point in the input cycle. In addition, the transistors in the main chain are n-type in FIG. **9**D, whereas the transistors of the main chain are p-type in FIG. **9**A. However, either of the embodiments of FIG. **9**A or FIG. **9**D could use n-type or p-type transistors in the main chain, with corresponding changes to the direction of connections as seen in the differences between FIG. **9**B and FIG. **9**C. FIGS. **9**C and **9**D don't show the additional transistor at the end of the rectifier that is shown as M_{n+1} in FIGS. **9**A and **9**B, but such an additional transistor may also be present in these figures. The additional transistor can be p-type or n-type.

[0078] To investigate the effectiveness of the adaptive blocks in reducing the leakage current, leakage current as a function of input power is plotted as shown in FIG. **11**A. The leakage current is simulated for the last transistor with a 1 MΩ load for different input power levels. As shown in FIG. **11**A, comparing the schemes i.e. hybrid, adaptive solid and adaptive diode-connected for power levels of 1 μW-100 μW and at an operating frequency of 915 MHz, the reverse leakage current is maximum for the hybrid scheme and increases with the input power. At an input power of 105 μW, the leakage current is 12 nA. The adaptive solid and the adaptive diode-connected scheme have an auxiliary transistor to control the reverse leakage current. At an input power of 1 μW, the auxiliary transistor does not have enough input power to turn ON and provide the required forward-compensation. Hence the leakage current is comparable with the hybrid scheme. At 1 μW (−30 dBm), the output transistor's leakage current is 6.96 nA for the adaptive solid scheme and 7.56 nA for the adaptive diode-connected scheme with a 1 MΩ load. As the input power increases, the leakage current for the adaptive scheme is drastically reduced as seen in FIG. **11**A. Also, the reverse leakage current for the adaptive schemes is relatively constant with increase in the input power. The leakage current also depends on the load resistance and increases with decrease in the load resistance. Another performance measuring parameter we have defined is the current transfer ratio which is the ratio of the forward-current to the reverse leakage current. FIG. **11**B shows the current transfer ratio as a function of input power. The current transfer ratio at an input power of 1 μW (−30 dBm) for the hybrid scheme is 270, for the adaptive solid scheme is 140 and for the adaptive diode-connected one is 108. The current transfer ratio for the adaptive solid scheme increases rapidly and outperforms the hybrid scheme from 2 μW (−27 dBm) while for the adaptive diode-connected one, the current transfer ratio intersects the hybrid curve at 11 μW. The current transfer ratio for the hybrid scheme initially increases with the input power and saturates to approximately 1500 at 100 μW of input power. The current transfer ratio for the adaptive solid scheme rapidly increases with the input power. At an input power of 90 μW, the current transfer ratio for the adaptive solid scheme is 9000 as shown in FIG. **11**B. The current transfer ratio for the adaptive diode-connected scheme is 4000 at 97 μW of input power. The current transfer ratio for the adaptive-solid scheme at an input power of 100 μW (−10 dBm) is 6 times while for the diode-connected one is 2.5 times better than the hybrid scheme.

[0079] The proposed adaptive scheme is effective in increasing the PCE with the addition of auxiliary adaptive blocks. The increase in the PCE with the addition of the adaptive auxiliary blocks for the adaptive diode-connected scheme is shown in FIG. **12**A. With the addition of the auxiliary block the maximum PCE increases to 12%. The auxiliary block is added for every alternate transistor. However, the additional power consumed by the auxiliary diode-connected transistor when the rectifying device in the main rectification chain is forward-biased prevents the scheme from achieving higher PCEs. Unlike the adaptive diode-connected scheme, the auxiliary adaptive block is added to all the transistors for the adaptive solid scheme. Except for the first and the last stage, all the transistors for the solid adaptive scheme have the adaptive blocks. The additional power consumption and the parasitic capacitance introduced by the auxiliary chain are much lower than the former circuit as the solid-connected scheme uses only one transistor of width 480 nm per adaptive block. Hence the PCE is relatively higher at extremely low power levels. As seen from FIG. **12**B, with the addition of adaptive blocks, the efficiency increases. FIG. **12**B shows the improvement in PCE of the rectifier with the addition of auxiliary block. The maximum PCE reaches to 33.5% when the number of adaptive blocks is 20, a figure much higher than the adaptive diode-connected scheme which has a maximum PCE of 12%. The PCE and the output voltage performance at a relatively higher power level 0 dBm to −20 dBm is much better compared with only the hybrid scheme.

[0080] Impedance Matching Technique

[0081] The overall PCE of the RF energy harvesting system can be improved by increasing the passive voltage amplification due to the matching network, which increases the overall voltage at the input of the rectifier. The rectifier is modeled as $R_{rec}-jX_{rec}$ where R_{rec} is the real part of the rectifier's impedance and X_{rec} is the imaginary part. An impedance matching circuit is designed to match it to 50Ω as shown in FIG. **13**.

[0082] The input capacitance is mainly associated with the parasitic capacitance of the transistors and the layout. As the number of stages increases, the input resistance decreases while the parasitic capacitance increases. Increase in the value of load current also decreases the input resistance. The major factor contributing to increase in load current is lower load resistance. Other factors affecting the input impedance is the aspect ratio (W/L) of the transistors, input power level. Increase in the aspect ratio, increases the parasitic capacitance whereas decreases the input resistance. At the output side, the rectifier can be modeled by a voltage source and an output resistance. Thus the non-linear characteristic of the rectifier circuit makes it difficult to analyze them. The input impedance of the rectifier can be assumed to be constant only when small signal input is applied. In this case, the transient signal is a large signal with a capacitive component at the input causing a phase difference component ϕ to appear. The derivation in appendix though considers the reactive impedance still assumes the linear VI relationship. In real case, the VI relationship is non-linear and the input resistance is to be found using CAD tools.

[0083] Experimental Results—Hybrid Scheme

[0084] Three rectifiers, named as "efficiency", "voltage" and "1-stage PMOS doubler" are designed and fabricated in a 0.13 μm 8-metal CMOS process. The active die areas for efficiency circuit, voltage circuit and the 1-stage PMOS doubler test circuit are 230 μm×810 μm, 230 μm×1050 μm and 160 μm×70 μm respectively. The chip is wire-bonded onto PCB board with FR4 substrate and tested with Agilent MXG-N5181 signal generator at a frequency of 915 MHz using a single-tone sinusoidal signal. The receiver power is calculated by finding the average power at the input of the rectifier. The performance of the designed efficiency and voltage circuit is measured for a range of input power levels.

[0085] Output DC voltage was measured for different peak-to-peak input voltages. For a 1 MΩ load, an input voltage of 170 mV results in 2.4 V and 2.8 V for efficiency and voltage circuit, respectively. The voltage multiplication ratio (VMR) which is the ratio of DC voltage to the peak-to-peak input voltage is 14 and 17 for efficiency and voltage circuit respectively. A 220 mV signal results in 3.1 V for efficiency circuit and 4.0 V for voltage circuit. Similarly for a 5 MΩ load, an output of 2.7 V (VMR=16) for efficiency circuit and 3.0 V (VMR=18) for voltage circuit was measured at an input peak-to-peak voltage of 170 mV. Thus, voltage circuit which has a higher level of compensation than efficiency circuit has a lower input voltage requirement.

[0086] Harvested power was measured for different received powers. From the results, it was observed that at low power-level, the harvested power has a higher dependence on load current. For a 1 MΩ load at power levels greater than −30 dBm, the rate of decay in the harvested power curve is higher than the 5 MΩ load. The roll-off in the harvested power for a 1 MΩ load starts at about −20 dBm while the roll-off point for a 5 MΩ load is approximately −30 dBm. Even for voltage circuit, the rate of decay in the harvested power curve is greater for a 1 MΩ load compared to a 5 MΩ load. A lower load value has a higher current requirement. The performance of both the circuits in terms of roll-off is similar for the same load value. The designed efficiency and voltage circuit outperforms the circuit in Papotto et al. especially at low power levels. The efficiency circuit delivers an output power of 4.7 μW at an input power of −16.8 dBm (20.9 μW) when loaded by 1 MΩ. With a 5 MΩ load, the output power is 1 μW for an input power of −17.5 dBm (17.7 μW). An output power of 3.4 μW at an input power of −14.8 dBm (33.1 μW) for a 1 MΩ load is supplied by voltage circuit. A graph of the measurements was found to be in close agreement with the simulation results.

[0087] Power efficiency was measured for different received power levels. The measured and the simulated power conversion efficiency for efficiency and voltage circuit for different load resistance values are further compared. The efficiency comparison is done while de-embedding the input reflections in Papotto et al. The output DC voltage was also measured for the efficiency and voltage circuits for different load resistance values. The PCE is optimized for low power levels. When loaded by 1 MΩ, efficiency circuit attains a maximum measured PCE of 22.6% at −16.8 dBm (20.9 μW) while delivering 2.2 V to the output. A maximum measured PCE of 21.6% is obtained by efficiency circuit while producing an output voltage of 1.1 V at an input power of −26.5 dBm (2.23 μW) for a 5 MΩ load. At an input power level of −14.8 dBm (33.1 μW), voltage circuit achieves a maximum measured PCE of 10.2% for a 1 MΩ load while delivering 1.8 V. Due to a lower load current requirement for a 5 MΩ load, the output voltage is higher compared to a 1 MΩ load. At an input power of −22.5 dBm (5.6 μW), voltage circuit has a measured output DC voltage of 1 V while efficiency circuit has a measured output DC voltage of 1.8 V at −24 dBm (4 μW) for a 5 MΩ load.

[0088] Experimental Results—Adaptive Scheme

[0089] Three RF-DC power converters named as "adaptive solid," "adaptive diode-connected" and the "hybrid" circuits were designed and fabricated side by side in a 0.13 μm metal CMOS process with eight layers of metallization. The active die areas for adaptive solid and adaptive diode-connected is 0.25 mm^2 and hybrid circuit is 0.15 mm^2. The chip is wire-bonded onto a 2-layer FR-4 PCB board and tested with Agilent MXG-N5181 signal generator using frequency modulated continuous signal in the 902-928 MHz industrial, scientific and medical (ISM) band. An off-chip L-section impedance matching network is implemented on the PCB to convert the RF-DC power converter's input impedance to 50Ω. The output DC voltage was obtained with an oscilloscope or a digital multimeter. The measured and the simulated PCE for the adaptive and the hybrid scheme were compared for a load resistance of 1 MΩ at different input power levels. The output DC voltage was measured for the adaptive and the hybrid scheme for a load resistance of 1 MΩ. The simulation is performed at a frequency of 915 MHz which is the center frequency for the 902-928 MHz ISM band. The adaptive RF-DC power converters are designed to provide high PCE and a large output DC voltage for input power levels of 1 μW-100 μW (−30 dBm to −10 dBm). At larger input power, even with lower PCE the available output power is large hence designing for high input power levels is not so crucial. The adaptive solid scheme attains a maximum PCE of 32% at an input power of 32 μW (−15 dBm) with an output DC voltage of 3.2 V for a 1 MΩ load. For the input power of 32 μW and 1 MΩ load, the hybrid scheme has a PCE of 18% and delivers 2.6 V to the output. The adaptive diode-connected scheme has a maximum PCE of 11.3% at an input power of 118 μW while delivering 3.7 V to the output. The additional power consumed by the auxiliary diode-connected transistor when the rectifying device in the main rectification chain is forward-biased prevents the adaptive diode-connected scheme from achieving higher PCEs. The PCE was measured for different input power for the adaptive solid scheme at different load resistances. The output DC voltage for the adaptive solid scheme was also measured as a function of input power for different load resistances. As the load resistance decreases, the peak conversion efficiency curve shifts towards similar efficiencies at higher input power. The maximum measured PCE is 33.4% for a load resistance of 500 kΩ load at an input power of 83 μW. The PCE for a 500 kΩ load is larger than a 1 MΩ load for input power levels greater than 50 μW (−13 dBm). The peak power conversion efficiency curve is a function of the load resistance and can provide peak efficiency at a much lower power levels for larger load resistances. With decrease in load resistance, the circuit provides a smaller output voltage than with the high load resistance due to the low load current requirement at high load resistances. The hybrid scheme provides a larger output voltage compared to the adaptive schemes for input power levels lesser than 15 μW. The adaptive scheme outperforms the hybrid scheme once the power-up threshold-requirement is met. A DC output voltage of 3.2 V is obtained at an input power of 64 μW (−12 dBm) for a 500 kΩ load.

[0090] The harvested power was measured as a function of input power for the adaptive and the hybrid scheme. When loaded by 1 MΩ, an output power of 10 μW is harvested with an input power of 250 μW (−6 dBm) for the hybrid scheme while the adaptive solid scheme harvests 10 μW with only 30 μW (−15.3 dBm) of input power. With a load resistance of 500 kΩ, 10 μW of output power is harvested at an input power of 42 μW (−13.8 dBm). An output power of 20 μW is harvested at an input power of 64 μW (−12 dBm) for a 500 kΩ load using adaptive solid scheme. For a 1 MΩ load, at larger input power, the PCE degrades much more rapidly compared to a 500 kΩ load. For a 1 MΩ load, 20 μW of power is harvested by the adaptive solid scheme at 164 μW (−7.9 dBm). Even with an input power of few milliwatts, the hybrid scheme is not effective in harvesting 20 μW of output power.

[0091] The adaptive scheme is highly effective when the input power is above 10 μW (−20 dBm) and a large output voltage (more than 1.5 V) is desired. The performance of the adaptive solid scheme is similar to the hybrid scheme for power levels of 1-10 μW (−30 dBm to −20 dBm). The sensitivity of the RF-DC power converter for obtaining an output voltage of 1 V with a 1 MΩ load for the adaptive solid scheme is 8.9 μW (−20.5 dBm). The sensitivity of the circuit using the hybrid scheme is 6.9 μW (−21.6 dBm). The adaptive diode-connected scheme gives similar performance as the adaptive solid scheme at input power levels greater than 100 μW (−10 dBm). For low power applications, the adaptive solid scheme should be preferred over the adaptive diode-connected scheme.

[0092] The forgoing description pertains to circuitry using MOSFET enhancement-mode transistors. Nothing in the description should be taken to exclude using other types of transistors with corresponding changes to the circuitry that would be obvious to a person skilled in the art.

[0093] Immaterial modifications may be made to the embodiments described here without departing from what is covered by the claims.

[0094] In the claims, the word "comprising" is used in its inclusive sense and does not exclude other elements being present. The indefinite articles "a" and "an" before a claim feature do not exclude more than one of the feature being present. Each one of the individual features described here may be used in one or more embodiments and is not, by virtue only of being described here, to be construed as essential to all embodiments as defined by the claims.

APPENDIX

[0095] In section II, the average input power was suggested consisting of real and imaginary parts.

$$V = V_{in}\sin\omega t,\ I = I_{in}(\sin\omega t - \phi)$$

$$P_{mean} = \frac{1}{T_2 - T_1}\int_{T_1}^{T_2} V_{in}\sin\omega t \cdot I_{in}(\sin\omega t - \phi)dt$$

$$Z_{in} = \frac{V_{in}^2}{2P_{mean}} \cdot \cos\phi + \frac{V_{in}^2}{4P_{mean}} \cdot \frac{\sin[2\omega(T_2 - T_1) - \phi]}{\omega(T_2 - T_1)}$$

[0096] Z_{in} consists of real and imaginary component. Equating the imaginary component to zero by considering the phase difference $\phi=0$.

$$R_{in} = \frac{V_{in}^2}{2P_{mean}},\ C_{in} = \frac{4P_{mean}(T_2 - T_1)}{V_{in}^2\sin[2\omega(T_2 - T_1) - \phi]}$$

1. A power conversion circuit comprising:

a first input line and a second input line, the first and second input lines configured to receive an alternating voltage differential between the first and second input lines;

a multi-stage rectifier comprising transistors arranged in series, each transistor having a gate, a source and a drain, adjacent transistors of the series being connected so that for adjacent p-type transistors the drain of the left p-type transistor is connected to the right adjacent source of the right p-type transistor and source of the right p-type transistor is connected to the left adjacent drain of the left p-type transistor, for adjacent n-type transistors the source of the left n-type transistor is connected to the right adjacent drain of the right n-type transistor and drain of the right n-type transistor is connected to the left adjacent source of the left n-type transistor, for a p-type transistor adjacent to an n-type transistor the source of the right p-type transistor is connected to the source of the left adjacent n-type transistor and drain of the left p-type transistor is connected to the drain of the right adjacent n-type transistor to form a junction, each junction being connected to one of the first input line and the second input line via a capacitor, with adjacent junctions having one junction of the adjacent junctions connected to the first input line and the other junction of the adjacent junctions connected to the second input line;

the gate of each transistor of the multi-stage rectifier being connected to a respective junction that is not a junction formed by the connection of that transistor to an adjacent transistor but a junction in the previous stage for the p-type transistor or later stage for the n-type transistor, at least one of the transistor in the multi-stage rectifier being p-type and at least one being n-type.

2. The power conversion circuit of claim **1** further comprising an auxiliary chain of p-type transistors, each auxiliary transistor having a gate, a source and a drain, each transistor of the auxiliary chain being connected to a p-type transistor of the multi-stage rectifier in the main chain, so that the gate of the respective auxiliary transistor is connected to the source of the respective transistor in the main chain, the source of the respective auxiliary transistor is connected to the gate of the transistor in the main chain and also connected to the previous stage junction of the main chain which is N transistors away from the respective transistor, and the drain of the respective auxiliary transistor is connected to the drain of the respective transistor or to the junction of the later stage transistor in the main chain.

3. The power conversion circuit of claim **1** further comprising an auxiliary chain of n-type transistors, each auxiliary transistor having a gate, a source and a drain, each transistor of the auxiliary chain being connected to an n-type transistor of the multi-stage rectifier in the main chain, so that the gate of the respective auxiliary transistor is connected to the source of the respective transistor in the main chain, the source of the respective auxiliary transistor is connected to the gate of the transistor in the main chain and also connected to the later stage junction of the main chain which is N transistors away from the respective transistor, and the drain of the respective auxiliary transistor is connected to the drain of the respective transistor or to the junction of the previous stage transistor in the main chain.

4. An auxiliary chain of transistors, each auxiliary transistor having a gate, a source and a drain, each transistor of the auxiliary chain being connected to a respective transistor of a main chain of a multi-stage rectifier in the main chain, so that the gate of the respective auxiliary transistor is connected to the source of the respective transistor in the main chain, the source of the respective auxiliary transistor is connected to the gate of the transistor in the main chain and also connected to the later stage junction of the main chain which is N transistors away from the respective transistor in the case that the respective transistor is n-type, or to the earlier stage junction of the main chain which is N transistors away from the respective transistor in the case that the respective transistor is p-type, and the drain of the respective auxiliary transistor is connected to the drain of the respective transistor or to the junction of a previous stage transistor in the main chain, in the case that the respective transistor is n-type, or to the junction of a later stage transistor in the main chain, in the case that the respective transistor is p-type.

5. The power conversion circuit of claim **4** in which the transistors of the main chain and auxiliary chain are n-type and further comprising a p-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of later stages, such that the gate of the additional p-type auxiliary transistor is connected to the gate of the n-type auxiliary transistor and the source of the n-type and p-type auxiliary transistors are connected, and the drain of the p-type auxiliary transistors is connected to junction of the main chain which is at 'N' later stages from the respective transistor.

6. The power conversion circuit of claim **4** in which the transistors of the main chain and auxiliary chain are p-type and further comprising an n-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of earlier stages, such that the gate of the additional n-type auxiliary transistor is connected to the gate of the p-type auxiliary transistor and the source of the n-type and p-type auxiliary transistors are connected, and the drain of the n-type auxiliary transistors is connected to junction of the main chain which is at 'N' earlier stages from the respective transistor.

7. The power conversion circuit of claim **2** further comprising a diode connected transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of previous stages which is N transistors away from the respective transistor.

8. The power conversion circuit of claim **3** further comprising a diode connected transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of later stages which is N transistors away from the respective transistor.

9. The power conversion circuit of claim **4** further comprising a diode connected p-type or n-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of later stages which is N transistors away from the respective transistor.

10. The power conversion circuit of claim **4** further comprising a diode connected p-type or n-type transistor on a line connecting the gate of each transistor in the main chain of the multistage rectifier to the junction of earlier stages which is N transistors away from the respective transistor.

11. The power conversion circuit of claim **1** further comprising an additional transistor connected in series with the multistage rectifier, the additional transistor having a gate, a source and a drain, the gate of the additional transistor being connected to the drain of the additional transistor.

12. The power conversion circuit of claim **1** in which the gate of each transistor in the main chain is connected to the junction of previous stages for p-type transistor and later stages for n-type transistor.

13. The power conversion circuit of claim **12** in which N of the transistors in the multistage rectifier arranged in series are n-type and all but N of the transistors in multistage rectifier arranged in series are p-type.

14. The power conversion circuit of claim **1** in which the second input line is grounded.

15. The power conversion circuit of claim **1** in which the body terminal of each p-type transistor is connected to the respective drain and the body terminal of each n-type transistor is either grounded or connected to the respective drain terminal.

* * * * *

March 25, 1969 W. C. BROWN ET AL 3,434,678

MICROWAVE TO DC CONVERTER

Filed May 5, 1965

Sheet 1 of 3

FIG. 1

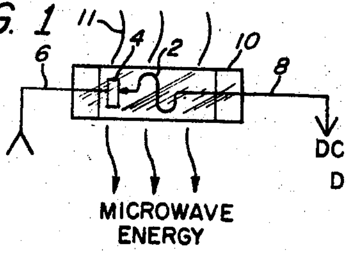
11
4
2
10
6
8
DC
MICROWAVE ENERGY

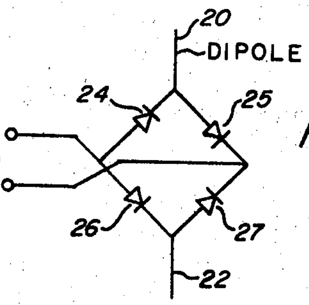
20
DIPOLE
24
25
DC LOAD
26
27
22

FIG. 2

FIG. 3

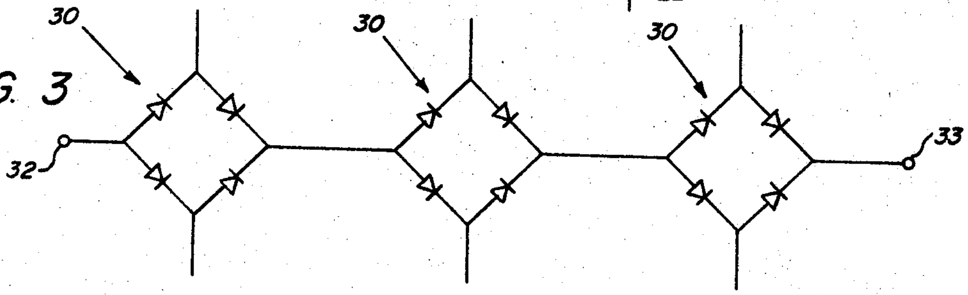
30
30
30
32
33

FIG. 4

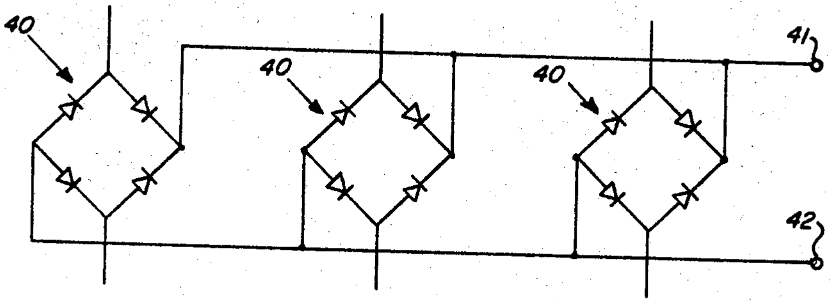
40
40
40
41
42

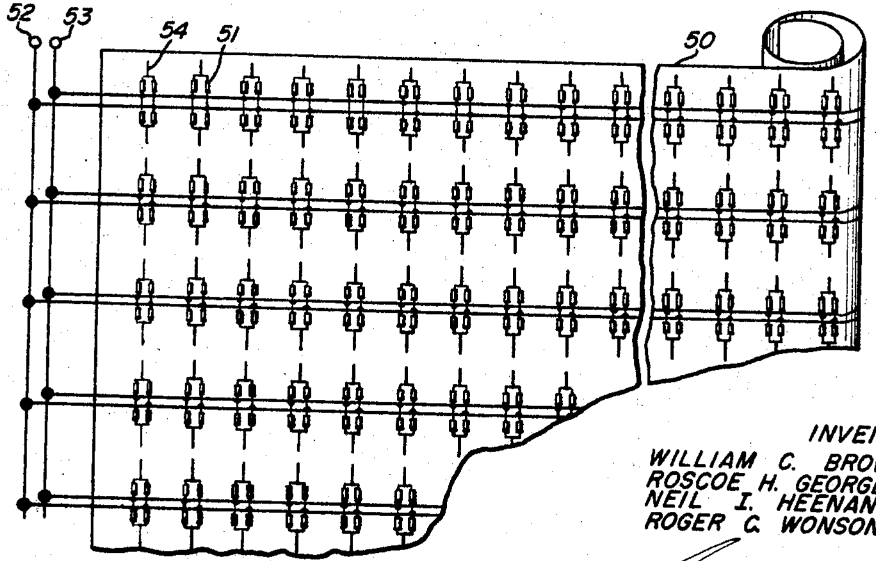
52
53
54
51
50

FIG. 5

INVENTORS
WILLIAM C. BROWN
ROSCOE H. GEORGE
NEIL I. HEENAN
ROGER C. WONSON

BY Edgar O. Rost
ATTORNEY

March 25, 1969 W. C. BROWN ET AL 3,434,678

MICROWAVE TO DC CONVERTER

Filed May 5, 1965

Sheet 2 of 3

FIG. 7

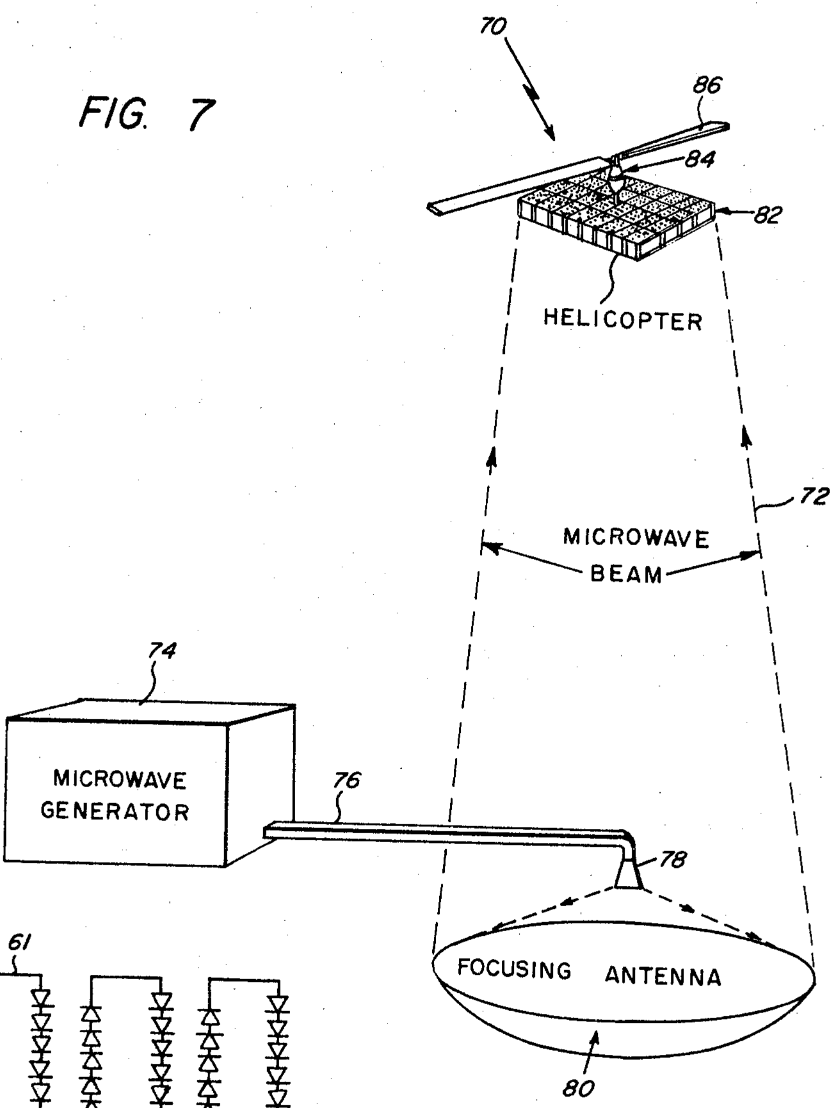
70
86
84
82
HELICOPTER
72
MICROWAVE
BEAM
74
MICROWAVE
GENERATOR
76
78
FOCUSING ANTENNA
80

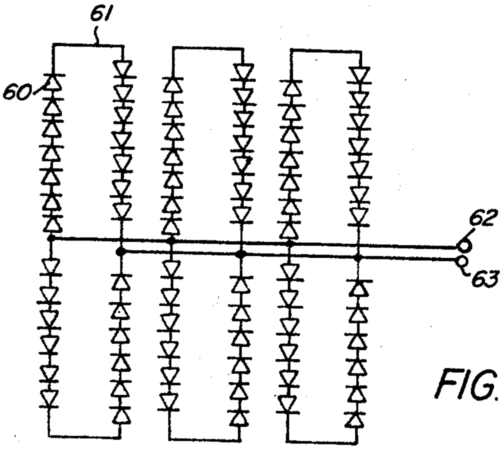

FIG. 6

INVENTORS
WILLIAM C. BROWN
ROSCOE H. GEORGE
NEIL I. HEENAN
ROGER C. WONSON

BY Edgar O. Rost

ATTORNEY

March 25, 1969 W. C. BROWN ET AL **3,434,678**

MICROWAVE TO DC CONVERTER

Filed May 5, 1965

Sheet 3 of 3

FIG. 8

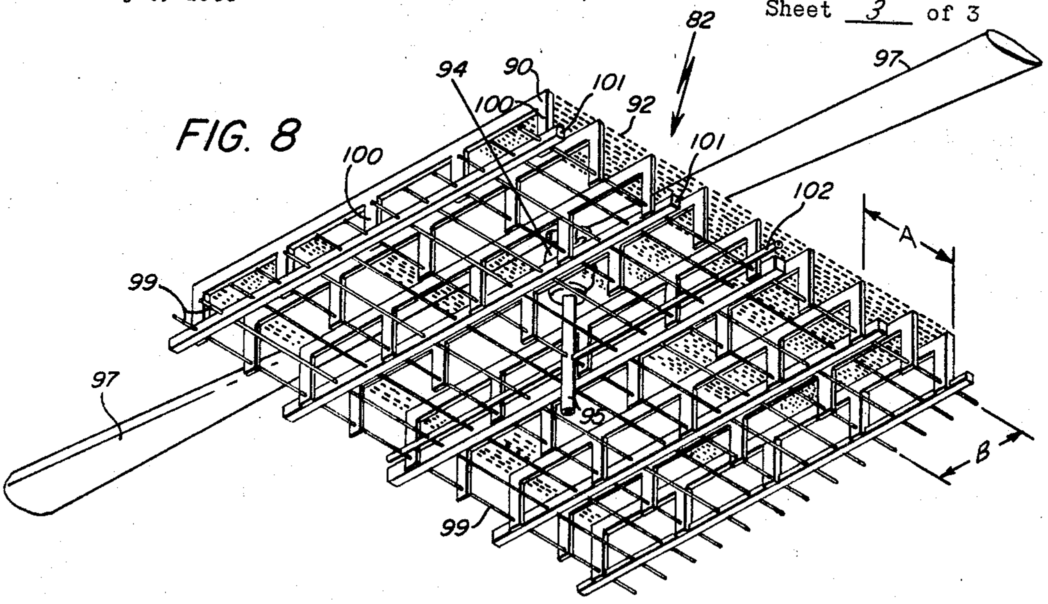

82
94
90
101
100
92
101
100
97
102
A
99
97
95
B
99

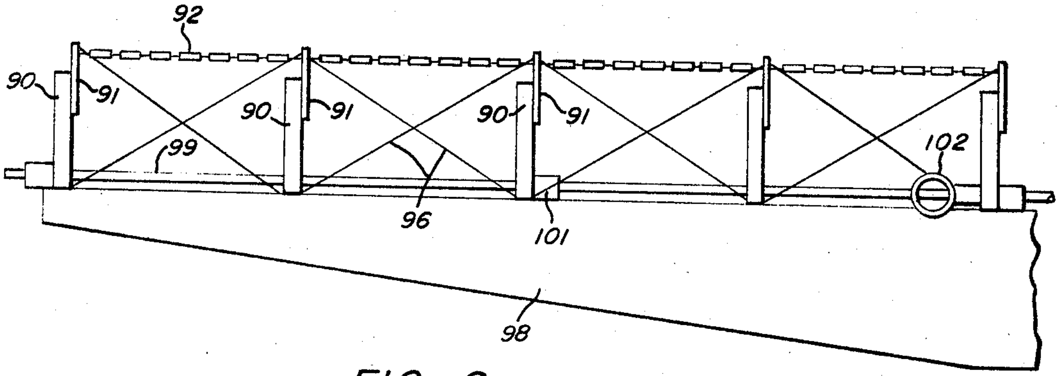

92
90
91
90
91
90
91
99
102
96
101
98

FIG. 9

INVENTORS
WILLIAM C. BROWN
ROSCOE H. GEORGE
NEIL I. HEENAN
ROGER C. WONSON

BY Edgar O. Rost
ATTORNEY

United States Patent Office 3,434,678
Patented Mar. 25, 1969

3,434,678

MICROWAVE TO DC CONVERTER

William C. Brown, Weston, Mass., Roscoe H. George, West Lafayette, Ind., and Neil I. Heenan, Needham, and Roger C. Wonson, Beverly, Mass., assignors to Raytheon Company, Lexington, Mass., a corporation of Delaware

Filed May 5, 1965, Ser. No. 453,415
Int. Cl. B64g *1/00*
U.S. Cl. 244—1 **3 Claims**

ABSTRACT OF THE DISCLOSURE

A combined antenna and conversion mechanism for reception of beamed high frequency electromagnetic energy in space including a large array of unidirectional current semiconductor rectifier devices. A self-supporting space vehicle utilizing the rectified DC electrical energy for propulsion is disclosed in an illustrative embodiment.

The present invention relates in general to the transfer of energy by means of an electromagnetic wave beam and more particularly to interception and rectification of such energy into low frequency electrical DC energy with a high degree of efficiency.

Improved technology in the field of microwave energy generation at superpower levels has resulted in the realization of electrical energy transmission over considerable distances for remote energization of devices or vehicles without the aid of wires. The transmission of microwave electromagnetic energy into space has been commonly employed in the radar pulse echo systems for the detection and orientation of desired objects within a predetermined scanning range of a transmitting antenna. Beams of a similar nature may now be employed for other useful purposes and the advantages attendant the utilization of electromagnetic energy in the microwave region in contrast with other wavelengths may now be enumerated.

Microwaves have been generally defined as high frequency radio waves whose wavelength is less than 30 centimeters, with a lower wavelength limit on the order of 1 millimeter sometimes being applied to what is commonly referred to as the "microwave region." The superiority of high frequency microwaves is due in part to the fact that it is generally desirable to focus the transmitted energy so as to achieve a high power density at a remote point or area with respect to a given power source. In accordance with the laws of optics, the sharpness of the microwave beam produced by a transmitting antenna varies as the ratio of antenna dimensions to the wavelength of the transmitted energy. Therefore, for a given or desired power density or beam sharpness, a decrease in the wavelength of the transmitted energy permits a corresponding decrease in the dimensions of the antenna. From the standpoint of mechanical considerations, it is desirable to employ small antennas and other components, and it is therefore advantageous to employ high frequency energy of very short wavelength. In addition, the difficulties encountered in long wave transmission as a result of natural and man-made interference or noise do not occur with any appreciable significance at microwave frequencies. Further, in aerospace applications with considerable distances separating the transmitter at an earth or mother planet location and the employment of shorter wavelength beamed energy is preferred since longer wave signals will generallly be reflected at certain altitudes by reflecting layers in the atmosphere.

In view of certain losses due to absorption which may occur in the atmosphere, microwaves in the region having the approximate bounds of 2 and 30 centimeters are readily adaptable to the convenient radiation of power to remote points without the utilization of wires. The preferred wavelengths are of the order of 5 or 10 centimeters to provide efficient focusing with existing transmitting antenna systems which may be maintained at a reasonable size. An illustrative device of the superpower high frequency microwave generators operative in the desired band is the so-called Amplitron which is an amplifier having a broad bandwidth and excellent performance characteristics for the focusing of the beam. Such devices are capable of producing 15 or 20 kilowatts of average continuous wave power in the neighborhood of 10 centimeters in wavelength with capabilities expected in the region of 500 kilowatts or more average power with 50 megawatts peak power. A complete description of such devices may be had by referring to Patent No. 2,933,723 issued Apr. 19, 1960 to William C. Brown and assigned to the assignee of the present invention.

With microwave energy capable of being generated and directed over longer distances conversion of such high frequency electromagnetic energy is of paramount concern. One conversion mechanism in the prior art involves direct conversion of such energy into heat which may then be utilized directly or indirectly for propulsion or generation of flight-producing forces. Examples of such devices for heat energy exchange as well as space vehicles utilizing such energy may be noted in Patent No. 3,174,705, issued Mar. 23, 1965, to D. Schiff et al., as well as U.S. Letters Patent No. 3,083,528, issued Apr. 2, 1963 and No. 3,114,517, issued Dec. 17, 1963, to William C. Brown. The heat exchanger method of conversion of electromagnetic energy into useful power is limited by the overall efficiencies of approximately 25 percent in the conversion of heat into mechanical or electrical work. Desirable, therefore, would be the direct rectification of the high frequency electromagnetic energy into low frequency electrical energy for the operation of many useful aerospace devices as well as systems.

The present invention has for its primary object the conversion of high frequency electromagnetic energy in the microwave region directly into low frequency electrical energy.

A further object of the present invention is the provision of a combined nondirectional receiving antenna and microwave electromagnetic energy to low frequency electrical energy conversion means in a unitary structure.

A still further object of the present invention is a provision of a new and novel combined nondirectional receiving antenna and microwave to DC energy converter for aerospace applications.

Another object of the present invention is the provision of a new and novel nondirectional receiving antenna and microwave to DC energy converter having a high degree of efficiency.

Still another object of the present invention is the provision of a new and novel aerospace vehicle with nondirectional receiving antenna and microwave to DC energy converter means with said vehicle being capable of being supported by its own energy generation means at a distance spaced apart from the power generation means.

In accordance with the teachings of the present invention, the above and other objects are achieved by the employment of efficient unidirectional microwave power rectifiers and dipole antenna means. Such rectifying devices, while being individually limited in power-handling capabilities, normally in the order of fractions of watts, have been found to be highly efficient means for the rectification of microwave power when assembled in large numbers in various arrays. It is interesting to note that the observed collective efficiency was on the order of 40 to 70 percent. In an illustrative embodiment, point-

contact semiconductor diodes were arranged in four arm bridge connected networks with the networks interconnected in various configurations such as series, parallel and series-parallel.

In discussing aerospace applications, an additional problem is encountered in the beaming of microwave energy to a remote point and the interception and utilization of such electrical energy. In such applications the advantages of a vehicle which may be maintained in space for indeterminate periods of time without employing a local fuel source are readily apparent. Such devices could readily provide communication networks, surveillance functions using radar techniques along with numerous other functions. The acapture of the beamed high frequency electromagnetic energy raises the need for an efficient antenna means capable of intersecting the beam at high altitudes. Conventional techniques employed in microwave radar usage such as receiving antenna horns are capable of intersecting only a small portion of the beam energy and add considerable weight in applications involving heavier-than-air vehicles. In an exemplary embodiment of the invention a space vehicle, namely a helicopter, is disclosed for either moving flight or a stationary location with self-supporting electrically operative propulsion means. The semi-conductor diode rectifier arrays have been demonstrated to fulfill the receiving antenna functions as well as the electrical energy rectification means in a highly efficient manner. Such combined antenna and rectifier means has also assisted in reduction of the weight problem in airborne devices. Further, it has provided a nondirectional means for the interception of the microwave energy to thereby reduce the problems of focusing inherent in prior art directional horn type receiving antennas.

With the above features, advantages and objects in mind the invention will now be described by reference to the following detailed description together with the accompanying drawings in which:

FIG. 1 is a perspective view of an illustrative diode rectifier;

FIG. 2 is a schematic circuit diagram of a bridge connected diode network with dipole antenna means;

FIG. 3 is a schematic circuit diagram of a plurality of bridge connected networks arranged in series;

FIG. 4 is a schematic circuit diagram of a parallel bridge connected network array;

FIG. 5 is a perspective view of an illustrative embodiment of a combined antenna and rectifier array in a folded or rolled up configuration;

FIG. 6 is a schematic circuit diagram illustrating the bridge connected diode array incorporated in the aerospace vehicle shown in FIG. 7;

FIG. 7 is a schematic representation in elevation illustrative of a heavier-than-air aerospace vehicle incorparating the structure of the present invention;

FIG. 8 is a perspective view of the aerospace vehicle embodiment as viewed from the under portion thereof; and

FIG. 9 is an enlarged partial view in elevation of a portion of the illustrative embodiment shown in FIG. 8.

FIG. 1 illustrates a point-contact semiconductor diode rectifier of the type employed in radar microwave receiver apparatus to rectify returned radar pulses. Any of the high burnout semiconductor diodes having high rectification characteristics are preferred and are commercially available, such as the 1N82 or 1N830. The rectifying junction is formed by whisker element **2** contacting the semiconductor element **4** respectively connected to leads **6** and **8**. Silicon is preferred over germanium for element **4** because of its ability to operate at higher temperatures and thereby handle higher powers. Envelope **10** houses the rectifying elements and may be of a hermetically sealed dielectric material or combination metal and ceramic composition. The inherent characteristic of such diode rectifiers is that the microwave energy is intercepted and rectified in a unidirectional manner and the line **11** indicate pictorially the rays of the beamed electromagnetic microwave energy in a plane normal to the envelope. In FIG. 2 a full-wave bridge connected diode network is illustrated with the forward direction of the rectified DC electric current indicated by the direction of the arrow symbols. The network shown consists of half-wave dipoles **20** and **22** each terminated with a diode rectifier element **24** to **27** in an arm of the bridge connected network. The dipole elements **20** and **22** are of the half-wave configuration and may be spaced apart from each other a one-half wavelength at the frequency of the beamed electromagnetic energy.

Referring now to FIG. 3, an array of bridge connected diode networks each with the half-wave dipoles are shown connected in series. Each network is referred to by the numeral **30** and is similar in the bridge connections to the single element network shown in FIG. 2. The DC output of the collective rectified energy is coupled by means of terminals **32** and **33**. In FIG. 4, a similar number of individual bridge connected diode-dipole networks are shown connected in a parallel array. Each network is indicated by the numeral **40**, and the output terminals are indicated as **41** and **42**.

Any number of diode-dipole networks may be provided and in FIG. 5 such a multi-element array is illustrated by mounting on a flexible material **50** which may be rolled or folded into any desired package or enclosed within a capsule to be launched and released at a predetermined point in space. Any flexible material which is pervious to electromagnetic energy is preferred. The total power desired would be the determining factor in a number of individual diode-dipole elements required. In this embodiment, the bridge connected networks **51** are connected in parallel to the output load indicated by terminals **52** and **53**, and representative measurements of electrical characteristics have shown that approximately five watts of DC electrical energy is realizable for each square foot of area of the combined antenna-rectifier. While the dipole elements **54** have been indicated in a particular array, it is within the scope of the invention to stagger the placement of such dipoles to increase the over-all efficiency of the antenna-rectifier.

To further increase the DC powder output, the full-wave bridge connected networks are preferably arranged with a plurality of diodes in series in each arm of the bridge. An illustrative schematic circuit diagram of such a configuration is shown in FIG. 6 wherein seven diodes **60** are shown in each arm of the bridge circuit and are connected in series for a total of twenty-eight diodes in each bridge network. The dipole members will then be the substantially U-shaped end portions **61** at the ends of each brace of seven diodes. In the illustration three such twenty-eight diode bridge networks are shown connected in parallel to terminals **62** and **63**. This closer spacing and compact arrangement has been shown to be a source of improved power output and is capable of a high degree of reliability through the redundant nature of the parallel series connections within each bridge network. If one of the diode rectifiers fails to function the over-all voltage drop across this element would be divided among the six remaining diode rectifiers. If any of the connecting wires between the diode elements should break, the adjacent arms of the other bridge assemblies would take the additional load due to the close proximity of the respective arms to each other. In addition, it is possible to have a number of open connections or inoperative diodes dispersed throughout the array without any serious impairment in performance.

In relation to the array concept to be hereinafter described it may be stated that within a six inch square area ten such individual bridge networks each containing twenty-eight diode rectifiers for a total of 280 diode rectifiers may be deployed in such a manner as to provide

maximum exposed area for each diode as well as the connecting leads. Such an arrangement will be hereinafter referred to as a "module" and a DC output in excess of fourteen watts has been measured for such a module. Any number of such modules could be connected provided for a desired power yield and this module concept readily lends itself to use in certain aerospace applications now to be described.

In FIG. 7 a propelled type of space vehicle **70** is shown wholly supported by means of the transfer and rectification of continuous wave electromagnetic energy via a microwave beam **72**. The source of the microwave energy which may be of the Amplitron type device as described in the aforementioned issued Patent No. 2,933,723 is indicated as **74**. This energy is fed by waveguide means **76** to a transmitting horn **78** to illuminate an ellipsoidal beam forming focusing antenna **80** for the transmission of the microwave beam **72**. It will be appreciated by those skilled in the art that the representations of the microwave generation and transmitting antenna means are pictorial representations to illustrate the usage of the invention in diagrammatic form and the present invention is not limited to any particular source of microwave energy or transmitting antenna assembly. It may be stated the reflector of the antenna assembly is considerably larger than most of the reflectors of the prior art in order to focus a large amount of the microwave power at high altitudes for use in the transfer of energy to space vehicles. Such antenna assemblies may be partially supported in a large hollowed area on the earth's surface or other convenient means of support.

The space vehicle or helicopter **70** can be described as a main body member supporting antenna-rectifier means **82** including a large number of the so-called modules connected together and rigidly supported in a planar parallel array. A motor **84** is supported by the combined body member and the receiving antenna-rectifier means and actuates the rotor **86** of conventional design employed in such self-propelled hovering vehicles. The disclosed vehicle provides for the illumination of the planar array of the semiconductor diode dipole elements by the microwave beam and the direct conversion of the microwave power transmitted by the beam into usable electrical energy for the self-propulsion of the device without any local fuel supply being required.

FIGS. 8 and 9 illustrate a space vehicle **82** comprising a plurality of the combined receiving antenna-rectifier module means for interception and rectification of the electromagnetic microwave energy beam emanating from an earth or mother planet source. A planar array of the antenna-rectifier modules is mechanically supported by means of structural members **90** of any lightweight wood or metal. Insulators **91** positioned coextensive with the members **90** support the diode rectifier array and avoid interference with the receiving and electrical performance characteristics by the structural support members. Carrying forward the module concept of 280 diode rectifiers to provide an approximate power output of 14 watts, it was noted that any number of such modules may be coupled together since the individual module outputs are relatively insensitive to a wide range of load resistances connected to the common output terminals. To achieve the desired electrical output of approximately 120 volts and 250 watts of power, subgroups of four modules each were assembled and parallel connected with an approximate 30 volts available for each subgroup. Four such subgroups were series-connected to result in a total of 4,480 diode rectifiers or 16 modules assembled in a two foot square self-supporting planar parallel array structure. The individual diode rectifiers connected in each arm of the bridge network are indicated by the numeral **92**. An exemplary module configuration would extend within the area delineated by the dotted lines and reference letter A on one side and similar dotted lines and reference letter B on the other side.

A motor **94** is connected to the DC side of the overall array and may be additionally supported by tubular member **95**. A shaft and propulsion means consisting of rotor blades **97** provide for the upward lift of the overall vehicle for the self-supporting of same in space applications. Additional structural support such as interlaced rigging **96** of a high tensile strength material such as nylon or steel wire, as well as bracing member **98**, may be employed for strengthening of the body means to withstand the vibrational forces and downwash from the propulsion means.

In accordance with the well known technology of microwave transmission the combined array of diode rectifiers and propulsion means presents a specific load impedance which must be suitably matched to the transmitted microwave energy beam to result in maximum efficiency. In aerospace applications a mismatch of approximately ten to one may be evident. Matching of the load impedance to a value of approximately 377 ohms as the free space value will be provided by a plurality of coplanar parallel metallic rod members **99** disposed in a grating array in front of the diode rectifiers a predetermined distance. Rod members **99** are linearly disposed and extend in a similar direction as the assembled diode rectifiers. A selected frontal spacing of one-quarter of the wavelength of the microwave frequency being transmitted has been experimentally determined to be suitable for impedance matching purposes. An approximate spacing of two inches between the respective members was preferred for a selected microwave frequency of 2,450 megacycles. Each of the members **90** are provided with lateral sections **100** to support the elongated bar members **101** which in turn maintain the rod members **99** in the desired position. A tubular member **102** of a lightweight metal may also be provided to combine with the motor support member **95** for structural support.

The combined antenna-rectifier array provides a source of electrical energy to render any space vehicle self-supporting. The diode rectifier elements when assembled in the antenna array have been found to be nondirectional with respect to interception of the beamed microwave energy. This represents a large step forward in the utilization of high power microwave energy over the prior art horn-type receiving antennas which must be accurately focused and pointed in a particular direction for the reception of any energy. The connections between the respective members of the diode rectifier array and deployment in the parallel configuration serves to provide maximum exposed area. Such connections and in particular the end loop portions adjacent the terminus of each arm of the bridge networks serves as an efficient dipole for the interception of the microwave energy.

Although it is not intended as a full explanation of the high degree of efficiency attained with the disclosed antenna-rectifier array, it is believed that the whisker elements within the semiconductor diodes themselves are a contributing factor and may function as additional dipole elements. The disclosed embodiment functioned efficiently when illuminated by microwave energy generating a vertically polarized beam. Hence, an efficient and light weight energy conversion apparatus is disclosed which may be self-supporting without the requirement of a large local fuel supply payload.

It may be within the purview of the invention to use the available rectified electrical energy for performing many functions in addition to the actuation of the propulsion means. Hence, communications' payloads may be maintained at predetermined positions in space in a hovering attitude utilizing a portion of the electrical energy available. Relay signals to other such vehicles or return signals to ground stations would then be within the realm of possibility. Such available energy may also be employed for servomechanisms, stabilizing and countertorque systems for the navigation of such vehicles.

The electrical efficiencies realized with the combined

receiving antenna and rectifier means have also provided certain weight advantages over other energy converters in aerospace applications. Examples of such converters would be heat exchangers or solar cells. In comparison to the present invention where five to eight pounds per kilowatt of energy realized is a normal characteristic, other energy conversion means weigh in the vicinity of 150 pounds per kilowatt of realizable energy. The inherent advantages of the present invention are therefore apparent. While the technology in the diode rectifier art is being continually advanced, new diode power rectifiers as well as integrated circuit techniques are readily available to future configurations of the present invention. The so-called Schottky barrier diodes could be employed to produce combined antenna-rectifier means weighing even less than two pounds per kilowatt of available energy.

Although the foregoing detailed description has referred to DC power rectification it will be evident that with suitable circuit components low frequency AC energy may also be made available. In addition, other propulsion means may be readily substituted using electrical energy. The embodiments disclosed herein are illustrative only and other modifications or alterations will be apparent to those skilled in the art which do not depart from the scope of the broadest aspects of the present invention as defined in the appended claims.

What is claimed is:

1. A space vehicle comprising:
body means;
said body means including spaced structural support members;
combined antenna and DC electrical energy rectification means for the interception and rectification of incident high frequency electromagnetic microwave energy carried by said support members in a planar parallel array;
said rectification means comprising a plurality of four

arm full-wave bridge connected rectifier circuit networks each having a plurality of unidirectional semiconductors in each arm;
said networks being electrically interconnected to common output terminals;
electrically operable propulsion means comprising a motor and rotor members carried by said body means and connected to said terminals for the utilization of said rectified DC energy; and
means for matching the load impedance of said combined antenna and electrical energy rectification means to the incident microwave energy.

2. A space vehicle according to claim 1 wherein said load impedance matching means are arranged in a coplanar array coextensive with said antenna and energy rectification means array, and spaced therefrom a distance of approximately one-quarter of a wavelength at the frequency of the microwave energy.

3. A space vehicle according to claim 2 wherein said load impedance matching means comprise a plurality of parallel disposed elongated metallic members.

References Cited

UNITED STATES PATENTS

1,217,149	2/1917	Caldwell	321—27 X
2,165,055	7/1939	Kafka	321—27
2,444,458	7/1948	Master	321—8
2,927,321	3/1960	Harris	343—6.8 X
3,098,971	7/1963	Richardson	325—592 X
3,174,705	3/1965	Schiff et al.	244—1

RODNEY D. BENNETT, *Primary Examiner.*

MALCOMB F. HUBLER, *Assistant Examiner.*

U.S. Cl. X.R.

307—151; 318—16; 321—27; 325—494; 343—100

United States Patent [19]

Carroll

[11] **Patent Number: 4,857,893**

[45] **Date of Patent: Aug. 15, 1989**

[54] **SINGLE CHIP TRANSPONDER DEVICE**

[75] Inventor: **Gary T. Carroll,** Boulder, Colo.

[73] Assignee: **Bi Inc.,** Boulder, Colo.

[21] Appl. No.: **153,237**

[22] Filed: **Feb. 8, 1988**

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 888,246, Jul. 18, 1986.

[51] **Int. Cl.⁴** **G08B 13/14**

[52] **U.S. Cl.** **340/572;** 340/825.54

[58] **Field of Search** 340/572, 825.54, 539, 340/571; 342/44, 51

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,465,724	9/1969	Broadbent	340/572
4,242,671	12/1980	Plows	340/572
4,260,983	4/1981	Falck et al.	340/572
4,333,072	6/1982	Beigel	340/825.54
4,463,353	7/1984	Kuzara	340/825.54
4,573,042	2/1986	Boyd et al.	340/539
4,630,044	12/1986	Polzer	340/825.54
4,646,090	2/1987	Mawhinney	340/572
4,656,463	4/1987	Anders et al.	340/572
4,724,427	2/1988	Carroll	340/572

OTHER PUBLICATIONS

Saleh et al., "Permalloy Thin-Film Inductors", Electronics Letters, vol. 6, No. 26, pp. 850-852, (Dec. 31, 1970).

Soohoo, "Magnetic Thin Film Inductors for Integrated Circuit Applications", IEEE Trans. on Magnetics, vol. Mag-15, No. 6, (Nov. 1979).

Primary Examiner—Joseph A. Orsino
Assistant Examiner—Jill D. Jackson
Attorney, Agent, or Firm—Fitch, Even, Tabin & Flannery

[57] **ABSTRACT**

A transponder device receives a carrier signal from an interrogator unit. This carrier signal, of frequeny F, is rectified by a rectifying circuit in order to generate operating power. Logic/timing circuits derive a clock signal and second carrier signal of frequency F/n from the received carrier signal. This clock signal reads a unique identifying data word from a programmable read only memory (PROM). The data word is encoded and mixed with the carrier signal in a balanced modulator circuit. The output of the balanced modulator is transmitted to the interrogator unit where it is decoded and used as an identifying signal. The identifying signal identifies the particular transponder device from which it originated. The rectifier and balanced modulator circuits are realized from the same diode elements. All electrical circuits of the transponder device are realized on the same monolithic semiconductor chip. In one embodiment, an antenna receiving/transmitting coil is also part of the chip, being placed around the periphery thereof. In alternative embodiments, various hybrid elements may be connected to the monolithic elements in order to realize additional functions, such as adjustable tuning of the receiving circuit, independent crystal frequency control, and battery-powered operation.

14 Claims, 10 Drawing Sheets

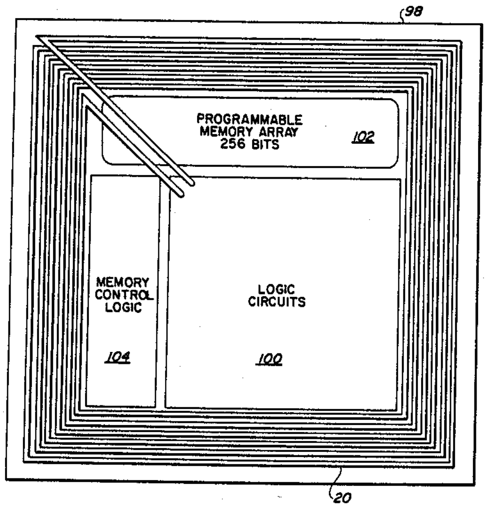

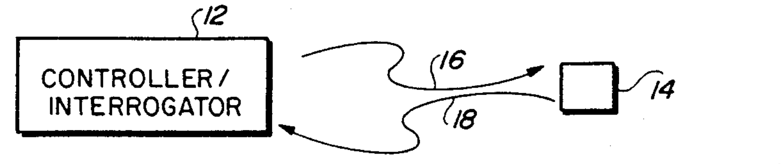
12
CONTROLLER / INTERROGATOR
16
18
14

FIG. 1

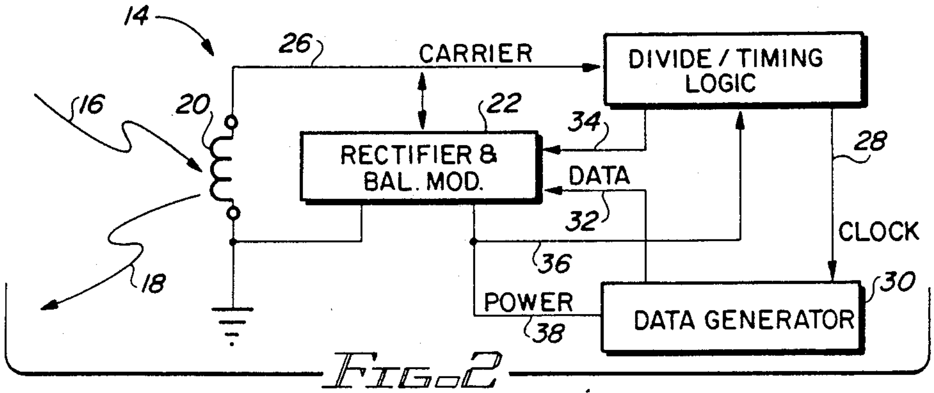
14
26
CARRIER
DIVIDE / TIMING LOGIC
16
20
22
RECTIFIER & BAL. MOD.
34
28
DATA
32
CLOCK
36
18
30
POWER
38
DATA GENERATOR

FIG. 2

FIG. 3

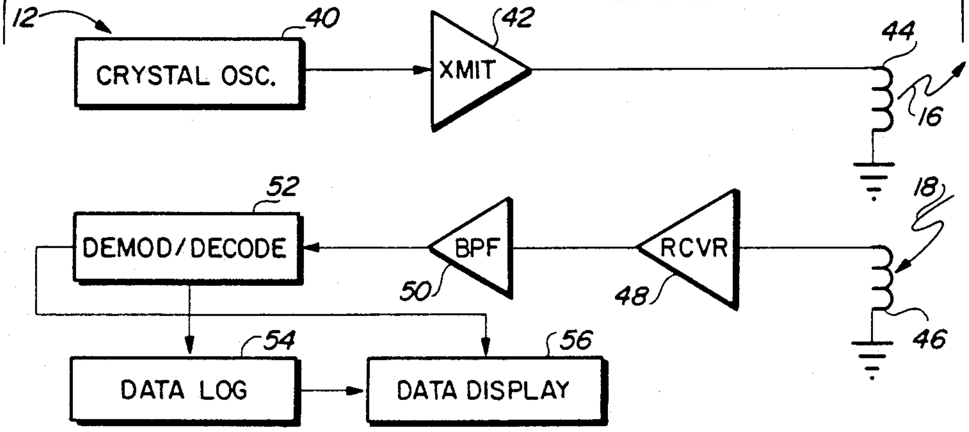
12
40
CRYSTAL OSC.
42
XMIT
44
16
52
18
DEMOD / DECODE
BPF
RCVR
50
48
46
54
56
DATA LOG
DATA DISPLAY

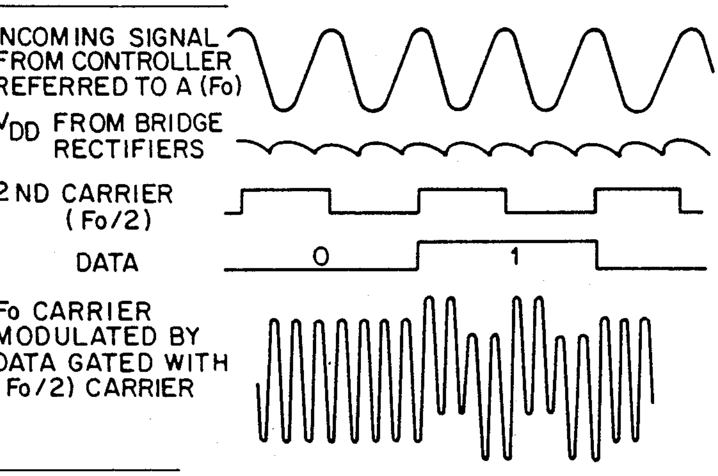

FIG. 5A

FIG. 4

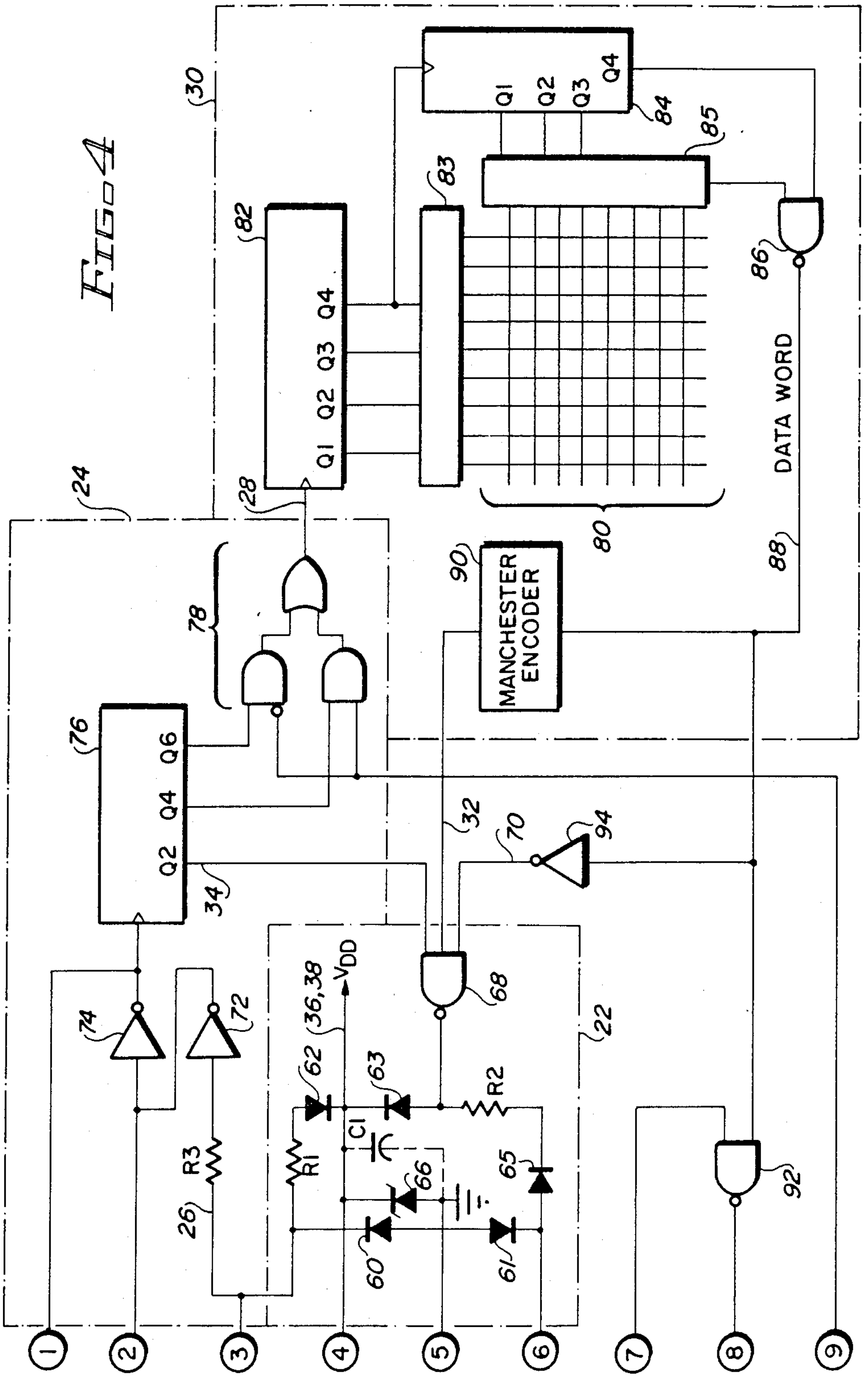
30
84
Q1
Q2
Q3
Q4
85
83
82
Q1
Q2
Q3
Q4
86
DATA WORD
80
88
24
28
78
90
MANCHESTER ENCODER
76
Q6
Q4
Q2
34
32
70
94
68
36,38
VDD
74
72
62
63
R2
22
C1
R3
R1
66
65
26
60
61
92
①
②
③
④
⑤
⑥
⑦
⑧
⑨

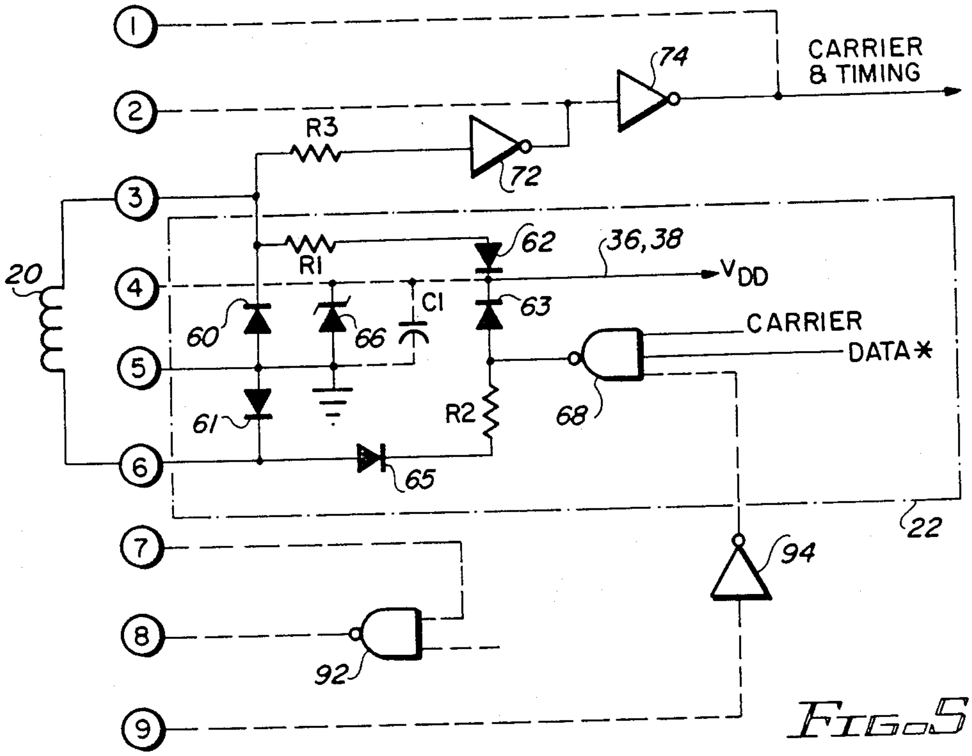
1
2
3
4
5
6
7
8
9
CARRIER & TIMING
74
R3
72
R1
62
36,38
VDD
20
C1
63
60
66
CARRIER
DATA*
61
R2
68
65
22
94
92

FIG. 5

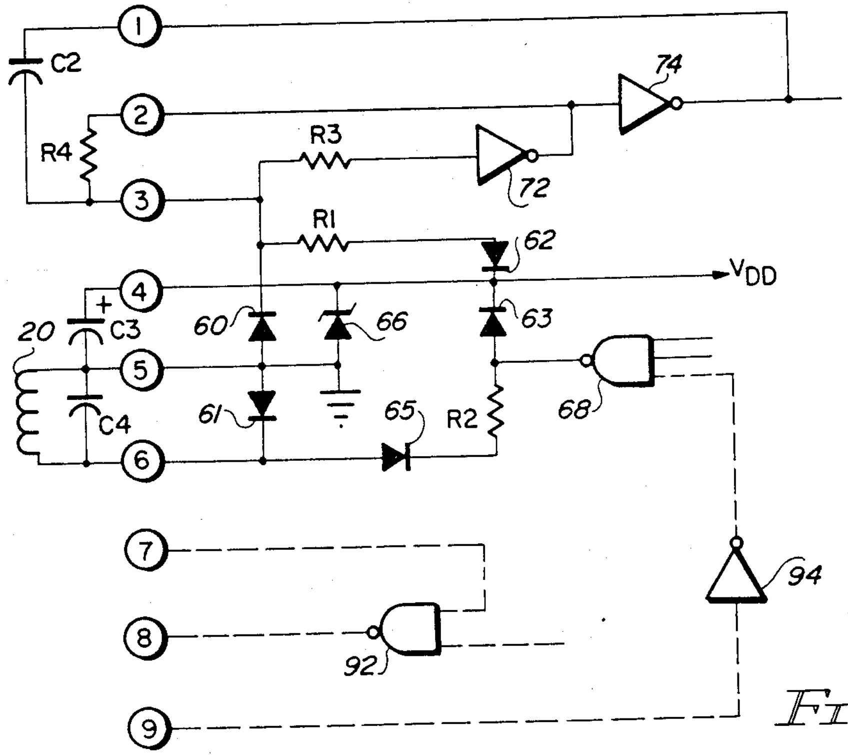
C2
R4
R3
R1
VDD
C3
C4
R2

FIG. 6

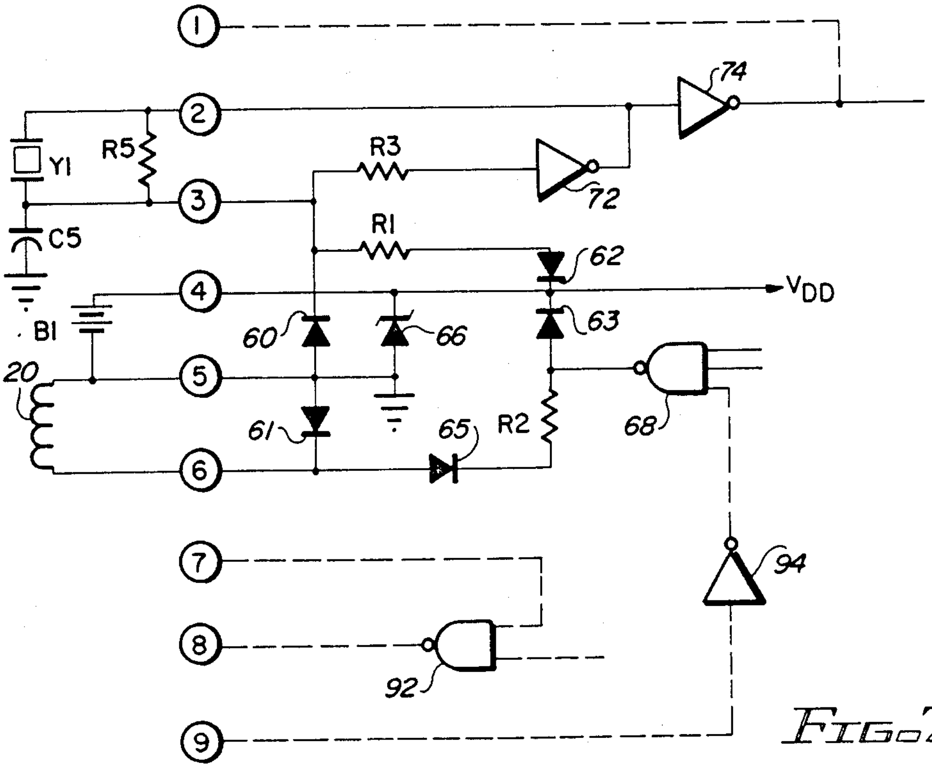
①
②
③
④
⑤
⑥
⑦
⑧
⑨
Y1
R5
C5
B1
20
R3
R1
R2
60
61
62
63
65
66
68
72
74
92
94
VDD

FIG. 7

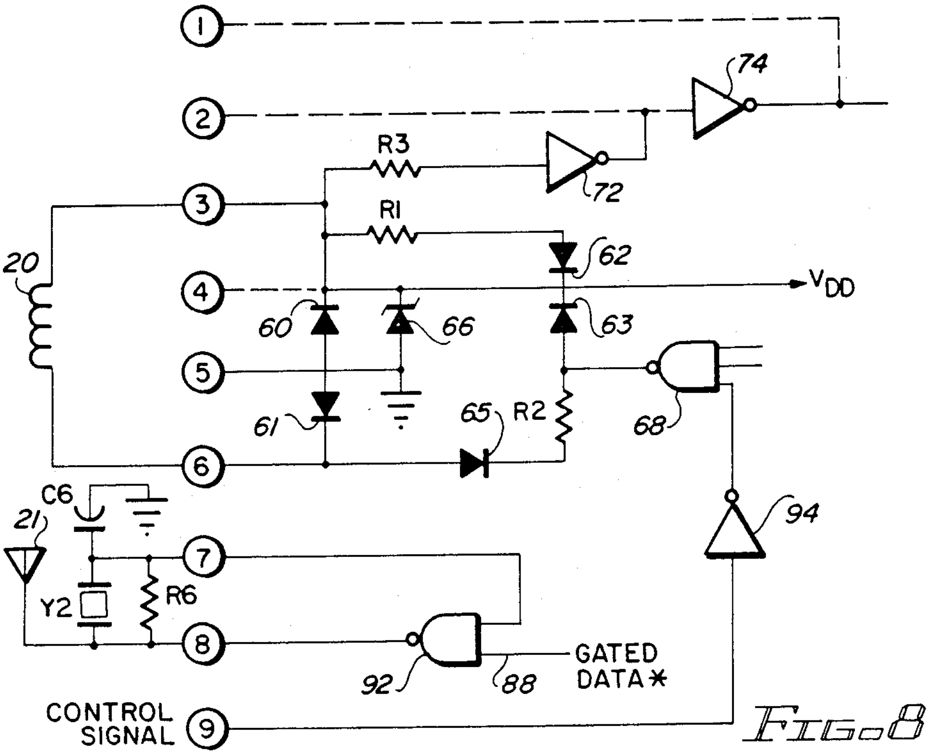
①
②
③
④
⑤
⑥
⑦
⑧
⑨
20
21
C6
Y2
R6
R3
R1
R2
60
61
62
63
65
66
68
72
74
88
92
94
VDD
GATED DATA*
CONTROL SIGNAL

FIG. 8

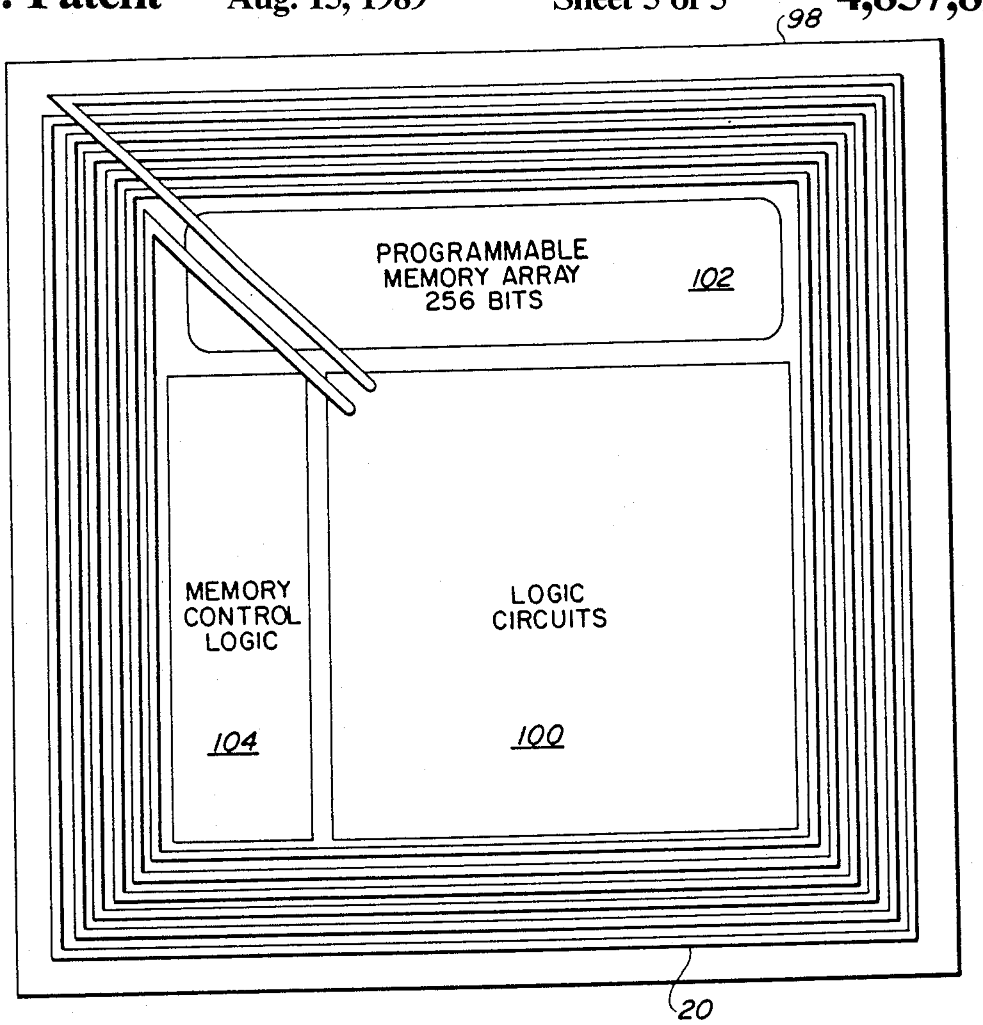
98
PROGRAMMABLE
MEMORY ARRAY
256 BITS
102
MEMORY
CONTROL
LOGIC
104
LOGIC
CIRCUITS
100
20

FIG. 9A

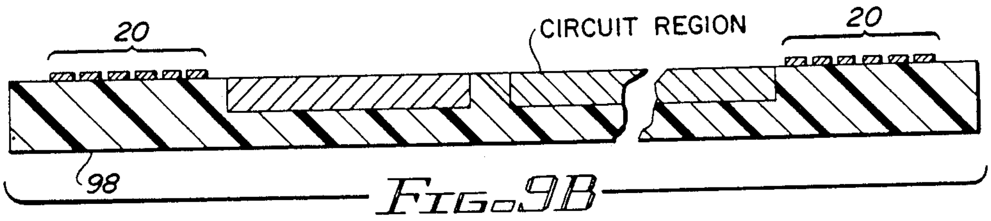
CIRCUIT REGION

FIG. 9B

SINGLE CHIP TRANSPONDER DEVICE

This is a continuation-in-part of copending application Ser. No. 888,246, filed on July 18, 1986.

This invention relates to electronic identification devices and systems, and more particularly to an electronic identification transponder device that is realized with miniaturized circuits built or fabricated on a single monolithic semiconductor chip. This invention also relates to the manner of processing signals within such a transponder device.

BACKGROUND OF THE INVENTION

Electronic identification systems known in the art are generally one of two types: (1) tuned/detuned systems; and (2) transponder systems.

Tuned/Detuned Identification Systems

Tuned/detuned systems are tuned circuits that are selectively detuned by an identification element. (Alternatively, detuned circuits may selectively be tuned by the identification element.) Such systems use a controller/interrogator that sets up some form of magnetic field. This field, when in close proximity to an identification element interacts with the identification element in such a way that the identification element can be identified. For example, in the absence of the identification element, the electromagnetic field created by the interrogator/controller exists in a tuned condition or state. However, as soon as the identification element is brought into close proximity to the field, it interacts with the field in a way to detune it. The manner and degree of detuning can then be measured and used to identify the particular identification element employed. By affixing unique identification elements (those which will interact with the field in a unique way) to a plurality of objects to be identified, such objects can be electronically identified whenever they are brought within the field of the controller/interrogator by simply monitoring the manner and degree of detuning that results. Advantageously, such identification elements may be totally passive, requiring no external or internal power for operation. Thus, they can be realized in a relatively small space. As such, they are especially well suited for use in small, lightweight, inexpensive identification tags that can be placed on the objects to be identified. U.S. Pat. Nos. 3,465,724 and 3,516,575 are examples of such a tuned/detuned electronic identification system.

It is significant to note that the identification element used in such tuned/detuned identification systems does not receive, process, generate, nor transmit any electronic signals. Rather, the identification element merely interacts with the magnetic field created by the interrogator/controller. This interaction can be thought of as a form of magnetic modulation that allows an identification to be made once the field is demodulated.

Because the identification element only interacts with the magnetic field created by the controller/interrogator, and is not required to receive and process any electronic signals transmitted by the interrogator/controller, and is not required to generate and transmit any responsive signals back to the interrogator/controller, the identification element of a tuned/detuned identification system may be realized very inexpensively and compactly. Unfortunately, however, because there are only a limited number of ways that the magnetic field from the interrogator/controller can be modified or modulated, and because the identification element must typically be in very close proximity to the interrogator/controller in order to measurably interact with the magnetic field, such tuned/detuned identification systems are limited in their use. Such systems will typically only be used in applications involving a relatively small number of elements or types of elements that are to be identified. Further, the elements to be identified must be movable to the extent that they can be brought in close proximity to the interrogator/controller.

Transponder Type Identification Systems

A second type of electronic identification system known in the art is the transponder or transceiver system. Such systems employ a transponder or transceiver identification element that is affixed to the objects to be identified. (It is noted that "objects", for purposes of this application and as used hereinafter when referring to items, things, persons, or animals being identified by an identification element of the prior art or the present invention, refers to any thing or item that needs identifying, and includes stationary, movable, inanimate, or animate objects.) This transponder or transceiver element receives an interrogation signal transmitted from a controller/interrogator unit. For purposes of this application, a transponder or transceiver element is any device that receives an incoming interrogation signal and responds thereto by generating and transmitting an outgoing responsive signal. The outgoing responsive signal, in turn, is modulated or otherwise encoded so as to uniquely identify or label the particular object to which the transponder element is affixed.

Because the amount of information that may be contained in the outgoing responsive signal is typically limited only by the type of encoding used, and because modern encoding techniques allow a large amount of information to be included therein, such transponder or transceiver electronic identification systems are especially useful where a relatively large number of objects or types of objects need to be identified. Applicant's U.S. Pat. No. 4,475,481 is an example of such a transponder/transceiver electronic identification system.

Further, because the incoming interrogation signal and the outgoing responsive signal are typically RF signals that have been transmitted from an appropriate RF transmitter circuit, and because such signals can generally be transmitted over further distances than the interactive magnetic field distances associated with detuned/tuned systems, the transponder type identification system is more adaptable to applications where the objects to be identified and the controller/interrogator unit are not necessary in close proximity to each other. (In this application, the term "transponder" will hereinafter be used to mean either "transponder" or "transceiver".)

Unfortunately, because RF transmitter and receiver circuits, as well as encoding and modulation circuits, must be employed within the identification element of a transponder type system, such elements must either carry their own independent power source (e.g., a battery) or must include additional circuitry that allows the needed operating power to be derived from the incoming RF signal. These additional power and circuit requirements have not heretofore allowed the transponder identification element of a transponder type identification system to be realized as inexpensively and compactly as would be desired for many identification applications.

What is needed, therefore, is a transponder type identification element that can be realized inexpensively and compactly in a very small space, thereby providing a transponder type identification system that includes all the advantages of similar prior art transponder type systems, but which eliminates the disadvantages thereof. The present invention addresses this and other needs.

SUMMARY OF THE INVENTION

In accordance with the teachings of the present invention, a very small, inexpensive, reliable transponder device may be realized on a single monolithic semiconductor chip. This chip, in turn, may be incorporated into a very small lightweight tag or label that can be readily and unobtrusively affixed to or carried by movable or ambulatory objects that are to be identified. Such tags or labels may thus function as the identification element of a transponder type identification system as described above.

Advantageously, because the transponder device of the present invention may be realized on a single semiconductor chip, the entire transponder identification element may be packaged in a housing that is much smaller than any similar elements heretofore available. Thus much smaller size opens the door to electronic identification applications not previously possible. For example, in accordance with one embodiment of the invention, the entire transponder device, including its receiving and transmitting antenna coil and power circuits, may be realized on a single semiconductor chip. The small size of a semiconductor chip allows it to be affixed to very small articles, such as vials in a chemical or pharmaceutical processing plant, the location, batch, date code, etc., of any individual vial of which can thereafter be electronically monitored.

The transponder device of the present invention includes receiving means for receiving a carrier signal from a controller/interrogator signal; and diode means for: (1) rectifying the received carrier signal in order to generate the operating power for the device, and (2) mixing an encoded data word (used to identify the particular transponder device being interrogated) with the carrier signal. Also included within the circuits of the transponder device are data generating means for deriving a clock signal that is used to extract a unique code word previously stored in a memory element of the device. This unique code word is thereafter appropriately encoded in order to generate the encoded data word that is mixed with the carrier element. The mixing of the carrier signal with the encoded data word creates sum and difference signals that are transmitted through appropriate transmitting means back to the controller/interrogator unit. Either the sum or difference signal thus serves as the responsive signal that is used to uniquely identify the particular transponder device being interrogated.

An important feature of the present invention is the use of diodes to function both as a rectifier circuit *and* as a balanced modulator circuit.

Another important feature of the invention is the ability to generate a clock signal and data word from the incoming carrier signal, thereby obviating the need for additional, bulky, power-consuming oscillator circuits and modulation/demodulation circuits. Frequency stability is maintained by using a crystal oscillator in the controller/interrogator unit to generate the carrier signal. This carrier signal, when received at the transponder device, is then divided by an appropriate integer in order to generate a stable and second carrier clock signal from which the identifying data word may be generated.

Still another important feature of the invention is the ability to realize all of the transponder circuits on a single monolithic semiconductor chip. In one embodiment, even the receiving/transmitting antenna coil is placed around the periphery of the chip, thereby providing a complete, self-contained, single chip transponder device.

The above and other features combine to provide a transponder device that is vastly improved over any prior art devices. For example, because the circuits can be realized on a single semiconductor chip, the transponder function can be realized for a significantly lower cost. Further, because only a single chip is involved, fewer components are needed to complete the identification system assembly, again resulting in reduced cost and significantly enhancing the system's reliability.

Alternative embodiments of the invention contemplate the use of various hybrid elements to complement and enhance the circuits of the single monolithic semiconductor chip. For example, the addition of a hybrid capacitor and/or resistor allows selective tuning of the device's front end circuits to the particular data rate that is employed. Further, the addition of a hybrid crystal and associated elements allows the device to generate its own stable carrier frequency independent of the carrier frequency from which operating power is derived. Also, the addition of a hybrid battery allows the device to be converted into a self-powered beacon device that periodically transmits its identifying encoded data word without the need for the presence of a carrier signal. For applications where longer distance is required, the coil may be separate from the chip and larger than the chip in order to gather more signal strength (power) from the controller/interrogator.

An electronic identification system is realized according to the teachings of the present invention through the use of a controller/interrogator unit and a transponder device carried by each object to be identified. The controller/interrogator unit includes means for generating and transmitting a stable RF carrier signal. The controller/interrogator unit also includes means for receiving a responsive signal back from an interrogated transponder unit. The transponder unit includes the elements above described that permit it to generate the responsive signal in response to receiving the RF carrier signal, i.e., means for receiving the carrier signal, diode means for rectifying the carrier signal in order to generate operating power for the transponder unit and for mixing the carrier signal with an encoded data word in order to generate the responsive signal, clock means for generating a clock signal that is derived from the carrier signal, data generating means for generating the encoded data word in synchrony with the clock signal, and means for transmitting the responsive signal back to the controller/interrogator unit.

Further, in accordance with the teachings of the present invention, a method is provided for processing a received power carrier signal in a transponder device used in an identification system. This method includes the steps of: (a) receiving the carrier signal, (b) rectifying the carrier signal in order to generate operating power for the transponder device, (c) processing the carrier signal in order to produce a clock signal and second carrier signal, (d) generating a unique data word signal using the clock signal, (e) mixing the second carrier with the power carrier signal in a balanced modulator circuit in order to generate sum and difference signals, the balanced modulator circuit being realized from the same components as are used to rectify the incoming carrier signal, and (f) transmitting the sum and difference signals. In the preferred manner of practicing this method, the clock signal is derived from the carrier signal by simply dividing the received carrier signal by an appropriate integer.

BRIEF DESCRIPTION OF THE DRAWINGS

The above features and advantages of the present invention, as well as other features and advantages, will be more apparent from the following more particular description thereof presented in conjunction with the following drawings, wherein:

FIG. 1 is a block diagram of an identification system, and illustrates the interaction between a controller/interrogator unit and a transponder device;

FIG. 2 is a block diagram of the transponder device of FIG. 1;

FIG. 3 is a block diagram of the controller/interrogator unit of FIG. 1;

FIG. 4 is a schematic diagram of the circuits of the transponder device of FIG. 1;

FIG. 5 is a schematic diagram of the front end of the transponder device of FIG. 1, and illustrates the diode bridge circuit that functions as a rectifier circuit and a balanced modulator circuit;

FIG. 5A illustrates some of the signal waveforms associated with the operation of the circuit of FIG. 5;

FIG. 6 is a schematic diagram of the front end of the transponder device of FIG. 1, and illustrates the use of additional hybrid components that can be connected to the transponder device in order to provide selective tuning of the carrier receiving circuits;

FIG. 7 is a schematic diagram of the front end of the transponder device of FIG. 1, and illustrates the use of a hybrid crystal and battery connected thereto in order to provide additional flexibility in its operation;

FIG. 8 is a schematic diagram of the front end of the transponder device of FIG. 1, and illustrates the use of still additional hybrid elements connected thereto in order to provide selective gating with a crystal oscillator to limit data at a high frequency unrelated to the power carrier frequency;

FIG. 9A is a top view of a monolithic semiconductor chip showing a generalized topographical layout of the various types of circuits used within the transponder device; and

FIG. 9B is a cross-sectional view of the chip of FIG. 9A.

DETAILED DESCRIPTION OF THE INVENTION

The following description is of the best presently contemplated mode of carrying out the invention. This description is not to be taken in a limiting sense but is made for the purpose of describing the general principles of the invention. The scope of the invention should be determined with reference to the appended claims.

Referring to FIG. 1 there is shown a simplified block diagram of an identification system. The system includes a controller/interrogator unit 12 and a plurality of transponder devices or units 14, only one of which is shown in FIG. 1. Each object to be identified carrier a transponder an interrogation signal 16 that is received by the transponder unit 14. After receiving and processing the interrogation signal 16, the transponder 14 transmits a response signal 18 back to the controller/interrogator 12. This response signal 18 has encoded therein an identifying data word that is unique to the particular transponder unit 14 from which is originated. Hence, by positioning or placing the transponder units 14 on the objects to be identified, it is possible to electronically identify the people, animals, or other objects being identified.

Referring next to FIG. 2, a block diagram of the transponder unit 14 is shown. The interrogation signal 16 is received by an antenna coil 20. In the preferred embodiment, this interrogation signal is simply an RF carrier signal that is applied to rectifier/balanced modulator circuit 22 and divide/timing logic circuit 24 over signal line 26. Divide/timing logic circuit 24 divides the carrier signal by an appropriate integer n in order to generate a clock signal and second carrier signal, on output signal line 28, having a frequency that is 1/nth of the frequency of the incoming carrier signal. The clock signal drives data generator circuit 30, which circuit generates an encoded data word. This encoded data word is presented over signal line 32 to the rectifier/balanced modulator 22. Also presented to the rectifier/balanced modulator 22, over signal line 34, is a buffered carrier signal. This buffered carrier signal can be either a "squared-up" version of the incoming carrier signal 16, or a modified carrier signal that has a frequency ½ or ¼ of the incoming carrier signal 16. The modified or buffered carrier signal and the encoded data word are mixed in the rectifier/balanced modulator 22 in order to produce sum and difference signals. These signals appear on signal line 26 and are presented to antenna coil 20, from which they are transmitted as the responsive signal 18.

The incoming carrier signal 16 is also rectified by the rectifier/balanced modulator circuit 22 in order to generate the operating power used by the divide/timing logic 24 and the data generator 30. This operating power is distributed to these circuits over power lines 36 and 38, respectively.

In FIG. 3, a block diagram of the controller/interrogator unit 12 is depicted. While the present invention is primarily directed to the transponder unit 14, and not to the controller/interrogator 12, some features of the controller/interrogator 12 enhance the operation of the transponder unit 14. For example, in the preferred embodiment, the carrier or interrogation signal 16 is generated using a crystal oscillator 40 and appropriate transmission circuitry 42, including transmission antenna or coil 44. This action causes the carrier signal 16 to have a very stable frequency. Because of this stable frequency, it is not necessary, in the preferred approach, to incorporate another crystal oscillator in the transponder unit 14.

The responsive signal 18 generated by the transponder unit 14 is received at the controller/interrogator 12 through antenna coil 46 and receiver circuit 48. These components are tuned to the particular frequency expected for this signal, which, as explained above in conjunction with FIG. 2, will be either the sum or difference frequency resulting from mixing the encoded data word signal with the carrier signal. Further, a band pass filter circuit 50 (which may be included within the receiver circuit 48) will further assure that only those frequencies of interest are passed on to the demodulation/decode circuits **52**. At the demodulation/decode circuits **52**, the responsive signal is demodulated and decoded using known demodulation and decoding techniques in order to extract the data word signal included in the responsive signal **18**. This data word signal serves to identify the particular transponder unit from which it originated. It can be stored in data log circuitry **54** and subsequently displayed by data display circuitry **56**. Alternatively, concurrently with the storage of the data word in data log circuitry **54**, the data word may be displayed by data display circuitry **56**, or data may be sent to a central processor for further processing or control.

Hence, from the block diagrams of FIGS. **1–3**, it is seen that in operation the controller interrogator unit **12** generates and transmits the carrier (interrogation) signal **16**. Upon receipt of this incoming carrier signal **16**, the transponder unit **14** generates a second carrier signal modulated by an encoded data word that is mixed with the carrier signal in rectifier/balanced modulator **22**. This action produces the responsive signal **18** that is transmitted back to the controller/interrogator **12**, which responsive signal **18** uniquely identifies the transponder unit **14** from which it originated.

A schematic diagram showing all usable signals brought out to the pins or transponder unit **14** is illustrated in FIG. **4**. This diagram depicts all of the circuits that are included on the single monolithic semiconductor chip, and further shows the nine pins of terminals, labeled **1–9**, that are used to interface with these circuits. While the antenna coil **20** is not included in the diagram of FIG. **4**, it is noted that it too can be included on the semiconductor chip as shown in FIG. **9**.

Still referring to FIG. **4**, it is seen that the circuitry is relatively simple and that not a lot of circuit elements are required. This simplicity, unavailable heretofore in similar transponder devices, makes possible the incorporation of the device as a self-contained unit on a single chip.

Referring next to FIG. **5**, the front end of the circuits of FIG. **4**—principally the rectifier/balanced modulator circuits **22**—will be explained. From FIG. **5** it is seen that the rectifier/balanced modulator **22** is realized with a diode bridge comprising diodes **60**, **61**, **62**, and **63**. The input terminals to this bridge circuit are pins **3** and **6**. An additional diode **65** is inserted between and in series with diodes **61** and **63**. This diode **65** allows the modulated signal, when applied to the rectifier/balanced modulator **22**, to swing more and thereby increase the percentage of modulation. Further resistors R1 and R2 are inserted into the legs of the bridge circuit as shown in order to distribute the current flow in an appropriate manner. In this regard, resistor R3 also serves to proportion the amount of signal current delivered to the divide/timing logic **24** through pin **3**. In the preferred configuration, R1 has a value that is about 1/10th the value of R2, and R2 has a value that is about 1/10th the value of R3. Hence, in operation, most of the signal current delivered through pin **3**, e.g., during a positive half cycle of the carrier signal, passes to the divide/timing logic **24** and not to the bridge circuit portion that includes diode **62**. However, when pin **6** is positive with respect to pin **3**, e.g., during a negative half cycle of the carrier cycle, then most of the signal current passes through diodes **65** and **63** in order to charge the supply voltage.

A zener diode **66** is incorporated into the bridge circuit in order to stabilize the voltage VDD that is generated. In the preferred embodiment, the voltage VDD is around six volts, although any suitable voltage level could be used. The bridge circuit uses the capacitance inherent in all the circuit elements present on the chip to help store the energy needed to maintain a sufficiently constant supply voltage VDD. This capacitance is shown in FIG. **4**, as C1, but it is to be understood that C1 represents all of the distributed and other internal capacitance present on the chip, and is not a separate element. In this regard, the inventor has found that it is not necessary that the voltage supply VDD he maintained at a constant level at all times. It is sufficient, for the type of synchronous operation herein used, if the supply voltage VDD is maintained at a prescribed level only when the circuits are active (i.e., drawing current). For CMOS circuits, such activity only exists during a data transition. Because the operation of the device is essentially synchronous—everything occurs in synchrony with the clock signal, which in turn is in synchrony with the received carrier signal—it is possible for the device to function so long as the requisite VDD voltage is present during the transitions of the carrier signal. Where the frequency of the carrier signal is sufficiently fast, as is the case here, the capacitance C1 doesn't have to hold a charge very long in order for the device to operate. Hence, C1 (which, again, is comprised of all the distributed and internal capacitance on the chip) does not have to be very large in order for the device to operate. It is noted, for example, that the static current drain of a typical CMOS chip is the picoamp range.

Further, even when semiconductor devices other than CMOS circuits are employed, such as TTL or GaAs devices, it is possible for the circuit to function during only a selected half cycle of the carrier signal by setting up the logic circuits so that they only change states during this selected half cycle. Thus, it is possible to design the VDD generation circuitry so that it is primarily charging up during, for example, the negative half cycle of the carrier signal, and design the logic circuits so that they are not active until the positive half cycle of the carrier signal, at which time sufficient charge will be available to provide the needed operating power.

FIG. **5A** illustrates some of the foregoing principles of operation relative to the circuit of FIG. **4** and **5**. In FIG. **5A**, the incoming carrier signal **16** from the controller/interrogator **12** is shown having a frequency Fo. After full-wave rectification, this signal substantially as a dc level having a ripple frequency of 2Fo. A second carrier signal is obtained by dividing the incoming carrier signal by two. This second carrier signal therefore has a frequency of Fo/2. This second carrier signal can, for desired embodiments, be used as a clock signal to generate a data signal, two bits of which, "01", are shown in FIG. **5A**. In turn, this data signal, gated with the second carrier signal Fo/2, is used to modulate the Fo carrier signal, realizing a modulated signal as shown at the bottom of FIG. **5A**.

Also included as part of the rectifier/balanced modulator circuit **22** of FIGS. **4** and **5** is NAND gate **68**. This gate **68** serves to combine the carrier signal (or a derivative of the carrier signal), on signal line **34**, with the encoded data, on signal line **32**, and apply this combination to the diodes **60–63** of the bridge circuit. This action causes these signals to be mixed, resulting in sum and difference frequencies that are presented to the output pins **3** and **6** (and hence to the antenna coil **20**) for transmission to the controller/interrogator **12** or other receiving device.

It is noted that gate **68** of the rectifier/balanced modulator **22** also has a control signal line **70** applied thereto, the function of which is explained below in conjunction with the discussion of FIG. **8**.

Referring next to the divide/timing logic portion **24** of FIG. **4**, it is seen that the incoming carrier signal (received through pin **3**) passes through resistor R3 and inverter gates **72** and **74**. The output of gate **72** is made available at pin **2**, and the output of gate **74** is made available at pin **1**. These pins can be used as explained below in connection with the discussion of FIGS. **6** and **7**.

The output of gate **74** is presented to the clock input of a register **76**. This register **76** divides the carrier signal by an appropriate factor. While the embodiment shown uses a simple series divide-by-2 arrangement, thereby providing outputs that have been divided by 2, 4, 8, 16, . . . , the invention contemplates that other dividing schemes could be employed such that the carrier signal could be divided by any desired integer n.

The desired division of the carrier signal in register **76** provides a clock signal that has a frequency that is 1/nth of the carrier frequency. Included in the divide/timing logic **24** are selection gates **78** that allow the selection of different clock frequencies as controlled by a control signal applied to pin **9**. The selected clock signal is then presented to the data generator circuitry **30** over signal line **28**.

Data generator circuitry **30** includes a programmable read only memory (PROM) array **80** which is accessed through suitable memory control logic elements **82–85**. Elements **82** and **84** are registers through which the clock signal cycles. After a prescribed number of clock cycles, these registers address certain rows and columns of the memory array **80** in order to extract the data that has been previously programmed therein. This data is passed serially to gate **86**, as is the signal Q4 from register **84**. As long as Q4 is high, the extracted data word is allowed to pass out of gate **86**, over signal line **88**, to Manchester encoder circuit **90**. Manchester encoder circuit **90** encodes the data word according to well known Manchester encoding techniques and presents the encoded data word to the rectifier/balanced modulator circuit **22** over signal line **32**. In the preferred embodiment, the encoded word is **64** bits long, although any suitable bit length could be employed. Further, encoding schemes other than a Manchester encoding scheme could also be employed.

The PROM memory array **80** may be realized using any suitable PROM configuration that can readily be incorporated on the same semiconductor chip as are the other transponder circuits. Preferably, an EEPROM (Erasable Electrically Programmable Read Only Memory) device could be used. Advantageously, during manufacture of the chip, or thereafter, the PROM can be easily programmed to contain an appropriate data word that would serve to uniquely identify that particular device. For example, a manufacturing date code in combination with a serial number could be employed for this purpose. If the identifying data word programmed into the device during manufacture did not suit the particular application, this data word could subsequently be erased using appropriate equipment and a new, more suitable, data word could be programmed thereinto.

Also shown in FIG. **4** is a NAND gate **92** and an inverter gate **94**. The inverter gate **94** is connected to pin **9** and is used to condition a control signal applied to pin **9** as discussed below in conjunction with FIG. **8**. The NAND gate **92** has one input connected to pin **7** and the other input connected to the data word signal output from the memory array **80** through gate **86**. The output of this NAND gate **92** is connected to pin **8**. This gate **92** can be selectively connected to additional hybrid elements as discussed below in connection with FIG. **8** in order to provide additional options for using the transponder device.

Referring next to FIGS. **6–8**, schematic diagrams of the front end circuits of the transponder unit **14** are shown, as in FIG. **5**, but further including additional hybrid elements connected thereto through the pin connections provided. In FIG. **6**, capacitor **C2** is connected between pins **1** and **3**, and resistor **R4** is connected between pins **2** and **3**. Further, hybrid capacitors **C3** and **C4** are connected in tandem as shown between pins **4** and **5**, and pins **5** and **6**, respectively. Also, in FIG. **6**, the antenna coil **20** is connected between ground, pin **5**, and pin **6**, in order to disassociate it from interacting with **R4** and **C2**. The values of capacitors **C3** and **C4** are selected as required in order to tune the front end circuit to the particular carrier frequency that is employed. The values of **R4** and **C2** are selected as needed in order to adjust the bit timing. Thus, flexibility is provided in order to make minor adjustments, i.e., tune the circuit, for the particular carrier frequency that is used.

In FIG. **7**, the single chip transponder unit **14** of the present invention is converted to a more conventional transponder unit through the use of a hybrid crystal oscillator circuit, comprising crystal **Y1**, capacitor **C5**, and resistor **R5**, and a battery **B1**. The crystal **Y1** and resistor **R5** are connected across pins **2** and **3**, while capacitor **C5** is shunted from pin **3** to ground (pin **5**). This configuration results in an oscillator circuit comprising the inverter gate **72**, resistors **R3** and **R5**, crystal **Y1**, and capacitor **C5**. This oscillator circuit allows the data word to be generated totally independent of the frequency of any incoming carrier signal. Further, when a battery **B1** is employed, as shown, the transponder device is no longer dependent on the carrier signal, and therefore becomes a transmitting circuit that continually transmits its identifying signal. When a battery **B1** is employed in this fashion, it will typically have a voltage level output that is slightly less than that of zener diode **66**, thereby keeping zener diode **66** turned off and effectively out of the circuit.

In FIG. **8**, a hybrid crystal **Y2**, capacitor **C6** and resistor **R6** are connected as shown to pins **7** and **8** in order to create a crystal oscillator circuit that selectively transmits the data word through a separate antenna **21**. The crystal oscillator is realized with internal-to-the-chip gate **92** and **Y2**, **C6**, and **R6**. Antenna coil **20** still receives a carrier signal and generates VDD therefrom. Further, the frequency of the carrier signal determines the clock rate, which in turn determines the rate at which the data word, appearing on signal line **88**, appears. The crystal oscillator output signal, presented to antenna **21**, is effectively modulated with the data word appearing on signal line **88**. Because gate **92** is only active when VDD is present, the circuit of FIG. **8** can thus function as a responding unit that transmits the data word on a carrier frequency independent of the incoming carrier signal. The use of a control signal applied to pin **9** further disables any signal from being mixed in the rectifier/balance modulator circuit **22**, and hence from being transmitted through antenna **20**, during the time that a signal is being transmitted from antenna **21**. Such a control signal can simply be realized by hardwiring pin **9** to a desired potential.

It is noted that the circuits of FIGS. **6–8** represent only a portion of the various circuit configurations and options that could be realized using the single chip transponder device of the present invention.

FIG. **9A** shows a topographical representation of the transponder chip **14** in an embodiment that includes the antenna coil **20** as part of the monolithic chip. In accordance with this embodiment, the coil **20** is etched or otherwise placed around the periphery of the chip substrate **98**, as described below. In the center of the coil **20** are found a custom logic circuit **100**, a programmable memory array **102**, and memory control logic **104**. Programmable memory array **102**, which may be realized for example from an EEPROM device, comprises the memory array **80** depicted in FIG. **4**. Memory control logic **104** comprises the memory control elements **82–85** of FIG. **4**, and the custom logic circuit **100** includes the balance of the circuitry shown in FIG. **4**. Thus, using the chip topography shown in FIG. **9A**, a functionally complete transponder unit **14** is realized on a single semiconductor chip.

Still with reference to FIG. **9A**, it is noted that placement of the logic circuit **100**, memory array **102**, and memory control logic **104** on the chip substrate **98** can be readily accomplished by those skilled in the art using known hybrid or monolithic circuit fabrication techniques. Not so readily known in the art, is the manner in which the antenna coil **20** is placed on the chip substrate **98**.

One approach for placing the coil **20** on the chip substrate **98** is to make the coil **20** in hybrid fashion, using e.g. a copper conductor, and then bonding the wound coil **20** to the chip using conventional bonding techniques. This approach would advantageously allow the coils to be effectively stacked, if needed, in order to increase the coil inductance without adding stray capacitance.

The preferred approach for placing the coil **20** on the chip substrate **98**, however, is to use a deposition technique that effectively creates a single monolithic chip assembly that includes all of the circuitry necessary to produce a functionally complete transponder unit **14**. While, the prior art teaches that it is extremely difficult to fabricate small air core inductors of sufficiently high inductance and Q for integrated circuit applications, the suggestion is made in the prior art that the use of magnetic film cores for such inductors can reduce the number of turns required, and thereby make the fabrication of such inductors feasible. See Soohoo, "Magnetic Thin Film Inductors For Integrated Circuit Applications", IEEE Transactions in Magnetic, Vol. MAG-15, No. 6, pp. 1803–1805 (Nov. 1979). One technique for using magnetic film cores is described in Salch and Qureshi, "Permalloy Thin-Film Inductors", *Electronic Letters*, Vol. 6, No. 26, pp. 850–852 (Dec. 31, 1970). There it is stated that a 10-turn square spiral coil for use at 10 MHz can be constructed having an outer diameter of 1 cm × 1 cm. The conducting path width is 0.005 inches. The spacing between turns is 0.001 in. The copper path is deposited by vacuum evaporation and then thickness is built up to about 25 micrometers by electroplating. Two permalloy magnetic films, having a thickness of from 1000–3000 Angstroms, surround the conductors, one on top, and the other on the bottom. The film is evaporated in an orienting magnetic field in such a way that the long dimension is parallel to the field, and thus it is the easy direction of magnetisation of the film. When a high-frequency current passes in the coil, the magnetic films are driven in a hard direction, and the two magnetic films around each conductor act as a magnetic core enclosing a 1-turn coil. The effect of the magnetic films is to increase the inductance of the coil in addition to its free-space inductance. The magnetic permeability is quite large, since the films are driven in the hard direction. Further, an insulting silicon-monoxide layer (SiO, 10,000 A thick) separates each magnetic film from the conducting path. For purposes of the present invention, the technique used to place the coil **20** on the substrate **98** is as follows:

(a) a wafer of semiconductor substrate material is processed, using conventional methods and technology, to place the IC circuits in the center of each desired chip area, as shown in FIGS. **9A** and **9B**;

(b) a conductive trace **110** is deposited on the substrate between the regions **112**, **114** (optionally, this step may be included in the fabrication of the circuit of step (a));

(c) a permalloy magnetic film, or other suitable material having a large magnetic permeability and electrical insulating properties, is applied over the wafer, leaving holes for contact to the return trace and the logic circuits at **106**, **108**, **112** and **114**.

(d) the coil **20**, with the exception of the trace **110** (which was previously deposited), is deposited and lithographed using conventional techniques,

(e) contacts are made to the logic circuit **100** through the holes **108**, **106**, and to the trace **110** through holes **112** and **114**;

(f) a second permalloy magnetic film, or other suitable material having desired properties, is applied over the wafer;

(g) the wafer is diced, thereby creating individual chips, each being a fully functional transponder device.

It is noted that the application of the magnetic film, in steps (c) and (f) involves appropriate evaporation (or equivalent techniques) in order to control the thickness thereof, which evaporation may occur in the presence of an orienting magnetic field, as previously described. Thus, this step may involve two sub-steps, a first to place and magnetically orient the magnetic film in a first direction; and a second to place and magnetically orient the film in a second direction (orthogonal to the first direction).

While the invention herein disclosed has been described by means of specific embodiments and applications thereof, numerous modifications and variations could be made thereto by those skilled in the art without departing from the spirit and scope of the present invention. It is to be understood therefore that within the scope of the appended claims, the invention may be practiced otherwise than as specifically described herein.

What is claimed is:

1. A transponder device comprising:

a single monolithic semiconductor chip;

receiving means for receiving a carrier signal from an external source;

diode means for rectifying the received carrier signal in order to generate operating power for the transponder device, and for mixing an encoded data word with the carrier signal in order to generate sum and difference signals;

digital circuit means for deriving a clock signal from said carrier signal;

data generating means responsive to said clock signal for generating the encoded data word that is mixed with the carrier signal in said diode means; and

transmitting means for transmitting the sum and difference signals to a location external from said transponder device;

said diode means, digital circuit means, and data generating means comprising electrical circuits that are realized on said single monolithic semiconductor chip.

2. The transponder device of claim **1** further including a plurality of select hybrid circuit elements and means for connecting said hybrid circuit elements to the electrical circuits of said monolithic semiconductor chip in order to tune said electrical circuits to the frequency of said carrier signal.

3. The transponder device of claim **1** further including a hybrid battery and means for connecting said hybrid battery to the electrical circuits of said monolithic semiconductor chip in order to provide operating power therefor independent of any power contained within said carrier signal.

4. The transponder device of claim **1** further including a hybrid crystal element and means for connecting said hybrid crystal element to the electrical circuits of said monolithic semiconductor chip in order to provide crystal frequency control of the signals used within said electrical circuits independent of the frequency of said carrier signal.

5. The transponder device of claim **1** wherein the receiving means and transmitting means comprise a single antenna coil that is placed around the periphery of the single monolithic semiconductor chip.

6. The transponder device of claim **5** wherein said monolithic semiconductor chip is no larger than 0.01 square inches.

7. The transponder device of claim **1** wherein said diode means comprises a bridge circuit that functions both as a rectifier circuit and a balanced modulator circuit.

8. The transponder device of claim **1** wherein said data generating means includes programmable read only memory (PROM) means, said PROM means having a unique code word previously programmed therein from which said encoded data word is derived.

9. The transponder device of claim **8** wherein said data generating means includes encoding means for encoding the unique code word programmed in said PROM means.

10. The transponder device of claim **9** wherein said encoded data word generated by the encoding means of said data generating means comprises an n bit digital word, wherein n is an integer greater than 7.

11. An electronic identification system for identifying a plurality of objects comprising:

a controller/interrogator unit for interrogating the objects to be identified and processing the responses received, said controller interrogator unit including:

means for generating an RF carrier signal, said RF carrier signal comprising an interrogation signal,

means for transmitting said interrogation signal towards an object to be identified,

means for receiving a responsive signal generated in response to said interrogation signal, and

means for decoding, storing, and displaying the response received; and

a transponder unit, carried by each object to be identified, comprising:

a single monolithic semiconductor chip;

receiving means on said semiconductor chip for receiving the interrogation signal from said controller-interrogator unit,

diode means on said semiconductor chip for rectifying the received interrogation signal in order to generate operating power for the transponder device, and for mixing an encoded data word with a digital carrier signal in order to generate sum and difference signals, said sum and difference signals comprising said responsive signal,

digital circuit means on said semiconductor chip for deriving a clock signal and said digital carrier signal from said interrogation signal,

data generating means on said semiconductor chip responsive to said clock signal for generating the encoded data word that is mixed with the digital carrier signal in said diode means, and

transmitting means on said semiconductor chip for transmitting the responsive signal to said controller/interrogator unit.

12. The identification system of claim **11** wherein the receiving means and transmitting means of said transponder unit include a single antenna coil that is placed around the periphery of said single monolithic semiconductor chip.

13. A transponder device consisting of:

a single semiconductor chip;

receiving means placed on said semiconductor chip for receiving a carrier signal from an external source;

diode means for rectifying the received carrier signal in order to generate operating power for the transponder device, and for mixing an encoded data word with the carrier signal in order to generate sum and difference signals, said diode means comprising an electrical circuit that is fabricated on said semiconductor chip;

digital circuit means for deriving a clock signal from said carrier signal, said digital circuit means comprising electrical circuits that are fabricated on said semiconductor chip;

data generating means responsive to said clock signal for generating the encoded data word that is mixed with the carrier signal in said diode means, said data generating means comprising an electrical circuit that is fabricated on said semiconductor chip; and

transmitting means placed on said semiconductor chip for transmitting the sum and difference signals to a location external from said transponder device.

14. The transponder device of claim **13** wherein the receiving means and transmitting means include an antenna coil that is deposited around the periphery of the single monolithic semiconductor chip.

* * * * *

US005223781A

United States Patent [19]

Criswell et al.

[11] **Patent Number: 5,223,781**

[45] **Date of Patent: Jun. 29, 1993**

[54] **POWER COLLECTION AND TRANSMISSION SYSTEM AND METHOD**

[76] Inventors: **David R. Criswell**, 4003 Camino Lindo, San Diego, Calif. 92122; **Robert D. Waldron**, 15339 Regalado St., Hacienda Height, Calif. 91745

[21] Appl. No.: **490,717**

[22] Filed: **Mar. 8, 1990**

Related U.S. Application Data

[63] Continuation of Ser. No. 157,081, Feb. 11, 1988, Pat. No. 5,019,768, which is a continuation of Ser. No. 732,352, May 8, 1985, abandoned, which is a continuation-in-part of Ser. No. 513,444, Jul. 13, 1983, abandoned.

[51] **Int. Cl.**^{5} **B64G 1/00; H02M 7/00**

[52] **U.S. Cl.** .... **322/2 R**; 290/1 R; 244/159

[58] **Field of Search** .... 322/2 R; 290/1 R; 244/159

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,225,208	12/1965	Wolfe	322/2 R X
3,999,182	12/1976	Moeller et al.	343/100 SA
4,078,747	3/1978	Minovitch	322/3 R X
4,251,679	2/1981	Zwan	322/2 R X
4,305,555	12/1981	Davis	322/3 R X
4,371,135	2/1983	Keigler	322/2 R X
4,490,668	12/1984	Sterzer	322/2 R
5,019,768	5/1991	Criswell et al.	322/2 R

Primary Examiner—R. J. Hickey
Attorney, Agent, or Firm—Brown, Martin, Haller & McClain

[57] **ABSTRACT**

A system for transmitting microwaves to one or more receiver assemblies comprises an array of separate microwave transmitting assemblies for emitting a plurality of microwave beams, the array being arranged to apparently fill a radiating aperture of predetermined shape and size when viewed from the direction of a receiver assembly, and a phase controlling assembly associated with the microwave transmitting assemblies for controlling the relative phase of the emitted beams to form at least one composite shaped microwave beam directed to at least one receiver assembly. The receiver assembly is located in the near field of the microwave beam. The transmitting assemblies are each associated with separate power sources comprising solar power collecting assemblies. In one arrangement the transmitting system is on the moon and the or each receiver assembly is on Earth.

9 Claims, 10 Drawing Sheets

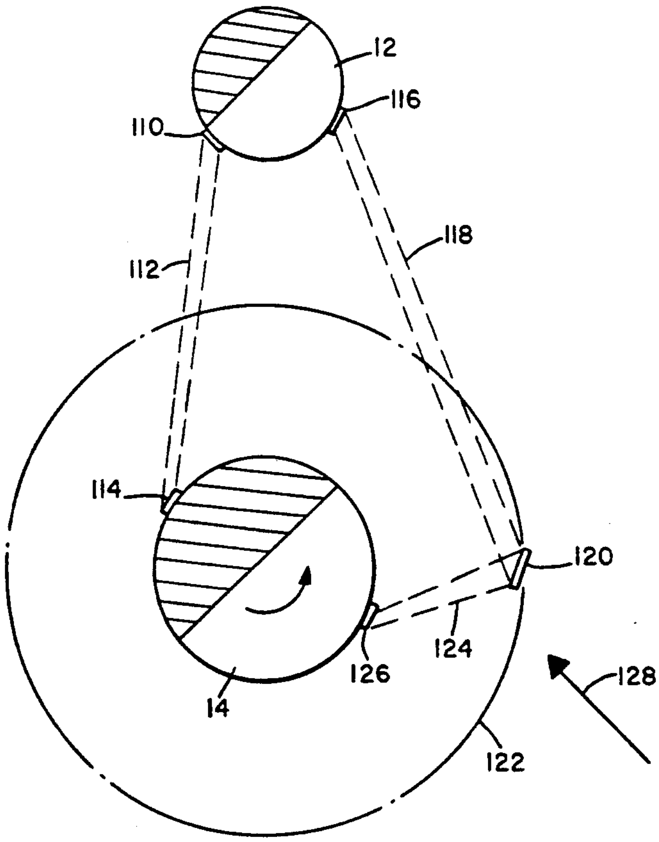

FIG. 1A
FM

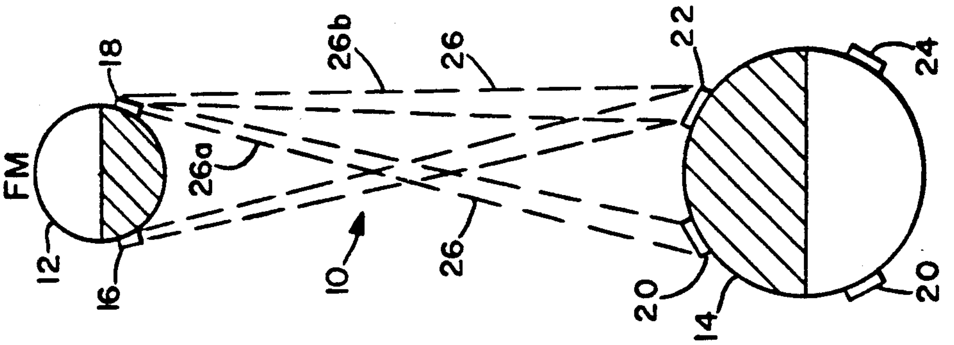

12
16
18
26a
26b
26
26
10
22
24
20
20
14

FIG. 1B
LQ

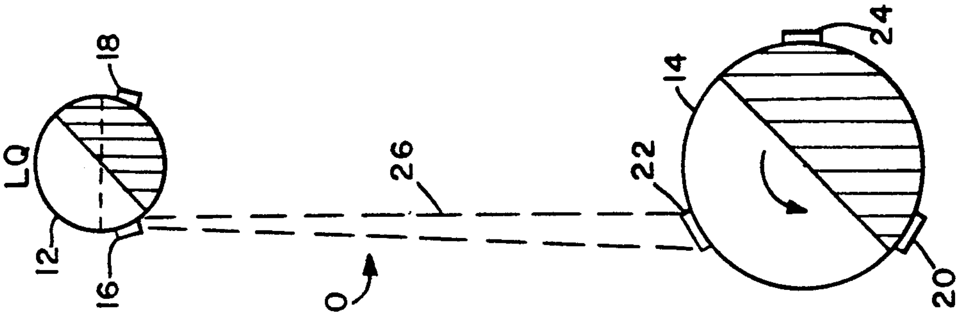

12
16
18
26
10
14
22
24
20

FIG. 1C
NM

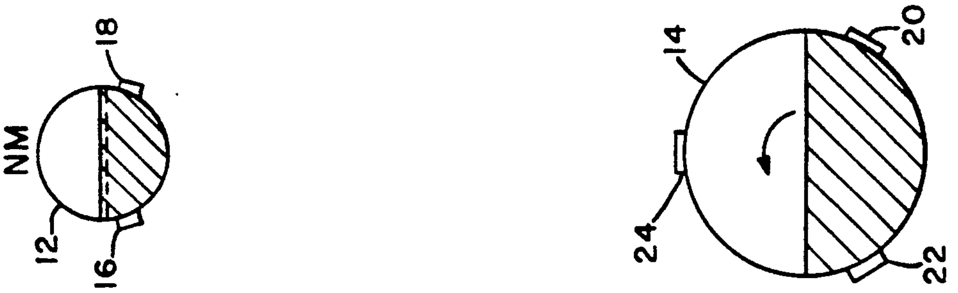

12
16
18
14
20
24
22

FIG. 1D
FQ

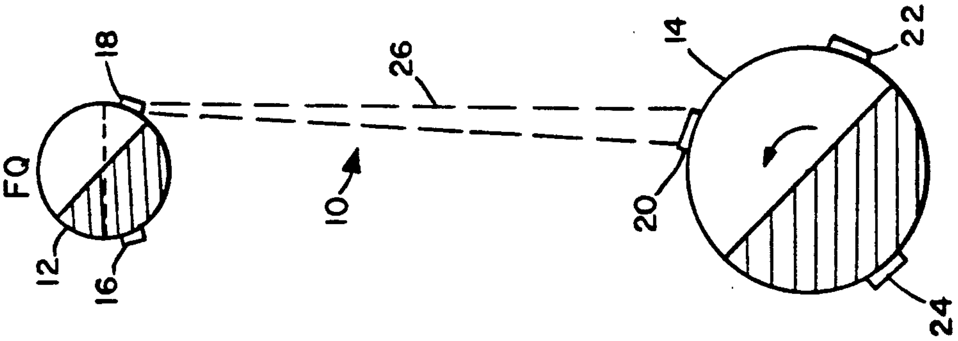

12
16
18
26
10
14
22
20
24

FIG. 1E
FM

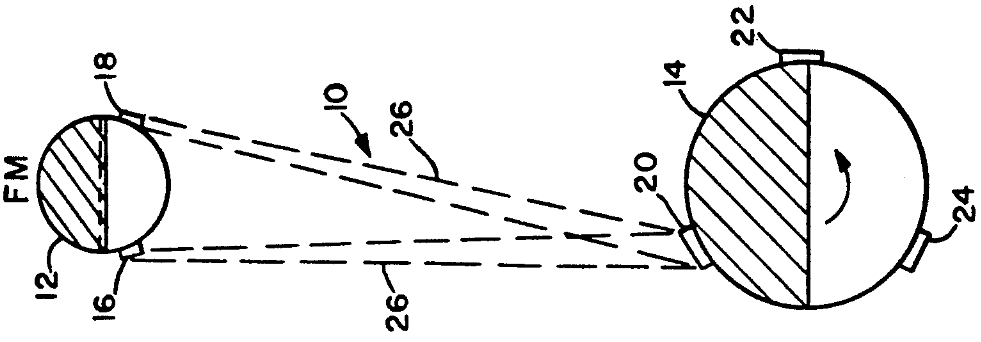

12
16
18
10
26
26
14
22
20
24

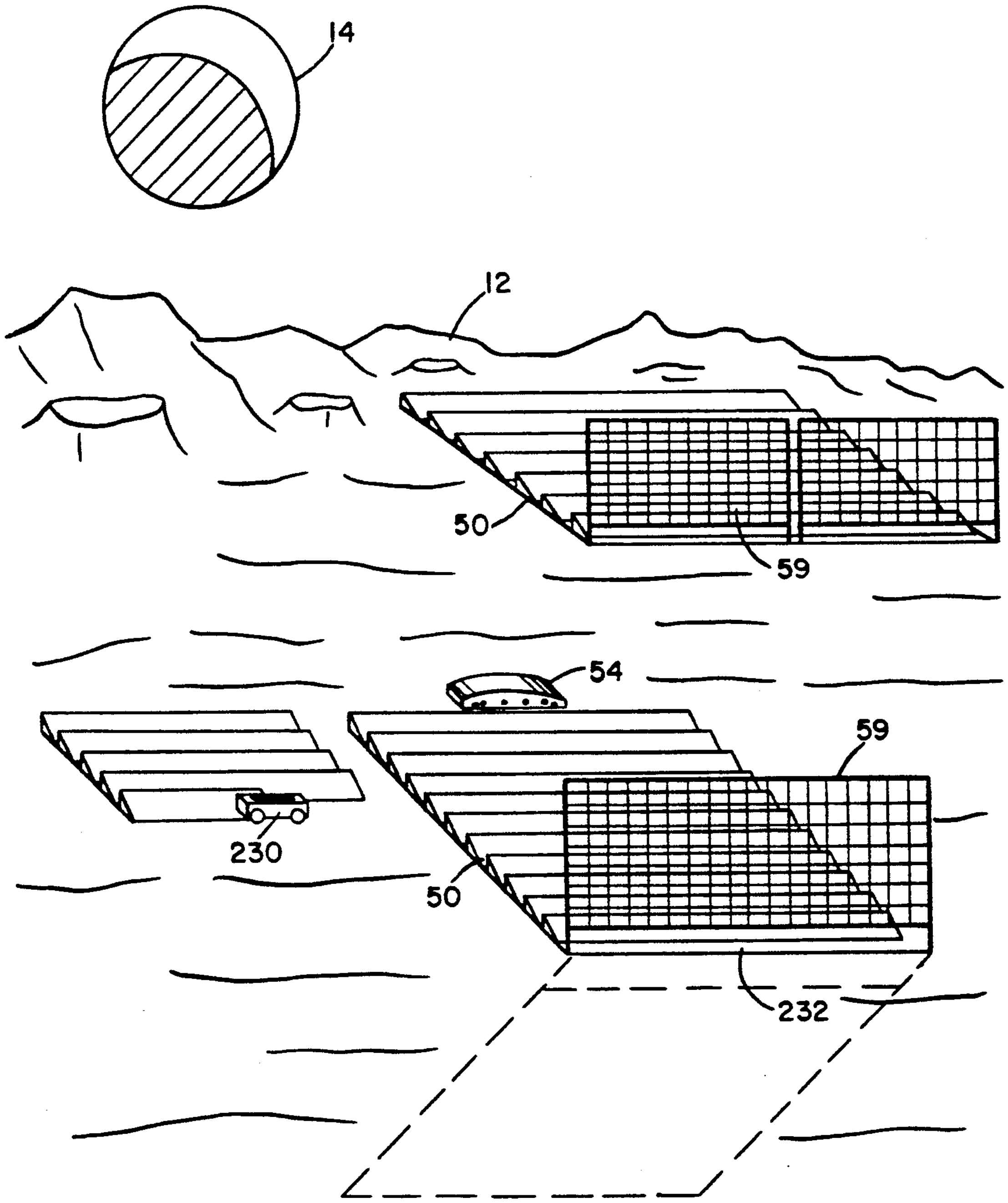
14
12
50
59
54
59
230
50
232

FIG. 2

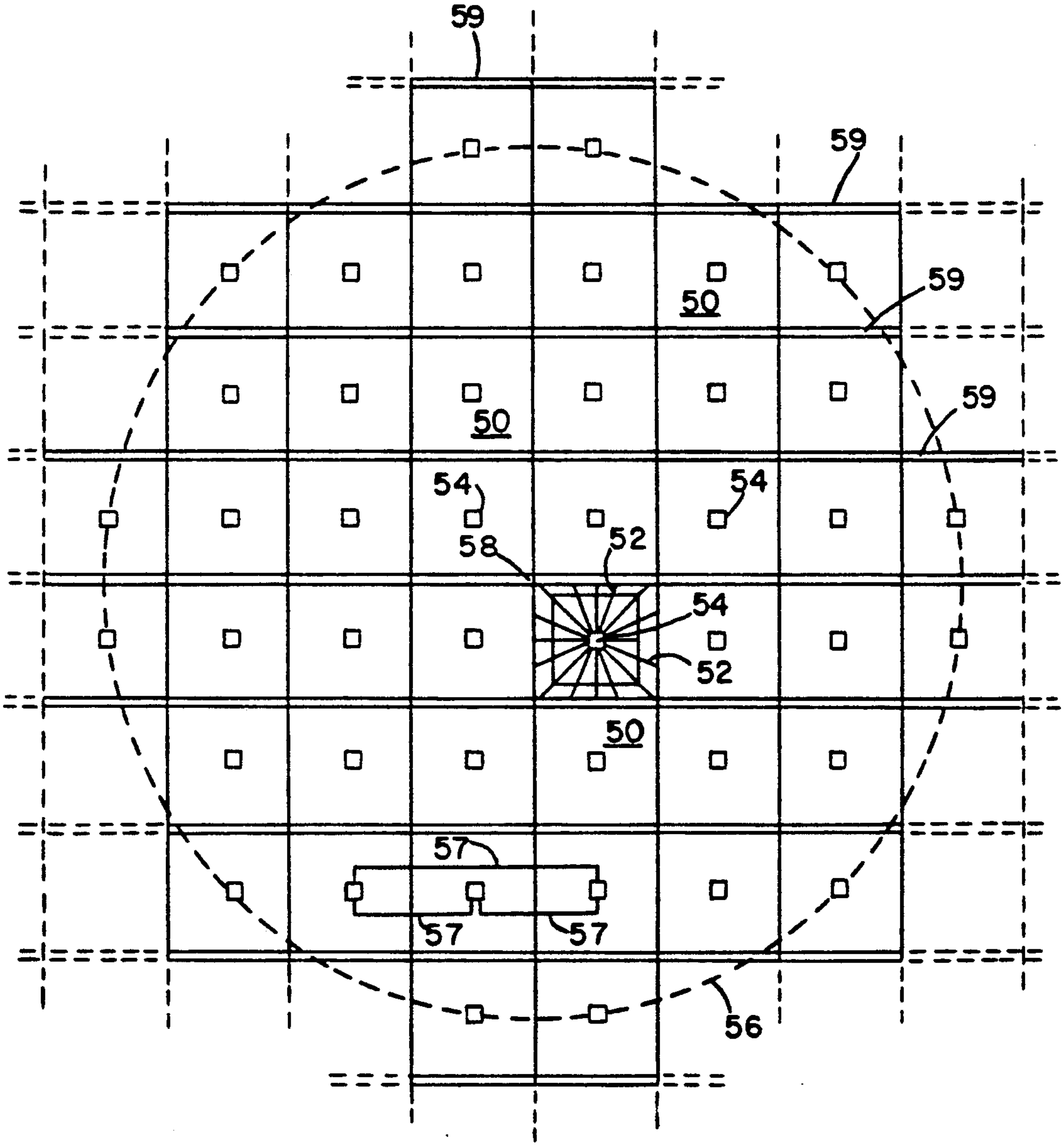
59
59
59
59
50
50
54
54
58
52
54
52
50
57
57
57
56

FIG. 3A

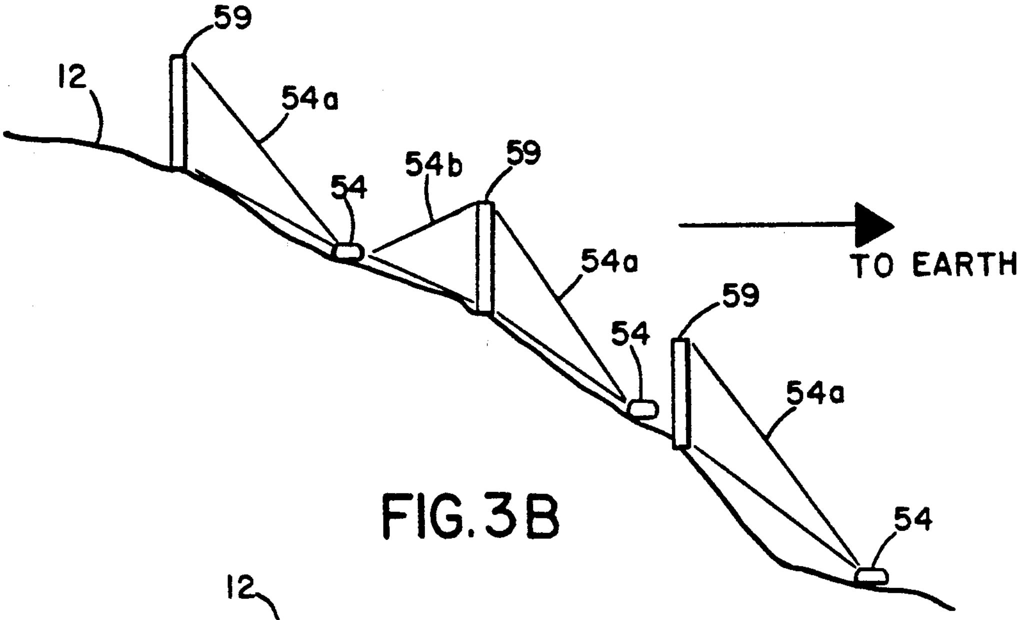
59
12
54a
54
54b
59
TO EARTH
54a
59
54
54a
54

FIG. 3B

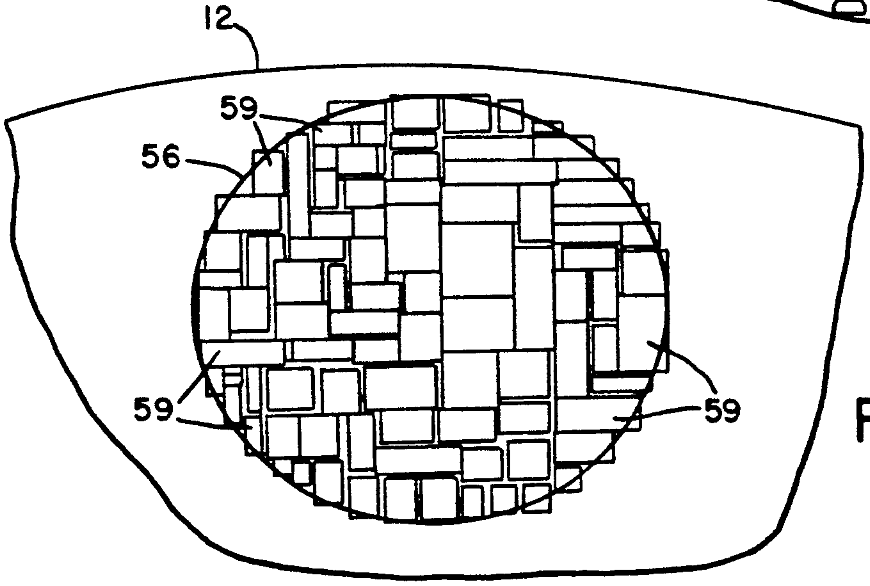
12
59
56
59
59

FIG. 3C

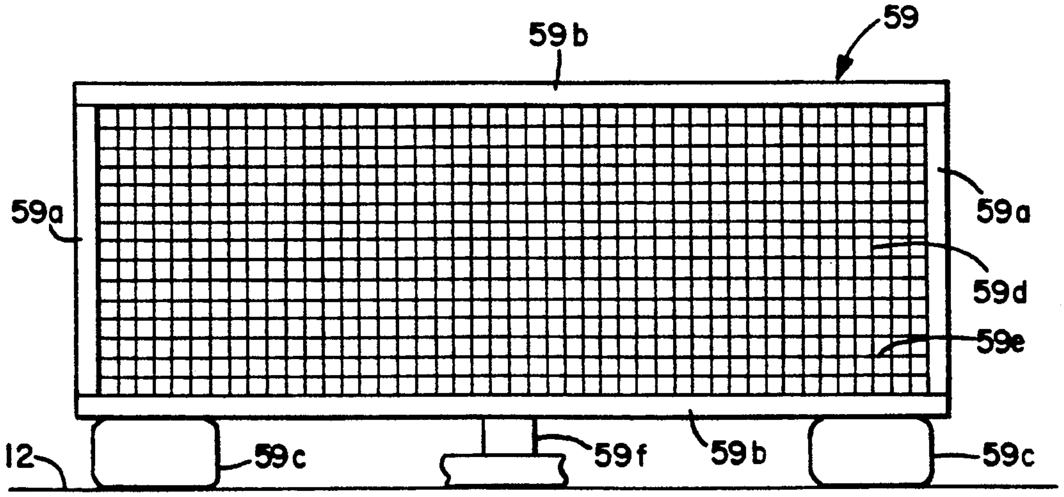
59b
59
59a
59a
59d
59e
12
59c
59f
59b
59c

FIG. 3D

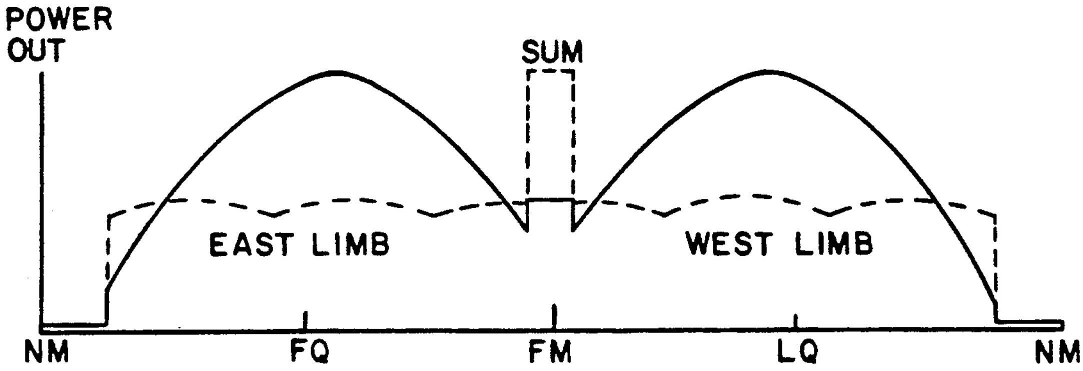
POWER OUT
SUM
EAST LIMB
WEST LIMB
NM
FQ
FM
LQ
NM

FIG.4

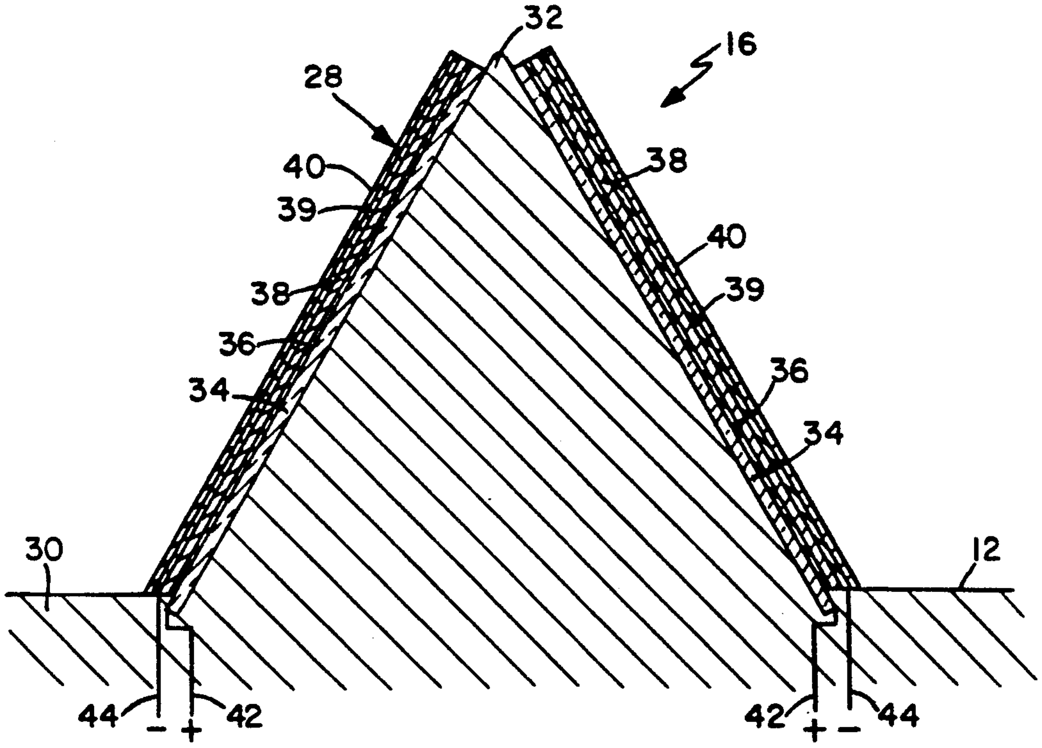
32
16
28
40
39
38
36
34
38
40
39
36
34
30
12
44
42
42
44
-
+
+
-

FIG. 5

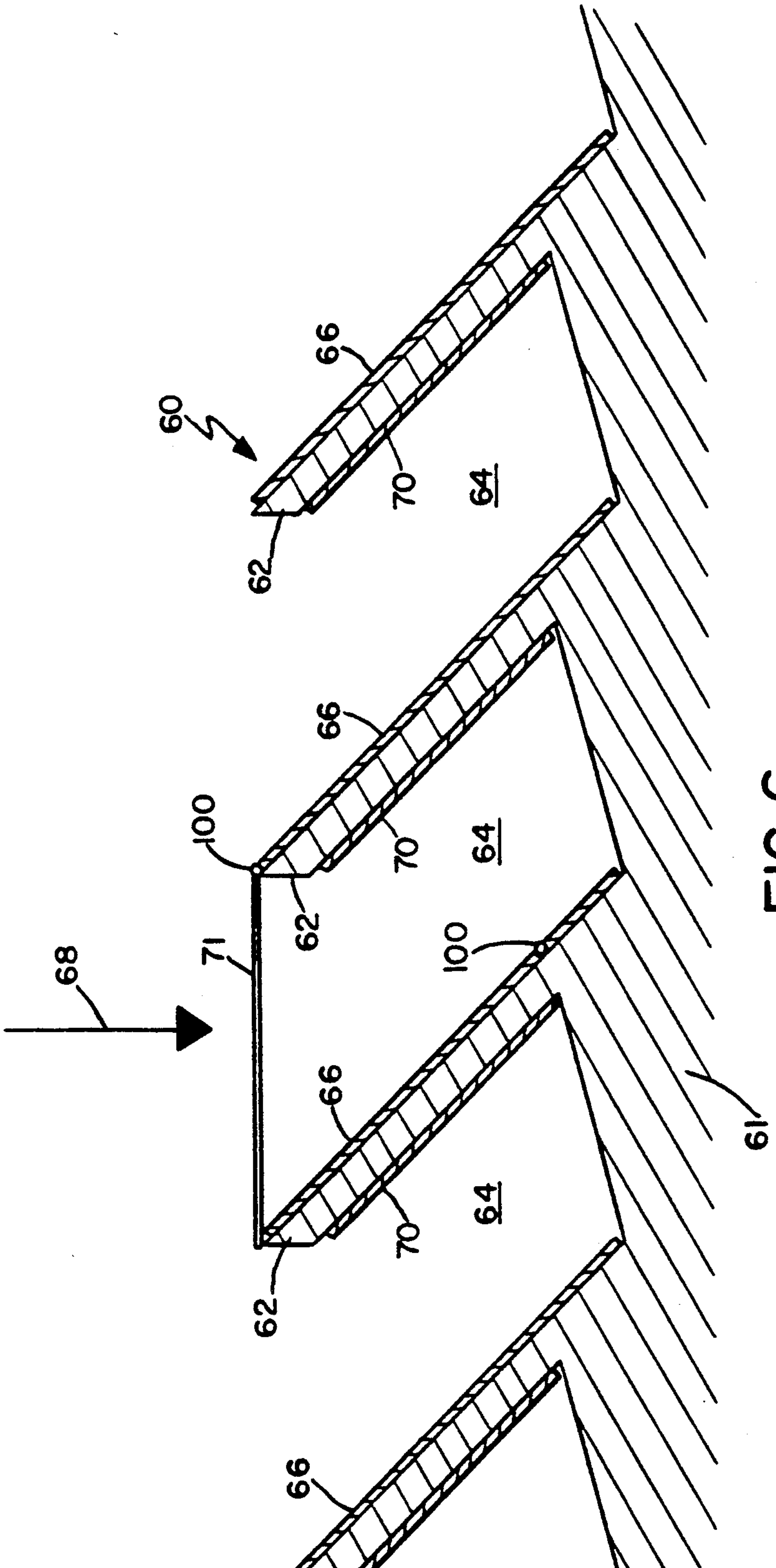
60
66
70
64
62
66
100
70
64
62
71
68
100
66
70
64
62
61
66
62

FIG. 6

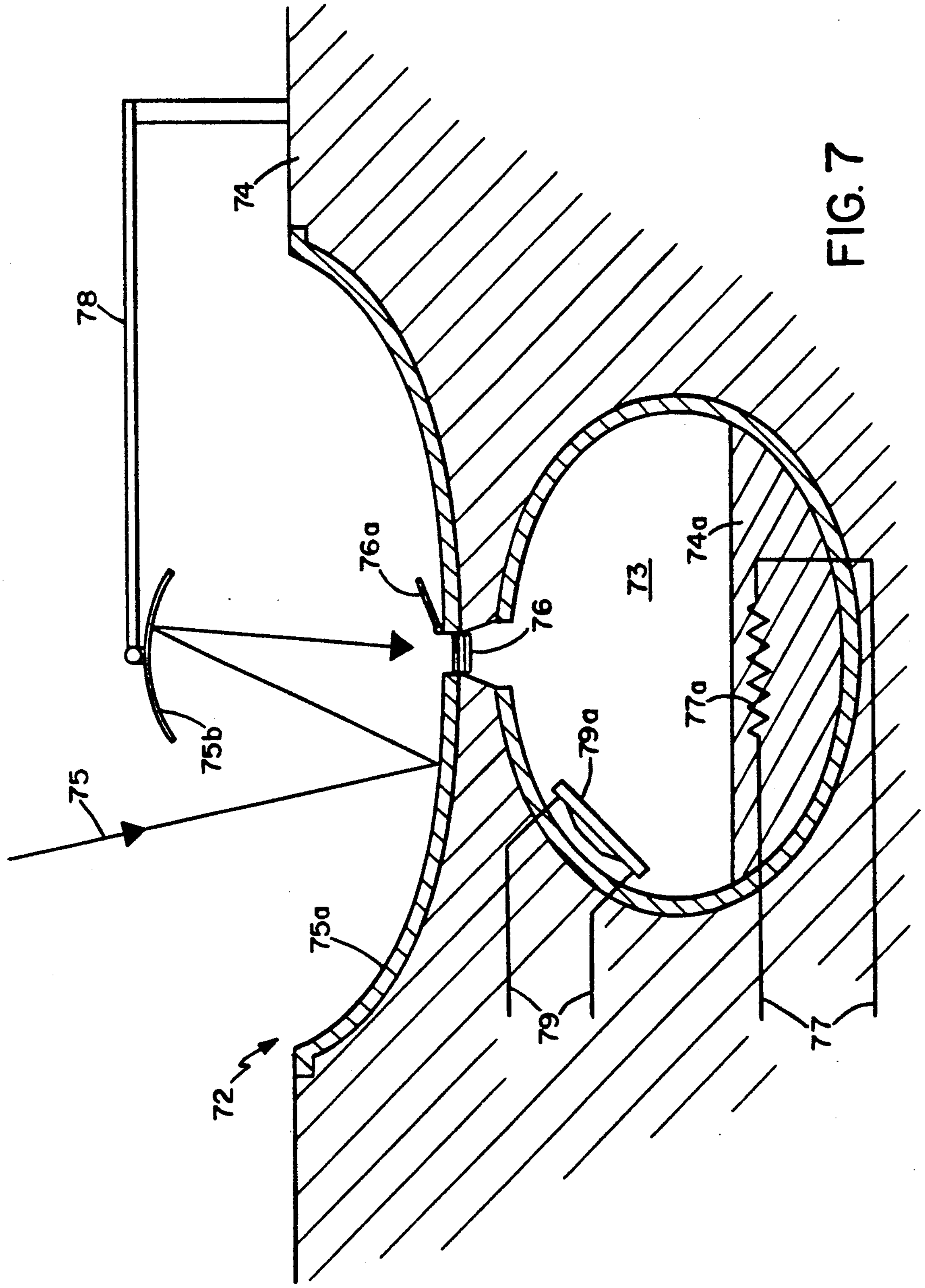
78
74
76a
76
73
74a
75b
75
77a
79a
75a
79
77
72

FIG. 7

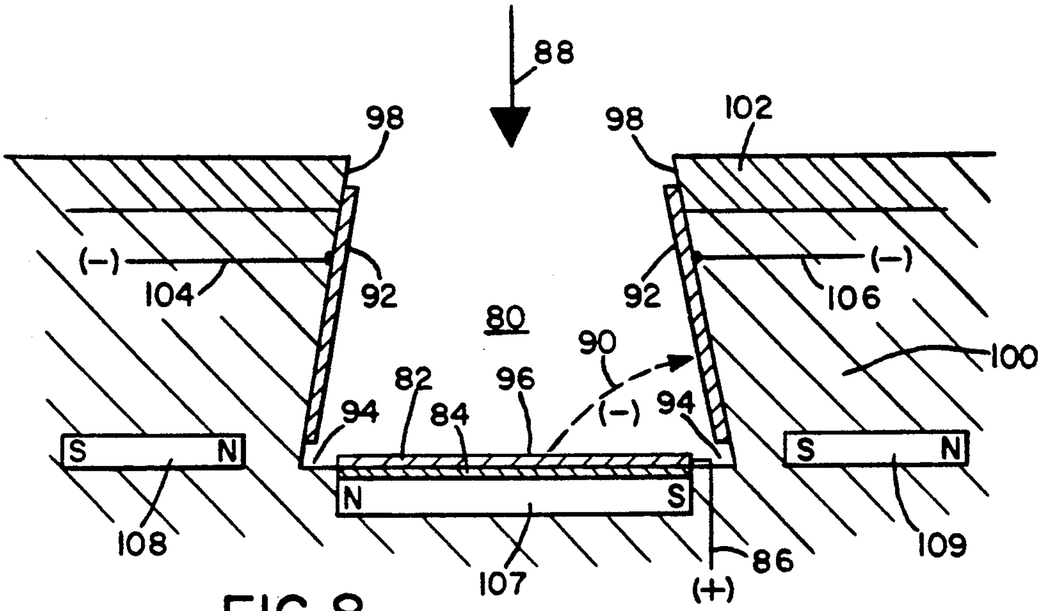
88
98
98
102
(-)
104
92
92
106
(-)
80
90
100
82
96
84
94
94
(-)
S
N
S
N
N
S
108
107
86
(+)
109

FIG. 8

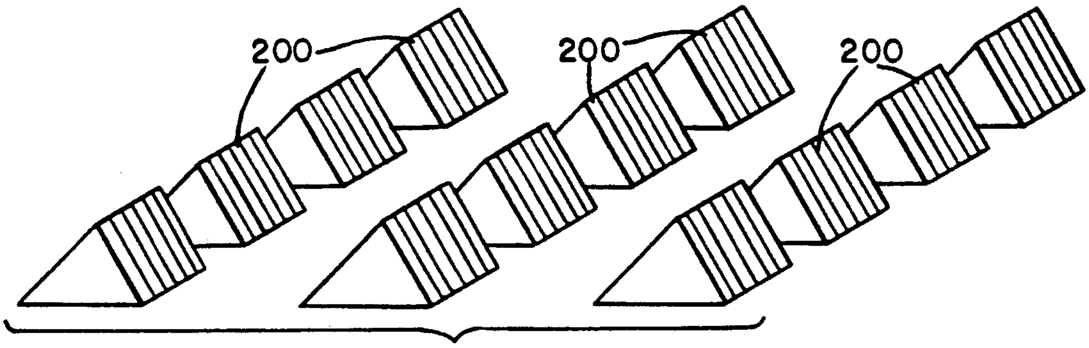

FIG. 9

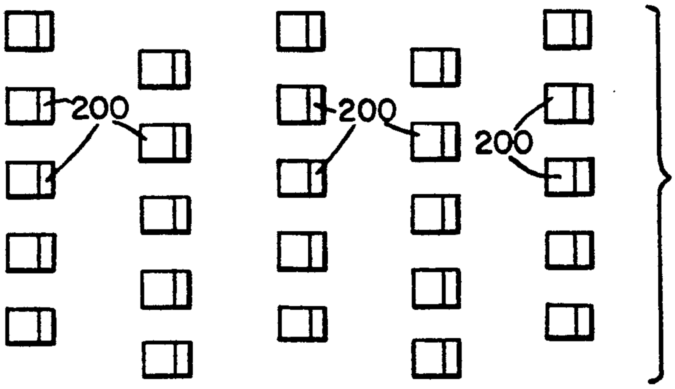

FIG. 10

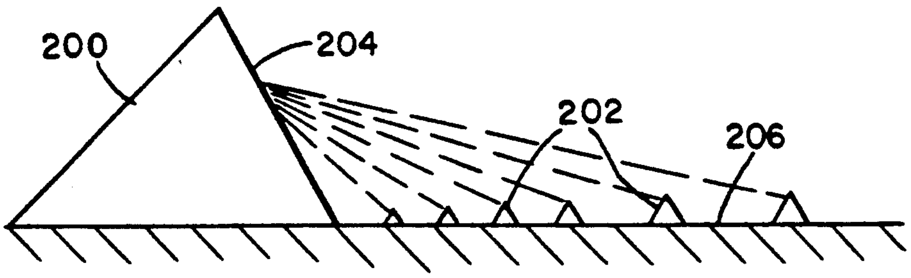
200
204
202
206

FIG. 11

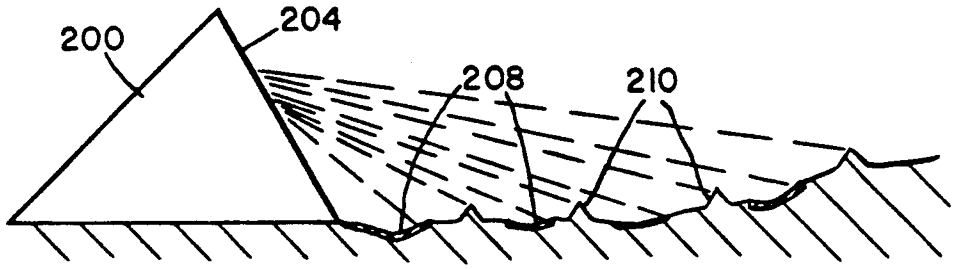
200
204
208
210

FIG. 12

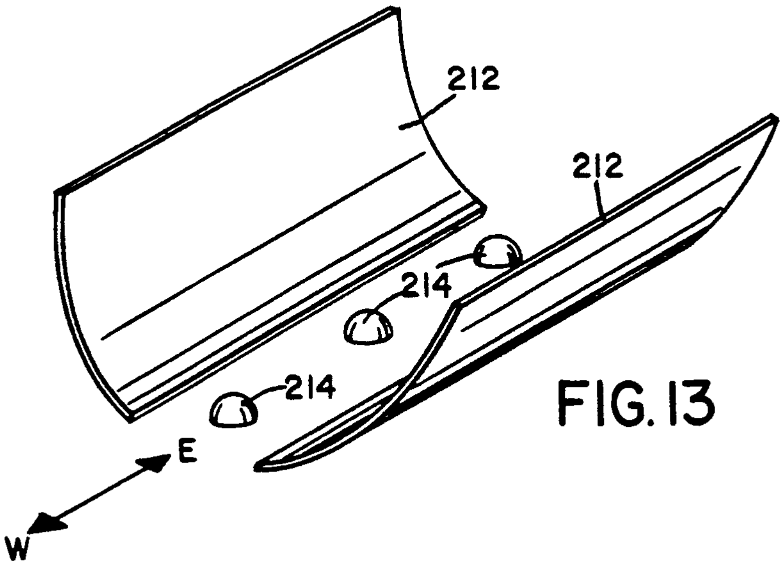
212
212
214
214
E
W

FIG. 13

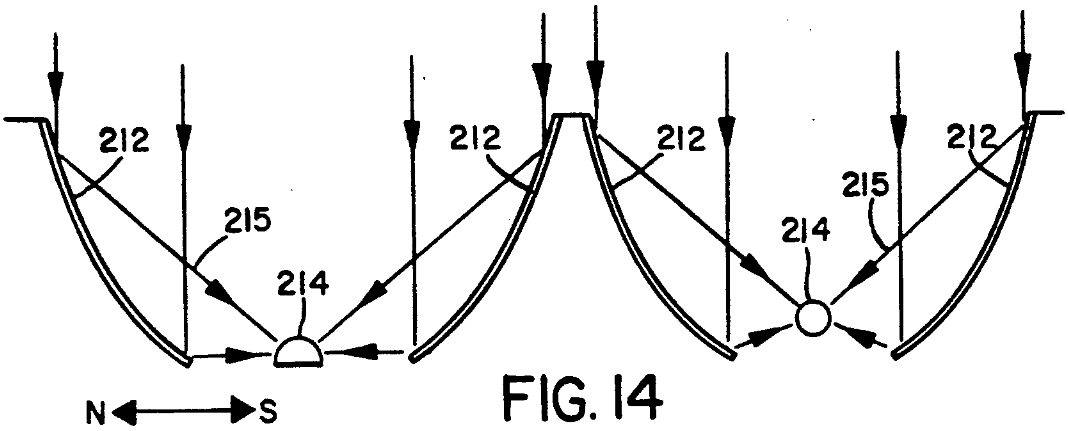
212
212
212
212
215
215
214
214
N
S

FIG. 14

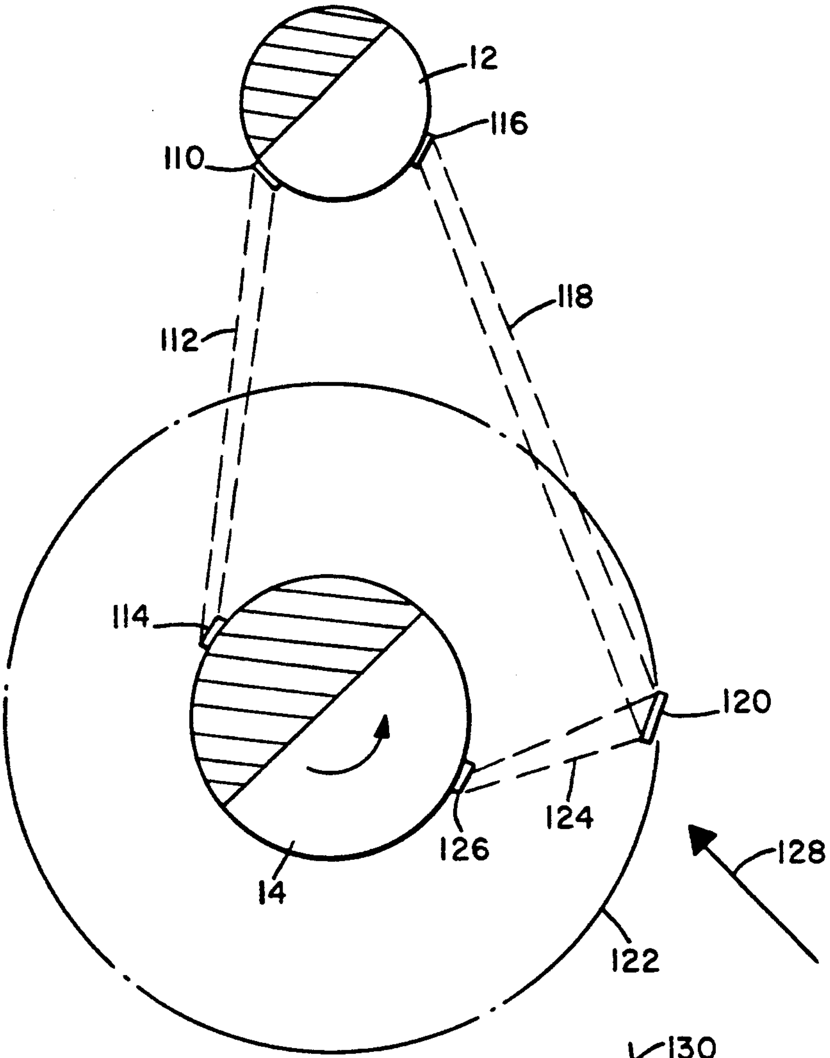

FIG. 15

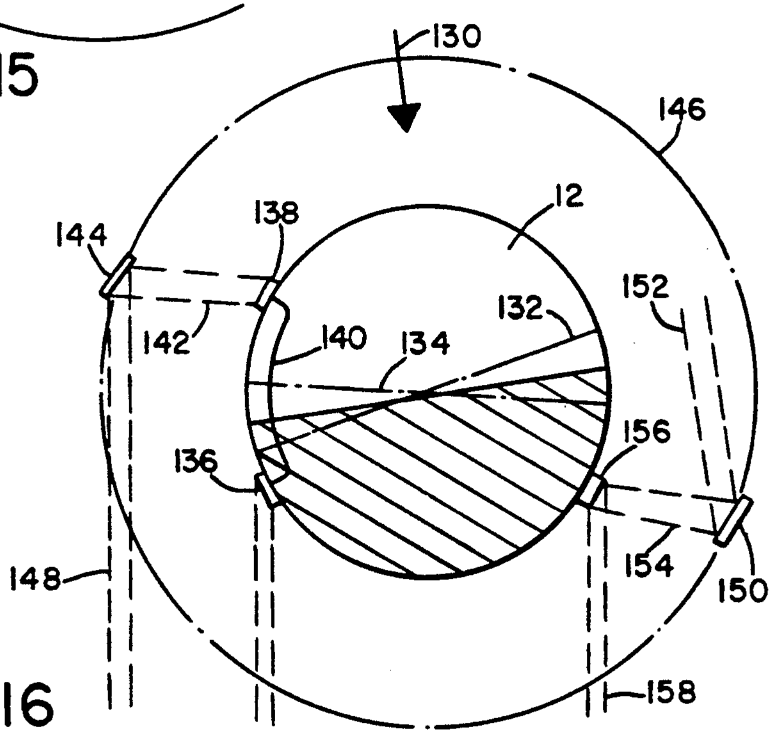

FIG. 16

POWER COLLECTION AND TRANSMISSION SYSTEM AND METHOD

This application is a continuation of application Ser. No. 07/157,081, filed Feb. 11, 1988, now U.S. Pat. No. 5,019,768, which was a continuation of application Ser. No. 06/732,352, filed May 8, 1985, now abandoned, which was a continuation-in-part of application Ser. No. 06/513,444 of the same applicants, filed Jul. 13, 1983, now abandoned, and entitled "Lunar Power System."

BACKGROUND OF THE INVENTION

The present invention relates generally to the collection and transmission of power, and is particularly concerned with the collection of solar power and its transmission in the form of a microwave beam.

Electric power supplies for most of the earth's inhabitants are currently provided by heat engines (steam turbines or the like) fueled by combustion of fossil fuels (coal, oil and/or natural gas). Progressive exhaustion of readily accessible fossil fuel sources, particularly oil and gas, has led to higher prices, shortages, and prospects for even more expensive and inaccessible fossil fuel supplies in the future.

Renewable energy sources including direct and indirect solar energy systems, the latter including hydroelectric, wind, ocean thermal, photosynthetic fuels (biomass), production of hydrogen and others currently furnish an important but minor fraction of the earth's electric energy. These resources should eventually furnish virtually the entire energy supply unless nuclear fusion energy is rendered economically and technically feasible in the future.

Most of the direct and indirect solar power systems are subject to geographic and climatological limitations. The solar power systems are also subject to variable and unpredictable changes or interruptions of output power. This has necessitated the use of reserve power systems (usually fossil fuel fired) or the use of restrictions on usage to only portions of the daily or long-term power usage cycles. These techniques unfavorably affect the economics of renewable energy use.

Despite these problems, ongoing development work is proceeding for making direct use of solar photovoltaic conversion systems, particularly for locations in latitudes with sunny climates. Current and near-term projected costs of conversion elements are far too high to be competitive with fossil fuel systems, even in the absence of power storage systems associated therewith. Coupled with this is the fact that the most advantageous applications of photoconversion are in economically underdeveloped and isolated regions on earth. There is hope that improvements in engineering and manufacturing methods may lead to systems which are cost competitive with future fossil fuel power systems.

There have been proposals in the past to collect solar power by means of earth orbiting satellites carrying solar collectors, and then to beam the power to suitable rectenna or receiver structures on earth, where the power is stored and distributed to users. Because of the many technological problems involved, none of these systems have yet been put into effect.

One proposal for such a system is described in U.S. Pat. No. 3,781,647 of Glaser, in which an orbiting satellite collects sunlight, converts it into electrical energy, converts the electrical energy to microwaves and beams the microwave power to earth from a planar phased array of microwave transmitters. One problem with this system is that it operates in the so-called "far field" or divergent area of the microwave beam, as is also the case in virtually all radar and high gain communications systems. The radiating near and far field regions of a wave field propagating from a radiating aperture (e.g. antenna) structure are separated by a near field transition length which is approximately equal to $L=D^2/w$ where D is the aperture diameter and w is the wavelength. In the radiating near field in Fresnel region ($L_n<L$), the maximum optical path difference from a field point to the distributed points of the aperture plane exceeds several wavelengths regardless of whether the field point is on or off axis of the system. Under such conditions, the radiated intensity is effectively averaged or smoothed for lateral displacement within the geometric optic beam and rapidly attenuates beyond the beam edge. Thus a beam is produced which closely approximates the geometric optics limit. In contrast, the radiating far field (Fraunhofer) region is divergent and has a maximum optical path difference of less than a wavelength between an axial field point and all points of the aperture plane. The radiated intensity decreases with lateral excursion from the beam center (with or without undulations depending on details of the aperture excitation.

Where an antenna system is used to transmit microwave power from a satellite, as taught by Glaser, the size of the antenna is limited, R is fixed, and the transition length L will be of the order of 10,000 km (for $D=1$ km and $w=10$ cm). Since geosynchronous satellites orbit at 36,000 km, the operating distance is more than three times the near field transition length for the above case. Thus the coherent microwave power of the beam will extend in a lobe-like Fraunhofer diffraction pattern across the earth. Because of the constrains on antenna sizes and operating frequencies for which sufficient atmosphere transparency exists, it would not be practical to operate in the near field in the Glaser system. The minimum beam diameter at the earth receiver would be about 3.6 km, and the approximate beam diameter would be 10 km. Thus the receiver must be very large in order to receive a substantial portion of the power adding considerable expense to the system.

While the Glaser system has certain merit in terms of providing the earth with needed electrical energy, there are major problems associated with the development of such a system. Primarily these considerations are both technological and economic. Technologically, there are many unanswered questions with regard to the feasibility and long term reliability of such terrestrial and space borne power generating systems. There are environmental problems restricting the total power which may be transmitted in this way. Such systems are also extremely expensive.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide an improved system for transmission of microwaves.

It is a further object of the present invention to provide microwave transmission system which operates in the radiating near field where the beam is substantially uniform and at least partially or slightly convergent and can be aimed more precisely at a relatively small receiving or rectenna structure.

According to one aspect of the present invention a microwave transmission system is provided which comprises an array of separate microwave radiators which are arranged to fill a radiating aperture of predetermined size when viewed from a microwave receiver assembly, each radiator emitting a separate microwave beam, a controller assembly for controlling the relative phases of the radiated beams to form a single, shaped composite microwave beam aimed at the receiver assembly, the receiver assembly being positioned in the radiating near field of the composite microwave beam, i.e. at a distance R which is less than about D^2/w where D is the diameter of the effective radiating aperture and w is the wavelength of the microwaves.

Preferably, the microwave transmission system is associated with a solar collector assembly comprising a plurality of solar collectors each associated with a respective one of the microwave radiators. Each collector preferably comprises an array of photoconversion devices and each microwave radiator may comprise an antenna array electrically connected to the collector array via a suitable underground grid, for example. Thus there is no single power source supplying power to all the individual radiators in the array, but there are a plurality of subassemblies each comprising a solar collector, a converter for conditioning and converting solar power to microwaves either directly or indirectly, and an individual microwave radiator. Thus if any one of the subassemblies should fail there will be little reduction in the overall power transmitted.

The controller assembly may be preprogrammed to shape and direct the composite microwave beam by suitable phase control of the individual radiators or microwave transmitters, e.g. by a controlled phase shifting network connected to the various microwave transmitters, as is well known in the development of phased array radars.

Such a system may be based on earth or extraterrestrially, for example on the moon.

According to another aspect of the invention a lunar power system may be provided which comprises a solar power collection device, means for converting solar power into microwaves, and an array of separate microwave transmitters or radiators which when viewed from a receiver structure on earth will appear to fill an aperture of predetermined size such that the receiver structure is in the near field of the microwave beam. A suitable phase controlling system is provided to direct and shape the beam, and this can be controlled to take into account the variations in the relative positions of the radiating aperture on the moon and the receiver system on the earth during the monthly and annual cycles of the two planetary bodies. The phasing is suitably controlled via integrated circuits as is done in phased array radars. Thus the array of microwave radiators is beam shaped to illuminate a fixed Earth receiver or rectenna site as the lunar elevation changes. The beam is steered electronically via the phase control mechanism.

By positioning the microwave antenna structure on the moon, the effective aperture can be made very large, and much larger than any satellite-borne antenna. Thus the beam can be aimed more accurately, the rectenna can be located in the near field, and more power can be transmitted.

Thus it is an advantage of this aspect of the invention, that a lunar based power system can be provided that minimizes the need for Earth facilities to redistribute power and the quantity of power beams passing through the atmosphere of Earth, thus reducing environmental problems.

The system may also be used to supply reliable power to mobile airborne, or space orbital structures and systems on the moon to permit operation of a number of civil, industrial, scientific or military operations at locations desirable for mission requirements.

The lunar power system may also be used to provide local power in space and on the moon to permit expansion and further industrial development in the vicinity of the power plant so as to eventually provide construction materials and manufacturing capability for use in space or for return to Earth.

The lunar power system may use native, beneficiated and refined lunar materials, with a minimal deployment of earth-supplied materials, equipment, and manpower. Large amounts of electrical power may be recovered from relatively simple structures made primarily of lunar materials emplaced on and in the lunar surface, and such power could be used on the moon, beamed back to Earth directly or by use of orbiting structures to redirect the beams. As a result of the above, large amounts of power could be returned to Earth without requiring the lifting of large masses from either the earth or the moon and without the necessity of constructing large, highly integrated, multipurpose space vehicles or other complex orbiting structures.

The solar energy receiving or collection device is derived primarily from materials available on the lunar surface. This receiving device preferably comprises separate arrays of photoconverters positioned at locations generally corresponding to the earthward portions of the east and west limbs of the moon's diameter. However other locations ae possible. The solar power collection device in one embodiment is a plurality of ridges of photoconverters, each of the ridges having a glass-like surface covering compacted lunar soil. Each microwave transmitter is associated with a predetermined plot of ridges, and a plurality of segmented microwave transmitters and reflectors are coordinated across the array so as to be convergent on a given target associated with a receiver structure normally situated on Earth. The receiver structure has the capability of converting the microwave beam into electrical power suitable for use on electrical grids on earth.

In alternative embodiments, the present invention further comprises a lunar reflector maintained in orbit generally about the moon. This lunar reflector directs the solar radiant power to the solar power collection device on the lunar surface. The present invention may also further comprise a lunar orbital reflector or retransmitter device maintained in orbit about the lunar surface for directing the microwave beam from the radiating aperture on the moon to one or more receiver structures which may be on earth, in space, or on the moon at other locations. The present invention may further include one or more earth reflectors maintained in orbit about the earth for directing the microwave beam to the receiver structure. Each earth reflector will switch from one receiver structure to another during its orbit about the earth.

BRIEF DESCRIPTIONS OF THE DRAWINGS

The above and other objects and advantages of the present invention will become clear from the following detailed description of some preferred embodiments of the invention, taken in conjunction with the accompanying drawings, in which like reference numerals refer to like parts and in which:

FIGS. **1** A–E show the arrangement of lunar power stations on the moon and the sequential delivery of power to microwave receivers on Earth in a power transmission system according to one embodiment of the invention. The sequence illustrated by FIGS. **1** A–E illustrates the operation of the lunar power station as a function of the moon's rotation about the Earth. This view is in a reference frame fixed to the centers of the moon and Earth;

FIG. **2** is a schematic representation of a lunar power station according to one embodiment of the invention;

FIG. **3***a* is a schematic representation of the independent power collection and transmission plots which can be phased to form a lunar power station and to cooperatively beam power in a selected direction from the moon;

FIG. **3***b* is a side view of the arrangement of microwave reflectors with respect to central transmission device;

FIG. **3***c* is a front view as seen from earth of the arrangement of microwave reflectors about the lunar surface, illustrating the effective transmission aperture;

FIG. **3***d* is a side elevational view of a microwave reflector of FIG. **3***a;*

FIG. **4** is a graphical representation of the power generation capability of the lunar power stations as a function of the position of the moon in relation to the sun.

FIG. **5** is a cross-sectional view of the ridge form of the solar converter in accordance with one preferred embodiment of the present invention;

FIG. **6** is a cross-sectional view of an alternative embodiment of the solar converter of the present invention. Specifically, FIG. **6** shows a set of Chevron-like structures fused into the lunar soil;

FIG. **7** is a cross-sectional view of an alternative embodiment of the solar converter of the present invention, which is a high temperature cavity form of photoconverter;

FIG. **8** is a cross-sectional view of an alternative embodiment of the solar converter according to the present invention. Specifically, FIG. **8** shows a photoelectron emission device impressed into the lunar soil;

FIG. **9** is a perspective view illustrating an alternative embodiment of the ridge form of solar collectors and converters, in which a plurality of spaced and offset rows of ridge-form solar collectors are provided;

FIG. **10** is a top plan view illustrating the staggered arrangement of the ridges in FIG. **9**;

FIG. **11** is a side view of an alternative embodiment in which a series of smaller ridges of gradually increasing size or spacing are located on one or both sides of each converter mound or ridge;

FIG. **12** shows another modification in which reflecting structures are provided in grooves on one or both sides of each ridge.

FIG. **13** is a perspective view of another alternative embodiment of the solar converter of the present invention, in which the reflector is in the form of a parabolic trough with a series of convertor structures spaced along the base of the trough;

FIG. **14** illustrates the projection of incident and reflected solar rays on the converter of FIG. **11**;

FIG. **15** is an alternative embodiment of the lunar power station of the present invention in which a microwave reflector is positioned in earth orbit to direct the energy beam from the moon to a desired location on earth; and

FIG. **16** shows an alternative embodiment of the present invention in which reflector units are placed in orbit about the moon so as to redirect either radiant energy to the lunar surface or microwave beams from the lunar surface.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

In all of the embodiments discussed below the provision of a power collection and transmission system on the moon for directing power to one or more receiver structures on the earth is described. However, it will be understood that the power transmission system may also be Earth-based, with suitable adjustments in the dimensions of the radiating aperture for transmitting power to an associated Earth-based receiver structure. In another alternative arrangement a lunar-based power transmission system may be associated with lunar, orbiting, or other non-Earth based receiver structures, either in addition to or in place of the Earth based receivers. The power transmission system need not be located on the moon but could be located on any suitable body in a sufficiently stable orbit about the Earth. Although the systems described below incorporate solar power collectors, other forms of power may be used for converting to microwave power, for example, fossil fuels.

The various embodiments of the invention illustrated in the drawings relate to the transmission of power in the form of a directed, convergent microwave beam **1** from one or more transmission systems **2** to the remote receiver devices **3**, the receiver devices being located in the near field of the microwave beam.

In the preferred embodiment of the invention one or more lunar power stations are provided for collecting solar power, converting it into an energy form suitable for transmission to earth, and transmitting the energy in a uniform and at least partially convergent beam directed at one or more receiver structures on Earth. The energy is preferably transmitted as a short, high frequency radio wave or microwave, but shorter wave visible, ultraviolet or infrared optical beams generated by coherent laser sources may alternatively be used.

FIGS. **1**A to **1**AE illustrate the positioning of two lunar power stations **16**, **18** on the moon **12** relative to receiver structures **20**, **22** and **24** on Earth during various phases of the moon. Each power station will comprise solar collector assemblies, converters for converting solar power to microwaves, and microwave transmitters or radiating antennas to transmit one or more microwave beams **26** towards the receiver structures on Earth. The converters may include devices for power accumulation, release and conditioning.

As shown in FIGS. **1**A to E, the solar power stations are located on the east and west limbs of the moon **12**, respectively. The receiver structures comprise microwave receiving units **20**, **22**, and **24** positioned about the surface of the earth. The microwave receiving units are positioned so as to receive the transmitted microwave beams from the transmitters on the moon. The operation of the lunar power system in relation to the relative motions between the moon **12** and the Earth **14** will now be described.

The operation of the lunar power station is sequentially illustrated by FIGS. **1** A–E. In these figures, a view of the moon **12** and the Earth **14** is shown from north of the ecliptic plane. The shaded areas are the night side of the two bodies. The moon is phase-locked with the Earth so that the same general side of the moon always faces the direction of the Earth. The moon rocks about the line between the centers by about plus or minus 7° or arc.

FIGS. 1A and 1E show the lunar power station system **10** of the present invention when the moon **12** is full as viewed from Earth. When the moon is full, the solar collecting assemblies in both power stations **16** and **18** collect solar radiant power since the sun acts on both units. The incident solar radiant power is converted to electrical power in D.C. or low frequency AC form. This photon to electricity conversion can be accomplished with direct photovoltaics, thermo-photovoltaics, heat engines, or other devices now known for such conversion. Some or all of the electrical power is reconverted and transmitted in the form of a coherent photon beam, such as microwave beam, to a single target on earth. This single target is microwave receiving unit **22** (FIG. 1A) located at a position on the moonward hemisphere of the earth. The microwave receiving unit **22** then converts the received microwave beam into energy usable in another form, such as electrical energy. As can be seen in FIG. 1A, the signals from the transmitters in power stations **16** and **18** are converged and directed toward the target area on Earth, as will be described in more detail below. While FIG. 1A shows the beams **26** as slightly divergent from the moon **12** to Earth **14**, it is preferable that beams **26** are convergent to a location on Earth **14**. The receiving area on Earth would not necessarily have to be larger than the transmission aperture on the moon. As illustrated in FIG. 1A, the amount of energy received and transmitted can be twice that available from a single unit of a given size located on the moon.

FIG. 1B shows the moon **12** in its last quarter. At this point in time, the solar collecting assemblies in power station **18** are in darkness while the power station **16** is exposed to solar radiant power. Since station **18** is not exposed to sunlight, no power is transmitted from unit **18** to Earth. However, each power station may be provided with suitable energy storage means, as described in more detail below, to permit transmission during periods of darkness. The transmitted microwave signal **26** shown in FIG. 1B is directed toward microwave receiving unit **22** on the Earth's surface during a portion of the time unit **22** is on the moonward side of the Earth.

FIG. 1C shows the lunar power station system **10** at new moon. No power is generated in this arrangement since both power stations **16** and **18** are in darkness. Hence, no power is transmitted in microwave form from the moon **12** to earth **14**. However, the power stations may include suitable energy storage devices to permit power to be stored during daylight periods for transmission during periods of darkness.

FIG. 1D shows the lunar power system **10** when the moon is in its first quarter. At this time, the power station **16** is in darkness and receiving unit **18** is exposed to sunlight. Receiving unit **18** transmits the solar energy in the form of a microwave signal **26** to microwave receiving unit **20** on Earth **14**.

Finally, FIG. 1E shows full moon **12**. As in FIG. 1A, both solar energy receiving units **16** and **18** focus and direct their microwave transmission **26** to receiving unit **20** on Earth. As shown in FIG. 1E, multiple beams **26*a*** and **26*b*** may be transmitted from a given unit such as **16** to two or more different receiving units, e.g. units **20** and **22**. As can be seen in a comparison of FIGS. 1A and 1E, the transmitter on the moon is synchronized so as to deliver power to the receiving units generally on the moonward side of the Earth. This will depend on the relative rotation of the Earth with respect to the moon. The receiving units on the far side of the Earth relative to the moon at any time may also be associated with suitable energy storage devices so that continuous power is available at any time. Many suitable storage methods are known, for example, pumped hydrostorage or compressed air stored in underground caverns.

FIG. 4 is a graphical representation of the power output curves for the lunar power station system **10**. In this graph, the total power generated by the lunar power station system is represented by the dotted line. The solid line represents the power that would be received by photoconverters of equal area placed flat on the lunar surface. This graph generally shows the power output as a function of the phase of the moon. As can be seen from FIG. 4, the maximum power is generated when the moon is full. This summing of transmitted power to high levels occurs in free space in the electromagnetic fields. The minimum power occurs near new moon. Between one or two days past new moon and just prior to full moon, the total power generated is relatively constant for a preferred collection geometry such as is shown in FIG. 5, for example (described in detail below). However, if continuous power generation is needed throughout the lunar cycle, one must either not transmit all the available power, storing the non-transmitted power for example by batteries, superconducting electrical coils and/or capacitors, flywheels, thermal or gravitational energy storage on the moon, or illuminate the solar collectors near new moon by using orbital sunlight reflectors or by relocating some of the collectors past the visible limb on the far side of the moon as viewed from Earth.

In the preferred embodiment of this invention, as illustrated in FIG. 1, the arrangement is to locate one transmitting array **18** on the east limb of the moon and another array **16** on the west limb of the moon. This will generally serve to provide a beam of power **26** toward Earth receivers **20**, **22** and **24** during most of the lunar month.

The specific form of each lunar power station **16**, **18** in the preferred embodiment of the invention will now be described. FIG. 2 is a schematic representation of one of the power stations during construction, and FIGS. 3*a* to 3*d* are schematic representations of solar power collectors and microwave transmitters in the power station. Each power station will comprise a plurality of separate solar collecting areas **50**, which need not be of equal size. Each collecting area is associated with a converter assembly for converting solar power into microwaves, and one or more microwave radiating or transmitting assemblies such as transmitters **54** and radiating antennas **59** for transmitting a microwave beam. Each transmitter **54** may comprise one or more individual microwave transmitters and may transmit a plurality of microwave beams. The transmitters may comprise microwave diodes, for example, which can be designed to transmit at least two simultaneous microwave beams. However, any suitable type of microwave transmitter may be used. In the preferred embodiment, the individual transmitters do not direct their transmitted beams in an earthward direction, but are associated with microwave reflectors or wavefront management devices **59** for directing the beams towards one or more receiver units on Earth. Each transmitter may be associated with one or more reflectors. The microwave transmitting system comprising the various microwave transmitters and associated reflectors is in form of an extended quasi-planar transmitting array. The relative phases of the transmitted microwave beams are controlled to form a composite microwave beam of predetermined shape and direction.

With the lunar power system shown, the receiver units can be located in the radiating near-field region of the transmitted composite microwave beam, and the various transmitted beams from the individual transmitters in the array can be controlled, for example by phase shifting integrated circuits as used in phased array radars, to shape and direct one or more composite microwave beams and to allow for the changes in the lunar elevation and the relative positions of the transmitter and receiver units as indicated in FIG. 1.

By positioning the transmitting system on the moon, there is little or no limitation on the overall size of the transmitting array. The radiating near and far field regions of a propagated wavefront are separated by the transition length, L, which is approximately given by:

$$L=D^2/w$$

in which D is the diameter of the radiating aperture and w is the wavelength. For microwaves, w is in the range from about 10^{-3} m. to about 1 m. For a wavelength of 10 cm. and a transmitting array diameter of 10 km, the transition distance to the far field is therefore 10^6 km. Thus, the Earth would be well within the near field of the transmitted composite microwave beam, allowing relatively precise directing and at least partial convergence of the beam onto a relatively small receiver unit. Because of the large lunar area availability, it would be feasible to construct much larger antenna arrays on the moon, for example of 100 km or more diameter, making the distance to the far field equal to 10^8 km. The Earth would then be even further within the near field, with consequent increased efficiency in power reception.

An oversized antenna aperture shared by many beams affords many methods for controlling multiple beams and the characteristics of any given beam such as diffraction limit, side lobe levels, position of beam center line, depth of field, and location of the focal region. The oversized array on the moon can use all the modern techniques of phased array electronics and geometric optics to control many separately directed beams simultaneously projected by each segmented array on the moon to form a shaped, convergent, composite beam. Thus, both small and large receivers on Earth or in space can be equally efficient in the reception of power. A 100 km diameter lunar antenna operating with 10 cm microwave would have a minimum central spot size at Earth of approximately 400 m. Most rectennas would be 2 or more times larger. A few hundreds of Megawatts of power could be received by a rectenna only a kilometer across.

FIG. 3 is a schematical representation of one embodiment of the lunar power station **16** or **18** incorporating an array of microwave transmitters located on the lunar surface. Generally, solar collector and converter devices or other energy sources are grouped into individual areas indicated by plots **50**. Electrically conducting wires **52** collect electrical power from the solar collectors and converters and conduct the electrical power to a central transmitter **54** associated with each plot **50**, which suitably comprises a plurality of microwave antennas. No particular wiring geometry is implied by wires **52**, as illustrated in FIG. 3. The general area **56** surrounding the array of microwave radiating assemblies, as shown in FIG. 3A, comprises a radiating or transmitting aperture. Either solid state or conventional means of generating coherent microwaves can be employed at transmitters **54**. Each transmitter **54** will normally comprise one or more individual microwave generators, such as microwave diodes, for example. Each plot **50** can be independent and separate from its neighbors for electrical collection. Electricity collection units and transmitters **54** in a given location may be interconnected by sub-surface wires **57** and switches. In this way, reliable transmission of power can be achieved even in the advent of a need to replace or repair solar collectors in a given plot **50**.

Each microwave transmitter **54** may be capable of producing more than one microwave signal. Transmitter **54** may contain microwave lenses and other devices for beam shaping. Multiple beams **26** may be generated by a given collection of radiating assemblies. These beams may be individually directed to different targets. The phasing of transmission is controlled by timing or reference signals from an external source so as to transmit a coherent composite microwave beam to microwave receiving units **20**, **22** and **24** located on earth **14**, or other locations in outer space.

In the preferred embodiment, an array of microwave transmitters **54** is provided which does not view the earthward direction of the transmitted beam. Instead, the energy from each microwave transmitter **54** is directed in a beam or beams **54***a* approximately in the opposite direction to one of a plurality of large microwave reflector surfaces **59** as shown in FIG. 3*b*. Each plot of collector assemblies **50** has one or more associated microwave reflectors **59**. Alternatively, the structures **59** may comprise microwave refractors and beams **54***b* may radiate through the refractors. Combinations of both techniques may be used.

The reflector is a structure substantially transparent to incident solar radiation, but which reflects microwaves. This arrangement permits the microwave radiating aperture to be interleaved with the solar power collectors while minimizing the blockage of sunlight to the collectors. FIGS. 3*b* and 3*c* illustrate this arrangement in relation to the lunar surface. The microwave reflectors **59**, when viewed from positions in the distance portions of the microwave beams will lie approximately within an area **56**. The reflectors **59** will cover enough of the area **56** when viewed from positions in a transmitted beam to define a transmitting aperture for the microwave beam. In FIG. 3*c*, the reflectors are shown as touching or slightly overlapping so that they fill the area or aperture **56** nearly completely. However, this is an exaggerated case and the reflectors need not overlap or touch as shown in FIG. 3*c*, but in fact the the total reflector size may be much less than the total area of the aperture **56**. The area **56** defined by microwave reflectors **59** has a generally circular shape as viewed in projection along beams **16** or **18**. This projected or multiple surface of microwave reflectors **59** produces a tighter and smoother beam than is possible with isolated small microwave transmitters.

Thus, a microwave transmission system is provided which comprises a plurality or array of separate microwave transmitters or radiators, whose virtual images approach or overlap when viewed from the microwave beam so as to present an apparently substantially contiguous radiating aperture **56** which may be 100 km or more in diameter. The aperture is preferably between about 100,000 and one million times the wavelength of the transmitted radiation diameter. Such a system may be used in other, Earth-based applications as well as in the lunar power system described with suitable reduction in the size of the array.

The use of a greatly expanded total lunar projected aperture and the operation in the geometrical optics—Fresnel diffraction region permits use of a truly convergent power beam with substantially uniform power flux density across the wave front which may be modified to allow differential convergence in transverse axes to shape the beam cross-section to match that of a fixed Earth receiver area. As the lunar elevation changes, the convergence may be accomplished by deactivating some of the array elements or transmitters **54** to change the active aperture perimeter **56**, by phase adjustment of radiating assemblies **54**, or by a combination of both these methods.

The polarization of the transmitted beam or beams may be controlled by one of two methods:

1. The polarization of the prime transmitting or radiating structures may be rotated by rotating said structures or by adjusting the phase and amplitude of 2 independent orthogonal sources.

2. The polarization may be changed from linear to circular or vice versa by using 2 orthogonal parallel wire reflecting layers spaced by about 1/8th wave length and fed by linear or circular polarized prime radiating structures respectively.

The Earth receiving units must effectively receive and rectify the transmitted beam over the range of azimuth, elevation and polarization planes characteristic of the process. The first two factors depend on time of transmission while the latter is adjustable by methods just described.

Each microwave radiator or transmitting assembly **54** in the preferred embodiment comprises one or more individual transmitters and a microwave reflecting surface or wavefront management device for reflecting microwaves from the or each transmitter to a remote receiver unit. Alternatively, an array of separate microwave transmitters may be provided which themselves direct the beam in the direction of a remote receiver unit.

FIG. **3*d*** shows an individual microwave reflector or wavefront management device **59**. Its appearance generally resembles that of a billboard. This microwave reflector **59** is constructed with foamed glass uprights **59*a*** and foamed glass horizontal members **59*b***. Uprights **59*a*** are anchored to the moon by pilings **59*c***. Crossed threads **59*d*** and **59*e***, formed of lunar glass or derived silicon, are extended taut across the frame so as to form a nearly flat or slightly curved surface. Threads **59*d*** and **59*e*** are coated with a thin film of electrical conducting metal, such as aluminum. This metal can be chemically refined from anorthositic lunar materials. The spacings between threads **59*d*** and **59*e*** must be significantly smaller than the wavelength of the photons used in the transmission system. The open space mesh reflectors will transmit at least 80 to 90 percent of the incident sunlight but reflect radiation of wavelength exceeding the average wire spacings by a factor of 5 or more.

It is possible to introduce more flexibility into power beaming by eliminating the fixed supports **59*c*** and by mounting the microwave reflector **59** on support **59*f***. Support **59*f*** is designed so as to rotate the reflectors **59** in azimuth and altitude. By phase controlling each of the individual microwave transmitters **54**, a wide range of beam transmission directions can be produced through the use of support **59*f***. The surface of the microwave reflector may be curved to aid in forming a beam.

The wavefront management devices **59** need not be reflectors as shown but may be of many different varieties. They may be lenses through which the microwave beam passes modified by controlled refraction. The wavefront management devices may be composed wholly or in part of active or passive phase delay elements. Hybrid systems with reflectors, lenses and active and passive multiple units may be used.

The wavefront management devices **59** may combine in their functions that of the transmitter **54**. Devices such as microstrip phased array transmitters can be made as large area devices which essentially fill the transmission aperture.

The wavefront management devices **59** in FIG. **3*c*** need not be placed in regular order but can be distributed pseudo-randomly so as to minimize power transmission in side lobes near the central beam direction. In addition, pseudo-random placements can permit transmitters which are less area-filling for given side lobes amplitude levels of the power beam.

The separate plots **50** of solar power collectors supplying power to the array of transmitters may be repeated over large areas of the lunar surface. They may be located across from each other on the lunar limbs or on the far side of the moon. They may be contiguous for the most part. In the preferred embodiment each plot is associated with a substantially central transmitting assembly **54**, but the transmitters may be positioned remote from the solar collector plots.

The radiating or transmitting assemblies **50** inside a given circle or radiating aperture, such as area **56** can be brought into phase so that the individual transmitters can act as a phased array which can generate one or more beams of power directed to one or more stations on Earth or to redirectors or reflectors in orbit about moon or Earth. Either signals from the earth or preprogrammed commands produced by phasing systems in space or on the moon can be used to define which ones of the plots **50** will perform as a coordinated array or aperture for a given time. Individual elements can be switched in and out during the operation of the array to change the shape or direction of the composite beam, for example, the center **58** of the transmitting array can be moved during transmission to change the path of the center of a given beam. The beam can be made convergent for a given target in the Earth-moon system because the effective aperture of the array is governed by the radius of the area **56** and can therefore be much larger than the aperture of the individual microwave transmitters **54**. The individual plots **50** do not have to be contiguous. However, as reflectors **59** fill up the projected area of the beam, then more photon energy is placed in the beam.

Each power station may include more than one transmitting array or aperture **56**. Different arrays can overlap so that the median of two areas **56** can contain plots **50** which may be side-by-side but are sending to different targets. The individual plots **50** do not have to be of the same size or the same power output. In addition, these plots **50** can be upgraded or repaired individually without having to shut down the entire system.

The array boundary **56** projected along the beam direction does not have to be circular, as in FIG. 3*c*, but can assume other shapes which are optimal for beam formation and focusing. More than one physically distinct array **56** can transmit to a target. The shapes of the individual plots **50** do not have to be square and the units do not have to be in contact. As an illustration of this concept, it is conceivable that an entire hemisphere of the moon could be used to produce a single beam if the individual units were present over a significant fraction of the lunar surface. The reflective surfaces or wavefront management devices **59** near the center of the lunar disk, as seen from earth, would be horizontal or almost horizontal to the local lunar surface. Transmitters **54** could be suspended above the reflectors **59** in towers.

The general physical orientation of transmitters and reflectors or wavefront management devices is somewhat dependent on their location on the moon and the direction of beaming from the moon. Horizontally oriented reflectors could be partially or fully covered by lunar soil to provide protection against temperature change, radiation and other disturbances. Temperature changes that occur on the lunar surface can induce physical dimensional changes and potential physical damage. Reflectors arranged generally parallel to and close to lunar slopes could also be partially or completely covered with lunar soil.

By operating with an at least partially convergent beam profile in the near field wave region, commonly described as the "Geometric optics—Fresnel diffraction" region, it is possible to control the cross-sectional area, shape and direction of the convergent microwave beam by altering the aperture perimeter by suitable switching of individual transmitters **54**, by adjusting the phases of transmission of the various transmitters, or by a combination of both of these methods. Thus, it is possible to ensure that the transmitted beam will fall on a fixed area of the Earth, or other remote receiver structure, in spite of the rotation of the Earth as viewed from the moon and orbiting of the moon relative to the Earth as long as the receiver site is on the moonward side of the Earth and not too close to the Earth's limb as seen from the moon.

As in standard phased array radar techniques, each element or radiator in the microwave array emits a quasi-spherical wave and the plurality of superimposed microwaves will combine constructively or destructively according to the phase differences. By suitable control of the signal phase the single composite microwave beam can be steered. Suitable phase shifters are used to shift the relative phases of the beam, and these may be microprocessor controlled.

The arrangement of two lunar power stations on the east and west limbs of the moon, as seen from the Earth, is used because of the fact that a given lunar station has useful illumination only when the sun's elevation is above some limiting lower value and hence less than 50% of the month or the year and similarly an Earth station could receive the transmitted power only when the lunar elevation exceeds some minimal value. So that continuous power can be delivered to a given Earth station, it thus becomes necessary to provide efficient, economical power storage or to add other operating elements to the system to extend operating time.

The cost of energy storage is composed of power dependent and energy dependent terms. Providing storage for the 15 days or more that a lunar base is in shadow or low elevation sunlight would be generally uneconomical. However, by using two lunar bases located close to the east and west equatorial limbs of the moon but displaced toward the sub-earth point sufficiently that lunar libration will not lower Earth's elevation angle below a minimum value needed to present needed aperture projected areas, the lunar bases could sequentially and cooperatively furnish power to one or more Earth receiving sites whenever the sites had sufficient lunar elevations, except for a short period near new moon. When both bases would be in shadow, the power interruption period would be total for about 1½ Earth days and the period of combined reduced and zero power would last up to about 3 days.

Provision may be made for energy storage at a single lunar power station for up to 3 Earth days' output or at both lunar stations for up to 1½ Earth days' output. This could provide for essentially continuous power beaming to a succession of Earth points following the Earth's rotation, except during lunar eclipse.

Energy may be stored either as thermal energy mechanical energy (e.g. kinetic energy of a flywheel or other structure), electrochemical energy (e.g. storage batteries, fuel cells), or gravitational potential energy, or electromagnetic energy (capacitors/inductors).

A possible gravitational potential energy storage method on the moon which would be equivalent to earth-based pumped hydrostorage methods would use lunar solids such as soil or rock. Two possible storage systems comprise mechanically transporting a mass of lunar soil from a lower to a high elevation during periods of solar illumination and recovering the energy by returning the mass to the lower elevation at night or low power periods. The transport mechanism may comprise a conveyor or pulley system, or a catapult method.

The period of zero or reduced power can also be lowered by siting the centers of solar power collecting plots **50** farther toward the lunar limbs than the centers of beam transmitting arrays at the expenses of increased lengths of power transmission lines (For every 370 km of combined displacement, the reduced power period is shortened by 24 hour). If the power conversion sites extend partially past the visible limb, it is possible to eliminate the periods of zero or reduced power.

Using either or both of the preceding methods to assure continuous lunar generation capability reduces the maximum duration of energy storage needed for a given Earth station to a fraction of a synodic day (24.84 hr) during which the lunar elevation is below a minimum useful value. For an equatorial Earth base, the storage duration may range from about 13 to 16 hours depending on cut-off angle, while for an Earth station at 45 degrees latitude, the average storage duration will be about 16.7 hr with an extreme value of about 19.3 hrs. Any conventional form of energy storage may be used. Alternatively, pumped hydrostorage, compressed air storage, or a combination of both, may be used. In one example, compressed air may be forced into a sealed reservoir which lifts the water to the surface. During energy recovery, the compressed air filled reservoir is expanded through a turbine to atmospheric pressure. The water is then allowed to flow to the lower reservoir through water turbines.

In addition to or instead of energy storage on Earth, orbital beam reflectors may be used to extend Earth base operation.

Some preferred forms of solar power collector and converter structures constructed in plots **50** to provide microwave power to the transmitter arrays will now be described. However, it will be understood that microwave power may be provided from other sources in alternative embodiments.

Collected solar power may be converted to microwaves either indirectly, by conversion of solar power to electricity and then from electricity into microwaves, or by conversion of thermal energy into electrical energy, or directly by conversion of sunlight into microwave photons for example by photoklystron devices.

FIG. **5** shows a first embodiment of an individual solar power collector and converter structure on the moon **12**. Specifically, this is a ridge-form of solar converter which can be formed on the moon **12** and used to receive solar radiant power from the sun. While this is the preferred embodiment of the conversion unit to be constructed on the moon, it should be noted that this is only one of a variety of possible devices for the conversion of sunlight to electricity or directly to microwaves (e.g. a photoklystron) which can be constructed on the moon primarily from lunar materials. A unique feature of the ridge-form of solar collector and converter structure **28** is that it uses the geometry of exposure to sunlight to even out the electric power production in the lunar day.

As can be seen in FIG. **5**, the lunar soil **30** is seived, smoothed, and formed into mounds **32**. A series of mounds **32** are aligned in the general north-south direction to cover the desired solar collecting area. These mounds **32** are nominally separated in the east-west direction by a distance sufficient to prevent excessive shadowing of each ridge by the next ridges in the lunar morning or evening.

The mounds or ridges need not be continuous but may comprise spaced rows of short ridges as shown in FIG. **9**. A thin layer **34** of native glass is applied to the ridge to bind together the sloping soil **32** as a cohesive layer. A thin layer **36** of iron or other electrically conductive, non-reactive (chemically) layer is vapor-deposited over the glass layer **34**. The iron layer **36** serves as the mechanical substrate and electrical pickup for a thin-film solid state photoconversion layer **38**. Photoconversion layer **38** is made of silicon. Other materials can also be utilized including those which could be imported from the Earth, (e.g. titanium, germanium, lower oxides of titanium, cadmium sulfide, lower sulfides of copper, selenides or tellurides of zinc, cadmium or copper, gallium arsenide, gallium phosphide, indium arsenide, manganese, iron, chromium, nickel or cobalt). A thin gridwork of iron **40** is vapor-deposited on top of the photoconversion layer **38** so as to complete the electrical circuit. A thin layer **39** of glass may be placed on top of photoconversion layer **38** so as to protect against solar wind, hydrogen atoms, and other forms of contamination. Electrical pickups are attached to the iron surfaces beneath the surface of the lunar soil **30** by leads **42** and **44**. Leads **42** and **44**, in turn, connect into the local power grid, as can be seen in FIG. **3**, for connection to associated microwave transmitting assembly **54**, which may comprise one or more microwave generators and antennas. The lunar iron in this solar converter structure **28** is acquired by beneficiating it in the form of micron-sized iron grains from the local soil, thermally separating it from adhering glass or minerals, and optionally electrorefining it on the moon. Silicon for use in the solar arrays can be obtained from the local lunar soil following the necessary chemical processing by means of machines brought from earth.

The ridges preferably take approximately the cross-sectional profile of an equilateral triangle. The spacing between the ridges is not critical. Such structures will effectively average the useful illumination and power collection over the period of lunar daylight.

The lunar environment is particularly suitable for the development of solar collectors and converters of this design. The lunar soils **30** are extremely good electrical and thermal insulators. There is no free oxygen or water in the lunar environment to damage, disturb, or interfere with the development and utilization of the solar converters. Finally, since the moon is seismically stable, the arrangement of ridges should remain free of significant geological disturbance.

The electrical output of a given inclined collector system can be augmented by diffusely reflected light. The contribution of diffusely reflected light will be greatest at local noon when the ground surface is at its brightest. The combined direct and diffuse illumination can be partially compensated by increasing the base angle of the ridge to greater than 60 degrees. The adjacent lunar soil surface reflectivity may also be increased by spreading a thin layer of white powder or glass material onto the surface, which can increase diffuse reflectivity to 90%. If this is applied for a distance of 3 face lengths in front of each ridge face and the base angle is also increased to 66 degrees, the noon illumination can be raised by about 24% and the average illumination may be raised by about 19% over solar elevations above 6 degrees.

FIG. **2** illustrates schematically an example of the construction of a lunar power station of the type shown in FIG. **3** made up of a series of areas or plots **50** of ridges as shown in FIG. **5** each associated with a microwave transmitter assembly and a reflector **59** of the billboard type. An automatic tractor **230** may be used to plow the lunar soil into a series of ridges. Power is brought by a grid of underground electrical wires from each ridge in a plot **50** to a collection of many solid state microwave transmission elements **54** at the center of each plot. The transmission elements **240** in a given plot illuminate the associated reflector **59** which is situated at the edge **232** of the plot or area remote from the Earth. The transmission elements in each plot are phased by local absolute clocks or a fiber optics network laid underground between the various transmitter sites in the array. FIG. **5** shows a ridge-form of solar collector and converter suitable for a lunar environment. However, there are other types of solar converters which would be suitable for the lunar environment. These types of solar converters could also make use of available lunar materials. As examples, alternative forms of solar collectors and converters are shown in FIGS. **6** and **7**.

FIG. **6** shows a chevron-like form **60** of solar collector and converter. Specifically, FIG. **6** shows a cross-sectional view of trenches created in the lunar soil by melting a small skin of lunar soil to form mechanically stable chevrons. The lunar soil **61** and the fused soil **62** form the body of this form of solar converter. The compacted lunar soil may have been stripped of various minerals and elemental components. The trenches **64** are of indeterminate length and can be adjusted to the most appropriate angle, size and length for the desired solar converter. A thin film of aluminum **66**, refined from lunar materials, is deposited on the spaceward facing surface so as to reflect direct sunlight **68** toward the back of the facing **70**. The back of facing **70** contains a photoconverter. Electrical attachments from the associated microwave transmitter to the photoconverter on facing **70** can be made through the soil **61** and the fused wall **62** by wires placed in the trench wall prior to the fusing of the soil. Connections between cells can be made by wires placed in the soil **61** under the converters and mirrors, thereby avoiding any contact with the plasma environment just above the lunar surface. Thin transparent cover sheets **71** can be placed over trenches **64** to provide additional isolation against contamination, solar wind, and cosmic rays. These transparent coversheets **71** are of the type which are less transparent in the infrared spectrum. These sheets **71** can also be used to smooth thermal changes at sunset.

In an alternative arrangement (not shown) the solar collectors and converters may be in the form of a series of V-shaped ridges and trenches. To produce power levelling in this type of structure, one side facing of each trench need not be flat but may be in the form of a series of short triangular ridges when viewed from above, similar to those shown in FIGS. **5** and **9** but rotated by nearly 90 degrees. The opposite side facing is a reflective surface forming a virtual image of the sun nearly in the horizontal plane, so triangular ridges with nearly vertical apex edges will experience nearly the same illumination sequence as the ridges of FIGS. **5** and **9**.

The trenches for the solar converters and mirror surfaces do not have to be open structures. An elaboration of the design is to make deep narrow cavities in which multiple light reflections are used to direct the solar energy to completely enclosed photoconverters. In this manner, a microenvironment can be created deeper in the lunar regolith to protect photoconverters from damaging photo and particle radiation, debris, sudden major temperature changes, etc. This shelter could be made predominantly from lunar material with simple thermal processing. The mirror surfaces do not have to be planar, but can be complicated shapes and even be optical gratings so that spectral splitting can be accomplished.

Mirrors **66** and optical cover **71** may be movable by means of thermally active hinges **100** or other solar tracking devices to allow one axis tracking of solar reflectors in orbit about the moon. They are preferably arranged to move so as to orient the surfaces as closely as possible to positions perpendicular to the incident solar rays throughout the lunar day.

Another form of the solar power collector and photoconverter structure is illustrated in cross-section at **72** in FIG. **7**. This structure **72** comprises a groove of any chosen length in the lunar surface connected to a high-temperature cavity **73** made of beneficiated native materials, such as anorthite or refined compounds or elements (e.g., Al_2, O_3 or Si). Cavities **73** can be buried under thermally insulating bulk or beneficiated lunar soils **74**. These cavities **73** can be partially filled with basaltic-like native or derived materials **74***a*. Materials **74***a* are of the type which melt at a significantly lower temperature than the materials forming cavities **73**. Solar energy **75** can be directed into these cavities **73** via simple reflectors **75***a* and **75***b*, either as concentrated sunlight (through windows **76** or other openings) or by electrical current through lines **77** and resistive heating element **77***a* during the times that regions **16** or **18** are illuminated. Since the axis of the moon is almost perpendicular to the ecliptic, relatively simple two-dimensional concentrators **75***a* and **75***b* with limited focus control **78** can achieve high concentration levels.

Power can be extracted from cavity **73** during the day and night by many means. Thermal energy could be extracted to provide process heat or drive turbines. High conversion efficiency output can be obtained by placing thermophotovoltaic solid state converters **79***a* in the cavity **73**. These thermophotovoltaics **79***a* are manufactured to convert photons, just off the peak energy density of the radiation in the cavity, into electricity. Non-converted photons are recycled and reequilibrated in cavity **73** for possible later conversion to electricity. Power is extracted through leads **79**. Cavity **73** can be thermally isolated at night by a cover sheet **76***a*. This cover sheet may be repositioned either automatically or mechanically by bimetallic devices over the entrance aperture. The cavity **73** may be permanently sealed by a glass lens or plate **76**.

This approach is practical on the moon because the lunar soil is an extremely good thermal insulator. Additionally, the lunar soil can be a source of supply for the components of the chamber and the molten material used for energy storage. Operating problems are reduced by the moisture and oxygen-free environment.

A similar approach may be used for thermal stage of power in systems of the type shown in FIGS. **5** and **6**. A large underground cavity of the type shown in FIG. **7** may be connected to the electrical grid of wires associated with one or more of the collector/converter arrays, and have a suitable heater for heating molten material in the cavity.

Another form of solar power collector and converter structure is illustrated in FIG. **8**. FIG. **8** shows in cross-section a structure which is suited for construction on the lunar surface from primarily native materials. A series of small, spaced trenches **80** are made in the lunar surface and a thin layer of photoemissive material **82** is vapor deposited over a thin film of aluminum **84** in the base of each trench. A thin film of aluminum **84** is connected by an electrical lead **86** to a larger grid of photoconverters. Sunlight **88** ejects photoelectrons **90** from the material. A small fraction of the photoelectrons **90** impact on either one of two anodes **92**. Anodes **92** are electrically separated by area **94** from the cathode **96**. Anodes **92** are made by vapor depositing a thin film of aluminum or other electrically conductive material on the slightly inward vertical walls **98** of trench **80**. The inward slope prevents direct illumination of the anodes **92** which suppresses back currents. The vertical walls are mechanically stabilized by fusing the soil **100** in a thin skin **102** after forming of the trench **80** in the soil **100**. The electron current flow is extracted by means of leads from the cathode **96** and the two anodes **92**. The leads **104** and **106** each introduce the power from this photoconverter section into the larger grid containing the array of microwave transmitters as in FIG. **3**.

In an alternative version of the photoconverter of FIG. **8**, the photoemissive layer **82** is vapor deposited on electrically separated small patches of electrical conductors in the bottom of the trench **80**. These conductors could be formed as small patches with smaller separating regions of non-photoemissive surfaces. The patches extend across the width of the trench. The width of each conductive line is approximately equal to the gyroradius of the photoelectrons (now perpendicular to path **90**) in the magnetic field of strip **107** which have energies of approximately 0.1 of the energy of the most populous of the differential photoelectron population of the emitted electrons. The electrons are made to have a gyroradius by layering native lunar irons in thin strips **107** under the patches or between the trenches **108** and **109** and magnetizing the iron layers so that a small magnetic field is impressed across the trenches parallel to the long axis of the photoconductive layers. Thus, most of the photoelectrons are made to step from one electrically conductive island to the next in the same direction down the long axis of the trench.

The cathode of the trench is the trench-end which the photoelectrons are moving away from, and will be the electrically positive lead end. The anode is at the other end of the trench. The device becomes more efficient as the separation distances between the edges of the islands decrease so that fewer electrons are impacted on non-conductive (electrically) soil. The efficacy of this approach is demonstrated by the fact that the magnetic fields on the lunar surface produce large potential differences by focusing the flow of solar wind electrons to the lunar surface and influencing the exact path of photoelectron return currents.

In another modification of the embodiment shown in FIG. **8**, multiple grids may be placed above the photoemissive layer **96** and biased electrically so as to force the photoelectrons to oscillate between the grids. This device is a so called photoklystron which permits the direct conversion of sunlight into microwave energy.

FIGS. **9** and **10** illustrate another alternative embodiment of the solar collector and converter structure. In this example, instead of spaced single continuous ridges, a series of rows of mounds or ridges **200** of the general type shown in FIG. **5** are provided. Each row is aligned in a generally north-south direction and separated from the next row in the east-west direction. As shown in FIG. **10**, the rows are staggered so that each mound in one row is aligned with the gaps between the mounds in the next adjacent row or rows. This can result in an equal or higher average illumination area for a given ground area.

The ground adjacent to ridges of the type shown in FIG. **5** or FIG. **9** may be provided with diffuse or specular reflection structures to increase average power collection and conversion. This is illustrated in FIGS. **11** and **12**. In FIG. **11** areas to the east and west of each mound of ridge **200** are provided with small ridges **202** of gradually increasing size. The ridges may be spaced so that a mid-point on the collector face **204** would see approximately equal fraction of the horizontal and inclined surfaces defined by the ridges **202**. White powder or glass material is applied over the surface of the ridge **area 206** so as to perform as an artificial "snow bank" or white sand surface. This raises the ground brightness and allows sunlight to be reflected from the ground area **206** onto the adjacent ridge or ridges. This arrangement may be arranged to produce an average ground brightness nearly independent of solar elevation.

Alternatively, as illustrated in FIG. 12, areas immediately east and west of the ridge faces **204** may comprise alternating specular and diffuse reflectors **208**, **210**. Specular reflectors **208** may comprise thin glass sheets or flakes with evaporated films of iron, nickel, aluminum, or other metals.

Alternate methods include many parallel layers of clear glass or molded prismatic glass sheets with V-shaped parallel grooves on the back side whose faces are inclined approximately + or −45 degrees to the planar front face. Horizontal specular reflectors can be conveniently prepared in-situ by applying radiant or other heat sources to smoothed lunar soil until it melts. Most lunar soils produce low viscosity glasses which level easily while the excellent thermal insulation qualities of the soil limit the depth of melt and serve as a container for the soil. After heat removal and solidification of the glassy melt, the surface may be metallized by vacuum evaporation of iron, aluminum, or other metal. The diffuse reflectors comprise layers of white powder to produce a snow bank effect.

Another example of suitable solar power collector and converter structure is shown in FIGS. **13** and **14**. In this example spaced split parabolic troughs **212** of specular reflectors lying along an east-west axis are prepared. The elevation of the bottom of the trough is conveniently set between about 30 degrees and 45 degrees while the elevation of the upper edge depends on the concentration ratio desired.

The projection of incident and reflected rays **215** on a north-south plane is independent of solar elevation and is shown in FIG. **14**. To achieve nearly uniform output power, solar converter structures such as photoconverters **214** comprising domed partial spherical structures or nearly equivalent polygonal solids are spaced along the base of each trough. The projected areas of the converters are nearly independent of direction. The east-west spacing of converter domes is established by minimum solar elevations for which no ray interferences is encountered, perhaps typical values of 3 to 10 converter diameters may be used.

Only a small fraction of the rays **215** striking the troughs reach the target converters at any one time. If power leveling is not needed or desired, one may use target converters of full length strips, cylindric or prismatic segments so that all rays which strike the trough will intercept the converters.

Solar radiation incident on the trough surfaces and reflected onto the converter surfaces will be suitably converted into electrical power for connection to microwave transmitters in the transmit array, as described above in connection with FIG. 3.

FIG. **15** illustrates an alternative system for transmitting microwaves from the moon to Earth. In this system, a microwave reflector or beam redirection device is provided in orbit about the Earth to direct microwave beams from the moon to receiving stations at high altitudes or angularly distant on Earth from the sub-lunar point. As shown in FIG. **15**, a power station **110** located on the earthward side the moon **12** beams a microwave (coherent photon) beam **112** toward the earth. The beam **12** is reflected onto a receiving station **114** on earth. At this receiving station **114**, the power in the photon beam **112** is converted into electrical current. On the opposite limb of the moon, a lunar power station **116** beams microwaves **118** to a microwave reflector or redirector device **120** in orbit **122** about the earth **14**. In this case the redirector device is a microwave reflector **120** which reflects beam **118**, as beam **124**, to a receiving station **126** on Earth. However, it may comprise any form of wavefront management device such as lenses, reflectors, active or passive phase delay devices, or reception and retransmission systems which re-transmit at the same or different frequencies to the input microwave beam. Device **120** may transmit more than one beam. It may transmit beams in directions corresponding to the angles implied by simple geometric reflection. It is important to note that as the number of receivers **126** and orbital microwave reflectors and re-directors **120** increase, more power can be transmitted to Earth.

The orbital device can be arranged to cooperate collectively with each other to direct power to all Earth stations at the level needed at a given time.

The rotation axis of the moon **12** is almost vertical to the ecliptic plane and makes an angle of approximately 1.5 degrees off the vertical. The rotation axis of the earth **14** is inclined approximately 23 degrees from the plane of the orbit of the earth about the sun. The direction of solar radiation is indicated by arrow **128** in FIG. **15**. The orbit **122** of the reflective mirror or re-director device **120** is established so as to most advantageously direct the microwave beam to the receiving station **126**, or to other users or devices in the vicinity of Earth. The orbit **122** is not necessarily a geosynchronous orbit or even in the plane or the equator of Earth. High inclination orbits can be established. The lunar power station **110** or **116** can supply power to the earth only when illuminated unless energy storage is provided. The earth is approximately fixed in the lunar sky of a given station. The lunar station is capable of directing the microwave beam to a given station on the earth as the station turns into view or directing the beam to reflectors in orbit about the earth **14**. Use of re-direction devices such as reflectors potentially enhances the ability to direct the microwave beam, to enhance the power transmission, and to deliver a more even flow of power to all stations on Earth from a minimum set of stations on the moon. The orbiting redirection devices or reflectors can become basic elements of a planet-wide power distribution system in which the power needed only enters the biosphere once.

Multiple re-direction devices or reflectors may also be used in orbit about the Earth. Power can be reflected or re-directed two or more times to reach users on Earth or in space. The reflectors **120** may also be configured as retransmitters. They could operate by receiving power at one frequency, converting it to another frequency, and then retransmitting it.

Another embodiment of the present invention incorporates the use of orbiting sunlight reflectors about the surface of the moon **12**. These orbiting reflectors can serve to increase the exposure of the lunar receiving units to sunlight and to increase the power transmission capabilities of the lunar power station. FIG. **16** is a view of the moon **12** from above the North Pole. The sunlight **130** is coming from the upper left and illuminating the moon hemisphere away from Earth. The dotted lines **132** and **134** define the extreme limits of the limb of the moon **12** as seen from the Earth. Lunar power station **136** is visible from the Earth at all times, whereas lunar power station **138** is invisible from the Earth at all times. The stations **136** and **138** are not necessarily on the equator of the moon and can be located nearer either pole than the equator and do not have to be distributed about the same latitudes. This figure reveals various ways that the power streams to the Earth can be smoothed in time.

The electrical energy at station **138**, as high voltage AC or DC, can be transmitted to station **136** by means of high voltage transmission line **140**. The power stream from the moon **12** to the Earth **14** will be smooth if two sets of these stations are located across the east and west limbs of the moon. At least one station of the four will always be illuminated and can send power to the corresponding earthward station for transmission.

The station **138** can direct a microwave beam to a microwave reflector mirror or other re-director device **144** in lunar orbit **146**. Microwave reflector **144** can then reflect the beam **142** into an earthward direction as beam **148**. Also, optically reflective mirror **150** in lunar orbit **146** (not necessarily the same orbit as for microwave mirror or re-director device **144**) can reflect sunlight **152** to the lunar surface as beam **154**. The sunlight is directed to the earth-side power station **156**, but may alternatively be directed to stations such as **138**. More than one mirror **150** can be used to illuminate a given lunar power station. From station **156** the microwave beam **158** is generated and directed toward Earth. Orbit **146** does not have to be circular or in the equatorial plane of the moon. Both types of redirector devices or mirrors **144** and **150** may be controlled by reaction engines, gravity gradients, tethers, moment of inertia changes or radiation pressure to point in the appropriate directions at the appropriate times to reflect beams. Device **144** and **150** could be the same type of re-director device, such as a mirror. Many mirrors and orbits can be used at the same time about the moon so as to minimize the need for multiple bases.

The variety or embodiments within the concepts as illustrated in FIG. **16** may be included to enhance the capabilities of this lunar power station in accordance with the present invention. The systems as shown in FIGS. **15** and **16** provide the capability to distribute power on a planet-wide scale so as to follow the diurnal patterns of power usage on Earth. Since the power only enters the biosphere once via the beams, the primary power production system can be minimized. As a result, this system provides for lowest costs and quickest completion. Intense power can be received almost anywhere on or above the earth via this system. It is not strongly dependent on local terrain, natural resources, or even the availability of land.

The use of a power station on the moon offers a number of advantages not found in prior actual or proposed extraterrestrial power systems. Of particular importance is the fact that the lunar power station could be constructed from materials found in abundance in the lunar soil. Even in the early stages of development, it would be possible to use substantial quantities of available lunar material for both the lunar power station and the production equipment used to build the station. One useful chemical element is iron, which can be separated by magnetic means from the bulk of the lunar soil with minimal complexity and minimal energy. This iron can be used for mechanical, electrical and magnetic functions. Other elements, such as the attached glass and assorted minerals may also be separated by several presently available techniques. The iron is present as tiny spheres and globes, many in the one to fifty micron diameter range, in both the soils of the lunar mare and highlands. Thermal heating can separate the iron from the glass using solar concentrators. It is advantageous in some applications to electrorefine the iron to remove minor amounts of nickel and cobalt. Because the lunar soil is pulverized, and due to the low lunar gravity, relatively little mechanical work is required to remove the iron metal from the local soil. The lunar soil and its derivatives may be used to provide structures, electrical and thermal insulations, transparent covers, and many other uses.

It should be noted that the various subsystems of a lunar power station as described above do not strongly interdigitate, as is the case with previously proposed satellite solar power stations (SPS). For example, SPS orbiting structures (microwave and/or optical collectors) must be optimized in mechanical properties to allow accurate pointing of portions toward earth targets and also to point another massive component toward the sun and, in addition, support interconnections between the power sources, the solar cells, and the mutually moving continuous antenna array. In the present invention, the solar conversion devices do not have to be optimized for minimal mass and maximum efficiency for use on the moon. In the present invention, these systems can be designed to be produced primarily out of the lunar feedstock with supplementation from minor-mass elements which may be needed from Earth. There is a minimal restriction to the use of solid state conversion devices for the lunar application. This lunar power station permits much more compact power receptors on earth than does the SPS system, may have improved reliability and is more immune than a satellite from meteoric of other space debris damage. It can also be readily modified and expanded with improving technology. Considerations of production and costs per watt become of direct relevance in the selection of the conversion device. Costs of transport from Earth and development of mass efficient devices (i.e. of low kg/kw) are of secondary concern since costs/lb are practically immaterial for mass sources already available at the point of use.

Advantages of solar based power stations as opposed to Earth based systems include reduced periods of low solar exposure during Earth winter. The synodic period for the moon is approximately 29.53 days. The lunar station is independent of cloudiness and is not exposed to such extreme weather conditions.

There are various other features of the lunar environment which are particularly conducive to the construction of lunar based solar power stations. These positive environmental factors include: (1) the very high vacuum and absence of disruptive chemical effects on thin-films; (2) the extremely stable mechanical environment of the lunar surface and its remoteness; (3) the complete dryness, high electrical resistivity and thermal insulating properties of the lunar soil which make it possible to build electronic components on and in the soil protect them from thermal changes, radiations, deliberate damage and plasmas; (4) vast areas are available for constructing large areas of converters and transmitters; (5) the moon is an exceptionally stable platform orbiting the Earth in a very regular and known manner; (6) chemically-free iron is available for the early stages of construction; (7) lunar derived materials can be used to support many phases of the space operations which support both construction of the lunar power system and other projects; and (8) simpler solar concentration systems due to the small lunar obliquity (1½ degrees). The Apollo, lunar science research, Skylab and shuttle programs have established the major portions of the research and development needs for the present invention. The tools necessary for high rates of production on the moon are within the state-of-the-art of present industrial capabilities.

Variations to this general approach are likely applicable to other natural objects in the solar system (e.g., Mercury, various asteroids, and the moons of Mars). Various non-microwave means of power transmissions are also clearly applicable (e.g., charged or neutral particle beams, lasers, etc.). On Earth, several different approaches to power reception are possible. These have been proposed for use in conjunction with the SPS. Such approaches to power reception could include heat engines driven by laser or particle beams. Photovoltaic arrays can be situated under the terrestrial rectenna elements of the microwave system.

For efficient power reception at the receiving units on Earth it is necessary that the amount of power reflected be minimized which requires that the gain and impedance of the receiving array antennas be properly matched to the input waves.

This can be done by individually rotating and aligning half wave or shorter dipole antennas of other directional antennas to proper directions to face the incoming power beams, but this is more costly than systems which use fixed structures.

For days in which the lunar maximum elevation is high and for non-equatorial Earth locations, it may be desirable and possible to collect power for lunar excursions of 120 degrees to as much as 180 degrees or more in celestial longitude (Right Ascension). Most fixed antenna systems cannot operate efficiently over a wide range of input angles. An exception exists for dipole or monopole antennas whose axes are perpendicular to the equatorial plane. These antennas have pattern symmetry about the polar axis but require input radiation polarized with electric vector parallel to the polar axis for maximum efficiency. This may be obtained by rotation of plane of transmission on the moon by +/−22 degrees to +/−25 degrees during a monthly cycle of operation. Actually the departure of a fixed direction of polarization (For example on the lunar or ecliptic pole) from ideal polarization is not serious and would result in a loss of about 7 to 9% of beam energy representing the orthogonal fraction.

A second alternative is to transmit a circular polarized wave which may be received using circular polarized omnidirectional antennas such as loop and monopole or slotted cylinder types with polar axes. For such systems, no adjustment of transmission polarization is required.

A further possible modification of receiver antenna systems is to use optical concentration by reflecting structures to raise the beam intensity and lower the actual active antenna aperture areas needed to process and rectify the power beam. While this would be a relatively simple matter for SPS Systems in which the beam direction does not change, it requires careful design for the lunar system with its variable beam direction. Most, if not all, high concentration ratio optical systems suffer severe spherical aberration errors if used over a large range of beam angles. One system previously described in the art which performs one axis concentration without measurable spherical or cylindrical aberrations is a circular cylinder with periodic reflector strips arranged along its inner surface. The angles between the strip change by only about one-fourth of the angular arc measured on the cylinder cross section. Input parallel rays are brought to a line focus on the projected extension of the cylinder, regardless of input direction. Use of such a concentrator requires either continual repositioning of the image receiver except when the moon is in the equatorial plane or else simultaneous rotation of the reflector strips. The reflector strips may be curved to correspond to surfaces on a torus to produce partial convergence and concentration along the other axis, but this will introduce some residual aberrations.

Beam concentration could result in radiation intensities which may be dangerous to bird and insect species in the receiver sites and require application of repellent techniques.

Beam reflector strips can be made of very low cost, open mesh reflectors.

The proposed lunar power system can use the moon, its environment, and its orientation in the earth-moon system as "gifts of nature", like coal or iron ore, to increase the net wealth of mankind with minimal destruction of natural resources.

The power transmitting system which permits transmission of power in a controllable microwave beam by means of an array of separate microwave transmitters is not limited in usefulness to a lunar environment but may also have applications to power transmission and receiving systems on Earth, in association with suitable power sources such as an assembly of solar power collectors.

A well defined microwave beam is possible because of the large transmitting aperture which may, for example, be synthesized from any appropriately positioned separate microwave transmitters, or combined transmitters and reflectors which completely or nearly completely fill the projected apparent aperture when viewed from a suitable receiver. Each reflector may serve as the reflection surface for hundreds, or even thousands, of individual microwave beams. When the segmented reflectors are used as part of a lunar power station, they are preferably arranged in an elliptical area near lunar limb so as to apparently overlap a filled, circular transmitting aperture 10 km to 100 km across when viewed from Earth.

To a first approximation, the segmented lunar antenna can be modeled as a set of spatially non-periodic radiators. The number (N) of radiators will be in the millions. The average intensity of stray power of a given beam will be 1/N of the intensity of the beam. It is conceivable that the stray microwave power of the LPS beaming 10,000 GWe could be less than a few billionths of a watt per square centimeter (few nanowatts/cm.**2) at Earth. This is a power level 10 to 100 times less intense than the human body radiates as incoherent power in the microwave region. If such a low level of incoherent stray power can be attained, then there is no biological risk from exposure to microwave outside the main beams. It may be possible to provide economic power via the lunar power system at beam intensities so low that heating of animals is negligible. Power would be cheaper at higher beam intensities.

Although some preferred embodiments of the invention have been described above by way of example, it will be understood by those skilled in the field that modifications may be made to the disclosed embodiments which are within the scope of the invention as defined by the appended claims.

We claim:

1. A solar power collection and transmission system comprising:

a plurality of solar power collecting assemblies for collecting solar radiation incident on said solar power collecting assemblies;

means defining a microwave radiating aperture of predetermined area comprising an array of separate, spaced apart microwave sources for emitting a plurality of separate microwave beams, and structurally separate microwave redirector means for redirecting the emitted microwave beams in a predetermined beam direction and expanding the effective microwave emitting area to substantially fill the microwave radiating aperture;

conversion means associated with each solar power collecting assembly for converting said collected solar power into a power form suitable for connection to said microwave sources;

each solar power collecting assembly and associated conversion means being connected to at least a respective one of said sources;

phase control means associated with said microwave sources for controlling the relative phase of microwave beams emitted from said sources to form a composite, shaped microwave beam directed from said microwave redirector means, the composite beam having separate radiating near field and far field regions; and

at least one microwave receiver assembly positioned in the radiating near field region of said microwave beam.

2. The system of claim 1, further including:

at least one receiver assembly remote from the radiating aperture and located in the near field of said composite beam;

said phase means further comprising means for directing said composite beam at said receiver assembly.

3. The system of claim 1, wherein said microwave redirector means comprise a plurality of microwave reflectors separate from said microwave sources for reflecting said microwave beams from said sources into said beam direction.

4. The system of claim 3, wherein said microwave reflectors are positioned so as to lie substantially within and cover substantially the whole area of the radiating aperture when viewed from a position in the composite microwave beam.

5. The system of claim 1 wherein each microwave source comprises at least one microwave generator and at least one associated radiating antenna, and said microwave redirecting means comprises a plurality of microwave reflective surfaces which are substantially transparent to incident sunlight and are positioned to reflect microwave beams emitted from said antennas in a predetermined direction, said microwave reflective surfaces being positioned to lie within said radiating aperture and cover substantially the whole area of said aperture when viewed from a position in said composite, shaped microwave beam.

6. The system of claim 1, wherein said solar power collecting assemblies each comprise an area of solar collector structures having photoconversion surfaces arranged to permit exposure to incident sunlight over a wide range of angles of incident sunlight and electrical connector means are provided for connecting said photoconversion surfaces in each area to a respective one of said microwave sources.

7. The system of claim 6 wherein each microwave source is positioned in a respective one of said solar collector areas.

8. A method of transmitting a microwave beam to at least one receiver located in the radiating near field region of the microwave beam, comprising:

collecting solar radiation incident on a plurality of solar power collecting assemblies;

forming a microwave radiating aperture of predetermined area by arranging an array of separate, spaced apart microwave sources to form a series of separate microwave beams directed at a series of separate microwave reflectors for reflecting and redirecting the separate beams into a common beam direction and expanding the effective microwave emitting area to substantially fill the microwave radiating aperture, the microwave reflectors being positioned so as to apparently define a radiating aperture of said predetermined area when viewed from the common beam direction;

connecting the microwave sources to conversion means associated with the solar power collecting assemblies for converting the collected power into a power form suitable for connection to said microwave sources, each solar power collecting assembly and associated conversion means being connected to at least a respective one of said microwave sources;

controlling the phase of the microwave beams emitted from the microwave sources so as to combine the separate beams into at least one composite, shaped microwave beam directed from said microwave reflectors, the composite beam having a separate radiating near field and far field region; and

receiving the microwave beam at at least one microwave receiver assembly located in the radiating near field region of the microwave beam.

9. The method of claim **8**, further including the step of:

positioning a microwave receiver structure in the near field of the shaped microwave beam.

* * * * *

US005733313A

United States Patent [19]

Barreras, Sr. et al.

[11] **Patent Number: 5,733,313**

[45] **Date of Patent: Mar. 31, 1998**

[54] **RF COUPLED, IMPLANTABLE MEDICAL DEVICE WITH RECHARGEABLE BACK-UP POWER SOURCE**

[75] Inventors: **Francisco Jose Barreras, Sr.**, Miami; **Oscar Jimenez**, Coral Gables, both of Fla.

[73] Assignee: **Exonix Corporation**, Miami, Fla.

[21] Appl. No.: **690,968**

[22] Filed: **Aug. 1, 1996**

[51] **Int. Cl.^{6}** **A61N 1/08**

[52] **U.S. Cl.** **607/33**; 607/60; 607/61; 128/903

[58] **Field of Search** 607/29–34, 60, 607/61; 128/903

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,942,535	3/1976	Schulman .	
4,041,955	8/1977	Kelly et al. .	
4,408,607	10/1983	Maurer	128/419 R
4,441,498	4/1984	Nordling .	
4,793,353	12/1988	Borkan	128/421
5,279,292	1/1994	Baumann et al.	607/137
5,314,453	5/1994	Jeutter	607/61
5,314,457	5/1994	Jeutter et al. .	
5,476,488	12/1995	Morgan et al.	607/32
5,480,415	1/1996	Cox et al.	607/32

Primary Examiner—William E. Kamm
Assistant Examiner—Kennedy J. Schaetzle
Attorney, Agent, or Firm—Thomas R. Vigil

[57] **ABSTRACT**

The implantable, electrically operated medical device system comprises an implanted radio frequency (RF) receiving unit (receiver) incorporating a back-up rechargeable power supply and an implanted, electrically operated device, and an external RF transmitting unit (transmitter). RF energy is transmitted by the transmitter and is coupled into the receiver which is used to power the implanted medical device and/or recharge the back-up power supply. The back-up power supply within the receiver has enough capacity to be able to, by itself, power the implanted device coupled to the receiver for at least 24 hours during continual delivery of medical therapy. The receiver is surgically implanted within the patient and the transmitter is worn externally by the patient. The transmitter can be powered by either a rechargeable or non-rechargeable battery. In a first mode of operation, the transmitter will supply power, via RF coupled energy, to operate the receiver and simultaneously recharge the back-up power supply. In a second mode of operation, the receiver can, automatically or upon external command from the transmitter, acquire its supply of power exclusively from the back-up power supply. Yet, in a third mode of operation, the receiver can, automatically or upon command from the transmitter, alternatively acquire its supply of power from either, RF energy coupled into the receiver or the internal back-up power supply.

28 Claims, 6 Drawing Sheets

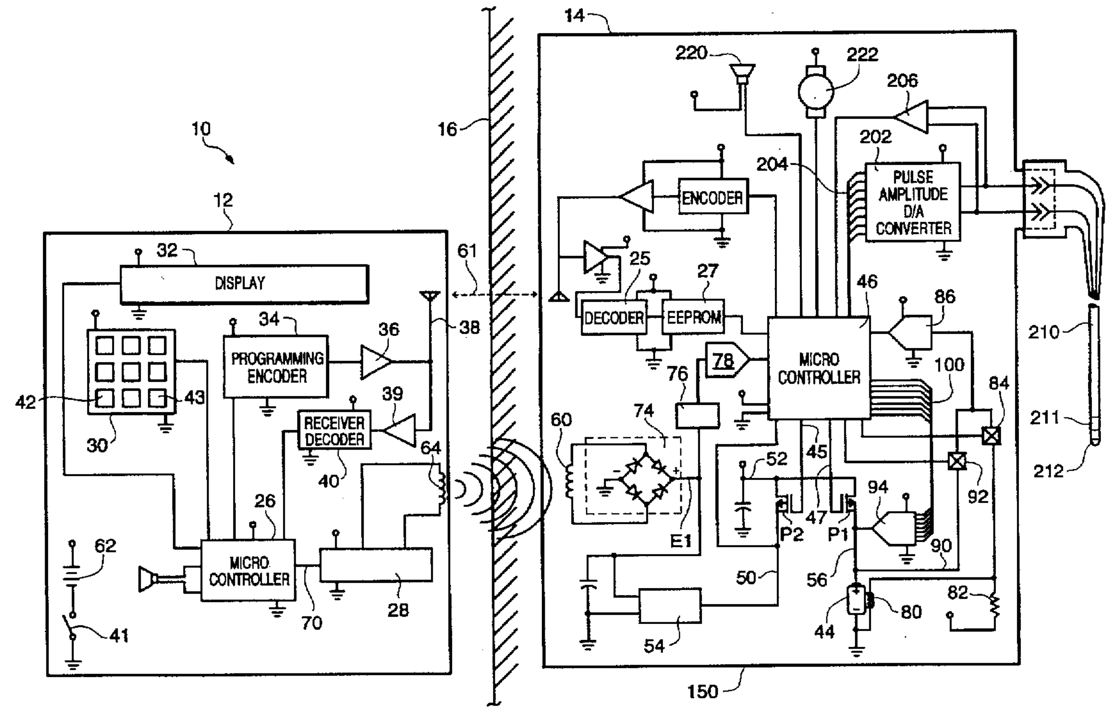
10
12
14
16
32
DISPLAY
34
PROGRAMMING ENCODER
36
38
61
39
RECEIVER DECODER
40
42
43
30
64
26
62
MICRO CONTROLLER
70
28
41
220
222
206
202
PULSE AMPLITUDE D/A CONVERTER
204
ENCODER
25
27
DECODER
EEPROM
46
86
78
76
MICRO CONTROLLER
100
84
74
60
52
45
94
47
P2
P1
92
90
E1
50
56
82
80
44
54
150
210
211
212

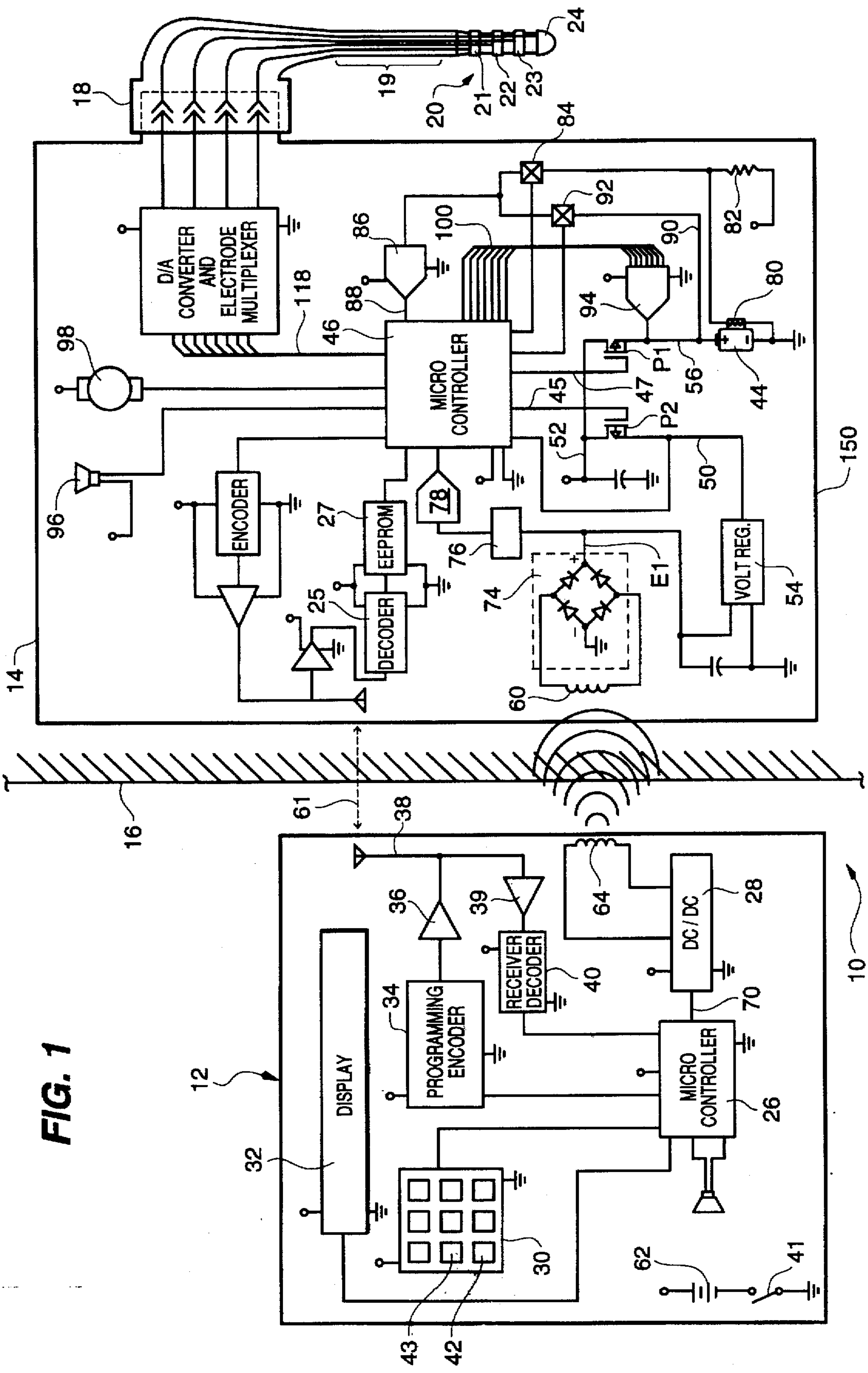
D/A CONVERTER AND ELECTRODE MULTIPLEXER
MICRO CONTROLLER
P1
P2
ENCODER
EEPROM
E1
VOLT REG.
DECODER
DC/DC
RECEIVER DECODER
PROGRAMMING ENCODER
DISPLAY
MICRO CONTROLLER

FIG. 1

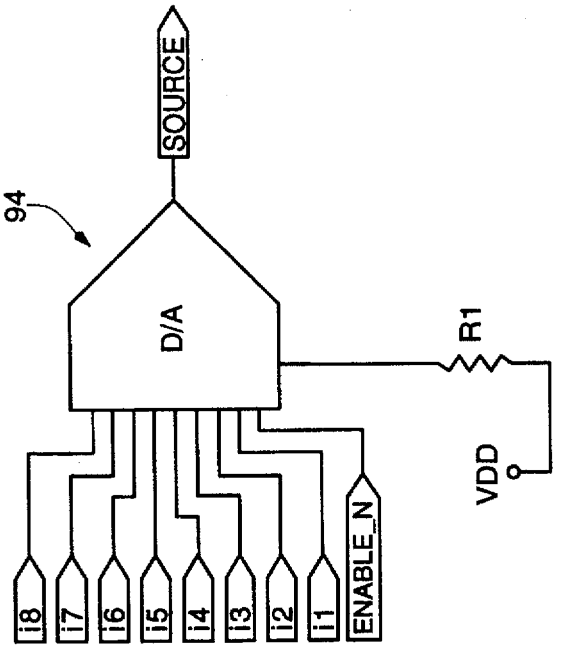
SOURCE
94
D/A
R1
VDD
i8
i7
i6
i5
i4
i3
i2
i1
ENABLE_N

FIG. 2A

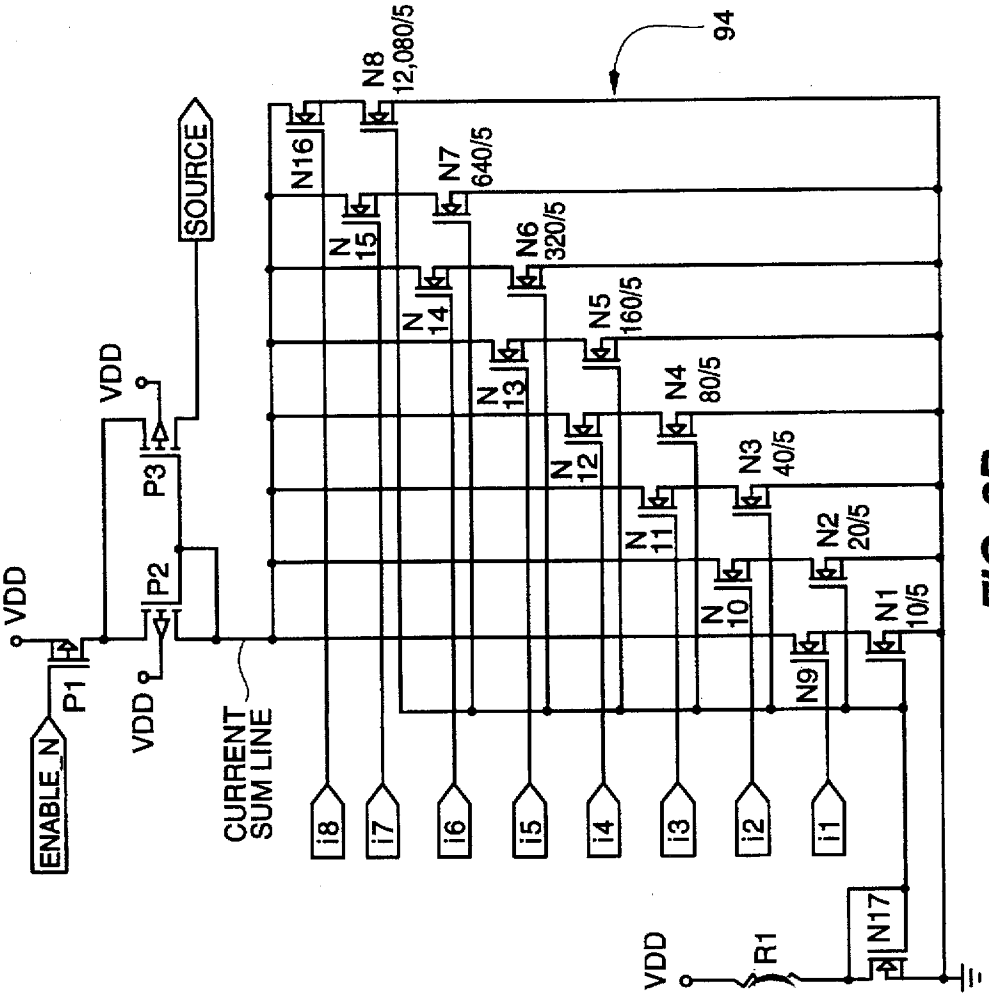
94
SOURCE
VDD
VDD
VDD
P1
P2
P3
ENABLE_N
CURRENT SUM LINE
N8 12,080/5
N16
N7 640/5
N 15
N6 320/5
N 14
N5 160/5
N 13
N4 80/5
N 12
N3 40/5
N 11
N2 20/5
N 10
N1 10/5
N9
i8
i7
i6
i5
i4
i3
i2
i1
N17
R1
VDD

FIG. 2B

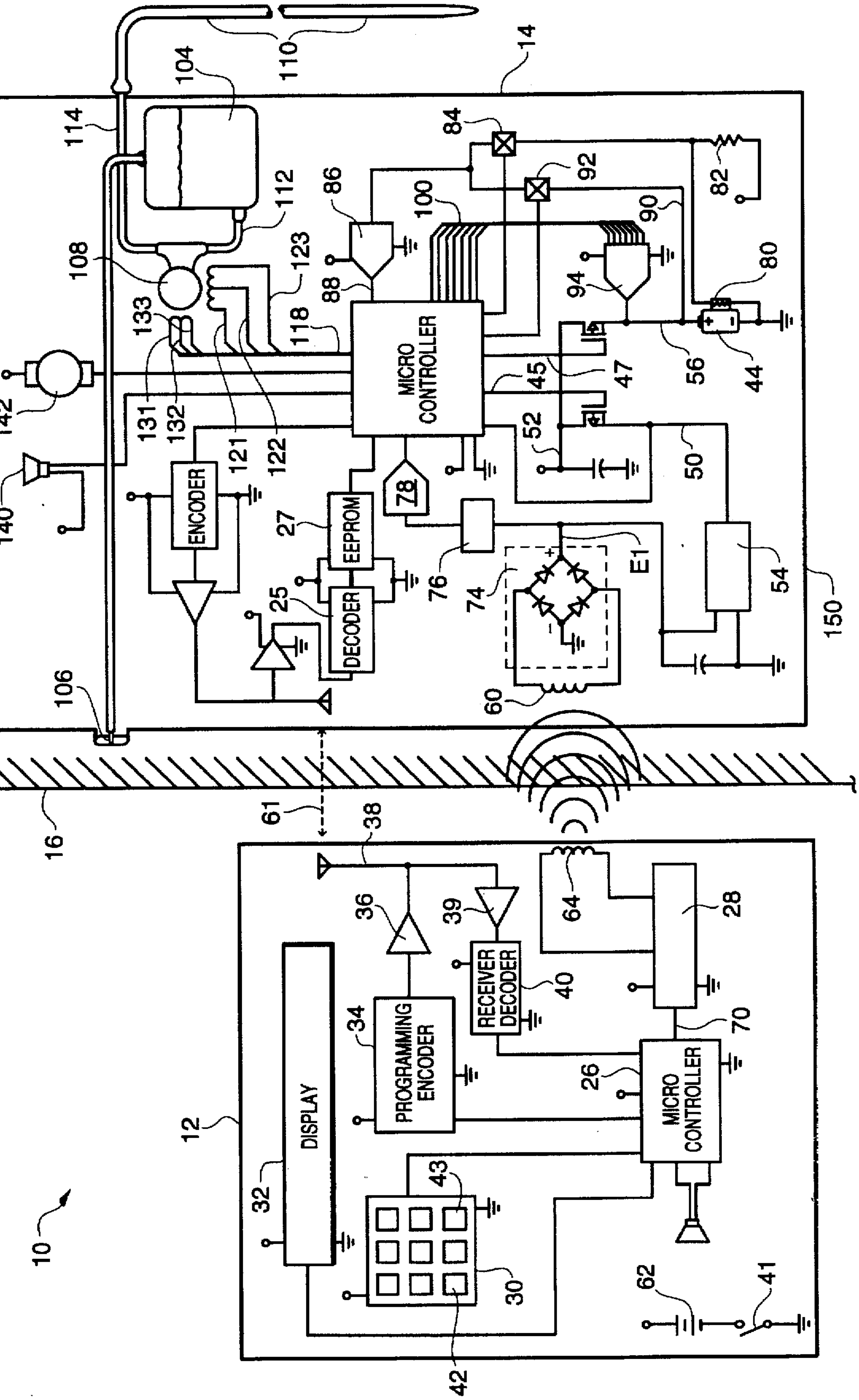
MICRO CONTROLLER
ENCODER
EEPROM
DECODER
E1
RECEIVER DECODER
PROGRAMMING ENCODER
MICRO CONTROLLER
DISPLAY

FIG. 3

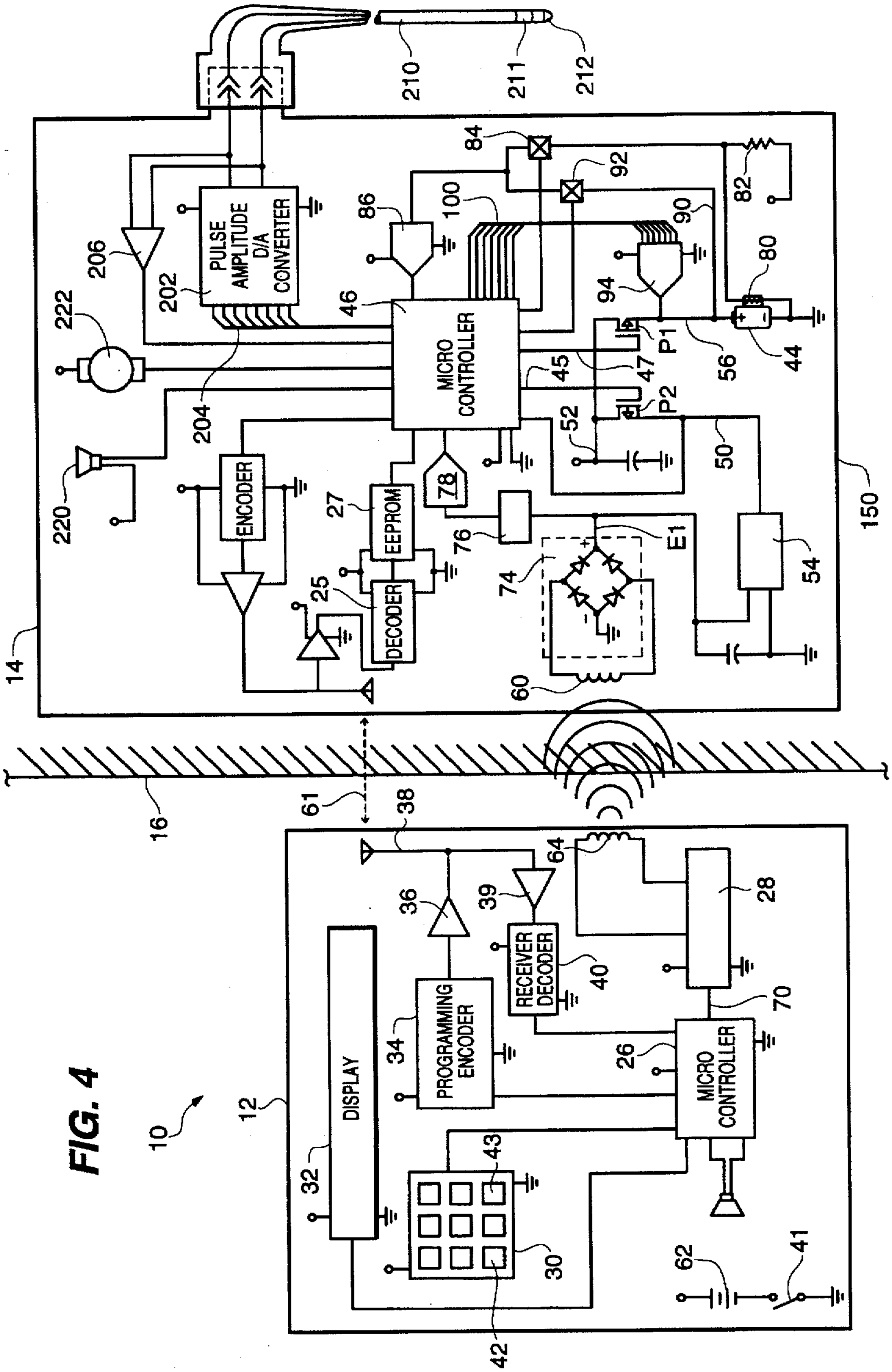

FIG. 4

E1
P1
P2
PULSE AMPLITUDE D/A CONVERTER
MICRO CONTROLLER
ENCODER
EEPROM
DECODER
DISPLAY
PROGRAMMING ENCODER
RECEIVER DECODER
MICRO CONTROLLER

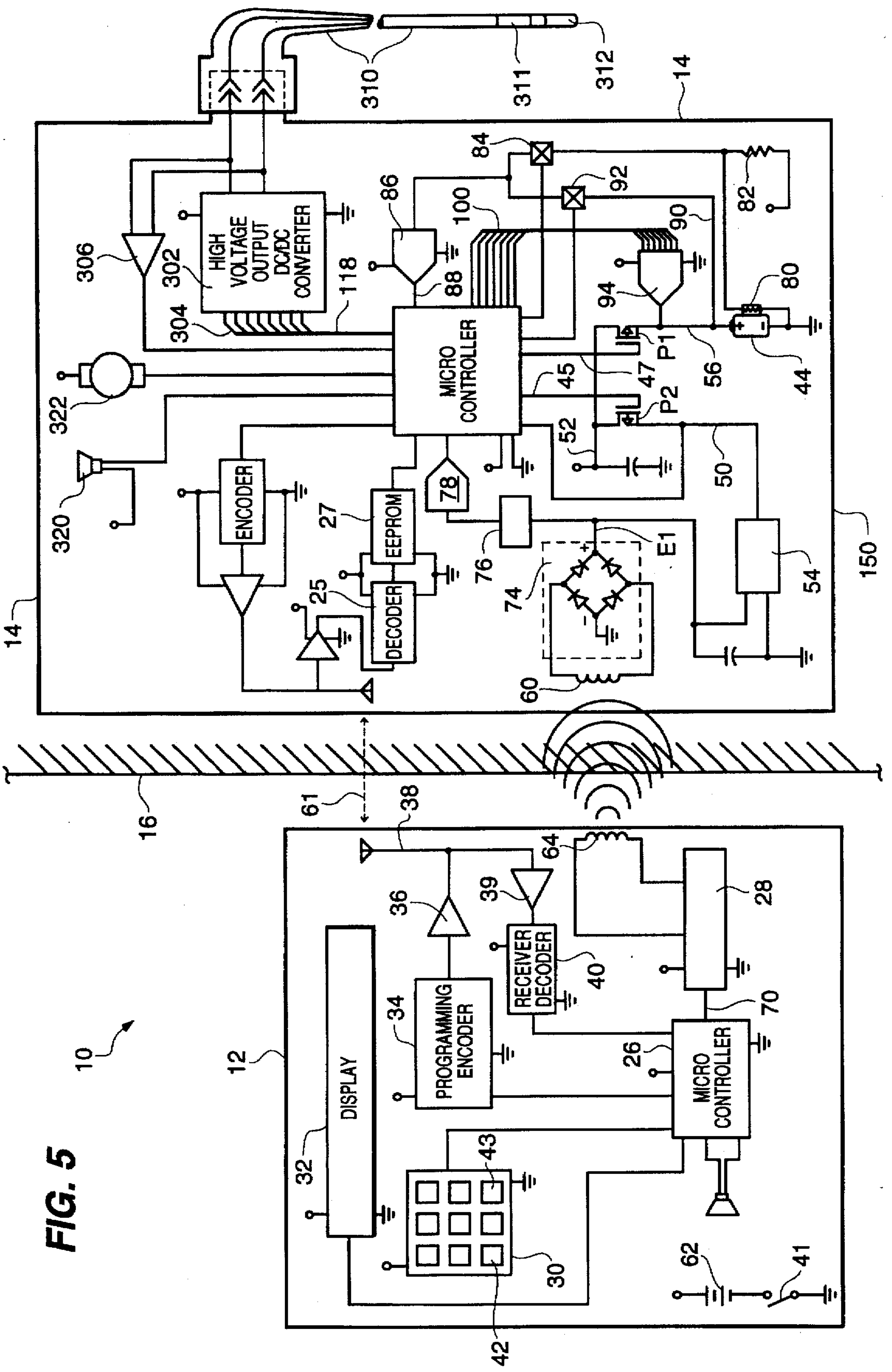
HIGH VOLTAGE OUTPUT DC/DC CONVERTER
MICRO CONTROLLER
P1
P2
ENCODER
EEPROM
E1
DECODER
RECEIVER DECODER
PROGRAMMING ENCODER
MICRO CONTROLLER
DISPLAY

FIG. 5

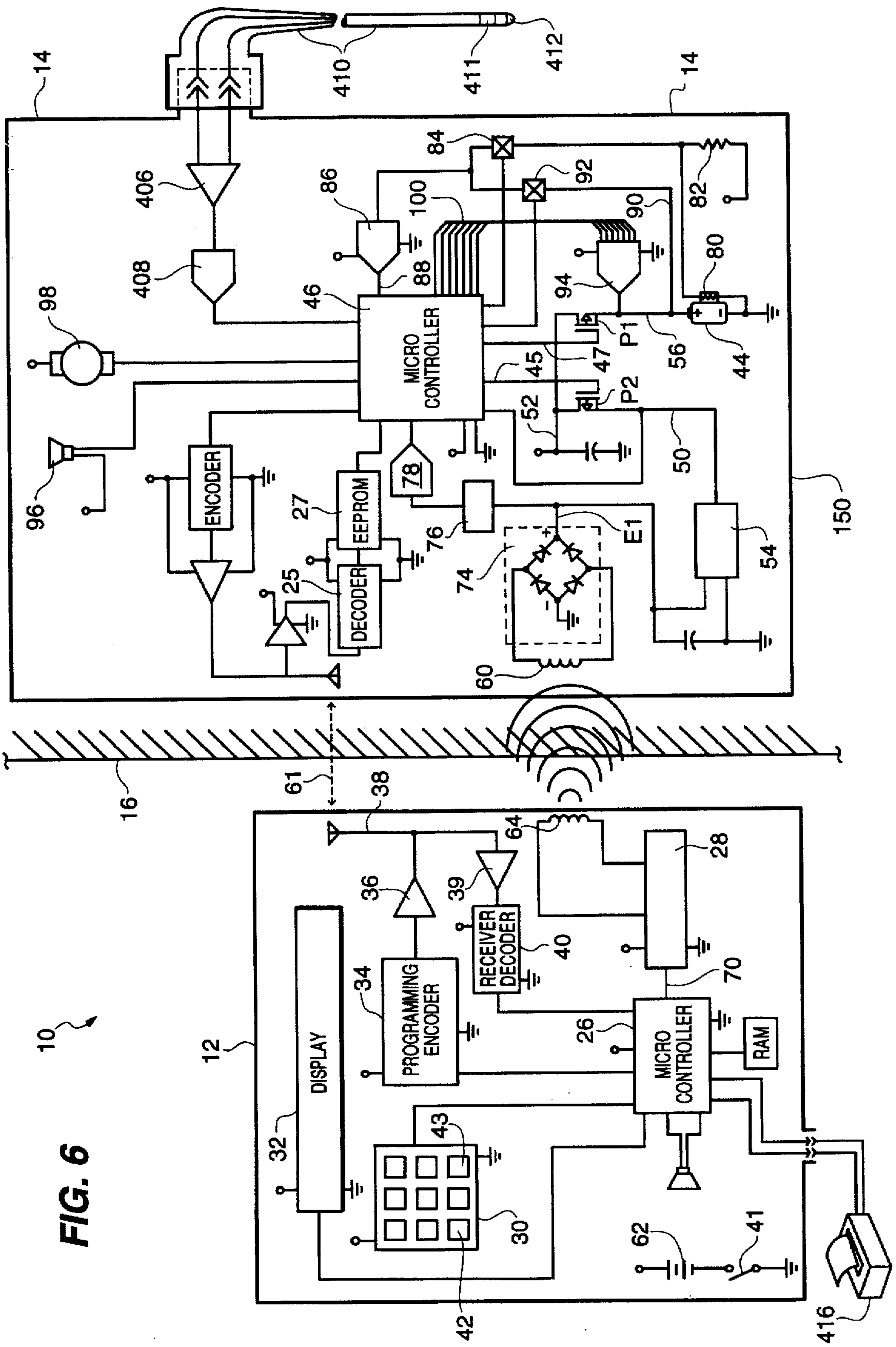
MICRO CONTROLLER
P1
P2
ENCODER
EEPROM
E1
DECODER
RECEIVER DECODER
PROGRAMMING ENCODER
DISPLAY
MICRO CONTROLLER
RAM

FIG. 6

RF COUPLED, IMPLANTABLE MEDICAL DEVICE WITH RECHARGEABLE BACK-UP POWER SOURCE

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to an implantable medical device including a rechargeable back-up power source and a charging unit for recharging the back-up power source via RF coupling.

2. Description of the Related Art Including Information Disclosed Under 37 CFR §§1.97–1.99

The concept of using an implantable, electrically operated medical device for treating specific diseases or physical disorders is well known. Examples of implantable, electrically operated medical devices are: cardiac pacemakers which restore a sick human heart to a normal rhythm, neural simulators which control nerve or brain response (such as pain or epileptic seizures), infusion pumps for subcutaneously drug delivery (such as insulin pump), and diagnostic devices for monitoring a patient's condition.

With respect to all of these implantable, electrically operated devices, it is necessary to provide power to the device implanted below the skin. Since the medical device is subcutaneously implanted in the patient, the power source must supply electrical energy for a reasonable period of time in order to reduce further surgical trauma to the patient and financial cost to the medical provider.

Implantable, electrically operated medical devices can be classified in three general areas: radio frequency (RF) coupled and partially powered implanted devices, non-rechargeable battery powered totally implanted devices, and devices which combine RF coupling and battery powered systems.

The RF coupled and powered devices do not carry or contain an independent power source. Therefore, the RF coupled device requires an external RF transmitter and a surgically implanted receiver. Such a device is an RF coupled neural stimulator. The RF link transfers stimulation pulses percutaneously through the skin and adjacent tissue layers of the patient from the external RF transmitter to the surgically implanted RF receiver and stimulator device. The transmitter sends, on a real-time basis, stimulation pulses to be applied ultimately to the implanted electrodes plus programming data defining the polarity of each electrode relative to each other to the implanted stimulation device. The implanted receiver obtains these stimulation pulses and programming data, converts pulses as necessary, delivers the energy contained in each transmitted stimulation pulse to the implanted electrodes as defined in the real-time programming data. The stimulation energy for each stimulation pulse is inductively coupled on a real-time basis from the external transmitter to the implanted receiver.

The common disadvantage of the RF coupled and powered stimulator is that the patient must always wear the external transmitter and antenna (even during sleep) in order for the implanted receiver to deliver stimulation pulses to the targeted tissue. Stimulation therapy ceases the moment the transmitter antenna is withdrawn just a few inches away from the implanted receiver. Although the RF coupled stimulator has this disadvantage, the service life of such an RF coupled and powered stimulator is not limited to the life of a battery contained within a fully implantable stimulation unit. Accordingly, the long term cost of the RF coupled and powered simulators is less than the non-rechargeable battery powered simulators because the service life of the former is much longer than that of the latter. RF coupled and powered simulators have been commercially marketed by Medtronic of Minneapolis, Avery Laboratories of New York and Neuromed of Dallas, Tex.

The battery powered stimulator utilizes a primary, non-rechargeable battery to power the implanted stimulator. The battery provides sole and exclusive power to the implanted stimulator continually while the stimulator generates one or more electric stimulation pulses, in a controlled manner, to the targeted tissue. Of course, the stimulation pulses are delivered to the targeted tissue via implanted leads. An external programmer may be used to non-invasively adjust the stimulation parameters, such as amplitude, pulse width or rate, or to control the duration of stimulation therapy each day. Programming may be provided through an RF telemetry link. After programming, the stimulator remembers the parameter values (the values are stored in an electronic memory) as long as the battery voltage remains above a minimum voltage level required by the electronics. Unfortunately, the service life of these battery powered implantable simulators is limited to the battery life. Accordingly, it is necessary to surgically remove and then replace the battery powered implantable simulators upon depletion of the electrochemically active materials in the battery. This disadvantage (i.e. surgical replacement) increases its long term cost to the medical provider relative to the aforementioned RF coupled and powered simulators. The battery powered implantable simulators do not require an external transmitter to deliver the stimulating electrical pulses. Accordingly, the battery powered implantable simulators are less complicated to use and more comfortable to wear than the RF coupled and powered simulators. Battery powered simulators have been marketed by Medtronic of Minneapolis and Exonix of Miami.

A third category of implantable, electrically operated devices include neural simulators which combine the RF coupled/powered systems with the battery powered implantable stimulator technology. These types of simulators enable the patient to obtain therapy without the necessity of having an external RF coupled unit proximate to the implant at all times. However, the stimulator must be surgically replaced after the battery is depleted if use of the external RF transmitter is not desired. This type of stimulator allows RF coupled stimulation at times when wearing the external transmitter is not objectionable, thereby extending battery life. Also, this type of stimulator may allow for continuing RF coupled stimulation after the internal power source is depleted, although some of these RF coupled and battery powered implantable simulators do not operate if the battery is completely depleted in the implanted stimulator.

Several examples of such previously proposed implantable devices are disclosed in the following U.S. patents:

U.S. Pat. No.	Patentee
4,408,607	Maurer
4,793,353	Borkan
5,279,292	Baumann et al.
5,314,453	Jeutter

U.S. Pat. No. 5,314,453 to Jeutter describes and claims an implant "locator" means to aid in locating an implanted device. The locator means comprises a reed switch affixed to the center of a transmitting coil and a magnet affixed within the implanted device. Transmission of high frequency RF energy is possible only when the reed switch is closed by the magnet within the implanted device, thus insuring some degree of good coupling between the transmitter and receiver. However, this patent describes a full wave rectifier along with rechargeable batteries. Accordingly, it incorporates a rechargeable battery. The present invention differs from the teachings of the Jeutter patent in the following respects:

1) Charging current is controlled by battery temperature to prevent gas generation by the battery and loss of battery electrolyte,
2) A real-time feedback system is provided between the receiver and the transmitter for real-time adjustment of the RF energy generated by the transmitter, thereby extending the service time of the transmitter's battery.
3) A low frequency RF coupling method (10 to 500,000 Hertz) is provided which allows RF coupling through a titanium encased receiver. The RF coupling system described by Jeutter operates at a very high frequency of 2,000,000 Hertz which is greatly attenuated by any metal enclosure. Jeutter describe an epoxy potted receiver which differs from a titanium encased receiver housing.
4) The receiver is capable of automatic switching the supply of power between RF coupled power upon sensing RF energy or battery power upon sensing loss of RF coupled power.

U.S. Pat. No. 5,279,292 to Joachim Baumann et al. teaches a series resonant circuit in an implantable device. The device disclosed hereinafter uses a parallel resonant circuit which can be tuned at low frequencies, 60 Hz to 500 kHz, such as, for example 8 kHz, which couples well through a titanium enclosure. Further this patent does not teach adjusting charging current as a function of battery temperature or controlling the charging current with the current output from a D/A converter.

U.S. Pat. No. 4,793,353 to William Borkan discloses a non-invasive multiprogramable tissue stimulator which utilizes RF coupling to charge and recharge a capacitor or other rechargeable voltage source. The Borkan circuit differs from the circuit disclosed hereinafter in several respects:

1) First of al in the Borkan circuit, while in the RF stimulation mode, each stimulation pulse must be generated and transmitted by the transmitter, on a real time basis, to the implanted receiver (in contrast, the receiver disclosed hereinafter incorporates all the elements required to autonomously generate and regulate the stimulation pulses);
2) Borkan teaches that a non-rechargeable battery can be used as an alternative power source to RF coupled stimulation, while the circuit disclosed hereinafter uses a rechargeable battery which can be "fast" or "trickle" recharged via low to medium frequency (10 to 500,000 Hertz) RF coupling while Borkan uses a much higher RF coupling (2,000,000 Hertz).

U.S. Pat. No. 4,408,607 to Donald Maurer teaches a capacitive energy source and associated circuitry for powering a medical apparatus. The Maurer circuitry differs from our the circuitry disclosed hereinafter in that:

1) The Maurer battery is non-rechargeable and only is used to power the implanted receiver during charging of the capacitor (the main power source);
2) Maurer provides no means for non-invasively recharging the battery;
3) Maurer does not control the charging current to a battery relative to the temperature of the battery to prevent gas generation by the battery and loss of battery electrolyte.

Moreover, Maurer does not teach the capability to switch automatically between RF coupled (power upon detection of RF energy) and battery power (upon sensing absence of RF energy).

SUMMARY OF THE INVENTION

According to the present invention there is provided an implantable, electrically operated medical device comprising an implanted radio frequency (RF) receiver incorporating a rechargeable back-up power source and an external RF transmitter. RF transmissions are coupled into the implanted receiver for powering delivery of medical therapy and simultaneous recharging of the back-up power source within the receiver. The back-up power source has enough capacity to be able to autonomously power the receiver for at least seven days during continual delivery of medical therapy. The receiver is surgically implanted within the patient and the transmitter is worn externally by the patient. The transmitter can be powered by either a rechargeable or non-rechargeable battery.

In a first mode of operation, the supply of power to operate the receiver, recharge the back-up power source and deliver therapy is acquired primarily from the RF energy coupled into the receiver, until the receiver senses a loss of RF power at which point it will automatically switch to the back-up power source in order to continue delivery of medical therapy. In this first mode of operation, the transmitter will transmit lower energy RF waves since the back-up power source is recharged at a slower rate.

In a second mode of operation, the receiver can, automatically or upon external command from the transmitter, acquire its supply of power exclusively from the back-up power source. In the second mode of operation, the transmitter will transmit high energy RF waves in order to recharge the back-up power source at a faster rate and will terminate the RF transmission upon receiving from the receiver a "termination command" which indicates that the back-up power source is fully charged.

Yet, in a third mode of operation, the receiver can, automatically or upon command from the transmitter, alternatively acquire its supply of power from either or both, RF energy coupled into the receiver and/or the internal back-up power source. This third mode of operation is useful in patients where it is desirable to power the receiver from RF coupled energy during periods of awakeness, but switch to the back-up power source during sleep for greater comfort. In this third mode of operation, a real time clock within the receiver can be programmed by the transmitter with different start/stop times for RF coupled power than for back-up battery power. In this fashion, the switching is done automatically and the external transmitter unit will alert the patient which is the source of power for the receiver at all times.

The receiver includes a mechanism for alerting the patient when the back-up power source is nearing depletion and needs to be recharged. The mechanism can include: 1) a vibrating device within the receiver; 2) an audible tone generating device within the receiver; or, 3) a specific message shown in the transmitter's display combined with a specific audible tone generated by the transmitter.

The transmitter incorporates a transmitting antenna which generates RF wave-fronts which are coupled into an inductor within the implanted receiver. This RF coupled power, which is alternating current or AC in nature, is rectified, filtered and converted into a high DC voltage within the receiver. Further, a voltage regulator within the receiver converts the high DC voltage into a lower precise DC voltage from which the receiver operates.

It is the objective of the present invention to provide an implantable, electrically operated medical device (receiver) capable of obtaining its source of electrical power from either, an external battery coupled via low level RF transmissions (transmitter), or a back-up rechargeable battery contained within the implanted receiver.

According to one aspect of the present invention, there is provided a method of supplying power, on an exclusive basis, from externally low energy RF coupled power (transmitter), to an implanted receiver during continual delivery of medical therapy.

According to another aspect of the present invention, there is provided circuitry for programming into the receiver the times of day (meaning a continuous 24 hour period) in which the receiver automatically switches it's supply of power from RF coupled energy to the back-up power source, and vice-versa.

According to still another aspect of the present invention, there is provided a method for recording into a non-volatile memory contained within the implanted receiver the stimulation values and other critical data, so that it will not be erased if the back-up power source is depleted and to eliminate the need for the transmitter to generate and regulate, on a real time basis, the delivery of medical therapy.

According to still another aspect of the present invention, there is provided a method of supplying power, on an exclusive basis, during at least a seven day cycle of substantially continual delivery of medical therapy utilizing a rechargeable power source.

According to still another aspect of the present invention, there is provided a method for non-invasively recharging the power source within the receiver, whereby the electrical energy contained in the battery powering the external transmitter is transferred into the rechargeable power source within the receiver utilizing an inductive RF power link between the external transmitter (recharging unit) and the implanted receiver (unit being recharged).

According to still another aspect of the present invention, there is provided a method for regulating the rate of recharging the back-up power source contained within the implanted receiver as a function of temperature of the back-up power source, in order to inhibit the power source from generating harmful gases and to prevent electrolyte loss, thereby enhancing the service life of the back-up power source and increasing the possible number of recharge cycles.

According to still another aspect of the present invention, there is provided a method, interactive between the transmitter and receiver, for regulating the RF energy generated by the external RF transmitter as a function of distance between the transmitting and receiving inductors.

With these methods, the level of RF energy coupled into the receiver antenna is monitored by the receiver which telemeters specific commands to the RF transmitter to adjust the RF energy level being generated. This real-time feedback system allows generation of just the minimum RF power needed at the receiver, thereby extending the service time of the battery powering the RF transmitter.

According to still another aspect of the present invention, there is provided in transmitter for the patient to select between RF powered operation only, back-up power source operation only, or a combination of both. If "RF power" is selected, the implanted receiver will only operate when the transmitter unit is proximate to the implanted receiver, and the transmitter will generate low level RF energy. If "back-up power" is selected, the implanted receiver will draw it's operating power exclusively from the back-up power source, and the transmitter will generate higher level RF energy when used to quickly recharge the back-up power source. If "combination power" is selected, the implanted receiver will draw it's operating power from the transmitter via low level RF coupled energy, but will automatically switch to draw it's operating power from the internal back-up power source when the transmitter is removed or turned off. The patient may select one of these options via a specific menu shown in the transmitter's display.

According to still another aspect of the present invention, there is provided a method for the receiver to automatically terminate the recharge cycle upon sensing a fully charged back-up battery by the transmission of a specific telemetry command to the transmitter, thereby enhancing the service life of the battery powering the RF transmitter.

According to still another aspect of the present invention, there is provided several receiver-initiated commands for alerting the patient when the rechargeable power source is nearing depletion and needs to be recharged. Some alerting signals are generated by the receiver itself and other alerting signals are generated by the transmitter.

According to still another aspect of the present invention, power sources and methods are utilized for non-invasively coupling energy and recharging the implanted battery, in combination with an implantable tissue stimulator.

According to still another aspect of the present invention, there is provided an implantable drug delivery system utilizing these power sources and methods for non-invasively coupling electric energy into the implanted drug delivery system and recharging its battery.

According to still another aspect of the present invention, there is provided an implantable monitor and diagnostic system utilizing these power sources and methods for non-invasively coupling electrical energy into the implanted monitor which is used to collect vital physical data from the patient which can be interrogated by a receiver external to the patient.

According to still another aspect of the present invention, there is provided an implantable cardiac pacemaker utilizing these power sources and methods for non-invasively coupling electrical energy into the implanted pacemaker and recharging its battery.

According to still another aspect of the present invention, there is provided an implantable cardioverter/defibrillator utilizing these power sources and methods for non-invasively coupling electrical energy into the implanted cardioverter/defibrillator and recharging its battery.

According to still another aspect of the present invention, there is provided a method for recharging power via an inductive RF power link between the recharging unit and the implant and through the hermetic titanium encasement of the implant.

According to still another aspect of the present invention, a patient is permitted to vary the delivery schedule and/or quantities of the medical therapy via the external RF transmitter/recharger unit.

According to still another aspect of the present invention, a medical physician, nurse or technician is allowed to program into the implanted medical device, via the external RF transmitter/recharger unit, the delivery schedule, values or quantities of the medical therapy and to set the limit for these schedules, values or quantities within which the patient may safely adjust them later on.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a block electrical, schematic circuit diagram of the overall system including a transmitting unit and an implanted receiver unit configured for an implantable, rechargeable tissue stimulator system.

FIG. **2A** is a block electrical schematic circuit of a digital to analog converter in the receiver unit

FIG. **2B** is a detailed electrical schematic circuit diagram of the digital to analog converter with current output shown in FIG. **2A** which is used, under control of a micro controller, to regulate the constant current rate for recharging the back-up power source within the implanted receiver unit.

FIG. **3** is a block electrical schematic circuit diagram of the overall system as configured for an implantable, rechargeable drug delivery system.

FIG. **4** is a block electrical schematic circuit diagram of the overall system as configured for an implantable, rechargeable cardiac pacemaker system.

FIG. **5** is a block electrical schematic circuit diagram of the overall system as configured for an implantable, rechargeable cardioverter/defibrillator.

FIG. **6** is a block electrical schematic circuit diagram of the overall system as configured for an implantable, rechargeable monitor and diagnostic system.

DESCRIPTION OF THE PREFERRED EMBODIMENT(S)

Referring to the drawings in greater detail, there is illustrated in FIG. **1** a block electrical schematic circuit diagram of an implantable, rechargeable tissue stimulator system **10**. The system **10** includes a transmitter **12** and a receiver **14**, the latter being surgically implanted beneath a patient's skin **16**. The receiver **14** is connected, via lead connector **18**, to an implanted medical device, which in this embodiment, is an implanted lead **19** which contains, at it's distal end **20**, stimulating electrodes **21–24**. These electrodes **21–24** are implanted adjacent to the target tissue to be stimulated (i.e., a specific nerve or a nerve bundle, a specific area of the brain or a specific muscle within the human body). The implanted receiver **14** receives therapy values, transmitted by the transmitter **12** via RF signals, which are decoded by decoder **25** and then stored in a non-volatile memory (EPROM) **27**.

Referring again to FIG. **1**, the major components of the transmitter **12** are a micro controller **26** which is used, via software, to: 1) control the output of a programmable DC to DC converter **28** in order to regulate the amount of RF energy to be coupled into the receiver **14**; 2) read data and command inputs inputted via a keyboard **30**, display messages and menus via a display **32**, transmit therapy parameter values, via a programming encoder **34**, a transmit driver **36** and an antenna **38**, to the implanted receiver **14**; 3) and receive commands and patient's diagnostic data, transmitted from the implanted receiver **14**, via the antenna **38**, an amplifier **39** and a decoder **40**.

As shown in FIG. **1**, when the transmitter **12** is powered up via a switch **41** and a "start therapy" key **42** on the keyboard **30** is pressed, the transmitter **12** will transmit the "start" command to the receiver **14** which will initiate delivery of therapy by the receiver **14**. Likewise, if a "stop therapy" key **43** on the keyboard **30** is pressed, the transmitter **12** will transmit the "stop" command to the receiver **14** which will cease delivery of therapy.

The implanted receiver **14** of FIG. **1** can be programmed by the physician or patient to obtain it's operating power from one of three sources: 1) RF coupled energy only; 2) back-up rechargeable power supply/source **44** only; or 3) a combination of both whereby the implanted receiver **14** alternates automatically from one to the other according to a preset schedule programmed via the transmitter **12**.

When "RF only" is selected from a menu displayed by the display **32** of the transmitter **12**, an output port **45** of a micro controller **46** in the receiver **14** is switched to a "0" and a port **47** is switched to a "1" which places pmos transistor P2 in a conducting state and pmos transistor P1 in a non-conducting state. This effectively connects a line conductor **50** to a line conductor **52**, making VDD equal to the output of a voltage regulator **54** which is at +3.0 vdc.

When "battery only" is selected from the same menu, the output port **47** is switched to a "0" and the port **45** is switched to a "1", thus effectively connecting line conductor **56** to line conductor **52** and making VDD equal to the voltage level at the rechargeable power source **44**.

As shown in FIG. **1**, when "combination" is selected from the same menu, the system **10** will automatically switch the source of VDD to the output of the voltage regulator **54** (line conductor **50**) when the transmitter **12** is proximal to the receiver **14**, or to the rechargeable power source **44** when the transmitter **12** is removed away from the receiver **14**. The automatic switching is performed by the micro controller **46** in response to the state of line conductor **50** which is at +3.0 volts when RF energy is being coupled into an inductor **60** (the transmitter **12** is proximal to the receiver **14**) or is 0 volts in the absence of RF energy (the transmitter **12** is away from the receiver **14**).

When the receiver **14** is programmed to "battery only" power acquisition mode, it's exclusive source of operating power becomes the rechargeable power source **44**. After prolonged use, the rechargeable power source **44** will reach a near depleted level, at which point the receiver **14** will transmit, via an RF communication link **61**, a "recharge" command to the transmitter **12**. This will cause the transmitter **12** to generate, via the battery **62**, the DC/DC converter **28** and an output inductor **64**, high energy RF waves which are coupled into the inductor **60** contained within the receiver **14**. The actual level of RF energy generated by the inductor **64** is regulated by an output port **70** of the micro controller **26** as a real-time response to data transmitted by the receiver **14** via the micro controller **46**, the data representing the voltage level E1 at the output of the rectifier **74** in the receiver **14** which is measured via an attenuator **76** and an analog to digital converter **78**. This feedback system extends the life of the battery **62** within the transmitter **12**, by adjusting, as a function of distance between the inductors **64** and **60**, the RF energy required to quickly recharge the rechargeable power source **44**. A close proximity requires much less RF energy to recharge the rechargeable power source **44** than a longer distance would, in the same time. During this recharging operation, the micro controller regulates, as a function of temperature, the current level used to recharge the rechargeable power source **44**. The temperature is measured by a thermistor **80** which is adhered to the rechargeable power source **44** during manufacturing. The junction between the thermistor **80** and a resistor **82** form a voltage divider which is fed through an analog switch **84** to an analog to digital converter **86** and, via a line conductor **88**, to the micro controller **46**. As the voltage rises, the ohmic value of the thermistor **80** drops proportionally to the temperature, thus reducing the voltage at the line conductor **88** to the micro controller **46**. This loop forms a temperature-controlled, current-regulated charging system which restricts the temperature rise of the rechargeable power source **44** during recharging, thus preventing the power source **44** from suffering electrolyte starvation and gas generation. Both of these phenomena will, if left unchecked, dramatically reduce the reliability and service life of the power source **44**. Also, during recharging of the power source **44**, the micro controller **46** will monitor the voltage level of the power source **44** via a line conductor **90**, analog switch **92**, the A/D converter **86** and, finally, the line conductor **88**. Upon sensing a fully charged state, the micro controller **46** will telemeter to transmitter **12**, via the RF communications link **61**, a "stop" recharging command and simultaneously will turn off a D/A converter **94** which will cut off the current needed to charge the rechargeable power source **44**. In this manner, the power source **44** cannot be overcharged, even if the "stop" command was not received by the transmitter **12** due to electromagnetic interference.

Referring to FIG. **1**, when the receiver **14** is programmed to "RF only", the power acquisition mode, it's exclusive source of operating power is the low level RF energy generated by transmitter **12** and coupled into the inductor **60** within the receiver **14**. The actual level of RF energy generated by the inductor **64** is regulated by the output port **70** of the micro controller **26** to the minimum level required to operate the receiver **14** and the rechargeable power source **44** is trickle charged, as a real-time response to data transmitted by the receiver **14** via the micro controller **26**, i.e., the data representing the voltage level E1 at the output of the rectifier **74** which is measured via the attenuator **76** and the analog to digital converter **78**. This feedback system extends the life of the battery **62** within the transmitter **12**, by adjusting the RF energy required to operate the receiver **14** and maintain a trickle charge to the rechargeable power source **44**, as a function of the distance between the inductors **64** and **60**. At close proximity, much less RF energy is required to accomplish these functions than at a longer distance.

During trickle charging, the micro controller **46** regulates, as a function of temperature, the current level used to trickle charge the power source **46**, by the same method already explained in the previous paragraph. Again, this prevents electrolyte starvation and gas generation within the rechargeable power source **44**. Also, during trickle charging, the micro controller **46** will monitor the charge level of the power source **44**, and upon sensing a fully charged state, the receiver **14** will telemeter to the transmitter **12** the "stop" recharging command and simultaneously will turn off the D/A converter **94** which will cut off the current needed to charge the power source **44**. In this manner, the power source **44** cannot be overcharged, even if the "stop" command was not received by the transmitter **12** due to electromagnetic interference.

When the receiver **14** is programmed to "combination" power acquisition mode, the micro controller **46** will automatically switch delivery of operating power to the receiver **14** to RF coupled energy upon detection of RF induced voltage at E1. Likewise, the micro controller **46** will switch delivery of operating power to the rechargeable power source **44** upon loss of RF induced voltage at E1. The patient may select, via a menu shown in the transmitter's display **32**, fast or trickle charge to the rechargeable power source

Upon sensing that the charge in the rechargeable power source **44** is below a predetermined level, the micro controller **46** signals the patient, via an audible alarm **96** and/or a vibrating alarm **98**, that the rechargeable power source should be recharged.

Referring to FIG. **2B**, a detailed electrical schematic of the digital to analog (D/A) converter **94** is provided. The D/A converter **94** is software programmable, precision current source whose output is regulated by the micro controller **46**. It should be noted that this type of D/A converter is best implemented into an integrated circuit where the electrical characteristics of the transistors can be precisely matched.

Referring again to FIG. **2B**, resistor **R1** is used to set the base bias current for the converter **94**. Transistors **N17** and **N1** form a 1:1 current mirror where the current into the transistor **N17** equals the current through the transistor **N1**, since both transistors have equal channel width and length. However, the channel width for the transistors **N2** through **N8** are binary weighted, so that the transistor **N2** has twice the width of the transistors **N1**, **N3** has twice the width of **N2**, and so forth. This binary scaling results in transistor **N2** conducting twice the current of transistor **N1** (assuming equal bias current), the transistor **N3** conducting twice the current of the transistor **N2**, and so forth. The transistors **N9** through **N16** are used as pass devices to allow the current available at the transistor **N1** through the transistor **N8** to pass to the current sum line **1**. The on-off state for the transistors **N9** through **N16** is governed by inputs **i1** through **i8** which, in one embodiment, are output ports from the micro controller **46** (bus line **100** in FIG. **1**). In this manner, the micro controller **46** is able to select any value of current between 1 and 256 times that of the current flowing through the resistor **R1** (bias current) to pass to the current sum line **1**.

Input line "enable.n" is used to turn on and off the D/A converter. When "enable.n" is a "0", the transistor **P1** conducts connecting VDD to the sources of the transistors **P2** and **P3** which form a 1:1 current mirror. Therefore, the current sourced by the transistor **P3** equals the current flowing through the transistors **P2**, **P3** being the output current device ("iSOURCE") for this D/A converter **94**.

In FIG. **3** is shown a block diagram for an implantable, rechargeable drug delivery system. The block diagram for the transmitter **12** is the same as for the transmitter **12** shown in FIG. **1**.

The block diagram for the receiver **14** contains the same power supply system, supply switching means and method for recharging the rechargeable power source **44** that has been already described above in connection with the description of FIG. **1**, but has been modified to incorporate the components required to assemble an implantable drug delivery system. These components are: 1) a drug reservoir **104** which contains the drug to be delivered by the receiver **14**; 2) a refill septum **106** used to percutaneously, via a hypodermic needle, refill the drug reservoir **104**; 3) a portioning pump **108** used to dispense a precise volume of drug to a catheter **110** by making one or more small injections (portions); 4) a pump inlet tube **112**; 5) a pump outlet tube **114**; 6) the drug delivery catheter **110** which is used to carry and deliver to the target tissue the drug volume dispensed by the pump **108**; and 7) a multi-wire cable **118** which carries the electrical signals for driving the pump **108**.

Referring again to FIG. **3**, the wire conductors **121–123** are used to drive the pump **108** and the wire conductors **131–133** are used to sense when a portion has been delivered. The micro controller **46** measures the time required for delivering each drug portion, and based on this time determines if the pump **108** is empty or contains fluid, since the former condition results is a faster time than the latter. Upon sensing a "pump empty" condition, the micro controller **46** signals the patient, via an audible alarm **140** and/or a vibrating alarm **142**, that the reservoir **104** is empty. It should be noted that the titanium housing **150** of the receiver **14** should be in close proximity to the audible alarm **140** in order to transmit the sound waves to outside the human body.

FIG. **4** is a block diagram of an implantable, rechargeable cardiac pacemaker system. The block diagram for the transmitter **12** is the same as for the block diagram of the transmitter **12** shown in FIG. **1**.

The block diagram for the receiver **14** contains the same power supply system, supply switching means and method for recharging the rechargeable power source **44** that already have been described above in connection with the description of FIG. **1**, but has been modified to incorporate the components required to assemble an implantable, rechargeable cardiac pacemaker system. These components are: 1) a pulse amplitude D/A converter **202** which is used to regulate, under command of the micro controller **46**, the amplitude of the stimulating pulses delivered to the human heart; 2) a bus **204** which carries the binary value for the amplitude from the micro controller **46** to the D/A converter **202**; 3) an amplifier and filter **206** which detects and amplifies the cardiac depolarization waves (such as R or P waves) and filters out other signal frequencies not related to cardiac activity; 4) an implanted lead **210** containing electrodes **211–212** which is used to deliver stimulating pulses to the heart in order to regulate the heart's rhythm, but which is used also to pick-up and carry the cardiac depolarization waves to the amplifier/filter **206**. These cardiac waves are used by the micro controller **46** to measure the intrinsic rate of the heart. The measurement is utilized to determine if electrical stimulation pulses are needed to speed-up the heart.

Upon sensing that the charge in the rechargeable power source **44** is below a predetermined level, the micro controller **46** signals the patient, via an audible alarm **220** and/or a vibrating alarm **222**, that the rechargeable power source **44** should be recharged.

In FIG. **5** is illustrated a block diagram of an implantable, rechargeable cardioverter/defibrillator. The block diagram for the transmitter **12** is the same as the block diagram for the transmitter **12** shown in FIG. **1**.

The block diagram for the receiver **14** contains the same power supply system, supply switching means and method for recharging the rechargeable power source **44** that already have been described above in connection with the description of FIG. **1**, but has been modified to incorporate the components required to assemble an implantable, rechargeable cardioverter/defibrillator system. These components are: 1) a high voltage output, DC to DC converter **302** used to convert the low voltage available at the rechargeable power source **44** to a relatively higher voltage required to cardiovert a fibrillating human heart; 2) a bus **304** which carries the binary value for the voltage amplitude from the micro controller **46** to the D/A converter **302**; 3) an amplifier/filter **306** used to detect the presence of a cardiac arrhythmia such as fibrillation or tachycardia; and, 4) an implanted cardioverting lead **310** containing cardioverting electrodes **311–312**.

As shown in FIG. **5**, the DC/DC converter **302** is used to generate either low voltage pulses to pace the human heart when needed or high voltage pulses to shock a large number of cardiac cells into synchrony, thereby restoring a normal cardiac rhythm. The micro controller **46**, via the bus **304**, regulates the timing and amplitude of low voltage pulses or high voltage shocks, depending if an arrhythmia is detected or not. Upon sensing that the charge in the rechargeable power source **44** is below a predetermined level, the micro controller **46** signals the patient, via an audible alarm **320** and/or a vibrating alarm **322**, that the rechargeable power source **44** should be recharged.

In FIG. **6** there is illustrated a block diagram of an implantable, rechargeable monitor and diagnostic system. The block diagram for the transmitter **12** is the same as the block diagram for the transmitter **12** shown in FIG. **1**.

The block diagram for the receiver **14** contains the same power supply system, supply switching means and method for recharging the rechargeable power source **44** that already have been described in connection with the description of FIG. **1**, but has been modified to incorporate the components required to assemble an implantable, rechargeable monitor and diagnostic system. These components are: 1) an amplifier/filter **406** used to amplify the desired biological signals and to filter out other undesirable signals; 2) an analog to digital converter **408** which is used to convert the biological signal into a digital value representative of frequency and amplitude of the biological signal; 3) a monitoring lead **410** containing electrodes **411–412** which are used to pick-up and carry the biological signals to the amplifier/filter **406**.

The mission of the monitor and diagnostic system shown in FIG. **6** is to monitor and record, in a non-volatile memory **414**, specific biological signals and events occurring adjacent to the monitoring electrodes **411–412**. Later, at a convenient time, these recordings can be telemetered to the transmitter **12** which will produce, via a graphic recorder **416**, a hard copy of the biological signals for the physician's examination and eventual diagnosis. Any time biological signals occur, they are scrutinized by the micro controller **46** for specific morphology which would cause the event to be stored into the memory **27** for later examination by the physician. An example of a typical use, would be to record dysfunctional endocardiac signals which, when inspected by a trained physician, may reveal the origin of a cardiac dysfunction not detected by conventional means, such as a surface EKG.

From the foregoing description, it will be apparent that the RF coupled, implantable medical system **10** with the rechargeable back-up power supply/source **44** of the present invention has a number of advantages, some of which have been described above and others of which are inherent in the invention. Also it will be understood that modifications can be made to the RF coupled, implantable medical system including the rechargeable back-up power supply/source **44** described above without departing from the teachings of the present invention. Accordingly, the scope of the invention is only to be limited as necessitated by the accompanying claims.

We claim:

1. An RF coupled implantable medical system comprising:

a transmitting unit;

a receiving unit including an implantable, electrically operated, medical device, RF energy receiving means, RF signal transmitting means and a rechargeable battery;

said transmitting unit including a power source, RF energy transmitting means, RF signal receiving means and first control means coupled to said RF energy transmitting means and to said RF signal receiving means for controlling the amount of RF energy transmitted to said receiving unit thereby to conserve on the amount of power obtained from said power source;

second control means coupled to said RF energy receiving means, to said rechargeable battery, to said RF signal transmitting means and to said implantable medical device, for adjusting the charging current flowing into said rechargeable battery, as a function of (a) the charge level of said rechargeable battery, (b) selected charging rate, and (c) selected power supply for the implantable medical device.

2. The system of claim **1** wherein said receiving unit includes a titanium housing enclosing said RF energy receiving means, said RF signal transmitting means, said rechargeable battery and said second control means.

3. The system of claim **1** wherein said RF energy transmitting means of said transmitting unit is constructed to transmit energy at a frequency as low as 10 Hz and up to at least 20,000 Hz.

4. The system of claim **1** wherein said rechargeable battery has a temperature sensor which is mounted closely adjacent thereto and which is coupled via said RF signal transmitting means to said first control means of said transmitting unit whereby the level of transmitted RF energy can be reduced proportionally to the reduction in charging rate of the rechargeable battery in said receiving unit, in order to reduce the power consumption from said power source powering said transmitting unit.

5. The system of claim **1** wherein said RF energy transmitting means of said transmitting unit includes mode selection means for recharging said rechargeable battery at a "fast" (high energy) rate or at a "trickle" (low to medium energy) rate.

6. The system of claim **1** wherein said transmitting unit includes power source selection means for setting said receiving unit to obtain its operating power from (1) RF coupled energy ("RF only" mode), (2) said rechargeable battery ("battery only" mode) or (3) automatically switch to "RF only" upon detection of said RF energy field, or "battery only" when said RF energy field is not detected ("combination" mode).

7. The system of claim **6** wherein said receiving unit includes: (a) means for rectifying said RF energy into a relatively high D.C. voltage, (b) means for regulating said high D.C. voltage into a lower D.C. voltage, and (c) means for detecting the presence of said RF energy field,

said receiving unit, when set to operate in said "RF coupled energy" mode, is operable to supply regulated electrical energy to said implantable device so long as said transmitting unit is located proximate to said receiving unit and said receiving unit is sensing transmitted RF energy.

8. The system of claim **6** wherein said receiving unit, when said transmitting unit is set to operate in said "battery only" mode, is operable, periodically, to supply electrical energy to said implantable device from said rechargeable power supply for a period of at least 24 hours.

9. The system of claim **6** wherein said receiving unit, when set to operate in said "combination" mode, is operable to supply regulated D.C. electrical energy to said implantable device, so long as said transmitting unit is located proximate to said receiving unit, and, separately, to "trickle charge" said rechargeable battery to maintain same fully charged.

10. The system of claim **1** wherein said first control means of said transmitting unit includes means for controlling the level of RF energy transfer from the transmitting unit to the receiving unit relative to one or more of one or more of the following parameters: (a) the charge level of said rechargeable battery, (b) selected charging rate, and (c) the selected power supply for said receiving unit.

11. The system of claim **1** wherein said receiving unit comprises means for measuring the charge level of said rechargeable battery and, upon sensing a fully charged battery, automatically up-links a coded signal which commands said transmitting unit to "stop" transmitting RF energy.

12. The system of claim **1** wherein said transmitting unit includes a visual display coupled to said first control means.

13. The system of claim **1** wherein said transmitting unit includes a keyboard coupled to said first control means.

14. The system of claim **13** wherein said keyboard includes keys to start and stop recharging of said rechargeable battery within the implantable medical device.

15. The system of claim **1** wherein said implanted medical device is a tissue stimulator.

16. The system of claim **1** wherein said implanted medical device is a drug delivery system.

17. The system of claim **1** wherein said implanted medical device is a cardiac pacemaker system.

18. The system of claim **1** wherein said implanted medical device is a cardioverter/defibrillator.

19. The system of claim **1** wherein said transmitting unit includes a battery, whereby said transmitting unit is portable and not dependant upon an a.c. power source.

20. The system of claim **19** wherein said battery can be a rechargeable battery or a non-rechargeable battery.

21. The system of claim **1** wherein said RF energy transmitting means of said transmitting unit includes mode selection means for setting said transmitting unit to operate in one of the following modes: "RF only", "battery only" or a "combination of both".

22. The system of claim **21** wherein said RF energy transmitting means of said transmitting unit controls the amount of RF energy transmitted and,

(a) when said implanted receiving unit is set to operate in said "RF only" mode, said transmitted RF energy is automatically adjusted to the amount of RF energy required (i) to operate said implanted device and (ii) to maintain said rechargeable battery, which is powering said implanted device, in a fully charged state;

(b) when said implanted receiving unit is set to operate in said "battery only" mode, said transmitted RF energy is automatically adjusted to the amount of RF energy required (i) to operate said implanted device and (ii) to recharge quickly said rechargeable battery which is powering said implanted device; and,

(c) when said implanted receiving unit is set to operate in said "combination of both" mode, said receiving unit is set to switch automatically to either said "RF only" mode upon detecting said transmitted RF energy, or to said "battery only" mode upon detecting a loss of said transmitted RF energy.

23. An RF coupled implantable medical system comprising:

a transmitting unit;

a receiving unit including an implantable, electrically operated, medical device;

said transmitting unit including RF energy transmitting means, RF signal receiving means and first control means coupled to said RF energy transmitting means and to said RF signal receiving means for controlling the amount of RF energy transmitted to said receiving unit;

said receiving unit including RF energy receiving means, RF signal transmitting means, a rechargeable power supply coupled to said RF energy receiving means and second control means for adjusting the charging current flowing into said rechargeable battery coupled to said rechargeable power supply means, to said RF energy receiving means, to said RF signal transmitting means and to said implanted medical device, and

mode selection means for setting said receiving unit to operate in one of the following modes: (1) "RF only", (2) "battery only" or (3) "combination of both".

24. The system of claim **23** wherein said receiving unit, when said transmitting unit is set to operate in said "RF only" mode, is operable to supply electrical energy to said implantable device, so long as said transmitting unit is located proximate to said receiving unit and said receiving unit is sensing transmitted RF energy.

25. The system of claim **23** wherein said receiving unit, when said transmitting unit is set to operate in said "battery only" mode, is operable, periodically, to supply electrical energy to said implantable device from said rechargeable power supply for a period of at least 7 days.

26. The system of claim **23** wherein said receiving unit, when said transmitting unit is set to operate in said "battery only" mode, is operable, periodically, to supply electrical energy to said implantable device from said rechargeable power supply for a period of at least 24 hours.

27. The system of claim **23** wherein said receiving unit, when said transmitting unit is set to operate in said "combination" mode, is operable to supply electrical energy to said implantable device through a rectifier directly to said implanted medical device, so long as said transmitting unit is located proximate to said receiving unit, and, separately, to "trickle charge" said rechargeable power supply.

28. An RF coupled implantable medical system comprising:

a transmitting unit;

a receiving unit including an implantable, electrically operated, medical device;

said transmitting unit including RF energy transmitting means, RF signal receiving means and first control means coupled to said RF energy transmitting means and to said RF signal receiving means for controlling the amount of RF energy transmitted to said receiving unit;

said receiving unit including RF energy receiving means, RF signal transmitting means, a rechargeable power supply coupled to said RF energy receiving means and second control means for adjusting the charging current flowing into said rechargeable battery coupled to said rechargeable power supply means, to said RF energy receiving means, to said RF signal transmitting means and to said implanted medical device; and,

said receiving unit comprising means for measuring the charge level of said rechargeable battery and, upon sensing a fully charged battery, automatically up-linking a coded signal which commands said transmitting unit to "stop" transmitting RF energy.

* * * * *

US008552597B2

(12) **United States Patent**
Song et al.

(10) **Patent No.: US 8,552,597 B2**
(45) **Date of Patent: Oct. 8, 2013**

(54) **PASSIVE RF ENERGY HARVESTING SCHEME FOR WIRELESS SENSOR**

(75) Inventors: **Zhen Song**, Plainsboro, NJ (US); **Chellury R. Sastry**, South Brunswick, NJ (US)

(73) Assignee: **Siemens Corporation**, Iselin, NJ (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 1542 days.

(21) Appl. No.: **11/691,545**

(22) Filed: **Mar. 27, 2007**

(65) **Prior Publication Data**

US 2012/0256492 A1 Oct. 11, 2012

Related U.S. Application Data

(60) Provisional application No. 60/744,028, filed on Mar. 31, 2006.

(51) **Int. Cl.**
H02J 17/00 (2006.01)

(52) **U.S. Cl.**
USPC **307/149**; 307/104

(58) **Field of Classification Search**
USPC 307/104, 149
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2,462,102	A	2/1949	Istvan	
2,554,270	A	5/1951	Rosenberg	
2,805,332	A	9/1957	Bell	
3,327,222	A	6/1967	King, Jr	
3,987,448	A *	10/1976	Scheppman	343/702
3,996,522	A	12/1976	Koch et al.	
4,229,742	A	10/1980	Rotunda	
5,075,670	A	12/1991	Bower et al.	
5,151,848	A	9/1992	Finello	
5,585,999	A	12/1996	De Long et al.	
6,037,743	A	3/2000	White et al.	
6,259,372	B1	7/2001	Taranowski et al.	
6,417,653	B1	7/2002	Massie et al.	
6,691,236	B1 *	2/2004	Atkinson	713/320
6,825,754	B1	11/2004	Rolin	
6,856,291	B2 *	2/2005	Mickle et al.	343/701
6,882,128	B1	4/2005	Rahmel et al.	

(Continued)

FOREIGN PATENT DOCUMENTS

WO	WO 01 37497	A1	5/2001

OTHER PUBLICATIONS

Pervasive Computing, Published by the IEEE CS and IEEE ComSoc, 1536-1268/05, copyright 2005 IEEE. Energy Scavenging for Mobile and Wireless Electronics, by Joseph A. Paradiso and Thad Starner. pp. 18-27.*

Primary Examiner — Fritz M Fleming

(57) **ABSTRACT**

A radio frequency energy harvester harvests RF energy from the environment. The energy harvester may include a local power source. The energy harvester may have an associated crystal radio antenna that harvests AM radio waves ubiquitously. The radio frequency energy harvested may be stored via a super capacitor. The stored energy may power a self-contained processing unit directly when the level of the energy harvested remains above a threshold level. However, if the energy harvested drops below the threshold level, the processing unit, which may include a radio, may switch to the local power source and/or a low power mode of operation. The low power mode of operation may reduce the amount of processing and/or transmitting by the radio. The processing unit may be part of a network of radios. The energy harvester may supply energy for a variety of applications, including building automation, industrial automation, power generation, and healthcare.

8 Claims, 9 Drawing Sheets

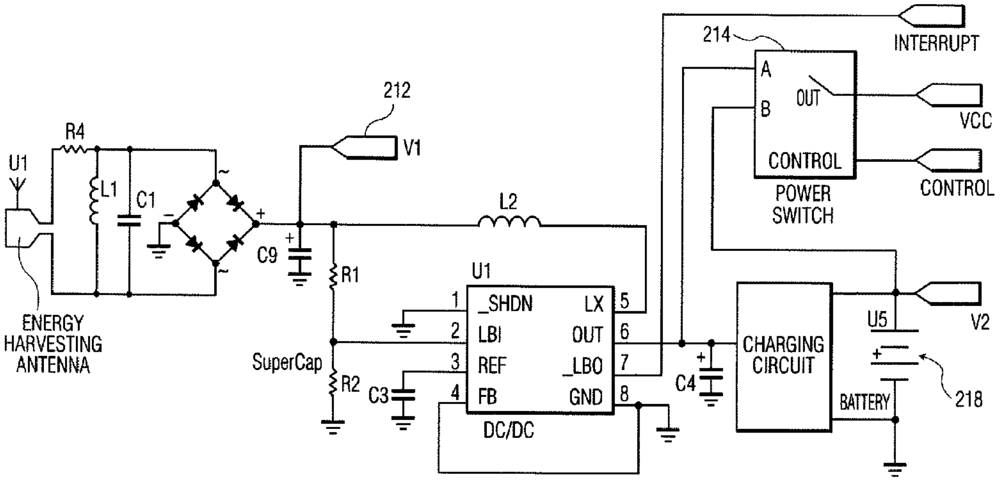

(56) References Cited

U.S. PATENT DOCUMENTS

6,924,772	B2	8/2005	Kiernan, Jr. et al.	
7,038,630	B1	5/2006	Bally et al.	
7,089,089	B2	8/2006	Cumming et al.	
7,132,757	B2 *	11/2006	Steigerwald et al.	290/1 R
7,139,239	B2	11/2006	McFarland	
2002/0116460	A1	8/2002	Treister et al.	
2005/0115600	A1 *	6/2005	DeSteese et al.	136/205

* cited by examiner

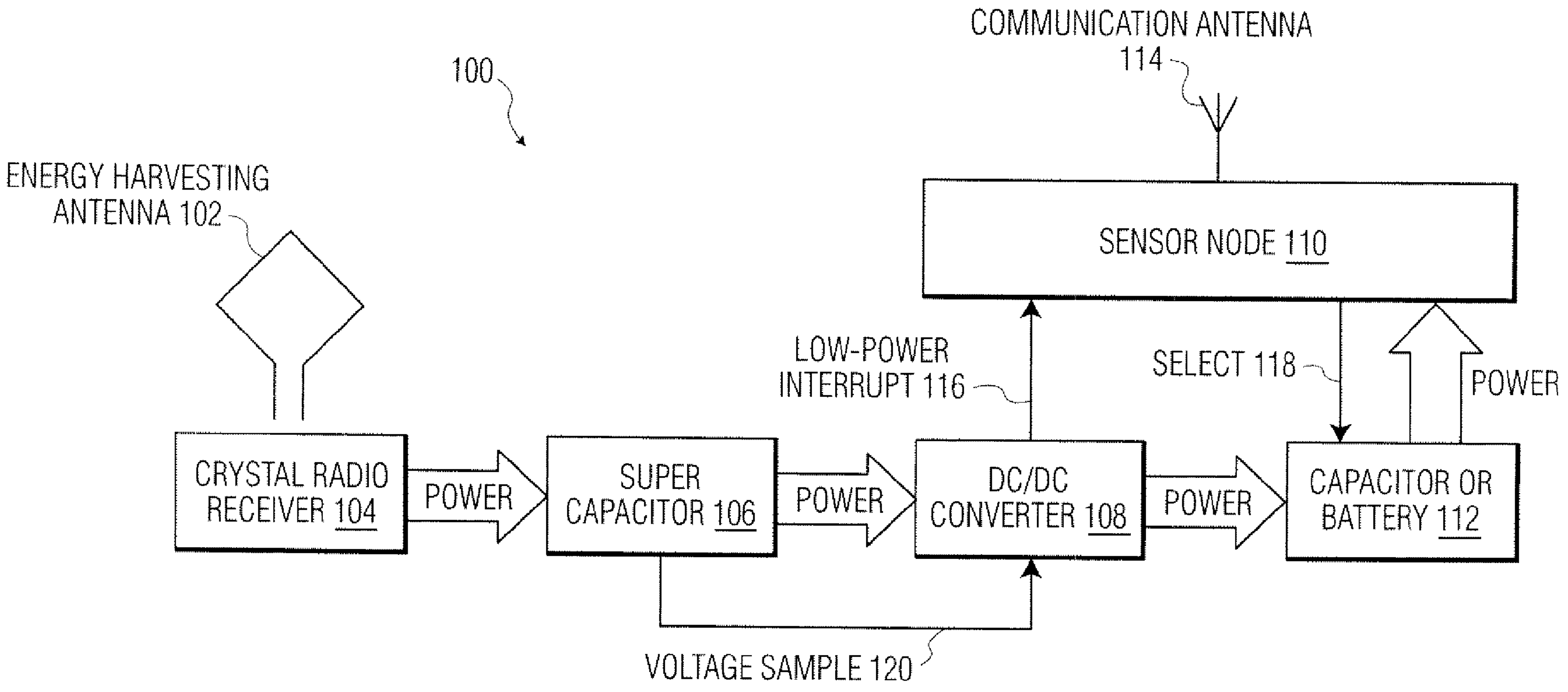
100
ENERGY HARVESTING ANTENNA 102
CRYSTAL RADIO RECEIVER 104
POWER
SUPER CAPACITOR 106
POWER
DC/DC CONVERTER 108
POWER
CAPACITOR OR BATTERY 112
POWER
COMMUNICATION ANTENNA 114
SENSOR NODE 110
LOW-POWER INTERRUPT 116
SELECT 118
VOLTAGE SAMPLE 120

FIG. 1

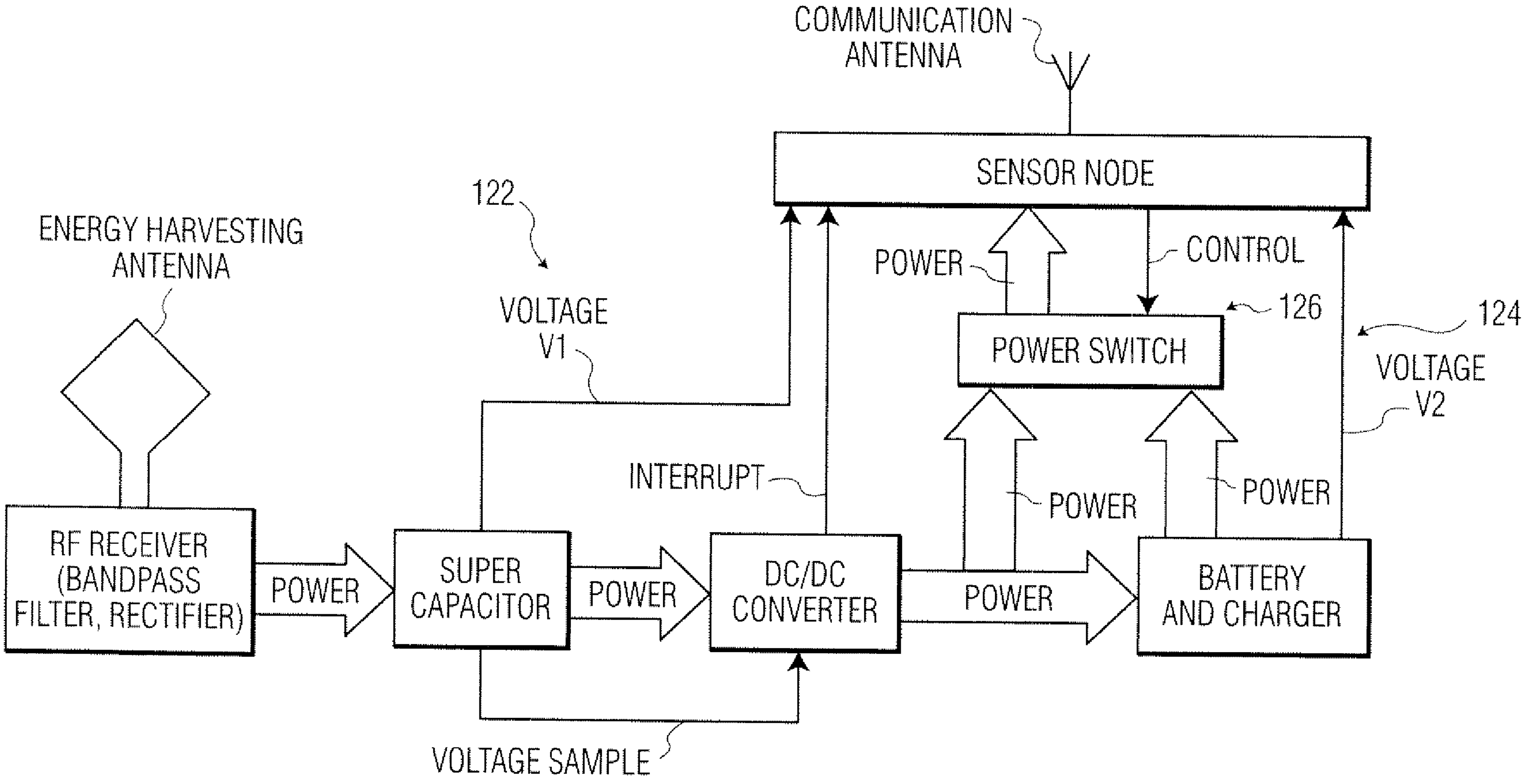
COMMUNICATION ANTENNA
SENSOR NODE
ENERGY HARVESTING ANTENNA
122
VOLTAGE V1
POWER
CONTROL
126
124
POWER SWITCH
VOLTAGE V2
INTERRUPT
POWER
POWER
RF RECEIVER (BANDPASS FILTER, RECTIFIER)
POWER
SUPER CAPACITOR
POWER
DC/DC CONVERTER
POWER
BATTERY AND CHARGER
VOLTAGE SAMPLE

FIG. 1a

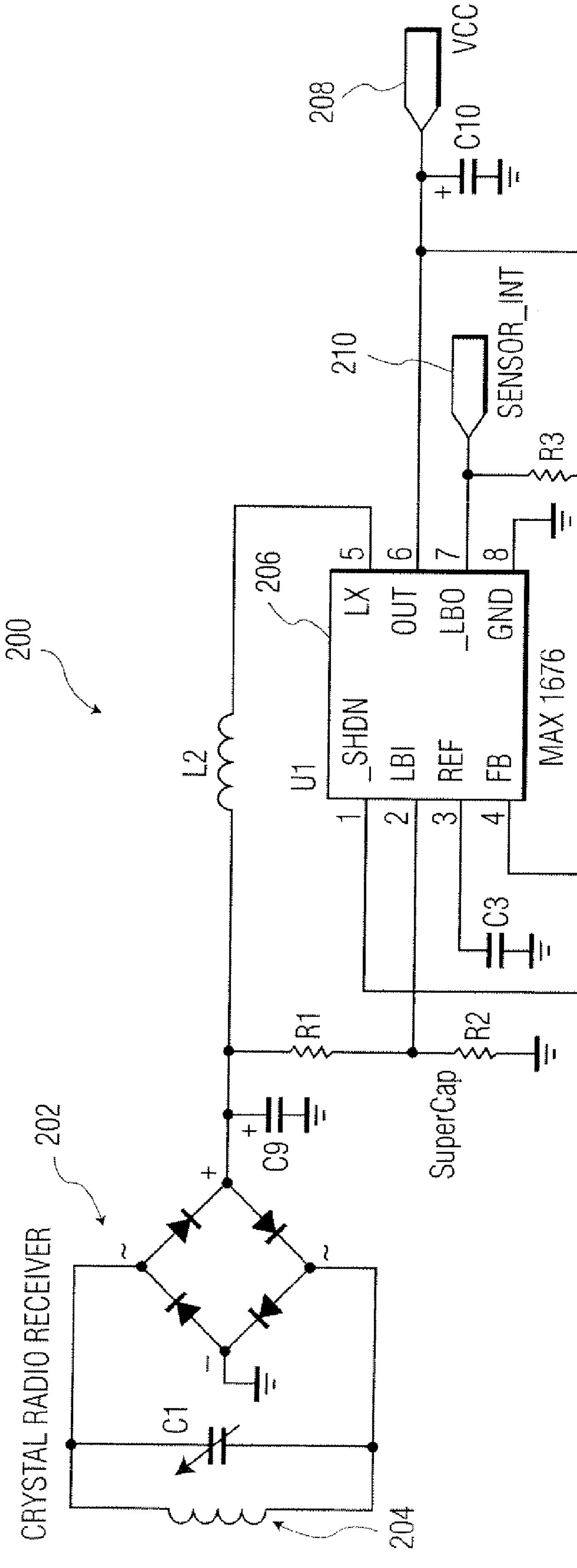
200
CRYSTAL RADIO RECEIVER
202
204
C1
~
~
+
-
C9
+
SuperCap
R1
R2
L2
U1
206
_SHDN
LBI
REF
FB
LX
OUT
_LBO
GND
MAX 1676
1
2
3
4
5
6
7
8
C3
R3
210
SENSOR_INT
C10
+
208
VCC

FIG. 2

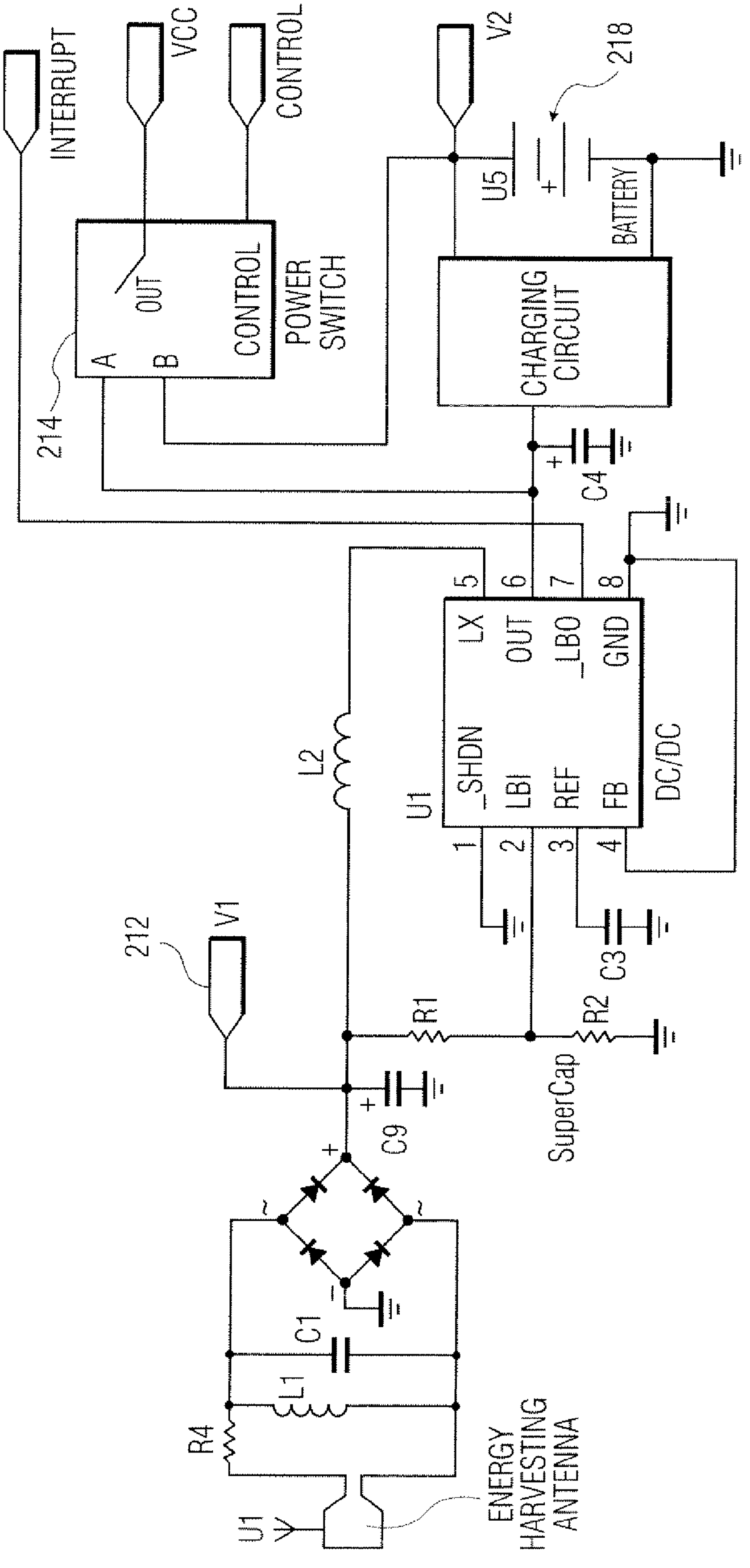
INTERRUPT
VCC
CONTROL
V2
218
U5
+
BATTERY
214
A
B
OUT
CONTROL
POWER SWITCH
CHARGING CIRCUIT
+
C4
5
6
7
8
LX
OUT
_LBO
GND
_SHDN
LBI
REF
FB
DC/DC
U1
1
2
3
4
L2
C3
212
V1
R1
R2
SuperCap
+
C9
+
~
~
-
C1
L1
R4
U1
ENERGY HARVESTING ANTENNA

FIG. 2a

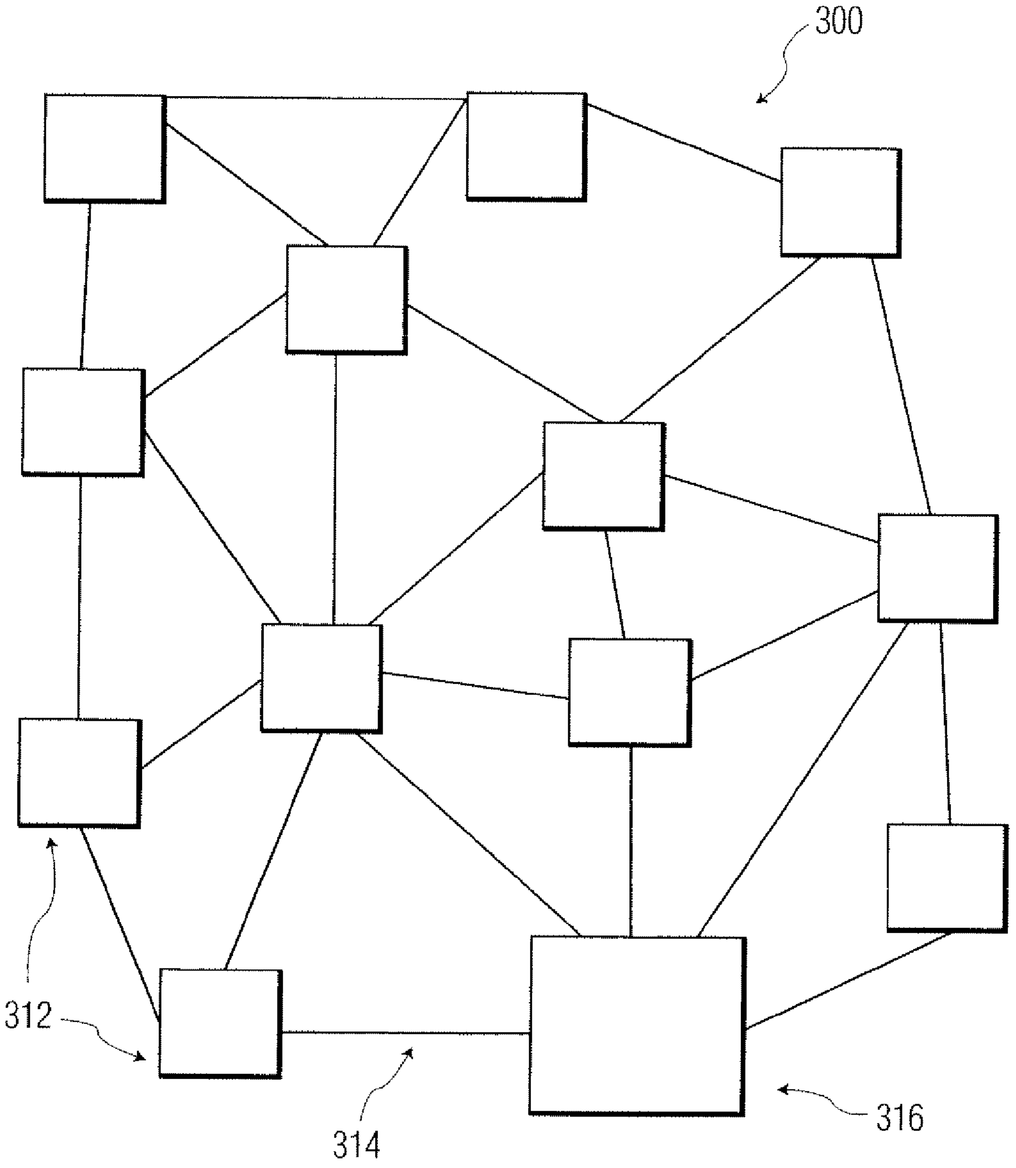
300
312
314
316

FIG. 3

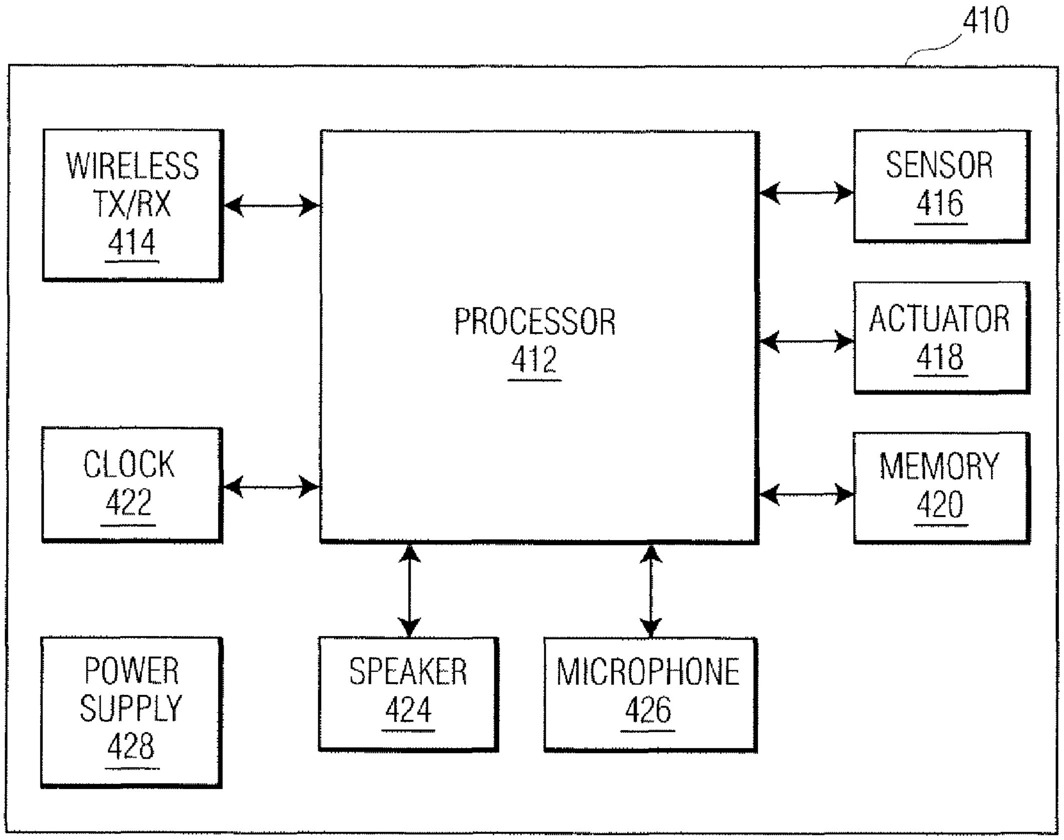
410
WIRELESS TX/RX 414
PROCESSOR 412
SENSOR 416
ACTUATOR 418
CLOCK 422
MEMORY 420
POWER SUPPLY 428
SPEAKER 424
MICROPHONE 426

FIG. 4

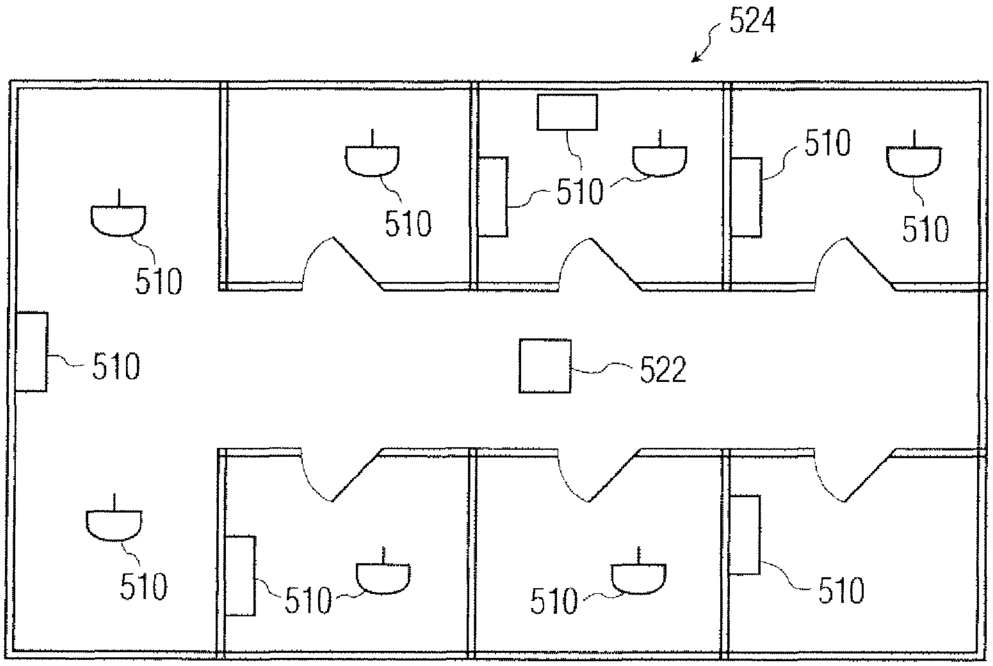

FIG. 5

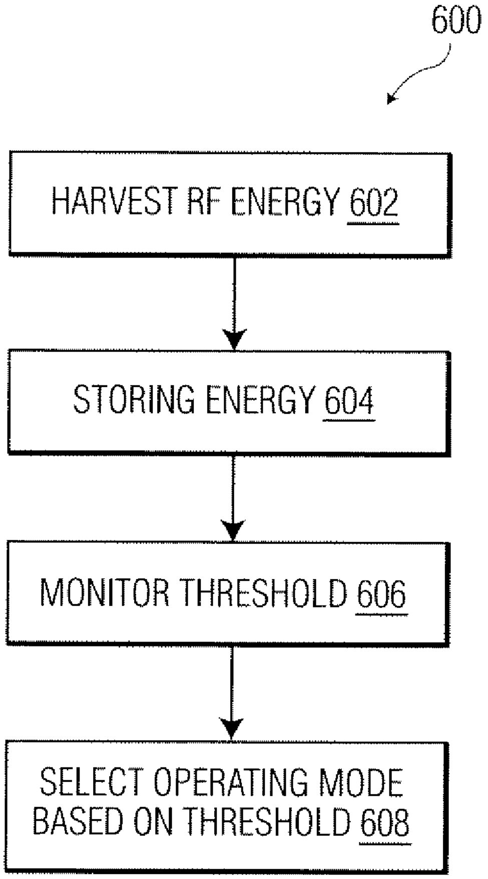
600
HARVEST RF ENERGY 602
STORING ENERGY 604
MONITOR THRESHOLD 606
SELECT OPERATING MODE BASED ON THRESHOLD 608

FIG. 6

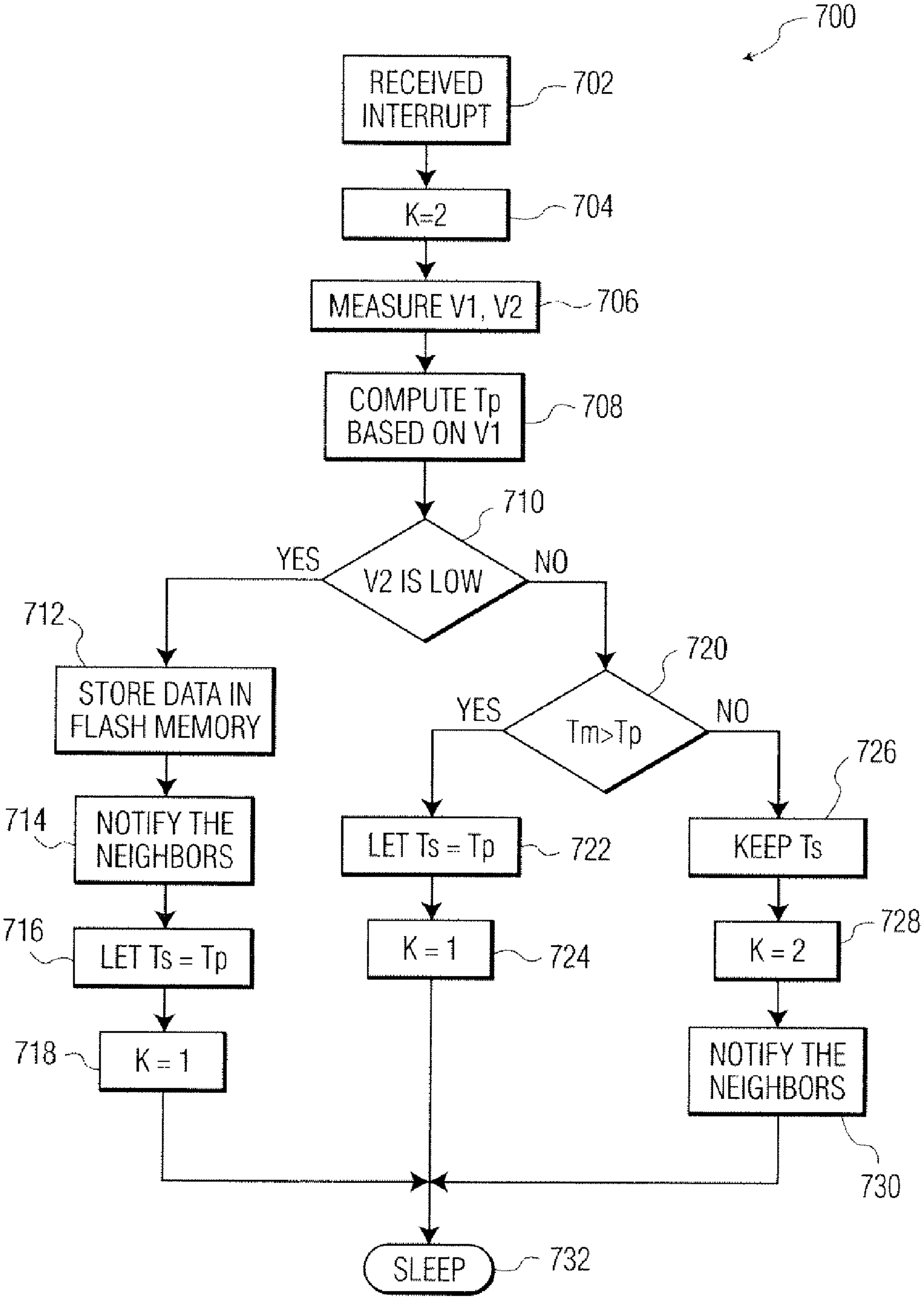
700
RECEIVED INTERRUPT
702
K=2
704
MEASURE V1, V2
706
COMPUTE Tp BASED ON V1
708
710
V2 IS LOW
YES
NO
712
STORE DATA IN FLASH MEMORY
714
NOTIFY THE NEIGHBORS
716
LET Ts = Tp
718
K = 1
720
Tm>Tp
YES
NO
LET Ts = Tp
722
K = 1
724
726
KEEP Ts
K = 2
728
NOTIFY THE NEIGHBORS
730
SLEEP
732

FIG. 7

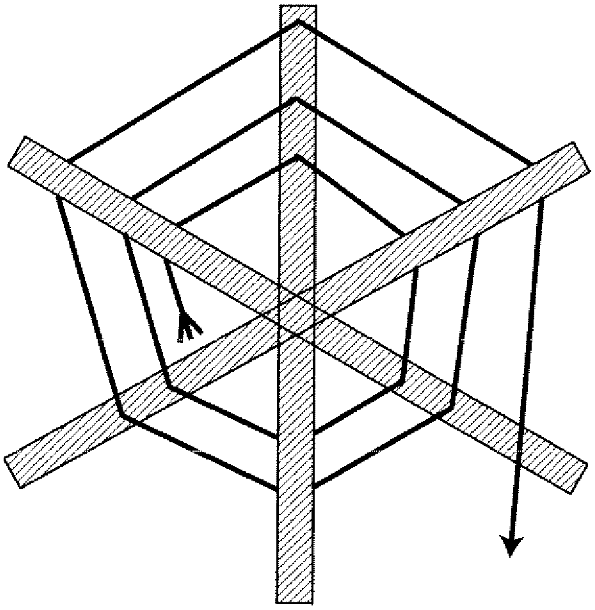

FIG. 8

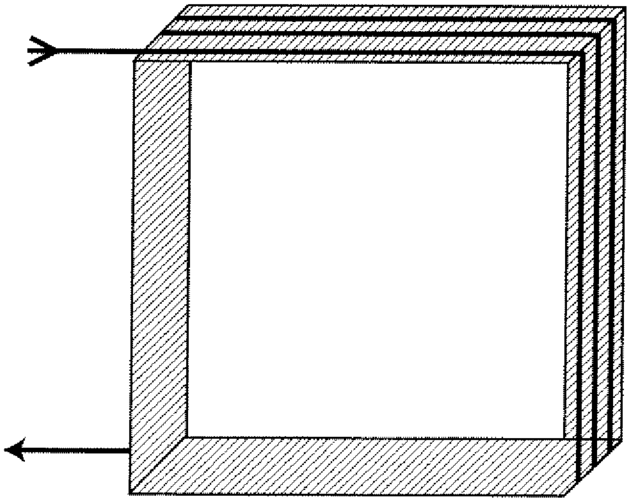

FIG. 9

PASSIVE RF ENERGY HARVESTING SCHEME FOR WIRELESS SENSOR

PRIORITY AND CROSS-REFERENCE TO RELATED APPLICATION

This application claims priority under 35 U.S.C. §119(e) to provisional application Ser. No. 60/744,028, filed on Mar. 31, 2006, which is incorporated by reference in its entirety herein.

BACKGROUND

The present embodiments relate generally to harvesting environmental energy. More particularly, the present embodiments relate to harvesting energy from radio frequency (RF) signals to power a sensor node and/or wireless radio.

Wireless sensor networks (WSNs) of so-called "smart" sensors can be used in remote monitoring and control applications. However, in almost all applications that involve WSNs, there is a trade off between the need to sense and communicate useful information and the requirement to do so in an energy efficient manner. A WSN, once deployed, may be intended to work for long periods of time with only minimal human intervention. For instance, remote monitoring applications may include individual sensor nodes, numbering into the hundreds or even thousands, that each run on a limited supply of energy, such as a battery. As a result, with conventional sensor nodes, excessive on-board processing and, perhaps more importantly, frequent invocation of the sensor node's radio may rapidly deplete the on-board energy source. Therefore, there is a need to develop sensor network protocols and algorithms with energy-efficient and self-organizing capabilities.

Additionally, replacing depleted batteries may be inefficient or impractical. A network may include a large number of individual sensor nodes that may be difficult to access due to being implanted within a wall, a ceiling, a larger device, or a patient. Accordingly, there is also a need for novel energy harvesting schemes operable to power wireless sensor nodes, as well as automatically replenish on-board energy storage units.

BRIEF SUMMARY

By way of introduction, the embodiments described below include methods, processes, apparatuses, instructions, or systems for harvesting energy from freely available radio frequency (RF) signals. The energy harvested may power processing units, such as self-contained processing units or sensor nodes. A "passive RF energy harvesting system," or "harvesting system," may be connected to a processing unit having a processor and delivers power supply to the processing unit. Each harvesting system may be associated with a crystal radio receiver and/or an all pass or band pass filter operable to harvest energy from all radio frequency bands simultaneously or ubiquitously. As a result, more of the available RF energy may be harvested from the existing electromagnetic signals. The harvesting is ubiquitous since no tuning is required if the harvester is moved from place to place. Most of the RF energy scavenged may be harvested from amplitude modulated (AM) radio signals. The harvesting system may have a super capacitor that temporarily stores the RF energy captured. If the level of RF energy being harvested falls below a threshold level, an interrupt signal may be generated that directs the processing unit to conserve energy. For instance, the processing unit may limit on-board processing or reduce the number and/or strength of signals transmitted via a wireless radio, or both.

In a first aspect, a method harvests passive radio frequency energy. The method includes harvesting radio frequency energy using a crystal receiver. The method also includes powering a processing unit directly from the radio frequency energy being harvested when the radio frequency energy being harvested or an associated signal is approximately equal to or above a threshold level, and alternatively powering the processing unit from an on-board power source when the radio frequency energy being harvested or the associated signal is below the threshold level.

In a second aspect, a method harvests passive radio frequency energy. The method includes harvesting radio frequency energy via a crystal receiver and storing the radio frequency energy harvested on a super capacitor for subsequent use.

In a third aspect, a system harvests passive radio frequency energy. The system includes a network of radios within a building operable to direct the operation of building equipment to control a building environment of a building and at least one radio within the network being a sensor node operable to harvest radio frequency energy via a crystal receiver, the radio frequency energy harvested being used to power the sensor node.

In a fourth aspect, a computer-readable medium having instructions executable on a computer stored thereon is described. The instructions include monitoring for an interrupt signal that is triggered when the radio frequency energy being harvested via a crystal receiver falls below certain threshold, the radio frequency energy being used to at least partially power a sensor node or charge the battery on the harvesting system. The instructions also include automatically reducing the amount of power being expended by the sensor node if the interrupt signal is detected.

The present invention is defined by the claims herein. Nothing in this section should be taken as a limitation on those claims. Further aspects and advantages of the invention are discussed below in conjunction with the preferred embodiments and may be later claimed independently or in combination.

BRIEF DESCRIPTION OF THE DRAWINGS

The components in the figures are not necessarily to scale, emphasis instead being placed upon illustrating the principles of the invention.

FIGS. **1** and **1***a* are system block diagrams of an exemplary passive RF energy harvesting systems;

FIGS. **2** and **2***a* are schematics of additional exemplary passive RF energy harvesting systems;

FIG. **3** is a schematic of an exemplary network of sensor nodes;

FIG. **4** is a block diagram of an exemplary sensor node;

FIG. **5** is a top plan view of an exemplary network of sensor nodes within a building;

FIG. **6** illustrates an exemplary method of harvesting radio frequency energy;

FIG. **7** illustrates an exemplary interrupt handling method; and

FIGS. **8** and **9** illustrate exemplary antennas.

DETAILED DESCRIPTION OF THE DRAWINGS AND PRESENTLY PREFERRED EMBODIMENTS

The present embodiments are directed toward methods, processes, apparatuses, instructions, or systems for harvest-

ing freely available radio frequency (RF) energy from the environment, such as signals generated from amplitude modulated (AM) and frequency modulated (FM) radio stations. The RF energy harvested is used to power a "sensor node." The sensor node may include a processor, a sensor, an actuator, a wireless radio, and/or other components. The sensor node may be part of a network of radios in direct or indirect communication with one another, such as a wireless sensor network (WSN).

The sensor node may have an associated crystal radio receiver. The crystal radio receiver may operate as an all pass filter to harvest energy from all radio frequency bands simultaneously or ubiquitously. As such, more of the available RF energy may be captured for the given electromagnetic signals being received by the sensor node at a specific time.

The energy harvesting system may have an associated super capacitor that temporarily stores the RF energy received or a related signal, such as a voltage, a current, or other electrical parameter or signal associated with the RF energy harvested. In one embodiment, if a voltage associated with the RF energy being harvested falls below a given threshold, an interrupt signal may be generated that directs the sensor node to switch to a "low-power" mode to conserve energy. In low-power mode, the sensor node may reduce the amount of processing that an on-board processor performs. Alternatively or additionally, in low-power mode, the operation of an on-board wireless radio or transmitter may be altered and/or limited. For example, the wake-up period of the sensor nodes may be increased for energy saving purposes. Other manners of reduced sensor node operation may be employed to conserve energy being expended by the sensor node.

I. Crystal Radio Receivers

Generally, the embodiments discussed herein provide energy harvesting or scavenging for sensor nodes and/or wireless sensor networks (WSNs). In particular, AM RF signals may be harvested via crystal radio receivers to provide the primary, or even sole, source of power for the sensor nodes. Other crystal receivers may be used.

Crystal radios were popular by the 1950's, but are now almost obsolete. One of the key features of crystal radios is that they do not require batteries. The earplugs and/or speakers of a crystal radio may be powered solely by amplitude modulated (AM) radio signals. The crystal radio may employ a piezo-electric crystal, dielectric disk, oscillating crystal, or another type of crystal. The crystal may oscillate or vibrate due to the radio frequency signals being received. Exemplary crystal radio receivers are disclosed by U.S. Pat. Nos. 2,554,270, 2,805,332, 3,327,222, and 3,996,522, which are all incorporated herein by reference in their entireties. Other crystal radio receivers may be used.

The crystal radio technology may be used in new applications due to the advancement of semi-conductors and reduced power requirements of sensor nodes. In other words, the available energy that may be harvested from radio frequency signals via a crystal radio receiver, may contain sufficient energy to power milliwatt level instruments and micro-devices. Additionally, the size of a crystal radio antenna may be configured to be relatively small, ranging from a few to tens of centimeters. An antenna of this size may be used with a number of today's wireless sensor node applications.

Harvesting small amounts of energy from radio signals has been studied before, although not in the context of powering nodes in a WSN by AM RF signals. Previous studies have concluded that RF energy is of little utility unless a receiver is very close to a transmitter, such as a cell phone station. An alternate approach would be to introduce an RF signal transmitter for the explicit purpose of transmitting RF energy to a receiver that then harvests the RF energy to perform certain tasks. In other embodiments, such an approach may not be practical in the context of WSNs.

The energy available from RF signals in the unlicensed band, for e.g., the 2.4 GHz spectrum, may be typically too limited to be of any practical use. Such RF power harvesting may be impractical due to (1) very limited power being available (for an antenna size of 100 cm^2, the received power may be less than 0.1 milliwatt), and (2) very limited range (not only may a transmitter be required, but the sensor also may have to be placed very close to the transmitter). On the other hand, the principles on which crystal radios operate may finally permit the successful harvesting of energy from freely available RF signals.

As AM radio signals may be strong enough to power an earplug or a speaker, the energy available from RF signals may be more than past estimates indicated. A typical sensor node may require about 20 milliamps in active mode and 5 microamps in sleep mode, and the operating voltage required may be about 3 volts. Thus, the required power for a typical sensor node is comparable to the power that may be harvested via RF signals. Additionally, low-duty low-energy sensors may not require a transmitter, further reducing the amount of energy needed to power the sensor node. Therefore, RF signals may be a viable "free" and persistent (available 24 hours a day) energy source for a considerable amount of a given radio broadcast receiving area.

II. Estimation of Received Radio Energy

Table 1 below illustrates an estimation of the available RF energy in Princeton, N.J. from AM and FM RF signals transmitted by area radio stations. Through this rough estimation analysis, the "free" RF energy in the Princeton area is shown to be within the 10 milliwatt level. The estimation is based upon the Friis propagation model,

$$P_r = \frac{p_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L},$$

TABLE 1

Radio Stations that can be Received in Princeton, N.J.

Call Sign	Frequency	Dist./Signal	City	Format	Estimated Received Energy (w)
WNJT	88.1 FM	7.1 mi.	Trenton, NJ	Public Radio	$2.8 * 10^{-6}$
WWFM	89.1 FM	6.7 mi.	Trenton, NJ	Classical	$3.1 * 10^{-6}$
WPST	94.5 RM	17.7 mi.	Trenton, NJ	Top-40	$0.40 * 10^{-6}$
WTHK	97.5 FM	12.5 mi.	Burlington, NJ	Classic Rock	$0.75 * 10^{-6}$
WMGQ	98.3 FM	10.7 mi.	New Brunswick, NJ	Adult Contemporary	$1.0 * 10^{-6}$

TABLE 1-continued

Radio Stations that can be Received in Princeton, N.J.

Call Sign	Frequency	Dist./Signal	City	Format	Estimated Received Energy (w)
WAWZ	99.1 FM	19.3 mi.	Zarephath, NJ	Christian Contemporary	$0.30 * 10^{-6}$
WKXW	101.5 FM	7.1 mi.	Trenton, NJ	News/Talk	$2.1 * 10^{-6}$
WPRB	103.3 FM	7.2 mi.	Princeton, NJ	College	$2.0 * 10^{-6}$
WWJZ	640 AM	24.5 mi.	Mount Holly, NJ	Children's	$9.1 * 10^{-3}$
WFAN	660 AM	55.0 mi.	New York, NY	Sports	$1.7 * 10^{-3}$
WOR	710 AM	40.4 mi.	New York, NY	Talk	$2.7 * 10^{-3}$
WABC	770 AM	46.1 mi.	New York, NY	News/Talk	$1.8 * 10^{-3}$
WCBS	880 AM	55.0 mi.	New York, NY	News/Talk	$0.95 * 10^{-3}$
WPHT	1210 AM	32.9 mi.	Philadelphia, PA	Talk	$1.4 * 10^{-3}$
WHWH	1350 AM	9.5 mi.	Princeton, NJ	Business News	$13.5 * 10^{-3}$

Here, p_t, p_r are the power of the transmitter and the receiver, respectively; G_t, G_r are the antenna gains of the transmitter and the receiver, respectively; λ is the wavelength of the RF signal in meters; d is the distance between the transmitter and the receiver; and L is the system losses ($L>1$). The estimation is based upon various assumptions, such as (1) radio station information (frequency and distance), (2) the transmitting power of AM radio stations being 10 kilowatt and FM stations being 5 kilowatt, (3) the antenna gains, G, are 1 (for a properly designed antenna, $G>1$), and (4) the system losses, L, are 1.

The results of the estimation indicate that (1) the total power received was 31.120 milliwatts, (2) the AM power received was 31.108 milliwatts, or 99.96% of the total power, and (3) the FM power received was 0.0126 milliwatt, or 0.04% of the total power. The estimation is based on ideal environments. In practice, L is more than 1 and the received power is less than the ideal case. Observing the Friis propagation model reveals that the effect of wavelength on the received power is more than that of the transmission power. Accordingly, signals with longer wavelength, such as AM signals, are much more suitable for RF energy harvesting.

A typical sensor node may require about 41 milliwatts in active mode and 1.5 microwatts in sleep mode. This power requirement is of the same order of magnitude as the power that is available from RF signals. With reasonable conditions, such as the sensor nodes being located near one or more AM radio stations and/or using relatively efficient antennas, the sensor nodes may be powered solely by the AM signal(s) and be kept running uninhibited all of the time. Alternatively, the sensor nodes may be powered primarily by the AM signal(s), if backed up by an on-board energy source.

III. Estimation of Available Energy from Other RF Signal Sources

As noted above, AM signals may include much more energy available for harvesting than FM signals. Energy from other RF signal sources also may be very small when compared with AM signals. For example, the RF signals from cell phones quickly attenuate as a function of distance. As a result, a cell phone signal may be used as a source of energy to power a sensor node only if the distance from the cell phone to the sensor node is about one meter or less.

Energy also may be received from IEEE 802.11.x protocol-based devices and television sets. As IEEE 802.11.x devices operate in the 2.4 GHz band, the energy available for harvesting may be even less than what is available from cell phones. Similarly, due to their high frequency, television signals may contribute very little towards harvestable energy. Therefore, AM signals may provide the best freely available RF source of power for wireless sensor nodes.

IV. Hybrid Scheme

A number of harvesting schemes may be utilized to capture energy from RF signals. As an example, harvesting energy from AM signals may not only be useful for indoor sensor nodes, but could also prove fruitful for outdoor sensor deployments.

In one aspect, the RF harvesting schemes discussed herein may combine two concepts. A sensor node may include crystal radio technology that is operable in conjunction with an energy storage unit, such as a battery and/or a capacitor. Once the output of the crystal radio receiver is connected to act as a source of energy and/or a "charger" to the energy storage unit, the system may be ready for use.

The sensor node may include a RF energy harvesting device that includes a so-called "energy harvesting antenna" operable to capture RF energy. The antenna may harvest AM, as well as FM, or RF signals. The antenna may be a spider antenna, braided antenna, omni-directional antenna, double-braided polyester rope, loop antenna, air coil loop antenna, sparrow antenna, Yagi-Uda antenna, or other type of antenna.

For example, FIG. **8** is an exemplary spider antenna. As shown, the antenna may have a number of legs of differing size. FIG. **9** is another exemplary antenna. Other shapes may be used, such as circular or disk shapes.

The antenna may be a laminated antenna having a number of wires etched onto a semi-conductor or other micro-device. The antenna wires may have either a flat or rounded shape, or be disk shaped. The antenna may have a number of spokes or legs, such as the spider antenna disclosed by U.S. Pat. No. 4,229,742, or the antenna may be a grid antenna similar to the one disclosed by U.S. Pat. No. 7,038,630, both of which are incorporated herein by reference in their entireties. As the antenna characteristics may be determined based upon the bandwidth and gain properties desired for a specific application, alternative antennas may be used.

In one embodiment, for both convenience and cost reduction purposes, a 1950's type of coil or wire antenna connected to the crystal radio receiver may be replaced by modern flexible printed circuit board. Printed circuit board similar to that used in conventional radio frequency identification (RFID) devices may be used. An exemplary RFID antenna is disclosed by U.S. Pat. No. 6,825,754, which is incorporated herein by reference in its entirety. Other types of printed circuit board material and shapes may be used as the antenna.

Once the RF energy is scavenged, the sensor node may regulate the energy within the crystal radio circuit and convert

the RF signals received to a DC signal, such as via a rectifier, without any frequency filtering. As an example, the RF signals harvested may not be filtered via low, high, or band pass filters. After the RF energy harvested passes through the rectifier, the energy may be stored on a super capacitor. However, since the voltage on the two sides of the super capacitor may not satisfy the requirements of the sensor node, a DC/DC converter and/or another capacitor may be employed (such as shown in FIG. **2**, discussed below).

Clean DC power may be available on both sides of the second capacitor. In addition, the DC/DC converter may sample the voltage on the super capacitor, and issue an interrupt to the sensor node once the voltage, which is an indication of the energy being stored and/or harvested, is lower than a certain threshold. The sensor node may then switch to the on-board power source of the harvesting system, such as a battery, a capacitor, or other energy storage device. If the energy being harvested provides a voltage level greater than that of an on-board energy storage device, the energy being harvested may power the sensor node directly, as well as provide a trickle charge to the on-board energy storage device. On the other hand, if the energy being harvested provides a voltage level lower than of the on-board energy storage device, the on-board energy storage device may provide power to the sensor node via a trickle discharge. The harvested RF energy is used to charge the battery under this case.

Therefore, the wireless and self-contained sensor nodes may employ a "hybrid" RF energy harvesting scheme. Each wireless and self-contained sensor node may be equipped with (1) a RE-based energy harvester and (2) an on-board power source. As long as there is enough available power from the RF energy harvester, the sensor node may operate using the RF power. Intermittently, however, the power from the harvester may not be at the threshold required to operate the sensor node. As a result, an interrupt may be generated that switches the sensor node to the on-board power source. The hybrid scheme ensures that the sensor node may be continuously operational while simultaneously greatly lengthening the life-time of the on-board power source.

There are some distinctions between RF energy harvesting as discussed herein and conventional harvesting schemes. With respect to solar cell techniques, the proposed RF harvesting schemes are not affected by the weather or time of day (darkness). The proposed RF harvesting schemes may provide a continuous power supply as most radio stations broadcast around the clock.

Additionally, the proposed RF harvesting schemes may provide more power than solar cells, such as with indoor environments, and at less cost, as typical solar cells may be more expensive than an RF antenna or coil. The costs associated with the proposed schemes may be further reduced if the antenna is implemented using technology similar to that of a RFID tag.

As compared to biomechanics-based harvesting schemes, the proposed RF harvesting schemes may be more suitable for sensor node applications. Conventional biomechanics harvesting methods tend to provide energy on an inconsistent basis and inconveniently toggle sensors periodically to recharge the power source. On the other hand, the proposed RF harvesting schemes may provide a consistent milliwatt-level power source. Toggling may be used with RF harvesting.

With respect to vibration or thermoelectric methods, it may be more difficult to find a vibration or heat source than it is to locate an AM radio station, especially in populated regions of the United States. When a free source is not available, the vibration or thermoelectric based schemes may require an artificial source. In many environments, such as an office or building, installation of such artificial energy sources may not be feasible.

In sum, new RF energy harvesting schemes are described that provide energy to self-contained sensor nodes. The schemes provide an improved crystal radio receiver and harvest AM RF signals as a source of power. The RF harvesting schemes are suitable for indoor applications that involve wireless sensor networks, such as building automation applications.

The RF harvesting schemes may be especially useful in areas having AM stations within about 20 kilometers to satisfy typical sensor node power requirements. However, the effective distance of the harvesting schemes may be extended. The antennas employed may be configured to harvest more energy and provide enhanced gain. Sensor nodes may be configured to be more energy efficient. For example, the sensor nodes may be programmed to operate more efficiently, such as spending more time in various reduced power modes of operation. Additionally, the sensor nodes equipped with RF energy harvesters may be supplemented with other types of energy harvesters, such as solar, piezo-electric, MEMS (micro-electro-mechanical systems), biomechanical, thermoelectric, and other harvesters. The RF harvesting schemes, except for requiring relatively nearby AM radio stations, may have no other substantial apparent limitations, such as the weather, periodic toggling, or other limits.

V. Exemplary RF Energy Harvesting Systems

FIG. **1** illustrates an exemplary passive RF energy harvesting system **100**. The system **100** may include an energy harvesting antenna **102**, a crystal radio receiver **104**, a super capacitor **106**, a DC/DC converter **108**, a sensor node **110**, a local power source **112**, and/or a communication antenna **114**. The system may include additional, fewer, or alternate components.

The crystal radio receiver **104** or crystal set may employ an antenna **102**. The crystal radio receiver **104** may not require a battery or power source except the power received from radio waves via the antenna **102**. The crystal radio receiver **104** may include an antenna wire, a tuning coil of copper wire, a rectifier, and other components. The crystal radio receiver **104** may be a passive radio receiver, not requiring an actively powered amplifier associated with conventional radios. The crystal radio receiver **104** may preserve electrical power from RF signals received via the antenna **102** and convert the RF signals to electronic power. The crystal radio receiver **104** may include various types of antennas and various "low-tech" or traditional electrical circuitry components, such as capacitors, resistors, inductors, and diodes. As such the crystal radio receiver **104** may be simpler and less expensive to manufacture than current "smart" radios, such as cellular phones.

The antenna **102** may be a conventional crystal radio antenna and include copper or silver wire coils, or a ferrite coil. The antenna **102** may be a Yagi, log-periodic, fiberglass, cable, repeater, or other type of antenna, including those discussed elsewhere herein. The antenna **102** may be manufactured from printed circuit board, aluminum, copper, silver, other metals, or other materials. The antenna may omni-directionally collect all of the available AM radio signals currently being received. The shape, size, circumference, and composition of the antenna **102** may be altered for a specific application and/or to enhance the gain associated with the antenna **102**.

The antenna **102** and crystal radio receiver **104** may act independently or in conjunction to harvest RF energy ubiquitously. The crystal radio receiver **104** may require no tuning and resonant with all of the available frequencies. As a result,

energy may be collected simultaneously from a number of radio stations, which may increase the average level of RF energy being harvested. This may be in contrast with typical receivers that may employ RC resonant circuits, which eliminate some RF signals from being harvested. In one embodiment, the antenna may have a physical length of between 10 and 20 cm.

The low power mode may include a number of algorithms and/or software instructions stored in the sensor node that direct the sensor node or an associated processor to conserve or reduce the amount of energy that the sensor node is using. A processing unit on the sensor node may perform reduced or limited processing in low power mode. The radio may change the message routing being employed based upon the amount of RF energy being sensed or harvested. A transmitter may transmit less often in low power mode, or not at all until the level of the RF energy being harvested returns to above a threshold or self-sustaining level. The transmitter may alter the type of signal being transmitted in low power mode. For example, the sensor node may not relay all the received communication packets, and/or the transmitter on the sensor node may transmit at a reduced power level. The transmission at a reduced power level may be intended to only reach the next or closest available sensor node or radio within the network.

The low power mode may include buffering or storing data, instead of constant or even intermittent transmission of information to other sensor nodes. With typical sensor nodes, the operation of the radio and/or transmitting information may require a substantial amount of energy. Therefore, reducing radio transmissions in low power mode may enhance the effectiveness of the RF harvesting schemes discussed herein.

The super capacitor **106** may store the RF energy being harvested. The capacitance of an electrode associated with the super capacitor **106** may be several hundred times the capacitance of a conventional double-layer capacitor. The super capacitor **106** may have one or more electrodes manufactured from titanium, palladium, aluminum, silicon, or other metals. Alternatively, the electrodes may be doped or have titanium, palladium, aluminum, silicon, or other metal coatings. The super capacitor **106** may be similar to those disclosed by U.S. Pat. Nos. 5,585,999 and 5,151,848, which are incorporated herein by reference in their entireties. Alternate types of super or other capacitors may be used.

The DC-to-DC converter **108** may operate to produce an output voltage signal and/or an interrupt signal **116** based upon a signal received from the super capacitor **106** or an electrical parameter of the super capacitor **106** monitored, such as a voltage, current, or other signal. The DC-to-DC converter **108** may be similar to the DC-to-DC converter disclosed by U.S. Pat. No. 6,417,653, which is incorporated herein by reference in its entirety. Alternate DC-to-DC converters may be used.

In one embodiment, the DC-to-DC converter **108** samples the voltage **120** on the super capacitor **106**. The voltage sampled may be a reflection the current level of RF energy being harvested crystal radio receiver **104** and/or stored by the super capacitor **106**. If the voltage sampled falls below a threshold level, such as a level required by the sensor node **110** to operate normally (i.e., with no power operating restrictions), then the DC-to-DC converter may generate a low-power interrupt **116** signal. The sensor node **110** may switch to a low or reduced power operating mode upon the receipt of the low-power interrupt **116**.

The on-board energy storage unit may be a capacitor or battery **112**. The on-board energy storage unit **112** may be on a trickle charge when the voltage sampled **120** from the super capacitor **106** remains above the threshold level (such as the actual battery or capacitor voltage) with the carryover RF energy harvested directly powering the sensor node **110**. Alternatively, the RF energy harvested may be sent directly from the DC-to-DC converter **108** to the sensor node **110** when the level of RF energy being harvested remains at or above the threshold level.

On the other hand, if the low-power interrupt **116** is sent to the sensor node **110**, the sensor node **110** may select **118** or switch to the on-board storage unit **112** as a direct source of power. Alternatively, when the voltage from the DC-to-DC converter falls below the threshold level (such as the actual capacitor or battery voltage), the sensor node **110** may automatically start to be powered by the on-board power storage unit **112** as the storage unit voltage would be higher than the output voltage of the DC-to-DC converter, as can be envisioned by FIG. **1**. Other manners of powering of the sensor node from the super capacitor, either solely or in conjunction with the storage unit, may be used.

The sensor node **110** may be mounted on a device that includes the other components shown in FIG. **1** or be associated with the other components in another fashion. The sensor node **110** may include a communication antenna **114** that provides for wireless communication with other sensor nodes. The sensor node **110** may include the components shown in FIG. **4** and discussed below, in addition to the components illustrated by FIG. **1**. The sensor node may include additional, fewer, or alternate components.

FIG. **1***a* illustrates another exemplary passive RF energy harvesting system **100**. In one embodiment, the system may be targeted to facilitate the powering of sensor nodes in an ad-hoc wireless sensor network. As each sensor node may have communication functionality through an on-board RF module, a message communication subsystem may not be required to be a part of the RF energy harvesting system **100**. Also, the RF energy harvesting system **100** may be purely passive, and thus, no registration with the FCC is required.

As shown in FIG. **1***a*, the system **100** may be connected to the sensor node and its communication antenna. The system **100** may be a band pass and/or all pass filter to match all AM frequencies in order to provide ubiquitous energy harvesting capability and also to harness the maximum available harvestable energy.

As shown in FIG. **1***a*, a super capacitor is introduced as an energy buffer to the rechargeable batteries. Super capacitors have large number of recharge cycles, but limited energy capacity when compared to rechargeable batteries. In one embodiment, since there is bound to be some variation of RF signal strength from time to time and place to place, the harvested RF energy may not be consistent. Thus, as can be seen from FIGS. **1** and **1***a*, harvested energy can be used to power the sensor node or charge the on-hoard power source (such as batteries) on each sensor node. Thus, the demands on recharge cycles of the batteries are relaxed and the battery lifetime may be improved.

FIG. **1***a* also illustrates the interrupt signal. If the harvested RF energy is not enough and falls below a certain threshold, the DC/DC converter (for example, a MAX1674/MAX1675) may trigger the sensor node via an interrupt signal as shown. The processor on the sensor node, upon receiving this interrupt, may measure the voltages of the super capacitor, V1 **122**, and that of the battery, V2 **124**. Depending upon what these measured values are, the sensor node may make intelligent decisions. In one embodiment, the sensor node controls a power switch **126** that is operable to switch between the super capacitor and the batteries as a source of power. An exemplary

flow chart of an interrupt handling process that may be implemented by software on the sensor node is illustrated by FIG. **7**.

FIG. **2** illustrates the circuit exemplary passive RF energy harvesting system **200**. The system **200** may include a crystal radio receiver **202**, an energy harvesting antenna **204**, a DC/DC converter **206**, a voltage output **208**, and an interrupt output **210**. The system may include additional, fewer, or alternate components.

As shown in FIG. **2**, the crystal radio receiver **202** may include a capacitor bank and a rectifier that converts the RF energy being received into a DC signal. The DC signal harvested may be stored upon the super capacitor C**9**. The DC/DC converter **206** may sense the voltage on the super capacitor C**9** and provide an output voltage VCC **210** to the sensor node, as well as the sensor interrupt **210** when the output voltage approximately reaches or falls below a threshold level. The sensor nodes may use RF energy harvesting systems having additional, fewer, or alternate components.

FIG. **2***a* illustrates another embodiment of the circuit of the exemplary passive RF energy harvesting system **200**. With respect to FIG. **1***a*, FIG. **2***a* illustrates in more detail from where V**1** **212** and V**2** **218** may be sensed. V**1** **212** may be sensed from the super capacitor C**2**. V**2** **218** may be sensed from the battery. FIG. **2***a* also shows the power switch **214**. The system may include additional, fewer, or alternate components.

VI. Exemplary Network

FIG. **3** illustrates an exemplary network **300** of wireless radios or sensor nodes **312**. The network **300** may utilize a dynamic routing algorithm that permits data transmitted to travel the shortest distance or link **314** between sensor nodes **312** to a destination, which decreases the required transmission time for a given message, as well as the required power level of that transmission. The destination may be another sensor node **312** or a control radio **316**. Each wireless and self-contained sensor node **312** and control radio **316** may have a dedicated processor, a receiver, and a transmitter. The network **300** may include additional, fewer, or alternate components.

Each sensor node **312** may communicate its associated routing information to every nearby or adjacent sensor node **312** or control radio **316**. After a sensor node **312** receives a data transmission, a processor of the sensor node **312** may determine what to do with that data, including whether to retransmit the data to an adjacent or nearby sensor node **312** or control radio **316**. The base station **316** may function as a network controller that directs the overall operation of the network **300**.

The network **300** may provide continuous communication with otherwise unavailable sensor nodes **312**. For instance, some sensor nodes **312** may become obstructed by obstacles, such as equipment, containers, furniture, or other items, or may fail. However, the network **300** may reconfigure itself around blocked paths by redirecting transmissions from one sensor node to the next until communication with a lost sensor node is re-established. The network **300** also may provide enhanced communication reliability between sensor nodes **312** as a single sensor node **312** may be in direct communication with a number of other sensor nodes **312**, as shown in FIG. **3**.

The network **300** may implement IEEE 802.15.4 protocols. Other protocol standards may be used. The network **300** may operate as a mesh network, as described in more detail below. Alternate control or routing algorithms/instructions may be used.

VII. Control of Building Equipment

One application of the sensor nodes described herein relates to a network that controls building equipment. The network may include multiple sensor nodes, each having a wireless radio, as well as one or more control radios or base stations that direct the network. Each sensor node may be a so-called "smart" sensor node that includes a receiver, a transmitter, a processor, memory, one or more sensors, an actuator, and/or other components. Each sensor node may transmit messages to a control radio acting as network controller. Alternatively, the network controller may be a dedicated processor. The network may have one or more network controllers and/or control radios. The term network herein may include the entire network, a sub-set of a network, a number of sensor nodes, one or more network controllers, one or more control radios, or a combination of sensor nodes with one or more network controllers or control radios.

A network controller may assimilate and analyze a number of messages received from a plurality of sensor nodes. In response to each of the messages received, the network controller may determine that a change in the currently operating building equipment, or the operating modes thereof, is in order. Subsequently, the network controller may transmit a message to one or more sensor nodes that direct the operation of building equipment. Upon receiving the message, a sensor node may alter the operation of building equipment.

The sensors associated with the sensor nodes/wireless radios may monitor specific parameters pertaining to building environmental conditions or specific operating equipment. The actuators associated with the sensor nodes may control the operation of certain building equipment. A sensor node may transmit the value of a parameter sensed by a sensor to the network. In response to the values of the parameters received, the network or another sensor node may automatically alter the operation of building equipment, such as by sending messages that operate the actuators that control the building equipment.

For example, the sensors may be temperature sensors that sense the temperature in an area of a building. Each temperature sensor may be connected with a sensor node having a wireless radio, the sensor nodes being dispersed throughout a building. Each sensor node having a temperature sensor may transmit a message to the network regarding the temperature sensed in the building area in which the sensor node is located. In response to the temperature information received, the network may direct that cooling, heating, ventilation, HVAC, emergency, or other building equipment be operated to alter the building environment of the building area in which the sensor node is located.

The network may employ multiple sensor nodes in each building area to monitor temperature. Conventional wall mourned temperature sensors and/or thermostats may be single point sources of information. However, the average value of individual temperature parameters received from a plurality of temperature sensors dispersed in a given building area may better reflect the actual temperature in the building area. Accordingly, the building environmental equipment may be directed to maintain the temperature of a building area closer to the desired temperature based upon the more accurate temperature information received.

The sensors also may be motion sensors that sense motion in a building area. Each motion sensor may be connected with a sensor node having a wireless radio, the sensor nodes being dispersed throughout a building. Each sensor node having a motion sensor may transmit a message to the network regarding the motion sensed in a building area. In response to the motion information received, the network may direct the operation of building equipment.

The motion detected may alert the network that a building area has recently become occupied or unoccupied. In response, the network may ensure that lighting equipment provides adequate light in or near the building area in which motion was sensed. The network may direct that building environmental equipment, such as cooling, heating, ventilation, HVAC, or other equipment, be operated to alter the building environment of the building area. The motion information received also may be used by the network to determine that a security breach has occurred. Accordingly, the network may trigger an alarm, secure passageways, and operate other security equipment in response to the security breach.

A sensor node having a wireless radio may be connected with an identification device located on an individual. After the sensor node located on the identification device transmits a message to the network, the network may determine the identification and/or location of the associated individual. In response, the network may transmit instructions to building environmental equipment to automatically alter the environmental conditions of the specific building area in which the individual is currently located based upon stored or transmitted environmental preferences associated with that individual.

In a building having numerous pieces of operating equipment, it may be desirable to automatically monitor various parameters associated with various pieces of equipment. For instance, in a power plant, refinery, factory, or other plant, it may be advantageous to monitor temperatures, pressures, alarms, tank levels, bilge levels, hydraulic levels, atmospheric conditions, operating pumps or fans, and other parameters. The change in various temperatures, pressures, levels, or equipment operating temperatures may indicate problematic conditions.

The network may automatically identify problematic conditions associated with operating building equipment. The various parameters monitored each may be sensed by a sensor located on a sensor node. The sensor node having a wireless radio may transmit the value of the parameter to the network, either periodically or upon being queried by the network or sensing an out of specification value. The sensor node may determine whether a parameter is within specification, i.e., a predetermined satisfactory range.

If a parameter is not within specification, the network may take corrective action to restore the parameter and/or building conditions to specification. For example, the running speed of a problematic piece of equipment may be shifted, increased, or decreased. The problematic piece of equipment also may be secured and an alternate piece of equipment may be started or placed on line to replace it. Additional, fewer, or alternate courses of action may be taken to correct problematic or out of specification parameters. Sensor nodes may be connected with or part of the equipment being controlled for wireless control of the equipment.

VIII. Mesh Network

In one embodiment, the network may include a number of sensor nodes, each having a wireless radio, arranged as a mesh network that also may be used to locate movable assets and/or operate building environmental equipment. The mesh network provides the capability of routing data and instructions between and among the network of radios. The mesh network permits data to be efficiently transmitted from one radio in the network to the next until the data reaches a desired destination.

The mesh network may be implemented over a wireless network or partially wireless network. Each sensor node having a radio within the network may function as a repeater that transmits data received from adjacent sensor nodes having radios to other nearby sensor nodes that are within range. The coverage area of the mesh network may be increased by adding additional radios. As a result, a network may be established that may cover an area of desired size, such as a floor of a building or an entire building.

Each sensor node having a radio within the mesh network is typically only required to transmit data as far as the next sensor node having a radio within the network. Hence, if a sensor node having a wireless radio has a limited power supply, the reduction in the distance that each sensor node is required to transmit permits lower power level transmissions, which may extend the operating life of the on-board power supply.

A number of protocols may be used to implement the mesh network. The sensor nodes may implement a protocol that uses low data rates and low power consumption. As noted above, the mesh network may employ devices that use very small amounts of power to facilitate significantly increased battery or power supply life. In some situations, power supply life may be extended by minimizing the time that the radio device of the sensor node is “awake” or in normal power using mode, as well as reducing the power at which a signal is transmitted.

Alternatively, the sensor nodes having wireless radios may implement a protocol that uses moderate or high data rates and power consumption. For instance, the sensor nodes may implement IEEE 802.11 protocols. An IEEE 802.11 LAN may be based on a cellular architecture where the system is subdivided into cells, where each cell is controlled by a base station. Other protocols may be implemented.

Additionally, by reducing the distance between sensor nodes, each sensor node having a wireless radio may be able to transmit signals at a reduced power level, which may extend the life of a power supply while the signals transmitted remain strong enough to reach an adjacent sensor node having a radio. The sensor nodes within the network may be synchronized such that each sensor node talks or listens at a particular time. Alternatively, one or more control radios may be generally active, while the remaining radios remain predominantly passive. The control radios may be hardwired directly to a power supply such that they are not confined by a limited power supply.

The mesh network may utilize the Zigbee protocol or other IEEE 802.15.4 Low-Rate Wireless Personal Area Network (WPAN) standards for wireless personal area networking. Zigbee is a published specification set of high level communication protocols designed for use with small, low power digital radios based upon the IEEE 802.15.4 standard. Other IEEE 802.15 standards also may be implemented, including those using Bluetooth or other WPAN or WLAN protocols or any other protocol.

The mesh network of sensor nodes may employ a dynamic routing algorithm. As a result, the mesh network may be self configuring and self mending. Each sensor node within the network may be able to identify neighboring sensor nodes. After receiving a message, a receiving sensor node may determine that it is not the sensor node closest to the destination and/or that it should not relay the message to another sensor node based upon the currently known configuration of operating sensor nodes. The receiving sensor node may wait a predetermined period and listen for another sensor node with a radio to relay the message. If after a predetermined time, the sensor node determines that the message has not been relayed as expected, the receiving sensor node may transmit or relay the message to a nearby sensor node.

By transmitting messages to only reach nearby or adjacent sensor nodes in the network, the messages within the network

may be transmitted at lower power. The low power transmission requires less energy from the on-board power supply of each sensor node. Additionally, the low power transmissions by the sensor nodes prevent one message from occupying the entire network and permits messages to be simultaneously transmitted from different sensor nodes and travel throughout the network of sensor nodes in parallel.

IX. Exemplary Sensor Nodes

FIG. **4** illustrates an exemplary sensor node or wireless radio **410**. The sensor node **410** includes a processor **412**, a wireless radio frequency transmitter and/or receiver **414**, a sensor **416**, an actuator **418**, a memory **420**, a clock **422**, a speaker **424**, a microphone **426**, and an on-board power supply **428**. The sensor node **410** may include additional, different, or fewer components.

For example, the sensor node **410** may be free of the sensor **416**, actuator **418**, memory **420**, clock **422**, speaker **424**, the microphone **426**, and/or power supply **428**. The sensor node **410** may include the processor **412** and the wireless transmitter and/or receiver **414**. For instance, the sensor node **410** may comprise the processor **412**, the wireless radio frequency transmitter and/or receiver **414**, the sensor **416**, the actuator **418**, and the power supply **428**. Or the sensor node **410** may comprise the processor **412**, the wireless radio frequency transmitter and/or receiver **414**, the sensor **416**, and the power supply **428**. The sensor node **410** may include other combinations employing additional, different, or fewer components.

In one embodiment, the sensor node **410** may be operable to automatically control building equipment and locate movable items within a building. In another embodiment, the sensor node **410** may be portable, such as in the case of being mounted upon a movable item, or affixed at a specific location or to an immovable item. The sensor node **410** may be a controller, actuator, sensor, locator or other device in a security, fire, environment control, HVAC, lighting, or other building automation system. The sensor node **410** may determine its present location, sense conditions within a building, report conditions within a building, generate a signal representative of a building condition, and/or respond to an interrogator. The sensor node **410** also or alternatively may actuate building control components. As a controller, the sensor node **410** may be free of the sensor **416** and/or the actuator **418**. In another embodiment, the sensor node **410** is a wireless device free of wired connections to other devices making the sensor node **410** self-contained and/or portable.

The sensor **416** may be a single sensor or include multiple sensors. The sensor **416** may be a temperature, pressure, humidity, fire, smoke, occupancy, air quality, flow, velocity, vibration, rotation, enthalpy, power, voltage, current, light, gas, CO_2, CO, N_2, O_2, chemical, radiation, fluid level, tank level, motion, Global Positioning System (GPS), infrared, or other sensor or combination thereof. The sensor **416** also may be a limit or proximity switch. Alternate sensors may be used.

The sensor **416** may be a motion sensor that detects when a portable sensor node **410** is moving. If it is sensed that the sensor node **410** is moving, the processor **412** may wake the sensor node **410** up from a sleep mode that draws less energy from the power supply **428**. Upon waking up, the sensor node **410** may transmit via the wireless transmitter **414** to the network a message indicating that sensor node **410** is moving.

The sensor **416** may be motion sensor that detects when there is movement within a predetermined distance. For example, the sensor **416** may be wall mounted to detect when an individual has entered a specific building area. If the building area was previously unoccupied, the sensor node **410** on which the sensor **416** is mounted may transmit a message to the network that the building area is no longer unoccupied. As a result, the network may direct that the environmental conditions be altered accordingly, such as increase the temperature during cold weather, decrease the temperature during hot weather, turn on one or additional lights, or adjust the room to the individual's personal preferences.

The sensor **416** may be a GPS unit capable of receiving GPS signals and determining the location of the sensor node **410**. The GPS unit may be able to determine the latitudinal and longitudinal coordinates, as well as the elevation, of the sensor node **410**. The location of the sensor node **410** determined by the GPS unit may be subsequently transmitted to the network via the wireless transmitter **414**.

The actuator **418** may be a single actuator or include multiple actuators. The actuator **418** may be a valve, relay, solenoid, speaker, bell, switch, motor, motor starter, turbine generator, motor generator, diesel generator, pneumatic device, damper, or pump actuating device or combinations thereof. For example, the actuator **412** may be a valve for controlling flow of fluid, gas, or steam in a pipe, or a damper controlling or redirecting air within an air duct. As another example, the actuator **412** may be a relay or other electrical control for opening and closing doors, releasing locks, actuating lights, or starting, stopping, and shifting motors and pumps. As a further example, the actuator **412** may be a solenoid that opens or closes valves, dampers, or doors, such as for altering the flow of fluid or air within piping or ducting. Alternate actuating devices also may be used.

The sensor node **410** may function as a controller or base station. The controller may be positioned at either a known or an unknown location. As a controller, the sensor node **410** interacts with other sensor nodes **410** for control or reporting functions.

The processor **412** is capable of processing data and/or controlling operation of the sensor node **410**. The processor **412** may be a general processor, digital signal processor, application-specific integrated circuit (ASIC), field programmable gate array, analog circuit, digital circuit, network of processors, programmable logic controller, or other processing device. The processor **412** may have an internal memory.

The sensor node **410** also may have a memory unit **420** external to the processor **412**. The memory unit **420** may store data and instructions for the operation and control of the sensor node **410**, such as discussed herein. Additional or alternate types of data also may be stored in the memory unit **420**.

A program may reside on the internal memory or the memory unit **420** and include one or more sequences of executable code or coded instructions that are executed by the processor **412**. The instructions may include monitoring a signal associated with radio frequency energy being harvested via a crystal radio receiver, the radio frequency energy being used to at least partially power a sensor node. The instructions also may include automatically reducing the amount of power being expended by the sensor node if the magnitude of the signal monitored approximately reaches or falls below a threshold level. The instructions may include directions to accomplish other functionality described herein.

The program may be loaded into the internal memory or memory unit **420** from a storage device. The processor **412** may execute one or more sequences of instructions of the program to process data. Data may be input to the data processor **412** with a data input device and/or received from a network. The program and other data may be stored on or read from machine-readable medium, including secondary storage devices such as hard disks, floppy disks, CD-ROMS, and

DVDs, or other forms of machine readable medium, either currently known or later developed.

The processor **412** is capable of directing the transmission or reception of data by the wireless transmitter or receiver **414**, the speaker **424** or the microphone **426**. For example, the processor **412** may direct the acoustic speaker **424** to transmit an ultrasound signal. The processor **412** may also direct the microphone **426** to receive an ultrasound signal and determine a distance from another device as a function of the received signal. Alternatively or additionally, the processor **412** may direct the wireless transmitter or receiver **414** to transmit data for determining the distance. Additionally or alternatively, the wireless transmitter **414** transmits a determined distance or distances as well as data regarding the processes and operation of the sensor **416** and/or the actuator **418**.

The wireless transmitter and receiver **414** or the speaker **424** may be alternate wireless transmitters capable of transmitting a signal for distance determination. Similarly, the wireless receiver **414** and microphone **426** may be alternative wireless receivers capable of transmitting a signal for distance determination.

FIG. **5** illustrates a floor layout for a network of sensor nodes having wireless radios **510** operating with one or more control radios **522** within a building **524**. The sensor nodes **510** may be dispersed throughout the building **524**. One or more of the sensor nodes **510** may be located in each room or other building area. Alternate dispersed arrangements of the sensor nodes **510** may be provided. While one control radio **522** is shown, a plurality of control radios **522** may be provided in other embodiments. Additional, different or fewer sensor nodes **510** and control radios **522** may be provided. While shown as a single floor of a building **524**, the network of sensor nodes **510** and control radios **522** may be distributed over multiple floors, a portion of the floor, a single room, a house, a structure, or any other building **524** or portion thereof.

The various sensor nodes **510** may be of the same configuration or a different configuration than each other. For example, some of the sensor nodes **510** may correspond to sensor arrangements, while other sensor nodes **510** may correspond to actuator arrangements. The same or different communication device, such as a wireless radio frequency transmitter and/or receiver, may be provided for each of the sensor nodes **510**. Alternatively, different communications mechanisms and/or protocols are provided for different groups of the sensor nodes **510**. The sensor nodes **510** may operate in an integrated manner for implementing one or multiple types of building automation control. Alternatively, different networks may be provided for different types of building automation, such as security, HVAC, heating, ventilation, and fire systems.

The sensor nodes powered by RF energy as discussed herein may be useful for a wide variety of applications, in addition to the building control functionality described above. For instance, the applications may include environmental and human body monitoring to military surveillance, building automation, and industrial monitoring. Each sensor node may facilitate one or more control, homeland security, reconnaissance, or other tasks. In one embodiment, the sensor node may be implanted internally within a patient, and sense and transmit data pertaining to a biological characteristic of the patient. The sensor nodes may be used for other applications.

X. Exemplary Method

FIG. **6** illustrates an exemplary RF energy harvesting method. The method may include harvesting RF energy **602**, storing the RF energy harvested **604**, monitoring a signal associated with the RF energy harvested **606**, and selecting an operating mode based upon the level of the signal monitored **608**. The method may include additional, fewer, or alternate actions.

The method may include harvesting RF energy **602**. RF energy may be harvested via a crystal radio receiver as discussed herein. Most of the energy harvested may be scavenged from AM radio signals. The RF energy may be harvested ubiquitously. The RF energy may be used to primarily or solely power a self-contained sensor node.

The method may include storing the RF energy harvested **604**. The RF energy or an associated signal may be measured from a super capacitor or other energy storage means. In one aspect, the RF energy harvestor may be used in unison with another type of energy harvester, such as a solar, biomechanical, thermoelectric, or vibration driven device.

The method may include monitoring a signal associated with the RF energy **606**. A voltage on the super capacitor may be monitored. Other signals related to the RF energy harvested also may be monitored.

The method may include selecting an operating mode based upon the level of the signal monitored **608**. The operating mode selected may relate to power constraints imposed upon the sensor node. If the level of the signal being monitored, such as a voltage signal or other signal, falls below a threshold level, the sensor node may be directed to operate in a low power mode, as discussed herein, to conserve energy until the amount of RF energy being harvested returns to a self-sustaining level with respect to the operating power requirements of the sensor node.

XI. Exemplary Method

FIG. **7** illustrates an exemplary interrupt handling method **700**. The method **700** may be implemented by software on the sensor node. As shown, the method **700** may use a state variable K having 2 states. When K=1, the power switch in FIG. **2***a* may connect pin A to pin Out, which in turn is connected to the pin labeled VCC. The voltage of pin VCC, which is connected to the sensor node, is equal to the output voltage of the DC/DC converter. Thus, in this case, the sensor node is powered through the energy harvester via the super capacitor and DC/DC converter. On the other hand, when K=2, the power switch connects pin B to pin Out and in this case, the sensor node is powered by the battery.

Ts in FIG. **7** is the wake-up period (which is configurable through software depending upon the application). The sensor node wakes up repetitively with time period of Ts. Each time when the sensor node wakes up, it may sense the environment, process the data, and transmit the date to the sink, if necessary. Then the sensor node may enter sleep mode, which costs almost no energy.

Tm in FIG. **7** is the maximum wake-up period allowed by the system. If Ts>Tm, the wake-up period is longer than the upper limit required by the system. Thus, the sensor node cannot provide spontaneous samples to satisfy the quality of service (QoS) requirements of the network.

Tp in FIG. **7** denotes a continuous block of time during which harvested energy is enough to support the continuous operation of the sensor node to execute for a time Te, which is the minimum time required by a sensor node to complete sensing tasks as required by an application. In other words, each time the sensor node wakes up, it may have to execute for at least a period of Te. And the sensor node may have to make a decision as to whether it powers itself from the energy harvester or through the battery for this amount of time. The

following brief analysis is presented to describe the novel sensor node operation that may be facilitated through the energy harvesting circuit.

Let the capacity of the super capacitor be represented by C. Then, the instantaneous energy E stored in the super capacitor is given by:

$$E=\frac{1}{2}CV_1^2.$$

If power rating of the sensor node is P, then the following equation holds

$$E+D'T_p=PT_e,$$

where E' is the derivative of E with respect to time Thus,

$$\begin{aligned} T_p &= \frac{PT_e - E}{E'} \\ &= \frac{PT_e - \frac{1}{2}CV_1^2}{CV_1V_1'}, \end{aligned}$$

where V_1' is the derivative of V**1** and V_1' can be estimated in practice by taking the difference of the V**1** values at consecutive time intervals.

The flow chart shown in FIG. **7** may be described as follows. After receiving the interrupt **702**, the sensor node sends a control signal to the power switch such that the sensor node is powered through the battery **704**, since the energy on the capacitor may be depleted soon. The voltages V**1** and V**2** are measured by the sensor node **706**. Tp is also computed based on V**1** **708**, as shown in the above equation.

V**2** is checked to see if it is low **710**. If it is, then all of the energy is going to be depleted soon. Recall that when the interrupt occurred, V**1** is already low. Thus, in this case, neither the energy harvester nor the battery can power the sensor node. The sensor node must take emergency actions, such as backup data **712** and notify its neighbors **714**. Finally, the sensor node sets Ts to be equal to Tp **716**, as well as sets K to 1 **718**. The rationale being that since the energy in the battery is going to be depleted soon, the node should wait in sleep mode **732** until enough energy has been harvested by the RF energy harvester.

When V**2** is not low, the decision is based on V**1**. If V**1** is low such that Tm≤Tp **720**, which means that the energy harvester cannot satisfy the sensor node energy requirements, then the sensor node once again has no choice but to use power from the battery source. Since there is enough battery energy (recall that V**2** is not low), the sensor node can be powered by the battery **728** through the current wake-up period Ts **726**. The neighbors also may be notified **730**. However, note that the energy from the harvester can still be used to charge the battery through the DC/DC converter. Thus, the system still benefits from the energy harvester even though the node is not powered directly by the energy harvester.

If V**1** is high enough to support the minimal energy requirements of the sensor during its wake-up period, i.e., Tm>Tp, then the sensor node can increase its wake-up period. In this case, Ts is set to be Tp **722** and the sensor node is powered by the energy harvester through the DC/DC converter **724**.

To summarize, the embodiments discussed herein facilitate the following complex sensor node behaviors: (1) switching the operating mode on the sensor node, such as standby or sleep modes; (2) sending a notification message to a sink via other sensor nodes; (3) saving important data to the flash memory on the sensor node, in case it is lost; (4) changing a sampling rate and transmission rate, and reducing the rates according to the voltages V**1** and V**2**, (5) adjusting an energy weighting factor in the network routing metric (if the voltages of the sensor node are not high enough, the use of this node as a router to route other wireless message packets is minimized or avoided), and (6) switching the input power source from the RF energy harvester to a battery.

The methods and systems discussed herein may use one, two, or more antennas. If two antennas are used, one antenna may be dedicated to the sensor node or communications unit, and the other antenna may be dedicated to the RF energy harvesting system, such as shown in the Figures. As both the sensor node and the energy harvester each may have their own dedicated antenna, switching a single antenna between the sensor node and the energy harvester may be alleviated.

The operational mode of the harvesting system and/or the sensor node may be switched by the sensor node itself. Additionally or alternatively, the RF energy harvester discussed herein may trigger the sensor node and provide voltage measurements. For energy conservation purposes, the harvesting system discussed herein may not switch operation modes based on. RF signals.

Therefore, the RF energy harvesters described herein may be directly connected to sensor nodes, purely passive and not needing to pass FCC requirements, and/or based on a band-pass or all pass filter for ubiquitous and wide spectrum RF energy harvesting. The systems discussed herein may facilitate advanced sensor behaviors. The system modes (active, standby and sleep) may be triggered by an interrupt signal and controlled by the sensor node, as compared to altering the operating mode according to received RF signals. The designs of the energy harvesting antenna may be different from known antennas. A super capacitor and/or a DC/DC converter may be employed by the energy harvester. Based on efficiency, size, and costs, the DC/DC converter may be provide benefits as compared to transformers. A buffered energy storage system may be constructed by the super capacitor and one or more rechargeable batteries. The harvesting system may improve battery life time.

While the invention has been described above by reference to various embodiments, it should be understood that many changes and modifications can be made without departing from the scope of the invention. The description and illustrations are by way of example only. Many more embodiments and implementations are possible within the scope of this invention and will be apparent to those of ordinary skill in the art. The various embodiments are not limited to the described environments, and have a wide variety of applications including integrated building control systems, environmental control, security detection, communications, industrial control, power distribution, and hazard reporting.

It is intended in the appended claims to cover all such changes and modifications which fall within the true spirit and scope of the invention. Therefore, the invention is not limited to the specific details, representative embodiments, and illustrated examples in this description. Accordingly, the invention is not to be restricted except in light as necessitated by the accompanying claims and their equivalents.

We claim:

1. A method of passive radio frequency energy harvesting, the method comprising:

harvesting radio frequency energy using a crystal receiver;

powering a processing unit only from the radio frequency energy being harvested when the radio frequency energy being harvested, or an indication of the amount of radio frequency energy being harvested, is above a threshold level;

generating an interrupt signal to switch the processing unit to a low power mode when the radio frequency energy

being harvested, or the indication of the amount of radio frequency energy being harvested, falls below the threshold level; and

alternatively powering the processing unit only from an on-board power source when the radio frequency energy being harvested, or the indication of the amount of radio frequency energy being harvested, is below the threshold level.

2. The method of claim **1**, wherein the crystal receiver employs an antenna that is fabricated from printed circuit board.

3. The method of claim **1**, wherein the radio frequency energy harvested is increased by a super capacitor before being used to power the processing unit.

4. The method of claim **1**, wherein the on-board power source is a battery or a capacitor.

5. The method of claim **1**, wherein the crystal receiver employs a ferrite coil antenna.

6. The method of claim **1**, wherein the radio frequency energy is harvested ubiquitously.

7. A method of passive radio frequency energy harvesting, the method comprising:

harvesting radio frequency energy using a crystal receiver;

powering a processing unit only from the radio frequency energy being harvested when the radio frequency energy being harvested, or an indication of the amount of radio frequency energy being harvested, is above a threshold level, wherein the processor unit is a self-contained sensor node operable to direct the operation of building equipment to control an environment of a building;

generating an interrupt signal to switch the processing unit to a low power mode when the radio frequency energy being harvested, or the indication of the amount of radio frequency energy being harvested, falls below the threshold level; and

alternatively powering the processing unit only from an on-board power source when the radio frequency energy being harvested, or the indication of the amount of radio frequency energy being harvested, is below the threshold level.

8. The method of claim **7**, wherein the processor unit has power requirements of greater than 10 milliwatts.

* * * * *

US008621245B2

(12) **United States Patent**
Shearer et al.

(10) **Patent No.: US 8,621,245 B2**
(45) **Date of Patent: Dec. 31, 2013**

(54) **POWERING DEVICES USING RF ENERGY HARVESTING**

(75) Inventors: **John G. Shearer**, Ligonier, PA (US); **Charles E. Greene**, Pittsburgh, PA (US); **Daniel W. Harrist**, Carnegie, PA (US)

(73) Assignee: **Powercast Corporation**, Ligonier, PA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 420 days.

(21) Appl. No.: **12/499,618**

(22) Filed: **Jul. 8, 2009**

(65) **Prior Publication Data**

US 2010/0090656 A1 Apr. 15, 2010

Related U.S. Application Data

(62) Division of application No. 11/447,412, filed on Jun. 6, 2006, now abandoned.

(60) Provisional application No. 60/688,587, filed on Jun. 8, 2005.

(51) **Int. Cl.**
G06F 1/26 (2006.01)

(52) **U.S. Cl.**
USPC **713/300**; 340/539.26; 340/693.1; 455/343.1

(58) **Field of Classification Search**
USPC 340/539.26, 693.1; 455/343.1; 713/300
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,297,618	B2	10/2001	Emori et al.	
6,447,448	B1	9/2002	Ishikawa et al.	
6,615,074	B2	9/2003	Mickle et al.	
6,784,358	B2	8/2004	Kukulka	
6,835,501	B2	12/2004	Morishita et al.	
6,836,095	B2	12/2004	Fogg	
6,858,970	B2	2/2005	Malkin et al.	
6,894,467	B2	5/2005	Pons et al.	
7,256,695	B2	8/2007	Hamel et al.	
7,268,517	B2	9/2007	Rahmel et al.	
7,561,866	B2 *	7/2009	Oliver et al.	455/343.3
7,956,572	B2 *	6/2011	Zane et al.	320/108
2002/0072784	A1	6/2002	Sheppard, Jr. et al.	
2004/0150529	A1	8/2004	Benoit et al.	
2004/0259604	A1	12/2004	Mickle et al.	

OTHER PUBLICATIONS

Non-Final Office Action for U.S. Appl. No. 11/447,412, mailed Jun. 3, 2008, 6 pages.
Final Office Action for U.S. Appl. No. 11/447,412, mailed Jan. 8, 2009, 10 pages.
International Search Report and Written Opinion for International Application No. PCT/US06/21940, mailed Sep. 25, 2007, 9 pages.
Examination Report for Australian Application No. 2006255054, mailed Nov. 27, 2009, 2 pages.

(Continued)

Primary Examiner — Albert Wang

(57) **ABSTRACT**

Disclosed is an apparatus for an application including a core device for the application. The apparatus includes a power (preferably RF energy) harvester connected to the core device to power the core device. Also disclosed is a method for an application. The method includes the steps of converting RF energy into usable energy. There is the step of powering the core device with the usable energy.

20 Claims, 90 Drawing Sheets

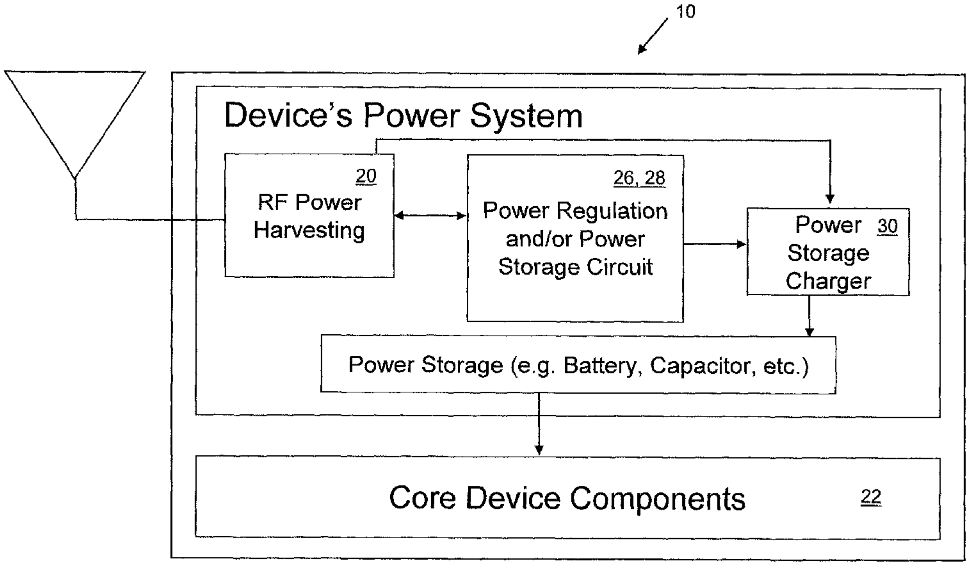
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

(56) **References Cited**

OTHER PUBLICATIONS

Office Action for Chinese Application No. 200680020092.3, received on Sep. 7, 2009, 5 pages.
Examination Report for Singapore Application No. 200718208-2, mailed Sep. 28, 2009, 6 pages.
Office Action for Mexican Patent Application No. MX/a/2007/015229, received on Mar. 4, 2010, 2 pages.
Examination Report for Australian Application No. 2006255054, mailed Jul. 7, 2011, 3 pages.
European Search Report for European Patent Application No. 06772310.6, mailed Jul. 27, 2011; 7 pages.

* cited by examiner

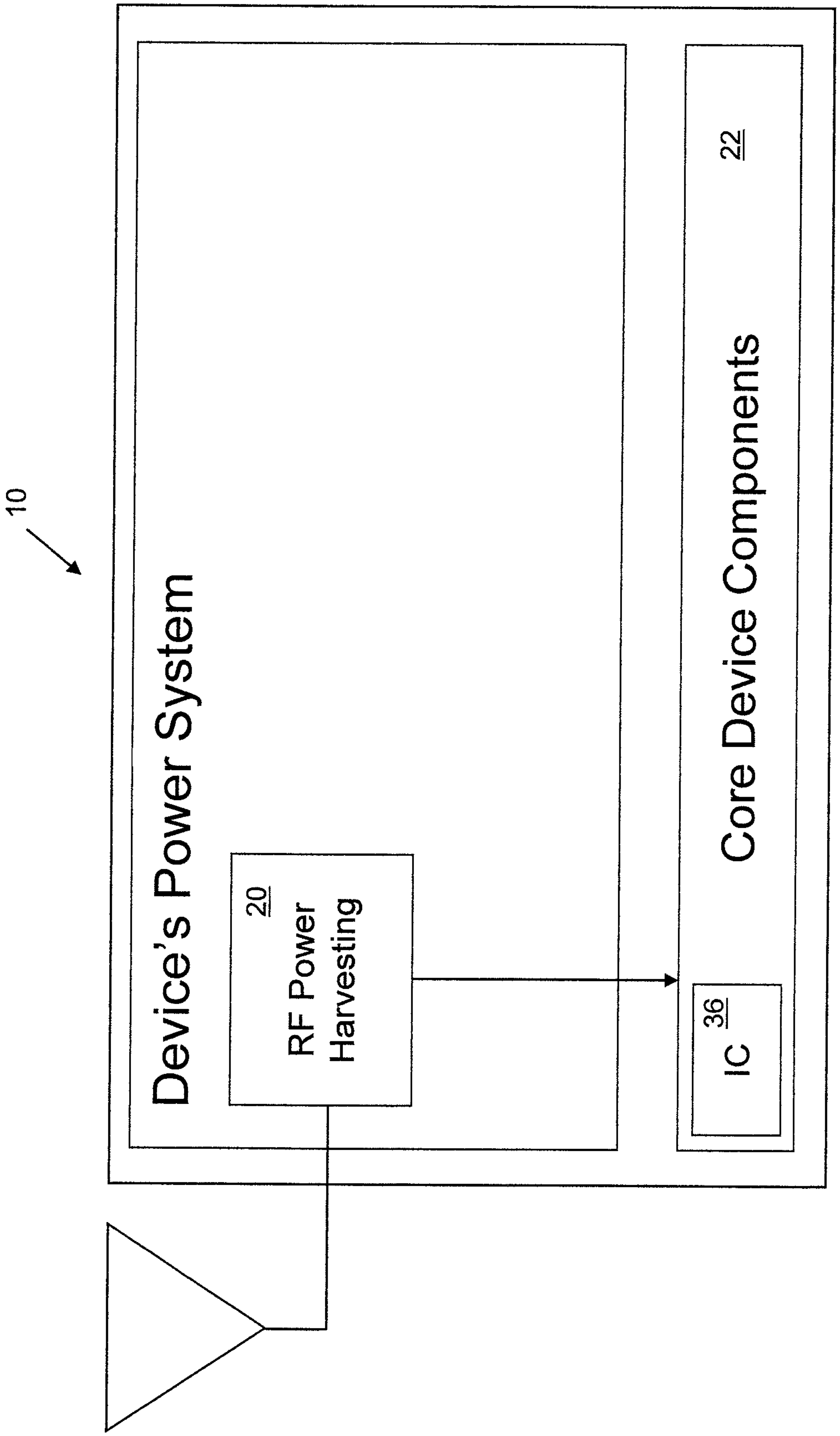
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
IC
36

Fig. 1

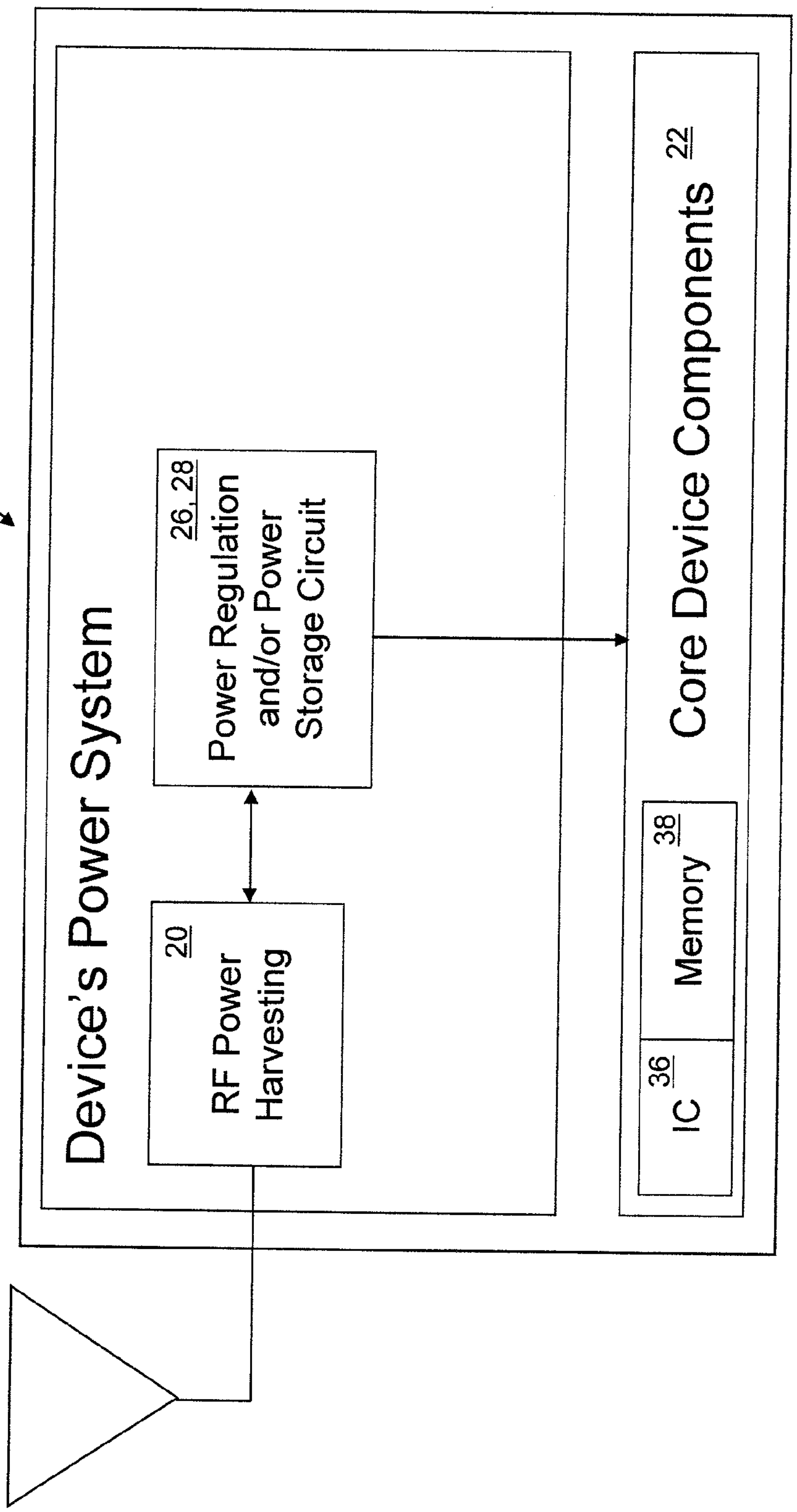
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Core Device Components
22
IC
36
Memory
38

Fig. 2

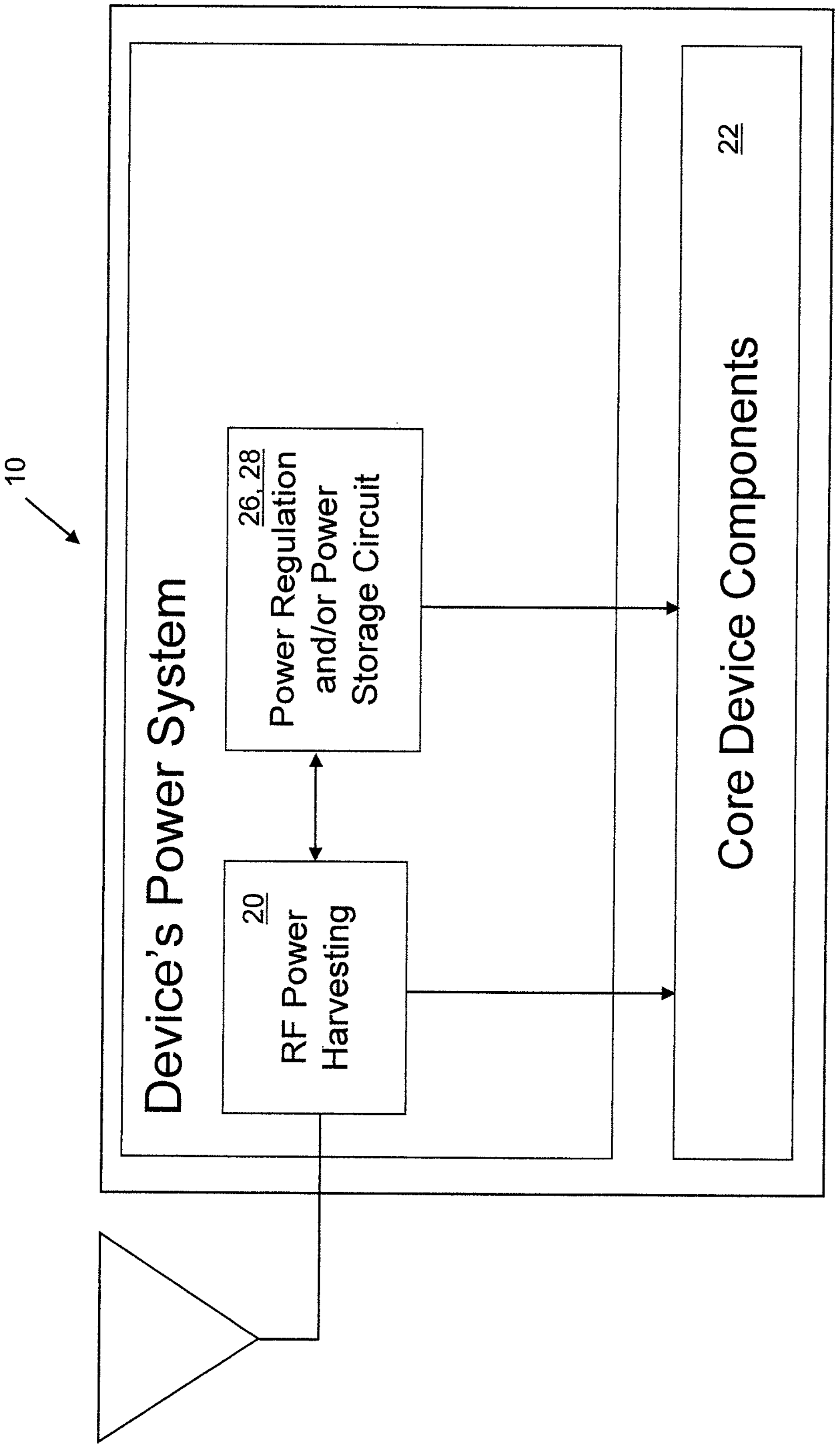
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Core Device Components
22

Fig. 3

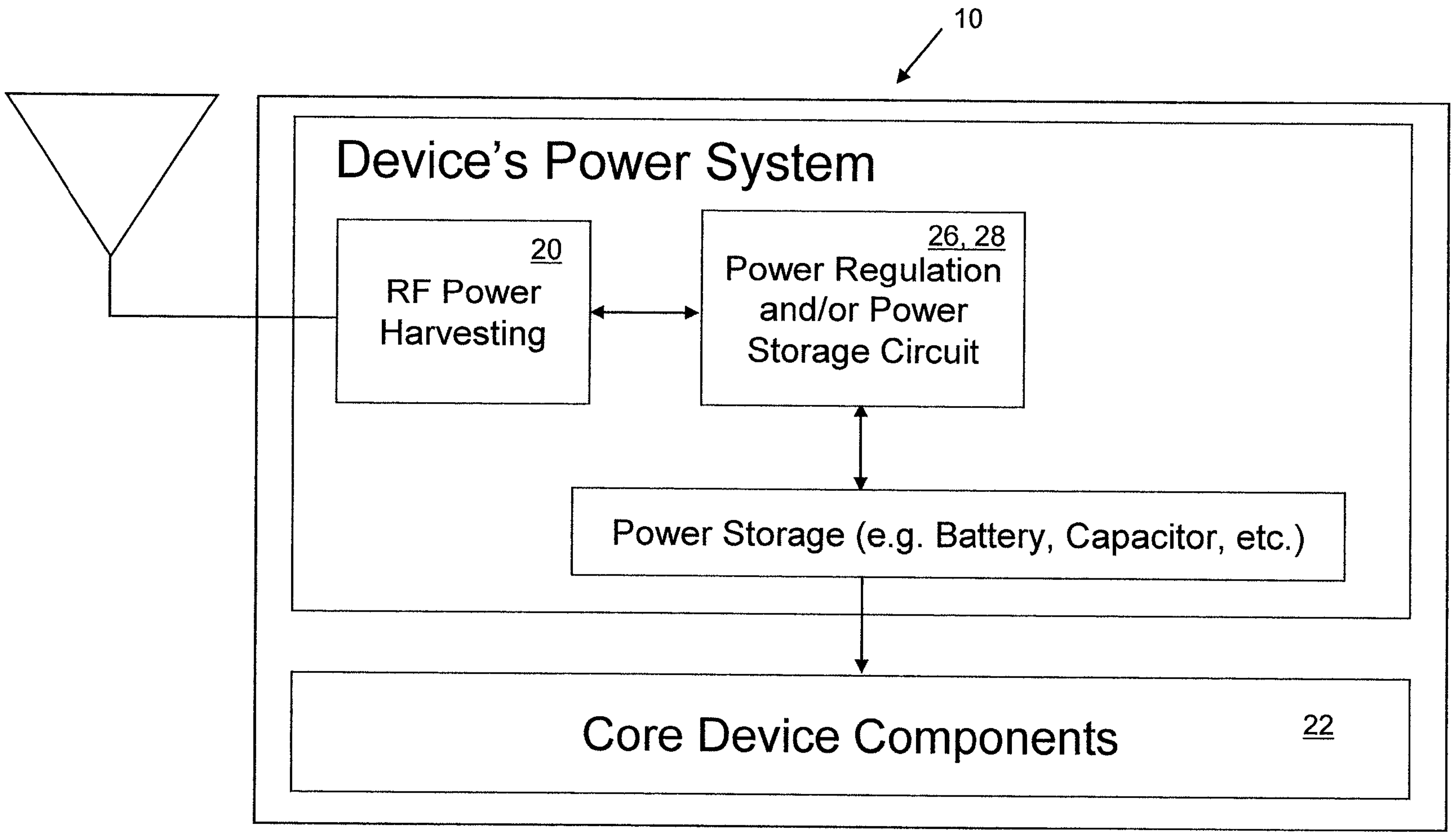
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 4

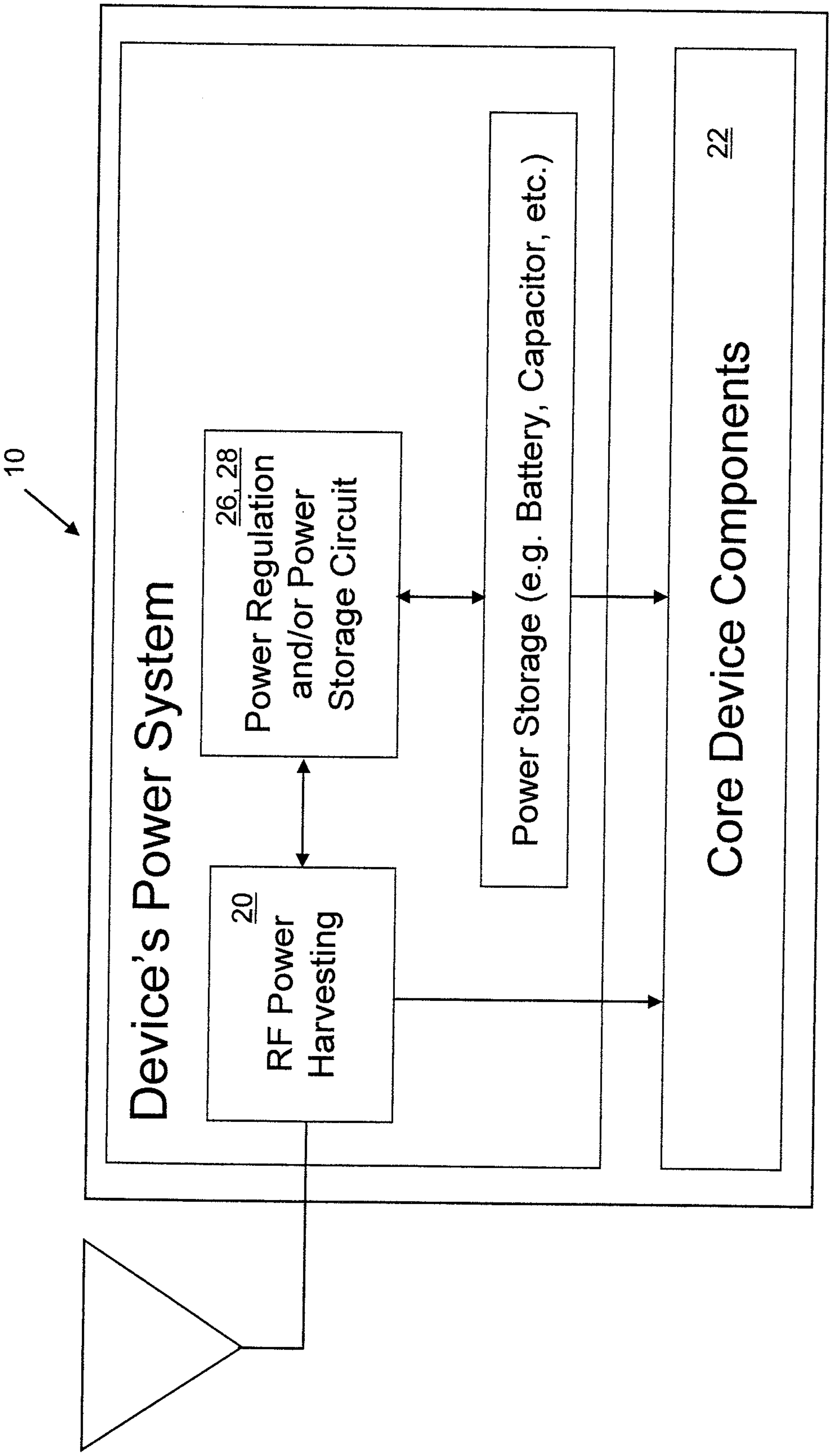
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 5

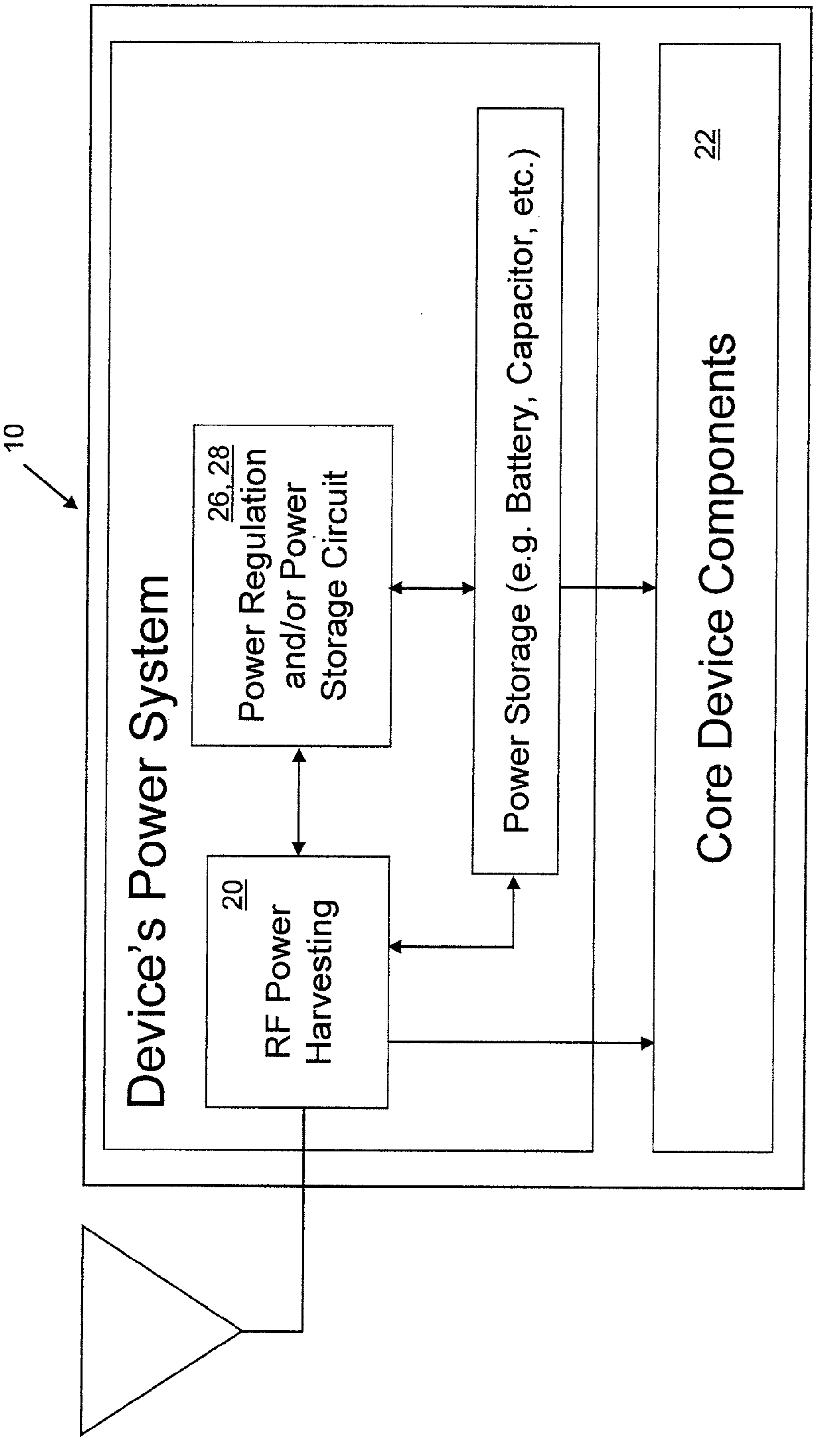
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 6

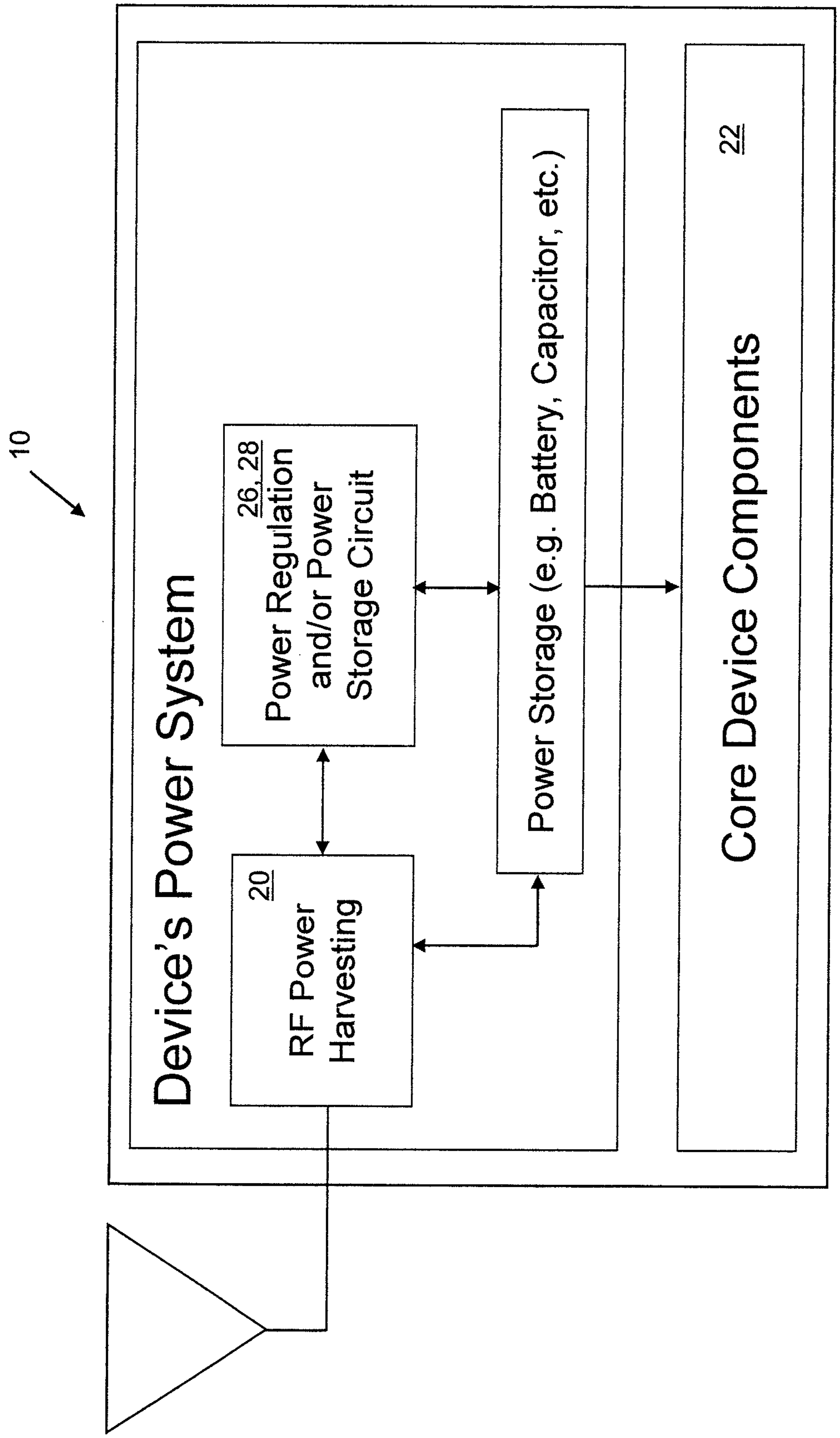
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 7

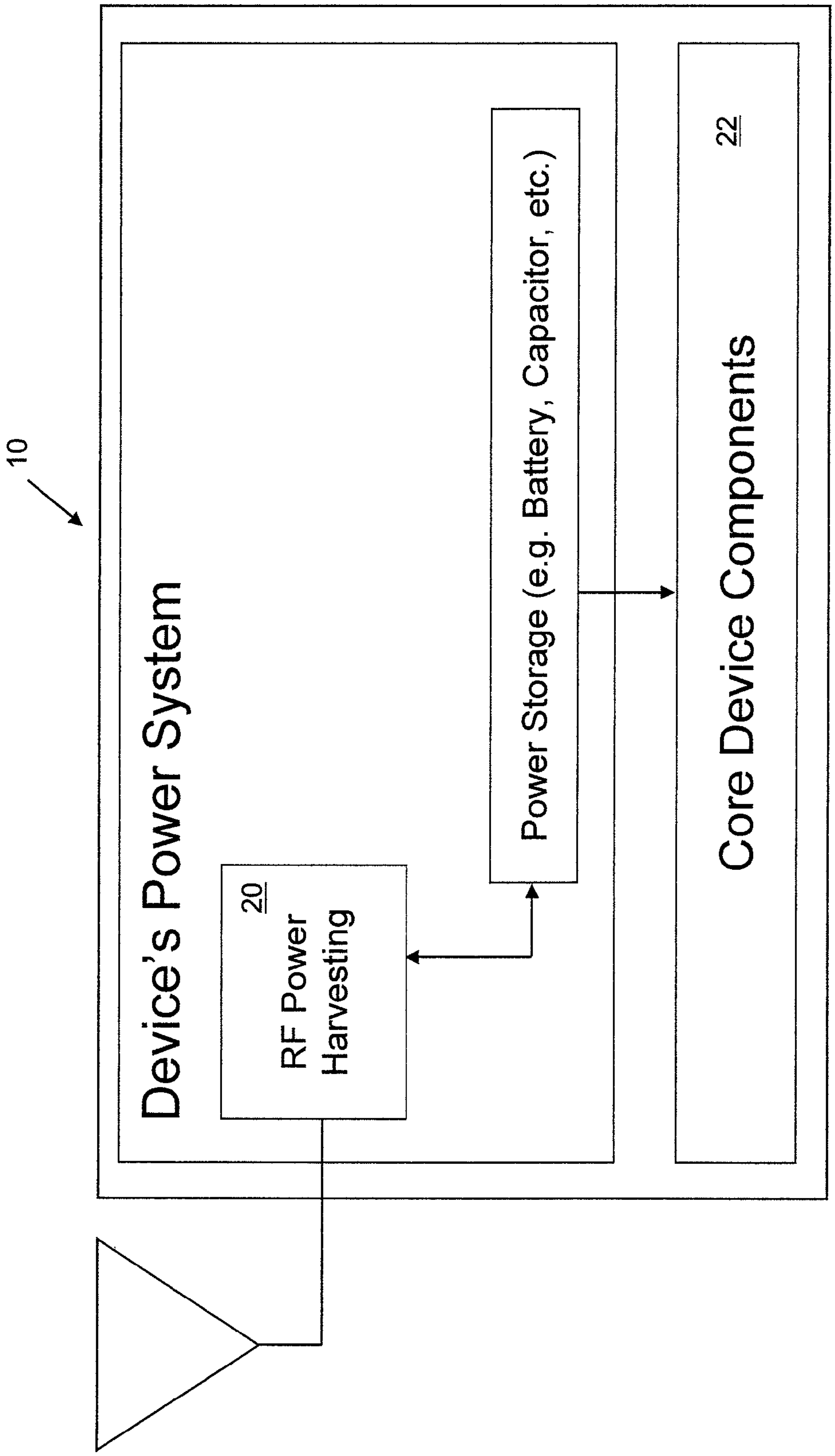
10
Device's Power System
RF Power Harvesting
20
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 8

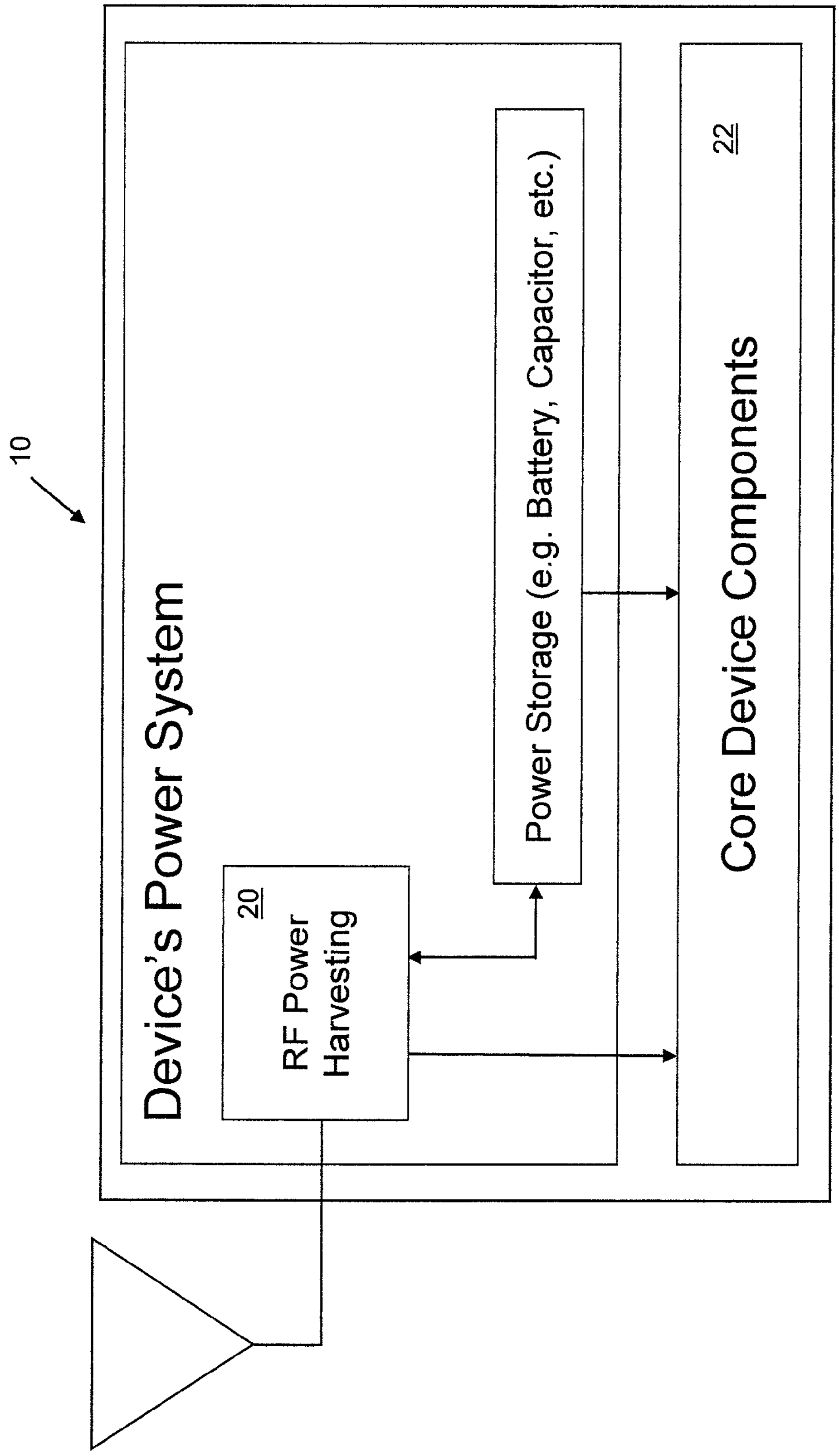
10
Device's Power System
20
RF Power Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 9

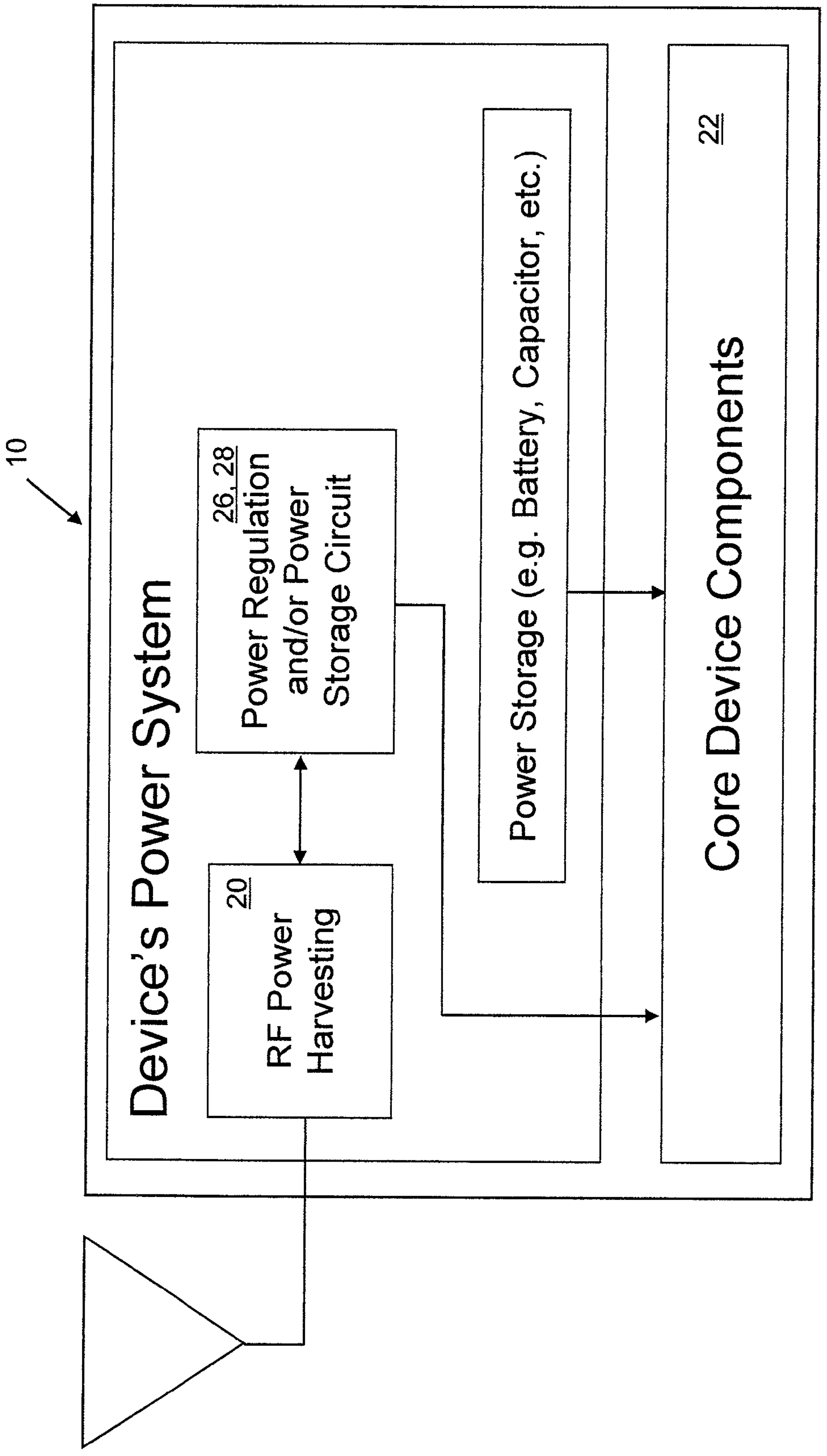
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 10

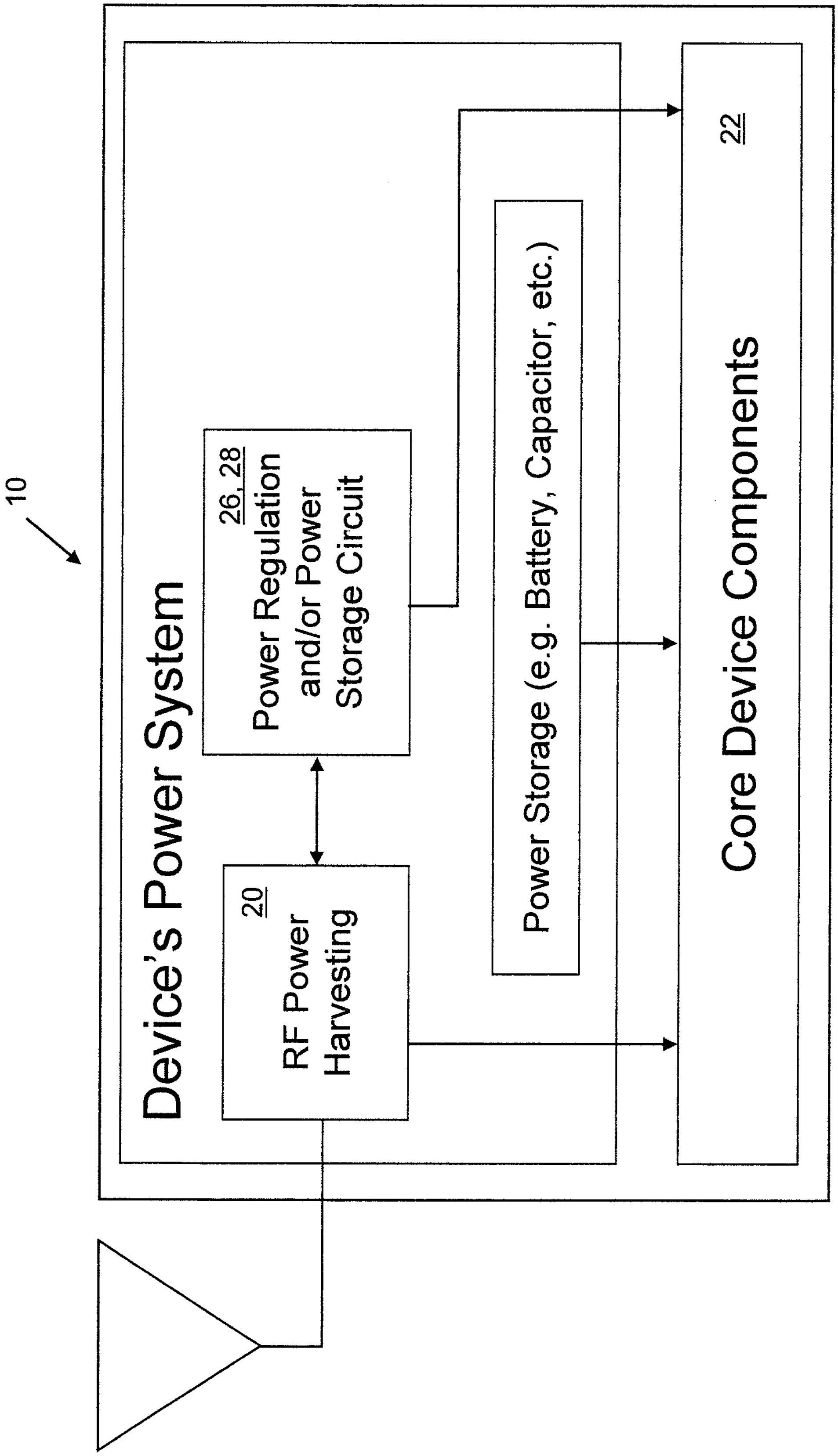
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 11

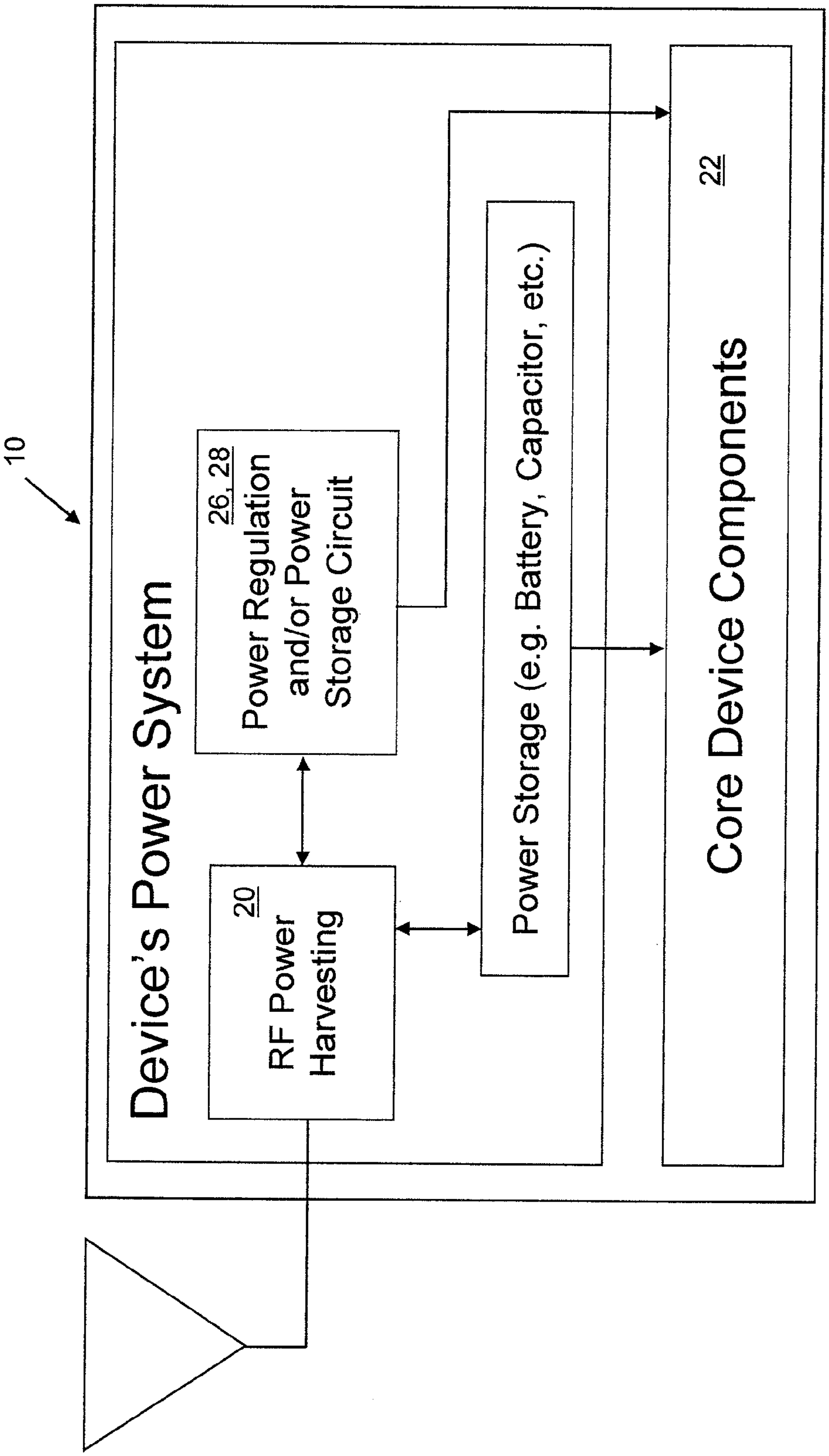
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 12

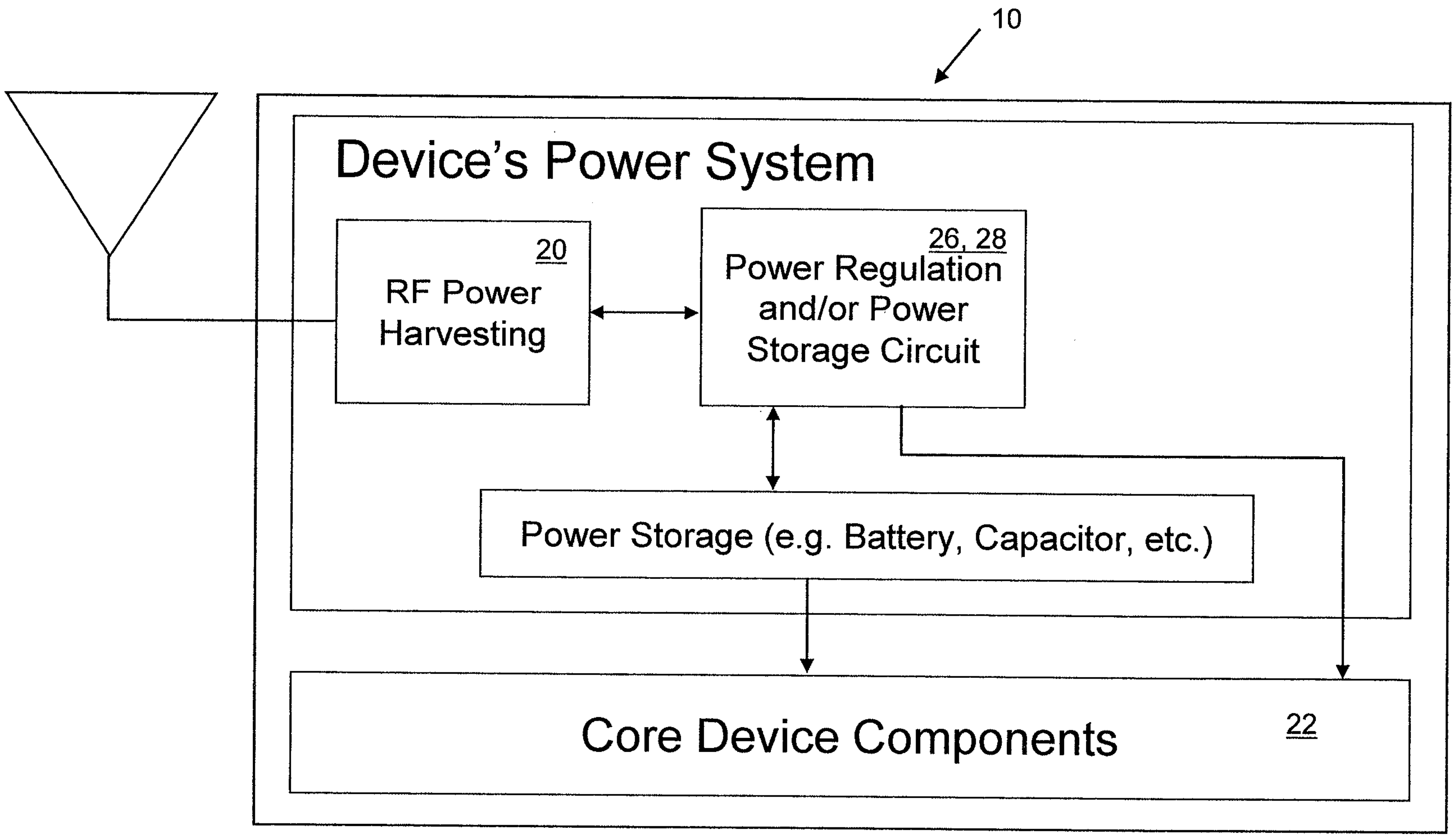
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 13

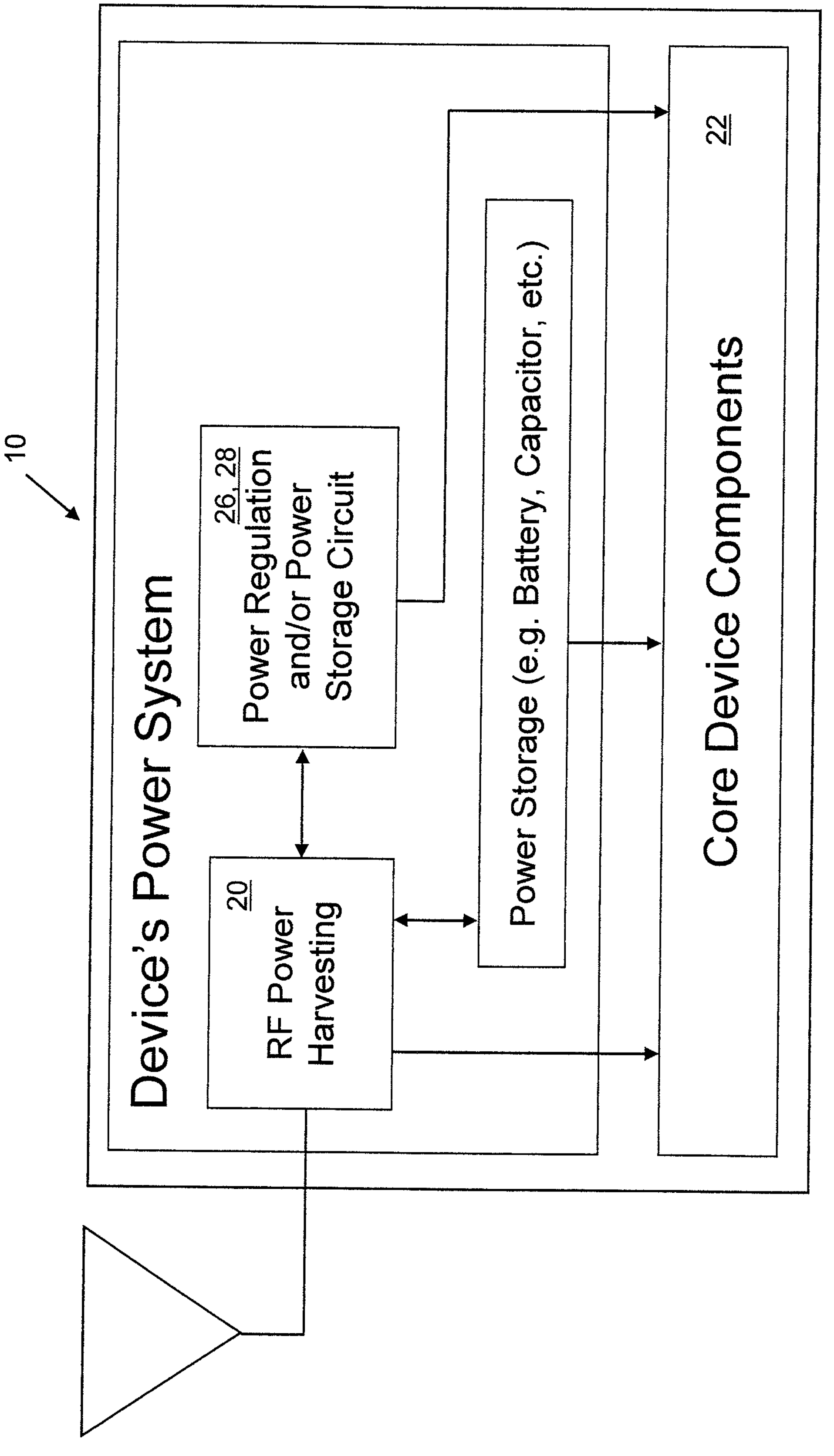
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 14

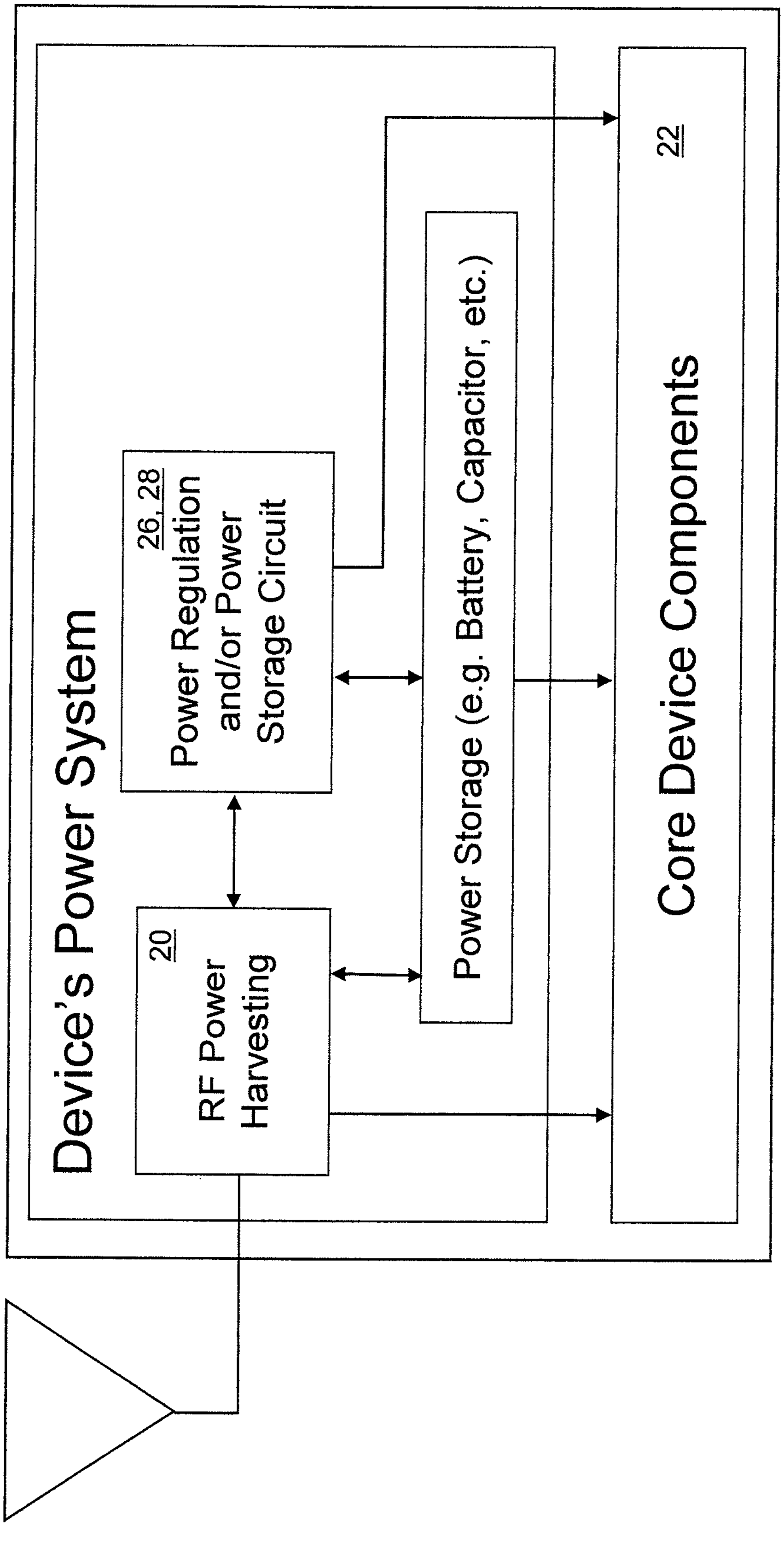
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 15

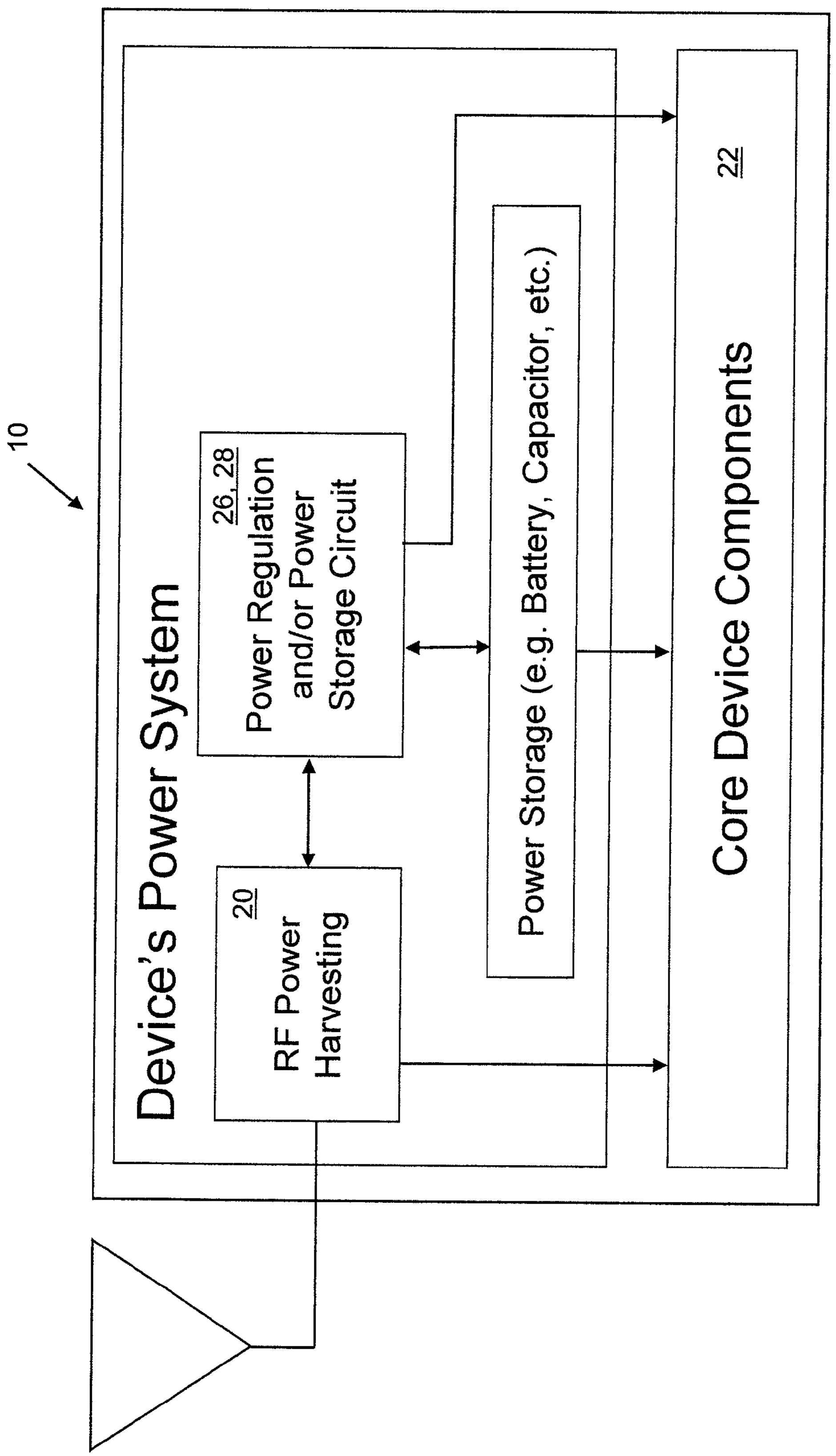
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 16

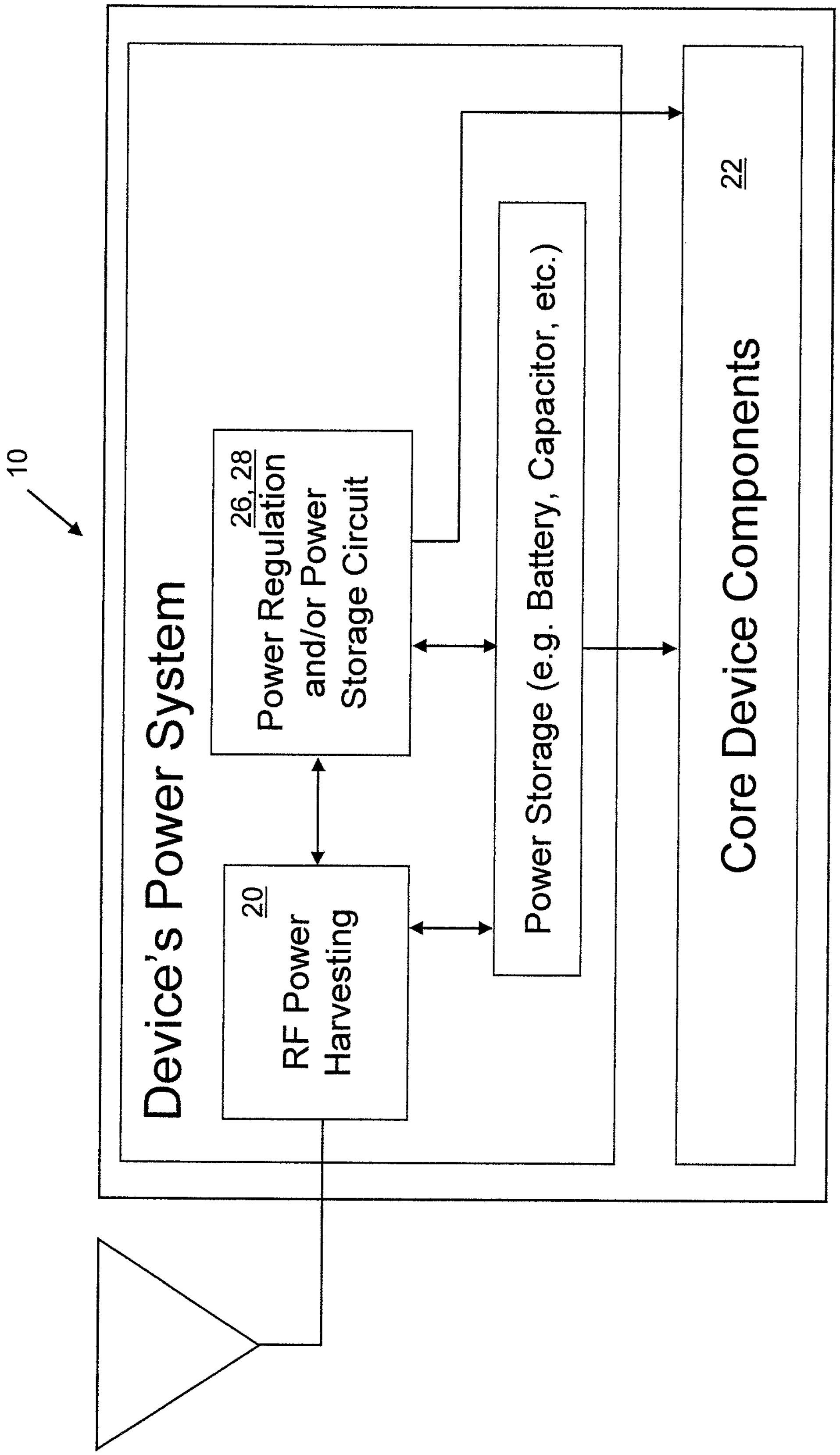
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 17

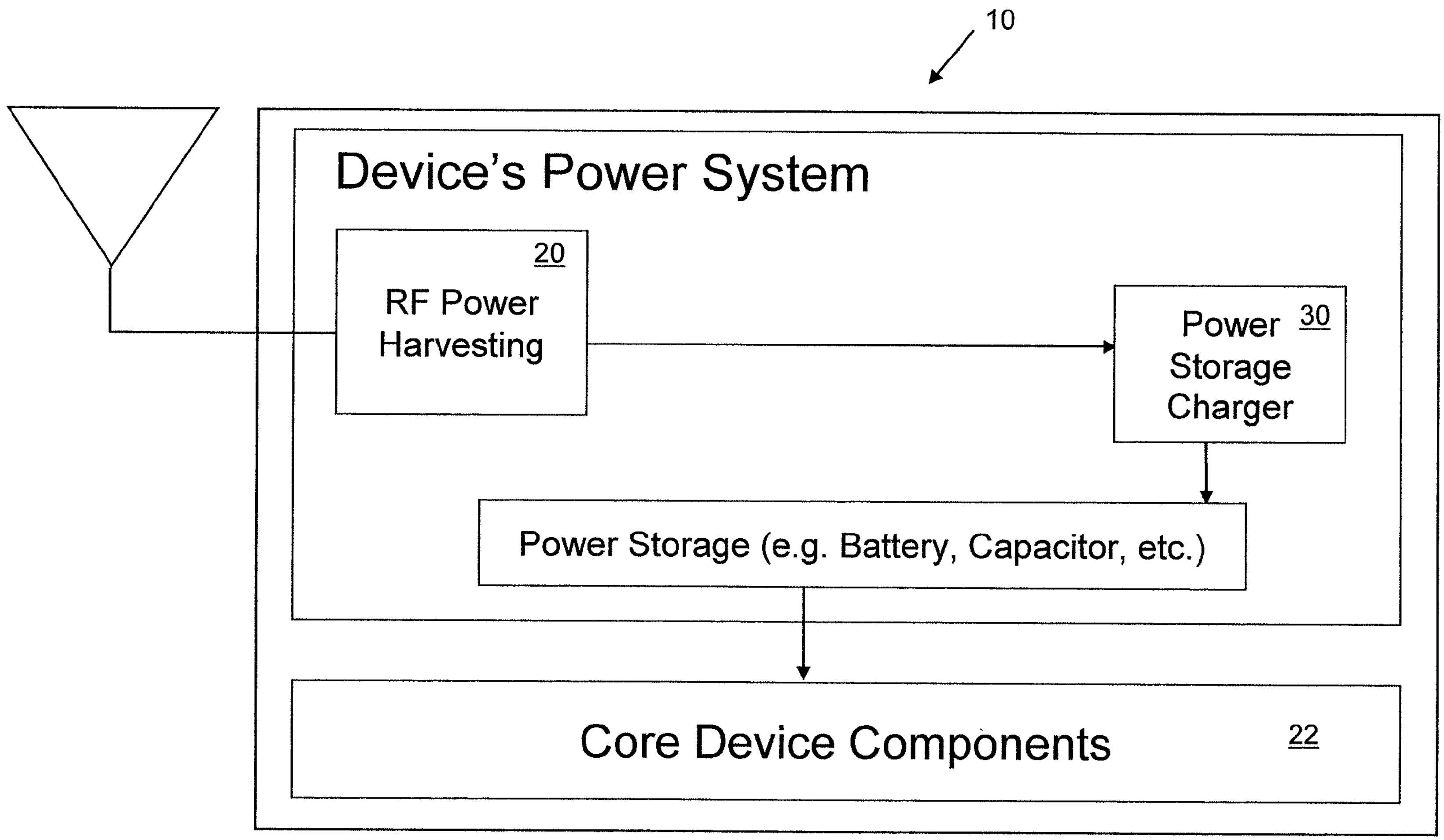
10
Device's Power System
20
RF Power Harvesting
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 18

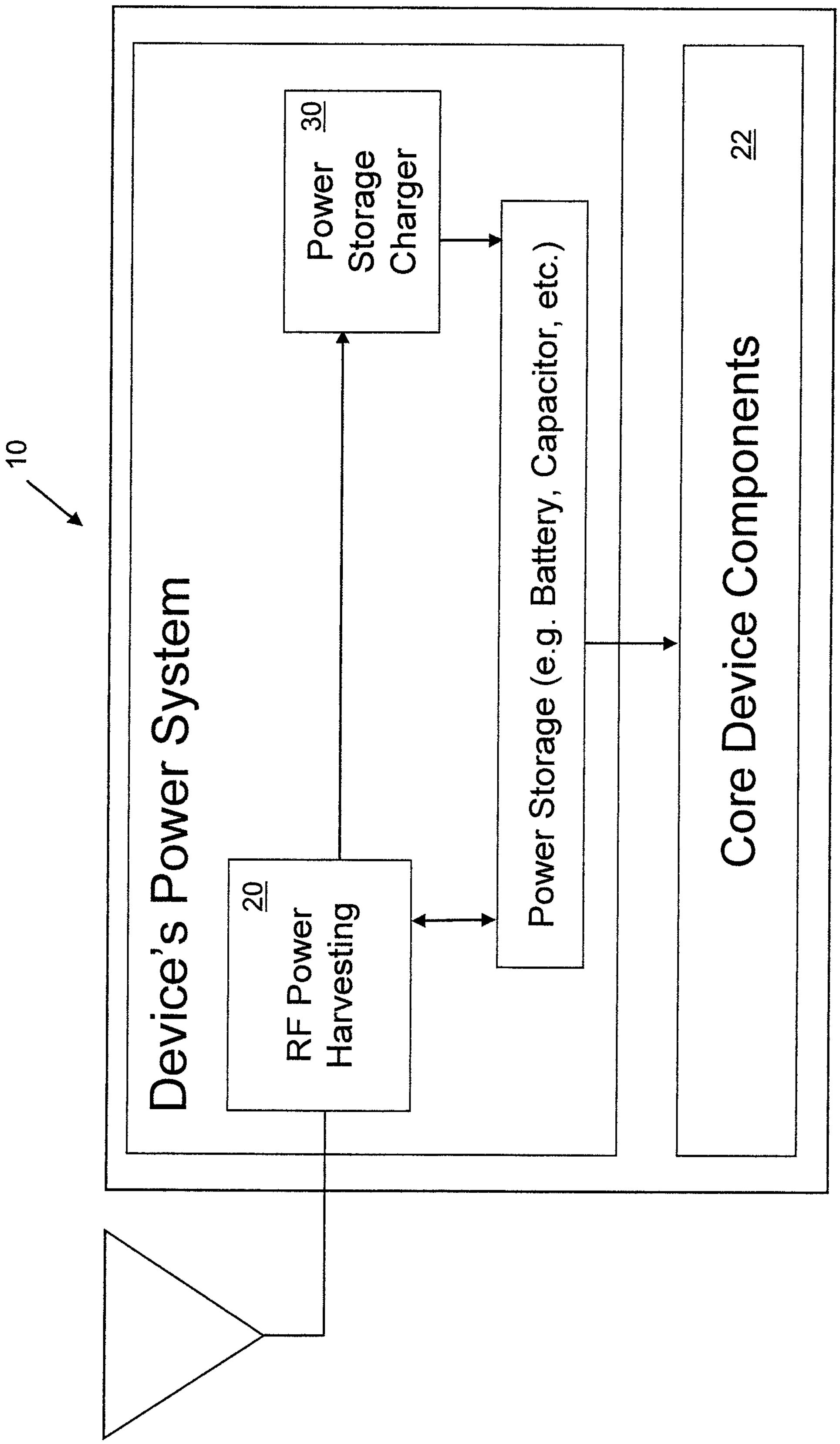
10
Device's Power System
RF Power Harvesting 20
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 19

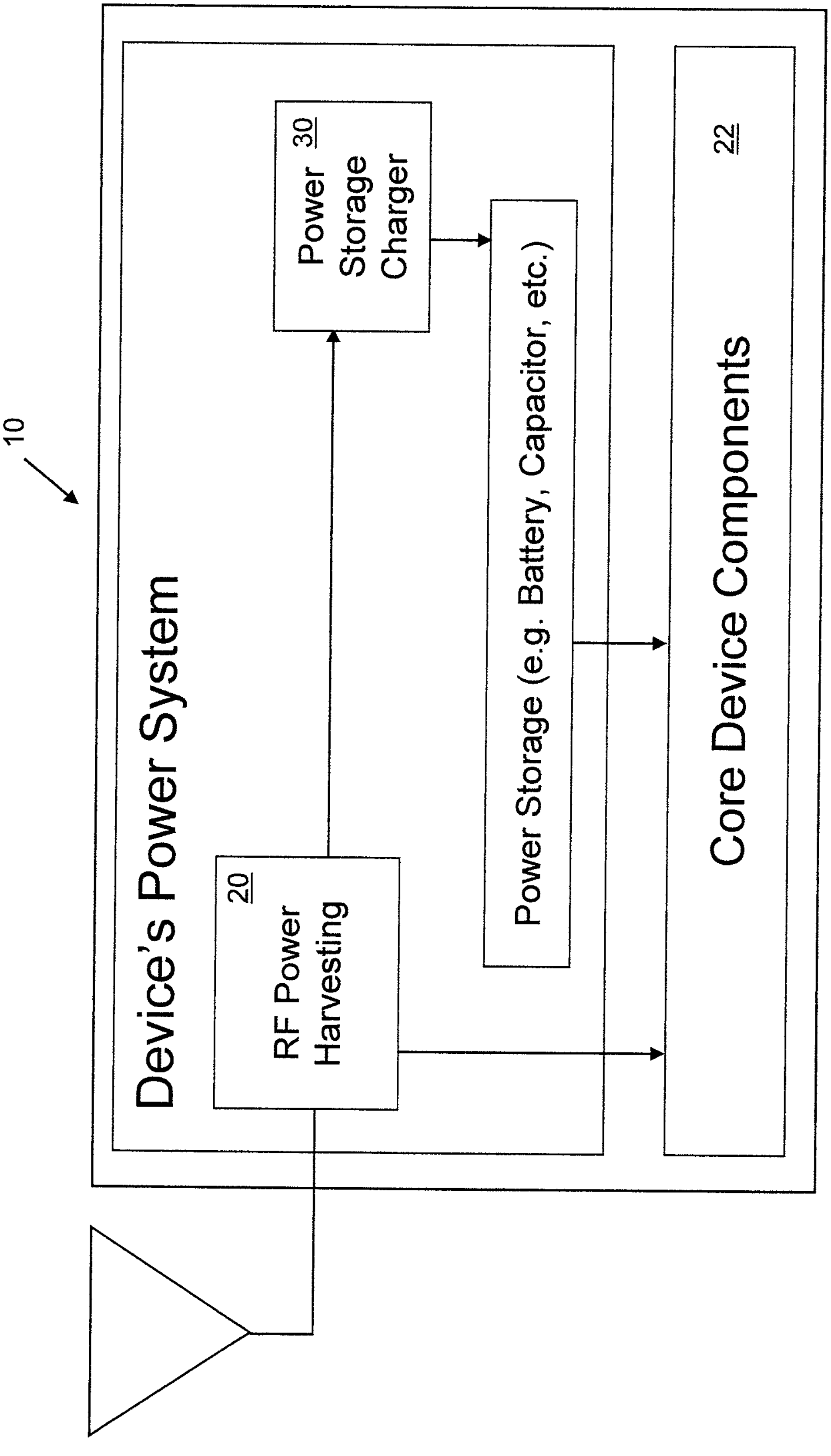
10
Device's Power System
20
RF Power Harvesting
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 20

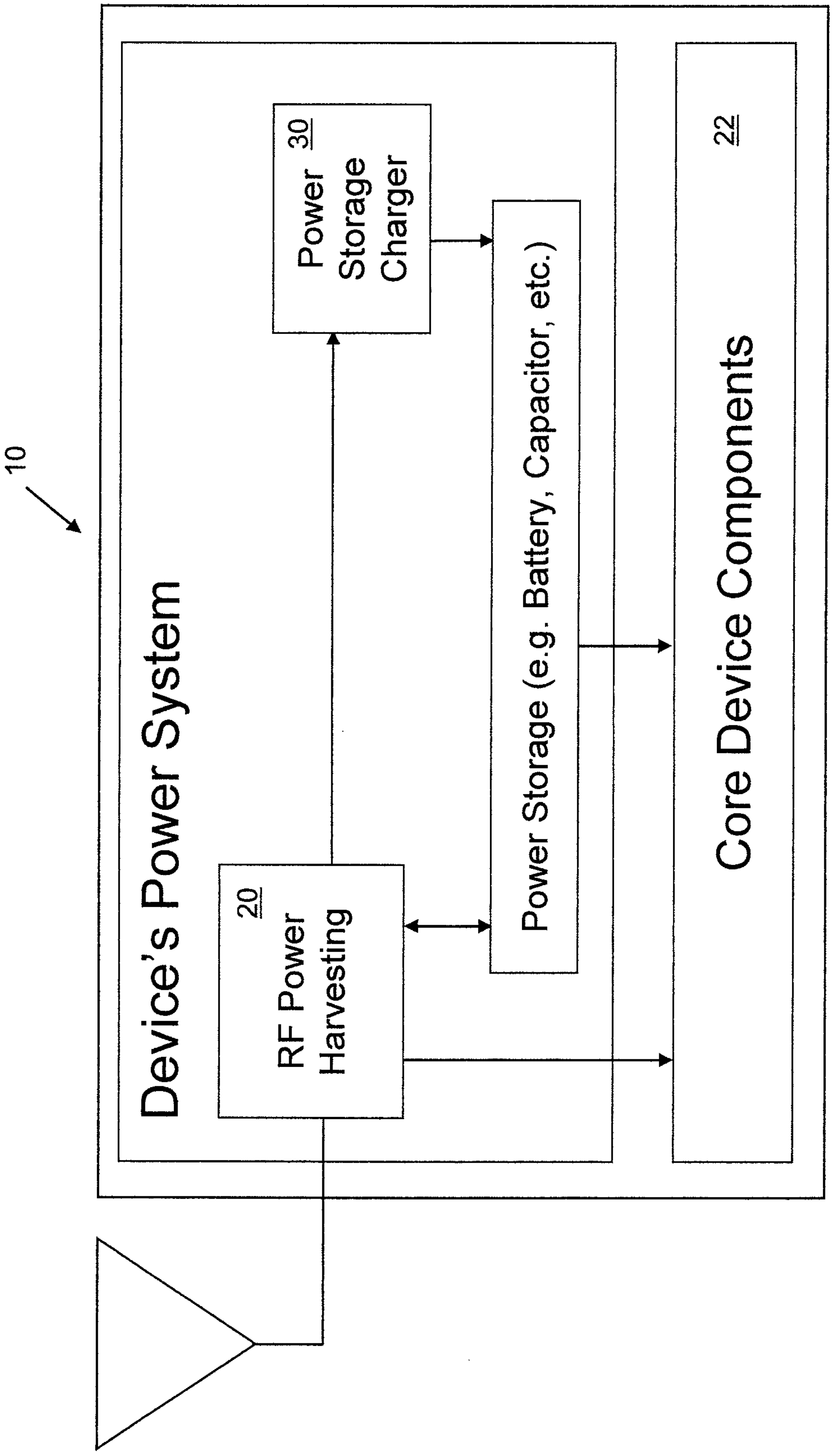
10
Device's Power System
20
RF Power Harvesting
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 21

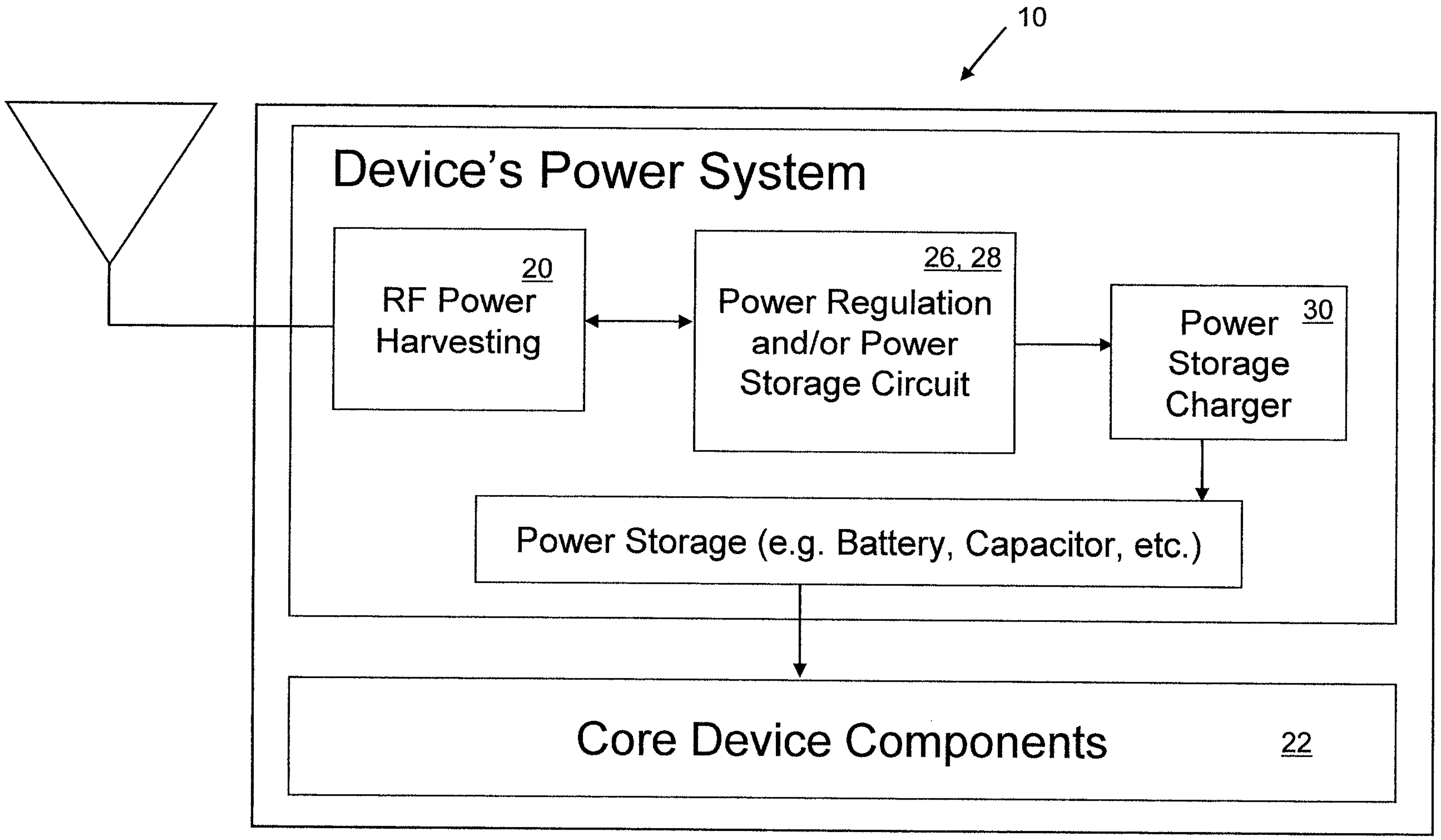
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 22

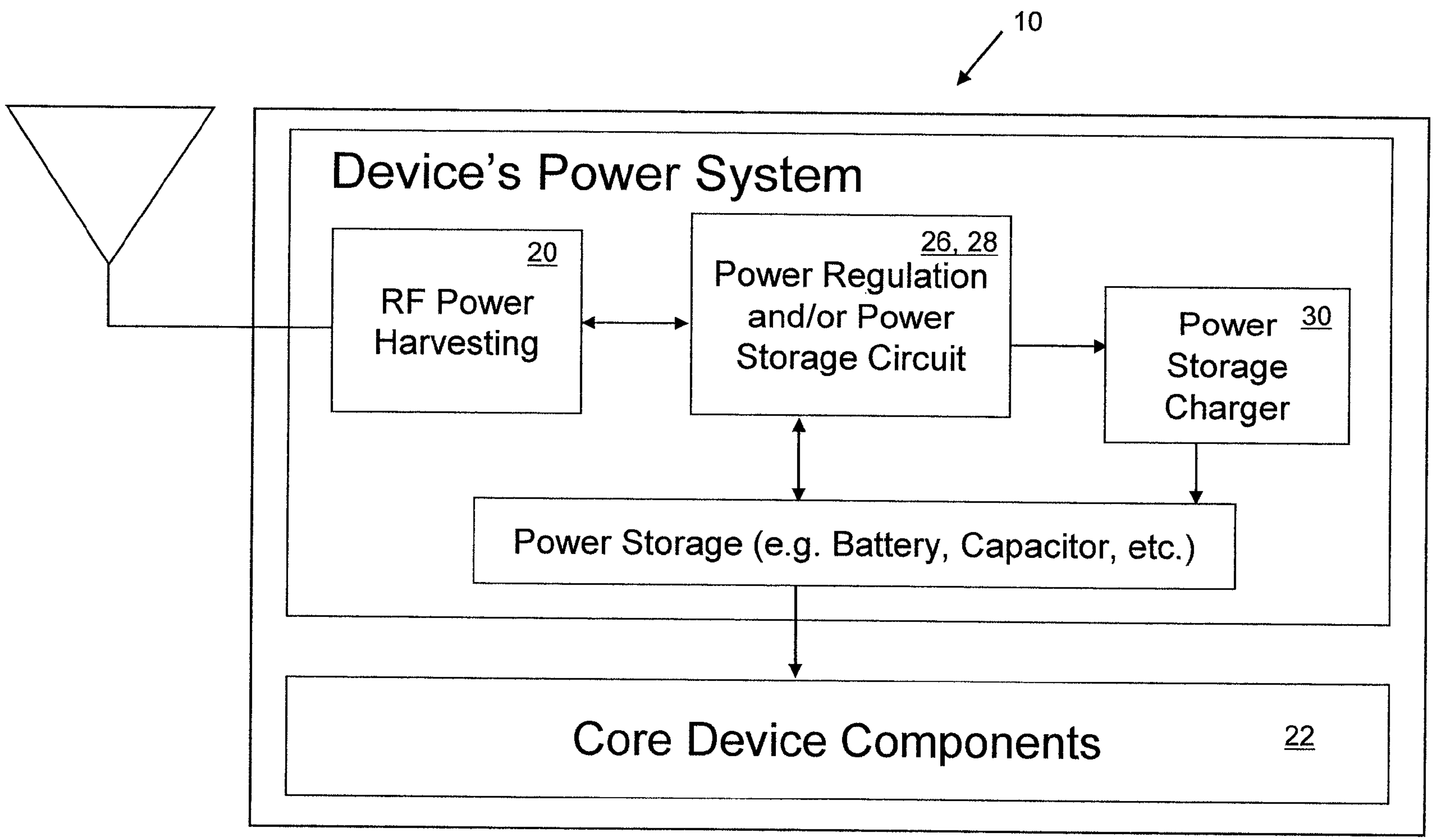
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 23

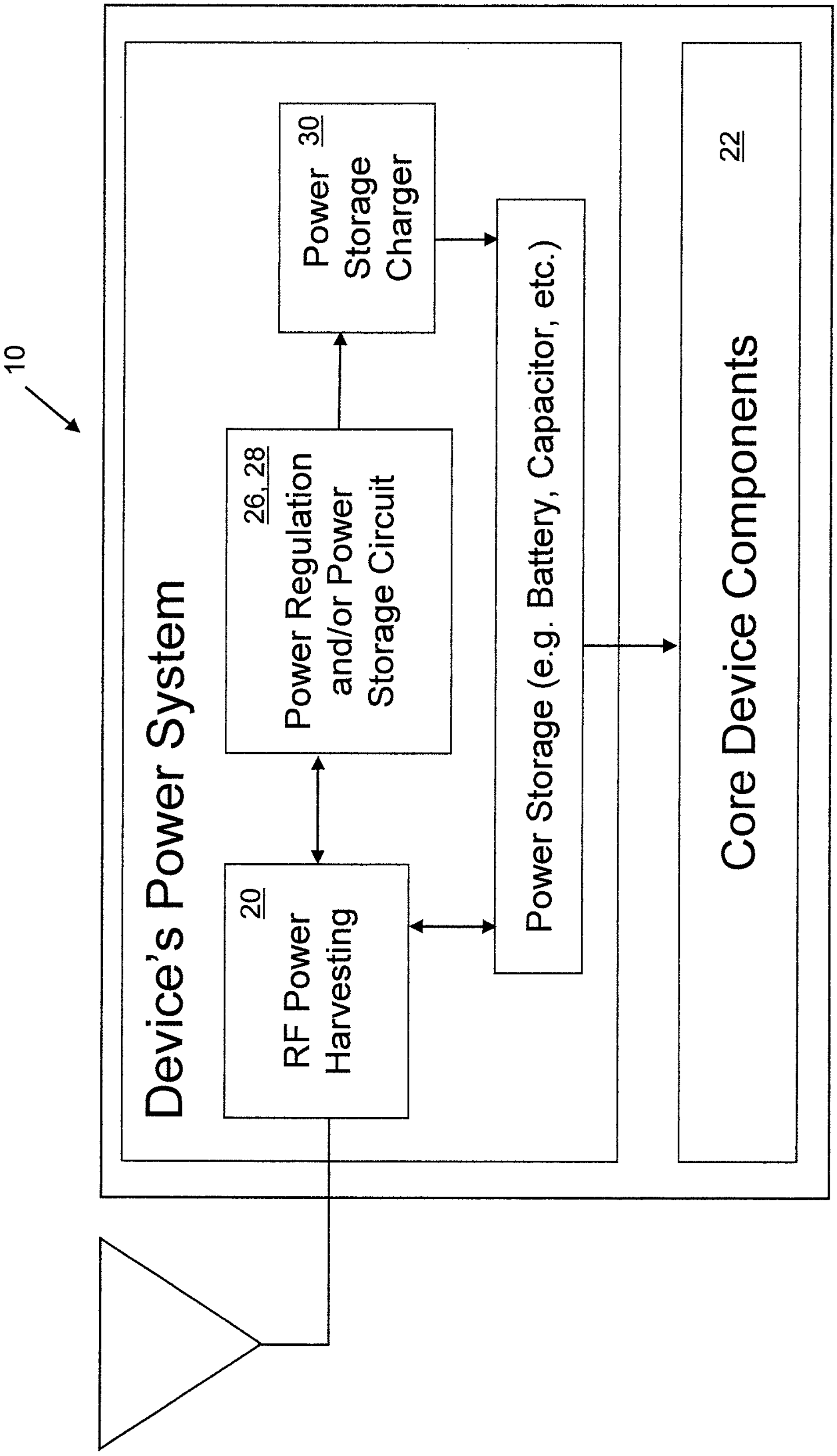
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 24

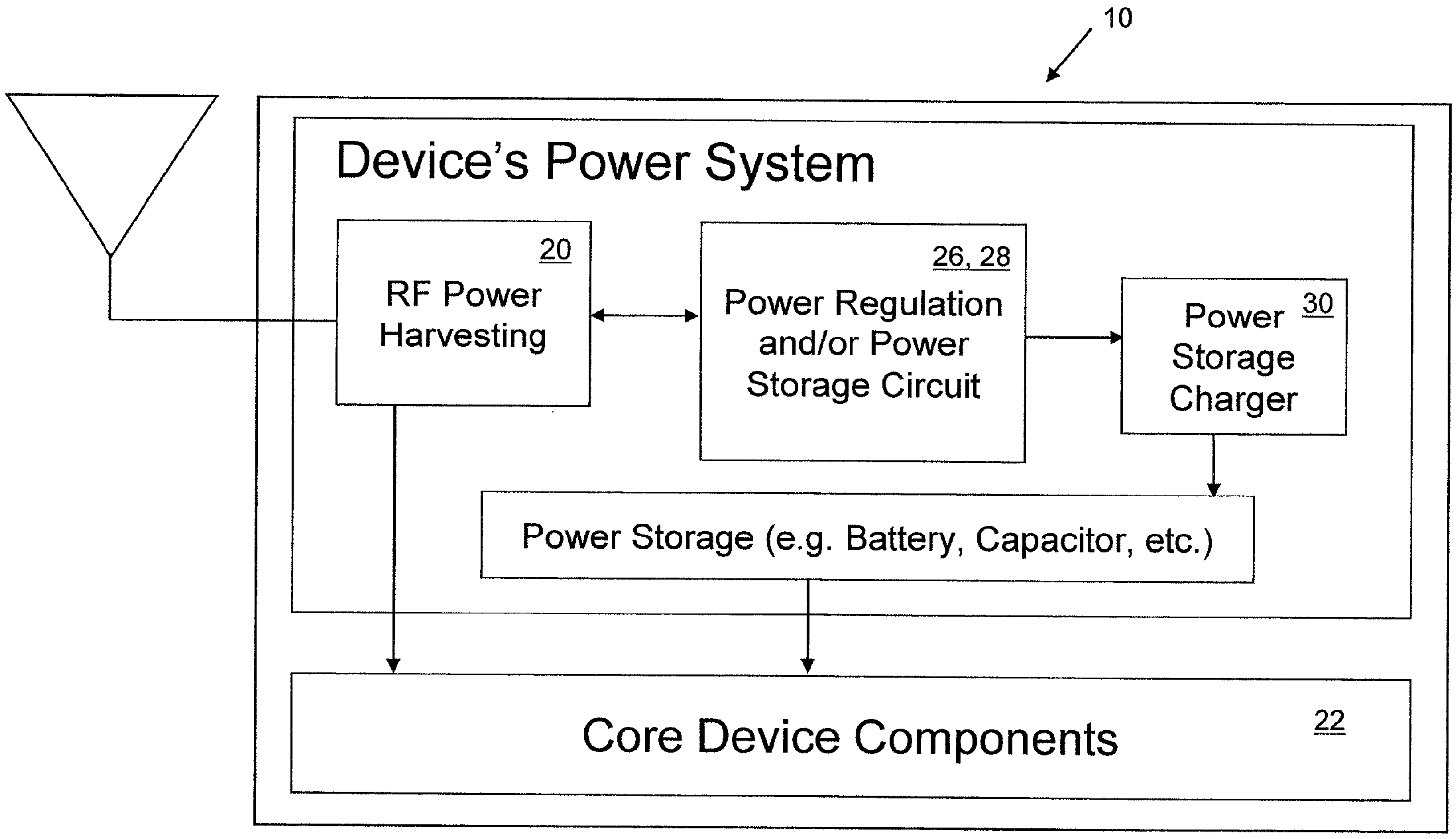
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 25

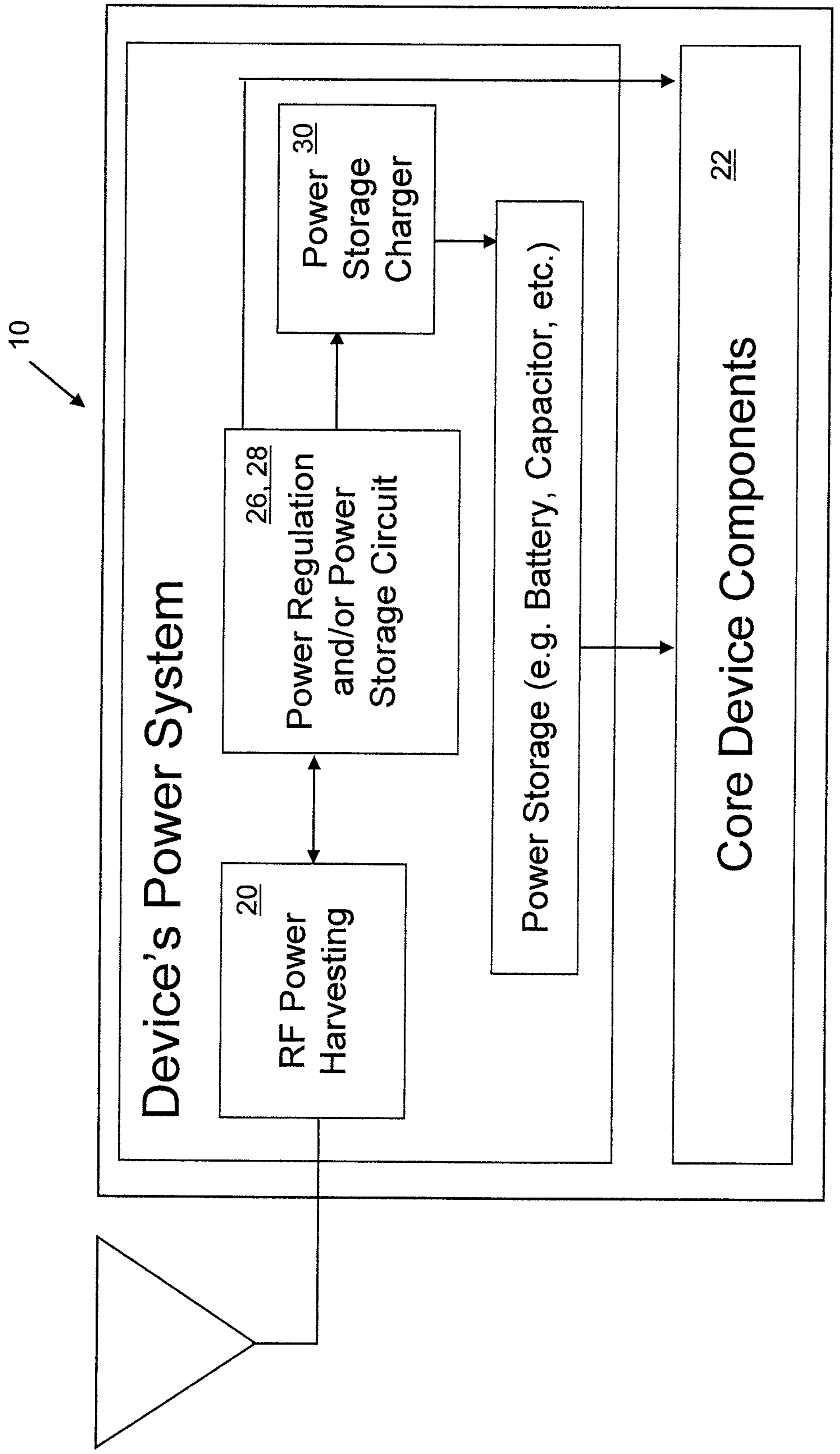
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 26

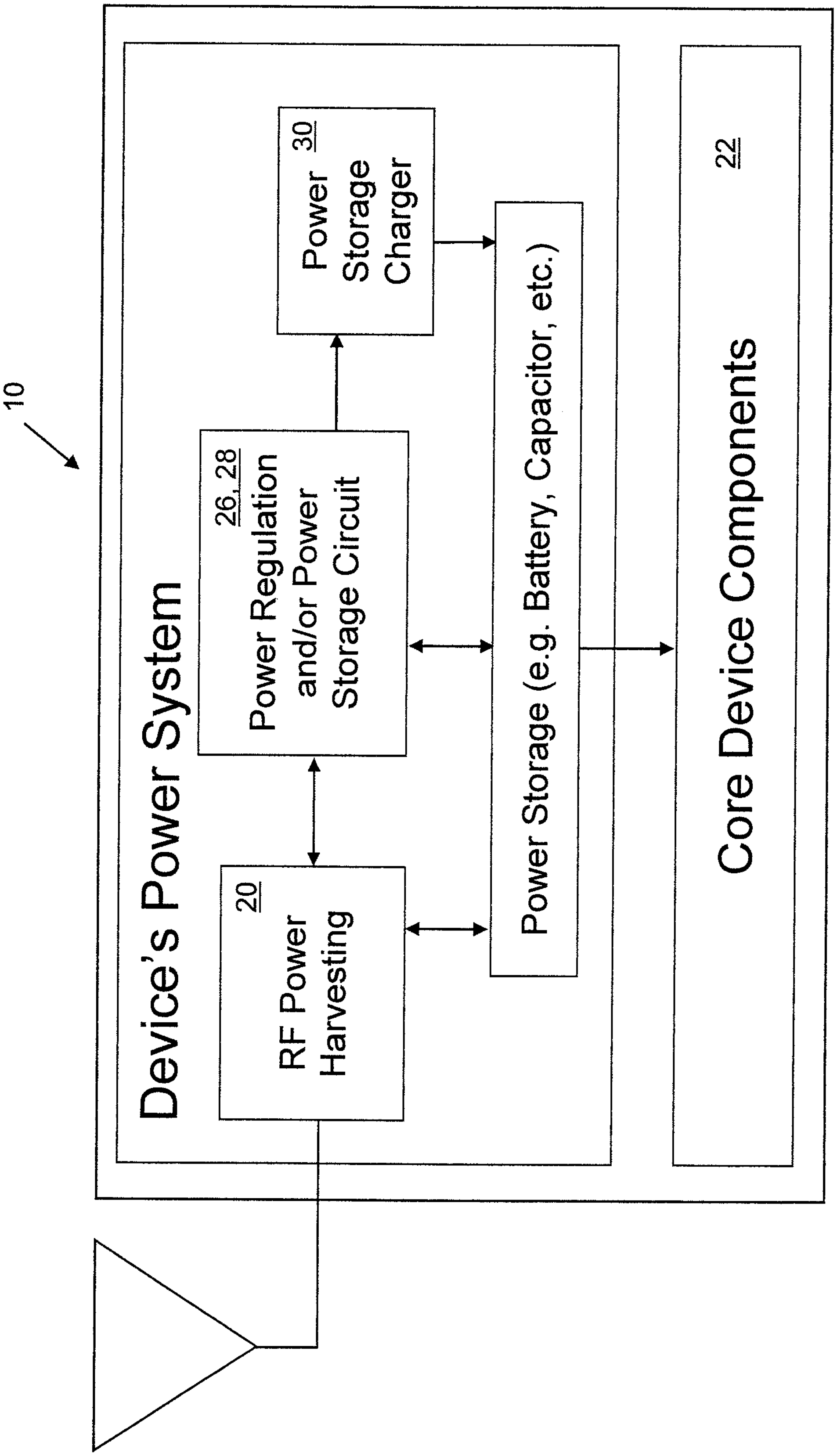
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 27

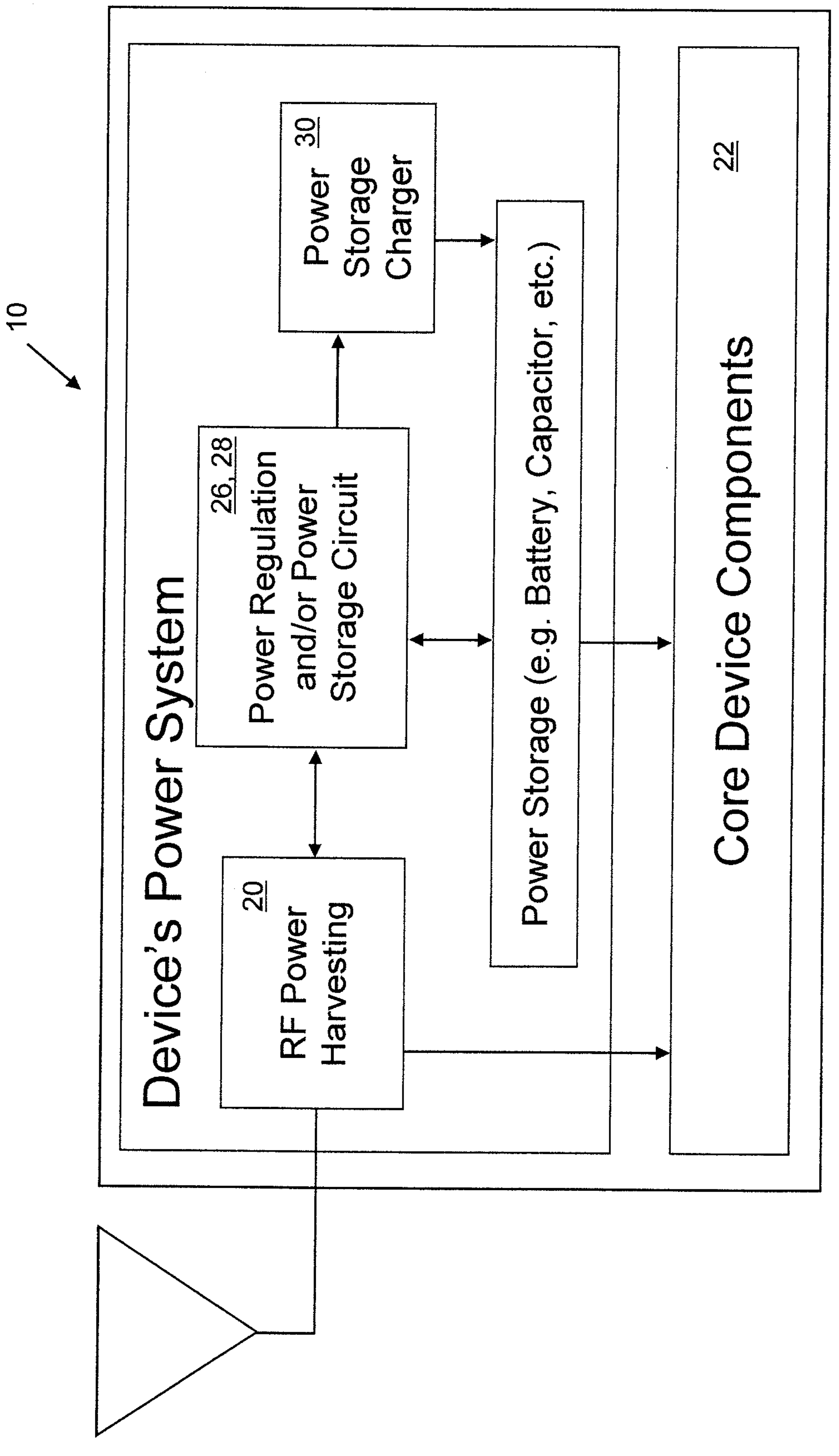
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 28

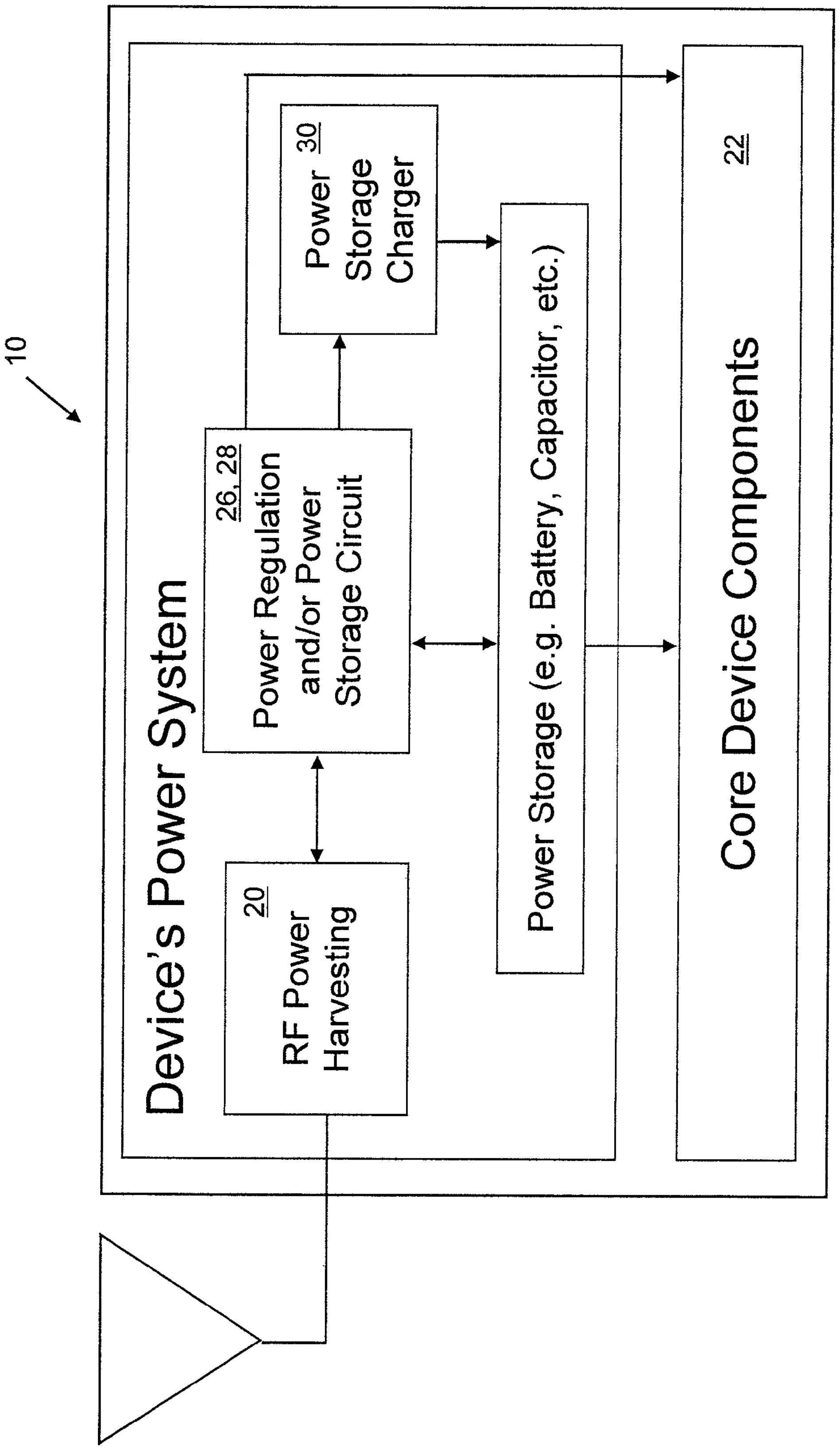
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 29

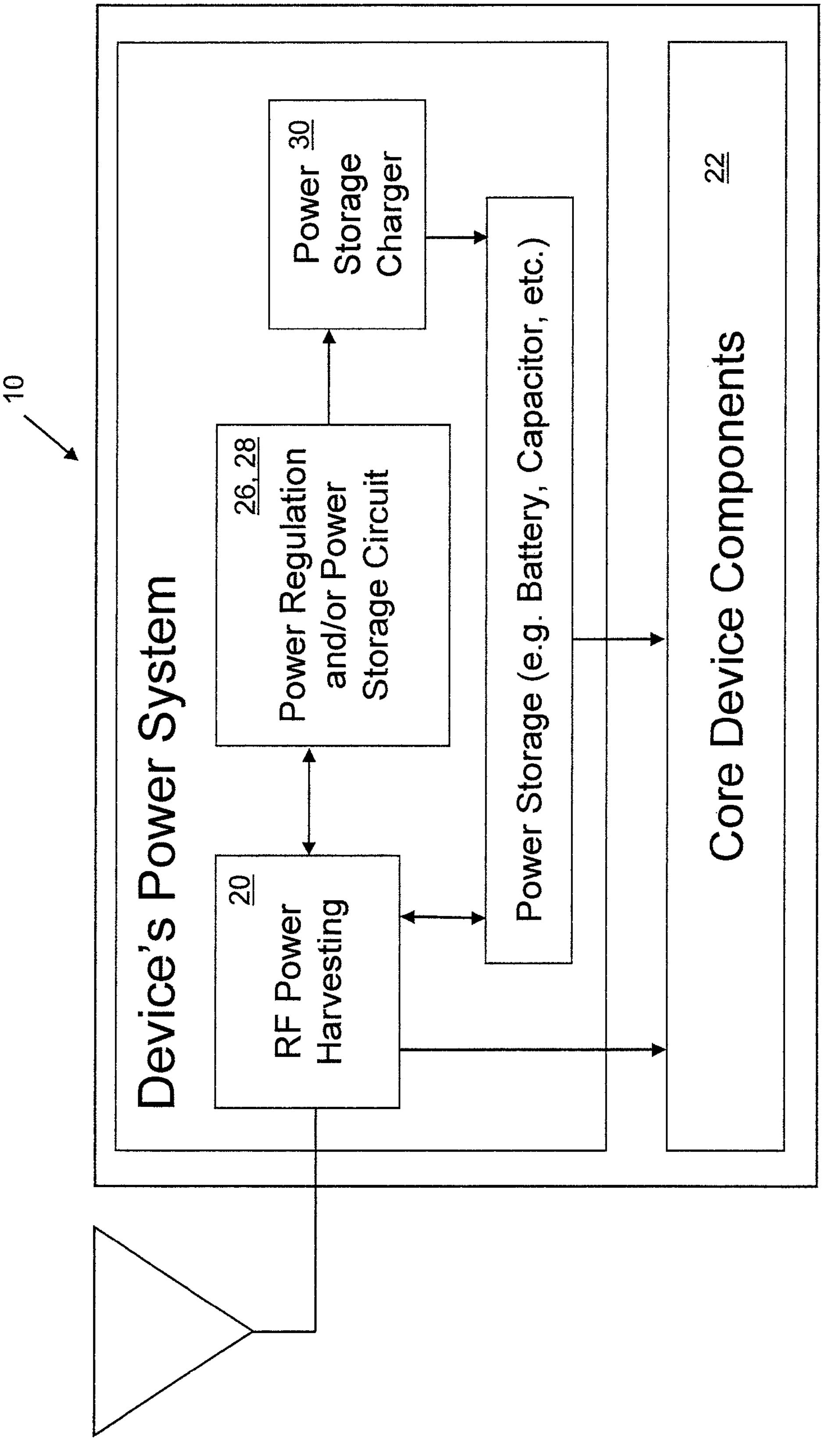
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 30

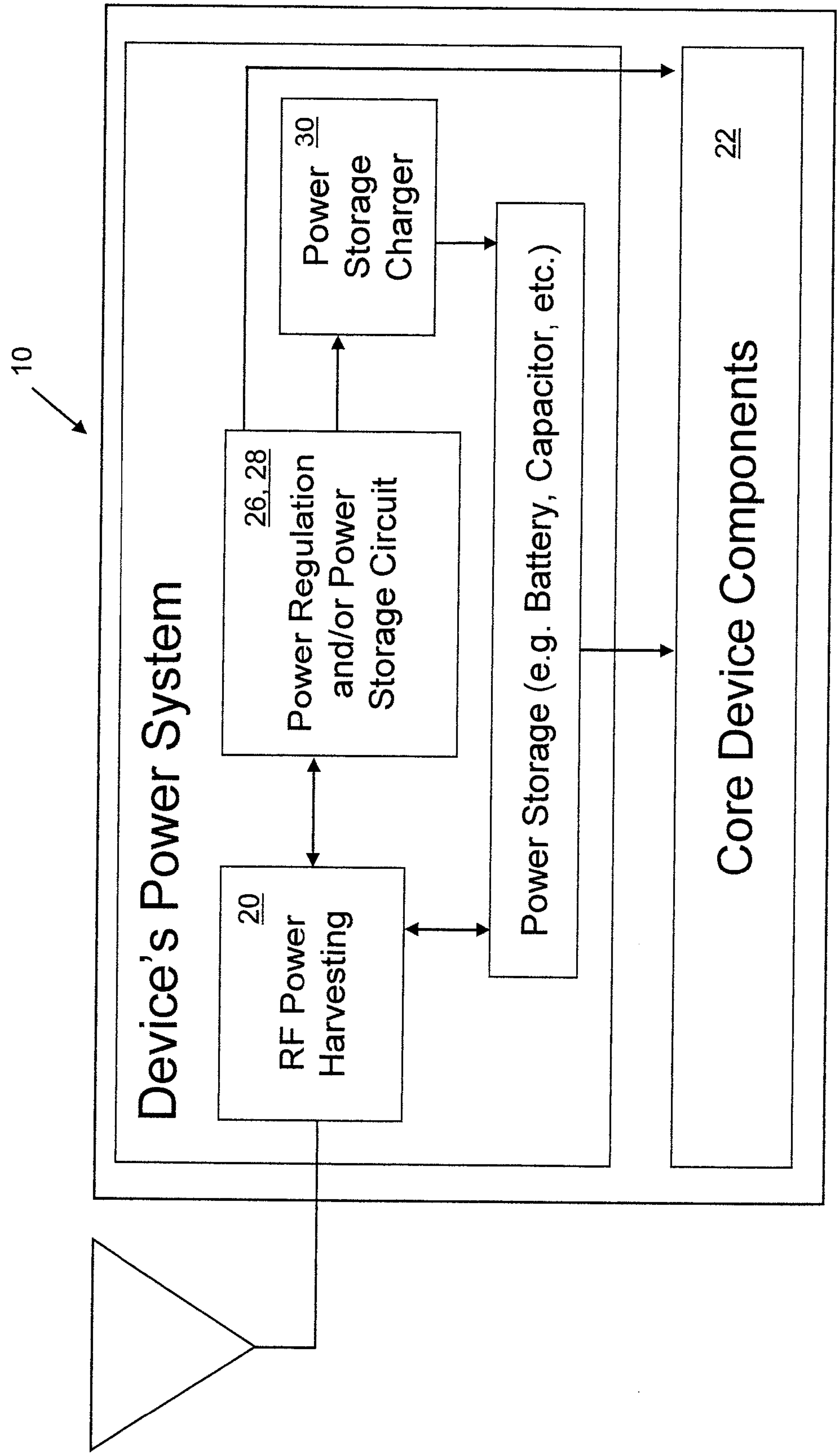
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 31

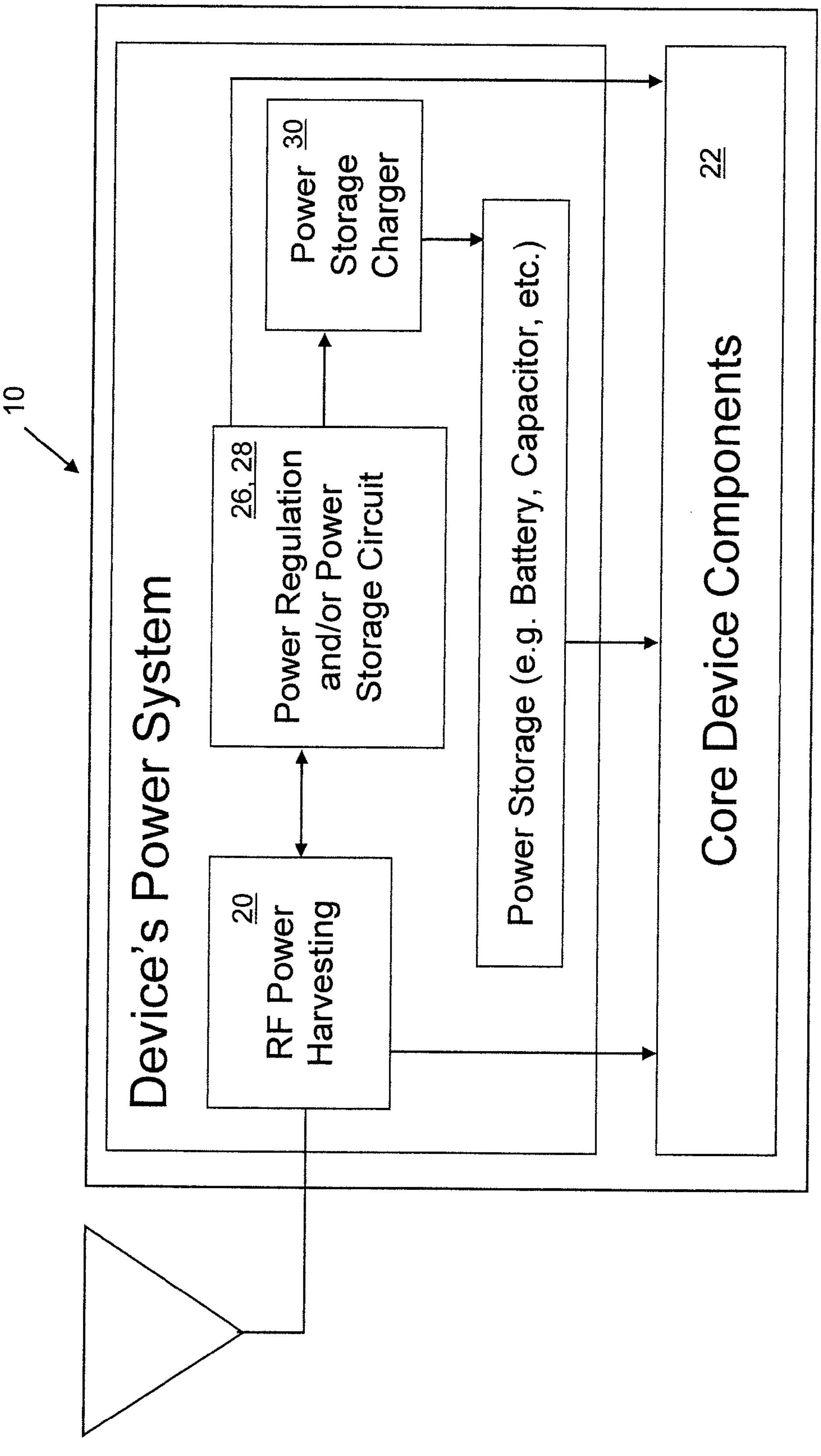
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 32

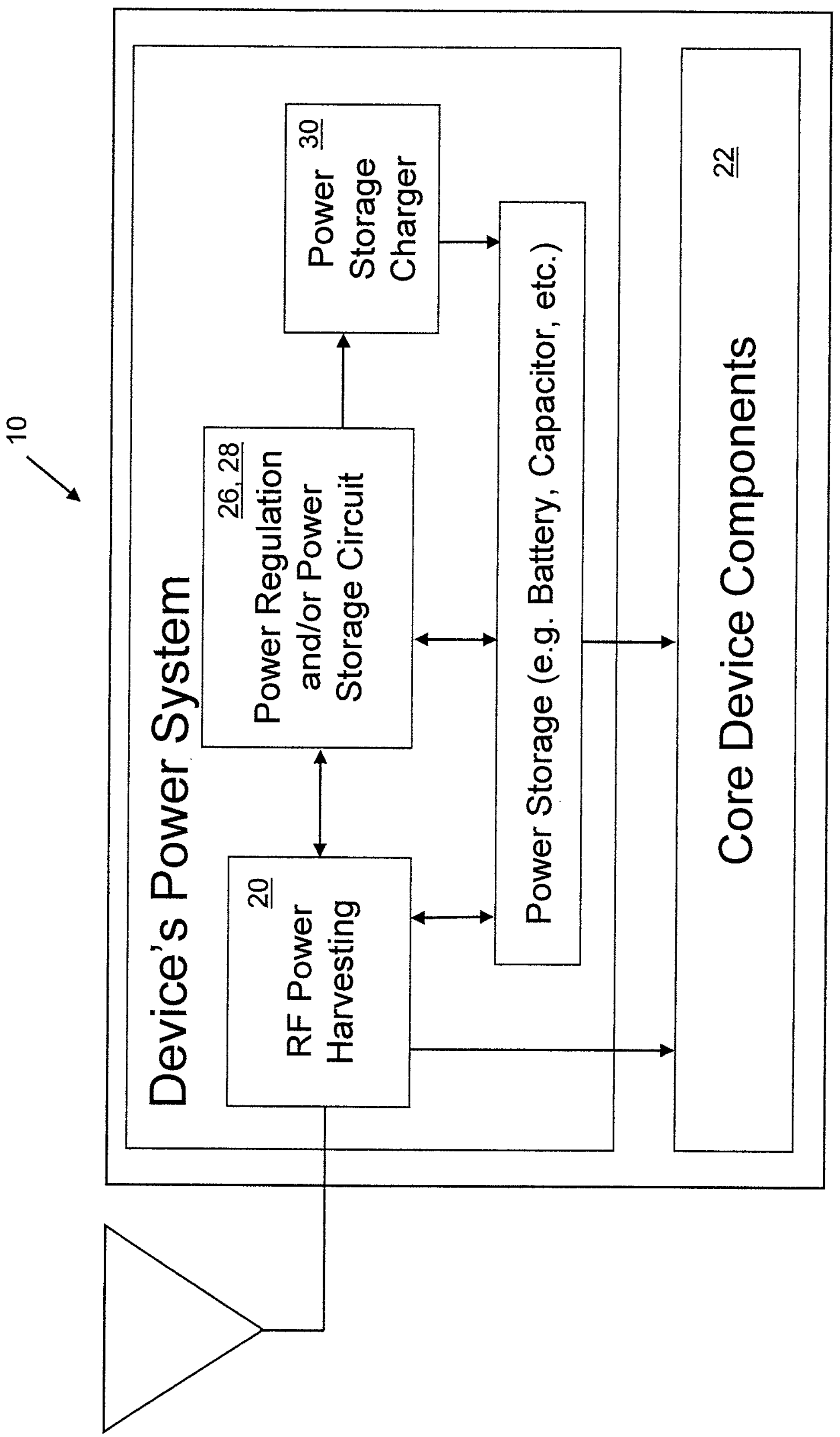
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 33

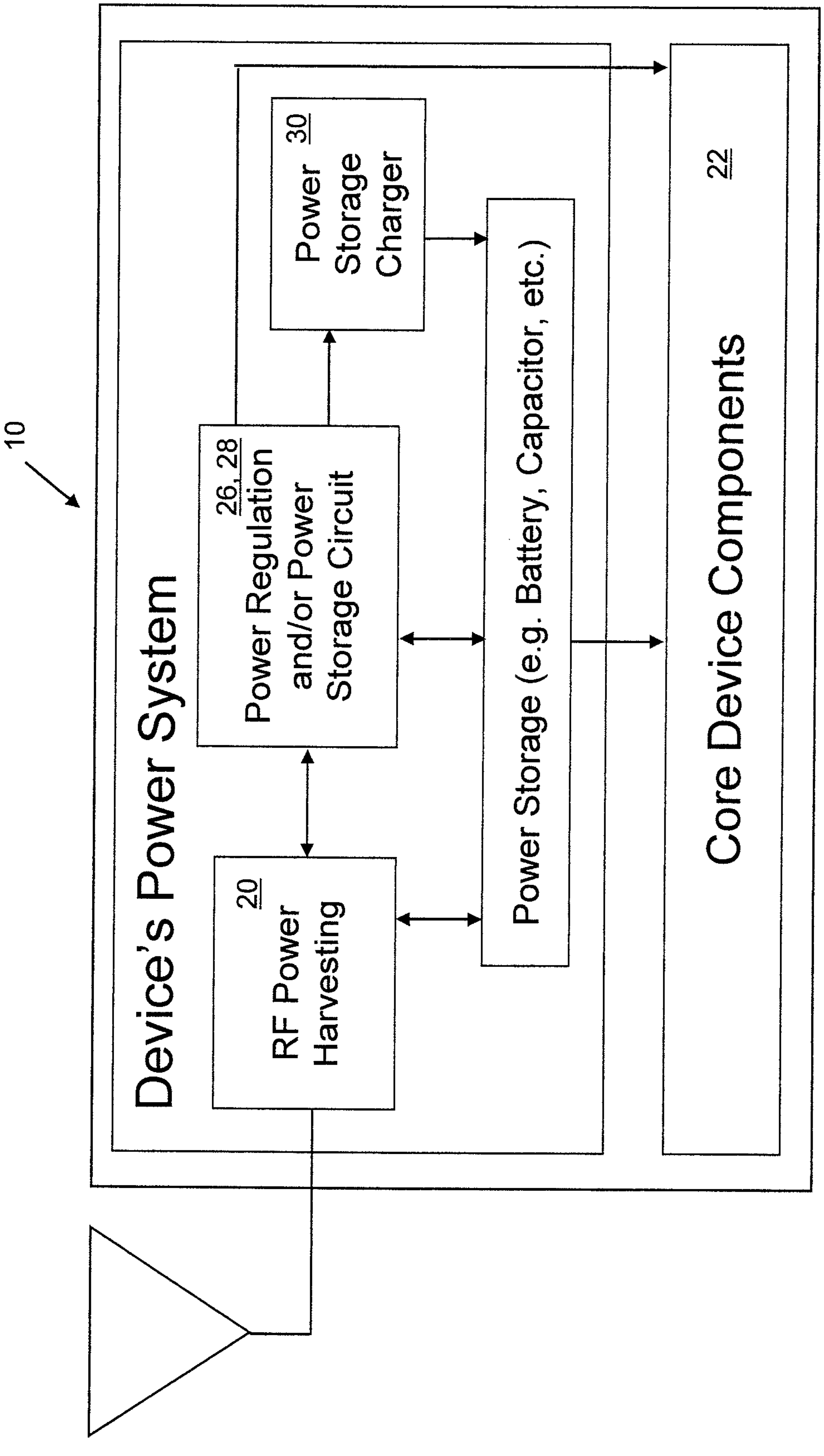
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 34

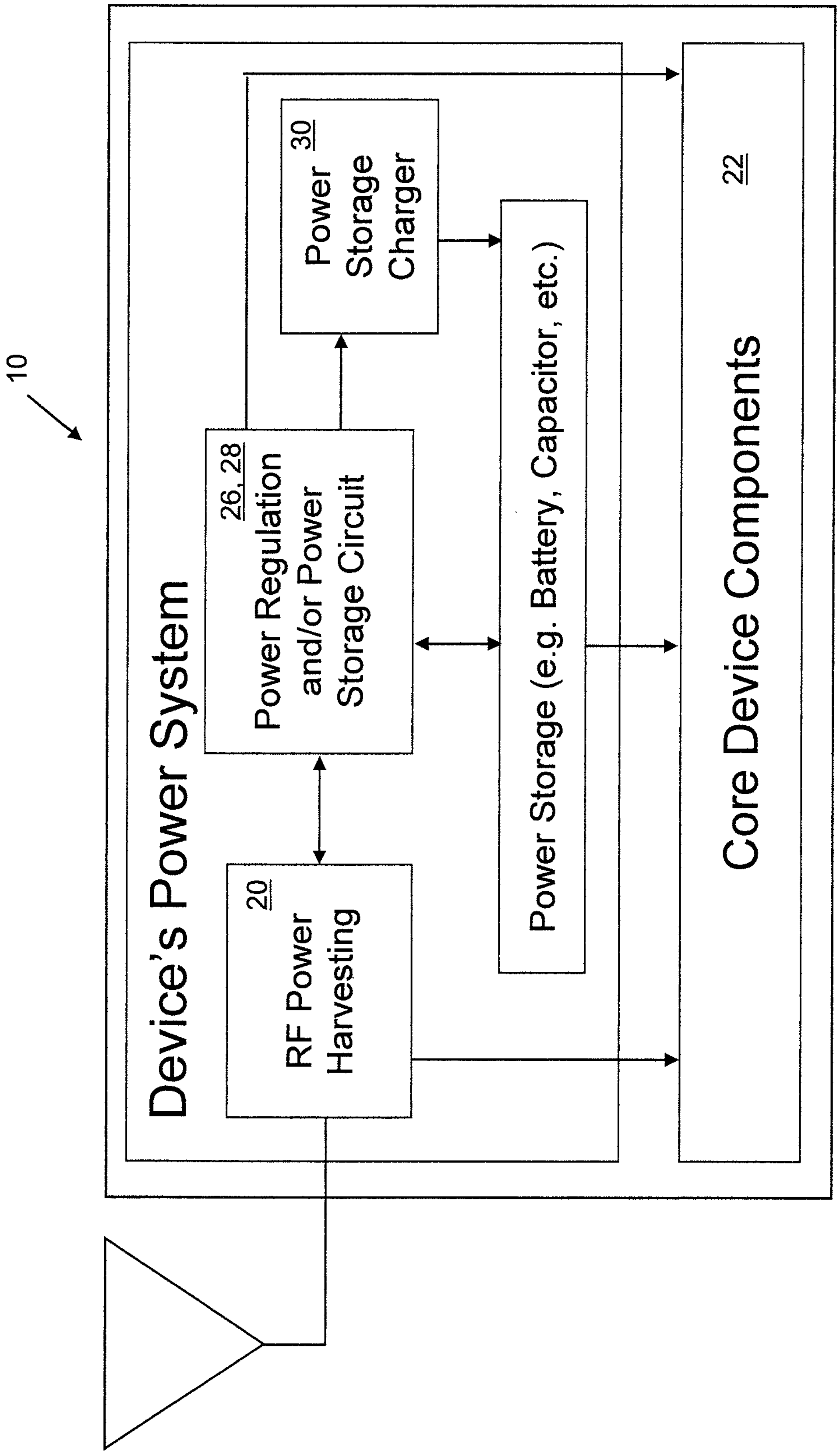
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 35

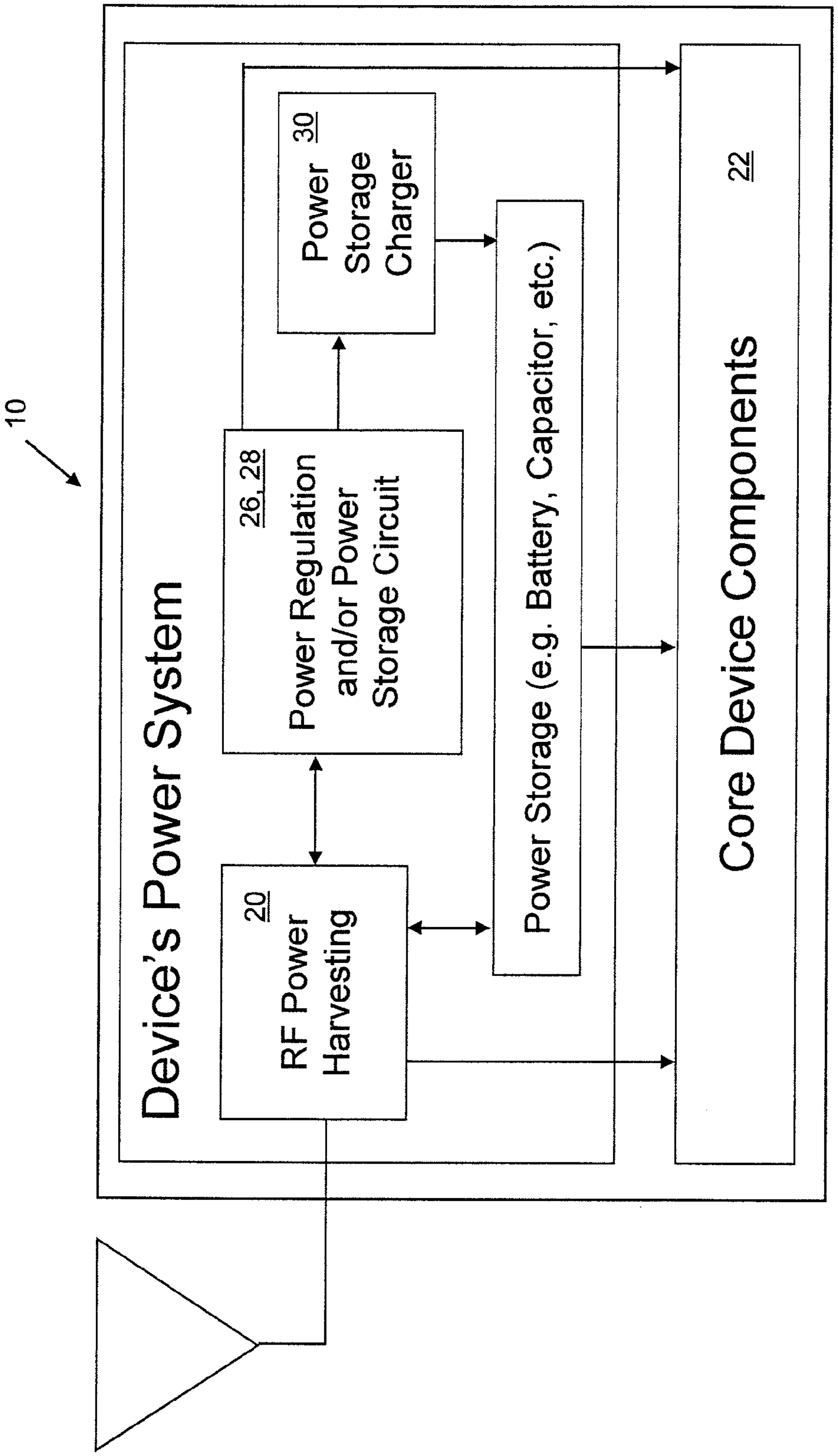
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 36

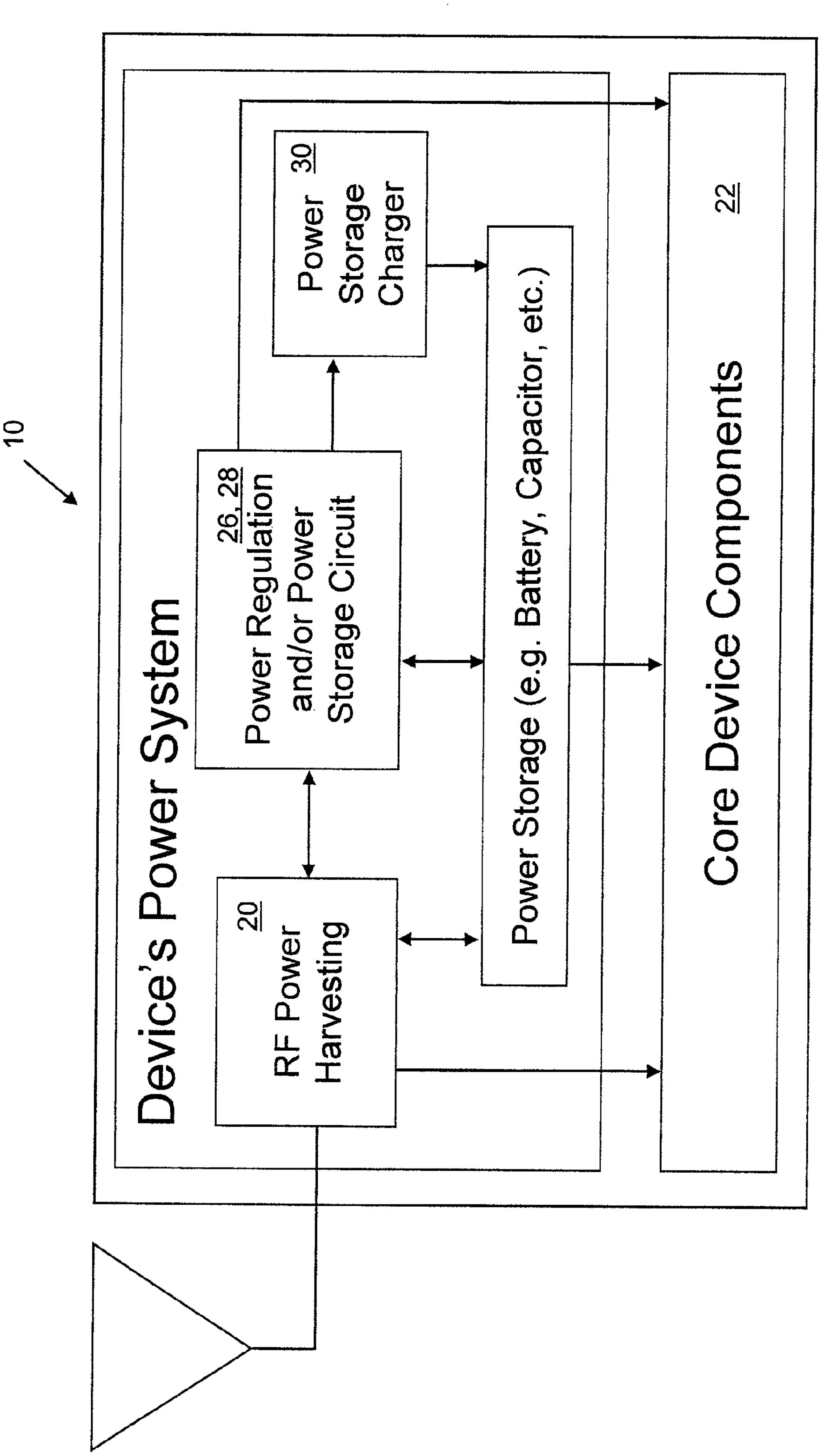
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 37

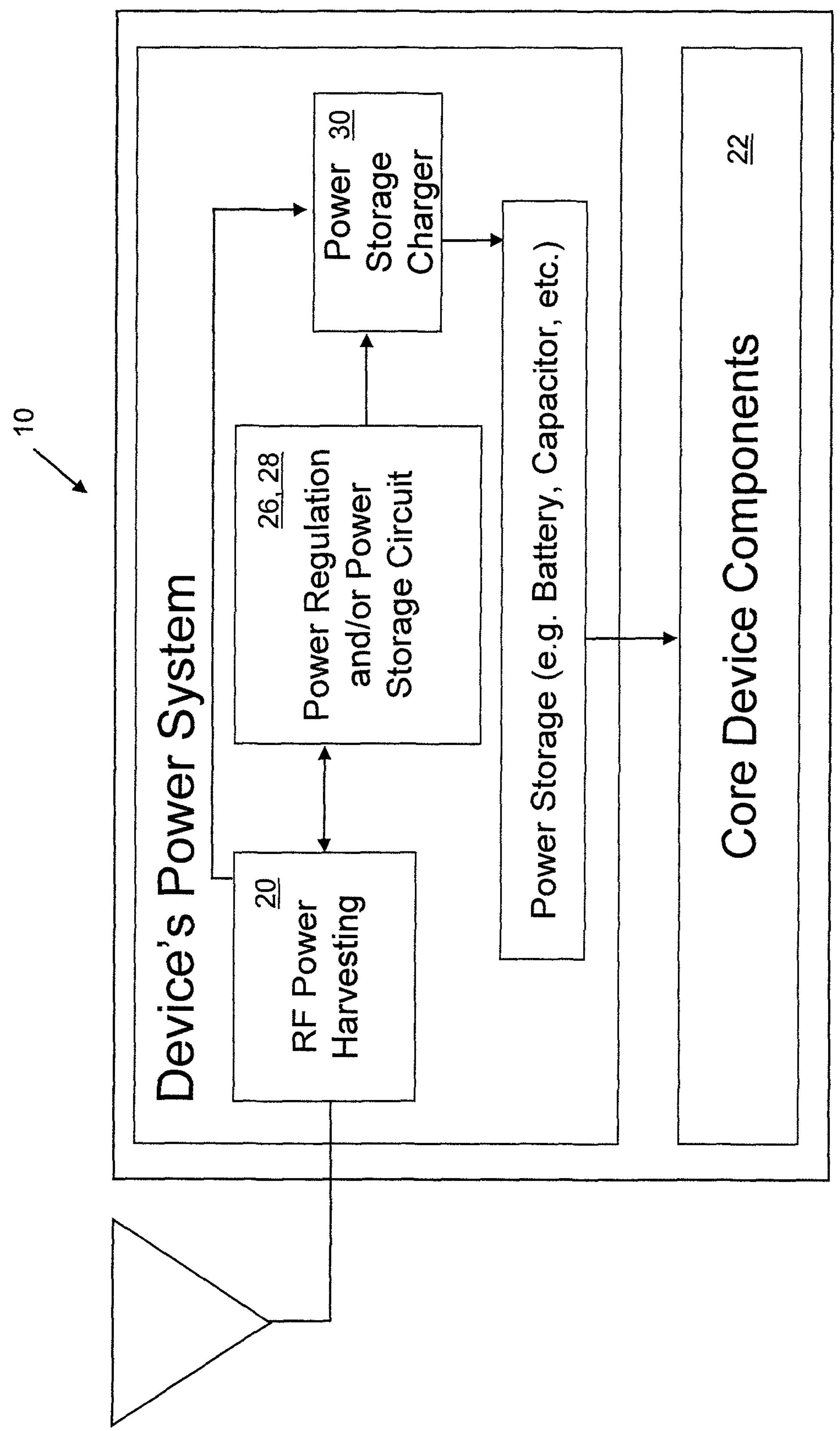
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 38

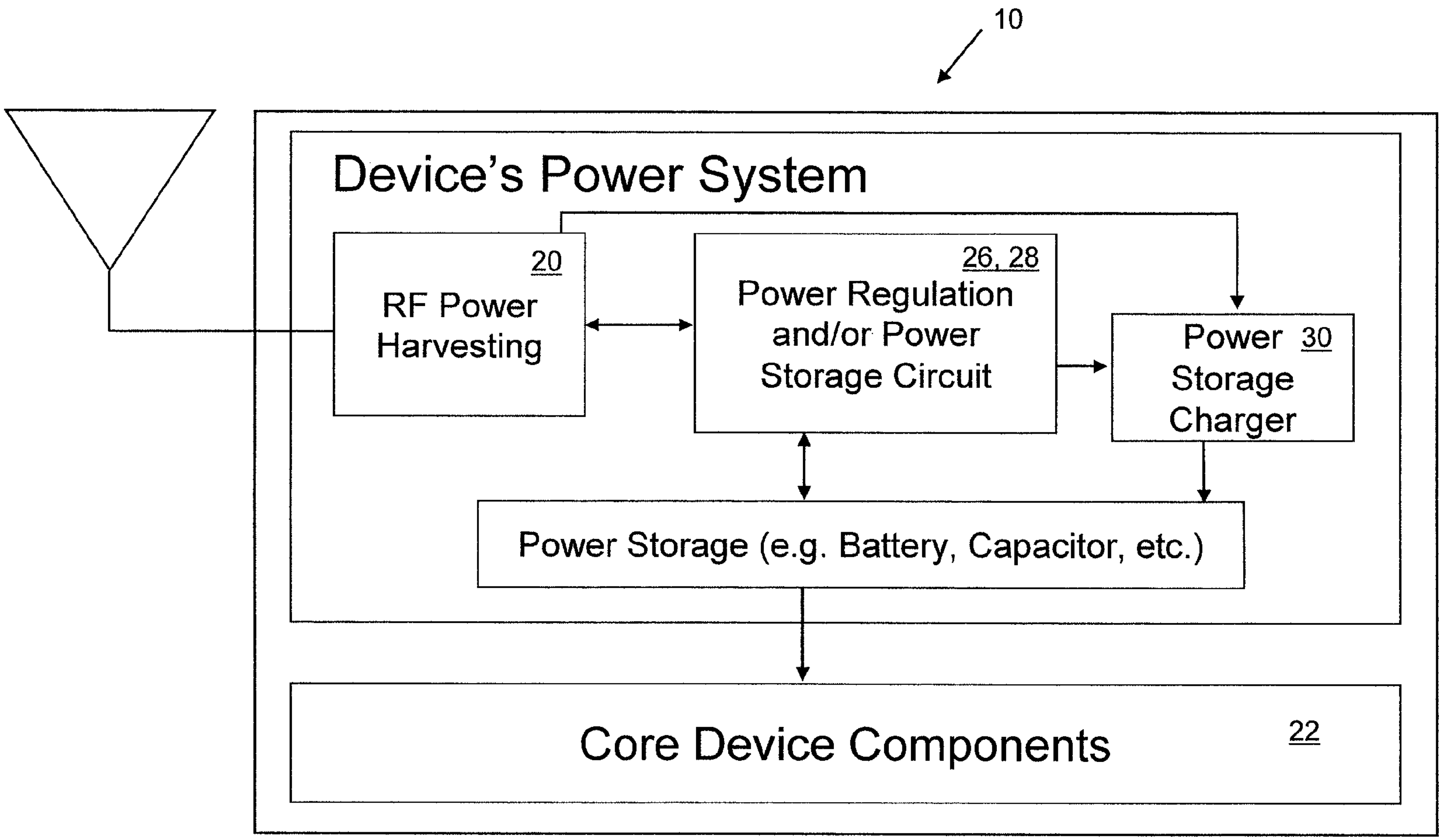
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30 Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 39

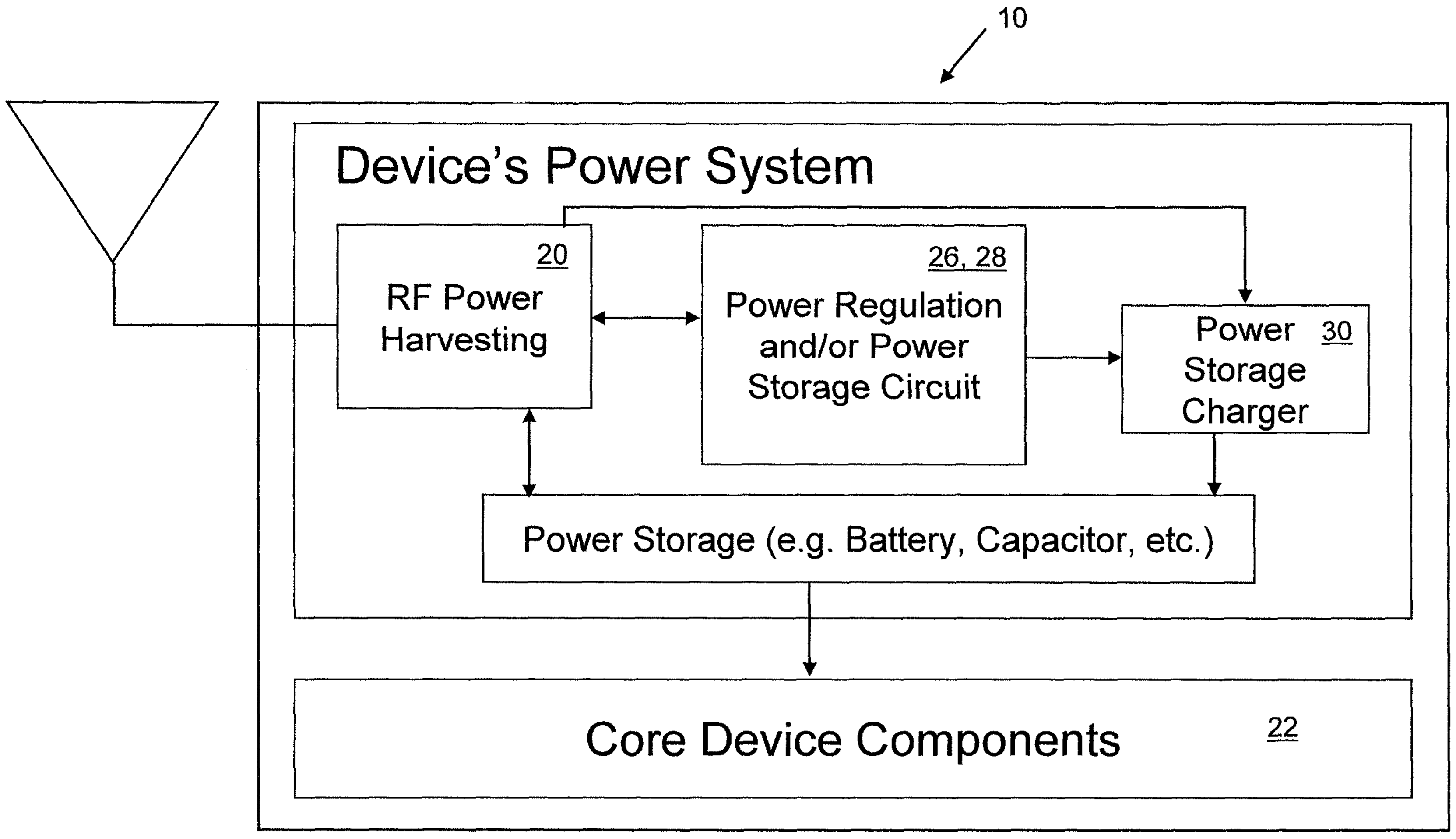
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 40

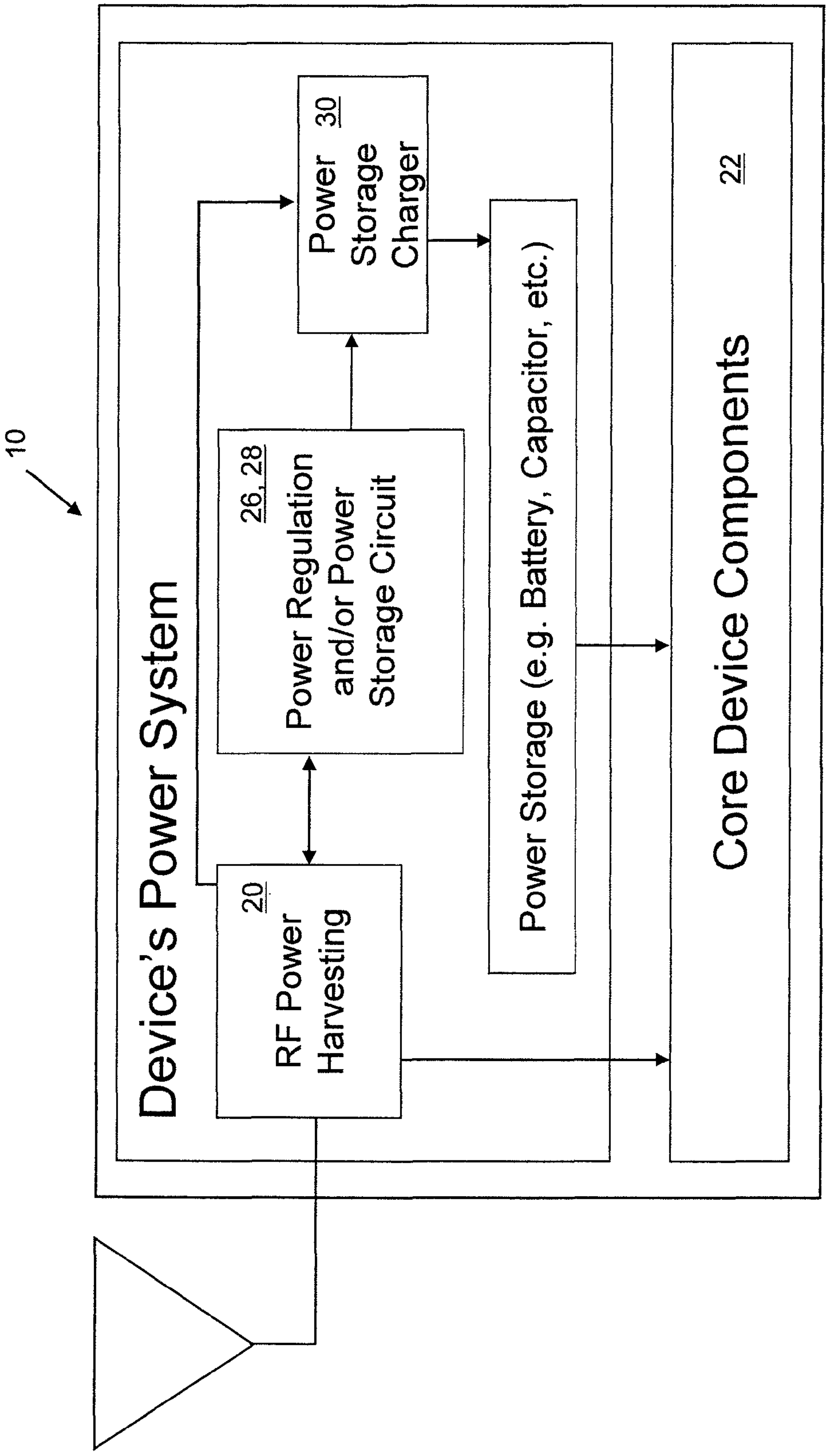
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 41

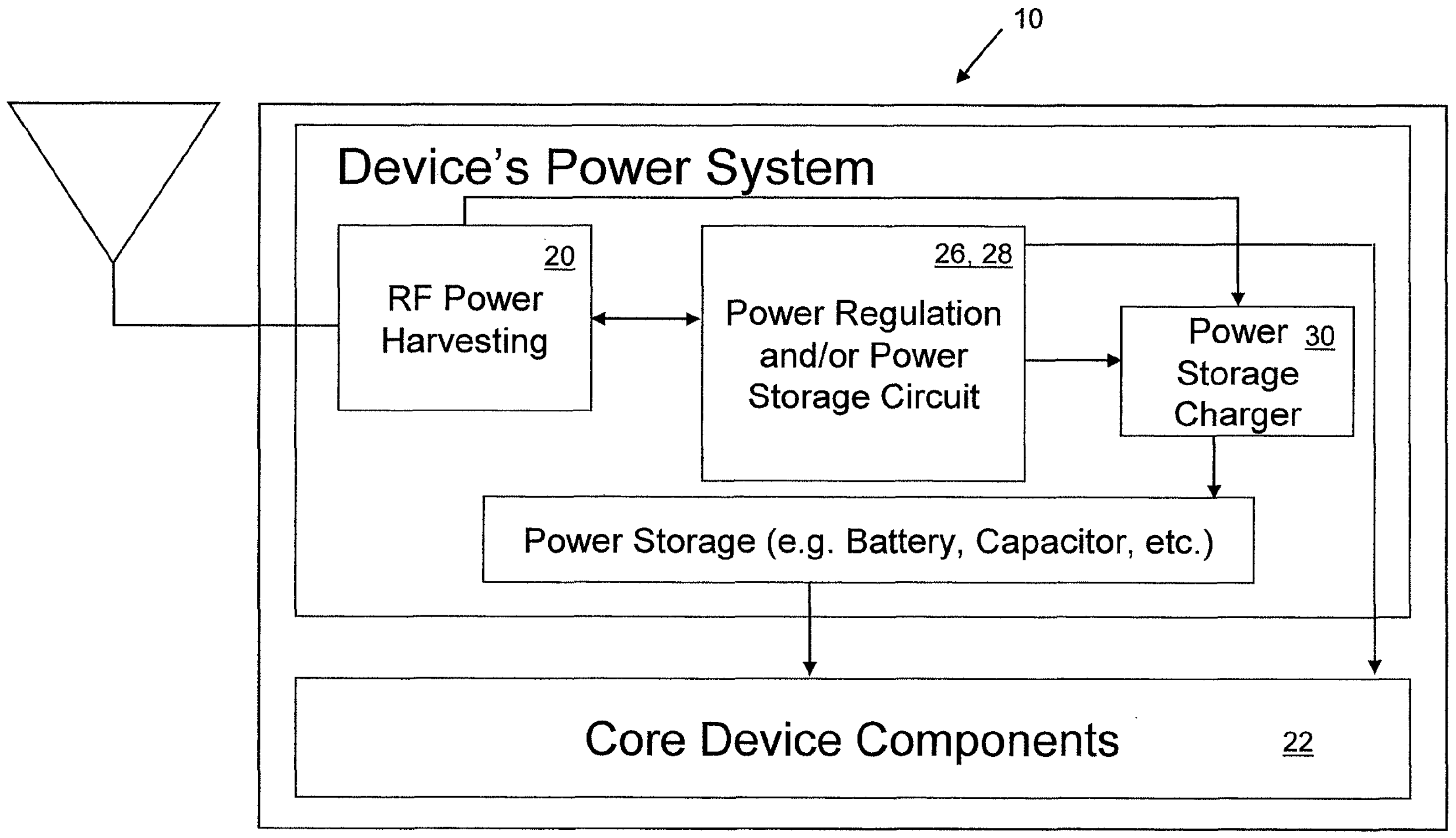
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 42

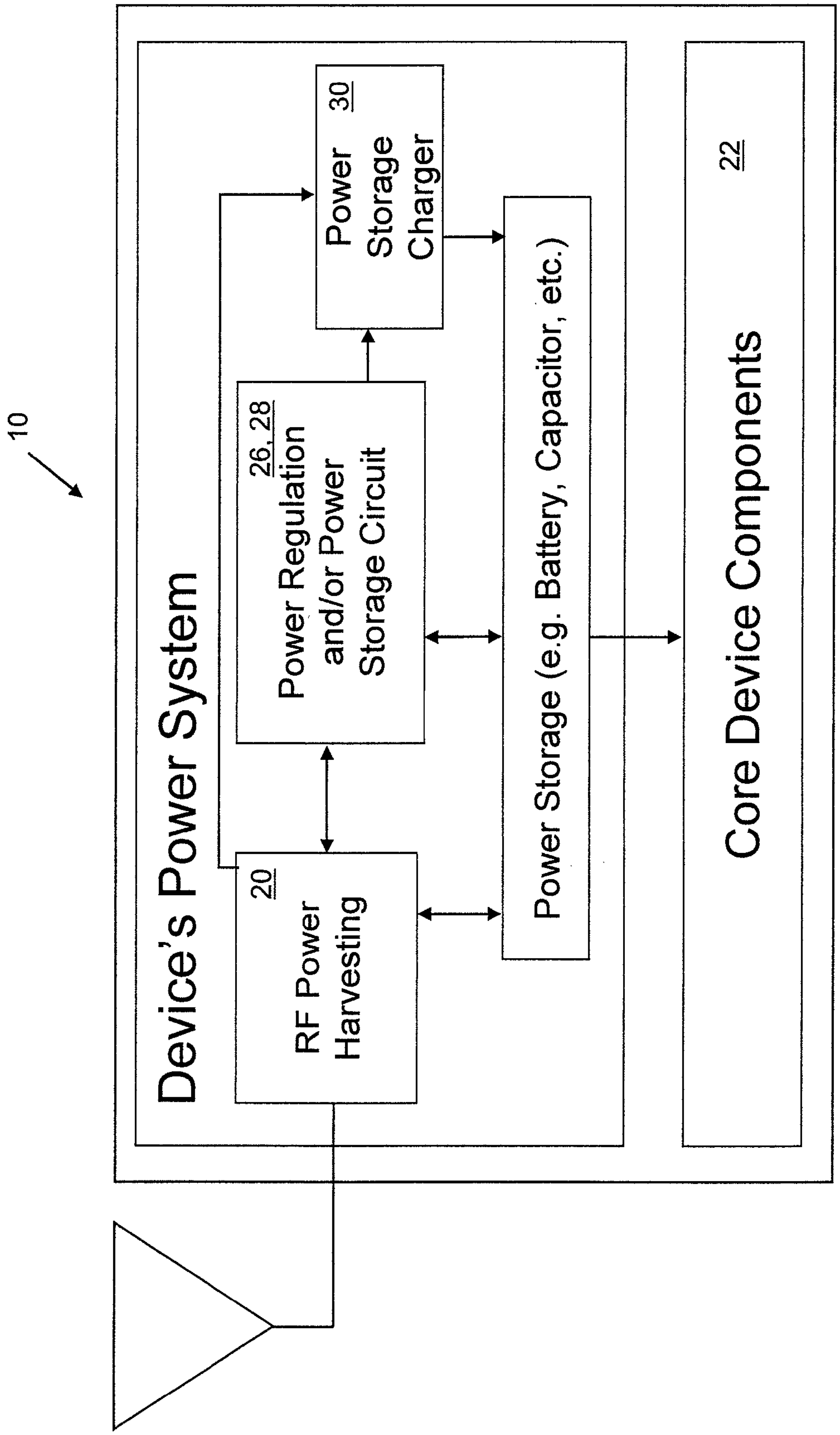
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 43

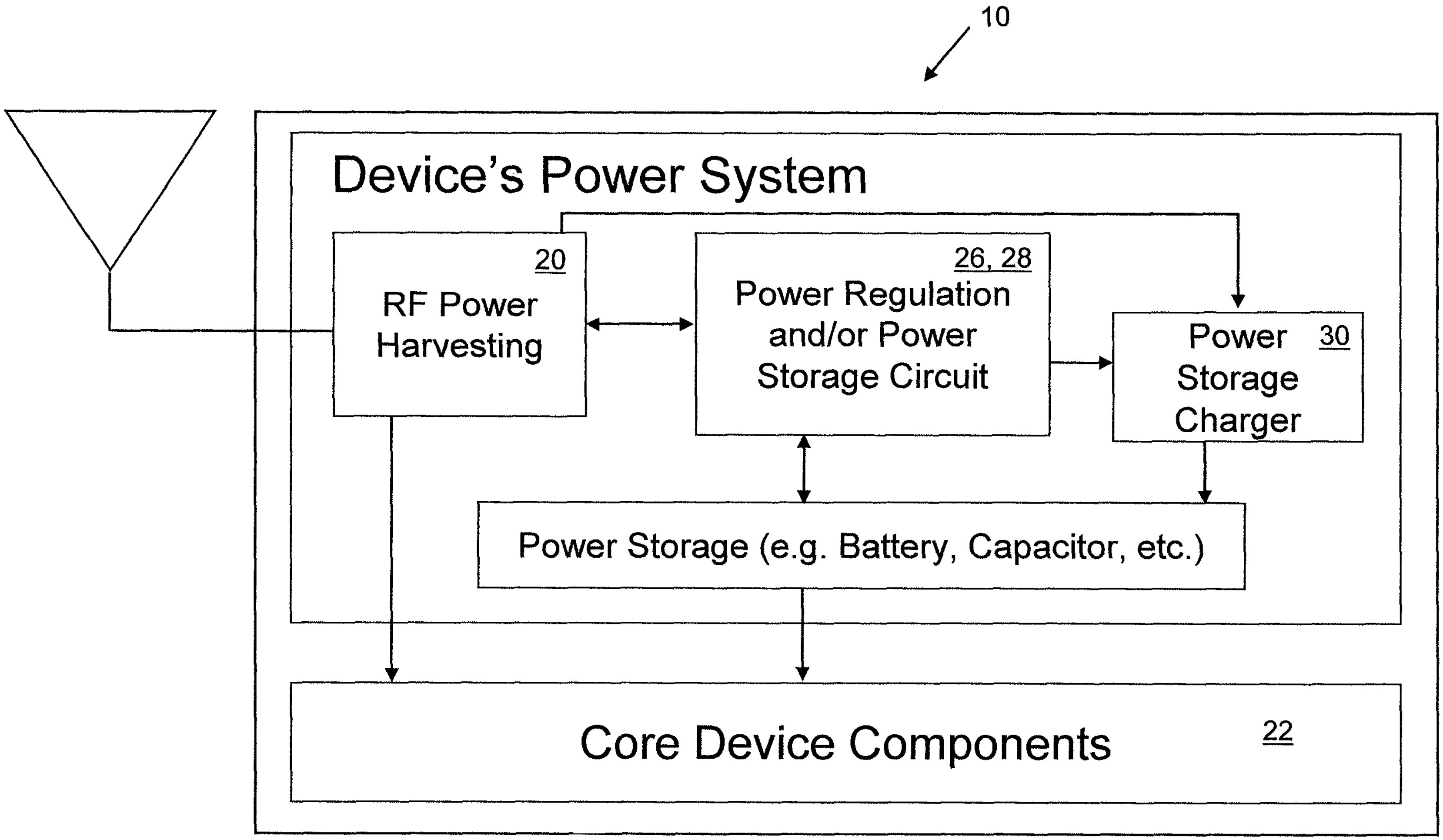
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 44

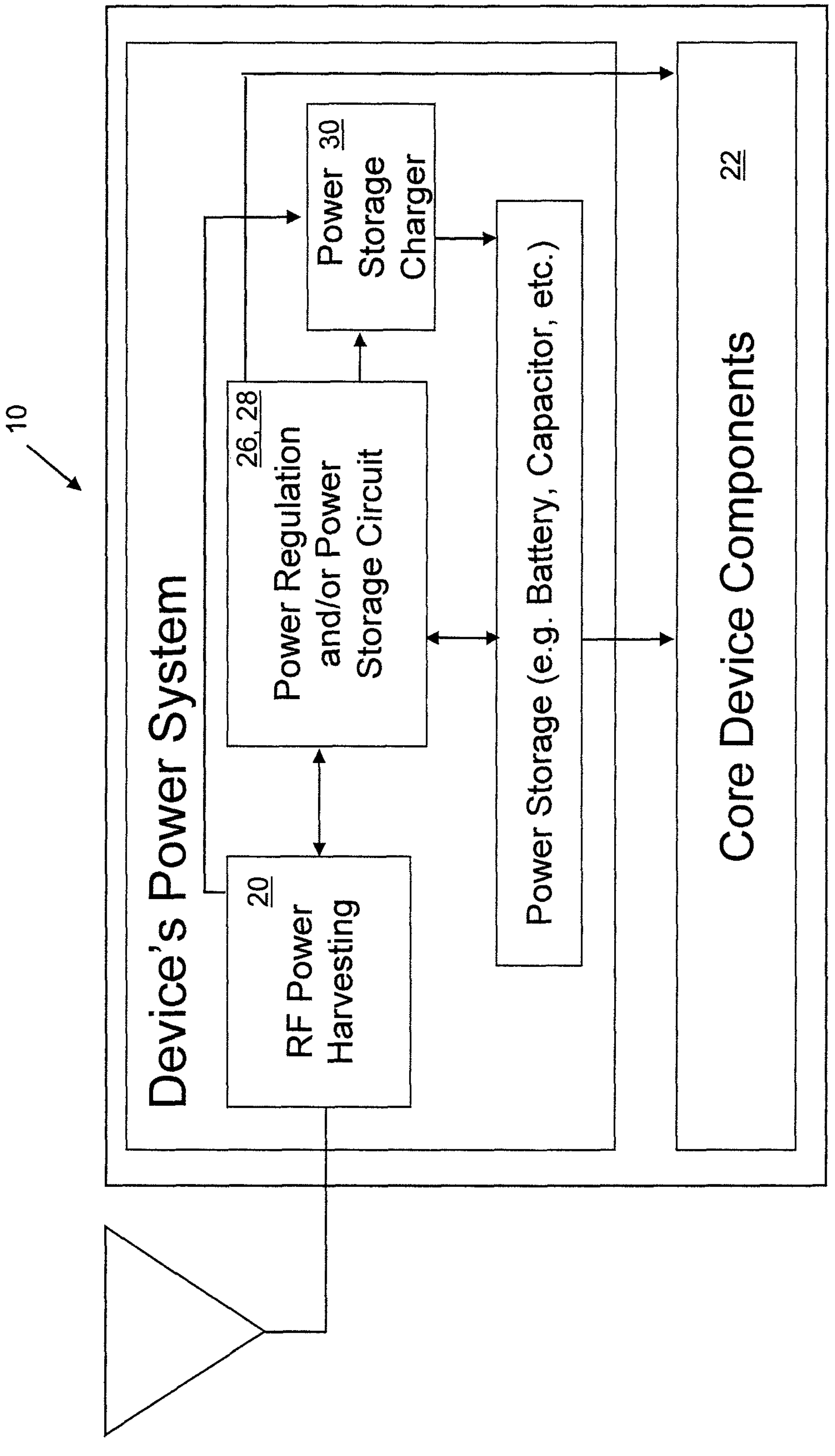
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 45

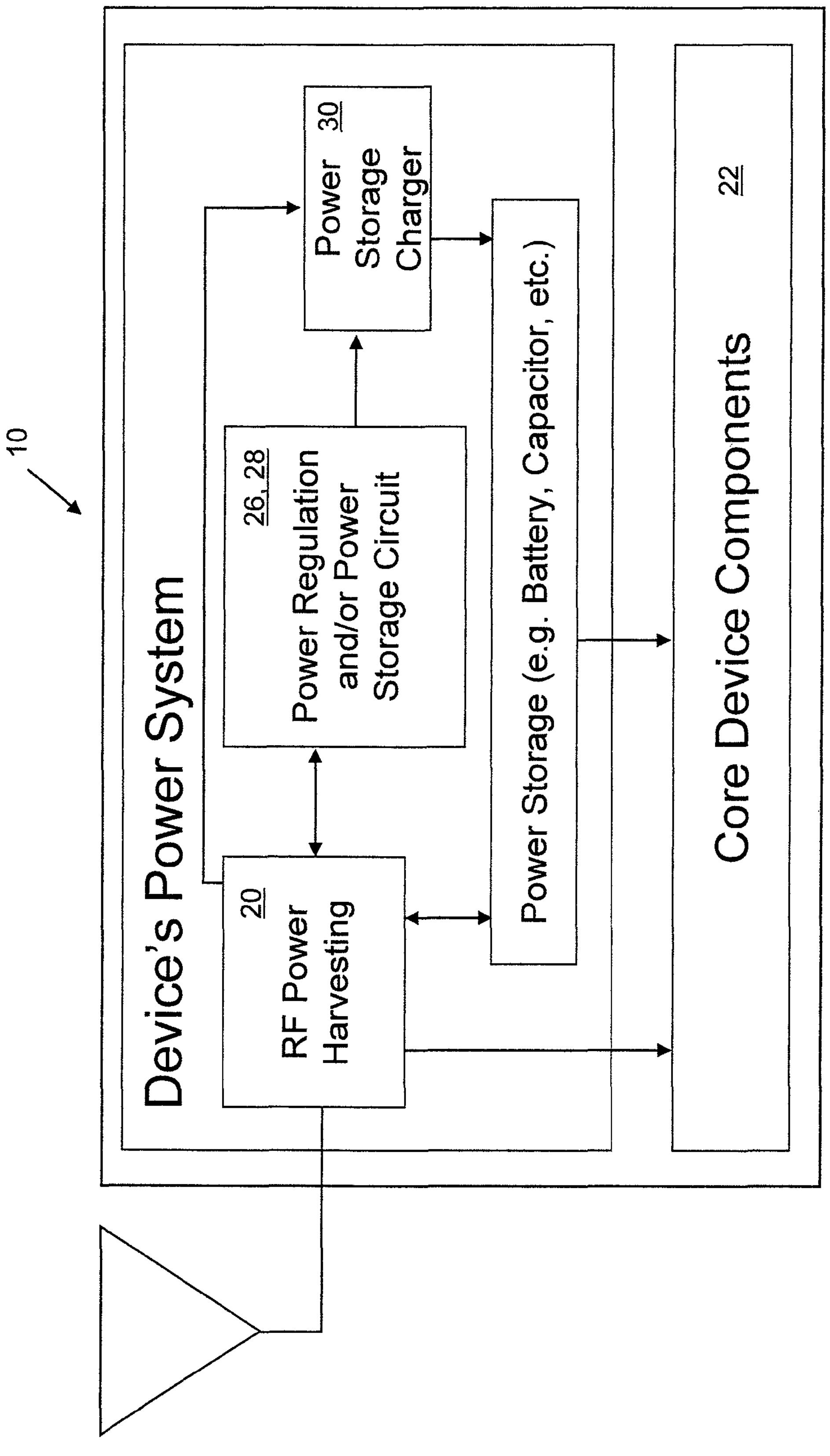
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 46

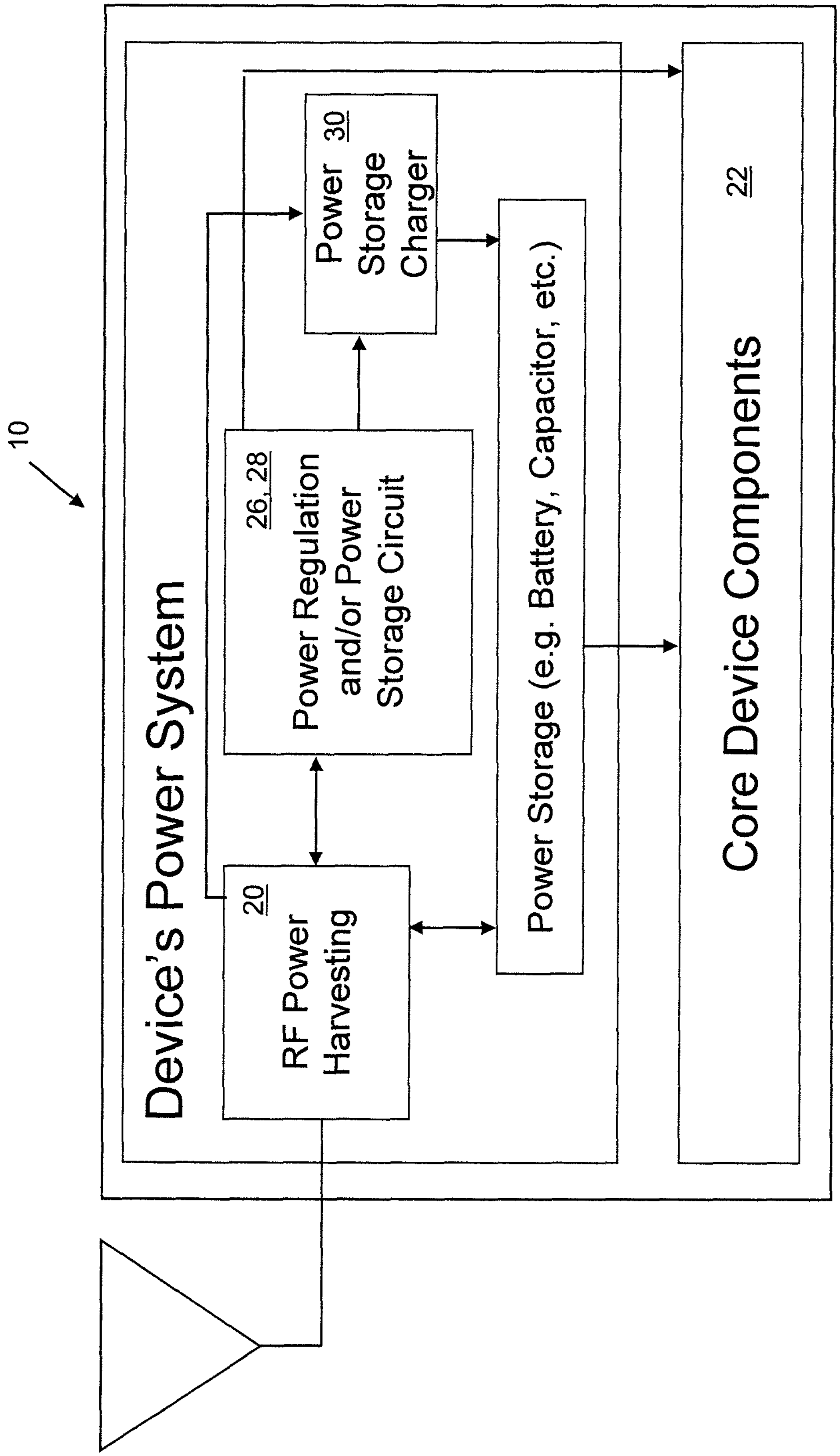
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 47

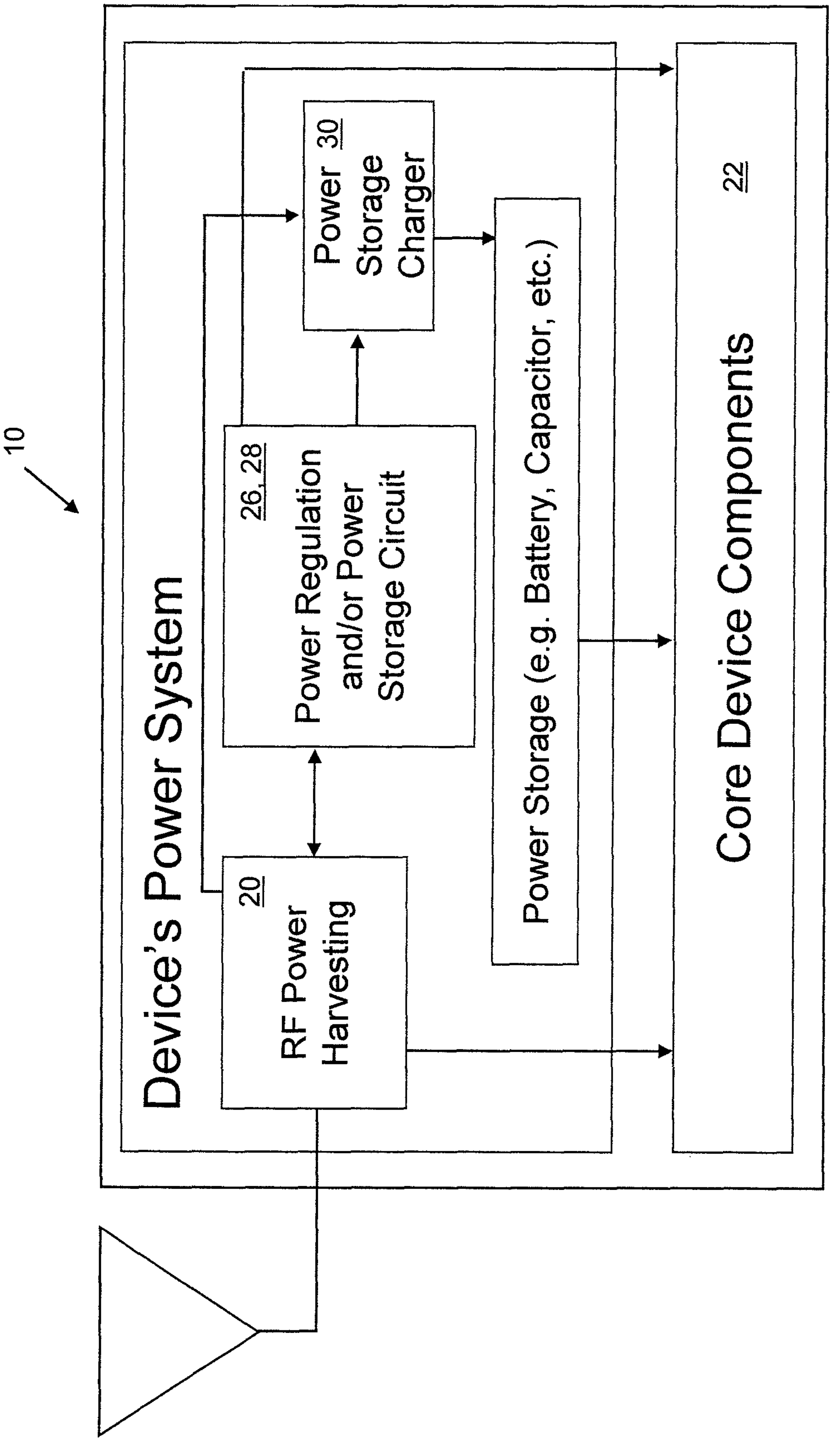
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 48

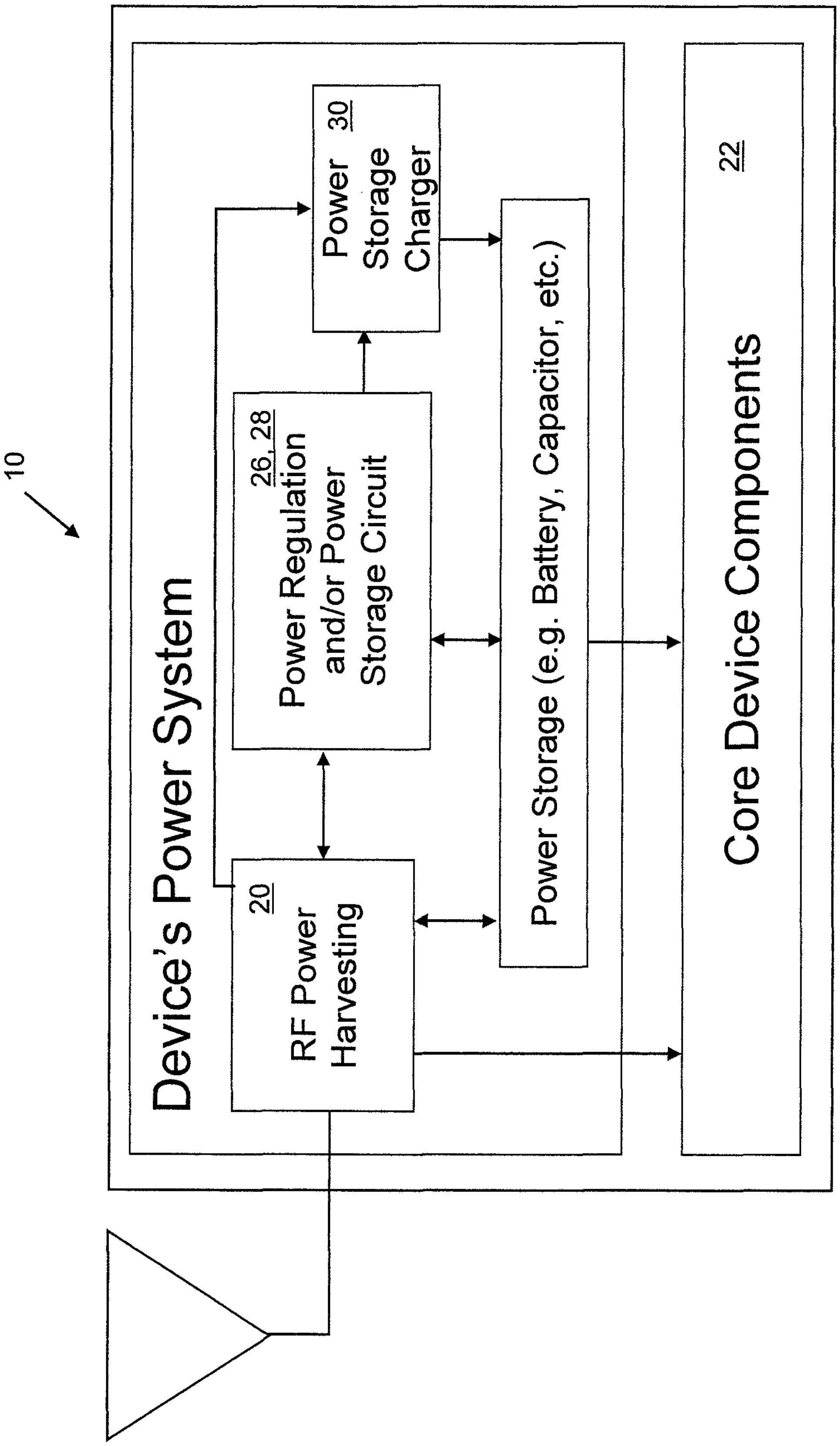
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 49

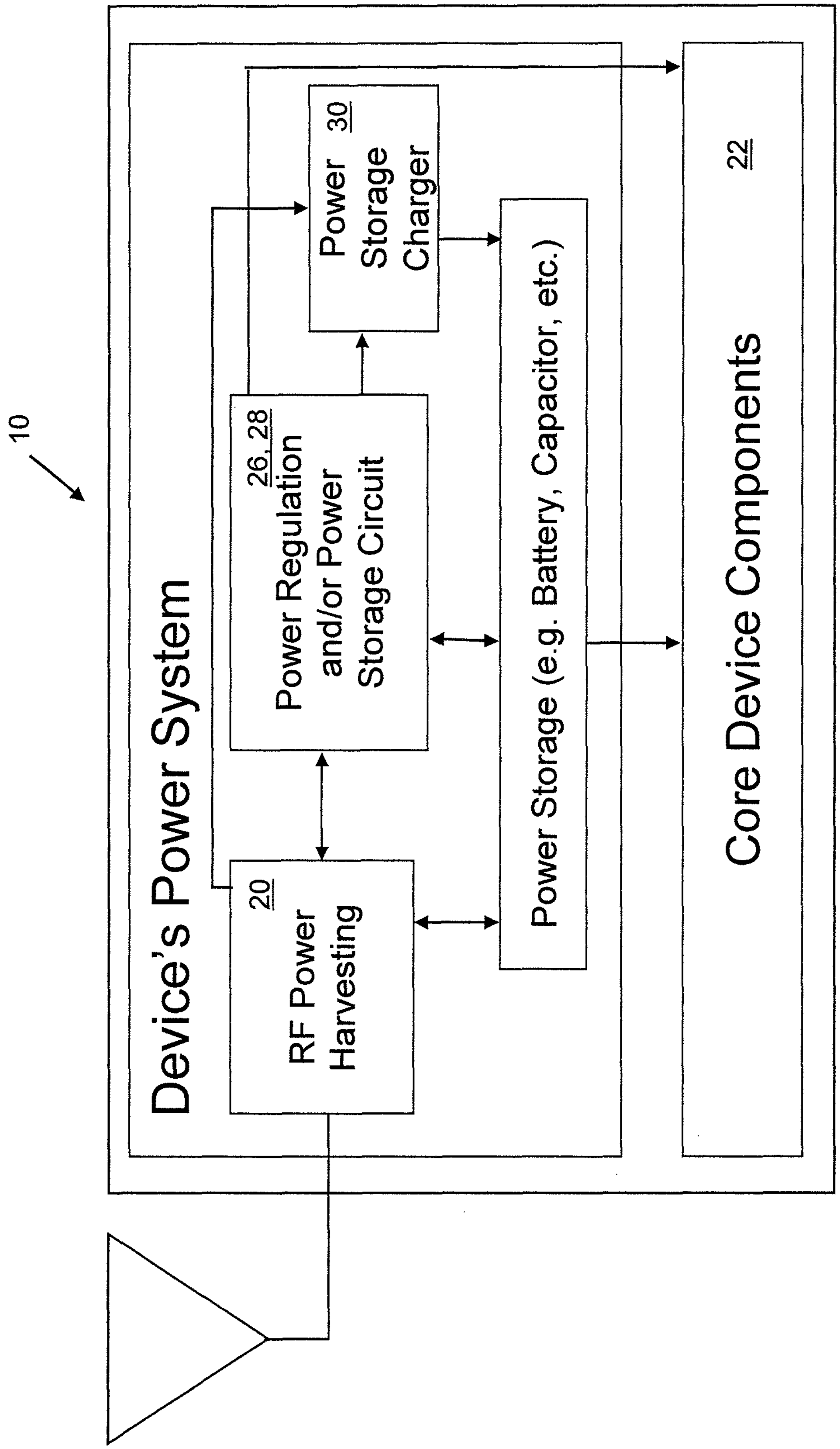
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 50

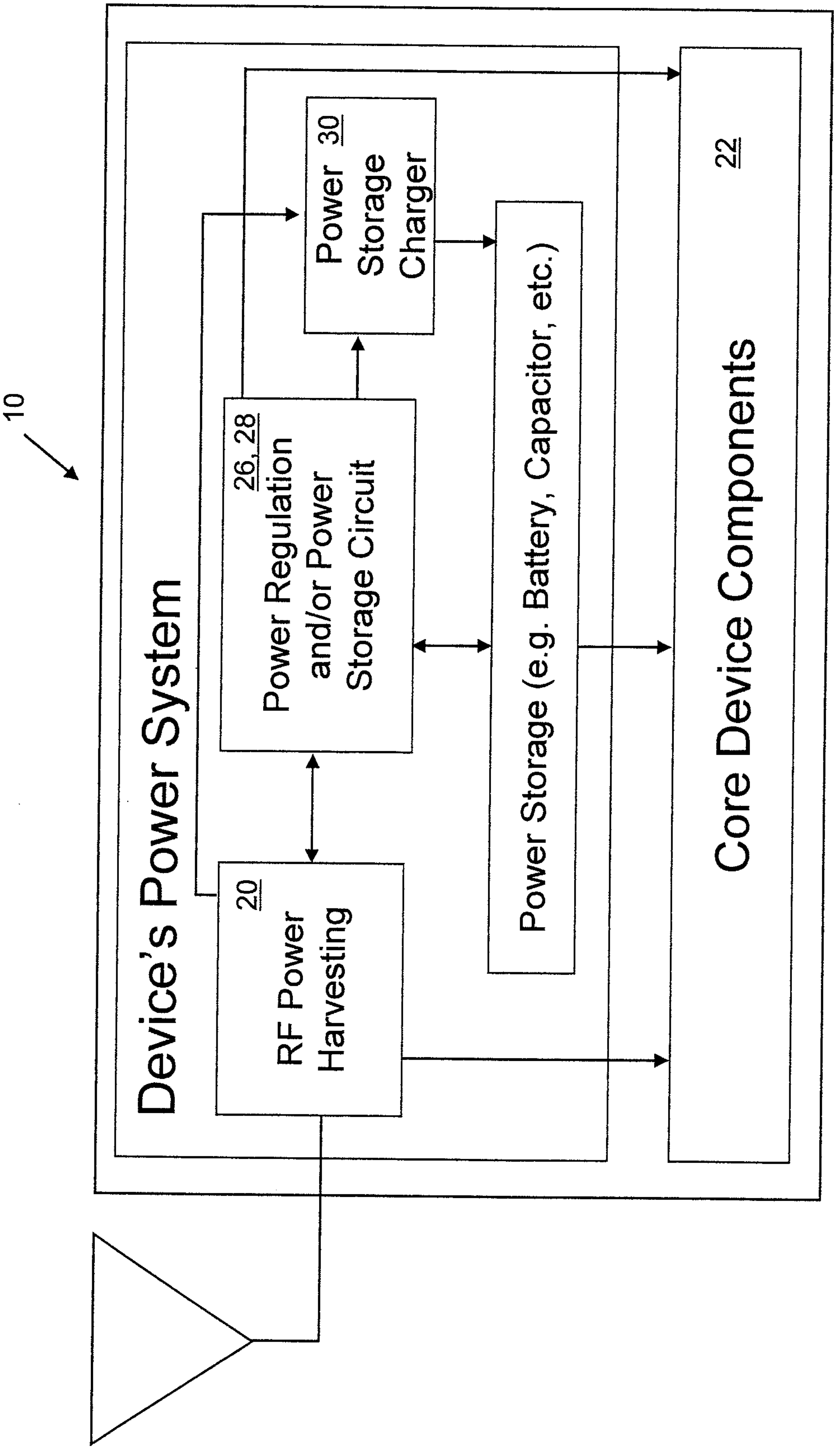
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage
Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 51

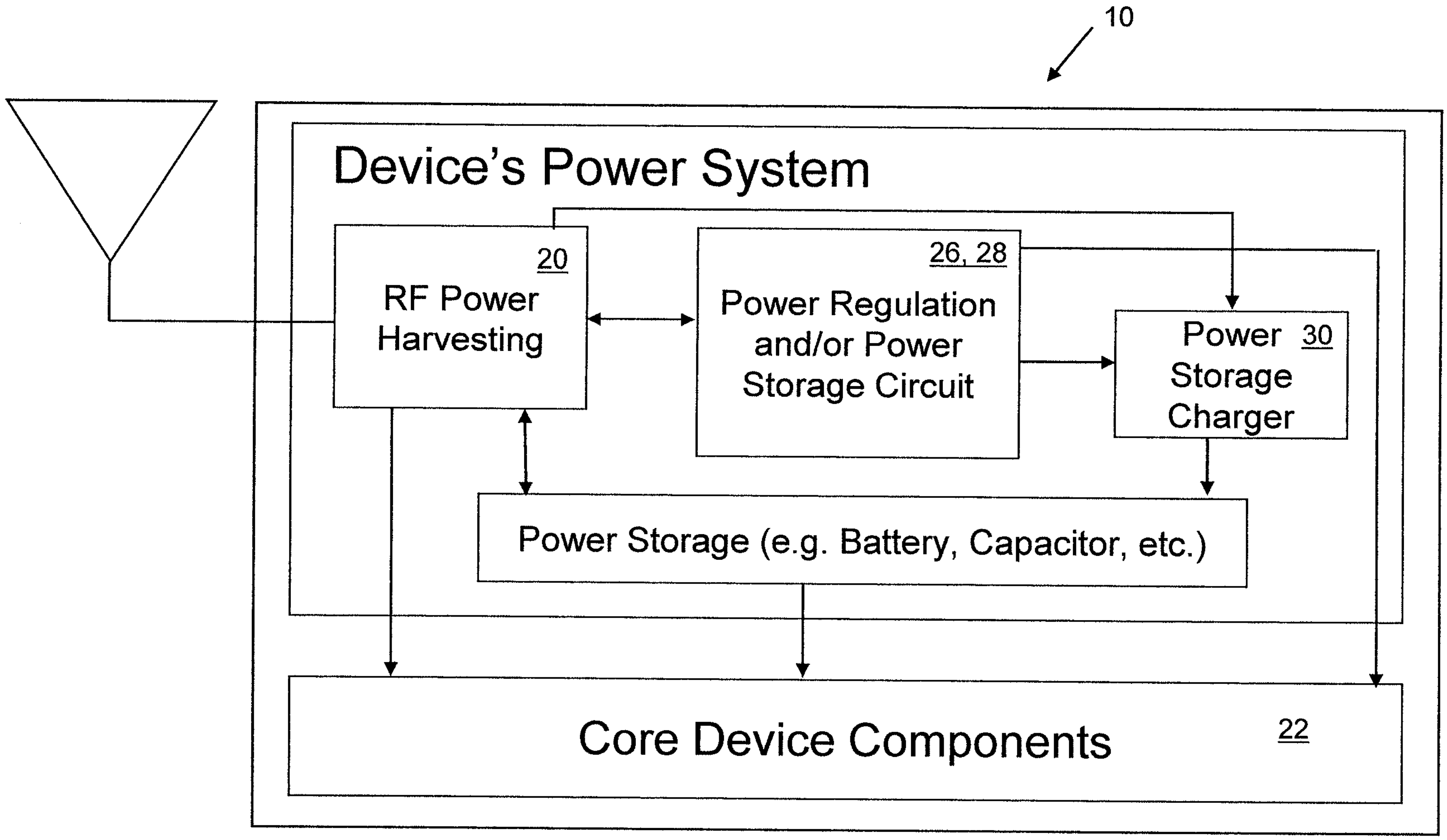
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30 Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 52

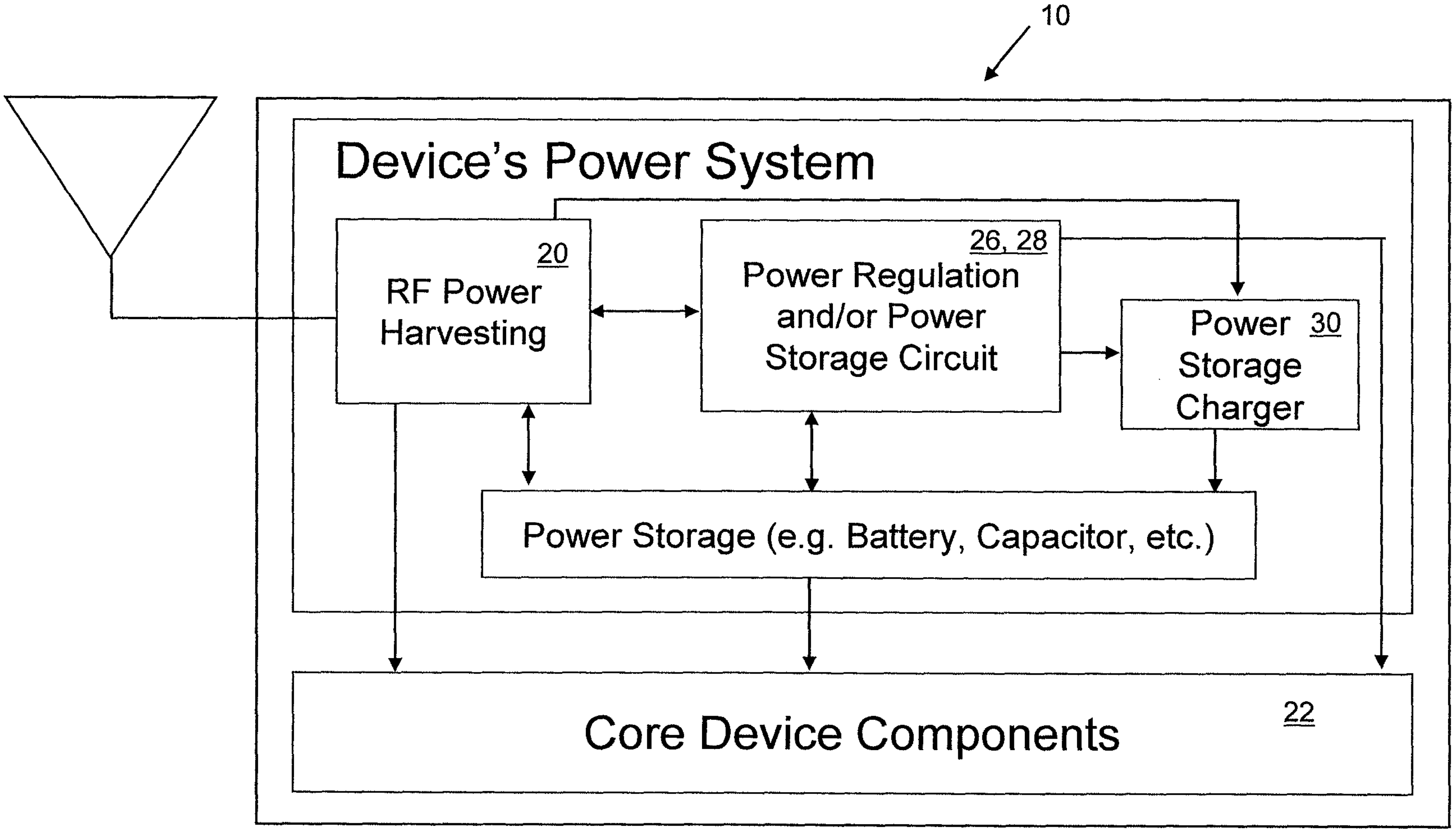
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 53

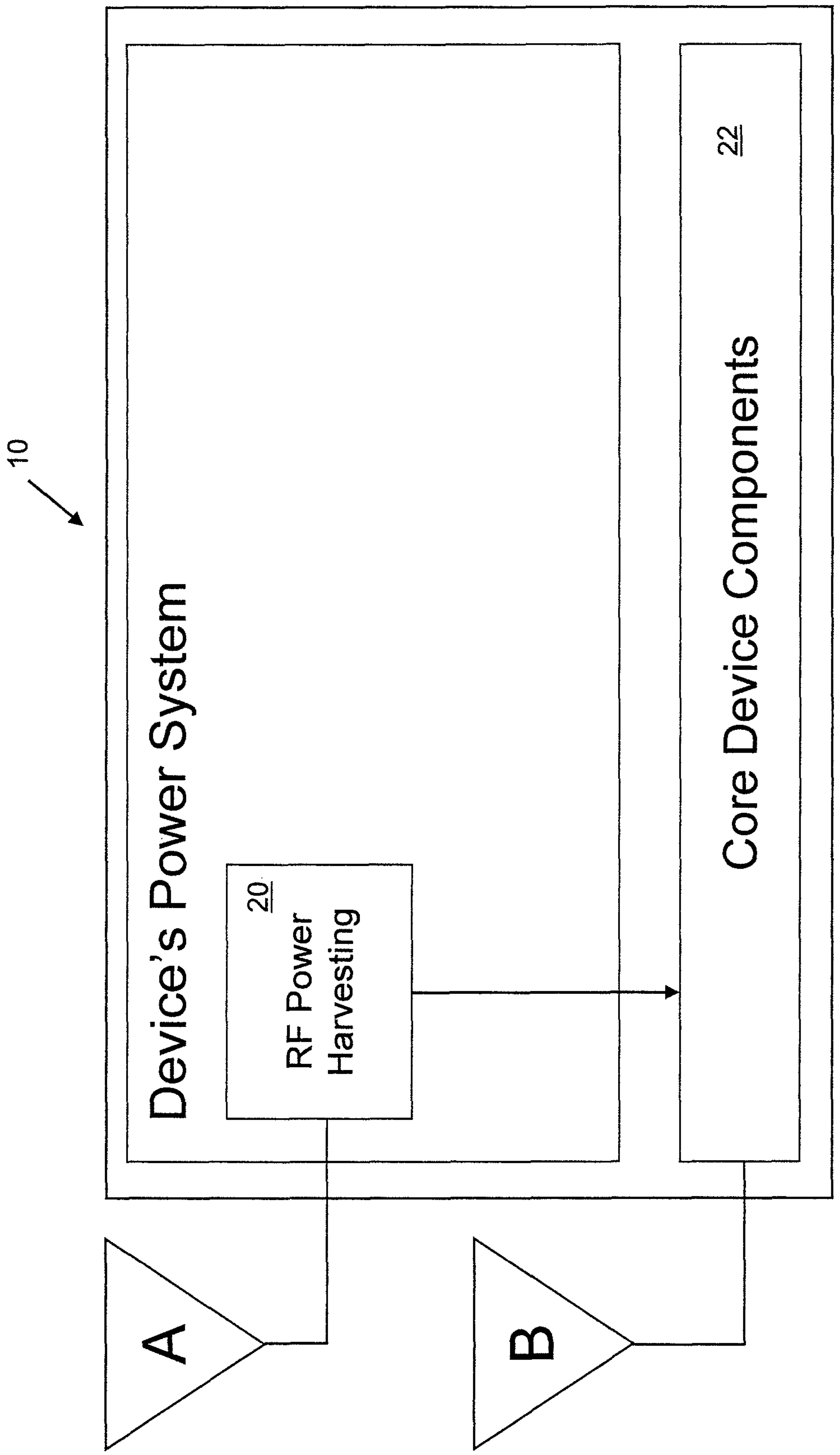
10
Device's Power System
RF Power Harvesting
20
Core Device Components
22
A
B

Fig. 54

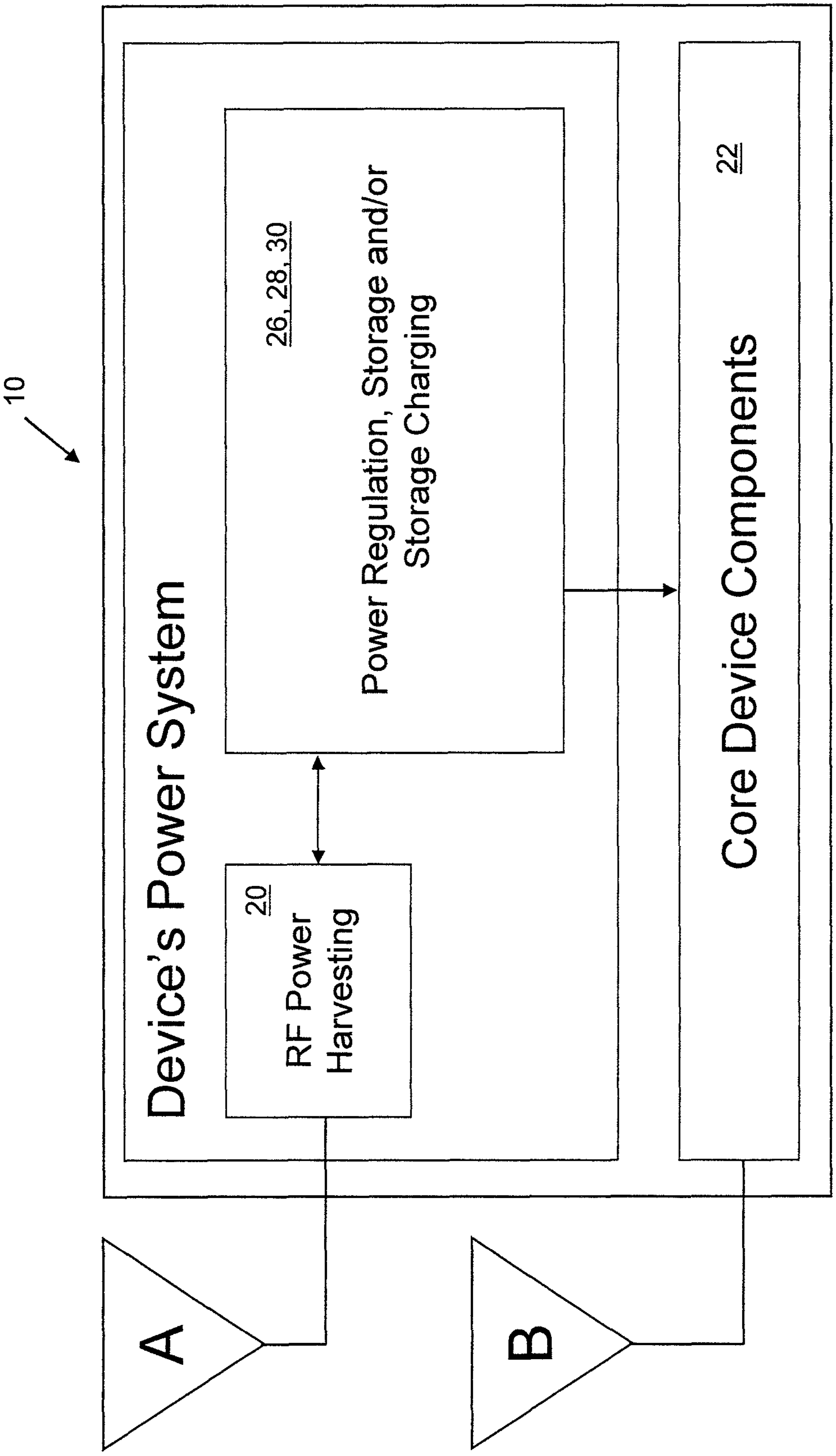
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
22
Core Device Components
A
B

Fig. 55

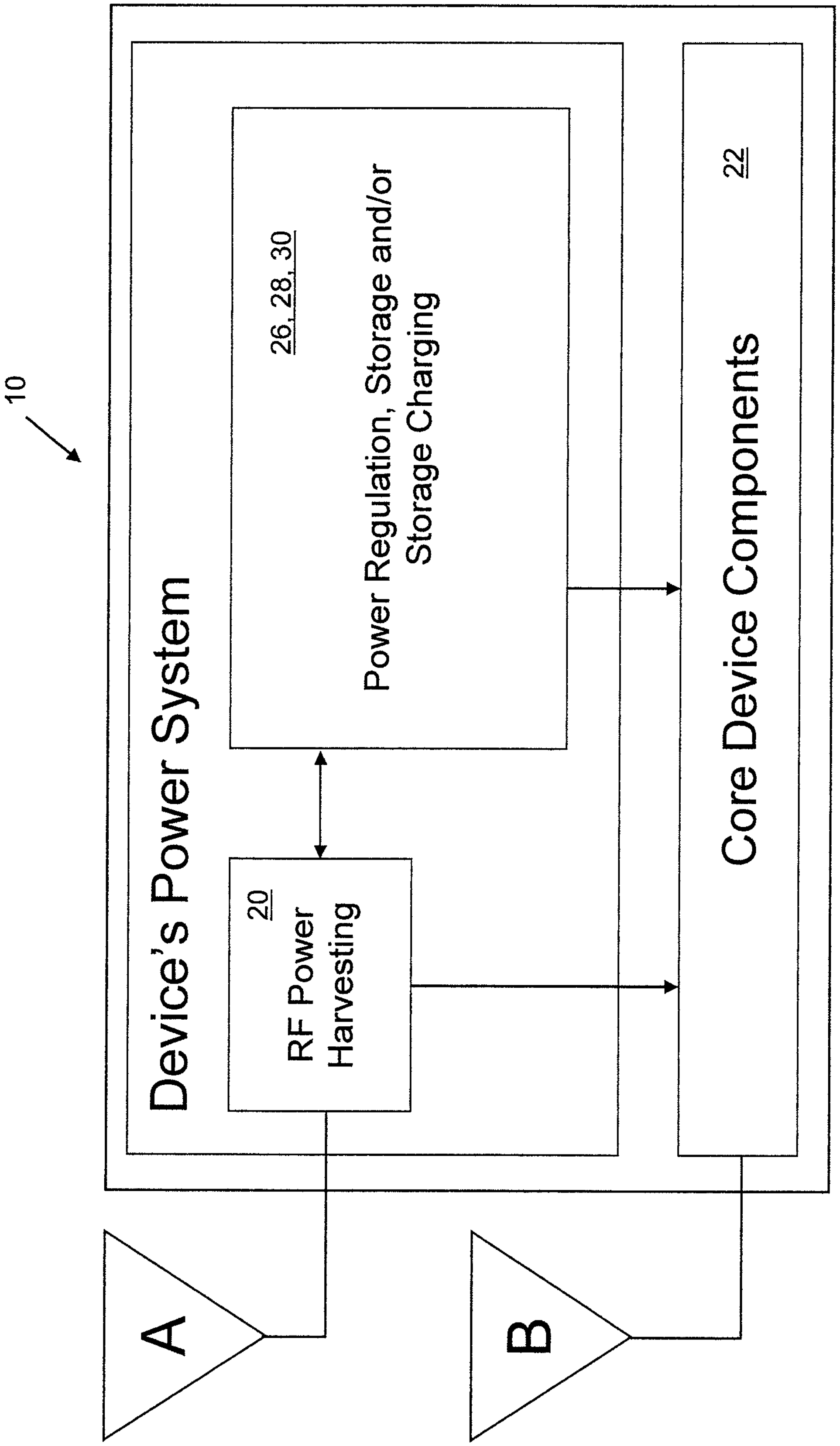
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
A
B

Fig. 56

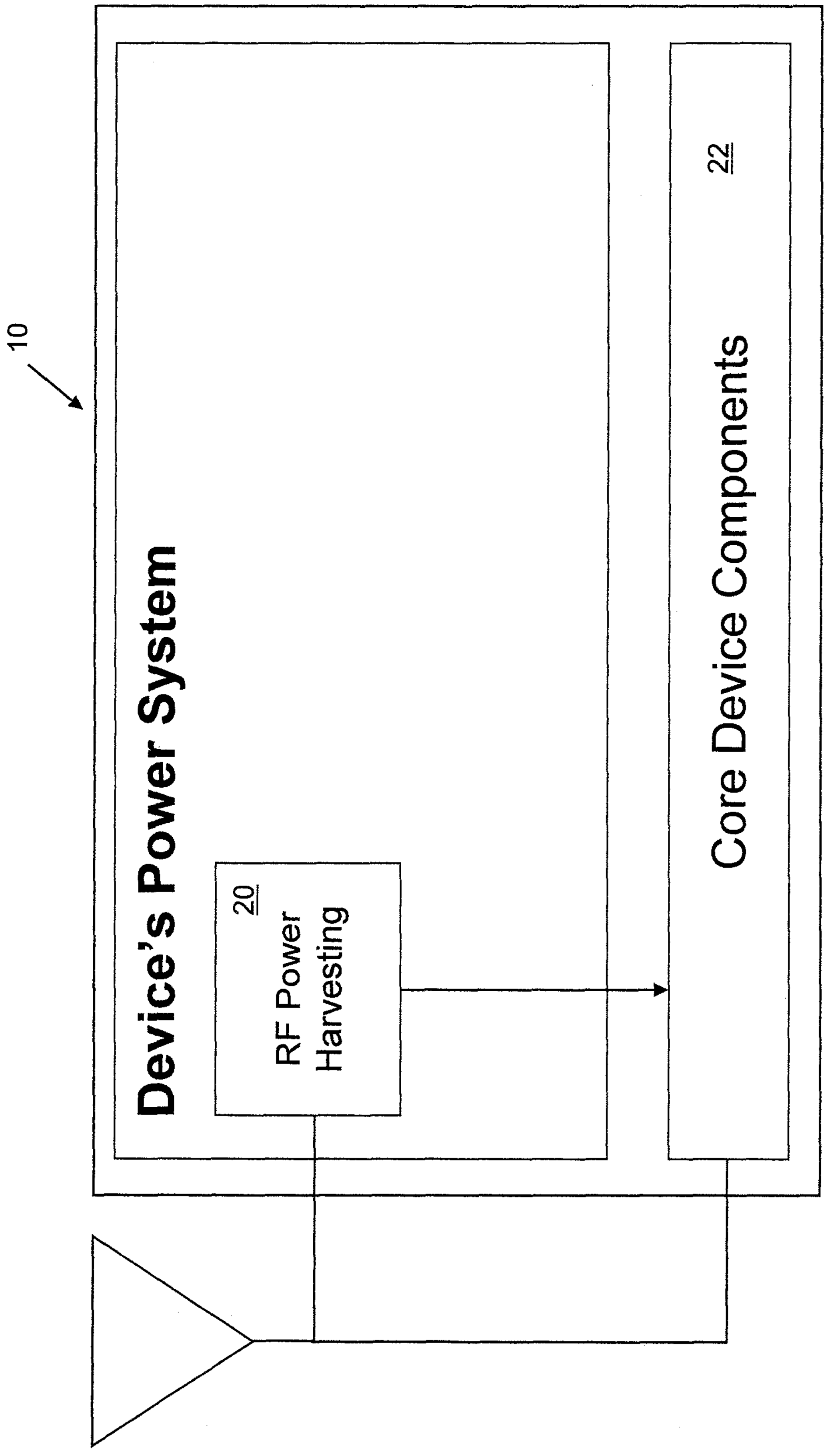
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22

Fig. 57

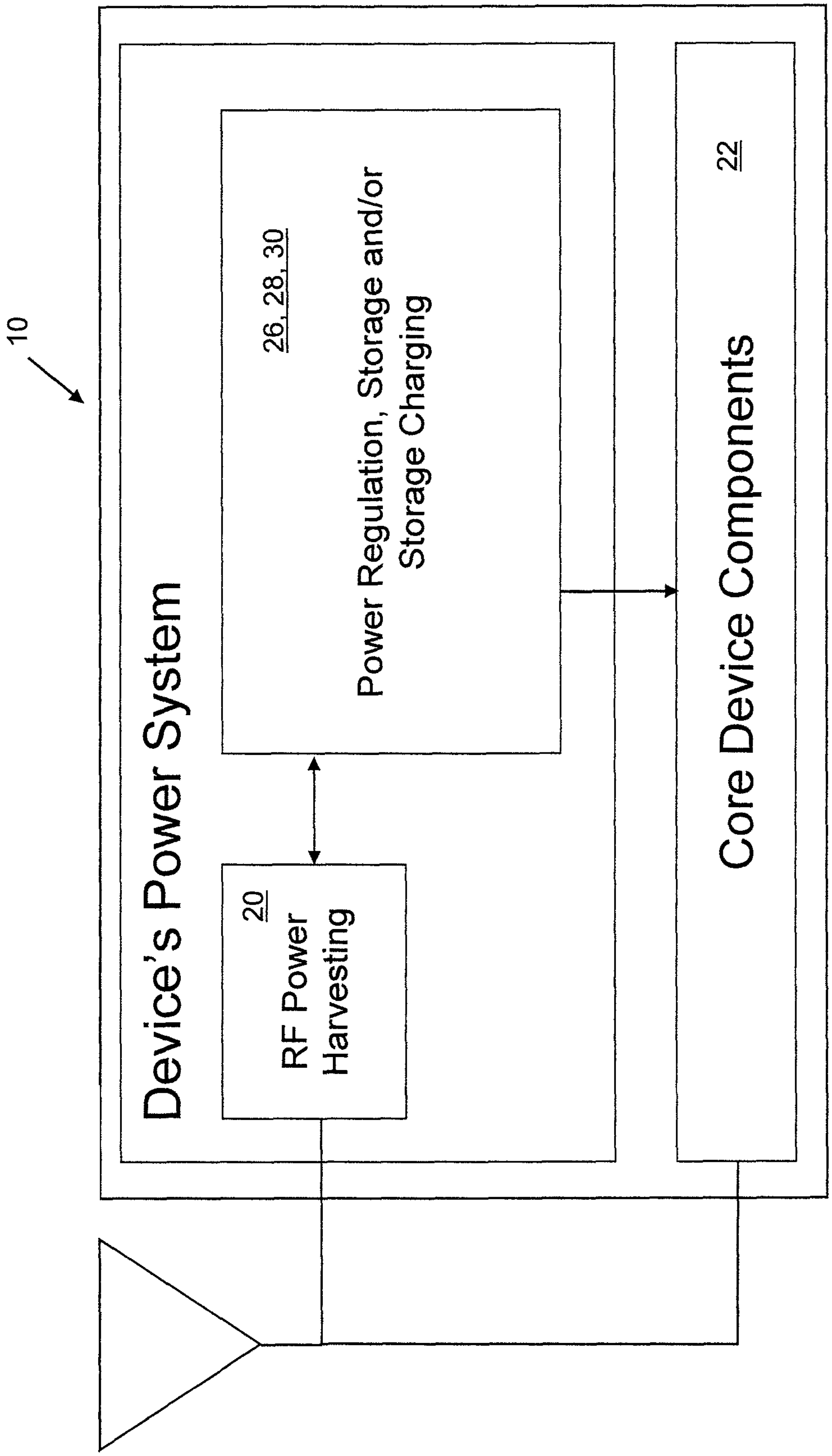
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22

Fig. 58

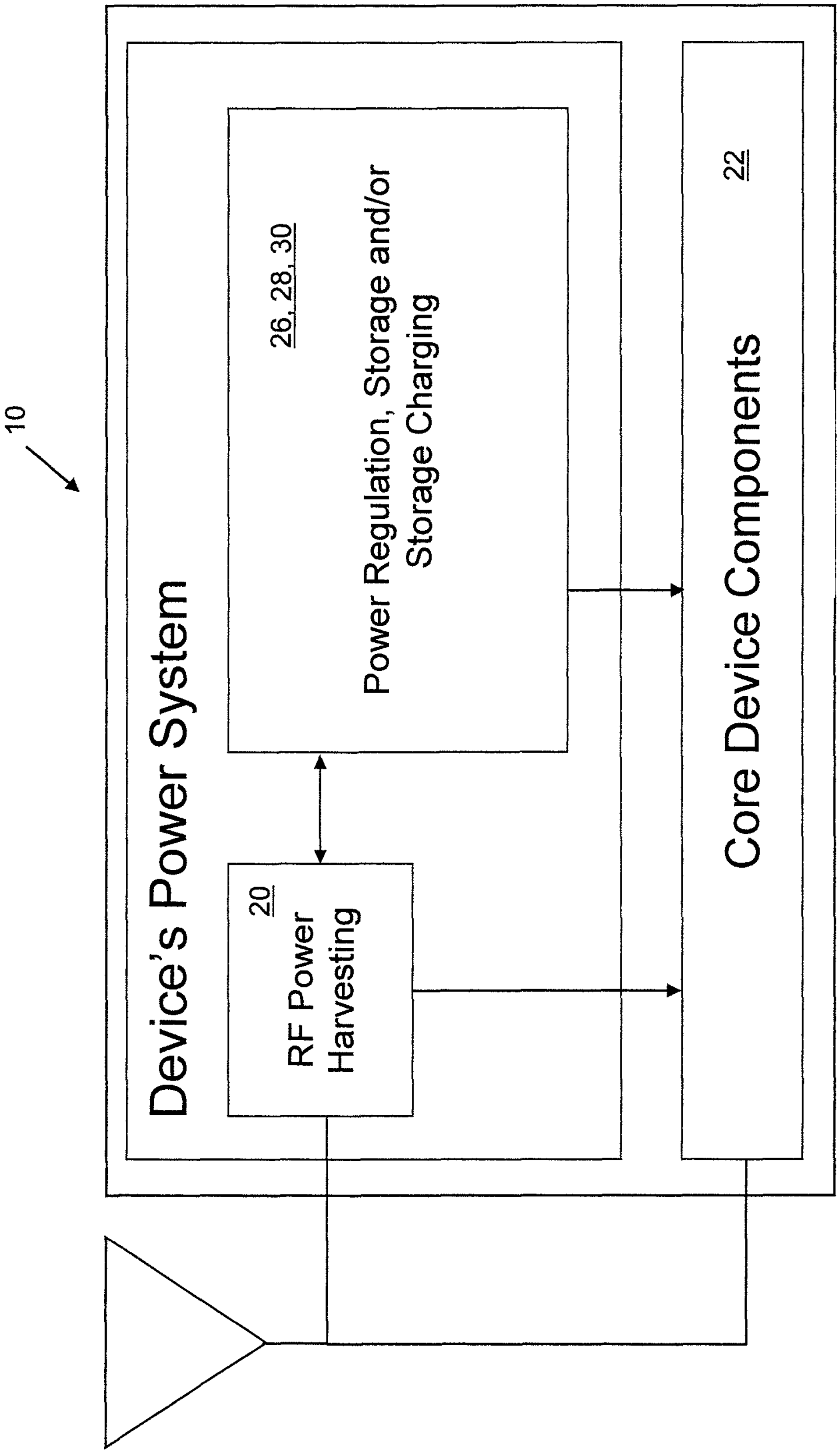
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22

Fig. 59

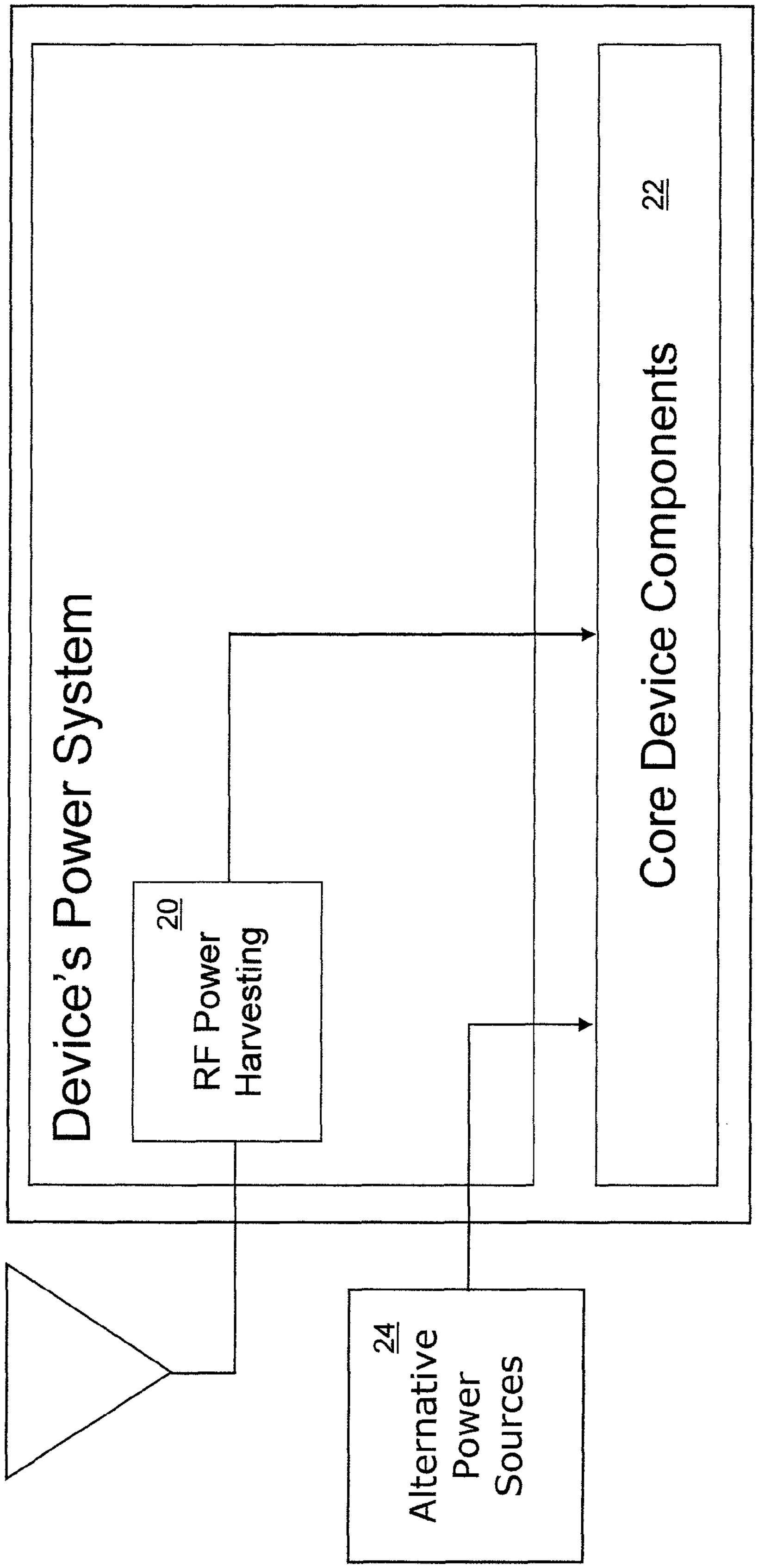
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 60

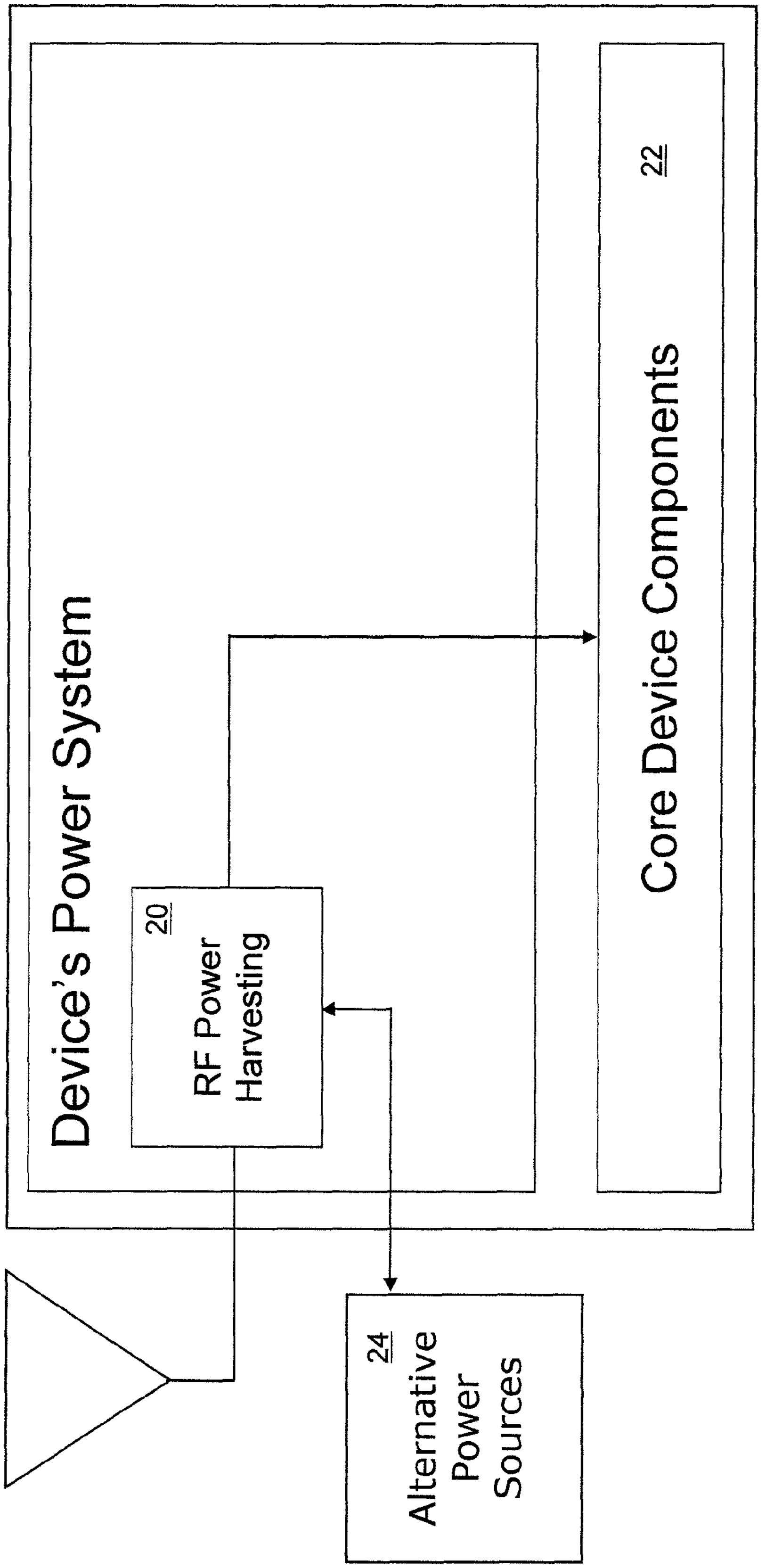
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 61

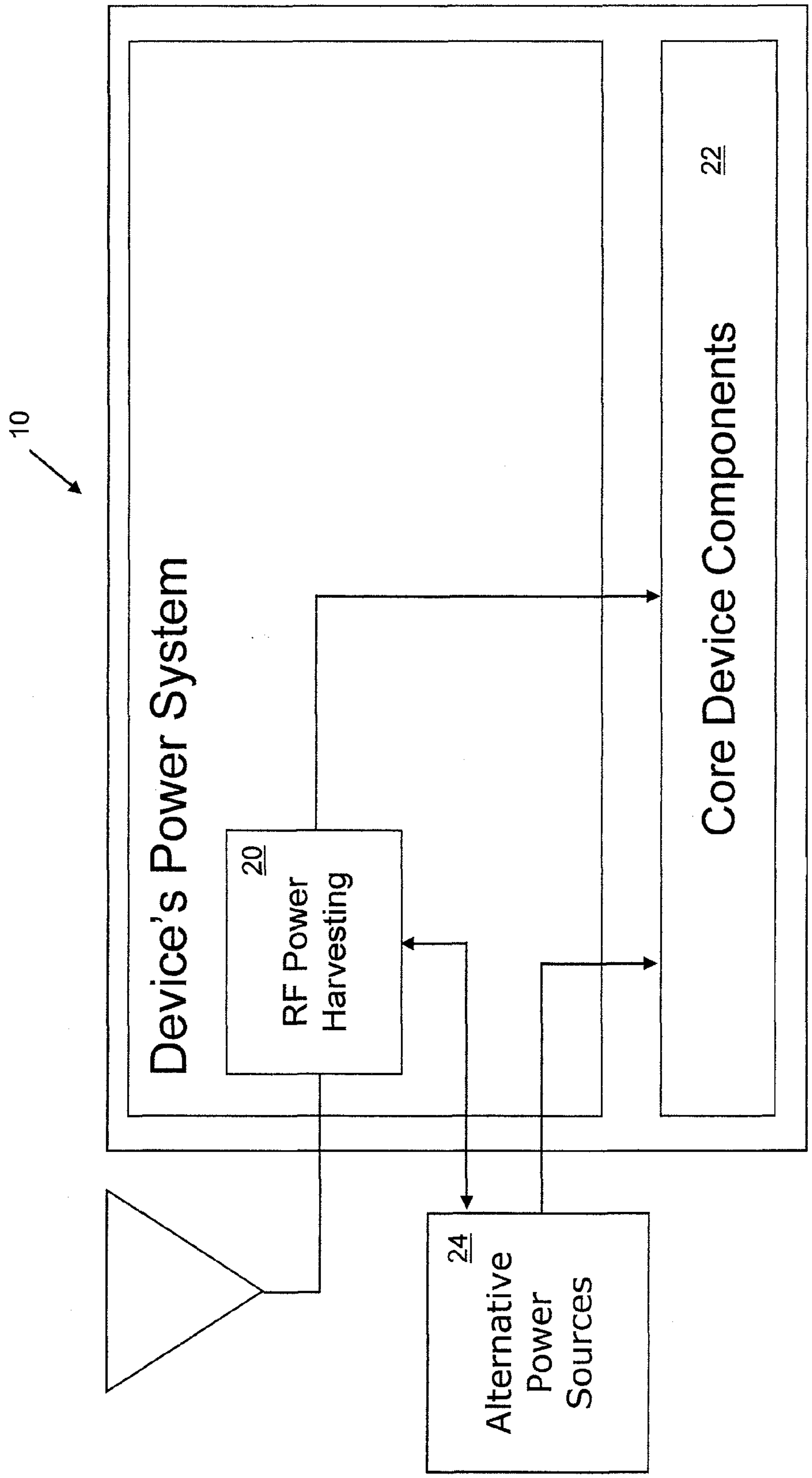
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 62

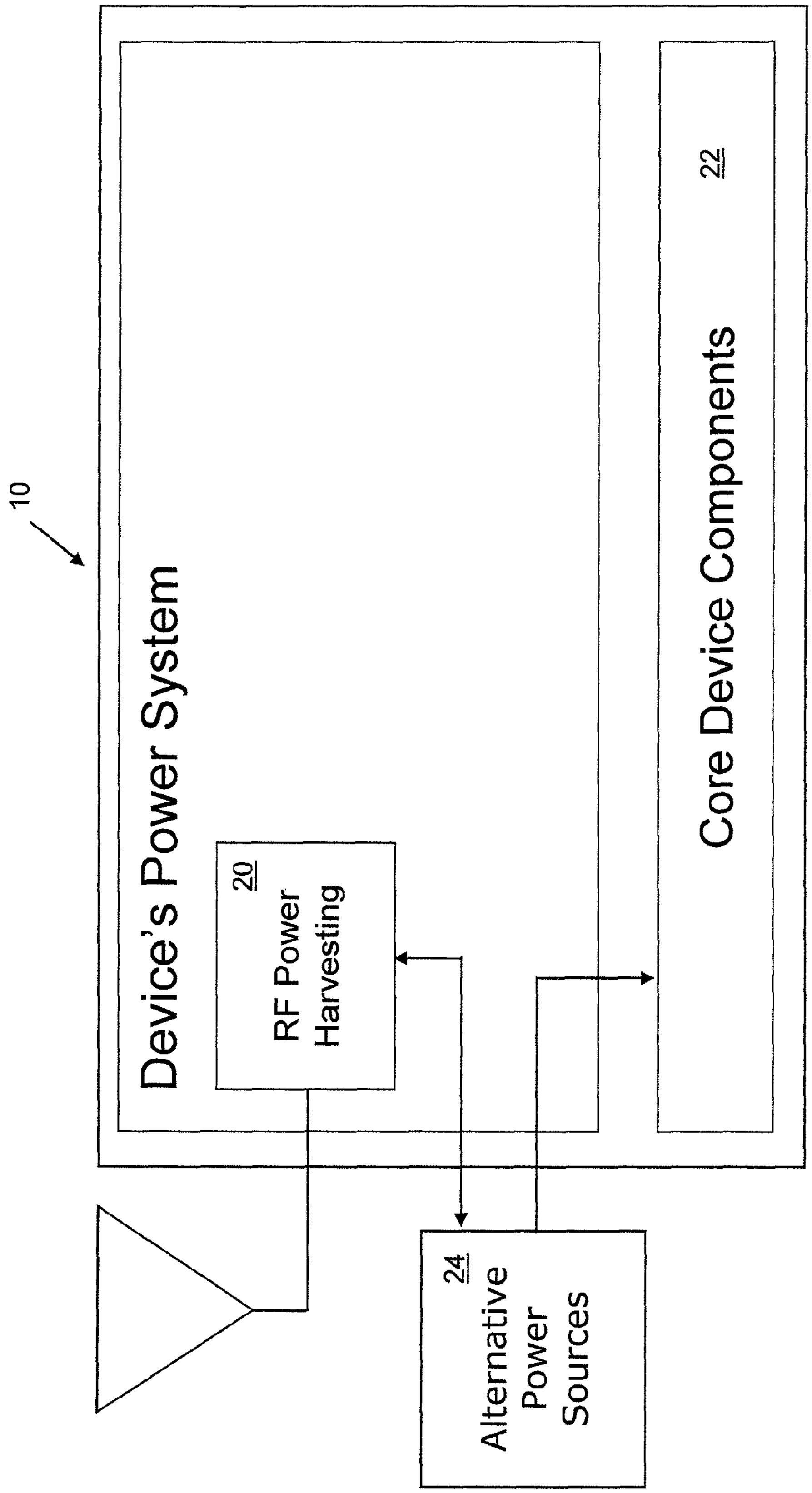
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 63

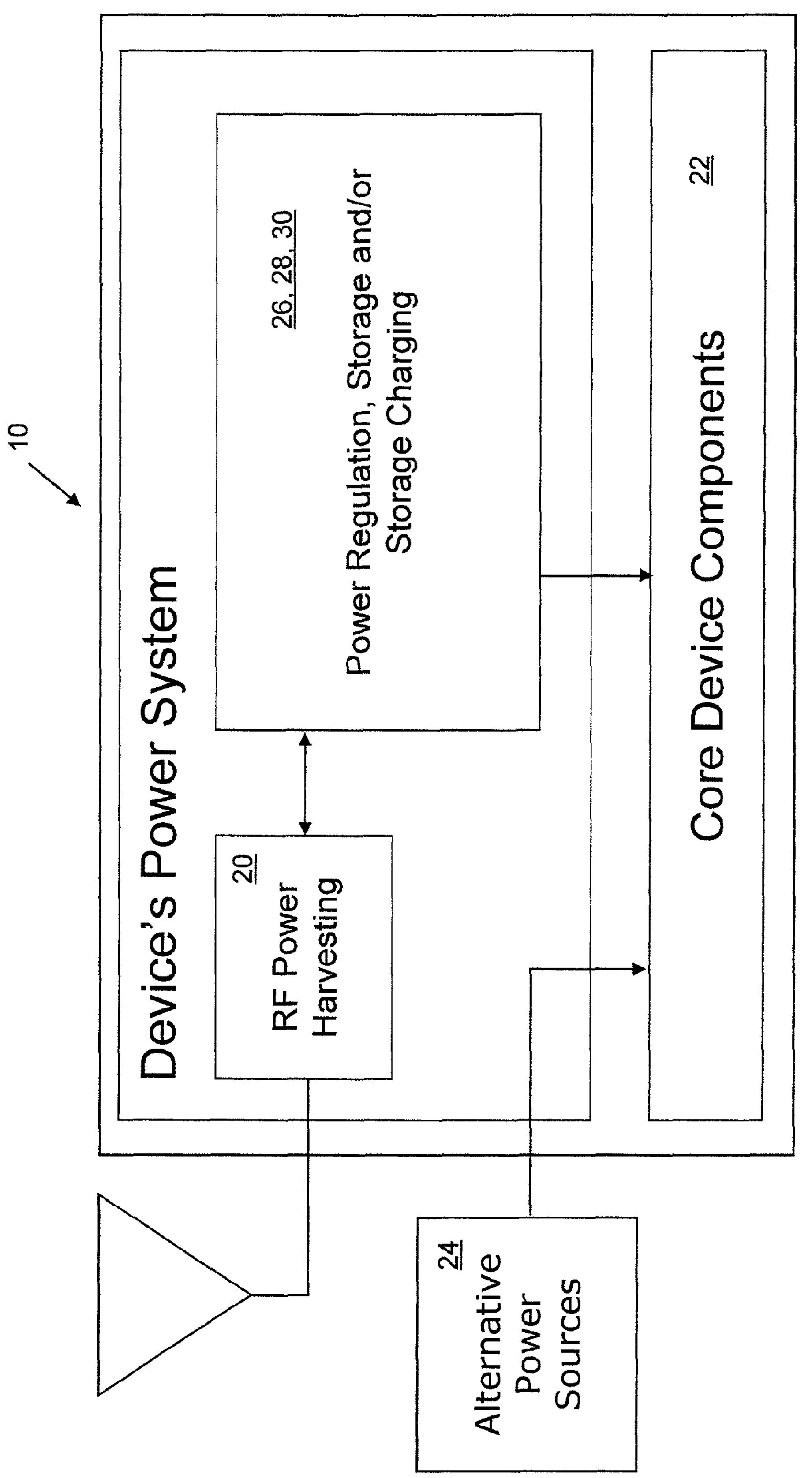
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
22
Core Device Components
24
Alternative Power Sources

Fig. 64

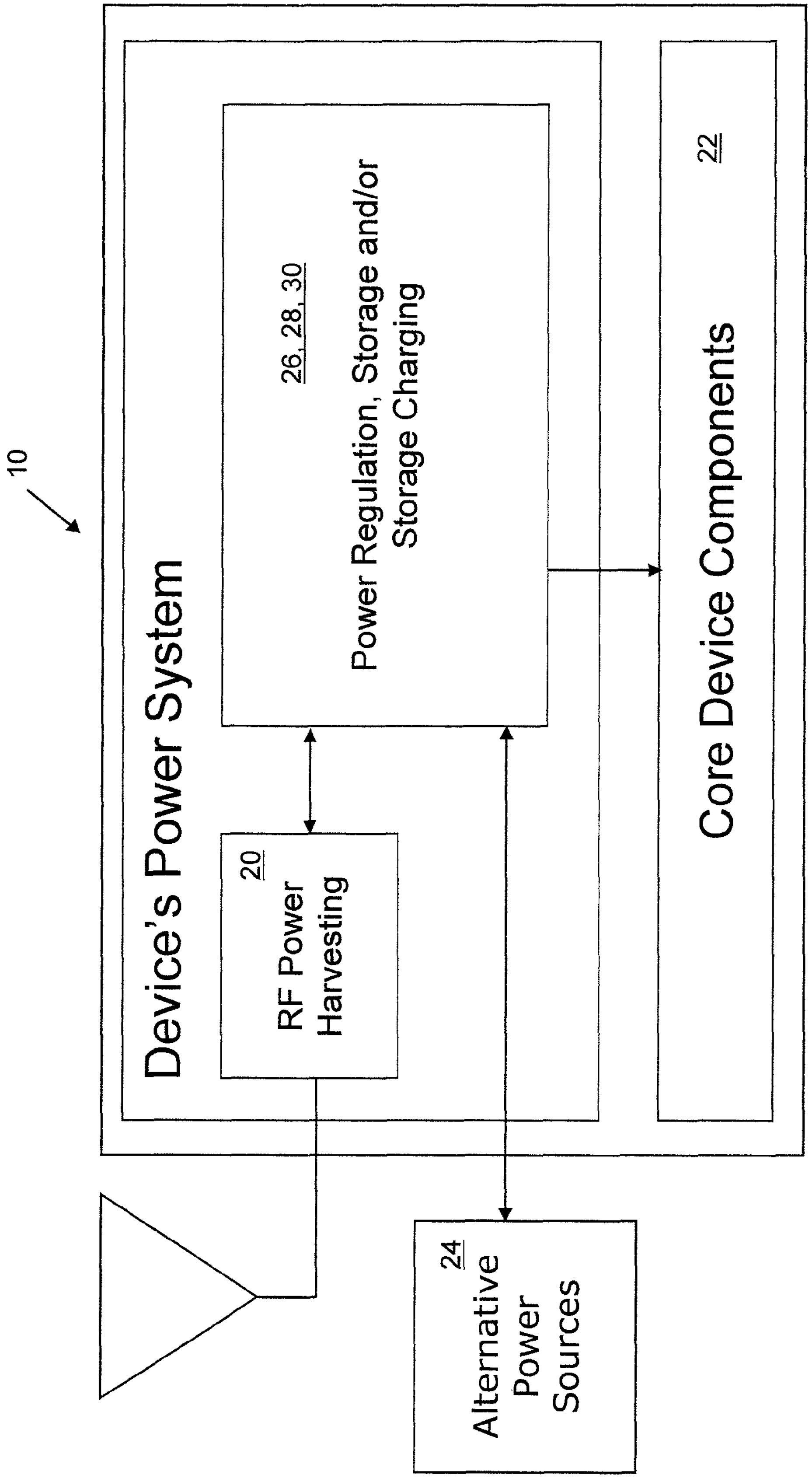
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 65

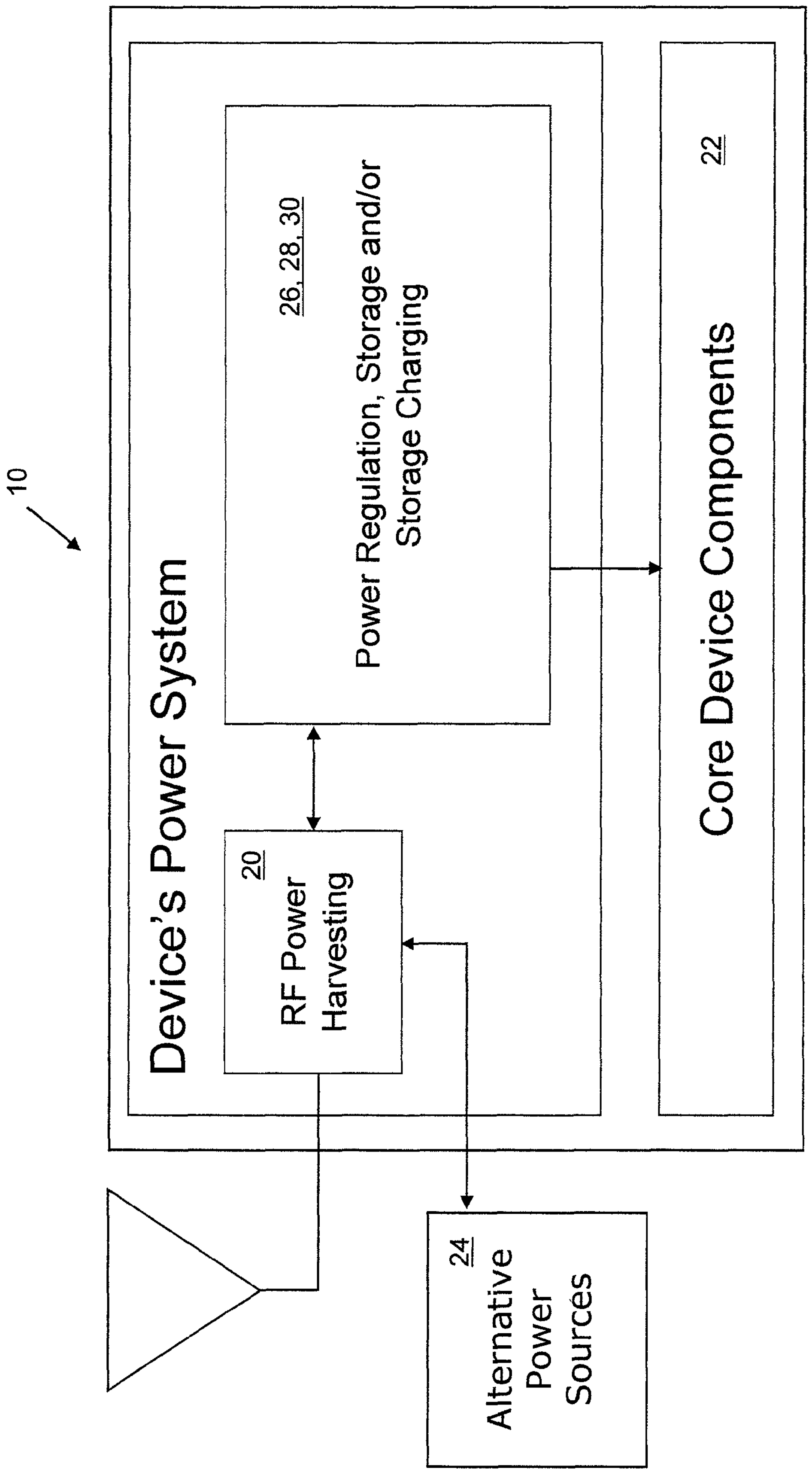
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 66

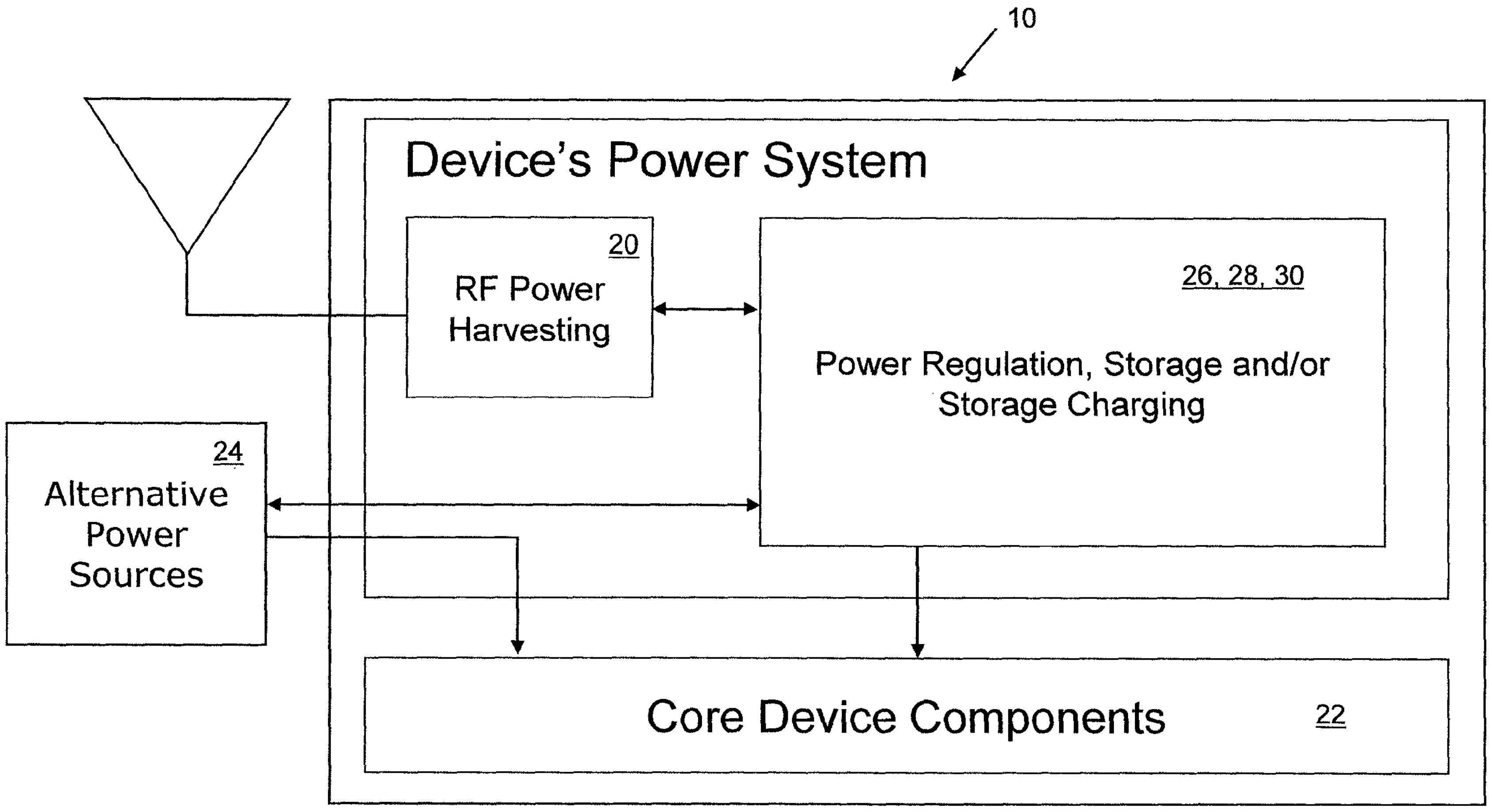
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 67

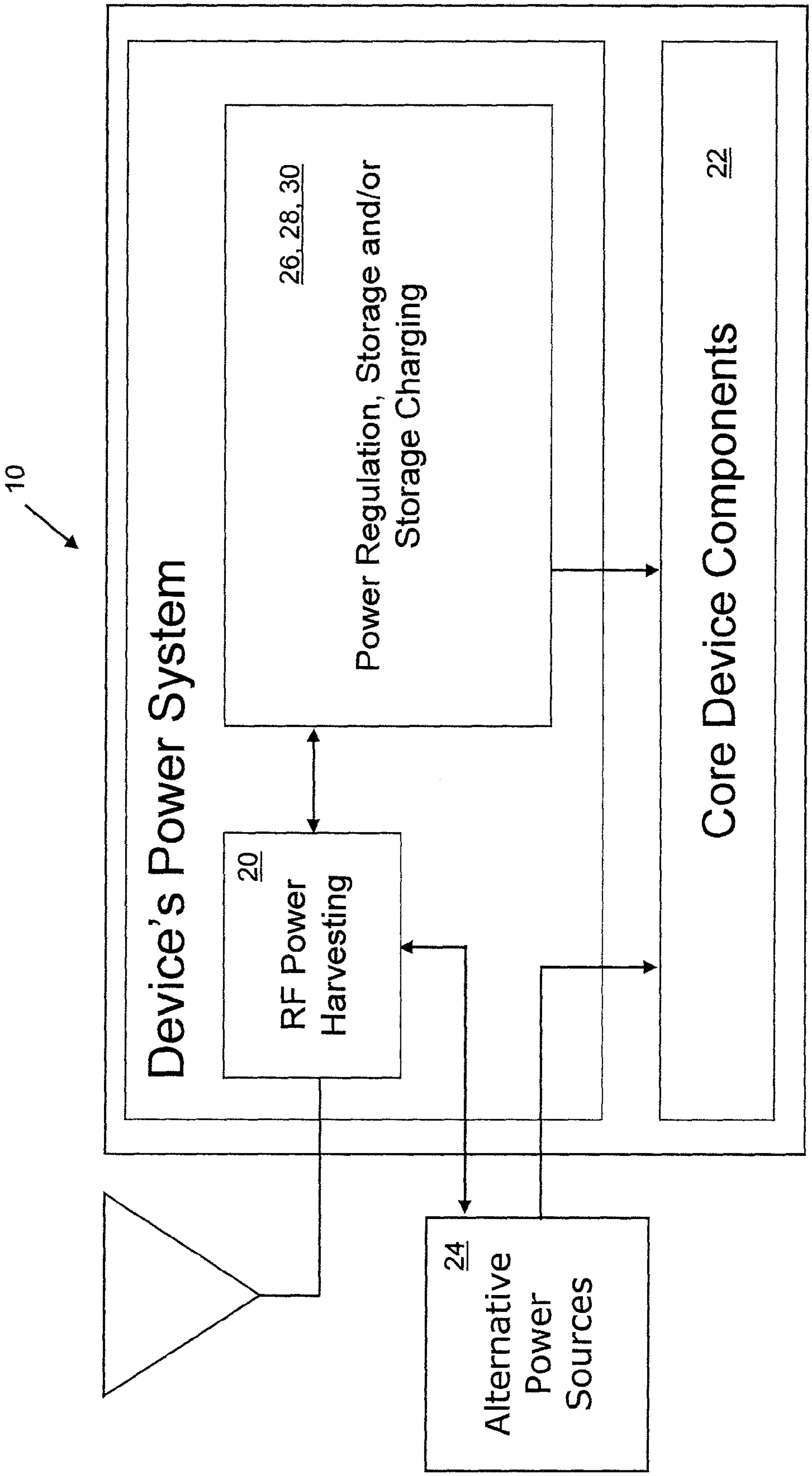
10
Device's Power System
Power Regulation, Storage and/or Storage Charging
26, 28, 30
RF Power Harvesting
20
Core Device Components
22
Alternative Power Sources
24

Fig. 68

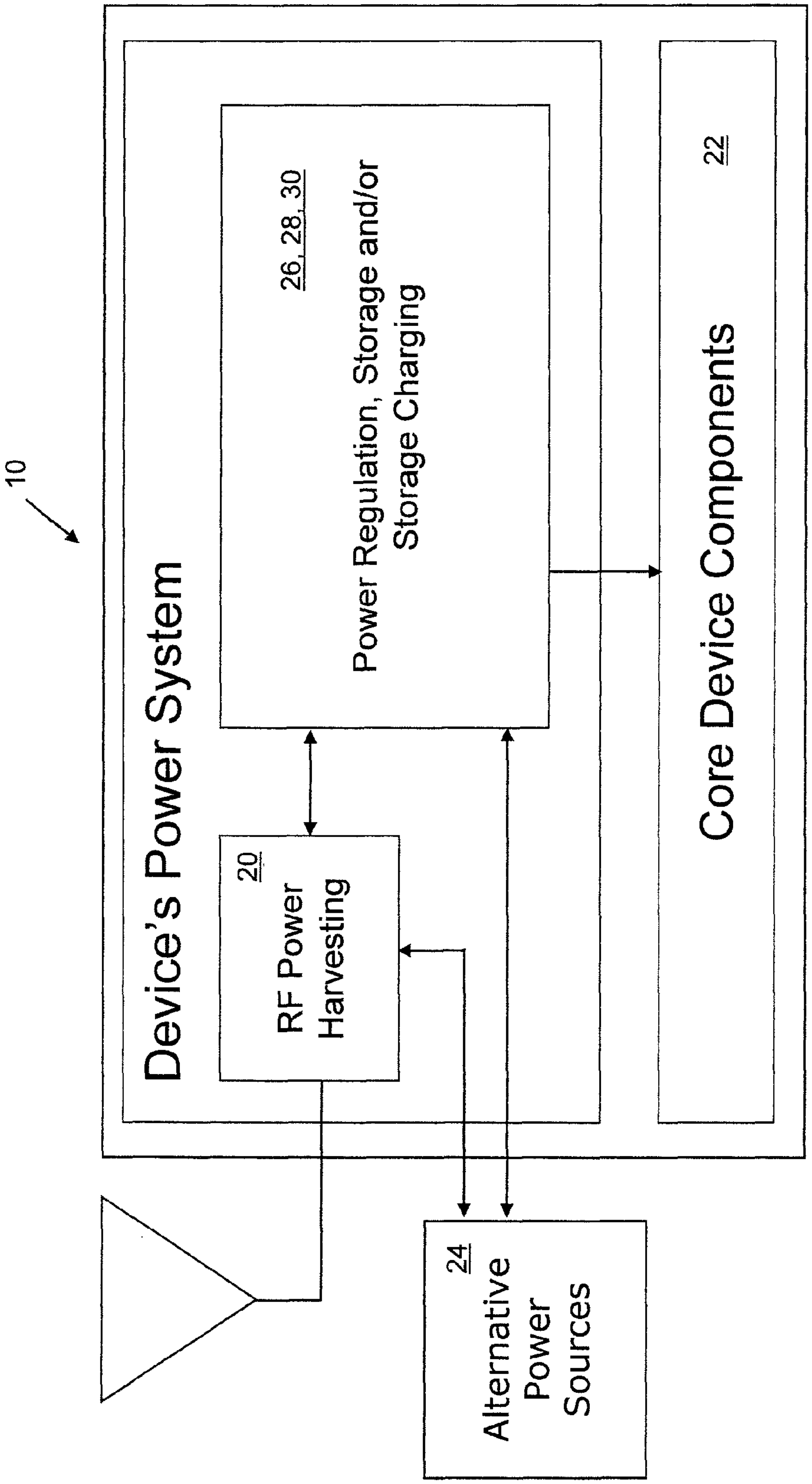
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 69

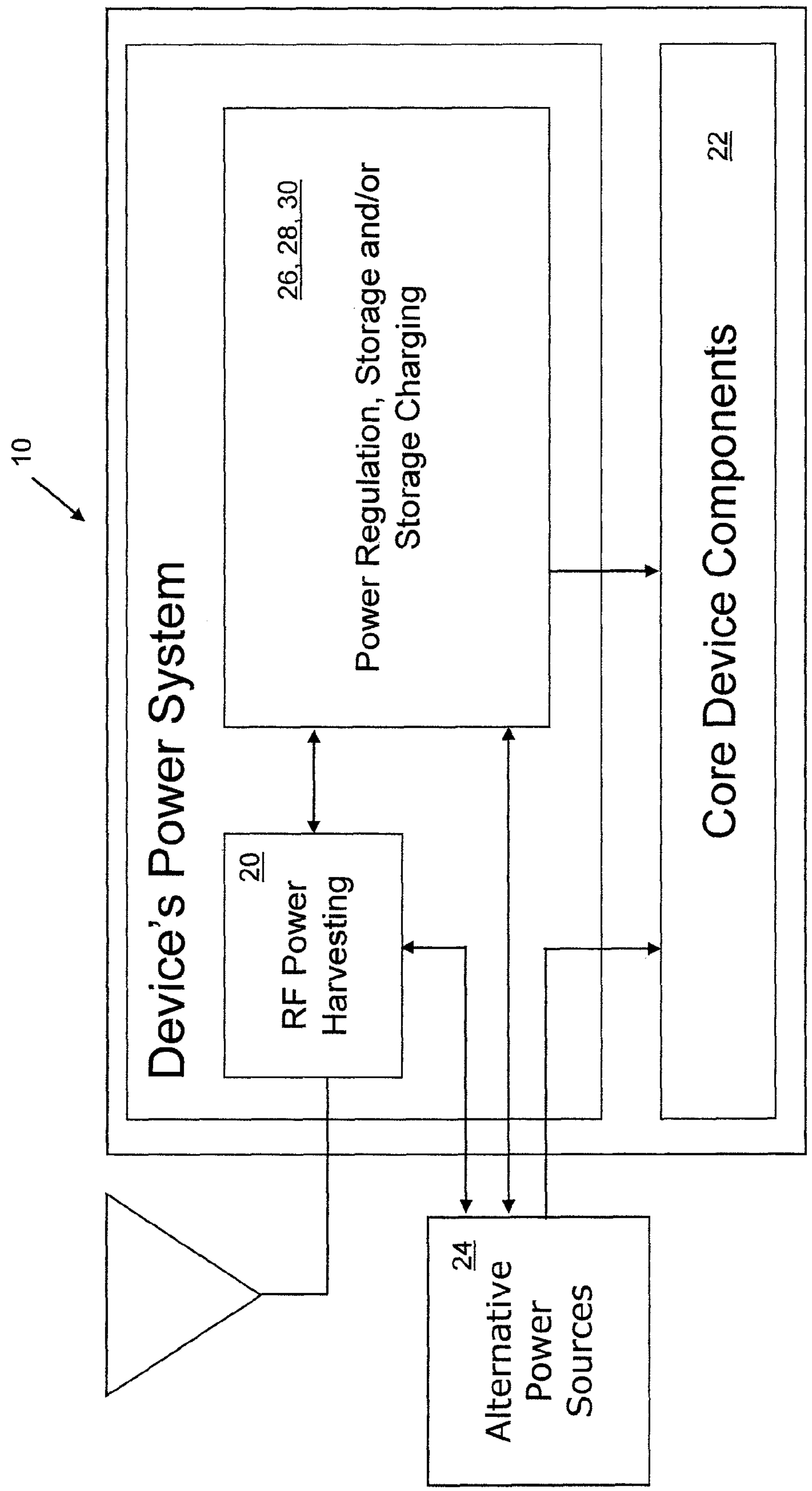
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 70

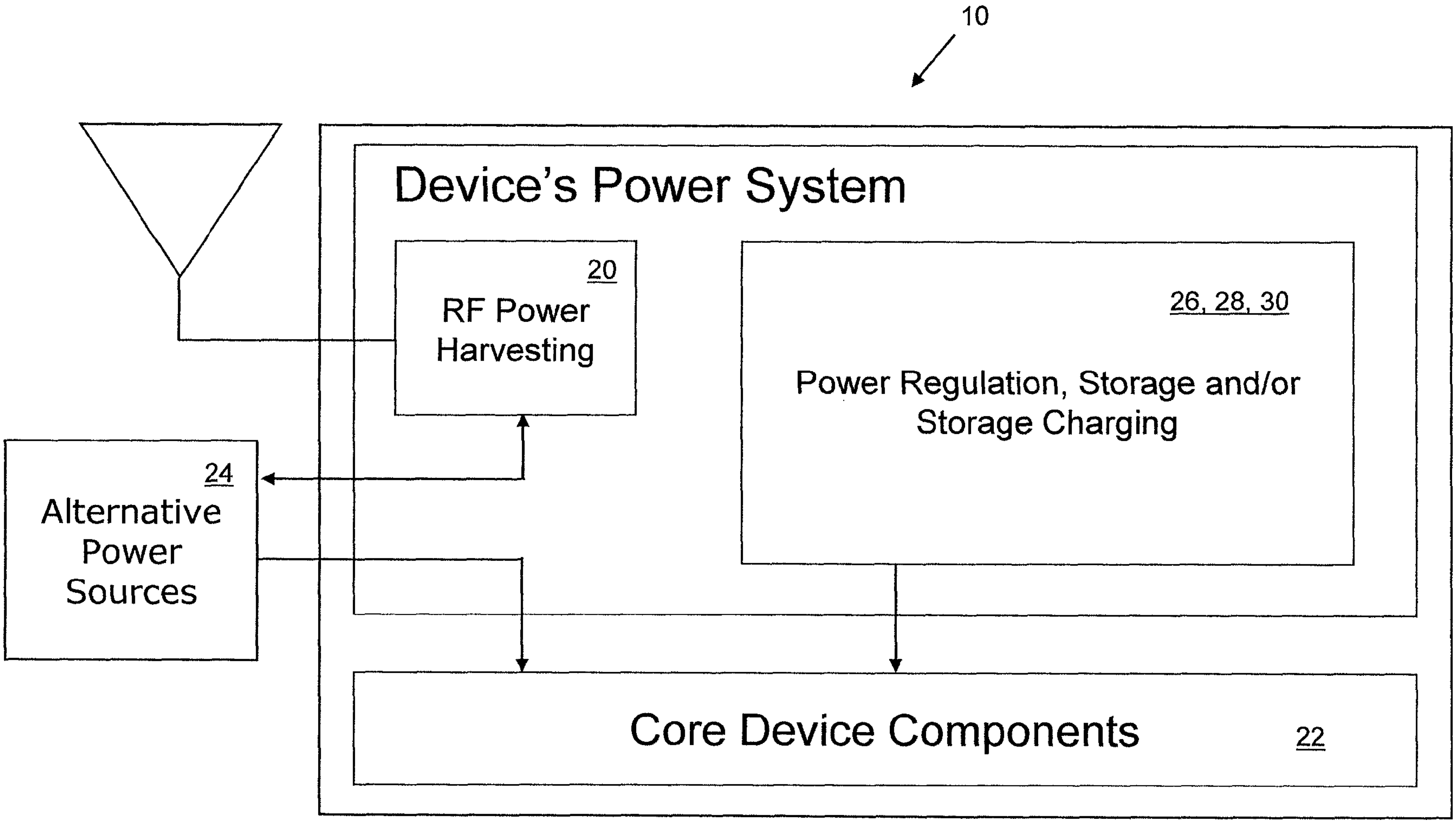
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 71

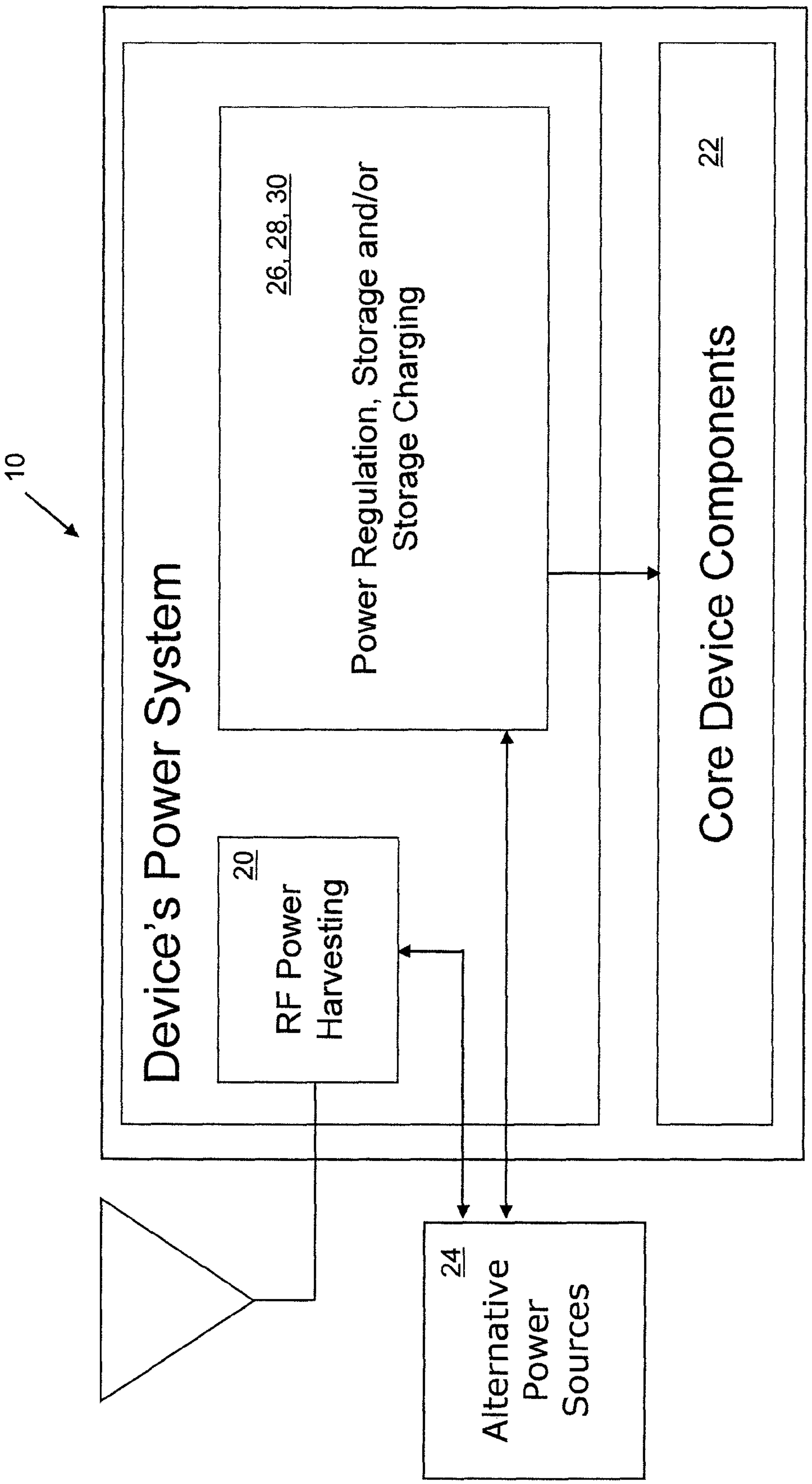
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
22
Core Device Components
24
Alternative Power Sources

Fig. 72

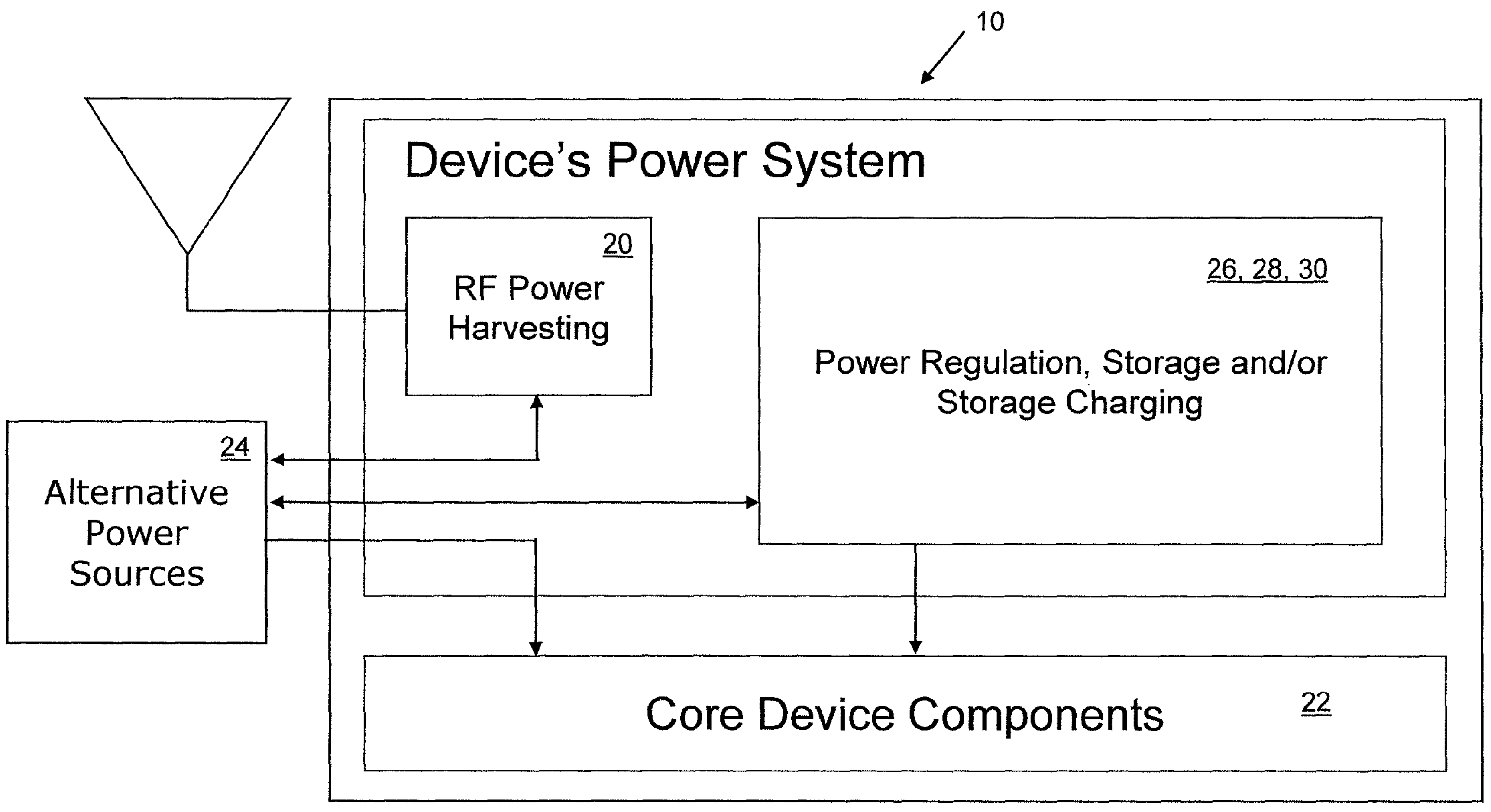
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 73

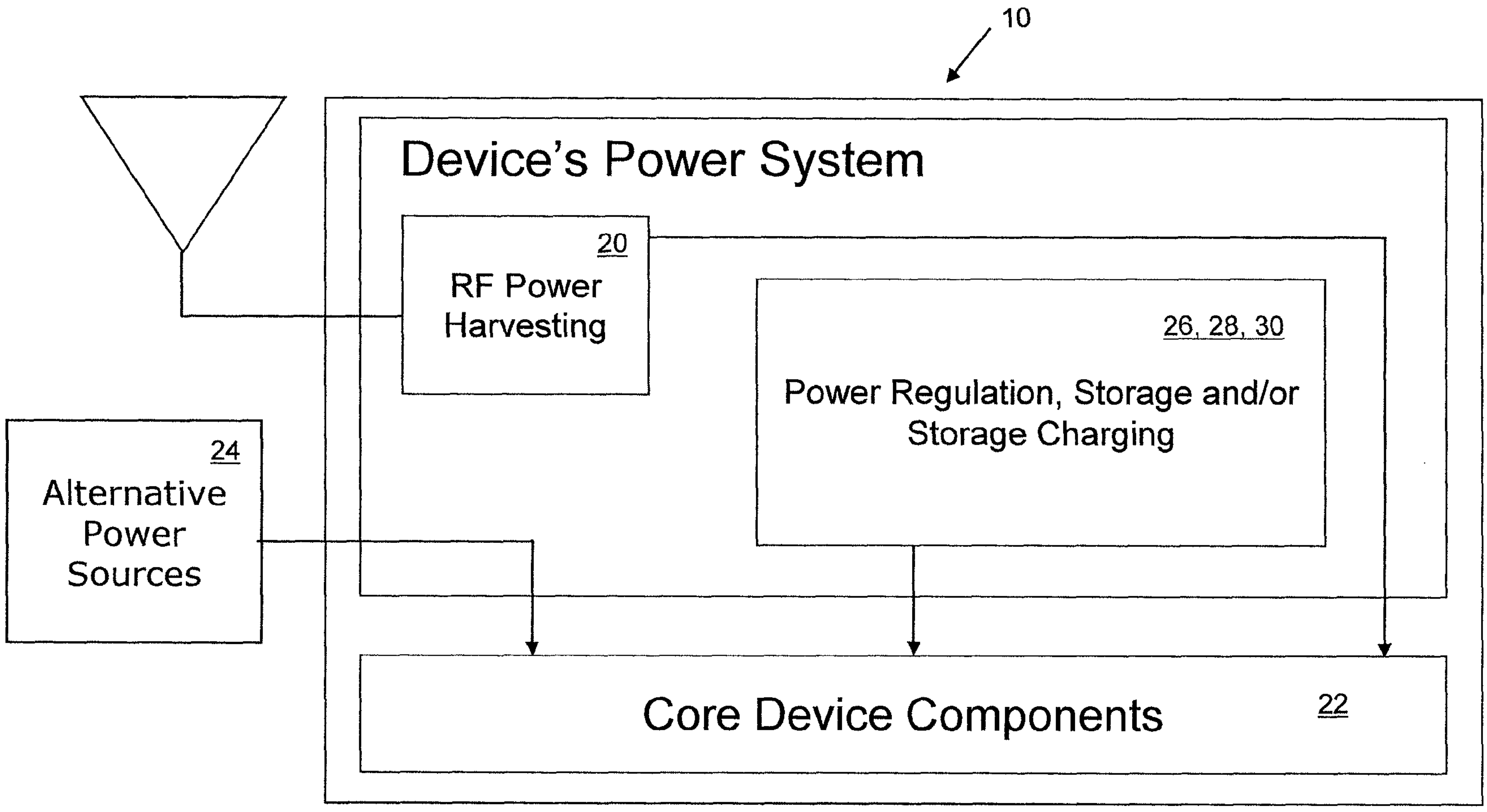
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 74

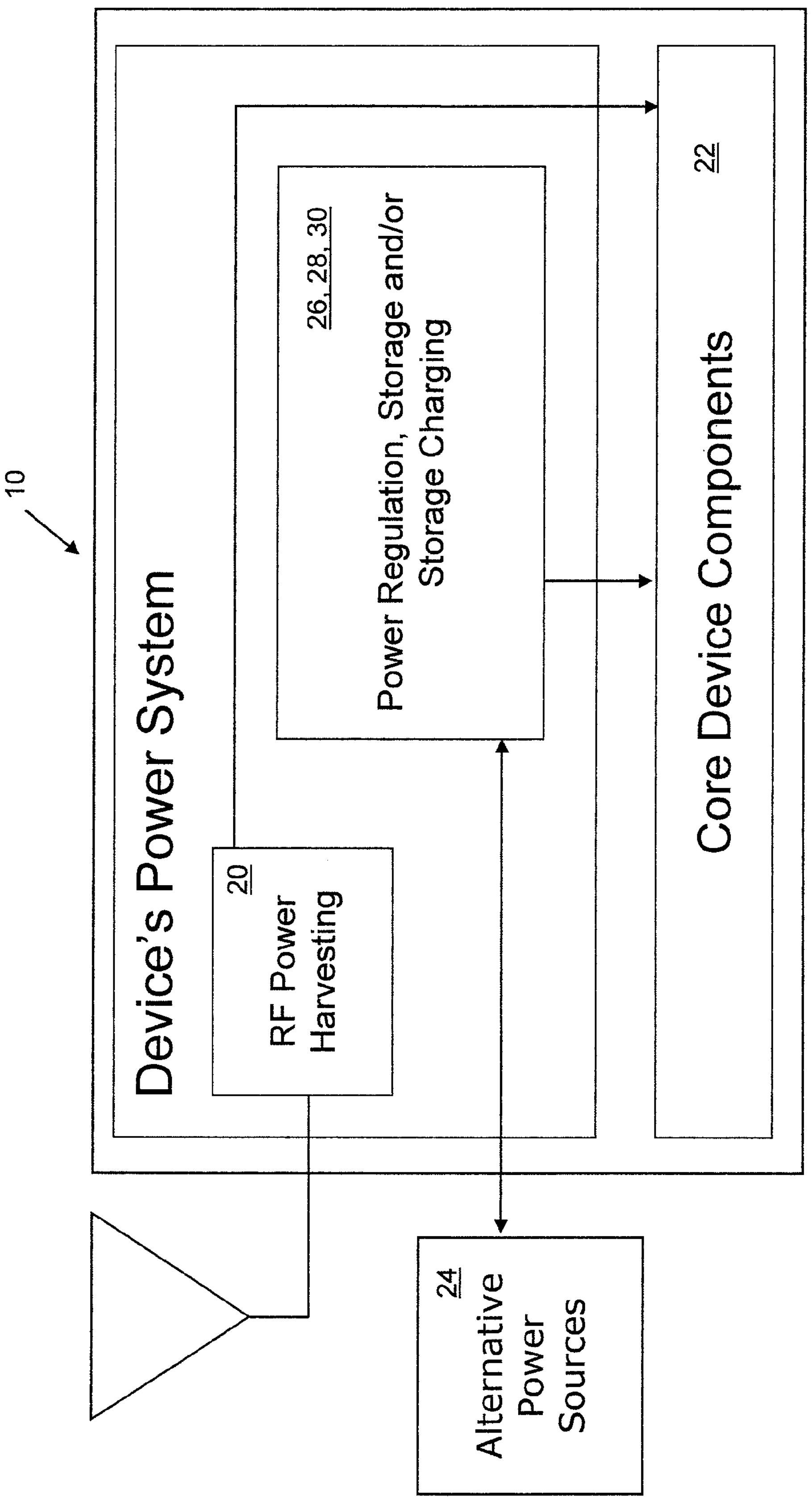
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 75

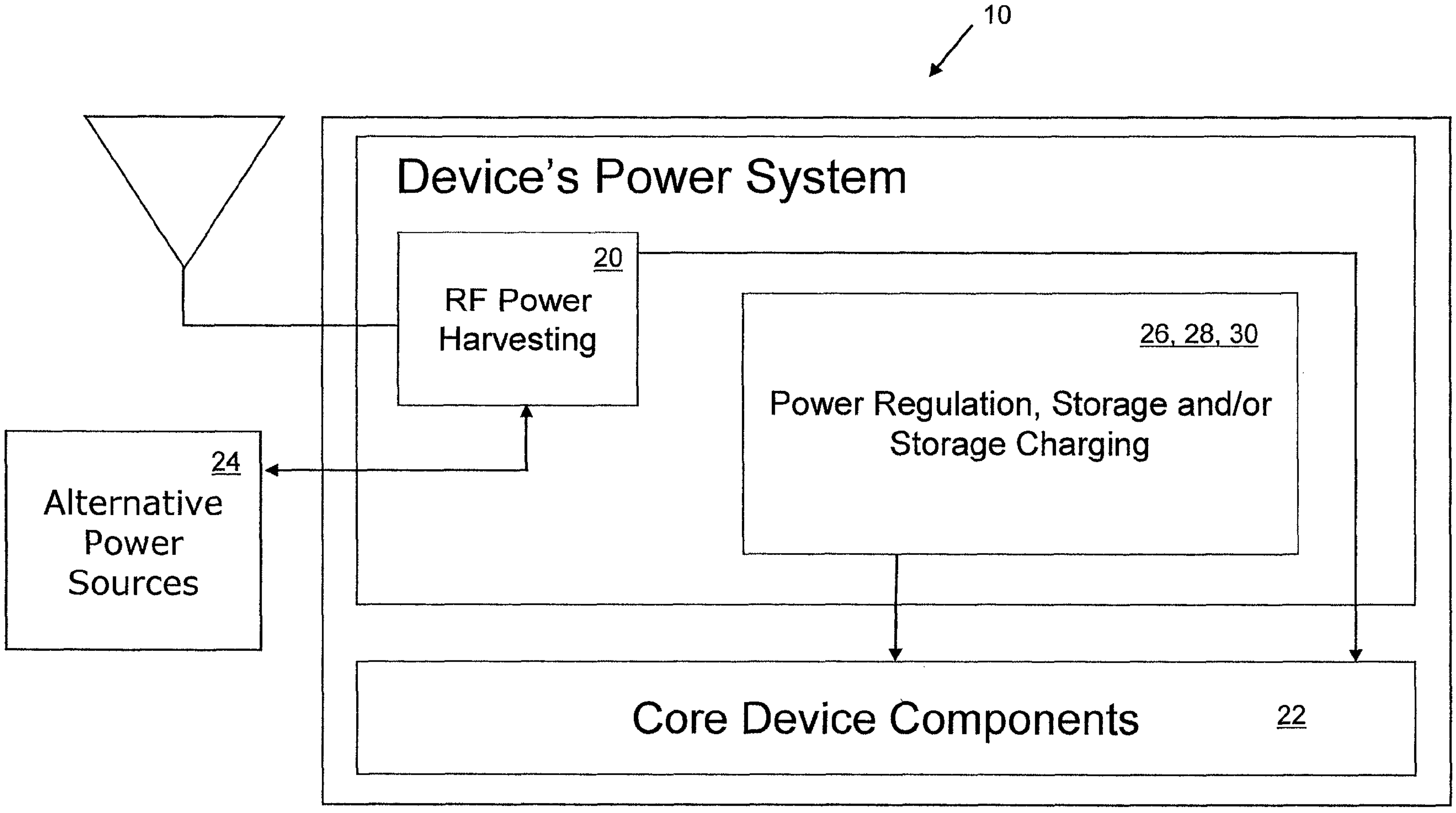
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 76

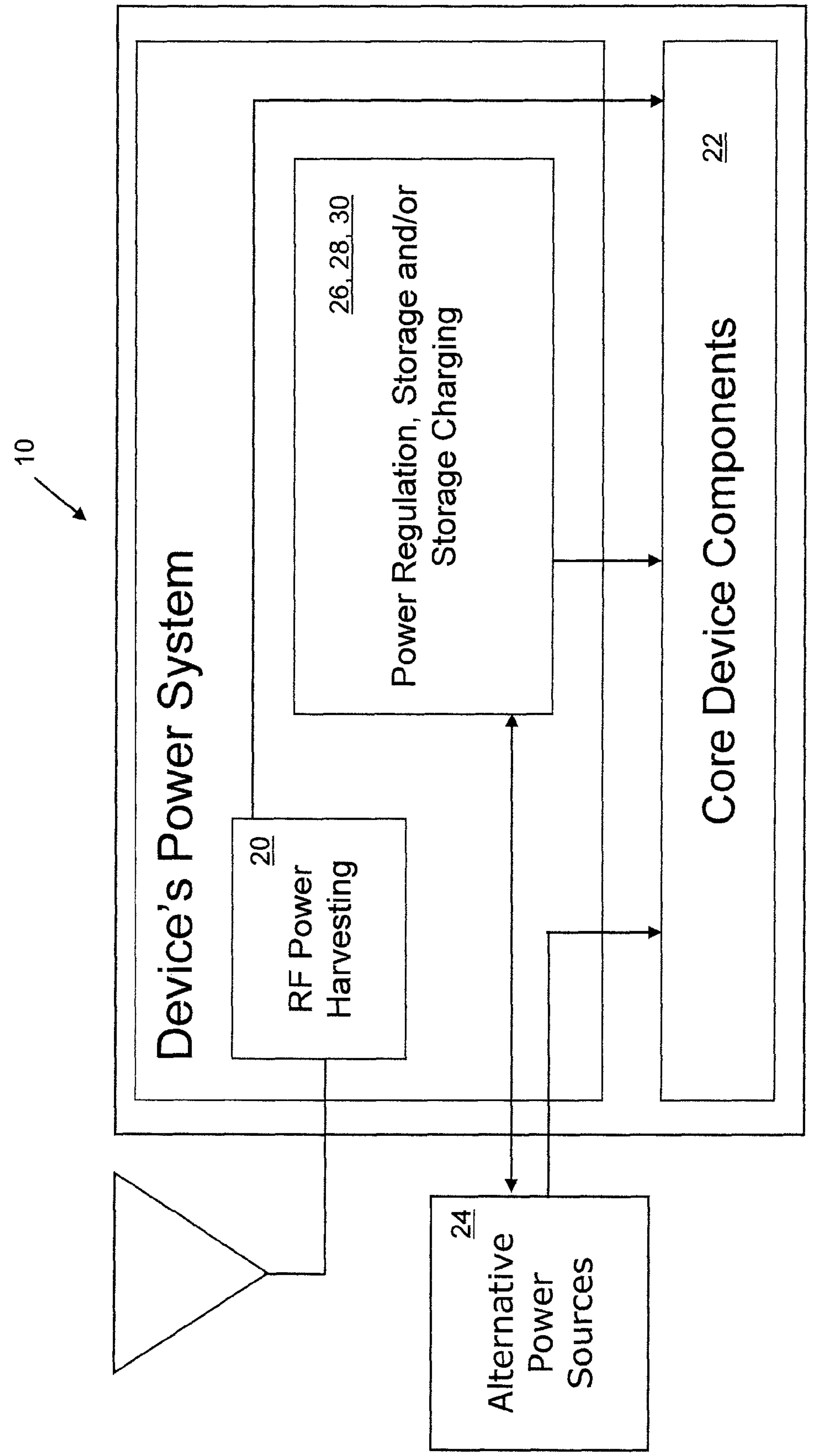
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 77

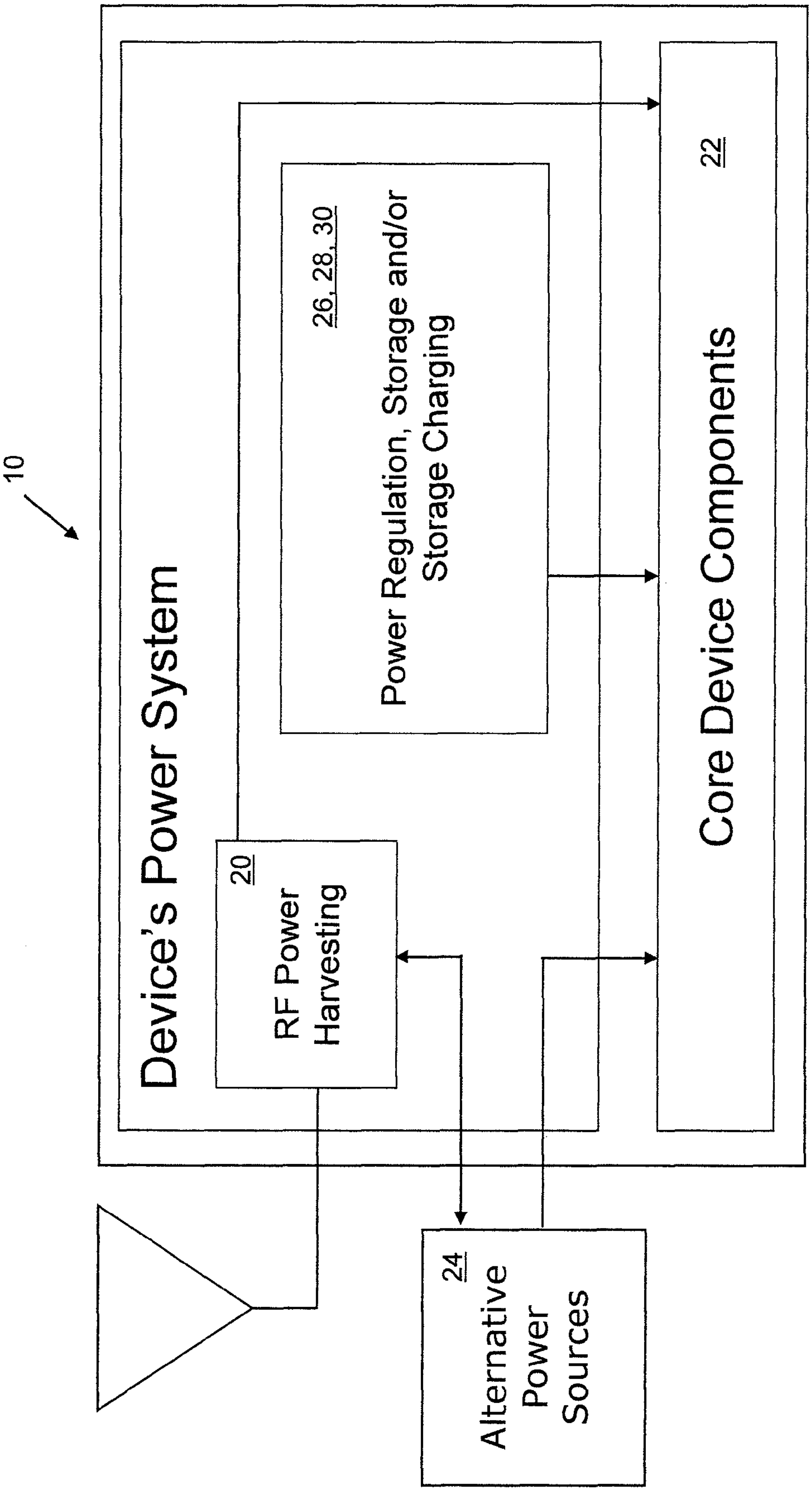
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 78

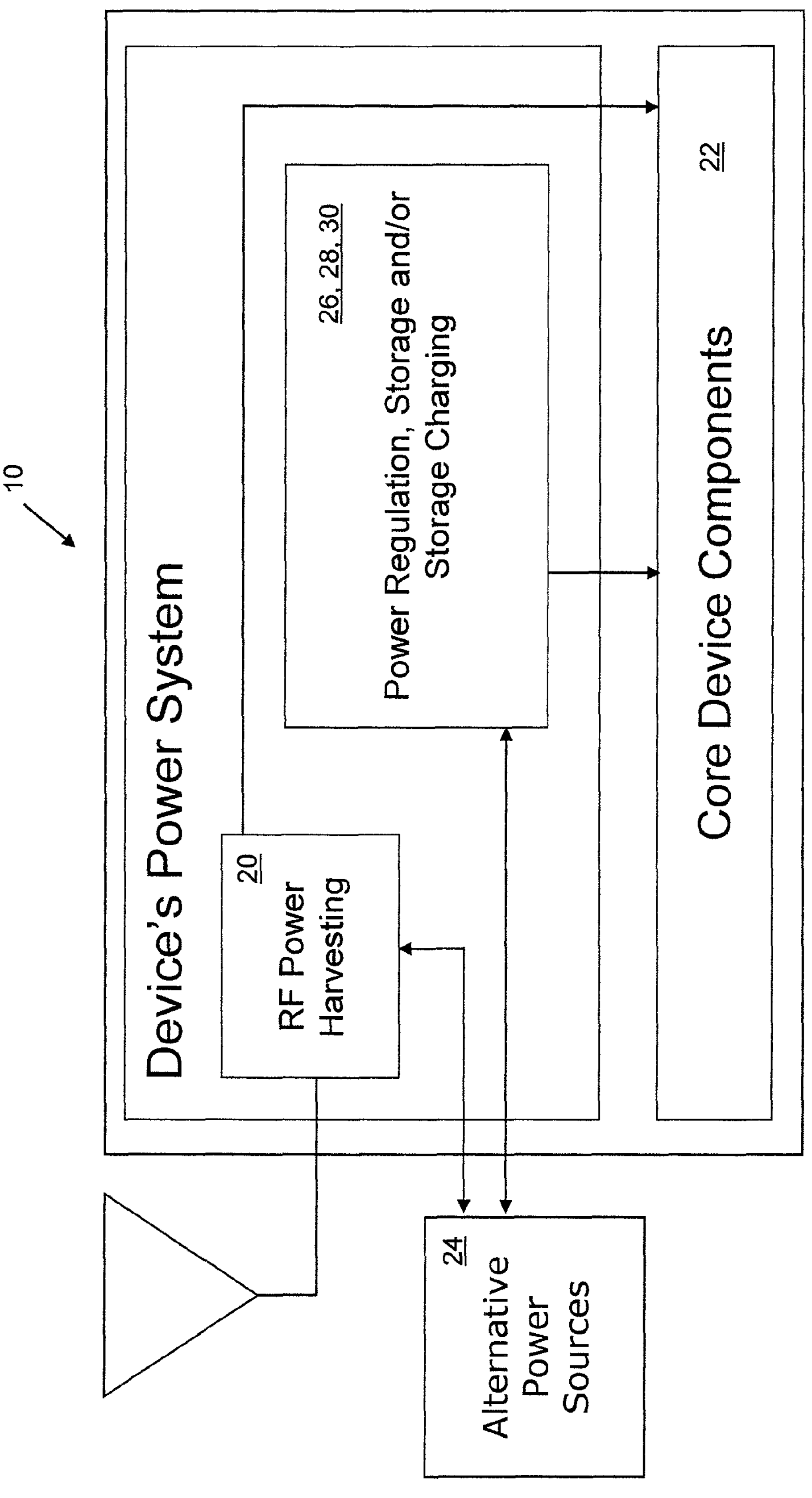
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 79

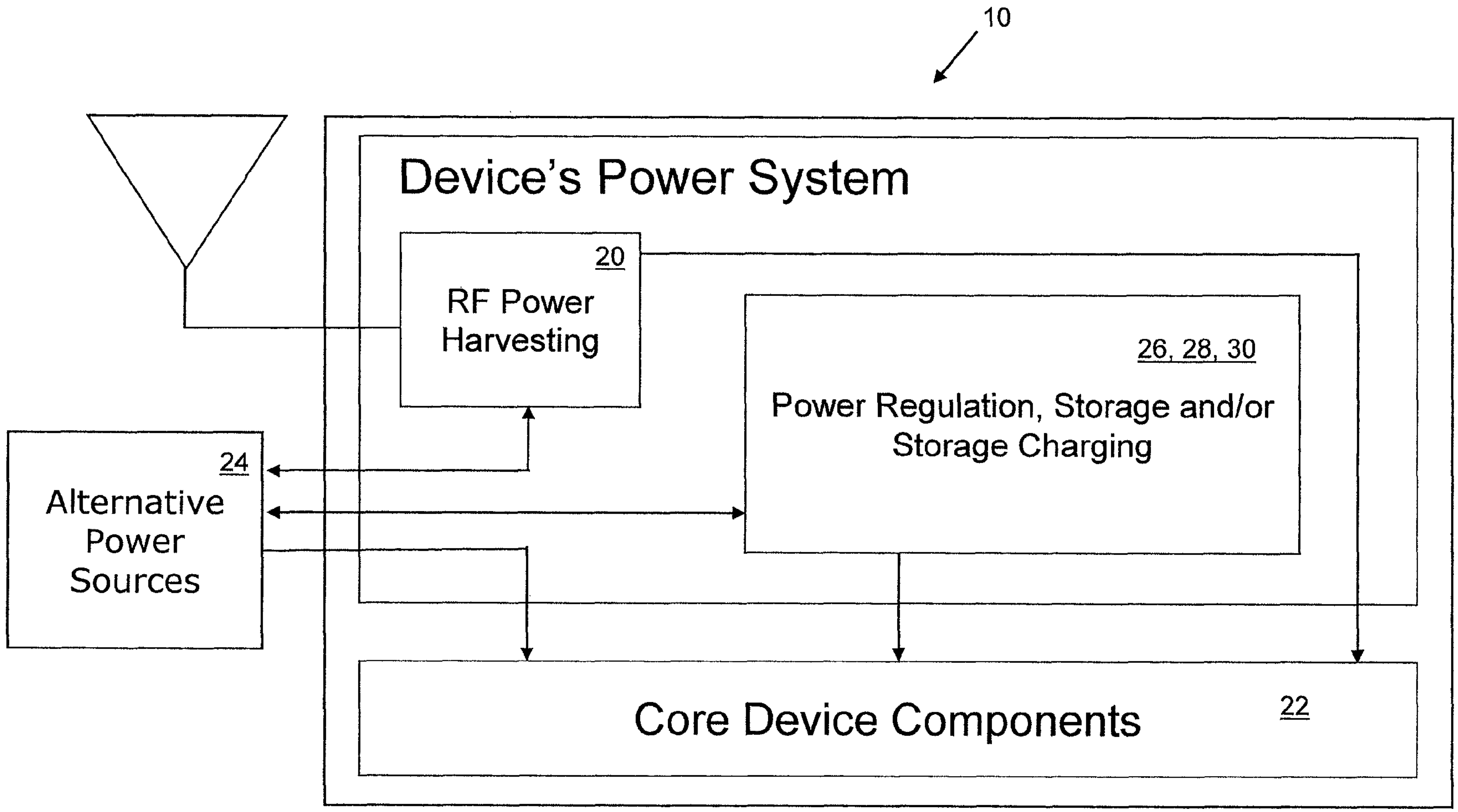
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 80

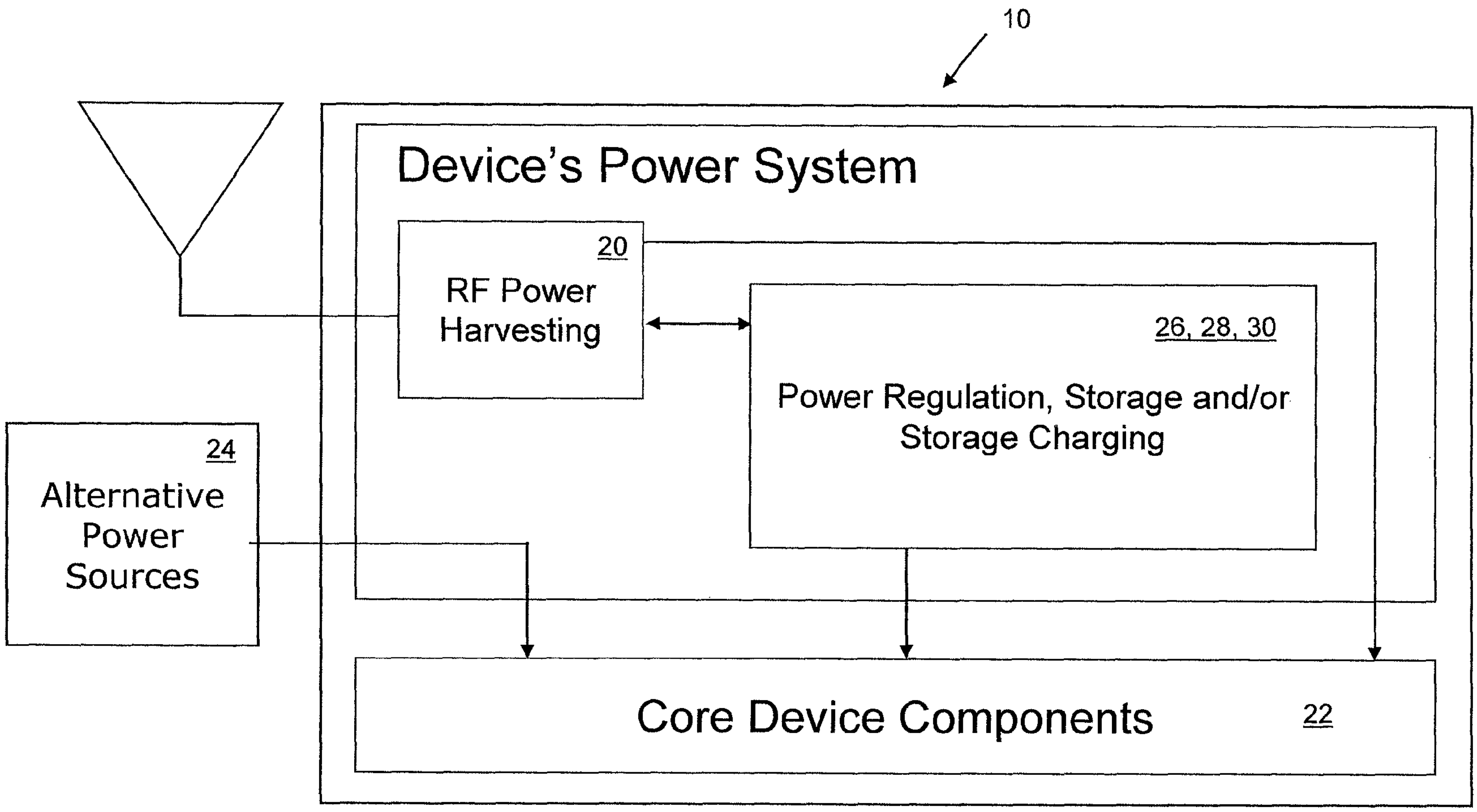
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 81

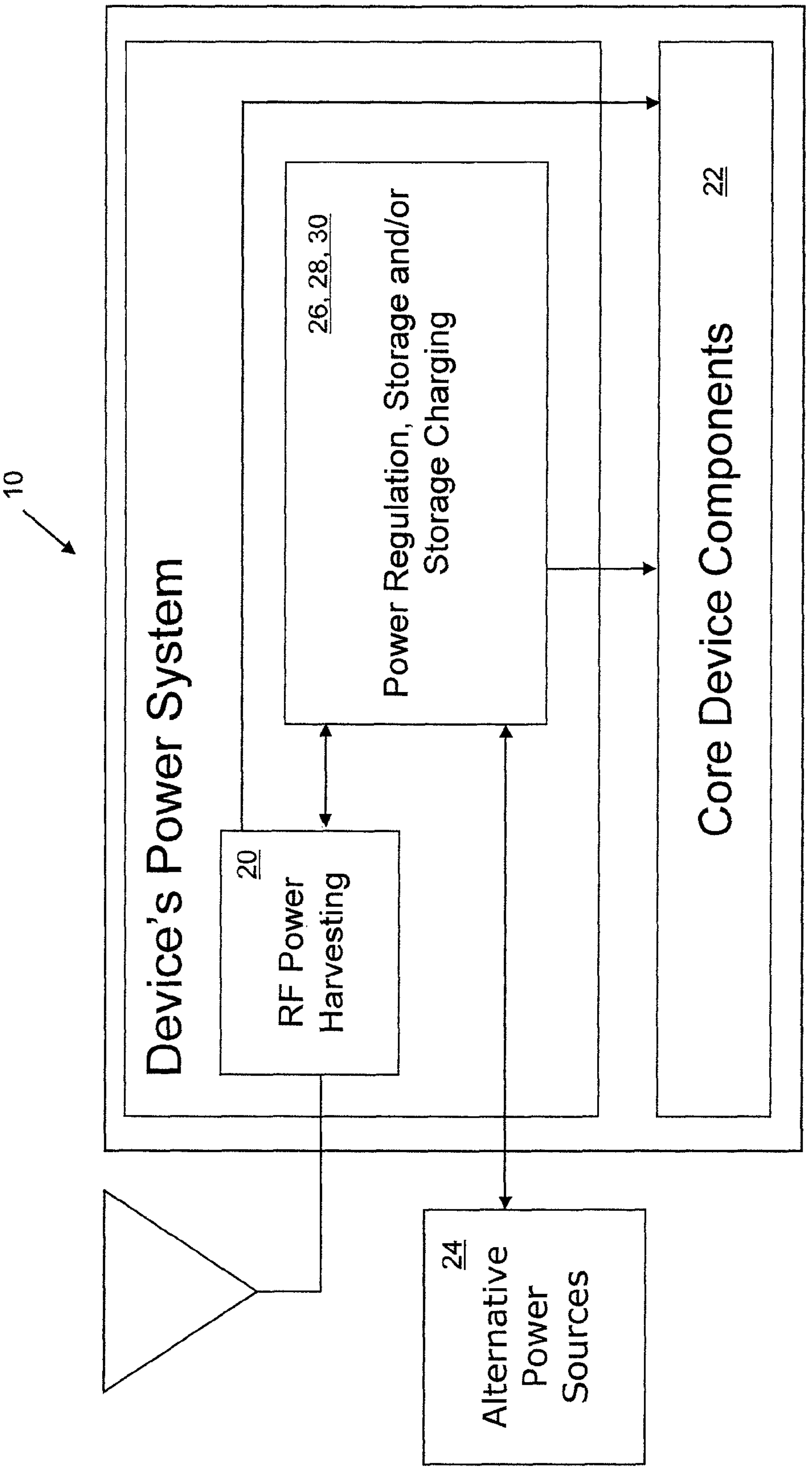
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 82

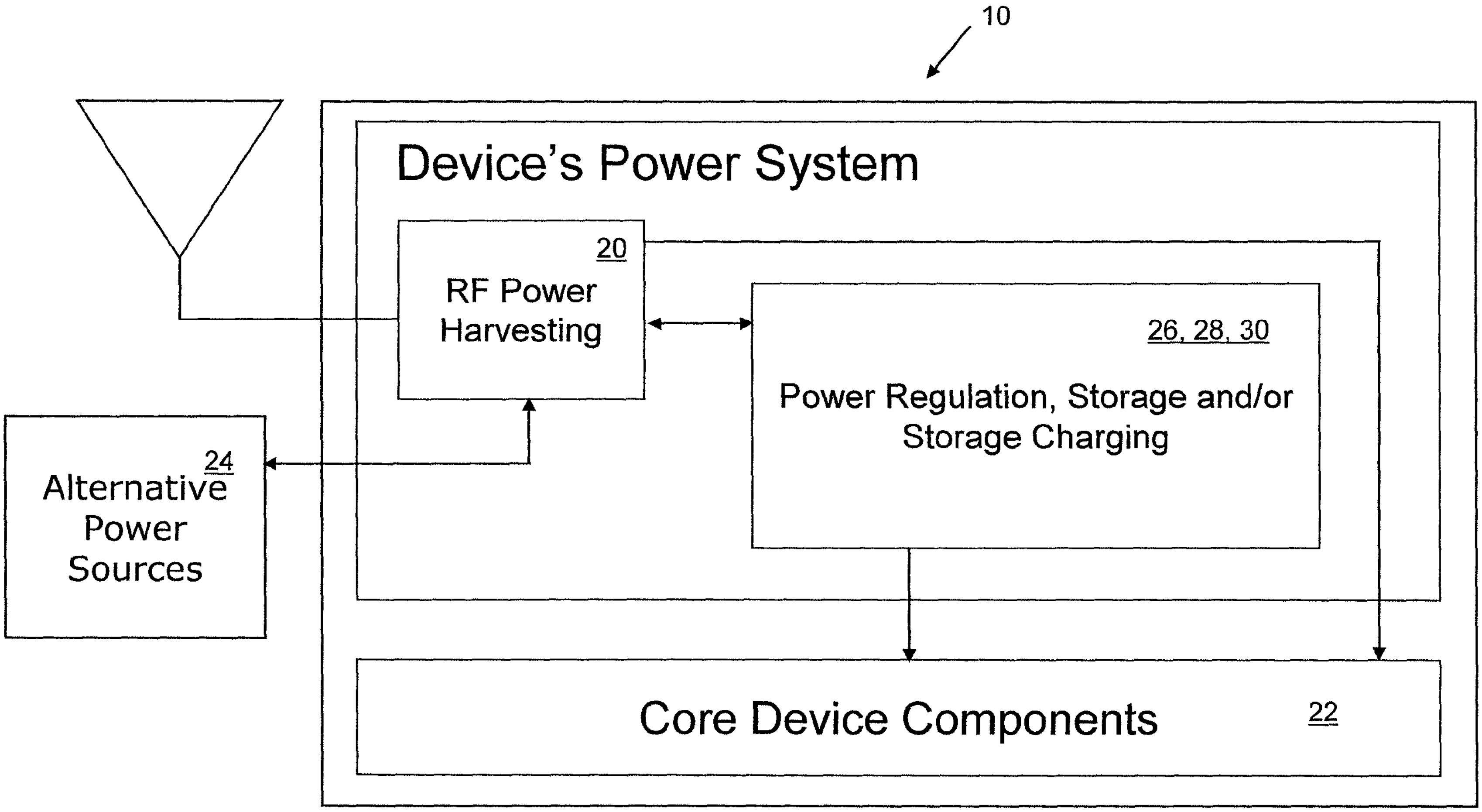
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 83

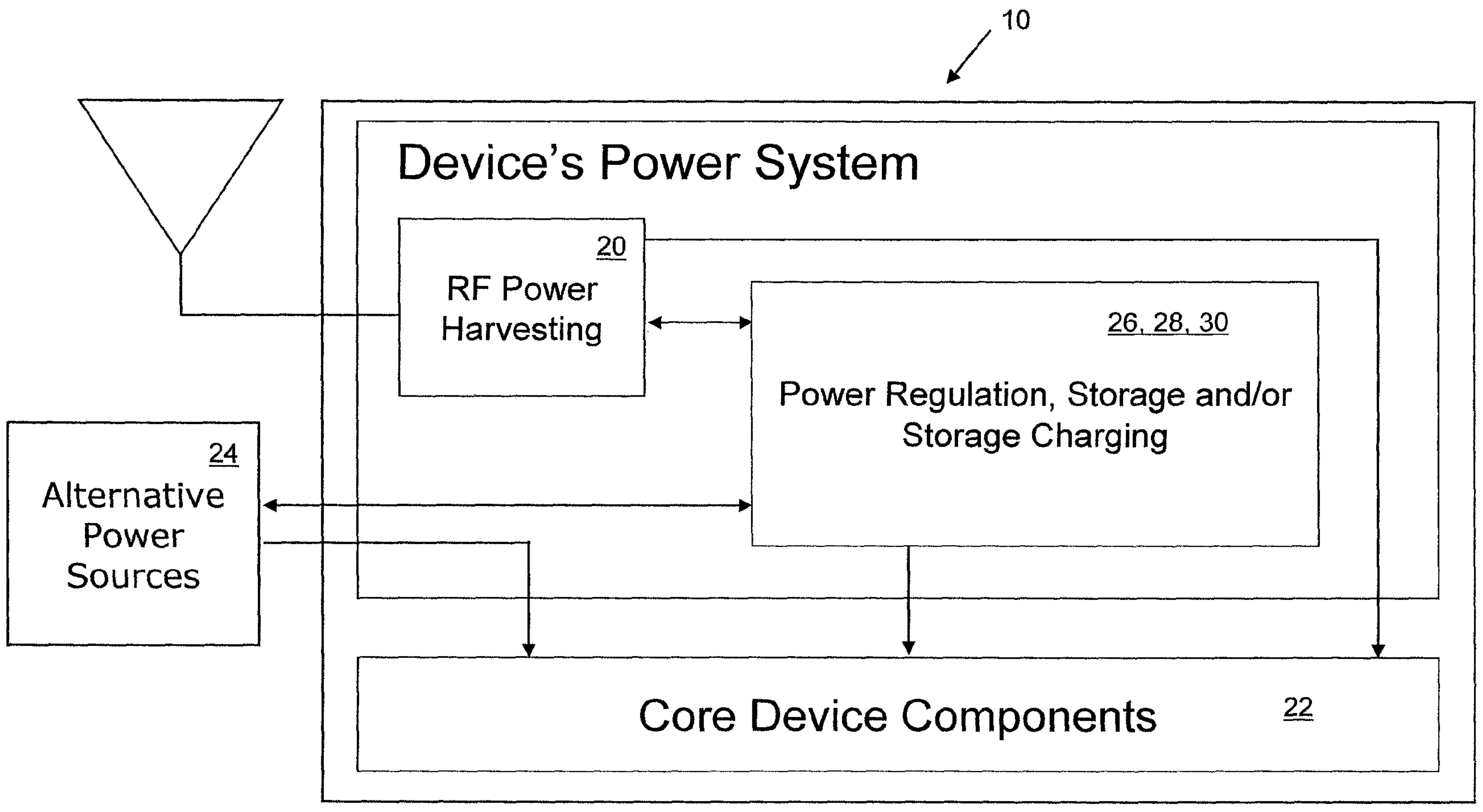
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 84

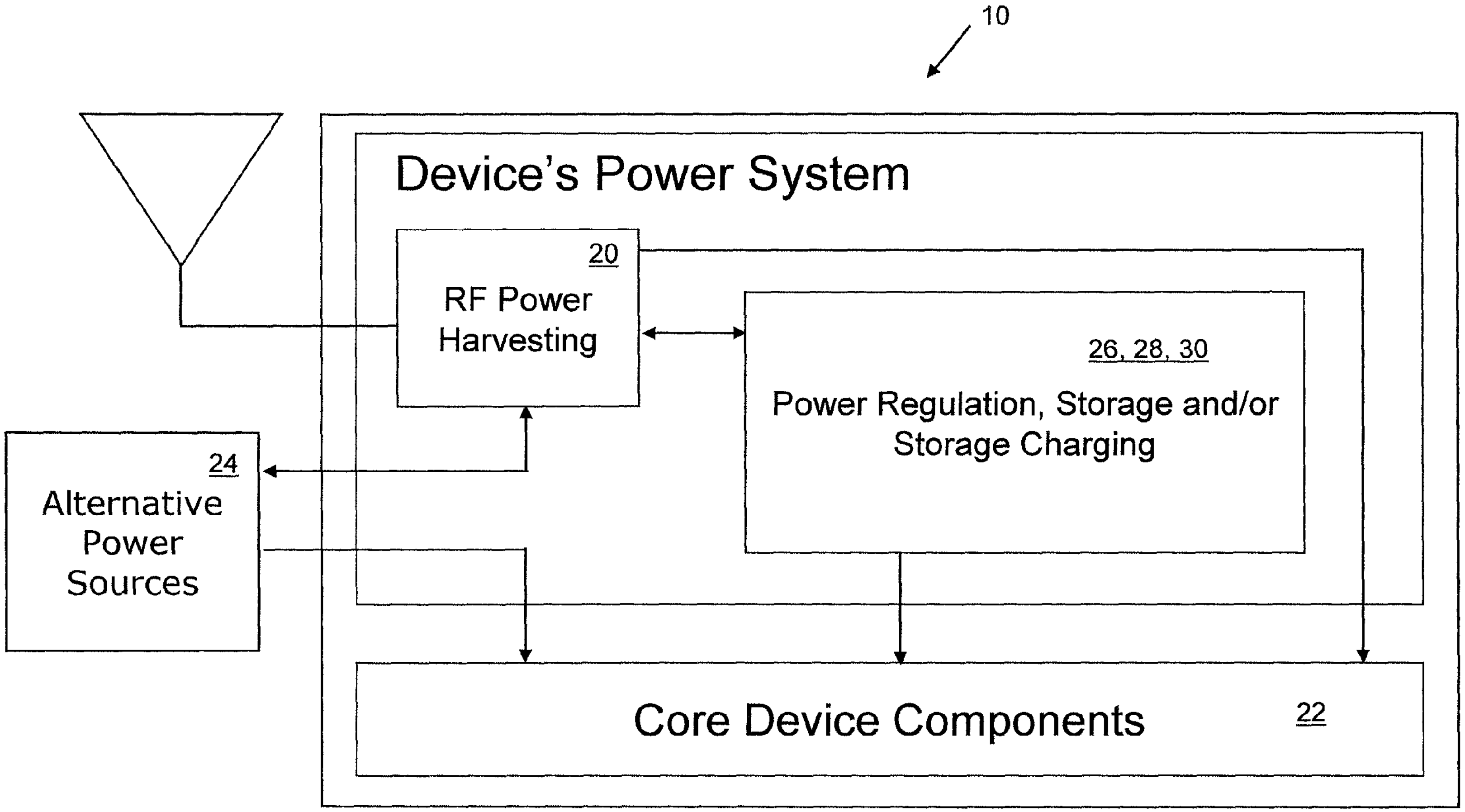
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 85

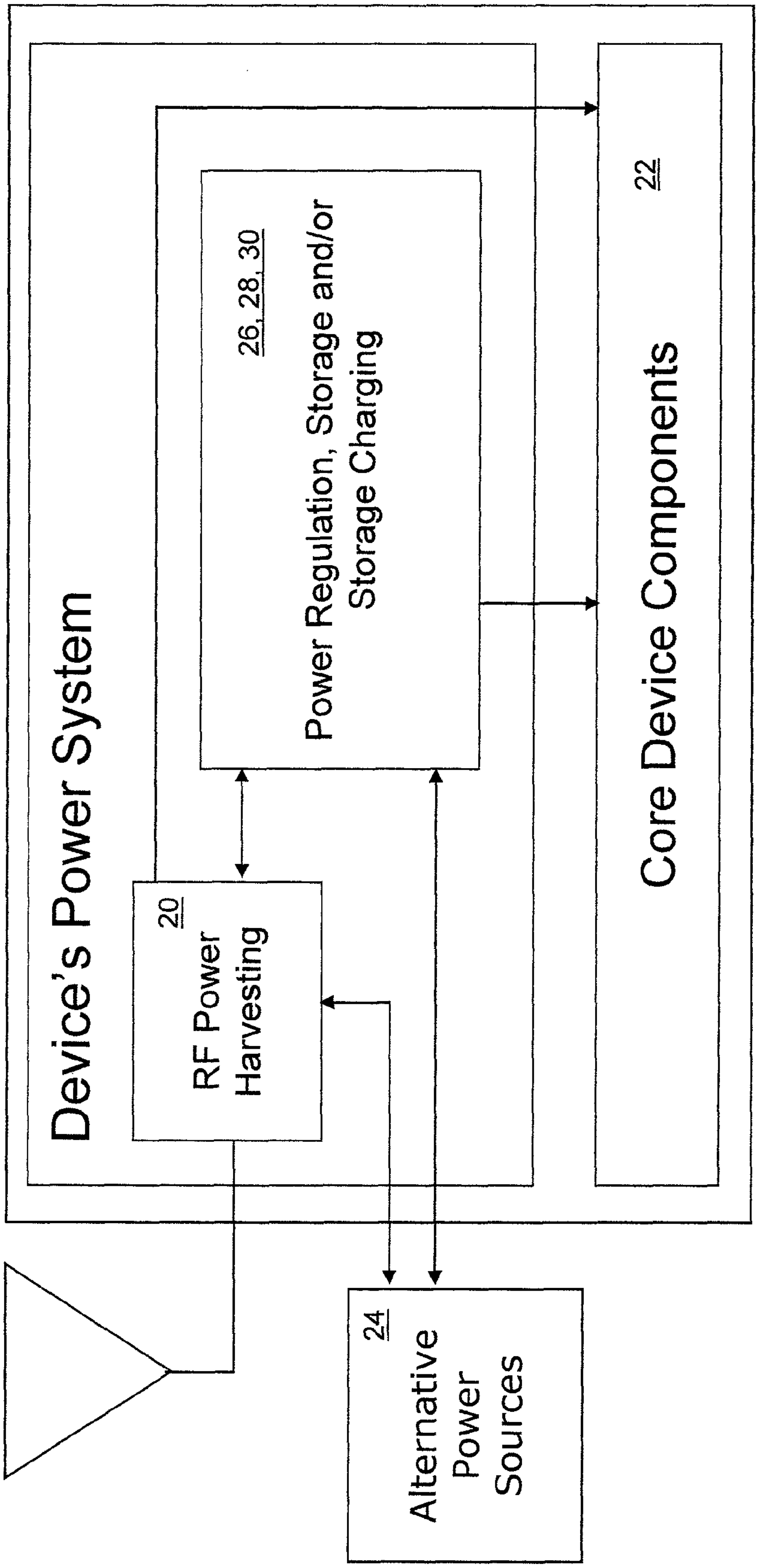
10
Device's Power System
RF Power Harvesting
20
Power Regulation, Storage and/or Storage Charging
26, 28, 30
Core Device Components
22
24
Alternative Power Sources

Fig. 86

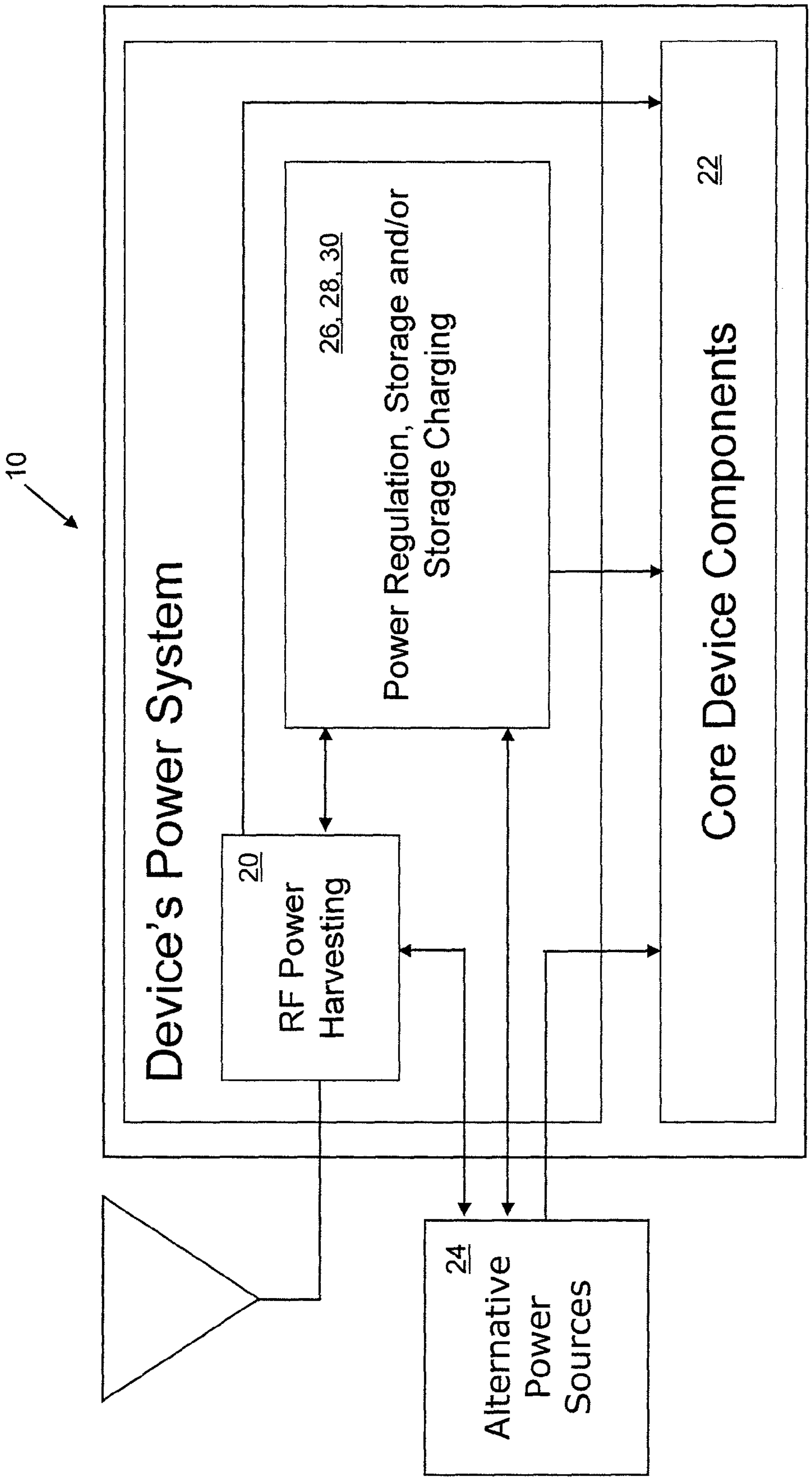
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 87

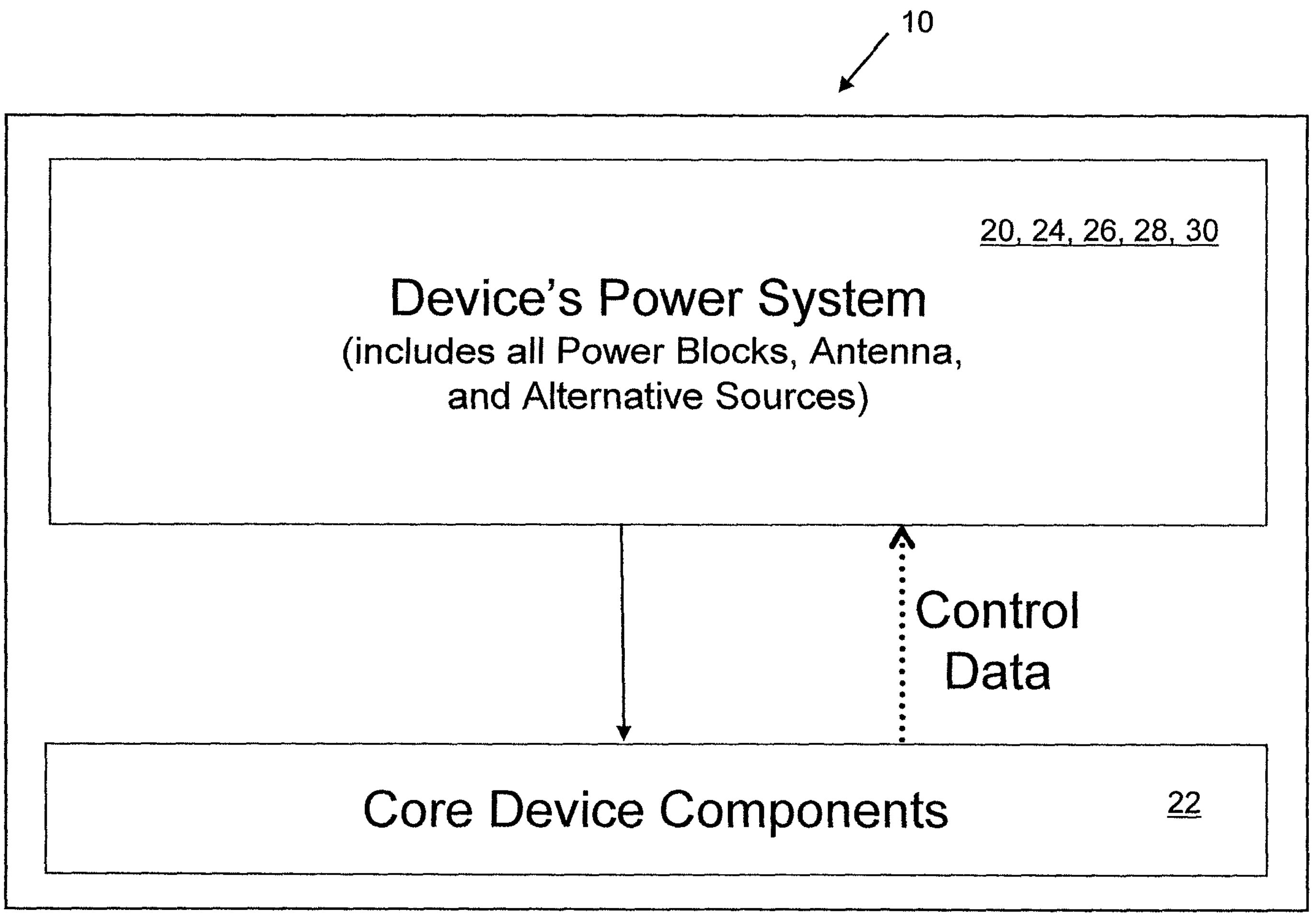
10
20, 24, 26, 28, 30
Device's Power System
(includes all Power Blocks, Antenna, and Alternative Sources)
Control Data
Core Device Components
22

Fig. 88

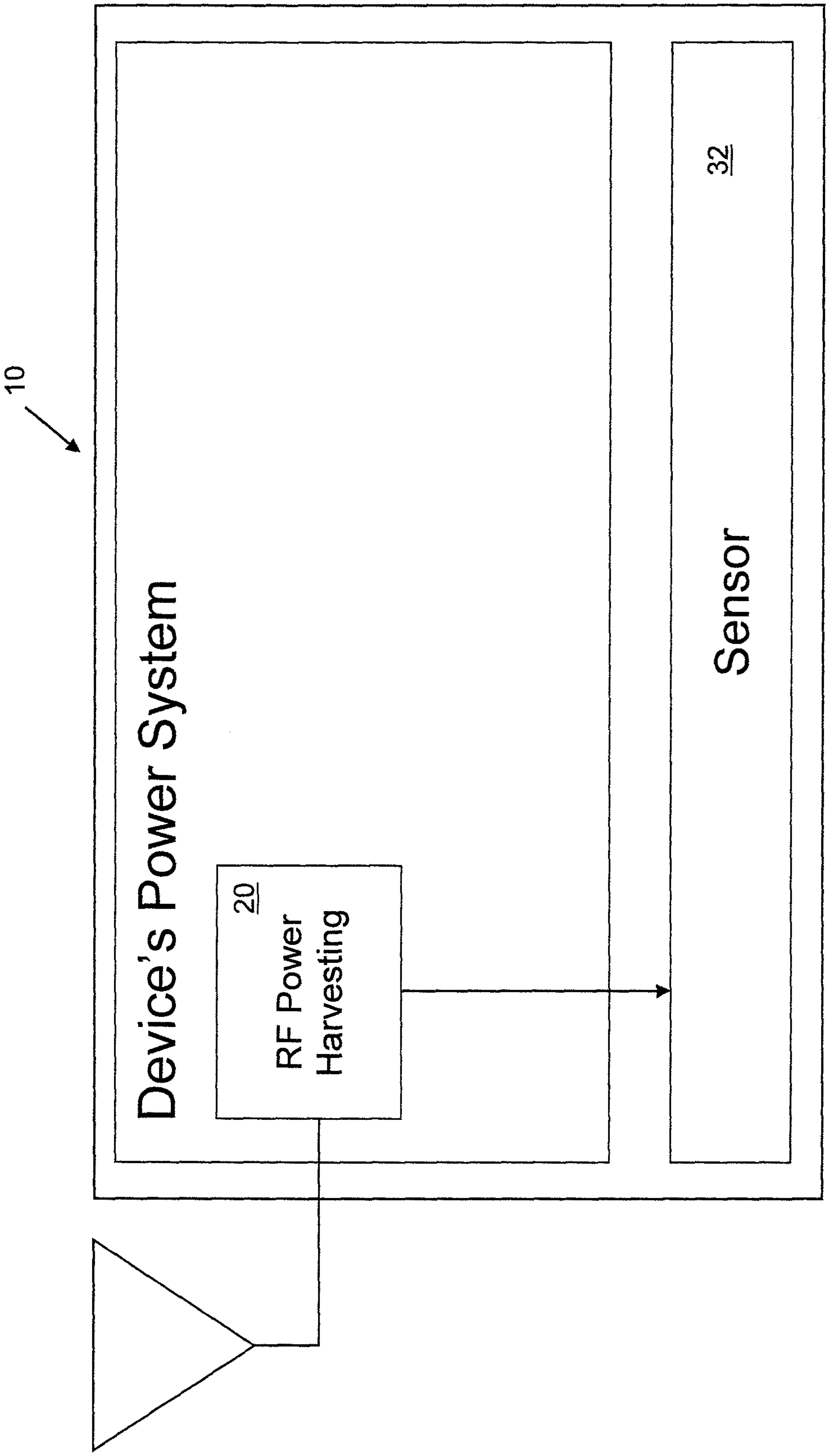
10
Device's Power System
20
RF Power Harvesting
Sensor
32

Fig. 89

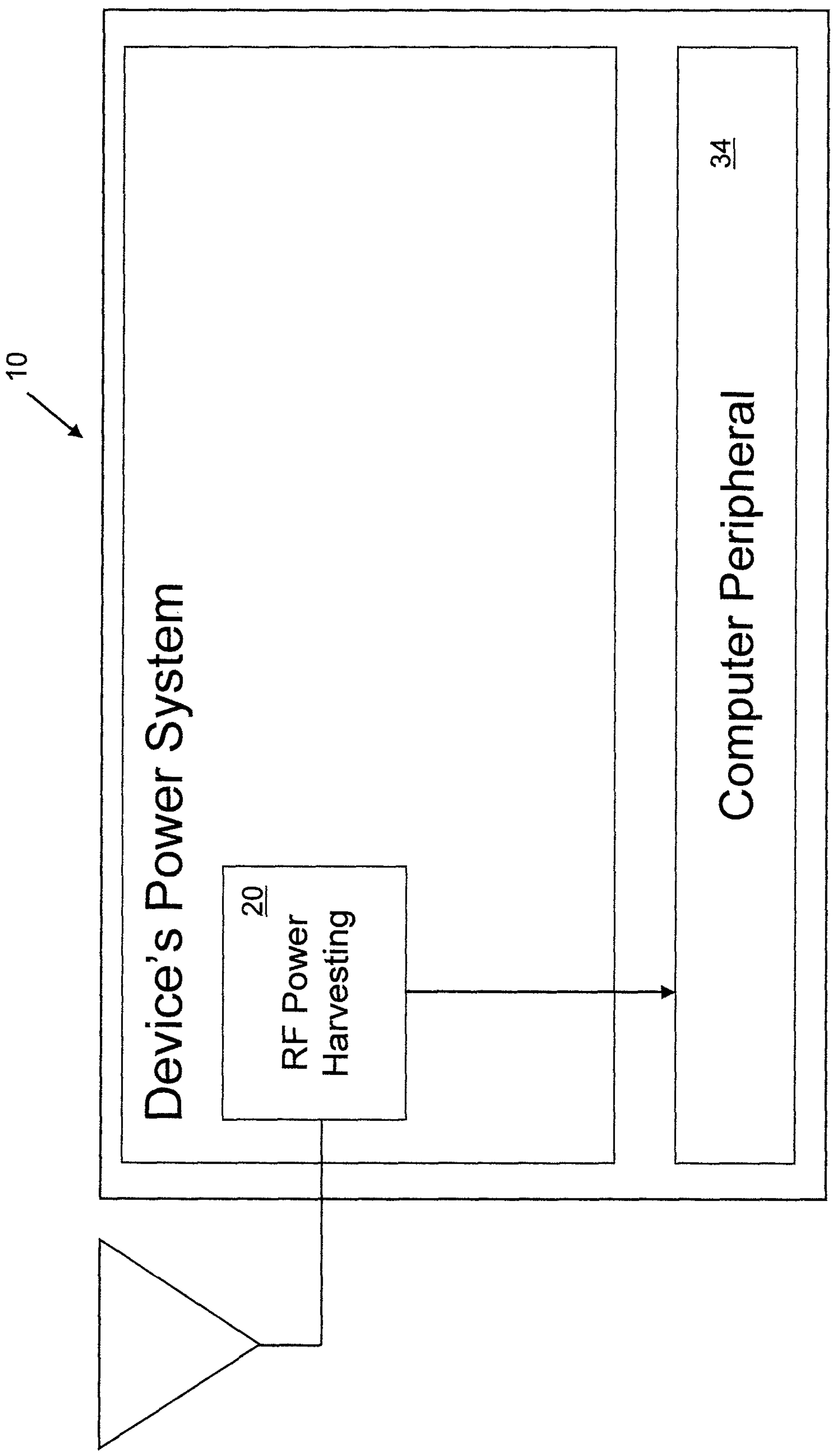
10
Device's Power System
20
RF Power Harvesting
34
Computer Peripheral

Fig. 90

POWERING DEVICES USING RF ENERGY HARVESTING

This application claims priority to and is a divisional application of U.S. patent application Ser. No. 11/447,412, filed Jun. 6, 2006, now abandoned and entitled "POWERING DEVICES USING RF ENERGY HARVESTING," which claims priority to U.S. Provisional Patent Application No. 60/688,587, filed on Jun. 8, 2005, and entitled "POWERING DEVICES USING RF ENERGY HARVESTING;" both of which are incorporated herein by references in their entireties.

FIELD OF THE INVENTION

The present invention is related to the wireless powering of devices. More specifically, the present invention is related to the wireless powering of devices with a power harvester.

BACKGROUND OF THE INVENTION

As processor capabilities have expanded and power requirements have decreased there has been an ongoing explosion of devices that operate completely independent of wires or power cords. These "untethered" devices range from cell phones, and wireless keyboards to building sensors and active RFID tags.

Engineers and designers of these untethered devices continue to have to deal with the limitations of portable power sources, primarily batteries as the key design parameter. While performance of processors and portable devices have been doubling every 18-24 months driven by Moore's law, battery technology in terms of capacity has only been growing at measly 6% per year. Even with power conscious designs and the latest in battery technology, many devices do not provide the lifetime cost and maintenance requirements for applications that require a large number of untethered devices such as logistics, and building automation. Today's devices that need two-way communication require scheduled maintenance every three to 18 months to replace or recharge the device's power source (typically a battery). One-way devices simply broadcasting their status (one-way) such as automated utility meter readers have a better battery life, typically requiring replacement within 10 years. For both device types, scheduled power-source maintenance is costly and disruptive to the entire system that a device is intended to monitor and/or control. Unscheduled maintenance trips are even more costly and disruptive. On a macro level, the relatively high cost associated with the internal battery also reduces the practical, or economically viable, number of devices that can be deployed.

The ideal solution to the power problem for untethered devices is a device or system that can collect and harness sufficient energy from the external environment. The harnessed energy would then either directly power an untethered device or augment a battery or other storage component. Directly powering an untethered device enables the device to be constructed without the need for a battery. Augmenting a storage component could be along two lines: 1) increasing the overall life of the device or 2) by providing more power to the device to increase the functionality of the device. The other parameters for an ideal solution is that the harnessing device could be used in a wide range of environments including harsh and sealed environments (e.g. nuclear reactors), would be inexpensive to produce, would be safe for humans, and would have a minimal effect on the basic size, weight and other physical characteristics of the untethered device.

BRIEF SUMMARY OF THE INVENTION

The present invention pertains to an apparatus for an application. The apparatus comprises a core device preferably having an integrated circuit for the application. The apparatus comprises a power harvester (preferably a radio frequency (RF) energy harvester) connected to the core device to power the core device.

The present invention pertains to an apparatus for an application. The apparatus comprises a core device preferably having an integrated circuit for the application. The apparatus comprises means for receiving energy wirelessly and providing power from the energy to the core device to power the integrated circuit of the core device. The receiving means is connected to the core device.

The present invention pertains to a method for an application. The method comprises the steps of converting RF energy into usable energy. There is the step of preferably powering an integrated circuit of the core device with the usable energy.

This invention pertains to a technique that uses radio frequency (RF) energy as a source of energy to directly power a device or augment a power storage component in a device. The present invention meets the requirements described in the previous "Background of the Invention" section.

Traditional RF receiving devices have focused on maximizing selectivity of the frequency to isolate and to be coherent without interference from other sources. In contrast, while the present invention operates at a specific frequency or range of frequencies, the device accepts any interference to supplement the output power of the device. Also, the research related to power harvesting that uses RF energy as a source has primarily focused on devices in close proximity of the source (inductive or near-field energy). In most cases, prior research assumed a dedicated or directed source of RF to power the device.

The invention should not be confused with power transfer by inductive coupling, which requires the device to be relatively close to the power transmission source. The RFID Handbook by the author Klaus Finkenzeller defines the inductive coupling region as distance between the transmitter and receiver of less than 0.16 times lambda where lambda is the wavelength of the RF wave. The invention can be implemented in the near field (sometimes referred to as inductive) region as well as the far-field region. The far-field region is distances greater than 0.16 times lambda.

It is an object of this invention to provide a method and apparatus to

1. remotely energize an untethered device without using direct wiring
2. power or augment the life of the power storage component so it matches the life of the device and, ultimately, powers the off-grid device with or without the use of batteries
3. allow untethered devices to be virtually maintenance free
4. provide augmentation for other energy harvesting technologies (solar, piezoelectric, etc.)
5. provide backup power to tethered devices

It is a further object of this invention to directly power a device or augment a power storage component in a device in conjunction with other power harvesting technologies and storage elements.

With this method and apparatus a device's power storage components do not require replacement, thus enabling the device to be permanently placed off-grid, where it may be physically impractical, costly, or dangerous (due to a harsh environment) to provide maintenance.

For devices on-grid (tethered) or with reliable power sources, RF power harvesting can be used as a backup in case the primary power source is lost.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING

In the accompanying drawings, the preferred embodiment of the invention and preferred methods of practicing the invention are illustrated in which:

FIG. **1** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

FIG. **2** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

FIG. **3** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

FIG. **4** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

FIG. **5** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

FIG. **6** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

FIG. **7** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

FIG. **8** is a block diagram of an RF Power Harvesting block in communication with a Power Storage block.

FIG. **9** is a block diagram of an RF Power Harvesting block in communication with a Power Storage block and used to supply power to Core Device Components.

FIG. **10** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

FIG. **11** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

FIG. **12** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block.

FIG. **13** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

FIG. **14** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

FIG. **15** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

FIG. **16** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

FIG. **17** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block.

FIG. **18** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger.

FIG. **19** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and the RF Power Harvesting block in communication with a Power Storage block.

FIG. **20** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and Core Device Components.

FIG. **21** is a block diagram of an RF Power Harvesting block supplying power to a Power Storage Charger and Core Device Components and the RF Power Harvesting block in communication with a Power Storage block.

FIG. **22** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

FIG. **23** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

FIG. **24** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

FIG. **25** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

FIG. **26** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

FIG. **27** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

FIG. **28** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the RF Power Harvesting block supplying power to Core Device Components.

FIG. **29** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

FIG. **30** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to Core Device Components.

FIG. **31** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

FIG. **32** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power Core Device Components.

FIG. **33** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

FIG. **34** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block.

FIG. **35** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components.

FIG. **36** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or

Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

FIG. **37** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components.

FIG. **38** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

FIG. **39** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

FIG. **40** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to the Power Storage Charger.

FIG. **41** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

FIG. **42** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

FIG. **43** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

FIG. **44** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the RF Power Harvesting block supplying power to Core Device Components and used to supply power to a Power Storage Charger.

FIG. **45** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to a Power Storage Charger.

FIG. **46** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

FIG. **47** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

FIG. **48** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power Core Device Components and used to supply power to a Power Storage Charger.

FIG. **49** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

FIG. **50** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to a Power Storage Charger.

FIG. **51** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

FIG. **52** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

FIG. **53** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and a Power Storage block and used to supply power to Core Device Components and used to supply power to a Power Storage Charger.

FIG. **54** is a block diagram of an RF Power Harvesting block using Antenna A to directly supply power to Core Device Components.

FIG. **55** is a block diagram of an RF Power Harvesting block using Antenna A to supply power to a Power Regulation, Storage and/or Storage Charging block.

FIG. **56** is a block diagram of an RF Power Harvesting block using Antenna A to supply power to a Power Regulation, Storage and/or Storage Charging block and used to supply power to Core Device Components.

FIG. **57** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

FIG. **58** is a block diagram of an RF Power Harvesting block used to supply power to a Power Regulation, Storage and/or Storage Charging block.

FIG. **59** is a block diagram of an RF Power Harvesting block used to supply power to a Power Regulation, Storage and/or Storage Charging block and used to supply power to Core Device Components.

FIG. **60** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components.

FIG. **61** is a block diagram of an RF Power Harvesting block used to directly supply power to Core Device Components and in communication with an Alternative Power Sources block.

FIG. **62** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and used to directly supply power to Core Device Components.

FIG. **63** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

FIG. **64** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

FIG. **65** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

FIG. **66** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

FIG. **67** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block.

FIG. **68** is a block diagram of an RF Power Harvesting block in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

FIG. **69** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

FIG. **70** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

FIG. **71** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

FIG. **72** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

FIG. **73** is a block diagram of an RF Power Harvesting block in communication with an Alternative Power Sources block.

FIG. **74** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

FIG. **75** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

FIG. **76** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

FIG. **77** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components.

FIG. **78** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

FIG. **79** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block.

FIG. **80** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources.

FIG. **81** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

FIG. **82** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

FIG. **83** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

FIG. **84** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block.

FIG. **85** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with a Power Regulation, Storage and/or Storage Charging block and an Alternative Power Sources block.

FIG. **86** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

FIG. **87** is a block diagram of an RF Power Harvesting block used to supply power to Core Device Components and in communication with an Alternative Power Sources block and a Power Regulation, Storage and/or Storage Charging block.

FIG. **88** is a block diagram of an entire power system for the device.

FIG. **89** is a block diagram of a power harvesting block used to supply power to a core device having a sensor.

FIG. **90** is a block diagram of a power harvesting block used to supply power to a core device having a computer peripheral.

DETAILED DESCRIPTION OF THE INVENTION

A complete understanding of the invention will be obtained from the following description when taken in connection with the accompanying drawing figures wherein like reference characters identify like parts throughout.

There is shown an apparatus **10** for an application. The apparatus **10** comprises a core device **22** preferably having an integrated circuit **36** for the application. The apparatus **10** comprises a power harvester **20** connected to the core device **22** to power the core device **22**.

The apparatus **10** preferably includes an alternative power source **24** connected to the core device **22** to power the core device **22** in conjunction with the power harvester **20**. Preferably, the apparatus **10** includes a power regulator **26** and/or power storage circuit **28** connected to the power harvester **20**. The apparatus **10** preferably includes a power storage charger **30** connected to the power harvester **20**. Preferably, the apparatus **10** includes a power storage connected to the power harvester **20**.

Preferably, the core device **22** includes a memory **38** connected to the integrated circuit **36** and to the power harvester **20** to power the memory **36**.

The core device **22** can include a sensor **32**, as shown in FIG. **89**. The sensor **32** can include a proximity sensor, an intrusion sensor, an environmental sensor, a chemical sensor, a biological sensor, a sensor in contact with an automobile, an occupancy sensor, a motion sensor, a position sensor, a metal detector, or a sensor **32** in contact with an aircraft. The sensor **32** can include an alarm connected to the power harvester **20** to power the alarm, a display connected to the power harvester **20** to power the display, a sensor **32** disposed in a building, an industrial automation sensor, a sensor **32** in contact with an elevator, a temperature sensor, a fire sensor, or an accelerometer.

The sensor **32** can include a level sensor, a gas level sensor, a fluid level sensor, a light level sensor. The sensor **32** can include a flow sensor, a gas flow sensor, a fluid flow sensor, a light flow sensor, or a plasma flow sensor.

The sensor **32** can include a pressure sensor, a gas pressure sensor, a fluid pressure sensor, a light sensor, an infrared light sensor, an ultraviolet light sensor, an x-ray sensor, a cosmic ray sensor, a visible light sensor, a gamma ray sensor, a stress sensor, a strain sensor, a depth sensor, or an electrical characteristic sensor.

The sensor **32** includes a voltage sensor, a current sensor, a viscosity sensor, an acoustical sensor, a sound sensor, a listening sensor, a thickness sensor, a density sensor, a surface quality sensor, a volume sensor, a physical sensor, a mass sensor, a weight sensor, a conductivity sensor, a distance sensor, an orientation sensor, or a vibration sensor.

The sensor **32** can include a radioactivity sensor, a field strength sensor, an electric field sensor or a magnetic field sensor, a smoke detector, a carbon monoxide detector, a radon detector, an air quality sensor, a humidity sensor, a glass breakage sensor, or a break beam detector. The sensor can include a thermal energy sensor, an electromagnetic sensor, a mechanical sensor, an optical sensor, a radiation sensor, a sensor in contact with a vehicle, or a sensor **32** in contact with a water craft.

The present invention pertains to an apparatus **10** for an application. The apparatus **10** comprises a core device **22** having an integrated circuit **36** for the application. The apparatus **10** comprises means for receiving energy wirelessly and providing power from the energy to the core device **22** to power the integrated circuit **36** of the core device **22**. The receiving means is connected to the core device **22**. Preferably, the core device **22** includes means for sensing.

The present invention pertains to a method for an application. The method comprises the steps of converting RF energy into usable energy. There is the step of preferably powering an integrated circuit **36** of the core device **22** with the usable energy.

Preferably, there is the step of regulating the usable energy provided to the core device **22**. There is preferably the step of storing the usable energy. Preferably, there is the step of providing power to the core device **22** from an alternative power source **24** in conjunction with the usable energy.

Alternatively, the core device **22** can include a computer peripheral **34**, as shown in FIG. **90**. The computer peripheral **34** can include a handheld game, a gaming system, a game controller, a controller, a keyboard, a mouse, a computer terminal, computer storage, or computer equipment.

The present invention can be implemented in numerous ways. A number of these ways are depicted in FIGS. **1-90**. These figures contain multiple blocks that are configured in multiple ways.

In the figures, an arrow represents the flow of power unless otherwise stated. Single-headed (one-way) arrows represent that the power is flowing from one block to another. The single-headed arrow may represent multiple wires that provide power from one block to multiple parts in the other block. Two-headed (two-way) arrows represent a single wire that can have power flow in either direction or multiple wires each having power flow in a single direction.

As an example, a two-headed arrow between an RF power harvesting **20** block and a power regulation **26** and/or power storage circuit **28** block can represent a single wire that allows harvested power to flow into a storage device such as a capacitor. The same block diagram can also represent two wires between the two blocks with the first wire allowing harvested power to flow into a voltage regulator **26**. The second wire can allow the regulated voltage to feedback to the RF power harvesting **20** block to provide power to internal components such as transistors to increase the performance of the RF power harvesting **20** block.

Each block is described in detail below. Each block represents the functionality described below associated with it. For instance, the RF power harvesting **20** block describes a power harvester **20**, and the power regulation **26** and/or power storage circuit **28** block describes a power regulator **26** and a power storage circuit **28**.

Device's Power System Block

The device's power system block includes all components/circuitry required to power the device. This block may include an RF power harvesting **20** block, a power regulation **26** and/or power storage circuit **28** block, a power storage block, and/or a power storage charger **30** block (all discussed below).

RF Power Harvesting **20** Block

The RF power harvesting **20** block is connected to an antenna. The antenna may or may not be used as the communications antenna for the core device **22** components. The RF power harvesting **20** block is used to convert the energy captured by the antenna into usable power, such as direct current (DC) voltage.

This block may include antenna matching, rectifying circuitry, voltage transforming circuitry, and/or other performance optimizing circuitry. The rectifying circuitry may include a diode(s), a transistor(s), or some other rectifying device or combination. Examples of the rectifying circuitry include, but are not limited to, half-wave, full-wave, and voltage doubling circuits.

The output of the RF power harvesting **20** block is a DC voltage or current. The RF power harvesting **20** block may accept feedback (or input) from other circuitry or blocks, which may be used to control the harvesting circuitry to improve the performance or vary the output. This feedback may include, but is not limited to, a DC voltage or a clock from the core device **22** components.

U.S. Pat. No. 6,615,074 (FIGS. 8, 9, 12a, 12b, 13, 14), incorporated by reference, herein, shows numerous examples of RF power harvesting circuits that can be used to implement the block and function described.

Power Regulation **26** and/or Power Storage Circuit **28** Block

It may be necessary to regulate the converted power (i.e., hold the power at a constant level) for specific devices. The devices that would need this block require a fairly constant voltage or current. Deviations from the required values may cause the device to not perform within its specifications.

The regulation can be implemented in many different ways. The block can be as simple as using a Zener diode, or as complicated as using an integrated circuit such as a linear voltage regulator **26** or switching regulator **26** to hold the voltage at a constant level.

Certain devices have a more tolerable power requirement. For these devices, the regulation stage may be excluded.

This block may also include, with or without the regulation, a storage device such as a capacitor, a battery, or some other device able to store charge. The output from the power regulation **26** and/or power storage circuit **28** block may be used as feedback to other blocks within the device's power system block or to the alternative power sources **24** block (described below), if they require a regulated supply voltage or stored power.

U.S. Pat. No. 6,894,467 (FIGS. 1, 3), Linear Voltage Regulator, incorporated by reference, herein, is an example of a practical application of implementing the regulation described in the block. U.S. Pat. No. 6,297,618 (FIGS. 1-4), Power storage device and method of measuring voltage of storage battery, incorporated by reference, herein, is an example of a practical application of implementing the storage described in the block.

Power Storage Charger **30** Block

The power storage charger **30** block may be needed if a power storage component of the device requires a special charging mechanism, such as pulse charging or trickle charging. This block controls how the captured and converted power is supplied to the power storage component.

U.S. Pat. No. 6,836,095 (FIGS. 1-3), Battery Charging Method and Apparatus, incorporated by reference herein, is an example of a practical application of implementing the special charging mechanism described in the block.

Power Storage Block

If a device has intermittent power requirements, it may be necessary to store the captured power for use at a later time. The power can be stored in the power storage block, which could include a battery, a capacitor, and/or another type of power storage component. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other power storage elements.

The output from the power storage block may be used as feedback to other blocks within the device's power system or to the alternative power sources **24** block, if they require a dedicated and predictable supply voltage.

U.S. Pat. No. 6,297,618 (FIGS. 1-4), Power Storage Device and Method of Measuring Voltage of Storage Battery, incorporated by reference herein, is an example of a practical application of implementing the storage described in the block. U.S. Pat. No. 6,835,501, Alkaline Rechargeable Battery, incorporated by reference herein, is also an example of a practical application of implementing the storage described in the block.

Core Device **22** Components Block

The core device **22** components block is the portion of the device that is receiving power from the device's power system. This block may be, but is not limited to, the devices listed in the subsequent pages of this document. It may be advantageous for the core device **22** components block to communicate with any of the blocks that are supplying power to it. This communication can include, but is not limited to, a feedback control signal such as a clock or an ON/OFF command. As an example, the device may want to turn off the alternative power sources **24** block if it is receiving sufficient power from the RF power harvesting **20** block.

Alternative Power Sources **24** Block

RF energy harvesting also has the ability to be augmented by other types of power harvesting, storage components, or dedicated sources (e.g. power line). The alternative power sources **24** block shows how this type of system could be implemented. The augmenting power harvesting technologies include, but are not limited to, solar, light (visible and non-visible), piezoelectric, vibration, acoustic, thermal, microgenerators, wind, and other environmental elements. This block can work independently or have communication with other blocks.

U.S. Pat. No. 6,784,358, Solar Cell Structure Utilizing an Amorphous Silicon Discrete By-Pass Diode, incorporated by reference herein, is an example of a practical application of implementing an alternative power source **24** described by the block. U.S. Pat. No. 6,858,970, Multi-Frequency Piezoelectric Energy Harvester, incorporated by reference herein, is also an example of a practical application of implementing an alternative power source **24** described by the block.

Power Regulation **26**, Storage **28** and/or Storage Charging **30** Block

The power regulation **26**, storage **28** and/or storage charging **30** block contains all the combinations of the power regulation **26** and/or power storage circuit **28** block, power storage charger **30** block, and power storage block. This block is used in the later figures to reduce the number of figures needed to show how the blocks can interconnect.

The disclosed invention is an apparatus and method for an application for retrieving radio frequency (RF) energy by an antenna, converting that energy into direct current (DC) power, regulating that energy using an optimized circuit, storing that energy in an optimized component, and/or supplying the power for a specific device.

Retrieval of RF Energy

The RF energy is retrieved from the environment by the use of the antenna. The antenna can be shared or standalone with respect to an antenna used for the device's wireless communication. FIGS. **54-56** show a device that has an antenna A for use by the RF power harvester **20** and an antenna B used for wireless one-way or two-way communication. FIGS. **57-59** show a device where the antenna is shared by both a device's communication module (within the core device **22** components) and the RF power harvester **20**. In terms of form factor, the antenna used by the apparatus **10** can be a separate component or integrated directly into the form factor of the device.

The antenna is able to capture two types of available RF energy. The first type of energy exists as ambient RF energy. This type of RF surrounds us in our day-to-day lives and is usually generated to carry one way or two-way combinations of voice, video and data communications. The sources that the antenna can harvest from include, but are not limited to, medium-frequency AM radio broadcast, very-high-frequency (VHF) FM radio broadcast and television broadcast, ultra-high-frequency (UHF) broadcast, cellular base stations, wireless data access points, super-high-frequency (SHF) frequencies, and the industrial, scientific, and medical (ISM) bands. These sources cover transmission frequencies from 300 kHz to 30 GHz.

The second type of energy available is directed RF energy. This type of RF energy is directed from a transmitter specifically designed to deliver RF energy for harvesting by the antenna. The transmitter can be configured as a standalone device or integrated into an existing device.

Conversion of the Energy into DC

The RF energy captured by the antenna must be converted into a useful form of energy for the specific device. This conversion is shown in block form in all FIGS. (**1-90**) as the RF power harvesting **20** block. The most common form of useable energy is DC energy. To perform this conversion, the block includes circuitry to rectify the captured alternating current (AC) energy to create DC energy. The rectification in this block can be done with a diode(s), a transistor(s), or some other rectifying device or combination.

Regulation of the Energy

It may be necessary to regulate the converted power (hold the power at a constant level) for specific devices. FIGS. **2-7**, **10-17**, and **22-53** show how this regulation can be implemented using a power regulation **26** and/or power storage circuit **28** block. The devices that would need this block require a fairly constant voltage or current. Deviations from the required values may cause the device to not perform within its specifications. The regulation can be implemented in many different ways. The block can be as simple as using a Zener diode, or as complicated as using an integrated circuit such as a linear voltage regulator **26** or a switching regulator **26** to hold the voltage at a constant level. Certain devices have a more tolerable power requirement. For these devices, the regulation stage may be excluded.

Storage of the Energy

If a device has intermittent power requirements, such as the devices exampled by FIGS. **2-53**, it may be necessary to store the captured power for use at a later time. The power can be stored in the power storage block or the power regulation **26** and/or power storage circuit **28** block. Storage devices can include, but are not limited to, a battery, a capacitor, or another type of power storage component. In certain applications, it may be necessary to include additional circuitry that controls how the power is transferred to the storage device. The power storage charger **30** block is shown in FIGS. **18-53**. This may be needed if the power storage component requires a special charging mechanism such as pulse charging or trickle charging. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other storage elements. There are devices that will not require storage. These devices can run directly off of the converted power. These devices also may or may not require regulation of the captured power.

Supplying the Power

The captured DC power, which may or may not be regulated and/or stored, is supplied to the device, which is represented by the core device **22** components block in the figures. This may be a single connection or it may supply multiple parts of the device with power.

RF energy harvesting also has the ability to be augmented by other types of power harvesting or storage components. Other power harvesting technologies include, but are not limited to, solar, light (visible and non-visible), piezoelectric, vibration, acoustic, thermal, microgenerators, wind, and other environmental elements. Power storage components include, but are not limited to, batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other storage elements. FIGS. **60-88** show how the alternative power sources **24** block can be connected to an RF energy harvesting system. These figure show how the RF energy harvesting **20** block and the alternative power sources **24** block can work independently or have communication with each other. The antenna configurations shown in FIGS. **54-59** are still applicable with the addition of an alternative power sources **24** block. These antenna configurations can be applied to FIGS. **60-88**.

RF energy harvesting also has the ability to provide a backup to devices on-grid (tethered) or with reliable power sources, which can be used in case the primary power source is lost. As an example, it may be mandated by regulations that a sensor has auxiliary power in case the primary supply is lost. It could be possible to use a rechargeable battery that obtains its charge from the primary supply when in operation. However, if the primary supply is lost for a time greater than the life of the rechargeable battery, the specification of uninterrupted power is not met. RF energy could be used to supply power to the described device while the primary supply is not available. The primary supply could include, but is not limited to, an on-grid connection, a generator, a battery, or other reliable power supply.

RF energy harvesting with or without alternative source augmentation is applicable to provide electric power directly or indirectly to a range of electronic components contained in any specific electrical or electronic device and includes, but is not limited to:

Passive electronic components, active electronic components;

Resistors, fixed resistors, variable resistors, thermistors, thyristor, thermocouple;

Capacitors, Electrolytic Capacitors, Tantalum Capacitors, Ceramic Capacitors, Multilayer Ceramic Capacitors, Polystyrene Film Capacitors, Electric Double Layer Capacitors (Super Capacitors), Polyester Film Capacitors, Polypropylene Capacitors, Mica Capacitors, Metallized Polyester Film Capacitors, Variable Capacitors;

Diodes, Voltage regulation diodes, light-emitting diodes, organic light-emitting diodes, Variable capacitance diodes, Rectification diodes, Switching diodes, Regulation Diodes, Diode bridges, Schottky barrier diodes, tunnel diodes, PIN diodes, Zener diodes, Avalanche diodes, TVSs;

Integrated circuits, microcontroller unit (MCU), microprocessor unit (MPU), logic circuits, memory, printed circuits, circuit boards, printed wiring boards;

Transistors, MOSFETs, FETs, BJTs, JFETs, IGBTs, Relays, Antennas, semiconductors, conductors, inductors, relays, diacs, triacs, SCRs, MOVs;

Fuses, circuit breakers;

Batteries, Non-rechargeable batteries, rechargeable batteries, coin cell batteries, button cell batteries, alkaline batteries, lithium batteries, lithium ion batteries, lithium polymer batteries, NIMH batteries, NICAD batteries, Lead acid batteries, Zinc air batteries, Manganese Lithium batteries, Niobium Titanium Lithium batteries, Vanadium Pentoxide Lithium batteries, Carbon Zinc batteries, Zinc Chloride batteries, Lithium Thionyl Chloride batteries, Manganese Dioxide batteries, Lithium Poly-Carbonmonofluoride batteries, Lithium Manganese Dioxide batteries, Lithium Chloride batteries, Lead Acid Calcium batteries, Lead Acid Tin batteries, Oxy Nickel batteries, Silver Oxide batteries, Magnesium batteries;

Inductors, Coils, High Frequency Coils, Toroidal Coils, Transformers, switches, chokes;

Motors, DC motors, stepper motor, AC motors, Fans;

Crystals, Oscillators, Clocks, Timers; and

Displays, LCDs, LED displays.

RF energy harvesting with or without alternative source augmentation is applicable across a range of markets and specific devices and includes but is not limited to:

Consumer Electronics:

Electronic equipment, wired devices, battery powered devices, wireless communication devices, cell phones, telephones, phones, cordless phones, portable phones, Bluetooth devices, Bluetooth headsets, hands-free headsets, headsets, headphones, Wireless headsets, radios, AM/FM radios, shortwave radios, weather radios, Two-way radios, portable radio, lights, lanterns, portable lights, flashlights, nightlights, spotlights, search lights, calculators, graphing calculators, desk calculators, clocks, alarm clocks, wall clocks, desk clocks, travel clocks, watches, wristwatches, pocket watches, stop watches, timers, voice recorders, Dictaphones, laser pointers, power tools, cordless power tools, electronic razors, electric razors, handheld games, gaming systems, game controllers, wireless game controllers, remote controls, battery chargers, computers, portable computers, keyless entry, toys, toy guns, toy laser guns, games, microphones, musical instruments, musical effects processors, musical instrument tuners, metronomes, electronic chord charts, door openers, garage door openers, PDA, Cameras, Video recorders, Multimeter, electronic test equipment, hand-held electronics, portable electronics, wireless pens, sound generators, noise generators, language translators, electric toothbrushes, portable televisions, pagers, transceivers, toy vehicles, remote control vehicles, toy planes, remote control planes, pet containment systems, invisible fence pet sensors, memory backup, base station battery backups, appliance battery backups, uninterrupted power supplies, GPS devices, memory retention power supplies, metal detectors, stud finders, metal stud finders, stun guns, tazers, wearable devices, baby monitors, intercoms, doorbells, wireless doorbells, electronic office supplies, electronic staplers, radar jammers, radar detectors, digital scales, microfilm cassettes, video head testers, compasses, noise canceling headphones, air samplers, depth finders, barometers, weather measurement instruments, data transfer devices, automatic distress signaling unit, Wireless audio speakers, Satellite radios, Police scanners, Car navigation systems (GPS devices), Decorative lights, Christmas lights, garden lights, lawn lights, ornamental lights, porch lights;

Multi-media players: MP3, DVD, analog music players, CD players, tape players, digital music players, digital video players, minidisc; and

Computer: keyboards, mice, peripherals, computer equipment, electronic computers, computer storage, computer terminals;

Building/Industrial Automation:

Sensors: Position, elevator, temperature, fire, accelerometers, level, gas level, fluid level, light level, flow, gas flow, fluid flow, light flow, plasma flow, pressure, gas pressure, fluid pressure, motion, light, infrared light, ultraviolet light, X-rays, cosmic rays, visible light, gamma rays, chemical, stress, strain, depth, electrical characteristics, voltage, current, viscosity, acoustical, sound, listening, thickness, density, surface quality, volume, physical, mass, weight, force, conductivity, distance, orientation, vibration, radioactivity, field strength, electric field strength, magnetic field strength, occupancy, smoke detector, carbon monoxide detector, radon detector, air quality, humidity, glass breakage, break beam detector;

Controls: Position, elevator, temperature, fire, accelerometers, level, gas level, fluid level, light level, flow, gas flow, fluid flow, light flow, plasma flow, pressure, gas pressure, fluid pressure, motion, light, infrared light, ultraviolet light, X-rays, cosmic rays, visible light, gamma rays, chemical, stress, strain, depth, electrical characteristics, voltage, current, viscosity, acoustical, sound, listening, thickness, density, surface quality, volume, physical, mass, weight, force, conductivity, distance, orientation, vibration, radioactivity, field strength, electric field strength, magnetic field strength, occupancy, smoke detector, carbon monoxide detector, radon detector, air quality, humidity; and

Devices: Thermostats, light switches, door locks, smart-card door locks, lighting, emergency lighting, motion lighting, safety lighting, highway lighting, construction lighting, sign lighting, roadway sign lighting, construction sign lighting, automatic flushing units, automatic soap dispenser, automatic paper towel dispenser, automatic faucets, automatic door sensors, identification reader, fingerprint reader, credit card readers, card readers, valve actuators, gauges, analog gauges, digital gauges, fire extinguishers, wireless switches, remotely operated inspection equipment, gas/oil pipeline monitoring systems, robotic pipeline inspection gauges, “auto-reclosers” for electric power lines, sonar buoys, telemetry systems, electronic record tracking systems, robbery tracking devices, interrogators, programmers, emergency exit alarms, alarms, flood alarms, gas alarms, electronic entry systems, security keypads, silo transducers, data recorders, signal tracers, anti-static strap testers, radiosonde weather balloons, utilities load controllers, profilometers, noise cancellation equipment, infrared beacons;

Military/Government:

Tracking tags: Weapons, vehicle, soldier, gear/assets, staff, general population, security access badges;

Sensors: Proximity, intrusion, environmental, chemical/biological; and

Equipment: Battery charger, surveillance, card readers, identification reader, fingerprint reader, retinal scanners, satellites, rockets, space vehicles, search and rescue transponders (SARTs), emergency position-indicating rescue beacons (EPIRBs), emergency locator transmitters (ELTs), military radios, electronic toll collection systems, postal tracking systems, communications, thermal imaging, night vision, training targets, field medical equipment, house arrest monitors, laser tags, electronic parking meters, multiple integrated laser engagement system, munitions and mines, ship sensors;

Utility:

Gas consumption meters, water consumption meters, and electric consumption meters;

Logistics & Supply Chain Management:

Radio-frequency identification devices (RFID), RFID readers;

Tracking: Asset tags, cargo container location beacons, transponders, transceivers; and

Devices: Smart price tags, smart shelving, handheld barcode scanner, barcode scanners, credit card readers, card readers, retail signage, hotel door locks;

Homeland Security:

Sensors: Occupancy, proximity, environmental, chemical/biological, motion, position; and

Metal detector wand;

Medical:

Implantable: cochlear implants, neural stimulators, pace makers, medication administration, defibrillator;

Body function monitors: pressure, temperature, respiration, blood oxygenation, insulin, hearing aid, pulse, EKG, heart, Holter;

Tracking tags: Patient, baby identification, assets, supplies, staff, medication, instruments;

Devices: Home healthcare equipment, ambulatory infusion pumps, blood analyzers, biofeedback systems, bone growth stimulators, thermometers, digital thermometers, stimulators, galvanic stimulators, muscle stimulators, pediatric scales; and

Agriculture—Livestock Tracking and Asset Tracking:

Tracking: livestock, asset, wildlife tracking devices; and

Equipment: cattle prods;

Automotive:

Automotive antennas, Automotive Audio Systems, Automotive Lighting, Automotive Video Systems, Computers, Processors, Controls, Switches, Electric Motors, Actuators, Ignition Systems, Starter Systems, Injection Systems, Powertrain Electronics, Radar Detectors, Proximity Detectors, Safety Systems, Security Systems, Sensors, Regulators, Distributors, Vehicle Control, Wiper Systems, Washer Systems, Radio, Video Systems, Entertainment Systems, Navigation Systems, GPS systems, Power Mirror Systems, Emission control systems;

Appliances:

Monitoring systems and control systems for major and small appliances including washing machines, dryers, refrigerators, freezers, coolers, air conditioners, humidifiers, dehumidifiers, air purifiers, air filters, fans, furnaces, water heaters, boilers, space heaters, sowing machines, ice makers, microwave ovens, convection ovens, ovens, toaster ovens, ranges, range hoods, cooktops, stoves, stovetops, crock pots, hot plates, dishwashers, garbage disposals, can openers, vacuum cleaners, blenders, mixers, food processors, irons, coffee makers, toasters, grills, hair dryers, electric tooth brushes, electric razors, electric drills, electric screwdrivers, chainsaws, lawnmowers, push mowers, riding mowers, trimmers, brush cutters, pruners, edgers, vending machines;

Ventilation, Heating, Air-Conditioning, and Commercial Refrigeration Equipment:

Monitoring systems, control systems;

Engine, Turbine, and Power Transmission Equipment;

Monitoring systems, control systems;

Other General Purpose Machinery Manufacturing:

Monitoring systems, control systems;

Telecommunications:

Monitoring systems, control systems;

Portable;

Aircraft:

Monitoring systems, control systems, actuator systems, sensors.

It should be noted that devices within a specific category may be applicable across multiple areas even if they are not specifically listed. (e.g., temperature sensors apply to Industrial and Building Automation and Appliances).

To retrofit or redesign the devices listed, it is possible to implement the described systems in numerous ways. It may be advantageous to leave the device design as-is including the existing power supply. As an example, a device may use non-rechargeable batteries to operate. The device will most likely have a protection circuit to prevent damage if the batteries are installed incorrectly. The protection mechanism is commonly a diode inline with the positive terminal of the battery. In this case, the RF power harvester **20** with or without the alternative power source **24** could be inserted, with an antenna, into the device. The power generated by the RF power harvester **20** (and the alternative power source **24**, if applicable) could be connected to the device after the protection mechanism described to avoid potential charging of a non-rechargeable battery.

Another way to configure the device's power system is to replace the non-rechargeable batteries with rechargeable batteries. In this instance, the output from the RF power harvester **20** (and alternative power source **24**) could be connected to either side of the protections device. If the connection is before the protection mechanism, the RF power harvester **20** (and the alternative power source **24**, if applicable) will recharge the battery and supply power to the device. If the connection is after the protection mechanism, the RF power harvester **20** (and the alternative power source **24**, if applicable) will supply power to the device and the battery will supply any extra power needed that could not be supplied by the RF power harvester **20** (and the alternative power source **24**, if applicable). It should be noted that the protection device in this case is not needed for proper operation. Its only function would be to protect the batteries from being installed incorrectly. An antenna could be contained inside or placed on the outside of the device.

Another configuration of the device's power system is to remove the existing batteries and install the RF power harvester **20** (and the alternative power source **24**) in the enclosure provided for the batteries. An antenna could be contained inside or placed on the outside of the device.

Yet another method of configuring the device's power system would be to reduce the number of batteries and replace them with the RF power harvester **20** (and the alternative power source **24**). In this case, the output from the system would be connected to the batteries in series or parallel depending on the original battery configuration. An antenna could be contained inside or placed on the outside of the device.

An additional option, would be to completely redesign the product and integrate the required circuitry and storage components into the device. This method is probably the most advantageous because it can fully take advantage of the benefits offered by the RF power harvester **20** (and the alternative power source **24**). An antenna could be contained inside or placed on the outside of the device.

If the RF power harvester **20** (and the alternative power source **24**) is used as a backup to the primary power supply, a switch could be implemented into the system in order to switch the RF power harvester **20** (and the alternative power source **24**) on when the primary source is lost. In this case, an antenna could be contained inside or placed on the outside of the device.

To show the flexibility of RF energy harvesting, several products were retrofitted to include an RF power harvester **20** (energy harvesting circuitry). These products include a wireless keyboard, a wall clock, and a desk calculator.

The wireless keyboard is an example of recharging and augmenting a battery to supply power to a device. This system is shown in FIG. **13**. The output from the regulation circuitry recharges the battery and supplies power to the keyboard. The battery is also used to supply power to the keyboard. The keyboard also includes a separate antenna for receiving power and for data communications. The antenna configuration can be seen in FIG. **55**.

The wall clock is an example of a direct powering system. The wall clock was retrofitted to include an RF power harvester **20** and the internal AA battery was removed. This system is shown in FIG. **2**. The wall clock did not need regulation, but did require a capacitor for storage to supply the pulse of power to move the second hand.

The calculator is an example of using RF energy harvesting with another energy harvesting technology. The calculator had an internal 1.5V coin cell battery and a small solar panel. The internal battery was removed, however, the solar panel was left intact. This system is shown in FIG. **60**. In this system, the calculator can receive power from both the solar panel and the RF power harvester **20** to eliminate the need for a battery.

As an additional example, an RF power harvester **20** (energy harvesting circuit) similar to the ones shown in U.S. Pat. No. 6,615,074 (FIGS. 8, 9, 12a, 12b, 13, 14), was connected in series with a 0.5V solar cell. Individually, the solar cell was able to provide 0.480V to a 10 kilo-ohm resistor, which was being used to simulate the core device **22** components. This corresponds to 23 microwatts. The RF power harvesting circuit by itself was able to provide 2.093V across the 10 kilo-ohm resistor when being supplied by 1 milliwatt of RF power. This corresponds to 438 microwatts. The two circuit outputs were then combined in series by connecting the output from the RF energy harvesting circuit to the ground of the solar cell. The output of the solar cell was then connected to the resistor. The other end of the resistor was connected to the ground of the RF energy harvesting circuit. The voltage across the resistor with the circuits connected as shown in FIG. **63** was 2.445V. This corresponds to 598 microwatts.

As can be seen, the combination of the two technologies produces a result higher than the addition of the individual powers. From this, it can be determined that the two technologies can cooperate in a way that produces favorable results.

In the example given, the solar cell produces current to supply the load and helps to bias the RF rectifying diodes, which allows the RF energy harvesting circuit to operate a higher efficiency. The solar cell also changes the impedance seen by the RF energy harvesting circuit, which produces a beneficial result. To be more specific, when examining the power output of the individual circuit (solar and RF power harvesting), the sum of the power captured by the individual circuits was 23 μW+438 μW=461 μW. However, when the two circuits are combined and are allowed to work in conjunction with one another, the output power becomes 598 μW. This result shows that combining the two power-harvesting technologies produces a 30 percent increase in the output power for this example. This same technique can be applied to multiple energy harvesting technologies to produce even greater output power. The equations for this example are shown below.

Individual Circuits

$$P_I=P_1+P_2+\ldots+P_N$$

Combined Circuits

$$P_C>P_I=P_1+P_2+\ldots+P_N$$

where P_I is the sum of the individual output powers

P_C is the output of the combined circuit

P_1 is the output power from the first power harvesting technology

P_2 is the output power from the second power harvesting technology

P_N is the output power from the N^{th} power harvesting technology

N is the number of power harvesting technologies or circuits.

As can be seen by the previous examples, RF energy harvesting can be used alone or in conjunction with alternative power sources to power a wide range of devices. The addition of RF energy harvesting technology to the device allows for increased battery life, increased functionality, or the removal of the primary battery.

Although the invention has been described in detail in the foregoing embodiments for the purpose of illustration, it is to be understood that such detail is solely for that purpose and that variations can be made therein by those skilled in the art without departing from the spirit and scope of the invention except as it may be described by the following claims.

The invention claimed is:

1. An apparatus, comprising:

a power harvester configured to convert energy received via an antenna to a DC voltage;

a power storage device configured to store a power associated with the DC voltage;

a power storage charger operatively coupled to the power harvester, the power harvester configured to provide an activation signal to the power storage charger to trigger storage of the power associated with the DC voltage at the power storage device;

a sensor configured to operate based on the DC voltage, the sensor configured to detect an environmental condition associated with a location of the sensor; and

an integrated circuit configured to operate based on the DC voltage, the integrated circuit configured to perform an application associated with the environmental condition detected by the sensor.

2. The apparatus of claim **1**, wherein:

the DC voltage is a first DC voltage, the integrated circuit is configured to operate based on the first DC voltage or a second DC voltage from a power source.

3. The apparatus of claim **1**, wherein:

the power associated with the DC voltage is a first power associated with the DC voltage; and

the integrated circuit is configured to produce a control signal that causes a second power associated with the DC voltage to be transferred from the power harvester to the integrated circuit.

4. The apparatus of claim **1**, wherein:

the sensor and the integrated circuit are each configured to operate based on the power associated with the DC voltage stored in the power storage device; and

the power harvester is configured to be disabled when the power associated with the DC voltage stored in the power storage device is sufficient for at least one of the sensor or the integrated circuit to operate.

5. The apparatus of claim **1**, wherein:

the power storage charger configured to control storage of the power associated with the DC voltage at the power storage device, the power storage charger configured to operate in a power pulse charging mode.

6. The apparatus of claim **1**, wherein:

the power storage charger configured to control storage of the power associated with the DC voltage at the power storage device, the power storage charger configured to operate in a power pulse charging mode or a power trickle charging mode.

7. The apparatus of claim **1**, wherein:

the power harvester is configured to receive at least one of an ambient radio frequency energy, a directed radio frequency energy, or an energy from a power source connected to the power harvester.

8. The apparatus of claim **1**, wherein:

the integrated circuit is configured to provide a deactivation signal to a power source when the power associated with the DC voltage is more than a predetermined power value, the deactivation signal configured to impede output of a power from the power source, the power source different from the power harvester and the power storage device.

9. The apparatus of claim **1**, wherein the power storage device is a first power storage device, the apparatus further comprising:

a second power storage device operatively coupled to the first power storage device and the integrated circuit, the second power storage device configured to store a portion of the power associated with the DC voltage and provide the portion of the power associated with the DC voltage to at least one of the sensor or the integrated circuit.

10. The apparatus of claim **1**, wherein the apparatus is part of a radio-frequency identification tag.

11. The apparatus of claim **1**, further comprising:

a display configured to operate, at least in part, based on the DC voltage.

12. The apparatus of claim **1**, wherein the sensor is at least one of a temperature sensor, a humidity sensor, a stress sensor, a strain sensor, a vibration sensor, an accelerometer, a level sensor, or a light sensor.

13. An apparatus, comprising:

a power harvester configured to convert energy received via an antenna to a DC voltage to produce a power associated with the DC voltage;

a regulator coupled to the power harvester and configured to regulate the power associated with the DC voltage, the regulator configured to transfer the regulated power to a power storage charger of the apparatus;

a power storage device configured to store power;

the power storage charger coupled to the power harvester and the regulator, the power storage charger configured to control the storage of the regulated power associated with the DC voltage to the power storage device; and

an integrated circuit configured to operate based on the power received from the power harvester or from the power storage device,

the power harvester configured to transfer power associated with the DC voltage to the regulator, the power harvester configured to transfer power associated with the DC voltage to the power storage device, the power harvester configured to transfer power associated with the DC voltage to the integrated circuit.

14. The apparatus of claim **13**, wherein:

the power storage charger is configured to operate in a power pulse charging mode or a power trickle charging mode.

15. The apparatus of claim **13**, further comprising:

a sensor configured to operate based on the power received from the power harvester or from the power storage device, the sensor configured to detect an environmental condition associated with a location of the sensor or a physical condition associated with an item coupled to the sensor, the integrated circuit configured to perform an application associated with the environmental condition or the physical condition detected by the sensor.

16. The apparatus of claim **13**, wherein:

the regulator is configured to send a control signal that causes the power harvester to vary an amount of the power associated with the DC voltage transferred from the power harvester to the regulator.

17. The apparatus of claim **13**, wherein:

the power storage device is configured to send a control signal that causes the power harvester to vary an amount of the power associated with the DC voltage transferred from the power harvester to the power storage device.

18. The apparatus of claim **13**, wherein the apparatus is part of a radio-frequency identification tag.

19. The apparatus of claim **13**, wherein the integrated circuit is in communication with a sensor configured to detect an environmental condition associated with a location of the sensor, the sensor being at least one of a temperature sensor, a humidity sensor, a stress sensor, a strain sensor, a vibration sensor, an accelerometer, a level sensor, or a light sensor.

20. The apparatus of claim **13**, further comprising:

a display configured to operate, at least in part, based on the DC voltage.

* * * * *

US 20110101789A1

(19) **United States**

(12) **Patent Application Publication**

Salter, JR. et al.

(10) **Pub. No.: US 2011/0101789 A1**

(43) **Pub. Date: May 5, 2011**

(54) RF POWER HARVESTING CIRCUIT

(76) Inventors: **Thomas Steven Salter, JR.**, Ellicott City, MD (US); **George M. Metze**, Millersville, MD (US); **Neil Goldsman**, Takoma Park, MD (US); **Kwangsik Choi**, College Park, MD (US); **Yves Ngu**, Essex, VT (US); **Zeynep Dilli**, Hyattsville, MD (US); **Martin Peckerar**, Silverspring, MD (US); **Li Bo**, College Park, MD (US)

(21) Appl. No.: **12/628,585**

(22) Filed: **Dec. 1, 2009**

Related U.S. Application Data

(60) Provisional application No. 61/118,808, filed on Dec. 1, 2008, provisional application No. 61/119,848, filed on Dec. 4, 2008.

Publication Classification

(51) **Int. Cl.**
H01F 38/14 (2006.01)

(52) **U.S. Cl.** **307/104**

(57) ABSTRACT

Provided is an RF power harvesting circuit with improved sensitivity to RF energy. The RF power harvesting device includes an inductor, a first capacitor connected to the inductor, a first MOSFET connected to a first node, and a second MOSFET connected to the first node. The inductor or the first capacitor are connected to the first node.

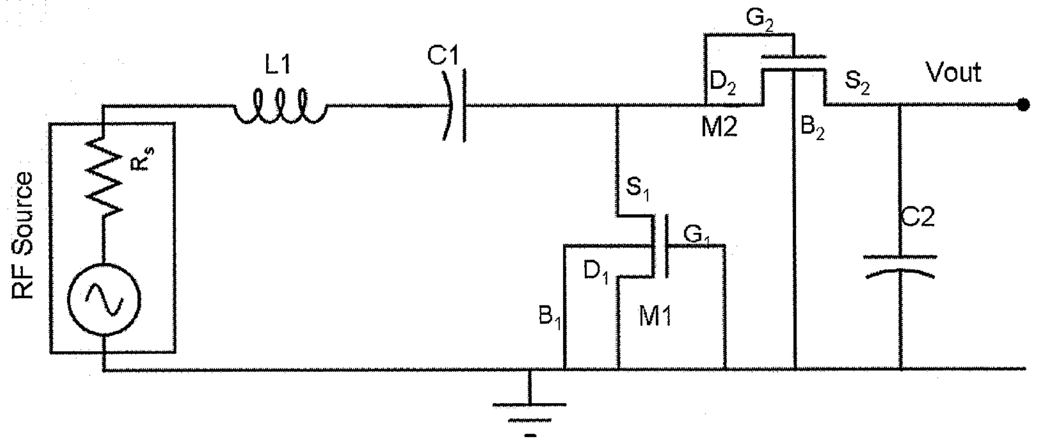

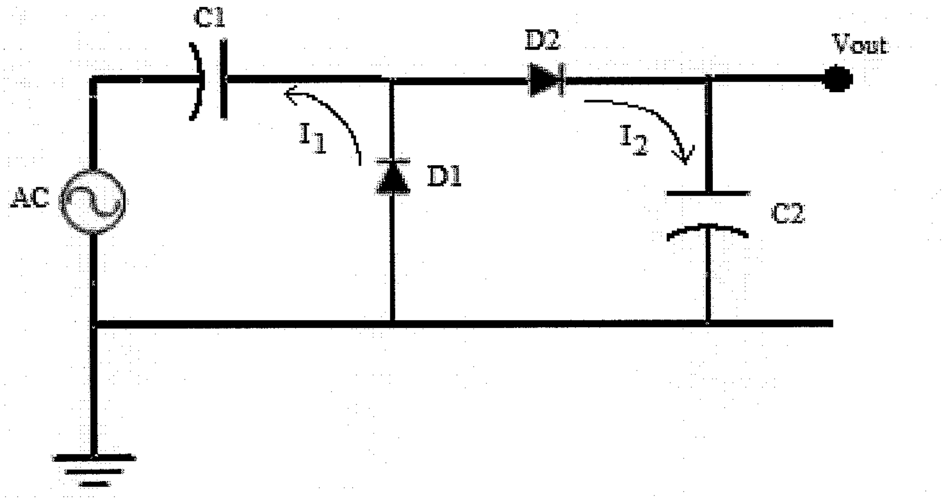
C1
D2
Vout
I_1
I_2
D1
AC
C2

FIG. 1

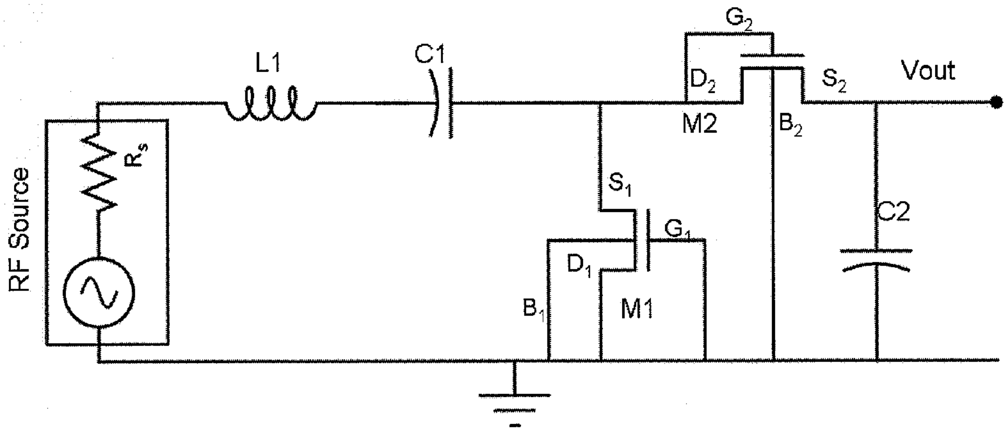
G_2
L1
C1
D_2
S_2
Vout
M2
B_2
RF Source
R_s
S_1
C2
G_1
D_1
B_1
M1

FIG. 2

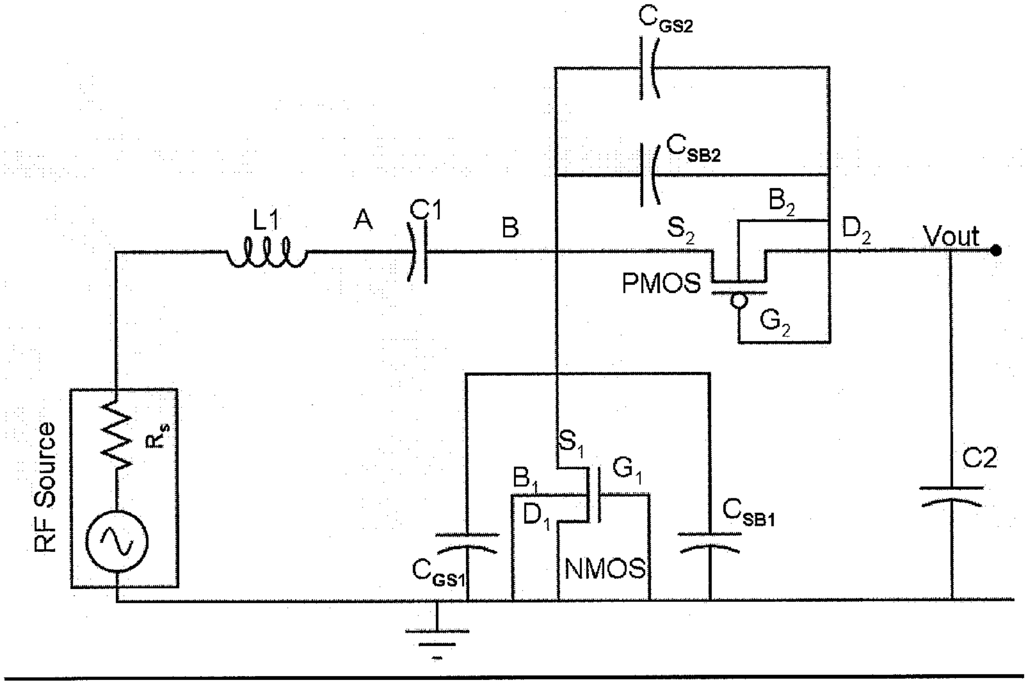
CGS2
CSB2
L1
A
C1
B
S2
B2
D2
Vout
PMOS
G2
RF Source
Rs
S1
G1
B1
D1
CGS1
NMOS
CSB1
C2

FIG. 3

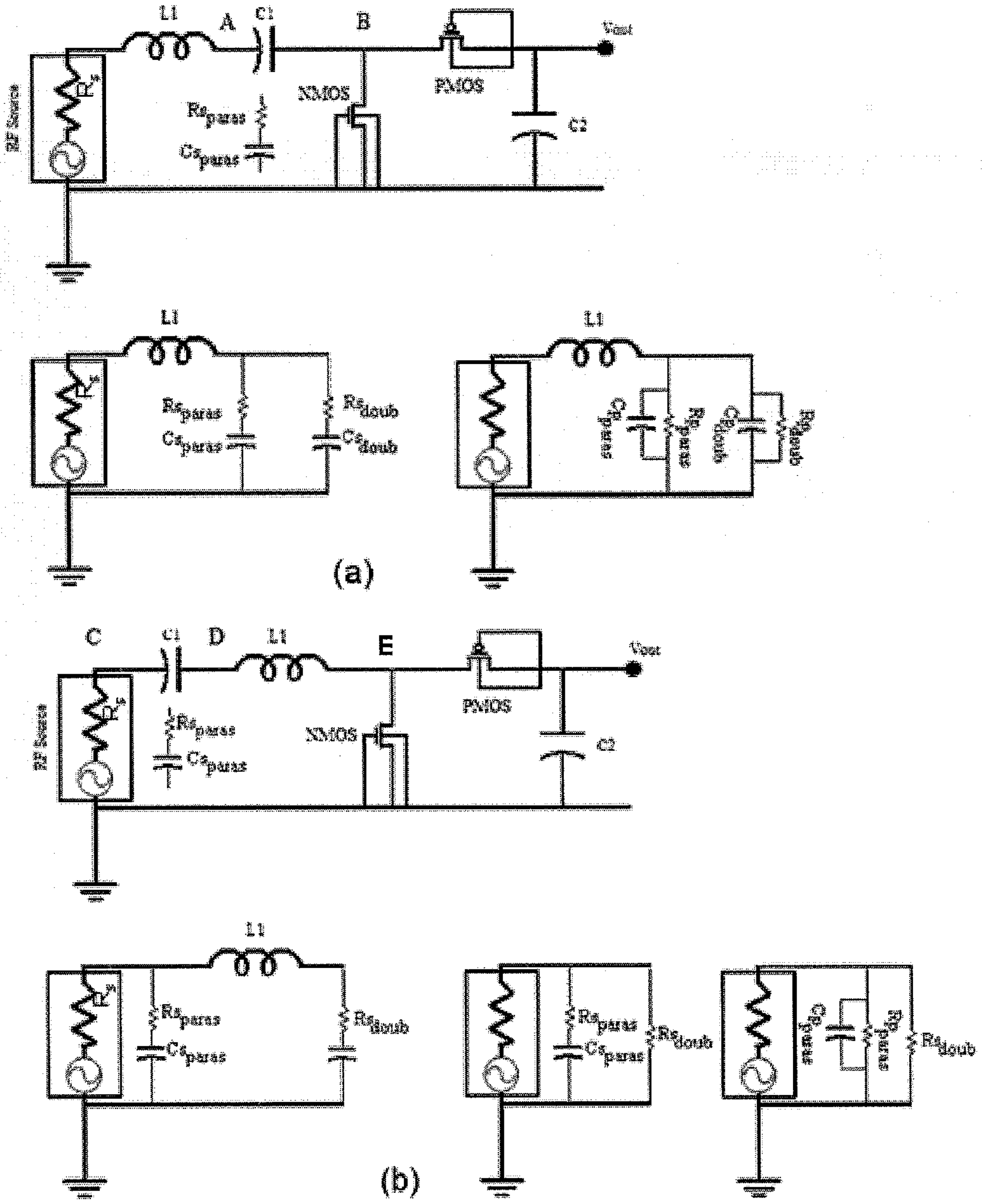
L1
A
C1
B
Vout
RF Source
R_s
NMOS
PMOS
Rs_{paras}
Cs_{paras}
C2
L1
R_s
Rs_{paras}
Cs_{paras}
Rs_{doub}
Cs_{doub}
L1
Cp_{paras}
Rp_{paras}
Cp_{doub}
Rp_{doub}
(a)
C
C1
D
L1
E
Vout
RF Source
R_s
Rs_{paras}
Cs_{paras}
NMOS
PMOS
C2
L1
R_s
Rs_{paras}
Cs_{paras}
Rs_{doub}
Rs_{paras}
Cs_{paras}
Rs_{doub}
Cp_{paras}
Rp_{paras}
Rs_{doub}
(b)

FIG. 4(a) and 4(b)

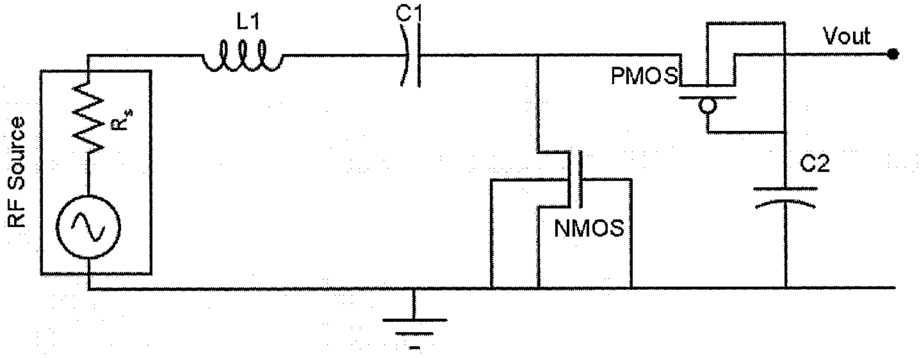
L1
C1
Vout
PMOS
RF Source
R_s
C2
NMOS

FIG. 5

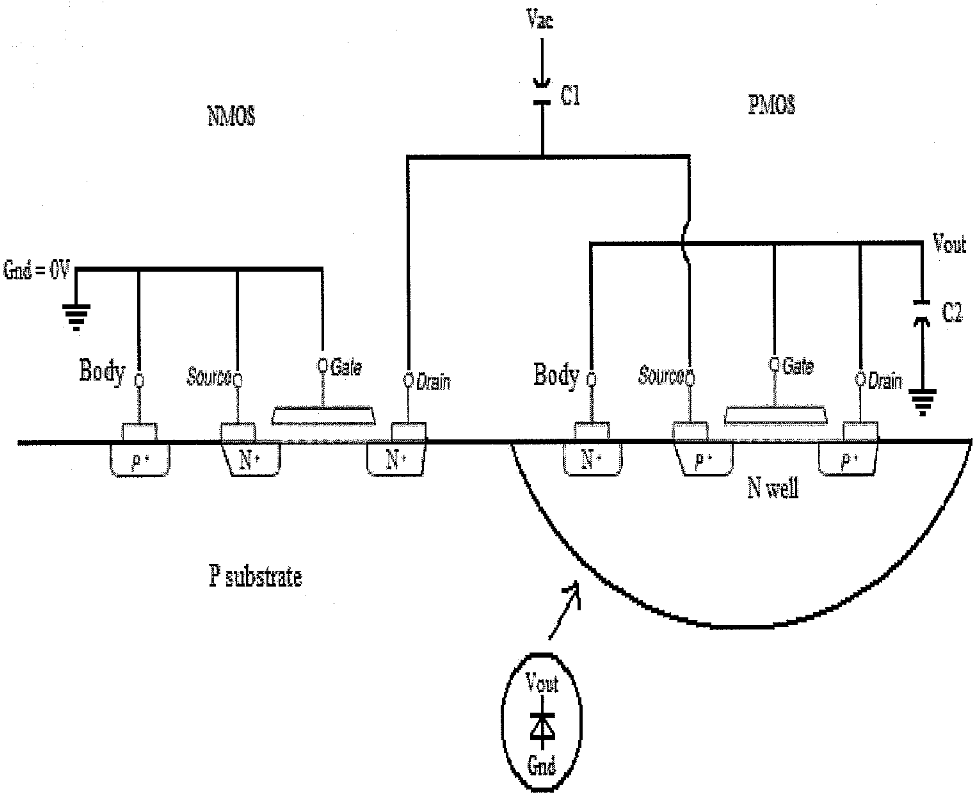
Vac
C1
NMOS
PMOS
Vout
Gnd = 0V
C2
Body
Source
Gate
Drain
Body
Source
Gate
Drain
P+
N+
N+
N+
P+
P+
N well
P substrate
Vout
Gnd

FIG. 6

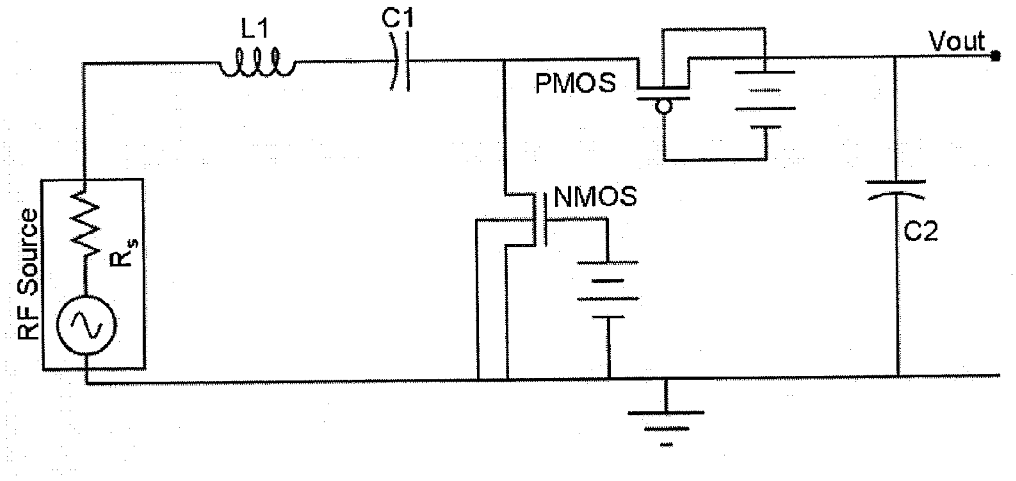
L1
C1
Vout
PMOS
NMOS
C2
RF Source
R_s

FIG. 7

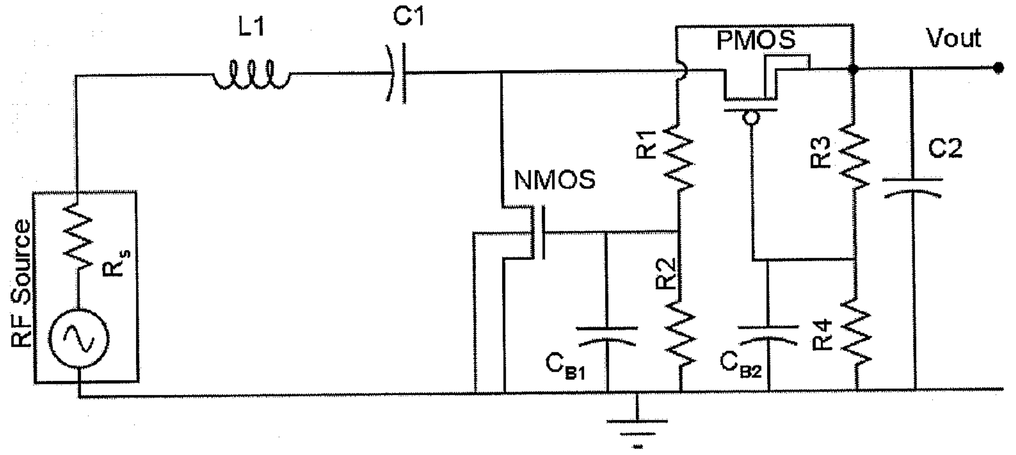
L1
C1
PMOS
Vout
R1
R3
C2
NMOS
R2
R4
RF Source
R_s
C_{B1}
C_{B2}

FIG. 8

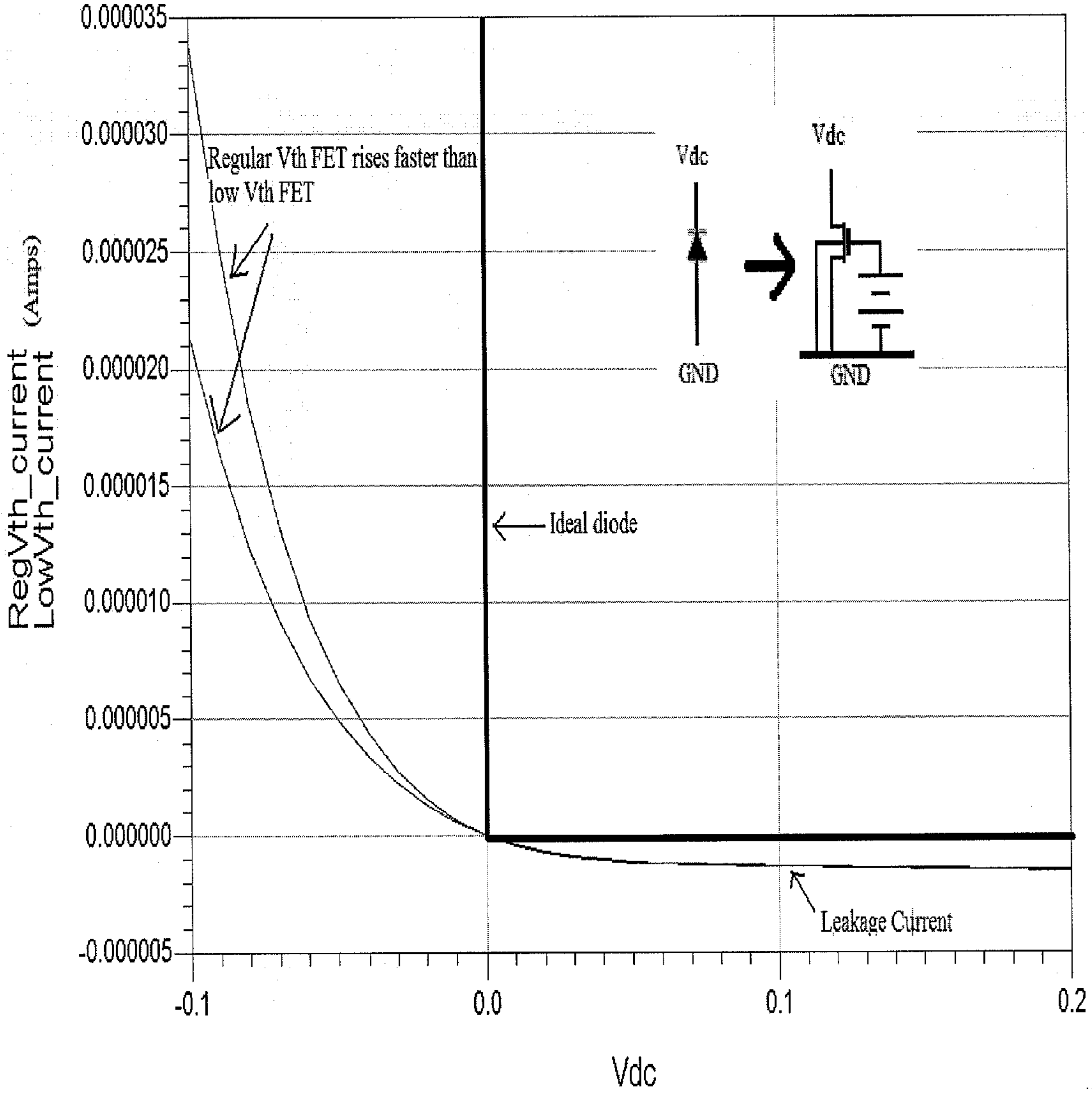
0.000035
0.000030
0.000025
0.000020
0.000015
0.000010
0.000005
0.000000
-0.000005
RegVth_current
LowVth_current
(Amps)
-0.1
0.0
0.1
0.2
Vdc
Regular Vth FET rises faster than low Vth FET
Ideal diode
Leakage Current
Vdc
GND
Vdc
GND

FIG. 9

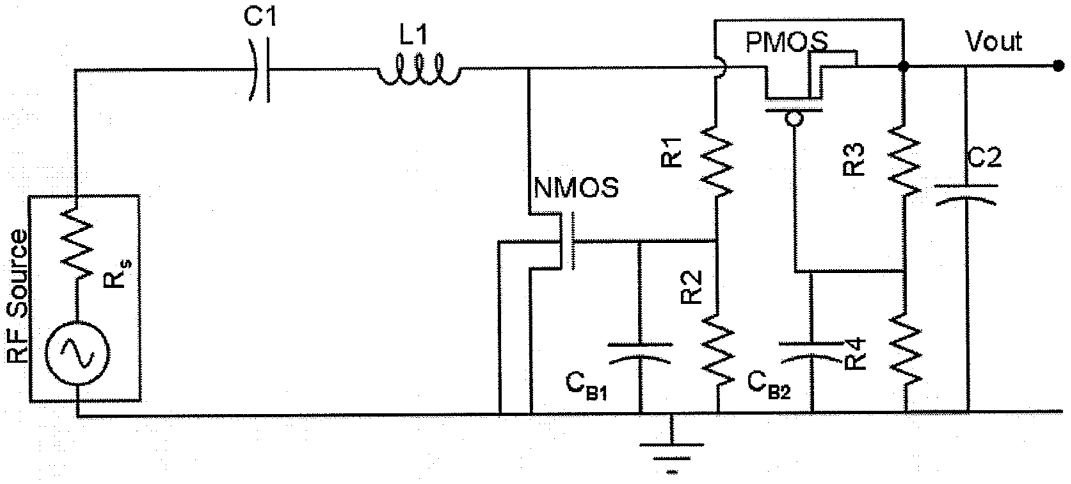
C1
L1
PMOS
Vout
R1
NMOS
R3
C2
RF Source
R_s
R2
C_{B1}
C_{B2}
R4

FIG. 10

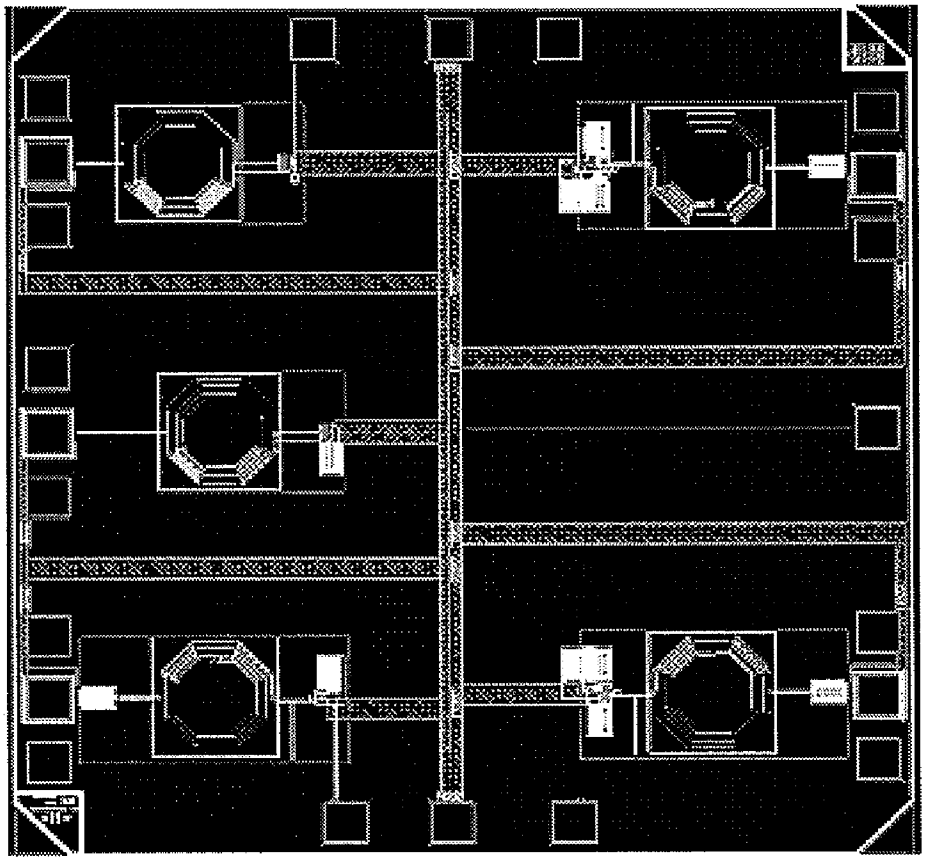

FIG. 11

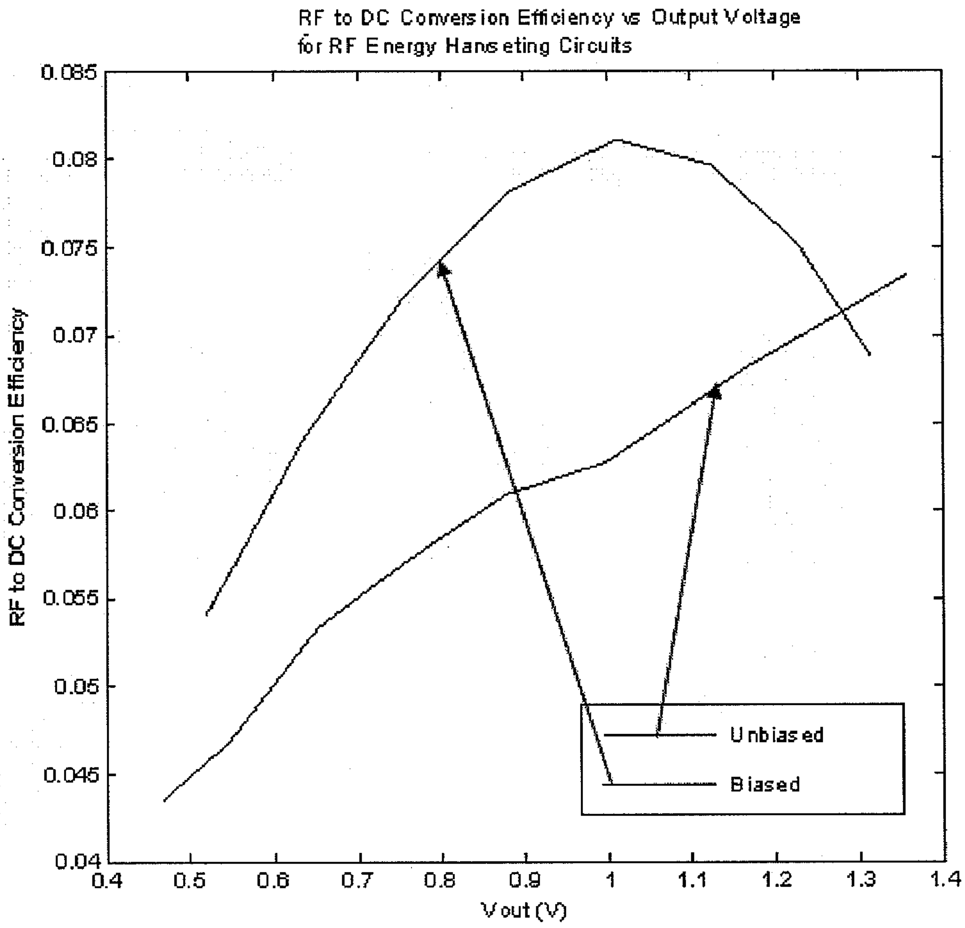
RF to DC Conversion Efficiency vs Output Voltage
for RF Energy Harvseting Circuits
RF to DC Conversion Efficiency
0.085
0.08
0.075
0.07
0.065
0.06
0.055
0.05
0.045
0.04
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
1.4
Vout (V)
Unbiased
Biased

FIG. 12

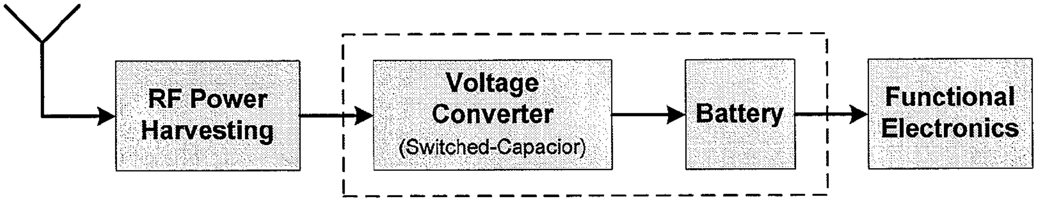
RF Power Harvesting
Voltage Converter
(Switched-Capacior)
Battery
Functional Electronics

FIG. 13

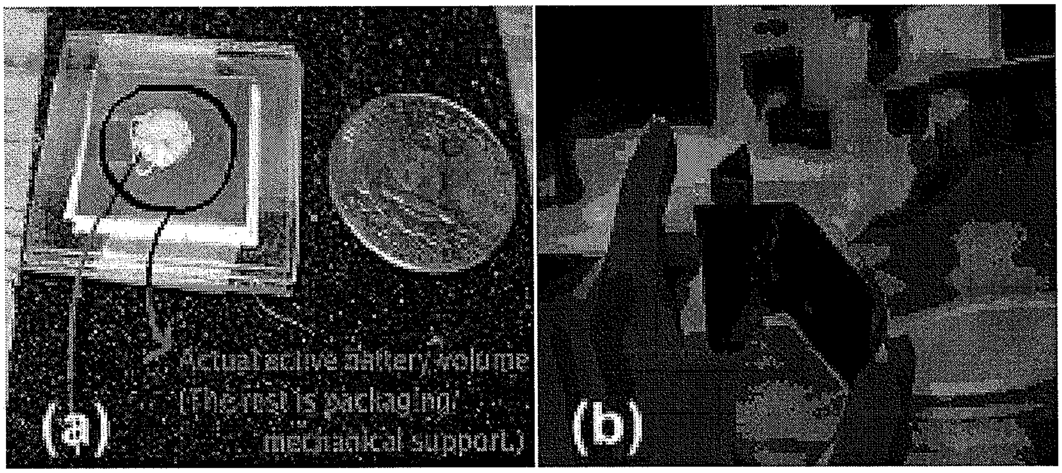
Actual active battery volume
(The rest is packaging/
mechanical support.)
(a)
(b)

FIGS. 14(a) and 14(b)

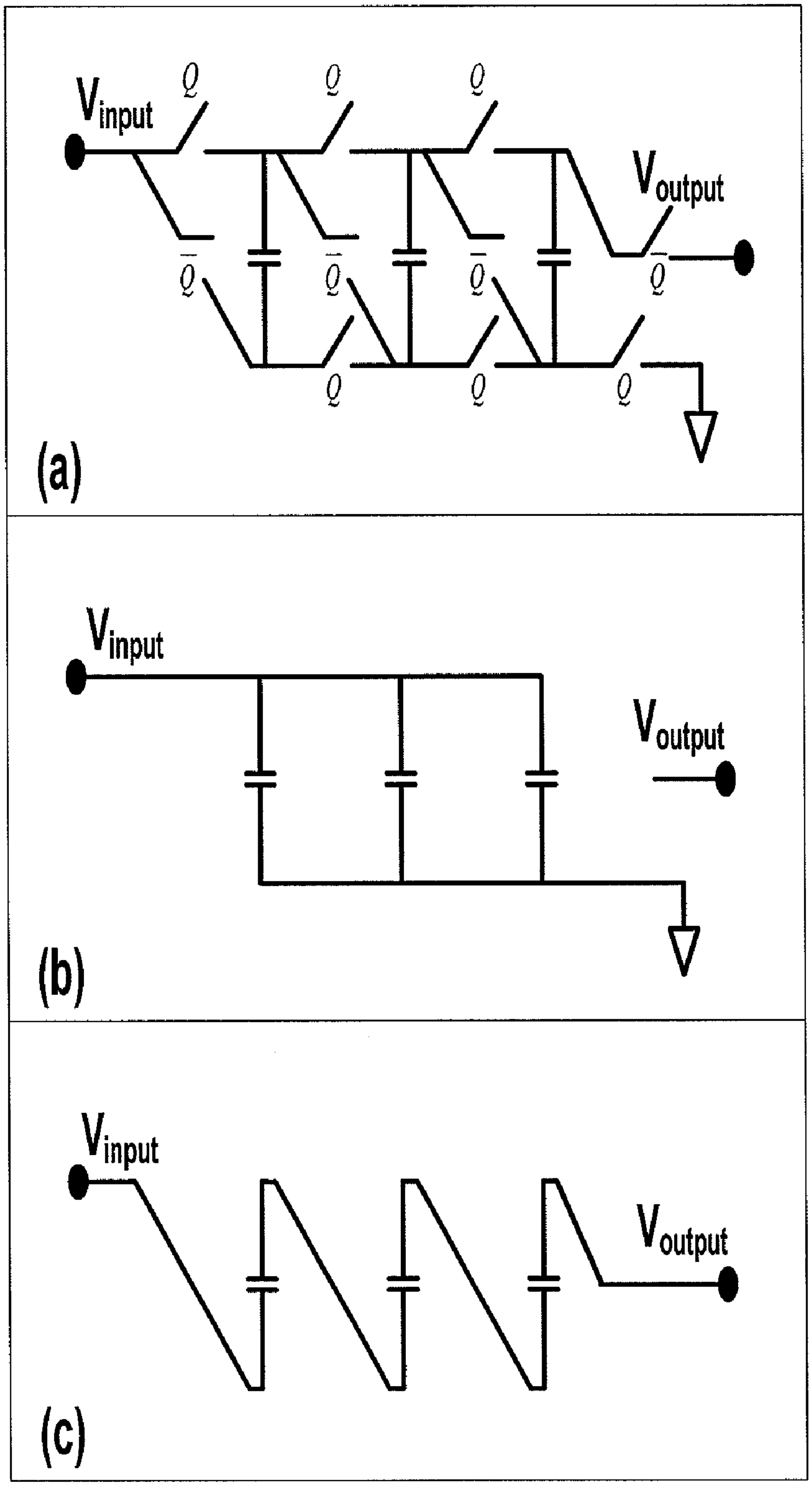
V_{input}
Q
Q
Q
V_{output}
$\bar{Q}$
$\bar{Q}$
$\bar{Q}$
$\bar{Q}$
Q
Q
Q
(a)
V_{input}
V_{output}
(b)
V_{input}
V_{output}
(c)

FIG. 15

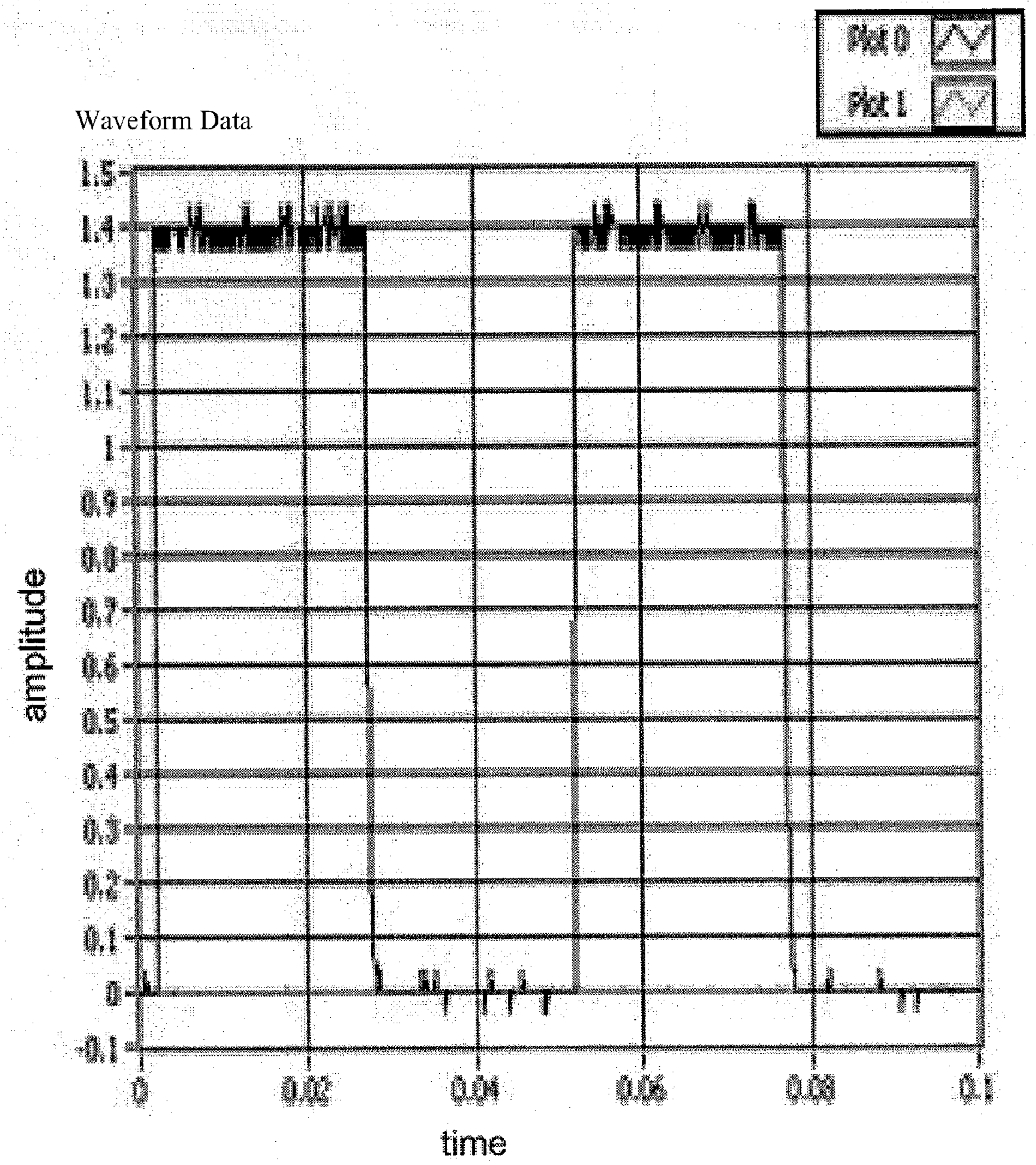
Waveform Data
Plot 0
Plot 1
amplitude
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
0
0.02
0.04
0.06
0.08
0.1
time

FIG. 16

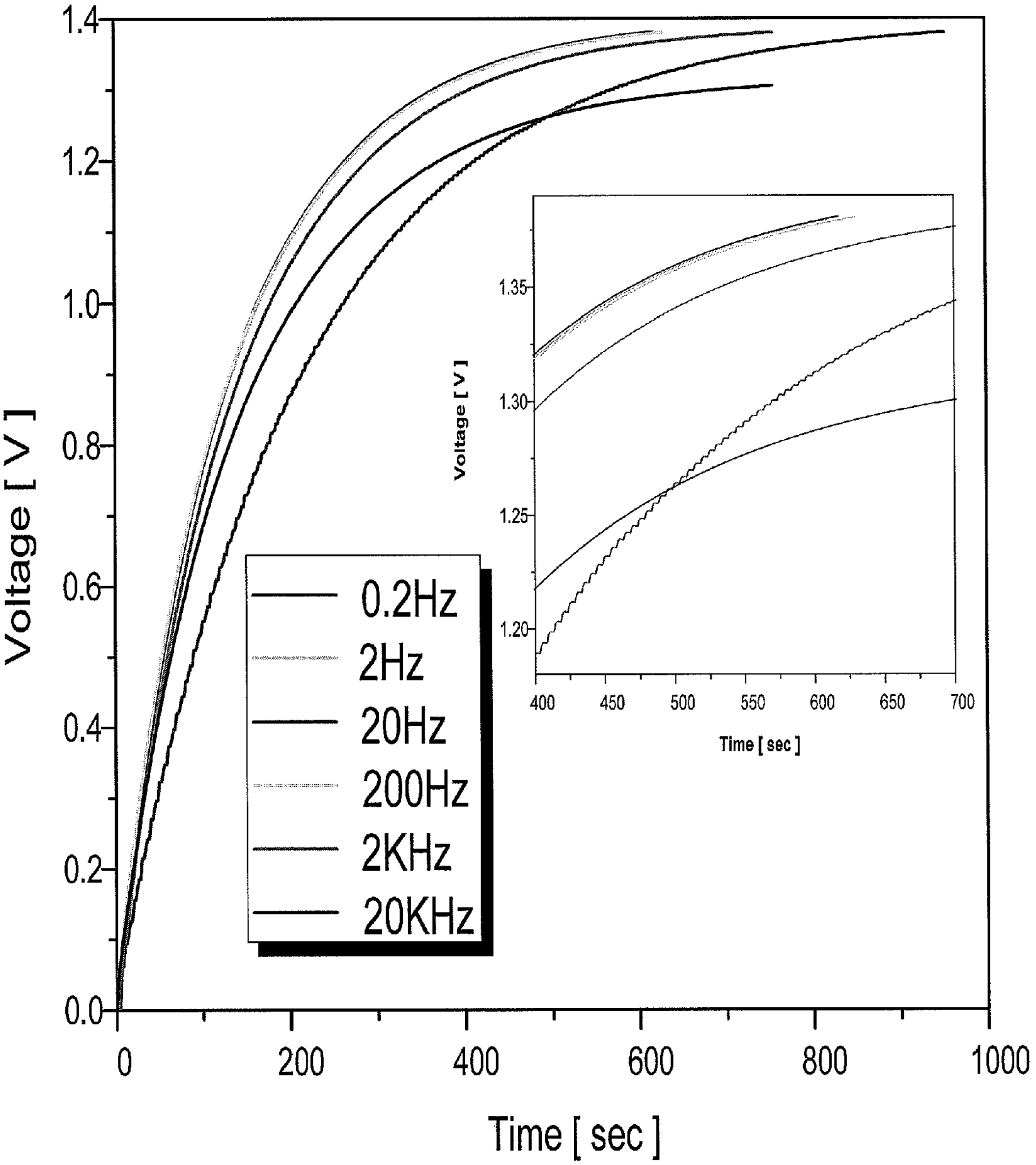
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Voltage [V]
0
200
400
600
800
1000
Time [sec]
0.2Hz
2Hz
20Hz
200Hz
2KHz
20KHz
1.35
1.30
1.25
1.20
Voltage [V]
400
450
500
550
600
650
700
Time [sec]

FIG. 17

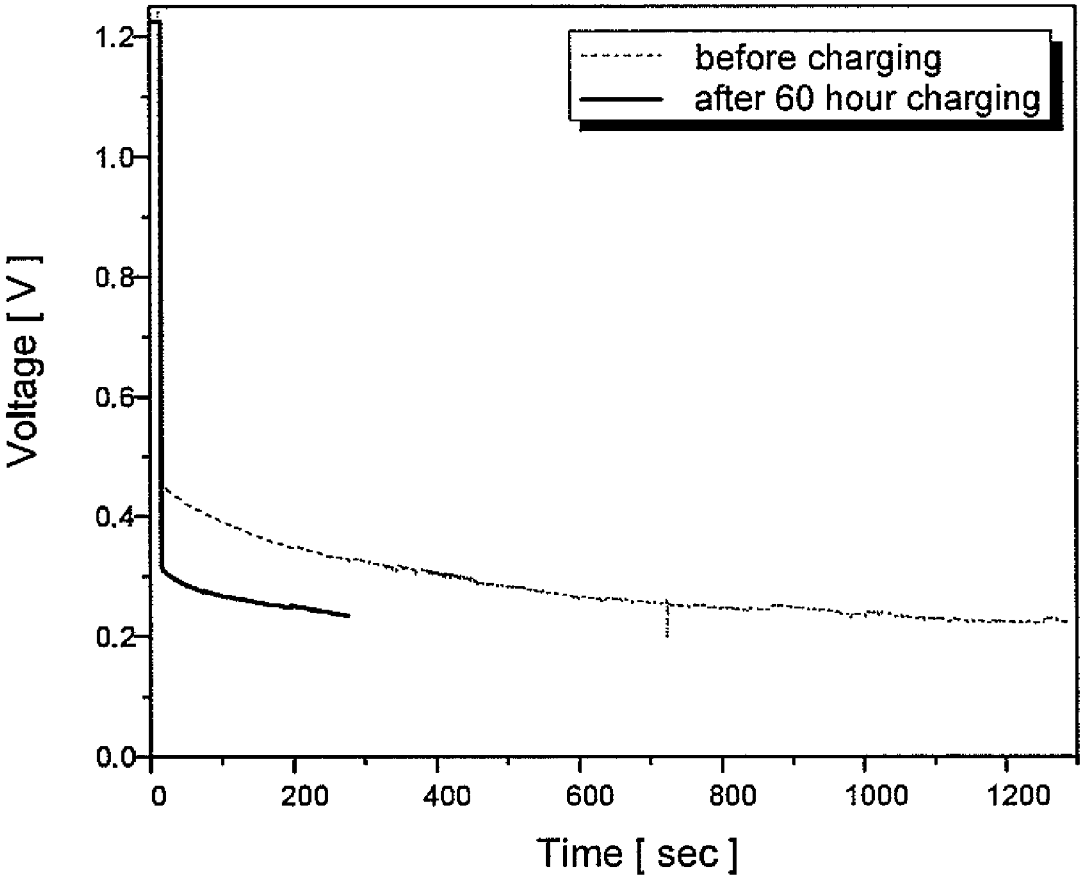
before charging
after 60 hour charging
Voltage [V]
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Time [sec]
0
200
400
600
800
1000
1200

FIG. 18

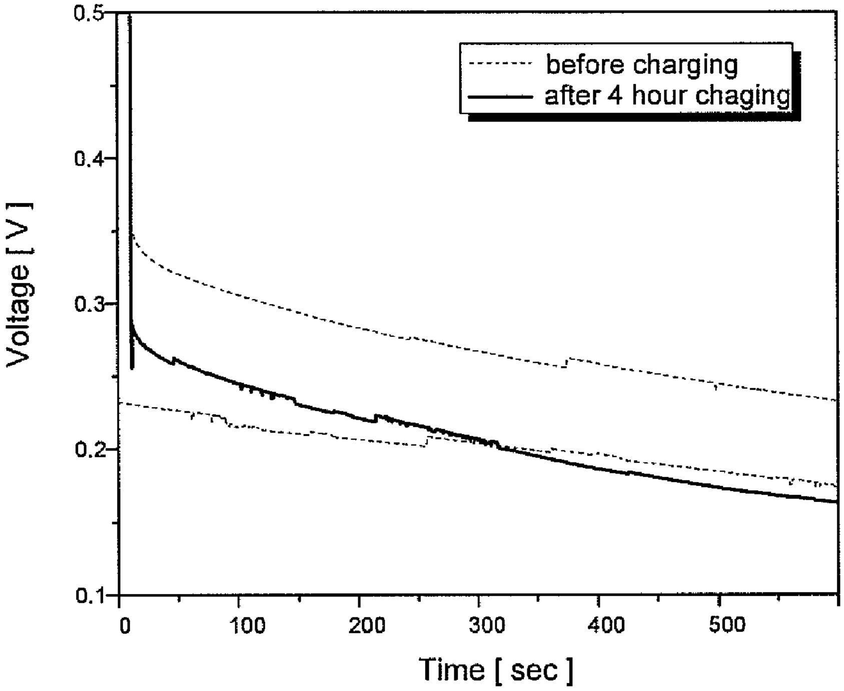
before charging
after 4 hour chaging
Voltage [V]
0.5
0.4
0.3
0.2
0.1
0
100
200
300
400
500
Time [sec]

FIG. 19

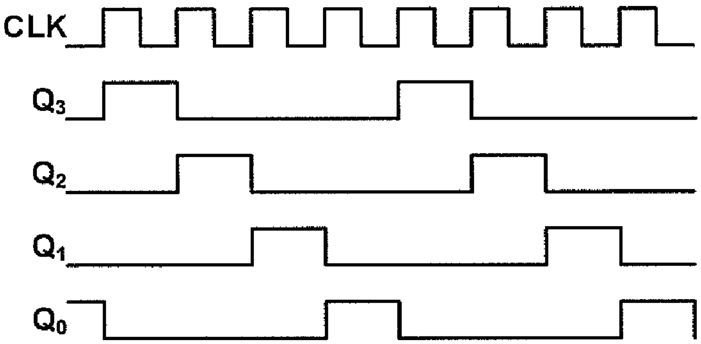
CLK
Q_3
Q_2
Q_1
Q_0

FIG. 20

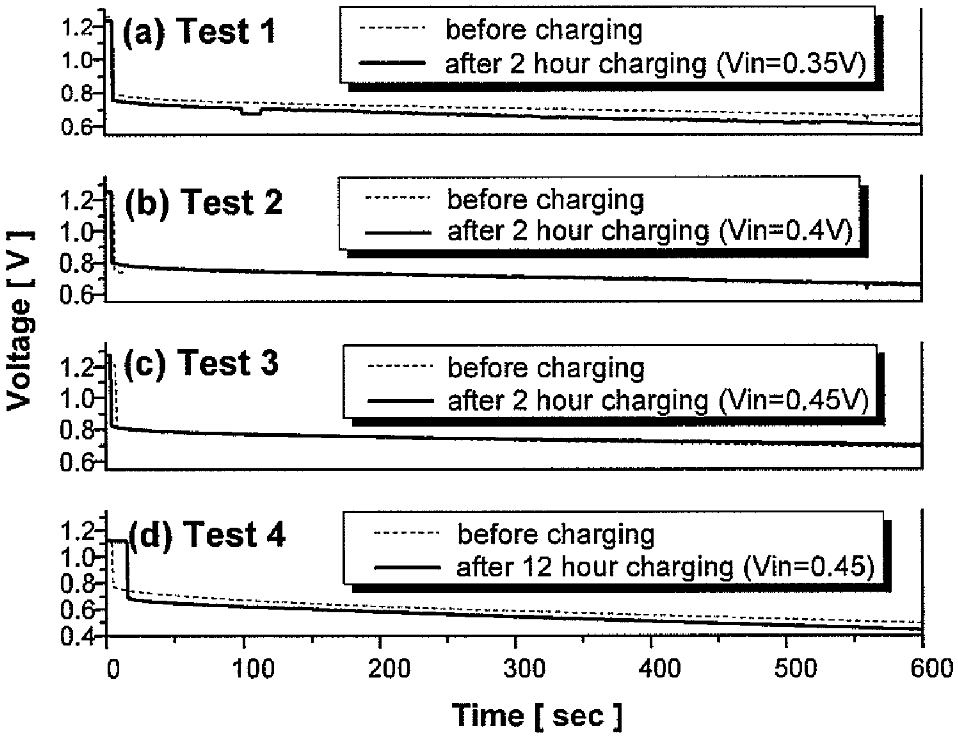
(a) Test 1
before charging
after 2 hour charging (Vin=0.35V)
(b) Test 2
before charging
after 2 hour charging (Vin=0.4V)
(c) Test 3
before charging
after 2 hour charging (Vin=0.45V)
(d) Test 4
before charging
after 12 hour charging (Vin=0.45)
Voltage [V]
1.2
1.0
0.8
0.6
0.4
0
100
200
300
400
500
600
Time [sec]

FIG. 21

RF POWER HARVESTING CIRCUIT

CROSS-REFERENCE TO RELATED PATENT APPLICATION

[0001] This application claims priority from U.S. Provisional Application No. 61/118,808 filed on Dec. 1, 2008, and U.S. Provisional Application No. 61/119,848 filed on Dec. 4, 2008, the disclosures of which are incorporated herein by reference in their entirety.

STATEMENT REGARDING FEDERALLY SPONSORED R&D

[0002] This invention was made with support under H9823004C0490 awarded by the National Security Agency. The government has certain rights in the invention.

BACKGROUND

[0003] 1. Technical Field

[0004] The present invention relates to an RF power harvesting circuit design, and more particularly, to an RF power harvesting circuit design which efficiently harvests power from RF energy.

[0005] 2. Description of the Related Art

[0006] Electronic devices are ubiquitous throughout the world today, with immeasurable applications and devices permeating all aspects of society. Energy that would have been wasted by such devices can now be recycled. For example, in some related art electronic devices (e.g., a mobile device) that generate radio frequency (RF), the RF is harvested back into the electronic device. All electronic devices require power to operate. Ambient energy in any form is an attractive source of power, particularly as power becomes more costly, or in areas where it is scarce. Ambient energy is also desirable where there is a need for electronic devices to operate for longer times between connection to readily available sources of power, or when there is a need to remotely recharge electronic devices, or to improve the overall efficiency of electronic devices. Of course, instead of scavenging for the ambient RF energy, intentional beaming of RF energy to a target device is also possible.

[0007] In this context, radio frequency (RF) energy represents an attractive source of power. Many electronic devices (e.g., communications devices) emit RF energy. For example, indoor power densities greater than 0.5 uW/cm^2 can be detected even a kilometer away from an FM radio tower. Comparable power densities can be detected at higher frequencies both domestically and internationally including GSM and ISM bands. Harvesting this energy from the environment can provide many benefits to electronic circuit and device designers. These benefits include, but are not limited to, extending operational life of electronic devices, providing new benefits such as remote recharging, reducing size of the electronic devices, and improving overall device efficiency.

[0008] RF power harvesting devices/RF energy scavenging devices have been constructed to capture this energy. An example of a related art power harvesting circuit is shown in FIG. **1**. FIG. **1** illustrates the basic Villard voltage doubler circuit. An intuitive understanding of this circuit can be gained by first examining what occurs when current flows in the direction of I_1. The diode D_2 blocks the flow of current through the capacitor C_2.

[0009] Therefore, all of the current goes across the capacitor C_1. This charges the capacitor C_1 up to roughly the same level as the peak of the AC voltage. Once the up-swing of the AC cycle (I_2) has been reached, the diode D_1 turns off, the diode D_2 turns on, and the voltage across both the AC source and C_1 drops across the capacitor C_2, charging it to approximately twice the peak voltage of the AC signal.

[0010] The related art RF scavenging circuits, however, require a minimum or threshold amount of incident RF energy to "turn on" and begin providing useful energy to other circuits and devices. Such devices can be devices that immediately perform a function or energy storage devices such as batteries and capacitors. Incident energy that falls below the minimum threshold is not captured, thereby reducing the efficiency and effectiveness of the scavenging circuit. Thus, there is a need for an RF power harvesting device that can capture and utilize ambient energy that falls below the minimum threshold for turning on the device.

SUMMARY

[0011] Embodiments of the disclosed RF energy harvesting circuit improve the sensitivity of RF energy harvesting circuits over the related art power harvesting circuits.

[0012] According to an aspect of the present invention, there is provided an RF power harvesting device including an inductor, a first capacitor connected to the inductor, a first MOSFET connected to a first node, and a second MOSFET connected to the first node, and the inductor or the first capacitor are connected to the first node.

[0013] In the RF power harvesting device, values of the first MOSFET and the second MOSFET are such that intrinsic capacitances of the first MOSFET and the second MOSFET and the inductor constitute a substantially resonant circuit during operation of the RF power harvesting device.

[0014] The RF power harvesting device of claim **1** further includes a second capacitor connected to a second node, and the second MOSFET is connected to the second node, wherein the second MOSFET is a PMOSFET.

[0015] The RF power harvesting device further includes a receiver which detects and receives ambient RF energy, and the first capacitor and the receiver are connected at a third node, the first capacitor and the inductor are connected in series at a fourth node, and the inductor is connected to the first node.

[0016] The RF power harvesting device further includes a first set of resisting elements connected to a first gate of the first MOSFET to apply a first DC bias to the first gate, and a second set of resisting elements connected to a second gate of the second MOSFET to apply a second DC bias to the second gate.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] The above and other features and advantages of the present invention will become more apparent by describing in detail exemplary embodiments thereof with reference to the attached drawings in which:

[0018] FIG. **1** illustrates a single stage of a related art Villard voltage doubler circuit.

[0019] FIG. **2**. illustrates a power harvesting device, according to an exemplary embodiment of the present invention.

[0020] FIG. **3** illustrates a power harvesting device, according to an exemplary embodiment of the present invention.

[0021] FIGS. **4**(*a*)-(*b*) illustrate different placements of a matching inductor and a DC blocking capacitor and the associated lossy parasitic capacitance and resistance, according to an exemplary embodiment of the present invention.

[0022] FIG. **5** illustrates a layout of a power harvesting device, according to an exemplary embodiment of the present invention.

[0023] FIG. **6** illustrates a power harvesting device, according to an exemplary embodiment of the present invention.

[0024] FIG. **7** illustrates biasing the gate of diode-connected MOSFETs, according to an exemplary embodiment of the present invention.

[0025] FIG. **8** illustrates a circuit design implementing sacrificial current biasing, according to an exemplary embodiment of the present invention.

[0026] FIG. **9** illustrates an ideal diode response of a diode connected MOSFET with regular FETs and a diode connected MOSFET using low threshold voltage FETs.

[0027] FIG. **10** shows a power harvesting device, according to an exemplary embodiment of the present invention.

[0028] FIG. **11** illustrates the layout of RF harvesting circuits to form a silicon integrated circuit, the layout including five RF power harvesting circuits, which are the physical implementations of the harvesting circuits shown in the circuit schematics shown in other figures, according to an exemplary embodiment of the present invention.

[0029] FIG. **12** illustrates a comparison of measured RF to DC conversion efficiency versus output voltage between a power harvesting device with sacrificial biasing, and a power harvesting device without biasing.

[0030] FIG. **13** illustrates a self-powered system.

[0031] FIGS. **14**(*a*)-(*b*) illustrate different types of batteries implemented in the system shown in FIG. **13**.

[0032] FIG. **15** illustrates a schematic of a switched capacitor DC-DC converter.

[0033] FIG. **16** illustrates a three stage converter operation for 0.35V input at 20 Hz switching using 100 uF external capacitors.

[0034] FIG. **17** illustrates the frequency dependency of a 1000 uF capacitor charging with 0.35V input and three 100 uF external storing capacitors.

[0035] FIG. **18** illustrates the discharging test results for 60 hour charging (1 KΩ) load, 0.2 Hz clock signal) with a rigid battery.

[0036] FIG. **19** illustrates the discharging test results for 4 hour charging (1 KΩ load, 25 Hz clock signal) using two chips with a rigid battery.

[0037] FIG. **20** illustrates a switching signal generated by a ring counter using four flip-flops.

[0038] FIG. **21** illustrates four converter charging results—(a) Vin=0.35V for a rigid battery, (b) Vin=0.4V for a rigid battery, (c) Vin=0.45V for a rigid battery, and (d) Vin=0.45V for a flexible battery.

DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS

[0039] The present invention will now be described more fully with reference to the accompanying drawings, in which exemplary embodiments of the invention are shown.

[0040] One embodiment of the present invention provides a modified power matched Villard voltage doubler circuit, compatible with modern complementary metal-oxide-semiconductor (CMOS) processes, which demonstrates the ability to harvest RF electromagnetic energy that is available at power levels as low as in the microwatt range. Revisions to the Villard voltage doubler have been made and accessory circuitry has been added to provide it with the ability to harvest energy from power levels as extremely low as tens of μWatts (micro) low power levels. In the related art, relatively large values of input power of hundreds of μWatts or even higher are required to utilize a Villard voltage doubler topology. The reason for this it that the general Villard voltage doubler topology utilizes rectifying diodes which need to be turned on in order to transform AC currents and voltages into DC currents and voltages, so that their intrinsic energy can be stored. However, turning on these diodes requires several tenths of a volt, and a relatively large amount of input power is typically required to generate these voltages.

[0041] The present invention provides several modifications to the Villard voltage doubler topology, which modifications allow it to harvest energy from very low power levels and go beyond the limitation previously set by the threshold level of the rectifying diodes. For instance, one embodiment of the present invention provides resonant reactive elements are incorporated in a CMOS circuit that allows for the generation of relatively large voltages, without the need for large power, in order to turn on the rectifying diodes. The Villard circuit topology is implemented in an actual CMOS process making it realizable for fabrication as an integrated circuit in modern technology. Certain parasitic elements of the CMOS process are utilized to increase voltage levels and improve circuit performance (typically, parasitic elements limit circuit performance).

[0042] Application specific circuit layout designs are disclosed so that the effect of other parasitic elements would be minimized. Diode connected CMOS devices are used instead of standard diodes to implement the Villard topology. Furthermore, in one embodiment, instead of using an n-type metal-oxide-semiconductor field-effect transistor (NMOSFET), one of the diode functions is implemented using a p-type MOSFET (PMOSFET) polarity. This embodiment mostly eliminates the body effect and allows for the diode connected MOSFETs to be turned on at a lower voltage, thereby giving rise to a lower threshold power in order to obtain AC to DC conversion and allow for energy storage at lower power levels. In another embodiment, a set of sacrificial bias resistors is used to increase the gate voltage of the MOSFETs, and thereby turn them on at a lower voltage which allows for further sensitivity and increases energy harvesting. In addition, in yet another embodiment, circuit blocks are stacked in series to generate even high DC voltage levels from microwatt levels of electromagnetic RF energy.

[0043] It is desirable to integrate the power harvesting circuits of the disclosed embodiments onto CMOS integrated circuits. To accomplish this, the diodes can be replaced with diode connected MOSFETS M**1** and M**2**, as illustrated in FIG. **2**. This modification has an advantage in that it can be naturally fabricated using a CMOS process. Furthermore, using diode connected MOSFETs gives added flexibility for establishing an adjustable effective diode turn on voltage.

[0044] An inductor L**1** can be added near the input power source (RF source) as shown in FIG. **2**. This helps give rise to a matching circuit. In addition to the standard concept of matching, which allows for maximum power transfer from a fixed source impedance, the inductor works in concert with the parasitic intrinsic capacitances usually associated with a MOSFET gate to generate a very large voltage across the reactive components (discussed below in further detail). The generation of such a large voltage is one advantage of using the power harvesting circuit shown in FIG. **1**. Specifically, the generation of the large voltage facilitates the turning on of the CMOS diode connected devices at low incident power levels in order to achieve efficient AC to DC conversion, and ultimately energy storage. Thus, instead of the parasitic intrinsic capacitances functioning to reduce the frequency response and be detrimental to CMOS circuit performance, their reactive nature when working in concordance with the added input inductor L**1** allows for generation of large voltages which will ultimately make the power harvesting device more sensitive to harvesting RF energy. This circuit has a complex impedance that mainly consists mainly of a reactive part.

[0045] A. Modification to Matching Inductor and use of Intrinsic Capacitance for Large Voltage Generation and Diode Turn-On.

[0046] As mentioned previously, in order to be able to utilize the Villard voltage doubler topology, it is necessary to achieve voltages sufficiently large to turn on the diodes (or diode connected MOSFETs) D**1** and D**2**. However, this is difficult if the power of the input signal is very low, in the microwatt range, for example. To overcome this problem, reactive components are utilized in the design which are able to accumulate relatively large amounts of AC power over time, and thus generate voltages that are sufficiently large to turn on the diodes, and thus harvest RF energy even at power levels as low as microwatts.

[0047] The reactive elements that generate the large voltages for turn on are shown in FIG. **3** (e.g., the inductor L**1**, and the intrinsic MOSFET capacitors C_{GS1}, C_{SB1}, C_{GS2}, and C_{SB2}). The inductor L**1** is used to resonate with the capacitive structures intrinsic to the MOSFETs themselves and the fabrication process. The inductor L**1** is an accessory element constructed on the chip. However, instead of adding an accessory capacitor network to generate the complementary reactive impedance, the capacitors that are typically deemed as parasitic and detrimental to circuit performance are utilized in the current embodiment. More specifically, the intrinsic source-gate capacitance (C_{GS1}, C_{GS2}), and the source-body capacitance (C_{SB1}, C_{SB2}) associated with the two diode connected MOSFETS (M**1** and M**2**), are used.

[0048] From FIG. **3**, it can be seen that that accessory inductor L**1** and the intrinsic capacitors C_{GS1}, C_{SB1}, C_{GS2}, and C_{SB2} together form a series resonant circuit. The values of these intrinsic capacitors are in the Femto-Farad range. Furthermore, the capacitors C**1** and C**2** are orders of magnitude larger than the aforementioned CMOS intrinsic capacitors, and thus have negligible impedance at the frequencies of interest (hundreds of MHz to GHz). As such, impedances of the capacitors C**1** and C**1** can be neglected. Considering the relative magnitudes of the intrinsic capacitors with respect to C**1** and C**2**, the intrinsic capacitors can be considered to be at AC ground. Since this is the case, the intrinsic capacitors are in parallel with the gate-source junction of the diode connected MOSFETs. Thus, when large voltages are generated across the capacitive reactive elements, large voltages are also generated across the gate-source voltages (V_{GS}) of the MOSFETs, thereby turning them on and allowing them to operate as rectifiers.

[0049] The optimal response of the circuit is obtained at an input frequency where the accessory matching circuit (comprising of L**1** and the intrinsic capacitors) is in resonance. This condition maximizes the voltage at the input (point B in FIG. **3**) of the circuit, and matches the source impedance of an attached antenna (or any power source) to the complex conjugate of the input impedance of the circuit. Assuming the input power source (i.e., the RF source) has a real source impedance, it can be derived that as long as the following expression (1) is true then the voltage at the input is maximized when the complex impedance (Z_{load}) at the input is matched to the source impedance Rs in FIG. **3** (Rsource in expression 1).

$$\sqrt{\frac{|Z_{load}|}{R_{source}\cos(\phi)}} > 2 \tag{1}$$

[0050] In expression 1, ϕ is the phase difference between the voltage and current wave form. This is equivalent to the phase difference between the real and imaginary components of the load impedance.

[0051] While achieving resonance is optimal for performance, the power harvesting circuit also performs well even when not at resonance. This is because, even at frequencies that are as much as 20% away from resonance, large voltages are still generated across the reactive components, specifically the intrinsic capacitors. These voltages are also in parallel with the gate-source voltage (V_{GS}) of the diode connected MOSFETs, thereby turning them on and allowing for energy harvesting even at power levels in the microwatt range.

[0052] B. Performance Improvement with Parasitic Aware Circuit Layout

[0053] While a schematic diagram of an electronic circuit is an abstraction, once that circuit is constructed into a silicon chip the performance of the current changes. One reason for this is that the physical implementation of passive structures, including inductors, capacitors and interconnects, gives rise to unintentional or parasitic elements. For example, laying out of an on-chip inductor gives rise to unintentional capacitors from the inductor to the substrate ground.

[0054] In one embodiment of the disclosed power harvesting device, performance enhancement can be achieved by switching the order of the impedance matching inductor L**1** and the DC blocking capacitor C**1**, as illustrated in FIG. **4*b***. While theoretically the schematic shown in FIG. **3** is independent of the order of the L**1**-C**1** pair, it becomes apparent after examination of the parasitics incurred during layout, that the schematics in FIG. **3** (also see FIGS. **4*a***) and **4*b*** are different. The schematic produced by the layout shown in FIG. **4*b*** is less sensitive to the negative parasitics associated with the capacitor C**1**.

[0055] In particular, the impedance to voltage doubler at point A in FIG. **3** is dominated by the intrinsic capacitances of the diode connected MOSFETs (C_{GS1}, C_{SB1}, C_{GS2} and C_{SB2}). Therefore, it appears as a short at the RF frequency and the impedance at point B is also highly reactive. The parasitic capacitance of C**1** results in an imaginary impedance parallel to the input impedance at point A and B. This imaginary impedance serves to reduce the magnitude of the complex impedance and reduce the phase difference between the imaginary and real part of the complex input impedance. It can be seen from the expression (2) below that this will reduce the magnitude of the voltage at the input to the voltage doubler.

$$|V_{load}| = \frac{V_{source}}{2}\sqrt{\frac{|Z_{load}|}{R_{source}\cos(\phi)}} \tag{2}$$

[0056] By switching the order of the capacitor C**1** and the inductor L**1**, as in FIG. **4***b*, the reactive part of the impedance at points D and E is tuned out generating a large resonant voltage across the MOSFET capacitors, resulting in a real impedance on the order of 10 to 100 ohms. This is orders of magnitude lower than any parasitic impedance to the substrate due to C**1**. Therefore, the circuit in FIG. **4***b* shows superior performance to the circuit in FIGS. **3** and **4***a* in the presence of parasitic capacitances from C**1**.

[0057] This improvement can also be understood as follows. If one considers the operation of the resonant voltage generator circuit, one realizes that the largest voltages are generated at point B and the lowest voltage exist at the node connecting inductor L**1** and the source resistor Rs in FIG. **3**. Therefore, in the current embodiment, the parasitic capacitor and parasitic series resistor are moved away from the highest potential in the circuit to the lowest potential in the power harvesting circuit by switching the order of the inductor L**1** and the capacitor C**1**. While this has no effect on the theoretical operation of the circuit with respect to a schematic (done without regard to actual layout and fabrication), it greatly improves actual operation by placing the parasitic elements at a lower voltage where they will draw substantially less current, and thereby minimize their negative impact on power harvesting circuit's operation.

[0058] Another implementation of this concept is performed with the layout of inductor L**1**. The inductor L**1** is implemented on two metal layers of the CMOS process. The lower layer, which is closer to the substrate, gives rise to larger parasitic capacitance and substrate resistance. Therefore, the inductor L**1** is fabricated with the lower layer location placed at the location where the circuit voltage is lowest. This location is the side of inductor L**1** that is closer to the input and specified by point D in FIG. **4***b*.

[0059] It should also be noted that modern processes such as the IBM 8 RF process require that a certain ratio of gate tiedowns to metal area be maintained. Connecting the inductor L**1** directly to the gate of NMOS **2** (e.g., M**2** in FIG. **2**) in a Villard voltage doubler would require an unrealistic number of tiedown contacts. Therefore, this technique is best suited for designs that use a PMOS in place of NMOS **2** as discussed below.

[0060] C. Use of PMOS to Reduce Body Effect

[0061] FIG. **5** illustrates a power harvesting circuit with an NMOSFET (e.g., M**2** in FIG. **2**) replaced with a PMOSFET (PMOS in FIG. **5**). In a steady state, an NMOS **2** (e.g., M in FIG. **2**) will have a source potential that is significantly higher than the body potential. This gives rise to an increase in the threshold voltage of NMOS **2** through the body effect. This causes the diode connected NMOS **2** to turn on at a higher voltage, thereby reducing efficiency. To minimize the body effect in the current embodiment, NMOS **2** is replaced with a diode connected PMOS, as shown in FIGS. **3** and **5**. The PMOS has the body, gate, and drain node connected to Vout and the source node connected to the voltage at the source terminal of the diode connected NMOS **1**. The body effect of NMOS **1** (e.g., M**1** in FIG. **2**) is already minimized and is treated similar to M**1** in FIG. **2**. The PMOS only conducts when the gate voltage is below the source voltage. Since the gate has been connected to Vout, this occurs when the AC voltage at the input is positive with respect to ground. Since both the body and the source of the PMOS are connected to Vout, no threshold voltage increase occurs due to the body effect. This improvement can be combined with the other embodiments discussed above, including the embodiment shown in FIG. **3** where the order of the capacitor C**1** and the inductor L**1** are transposed to reduce parasitic capacitive and resistive losses.

[0062] In order for the design to work, the PMOS needs to be sufficiently isolated to allow connecting the body of the PMOS to Vout without any current flowing from the PMOS body to the bulk substrate. Connecting the body of a diode connected PMOS to Vout is possible due to the fact that PMOS FETs are placed in an n doped well inside of the p substrate.

[0063] As seen in FIG. **6**, the PMOS is built in an N well. Since the junction from the N well to the P substrate acts as a diode, if the voltage potential in the N well is higher than the voltage potential in the P substrate, current will not flow. This is what allows the body to be connected to Vout which is higher than the potential of the substrate. The junction capacitance between the N well and the substrate is parallel to and several orders of magnitude lower than the output capacitance C**2** of the power harvesting circuit. Therefore, the junction capacitance has no significant effect on the circuit performance.

[0064] D. Sacrificial Biasing

[0065] The output voltage of the related art voltage doubler is two times the input voltage minus the threshold voltage of both diodes. Therefore, minimizing the threshold voltage of the diodes maximizes the voltage at the output of the voltage doubler. Previous works in the related art have attempted to improve RF power harvesting efficiency by using special MOSFETS with reduced gate threshold voltages.

[0066] In an exemplary embodiment of the present invention shown in FIGS. **7** and **8**, an alternative technique uses biasing at the gate combined with the use of a PMOS in place of the output NMOSFET to eliminate the threshold voltage for both diode connected MOSFETs.

[0067] For example, in the embodiment shown in FIG. **7**, voltage sources are directly connected to the gates of the diode connected MOSFETs in the RF power harvesting circuit.. An alternative approach is to use the desired output voltage to create the bias voltages through a voltage divider network, as shown in FIG. **8**. This technique sacrifices some current to create an overall improvement to system efficiency. In this embodiment, the NMOS **2** is replaced by a PMOS IN FIG. **8** for two reasons. First, the threshold voltage of the output diode connected MOSFET must be below Vout to be obtainable through a voltage divider network. Second, the gate is connected to a DC node in the circuit. Sufficiently larger resistors R**1**, R**2**, R**3**, and R**4** are used in the divider network IN FIG. **8** to reduce power dissipated due to the bias current. The resistors are chosen to be 1 MΩ to 10 MΩ for 10 μW input power. Generally, the smaller the input power, the higher the resistances. This power dissipation can be designed to be orders or magnitude below the increase in output power due to the bias voltages. Practical limitations are placed on the size of the resistors used in the divider network due to a physical limitation of the resistor size in the circuit layout.

[0068] In the current embodiment, traditional MOSFETs outperform low threshold voltage MOSFETs when sacrificial current biasing is utilized. This is in contrast to related art literature that has concluded that the use of low threshold voltage MOSFETs provides the best performance for RF power harvesting. So, instead of using low threshold voltage MOSFETs, we have introduced sacrificial biasing while using regular MOSFETs and have achieved improved results. The improvement is because we achieve the effect of easily turning on the MOSFETs, while retaining the improved performance of regular MOSFETs with respect to the turn on characteristics as shown in FIG. **9**. Regular MOSFETs, have a sharper turn-on transition than low threshold voltage MOSFETs resulting in a faster turn on when transitioning to the open state, showing better RF conversion efficiency.

[0069] FIG. **10** illustrates a comprehensive design of a power harvesting device which includes the transposition of the inductor L**1** and the capacitor C**1** of FIG. **4***b*, the implementation of the PMOS device in place of M**2** as shown in FIG. **5**, inclusion of sacrificial bias resistors R**1**-R**4** as shown in FIG. **8**, as well as implicit utilization of MOSFET intrinsic capacitors for achieving maximum turn on voltages with low power input as shown in FIG. **3**.

[0070] FIG. **11** is the layout of the harvesters to form a silicon integrated circuit. The layout shows five RF power harvesting circuits, which are the physical implementations of the harvesting circuits shown in the circuit schematics shown in other figures. The circuits have been used to compare and test the various designs of the energy harvester. The testing provided the data that has lead to the conclusion and claims in this patent application.

[0071] FIG. **12** illustrates a comparison of measured RF to DC conversion efficiency versus output voltage between a power harvesting device with sacrificial biasing, and a power harvesting device without biasing. The harvester with sacrificial biasing achieves high efficiency compared with unbiased design for 1V output voltage. The higher curve represents a higher conversion efficiency for RF energy into DC energy. That is, the results show that for the given output voltages, the sacrificial biasing provides improved efficiency over the unbiased design.

[0072] FIGS. **13**-**21** relate to self-powered systems. An RF power harvesting device discussed with respect to the above embodiments is incorporated. Available RF power is sometimes very weak, so it is not easy to generate a sufficient output for direct use. Using a combination block, converter, and battery provides an attractive solution.

[0073] FIG. **13** shows a block diagram of self-powered system including an RF power harvesting block **1**, a voltage converter **2**, a battery **3**, a combined block **4**, and functional electronics **5**. Here, the RF power harvesting block **1** could be one of the RF power harvesting devices discussed in the above embodiments.

[0074] In one embodiment, the converter **2** is a switched-capacitor DC-DC voltage converter. The converter is designed to generate maximum six times of input voltage at output using external capacitors. Through capacitor charging tests, it shows up to 40.5% energy transfer efficiency using 0.35V input for a 1000 uF capacitor charging to 1.4V during 10 minute with a three stage converter. The battery **3** can be a rigid type battery or a flexible type battery. Both types of batteries are successfully recharged using the converter **2**. In one embodiment, two hour charging with four parallel converters recharges a rigid battery fully with 0.4V input, so it returns the battery potential to the initial condition before discharging of a 9.85 KΩ load resistor during 10 minutes.

[0075] The rigid type electrochemical battery comprises hydrated Ruthenium Oxide ($RuO_2.xH_2O$) and activated Carbon (CA). These cells have high current capacity, rechargeability, and even flexibility. One of the targets of the battery **3** is playing a role of the power source for distributed sensor networks, or Smart-dust nodes, for achieving a stand-alone system. Much research in the related art has been focused on implementing self-powered systems such as a node of low-power ad hoc distributed networks using an RF power harvesting device. However, when the input power at RF harvesting block **1** is very small, the output is not enough to drive a system or node (e.g., the functional electronics **5**). This problem can be solved by using a voltage converter **2** and rechargeable battery **3** combination block **4**, as shown in FIG. **13**. For example, a DC voltage converter **2** stacks the small voltage of RF energy harvesting block **1** up to the necessary charging voltage for battery cell and charges a battery **3**. Then, the battery **3** provides enough bias or power for the system (e.g., functional electronics **5**) only when it needs to be working. In the current embodiment, a switched-capacitor DC voltage converter is used as the converter **2** for battery charging. This converter **2** and battery **3** combined block **4** bridges a gap between RF power harvesting block **1** and functional electronics **5**, e.g., in insufficient RF power environments. So, it can help to implement a self-powered or stand alone system more effectively.

[0076] Usually, the developed battery cells have 1.1V~1.2V built in potential after fabrication without any other charging or discharging steps, and slightly higher charging voltage than the built in potential was fine for battery recharging. Actually, charging voltage and time for constant voltage charging scheme show a trade-off relation. A 1.4 charging voltage of DC power supply is used during two hours for full recharging. FIG. **14** shows two types of batteries, a rigid battery (FIG. **14**(*a*)), and a flexible battery (FIG. **14**(*b*)).

[0077] Switched-Capacitor DC-DC Converter

[0078] There are two reasons why we choose a switched-capacitor scheme for DC-DC converter among several different types. The first is its simplicity of implementation, and the second is it needs capacitors for storing energy. The second factor is advantageous because the battery cells can be used as a capacitor also by some modifications. So the switched-capacitor method is a very attractive one.

[0079] FIG. **15**(*a*) shows the schematic of a switched capacitor DC-DC converter with three stages. A single chip is implemented using IBM8RF 0.13 um CMOS process including every switch for a five stage converter, except external capacitors. That is, maximum available output voltage is six times of input voltage through five external capacitors, and the output can be selected by a user from minimum (input×1) to maximum (input×6) just connecting an external capacitor's plus pad to input.

[0080] In the schematic, the top switches are made by nMOS only, but the middle and bottom switches are made by transmission gate, combining nMOS & pMOS to transfer the accumulated potential on each capacitor to the next stage without any loss. The converter works on two different phases. For capacitor storing phase ($Q=1$, $\overline{Q}=0$ FIG. **15**(*b*)), the input node is connected to every external capacitor, so all capacitors are in parallel. However, the output is disconnected from the converter. In this step, external capacitors are charged up through input. Then, during battery charging phase (Q=0, $\overline{Q}$=1,

[0081] FIG. **15**(*c*)), the parallel configuration of capacitors is changed into series, and the output is connected to the last capacitor's plus node. Note the input can be used in both phases without any waiting or sleeping mode. So, it can decrease the charging time by fully using the input source. FIG. **16** shows the output (~1.4V) of the converter using a three stage converter for 0.35V input at 20 Hz clock signal using 100 uF three external capacitors. Note the body contacts of pMOS in converter switches are connected into external 1.5V DC power supply. This external source can be replaced with a battery for fully self-powered system implementation.

[0082] Big Capacitor Charging

[0083] At the initial tests of the converter, a big capacitor, 1000 uF, replaces a battery for charging test to verify functionality of the converter, just using three stages. Also, frequency dependency is checked for different clock signals of 50% duty cycle. The external capacitors, storing energy from input voltage, and input voltage are fixed as 100 uF and 0.35V, respectively. FIG. **17** shows the results of capacitor charging for different clock signals, and the inner graph is a magnified version of 400~700 sec section. All tests are stopped when the capacitor voltages reach 1.38V, except 20 KHz test. The plots for 2 Hz, 20 Hz, and 200 Hz test are almost overlapped. As listed in Table 1, 2 Hz and 20 Hz test have almost same results for energy transfer efficiency, around 40%, and charging time about 10 minutes, where the supplied energy is calculated by integrating the measured current from power supply and the stored energy is just calculated value using ½CV^2. As the clock frequency goes higher, the saturation voltage goes lower. It is caused by frequency dependency of dynamic power consumption of CMOS switches [6]. By frequency dependency tests, a 0.2 Hz~20 Hz clock frequency range is used for battery charging tests.

TABLE 1

Frequency dependency of 1000 uF capacitor charging properties

Clock Frequency [Hz]	Charging Time [sec]	Supplied Energy [mJ]	Stored Energy [mJ]	Energy Transfer Efficiency [%]
0.2	953.5	2.352	0.952	40.48
2	611.5	2.349	0.952	40.53
20	617.5	2.394	0.952	39.77
200	629.5	2.555	0.953	37.30
2K	755.5	2.755	0.953	34.59
20K*	755.5	3.010	0.852	28.31

*For 20 KHz test, 1.3 V saturated voltage is used for calculations.

[0084] Battery Charging

[0085] The test procedure follows three steps for battery charging: 1) initial discharging, 2) cell charging, and 3) discharging.

[0086] The first charging test set up is 0.3V input from power supply, five 100 uF external storing capacitors for a five stage converter, which provides 1.8V output without load, and 50% duty cycle 0.2 Hz clock signal. Actually, a test with 0.25V input, making 1.5V output by five stages, was performed, but the charging time took over several days. So, to reduce charging time, input voltage is increased from 0.25V to 0.3V. FIG. **18** shows the test results. After the initial discharging (dashed line) during 1280 seconds, the load voltage was dropped from 0.485V to 0.223V for 1 KΩ load resistor. The black solid line shows the discharging after 60 hours charging. 60 hours charging gives 0.316V initial load voltage, when 1 KΩ load is connected to the battery after charging. Though the charging time, 60 hours, is a little long, we demonstrate a DC-DC converter is working properly for the battery charging.

[0087] To reduce charging time, the input voltage is increased from 0.3V to 0.34V and two converters are used in parallel for another test. Clock signals are applied for two converters to operate in opposite mode at the same time. That is, if a converter is in capacitor storing mode, the other is in battery charging mode. The test setup is 0.34V input, generating 2V output without load due to using five stage, and 25 Hz clock signal with 50% duty cycle. The battery cell is discharged with 1 KΩ load during 20 minute before charging, as shown by the dashed line in FIG. **19**, and load voltage drop is measured from 0.358V to 0.174V. Then, two converters charge the cell for 4 hours. This leads to 0.29V initial load voltage at discharging after charging (solid line in FIG. **19**). Using higher input voltage and two converters, the charging time is effectively reduced from 60 hour to 4 hour.

[0088] Another charging test is performed using four converters for further decreasing charging time. To run four converters together, switching signals are generated using a ring counter, which is made by four D Flip-Flops. FIG. **20** shows the generated timing diagram of the counter, Q_0~Q_3. Simply, inverters inverse each timing signal for the reverse switching signal inputs, $\overline{Q_0}$~$\overline{Q_3}$. Using the switching signals, if Q_0 is high, only one output of converters is connected to battery cell, and the others are in capacitor storing mode.

[0089] During capacitor storing mode, external capacitors are in parallel, so it gives a high total capacitance. Accordingly, it makes a long RC time constant for capacitor storing. Previous tests used five stages with 100 uF external capacitor, which gave 500 uF total capacitance for capacitor storing mode. For quick charging up capacitors, only three stages are used for other tests, which need three external capacitors for each converter. However, reducing the number of stage affects the output voltage directly. To compensate for this, higher input voltage is used. Also, higher input voltage can reduce the charging time further by providing more charges on storing capacitor, if the current capacity of the input signal is not limited. In the following tests, 0.35V, 0.4V, and 0.45V input are used for comparison.

[0090] Four charging tests are performed using a rigid (Test 1~3) and flexible (Test 4) type battery. For a rigid battery, charging time is limited to 2 hour, but the flexible cell is charged for 12 hours due to the cell's own property. Usually, the flexible battery consumed more current at the same charging voltage compared with rigid one. Furthermore, 9.85 KΩ and 100.2Ω load resistor are used at discharging for rigid and flexible battery, respectively. The test results and setup are shown in FIG. **21** and Table 2.

[0091] Every charging test shows energy recovery of battery. Although 2 hour charging with 0.35V input is not sufficient for full recovery before discharging, other two tests, Test 2 & 3, show the fully recovered battery potential compared with discharging before charging tests. However, Test 3 does not give any improvement in spite of higher input voltage, 0.45V. This may be due to the aging of the cell through several charging and discharging tests. Test 4 shows 12 hour charging result for flexible cell.

TABLE 2

Charging test setup and results

			Charging (three stage converter)			Discharging during 10 minutes				
							Before charging [V]		After charging [V]	
Test #	Battery Cell Type	Battery Cell Size	Input [V]	Clock [Hz]	Charging time [Hour]	Load resistor [Ω]	Initial load voltage	Final load voltage	Initial load voltage	Final load voltage
1	Rigid	4 cm^2	0.35	20	2	9.85K	0.804	0.659	0.759	0.612
2	Rigid	4 cm^2	0.4	20	2	9.85K	0.790*	0.648	0.804	0.659
3	Rigid	4 cm^2	0.45	20	2	9.85K	0.818	0.688	0.828	0.701
4	Flexible	4 cm^2	0.45	2	12	100.2	0.775	0.497	0.689	0.445

*Initial load voltage for Test 3 before charging is estimated value because of a slacken load connection.

[0092] As described above with respect to FIGS. **13-21**, a switched-capacitor DC-DC converter is implemented using a CMOS process for a battery cell charging application. Up to 40% energy transfer efficiency is obtained through capacitor charging tests. For battery cell charging tests, the functionality of the converter is verified for both rigid and flexible type battery cells using much smaller input voltage compared with general high charging voltage, more than 1.2V for the batteries discussed with respect to FIGS. **13-21**.

[0093] The techniques described above show a dramatic increase in power harvesting efficiency as compared to existing RF Power Harvesting designs and, enable recharging portable wireless electric using ambient RF energy sources. Utilizing these improvements, a design goal of generating 1V output voltage with a greater than 20% RF to DC conversion efficiency from RF energy levels measured in the environment (66 uW) was met. This represents better than double the RF to DC conversion efficiency of the related art power matched RF energy harvesting circuit based on a Villard voltage doubler.

[0094] Several design improvements, novel to RF power harvesting circuits, have been disclosed. These improvement include RF to DC conversion efficiency through a reduction in the body effect of diode connected MOSFETs, reduction in the threshold voltage and by reducing the affects of circuit parasitics. These circuit improvements have been simulated to show a better than 60% improvement to RF to DC conversion efficiency.

[0095] While the present invention has been particularly shown and described with reference to exemplary embodiments thereof, it will be understood by one of ordinary skill in the art that various changes in form and details may be made therein without departing from the spirit and scope of the present invention as defined by the following claims.

1. An RF power harvesting device comprising:
an inductor;
a first capacitor connected to the inductor;
a first MOSFET connected to a first node, and
a second MOSFET connected to the first node,
wherein the inductor or the first capacitor are connected to the first node.

2. The RF power harvesting device of claim **1**, wherein values of the first MOSFET and the second MOSFET are such that intrinsic capacitances of the first MOSFET and the second MOSFET and the inductor constitute a substantially resonant circuit during operation of the RF power harvesting device.

3. The RF power harvesting device of claim **1**, further comprising:
a second capacitor connected to a second node, and the second MOSFET is connected to the second node,
wherein the second MOSFET is a PMOSFET.

4. The RF power harvesting device of claim **1**, further comprising:
a receiver which detects and receives ambient RF energy,
wherein the first capacitor and the receiver are connected at a third node, the first capacitor and the inductor are connected in series at a fourth node, and the inductor is connected to the first node.

5. The RF power harvesting device of claim **1**, further comprising:
a first set of resisting elements connected to a first gate of the first MOSFET to apply a first DC bias to the first gate; and
a second set of resisting elements connected to a second gate of the second MOSFET to apply a second DC bias to the second gate.

6. A receiver comprising:
an impedance matching circuit connected to a first node;
a first MOSFET connected to the first node; and
a second MOSFET connected to the first node, wherein
the first MOSFET is a diode-connected p type MOSFET.

7. The receiver of claim **6**, wherein the impedance matching circuit comprises an inductor and a capacitor in series, the inductor connected to the first node.

8. The receiver of claim **6**, wherein a resistive biasing network provides a bias voltage to gate terminals of the first MOSFET and the second MOSFET.

9. A system comprising:
a receiver;
a voltage converter; and
an energy storage device, wherein
the receiver is configured to receive a high frequency signal and supply a first voltage to the voltage converter, and
the voltage converter is configured to amplify the first voltage supplied by the receiver and supply the amplified voltage to the energy storage device.

10. The system of claim **9**, wherein the voltage converter includes a switched capacitor circuit.

* * * * *

US 20110242863A1

(19) **United States**

(12) **Patent Application Publication**

PARK et al.

(10) **Pub. No.: US 2011/0242863 A1**

(43) **Pub. Date: Oct. 6, 2011**

(54) **PATCH ANTENNA AND RECTENNA USING THE SAME**

(75) Inventors: **Jun-seok PARK**, Seoul (KR); **Ha-Ryoung OH**, Goyang-si (KR); **Yeong-Rak SEONG**, Seoul (KR); **Hyoung-Hwan ROH**, Seoul (KR); **Ki-won SEO**, Seoul (KR)

(73) Assignee: **Kookmin University Industry Academy Cooperation Foundation**, Seoul (KR)

(21) Appl. No.: **12/758,987**

(22) Filed: **Apr. 13, 2010**

(30) **Foreign Application Priority Data**

Mar. 31, 2010 (KR) 10-2010-0029261
Mar. 31, 2010 (KR) 10-2010-0029262
Mar. 31, 2010 (KR) 10-2010-0029263

Publication Classification

(51) **Int. Cl.**
H02M 7/06 (2006.01)
H01Q 1/50 (2006.01)
H01Q 9/04 (2006.01)

(52) **U.S. Cl.** .... **363/126**; 343/860; 343/700 MS

(57) **ABSTRACT**

A patch antenna for receiving high frequency wireless signal and a rectenna using the same, more particularly, an impedance-matched patch antenna adopting a slot capacitive coupling structure and a rectenna capable of generating electrical energy from the wireless signals having different frequency band. A rectenna for receiving an A.C. wireless signal carrying electrical energy and converting the wireless signal into a D.C. electrical energy, is comprised of: a patch antenna for receiving the wireless signal comprising an dielectric substrate, a patch that is formed at the upper area of the surface of the dielectric substrate and providing the first frequency response characteristics, a ground plane formed on the other surface of the dielectric substrate, and an impedance matching means providing the second frequency response characteristics; and a rectifying unit that converts the wireless signal, received via the patch antenna, into a D.C. electrical energy by rectifying the wireless signal.

52

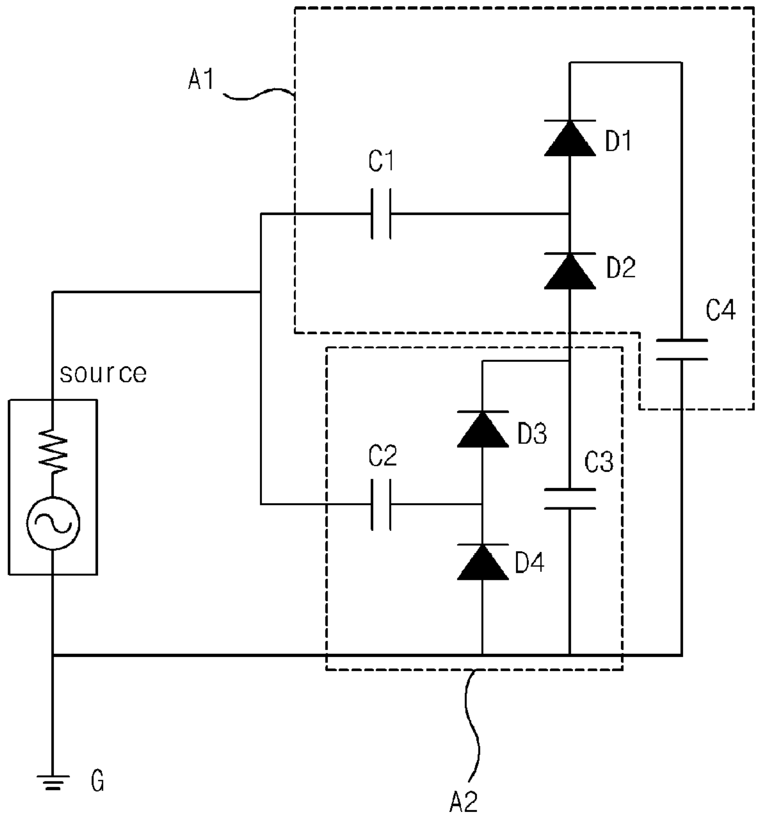

FIG. 1

Prior Art

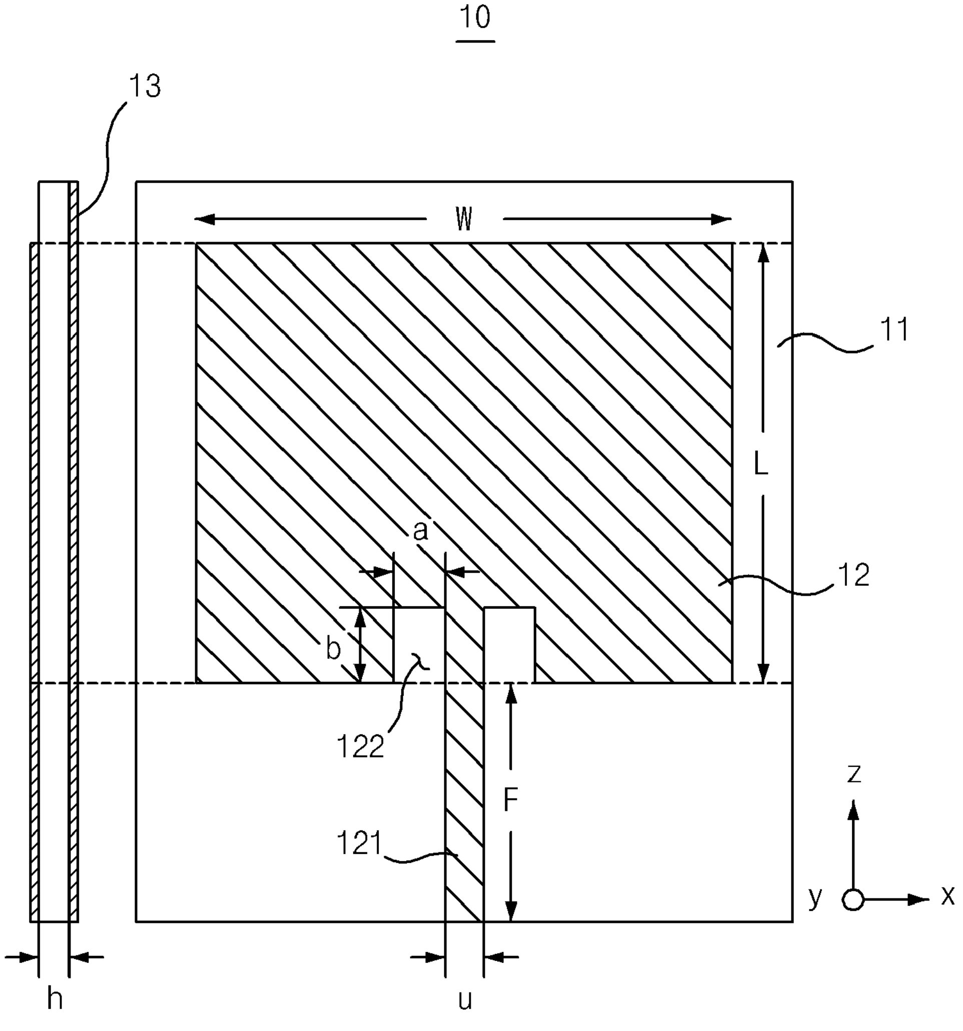
10
13
W
11
L
a
12
b
122
F
121
z
y
x
h
u

FIG. 2

Prior Art

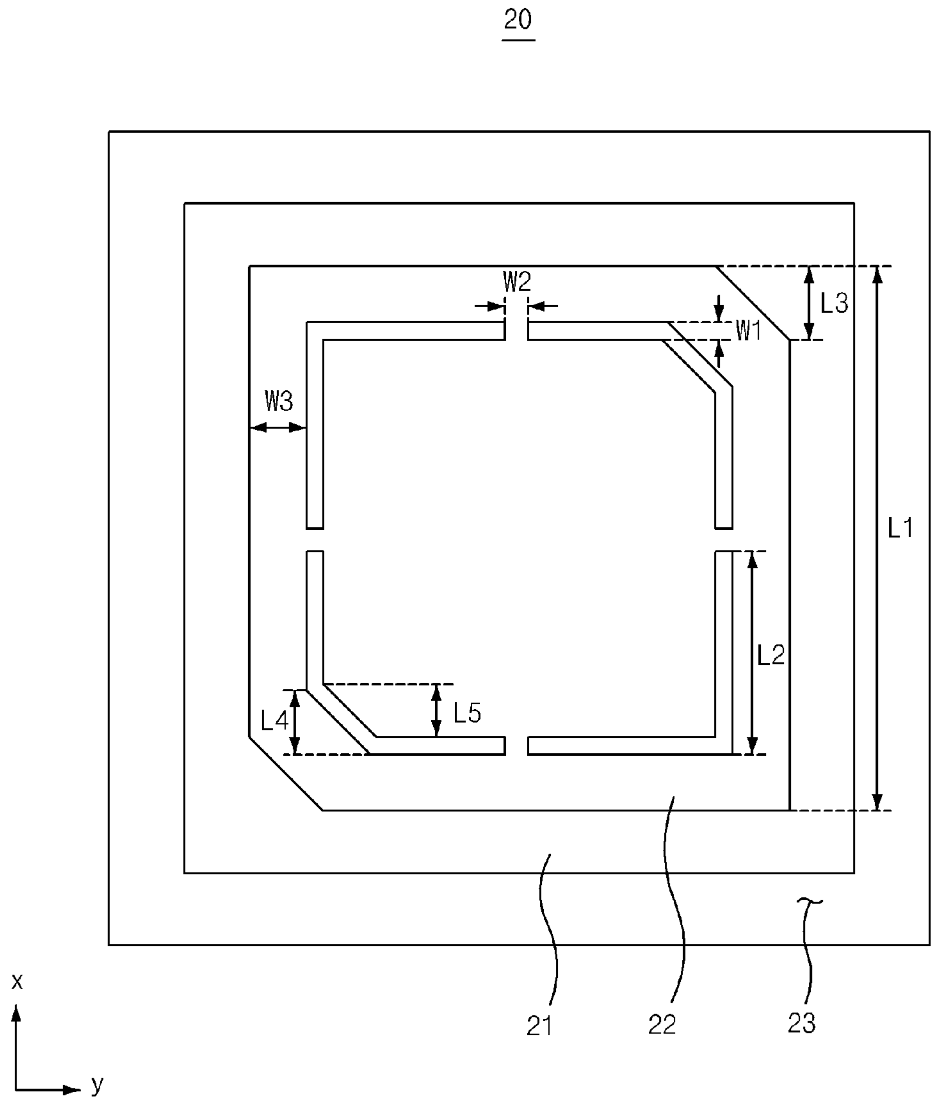
20
W2
L3
W1
W3
L1
L2
L5
L4
21
22
23
x
y

FIG. 3

Prior Art

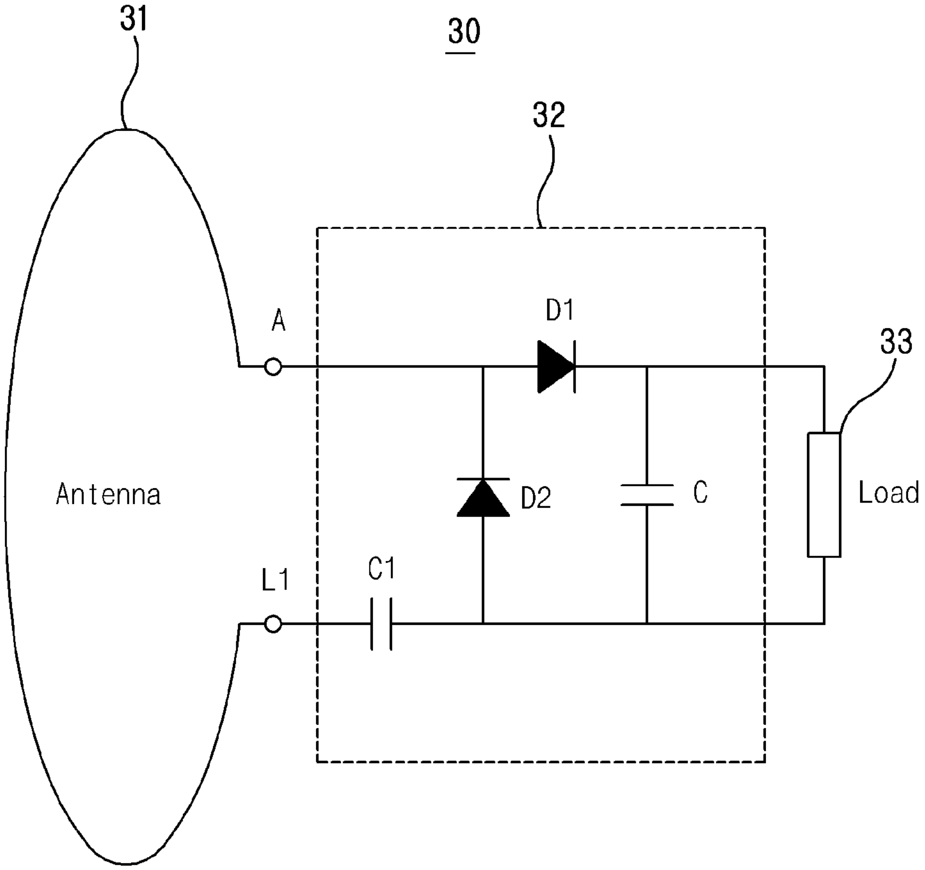
31
30
32
A
D1
33
Antenna
D2
C
Load
L1
C1

FIG. 4

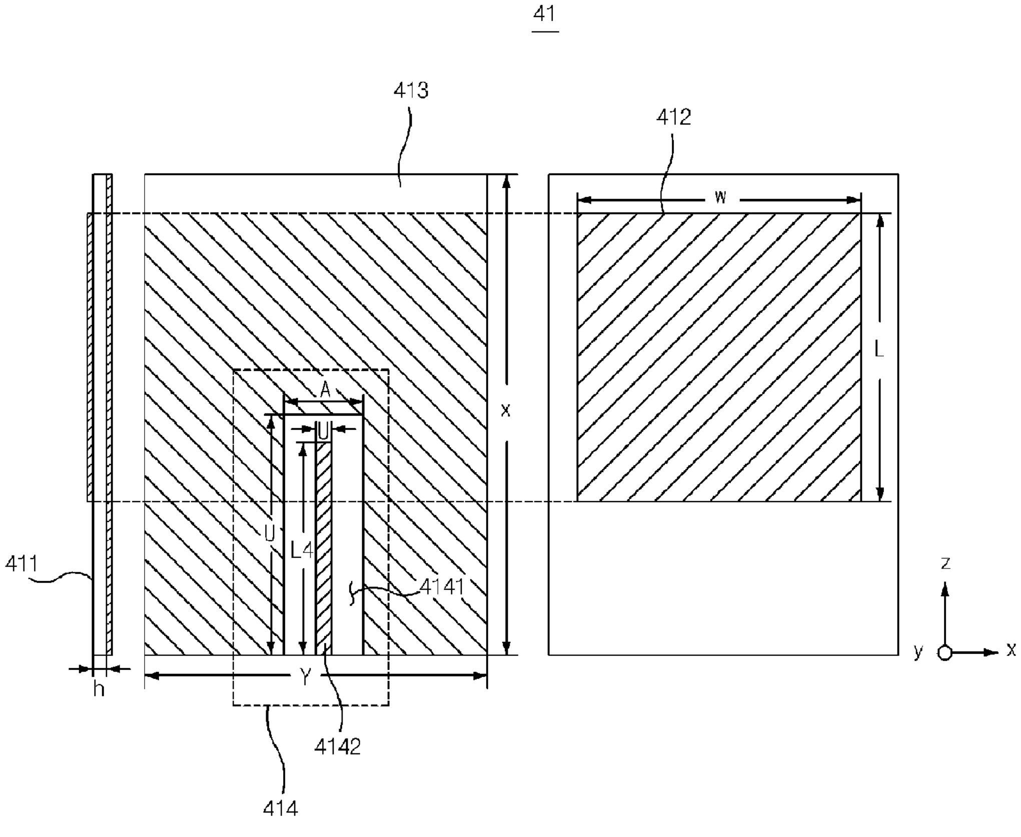
41
413
412
w
L
A
U
U
L4
x
4141
411
h
Y
4142
414
z
y
x

FIG. 5a

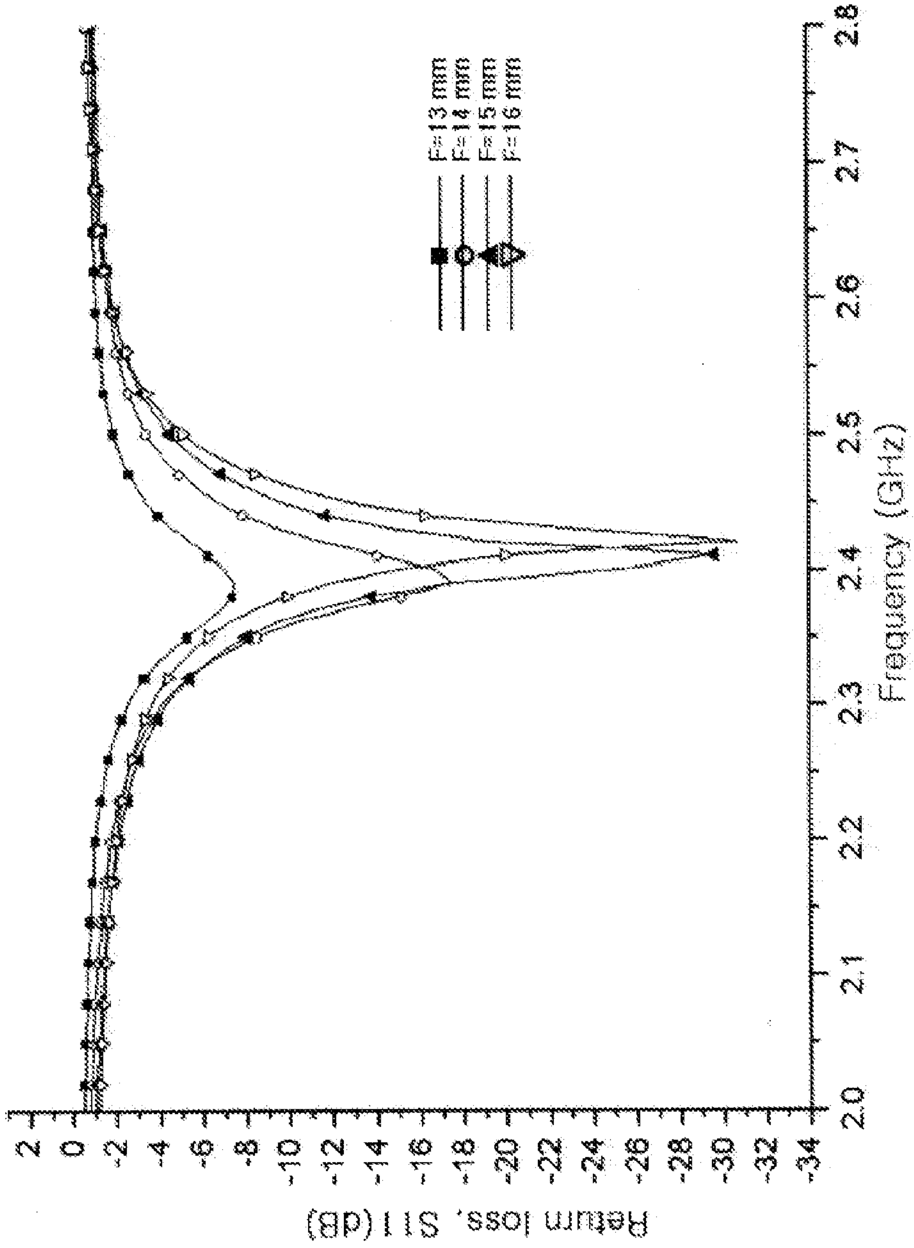
Return loss, S11 (dB)
2
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
-20
-22
-24
-26
-28
-30
-32
-34
Frequency (GHz)
2.0
2.1
2.2
2.3
2.4
2.5
2.6
2.7
2.8
F=13 mm
F=14 mm
F=15 mm
F=16 mm

FIG. 5b

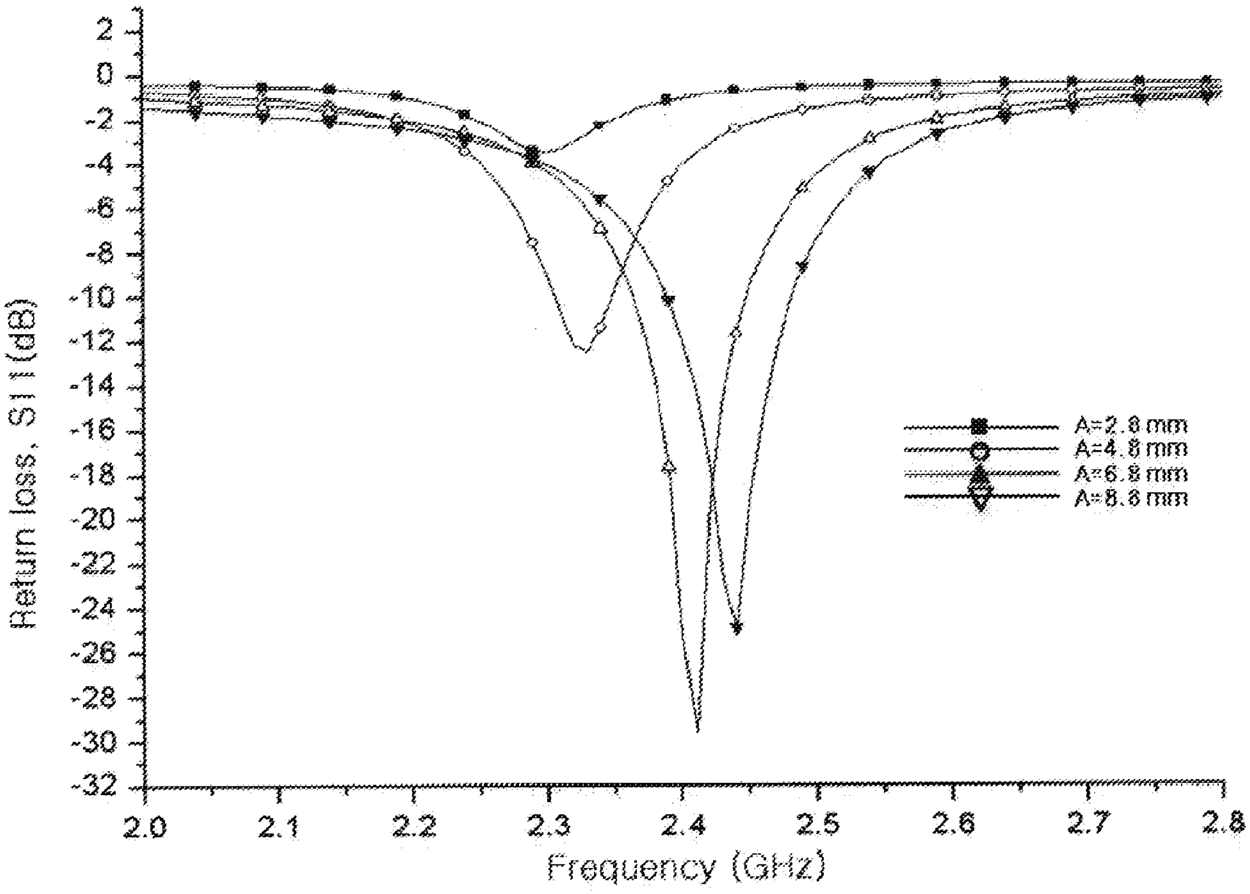
2
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
-20
-22
-24
-26
-28
-30
-32
Return loss, S11(dB)
2.0
2.1
2.2
2.3
2.4
2.5
2.6
2.7
2.8
Frequency (GHz)
A=2.8 mm
A=4.8 mm
A=6.8 mm
A=8.8 mm

FIG. 6

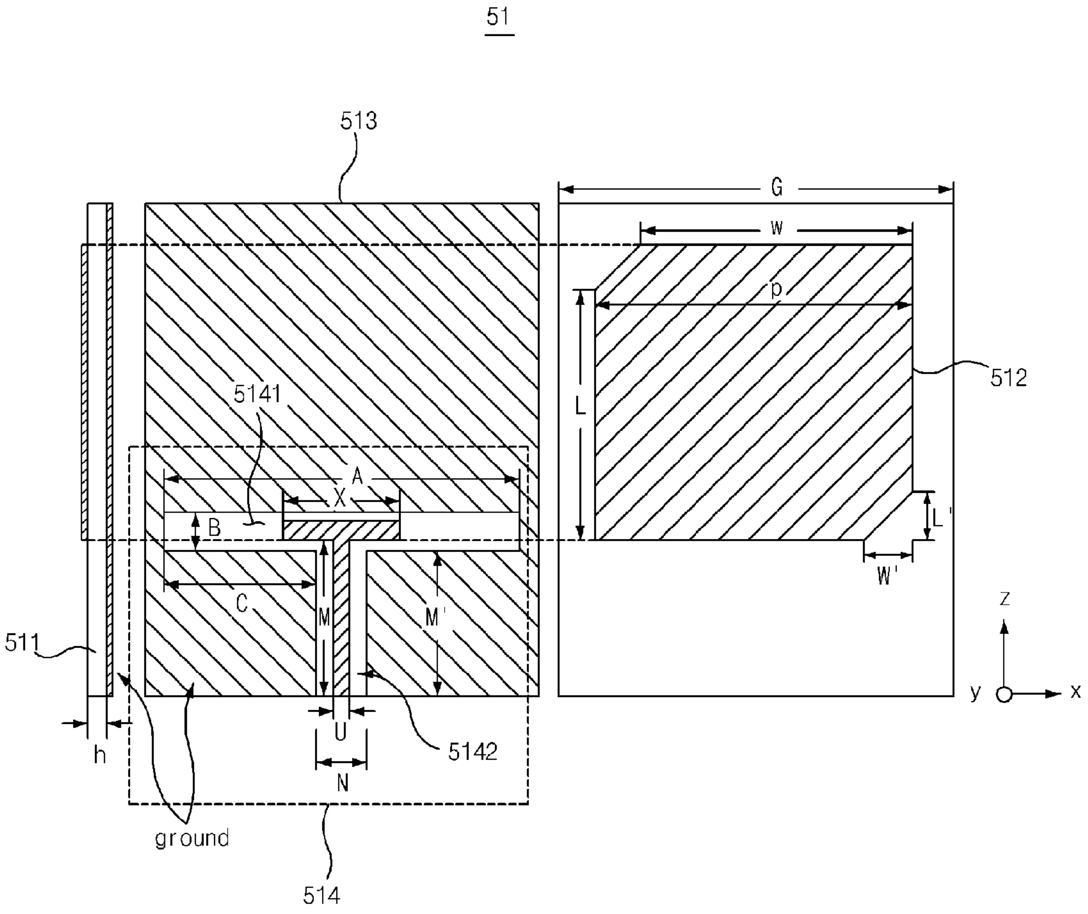
51
513
G
w
p
512
L
L'
W'
5141
A
X
B
C
M
M'
511
h
U
N
5142
ground
514
z
y
x

FIG. 7

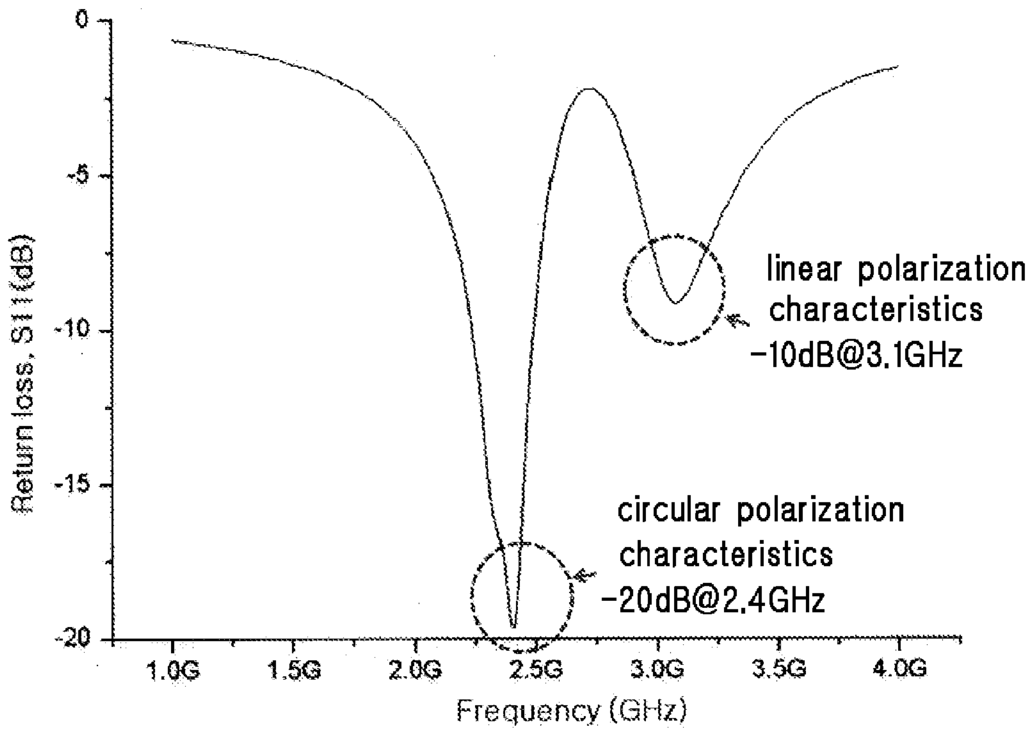
0
-5
-10
-15
-20
Return loss, S11(dB)
1.0G
1.5G
2.0G
2.5G
3.0G
3.5G
4.0G
Frequency (GHz)
linear polarization characteristics -10dB@3.1GHz
circular polarization characteristics -20dB@2.4GHz

FIG. 8

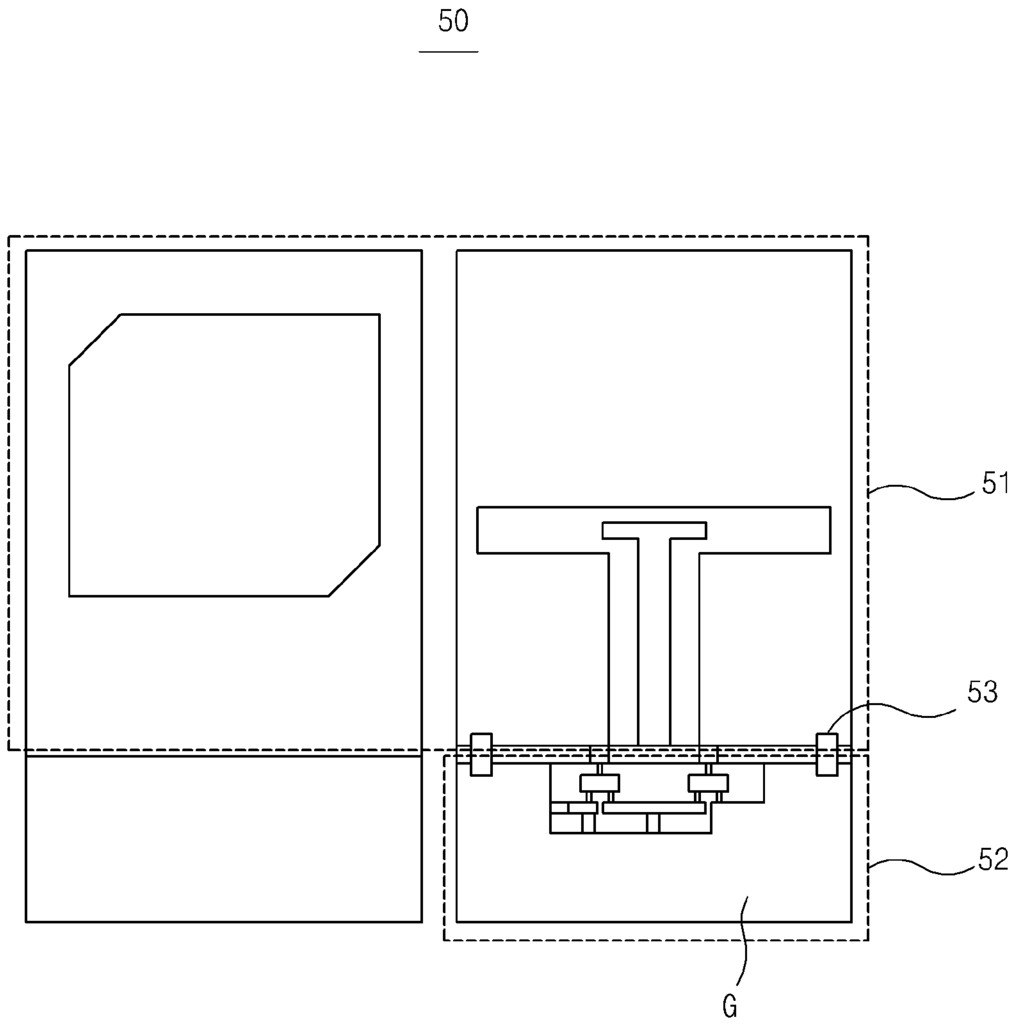

FIG. 9

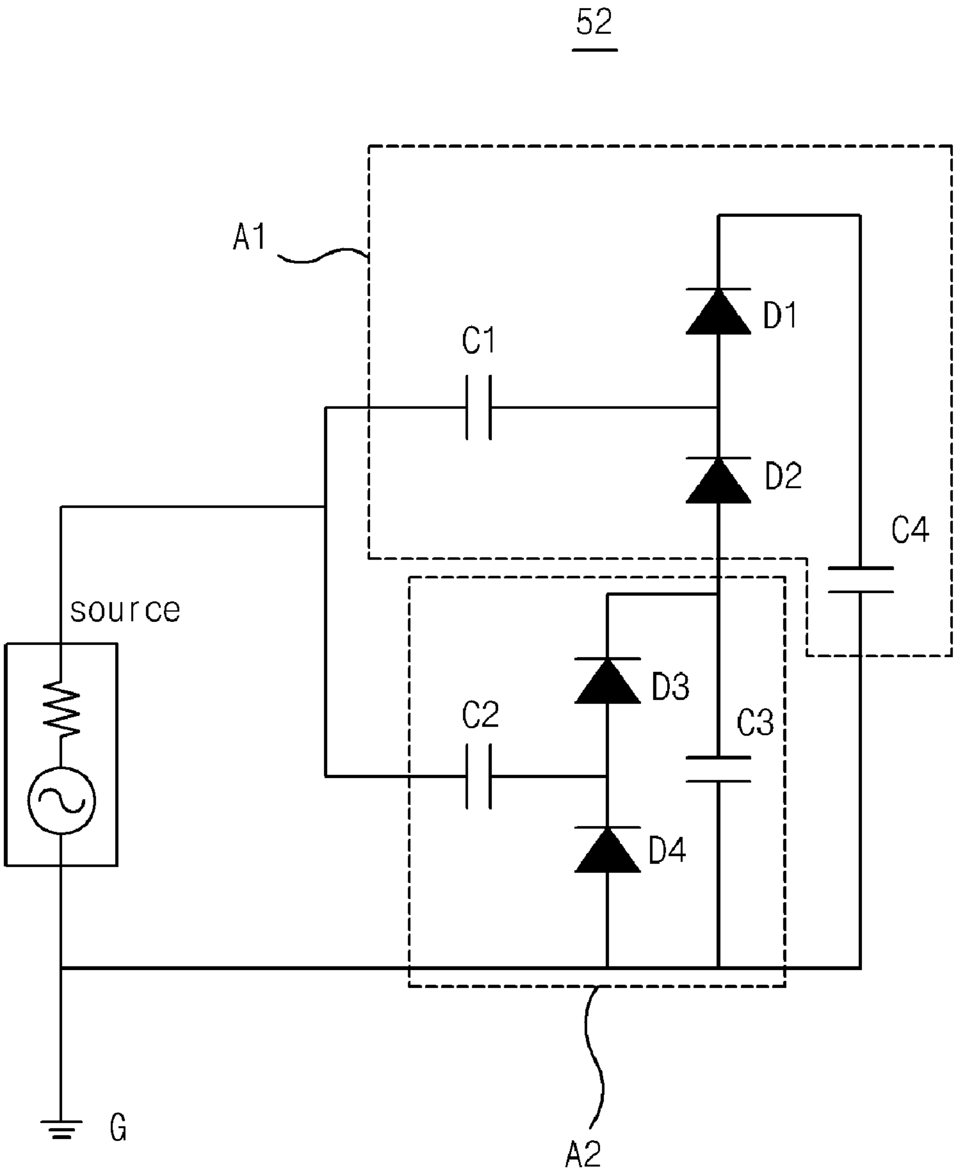
52
A1
D1
C1
D2
C4
source
D3
C2
C3
D4
G
A2

FIG. 10

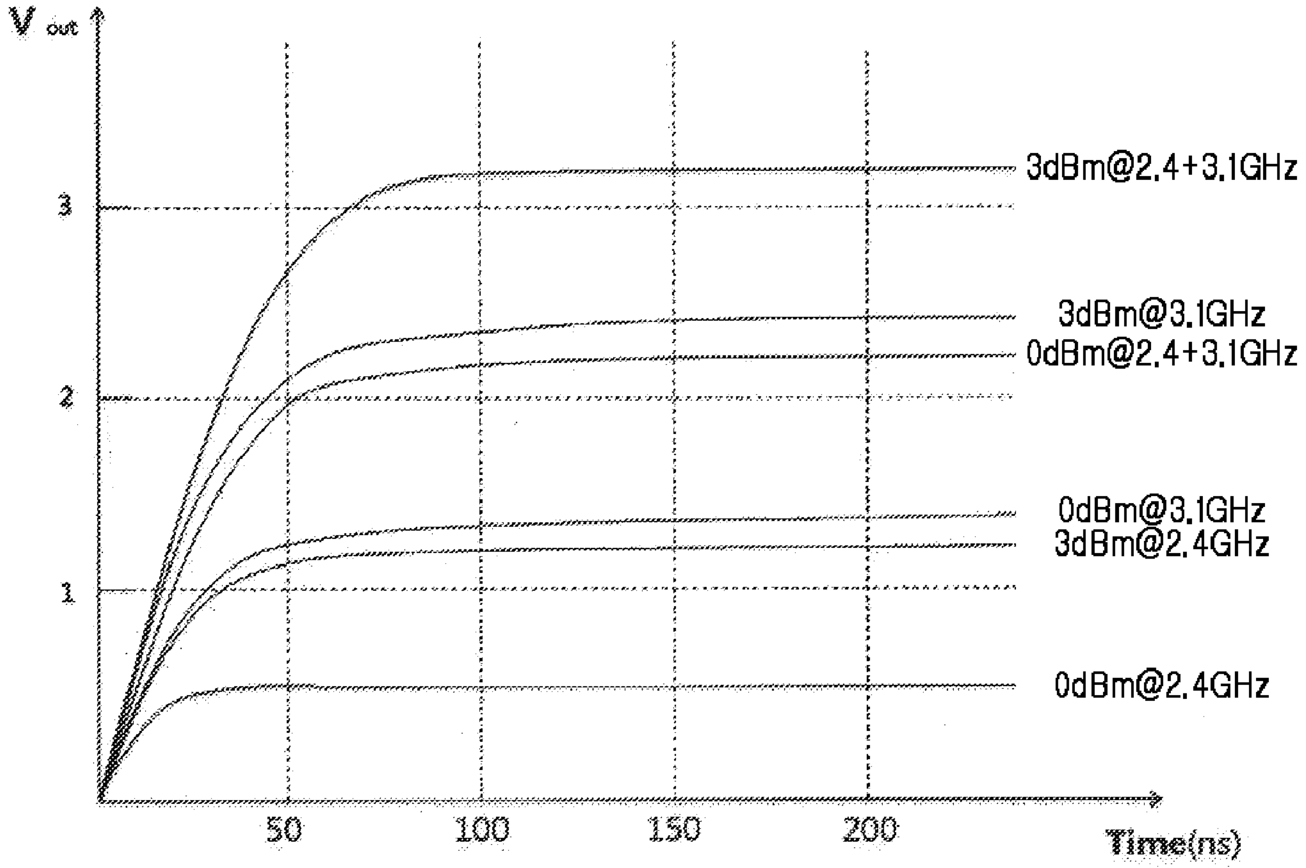
V out
3
2
1
50
100
150
200
Time(ns)
3dBm@2.4+3.1GHz
3dBm@3.1GHz
0dBm@2.4+3.1GHz
0dBm@3.1GHz
3dBm@2.4GHz
0dBm@2.4GHz

PATCH ANTENNA AND RECTENNA USING THE SAME

BACKGROUND OF THE INVENTION

[0001] 1. Field of the Invention

[0002] This invention relates to a patch antenna for receiving high frequency wireless signal and a rectenna using the same, more particularly, to an impedance-matched patch antenna adopting a slot capacitive coupling structure and a rectenna capable of generating electrical energy from the wireless signals having different frequency band.

[0003] 2. Description of the Related Art

[0004] A patch antenna according to the first example of prior art is illustrated in FIG. **1**. A patch antenna **10** illustrated in FIG. **1** adopts an in-set structure, wherein impedance matching is achieved by adjusting design parameters (a, b) of the gap **122** formed at each side of the feeder line **121** that is a partial extension from the bottom side of the patch **12** which is formed on a surface of the dielectric substrate **11**.

[0005] According to the first example of prior art, forming the minute gaps at both sides of the feeder line **121** is a very difficult process; furthermore, the impedance matching condition gets easily out of control due to a small variation in process parameters during actual fabrication of a patch antenna.

[0006] In addition, since the gaps **122**, located at the both sides of the feeder line **121** which is an impedance matching means, are formed inside the patch **12** as an integrated part thereof, the expensive patch **12** has to be newly fabricated in case when a problem occurs during the gap forming process.

[0007] A patch antenna according to the second example of prior art is illustrated in FIG. **2**. The patch antenna **20**, having a dual frequency band characteristics, of the second example of prior art is comprised of an dielectric substrate **21**, an external patch **22**, and an internal patch **23** as illustrated in FIG. **2**.

[0008] The patch antenna **20** of the second example of prior art is capable of transmitting and receiving both signals having two different frequency bands by adopting an external patch **22** and an internal patch **23** having different frequency characteristics.

[0009] But, since this type of patch antenna utilizes two patches in transmitting and receiving signals having two different frequency bands, it requires high production cost and large antenna size to integrate two types of patch into a stacked form.

[0010] A rectenna according to an example of the prior art is illustrated in FIG. **3**. The rectenna **30**, shown in FIG. **3**, relates to a rectenna that utilizes a loop antenna; it is largely comprised of a loop antenna **31** and a rectifying circuit **32**; it converts an A.C. wireless signal received via the loop antenna **31** into a D.C. electrical energy by rectifying the A.C. wireless signal in the rectifying circuit **32**, and supplies this electrical energy to the load **33**.

[0011] Since such rectenna **30** of the prior art has only one frequency characteristics depending on the characteristics of the loop antenna **31**, the electrical energy cannot be supplied to the load **33** if a wireless signal having the frequency band matched in frequency characteristics of the rectenna **30** is not received.

[0012] Furthermore, the rectenna **30** of the prior art simply rectifies the received wireless signal, so the wireless signal must be amplified before transmitting the wireless signal to the rectenna **30** when the load **33** needs a high voltage.

BRIEF SUMMARY OF THE INVENTION

[0013] The first objective of this invention, to solve above-described problems, is to provide a patch antenna having structures for an easy impedance matching.

[0014] The second objective of this invention is to provide a dual-band patch antenna having dual-band signal characteristics using a single patch.

[0015] The third objective of this invention is to provide a dual-band patch antenna capable of transmitting and receiving a circularly polarized wave and a linearly polarized wave in two different frequency bands.

[0016] The fourth objective of this invention is to provide a rectenna capable of generating an additional electrical energy by receiving non-relevant signals such as Wi-Fi or Bluetooth signal existing in free space mixed with the signals which are generally utilized for wireless power transmission.

[0017] The fifth objective of this invention is to provide a rectenna capable of generating higher voltage than that of the rectified energy obtained by rectifying the input signal from the voltage multiplying circuit.

[0018] Other objectives and advantages of the present invention will be described hereinafter and become apparent through the exemplary embodiments of the present invention. The objectives and advantages of the present invention can be accomplished through the means and combination of the elements of the claims.

[0019] A patch antenna in accordance with a preferred exemplary embodiment of the present invention to solve above-described problems is comprised of: an dielectric substrate; a patch that is formed on the first surface of the dielectric substrate and assigns the first frequency characteristics; and a ground plane that is formed on the second surface of the dielectric substrate, wherein an impedance matching means, that is capacitively coupled with the patch, is formed on the one side of the ground plane.

[0020] According to the patch antenna of the present invention, it is preferred that the impedance matching means is comprised of: a slot formed in the bottom area of the ground plane; and a feeder line formed inside of the slot and being separated by a predetermined distance therefrom.

[0021] According to the patch antenna of the present invention, it is preferred that the feeder line is formed at a location where the feeder line can face the bottom area of the patch with the dielectric substrate interposed between one end of the feeder line and the patch.

[0022] According to the patch antenna of the present invention, it is preferred that the slot and the feeder line is to be formed in a T-shaped pattern, wherein the feeder line is formed at a location where the bar-shaped head of the feeder line can face the bottom area of the patch so as to provide the second frequency response by adjusting the size of the slot and the feeder line.

[0023] According to the patch antenna of the present invention, the patch may be a hexagonal shape formed by etching off two facing corners of a square.

[0024] Meanwhile, a rectenna in accordance with a preferred exemplary embodiment of the present invention, for receiving an A.C. wireless signal carrying electrical energy and converting the wireless signal into a D.C. electrical energy, is comprised of: a patch antenna for receiving the wireless signal comprising an dielectric substrate, a patch that is formed at the upper area of the surface of the dielectric substrate and characterizing the first frequency response characteristics, a ground plane formed on the other surface of the dielectric substrate, and an impedance matching means characterizing the second frequency response characteristics; and a rectifying unit that converts the wireless signal, received via the patch antenna, into a D.C. electrical energy by rectifying the wireless signal.

[0025] According to the rectenna of the present invention, it is preferred that the rectifying unit is comprised of a voltage multiplying circuit that comprises multiple capacitors and diodes and at least two of serially connected rectifying and multiplying stages so as to convert the wireless signal into an electrical energy by rectifying the wireless signal and multiply the voltage of the converted electrical energy simultaneously.

[0026] Besides, according to the rectenna of the present invention, it is preferred that the rectifying unit is further comprised of: a grounding area that is being separated from the ground plane of the patch antenna; and a bead for suppressing the effect of the A.C. signal and located between the grounding area and the ground plane of the patch antenna.

[0027] According to the rectenna of the present invention, the bead may further include an inductor.

[0028] According to the rectenna of the present invention, it is preferred that the rectifying unit is installed on the same plane of the ground plane of the patch antenna.

[0029] According to the antenna of the present invention, a capacitive coupling between the patch and the feeder line is obtained through the structure wherein the slot is formed on the ground plane and the feeder line is formed inside of the slot; when compared to the patch antennas of conventional technology, more easy impedance matching can be achieved by adjusting the design parameters of the slot and the feeder line on the ground plane; therefore, additional fabrication of another expensive patch can be avoided even in the worst-case.

[0030] According to the antenna of the present invention, it has the advantage that the production cost can be reduced by using a single patch instead of two patches; even though it has a dual-band frequency response, the size of the antenna may not be increased because another frequency response different than that of the patch is provided through adjustment of the design parameters of the T-shaped slot and the T-shaped feeder line.

[0031] According to the antenna of the present invention, it has the advantage that it can handle various signals since a circularly polarized wave and a linearly polarized wave can be transmitted and received in two different frequency bands.

[0032] According to the rectenna of the present invention, it has the advantage that it can supply more stable power compare to the conventional rectenna since it can generate an additional electrical energy by receiving non-relevant signals such as Wi-Fi or Bluetooth signal existing in free space mixed with the signals which are generally utilized for wireless power transmission.

[0033] According to the rectenna of the present invention, it has the advantage that it can be applied to various equipments because higher voltage than that of the input signal can be obtained through two-stage voltage multiplication in the rectifying unit.

[0034] According to the rectenna of the present invention, it has the advantage that the space required to install a rectenna can be reduced since the overall size can be reduced by forming the rectifying unit on the ground plane of the antenna.

BRIEF DESCRIPTION OF THE DRAWINGS

[0035] FIG. **1** is a schematic diagram of a patch antenna in accordance with a conventional technology;

[0036] FIG. **2** is a schematic diagram of a patch antenna in accordance with another conventional technology;

[0037] FIG. **3** is a schematic diagram of a rectenna in accordance with a conventional technology;

[0038] FIG. **4** is a schematic diagram of a patch antenna in accordance with a preferred exemplary embodiment of the present invention;

[0039] FIGS. **5***a* and **5***b* are example graphs illustrating simulated return loss values versus operating frequencies for different design parameters of a patch antenna in accordance with a preferred exemplary embodiment of the present invention;

[0040] FIG. **6** is a schematic diagram of a patch antenna in accordance with another preferred exemplary embodiment of the present invention;

[0041] FIG. **7** is an example graph illustrating simulation and measured return loss values of a patch antenna in accordance with another preferred exemplary embodiment of the present invention;

[0042] FIG. **8** is a schematic diagram of a rectenna in accordance with a preferred exemplary embodiment of the present invention;

[0043] FIG. **9** is a detailed circuit diagram of the rectifying unit shown in FIG. **8**; and

[0044] FIG. **10** is an example graph illustrating the simulated output values of the rectifying unit as a result of the voltage multiplication.

DETAILED DESCRIPTION OF THE INVENTION

[0045] Hereinafter, preferred exemplary embodiments of the present invention will be described in detail with reference to the attached drawings. Terminology and word used in the description and the claims of the present invention should not be construed as a typical or a lexical meaning, but construed as a meaning and concept in accordance with the technical spirit of the present invention on the basis of the principle that the inventors may properly define the concept of a terminology in order to best describe his/her invention.

[0046] Therefore, it should be understood that a variety of equivalents and modification examples that can replace the exemplary embodiments and configurations described in the drawings may exist at the time of this patent application since these exemplary embodiments and configurations are only a few examples of the present invention and not representing all the technical spirits of the present invention.

[0047] Hereinafter, a patch antenna and a rectenna in accordance with the present invention will be described with reference to the attached drawings.

[0048] FIG. **4** is a schematic diagram of a patch antenna in accordance with a preferred exemplary embodiment of the present invention. As illustrated in FIG. **4**, the patch antenna **41** in accordance with a preferred exemplary embodiment of the present invention, similar to conventional patch antenna, is comprised of an dielectric substrate **411**, a patch **412** that is formed on a surface of the dielectric substrate **411**, and a ground plane **413** that is formed on the other surface of the dielectric substrate **411**.

[0049] The dielectric substrate **411**, where a patch **412** and a ground plane **413** is formed on the opposite surface thereof facing each other, is made of a non-conducting dielectric material as the name implies.

[0050] Since this exemplary embodiment is based on the dielectric substrate **411** having thickness of 1 mm, the result of the following experiment can be affected if the thickness of the dielectric substrate **411** is different.

[0051] The patch **412** is the most essential element that determines the operating frequency of the patch antenna **41** according to the size and the shape of the patch **412**; it is formed in the upper area of the first surface of the dielectric substrate **411** and provides the frequency response of the patch antenna **41**. The patch antenna **41**, for example, having a 29 mm square patch **412** may exhibit a linear polarized wave characteristics for frequencies around 2.4 GHz.

[0052] The patch antenna **41** of the present invention is for transmitting and receiving linearly polarized waves, and it is preferred to form the patch **412** in a square shape for transmitting and receiving such linearly polarized waves.

[0053] The ground plane **413** is formed on the other side of the patch **412**, in other words, on the other surface of the dielectric substrate **411**; in the present invention, the feeder line **4142** is formed on the same plane as the ground plane **413** in order to form a coplanar waveguide (CPW) structure.

[0054] In a preferred exemplary embodiment of the present invention as shown in FIG. **4**, an impedance matching means **414** is formed inside of the ground plane **413**, and the impedance matching means **414** is comprised of: a slot **4141** formed along the length direction of the ground plane **413**; and a feeder line **4142** formed inside of the slot **4141** and being separated by a predetermined distance therefrom.

[0055] In here, the feeder line **4142** should be formed at a location where the feeder line can face the bottom area of the patch **412** with the dielectric substrate **411** interposed between one end of the feeder line and the patch.

[0056] Through above-described configuration of the present invention, the capacitive coupling between the feeder line and the patch is being accomplished.

[0057] According to the exemplary embodiment of the present invention, an easy impedance matching can be obtained by adjusting the design parameters of the ground plane using such configuration.

[0058] FIGS. **5***a* and **5***b* illustrate an example of return loss varying with the values of F and A, shown in FIG. **4**, that are the design parameters of the antenna in accordance with the present invention.

[0059] As shown in FIGS. **5***a* and **5***b*, impedance matching can be achieved at frequencies around 2.4 GHz by adjusting the values of design parameters of F and A of the slot **4141** and the feeder line **4142**; according to the simulation results, a return loss of near −30 dB can be obtained at frequencies around 2.4 GHz.

[0060] Table 1 provides examples of values of design parameters that are being adjusted so that a patch antenna in accordance with the present invention could have a linearly polarized wave characteristics at frequencies around 2.4 GHz

TABLE 1

design parameter	value [mm]	design parameter	value [mm]
W	29	A	7
L	29	B	17
X	43	U	1.865
Y	36	F	16

[0061] According to a patch antenna in accordance with a preferred exemplary embodiment of the present invention, it has the advantage that the variations in the frequency response characteristics caused by the tolerances are less pronounced compare to conventional in-set type structures; impedance matching can be achieved by simply adjusting the design parameters of the slot and the feeder line on the ground plane; therefore, additional fabrication of another patch can be avoided even when an impedance matching is failed.

[0062] A patch antenna **51** in accordance with another preferred exemplary embodiment of the present invention is illustrated in FIG. **6**.

[0063] A patch antenna having two kinds of frequency band characteristics is provided in the another preferred exemplary embodiment of the present invention.

[0064] To achieve above-described capability, a patch antenna **51** in accordance with another preferred exemplary embodiment of the present invention is comprised of an impedance matching means **514** that is formed on the ground plane **513** and comprising a T-shaped slot **5141** and a T-shaped feeder line **5142** as shown in FIG. **6**.

[0065] A patch antenna **51** in accordance with another preferred exemplary embodiment of the present invention enables an impedance matching by adjusting the design parameters of the T-shaped slot **5141** and the T-shaped feeder line **5142** as shown in FIG. **6**; in addition, a dual-band patch antenna in accordance with the present invention can obtain previously mentioned second frequency response characteristics by adjusting the design parameters of the T-shaped slot **5141** and the T-shaped feeder line **5142**, more particularly, by adjusting the values of the design parameters B, X, and M.

[0066] Table 2 provides examples of values of antenna design parameters that are being adjusted so that the patch antenna could have an additional frequency response characteristics at frequencies around 3.1 GHz

TABLE 2

design parameter	value [mm]	design parameter	value [mm]	design parameter	value [mm]
W	23.5	A	33	y	1.5
L	23.5	B	4	U	1.865
W'	3.5	C	14	M	15.4
L'	3.5	X	8	M'	14.5

[0067] According to another preferred exemplary embodiment of the present invention, a patch **512** can be a hexagonal shape formed by etching off two facing corners of a square as shown in FIG. **6**; a patch antenna adopting such type of patch possesses a circularly polarized wave characteristics at frequencies around 2.4 GHz reflecting the design parameters of the particular patch shape, and possesses a linearly polarized wave characteristics at frequencies around 3.1 GHz reflecting the design parameters of the ground plane, but showing independence in patch size.

[0068] Through the above-described configuration, a patch antenna in accordance with another preferred exemplary embodiment of the present invention can handle not only two different frequency band signals, but also two different polarization types (i.e. a circular and a linear polarization).

[0069] Hereinafter, a rectenna in accordance with a preferred exemplary embodiment of the present invention will be described with reference to FIG. **8**. In the following description, an element with the symbol number quoted from FIG. **6** is considered as equivalent as a corresponding element in FIG. **4** performing same function as the function of the corresponding element in FIG. **6**, and vice versa.

[0070] FIG. **8** is a schematic diagram of a rectenna in accordance with a preferred exemplary embodiment of the present invention. A rectenna **50** in accordance with a preferred exemplary embodiment of the present invention is comprised of a patch antenna **51** and a rectifying unit **52** as shown in FIG. **8**.

[0071] Since the patch antenna **51** in FIG. **8** has same function and configuration as the patch antenna shown in FIGS. **4** and **6**, it will not be described hereinafter.

[0072] The rectifying unit **52** rectifies wireless signals received via the patch antenna **51**, and converts the rectified signal into an electrical energy.

[0073] The rectifying unit **52** can be integrated on the same planar surface with the patch antenna **51** using the bottom area of the ground plane **513** of the patch antenna **51** as shown in FIG. **8**; using such configuration, the total area of the rectenna **50** of the present invention can be reduced.

[0074] Meanwhile, a detailed circuit diagram of the rectifying unit **52** is shown in FIG. **9**. As shown in FIG. **9**, the rectifying unit **52** is comprised of multiple capacitors C**1** to C**4** and multiple diodes D**1**, to D**4**, wherein at least two rectifying and multiplying stages A**1** and A**2**, storing electrical energy by rectification effect, are serially connected, and the wireless signal received via the patch antenna **51** is rectified and converted into a D.C. electrical energy and the voltage is multiplied simultaneously.

[0075] More particularly, the rectifying unit **52** in accordance with the present invention is comprised of at least two rectifying and multiplying stages A**1** and A**2**.

[0076] Individual rectifying and multiplying stages have capacitors C**1**, C**2** and C**3**, connected between the serially connected diodes for temporary storage, and the storage capacitor C**4**; in addition, the rectifying and multiplying stages have forward biased diodes D**1** and D**3**, and reverse biased diodes D**2** and D**4** with respect to the ground G.

[0077] When the input voltage is reverse direction (i.e. negative potential with respect to the ground), diodes D**2** and D**4** are turned on so as to charge C**1** and C**2** with electrical energy; later, when the input voltage is forward direction (i.e. positive potential with respect to the ground), electrical energy input coming through the forward biased diodes D**1** and D**3** is simultaneously superimposed on the discharging electrical energy from the capacitors C**1** and C**2** thereby voltage multiplication is achieved.

[0078] Each output voltage from each rectifying and multiplying stage is stored in the capacitors C**3** and C**4**, and the electrical energy stored in the capacitor C**3** eventually flows into the capacitor C**4** due to the serially connected circuit configuration; therefore, the final voltage of the electrical energy stored in the capacitor C**4** is much higher than the initial input voltage to the rectifying unit **52**.

[0079] Although a rectifying unit configuration having only two rectifying and multiplying stages connected each other is described in this exemplary embodiment of the present invention, higher output voltage can be easily achieved by simply connecting more number of rectifying and multiplying stages.

[0080] The graphs illustrating the results of rectification and multiplication of the received wireless signal is shown in FIG. **10**.

[0081] As shown in FIG. **10**, a higher output voltage can be obtained through voltage multiplication by using wireless signal in each individual frequency band, i.e. 2.4 GHz or 3.1 GHz; much higher output voltage can be obtained by using wireless signals in both frequency bands.

[0082] Meanwhile, the rectifying unit **52** comprises a grounding area G that is isolated from the ground plane **513** of the patch antenna **51**, and may further comprise a bead **53** between the grounding area G of the rectifying unit **52** and the ground plane **513** of the patch antenna **51** as shown in FIG. **8**.

[0083] The bead **53** may comprise an inductor; undesirable effects between RF and DC signals can be suppressed by such bead.

[0084] In the rectenna **50**, for a good grounding of the rectifying unit **52**, it is desirable that the grounding area G of the rectifying unit **51** is connected to the ground plane **513** that is the largest conductor; but direct connection of the ground of the patch antenna **51** that handles RF signal to the grounding area G of the rectifying unit **52** that handles D.C. signal could cause a change in the circuit characteristics deviating from the designers intention.

[0085] To prevent such problems, the ground plane **513** and the grounding area G is connected by an inductor so that the ground plane **513** and the grounding area G is electrically disconnected from each other for an A.C. signal while the ground plane **513** and the grounding area G is electrically connected to each other for a D.C. signal.

[0086] Through the above-described configuration, the rectifying unit **52** can be effectively grounded and undesirable changes in the circuit characteristics of the patch antenna **51** can be prevented too.

[0087] According to the previously described rectenna of the present invention, the rectenna converts received main wireless signal of linearly polarized wave into an electrical energy for supplying power in 3.1 GHz frequency band; at the same time, the rectenna additionally generates an electrical energy by receiving circularly polarized wireless signals in 2.4 GHz frequency band such as Wi-Fi, Bluetooth, or DMB signal; therefore, more stable and smooth power can be supplied even in case the reception of the main wireless signal is rather poor.

[0088] Although the present invention herein has been described in the above with reference to the exemplary embodiments, it will be apparent to those skilled in the art that various changes and modification may be made to the above described embodiments, without departing from the scope and spirit of the present invention as disclosed in the accompanying claims.

What is claimed is:

1. A patch antenna comprising:

an dielectric substrate;

a patch that is formed on the first surface of the dielectric substrate and provides the first frequency characteristics; and

a ground plane that is formed on the second surface of the dielectric substrate, wherein an impedance matching means, that is capacitively coupled with the patch, is formed on the one side of the ground plane.

2. The patch antenna according to claim **1**, wherein the impedance matching means is comprised of:

a slot formed in the bottom area of the ground plane; and

a feeder line formed inside of the slot and being separated by a predetermined distance therefrom.

3. The patch antenna according to claim **2**, wherein the feeder line is formed at a location where the feeder line can face the bottom area of the patch with the dielectric substrate interposed between one end of the feeder line and the patch.

4. The patch antenna according to claim **3**, wherein the slot and the feeder line is to be formed in a T-shaped pattern,

wherein the feeder line is formed at a location where the bar-shaped head of the feeder line can face the bottom area of the patch so as to provide the second frequency characteristics by adjusting the size of the slot and the feeder line.

5. The patch antenna according to claim **4**, wherein the patch is a hexagonal shape formed by etching off two facing corners of a square.

6. A rectenna for receiving an A.C. wireless signal carrying electrical energy, and converting the wireless signal into a D.C. electrical energy, comprising:

a patch antenna for receiving the wireless signal comprised of a dielectric substrate, a patch that is formed at the upper area of the surface of the dielectric substrate and provides the first frequency response characteristics, a ground plane formed on the other surface of the dielectric substrate, and an impedance matching means that provides the second frequency response characteristics; and

a rectifying unit for converting the wireless signal, received via the patch antenna, into a D.C. electrical energy by rectifying the wireless signal.

7. The rectenna to claim **6**, wherein the rectifying unit is comprised of a voltage multiplying circuit that comprises at least one capacitor and diode and at least two of serially connected rectifying and multiplying stages capable of storing energy temporarily so as to convert the wireless signal into an electrical energy by rectifying the wireless signal and multiply the voltage of the converted electrical energy simultaneously.

8. The patch antenna according to claim **6**, wherein the rectifying unit is further comprised of:

a grounding area that is being separated from the ground plane of the patch antenna; and

a bead that is located between the grounding area and the ground plane of the patch antenna for suppressing the effect of the A.C. signal.

* * * * *

US 20150117281A1

(19) **United States**

(12) **Patent Application Publication**

Khlat et al.

(10) **Pub. No.: US 2015/0117281 A1**

(43) **Pub. Date: Apr. 30, 2015**

(54) **RF DIPLEXER**

(71) Applicant: **RF Micro Devices, Inc.**, Greensboro, NC (US)

(72) Inventors: **Nadim Khlat**, Cugnaux (FR); **Marcus Granger-Jones**, Scotts Valley, CA (US)

(21) Appl. No.: **14/547,502**

(22) Filed: **Nov. 19, 2014**

Related U.S. Application Data

(63) Continuation-in-part of application No. 14/523,065, filed on Oct. 24, 2014.

(60) Provisional application No. 61/906,622, filed on Nov. 20, 2013, provisional application No. 61/895,386, filed on Oct. 24, 2013.

Publication Classification

(51) **Int. Cl.**
H04L 5/08 (2006.01)

(52) **U.S. Cl.**
CPC ***H04L 5/08*** (2013.01)

(57) **ABSTRACT**

Embodiments of a tunable radio frequency (RF) diplexer and methods of operating the same are disclosed. In one embodiment, the RF diplexer includes a first hybrid coupler, a second hybrid coupler, an RF filter circuit, and a phase inversion component. Both the RF filter circuit and the phase inversion component are connected between the first hybrid coupler and the second hybrid coupler. The phase inversion component is configured to provide approximately a differential phase. The RF filter circuit is configured to provide a passband and a notch. The RF filter circuit is tunable to provide the notch on both a high-frequency side of the passband and a low frequency side of the passband. Accordingly, the tunable RF diplexer provides lower insertion losses and higher isolation regardless of whether the one of the diplexed frequency bands is set at higher frequencies or lower frequencies than the other diplexed frequency band.

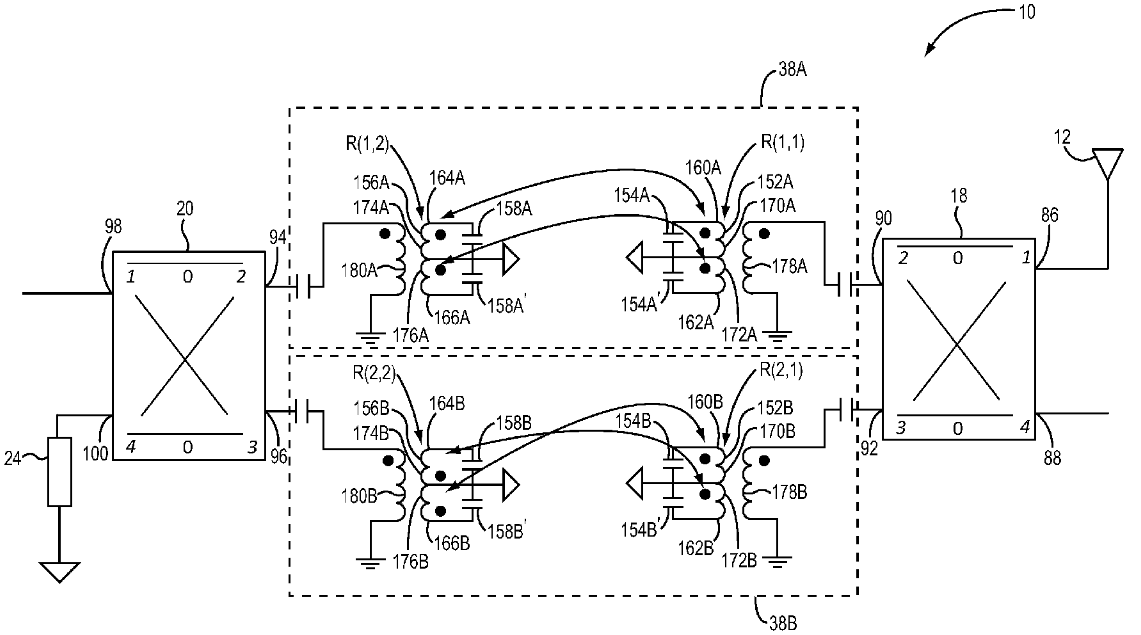
10
38A
12
R(1,2)
R(1,1)
156A
164A
160A
152A
174A
158A
154A
170A
20
18
98
94
90
86
1
0
2
180A
178A
2
0
1
158A'
154A'
166A
162A
176A
172A
R(2,2)
R(2,1)
156B
164B
160B
152B
174B
158B
154B
170B
24
100
4
0
3
96
92
3
0
4
88
180B
178B
158B'
154B'
166B
162B
176B
172B
38B

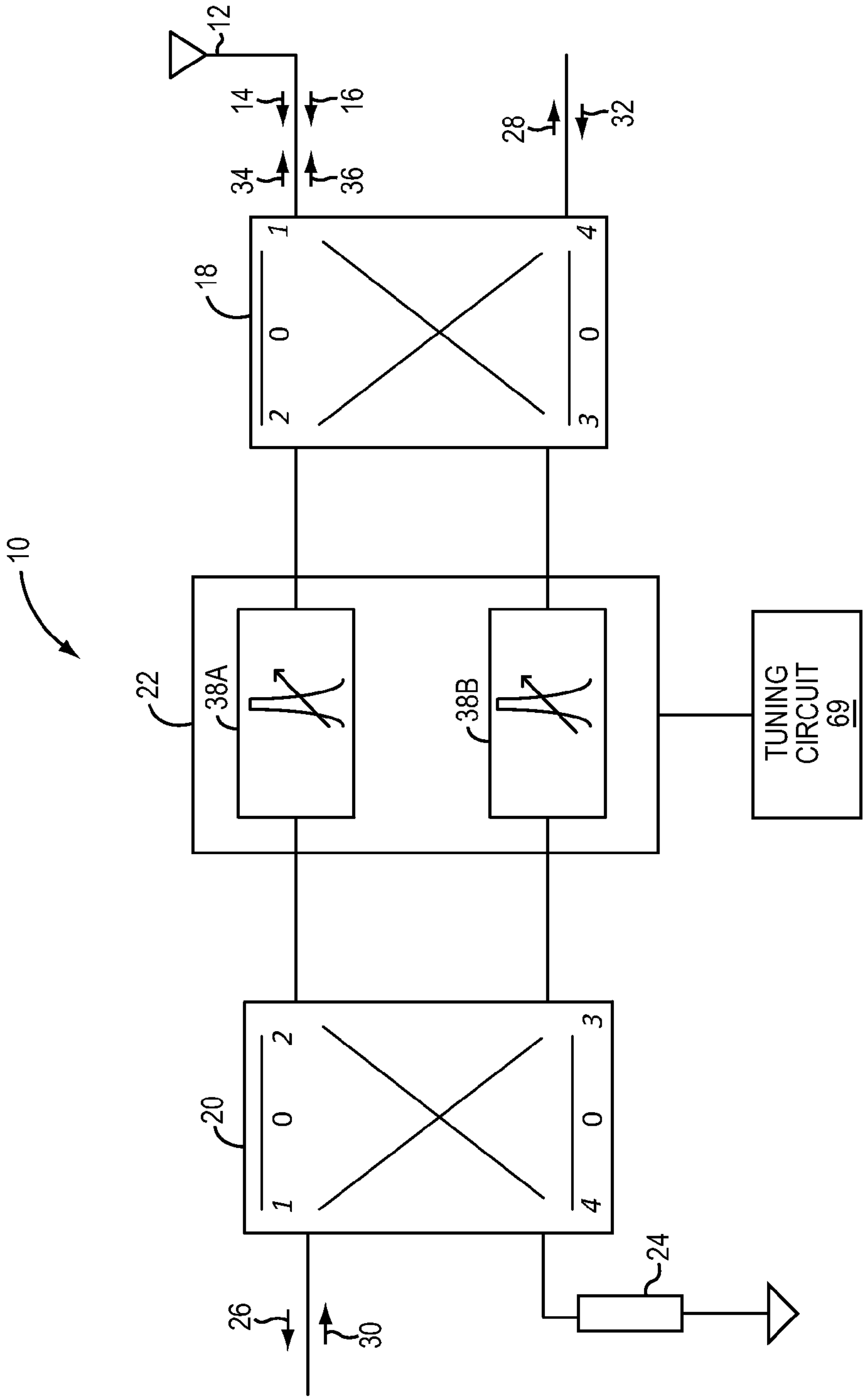
TUNING CIRCUIT 69

FIG. 1

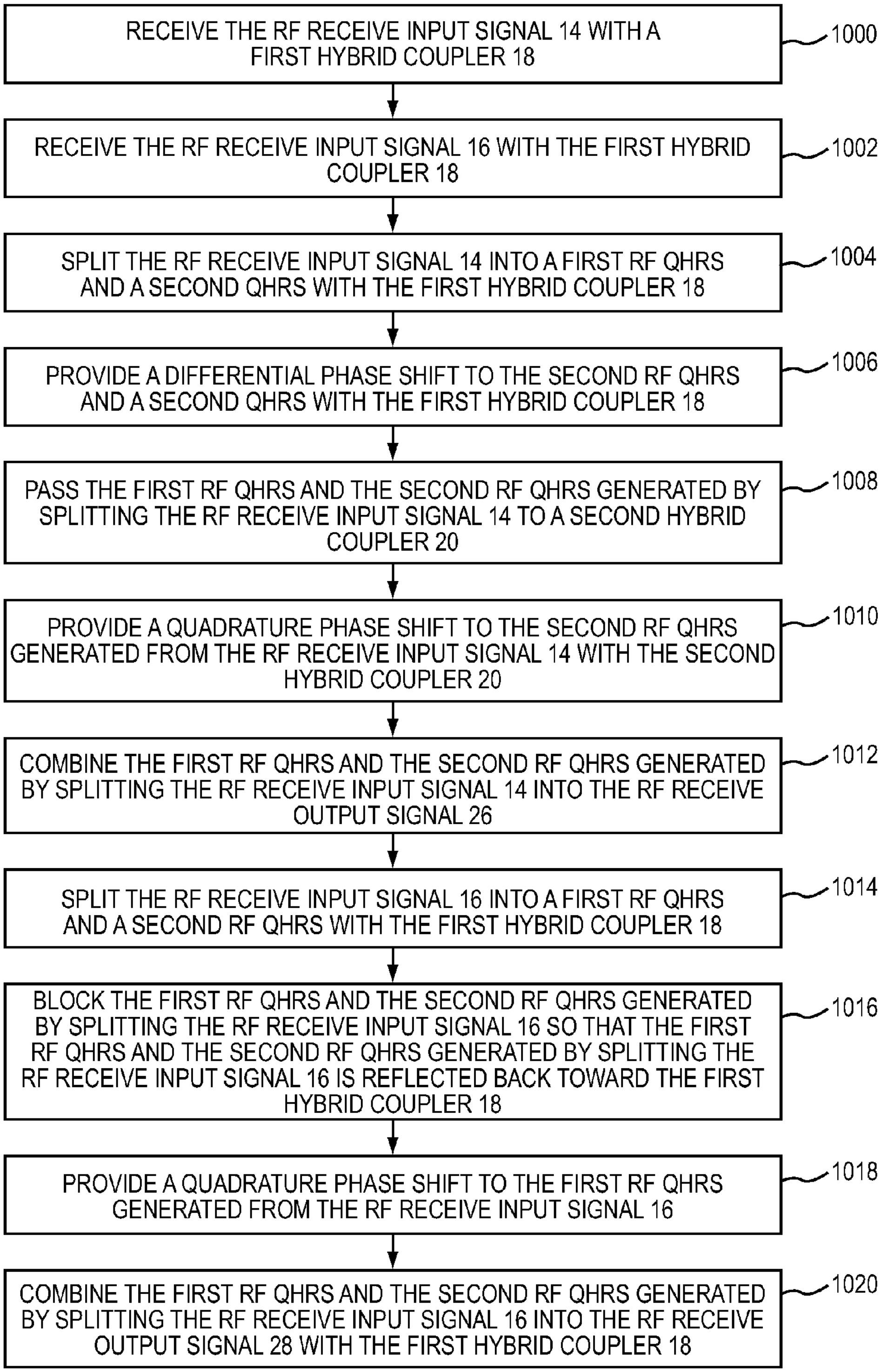
RECEIVE THE RF RECEIVE INPUT SIGNAL 14 WITH A FIRST HYBRID COUPLER 18
1000
RECEIVE THE RF RECEIVE INPUT SIGNAL 16 WITH THE FIRST HYBRID COUPLER 18
1002
SPLIT THE RF RECEIVE INPUT SIGNAL 14 INTO A FIRST RF QHRS AND A SECOND QHRS WITH THE FIRST HYBRID COUPLER 18
1004
PROVIDE A DIFFERENTIAL PHASE SHIFT TO THE SECOND RF QHRS AND A SECOND QHRS WITH THE FIRST HYBRID COUPLER 18
1006
PASS THE FIRST RF QHRS AND THE SECOND RF QHRS GENERATED BY SPLITTING THE RF RECEIVE INPUT SIGNAL 14 TO A SECOND HYBRID COUPLER 20
1008
PROVIDE A QUADRATURE PHASE SHIFT TO THE SECOND RF QHRS GENERATED FROM THE RF RECEIVE INPUT SIGNAL 14 WITH THE SECOND HYBRID COUPLER 20
1010
COMBINE THE FIRST RF QHRS AND THE SECOND RF QHRS GENERATED BY SPLITTING THE RF RECEIVE INPUT SIGNAL 14 INTO THE RF RECEIVE OUTPUT SIGNAL 26
1012
SPLIT THE RF RECEIVE INPUT SIGNAL 16 INTO A FIRST RF QHRS AND A SECOND RF QHRS WITH THE FIRST HYBRID COUPLER 18
1014
BLOCK THE FIRST RF QHRS AND THE SECOND RF QHRS GENERATED BY SPLITTING THE RF RECEIVE INPUT SIGNAL 16 SO THAT THE FIRST RF QHRS AND THE SECOND RF QHRS GENERATED BY SPLITTING THE RF RECEIVE INPUT SIGNAL 16 IS REFLECTED BACK TOWARD THE FIRST HYBRID COUPLER 18
1016
PROVIDE A QUADRATURE PHASE SHIFT TO THE FIRST RF QHRS GENERATED FROM THE RF RECEIVE INPUT SIGNAL 16
1018
COMBINE THE FIRST RF QHRS AND THE SECOND RF QHRS GENERATED BY SPLITTING THE RF RECEIVE INPUT SIGNAL 16 INTO THE RF RECEIVE OUTPUT SIGNAL 28 WITH THE FIRST HYBRID COUPLER 18
1020

FIG. 2A

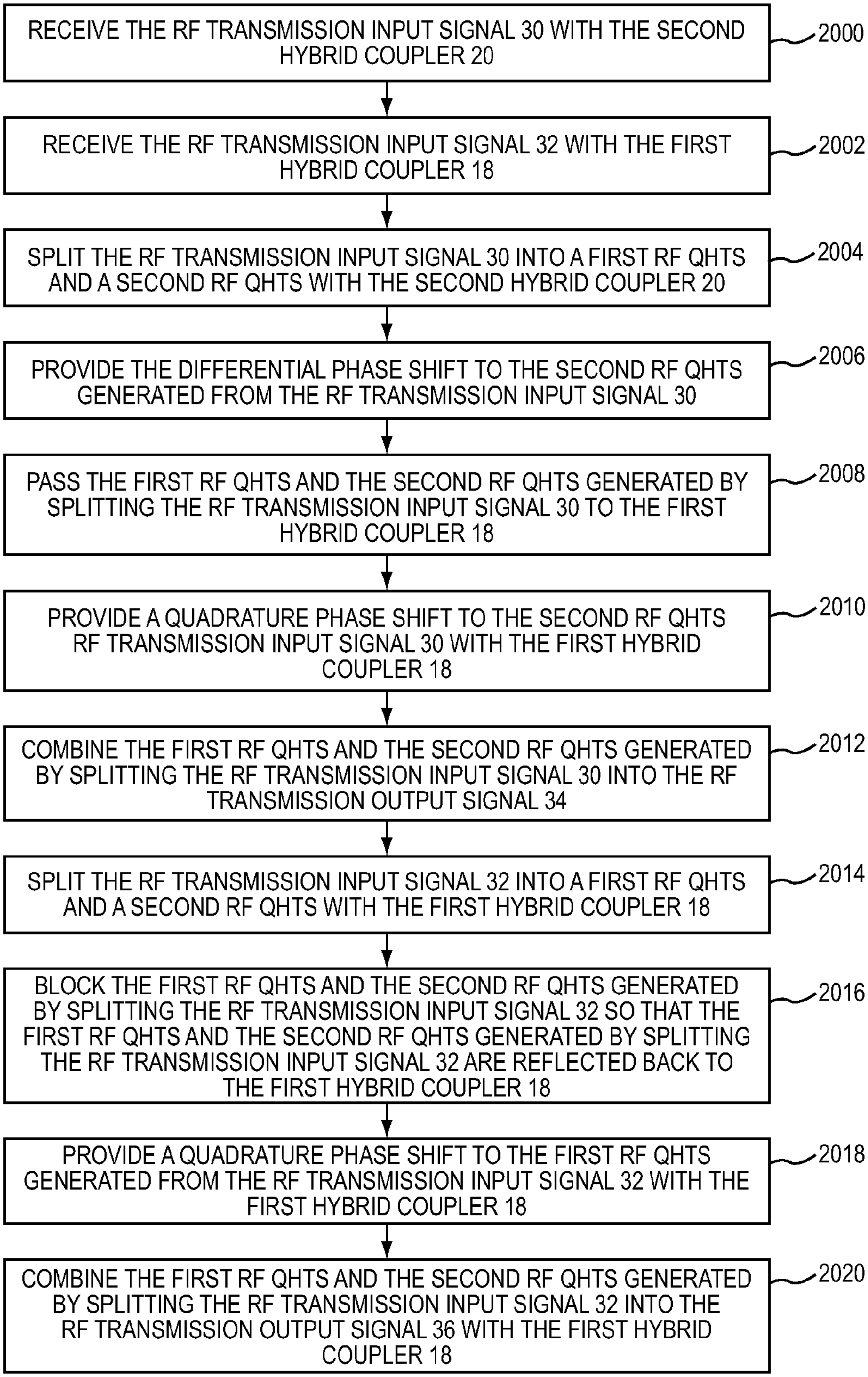

2000 RECEIVE THE RF TRANSMISSION INPUT SIGNAL 30 WITH THE SECOND HYBRID COUPLER 20

2002 RECEIVE THE RF TRANSMISSION INPUT SIGNAL 32 WITH THE FIRST HYBRID COUPLER 18

2004 SPLIT THE RF TRANSMISSION INPUT SIGNAL 30 INTO A FIRST RF QHTS AND A SECOND RF QHTS WITH THE SECOND HYBRID COUPLER 20

2006 PROVIDE THE DIFFERENTIAL PHASE SHIFT TO THE SECOND RF QHTS GENERATED FROM THE RF TRANSMISSION INPUT SIGNAL 30

2008 PASS THE FIRST RF QHTS AND THE SECOND RF QHTS GENERATED BY SPLITTING THE RF TRANSMISSION INPUT SIGNAL 30 TO THE FIRST HYBRID COUPLER 18

2010 PROVIDE A QUADRATURE PHASE SHIFT TO THE SECOND RF QHTS RF TRANSMISSION INPUT SIGNAL 30 WITH THE FIRST HYBRID COUPLER 18

2012 COMBINE THE FIRST RF QHTS AND THE SECOND RF QHTS GENERATED BY SPLITTING THE RF TRANSMISSION INPUT SIGNAL 30 INTO THE RF TRANSMISSION OUTPUT SIGNAL 34

2014 SPLIT THE RF TRANSMISSION INPUT SIGNAL 32 INTO A FIRST RF QHTS AND A SECOND RF QHTS WITH THE FIRST HYBRID COUPLER 18

2016 BLOCK THE FIRST RF QHTS AND THE SECOND RF QHTS GENERATED BY SPLITTING THE RF TRANSMISSION INPUT SIGNAL 32 SO THAT THE FIRST RF QHTS AND THE SECOND RF QHTS GENERATED BY SPLITTING THE RF TRANSMISSION INPUT SIGNAL 32 ARE REFLECTED BACK TO THE FIRST HYBRID COUPLER 18

2018 PROVIDE A QUADRATURE PHASE SHIFT TO THE FIRST RF QHTS GENERATED FROM THE RF TRANSMISSION INPUT SIGNAL 32 WITH THE FIRST HYBRID COUPLER 18

2020 COMBINE THE FIRST RF QHTS AND THE SECOND RF QHTS GENERATED BY SPLITTING THE RF TRANSMISSION INPUT SIGNAL 32 INTO THE RF TRANSMISSION OUTPUT SIGNAL 36 WITH THE FIRST HYBRID COUPLER 18

FIG. 2B

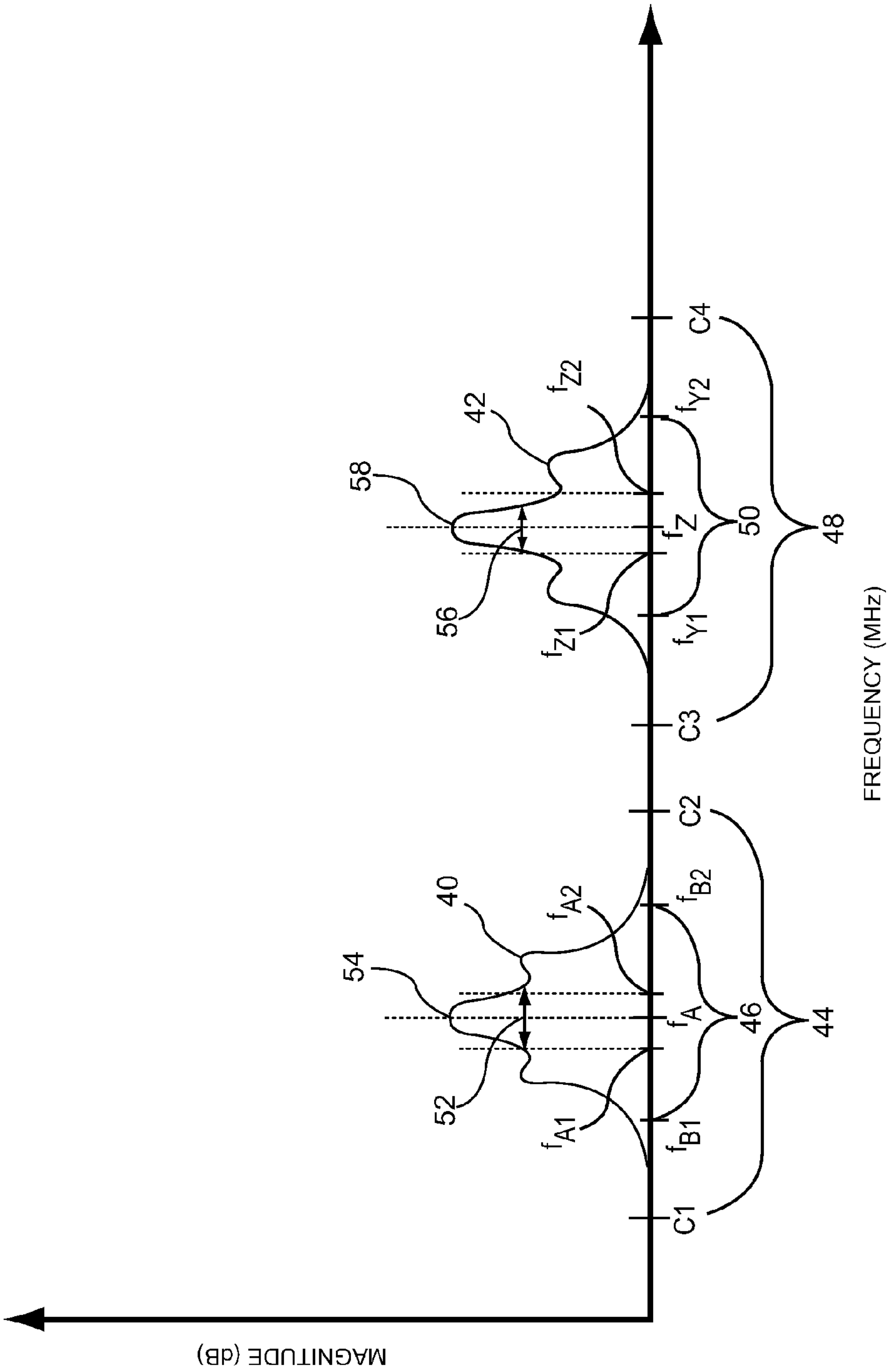
MAGNITUDE (dB)
FREQUENCY (MHz)
C1
C2
C3
C4
f_{A1}
f_{A2}
f_A
f_{B1}
f_{B2}
f_{Z1}
f_{Z2}
f_Z
f_{Y1}
f_{Y2}
40
42
44
46
48
50
52
54
56
58

FIG. 2C

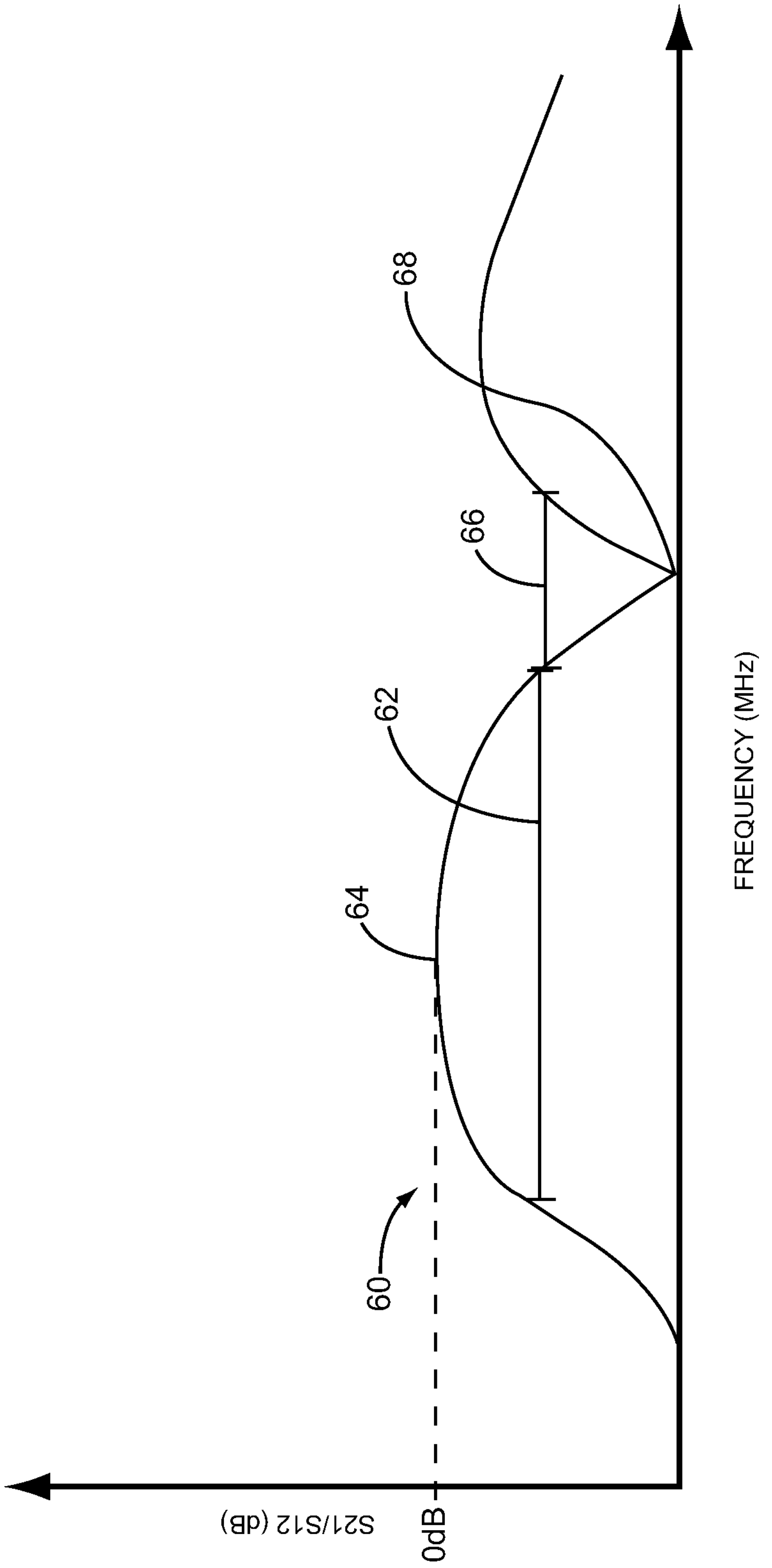

FIG. 2D

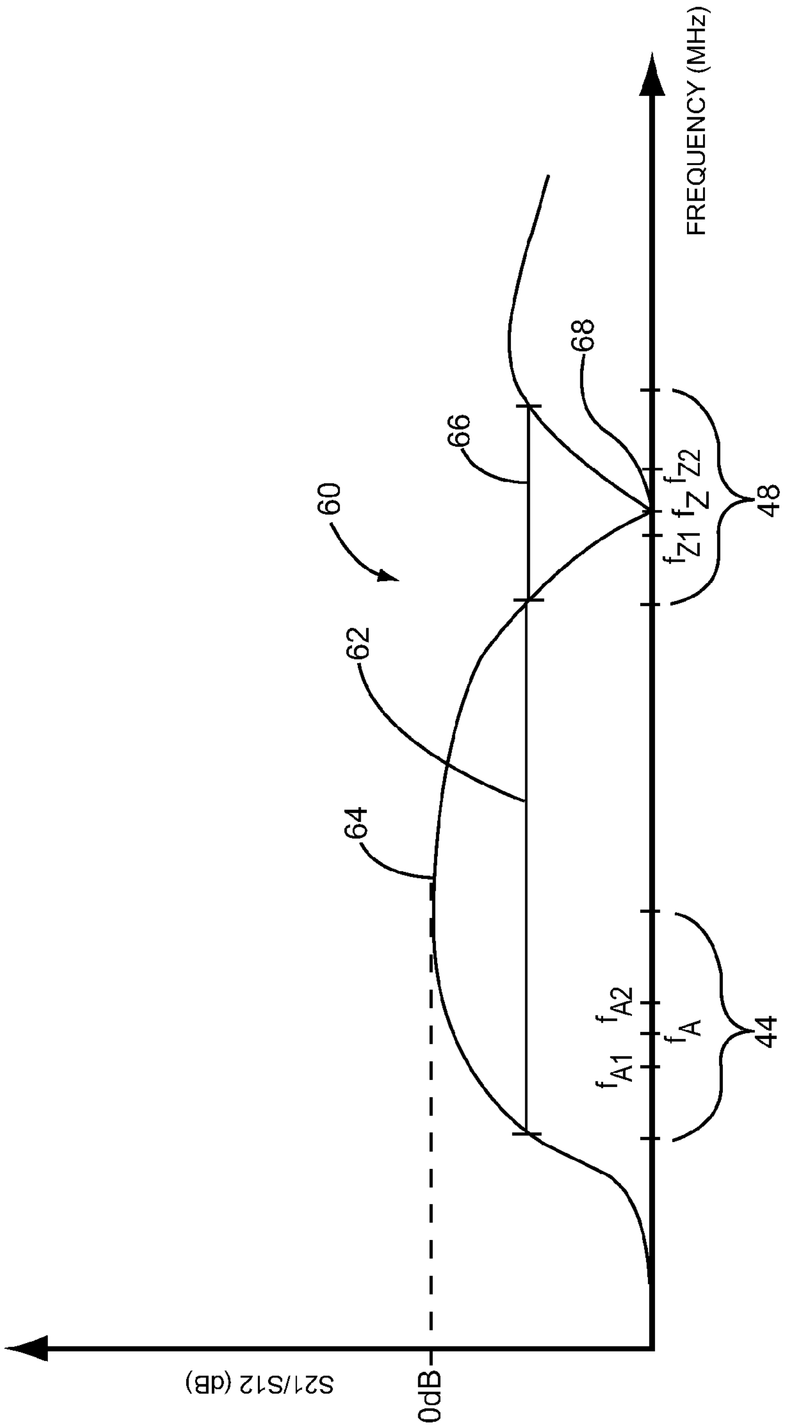
FREQUENCY (MHz)
S21/S12 (dB)
0dB
60
62
64
66
68
f_{A1}
f_A
f_{A2}
44
f_{Z1}
f_Z
f_{Z2}
48

FIG. 2E

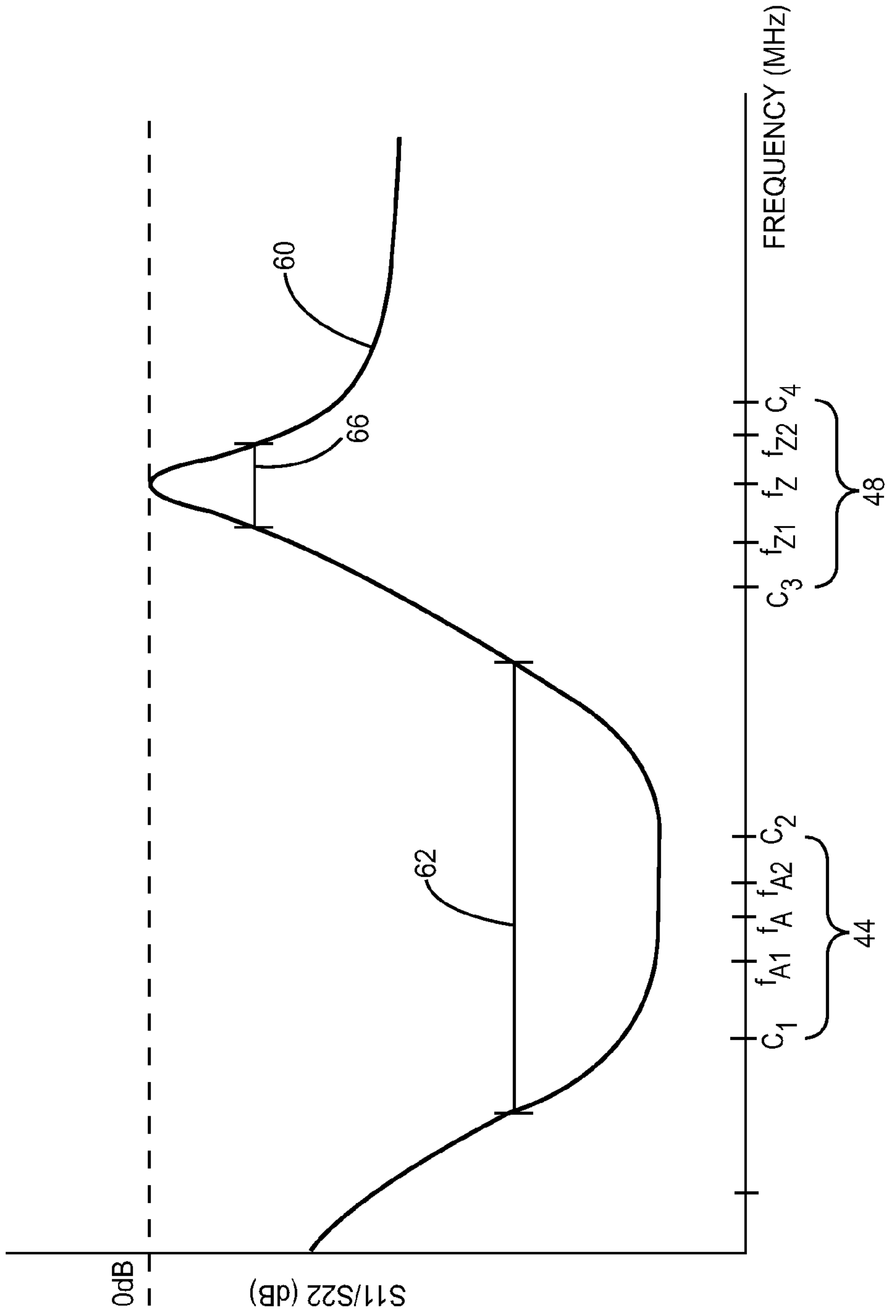
0dB
S11/S22 (dB)
FREQUENCY (MHz)
60
66
62
C_1
f_{A1}
f_A
f_{A2}
C_2
44
C_3
f_{Z1}
f_Z
f_{Z2}
C_4
48

FIG. 2F

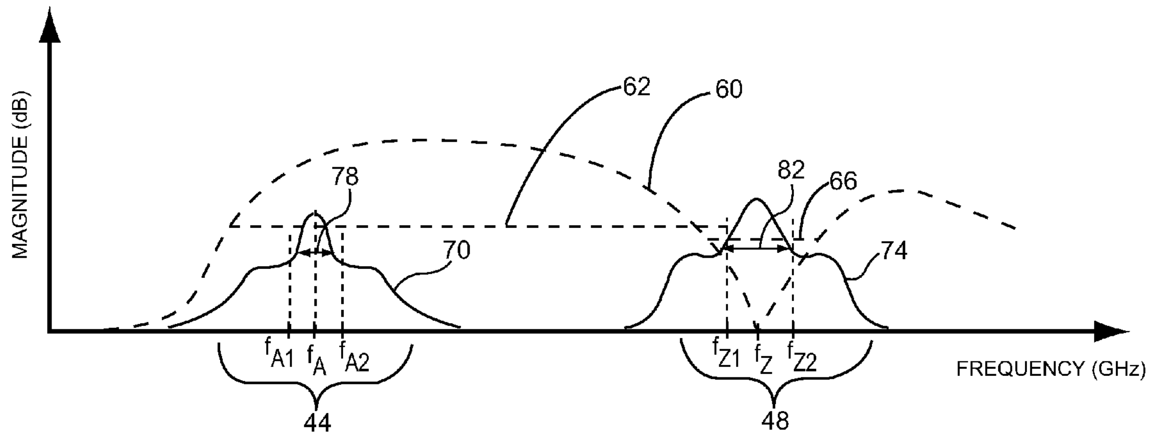
MAGNITUDE (dB)
62
60
78
82
66
70
74
f_{A1}
f_A
f_{A2}
f_{Z1}
f_Z
f_{Z2}
FREQUENCY (GHz)
44
48

FIG. 2G

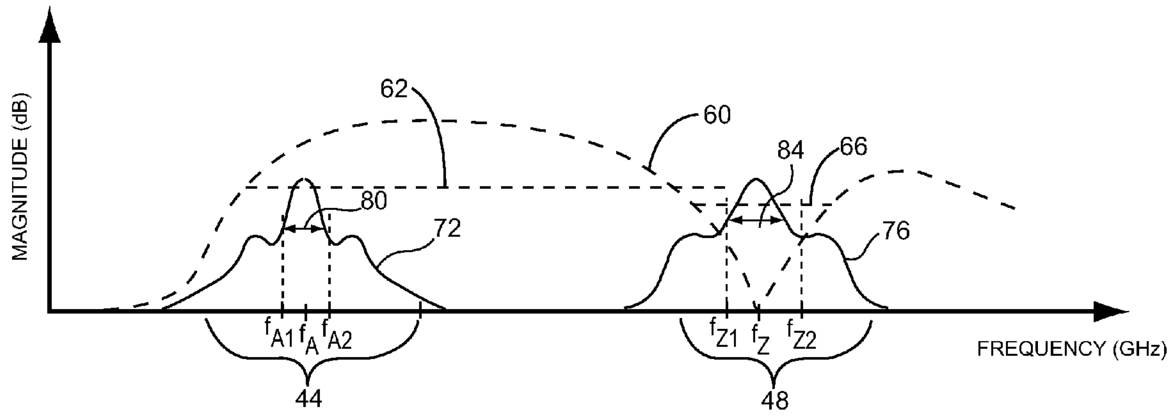
MAGNITUDE (dB)
62
60
84
66
80
72
76
f_{A1}
f_A
f_{A2}
f_{Z1}
f_Z
f_{Z2}
FREQUENCY (GHz)
44
48

FIG. 2H

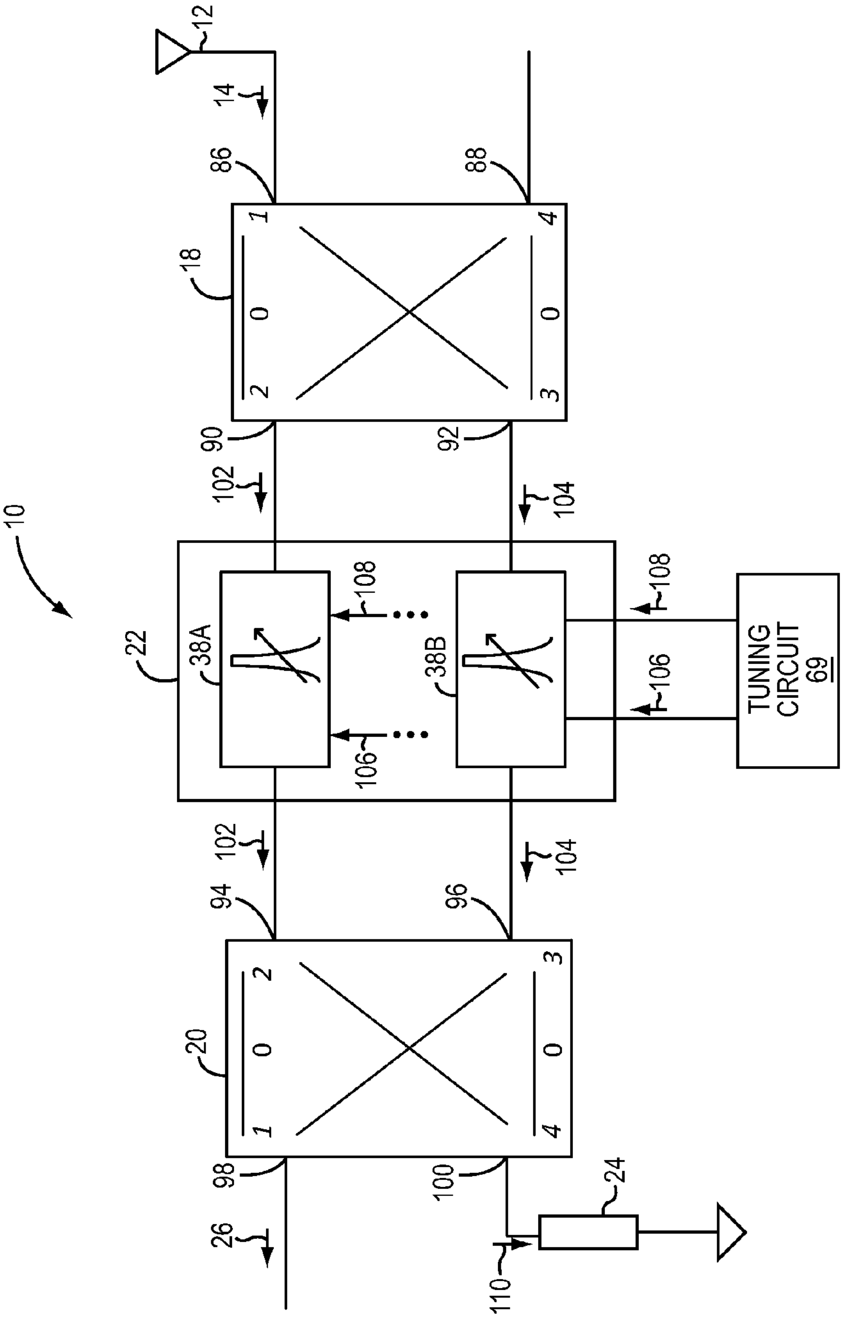
TUNING CIRCUIT 69

FIG. 3A

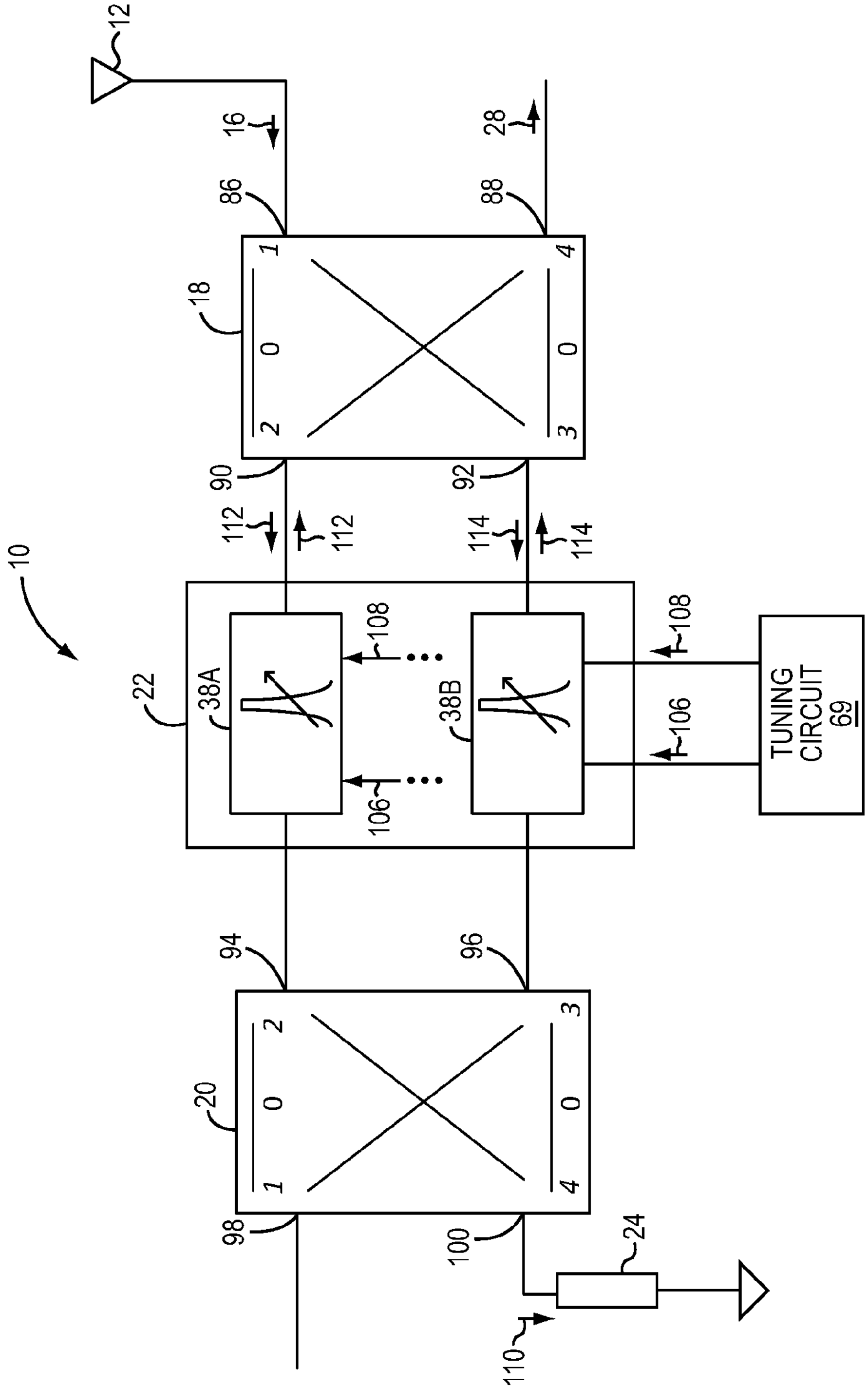
TUNING CIRCUIT 69

FIG. 3B

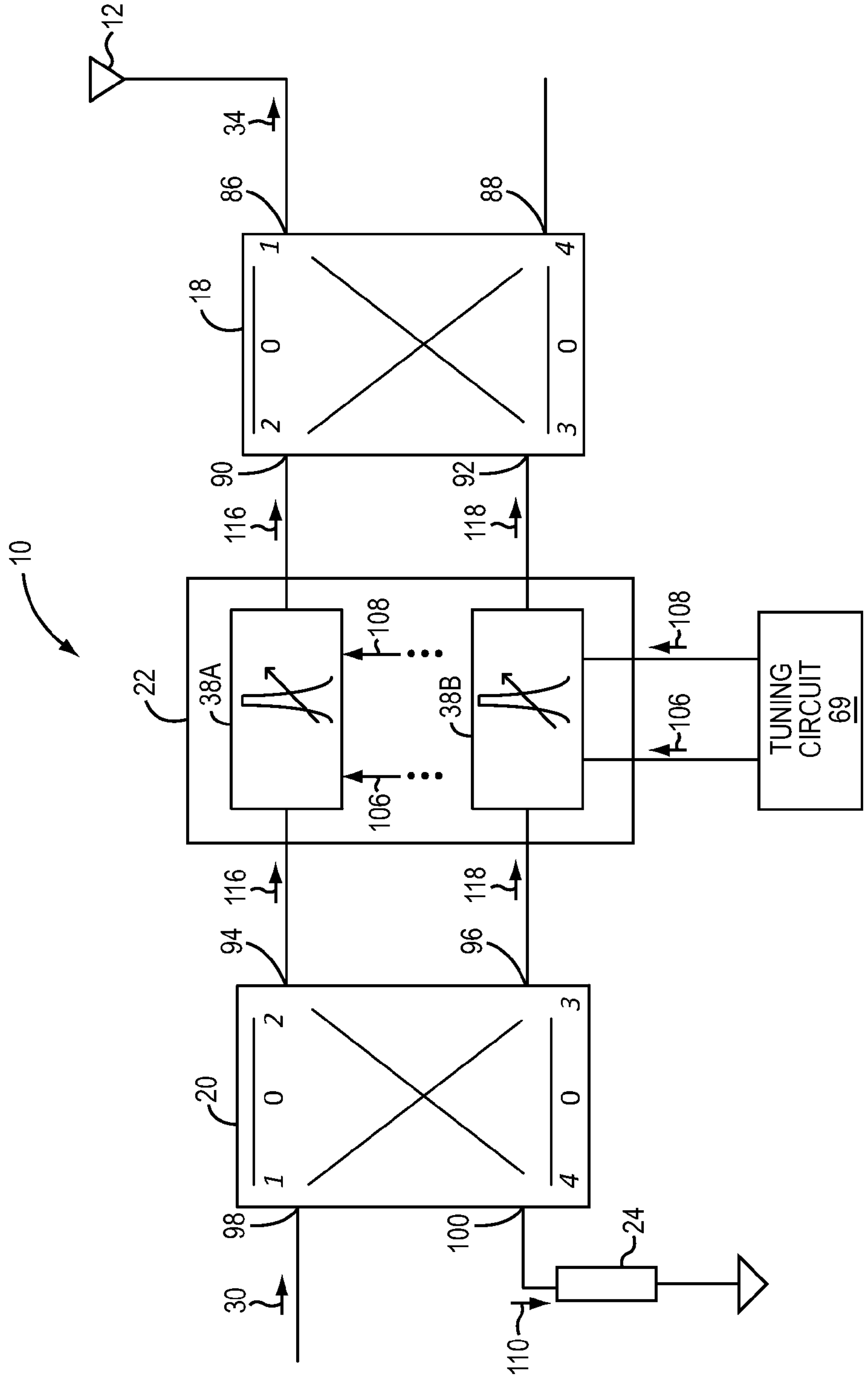
TUNING CIRCUIT 69

FIG. 3C

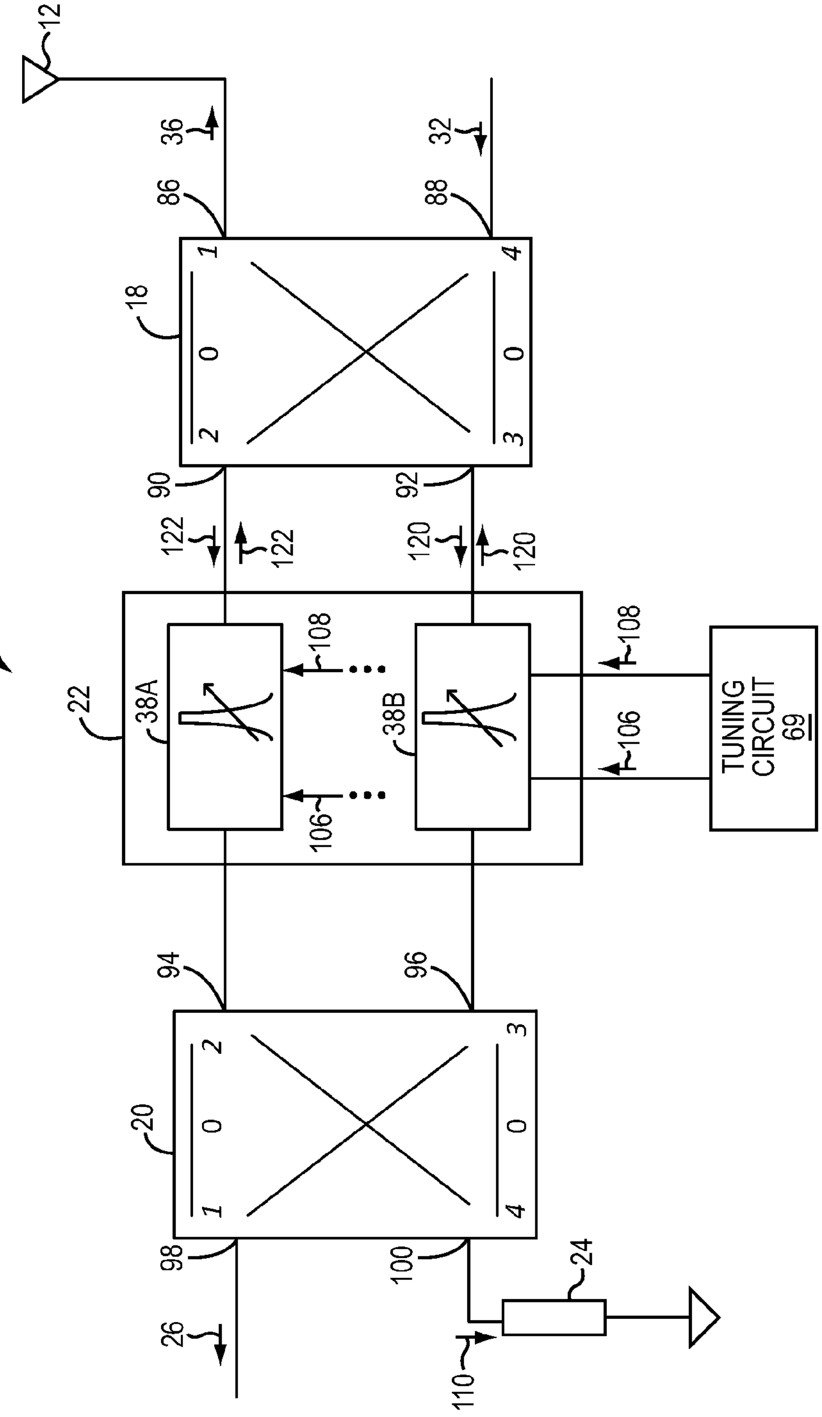
TUNING CIRCUIT 69

FIG. 3D

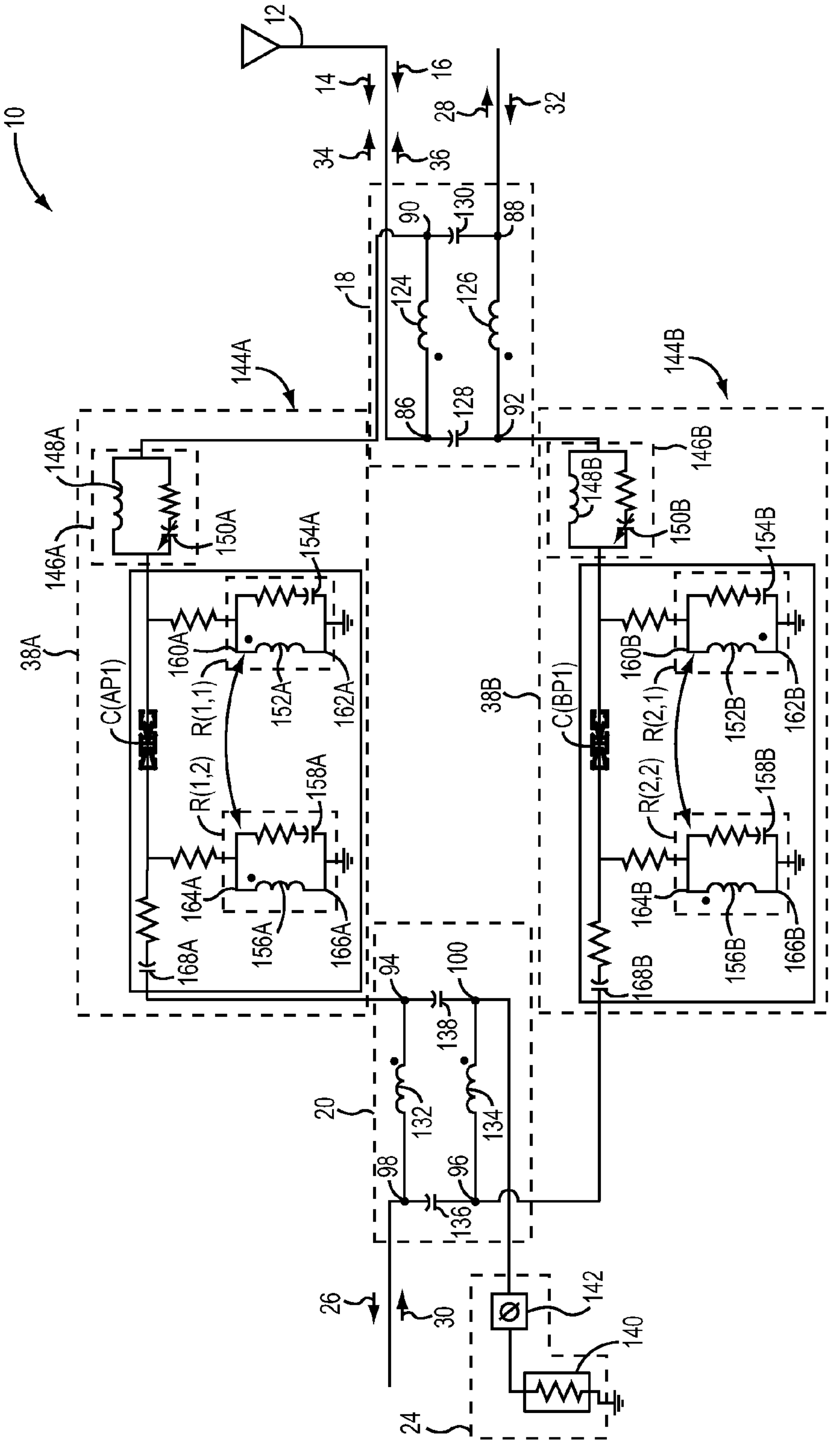
C(AP1)
R(1,1)
R(1,2)
C(BP1)
R(2,1)
R(2,2)
Ø

FIG. 4

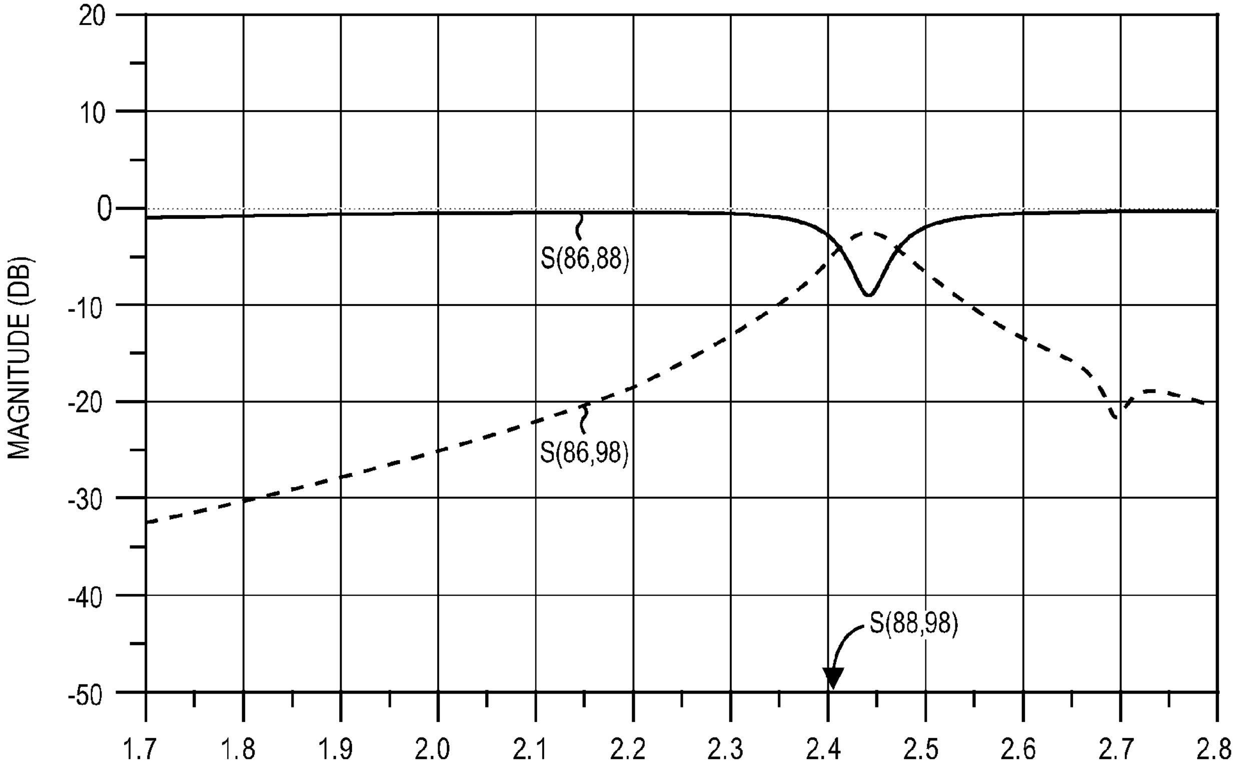
20
10
0
-10
-20
-30
-40
-50
MAGNITUDE (DB)
1.7
1.8
1.9
2.0
2.1
2.2
2.3
2.4
2.5
2.6
2.7
2.8
S(86,88)
S(86,98)
S(88,98)

FIG. 5

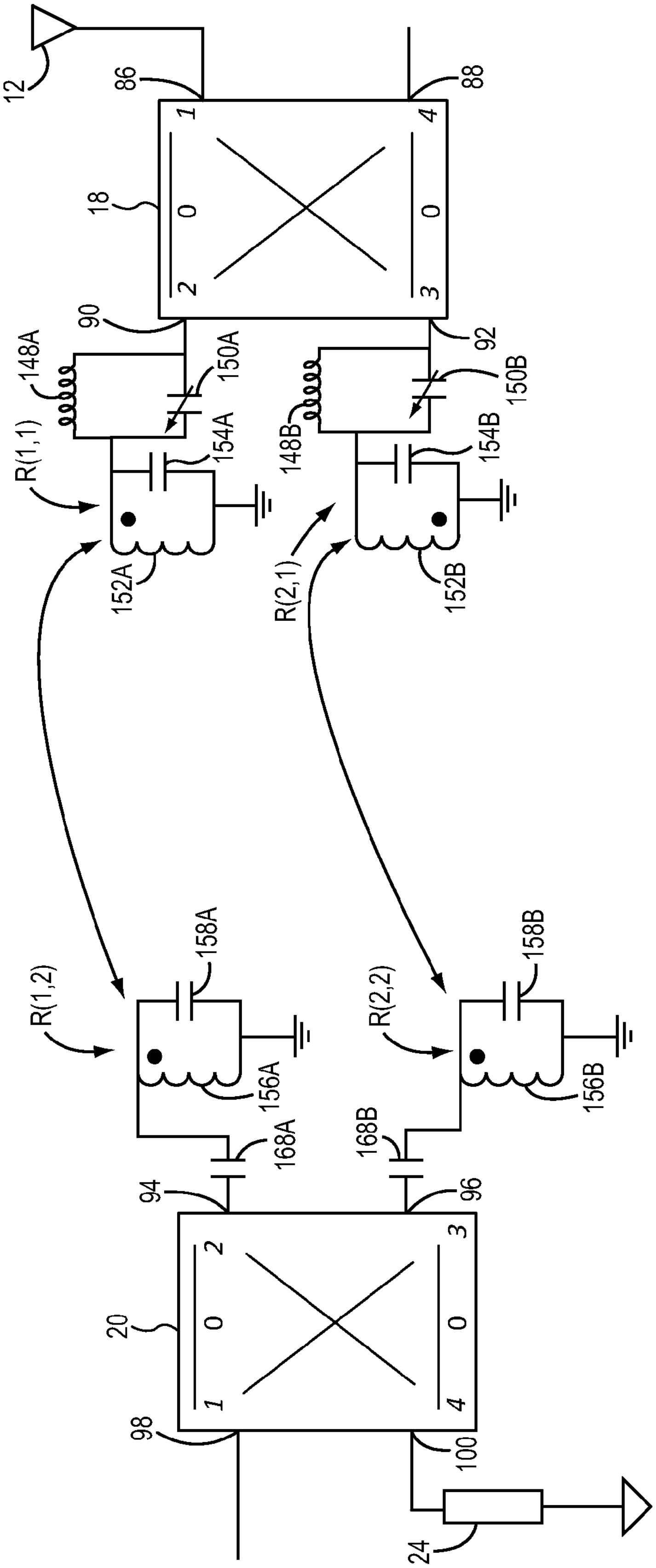
R(1,1)
R(2,1)
R(1,2)
R(2,2)

FIG. 6

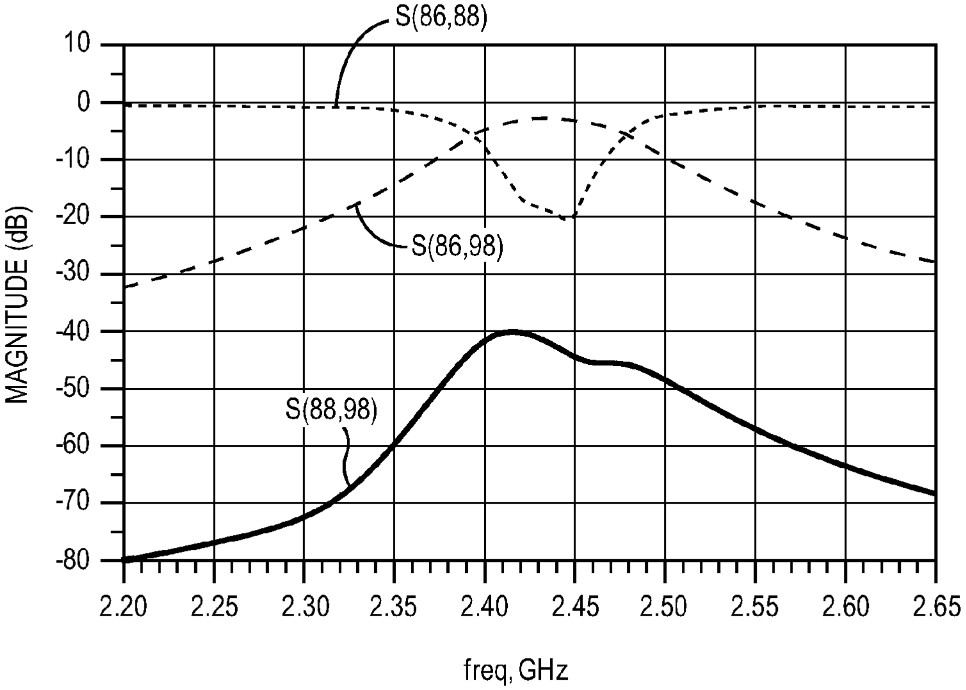
S(86,88)
S(86,98)
S(88,98)
MAGNITUDE (dB)
10
0
-10
-20
-30
-40
-50
-60
-70
-80
2.20
2.25
2.30
2.35
2.40
2.45
2.50
2.55
2.60
2.65
freq, GHz

FIG. 7

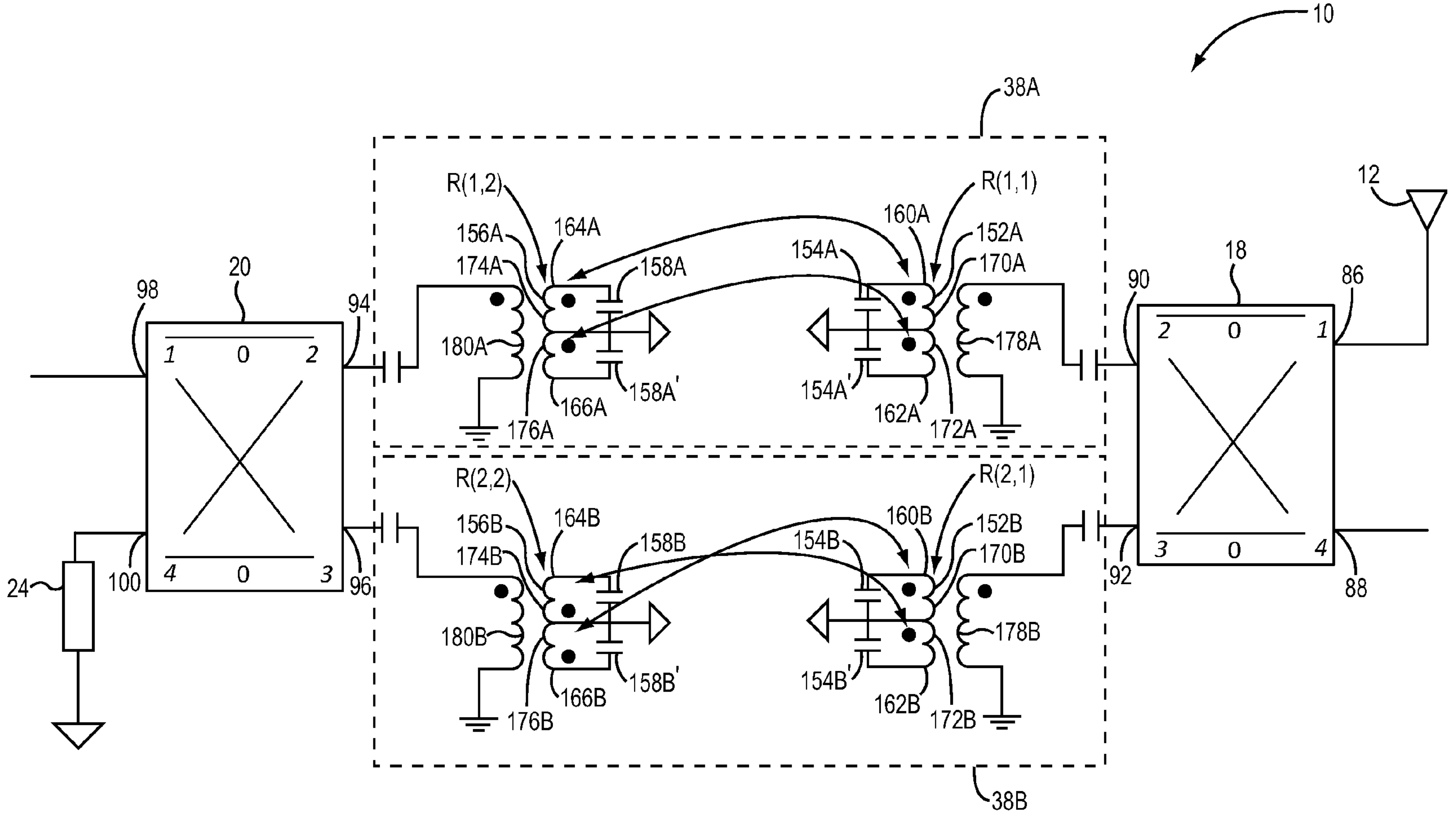
R(1,2)
R(1,1)
R(2,2)
R(2,1)

FIG. 8

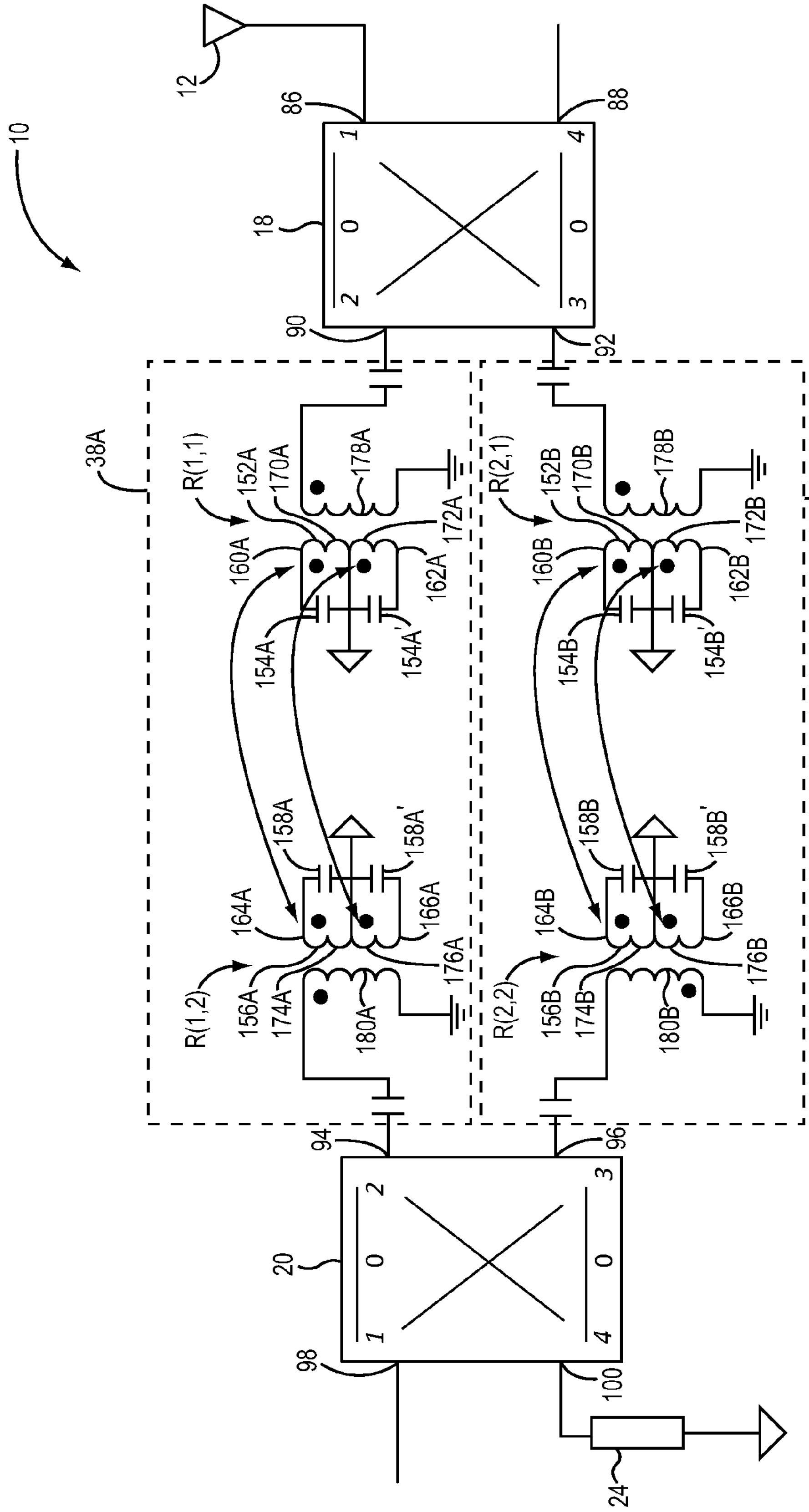
R(1,1)
R(1,2)
R(2,1)
R(2,2)

FIG. 9

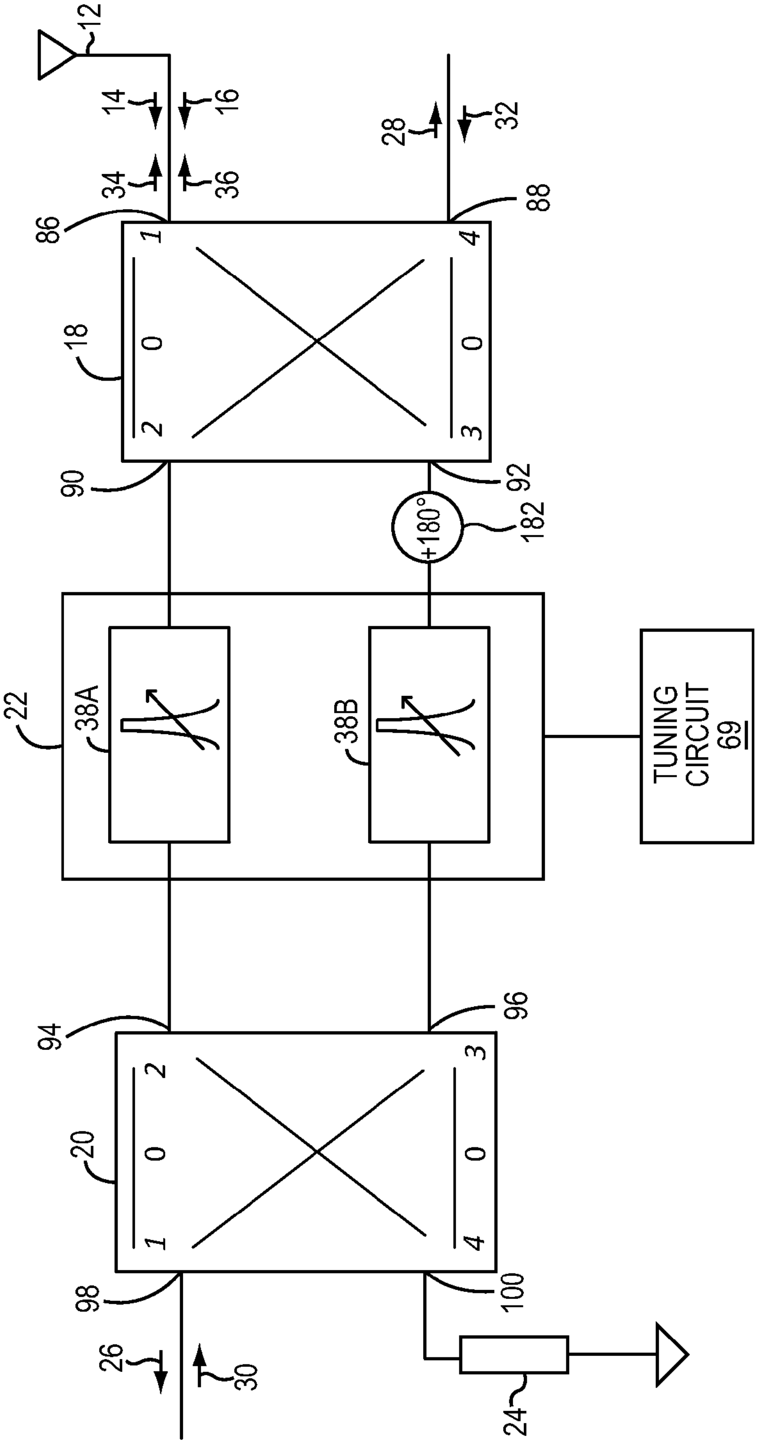
+180°
TUNING CIRCUIT 69

FIG. 10

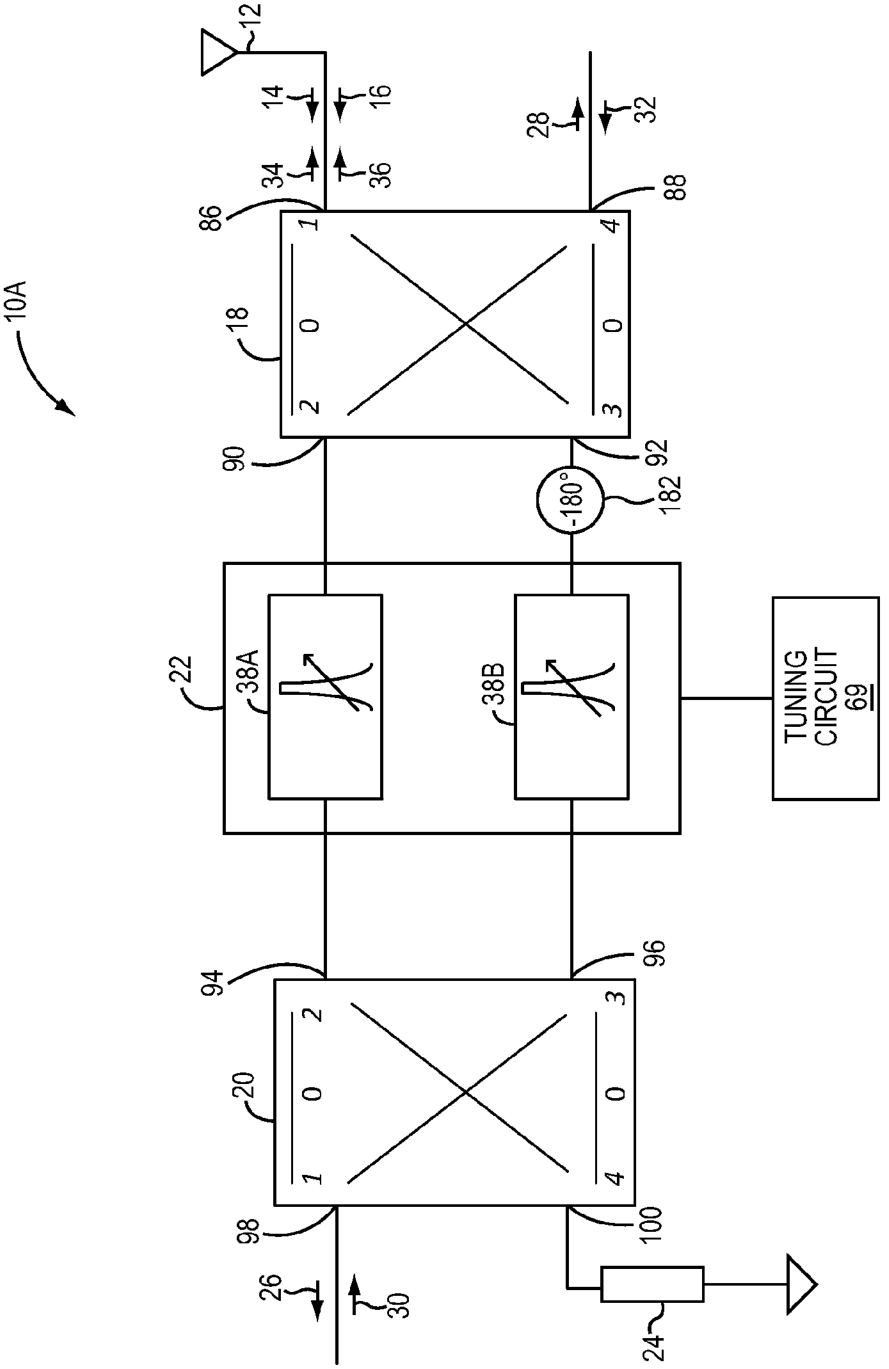
-180°
TUNING
CIRCUIT

FIG. 11

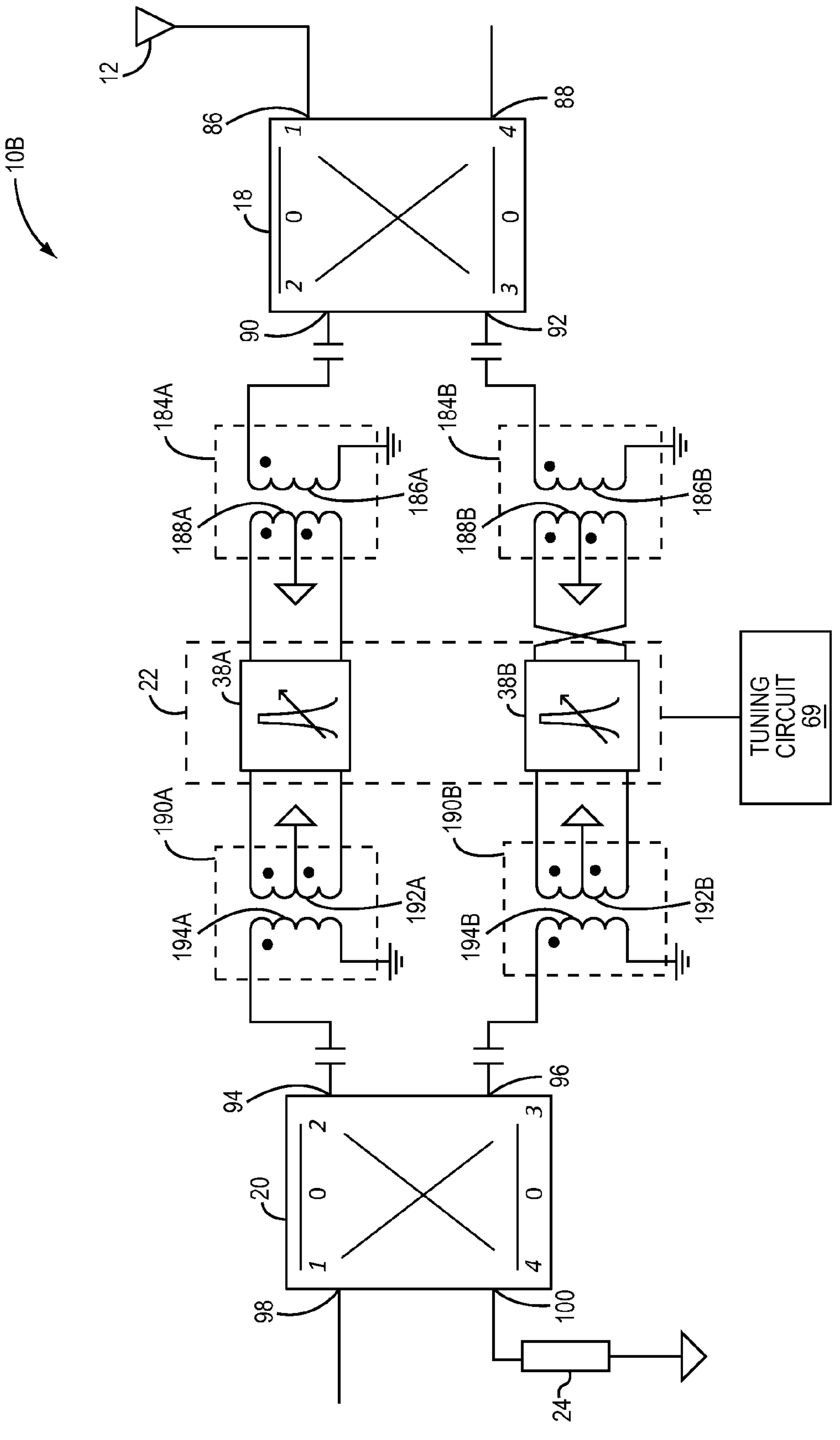
TUNING
CIRCUIT

FIG. 12

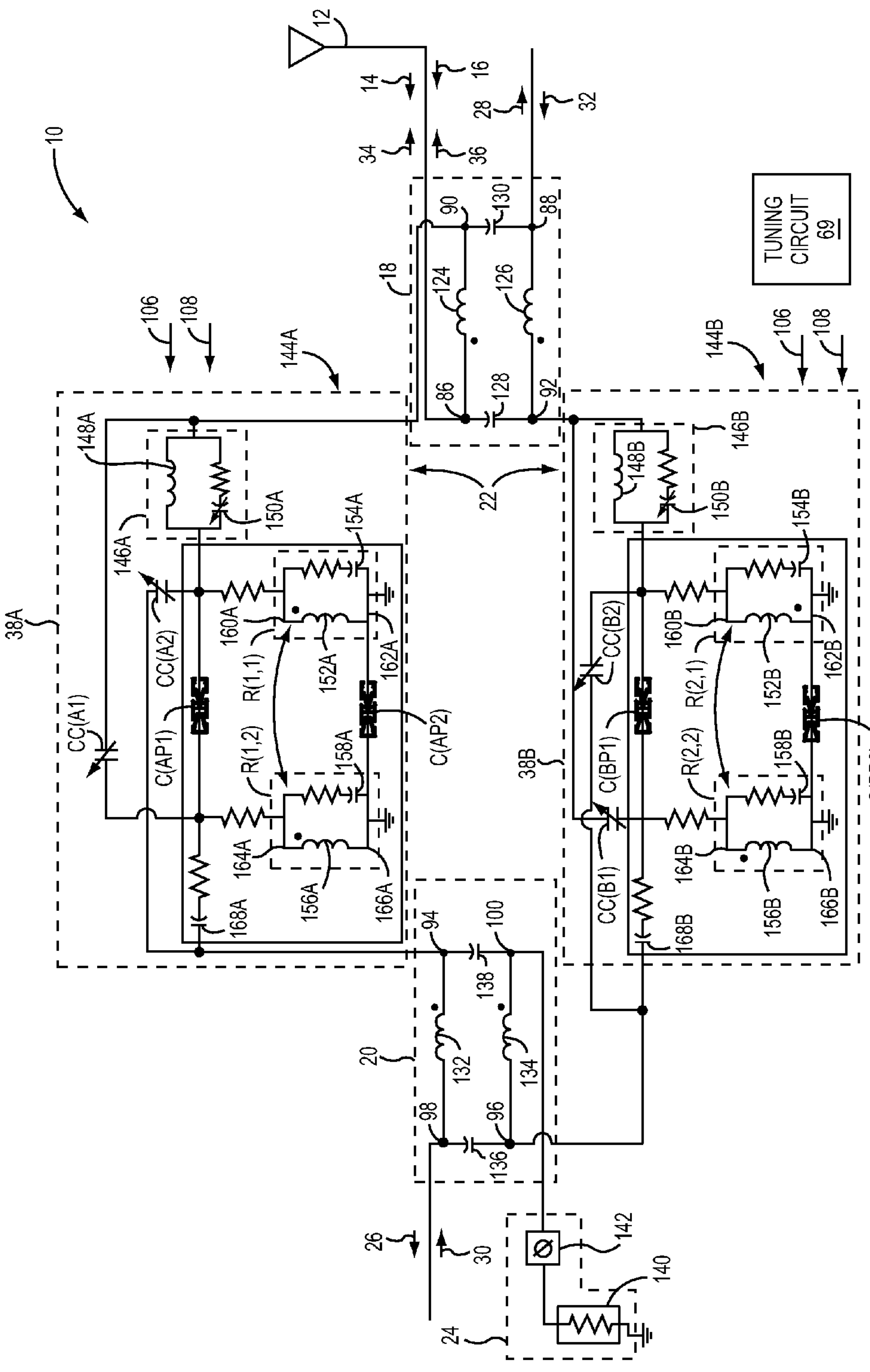
TUNING CIRCUIT 69
CC(A2)
CC(A1)
C(AP1)
R(1,1)
R(1,2)
C(AP2)
CC(B2)
R(2,1)
R(2,2)
C(BP1)
C(BP2)
CC(B1)

FIG. 13

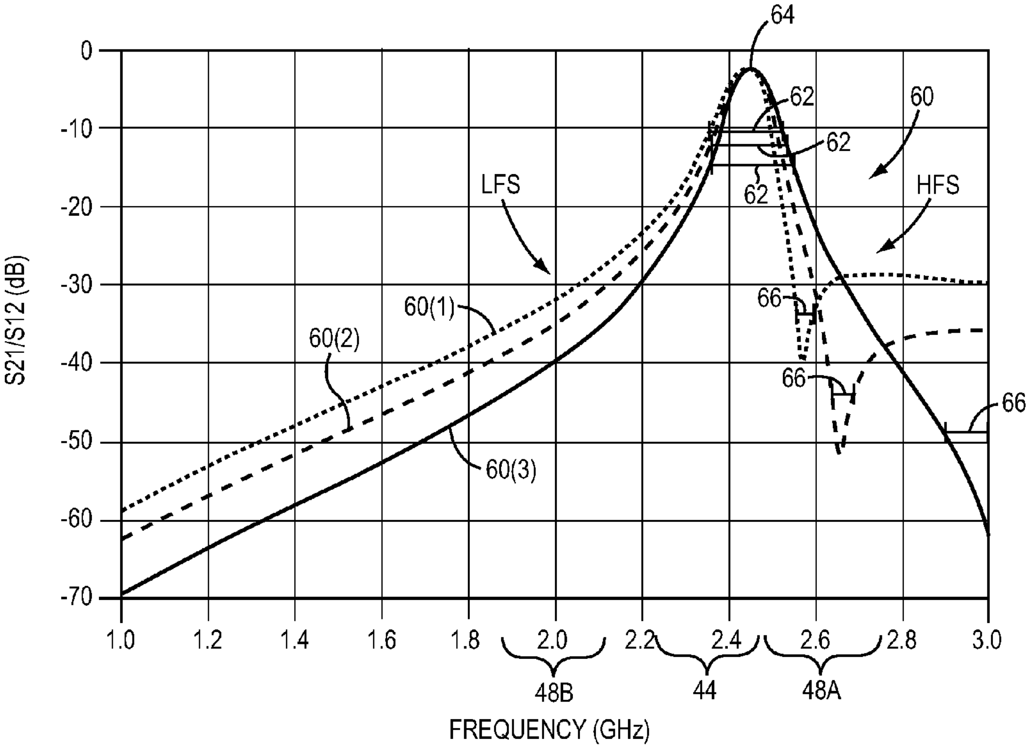
64
62
62
62
60
LFS
HFS
60(1)
60(2)
60(3)
66
66
66
S21/S12 (dB)
0
-10
-20
-30
-40
-50
-60
-70
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0
48B
44
48A
FREQUENCY (GHz)

FIG. 14

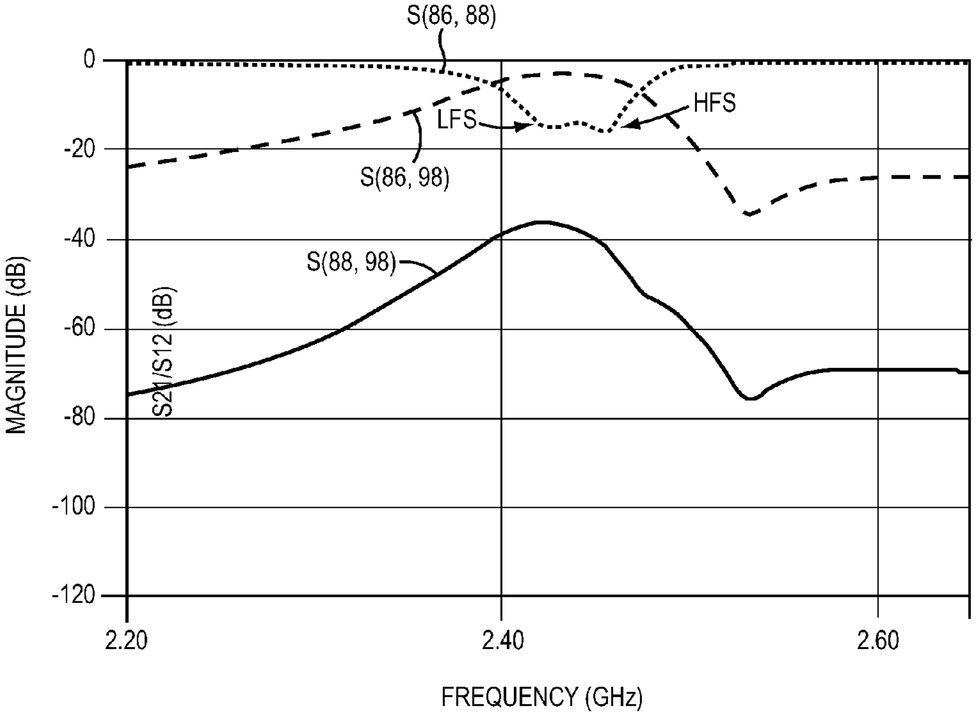
S(86, 88)
0
-20
-40
-60
-80
-100
-120
LFS
HFS
S(86, 98)
S(88, 98)
MAGNITUDE (dB)
S21/S12 (dB)
2.20
2.40
2.60
FREQUENCY (GHz)

FIG. 15

RF DIPLEXER

RELATED APPLICATIONS

[0001] This application claims the benefit of provisional patent application Ser. No. 61/906,622, filed Nov. 20, 2013, the disclosure of which is hereby incorporated herein by reference in its entirety.

[0002] This application is a continuation-in-part of U.S. patent application Ser. No. 14/523,065, filed Oct. 24, 2014; which claims the benefit of provisional patent application 61/895,386, filed Oct. 24, 2013; the disclosures of which are hereby incorporated herein by reference in their entirety.

FIELD OF THE DISCLOSURE

[0003] This disclosure relates generally to radio frequency (RF) diplexers and diplexing methods for RF front-end modules.

BACKGROUND

[0004] A radio frequency (RF) diplexer is a device that facilitates communication of two different RF signals within two different RF communication bands by a common antenna. For example, with respect to RF receive diplexing, an RF diplexer is configured to receive two different RF receive signals on a common antenna and then route each of the respective RF receive signals to their appropriate ports. With respect to RF transmission diplexing, an RF diplexer is configured to transmit two different RF transmission signals from different ports and route both to the common antenna. In order to facilitate simultaneous diplexing over the antenna, the RF diplexer needs to be designed for operation at both RF communication bands while providing adequate isolation between the different RF communication bands and the different RF signals. Many previously known RF diplexer topologies struggle to provide this isolation, particularly at the high RF communication bands required by modern RF front-ends. Furthermore, while there are tunable RF diplexers that provide diplexing with respect to various combinations of different RF communication bands, these tunable RF diplexers struggle to provide adequate broadband isolation across large portions of the RF frequency spectrum.

[0005] Accordingly, RF diplexers are needed that not only provide high isolation but also maintain this high isolation along greater portions of the RF frequency spectrum.

SUMMARY

[0006] Embodiments of a tunable radio frequency (RF) diplexer and methods of diplexing transmission and receive signals are disclosed. In one embodiment, the RF diplexer includes a first hybrid coupler, a second hybrid coupler, an RF filter circuit, and a phase inversion component. Both the RF filter circuit and the phase inversion component are connected between the first hybrid coupler and the second hybrid coupler. In some embodiments, the phase inversion component is provided by the RF filter circuit, while in other embodiments, the phase inversion component is provided separately. In one embodiment, the RF filter circuit is configured to provide a passband and a notch. The RF filter circuit is tunable to provide the notch on both a high-frequency side of the passband and a low frequency side of the passband. The benefit of allowing the notch to be placed on both the high-frequency side and the low-frequency side of the passband is that the tunable RF diplexer provides lower insertion losses and higher isolation regardless of whether the one of the diplexed frequency bands is provided at higher frequencies or lower frequencies than the other diplexed frequency band.

[0007] Those skilled in the art will appreciate the scope of the present disclosure and realize additional aspects thereof after reading the following detailed description of the preferred embodiments in association with the accompanying drawing figures.

BRIEF DESCRIPTION OF THE DRAWING FIGURES

[0008] The accompanying drawing figures incorporated in and forming a part of this specification illustrate several aspects of the disclosure, and together with the description serve to explain the principles of the disclosure.

[0009] FIG. **1** illustrates one embodiment of a tunable radio frequency (RF) diplexer, wherein a phase inversion component is provided by an RF filter between a first hybrid coupler and a second hybrid coupler.

[0010] FIG. **2**A illustrates exemplary procedures that may be implemented by the tunable RF diplexer shown in FIG. **1** so that the tunable RF diplexer can provide RF receive diplexing to two different RF receive signals.

[0011] FIG. **2**B illustrates exemplary procedures that may be implemented by the tunable RF diplexer shown in FIG. **1** so that the tunable RF diplexer can provide RF transmission diplexing to two different RF transmission signals.

[0012] FIG. **2**C illustrates two different RF input signals in two different RF communication bands.

[0013] FIG. **2**D illustrates a frequency response of the RF filter circuit shown in FIG. **1**, where the frequency response defines a passband and a stopband.

[0014] FIG. **2**E illustrates the frequency response of the RF filter circuit shown in FIG. **1**, where the passband has been shifted into one of RF communication bands shown in FIG. **2**C, and the stopband has been shifted into the other RF communication band shown in FIG. **2**C.

[0015] FIG. **2**F is another representation of the frequency response shown in FIG. **2**E.

[0016] FIG. **2**G and FIG. **2**H illustrate quadrature hybrid signals generated by splitting the two different RF input signal shown in FIG. **2**C.

[0017] FIG. **3**A illustrates a receive signal flow for one of the RF receive signals described in FIG. **2**A.

[0018] FIG. **3**B illustrates a receive signal flow for the other RF receive signal described in FIG. **2**A.

[0019] FIG. **3**C illustrates a transmission signal flow for one of the RF transmission signals described in FIG. **2**B.

[0020] FIG. **3**D illustrates a transmission signal flow for the other RF transmission signal described in FIG. **2**B

[0021] FIG. **4** illustrates an example of the tunable RF diplexer shown in FIG. **1**, wherein the phase inversion component is provided by weakly coupled resonators within an RF filter path.

[0022] FIG. **5** is a frequency domain representation of S parameters for the tunable RF diplexer shown in FIG. **4**.

[0023] FIG. **6** illustrates another example of the tunable RF diplexer shown in FIG. **4** wherein the phase inversion component is provided by a single-ended transformer in the RF filter circuit.

[0024] FIG. **7** illustrates a frequency domain representation of S parameters for the tunable RF diplexer shown in FIG. **6**.

[0025] FIG. **8** illustrates yet another example of the tunable RF diplexer shown in FIG. **4** wherein the phase inversion component is provided by a differential transformer in the RF filter circuit.

[0026] FIG. **9** illustrates yet another example of the tunable RF diplexer shown in FIG. **4** wherein the phase inversion component is provided by a single-ended to differential transformer in the RF filter circuit.

[0027] FIG. **10** illustrates another embodiment of a tunable RF diplexer, wherein the phase inversion component is provided by a phase shifter configured to provide a positive differential phase shift.

[0028] FIG. **11** illustrates yet another embodiment of a tunable RF diplexer, wherein the phase inversion component is provided by a phase shifter configured to provide a negative differential phase shift.

[0029] FIG. **12** illustrates still another embodiment of the tunable RF diplexer where the phase inversion component is provided by a single-ended to differential transformer connected to the RF filter circuit.

[0030] FIG. **13** illustrates yet another exemplary embodiment of the tunable RF diplexer that is similar to the embodiment in FIG. **4** and where the RF filter circuit is tunable to provide a notch both at a high-frequency side and a low-frequency side of a passband.

[0031] FIG. **14** illustrates a frequency response of the RF filter circuit shown in FIG. **13**.

[0032] FIG. **15** illustrates a frequency domain representation of S parameters for the tunable RF diplexer shown in FIG. **13**.

DETAILED DESCRIPTION

[0033] The embodiments set forth below represent the necessary information to enable those skilled in the art to practice the embodiments and illustrate the best mode of practicing the embodiments. Upon reading the following description in light of the accompanying drawing figures, those skilled in the art will understand the concepts of the disclosure and will recognize applications of these concepts not particularly addressed herein. It should be understood that these concepts and applications fall within the scope of the disclosure and the accompanying claims.

[0034] Note that relational terminology such as "substantially," "approximately," and/or the like, should be interpreted objectively in accordance with the communication device and technological environment in which the radio frequency (RF) diplexer is employed and, in addition, the performance parameters relevant to the operation of the RF diplexer for at least one the particular application of the RF diplexer within the communication device [or at least one prospective communication device] and the technological environment [or at least one prospective technological environment].

[0035] This disclosure describes different embodiments of a tunable RF diplexer and methods of operating the same. Embodiments of the tunable RF diplexer can be used to provide receive diplexing and/or transmission diplexing. To provide diplexing, the tunable RF diplexer includes a first hybrid coupler, a second hybrid coupler, an RF filter circuit, and a phase inversion component. In some embodiments, the phase inversion component may be provided by the first hybrid coupler, the second hybrid coupler, or the RF filter circuit. In other embodiments, the phase inversion component is provided as a separate component. As explained in further detail below, the first hybrid coupler, the second hybrid coupler, or the RF filter circuit is provided to split and route RF signals to and from the appropriate ports in order to provide diplexing. The phase inversion component, however, is configured to provide a differential phase shift of approximately 180 degrees in one of the paths of the tunable RF diplexer. The benefit of introducing the differential phase shift is that it provides increased broadband isolation between the different frequency bands being diplexed by the tunable RF diplexer.

[0036] FIG. **1** illustrates an embodiment of a tunable radio frequency (RF) diplexer **10**. An antenna **12** is operably associated with the tunable RF diplexer **10** and is capable of emitting or radiating RF signals in different RF frequency bands. In order to prevent out-of-band noise and spurious emissions from distorting RF signals in the different RF frequency band, the tunable RF diplexer **10** provides isolation between RF signals as well as out-of-band filtering. Accordingly, the tunable RF diplexer **10** allows for an RF receive input signal **14** and an RF receive input signal **16** to be received simultaneously by the antenna **12**. The RF receive input signal **14** and the RF receive input signal **16** may be provided in different frequency bands.

[0037] The tunable RF diplexer **10** is operable to separate (i.e., split) the RF receive input signal **14** and the RF receive input signal **16** and route them appropriately. In this embodiment, the tunable RF diplexer **10** includes a first hybrid coupler **18**, a second hybrid coupler **20**, an RF filter circuit **22**, and an impedance termination **24**. The first hybrid coupler **18** receives the RF receive input signal **14** and the RF receive input signal **16** from the antenna **12**. An RF receive output signal **26** is provided by the second hybrid coupler **20** and output to downstream circuitry, such as baseband circuitry. The RF receive output signal **26** is essentially the RF receive input signal **14** after the tunable RF diplexer **10** has filtered and routed the RF receive input signal **14**. An RF receive output signal **26** is provided by the second hybrid coupler **20** and output to downstream circuitry, such as baseband circuitry. The RF receive output signal **26** is essentially the RF receive input signal **14** after the tunable RF diplexer **10** has filtered and routed the RF receive input signal **14**. An RF receive output signal **28** is also provided by the second hybrid coupler **20** and output to downstream circuitry, such as baseband circuitry. The RF receive output signal **28** is essentially the RF receive input signal **16** after the tunable RF diplexer **10** has filtered and routed the RF receive input signal **16**.

[0038] The tunable RF diplexer **10** also allows for an RF transmission input signal **30** and an RF transmission input signal **32** to be combined for transmission by the antenna **12** simultaneously. The second hybrid coupler **20** is configured to receive the RF transmission input signal **30**. In response, the first hybrid coupler **18** outputs an RF transmission output signal **34** towards the antenna **12**. The RF transmission output signal **34** is simply the RF transmission input signal **30** after filtering and routing by the tunable RF diplexer **10**. Additionally, the first hybrid coupler **18** is configured to receive the RF transmission input signal **32**. In response, the first hybrid coupler **18** outputs an RF transmission output signal **36** towards the antenna **12**. The RF transmission output signal **36** is simply the RF transmission input signal **32** after filtering and routing by the tunable RF diplexer **10**. The RF transmission input signal **30** and the RF transmission input signal **32** may each be received from upstream circuitry, such as a power amplifier.

[0039] As explained in further detail below, the first hybrid coupler **18**, the RF filter circuit **22**, and the second hybrid coupler **20** provide the appropriate isolation between the RF receive input signals **14**, **16** and the RF frequency bands. The first hybrid coupler **18**, the RF filter circuit **22**, and the second hybrid coupler **20** provide the appropriate isolation between the RF transmission input signals **30**, **32** and the RF frequency bands. The RF filter circuit **22** includes a first RF filter **38**A across its top ports and an second RF filter **38**B across its bottom ports. Thus, the RF filter circuit **22** is a four-port network. However, the first RF filter **38**A and the second RF filter **38**B are independent of one another and each of the RF filters **38**A, **38**B operates as a two-port network. Also, the first RF filter **38**A and the second RF filter **38**B may be identical, and thus may each have the same individual frequency response. In this embodiment, the first RF filter **38**A is a RF bandpass filter that defines a transfer response with a passband. The passband is symmetrical in that it is the same in both directions of propagation across the first RF filter **38**A to and from the first hybrid coupler **18** and the second hybrid coupler **20**. The second RF filter **38**B is another RF bandpass filter that defines a transfer response with a passband. The passband is also symmetrical in that it is the same in both directions of propagation across the second RF filter **38**B to and from the first hybrid coupler **18** and the second hybrid coupler **20**. In this case, the passband of the first RF filter **38**A and the passband of the second RF filter **38**B are identical. As explained in further detail below, the second RF filter **38**B includes the phase inversion component.

[0040] Referring now to FIGS. **2**A and **2**B, FIG. **2**A illustrates exemplary procedures that may be implemented to provide RF diplexing with respect to receiving the RF receive input signal **14** and the RF receive input signal **16** at the antenna **12**. FIG. **2**B illustrates exemplary procedures that may be implemented to provide RF diplexing with respect to transmitting the RF transmission output signal **34** and the RF transmission output signal **36** from the antenna. As explained in further detail below, the procedures described in FIG. **2** are implemented by the tunable RF diplexer **10** shown in FIG. **1**. Different embodiments of these exemplary procedures may be implemented depending on a particular component structure of a tunable RF diplexer **10**. Furthermore, the order in which the procedures are presented is not intended to imply a required sequence for the procedures. Rather, the procedures may be implemented in a different sequence and/or some or all of the procedures may be implemented simultaneously.

[0041] As shown in FIG. **2**A, the tunable RF diplexer **10** (shown in FIG. **1**) receives the RF receive input signal **14** (procedure **1000**). The RF receive input signal **14** was initially intercepted by the antenna **12** (shown in FIG. **1**). Additionally, the tunable RF diplexer **10** receives the RF receive input signal **16** with the first hybrid coupler **18** (procedure **1002**). The RF receive input signal **16** was also initially intercepted by the antenna **12**. In one embodiment, the tunable RF diplexer **10** is configured to provide diplexing when the RF receive input signal **14** is provided within an industrial, scientific, and medical (ISM) frequency band having a frequency range between 2.4 GHz and 2.5 GHz and a center frequency of around 2.450 GHz. Thus, the RF receive output signal **26** is also found in the same ISM frequency band. The RF receive input signal **16** can be provided within any one of multiple Medium Band (MB) and High Band (HB) frequency bands. For example, the RF receive input signal **16** can be provided in B1-B7 frequency bands which extend between 1.810 GHz to 2.690 GHz. The RF receive output signal **28** is thus provided in the same MB/HB frequency band.

[0042] With respect to FIG. **2**B, the tunable RF diplexer **10** receives the RF transmission input signal **30** with the second hybrid coupler **20** (procedure **2000**). The RF transmission input signal **30** was initially received from upstream circuitry (e.g., a power amplifier) by the second hybrid coupler **20**. Additionally, the tunable RF diplexer **10** receives the RF transmission input signal **32** (procedure **2002**). The RF transmission input signal **32** was initially received from upstream circuitry (e.g., a power amplifier) by the second hybrid coupler **20**. In one embodiment, the tunable RF diplexer **10** is configured to provide diplexing when the RF transmission input signal **30** is provided within the same ISM frequency band as the RF receive input signal **14**. Thus, the RF transmission output signal **34** is also provided in the same ISM frequency band. The RF transmission input signal **32** can be provided within any one of multiple MB and HB frequency bands. For example, the RF transmission input signal **32** can be provided in B1-B7 frequency bands which extend between 1.810 GHz to 2.690 GHz. The RF transmission output signal **36** is thus provided in the same corresponding frequency band as the RF transmission input signal **32**.

[0043] Accordingly, the tunable RF diplexer **10** provides diplexing so that RF signals in different ISM and MB/HB frequency bands can share the antenna **12**. In addition, the tunable RF diplexer **10** is configured to allow for carrier aggregation and combine the RF transmission output signals **34**, **36** in the separate ISM and Cellular MB/HB frequency bands to emit the RF transmission output signals **34**, **36** are simultaneously emitted by the antenna **12**. Again, to operate optimally, the tunable RF diplexer **10** is required to provide isolation across broadband frequency ranges from 1810 MHz to 2690 MHz. The use of tunable of ISM/MB-HB cellular diplexer provides other benefits, for example, when operating into cellular mode only, the ISM bandpass filter is retuned out of the ISM band frequency into either a lower frequency or higher frequency to reduce the IL of the operating cellular frequency as shown in FIG. **23**.

[0044] FIG. **2**C illustrates a graph of one embodiment of an RF input signal **40** and an RF input signal **42** in the frequency domain. With respect to the procedures described in FIG. **2**A, the RF input signal **40** is the RF receive input signal **14**, and the RF input signal **42** is the RF receive input signal **16**. However, with respect to the procedures described in FIG. **2**B, the RF input signal **40** is the RF transmission input signal **30**, and the RF input signal **42** is the RF transmission input signal **32**. The RF input signal **40** operates in an RF communication band **44**, which is one embodiment of the ISM frequency band. In this embodiment, the RF communication band **44** is the set of frequencies between a cutoff frequency C1 and a cutoff frequency C2. Note that the RF communication band **44** may include additional RF subbands, such as the RF subband **46**. The RF subband **46** may be a transmission subband, a receive subband, and/or the like. The RF subband **46** is defined as the set of frequencies between a cutoff frequency f_{B1} and a cutoff frequency f_{B2}. With respect to the RF input signal **42**, the RF input signal **42** operates in an RF communication band **48**, which is one embodiment of the MB/HB frequency bands. In this embodiment, the RF communication band **48** is the set of frequencies between a cutoff frequency C3 and a cutoff frequency C4. Note that the RF communication band **48** may include additional RF subbands, such as the RF subband **50**. The RF subband **50** may be a transmission subband, a receive subband, and/or the like.

The RF subband **50** is defined as the set of frequencies between a cutoff frequency f_{Y1} and a cutoff frequency f_{Y2}.

[0045] In this example, a signal bandwidth **52** of the RF input signal **40** is the set of frequencies that corresponds to the portion of the RF input signal **40** within 3 dB of a maximum magnitude **54**. The RF input signal **40** shown in FIG. **2**C operates at a frequency f_A. This frequency f_A corresponds to the maximum magnitude **54** of the RF input signal **40**. For example, the frequency f_A may be a carrier frequency of the RF input signal **40**. The signal bandwidth **52** reaches a cutoff frequency f_{A1}, since the frequency f_{A1} corresponds to a value of the RF input signal **40** that is 3 dB from the maximum magnitude **54**. The signal bandwidth **52** reaches a cutoff frequency f_{A2}, because the cutoff frequency f_{A2} corresponds to a value of the RF input signal **40** that is 3 dB from the maximum magnitude **54**.

[0046] Also, in this example, a signal bandwidth **56** of the RF input signal **42** is the set of frequencies that corresponds to the portion of the RF input signal **42** within 3 dB of a maximum magnitude **58**. The RF input signal **42** shown in FIG. **2**A operates at a frequency f_Z. This frequency f_Z corresponds to the maximum magnitude **58** of the RF input signal **42**. For example, the frequency f_Z may be a carrier frequency of the RF input signal **42**. The signal bandwidth **56** reaches a cutoff frequency f_{Z1}, since the frequency f_{Z1} corresponds to a value of the RF input signal **42** that is 3 dB from the maximum magnitude **58**. The signal bandwidth **56** reaches a cutoff frequency f_{Z2}, because the cutoff frequency f_{Z2} corresponds to a value of the RF input signal **42** that is 3 dB from the maximum magnitude **58**.

[0047] Portions of the RF input signal **40** and the RF input signal **42** outside of their respective signal bandwidths **52**, **56** may be considered spurious emissions. In other words, the portions of the RF input signal **40** and the RF input signal **42** may be reduced or eliminated without affecting the corresponding information or data in the RF input signal **40** and the RF input signal **42**. Spurious emissions include parasitic emissions, intermodulation, interference, harmonic emissions, and frequency conversion products. The signal bandwidth **52** and the signal bandwidth **56** are defined as 3 dB bandwidths for pragmatic purposes. Generally speaking, at least for the types of signals being shown in FIG. **2**C, the signal bandwidths **52**, **56** are measured by finding 3 dB magnitudes from the maximum magnitudes **54**, **58**, as explained above. However, more accurately, a necessary signal bandwidth is an exact amount of signal bandwidth required to carry the information or data of a signal. Anything outside of this necessary bandwidth would be considered spurious emissions. Thus, the signal bandwidth **52** and the signal bandwidth **56** may or may not include a small portion of the spurious emissions. The necessary signal bandwidths may be slightly smaller or slightly greater than the signal bandwidths **52** and **56**.

[0048] Finally, it should be noted that the RF input signal **40** and the RF input signal **42** may each be narrow-band signals. Accordingly, the RF input signal **40** and the RF input signal **42** may represent time division multiplexing (TDM) signals, frequency division multiplexing (FDM) signals, space division multiplexing (SDM) signals, and/or the like. Accordingly, these narrow-band signals may be said to operate at a particular frequency, which for the RF input signal **40** is the frequency f_A and for the RF input signal **42** is the frequency f_Z. The RF subband **46** is thus an RF channel within the RF communication band **44**, while the RF subband **50** is an RF channel within the RF communication band **48**.

[0049] However, this disclosure is not limited to narrow-band signals and the examples given in FIG. **2**C and throughout this disclosure are not intended to be limited in this manner. Rather, embodiments of the tunable RF diplexer **10** and the method shown in FIGS. **2**A and **2**B may be provided for wide-band signals, and also for both wide-band and narrow-band signals. With wide-band signals, such as orthogonal frequency division multiple access (OFDMA) signals or Code Division Multiple Access (CDMA) signals, information or data is coded and spread across a larger portion of the spectrum. Thus, there would be no signal with a single carrier frequency that has all of the information or data, but rather there may be various carriers carrying different coded portions of the information. As such, the RF communication band **44** and/or the RF communication band **48** may include various RF channels. Similarly, the RF communication band **48** may include various RF channels. With CDMA signals and other wide-band spectrum signals, it is more practical to define the bandwidths by simply using the necessary bandwidth, as is known in the art.

[0050] FIG. **2**D illustrates one embodiment of a frequency response **60** provided by the RF filter circuit **22**. In FIG. **2**D, the frequency response **60** represents S21 and S12 parameters of each of the RF filters **38**A, **38**B (shown in FIG. **1**) individually, as a function of frequency. Thus, the two-port S21 and S12 parameter of the first RF filter **38**A is represented by the frequency response **60** in FIG. **2**D. Since the second RF filter **38**B is identical to the first RF filter **38**A, the two-port S21 and S12 parameter of the second RF filter **38**B is also represented by the frequency response **60** in FIG. **2**D. The two-port S21 and S12 parameter represents the forward and reverse transmission of the RF filters **38**A, **38**B, as a function of frequency. A passband **62** corresponds to S21/S12 values in the frequency response **60** that are within 3 dB of a maxima **64**. A stopband **66** is determined relative to a minima **68**. The maxima **64** and the minima **68** are set by the poles and zeros of the frequency response **60**. The stopband **66** is a set of frequencies that correspond to S21/S12 values within 3 dB of the minima **68**. In this embodiment, the frequency response **60** defines the stopband **66** as a notch. As explained in further detail below, the RF filter circuit **22** is tunable so as to shift the passband **62** and the stopband **66**. Thus, by tuning the RF filter circuit **22**, the frequency response **60** may be transposed, so that the passband **62** and the stopband **66** are provided at the desired frequency bands.

[0051] Referring now to FIGS. **2**A, **2**B, and **2**E, FIG. **2**E illustrates the frequency response **60** of the RF filter circuit **22** once the passband **62** is shifted to include the RF communication band **44**, and once the stopband **66** is shifted to include the RF communication band **48**. The tuning circuit **69** (shown in FIG. **1**) tunes the passband **62** so that the passband **62** includes the RF communication band **44**. The tuning circuit **69** may also tune the stopband **66** so that the stopband **66** includes the RF communication band **48**. In this manner, signals that operate in the RF communication band **44** are passed by the RF filter circuit **22**, while signals that operate in the RF communication band **48** are blocked by the RF filter circuit **22**.

[0052] Referring now to FIGS. **2**A, **2**B, and **2**F, FIG. **2**F also illustrates the frequency response **60** of the RF filter circuit **22** once the passband **62** is shifted to include the RF communication band **44**, and once the stopband **66** is shifted to include the RF communication band **48**. However, in FIG. **2**F, the frequency response **60** represents S11 and S22 parameters of each of the RF filters **38**A, **38**B (shown in FIG. **1**) individually, as a function of frequency. Thus, the two-port S11 and S22 parameter of the first RF filter **38**A is represented by the frequency response **60** in FIG. **2**F. Since the second RF filter **38**B is identical to the first RF filter **38**A, the two-port S11 and S22 parameter of the second RF filter **38**B is also represented by the frequency response **60** in FIG. **2**F. The two-port S11 and S22 parameter represents the forward and reverse return loss of the RF filters **38**A, **38**B, as a function of frequency. Note that in this embodiment, the S11/S22 values of the stopband **66** are at or near zero (0) dB in the RF communication band **48**. By placing the stopband **66** at or near zero (0) dB, reflections in the stopband **66** are maximized, while the insertion losses within the RF communication band **48** are minimized. Filtering thus removes noise outside of the RF communication band **48** while minimizing losses of reflected signals.

[0053] Referring now to FIGS. **2**G and **2**H, the tunable RF diplexer **10** is configured to split the RF input signal **40** into a first RF quadrature hybrid signal (QHS) **70** and a second RF QHS **72**. As such, the first RF QHS **70** is 90 degrees or π/2 radians out of phase with the second RF QHS **72**. Also, the tunable RF diplexer **10** is configured to split the RF input signal **42** into a first RF QHS **74** and a second RF QHS **76**. FIGS. **2**G and **2**H illustrate the first RF QHS **70** (shown in FIG. **2**G) and the second RF QHS **72** (shown in FIG. **2**H) generated by splitting the RF input signal **40**. The first RF QHS **70** and the second RF QHS **72** have substantially identical magnitude characteristics as the RF input signal **40**. However, both the first RF QHS **70** and the second RF QHS **72** have a power spectral density that is at a power ratio of the power spectral density of the RF input signal **40**. In this example, the power ratio is 3 dB, and thus the first RF QHS **70** and the second RF QHS **72** have approximately one half of the power of the RF input signal **40**. The first RF QHS **70** and the second RF QHS **72** are quadrature hybrids, since there is approximately a 90-degree or π/2 radians phase difference between the signals.

[0054] Of course, non-ideal characteristics of the tunable RF diplexer **10**, such as parasitic impedances, may result in the first RF QHS **70** and the second RF QHS **72** being slightly unbalanced with respect to one another, or having slightly less than half the power of the RF input signal **40**. Also, non-ideal characteristics can result in the phase difference between the first RF QHS **70** and the second RF QHS **72** fluctuating somewhat from a 90-degree or π/2 radians phase difference. These types of errors are acceptable so long as the first RF QHS **70** and the second RF QHS **72** can be generated to comply with spectrum requirements for the RF communication standard with respect to the RF input signal **40**. As explained in further detail below, the tunable RF diplexer **10** is configured to provide a differential phase shift to the second RF QHS **72** after the RF input signal **40** has been split. The differential phase shift is approximately 180 degrees or π radians. In this embodiment, the second RF filter **38**B (shown in FIG. **1**) is configured to provide the differential phase shift and includes the phase inversion component that provides the differential phase shift. With regard to the ISM frequency band and MB/HB frequency band embodiment, IM3 cancellation is provided by the architecture of the back-to-back hybrid couplers, **20**.

[0055] With respect to the procedures in FIG. **2**A, the RF input signal **40** is the RF receive input signal **14**, shown in FIG. **1**. Accordingly, the first hybrid coupler **18** is configured to split the RF receive input signal **14** into a first RF quadrature hybrid receive signal (QHRS) (which in the procedures of FIG. **2**A is the first RF QHS **70** shown in FIG. **2**G) and a second RF QHRS (which in the procedures of FIG. **2**A is the second RF QHS **72** shown in FIG. **2**H) (procedure **1004**). As such, the first RF QHRS is 90 degrees or π/2 radians out of phase with the second RF QHRS when the first hybrid coupler **18** splits the RF receive input signal **14**. The RF filter circuit **22** then filters the first RF QHRS and the second RF QHRS. More specifically, the second RF filter **38**B provides the differential phase shift to the second RF QHRS (procedure **1006**), and the RF filter circuit **22** passes the first RF QHRS and the second RF QHRS to the second hybrid coupler **20** (procedure **1008**). As explained in further detail below, the differential phase shift may be provided by other phase inversion components, such as phase shifter, transformers, and the like. The second hybrid coupler **20** provides a quadrature phase shift (approximately 90-degree or π/2 radians phase shift) to the second RF QHRS (procedure **1010**). In this manner, the second hybrid coupler **20** combines the first RF QHRS and the second RF QHRS into the RF receive output signal **26** (procedure **1012**). These procedures are further explained in Figures below.

[0056] With respect to the procedures in FIG. **2**B, the RF input signal **40** is the RF transmission input signal **30**, shown in FIG. **1**. Accordingly, the second hybrid coupler **20** is configured to split the RF transmission input signal **30** into a first RF quadrature hybrid transmission signal (QHTS) (which in the procedures of FIG. **2**B is the first RF QHS **70** shown in FIG. **2**G) and a second RF QHTS (which in the procedures of FIG. **2**B is the second RF QHS **72** shown in FIG. **2**H) (procedure **2004**). As such, the first RF QHTS is 90 degrees or π/2 radians out of phase with the second RF QHTS. The RF filter circuit **22** then filters the first RF QHTS and the second RF QHTS. More specifically, the second RF filter **38**B provides the differential phase shift to the second RF QHTS (procedure **2006**), and the RF filter circuit **22** passes the first RF QHTS and the second RF QHTS to the first hybrid coupler **18** (procedure **2008**). The first hybrid coupler **18** provides a quadrature phase shift (approximately 90-degree or π/2 radian phase shift) to the second RF QHTS (procedure **2010**). In this manner, the first hybrid coupler **18** also combines the first RF QHTS and the second RF QHTS into the RF transmission output signal **34** (procedure **2012**).

[0057] Referring again to FIGS. **2**G and **2**H, the tunable RF diplexer **10** is also configured to split the RF input signal **42** into the first RF QHS **74** and the second RF QHS **76**, as mentioned above. As such, the first RF QHS **74** is 90 degrees or π/2 radians out of phase with the second RF QHS **76**. FIGS. **2**G and **2**H illustrate the first RF QHS **74** (shown in FIG. **2**G) and the second RF QHS **76** (shown in FIG. **2**H) generated by splitting the RF input signal **42**. The first RF QHS **74** and the second RF QHS **76** have substantially identical magnitude characteristics as the RF input signal **42**. However, both the first RF QHS **74** and the second RF QHS **76** have a power spectral density that is at a power ratio of the power spectral density of the RF input signal **42**. In this example, the power ratio is 3 dB, and thus the first RF QHS **74** and the second RF QHS **76** have approximately one half of the power of the RF input signal **42**. The first RF QHS **74** and the second RF QHS

76 are quadrature hybrids, since there is approximately a 90-degree or $\pi/2$ radians phase difference between the signals.

[0058] Of course, non-ideal characteristics of the tunable RF diplexer **10**, such as parasitic impedances, may result in the first RF QHS **74** and the second RF QHS **76** being slightly unbalanced with respect to one another, or having slightly less than half the power of the RF input signal **42**. Also, non-ideal characteristics can result in the phase difference between the first RF QHS **74** and the second RF QHS **76** fluctuating somewhat from a 90-degree or $\pi/2$ radians phase difference. These types of errors are acceptable so long as the first RF QHS **74** and the second RF QHS **76** can be generated to comply with spectrum requirements for the RF communication standard with respect to the RF input signal **42**. As explained in further detail below, the tunable RF diplexer **10** is configured to provide a differential phase shift to either the first RF QHS **74** or the second RF QHS **76** after the RF input signal **42** has been split. The differential phase shift is approximately 180 degrees or π radians. As mentioned above, in this embodiment, the second RF filter **38**B (shown in FIG. **1**) is configured to provide the differential phase shift.

[0059] With respect to the procedures in FIG. **2**A, the RF input signal **42** is the RF receive input signal **16**, shown in FIG. **1**. Accordingly, the first hybrid coupler **18** is configured to split the RF receive input signal **16** into a first RF QHRS (which in the procedures of FIG. **2**A is the first RF QHS **74** shown in FIG. **2**G) and a second RF QHRS (which in the procedures of FIG. **2**A is the second RF QHS **76** shown in FIG. **2**H) (procedure **1014**). As such, the first RF QHRS is 90 degrees or $\pi/2$ radians out of phase with the second RF QHRS. The RF filter circuit **22** then filters the first RF QHRS and the second RF QHRS. More specifically, the RF filter circuit **22** blocks the first RF QHRS and the second RF QHRS so that the first RF QHRS and the second RF QHRS is reflected back toward the first hybrid coupler **18** (procedure **1016**). The first hybrid coupler **18** then provides a quadrature phase shift (approximately 90-degree or $\pi/2$ radians phase shift) to the first RF QHRS (procedure **1018**). In this manner, the first hybrid coupler **18** combines the first RF QHRS and the second RF QHRS generated by splitting the RF input signal **42** into the RF receive output signal **28** (procedure **1020**). These procedures are further explained in Figures below. Furthermore, note that the procedures **1000**, **1004-1012** may all occur simultaneously with procedures **1002**, **1014-1020**.

[0060] With respect to the procedures in FIG. **2**B, the RF input signal **42** is the RF transmission input signal **32**, shown in FIG. **1**. Accordingly, the first hybrid coupler **18** is configured to split the RF transmission input signal **32** into a first RF QHTS (which in the procedures of FIG. **2**B is the first RF QHS **74** shown in FIG. **2**G) and a second RF QHTS (which in the procedures of FIG. **2**B is the second RF QHS **76** shown in FIG. **2**H) (procedure **2014**). As such, the first RF QHTS is 90 degrees or $\pi/2$ radians out of phase with the second RF QHTS. The RF filter circuit **22** then filters the first RF QHTS and the second RF QHTS generated by splitting the RF transmission input signal **32**. More specifically, the RF filter circuit **22** blocks the first RF QHTS and the second RF QHTS generated by splitting the RF transmission input signal **32** so that the first RF QHTS and the second RF QHTS generated by splitting the RF transmission input signal **32** are reflected back to the first hybrid coupler **18** (procedure **2016**). The first hybrid coupler **18** then provides a quadrature phase shift (approximately 90-degree or $\pi/2$ radians phase shift) to the first RF QHTS (procedure **2018**). In this manner, the first hybrid coupler **18** combines the first RF QHTS and the second RF QHTS generated by splitting the RF transmission input signal **32** into the RF transmission output signal **36** (procedure **2020**). These procedures are further explained in Figures below. Furthermore, note that the procedures **2000**, **2004-2012** may all occur simultaneously with procedures **2002**, **2014-2020**.

[0061] Referring again to FIGS. **2**G and **2**H, since the RF input signal **40** is split into the first RF QHS **70** and the second RF QHS **72**, both the first RF QHS **70** and the second RF QHS **72** operate in the RF communication band **44**. When the hybrid couplers **18**, **20** output the first RF QHS **70** and the second RF QHS **72**, the RF filter circuit **22** filters the first RF QHS **70** and the second RF QHS **72** to pass the first RF QHS **70** and the second RF QHS **72** within the passband **62**. As shown in FIGS. **2**G and **2**H, the frequency response **60** of the RF filter circuit **22** has been transposed so that the passband **62** and the stopband **66** are shifted into the RF communication band **44**. More particularly, the passband **62** is shifted so that the RF communication band **44** is in the passband **62**.

[0062] The stopband **66** has been shifted so that the RF communication band **48** is in the stopband **66**. Since the first RF QHS **70** and the second RF QHS **72** are within the passband **62**, the RF filter circuit **22** is configured to pass the first RF QHS **70** and the second RF QHS **72**. However, the RF filter circuit **22** filters the first RF QHS **74** and the second RF QHS **76** to reflect the first RF QHS **74** and the second RF QHS **76** within the stopband **66**. The first RF QHS **74** and the second RF QHS **76** operate within the RF communication band **48**. Since the stopband **66** is in the RF communication band **48**, the RF filter circuit **22** has been tuned to block the first RF QHS **74** and the second RF QHS **76**.

[0063] As shown in FIGS. **2**G and **2**H, the first RF QHS **70** has a signal bandwidth **78**, and the second RF QHS **72** has a signal bandwidth **80**. Although the first RF QHS **70** and the second RF QHS **72** have approximately half the spectral power density (excluding losses) of the RF input signal **40** shown in FIG. **2**C, the signal bandwidth **78** of the first RF QHS **70** and the signal bandwidth **80** of the second RF QHS **72** have approximately the same size and are located at approximately the same frequencies. In FIGS. **2**E and **2**F, the signal bandwidth **78** of the first RF QHS **70** and the signal bandwidth **80** of the second RF QHS **72** are between the frequencies f_{A1} and f_{A2}. However, non-ideal circuit behavior may result in either slight misalignments and/or slight size differences in the signal bandwidth **78** and the signal bandwidth **80**. The amount of error that is permissible may depend on the particular application and spectrum requirements. The tuning circuit **69** is configured to tune the frequency response **60** of the RF filter circuit **22** so that the signal bandwidth **78** of the first RF QHS **70** and the signal bandwidth **80** of the second RF QHS **72** are each within the passband **62**. Thus, the RF filter circuit **22** filters the first RF QHS **70** and the second RF QHS **72** to pass the first RF QHS **70** and the second RF QHS **72** within the passband **62** to the second hybrid coupler **20**.

[0064] As shown in FIGS. **2**G and **2**H, the first RF QHS **74** has a signal bandwidth **82**, while the second RF QHS **76** has a signal bandwidth **84**. Although the first RF QHS **74** and the second RF QHS **76** have approximately half the power spectral density of the RF input signal **42** (excluding losses) shown in FIG. **2**C. The signal bandwidth **82** of the first RF QHS **74** and the signal bandwidth **84** of the second RF QHS **76** are configured to be approximately the same as the signal bandwidth **82** of the RF input signal **42**. In the embodiments illustrated in FIGS. **2**G and **2**H, the signal bandwidth **82** and the signal bandwidth **84** are both between the frequencies f_{Z1} and f_{Z2}. However, non-ideal circuit behavior may result in slight misalignments or size differences in the signal bandwidth **82** and the signal bandwidth **84**. The amount of error that is permissible may depend on the particular application and spectrum requirements. The tuning circuit **69** is configured to tune the frequency response **60** of the RF filter circuit **22** so that the signal bandwidth **82** of the first RF QHS **74** and the signal bandwidth **84** of the second RF QHS **76** are each within the stopband **66**. In this manner, the RF filter circuit **22** blocks the first RF QHS **74** and the second RF QHS **76** such that the first RF QHS **74** and the second RF QHS **76** are reflected back to the second hybrid coupler **20**.

[0065] FIG. **3**A illustrates the tunable RF diplexer **10** along with a receive signal flow of the tunable RF diplexer **10** with respect to the procedures **1000**, **1004-1012** for the RF receive input signal **14** and the RF receive output signal **26**. As shown in FIG. **3**A, the first hybrid coupler **18** has a first port **86**, a second port **88**, a third port **90**, and a fourth port **92**, while the second hybrid coupler **20** has a fifth port **94**, a sixth port **96**, a seventh port **98**, and an eighth port **100**.

[0066] The antenna **12** intercepts the RF receive input signal **14** as electromagnetic waves in free space. These electromagnetic waves result in excitations within the antenna **12**, thereby converting the electromagnetic waves into the RF receive input signal **14**. The first hybrid coupler **18** is operable to receive the RF receive input signal **14**. In this particular embodiment, the first hybrid coupler **18** is coupled to receive the RF receive input signal **14** at the first port **86** from the antenna **12**. The first hybrid coupler **18** is operable to split the RF receive input signal **14** into a first RF QHRS **102** and the second RF QHRS **104**. In this manner, the first RF QHRS **102** and the second RF QHRS **104** have approximately the same power ratio with respect to the RF receive input signal **14**, but have a quadrature phase difference of approximately 90 degrees or π/2 radians with respect to one another.

[0067] With regard to the first hybrid coupler **18** shown in FIG. **3**A, the first port **86** is phase-aligned with the third port **90**, while the first port **86** has a quadrature phase shift with respect to the fourth port **92**. Thus, the first RF QHRS **102** at the third port **90** is approximately phase-aligned with the RF receive input signal **14** at the first port **86**, but there is a quadrature phase difference between the RF receive input signal **14** at the first port **86** and the second RF QHRS **104** at the fourth port **92**.

[0068] Note that in alternative embodiments, this may or may not be the case. For example, there may be a phase shift between the first port **86** and the third port **90** (such as +45 degrees or +π/4 radians). The phase shift between the first port **86** and the fourth port **92** may then be equal to this phase shift plus or minus 90 degrees (such as +135 degrees or +3π/4 radians, −45 degrees or −π/4 radians). Accordingly, so long as the phase difference between the first RF QHRS **102** and the second RF QHRS **104** is about 90 degrees or π/2 radians, phase alignment between the third port **90** and the first port **86**, and between the fourth port **92** and the first port **86**, can vary. The first RF QHRS **102** is output at the third port **90** toward the first RF filter **38**A of the RF filter circuit **22**. The first RF filter **38**A is connected between the third port **90** and the fifth port **94**. Additionally, the second RF QHRS **104** is output at the fourth port **92** toward the second RF filter **38**B of the RF filter circuit **22**. The second RF filter **38**B is connected between the fourth port **92** and the sixth port **96**.

[0069] The tunable RF diplexer **10** includes a phase inversion component that is connected between the first hybrid coupler **18** and the second hybrid coupler **20**. The phase inversion component is configured to provide a differential phase shift to the second RF QHRS **104**. In this embodiment, the phase inversion component is the second RF filter **38**B in the RF filter circuit **22**. As such, the second RF filter **38**B shown in FIG. **3**A is configured to provide the differential phase shift (i.e. approximately 180 degree/π radian phase shift) to the second RF QHRS **104**. However, as explained in further detail below, the phase inversion component may be provided by any phase inversion component capable of providing a differential phase shift, such as a phase shifter, a transformer, and/or the like. The benefit of introducing the differential phase shift is that it provides increased isolation between diplexed RF communication bands over a greater frequency range. For example, if the RF receive input signal **14** and the RF receive output signal **26** are provided in the ISM frequency band, and the RF receive input signal **16** (shown in FIG. **1**) and the RF receive output signal **28** (shown in FIG. **1**) are provided an MB/HB frequency band, the cancellation of the RF receive input signal **16** (shown in FIG. **1**) and the RF receive output signal **28** (shown in FIG. **1**) is provided through the differential phase shift, rather than simply on a precise power split. This differential phase shift is much easier to maintain over a broad frequency range.

[0070] Referring again to FIG. **3**A, the tuning circuit **69** is configured to tune the frequency response of the RF filter circuit **22** so that the passband **62** of the first RF filter **38**A and the passband **62** of the second RF filter **38**B includes the RF communication band of the RF receive input signal **14** and the RF receive output signal **26**. In this manner, the RF filter circuit **22** is operable to pass the first RF QHRS **102** and the second RF QHRS **104** to the second hybrid coupler **20**. The manner of tuning the frequency response may depend on the topology of the RF filter circuit **22**. For example, the first RF filter **38**A and the second RF filter **38**B shown in FIG. **3**A may both be variable passive filters. Accordingly, one or more reactive impedance components (inductive, capacitive, or both) in each of the first and second RF filters **38**A, **38**B may have a variable reactive impedance level. By varying these variable reactive impedance levels, the poles and zeros of the individual frequency responses provided by each of the first and second RF filters **38**A, **38**B are adjusted. This thereby shifts the passband **62** in the frequency domain. As explained in further detail below, the first and second RF filters **38**A, **38**B may include weakly coupled resonators to provide filtering and the differential phase shift.

[0071] In this embodiment, the tuning circuit **69** illustrated in FIG. **3**A generates a passband tuning control output **106** and a stopband tuning control output **108**. The variable reactive impedance components in both the first RF filter **38**A and the second RF filter **38**B are set in accordance with one or more signal levels of the passband tuning control output **106**. In this manner, the stopband **66** is shifted to include the RF transmission band in accordance with one or more signal levels of the stopband tuning control output **108**. Similarly, reactive impedance levels of variable reactive components in the first RF filter **38**A and the second RF filter **38**B are set in accordance with one or more signal levels of the passband tuning control output **106**. The RF filter circuit **22** may also include active RF filters, Surface Acoustic Wave (SAW) filters, or any other type of RF filter or combination of RF filters that is suitable to provide a desired frequency response. As such, the tuning circuit **69** may employ various types of tuning topologies, depending on the particular filtering topology being employed by the RF filter circuit **22**.

[0072] By placing the passband **62** of the first RF filter **38**A and the second RF filter **38**B in the RF communication band, the RF filter circuit **22** passes the first RF QHRS **102** and the second RF QHRS **104** to the second hybrid coupler **20**. Due to the differential phase shift provided by the second RF filter **38**B, the second RF QHRS **104** has an approximately 270 degree or 3/2π radian phase difference with respect to the first RF QHRS **102** once the first RF QHRS **102** and the second RF QHRS **104** are output from the RF filter circuit **22**.

[0073] Referring again to FIG. **3**A, the second hybrid coupler **20** receives the first RF QHRS **102** from the first RF filter **38**A at the fifth port **94**. The second RF QHRS **104** is received by the second hybrid coupler **20** from the second RF filter **38**B at the sixth port **96**. As discussed above, the phase difference between the first RF QHRS **102** and the second RF QHRS **104** have about 270 degrees or 3π/2 radians. Thus, for example, if the first RF QHRS **102** has a phase of zero degrees, the second RF QHRS **104** would have a phase of approximately 270 degrees (or 3π/2 radians). From the fifth port **94** to the seventh port **98**, the second hybrid coupler **20** provides no phase shift. Alternatively, the second hybrid coupler **20** may be configured to provide a phase shift from the fifth port **94** to the seventh port **98** of Δ (i.e., such as +45 degrees or π/4 radians).

[0074] The second hybrid coupler **20** is configured to output the RF receive output signal **26** from the seventh port **98** in response to the first RF QHRS **102** being received from the RF filter circuit **22** at the fifth port **94** and the second RF QHRS **104** being received from the RF filter circuit **22** at the sixth port **96**. In this particular embodiment, the second hybrid coupler **20** is configured to pass the second RF QHRS **104** from the sixth port **96** to the seventh port **98** by providing another quadrature phase shift. Thus, as seen at the seventh port **98**, the second RF QHRS **104** has a phase of 360 degree or 2π radian phase. The second hybrid coupler **20** provides no phase shift to the first RF QHRS **102** from the sixth port **96** to the seventh port **98**. The first RF QHRS **102** is thus passed with a phase of 0 degrees to the seventh port **98**. Thus, the interference between the first RF QHRS **102** and the second RF QHRS **104** at the seventh port **98** is constructive. Accordingly, the second hybrid coupler **20** is configured to combine the first RF QHRS **102** and the second RF QHRS **104** into the RF receive output signal **26**. The RF receive output signal **26** is then transmitted to external circuitry in an RF front end, such as an RF receive chain.

[0075] The second hybrid coupler **20** shown in FIG. **3**A is configured to provide a quadrature phase shift from the fifth port **94** to the eighth port **100**. In this example, the phase shift is 90 degrees (or π/2 radians), and thus the second RF QHRS **104** has a phase, as seen from the eighth port **100**, of 90 degrees (note that the first RF QHRS **102** was received at the fifth port **94** with a phase of 0 degrees in this example). The second RF QHRS **104** is received with a phase of 270 degrees at the sixth port **96** and the second hybrid coupler **20** provides no phase shift between the sixth port **96** and the eighth port **100**. Thus, the phase difference between the first RF QHRS **102** and the second RF QHRS **104** as seen from the eighth port **100** is about 180 degrees. Accordingly, the quadrature phase shift at the eighth port **100** from the fifth port **94** results in destructive interference between the first RF QHRS **102** and the second RF QHRS **104** at the eighth port **100**. As a result, the first RF QHRS **102** and the second RF QHRS **104** are substantially cancelled at the eighth port **100**. In this manner, the eighth port **100** is substantially isolated from receive signal flow.

[0076] An impedance termination **24** is coupled to the eighth port **100** of the second hybrid coupler **20**. The impedance termination **24** may be a 50 Ohm load. Due to the phase-shifting provided by the first hybrid coupler **18**, the first RF filter **38**A and the second hybrid coupler **20**, spurious emissions from the second port **88** of the first hybrid coupler **18** would see a very high (theoretically infinite) impedance at the seventh port **98** of the second hybrid coupler **20** but only the impedance termination **24** at the eighth port **100**. Thus, the spurious emissions are aggregated to be an aggregated noise signal **110** at the eighth port **100**. This aggregated noise signal **110** is dissipated by the impedance termination **24**. Additionally, the eighth port **100** is isolated from the seventh port **98**. As such, the seventh port **98** is substantially unresponsive to signals incident at the eighth port **100**, and the seventh port **98** is substantially unresponsive to signals incident at the eighth port **100**. An impedance of the impedance termination **24** can be varied by the tuning circuit **69** in order to maintain the impedance termination **24** at a characteristic 50 Ohms.

[0077] FIG. **3**B illustrates the tunable RF diplexer **10** along with the receive signal flow of the tunable RF diplexer **10** with respect to the procedures **1002**, **1014**-**1020** shown in FIG. **2**A for the RF receive input signal **16** and the RF receive output signal **28**. The procedures **1000**, **1004**-**1012** and the procedures for the RF receive input signal **14** and the RF receive output signal **26** shown in FIG. **2**A and illustrated in FIG. **3**A can performed simultaneously with the procedures **1002**, **1014**-**1020** shown FIG. **2**A and illustrated in FIG. **3**B. Thus, FIG. **3**A and FIG. **3**B together illustrate receive flow diplexing for the tunable RF diplexer **10**. As shown in FIG. **3**B, the antenna **12** intercepts the RF receive input signal **16** as electromagnetic waves in free space. These electromagnetic waves result in excitations within the antenna **12**, thereby converting the electromagnetic waves into the RF receive input signal **16**. The first hybrid coupler **18** is configured to receive the RF receive input signal **16** at the first port **86**. The RF receive input signal **16** operates in the same RF communication band as the RF transmission input signal **32** (shown in FIG. **1**). The first hybrid coupler **18** is operable to split the RF receive input signal **16** into the first RF QHRS **112** and the second RF QHRS **114**. Since the first RF QHRS **112** and the second RF QHRS **114** are quadrature hybrids, the first RF QHRS **112** and the second RF QHRS **114** are approximately equal in power but have a quadrature phase difference of 90 degrees or π/2 radians. The first hybrid coupler **18** outputs the first RF QHRS **112** from the third port **90** and outputs the second RF QHRS **114** from the fourth port **92** in response to receiving the RF receive input signal **16** at the first port **86**.

[0078] In the embodiment illustrated in FIG. **3**B, the first RF QHRS **112** is phase-aligned with the RF receive input signal **16**, while the second RF QHRS **114** has a phase difference of about 90 degrees with respect to the RF receive input signal **16**. It should be noted that this may or may not be the case. For example, in alternative embodiments, a phase shift of Δ may be provided between the first port **86** and the third port **90**, and thus, a phase shift of Δ+90 degrees (or π/2 radians) would be provided between the first port **86** and the fourth port **92**.

[0079] The RF filter circuit **22** is operable to reflect the first RF QHRS **112** and the second RF QHRS **114**. As discussed above, the frequency response of the RF filter circuit **22** defines the stopband **66** and the RF filter circuit **22** is tunable so as to shift the stopband **66**. For example, the stopband **66** may be a notch that is shiftable. The tuning circuit **69** is configured to tune the frequency response of the RF filter circuit **22** so that the signal bandwidth of the first RF QHRS **112** and the signal bandwidth of the second RF QHRS **114** are each within the stopband. More specifically, the tuning circuit **69** is configured to shift the stopband **66** of the first RF filter **38**A and the stopband **66** of the second RF filter **38**B into the frequency band of the RF receive input signal **16**. For instance, the tuning circuit **69** may be configured to place the notch within the frequency band of the RF receive input signal **16** so that the notch is centered at the RF receive signal frequency. In this embodiment, the tuning circuit **69** generates the stopband tuning control output **108**. Variable reactive impedance components in both the first RF filter **38**A and the second RF filter **38**B are responsive to the signal level of the stopband tuning control output **108** so as to adjust the variable impedance levels based on the signal level of the stopband tuning control output **108**. As a result, the notch defined by the individual frequency response of the first RF filter **38**A is shifted to include the signal bandwidth of the second RF QHRS **114**. Also, the notch defined by the individual frequency response of the second RF filter **38**B is shifted to include the signal bandwidth of the first RF QHRS **112**. In other words, the notches defined by the individual frequency responses of the first RF filter **38**A and the second RF filter **38**B are placed in the frequency band of the RF receive input signal **16**.

[0080] Since the tuning circuit **69** has tuned the frequency response of the RF filter circuit **22** so that the stopband **66** includes the frequency band of the RF receive input signal **16**, the RF filter circuit **22** blocks the first RF QHRS **112** and the second RF QHRS **114**. Accordingly, the seventh port **98** is substantially isolated from the receive signal flow of the RF receive input signal **16**. The RF filter circuit **22** then reflects the first RF QHRS **112** and the second RF QHRS **114** back to the first hybrid coupler **18**. In the embodiment illustrated in FIG. 3B, the first RF filter **38**A reflects the first RF QHRS **112** back to the first hybrid coupler **18** at the third port **90**. The second RF filter **38**B reflects the second RF QHRS **114** back to the first hybrid coupler **18** at the fourth port **92**.

[0081] The first hybrid coupler **18** is configured to combine the first RF QHRS **112** and the second RF QHRS **114** into the RF receive output signal **28** at the second port **88**. To combine the first RF QHRS **112** and the second RF QHRS **114** into the RF receive output signal **28**, the first hybrid coupler **18** is configured to pass the second RF QHRS **114** from the fourth port **92** to the second port **88**. Additionally, the second hybrid coupler is configured to pass the first RF QHRS **112** from the third port **90** to the second port **88**. However, the first hybrid coupler **18** provides a quadrature phase shift to the first RF QHRS **112** at the second port **88**. Thus, the first RF QHRS **112** is provided substantially as a duplicate of the second RF QHRS **114** at the second port **88**. For example, the phase of the first RF QHRS **112** at the third port **90** is approximately zero degrees. If the phase of the second RF QHRS **114** is 90 degrees at the fourth port **92**, the second RF QHRS **114** has a phase of 90 degrees at the second port **88** since the first hybrid coupler **18** provides no phase shift from the fourth port **92** to the second port **88**. However, due to the quadrature phase shift between the third port **90** and the second port **88**, the first RF QHRS **112** has a phase of about 90 degrees at the second port **88**. Accordingly, the first RF QHRS **112** and the second RF QHRS **114** constructively interfere at the second port **88** to output the RF receive output signal **28** from the second port **88**.

[0082] Also, note that the first hybrid coupler **18** is configured such that the quadrature phase shift at the first port **86** results in destructive interference between the first RF QHRS **112** and the second RF QHRS **114**. Referring again to the previous example, upon reflection of the first RF QHRS **112** and the second RF QHRS **114** by the RF filter circuit **22**, the first RF QHRS **112** appears to have a phase of zero degrees at the first port **86**, but the second RF QHRS **114** appears to have a phase of 180 degrees at the first port **86** due to the quadrature phase shift provided by the first hybrid coupler **18** from the fourth port **92** to the first port **86**. As a result, the first RF QHRS **112** and the second RF QHRS **114** are substantially cancelled at the first port **86**. Consequently, most, if not all, of the power of the first RF QHRS **112** and the second RF QHRS **114** is transferred to the second port **88** and provided in the RF receive output signal **28**. The first hybrid coupler **18** is thus configured to output the RF receive output signal **28** from the second port **88** in response to the first RF QHRS **112** being reflected back by the RF filter circuit **22** to the third port **90** and the second RF QHRS **114** being reflected back by the RF filter circuit **22** to the fourth port **92**. The configuration of the first hybrid coupler **18** and the RF filter circuit **22** thus provides substantial isolation between the second port **88** and the first port **86**.

[0083] It should be noted that due to the differential phase shift provided by the second RF filter **38**B, isolation is provided between the second port **88** for one of the RF communication bands and the seventh port **98** for the other one of the diplexed RF frequency bands. For the following explanation, it is assumed that the first port **86**, the third port **90**, the fourth port **92**, and the second port **88** are port 1, port 2, port 3, and port 4 respectively of the first hybrid coupler **18**. Also, it is assumed that the seventh port **98**, the fifth port **94**, the sixth port **96**, and the eighth port **100** are port 1, port 2, port 3, and port 4 of the second hybrid coupler **20**. Subscripts on the S parameters indicate that the S parameter is either for the first hybrid coupler **18** or the second hybrid coupler **20**. Now, given that the differential phase shift is provided between the fourth port **92** of the first hybrid coupler **18** and the sixth port **96** of the second hybrid coupler **20**, the tunable RF diplexer **10** provides isolation between the second port **88** and the seventh port **98** by assuring that matching of $S(2,4)_{18}*S(1,2)_{20}=S(3,4)_{18}*S(1,3)_{20}$. As a result of the differential phase shift provided by the second RF filter **38**B, the isolation relies on the same sum term of the receive signal flow, which remains approximately true over broad frequency range and thus provides better broadband isolation.

[0084] FIG. 3C illustrates the tunable RF diplexer **10** along with a transmission signal flow of the tunable RF diplexer **10** with respect to the procedures **2000**, **2004-2012** (shown in FIG. 2B) for the RF transmission input signal **30** and the RF transmission output signal **34**. The second hybrid coupler **20** is configured to receive the RF transmission input signal **30** at the seventh port **98**. The eighth port **100** of the second hybrid coupler **20** is isolated from the seventh port **98**. This means that the seventh port **98** is substantially unresponsive to signals incident at the eighth port **100**, and the eighth port **100** is substantially unresponsive to signals incident at the seventh port **98**. As a result, the eighth port **100** is substantially unresponsive to the RF transmission input signal **30** incident at the seventh port **98**.

[0085] The second hybrid coupler **20** is operable to split the RF transmission input signal **30** into a first RF QHTS **116** and a second RF QHTS **118**. In this manner, the first RF QHTS **116** and the second RF QHTS **118** have approximately the same power ratio with respect to the RF transmission input signal **30**, but have a quadrature phase difference of approximately 90 degrees or π/2 radians. The first RF QHTS **116** is output from the fifth port **94**, and the second RF QHTS **118** is output from the sixth port **96**. In the embodiment shown in FIG. **3**C, the seventh port **98** is phase-aligned with the fifth port **94**, while the seventh port **98** has a quadrature phase shift with respect to the sixth port **96**. Thus, the first RF QHTS **116** is approximately phase-aligned with the RF transmission input signal **30**, but there is a quadrature phase difference between the RF transmission input signal **30** and the second RF QHTS **118**.

[0086] Note that in alternative embodiments, this may or may not be the case. For example, there may be a phase shift between the seventh port **98** and the fifth port **94**. The phase shift between the seventh port **98** and the sixth port **96** may then be equal to this phase shift plus the quadrature phase shift of approximately 90 degrees or π/2 radians. Accordingly, so long as the phase difference between the first RF QHTS **116** and the second RF QHTS **118** is about 90 degrees or π/2 radians, phase alignment between the fifth port **94** and the seventh port **98**, and between the sixth port **96** and the seventh port **98**, can vary.

[0087] The first RF QHTS **116** is output at the fifth port **94** to the first RF filter **38**A of the RF filter circuit **22**. Additionally, the second RF QHTS **118** is output at the sixth port **96** to the second RF filter **38**B of the RF filter circuit **22**. The first RF filter **38**A is coupled to the fifth port **94** so as to receive the first RF QHTS **116** from the second hybrid coupler **20**. The second RF filter **38**B is coupled to the sixth port **96** so as to receive the second RF QHTS **118** from the second hybrid coupler **20**. In this embodiment, the second RF filter **38**B is configured to provide a differential phase shift to the second RF QHTS **118**. Accordingly, when the first RF QHTS **116** and the second RF QHTS **118** are output from the RF filter circuit **22**, the phase difference between the first RF QHTS **116** and the second RF QHTS **118** is approximately 270 degrees or 3/2π radians.

[0088] Referring again to FIG. **3**C, the tuning circuit **69** is configured to tune the frequency response of the RF filter circuit **22** so that the passband **62** includes the RF communication band of the RF transmission input signal **30**. The tuning circuit **69** thus shifts the passband **62** of the first RF filter **38**A and the second RF filter **38**B to include the RF communication band of the RF transmission input signal **30**. In this manner, the RF filter circuit **22** is operable to pass the first RF QHTS **116** and the second RF QHTS **118** to the first hybrid coupler **18**. More specifically, by placing the passband **62** in the RF transmission band, the RF filter circuit **22** passes the first RF QHTS **116** and the second RF QHTS **118** to the first hybrid coupler **18**. In this particular embodiment, the first RF filter **38**A passes the first RF QHTS **116** to the first hybrid coupler **18**, while the second RF filter **38**B passes the second RF QHTS **118** to the first hybrid coupler **18**. Spurious emissions of the first RF QHTS **116** and the second RF QHTS **118** are reflected back to the second hybrid coupler **20**. The impedance termination **24** is coupled to the eighth port **100** of the second hybrid coupler **20**. The impedance termination **24** may be a 50 Ohm load. Due to the phase-shifting provided by the second hybrid coupler **20**, the spurious emissions in the RF receive band see a very high (theoretically infinite) impedance level at the seventh port **98**, but only the impedance termination **24** at the eighth port **100**. Thus, the spurious emissions are aggregated to be the aggregated noise signal **110** at the eighth port **100**. This aggregated noise signal **110** is dissipated by the impedance termination **24**.

[0089] Referring again to FIG. **3**C, the first hybrid coupler **18** receives the first RF QHTS **116** from the first RF filter **38**A at the third port **90**. The second RF QHTS **118** is received by the first hybrid coupler **18** from the second RF filter **38**B at the fourth port **92**. As discussed above, the first RF QHTS **116** and the second RF QHTS **118** have a phase difference of about 270 degrees or 3π/2 radians when received by the first hybrid coupler **18**. Thus, for example, if the first RF QHTS **116** has a phase of zero degrees, the second RF QHTS **118** would have a phase of approximately 270 degrees (or 3π/2 radians). From the fourth port **92** to the second port **88**, the first hybrid coupler **18** provides no phase shift. Alternatively, the first hybrid coupler **18** may be configured to provide a phase shift from the third port **90** to the first port **86** of Δ.

[0090] The first hybrid coupler **18** shown in FIG. **3**C is configured to provide a quadrature phase shift from the third port **90** to the second port **88**. In this example, the phase shift is 90 degrees (or π/2 radians), and thus the second RF QHTS **118** has a phase, as seen from the second port **88**, of 270 degrees (note that the second RF QHTS **118** was received with a phase of 270 degrees in this example, and thus is seen with a phase of 270 degrees). Alternatively, the phase shift from the fourth port **92** to the second port **88** may be Δ+90 degrees (or π/2 radians). In any case, the phase difference between the first RF QHTS **116** and the second RF QHTS **118** as seen from the second port **88** is about 180 degrees (note that the first RF QHTS **116** was received with a phase of 0 degrees and thus is seen with a phase of 90 degrees at the second port **88** with quadrature phase shift of 90 degrees between the third port **90** and the second port **88**). Accordingly, the quadrature phase shift at the second port **88** from the third port **90** and the differential phase shift of the second RF filter **38**B results in destructive interference between the first RF QHTS **116** and the second RF QHTS **118** at the second port **88**. As a result, the first RF QHTS **116** and the second RF QHTS **118** are substantially cancelled at the second port **88**. In this manner, the second port **88** is substantially isolated from transmission signal flow for the RF transmission input signal **30** and the RF transmission output signal **34**.

[0091] The first hybrid coupler **18** is configured to output the RF transmission output signal **34** from the first port **86** in response to the first RF QHTS **116** being received from the RF filter circuit **22** at the third port **90** and the second RF QHTS **118** being received from the RF filter circuit **22** at the fourth port **92**. In this particular embodiment, the first hybrid coupler **18** is configured to pass the second RF QHTS **118** from the fourth port **92** to the first port **86**. The first hybrid coupler **18** provides a quadrature phase shift to the second RF QHTS **118** from the fourth port **92** to the first port **86**. The second RF QHTS **118** is thus passed with a phase of 360 degrees to the first port **86** since the second RF QHTS **118** was received at the fourth port **92** with a phase of approximately 270 degrees. Alternatively, the first hybrid coupler **18** may provide a phase shift of Δ to the second RF QHTS **118** when passed from the fourth port **92** to the first port **86**. The first hybrid coupler **18**

is configured to pass the first RF QHTS **116** from the third port **90** to the first port **86**. The first hybrid coupler **18** provides no phase shift to the first RF QHTS **116** at the first port **86**. Thus, the first RF QHTS **116** has a phase of approximately zero at the first port **86**. Alternatively, if a phase shift of Δ+90 degrees (or π/2 radians) 90 degrees (or π/2 radians) were provided to the second RF QHTS **118** from the fourth port **92** to the first port **86**, the quadrature phase shift would be Δ.

[0092] Accordingly, the first RF QHTS **116** is provided substantially as a duplicate of the second RF QHTS **118** at the first port **86**. As a result, the first RF QHTS **116** and the second RF QHTS **118** constructively interfere at the first port **86** to output the RF transmission output signal **34** from the first port **86**. Note that since the first RF QHTS **116** and the second RF QHTS **118** substantially cancel at the second port **88** due to destructive interference, very little or no power is transferred from the first RF QHTS **116** and the second RF QHTS **118** to the second port **88**. Instead most, if not all, of the power in the first RF QHTS **116** and the second RF QHTS **118** is transferred to the first port **86** and provided in the RF transmission output signal **34**.

[0093] FIG. **3**D illustrates the tunable RF diplexer **10** along with the transmission signal flow of the tunable RF diplexer **10** with respect to the procedures **2002**, **2014-2020** shown in FIG. **2**B for the RF transmission input signal **32** and the RF transmission output signal **36**. The procedures **2000**, **2004-2012** and the procedures for the RF transmission input signal **30** and the RF transmission output signal **34** shown in FIG. **2**B and illustrated in FIG. **3**A can performed simultaneously with the procedures **2002**, **2014-2020** shown in FIG. **2**B and illustrated in FIG. **3**D. Thus, FIG. **3**C and FIG. **3**D together illustrates transmission flow diplexing for the tunable RF diplexer **10**. As shown in FIG. **3**B, the first hybrid coupler **18** is operable to receive the RF transmission input signal **32** at the second port **88**. The RF transmission input signal **32** operates within the same RF communication band as the RF receive band of the RF receive input signal **16** (see FIG. **3**B). The first hybrid coupler **18** is operable to split the RF transmission input signal **32** into the first RF QHTS **120** and the second RF QHTS **122**. Since the first RF QHTS **120** and the second RF QHTS **122** are quadrature hybrids, the first RF QHTS **120** and the second RF QHTS **122** are approximately equal in power, but have a quadrature phase difference of 90 degrees or π/2 radians. The first hybrid coupler **18** is operable to output the first RF QHTS **120** from the fourth port **92** and to output the second RF QHTS **122** from the third port **90** in response to receiving the RF transmission input signal **32** at the second port **88**.

[0094] In the embodiment illustrated in FIG. **3**D, the first RF QHTS **120** is phase-aligned with the RF transmission input signal **32**, while the second RF QHTS **122** has a phase difference of about 90 degrees with respect to the RF transmission input signal **32**. It should be noted that this may or may not be the case. For example, in alternative embodiments, a phase shift of Δ (i.e., +45 degrees or π/4 radians) may be provided between the second port **88** and the fourth port **92**, and thus, a phase shift of Δ±90 degrees (or π/2 radians) would be provided between the second port **88** and the third port **90**.

[0095] The RF filter circuit **22** is operable to reflect the first RF QHTS **120** and the second RF QHTS **122**. As discussed above, the frequency response of the RF filter circuit **22** defines the stopband **66**, and the RF filter circuit **22** is tunable so as to shift the stopband. For example, the stopband **66** may be a notch that is shiftable. The tuning circuit **69** is configured to tune the frequency response of the first RF filter **38**A and the second RF filter **38**B of RF filter circuit **22** so that the signal bandwidth of the first RF QHTS **120** and the signal bandwidth of the second RF QHTS **122** are each within the stopband. For instance, the tuning circuit **69** may be configured to place the notch within the RF communication band of the RF transmission input signal **32** so that the notch is centered at the RF transmission signal frequency of the RF transmission input signal **32**. In this embodiment, the tuning circuit **69** generates the stopband tuning control output **108**.

[0096] Since the tuning circuit **69** has tuned the frequency response of the RF filter circuit **22** so that the stopband **66** includes the RF transmission signal band, the RF filter circuit **22** blocks the first RF QHTS **120** and the second RF QHTS **122**. Accordingly, the second hybrid coupler **20** is substantially isolated from the transmission signal flow of the RF transmission input signal **32** and the RF transmission output signal **34**. The second RF filter **38**B then reflects the first RF QHTS **120**, and the first RF filter **38**A reflects the second RF QHTS **122** back to the first hybrid coupler **18**. In the embodiment illustrated in FIG. **3**D, the second RF filter **38**B reflects the first RF QHTS **120** back to the first hybrid coupler **18** at the fourth port **92**. The first RF filter **38**A reflects the second RF QHTS **122** back to the first hybrid coupler **18** at the third port **90**.

[0097] The first hybrid coupler **18** is configured to combine the first RF QHTS **120** and the second RF QHTS **122** into the RF transmission output signal **36** at the first port **86**. To combine the first RF QHTS **120** and the second RF QHTS **122** into the RF transmission output signal **36**, the first hybrid coupler **18** is configured to pass the first RF QHTS **120** from the fourth port **92** to the first port **86**. Additionally, the first hybrid coupler **18** is configured to pass the second RF QHTS **122** from the third port **90** to the first port **86**. However, the first hybrid coupler **18** provides a quadrature phase shift to the first RF QHTS **120** from the fourth port **92** to the first port **86**. Thus, the first RF QHTS **120** is provided substantially as a duplicate of the second RF QHTS **122** at the first port **86**. For example, if the phase of the second RF QHTS **122** is 90 degrees at the third port **90**, the first RF QHTS **120** has a phase of 90 degrees at the first port **86**, since the phase of the first RF QHTS **120** was zero degrees at the fourth port **92**. More specifically, due to the quadrature phase shift between the fourth port **92** and the first port **86**, the first RF QHTS **120** has a phase of about 90 degrees at the first port **86** just like the second RF QHTS **122**. Accordingly, the first RF QHTS **120** and the second RF QHTS **122** constructively interfere at the first port **86** to output the RF transmission output signal **36** from the first port **86**.

[0098] Also, note that the first hybrid coupler **18** is configured such that the quadrature phase shift at the second port **88** results in destructive interference between the first RF QHTS **120** and the second RF QHTS **122**. Referring again to the previous example provided, at the second port **88**, the first RF QHTS **120** appears to have a phase of zero degrees, but the second RF QHTS **122** appears to have a phase of 180 degrees. As a result, the first RF QHTS **120** and the second RF QHTS **122** are substantially cancelled at the second port **88**. Consequently, most, if not all, of the power of the first RF QHTS **120** and the second RF QHTS **122** is transferred to the first port **86** and provided in the RF transmission output signal **36**. The first hybrid coupler **18** is thus configured to output the RF transmission output signal **36** from the first port **86** in response to the first RF QHTS **120** being reflected back by the RF filter circuit **22** to the fourth port **92** and the second RF QHTS **122** being reflected back by the RF filter circuit **22** to the third port **90**.

[0099] It should be noted that due to the differential phase shift provided by the second RF filter **38**B, isolation is provided between the second port **88** for one of the RF communication bands and the seventh port **98** for the other one of the diplexed RF frequency bands. Given that the differential phase shift is provided between the fourth port **92** of the first hybrid coupler **18** and the sixth port **96** of the second hybrid coupler **20**, the tunable RF diplexer **10** provides isolation between the second port **88** and the seventh port **98** by assuring that matching of $S(2,4)_{18}*S(1,2)_{20}=S(3,4)_{18}*S(1,3)_{20}$. As a result of the differential phase shift provided by the second RF filter **38**B, the isolation relies on the same sum term of the transmission signal flow which remains approximately true over broad frequency range and thus provides better broadband isolation for transmission diplexing. In addition to the improved isolation for receive and transmission diplexing, providing the differential phase shift improves the performances of the tunable RF diplexer **10** by tuning the impedance termination **24** to match a load of the antenna **12**.

[0100] FIG. **4** illustrates a circuit diagram of one embodiment of the tunable RF diplexer **10** shown in FIG. **1** and FIGS. **3**A-**3**D. In this embodiment, the first hybrid coupler **18** and the second hybrid coupler **20** are both lumped-element circuits formed as an LCCL network. With regards to the first hybrid coupler **18**, the first port **86**, the second port **88**, the third port **90**, and the fourth port **92** are each provided as nodes. An inductor **124** is connected between the first port **86** and the third port **90** while an inductor **126** is coupled between the second port **88** and the fourth port **92**. The inductor **124** and the inductor **126** are mutually magnetically coupled to one another. A capacitor **128** is connected between the first port **86** and the fourth port **92** while a capacitor **130** is connected between the third port **90** and the second port **88**. In this manner, the appropriate quadrature phase shifts are provided, as described above with respect to FIGS. **3**A-**3**D. With regards to the second hybrid coupler **20**, the fifth port **94**, the sixth port **96**, the seventh port **98**, and the eighth port **100** are also each provided as nodes. An inductor **132** is connected between the fifth port **94** and the seventh port **98**, while an inductor **134** is coupled between the sixth port **96** and the eighth port **100**. The inductor **132** and the inductor **134** are mutually magnetically coupled to one another. A capacitor **136** is connected between the seventh port **98** and the sixth port **96** while a capacitor **138** is connected between the fifth port **94** and the eighth port **100**. In this manner, the appropriate quadrature phase shifts are provided, as described above with respect to FIGS. **3**A-**3**D. An embodiment of the impedance termination **24** is also shown. In this example, the impedance termination **24** is provided as a resistor **140** and a phase shifter **142**. By adjusting a phase of the phase shifter **142**, the impedance load of the impedance termination **24** can be kept at the appropriate characteristic impedance.

[0101] An embodiment of the first RF filter **38**A and the second RF filter **38**B are also shown in FIG. **4**. The first RF filter **38**A includes a parallel resonator **146**A coupled in series with the third port **90**. Additionally, the second RF filter **38**B includes a parallel resonator **146**B coupled in series with the fourth port **92**. As such, the stopband **66** of both the first RF filter **38**A and the second RF filter **38**B are each provided as a notch. The parallel resonator **146**A and the parallel resonator **146**B are thus each an RF notch filter. To provide the passband **62**, both the first RF filter **38**A and the second RF filter **38**B use weakly coupled resonators. With regard to the first RF filter **38**A, the first RF filter **38**A forms a tunable RF filter path **144**A shown in FIG. **4** connected between the third port **90** of the first hybrid coupler **18** and the fifth port **94** of the second hybrid coupler **20**. In order to provide the stopband **66** of the first RF filter **38**A, the first RF filter **38**A includes the parallel resonator **146**A coupled in series with the third port **90** of the first hybrid coupler **18**. The parallel resonator **146**A includes an inductor **148**A and a variable capacitive structure **150**A. As such, the stopband **66** of the first RF filter **38**A is provided as a notch. The variable capacitive structure **150**A has a variable capacitance that is adjusted by the stopband tuning control output **108**. In this manner, the stopband **66** of the first RF filter **38**A can be shifted into the RF communication band of the RF receive input signal **16**, the RF receive output signal **28**, the RF transmission input signal **32**, and the RF transmission output signal **36**.

[0102] With regards to providing the passband **62** of the first RF filter **38**A, the tunable RF filter path **144**A includes an embodiment of a resonator R(1,1) and an embodiment of a resonator R(1,2). The resonator R(1,1) and the resonator R(1,2) are weakly coupled to one another. More specifically, the resonator R(1,1) includes an inductor **152**A and a capacitive structure **154**A. The resonator R(1,2) includes an inductor **156**A and a capacitive structure **158**A.

[0103] The resonator R(1,1) and the resonator R(1,2) are a pair of weakly coupled resonators. The resonator R(1,1) and the resonator R(1,2) are weakly coupled by providing the inductor **152**A and the inductor **156**A such that the inductor **152**A and the inductor **156**A have a weak mutual coupling. Although the resonator R(1,1) and the resonator R(1,2) are weakly coupled, the inductor **156**A has a maximum lateral width and a displacement from the inductor **152**A that is less than or equal to half the maximum lateral width of the inductor **156**A. As such, the inductor **152**A and the inductor **156**A are relatively close to one another. The displacement between the inductor **152**A and the inductor **156**A may be measured from a geometric centroid of the inductor **152**A to a geometric centroid of the inductor **156**A. The maximum lateral width may be a maximum dimension of the inductor **156**A along a plane defined by its largest winding. The weak coupling between the inductor **152**A and the inductor **156**A is obtained through topological techniques. For example, the inductor **152**A and the inductor **156**A may be fully or partially aligned, where winding(s) of the inductor **152**A and winding(s) of the inductor **156**A are configured to provide weak coupling through cancellation. Alternatively or additionally, a plane defining an orientation of the winding(s) of the inductor **152**A and a plane defining an orientation of the winding(s) of the inductor **156**A may be fully or partially orthogonal to one another. Some of the magnetic couplings between the resonators R can be unidirectional (passive or active). This can significantly improve isolation (e.g., isolation in diplexers). In other embodiments, topological techniques are not used to provide weak mutual coupling between the inductor **152**A and the inductor **156**A. Rather, the inductor **152**A and the inductor **156**A are weakly mutually coupled by being sufficiently far apart.

[0104] To maximize the quality (Q) factor of the tunable RF filter paths **144**A through **68**, most of the total mutual coupling should be realized magnetically, and only fine-tuning is provided electrically. While the magnetic coupling can be adjusted only statically, with a new layout design, the electric coupling can be tuned on the fly (after fabrication). The filter characteristics (e.g., bias network structure, resonator capacitance) can be adjusted based on given coupling coefficients to maximize filter performance.

[0105] To provide a tuning range to tune the passband **62** of the tunable RF filter path **144**A and provide a fast roll-off from a low-frequency side to a high-frequency side of the transfer function, the tunable RF filter path **144**A is configured to change a sign of a total mutual coupling coefficient between the resonator R(1,1) and the resonator R(1,2). Accordingly, the tunable RF filter path **144**A includes a cross-coupling capacitive structure C(AP1). As shown in FIG. **4**, the cross-coupling capacitive structure C(AP1) is electrically connected in the tunable RF filter path **144**A between the resonator R(1,1) and the resonator R(1,2) so as to provide a variable coupling coefficient between the resonator R(1,1) and the resonator R(1,2). The cross-coupling capacitive structure C(AP1) is a variable cross-coupling capacitive structure configured to vary the coupling coefficient provided between the resonator R(1,1) and the resonator R(1,2). The passband tuning control output **106** is provided to the cross-coupling capacitive structure C(AP1) to adjust the variable coupling coefficient and thereby shift the passband **62** into the RF communication band of the RF receive input signal **14**, the RF receive output signal **26**, the RF transmission input signal **30**, and the RF transmission output signal **34**.

[0106] In the resonator R(1,1), the inductor **152**A and the capacitive structure **154**A are electrically connected in parallel. More specifically, the inductor **152**A has an end **160**A and an end **162**A, which are disposed opposite to one another. The ends **160**A, **162**A are each electrically connected to the capacitive structure **154**A, which is grounded. Thus, the resonator R(1,1) is a single-ended resonator. The inductor **156**A is also electrically connected in parallel to the capacitive structure **158**A. More specifically, the inductor **156**A has an end **164**A and an end **166**A, which are disposed opposite to one another. The ends **164**A, **166**A are each electrically connected to the capacitive structure **158**A, which is grounded.

[0107] As shown in FIG. **4**, the inductor **152**A and the inductor **156**A are mutually magnetically coupled so as to have a non-inverting mutual coupling and the no phase inversion is provided by the first RF filter **38**A. In this embodiment, the resonator R(1,1) and the resonator R(1,2) are single-ended resonators. The inductor **152**A is magnetically coupled to the inductor **156**A such that an RF signal received at the end **160**A of the inductor **152**A with a voltage polarity (i.e., either a positive voltage polarity or a negative voltage polarity) results in a filtered RF signal being transmitted out the end **164**A of the inductor **156**A with the same voltage polarity. Also, the inductor **156**A is magnetically coupled to the inductor **152**A such that an RF signal received at the end **164**A of the inductor **156**A with a voltage polarity (i.e., either a positive voltage polarity or a negative voltage polarity) results in a filtered RF signal being transmitted out the end **160**A of the inductor **152**A with the same voltage polarity. This is indicated in FIG. **4** by the dot convention where a dot is placed at the end **160**A of the inductor **152**A, and a dot is placed at the end **164**A of the inductor **156**A. Additionally, the inductors **152**A, **156**A may be magnetically coupled so as to have a low magnetic coupling coefficient through field cancellation, with the variable coupling coefficient. In this case, the inductor **152**A and the inductor **156**A are arranged such that the inductor **156**A reduces a mutual magnetic coupling coefficient of the inductor **152**A.

[0108] The resonator R(1,2) is operably associated with the resonator R(1,1) such that an energy transfer factor between the resonator R(1,1) and the resonator R(1,2) is less than 10%. Thus, the resonator R(1,1) and the resonator R(1,2) are weakly coupled. A total mutual coupling between the resonator R(1,1) and the resonator R(1,2) is provided by a sum total of the mutual magnetic factor between the resonator R(1,1) and the resonator R(1,2) and the mutual electric coupling coefficient is between the resonator R(1,1) and the resonator R(1,2). In this embodiment, the mutual magnetic coupling coefficient between the inductor **152**A and the inductor **156**A is a fixed mutual magnetic coupling coefficient. Although embodiments of the resonators R(1,1), R(1,2) may be provided so as to provide a variable magnetic coupling coefficient between the resonators R(1,1), R(1,2), embodiments of the resonators R(1,1), R(1,2) that provide variable magnetic couplings can be costly and difficult to realize. However, providing variable electric coupling coefficients is easier and more economical. Thus, using the cross-coupling capacitive structure C(AP1) to provide the variable electric coupling coefficient is an economical technique for providing a tunable filter characteristic between the resonators R(1,1), R(1,2). Furthermore, since the mutual magnetic coupling coefficient between the inductor **152**A and the inductor **156**A is fixed, the tunable RF filter path **144**A has lower insertion losses. A matching capacitor **168**A is connected in series with the fifth port **94** of the first hybrid coupler **18**.

[0109] With regard to the second RF filter **38**B, the second RF filter **38**B forms a tunable RF filter path **144**B shown in FIG. **4** connected between the fourth port **92** of the first hybrid coupler **18** and the sixth port **96** of the second hybrid coupler **20**. In order to provide the stopband **66** of the second RF filter **38**B, the second RF filter **38**B includes a parallel resonator **146**B coupled in series with the third port **90** of the first hybrid coupler **18**. The parallel resonator **146**B includes an inductor **148**B and a variable capacitive structure **150**B. As such the stopband **66** is provided as a notch. The variable capacitive structure **150**B has a variable capacitance that is adjusted by the stopband tuning control output **108**. In this manner, the stopband **66** can be shifted into the RF communication band of the RF receive input signal **16**, the RF receive output signal **28**, the RF transmission input signal **32**, and the RF transmission output signal **36**.

[0110] With regards to providing the passband **62** of the second RF filter **38**B, the tunable RF filter path **144**B includes an embodiment of a resonator R(2,1) and an embodiment of a resonator R(2,2). The resonator R(2,1) and the resonator R(2,2) are weakly coupled to one another. More specifically, the resonator R(2,1) includes an inductor **152**B and a capacitive structure **154**B. The resonator R(2,2) includes an inductor **156**B and a capacitive structure **158**B.

[0111] The resonator R(2,1) and the resonator R(2,2) are a pair of weakly coupled resonators. The resonator R(2,1) and the resonator R(2,2) are weakly coupled by providing the inductor **152**B and the inductor **156**B such that the inductor **152**B and the inductor **156**B have a weak mutual coupling. Although the resonator R(2,1) and the resonator R(2,2) are weakly coupled, the inductor **156**B has a maximum lateral width and a displacement from the inductor **152**B that is less than or equal to half the maximum lateral width of the inductor **156**B. As such, the inductor **152**B and the inductor **156**B are relatively close to one another. The displacement between the inductor **152**B and the inductor **156**B may be measured from a geometric centroid of the inductor **152**B to a geometric centroid of the inductor **156**B. The maximum lateral width may be a maximum dimension of the inductor **156**B along a plane defined by its largest winding. The weak coupling between the inductor **152**B and the inductor **156**B is obtained through topological techniques. For example, the inductor **152**B and the inductor **156**B may be fully or partially aligned, where winding(s) of the inductor **152**B and winding(s) of the inductor **156**B are configured to provide weak coupling through cancellation. Alternatively or additionally, a plane defining an orientation of the winding(s) of the inductor **152**B and a plane defining an orientation of the winding(s) of the inductor **156**B may be fully or partially orthogonal to one another. Some of the magnetic couplings between the resonators R can be unidirectional (passive or active). This can significantly improve isolation (e.g., isolation in diplexers). In other embodiments, topological techniques are not used to provide weak mutual coupling between the inductor **152**B and the inductor **156**B. Rather, the inductor **152**B and the inductor **156**B are weakly mutually coupled by being sufficiently far apart.

[0112] To maximize the quality (Q) factor of the tunable RF filter paths **144**B through **68**, most of the total mutual coupling should be realized magnetically, and only fine-tuning is provided electrically. This also heLFS to reduce common-mode signal transfer in the differential resonators and thus keeps the Q factor high. While the magnetic coupling can be adjusted only statically, with a new layout design, the electric coupling can be tuned on the fly (after fabrication). The filter characteristics (e.g., bias network structure, resonator capacitance) can be adjusted based on given coupling coefficients to maximize filter performance.

[0113] To provide a tuning range to tune the passband **62** of the tunable RF filter path **144**B and provide a fast roll-off from a low-frequency side to a high-frequency side of the transfer function, the tunable RF filter path **144**B is configured to change a sign of a total mutual coupling coefficient between the resonator R(2,1) and the resonator R(2,2). Accordingly, the tunable RF filter path **144**B includes a cross-coupling capacitive structure C(BP1). As shown in FIG. **4**, the cross-coupling capacitive structure C(BP1) is electrically connected in the tunable RF filter path **144**B between the resonator R(2,1) and the resonator R(2,2) so as to provide a variable coupling coefficient between the resonator R(2,1) and the resonator R(2,2). The cross-coupling capacitive structure C(BP1) is a variable cross-coupling capacitive structure configured to vary the coupling coefficient provided between the resonator R(2,1) and the resonator R(2,2). The passband tuning control output **106** is provided to the cross-coupling capacitive structure C(BP1) to adjust the variable coupling coefficient and thereby shift the passband **62** into the RF communication band of the RF receive input signal **14**, the RF receive output signal **26**, the RF transmission input signal **30**, and the RF transmission output signal **34**.

[0114] In the resonator R(2,1), the inductor **152**B and the capacitive structure **154**B are electrically connected in parallel. More specifically, the inductor **152**B has an end **160**B and an end **162**B, which are disposed opposite to one another. The ends **160**B, **162**B are each electrically connected to the capacitive structure **154**B, which is grounded. Thus, the resonator R(2,1) is a single-ended resonator. The inductor **156**B is also electrically connected in parallel to the capacitive structure **158**B. The inductor **156**B has an end **164**B and an end **166**B, which are disposed opposite to one another. The ends **164**B, **166**B are each electrically connected to the capacitive structure **158**B, which is grounded.

[0115] As shown in FIG. **4**, the inductor **152**B and the inductor **156**B are mutually magnetically coupled so as to have a inverting mutual coupling and provide the differential phase inversion with the second RF filter **38**B. The inductor **152**B is magnetically coupled to the inductor **156**B such that an RF signal received at the end **160**B of the inductor **152**B with a voltage polarity (i.e., either a positive voltage polarity or a negative voltage polarity) results in a filtered RF signal being transmitted out the end **166**B of the inductor **156**B with the same voltage polarity. Also, the inductor **156**B is magnetically coupled to the inductor **152**B such that an RF signal received at the end **164**B of the inductor **156**B with a voltage polarity (i.e., either a positive voltage polarity or a negative voltage polarity) results in a filtered RF signal being transmitted out the end **162**B of the inductor **152**B with the same voltage polarity. This is indicated in FIG. **4** by the dot convention where a dot is placed at the end **160**B of the inductor **152**B and a dot is placed at the end **164**B of the inductor **156**B. Additionally, the inductors **152**B, **156**B may be magnetically coupled so as to have a low magnetic coupling coefficient through field cancellation, with the variable coupling coefficient. In this case, the inductor **152**B and the inductor **156**B are arranged such that the inductor **156**B reduces a mutual magnetic coupling coefficient of the inductor **152**B.

[0116] The resonator R(2,2) is operably associated with the resonator R(2,1) such that an energy transfer factor between the resonator R(2,1) and the resonator R(2,2) is less than 10%. Thus, the resonator R(2,1) and the resonator R(2,2) are weakly coupled. A total mutual coupling between the resonator R(2,1) and the resonator R(2,2) is provided by a sum total of the mutual magnetic factor between the resonator R(2,1) and the resonator R(2,2) and the mutual electric coupling coefficients between the resonator R(2,1) and the resonator R(2,2). In this embodiment, the mutual magnetic coupling coefficient between the inductor **152**B and the inductor **156**B is a fixed mutual magnetic coupling coefficient. Although embodiments of the resonators R(2,1), R(2,2) may be provided so as to provide a variable magnetic coupling coefficient between the resonators R(2,1), R(2,2), embodiments of the resonators R(2,1), R(2,2) that provide variable magnetic couplings can be costly and difficult to realize. However, providing variable electric coupling coefficients is easier and more economical. Thus, using the cross-coupling capacitive structure C(BP1) to provide the variable electric coupling coefficients is an economical technique for providing a tunable filter characteristic between the resonators R(2, 1), R(2,2). Furthermore, since the mutual magnetic coupling coefficient between the inductor **152**B and the inductor **156**B is fixed, the tunable RF filter path **144**B has lower insertion losses. A matching capacitor **168**B is connected in series with the sixth port **96** of the second hybrid coupler **20**. In this embodiment, the resonator R(2,1) and the resonator R(2,2) are single-ended resonators.

[0117] FIG. **5** illustrates S parameters for the tunable RF diplexer **10** shown in FIG. **4** where the tunable RF diplexer **10** has been tuned to center the passband **62** of the first RF filter **38**A and the second RF filter **38**B in the ISM frequency band centered around 2.44 GHz. The stopband **66** of the first RF filter **38**A and the second RF filter **38**B is centered in MB/HB frequency bands from 1.7 GHz to 2.9 GHz. An S(86, 88) parameter is an S parameter of the tunable RF diplexer **10** between the first port **86** and the second port **88**. The S(86,88) parameter shows that the transfer response of the tunable RF diplexer **10** remains at around 0 dB except around the ISM frequency band. Accordingly, the tunable RF diplexer **10** is configured to have small insertion losses in the MB/HB frequency bands between the first port **86** and the second port **88**. An S(86, 98) parameter is an S parameter of the tunable RF diplexer **10** between the first port **86** and the seventh port **98**. The S(86,98) parameter shows that the transfer response of the tunable RF diplexer **10** increases to around 0 dB in the ISM frequency band and decreases relatively quickly outside the ISM frequency band. Accordingly, the tunable RF diplexer **10** is configured to have small insertion losses in the ISM frequency bands between the first port **86** and the seventh port **98**. An S(88, 98) parameter is an S parameter of the tunable RF diplexer **10** between the second port **88** and the seventh port **98**. The S(88, 98) parameter is not explicitly shown because it remains below 50 dB and is −51.077 dB at the center frequency of the ISM band. The S(88,98) parameter shows that the transfer response of the tunable RF diplexer **10** provides improved isolation between the second port **88** for the MB/HB frequency bands and the seventh port **98** for the ISM frequency band due to the differential phase shift provided by the second RF filter **38**B. Accordingly, the performance of the tunable RF diplexer **10** is enhanced by the differential phase shift.

[0118] FIG. **6** illustrates another embodiment of the tunable RF diplexer **10** shown in FIG. **1**. The tunable RF diplexer **10** shown in FIG. **1** is similar to the embodiment shown in FIG. **4**. However, in this embodiment, the inductor **152**A of the resonator R(1,1) and the inductor **156**A of the resonator R(1,2) form a first single-ended transformer. The resonator R(1,1) and the resonator R(1,2) are weakly coupled and are single-ended resonators, as explained above with respect to FIG. **4**. Similarly, the inductor **152**B of the resonator R(2,1) and the inductor **156**B of the resonator R(2,2) also form a second single-ended transformer. The resonator R(2,1) and the resonator R(2,2) are weakly coupled, as explained above with respect to FIG. **4**. The second single-ended transformer provided by the inductor **152**B and the inductor **156**B provides the differential phase shift between the fourth port **92** of the first hybrid coupler **18** and the sixth port **96** of the second hybrid coupler **20**, as described above with respect to FIGS. **3**A-**3**D. Since the first single-ended transformer and the second single ended transformer are provided in the first RF filter **38**A and the second RF filter **38**B, the capacitive structure C(PA1) (shown in FIG. **4**) and the capacitive structure C(PB1) (shown in FIG. **4**) are not provided.

[0119] FIG. **7** illustrates S parameters for the tunable RF diplexer **10** shown in FIG. **6** where the tunable RF diplexer **10** has been provided to center the passband **62** of the first RF filter **38**A, and the second RF filter **38**B in the ISM frequency band is centered around 2.44 GHz. The stopband **66** of the first RF filter **38**A and the second RF filter **38**B is centered in MB/HB frequency bands from 1.7 GHz to 2.9 GHz. The S(86,88) parameter shows that the transfer response of the tunable RF diplexer **10** remains at around 0 dB except around the ISM frequency band. Accordingly, the tunable RF diplexer **10** is configured to have small insertion losses in the MB/HB frequency bands between the first port **86** and the second port **88**. The S(86,98) parameter shows that the transfer response of the tunable RF diplexer **10** increases to around 0 dB in the ISM frequency band and decreases relatively quickly outside the ISM frequency band. Accordingly, the tunable RF diplexer **10** is configured to have small insertion losses in the ISM frequency bands between the first port **86** and the seventh port **98**. The S(88, 98) remains below 40 dB and shows that the transfer response of the tunable RF diplexer **10** provides improved isolation between the second port **88** for the MB/HB frequency bands and the seventh port **98** for the ISM frequency band due to the differential phase shift provided by the second RF filter **38**B. Accordingly, the performance of the tunable RF diplexer **10** is enhanced by the differential phase shift.

[0120] FIG. **8** illustrates another embodiment of the tunable RF diplexer **10** shown in FIG. **1** and is similar to the embodiment illustrated in FIG. **4**. With respect to the first RF filter **38**A, the resonator R(1,1) and the resonator R(1,2) are weakly coupled resonators, as explained above. However, in this embodiment, the resonator R(1,1) is a differential resonator, and the resonator R(1,2) is a differential resonator. With regard to the resonator R(1,1) shown in FIG. **8**, the inductor **152**A has a grounded center tap so that the inductor **152**A has an inductor portion **170**A between the end **160**A and the center tap, and an inductor portion **172**A between the center tap and the end **162**A. In this embodiment, the capacitive structure **154**A is connected between the end **160**A of the inductor **152**A and ground. A capacitive structure **154**A' is connected between the end **162**A of the inductor **152**A and ground. The resonator R(1,1) thus forms a differential resonator.

[0121] With regard to the resonator R(1,2) shown in FIG. **8**, the inductor **156**A has a grounded center tap so that the inductor **156**A has an inductor portion **174**A between the end **164**A and the center tap, and an inductor portion **176**A between the center tap and the end **166**A. In this embodiment, the capacitive structure **158**A is connected between the end **164**A of the inductor **156**A and ground. A capacitive structure **158**A' is connected between the end **166**A of the inductor **156**A and ground. The resonator R(1,2) thus also forms a differential resonator.

[0122] The inductor **152**A of the resonator R(1,1) and the inductor **156**A of the resonator (1,2) form a differential transformer. The differential transformer provides no phase shift. More specifically, the inductor portion **170**A of the inductor **152**A is magnetically coupled to the inductor portion **174**A of the inductor **156**A. The inductor portion **172**A of the inductor **152**A is magnetically coupled to the inductor portion **176**A of the inductor **156**A. As such, a differential RF signal having a positive side signal received at the end **160**A of the inductor **152**A and a negative side being transmitted out of the end **162**A of the inductor **152**A results in a filtered differential RF signal having a positive side transmitted out the end **164**A of the inductor **156**A and a negative side transmitted into the end **166**A of the inductor **156**A. In this embodiment, the first RF filter **38**A includes an inductor **178**A that is coupled to the third port **90** of the first hybrid coupler **18**. The inductor **178**A is magnetically coupled to the inductor **154**A of the resonator R(1,1) so that the inductor **178**A and the inductor **154**A form a single-ended to differential transformer. Similarly, the first RF filter **38**A includes an inductor **180**A coupled to the fifth port **94**. The inductor **180**A is magnetically coupled to the inductor **156**A of the resonator R(1,2) so that the inductor **180**A and the inductor **156**A form another single-ended to differential transformer.

[0123] With respect to the second RF filter **38**B, the resonator R(2,1) and the resonator R(2,2) are weakly coupled resonators, as explained above. However, in this embodiment, the resonator R(2,1) is a differential resonator, and the resonator R(2,2) is a differential resonator. With regard to the resonator R(2,1) shown in FIG. **8**, the inductor **152**B has a grounded center tap so that the inductor **152**B has an inductor portion **170**B between the end **160**B and the center tap, and an inductor portion **172**B between the center tap and the end **162**B. In this embodiment, the capacitive structure **154**B is connected between the end **160**B of the inductor **152**B and ground. A capacitive structure **154**B' is connected between the end **162**B of the inductor **152**B and ground. The resonator R(2,1) thus forms a differential resonator.

[0124] With regard to the resonator R(2,2) shown in FIG. **8**, the inductor **156**B has a grounded center tap so that the inductor **156**B has an inductor portion **174**B between the end **164**B and the center tap, and an inductor portion **176**B between the center tap and the end **166**B. In this embodiment, the capacitive structure **158**B is connected between the end **164**B of the inductor **156**B and ground. A capacitive structure **158**B' is connected between the end **166**B of the inductor **156**B and ground. The resonator R(2,2) thus also forms a differential resonator.

[0125] The inductor **152**B of the resonator R(2,1) and the inductor **156**B of the resonator (2,2) form a differential transformer. The differential transformer provides the differential phase shift discussed above with respect to FIGS. **3**A-**3**D. More specifically, the inductor portion **170**B of the inductor **152**B is magnetically coupled to the inductor portion **176**B of the inductor **156**B. The inductor portion **172**B of the inductor **152**B is magnetically coupled to the inductor portion **174**B of the inductor **156**B. As such, a differential RF signal having a positive side signal received at the end **160**B of the inductor **152**B and a negative side being transmitted out of the end **162**B of the inductor **152**B results in a filtered differential RF signal having a positive side transmitted into the end **166**B of the inductor **156**B and a negative side transmitted out of the end **164**B of the inductor **156**B. Providing the differential phase shift with the differential transformer formed by the inductor **152**B and the inductor **156**B improves the performance of the tunable RF diplexer **10** shown in FIG. **9** since the differential transformer has less dependency on common mode inductances.

[0126] In this embodiment, the second RF filter **38**B includes an inductor **178**B that is coupled to the fourth port **92** of the first hybrid coupler **18**. The inductor **178**B is magnetically coupled to the inductor **152**B of the resonator R(2,1) so that the inductor **178**B and the inductor **152**B form a single-ended to differential transformer. Similarly, the second RF filter **38**B includes an inductor **180**B coupled to the sixth port **96** of the second hybrid coupler **20**. The inductor **180**B is magnetically coupled to the inductor **156**B of the resonator R(2,2) so that the inductor **180**B and the inductor **156**B form another single-ended to differential transformer.

[0127] FIG. **9** illustrates another embodiment of the tunable RF diplexer **10** shown in FIG. **1**, which is the same as the embodiment shown in FIG. **8** except in the manner that the second RF filter **38**B provides the differential phase shift. Unlike the embodiment shown in FIG. **8**, the differential transformer formed through the weak magnetic coupling of the inductor **152**B of the resonator R(2,1), and the inductor **156**B of the resonator R(2,2) shown in FIG. **9** provides approximately no phase shift. In this embodiment, the differential transformer is provided by the inductor **152**B of the resonator R(2,1) and the inductor **156**B of the resonator R(2, 2). Accordingly, in this embodiment, the inductor portion **170**B of the inductor **152**B is magnetically coupled to the inductor portion **174**B of the inductor **156**B. The inductor portion **172**B of the inductor **152**B is magnetically coupled to the inductor portion **176**B of the inductor **156**B. As such, a differential RF signal having a positive side signal received at the end **160**B of the inductor **152**B and a negative side being transmitted out of the end **162**B of the inductor **152**B results in a filtered differential RF signal having a positive side transmitted out the end **164**B of the inductor **156**B and a negative side transmitted into the end **166**B of the inductor **156**B.

[0128] Instead, the differential phase shift is provided by the single to differential transformer formed by the inductor **156**B of the resonator R(2,2) and the inductor **180**B. More specifically, the inductor **156**B and the inductor **180**B are magnetically coupled such that the single to differential transformer provides the differential phase shift, as shown in FIG. **9**. In alternative embodiments, the single to differential transformer formed by the inductor **178**B and the inductor **152**B of the resonator R(1,1) provide the differential shift. More specifically, in these alternative embodiments, the inductor **178**B and the inductor **152**B of the resonator R(1,1) are magnetically coupled such that the single to differential transformer provides the differential phase shift.

[0129] FIG. **10** illustrates another embodiment of a tunable RF diplexer **10**A. The tunable RF diplexer **10**A provides diplexing to the RF input receive signal **14**, the RF receive input signal **16**, the RF receive output signal **26**, the RF receive output signal **28**, the RF transmission input signal **30**, the RF transmission input signal **32**, the RF transmission output signal **34**, and the RF transmission output signal **36** in the same manner described above with respect to FIGS. **1**, **2**A-**2**B, and **3**A-**3**D, except that in this embodiment the second RF filter **38**B does not provide the differential phase shift. Instead, the phase inversion component is provided by a phase shifter **182**. Thus, while the second RF filter **38**B in FIG. **10** provides no phase shift, the phase shifter **182** shown in FIG. **10** provides a differential phase shift. The phase shifter **182** is thus the phase inversion component. In this embodiment, the phase shifter **182** is configured to provide a positive differential phase shift of approximately +180 degrees or π radians.

[0130] As shown in FIG. **10**, the phase shifter **182** is connected between the fourth port **92** of the first hybrid coupler **18** and the second RF filter **38**B of the RF filter circuit **22**. Using the phase shifter **182** to provide the differential phase shift (e.g., the positive differential phase shift) is advantageous since the phase shifter **182** can be included to provide the differential phase shift over a wide frequency range. Thus, the tunable RF diplexer **10**A in FIG. **10** is operable to provide broadband isolation from the seventh port **98** to the second port **88**.

[0131] FIG. **11** illustrates another embodiment of the tunable RF diplexer **10**A. The tunable RF diplexer **10**A shown in FIG. **11** is the same as the embodiment shown in FIG. **11**, except in this embodiment, the phase shifter provides a negative differential phase shift of approximately 180 degrees or 1π radians. The phase shifter **182** is thus the differential phase shifting component. The tunable RF diplexer **10**A in FIG. **11** is also operable to provide broadband isolation from the seventh port **98** to the second port **88**.

[0132] FIG. **12** illustrates another embodiment of a tunable RF diplexer **10**B. The tunable RF diplexer **10**B provides diplexing to the RF input receive signal **14**, the RF receive input signal **16**, the RF receive output signal **26**, the RF receive output signal **28**, the RF transmission input signal **30**, the RF transmission input signal **32**, the RF transmission output signal **34**, and the RF transmission output signal **34** in the same manner described above with respect to FIGS. **1**, **2**A-**2**B, and **3**A-**3**D, except that in this embodiment the second RF filter **38**B also does not provide the differential phase shift. Instead, the phase inversion component is provided as a single to differential transformer connected to the second RF filter **38**B, as explained in further detail below.

[0133] In this embodiment, a single to differential transformer **184**A is connected between the third port **90** and the first RF filter **38**A. The single to differential transformer **184**A includes an inductor **186**A connected to the third port **90**. An inductor **188**A is magnetically coupled to the inductor **186**A has a grounded center tapped. The single to differential transformer **184**A shown in FIG. **12** does not provide a differential phase shift. In this embodiment, the inductor **186**A is connected to the first RF filter **38**A. Another single to differential transformer **190**A is connected between the first RF filter **38**A and the fifth port **94** of the second hybrid coupler **20**. The single to differential transformer **190**A includes an inductor **192**A connected to the first RF filter **38**A. An inductor **194**A is magnetically coupled to the inductor **192**A and has a grounded center tap. The inductor **194**A is connected to the fifth port **94** of the second hybrid coupler **20**. The single to differential transformer **190**A shown in FIG. **12** does not provide a phase shift.

[0134] Also in this embodiment, a single to differential transformer **184**B is connected between the fourth port **92** of the first hybrid coupler **18** and the second RF filter **38**B. The single to differential transformer **184**B includes an inductor **186**B connected to the fourth port **92**. An inductor **188**B is magnetically coupled to the inductor **186**B has a grounded center tapped. The single to differential transformer **184**B shown in FIG. **12** provides a differential phase shift and is thus the phase inversion component. In this embodiment, the inductor **186**B is connected to the second RF filter **38**B such that there is an inversion of the differential polarity provided to the second RF filter **38**B. As such, the single to differential transformer **184**B provides the differential phase shift and is the phase inversion component. Another single to differential transformer **190**B is connected between the second RF filter **38**B and the sixth port **96**. The single to differential transformer **190**B includes an inductor **192**B connected to the second RF filter **38**B. An inductor **194**B is magnetically coupled to the inductor **192**B and has a grounded center tap. The inductor **194**B is connected to the sixth port **96** of the second hybrid coupler **20**. The single to differential transformer **190**B shown in FIG. **12** does not provide a phase shift. In alternative embodiment, the single to differential transformer **190**B is provided as the phase inversion component to provide the differential phase shift. For example, the inductor **192**B may be connected to the second RF filter **38**B such that there is an inversion of the differential polarity provided from the second RF filter **38**B.

[0135] FIG. **13** illustrates a circuit diagram of another embodiment of the tunable RF diplexer **10** shown in FIG. **1** and FIGS. **3**A-**3**D, wherein the tunable RF diplexer **10** shown in FIG. **13** is the same as the embodiment shown in FIG. **4**, except in this embodiment, the tunable RF diplexer **10** further includes a cross-coupling capacitive structure C(AP2), a cross-coupling capacitive structure C(BP2), a cross-coupling capacitive structure CC(A1), a cross-coupling capacitive structure CC(A2), a cross-coupling capacitive structure CC(B1), and a cross-coupling capacitive structure CC(B2). As explained in further detail below, by providing the cross-coupling capacitive structure CC(A2), the cross-coupling capacitive structure CC(B1), and the cross-coupling capacitive structure CC(B2), the RF filter circuit **22** is tunable to provide the stopband **66**, which in this embodiment is a notch, on both a high-frequency side of the passband **62** and a low-frequency side of the passband **62**. This can resolve problems that may result when the RF communication band **44** for operation at the second port **88** gets too close to the RF communication band **48** of operation at the seventh port **98**.

[0136] For example, when the RF communication band **44** is a B7 transmission frequency band centered around 2.504 GHz and the RF communication band **48** is the ISM frequency band centered around 2.44 GHz, it becomes more difficult for the tunable RF circuit **22** to pass and reflect appropriately as described in FIGS. **3**A-**3**D. This can result in the tunable RF circuit **22** experiencing high insertion losses. Additionally, this can also result in degradation of isolation in the tunable RF diplexer **10**. To help correct this and provide better performance, the tuning circuit **69** is configured to tune the RF filter circuit **22** to switch the stopband **66** (e.g., in this case, the notch) (shown in FIG. **2**D) from the high-frequency side of the passband **62** (shown in FIG. **2**D) to the low-frequency side of the passband **62** as a result of the RF communication band **48** (shown in FIG. **2**D) for operation at the second port **88** being changed from HB frequency bands to MB frequency bands that are close to the ISM frequency band, such as the B7 transmission frequency band. This allows for the tunable RF diplexer **10** shown in FIG. **13** to have less insertion loss and better isolation when the RF communication band **48** is in the B7 transmission frequency band. The tuning circuit **69** is configured also to tune the RF filter circuit **22** to switch the stopband **66** (e.g., in this case, the notch) from the low-frequency side of the passband **62** to the high-frequency side of the passband **62** as a result of the RF communication band **48** for operation at the second port **88** being changed from MB frequency bands that are close to the ISM frequency band to HB frequency bands. The tuning circuit **69** is operable to switch the stopband **66** (e.g., in this case, the notch) from to and from the low-frequency side and the high-frequency side of the passband **62** by adjusting variable capacitances in the cross-coupling capacitive structure CC(A2), the cross-coupling capacitive structure CC(B1), and the cross-coupling capacitive structure CC(B2), as explained in further detail below.

[0137] With respect to the RF filter **38**A shown in FIG. **13**, the tunable RF filter path **144**A further includes the cross-coupling capacitive structure C(AP2). As shown in FIG. **13**, the cross-coupling capacitive structure C(AP2) is electrically connected in the tunable RF filter path **144**A between the resonator R(1,1) and the resonator R(1,2) so as to provide a variable coupling coefficient between the resonator R(1,1) and the resonator R(1,2). The cross-coupling capacitive structure C(AP2) is a variable cross-coupling capacitive structure configured to vary the coupling coefficient provided between the resonator R(1,1) and the resonator R(1,2). The passband tuning control output **106** is provided to the cross-coupling capacitive structure C(AP2) to adjust the variable coupling coefficient and thereby shift the passband **62** into the RF communication band of the RF receive input signal **14**, the RF receive output signal **26**, the RF transmission input signal **30**, and the RF transmission output signal **34**. As described above, the resonator R(1,1) and the resonator R(1,2) are connected to provide the passband **62** of the tunable RF filter path **144**A, and thus a bandpass filter is formed with the resonator R(1,1) and the resonator R(1,2). The stopband **66** of the RF filter **38**A in this case is a notch provided by the parallel resonator **146**A connected in series at the third port **90**. Accordingly, the parallel resonator **146**A is a notch filter.

[0138] The RF filter **38**A is tunable to provide the stopband **66** (e.g., in this case, the notch) of the RF filter **38**A on both a high-frequency side of the passband **62** and a low-frequency side of the passband **62**. As in FIG. **4**, the RF filter **38**A in FIG. **13** is provided by the tunable RF filter path **144**A. In order for the RF filter **38**A to be tunable to provide the stopband **66** (e.g., in this case, the notch) of the RF filter **38**A on both the high-frequency side of the passband **62** and the low-frequency side of the passband **62**, the RF filter **38**A includes the cross-coupling capacitive structure CC(A1) and the cross-coupling capacitive structure CC(A2). The cross-coupling capacitive structure CC(A1) has a variable capacitance, and the cross-coupling capacitive structure CC(A2) also has a variable capacitance. The stopband tuning control output **108** is provided to the cross-coupling capacitive structure CC(A1) and the cross-coupling capacitive structure CC(A2) to adjust their variable capacitances and so that the stopband **66** (e.g., in this case, the notch) is transposed to and from the low-frequency side and the high frequency side of the passband **62** provided by the RF filter **38**A. In essence, the cross-coupling capacitive structure CC(A1) and the cross-coupling capacitive structure CC(A2) are connected to introduce factors into an equation describing the relationship between the resonators R(1,1), R(1,2) and the variable coupling coefficients provided by the capacitive structures C(AP1), C(AP2) that allow for the stopband **66** (e.g., in this case, the notch) to be switched to and from the high-frequency side and the low-frequency side of the passband **62** of the tunable RF filter path **144**A.

[0139] Accordingly, the cross-coupling capacitive structure CC(A1) and the cross-coupling capacitive structure CC(A2) are coupled such that the tunable RF filter path **144**A is tunable to provide the stopband **66** (e.g., in this case, the notch) of the RF filter **38**A on both the high-frequency side of the first passband and the low-frequency side of the first passband by varying the first variable capacitance and the second variable capacitance. In this embodiment, the cross-coupling capacitive structure CC(A1) is electrically connected from the third port **90** to the resonator R(2,2). The resonator R(2,2) in the tunable RF filter path **144**A is connected closest to the fifth port **94** of the second hybrid coupler **20**. Furthermore, the cross-coupling capacitive structure CC(A2) is electrically connected from the fifth port **94** to the resonator R(1,1). The resonator R(1,1) in the tunable RF filter path **144**A is connected closest to the third port **90** of the first hybrid coupler **18**.

[0140] With respect to the RF filter **38**B shown in FIG. **13**, the tunable RF filter path **144**B further includes a cross-coupling capacitive structure C(BP2). As shown in FIG. **13**, the cross-coupling capacitive structure C(BP2) is electrically connected in the tunable RF filter path **144**B between the resonator R(2,1) and the resonator R(2,2) so as to provide a variable coupling coefficient between the resonator R(2,1) and the resonator R(2,2). The cross-coupling capacitive structure C(BP2) is a variable cross-coupling capacitive structure configured to vary the coupling coefficient provided between the resonator R(2,1) and the resonator R(2,2). The passband tuning control output **106** is provided to the cross-coupling capacitive structure C(BP2) to adjust the variable coupling coefficient and thereby shift the passband **62** into the RF communication band of the RF receive input signal **14**, the RF receive output signal **26**, the RF transmission input signal **30**, and the RF transmission output signal **34**. As described above, the resonator R(2,1) and the resonator R(2,2) are connected to provide the passband **62** of the tunable RF filter path **144**B, and thus a bandpass filter is formed with the resonator R(2,1) and the resonator R(2,2). The stopband **66** of the RF filter **38**B in this case is a notch provided by the parallel resonator **146**B connected in series at the fourth port **92**. Accordingly, the parallel resonator **146**B is a notch filter.

[0141] The RF filter **38**B is tunable to provide the stopband **66** (e.g., in this case, the notch) of the RF filter **38**B on both a high-frequency side of the passband **62** and a low-frequency side of the passband **62**. As in FIG. **4**, the RF filter **38**B in FIG. **13** is provided by the tunable RF filter path **144**B. In order for the RF filter **38**B to be tunable to provide the stopband **66** (e.g., in this case, the notch) of the RF filter **38**B on both the high-frequency side of the passband **62** and the low-frequency side of the passband **62**, the RF filter **38**B includes the cross-coupling capacitive structure CC(B1) and the cross-coupling capacitive structure CC(B2). The cross-coupling capacitive structure CC(B1) has a variable capacitance, and the cross-coupling capacitive structure CC(B2) also has a variable capacitance. The stopband tuning control output **108** is provided to the cross-coupling capacitive structure CC(B1) and the cross-coupling capacitive structure CC(B2) to adjust their variable capacitances and so that the stopband **66** (e.g., in this case, the notch) is transposed to and from the low-frequency side and the high frequency side of the passband **62** provided by the RF filter **38**B. In essence, the cross-coupling capacitive structure CC(B1) and the cross-coupling capacitive structure CC(B2) are connected to introduce factors into an equation describing the relationship between the resonators R(2,1), R(2,2) and the variable coupling coefficients provided by the capacitive structures C(BP1), C(BP2) that allow for the stopband **66** (e.g., in this case, the notch) to be switched to and from the high-frequency side and the low-frequency side of the passband **62** of the tunable RF filter path **144**B.

[0142] Accordingly, the cross-coupling capacitive structure CC(B1) and the cross-coupling capacitive structure CC(B2) are coupled such that the tunable RF filter path **144**B is tunable to provide the stopband **66** (e.g., in this case, the notch) of the RF filter **38**B on both the high-frequency side of the first passband and the low-frequency side of the first passband by varying the first variable capacitance and the second variable capacitance. In this embodiment, the cross-coupling capacitive structure CC(B1) is electrically connected from the fourth port **92** to the resonator R(2,2). The resonator R(2,2) in the tunable RF filter path **144**B is connected closest to the sixth port **96** of the second hybrid coupler **20**. Furthermore, the cross-coupling capacitive structure CC(B2) is electrically connected from the sixth port **96** to the resonator R(2,1). The resonator R(2,1) in the tunable RF filter path **144**B is connected closest to the fourth port **92** of the first hybrid coupler **18**.

[0143] FIG. **14** illustrates different exemplary frequency response curves **60(1)**, **60(2)**, **60(3)** for the frequency response **60** provided by the RF filter circuit **22** shown in FIG. **13**. In FIG. **14**, the frequency response **60** (and the frequency response curves **60(1)**, **60(2)**, **60(3)**) represent S21 and S12 parameters of each of the RF filters **38**A, **38**B (shown in FIG. **13**) individually as a function of frequency. Thus, the two-port S21 and S12 parameter of the first RF filter **38**A is represented by the frequency response **60** (and the frequency response curves **60(1)**, **60(2)**, **60(3)**) in FIG. **14**. Since the second RF filter **38**B is identical to the first RF filter **38**A, the two-port S21 and S12 parameter of the second RF filter **38**B is also represented by the frequency response **60** (and the frequency response curves **60(1)**, **60(2)**, **60(3)**) in FIG. **14**. The two-port S21 and S12 parameter represents the forward and reverse transmission of the RF filters **38**A, **38**B, as a function of frequency.

[0144] In FIG. **14**, the frequency response curve **60(1)** is the frequency response **60** of the RF filter circuit **22** when the variable capacitances of each of the cross-coupling capacitive structures CC(A1), CC(A2), CC(B1), CC(B2) are set to 50 fF. The frequency response curve **60(2)** is the frequency response **60** of the RF filter circuit **22** when the variable capacitances of each of the cross-coupling capacitive structures CC(A1), CC(A2), CC(B1), CC(B2) are set to 30 fF. Finally, the frequency response curve **60(3)** is the frequency response **60** of the RF filter circuit **22** when the variable capacitances of each of the cross-coupling capacitive structures CC(A1), CC(A2), CC(B1), CC(B2) are set to 10 fF. In this example of the frequency response curves **60(1)**, **60(2)**, **60(3)**, the notch **66** is provided on a high-frequency side HFS of the passband **62** because the notch **66** is provided as the passband **62** rolls off on the side provided at higher frequencies. With regard to the frequency response **60**, the high-frequency side HFS may be considered the side at higher frequencies than the maxima **64** of the passband **62**. A low-frequency side LFS of the passband **62** is the side of the passband **62** that rolls off at lower frequencies. With regard to the frequency response **60**, the low-frequency side LFS may be considered the side at lower frequencies than the maxima **64** of the passband **62**.

[0145] Note from the frequency response curves **60(1)**, **60(2)**, **60(3)**, the notch **66** is transposed closer to the low-frequency side LPF as the variable capacitances of the cross-coupling capacitive structures CC(A1), CC(A2), CC(B1), CC(B2) increases. The notch **66** also becomes shallower as the variable capacitances of the cross-coupling capacitive structures CC(A1), CC(A2), CC(B1), CC(B2) increases. The variable capacitances of the cross-coupling capacitive structures CC(A1), CC(A2), CC(B1), CC(B2) can be continued to be increased until the notch **66** is provided on the low-frequency side LFS of the passband **62**.

[0146] For example, the tuning circuit **69** may be configured to place the notch **66** of each of the RF filters **38**A, **38**B on the high-frequency side HPF when an RF communication band **48**A (e.g., an embodiment of the RF communication band **48** shown in FIG. **2**D) is selected as a HB frequency band in order to reduce insertion loss and improve isolation. The RF communication band **48**A is at a higher frequency range than the RF communication band **44**, which in this example is the ISM frequency band. However, when an RF frequency band **48**B (e.g., an embodiment of the RF communication band **48** shown in FIG. **2**D) is selected is selected as a MB frequency band, the tuning circuit **69** may be configured to place the notch **66** on the low-frequency side LFS of the passband **62** in order to reduce insertion loss and improve isolation and improve isolation. As explained with respect to FIG. **13**, the tuning circuit **69** sets the variable capacitances of the cross-coupling capacitive structures CC(A1), CC(A2), CC(B1), CC(B2) so that the notch **66** is on the high-frequency side HFS and the low-frequency side LFS of the passband **62** of each of the RF filters **38**A, **38**B.

[0147] FIG. **15** illustrates S parameters for the tunable RF diplexer **10** shown in FIG. **13** where the tunable RF diplexer **10** has been provided to center the passband **62** of the first RF filter **38**A, and the second RF filter **38**B in the ISM frequency band is centered around a center frequency of 2.44 GHz. The notch **66** (shown in FIG. **14**) of the first RF filter **38**A and the second RF filter **38**B can be centered in MB/HB frequency bands with center frequencies from 1.7 GHz to 2.9 GHz. In particular, the B7 transmission frequency band is centered at a center frequency of 2.504 GHz. The S(86,88) parameter shows that the transfer response of the tunable RF diplexer **10** remains at around 0 dB except around the ISM frequency band. Accordingly, the tunable RF diplexer **10** is configured to have small insertion losses in the MB/HB frequency bands between the first port **86** and the second port **88**. The S(86,98) parameter shows that the transfer response of the tunable RF diplexer **10** increases to around 0 dB in the ISM frequency band and decreases relatively quickly outside the ISM frequency band. Accordingly, the tunable RF diplexer **10** is configured to have small insertion losses in the ISM frequency bands between the first port **86** and the seventh port **98**. The insertion losses at the center frequency of the B7 transmission frequency band are reduced and the greater isolation is provided at higher HB frequency bands due to the extra filtering provided on the high-frequency side HPF. Note if an MB frequency band is selected that is on the low-frequency side LFS (i.e. below the center frequency of 2.4 GHz of the ISM frequency band), the tuning circuit **69** is configured to adjust the variable capacitances of the cross-coupling capacitive structures CC(A1), CC(A2), CC(B1), CC(B2) so that the notch **66** (shown in FIG. **14**) is placed on the low-frequency side LFS of the passband **62**. Otherwise if an MB/HB frequency band is selected that is on the high-frequency side HPF, the tuning circuit **69** is configured to adjust the variable capacitances of the cross-coupling capacitive structures CC(A1), CC(A2), CC(B1), CC(B2) so that the notch **66** (shown in FIG. **14**) is placed on the high-frequency side HFS. In this manner, less insertion loss and greater isolation is provided regardless of whether the MB/HB frequency band is on the high-frequency side HFS or the low-frequency side LFS.

[0148] Those skilled in the art will recognize improvements and modifications to the preferred embodiments of the present disclosure. All such improvements and modifications are considered within the scope of the concepts disclosed herein and the claims that follow.

What is claimed is:

1. A radio frequency (RF) diplexer comprising:

a first hybrid coupler;

a second hybrid coupler;

a RF filter circuit connected between the first hybrid coupler and the second hybrid coupler, wherein the RF filter circuit defines a passband and a notch, the RF filter circuit is tunable to provide the notch on both a high-frequency side of the passband and a low-frequency side of the passband; and a phase inversion component connected between the first hybrid coupler and the second hybrid coupler and configured to provide approximately a differential phase shift.

2. The RF diplexer of claim **1** wherein:
the RF filter circuit comprises a first RF filter and a second RF filter;
the first RF filter is connected between the first hybrid coupler and the second hybrid coupler; and
the second RF filter comprises is connected between the first hybrid coupler and the second hybrid coupler, wherein the second RF filter comprises the phase inversion component.

3. The RF diplexer of claim **2** wherein the first RF filter defines a tunable RF filter path, the tunable RF filter path comprises:
at least a pair of weakly coupled resonators;
a first cross-coupling capacitive structure having a first variable capacitance and connected from the first hybrid coupler to one of the pair of weakly coupled resonators;
a second cross-coupling capacitive structure having a second variable capacitance and connected from the second hybrid coupler to an other one of the pair of weakly coupled resonators.

4. The RF diplexer of claim **3** wherein the tunable RF filter path further comprises a notch filter.

5. The RF diplexer of claim **2** wherein the second RF filter defines a tunable RF filter path, the tunable RF filter path comprises:
at least a pair of weakly coupled resonators, wherein the pair of weakly coupled resonators form the phase inversion component;
a first cross-coupling capacitive structure having a first variable capacitance and connected from the first hybrid coupler to one of the pair of weakly coupled resonators;
a second cross-coupling capacitive structure having a second variable capacitance and connected from the second hybrid coupler to an other one of the pair of weakly coupled resonators.

6. The RF diplexer of claim **5** wherein the tunable RF filter path further comprises a notch filter.

7. The RF diplexer of claim **2** wherein:
the first RF filter defines a first tunable RF filter path, the first tunable RF filter path comprises:
at least a first pair of weakly coupled resonators;
a first cross-coupling capacitive structure having a first variable capacitance and connected from the first hybrid coupler to one of the first pair of weakly coupled resonators;
a second cross-coupling capacitive structure having a second variable capacitance and connected from the second hybrid coupler to an other one of the first pair of weakly coupled resonators;
the second RF filter defines a second tunable RF filter path, the second tunable RF filter path comprises:
at least a second pair of weakly coupled resonators, wherein the second pair of weakly coupled resonators form the phase inversion component;
a third cross-coupling capacitive structure having a third variable capacitance and connected from the first hybrid coupler to one of the second pair of weakly coupled resonators;
a fourth cross-coupling capacitive structure having a fourth variable capacitance and connected from the second hybrid coupler to an other one of the second pair of weakly coupled resonators.

8. The RF diplexer of claim **7** wherein the first cross-coupling capacitive structure, the second cross-coupling capacitive structure, the third cross-coupling capacitive structure, and the fourth cross-coupling capacitive structure are coupled such that the RF filter circuit is tunable to provide the notch on both the high-frequency side of the passband and the low-frequency side of the passband by adjusting the first variable capacitance, the second variable capacitance, the third variable capacitance and the fourth variable capacitance.

9. The RF diplexer of claim **8** wherein the first tunable RF filter path further comprises a first notch filter and the second tunable RF filter path further comprises a second notch filter.

10. The RF diplexer of claim **8** wherein:
the first hybrid coupler comprises:
a first port configured to be coupled to an antenna;
a second port configured to be coupled to external circuitry;
a third port coupled to the first RF filter, wherein the first cross-coupling capacitive structure is connected from the third port to the one of the first pair of weakly coupled resonators;
a fourth port coupled to the second RF filter, the third cross-coupling capacitive structure is connected from the fourth port to the one of the second pair of weakly coupled resonators;
the second hybrid coupler comprises:
a fifth port coupled to the first RF filter, wherein the second cross-coupling capacitive structure is connected from the fifth port to the other one of the first pair of weakly coupled resonators;
a sixth port coupled to the second RF filter, wherein the fourth cross-coupling capacitive structure is connected from the sixth port to the other one of the second pair of weakly coupled resonators;
a seventh port coupled to other external circuitry; and
an eighth port configured to be coupled to an impedance termination.

11. A radio frequency (RF) diplexer comprising:
a first hybrid coupler;
a second hybrid coupler;
a first tunable RF filter path comprising at least a first pair of weakly coupled resonators, where the first tunable RF filter path is connected between the first hybrid coupler and the second hybrid coupler and the first tunable RF filter path defines a first passband and a first notch, the first tunable RF filter path is tunable to provide the first notch on both a high-frequency side of the first passband and a low-frequency side of the first passband;
a phase inversion component connected between the first hybrid coupler and the second hybrid coupler and configured to provide approximately a differential phase shift.

12. The RF diplexer of claim **11** further comprising a second tunable RF filter path comprising at least a second pair of weakly coupled resonators, where the second tunable RF filter path is connected between the first hybrid coupler and the second hybrid coupler and the second tunable RF filter path defines a second passband and a second notch, the second tunable RF filter path is tunable to provide the second notch on both a high-frequency side of the second passband and a low-frequency side of the second passband.

13. The RF diplexer of claim **12** wherein the second tunable RF filter path forms the phase inversion component with the second pair of weakly coupled resonators.

14. The RF diplexer of claim **12** wherein the first tunable RF filter path further comprises:

the first hybrid coupler comprises:
  a first port configured to be coupled to an antenna;
  a second port configured to be coupled to external circuitry;
  a third port coupled to the first tunable RF filter path;
  a fourth port coupled to the second tunable RF filter path;
the second hybrid coupler comprises:
  a fifth port coupled to the first tunable RF filter path;
  a sixth port coupled to the second tunable RF filter path,
  a seventh port coupled to other external circuitry; and
  an eighth port configured to be coupled to an impedance termination.

15. The RF diplexer of claim **14** wherein:

the first tunable RF filter path further comprises:
  a first cross-coupling capacitive structure having a first variable capacitance and connected from the third port to one of the first pair of weakly coupled resonators;
  a second cross-coupling capacitive structure having a second variable capacitance and connected from the fifth port to an other one of the first pair of weakly coupled resonators;
the second tunable RF filter path further comprises:
  a third cross-coupling capacitive structure having a third variable capacitance and connected from the fourth port to one of the second pair of weakly coupled resonators;
  a fourth cross-coupling capacitive structure having a fourth variable capacitance and connected from the sixth port to an other one of the second pair of weakly coupled resonators.

16. The RF diplexer of claim **15** wherein:

the first tunable RF filter path further comprises:
  a first bandpass filter formed with the first pair of weakly coupled resonators and configured to provide the first passband;
  a first notch filter connected in series at the third port and configured to provide the first notch;
the second tunable RF filter path further comprises:
  a second bandpass filter formed with the second pair of weakly coupled resonators and configured to provide the second passband;
  a second notch filter connected in series at the fourth port and configured to provide the second notch.

17. The RF diplexer of claim **16** wherein:

the first cross-coupling capacitive structure and the second cross-coupling capacitive structure are coupled such that the first tunable RF filter path is tunable to provide the first notch on both the high-frequency side of the first passband and the low-frequency side of the first passband by varying the first variable capacitance and the second variable capacitance;
the third cross-coupling capacitive structure and the fourth cross-coupling capacitive structure are coupled such that the second tunable RF filter path is tunable to provide the second notch on both the high-frequency side of the second passband and the low-frequency side of the second passband by varying the third variable capacitance and the fourth variable capacitance.

18. The RF diplexer of claim **11** wherein the first tunable RF filter path forms the phase inversion component with the first pair of weakly coupled resonators.

19. The RF diplexer of claim **18** wherein the first tunable RF filter path further comprises:

a first cross-coupling capacitive structure having a first variable capacitance and connected from the first hybrid coupler to one of the first pair of weakly coupled resonators;
a second cross-coupling capacitive structure having a second variable capacitance and connected from the second hybrid coupler to an other one of the first pair of weakly coupled resonators.

20. The RF diplexer of claim **19** wherein the first cross-coupling capacitive structure and the second cross-coupling capacitive structure are coupled such that the first tunable RF filter path is tunable to provide the first notch on both the high-frequency side of the first passband and the low-frequency side of the first passband by varying the first variable capacitance and the second variable capacitance.

* * * * *

G+ More Next Blog» Create Blog Sign In

wireless communication t[illegible]y

Thursday, 9 July 2015

RF Energy Harvesting System from Cell Towers in 900MHz Band :

Available RF energy in the ambient or areas close to transmission towers provides an opportunity to harvest that energy. Some of the most prominent sources are FM radio systems ((88-108 MHz, transmitted power few tens of KW), TV Transmission (180-220 MHz, transmitted power few tens of KW), Cell Tower Transmission (10 to 20 W per carrier), Wi-Fi (2.45GHz, 5.8GHz), AM Transmission (540-1600 KHz, transmitted power few hundred KW) and mobile phones (transmitted power 1W to 2W), etc.

Cell towers can be used as a continuous source of renewable energy as they transmit 24 hours. In India cell towers transmit in the frequency range of 869- 890 MHz in CDMA, 935-960 MHz in GSM 900 and 1810-1880MHz in GSM 1800 bands.
It transmits 10 to 20W per carrier; there maybe 3 to 4 carriers and 3 to 4 operators on a single tower or spread over the roof top of buildings. Gain of the cell tower transmitter antenna is typically 17dB. The half power beam width (HPBW) of the antenna in horizontal direction maybe between 60o to 90o and in vertical direction varies between 5oto 10o. Maximum power is received when the receiver is in the main beam. For cell site consisting of transmitting towers of GSM900 band, signal strengths are calculated in Table I at various distances according to Friis transmission equation.

Pr = PtGtGr[=(4 R)]^2

Where,
Pr = Received Power
Pt= Transmitted Power
Gt= Gain of the transmitted Antenna
Gr= Gain of the receiver Antenna
R= Distance between the transmitter and receiver antennas

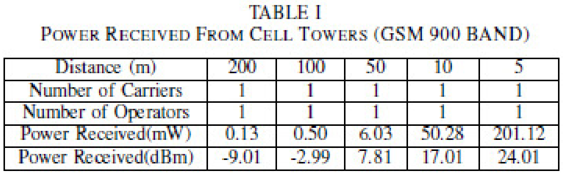

TABLE I
POWER RECEIVED FROM CELL TOWERS (GSM 900 BAND)

Distance (m)	200	100	50	10	5
Number of Carriers	1	1	1	1	1
Number of Operators	1	1	1	1	1
Power Received(mW)	0.13	0.50	6.03	50.28	201.12
Power Received(dBm)	-9.01	-2.99	7.81	17.01	24.01

The gain of the cell tower transmitter antenna is 17dB. Receiver antenna gain is taken as 9dB according to the fabricated microstrip antenna gain. Transmitted power is 20W from the cell towers. Frequency is taken as 950MHz which is approximately middle value of GSM 900 band.
For full signal strength, a cell phone requires only -69dBm power. In a radius within 50m from a cell tower, power level is very high. Such high power levels cause various health hazards, such as, headache, memory loss, nausea, dizziness, tingling, altered reflexes, muscle and joint pain, and ultimately leading to cancer [1].
There are more than 4 lakhs cell towers in India, therefore harvesting ambient RF energy would provide an alternate energy source for various applications and reduce radiation

About Me

ravigandhi arul

G+ Follow 14

View my complete profile

Blog Archive

- ▼ 2015 (1)
  - ▼ July (1)
    - RF Energy Harvesting System from Cell Towers in 90...

health hazards. Fig.1 shows the basic block diagram of RF Energy harvesting system.

However, harvesting energy from them poses several challenges.

1. Available power varies with distance, direction and gain of the receiver antenna; therefore high gain antenna for all frequency bands would be required.

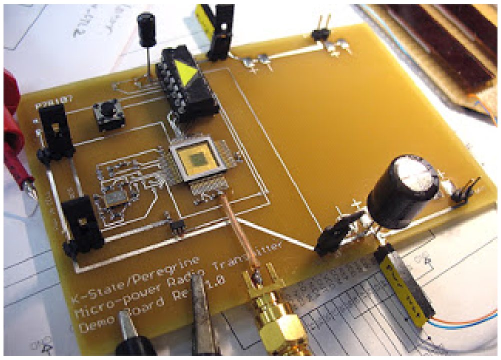

2. Due to non-linear dependence of the rectifier impedance on the frequency and power, broadband impedance matching network is essential for maximum power transfer.

3. High efficiency of RF-DC conversion and low power DCDC converter is required.

Minimum numbers of components are used in the design of the RF Energy harvesting system to reduce power dissipation in the circuit itself.

II. MICROSTRIP ANTENNA DESIGN

Antenna is one of the most crucial components in RF energy harvesting system to extract maximum power. Various antenna topologies have been reported in [2-4] for RF energy harvesting but high gain and bandwidth have not been achieved simultaneously.

In this paper, broadband electromagnetically coupled Square microstrip antenna (SMSA) [6] is used for the proposed RF energy harvesting system. It uses stacked multi-resonator configuration for broadband operation. Only the bottom patch is fed and top patch is coupled electromagnetically.

Antenna has a stacked configuration. Two 2mm thick aluminum radiating patches are electromagnetically coupled through an air gap of 1.3cm. First and second square patch are of length 15.2cm and 12.8cm respectively. The ground plane size is 24 cm X 24 cm. The total height of the antenna is 2.6cm. Fig. 2 shows fabricated electromagnetically coupled SMSA designed for GSM 900 band using IE3D software from Zeland.

Simulated VSWR and input impedance plots of electromagnetically coupled SMSA are shown Fig. 3(a) and (b).

Bandwidth for VSWR less than 2 is from 877 to 998 MHz

Fig.4 shows the gain plot of the antenna. The gain of antenna is 9.1 dB. Fig.5 shows measured VSWR plot and input impedance plot of SMSA. The antenna is designed for the 935-960 MHz (GSM 900) band but it can also covers part of CDMA transmission band (869-890MHz).To enhance the gain, array of SMSA can be used.

III. RF TO DC CONVERSION USING SCHOTTKY DIODES

A voltage doubler is designed using Dickson Multiplier topology. Silicon based Schottky diode having threshold voltage of 230 mV and diode capacitance of 0.26 pF is chosen.

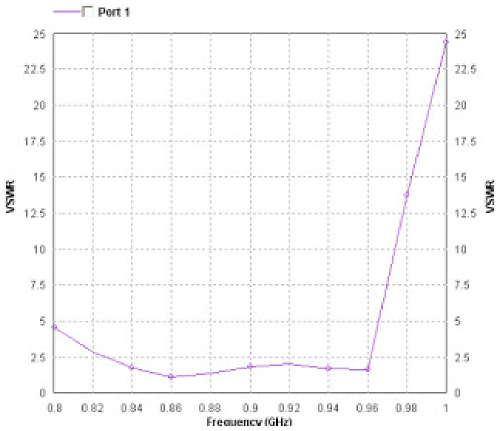

(a) VSWR of Electromagnetically Coupled SMSA
(b) Input Impedance of Electromagnetically Coupled SMSA

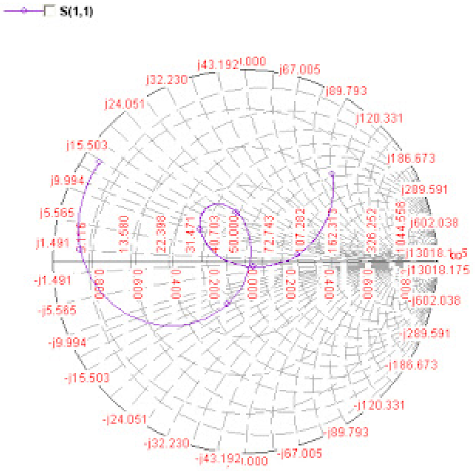

Fig. 3. Simulated (a) VSWR and (b) Input impedance of Electromagnetically Coupled SMSA

At microwave frequency, the non-linear capacitance of the diode governs the maximum power transfer to the load and amplitude of the rectified output as input impedance of the rectifier changes with frequency. Fig. 6 (a) and (b) shows a single stage voltage doubler and 6 stage voltage doubler. Both single stage and 6 stage voltage doubler circuits are fabricated. Fig.7 (a) and (b) show fabricated single stage voltage doubler and 6-stage voltage doubler circuits respectively. Six-stage voltage doubler is implemented to measure the power dependence of the rectifier for lower power level as the voltage measured at single stage is very low. Single stage voltage doubler is used for rectenna measurements inside the lab and near the cell towers as the power level are higher at

these places.

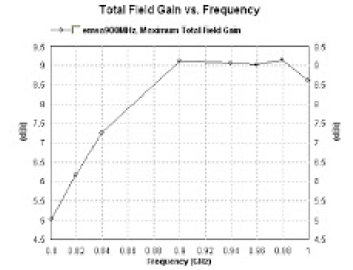

Fig. 4. Simulated gain plot of electromagnetically coupled SMSA

TABLE II
VOLTAGE AT VARIOUS STAGES OF SIX STAGE RECTIFIER WITH 2.2 KOHM LOAD AT 900 MHZ FOR DIFFERENT INPUT POWER LEVELS

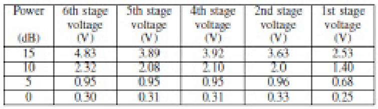

Power (dB)	6th stage voltage (V)	5th stage voltage (V)	4th stage voltage (V)	2nd stage voltage (V)	1st stage voltage (V)
15	4.83	3.89	3.92	3.63	2.53
10	2.32	2.08	2.10	2.0	1.40
5	0.95	0.95	0.95	0.96	0.68
0	0.30	0.31	0.31	0.33	0.25

Input impedance of the diode changes with the input poweras the dc operating point of the I-V curve moves in a nonlinear fashion [5]. The impedance variation of a six stage rectifier at 900 MHz with change in power level is shown in Fig.8. Impedance variation is taken at 900MHz as it is the approximate middle value of the CDMA and GSM 900 bands. Large variation of input impedance with input power and frequency makes designing a broadband conjugate matching network complex. Therefore, a resistive impedance matching network, as shown in Fig.9, is used to allow maximum power transfer.

The value of Rmatch depends on the input impedance of the antenna. This approach provides an approximate and simple matching between the antenna and rectifier.

The 6- stage rectifier is tested for its performance by feeding it directly from a RF source. Table II shows output voltage variation of six stage voltage doubler with 2.2 Kohm load at 900 MHz using a RF source at various power levels. Table III shows the output voltage variation of six stage voltage doubler for no load condition.

Rectenna is a combination of rectifier and antenna. Fig.10. shows the experimental setup of rectenna. It consists of an RF source transmitting 900 MHz (approximately middle value of CDMA and GSM 900 band) with output power of 17dBm (50mW). A horn antenna with a gain of 11dB is connected to the RF source to simulate cell towers inside the lab, and voltages at output of rectenna at various distances are measured.

Table IV summarizes the measured results at 900MHz using a single stage voltage doubler with 100K ohm and 10uF load.

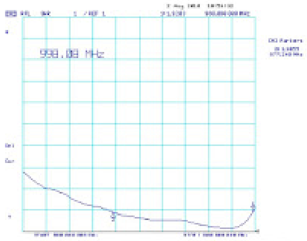

(a) Measured VSWR of Electromagnetically Coupled SMSA

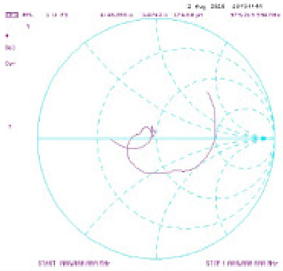

(b) Measured Input Impedance of Electromagnetically Coupled SMSA

Fig. 5. Measured (a) VSWR and (b) Input impedance of Electromagnetically Coupled SMSA

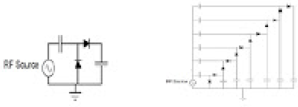

(a) Single stage voltage doubler (b) 6-stage voltage doubler

Fig. 6. (a)Single stage voltage doubler and (b) 6-stage voltage doubler

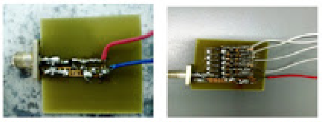

(a) Fabricated Single stage voltage doubler (b) Fabricated 6-stage voltage doubler

Fig. 7. Fabricated (a) Single stage and (b) 6 stage voltage doubler circuit

IV. MEASUREMENT AT ACTUAL SITE

Voltage levels are measured at actual site consisting of transmission cell towers at Indian Institute of Technology Bombay (IIT-B) using the fabricated electromagnetically coupled microstrip antenna and single stage voltage doubler. Figs.11 and 12 show the experimental setup for rectenna measurement at 50 m and 10 m from cell towers respectively. The voltage obtained at 50 m was 0.87V.This voltage can be stepped up using a boost converter. A stand alone boost converter from Texas Instrument TPS61220 has been designed to work from an input voltage of 0.7V. Voltage at 50m after stepping up can be used to charge a battery using a charging circuit. At a distance of 10m voltage of 2.78V is obtained. This large voltage can be used to charge a battery using a charging circuit. The power level observed using a hand-held radiation monitor was 4dBm and 5dBm at 50m and 10m respectively. Lower power levels are observed as the rectenna is not in the main beam of the cell towers antenna.

V. CONCLUSION

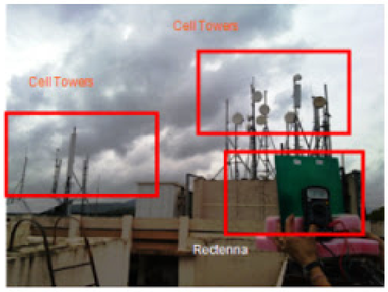

Fig. 12. Experimental setup for Rectenna measurement at 10m from cell tower (IIT-B)

In this article, a RF energy harvesting system from cell towers is presented. A high gain electromagnetically coupled antenna is developed. A silicon based schottky diode single stage and 6 stage rectifier are also designed. A voltage of 2.78V is measured at a distance of 10m from the cell tower. Due to the increase in the population density, many people live within 10m from the cell tower. RF energy harvesting from cell towers would therefore provide two fold benefits: provide an alternative source of energy and protect people living in close vicinity of the tower from radiation health hazards.

Posted by ravigandhi arul at 02:25 No comments:

Home

Subscribe to: Posts (Atom)

Awesome Inc. theme. Theme images by molotovcoketail. Powered by Blogger.

Current Developments of RF Energy Harvesting System for Wireless Sensor Networks

[1]Zahriladha Zakaria, [1]Nur Aishah Zainuddin, [1]Mohd Nor Husain, [1]Mohamad Zoinol Abidin Abd Aziz, [1]Mohamad Ariffin Mutalib, [1]Abdul Rani Othman
[1]Centre for Telecommunication Research and Innovation (CeTRI), Faculty of Electronic and Computer Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
zahriladha@utem.edu.my

Abstract

Energy harvesting or energy scavenging is basically a conversion process of the ambient energy into the electrical energy. The ambient energy exists around us in many different forms including thermal, chemical, electrical and radio frequency (RF). This technique significantly reduces the costs of replacing batteries periodically. Hence, energy harvesting offers various environmental friendly alternative energy sources, which include the vibration, electromagnetic wave, wind energy and solar power. This study will focus on RF energy harvesting that involves the generating of a small amount of the electrical power to drive circuits in wireless communication electronics devices. Recently, wireless sensor network (WSN) has been a crucial part of our daily life. The importance of WSN can be described by the use of sensors in many devices for home security including light sensors, room thermostat and alarm systems. This paper presents an overview and the progress achieved in RF energy harvesting, which involves the integration of antenna with rectifying circuit. Different combinations of antenna and rectifier topologies yield diverse results. Therefore, this study is expected to give an indication on the appropriate techniques to develop an efficient RF energy harvesting system.

Keywords: *Antenna, Rectifier, Energy Harvesting, Wireless Sensor Network*

1. Introduction

In few years back, many new technologies have been applied in order to enhance our life quality. One of these technologies is the cost-saving system in a house which uses self-powered devices. For instance, Figure 1 shows a domestic lighting application which uses a sensor system to operate. The light is automatically turned on once the sensor detects a specific level of darkness in the room. The smart system not only offers a more comfortable living, but also a large scale of energy saving and a lower cost solution.

An energy scavenging method has been applied in small devices based on wireless sensor technology. For instance, battery-less remote control and mobile phone charger which operates using harvested energy is currently having a great demand from users since they contribute to reduce cost and saving time.

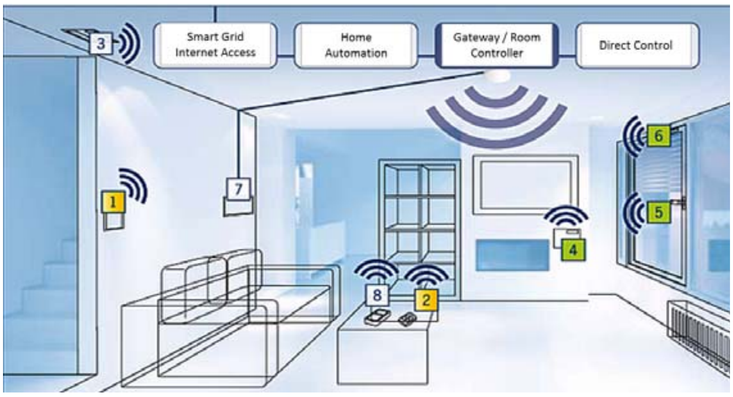

Figure 1. Energy harvesting wireless system in a smart house [1]

In recent years, there has been a growing usage of sensor-based wireless networks and applications which directly leads to the increased battery usage [2]. Until today, the limitation of energy supplies has become a crucial issue for the lifetime of these sensors since they operate on conventional batteries with a limited lifespan and fixed energy rate [3]. Furthermore, the deployment of a large number of sensor nodes results in periodic battery replacements which is impractical and cost-consumed [4].

One of the solutions proposed to overcome these issues is to opt for energy harvesting [3]. Energy harvesting is mostly applied to enhance the efficiency of a system and to enable new technology. Harvesters utilize different radiating sources such as solar, motion and vibration, temperature scale and radio frequency, depending on their type. These different radiating sources are described as in Figure 2 [2] .

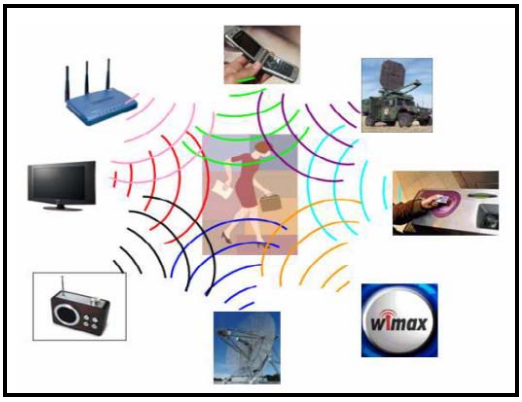

Figure 2. Examples of different radiating sources [2]

Similarly, the ambient energy can be extracted and converted into electrical energy that can be used to power up the sensors. Figure 3 shows the basic block diagram of RF-DC conversion system.

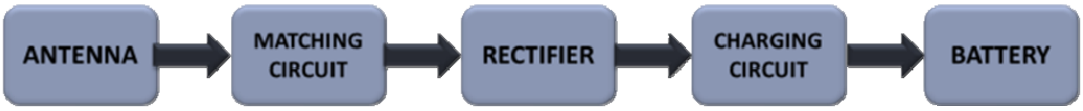

Figure 3. Basic block diagram of RF energy harvesting system

Despite the challenges possessed in RF energy harvesting [5], this system is still preferred due to the increasing availability of free RF energy from the advent of wireless broadcasting and communication systems.

The antenna is the key element of an RF energy harvesting system to capture the radiated RF energy and extract maximum power. The aim is to design antenna which can produce large DC output voltage. Until today, various antenna topologies have been recorded [3][4][6] for RF energy harvesting; nevertheless, only a few achieves a good performance in both gain and bandwidth.

Antennas can be designed in many forms, for instance patch antennas, dipole antennas, planar antennas, microstrip antennas and uniplanar antennas. Several types of patch antennas have been designed in various frequency regions focusing on bandwidth improvement [7]. Patch antenna with spiral structure [8] are also introduced where 3 types of spiral were compared to find the optimal structure.

Printed dipole antennas possess numerous advantages including low cost and large bandwidth. Therefore a dipole antenna was designed for optimum usage in GSM band applications [9]. Another antenna was designed for voltage boosting purposes in order to enhance DC output of RF energy harvesting system, specifically in a weak signal reception environment [10]. A study on improving antenna to exhibit dualband response for the energy harvesting application has been done in [11]. The other researchers are focused mainly on single band applications.

This paper reviews several techniques in obtaining RF energy from the ambient sources to be converted into the DC signal. In particular, the study focuses on the topology of the antennas and rectifying circuits to obtain optimum efficiency of the RF-DC conversion.

2. Energy harvesting system

An energy harvesting system consists of two main subsystems. The first one is the receiving antenna, which functions to capture ambient RF energy to power up the integrated embedded system. The second subsystem is the rectification circuitry, which converts the input RF power into DC output power efficiently.

The antenna captures the RF signals, and subsequently the rectifier circuit extracts the power of those signals and converts them into DC voltage. Therefore, an antenna with high efficiency is needed to transfer wireless power effectively.

2.1. Antenna Design

Recently, several efforts have been given to enhance the efficiency of the antenna in power scavenging [12][13]. Rectifying antenna, or literally called a rectenna, is one of the most important components in the wireless power transmission. It comprises a receiving antenna and a circuit which converts radio frequency to direct current (RF-DC) [14][15][16]. An efficient wireless power transmission system requires a rectenna with high RF-DC conversion efficiency. Several antenna topologies have been reported [5] for RF energy harvesting.

Most early research on radio frequency (RF) energy harvester designs focus on harvesting RF energy in a single RF band [17][18]. However, recently, there have been designs and implementations of a dual band [11] and triple band [19][6] antenna design for RF energy harvesting applications. The design of compact multiple band microstrip antennas for wireless applications also gains a highlight and received much attention in [20][21].

The planar monopole antenna is used to provide required impedance matching at two bands with easier integration [11]. The antenna consists of a ground plane on the back of the PCB, and three connected microstrip lines on the top of the PCB. All the microstrip lines shared the same feeding point with the RF input pin. The antenna was optimized by altering the length of the microstrip lines which also managed to provide required impedance. The same antenna structure with different antenna parameters was also designed. The purpose is to use the antenna for different implementations where Antenna A is for 900MHz & 1900MHz band application while Antenna B is for 900MHz & 2000MHz band application. Antenna B obtained better return loss with -35dB in comparison to Antenna A with -25dB of return loss. However, both antennas are still suitable to be adapted to general dual band applications which require a matching to different impedances at two frequencies.

Another study on dual band antenna has been done, but this time a microstrip patch was used [22]. The structure consists of a square-shaped ring patch and it is connected to a microstrip feedline. This antenna potentially work well with RF energy harvesting system due to its promising gain and bandwidth values. Furthermore, miniaturized microstrip antennas are able to improve their performance and reduce the overall size of wireless devices [27].

The study about ring patches proves that various ring shapes managed to obtain better bandwidth and higher radiation resistance when the higher order modes are excited [23][24][25]. Another different study found that the ring antenna is excited with a capacitive strip in another layer in a stacked configuration [26]. Paper [25] describes the impedance properties of a dual frequency concentric annular ring microstrip patch antenna. Meanwhile, a parametric study of a microstrip square ring antenna has been done in [22]. From the investigation, the bandwidth and gain of the microstrip square ring antenna increased with the increment the ring’s width even though the resonance frequency is slightly shifted to a lower frequency.

Printed dipole was also a type of antenna that was used for energy harvesting due to its advantageous features such as low cost and provide large bandwidth. In [9], printed dipole antenna is designed and optimized specifically for GSM band application at 915 MHz. This antenna yields a very omnidirectional radiation pattern which is very suitable for RF energy harvesting to capture signals in all directions. A mini reflector was used to triple up the gain of this antenna from 2.49 dB to 6.28 dB.

In [8], several spiral antennas for RF energy harvesting in 0.2 GHz to 2.4 GHz frequency range is compared. The spiral was known as one kind of antenna structure that possess radiation pattern over a wide range of frequency. However, it has less packing conformity when built in printed circuit board, which caused a drawback of many excessive areas on the board. 3 types of spiral antennas have been compared which are Archimedean spiral antenna, square spiral antenna and hexagonal spiral antenna as shown in Figure 4.

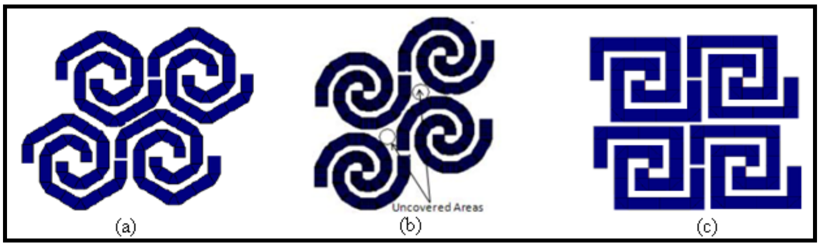

Figure 4. Array for (a) Hexagonal spiral antenna (b) Archimedean spiral antenna (c) Square spiral antenna [8]

Archimedean and square spiral antenna obtained better gain value than hexagonal spiral antenna with 5dB and 4.8dB, both at 2.4GHz frequency. However hexagonal spiral tends to approach the circular shape which optimizes the PCB cover area by accommodating a number of spirals. It also generates better broadband while maintaining packing conformality.

Patch antenna most likely has been chosen to be designed and implemented with RF energy harvesting. Various structures in the form of patch antennas have been designed such as L-shape, U-shape, star shape, E-shape, circular patch and square shaped to be used in the 900MHz, 2.4GHz, 4GHz, 5GHz, and 6GHz frequency [28][29][30][31]. Despite of its advantages of being low profile and inexpensive to fabricate, patch antennas has a drawback of narrow bandwidth, low efficiency, and low gain [7][4][19]. In order to overcome the problem, some methods have been investigated and proposed. These include embedding slots on the patch antenna and using various impedance matching and feeding techniques [32][33][34].

Antenna array design has also been investigated for RF energy harvesting with expectations to scavenge power effectively in [12][13]. Author of paper [4] proposed a microstrip stacked patch antenna array with lower power consumption. It shows better VSWR, return loss and efficiency compared to common inset fed patch antenna at 2.45 GHz frequency with improvement of return loss of about 30dB. Apart from that, the high gain value of 11.98 dB also was a factor that contributes to the antenna's ability to achieve better voltage levels. However this antenna has a drawback of larger antenna size even though the use of antenna arrays can increase the RF and DC power [2]. A single patch itself has a dimension of 40.49mm x 40.49mm.

2.2. Rectifier Design

Rectifier, which is also called as pump device, has the function of converting an input RF signal received by the antenna into suitable DC supply voltage. Rectifier converts a part of the incoming power supply to DC for power supply. It takes advantage of the diode's behavior to rectify a signal; depending on the signal level, the diode allows or does not allow the current to pass.

The basic or simplest rectifier consists of two different positions of the capacitor; parallel or series. Figure 5 (a) shows a rectifier using a series diode where the capacitor is connected in series followed by a parallel connected diode.

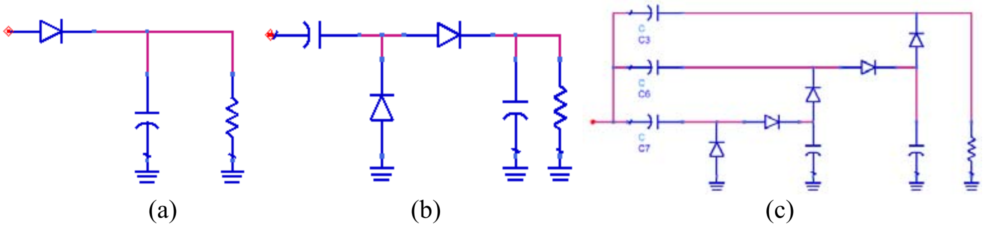

Figure 5. Type of rectifier: (a) Basic rectifier (b) Voltage doubler (c) Voltage multiplier

From the basic rectifier, a voltage doubler circuit can be designed by using two diodes and two capacitors as shown in Figure 5 (b). The voltage doubler configuration is chosen to be used as the rectifier circuit in [35] and [36]. Paper [37] has used conventional voltage doubler rectifier together with floating-gate voltage doubler rectifier. The function of voltage doubler rectifier is to rectify the full-wave peak-to-peak voltage of the incoming RF signal. It can be arranged in cascade to increase the output voltage.

Additional sections of diodes and capacitors may be cascaded in order to obtain voltage multiplier circuits of higher order. Voltage multiplier consists of N-stage which indicates the number of stages of particular voltage multiplier. Figure 5 (c) shows the design of a voltage multiplier.

The design of voltage multiplier circuit has been described in various literatures [38][39][40], for instance Cockroft Walton, Villard Cascade and Dickson voltage multiplier. Cockroft Walton voltage multiplier is capable to produce DC output voltage potential tens of times that of the peak AC input voltage. However, the drawback of this kind of rectifier is the limitation in current capacity and regulation. Table 1 shows the comparison of the rectifiers.

Table 1. Comparison of rectifiers

Type of rectifier	Structure	Rectifier Topology
Basic rectifier	A diode connected in series with a load. A capacitor acted as a filter to smoothen the ripple in the output. Commonly called as single-stage rectifier.	Half-wave rectifier, full-wave rectifier
Voltage doubler	Uses two stages to approximately double up the DC voltage.	Villard circuit, Greinacher circuit, bridge circuit, Dickson charge pump voltage-doubler
Voltage multiplier	Converts RF energy into DC voltage using a network of capacitors and diodes.	Villard cascade voltage multiplier, Dickson multiplier, Cockroft Walton voltage multiplier

3. Discussion

The integration of the antenna and the rectifier is also known as rectenna. The rectenna is a passive element with a diode than can receive and rectify microwave power to direct current, which results in its ability to operate without any power source. Many research works on rectenna elements have been published. For example, the finding of various antenna types to be used in rectenna system such as dipole, Yagi-Uda antenna, microstrip antenna, monopole, coplanar patch, spiral antenna, and even parabolic antenna. The antennas can also be integrated with any type of rectify circuit (rectifier) such as single shunt full-wave rectifier, full-wave bridge rectifier, voltage multiplier and hybrid rectifiers. The circuit, especially the diode, determines the RF-DC conversion efficiency. Common diode used in most rectennas is silicon Schottky barrier diodes.

RF energy harvesting basically has several different types of system. In [2], two systems have been studied to recover the RF energy which is a broadband system and a narrowband system. For both systems, a broadband antenna is needed and therefore a spiral antenna is used. Unlike any spiral antenna structure in [8], an equiangular spiral antenna in [2] is used due to its advantageous features including possible dual polarization and its ability to operate under a broad frequency band.

In broadband system, it consists of rectifier without matching circuit and broadband antenna. Therefore the antenna used must be able to maximize the DC power harvested and recover all signals available. For that purpose, an omni-directional broadband antenna is needed. Return loss of the proposed antenna is lower than -10dB which means the efficiency of this antenna will be more than 90% overall. The antenna also managed to obtain a quasi omni-directional radiation pattern at a frequency of around 1GHz.

While for the narrowband system, the same antenna structure was used together with matching circuit which resulted an increment to DC power. However, the drawback of this system is the low efficiency of conversion RF/DC that occurred (0.6%).

Another paper that is interested in spiral antenna as a receiver medium is stated in [37][38]. It presents a constant gain and input impedance in the interest band with circular polarization. The input recoverable power increases increasing the antenna gain. Antenna gain plays an important role on the RF to DC conversion efficiency. In addition, the antenna with directional radiation pattern contributes higher gain achievement, which eventually will increase the antenna's receiving power. Therefore, most researchers use array and stacked patch antenna technique to increase the antenna gain, approximately until 10dB and above.

Several designs of antenna that are operate under a broad frequency band haven been identified from selected sources. These antennas have been redesigned and optimized using Computer Simulation Tool (CST). Some modifications have been made to the antenna structure to investigate the new outcome. Most of the antenna uses a FR-4 board material with different type such as microstrip, monopole, dipole and planar and embedded with various antenna structure as summarized in Table 2.

Table 2. Comparison of antenna structure

Antenna Type	**Antenna Structure**	**Remarks**	**Ref.**
Monopole	Microstrip line	• Well suited to be adapted to general dual band applications. • Has a flexible design where required impedance can be achieved by altering length and geometries of the microstrip lines. Return loss achieved is -25dB. • Helps provide the ability to operate at more RF bands and harvest energy for remote power applications.	[14]
Microstrip	Square ring	• This antenna has a drawback of a very narrow bandwidth. • Bandwidth increased by increasing the width of the ring, however, another drawback is resonant frequency is also slightly shifted to lower frequency when the width of ring increases. Therefore only the width of two sides away from the feedline is increased with return loss achieved until -44.8dB and a gain of 6.12dB.	[29]
Dipole	Similar to golf clubs shape	• The structure of suggested GSM printed dipole antenna is geometrically optimized to reduce antenna size, increase impedance adaptation, to center the resonant frequency and to increase gain. • Have an omnidirectional radiation pattern which is very good for energy harvesting antenna. • Antenna gain is improved by 252% by using a mini reflector behind the antenna at a distance of $\lambda/4 \approx 75$mm.	[12]
Planar	Spiral	• The rectangular patch antenna has a drawback of narrow bandwidth. However, instead of that, it has good radiation pattern over a wide range of frequency. 3 types of spiral antenna is compared – Archimedean spiral, square spiral, and hexagonal spiral. • Among all 3 designs, hexagonal spiral generates better broadband and gain. More number of spirals also can be used by using hexagonal spiral.	[11]
Microstrip	Rectangular	• Patch antenna has the advantages of low profile and suitable for high-frequency application despite of their narrow bandwidth. • Antenna structure has a patch, pi matching network, and partial ground plane on the other side. The edge fiddling technique is used. • Wide bandwidth can be achieved by adjusting antenna's parameter.	[8]
Planar	Stacked patch with	• Antenna consists of two stacked rectangular patches and a ground plane where the upper four patches have four slots	[3]

	slots	on the edges. • The antenna has a drawback in terms of size. It is bigger than other conventional antenna for RF energy harvesting. • However, it has a very high gain value of 11.98dB.	
Microstrip	Equiangular spiral	• For maximal power transfer, the antenna impedance must be matched to the optimal diode impedance for all frequencies. • Equiangular antenna has a possible dual polarization, broadband antenna and omni-directional radiation pattern.	[1]

There are a few requirements need to be considered when designing a rectifier circuit [41]. For instance, several criteria of rectifying circuit are evaluated in [42] including the choice of diodes, number of stages, effect of load impedance as well as the efficiency of RF-DC conversion.

We need to bear in mind that the most crucial requirement for a harvesting energy circuit is to be able to operate with weak input RF power. Therefore, a diode with a very fast switching time need to be used for the energy harvesting circuit is operating at high frequencies. Schottky diode is one of the popular diodes used in rectifier circuit. It uses a metal-semiconductor junction which allows the junction to operate much faster and gives forward voltage drop. Saturation current is another critical parameter that impacts the efficiency of diodes.

The number of rectifier stages has a major influence on the output voltage of the energy harvesting circuit where the output voltage is directly proportional to the number of stages used in the energy harvesting circuit. A rectifier with too few rectifier stages would yield an insufficient output voltage while too many rectifier stages may damp out the effect of the high Q-resonator. Figure 6 and Figure 7 shows the relation of number of rectifier stages on the energy harvesting efficiency and output voltage.

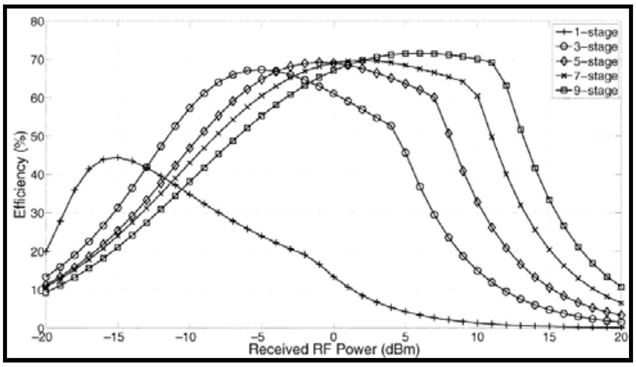

Figure 6. Effect of number of stages on the efficiency of energy harvesting circuit [42]

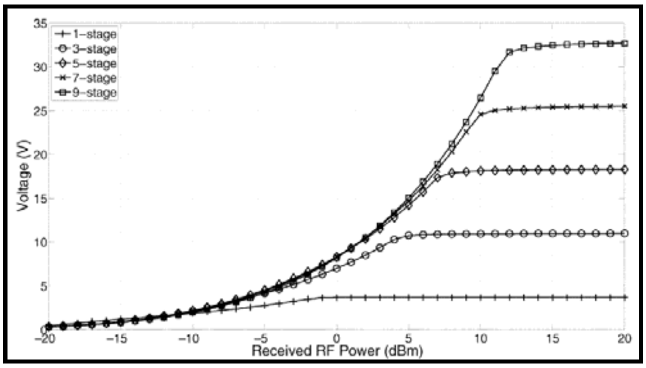

Figure 7. Effect of number of stages on the output voltage of energy harvesting circuit [42]

As mentioned previously, there have been several types of system in RF energy harvesting. Researchers in [2] came out with two types of systems. The first one is a broadband system without using any matching circuit while the second one is a narrowband system which includes a matching circuit.

The DC power scavenged using the narrowband system [2] can be attended until 400pW. The RF/DC converter used is a voltage doubler which uses the harmonic balance method. It was then optimized by using a commercial zero biased Schottky diode. Figure 8 shows the variations of the output DC power versus the input RF power at 1.5GHz which resulting the RF power more than -32dBm for both simulation and measurement.

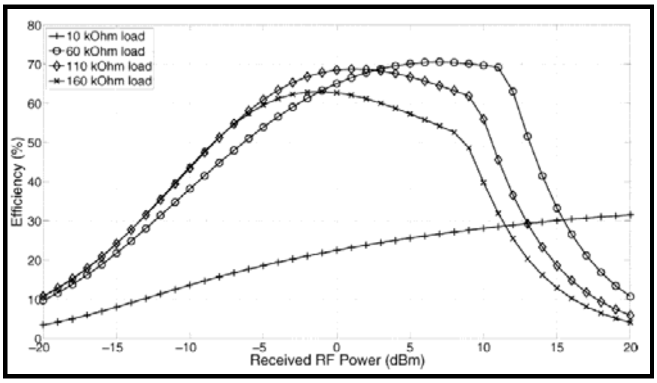

Figure 8. Effect of load impedance on the efficiency of energy harvesting circuit [42]

Based on the review gained in this study, it is believed that there is still a room for further researches and developments before the technique become mature. Thus, it is recommended to further investigate and discover the optimum antenna and rectifier topology at GSM850, GSM900, GSM1800 and ISM frequency band particularly in Malaysia. It also involves designing several types of antennas to cover the above-mentioned frequency range in order to match the criteria for the overall system.

For the RF energy harvesting, the antenna with directional radiation pattern is most suitable to integrate with rectifying circuit because it radiates or receives more power in particular directions. However, antennas with an omnidirectional radiation pattern can still be considered since it provides and receives equal signals in all directions even though the signal strength might be weaker. Another factor to consider is to design an antenna with higher gain since the gain relates to the antenna's ability to capture ambient signals significantly. Besides, the antenna itself must have a return loss better than -10 dB which contributes to 90% of antenna efficiency. Larger return loss indicates higher power being radiated by the antenna which eventually increases the antenna's gain. Therefore, printed circuit board or PCB-based antenna will be used in this project since it is easy to fabricate, low cost and able to achieve high gain.

While the optimum rectifying circuit topologies must be comprehended to fulfill the requirement of supplying sufficient energy harvested and thus, it has been recognized that voltage multiplier is capable to realize a maximum conversion of RF signal into DC electrical signal.

Figure 9 shows the measurement setup for an energy harvesting system. The propagation model as studied in [43] can also be used to determine the path loss between the transmitter and receiver. Then, the intelligent technique developed by [44] is suggested to be applied in this work to improve the energy-efficiency for wireless sensor networks.

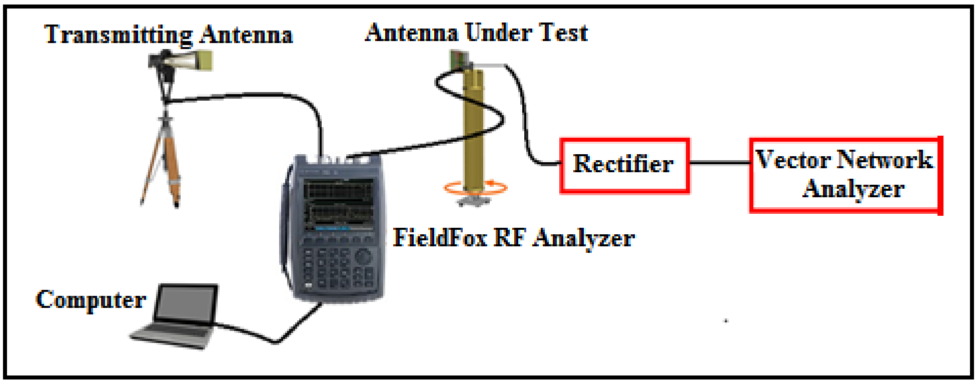

Figure 9. Antenna and rectifier measurement setup

4. Conclusion

A study on energy harvesting based on microwave signals has been reviewed. There are some design issues occurred in RF energy harvesting that need to be considered including low conversion efficiency and low rectified power. Improvements in efficiency of the RF signal harvesting is important. In general, many published papers on RF power to DC conversion - have presented circuits that are capable of converting input or incident power as low as -20dBm. This means that, if an RF survey or scan finds signals in space, with power spectrum levels around -20dBm, then it is potentially viable to harvest such signal power. The verification process of the proposed antenna and rectifier design that operates particularly in GSM and ISM band will be further explored and implemented through an experiment in laboratory and field works. It is believed that these discoveries will motivate further researches towards the usage of RF energy harvesting particularly in wireless sensor networks and their applications.

5. Acknowledgement

The authors would like to thank UTeM for sponsoring this work under the CoE, research grant UTeM, PJP/2012/CeTRI/Y00001.

6. References

[1] “Smart Homes and Home Automation by EnOcean.” [Online]. Available: http://www.enocean-alliance.org/en/smart-homes-residential-buildings/.
[2] Bouchouicha, D., F. Dupont, M. Latrach and L. Ventura, 2010. Ambient RF Energy Harvesting. International Conference on Renewable Energies and Power Quality, pp. 2–6.
[3] Sim, Z. W., R. Shuttleworth and B. Grieve, 2009. Investigation of PCB Microstrip Patch Receiving Antenna for Outdoor RF Energy Harvesting in Wireless Sensor Networks: 129–132.
[4] Patel, A. C., M. P. Vaghela, H. Bajwa, and P. K. Patra, 2009. Power Harvesting for Low Power Wireless Sensor Network: 633–636.
[5] Thomas, J. P., M. A. Qidwai, and J. C. Kellogg, 2006. Energy scavenging for small-scale unmanned systems. Journal of Power Sources, 159(2): 1494–1509.
[6] Arrawatia, M., M. S. Baghini, and G. Kumar, 2010. RF Energy Harvesting System at 2 . 67 and 5.8 GHz Microwave Conference Proceedings (APMC), 2010 Asia-Pacific, pp: 900–903.
[7] Rizzoli, V., G. Bichicchi, A. Costanzo, F. Donzelli, and D. Masotti, 2009. CAD of Multi-resonator rectenna for micro-power generation. Microwave Integrated Circuits Conference, 2009. EuMIC 2009. European, pp: 8–11.

[8] Devi, K. K. A., and S. Sadasivam, 2011. Design of a 377 Ω Patch Antenna for Ambient RF Energy Harvesting at Downlink Frequency of GSM 900. Communications (APCC), 2011 17th Asia-Pacific Conference, pp: 492–495.
[9] Mandal, S., and S. K. Giri, 2011. Comparison of antennas for Radio Frequency Energy Harvesting in 0.2–2.4 GHz range. 2011 3rd International Conference on Electronics Computer Technology, pp: 93–97.
[10] Alaeldine, A., M. Latrach, H. Raggad, and Z. Sayegh, 2011. Design and Optimization of a GSM Printed Dipole Antenna for Energy Harvesting Applications. 2011 IEEE International Symposium on Antennas and Propagation (APSURSI), pp: 1780–1783.
[11] Arai, H., M. Chomora, and M. Yoshida, 2012. A voltage-boosting antenna for RF energy harvesting. 2012 IEEE International Workshop on Antenna Technology (IWAT), pp: 169–172.
[12] Shao, X., B. Li, N. Shahshahan, N. Goldsman, T. S. Salter, and G. M. Metze, 2011. A Planar Dual-Band Antenna Design for RF Energy Harvesting Applications. Semiconductor Device Research Symposium (ISDRS), 2011 International, 41(1): 10–11.
[13] Pozar, D. M., 1992. Microstrip Antenna. Proceedings of The IEEE, 80(1) : 79–81.
[14] Balanis, A., 2005. Antenna Theory, 3rd Edition. John Willey & Sons.
[15] Tsunayuki, Y., F. Kazuhiro, S. Minoru, and N. Shigeji, 2008. Design of RF-DC Conversion Circuit Composed of Chip Devices, Microwave Conference, 2008. APMC 2008. Asia-Pacific, pp: 1-4.
[16] Selvakumaran, W., W. Liu, H. S. Boon, M. Luo, and Y. L. Sum, 2009. Design of Low Power Rectenna for Wireless Power Transfer. TENCON 2009 – 2009 IEEE Region 10 Conference, IEEE Conferences, pp: 1-5.
[17] Georgiadis, A., G. Andia, and A. Collado, 2010. Rectenna Design and Optimization Using Reciprocity Theory and Harmonic Balance Analysis for Electromagnetic (EM) Energy Harvesting. Antennas and Wireless Propagation Letters, IEEE Journal, 2010, 9: 444-446.
[18] Umeda, T., 2006. A 950-MHz Rectifier Circuit for Sensor Network Tags With 10-M Distance. IEEE Journal of Solid-State Circuits, 41(1): 35-41.
[19] Salter, T., 2009. RF Energy Scavenging System Utilizing Switched Capacitor DC-DC Converter. Electronic Letters, 45(7): 374–376.
[20] Alkanhal, M. A. S., 2009. Composite Compact Triple Band Microstrip Antennas. Progress In Electromagnetics Research, 93: 221–236.
[21] Rowell, C. R., and R. D. Murch, 1998. A Compact PIFA Suitable for Dual-Frequency 900/1800 MHz Operation. IEEE Trans. Antennas Propagation, 46: 596-598.
[22] Sim, C. Y. D., 2007. A Novel Dual Frequency PIFA Design for Ease of Manufacturing Journal of Electromagnetic Waves and Applications, 21(3): 409-419.
[23] Behera, S., and K. J. Vinoy, 2009. Microstrip Square Ring Antenna For Dual-Band Operation. Progress In Electromagnetics Research, 63: 41–56.
[24] Mayhew-Rydgers, G., J. W. Avondale, and J. Joubert, 2000. New Feeding Mechanism for Annular-Ring Microstrip Antenna. Electronics Letters, 36: 605-606.
[25] Nurie, N. S., and R. J. Langley, 1990. Input Impedance of Concentric Ring Microstrip Antennas for Dual Frequency Band Operation Including Surface Wave Coupling. IEE Proceedings H Microwaves, Antennas and Propagation, 137(6): 331.
[26] Misra, I., S. K. Chowdhury, and S. Member, 1998. Study of Impedance and Radiation Properties of a Concentric Microstrip Triangular-Ring Antenna and Its Modeling Techniques Using FDTD Method, 46(4): 531–537.
[27] Latif, S. I., and L. Shafai, 2005. Dual-Layer Square-Ring Antenna (DLSRA) for Circular Polarization. 2005 IEEE Antennas and Propagation Society International Symposium, 2: 525–528.
[28] Garg, R., P. Bhartia, I. J. Bhal, and A. Ittipiboon, 2001. Microstrip Antenna Design Handbook. Artech House, Boston.
[29] Yang, F., S. Member, X. Zhang, X. Ye, and Y. Rahmat-samii, 2001. Wide-Band E-Shaped Patch Antennas for Wireless. Antenna and Propagation, IEEE Transaction, 49(7): 1094–1100.
[30] Chair, R., C. L. Mak, K. F. Lell, K. M. Luk, and A. Kishk, 2005. Miniature wideband half U-slot and half E-shaped patch antennas. IEEE Trans. Antenna Propagation, 53: 2645-2652.
[31] Xiong, J., Z. Ying, and S. He, 2008. A Broadband E-Shaped Patch Antenna of Compact Size and Low Profile. IEEE Antennas and Propagation Society International Symposium, pp: 1-4 .

[32] Bulja, S., 2006. Broadbanding Microstrip Antennas Using Unequal Patches, IEEE Antennas and Propagation Society International Symposium, pp: 3731–3734.
[33] Albooyeh, M., N. Kamjani, and M. Shobeyri, 2008. A Novel Cross-Slot Geometry To Improve Impedance Bandwidth of Microstrip Antennas. Progress In Electromagnetics Research Letters, 4: 63-72.
[34] Kasabegoudar, V. G., and K. J. Vinoy, 2009. A Broadband Suspended Micostrip Antenna for Circular Polarization. Progress In Electromagnetics Research, PIER 90: 353-368.
[35] Kordzadeh, A., and F. H. Kashani, 2009. A New Reduced Size Microstrip Patch Antenna With Fractal Shaped Defects. Progress In Electromagnetics Research B, 11: 29-37.
[36] Le, T., J. Han, A. V. Jouanne, K. Mayaram, and T. S. Fiez, 2003. Piezoelectric Power Generation Interface Circuits. Proceeding IEEE Custom Integrated Circuit Conference (CICC). San Jose, California, USA, pp: 489-492.
[37] Sivaramakrishnan A., and K. J. Jegadishkumar, October 2011. A Highly Efficient Power Management System for Charging Mobile Phones using RF Energy Harvesting, International Journal of Information Technology Convergence and Services (IJITCS) Vol.1, No.5, pp: 21–30.
[38] Le, T., K. Mayaram, and T. S. Fiez, 2008. Efficient Far-Field Radio Frequency Energy Harvesting for Passively Powered Sensor Networks. IEEE Journal of Solid-State Circuits, 43(5): 1287- 1302.
[39] Minhong M., M. H Mickle, C. Capelli, and H. Swift, 2005. RF Energy Harvesting with Multiple Antennas In The Same Space. IEEE Antennas and Propagation Magazine, 47(5): 100-106.
[40] Govardhani I., M. S. R. S Tejaswi, M. Venkata, N. Narayana, A. Babu., G. Anupama., and K. R. T. Venkata, 2011. Design of Coaxial Fed Microstrip Patch Antenna for 2.4 GHz Bluetooth Applications. Journal of Emerging Trends in Computing and Information Sciences, 2(12): 686-690.
[41] Brown, W. C., 1984. The History of Power Transmission by Radio Waves. Microwave Theory and Techniques, IEEE Transactions, 32(9): 1230-1242.
[42] Escala, O. A. C., 2010. Study of The Efficiency of Rectifying Antenna Systems for Electromagnetic Energy Harvesting. Degree. Thesis. Universidad Politecnica de Catalunya. Peru.
[43] Tan T. D., Anh N. D., Thang C. M., Linh B. L. P., Quynh T. T. T., “ An Indoor Propagation Case Study at Frequency 915 MHz”, IJEI: International Journal of Engineering and Industries, Vol. 3, No. 4, pp. 9 ~ 19, 2012.
[44] Zhenjun Luo, Yingjiang Zhang, Zhangfan Pang, "Research on Energy-Efficient Intelligent Method for WSN", JCIT: Journal of Convergence Information Technology, Vol. 8, No. 1, pp. 94 ~ 102, 2013.

Low Profile Design for Intrepid to Assess, Measure and Harvest RF Energy

Manish Upadhyay[*]

Electronics and Communication Engineering Department
Bharati Vidyapeeth's College of Engineering
New Delhi, India
**Corresponding author E-mail:polarizadmax@gmail.com*

Abstract

A low profile design which could be used in order to analyze RF energy in near home premises is presented in this paper. The system design is intended to practice in the real world in an easy, economical and DIY way. The idea was to detect the possible source of RF energy that could be harvested even for a short span. A monitoring unit which is microcontroller with two patch antennas, each with a two stage schottky diode voltage multiplier acting as its sensors look for a potential RF energy (Bi-directional only) as soon unit powered up, then unit compare and record the strength of nearby RF energy which is provided as feedback to a servo motor. A rectenna, which is mounted on the top of the servo motor, is made responsible to measure and harvest RF energy. The DC power from the harvesting rectenna is used to power digital thermometer with an LCD module. A low profile prototype is implemented, tested locally and found with acceptable performance.

Keywords: *Energy harvesting, Measurement, Rectennas, Schottky diode, Sensor System.*

---

1. Introduction

Emerging self powered systems challenge and dictate the direction of research in energy harvesting (EH). State of the art in energy harvesting is being applied in various fields using different single energy sources or a combination of two or more sources. This kind of environmental friendly energy sources include energy harvesting from rectennas, passive human power, wind energy and solar power. Power we can extract from those techniques is limited by regulations and free-space path loss. As a general idea, small dimensions are a basic feature of portable devices, so the rectenna should be the same way. [17] Smaller the size results in the received power to be lower. This made wireless power transfer suitable for the low-power applications, e.g., a low-power wireless sensor.

The way technology advance every year allow the decrease of certain characteristics in digital systems, like size and power consumption, that will lead to the gain of new ways of computing and use of electronics, as an example we have wearable devices and wireless sensor networks. Currently, these devices are more powered by batteries, however, they present many disadvantages such as: the need to either replace them or recharge them periodically and their big size and weight compared to high technology electronics. A solution proposed to this problem was stated before: to extract (harvest) energy from the environment to either recharge a battery, or even to directly power the electronic device. [3][20]

Targeted RF energy is in the ambient or areas close to transmission towers provides an opportunity to harvest that energy. Some of the most prominent sources are FM radio systems ((88-108 MHz, transmitted power few tens of KW), TV Transmission (180-220 MHz, transmitted power few tens of KW), Cell Tower Transmission (10 to 20 W per carrier), Wi-Fi (2.45GHz, 5.8GHz), AM Transmission (540-1600 KHz, transmitted power few hundred KW) and mobile phones (transmitted power 1W to 2W), etc. [15] Cell towers can be used as a continuous source of renewable energy as they transmit 24 hours.

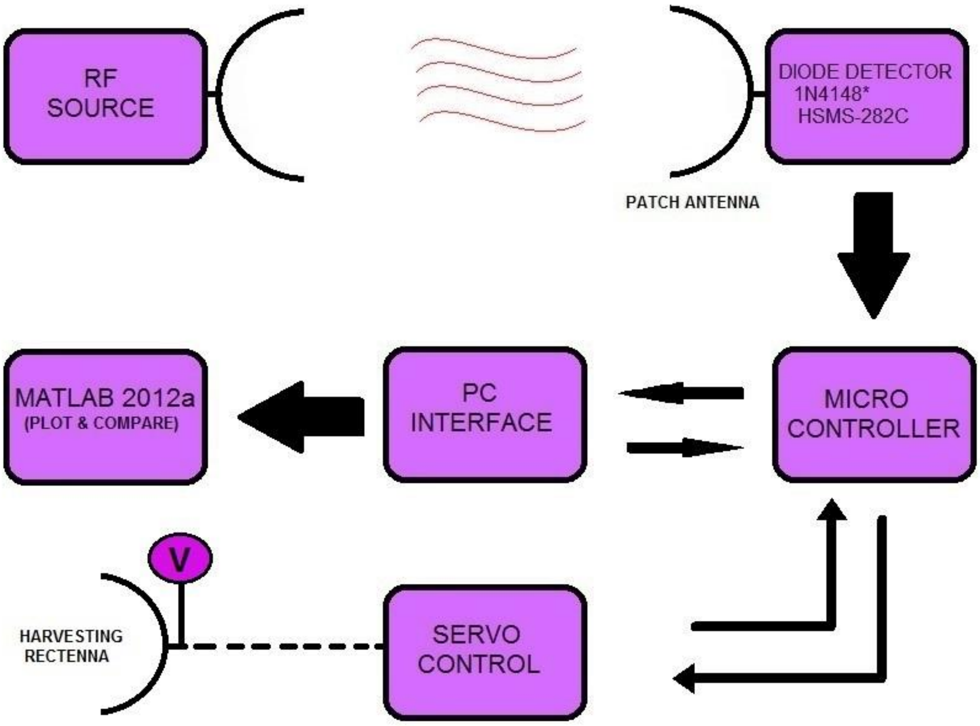

Fig. 1: a) Detailed Block Diagram for proposed system. (V-voltmeter)

Fig 1 presents the basic functionality of the system:

1. First, an antenna is in charge to capture the RF signals from the possible source of interest (CDMA-GSM Band).
2. Then a zero bias diode detector circuit (voltage multiplier) will convert RF-DC (carrier power) which will be input to the monitoring unit.
3. The Arduino Uno (open source hardware as monitoring unit) is used to read real time analog data from the voltage multiplier and will send it to PC via an interface.
4. Matlab R2012a software is used to read digital data values available at pc port and to plot it in terms of voltage, compare the signal and put command to control servo motor of the measuring Rectenna.
5. Servo motor will get a command from Matlab R2012a via microcontroller unit and also sent back feedback voltage for current position.
6. Finally the measuring rectenna will be on target RF source to measure electric field strength and display it in terms of voltage on the voltmeter.

Several operating frequencies for the rectenna have been investigated in the literature. Traditionally the GSM900 band has been utilized due to the presence of 2G networks; additionally the 1800 MHz band has also been considered which implies a smaller antenna aperture area than that of 900MHz. Both frequency bands present similar advantages because they have comparably low atmospheric loss, cheap components availability, and high conversion efficiency.[4][8] Hence the frequency bands corresponding to mobile telephone systems such as 800 MHz, 900 MHz and 1800MHz present good alternatives for electromagnetic energy harvesting systems, although they require a larger antenna size.

2. Design

In order to be able to utilize wireless power efficiently rectenna is needed. This paper presents a rectenna system with numerical and test observations. This provides a good insight of low profile system for the wireless power assessment and utilization. The incident RF power is converted into dc power by the voltage multiplier. The matching network, composed of inductive and capacitive elements, ensures the maximum power delivery from antenna to voltage multiplier. The energy storage (output capacitor) ensures smooth power delivery to the load and as a reserve for durations when external energy is unavailable. Such a design needs to be carefully crafted: increasing the number of multiplier stages gives higher voltage at the load, and yet reduces the current through the final load branch. [17]

This work has three major goals:

1. Measurement of stray radiation in term of Electric field intensity (V/ m).
2. RF energy harvesting/recycling.
3. Powering of low-power applications.

2.1. Voltage multiplier and diode selection

The forward voltage drop (V_f), reverse-recovery time (T_{rr}), and junction capacitance (C_j) of Schottky diodes are closer to ideal than the average “rectifying” diode. This makes them well suited for high-frequency applications. Unfortunately, though, Schottky diodes typically have lower forward current (If) and reverse voltage (V_{rrm} and V_{dc}) ratings than rectifying diodes and are thus unsuitable for applications involving substantial amounts of power, Though they are used in low voltage switching regulator power supplies. [23]

The Schottky diodes used are characterized by a low parasitic capacitance C_{jo} (0.14 pF) and a low serial resistor R_s (20 Ω). At microwave frequencies, rectifiers are highly nonlinear and difficult to design based upon purely analytic equations. Commercially available harmonic-balance simulators are useful at low power levels.

When looking for a rectifying diode in the context of RF recycling, a diode with high conversion efficiency, even for very small incident power levels is required. One of the crucial requirements for the energy harvesting circuit is to be able to operate with weak input RF power. For a typical 50-ohm antenna, the -20 dBm received RF signal power means amplitude of 32 mW. As the peak voltage of the ac signal obtained at the antenna is generally much smaller than the diode threshold [12], diodes with lowest possible turn on voltage are preferable. Moreover, since the energy harvesting circuit is operating in high frequencies, diodes with a very fast switching time need to be used. [17] Schottky diodes use a metal–semiconductor junction instead of a semiconductor–semiconductor junction. This allows the junction to operate much faster, and gives a forward voltage drop of as low as 0.15 V.

Earlier the choice was using Fairchild’s 1N4148 small signal diodes and 1N5819 Schottky rectifiers. But the study presented in [22] gave a broad view of there operation under various condition determine that the previous diodes having higher forward voltage requirement as compared to Hsms 282c and the capacitance of the diode’s is higher as well .The final rectifier circuit is designed using Hsms 282c Schottky Diode. Avago’s Hsms-282x family of zero bias Schottky detector diodes are already designed and optimized by manufacturer for use in small signal (Pin >-20 dBm). They are ideal for RFID and RF Tag applications where primary (DC bias) power is not available. [13][14]

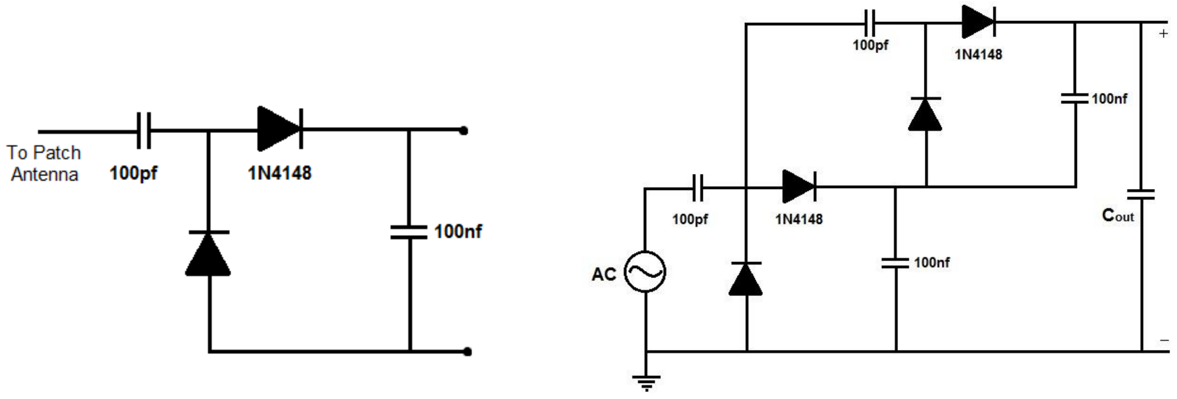

a) **b)**

Fig. 2: a) Single stage diode detector. b) Two stage diode detector with output capacitor.

The fig. 2 above presents the two types of diode detectors on of which (b) is full wave rectifier (known also as a two stage diode voltage doubler). Such a circuit offers several advantages. First the voltage outputs of two diodes are added in series, increasing the overall value of voltage sensitivity for the network (compared to a single diode detector). Second, the RF impedances of the two diodes are added in parallel, making the job of reactive matching a bit easier. [14]

The number of rectifier stages has a major influence on the output voltage of the energy harvesting circuit. [17] Each stage here is a modified voltage multiplier, arranged in series. The output voltage is directly proportional to the number of stages used in the energy harvesting circuit. However, practical constraints force a limit on the number of permissible stages, and in turn, the output voltage. Here, the voltage gain decreases as number of stages increases due to parasitic effect of the constituent capacitors of each stage, and finally it becomes negligible.[13]

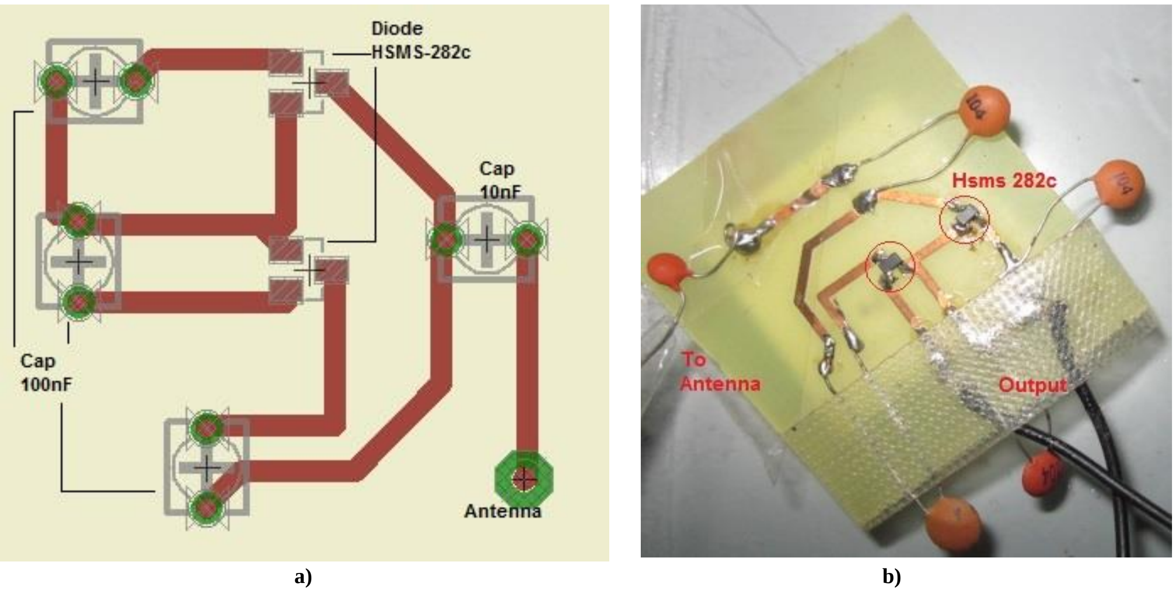
Diode
HSMS-282c
Cap
10nF
Cap
100nF
Antenna
Hsms 282c
To
Antenna
Output

a) b)

Fig. 3: a) PCB layout for voltage multiplier b) Fabricated two stage voltage multiplier with HSMS-282c.

2.2. Rectenna

The word rectenna is composed of rectifying circuit and antenna. The rectenna and its word were invented by W. C. Brown in 1960's. [2] The rectenna can receive and rectify a microwave power to DC, is passive element with a rectifying diode, operated without any power source. The antenna of rectenna can be any type such as dipole, Yagi-Uda antenna, microstrip antenna, monopole, coplanar patch, spiral antenna, or even parabolic antenna. The rectenna can also take any type of rectifying circuit such as single shunt full-wave rectifier, full-wave bridge rectifier, or other hybrid rectifiers. The circuit, especially diode, mainly determines the RF-DC conversion efficiency; rectennas with FET or HEMT appear in recent years. (The rectenna using the active devices is not passive element). [2]

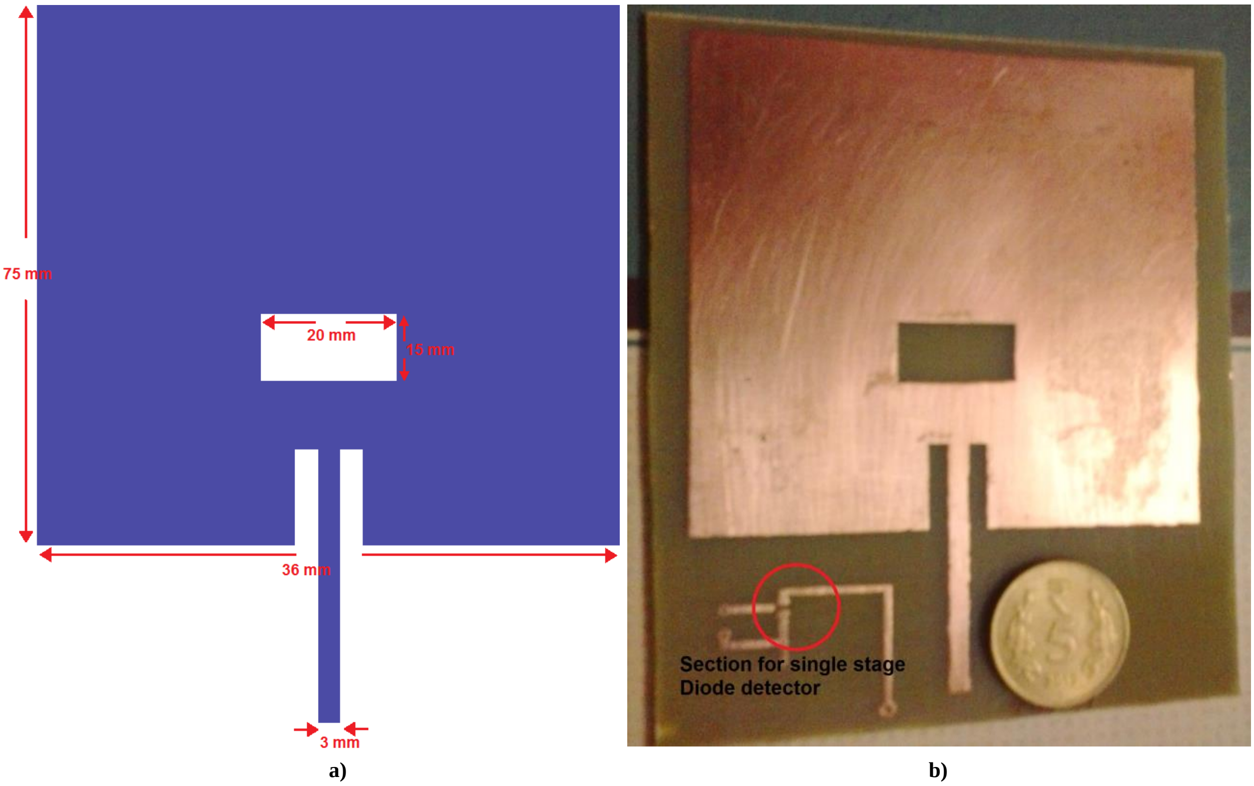
75 mm
20 mm
15 mm
36 mm
3 mm
Section for single stage
Diode detector

a) b)

Fig. 4: Modified Inset-feed Patch Antenna with Dimensions a) CAD design with dimension b) Fabricated patch on FR4

A circular polarized microstrip patch antenna in fig 4 a) and b) above (86 mm x 75 mm) has been associated with the rectifier to obtain the complete rectenna design. The antenna simulation was carried out using the electromagnetic simulator Ansoft HFSS v.11 [9]. The closest to desired resonant frequency has been achieved and matched in a single simulation. It showed a good input matching level at 886 MHz.

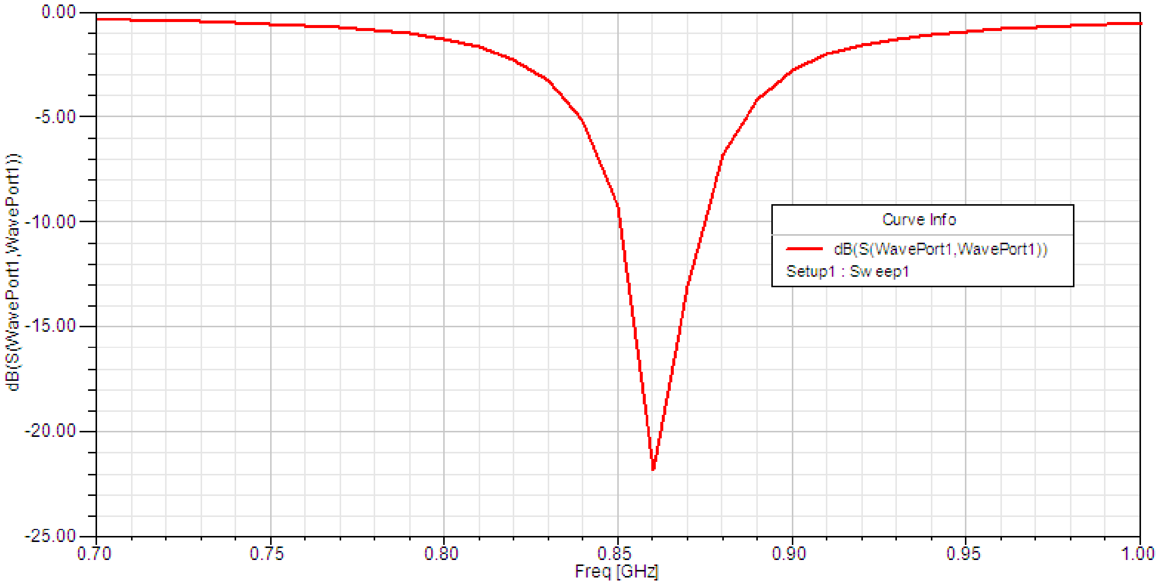

Figure 5: Return Loss (S_{11}) of Microstrip antenna.

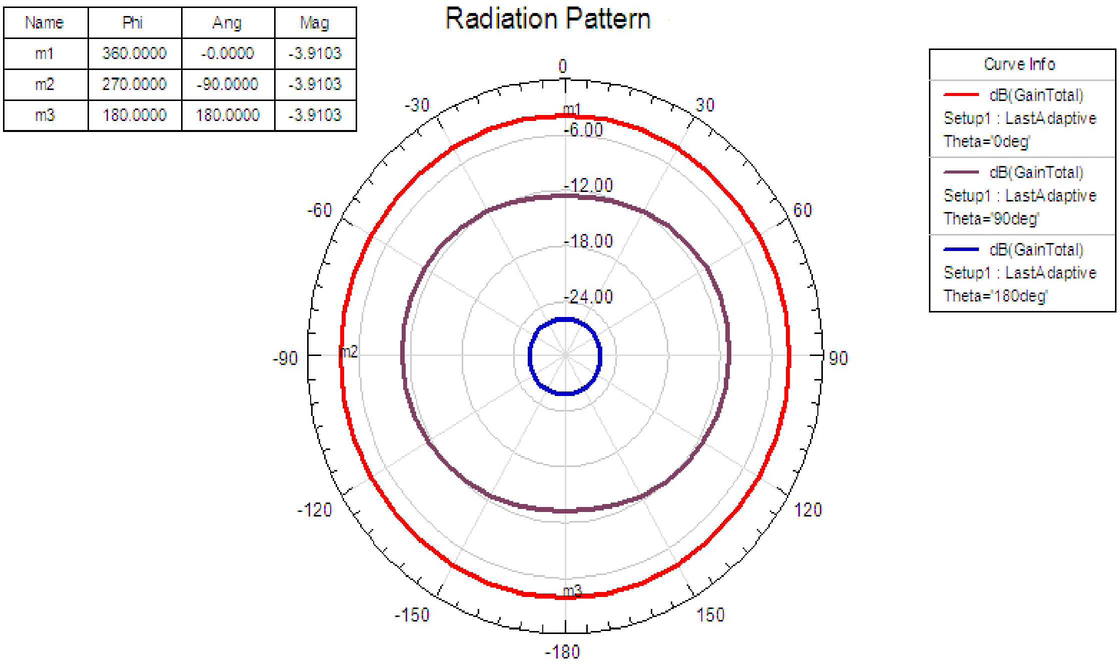

Name	Phi	Ang	Mag
m1	360.0000	-0.0000	-3.9103
m2	270.0000	-90.0000	-3.9103
m3	180.0000	180.0000	-3.9103

Figure 6: Radiation Gain plot of Microstrip antenna.

After simulating the structure figure 5 above showing return loss (S_{11}) of the patch antenna structure .The plot gives return loss of ~21.9 dB at 886 MHz. Figure 6 is the radiation gain plot of the microstrip antenna at 886 MHz from three major angles of 0, 90 and 180 degree which is showing the maximum gain as -3.91dB.Here point t ponder is although the bandwidth of antenna is ~25 MHz still it could be used to beyond that band. The problem that will occur using this antenna beyond its resonant band is mismatched impedance and it will not produce maximum/efficient output.

The purpose of the inset cut in the patch is to match the impedance of the feed line to the patch without the need for any additional matching element. This is achieved by properly controlling the inset position. Hence this is an easy feeding scheme, since it provides ease of fabrication and simplicity in modeling as well as impedance matching. However as the thickness of the dielectric substrate being used, increases, surface waves and spurious feed radiation also increases. The feed radiation also leads to undesired cross polarized radiation. [17][11]

The three essential parameters for the design of a rectangular Microstrip Patch Antenna are:

- Frequency of operation (fo) : 886 MHz
- Dielectric constant of the substrate (εr) : 4.4
- Height of dielectric substrate (h) : 1.6 mm.

The parameters obtained using the formulas for inset feed [11] are:

- Total width of the antenna (W) : 86 mm
- Total length of the antenna (L) : 75 mm
- Effective Dielectric Constant (εreff) : 3.9262

At higher frequencies such as 800 or 1900MHz the low values of inductors are difficult to construct especially at board level circuit design. But using the inductor along with capacitor at integrated circuit level design greatly improves the performance. Resonant frequency is also influenced by diode capacitance as it is related with reverse diode voltage and input voltage. [6][23]

In order to test single and double stage voltage multiplier output both are fabricated on a Copper clad sheet (FR4).the single stage detector is embedded with the patch antenna as in fig.7 a) while the two stage diode detector is attached with patch antenna on it back side can be seen in inset of fig.7 b).

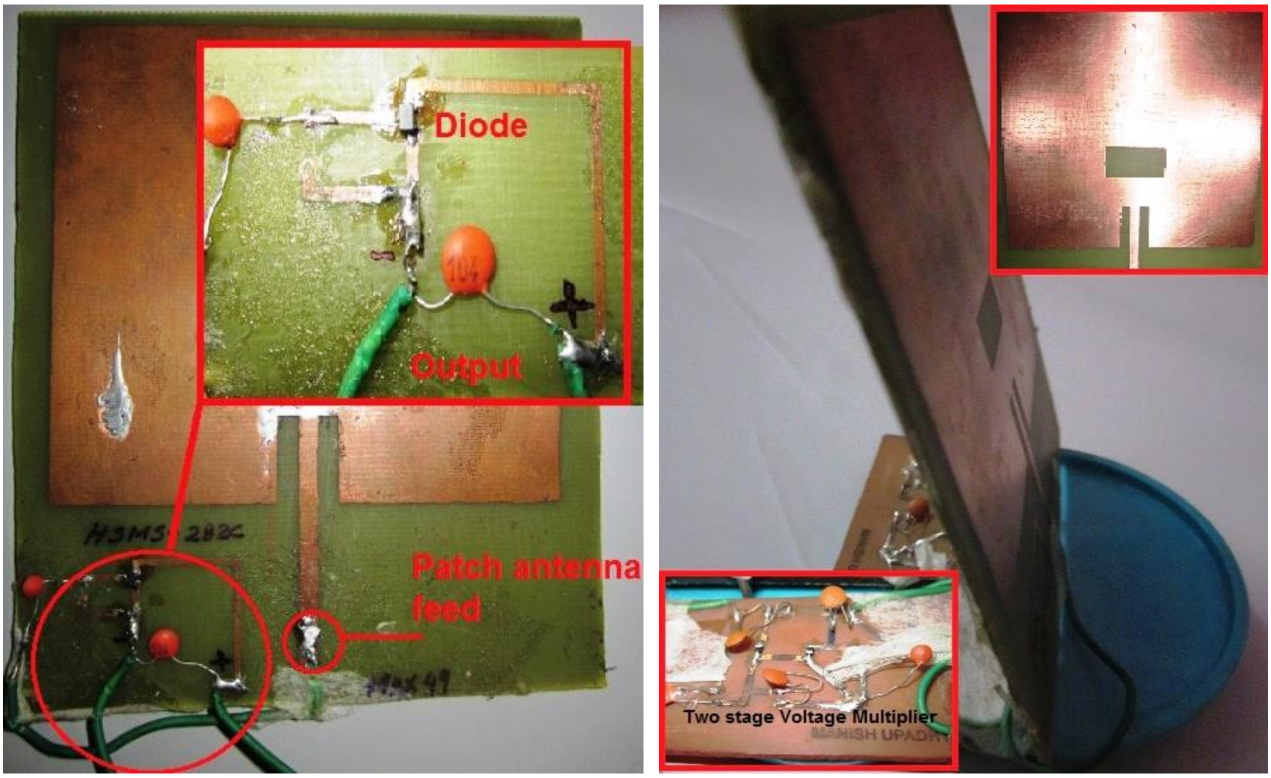

a) b)

Figure 7: Fabricated patch antenna with a) single stage diode (HSMS-282c) detector b) two stage diode (BAT-16) detector (voltage multiplier).

Some effects were previously observed by many researchers is that the circuit yields the optimal efficiency at a particular load value, that is, the circuit's efficiency decreases dramatically if the load value is too low or too high. Since the energy harvesting circuit consists of diodes, which are nonlinear devices, the circuit itself exhibits nonlinearity. This implies that the impedance of the energy harvesting circuit varies with the amount of power received from the antenna. Since the maximum power transfer occurs when the circuit is matched with the antenna, the impedance matching is usually performed at the particular input power. [17]

The RF signals can be captured using Multiband antenna such as quad band are easily available in market and usually work at 900MHz/1800MHz/1900MHz/2.4GHz. These are of usually whip type, but small size such as printed, patch, spiral antennas are also testable. A tuning capacitor (C_{tune}) is used to capture the maximum power at required frequency range. Using only Clamp capacitor in antenna matching circuit we can tune the antenna at its resonant frequency. The tuning Capacitor (C_{tune}) can be verified by using the following formula to resonate with antenna inductance ($L_{antenna}$), where f is the frequency of operation.

$$2\pi f = \frac{1}{\sqrt{L_{antenna} \times C_{tune}}} \qquad 1.$$

f = Reasonant Frequency
C_{tune} = Tuning Capacitor
$L_{antenna}$ = Antenna Inductance

3. Test and Observation discussion

As per the block diagram representation the system required a (CDMA-GSM) RF source with known power level that is served by BTS tower near the lab. It is well known that in India the mobile tower have major band of operation in 877MHz-915MHz. For measurement it needs to determine the power at specified distance that is measured by Spectrum Analyzer and my four stages diode detector using 1N4148 is used earlier. Antenna for transmitting is assumed omni-directional whip antenna. While the antenna used in case of measurement of standard emission by RF source is measured using helical wire (unmatched) antenna on spectrum analyzer port and a copper patch for IN4148 diode detector.

Since the test receiver/detector used was a very low profile diode detector using 1N4148 (Semiconductor-semiconductor junction) which is having forward voltage drop ~620mV at 1mA and 1V at 15mA which is not an efficient value for the required application. For the final tests 1N4148 design is neglected and Schottkey Diodes (zero bias) like HSMS-282C AND BAT-16 that is designed for having very low forward voltage drop (approx. 340mV & 580mV respectively) and high switching speed is used.

3.1. Testing at external site

Setup shown in fig. 8 is a copper patch attached with two stage diode voltage multiplier to test electric field strength in vicinity of mobile phone tower. In order to make comparison of the electric field strength (V/m) near mobile phone tower antenna one needed to calculate ideal value of power that should be received and then the practical value measurement.

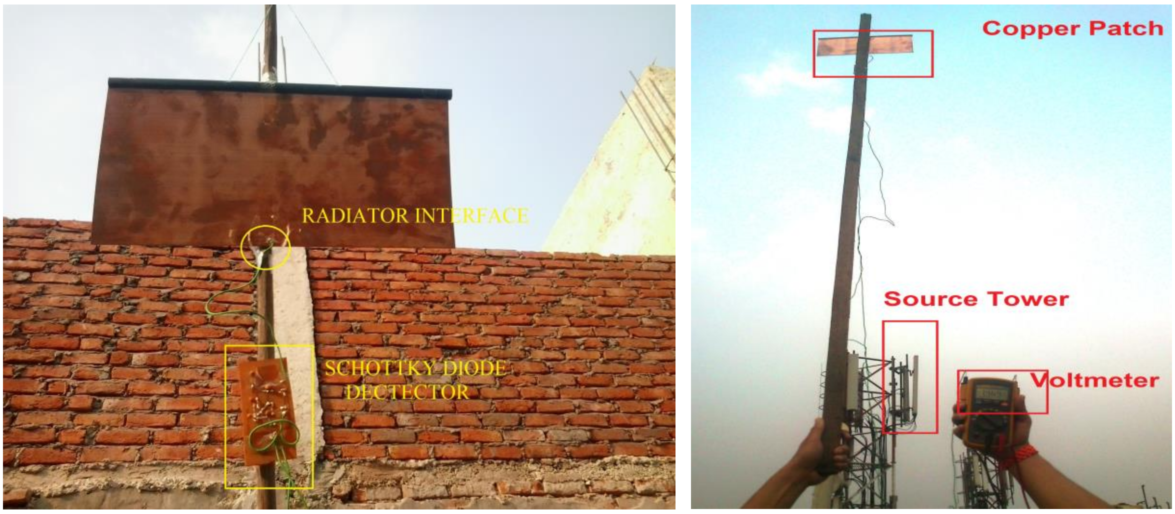

a) **b)**

Fig 8: Rectenna
a) using two stage schottkey diode detector b) Testing at external site (cell tower)

For calculation of ideal radiation figures/component certain parameters are assumed (on the basis of previous presented reports and data from Indian ministry of telecommunication site). These considerations could have deviation from standard one which may be the possible cause of difference reflected in practical measurement. Below present equation 2 to 7 are standard equations that are used with FRISS Model to calculate ideal values and to convert measured value. [17]

Parameters:

Frequency of interest (f) = 900 MHz (Between CDMA and GSM band)

Largest radiator dimension (D) = approx. 1 meter

Wavelength in meters (λ) = 0.33 meters

Typical Base station power (single carrier per operator) (Pt) = 10 watts/2 watt (amended Indian limit for radiation).

Typical Transmitter Antenna gain (Gt) = 17 dB

Receiver Antenna gain assumed to be (Gr) = -3.9 dB = -1.75 dBi (calculated with simulation model)

$$\mathrm{P_r} = Ae.\mathrm{E}.\mathrm{H} \quad 2.$$

$$Ae = g\frac{\lambda^2}{4\pi} \quad 3.$$

$$E = Z_o.\mathrm{H} \quad 4.$$

$$Z_o = 120\pi = 377\Omega \quad 5.$$

$$S = \frac{E^2}{Z_o} \quad 6.$$

$\mathrm{P_r}$ = Power received (W)
Ae = Antenna aperture (m)
E = Electric field strength (V/m)
H = Magnetic field strength (A/m)
λ = Wavelength (m)
g = Antenna gain
Z_O = Impedance (Ohm)
S = Power density (W/m2)

The **Fraunhofer** distance (d):

$$d = \frac{2.\mathrm{D}^2}{\lambda} = 6m \quad 7.$$

During measurement practice its is to be ensure that measurement is taken by maintaining minimum distance from cell tower because the Friss model of calculation that is to be used to estimate power received valid over far field only. The observations are presented in the table 1.Some quick conclusion that could be made is the test site reading are closer to ideal value one with the 2W/carrier standard.

Distance from Source	Power & Electric Field density (Ideal)			Power & Electric Field density strength Received (Test site)		
15 meter	177.3 mW/m^2	8.16 V/m	18.94 dBm	15 mW/m^2	2.44 V/m	-10.31 dBm
	35 mW/m^2	3.63 V/m	11.90 dBm			
30 meter	44 mW/m^2	4.07 V/m	12.89 dBm	6.7 mW/m^2	1.62 V/m	-13.97 dBm
	8.8 mW/m^2	1.82 V/m	5.90 dBm			
50 meter	15.9 mW/m^2	2.35 V/m	8.22 dBm	3.2 mW/m^2	1.1 V/m	-17.23 dBm
	3.1 mW/m^2	1.08 V/m	1.37 dBm			

Table 1: Test results comparison with ideal calculated values while mobile tower as RF source and two stage diode detector as receiver. Ideal transmitted power values correspond to RED = 10W/carrier & BLUE = 2W/carrier

3.2. Testing with the combined system

The combined system is presented in fig 9 a) which includes a glass container used to house two patch antennas connected with separate voltage multiplier which can be seen in fig 9 b) are connected to arduino uno (open source hardware platform with atmega-328) to its analog input pins. The measuring rectenna is mounted over a servo motor, which is controlled by the monitoring unit (microcontroller atmega-328).The voltmeter which is designed to measure electric field strength (V/m) is connected to the output of measuring antenna. The data to the analog pin of arduino uno is converted to equivalent digital bits and fed to laptop. The Matlab 2012a program on the laptop will collect data from arduino uno.

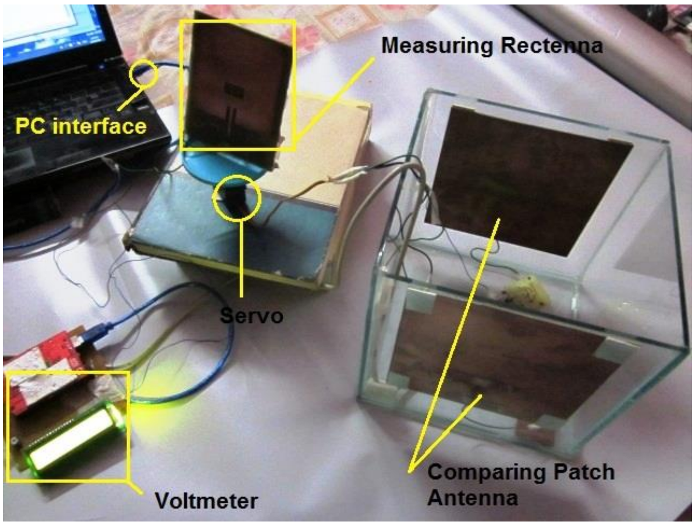

a)

When an RF source (CDMA-GSM band) like an operating cell phone with call initiated is brought near the measuring rectenna will result into an output voltage on voltmeter (field strength) as shown in fig 9. A program file written in the Matlab 2012a software which is for comparing output from the monitoring unit if executed, will start the arduino uno (monitoring unit); patch antenna on both the side of glass structure will start working to provide the strength of nearby field which will be input to the analog pins of arduino uno. Further the monitoring data will be fed to Matlab 2012a program executing on its way. The program will start plotting a real time graph of two voltages (for both the side of glass structure); on the basis of the relative strength of both antennas the program will command servo via arduino uno to move/target measuring rectenna toward the major RF source. In Fig. 9 c) the cell phone with a call initiated brought near the measuring rectenna, which result into an output of ~4.9V as pulsating DC is displayed on voltmeter LCD. Once the program is manually started in Matlab 2012a its activate both the patch antenna attached with monitoring unit as a result we can see the fig. 10 .The patch antennas are marked as right and left are assigned with color marker in Matlab 2012a graph plot. In Fig. 10 we can see that initially voltage plot is high due to the cell phone setting up communication with nearest tower. Another point we can notice is that the cell phone and the rectenna facing to the left side of the glass structure result into the left patch receiving most of the RF radiation.

Once we move RF source (cell phone) away from left side result into decreasing strength of RF radiation recorded by the blue line (left patch antenna) on the plot Eventually moving toward the right of the glass structure rise the RF strength recorded by right Patch antenna which will result into high red line strength on the plot. When the red line (right patch) record higher strength than a blue line (left patch),

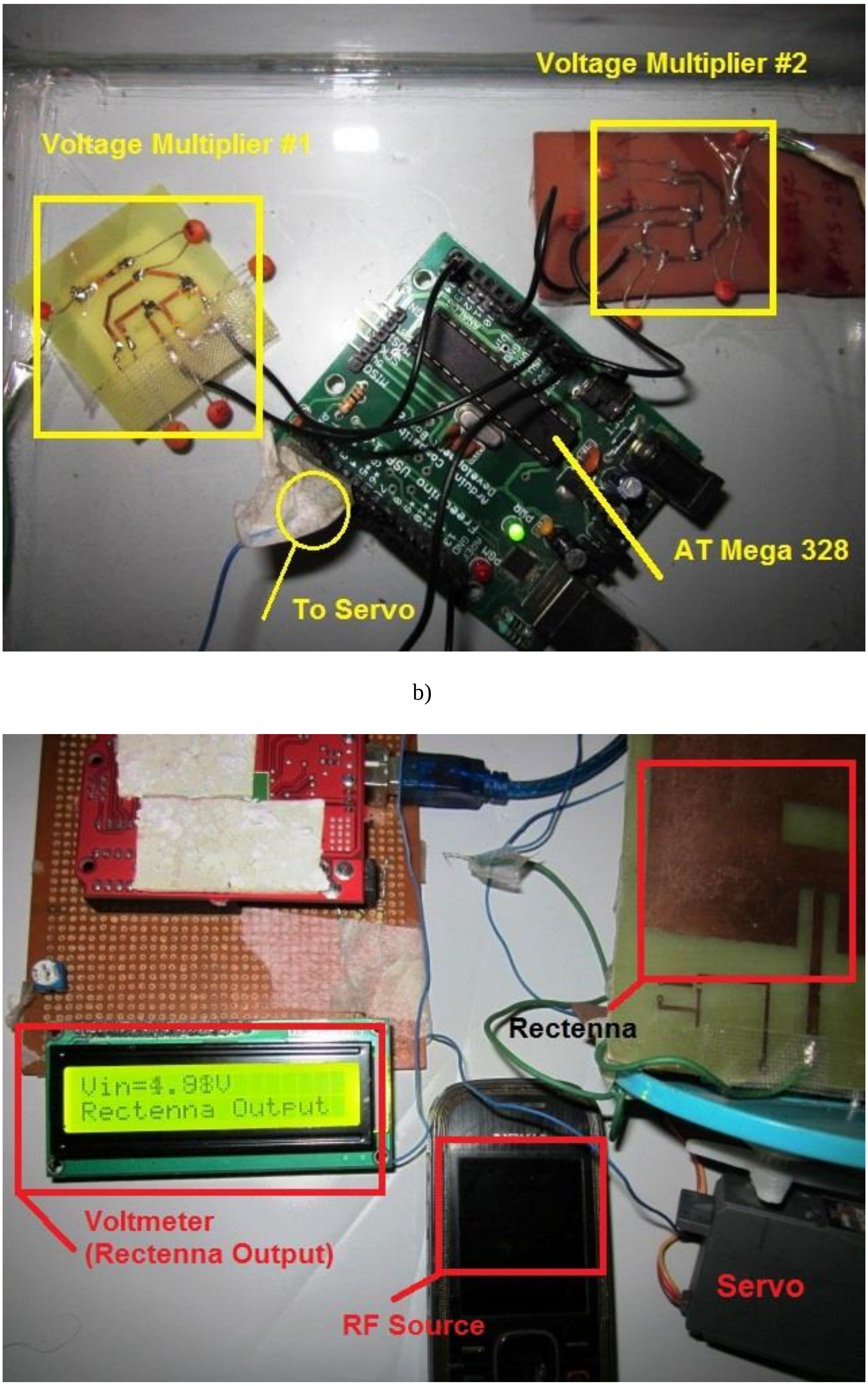
Voltage Multiplier #2
Voltage Multiplier #1
AT Mega 328
To Servo
Vin=4.98V
Rectenna Output
Rectenna
Voltmeter
(Rectenna Output)
Servo
RF Source

b)

c)

Figure 9: Combined system testing a) complete prototype operation, b) circuit inside glass structure, c) measuring rectenna output on voltmeter.

The program will command servo to move measuring rectenna position (facing) from left to Right side .In this way, system will keep on comparing the strength and will automatically move measuring antenna toward more RF strength.

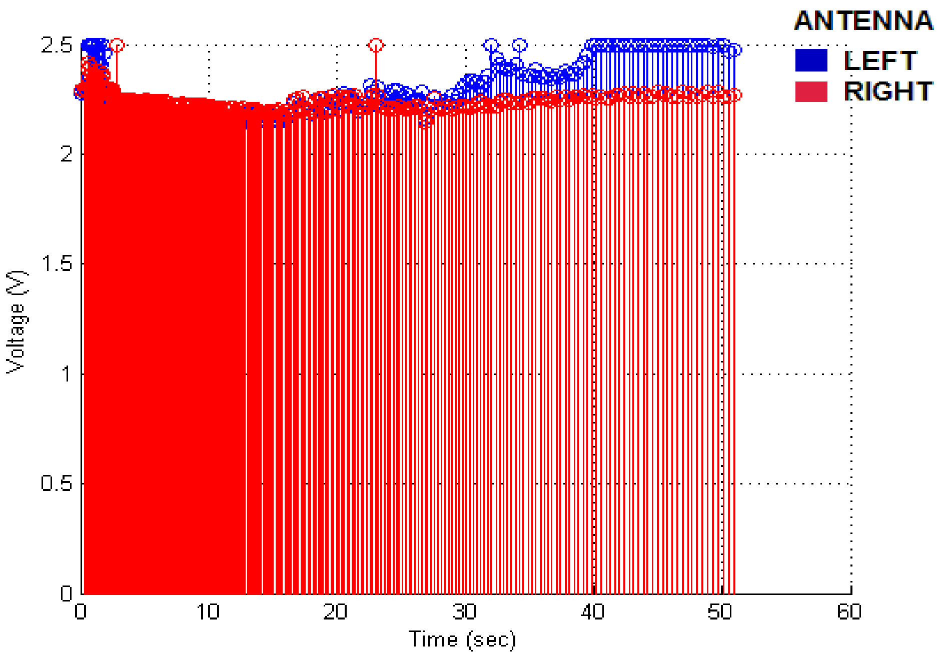

Figure 10: Plot for patch antenna on left (blue) and right (red) side of monitoring unit.

3.3. Powering LCD thermometer module

In fig 11 above presented a practice to power up an LCD thermometer module using the rectenna and a cell phone. As in fig 11 b) we can see that the battery from the thermometer module is replaced by rectenna output. Here in this case the wire connected to the battery section of the thermometer module is nothing but the output of single stage zero biased diode detector using (HSMS-282c), which will convert RF current to DC. When we brought a cell phone with an initiated call result in rectenna to provide dc voltage to the thermometer module as could be seen in fig. 11 a) the LCD on thermometer module start operating. Once the call terminated the rectenna stopped providing dc output result in the power to Thermometer module get down in very short span. Although the thermometer is able to operate for the very short span during an operating cell phone call still it is verified that power is enough to operate module but a consistent source of RF energy or a capacitor with low leakage and possibly with high discharge time.

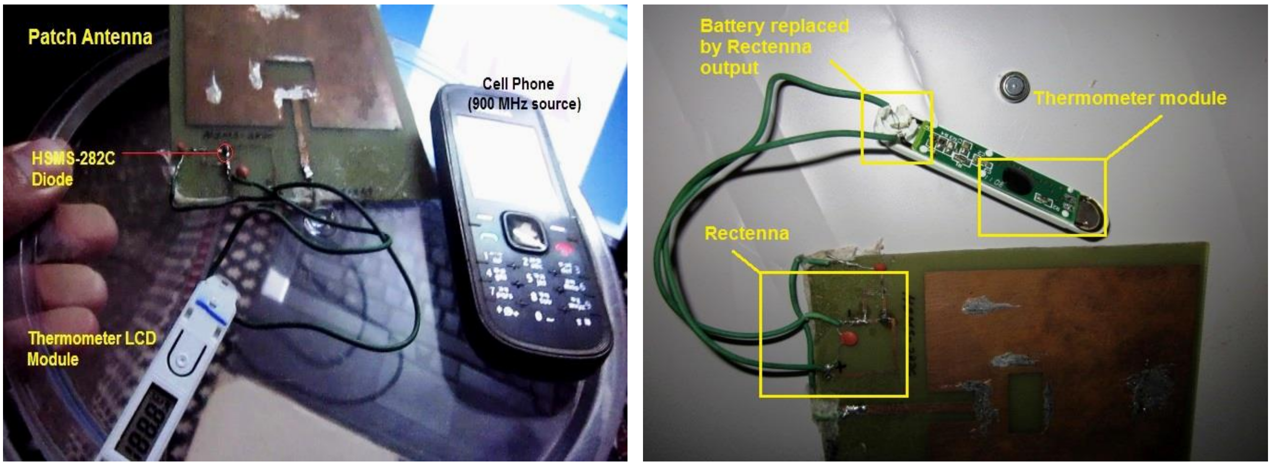

a)

b)

Figure 11: LCD thermometer module a) powered by rectenna using cell phone as RF source, b) Battery is replaced by rectenna output.

4. Conclusion

A low profile RF (CDMA-GSM) energy measuring and harvesting system has been analyzed, designed and tested. The novel idea is to track/target the RF energy in real time to measure and harvest it. Voltage multiplier configurations, has been verified and optimized by means of test series conducted in lab and external site. Although no input low pass filter is used for rejection of harmonics at the input power of the rectifier still results gives good agreement to theoretical calculations.

A low profile inset-feed modified patch antenna is developed and used. A silicon based schottky diode single stage and 2 stage rectifiers are designed and fabricated on copper clad sheet (FR4). Comparison of ideal and RF energy measured with rectenna at test site with cell towers is presented.

Low series resistance Schottky diode in single stage detector is used to implement rectenna which could successfully power up the LCD thermometer. Although the power is not consistent for long use but system could be improved using super capacitor as output load of rectenna to operate low power devices.

Finally such system could be developed to be prove effective in charging of mobile handset (primary device) by placing to be charged mobile handset (secondary device) very close to fully charged mobile handset and draining primary device energy for charging the secondary device. Higher performance can be achieved by implemented the circuit in low sub micron manufacturing processes. Output voltage level can be increased using multiple stages of the multiplier circuit.

Acknowledgements

This project is using the part of the study carried out by me for the thesis submitted to GGSIPU, Delhi. I would like to express my thanks to the Asst. Professor Mr. S.B. Kumar and Ms. Jolly Parikh of Dept. of Electronics and Communication Engineering, BVCOE, New Delhi, for their guidance and support throughout this study.

References

[1] Nicolas Tesla. Experiments with Alternate Current of High Potential and High Frequency, McGraw, 1904.

[2] William C. Brown. “The history of power transmission by radio waves”, IEEE Trans.MTT, 32(9):1230-1242, 1984.

[3] N. Shinohara.”Wireless power transmission for solar power satellite” (sps second draft). 2006. http://www.sspi.gatech.edu/wptshinohara.pdf.

[4] W. C. Brown. “An experimental low power density rectenna”. IEEE MIT-S Int. Microwave Symp, 1991.

[5] David M. Pozar. “Microwave Engineering”, John Wiley &Sons, Inc. 1998

[6] J. A. G. Akkermans, M. C. van Beurden, G. J. N. Doodeman, and H. J. Visser, “Analytical models for low-power rectenna design,” IEEE Antennas Wireless Propag. Lett., vol. 4, pp. 187–190, 2005.

[7] J. Zbitou, M. Latrach and S. Toutain, “Hybrid Rectenna and Monolithic Integrated Zero-Bias Microwave Rectifier”, IEEE Trans. on Microw. Theory and Tech., vol. 54, no.1, pp 147-152, Jan. 2006.

[8] A. Douyere, J.D. Lan Sun Luk and F. Alicalapa, “High efficiency microwave rectenna circuit: modeling and design,” Electronics Letters, vol. 44, no. 24, Nov. 2008.

[9] Ansys HFSS Software Version 10, Ansys Corporation.

[10] Hewlett Packard, Technical Data “Surface Mount Microwave Schottky Detector Diodes”.

[11] JAMES J.R., and HALL P.S., Handbook of Microstrip Antennas, Peter Peregrinus Ltd., London, UK, 1989.

[12] Balanis, C.A., ‘Antenna Theory: Analysis and Design’, Third Edition, John Wiley & Sons, Inc, 2005.

[13] Qijuan He and Changjun Liu. “An Enhanced Microwave Rectifying Circuit Using HSMS-282x”, School of Electronics and Information Engineering, Sichuan University, Chengdu 610064, China.

[14] Stewart Jenvey, Jonas Gustafsson and Fredrik Henriksson “A Portable Monopulse Tracking Antenna for UAV Communication”, in 22nd International Unmanned Air Vehicle Systems Conference, April 2007.

[15] Avago Technologies. “Surface Mount Zero Bias Schottky Detector diode” Technical Data Sheet.

[16] Prof. Girish Kumar (IIT-B) “Report on cell tower radiation”, to Dept. of Telecomm. Delhi, Dec 2010.

[17] M. Guardiola, B. Monsalve, I. Calafell and G. Roqueta “Eco-Friendly Approach for Young Intrepid Students on the Fabrication and Measurement of Home-Made Standard Antennas”, IEEE Antennas and Propagation Society, DIY team ,May 2010.

[18] P. Nintanavongsa, U. Muncuk, D.R. Lewis and K.R. Chowdhury “Design Optimization and Implementation for RF Energy Harvesting Circuits, Vol. 2, March 2012.

[19] Fairchild Semiconductor Corporation, “1N4148 Small Signal Diode - Rev. B.

[20] Source for number of BTS: http://telecomtalk.info.

[21] Omar André Campana Escala, “Study of the efficiency of rectifying antenna systems for electromagnetic energy harvesting”, Pontificia Universidad Católica del Perú, Oct 2010.

[22] Richard Patterson, Ahmad Hammoud, Malik Elbuluk, Scott Gerber, “Performance of Various Types of Diodes at Cryogenic Temperatures”, NASA Electronic Parts and Packaging Program.

[23] M.Arrawatia, M.S.Baghini, Girish Kumar, “RF Energy Harvesting System from Cell Tower in 900MHz Band”, Indian Institute of Technology Bombay, India.

[24] Special-purpose diode, Volume-III semiconductors: www.allaboutcircuits.com/vol_3/chpt_3/12.html

COMMSCOPE®

UNDERSTANDING THE RF PATH

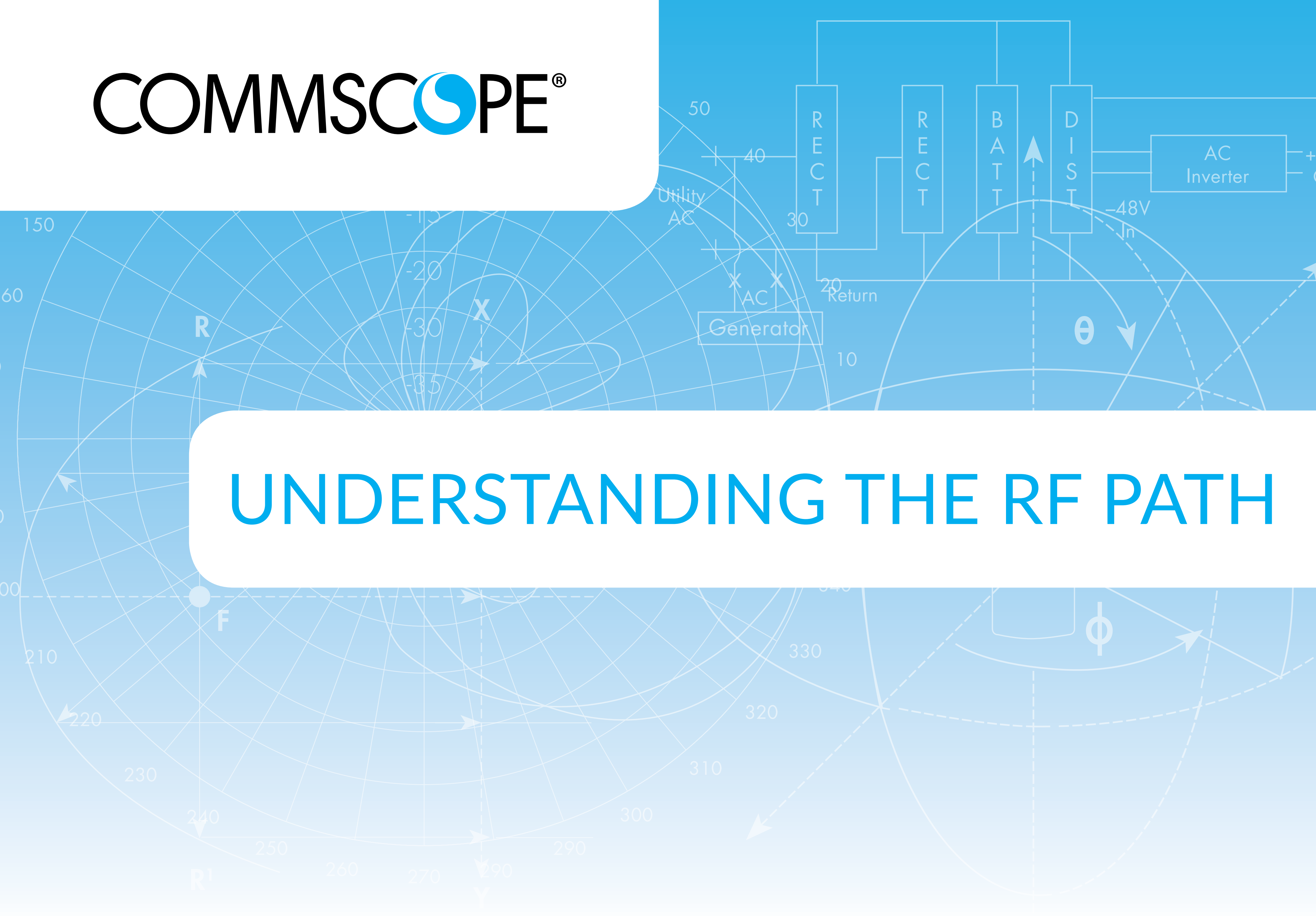
50
40
30
20
10
150
60
210
220
230
240
250
260
270
290
300
310
320
330
-15
-20
-30
-35
R
X
F
R'
Y
Utility
AC
AC
Generator
RECT
RECT
BATT
DIST
AC
Inverter
–48V
In
Return
θ
ϕ

CommScope: understanding the RF path

Chapter 1	**Introduction: welcome to RF communications**	**1.1**
Chapter 2	**Where RF theory meets best practices: cell site development and construction**	**2.1**
Chapter 3	**Getting the signal across: base station antennas**	**3.1**
Chapter 4	**Working within the limits: co-siting solutions**	**4.1**
Chapter 5	**Talking and listening at the same time: transmission and receiving isolation systems**	**5.1**
Chapter 6	**The many ways to connect: RF transmission lines**	**6.1**
Chapter 7	**Making every connection count: passive intermodulation (PIM) fundamentals**	**7.1**
Chapter 8	**The infrastructure behind the connection: microwave backhaul**	**8.1**
Chapter 9	**The energy that connects the world: powering wireless networks**	**9.1**
Chapter 10	**Successfully planning against failure: reliability in wireless systems**	**10.1**
Chapter 11	**Covering all the bases: distributed antenna systems**	**11.1**
Chapter 12	**Going to ground: lightning protection**	**12.1**
Appendix A	**Spectrum configurations around the world**	
Appendix B	**Biographies**	
Appendix C	**Glossary**	

Chapter 1

Introduction:
Welcome to RF communications

As a trusted advisor for communications networks around the world, **CommScope** invests in people as much as our products and our industry. Our comprehensive training and education programs, participation in working groups on specs and standards, educational conferences, and extensive research are long-standing traditions that benefit everyone in our industry.

That's why we're pleased to present this book on the fundamentals of radio frequency (RF) communications for the wireless industry. We hope this valuable information will help foster a greater understanding of, and appreciation for, the technology, science and business behind modern RF communications.

The technologies and theories explored here are technical in nature, so we have made every effort to make the science of radio systems more accessible for a wider audience—whether you are an engineer or simply someone who works in this industry and wants to learn more about the basics of how RF communication works, including the issues involved in the planning, deployment and maintenance of communication networks.

Let's start by taking a look at the storied history behind the technologies that have shaped our on-the-go, wireless world.

RF systems, then and now

Wireless communications systems have been in commercial use since the 1940s. Some of the earliest implementations included community repeaters, paging systems, point-to-point links and specialized mobile radio (trunked) systems. More recent innovations and uses of the RF spectrum include cellular radio networks, which originated in the 1980s and now drive the cell phone and mobile device industry. Even more recent evolutions have given us familiar names like WiFi and WiMax, which let you connect to a network anywhere in your house, or virtually anywhere you go.

The shared link

This storied list of applications is incredibly diverse in purpose and design, yet every item shares at least one unique characteristic: they all utilize radio frequencies between 30 MHz and 2.6 GHz to move information between base locations and remote users.

The chain of components required to make this movement of information possible is a complex and varied one. As a proud supplier to the world's communications networks, **CommScope** is uniquely able to explore and explain these components, how they work, and more importantly, how they work together.

Doing our part

CommScope goes to great lengths to provide the best possible solution for every application. For simple or complex RF communications systems, each part chosen by an engineer must fulfill a specific role, determined by the expectations and limits placed on the installation itself. This book will explore several common components and how they interact, including:

- Antennas
- Coaxial cables
- Filters
- Duplexers and diplexers
- Amplifiers
- Remote radio heads
- Enclosures
- Power backups

... and many others

RF communications: the early years

While the full story of RF communications is far from complete, the capabilities of modern RF technologies exceed the wildest dreams of the field's earliest pioneers. While we are all familiar with—and indeed, take for granted—living in a globally connected world, the early systems that laid the foundation for our networked existence were much less powerful.

The first RF systems featured a base radio using an omnidirectional antenna to communicate with one or more mobile users. Then, as now, the effective coverage radius of that base radio was limited by certain characteristics, such as RF power, antenna height, and the sensitivity of mobile receivers that were vulnerable to thermal noise and other interference sources. These systems were also limited by the fact that certain frequencies could only be used once in a particular geographical area. Once a mobile user left that area, no communication was possible.

The engineers of AT&T Bell Laboratories envisioned a future that would require much higher RF capacities to service thousands or millions of users at once. To deliver this future, they developed the cellular concept: a wireless network that uses lower antenna heights and transmission power levels to create limited-radius coverage areas that used and reused the same frequencies within its coverage area. Voice and data calls could be seamlessly handed off to neighboring cells as the user moved from one coverage area or cell to the next. The close-proximity reuse of radio channels is the fundamental concept of cellular telephones, and this is how today's wireless networks carry enormous traffic loads in spite of limited spectrum. The data riding these RF frequencies could be voice, data, or a combination of both; by intelligently managing and reusing the limited available bandwidth, the amount of available service increased by many orders of magnitude.

The business case for technical expertise

Cell providers monitor key performance indicators within their territories to identify coverage problems and assure customer satisfaction. These indicators include quality of service (QOS), dropped calls, failed access attempts, and other criteria. Their engineers are on the front lines of this battle for quality and constantly optimizing network performance as traffic grows. There are sophisticated parameters and advanced settings that help the RF engineer optimize a wireless network. However, successful wireless network optimization requires a solid foundation of physical components across the RF path (figure 1.1).

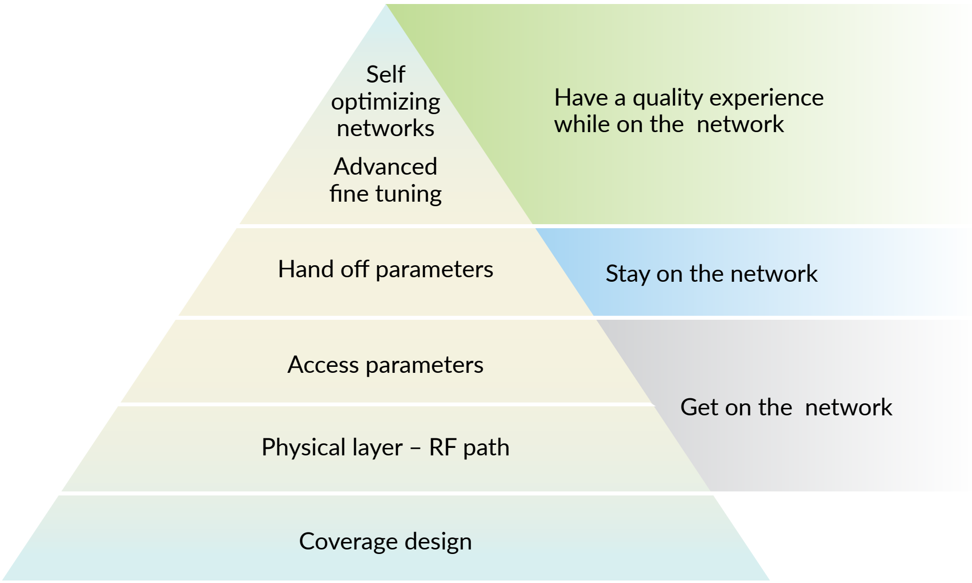

1.1: The interconnection of technology, design and optimization

The tuning and optimization of parameters can never be maximized without a high-performance RF path implemented with top-quality RF components. Conversely, wireless systems with weak RF paths perform poorly, resulting in expensive problems on both the operational and customer service fronts. Once a network is active, taking down a part of a cell network for maintenance is a costly and disruptive interruption; and, in most cases, one that could have been prevented by a better understanding of the site's requirements and the component choices available.

The path begins

Now that we understand the importance of foresight in planning, and insight in selection, we can begin this brief exploration of the many ways a modern RF communications network can come together. Thank you for joining **CommScope** on this journey.

Chapter 2

Where RF theory meets best practices:

Cell site development and construction

Building a new cell site raises some of the same questions as building a new house, such as deciding what materials to use, who to entrust with the construction, and how to get the best results for your money.

Of course, these problems are magnified when applied to cell site construction because there are so many different people, companies, municipalities and regulatory agencies involved. It takes a lot of experience to plan for every obstacle. That's why **CommScope** offers the following proven recommendations.

These general guidelines are the result of decades of successful, practical field work. While no guide can cover every aspect of cell site development, we'll cover the most commonly experienced technical issues here.

Service company
A cell site development partner responsible for actual construction on the site, including antenna towers, concrete footers and pads, security fencing, and equipment shelters.

You will work with several different companies and agencies in the process of developing your new cell site. You must be able to count on their competence and expertise. For each stage of development, you should become familiar with the challenges and requirements each partner will face. Knowing these factors will help you judge how well your partners are doing their work and the impact those factors will have on your network when the project is done.

Step one: choosing a service company

A ***service company*** will help you build the actual structure of your cell site. This includes the tower, shelters, cabinets, and other physical infrastructure to support the site's operation (figure 2.1). This can involve heavy, demanding construction work and the precision required by modern cell networks means a lot can go wrong at this stage.

2.1: A services company building a cell site tower

As with building any structure, it pays to research your construction company in advance. Here are some of the key facts you should learn before selecting your services company:

- **How long has the company been in business, and in the wireless field?** This is one of the first questions to ask. Experience matters as much as, if not more than, expertise.
- **How many employees do they have? How many of them work in the field and how many in their home offices?** Their staffing levels can directly affect how they handle your project.
- **How financially sound is the company?** For legal and logistical reasons, it's critical that they don't disappear in the middle of construction. Plus, they will need to maintain a certain minimum cash flow and credit to procure materials.
- **What are their expected payment terms?** Some require payment within 30 days. Others may allow as many as 90 days. Make sure their terms work with your cash flow, too.
- **What is their recent bond history?** For everyone's protection, all craftsmen in the field should be bonded. Looking at their bond history can give you a good picture of how well they perform.
- **What kind of insurance do they carry?** The company should offer to share their certificates of insurance. It's also smart to request a report on their recent claims history.
- **How does their safety plan hold up?** Unfortunately, some companies keep a generic plan on file simply as a token compliance effort. Speaking with the company's safety manager will tell you if they really understand the written plan.
- **What's in their OSHA logs?** The Occupational Safety and Health Administration (OSHA) monitors workplace safety; your potential services company should have a log of recent incidents, claims and actions. As a minimum, the previous three years should be reviewed for diligence and completeness.
- **Are their field employees properly certified?** All field employees should carry current certifications for first aid, CPR, the OSHA 10-hour training course, Competent Climber and Tower Rescue operations.
- **Is their gear properly certified for safety?** You should request copies of current climbing gear inspection certificates.
- **How is their workplace recordkeeping?** To see what you could expect for your project, ask to see their job site analysis worksheets.
- **What kind of vehicles will they be using?** If they use vehicles regulated by the Department of Transportation (DOT), request a copy of the DOT Carrier Safety Measurement System (CSMS) rating. Prior to 2011, this rating was called SafeStat, and some companies may still refer to it by that name.

In general, finding the answers to these questions is simply a matter of thorough due diligence. With so many certifications and other qualifications in play, you can see how important it is to know your services company partner as well as possible.

Coaxial cable
A type of cable featuring an inner conductive core, an outer conductive layer, and a dielectric, or insulating, space between them. Coaxial cable connects antennas to their base stations.

Step two: choosing coaxial cable and connectors

Efficient cell site operation relies on the precise pairing of components. As we will discuss in detail in chapter 3, certain cable types are designed to work with certain antennas. As we will explore in chapter 7, those cables must interface with their systems via connectors built for certain frequencies and power levels (figure 2.2).

An out-of-specification component or an improper installation of the correct component can cripple an entire system. For this reason, it's wise to contract with skilled engineers for the installation.

2.2: Coaxial cable examples; different types are available for different applications

Handling the cable and connectors during these installations is a delicate business. In general, observing these tips can help assure a trouble-free installation and dependable, long-term operation:

- **Use the right tool for the job.** Using the appropriate cable prep tool, usually available from the cable's manufacturer, is the only way to cut and prep cable ends for use in connections. Never use a saw; they leave metal filings behind, which cause poor electrical performance and problems with passive intermodulation (PIM).
- **Watch those tricky curves.** Different cable types have different degrees of allowable bend radii, or flexibility, so you must observe the manufacturer's prescribed bend radius for your particular cable. Bending too tightly can lead to poor electrical performance and failure in the RF path.
- **Keep your cables consistent.** If at all possible, use RF jumper cables from the same manufacturer to make those tight connections. Doing so provides consistent electrical performance and guarantees PIM performance.
- **Ensure proper cable support.** Manufacturers publish specifications describing how to support lengths of cable, both vertically and horizontally. Your specific guidelines will depend on your cable's construction, size and weight. If possible, use support clamps from the same manufacturer to avoid damage to the cable and loss of performance. Using third-party clamps may also invalidate your warranty.
- **Lift smart.** Getting cables up an antenna tower is difficult. Fortunately, using the correct hoisting grip will let you put that cable where it needs to be without damaging it. Hoisting grips come in several types and sizes, so make sure yours matches your cable's specifications.
- **Go to ground.** Grounding the cable is very important to prevent damage from lightning strikes. Best practices dictate at least three grounding points: at the top of the tower, at the bottom of the tower and just outside the entrance to the outbuilding, shelter or cabinet.
- **Put on the pressure.** Air dielectric cables use an air-filled gap to insulate the inner and outer conductors of a coaxial cable (chapter 7). This gap must be pressurized, like a bicycle tire, for the cable to hold its shape and prevent damage to the conductors inside. When installing air dielectric cable, it must be pressurized immediately. Leaving it overnight can lead to moisture condensation in the cable, which will degrade performance and is almost impossible to remove once introduced.
- **Finish with the seal of approval.** Connectors are particularly vulnerable to infiltration by weather and moisture. As soon as the connections are made, you should weatherproof them. Butyl tape is the preferred method, but in tight connection spaces, like those atop the antenna tower, you can opt for heat-shrink tubing applied with a heat gun.

By following these recommendations, you can help ensure that your cell site will operate at peak efficiency with minimal maintenance.

Electrical tilt antenna

An antenna fitted with actuators that can adjust its tilt relative to the ground. Changing tilt affects gain, or performance, of the antenna within defined geographical areas.

Step three: setting and troubleshooting remote electrical tilt antennas

Of the many ways you can improve performance from a cell site's antennas, a particularly effective method is beamtilt. It involves physically tilting the orientation of the antennas below the horizon, placing its greatest gain—its operational power—where it's needed most (chapter 3).

Efficiently changing the orientation of an antenna to realize this benefit is a matter of using electrical tilting mechanisms, or actuators, which can be operated from a remote location (figures 2.3, 2.4, and 2.5).

These mechanisms are controlled by an Antenna Interface Standards Group (AISG) Remote Electrical Tilt (RET) controller, which connects via AISG cables at the cell site for adjustment (figure 2.5).

2.3: An electrical tilt mechanism, or actuator, for a cell site antenna

2.4: A base station antenna (BSA) with integrated RET actuators

2.5: An example of an AISG RET controller module

As precision equipment, these tilting antennas can be a challenge to install and adjust properly. Getting the best result is a matter of understanding the software just as much as the hardware, but there are several ways to avoid common pitfalls:

- **Install the software first.** Before your crew goes to the cell site, install the manufacturer's software and become familiar with the controller's operation. This early training will help your team hit the ground running when they arrive.
- **Check for program updates.** Driver software for your electronic tilt actuator system is constantly updated to work with an ever-growing number of different antennas. Make sure your software is current by checking for updates on the manufacturer's website.
- **Understand the naming conventions.** To prevent onsite confusion, use conventions for the configuration of actuators that everyone will understand.
- **Test before installing.** For new installations, a little upfront effort can prevent big headaches later. Test the actuators, cables and other components before installing them on the tower. It's much easier to address problems when the components, and you, are on the ground.
- **Match antennas and tilts.** Not every antenna has the same tilt range, so be sure you select the correct one from the database before adjusting it. Each antenna's address is based on its product serial number, so be sure to keep a written record. You should double-check your tilts through tab reports generated by the controller.
- **Keep a spare cable on hand.** Bring a spare cable to the site in case you need to troubleshoot a non-reporting actuator. It's the fastest, surest way to tell if the problem is a faulty actuator or just a bad cable.
- **Check before tilting.** Before making any new tilt adjustments, pre-scan the other antennas to determine their tilt values.
- **Double-check your work.** After making the adjustment, you should perform a post-scan to confirm the new settings have been correctly applied.
- **Don't weatherproof cables and connectors.** Using electrical tape won't keep moisture out, in fact, it gives water a place to accumulate in the connector, where it can cause shorts.
- **Protect against lightning**. Lightning protection units should be installed at the base of the tower, or just before the cable enters the shelter or cabinet. Also, as stated above, it should be grounded in at least three locations: at the top of the tower, at the bottom of the tower, and just before entering the shelter or cabinet.
- **Don't splice in a ground lead.** Cutting into the jacket to attach a ground to the thin foil tape inside will cause water migration, damaging the conductors below the foil.
- **Go right to the source for cable.** It's considered good practice to purchase your cable directly from the manufacturer rather than obtaining it from a third party. Each manufacturer's system requires specific electrical conductors, and using a mismatched cable may lead to actuator failure, voiding your warranty.
- **Make the right connections.** The home run cable's male connector—the end with the pins—is the end that connects to the controller. Also, be careful not to cross-thread actuator cables at the controller or on the actuator itself. They should be hand-tightened only. Never use a wrench.
- **Cycle the actuators when you're done.** After addressing each actuator, cycle it fully to confirm there are no hidden problems.
- **Check for cable stress.** All cables should be free of stress and secured in intervals of 18 to 24 inches.

Thorough planning and clear procedures like these will ensure that your cell site reaches and maintains its maximum potential while also allowing you to make the proper adjustments as your network evolves.

Best practices always yield the best results

Like building a house, there are countless ways for things to go wrong when building a cell site—which makes it all the more important to work with a partner who has the expertise and experience you need to do the job right. **CommScope's** long history of expertise can turn your next project into your next success.

Chapter 2 summary

Cell site development and construction:

- Research and choose a reliable services company
- Know your cables and connectors
- Maximize gain with electrical tilt antennas

Getting the signal across:
Base station antennas

Today, the quest for a stronger signal strength—or for some, any signal strength at al—has become a routine part of our daily lives. We're always searching for a way to get more "bars" on our cell phones or faster Wi-Fi connections on our computers. Whether you're at home, at work or on the go, you need good reception to communicate and good reception depends on antennas.

The antenna is one of the most critical parts of both transmitters and receivers, and often, it's the most visible. You can see big antennas mounted on tall towers, and small ones attached to Wi-Fi adapters or cell phones. Antennas come in all shapes and sizes because each one is built for a specific purpose. However, all antennas share a common link: they are the key to how well and how far communications can be shared.

CommScope invests significant research and resources in the development of reliable, high-performance antennas, technology and innovation to better support evolving networks around the world.

Antenna
The portion of an RF system that radiates radio energy into space and collects it from space.

Dummy load
A simulated power load applied to an electrical system for testing purposes.

What is an antenna?

At its most basic level, an *antenna* is the portion of a radio system that can:

1. Take radio energy from a transmission line and radiate it into space in a predictable pattern, and
2. Receive radio energy from open space and feed it back down the transmission line.

Antennas are surprisingly efficient in this line-to-space and space-to-line energy conversion process. In fact, when properly configured with the right components, antennas can yield 80 percent efficiency or greater—a remarkably high figure in engineering terms. By way of comparison, consider the common incandescent light bulb, which yields only 20 percent efficiency. This means that, of the amount of energy put into a bulb as electricity, only 20 percent of that energy is put out as light. An important consideration to maintain an antenna's extraordinary efficiency lies in the transmission cable that connects it to the transmitter.

Matching the line

To get maximum efficiency from a radio transmission's power, the antenna and cable must share certain characteristics to avoid wasted energy. For example, if a radio system uses an industry standard coaxial cable fixed at 50 ohms to connect the antenna and its transmitter, the antenna itself must rate reasonably close to 50 ohms as well.

Testing this configuration is a simple task. We connect the coaxial cable to the transmitter and place a 50-ohm *"dummy load"* on the other end to simulate an antenna. Using a watt meter will reveal two important factors that measure the efficiency of the system:

1. The amount of power entering the cable from the transmitter, and
2. The amount of power reaching the dummy load.

The difference between these two measurements represents the power lost in the line itself. The better matched the cable, the smaller the difference, and the more power reaches our simulated antenna.

If we reduce the simulated antenna's load from 50 ohms to just 25 ohms, 11 percent of the energy sent through the coaxial cable would be uselessly returned to the transmitter. That would yield very low efficiency unless we were to replace the 50-ohm coaxial cable with one rated at near 25 ohms, thereby restoring the balance. However, the 25 ohm cable would move mismatch to the source end where it connects to the transmitter.

Like water pouring through a funnel, the amount of throughput is dictated by the tightest portion of the route. In a radio system, the excess energy bounces between the transmitter and the antenna, which must reject all power above its capacity. This endlessly reflected power creates a measurable wave pattern in the cable, an effect called the voltage standing wave ratio (VSWR).

VSWR is the measurement of how well-matched a transmission line is to its antenna. Expressed as a ratio, a VSWR of 1.0:1 indicates a perfect match. Likewise, a VSWR of 1.5:1 indicates a 4 percent power reflection, which is another way of describing 96 percent efficiency, where 96 percent of the power output from the transmitter actually makes it to the antenna (table 3.1).

Voltage standing wave ratio (VSWR)
A measurement of the power reflected between transmitter and antenna in a transmission line that connects the two. This figure yields the system's transmission efficiency.

Calculating voltage standing wave ratio (VSWR)

VSWR	Return loss (dB)	Reflected power (%)	Through power (%)
1.10	26.5	0.2	99.8
1.25	19.1	1.2	98.8
1.50	14.1	4.0	96.0
1.75	11.6	7.4	92.6
2.00	10.0	11.0	89.0

VSWR = $[1 + 10^{((-\text{Return Loss})/20)}]/[1 - 10^{((-\text{Return Loss})/20)}]$

3.1: Calculating VSWR, and some sample efficiencies

Velocity, frequency and wavelength

Like all forms of radiation, including visible light, radio waves travel about 186,000 miles, or nearly a billion feet, per second. Like other forms of radiation, radio waves oscillate, or flip back and forth, between plus and minus at a predictable rate. Each complete flip is called a cycle, and cycles are expressed in hertz (figure 3.2). Measuring how many cycles, or hertz, a signal oscillates per second gives us its frequency—literally, how "frequently" the signal oscillates in one second.

Knowing a signal's speed and its frequency, we can divide the first by the second to determine its wavelength—the distance the signal travels while completing one full cycle. Wavelengths are usually measured in feet or inches, and are useful in understanding what it means to be "in phase" or "out of phase," which we'll explore later in this chapter.

Antennas are two-way streets

In theory, antennas transmit and receive in precisely the same way; the same processes occur both ways. Only the direction is reversed. In actual practice, however, a number of complicating factors, particularly on the receiving end, can impact the efficiency with which the antenna operates.

To demonstrate, it is perhaps easiest to explore the most basic of antennas: the half-wave dipole.

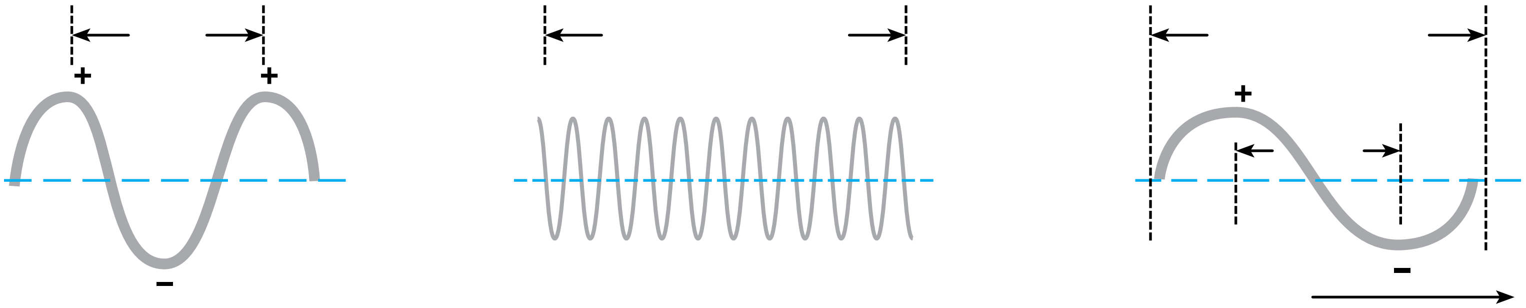

A cycle completes itself traveling from plus to minus back to plus.

The number of cycles in one second gives the frequency.

The speed divided by the frequency gives the distance the wave travels in one cycle. This is called the wavelength.

3.2: Relationship of time, frequency, and wavelength

Half-wave dipole

The half-wave dipole radiator antenna, often just called a "dipole," is the most basic one used in two-way base station applications. It is essentially nothing more than a straight conductor made of wire, rod or tubing that measures exactly half of its assigned frequency's wavelength. A rule of thumb for determining the correct length is:

Length (in inches) = 492 divided by the desired frequency in MHz

As a result, dipole antenna length can be highly variable. It could be just 2.4 inches in length for a frequency of 2500 MHz, or 20 feet long for a frequency of 25 MHz. The table below provides more examples (table 3.3).

Generally, the feeder line is connected at the midpoint, so the antenna radiates at maximum intensity in the middle of the dipole. (figure 3.4).

Frequency (MHz)	1/2 Wavelength (inches)	Frequency (MHz)	1/2 Wavelength (inches)
30	196.8	800	7.4
50	118.1	900	6.6
74	79.8	1700	3.5
150	39.4	1900	3.1
220	26.8	2100	2.8
450	13.1	2500	2.4
750	7.9	3500	1.7

3.3: Half wavelengths of two-way frequencies

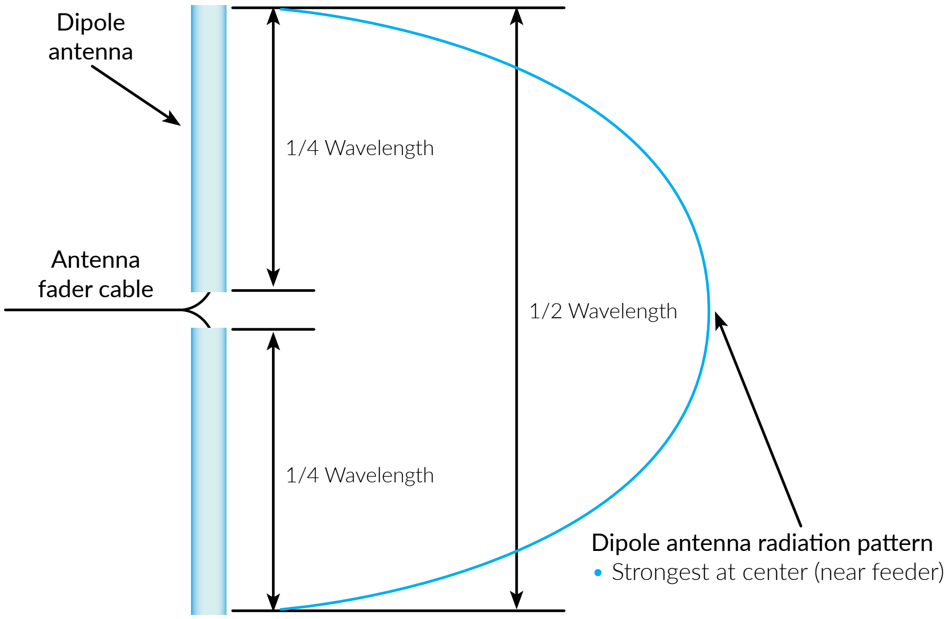
Dipole antenna
1/4 Wavelength
Antenna fader cable
1/2 Wavelength
1/4 Wavelength
Dipole antenna radiation pattern
• Strongest at center (near feeder)

3.4: Dipole antenna construction and radiation pattern

Vertical and horizontal antenna radiation patterns

All antennas, regardless of polarization, have three-dimensional radiation patterns. If the pattern is extended in all directions equally, the resulting shape would be a sphere, with the antenna at its center. The polarization of the antenna determines which portion of that sphere represents an antenna's actual pattern. Slicing the sphere vertically yields a vertical circle while a horizontal slice would reveal a horizontal circle.

These theoretical descriptions of the two polarization patterns appear to be omnidirectional within their planes, but that's not quite the case. In practice, there are no truly omnidirectional antennas. Our example, half-wave dipole antenna, for instance, reveals the truth (figure 3.5). The pattern appears circular, like a doughnut, on a horizontal plane, but forms a figure-8 in the vertical plane.

As we will see later in this section, most real-world antennas consist of a vertical array of radiating elements and elevation pattern shaping has become quite important for interference minimization (figure 3.18).

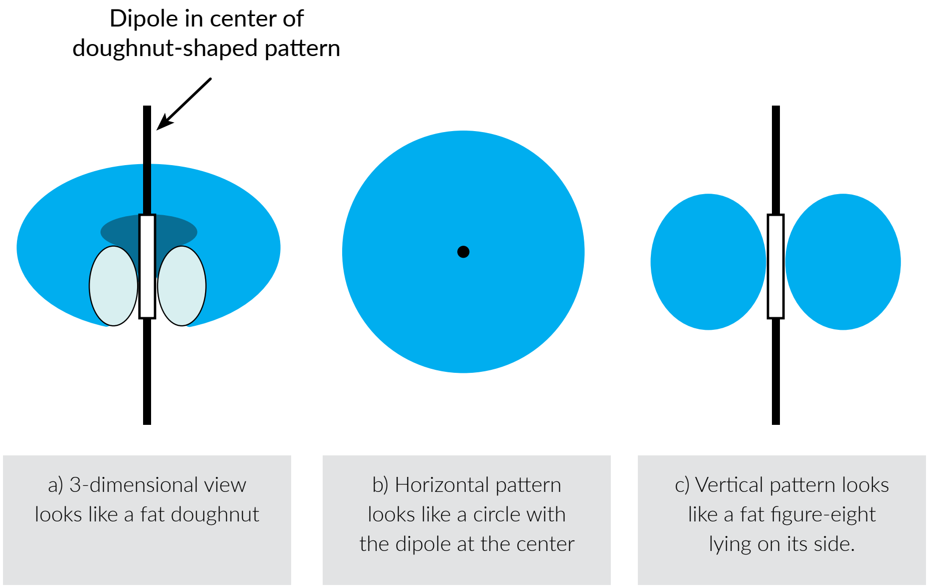

Dipole in center of
doughnut-shaped pattern

a) 3-dimensional view looks like a fat doughnut

b) Horizontal pattern looks like a circle with the dipole at the center

c) Vertical pattern looks like a fat figure-eight lying on its side.

3.5: Radiation patterns for a dipole antenna

Antenna gain

Using our half-wave dipole as our reference, we measure gain in decibels (dB). Decibels are used to compare one quantity of power to another. So, taking our dipole as our starting point, we say it has a gain of one, or that it has a value of 0 dB, as it has “zero difference” from itself. If we have an antenna with double the power of our reference dipole, we find that its power ratio of 2.00 yields 3 dB in gain (table 3.6).

Power ratio	dB	Power ratio	dB
0.10	–10	1.00	0
0.13	– 9	1.26	1
0.16	– 8	1.58	2
0.20	– 7	2.00	3
0.25	– 6	2.50	4
0.32	– 5	3.16	5
0.40	– 4	4.00	6
0.50	– 3	5.00	7
0.63	– 2	6.30	8
0.79	– 1	8.00	9
1.00	– 0	10.00	10

3.6: Deriving gain in dB from power ratios

Boosting gain

Theoretically, there are two ways to increase antenna gain.

1. We can increase the power or current density in the antenna so it will radiate its pattern with greater intensity. But as we discovered earlier, we cannot raise the power without fundamentally altering the mechanics of the antenna, cable and transmitter.
2. Alternately, we can change the radiation pattern, tightening its focus, so more of the existing power is directed where it will be used. This can be done without changing out hardware.

Consider again the circular, doughnut-shape pattern of our dipole antenna (figure 3.7). By “squashing” the doughnut vertically, we produce a denser, flatter, rounder pattern. As a consequence, the circle also grows larger as vertical space is traded for horizontal space. Since verticality is rarely of importance, this is a very profitable exchange.

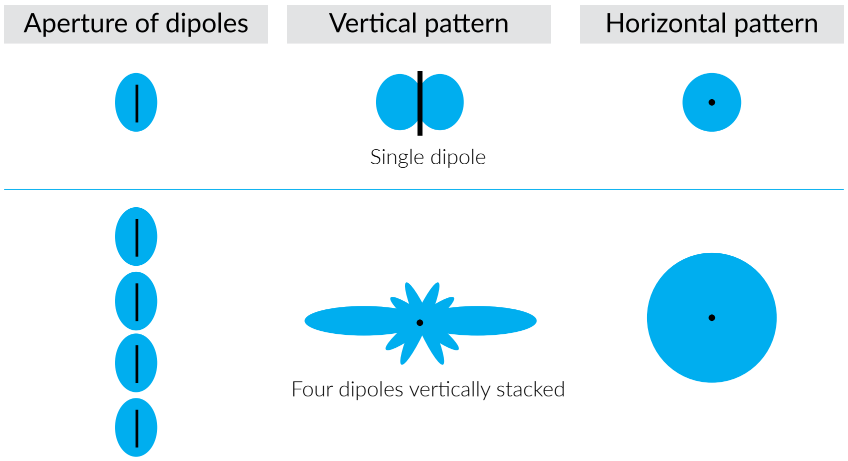

3.7: This figure illustrates how stacking four dipoles vertically in line changes the pattern shape (squashes the doughnut and increases the gain over a single dipole). The area of the horizontal pattern measures the gain. The small lobes in the lower center section are secondary minor lobes.

In phase
Multiple antennas radiating together at precisely the same time and rate.

Omnidirectional pattern gain antennas

To achieve greater gain in this circular (or omnidirectional) pattern, we can stack multiple vertical dipole antennas above each other, as shown in figure 3.7. This increases the vertical size of the antenna. Then, we feed power to the dipoles in such a way that they add together at a distant point—again, with transmission lines matching their radiation power limits for greatest efficiency. By feeding equal amounts of power that arrive at each dipole at the same instant, the dipoles radiate ***"in phase,"*** or in synchronicity, for improved gain by virtue of its pattern.

This type of antenna is called a vertical collinear phased array, and it is the most commonly used type of base station antenna.

Aperture

Aperture, or beam width, determines the gain of an antenna. Like an adjustable nozzle on a garden hose, aperture describes the degree to which the signal is focused: the tighter the focus, the greater the gain within that area of focus.

Spacing of dipole elements

In a vertical collinear array, each dipole or sub-array of dipoles is connected in parallel to the common feed point by a separate transmission line. This means it's possible to locate the dipoles so that their vertical separation tightens overall aperture to boost gain. This separation is usually something less than a single wavelength of the assigned frequency being transmitted. Anything less tends to reduce the improvements in gain.

Feeding the array

In a vertical collinear array of two or more dipoles, the most common means of feeding power via the coaxial transmission lines are:

> **Parallel (shunt) feed.** Power is fed along individual lines to each dipole or sub-array of dipoles. Using matching transformers and junctions, the cables connect to the line running down the tower. This allows the array to be fed from the center, equalizing the effectiveness of each array element and preventing the beam tilt that affects series-fed installations.

In either application, the physical length of the aperture directly determines the amount of gain.

Directional gain antennas

While omnidirectional gain antennas like the vertical collinear array achieve greater gain by compressing its vertical pattern into a flatter circular shape, there are other types of antenna that modify their horizontal patterns to accomplish the same gain improvements.

Dipole and reflector

As we've shown, a vertical dipole antenna has a circular horizontal pattern. However, if we position it in front of a metal screen or wire mesh, we see that radiation going to the rear will be blocked (figure 3.8). If this blocked radiation is reflected off this screen, the horizontal pattern will no longer be circular, but directional.

If a dipole is positioned exactly one-quarter wavelength from this reflection screen, the radiation that would ordinarily go to the rear is redirected to the front to form what is called a directional lobe. It's the same effect as that of the reflective mirror behind a flashlight's light bulb, which redirects the circular light pattern into a single direction. The larger the screen, the greater the reflection and the narrower the directional lobe becomes—and just as the omnidirectional antenna increases gain by compressing vertically, this directional antenna increases gain by compressing horizontally and directing all its power in a single direction.

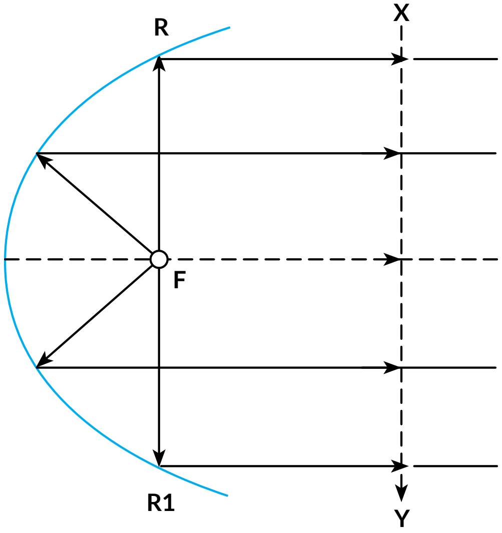

3.8: The omnidirectional pattern of a dipole can be made directional

Corner reflector and parabolic reflector

Corner reflectors are right-angled V shapes, with their vertices spaced a certain distance from the dipole (figure 3.9). The resulting reduced beam width and the corresponding increased gain is determined by the size, shape and distance of the screen from the dipole.

Modifying this corner reflector into a rounded bowl shape creates a parabolic reflector, with the dipole placed at its focal point. In either application, the reflector may be built of closely spaced rods instead of wire mesh, increasing the strength and reducing the wind drag of the reflector.

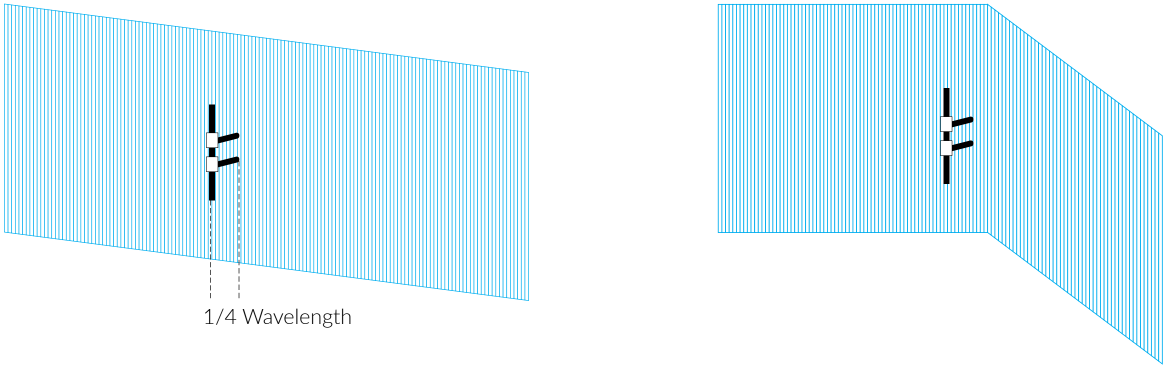

Flat screen
Radiation to the rear is blocked

Corner screen
The screen behind the dipole cuts off radiation to the back and reflects it forward to form a beam

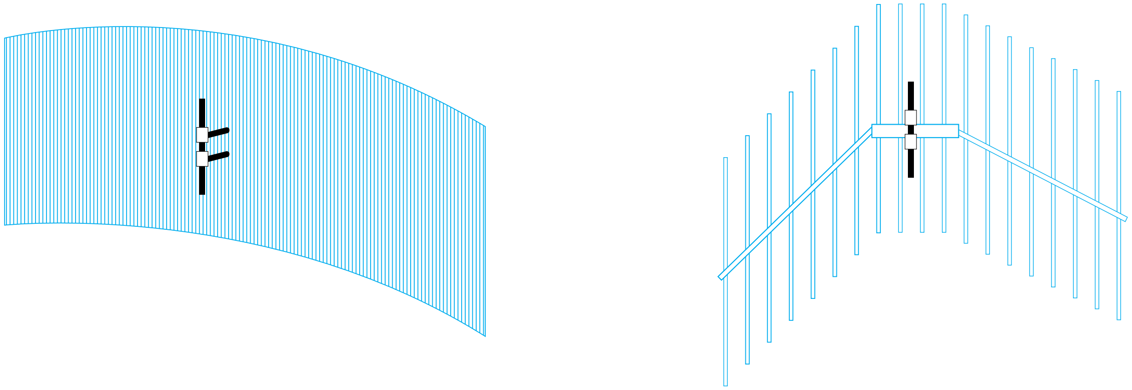

Parabolic screen
The parabolic screen focuses on the beam

Corner reflector
Vertical rods replace the scree of the corner screen to make a reflector of stronger mechanical design

3.9: Dipole antenna reflector shapes

The Yagi antenna

Perhaps the most commonly used directional antenna is the Yagi antenna. The Yagi offers many variations and forms, but it generally consists of three elements (figure 3.10):

1. Radiator
2. Reflector
3. Director

These components are arranged such that the director is in the front, the radiator behind that, and the reflector behind both. Generally, the director is the shortest component, and the reflector the longest. The spatial relationships between the three elements determine the power that goes into the Yagi's directional lobe, and therefore also determine the Yagi's gain. It is an efficient and flexible design, offering high gain, low weight, minimal wind drag and modest cost, resulting in its popularity for two-way radio communications.

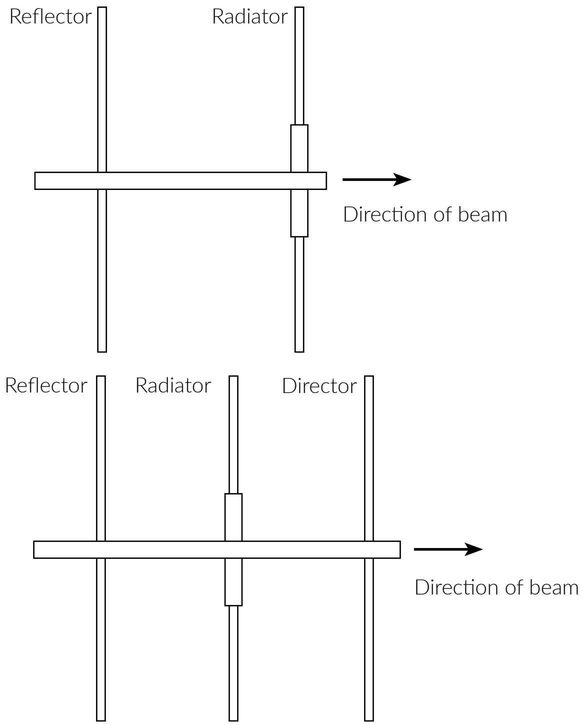

3.10: Two-element and three-element Yagi directional antenna

Yagi antenna
Also known as an Yagi-Uda antenna, this is a common type of directional antenna, first created in Japan in 1926 by Hidetsugu Yagi and Shintaro Uda.

The bandwidth factor

Another consideration in antenna construction is that of bandwidth. As we discussed in our description of the dipole, antennas must be built to lengths that are determined by their operating frequency wavelengths. As you may recall, the dipole antenna must be half the length of its assigned wavelength. However, it is possible to build an antenna that covers a range of frequency bands or bandwidth centered around a particular frequency.

Indeed, nearly every antenna in use today affords a fairly wide percentage of bandwidth. In fact, certain antenna designs in the 1900 MHz frequency range can offer over 45 percent bandwidth (1710-2690 MHz). These are known as ultra-wideband antennas.

The bottom line of antenna design

We've just covered the basics of antenna design, but it's important to keep in mind that there are no magic formulas or one "ideal" antenna configuration. Improving one aspect of operation always comes at the expense of another aspect. The best design is always one that is driven by the specific operational objectives of the antenna.

Cellular antenna concepts

Now that you have a basic working knowledge of antennas and how specific configurations can help them perform even better—that is, by improving gain. Let's look at a much more specialized and sophisticated category: the cellular antenna.

Cellular antennas are a familiar feature in nearly every corner of the world. In many cases, these are cellular networks that bring new connectivity where it had never before been possible, and these connections depend on cellular base station antennas.

In cellular base stations, there are two basic antenna types currently in use (figure 3.11):

1. **Omnidirectional antennas,** which we defined previously as antennas that exhibit a circular radiation pattern and operate in virtually all directions, and
2. **Directional (or sector) antennas,** which operate in a specific direction, most commonly covering an arc of 120 degrees or less, depending on capacity requirements.

Figure 3.11 shows older legacy sites using vertically polarized antennas. Rural sites typically used 90° horizontal beamwidth models, suburban sites used 65° models and urban sites used models ranging from 33° to 65°. In these cases, two Rx antennas were required per sector to support Rx diversity. For modern sites, a single Dual-pol (±45° polarization) model with the appropriate horizontal beamwidth supports Rx diversity.

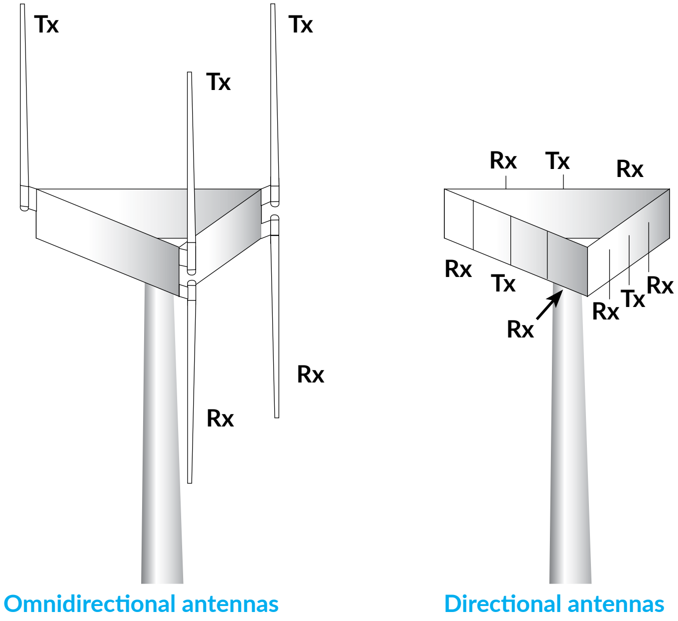

3.11: The two types of commonly-used cellular base station antennas

Cell reuse

What makes cellular networks different from other types of communications is the principle of cell reuse. Cell reuse is a way of increasing network capacity by “reusing,” or reassigning, individual frequencies on the fly within a particular cell.

To see this process in action, consider the shape of cells and how they fit together. Typically, cells are represented as interlocking hexagons, as seen below (figure 3.12). Depending on the density of the area served, these hexagons can be miles across or cover just a few hundred feet.

As a result of this incredible flexibility, channel sensitivity is limited by external interference rather than noise issues, as older radio communications have traditionally been. The specialized pattern shaping available with directional antennas, both in azimuth (horizontal direction) and elevation (vertical space), allows incredibly precise coverage that doesn’t interfere with neighboring cells.

Antenna characteristics

A cellular base station’s antenna is the most critical consideration in an efficient cellular network, and it all depends on choosing the antenna with exactly the right physical characteristics for a specific application. These characteristics relate to radiation pattern, antenna gain, front-to-back ratio and a number of other critical factors.

In the real world, defining, choosing and testing these characteristics requires a great deal of technical expertise and mathematical skill, so for the purposes of this discussion, we will cover the basics with a far more generalized approach than an engineer would use in an actual evaluation.

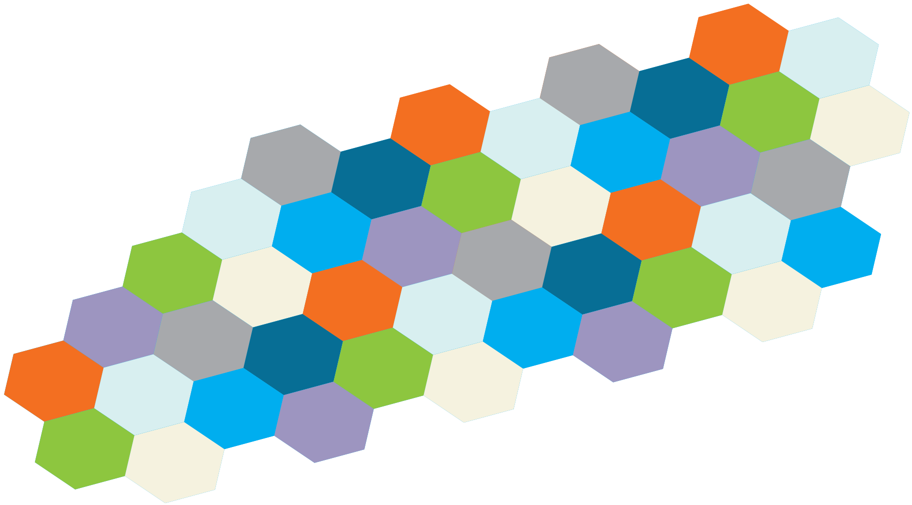

3.12: Cell reuse in a sample map. The entire map can be covered with just 7 unique cells and still provide adequate channel isolation between cells.

Radiation pattern
The three-dimensional shape of an antenna's strongest signal transmission.

Spherical coordinate system
A geometric polar coordinate system used to mathematically map the radiation pattern of antennas.

Azimuth coordinate system
The polar coordinate system used in the field by RF engineers and surveyors to map the radiation pattern of antennas.

Radiation pattern

Perhaps the most obvious and important characteristic to understand is an antenna's radiation pattern. If a particular application calls for coverage in all directions, you would choose an antenna with a circular or omnidirectional radiation pattern. If your installation requires a more focused signal, a directional antenna's radiation pattern would satisfy your needs.

Mapping an antenna's radiation pattern is a fairly simple task. Connecting a probe antenna to a receiver and moving it around your tested antenna at a fixed distance allows you see the variations in signal strength. Mapping these readings with polar coordinates yields a three-dimensional map showing in which directions the antenna transmits most strongly (figure 3.13).

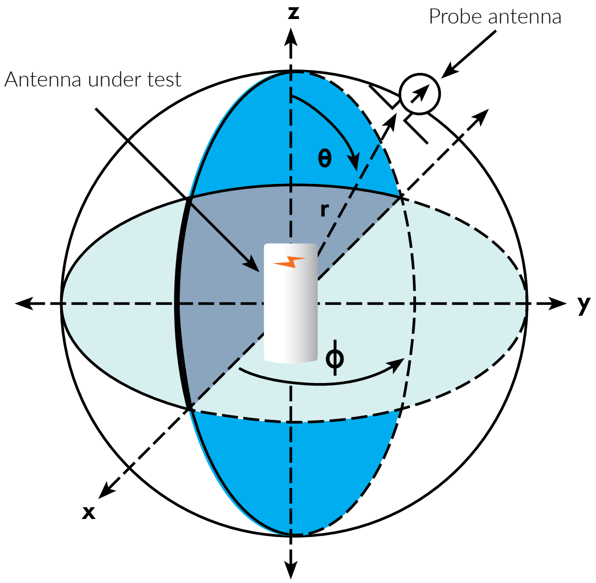

3.13: Moving a probe antenna around the tested antenna at a fixed distance yields a three-dimensional map of its radiation pattern

At the same time, a pattern can also be expressed as a conventional rectangular plot with angular position on the X-axis and signal strength on the Y-axis. Examples of both are shown below. Depending on the design of the antenna, the radiation pattern can display any number of shapes. The isotropic dBi reference is a theoretical "point source" and thus generates a pattern covering all directions of a sphere. As seen previously, the ½ wave dipole dBd reference pattern has nulls above and below the dipole and thus from a conservation of energy standpoint must have more gain on the horizon than the dBi reference. The absolute difference of these two standards is 2.14 dB and today most manufacturers rate their products in both dBi and dBd. Since an antenna's gain is determined by comparing it to one of these standards, the dBi rating will always be 2.14 dB greater than the dBd rating.

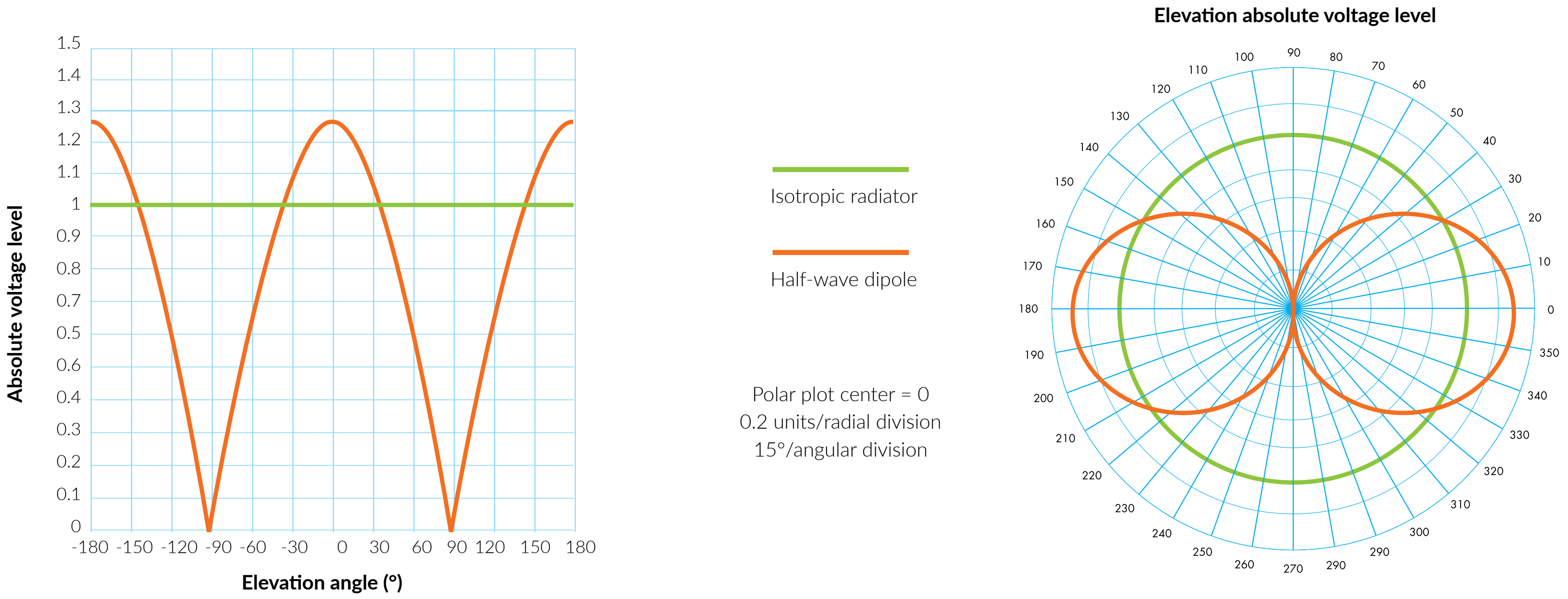

3.14: A system controller interface displaying voltage, amperage and alerts

Antenna gain

As discussed earlier, an antenna’s radiation pattern is directly connected to its gain, or performance power. While we cannot improve gain by increasing transmitter power without rebuilding the system entirely, we can use radiation patterns to achieve the same objective.

Gain is measured in decibels (dB) and rises as a function of increasing aperture size, which in most cases means increasing the physical size of the antenna. As a general rule, a doubling of aperture results in a doubling of gain. Though, as a practical matter, larger antennas introduce efficiency-reducing power losses that can diminish these improvements.

Front-to-back ratio

The ratio of a directional antenna’s maximum directivity to “front” (where its main lobe appears) to its “back” where its reflector is located is called, appropriately, the antenna’s front-to-back ratio (figure 3.17).

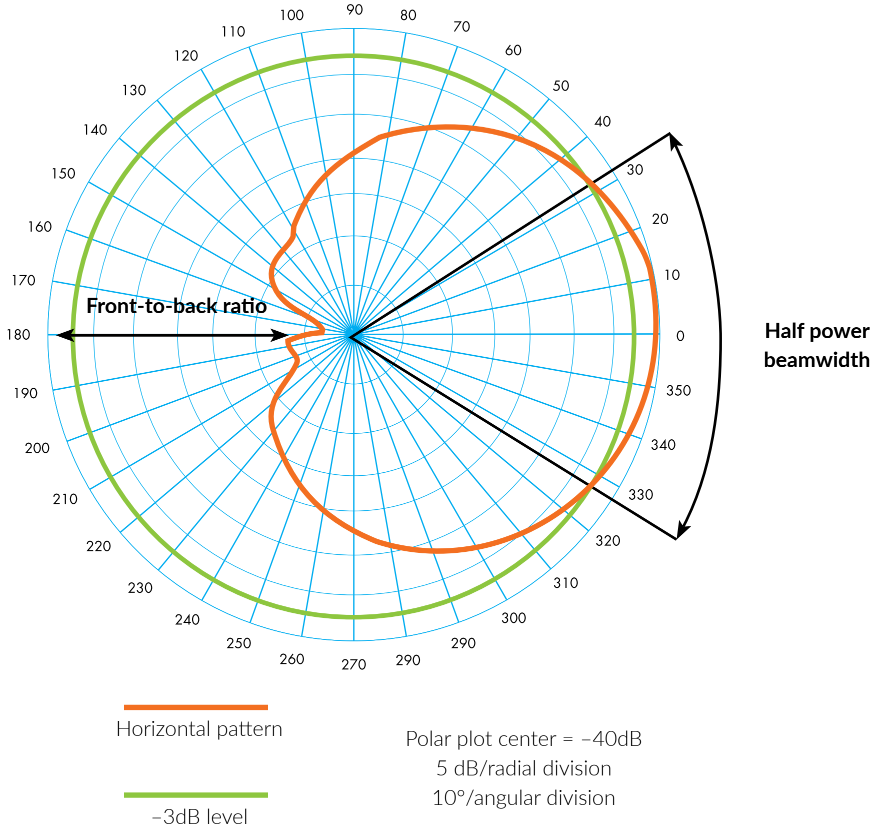

3.17: A polar representation of a directional antenna’s front-to-back ratio

Side lobes and nulls

Apart from a radiation pattern's main lobe, there can also exist side lobes and nulls. Side lobes are extraneous areas of strong signal, and nulls are the low-energy spaces between them (figure 3.18). Nulls may exhibit 30 dB or more of attenuation, meaning signals found there can be as weak as one one-thousandth of the power of the main lobe.

There are ways of redirecting side lobe power back into the main lobe. This process, called null fill, can result in the widening of the main lobe and reduce gain accordingly.

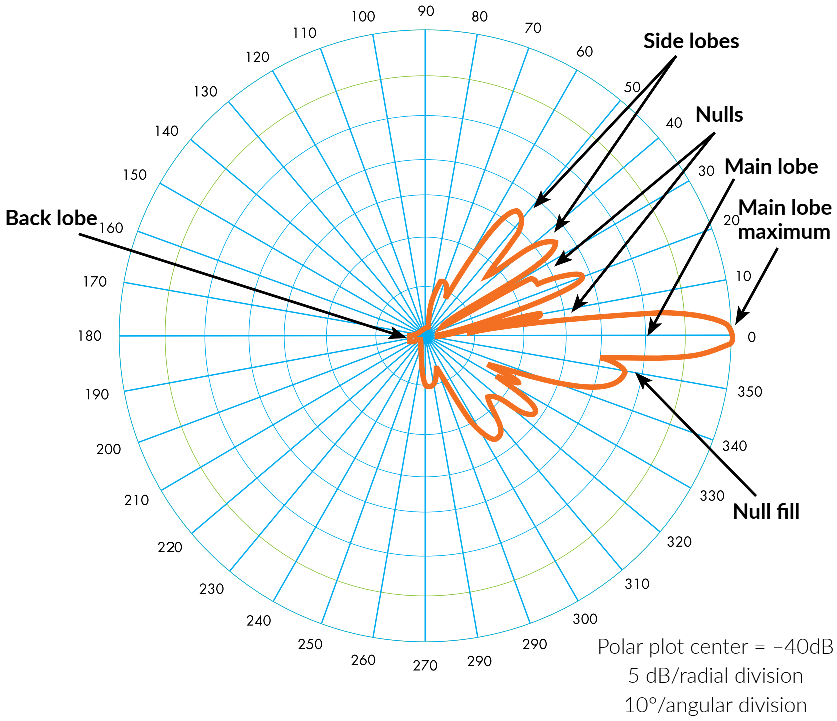

Polar plot center = –40dB
5 dB/radial division
10°/angular division

3.18: *A polar representation of a vertical pattern including side lobes and nulls*

Polarization

Polarization is a property of the wave produced by an antenna that describes the way that wave varies in space over time. In simpler terms, it describes the orientation of that wave, such as vertical or horizontal or slant 45 degrees (Dual-pol).

Cross-polarization ratio

This characteristic measures the performance of a dual-polarized array in distinguishing between orthogonal waves (two signals broadcast perpendicular to one another, such as horizontally and vertically). This figure is calculated as the ratio of co-polarization to cross-polarization occurring in the antenna's main lobe (figure 3.19).

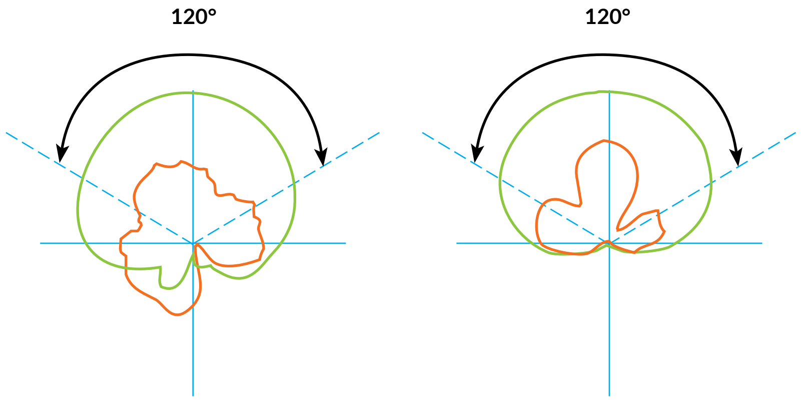

3.19: Typical and directed dipole cross-polarization ratios

Sector power ratio

Basically, sector power ratio is a comparison of signal power registered outside and inside a desired receiving area as a consequence of an antenna's radiation pattern (figure 3.20). The higher the ratio, the better the antenna's interference performance.

As a practical matter, particularly in cellular network applications, lower sector power ratios indicate a higher amount of interference between antennas in adjacent coverage areas. When competing signals overlap, interference can increase and reduce performance, such as dropping a cell phone call while moving from one cell to another. Cellular networks require precise sectorized planning to prevent this kind of problem.

Beamtilt

As capacity requirements increase, one solution is to split the hexagons shown in figure 3.12 allowing the addition of more sites and reducing the coverage radius of the original site. To accomplish this, elevation beam downtilt is commonly used to reduce the gain on the horizon (and thus the coverage radius) as shown in figure 3.21. Mechanical downtilt results in undesirable pattern distortion on the horizon while electrical downtilt maintains the desired pattern shape.

Early antennas incorporated fixed electrical downtilt, but this required multiple different models.

Cellular antennas on a practical level

When we move beyond the drawing board of theoretical antenna design to the real world, we soon discover that the laws of physics are not the only limiting factors affecting an actual installation. These issues include everything from tower weight and wind limits to local zoning board approvals for antenna size, shape, height and appearance. In most installations, compromises are necessary to satisfy all the competing interests.

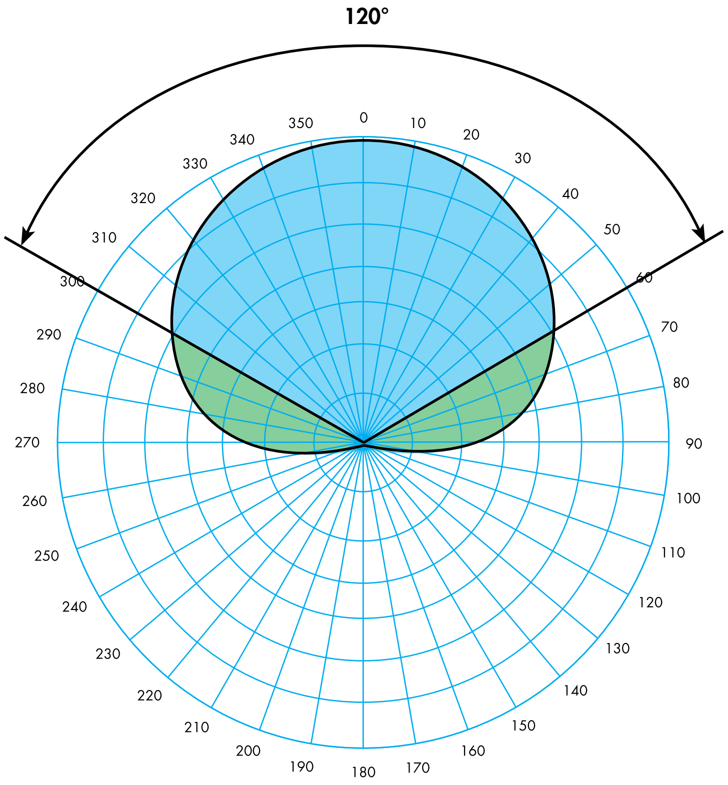

3.20: Graphic and mathematical representations of an antenna's sector power ratio

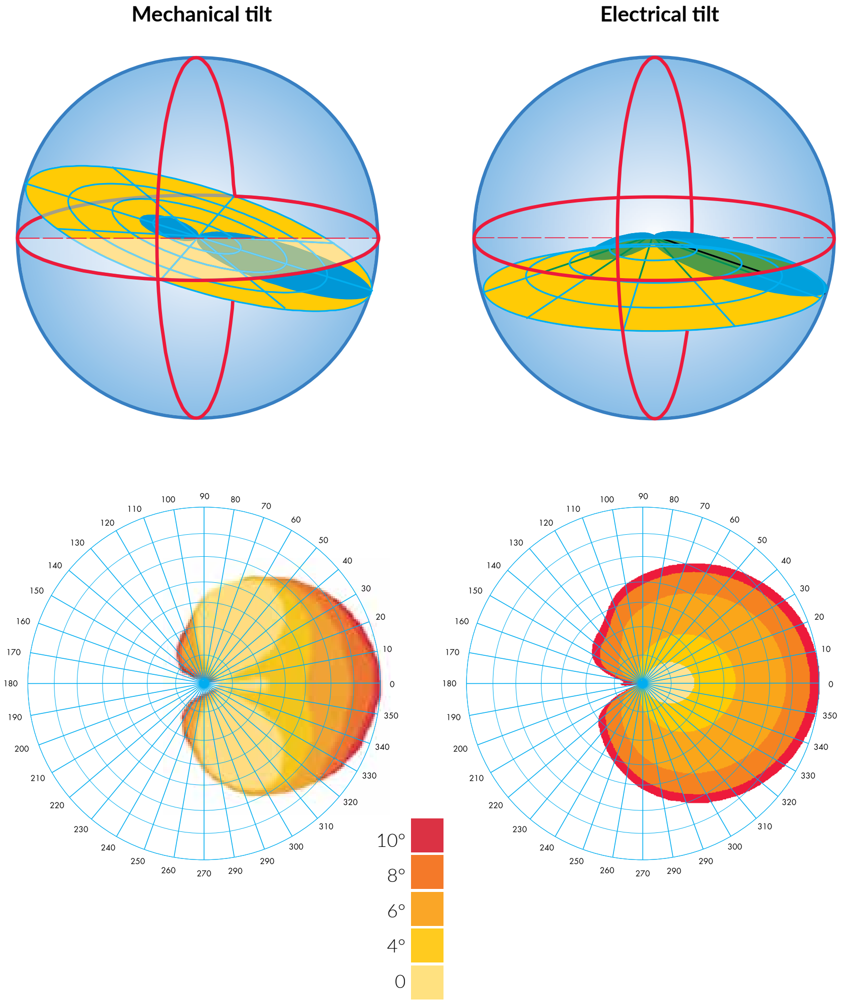

Mechanical tilt

Electrical tilt

3.21: Tilting the antenna changes the shape of the lobe at ground level, reducing gain

Most cellular antennas are produced in a variety of physical sizes to offer the best performance while conforming to other requirements. Chances are that you've seen cellular antennas mounted in a number of ways, featuring diverse sizes and designs, such as the commonly used lengths of 4, 6 and 8 feet.

Materials and environment

Cellular base station antennas are only as reliable as the materials that go into their construction, and the construction of their arrays. When it comes to working with the physical limitations of an antenna's location, matching the right materials to the environment is a critical consideration. Here are just a few examples.

In the antenna array itself:

- Aluminum alloys offer lightweight strength, but can be vulnerable to the elements
- Pressure cast aluminum is well suited to bases, sockets, mounts and clamps, where its hardness and resistance to corrosion are critical
- In circumstances where weight is not a serious factor, copper and brass are used for their easy plating and conductivity properties

Antenna radomes:

- High-strength, low-RF loss materials such as fiberglass offer protection from the elements
- Materials must offer UV protection to prevent deterioration due to sunlight exposure

Tower appearance:

- For purposes of appearance and zone compliance, non-metallic paint can be applied to the entire structure
- For better wear, smooth surfaces should be roughed prior to painting

These are just a few of the more obvious physical considerations. Other matters in cable selection, connector choice and termination options demand close attention as well.

More capacity with fewer antennas

As mentioned earlier, cellular antennas are directional, often covering 120 degrees, or one-third of a complete circle. Mounted together on a triangular tower, three sets of these antennas can cover all directions. But in densely urban areas that require more capacity, narrower focus-antennas (called a six-sector scheme) can handle additional traffic along with the cost of adding more antennas. Having so many antennas in a single location makes it more likely to run afoul of local zoning codes.

TwinBeam

One recent solution to this problem involves the TwinBeam antenna from **CommScope**, which produces two separate 35-degree beams with centers separated by 60 degrees. As the illustration below shows, this dual-lobe approach provides excellent coverage and, unlike the six-sector solution, only requires three antennas instead of six (figure 3.22).

'Six-Sector' sites

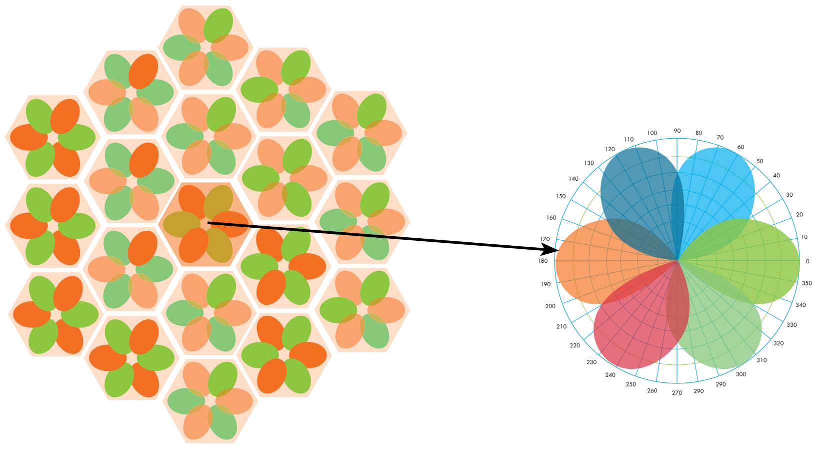

65° Antenna vs. Twin Beam antenna

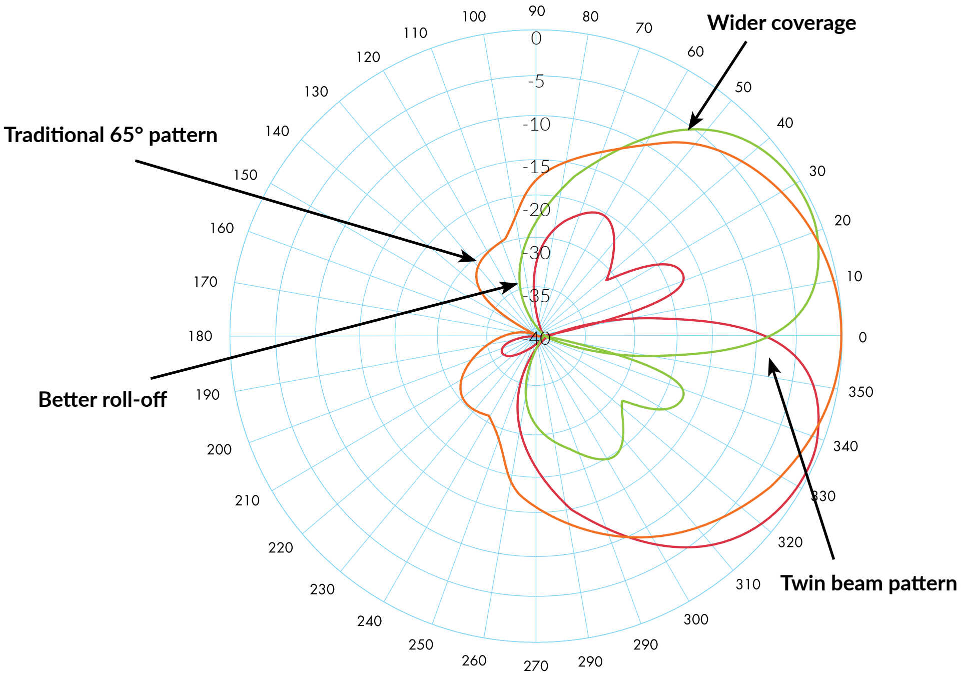

Twin beam advantages

- Back-to-back broadcast channel (BCCH) re-use capability
- Enhanced RF footprint (higher gain)
- Better interference containment (better roll-off)
- Easier zoning and faster site deployment

3.22: Same number of antennas, double the coverage—a six-sector pattern compared to TwinBeam installation

SmartBeam®

Another type of antenna addresses growing capacity needs by intelligently steering themselves for maximum efficiency.

The **CommScope** solution is called SmartBeam. In addition to electronic downtilt, these multiple-degree-of-freedom antennas incorporate azimuth beam steering plus or minus 30 degrees and azimuth beam adjustment from 35 to 105 degrees (figure 3.23).

SmartBeam technology video

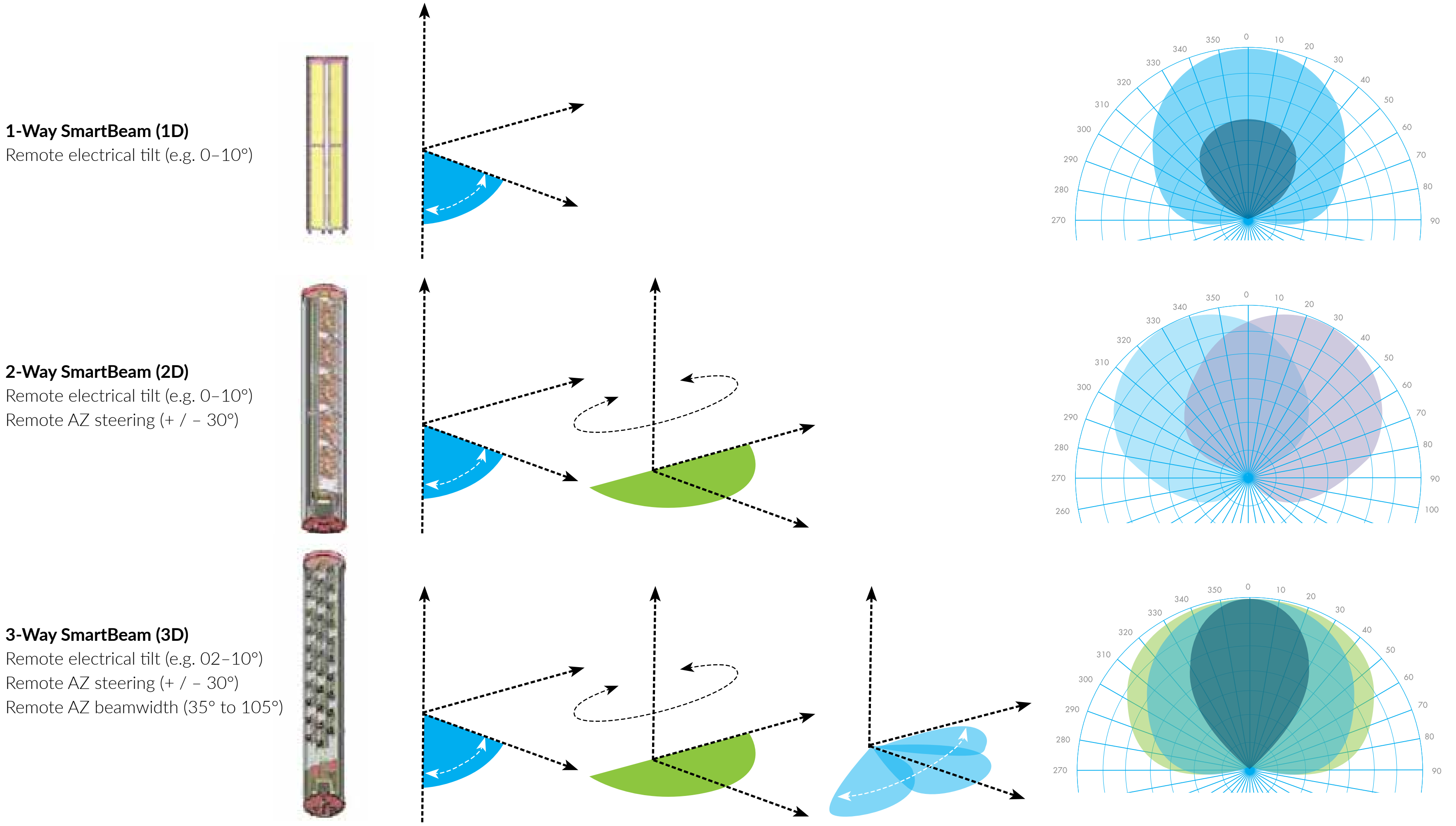
1-Way SmartBeam (1D)
Remote electrical tilt (e.g. 0–10°)
2-Way SmartBeam (2D)
Remote electrical tilt (e.g. 0–10°)
Remote AZ steering (+ / – 30°)
3-Way SmartBeam (3D)
Remote electrical tilt (e.g. 02–10°)
Remote AZ steering (+ / – 30°)
Remote AZ beamwidth (35° to 105°)
0
10
20
30
40
50
60
70
80
90
100
350
340
330
320
310
300
290
280
270
260

3.23

Shannon's Law

Created by Claude Shannon and Ralph Hartley, this law establishes a theoretical limit to how much data can be reliably pushed through a given amount of bandwidth.

Adaptive array

A third capacity-boosting option incorporates several vertical elements that steer a beam toward each user on a tightly managed time-division basis. In this application, each user owns a particular time slot to move his or her traffic. Of course, managing this system for a large number of users requires powerful and sophisticated digital processing, but it also holds the potential to effectively "null out" nearby interference for better high-speed throughput (figure 3.24).

The next evolution

New technologies are being developed and deployed at a dizzying rate. The current field of cutting-edge networks is collectively known as long-term evolution networks (LTE). LTE has the potential to completely reshape how networks can perform, because it incorporates a concept called multiple input, multiple output (MIMO), which splits data transmission into multiple streams and sends them at the same time on the same frequency using multiple de-correlated RF ports. The expression 2x4 MIMO means that there are 2 de-correlated paths in the downlink and 4 de-correlated paths in the uplink.

What makes this development so exciting is that MIMO offers a way around a classic limiting factor of RF communications known as Shannon's Law, which dictates how much throughput can be delivered down a given amount of bandwidth. As figure 3.24 shows you can only expect to get to within 3 dB of a bandwidth's theoretical maximum in a practical application.

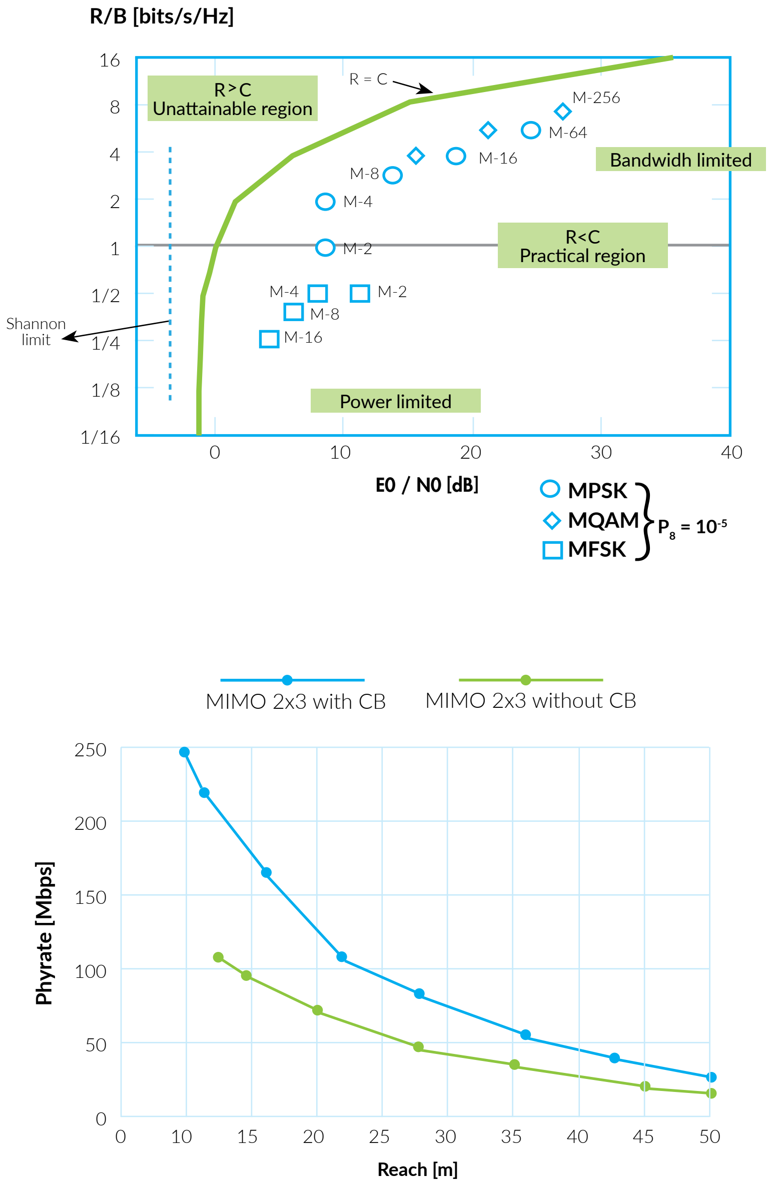

3.24: A traditional plot of Shannon's limiting law

MIMO circumvents this limit through digital signal processing (DSP), which can distinguish between the two split signal paths and reassemble them into the original data on the receiving end. This workaround literally doubles the theoretical limits defined by Shannon's Law when applied in a 2x2 MIMO configuration with two transmit and two receive antennas (figure 3.25). It is quadrupled in a 4x4 MIMO configuration with 4 transmit and 4 receive antennas. Actual throughput improvements do not quite achieve this degree of volume, but that differential is to be expected in any practical application of theoretical performance.

MIMO systems
2 x 2 SU-MIMO: Spatial multiplexing

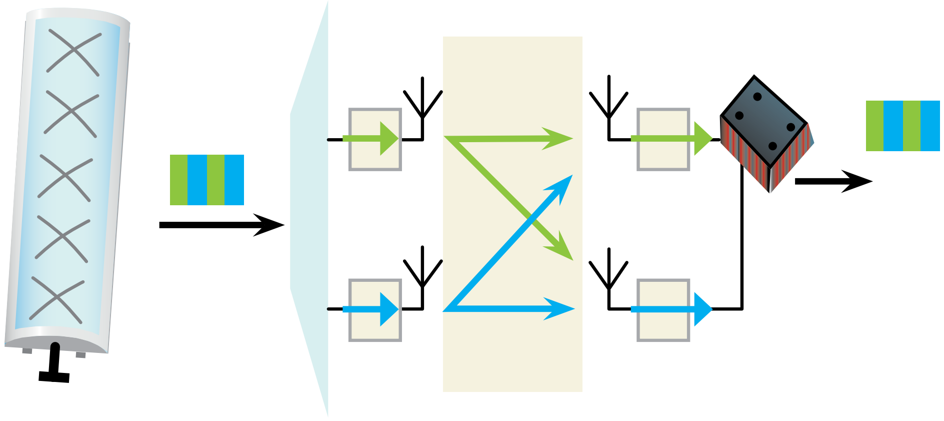

Same time and frequency resource

- Multiple input multiple output
- Capacity gains due to multiple antennas at both ends of the link
- Multipath provides additional channel using DSP
- LTE supports 1x2, 2x2, 4x2, 4x4
- Spatial multiplexing requires a multi-path environment

Different data streams

- Space Time Block Coding is a transmit diversity mode used S/N cannot support Spatial multiplexing
- Decorrelation between antennas and propagation paths required for Spatial multiplexing
- A dual polarized BSZ for 2x2 MIMO; two separated for 4x2 or 4x4 MIMO
- Alternatively vertically polarized antennas can be used with spatial separation

3.25: This 2x2 MIMO system uses digital signal processing to circumvent theoretical throughput limits

The basics of base station antennas

Incredibly diverse and remarkably efficient, antennas are the most critical link in any communications network. Radiating radio energy into space and collecting it from space, they can connect a single network backbone or thousands of individual users.

By virtue of their design, antennas can cover virtually any desired area of any shape. But it takes a lot of insight, knowledge and planning to get the most out of every watt. It all comes down to understanding your application's needs and its limitations.

With your new understanding of how antennas work, and how their performance is measured and compared, you may think twice the next time you are searching for more “bars” on your cell phone or a faster Wi-Fi connection for your computer. Then look up at the next TV aerial or cell phone tower you see and remember the complexity of the invisible processes that make modern communications possible at home, at work and on the go around the world everyday.

Chapter 3 summary

Antennas:

- Structures that radiate and receive radio energy
- Can achieve 80 percent efficiency or greater
- Directional (sectorized) or omnidirectional

Performance characteristics:

- Radiation pattern
- Polarization
- Gain
- Aperture

Enhancements through design:

- Vertical stacking
- Element spacing
- Horizontal pattern shaping
- Downtilt

Cellular base station antennas:

- Sectorized, grouped antennas commonly covering 120 degrees or less
- MIMO and LTE technology represent next step in speed and reliability

Working within the limits: Co-siting solutions

If you've ever tried to get more use out of the space in your home by combining a home office with a guest bedroom, you may not have been entirely satisfied with the results. Sure, you've saved space by making one room do the job of two, but you probably found that it can't do either job quite as well as a dedicated space would have allowed.

This tradeoff of space for utility is also the guiding principle behind co-siting a cellular installation. With space at a premium, there are real incentives to reducing your equipment footprint but every square foot saved places new constraints on the way that base station operates. Since every site has unique limitations, it can be a challenge to identify and implement the best *co-siting solutions*.

Whatever the specifics of a given cellular installation may be, **CommScope** offers a wide range of solutions that meet virtually any installation requirement. It takes a combination of technology and insight to make the best of every situation.

Co-siting solutions
The technology and techniques that allow cellular base stations and air interfaces to share architecture and operate within limiting factors of their locations.

Dealing with the realities

Just as it would be ideal to have an unlimited number of rooms in your home for every possible purpose, it would be ideal for cellular base stations to be equipped with their own dedicated antennas and feeders at every cell site (figure 4.1).

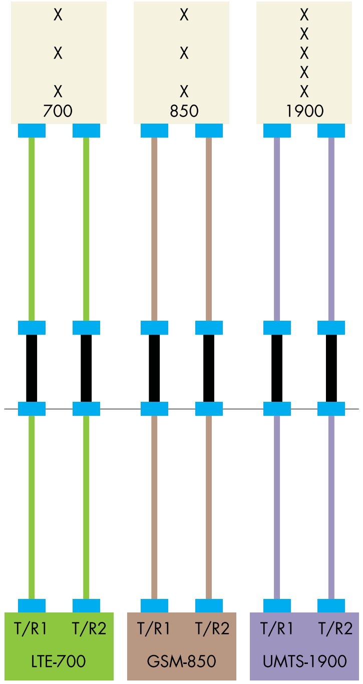

4.1: Multiband sector with separate feeders

If such an arrangement were possible in every installation, the benefits could include:

- Individually optimized antenna pattern, azimuth direction and downtilt angle
- Minimal RF path loss and signal mismatch
- Reduced interference and intermodulation between systems
- The ability to perform maintenance on one system without impacting the others

Sadly, this arrangement isn't a practical option for most real-world designs. When a cellular base station makes the move from the drawing board to the tower installation, its design becomes subject to an incredible number of variables and limiting factors. Some of the more common limits are:

- Local zoning ordinances that restrict quantity, size and location of antennas
- The tower's structural weight limits and wind load restrictions
- Budget constraints that limit both the initial and ongoing costs
- Scheduling demands that require accelerated service rollouts

Making the most of every watt

To address these limits and wring greater performance from every watt of power, co-siting solutions can help different technologies operate on a single architecture. Even networks operated by competing companies can realize mutual benefits by sharing site equipment, much as competing airlines will honor each other's tickets in the event of flight cancellations. In fact, such sharing agreements are now the norm with consumers benefiting from the providers' reduced operational costs in the form of better transmission speeds and reduced cell phone and data bills.

Co-siting solutions are usually based on specific equipment and configurations designed to improve performance within a defined set of circumstances under a defined set of limitations.

Multi-band combining

One frequently used technique is called *multi-band combining*, a method of frequency multiplexing. It takes advantage of the fact that feeder cables are naturally well suited to being shared by multiple frequency bands. In other words, multiple base station services can be funneled into a single feeder cable that runs up the tower to the antennas. Those services can then be split away from that one cable directly beneath the antennas.

To visualize this concept, think of how you bundle your home or office computer's wires into a single plastic cable wrap; at one end, the cables separate into various ports on the back of your computer. On the other end, the cables separate into your keyboard, mouse, network and printer connections.

To achieve the benefits of frequency multiplexing, the feeder cable must be equipped with the correct combining devices. Two or more frequency bands can be combined using multi-band combiners. Multi-band combiners are often added to a system as separate components, but they can also be built directly into other components such as antennas.

Widely known as crossband couplers, these combiners may be referred to as diplexers (two frequencies), triplexers (three frequencies), and so forth according to the number of frequency paths involved (figures 4.2 and 4.3).

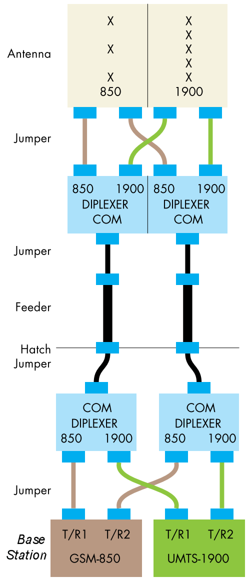
Antenna
X X X
850
X X X X X
1900
Jumper
850 1900 DIPLEXER COM
850 1900 DIPLEXER COM
Jumper
Feeder
Hatch
Jumper
COM DIPLEXER 850 1900
COM DIPLEXER 850 1900
Jumper
Base Station
T/R1 T/R2 GSM-850
T/R1 T/R2 UMTS-1900

4.2: *Shared feeders using diplex crossband couplers*

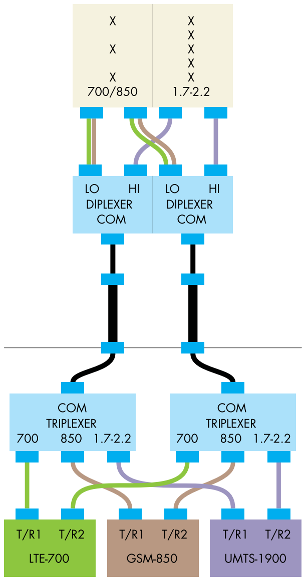
X X X
700/850
X X X X X
1.7-2.2
LO HI DIPLEXER COM
LO HI DIPLEXER COM
COM TRIPLEXER 700 850 1.7-2.2
COM TRIPLEXER 700 850 1.7-2.2
T/R1 T/R2 LTE-700
T/R1 T/R2 GSM-850
T/R1 T/R2 UMTS-1900

4.3: Shared feeders using triplex crossband couplers, with broadband antennas using diplex crossband couplers

Multi-band combining
A configuration that connects multiple base station services that operate in separate bands to multiple antennas via a single feeder cable and its associated couplers.

The kind of crossband couplers required in a particular application is largely determined by the frequencies the system uses, and, more specifically, how far apart from each other those frequencies are. In systems with wide frequency separation such as 700-1000 MHz, 1700-2200 MHz and 2400-2700 MHz the needed crossband couplers are likely to be low-cost, compact devices that introduce virtually no loss or mismatch.

However, when dealing with frequencies that are relatively close to one another, such as 700 MHz and 850 MHz, the appropriate crossband coupler grows more complex, bulky and expensive (figure 4.4).

On the antenna side of the connection, additional efficiencies can be gained among broadband antennas that can accept more than one frequency through a single port. This allows it to operate over a range of bands through one feeder cable, as shown in figure 4.3.

Like the other circumstances involved in planning an efficient and compliant base station site, antenna selection and the base stations' assigned frequencies can play a large part in how a particular co-siting solution comes together.

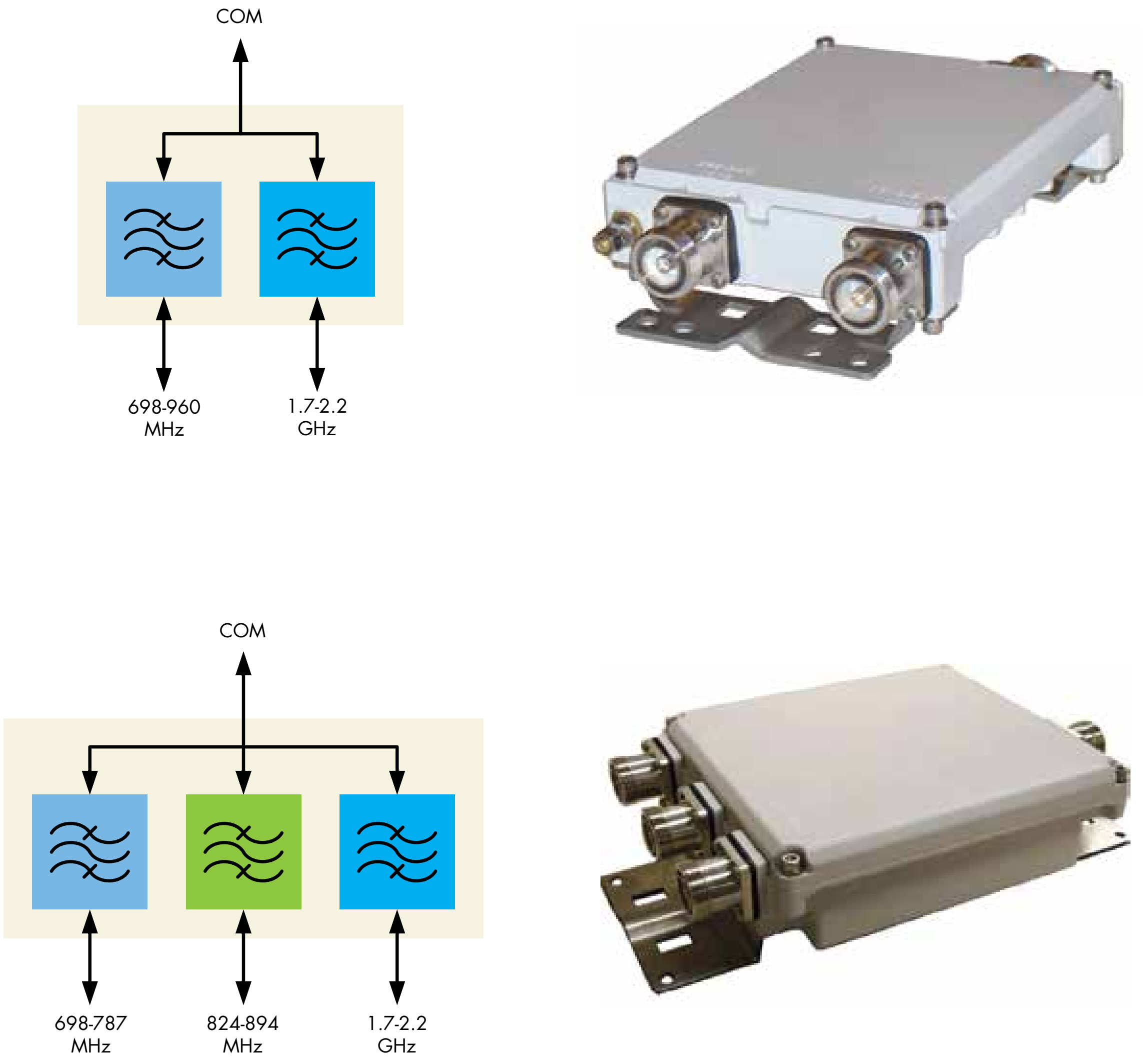

4.4: Compact diplex and triplex crossband couplers, with example frequency differentiation

Same-band combining

In some instances, multiple services require the use of the same frequency band. When this happens, regular crossband couplers, which are designed to suit specific frequency separation, don't provide the solution we need. Instead, we can use a variety of *same-band combining (SBC)* options, which can allow different services to share the same space on the electromagnetic spectrum.

In some applications, same-band combining is even used for single-service systems—not to allow other services, but to increase the channels available to the one operating service. In all cases, the idea is to combine transmit signals (TX) and divide receive signals (RX). The best way to achieve this depends on the specifics of the application.

Now let's look at some of the more commonly used techniques.

Hybrid combining

Hybrid combiners offer a low-cost means of combining TX signals and dividing RX signals (figure 4.5), but this advantage comes at the cost of other operational restrictions inherent in its design.

The main disadvantage of this technique is the high rate of loss experienced in both directions. This loss increases with the number of ports involved, so hybrid combiners are generally used only in two-port applications.

Another consideration is the significant heat it generates which must be dissipated, adding costs and creating even more design limitations. These drawbacks limit the practicality of hybrid combining to in-building coverage and similar uses. It is rarely used in cellular sites.

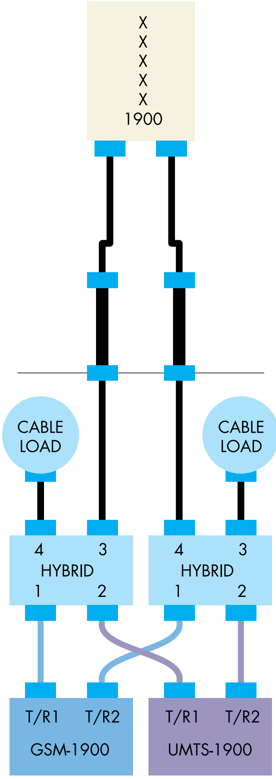

4.5: A hybrid combiner, using cable load to lower passive intermodulation

Same-band combining (SBC)
A base station configuration that allows multiple services to share the same bands.

Guard bands

Narrow gaps inserted into the bandwidths managed by the LLC to distinguish different signals riding on the same bands.

Low loss combiner-multiplexer

The low loss combiner (LLC) offers a different way to combine base station transmitters. Commonly employed for combining TX signals, integrating a duplexer allows for distribution of RX signals as well (figure 4.6).

Like crossband couplers discussed earlier, the LLC is a filter multiplexer. However, unlike a crossband coupler that requires spaces between bands—you'll recall that the bigger the spaces, the better the coupler operates—the LLC handles frequencies inside the same bandwidth.

This is possible due to the addition of guard bands, which act as very small gaps within the band. They create boundary spaces between the frequencies, allowing them to be distinguished from one another.

Including these tiny *guard bands* often requires those narrow frequencies to be left unused, which adds up to slight bandwidth loss. In LLC design, smaller guard bands incur greater cost, size and complexity, so an economical alternative is to re-use the "lost" guard band space with a second feeder and antenna.

LLC design significantly reduces insertion loss over that of a hybrid combiner, but its reliance on filter multiplexing places significant restrictions on its scalability. As technology develops, networks require constant upgrading, adjusting and scaling which often means the adjustment or replacement of the LLC component. Recent developments in remotely tuned LLC hardware have helped reduce this limitation, but it remains a significant drawback for many applications.

Several examples of LLC realizations are shown in figures 4.7 and 4.8.

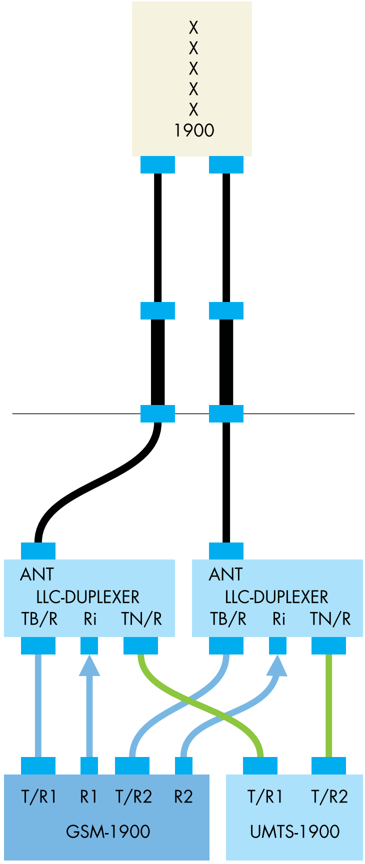
X
X
X
X
X
1900
ANT
LLC-DUPLEXER
TB/R
Ri
TN/R
ANT
LLC-DUPLEXER
TB/R
Ri
TN/R
T/R1
R1
T/R2
R2
GSM-1900
T/R1
T/R2
UMTS-1900

4.6: An LLC with integrated duplexer; RX distribution from GSM BTS

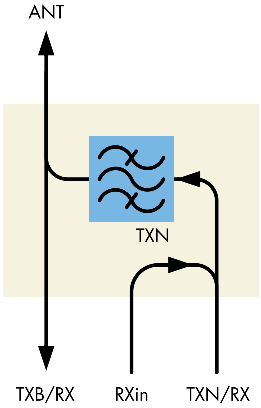
ANT
TXN
TXB/RX
RXin
TXN/RX

4.7: An LLC combines a narrow portion of TX band into broadband path; includes duplexer for RX re-injection

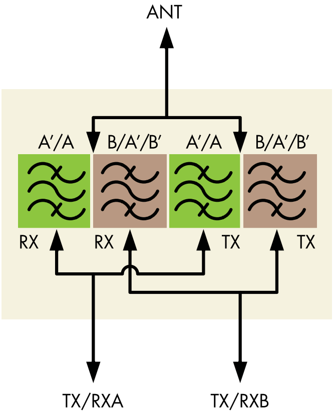
ANT
A′/A
B/A′/B′
A′/A
B/A′/B′
RX
RX
TX
TX
TX/RXA
TX/RXB

4.8: Filter multiplexer for downlink and uplink—a quadruplexer

Amplification

Another technology that makes co-siting possible is amplification. There are several ways amplification can be used in support of the devices listed above, expanding their utility, power and range.

Single carrier power amplifier. You will recall that one of the hybrid combiner's drawbacks was its high rate of insertion loss for both TX and RX signals. One way to compensate for this is to add a single carrier power amplifier (SCPA) (figure 4.9).

The SCPA is highly efficient in regard to power consumption, but is only suited to certain engineering standards, such as the Global System for Mobile Communications (GSM) established by the European Telecommunications Standards Institute (ETSI). Still, in those applications where the SCPA is appropriate, the SCPA offers a low-cost means of improving hybrid combiner performance (figure 4.10).

Multi-carrier power amplifier. Like the SCPA, the multi-carrier power amplifier (MCPA) is a high-power amplifier for carrier signals. Unlike the SCPA, however, the MCPA can also combine multiple RF signals into a single output. Its input circuits can be expanded to accommodate from two to eight ports, and sometimes even more (figure 4.11).

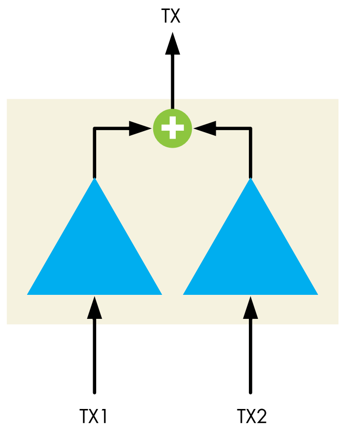

4.9: An SCPA module amplifies and combines two transmitters into one path

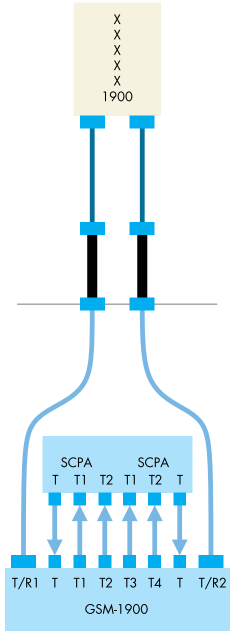

4.10: An SCPA installed between radios and BTS duplexer

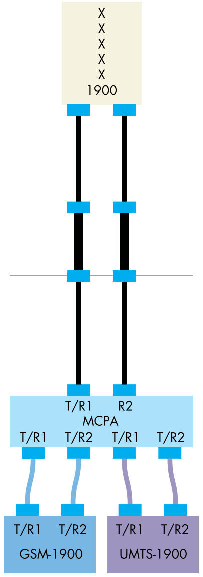

4.11: An MCPA with integrated RX distribution

The MCPA's strong suit is boosting transmit power to increase the coverage or capacity of a particular base station. It also offers complete frequency agility, allowing free use of all frequencies within a license band. This makes it a very easy system to interface to other technologies. Its design generally incorporates one or more amplifier "bricks" working in parallel to provide the necessary power (figures 4.12 and 4.13).

An MCPA's demanding power consumption is its most notable drawback, which leads to elevated implementation and operational costs.

Receiver multicoupler. As its name implies, the receiver multicoupler (RXMC) distributes RX signals from shared antennas to multiple receivers. By splitting the signal this way, a natural side effect is some loss of power. To compensate, the RX input first crosses a low-noise amplifier (LNA) which preserves signal strength, preceded by a preselector filter. Inclusion of an LNA is recommended for most applications with more than two receivers.

As a rule, the RXMC distributes the full RX frequency band to all outputs with the same degree of gain across the board. To individualize distributions, the RXMC may allocate specific signal strength to each receiver by unequally dividing the gain. This is a useful option when dealing with different locations, or with a receiver that will further divide its signal to other receivers down the line.

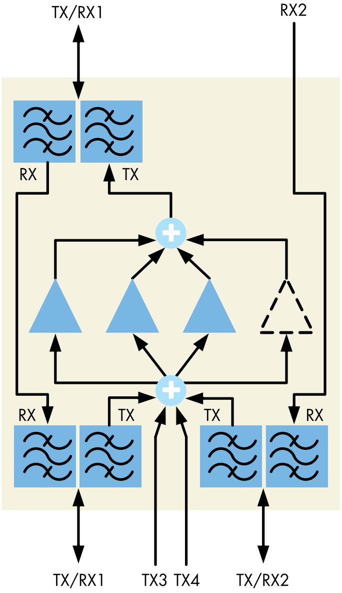

4.12: An MCPA for one sector with two duplexed inputs and two simplex inputs; three amplifier bricks are working in parallel

4.13: An MCPA for three sectors, two duplexed inputs and six simplex inputs per sector, one amplifier brick per sector, plus one hot standby brick

Other integrated configurations. The following figures illustrate other same-band combining technologies (figures 4.14 through 4.18).

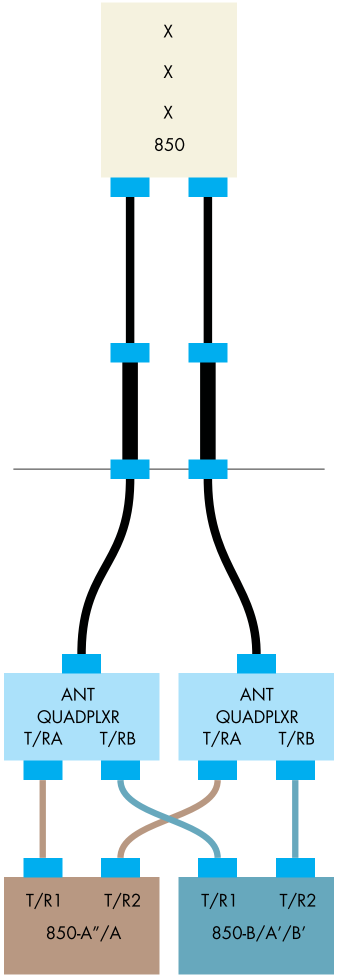
X
X
X
850
ANT
QUADPLXR
T/RA
T/RB
ANT
QUADPLXR
T/RA
T/RB
T/R1
T/R2
850-A"/A
T/R1
T/R2
850-B/A'/B'

4.14: *Antenna sharing accomplished with a TX/RX quadruplexer*

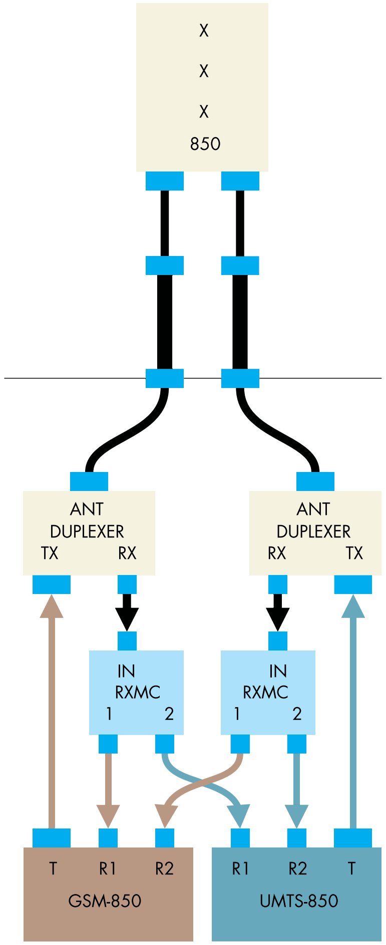
X
X
X
850
ANT
DUPLEXER
TX
RX
ANT
DUPLEXER
RX
TX
IN
RXMC
1
2
IN
RXMC
1
2
T
R1
R2
GSM-850
R1
R2
T
UMTS-850

4.15: *RX distribution to simplex BTS using duplexers and RXMC*

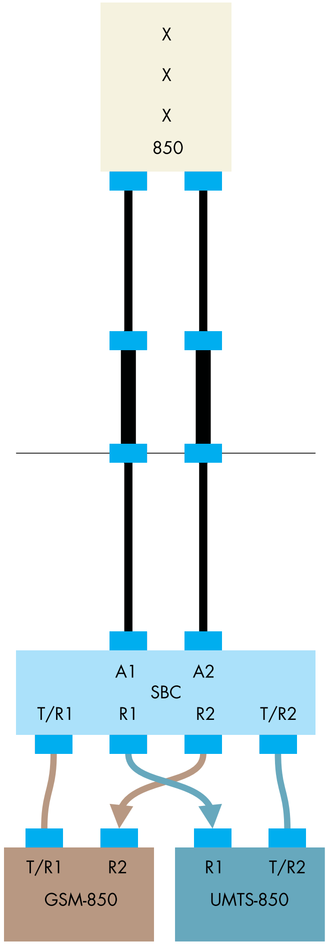
X
X
X
850
A1
A2
SBC
T/R1
R1
R2
T/R2
T/R1
R2
GSM-850
R1
T/R2
UMTS-850

4.16: *Antenna sharing accomplished with integrated Duplex/RX SBC*

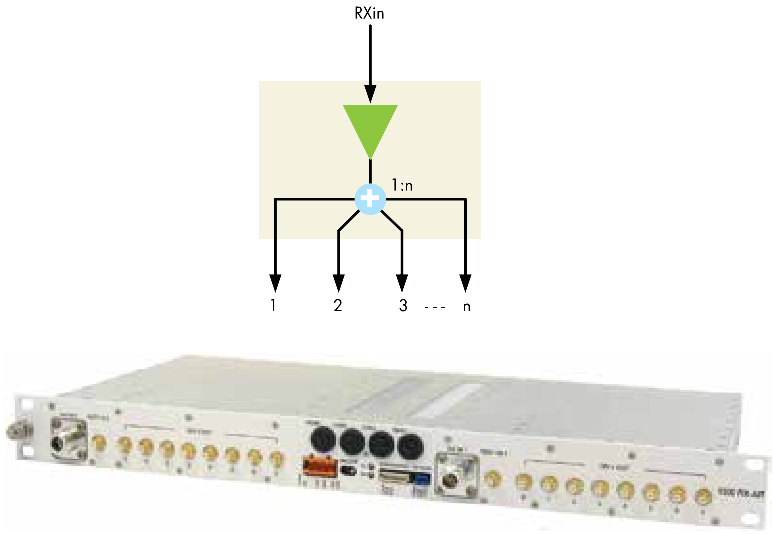

4.17: *Dual RXMC with eight outputs per channel*

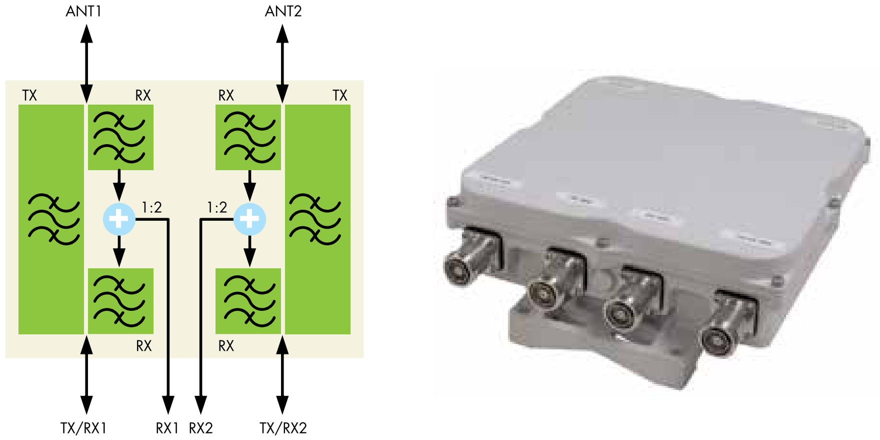

4.18: Integrated SBC

Tower-mounted amplifiers. Working with these other co-siting solutions, a tower-mounted amplifier (TMA) improves the base station's sensitivity the same way a hearing aid can improve diminished hearing. It works to offset the losses experienced by RX signals as they travel to the receiver. This improvement in signal clarity is seen in the carrier-to-noise ratio (C/N), measured in decibels (dB). Adding a TMA to the RX circuit on a cellular base station can yield a typical improvement of 5-6 dB.

TMAs are a key part of RF path technology. Properly implemented and configured, they improve:

Coverage. They boost the effective service radius of a cell base station while improving signals in weak spots, such as indoors.

Accessibility. They significantly reduce failed access attempts.

Retainability. They improve a site's retainability, or its ability to maintain connections within and across cells for fewer dropped calls.

Co-channel interference. They improve call capacity in spread spectrum systems.

Data throughput. They enable higher order modulations for increased traffic capacity.

Handset battery life. Because less power is required from the cell phone's transmitter, TMAs prevent unnecessary battery drain.

With the right adjustments, TMAs offer all these advantages that help operators enhance their network capacity and coverage.

TMA benefits

To gain the maximum advantage, it is important to boost the RX signal level before it becomes degraded by feeder loss between the tower and the base station receiver. This is the reason the TMA should be mounted at the tower top, as close to the RX antenna as possible.

An LNA may also be installed at the ground level and is then called a ground-mounted amplifier (GMA). When it reaches the GMA, the RX signal will already be weaker and noisier than it originally was at the antenna. Therefore, the sensitivity improvement a GMA can provide is limited at it's improving an already weakened signal. On the other hand, its main advantage is that it is easier to install than a TMA, and does not occupy precious tower space. Because it doesn't introduce the size and weight concerns of a tower-mounted TMA, the GMA design can focus on performance. Superconductors and cryogenic cooling can be used to maximize performance and partially offset the disadvantage of its ground-level location.

Many factors contribute to feeder loss. Smaller cable diameter, longer feeder runs, and higher operating frequency all influence performance, and the TMA offers a single method of offsetting all these factors with one remedial measure. You can see how TMA and GMA implementations affect the noise levels in a system in figure 4.19.

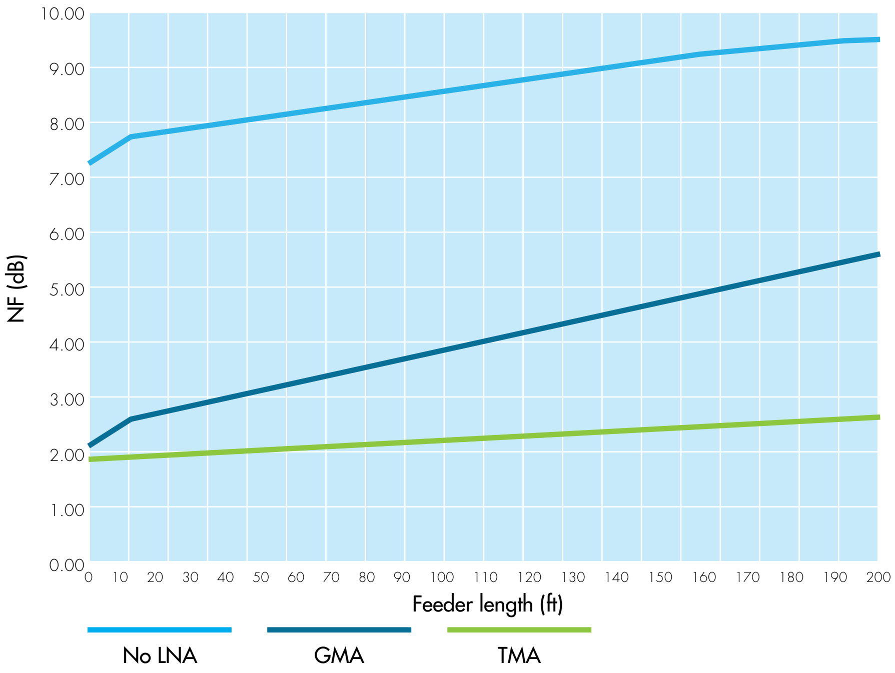

4.19: Noise figure in a system with no LNA, with a GMA and with a TMA; the superior performance of the TMA is evident in its lowest noise figure

TMA configurations

As explained in our earlier look at the receiver multicoupler (RXMC), a tower-mounted amplifier employs an LNA (with a preselector filter) to boost signal strength to compensate for division loss. Since modern TMAs often employ a dual duplex configuration, which allows the use of duplexed feeders, two additional filters are needed to pass the signals between the BTS and antenna.

Dual-band TMAs are essentially a pair of single-band TMAs integrated into one device. Some types feature separate RF paths for each band, while others diplex the bands into a single path at the BTS port or at the ANT port. Interestingly, diplexers can also be integrated into single-band TMAs to offer an additional, non-amplified path. These devices are called *bypass or pass-through configurations*.

As a result of growing demand for reliable cellular network service, coupled with frequent restrictions on the amount of equipment permitted on a particular cell site, dual-band TMAs are becoming more and more popular as a co-siting solution. You can see examples of some different TMA architectures in figures 4.20 through 4.26 beginning on this page and continuing on page 14.

Bypass or pass-through configuration

A single-band tower-mounted antenna with an integrated diplexer that adds a secondary, non-amplified RF path to the system.

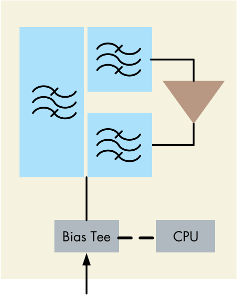

4.20: Single-band TMA

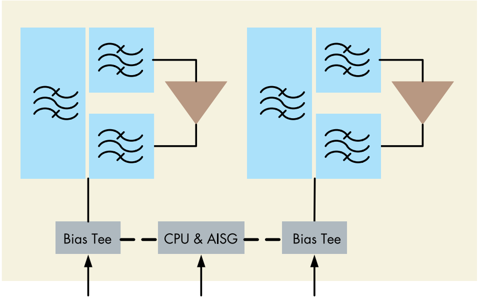

4.21: Twin single-band TMA with AISG support

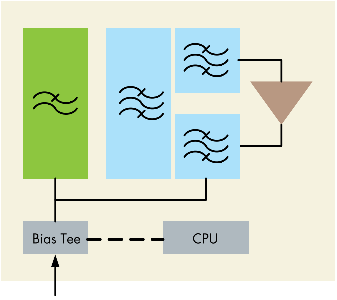

4.22: TMA with integrated diplexer, bypass path

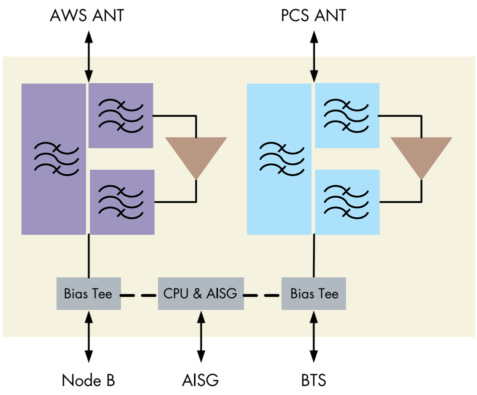
AWS ANT
PCS ANT
Bias Tee
CPU & AISG
Bias Tee
Node B
AISG
BTS

4.23: Dual-band TMA with AISG support

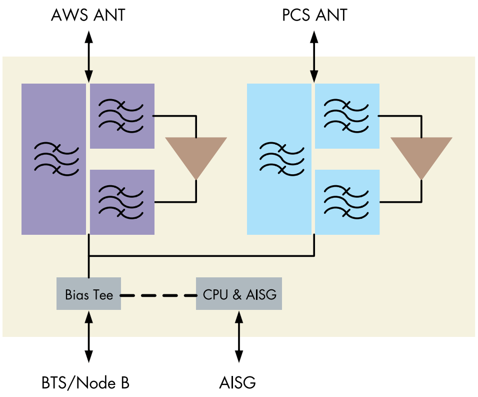
AWS ANT
PCS ANT
Bias Tee
CPU & AISG
BTS/Node B
AISG

4.24: Diplexed dual-band TMA with AISG support

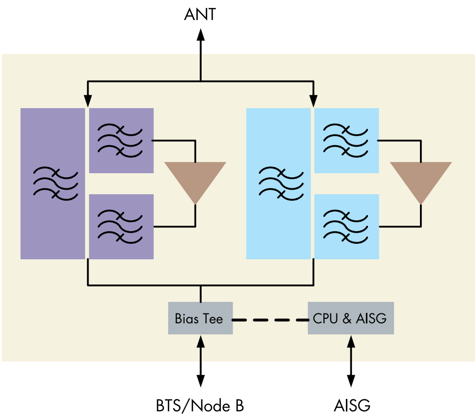
ANT
Bias Tee
CPU & AISG
BTS/Node B
AISG

4.25: Dual diplexed dual-band TMA with AISG

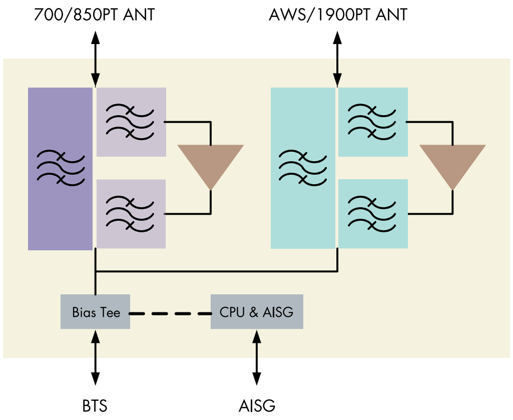
700/850PT ANT
AWS/1900PT ANT
Bias Tee
CPU & AISG
BTS
AISG

4.26: Diplexed dual-band TMA with pass-through and AISG

Enhanced features: AISG

The *Antenna Interface Standards Group (AISG)* is comprised of representatives from the world's leading wireless equipment manufacturers and service providers, including **CommScope.** AISG's standards have led to improvements in remote control and monitoring of antenna downtilt, TMA alarms, and other important advances. Adoption of these protocols has advanced performance and assured interoperability between various systems—a major boon for co-siting solutions like the ones discussed above.

Contemporary co-siting components must accommodate AISG communication, whether they process the signals or simply pass them along to other devices. The AISG protocol currently includes procedures for monitoring and controlling TMAs, such as setting gain and reporting alarms. Additional procedures may be defined in the future to support a wider range of "smart" co-siting devices that will help optimize performance, enable remote supervision of the RF path, and reduce cost of ownership.

Antenna Interface Standards Group (AISG)

Founded in 2001, the AISG represents more than 40 top manufacturers and service providers from all over the world. AISG publishes universally accepted industry protocols for communications between base stations and tower-based equipment, such as antennas and TMAs.

Making the most of available space and power

The design of a cellular communications system reflects many choices and compromises. The result is that no two are exactly alike. Certain preferred characteristics come at the expense of other characteristics; those choices are always made with an eye toward conserving space, reducing costs and operating within constraints.

By employing the kind of innovative solutions discussed in this chapter, wireless operators are able to improve service and reduce costs by working together to share infrastructure wherever possible. We see the benefits in the form of lower cell phone bills, clearer and more reliable calls, and faster data downloads and Web surfing on our mobile devices.

Chapter 4 summary

Co-siting solutions:

- The technology and techniques that allow more performance in less space
- Driven by limits on amount, weight and cost of base and antenna-mounted equipment

Multi-band combining:

- Leverages feeder cable's capacity for multiple frequencies
- Requires frequencies in separated bands
- Uses crossband couplers in diplex, triplex or more complex configurations

Same-band combining:

- Hybrid combining: inexpensive but lossy
- Low loss combiner: efficient but constrained frequency applicability

Amplification:

- Single carrier power amplifier: low cost, high efficiency, limited applications
- Multi-carrier power amplifier: scalable, flexible but expensive to install and operate
- Tower-mounted amplifiers: broad-based receiver performance boost

Chapter 5

Talking and listening at the same time:
Transmission and receiving isolation systems

Right now, millions of people around the world are downloading music, surfing the web, texting, talking and listening on their mobile devices. It's probably safe to say that they are *not* thinking about the science or technology that enables every download, text or conversation. Mobile devices are simply a way of life.

Here at **CommScope,** we're continually fascinated by the technical innovation and principles behind wireless communication. Take transmission and receiving isolation systems, for example. Unlike conventional landline phones, cell phones are actually radio receivers and transmitters, so maintaining simultaneous two-way communication—talking and listening during a call—is more complex than it appears.

Duplex communications
A transmitter and receiver that work at the same time on the same RF device.

Isolation
The amount of separation achieved between the transmitter and receiver in a duplex communication system. In general, more isolation translates to less interference and clearer communications.

An RF communications system that employs this simultaneous, two-way flow of voice, data or other information is called a *duplex system*. Duplex communications systems combine multiple transmit and receive channels on a shared antenna, with information flowing both ways at the same time.

Imagine the simultaneous flow of traffic on a busy two-way street. You immediately see the importance of keeping the two different directions of traffic separated. Just as vehicles on a busy, two-way street require clear lane markings to avoid collisions with oncoming vehicles, duplex RF channels also must be “isolated” from each other to avoid interference.

In RF terms, *isolation* is measured as the loss between two channel ports, either transmitter-to-transmitter or transmitter-to-receiver ports. The higher the loss, or isolation, between the two ports, the cleaner the signal.

To illustrate this concept, think about making a cell phone call from your car. This simplest of duplex systems—one transmitter and receiver pair communicating with another transmitter and receiver pair—requires that both the phone and receiving station be able to receive and transmit at the same time, allowing a normal telephone conversation to take place (figure 5.1).

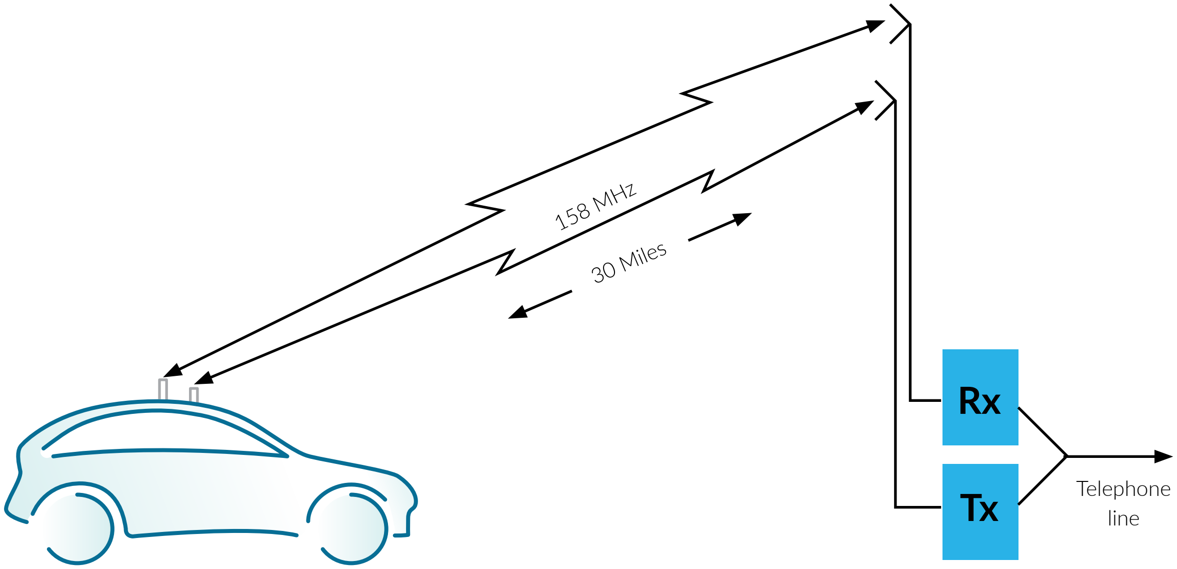

5.1: Duplex operation between two pairs of transmitters and receivers

To allow this communication to flow on a single antenna, a *duplexer* must be used with adequate isolation measures (figure 5.2). Measured in dB, isolation is a critical consideration in the design of any duplex system. Without proper isolation, a transmitter will adversely affect the performance of its associated receiver, even though they may operate on different frequencies.

Duplexer
A device situated between a duplexed antenna and its associated transmitter and receiver. A duplexer's function is to provide isolation between the signals.

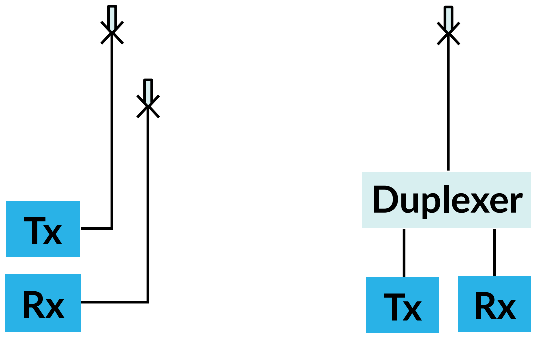

5.2: Two solutions: Use two antennas, or a single antenna with a duplexer

The specifications covering a particular receiver, for instance, may indicate that any RF signal outside the receiver's passband (which can be as narrow as 15 kHz) will be attenuated, or weakened, by as much as 100 dB. That means that the transmission's power will be reduced to 1/10,000,000,000th of its original strength, making the communication unintelligible and useless in most cases.

You might think that such a selective receiver would prevent interference from a transmitter operating on a frequency far outside the receiver's passband. After all, if the interfering signal is 5 MHz away, how could it create complications when just being 5 kHz off the mark reduces the transmitter's signal to virtually nothing? The answer lies in the characteristics of modern receivers, and the way they can step high-frequency signals downward to achieve such precise frequency selectivity.

Receiver desensitization
Interference caused by unwanted frequencies entering a receiver's upper-stage passbands. These errant signals create electrical variances that impede the receiver's operation.

The first challenge: receiver desensitization

Receiver desensitization is an inherent side effect of modern receiver design, which receives relatively high-frequency signals (often between 700 MHz and 3500 MHz). These signals pass through frequency-lowering stages in the receivers, which allow the receivers to feature such narrow, selective passbands (figure 5.3). Once the signal has been lowered enough, only a small band remains and the circuitry can reject other bands within a margin measured in dB. A receiver's specification sheet will include this measurement of overall selectivity.

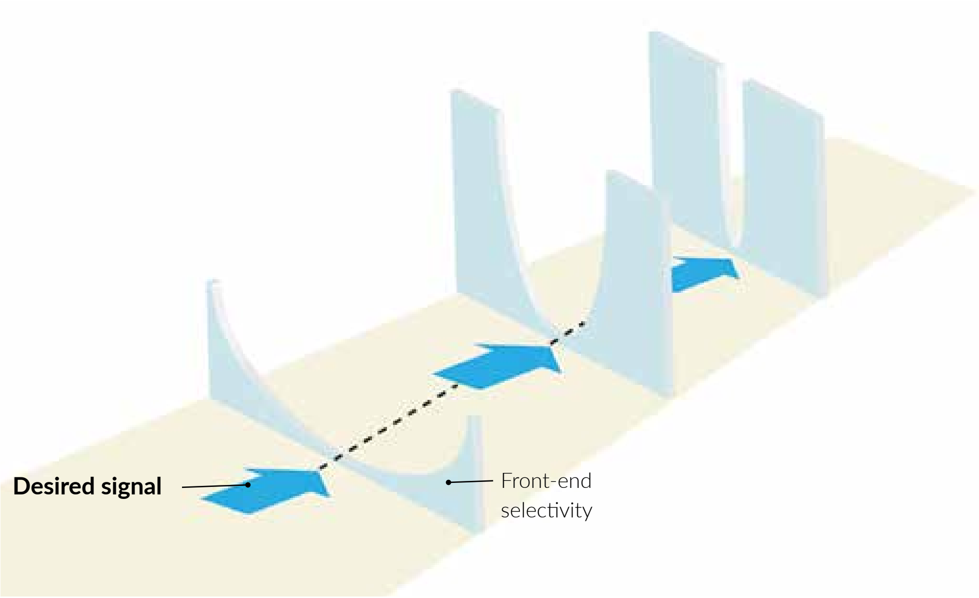

5.3: In a receiver, high-frequency signals are reduced in stages

The vulnerability is not at the end of this reducing process, but at its beginning. Remember that the initial signal was of higher frequency, and only after multiple stages of reduction was it lowered to the point where the receiver could use it. The receiver's earlier, broader stages cannot completely reject errant signals, even those several MHz away from the receiver's operating frequency.

For optimum performance, critical voltage and current levels exist at certain points throughout the front-end stages of a receiver. If these levels change significantly, the performance of the receiver suffers. This happens when a nearby transmitter's off-frequency signal enters the front-end stage.

Such signals can be several MHz away from a receiving frequency, and radiate from sources several thousand feet away, and still cause significant interference (figure 5.4).

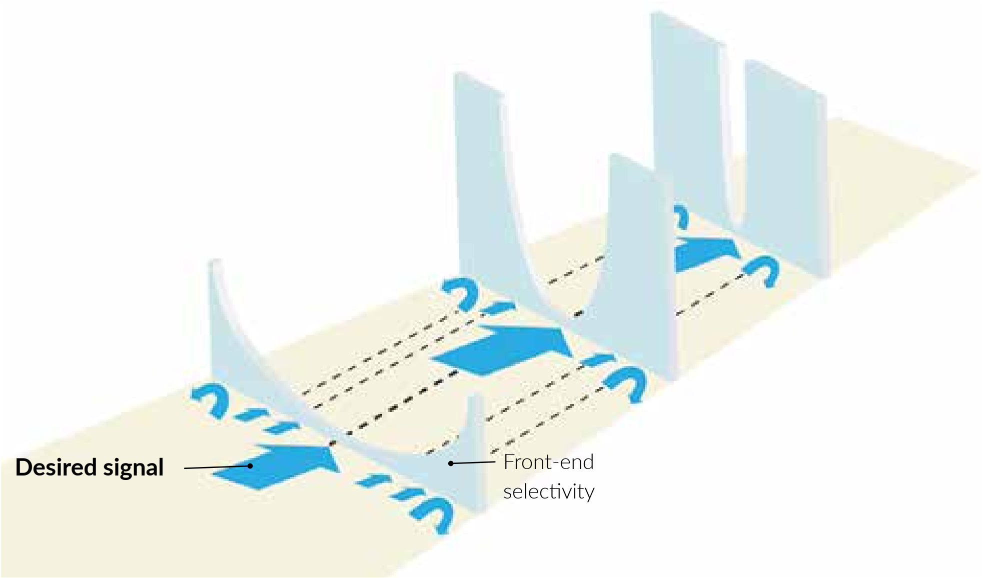

5.4: Unwanted frequencies (shown here as reflected arrows) can alter critical receiver voltage and current levels

Transmitter noise
Interference experienced by a receiver as a result of transmission power "leaking" into other nearby frequencies.

The second challenge: transmitter noise

Transmitter noise is interference caused by carrier signals just outside of a transmitter's assigned frequency. In an ideal world, a transmitter would channel 100 percent of its signal power into the narrow band of frequencies assigned to its transmission channel. In the real world, however, this level of precision is simply not possible, and the result is called transmitter broadband noise radiation, or more commonly, transmitter noise. While the vast majority of transmission power remains within the assigned channel, there remains a small fraction that "leaks" into channels above and below the intended carrier frequency.

Modern transmitters are equipped with filter circuits that eliminate a large portion of these errant signals, but even with these measures in place, enough transmitter noise escapes to degrade the performance of a receiver. As the chart below illustrates, this interference effect is most pronounced at frequencies closest to the transmitter's carrier frequency (figure 5.5), but can also impact receivers operating several MHz away.

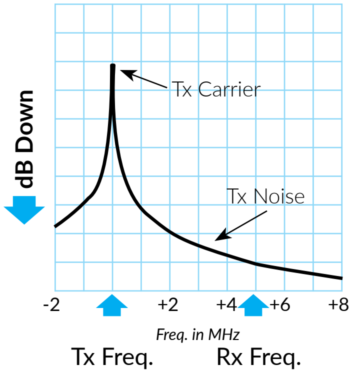

5.5: Transmitter interference is most pronounced near the assigned frequency (shown here as Tx frequency, located at zero on horizontal axis)

We hear transmitter noise in a receiver as "on-channel" noise interference. Because it falls within the receiver's operating frequency, it competes with the desired signal and cannot be filtered out.

To illustrate this kind of interference, imagine having a conversation with someone in a crowded room. If everyone else is talking, you'll notice how hard it is to understand the other person, even if the overall noise level in the room is relatively low. That's because other voices, like unwanted transmitter noise, are similar to the voice you're trying to hear.

This is a key distinction between transmitter noise and receiver desensitization, which you'll recall comes from signals far from the operating frequency of the receiver. Consider again the illustration of having a conversation. Receiver desensitization is more like loud, disruptive sounds coming from a construction site next door. The interference is not similar to the voice you're trying to hear, but it still distracts you from the other person's voice.

How isolation helps duplex communications overcome both challenges

In duplex RF systems, transmitting and receiving frequencies are close to each other. In addition, the antennas will also be physically close, or even share a single antenna. Now that we understand the source and nature of the two interfering elements—receiver desensitization and transmitter noise—how can we overcome these interfering influences and assure reliable operation of our paired transmitters and receivers?

The answer, as you may have guessed, is proper isolation.

Earlier in this chapter, we explored how a duplex RF system required isolation between transmitter and receiver using the example of a call on a mobile device. But when applying that theory to practical application, adding isolation to the system requires some planning and a bit of math. Remember that we have not one but two sources of interference to overcome—receiver desensitization and transmitter noise—and each requires its own solution.

It boils down to two simple questions:

1. How much isolation is required **to prevent receiver desensitization** from the transmitter's carrier, and;
2. How much isolation is required **to reduce transmitter noise** to a lower, even negligible level?

While these are simple questions, each one has many more questions built into it, such as, but not limited to:

- How close together are the transmitter and receiver frequencies?
- What frequency band are we using?
- What is the transmitter's power output?
- What are the unique product specifications for the particular transmitter and receiver we're using?

While each application will have very different answers for these and other considerations, you can usually find the answers in the equipment manufacturer's data. For the purposes of this discussion, we'll focus instead on the broader use of isolation in duplex systems.

Determining the amount of required isolation is a matter of examining both sources of interference and identifying the optimal isolation level. As shown below (figure 5.6), we see the effect of frequency on both interfering influences, receiver desensitization (dotted line) and transmitter noise (solid line).

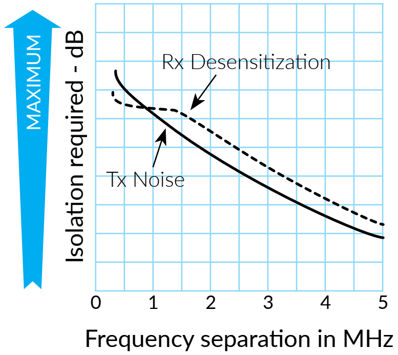

5.6: The effect of frequency separation on receiver desensitization and transmitter noise

In short, the closer the frequencies are to one another, the greater the need for isolation. For instance, the chart shows that reducing the frequency separation from 5 MHz to 1 MHz requires double the isolation to assure that the receiver will not be sensitized and that transmitter noise will be reduced to negligible levels.

Horizontal separation
The practice of placing a transmitter's antenna a certain distance from the same device's receiving antenna to achieve the necessary isolation.

Achieving sufficient isolation

Once we have determined the requisite amount of isolation by answering our specification questions and applying what we learned from figure 5.6, we can implement the correct degree of isolation by one of two methods:

1. Use two antennas, physically separated by a given distance, or;
2. Use the appropriate duplexer with a single-antenna system.

Let's examine the first option of two physically separated antennas. Within this option, there are two ways of achieving the desired result: horizontal and vertical separation.

Method 1: Two antennas—horizontal separation

If you've ever driven cross-country with the car radio on and heard a favorite song fade to static in mid-chorus, you've experienced an effect called propagation loss. Propagation loss describes the way an RF signal loses intensity and weakens (or attenuates) as it travels across distance. This effect means that placing the two antennas apart—creating *horizontal separation*—yields a certain amount of isolation, simply by virtue of signal attenuation in the space between them (figure 5.7).

With enough distance, we can achieve virtually perfect isolation and total protection from both receiver desensitization and transmitter noise. However, even the most isolated RF system is vulnerable to interference from outside sources located nearby.

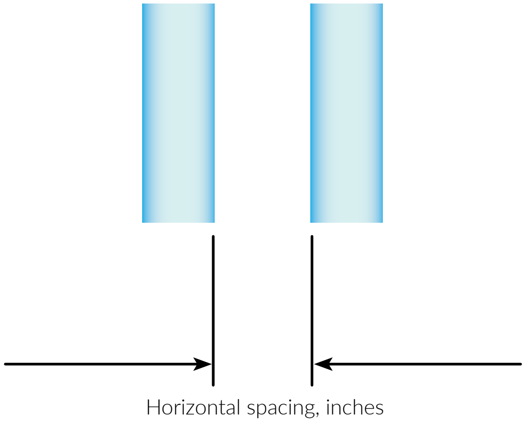

5.7: Horizontal antenna spacing and isolation

Method 1 alternative: Two antennas—vertical separation

Alternately, one may achieve the same isolating effect by separating the transmitter and receiver vertically, a practice called *vertical separation*. In real-world applications, this option is more convenient and efficient as it allows both transmitter and receiver to be mounted on a single tower, one above the other, separated by the requisite distance to achieve sufficient isolation.

A secondary benefit of vertical separation is that this arrangement takes advantage of what is known as the “cone of silence” that exists between vertically stacked antennas (figures 5.8A and 5.8B).

The cone of silence is a dead zone (technically known as a null or lack of gain) that extends above and below communications antennas, allowing each to operate in the other’s shadow, so to speak.

Vertical separation
The practice of placing a transmitter and receiver in separate locations on a single antenna, allowing the height difference to achieve the necessary isolation.

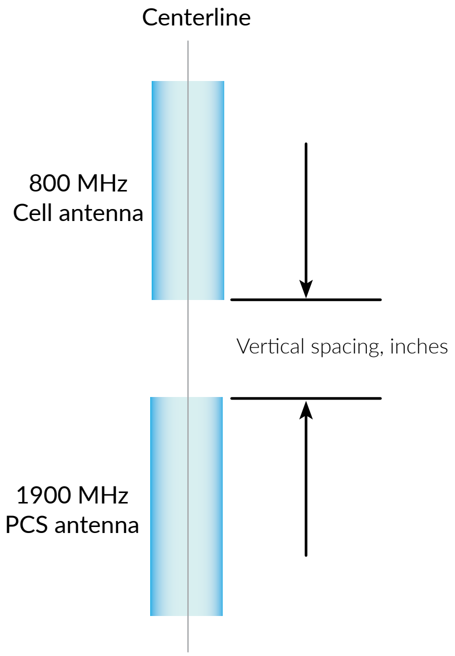
Centerline
800 MHz
Cell antenna
Vertical spacing, inches
1900 MHz
PCS antenna

5.8A: Vertical antenna separation and isolation

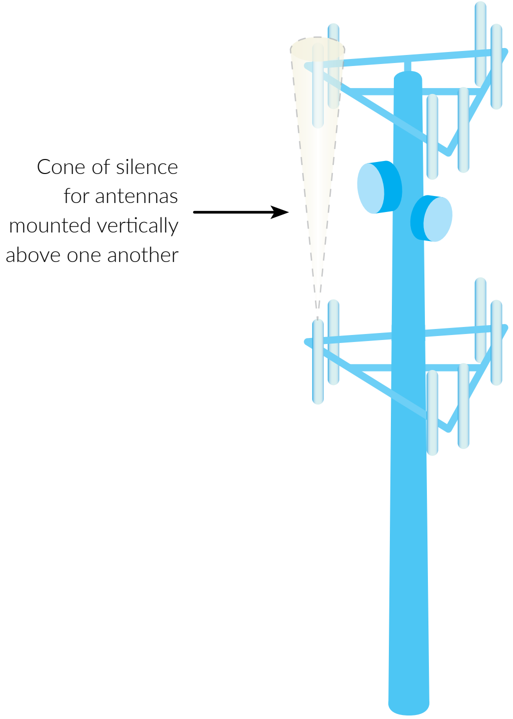
Cone of silence
for antennas
mounted vertically
above one another

5.8B: The “cone of silence” isolates antennas directly above and below each other

Comparing the characteristics of each quickly reveals the superiority of vertical separation in practical applications (figure 5.8C).

It should also be noted, however, that the effects of horizontal and vertical separation are not directly additive. In other words, using both methods on the same system will not yield the full, combined isolation of each. Antenna manufacturers can supply specific figures on what you can expect from combining methods in any particular application.

● Vertical spacing 0 deg d/t

● Vertical spacing 10 deg d/t

● Horizontal spacing 0 deg d/t

5.8C: Graphs at right describing attenuation (in dB) against separation (in feet) for both horizontally and vertically separated antenna pairs

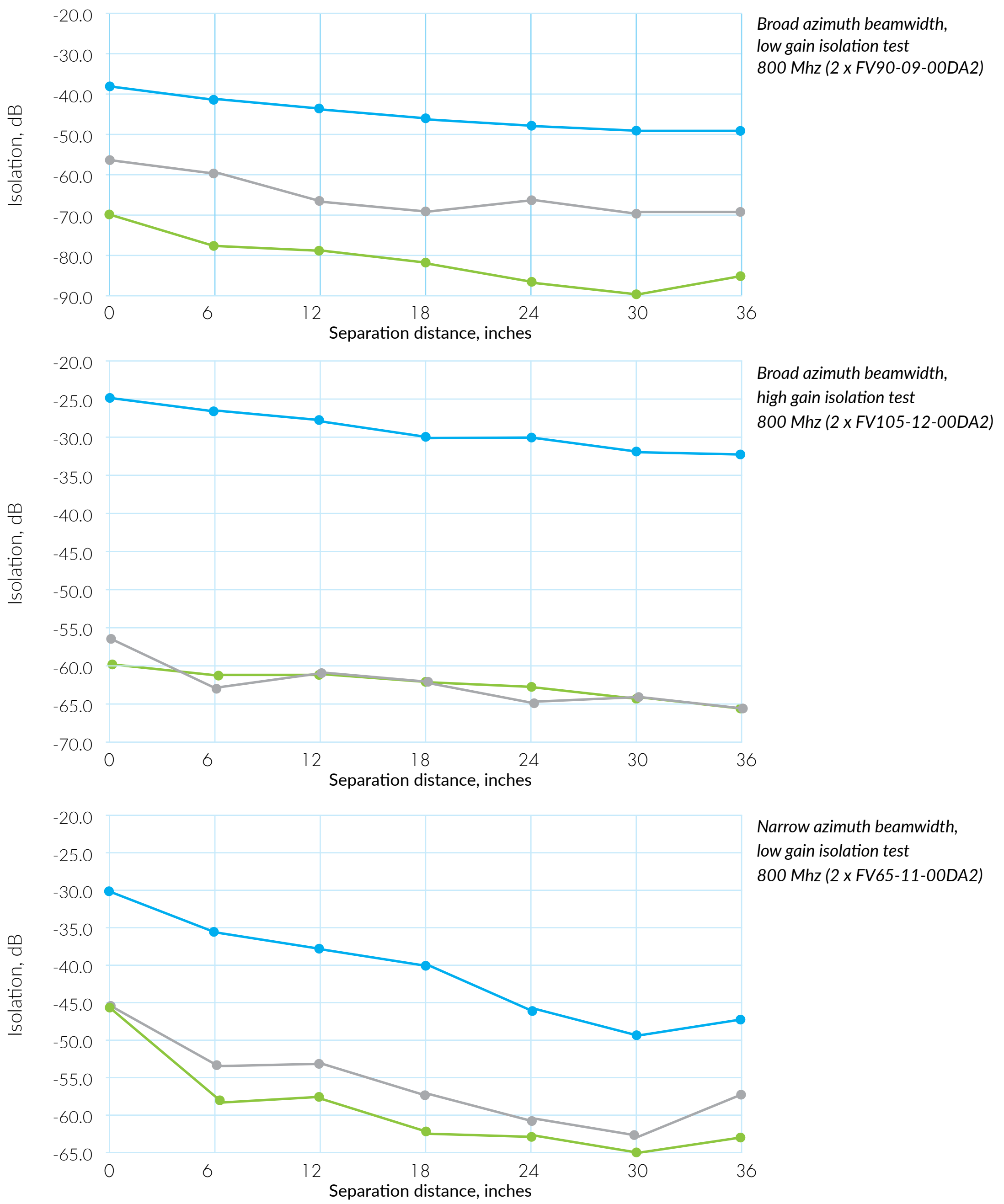

Broad azimuth beamwidth, low gain isolation test 800 Mhz (2 x FV90-09-00DA2)

Broad azimuth beamwidth, high gain isolation test 800 Mhz (2 x FV105-12-00DA2)

Narrow azimuth beamwidth, low gain isolation test 800 Mhz (2 x FV65-11-00DA2)

Method 2: One antenna with a duplexer

The other method of achieving the required isolation between transmitter and receiver is the use of a duplexer in a single-antenna system. A duplexer replaces one of the two antennas and two lengths of coaxial cable by allowing both transmitter and receiver to operate at the same time, on the same antenna (figure 5.9).

The cost benefits from this option can be significant, as a duplexer cuts the needed infrastructure in half. But the cost benefits are secondary to the other advantages, including:

Isolation. A duplexer reliably isolates transmitter and receiver, regardless of external circumstances or terrain.

Antenna pattern. Without a duplexer, two separated antennas are required. Whether arranged in a horizontally or vertically separated configuration, they cannot occupy the same space. This separation means that the coverage area of either the transmitter or receiver may be larger or smaller than the other, a variable that a duplexer eliminates.

Tower space. Leasing tower space is expensive and space on the tower is at a premium. By building or leasing only one tower instead of two, operators can realize lower total cost of ownership.

These are all good reasons to opt for a single antenna with a duplexer, but, as with every advantage in engineering, there are drawbacks as well. We'll explore these considerations a bit further on, but, for now, let's examine how a duplexer actually works and how to choose the right one for a particular application.

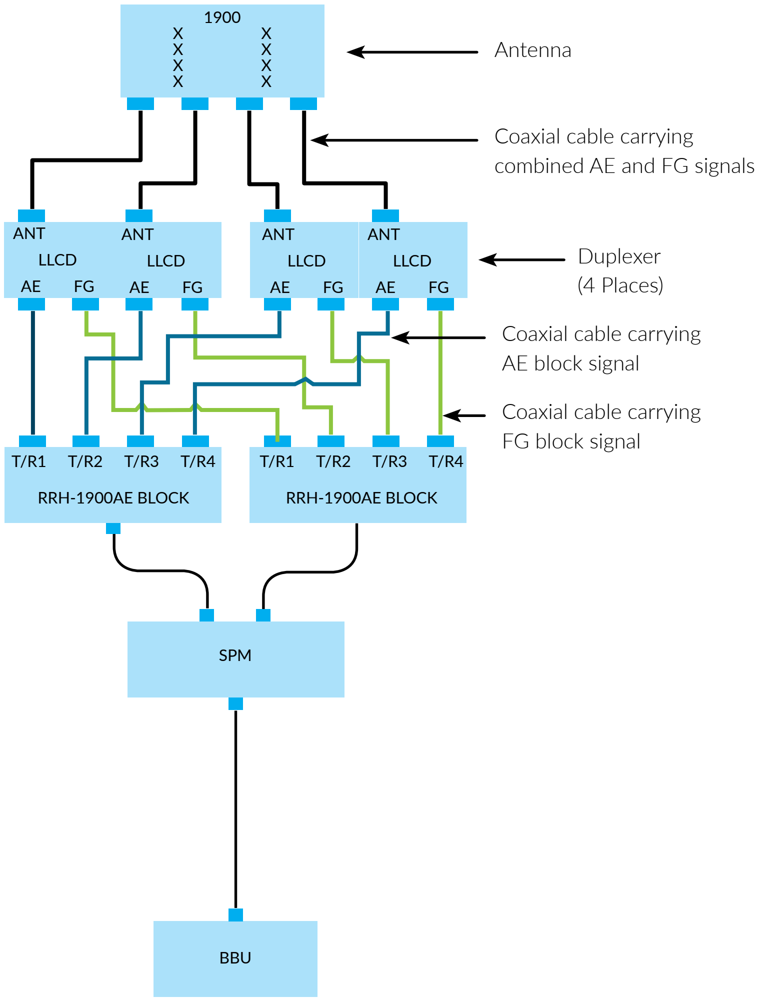
1900
X X X X
X X X X
Antenna
Coaxial cable carrying combined AE and FG signals
ANT
LLCD
AE
FG
Duplexer (4 Places)
Coaxial cable carrying AE block signal
Coaxial cable carrying FG block signal
T/R1
T/R2
T/R3
T/R4
RRH-1900AE BLOCK
RRH-1900AE BLOCK
SPM
BBU

5.9: A typical cell site duplexer configuration

Considerations in choosing a duplexer

There are two distinct families of duplexers used in two-way RF communications: the bandpass duplexer and the band-reject duplexer. To choose the right option, it's important to consider the requirements of the system and what the correct duplexer must do. For the best results, a duplexer must:

1. Be designed to operate in the system's frequency band
2. Be able to handle the transmitter's power output
3. Operate at or below the system's frequency separation
4. Create minimal power loss to transmit and receive signals
5. Provide at least a minimum level of transmitter noise rejection
6. Deliver sufficient isolation to prevent receiver desensitization

These last two factors relate specifically to isolation and prevention of interference. In both cases, protection must meet a minimum threshold, but there is no hard upper limit, and no harm in exceeding specified isolation levels.

Losses through the duplexer

As a matter of course, the signal strength of both the transmitter and the receiver are reduced slightly in the process of passing through the duplexer. These losses are called *insertion loss: transmitter to antenna* and *insertion loss: receiver to antenna*. Like losses caused by other forms of attenuation, these duplexer losses are measured in dB and tend to increase as frequency separation between transmitter and receiver decreases (figure 5.10).

Insertion loss: transmitter to antenna	
Insertion loss	Reduced power (watts)
0.5 dB	11%
1.0 dB	20%
2.0 dB	37%

Insertion loss: receiver to antenna	
Insertion Loss	Reduced strength (microvolts)
0.5 dB	5%
1.0 dB	11%
2.0 dB	20%

5.10: Equivalent signal power loss at discrete duplexer insertion loss levels

The bandpass cavity

The distinguishing feature of the bandpass duplexer's design is the *bandpass cavity*. The bandpass cavity works as a filter of RF frequencies, allowing a narrow band of desired frequencies to reach the receiver and attenuating frequencies outside this band. Energy is fed into the cavity by means of a coupling loop. This energizes the resonant circuit formed by the inner and outer conductors. A second loop couples energy from the resonant circuit to the output (figure 5.11). The loops determine the selectivity of the bandpass cavity.

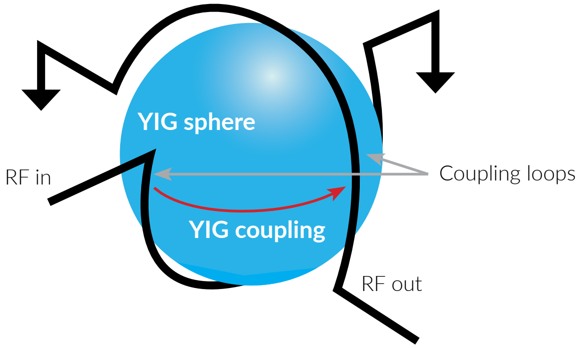

5.11: The operation of a coupling loop

The narrow band of desired frequencies that pass through the cavity experience only slight loss and are all within a few thousand cycles of the cavity's *resonant frequency*. The effect of multiple frequencies, transmitted at equal power, on a bandpass cavity is illustrated below (figure 5.12).

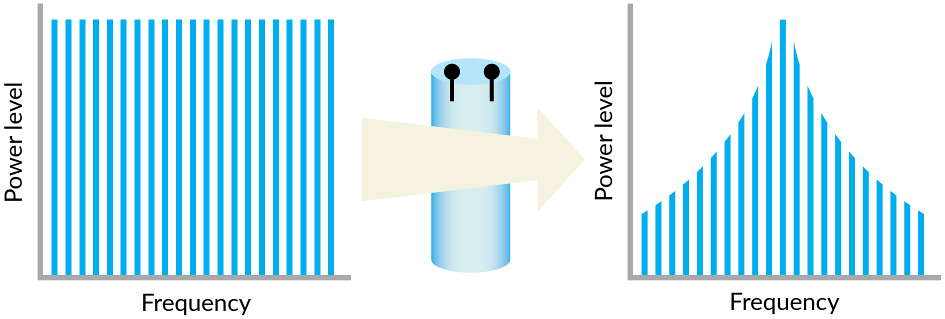

5.12: Attenuation of undesired frequencies passing through a bandpass duplexer

Bandpass cavity
A "frequency filter" that limits the channels that pass through the filter to a select set of frequencies. Other frequencies are prevented from passing. Most devices have multi-stage bandpass cavities that filter out different frequencies at each stage.

Resonant frequency
The natural tendency of a system to oscillate with larger amplitude at particular frequencies. At these frequencies, even small periodic driving forces can produce large amplitude oscillations.

The selectivity of a bandpass cavity is usually illustrated in a frequency response curve. This curve describes the degree of attenuation provided by the cavity at discrete frequencies above and below the cavity's resonant frequency. The curve also shows the insertion loss to the desired signal at the cavity's resonant frequency (figure 5.13).

In cases where a single bandpass cavity cannot provide enough rejection to undesired frequencies, the addition of more cavities in sequence can further refine selectivity. While this results in additional insertion loss, selectivity can be increased substantially.

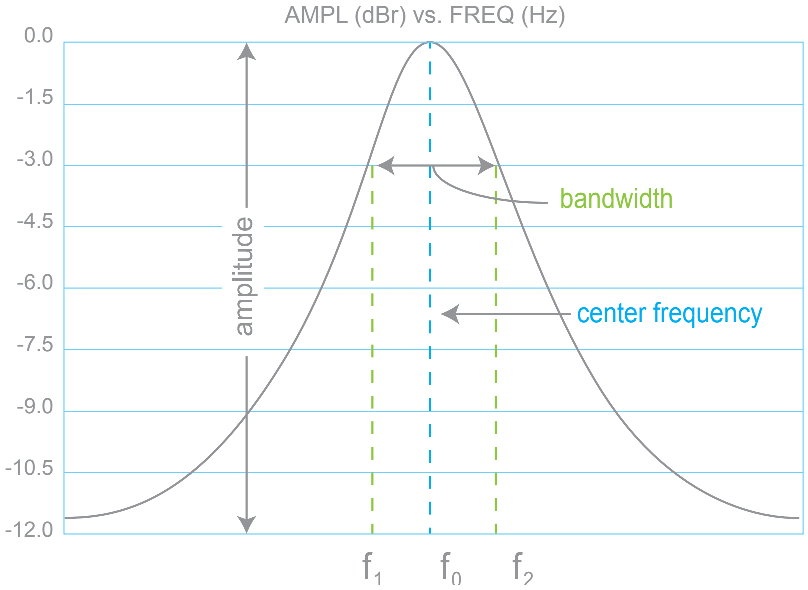

5.13: The effect of bandpass filter on multiple frequencies

The bandpass duplexer

It is the combination of two or more of these bandpass cavities—interconnected in a duplex configuration—that makes a *bandpass duplexer* work. One or more of the cavities are placed in the transmitter part of the duplexer, tuned to allow only the narrow band of transmitting frequencies to pass freely. Similarly, those cavities in the receiving part of the duplexer are tuned to the narrow band of receiving frequencies (figure 5.14).

The transmitter's output signal is filtered through the transmitter bandpass cavities of the duplexer on its way to the antenna, with the desired frequencies experiencing very little loss. At the same time, undesired frequencies are attenuated significantly, reducing the transmitter noise that could otherwise interfere with the signal.

As an added bonus, this reduced transmitter noise not only improves the signal to our own receivers, but those operating on different frequencies as well. By limiting errant frequencies, we reduce the likelihood of receiver desensitization for other users on entirely different channels.

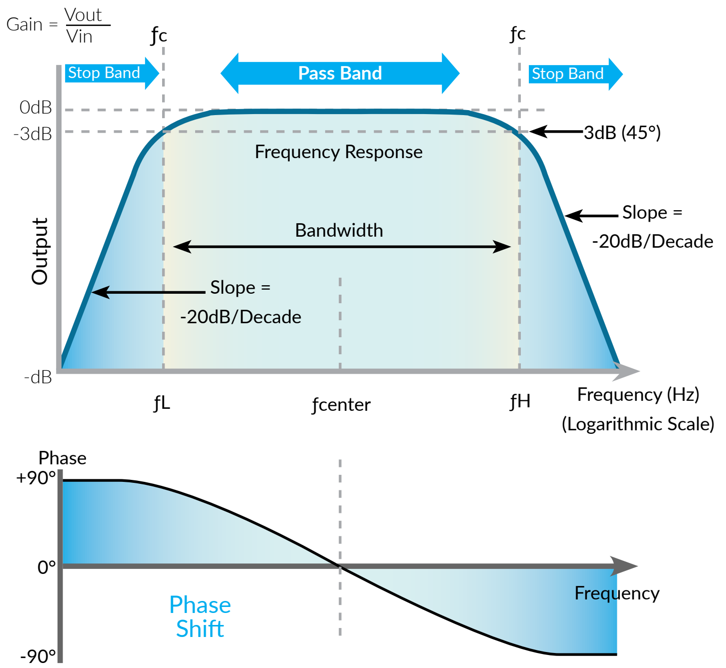

5.14: The effect of bandpass filter on multiple frequencies

Bandpass duplexer
A duplexer that uses multiple bandpass cavities to separate transmitter and receiver signals, allowing for simultaneous two-way communications.

Similarly, the bandpass cavities on the receiver part of the duplexer (again, usually two or more cavities in a duplex configuration) are resonant to receive only assigned frequencies. As with the transmitter's bandpass cavities, there is a modest loss of power in the process of receiving, but unwanted frequencies are attenuated to negligible levels.

The net effect is that off-frequency signals are virtually invisible to the receiver, protecting it from desensitization—not only from its own corresponding transmitter, but from others operating on completely different frequencies.

Isolating the best solution

Modern two-way communications networks must contend with the interfering effects of both receiver desensitization and transmitter noise. While a two-antenna solution is one way to address these factors, most practical applications must contend with space, cost and antenna availability limits. In most cases, a bandpass duplexer provides the requisite isolation between transmitter and receiver, even when operating on the same antenna.

With the isolating properties afforded by a bandpass duplexer, both transmitter and receiver can operate efficiently while reducing transmitter noise and receiver desensitization. The result is a compact, efficient and reliable communications network that easily accommodates two-way communication of voice and data.

Behind every simple call or text on millions of mobile devices at any given moment is a world of complex science and technology at work—and now you have a better understanding of the important role that transmission and receiving isolation systems play in RF communications.

Chapter 5 summary

Duplex RF communications:

- Allow simultaneous two-way signal traffic
- Inherently vulnerable to interference and require isolation to work efficiently

Sources of interference:

- Receiver desensitization
- Transmitter noise

Isolation:

- Techniques that prevent both kinds of interference
- Measured in dB; the higher the dB loss, or attenuation, the clearer the signal

Antenna solutions:

- Polarization separation
- Horizontal separation
- Vertical separation
- Addition of duplexer

Duplexer choices:

- Bandpass duplexer
- Band-reject duplexer

Chapter 6

The many ways to connect:
RF transmission lines

Look around your home and office and you'll see wires, cords and cables everywhere.
In your office, network cables connect your computer to the outside world. In your living room, coaxial cables bring in premium programming and high-definition video cables feed it to your flat-screen TV. In any room of the house, you may have a reliable land-line telephone that can reach out to virtually any person on the planet, all through a slender phone line of twisted copper.

Indoors and out, these connections manage the flow of information that drives our daily lives. **CommScope** is dedicated to the continuous improvement of cable technologies that have an impact on every life, every day.

Different cables are made for an amazing variety of uses, but they all have one thing in common: *they transmit power and patterns from a transmitting source to a receiving destination*. In RF applications, these cables are the hard links that connect antennas to base stations.
They are collectively known as transmission lines.

Transmission line
In RF applications, the physical medium that conducts RF power from one point to another, usually between a base station and an antenna.

Long ago, *transmission lines* were primarily used for the movement of electrical energy. Multi-conductor transmission lines could efficiently connect a power source, like a generator or battery, to a device that would consume that energy. This kind of configuration is common even today. You create a small-scale version of it every time you use an extension cord to connect an outlet to an appliance.

As telephone technology emerged, the limitations of this technique soon became apparent. When passing multiple circuits along a single transmission line, the signals proved highly vulnerable to external interference. To address this problem, Bell Telephone Laboratories developed a new type of cable in the 1930s. It was a shielded cable consisting of an inner wire surrounded by non-conductive material called a dielectric. This nonconductive material was then surrounded by an outer, sleeve-shaped conductor, and the whole assembly was finally encased in an insulating cover. If this design sounds familiar, it's because you've seen it before. This was the first coaxial cable, essentially the same design used today for data transmission (figure 6.1).

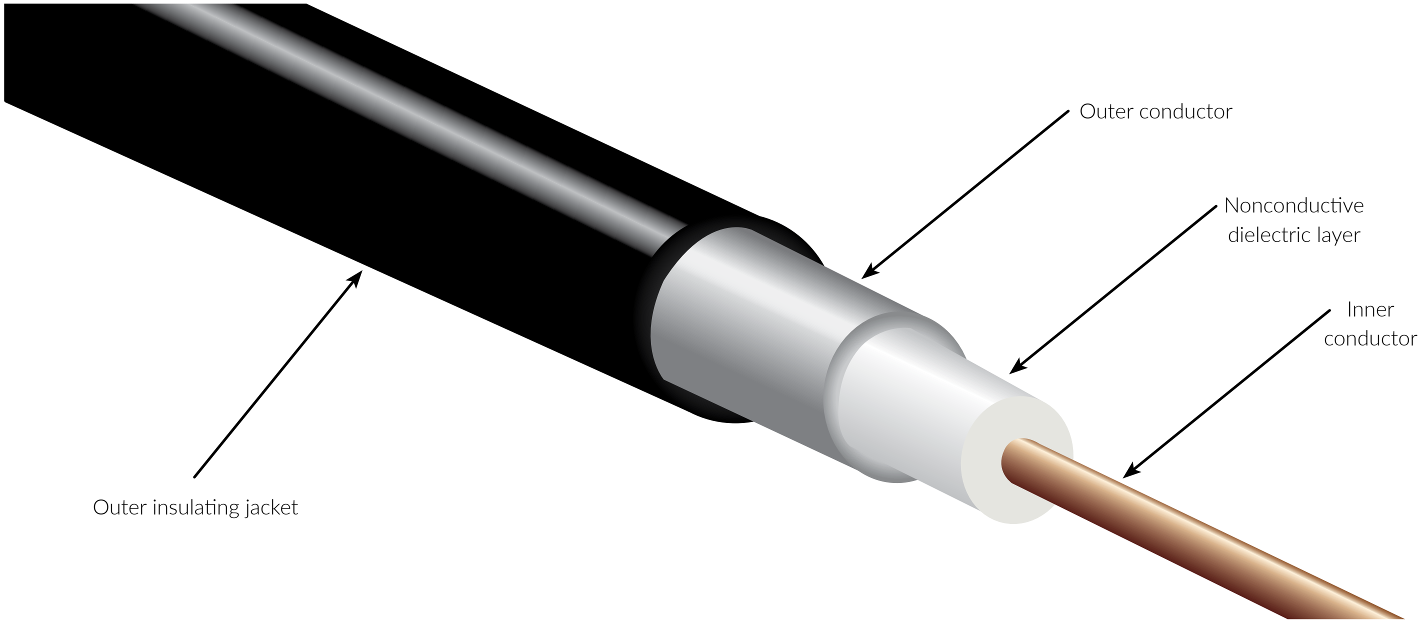

6.1: The basic construction of a coaxial cable

In RF applications, *coaxial cable* is used as a transmission line for radio frequencies that only penetrate the outer layer of a solid conductor, a transfer known as the "skin effect." The benefit of this arrangement is that it allows the outer surface of the outer conductor to be grounded.

Signals pass along a coaxial cable by riding the outer surface of the interior conductor and the inner surface of the outer conductor with a nonconductive dielectric layer between them. As a result, the only escape points for the energy carried on the line are at either end—exactly where they're needed for clear transmission.

Coaxial cable
A transmission line built to prevent interference while carrying multiple signals. It consists of an inner core conductor and an outer sleeve conductor, separated by a nonconductive dielectric layer.

Coaxial cable types

Modern coaxial cables used in RF transmission can be grouped into three main categories: solid dielectric, air dielectric and foam dielectric. The construction of each category makes each of them suited to particular uses.

Advantages:	**Disadvantages:**
Flexible Easy to install Inexpensive No pressurization required	High signal loss Prone to deterioration RF signal leakage through outer conductor
Low signal loss High power and frequency capacity Long operational life	High initial costs Pressurization logistics Vulnerability to moisture
Reduced power loss No pressurization required Moderately priced Long operational life Enhanced crush resistance	Slightly more loss than air dielectric More expensive than solid dielectric

Solid dielectric cables employ a flexible inner conductor (stranded or woven, as opposed to a solid wire), covered by solid extruded polyethylene insulation. The outer conductor is braided, and multiple layers can be stacked with shielding foil between them. The outer insulation is a polyethylene jacket.

Air dielectric cables are similar to the solid variety except that they employ open space as the inner nonconductive layer. This cavity is supported with small insulating spacers that maintain the open channel and are pressurized to keep out moisture.

Foam dielectric cables employ a solid, as opposed to stranded, copper wire core. The outer conductor is generally smooth aluminum, corrugated aluminum or corrugated copper. The inner nonconductive layer is made of polymer foam, which combines several key advantages from both solid and air dielectric cable varieties. Its power loss and cost characteristics lay between the two other options, but foam also offers practical advantages that make it the preferred choice in many modern two-way RF applications.

The mechanical elements of coaxial cable

Several material choices are available for both the conductive and nonconductive elements of coaxial cable. The specific needs of a particular use determine which combination is most efficient and affordable. Below, you can see two examples of the many varieties available (figure 6.2). Note how they feature different conductive components, but both have the same solid dielectric insulator and outer jacket.

Signal energy is carried along the inner and outer conductor. You will notice that, in both cases, the surface area of the outer conductor is much greater than that of the inner conductor. Therefore, the conductive properties of the inner conductor must be as efficient as possible. That's why highly conductive copper is almost universally preferred.

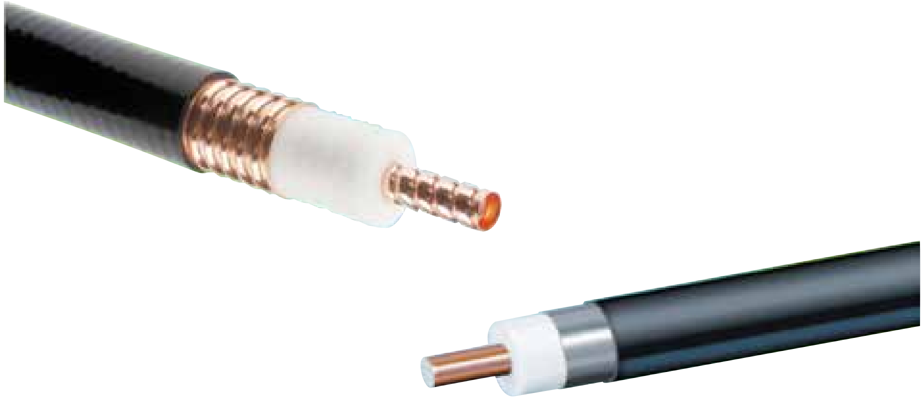

6.2: Corrugated copper (left) and smooth-wall aluminum (right) coaxial cables

Braided copper is the most commonly used outer conductor on solid dielectric cables. Copper is again chosen for its exceptional conductivity, and it is used in a braided form to improve its flexibility. Solid copper or aluminum material, either corrugated or smooth-walled, is most often used for foam or air dielectric cables. The choice between aluminum and copper often comes down to cost. While aluminum is less expensive than copper, it also has lower conductivity.

In RF transmission lines, the preferred dielectric material is polyethylene due to its low loss characteristics and long life span. This material can be used in either solid or foam dielectric constructions, or as the spacers in an air dielectric design. For high-power applications, Teflon is substituted for its high melting temperature due to the higher operational temperatures. Teflon is more expensive, so there are various other materials with costs and temperature resistances between those of polyethylene and Teflon.

Electrical properties of coaxial cable

Signal loss, or *attenuation*, is a significant consideration in the design of a cable. The loss occurs in three ways:

1. **Conductor loss** is direct function of the conductive properties of the cable's materials.
2. **RF leakage** is a measure of the effectiveness of a cable's shielding.
3. **Insulation** loss is a fixed degree of attenuation inherent in the material of the cable's dielectric layer.

Attenuation is measured in decibels; in the specific case of transmission lines, attenuation is expressed in either decibels per 100 feet of cable length or decibels per 100 meters of cable length.

How well these losses are managed depends on such factors as the size and length of the cable, the conductivity of the materials used in the cable, the frequencies traveling along the cable and the effectiveness of its shielding. There are general physical rules governing how these factors impact attenuation, such as:

Cable size. As a rule, a cable's conductor loss will decrease as its size increases. This is due to a larger cable's broader cross-section and its corresponding increase in conductive area.

Cable design. Solid outer conductors allow less RF leakage than braided ones, though at the expense of flexibility.

Dielectric material choice. By choosing any particular dielectric material, you can anticipate a predictable level of insulation loss. As explained earlier, air dielectric offers the lowest insulation loss, while solid dielectric comes with the highest loss.

Assigned frequency. All three types of attenuation directly increase as a function of the frequency of the cable's signal. The higher the frequency and the shorter the wavelength, the greater the loss in any given cable.

This complex balancing act of performance, ease-of-handling and cost means no single transmission line design is ideal for all, or even most, circumstances. Each design is an exercise in compromise between these factors.

Attenuation
Measured in decibels (dB), this is the loss of power experienced by an RF signal as it moves from one point to another. Transmission line attenuation is expressed in either decibels per 100 feet (dB/100 feet) of cable length, or decibels per 100 meters (dB/100m) of cable length.

Ohm
The unit of electrical resistance. In terms of RF transmission lines, ohms refer to the inherent, or characteristic, loss over a length of cable.

Characteristic impedance

Characteristic impedance, commonly called cable impedance, is a measurement of the electrical resistance of an RF transmission line as measured in *ohms*. The figure is derived by a complex formula involving the ratio between the cable's two conductors. As a general rule the industry standard impedance for RF cable is usually 50 ohms (though some applications require 75 ohms).

This expected degree of impedance can be affected by imperfections or damage in the cable itself. A deep dent in the outer wall of a coaxial cable can cause its impedance to vary from its standard level. This disruption is called a discontinuity, or a change in the distance between the inner and outer conductors, as you might see from a squashed cable. The signal refelects within the cable, creating the same loss of performance as a mismatch between cable and antenna (chapter 3).

This is one reason that a cable's flexibility and crush resistance are such crucial factors. Damage during installation is a frequent source of discontinuity and can be expensive and time-consuming to remedy.

Velocity of propagation

Simply stated, the velocity of propagation within a coaxial cable is the speed at which a signal can travel along that cable. Velocity is governed by the amount and type of dielectric used, and is expressed as a percentage of the speed of light. It can range from 67 percent for solid dielectric cables up to 92 percent for air dielectric cables. However, since the speed of light is more than 670 million miles per hour, velocity is rarely a concern in itself, though there are exceptions. For example, velocity becomes significant in cases where phasing is required (chapter 3).

Power handling capability

The amount of power that a particular transmission line can handle is dictated by two measurements of temperature: the ambient temperature (the temperature of the air surrounding the cable) and the temperature of the cable itself under ordinary operation.

As we've seen, power loss is inherent in any cable design, and is dependent upon the kind of dielectric used in the line. This lost RF power translates to heat, so the greater the attenuation within a cable, the more heat it will generate from that lost energy. Likewise, the greater the frequency passing along any given cable, the more heat it will generate as a function of loss.

Heat resistance is a critical factor in cable design. For instance, foam dielectrics begin to soften near 180 degrees Fahrenheit, so an engineer choosing the right transmission line will need to be certain that the combined internal and ambient temperatures won't exceed 180 degrees. If the cable exceeds its limit, the softened dielectric will allow the inner conductor to shift, creating a discontinuity. If it should contact the outer conductor, the result would be a shorted cable. To help engineers make the right choice and prevent such failures, cable manufacturers like **CommScope** rate each type of cable for certain power levels at certain ambient temperatures.

RF leakage

As its name suggests, RF leakage is a function of the physical ways an RF signal can "leak" out of a transmission line. In the case of a cable with a braided outer conductor, there are countless tiny openings in the cable. As with a leaky garden hose, the more power or pressure you apply, the more significantly those leaks affect performance.

In addition to attenuation, RF leakage causes another challenge when several high-power braided coaxial cables are arrayed in close proximity to one another. The leakage can cause interference between the cables at their endpoints, such as when they connect to antennas, multi-couplers, duplexers and so forth. Faulty connections can make this problem worse.

As with power handling, RF leakage is included in a cable's specifications to help engineers choose the best option for any particular configuration, especially in circumstances where many cables terminate close together.

Environmental factors

The climate and setting of a cable installation dictates what kind of cable should be used. Considerations include:

- Sunlight UV exposure
- Humidity and moisture
- Temperature extremes

Cable life expectancy

The expected useful life span of a coaxial transmission line depends largely upon *environmental factors*. Since engineers have little control over these factors, they must compensate by choosing the right design with the right materials to assure the longest possible life span.

The composition of the cable's outer jacket is one of the more obvious considerations. Most flexible and semi-flexible coaxial line jackets use polyethylene, polypropylene, or polyvinyl chloride (PVC). All three options are vulnerable to long-term sun exposure, so manufacturers incorporate carbon black into the resin to improve the jacket's resistance to aging under ultraviolet light, which can extend their operational lives up to 20 years of service. This improvement is why so much coaxial cable you see in household use is black.

Moisture and humidity are important factors as well. While water can infiltrate through tiny nicks, cuts or age cracks in the cable, the single most common form of moisture infiltration is through improperly sealed connectors on the ends of the cable. Even humid air present inside the connector can condense as temperatures fall, resulting in liquid water that wicks deeper into the cable along the outer conductor's braid. This can potentially corrupt the entire cable and short the inner and outer conductors, particularly in the connectors themselves. The result is increased signal reflections within the cable and degraded passive modulation performance.

Connectors

As the number of modern RF applications has grown, the technology used to connect a cable to its terminus has evolved. The simple designs created in the 1940s for military uses have diversified and improved into a variety of types such as these (figures 6.3 through 6.8).

UHF connectors are the oldest and most popular type still in use for two-way communications. They are rugged, reliable and easy to install, which is why they are the preferred choice for applications with frequencies up to 300 MHz.

BNC connectors are small, quick-disconnect versions with a bayonet-style locking coupling. These are often used on narrow cables connecting equipment.

TNC connectors are similar to BNC connectors, but include threaded connections that keep them secure in environments where vibration is a concern.

Type-N connectors are an industry favorite for RF communications with frequencies above 300 MHz, where UHF connectors are not suitable. Type-N connectors may be rated to perform at 10 GHz or even higher.

EIA flanges are used primarily on pressurized air dielectric cables operating above 450 MHz. These connectors offer the standard 50 ohms of impedance and typically offer higher voltage characteristics than Type-N connectors.

DIN (Deutsche Industrie Normenausschuss) connectors are available in several sizes and have come to dominate the RF communication industry as a whole. They have a larger cross-section than Type-N connectors, and better withstand the rigors of field installation.

For exceptionally congested installations with multiple service carriers operating in the 700, 800, 900 and 1900 MHz bands, silver-plated Type-N or DIN connectors are required.

6.3: DIN coax connector

6.4: BNC coax connector

6.5: TNC coax connector

6.6: Type N connector

6.7: EIA flange coax connector

6.8: UHF coax connector

Voltage standing wave ratio (VSWR)
A key measurement of cable performance and signal quality. It quantifies the amount of signal reflected backward along a cable to its source. Theoretically, perfect operation yields a VSWR value of 1.0, or "unity," meaning zero reflections.

Installation step 1: cable choice

As discussed above, choosing an appropriate cable depends on knowing:

- Which frequencies it will carry
- How much loss is tolerable
- Where it will be installed
- What kind of budget limits exist

In most cases, there will be more than one acceptable cable solution for any one of these four criteria, but the key is to get the greatest possible benefit against all four. For instance, if we look at cost as the primary driver, we see that smaller-diameter cables cost less to purchase and install, but will need more upkeep and eventual replacement. We may also consider a less expensive option with a higher rate of loss, intending to compensate for that loss through greater RF power generation on one end of the cable or increased antenna gain on the other.

Installation step 2: field testing

Once you have selected and installed the cable that best performs to your application's priorities, the process of fine-tuning that performance can begin. There are three tests you would likely perform: inner and outer conductor continuity, shorts between conductors, and a *voltage standing wave ratio (VSWR)* test.

The first two tests are simple and direct measurements of impedance in the cable, performed with an ohmmeter. Any physical disruption of the cable's integrity would be revealed in a non-standard level of impedance, and you could begin inspecting the cable for the source of the problem.

The third option, the VSWR test, is an indirect but ultimately more revealing measurement of overall line performance. Basically, the VSWR test measures the amount of signal reflection taking place within the cable. Measuring both forward and reflected power with a wattmeter, you can compare the values against the cable manufacturer's conversion chart for that particular type of cable. If everything is functioning correctly, the observed amount of reflected energy should fall within expected limits. Poorly made connections or connections with mismatched impedance will quickly become obvious. This process is explained in more detail early in chapter three.

Installation step 3: troubleshooting

Reduced RF communication performance can be rooted in any number of problems, occurring in any component in the system. When the antenna and transmitter have been ruled out as potential trouble spots, it's time to examine the transmission lines because a lot of things can go wrong with cables.

Here are just a few things that can cause system performance to drop suddenly:

Weather. A good place to start is to visit the site itself and speak to those familiar with recent weather trends. A bad storm, lightning, hail or high winds can damage cables and loosen connectors.

Local phenomena. In addition to weather, other local events can impact performance. Explosions from nearby mining operations, small earthquakes, even a stray bullet from a hunter's gun have been identified as culprits.

Water infiltration. As discussed earlier, water is perhaps a cable's greatest enemy. Checking connectors for signs of moisture, double-checking their seals, and examining the cable itself for any new damage will help confirm or rule out water as a cause.

In any event, once the damage has been identified, that section of cable cannot be taped or otherwise repaired. It must be replaced. Long-term system performance degradation can be just as serious, and is often caused by cable aging. While metal-sheath cables are almost impervious to aging when properly installed, inferior cables can age and crack with extended exposure to the sun's UV rays and extreme temperatures.

Localizing the problem

If your VSWR measurements reveal a high level of reflection—say, 20 percent above the level indicated by the manufacturer's table—then, most likely, your cable is experiencing an open, a short or a partial short somewhere along its length or in a connector. To confirm this, you could perform the following tests:

1. Open the top of the cable and short the inner and outer conductors (the cable ground should be removed for this). Measure impedance between the conductors with an ohmmeter. An intact cable will show low impedance between the two, while high impedance will reveal damage to the outer conductor. This kind of damage is hard to locate. If economically feasible, replacement may be the best option.

2. Remove the short between the conductors and test impedance again. In this instance, an intact cable will show high impedance, while low impedance may indicate damage to the inner conductor that creates a short somewhere within the cable. The damage required to cause this kind of fault often leaves more obvious traces on the outer jacket and is easier to identify.

3. Examine the connectors themselves. Type-N connectors are particularly vulnerable to misalignment and pin breakage, which can result in a short. Also, as the primary source for any potential water infiltration, it's a good idea to examine any type of connector for signs of moisture. For best results, check connectors during cool weather or at night, where any trapped vapor will have condensed into more visible droplets.

4. Practice good preventative maintenance. Proper installations reduce the need for ongoing maintenance, but vigilance is always to your benefit. Any time an installation is realigned or painted, it's smart to inspect the cables and connectors. Identifying small problems before they become big problems can save a great deal of time and money and minimize lost performance.

In summary, a solid understanding of the construction of cables helps you understand their best applications, where they may be vulnerable, and where to look when a fault is suspected.

Transmission lines connect the world

While so much of modern RF communication is comprised of radio energy radiated through the air, the critical links on either end depend on the right kind of transmission line cable and the right connectors between base station and antenna.

As a physical link, these cables must be able to flex where they're needed, withstand the punishing elements, and faithfully carry the frequencies that eventually reach you as your internet connection, land-line call, or mobile phone call virtually anywhere in the world.

You live among small-scale examples of the same technology at home. From the USB cord on your computer mouse to the century-old design of your telephone's wall cord, each cable is designed to carry specific frequencies over specific lengths, each one for a unique purpose. As a leading provider of coaxial and other transmission line products for networks all over the world, **CommScope** is at the forefront of the race to develop innovative solutions that will address the future of technology tomorrow.

Chapter 6 summary

RF transmission lines:

- The bridge between base station and antenna
- Adapted from designs once used to carry simple electricity

Coaxial cable:

- Characterized by insulated inner and outer conductors
- Solid dielectric
- Air dielectric
- Foam dielectric

Mechanics and materials:

- Copper used for inner conductor
- Copper or aluminum used for outer conductor
- Polyethylene used for dielectric insulation
- Various connectors for different applications

Design considerations:

- Cable impedance
- Velocity of propagation
- Power-handling capability
- Heat and frequency limits

Measuring performance:

- Conductor loss
- RF leakage
- Insulation loss
- VSWR testing and fault detection

Chapter 7

Making every connection count:
Passive intermodulation (PIM) fundamentals

It's a fact of life for electronic devices: when something isn't working correctly, we check the connections. From the most complex cellular transmission system to the simplest toaster, junctions between cables and components are the most likely place for problems to occur.

But beyond the obvious culprits like poor connections, water infiltration or mechanical issues, a communications system's connections can also play host to other problems. One example is *passive intermodulation* (PIM), an inherent product of the system's frequencies and their associated harmonics. PIM can create undesirable sideband frequencies that interfere with the system's assigned frequencies. It takes a knowledgeable partner like **CommScope** to choose the right components and assist in the design and installation of a communications system.

Passive intermodulation (PIM)
A potential side effect of having more than one high-powered signal operating on a passive device such as a cable or antenna. PIM occurs at non-linear points in a system such as junctions, connections or interfaces between dissimilar metal conductors, creating interfering frequencies that can decrease efficiency. The higher the signal amplitude, or power, the greater the effect.

Intermodulation explained

The growing demand for wireless services has increased the complexity of system design and the resources to support that demand. As a result, there are more RF components, configurations and spectrums utilized in the RF path. Included in these additional components are passive devices that can contribute to PIM. With a good understanding of these PIM contributors, we can proactively address PIM and its impact on the system noise floor.

Because PIM will act as a component of the overall noise floor, we will first discuss the components of system noise and how noise impacts network performance. The system uplink noise floor is dynamic and affected by PIM and other factors. It is also driven by the number of active users. For every new user added to the site, additional noise is added to the network, causing a "noise rise". As new users raise the noise floor, each user must deliver a higher power transmit signal to overcome the increased noise. This means the user must be closer to the receiving cell tower. Consequently, the cell coverage area is reduced.

There is a theoretical limit to "noise rise" and a consequent corresponding limit to the number of users that can be added. This theoretical maximum number of users is referred to as the "pole capacity" of the network.

Uplink pole capacity is proportional to the system signal-to-interference/noise ratio (SIR) and several other variables and, for WCDMA networks, can be estimated using the following widely accepted formula:

$$Np=(W / R)/((1+ f)*AF*10^{(EbNo/10)})$$

where

Np = Pole capacity

W = Spreading bandwidth

R = Users' radio bit-rate

f = Ratio of other-cell interference to in-cell interference

AF = Activity factor

Eb/No = Ratio of energy per information bit to power spectral density of noise + interference

Notes:

1) Basically, Eb/No represents the amount of gain that must be provided above noise and interference for acceptable system performance

2) Eb/No can be expressed as Eb/No = Processing Gain + SIR, where SIR is the Signal-to-Interference/Noise ratio

3) For 3G voice at a CS12.2K data rate, the Eb/No requirement is typically 5 dB

An example pole capacity calculation estimating the uplink capacity for a WCDMA cell site is shown as follows:

Np = Pole capacity

W = 3.84 Mhz (value for 3G UMTS)

R = 12,200 bits/second (value for CS12.2K for 3G UMTS voice)

f = 65% (typical network value)

AF = 50% (typical network value)

Eb/No = 5 dB (typical value for 3G UMTS)

Np = (3840000 Mhz / 12200 bit/s) / ((1+0.65) * 0.50 * 10^(5 / 10))

Np = 20.6 users maximum

We can see from the above pole capacity equations that increases in the system noise floor can degrade system performance, increase system power consumption and reduce cell coverage area. Because PIM, when not mitigated, can be a significant contributor to noise, it is clear now the serious impact it can have on wireless network performance.

Now that we have developed an understanding of the system level impact of PIM, we will describe specifically how PIM is generated and how it can be mitigated.

Generally speaking, intermodulation is the result of two or more frequencies (often, a duplexed system's receiving and transmitting frequencies) interacting with one another according to certain mathematical relationships related to their specific frequencies. The effect creates errant signals that contribute to noise and interfere with the system's operation.

Passive intermodulation is a particular kind of intermodulation that takes place in the passive parts of a system—such as cables and antennas—often at connections that create nonlinearity in the system.

As the complexity of communications systems has increased, so has the potential for PIM, making it a top priority for service providers to manage its effects. In a cellular base station, for example, a transmitting frequency can create PIM interference in its own receiving frequency, or vice versa. Where PIM occurs depends on the separation of the two main frequencies, as shown in the chart below (figure 7.1).

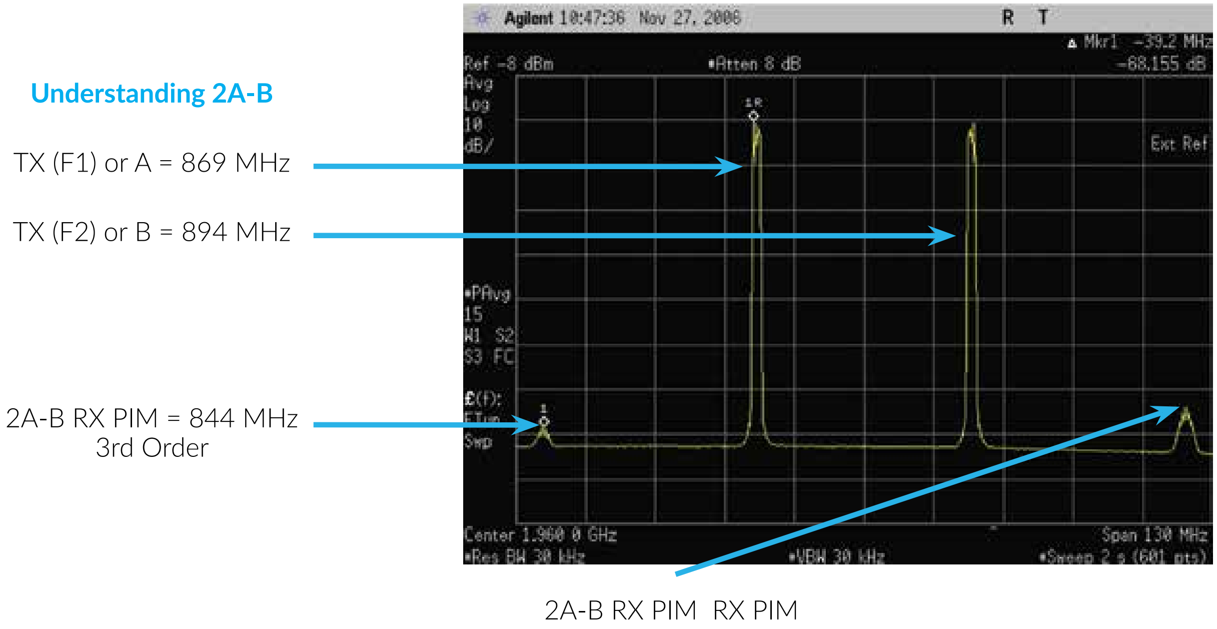

7.1: Calculating where PIM will occur, based on two example frequencies, A and B; in this example, PIM occurs at 2A-B (a common 3rd order product) and again at 2B-A

Calculating potential PIM

In this example, transmission frequency A (869 MHz) and transmission frequency B (894 MHz) create potential PIM problems at 2A-B (844 MHz) and at 2B-A (919 MHz). This means we must expect troublesome interference if our reception frequencies include 844 MHz or 919 Mhz.

Nonlinearity

A location within an electrical circuit where voltage does not remain consistently proportional to power. This effect is caused by imperfect connections between components and cables.

Managing PIM

Since we don't always have the freedom to select frequencies that avoid PIM issues, we need to look at other methods of reducing its influence in our communications system. Reducing PIM levels starts with reducing nonlinearity in the circuit.

Nonlinearity in a passive RF circuit typically results from current rectification at the conductor joints and mechanical junctions. Resolving nonlinearity generally means improving connections throughout the RF path of the system. This means addressing problems such as:

- Improper connector attachment
- Poorly torqued connections with incorrect contact pressure
- Contamination or corrosion of conducting surfaces
- Inadequate plating on rust-prone ferromagnetic components
- Poor connections due to cold solder joints

The most common and visible contributors to high PIM levels in the system are associated with the mechanical and physical integrity of the connections in the RF path. Therefore, those who deal directly with the RF components in the field, such as installers, service technicians and test engineers are integral in the battle against PIM. They will require training in proper field installation, proper use of PIM test equipment and a fundamental understanding of PIM.

In order to manage PIM, we also need to understand the environment in which we are performing the PIM testing. For a manufacturer, it is now common for PIM to be tested in a controlled RF environment such as an anechoic chamber. This is done to ensure the device under test has an isolated and "quiet" RF environment to confirm the PIM performance as it relates to the published manufacturer specification.

Field testing presents an entirely different set of challenges if PIM performance testing is required. This is not an isolated or "quiet" RF environment. This noisy RF environment can negatively impact PIM or provide false PIM levels that are a direct result of the environment. In other words, you may obtain PIM levels in your testing that are not reflective of the actual device under test, but rather the RF environment in which it is tested. Shown on the next page are some examples of base station antenna testing where the environment impacted the PIM measured in the field.

Although conditions such as external RF signals or the test environment may not be controlled, the selected components, installation practices and testing are within our control. Education and awareness are the first steps in managing the difficulties associated with field PIM testing.

System PIM calculators

Reducing PIM in the RF path is important and challenging. CommScope provides several PIM system calculators to support your RF path analysis.

Download the latest tools below.

PIM

PIM and VSWR System Calculator (Global)

Learn more

PIM

Band and Block Second and Third Order PIM Calculator (North America)

Learn more

PIM

Band and Block Second and Third Order PIM Calculator (EMEA)

Learn more

Clear sky
-123 dBm (-145 dBc)

Towards forklift
-84 dBm (-127 dBc)

Person nearby with phone, keys, adapters, and badge
-94 dBm (-137 dBc)

Near shelter
-102 dBm (-145 dBc)

Pointed at fence
-102 dBm (-145 dBc)

Near cabinet and test equipment
-96 dBm (-139 dBc)

7.2: PIM measurements—field observations

- *On-site antenna measurements with iQA200 portable PIM*
- *Clear sky RF field-of-view required to avoid any secondary PIM sources which could cause false PIM failure*

Researching, testing and installing

The severity of PIM is directly related to signal amplitude, or power. That's why PIM performance of different RF components specify power levels along with their other properties. Typically, this figure is calculated by applying two signals of close frequencies at 20 watts (+43 dBm), the industry standard for PIM testing. This standard helps assure that comparisons between different products yield meaningful answers.

Understanding the PIM properties of components like cables, connectors, combiners, filters, tower-mounted amplifiers and antennas allows you to design a system with minimal exposure to potential PIM issues. It's a meticulous process, but it's essential to preventing potentially crippling PIM problems later on:

- Choose a knowledgeable provider that has demonstrated experience in the PIM specification of their products—one who can help you make the right component choices. **CommScope** makes this expertise freely available to our customers.
- Test your component performance against PIM specifications to assure trouble-free operation later on.
- Use trained installers certified in preparation and installation techniques. Since they will be managing connections, and connections are the source of PIM, their skill is your best guarantee against problems.

By following these recommendations, you can count on an RF system that will operate efficiently, virtually free of troublesome PIM effects.

PIM: best addressed at the planning stage

In engineering, connections are perennial trouble spots. Each connection, junction and interface is an opportunity for something to go wrong, including passive intermodulation (PIM).

The only way to design the system to mitigate PIM is to study and test the PIM specifications of the components comprising that system, which is why partnership with **CommScope** is so vital. We provide the experience and insight to spot potential trouble early in the process. Visit the PIM-dedicated portal of our website, to learn how we can assist you with awareness, prevention, identification, resolution and support.

Chapter 7 summary

Passive intermodulation (PIM):

- Caused by interactions between multiple frequencies and their harmonics
- Can cause transmission or receiver frequencies to interfere with each other
- Occurs as a byproduct of nonlinearity in a circuit, such as at a connection
- A critical consideration in the selection of components and installation of those components in an RF system

Chapter 8

The infrastructure behind the connection: Microwave backhaul

Imagine two kids, each standing in their own backyards, talking to each other on soup cans connected by a string. This is the simplest of connections, nothing more than two users on a direct, point-to-point, dedicated line. This same simplicity applies if we replace the cans with walkie-talkies—the communications system is still reliant on just two points of contact.

Once the system becomes more complex, like modern distributed communications networks with millions of users, these simple connections are no longer practical or possible. A centralized processing point is needed to route many conversations simultaneously just to make the correct phone ring when you dial its number.

This centralized processing is called *backhaul* and it's the infrastructure behind the connection where you'll find many **CommScope** solutions.

Backhaul
The process of connecting two ends of a transmission through a central routing point.

To imagine backhaul in action, consider the classic image of an early 20th-century telephone operator, manually connecting calls at a switchboard. As more and more people call at the same time, the task soon grows very complex—eventually, too complex for humans to perform.

Today's modern cell phone networks require much more complex connectivity—collectively called backhaul—than even the fastest human operator could provide. The process of routing network traffic for a cell phone call requires many steps to complete, and looks something like this:

1. Mary makes a call to John on her cell phone from her office
2. Mary's outbound call is picked up by the nearest cell tower
3. The tower routes Mary's call to the area's regional network
4. The regional network sends Mary's call to the national network
5. The national network routes Mary's call to Bill's regional network
6. The regional network broadcasts Mary's call from the nearest cell tower
7. Bill's cell phone rings, he answers and the call connects

Backhaul is the process of routing Mary's cell call—and all network traffic—up to, then down from the core processing backbone between the cell tower nearest to Mary and the one nearest to Bill. While backhaul can be achieved over a number of different materials, such as twisted-pair copper cable, fiber optics or coaxial cable, microwave backhaul offers a time- and cost-efficient backhaul link, ideally suited to cell phone networks that move calls and data across the country and across the world millions of times every day.

Microwaves and the electromagnetic spectrum

The electromagnetic spectrum includes an incredible variety of radiation types, all expressed in Hertz (Hz), a measurement of a particular radiation's frequency. Most frequencies used in electronics are expressed in thousands of Hertz, or kilohertz (KHz); millions of Hertz, or megahertz (MHz); billions of Hertz, gigahertz (GHz) or even trillions of Hertz, terahertz (THz). Some of the more familiar types are listed below (table 8.1).

Frequency	Wavelength	Application
50–60Hz	6000–5000km	AC electricity transmission
3–30kHz	100–10km	Sub-marine communication
30–300kHz	10–1km	Long-wave radio broadcast
180–1600kHz	1.7km–188m	AM radio broadcast
1.8–30MHz	167–10m	Shortwave radio
88–108MHz	3.4–2.7m	FM broadcast
300–3000MHz	1–0.1m	UHF point-to-point
700–2700MHz	0.43–0.11m	Mobile base station
0.3–300GHz	1–0.001m	Microwave/Millimeter Backhaul
352, 230, 193THz	1550, 1300, 850nm	Fiber-optic links
420–750THz	714–400nm	Visible light

8.1: Applications associated with different radiation wavelengths within the electromagnetic spectrum

Microwave backhaul frequencies

The International Telecommunications Union (ITU) regulates the microwave spectrum available for backhaul applications. Each band is best suited to particular situations of topography, climate, bandwidth requirements and cost, determined in large part by their typical hop lengths, or the distance each band can travel in open air between points (table 8.2).

You may notice that the table includes several "unlicensed" frequency bands. While most microwave frequencies are regulated to prevent interference during hops, unlicensed frequencies are unregulated, requiring cooperation between operators to prevent interference.

The highest bands, 60 GHz and 80 GHz, possess unique propagation characteristics, as shown in their very small maximum hop length. They serve a special purpose which we will discuss later in this chapter.

Frequency band	Frequencies (GHz)	Typical maximum hop length (km)	Typical minimum hop length (km)
0.9 (unlicensed)	0.902–0.928	100	–
2.4 (unlicensed)	2.4–2.5	100	–
4	3.6–4.2	70	24
5	4.4–5.0	60	16
5 (unlicensed)	5.3, 5.4 & 5.8	50	–
L6	5.925–6.425	50	16
U6	6.425–7.125	50	16
L7	7.1–7.75	50	10
U8	7.75–8.5	50	10
10	10–10.7	20	10
11	10.7–11.7	20	10
13	12.7–13.25	20	6
15	14.4–15.35	20	6
18	17.7–19.7	20	2
23	21.2–23.6	20	2
26	24.25–26.5	20	2
28	27.5–29.5	15	2
32	31.0–33.4	10	1.5
38	37.0–40.0	10	1
42	40.5–43.5	10	1
60 (unlicensed)	57.0–66.0	1	–
80	71–76/81–86/92–95	5	–

8.2: Typical hop lengths for microwave frequency bands

Duplex communications
A transmitter and receiver that work in different time slots or frequency slots on the same device.

Microwave advantages

The microwave band possesses characteristics ideal for radio transmissions, including those from cell phones:

- They help directional antennas make hops from point to point in narrow beams that don't interfere with each other
- The low end of the band (below 11 GHz) propagates over long distances, ideal for long-haul connections to users in remote locations
- The higher end of the band (above 11 GHz) propagates over shorter distances, ideal for short-haul connectivity required in urban locations

In addition to their technical characteristics, microwave links also offer practical and financial advantages over cable solutions, which have made them the preferred choice for modern communications networks:

- Less expensive to establish and operate
- Faster deployment
- Greater bandwidth than twisted-pair copper, critical to high-traffic applications like modern cell phone networks
- Readily scalable to handle more traffic as required
- Very reliable to operate

Microwave backhaul in action

Let's revisit the process of making a cell phone call. The user dials the call and the cell phone connects with the nearest cell tower, operating on the network's radio frequencies (within the 700–2700 MHz part of the spectrum). The cell tower hands off the call to a microwave transmitter (typically into one of the bands between 6–40 GHz) which directs it via a directional beam—a hop—to a collection or aggregation center which, in turn, makes the connection to the mobile network's core network. The traffic is then routed across the network to the region closest to the receiver's location, where it will be transmitted again by microwave to the nearest cell tower.

The receiving cell tower station down-converts the microwave signal back to the network's radio frequencies for its final journey to the target cell phone. The process is reversed for traffic moving in the opposite direction.

This two-way microwave backhaul generally uses a frequency-division duplex (FDD) system that allocates frequency channel pairs for simultaneous two-way, or *duplex communication* (figure 8.3).

Another way to achieve duplex communication is via a time-division duplex (TDD), which achieves the same goal by switching the required direction of transmission in a very fast, precise manner (figure 8.4). While a more efficient option, TDD requires very careful timing control and is not the preferred system for microwave use.

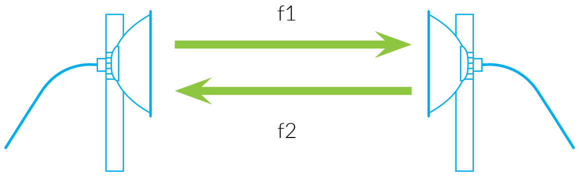

8.3: Frequency-division duplex (FDD) system with separate go/return frequencies

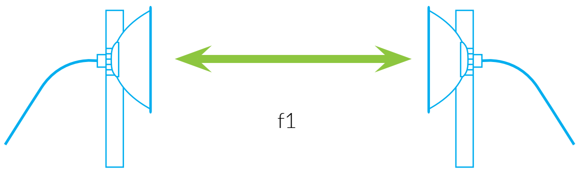

8.4: Time-division duplex (TDD) system using just one go/return frequency

Typical microwave deployments

Hop connections between microwave transmitters and receivers usually require *line-of-sight* (LOS) clearance. LOS means both ends can literally “see” one another without any obstructions in between—including the curvature of the Earth’s surface. For this reason, microwave antennas are often mounted on towers or at the top of high buildings for best LOS clearance.

To maximize the value of these choice locations and to reduce the costs for leasing these locations, microwave backhaul antennas are often mounted adjacent to base station antennas, which also rely on altitude for efficient operation (figure 8.5).

Line of sight (LOS)
The unobstructed space between transmitter and receiver. Longer hops must even account for the curve of the Earth as an obstruction.

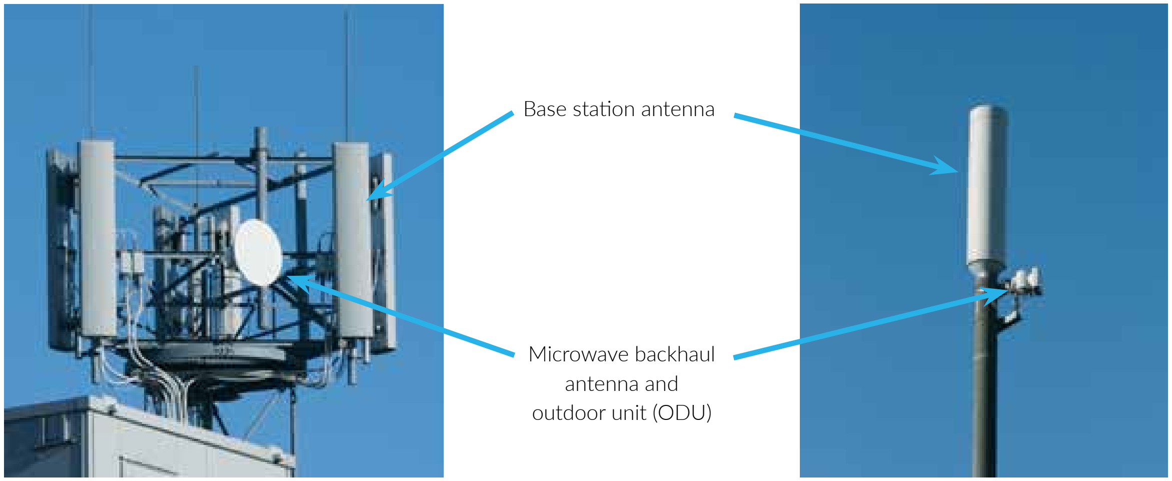

8.5: Typical microwave backhaul antenna integrated into a base station antenna location

Split-mount radio system
A two-stage connection that lets microwave radios located in an indoor unit (IDU) receive and transmit through an antenna fitted with an outdoor unit (ODU).

Many short-haul installations, typically operating above 11 GHz, use a *split-mount radio system*, which divides the radio into an outdoor unit (ODU) and an indoor unit (IDU). The ODU houses the microwave circuitry, including the go/return microwave signal separating diplexer and the up/down frequency converters. It is mounted in an enclosure adjacent to the antenna, or more frequently integrated into the antenna assembly itself (figure 8.6).

The IDU contains the modulator/demodulator, more commonly known as a modem, and the control circuitry necessary for translating the cell phone traffic into a form suitable for microwave transmission.

For high-density traffic and long-haul hops, multiple radios are typically housed in a remote radio room adjacent to the base of the tower. Generally, these hops use larger antennas operating at frequencies under 11 GHz (figure 8.7).

Connections between the antenna and the radios are made by coaxial cable, elliptical waveguide or circular waveguide transmission lines, depending on the frequencies involved (figure 8.8). Chapter 6 provides detailed information on transmission lines.

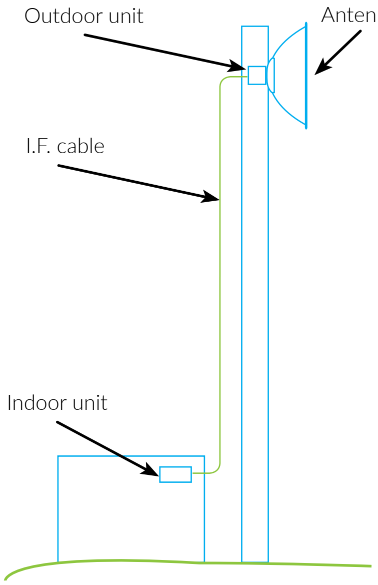

8.6: Typical split-mount microwave radio system showing IDU and ODU

8.7: A typical lower frequency aggregation point

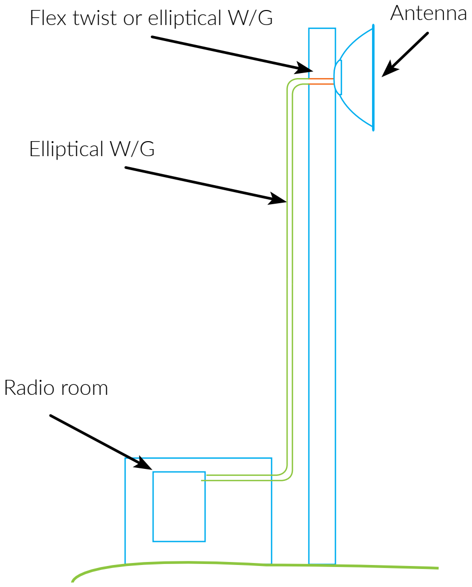

8.8: Connections between remote microwave radios and a microwave antenna

Planning a microwave link

When considering a new microwave backhaul path, it is important to consider potential interference issues during the design process. The planned link must not interfere with adjacent links or other operators in the area. To prevent conflicts and other problems, you must consider:

- Frequency coordination with other links in the vicinity
- Radio and radiation characteristics of the antenna
- Transmission power levels

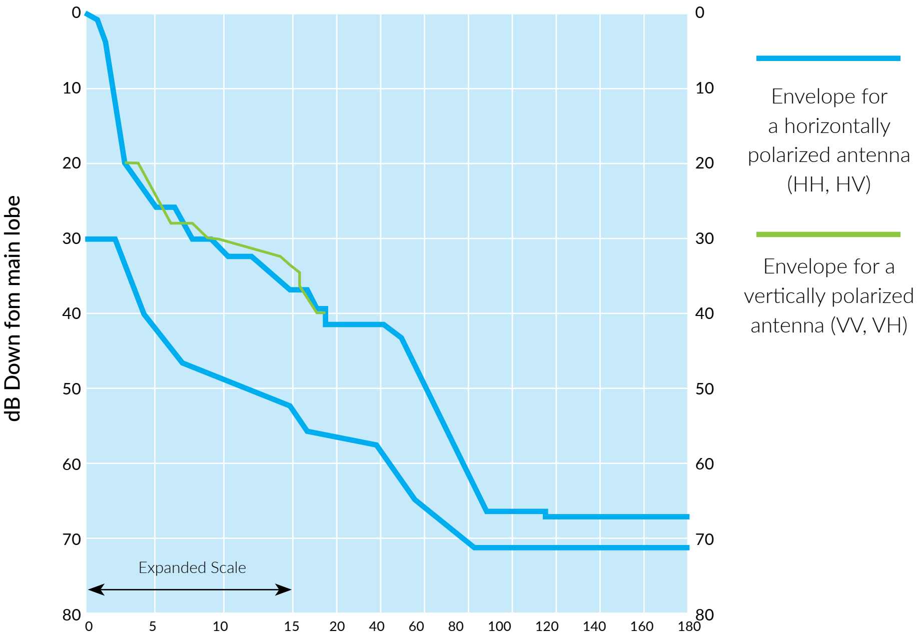

8.9: A typical radiation pattern envelope (RPE) document

To avoid these issues, industry standard software tools such as iQ-link®XG from **CommScope** can smooth the planning process and assist in regional overview. Antenna manufacturers offer assistance as well, providing planners with the radiation pattern envelope (RPE). An RPE document includes a performance summary and the key specifications related to antenna gain, beam width, cross-polar performance and radiation patterns (figure 8.9).

The chart describes the directional properties of the antenna by mapping it directionality (in dB) against its azimuth angle. As this chart shows, the envelope has a main beam area at zero degrees, corresponding to the electrical axis of the antenna. This is the line-of-sight direction, where the directionality is at its maximum.

Away from the main beam, the directionality quickly decreases. This corresponds to a drop-off in antenna sensitivity, whereby signals transmitted or received away from the on-axis direction reduce rapidly. The link planner uses this information to determine how much of their new proposed link signal will deviate from the intended direction and assess whether this is likely to present problems (e.g., interference, threshold degradation, etc.

Because of their importance in the planning process, RPE documents are strictly regulated. In Europe, the European Telecommunications Standards Institute (ETSI) publishes several classes of envelope standards that all antennas must satisfy. A Class 2 antenna may be permissible in locations where interference is not an issue, but cannot be used where a Class 3 antenna's stricter standards are required. The importance of using Compliant Class 3 or better specification antennas is discussed in further detail in a later section.

Most regions across the globe have adopted ETSI standards with the notable exceptions of Canada, Australia and the United States, which have their own regulatory envelope minimum standards.

Signal polarization
The orientation of a signal's electric field relative to the ground. It may be horizontal or vertical.

Protecting microwave systems from the elements

Like every stage in the chain of communication, backhaul must be reliable and available at all times. Downtime means lost revenues, irritated customers and expensive repairs. In practical terms, downtime is measured as a percentage, or in minutes per year (table 8.10). A detailed explanation of reliability predictions and measurement can be found in chapter 10.

Downtime, minutes per year	Availability
525.6	99.9000%
52.56	99.9900%
26.28	99.9950%
5.256	99.9990%

8.10: Downtime in minutes per year and corresponding availability percentages

This growing need for reliable communications within realistic budget constraints is partly what drives demand for microwave backhaul. However, due to its open exposure to the elements, certain reliability-limiting factors are unavoidable:

- Precipitation and moisture
- High winds
- Temperature variances
- Lightning strikes
- Atmospheric refraction

Fortunately, each challenge to reliability has an available mitigating measure.

Rain and snow

As mentioned earlier, lower-frequency microwave bands propagate very well across long distances, allowing hops of 50 km or more. In fact, the most significant limiting factor is not distance itself, but atmospheric conditions. Rain falling through the signal path reduces signal strength, an attenuating phenomenon known as "fade."

In frequencies above 11 GHz, rain-induced attenuation becomes more pronounced, reducing hop distances accordingly. Rain, and to a lesser extent, snow, can scatter signals in these frequencies. The impact depends on the rate of precipitation, the frequency involved and the signal polarization (the orientation of the signal wave, which may be horizontal or vertical).

Horizontal signals are more adversely affected by rainfall due to the shape of raindrops as they fall, so vertical polarization is the preferred choice for any link-planning.

Fortunately, it is possible to mitigate these effects based on the calculations of rain outage models, building a safety margin into the transmission's power levels to compensate for expected loss and assure a reliable hop between stations. Modern microwave radios will even adjust power on the fly when needed, using an automatic transmit power control (ATPC) system.

Precipitation also interferes with polarized transmissions through an effect called polarization rotation, which essentially turns a signal's polarity enough to interfere with other signals. To counter this effect, a cross-polar interference canceller (XPIC) samples signals in both polarities in order to produce a wave that cancels out the interfering, "rotated" part of the signal.

Adaptive modulation

Another technique gaining widespread acceptance in microwave backhaul applications is called adaptive modulation. In addition to compressing, or modulating, network traffic into smaller bandwidths at higher signal levels, adaptive modulation adjusts the amount of modulation in response to any link impediments. The result is that adaptive modulation can dynamically reduce traffic to compensate for the impaired signal level while still maintaining the link, albeit with lower capacity.

Mitigation methods can also be built into link designs themselves. In multiple-hop situations, mesh and ring topologies provide alternative signal paths that bypass problematic hops by rerouting around them. Path selection is dynamic and adapts on the fly to changing conditions (figure 8.11).

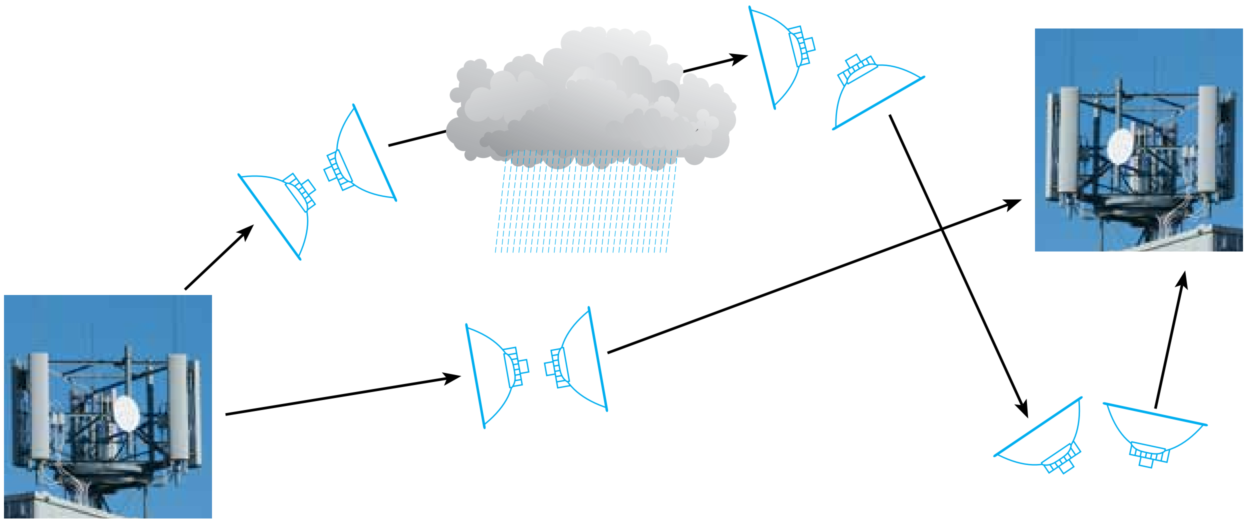

8.11: Network topology with dynamic routing paths

Automatic transmission power control (ATPC)
A system that dynamically raises transmission power to overcome the effects of interference.

The Sentinel Performance Advantage:

Microwave antennas with specialized adaptive modulation schemes

Radome
A wind- and water-proofed fabric or plastic cover that protects an antenna from the elements.

Fog

Fog only presents a challenge to the highest microwave bands above 60 GHz. Unlike rain, snow and other precipitation, it presents no real obstacle to lower, more commonly used microwave frequencies.

Temperature

By itself, temperature has little effect on microwave signals. However, if water vapor is present in transmission lines, it can condense there and impede performance when the temperature drops. The effect is similar to the attenuation caused by rain.

The hardware effect

In addition to climate effects on transmission, weather also affects a microwave installation's physical integrity. Snow or ice accumulation on an antenna introduces critical weight considerations for the antenna's design, as well as for the components installed there.

Long-haul antennas generally employ a fabric enclosure, or *radome*, that protects sensitive components from wind or moisture infiltration. Ice shields offer additional protection from falling ice, either from the antenna itself or from positions above it (figure 8.12).

8.12: Ice accumulation on a radome-protected installation

Wind

The force of wind on an antenna structure is called wind load, and it can present a serious threat to tower-mounted equipment. Wind speeds rise with altitude, so a breeze at ground level can become a gale a few hundred feet up. This is why antennas are designed to ensure mechanical integrity under all anticipated environmental conditions, typically able to withstand 180 km/h (112 mph) winds without moving on their mounts. Above this speed, some flexibility may be permitted, with a topmost survival rating approaching 250 km/h (155 mph) wind speeds.

Far below these extreme limits, an antenna will flex and move slightly, even in moderate winds, within its engineering standards. While this helps maintain the integrity of the structure, it introduces an obvious challenge to LOS transmissions that depend on precise targeting between links. To ensure that this slight movement doesn't interfere with a successful hop between stations, these temporary misalignments must be factored into the antenna's power budget.

Lightning strikes

Installed in open, unobstructed locations, microwave antenna towers are natural targets for lightning strikes. If a strike passed through the antenna itself, serious damage to sensitive components would result. To avoid this danger, low-resistance earth-paths—in effect, lightning rods—are installed to direct lightning strikes away from critical components. For a full exploration of how to guard against lightning, see chapter 11.

Atmospheric effects

Atmospheric effects can disrupt reception, particularly for lower-frequency signals. Under some circumstances, a signal may essentially be received twice, first by its intended LOS connection, and then again as a slightly delayed echo of itself as a result of atmospheric refraction or ground reflection. The tiny timing difference can mean the signal arrives out of phase, in effect interfering with itself. This effect is known as multipath fading or dispersive fading, as illustrated at the right (figure 8.13).

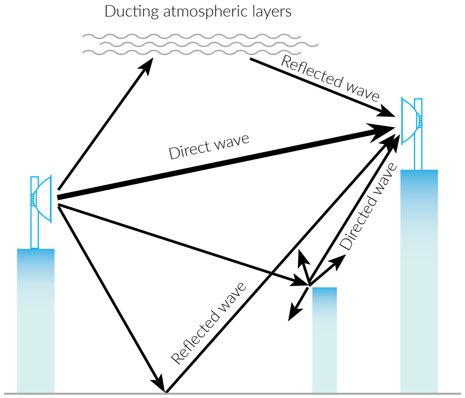

8.13: Dispersive fading creating out-of-phase signal echoes—delayed signals bounce off the ground and vertical obstacles, and also refract in the atmosphere

8.14: A typical vertical space diversity arrangement

Option 1: space diversity

To counter the effects of dispersive fading, we can add a second, uncorrelated parallel microwave transmission path, separated vertically (figure 8.14) or horizontally.

Because the two paths don't share exactly the same space, their signals don't encounter exactly the same fading effects. The practical result is that the receiver has the option of accepting its signal from the path that happens to be less disrupted at the time.

Option 2: angle diversity

Another countermeasure is angle diversity, which requires only one antenna instead of the two. However, that antenna requires two vertically separated feed systems. While a less expensive option, angle diversity is also less effective than space diversity and is used only where a second antenna cannot be added to the installation.

Option 3: frequency diversity

Frequency diversity is another means of combating atmospheric or dispersive signal loss. A secondary, standby channel operates at a different frequency from the main channel. Since different frequencies propagate differently, two signals of different frequencies don't experience the same attenuation, doubling the chances of clear reception.

Flat fading
Total signal loss caused by atmospheric refraction. It is the result of a signal being bent completely out of its LOS connection with its receiver.

Flat fading

Flat fading is another atmospheric hazard for low-frequency signals. Unlike dispersive fading, the entire channel is attenuated because refraction has disrupted the LOS connection over the hop, so it misses the receiver entirely. A second, uncorrelated microwave path or increased power via ATPC can counter this effect.

Improving network reliability

Some of the reliability-limiting factors can be avoided by using compliant spec antennas from a reputable manufacturer. With the pressure to minimize total cost of ownership, it is tempting to purchase the lowest price antenna option, but you must also consider ongoing operations expenses in budget calculations. An inexpensive install may end up costing far more than a more expensive antenna with better characteristics for your application.

Noncompliant antennas are not simply identified by cheaper prices. They also include design problems that can spell trouble from day one (figure 8.15). Generally, they are:

- Not tested and feature non-repeatable designs
- Prone to deteriorate too quickly, resulting in RF leakage, moisture ingress and degrading metallic components
- Used with third-party add-ons for normal operation, which introduce new opportunities for corrosion, vibration and other loss of integrity.

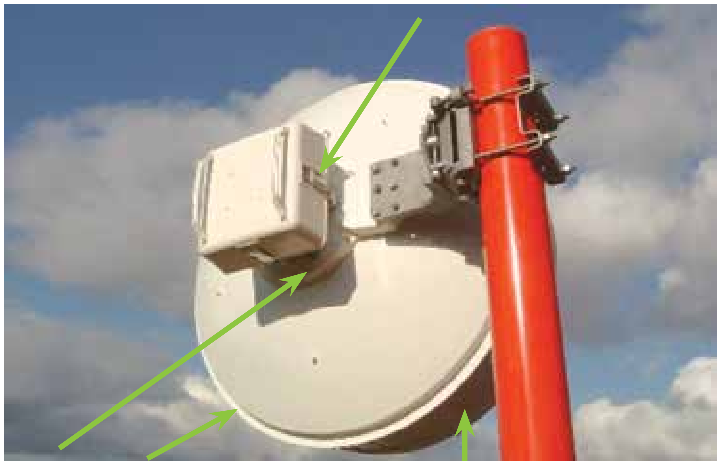

8.15: Example of RF leakage in a noncompliant microwave antenna

The hidden costs of noncompliance

CommScope recently conducted a study to measure the cost and benefits of using lower-quality, noncompliant antennas. In this study, we examined three types in the 15, 18 and 23 GHz bands. In actual operation, the true costs began to emerge:

- **19% of the backhaul network failed** in the 15 GHz band
- **29% of the backhaul network failed** in the 18 GHz band
- **21% of the backhaul network failed** in the 23 GHz band

Obviously, these failure rates are unacceptable to most modern applications. Choosing a compliant antenna, even a more expensive one, offers benefits that can save money and hassles once the antennas are in operation, such as:

- Less potential external noise
- Less RX threshold degradation
- Longer hops
- More link availability
- More capacity
- More efficient use of the spectrum

In addition, there are other initial cost savings that aren't necessarily reflected in the antenna price itself, such as:

- Smaller antennas with lower shipping costs
- Lower landed cost
- Lower tower leasing cost

It all adds up to lower total cost of ownership, which makes good sense and means good business. There are a few common-sense steps you can take to avoid noncompliant antennas and assure your network's availability and reliability. They all start with choosing a reputable supplier.

1. Take advantage of your supplier's resources to demonstrate structural integrity under all anticipated environmental conditions.
2. Review the supplier's antenna interface design data and test range facilities for each integration type.
3. Avoid third-party add-ons that don't qualify at the integration level; the third-party supplier won't have this information, but your antenna supplier should.

This basic due diligence will pay off in reliability, speed and total cost of ownership.

Co-channel dual-polar (CCDP) operation
Using both horizontal and vertical polarity of a single frequency to double available bandwidth.

Network capacity and managing demand

Simply put, capacity is a network's ability to handle transmission traffic. In the case of cell communications, this traffic means voice and data—often, a great deal of data. As capacity demands continue to rise with the spread of long-term evolution (LTE) and 4G mobile devices, smart planning becomes even more important to assure headroom for tomorrow's data-hungry applications.

Modulation. One way to boost capacity along a microwave link is called modulation. By employing different modulation schemes, more traffic can be squeezed into the limited bandwidth available when needed. The tradeoff is that higher modulation schemes require higher signal-to-noise performance to maintain the integrity of the data, boosting operating costs. Plus, any disruptive effects like those described above create bigger problems than they do for un-modulated traffic.

Adaptive modulation. Recently, adaptive modulation has become a universally adopted technique to help operators balance traffic and reliability needs. Adaptive modulation scales the amount of signal modulation employed as a function of the link's condition. So, if rain or other factors are present, modulation is dialed down to maintain error-free, if somewhat slower, traffic rates. When the link condition improves, modulation is automatically increased to take advantage of prevailing conditions.

Co-channel dual-polar operation. Another capacity-boosting technique leverages the polarization characteristics of microwaves themselves, which allow two streams of traffic to travel the same bandwidth at the same time—one vertically, one horizontally. This technique is called *co-channel dual-polar* (CCDP) operation.

CCDP is often used in short-haul antenna systems where an integrated dual-polarized antenna is created using an Ortho-mode transducer (OMT) to attach two ODU radios to a single antenna. This arrangement maintains a high level of isolation between the two signals for maximum clarity (figure 8.16).

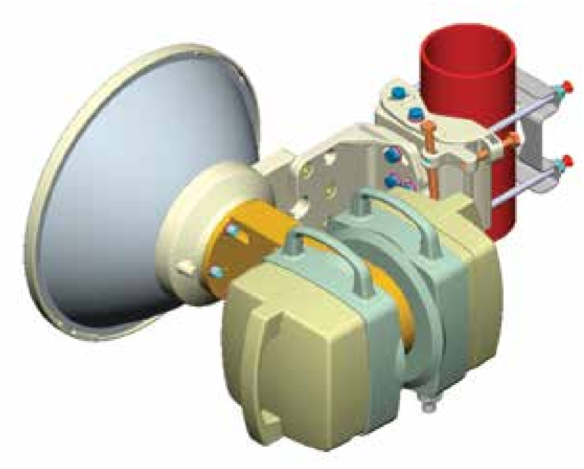

8.16: An example of an integrated dual-polarized antenna

Microwave capacity

With all these capacity-boosting tools at our disposal, modern microwave backhaul data rates typically range between 32 and 155 Mbps (megabits of data per second).

However, as technology and techniques continue to improve, much higher data rates are becoming possible and economical. In the near term, 1 Gbps (gigabits of data per second) single-channel systems are coming online now.

The future of microwave backhaul

Explosive demand for mobile communications drives the need for cost-effective microwave backhaul, which, in turn drives new innovations. Old technologies are replaced by new, more efficient ways of moving data faster, more reliably and at less cost. Microwaves form the backbone of these new technologies.

For instance, legacy networks built on circuit-based transmission protocols maintain their connection regardless of how much, if any, traffic is actually being transmitted at any given time. This was once necessary to carry voice communications, but the world has moved on and this technology isn't efficient in our on-demand data world. Packet-based microwave radios encode traffic from multiple sources and routes it through IP over Ethernet, so it only utilizes bandwidth as it's needed, reducing wasted energy and capacity.

Capacity and coverage

Improved network coverage is another critical requirement for emerging mobile technologies, such as long-term evolution (LTE) and 4G mobile networks. Customer-level access is required in all regions if high-speed connectivity is to be available as a constant resource, and the ongoing rollout of base stations (known as macro-cells) offers more and more coverage to broad geographical areas.

At the same time, the capacity available from these macro-cells diminishes with distance from the base stations, requiring the addition of smaller coverage micro-cells to sustain capacity. Pico-cells represent a further layer of coverage for built-up urban areas. Both micro- and pico-cells require backhaul connectivity, and that means microwave links.

The techniques of tomorrow

There are currently a number of studies researching which type of backhaul medium will be most suitable for these new scenarios, and it seems likely that a mix of the various technologies will form the networks of tomorrow. Non-line-of-sight (NLOS) schemes using unlicensed bands offer one method of small-cell backhaul, allowing signals to turn corners and avoid obstacles in urban environments where LOS systems aren't as effective as they are in the open.

In other exciting developments, there seems to be great promise in the recently opened 60 GHz unlicensed band. At these high frequencies, the oxygen in the air itself can absorb signal power, making it suitable for short hops, often less than a single kilometer. While this may seem like a limitation rather than an advantage, an urban environment benefits from this short-hop option because they offer high data traffic rates and have such limited ability to interfere with one another from one pico- or micro-cell to another.

Lastly, the recent opening of the E-band spectrum (71-76 GHz, 81-86 GHz and 92-95 GHz) promises to open new avenues for high-capacity microwave backhaul. Operating under a light-licensing regime, very wide channel assignments (n × 250 MHz) are available to operators, making 1 Gbps data rates a real possibility without the restrictions of high-level modulation schemes.

With so many advances in recent years, and so many still to come, this is truly an exciting time in communications.

Backhaul makes modern communications possible

The ever-increasing complexity of modern communications networks demands more efficient and innovative ways of managing backhaul. In our wireless age, microwave backhaul moves the information that moves the world forward.

Though an efficient and cost-effective means of moving data through central processing, microwaves face challenges and limitations from weather, topography and interference from other nearby links. Smart planning and the right equipment can overcome these obstacles, allowing microwaves to connect us to one another wherever we may be.

As children, we first experienced the idea of backhaul as a length of string between two soup cans or the scratchy sound of a walkie-talkie. As adults, backhaul has become a key part of daily life as we depend on our cell phones, computers and other devices to connect with friends, family, colleagues and the world at large.

From strings to radio waves to microwaves, backhaul is the technology that keeps us all connected.

Chapter 8 summary

Microwave backhaul

- Communications moving through a centralized routing center
- Accomplished via links or “hops” from station to station

Microwave deployments

- Rely on line-of-sight between hop points
- Can connect multiple radios to antenna via paired ODUs

Issues affecting reliability

- Rain, snow and hail
- Wind loads
- Ice buildup on antennas
- Lightning strikes
- Dispersive fading

Avoiding external interference

- Coordinate with other links in vicinity
- Use manufacturer RPE to choose specifications

Avoiding trouble

- Choose a reputable supplier
- Look at entire cost, not just the purchase price

Maximizing capacity

- Adaptive modulation
- Co-channel dual-polar operation

Chapter 9

The energy that connects the world:
Powering wireless networks

Every year, our reliance on always-on technology grows. We expect to be able to place a call or surf the Internet with our cell phones at any time, under any circumstances. However, the electrical infrastructure that powers our networks has not kept pace with the explosive growth of cellular access—and certainly hasn't kept pace with our expectations of 24/7 availability.

In previous chapters, we have explored some of the intricate and complex ways different components in a cellular base station come together to work efficiently and reliably. So far, the idea of powering these stations has been taken for granted. In real life, however, we don't have this luxury. Power supply connection is a very real challenge, and planning for the inevitable outages and interruptions is critical to keeping the network operating, no matter what complications may arise.

Direct current (DC)
An electrical current that runs continuously in a single direction, making it well suited for use in motors and electronic components such as semiconductors. Batteries also produce DC current.

Alternating current (AC)
An electrical current that changes polarity (direction) 50 to 60 times per second. It offers significant efficiencies when transmitted across power lines, making it the standard current for household use.

DC vs. AC power

The core technology that drives modern communications runs on *direct current* (DC) electricity. DC is different than the *alternating current* (AC) we use in our homes and offices. Since electrical power sent across transmission lines from power plants is AC, a device called a rectifier is needed to convert the AC current to DC before it can be used by the communications equipment.

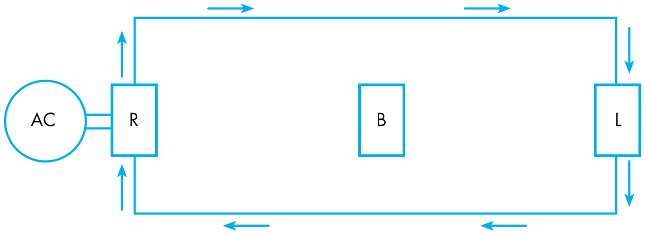

9.1: Basic power system diagram—arrows indicate the DC current's direction

The rectifiers' output is connected to both the radio and its transmission equipment—the "load" for the current—as well as the backup battery equipment (figure 9.1).

Why go to the extra trouble of converting AC to DC? There are several reasons. First, most communications equipment includes semiconductors and other integrated circuitry that are specifically designed to operate with DC, such as:

- Telephone switches
- Microwave transmitters
- Fiber-optic transmitters
- Mobile radio and cellular systems

Another reason DC power is preferred for communications systems is its reliability advantage. Even the most advanced electrical grid can fail from time to time, and no one is immune to the possibility of power interruptions that may last for hours or days. When the outage occurs at a cell station, shutting down is not an option. So battery or fuel-cell backups are installed to allow continuous operation and these power sources produce DC power. This system is used for both cellular and conventional land-line telephone service, which is why you still get a dial tone even when your home's AC power is out.

DC power by the numbers

Communication equipment requires specific voltages of DC current. Most commonly, the requirements are positive 24 volts (+24V) and negative 48 volts (–48V). These high voltage values reduce their associated current and lighter current requirements allow for smaller and less expensive fuses, circuit breakers and cables.

The +24V value evolved from the early mobile radio industry, back when equipment was designed to draw power from a +12V automobile battery or +24V truck battery, the same kinds of batteries used today.

Going beyond the basics

As mentioned above, the basic core components of a communication system's power connection are the rectifier, which converts AC current to usable DC current, and the batteries which assume the load when external power is interrupted. Once DC is online, however, we must consider the specific power needs, or loads, of different equipment. We need a means to distribute the correct voltages to each piece of equipment.

To make these adjustments, devices called DC-DC converters can modify the primary DC voltage to suit the needs of a piece of equipment that demands a different voltage. In the event that some equipment requires AC current, we can also include a device called an inverter, which changes DC back into AC. Because this all takes place behind the station's main rectifier, this reconverted AC power isn't subject to interruptions from the external power supply. A sample diagram showing all these components appears below (figure 9.2).

You may also notice an important new element appearing in this diagram: the generator. Generators supply AC power for the station's rectifiers if external power is interrupted. This keeps all systems operating and the batteries fully charged. There are several kinds of batteries and generators in common use. Selecting the right generator often depends on space, budget and operational expectations.

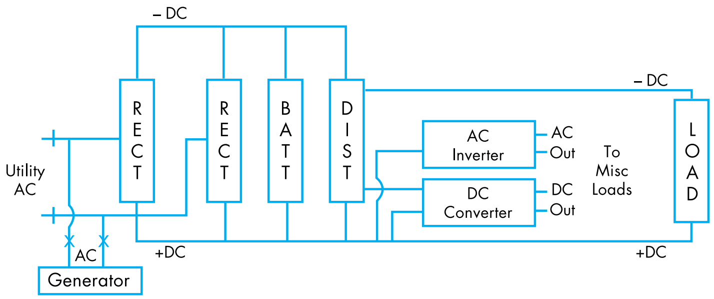

9.2: Block diagram of a basic telecommunications power system

Volt (V)

A measurement of electric potential difference between two points in a path. Voltage is sometimes referred to as "pressure," because it shares many characteristics with pressure in a water pipe.

The chemistry of batteries

At its most basic, a battery is an electro-chemical method of storing and releasing electrical energy. A battery contains chemicals that react with one another, producing DC electrical current as a byproduct. The specific chemicals and processes may vary—we will examine some common varieties below—but, on a fundamental level, all batteries operate on this core principle.

While the battery is receiving a charge from a fully operational system, it does not discharge any current. Rather, it consumes a small amount of DC from the power relayed by the rectifier. When this current is interrupted, natural processes kick in and the batteries discharge their stored DC power in accordance with the load placed on the circuit. Most importantly, this shift occurs seamlessly and leaves no interruption in current.

Reliable backup times usually range from two to eight hours, depending on the load. Since telecommunications providers must be able to specify expected operational times, choosing the best type and configuration of batteries is critical.

Lead-acid batteries

These are commonly used as backups for telecom power systems. They are highly compact relative to their output, and are the same kind you would find under the hood of your car. They are available in vented and valve-regulated forms (figure 9.3).

Flooded battery

VRLA battery

9.3: Vented (flooded) and valve-regulated (VRLA) batteries

Vented (also known as **wet** or **flooded**) **batteries** are a mainstay of telecom central offices and switching centers. They maintain a charge for up to 20 years or longer. However, they demand a great deal of costly maintenance such as water treatment, spill containment and forced-air ventilation. These drawbacks make them less suited to remote cell base stations.

Valve-regulated lead acid batteries (VRLAs) are recombinant batteries, which means oxygen and hydrogen can recombine to prevent water loss. Because they don't need added water and can therefore be built in spill-proof containers with pressure relief valves, these are often called sealed lead-acid batteries. They are easier to ship, maintain and install, and cannot freeze like a flooded battery, making them the preferred choice for cell base station use.

Battery cell configurations

Each battery is comprised of cells arrayed in electrical series. The number of cells used in a particular battery is determined by the overall required operating voltage and the voltage of each cell. For VRLA batteries, a standard voltage of 2V per cell is the norm so, to power a +24V system, 12 such cells must be connected in a series. Likewise, a –48V system requires 24 cells. Often, VRLA batteries are available in 12V blocks, like those seen in figure 9.4.

The arrangement of these batteries in series determines the voltage polarity. If each of the 12V batteries shown above is rated for 100 amp-hours, then each series string of batteries could be expected to produce 100 amps of current for one hour. Capacity is directly related to the size of the battery, but rather than spending more on larger batteries, we can achieve the same boost to capacity by adding more strings of batteries to the system in parallel as opposed to adding them in series. This option also provides safeguards against the failure of an individual battery, which would remove its string from the system altogether. By connecting in parallel, the spare capacity is already online and ready to maintain the current for its rated length of time. A diagram of sample parallel battery strings rated at 240 amp-hours, appears below. Note that if a battery failed in any of the three series strings, the remaining two strings would continue to supply steady power at 160 amp-hours (figure 9.5).

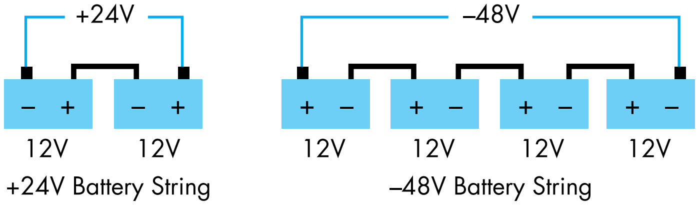

9.4: VRLA batteries configured in series for +24V and –48V applications

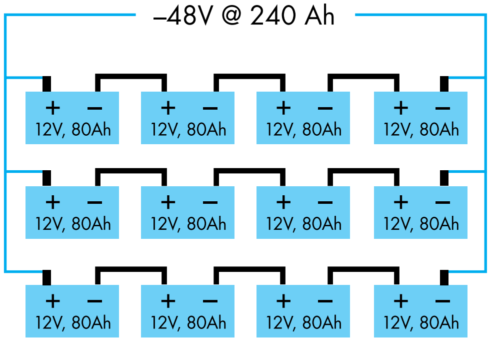

9.5: A multi-string battery configuration offering additional redundancy

This configuration also provides a convenient means of maintaining the batteries. Often, these strings will be installed with separate disconnection breakers, which make it easier to locate failures and isolate problems that could otherwise cripple the entire system.

Voltage polarity: positive (+) and negative (–) voltage

The + and – designations of +24V and –48V refer to which polarity of the battery circuit is measured; in terms of actual power produced, the distinction is meaningless.

Generators: the first line of defense

If batteries are the last line of defense against service interruption, generators are the first. Since batteries alone can only maintain operations for a few hours, longer AC service interruptions require a longer-term solution and that means generating our own power.

Unlike batteries, generators provide power by burning fuel. Like batteries, there are different types and configurations available. Which one you install depends on factors like space, cost and service expectations.

Since they operate outside the cell station's internal DC system, generators aren't considered part of that system. Because they supply the DC system's rectifiers with the AC they need, however, they're a vital link in assuring reliable operation. In the event the station must switch from external to generator-supplied AC power, an electrical device called a transfer switch shunts the load to the generator.

For stations with permanently installed standby generators, an automatic transfer switch (ATS) monitors external power levels and performs the changeover to generator power if the voltage drops below acceptable levels. View an example of AC generator. (Figure 9.6)

AC permanent genset

9.6: AC permanent generator

Generator fuel sources

You may have an emergency generator at home, or use one for electrical service in a remote location. Chances are that your generator runs on ordinary gasoline. Larger generators may also use compressed natural gas or diesel fuel, but the optimal generator for a DC system like a cellular base station is hydrogen fuel cell technology, for several good reasons.

Hydrogen fuel cells use proton electrolyte membrane (PEM) technology to generate DC power. We'll explore this later, but from a practical standpoint, hydrogen fuel cells offer several key advantages over their petroluem-fueled cousins.

Hydrogen fuel cells:

- Run on inexpensive, readily available hydrogen
- Produce very little heat, preventing damage to other components
- Offer higher efficiency than typical diesel generators
- Operate in ambient temperatures from −40°F to 122°F
- Are more compact than other fuel cell types
- Are quiet and non-polluting
- Have no moving parts, extending their operational life spans
- Require very short warm-up periods so they come online quickly

Generators in general, and hydrogen fuel cells in particular, can provide large amounts of steady, predictable power for extended periods of time.

The science of fuel cells

In a PEM fuel cell, hydrogen atoms are stripped of their negatively charged electrons at the anode end of the battery. This ionizes the hydrogen atoms, giving them a positive charge. The liberated electrons provide the actual DC to the connected equipment. Meanwhile, at the cathode end of the battery, oxygen combines with the hydrogen and electrons returning from the circuit, creating water (figure 9.7).

As long as the system is supplied with hydrogen and oxygen, the fuel cell will produce a predictable amount of current.

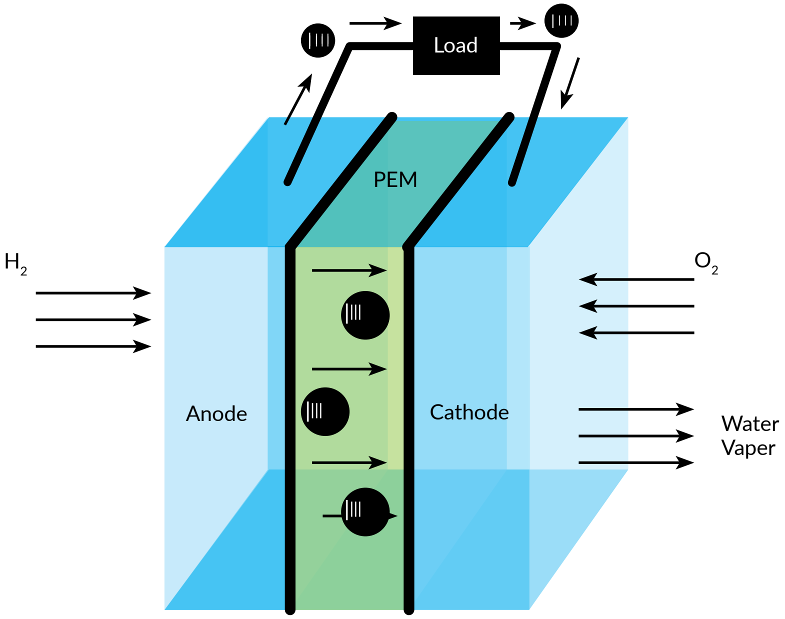

9.7: Basic operation of a hydrogen fuel cell

Hydrogen fuel cell installation

Hyrdogen fuel cell installations include three main components: hydrogen storage, a fuel cell stack and power module, and an integrated bridge battery. A typical layout is shown in figure 9.8, along with a view of an actual installation in figure 9.9.

Hydrogen storage

Like many liquid or gaseous fuels, hydrogen is stored in tanks. For a fuel cell deployment at a cellular base station, this storage would usually comprise 8 to 16 cylinders of compressed hydrogen.

Fuel stack and power module

Because of the power demands of base station equipment, a single fuel cell cannot provide enough power by itself, so multiple cells, called a stack, are linked together in a fuel cell power module (figure 9.10). This linked arrangement is similar to the multiple battery configurations discussed earlier.

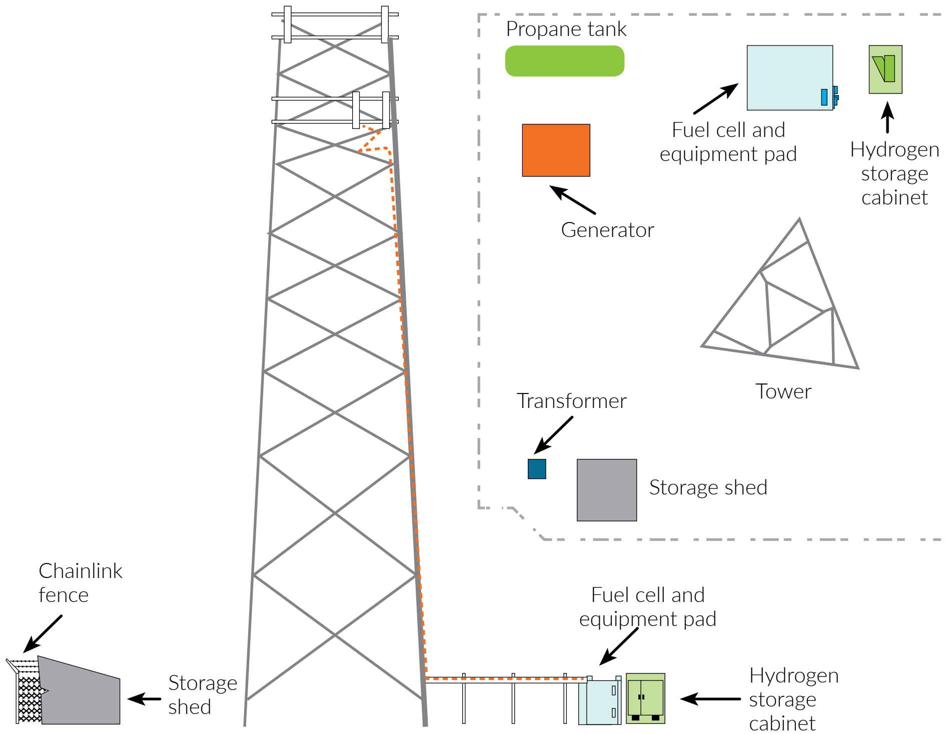
Propane tank
Fuel cell and equipment pad
Hydrogen storage cabinet
Generator
Tower
Transformer
Storage shed
Chainlink fence
Storage shed
Fuel cell and equipment pad
Hydrogen storage cabinet

9.8: *A typical layout for a base station's fuel cell power plant*

9.9: *An actual fuel cell and hydrogen storage cabinet installation*

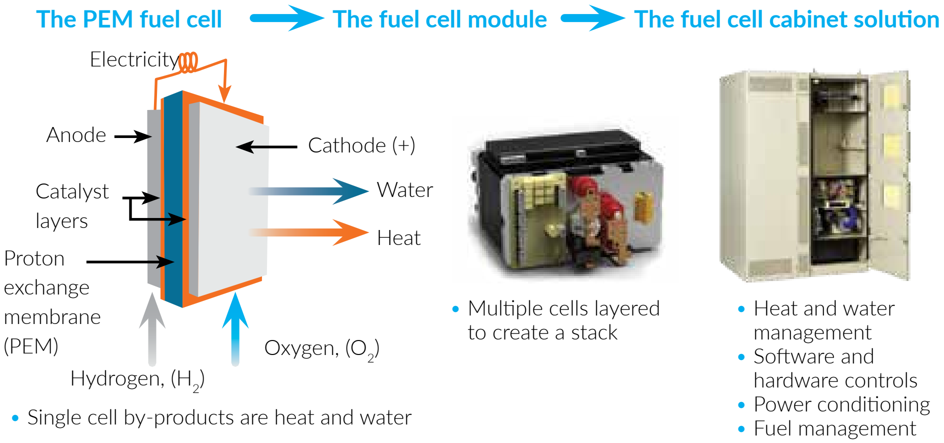
The PEM fuel cell
The fuel cell module
The fuel cell cabinet solution
Electricity
Anode
Cathode (+)
Catalyst layers
Water
Heat
Proton exchange membrane (PEM)
Oxygen, (O_2)
Hydrogen, (H_2)
• Single cell by-products are heat and water
• Multiple cells layered to create a stack
• Heat and water management
• Software and hardware controls
• Power conditioning
• Fuel management

9.10: *The connection between fuel cell, fuel cell stack and fuel cell power module*

Bridge batteries

While hydrogen fuel cell generators warm up quickly and are ready for service shortly after an AC power interruption, there is a brief bridging period when batteries must carry the load alone. These integrated battery backups provide the necessary power to cover the brief time needed for a fuel cell to come online.

Fuel supply and maintenance

Hydrogen fuel cells are efficient, quiet, reliable and long lasting. But, like any generator, they require a supply of fuel to operate. When the hydrogen supply dwindles, an automatic alarm alerts the system's operator that fresh supplies are needed. To preserve hydrogen, fuel cells are designed to shut down once external AC power has been restored. They are then ready to engage the next time an interruption occurs.

With no moving internal parts to wear out and very little heat radiation, fuel cells require very infrequent maintenance, often amounting only to an annual inspection of their oxygen intake filters and radiator fluid levels.

Rectifiers: the AC-DC interface

Once beyond the external source of power, whether it's commercial AC electrical service or an AC generator, the rectifier is the core of the system's DC power distribution needs.

The rectifier is designed to provide an output DC voltage level that maintains the battery charge. This level is called float voltage, and it supplies the equipment load as well as a trickle charge to the battery. In the event of an AC power interruption, the rectifiers go offline and the batteries automatically kick in, providing the same level of power to the rest of the system. When external AC power is restored, the rectifiers re-engage and the batteries return to their trickle-charging state (figure 9.11).

In practical deployments, multiple rectifier modules are usually required to supply power for the station's load. Rectifier modules are connected in parallel, letting each one share an equal part of the load, a practice known as load sharing. With load sharing, operators can design in a degree of redundancy to guard against individual rectifier module failures.

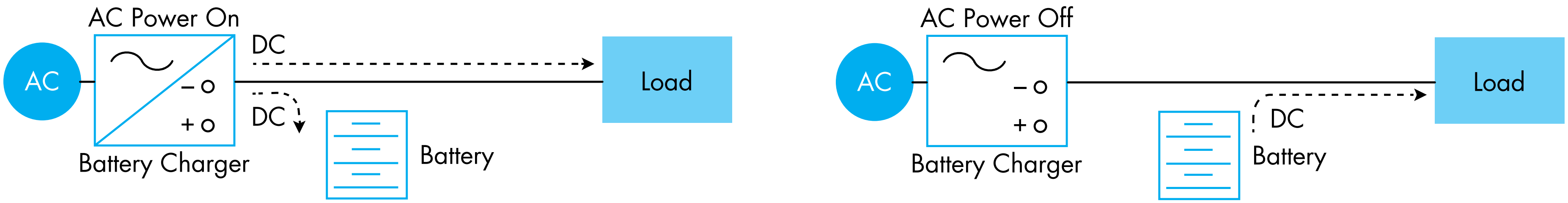

9.11: External AC power on (float mode) and off (discharge mode)

How redundancy adds up to reliability

For a –48V application with 200 amps of load, an operator may choose to install five 50-amp rectifiers.

Why add a fifth, when four would provide the requisite 200 amps? To ensure N+1 redundancy. Arranging multiple rectifiers in parallel allows load sharing to even out and shift the load.

Choosing the right rectifier

The first consideration in deciding which rectifier will best suit a given installation is the kind of AC power it will receive. Switchmode rectifiers are the preferred choice for cell and microwave sites since they can support multiple AC inputs and have a broader operating range—from single-phase to three-phase inputs. This flexibility means fewer rectifiers are required, saving money, space and maintenance.

Another consideration is the required output voltage. As we've seen, the rectifier must be able to output the voltage required by the station's equipment, that is, +24V or –48V. The amount of current must also be sufficient to power all the loads, such as the radio, transmission equipment and battery backup.

83A Switchmode
6.3H x 3.4W x 11.8D

9.12: Example of a typical rectifier

Distributing the power

The rectifier is merely the first stage in the DC power system. Once converted, the power must be distributed to the many component loads within the system. As mentioned above, these loads include core elements like the radio, transmitter and battery backup, but they can also include secondary systems like lighting, security networks and HVAC systems (figure 9.13).

In the most complex installations there may be so many components that up to 80 circuit breakers are required to manage them.

The most visible parts of the distribution system are the fuses or circuit breakers, which safely distribute DC power from the rectifier to each individual load. Fuses and breakers are connected in series between the power system and their loads, protecting them from short circuits and damage from overload conditions. They also provide a safe, simple way to manually shut down individual components or batteries for service, maintenance or replacement.

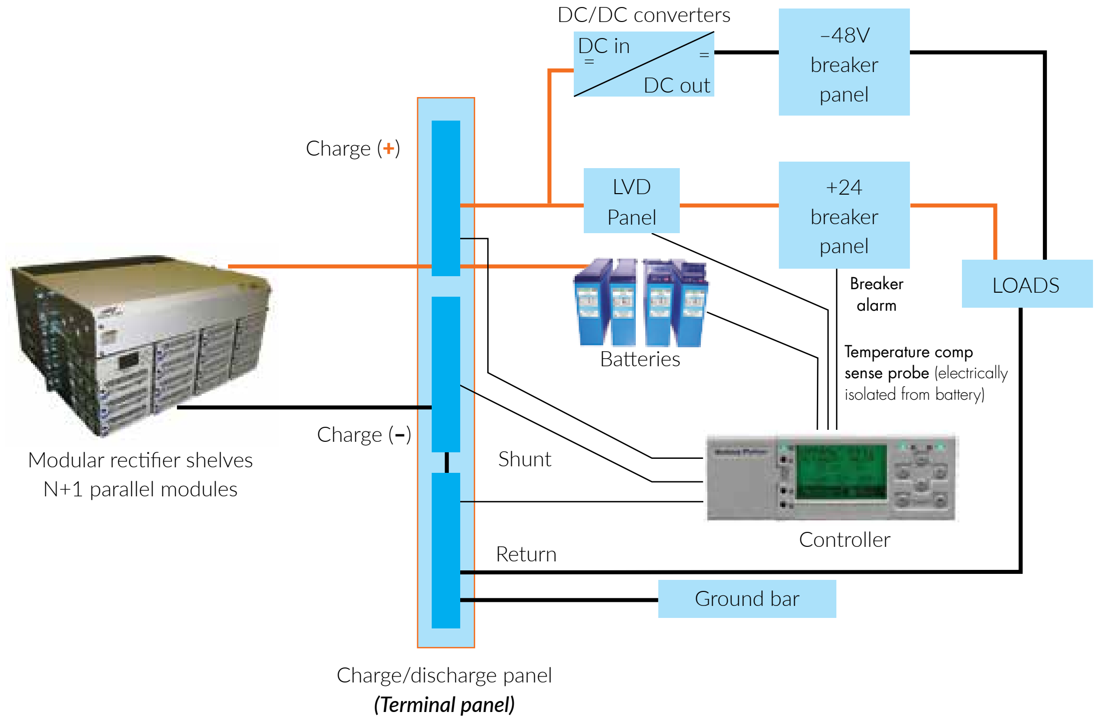
DC/DC converters
DC in
DC out
–48V breaker panel
Charge (+)
LVD Panel
+24 breaker panel
LOADS
Batteries
Breaker alarm
Temperature comp sense probe (electrically isolated from battery)
Charge (–)
Shunt
Modular rectifier shelves
N+1 parallel modules
Controller
Return
Ground bar
Charge/discharge panel
(Terminal panel)

9.13: A diagram of a typical +24V power distribution system

Images and illustration courtesy of GE.

Fuses vs. circuit breakers

While both fuses and breakers provide overcurrent protection, they do it in different ways. Fuses are designed to melt under unsafe currents, physically breaking the connection between the power source and the load. Circuit breakers have internal switches that pop to the "off" position under unsafe conditions, again providing a physical break in the circuit.

Sensitive wireless equipment requires "fast blow" fuses or short delay-curve breakers to provide the needed protection. Fuses are generally used for lower loads and offer the advantages of lower cost, greater flexibility and fast action. Circuit breakers are preferred for larger loads and do not require replacement every time they are tripped. Some typical examples are shown to the right (figure 9.14).

Because complex cell sites can require from 24 to 80 breakers of varying capacities, wireless carriers have adopted modular distribution structures that can accept a wide range of breaker sizes.

9.14: Common types of telecom fuses and circuit breakers

Surge protection

Typical variations in AC power are not the only threat to a cell site. Electrical events like lightning can also produce excessive voltages and currents, events known as electrical surges. Surge protection devices (SPDs) are incorporated to reduce the effects of these surges on sensitive electronics (figure 9.15).

An SPD features a non-linear voltage-current characteristic that reduces unsafe voltages by increasing the conducted current. In this case, a cell site's SPD operates on the same principle as a surge protector does in your home, safeguarding expensive electronics from lightning-induced surges.

9.15: Typical surge protector devices (SPDs); image courtesy of Raycap Corporation

Bus bar conductors

The cell site's distribution system is supported by the bus bar conductors, which physically connect the rectifiers to the batteries and DC loads. There are two bus bar connectors: the **charge bus** and **battery return bus.**

> The **charge bus** is a current-carrying conductor that connects the rectifier's output to the battery string. For instance, in a –48V system the negative rectifier lead would terminate on the charge bus along with the corresponding negative lead of the battery.
>
> The **battery return bus** provides a common return point for the loads connected to the power system. This common point is grounded to provide a low-impedance path for transients and noise, and offers a ground reference to all connected equipment.

The shunt

In a distribution system, a shunt is a low-resistance resistor designed to provide a specific voltage drop at a particular level of current. As the current passes through the shunt, it develops a small voltage proportional to the amount of that current. This voltage drop allows an operator to calculate the current flow in the system.

Battery disconnects

Battery disconnects are switches installed on a battery string that allow easy disconnection for maintenance or replacement. Some disconnects incorporate safety measures such as overcurrent fusing or breakers.

Load disconnects

Low-voltage disconnects (LVDs) are designed to respond to low voltage conditions in the circuit. Low-voltage load disconnects (LVLDs) can disconnect individual loads, while low-voltage battery disconnects (LVBDs) can disconnect a fully discharged battery.

LVDs serve three main protective functions:

1. They prevent damage to sensitive electronics caused by low-voltage (and hence, high-current) conditions
2. They prevent permanent damage to the battery from over-discharging
3. They prioritize which components are disconnected, and in which order, preserving limited function when necessary

Supervision, monitoring and control

Modern telecommunication power plants are equipped with electronic monitoring and control systems, generally called **controllers.** They keep track of system voltages, currents, temperatures and other key indicators. They also allow operators to make adjustments from a central monitoring point, usually on the power plant itself, on the distribution cabinet or in a rectifier slot (figure 9.16).

54.00V
75A
Distribution Fuse Alarm
Cordex
CXC
RESET
SYSTEM OK
MINOR ALARM
MAJOR ALARM

9.16: A system controller interface displaying voltage, amperage and alerts

With a modern controller, there is very little you can't learn about your cell site's power system performance.

This controller centralizes several key functions in a single, simplified interface, including:

Plant control. Control functions are extended from the supervisory panel to control other power system components. These panels communicate directly with the rectifiers, and in some cases can coordinate the sequenced restart of all rectifiers to prevent power surges during switchovers from external AC to a backup power source.

Manual equalizing. This allows a user to engage all rectifiers in equalize mode at once. This is useful for maintenance on VRLA batteries, equalizing cell voltage within a battery string.

High-voltage shutdown/overvoltage protection (HSVD/OVP). Controllers can automatically shut down rectifiers when DC output overvoltage conditions are detected, avoiding costly damage to load components.

Low-voltage disconnect (LVD). If a low voltage condition is detected in the backup batteries, the controller can open additional contacts to equalize voltage and close them again when levels equalize.

Battery temperature compensation. The controller can adjust rectifier output to meet the temperature-driven voltage needs of the batteries.

Charge current control. This feature limits the current flow to a battery when it begins recharging after a power interruption. By keeping the battery from recharging too quickly, it prevents overheating and prolongs life.

Battery diagnostics. The controller can estimate the "health" of the battery and predict how long it will provide power based on its charge status.

Alarm monitoring. The controller monitors critical functions like distribution and battery fuse alarms, rectifier failures, converter failures and so forth. It reports these alarms by way of network backhaul interfaces and LED indicators. Some units include audible alarms as well.

Status monitoring. The controller can measure and compare the battery charge to the system load via an external shunt.

Plant history. Controllers can log power system details over a span of time, including such statistics as thermal performance of outdoor enclosures, battery cell states, or variations in AC input experienced by the rectifiers.

DC-DC power conversion

Some wireless sites require multiple DC voltage outputs, such as +24 VDC and −48 VDC. One solution is to install a second rectifier plant, but doing so comes with the burden of including a second battery backup array as well, which consumes considerable space and adds cost.

Another solution is to use a DC-DC converter system, an electronic power conversion device that changes a DC input voltage to a different DC output voltage. Below, you can see where a DC-DC converter system is connected in series between the main DC power system and the site's load (figure 9.17).

A DC-DC converter system actually consists of multiple DC-DC converters arranged in parallel. It may also incorporate many of the same functions as the primary DC power system, such as distribution. It also has dedicated fuses or circuit breakers, isolating it from the rest of the system.

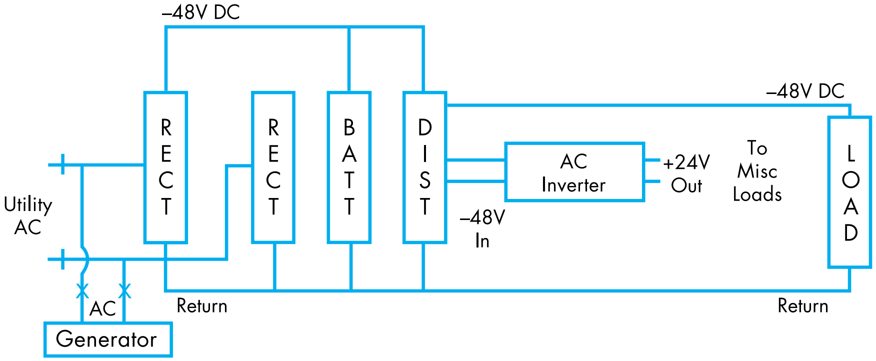
−48V DC
R E C T
R E C T
B A T T
D I S T
−48V DC
AC Inverter
+24V Out
To Misc Loads
L O A D
−48V In
Utility AC
AC
Generator
Return
Return

9.17: A DC-DC converter system connected in series to a cell site's power system

Since a DC-DC converter system does not have an associated battery connected to its output, it isn't bound by a battery system's requirement for precise output voltage. However, since it is necessarily energized by the primary DC power system, that demand must be figured into the power system's initial design.

Dual-voltage power plant using DC-DC converters

Ordinarily, a DC power plant's rectifiers and batteries supply consistent power for a single voltage requirement. If two requirements exist at the site and the demand is heavily skewed toward one voltage or the other—either predominantly +24V or −48V—the secondary voltage may be supplied through the addition of a DC-DC converter system.

Because the conversion takes place between the rectifiers/batteries and the load, the secondary voltage will still receive all the backup support of the primary voltage in the event of external AC service interruption—and only one set of backup batteries is required (figure 9.18).

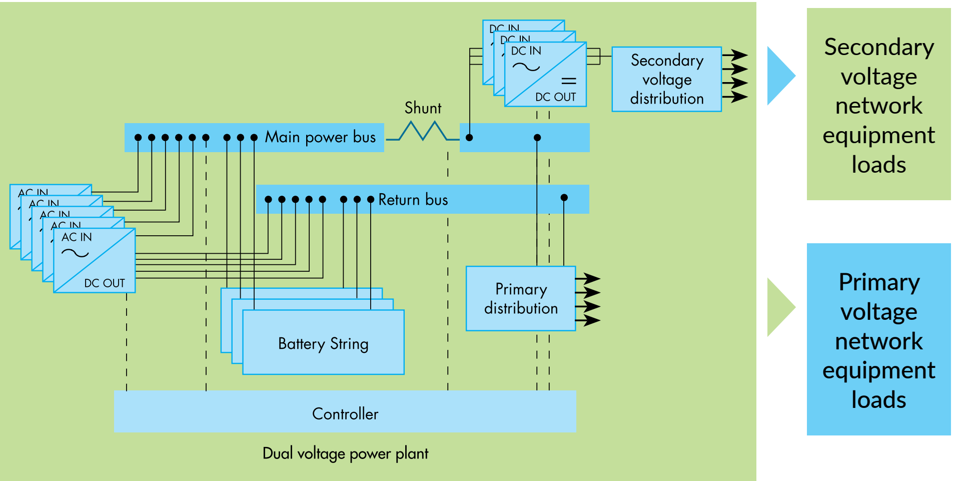
DC IN
DC IN
DC IN
DC OUT
Secondary voltage distribution
Shunt
Main power bus
Return bus
AC IN
AC IN
AC IN
AC IN
AC IN
DC OUT
Primary distribution
Battery String
Controller
Dual voltage power plant
Secondary voltage network equipment loads
Primary voltage network equipment loads

9.18: Supporting two voltages at a single site using one power plant and converters

Integrated power systems
Space-saving combinations of related components, built into a single device for easy installation.

Advantages of converting voltage

Modern DC-DC converters are essentially “plug and play” devices that are designed to fit in the racks alongside rectifiers and other converters (figure 9.19).

This approach offers communications providers the greatest flexibility in adopting next-generation technology, offering new services while maintaining older standards. As the deployment of the newer technology grows, additional converters can be added to the system to increase the secondary voltage’s share of the power supply. It should be noted, however, that doing so may require upgrading the rectifier capacity as well, while making sure to maintain the usual redundancy margin.

Disadvantages of converting voltage

On the downside, converting to a given voltage is inherently less efficient than drawing that voltage directly from the rectifiers, so losses increase as more and more DC power is converted away from the primary voltage.

Also, many traditional power plants cannot support the direct addition of converters to their systems. To draw two voltages in these instances, a separate converter plant must be installed. The converters draw power from the existing DC power plant via fuse or breaker positions in the distribution system, just like any other load device. This reduces the number of available distributions for the primary voltage equipment at the site.

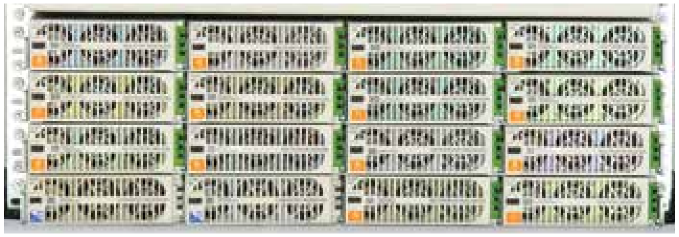

9.19: Converters and rectifiers occupying adjacent power shelf slots

Mapping the positions

Since a single power plant can generate varying amounts of both primary and secondary voltages, the need arises to assign numbers to the distribution positions of each voltage. A selectable voltage distribution panel makes this organization possible (figure 9.20).

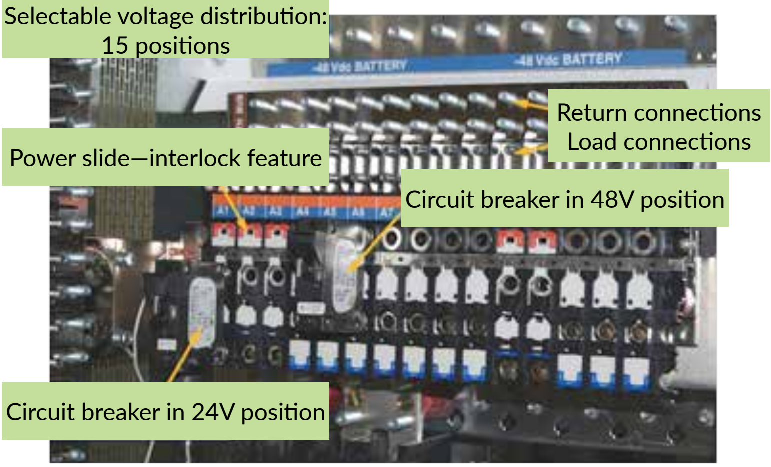
Selectable voltage distribution: 15 positions
-48 Vdc BATTERY
-48 Vdc BATTERY
Return connections
Load connections
Power slide—interlock feature
Circuit breaker in 48V position
Circuit breaker in 24V position

9.20: A selectable voltage distribution panel, showing both +24V and –48V applications

Integrated power systems

So far, we have focused on the individual components that comprise a cell site's power system. With so many components, you can imagine how quickly the space limitations of a site's shelter or cabinet become an obstacle.

To address these space limits, **CommScope** produces *integrated power systems* with several components built into a single device, and suited for installation in a single rack. This approach is increasingly common in modern cell sites.

A typical integrated cell site power system includes one or more shelves of rectifiers along with one or more shelves of DC-DC converters. This integrates power conversion and power distribution functions, connecting them with bus conductors. The distribution system contains an integrated DC bus, fuses or breakers and cabling tie-downs to distribute power to the load (figure 9.21).

Cell site power requirements range from 100 amps to several hundred amps of primary DC power. In +24V applications, these systems are configured to deliver 1,000 to 1,200 amps of primary power; −48V applications, about 600 amps. This may seem like a great deal of capacity, but the actual power consumption is typically far less than the system's theoretical capacity.

Since integrated power systems feature modular design, additional power can be added simply by plugging an additional rectifier into a vacant rack slot. Likewise, if there are secondary loads that use different voltages such as adding a +24V radio to a −48V system then DC-DC converters can be added to open slots to accommodate those needs. This is particularly useful in an age where radio technologies evolve at a rapid pace.

DC-AC inverter

Some of the equipment operating at a cell site may require AC current from battery backup supplies. Since the entire system is built around DC power, a DC-AC inverter is needed to provide the necessary AC voltage (figure 9.22).

There are two basic types of inverters:

> **Offline inverters** feature an AC input and an AC output with a standby DC line connection available. This is the type generally used in cell site applications.
>
> **Online inverters** feature a DC input and an AC output with an optional AC standby line available.

Like DC-DC converters, the input for a DC-AC inverter is supplied by the primary power plant. Like converters and rectifiers, inverters are often installed and configured for redundancy. A static switch maintains equalized voltage to the load by switching automatically between external AC power and the inverter's AC power. This switching is done instantaneously, assuring no interruption in operation.

Along with the static switch, many systems also include a maintenance bypass switching panel in an inverter installation, which allows an operator to power down an inverter for servicing or replacement without disturbing the system's load. During this power-down time, however, the system load is entirely dependent upon external AC power.

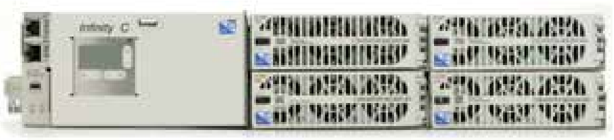

9.21: An integrated rectifier, DC distribution and controller power system

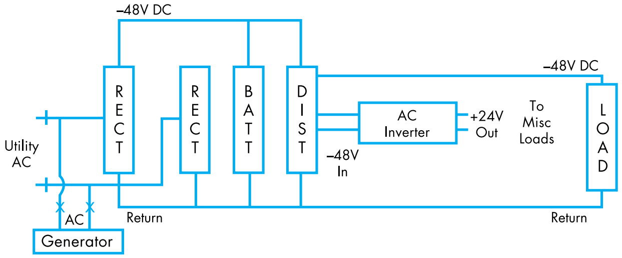

9.22: A DC-DC converter system connected in series to a cell site's power system

Remote radio head (RRH)
A recent design advance in base station architecture that separates a cell site base station's RF and baseband functions for improved efficiency.

AC power sourcing flexibility

With ever-increasing utility costs, the ability to combine power from renewable sources with utility power is another aspect of power flexibility. To address this, some rectifiers can accept AC power from attached solar panels or wind turbines just as easily as they can draw it from the utility's transmission lines. An illustration of this flexibility is shown below (figure 9.23).

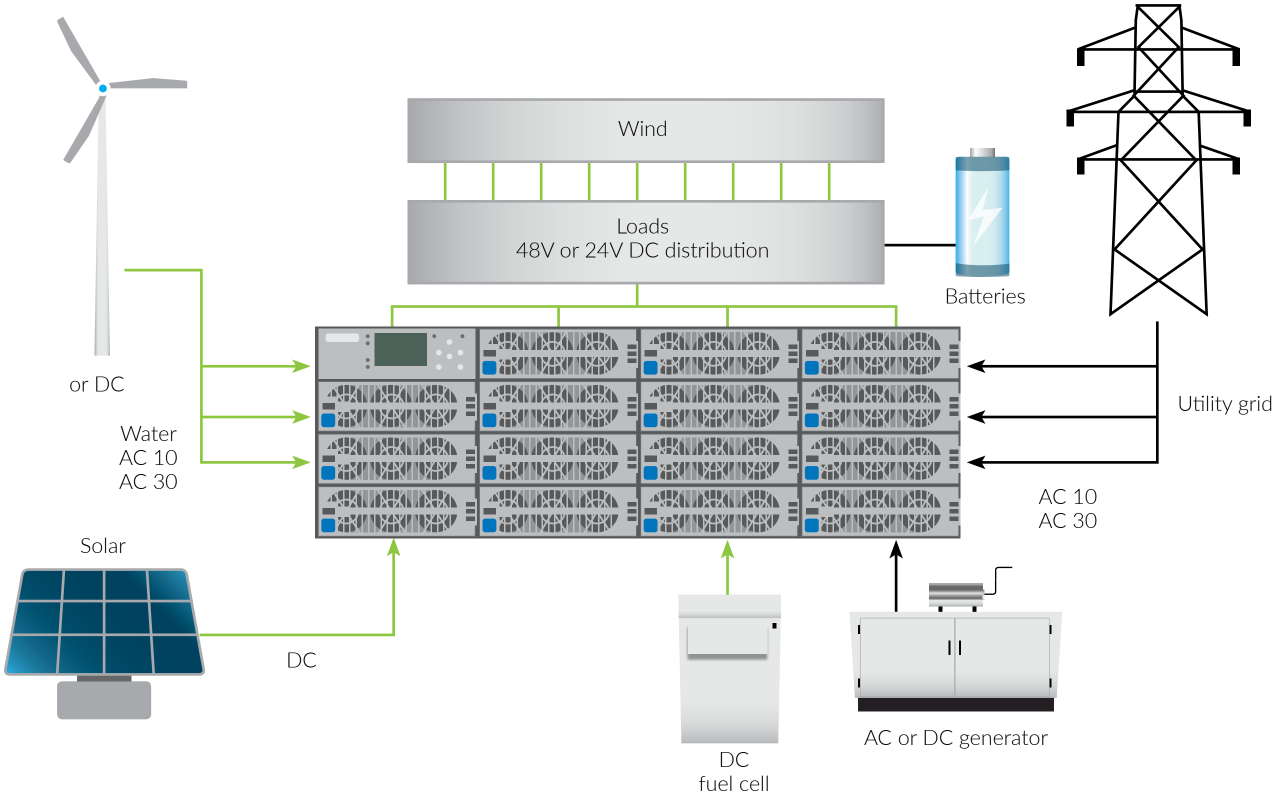
Wind
Loads
48V or 24V DC distribution
Batteries
Utility grid
or DC
Water
AC 10
AC 30
AC 10
AC 30
Solar
DC
DC
fuel cell
AC or DC generator

9.23: Power source flexibility lets rectifiers draw from conventional or renewable sources

Architectural improvements to power management

Significant power losses occur as a signal moves from the radio transmission equipment in a cell site's base station shelter to the antenna up on the tower. These losses are a natural consequence of traversing long stretches of coaxial cable. However, the simple architectural change of moving the transmitter and amplifier—known collectively as the radio head—from the shelter to the tower eliminates these losses and reduces power requirements. This design is called the remote radio head (RRH).

The baseband equipment remains on the ground and external AC power still enters at the cabinet or shelter. But with the transmitter amplifier mounted adjacent to the antenna on the tower, more space is freed up and less heat management is required.

Because of its exposed location, the tower top is not suitable for the battery backups, but other RRH equipment can be easily sealed against the elements. Instead of coaxial cable running up the tower, now only power transmission lines are needed. Below are two possible methods of providing power to the RRH (figure 9.24).

As useful as this design is, it also introduces new challenges. For instance, since the battery backup is now located far away from the critical components in the RRH, a heavier gauge of power transmission line is needed unless the power is converted to a higher voltage down at ground level and then converted back to the needed voltage (+24V or −48V) at the RRH itself.

This arrangement sometimes uses "line power" equipment, which is now available for these applications. With line power, the ground-level voltage is increased to +/−190 V DC, reducing the current and thereby reducing transmission loss (figure 9.25).

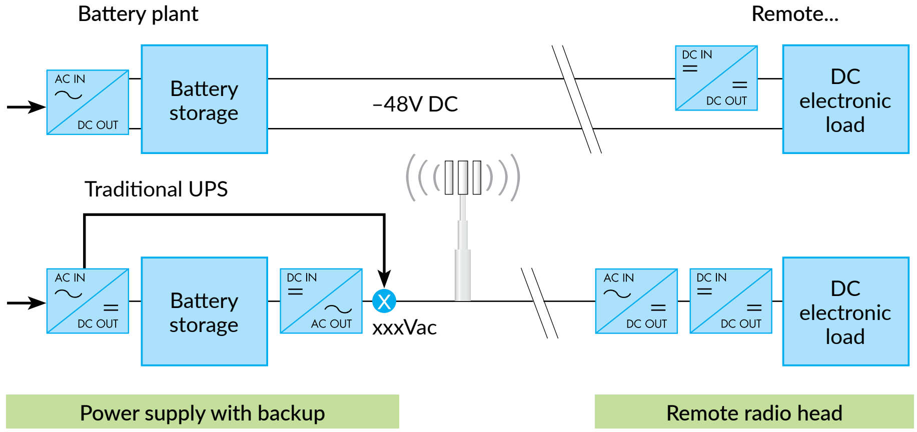
Battery plant
Remote...
AC IN
DC OUT
Battery storage
−48V DC
DC IN
DC OUT
DC electronic load
Traditional UPS
AC IN
DC OUT
Battery storage
DC IN
AC OUT
xxxVac
AC IN
DC OUT
DC IN
DC OUT
DC electronic load
Power supply with backup
Remote radio head

9.24: *DC powering options for remote radio heads (RRHs)*

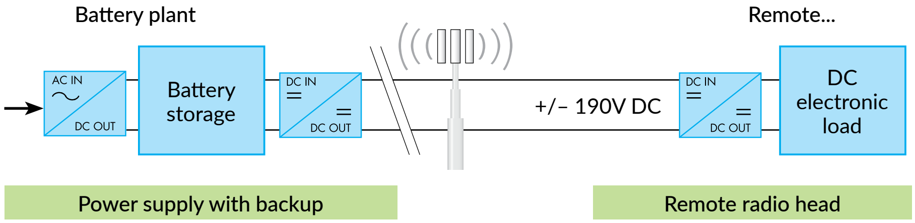
Battery plant
Remote...
AC IN
DC OUT
Battery storage
DC IN
DC OUT
+/− 190V DC
DC IN
DC OUT
DC electronic load
Power supply with backup
Remote radio head

9.25: *Line power voltage conversion used for an RRH application*

Cell site power system configurations

Different cell sites store their power equipment in different ways. Some sites have equipment shelters at the base of the cell tower. Inside these climate-controlled enclosures, equipment is mounted in equipment racks, with an integrated power system in one rack and battery strings in another, and radio equipment in still another.

Other sites place integrated compact equipment in outside plant (OSP) cabinets. This approach offers less space and fewer opportunities for environmental management, so only components specifically designed for OSP cabinet environments can be installed. It is common practice to keep power and radio systems in separate cabinets. Because of their size, often batteries will have a dedicated cabinet as well. A few examples of OSP cabinets are shown in figure 9.26 and 9.27.

Keeping the power and communication flowing

A lot of thought, planning and effort is required to deliver and support our always-on, 24/7 world of communications. When power outages occur, as they inevitably do, it falls to forward-thinking engineers and companies like **CommScope** to put effective, reliable contingencies in place.

Many of the same systems that keep your home electronics in good working order like battery backups, circuit breakers, DC converters and surge protectors also keep our national cellular backbone in a state of constant readiness.

Interruptions in power shouldn't mean interruptions in our lives, and with the technology and expertise of **CommScope** and our partners in the communications industry, they don't have to.

9.26: Integrated cell site battery only cabinet

9.27: Integrated cell site power and battery cabinet

Chapter 9 summary

Powering wireless networks

- DC vs. AC power
- Cell site equipment requires DC power
- Rectifiers convert AC to DC

Batteries

- Lead-acid batteries are available in flooded and valve-regulated versions
- Backup systems built of batteries stringed in series and in parallel

Generators

- On-site power generation
- Hydrogen fuel cells

Distribution

- Fuses, circuit breakers
- Surge protectors
- Shunts
- Battery and load disconnects

Integrated power systems

- Combine multiple elements in one device

New efficiencies

- Dual-voltage power plants
- Remote radio heads
- "Line power" voltage conversion

Chapter 10

Successfully planning against failure:
Reliability in wireless systems

It's simply a fact of life that items left out in the elements will become more susceptible to problems as a result of such exposure. Outdoor furniture ages more quickly than indoor furniture, the car parked at the curb shows more wear than the car kept in the garage. As a matter of necessity, a home's exterior paint will need refreshing more often than its interior. The elements, as a rule, are harsh.

Planning for environmental punishment is also a key concern for cell site operators, as new efficiencies that wring more work from every watt often mean placing components farther out into the network, and that means placing them outdoors, high on antenna towers. The same degrading effects that peel a house's paint work relentlessly against the sensitive electronics that drive modern cellular communications.

The precise balancing act of increased component failure rates against operational efficiencies has led to a revolution in how cell towers and cell systems are developed and built.

CommScope is at the forefront of this new network architecture and its impact on reliability. We offer the tools and expertise to help operators maximize redundancy, improve weatherizing and plan for system component failure and systems all over the world.

Reliability
The probability of a device working correctly over a defined length of time, operating under specified conditions.

Reliability, objectively defined

In everyday conversation, reliability usually refers to your experience with a device or service. It's somewhat subjective in that no two people have exactly the same experience, and they don't react in the same way when a problem does occur.

For instance, your home Internet service may occasionally go down for a few minutes in the middle of the night, but since you're not using the Internet when it happens, you would refer to your service as highly reliable even though it occasionally fails. Another person might need a stable connection during a particular time of day and experience an Internet outage during that critical moment. Even though the interruption may last only seconds, that person would describe his connection as unreliable.

In engineering terms, however, reliability takes on an entirely new meaning. It has an objective, empirical value based on study, analytics and calculation. In this sense, reliability is defined as the probability that a product or service will perform as it should at any given time, under specific conditions. In engineering, "failure" is something to be anticipated, controlled and compensated for via system redundancy.

Generally, reliability rates can be defined with a classic bathtub curve, so named for its bowl-like shape (figure 10.1). As you can see, it predicts higher failure rates at the beginning and ending of a device's life cycle. For any given device, this graph provides a general guide as to its rate of failure—decreasing initially, and increasing during end of life.

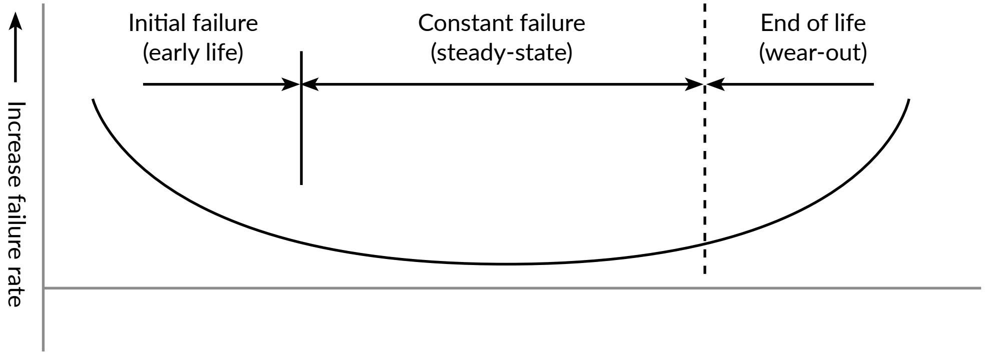

10.1: The reliability "bathtub curve," showing the failure rate over the operational life of a device

The three stages of component life introduce three common causes of failure:

1. **Early-life initial failure** may be due to manufacturing problems, incorrect installation or damage during shipping.
2. **Steady-state constant failure** indicates random failure as a normal function of operation. This is the stage we are most concerned with in this chapter.
3. **End-of-life wear-out failure** occurs when fatigue, corrosion or other factors accumulate to the point where failure becomes more and more likely.

Determining when a device's end-of-life stage occurs depends on complex computations. For example, one must take into account known life expectancy of the device's main components, such as motors and fans with moving parts that are eventually subject to mechanical wear.

Similarly, electronic components also have life expectancies. For example, electrolytic capacitors used in wireless electronics are subject to degradation from high temperatures and AC ripple currents.

Quantitative reliability predictions

As you might expect from a complex wireless communications system, predicting likelihood of failure is a complicated process, but a necessary one. There are several methods:

1. Collection of empirical field data from customers
2. Accelerated life testing data
3. Prediction models based on "parts count" method
4. System availability models for large systems with internal redundancy

Each method offers different advantages. The "parts count" method, for instance, is particularly useful for new product designs, even before the product moves beyond its design stages. This method applies established life cycle information for the components used in the design—the steady-state failure rates indicated in figure 10.2—to create an aggregated model of potential failures.

This computation relies on industry software reliability prediction tools such as Telcordia SR-332 Reliability Procedure for Electronic Equipment. It adds up individual component failure rates and applies designer-specified multipliers accounting for specific temperature, electrical stress, production quality and environmental conditions to yield a final, steady-state failure rate for the component. The various stress parameters are:

- Stress factor for operation is de-rated from specified limits
- Temperature factor (often adjusted up or down from a reference point of 40°C)
- Quality factor accounting for supplier and process controls
- Environmental factor accounting for indoor vs. outdoor conditions

From such prediction tools, a designer can compute a predicted failure rate with a 90 percent or greater confidence limit, which means at least a 9-out-of-10 chance that the actual failure rate will be no higher than predicted. These estimates usually reflect conservative numbers, making them highly reliable predictors.

Stress factors

Of the several stress parameters, perhaps the most critical factors in predicting reliability in the wireless communications industry are temperature and environmental stress.

Temperature factor, more specifically operating temperature, is the sum of the ambient temperature and the temperature of the heat produced by the component itself. In practice, a 10°C increase in operating temperature can double the likely failure rate. Likewise, reducing the temperature by a similar amount can reduce predicted failure rates by up to 50 percent.

Environmental factor is just as important as a predictive element. For example, an outdoor environment introduces a multiplier of 1.5 to 2.0, depending on the outdoor application. This factor accounts for variations in temperature, vibration and other environmental variables in an uncontrolled outdoor deployment versus the same equipment in a climate-controlled enclosure. Recent data suggest that a 1.5 factor is typical for outdoor wireless equipment such as tower-mounted antennas and remote radio head (RRH) equipment. Much of this data was collected by monitored RRHs.

Water ingress protection starts with the careful analysis of points of ingress, design considerations for protection of critical RF connection points, formulation of condensation and management of condensation. Addressing each element often means tradeoffs between cost and efficiency in the design.

The final product of the reliability prediction tool includes detailed, part-by-part information such as that shown on the right (table 10.2).

Part number	Category	Unit FR (FITs)	Quantity	90% CL Failure rate (FITs)	Ref des
7094037	Capacitor	0.21	1	0.26	C37
7131706	Resistor	0.57	15	9.42	R3, R44, R6, R20
7131748	Resistor	0.57	3	2.10	R150, R151, R152
7131797	Resistor	0.57	9	5.80	R1022, R243
7144735	Capacitor	0.21	10	2.24	C152, C162
7144739	Capacitor	0.21	2	0.49	C58, C740
7164258	Miscellaneous	3.80	5	25.82	AT1, AT2
7165048	Resistor	0.57	1	0.80	R126
7500917	Resistor	0.57	1	0.80	R23
7501483	Resistor	0.57	1	0.80	R135
7512949	Capacitor	0.21	1	0.26	C1053
7541771	Integrated circuit	6.02	2	16.66	U1007, U1011
7563383	Integrated circuit	6.02	1	9.31	U1017

10.2: Example of a component reliability table

Measuring reliability

As mentioned above, reliability is the probability that a device will perform correctly under defined operational conditions over a specific span of time. But supporting this general definition are several practical ways of measuring reliability in real-world applications.

Mean time between failure (MTBF) is the time between two consecutive failures. This is the most common definition for reliability. MTBF is expressed as the inverse of the failure rate.

MTBF = 1 / Failure rate = 109 / FITs.

Mean time to repair (MTTR), sometimes called mean time to restore, is the time needed to repair or replace a failed component and restore its function. This includes procurement and travel time, so the figure is comprehensive in its scope.

Availability (A) is the percentage of time the system as a whole operates normally. When someone refers to "4 nines of reliability," they mean "99.99% uptime" (table 10.3). Availability is a function not only of how often a component fails, but also how long it takes to restore service when it does fail, so it figures in both MTBF and MTTR.

A = MTBF / [MTBF + MTTR]

When describing the reliability of an entire system, availability is a more useful measurement. That's because redundancies built into the design may tolerate some individual failures without seriously compromising the function of the system as a whole.

Unavailability (U) is the flipside of availability, in that it expresses the percentage of time a system is not working properly. Also, like availability, it can be calculated as a function of MTBF and MTTR.

U = MTTR / [MTBF + MTTR]

As you may expect, A + U always equals 100%.

Downtime (DT) is derived from unavailability and expressed as the average amount of time per year the system will be in an unavailable state. Since it measures failures that often last only minutes, and expresses them as a percentage of a full year, we simply multiply the unavailability percentage by the number of minutes in a year, or 525,600.

DT = U x 525,600

To illustrate, consider a system with 99.99% availability, or "4 nines." That means its mathematical value is .9999, resulting in a U value of .0001. Multiplying .0001 by 525,600 yields an expected annual downtime of 52.5 minutes per year, or less than a full hour. This is a key indicator in overall service quality (table 10.3).

Percent availability	Number of nines	Downtime (minutes/ year)	Service quality level
99%	2-Nines	5,000 m/y	Moderate
99.9%	3-Nines	500 m/y	Well managed
99.99%	4-Nines	50 m/y	High availability
99.999%	5-Nines	5 m/y	Very high availability

10.3: Service quality measured by uptime; more nines mean more availability

Failures in time (FITs)
The number of expected component failures per billion operating hours.

Availability on a systemwide scale

Because of the vast number of components that make up a wireless communications system—each with their own failure rates, redundancies and importance to overall operation of that system—predicting long-term performance of the system as a whole can be quite a challenge.

From a design standpoint, cost constraints make it impossible to build components that are not subject to some degree of failure. Therefore, the design must incorporate redundant subsystems to avoid complete system unavailability due to any single component's failure. Properly implemented, these subsystems will allow the system to continue functioning at full performance, even in its degraded state, until repairs can be economically performed on the failed component.

Budgeting for failure

Dividing a system's functionality into subsystems allows a "reliability budget" to emerge. By breaking down the complex whole into manageable segments, subsystem reliability can be more easily modeled based on its parts count or by another appropriate method. Then, the likelihood of failure of the entire subsystem can be modeled to learn its effect on the overall function of the system itself.

In a practical example, consider the mass of electronic components mounted atop a cell site tower. The system can be broken down into subsystems involved in the transmit path, the receive path, the power system and other related functions. Each of these will have a maximum allowable failure rate assigned based on its importance to the operation of the system. So the site's designers can plan accordingly, the overall reliability budget is split up and allocated where it is needed most.

In practical applications, these models are more concerned with the functioning state of a system or subsystem, rather than with the actual hardware or software itself. At this level of planning, MTBF and MTTR become more meaningful descriptors of reliability than failure rate alone.

Failure mode, effects and analysis (FMEA)

Given enough time, component failure is a certainty. Where and when it occurs, however, is a variable that must be modeled to be predicted. That means a lot of what-if scenarios, not only of component failure but the effect of that failure on its subsystem and the effect of that subsystem on the system as a whole. For this kind of analysis, failure mode, effects and analysis (FMEA) is a simple, table-based method of measuring these variables together.

The FMEA for a particular system lists each failure mode and its effect on overall system performance. Failures that result in total loss of service are combined to calculate the system's total availability, while failures that cause only minor effects on service are combined to calculate the system's partial availability.

Fault-tolerant design

Redundancy is the key to boosting availability within a system without requiring the subsystems to be more reliable themselves. Redundancy schemes vary by application, but they all have one thing in common: on-demand access to a device or service that can assume the function of a failed device until it can be repaired or replaced. Sometimes this includes a spare component; other times it means shifting load to other systems. Common redundancy schemes include:

- **Active/hot standby,** a spare component built into the system that operates all the time and can assume more load when needed due to primary component failure
- **Active/cold standby,** a spare component built into the system, which only comes online in the event of failure, with a possible interruption in service
- **1 + 1 load sharing,** providing two active routes for communications so one will be available in the event that component failure causes an outage in the other
- **N+1 load sharing,** providing a standby alternate route for communications to assume the load in the event that component failure causes an outage in any of the other routes

The downside of adding redundancy is increased cost. Therefore, redundancy should be added in circumstances where benefits outweigh costs.

The reliability block diagrams

Another widely used approach to measuring system reliability is the reliability block diagram (RBD). Unlike the lists of the FMEA table, it graphically shows the interconnections between subsystems on a conceptual level and how redundancy measures are integrated. Also, unlike the FMEA table, these element "blocks" are described purely by function, not by individual component; the system's reliability depends on how these blocks are connected. Arrows represent the direction of information flow, but may not necessarily correspond to the physical direction of current in the system.

How the RBD shapes up depends on the kind of system architecture under consideration. A typical architecture may include both redundant and non-redundant subsystems, as shown in figure 10.4.

An RBD is extremely useful in predicting system reliability, but it does have disadvantages. The main limitation is its static nature: it can only predict individual failures without accounting for cascading effects throughout the system as it continues to operate in a degraded state.

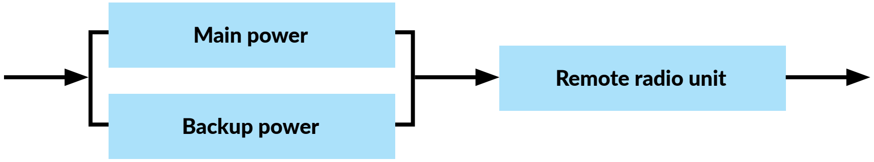

10.4: A simple RBD showing redundant power and non-redundant radio unit

State transition diagram (Markov Model)

In non-redundant systems, there are two states of being: working and not working. Transitions between these two states are defined by failure rates (1/MTBF) and repair rates (1/MTTR). Between these two measurements, availability can be easily determined.

However, more complex and more fault-tolerant systems have many levels of operational efficiency. We have referred to systems operating in degraded states, or a state of partial failure. To measure the reliability of these complex systems, the Markov Model defines all possible degrees of a system's operation and maps every state transition involved in making those states occur.

To illustrate, consider a simple system with two subsystems, A and B, each with the same failure rate (figure 10.5). As you can see, there are three possible operating states: fully operational, partially degraded and complete unavailability.

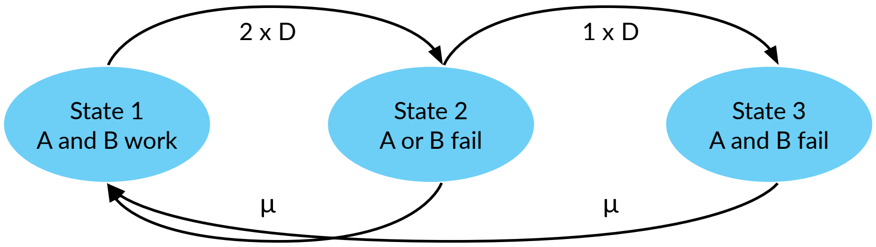

10.5: A basic Markov Model describing the three possible states for a two-subsystem design

The arrows indicate potential failure and repair transitions between states with the failure and repair rates for each. Different failure rates among subsystems naturally introduce additional variables, but the computations remain the same.

Markov Models can account for multiple combinations of failure conditions and the effect each has on system performance. This offers a better view of the comparative severity of different subsystem failures and what kind of degraded performance can be expected.

Markov Models are very useful in calculating the cost/benefit analysis of steps designed to reduce failure rates at various places within a system—essentially putting a "time and trouble" cost on any possible subsystem failure, which is particularly valuable when considering how difficult it is to service tower-mounted wireless communications equipment. It can also inform design decisions at the planning stage, taking into account accessibility factors early in the process. The downside to the Markov Model, however, is that it cannot assign a single MTBF value or failure rate to the system as a whole.

Reliability factors

Now that we are able to determine system reliability by multiple methods, we can examine what can be done to improve that reliability. These measures start in the design phase and carry through to installation and maintenance practices.

Product complexity is a primary factor to consider. Simply put, adding more components means adding more opportunities for failure. Take the integrated remote radio head (RRH) (figure 10.6).

With thousands of electronic parts built into a single device mounted in an outdoor environment, the predicted failure rate may be as high as 5 percent per year. This rate takes into account the ambient temperature, internal temperature and the RRH's heat dissipation features, as defined by Telcordia SR-322 software. Reducing the number of components in this or any device will directly increase its reliability.

10.6: A compact, tower-mounted remote radio head

Redundant hardware improves reliability by increasing the number of states in which the system may operate, adding flexibility to address service levels and repair schedules. However, redundant hardware will not reduce maintenance costs and it requires a greater up-front investment.

Heat dissipation is vital to the long-term reliability of any electronic system or subsystem, whether mounted atop a tower or located in a ground enclosure. For instance, a 60-watt amplifier may generate an internal temperature rise of 25°C to 30°C, which, when added to an ambient temperature of 25°C, yields up to 55°C of heat, exceeding the default operating temperature of 40°C found in the Telcordia SR-322. Recall that reducing operating temperature by just 10°C may reduce failure rates by 50 percent.

Thermal limitations relate to how heat dissipation is handled in a device. In the example of the RRH, front- and rear-mounted heat sinks arrayed in fins dissipate internal temperature rise into the air. The larger the fins, the more heat can be transmitted away. Limitations appear in the form of mounting orientation, available space on the mount and the physical size of the RRH itself.

Environmental issues are a key factor when dealing with electronics mounted outdoors atop a cell tower. Temperature variations, moisture, lightning strikes and other local conditions all play a part in how reliability is measured and improved. Each consideration should be thoroughly qualification tested and take into account prevailing industry standards.

Thermal design considerations

- Robust margins for thermal tolerance
- Design to conform to outdoor cabinet specifications
- Integrated thermal protection against over-thermal conditions

Mechanical considerations

- Resistance to high winds and vibrations on rigid mounting
- Accommodation of expansion and contraction
- Mechanical change-induced drift compensation

Atmospheric considerations

- Resistance to water infiltration
- Resistance to corrosion, fading and peeling
- Connectors, seals and gasket design
- Proper lightning mitigation (shielding and grounding)

Installation practices are just as important as design factors when it comes to ensuring reliability. As discussed in chapter two, it’s vitally important to work with a competent and experienced cell site services company with well-documented safety records and tower climb-certified technicians to handle both mechanical and electrical services.

Qualified technicians will reduce the chances of improper lightning protection, poor connections, mishandled feeder cable and weatherproofing problems. In the long run, maintenance and troubleshooting are much easier and less disruptive to your network when trained professionals handle your work.

Reliability testing

A number of reliability test programs are designed to improve product reliability from early design prototype to deployment. Such tests include the following.

Design verification testing (DVT)

Products are tested to electrical and mechanical specifications contained in their product specifications. Testing includes, but is not limited to:

	Temperature and humidity exposure	
	Antennas/RF components	**Outdoor cabinets**
Cold exposure	Test per IEC 60068-2-1 at -40°C for 24 hr.	Test to -40°C per GR-487-CORE.
Heat exposure	IEC 60068-2-2 at +70°C for 24 hr.	Test to +46°C per GR-487-CORE.
Temperature cycling	IEC 600068-2-14 cycling between -40°C and 70°C for 96 hrs.; dwell times at limits shall be 1 hr. past chamber equilibrium and transition times between limits shall be less than 2 hrs.	Test from -40°C to +50°C per GR-487-CORE.
Damp heat (humidity)	Test per IEC 600068-2-30, Test Db; one 24-hr. cycle; 9 hr. at 25°C / 95% RH and 9 hr. at 40°C / 90% RH with 3-hr. ramp transitions.	Perform 7-day temperature/humidity cycling per GR-487-CORE for outdoor cabinets.

	Corrosion (salt mist)	
	Antennas/RF components	**Outdoor cabinets**
Continuous exposure	IEC 60068-2-11 Test Ka (ASTM B117) for a minimum of 720 hrs at 35°C with a salt mist concentration of 5 wt% NaCl	720 hrs. per ASTM B117 and GR-487-CORE
Cyclic exposure	IEC 60068-2-52 Test Kb (Severity Level 4 – 2 cycles) at 35°C with a salt mist concentration of 5 wt% NaCl	

	Ingress protection	
	Antennas/RF components	**Outdoor cabinets**
Water ingress	IEC 529; for IP67 mated connectors: 0.5-hr. immersion under 1 meter of water (9.78 kPa at 25°C). (Unless sand & dust tests are required, physical ingress rating is inferred from design and water ingress rating).	
Sand and dust ingress	IEC 60068-2-68, or per Telcordia GR-487-CORE section 3.28.4. Enclosure door(s) opened and closed 50 times before exposure and dust collectors (consisting of 1.3 cm thick, 2.5 cm X 2.5 cm black conductive, high-density polyurethane foam pads) placed inside. 0.9kg (2 lbs) of simulated dust (consisting of 325 mesh white hydrated alumina silicate, or equivalent) is blown into the enclosure at minimum velocity of 27 m/sec (60 mph) for 1 hr. The foam collectors are then removed and visually inspected for dust accumulation.	Telcordia GR-487-CORE section 3.28.4. Enclosure door(s) opened and closed 50 times before exposure and dust collectors (consisting of 1.3 cm thick, 2.5 cm X 2.5 cm black conductive, high-density polyurethane foam pads) placed inside. 0.9kg (2 lbs) of simulated dust (consisting of 325 mesh white hydrated alumina silicate, or equivalent) is blown into the enclosure at minimum velocity of 27 m/sec (60 mph) for 1 hr. The foam collectors are then removed and visually inspected for dust accumulation.
Wind-driven rain	IEC 60068-2-18 Test Ra, Method 1, using a four-quadrant spray chamber simulating 40-mph, 70-in/hr rain for a minimum of 4 hrs.	70-mph wind and 5.8-in/hr rain per GR-487-CORE

10

Other tests include UV weathering effects from sun exposure:

- UV-A exposure with fluorescent lamps per IEC 60068-2-5, procedure B at 55°C for a minimum 240 hours
- Full spectrum UV-A/B exposure with xenon arc lamps per ASTM G155
- Other multi-year outdoor weathering tests in urban environments

Lightning protection:

- Test per IEC 61000-4-5, 1.2/50µs Voltage 8/20µs Current Combination Waveform, 10 repetitions @ ±6kV, ±3kA.

Some examples of testing and analysis in action are shown below (figures 10.7 through 10.10).

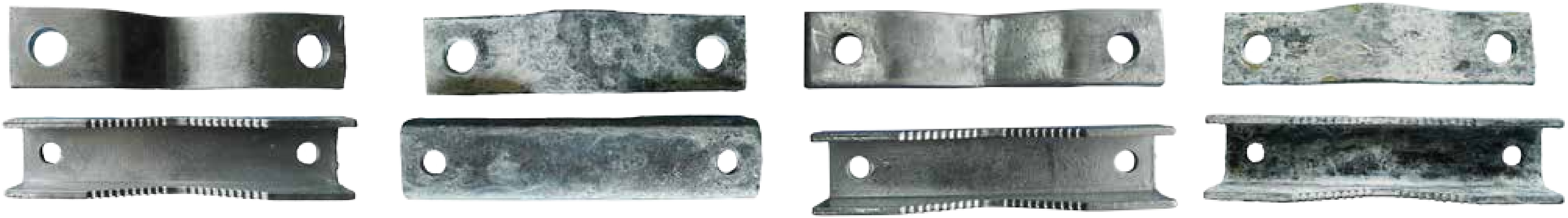

Group 1 *Group 1, 1000 hours of corrosion* *Group 2* *Group 2, 1000 hours of corrosion*

10.7: Weathering and corrosion test of silver-plated steel products

10.8: UV and weathering test of base station antenna assemblies

10.9: Outdoor wireless cabinet thermal testing

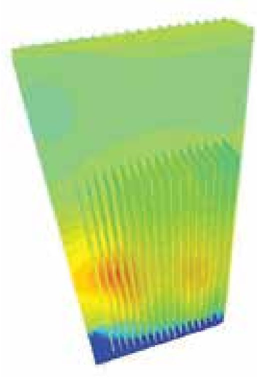

10.10: Remote radio head subassembly thermal simulation

Robustness and life testing

Accelerated life testing (ALT)

ALT is performed to demonstrate the long-term reliability of a product. During ALT, a sample of units is subjected to more severe thermal conditions than would normally be experienced in the field. This is implemented with a greater temperature excursion and an increased frequency of thermal cycles. Though product variable, a typical ALT includes high temperature dwells and thermal cycling to stress the unit's electrical components as well as their mechanical attachments. The duration of the testing is again product variable, but typical testing can last 60-100 days to evaluate the product's long-term reliability.

Highly accelerated life testing (HALT)

HALT reveals latent defects in design components and manufacturing that would not otherwise be found by conventional test methods. HALT stresses the products to failure in order to assess design robustness and marginality. Our regime includes:

- Step temperature stress
- Voltage stress
- Thermal dwell stress
- Rapid thermal cycling stress
- Random vibration stress
- Combined thermal cycling-vibration stress
- Other stress tests that may be product applicable

Recent developments in reliability

While it's easy to establish areas of attack when it comes to improving reliability turning those insights into real-life design advances requires a great deal of experimentation and testing. For every theoretical opportunity, there is a practical obstacle, but **CommScope** is dedicated to leading the charge for improved reliability on every available front.

Thermal design

In the arena of thermal design, **CommScope** is always working to develop new designs and alternative materials to make heat sinks that more effectively transfer heat from components to the air. This approach means less reliance on costly and failure-prone cooling fans. As mentioned earlier, every 10°C reduction in operating temperature doubles reliability.

Internal redundancy

CommScope is focused on load-sharing redundancy to improve system availability. While not every single subsystem can be built to ideal fault tolerances, we're making every effort to cover as many subsystems as possible.

Field data analysis

CommScope continuously monitors field returns and performs root-cause analyses on those returns. These analyses inform our design and manufacturing processes, so we can prevent potential problems at the source rather than on the tower—and field results prove that the process works, showing continuous improvement over time. We're building a whole new layer of reliability right into each product.

Industry forums

In 2008, the International Wireless Packaging Consortium convened the Tower Top Reliability Working Group to address carrier concerns over the reliability of tower-mounted equipment.

The best minds from carrier companies, equipment suppliers and other industry experts split into subgroups in order to draft a comprehensive best practices document. Results will be published at future IWPC proceedings and early results suggest that the study will become an ongoing fixture in the development of industry standards.

CommScope is proud to share our expertise, as we are well-represented in several key subgroups and lead the team dealing with reliability prediction.

Ensuring a reliable network

In wireless communications, every design choice involves a tradeoff. In exchange for more efficient use of power and space in cell site deployments, there exists a greater risk of component failure. Such failures are a part of life, but, they have to be part of the plan.

Predicting and measuring reliability can be a complex process with many competing aspects. Determining the reliability of a component, a subsystem or an entire cell site depends heavily on what matters most: maintenance time, upkeep costs, fault tolerance and a host of other considerations. There are ways to improve reliability, but the tradeoff in cost may not always be worth it.

In modern communications, there are no "one-size-fits-all solutions." Every step to improve reliability represents a careful balancing act between performance expectations, installation, and maintenance budgets and risk tolerance. **CommScope** helps make those decisions easier with the technology and insight that lets you choose the right solution from the best available options.

Chapter 10 summary

Reliability in wireless systems

- More tower-mounted equipment improves efficiency but poses challenges to reliability and maintenance.
- Reliability over life span defined by bathtub curve

Reliability stress factors

- Temperature extremes
- Environmental stress
- Heat and heat dissipation

Measurements of reliability

- Failure rate
- MTBF
- MTTR
- Availability
- Unavailability
- Downtime

Reliability prediction tools

- FMEA
- RBD
- Markov Model

Testing regimens

- Design verifications testing (DVT)
- Accelerated life testing (ALT)
- Highly accelerated life testing (HALT)

Reliability improvement opportunities

- Product simplification
- Redundant hardware
- Better heat dissipation
- Installation best practices
- Improved prediction tools
- Field data analysis and integration of findings into production processes
- Industry forum leadership

Chapter 11

Covering all the bases:
Distributed antenna systems

A population map of the United States illustrates what you already know: people are not evenly distributed across the country, or even across a particular state (figure 11.1).

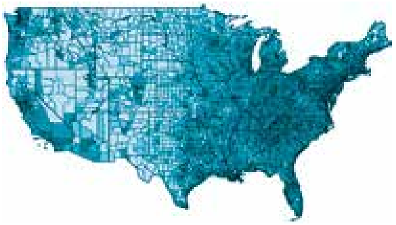

11.1: A map of the United States showing relative population density

Since cell network service is driven in large part by the population it serves, it makes sense that coverage must be denser in those places where population is denser. While this map shows how population density changes from state to state and from county to county, it doesn't show how density changes from one neighborhood to another, or even from one building to another. These localized differences matter in cell network planning, defining cell size, shape and power requirements. A key part of that plan is the *distributed antenna system (DAS)*, which serves geographical areas and buildings with the highest demands.

Distributed antenna system (DAS)
A network of nodes serving a specific place or building. They connect through a central base station for backhaul out to the public network.

A DAS is a network of spatially separated antenna nodes—called microcells—arranged to support cell network service in a particular place, often a single building or a campus of buildings (figure 11.2).

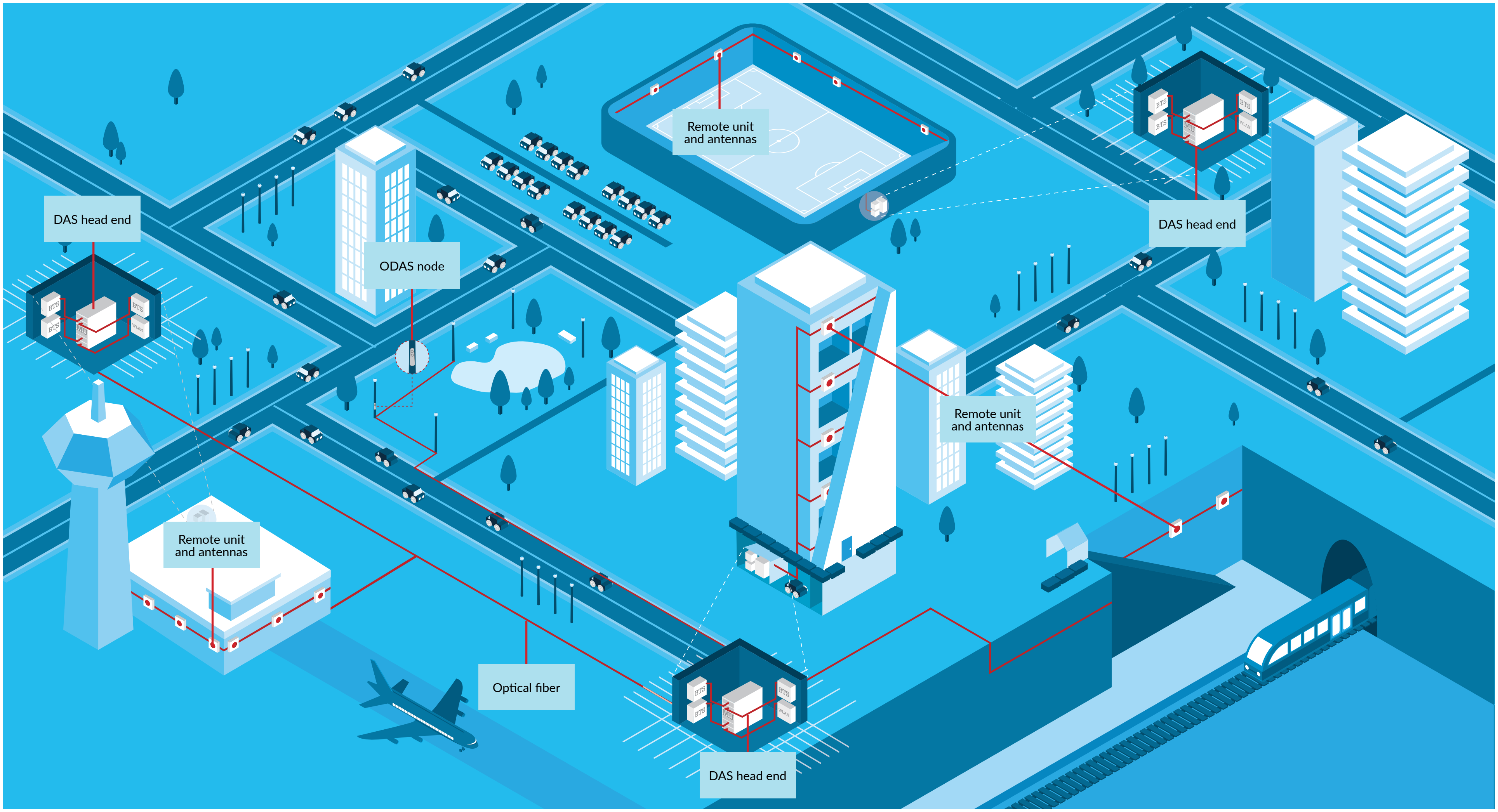
Remote unit and antennas
DAS head end
DAS head end
ODAS node
Remote unit and antennas
Remote unit and antennas
Optical fiber
DAS head end
BTS
RAU
WLAN

11.2: An overview of a typical DAS deployment

These nodes are compact, so they can be placed almost anywhere. DAS nodes typically have smaller footprints than macrocells. You can see a diagram of DAS architecture in figure 11.3.

DAS systems can be used to support an amazing variety of different communication and data network services, from analog to cutting-edge LTE and WiMAX (table 11.4).

This flexibility, combined with their small size and ability to cover small, specific areas, makes DAS an ideal means to extend or improve cellular service into locations that lack adequate coverage. Sometimes this need is a result of problematic location or geography, and; other times by limited network capacity. In either event, a DAS system can be a direct connection to the public network for locations unserviceable by macrocells.

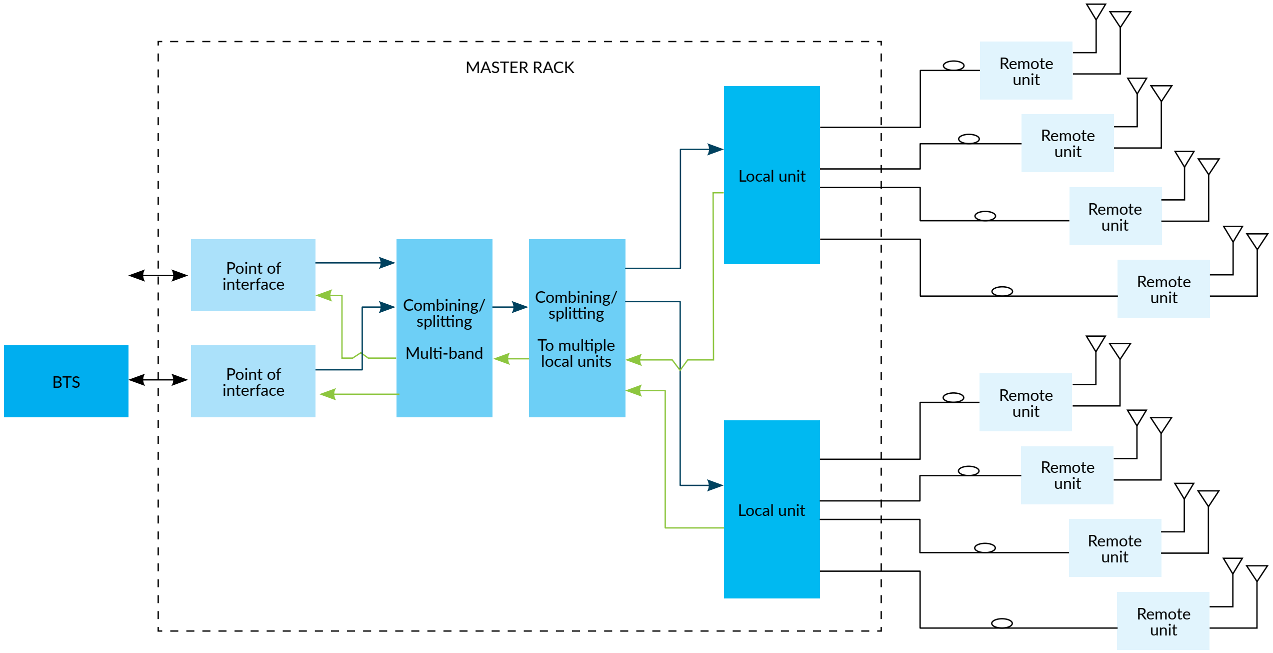
MASTER RACK
BTS
Point of interface
Point of interface
Combining/ splitting
Multi-band
Combining/ splitting
To multiple local units
Local unit
Local unit
Remote unit
Remote unit
Remote unit
Remote unit
Remote unit
Remote unit
Remote unit
Remote unit

11.3: A summary of DAS-supported frequencies and network services

Frequency range	Service types
700 MHz	Analog, GSM, LTE
800 – 850 MHz	Analog, GSM, iDEN, CDMA
900 MHz	Analog, iDEN
1700 MHz	CDMA, W-CDMA, LTE
1900 MHz	Analog, CDMA, W-CDMA
2600 MHz	WiMAX

11.4: A summary of DAS-supported frequencies and network services

DAS can be deployed indoors, outdoors and even in places that are a combination of the two, such as:

Indoor

- High-rise apartment or condo buildings
- Large corporate offices
- Exhibition halls and shopping centers
- Hotels, hospitals and restaurants

In these installations, nodes are arranged to provide even coverage across each area or each floor. An example of an indoor DAS layout appears below (figure 11.5).

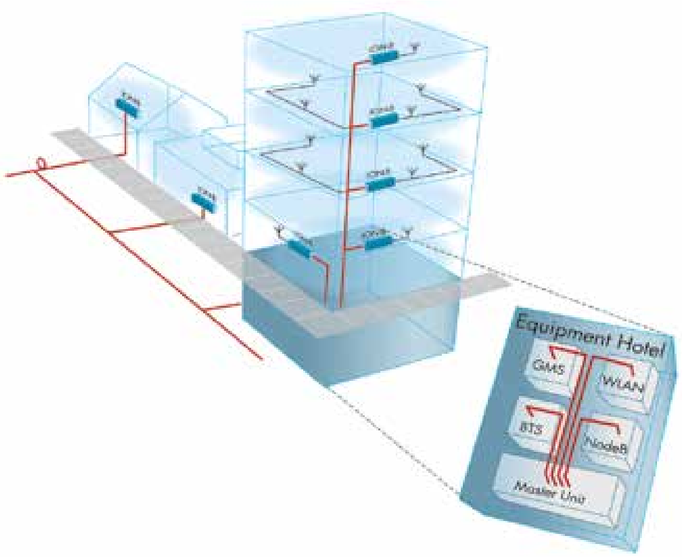

11.5: An indoor DAS layout for a building—remote nodes connect to the base station

Outdoor

- Open metropolitan areas
- Railways

Outdoor deployments create service areas in the open, but the architecture remains basically the same as an indoor DAS (figure 11.6).

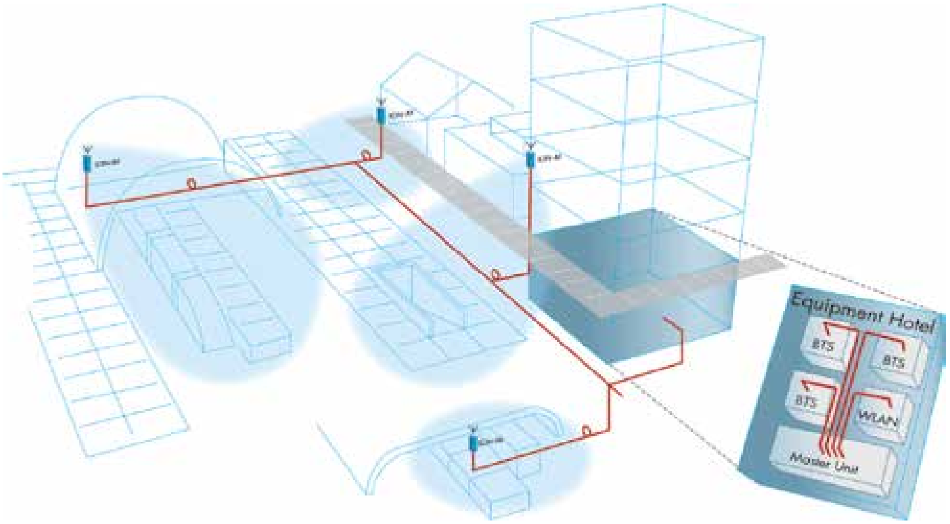

11.6: Outdoor DAS layout—nodes create service areas outside

Combination

- Corporate campuses
- Industrial parks
- Airports
- Stadiums
- Subways, tunnels and trains

Combination DAS layouts include elements of both indoor and outdoor designs.

DAS components

A DAS network layout is based on three core components:

1. **Master unit.** This is the DAS interface to the outside world, connected to a nearby macrocell base station that provides access to the larger network.
2. **DAS remotes.** These distributed devices receive traffic from the master unit via optical cable and relay RF transmission to and from the antenna via coaxial cable.
3. **DAS antennas.** These are the entry and exit points between the DAS and individual network users. They relay traffic to and from the remote.

Seen together, you'll notice that the entire system starts to resemble the topology of an ordinary computer Ethernet network, in which a central server connects to individual workstations as well as to any outside networks, such as the Internet (figure 11.7).

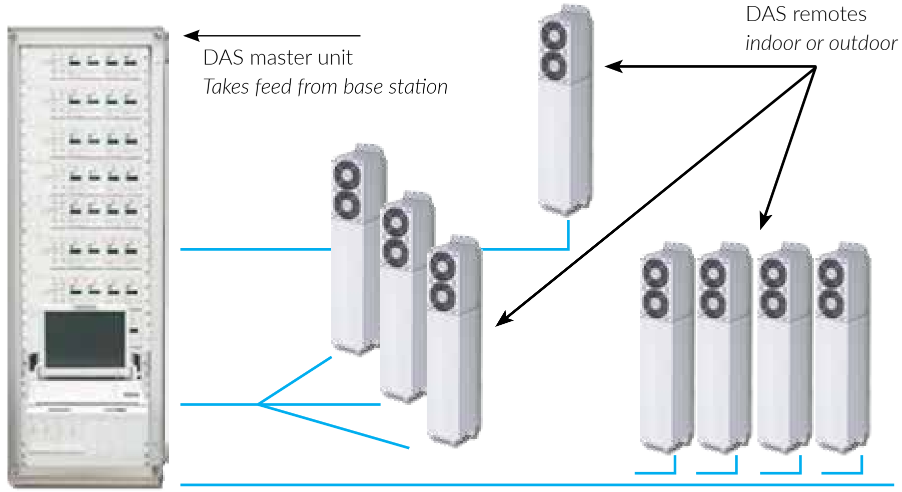

11.7: Typical DAS components—the antennas connect via coax to remotes

DAS in practice

Perhaps the best way to understand how a DAS can bring enhanced network service to a specific area is to look at a few real-world examples.

Indoor DAS deployment: Rome Telecom Italia Mobile

A corporate office building provides all the networking challenges DAS was designed for. In general, these environments are characterized by:

- **Large, dense populations of users.** The DAS must meet and maintain reliable service in high-traffic situations, whatever the circumstances.
- **Scaling needs for service.** The DAS must be able to offer seamless coverage from one node to the next and scale itself to meet the changing capacity requirements of users as they move through the building.
- **Ever-evolving service needs.** As technology changes and future-readying becomes a necessity, the DAS must be able to add or remove services without disrupting access or degrading performance, including those shown above in table 11.4.

A good example of a DAS that meets all these needs is the ION-B system offered by **CommScope**. The ION-B series can manage multiple technologies over several bands simultaneously. Depending on the situation, a designer may choose single-band, dual-band or triple-band versions with wireless LAN (WLAN) on an auxiliary channel.

Rome Telecom Italiaia Mobile (TIM) needed a reliable DAS solution that would link users at their headquarters.

CommScope delivered a DAS network that met all these requirements, with a slim public profile.

The layout of the DAS reveals the structure of data flow throughout the two buildings, including in hard-to-serve places like the two underground parking levels and stairwells (figure 11.8). The system was installed and optimized in just four weeks and featured four IP-based base transceiver stations to support the high traffic requirements between the DAS and the outside network.

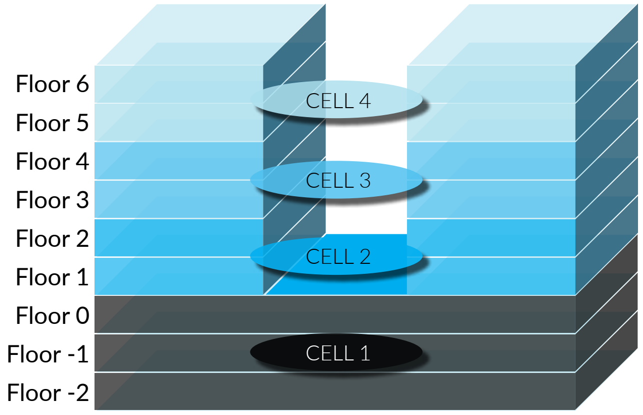

11.8: Layout of the TIM HQ system, connecting two buildings through a single DAS

Combination indoor/outdoor deployment:

Dallas/Fort Worth International Airport

Airports are high-traffic areas in more ways than one. In the dynamic world of air travel, expansion and evolution are inevitable. DAS serving an airport's personnel and its travelers must be able to expand and evolve as well. Like most indoor/outdoor installations, an airport DAS must offer:

- **Flexible service support.** With so many different technologies on the airwaves, an indoor/outdoor DAS must support the standards of today and tomorrow.
- **Broader coverage areas.** By definition, indoor/outdoor deployments tend to cover wider spaces and more diverse terrain. Expanding the DAS range cannot come at the cost of reliability or speed.
- **Steeper demands for on-the-fly scalability.** Wider coverage areas have greater potential traffic needs. For efficient operation, the DAS must be able to dynamically allocate capacity where and when it is needed.

Again, the ION-B® series from **CommScope** proved an effective solution for these challenges, as well as other specific circumstances related to our DAS installation at the Dallas/Fort Worth International Airport.

Dallas/Fort Worth International Airport:

- A busy cargo and passenger hub for multiple carriers
- 5 large terminals dispersed over an area of 18,000 acres
- More than 28,000 parking spaces in ramps and lots
- Over 56 million passengers and 630,000 flights annually

The DAS solution would need to:

- Cover vast spaces between terminals and parking areas
- Include capacity for public safety traffic
- Scale with airport growth
- Allow flexible leasing by airport management to multiple participating partners
- Be designed and installed quickly with minimal environmental impact
- Not interfere with other RF traffic in the area

The CommScope solution:

- Phase 1
  - 128 remote units
  - 218 antennas
  - 150,000 meters of fiber-optic cable
  - 30,000 meters of coaxial cable
- Phase 2
  - 58 remote units
  - 182 antennas
  - 5,500 meters of fiber-optic cable
  - 10,000 meters of coaxial cable
  - 10,000 meters of composite cable

Andrew integrated management and operating system (A.I.M.O.S.)
The network management solution for all **CommScope** distributed antenna systems. Part of our Andrew Solutions portfolio, it performs configuration fault and inventory via SNMP for alarm forwarding.

CommScope provided the right solution in a ION-B system with all the power and flexibility needed to meet the airport's present and anticipated future challenges.

To address the long distances between components, the optical links between remotes and master units were rated for signal integrity up to 20 kilometers. This meant remote units could be placed literally anywhere on the property and still communicate effectively with the master unit.

The ION-B system is also transparent over its full operating bandwidth (from 800 MHz to 2500 MHz) so each licensed user within the airport's hierarchy could count on top performance without interfering with other RF traffic in the area.

Cellular and other wireless signals from passengers, personnel and safety officials are distributed down hallways, into alcoves and through corridors. Intelligent network management also permits priority service for emergency responders.

To control maintenance and monitoring costs, the ION-B also allows remote, Web-based supervision by such common protocols as TCP/IP, SNMPv2, FTP or Telnet with its integrated *Andrew integrated management and operating system (A.I.M.O.S.).* To date, four wireless providers have leased access to the network.

Combination indoor/outdoor deployment:
Allianz Arena Munich

While the installation at Dallas-Ft. Worth International Airport (DFW) presented unique challenges in regard to area of coverage, stadiums and other indoor/outdoor applications require special attention to matters of variable capacity. While meeting all the requirements of the DFW example, a stadium installation must also deal with the additional task of assuring QoS in hard-to-serve indoor spaces as well as open-air spaces above. This is not only for the fans in the seats, but the enormous media presence required to broadcast the biggest events to a global audience.

For the Allianz Arena in Munich, **CommScope** deployed the ION-M series precisely because it could address these specific requirements.

Because the **CommScope** ION-M series can transmit on several bands simultaneously – with no restriction on the number of carriers – it was the natural choice for this application.

The combination of high-powered remotes and a flexible master unit allowed easy customization to penetrate even the hardest-to-serve locations. To scale along with the significance of the events taking place, ION-M also supports multiple configurations, like point-to-point, star and daisy-chain.

If fiber-optic cable access is limited, both coarse wavelength division multiplexing (CWDM) and dense wavelength division multiplexing (DWDM) can be used to establish multi-sectored sites with a single fiber connection. With all the maintenance and monitoring capabilities of A.I.M.O.S., the system offers cost-effective upkeep.

To date, four operators have signed contracts to use the system, which has proven equal to the task of keeping fans and the media connected – even for the most important tournaments.

Allianz Area:
- Home field for two German soccer teams
- Sitting and standing capacity of nearly 70,000 fans
- Hosted the globally broadcast FIFA World Cup in 2006
- Will play host to UEFA Championship Finals in 2012

The DAS solution would need to:
- Accommodate a full house of users without overloading the network
- Support multiple services such as GSM900, GSM 1800 and UMTS

The CommScope solution:
- 28 remote units
- 350 antennas
- 150 meters of radiating cable

Coarse/Dense Wavelength Division Multiplexing (CWDM/ DWDM)
Methods of duplexing signals in an optical cable by using different colors of laser light for increased capacity.

Swiss high-speed rail project:

- Route between Heitersberg and Däniken covers 20 km
- High-speed trains may travel up to 300 km/hr

The DAS solution would need to:

- Accommodate seamless handoffs between cells
- Manage calls through multiple operators
- Support GSM900, GSM1800 and UMTS services
- Cost significantly less than a macrocell solution

The CommScope solution:

- 24 remote units
- 48 antennas

Outdoor DAS deployment:

Swiss Rail Corridor from Heitersberg to Däniken

Full outdoor deployments place unique demands on the design of a DAS. It becomes even more challenging when the average network user is passing through your service area at more than 300 kilometers an hour, as they do on European bullet trains. For passengers to connect while they ride the rails, they need a DAS that offers:

- **Seamless coverage and capacity.** Handoffs between cells is tricky when the user is moving through at bullet train speeds, yet dropped calls are not an option.
- **Closed coverage.** Clean connections to the macro network keep the signal clear no matter what kind of interference sources the train is passing at the moment.
- **Remote administration and maintenance.** When the DAS is miles long, remote management capabilities are critical.
- **Multiple carrier support.** To keep development costs reasonable, the DAS must offer shared infrastructure to support multiple carriers simultaneously.

When the Swiss rail lines requested that **CommScope** build a DAS that would keep their passengers as connected as they were comfortable, even at the highest speeds, these stringent criteria put us to the test.

The ION-M system proved the best solution. By creating several radiation points along the length of the track, the cells comprising the DAS elongate and stretch in just the right direction to perform reliable handoffs, even at top speed. The coverage is dedicated and localized, keeping outside interference at bay and assuring enough capacity for the most crowded passenger trains.

All master unit equipment is concentrated in the BTS interface located in the middle of the track, while remote units are spaced along the length of the track. By elongating the cells, **CommScope** was able to achieve distances between remotes of up to 2 km, and with the integrated A.I.M.O.S. capability, system monitoring and adjustment can be performed remotely.

Outdoor DAS deployment:
The Moscow Metro underground

When the same technology is applied to subterranean settings with limited access, distances between remotes and master units become even more important. Both the ION-B and ION-M series guarantee signal quality for distances up to 3 km. By applying the optical link option, theoretical distances of up to 20 km are possible, helping future-ready the system against growing capacity needs.

As users and their associated network traffic move from one part of the DAS to another, the remotes must be powerful enough to manage dynamic changes in capacity while remaining small enough for safe installation in tight spaces found underground. In the case of the Moscow Underground, we find that reliable communications are much more than a matter of passenger convenience. They can make the difference in the event of an emergency situation as well.

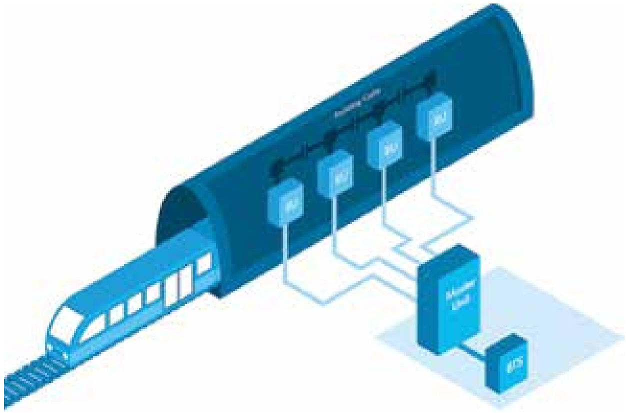

11.9: The ION-M deployment in the Moscow Metro underground

The Russian government ordered a DAS system that would provide seamless communication through the vast underground network, both for passenger convenience and for emergency responders, due to the threat of terrorist attacks. This capability was also needed for more commonplace emergencies such as fires, accidents and other injuries. As a government directive, installation time was of the essence.

CommScope designed and installed a dedicated ION-M DAS using remote units that could be installed without damage to the legendary décor of the Metro's elegant stations, which are often called "underground palaces" (figure 11.9).

Indoors, outdoors or both—the right solution makes the connection work

All these examples bring dedicated cellular services to locations that were not otherwise accessible, or at least not reliably accessible. A DAS network can achieve this primary objective by several methods, but the result is always the same: better, more reliable coverage, indoors and out.

Moscow Metro underground:

- 12 lines with 182 stations and a total rail length of 301 km
- 11 million riders daily
- A historically high-profile target for terror attacks

The DAS solution would need to:

- Increase security efficiency
- Provide highest reliability
- Offer fast access for emergency responders
- Present little or no visibility
- Be delivered on a very aggressive schedule

The CommScope solution:

- 260 remote units
- 10 master units
- 400 km of radiating cable

Capacity on demand, wherever it's needed

If the world's population were evenly distributed, designing the perfect cellular network would be as simple as drawing circles on a map. When the Swiss rail lines requested that CommScope build a DAS that would keep their passengers as connected as they were comfortable, even at the highest speeds, these stringent criteria put us to the test. Since the real world is full of high-density and low-density population centers, as well as variations in geography, government oversight and terrain, we need to adjust our methods. It's the only way people can connect…no matter where they are.

As more and more places are connected to accommodate our on-demand world, companies like **CommScope** continue to innovate new ways to improve service while reducing visibility, bringing access without eyesores.

Distributed antenna systems help service operators reach their customers in crowded buildings, deep underground and even as they cross the countryside at hundreds of miles an hour. They can connect a stadium of die-hard fans to their loved ones at home and let the whole world watch what's happening on the field below in real time—one of the many benefits of putting the right communication infrastructure to work.

Chapter 11 summary

Distributed antenna systems (DAS)

- Microcell solutions
- Compact, flexible and scalable
- Support for multiple services

Best circumstances for DAS

- Poor existing coverage
- Variable capacity requirements
- Exceptionally dense traffic areas
- No cost-effective macrocell solution

Indoor deployments

- Ideal for office buildings, shopping centers, apartments, expo halls, hotels and hospitals

Outdoor deployments

- Ideal for railways, metro areas and other outdoor spaces

Combo in/outdoor deployment

- Ideal for office campuses, airports, stadiums, subways and moving trains

DAS components

- Master units
- DAS remotes
- DAS antennas

Chapter 12

Going to ground:
Lightning protection

Even in the 21st century, the source of atmospheric lightning is the subject of scientific debate. Different theories assign different mechanisms to the creation of lightning: wind and friction, ice formation inside clouds—even the accumulation of charged particles from solar winds.

Far better understood is the behavior and power of lightning. We've all been cautioned not to stand out in the open during a lightning storm—and for good reason. A lightning bolt can reach temperatures of 54,000° Fahrenheit, five times the temperature of the sun's surface and hot enough to fuse loose sand into hard glass in an instant. Superheated air around the bolt expands violently as it passes, creating the familiar deep rumble of thunder.

Like any electrical discharge, lightning always seeks the path of least resistance to the ground. Often, this is through the tallest or most electrically conductive object available, which is why you don't want to stand in an open field during a storm. The human body presents the shortest path into the earth, boosting conductivity by shaving five or six feet off the distance a bolt must travel through the air.

Lightning by the numbers
In the 30 microseconds it exists, an average lightning bolt (at its peak discharge level) can carry:

- 30,000 amperes of current
- 1 trillion watts of electricity
- 500 megajoules of energy

But how do we deal with sensitive electronics that can't take shelter from the storm? One look at a cell antenna tower will tell you that, by virtue of its metallic composition as well as its height, it's a prime target for lightning strikes. A number of components are particularly attractive to lightning, including:

- Antennas and their support structures
- Coaxial lines and waveguides
- Steel buildings, cabinets and other equipment housing
- Connected communication and power lines

This exposure opens up the installation to expensive damage, maintenance and downtime, so it's vitally important that we take protective measures to minimize the risk of lightning damage.

Understanding the risks

Unfortunately for planners, most of the risk factors for lightning strikes are the same characteristics that make for a good cell site: open land and high elevation. Since there is little that can be done about location, lightning mitigation efforts must be directed elsewhere. Let's look first at the two types of meteorological events that present the greatest risks:

- Convection storms are caused by the heating of air near the ground and its interaction with cooler air above. These create the localized, short-lived storms we see most often in the summer months.
- Frontal storms are created by warm and cool fronts meeting. These storms can extend hundreds of miles and regenerate their strength over and over again, allowing them to persist for days and affect enormous areas. Frontal storms present the greater lightning risk.

The overall pattern of these storm types can be anticipated by season and location. Historical trends are accurate predictors of future activity. While government statistics don't include specific numbers regarding the number or severity of lightning strikes, they do provide overall counts of thunderstorm days in a given area (figure 12.1).

The nature of lightning

Lightning occurs in two common forms:

> **Cloud-to-cloud lightning** discharges itself by equalizing its charge with another cloud, remaining high above the ground in the process.
>
> **Cloud-to-ground** lightning seeks discharge through the earth. This is the kind that creates problems for objects on the ground, including cell sites.

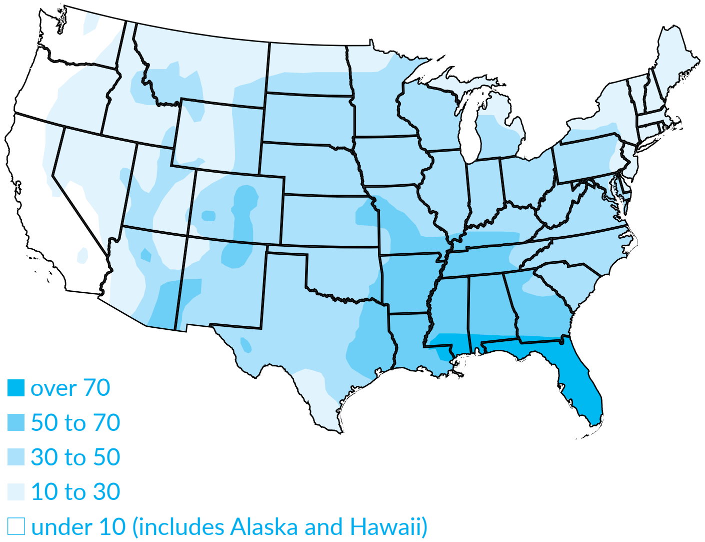

12.1: A meteorological map showing the annual number of thunderstorm days

In both cases, the lightning occurs when a difference in electrical charge—the *electrical potential*—exists. When this difference grows to a magnitude that overcomes the natural insulating properties of the air, the electrical difference seeks equilibrium by discharging itself along the path of least electrical resistance. For cloud-to-ground lightning, the less distance traveled in the air, the easier it is to discharge. That's why it seeks a more conductive object on the ground as its preferred path.

In most cases, this discharge represents a negative charge seeking a positive charge and may represent an electrical potential of as much as 100 million volts.

Electrical potential
Measured in volts, this is the difference in electrical charge between two points in space. The greater the difference, the higher the potential, and therefore, the greater the voltage.

Coaxial cable
A type of cable featuring an inner conductive core, an outer conductive layer and a dielectric, or insulating, space between them. Coaxial cable connects antennas to their base stations.

The birth of a bolt

As we can see in the three-step illustration below, cloud-to-ground lightning begins as a faint or invisible "pilot leader" high in the cloud. As it progresses downward, it establishes the first phase of the strike path. This pilot leader is followed by the "step leader," a surge in current following the new path. The step leader jumps in roughly 100-foot increments, or steps, until it approaches the positively charged point on the ground.

At this point, something incredible happens: a secondary discharge extends upwards from the ground, meeting the bolt in midair and completing the circuit. It happens so fast that the human eye only sees the bolt descending from the sky, not the one reaching up from the ground (figure 12.2).

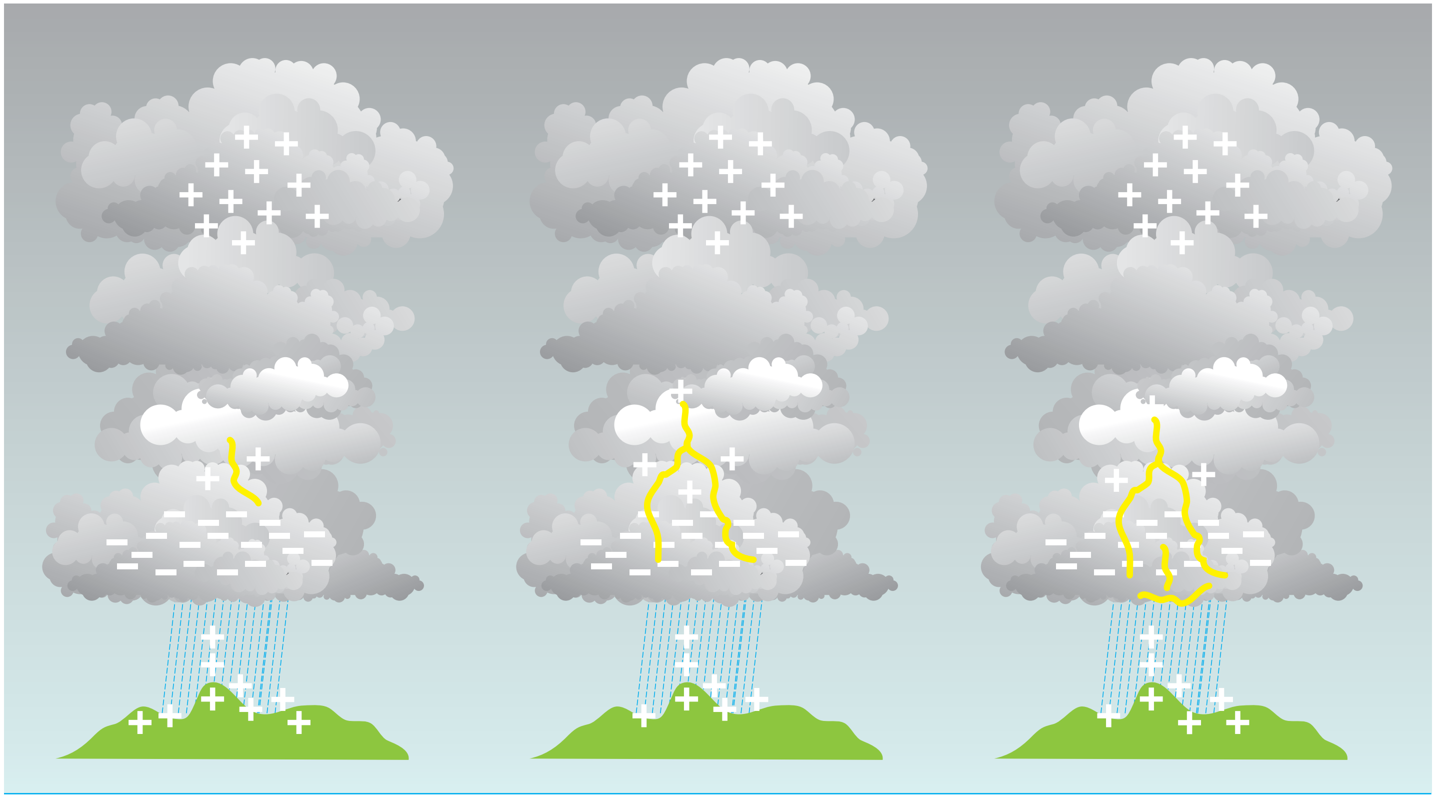

12.2: The origination and path of a cloud-to-ground strike

The intense light of a lightning bolt is created by molecules of air energized by the current passing through them. The shape of the visible lightning can help you identify its type:

Streak lightning is the most commonly seen type, characterized by a single line running from cloud to cloud or cloud to ground.

Forked lightning reveals the full conductive channel as smaller tributaries branching off the main line.

Sheet lightning is a shapeless, wide-area illumination commonly seen in cloud-to-cloud discharges.

Ribbon lightning is a streak that seems to repeat itself in a parallel path. This is due to high winds moving the air in the midst of the strike.

Beaded lightning, also called chain lightning, appears to break up into separate branches and persist longer than the main strike.

Heat lightning is not truly a lightning type, but the red-tinted appearance of other lightning types visible on a distant horizon. The coloration is due to atmospheric reflections and light scattering between the lightning and the observer.

Dealing with lightning

Now that we know a little about the challenges we face, let's look at some of the ways we can guard against the damaging effects of lightning.

The science of grounding

All electrical facilities are inherently connected to the ground, either by design or by circumstance. The earth itself represents the common electrical potential, or voltage, that other electrical sources naturally seek for equilibrium. By improving the way these discharges reach the earth, we can control the path and divert its damaging power away from equipment and structures that would otherwise be harmed.

When you imagine an electrical *grounding* system, you may have an image of a simple lightning rod with a wired connection to the ground. In the case of cell site installations, a grounding system is much more complex and serves purposes other than simply diverting lightning strikes; it also minimizes the chance of shock from the equipment itself, reduces noisy voltages that interfere with signals and protects sensitive electronics from damaging overvoltage conditions from all sources.

Grounding
Measures taken to control and facilitate the path of an electrical discharge from its source to the ground, avoiding potential damage to sensitive equipment along the way.

Grounding limits

As important as effective grounding is, it's often not enough by itself. Any path you install to ground a discharge has a certain physical limit to the voltage it can handle. Even the most substantial methods, like water pipes and specially designed grounding rods, are restricted as to how much voltage they can pass to the ground. To address these limits, it's wise to design in multiple paths so the grounding system can dissipate the most voltage possible.

Protecting the tower

The antenna tower presents the most obvious electrical target, as well as the best opportunity to protect the rest of the installation. That's because drawing lightning to the tower for safe discharge also gains us valuable insurance for the harder-to-protect components down on the ground and those connected by transmission lines.

By its very nature, a tall metallic tower can conduct lightning current into the ground. The danger arises when the voltage exceeds the structure's ability to dissipate it and electrical arcing occurs. This current can damage microwave antennas and, in particularly powerful strikes, fuse the dipole elements of two-way radio antennas.

To protect these and other components mounted on metallic towers, lightning rods should be affixed directly to the tower above the components to assure safe interception of the strike. It's also important to ensure that the tower's base, footings and any guy wires are also properly grounded.

Additional protective measures include insulating gaps built into the design and devices called shorting stubs that can be added to allow a short circuit at lightning's natural frequencies. We'll dig deeper into these measures on the following pages.

Wooden tower structures

In wooden structures, conductive paths must be added to give strikes a direct route to the ground. If wooden support poles provide a nonconductive obstacle from a metal tower, additional lightning rods should be affixed atop these poles to prevent them from splitting under the force of a strike and potentially collapsing the tower they support.

Rods for this kind of application are typically #6 AWG bare copper stapled to the pole on the side opposite the antenna's transmission line. This ground line should be connected to all equipment on top of the pole as well as any lines leading away to a connected shed or cabinet (figure 12.3).

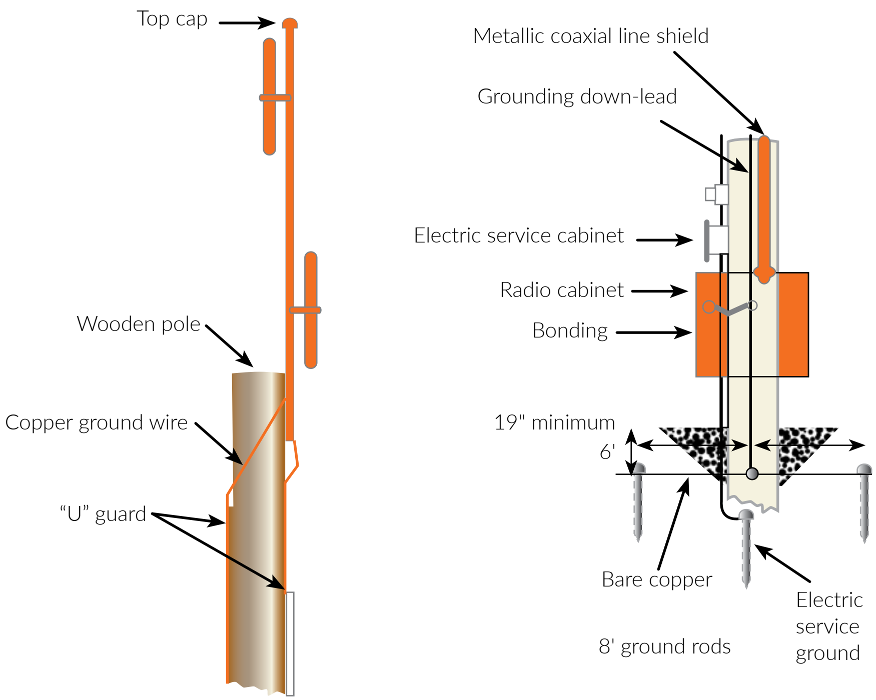

12.3: Grounding details for a wooden pole-mounted antenna

Two-way radio antennas

Conventional coaxial dipole antennas are often fitted with a serrated washer that forms a physical gap between the dipole whip and its support. Should lightning strike this point, the gaps in the serrated washer force the current to arc across the open space, creating a short circuit and dissipating the discharge.

Another protective measure is the insertion of a *quarter-wave shorting stub* in the coaxial transmission line at the base of the antenna. They're called quarter-wave shorting stubs because their place in the circuit does not impact normal operating frequencies of the site, and their length (a quarter of one wavelength), will cause an immediate short circuit for electrical frequencies associated with lightning. In this, they act as a sort of electrical "release valve" that will only divert a certain kind of dangerous current away from the system.

Other types of antennas tend to be self-protecting, such as folded dipoles, ground plane and Yagi antennas (chapter three for more information on different antenna configurations). These types are generally constructed of materials capable of handling most strikes, and their transmission lines are adequately shielded to direct any lightning current to the ground by other, easier paths.

Microwave antennas

Common types of microwave antennas, such as the paraboloid (dish-shaped) and horn reflector varieties are generally rugged enough to sustain normal lightning strikes without damage. However, the warning lights visible atop these installations are not so durable. To protect these regulatory-mandated devices, lightning rods are used to divert lightning discharges away from their more delicate wiring.

These protective systems may seem like a lot of expense to protect what are essentially blinking red lights, but the labor involved in replacing them after a lightning storm quickly becomes a costly maintenance situation.

Two-way radio antenna support structures

The buried end of a ground line can take several forms. Ideally, you would want the buried end to extend deep into the earth, providing a more reliable interface for dissipating the voltage. In some locations, such as rocky mountaintops, these depths aren't available. In these cases, the support structure can be protected by laying in multiple ground lines in a radial pattern to achieve horizontally what a single deep line would achieve vertically.

Quarter-wave shorting stub
A device inserted into the connection between transmission line and antenna that does not affect normal frequencies, but will immediately short and dissipate energy when lightning frequencies cross.

For antenna structures installed atop buildings—as is more frequently the case these days—all equipment, transmission lines and other conductive objects within a six-foot radius of the base should be commonly connected. The entire array should attach to a separate conductor with a minimum of two conductor systems integrated into the building itself. These conductors may be water pipes, steel framing or other electrically sturdy materials having direct contact with the ground below.

For those antenna structures mounted on metallic towers, grounding is a much simpler matter, since the tower structure itself provides a clear path to ground in the event of a lightning strike. Of course, this all depends on properly grounded base supports. Four methods of assuring this effect are shown to the right (figures 12.4 through 12.7).

For ground lines terminating through concrete bases and guy wire anchors, good conductive continuity inside the concrete itself should be virtually immune to any negative effects from lightning strikes. This is true for towers built on the ground as well as structures installed on top of buildings.

However, inadequate welds within the concrete may lead to electrical discontinuity, which can cause electrical arcing within the support…with potentially explosive results. For maximum safety, a secondary ground line is highly recommended.

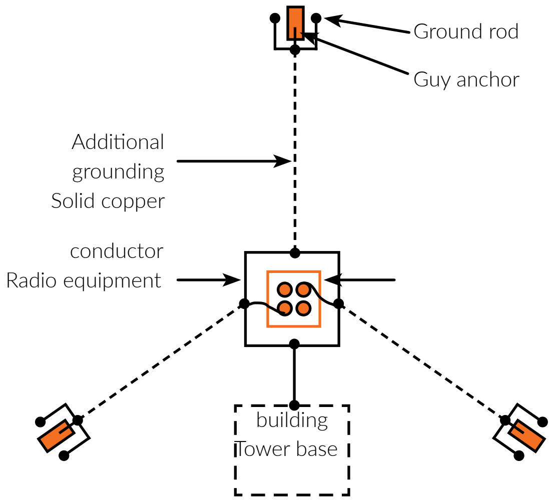
Ground rod
Guy anchor
Additional
grounding
Solid copper
conductor
Radio equipment
building
Tower base

12.4: *Support grounding installed in a tower pier*

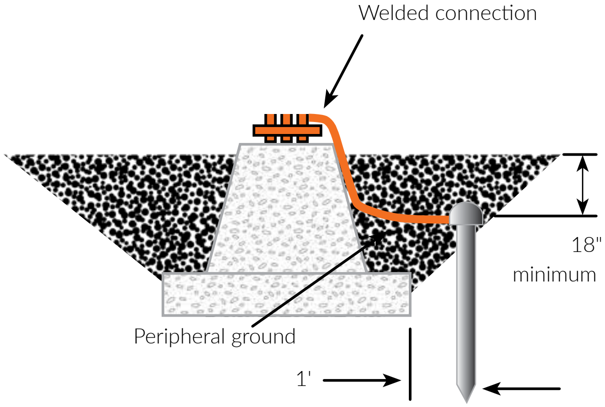
Welded connection
18"
minimum
Peripheral ground
1'

12.5: *Support grounding achieved via guy wires*

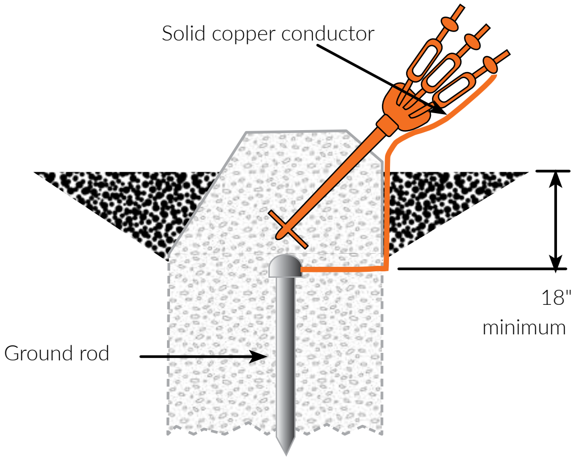
Solid copper conductor
18"
minimum
Ground rod

12.6: Support grounding installed in a tower's guy wire anchors

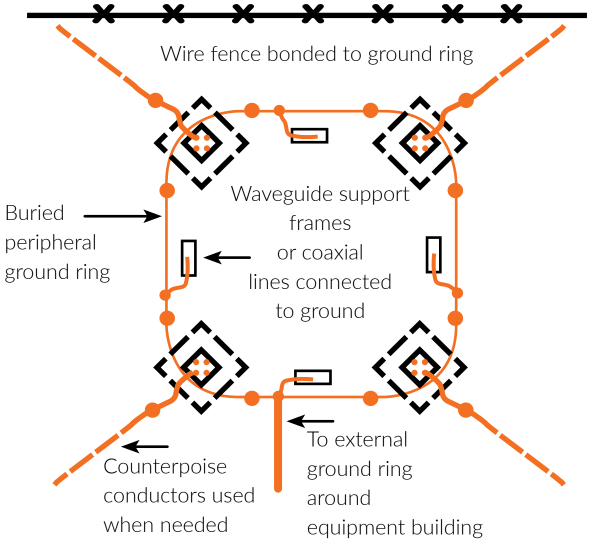
Wire fence bonded to ground ring
Buried
peripheral
ground ring
Waveguide support
frames
or coaxial
lines connected
to ground
To external
ground ring
around
equipment building
Counterpoise
conductors used
when needed

12.7: Support grounding for a freestanding tower structure

Coaxial transmission lines

Coaxial cables are subject to two potential hazards from surge currents reaching the outer conductor layer:

1. Damage to the insulating *dielectric layer* between the inner and outer conductors, which may destroy the cable and damage any equipment connected to it.
2. Mechanical crush forces that are associated with surge currents. While solid dielectric and larger cables (7/8" and up) are resistant to this effect, air dielectric and smaller-diameter cables have been known to be physically crushed by the magnetic effects of lightning-induced surges.

Providing a shunt path from the antenna to the ground line will usually prevent both kinds of damage. This danger highlights the rationale for securing the cables at such frequent intervals to prevent arcing between cable and tower.

Protecting the DC power system

As discussed in chapter 9, cell sites generally operate on DC power provided by the site's rectifiers. This energized part of the system presents an attractive target for lightning. The devices used to prevent overvoltage conditions are surge protectors. They work much like the consumer-grade version you may have installed in your home to protect sensitive electronics like computers or televisions.

Surge protector devices (SPDs) are in-line devices that feature a non-linear voltage-current characteristic which mitigates high voltages by increasing the associated current (figure 12.8).

12.8: Typical surge protection devices (SPDs)

Dielectric layer
The insulated, tube-shaped layer separating a coaxial cable's inner and outer conductors. Dielectrics may be made of solid material, flexible foam or open air channels supported by nonconductive spacers. If the dielectric becomes damaged, the cable will short.

SPDs are used to safeguard all cell site components connected to the DC power system, but any devices connected to an external metallic conductor will require separate protection.

The most vulnerable components connected to the DC power system are remote radio heads (RRHs), which we discussed in chapter nine. Mounted as they are atop the cell site's tower and adjacent to the antenna, the RRH and its copper power cable are natural targets for lightning, much like the power distribution equipment at the tower's base. Insertion of SPDs as close as possible to these locations protects the RRH units from overvoltage damage (figures 12.9 and 12.10).

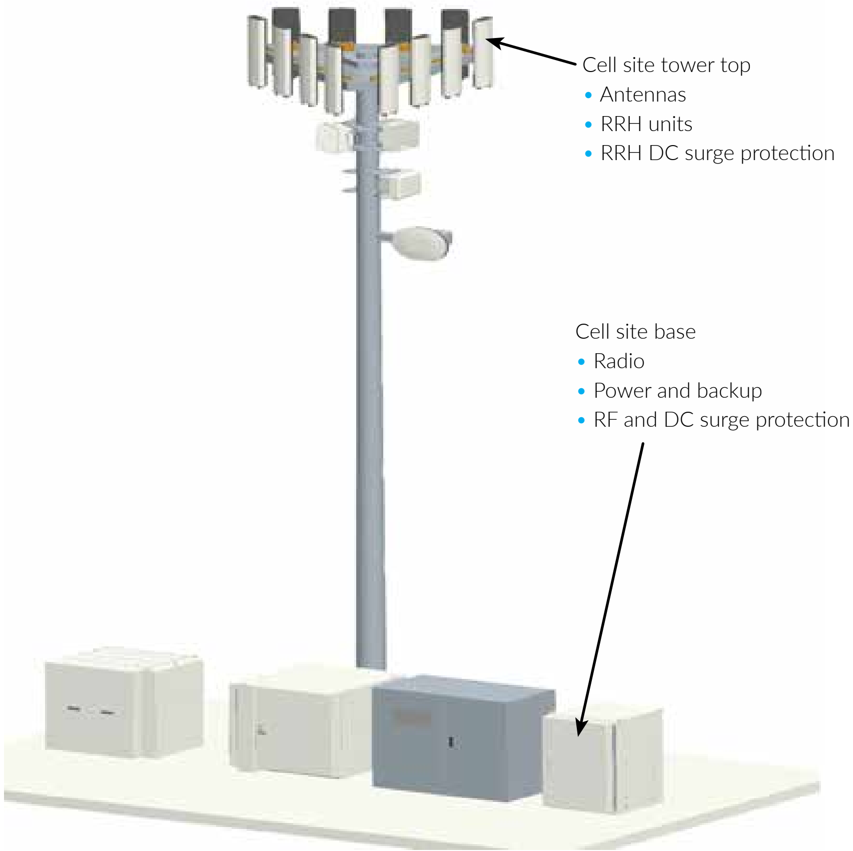
Cell site tower top
• Antennas
• RRH units
• RRH DC surge protection
Cell site base
• Radio
• Power and backup
• RF and DC surge protection

12.9: Vulnerable components at the top and base of a cell tower

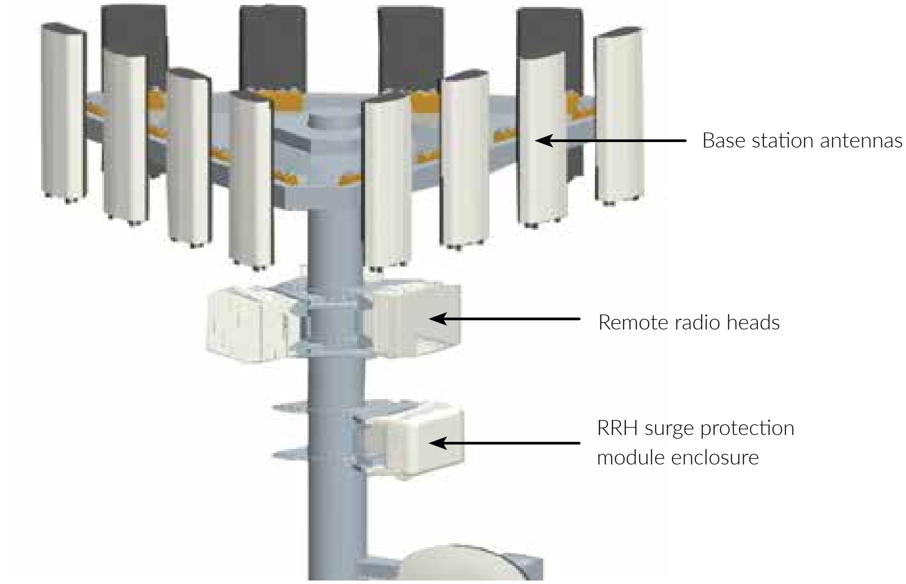
Base station antennas
Remote radio heads
RRH surge protection module enclosure

12.10: A surge protector mounted alongside the RRH

An SPD capable of withstanding not just one, but multiple lightning strikes must possess a robust resiliency to avoid costly maintenance and downtime for replacement. A detailed look at the inner layout of a sample RRH surge protection module appears below (figure 12.11).

Equipment at the base of the tower requires protection as well. Key components like the radio, the transmitter and backup battery systems, visible at the bottom of figure 12.9, are all vulnerable to damage from lightning-induced overvoltage. To prevent this, rack-mounted SPD units safeguard the power distribution system's connections between the RRH power cables and the rest of the equipment installed at the base (figures 12.12 and 12.13).

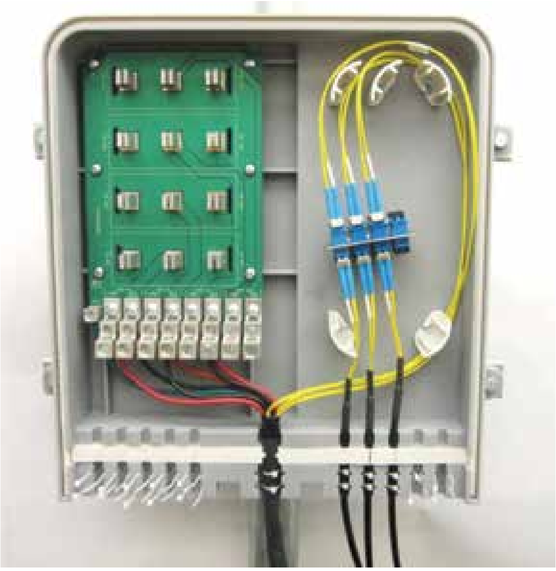

12.11: *Inside an RRH surge protector module, cover removed*

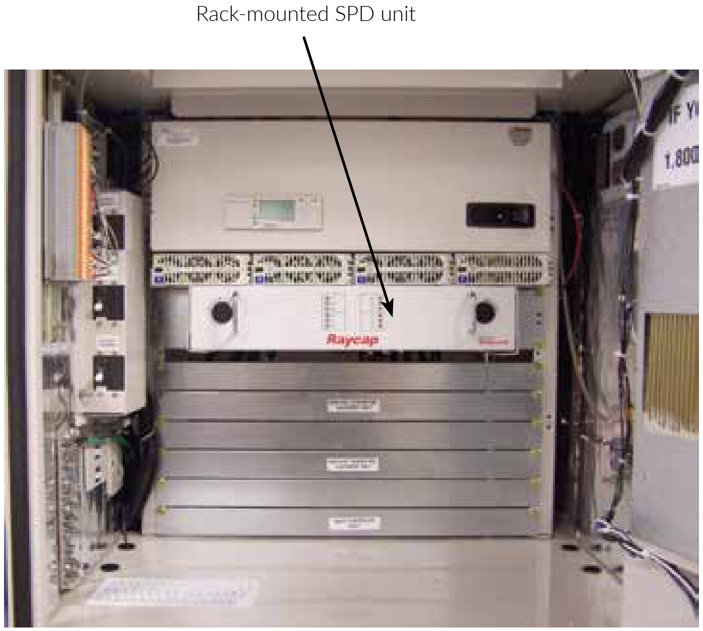

12.12: *An SPD assembly mounted in a DC power and battery enclosure*

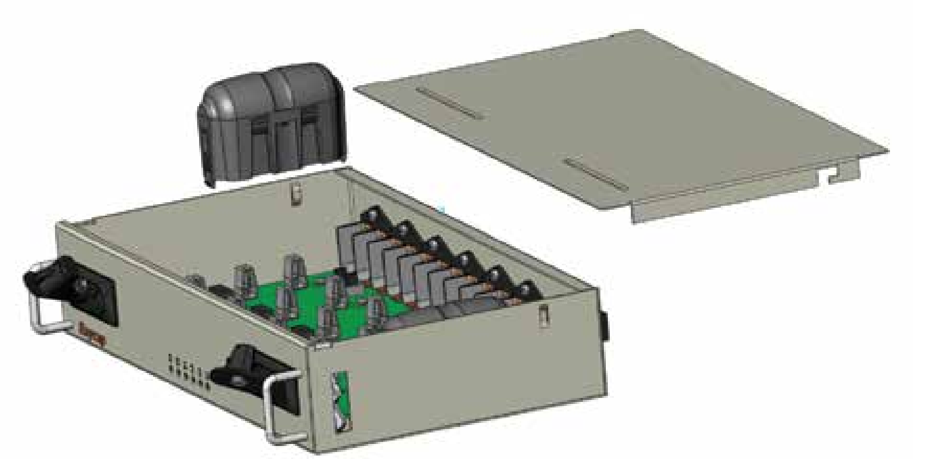

12.13: *The interior detail of an SPD unit designed to protect an RRH*

Other forms of overvoltage protection: fuses and breakers

SPDs serve functions other than simply protecting components from lightning strike-induced overvoltage situations. As we learned in chapter 9, SPDs also offer protection in the form of fuses and circuit breakers that keep more common forms of overvoltage from damaging components and batteries.

In a cell site's DC power distribution system, as many as 80 circuit breakers may regulate DC power from the rectifier to all the connected loads, both in the enclosure and on the tower (figure 12.14).

Both fuses and circuit breakers perform the same basic function, which is to interrupt power to the load when levels grow unsafe. Examples of each are shown below (figure 12.15).

- Fuses incorporate conductors designed to melt under the stress of overcurrent conditions, breaking the circuit and protecting the load on the other end
- Circuit breakers incorporate short-delay curve or "fast-blow" fuses to break the circuit and protect the load

Because of the large number of circuit breakers required—often from 24 to 80 per cell site—wireless carriers have adopted modular distribution structures that accept a wide variety of circuit breaker sizes. You can learn more about these features in chapter 9.

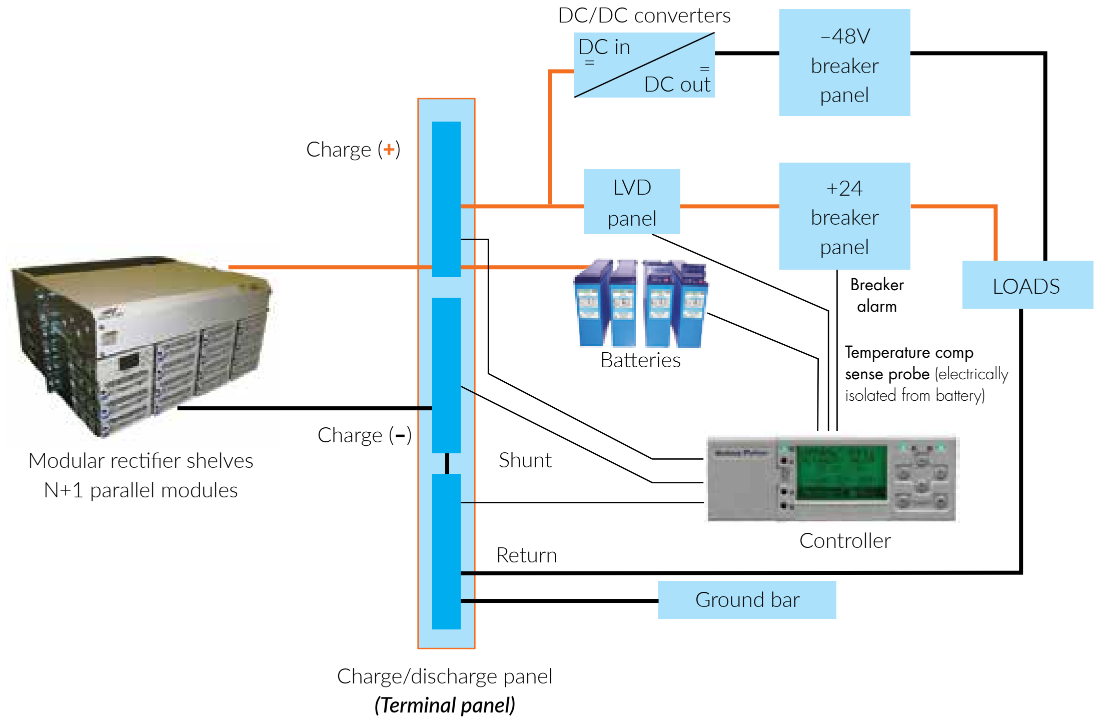
DC/DC converters
DC in
DC out
−48V breaker panel
Charge (+)
LVD panel
+24 breaker panel
LOADS
Batteries
Breaker alarm
Temperature comp sense probe (electrically isolated from battery)
Modular rectifier shelves
N+1 parallel modules
Charge (−)
Shunt
Controller
Return
Ground bar
Charge/discharge panel
(Terminal panel)

12.14: Block diagram of +24V power and distribution system, typical of a cell site installation

Images and illustration courtesy of GE.

GTM fuse TPL fuse/holder Plug-in circuit breakers

12.15: Commonly used telecom fuses and circuit breakers

Taming lightning by controlling the current

No matter how many times we witness it, the incredible power of a lightning bolt can instill fear and awe. When lightning strikes exposed cellular installations, only careful planning can divert its devastating power away from the sensitive components that keep modern networks operating.

With the right components, design and experience, a cell site can become virtually immune to the damaging effects of lightning as well as other electrically dangerous situations. **CommScope** supplies networks with the equipment and expertise required to harness lightning and direct it harmlessly away, keeping the world's key communications systems operating safely.

Chapter 12 summary

Lightning protection

- Lightning seeks a ground
- Providing safe paths around equipment protects cell sites

Lightning types

- Cloud-to-cloud
- Cloud-to-ground

Grounding an installation

- Metal towers can be self-grounding
- Wooden structures require paths
- Ground wire can travel deep or radiate horizontally

Tower and base station protection

- Remote radio heads (RRHs)
- Power distribution
- Radio, transmitter, battery backups

DC power system

- Surge protection devices
- Fuses and circuit breakers

Appendix A: Spectrum configurations around the world

			Applicable technologies				
Region	**Countries**	**Frequency bands**	**GSM**	**CDMA**	**UMTS**	**LTE**	**WIMAX**
Africa, Eurasia, South America	Argentina, Belarus, Czech Republic, Denmark, Georgia, Ireland, Latvia, Madagascar, Moldova, Norway, Peru, Portugal, Romania, Russia, Sweden, Tajikistan, Tanzania, Ukraine, Uzbekistan, Vietnam	450 MHz		✓			
Eurasia	Czech Republic, Hungary, Latvia, Mongolia, Romania, Russia	450 MHz				✓	
Africa, Eurasia	Cameroon, Kazakhstan, Moldova	450/800 MHz		✓			
Middle East	Iraq	450/800/1900 MHz				✓	
Asia, North America, South America	Canada, Kazakhstan, Peru, United States of America (USA)	700 MHz		✓			
Asia	Taiwan, Uzbekistan	700 MHz/2.6 GHz				✓	
North America	Canada, USA	AWS & 700 MHz				✓	
Europe	Germany	800 MHz				✓	
Africa, Asia, Caribbean, Central America, Middle East, Oceania, South America	Afghanistan, Angola, Australia, Bahamas, Bangladesh, Bermuda, Brazil, Cambodia, Cayman Islands, China, Colombia, Cote d'Ivoire, Ecuador, Fiji, Ghana, Guam, Hong Kong, India, Indonesia, Israel, Jamaica, Japan, Macau, Mongolia, Nepal, New Zealand, Nicaragua, Panama, Peru, South Africa, South Korea, Sudan, Taiwan, Thailand, Ukraine, Uzbekistan, Venezuela, Vietnam	800 MHz		✓			
Asia	Japan	800 MHz & 1.5 GHz				✓	
Europe	France, Germany, United Kingdom	800 MHz & 2.6 GHz				✓	
Europe	Germany	790 - 862 MHz				✓	
Caribbean, North America, Oceania, South America	Anguilla, Antigua and Barbuda, Argentina, Bolivia, Canada, Colombia, Ecuador, Guam, Honduras, Nicaragua, Northern Mariana Islands, Paraguay, Peru, Puerto Rico, USA, Virgin Islands (USA)	850 MHz/1900	✓				
North America, South America	Argentina, Canada, Colombia, El Salvador, Mexico, Nicaragua, Panama, Puerto Rico, Uruguay, Venezuela	850 MHz/1900			✓		
Asia, North America, Oceania, South America	Brazil, New Zealand, Pelephone, Philippines, Uruguay	850 MHz/2100 MHz			✓		
Caribbean, Central America, Europe, South America	Aruba, Belarus, British Virgin Islands, El Salvador, Honduras, Trinidad/Tobago, Ukraine, Venezuela	850 MHz	✓				

			Applicable technologies				
Region	**Countries**	**Frequency bands**	**GSM**	**CDMA**	**UMTS**	**LTE**	**WIMAX**
Africa, Asia-Pacific, Caribbean, Central America, Europe, Middle East, Oceania, South America	Andorra, Burkina Faso, Burundi, Cameroon, China, Comoros, Cook Islands, Cuba, Djibouti, East Timor, East Timor, Ethiopia, Falkland Islands, Faroe Islands, Fiji, French Polynesia, Ghana, Gibraltar, Greenland, Guyana, Iraq, Kazakhstan, Libya, Mali, Micronesia, Morocco, Myanmar, Niger, Norfolk Island, North Korea, Papua New Guinea, Rwanda, Saint Pierre and Miquelon, Samoa, Sao Tome and Principal, Saudi Arabia, Senegal, Solomon Islands, Swaziland, Togo, Tonga, Vanuatu, West Bank/Gaza Strip, Yemen, Zambia, Zimbabwe	900 MHz	✓				
Africa, Asia-Pacific, Caribbean, Europe, Middle East, Oceania	Afghanistan, Albania, Algeria, Angola, Armenia, Aruba, Australia, Austria, Azerbaijan, Bahrain, Bangladesh, Belarus, Belgium, Benin, Bhutan, Bosnia and Herzegovina, Bulgaria, Cameroon, Cape Verde, Central African Republic, Congo, Cote d'Ivoire, Croatia, Cyprus, Czech Republic, Denmark, Egypt, Eritrea, Estonia, Finland, France, French Guiana, French Westside, Gabon, Gambia, Gaza Strip/West Bank, Georgia, Germany, Greece, Guinea-Bissau, Guinea, Holy See, Hong Kong, Hungary, Iceland, India, Indonesia, Iran, Ireland, Isle of Man, Israel, Italy, Jamaica, Jersey, Jordan, Kenya, Kuwait, Kyrgyzstan, Laos, Latvia, Liberia, Liechtenstein, Lithuania, Luxembourg, Macau, Macau, Macedonia, Madagascar, Malawi, Malaysia, Maldives, Malta, Mayotte, Mongolia, Montenegro, Mozambique, Namibia, Nepal, Netherlands Antilles, Netherlands, New Zealand, Nigeria, Norway, Pakistan, Palau, Philippines, Poland, Portugal, Qatar, Reunion, Romania, Russia, San Marino, Serbia, Seychelles, Sierra Leone, Singapore, Slovakia, Slovenia, Somalia, South Africa, Spain, Sri Lanka, Sudan, Suriname, Sweden, Switzerland, Syria, Taiwan, Tajikistan, Tanzania, Tunisia, Turkey, Turkmenistan, Uganda, Ukraine, United Arab Emirates, United Kingdom, Uzbekistan, Vietnam	900 MHz/1800 MHz	✓			✓	
Asia, Caribbean	Aruba, Barbados, Dominican Republic, Thailand	900 MHz/1800 MHz/1900 MHz	✓				
Caribbean	Antigua and Barbuda, Dominica	900 MHz/850 MHz/1900 MHz	✓				
Caribbean, South America	Brazil, British Virgin Islands, Cayman Islands, Dominican Republic, Jamaica, Saint Kitts and Nevis, Saint Lucia, Turks and Caicos Islands	900 MHz/850 MHz/1900 MHz/1800 MHz	✓				
Island - British Channel	Guernsey	900 MHz/850 MHz/1800 MHz	✓				
Eurasia, Oceania	Armenia, Croatia, Denmark, Estonia, Finland, France, Hong Kong, Poland, New Zealand, Romania, Slovenia, Sweden, Australia	900 MHz/2100 MHz			✓		

			Applicable technologies				
Region	**Countries**	**Frequency bands**	**GSM**	**CDMA**	**UMTS**	**LTE**	**WIMAX**
Asia	Japan	1500 MHz				✓	
North America	USA	1.6 GHz				✓	
Asia	Japan	1.7 GHz				✓	
Asia, North America	Canada, Japan, USA	1700 MHz			✓		
North America	USA	1.7 GHz/2.1GHz				✓	
Europe	Poland	1800 MHz	✓				
Caribbean, North America, Oceania, South America	American Samoa, Bahamas, Belize, Bermuda, Brazil, Chile, El Salvador, Guatemala, Mexico	1900 MHz	✓				
Africa, Asia, Caribbean	Algeria, Argentina, Aruba, Barbados, Bermuda, Brazil, Cayman Islands, Chile, Dominican Republic, Ecuador, El Salvador, Guam, Guatemala, Jamaica, Philippines, Puerto Rico, Uruguay	1900 MHz		✓			
Asia, Caribbean, North America, South America	Barbados, Canada, Indonesia, Mexico, Trinidad/Tobago, USA	800 MHz/1900 MHz		✓			
Africa	Nigeria	cdma2000 1x/EV-DO		✓			
Africa, Caribbean, Eurasia, Middle East, Oceania, South America	Albania, Andorra, Angola, Armenia, Aruba, Australia, Austria, Azerbaijan, Bahrain, Bangladesh, Belarus, Belgium, Bosnia, Brazil, Brunei, Bulgaria, Cambodia, China, Croatia, Cyprus, Czech Republic, Denmark, Egypt, Estonia, Finland, Georgia, Germany, Ghana, Gibraltar, Greece, Guernsey, Hong Kong, Hungary, Iceland, India, Indonesia, Iraq, Ireland, Isle of Man, Israel, Italy, Japan, Jersey, Kenya, Kuwait, Laos, Latvia, Lauritius, Libya, Liechtenstein, Lithuania, Luxembourg, Macau, Macedonia, Malaysia, Malta, Moldova, Monaco, Montenegro, Morocco, Namibia, Nepal, Netherlands, Nigeria, Norway, Pakistan, Philippines, Poland, Portugal, Qatar, Romania, Russia, Saudi Arabia, Serbia, Seychelles, Singapore, Slovak Republic, Slovenia, South Africa, South Korea, Spain, Sri Lanka, Sudan, Sweden, Switzerland, Syria, Taiwan, Tajikistan, Tanzania, Thailand, Turkey, Uganda, Ukraine, United Arab Emirates, United Kingdom, Zimbabwe	2100 MHz			✓		
Europe	Germany	2.1 GHz/800 MHz/ 2.6 GHz/1800 MHz				✓	

			Applicable technologies				
Region	**Countries**	**Frequency bands**	**GSM**	**CDMA**	**UMTS**	**LTE**	**WIMAX**
Oceania	Australia	2.1 GHz/1800 MHz				✓	
Asia	India	2.3 GHz				✓	
Eurasia	Georgia, Indonesia, Malaysia, Russia, Tajikistan	2.3 GHz					✓
Eurasia	Russia	2.3 GHz/2.4 GHz				✓	
Asia	China	2.3 GHz/2.5 GHz				✓	
Europe	Denmark	2.5 GHz				✓	
Africa, Eurasia, North America	Angola, Canada, Russia, USA	2.5 GHz					✓
Africa, Central America, Eurasia, South America	South Africa, Argentina, Chile, Colombia, Hong Kong, Malaysia, Lithuania, Uzbekistan, Austria, Finland, Germany, Jersey, Netherlands, Sweden	2.6 GHz				✓	
Europe	Finland	2.6 GHz/1.8 GHz				✓	
Europe	Germany	2.6GHz/2.1GHz				✓	
Asia	Indonesia	3.3 GHz					✓
Europe	Germany	3.4 GHz-3.6 GHz					✓
Africa, Eurasia, North America, Oceania	Italy, Bulgaria, Canada, Colombia, Croatia, Dominican Republic, Ethiopia, Macedonia, Malta, The Netherlands, New Zealand, Nigeria, Russia, Sri Lanka	3.5 GHz					✓
Europe	Estonia	3.5 GHz – 3.6 GHz					✓
Oceania	Australia	5 GHz					✓
Asia, Caribbean, Middle East	Cayman Islands, Indonesia, Jamaica, Saudi Arabia	802.16d					✓
Africa, Caribbean, Eurasia, North America	Cayman Islands, Bangladesh, Canada, Indonesia, Libya, Russia	802.16e					✓

Biographies

Robert Cameron

Applications engineer

Robert is an invaluable resource to CommScope customers, providing insight and training on the latest solutions and how best to implement them. He is also responsible for providing background technical information and troubleshooting. He has spent the past 17 years in the RF industry, working in amplifier, antenna and RF system design. Robert's career began as part of an ambitious startup company, where he designed cellular network repeaters before moving into system design and project management to round out his expertise. He has managed several high-profile projects, including the two stadiums where the Kansas City's Chiefs and Royals play football and baseball, respectively, as well as in-building systems for the Library of Congress and the World Bank. He has also managed installations for the US military and several key airports and large banks. Robert holds a B.S. in Electrical and Computer Engineering from Ohio State University, with a focus on RF.

Fred Hawley

Principal reliability engineer

Fred has spent more than 35 years in telecommunications and military reliability engineering. Currently, he focuses on reliability prediction and modeling for CommScope products. His expertise helps direct the development of such diverse solutions as power amplifiers, remote radio heads and other base station- and tower-mounted wireless communication equipment. His work helps ensure superior performance from every solution CommScope builds. Fred holds a MSEE from Columbia University in New York, and is an active member of IEEE.

Mark Hendrix, P.E.

Engineering manager, new product development

Mark is responsible for directing innovation efforts in wireless communications systems and modular data centers for CommScope, with particular focus on thermal design, power systems and other efficiency drivers. He brings 27 years of expertise in electronic packaging, with strong backgrounds in both defense electronics and telecommunications for such names as Texas Instruments, Fujitsu and Xtera Communications. Mark holds eight U.S. patents and is a registered Professional Engineer in the State of Texas. He holds a B.S. in Mechanical Engineering from Clemson University, and an M.S. in Mechanical Engineering from Southern Methodist University.

Chris Hills
Technical director, microwave systems

Chris is an important part of CommScope's Andrew Solutions team, offering leadership in the field of microwave antennas and array antennas and aligning new technology to marketplace conditions. In an earlier role as RF Engineering Manager for Andrew, Ltd., Chris devoted nearly 20 years to advancing antenna design, resulting in the development of the highly successful ValuLine family of antennas used the world over for point-to-point communications, as well as other significant intellectual property. Chris earned his B.S. degree with Honors in Electronic and Electrical Engineering from the University of Surrey in the United Kingdom. He holds multiple patents and serves as an active member of IEEE and IET.

Erik Lilieholm
Applications engineering manager, wireless network solutions

Erik's diverse background in North America's wireless communications industry dates back to the launch of the first cellular communications networks. He has built his expertise with Allen Telecom, LGP Telecom and Ericsson. With more than 25 years in RF design, product management and technical marketing, Erik provides critical leadership to CommScope's families of wireless solutions, helping each product fulfill its specific role and customer need. He holds several patents in the field of RF filter technology. Erik earned a Master of Science degree in Electrical Engineering from the Royal Institute of Technology in Stockholm, Sweden, and an MBA from the University of Nevada, Reno.

Louis Meyer, P.E.
Director of technical marketing—RF path

Louis has spent a lifetime advancing RF technology, taking it from the drawing board to practical use. Over the years in various roles with Allen Telecom, Andrew Ltd. and CommScope, Louis was responsible for supporting the sales teams for such solutions as remote antenna control systems, transmission lines, diplexers and other important components. Prior to joining Allen Telecom, Louis worked with Decibel Products as V.P. Antenna Design and V.P. International OEM Relations. Earlier, Louis worked with Harris Corporation in RF communications and Bendix Corporation in their missile systems division. Louis holds five patents and has been active as a chair and vice-chair of the TIA's TR-8.11 Antenna Standards sub-committee. He earned his B.S. in Electrical Engineering from Marquette University in Milwaukee, Wisconsin and is currently a registered Professional Engineer in the state of Texas.

Larry Seper
Director of construction services

Larry provides telecommunications installation solutions for CommScope's customers, supporting products and systems such as antennas and transmission lines, power amplifiers, remote radio heads and PIM/Sweep testing procedures. He also focuses on civil site work, helping establish the physical foundations of wireless communication as well as its technological foundations. Larry brings more than 36 years of financial and operations management experience to bear for CommScope's customers, and is a key player in making certain that every CommScope solution is the right solution. Larry holds a B.S. in Accounting from Marquette University in Milwaukee, Wisconsin and is an active member of AICPA.

Tom Sullivan

Director, global new product introductions

Tom currently oversees new product introductions for CommScope, and he's served in many roles since joining Andrew, Ltd. in 1982 as an antenna design engineer. In that position, Tom developed the antennas and supporting devices customers needed to build point-to-point communications networks. He has also served as a Global Sales Manager for OEMs including Ericsson, Nokia, Lucent and Motorola. He also worked in quality control and product line management in support of CommScope's successful HELIAX cable and connector solutions. Tom is an active member of IEEE, and holds a B.S. in Electronic Engineering Technology from DeVry Institute of Technology, and an MBA in International Business from St. Xavier University in Chicago, Illinois.

Dr. Junaid Syed

Electrical engineering manager, microwave systems

Junaid directs new product development for CommScope in the areas of microwave and millimeter wave antenna systems, flexi waveguides and waveguide components that support mobile backhaul systems. He brings 26 years of international experience in the telecommunication and defense industries. He holds seven patents and is a current member of SE Scotland IET and Technical committees. He also represents CommScope as a technical committee member with ETSI. Junaid earned his B.S. in Aero Sciences from Punjab University with Silver Medal honors, and a Bachelor of Engineering degree in Electronics/Avionics from NED University of Engineering and Technology with Gold Medal honors, both in Pakistan. He earned his Ph.D. in Microwave and Millimeter Wave from the University of London and conducted his post-doctoral research on reflect array antenna design at Queen's University Belfast, both in the United Kingdom.

Dr. Keith Tappin—In Memoriam

Director of engineering, microwave systems

It is with heavy hearts that we share the news of Dr. Keith Tappin's sudden passing. His significant influence on this book is one of many positive contributions Keith made as a part of the CommScope family. He was a first-rate engineer, manager and leader—a warm man with a great sense of humor and a genuine passion for his profession. Please join us as we offer our condolences to Keith's family, friends and colleagues—and our gratitude for the life he shared with us.

Keith was responsible for developing new products and processes related to advanced passive microwave applications for CommScope's Microwave Systems group. Keith built his expertise in microwave components and antenna design for 15 years, focusing heavily on defense and commercial communications systems. Prior to joining CommScope, Keith conducted research at the University of Illinois (U.S.) and The University of Birmingham (UK), where he investigated metallic radiation interactions and developed high-temperature alloys for aerospace applications. Keith published more than 20 technical papers and held two patents. He earned his degree in Metallurgy and Materials Science from the University of Liverpool in the UK, where he later completed his Ph.D.

Glossary

alternating current (AC)
An electrical current that changes polarity (i.e., direction) 50 to 60 times per second. It offers significant efficiencies when transmitted across power lines, making it the standard current for household use. *See also: direct current*.

Andrew integrated management and operating system (A.I.M.O.S.)
The network management software solution for all CommScope-distributed antenna systems. Part of the Andrew Solutions portfolio, A.I.M.O.S. performs configuration fault and inventory via SNMP for alarm forwarding.

antenna
The portion of an RF system that radiates radio energy into space and collects it from space.

Antenna Interface Standards Group (AISG)
An industry group comprised of more than 40 top manufacturers and service providers from all over the world. AISG was founded in 2001 and publishes universally accepted industry protocols for communications between base stations and tower-based equipment, such as antennas and tower-mounted amplifiers.

attenuation
Measured in decibels (dB), attenuation is the loss of power experienced by an RF signal as it moves from one point to another. Transmission line attenuation is expressed in either decibels per 100 feet (dB/100 feet) or decibels per 100 meters (dB/100m) of cable length.

automatic transmission power control (ATPC)
A system that dynamically raises transmission power to overcome the effects of interference.

azimuth coordinate system
The polar coordinate system used in the field by RF engineers and surveyors to map the radiation pattern of antennas. *See also: radiation pattern, spherical coordinate system.*

backhaul
The process of connecting two ends of a transmission through a central routing point.

bandpass cavity
A “frequency filter” that limits the channels that pass through the filter to a radio receiver’s select set of frequencies. Other frequencies are prevented from passing. Most devices have multi-stage bandpass cavities that filter out different frequencies at each stage.

bandpass duplexer
A duplexer that uses multiple bandpass cavities to separate transmitter and receiver signals, allowing for simultaneous two-way communications. *See also: Duplexer, Duplex Communications, Bandpass Cavity.*

bypass (pass-through) configuration
A single-band tower-mounted antenna with an integrated diplexer that adds a secondary, non-amplified RF path to the system.

co-channel dual-polar (CCDP) operation
Using both horizontal and vertical polarity of a single frequency to double available bandwidth.

co-siting solutions
The technology and techniques that allow cellular base stations and air interfaces to share architecture and operate within limiting factors of their locations.

coaxial cable
A transmission line built to prevent interference while carrying multiple signals. Coaxial cable consists of an inner core conductor and an outer sleeve conductor, separated by a nonconductive dielectric layer. Coaxial cable is often used to connect antennas to base stations. *See also: dielectric layer.*

dielectric layer
The insulated, tube-shaped layer separating a coaxial cable’s inner and outer conductors. Dielectrics may be made of solid material, flexible foam or open air channels supported by nonconductive spacers. *See also: coaxial cable.*

direct current (DC)
An electrical current that runs continuously in a single direction, making it well suited for use in motors and electronic components such as semiconductors. Batteries also produce DC current. *See also: alternating current*.

distributed antenna system (DAS)
A network of nodes serving a specific place, area or building. They connect through a central base station for backhaul out to the public network.

dummy load
A simulated power load applied to an electrical system for testing purposes. *See also: voltage standing wave ratio (VSWR).*

duplex communications
A transmitter and receiver that work at the same time on the same RF device, allowing two-way communications. *See also: duplexer.*

duplexer
A device situated between a duplexed antenna and its associated transmitter and receiver that provides isolation between the two signals. See also: Duplex Communications.

electrical potential
The difference in electrical charge between two points in space, measured in volts. The greater the difference, the higher the potential – and therefore, the greater the voltage. *See also: volt.*

electrical tilt antenna
An antenna fitted with actuators that can adjust its tilt relative to the ground. These adjustments affect gain, or performance, of the antenna within defined geographical areas.

environmental factors
Circumstances of temperature, sunlight exposure, humidity and other specific characteristics of an installation. Environmental factors play a large role in determining what kind of antenna, transmission line, power and other components are ideal for use in a particular location.

failures in time (FITs)
The number of expected component failures per billion operating hours. See also: reliability.

flat fading
Total signal loss caused by atmospheric refraction. It is the result of a signal being bent completely out of its LOS connection with its receiver.

frequency multiplexing
A configuration that connects multiple base station services that operate in separate bands to multiple antennas via a single feeder cable and its associated couplers.

grounding
Measures taken to control and facilitate the path of an electrical discharge from its source to the ground, avoiding potential damage to sensitive equipment along the way. Grounding is a key element in protecting an installation from damage by lightning strike or other hazards.

guard bands
Narrow gaps inserted into the bandwidths managed by the low loss combiner (LLC) to distinguish between different signals riding on the same bands. *See also: isolation, low loss combiner, transmitter noise.*

horizontal separation
The practice of placing a transmitter's antenna a certain distance from the same device's receiving antenna to achieve the necessary isolation. *See also: duplex communications, isolation, vertical separation.*

in phase
A state of operation referring to multiple antennas radiating together at precisely the same time and rate.

integrated power systems
Space-saving combinations of related components, built into a single device for easy installation.

interoperability testing (IOT)
The practice of testing for operational problems in a cell site that employs equipment from multiple manufacturers.

isolation
The amount of separation achieved between the transmitter and receiver in a duplex communication system. In general, more isolation translates to less interference between the two functions, and correspondingly clearer communications. *See also: duplex communications, horizontal separation, vertical separation.*

line of sight (LOS)
The unobstructed space between transmitter and receiver. Longer hops must even account for the curve of the Earth as an obstruction.

low loss combiner (LLC)
A device in the RF path that permits the simultaneous operation of multiple transmitters on a single antenna. It applies guard bands and bandpass cavities to provide the necessary isolation between signals. *See also: bandpass cavity, guard bands, isolation.*

nonlinearity
A location within an electrical circuit where voltage does not remain consistently proportional to power, generally caused by imperfect connections between components and cables or damage to a cable's structure.

ohm
The unit of measurement of a material's electrical resistance. When applied to discussions of RF transmission lines, ohms refer to the inherent, or characteristic, loss of strength a signal encounters as it passes along a length of cable.

pass-through (bypass) configuration
A single-band tower-mounted antenna with an integrated diplexer that adds a secondary, non-amplified RF path to the system.

passive intermodulation (PIM)
A potential side effect of having more than one high-powered signal operating on a passive device such as a cable or antenna. PIM occurs at non-linear points in a system such as junctions, connections or interfaces between dissimilar metal conductors, creating interfering frequencies that can decrease efficiency. The higher the signal amplitude, or power, the greater the effect. *See also: nonlinearity.*

quarter-wave shorting stub
A device inserted into the connection between transmission line and antenna that does not affect normal frequencies, but will immediately short—and safely dissipate energy—when lightning frequencies attempt to cross. *See also: grounding.*

radiation pattern
The three-dimensional shape of an antenna's strongest signal transmission.

radome
A wind- and water-proofed fabric or plastic cover that protects an antenna from the elements.

receiver desensitization
Interference caused by unwanted frequencies entering a receiver's upper stage passbands. These errant signals create electrical variances that impede the receiver's operation. *See also: bandpass cavity.*

reliability
The probability of a device working correctly over a defined length of time, operating under specified conditions. *See also: failures in time (FITs).*

remote radio head (RRH)
A recent advance in base station architecture that separates a cell site base station's RF and baseband functions for improved efficiency. RRH advantages include no active cooling requirement, lower overall power loss, less weight on the tower and compact size.

resonant frequency
The natural tendency of a system to oscillate with larger amplitude at particular frequencies. At these frequencies, even small periodic driving forces can produce large amplitude oscillations.

same-band combining (SBC)
A base station configuration that allows multiple services to share the same bands.

service company
A cell site development partner responsible for actual construction on the site, including antenna towers, concrete footers and pads, security fencing, and equipment shelters.

Shannon's Law
Created by Claude Shannon and Ralph Hartley, this law establishes a theoretical limit to how much data can be reliably pushed through a given amount of bandwidth.

signal polarization
The orientation of a signal's electric field relative to the ground. It may be horizontal or vertical.

spherical coordinate system
A geometric polar coordinate system used to mathematically map the radiation pattern of antennas. *See also: azimuth coordinate system, radiation pattern.*

split-mount radio system
A two-stage connection that lets microwave radios located in an indoor unit (IDU) receive and transmit through an antenna fitted with an outdoor unit (ODU).

transmission lines
In RF applications, the physical medium that conducts RF power from one point to another, usually between a base station and an antenna.

transmitter noise
Interference experienced by a receiver as a result of transmission power "leaking" into other nearby frequencies.

vertical separation
The practice of placing a transmitter and receiver in separate locations on a single antenna, allowing the height difference to achieve the necessary isolation. *See also: duplex communications, horizontal separation, isolation.*

volt
A measurement of electric potential difference between two points in a path. Voltage is sometimes referred to as "pressure," because it shares many characteristics with pressure in a water pipe.

voltage polarity (+ and –)
The positive (+) and negative (–) designations of voltage refer to which polarity of a circuit is measured; in terms of actual power produced, the distinction is meaningless.

voltage standing wave ratio (VSWR)
A key measurement of cable performance and signal quality. It quantifies the amount of signal reflected backward along a cable to its source. Theoretically, perfect operation yields a VSWR value of 1.0, or "unity," meaning zero reflections.

Yagi antenna
Also known as a Yagi-Uda antenna, this is a common type of directional antenna, first created in Japan in 1926 by Hidetsugu Yagi and Shintaro Uda.

COMMSCOPE®

INFRASTRUCTURE ACADEMY

Take your learning to the next level with the CommScope Infrastructure Academy.

The CommScope Infrastructure Academy SP6000 is your next step in understanding the RF Path. This affordable, online resource is available now—so you can learn even more about the science, the technology and the business of RF communications on your own schedule. SP6000 covers such important topics as:

- Cell site development and construction
- Base station antennas
- Co-siting solutions
- Remote radio heads
- Transmission and receiving isolation systems
- Transmission lines
- Passive intermodulation (PIM) fundamentals
- Microwave backhaul
- Powering wireless networks
- Distributed antenna systems
- Lightning protection

To learn more about the CommScope Infrastructure Academy or to register for a class, please visit www.commscopetraining.com.

Technology is reshaping the way we live, learn and thrive. The epicenter of this transformation is the communications network—our passion. CommScope designs and builds the connectivity for the world's most advanced networks.

COMMSCOPE®

commscope.com

Visit our website or contact your local CommScope representative for more information.

© 2017 CommScope, Inc. All rights reserved.

All trademarks identified by ® or ™ are registered trademarks or trademarks, respectively, of CommScope, Inc. This document is for planning purposes only and is not intended to modify or supplement any specifications or warranties relating to CommScope products or services. CommScope is committed to the highest standards of business integrity and environmental sustainability, with a number of CommScope's facilities across the globe certified in accordance with international standards, including ISO 9001, TL 9000, and ISO 14001. Further information regarding CommScope's commitment can be found at www.commscope.com/About-Us/Corporate-Responsibility-and-Sustainability.

advertisement

Microwave Energy Detector & Wireless Energy Harvesting

By tbudka (/member/tbudka/) in Technology (/technology/) > Electronics (/technology/electronics/) 20,966 282 19 Featured

PUBLIC DOMAIN Download Favorite

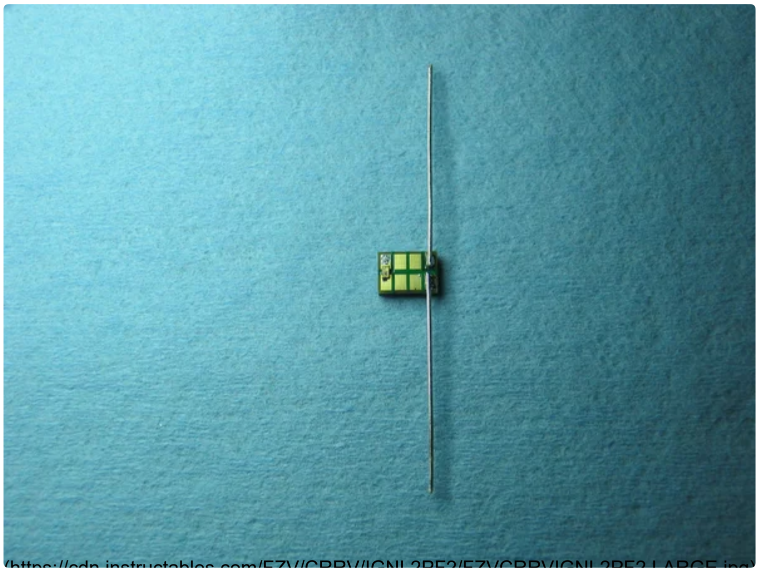
(https://cdn.instructables.com/FZV/CBRV/ICNL2BF2/FZVCBRVICNL2BF2.LARGE.jpg)

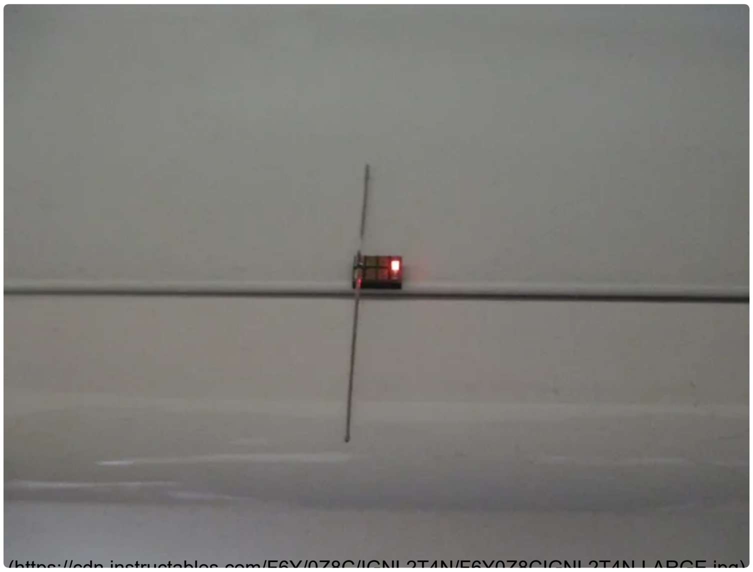
(https://cdn.instructables.com/F6Y/0Z8C/ICNL2T4N/F6Y0Z8CICNL2T4N.LARGE.jpg)

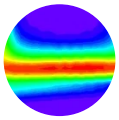
(/member/tbudka/)
By **tbudka**
(/member/tbudka/)
RF Diagnostics, LLC
(http://www.rfdiagnostics.com)

Follow

More by the author:

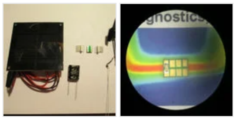

About: Microwave Engineer, Consultant, Design Contractor, Electronics Enthusiast, Power Amplifier Designer, Filter Design, Electromagnetic Simulations, Orcad PCB Editor Design More About tbudka » (/member/tbudka/)

Description

This article describes how to build a wireless harvesting device that will capture stray 2.5GHz energy from a microwave oven and convert it into a useable voltage that will light up a red light emitting diode (LED). Hobbyists can modify the antenna with longer wires and try to pick up other signals such as AM/FM radio, mobile phone signals and other wireless energy. The LED's light will be visible when the module picks up 1 milliwatt-10 milliwatts of microwave power. Usually the most power leaks out of the door seam of a microwave oven.

What you need:

1 RFD102A (RF-DC Converter) module available at

http://www.rfdiagnostics.com/product/rfd102a/

1 ¼ watt leaded resistor (any value) or two wires that are 1 1/8 inches long (28.6mm)

1 Kingbright Surface mount red LED (part number: APT1608EC) available at Digikey/Mouser/Newark

Fine tipped soldering iron

Solder paste or Fine Solder

Add Tip Ask Question Comment Download

LA TOTALMENTE NUEVA
SILVERADO
CHEVROLET
Explora Silverado
Ad ✕
advertisement

Step 1: Assembly Instructions

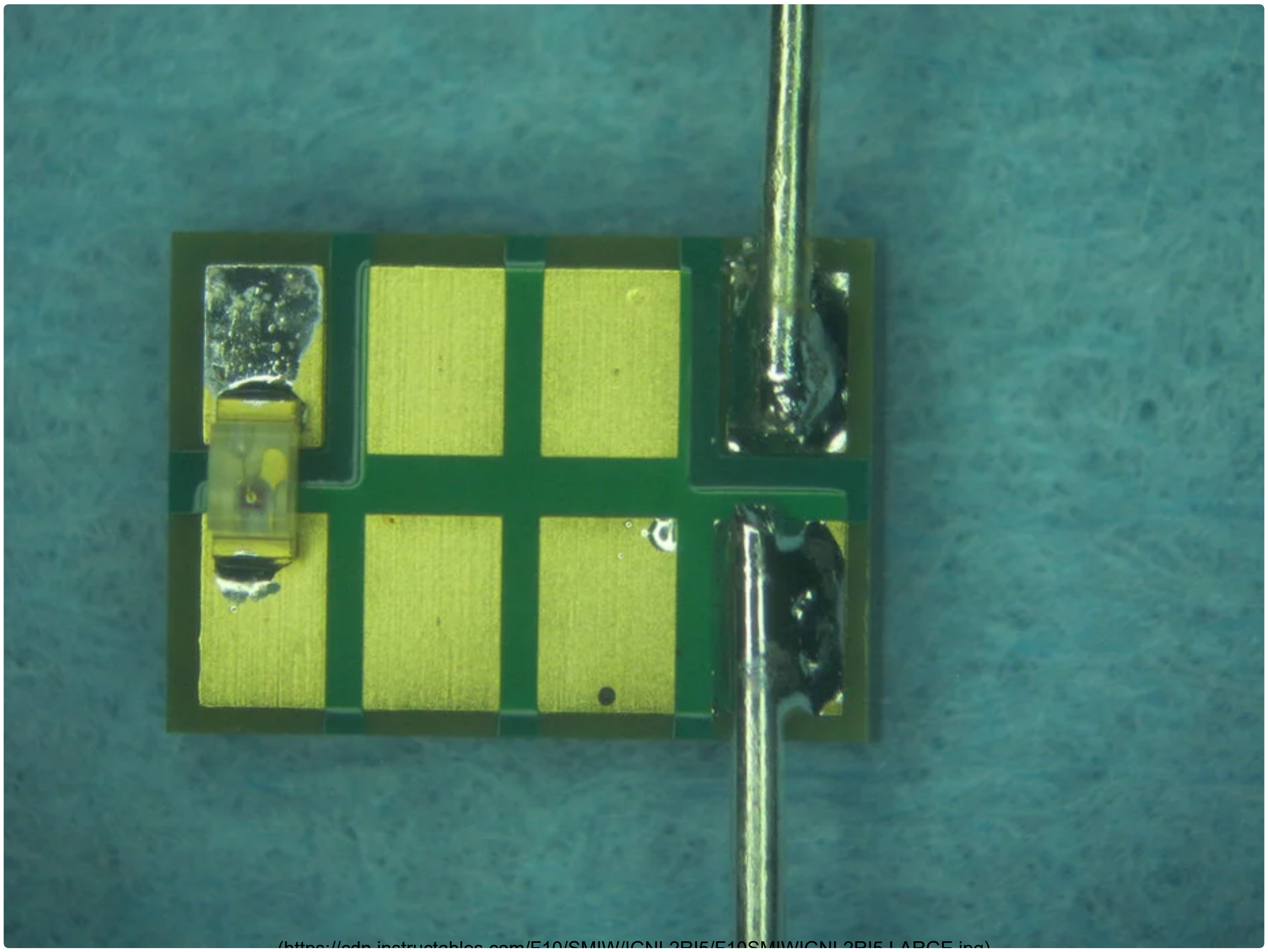

(https://cdn.instructables.com/F10/SMIW/JCNL2BI5/F10SMIWJCNL2BI5.LARGE.jpg)

Cut the resistor wires off next to the resistor. These are just the right size at 1 1/8” long for a 2.5GHz dipole. Throw away the resistor and keep the wires.

Put solder paste on the module at pins 1 & 8 and at pins 4 and 5.Place the wires on pins 4 and 5 and solder carefully using tweezers to hold the wires (it will burn you otherwise). Solder at the lowest soldering temperature possible to avoid damaging the module. If the iron is too hot then you may damage the internal connections inside the module. Use a minimum of time for soldering (<10secs). The wires work as a dipole antenna to collect the 2.5GHz energy into the RF (Radio Frequency) Input of the module.

Place the LED with the anode (positive side) onto pin 1 and the cathode (negative side) on pin 8 and solder carefully. For those not familiar with LEDs, the triangle symbol of the diode should point to the ground pin of the module (pin 8). Your final microwave harvester should look like figure 2

Add Tip Ask Question Comment Download

Step 2: Test on Your Microwave Oven

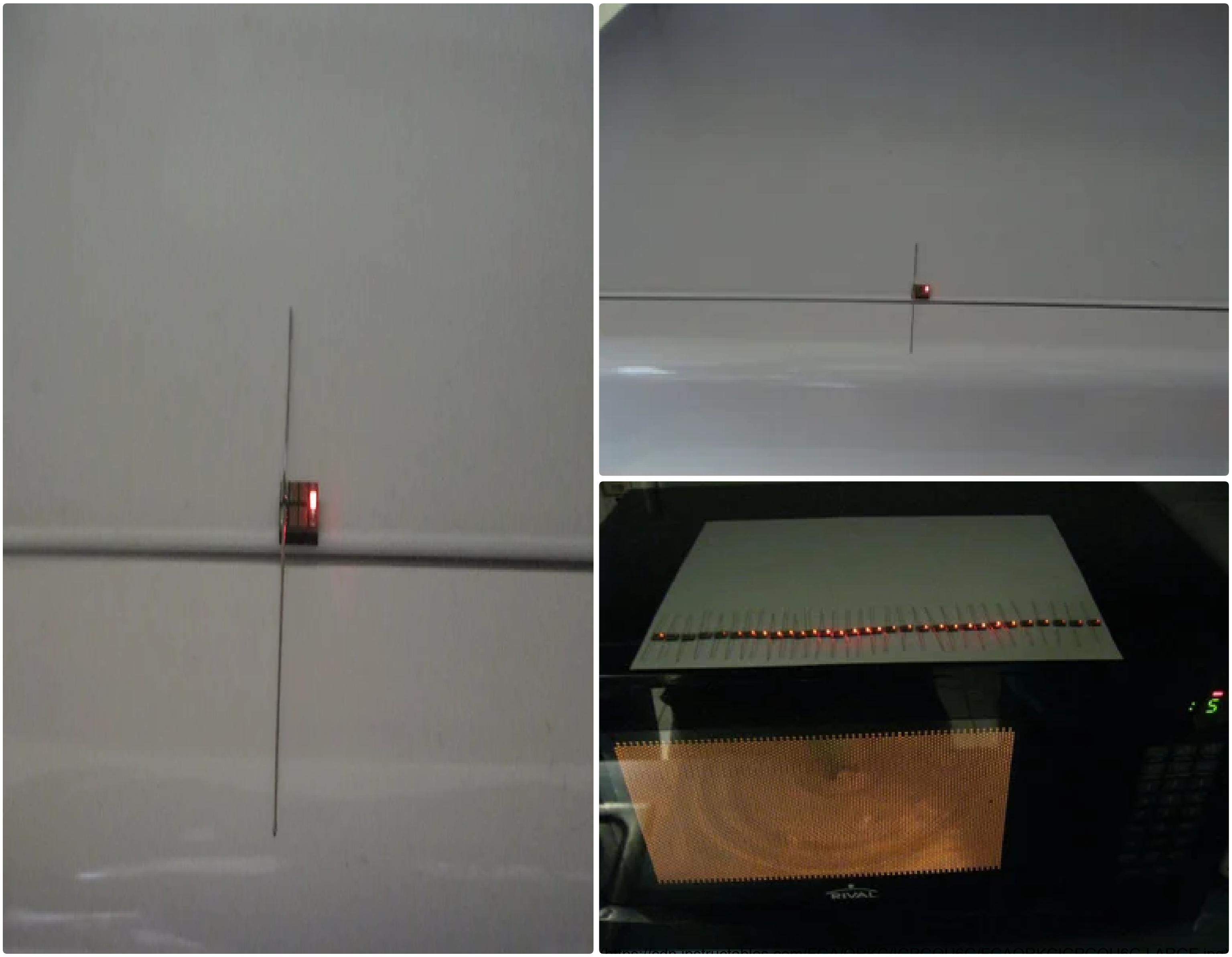

Now put some water in a mug and put it in the microwave oven and cook on high for 2 minutes. While the microwave is running, hold the module and move the module antenna around to find a hot spot on the microwave. Tape the module on the ground wire to the microwave to keep it in place. If all is well then you will see the LED light up and dim as the electric fields change while the mug rotates in the microwave oven. The above figures show the microwave energy harvester in action. Experiment with antenna shapes and sizes. If you go too big or too small with the antenna then less energy will be collected. You can also put the assembled module near an 802.11b/g/n wireless LAN router and watch it flicker as the router transmits and receives 2.4-2.5GHz energy. Enjoy!

Other Project Ideas:

- Harvest AM/FM Radio Signals

FAQ

- What is the maximum current this module can provide? Ans: Typical current is ~0.5-5mA. Max current is 18mA.

- What is the efficiency of converting RF energy to a DC voltage? Ans: If the input power to the module is around 10-50mW then the RFD102A can achieve >50% efficiency if the resistance seen by the DC output is ~10kOhm.

- I don't have a soldering iron and am all thumbs. Where can I get a detector like this already built up? Ans. The RFD102A-DET is available http://www.rfdiagnostics.com/product/microwave-oven-window-energy-harvesting-kit/

- What is the maximum output voltage this module can produce? Ans: With 0.5W input power the module can produce up to 37V at 915MHz.

- Can I charge my mobile phone with this? Ans: It is possible with 4 RFD102A modules and antennas to produce enough current to charge a mobile phone. The hard part is the wireless source would need to be so powerful that you wouldn't want to be in the same room during charging. You may hurt your eyes by cooking them.

Add Tip Ask Question Comment Download

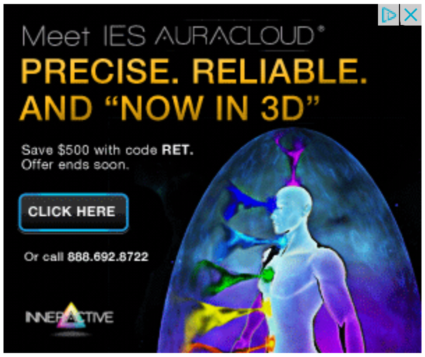
Meet IES AURACLOUD®
PRECISE. RELIABLE. AND "NOW IN 3D"
Save $500 with code RET. Offer ends soon.
CLICK HERE
Or call 888.692.8722
INNERACTIVE
advertisement

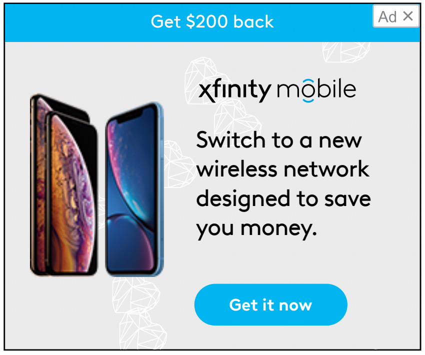
Get $200 back
Ad
xfinity mobile
Switch to a new wireless network designed to save you money.
Get it now
advertisement

Share

Did you make this project? Share it with us!

I Made It!

RF Energy Harvesting System And Rectennas

Riboy Cheriyan, Amsu Thejus. M, Jagadish Chandran G, Dr. K P Zacharia

Abstract —Recent advances in ultra-low power wireless communications and energy-harvesting technologies have made self-sustainable devices feasible. Typically, the major concern for these devices is battery life and replacement. Applying energy harvesting techniques to these devices can significantly extend battery life and sometimes even eliminate the need for a battery. This technique is widely applicable in low power circuits such as wireless sensors. One emerging wireless sensor networks (WSN) application is in agriculture sector, where sensor nodes are deployed in fields to monitor humidity, temperature and soil moisture. Energy supply to such sensors is an issue as they are typically powered by conventional batteries which have a limited lifespan. Cost is often prohibitive when replacing exhausted batteries since the sensor devices need to be unearthed. An attractive solution is to use radio frequency (RF) energy harvesting, in which the radiated RF energy from ambient is extracted and converted into usable energy to power up the sensors.

Index Terms — Harvesting energy, Patch antenna, Wireless sensors, Ambient energy , Schottky diode, Resonant circuit.

◆

1. INTRODUCTION

Radio frequency energy is emitted by sources that generate high electromagnetic fields such as TV signals, wireless radio networks and cell phone towers, but through using a power generating circuit linked to a receiving antenna; this free flowing energy can be captured and converted into usable DC voltage. Most commonly used as an application for radio frequency identification tags in which the sensing device wirelessly sends a radio frequency to a harvesting device which supplies just enough power to send back identification information specific to the item of interest. The circuit systems which receive the detected radio frequency from the antenna are made on a fraction of a micrometer scale but can convert the propagated electromagnetic waves to low voltage DC power at distances up to 100 meters. Depending on concentration levels which can differ through the day, the power conversion circuit may be attached to a capacitor which can disperse a constant required voltage for the sensor and circuit when there isn’t a sufficient supply of incoming energy.

We are being bombarded with energy waves every second of the day. Radio and television towers, satellites orbiting earth, and even the cellular phone antennas are constantly transmitting energy. What if there was a way we could harvest the energy that is being transmitted and use it as a source of power? If it could be possible to gather the energy and store it, we could potentially use it to power other circuits. In the case of the cellular phone, this power could be used to recharge a battery that is constantly being depleted. The potential exists for cellular phones, and even more complicated devices - i.e. pocket organizers, person digital assistants (PDAs), and even notebook computers - to become completely wireless.

Of course, right now this is all theoretical. There are many complications to be dealt with. The first major obstacle is that it is not a trivial problem to capture energy from the air. We will use a concept called energy harvesting. Energy harvesting is the idea of gathering transmitted energy and either using it to power a circuit or storing it for later use. The concept needs an efficient antenna along with a circuit capable of converting alternating-current (AC) voltage to direct-current (DC) voltage. The efficiency of an antenna, as being discussed here, is related to the shape and impedance of the antenna and the impedance of the circuit. If the two impedances aren’t matched then there is reflection of the power back into the antenna meaning that the circuit was unable to receive all the available power. Matching of the impedances means that the impedance of the antenna is the complex conjugate of the impedance of the circuit.

This paper is considered to be one of the first steps towards what could become a standard circuit included in every wireless sensors, and quite possibly every electronic device made. A way to charge the battery of an electric circuit without plugging it into the wall would change the way people use wireless systems. However, this technology needs to be proven first. It was decided to begin the project with a cellular phone because of the relative simplicity of the battery system. Also, after we prove that the technology will work in the manner suggested, cellular phones would most likely be the first devices to have such circuitry implemented on a wide scale. This advancement coupled with a better overall wireless service can be expected to lead to the mainstream use of cell phones as people’s only phones. This paper is an empirical study of whether or not this idea is feasible. This first step is to get an external wireless circuit to work with an existing phone by transmitting energy to the phone (battery) through the air.

2. POWER SOURCES

The main technological advancement that has allowed these harvesting devices to generate sufficient power is through the development of receivers which can sense broad ranges of frequencies, not just limited to television UHF signals, while capturing the highest concentration of generated waves. Since the propagation of wave energy dissipates with distance from source, sensors located at far distances must extract energy at low power density which is related to the distance through 1/d^2 [1]. A typical television antenna linked to the correct conversion circuit if located approximately 4 meters from the source of radio frequency generation of 677 mega Hertz and 960 kilowatts of effective radiation power is able to produce a detected voltage of 0.7 across an 8 kilo ohm load which is approximately 60 microwatts of harvested power and enough to power an LCD display thermometer[1].

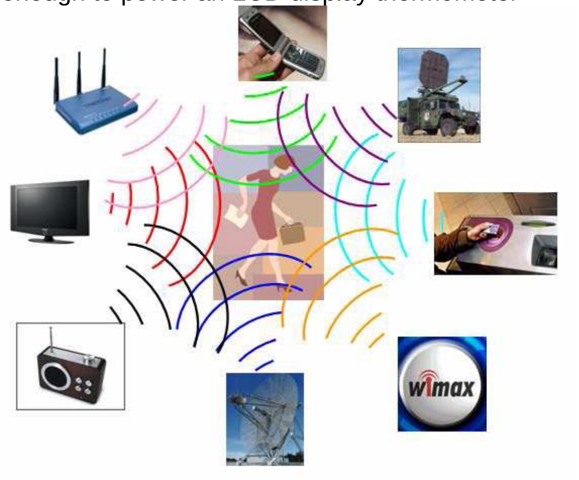

Fig 1. Ambient power sources.

The output power of is limited by regulations, such as Federal Communication Commission (FCC), USA due to safety and health concern offered by EM radiations. The maximum theoretical power available for RF energy harvesting is 7.0 *u*W and 1.0 uW for 900 MHz and 2.4 GHz frequencies respectively for a free space distance of 40 meters.

A 1V/m electric field can yield power density of about 0.26uW/cm2. RF emissions from 800MHz range mobile equipments are significantly stronger signal than others, such as 2.4GHz range, which is mostly used by wireless LAN. A preferable range for RF energy harvesting is between 500 megahertz and 10 gigahertz, in which many different radio communication signals lies[4].

3. BASIC BLOCK DIAGRAM

Figure 2 shows the basic block diagram of RF energy harvesting system.

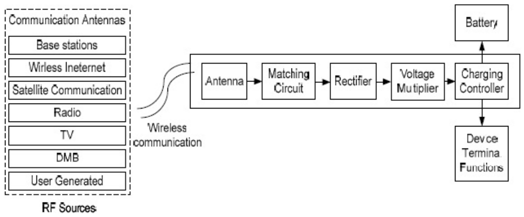

Fig. 2. Basic block diagram

Here we harvest energy from different sources such as mobile base stations, Wi-Fi, Radio waves etc. An antenna is used to capture signals, it may be a broadband antenna or separate antennas for different frequencies[2] .At high frequency we can use small antennas like patch antenna, but at lower frequencies we need a large antenna such as telescopic antenna in our FM radios[2]. Matching circuit consists of a variable capacitor for giving good impedance matching at the output of antenna at different frequencies. Rectifier is used to convert the ac at the output of antenna into DC. Here we prefer Schottky diodes because of its high frequency capability. Voltage multiplier consists of capacitor diode network which is used to increase the voltage level for charging the battery.

4. A SIMPLE GERMANIUM DIODE BASED ENERGY HARVESTING SYSTEM

Here a germanium diode based simple RF energy harvesting is shown in figure 3. Antenna used here is a stick antenna that is used in FM radios. The Gang capacitor and MW coil act as a matching circuit, resonating at a particular frequency.

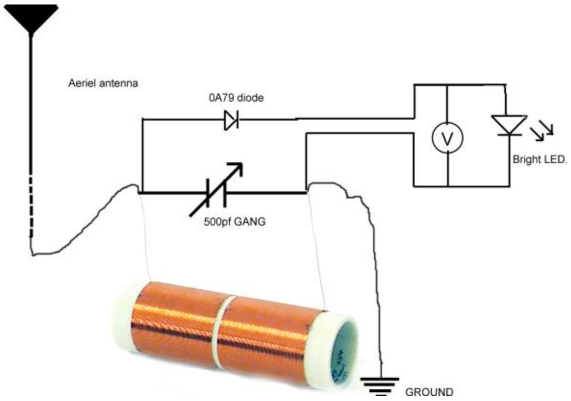

Fig. 3 OA79 Germanium diode based energy harvesting circuit.

OA79 Diode is used as rectifier because of its low turn on voltage. The gang capacitor tuned in such a way that the matching circuit resonates with the strong frequency that the antenna can capture. If the RF signal is strong, the circuit can drive a LED.

5. RF ENERGY HARVESTING SYSTEM BASED ON SCHOTTKY DIODE

Schottky diode offer low forward voltage and high switching speed, and consider as an ideal component for RF energy harvesting.

This circuit uses HSMS2820 schottky diode. RF signals can be captured using multiband antenna as shown in figure. It is capable of receiving 900Mhz/1800MHz/1900MHz/2.4GHz. Small patch antennas can be used at higher frequencies.

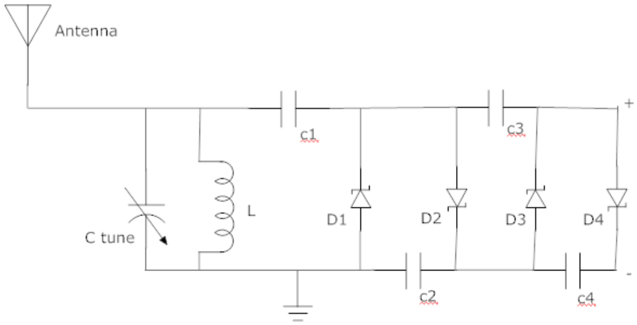

Fig. 4. Schottky diode based RF energy harvesting circuit with voltage multiplier.

Here combination of variable capacitor Ctune and inductor L forms the matching circuit for antenna at different frequencies. Using the Ctune in antenna matching circuit we can tune the antenna at its resonant frequency. The capacitor Ctune in tuning circuit can be varied using following formula to resonate with the antenna inductance L , were f is the frequency of operation.

$$2\pi f = \frac{1}{\sqrt{L_{antenna} \times C_{tune}}}$$

At higher frequencies such as 800 or 1900MHz the low values of inductors are difficult to construct especially at board level circuit design. But using the inductor along with capacitor at integrated circuit level design greatly improves the performance. Resonant frequency is also influenced by diode capacitance as it is related with reverse diode voltage and input voltage[4].

For converting Radio frequency to DC, a rectifier circuit is used .A charge pump circuit such as a voltage multiplier circuit can be used as rectifier. In figure there is a two stage voltage multiplier consists of two diodes and two capacitors in each stage. Here the diode used is a schottky diode (HSMS2820) because of its low forward voltage and high switching speed.

The voltage output of multiplier is twice input peak voltage minus the diode twice the diode threshold voltage.[4] The charge drained (Delta q) by the load current (I load) per period, where f is input signal frequency is

$$\Delta q = I_{load}/f$$

The circuit can be extended to n stages, producing the required output level. For a n stage multiplier the output capacitance can be calculated as

$$V_o = nV_{in} - \frac{n-1}{fC} I_{load}$$

From this, there should be a diode with very low turn on voltage and high operating frequency. Schottky diode has all this parameters so here a schottky diode is used.

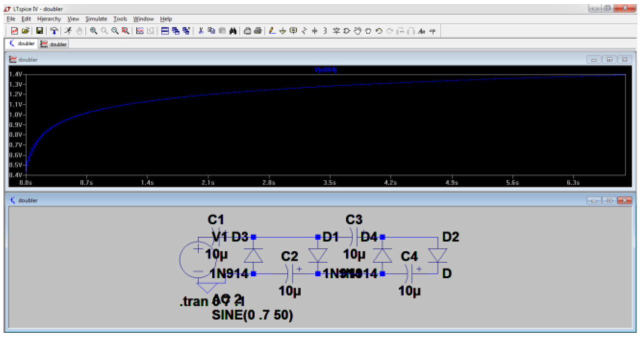

Fig. 5. Simulation result of two stage voltage multiplier.

Figure 5 shows the simulation result of a two stage multiplier with 0.7V input using LT Spice.

6. RECTENNAS

The rectenna termed as rectifying antenna, is combination of an antenna and a nonlinear rectifying element (Schottky diode, IMPATT diode…etc.) where the two elements are integrated into a single circuit. Since the rectenna is a receiving module collecting power from a radiation field, its sensitivity is defined by

$$S_R = \frac{output\ voltage}{power\ density\ at\ the\ antenna}$$

Division of this value with the effective antenna aperture leads to the normally used quantity for detectors.

$$S_D = \frac{output\ voltage}{input\ power}$$

This characterizes the nonlinear element with its matching.

A rectenna is useful as the receiving terminal of a power transmission system where dc power needs to be delivered to a load, through free space, for which physical transmission lines are not feasible. It is also suitable in applications where dc power needs to be distributed to more numbers of load elements that are spatially distributed. Such power distribution is achieved by the dispersive nature of microwave energy in space, eliminating the need for physical interconnects to individual load elements.

Figure 6 shows a rectenna for 899 Mhz. Here antenna consists of ainset fed rectangular patch with ground plane. It uses HSMS 2820 schottky diode as rectifying element

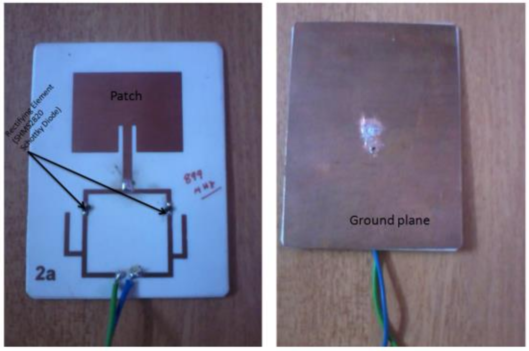

Fig. 6. Rectenna For 899 Mhz.

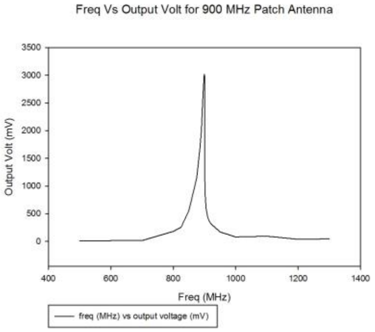

Fig. 7. Frequency Response of a 899 MHz Rectenna .

Figure 7 shows the frequency response of a 899Mhz rectenna. Here the peak is at 895 MHz. At this frequency we obtained an output dc voltage of 3.2V. In general it is difficult to predict how the rectenna system is optimized for the maximum conversion efficiency. However, there are several theoretical methods to overcome this problem.

The main objective with the design of rectenna is to obtain high conversion efficiency, and there are two approaches to achieve this goal. In first maximum power is collected and delivered to the rectifying diode, while in the second one harmonics generated by the diode are suppressed, which re-radiate from the antenna as power lost. In order to increase conversion efficiency by the first method, several broadband antennas, large antenna arrays and circularly polarized antennas have been designed and experimented. The broadband antenna enables relatively high RF power to be received from various sources while the antenna array can increase incident power delivered to the diode by enlarging antenna aperture and antenna gain. Antenna array is an effective means to increase the receiving power for rectification. However, a trade-off arises between the antenna size and the radiation gain. The circularly polarized antenna offers power reception with less polarization mismatch. However in second method, the rectenna consist of an LPF between the antenna and the diode, as well as an additional LPF on the dc output side of the diode. The main reason for the rise in the efficiency was the improvement in the diode and circuit construction for high input power levels.

7. CONCLUSION

In this paper, a study of various methods used for RFenergy harvesting has been made. It is found that we can harvest energy in micro watt range from ambient RF sources. Here the harvested power highly depends on the distance between the transmitter and RF harvesting system. By using an array of harvesting antennas we can harvest considerable amount of power.

REFERENCES

[1] D. Bouchouicha, F. Dupont, M. Latrach, L.ventura, “Ambient RF Energy Harvesting”, International Conference on Renewable Energies and Power Quality, Granada (Spain), 23th to 25th March, 2010.

[2] James O, MC Spadden, Taewhan Yoo, Kai Chang, Theoretical and Experimental Investigation of Rectenna Element for Micro Wave Power Transmission ”, IEEE Transactions On Microwave theory and Technique, Vol. 40, No. 12, December 1992.

[3] Rakesh Kumar Yadav, Sushrut Das and R. L. Yadava, “Rectenna Design Development and Application” International Journal of Engineering Science and Technology, Vol. 3 No. 10 October 2011.

[4] Hamid Jabbar Student Member, IEEE,Young. S. Song Member, IEEE,Taikyeong Ted. Jeong,” RF Energy Harvesting System and Circuits for Charging of Mobile Devices” IEEE Transactions on Consumer Electronics, Vol. 56, No. 1, February 2010.

Progetti dei lettori - Le radio di Sophie

Ricevitore a cristallo per FM

Un progetto di Giacomo Cavuoti

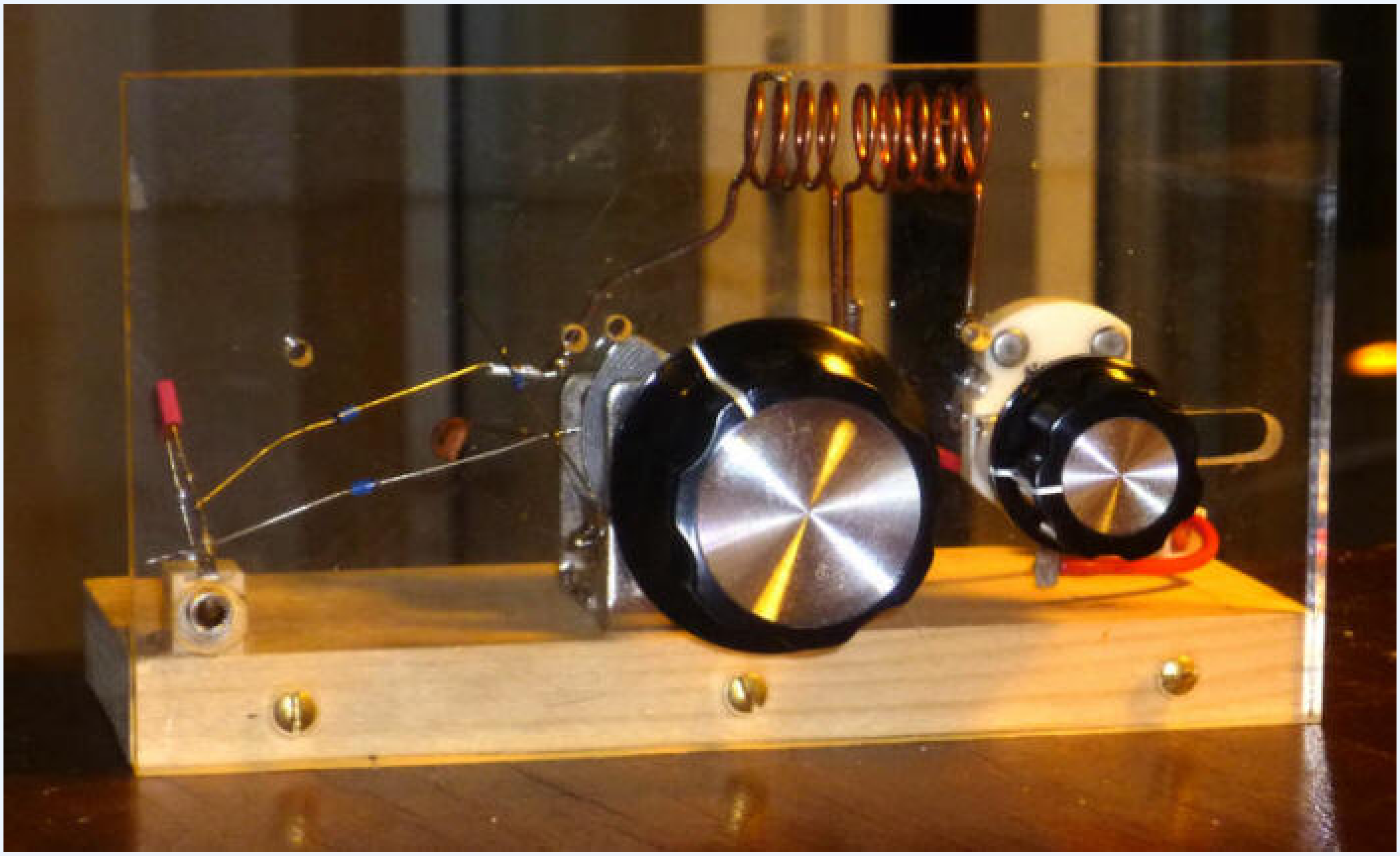

Seguendo l'invito di Leonardo ho realizzato questo semplice ricevitore a cristallo per FM (fare clic sullo schema qua sotto per vederlo ingrandito).

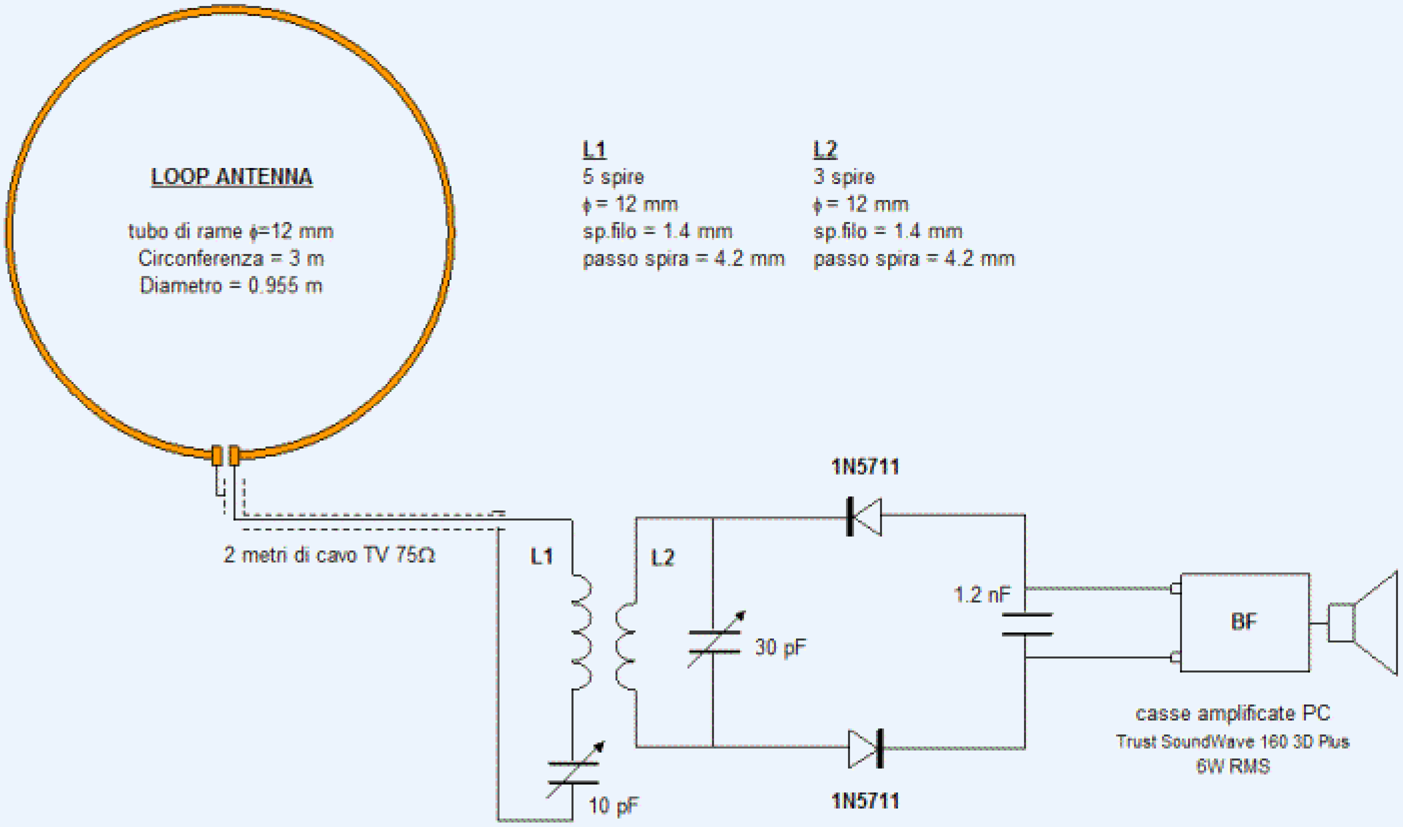
LOOP ANTENNA
tubo di rame ϕ=12 mm
Circonferenza = 3 m
Diametro = 0.955 m
L1
5 spire
ϕ = 12 mm
sp.filo = 1.4 mm
passo spira = 4.2 mm
L2
3 spire
ϕ = 12 mm
sp.filo = 1.4 mm
passo spira = 4.2 mm
1N5711
2 metri di cavo TV 75Ω
L1
L2
30 pF
1.2 nF
BF
casse amplificate PC
Trust SoundWave 160 3D Plus
6W RMS
10 pF
1N5711

Il ricevitore è "alimentato" da una antenna a loop ad onda intera costituita da una sola spira circolare realizzata con 3 metri di tubo di rame per impianti idraulici (induttanza misurata pari a circa 3 uH). Il supporto è un semplice listello in legno di abete lungo 2 metri, largo 3 cm e spesso 1 cm, a cui il cerchio di rame è stato fissato con nastro isolante. Un'altra soluzione per realizzare l'antenna a loop in modo semplice ed istantaneo è utilizzare una piattina di alluminio di 2 mm di spessore e 2 cm di larghezza. I rivelatori sono due diodi *Schottky* tipo 1N5711, particolarmente adatti per VHF grazie alla bassa capacità (circa 2pF). Sono attualmente in produzione e quindi facili da trovare. Il ricevitore è stato collaudato in una situazione "stiff" con l'antenna esposta in un balcone incassato (3 x 1.6 metri) al quarto piano di un palazzo in cemento armato di 6 piani circondato da altri edifici realizzati con pareti intere in cemento armato. Come si può vedere e sentire dai due filmati che ho caricato su Youtube:

http://www.youtube.com/watch?v=YNiuRoIp36E

http://www.youtube.com/watch?v=SJwx8_3mCfI

la ricezione è buona come anche la selettività, grazie all'accoppiamento appropriato, ma molto facile da trovare, tra la bobina di antenna L1 e quella di sintonia L2. L'amplificazione audio è stata affidata a due piccole casse amplificate da PC. Le foto seguenti mostrano il primo prototipo funzionante montato "in aria" (fare clic sulle immagini per vederle meglio). La foto sotto il titolo mostra invece una realizzazione esteticamente più accurata.

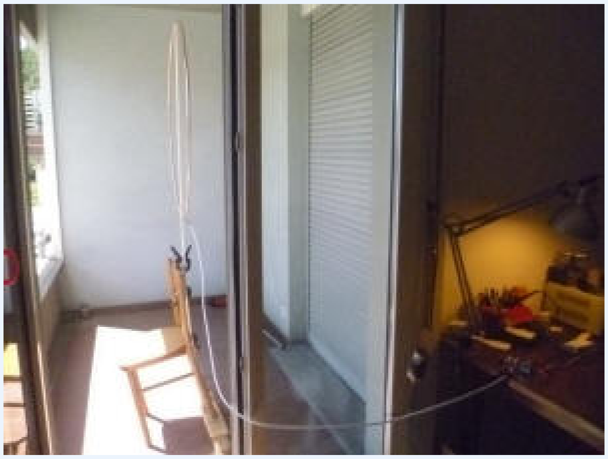

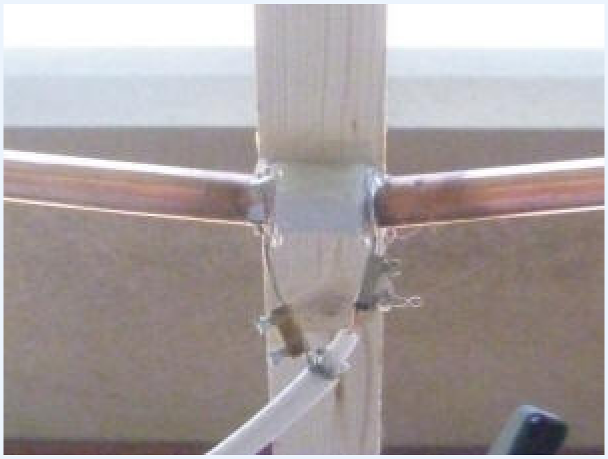

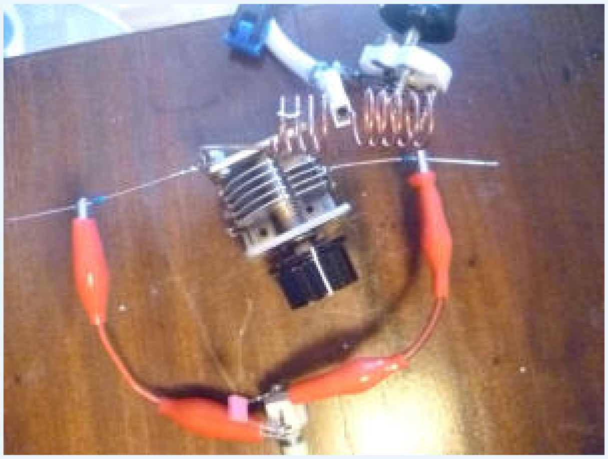

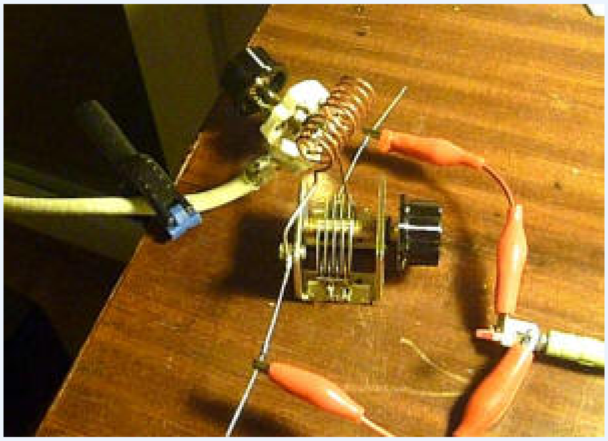

Per ogni chiarimento non esitate a contattarmi. Giacomo Cavuoti

Progetti dei lettori - Radio a cristallo
Torna alla Pagina Principale

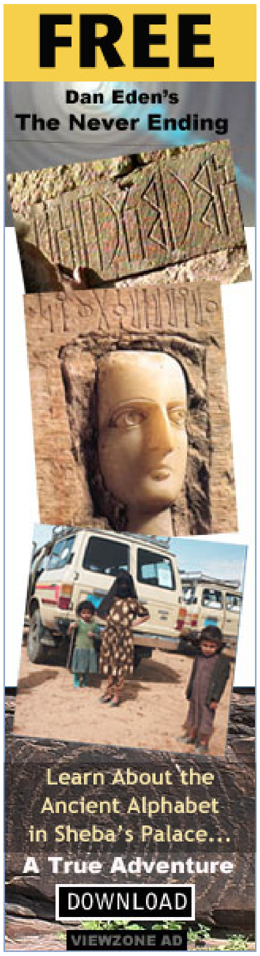

FREE
Dan Eden's
The Never Ending
Learn About the Ancient Alphabet in Sheba's Palace...
A True Adventure
DOWNLOAD
VIEWZONE AD

Are you ready? We'll help you find your way.
Start Now
ODUOnline
OLD DOMINION UNIVERSITY

An Invention of Free Energy : 100 Years Ago!

Roy Meyers' Absorber -- "Free Energy"

by Robert A. Nelson

As planet Earth suffers under the relentless onslaught of ever increasing human consumption, we are becoming increasingly desperate for sources of energy with which to power our pretensions of civilization. There is not enough oil or coal available to sustain the present rate of growth, and the apparently inevitable result will be endless wars over dwindling resources. There are, however, numerous marvelous alternative energy technologies that have been suppressed by the powers that be.

We could easily be enjoying the benefits of clean, cheap (and even free) energy in various forms, were it not for the insatiable greed of our corporate masters, who rule the world without regard for humanity or the environment.

There are dozens of stories that can be told of inventors who presented wonderful liberating technologies, only to be ruined for their efforts. Many people have heard of various types of so-called "free energy" or "over-unity" generators that tap what scientists call Zero Point Energy (and other names, such as the aether) without need for fuel and without producing any pollution. Detractors like to label such devices as impossible "perpetual motion machines", even though some of them are solid state electronic systems that have been built and shown to work. Readers who are interested in such possibilities need search no farther than the impeccable history of Thomas Henry Moray, which is widely published on the internet (e.g., http://www.rexresearch.com/moray.moray.htm).

If that is too much bother for you, then just continue reading here. This is the little-known story of **Roy Jerome Meyers**, who invented an "Absorber" that produced limitless electrical power from the terrestrial magnetic field.

Almost 100 years ago, Meyers served a 3-1/2 year sentence in the Arizona state prison at Florence. Before his incarceration, he had already invented an improved trolley wheel head that prevented it from jumping off the wire. While he was imprisoned, he was inspired to invent a device that would draw electricity from the atmosphere. He conferred with Superintendent Sims and Parole Clerk Sanders, and convinced them to give him the opportunity to develop his idea in a wooden shed on the grounds of the penitentiary. Within a few weeks, using easily available materials (chrome steel magnets and iron wire), he constructed the first crude working model, and used it to spark the gas engines of the prison's pump house. His second model developed 8 volts.

Miss Kate Barnard, who was State Commissioner of Charities and Corrections of Oklahoma, was a guest of Superintendent Sims at the time, and she saw the machine in operation. Miss Barnard was so impressed by it and by Meyers' essential integrity (despite the lapse that had gotten him imprisoned) that she told the story of Roy Meyers when later she appeared before the Arizona legislature to address them concerning prison reform. The legislature and Governor Hunt were convinced to grant Meyers an unsupervised leave of absence for 30 days to travel to Washington DC in order to apply for a patent.

Meyers gave this account of his trip:

> **"When I arrived in Washington and laid my plans before the patent office experts, they merely smiled and told me that I would have to build a model and demonstrate my claims --- that it seemed strange that I, unknown as I am in the electrical world, should have**

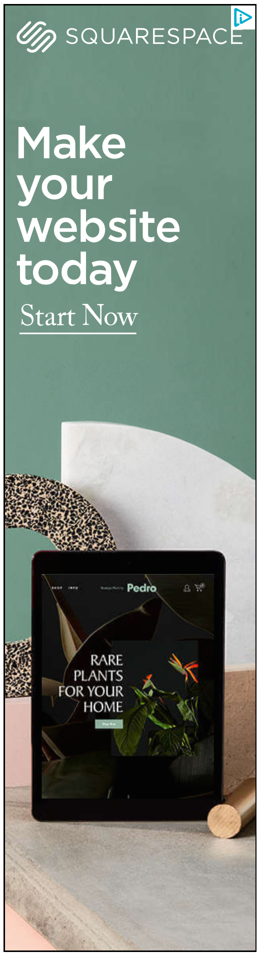

Are you interested in 70-448? Get our self paced MB5-858 and 1Y0-A15 packages to pass your exam 70-667 exam without any difficulty in cisco 642-647 guides.

> **accomplished the things for which Edison, Tesla and other experts have been striving for years.**
>
> **"They could not grasp the meaning of my drawings nor the explanation I tried to make to them. There was little time to spare, as I had only 20 days left of my leave, but I set to work in a few days was able to take a crude model around to the patent office to make a demonstration.**
>
> **"Arriving at the patent office I telephoned to a friend who had been so kind as to introduce me and aid me in reaching the proper officials. The absorber was hoisted on two short poles and made to work. While they were as yet unable to understand the principles involved and hardly willing to believe their eyes, they were forced to admit that I had something new and different, and they told me that there would be no further objection; that I might file my application without further delay.**
>
> **"I hope to construct my first large machine right here in Phoenix. I feel grateful; to Governor Hunt and others for what they have done for me and to the help they have given in securing protection I might not otherwise have had, and I am desirous of demonstrating this gratitude. I am going back to Florence today to resume the serving of my sentence, which will expire in 10 months. Then, here in Phoenix, I will begin the work of making my machines."** (*Technology World Magazine,* 1912)

Meyers planned to build a 200 foot tower upon which to place an "Absorber" (as he called it) with sufficient power to light the city of Phoenix. According to him, it comprised "a series of magnetized steel plates set in a circle and this mechanism attracts the electricity from the atmosphere. This is carried by wires to a transformer in the engine house below and thence is applied to produce either power or light after the usual manner.

> "The flow of electricity is constant. When it emerges into the transformer it is in the form of a direct current. It will absorb the electricity day and night and will work whenever the wireless will work. I can put up a plant to supply such a building as the Adams Hotel for about $1500, and one of the principal items of the expense is the cost of the towers, the wires, the magnetizing of one set of plates, which is part of the secret of the treatment which makes it respond to the accumulations of the atmosphere.
>
> "For use in the case of an electrical storm I have made what I call a modified form of circuit breaker, such as is commonly used as a lightning arrester on telegraph lines. In case of a storm the accumulator would suddenly become overcharged, possibly, and as the electricity would not of itself flow back into the air, the result might be disastrous. So I send it down into the ground, whenever the voltage rises above a certain amount."

And the story ends there. It is not known what became of Roy J. Meyers. No "Absorber" was built to power Phoenix, and it appears that a US Patent was not granted. This writer first learned of Meyers' Absorber in 1986, and repeatedly searched for it in the US Patent Gazette, to no avail. It was not until 2006 that Meyers' British patent for "Improvements in and Relating to Apparatus for Producing Electricity" was located after several periodic advanced searches of the European Patent Office online (http://ep.espacenet.com). The patent has long since expired and entered the public domain. It is reprinted below for your edification.

It should be obvious from this brief but indisputable account, and the description given in the patent, that "free energy" is a reality. Anyone who thinks otherwise is merely ignorant or in denial.

Roy J. Meyers & the "Absorber"

References

(1) Charlton Lawrence Edholm: "Picks Power from the Air"; *Technology World Magazine*, p. 279-281 (1912)*Electronic World* (November 1912)

(2) Roy Jerome Meyers: British Patent # 1098 (January 14, 1913); Improvements in and Relating to Apparatus for Producing Electricity

British Patent # 1098

(January 14, 1913)

Improvements in and Relating to Apparatus for Producing Electricity

Roy Jerome Meyers

This invention relates to improvements in apparatus for the production of electrical currents, and the primary object in view is the production of a commercially serviceable electrical current without the employment of mechanical or chemical action. To this end the invention comprises means for producing what I believe to be dynamic electricity from the earth and its ambient elements.

I am, of course aware that it has been proposed to obtain static charges from upper strata of the atmosphere, but such charges are recognized as of widely variant potential and have thus far proved of no practical commercial value, and the present invention is distinguished from all such apparatus as has heretofore been employed for attracting static charges by the fact that this improved apparatus is not designed or employed to produce or generate irregular, fluctuating or other electrical charges which lack constancy, but on the other hand I have by actual test been able to produce from a very small apparatus at comparatively low elevation, say about 50 or 60 feet above the earth's surface, a substantially constant current at a commercially usable voltage and amperage. This current I ascertained by repeated tests is capable of being readily increased by additions of the unit elements in the apparatus hereinafter set forth, and I am convinced from the constancy of the current obtained and its comparatively low potential that the current is dynamic and not static, although, of course, it is not impossible that certain static discharges occur and, in fact, I have found occasion to provide against the damage which might result from such discharge by the provision of lightning arresters and cut-out apparatus which assist in rendering the obtained current stable by eliminating sudden fluctuations which sometimes occur during conditions of high humidity from what I consider static discharges. The nature of my invention is obviously such that I have been unable to establish authoritatively all of the principles involved, and some of the theories herein expressed may possibly prove erroneous, but I do know and am able to demonstrate that the apparatus which I have discovered does produce, generate, or otherwise acquire a difference of potential representing a current amperage above stated, or varied therefrom at the will of the operator according to the uses which the current is to be subjected.

The invention comprises generically means for producing electrical currents of serviceable potential substantially without the employment of mechanical or chemical action, and in this connection I have been able to observe no chemical action whatever on the parts utilized although deterioration may possibly occur in some of the parts, but so far as I am able to determine such deterioration does not add to the current supply but is merely incidental to the effect of climatic action.

The invention more specifically comprises the employment of a magnet or magnets and a co-operating element, such as zinc disposed adjacent to the magnet or magnets and connected in such manner and arranged relative to the earth so as to produce current, my observation being that current is produced only when such magnets have their poles facing substantially to the north and south and the zincs are disposed substantially along the magnets.

The invention also comprehends other details of construction, combinations and arrangements of parts as will hereinafter be fully set forth and claimed.

In the accompanying drawings:

Figure 1 is a top plan view of an apparatus embodying the features of the present invention, the arrow accompanying the figure indicating substantially the geographical north, parts of the figure being diagrammatic for condensing the showing.

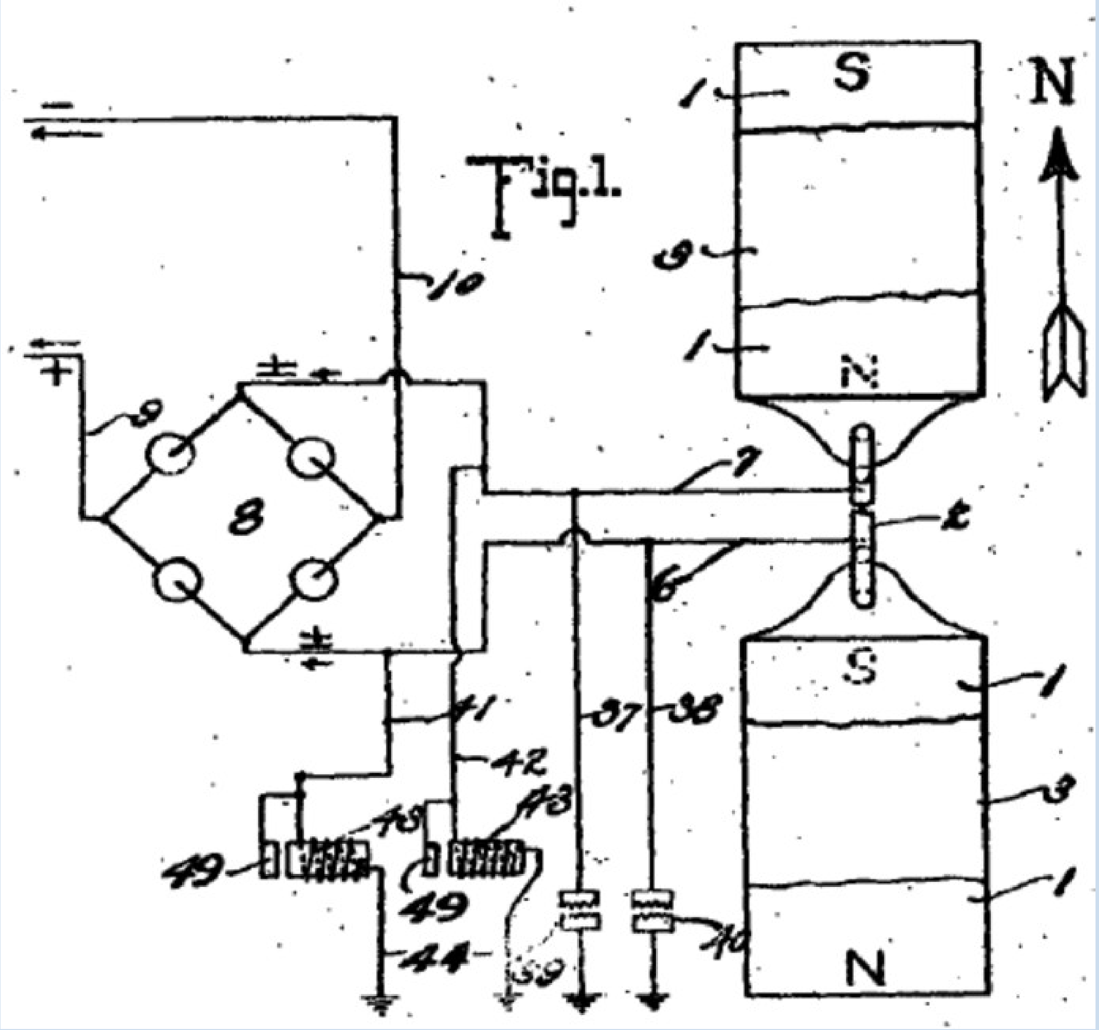

Fig.1.

Figure 2 is a view is side elevation of the parts seen in plan in Figure 1.

Fig.2.

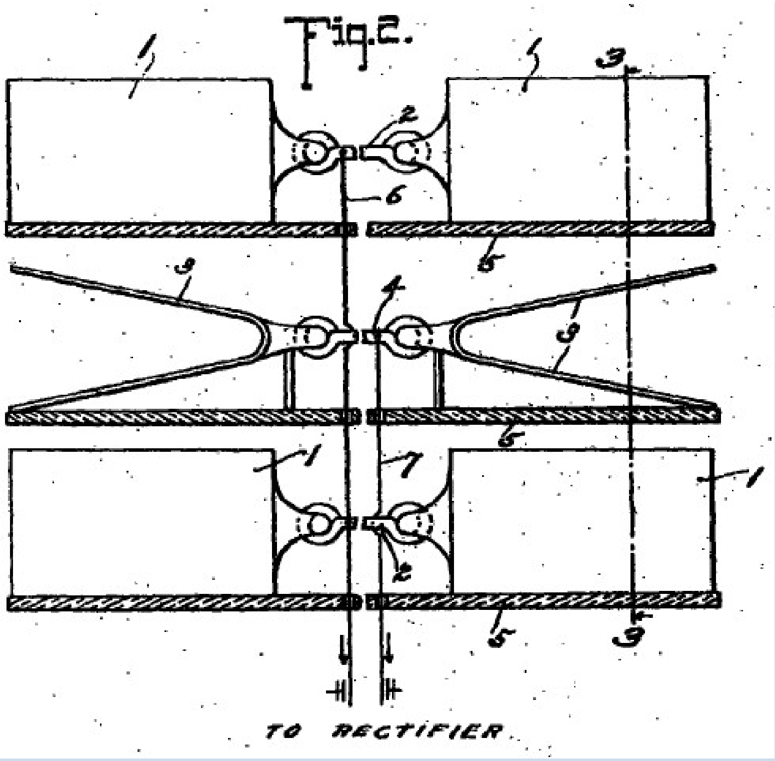
1
1
3
2
6
5
9
4
9
6
1
7
1
2
5
3
TO RECTIFIER

Figure 3 is a vertical section taken on the plane indicated by the line 3-3 of Figure 2 and looking in the direction indicated by the arrow.

Fig.5.

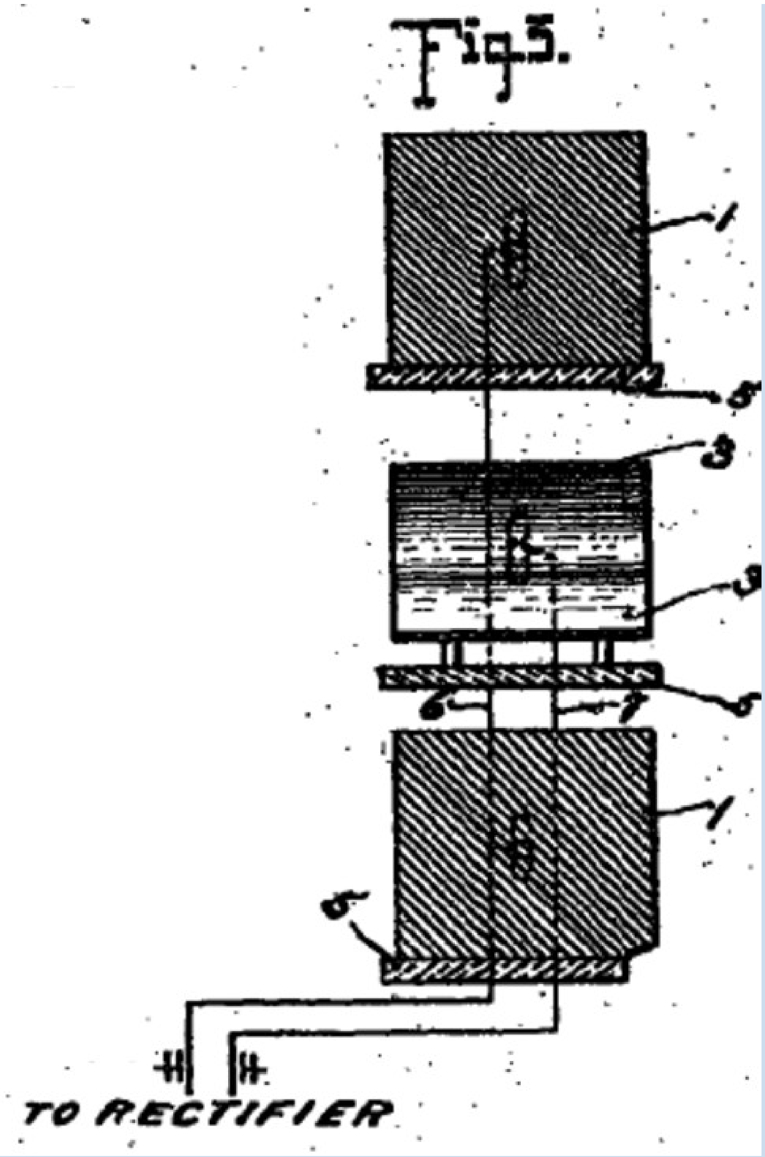

Figure 4 is a detail view partly in elevation and partly in section showing the detail connections of the converter and intensifier.

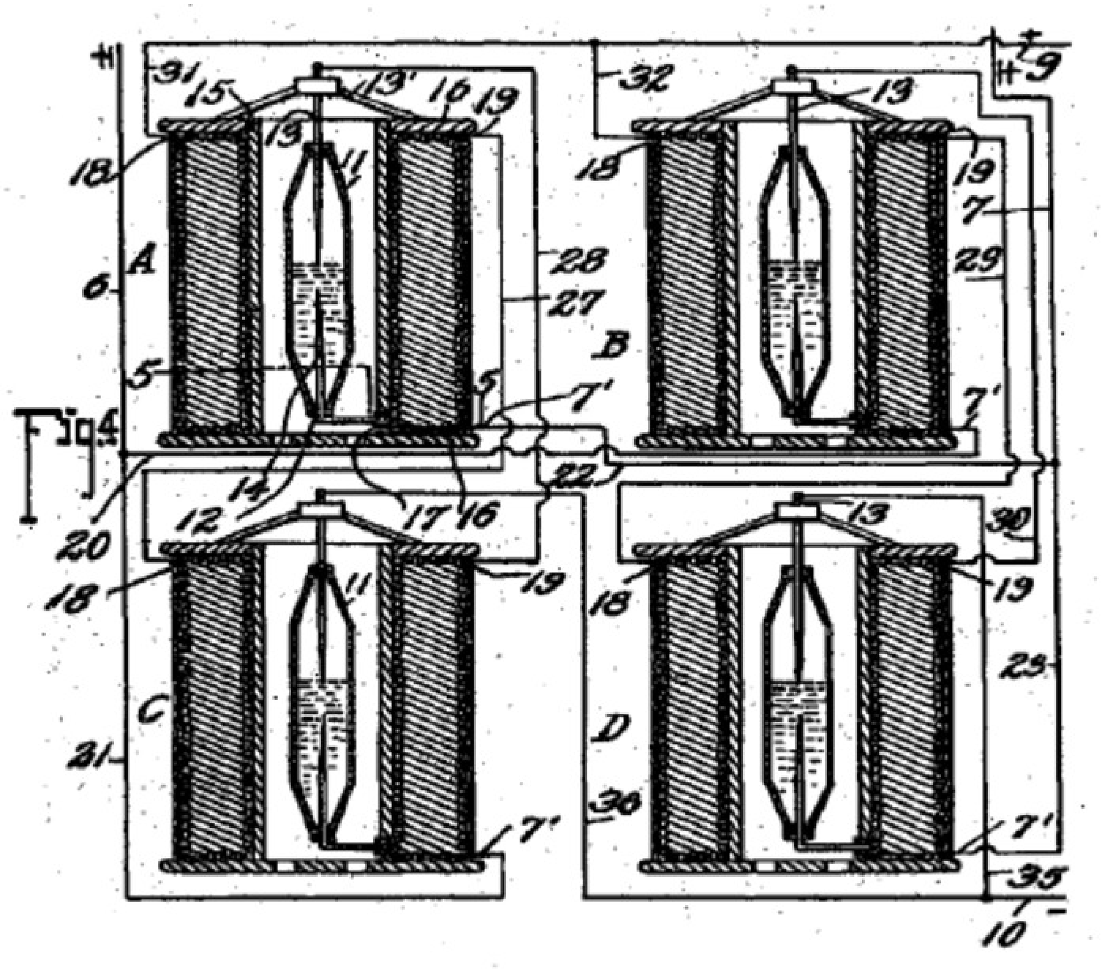

Fig. 4

+
31
15
13'
16
19
13
18
11
A
6
28
27
B
5
5
7'
20
14
12
17
16
22
32
13
18
19
7
29
7'
9
+
13
30
18
19
11
C
21
7'
D
36
18
19
23
7'
35
10
-

Figure 5 is a transverse section taken on the planes indicated by line 5-5 of Figure 4 and looking downwardly.

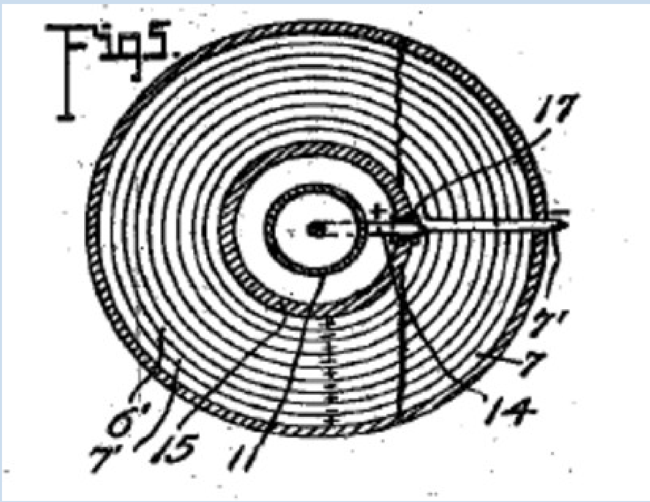

Fig. 5.

17
7'
7
14
6'
7'
15
11

Figure 6 is an enlarged detail fragmentary section illustrating the parts at the juncture of the conductors and one of the intensifiers.

Fig.6

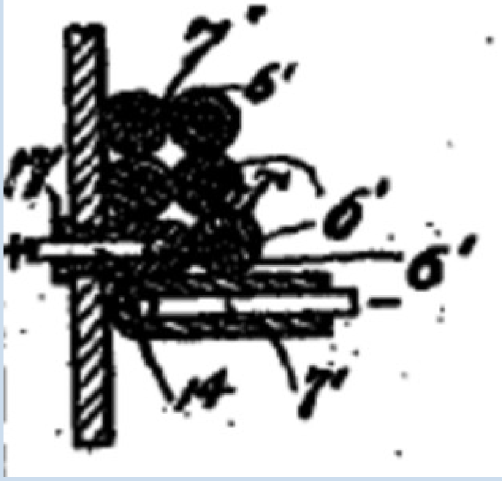

Figure 7 is an enlarged detail view partly in elevation and partly in section of one of the automatic cut-outs and

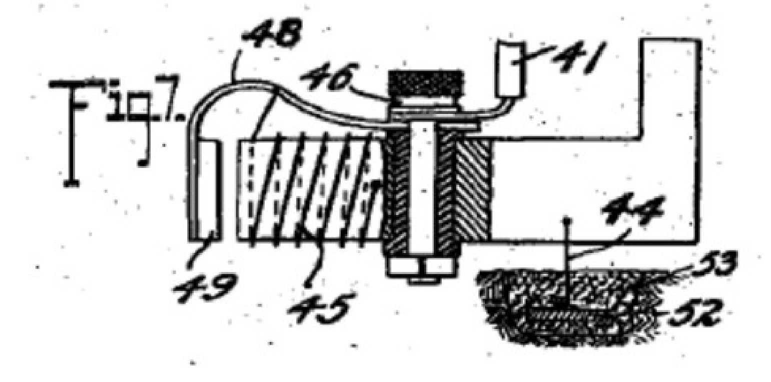
Fig.7.
48
46
41
44
53
52
49
45

Figure 8 is a diagrammatic view of one of the simplest forms of embodiment of the invention.

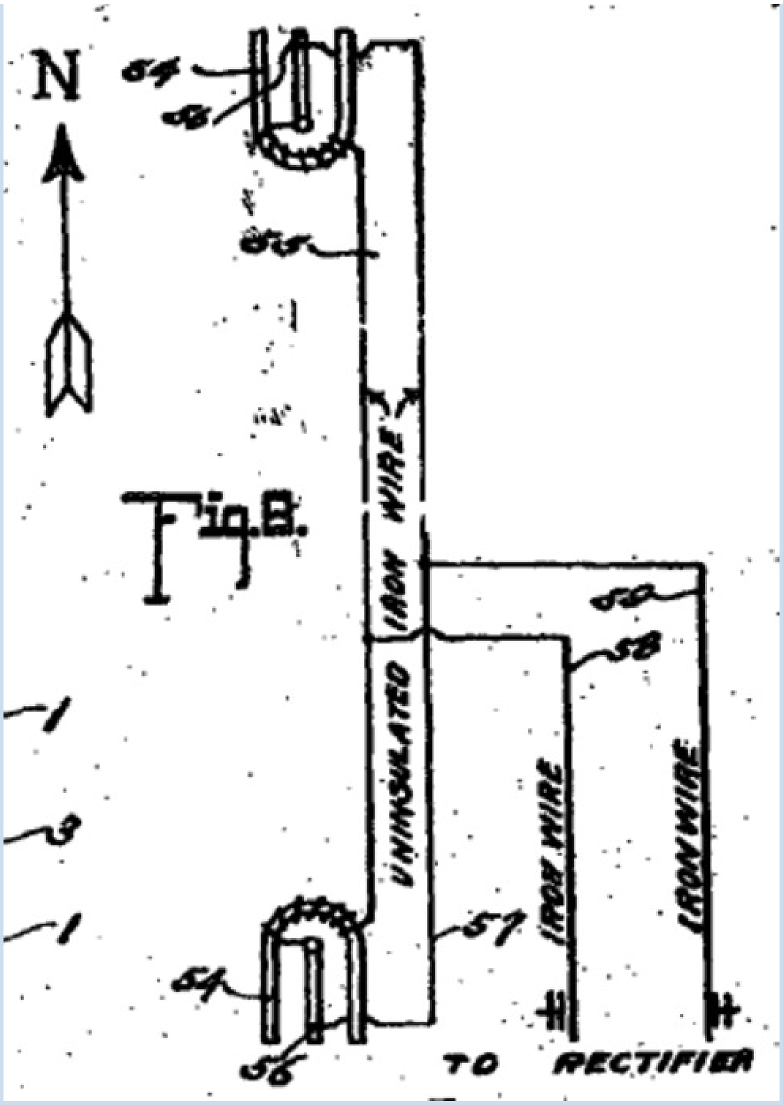

Fig. 8.

Referring to the drawing by numerals, 1,1 indicates magnets connected by a magnetic substance 2, preferably an iron wire. The magnets 1 are arranged in pairs, one pair being spaced beneath the other, and interposed between the magnets are zinc plates 3,3 connected by an iron wire conductor 4. Suitable insulating supports 5 are arranged for sustaining the respective magnets 1 and plates 3,3. Each plate 3 is preferably bent substantially into V form, as clearly seen in Figure 1, and the V1 of one of the plates opens or faces toward the north and the V of the other plate to the South. I have determined by experimentation that it is essential that the plates 3 be disposed substantially north and south with their flat faces approximately parallel to the adjacent faces of the co-operating magnets, although by experience I have not discovered any material difference in the current obtained when the plates are disposed slightly to one side of north and south, as for instance when the plates are disposed slightly to one side of north and south, as for instance when disposed in the line of the magnetic polarity of the earth. The same is true with respect to the magnets 1, the said magnets being disposed substantially north and south for operative purposes, although I find that it is immaterial whether the north pole of one of the magnets is disposed to the north and the south pole to the south, or vice versa, and it is my conviction from experience that it is essential to have the magnets of each pair connected by magnetic material so that the magnets substantially become one with a pole exposed to the north and a pole exposed to the north. In Figure 1, I have indicated in full lines by the letters 8 and N the respective polarities of the magnets 1, and have indicated in dotted lines the other pole of those magnets when the connection 2 is severed. I have found that the magnets and zinc plates operate to produce, whether by collection or generation I am not certain, electrical currents when disposed substantially north and south, but when disposed substantially east and west no such currents are produced. I also find that the question of elevation is by no means vital, but it is true that more efficient results are obtained by placing the zincs and magnets on elevated supports. I furthermore find from tests that it is possible to obtain currents from the apparatus with the zincs and magnets disposed in a building or otherwise enclosed, although more efficient results are obtained by having the said elements arranged in the open.

While in Figures 1, 2, and 3, I have shown the magnets and the zinc plates as superimposed, it will be apparent, as hereinafter fully set forth, that these elements may be juxtaposed in horizontal planes, and substantially the same results will be secured. Furthermore, the magnets 1 with the interposed zincs 3, as shown in Figures 1, 2 and 3 merely represent a unit which may be repeated either horizontally or vertically for increasing the current supply, and when the unit is repeated the zinc plates are arranged alternating with the magnets throughout the entire series as hereinafter indicated.

A conductor 6 is connected in multiple with the conductors 2 and a conductor 7 is connected with conductor 4, the conductor 6 extending to one terminal of a rectifier which I have indicated by the general reference character 8, and the conductor 7 extending to the other terminal of said rectifier. The rectifier as seen in diagram in Figure 1 may assume any of several well known embodiments of the electrical valve type and may consist of four asymmetric cells or Cooper-Hewitt mercury vapor lamps connected as indicated in Figure 1 for permitting communication of the positive impulses from the conductor 6 only to the line conductor 9 and the negative impulses from conductor 6 on only to the line conductor 10. The current from this rectifier may be delivered through the conductors 9 and 10 to any suitable source for consumption.

While the said rectifier 8 may consist of any of the known types, as above outlined, it preferably consists of a specially constructed rectifier which also has the capacity of intensifying the current and comprises specifically the elements shown in detail in Figures 4, 5, and 6 wherein I have disclosed the detail wiring of the rectifier when composed of four of the rectifying and intensify in elements instead of asymmetric cells or simple mercury vapor valves. As each of these structures is an exact embodiment of all the others, one only will be described, and the description will apply to all. The rectifying element of each construction consists of a mercury tube 11 which is preferably formed of glass or other suitable material, and comprises a cylinder having its end portions tapered and each terminating in an insulating plug or stopper 12. Through the upper stopper 12 is extended the electrode 13 which extends well into the tube and preferably substantially one-half the length thereof to a point adjacent the inner end of an opposing electrode 14 which latter electrode extends thence downwardly through the insulation 12 at the lower end of the tube. The tube 11 is supplied with mercury and is adapted to operate on the principle of the mercury vapor lamp, serving to rectify current by checking back impulses of one sign and permitting passage of impulses of the other. To avoid the necessity for utilizing a starter, as is common with the lamp type of electrical valve, the supply of mercury within the tube may be sufficient to contact with the lower end of the electrode 13 when current is not being supplied, so that as soon as current is passed from one electrode to the other sufficiently for volatilizing that portion of the mercury immediately adjacent the lower end of electrode 13, the structure begins its operation as a rectifier. The tube 11 is surrounded by a tube 15 which is preferably spaced from tube 11 sufficiently for allowing atmospheric or other cooling circulation to pass the tube 11. In some instances, it may be desirable to cool the tube 11 by a surrounding body of liquid, as hereinafter indicated. The tube 15 may be of insulating material but I find efficient results attained by the employment of a steel tube, and fixed to the ends of the of the tube are insulating disks 16, 16 forming a spool on which are wound twin wires 6^1 and 7^1, the wire 6^1 being connected at the inner helix of the coil with the outer end of the electrode 14, the lower portion of said electrode being extended to one side of the tube 11 and passed through an insulating sleeve 17 extending through the tube 15, and at its outer end merging into the adjacent end of the wire 6^1. The wire 7^1 extends directly from the outer portion of the spool through the several helices to a point adjacent the juncture of the electrode 14 with wire 6^1 and thence extends in mechanical parallelism with the wire throughout the coil, the wire 6^1 ending in a terminal 18 and the wire 7^1 ending in a terminal 19. For the sake of convenience of description and of tracing the circuits, each of the apparatus just above described and herein known as an intensifier and rectifier will be mentioned as A, B, C and D, respectively. Conductor 6 is formed with branches 20 and 21 and conductor 7 is formed with similar branches 22 and 23. Branch 20 from conductor 6 connects with conductor 7^1 of intensifier B and branch 21 of conductor 6 connects with the conductor 7^1 of intensifier C, while branch 22 of conductor 7 of intensifier C, while branch 22 of conductor 7 connects with conductor 7^1 of intensifier D. A conductor 27 is connected with terminal 19 of intensifier A and extends to and is connected with the terminal 18 of intensifier C, and a conductor 7 connects with conductor 7^1 of intensifier D. A conductor 27 is connected with terminal 19 of intensifier A. and extends to and is connected with terminal 18 of intensifier C, and a conductor 28 is connected with the terminal 19 of intensifier C and extends from the terminal 19 of intensifier B to the terminal 18 of intensifier D to electrode 13 of intensifier B. Each electrode 13 is supported on a spider 13^1 resting on the upper disk 16 of the respective intensifier. Conductors 31 and 32 are connected with the terminals 18 of intensifiers A and B and are united to form the positive line wire 9 which co-operates with the negative line wire 10 and extends to any suitable point of consumption. The line wire 10 is provided with branches 35 and 36 extending to the electrodes 13 of intensifiers C and D for completing the negative side of the circuit.

Thus it will be seen that alternating currents produced in the wires 6 and 7 will be rectified and delivered in the form of a direct current through the line wires 9 and 10, and I find by experiment that the wires 6 and 7 should be of iron, preferably soft, and may of course be insulated, the other wiring not specified as iron being of copper or other suitable material.

In carrying out the operation as stated, the circuits may be traced as follows: A positive impulse starting at the zincs 3 is directed along conductor 7 to branch 23 to conductor 7^1 and the winding of the rectifier of intensifier B through said rectifier to the conductor 6^1, through the winding thereof to the contact 18, conductor 32 and to the line wire 9. The next or negative impulse directed along conductor 7 cannot find its way along branch 23 and the circuit just above traced because it cannot pass across the rectifier of intensifier B but instead the negative impulse passes along conductor 22 to conductor 7 of intensifier A and the winding thereof to the contact 19 and to conductor 27 to contact 18 of intensifier C, to the winding of the wire 6^1 thereof to the electrode 14 through the rectifier to the of the electrode 13 and conductor of intensifier A, electrode 14 thereof and conductor 6^1 to contact 18 and wire 31 to line wire 9. Obviously the positive impulse cannot pass along the wire 20 because of its inverse approach to the rectifier of intensifier B. The next impulse or negative impulse delivered to conductor 6 cannot pass along conductor 21 because of its connection with electrode 13 of the rectifier of intensifier A, but instead passes along conductor 20 to the wire 7^1 and its winding forming part of intensifier B to the contact 19 and conductor 29 to contact 18 and the winding of wire 6^1 of intensifier D to the electrode 14 and through the rectifier to the electrode 13 and conductor 35 to line wire 10. Thus the current is rectified and all positive impulses directed along one line and all negative impulses along the other lie s that the potential difference between the two lines will be maximum for the given current of the alternating circuit. It is, of course, apparent that a less number of intensifiers with their accompanying rectifier elements may be employed with a sacrifice of the impulses which are checked back from a lack of ability to pass the respective rectifier elements, and in fact I have secured efficient results by the use of a single intensifier with its rectifier elements, as hereinafter set forth.

Grounding conductors 37 and 38 are connected respectively with the conductors 6 and 7 and are provided with the ordinary lightning arresters 39 and 40 respectively for protecting the circuit against high tension static charges.

Conductors 41 and 42 are connected respectively with the conductors 6 and 7 and each connects with an automatic cutout 43 which is grounded as at 4. Each of said automatic cutouts is exactly like the other and one of the same is shown in detail in Figure 7 and comprises the inductive resistance 45 provided with an insulated binding post 46 wit which the respective conductor 6 or 7 is connected, said post also supporting a spring 48 which sustains an armature 49 adjacent to the core of the resistance 45. The helix of resistance 45 is connected preferably through the spring to the binding post at one end and at the other end is grounded on the core of the resistance, the said core being grounded by ground conductor 44 which extends to the metallic plate 52 embedded in moist carbon or other inductive material buried in the earth. Each of the conductors 41, 42 and 44 is of iron, and in this connection I wish it understood that where I state the specific substance I am able to verify the accuracy of the statement by the results of tests which I have made, but of course I wish to include along with such substances al equivalents, as for instance, where iron is mentioned its byproducts, such as steel, and its equivalents such as nickel and other magnetic substances are intended to be comprehended. The cutout apparatus seen in detail in Figure 7 is employed particularly for insuring against high tension currents, it being obvious from the structure shown that when potential rises beyond the limit established by the tension of the spring sustaining the armature 40, the armature will be moved to a position contacting with the core of the cutout device and thereby directly close the ground connection for line wire 41 with conductor 44, eliminating the resistance of winding 45 and allowing the high tension current to be discharged to the ground. Immediately upon such discharge the winding 45 losing its current will allow the core to become demagnetized and release the armature 49 whereby the ground connection is substantially broken leaving only the connection through the winding 45 the resistance of which is sufficient for insuring against loss of low tension current.

In Figure 8 I have illustrated an apparatus which though apparently primitive in construction and arrangement comprehends the first successful embodiment which I produced in the course of discovery of the present invention, and it will be observed that the essential features of the invention are therein disclosed. The structure delineated in said figure consists of horseshoe magnets 54, 55, one facing north and the other south, that is, each opening in the respective directions indicated and the two being connected by an iron wire 55 which is uninsulated and wrapped about the respective magnets each end portion of the wire 55 being extended from the respective magnets to and connected with, as by being soldered to, a zinc plate 56, there being a plate 56 for each magnet and each plate being arranged longitudinally substantially parallel with the legs of the magnet and with the faces of the plate exposed toward the respective legs of the magnet, the plate being thus arranged endwise toward the north and south. An iron wire 57 connects the plates 56, the ends of the wire being preferably connected adjacent the outer ends of the plates but from experiment I find that the wire may be connected at practically any point to the plate. Lead wires 58 and 59 are connected respectively with the wires 55 and 57 and supply an alternating current at a comparatively low tension, and to control such current the wires 58 and 59 may be extended to a rectifier or combined rectifier and intensifier, as above set forth.

The tests which I have found successful with the apparatus seen in Figure 8 were carried out by the employment first of horseshoe magnets approximately 4 inches in length, the bar comprising the horseshoe being about one inch square, the zincs being dimensioned proportionately and from this apparatus with the employment of a single intensifier and rectifier, as above stated, I was able to obtain a constant current of 8 volts.

It should be obvious that the magnets forming one of the electrodes of this apparatus may be permanent or may be electromagnets, or a combination of the two.

While the magnets mentioned throughout the above may be formed of any magnetic substance, I find the best results obtained by the employment of the nickel chrome steel.

While the successful operation of the various devices which I have constructed embodying the present invention have not enabled me to arrive definitely and positively at fixed conclusion relative to the principles and theories of operation and the source from which current is supplied, I wish it to be understood that I consider myself as the first inventor of the general type hereinbefore described capable of producing commercially serviceable electricity, for which reason my claims hereinafter appended contemplate that I may utilize a wide range of equivalents so far as concerns details of construction suggested as preferably employed.

The current which I am able to obtain is dynamic in the sense that it is not static and its production is accomplished without chemical or mechanical action either incident to the actual chemical or mechanical motion or incident to changing caloric conditions so that the elimination of necessity for the use of chemical or mechanical action is to be considered as including the elimination of the necessity for the use of heat or varying degrees thereof.

Back to Viewzone || Comments?

COMMENTS:

The Meyers invention is a simple antenna. The magnets attract and concentrate the electro-magnetic field and the tin horns collect it.

In Meyers' day he only had the natural magnetic field of the earth to harvest. Today we have filled the E-M spectrum with radio, TV and cell phones. If you were to build this device today it would interfere with all of the above. This would make a lot of people upset with you and probably get you in trouble with the government agencies which regulate these utilities.

If you want to harvest free range electricity, I would recommend using a ground antenna like Nathan Stubblefield or a metallic pyramid like Les Brown.

Robert

This alternate energy source is much like a static energy collector designed and patented by Nikola Tesla in 1901. Patent numbers: 685957 and 685958. However, Tesla didn't find it necessary to use magnets, so there is that difference. There are many inventors on YouTube that mimic his design and show great results. I've seen one with a 40' antennae pole with a wire that generates 135 volts.

I'm currently researching the subject and when I saw your blog concerning Meyers I thought I'd give you the knowledge about Tesla's invention that was years prior to Meyers. Scott

Device for energy generation from electric field of atmosphere

Abstract

FIELD: electrical engineering.
SUBSTANCE: proposed device designed for generating electrical energy of desired amount from electric field of atmosphere and for no-break power supply to off-line electrical equipment, such as computerized weather stations and space probes, has electrodes, bearing structure, and atmospheric gas ionizers. Electrodes are spaced along electric field lines of force. Electrodes surrounded by atmosphere are disposed on bearing structure. Outer surface of these electrodes is aligned with atmospheric gas analyzers. Bottom electrode functions as grounding conductor. Load is connected between electrodes spaced along lines of force of atmospheric electric field. In absence of contact between device and the Earth external surface of all electrodes surrounded by atmosphere is physically aligned with atmospheric gas analyzers. Energized current-carrying parts of device are electrically isolated from bearing surface.
EFFECT: reduced mass of device dispensing with moving parts, simplified design and easy shipment of its components.
1 cl, 2 dwg

Images (1)

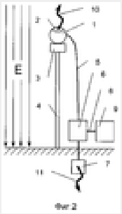

RU2245606C2

RU Grant

Find Prior Art Similar

Other languages: Russian

Inventor: П.А. Кучер (RU), П.А. Кучер, В.И. Коломиец (RU), В.И. Коломиец

Original Assignee: Кучер Павел Алексеевич, Коломиец Валентин Иванович

Priority date : 2003-03-11

Family: RU (1)

Date	App/Pub Number	Status
2003-03-11	RU2003106714A	
2004-09-27	RU2003106714A	Application
2005-01-27	RU2245606C2	Grant

Info: Non-patent citations (2), Similar documents, Priority and Related Applications

External links: Espacenet, Global Dossier, Discuss

Claims (1)

translated from Russian

1. An apparatus for recovering energy from the atmospheric electric field, comprising electrodes spaced along the field lines, the bottom of which is the earth electrode, the supporting structure with encircled atmosphere electrodes ionizers atmospheric gas, structurally combined with the outer surface of these electrodes, characterized in that between spaced along field line electrodes included a load in the absence of the device in contact with the outer surface of the planet atmosphere surrounded electrodes constructive about aligned with the ionization of atmospheric gas, the conductive portion of the device under voltage, electrically insulated from the supporting structure.

Description

translated from Russian

The invention relates to electrical engineering and is intended to ensure uninterrupted autonomous electrical energy, such as automatic weather station or space probes on the Earth and other planets with the atmospheric electric field.

Aware of the difficulties associated with the uninterrupted supply of energy autonomous electrical equipment. Internal power sources (batteries, accumulators, fuel cells) require periodic replacement, recharging or refueling, i.e. qualified service manual.

External sources of energy (flowing water, wind, sunlight) require the use of massive large-sized structures, transmitters with moving parts and are not always available.

To solve the problem let low-power needs, but reliable anytime, anywhere uninterrupted power source. Subject to the conditions of autonomous electrical operation it should be easy, structurally simple and easy to transport to any distance.

Known devices for indirect energy from the atmosphere by water wheels, hydroelectric turbines and other installations, and also to direct energy from the atmosphere by means of wind turbines, wind turbines and other installations. In these devices, atmospheric energy is converted into heat, work or electricity.

Uninterrupted flow generated with their help it is impossible to provide energy. It depends on the state of the atmosphere, climate and time of year. (. Chaly G. Energy yesterday, today and tomorrow - Chisinau: Map Moldovennske, 1977. - 202, IL, pp 44-64,...).

For the prototype adopted diverter, which is designed to neutralize the atmospheric electric field energy. A lightning arrester comprising a vertically oriented supporting structure rising above the terrain, and spaced along the lines of the field electrodes, the collector interconnected. At the top structure is surrounded by an electrode mounted atmosphere - lightning arrester in the form of a massive metal rod. It can be combined with the atmospheric gas ionizer. The lower electrode is an earth electrode and provides an electrical device contact with a surface of the planet (V.V.Bazutkin and other high voltage equipment -.. M .: Energoatomizdat, Textbook for high schools, 1986. - 464 p, yl, pp 219-220...). The design of a lightning conductor does not include a payload. lightning energy is used almost entirely outside for heating the soil surrounding the earthing switch, heating and ionization of the air.

Signs prototype match the essential features of the invention are as follows. A lightning arrester comprising electrodes, a supporting structure and the atmospheric gas ionizers. Electrodes spaced along the electric field lines. Surrounded atmosphere electrodes are arranged on the supporting structure. The outer surface of these electrodes structurally combined with the atmospheric gas ionizers. The lower electrode is an earth electrode.

The reasons impeding obtaining a desired technical result, the prototype, are as follows. Lightning conductor provides air flow through the electric current only in the form of intermittent spark discharge. the passage of this current time is unpredictable, and the value is random and can not be adjusted. Lightning conductor contains no payload. It is designed to work in an abnormally strong electric field, and most of the time useless idle. Lightning conductor does not work in places where there are no storm clouds. A lightning arrester does not work in the absence of contact with the planet. Lightning conductor parts have electrical isolation from the supporting structure.

The technical result - uninterrupted energy from the electric field of the atmosphere in the required amount and for this purpose create the proposed device.

Apparatus favorably with ease, lack of moving parts, ease of required components and their ease of transportation. It is more reliable than all hitherto known independent power supply devices, including wind turbines and solar panels, since the electric field of the atmosphere is weakly dependent on the time of year, there is round the clock and is available anywhere in the world.

The technical result is achieved as follows. Between the electrodes spaced along the lines of the atmospheric electric field is turned on load. In the absence of the device in contact with the outer surface of the planet surrounded electrodes atmosphere structurally combined with the atmospheric gas ionizers. The conductive portion of the device under voltage, electrically insulated from the supporting structure.

The essential features of the claimed invention as follows. An apparatus for recovering energy from the atmospheric electric field comprises electrodes spaced along the electric field lines, the lower electrode is an earth electrode. The electrodes are surrounded by the atmosphere, it is placed on the supporting structure. Ionizers atmospheric gas structurally aligned with the outer surface surrounded by electrodes atmosphere.

Unlike the prototype between spaced along the electric field lines included electrodes load. In the absence of the device in contact with the outer surface of the planet electrodes structurally combined with the atmospheric gas ionizers. Energized live parts of the device are electrically isolated from the supporting structure.

The farther apart the electrodes are spaced, the higher the electric field strength near the surface and more spreading velocity of free charge carriers in the surrounding gas. Surrounded by an atmosphere favorable to place electrodes in the most remote points of the vertical supporting structure. Spreading of charge carriers in the atmosphere promotes free movement of gas at the electrodes and around the device no other stress raisers.

In the presence of the device in contact with the planet earth electrode is a lower electrode. This provides a low electric resistance to the passage of atmospheric electric current through the soil. In the absence of the device in contact with the planet atmospheric electrical current generated by them can pass only through the gas discharge channel. In this case, all the electrodes are surrounded by an atmosphere (both upper and lower) structurally combined with the atmospheric gas ionizers. Thus, a smooth transformation of the atmospheric electric field energy into the energy of electric current flowing through the load. Further useful conversion of this energy into heat, work or electricity by selecting the type of load (heater, motor or other electrical equipment).

Influence of essential features of the claimed technical effect obtained in the following. The supporting structure provides a proper orientation device in the atmospheric electric field coincides with the direction of its lines of force. It keeps spaced along the field lines of the electrodes at the required distance for operation, the device provides mechanical strength and its separate parts united into a whole. Electrodes should move apart from each other by a technically maximum possible distance. Proportional to the distance increases, the electric field at their outer surface, which increases mobility of charge carriers in the atmosphere surrounding the electrodes and facilitates the flow of electrical current therethrough. Mutually remote position of the electrodes eliminates obstacles to free movement of the atmospheric gas streams.

In the presence of the device in contact with the planet earth electrode is a lower electrode. It is the most simple, cheap and reliable way to ensure low resistance to flow through it into the ground electric current. Surrounded atmosphere electrodes are mounted on the supporting structure. This ensures the stability of their spatial position in an external electric field. Ionizers atmospheric gas structurally aligned with the outer surface surrounded by electrodes atmosphere. This ensures the constant presence of a sufficient number of free charge carriers in the space adjacent to the electrodes and the free movement of these charges along the electric field lines. Thus, electrical contact is provided with a low resistance between the solid electrodes and the atmospheric gas. The force transmitted through such contact current over a wide range does not depend on the external electric field and the capacity is determined only by the ionizers. This ensures continuity of energy, regardless of external factors, minimizes power consumption for maintaining the non-self-electric discharge channel through the neutral atmosphere and prevents current surges through the device in a thunderstorm activity.

In the absence of the device in contact with the planet atmospheric gas ionizers structurally aligned with the outer surface of the electrode surrounded by the atmosphere, both upper and lower. This is the only method available to ensure the flow of electric current through the atmosphere on either side of the deployed device along the lines of the field.

The load is connected between the spaced apart along the field lines of the electric field electrodes. This provides serial connection it (together with the electrodes) in a circuit current passing through the electric discharge atmosphere nonself. The product of the voltage drop across the load and the amount of current passing through it determines the useful power received from a device outside the electric field. useful devices coefficient determined by the balance between the load resistance and the transfer resistance of the electrodes with the surrounding space (ionizers capacity). The lower the contact resistance of the electrodes, the greater part of the energy released in the load.

Insulation of live parts of the device from the bearing structure ensures optimal arrangement of the electric field lines in the surrounding area and to prevent the flow of electric current through the conductive parts of the device, unrelated to its circuitry.

Uninterrupted power supply provided by the fact that the electric field of the atmosphere is not dependent on the time of year, there is a clock in all weather, anywhere in the description of the troposphere. This energy resource atmosphere is constantly fueled by all the power of the planetary mechanism of separation of electrical charges.

The reliability of energy supply is provided by the simplicity of the device, in the absence of its most important elements of moving parts. Energy from the atmospheric electric field does not require large structures (dams, towers high altitude) and complex technological methods. Maintenance of such a device is much easier than in prior art.

Simplicity of design elements of the device for obtaining energy from the electric field of the atmosphere arises from the fact that it comprises only standard electrical components without moving parts, requiring no setting adjustment and regular maintenance. Details of apparatus and parts of the supporting structure does not require careful manufacture.

Easy to transport device is achieved in that it is made with a minimum margin of safety, since it is not experiencing dynamic loads and, therefore, for its production does not require massive, large parts and assemblies. Apparatus performed foldable or collapsible.

1 is a perspective view of an apparatus for recovering energy from the electric field of the atmosphere in the absence of contact with the planet.

The apparatus comprises a vertically oriented support structure, such as balloon 1, electrodes 2, 3, the air ionizers 4 and 5, the insulators 6 and 7, the cables 8 and 9, the voltage converter 10 and the load, e.g. radiosonde balloon 11. The surrounding space tested Power outside the electric field lines E. The electrodes are structurally combined with air ionizers are fixed on insulators. The balloon holds the electrodes at a sufficient distance from each other and the cables connecting them to the voltage converter. voltage converter associated with a payload and electrically integrated common housing.

The apparatus operates as follows. After switching on the ionizer 4, 5 saturate the air surrounding the electrodes 2, 3, the free charge carriers. They begin to drift through the neutral air moving along the lines of force of the atmospheric electric field E, is further amplified spaced position electrodes on the balloon 1. Leakage of charge carriers from the electrodes is compensated for constant work ionizers. By nonself gas discharge channels 12 and 13 from the electrodes through the atmosphere there is a constant electric current. Between spaced electrodes a potential difference. They retain its through insulators 6 and 7. When the electrodes at the working potential difference (≈5 kV) voltage switched converter 10. Atmospheric electrical current therethrough closes through cables 8 and 9. The converter transforms the input current into high voltage DC output voltage feeding payload 11. The process continues until there is an atmospheric electric field and act ionizer electrodes.

Determine the electrical power necessary for independent uninterruptible power payload. For example, modern equipment for radiosonde balloon suspended to sufficiently constant electric power of 10 watts. Determine limiting the electrical power required for operation of accessories and air ionizers. For example, in an amount of not more than 150% of the usable, i.e. 15 watts. Considering the operating conditions of current collectors define limiting the potential difference between the exposed conductive parts of the installation. For example, it is recommended to 10 kV and 5 kV actually selected. Calculate the maximum current in the discharge channel. In this case, not more than (10 W + 15 W) / 5 mA = 5 kV. This value specifies ionizers performance, structurally combined with the electrodes. Determine the conduction current density, and the vertical field strength in the troposphere to the calculated flying height. For example, the conduction current is not more than 1.5 mA / m ^{2,} the electric field intensity 2,2-3,5 V / m. Calculate the distance between the electrodes, their shape and structure to ensure effective spreading of charge carriers generated by the ionization of the atmosphere. For example, the distance 50 m, the shape of the receiving electrodes - the ball, the area of each of at least 0.5 m $^{2.}$ A balloon attached and suspension system with insulators electrodes, lay the cable, are mounted in the container with the equipment voltage power converter.

FIG. 2 is a perspective view of an apparatus for recovering energy from the electric field of the atmosphere in contact with the planet apparatus.

The apparatus comprises an electrode 1, ionizer 2, an insulator 3, a supporting structure (rack) 4, the cable 5, the voltage converter 6, the ground connection 7, the connection cable 8 and a payload (weather station) surrounding space 9. tested to the ground surface outside power line electric field E. The electrode is structurally combined with the air ionizer is fixed on the insulator. Stand holds the electrode at a sufficient height, and the cable connecting it to the voltage converter. The voltage converter is electrically connected with the ground through the grounding and the payload via a connection cable.

The apparatus operates as follows:

After switching on the ionizer 2 saturates the air surrounding the electrode 1, free charge carriers. They begin to drift through the neutral air moving along the lines of force of the atmospheric electric field E, is further amplified elevated position on the counter electrode 4. Leakage of charge carriers from the receiving electrode is compensated by regular job ionizer. Via nonself-gas discharge from the electrode 10 through the atmosphere there is a constant electric current. Electrode acquires the electric potential relative to the ground surface and stores it through the insulator 3. When the potential of the working electrode (≈25 kV) is switched voltage converter 6. Atmospheric electrical current therethrough is closed via the cable 5 on the ground 7. The converter transforms the input current at high voltage constant output voltage 27 V, supplied through the connection cable 8 to the power payload 9. The process continues until there is an atmospheric electric field and acts ionizer electrode.

Determine the electrical power necessary for independent uninterruptible power payload. For example, for modern automatic weather station fairly constant electric power of 100 W at rated voltage of 27 V. Determine limiting the electrical power required for operation of accessories and air ionizers. For example, in an amount not more than 50% of the useful, i.e. 50 watts. From the climatic conditions and characteristics of the susceptor structure is determined limiting the potential difference between the exposed conductive parts of the installation. For example, it is recommended not more than 30 kV, and 25 kV actually selected. Calculate the maximum current in the discharge channel. In this case, not more than (100 W + 50 W) / 6 = 25 kV mA. This value sets the performance of the ionizer, structurally combined with the electrode. Measured or found from tables of the electrical activity of the atmosphere (vertical conduction current density and field strength) in the locality. For example, the conduction current is not more than 0.1 mA / m ^{2,} the electric fields 110-250 V / m. Calculate the electrode lifting height above the terrain, its shape and surface working, sufficient for effective spreading of charge carriers generated in the ionizer the surface atmosphere. For example, the height of not less than 10 m, the shape - sphere, area of at least 1 m $^{2.}$ In the calculations take into account the wind speed, the presence of other stress raisers, the geological structure of the soil and other relevant factors. Install the required height rack with an electrode on the insulator, organize ground, mounted auxiliary electric power and voltage converter.

This ensures an uninterrupted supply of electrical energy due to autonomous gratuitous force of the atmospheric electric field. The apparatus advantageously distinguished by the absence of moving parts, ease of required components and their ease of transportation. It is more reliable than all hitherto known independent power supply devices, including wind turbines and solar panels, since the electric field of the atmosphere is weakly dependent on the time of year, there is round the clock and is available at any point of the globe.

Non-Patent Citations (2)

Title
А.П.МЕРКУЛОВ. Магнитные поля – труженики. – М.: Машиностроение, 1978, с. 142-144. *
В.В.БАЗУТКИН и др. Техника высоких напряжений. – М.: Энергоатомиздат. Учебник для вузов. 1986, с. 219-220. *

* Cited by examiner, † Cited by third party

Similar Documents

Publication	Publication Date	Title
US7105940B2	2006-09-12	Mobile renewable energy generator
US20120235410A1	2012-09-20	Lighter than air wind and solar energy conversion system
US20110025070A1	2011-02-03	Utility grid vertical axis wind turbine system
US20120025750A1	2012-02-02	Portable Solar Power System
Ramakumar et al.	1993	Photovoltaic systems
US7478712B2	2009-01-20	Energy collection
US20110100433A1	2011-05-05	System and method for grounding photovoltaic modules
GB2449102A	2008-11-12	Urban power generation platform for use on street lampposts
US4388487A	1983-06-14	Arrangement for preventing the formation of a foreign layer on a high-voltage insulator
US8006933B2	2011-08-30	Airborne power station
Fritts	1967	Special applications of thermoelectric generators
Lings	2005	Overview of transmission lines above 700 kV
DE4205521A1	1993-04-01	Receiving accumulation system for lightning energy - picks up atmospheric thunderstorm

		cloud discharge lightning and stores its energy
US8193760B2	2012-06-05	Portable cart for solar energy conversion
Vasisht et al.	2016	Performance of solar photovoltaic installations: Effect of seasonal variations
CN201562024U	2010-08-25	Test device for ultra-high voltage AC/DC gas-insulated metal-enclosed power transmission line
Djafour et al.	2011	Photovoltaic-assisted alkaline water electrolysis: Basic principles
US20030192586A1	2003-10-16	Space photovoltaic power generation system, portable small power electronic machine, reception antenna apparatus, and electric power system
CN201393183Y	2010-01-27	Area grid-connected type photovoltaic system
US20120118353A1	2012-05-17	Electrical grid solar energy harvesting system
Gomaa et al.	1995	Design and analysis of photovoltaic and wind energy hybrid systems in Alexandria, Egypt
US8045314B2	2011-10-25	Method of atmospheric discharge energy conversion, storage and distribution
MacLeod et al.	2010	Connection of renewable energy sources through grid constraint points using HVDC power transmission systems
KR20100117183A	2010-11-03	A device for solar generation of electric power on water supporting by a weight
CN1657771A	2005-08-24	Height adaptive renewable source of energy generation method and its generating system

Priority And Related Applications

Priority Applications (1)

Application	Priority date	Filing date	Title
RU2003106714A	2003-03-11	2003-03-11	Device for energy generation from electric field of atmosphere

Applications Claiming Priority (1)

Application	Filing date	Title
RU2003106714A	2003-03-11	Device for energy generation from electric field of atmosphere

rexresearch.com

Scalar Fields

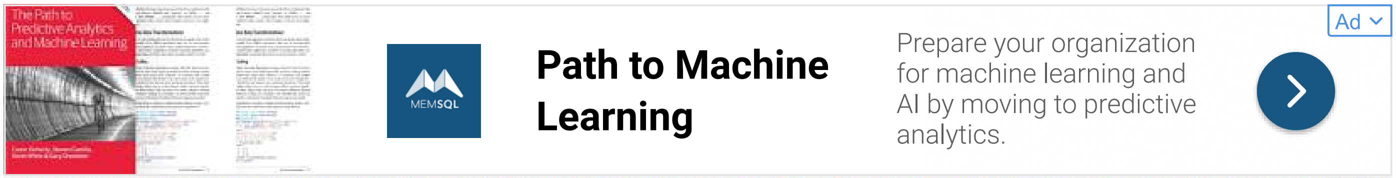
Path to Machine Learning

Prepare your organization for machine learning and AI by moving to predictive analytics.

Ad

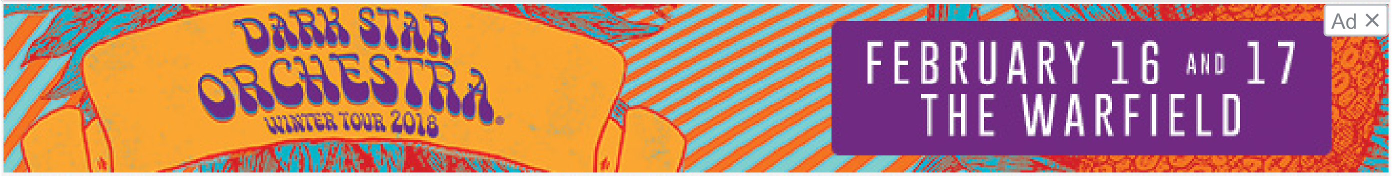
DARK STAR ORCHESTRA WINTER TOUR 2018

FEBRUARY 16 AND 17 THE WARFIELD

Ad ✕

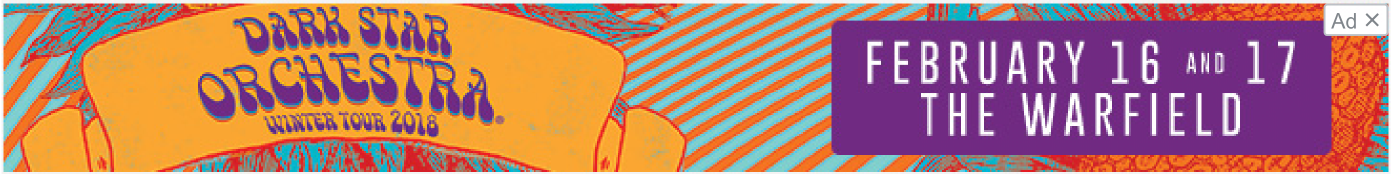
DARK STAR ORCHESTRA WINTER TOUR 2018

FEBRUARY 16 AND 17 THE WARFIELD

Ad ✕

PDF Files :

Capt. Robert Collins : "Soviet Research on Unified Field Theories, False Vacuum States, and Antigravity (U)", Appendix 1 -- Far Field Equations of Maxwell's Equations.

Warren York : "Scalar Wave Technology" ; *Extraordinary Science* (Jan/Feb/Mar 1989).

Thomas Bearden : "Basic Scalar EM Concepts and a Scalar Wave Detector" ; *U.S. Psychotronics Assoc. Newsletter* 5 (4): 4-7.

Jack Dea : "Fundamental Fields and Phase Information" ; *Planetary Association for Clean Energy* 4 (5): 12-15.

J. B. Bronzan : "The Magnetic Scalar Potential" ; *Amer. J. Physics* 39 : 1357-1359 (Nov. 1971).

E.T. Whittaker : "On an Expression of the Electromagnetic Field Due to Electrons by Means of Two Scalar Potential Functions" ; *Proc. London Math. Soc.*, Series 2, 1 : 367-372 (1904).

A. Dolgov & I. Khriplovich : "Propagation of Electromagnetic Waves in a Gravitational Field" ; *Soviet Physics JETP* 58 (4): 671-674 (Oct 1983).

L. Kofman, *et al.* : "Inflationary Universe Generated by the Combined Action of a Scalar Field and Gravitational Vacuum Polarization" ; *Physics Letters* 157-B (5, 6) : 361-366 (25 July 1985).

Gregory Hodowanec : "Do Intense Scalar Fields Affect Life Processes?"

---

http://en.wikipedia.org/wiki/Scalar_field

Scalar Field

In mathematics and physics, a scalar field associates a scalar value, which can be either mathematical in definition, or physical, to every point in space. Scalar fields are often used in physics, for instance to indicate the temperature distribution throughout space, or the air pressure. In mathematics, or more specifically, differential geometry, the set of functions defined on a manifold define the commutative ring of functions.

Just as the concept of a scalar in mathematics is identical to the concept of a scalar in physics, so also the scalar field defined in differential geometry is identical to, in the abstract, to the (unquantized) scalar fields of physics.

Definition

A scalar field is a function from Rn to R. That is, it is a function defined on the n-dimensional Euclidean space with real values. Often it is required to be continuous, or one or more times differentiable, that is, a function of class Ck.

The scalar field can be visualized as a n-dimensional space with a real or complex number attached to each point in the space.

The derivative of a scalar field results in a vector field called the gradient.

Differential geometry

A scalar field on a Ck-manifold is a Ck function to the real numbers. Taking Rn as manifold gives back the special case of vector calculus.

A scalar field is also a 0-form. The set of all scalar fields on a manifold forms a commutative ring, under the natural operations of multiplication and addition, point by point.

Uses in physics

In physics, scalar fields can be used to ascribe forces (which are usually vector fields) to a more general scalar field, the gradient of which describes the force.

* Potential fields, such as the Newtonian gravitational potential field for gravitation, or the electric potential in electrostatics, are scalar fields which describes the more familiar forces.

* A temperature, humidity or pressure field, such as those used in meteorology. Note that when modeling weather on a global basis, the surface of the Earth is not flat, and thus the general language of curvature in differential geometry plays a role. Dopplerized weather radar generates a projection of a vector field onto a scalar field.

Examples in quantum theory and relativity

* In quantum field theory, a scalar field is associated with spin 0 particles, such as mesons or bosons. The scalar field may be real or complex valued (depending on whether it will associate a real or complex number to every point of space-time). Complex scalar fields represent charged particles. These include the Higgs field of the Standard Model, as well as the pion field mediating the strong nuclear interaction.

* In the Standard Model of elementary particles, a scalar field is used to give the leptons their mass, via a combination of the Yukawa interaction and the spontaneous symmetry breaking. This mechanism is

known as the Higgs mechanism [1]. This supposes the existence of a (still hypothetical) spin 0 particle called Higgs boson.

* In scalar theories of gravitation scalar fields are used to describe the gravitational field.

* scalar-tensor theories represent the gravitational interaction through both a tensor and a scalar. Such attempts are for example the Jordan theory [2] as a generalization of the Kaluza-Klein theory and the Brans-Dicke theory [3].

* Scalar fields like the Higgs field can be found within scalar-tensor theories, using as scalar field the Higgs field of the Standard Model [4], [5]. This field interacts gravitatively and Yukawa-like (short-ranged) with the particles that get mass through it [6].

* Scalar fields are found within superstring theories as dilaton fields, breaking the conformal symmetry of the string, though balancing the quantum anomalies of this tensor [7].

* Scalar fields are supposed to cause the accelerated expansion of the universe (inflation [8]), helping to solve the horizon problem and giving an hypothetical reason for the non-vanishing cosmological constant of cosmology. Massless (i.e. long-ranged) scalar fields in this context are known are inflatons. Massive (i.e. short-ranged) scalar fields are proposed, too, using for example Higgs-like fields (e.g. [9]).

Other kinds of fields

* Vector fields, which associate a vector to every point in space. Some examples of vector fields include the electromagnetic field and the Newtonian gravitational field.

* Tensor fields, which associate a tensor to every point in space. For example, in general relativity gravitation is associated with a tensor field (in particular, with the Riemann curvature tensor). In Kaluza-Klein theory, spacetime is extended to five dimensions and its Riemann curvature tensor can be separated out into ordinary four-dimensional gravitation plus an extra set, which is equivalent to Maxwell's equations for the electromagnetic field, plus an extra scalar field known as the "dilaton".

References

1. ^ P.W. Higgs; Phys. Rev. Lett. 13(16): 508, Oct. 1964.
2. ^ P. Jordanm Schwerkraft und Weltall, Vieweg (Braunschweig) 1955.
3. ^ C. Brans and R. Dicke; Phis. Rev. 124(3): 925, 1961.
4. ^ A. Zee; Phys. Rev. Lett. 42(7): 417, 1979.
5. ^ H. Dehnen et al.; Int. J. of Theor. Phys. 31(1): 109, 1992.
6. ^ H. Dehnen and H. Frommmert, Int. J. of Theor. Phys. 30(7): 987, 1991.
7. ^ C.H. Brans; "The Roots of scalar-tensor theory", arXiv:gr-qc/0506063v1, June 2005.
8. ^ A. Guth; Pys. Rev. D23: 346, 1981.
9. ^ J.L. Cervantes-Cota and H. Dehnen; Phys. Rev. D51, 395, 1995.

---

http://en.wikipedia.org/wiki/Scalar_field_theory

Scalar Field Theory

In theoretical physics, scalar field theory can refer to a classical or quantum theory of scalar fields. Such a field is distinguished by its invariance under a Lorentz transformation, hence the name "scalar", in contrast to a vector or tensor field. The quanta of the quantized scalar field are spin-zero particles, and as such are bosons.

No fundamental scalar fields have been observed in nature, though the Higgs boson may yet prove the first example. However, scalar fields certainly do appear in the effective field theory descriptions of many physical phenomena. Because of the relative simplicity of the mathematics involved, scalar fields are often the first field introduced to a student of classical or quantum field theory.

Although scalar fields are Lorentz scalars, they may transform nontrivially under other symmetries, such as flavour or isospin. For example, the pion is invariant under the restricted Lorentz group, but is an isospin triplet (meaning it transforms like a three component vector under the SU(2) isospin symmetry). Furthermore, it picks up a negative phase under parity inversion, so it transforms nontrivially under the full Lorentz group; such particles are called pseudoscalar rather than scalar. Most mesons are pseudoscalar particles.

In this article, the repeated index notation indicates the Einstein summation convention for summation over repeated indices. The theories described are defined in flat, D-dimensional Minkowski space, with (D-1) spatial dimension and one time dimension and are, by construction, relativistically covariant. The Minkowski space metric, has a particularly simple form: it is diagonal, and here we use the + ? ? ? sign convention.

---

http://en.wikipedia.org/wiki/Scalar_field_theory_(pseudoscience)

Scalar field theory (pseudoscience)

For the quantum mechanical "scalar field theory" which is a field theory of spinless particles, see "Scalar field theory"

Scalar field theory (SFT) is a set of theories in a model which posits that there is a basic mechanism that produces the electric field and the magnetic field. Proponents of this theory claim that it advances the standard electromagnetic theory.

Scalar waves in these theories (as opposed to a scalar field in mainstream physics) are hypothetical waves, which differ from the conventional electromagnetic transverse waves by having one oscillation level parallel to the direction of propagation; they thus have characteristics of longitudinal waves. Their existence however, as presupposed in numerous theories, has not been supported by experiment. Scalar waves are called also "electromagnetic longitudinal waves", "Maxwellian waves", or "Teslawellen" (tr., "Tesla waves"). Variants of the theory claim that Scalar electromagnetics (also known as scalar energy) is the background quantum mechanical fluctuations and associated zero-point energies (in contrast to "vector energies" which sum to zero).[citation needed]

Description

Terminology

The basic understanding of scalar field theory begins with several definition of terms within the theory, which are also used in academic physics, but assigns them other meanings. A "scalar field" is a set of assigned observable magnitudes at every point in n-dimensional space (compare this with the standard definition; n is also 4 or greater). [1] An "electric field" is composed of the spinning charged mass, in motion through a finite change in electrostatic scalar potential (compare this with the current academic definition). A "potential" is pure energy and, is any ordering (static or dynamic) in the vacuum (eg., the position of the object relative to other objects). A "scalar potential" is the stationary ordering in the virtual particle flux of the vacuum (compare this with the current academic definition). A "vector potential" is any nonstationary ordering in the virtual particle flux of vacuum (compare this with the current academic definition). Scalar potentials and vector potentials are thus defined as: "contained" inside the energy domain.[citation needed]

Magnetic fields interaction

SFT is based on "non-symmetrical regauging" potentials, demonstrated by the interaction of two magnetic fields.

When the field lines oppose each other, the magnets are pulled together. When the fields are aligned in the same direction, the magnets push apart. When two magnets strongly oppose each other but are not permitted to move apart, the force between them is said to create a "scalar bubble" between the magnets.

The greater the repulsive force, the larger this scalar bubble becomes. As the magnets move away and the pushing force decreases, the scalar bubble shrinks in size and strength.

In a similar manner, two magnets that are strongly attracted create a "scalar void" between them that grows larger the closer the two magnets become. Two magnets powerfully attracted to one another create a very large scalar void, that decreases as the attracting magnets are moved apart.

Despite the claims of its proponents, no repeatable experiments were able to show the existence of the scalar field. All observed effects were shown to comply to the standard physical laws of electrodynamics. The observations are in spectacular agreement not only with classical electromagnetics, but also with quantum electrodynamics, both of which are fields of physics.

Field effects of scalar energy

SFT suggests that scalar energy can move through space much like an electromagnetic wave. However, the operating principles are different. The regular expansion and contraction of a scalar bubble/void is like rhythmicly splashing water on a pond. It sends out ripples through the general scalar field that can subtly affect the size and strength of distant scalar bubbles/voids.

This means that a pair of magnets that are rhythmically opposing/attracting each other are sending out scalar ripples through space that will slightly perturb the scalar bubble/void between a second pair magnets nearby. The net effect is that the attraction and/or repulsion between the second pair of magnets exhibits a change in strength, even though the magnets and fields themselves are motionless.

According to skeptics, the following description given for an application to a communication system reportedly failed to give reproduceable results.

A basic scalar communications system

A scalar communications broadcast antenna does not make any sense according to normal electromagnetic theory. The goal of a scalar broadcast antenna is to create powerful repulsion/attraction between two magnetic fields, to create large scalar bubbles/voids. This is done by using a broadcast antenna with two opposing electromagnetic coils that effectively cancel out as much of each other's magnetic field as possible. An ideal scalar broadcast antenna will emit no electromagnetic field (or as little as possible), since all power is being focused into the repulsion/attraction between the two opposing magnetic fields. Normal electromagnetic theory suggests that since such a device emits no measurable electromagnetic field, it is useless and will only heat up. For a scalar broadcast antenna, any normal RF emission is wasted energy.

A scalar reception antenna similarly excludes normal electromagnetic waves and only measures changes in magnetic field attraction and repulsion. This will typically be a two-coil powered antenna that sets up a static opposing or attracting magnetic field between the coils. The coils are counter-wound so that any normal RF signal will be picked up by both coils simultaneously and effectively cancel itself out, leaving only the scalar component.

Scalar field detection by normal RF antennas

Even though a scalar wave train does not contain the regular EM components that are used by radio frequency communications, it can still be detected by a normal RF antenna, if that antenna is in the presence of some other static magnetic field. When the scalar wave train passes through, it will create a disturbance in the field surrounding that magnet and make the field lines move, which will impart a small electrical current in the standard RF antenna, as if the magnet itself were moved.

Since all normal RF antennas are immersed in the magnetic field of the planet, they can serve as crude scalar detectors, though the reception will be extremely weak and washed out by any normal RF in the vicinity. Detection ability is greatly increased by enclosing the antenna and circuitry in a faraday cage, and by placing a very strong magnet near the antenna inside the cage.

Scalar antennas and detectors

Various proponents have claimed to have developed instruments with characteristics and specifications for different designs. Scalar antenna and detector examples include:

Types [2]

* Magnetostatic Detectors
* Electrostatic Detectors
* Barkhausen Detector

Windings [3]

* single-wire bifilar
* dual-wire bifilar
* pancake bifilar
* cone bifiliar

Proponents of the theory have constructed bifiliar test antennas as isolation transformers. These have taken the form of a ferrite rod, ferrite ring, an air core, or the common square transformer shape.

References

1. ^ T. E. Bearden, "Scalar field". A partial glossary for scalar electromagnetics and subtle phenomena, 1988 (updated 2000).
2. ^ Robert Shannon, "Notes on Scalar Detector Designs". 1996.
3. ^ More Electromagnetism : Scalar Waves". RMCybernetics, 2005.

---

http://www.rense.com/general84/ben.htm

Benjamin Fulford's 'Creature' (?)

From **Brad Steiger**
1-26-9

Absolutely Bizarre 'Creature' In Benjamin Fulford's Spine

On Friday, January 23rd, Benjamin Fulford shocked the audience of the Jeff Rense Program by emailing the claim, accompanied by a photograph, that surgeons had removed a bizarre life form, very much resembling a salamander, from his spine. Those who could view the strange creature were shocked and horrified. Immediately, everyone who saw the "thing" wondered what it was--and if it were real.

Whether the creature was real or whether it was an ill-advised hoax, I wanted to find out if it was possible to place or to inject such a life form into an unsuspecting victim. One of the experts that I queried was a top-research scientist and inventor, who quickly responded that not only was it possible to inject the embryo of a parasite into a targeted victim without his or her knowledge, but that in the world of deep black operations, Russian scientists had already perfected such grotesque methodology.

I asked the scientist to come on the Rense Program and discuss the science of bioenergetics being developed for Black Ops projects, but he informed me that he had been placed under "very strict warnings" not to make any appearances on any radio or television program. What he did do was to write a very informative email on the incredible world of bioenergetics and psychoenergetics which I reproduce below with his special permission. At the present time, he prefers to remain anonymous.

Yes, Brad, the "embryo" for such a "new living parasite critter" could indeed be developed and inserted into a targeted victim, including at a great distance, but only by the highly "clandestine" world of the off-budget deep black operations -- particularly those in Russia, which are still far ahead of anything in the West.

The Russian science of "energetics" -- developed shortly after WW II by a tremendously massive Soviet effort of some of the best nonlinear scientists on earth -- did solve some of the perplexities of our modern physics, of psychology and biology, etc. And under Stalin's iron boot, the new and spectacular Russian scientific developments were immediately placed into extended weapons development.

To this day, our own more-materialistic scientists still do not really comprehend Soviet energetics. (And they also still do not even know what a mind actually is, what a "dead" (separated) mind is, or what a "mind and body linked living system" really is).

Soviet energetics had three branches: (1) the science of inert matter, called by the same name "energetics". (2) the science of living biological systems, where a mind is linked to a living body. This branch is called "bioenergetics". (3) the science of the mind and mental phenomena, which is called "psychoenergetics".

Western science, being so materialistic, has not yet comprehended the most important thing about "mind": that is, that the mind and mental things exist in time, but not in 3-space.

If one turns to modern quantum field theory, then there are now FOUR photons available, not the two transverse photons so well known to classical electrodynamicists etc. In the following discussion, the travel direction of the photon in 3-space is assumed along the Z axis. The "regions of vibration" of the photon -- i.e., the type and direction of its oscillation -- may occur in four ways. These are (1, 2) the two "orthogonal" oscillation polarizations, one along the X-axis and the other along the Y-axis. Then there is (3) the longitudinal (i.e., longitudinally-polarized) photon, which oscillates along the Z-direction of travel itself. And finally there is (4) the time-polarized or "scalar" photon, which oscillates along the time axis. These four photons are all in standard quantum field theory, even in the West.

Now note that only the two transverse-polarized photons (1) and (2) are observable by our instruments. The longitudinal (3) photon is nonobservable, as is the time-polarized (4) photon.

Here's the first "magic":

A combination of a longitudinal photon (3) and a time-polarized photon (4) is OBSERVABLE AS A SPIKE OF VOLTAGE -- more precisely, a spike of electrostatic scalar potential.

Now suppose you have a hidden (nonobservable) electromagnetic "system" you have built in the time-domain, so that it is built only of time-polarized photons. The system is not observable, nor is any part of it observable.

But then suppose you also build a corresponding set of operations (a "system") in 3-space matter so that this new EM system consists only of longitudinal photons. And suppose you have made this "longitudinal" photon system's operations so that they exactly parallel the time-domain system's operations. In short, you now can superpose your longitudinally-polarized EM wave system and the time-polarized EM wave system. And your instruments will now observe the combination -- i.e., the actual ongoing linkage system's operation -- as a very strange but highly organized system of voltage spikes!

The West has totally missed the fundamental characteristic of mind: Mind occupies time but not 3-space. Hence, mind is a totally "electromagnetic" system, but it is comprised of only non-observable time-polarized photon operations!

Well, how could you possibly couple that mind to a biological (3-space) material body?

You would have to have a parallel longitudinally-polarized set of electromagnetic operations ongoing in that 3-space body.

And the direct coupling of those two "operational systems" would be the mechanism of the coupling of the mind to the body -- and it would and could be directly observed by our instruments as a correlated system of voltage spikes!

And that is why the dendrite cells of the body all have so many endings and "continual and incredible voltage spikings"! In looking at the entire set of spiking, you are looking at the "coupling mechanism" (that couples the mind to the body) in action! That set of spikings is what links or couples (coherently) the living temporal mind to the living biological body!

Now if you place great scientific effort on that area and in that fashion, for some decades, you will eventually be able to (1) directly produce an operating living mind in the laboratory, including a human mind and any version of it, and (2) link it to a properly designed biological body (to a proper set of longitudinal operations therein), and this will give you a living biological system created directly in the laboratory.

In about 1990, the Russian hidden psychoenergetics science had progressed to the point that living minds were being designed and produced in the laboratory, including various "versions" (personalities) of a given mind.

If one knows where to look, there is and has been quite a bit of external "testing" by the Soviets to stimulate the West and see if we recognize what is being done (in short, to see whether the West also has developed such science). Sadly, the provocative tests have clearly shown the Soviets that we do not have such capabilities or such scientific development. Indeed, we still do not even recognize that "testing" is even going on!

One of the New Energy scientists -- Tom Bearden -- stumbled onto a portion of this Soviet energetic work. In his 2005 book, Oblivion: America at the Brink, Bearden gave an abbreviated presentation of the development of the energetics weaponry and some of the actual "evocative testing" examples. There are many more weapons and examples that he did not run across.

Presently, the Russians can clandestinely duplicate the mind of a targeted individual, in the duplicate mind they can "alter" portions to produce "different behavior" if that alternate mind is inserted into the targeted individual in the West, and then they can do several things. (1) they can just leave it in that individual as an extra "multiple personality," so they have the 'perfect spy" that hears everything the person's ears hear, sees everything his eyes see, and also receives all the thoughts that his true mind thinks. And they can put that on the TV screen in Moscow if they wish. With the portion for which he actually encountered evidence, Bearden gave examples of the military use of such capabilities in Oblivion.

Sadly, the West still thinks that the human "mind" is just some electron wiggles in the brain, and that we are like a giant "nonliving" computer.

There is also a way to directly engineer any physical observable system one wishes, in any part of it, living or nonliving. One can directly engineer and change inert matter or living matter, etc. I have learned that Bearden plans a future book (hopefully this year) dealing with this scientific methodology, which he calls "precursor engineering", but which the Soviets still have embedded in its term "energetics" with its associate science (and weapons).

Once you understand something about this 3-fold secret weapons science of energetics that the Soviets have highly developed, then you can understand how -- with such developed technology -- one could directly build both the physical body system (its correlated longitudinal EM photon system) and the "mind system" of such a "living critter" one creates in the laboratory, and how one could then use such capability (such weird and artificially-created critters) by sudden surprise on an unsuspecting enemy to overwhelm and crush him almost instantly.

As you can see, we are referring to a gigantic leap forward in what the West calls "biological weapons" and in "biological warfare".

Well, one should be looking for "very strange incidents" suggesting the development of such "critters" and parasites. And here is at least one "candidate" occurrence.

This present "incredible tumor" discovery [in Benjamin Fulford's spin] may be the discovery of one "provocative test" of these capabilities, just to see what our western science and biological warfare experts make of it.

And if it is a "provocative test" to see what we know, of if we know anything of it at all, we have certainly once again assured them we have not the foggiest notion of the true reach of Soviet energetics science.

WO2008092205
METHOD AND APPARATUS FOR ANALYSING GEOLOGICAL FEATURES

Inventor(s): DUNCAN ANDREW [AU]
Classification: - international: G01V3/00; G01V3/15; G01V3/16; G01V3/17; G01V3/00; G01V3/15 - European: G01V3/165
Abstract -- An apparatus (10) for analysing geological features comprises a receiver (20) for measuring a magnetic field received from adjacent geological features (18) excited by a periodic transmitted electromagnetic signal, wherein the measured magnetic field is a scalar amplitude of the magnetic field or a scalar amplitude of the magnetic field is derivable from the measured magnetic field, wherein the receiver generates a received signal from the measured magnetic field; and a processor (28) for filtering unwanted signal components which are substantially synchronous with the periodic transmitted electromagnetic signal from the scalar amplitude of the received signal or the scalar amplitude derived from the received signal, such that target geological features are able to be analysed using the filtered scalar amplitude.

WO2008060213
METHOD FOR PREDICTING WHERE THE NEXT MAJOR EARTHQUAKE WILL TAKE PLACE WITHIN AN AREA

Inventor(s): SLUNGA RAGNAR [SE]
Classification: - international: G01V1/28; G01V1/00; G01V1/28; G01V1/00- European: G01V1/00E
Also published as: WO2008060213 // SE0602417 (A) // SE530569 (C2)
Abstract -- The present invention relates to a method of predicting where the next major earth- quake will occur within a area based on knowledge of the stress tensor field in the area, including determining stress tensors that have caused a shear slip in the form of an earthquake. It is first assumed that said first shear slip is the only one that is not stable according the Mohr-Coulomb slip criterion applied to contenplated fault planes with all conceivable orientations and calculating according to the Mohr-Coulomb slip criterion the principal stress directions as a function of the friction coefficient f. After that, it is established according to the Mohr-Coulomb slip criterion a relationship between two of the principal stresses.; Moreover the normal stress sv in a known direction Sv is determined and, according to the elasticity theory, a relationship between the normal stress sv and the principal stresses is established. Then expressions of the three principal stresses as a function of a scalar parameter is established, and a function of the elastic deformation energy per unit of volume relative to an isotropic reference stress state with the pressure sv based on the expressions of the principal stresses is established. Finally, the remaining degree of freedom is eliminated by determining the value of said scalar parameter which minimises the function of the elastic deformation energy and the value of the scalar parameter in the expressions of the principal stresses is inserted.

WO2008050137
MAGNETIC SIGNATURE ASSESSMENT

Inventor(s): TWELVETREES ROGER [GB]; COX EMILY [GB]; GANDERTON CARL [GB]; RAWLINS PAUL [GB]; WATSON STEPHEN MILES [GB]

Classification: - international: B63G9/06; B63G9/00 - European: B63G9/06
Also published as: GB2443265 (A)
Abstract -- Magnetic signature assessment apparatus for a vehicle comprising sensors for incorporation in the vehicle to measure the magnetic field normal to a closed surface at least approximately bounding the vessel and processing means for calculating from the normal field measurements a scalar magnetic potential outside the surface. Apparatus may be provided on the vehicle to generate a magnetic field to suppress the magnetic signature corresponding to the scalar potential. The invention also extends to corresponding methods and to programs for implementing those methods.

RU2005105859
LONGITUDINAL-SCALAR ELECTROMAGNETIC WAVE RADIATING DEVICE

Applicant(s): KUZNETSOV JURIJ NIKOLAEVICH
Classification: - international: H01Q13/02; H01Q13/00
Also published as: RU2287212
Abstract -- FIELD: radio communications and radiolocation, geological prospecting, and medicine. ^ SUBSTANCE: proposed radiating device is provided with second waveguide and second exciting component; exciting components are installed in waveguides which have different length and are brought to single plane through their open ends and further through return horn to equally shaped adding waveguide terminating in horn. Waveguides differ by half-length of common radiated electromagnetic wave. In planar-wave approximation strength vectors of electric and magnetic fields are oriented in longitudinal-scalar electromagnetic wave in open space collinearly to vector of electromagnetic energy flux density (S = Ei ¦Hi¦ + Hi ¦Ei¦).Irrotational electric field of longitudinal electromagnetic wave builds up potential difference across open electric conductor longitudinally oriented to this field.; Magnetic field of longitudinal-scalar electromagnetic wave acts upon moving electric charges by force directed along their velocity vector. ^ EFFECT: enhanced reliability due to use of irrotational longitudinal-scalar electromagnetic waves apart from rotational cross-vector ones.

US2005197808
Method for determining electrical and magnetic field effects

Inventor(s): KUO, AN-YU
Classification: - international: G06F17/10; G06F17/10; (IPC1-7): G06F17/10
Abstract -- A method for determining electrical and magnetic field effects determines Lanczos matrices by performing a preconditioned conjugate gradient method using a nested multi-grid, vector and scalar potential preconditioner so that Pade via Lanczos frequency expansion may be used to determine the electrical and magnetic field effects over a frequency range without having to perform computationally slow and memory intensive matrix decomposition.

US2004046553
Vector Measurement of a Magnetic Field

Inventor(s): LEGER JEAN-MICHEL [FR]; GRAVRAND OLIVIER [FR]; BERTRAND FRANCOIS
Classification: - international: G01R33/24; G01R33/24; (IPC1-7): G01V3/00 - European: G01R33/24
Also published as: US6844726 // FR2815416 // FR2815416 // WO0233434 // EP1344075
Abstract -- A vectorial magnetometer (1), measures the components of a magnetic field in three directions (Oxyz) using a scalar magnetometer (2). The field is periodically modulated in each of the directions by generators (Gx, Gy, Gz) which have a specific frequency for each direction and that power

coils (Ex, Ey, Ez). Synchronous demodulation of the of the output signal of the scalar magnetometer (2) for each of the three frequencies permits the relative continuous component of each axis to be found.; The vectorial magnetometer (1) is characterised in that it has means (Dx D'x, Dy D'y, Dz D'z) that can carry out a double demodulation for phase and quadrature for each of the frequencies and processing means (70) that use the continuous component modules for phase and quadrature to calculate a transfer function of the scalar magnetometer at the frequency in question, and to apply this function to the correction of the components.

EP0964260
Device for Measuring Magnetic Field Components Comprising a Scalar Magnetometer

LEGER, JEAN-MICHEL
Classification: - international: G01R33/26; G01R33/24; (IPC1-7): G01R33/26 - European: G01R33/26
Also published as: FR2779530 // US6313628 // CA2274623
Abstract -- The system uses a scalar magnetometer with additional fields applied by external coils, which are arranged with mutually orthogonal axes. The system for measuring the components of a magnetic field includes a scalar magnetometer (10,14,16,21,24,26,30,40,46,56). This magnetometer provides an output signal corresponding to the modulus of the applied magnetic field. There are also at least two conducting coils (Ex,Ey,Ez) disposed around the scalar magnetometer with mutually orthogonal axes. Each coil is supplied with a current at a given frequency. Processing circuits (Dx,Dy,Dz) receive the signal produced by the magnetometer and effect synchronous demodulation of the signal at the respective frequencies, to provide signals (Bx,By,Bz) corresponding to the components of the magnetic field which exists in this region.

JP11306328
METHOD AND DEVICE FOR GENERATING TWO-DIMENSIONAL SCALAR FIELD

KAWAI, NAOKI
Classification: - international: G03F1/00; G06T1/00; G03F1/00; G06T1/00; (IPC1-7): G06T1/00; G03F1/00
Abstract -- PROBLEM TO BE SOLVED: To extract flow of each conduit from a conduit picture to generate a two-dimensional scalar field. SOLUTION: Conduits are extracted from the conduit picture one by one (S2), and representative points and values of flow are defined for each conduit, and respective values of flow are registered in positions corresponding to these representative points in a separately prepared two-dimensional area, and values of flow in other positions of the two-dimensional area are interpolated based on registered values of flow and their positions (S5), thus generating a two-dimensional scalar field.

US5845220
Communication method and apparatus with signals comprising scalar and vector potentials without electromagnetic fields

PUTHOFF, HAROLD
Classification: - international: H01Q7/00; H04B5/00; H01Q7/00; H04B5/00; (IPC1-7): H04B1/00
- European: H01Q7/00; H04B5/00
Abstract of US 5845220 (A)
Information that changes as a function of time is communicated from a transmitting site to a receiving site

by transmitting a signal comprising scalar and vector potentials without including ay electromagnetic field. The potentials vary as a function of time in accordance with the information.

US3924210
Field Shaping Magnet Structure

1975-12-02
Inventor(s): DIONNE NORMAN J
Classification: - international: H01J31/50; H01F7/02; H01F7/20; H01J29/64; H01J31/08; H01F7/02; H01F7/20; H01J29/58; (IPC1-7): H01F7/02- European: H01F7/02C1; H01F7/20
Also published as: NL7512795 // NL7512795 // NL172496 // NL172496 // JP51067961

Abstract -- A magnet structure having a pair of spaced pole pieces interconnected by a reluctor circuit comprising an operatively aligned array of soft magnetic members suitably spaced apart by interposed nonmagnetic material to provide between the pole pieces a preferred magnetic scalar potential and an associated magnetic flux distributed as desired over a relatively large volume adjacent the array.

---

http://amasci.com/hum/adetect.txt

Date: 10 Feb 97 18:50:04 EST
From: "Sara T. Allen"
To: Taos Hum list serve <taoshum-l@eskimo.com>

Subject: Scalar fields - Honeywell Patent Numbers

List of patents by Raymond C. Gelinas but assigned to Honeywell:
(note that a Vector Potential is a scalar field)

4,429,280
31 Jan 1984
Apparatus and Method for Demodulation of a Modulated Curl-Free Magnetic Vector Potential.

4,429,288
31 Jan 1984
Apparatus and Method for Modulation of a Curl-Free Magnetic Vector Potential Field.

4,432,098
14 Feb 1984
Apparatus and Method for Transfer of Information by Means of a Curl-Free Magnetic Vector Potential Field.

4,447,779
8 May 1984
Apparatus and Method for Determination of a Receiving Device Utilizing a Curl-Free Magnetic Vector Potential Field.

4,605,897
12 Aug 1986
Apparatus and Method for Distance Determination Between a Receiving Device and a Transmitting Device Utilizing a Curl-Free Magnetic Vector Potential Field.

4,491,795
1 Jan 1985
Josephson Junction Interferometer Device for Detection of Curl-Free Magnetic Vector Potential Fields.

Also, you can find a good article in Scientific American of April 1989, pp. 56-62, "Quantum Interference and the Aharonov-Bohm Effect" by Yoseph Imry & Richard Webb.

Another one:

US5845220: Communication method and apparatus with signals comprising
scalar and vector potentials without electromagnetic fields , H. Puthoff

---

http://www.unusualresearch.com/scalarbib/scalarbib.htm

Y. Aharonov and D. Bohm, "Significance of Electromagnetic Potentials in in the Quantum Theory," The Physical Review, vol. 115, no. 3, Aug. 1959.
Abstract: In this paper, we discuss some interesting properties of the electromagneticpotentials in the quantum domain. We shall show that, contrary to theconclusions of classical mechanics, there exists effects of potentials oncharged particles, even in the region where all the fields (and therefore theforces on the particles) vanish. We shall then discuss possible experiments totest these conclusions; and, finally, we shall suggest further possibledevelopments in the interpretation of the potentials.

Yoseph Imry and Richard A. Webb, "Quantum Interference and the Aharonov-Bohm Effect," Scientific American, vol. 260, no. 4, Apr. 1989.
Abstract: Can electrons be influenced by a nearby magnet so well shieldedthat its force field cannot be detected? The counterintuitive answer is yes:an energy emanation from the magnet known as the potential does indeed affectthe electrons' wave function. This quantum-mechanical effect is being brought tobear on the development of new microelectronic devices.

Capt. Robert M. Collins (TQTR), "Soviet Research On The A-Vector Potential and Scalar Waves (U)," Unknown.
Abstract: Active in the areas of the Aharonov-Bohm effect as applied to the A-vector potential and scalar fields as applied to solving force related problems.

Capt. Robert M. Collins (TQTR), "Soviet Research On Unified Field Theories, False Vacuum States, and Antigravity (U)," Unknown.
Abstract: Theoretical progress in dealing with unified field theories...new concepts in weapons, transportation, propulsion

Dr. Jack Dea, "Fundamental Fields and Phase Information," P.A.C.E. Newsletter, vol. 4, no. 3 (Planetary Assoc. f. Clean Energy)

K.J. van Vlaenderen, "Electrodynamics with the scalar field," Sep. 2001.
Abstract: The expressions of the electric and magnetic fields, that can be derived from the Li'enard-Wiechert potentials, are in good agreement with experiments. Surprisingly the Li'enard-Wiechert potentials do not satisfy the Lorenz gauge condition, therefore they cannot be regarded as solutions of the Maxwell equations in the Lorenz gauge, but only as solutions of the inhomogeneous potential wave equations. Therefore, the inhomogeneous potential wave equations should be regarded as a separate set of differential equation that are unconnected with Maxwell's equations. In order to simplify this theory of classical electrodynamics we propose a generalization of Maxwell's equations, such that a gauge of the potentials is unnecessary in order to derive the Lorenz inhomogeneous potential wave equations. This generalization can be described as a conditional current regauge that does not violate the conservation of charge. This has several consequences: - the generalized Maxwell equations also contain scalar field terms. - the prediction of a longitudinal electro-scalar wave (LES wave) in vacuum. - a generalized Lorentz force expression that contains an extra scalar term. - generalized energy and momentum theorems, with an extra power flow term associated with LES waves.

E. T. Whittaker in Cambridge, "On the partial differential equations of mathematical physics," vol. 57, pp. 333-355, Nov. 1903.
Although as much care as possible was taken in its reproduction, no warranty is expressed or implied as to the suitability or accuracy of this document for any purpose whatsoever.

CAVEAT: There are what appear to be errors in the original. Those that are changed are noted here. Spelling and other obvious errors remain as in the original.
On Page 346: The first double integral was missing the 'x' on the 'x sin u cos v' term of the first argument of f().
Frank Lofaro gives the 1903 paper this description: 1903 paper on potentials being composed of harmonic phase-conjugate scalar pairs.
He also gives the 1904 paper this description: 1904 paper on EM waves being composed of two scalars.

E. T. Whittaker in Cambridge, "The Electromagnetic Field Due to Electrons by Means of Two Scalar Potential," pp. 367-372, Nov. 1904.
Although as much care as possible was taken in its reproduction, no warranty is expressed or implied as to the suitability or accuracy of this document for any purpose whatsoever.
CAVEAT: There are what appear to be errors in the original. Those that are changed are noted here. Spelling and other obvious errors remain as in the original.
The equation in the middle of page 368 appears to be in error in the original. The "c-square, del-square, a-sub y" term is missing the superscript 2 on the del in the original paper. From looking at the other equations in the group, it seems like it should be squared also.

Concetto R. Giuliano, "Applications of optical phase conjugation," Physics Today, vol. 34, no. 4, pp. 27-35, Apr. 1981.
Abstract: Light waves that are, in effect, time-reversed images of their original can serve to restore severely aberrated waves to their original state.

Dr. Harry E. Stockman, "Plasma-diode experiments," Ham Radio, Feb. 1980.

GSC-12645, "Precise Phase Comparator for Nearly Equal Frequencies," NASA Tech Briefs, Fall/Winter 1981, vol. 6, no. 3, Nov. 1981.

Dwight W. Batteau and Peter R. Markey, "Man/Dolphin Communication Final Report," Dec. 1966.
Abstract: Research intended to determine the feasibility of establishing a language, approaching English, between man and dolphin. Neurophone like circuit in Appendix-A.

http://www.eskimo.com/~billb/freenrg/bark.html

http://www.textfiles.com/bbs/KEELYNET/ENERGY/reply2.asc

http://www.textfiles.com/bbs/KEELYNET/ENERGY/replytvq.asc

http://www.textfiles.com/bbs/KEELYNET/BIOLOGY/health1.asc

http://www.textfiles.com/bbs/KEELYNET/BIOLOGY/warning1.asc

Barkhausen Effect Scalar Detector

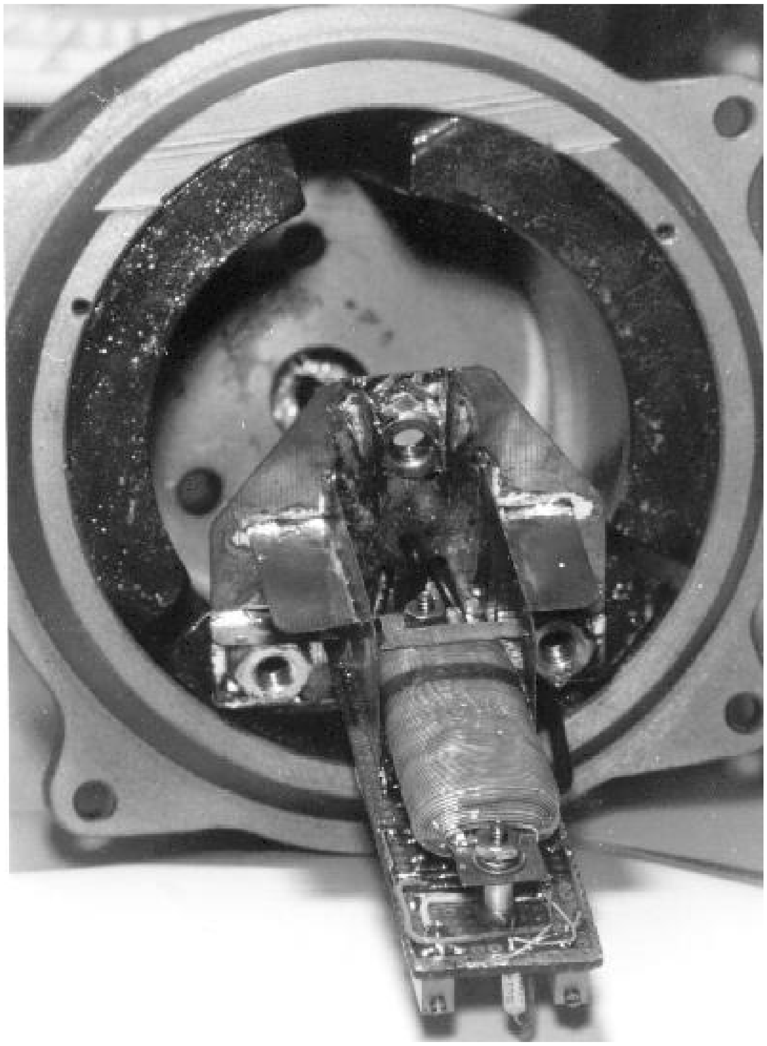

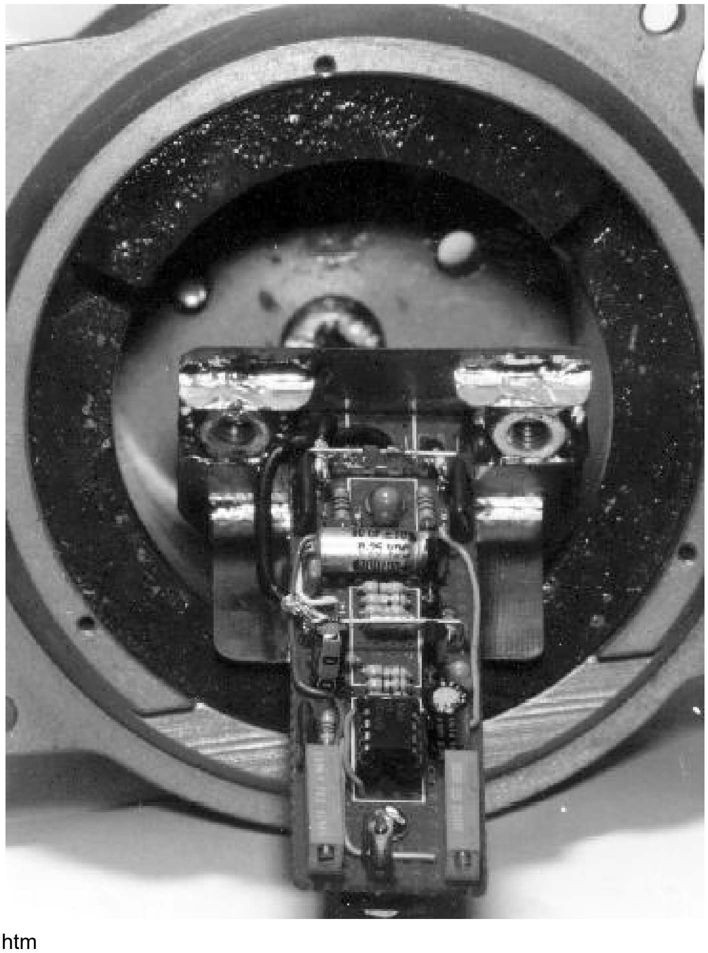

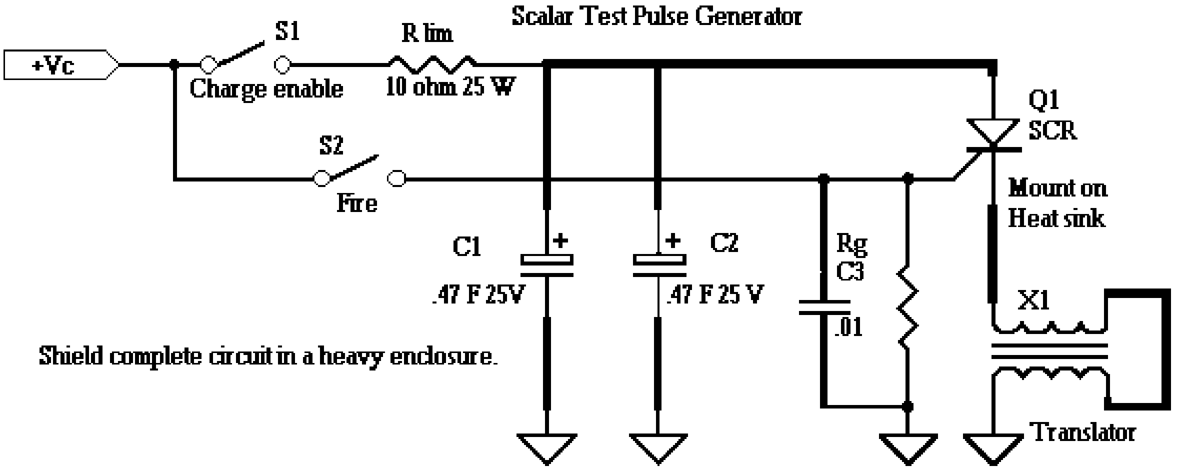
Scalar Test Pulse Generator
+Vc
S1
Charge enable
R lim
10 ohm 25 W
Q1
SCR
S2
Fire
Mount on
Heat sink
C1
.47 F 25V
C2
.47 F 25 V
Rg
C3
.01
X1
Shield complete circuit in a heavy enclosure.
Translator

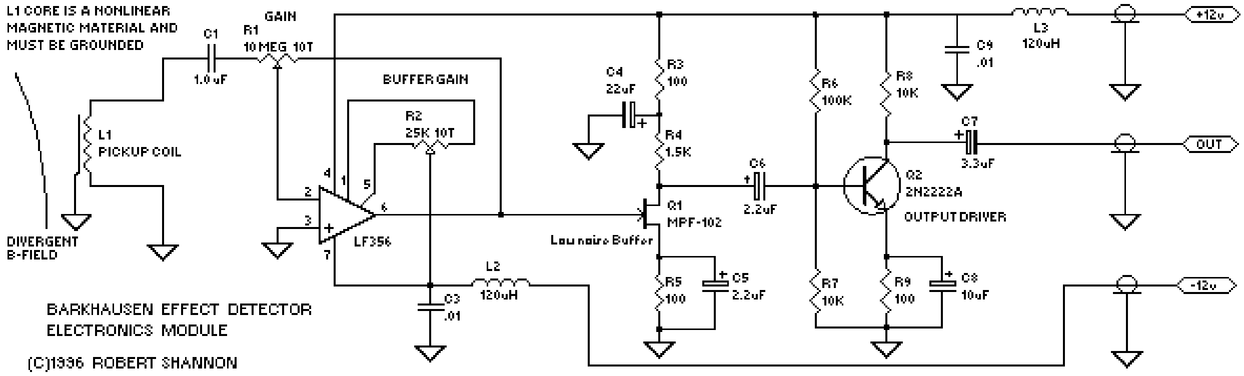
L1 CORE IS A NONLINEAR
MAGNETIC MATERIAL AND
MUST BE GROUNDED
C1
1.0 uF
R1
10 MEG 10T
GAIN
BUFFER GAIN
R2
25K 10T
L1
PICKUP COIL
DIVERGENT
B-FIELD
4
1
5
2
3
6
7
LF356
C3
.01
L2
120uH
BARKHAUSEN EFFECT DETECTOR
ELECTRONICS MODULE
(C)1996 ROBERT SHANNON
C4
22uF
R3
100
R4
1.5K
Q1
MPF-102
Lounaire Buffer
R5
100
C5
2.2uF
C6
2.2uF
R6
100K
R7
10K
Q2
2N2222A
OUTPUT DRIVER
R8
10K
R9
100
C8
10uF
C7
3.3uF
C9
.01
L3
120uH
+12v
OUT
-12v

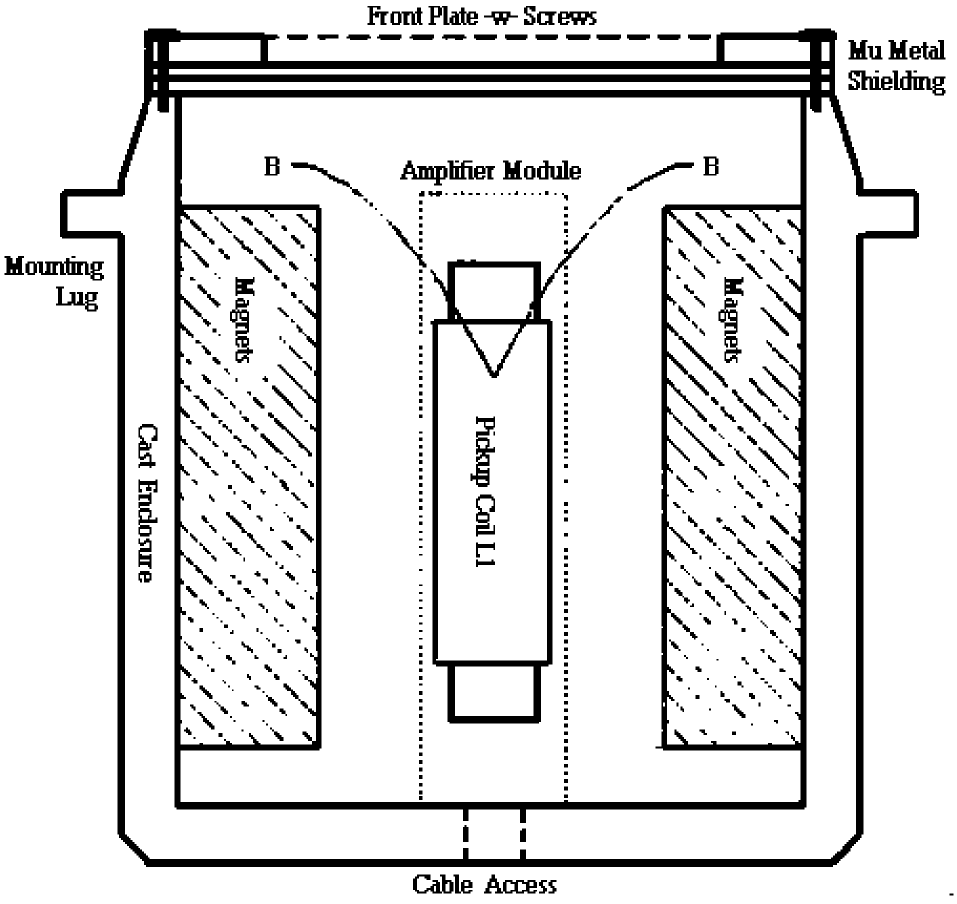
Front Plate -w- Screws
Mu Metal
Shielding
B
Amplifier Module
B
Mounting
Lug
Magnets
Magnets
Pickup Coil L1
Cast Enclosure
Cable Access

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

**... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !**

ORDER PAGE

US007960867B2

(12) **United States Patent**
Issa et al.

(10) **Patent No.:** **US 7,960,867 B2**
(45) **Date of Patent:** **Jun. 14, 2011**

(54) METHODS AND SYSTEMS FOR WIRELESS ENERGY AND DATA TRANSMISSION

(75) Inventors: **Anthony Francis Issa**, Orlando, FL (US); **Jeffrey Messer**, San Bernadino, CA (US); **John Michael Tobias**, Toms River, NJ (US)

(73) Assignee: **Extremely Ingenious Engineering**, Orlando, FL (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 266 days.

(21) Appl. No.: **12/288,586**

(22) Filed: **Oct. 21, 2008**

(65) **Prior Publication Data**

US 2009/0134711 A1 May 28, 2009

Related U.S. Application Data

(60) Provisional application No. 61/004,373, filed on Nov. 27, 2007.

(51) **Int. Cl.**
H01F 27/28 (2006.01)
H01F 38/00 (2006.01)

(52) **U.S. Cl.** **307/104**
(58) **Field of Classification Search** 307/104
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

514,168	A	2/1894	Tesla
593,138	A	11/1897	Tesla
645,576	A	3/1900	Telsa
649,621	A	5/1900	Tesla
685,953	A	11/1901	Tesla
685,954	A	11/1901	Tesla
685,955	A	11/1901	Tesla
685,956	A	11/1901	Tesla
685,957	A	11/1901	Tesla
685,958	A	11/1901	Tesla
787,412	A	4/1905	Tesla
1,119,732	A	12/1914	Tesla
2,205,204	A	6/1938	King
3,432,664	A	3/1969	Robison
3,758,869	A	9/1973	Wagner
3,781,647	A	12/1973	Glaser
3,909,736	A	9/1975	Huchital et al.
4,379,253	A	4/1983	Myer
4,485,426	A	11/1984	Kerls
4,685,047	A	8/1987	Phillips, Sr.
4,717,889	A	1/1988	Engelmann
4,727,297	A	2/1988	Wolze
4,872,100	A	10/1989	Diaz

(Continued)

OTHER PUBLICATIONS

International Preliminary Report for PCT/US2008/007655 dated Dec. 22, 2009.

(Continued)

Primary Examiner — Jared J Fureman
Assistant Examiner — Hal I Kaplan
(74) *Attorney, Agent, or Firm* — Milbank Tweed Hadley & McCloy LLP

(57) ABSTRACT

System and methods for wireless energy and data transmission are described. One aspect includes an emitter and receiver. The emitter may include primary and secondary emitting windings. The receiver may include primary and secondary receiver windings. In further aspects, the emitter may include a tertiary emitting winding and the receiver may include a tertiary receiver winding. In one aspect, the emitter uses driving circuitry and the receiver uses rectifying circuitry and one or more capacitors. In further aspects, the emitter and receiver are reversible.

39 Claims, 16 Drawing Sheets

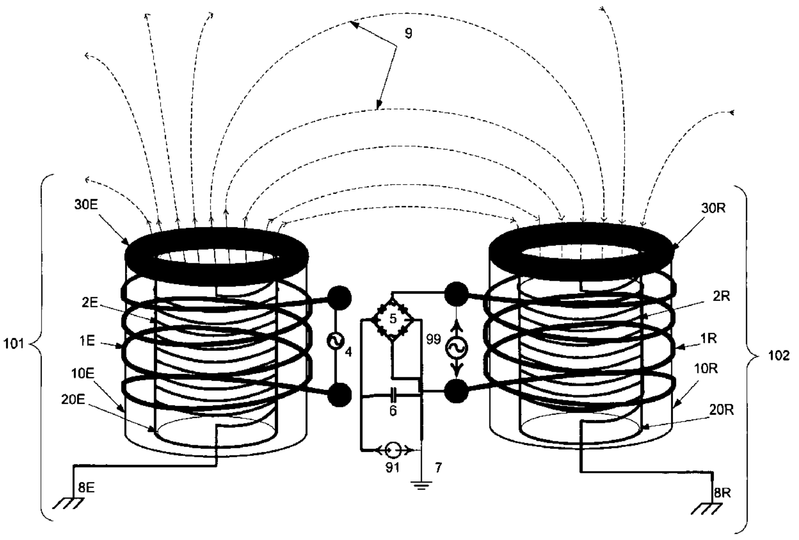

U.S. PATENT DOCUMENTS

4,916,379	A	4/1990	Wand et al.	
4,937,832	A	6/1990	Rocca	
4,945,721	A	8/1990	Cornwell et al.	
4,956,579	A	9/1990	Albright	
4,963,792	A	10/1990	Parker	
5,103,366	A	4/1992	Battochi	
5,173,643	A	12/1992	Sullivan et al.	
5,225,623	A	7/1993	Krasnow	
5,276,281	A	1/1994	Sato et al.	
5,281,898	A	1/1994	Albright	
5,739,997	A	4/1998	Gross	
5,818,180	A	10/1998	Canclini	
5,841,239	A	11/1998	Sullivan et al.	
5,864,212	A	1/1999	Sullivan	
6,052,017	A	4/2000	Pidutti et al.	
6,118,229	A	9/2000	Lee	
6,166,869	A	12/2000	Pidutti et al.	
6,198,335	B1	3/2001	Pakriswamy	
6,259,305	B1	7/2001	Pakriswamy	
6,272,694	B1	8/2001	Weaver et al.	
6,320,508	B1	11/2001	Giesler et al.	
6,522,089	B1	2/2003	Duong et al.	
6,549,044	B2	4/2003	Brambilla et al.	
6,798,716	B1	9/2004	Charych	
6,883,509	B2	4/2005	Porter et al.	
6,906,486	B2	6/2005	Berroth et al.	
6,906,495	B2	6/2005	Cheng et al.	
6,911,848	B2	6/2005	Vinciarelli	
6,930,893	B2	8/2005	Vinciarelli	
6,934,166	B2	8/2005	Vinciarelli	
6,961,227	B1	11/2005	Whiton et al.	
7,027,311	B2	4/2006	Vanderelli et al.	
7,235,945	B2	6/2007	Correa et al.	
7,400,487	B1	7/2008	Bitar	
7,675,197	B2 *	3/2010	Tetlow	307/104
2003/0011324	A1	1/2003	Lee	
2004/0248742	A1	12/2004	Terashima et al.	
2005/0046387	A1	3/2005	Aker et al.	
2005/0083059	A1	4/2005	Morita et al.	
2005/0148864	A1	7/2005	Slade et al.	
2005/0184689	A1	8/2005	Maslov et al.	
2006/0098378	A1	5/2006	Gross	
2006/0228548	A1	10/2006	Ayai et al.	
2007/0018629	A1	1/2007	Potvin et al.	
2007/0075053	A1	4/2007	Smith et al.	
2007/0145018	A1	6/2007	Smith et al.	
2007/0195561	A1	8/2007	Suenaga et al.	
2007/0222426	A1	9/2007	Waffenschmidt et al.	
2007/0263415	A1	11/2007	Jansen et al.	

OTHER PUBLICATIONS

U.S. Appl. No. 12/152,545, filed May 15, 2008, Messer.

U.S. Appl. No. 12/152,525, filed May 15, 2008, Messer.

U.S. Appl. No. 12/313,792, filed Nov. 25, 2008, Issa, et al.

U.S. Appl. No. 12/214,655, filed Jun. 20, 2008, Issa.

U.S. Appl. No. 12/152,977, filed May 15, 2008, Messer.

U.S. Appl. No. 12/152,976, filed May 15, 2008, Messer.

U.S. Appl. No. 60/930,221, filed May 15, 2007, Messer, et al.

U.S. Appl. No. 60/936,506, filed Jun. 20, 2007, Messer, et al.

U.S. Appl. No. 61/004,373, filed Nov. 27, 2007, Messer, et al.

Tracy V. Wilson, How Wireless Power Works, www.howstuffworks.com, Nov. 20, 2008.

Minoru Watanabe, et al. A Study on High Temperature Superconductin Coil of Different Coil Arrangements, IEEE Transactions on Applied Superconductivity, Jun. 2007, vol. 17, No. 2.

Power Requirements for Cooling versus Temperature in Kelvin,www.metox.biz/super_roadmap, Jun. 8, 2008.

Cryostat, Wikipedia Encyclopedia,Jun. 1, 2008.

Resonator, Wikipedia Encyclopedia, Jun. 3, 2008.

Zenergy Receives Order for HTS Induction Heater, Superconductor Week, Oct. 15, 2008, vol. 21 No. 18, Peregrine Communications.

Superconductivity Roadmap, www.metox.biz/super_roadmap, Jun. 8, 2008.

Miyagi, et al., AC Loss Characteristics of an assembled conductor of round Bi2223 Ag sheated wires, Elsivier Science B.V., Aug. 2002, vol. 372-376 Part 3 pp. 1727-1729.

Daumling, et al., Ac loss in superconducting power cables, Studies in High Temperature Superconductors, vol. 33.

O'Miura, et al, The development of a 2.5 T/100 kV A AC superconducting magnet using a high-J NbTi superconducting wire having Nb artificial pins, Supercond. Sci. Techno. Oct. 1993.

HTS-110 Leaders in HTS Magnetic Solutions, www.hts-110.co.nz, Jun. 8, 2008.

Michael Walker, et al., Performance of Coils Wound from Long Lengths of Surface-Coated, Reactedm BSCC0-2212 Conductor, IEEE Transactions on Applied Supercond. Jun. 1997, vol. 7 No. 2.

K. Hayashi, et al., Development of Ag-Sheated Bi2223 Superconducting Wires & Thier Applications, IEEE Transactions on Applied Superconductivity, Mar. 2001, vol. 11 No. 1.

A.B. Sneary,et al., Development og High Temperature Superconducting Coils Using Bi-2223/Ag Tapes, IEEE Transactions on Applied Superconductivity vol. 9 No. 2, Jun. 1999.

Mark S. Newson, et al., Progress on the Design and Operation of High-Tc Coils Using Dip-Coat BSCCO-2212/Ag Tape, IEEE Transactions on Applied Superconductivity, vol. 12 No. 1.

L. Bigoni, et al., Development of Bi-2223/Ag Tapes and Coils, IEEE Transactions on Applied Superconductivity, vol. 7 No. 2, Jun. 1997.

Pradeep Halder, et al., Developement of Bi-2223 HTS High Field Coils and Magnets, IEEE Transactions on Applied Superconductivity, Jun. 1995, vol. 5 No. 2.

PJ Masson, et al., HTS Machines as enabling technology for all-electric airborne vehicles, Superconductor Science & Technology, 2007, 748-756.

M. Ariante. et al., Characteristics of Coils wound with Mono and Multifilamentary Bi-2212/Ag from 4 to 80 K, IEEE Transactions on Magnetics vol. 32 No. 4, Jul. 1996.

Shinichi Nomura, et al., Helically Wound Coils for High Field Magnets, IEEE Transactions on Applied Superconductivity, vol. 14 No. 2, Jun. 2004.

Dan Combine, Radiant Energy and Over-Unity, Version 6, Oct. 2006.

Tomorrow's High-Temperature Superconducting Cables, www.innovations-report.com, Jun. 27, 2001.

Material Safety Data Sheet for SCS4050, SF12050, SF12100; High Temperature Superconductor Tape, Jun. 12, 2007, SuperPower Inc.

International Search Report for PCT/US2008/12013 dated Feb. 4, 2009.

International Search Report for PCT/US2008/06239 dated Aug. 26, 2008.

International Search Report for PCT/US2008/06176 dated Sep. 17, 2008.

International Search Report for PCT/US2008/13097 dated Feb. 4, 2009.

International Search Report for PCT/US2008/07655 dated Sep. 25, 2008.

Wireless Power Tesla Coil Demo, Last Accessed Sep. 14, 2009, http://www.youtube.com/watch?v=aG-We7A-i9U.

Miniature Vacuum Tube Tesla Coil, Last Accessed, Nov. 17, 2009, www.capturedlightnin•g.org/hot-streamer/electrontube/minivttc.htm.

MINIVTTC.Bmp Image http://www.capturedlightning.org/hot-streamer/electrontube/MINIVTTC.bmp.

NASA SBIR 2009 Phase 1 Form C-Budget Summary.

Information about Principal Investigators/Project directors and co-Principal Investigators/co-Project Directors, Dec. 3, 2009.

Page, Lewis, Lightning-gun tech approaching weaponisation, www.theregister.co.uk, Aug. 26, 2009.

Tobias, John and Grebel H., Self-Imaging in photonic crystals in a subwavelength range, Optice Letters, Dec. 1, 1999, vol. 24, No. 23.

Tobias, John, Ajgaonkar M., Grebel H., Morphology-dependent transmission through photonic crystals, J. Opt. Soc.Am.B., Mar. 2002, vol. 19, No. 3.
Qels 2002, Tuesday Afternoon.
Mar. 15, 2010 Non-final Rejection for U.S. Appl. No. 12/152,976.
Apr. 16, 2010 Notice of Allowance, for U.S. Appl. No. 12/152,977.
Jul. 19, 2010 Notice of Allowance, for U.S. Appl. No. 12/152,977.
Aug. 26, 2010 Notice of Allowance, for U.S. Appl. No. 12/152,977.
Nov. 10, 2010 Ex Parte Quayle Action, for U.S. Appl. No. 12/313,792.
Nov. 5, 2010 Non-final Rejection, for U.S. Appl. No. 12/152,545.
Dec. 8, 2010 Response to Office Action, for U.S. Appl. No. 12/313,792.
Feb. 2, 2011 Notice of Allowance, for U.S. Appl. No. 12/313,792.

* cited by examiner

FIG. 1A

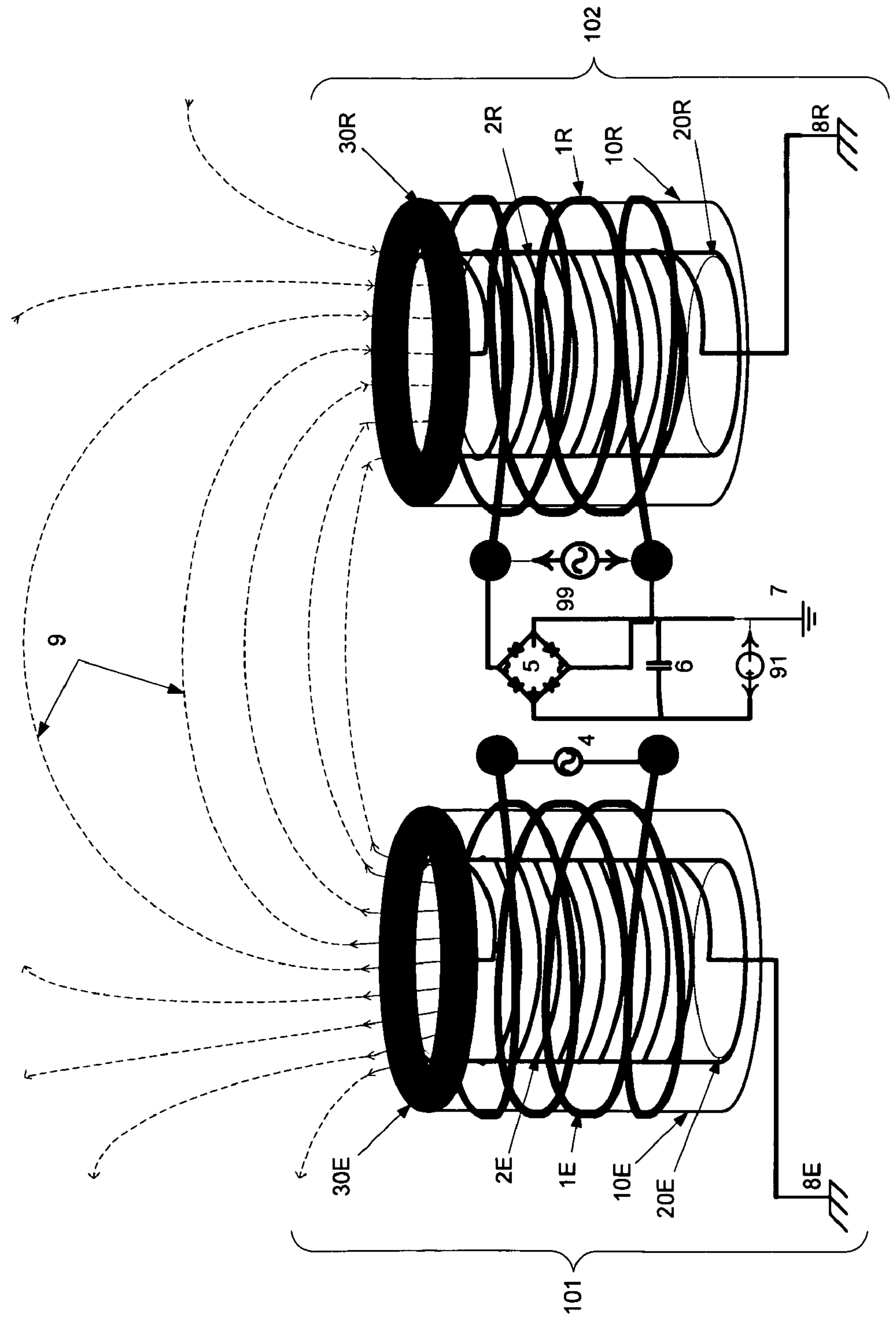
102
30R
2R
1R
10R
20R
8R
99
7
5
6
91
4
9
30E
2E
1E
10E
20E
8E
101

FIG. 1B

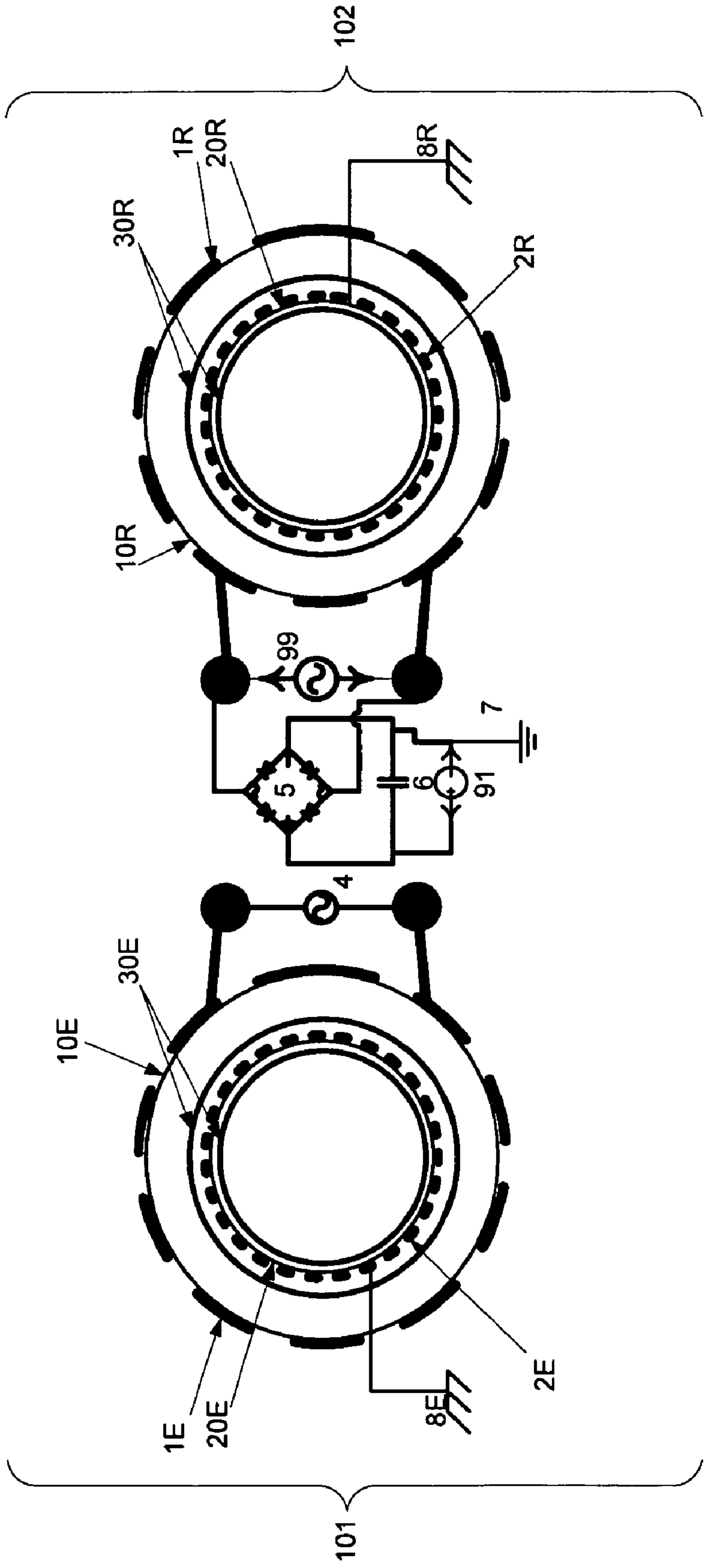
102
1R
20R
8R
30R
2R
10R
99
7
5
6
91
4
30E
10E
1E
20E
8E
2E
101

FIG. 1C

$$R_{fq} = 1 / (2\pi \mathrm{sqrt}(LC))$$

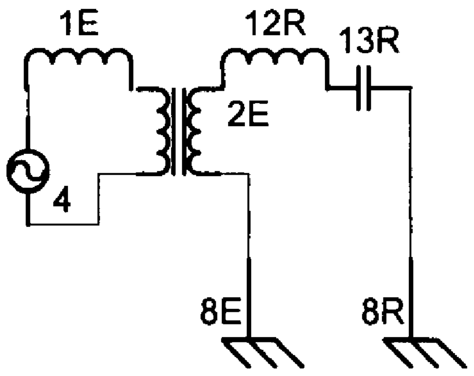
1E
12R
13R
2E
4
8E
8R

FIG. 1D

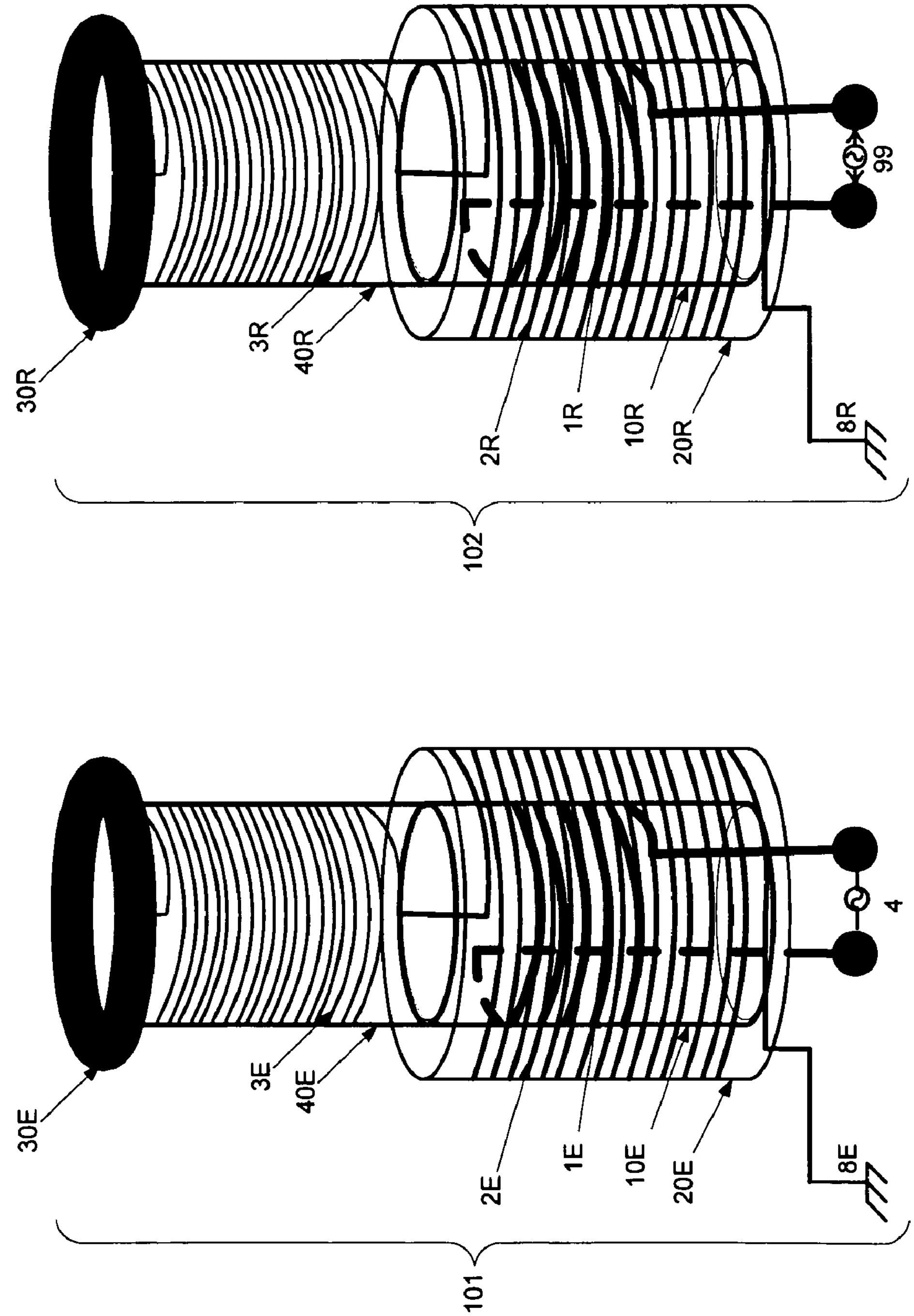
30R
3R
40R
2R
1R
10R
20R
8R
99
102
30E
3E
40E
2E
1E
10E
20E
8E
4
101

FIG. 1E

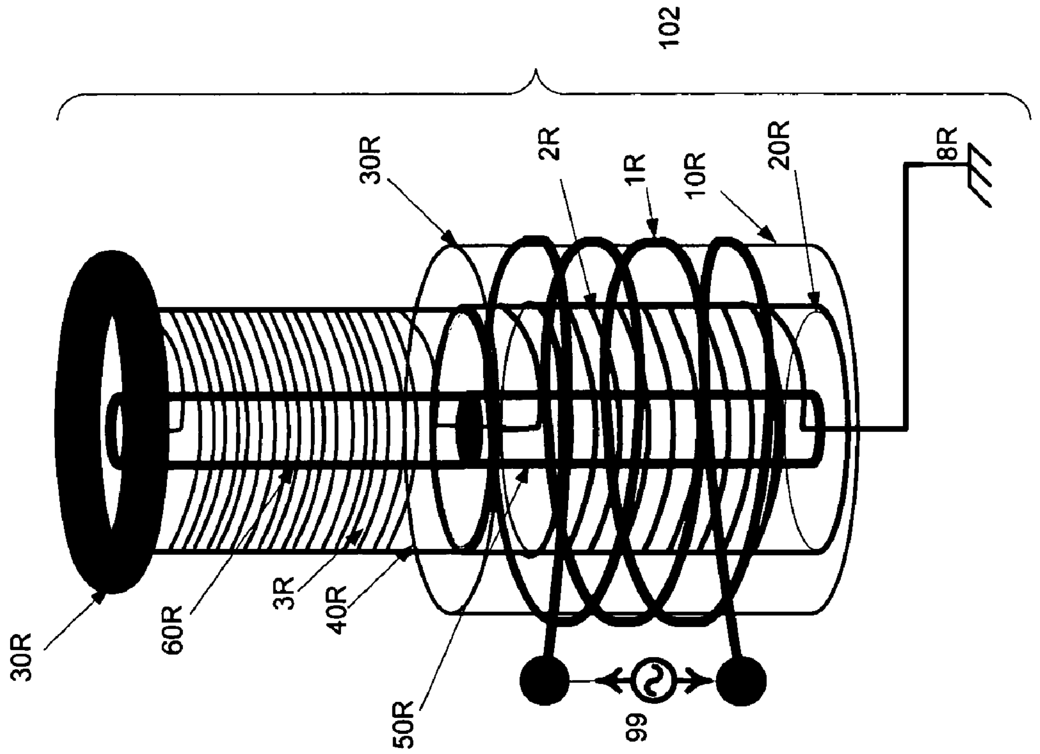
102
30R
2R
1R
10R
20R
8R
30R
60R
3R
40R
50R
99

FIG. 2

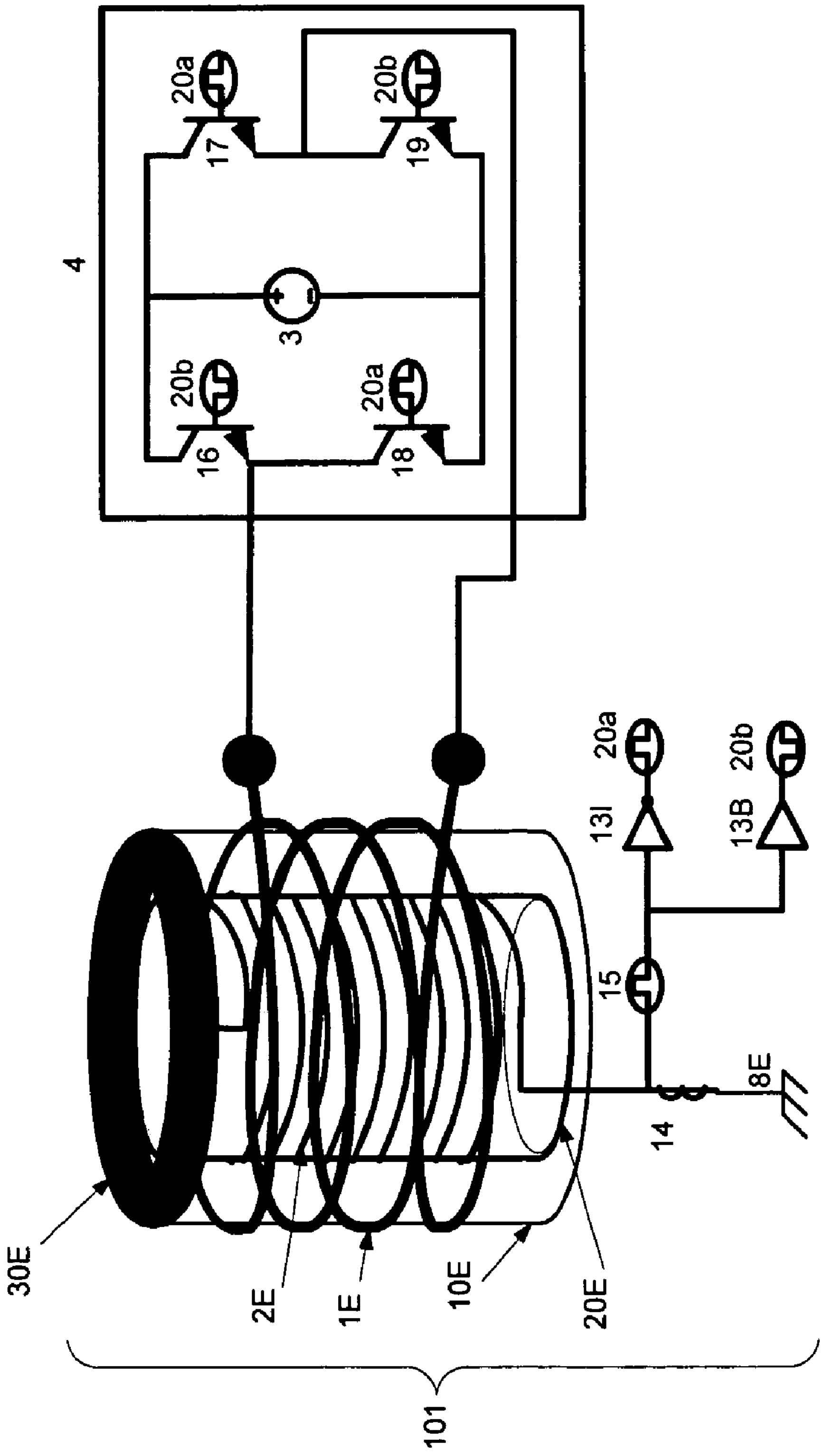
13I
8E
2E
1E

FIG. 3

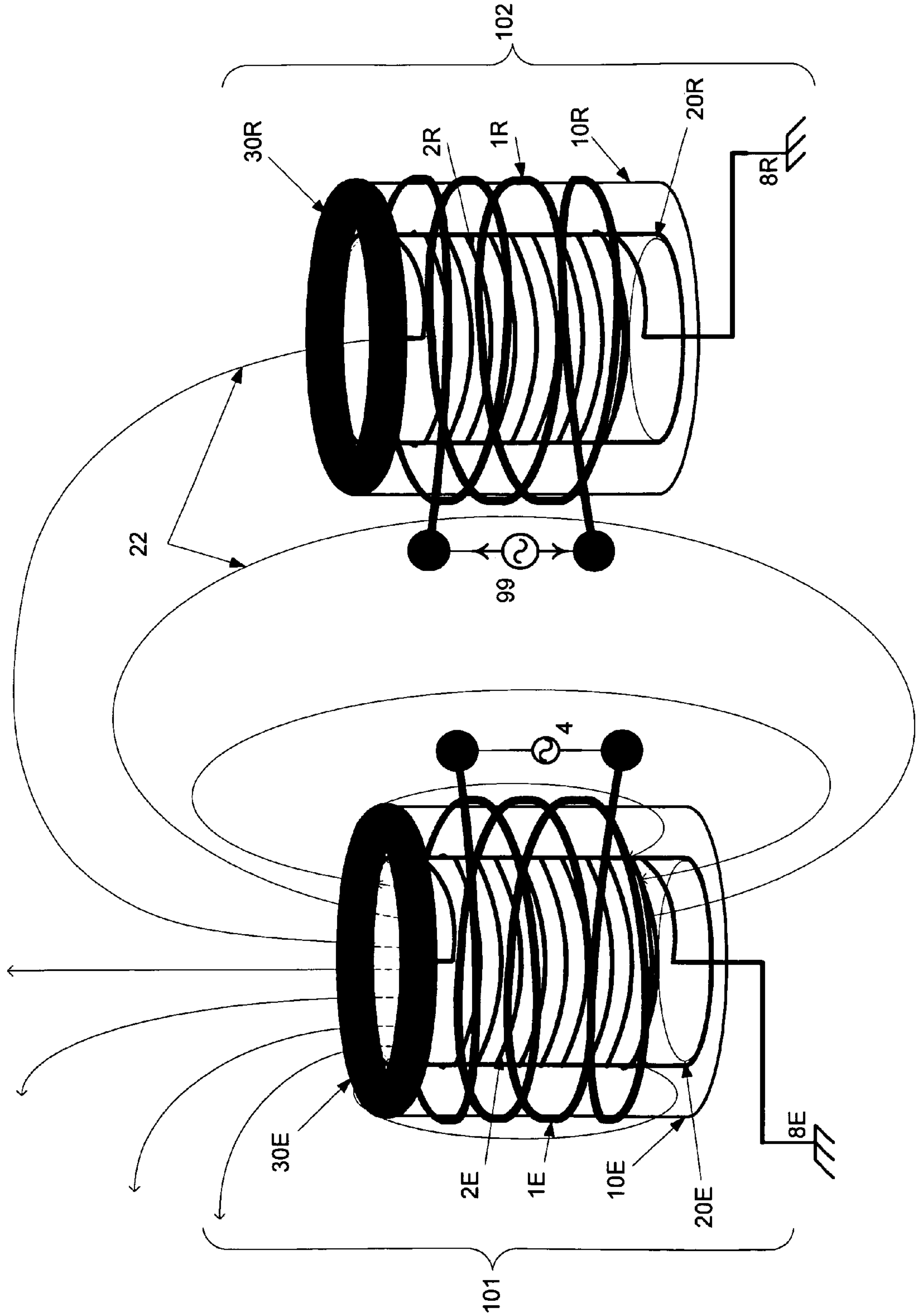
102
30R
2R
1R
10R
20R
8R
22
99
4
30E
2E
1E
10E
20E
8E
101

FIG. 4

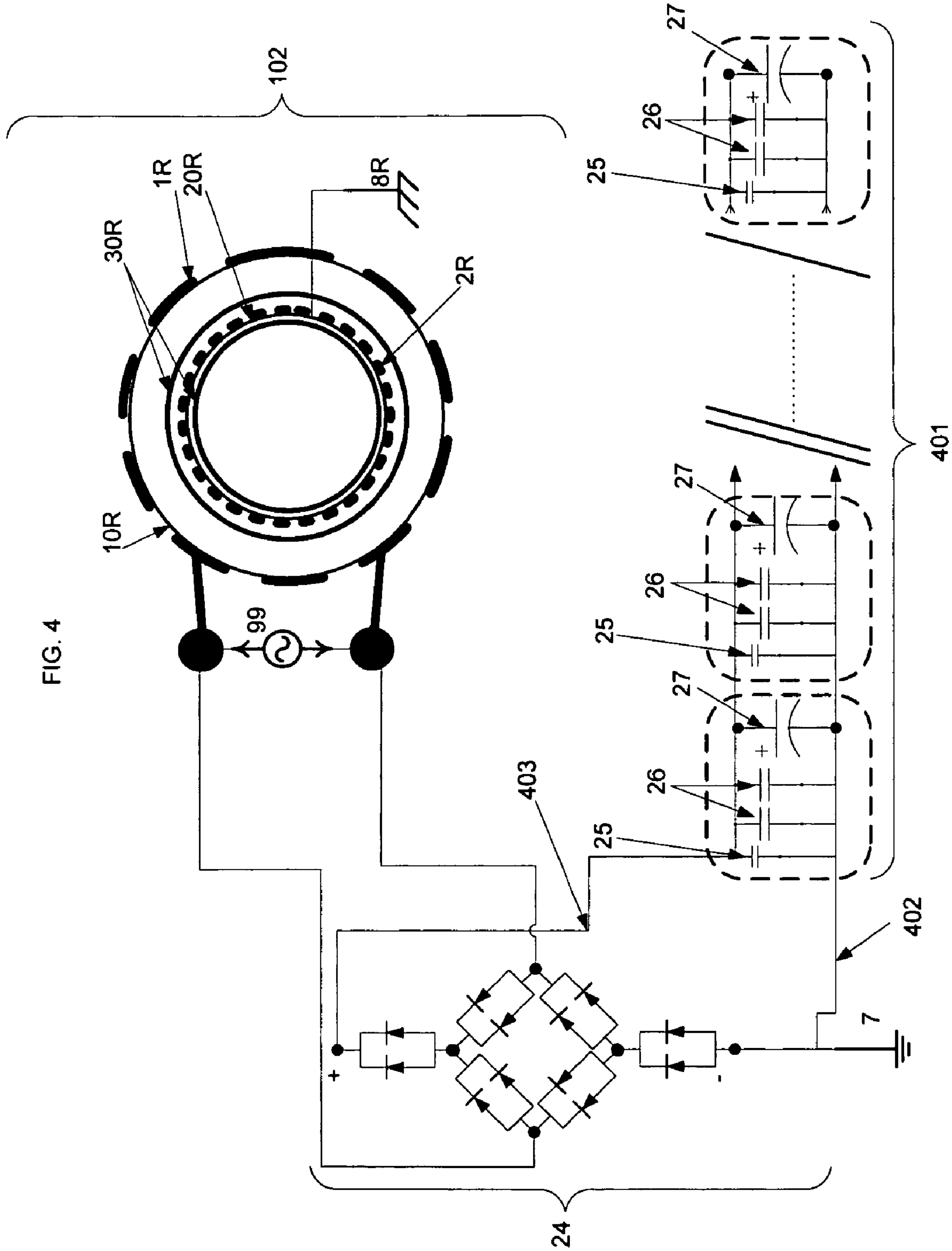
102
1R
20R
30R
8R
2R
10R
99
403
25
26
27
401
402
7
+
-
24

FIG. 5

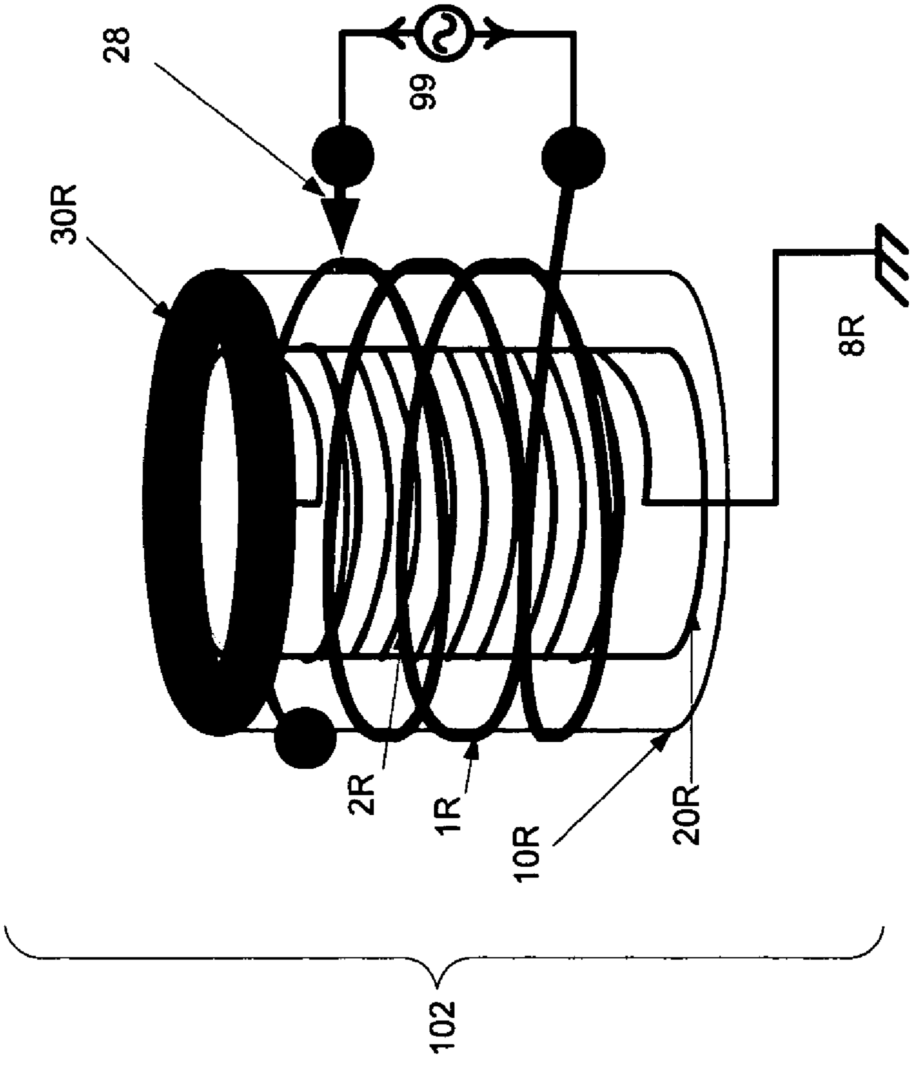
28
99
30R
8R
2R
1R
10R
20R
102

FIG. 6

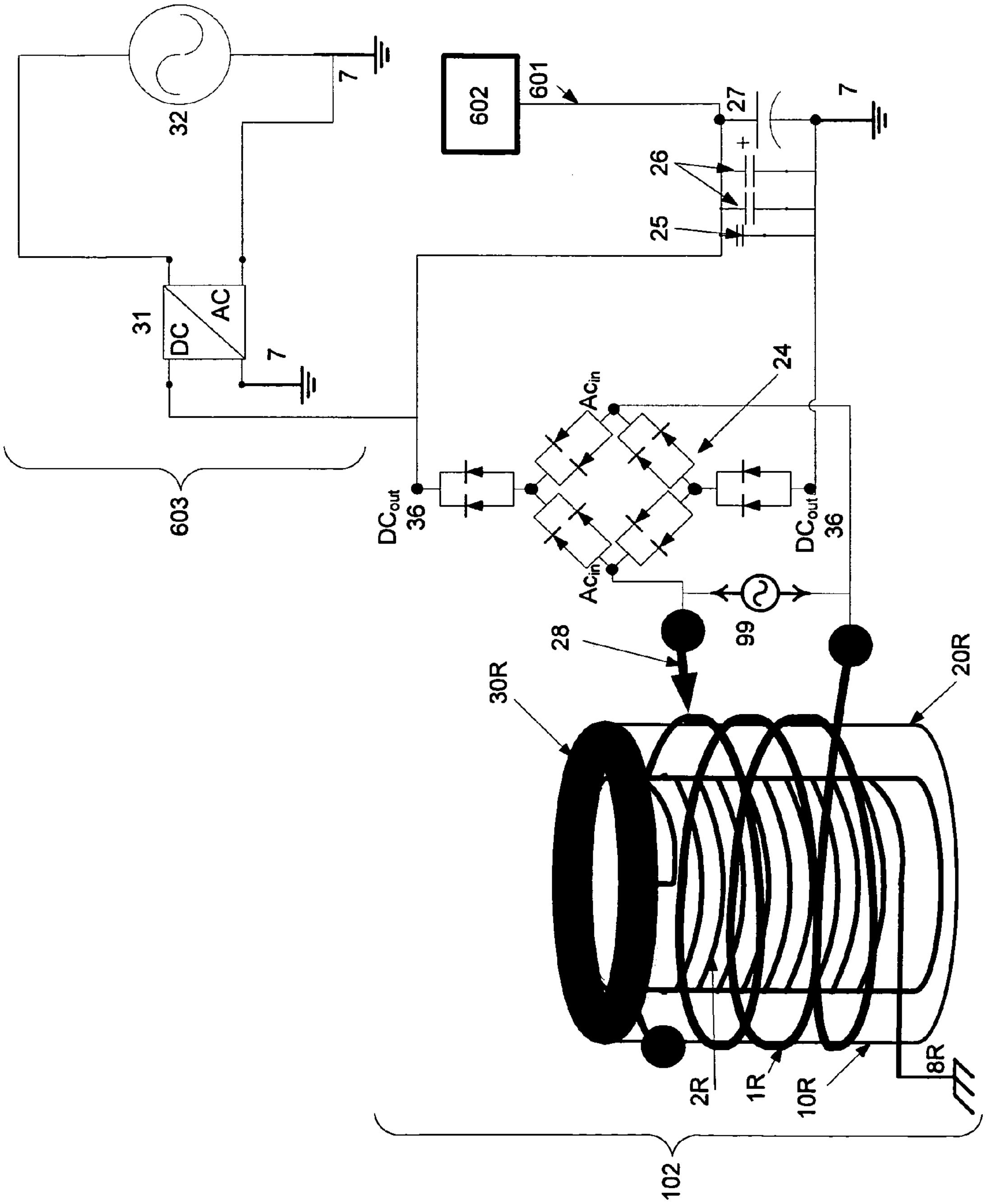
32
7
602
601
27
+
7
26
25
31
DC
AC
7
24
Ac_in
603
DC_out
36
DC_out
36
Ac_in
28
99
30R
20R
2R
1R
10R
8R
102

FIG. 7

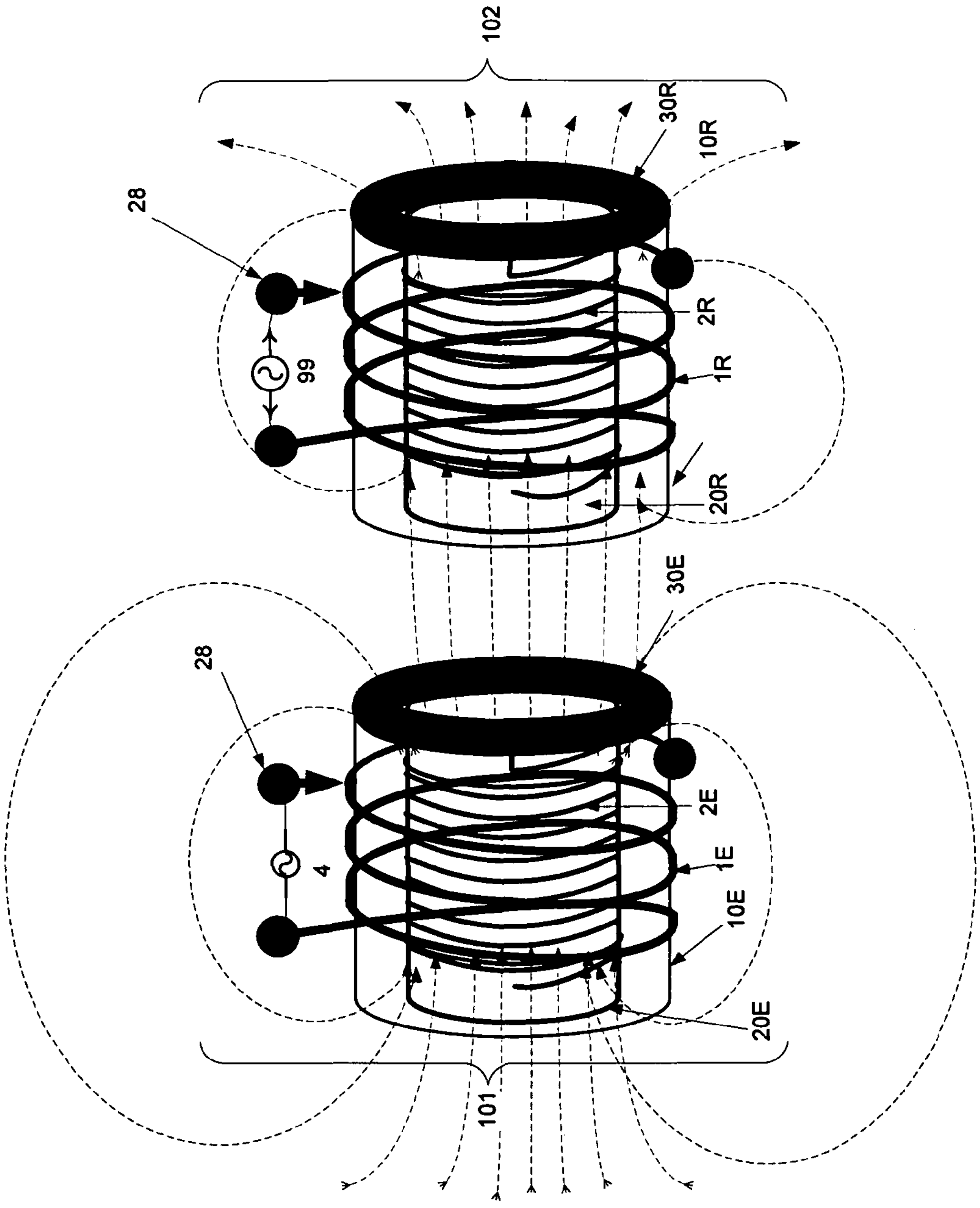
102
30R
10R
28
2R
99
1R
20R
30E
28
2E
4
1E
10E
20E
101

FIG. 8

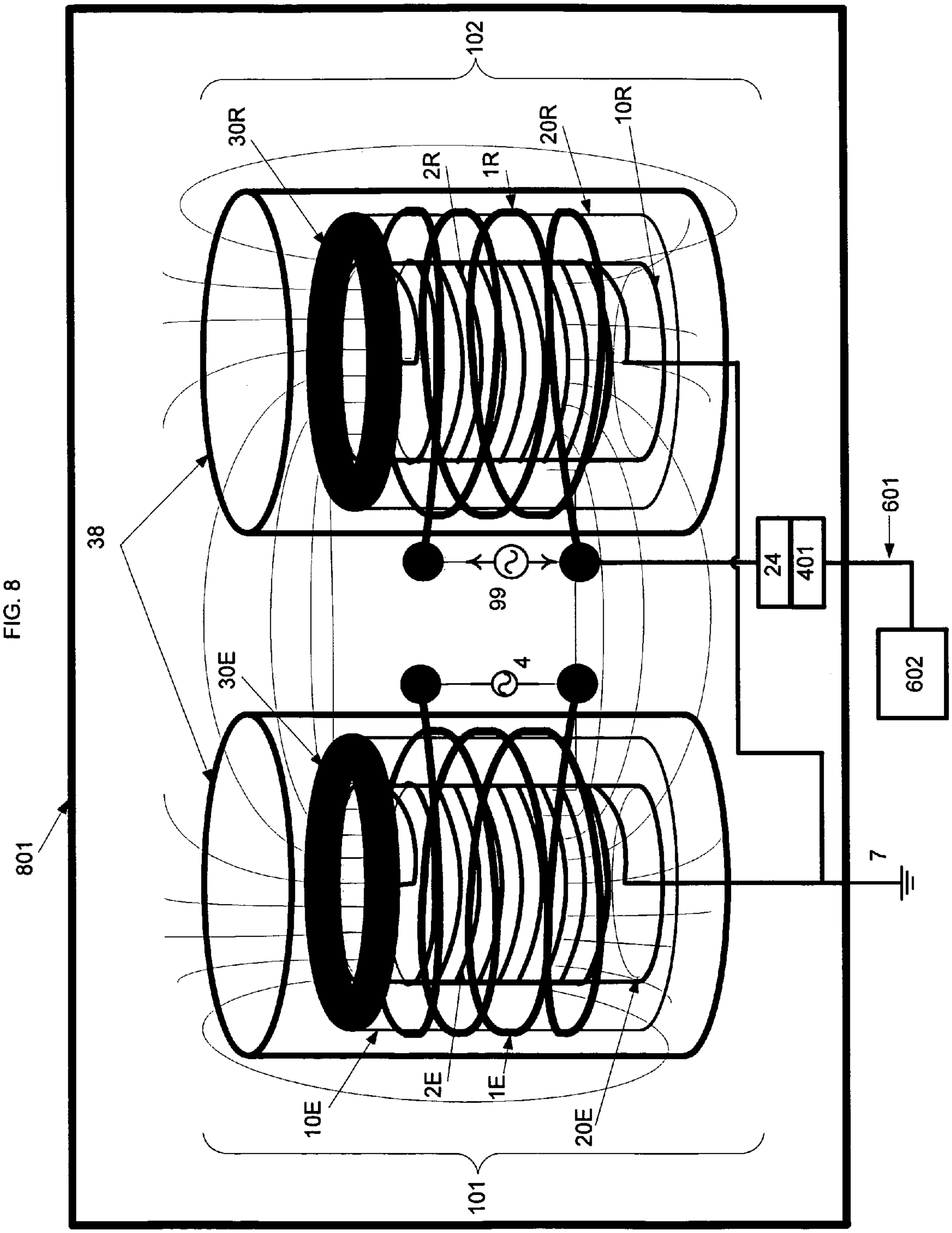
30R
2R
1R
20R
10R
2E
1E

FIG. 9A

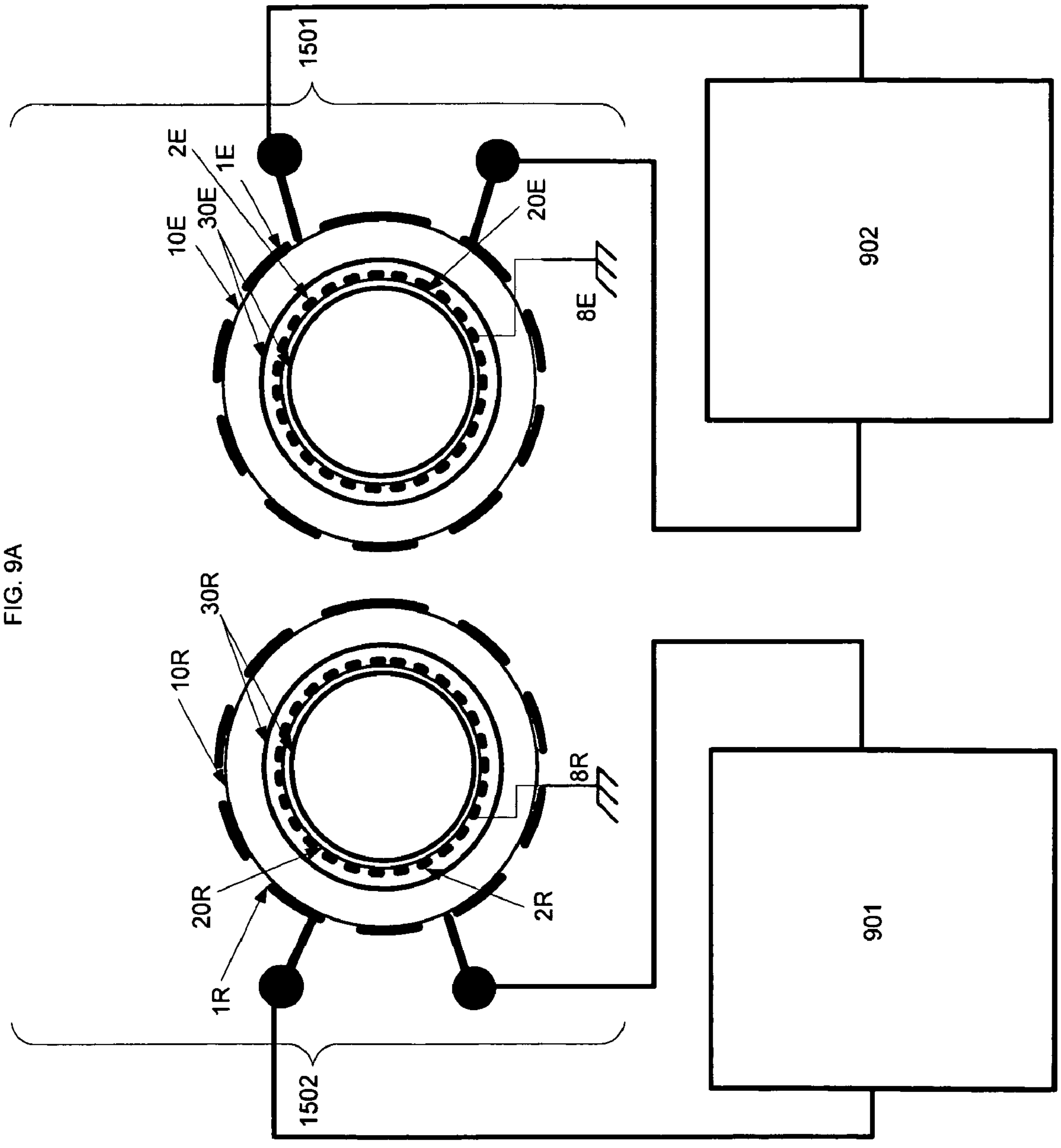
1501
2E
1E
10E
30E
20E
8E
902
30R
10R
8R
20R
2R
1R
901
1502

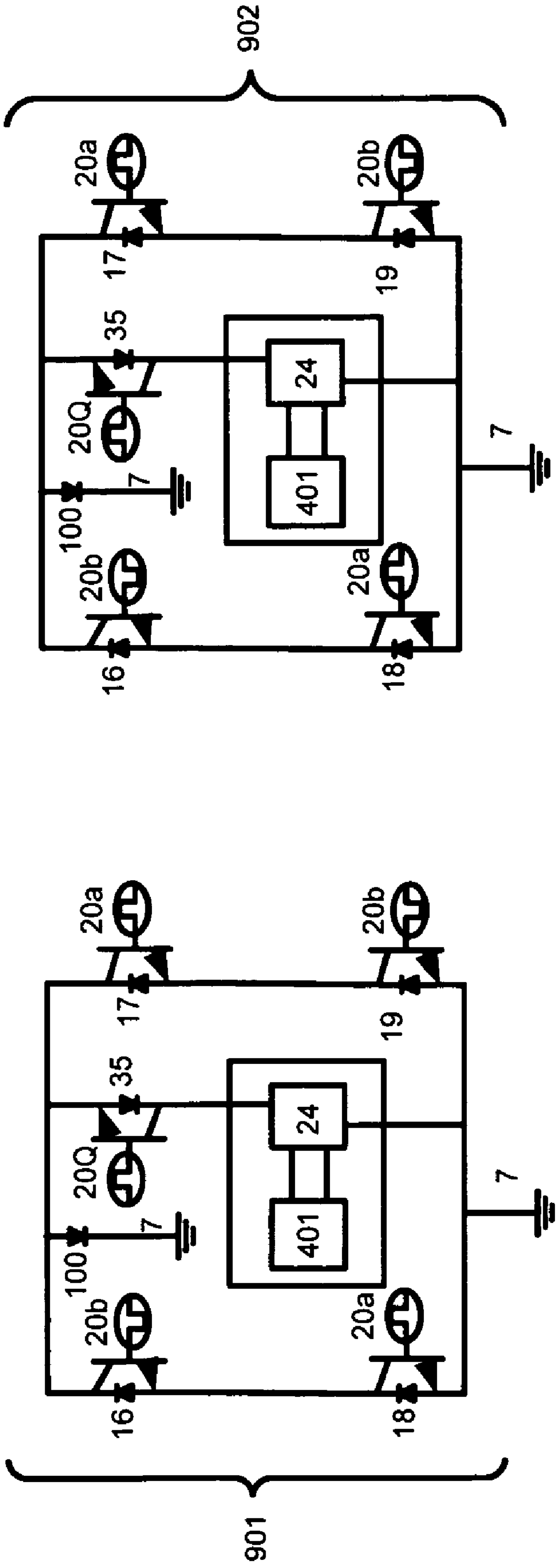
20Q
20Q

FIG. 9B

FIG. 10

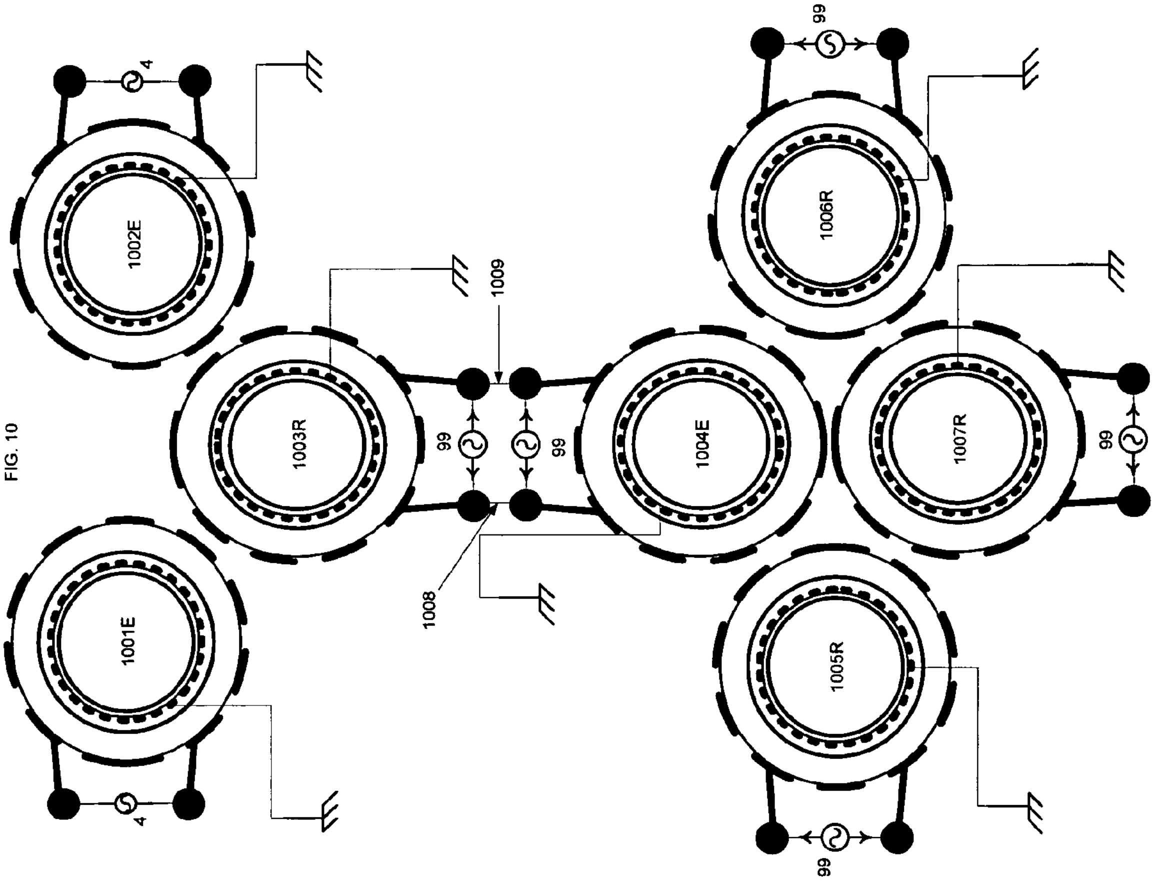
1003R
1006R
1007R
1005R

FIG. 11

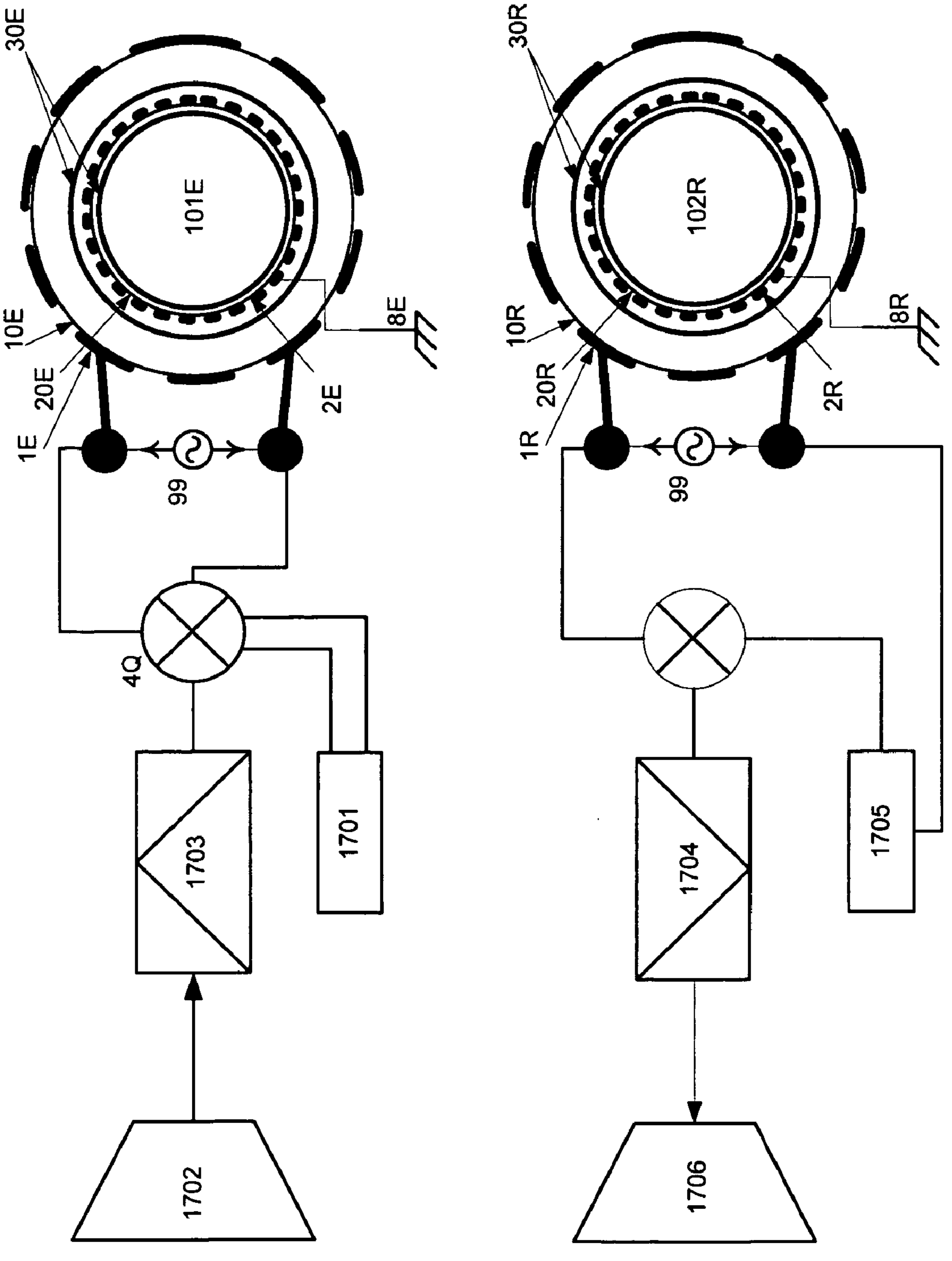
30E
101E
10E
20E
1E
2E
8E
99
4Q
1703
1701
1702
30R
102R
10R
20R
1R
2R
8R
99
1704
1705
1706

METHODS AND SYSTEMS FOR WIRELESS ENERGY AND DATA TRANSMISSION

This application claims the benefit of provisional patent application Ser. No. 61/004,373, filed Nov. 27, 2007. This application hereby incorporates by reference the entire contents of each of U.S. non-provisional patent application Ser. No. 12/152,545 titled "System and Method for Forming and Controlling Electric Arcs," filed May 15, 2008; U.S. non-provisional patent application Ser. No. 12/152,525 titled "System and Method for Controlling an Electromagnetic Field Generator," filed May 15, 2008; and U.S. non-provisional patent application Ser. No. 12/214,655 titled "System and Method for Using a Vacuum Core High Temperature Superconducting Resonator," filed Jun. 20, 2008.

FIELD OF THE INVENTION

The present invention relates to methods and systems for wireless energy and data transmission.

BACKGROUND OF THE INVENTION

It may be desirable to have a system and method for wireless energy and data transmission, especially in fields including, but not limited to, manufacturing, robotics, and consumer devices. By powering devices through wireless transmission of energy, the requirement for batteries or other chemical energy storage means may be reduced. In turn this may reduce the need for copper or aluminum cabling manufacturing costs and battery waste, thereby resulting in more environmentally friendly products.

The use of wireless transmission of energy may also simplify installation, as said devices may be moved without labor intensive rewiring. An additional benefit may be gained from the simultaneous transmission of data over the same carrier wave that transmits the energy for similar reasons of utility, device simplicity, and cost reduction through the use of fewer discrete devices in a given device design. Such use may also conserve bandwidth in other wireless pure data transmission regimes, which may lead to increased network efficiencies.

SUMMARY OF THE INVENTION

In certain aspects, the present invention may provide a system for wirelessly transmitting and receiving energy. In one aspect, the system may include an emitter and a receiver. The emitter may include a primary emitting winding wrapped around a first support and having first and second ends coupled to a driver and a secondary emitting winding wrapped around a second support and having a first end of the secondary emitting winding coupled to a first top load and a second end. The receiver may include a primary receiving winding wrapped around a third support and having first and second ends coupled to first and second inputs of a rectifying circuitry, respectively. The first output of the rectifying circuitry may be coupled to a first terminal of one or more capacitors and a second output of the rectifying circuitry may be coupled a second terminal of the one or more capacitors. The receiver may further include a secondary receiving winding wrapped around a fourth support and having a first end coupled to a second top load and a second end.

In one aspect, the second support may be located within the first support and the fourth support may be located within the third support. In another aspect, the first support may be located with the second support and the third support may be located within the fourth support.

In one aspect, the emitter may include a primary emitting winding wrapped around a first support and having first and second ends coupled to a driver, a secondary emitting winding wrapped around a second support and having first and second ends, and a tertiary emitting winding wrapped around a third support and having a first end of the tertiary emitting winding coupled to the first end of the secondary emitting winding and the second end of the tertiary emitting winding coupled to a first top load. In one aspect, the receiver may include a primary receiving winding wrapped around a fourth support and having first and second ends coupled to first and second inputs of a rectifying circuitry, respectively, where a first output of the rectifying circuitry is coupled to a first terminal of one or more capacitors and a second output of the rectifying circuitry is coupled to a second terminal of the one or more capacitors. The receiver may further include a secondary receiving winding wrapped around a fifth support and having first and second ends and a tertiary receiving winding wrapped around a sixth support and having a first end of the tertiary receiving winding coupled to the first end of the secondary receiving winding and the second end of the tertiary receiving winding coupled to a second top load.

In further aspects, the first support may be located within the second support, the fourth support may be located within the fifth support, the third support may be located above the first and second supports, and the sixth support may be located above the fourth and fifth supports.

In another aspect, the second support may be located with the first support, the fifth support may be located within the fourth support, the third support may be located above the first and second supports, and the sixth support may be located above the fourth and fifth supports.

In one aspect, the present invention may provide for a method of wirelessly transferring data and energy. The method may include inputting a data signal, inputting power to a driver, generating a carrier wave with the driver at a resonant frequency of an emitter, modulating the carrier wave based on the data signal to create a combined signal, driving the combined signal to an AC input on the emitter, electromagnetically coupling the emitter to a receiver, reproducing the combined signal in an AC output of the receiver, rectifying the AC output into DC power, demodulating the combined signal into the data signal, and transmitting the data signal to a device.

In one aspect, the present invention may provide for a method of wirelessly transmitting and receiving energy. The method may include transmitting an alternating current into a driver coupled to a primary emitting winding of an emitter, inducing the alternating current into a secondary emitting winding of the emitter, sensing a frequency of the alternating current on the secondary emitting winding, transmitting a feedback signal to the driver based on the frequency, driving the AC current into the primary emitting winding at the resonant frequency of the secondary emitting winding, electromagnetically coupling the emitter to a receiver, inducing the alternating current into a secondary receiving winding of the receiver, inducing the alternating current into a primary receiving winding of the receiver, outputting the alternating current into a rectifying circuitry, rectifying the alternating current into a direct current; reducing ripples on the direct current, and transmitting the direct current to a device.

BRIEF DESCRIPTION OF THE DRAWINGS

Features and other aspects of embodiments of the present invention are explained in the following description taken in conjunction with the accompanying drawings, wherein:

FIG. **1**A illustrates a general system including a Tesla coil emitter and receiver according to one aspect of the system and method of the present disclosure;

FIG. **1**B illustrates a general system including a Tesla coil emitter and receiver in plan view according to one aspect of the system and method of the present disclosure;

FIG. **1**C illustrates a schematic representation of a Tesla coil emitter and receiver according to one aspect of the system and method of the present disclosure;

FIG. **1**D illustrates a general system including Tesla coil emitter and receiver each with tertiary windings and the primary windings located within the secondary windings according to one aspect of the system and method of the present disclosure;

FIG. **1**E illustrates a receiver including materials of high magnetic permeability according to one aspect of the system and method of the present disclosure;

FIG. **2** illustrates a driving circuit detail according to one aspect of the system and method of the present disclosure;

FIG. **3** illustrates a magnetic field coupling of the coils according to one aspect of the system and method of the present disclosure;

FIG. **4** illustrates a receiver rectification circuit detail with capacitor bank according to one aspect of the system and method of the present disclosure;

FIG. **5** illustrates a variable tap implemented on Tesla coil according to one aspect of the system and method of the present disclosure;

FIG. **6** illustrates an example of application of a system to provide power for utilization according to one aspect of the system and method of the present disclosure;

FIG. **7** illustrates a coaxial orientation of emitter and receiver according to one aspect of the system and method of the present disclosure;

FIG. **8** illustrates a Tesla coil emitter and receiver according to one aspect of the system and method of the present disclosure;

FIG. **9**A illustrates a reversible Tesla coil system where the function of emitter and receiver may alternate according to one aspect of the system and method of the present disclosure;

FIG. **9**B illustrates circuitry for a reversible Tesla coil system where the function of emitter and receiver may alternate according to one aspect of the system and method of the present disclosure

FIG. **10** illustrates a representative example of a network of Tesla coil emitters and receivers according to one aspect of the system and method of the present disclosure;

FIG. **11** illustrates a representative method of information and power transmission using a Tesla coil system according to one aspect of the system and method of the present disclosure;

The drawings are exemplary, not limiting. It is intended for items that are labeled with the same number in multiple figures to refer to the same item throughout the figures.

DETAILED DESCRIPTION

Various embodiments of the present invention will now be described in greater detail with reference to the drawings.

As shown in FIG. **1**A, one aspect may include two Tesla resonators **101** and **102**, which may be electromagnetically coupled through emitting a time-varying magnetic and/or electric field **9**. Tesla resonators **101** and **102** may be oriented parallel to each other. The emitting Tesla resonator apparatus ("emitter") **101** may emit the field, while the receiving Tesla coil apparatus ("receiver") **102** may subtend the magnetic field from the emitter **101**. Parallel orientation of emitter **101** and receiver **102** may ensure maximum flux coupling between them. Emitter **101** may include primary winding **1**E, secondary winding **2**E, support apparatus for primary winding **10**E, and support apparatus for secondary winding **20**E; receiver **102** may include primary winding **1**R, secondary winding **2**R, support apparatus for primary winding **10**R, and support apparatus for secondary winding **20**R. Primary and secondary windings **1**E, **1**R, **2**E, and **2**R may be composed of any common wiring material used in the implementation or construction of coils and transformers. Other aspects may use other materials. Primary structural supports **10**E and **10**R and secondary structural supports **20**E and **20**R may be composed of ceramic, plastic, Plexiglas.RTM., or any other insulating or nonconductive (e.g., dielectric) material. Primary winding **1**E and **1**R may be wrapped around primary structural supports **10**E and **10**R, respectively, and secondary windings **2**E and **2**R may be wrapped around secondary structural supports **20**E and **20**R, respectively. Windings **1**E, **1**R, **2**E, and **2**R may be wrapped in a helically-coiled fashion, where each primary winding **1**E and **1**R is oriented similarly as its respective secondary winding **2**E and **2**R. In such aspects, the ends of each winding may not be helically-coiled, as, in various aspects, they may be coupled to other windings, circuitry, or to ground. Primary and secondary structural supports **10**E, **10**R, **20**E, and **20**R may be cylindrical in shape. With respect to structural supports and windings, emitter **101** may be substantially identical to, or identically arranged as, receiver **102**.

In one aspect, primary windings **1**E and **1**R may be 6 AWG wire; secondary windings **2**E and **2**R may be 24 AWG wire; and tertiary windings **3**E and **3**R may be 28 AWG wire. In another aspect, secondary windings **2**E and **2**R may have 1,800 turns; tertiary windings **3**E and **3**R may have 2,400 turns; primary receiver winding **1**R may have 12 turns; and primary emitter winding **1**E may have 28 turns. In other aspects, other wire and different number of turns may be used. The drawings of the windings in the figures are not intended to show the exact number of turns or ratio of turns used in aspects of the present invention. In one aspect, primary structural supports **10**E and **10**R may be about 12 inches each in diameter; secondary structural supports **20**E and **20**R may be about 8 inches each in diameter; and tertiary structural supports **40**E and **40**R may be each about 6 inches in diameter. In one aspect, top loads **30**E and **30**R may be each about 2.5 inches thick, about 8 inches in outer diameter, and about 6 inches in inner diameter. In one aspect having primary windings **1**E and **1**R located within secondary windings **2**E and **2**R, respectively (as shown in FIG. **1**D according to one aspect), primary structural supports **10**E and **10**R may have a smaller diameter than secondary structural supports **20**E and **20**R, for example, 6 inches in outer diameter. In one aspect having tertiary windings **3**E and **3**R, an end cap made of a nonconductive material may be used on the larger of the primary and secondary structural supports **10**E, **10**R, **20**E, and **20**R in order to provide a base for tertiary structural supports **40**E and **40**R. In other aspects, other sized supports and top loads may be used.

In one aspect, top load **30**E and/or **30**R may be provided on the secondary winding **2**E and/or **2**R that may reduce eddy currents and corona and thereby may reduce electric leakage of the device through corona. Secondary top load **30**E and **30**R may be comprised of a toroidal metal shape with a hole cut out of the center of the toroid that is penetrated by the inner diameter of secondary support **20**E and/or **20**R of secondary winding **2**E and/or **2**R. In these aspects, the final coil in the series (whether secondary or tertiary) may receive top load **30**E and/or **30**R. Using top load **30**E and/or **30**R may reduce energy loss by about 5-10% by reducing eddy currents. In another aspect, toroidal top load **30**E and/or **30**R may also use a vertical slit as opposed to the circular hole for similar purposes and effects.

As shown in FIG. **1**D, one aspect may include primary windings **1**E and **1**R wrapped around primary structural supports **10**E and **10**R, respectively, and have primary structural supports **10**E and **10**R located within secondary structural supports **20**E and **20**R, respectively. In further aspects, one or more tertiary windings **3**E and **3**R may be wrapped around tertiary structural supports **40**E and **40**R, respectively, which sit on top of primary and secondary structural supports **10**E, **20**E, **10**R, and **20**R. In further aspects, tertiary emitting winding **3**E having two ends wrapped around tertiary support structure **40**E may be coupled to the end of secondary emitting winding **2**E that is not connected to ground. The other end of the tertiary emitting winding **3**E may be connected to top load **30**E (or, in other aspects, to another emitting winding in the same pattern (not shown)). In further aspects, tertiary emitting winding **3**E may have more windings than secondary emitting winding **2**R in order to create more voltage if more voltage is desired.

In further aspects, tertiary receiving winding **3**R having two ends wrapped around tertiary support structure **40**R may be connected to the end of secondary receiving winding **2**R that is not connected to ground. The other end of tertiary receiving winding **3**R may be connected to top load **30**R (or, in other aspects, to another emitting winding in the same pattern (not shown)). In further aspects, tertiary receiving winding **3**R may have more windings than secondary receiving winding **2**R in order to create more voltage if more voltage is desired.

In further aspects, the system may include one or more tertiary Tesla resonators (not shown) or emitters (not shown). In further aspects, additional windings may be used; therefore, aspects are not intended to be limited to merely two or three windings.

In further aspects, AC voltage **4** may be presented through application of a solid state driving system that drives primary emitting winding **1**E at the resonant frequency of Tesla resonators' **101** and **102**. Receiver **102** may receive energy through electromagnetic coupling to emitter **101**.

In further aspects, the coil structure of emitter **101** and receiver **102** may be substantially identical, with one difference being the driving circuitry of emitter **101** and the rectification circuitry of receiver **102**. In further aspects, emitter **101** and receiver **102** may be provided with the rectification and switching circuits along with circuitry to switch between the two, so that energy may be transferred back and forth between emitter **101** and receiver **102**. In further aspects, an electromagnetic field couples emitter **101** to receiver **102**. Current from primary winding emitter **1**E ultimately may couple through the exchange of electromagnetic energy into receiver **102**. That is, primary emitting winding **1**E may induce AC voltage **99** on primary receiving winding **1**R through electromagnetic coupling between emitter **101** and receiver **102**. AC out **99** may be rectified by full bridge rectifier **5** and may charge capacitor **6** with DC out **91** voltage.

In aspects, emitter **101** and receiver **102** may be spatially distributed, either in fixed positions or mobile. These fixed positions may not be in set locations, so long as receiver **102** is within a certain range of emitter **101** that may be determined by the electromagnetic field intensity of a particular device. Likewise, a mobile receiver **102** may not be required to be constrained to a track or fixed path but so long as it remains within a certain range of one or more emitters **101**. In further aspects, multiple emitters **101** and receivers **102** may be used, where several receivers **102** may draw energy from one emitter **101** or a single receiver **102** may draw energy from multiple emitters **101** (and so forth).

FIG. **1**B displays emitter **101** and receiver **102** in plan view according to one aspect of the system of the present disclosure.

FIG. **1**C shows an electrical schematic according to one aspect The resonant frequency is given by equation Fr=1/(2 pi sqrt(LC)), where C is capacitive value of **13**R and L is inductive value of lumped inductance **12**R of secondary emitting winding **2**E.

As shown in FIG. **1**C, further aspects may include secondary emitting winding **2**E and loading capacitor **13**R, which may be attached to ground **8**E and **8**R. In further aspects, secondary emitting winding **2**E may be floating, that is, not connected to electrical ground. In further aspects, loading capacitor **13**R may be floating.

As shown in FIG. **1**E, one aspect may include cylinders of high permeability material (e.g., iron, Mu metal, HyMu80) **50**R and **60**R in structural supports **30**R and **40**R, respectively. Use of cylinders of high permeability material **50**R and **60**R in receiver **102** may make the power transfer more efficient between emitter **101** and receiver **102** by concentrating the magnetic flux within the windings surrounding the cylinders of high permeability material **50**R and **60**R. In one aspect, cylinders of high permeability material **50**R and **60**R may be solid. In another aspect, cylinders of high permeability material **50**R and **60**R may be hollow, such as, a sheet of such material wrapped into a cylindrical form. In other aspects, high permeability material **50**R and **60**R may be a form in shapes other than a cylinder.

FIG. **2** shows the driving circuit of emitter **101** according to one aspect of the system of the present disclosure. Primary emitting winding lE may be driven by insulated gate bipolar transistors ("IGBT"s) **16**, **17**, **18**, and **19** in an H-bridge configuration with a power source **3**. The AC signal to primary emitting winding **1**E may thereby induce a voltage in secondary winding **2**E, initiating the Tesla coil effect.

Further aspects may include a feedback system that monitors the direction of current via current drive transformer **14**. In one aspect, current drive transformer **14** is coupled to secondary emitting winding **20**E. Current drive transformer **14** may sense current on secondary emitting winding **20**E, where current drive transformer **14** may provide voltage in response to oscillation at the resonant frequency of secondary emitting winding **20**E. Digital feedback signal **15** may be a logic "1" signifying current flowing in one direction and a logic "0" signifying current flowing in the opposite direction. Feedback signal **15** may allow the H-bridge to determine when to switch its transistors such that induced current flows into emitter **101** from primary emitting winding **1**E may constructively interfere to increase magnetic flux in emitter **101**. In further aspects, the feedback system drives primary emitting winding **1**E at the resonant frequency of secondary emitting winding **2**E.

In further aspects, feedback signal **15**, filtered with inverting **131** and non-inverting **13**B IGBT gate drive system, may create logic signals **20***a* and **20***b*. Logic signals **20***a* and **20***b* may drive the H-bridge with the frequency, phase, and polarity for an AC in wave to flow to primary emitter winding **1** E to amplify the resonating field in emitter **101**. Thus, the device may be self-correcting in phase and frequency.

FIG. **3** shows a magnetic field coupling between emitter **101** and receiver **102** according to one aspect. Magnetic flux lines **22** are intended to be drawn to show that secondary receiving winding **2**R of receiver **102** is immersed in field **22**. In further aspects, AC voltage **99** may result from the magnetic field coupling and the effect of secondary receiving winding **2**R and primary receiving winding **1**R. Because of the physical properties of Tesla coils, when an exciting voltage is presented to **99**, receiver **102** may be capable of acting as an emitter and transferring energy to emitter **101**.

Wireless energy transfer may allow the conversion of the AC signal **99** from primary receiving winding **1**R into a useful form. As shown in FIG. **4**, according to one aspect of the system of the present disclosure, the conversion of AC voltage output **99** to DC voltage across capacitor **27** may occur from induced current in primary receiver winding **1**R, which was demonstrated as induced by a coupled magnetic field between Tesla resonators **101** and **102**. DC voltage exiting diode array **24** may be imperfect because of possible ripples. Output current **99** may be rectified via high recovery speed diodes and diode array **24** and stored into, for example, capacitor bank **401**, as shown according to one aspect. In further aspects, high recovery speed diodes and diode array **24** may be rated at, for example, 1200V, 60 A with a reverse recovery time of 30 ns and a speed for fast recovery =<500 ns and =>200 mA. In one aspect, adding diodes (not shown) to diode array **24** may allow the system to handle higher voltages.

In further aspects, capacitor bank **401** may be a system of capacitors for the purpose of storing energy. The system may include low inductance polymeric capacitor **26** and snubber capacitor **25** acting as filters connected in parallel to electrolytic capacitor **27** acting as a storage capacitor for fast charging and discharging of capacitors **25**, **26**, and **27**. Low internal and external stray inductance may be achieved by minimum length of wires interconnecting capacitors **25** and **26** and capacitor **27**. Diode array **24** may be connected to inputs of capacitor bank **401** through leads **402** and **403** between the positive (+) and negative (−) outputs on diode array **24** and the positive (+) and negative (−) inputs on capacitor bank **401**. In further aspects, minimizing and equalizing the inductance of leads **402** and **403** may be accomplished by ensuring leads **402** and **403** are of equal length, and as short as possible.

As shown in FIG. **5**, further aspects may include variable tap **28** on primary receiving winding **1**R of receiver **102**. Variable tap **28** may be a moveable and changeable connection on a transformer through which the voltage may be changed. Variable tap **28** may change the voltage output to the voltage input ratio, similar to the ways an iron core transformer with multiple primary taps may change in voltage with relative voltage output to voltage input ratio change. Because in aspects, the system is reversible, variable tap **28** may be coupled to primary emitting winding **1**E or emitter **101**.

As shown in FIG. **6**, further aspects may include receiver **102** with primary receiver winding **1**R and variable pick-up tap **28**. For example, if it is desired to have 12 VDC for a standard DC-to-AC (60 hz) inverter, primary receiver winding **1**R may be tapped by variable tap **28**, such that the RF output may be about ~10 Vrms, which yields about ~14 Vdc, so up to about 2 Vdc may be lost by forward diode drops over rectifying diodes **24**. DC rectifying array **24** and capacitors **25**, **26**, and **27** may allow stable and constant DC power to inverter **31**. In further aspects, this may be a particularly useful configuration due to the widespread use of 60 Hz 120 VAC for power applications. However, many different DC and AC outputs could be synthesized from the signal generated by primary receiving winding **1**R. In one aspect, device **602** may receive DC power from the system by being connected to positive (+) output **27** through wire **601**, and device **32** may receive AC power from DC inverter **31**.

As shown in FIG. **7** according to aspects, emitter **101** and receiver **102** may be oriented in a manner whereby the windings could remain parallel but share a common coil axis. Alternatively, in further aspects, the coils may be oriented in a similar configuration, for example, as shown in FIG. **1**A, but without a common coil axis (not shown), for example, if two coils were placed on different floors of a building (like at the top-left corner of a second floor and bottom right corner of a first floor). In other aspects, emitter **101** and receiver **102** may be transposed such that top loads **30**E and **30**R are facing each other.

As shown in FIG. **8** according to further aspects, because it is possible to achieve high voltages on secondary windings **2**E and **2**R, an arc suppression system may be implemented on emitter **101**. FIG. **8** is illustrated to show such an aspect, given the cylinders labeled **38** surrounding windings **1**R, **2**R, **1**E, and **2**E are constructed as, for example, a gas tight enclosure. In another aspect, emitter **101** and/or receiver **102** may be immersed in an atmosphere of higher dielectric breakdown strength compared to air (for example, sulfur hexafluoride) to suppress corona and breakdown arcs. In one aspect, cylinders **38** may be vacuums.

In one aspect, emitter **101** and receiver **102** may be screened by a Faraday-cage type of enclosure **801** to reduce electric fields around the system. This aspect may operate as a transformer. Driving circuitry **4** may be inside or outside Faraday-cage type of enclosure **801**, but the power supply to driving circuitry **4** is outside Faraday-cage type of enclosure **801**. Wire **601** may transfer DC power to device **602** outside Faraday-cage type of enclosure **801**. In one aspect, Faraday-cage type of enclosure **801** is connected to ground **7**. In one aspect, secondary windings **2**E and **2**R are floating and, therefore, are not connected to grounds **8**E and **8**R.

As shown in FIGS. **9**A and **9**B, according to one aspect of the present disclosure, Tesla resonators **1501** and **1502** may couple energy to each other and may be "reversible," i.e., each Tesla resonator **1501** and **1502** may emit and receive energy so either one may be used. In FIG. **9**A, resonators **1501** and **1502** are each coupled to driving circuits **901** and **902** so each may act as an emitter and/or a receiver. In one aspect, if IGBT 35 of circuit **901** is set to closed by logic signal **200**, then circuit **901** would function as downstream circuitry of receiver **102** and, therefore, would ultimately produce DC current at DC out **36** (as shown in FIG. **6**). In one aspect, if IGBT 35 of circuit **901** is set to open by logic signal **20Q**, then circuit **901** would function as emitter **101**.

As shown in FIG. **10**, in one aspect, the system of the present disclosure may include receivers **1003**R, **1005**R, and **1006**R and emitters **1001**E, **1002**E, and **1004**E may be arranged to relay energy. In one aspect, two emitters **1001** and **1002** transmit energy wirelessly to a receiver **1003**R that is connected through, for example, wires **1008** and **1009** to a third emitter **1004**E. Emitter **1004**E may then wirelessly power three separate receivers **1005**R, **1006**R, and **1007**R. It should be noted that this illustration is provided as an example of a repeater winding network and not as a limiting design.

As shown in FIG. **11**, another aspect of the system of the present disclosure may include the use of imposing a modulation upon the basic carrier wave emitted from emitter **101**. Any number of digital or analog modulation methods or schemes may be applied to a basic carrier wave. After reception by receiver **102**, the carrier wave may be demodulated to provide information transfer while using the same carrier wave for power transfer. This simultaneous transfer of information and power may be effectuated by using the same devices discussed in aspects of the present invention, e.g., emitter **101** and receiver **102**.

For example, in one aspect, the system may transmit energy along with an audio signal. The audio signal may be modulated, such as with pulse width modulation (PWM) or amplitude modulation (AM) onto, for example, an 80 kHz power carrier signal from emitter **101**E to receiver **102**R embodied within a speaker assembly. One of ordinary skill in the art would understand that alternative modulation schemes, such as, frequency modulation (FM) or phase modulation, may be used with the system as described in aspects of the present disclosure.

In one aspect, signal input **1702** is modulated using modulator **1703** and combined with power input **1701** through the driving circuit **4**Q (as shown according in FIG. **2** according to one aspect of the present invention). In this aspect, the high-frequency side of the input signal becomes a mixer, imposing the data signal on AC input **99**, which is the carrier wave frequency, generated by driving circuit **4**Q. The speaker assembly with emitter **101** E may be placed anywhere within a range of receiver **102**R and be self-powered by reception of power from emitter **101**E to drive electronics to demodulate the carrier signal to reproduce the audio signal.

In further aspects, demodulator **1704** may be placed to sense combined signal **99** reproduced at the output of the receiver **102**R. The signal may be sensed by direct connection or by other means familiar to one of ordinary skill in the art. Data signal **1706** is then produced by demodulator **1704**. The carrier component of signal **99** is fed into circuitry **1705**, which may include diode array **24** (as shown in FIG. **4** according to one aspect), capacitor bank **401** (as shown in FIG. **4** according to one aspect), and circuitry **603** (as shown in FIG. **6** according to one aspect) as necessary to produce power.

In one aspect, circuitry for performing driving circuitry **4**Q and modulator **1703** are described in U.S. non-provisional patent application Ser. No. 12/152,525 titled "System and Method for Controlling an Electromagnetic Field Generator," filed May 15, 2008.

Although illustrative embodiments have been shown and described herein in detail, it should be noted and will be appreciated by those skilled in the art that there may be numerous variations and other embodiments that may be equivalent to those explicitly shown and described. For example, the scope of the present invention is not necessarily limited in all cases to execution of the aforementioned steps in the order discussed. Unless otherwise specifically stated, terms and expressions have been used herein as terms of description, not of limitation. Accordingly, the invention is not to be limited by the specific illustrated and described embodiments (or the terms or expressions used to describe them) but only by the scope of claims.

The invention claimed is:

1. A system for wirelessly transmitting and receiving energy, comprising:

an emitter, comprising:

a primary emitting winding wrapped around a first support and having first and second ends coupled to a driver; and

a secondary emitting winding wrapped around a second support and having a first end of the secondary emitting winding coupled to a first top load and a second end; and

a receiver, comprising:

a primary receiving winding wrapped around a third support and having first and second ends coupled to first and second inputs of a rectifying circuitry, respectively, where a first output of the rectifying circuitry is coupled to a first terminal of one or more capacitors and a second output of the rectifying circuitry is coupled to a second terminal of the one or more capacitors; and

a secondary receiving winding wrapped around a fourth support and having a first end coupled to a second top load and a second end;

wherein the second support is located within the first support and the fourth support is located within the third support.

2. The system of claim **1**, wherein the second support is approximately concentric with the first support and the fourth support is approximately concentric with the third support.

3. The system of claim **1**, further comprising a Faraday cage containing the emitter and the receiver.

4. The system of claim **1**, wherein the rectifying circuitry includes a diode array.

5. The system of claim **1**, wherein the one or more capacitors includes one or more snubber capacitors, low inductance polymeric capacitors, and electrolytic capacitors.

6. The system of claim **1**, wherein a device is coupled to a positive output of the one or more capacitors.

7. The system of claim **1**, wherein the driver includes a feedback system and insulated gate bipolar transistors in an H-bridge configuration.

8. The system of claim **1**, further comprising a variable tap coupled to the primary receiving winding.

9. The system of claim **1**, further comprising a first gas-tight container surrounding the emitter and a second gas-tight container surrounding the receiver.

10. The system of claim **9**, wherein the first and second gas-tight containers include a gas with a high-dielectric atmosphere.

11. The system of claim **9**, wherein the first and second gas-tight containers are vacuums.

12. The system of claim **1**, wherein the first, second, third, and fourth supports are non-conductive tubes.

13. The system of claim **1**, wherein the second end of the secondary emitting winding is coupled to a first ground.

14. The system of claim **1**, wherein the second end of the secondary receiving winding is coupled to a first ground.

15. The system of claim **13**, wherein the second end of the secondary receiving winding is coupled to a second ground.

16. The system of claim **1**, wherein the one or more capacitors comprises a capacitor bank.

17. A system for wirelessly transmitting and receiving energy, comprising:

an emitter, comprising:

a primary emitting winding wrapped around a first support and having first and second ends coupled to a driver;

a secondary emitting winding wrapped around a second support and having first and second ends; and

a tertiary emitting winding wrapped around a third support and having a first end of the tertiary emitting winding coupled to the first end of the secondary emitting winding and the second end of the tertiary emitting winding coupled to a first top load; and

a receiver, comprising:

a primary receiving winding wrapped around a fourth support and having first and second ends coupled to first and second inputs of a rectifying circuitry, respectively, where a first output of the rectifying circuitry is coupled to a first terminal of one or more capacitors and a second output of the rectifying circuitry is coupled to a second terminal of the one or more capacitors;

a secondary receiving winding wrapped around a fifth support and having first and second ends; and

a tertiary receiving winding wrapped around a sixth support and having a first end of the tertiary receiving winding coupled to the first end of the secondary receiving winding and the second end of the tertiary receiving winding coupled to a second top load;

wherein the first support is located within the second support, the fourth support is located within the fifth support, the third support is located above the first and second supports, and the sixth support is located above the fourth and fifth supports.

18. The system of claim **17**, wherein the second end of the secondary emitting winding is coupled to a first ground.

19. The system of claim **17**, wherein the second end of the secondary receiving winding is coupled to a first ground.

20. The system of claim **18**, wherein the second end of the secondary emitting winding is coupled to a first ground.

21. A method of wirelessly transferring data and energy, comprising:

(a) inputting a data signal;

(b) inputting power to a driver;

(c) generating a carrier wave with the driver at a resonant frequency of an emitter;

(d) modulating the carrier wave based on the data signal to create a combined signal;

(e) driving the combined signal to an AC input on the emitter;

(f) electromagnetically coupling the emitter to a receiver;

(g) reproducing the combined signal in an AC output of the receiver;

(h) rectifying the AC output into DC power;

(i) demodulating the combined signal into the data signal; and

(j) transmitting the data signal to a device.

22. The method of claim **21**, wherein the data signal is an audio signal.

23. The method claim **21**, wherein the data signal is modulated using one of amplitude modulation, pulse width modulation, frequency modulation, and phase modulation.

24. The method of claim **21**, wherein the device is an audio speaker.

25. The method of claim **21**, wherein the rectifying the AC output into DC power is done through a rectifying circuitry and a capacitor bank.

26. The method claim **21**, wherein the driver includes a feedback system and insulated gate bipolar transistors in an H-bridge configuration.

27. The method of claim **21**, wherein the modulating step further comprises modulating the amplitude of the carrier wave.

28. The method of claim **21**, wherein,

the emitter comprises:

a primary emitting winding wrapped around a first support and having first and second ends coupled to a driver;

a secondary emitting winding wrapped around a second support and having first and second ends; and

a tertiary emitting winding wrapped around a third support and having a first end of the tertiary emitting winding coupled to the first end of the secondary emitting winding and the second end of the tertiary emitting winding coupled to a first top load; and

the receiver comprises:

a primary receiving winding wrapped around a fourth support and having first and second ends coupled to first and second inputs of a rectifying circuitry, respectively, where a first output of the rectifying circuitry is coupled to a first terminal of one or more capacitors and a second output of the rectifying circuitry is coupled to a second terminal of the one or more capacitors;

a secondary receiving winding wrapped around a fifth support and having first and second ends; and

a tertiary receiving winding wrapped around a sixth support and having a first end of the tertiary receiving winding coupled to the first end of the secondary receiving winding and the second end of the tertiary receiving winding coupled to a second top load.

29. The method of claim **21**, wherein

the emitter comprises:

a primary emitting winding wrapped around a first support and having first and second ends coupled to a driver;

a secondary emitting winding wrapped around a second support and having a first end of the secondary emitting winding coupled to a first top load and a second end; and

the receiver comprises:

a primary receiving winding wrapped around a third support and having first and second ends coupled to first and second inputs of a rectifying circuitry, respectively, where a first output of the rectifying circuitry is coupled to a first terminal of one or more capacitors and a second output of the rectifying circuitry is coupled to a second terminal of the one or more capacitors; and

a secondary receiving winding wrapped around a fourth support and having a first end coupled to a second top load and a second end.

30. A system for wirelessly transmitting and receiving energy, comprising:

an emitter, comprising:

a primary emitting winding wrapped around a first support and having first and second ends coupled to a first driver, wherein the first driver is coupled to first and second inputs of a rectifying circuitry respectively, where a first output of the rectifying circuitry is coupled to one or more capacitors through a first terminal of the one or more capacitors and a second output of the rectifying circuitry is connected to the one or more capacitors through a second terminal of the one or more capacitors; and

a secondary emitting winding wrapped around a second support and having a first end of the secondary emitting winding coupled to a first top load and a second end; and

a receiver, comprising:

a primary receiving winding wrapped around a third support and having first and second ends coupled to a second driver, wherein the second driver is coupled to first and second inputs of a rectifying circuitry respectively, where a first output of the rectifying circuitry is coupled to one or more capacitors through a first terminal of the one or more capacitors and a second output of the rectifying circuitry is connected to the one or more capacitors through a second terminal of the one or more capacitors; and

a secondary receiving winding wrapped around a fourth support and having a first end coupled to a second top load and a second end;

wherein the first support is located within the second support and the third support is located within the fourth support.

31. A method for wirelessly transmitting and receiving energy, comprising:

(a) transmitting an alternating current into a driver coupled to a primary emitting winding of an emitter;

(b) inducing the alternating current into a secondary emitting winding of the emitter;

(c) sensing a frequency of the alternating current on the secondary emitting winding;

(d) transmitting a feedback signal to the driver based on the frequency;

(e) driving the AC current into the primary emitting winding at the resonant frequency of the secondary emitting winding;

(f) electromagnetically coupling the emitter to a receiver;

(g) inducing the alternating current into a secondary receiving winding of the receiver;

(h) inducing the alternating current into a primary receiving winding of the receiver;

(i) outputting the alternating current into a rectifying circuitry;

(j) rectifying the alternating current into a direct current;

(k) reducing ripples on the direct current; and

(l) transmitting the direct current to a device.

32. A system for wirelessly transmitting and receiving energy, comprising:

a reversible emitter, comprising:

a primary emitting winding wrapped around a first support and having first and second ends coupled to a first driver, wherein the first driver is coupled to a first rectifying circuitry, wherein the first rectifying circuitry is coupled to a first one or more capacitors;

a secondary emitting winding wrapped around a second support and having first and second ends; and

a tertiary emitting winding wrapped around a third support and having a first end of the tertiary emitting winding coupled to the first end of the secondary emitting winding and the second end of the tertiary emitting winding coupled to a first top load; and

a reversible receiver, comprising:

a primary receiving winding wrapped around a fourth support and having first and second ends coupled to a second driver, wherein the second driver is coupled to a second rectifying circuitry, wherein the second rectifying circuitry is coupled to a second one or more capacitors;

a secondary receiving winding wrapped around a fifth support and having first and second ends; and

a tertiary receiving winding wrapped around a sixth support and having a first end of the tertiary receiving winding coupled to the first end of the secondary receiving winding and the second end of the tertiary receiving winding coupled to a second top load;

wherein the second support is located within the first support, the fifth support is located within the fourth support, the third support is located above the first and second supports, and the sixth support is located above the fourth and fifth supports.

33. The system of claim **32**, wherein

the first driver comprises:

insulated gate bipolar transistors in an H-bridge configuration; and

a switch for activating the first rectifying circuitry and the first one or more capacitors; and

the second driver comprises:

insulated gate bipolar transistors in an H-bridge configuration; and

a switch for activating the second rectifying circuitry and the second one or more capacitors.

34. The system of claim **1**, wherein the fourth support surrounds a form including a high permeability material.

35. The system of claim **17**, wherein the fifth and sixth supports each surround forms including a high permeability material.

36. The method of claim **28**, wherein the fifth and sixth supports each surround forms including a high permeability material.

37. The system of claim **29**, wherein the fourth support surrounds a form including a high permeability material.

38. The system of claim **30**, wherein the fourth support surrounds a form including a high permeability material.

39. The system of claim **32**, wherein the fifth and sixth supports each surround forms including a high permeability material.

* * * * *

US 20160350458A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2016/0350458 A1**

Hively (43) **Pub. Date: Dec. 1, 2016**

(54) **SYSTEMS, APPARATUSES, AND METHODS FOR GENERATING AND/OR UTILIZING SCALAR-LONGITUDINAL WAVES**

(71) Applicant: **GRADIENT DYNAMICS LLC**, McLean, VA (US)

(72) Inventor: **Lee M. Hively**, Philadelphia, TN (US)

(73) Assignee: **GRADIENT DYNAMICS LLC**

(21) Appl. No.: **14/738,775**

(22) Filed: **Jun. 12, 2015**

Related U.S. Application Data

(63) Continuation of application No. 14/726,305, filed on May 29, 2015, now Pat. No. 9,306,527.

Publication Classification

(51) **Int. Cl.**

G06F 17/50	(2006.01)
H04W 72/04	(2006.01)
G06F 17/10	(2006.01)

(52) **U.S. Cl.**

CPC ***G06F 17/5009*** (2013.01); ***G06F 17/10*** (2013.01); ***H04W 72/0453*** (2013.01)

(57) **ABSTRACT**

Scalar-longitudinal waves (SLWs) may be transmitted and/or received. A first apparatus configured to transmit and/or receive SLWs may include a linear first conductor configured to operate as a linear monopole antenna at a first operating frequency. The first apparatus may include a tubular second conductor coaxially aligned with the first conductor and an annular balun configured to cancel most or all return current on an outer surface of the second conductor during operation such that the first conductor transmits or receives SLWs. A second apparatus configured to transmit and/or receive scalar-longitudinal waves may include a bifilar coil formed in an alternating fashion of a first conductor and a second conductor such that an electrical current in the coil will propagate in opposite directions in adjacent turns of the coil thereby cancelling any magnetic field so that during operation the coil transmits or receives SLWs.

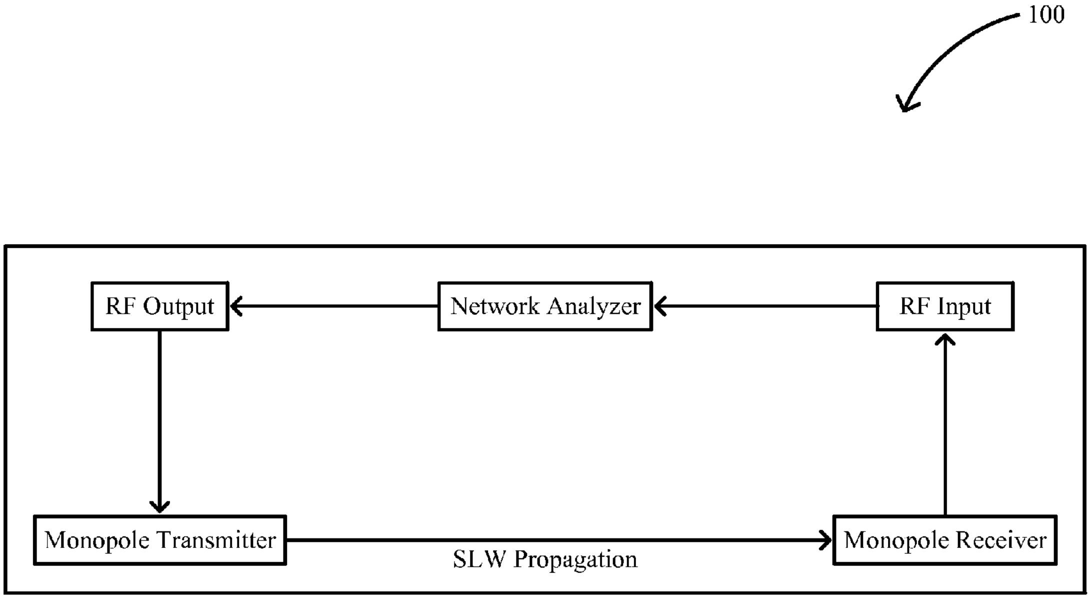

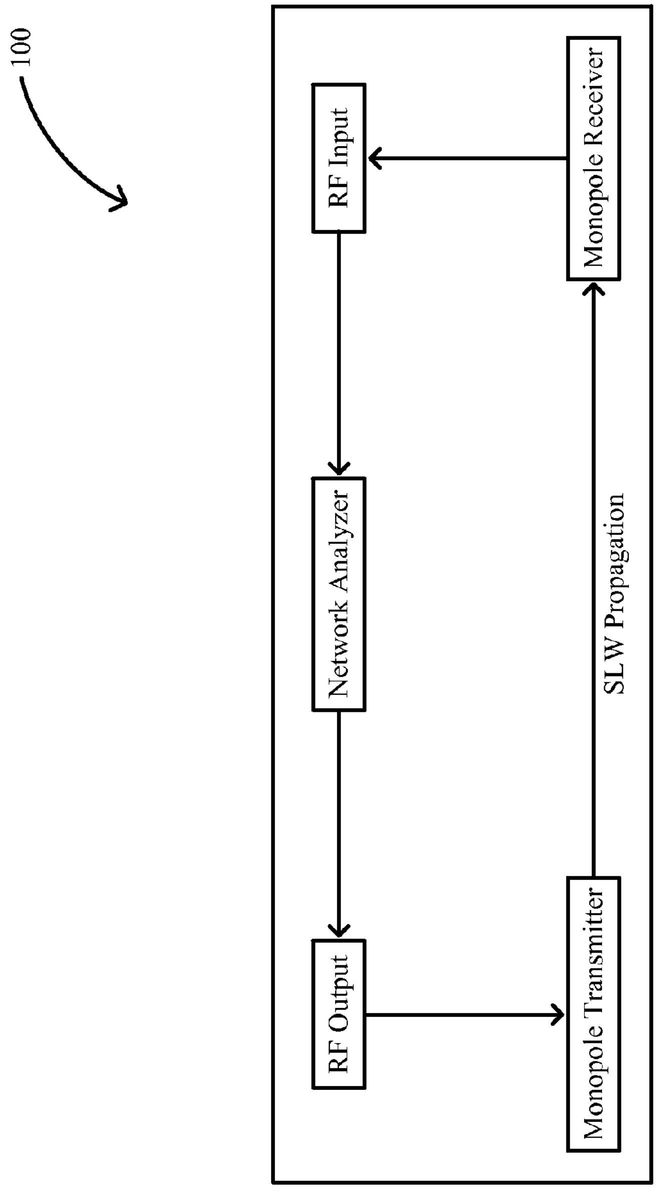
100
RF Input
Monopole Receiver
Network Analyzer
SLW Propagation
RF Output
Monopole Transmitter

FIG. 1

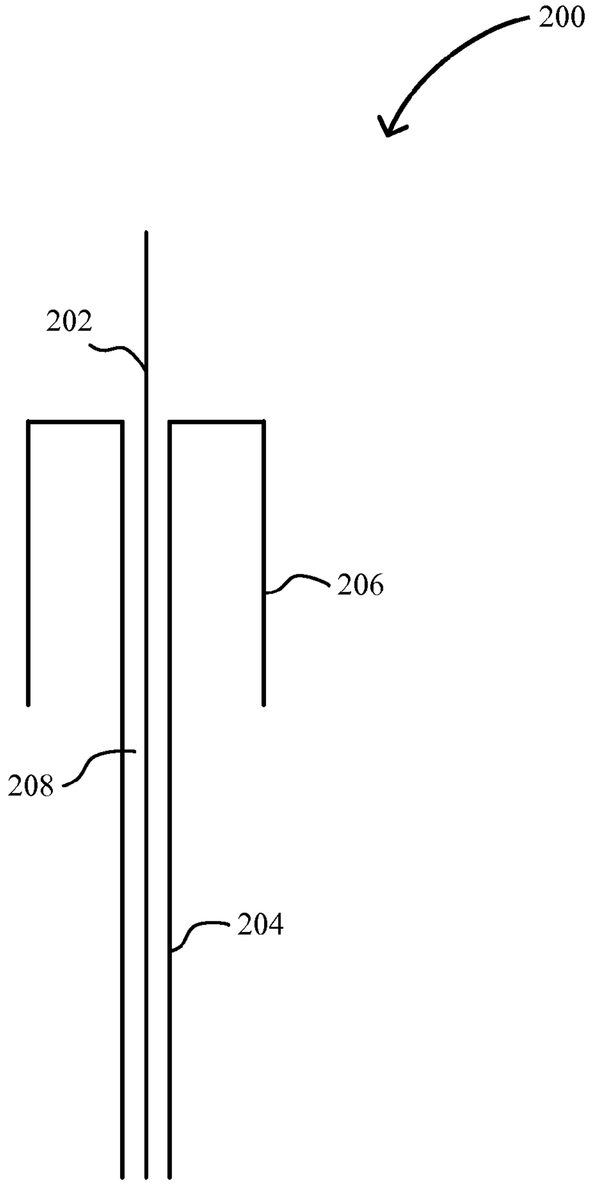
200
202
206
208
204

FIG. 2A

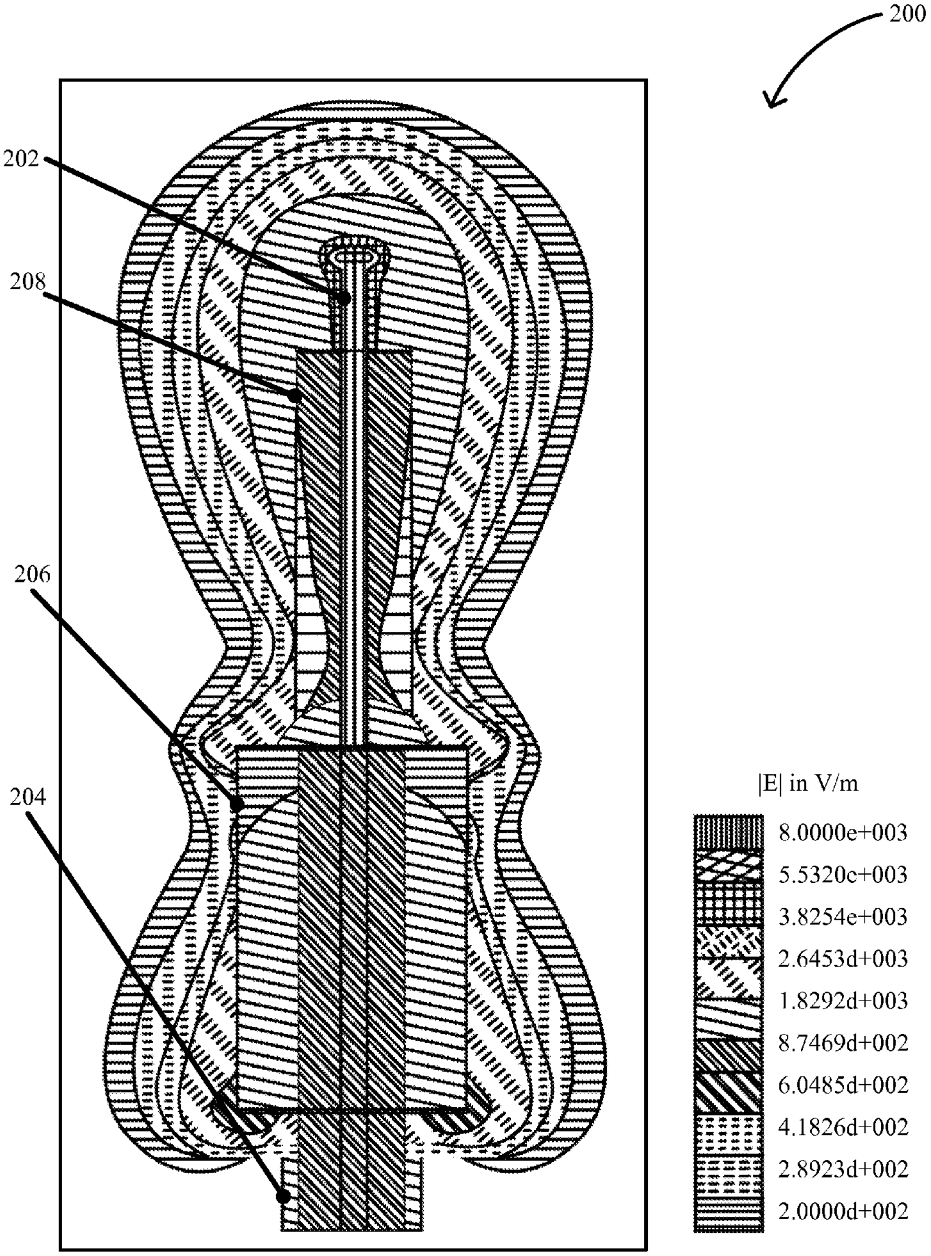
200
202
208
206
204
|E| in V/m
8.0000e+003
5.5320c+003
3.8254e+003
2.6453d+003
1.8292d+003
8.7469d+002
6.0485d+002
4.1826d+002
2.8923d+002
2.0000d+002

FIG. 2B

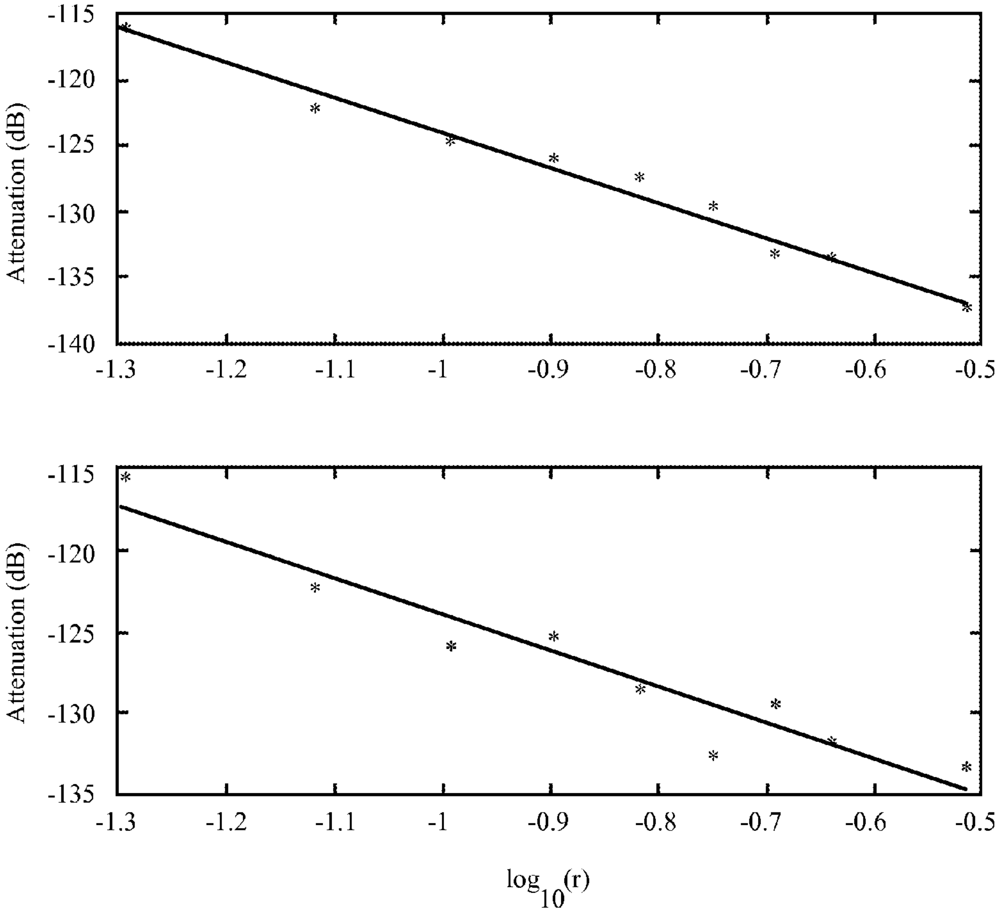
-115
-120
-125
-130
-135
-140
Attenuation (dB)
-1.3
-1.2
-1.1
-1
-0.9
-0.8
-0.7
-0.6
-0.5
-115
-120
-125
-130
-135
Attenuation (dB)
-1.3
-1.2
-1.1
-1
-0.9
-0.8
-0.7
-0.6
-0.5
$\log_{10}(r)$

FIG. 3

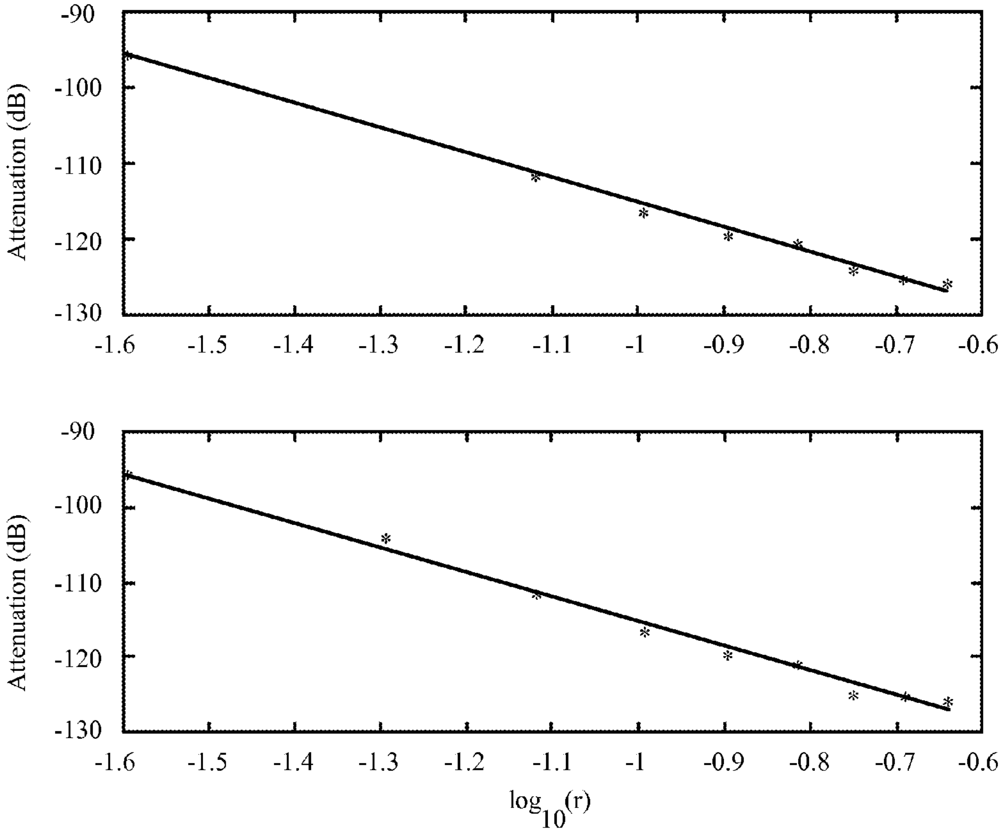
-90
-100
-110
-120
-130
Attenuation (dB)
-1.6
-1.5
-1.4
-1.3
-1.2
-1.1
-1
-0.9
-0.8
-0.7
-0.6
-90
-100
-110
-120
-130
Attenuation (dB)
-1.6
-1.5
-1.4
-1.3
-1.2
-1.1
-1
-0.9
-0.8
-0.7
-0.6
log$_{10}$(r)

FIG. 4

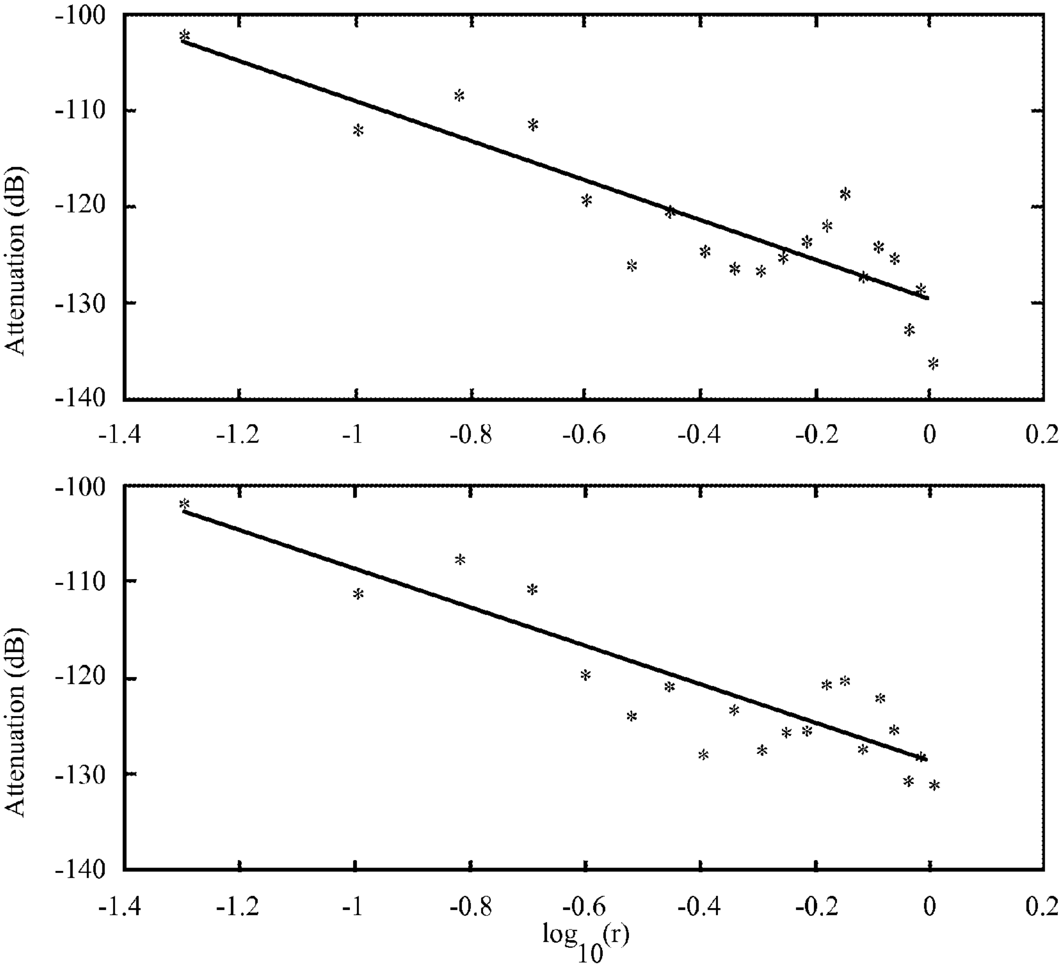
-100
-110
-120
-130
-140
Attenuation (dB)
-1.4
-1.2
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
-100
-110
-120
-130
-140
Attenuation (dB)
-1.4
-1.2
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
$\log_{10}(r)$

FIG. 5

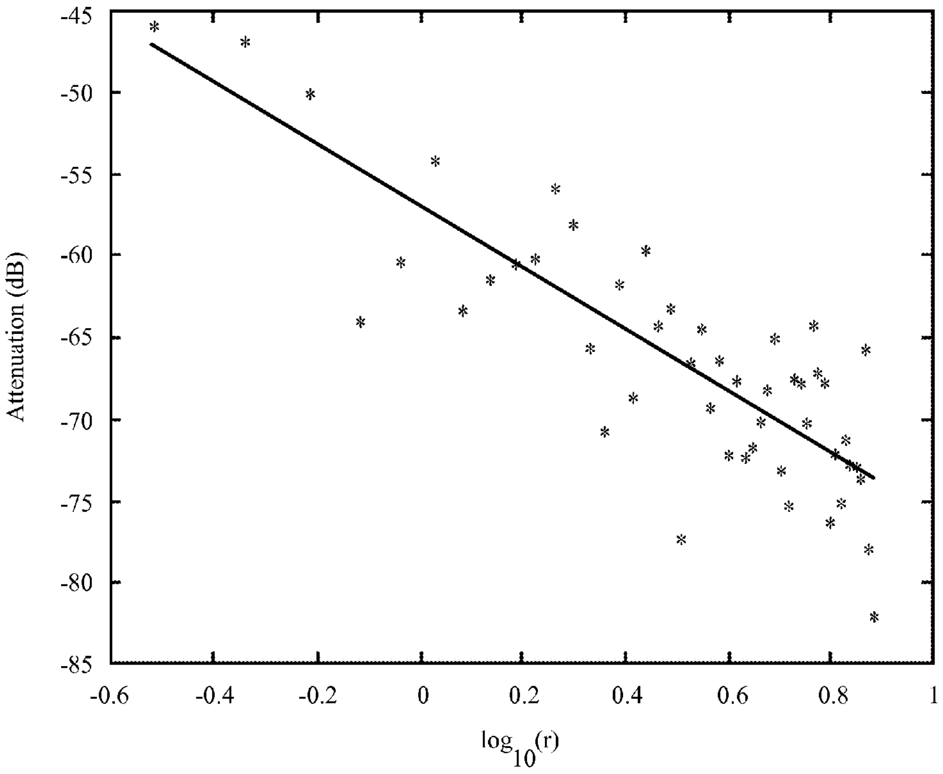
-45
-50
-55
-60
-65
-70
-75
-80
-85
Attenuation (dB)
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
$\log_{10}(r)$

FIG. 6

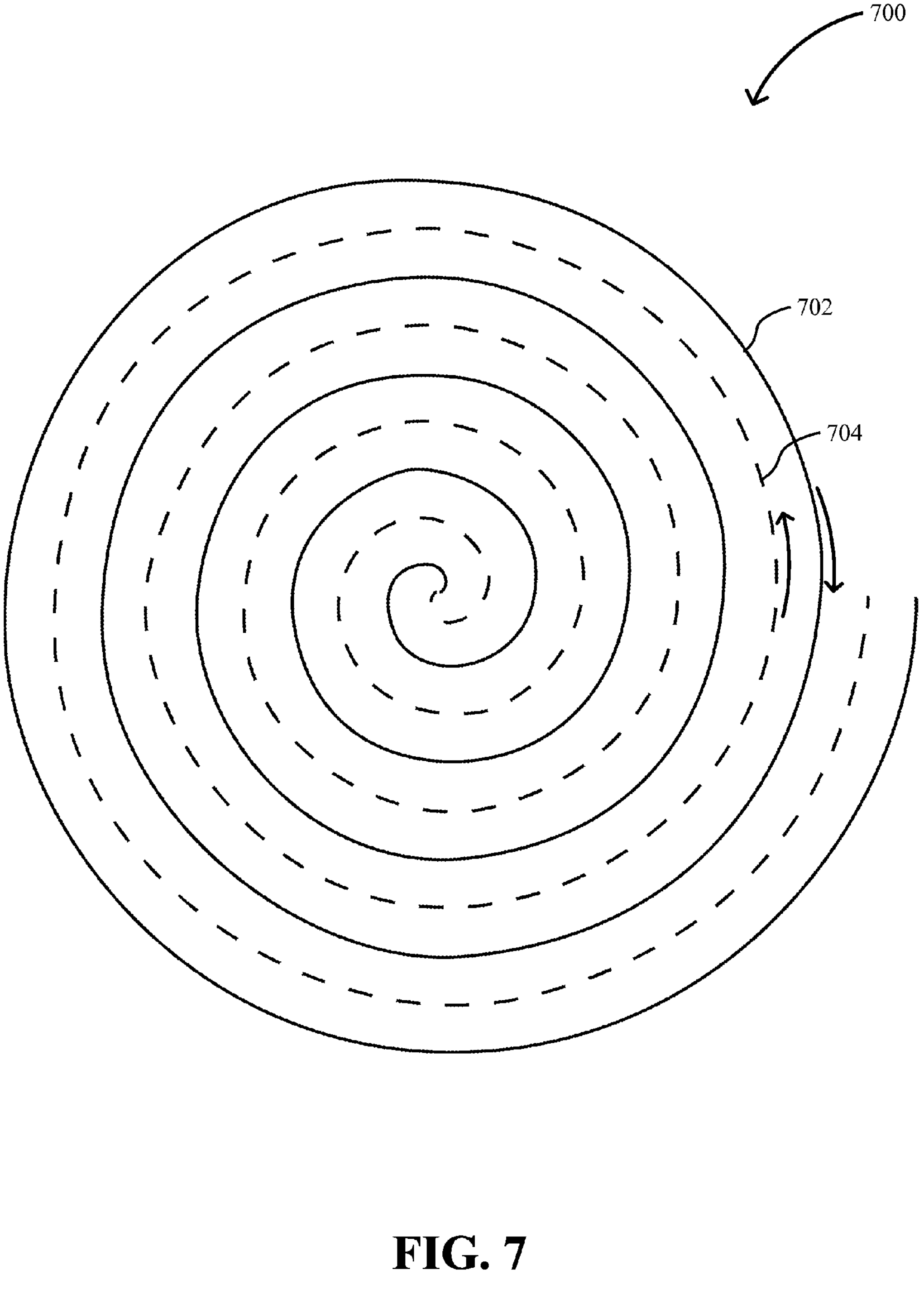

FIG. 7

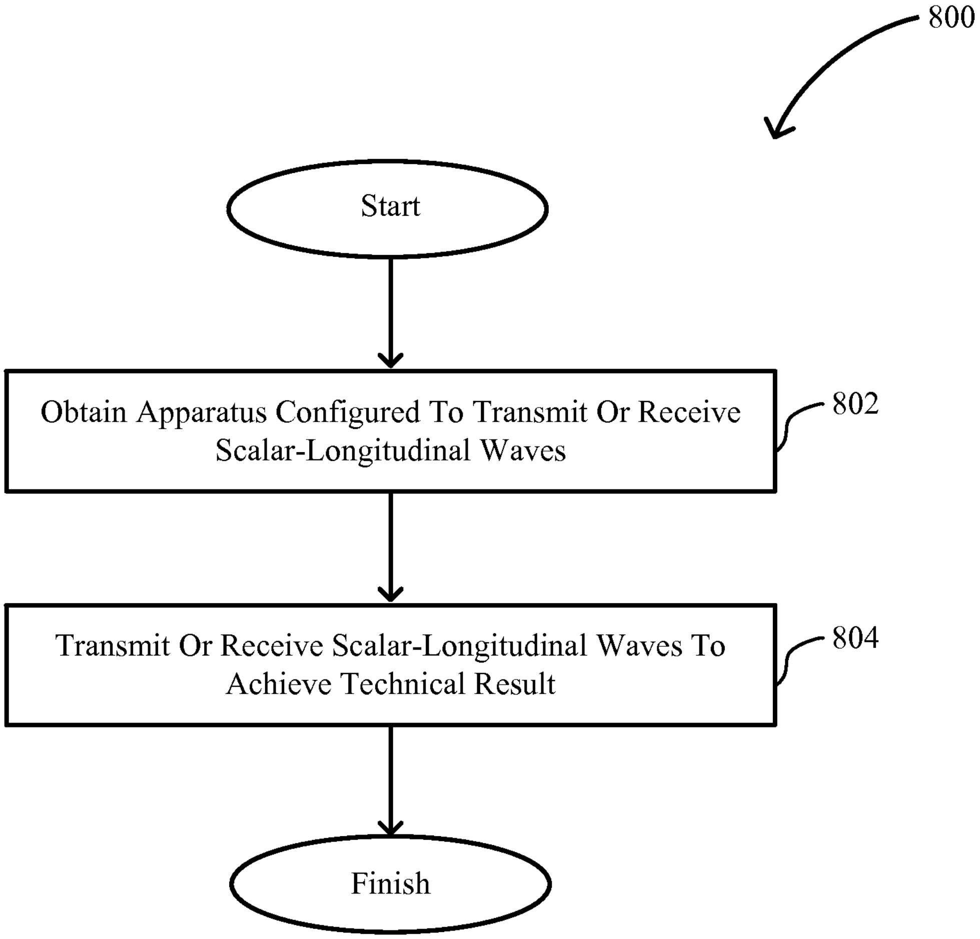
800
Start
Obtain Apparatus Configured To Transmit Or Receive Scalar-Longitudinal Waves
802
Transmit Or Receive Scalar-Longitudinal Waves To Achieve Technical Result
804
Finish

FIG. 8

SYSTEMS, APPARATUSES, AND METHODS FOR GENERATING AND/OR UTILIZING SCALAR-LONGITUDINAL WAVES

FIELD OF THE DISCLOSURE

[0001] This disclosure relates to systems, apparatuses, and methods for generating and/or utilizing scalar-longitudinal waves.

BACKGROUND

[0002] Classical electrodynamic theory may be regarded as central to physics. An electric field (E) arises from an electric charge density (ρ). Charge motion creates an electrical current density (J) that drives dynamical changes in E and the magnetic field, B. The classical electrodynamic model is based on a coupled set of partial-differential equations for the quantities (B, E, J, ρ).

[0003] Classical electromagnetics predicts no wave creation by radial motion of a charged sphere. More specifically, spherical symmetry of the radial electric field on a charged, oscillating sphere implies a curl-free electric field ($\nabla\times E=0$), which in turn yields no variation in magnetic field from Faraday's law ($\nabla\times E=-\partial B/\partial t=0$), corresponding to no magnetic wave. Thus, the Poynting vector, $E\times B/\mu$, is zero, resulting in no classical electromagnetic radiation. This statement applies more generally to no creation of electrical waves by radial motion of any extended charge distribution. An electrically equivalent antenna may include a classical linear monopole that is driven by a sinusoidal current to put charge onto and remove charge from the linear conductor (antenna).

SUMMARY

[0004] Exemplary implementations of the disclosure provide and/or facilitate transmission and/or reception of scalar-longitudinal waves (SLW), together with technology and/or applications using those waves. More specifically, this disclosure includes inter alia: (1) a more complete electrodynamics (MCE) model that may remove and/or lessen incompleteness and/or inconsistency in classical electrodynamics; (2) verification of a scalar-longitudinal wave (SLW) that arises from a gradient-driven current density; (3) SLW antenna apparatus designs; (4) experimental data demonstrating that the SLW exists and can be transmitted and received by SLW antenna apparatuses; (5) experimental data showing the SLW is not subject to the classical skin effect, as predicted by the MCE theory; (6) technology applications of the scalar-longitudinal waves; and (7) additional applications that arise from MCE.

[0005] One aspect of the disclosure relates to an apparatus configured to transmit or receive scalar-longitudinal waves. The apparatus may include a linear first conductor configured to operate as a linear monopole antenna at a first operating frequency. The apparatus may include a tubular second conductor coaxially aligned with the first conductor such that the first conductor extends out in a first direction from within the second conductor. The apparatus may include an annular skirt balun disposed at an end of the second conductor from which the first conductor extends. The balun may have a larger diameter than the second conductor. The balun may extend in a second direction opposite the first direction. The balun may be configured to cancel most or all return current on an outer surface of the second conductor during operation such that the first conductor transmits or receives scalar-longitudinal waves.

[0006] Another aspect of the disclosure relates to an apparatus configured to transmit or receive scalar-longitudinal waves. The apparatus may include a bifilar coil formed in an alternating fashion of a first conductor and a second conductor such that a given turn of the coil that is made of the first conductor is adjacent on either side to turns of the coil made of the second conductor. The first conductor and the second conductor may be conductively coupled such that an electrical current in the coil will propagate in opposite directions in adjacent turns of the coil thereby cancelling any magnetic field so that during operation the coil transmits or receives scalar-longitudinal waves.

[0007] These and other features, and characteristics of the present technology, as well as the methods of operation and functions of the related elements of structure and the combination of parts and economies of manufacture, will become more apparent upon consideration of the following description and the appended claims with reference to the accompanying drawings, all of which form a part of this specification, wherein like reference numerals designate corresponding parts in the various figures. It is to be expressly understood, however, that the drawings are for the purpose of illustration and description only and are not intended as a definition of the limits of the invention. Also, it is to be expressly understood that permissive language (e.g., "may") used in the specification to describe the present technology conveys a present understanding of the underlying science, but any inadequacies in that understanding should not be used to limit the claims. As used in the specification and in the claims, the singular form of "a", "an", and "the" include plural referents unless the context clearly dictates otherwise.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] FIG. **1** illustrates a system configured to transmit and/or receive scalar-longitudinal waves, in accordance with one or more implementations.

[0009] FIG. **2**A illustrates a cross-sectional view of a linear monopole antenna apparatus **200** configured to transmit and/or receive scalar-longitudinal waves, in accordance with one or more implementations.

[0010] FIG. **2**B illustrates a cross-sectional view of constant electric field magnitude contours for a monopole antenna with a skirt balun, in accordance with one or more implementations.

[0011] FIGS. **3**, **4**, **5**, and **6** show experimental results for scalar-longitudinal wave attenuation under various conditions.

[0012] FIG. **7** illustrates a bifilar coil apparatus configured to transmit and/or receive scalar-longitudinal waves, in accordance with one or more implementations.

[0013] FIG. **8** illustrates a method for utilizing scalar-longitudinal waves, in accordance with one or more implementations.

DETAILED DESCRIPTION

[0014] This disclosure may demonstrate measurement of scalar-longitudinal wave (SLW). The SLW may have $C=\nabla\cdot A+\epsilon\mu\partial\Phi/\partial t$ as a non-zero, dynamical, scalar field, together with a longitudinal E-field. A and Φ may represent the vector and scalar electrical potentials, respectively. Here,

∈ and μ may represent the electrical permittivity and permeability (not necessarily vacuum values), respectively. The SLW may have no magnetic component. Exemplary implementations may provide and/or facilitate transmission and/or reception of the SLW via linear monopole antennas and/or tightly-wound, bifilar, helical coils. Exemplary implementations may facilitate a 1S to 2S atomic transition (mS to nS transition in general). The measurements show that the SLW can be transmitted through a thick Faraday cage or box (thousands of classical skin depths thick) to a companion SLW receiver, either enclosed in a separate Faraday box or without a Faraday box. This result may not be explainable by radiation of classical transverse waves because such waves are eliminated via Faraday cage(s) around the transmitting and/or receiving antennas.

[0015] The need for an additional (arbitrary) assumption to solve Maxwell's equations was recognized in the late 1800s. A word equivalent to "arbitrary" was chosen, analogous to the width variation (gauge) of railroad tracks at that time. The gauge function (Λ) comes from the vector-field solution for the magnetic field, $B=\nabla\times A$. Here, A may describe the vector potential with infinitely many choices, $A\rightarrow A+\nabla\Lambda$, while B is unchanged. The electric field may be represented by $E=-\nabla\Phi-\partial A/\partial t$. Here, Φ may represent the electric potential with an arbitrariness for the same reason, $\Phi\rightarrow\Phi-\partial\Lambda/\partial t$, while E is unchanged. An example may include $\nabla\cdot A+\alpha\in\mu\ \partial\Phi/\partial t=0$. The permittivity and permeability may be represented by ∈ and μ (not necessarily vacuum), respectively. The Lorenz gauge ($\alpha=1$) may imply that the effect of a charge source propagates at the speed of light, c. The Coulomb gauge ($\alpha=0$) may yield electrostatics with Φ propagation at infinite speed. The velocity gauge ($0<\alpha<1$) may imply Φ propagation at a speed, c/α. The need for a gauge condition may imply that classical electrodynamics is incomplete.

[0016] Insight into the incompleteness of classical electrodynamics may begin with the Helmholtz theorem, which states that a sufficiently smooth three-dimensional vector field (F) can be uniquely decomposed into two parts, $F=-\nabla\Phi+\nabla\times A$. A generalized theorem may exist for unique decomposition of a sufficiently smooth, Minkowski four-vector field (three spatial dimensions, plus time) into four-irrotational and four-solenoidal parts, together with the tangential and normal components on the bounding three-surface. With this background, a general electromagnetic Lagrangian density may be written as:

$$\mathcal{L}=-\frac{\varepsilon c^2}{4}F_{\mu\nu}F^{\mu\nu}+J_\mu A^\mu-\frac{\gamma\varepsilon c^2}{2}(\partial_\mu A^\mu)^2-\frac{\varepsilon c^2k^2}{2}(A_\mu A^\mu). \tag{1}$$

$F^{\mu\nu}$ may represent the Maxwell field tensor; ∈ may represent the medium's permittivity (not necessarily vacuum); c may represent the speed of light with $c^2=1/\in\mu$ (not necessarily vacuum); the four-current may be represented by $J_\mu=(\rho c,J)$; $A_\mu=(\Phi/c, A)$ may represent the four-potential; the Compton wave number for a photon with mass, m, may be $k=2\pi mc/h$; and h may represent Planck's constant. For $\gamma=0$ and $m>0$, EQN. 1 may yield the Maxwell-Proca theory, for which recent tests set an upper bound of $m\leq10^{-54}$ kg, consistent with massless photons. For $\gamma=1$ and $m=0$, EQN. 1 may be written in terms of the potentials:

$$\mathcal{L}_{EM}=\frac{\varepsilon c^2}{2}\left[\frac{1}{c^2}\left(\nabla\Phi+\frac{\partial A}{\partial t}\right)^2-(\nabla\times A)^2\right]-\rho\Phi+J\cdot A-\frac{\varepsilon c^2}{2}\left(\frac{1}{c^2}\frac{\partial\Phi}{\partial t}+\nabla\cdot A\right)^2. \tag{2}$$

[0017] EQN. 2 may allow two potentially physical classes of four-vector fields. One class may have zero divergence, $C=\partial_\mu A_\mu=0$, consistent with classical electrodynamics. The second class may have zero curl of A^μ:$F^{\mu\nu}=\partial^\mu A^\nu-\partial^\nu A^\mu=\partial^\mu\partial^\nu\Lambda-\partial^\nu\partial^\mu\Lambda=0$ with a solution, $A^\mu=\partial^\mu\Lambda$, and a dynamical quantity, $C=\partial_\mu A^\mu=\partial_\mu\partial^\mu\Lambda$. Here, Λ may represent a scalar function of space and time. A more complete electrodynamic model (MCE) may be derived from EQN. 2 with $C\neq0$:

$$E=-\nabla\Phi-\frac{\partial A}{\partial t}; \tag{3}$$

$$B=\nabla\times A; \tag{4}$$

$$C=\nabla\cdot A+\frac{1}{c^2}\frac{\partial\Phi}{\partial t}; \tag{5}$$

$$\nabla\times B-\frac{1}{c^2}\frac{\partial E}{\partial t}-\nabla C=\mu J; \tag{6}$$

$$\nabla\cdot E=\frac{\partial C}{\partial t}=\frac{\rho}{\varepsilon}. \tag{7}$$

[0018] The homogeneous equations ($\nabla\cdot B=0$; $\nabla\times E+\partial B/\partial t=0$) may be unchanged from the classical model. EQNS. 3-7 may have the caveat that EQN. 2 is based on the least-action principle, requiring a finite, lower bound on the Lagrangian density. However, EQN. 2 has ($-C^2/2\mu$), which may imply that an arbitrarily fast change in Φ over time (or an arbitrarily rapid change in A over space) can make the action arbitrarily large and negative, in violation of the least-action principle. This issue may be resolved by noting that EQNS. 1-7 are based on partial derivatives, i.e., infinitesimal limits over time and space. However, the Planck scale may provide finite limits, where quantum effects of gravity become strong, corresponding to time and length scales of 5.4×10^{-44} s and 1.6×10^{-35} m, respectively. The Planck scales may be experimentally inaccessible and may be indistinguishable from infinitesimal, thus providing a finite lower bound for EQN. 2. Moreover, EQN. 2 without the new term still has $-(\nabla\times A)^2/2\mu=-B^2/2\mu$, to which a finite lower bound may apply for least action. Classical electrodynamics has been well validated against experiments, so the presence of the new term, ($-C^2/2\mu$, may not require any modification in the model of EQNS. 1-7.

[0019] EQNS. 3-7 may lead to important predictions: (1) relativistic covariance; (2) classical fields (B and E) in terms of the usual classical potentials (A and Φ); (3) classical wave equations for A, B, E, and Φ without use of a gauge condition; and (4) a scalar-longitudinal wave (SLW), composed of the scalar and longitudinal-electric fields. Regarding item (3), the MCE theory may produce cancellation of $\partial C/\partial t$ and $-\nabla C$ in the classical wave equations for Φ and A, thus eliminating the need for a gauge condition (and its attendant incompleteness) in the classical electrodynamics. A necessary and sufficient condition for the SLW may be that $B=0$. The wave equation for a null magnetic field may be shown as ${}^2B=\mu\nabla\times J=0$, which may imply that $J=\nabla\kappa$ as a result of the vector calculus identity, $\nabla\times\nabla\kappa=0$. Here, κ may be a scalar function of space and time. J for the SLW may be gradient driven and thus may be uniquely detectable, in contrast to classical waves that arise from a solenoidal current density ($\nabla\times J\neq 0$). Moreover, the gradient-driven current density may correspond to a longitudinal E-field in linearly conductive media, since $E=J/\sigma=\nabla\kappa/\sigma$.

[0020] A wave equation for C may arise by use of $\in\mu(\partial/\partial t)$ on EQN. 7, added to the divergence of EQN. 6:

$$\frac{\partial^2 C}{\partial c^2 t^2}-\nabla^2 C\equiv\Box^2 C=\mu\left(\frac{\partial\rho}{\partial t}+\nabla\cdot J\right). \tag{8}$$

The D'Alembertian may be represented by ^{2}; $\in\mu$ may be $1/c^2$ in the propagation medium (not necessarily vacuum). Use of C from EQN. 5 in EQN. 8 may yield an identity via the classical wave equations for Φ and A, and the vector calculus identity for $\nabla\times\nabla\times A=\nabla(\nabla\cdot A)-\nabla^2 A=0$ to give $\nabla^2\nabla\cdot A=\nabla\cdot\nabla^2 A$ since $B=0$ for the SLW. Charge conservation may give zero on the right-hand side (RHS) of EQN. 8:

$$\Box^2 C=0. \tag{9}$$

[0021] EQN. 9 may provide wave-like solutions, with the lowest-order form in a spherically symmetric geometry at a distance (r), $C=C_o\exp[j(kr-\omega t)]/r$. The boundary condition, $C(r\rightarrow\infty)\rightarrow 0$, may be trivially satisfied. The scalar field's energy density may be $(C^2/2\mu)$, yielding a constant energy, $4\pi r^2(C^2/2\mu)$, through a spherical boundary in arbitrary media. Classical electrodynamics may forbid a spherically symmetric, transverse wave. This constraint may be absent under the MCE theory, because the SLW may correspond to a gradient-driven current. The divergence theorem on EQN. 8 may yield interface matching in the normal component ('n') of $\nabla C/\mu$:

$$\left(\frac{\nabla C}{\mu}\right)_{1n}=\left(\frac{\nabla C}{\mu}\right)_{2n}. \tag{10}$$

The subscripts in EQN. 10 may denote ∇C in medium 1 or medium 2, respectively.

[0022] The wave equation for E may come from the curl of Faraday's law, use of $\nabla\times B$ from EQN. 6, and substitution for $\nabla\cdot E$ from EQN. 7 with cancellation of the terms $\nabla(\partial C/\partial t)-(\partial/\partial t)\nabla C$:

$$\frac{\partial^2 E}{\partial c^2 t^2}-\nabla^2 E=\Box^2 C=\mu\frac{\partial J}{\partial t}-\frac{\nabla\rho}{\varepsilon}. \tag{11}$$

EQN. 11 may represent the classical E-wave form. A time derivative of EQN. 11, and the use of classical charge conservation ($\partial\rho/\partial t=-\nabla\cdot J$) may yield ${}^2\dot{E}=-\mu\ddot{J}+\nabla(\nabla\cdot J)/\in$. Here, the over-dot(s) may indicate partial time derivative(s). $B=0$ for the SLW may imply ${}^2B=\mu\nabla\times J=0$, allowing use of the vector calculus identity, $\nabla\times\nabla\times J=0=\nabla(\nabla\cdot J)-\nabla^2 J$, giving $\nabla(\nabla\cdot J)=\nabla^2 J$, which may imply ${}^2\,(\dot{E}+J/\in)=0$. Linear electrical conductivity (σ), $J=\sigma E$, then may give ${}^2(\dot{E}+\sigma E/\in)=0$. A very rapidly decaying, transient solution may arise by setting the terms inside the parentheses to zero, giving $E=E_o\exp(-\in t/\sigma)$. Here, E_o may be the initial value of E.

[0023] A second solution may use the non-transient form, $E=E_o(r)\exp(-j\omega t)$, which yields:

$$\Box^2 E=0. \tag{12}$$

The lowest order, outgoing, spherical wave may be, $E=E_o\hat{r}\exp[j(kr-\omega t)]/r$, where $\hat{r}$ represents the unit vector in the radial direction and r represents the radial distance. As before, the electric wave's energy, $4\pi r^2(\in E^2/2)$, may be constant through a spherical boundary of arbitrary radius and $E(r\rightarrow\infty)=0$. Substitution of $J=E/\sigma$ into EQN. 12 may yield an equivalent form: ${}^2J=0$. The SLW equations for E and, J may be remarkable for several reasons. First, the vector SLW equations for E and J may be fully captured in one wave-equation for the scalar function (κ): ${}^2\kappa=0$. This form may be obtained from ${}^2J=0$ by substitution of $J=\nabla\kappa$ into the above identity, $\nabla^2 J=\nabla^2\nabla\kappa=\nabla(\nabla\cdot J)=\nabla(\nabla^2\kappa)$. Second, these forms may be like ${}^2C=0$ in EQN. 9. Third, these equations may have zero on the RHS for propagation in conductive media. This last result may arise from $B=0$ for the SLW, implying no back-electromagnetic field from $\dot{B}$ in Faraday's law that in turn may give no (circulating) eddy currents. Consequently, the SLW may not be subject to the skin effect in media with linear electrical conductivity.

[0024] C, E, and J may interact via EQN. 6, which can be rewritten using $J=\sigma E$, and identifying the electrical conductivity with the imaginary part of the complex permittivity $\in\underline{''}\sigma/\in_o\omega$:

$$\frac{C}{E_r}=\frac{\varepsilon'}{c}\left(\frac{1+j\alpha}{1+j/kr}\right) \tag{13}$$

Spherical waves may be assumed: $E=E_r\hat{r}\exp[j(kr-\omega t)]/r$ and $C=C_0\exp[j(kr-\omega t)]/r$, $\in_o$ may represent the free space permittivity; f may represent the frequency; λ, c, and ∈', may represent wavelength, speed of light, and real part of the dielectric constant in the propagation medium, respectively; $k=2\pi/\lambda$ and $\omega=2\pi f$. EQN. 13 may predict that $|C/E_r|\sim\in'/c$ in a low-conductivity medium ($a\equiv\in/`\in'<<1$), with a phase shift of a between C and $E_r$ in the far field. This ratio may be $|C/E_r|\sim\in/`\in'c$ in a good conductor ($\in/`\in'>>1$) with a phase shift $\sim\pi/2$ in the far field. EQN. 13 may be consistent with the ratio for transverse magnetic and electric fields, $|B|/|E|\sim 1/c$. C may have the same units as the magnetic field. The energy balance equation for the MCE theory may be shown as:

$$\frac{\partial}{\partial t}\left(\frac{B^2}{2\mu}+\frac{C^2}{2\mu}+\frac{\varepsilon E^2}{2}\right)+\nabla\cdot\left(\frac{E\times B}{\mu}+\frac{CE}{\mu}\right)+J\cdot E-\frac{\rho C}{\varepsilon\mu}=0. \tag{14}$$

[0025] Use of the spherical wave forms for E and C in EQN. 14 with $B=0$ and the ratio of (C/E_r) from EQN. 13 may yield an identity, 0=0; the same result may arise for plane waves, thus explicitly verifying the no-skin-effect prediction. Table 1 shows the unique and testable features of the MCE theory, which predicts transmission and reception of the SLW.

[0026] The radiated SLW power (P_{OUT}) may be obtained, e.g., as follows. The antenna may be short, linear monopole (length=L) along the z-axis with a gradient-driven current density (J) that is maximal at the feed point (z=0) and zero at the end (z=L). A and Φ may be obtained from the retarded potentials. E and C may be derived from EQNS. 3 and 5. The radiated power may come from the time-average of the radial component of CE/μ in EQN. 14.

TABLE 1

Unique and Testable SLW Properties

Item number and specific SLW property	Equation(s)
1) C is a dynamical quantity (scalar field).	EQN. 3-7
2) C is driven by gradients in J and $\partial E/\partial t$.	EQN. 6
3) SLW propagates in conductive media without the skin effect.	EQNS. 8-12
4) Interface matching involves continuity in $(\nabla C/\mu)_n$.	EQN. 10
5) A longitudinal E-field accompanies C, as an SLW.	EQNS. 11-17
6) C, E, and J exchange energy in conductive media.	EQNS. 11-17
7) $\lvert C\rvert/\lvert E\rvert \geq 1/c$ in conductive media, allowing normal instrumentation.	EQN. 13

[0027] Here, the impedance of the propagation medium may be $Z=(\mu/\in)^{1/2}$, which is 376.73Ω in free space. Terms on the order of $(kr)^{-1}$ and higher may be neglected. The resultant form for P_{OUT} may be shown as:

$$P_{OUT} = \frac{I^2}{4\pi}\sqrt{\frac{\mu}{\varepsilon}}. \tag{15}$$

[0028] EQN. 15 may be obtained from the classical, retarded potentials for a gradient-driven current. Then, a paradox arises, since C is a non-zero dynamical field, in contrast to the assumption under which the classical, retarded potentials were obtained via the Lorenz gauge (C=0). The paradox may be resolved by the MCE theory, which predicts explicitly that C is a dynamical field without a gauge assumption.

[0029] FIG. **1** illustrates a system **100** configured to transmit and/or receive scalar-longitudinal waves, in accordance with one or more implementations. Some implementations may include an Agilent Technologies E5071C network analyzer (300 kHz-20 GHz). The transmitting and receiving antennas may be identical, because the reciprocity theorem guarantees that the transmitter geometry can also act as a receiver. This simple layout is to facilitate experimental replication in any laboratory with the appropriate facilities and equipment.

[0030] FIG. **2**A illustrates a cross-sectional view of a linear monopole antenna apparatus **200** configured to transmit and/or receive scalar-longitudinal waves, in accordance with one or more implementations. The apparatus **200** may include a linear first conductor **202**, a tubular second conductor **204**, an annular skirt balun **206**, and/or other components. The first conductor **202** may extend from a core of a coaxial cable. The second conductor **204** may extend from an outer conductor of a coaxial cable. The first conductor **202** may be configured to operate as a linear monopole antenna at a first operating frequency. The second conductor **204** may be coaxially aligned with the first conductor **202** such that the first conductor **202** extends out in a first direction from within the second conductor **204**. The skirt balun **206** may be disposed at an end of the second conductor **204** from which the first conductor **202** extends. The balun **206** may have a larger diameter than the second conductor **204**. The balun **206** may extend in a second direction opposite the first direction. The balun **206** may be configured to cancel most or all return current on an outer surface of the second conductor during operation such that the first conductor transmits or receives scalar-longitudinal waves. Some implementations of apparatus **200** may include a tubular dielectric **208** coaxially disposed between the first conductor and the second conductor, the tubular dielectric extending out in the first direction from within the second conductor at least part way up the first conductor.

[0031] The configuration of apparatus **200** is for illustrative purposes and should not be viewed as limiting as other configurations are contemplated and are within the scope of the disclosure. In some implementations, the length of the balun **206** extending in the second direction may be approximately one fourth of a wavelength corresponding to the first operating frequency. During operation an electrical current wave on the balun **206** may be approximately 180 degrees out of phase relative to an electrical current wave on the outer surface of the second conductor **204** adjacent to the balun thereby cancelling most or all of the return current on the outer surface of the second conductor.

[0032] Attenuation of scalar-longitudinal waves transmitted by the first conductor **202** may be inversely proportional to the square of a distance from a center of the first conductor **202** in free space. Scalar-longitudinal waves transmitted by the first conductor propagate through a conductive medium with substantially lower attenuation relative to a classical skin-depth attenuation. The conductive medium may include a solid-copper Faraday box. Some implementations of apparatus **200** may include a solid-copper Faraday box (not shown) enclosing the first conductor **202**, the second conductor **204**, and the balun **206**. The Faraday box may be configured to block most or all transverse electromagnetic waves impinging the on Faraday box.

[0033] FIG. **2**B shows contours of constant |E| for an electrodynamic simulation of a linear monopole antenna apparatus (see, e.g., FIG. **2**A). A three-λ-diameter ground-plane disk at the feed-point may give essentially the same |E| contours in the transmitter-to-receiver direction, thus confirming the linear monopolar, counter-poise design of apparatus **200** in FIG. **2**A.

[0034] High frequency (8 GHz) experiments may allow the use of an indoor, controlled test environment. Electrodynamic simulations predict that the skirt balun **206** in FIG. **2**A attenuates the return current on the outside of the coaxial cable (**204**) by −67.5 db, to make the antenna radiate like a monopole. Grounding to a single point may eliminate current loops. Digital instrumentation may allow accurate measurement of field amplitudes and polarizations. The transmitter-to-receiver distance may be accurately measured to <1 mm. MCE theory may give radical predictions, requiring extraordinary evidence, as shown below.

[0035] According to exemplary implementations, an antenna may be a linear, coaxial center conductor with the outer coaxial conductor electrically connected to the top of the skirt balun (see, e.g., FIGS. **2**A and **2**B). The currents on the coax center conductor and on the outside of the skirt balun may be both in +Z (positive vertical) direction, forming a monopole radiation pattern. The skirt balun length (λ/4) may cause a phase shift in the current flow along the guided path from the bottom (inside surface) of the outer balun conductor (0°) to the top (inside surface) of the skirt balun (90°) and back down the outer surface of the coax outer conductor to the end of the balun (180°). This 180°-phase shift may cause cancellation of the return current flow along the outside of the outer coax conductor. Return-current attenuation may allow the monopole antenna to draw charge from the ground plane (top of the skirt balun) and also may create a close match between the antenna-balun impedance (49.76−j0.24Ω) and the source (50Ω). The measured return loss was −22 db for a single skirt balun (see, e.g., FIGS. 2A and 2B) and −42 db for a double balun (not shown). An exemplary monopole antenna may create an oscillating, gradient-driven current to create the SLW. Second, the skirt balun may dramatically reduce the return current to the source, so that essentially all of the electrical current goes into charging and discharging the antenna. The far-field contours of constant |E| may be mostly spherical on the receiver side (top of the diagram). The RG-405/U coaxial cabling may use a solid, outer conductor to minimize stray fields. The presence or absence of an outer insulating jacket may make no difference in the results of the electrodynamic simulation.

[0036] The transmitting antenna may remain in a fixed location, while the receiving antenna may be moved to a transmitter-to-receiver distance (r) in the horizontal plane. The source frequency may be 8 GHz, corresponding to a free-space wavelength, λ=3.75 cm. This choice may allow free space measurements for r over many wavelengths in a well-controlled, laboratory test. The measurements may include: the value of r; and the signal attenuation (dB) between the transmitter and receiver. Repetitions of the test provide an estimate of the statistical error. The measurement results are saved to a database.

[0037] FIGS. **3**, **4**, **5**, and **6** show experimental results for scalar-longitudinal wave attenuation under various conditions. FIG. **3** shows the attenuation versus separation distance (r) for two, facing, collinear SLW antennas inside an anechoic, Faraday chamber. Both antennas were enclosed by a cylindrical copper pipe with both ends of the copper pipe soldered to hemispherical, copper end caps. The outer conductor of the antenna's coaxial cable was soldered to a hole in one end cap. The attenuation-versus-distance plots show the SLW propagating through both Faraday cages at 8 GHz. The thickness of the two spherical caps corresponds to 2885 skin depths, which should produce a classical attenuation of −25,064 db (down by a factor of 10^{-1253}). The test measurement yielded an attenuation between −115 db at r=2 cm separation to −137 db at r=30 cm. Extrapolating the straight-line fit to a separation of 2 mm (where the Faraday-caged antennas are barely separated) gives −79 db and −86 db (down by a factor of 10^{-4}). The difference between the measured values and the classical estimate is −24,985 db, corresponding to >1249 orders of magnitude between attenuation for classical waves and the measured attenuation for the SLW.

[0038] FIG. **4** shows the attenuation versus separation distance (r) for two, facing, collinear SLW antennas. The SLW on port-**1** was surrounded by a cylindrical copper pipe with hemispherical end caps as a Faraday cage; the SLW antenna on port-**2** had no Faraday cage. Again, the attenuation-versus-distance plots show clear evidence for the SLW propagating through the transmitter's Faraday cage to/from the port-**2** SLW antenna at 8 GHz.

[0039] The same attenuation-versus-distance test was performed outside the Faraday-shielded anechoic chamber, on the ground floor of an office building. One SLW antenna had had no Faraday cage; the other SLW had a Faraday cage, as described above. FIG. **5** shows representative results, which varied widely from replicate to replicate, due to irreproducible factors (e.g., image charges and image currents in surrounding conductors, consistent with classical electrodynamics). A key result is propagation of the SLW through the **1442** skin depth of the Faraday box, which attenuates classical transverse waves to an undetectable level.

[0040] FIG. **6** shows the attenuation versus separation distance (r) for two, vertical SLW antennas at 8 GHz. The SLW antennas were bare (no Faraday cage around either the transmitter or receiver), and both were oriented vertically for comparison with the attenuation measurements for double- and single-Faraday-boxed antennas. This test was performed outside the Faraday-shielded anechoic chamber, on the ground floor of an office building, and consequently displays much variability due to scattering and reflections from nearby conductors. The results in FIGS. **3**-**5** provide clear evidence for a non-classical wave that is not constrained by the classical skin effect.

[0041] FIG. **7** illustrates a bifilar coil apparatus **700** configured to transmit and/or receive scalar-longitudinal waves, in accordance with one or more implementations. The apparatus **700** may include a bifilar coil formed in an alternating fashion of a first conductor **702** and a second conductor **704** such that a given turn of the coil that is made of the first conductor **702** is adjacent on either side to turns of the coil made of the second conductor **704**. The first conductor **702** and the second conductor **704** may be conductively coupled such that an electrical current in the coil will propagate in opposite directions in adjacent turns of the coil thereby cancelling any magnetic field so that during operation the coil transmits or receives scalar-longitudinal waves. The coil may be configured to create a gradient driven current. There may be zero or approximately zero inductance associated with the coil as a result of magnetic-field cancellation by counter-going electrical currents in adjacent turns of the coil. There may be zero or approximately zero capacitance associated with the coil as a result of adjacent turns of the coil having the same or approximately the same electrical charge density. An electrical resistance of the coil approximately matches a source impedance for maximal transmission of scalar-longitudinal waves.

[0042] The configuration of apparatus **700** is for illustrative purposes and should not be viewed as limiting as other configurations are contemplated and are within the scope of the disclosure. In some implementations, the first conductor **702** and the second conductor **704** may be conductively coupled proximate to the center of the coil. In some implementations, the first conductor **702** and the second conductor **704** may be conductivity coupled proximate to an outer edge of the coil. Although shown in FIG. **7** as being spaced apart, in exemplary implementations the first conductor **702** and the second conductor **704** may be tightly wound together to form the coil. In some implementations, the coil may be substantially planar. In some implementations, the coil is formed in a volumetric shape (e.g., a sphere or toroid). The first conductor **702** and the second conductor **704** may include conducting wires, conducting ribbons, and/or other shapes of conducting materials.

[0043] Attenuation of scalar-longitudinal waves transmitted by the coil shown in FIG. **7** may be inversely proportional to the square of a distance from the center of the coil in free space. Scalar-longitudinal waves transmitted by the coil may propagate through a conductive medium with substantially lower attenuation relative to a classical skin-depth attenuation. The conductive medium may include a solid-copper Faraday box. In some implementations, apparatus **700** may include a solid copper Faraday box (not shown) enclosing the coil. The Faraday box may be configured to block most or all transverse electromagnetic waves impinging on the Faraday box.

[0044] According to some implementations, the coil shown in FIG. **7** may represent (in cross sectional view) of a three-dimensional accumulator of charge, when each conductor is a flat, conductive sheet. The conductive sheet can be wound into a bifilar, helical pancake configuration, like a modern, cylindrical, super-capacitor. However, unlike a super-capacitor, the apparatus **700** of FIG. **7** may allow continuous current flow though the coil windings over a broad range of frequencies. As one normally skilled in the art will appreciate, other geometries are also possible for three-dimensional (volumetric), wire-wound, charge accumulation, such as a sphere and toroid. The apparatus **700** of FIG. **7** may be scaled to higher (or lower) frequencies by limiting the total length of the conductor to $L \ll \lambda$, or more specifically $L \leq \lambda/100$. Table 2 shows examples of the conductor lengths (L) versus frequency for free-space propagation.

[0045] Note that macro-scale antennas ($L \geq 3$ cm) may occur for frequencies less than or equal to 100 MHz, meaning that a person normally skilled in the art may build such an antenna with the unaided eye. Micro-scale antennas (3 $\mu m \leq L \leq cm$) may occur for frequencies in the range of 100 MHz to 1 THz, meaning that a person normally skilled in the art may need a microscope to build such an antenna. Nano-scale antennas ($L \leq \mu m$) occur for frequencies of >1 THz, meaning that a person normally skilled in the art may need an electron-microscope (or equivalent) to build such an antenna. High frequencies (>1 THz) may correspond, for example, to atomic transitions from a 1S to a 2S orbital, which are forbidden by classical quantum mechanics on the basis of classical electrodynamics, as discussed above. As one normally skilled in the art can appreciate, analogous molecular, nuclear, and sub-atomic transitions, also exist. Note further that essentially all transverse-wave transmission or reception may be eliminated by enclosing a SLW antenna (e.g., apparatus **200** or apparatus **700**) inside a Faraday cage or box (e.g., a copper or aluminum casing not unlike that for a modern super-capacitor).

TABLE 2

Wavelength and L versus frequency

Frequency (f in Hz)	Wavelength ($\lambda = 3 \times 10^8/f$ in meters)	$L \leq 0.01\lambda$ (in meters)
1	3×10^8 (300,000 km)	3,000 km
10^2	3×10^6 (3,000 km)	30 km
10^4	3×10^4 (30 km)	300 m
10^6	3×10^2 (300 m)	3 m
10^8	3×1 (3 m)	3 cm
10^{10}	3×10^{-2} (3 cm)	0.3 mm
10^{12}	3×10^{-4} (0.3 mm)	3 μm
10^{14}	3×10^{-6} (3 μm)	30 nm

[0046] More complete electrodynamics (MCE) may be important for several reasons. First, the MCE theory may involve a radical revision of Maxwell's equations with one new term $\partial c/\partial t$) in Gauss' law and one new term ($-\nabla C$) in Ampere's law. These new terms may arise from ($—C^2/2\mu$) in the Stueckelberg Lagrangian. Second, the MCE theory may give relativistic covariance; preservation of the fields (B and E) in terms of the classical potentials (A and Φ); and the classical wave equations for A and Φ without a gauge condition. Third, the MCE theory may predict new force terms in the MCE momentum balance equation that might explain "dark matter" as a placeholder for unexplained cosmological attractive forces. Fourth, new terms in the MCE energy balance (EQN. 15) may explain "dark energy" as a placeholder for unexplained repulsive cosmological forces. Fifth, the MCE theory (along with classical theory) may predict that a gradient-driven current produces a scalar-longitudinal photon, consisting of both scalar (C) and longitudinal E-field components. This last prediction may make the SLW wave uniquely detectable via a gradient-driven current density in the novel antenna, distinct from classical transverse photons that require a circulating current ($\nabla \times J \neq 0$). The existence of dark matter and dark energy may signify that our physics understanding is incomplete, likely requiring a new idea as profound as general relativity. Scalar-longitudinal waves/photons may be that new idea, as validated by our experimental results.

[0047] In some implementations, system **100** (see FIG. **1**) may be configured for providing a computational simulator based on scalar-longitudinal waves. The system **100** may include one or more hardware processors (not depicted) configured by machine-readable instructions. The machine-readable instructions may include a simulation component, a classical transverse electromagnetic wave component, a scalar-longitudinal wave component, an evaluation component, an optimization component, and/or other components. The simulation component may be configured to provide a computerized physical simulation environment in which electromagnetic simulations of an antenna or device are performed. In some implementations, the computerized physical simulation environment may include a reflector added to the antenna or device to form a directed beam for transmission and/or reception of scalar-longitudinal waves. The classical transverse electromagnetic wave component may be configured to provide simulated classical transverse electromagnetic waves that are received or transmitted in the electromagnetic simulation of the antenna or device. The scalar-longitudinal wave component may be configured to provide simulated scalar-longitudinal waves that are received or transmitted in the electromagnetic simulations of the antenna or device. The evaluation component may be configured to evaluate characteristics of the antenna or device based on information associated with simulated classical transverse electromagnetic waves and/or simulated scalar-longitudinal waves. The optimization component may be configured to optimize one or more characteristics of the antenna or device based on the evaluation of the characteristics.

[0048] A given processor may be configured to provide information processing capabilities in system **100**. As such, the given processor may include one or more of a digital processor, an analog processor, a digital circuit designed to process information, an analog circuit designed to process information, a state machine, and/or other mechanisms for electronically processing information. In some implementations, system **100** may include a plurality of processing units. These processing units may be physically located within the same device, or the given processor may represent processing functionality of a plurality of devices operating in coordination. The given processor may be configured to execute machine-readable instructions include the simulation component, the classical transverse electromagnetic wave component, the scalar-longitudinal wave component, the evaluation component, the optimization component, and/or other components of machine-readable instructions. The given processor may execute machine-readable instructions by software; hardware; firmware; some combination of software, hardware, and/or firmware; and/or other mechanisms for configuring processing capabilities on the given processor.

[0049] It should be appreciated that the description of the functionality provided by the different machine-readable instruction components described herein is for illustrative purposes, and is not intended to be limiting, as any of the machine-readable instruction components may provide more or less functionality than is described. For example, one or more of the machine-readable instruction components may be eliminated, and some or all of its functionality may be provided by other ones of the machine-readable instruction components. As another example, the given processor may be configured to execute one or more additional machine-readable instruction components that may perform some or all of the functionality attributed herein to one of the machine-readable instruction components.

[0050] The system **100** may include electronic storage (not depicted). The electronic storage may store machine-readable instructions and/or other information. Electronic storage may comprise non-transitory storage media that electronically stores information. The electronic storage media of electronic storage may include one or both of system storage that is provided integrally (i.e., substantially non-removable) with a physical computing platform and/or removable storage that is removably connectable to a physical computing platform via, for example, a port (e.g., a USB port, a firewire port, etc.) or a drive (e.g., a disk drive, etc.). Electronic storage may include one or more of optically readable storage media (e.g., optical disks, etc.), magnetically readable storage media (e.g., magnetic tape, magnetic hard drive, floppy drive, etc.), electrical charge-based storage media (e.g., EEPROM, RAM, etc.), solid-state storage media (e.g., flash drive, etc.), and/or other electronically readable storage media. Electronic storage may include one or more virtual storage resources (e.g., cloud storage, a virtual private network, and/or other virtual storage resources). Electronic storage may store software algorithms, information determined by processors, and/or other information that enables system **100** to function as described herein.

[0051] FIG. **8** illustrates a method **800** for utilizing scalar-longitudinal waves, in accordance with one or more implementations. The operations of method **800** presented below are intended to be illustrative. In some implementations, method **800** may be accomplished with one or more additional operations not described, and/or without one or more of the operations discussed. Additionally, the order in which the operations of method **800** are illustrated in FIG. **8** and described below is not intended to be limiting.

[0052] In some implementations, one or more operations of method **800** may be implemented in one or more processing devices (e.g., a digital processor, an analog processor, a digital circuit designed to process information, an analog circuit designed to process information, a state machine, and/or other mechanisms for electronically processing information). The one or more processing devices may include one or more devices executing some or all of the operations of method **800** in response to instructions stored electronically on an electronic storage medium. The one or more processing devices may include one or more devices configured through hardware, firmware, and/or software to be specifically designed for execution of one or more of the operations of method **800**.

[0053] At an operation **802**, a first apparatus or a second apparatus configured to transmit and/or receive scalar-longitudinal waves may be obtained. The first apparatus (see, e.g., apparatus **200** of FIG. **2**A) may include a linear first conductor, a tubular second conductor, and an annular skirt balun. The first conductor may configured to operate as a linear monopole antenna at a first operating frequency. The second conductor may be coaxially aligned with the first conductor such that the first conductor extends out in a first direction from within the second conductor. The balun may disposed at an end of the second conductor from which the first conductor extends. The balun may have a larger diameter than the second conductor. The balun may extend in a second direction opposite the first direction. The balun may be configured to cancel most or all return current on an outer surface of the second conductor during operation such that the first conductor transmits or receives scalar-longitudinal waves. The second apparatus (see, e.g., apparatus **700** of FIG. **7**) may include a bifilar coil formed in an alternating fashion of a first conductor and a second conductor such that a given turn of the coil that is made of the first conductor is adjacent on either side to turns of the coil made of the second conductor. The first conductor and the second conductor may be conductively coupled such that an electrical current in the coil will propagate in opposite directions in adjacent turns of the coil thereby cancelling any magnetic field so that during operation the coil transmits or receives scalar-longitudinal waves.

[0054] At an operation **804**, scalar-longitudinal waves may be transmitted and/or received using the first apparatus or the second apparatus in order to achieve a technical result. Exemplary technical results are described herein but should not be viewed as limiting as other technical results involving scalar-longitudinal waves are contemplated and are within the scope of the disclosure.

[0055] In some implementations, the technical result of method **800** may include communicating and/or sensing information underwater.

[0056] In some implementations, the technical result of method **800** may include communicating and/or sensing information underground.

[0057] In some implementations, the technical result of method **800** may include enhancing (or de-enhancing) a decay rate of a radioactive material. (De)enhancement of radio-active decay rates may be achieved because the Stueckelberg Lagrangian density in EQN. 2 corresponds to new terms (involving C) in the electrodynamic Hamiltonian:

$$\mathcal{H}_{EM} = \left(\frac{\varepsilon E^2}{2} + \frac{B^2}{2\mu}\right) + (\rho - \varepsilon\nabla\cdot E)\Phi - J\cdot A + \frac{C^2}{2\mu} + \frac{C\nabla\cdot A}{\mu}. \tag{16}$$

[0058] This MCE Hamiltonian may modify the charged-particle interactions via the SLW (e.g., orbital electrons and nuclear protons in electron-capture decay, and bound electrons and protons in beta decay of neutrons). These new terms may modulate the nuclear barrier potential, causing decay-rate variations in proportion to the SLW power. Indeed, time-variable radioactive-decay rates have been reported (typically ±0.3%) in ^{3}H, ^{22}Na, ^{36}Cl, ^{44}Ti, ^{54}Mn, ^{60}Co, ^{85}Kr, ^{90}Sr, ^{108m}Ag, ^{133}Ba, ^{137}Cs, ^{152}Eu, ^{154}Eu, ^{222}Rn, ^{236}Ra, and ^{239}Pu. Typical periods in the decay rate may include: one day, 12.08/year (solar rotation rate), one year, and ~12 years (sun-spot cycle). Classical low-energy nuclear theory is a collection of ad hoc models whose predictions cannot explain these observations. The sun is a sphere of charged particles (plasma) that is well-known to oscillate radially (breathing mode) and in multi-pole modes. The oscillation amplitude of ions is much different from electrons, giving rise to a net radial current density that creates the SLW, which in turn may modulate the radioactive decay rate according to the new terms in EQN. 16. One specific application is use of SLW power to enhance decay of radioactive fission-waste products from a nuclear reactor, and isotopes of proliferation concern.

[0059] In some implementations, the technical result of method **800** may include enhancing a fusion rate reaction to produce heat and/or electrical power. Classical methods and apparatus for controlled fusion typically involve maintaining a high enough fuel density (e.g., deuterium and tritium) at a sufficient temperature (e.g., 100 million degrees K) for a long enough time (e.g., many seconds). EQN. 16 predicts modulation of the nuclear barrier potential, allowing fusion reactions at room temperature. Nuclear reactions could be enhanced directly (e.g., $^{2}D_{1}+^{2}D_{1}+SLW\rightarrow{}_{2}He^{4}$+energy) via cold fusion of D_2O without intermediate steps (and correspondingly complex infrastructure).

[0060] Existing electromagnetic/multi-physics simulators use the classical version of Maxwell's equations. A specific application of the MCE theory (EQNS. 1-17) may be a more complete simulator for detailed design of antennas and other electrical/electronic devices that use SLW technology.

[0061] A specific application of exemplary implementations may include a focusing SLW antenna. The above tests showed scattering and reflections of the SLW. This observation is consistent with classical electrodynamics, which predicts scattering of electric fields from conductors due to image-charges and image-currents. This observation implies that MCE modifications of electrodynamic simulators may be used for development of SLW antenna(s) for focused transmission (and reception). Such antennas may reduce power, weight, and cost in practical applications.

[0062] In some implementations, the technical result of method **800** may include detecting scalar-longitudinal waves emitted from a chemical-bond-breaking process. The chemical-bond-breaking may be caused by seismic activity associated with an earthquake, a failure of a manmade structure, and/or other processes. Earthquake prediction has been sought for decades, and typically may depend on quantitative measurement of underground motion and/or slip-stick stress at tectonic plate boundaries. The seismic activity causes grinding of rock to powder, which may generate high voltages by molecular bond breaking. The voltage corresponds to an electric field, which drives a current gradient, as the SLW driver. These signals occur well in advance of the slip events, and may allow prediction of the time and location of events with suitable SLW detection/imaging. "Earthquake clouds" and electromagnetic precursors of seismic events have been reported. Geophysicists recently discovered low-frequency toroidal oscillations with a period of 2 to 5 minutes; these low-frequency waves may cause excited animal behavior prior to an earthquake. The peeling of tape is another example of bond breaking. The specific applications may include passive detection/prediction of structural failures of all kinds, such as bridges, buildings, critical equipment, and seismic activity.

[0063] In some implementations, the technical result of method **800** may include passive imaging of a living organism based on gradient-driven currents across cellular membranes. In some implementations, the technical result of method **800** may include transmission of scalar-longitudinal waves into a living organism to enhance health and/or treat a disease via gradient-driven currents across cellular membranes. Generally speaking, living processes are driven by charged ion transport across the cell membranes. The ion transport is driven, in turn by concentration gradients in the intra- and extra-cellular media. This gradient-driven transport of charged ions creates a gradient-driven electrical current, which is the basis for SLW creation. Consequently, all living organisms create SLWs, which can be imaged by a phased-array of receivers. This new imaging modality may allow passive imaging of living organisms (including people) for research and disease diagnosis and treatment. As one who is normally skilled in the art can appreciate, standard techniques may be used to convert variation in line-of-sight SLW amplitude into an image, not unlike a CT scan. One application is passive imaging of live animals and humans (e.g, brain, heart, lungs). Human electrophysiology has a typical frequency range of 0.5-1000 Hz, implying SLW might be efficacious in this frequency range.

[0064] In some implementations, the technical result of method **800** may include imaging an object or a void. The method **800** may include providing a phased-array of scalar-longitudinal waves for the imaging. A phased array of SLW receivers may passively image objects of interest or voids (e.g., underground tunnels, facilities, and pipelines) using the background solar SLW flux for illumination. One specific example is imaging of buildings' interiors, which would be bathed in solar SLW. Detection of underground nuclear tests is part of the nuclear test-ban treaty verification. A nuclear explosion ejects concentric, radially-expanding shells of fast electrons (outer shell) and slower-moving positive-ions (inner shell). These charged shells form a spherical capacitor with a radial E-field (gradient-driven current), thus creating a SLW. The sun is a hot ball of ions and electrons (in the form of a plasma) that oscillates radially, thus creating the SLW that then image solar storms for prediction/mitigation of adverse events (e.g., power outages). More generally, the SLW may be used to create three-dimension images (e.g., via binocular image) that sees through fog, clouds, dust, rain, and building fires during the day or night. Another application is astronomical imaging in the across the entire frequency spectrum.

[0065] In some implementations, the technical result of method **800** may include transmission and/or reception of scalar-longitudinal waves for radar imaging of an object and/or a void. The sun emits low-loss SLW, which may be used to form passive images of underwater objects via a phased array that looks upward from the ocean floor. An active, phased-array (or synthetic-aperture) SLW-RADAR from ships, aircraft, and satellites may detect and identify underwater objects, underwater vehicles, underground tunnels, pipelines, underground facilities, stealth aircraft under adverse weather conditions, and/or other objects. The SLW transmission may not need to be limited to one frequency. The SLW transmission may be hyper-spectral (e.g., MHz to THz). A space-based implementation may use satellite arrays to transmit the SLW signal at many frequencies, receive the reflections in synthetic-aperture mode, and process the results on-board for real-time imaging. SLW RADAR can be used for detection of improvised explosive devices (IEDs) on the battlefield.

[0066] SLW propagation through conductive media may include ionized plasma around a space vehicle that re-enters the earth's atmosphere. More specifically, SLW-RADAR may be used to characterize space vehicles that re-enter the earth's atmosphere, while surrounded by a hot sheath of plasma. Classical (transverse) electromagnetic waves cannot penetrate the plasma sheath.

[0067] In some implementations, the technical result of method **800** may include reception of solar-generated scalar-longitudinal waves to produce electrical power. The MCE theory may predict that a charged sphere, oscillating in a ballooning (monopolar) mode (expanding and contracting radially) will radiate the SLW. The MCE theory may predict that higher-order (multi-pole) oscillations will create the SLW. The sun is a very hot ball of charged particles (electrons and ions in the form of a plasma) that undergoes such oscillations. Consequently, SLW power reaches the earth, just as sunlight does. A specific application may include conversion of solar SLW power into electric power for (re)charging batteries and/or powering electrical device, such as electric vehicles could be replaced by power converters for the SLW, which is not limited by the skin effect. Harvesting of this solar power may be scalable via advanced photovoltaics that convert the variable-frequency SLW to direct current, then invert the DC power to stable 60 Hz alternating current (for example). An extension of this approach is wireless SLW power transmission that could then be converted to usable electrical power.

[0068] A specific application of exemplary implementations may involve electrical power generation from solar SLWs, on the basis of new terms in the MCE momentum balance equation:

$$\varepsilon\mu\frac{\partial}{\partial t}\left(\frac{E\times B}{\mu}-\frac{CE}{\mu}\right)+J\times B+\rho E-CJ+\nabla\times BC=\nabla\cdot\overline{\overline{T}}+\frac{\nabla C^2}{\mu}. \tag{17}$$

[0069] $\overline{\mathrm{T}}$ may represent the Maxwell stress tensor. More specifically, electrical power may be generated by charging a flat-plate capacitor to give a large, directed E-field. SLW emission from the sun may generate force variations across the capacitor plates via the term, (EC/μ) in EQN. 17, corresponding to a voltage to drive a power-producing current. This power may be proportional to E (and therefore the capacitor voltage) and/or may be proportional to the amplitude of the solar SLW emissions (C). The variable-frequency power may be rectified, and subsequently converted to alternating current via an alternator. More specifically, the sun is an oscillating sphere (monopolar antenna) of charged matter (plasma) that will produce SLW under the MCE theory, because the oscillation distance for ions is different than for electrons for various plasma waves. Earth is a spherical conductor (monopole antenna) that would receive the solar SLW emissions, and would re-radiate them (along with other planets, comets, asteroids). So, near-stellar regions are bathed in SLW emissions, day and night, allowing SLW power-conversion to operate day and night, rain or shine.

[0070] The term (EC/μ) in EQN. 17 may be an additional term that forms a generalized Poynting vector, corresponding to the magnitude and direction of power density transmission by the SLW. This new term may imply that SLW power can be transmitted wirelessly over large distances in a directed fashion, for example to power satellites or aircraft from the ground, electrical power transmission, and advanced forms of (directed) beams.

[0071] MCE theory may predict new terms in electromagnetic momentum balance and power balance, EQNS. 14 and 17. The term CE/μ may correspond to an increase (or decrease) in longitudinal electrodynamic momentum in EQN. 17 along the direction of motion, with a concomitant decrease (increase) in electrical power per EQN. 14. This sign change may be important because longitudinal electrodynamic power loss (or gain) may drive a corresponding kinetic energy gain (loss) in the physically massive object that is emitting these waves. Consequently, the MCE theory may predict a propulsion mechanism without the use of propellant mass, which is a severe constraint on all transportation systems. This mechanism may depend on adequate energy to create the SLW. The potential applications include all transportation modes on land, sea, air, and space for propellant-less propulsion.

[0072] MCE theory may predict a new term, CJ, in EQN. 17. Emission of the SLW from a physically massive object may have at least two components: the longitudinal electric field (E) and the scalar field (C). The electric field may induce an electrical current density (J) in any (distant) conductive object in its path, according to J=σE. The concomitant presence of the scalar field (C) may interact with this current to product a force (CJ) on the distant object. By use of a phased array of SLW emitters, the relative of phase of E (and thus J) may be shifted relative to the phase of C. The resultant force, CJ, may be adjusted to have a positive (repulsive) force or a negative (attractive) force, commonly called a "tractor beam." Potential applications may range from the nano- to macro-scales on any conductive object (e.g., sub-atomic particles, molecules, living cells, people, animals, vehicles, comets, asteroids, planets, stars, galaxies).

[0073] A mathematical theorem states that nonlinear quantum systems can be used to solve the hardest, non-deterministic, polynomial-time (NP-hard) problems in deterministic polynomial time. The Hamiltonian in EQN. 16 may be inherently nonlinear, and therefore may provide a path to construct such a computer, which would then enable the solution of grand-challenge class problems in a very finite time (minutes to hours, instead of years or more).

[0074] High-temperature superconductivity was (HTS) discovered in 1986. The highest critical temperature for HTS is >150K. However, the physical mechanism for HTS is one of the major unsolved problems in theoretical condensed matter physics, in part because the materials are very complex, multi-layered crystals. Moreover, this theoretic effort uses classical electrodynamic interactions in condensed matter, while the MCE theory may provide an explanation on the basis of gradient-driven currents between (or among) the crystal layers. The well-known London model of superconductivity is not gauge invariant. Specifically, the London model works only for the Coulomb gauge, ∇·A=0. Recent experiments show evidence for a Higgs-like mode in two-dimensional superconductors, namely excess absorption of THz radiation. However, the Higgs-like mode may not be an actual particle, but a collective quantum mode. The new Hamiltonian of EQN. 16 may include the SLW due to gradient-driven currents among the crystalline layers, as an explanation of HTS. Many commercial and military applications exist, including sensitive magnetometers based on SQUIDS, fast digit circuits, rapid single-flux quantum technology, maglev trains, MRI imaging, magnetic confinement fusion, magnetics in particle accelerators, microware filters, high sensitivity particle detectors, nanowire single-photon detector, railguns and coilguns, electric motors and generators, fault current limiters, and electrical power storage and transmission.

[0075] Recent research has investigated a connection between the SLW, high-temperature superconductors, and gravity. Initial work placed a small, non-conducting, non-magnetic mass (5.48 g) over a levitating, rotating, superconducting disk at <77 K; the mass weighed 0.05-0.3% less, depending on the disk's rotation speed. Subsequent work used a rotating, toroidal, current-carrying, superconducting disk at <70 K; masses placed over the disk weighed 0.3-0.5% less initially and 1.9-2.1% less as the rotation speed was reduced. Further work used a superconducting electrode at 40 K to generate discharges (10^4 amps at >1 MV with a trapped magnetic field of ≤1 T). The resultant focused beam propagated without noticeable attenuation through different materials and exerted a short repulsive force on small movable objects in proportion to their mass, independent of the sample's composition. More recent work used a superconducting cathode at 50-70 K and a copper anode to create discharges (10^4 amps at ≤2 MV) in low pressure gases. The discharge changed from a spark to a flat glowing plasma that originated from the superconducting cathode at >500 kV. A collimated, non-electromagnetic "radiation pulse" propagated from the cathode, toward and beyond the anode, apparently without attenuation. Recent work used this device to measure the scattering of laser light whose attenuation lasted 34-48 ns and increased with discharge voltage up to 7% at 2 MV. The radiation-pulse propagation speed was measured by two piezoelectric crystals over 1211 m with a time delay of 63±1 ns, corresponding to 64 times the speed of light. Different targets (ballistic pendulums, photons, piezoelectric crystals) are affected differently by the radiation pulse, possibly reacting to beam components at different speeds. This approach may be the basis for a SLW laser.

[0076] Spectrum allocation of the SLW may be important for greater data rates over less bandwidth. Examples for higher data rates may be applied to the SLW, such as frequency hopping/re-use, spread-spectrum technology, polarization, code division, and/or other examples. The SLW may be independent of TEM spectrum, effectively doubling the present transmission/reception capacity.

[0077] Although the present technology has been described in detail for the purpose of illustration based on what is currently considered to be the most practical and preferred implementations, it is to be understood that such detail is solely for that purpose and that the technology is not limited to the disclosed implementations, but, on the contrary, is intended to cover modifications and equivalent arrangements that are within the spirit and scope of the appended claims. For example, it is to be understood that the present technology contemplates that, to the extent possible, one or more features of any implementation can be combined with one or more features of any other implementation.

1. A system for providing a computational simulator based on scalar-longitudinal waves, the system comprising:
one or more hardware processors configured by machine-readable instruction to:
provide a computerized physical simulation environment in which electromagnetic simulations of an antenna or device are performed;
provide simulated classical transverse electromagnetic waves that are received or transmitted in the electromagnetic simulation of the antenna or device;
provide simulated scalar-longitudinal waves that are received or transmitted in the electromagnetic simulations of the antenna or device;
evaluate characteristics of the antenna or device based on information associated with simulated classical transverse electromagnetic waves and/or simulated scalar-longitudinal waves; and
optimize one or more characteristics of the antenna or device based on the evaluation of the characteristics.

2. The system of claim **1**, wherein the computerized physical simulation environment includes a reflector added to the antenna or device to form a directed beam for transmission and/or reception of scalar-longitudinal waves.

3. A process for allocating spectral frequencies of scalar-longitudinal waves, the process comprising:
allocating bands of spectral frequencies of scalar-longitudinal waves to one or more entities according to rules comprising:
no entity shall transmit at frequencies reserved for radio astronomy to avoid interference at radio telescopes;
a given entity is allowed to transmit over open-spectrum bands so long as the given entity adheres to restrictions on transmission power and any other relevant constraints;
only licensed entities of a given spectral frequency band are allowed to transmit so long as individual ones of the licensed entities of the given spectral frequency transmit in non-overlapping geographic regions; and
only licensed entities of spectral frequency bands corresponding to amateur radio frequencies are allowed to transmit so long as the licensed entities of spectral frequency bands corresponding to amateur radio frequencies use a listen-before-talk contention-based protocol.

* * * * *

Dec. 26, 1961 E. M. TURNER ETAL 3,015,101

SCIMITAR ANTENNA

Filed Oct. 31, 1958 14 Sheets-Sheet 1

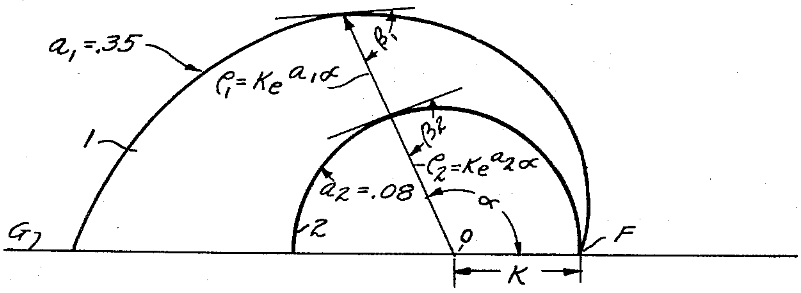

$a_1=.35$
$\rho_1=Ke^{a_1\alpha}$
β_1
β_2
$\rho_2=Ke^{a_2\alpha}$
$a_2=.08$
α
1
2
G
O
F
K

FIG. 1

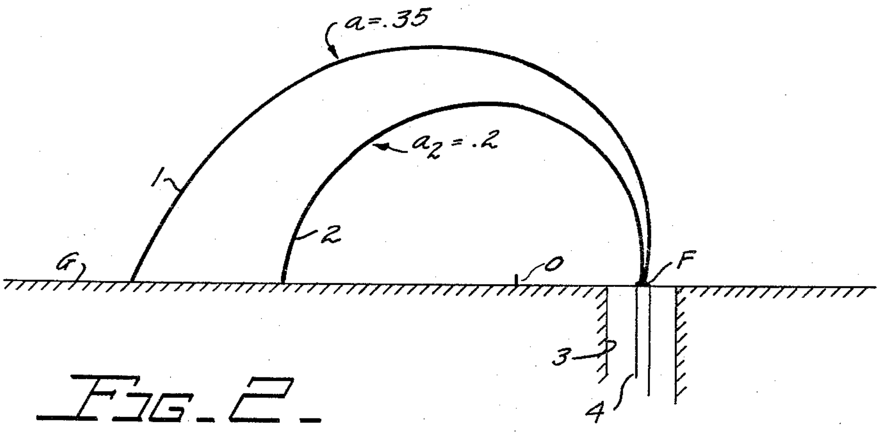

$a=.35$
$a_2=.2$
1
2
G
O
F
3
4

FIG. 2

INVENTORS.
W. P. TURNER
E. M. TURNER
BY Wade Koontz
ATTORNEY
James S. Shannon
AGENT

Dec. 26, 1961

E. M. TURNER ETAL

3,015,101

SCIMITAR ANTENNA

Filed Oct. 31, 1958

14 Sheets-Sheet 2

FIG. 3.

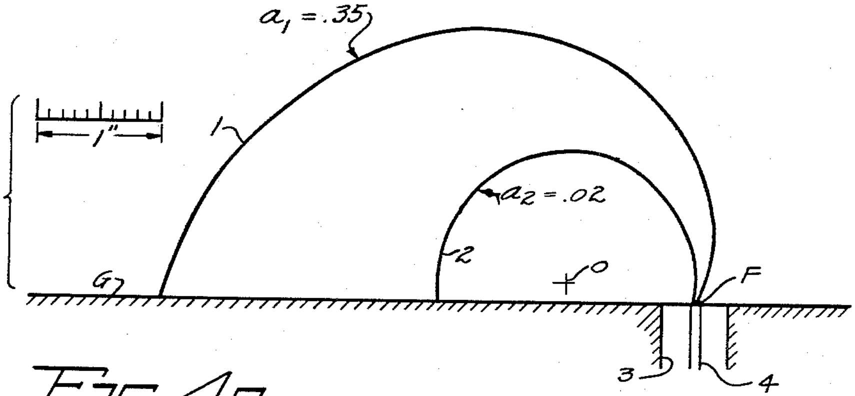

$a_1 = .35$
1
1″
$a_2 = .02$
2
O
G
F
3
4

FIG. 4a.

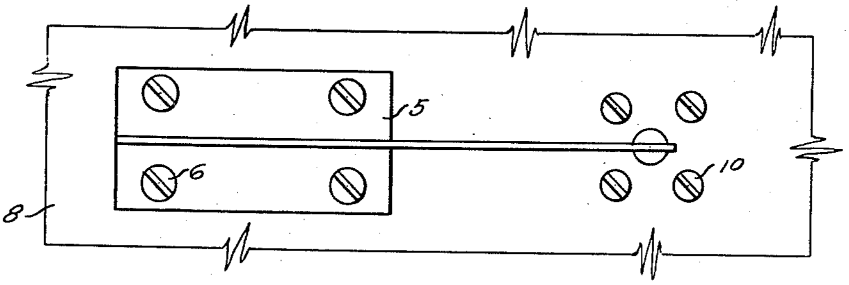

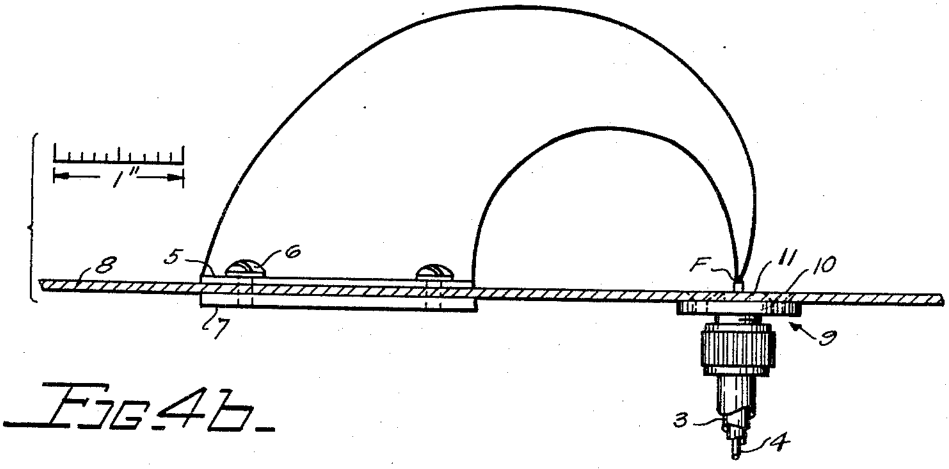

1″
8
5
6
F
11
10
7
9
3
4

FIG. 4b.

INVENTORS.
W. P. TURNER
E. M. TURNER
BY Wade Koontz
ATTORNEY
James S. Shannon
AGENT

Dec. 26, 1961 E. M. TURNER ETAL 3,015,101

SCIMITAR ANTENNA

Filed Oct. 31, 1958 14 Sheets-Sheet 3

FIG. 3a

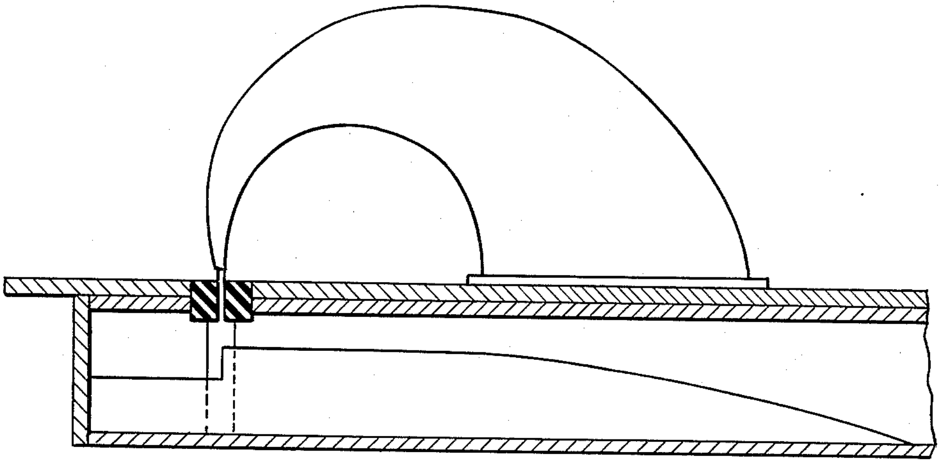

FIG. 3b

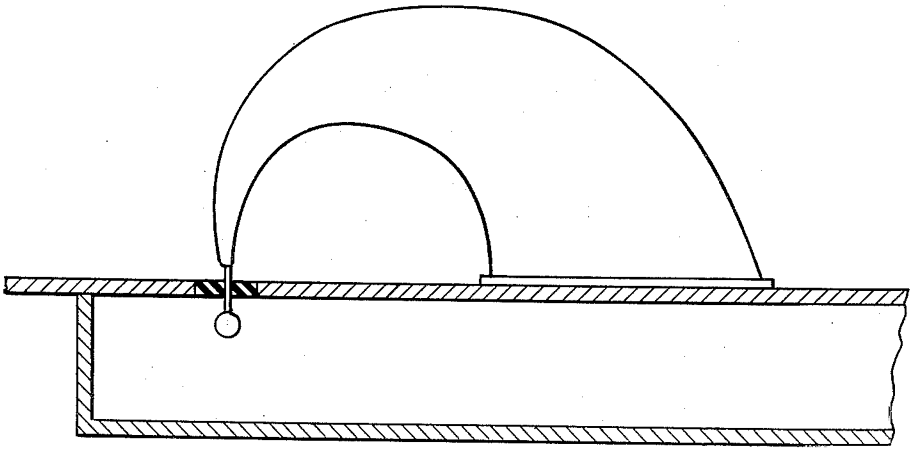

INVENTORS
W. P. TURNER
E. M. TURNER
BY Wade Koontz
ATTORNEY

Dec. 26, 1961 E. M. TURNER ETAL **3,015,101**

SCIMITAR ANTENNA

Filed Oct. 31, 1958

14 Sheets-Sheet 4

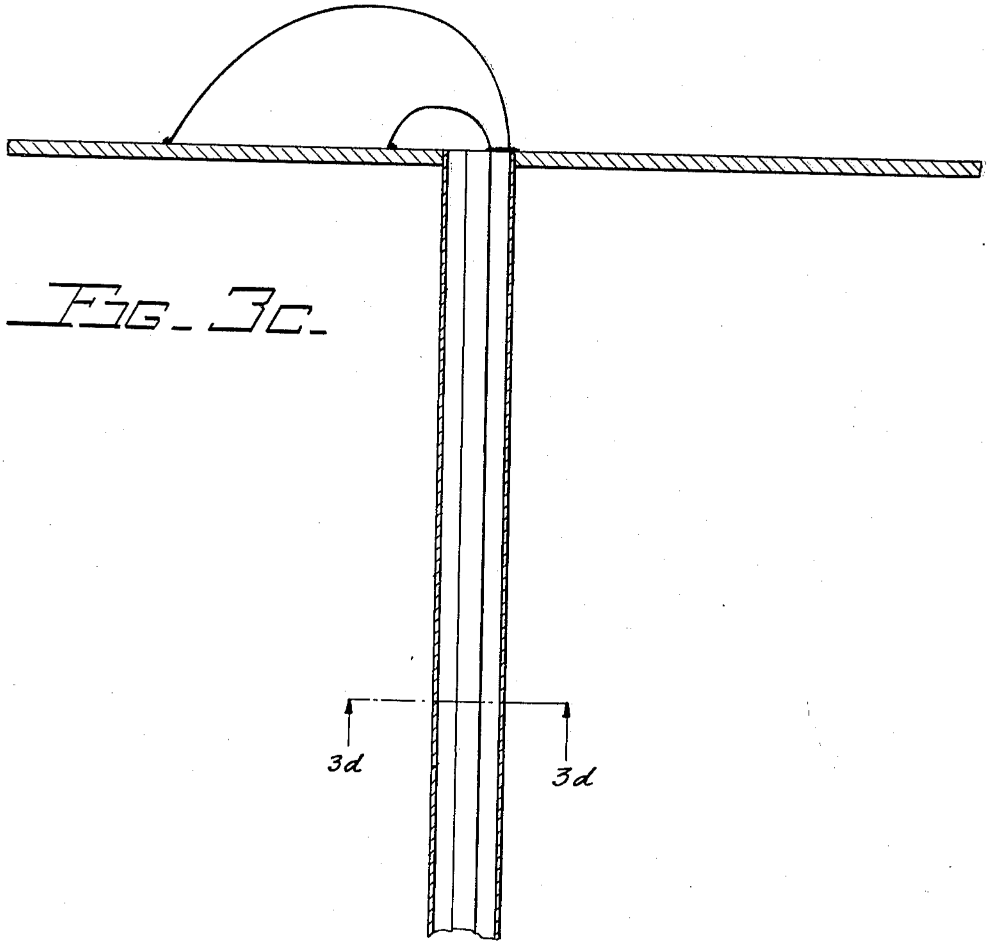

FIG. 3c

FIG. 3d

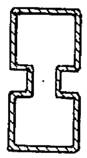

INVENTORS.
W. P. TURNER
E. M. TURNER
BY
Wade Koontz
ATTORNEY and
James [illegible] Shannon
AGENT

FIG. 5

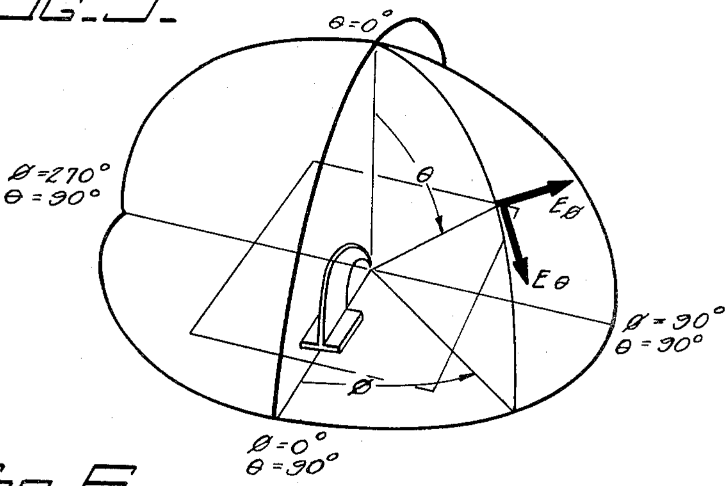
θ=0°
φ=270°
θ=90°
θ
Eφ
Eθ
φ=90°
θ=90°
φ
φ=0°
θ=90°

FIG. 6

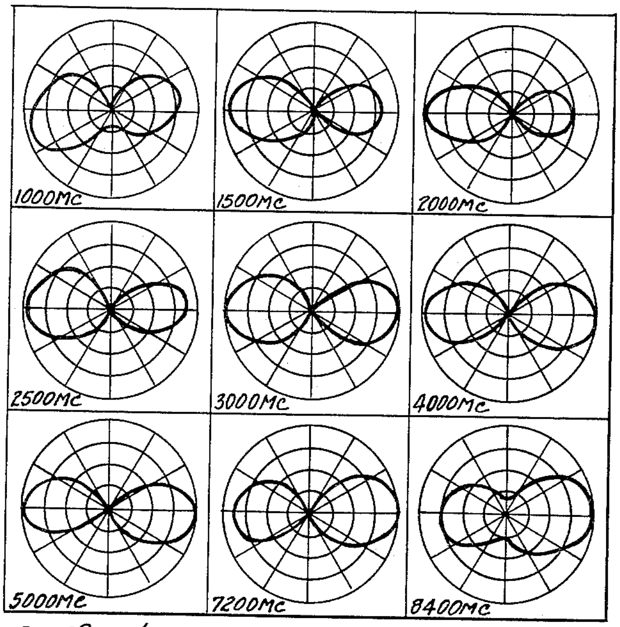

θ=0° φ CUT
POLARIZATION Eθ

INVENTORS
W. P. TURNER
E. M. TURNER
BY Wade Koontz
ATTORNEY
James S. Shannon
AGENT

Dec. 26, 1961

E. M. TURNER ETAL

3,015,101

SCIMITAR ANTENNA

Filed Oct. 31, 1958

14 Sheets-Sheet 6

FIG. 7

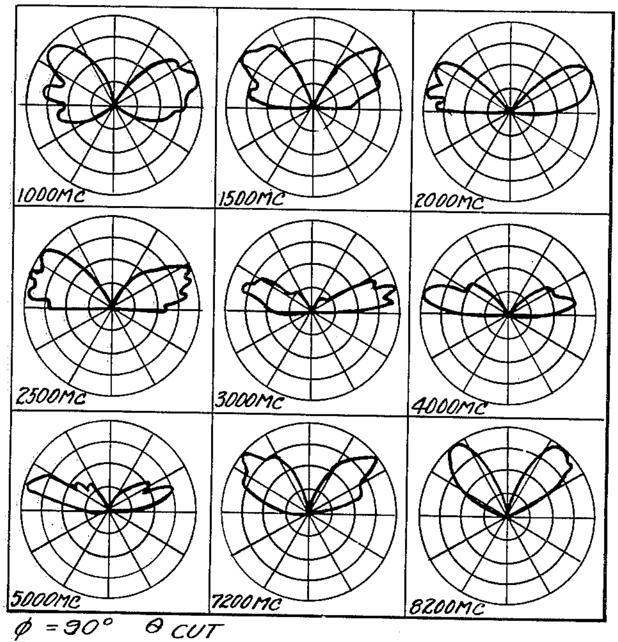

$\phi = 90°$ θ CUT
POLARIZATION E_θ

FIG. 8

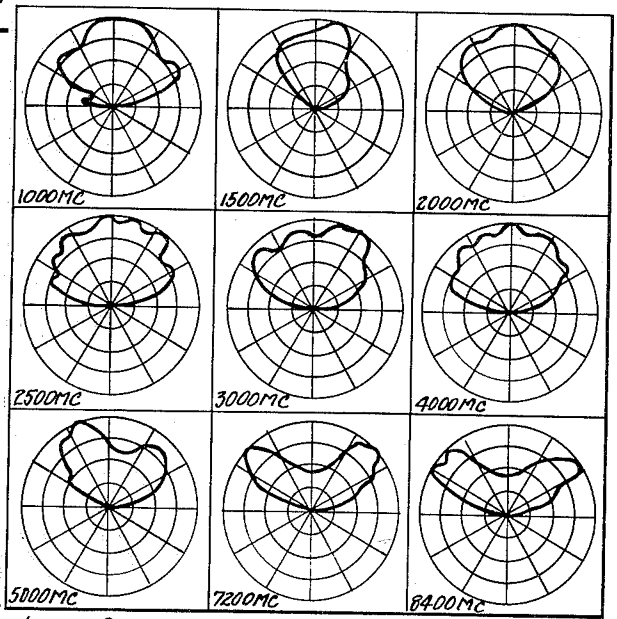

$\phi = 90°$ θ CUT
POLARIZATION E_ϕ

INVENTORS.
W.P. TURNER
E.M. TURNER
BY Wade Koontz
ATTORNEY and
James S. Shannon
AGENT

Dec. 26, 1961

E. M. TURNER ETAL

3,015,101

SCIMITAR ANTENNA

Filed Oct. 31, 1958

14 Sheets-Sheet 7

FIG. 9

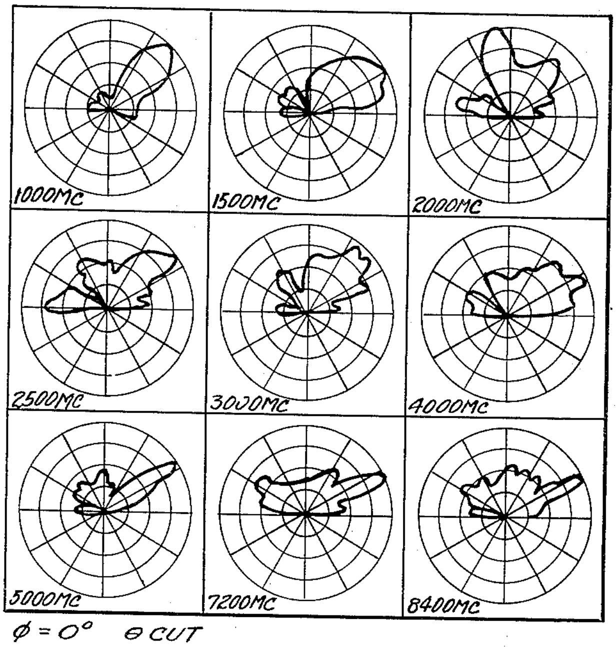

$\phi = 0°$ θ CUT
POLARIZATION E_θ

INVENTORS
W. P. TURNER
E. M. TURNER

BY Wade Koontz
ATTORNEY
and
James S. Shannon
AGENT

FIG. 10

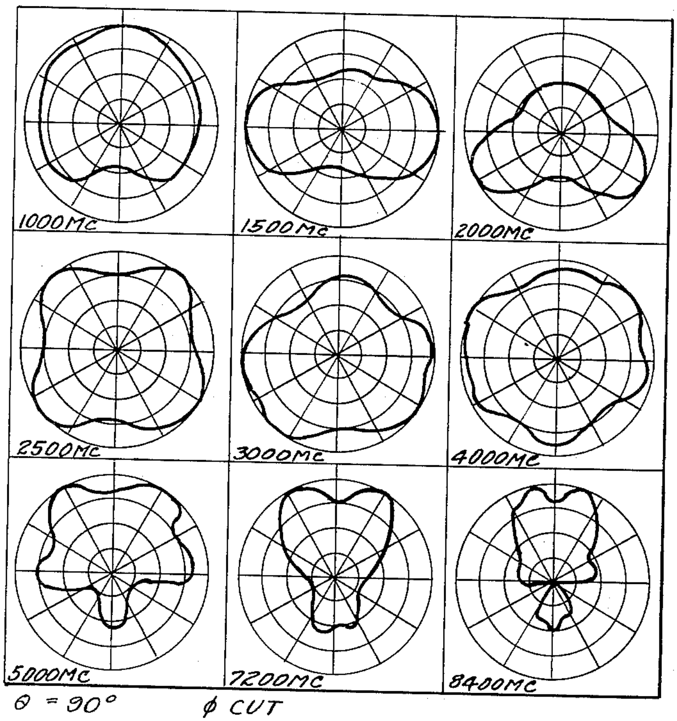

θ = 90° φ CUT
POLARIZATION Eθ

FIG. 11

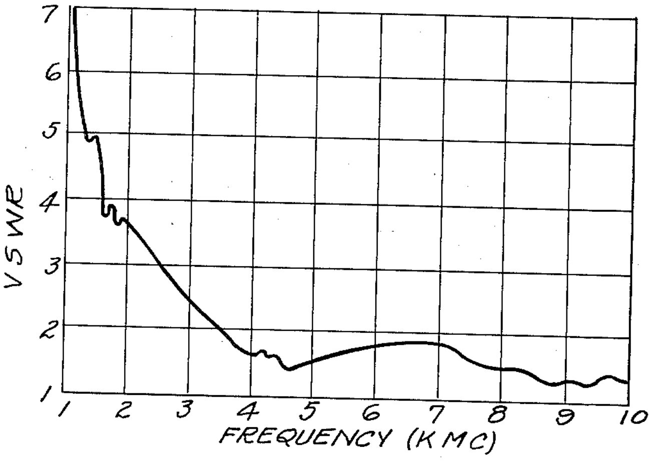

INVENTORS
W. P. TURNER
E. M. TURNER
BY Wade Koontz
ATTORNEY
James S. Shannon
AGENT

Dec. 26, 1961 E. M. TURNER ETAL 3,015,101

SCIMITAR ANTENNA

Filed Oct. 31, 1958

14 Sheets-Sheet 9

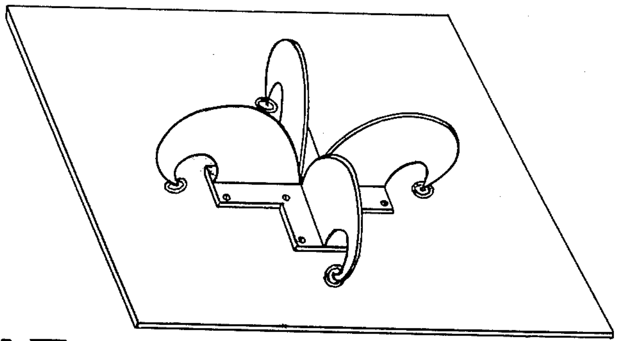

FIG. 12a.

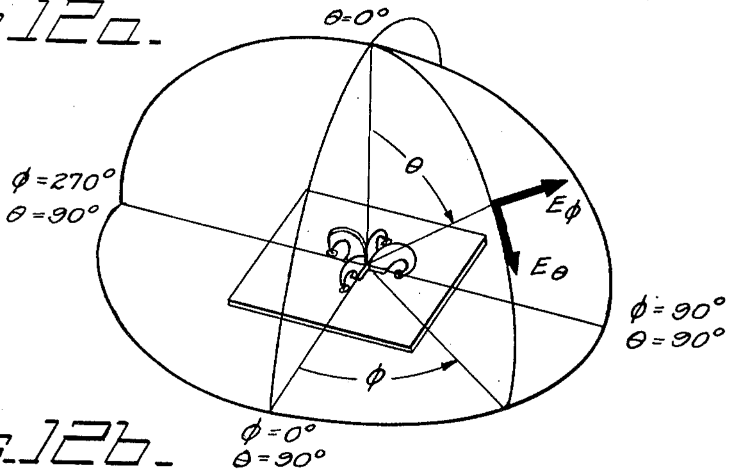
θ=0°
φ=270°
θ=90°
θ
Eφ
Eθ
φ=90°
θ=90°
φ
φ=0°
θ=90°

FIG. 12b.

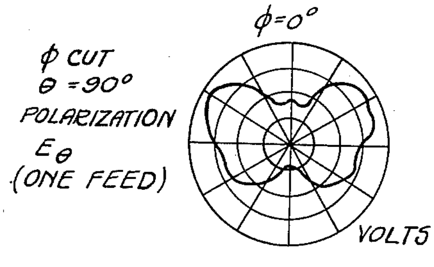

FIG. 12c.

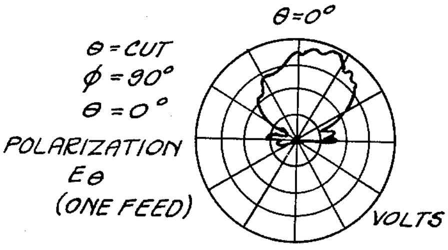

FIG. 12d.

INVENTORS
W. P. TURNER
E. M. TURNER
BY Wade Koontz
ATTORNEY
James L. Shannon
AGENT

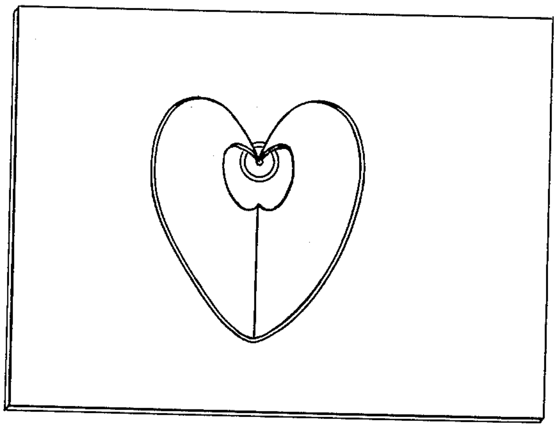

FIG. 13a

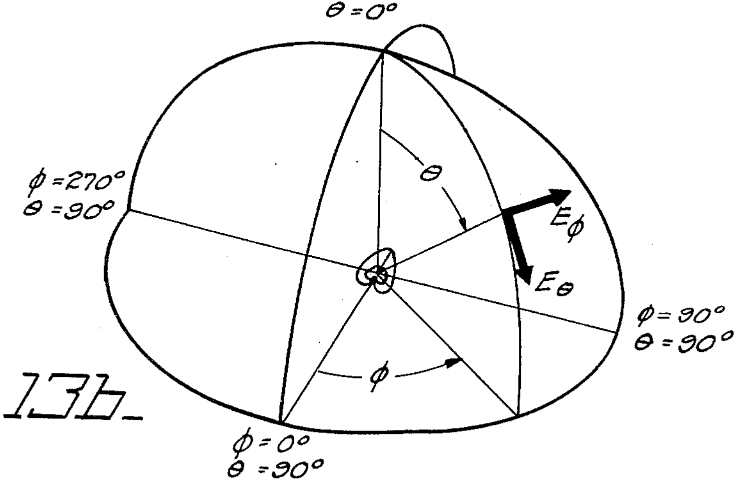
θ = 0°
φ = 270°
θ = 90°
θ
Eφ
Eθ
φ = 90°
θ = 90°
φ
φ = 0°
θ = 90°

FIG. 13b

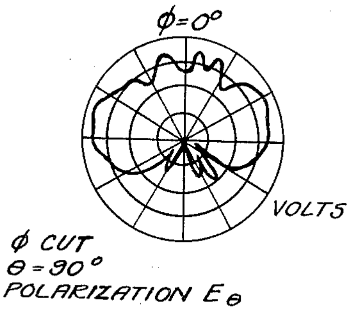

FIG. 13c

FIG. 13d

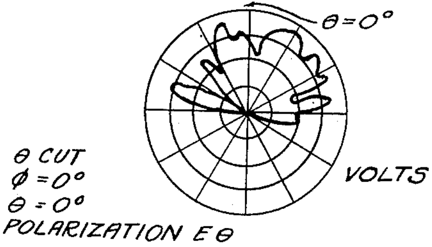

INVENTORS
W. P. TURNER
E. M. TURNER
BY
Wade Koontz
ATTORNEY and
James A. Shannon
AGENT

Dec. 26, 1961 E. M. TURNER ETAL 3,015,101

SCIMITAR ANTENNA

Filed Oct. 31, 1958

14 Sheets-Sheet 11

FIG. 14a

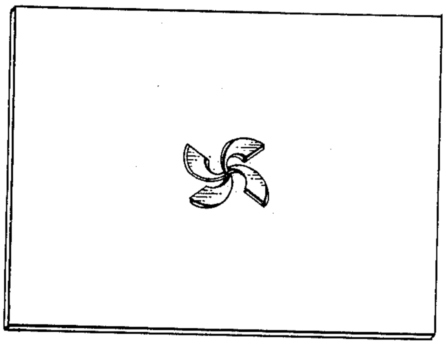

FIG. 14b

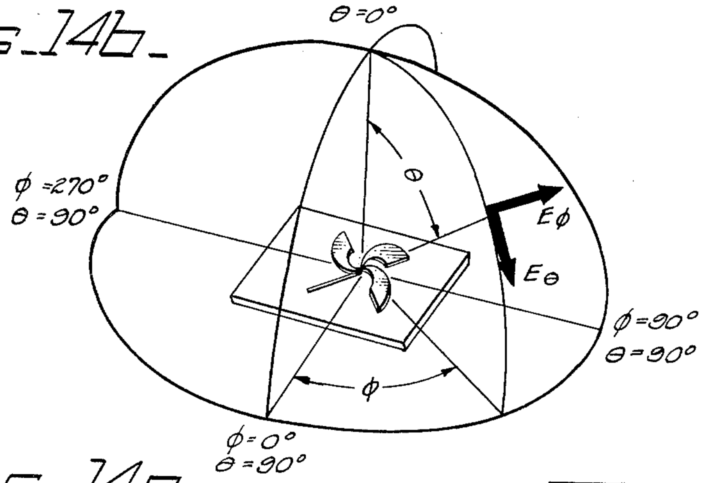
θ=0°
θ
φ=270°
θ=90°
Eφ
Eθ
φ=90°
θ=90°
φ
φ=0°
θ=90°

FIG. 14c

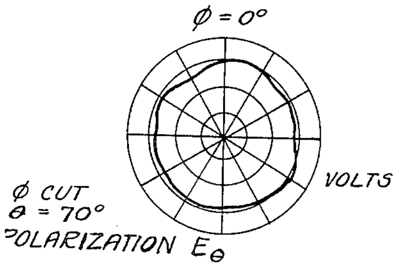

FIG. 14d

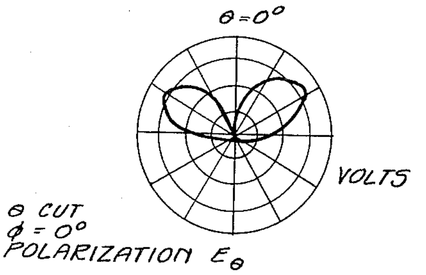

INVENTORS
W. P. TURNER
E. M. TURNER
BY Wade Koontz
ATTORNEY and
James S. Shannon
AGENT

Fig. 15a

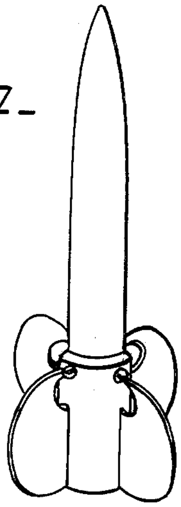

Fig. 15b

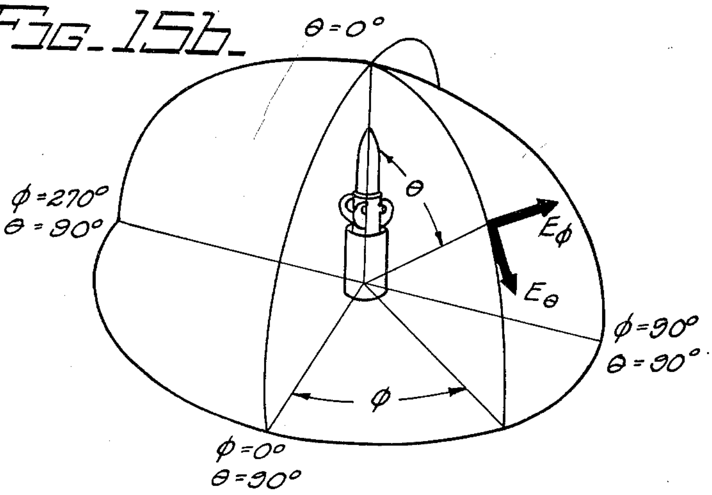
θ=0°
θ
Eφ
Eθ
φ=270°
θ=90°
φ=90°
θ=90°
φ
φ=0°
θ=90°

Fig. 15c

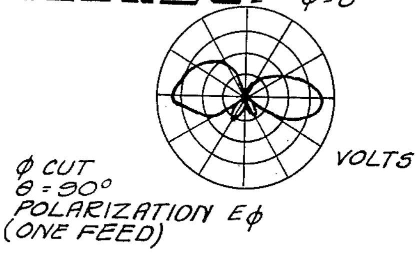

Fig. 15d

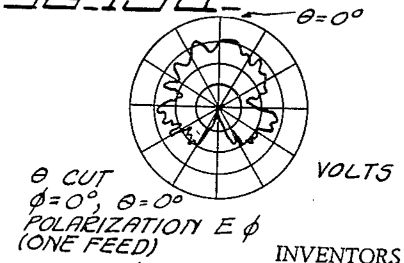

INVENTORS
W. P. TURNER
E. M. TURNER
BY Wade Koontz
ATTORNEY and
James S. Shannon
AGENT

Dec. 26, 1961

E. M. TURNER ET AL

3,015,101

SCIMITAR ANTENNA

Filed Oct. 31, 1958

14 Sheets-Sheet 13

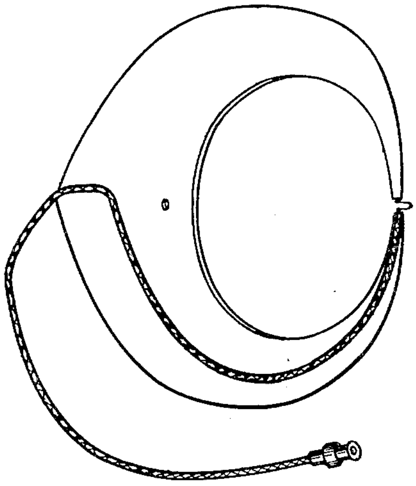

Fig. 17a.

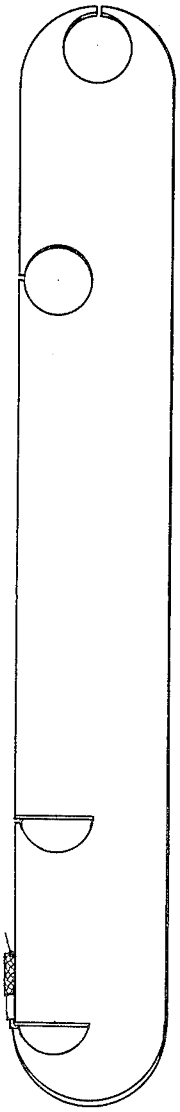

Fig. 16.

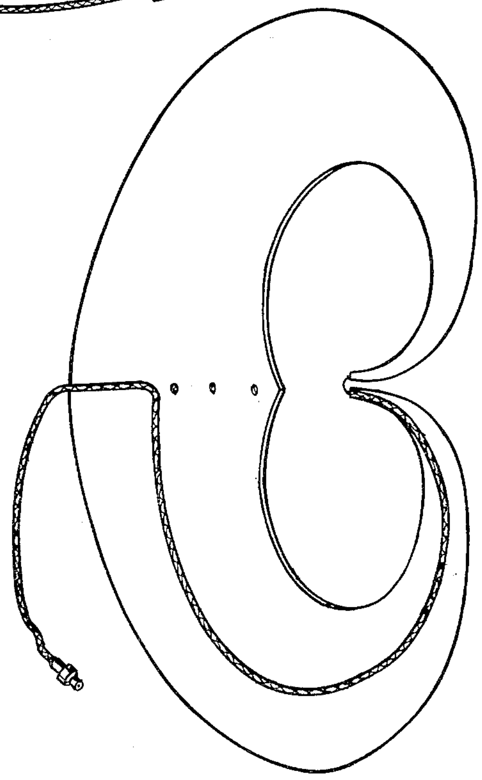

Fig. 17b.

INVENTORS
W.P. TURNER
E.M. TURNER
BY
Wade Koontz
ATTORNEY and
James S. Shannon
AGENT

Dec. 26, 1961

E. M. TURNER ETAL

3,015,101

SCIMITAR ANTENNA

Filed Oct. 31, 1958

14 Sheets-Sheet 14

FIG. 18a

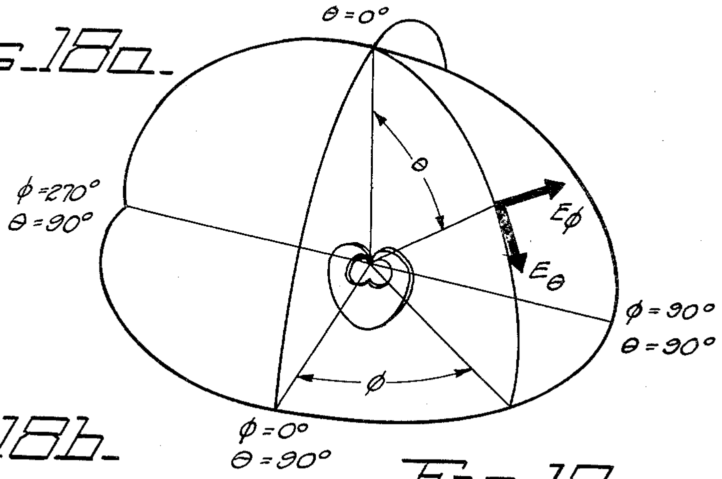
θ = 0°
θ
φ = 270°
θ = 90°
Eφ
Eθ
φ = 90°
θ = 90°
φ
φ = 0°
θ = 90°

FIG. 18b

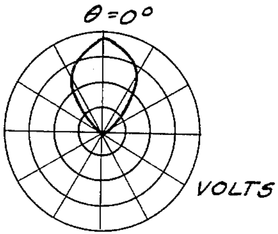

θ CUT, φ = 90°
θ = 0°
POLARIZATION Eθ
NO GROUND PLANE

FIG. 18c

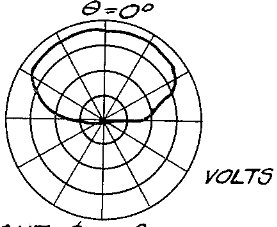

θ CUT, φ = 0°
θ = 0°
POLARIZATION = Eφ
NO GROUND PLANE

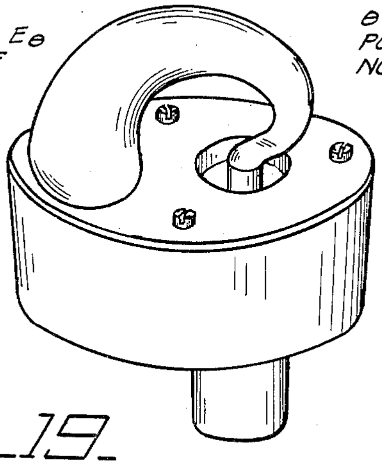

FIG. 19

INVENTORS
W. P. TURNER
E. M. TURNER
BY Wade Koontz
ATTORNEY
James S. Shannon
AGENT

United States Patent Office

3,015,101

Patented Dec. 26, 1961

3,015,101
SCIMITAR ANTENNA

Edwin M. Turner and William P. Turner, Dayton, Ohio, assignors to the United States of America as represented by the Secretary of the Air Force

Filed Oct. 31, 1958, Ser. No. 771,168
5 Claims. (Cl. 343—848)
(Granted under Title 35, U.S. Code (1952), sec. 266)

The invention described herein may be manufactured and used by or for the United States Government for governmental purposes without payment to us of any royalty thereon.

This invention relates to antennas operable in the 30 to 30,000 mc./s. frequency range and is particularly concerned with the provision of an antenna for this frequency range having the desirable electrical properties of pattern stability, impedance stability and high efficiency over a bandwidth of the order of 10 to 1, and having physical properties such as ruggedness, simplicity, ease of mounting and low drag that make it particularly suited to use on aircraft.

The antenna consists of one or more elements each essentially a coplanar equiangular stub antenna with a folded over shorted base, the general configuration being that of a scimitar blade having its broad end electrically and mechanically attached to the ground plane, which may be the surface of the aircraft, and its point coupled to the feed line.

A more detailed description of the antenna will be given with reference to the specific embodiments thereof shown in the accompanying drawing in which

FIGS. 1, 2 and 3 show various forms of the antenna with a coaxial feed line,

FIGS. 3*a*, 3*b*, 3*c* and 3*d* show waveguide feeds for an antenna element,

FIGS. 4*a* and 4*b* show plan and elevation views, respectively, of a physical embodiment of the antenna,

FIG. 5 illustrates the coordinate system used in portraying antenna patterns,

FIGS. 6, 7, 8, 9 and 10 show various field patterns of an antenna such as shown in FIGS. 4*a*—4*b* based on the coordinate system of FIG. 5,

FIG. 11 is a graph illustrating the relatively stable impedance of an antenna of the type described,

FIGS. 12*a*, 12*b*, 12*c*, 12*d*, 13*a*, 13*b*, 13*c*, 13*d*, 14*a*, 14*b*, 14*c*, 14*d* and 15*a*, 15*b*, 15*c*, 15*d* show forms of the antenna using more than one scimitar element,

FIG. 16 shows the manner in which the antenna may be incorporated in a flat plate such as an aerodynamic control element,

FIGS. 17*a* and 17*b* show other forms of the antenna using dual elements,

FIGS. 18*a*, 18*b* and 18*c* show the characteristics of antenna 17*a*, and

FIG. 19 is form of the antenna in which the element has a circular cross section.

FIG. 1 shows the general configuration of the antenna element. The outer and inner boundaries of the antenna element, which may be cut from a sheet of suitable conductive material such as copper or brass, silverplated if desired, to increase surface conductivity, follow diverging spiral paths 1 and 2. These start at feed point F and end at a conductive ground plane G to which the broad end of the element is electrically connected. The curves 1 and 2 are logarithmic spirals having their origin at O which, in this case, lies in the ground plane G. The general equation of these curves is

$$\rho = ke^{a\alpha} \tag{1}$$

where ρ is the distance from origin to curve, α is the direction of ρ measured from the starting point, in this case the feed point F, e is the natural logarithm base, and k and a are constants which provide the parameters through which various forms of the antenna element may be derived. The constant k is the same for both inner and outer boundaries and determines the overall size of the antenna element. The constant a determines the rate at which the spiral moves outward from the origin O. As a approaches zero, the spiral approaches a circle. In FIG. 1, the values of a for the outer boundary, designated a_1, is 0.35 while that for the inner boundary, designated a_2 is 0.08. A characteristic of the logarithmic spiral is that the angle (β_1 or β_2) between the tangent to any point and the line from that point to the origin is constant for all points on the curve.

The difference between a_1 and a_2 determines the rate of divergence of the outer and inner boundaries of the antenna element. Decreasing the difference causes the rate of divergence to decrease as may be seen by comparing FIG. 2, where $a_2=.2$, with FIG. 1. Increasing this difference causes the rate of divergence to increase as may be seen by comparing the antenna element of FIG. 3, for which $\alpha_2=0.02$, with FIG. 1. FIG. 3 also illustrates the effect of placing the origin for the outer and inner boundaries above the ground plane rather than in the ground plane as in FIGS. 1 and 2. The effect is to increase both the length of the antenna element and its height above the ground plane.

The antenna may be fed by a coaxial transmission line having its outer conductor connected to the surface forming the ground plane and its inner conductor extending through a hole in this surface to the feed point F of the antenna element. This is illustrated in FIGS. 2 and 3 where 3 and 4 are the outer and inner conductors, respectively, of the coaxial line. The antenna element may also be fed from a waveguide as shown in FIGS. 3*a*, 3*b*, 3*c* and 3*d*.

FIGS. 4*a* and 4*b* show a practical embodiment of a scimitar antenna in which the antenna element has the same parameters as that in FIG. 3. The antenna element is shown as having a thickness of 0.0625" and has a base plate 5 joined to its broad end for convenient attachment, by means of screws 6 and subplate 7, to the metallic aircraft skin 8 which serves as the ground plane. A suitable coaxial line fitting 9 is attached to the under side of skin 8 by screws 10 and serves to connect the outer conductor 3 of the coaxial line to the metallic surface 8. The center conductor 4 of the transmission line extends through a hole 11 in skin 8 for connection, as by soldering, to the feed point F of the antenna element.

The constants a_1, a_2 and k, the position of origin O relative to the ground surface, and the shape of the ground surface are all parameters affecting the electrical characteristics of the antenna. These may be varied experimentally to obtain the characteristics desired. FIGS. 6–10 show field strength patterns, on a voltage basis, extending over a frequency range of 1000 mc./s. to 8400 mc./s. for an antenna having substantially the same parameters as that shown in FIGS. 4*a*–4*b* and using a flat 14" x 14" ground plane. These patterns are with respect to the coordinate system shown in FIG. 5, E_θ and E_ϕ being the vertical and horizontal components, respectively, of the electric vector.

FIG. 11 shows the voltage standing wave ratio (VSWR) measured on a transmission line feeding the typical scimitar antenna. These measurements are made over a wide frequency band and show the relatively low value and relatively small variation of the VSWR of the greater portion of the band.

FIG. 12*a* shows an antenna made up of four basic scimitar elements. All or any number of these elements may be fed in order to achieve the desired directivity pattern. FIGS. 12*c* and 12*d* show patterns, based on the coordinate system of FIG. 12*b*, as specified for a single feed.

FIG. 13*a* shows a "butterfly" antenna arrangement consisting of two scimitar antennas joined at their bases and feed points and making equal angles with the ground plane. FIGS. 13*c* and 13*d* show patterns as specified based on the coordinate system of FIG. 13*b*.

FIG. 14*a* shows a "cloverleaf" arrangement consisting of four basic scimitar elements equally spaced and equally inclined to the ground plane. All elements are fed from a common point. FIGS. 14*c* and 14*d* show patterns as specified based on the coordinate system of FIG. 14*b*.

FIG. 15*a* shows the manner in which the stabilizing fins of a missile may be shaped to perform the functions of radio antennas. Each fin constitutes a basic scimitar antenna fed through an opening in the missile body. FIGS. 15*c* and 15*d* show patterns as specified for a single antenna element based on the coordinate system of FIG. 15*b*.

FIG. 16 shows various ways in which antennas of the scimitar type or similar thereto may be incorporated in a thin metallic member which may be, for example, an aerodynamic control element, on a missile or similar device. The method of feeding one of the antennas with a coaxial transmission line is shown. The other elements may be similarly fed.

FIGS. 17*a* and 17*b* show two forms of a dual element antenna in which the two scimitar elements lie in the same plane. FIGS. 18*b* and 18*c* show patterns as specified for the antenna of FIG. 17*a* based on the coordinate system of FIG. 18*a*.

Although it is conventient and in many cases desirable to make the scimitar element from a flat relatively thin piece of metal, it may also be given other than a thin rectangular cross section as shown in FIG. 19 in which the antenna element has a circular cross section.

We claim:

1. An antenna comprising means forming a conductive ground surface and a curved antenna element attached and electrically connected at one end to said ground surface and extending to a feed point situated substantially in the plane of said ground surface, the farther and nearer boundaries of said element relative to said surface being two diverging co-planar logarithmic spirals starting at said feed point and ending at said surface.

2. An antenna as claimed in claim 1 in which said logarithmic spirals have a common origin.

3. An antenna comprising means forming a conductive ground surface and a curved antenna element attached to and electrically connected at one end to said ground surface and extending to a feed point situated substantially in the plane of said ground surface, said element having a uniform thickness and having farther and nearer boundaries relative to said surface that are diverging logarithmic spirals starting at said feed point and ending at said surface.

4. An antenna comprising means forming a conductive ground surface, a plurality of curved antenna elements each attached to and electrically connected at one end to said ground surface and extending to a common feed point situated substantially in the plane of said ground surface, each of said elements having a uniform thickness and having farther and nearer boundaries relative to said ground surface that are two diverging co-planar logarithmic spirals starting at said feed point and ending at said ground surface, the plane of said spirals being at an acute angle to said ground surface, and a transmission line situated beneath said ground surface and coupled through a hole in said ground surface to said antenna elements at said feed point.

5. Apparatus as claimed in claim 4 in which said elements are of equal size and are equally spaced about said feed point.

References Cited in the file of this patent

UNITED STATES PATENTS

2,615,134	Carter	Oct. 21, 1952
2,755,465	Ramsay	July 17, 1956

FOREIGN PATENTS

1,134,578	France	Apr. 15, 1957

OTHER REFERENCES

I.R.E. Convention Record, part I, vol. 5, March 18, 1957.

"Airborne Spiral Antennas Minimize Drag," by P. J. Klass, Aviation Week, July 14, 1958, pp. 75 to 82.

US007688036B2

(12) **United States Patent**
Yarger et al.

(10) **Patent No.: US 7,688,036 B2**
(45) **Date of Patent: Mar. 30, 2010**

(54) **SYSTEM AND METHOD FOR STORING ENERGY**

(75) Inventors: **Eric Jay Yarger**, Rigby, ID (US); **John Morrison**, Butte, MT (US); **John Grant Richardson**, Idaho Falls, ID (US); **David Frazer Spencer**, Idaho Falls, ID (US); **Dale W. Christiansen**, Blackfoot, ID (US)

(73) Assignee: **Battelle Energy Alliance, LLC**, Idaho Falls, ID (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 318 days.

(21) Appl. No.: **11/475,564**

(22) Filed: **Jun. 26, 2006**

(65) **Prior Publication Data**

US 2008/0074083 A1 Mar. 27, 2008

(51) **Int. Cl.**

H02J 7/00	(2006.01)
H04B 1/38	(2006.01)
G11B 13/00	(2006.01)
H01F 27/02	(2006.01)

(52) **U.S. Cl.** **320/137**; 320/138; 455/573; 369/13.02; 336/90

(58) **Field of Classification Search** 320/137, 320/138; 455/573
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2,209,809	A	7/1940	Burton	175/21
2,302,687	A	11/1942	Dyer	290/37
3,022,400	A	2/1962	von Ahlefeldt	200/93
3,031,802	A	5/1962	Leliaert	
3,549,818	A	12/1970	Turner	179/82
3,607,403	A	9/1971	Arrance	136/6
3,660,760	A	5/1972	Schaad et al.	325/28
3,665,123	A	5/1972	Ikeda	179/100.41 K
3,665,597	A	5/1972	Tan et al.	
3,809,825	A	5/1974	Schaad et al.	179/82
3,925,694	A	12/1975	Richter et al.	
3,963,854	A	6/1976	Fowler	174/36
3,973,269	A	8/1976	Koch et al.	
4,046,244	A	9/1977	Velazquez	
4,080,663	A	3/1978	Wik	366/274

(Continued)

FOREIGN PATENT DOCUMENTS

AU	20022585 A4	6/2002

(Continued)

OTHER PUBLICATIONS

Callahan et al., "Free-Piston Engine Linear Generator for Hybrid Vehicles Modeling Study," Advanced Research Projects Agency, Arlington, VA, May 1995.

(Continued)

Primary Examiner—Edward Tso
Assistant Examiner—Yalkew Fantu
(74) *Attorney, Agent, or Firm*—Seed IP Law Group PLLC

(57) **ABSTRACT**

A self-recharging battery comprising a generator and an energy storage device contained within the battery case. The generator comprises a magnetic structure configured to generate a compressed magnetic field and a coil configured to focus the compressed magnetic field in electrical conductive elements of the coil.

12 Claims, 39 Drawing Sheets
(3 of 39 Drawing Sheet(s) Filed in Color)

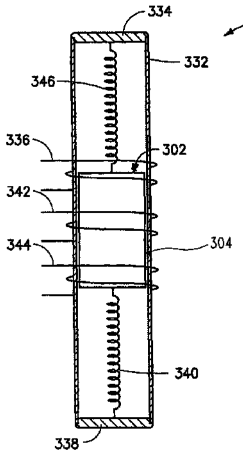

U.S. PATENT DOCUMENTS

4,217,171	A	8/1980	Schaffer	
4,218,712	A	8/1980	Clymer et al.	358/229
4,245,640	A	1/1981	Hunt	128/419 B
4,290,663	A	9/1981	Fowler et al.	339/143 R
4,304,136	A	12/1981	McCabe et al.	
4,320,927	A	3/1982	Sertich	
4,331,285	A	5/1982	Gottwals	
4,363,902	A	12/1982	Kurz	
4,372,214	A	2/1983	Toton et al.	
4,421,118	A	12/1983	Dow et al.	
4,430,679	A	2/1984	van de Bult	360/106
4,460,855	A	7/1984	Kelly	
4,563,727	A	1/1986	Curiel	362/183
4,636,902	A	1/1987	Dalziel et al.	360/128
4,639,667	A	1/1987	Andresen	
4,641,118	A	2/1987	Hirose et al.	
4,646,141	A	2/1987	Timmermans et al.	358/50
4,648,013	A	3/1987	Curiel	362/183
4,890,812	A	1/1990	Chechile et al.	248/674
4,959,504	A	9/1990	Yarger et al.	
4,992,624	A	2/1991	Benson et al.	174/35 R
5,075,809	A	12/1991	Heinz et al.	360/128
5,146,123	A	9/1992	Yarr	
5,161,414	A	11/1992	Rubbelke	73/658
5,181,004	A	1/1993	Johnson	
5,214,710	A	5/1993	Ziegenberg et al.	
5,319,332	A	6/1994	Espinosa	335/214
5,347,186	A	9/1994	Konotchick	
5,463,268	A	10/1995	Schroeder	313/293
5,502,901	A	4/1996	Brown	36/28
5,650,681	A	7/1997	DeLerno	
5,770,945	A	6/1998	Constable	324/350
5,782,512	A	7/1998	Cargnoni	
5,838,138	A	11/1998	Henty	320/107
5,951,459	A	9/1999	Blackwell	600/13
5,975,714	A	11/1999	Vetorino et al.	
5,978,446	A	11/1999	Resnick	
5,997,464	A	12/1999	Blackwell	600/13
6,111,257	A	8/2000	Shand et al.	250/368
6,174,276	B1	1/2001	Blackwell	600/13
6,179,772	B1	1/2001	Blackwell	600/13
6,186,941	B1	2/2001	Blackwell	600/13
6,215,303	B1	4/2001	Weinstock et al.	324/263
6,220,719	B1	4/2001	Vetorino et al.	
6,313,624	B1	11/2001	Alhorn et al.	324/207.17
6,489,772	B1	12/2002	Holladay et al.	324/339
6,550,306	B1	4/2003	Bensley	73/12.04
6,554,097	B2	4/2003	König	181/129
6,600,399	B1	7/2003	Trandafir	335/222
6,603,224	B1	8/2003	Hollingsworth et al.	
6,664,759	B1	12/2003	Goris	
6,731,968	B2	5/2004	Buchanan	600/409
6,759,755	B2	7/2004	Sagov	
6,791,205	B2	9/2004	Woodbridge	
6,798,090	B2	9/2004	Cheung et al.	
6,864,592	B1	3/2005	Kelly	
6,903,725	B2	6/2005	Nacson	345/163
6,915,145	B2 *	7/2005	Tu et al.	455/573
6,936,937	B2	8/2005	Tu et al.	
6,937,881	B2	8/2005	König	455/575.1
6,952,060	B2	10/2005	Goldner et al.	
7,009,310	B2	3/2006	Cheung et al.	
7,327,046	B2	2/2008	Biamonte	290/1 R
7,501,726	B1	3/2009	Waters et al.	
7,573,163	B2	8/2009	Tu et al.	
2001/0033175	A1	10/2001	Marbach et al.	
2002/0066617	A1	6/2002	Konig	181/129
2002/0074881	A1	6/2002	Imlach	
2002/0074882	A1	6/2002	Werfel et al.	
2002/0148625	A1	10/2002	Darling et al.	174/35 R
2002/0154518	A1	10/2002	Elferich et al.	363/15
2002/0160726	A1	10/2002	Konig	455/90
2003/0168924	A1	9/2003	Murakami et al.	
2003/0232627	A1 *	12/2003	Tu et al.	455/550.1
2004/0070482	A1	4/2004	Golan	336/223
2004/0133130	A1	7/2004	Ferry et al.	
2004/0140880	A1	7/2004	Haugs et al.	
2005/0001703	A1	1/2005	Zimmerling	
2005/0067907	A1	3/2005	Takeuchi	
2005/0127765	A1	6/2005	Kummeth	
2005/0141742	A1	6/2005	Kasztelan	381/331
2005/0168307	A1	8/2005	Reynolds et al.	335/220

FOREIGN PATENT DOCUMENTS

CN	2457747		10/2001
CN	1444234	A	9/2003
DE	19702056	A1	7/1998
DE	19947119	A1	5/2001
DE	10132881	A1	1/2003
EP	0192860	A1	9/1986
EP	0290343	A1	11/1988
EP	0192860	B1	3/1990
EP	0523450	A1	1/1993
EP	0290343	B1	2/1993
EP	0786849	B1	7/1997
EP	0933980	A2	8/1999
EP	0933980	A3	4/2000
EP	1143461	A1	10/2001
GB	115789	A	5/1918
GB	424549	A	2/1935
GB	424669	A	2/1935
GB	435416		9/1935
GB	2052886	A	1/1981
GB	8315514	A	7/1983
GB	2181306	A	4/1987
GB	2263386	A	7/1993
GB	2314470	A	12/1997
GB	2354646	A	3/2001
JP	2114438	A	4/1990
JP	3119638	A	5/1991
JP	6207965	A	7/1994
JP	10283079	A	10/1998
JP	2000036418	A	2/2000
JP	2000092783	A	3/2000
SU	546951	A	3/1977
WO	WO 86/04132		7/1986
WO	WO 91/17611		11/1991
WO	WO 92/01362		1/1992
WO	WO 02/15658	A1	2/2002
WO	WO 02/052671	A1	7/2002
WO	WO 02/060041	A2	8/2002
WO	WO 02/060041	A3	8/2002
WO	WO 03/085408	A2	10/2003
WO	WO 03/085408	A3	10/2003
WO	2005031952	A1	4/2005
WO	2009126188	A2	10/2009

OTHER PUBLICATIONS

English Abstract for SU546951A.
English Translation of SU546951A.
United States Patent and Trademark Office Action dated Jul. 10, 2009 entered in co-pending U.S. Appl. No. 11/475,389.

* cited by examiner

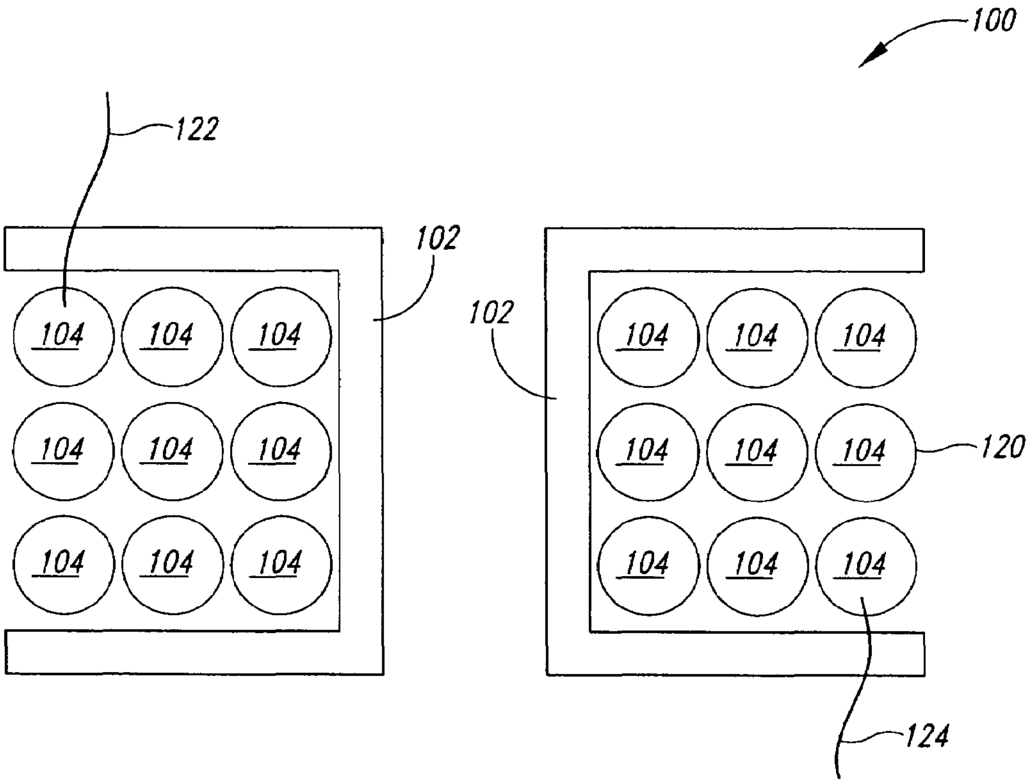
100
122
102
102
104
104
104
104
104
104
104
104
104
104
104
104
104
104
104
104
104
104
120
124

FIG. 1
(Prior Art)

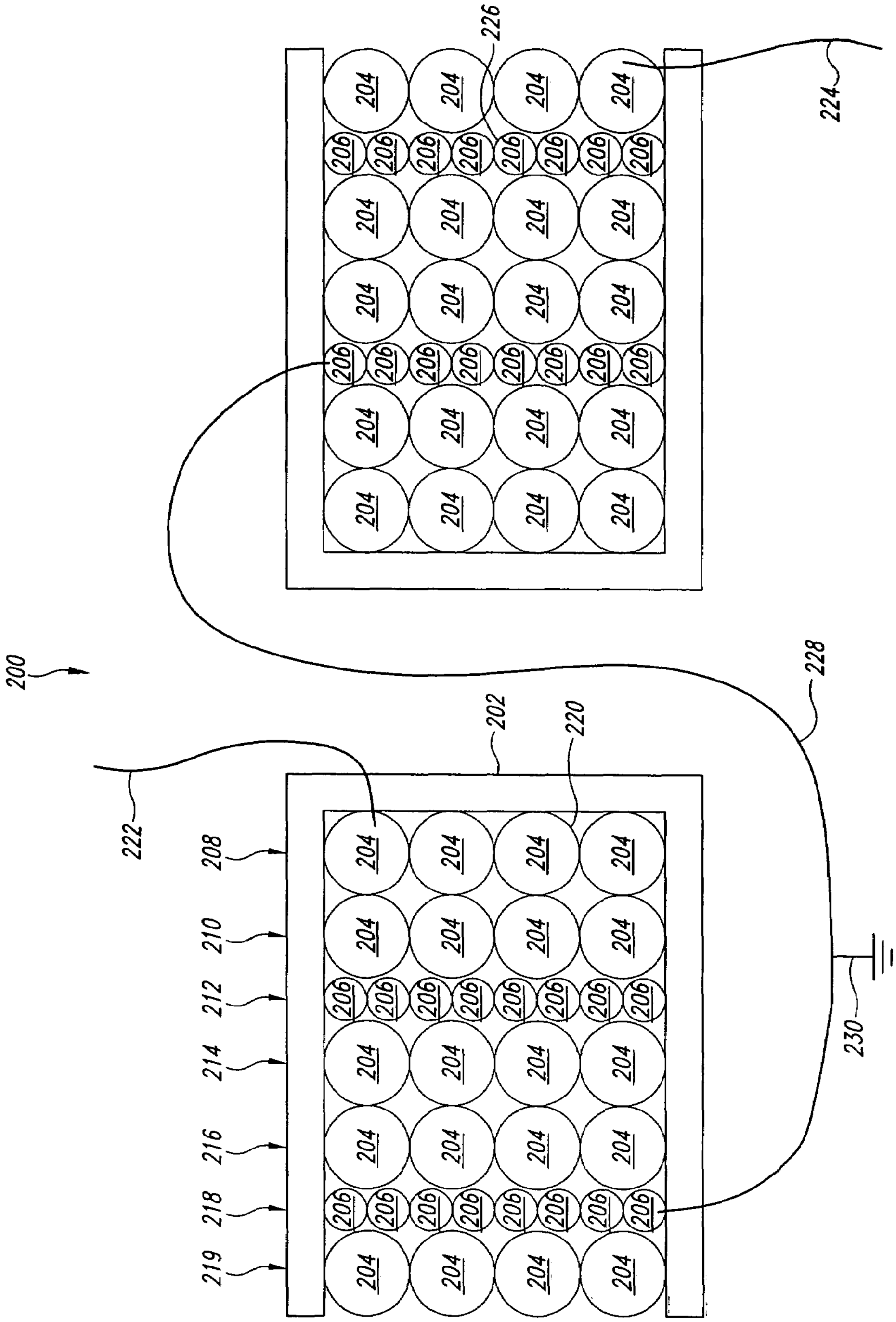

FIG. 2

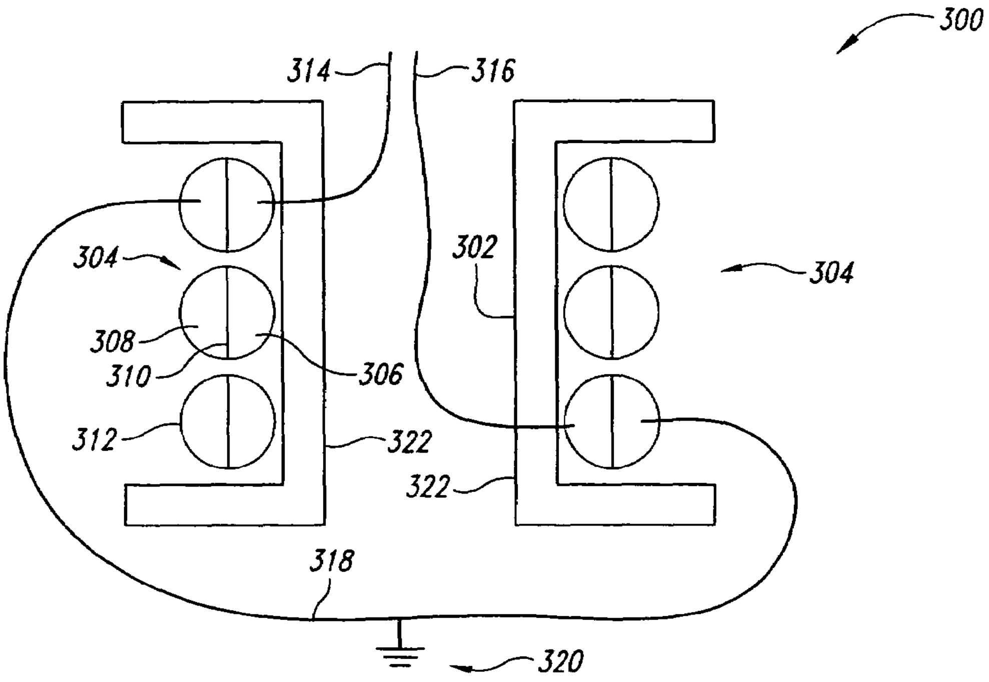

FIG. 3

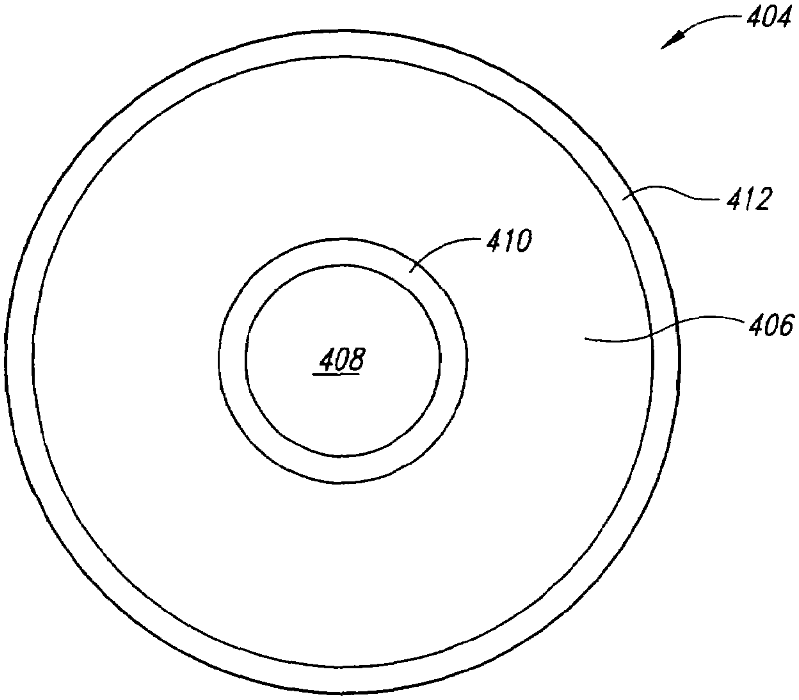

FIG. 4

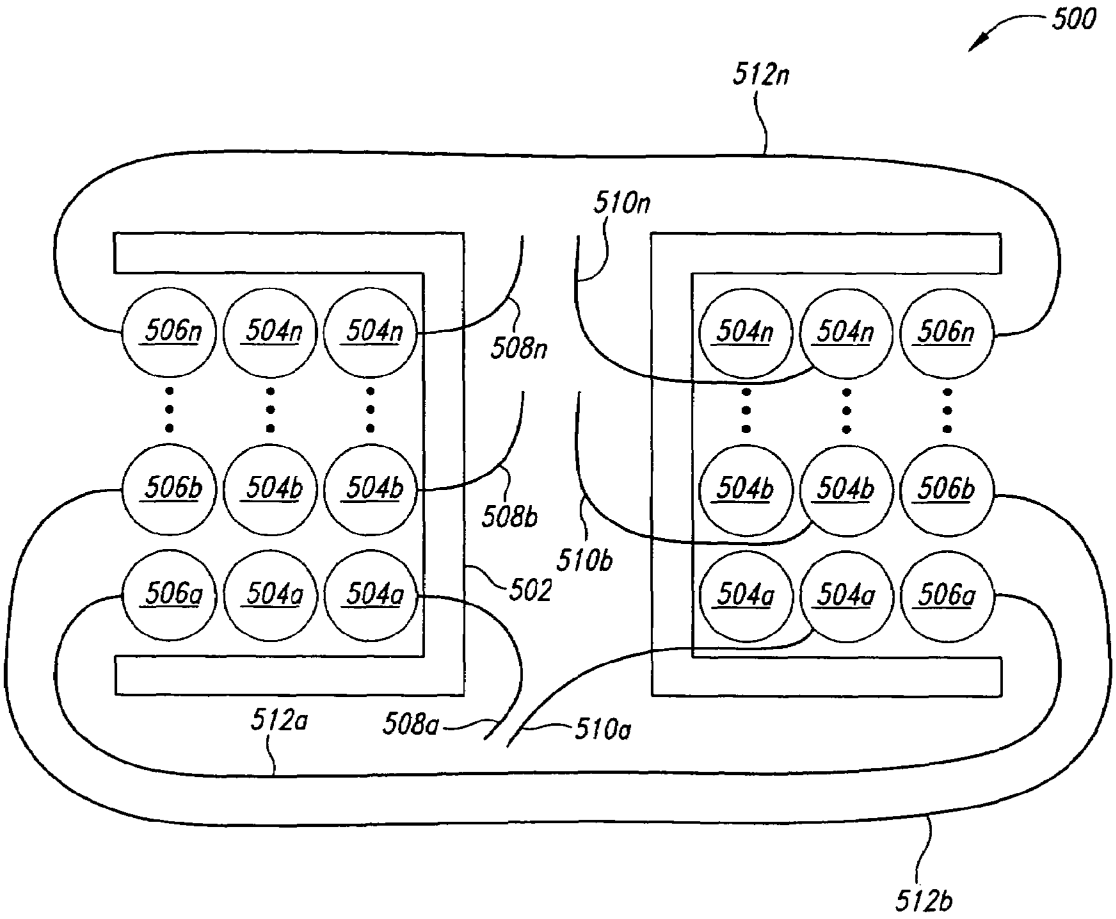
512n
510n
506n
504n
504n
508n
504n
504n
506n

FIG. 5

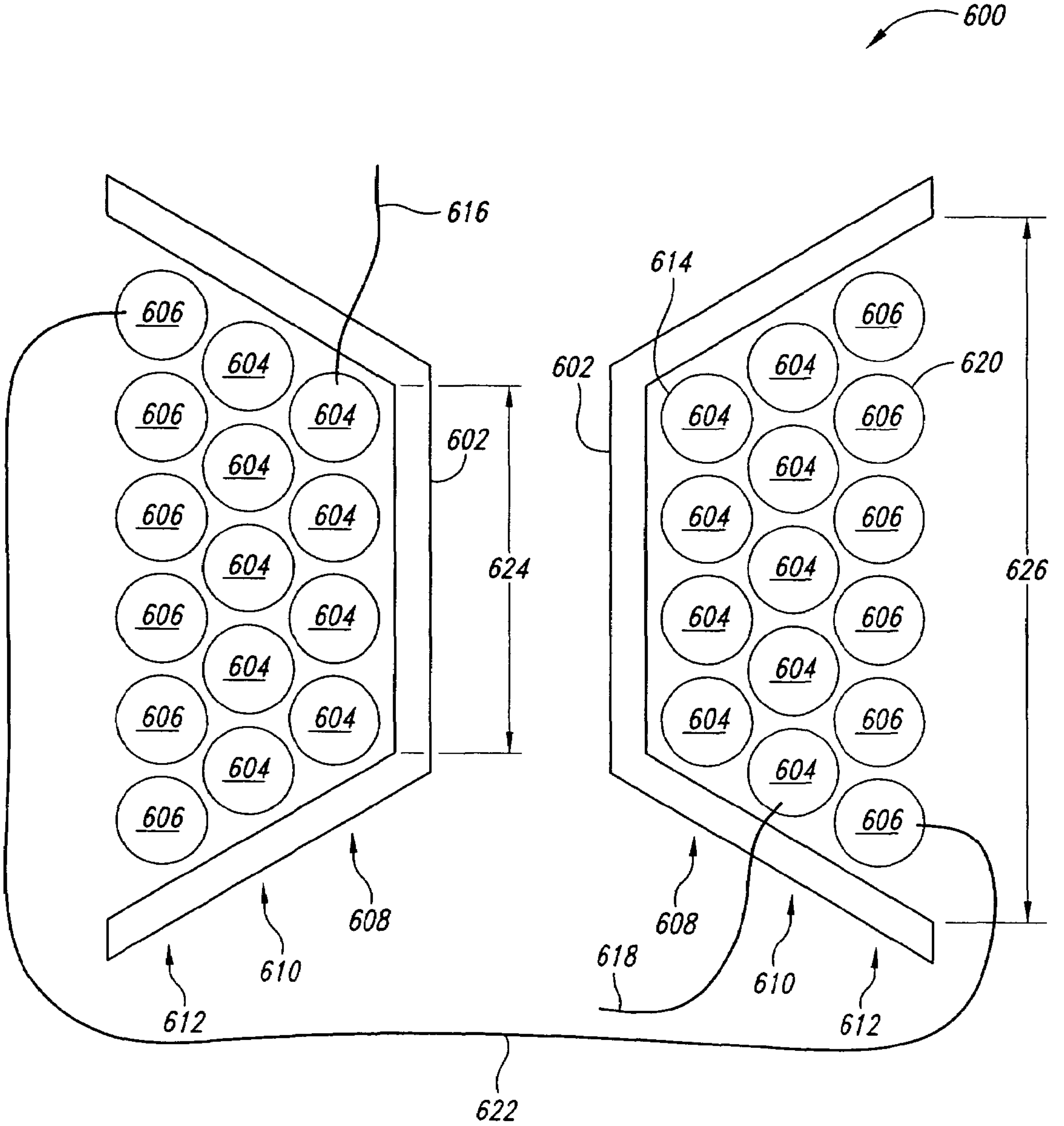
600
616
614
602
620
606
604
602
624
626
608
610
612
618
622

FIG. 6

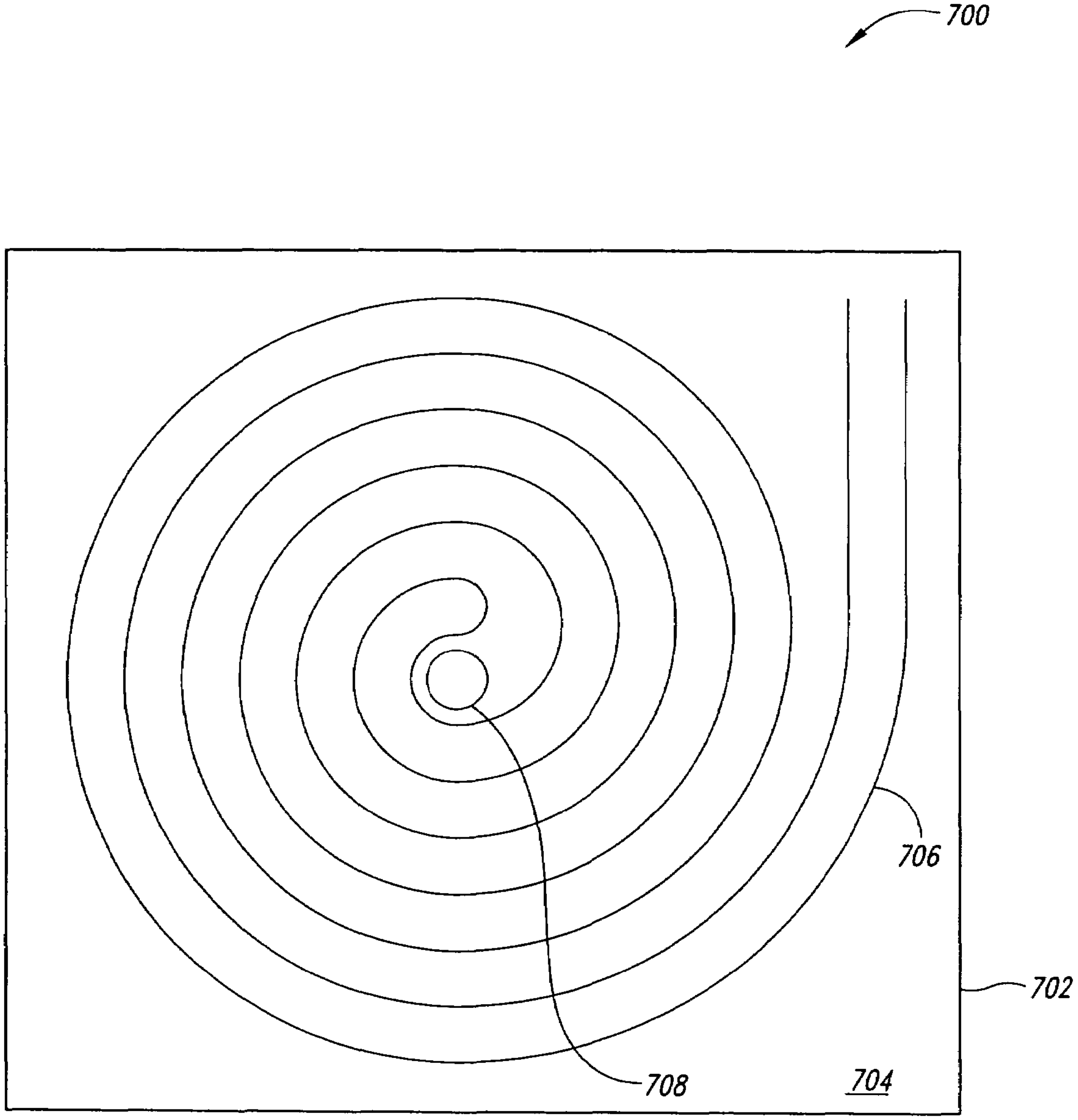
700
706
702
708
704

FIG. 7

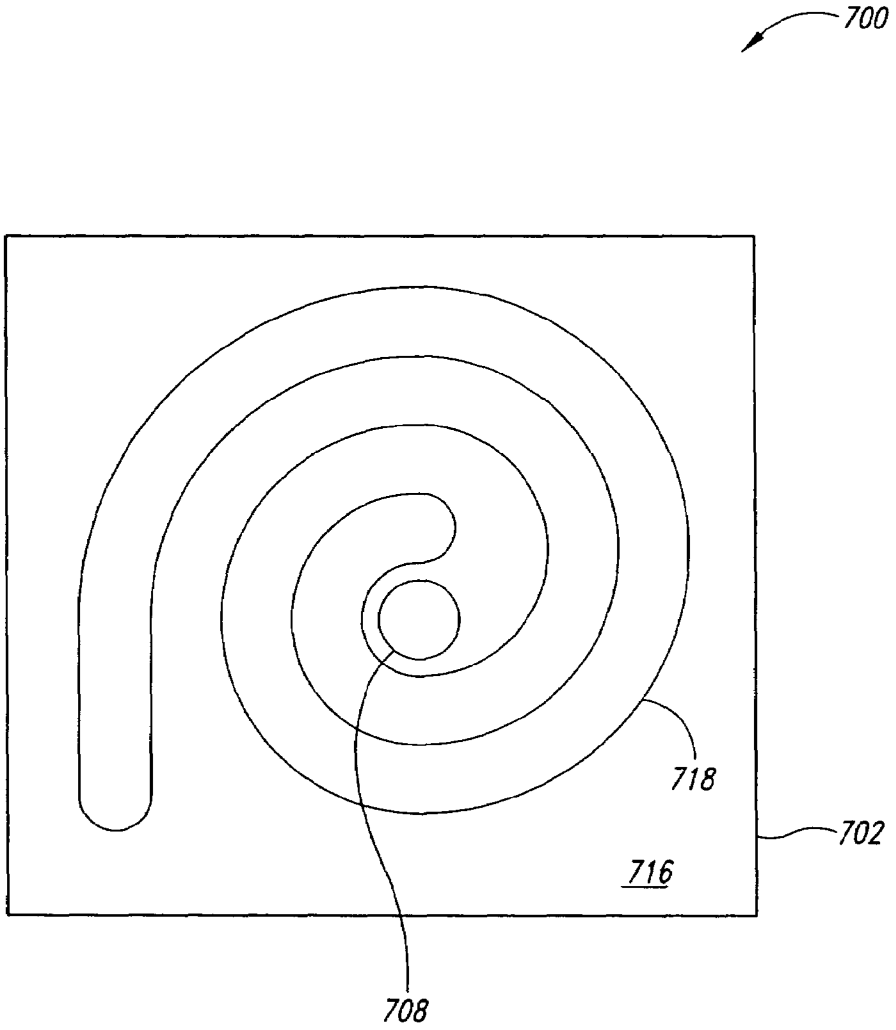
700
718
702
716
708

FIG. 8

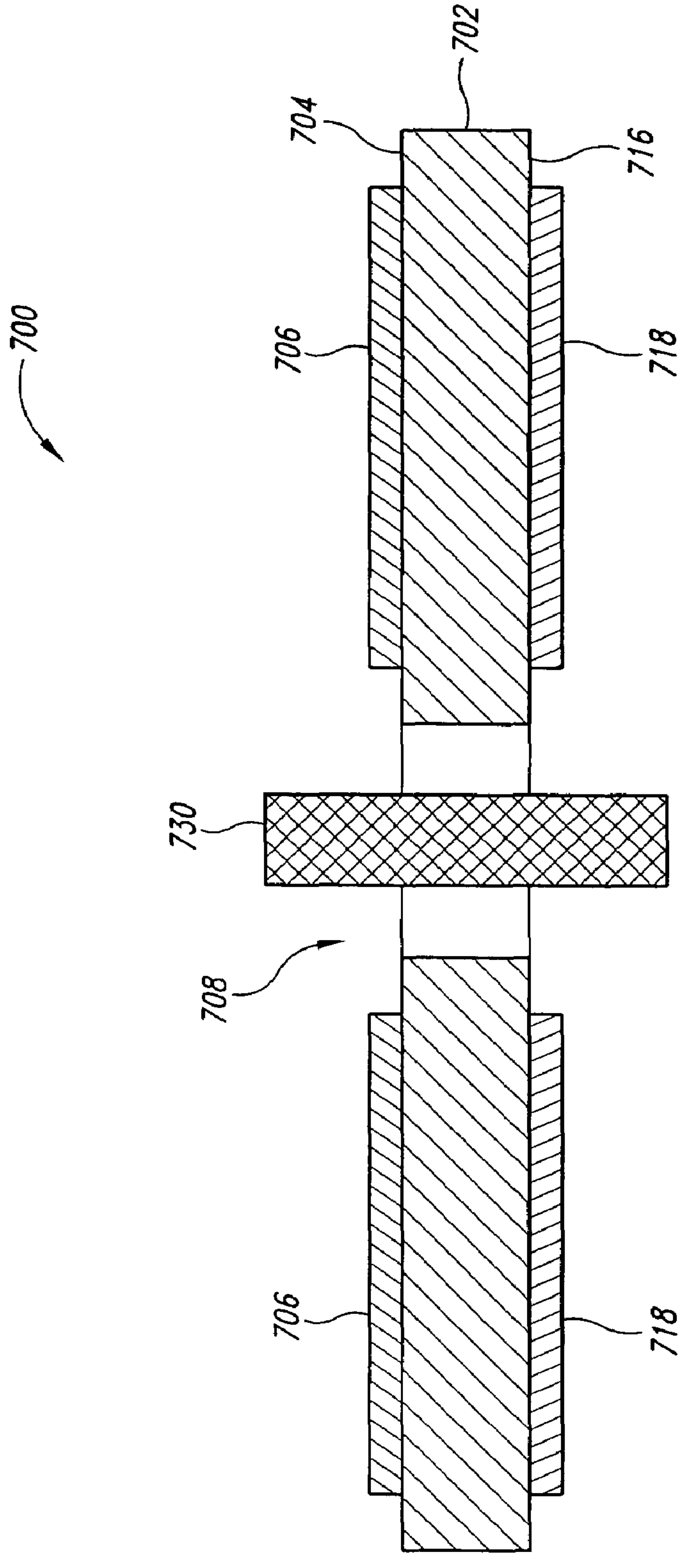
700
702
704
716
706
718
730
708
706
718

FIG. 9

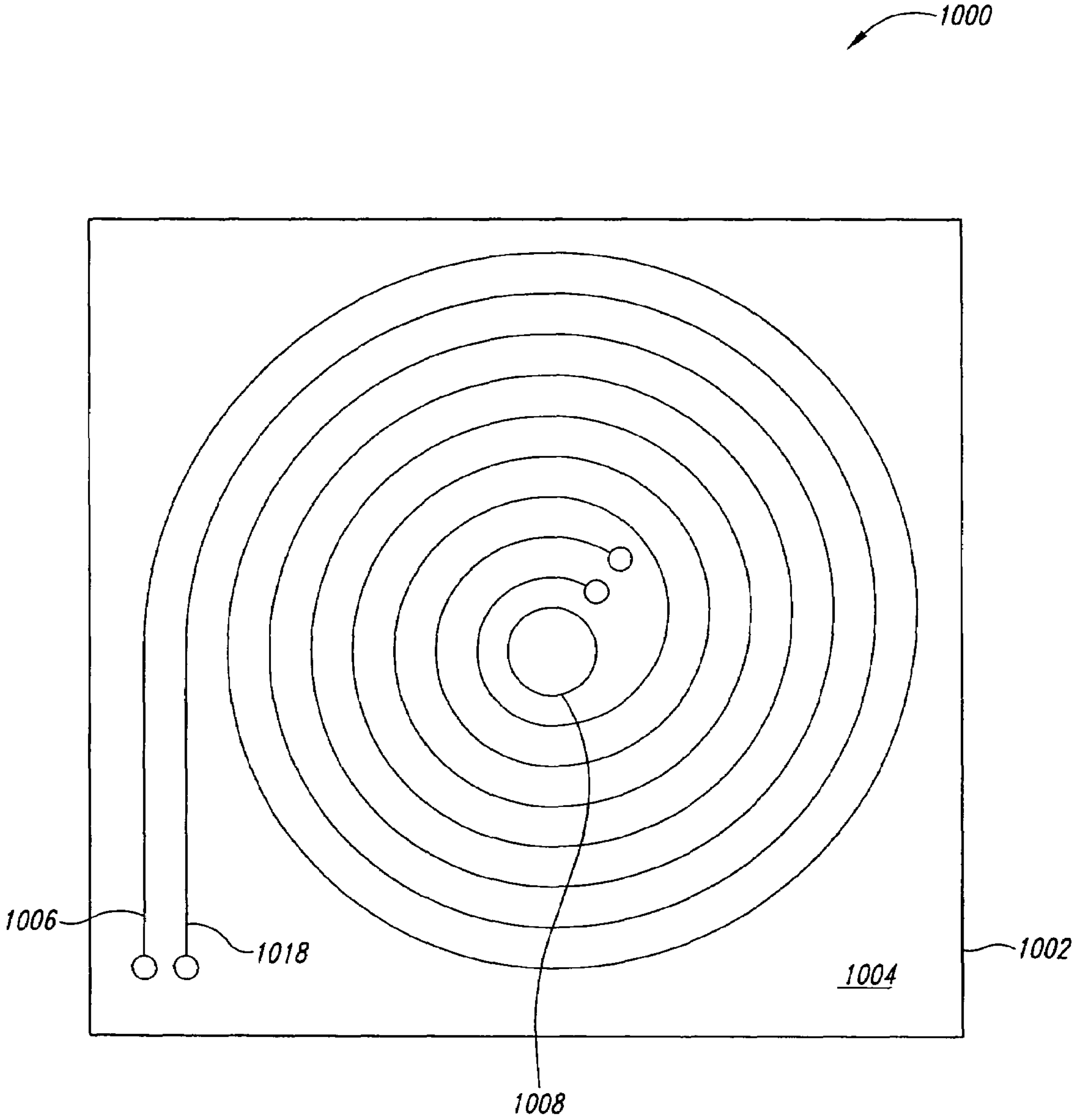
1000
1006
1018
1002
1004
1008

FIG. 10

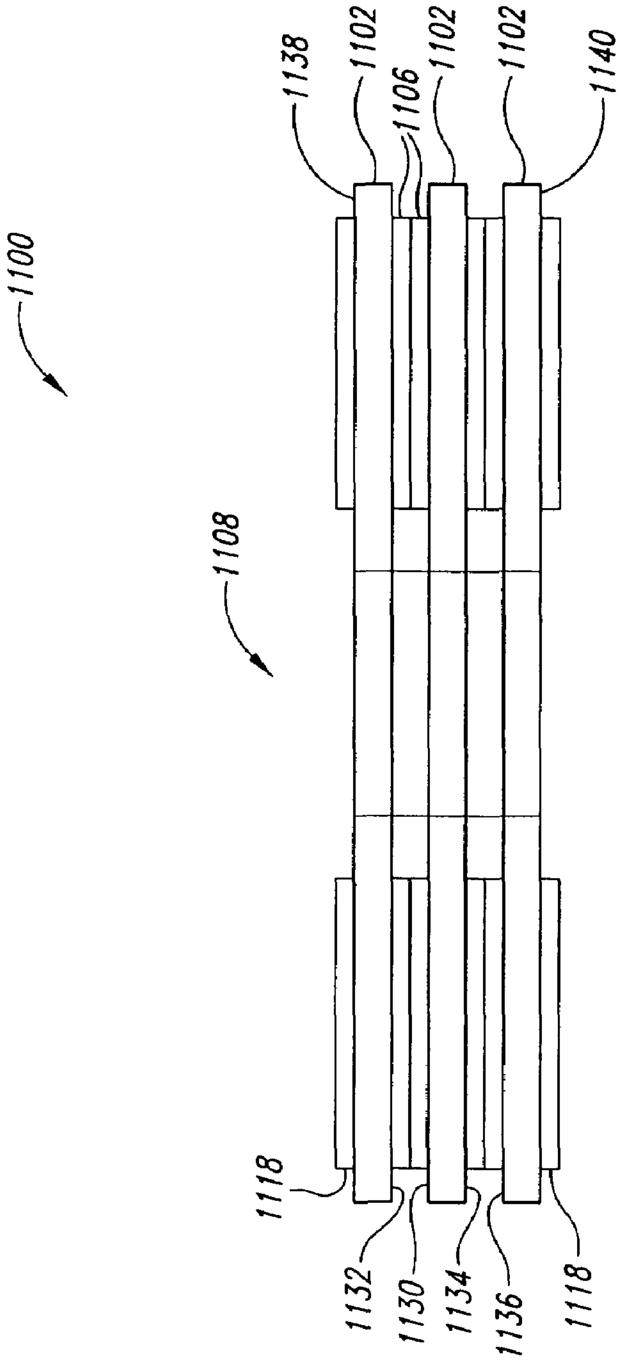

FIG. 11

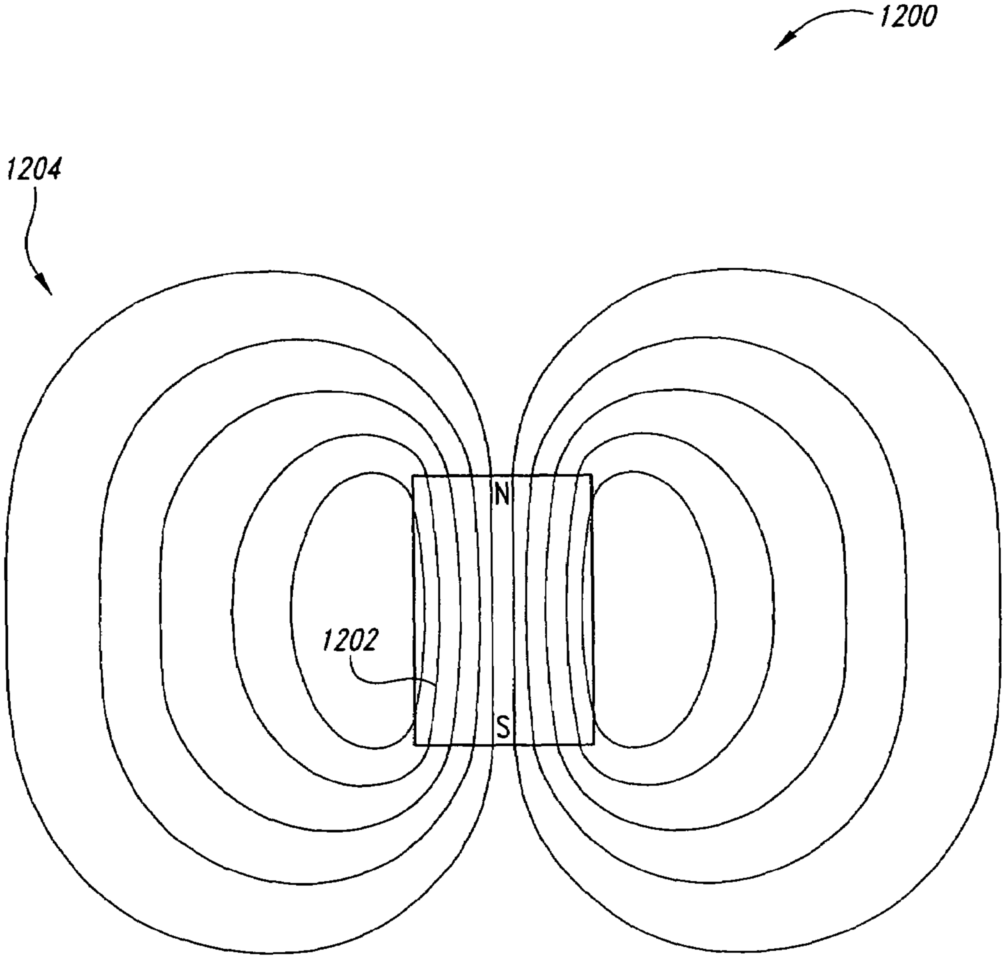
1200
1204
N
1202
S

FIG. 12
(Prior Art)

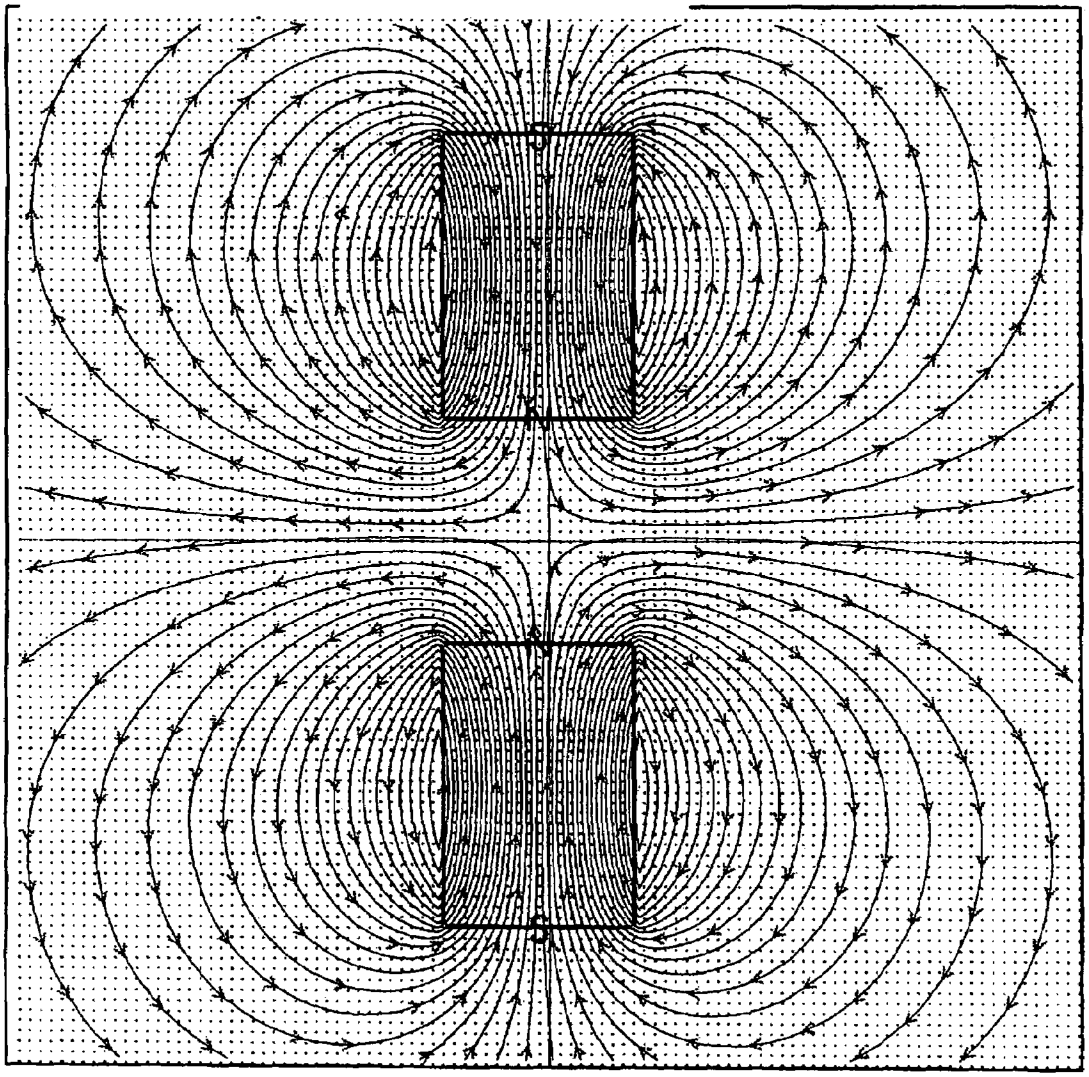

FIG. 13A

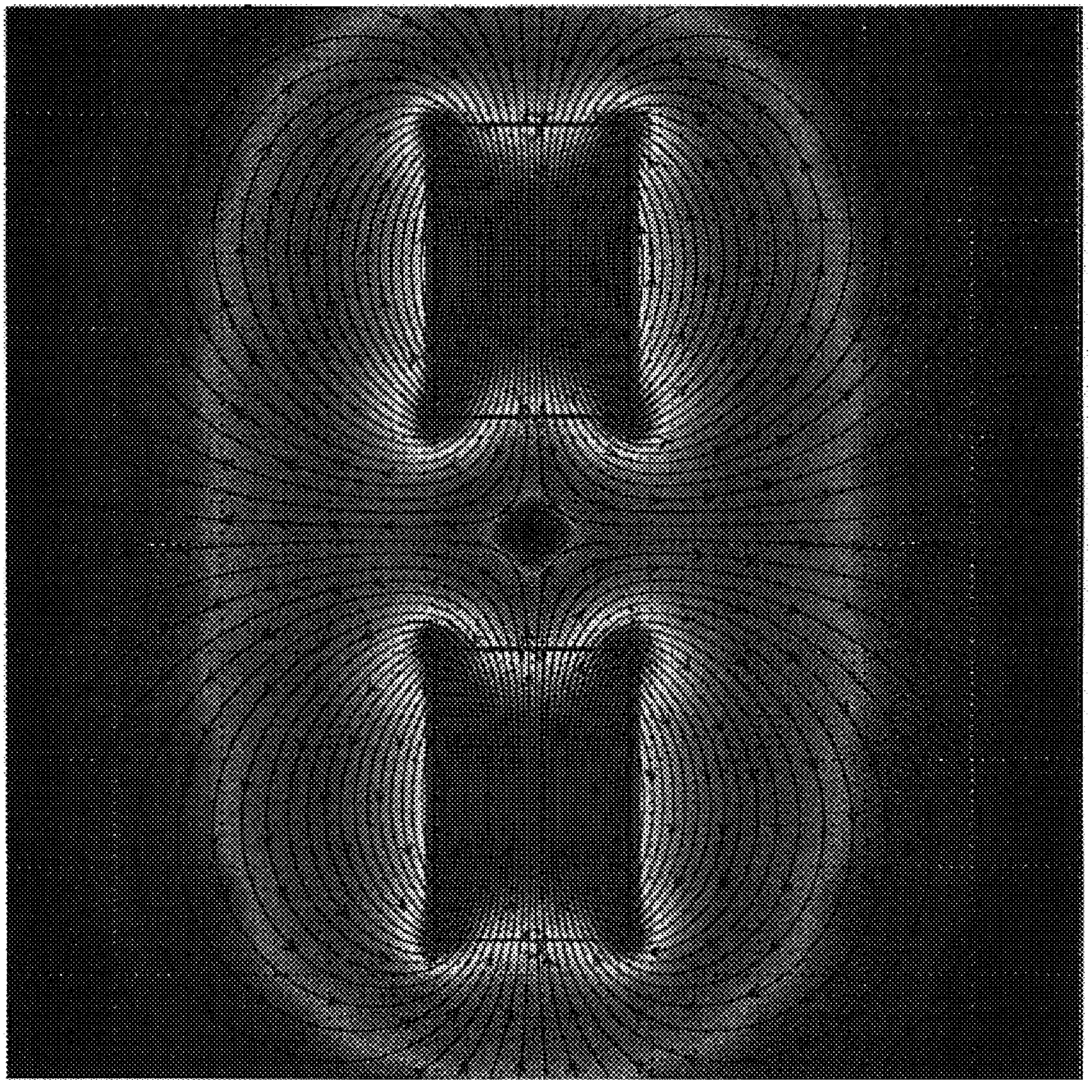

FIG. 13B

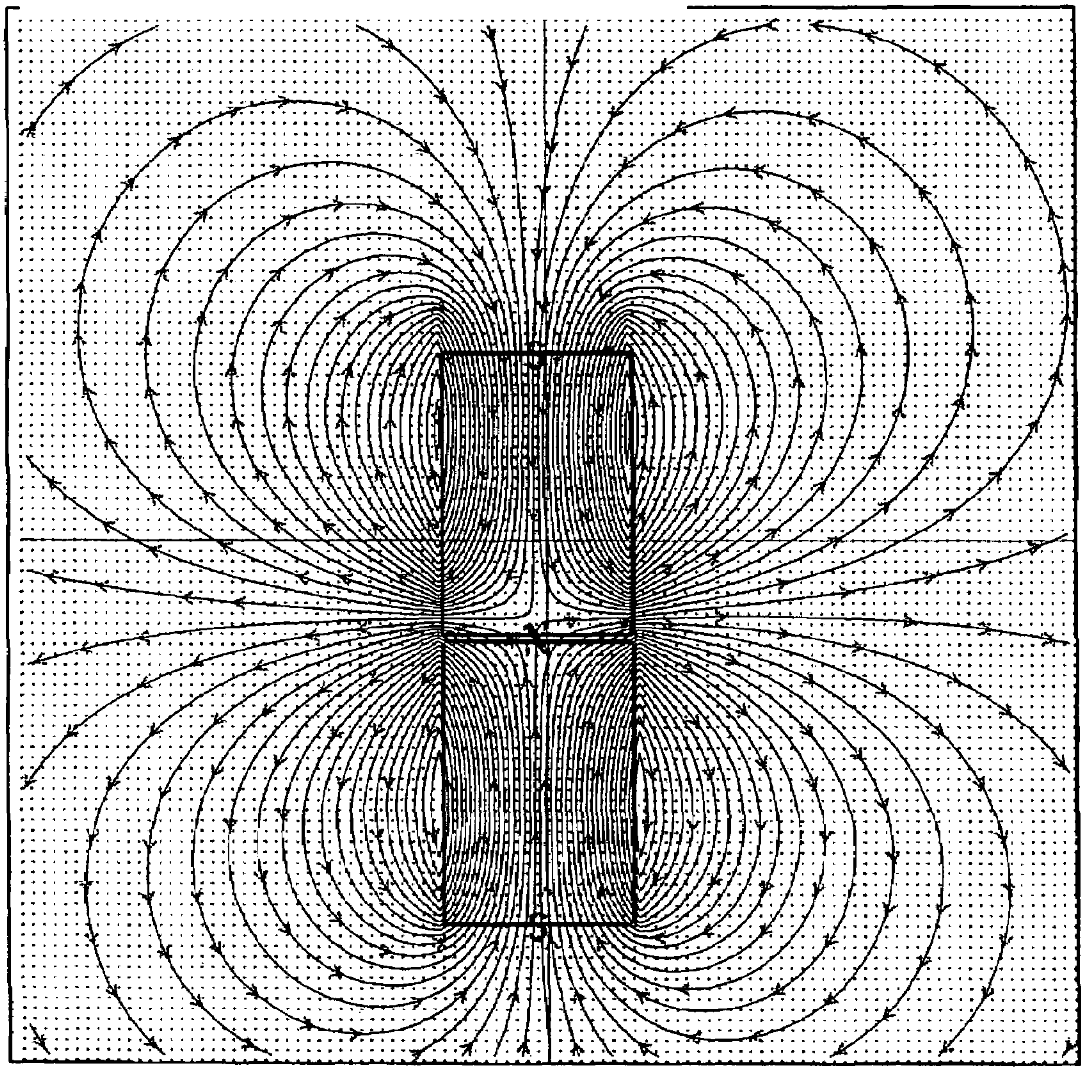

FIG. 14A

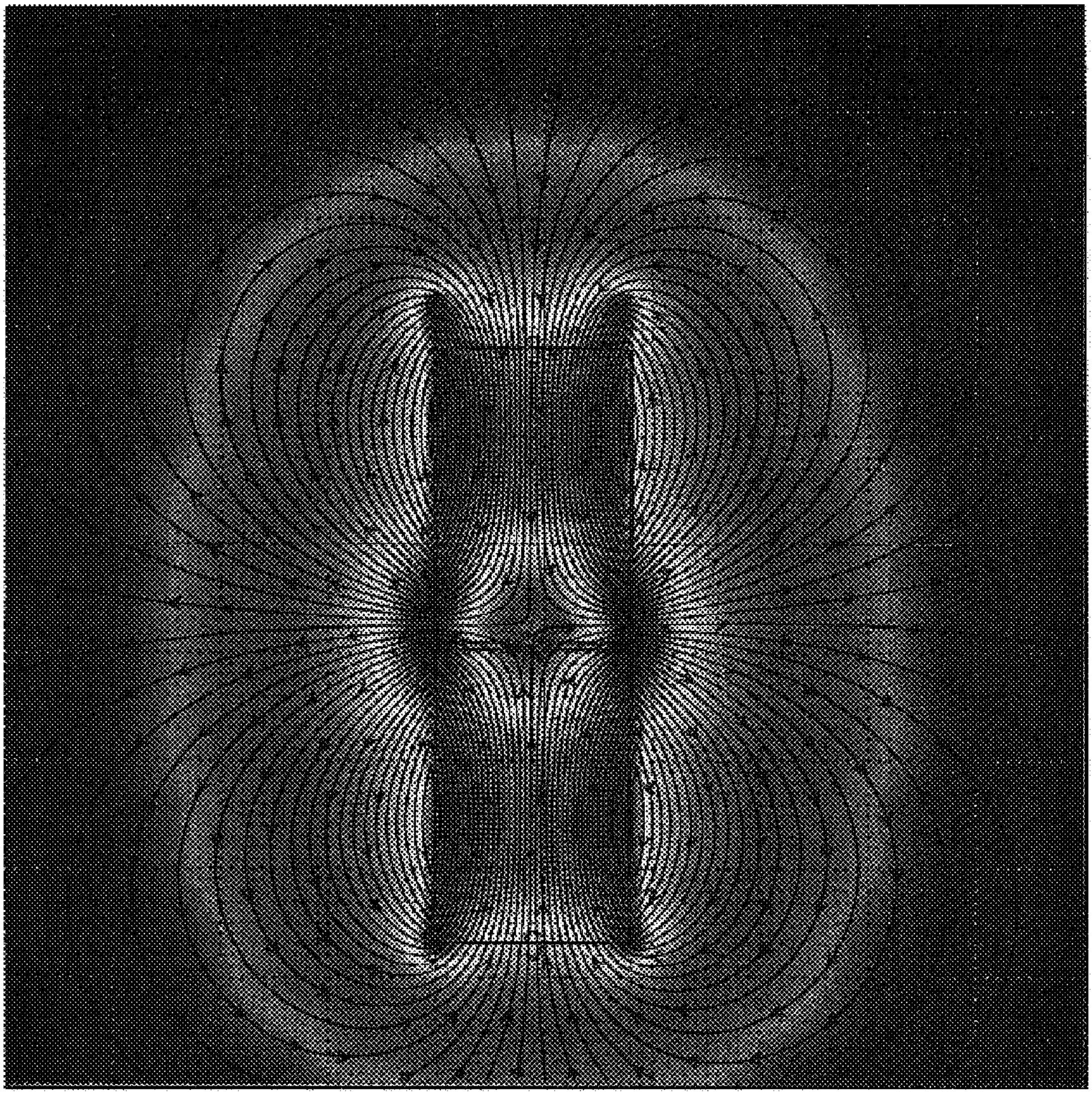

FIG. 14B

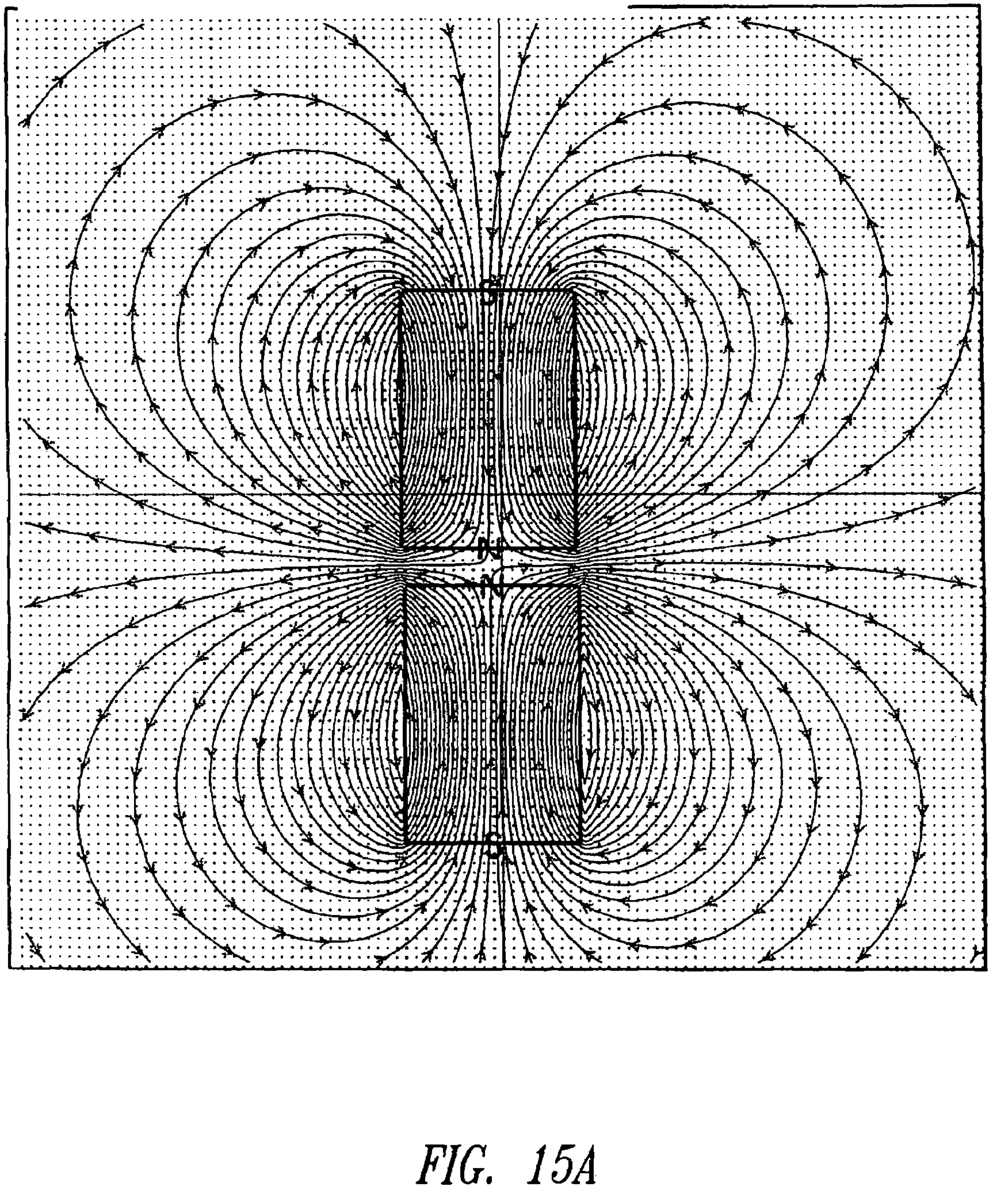

FIG. 15A

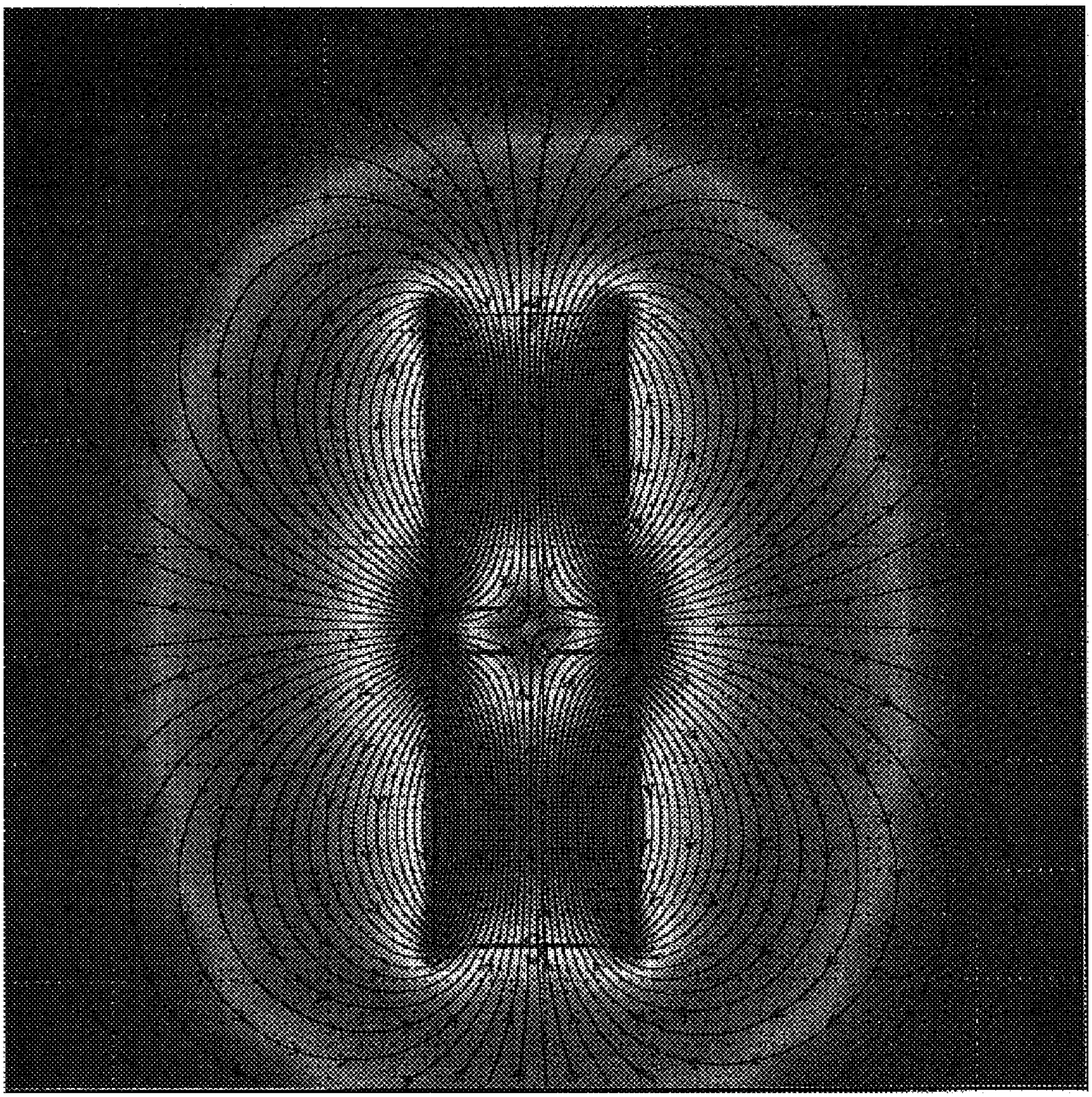

FIG. 15B

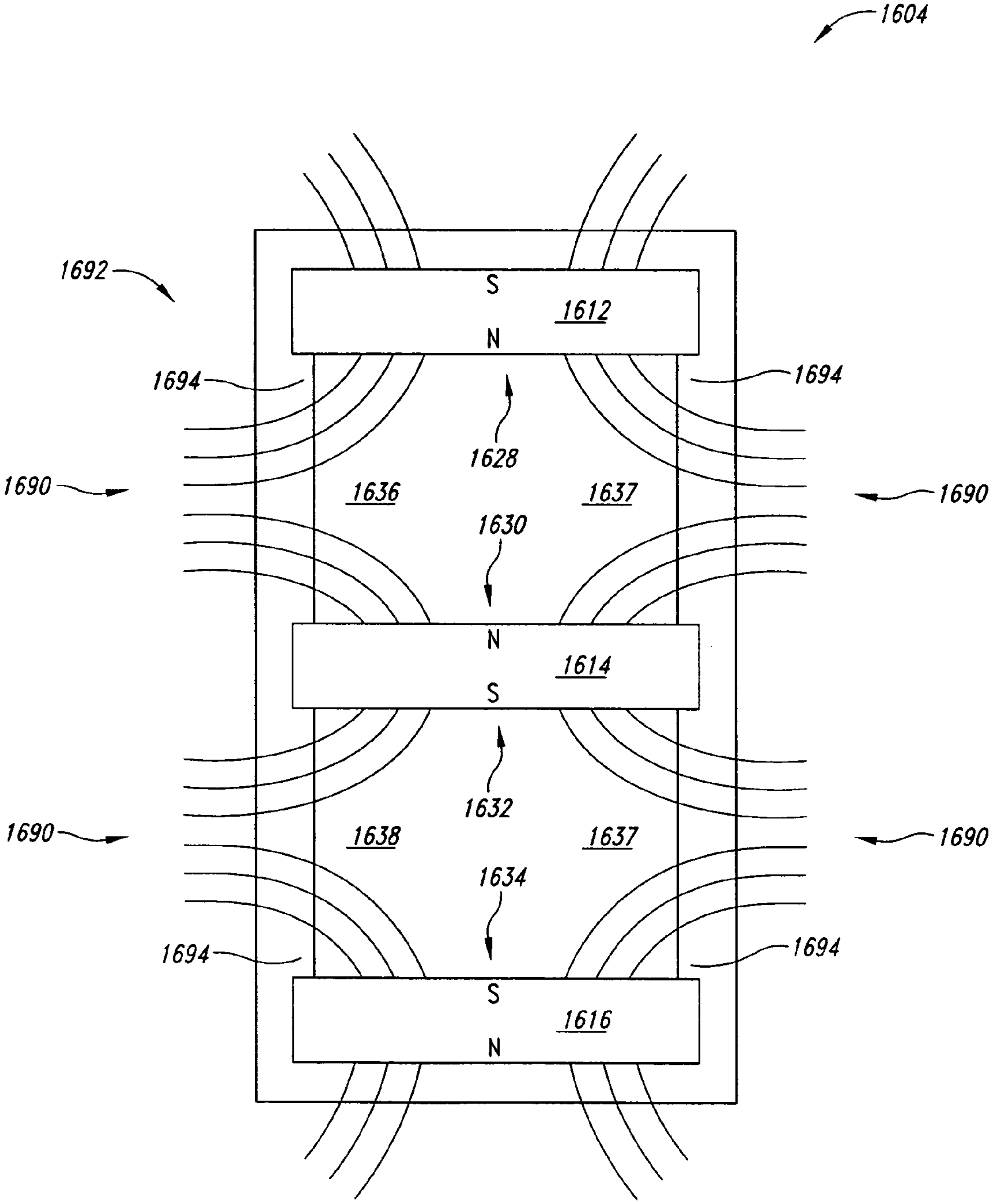
1604
1692
S
1612
N
1694
1694
1628
1690
1636
1637
1690
1630
N
1614
S
1632
1690
1638
1637
1690
1634
1694
1694
S
1616
N

FIG. 16

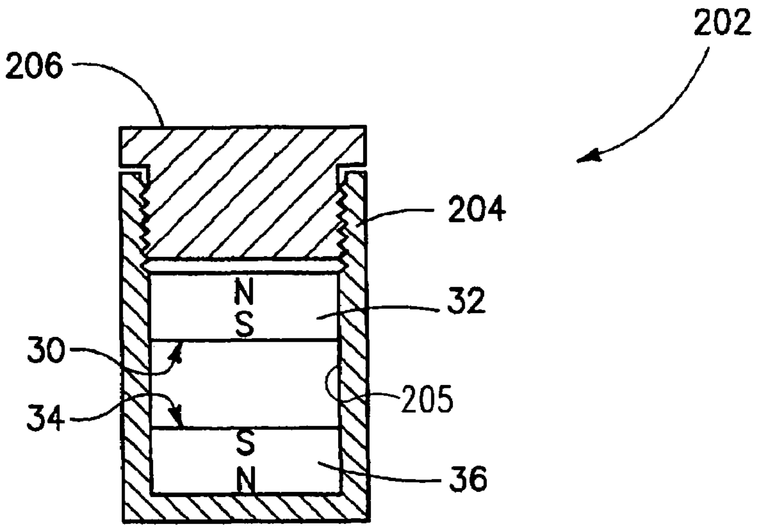
N
S
S
N

FIG. 17

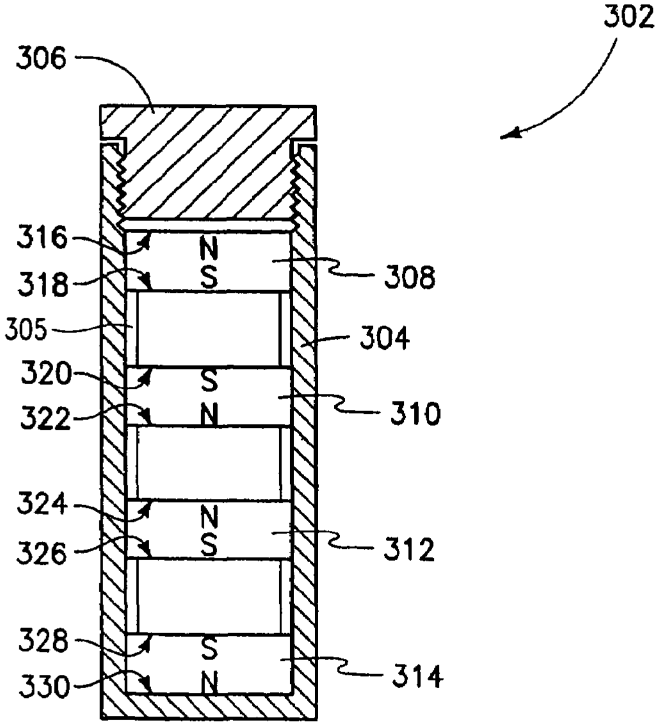
N
S
S
N
N
S
S
N

FIG. 18

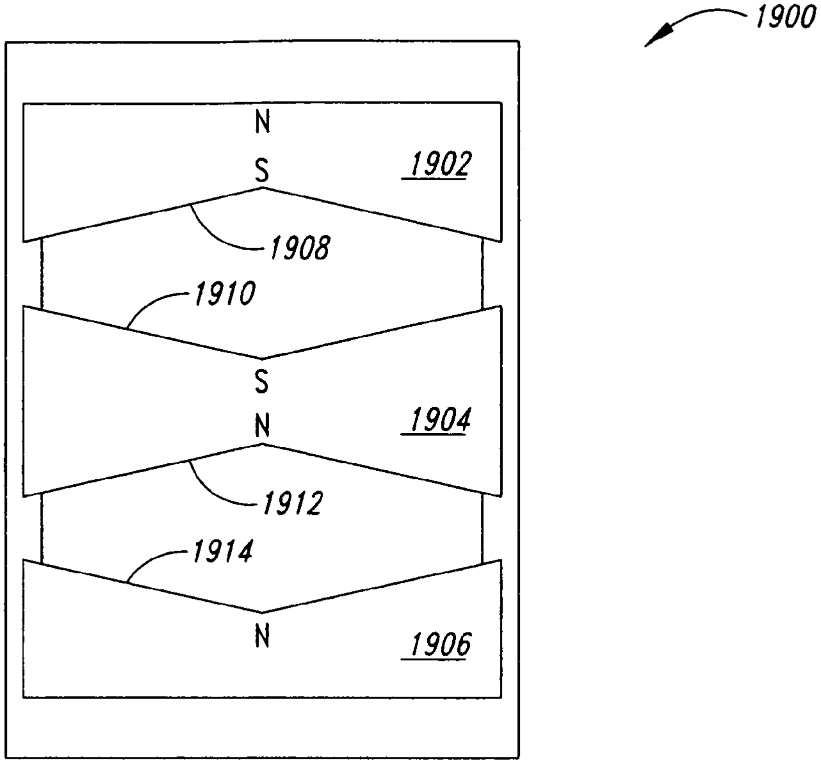
1900
N
S
1902
1908
1910
S
N
1904
1912
1914
N
1906

FIG. 19

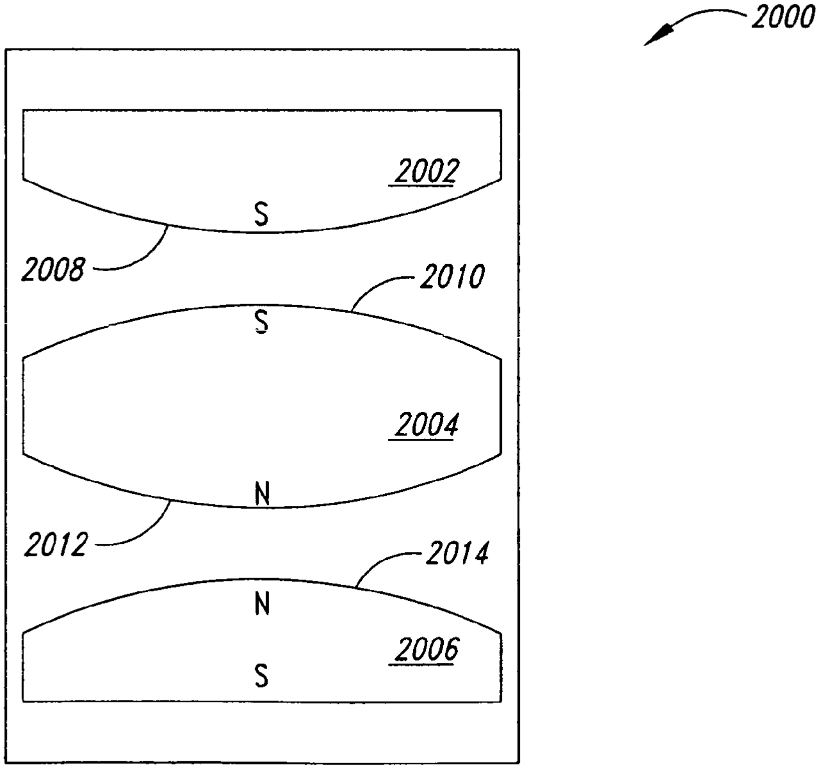
2000
2002
S
2008
2010
S
2004
N
2012
2014
N
2006
S

FIG. 20

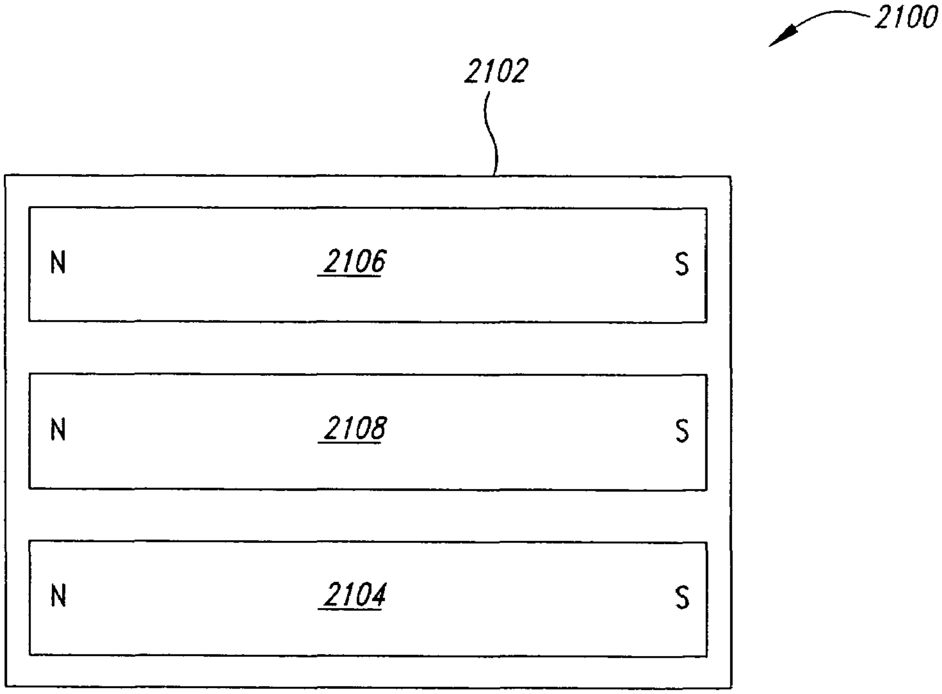
2100
2102
N
2106
S
N
2108
S
N
2104
S

FIG. 21

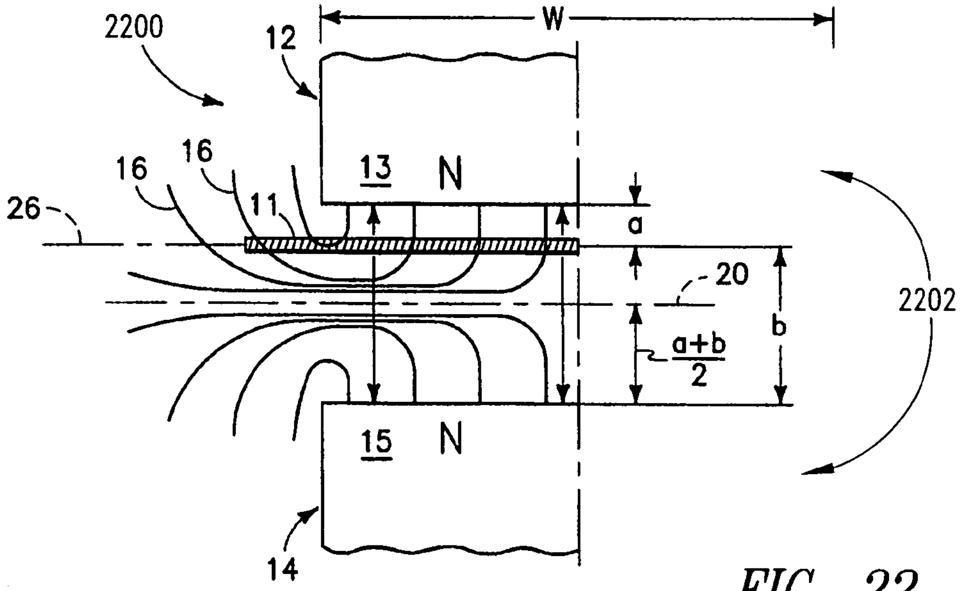
2200
12
W
16
16
11
13
N
26
a
20
b
2202
$\frac{a+b}{2}$
15
N
14

FIG. 22

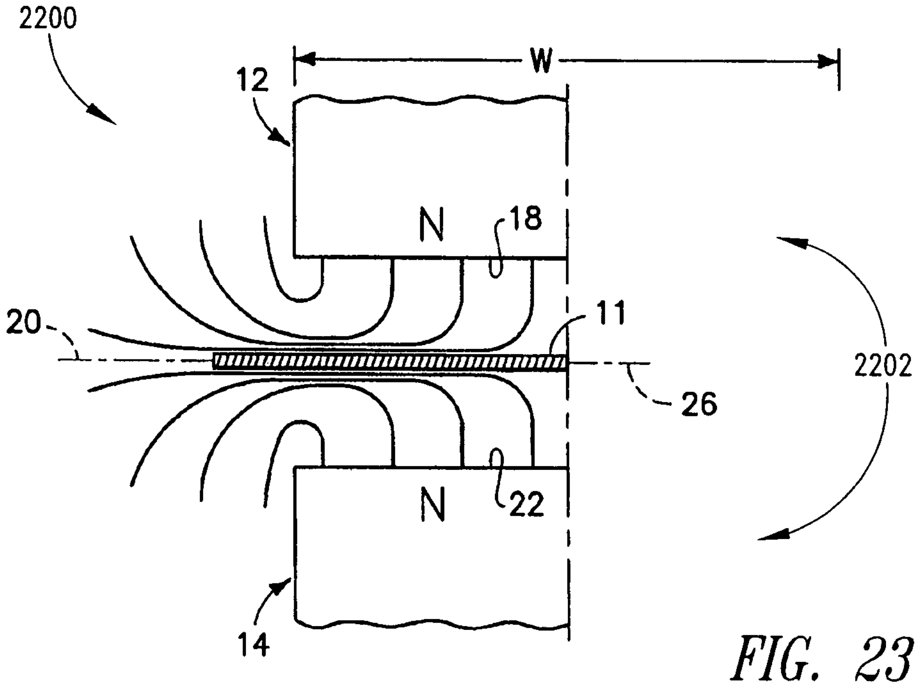
2200
12
W
N
18
20
11
26
2202
N
22
14

FIG. 23

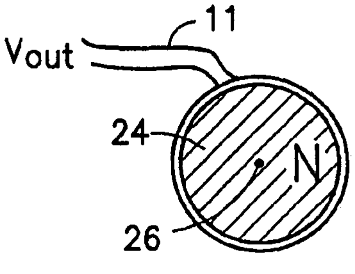
Vout
11
24
N
26

FIG. 24

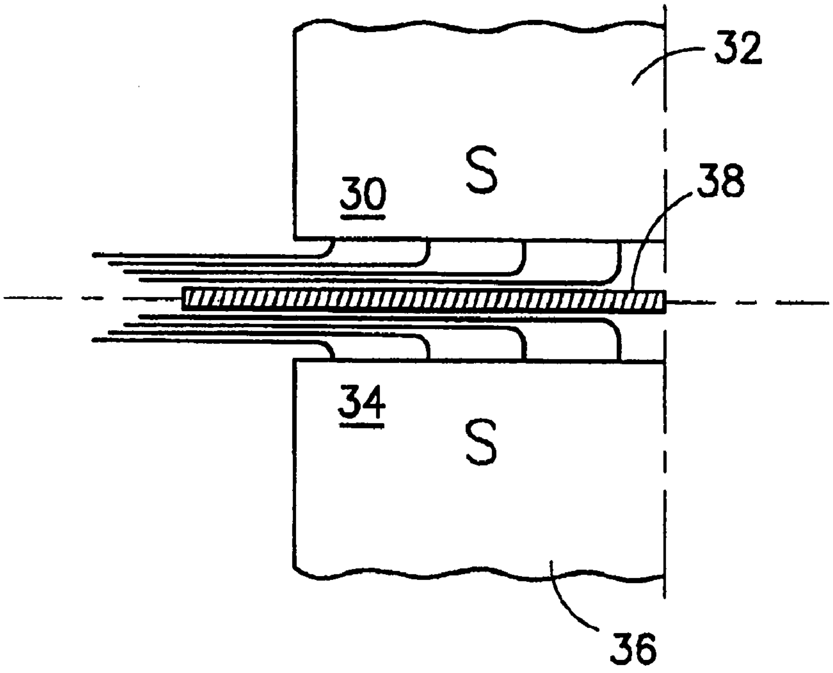
32
S
30
38
34
S
36

FIG. 25

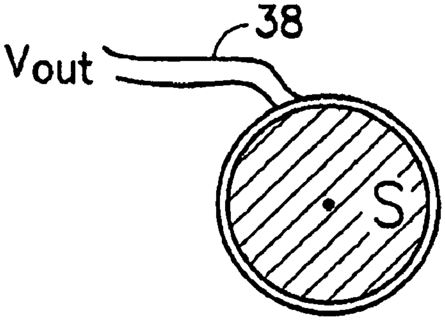

FIG. 26

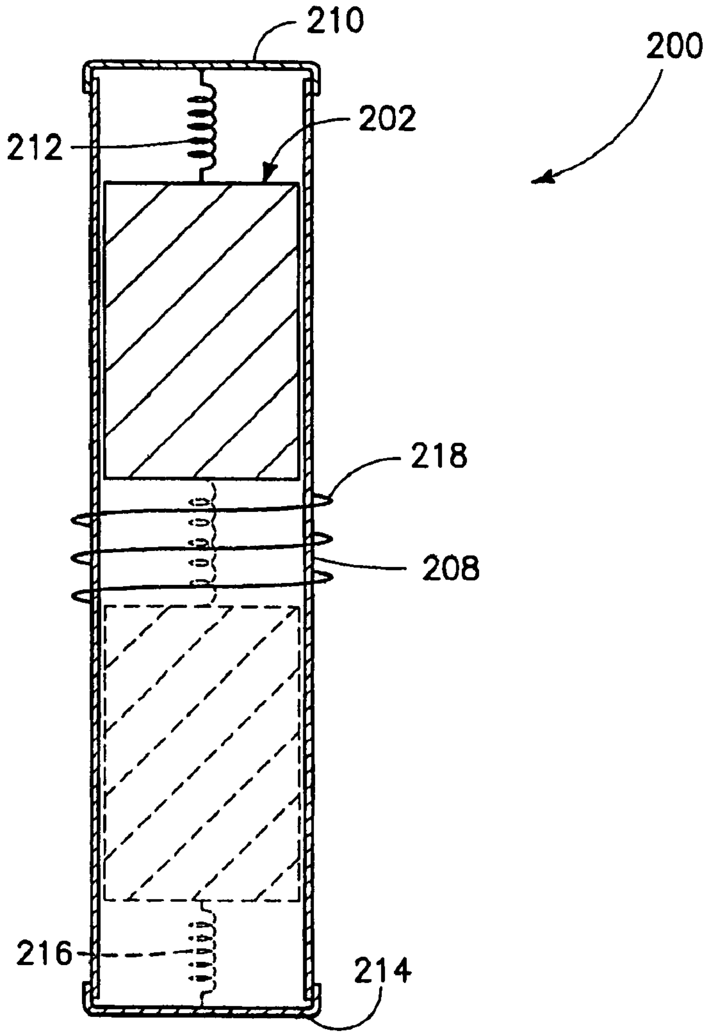
210
200
202
212
218
208
216
214

FIG. 27

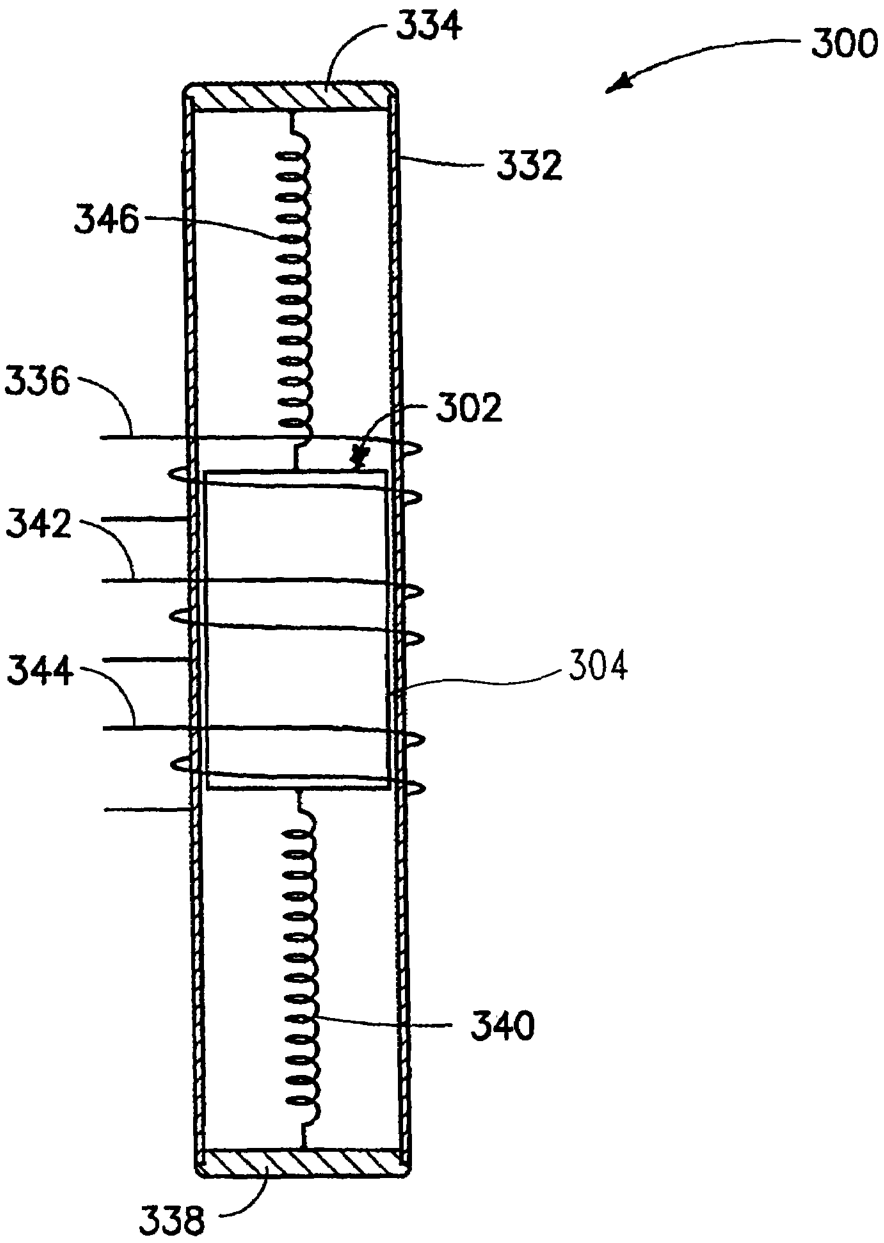
334
300
332
346
336
302
342
344
304
340
338

FIG. 28

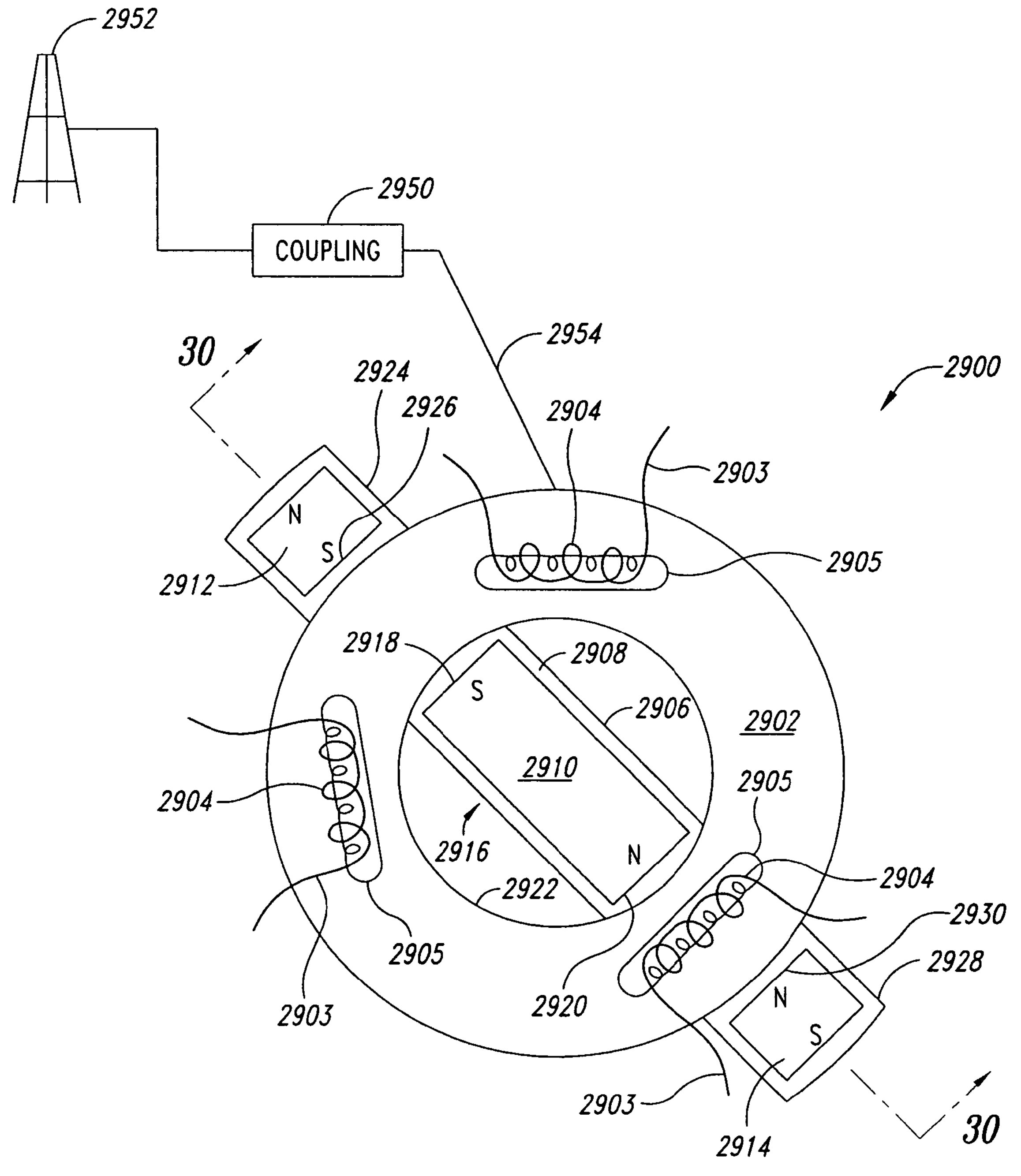
COUPLING
N
S
S
N
N
S

FIG. 29

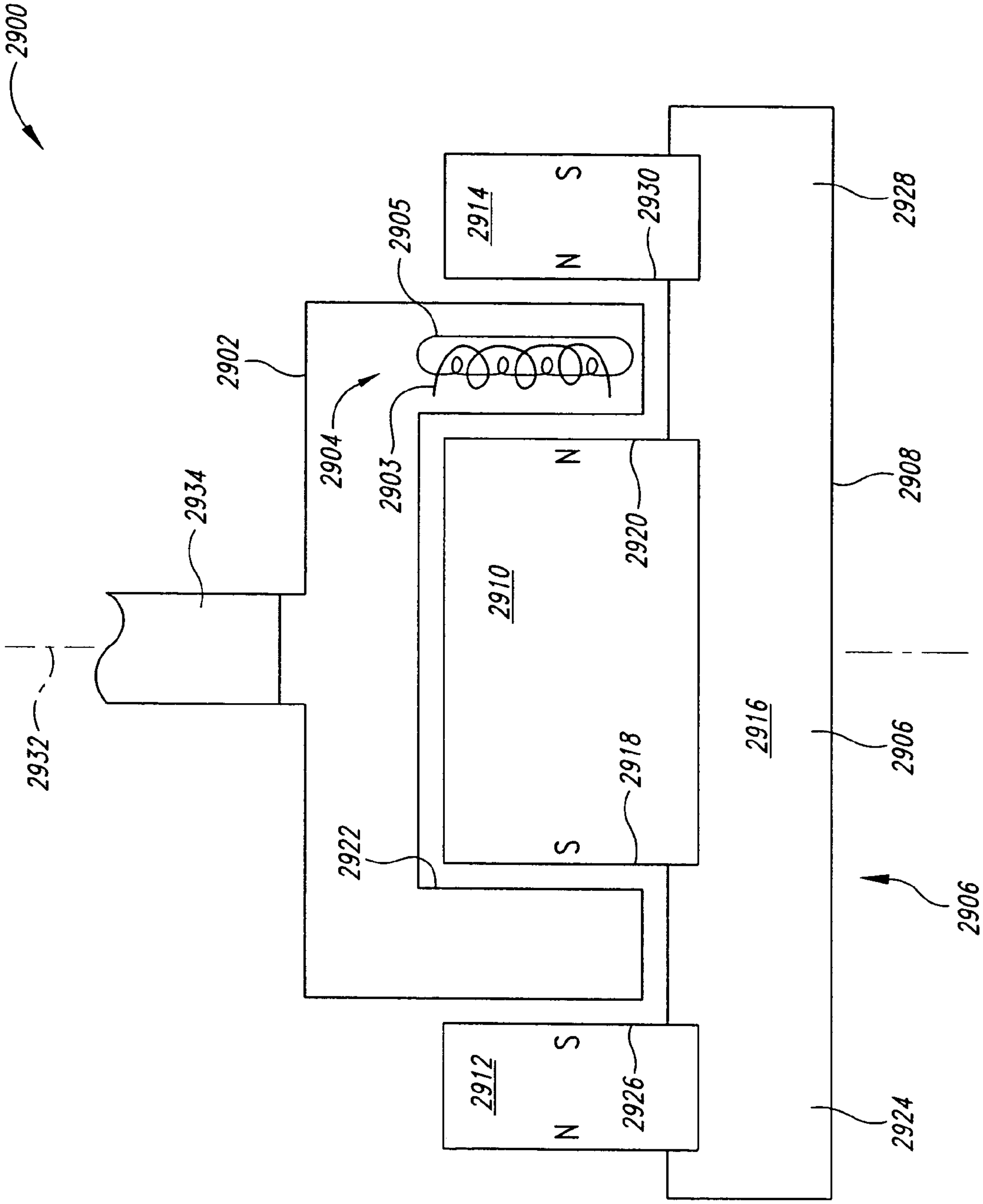
N
S
N
S
S
N

FIG. 30

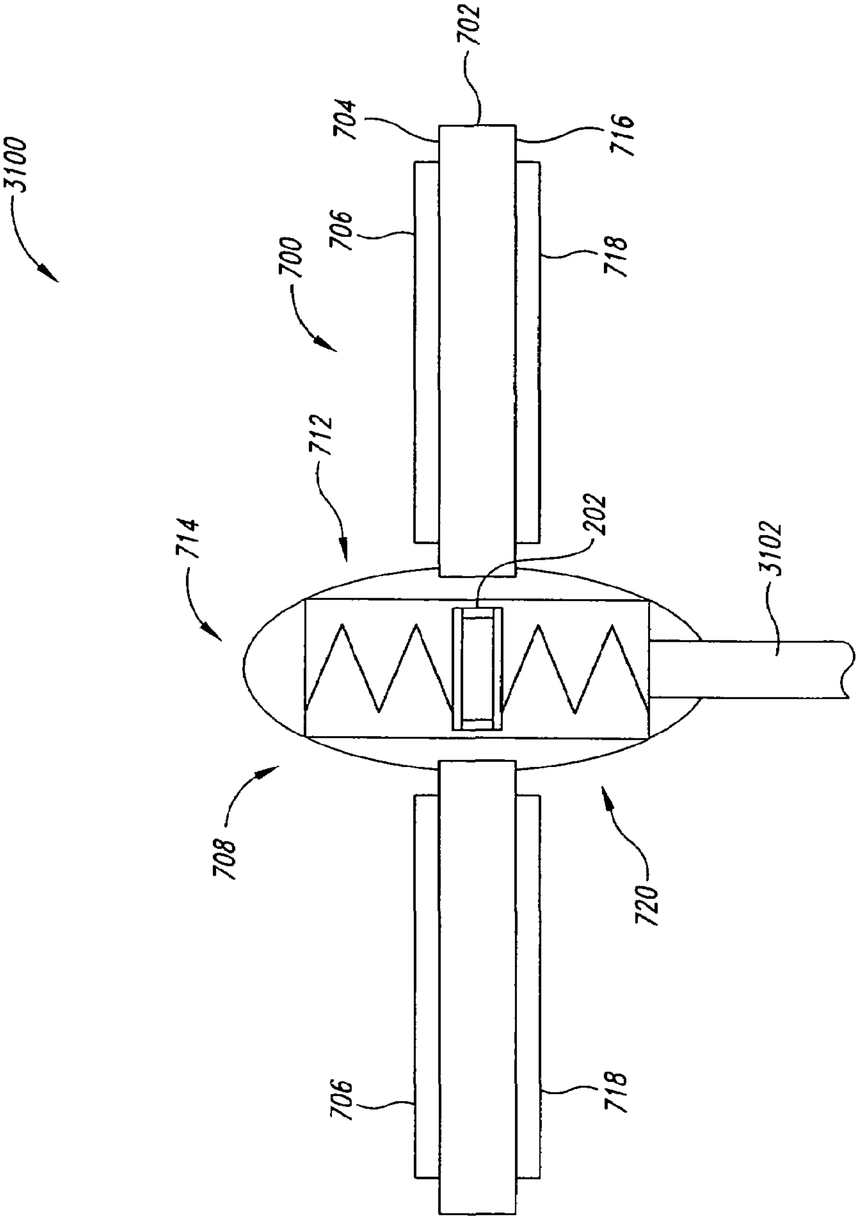
3100
702
704
716
706
718
700
712
714
202
3102
708
720
706
718

FIG. 31

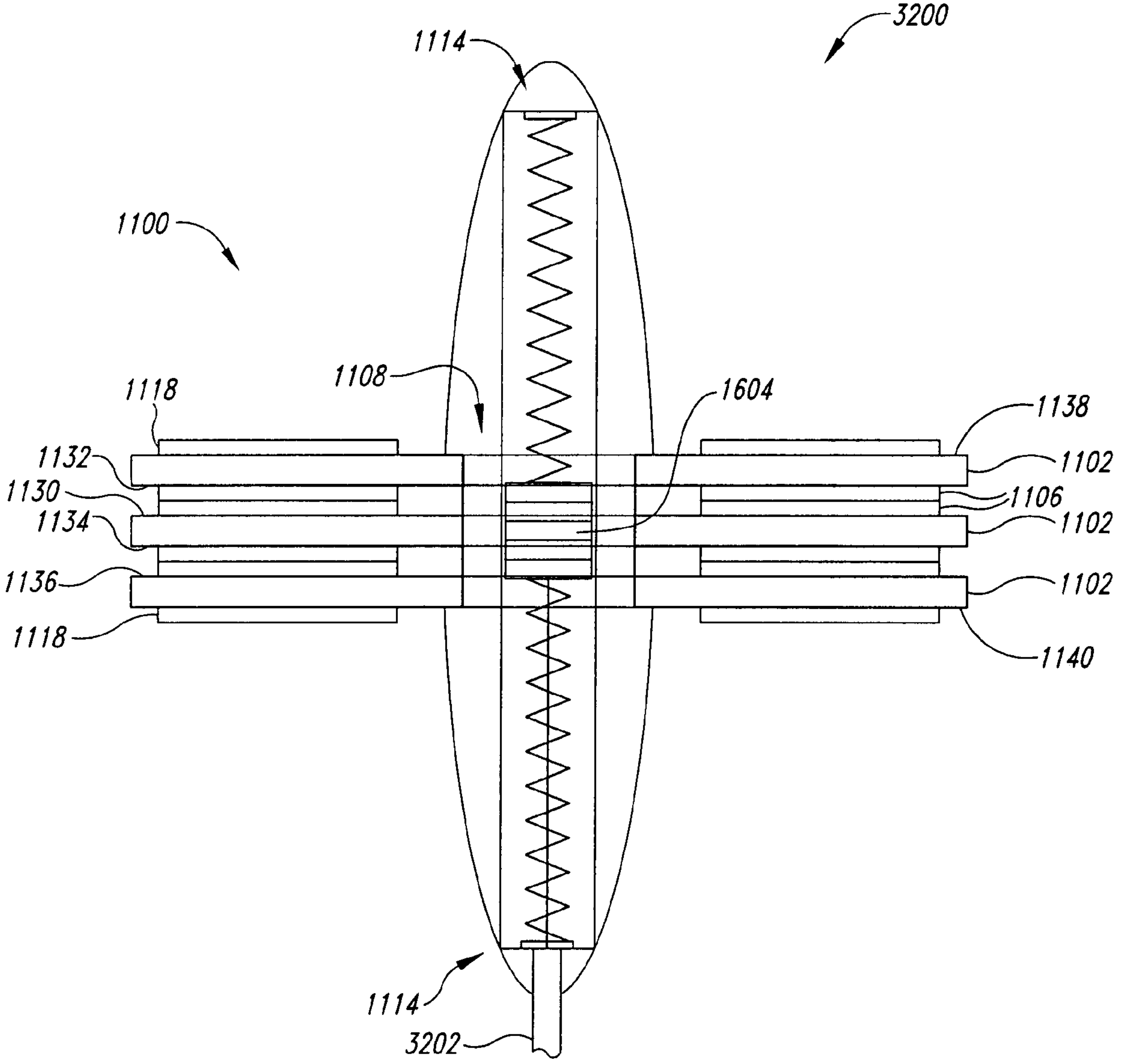
1604

FIG. 32

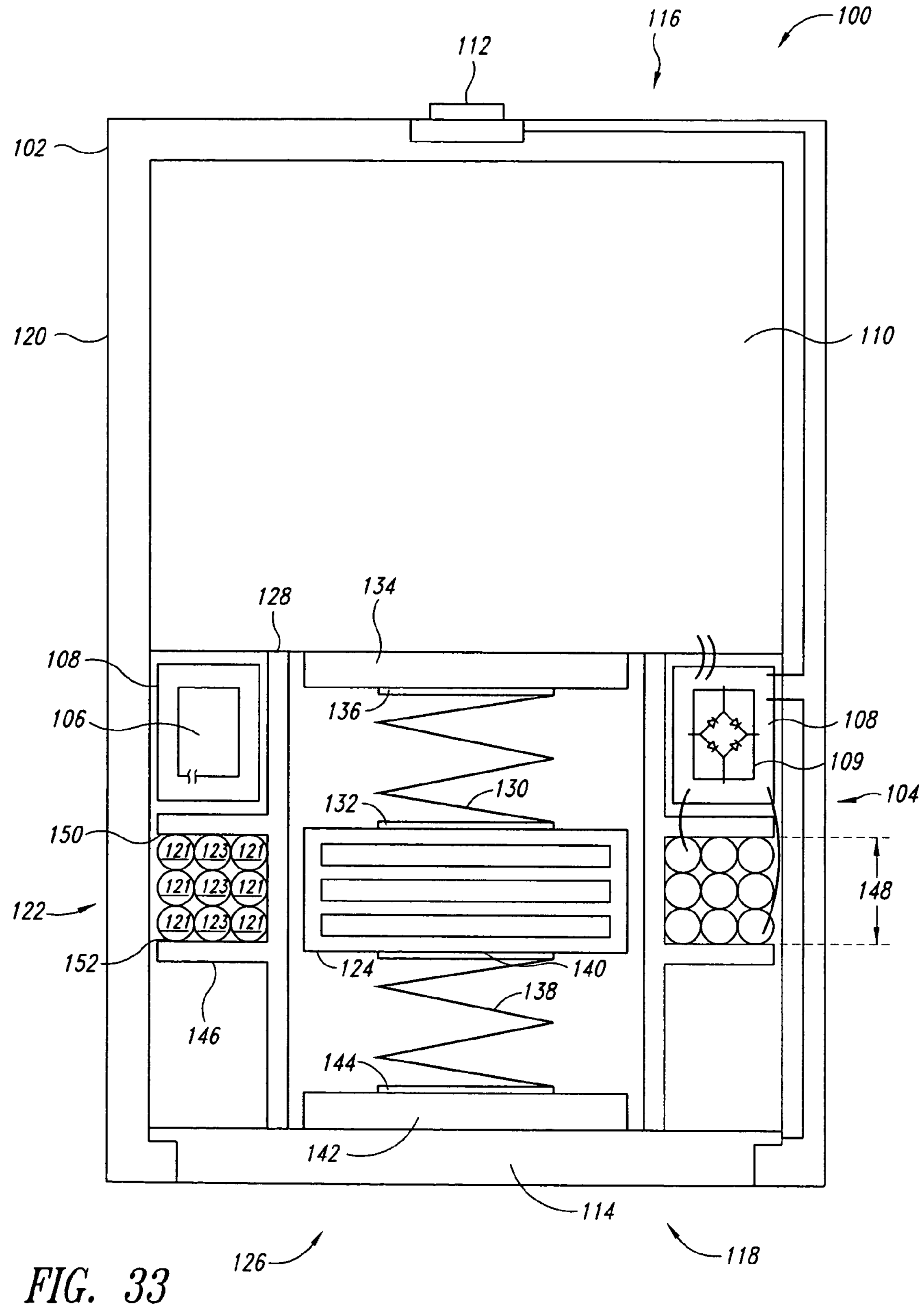

FIG. 33

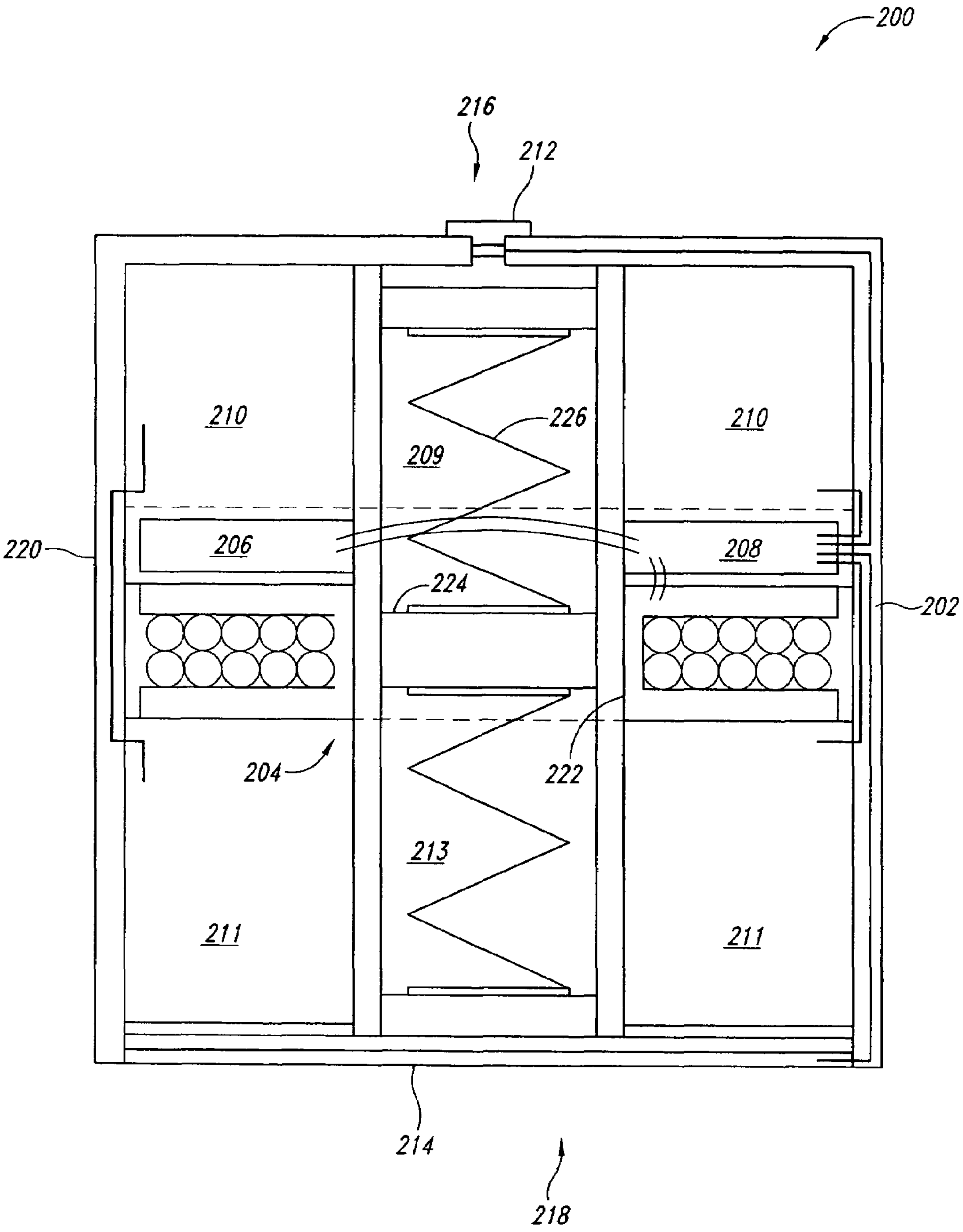

FIG. 34

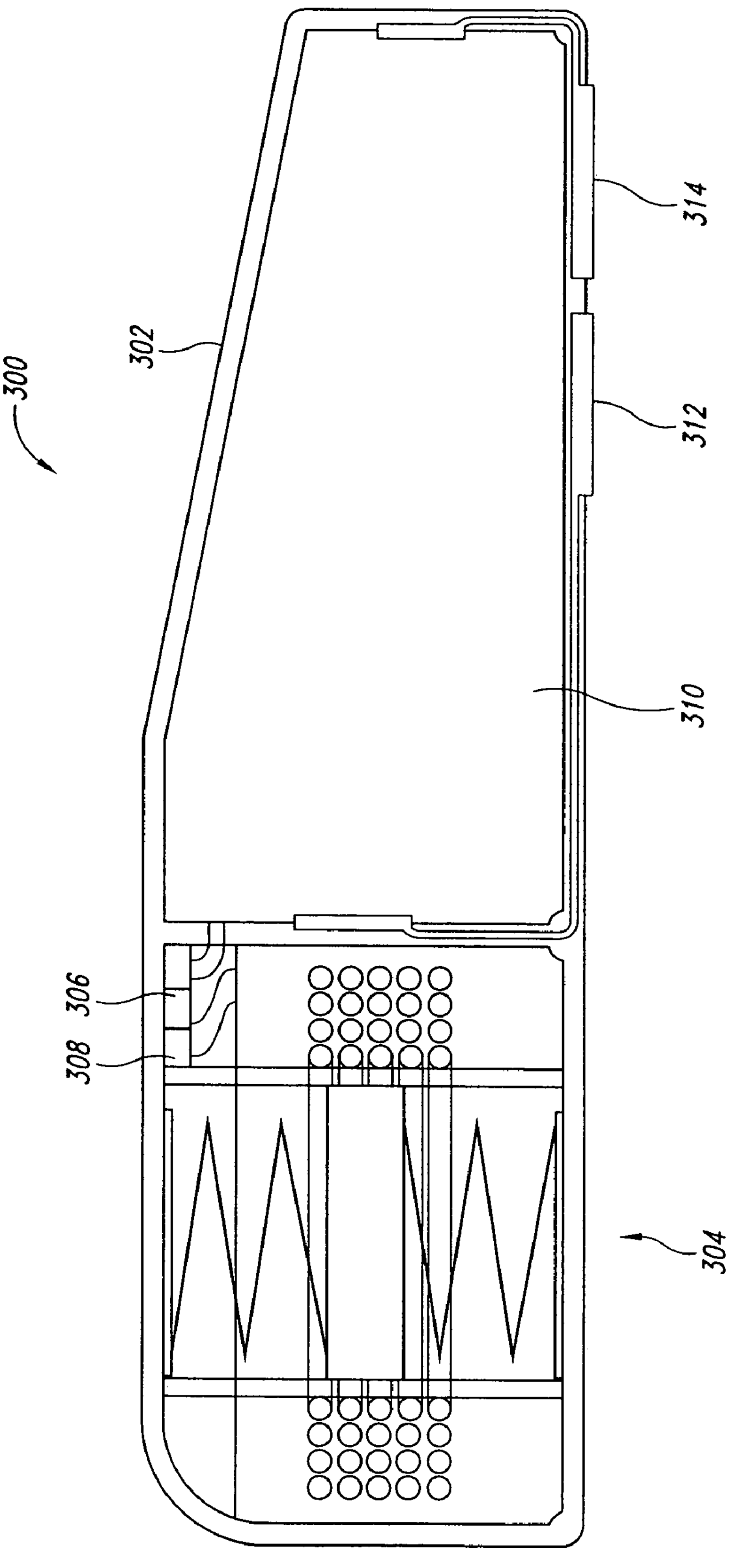
300
302
304
306
308
310
312
314

FIG. 35

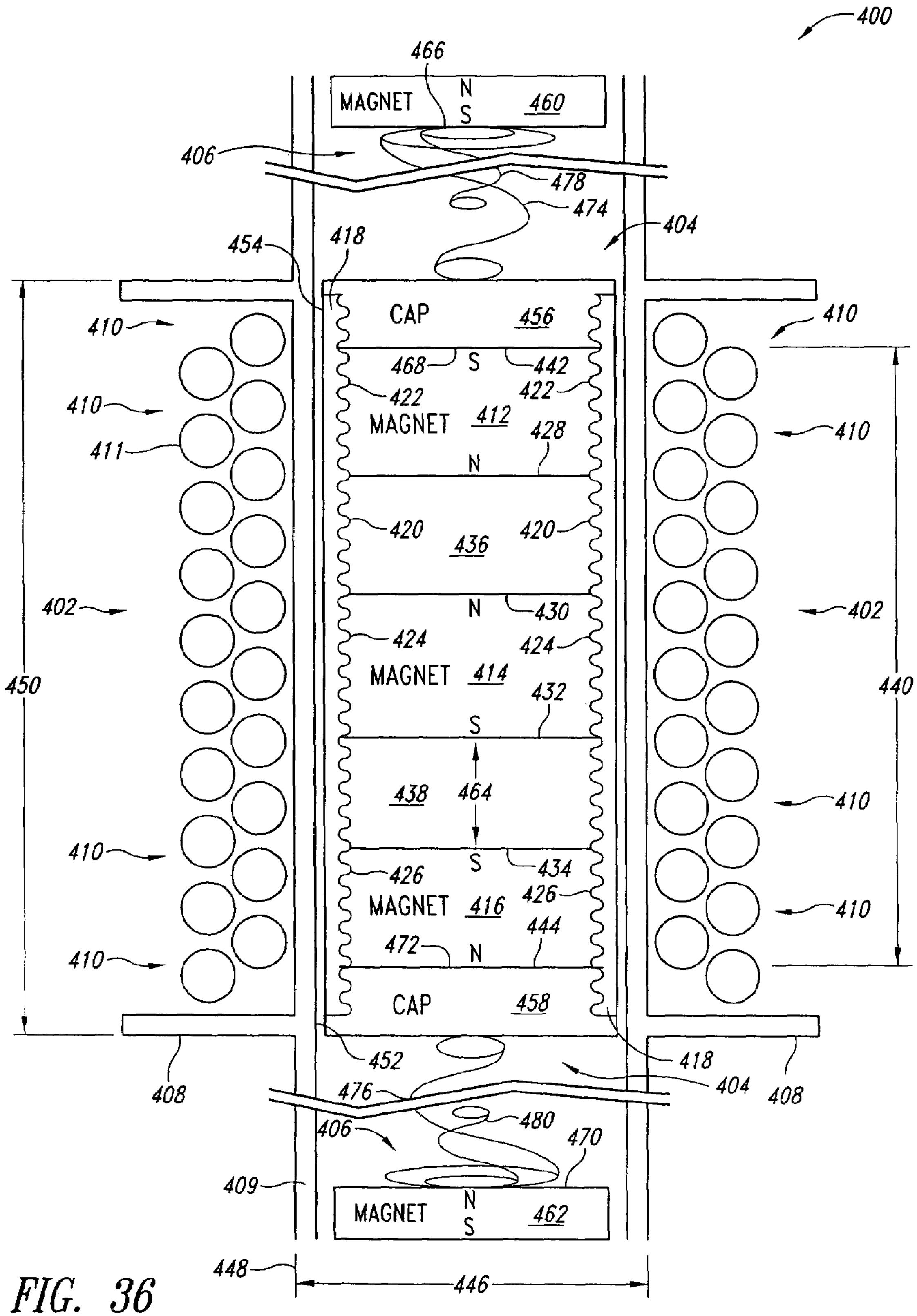
MAGNET
N
S
CAP
S
MAGNET
N
N
MAGNET
S
S
MAGNET
N
CAP
MAGNET
N
S

FIG. 36

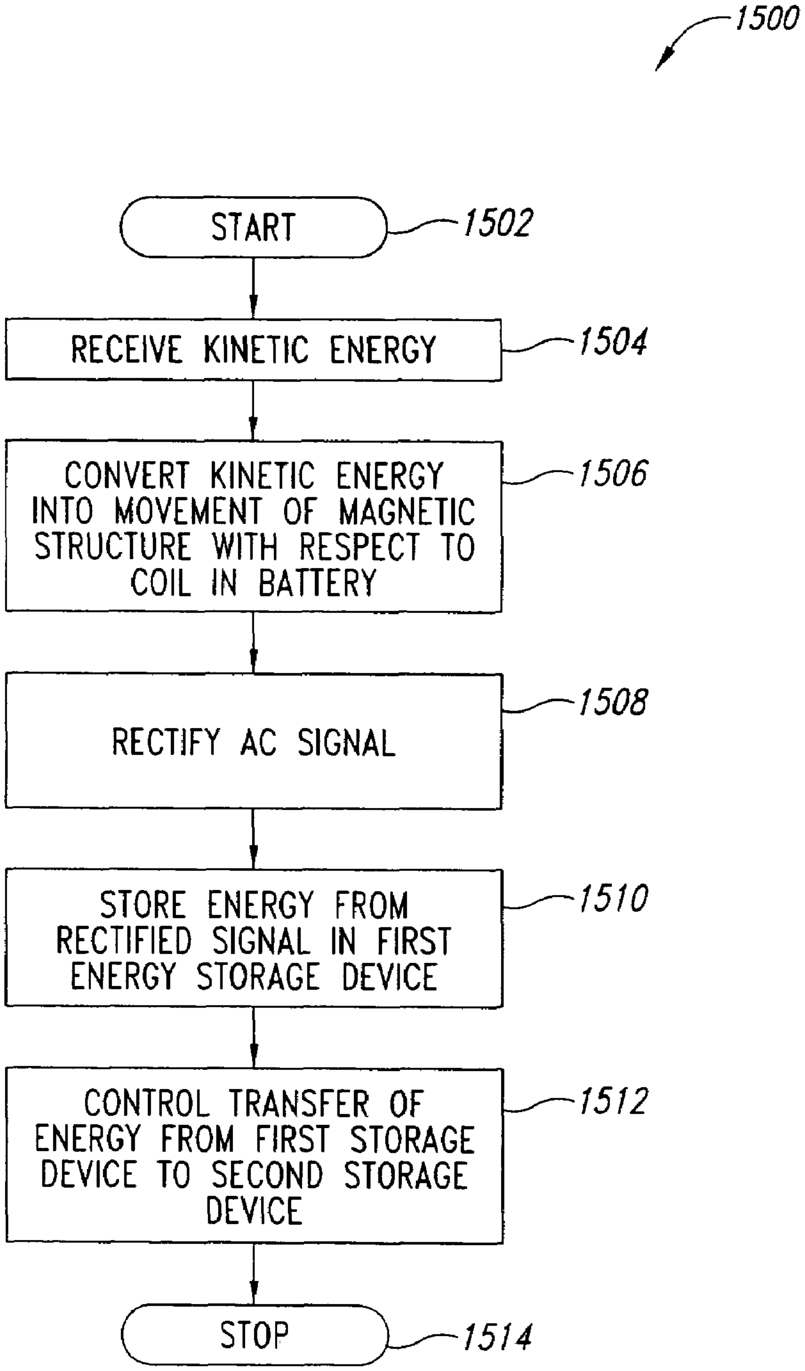
1500
START
1502
RECEIVE KINETIC ENERGY
1504
CONVERT KINETIC ENERGY INTO MOVEMENT OF MAGNETIC STRUCTURE WITH RESPECT TO COIL IN BATTERY
1506
RECTIFY AC SIGNAL
1508
STORE ENERGY FROM RECTIFIED SIGNAL IN FIRST ENERGY STORAGE DEVICE
1510
CONTROL TRANSFER OF ENERGY FROM FIRST STORAGE DEVICE TO SECOND STORAGE DEVICE
1512
STOP
1514

FIG. 37

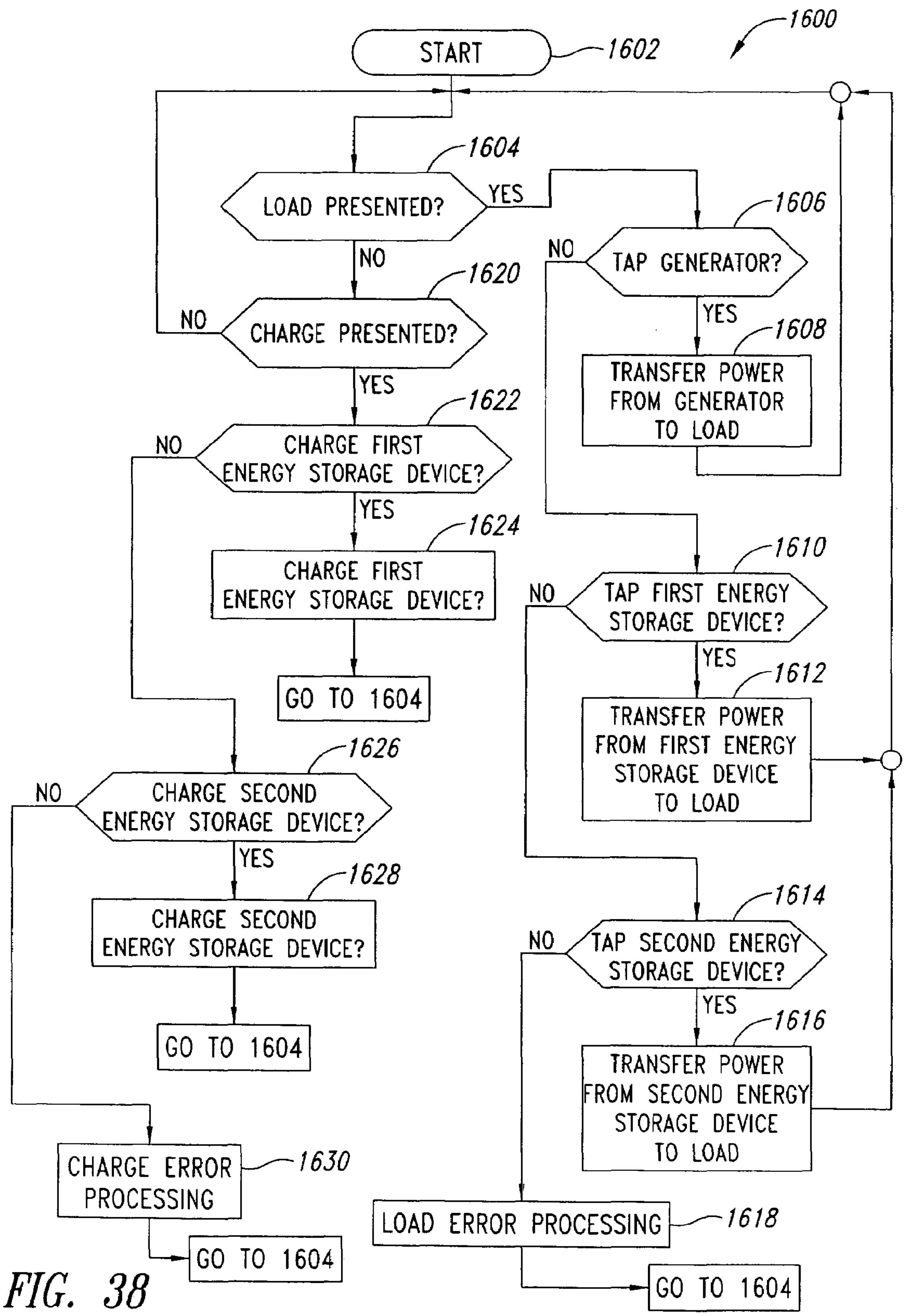
1600
START
1602
1604
LOAD PRESENTED?
YES
NO
1606
NO
TAP GENERATOR?
YES
1608
TRANSFER POWER FROM GENERATOR TO LOAD
1620
NO
CHARGE PRESENTED?
YES
1622
NO
CHARGE FIRST ENERGY STORAGE DEVICE?
YES
1624
CHARGE FIRST ENERGY STORAGE DEVICE?
GO TO 1604
1610
NO
TAP FIRST ENERGY STORAGE DEVICE?
YES
1612
TRANSFER POWER FROM FIRST ENERGY STORAGE DEVICE TO LOAD
1626
NO
CHARGE SECOND ENERGY STORAGE DEVICE?
YES
1628
CHARGE SECOND ENERGY STORAGE DEVICE?
GO TO 1604
1614
NO
TAP SECOND ENERGY STORAGE DEVICE?
YES
1616
TRANSFER POWER FROM SECOND ENERGY STORAGE DEVICE TO LOAD
CHARGE ERROR PROCESSING
1630
GO TO 1604
LOAD ERROR PROCESSING
1618
GO TO 1604

FIG. 38

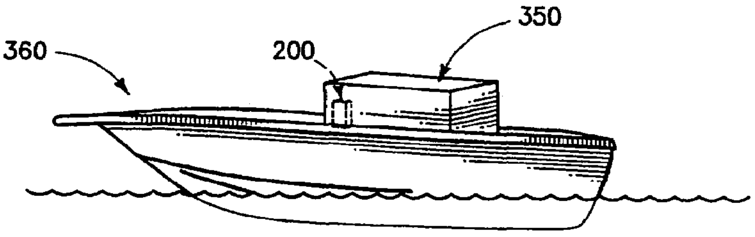

FIG. 39

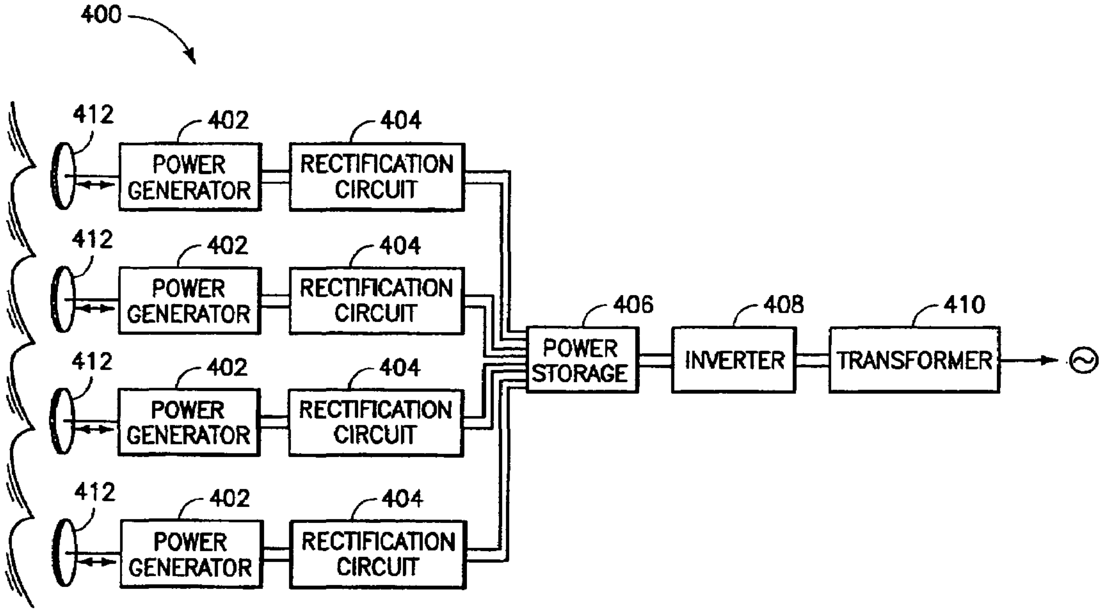
400
412
402
POWER GENERATOR
404
RECTIFICATION CIRCUIT
412
402
POWER GENERATOR
404
RECTIFICATION CIRCUIT
412
402
POWER GENERATOR
404
RECTIFICATION CIRCUIT
412
402
POWER GENERATOR
404
RECTIFICATION CIRCUIT
406
POWER STORAGE
408
INVERTER
410
TRANSFORMER

FIG. 40

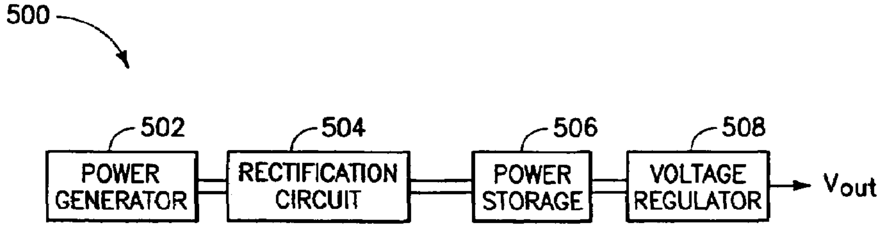
500
502
POWER GENERATOR
504
RECTIFICATION CIRCUIT
506
POWER STORAGE
508
VOLTAGE REGULATOR
Vout

FIG. 41

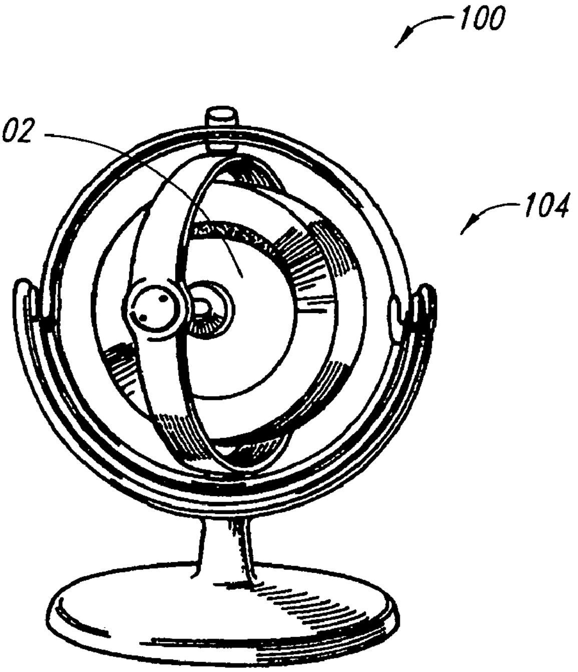

FIG. 42

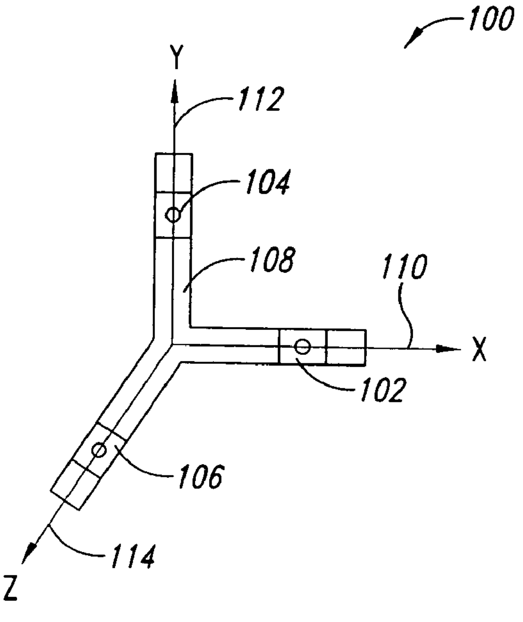
100
Y
112
104
108
110
X
102
106
114
Z

FIG. 43

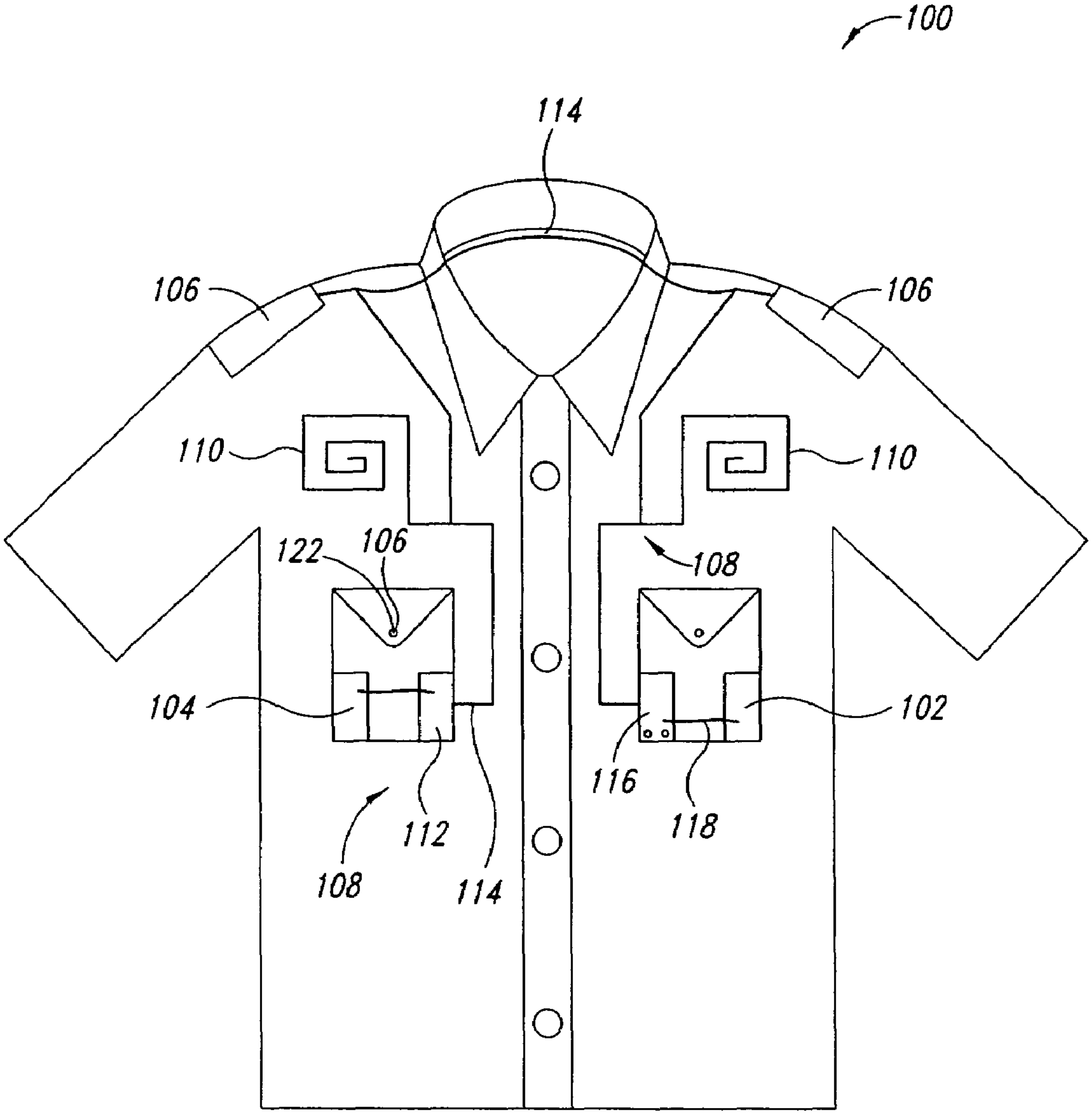
100
114
106
106
110
110
122
106
108
104
102
116
118
112
108
114

FIG. 44

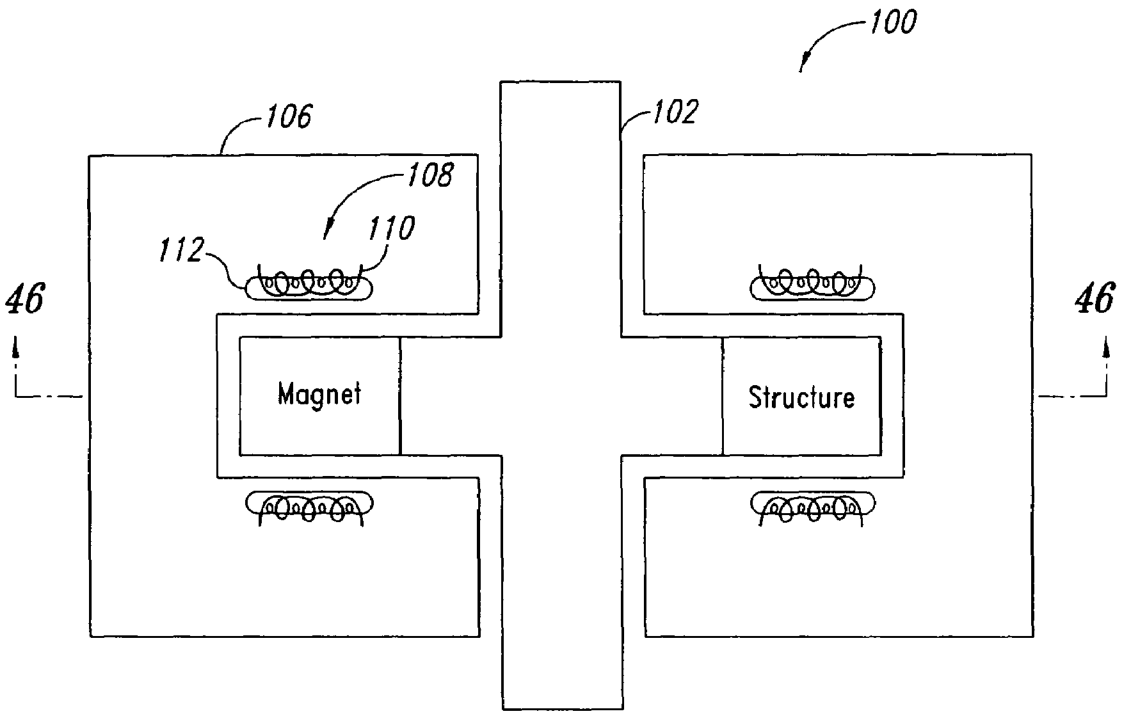
Magnet
Structure

FIG. 45

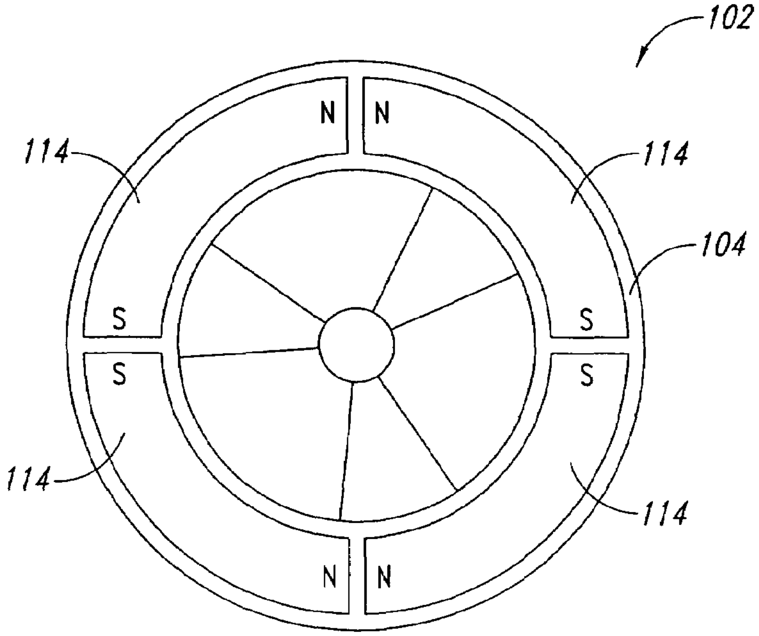
102
114
114
104
N
N
S
S
S
S
N
N
114
114

FIG. 46

SYSTEM AND METHOD FOR STORING ENERGY

STATEMENT REGARDING GOVERNMENT INTEREST

This invention was made with United States Government support under Contract No. DE-AC07-05-ID14517 awarded by the United States Department of Energy. The United States Government has certain rights in the invention.

BACKGROUND OF THE INVENTION

1. Field of the Invention

This disclosure generally relates to a system and method for storing energy and more particularly to a system and method for portably converting energy into an electrical current and storing the energy.

2. Description of the Related Art

Conventional portable energy storage devices, such as conventional batteries and storage capacitors, are either disposable or rechargeable by coupling the device to a remote source of electrical energy. Disposable devices inherently limit the amount of portable stored energy available to a user. Carrying extra disposable devices is expensive, and a user must accommodate associated weight, storage space, and disposal requirements. They also are environmentally unfriendly. Conventional rechargeable devices are more environmentally friendly than non-rechargeable devices, but a remote source of energy is required to recharge the devices and the energy available between accesses to the remote source of energy is limited. In addition, recharging requires user intervention.

Conventional shake-flashlights and similar shake-activated devices provide for storing limited amounts of energy, enough to power, for example, an LED for a brief period of time, but typically do not produce enough energy to power a conventional flashlight light bulb, and other high current draw devices, for example, a cell-phone, a camera, a GPS system or a conventional flashlight. They also are bulky and require physical activity expressly directed to charging the device. In addition, conventional shake-activated devices are not readily deployable to power another portable device. They also produce objectionable levels of magnetic fields that can potentially interfere with the operation of electronic devices, such as cell-phones and health-related devices, such as pacemakers.

Conventional crank-powered devices have much more energy-storage capabilities than conventional shake-activated devices, but similarly are bulky, require physical activity expressly directed to recharging the device, are not readily deployable to power another portable device, and may generate objectionable levels of magnetic fields.

Electro-magnetic and electro-mechanical devices and applications, such as, for example, motors, generators and alternators, typically employ coils and/or magnets. Conventional magnetic structures employ a single magnet to generate a magnetic field, or a plurality of magnets arranged to generate a magnetic field. The magnets are typically permanent magnets or electromagnets. The efficiency of many applications is dependant on the gradient of the magnetic field generated by the magnetic structure.

When an increase in output or performance was desired, conventionally the size or number of coils was increased or the size or strength of the magnets would be increased. These approaches introduce weight, cost, size and durability issues. These approaches also are not practical for many applications. Therefore it can be appreciated that there is a need for improved coils and magnets for use in electro-magnetic and electromechanical devices and applications.

BRIEF SUMMARY OF THE INVENTION

In one embodiment, a coil comprises an electrical conductive winding and a magnetic conductive winding configured to focus magnetic flux in the electrical conductive winding. In one embodiment, the coil further comprises a winding form. In one embodiment, a first layer on the winding form comprises a layer of the electrical conductive winding. In one embodiment, a second layer of the electrical conductive winding is adjacent to the first layer on the winding form. In one embodiment, a layer of the magnetic conductive winding is adjacent to the second layer of the electrical conductive winding. In one embodiment, a layer of the magnetic conductive winding is adjacent to the first layer on the winding form. In one embodiment, a last layer on the winding form comprises a layer of the electrical conductive winding. In one embodiment, a last layer on the winding form comprises a layer of the electrical conductive winding. In one embodiment, a layer of the magnetic conductive winding is between two layers of the electric conductive winding. In one embodiment, a plurality of layers of the magnetic conductive winding is between two layers of the electric conductive winding. In one embodiment, the magnetic conductive winding forms a closed loop. In one embodiment, the coil has a trapezoidal-shaped portion. In one embodiment, the coil is wound around a core. In one embodiment, the electrical conductive winding and the magnetic conductive winding together comprise a dual-conductor winding. In one embodiment, the magnetic conductive winding comprises a silver/nickel alloy. In one embodiment, the coil further comprises a layer of insulating material, wherein the electrical conductive winding comprises a trace formed on the layer of insulation material.

In one embodiment, a winding comprises an electrical conductive wire and a magnetic conductive wire insulated from and secured to the electrical conductive wire and configured to focus magnetic flux in the electrical conductive wire. In one embodiment, the magnetic conductive wire forms a closed loop. In one embodiment, the magnetic conductive wire is secured to the magnetic conductive wire by an insulating material. In one embodiment, the magnetic conductive wire forms a core of the winding and is surrounded by an insulating layer and the electrical conductive wire surrounds the insulating layer. In one embodiment, the electrical conductive wire comprises a stranded wire.

In one embodiment, a system comprises a magnetic structure and a coil, the coil comprising an electrical conductive winding and a magnetic conductive winding configured to focus magnetic flux in the electrical conductive winding. In one embodiment, the system is configured to receive energy and to generate an electrical signal in response to the receipt of the energy. In one embodiment, the system further comprises a mechanical transmission system configured to receive the energy. In one embodiment, the mechanical transmission system is coupled to the magnetic structure and configured to move the magnet structure with respect to the coil in response to the receipt of energy. In one embodiment, the mechanical transmission system is configured to move the magnetic structure in a linear manner. In one embodiment, the mechanical transmission system is configured to move the magnetic structure in a rotational manner. In one embodiment, the mechanical transmission system is configured to move the magnetic structure in a radial manner. In one embodiment, the mechanical transmission system is coupled to the coil and configured to move the coil with respect to the magnetic structure in response to the receipt of energy. In one embodiment, the coil is configured to receive an electrical signal and the system is configured to generate mechanical force in response to the receipt of the electrical signal. In one embodiment, the system further comprises a mechanical transmission system.

In one embodiment, a system comprises a coil comprising means for conducting an electric signal and means for focusing magnetic flux in the means for conducting an electrical signal, and a magnetic structure. In one embodiment, the means for focusing magnetic flux comprises a winding comprising a silver/nickel alloy. In one embodiment, the means for conducting an electrical signal comprises a stranded copper wire. In one embodiment, the coil further comprises a first insulating substrate and the means for conducting an electrical signal comprises an electrical conductive trace formed on the first insulating substrate. In one embodiment, the means for focusing magnetic flux comprises a magnetic conductive trace formed on the first insulating substrate. In one embodiment, the electrical conductive trace is formed on a first surface of the first insulating substrate and the magnetic conductive trace is formed on the first surface of the first insulating substrate. In one embodiment, the coil further comprises a plurality of insulating substrates, and the means for conducting an electric signal comprises a plurality of electrical conductive traces formed on selected substrates in the plurality of substrates and the means for focusing magnetic flux comprises a plurality of magnetic conductive traces formed on selected substrates in the plurality of substrates.

In one embodiment, a method for generating an electrical signal comprises causing relative movement between a magnetic structure and an electrical conductive winding and focusing magnetic flux generated by the magnetic structure in the electrical conductive winding using a magnetic conductive winding. In one embodiment, the method further comprises forming a closed loop with the magnetic conductive winding.

In one embodiment, a coil comprises a plurality of insulating substrates, a plurality of electrical conductive traces formed on a first set of selected substrates in the plurality of substrates and a plurality of magnetic conductive traces formed on a second set of selected substrates in the plurality of substrates. In one embodiment, the first set of selected substrates comprises every other insulating substrate in the plurality of insulating substrates and the plurality of electrical conductive traces consists of an electrical conductive trace formed on each of the plurality of insulating substrates in the first set of selected substrates. In one embodiment, the plurality of electrical conductive traces are electrically coupled in series. In one embodiment, the plurality of magnetic conductive traces are electrically coupled together to form a closed loop.

In one embodiment, a method of generating mechanical force comprises generating a magnetic field, focusing the magnetic field in an electrical conductive element and conducting a current through the electrical conductive element. In one embodiment, the current is an alternating current. In one embodiment, the method further comprises applying the mechanical force so as to generate a linear movement in a transmission system. In one embodiment, the method further comprises applying the mechanical force so as to generate a rotational movement in a transmission system. In one embodiment, the method further comprises applying the mechanical force so as to generate a radial movement in a transmission system. In one embodiment, the current is a direct current. In one embodiment, the electrical conductive element comprises layers of an electrical conductive winding and focusing the magnetic field in the electrical conductive winding comprises inserting a magnetic conductive winding between two layers of the electrical conductive winding. In one embodiment, the magnetic conductive winding forms a closed loop.

In one embodiment, a system comprises a first magnet housing, a first magnet secured within the first magnet housing and having a first pole of a first polarity and a second pole of a second polarity, and a second magnet having a first pole of the first polarity and a second pole of the second polarity, secured within the first magnet housing such that the first pole of the second magnet is held spaced apart a distance from and generally facing the first pole of the first magnet, so as to generate a compressed magnetic field. In one embodiment, the first magnet comprises a rare earth magnet. In one embodiment, the system further comprises a coil. In one embodiment, the system is configured to receive energy and to generate an electrical signal in response to the receipt of the energy. In one embodiment, the system is configured to receive an electrical signal and to generate mechanical force in response to the electrical signal. In one embodiment, the system further comprises a mechanical transmission system. In one embodiment, the mechanical transmission system is coupled to the first magnet housing and configured to move the first magnet housing with respect to the coil in response to the receipt of the energy. In one embodiment, the mechanical transmission system is configured to move the first magnet housing in a linear manner. In one embodiment, the mechanical transmission system is configured to rotate the first magnet housing. In one embodiment, the system further comprises a third magnet having a first pole of the first polarity and a second pole of the second polarity, secured within the first magnet housing such that the second pole of the third magnet is held spaced apart a distance from and generally facing the second pole of the first magnet, so as to generate a compressed magnetic field. In one embodiment, the coil is configured to pass between the first and second magnets as the first magnet housing is rotated. In one embodiment, the mechanical transmission system is coupled to the coil and configured to move the coil with respect to the first magnet housing in response to a receipt of energy. In one embodiment, the mechanical transmission system comprises a repelling magnet. In one embodiment, the mechanical transmission system comprises a mechanical repelling system. In one embodiment, the coil is configured to receive an electrical signal and the system is configured to move the first magnet housing with respect to the coil in response to the receipt of the electrical signal. In one embodiment, the system is configured to receive energy and to move the first magnet housing with respect to the coil in response to the receipt of the energy. In one embodiment, the system further comprises a second coil. In one embodiment, the coil has an axis that is at least generally aligned with an axis along which the first magnet housing is configured to move relative to the coil. In one embodiment, the system further comprises a second magnet housing, a third magnet secured within the second magnet housing and having a first pole of a first polarity and a second pole of a second polarity, and a fourth magnet having a first pole of the first polarity and a second pole of the second polarity, secured within the second magnet housing such that the first pole of the third magnet is held spaced apart a distance from and generally facing the first pole of the fourth magnet, so as to generate a compressed magnetic field. In one embodiment, the second magnet housing is substantially perpendicular to the first magnet housing. In one embodiment, the system further comprises a second coil. In one embodiment, the first magnet housing is gimbaled. In one embodiment, the electrical signal comprises a

DC current. In one embodiment, the electrical signal comprises an AC current and the system further comprises rectification circuitry coupled to the coil and configured to convert the AC current to a DC current. In one embodiment, the system further comprises a power storage system coupled to the rectification circuitry for accumulating and storing power generated by the system. In one embodiment, the system further comprises an inverter coupled to the power storage system and configured to supply alternating current to an electricity distribution system. In one embodiment, the system is configured to convert energy from waves into an electrical signal.

In one embodiment, a magnetic structure comprises a magnet housing, a first magnet secured within the magnet housing, and having a first pole of a first polarity and a second pole of a second polarity, and a second magnet having a first pole of the first polarity and a second pole of the second polarity, secured within the magnet housing such that the first pole of the second magnet is held spaced apart a distance from and generally facing the first pole of the first magnet, so as to generate a compressed magnetic field. In one embodiment, the first magnet comprises a permanent magnet. In one embodiment, the first magnet comprises a rare earth magnet. In one embodiment, the first magnet comprises an electromagnet. In one embodiment, a space between the first and second magnets is substantially filled with a non-magnet substance. In one embodiment, the non-magnetic substance comprises air. In one embodiment, the non-magnetic substance comprises a fluoropolymer resin. In one embodiment, the magnetic structure further comprises a third magnet having a first pole of the first polarity and a second pole of the second polarity, secured within the magnet housing such that the second pole of the third magnet is held spaced apart a distance from and generally facing the second pole of the first magnet, so as to generate a compressed magnetic field. In one embodiment, the first polarity is a north polarity. In one embodiment, a face of the first pole of the first magnet is at least generally planar. In one embodiment, a face of the first pole of the second magnet is at least generally planar. In one embodiment, a face of the first pole of the first magnet is at least generally convex. In one embodiment, a face of the first pole of the first magnet is at least generally concave. In one embodiment, the first magnet is generally rectangular. In one embodiment, the first magnet is generally spherical. In one embodiment, the magnetic structure further comprises a suspension system. In one embodiment, the suspension system is gimbaled. In one embodiment, gravitational forces are used to position the magnetic structure within the suspension system. In one embodiment, the suspension system is configured to employ gyroscopic principles to position the magnetic structure. In one embodiment, the magnet housing is evacuated and hermetically sealed.

In one embodiment, a magnetic structure comprises a plurality of magnets and means for holding the magnets spaced apart with respect to each other and configured so as to generate a compressed magnetic field. In one embodiment, the means for holding the magnets comprises a magnet housing having a threaded inner surface. In one embodiment, the means for holding the magnets comprises tabs configured to hold the plurality of magnets in fixed positions with respect to each other. In one embodiment, the magnetic structure further comprises means for transmitting mechanical energy coupled to the means for holding the magnets.

In one embodiment, a method of generating power comprises generating a compressed magnetic field using a plurality of spaced-apart magnets and causing relative movement between an electrical conductive winding and the compressed magnetic field. In one embodiment, generating the compressed magnetic field comprises holding the plurality of magnets spaced apart in a fixed position with respect to each other such that like poles of the magnets face each other so as to generate the compressed magnetic field. In one embodiment, the plurality of magnets consists of two magnets and a distance between the two magnets is less than an ambient distance. In one embodiment, the method further comprises rectifying a current generated in the electrical conductive winding. In one embodiment, the method further comprises storing the rectified current in an energy storage system. In one embodiment, causing relative movement comprises moving the electrical conductive winding with respect to the plurality of magnets. In one embodiment, causing relative movement comprises moving the plurality of magnets with respect to the electrical conductive winding. In one embodiment, moving the plurality of magnets with respect to the electrical conductive winding comprises moving the plurality of magnets along a generally linear path. In one embodiment, moving the plurality of magnets with respect to the electrical conductive winding comprises moving the plurality of magnets along a generally radial path. In one embodiment, moving the plurality of magnets with respect to the electrical conductive winding comprises rotating the plurality of magnets. In one embodiment, the method further comprises optimizing a gradient of the compressed magnetic field.

In one embodiment, a method of generating mechanical force comprises generating a compressed magnetic field and conducting a current through an electrical conductive winding in the compressed magnetic field. In one embodiment, generating the compressed magnetic field comprises holding a plurality of magnets spaced apart in a fixed position with respect to each other such that like poles of the magnets face each other so as to generate the compressed magnetic field. In one embodiment, the plurality of magnets consists of two magnets and a distance between the two magnets is less than an ambient distance. In one embodiment, the current is an alternating current. In one embodiment, the current is a direct current. In one embodiment, the method further comprises applying the mechanical force so as to cause a generally linear movement in a transmission system. In one embodiment, the method further comprises applying the mechanical force so as to cause a generally rotational movement in a transmission system.

In one embodiment, a system comprises a coil having an electrical conductive winding and a magnetic conductive winding configured to focus magnetic flux in the electrical conductive winding, and a magnetic structure configured to generate a compressed magnetic field. In one embodiment, the magnetic structure comprises a first magnet housing, a first magnet secured within the first magnet housing, and having a first pole of a first polarity and a second pole of a second polarity, and a second magnet having a first pole of the first polarity and a second pole of the second polarity, secured within the first magnet housing such that the first pole of the second magnet is held spaced apart a distance from and generally facing the first pole of the first magnet, so as to generate the compressed magnetic field. In one embodiment, the system is configured to receive energy and to generate an electrical signal in response to the receipt of the energy. In one embodiment, the electrical signal comprises an AC current and the system further comprises rectification circuitry coupled to the coil and configured to convert the AC current to a DC current. In one embodiment, the electrical signal comprises a DC current. In one embodiment, the system is configured to receive an electrical signal and to generate mechanical force in response to the electrical signal. In one embodiment, the system further comprises a mechanical transmission system. In one embodiment, the mechanical transmission system is coupled to the magnetic structure and configured to move the magnetic structure with respect to the coil in response to a receipt of energy. In one embodiment, the mechanical transmission system is configured to move the magnetic structure in a linear manner. In one embodiment, the mechanical transmission system is configured to rotate the magnetic structure. In one embodiment, the mechanical transmission system is configured to move the magnetic structure along a radial path. In one embodiment, the mechanical transmission system is coupled to the coil and configured to move the coil with respect to the magnetic structure in response to a receipt of energy. In one embodiment, the coil is configured to receive an electrical signal and the system is configured to move the magnetic structure with respect to the coil in response to the receipt of the electrical signal. In one embodiment, the system is configured to receive energy and to move the magnetic structure with respect to the coil in response to the receipt of the energy. In one embodiment, the system is configured to receive energy and to move the coil with respect to the magnetic structure in response to the receipt of the energy. In one embodiment, the coil has an axis that is at least generally aligned with an axis along which the magnetic structure is configured to move relative to the coil. In one embodiment, the system further comprises a gimbaled suspension system. In one embodiment, the system is configured to convert energy from waves into an electrical signal. In one embodiment, the magnetic conductive winding is configured as a closed loop. In one embodiment, the system further comprises an article of clothing configured for coupling the system to a person. In one embodiment, the system further comprises a coupler configured to couple the coil to an electrical transmission grid.

In one embodiment, a method of generating power comprises generating a compressed magnetic field using a plurality of spaced-apart magnets, moving an electrical conductive winding with respect to the compressed magnetic field, and focusing magnetic flux in the electrical conductive winding using a magnetic conductive winding. In one embodiment, generating the compressed magnetic field comprises holding the plurality of magnets spaced apart in a fixed position with respect to each other such that like poles of the magnets face each other so as to generate the compressed magnetic field. In one embodiment, the plurality of magnets consists of two magnets and a distance between the two magnets is less than an ambient distance. In one embodiment, the method further comprises rectifying a current generated in the electrical conductive winding. In one embodiment, the method further comprises storing the rectified current in an energy storage system. In one embodiment, moving the electrical conductive winding with respect to the compressed magnetic field comprises moving the electrical conductive winding with respect to the plurality of magnets. In one embodiment, moving the electrical conductive winding with respect to the compressed magnetic field comprises moving the plurality of magnets with respect to the electrical conductive winding. In one embodiment, moving the plurality of magnets with respect to the electrical conductive winding comprises moving the plurality of magnets along a generally linear path. In one embodiment, moving the plurality of magnets with respect to the electrical conductive winding comprises rotating the plurality of magnets. In one embodiment, the method further comprises optimizing a gradient of the compressed magnetic field. In one embodiment, the magnetic conductive winding forms a closed loop. In one embodiment, the method further comprises coupling the electrical conductive winding to an electrical transmission grid. In one embodiment, the method further comprises generating an alternating current in the electrical conductive winding. In one embodiment, the method further comprises generating a direct current in the electrical conductive winding.

In one embodiment, a method of generating mechanical force comprises generating a compressed magnetic field, focusing magnetic flux in an electrical conductive winding using a magnetic conductive winding, and conducting a current through the electrical conductive winding in the compressed magnetic field. In one embodiment, generating the compressed magnetic field comprises holding a plurality of magnets spaced apart in a fixed position with respect to each other such that like poles of the magnets face each other so as to generate the compressed magnetic field. In one embodiment, the plurality of magnets consists of two magnets and a distance between the two magnets is less than an ambient distance. In one embodiment, the current is an alternating current. In one embodiment, the current is a direct current. In one embodiment, the method further comprises applying the mechanical force so as to cause a generally linear movement in a transmission system. In one embodiment, the method further comprises applying the mechanical force so as to cause a generally rotational movement in a transmission system. In one embodiment, the magnetic conductive winding forms a closed loop.

In one embodiment, an article of clothing comprises a coil having an electrical conductive winding and a magnetic conductive winding configured to focus magnetic flux in the electrical conductive winding, and a magnetic structure configured to generate a compressed magnetic field. In one embodiment, the magnetic structure comprises a first magnet housing, a first magnet secured within the first magnet housing, and having a first pole of a first polarity and a second pole of a second polarity, and a second magnet having a first pole of the first polarity and a second pole of the second polarity, secured within the first magnet housing such that the first pole of the second magnet is held spaced apart a distance from and generally facing the first pole of the first magnet, so as to generate the compressed magnetic field. In one embodiment, the magnetic conductive winding forms a closed loop. In one embodiment, the magnetic structure and the coil are contained within a battery case.

In one embodiment, a system comprises a coil comprising means for conducting an electrical current in response to changes in magnetic flux and means for focusing magnetic flux in the means for conducting the electrical current, and means for generating a compressed magnetic field. In one embodiment, the means for conducting the electrical current comprises an electrical conductive winding and the means for focusing magnetic flux comprises a magnetic conductive winding. In one embodiment, the magnetic conductive winding forms a closed loop. In one embodiment, the means for generating the compressed magnetic field comprises a first magnet housing, a first magnet secured within the first magnet housing, and having a first pole of a first polarity and a second pole of a second polarity, and a second magnet having a first pole of the first polarity and a second pole of the second polarity, secured within the first magnet housing such that the first pole of the second magnet is held spaced apart a distance from and generally facing the first pole of the first magnet, so as to generate the compressed magnetic field.

In one embodiment, a battery comprises a case, a first generator contained within the case and configured to convert energy received by the battery into electrical energy, a first energy storage device contained within the case, a second energy storage device contained within the case, a control module contained within the case, coupled to the first and second energy storage devices, and configured to control a transfer of the electrical energy from the first energy storage device to the second energy storage device, and a plurality of contact terminals. In one embodiment, the first energy storage device comprises an ultracapacitor and the second energy storage device comprises a lithium cell. In one embodiment, the battery further comprises a third energy storage device. In one embodiment, the third energy storage device is coupled in series with the second energy storage device. In one embodiment, the third energy storage device is coupled in parallel with the first energy storage device. In one embodiment, the battery further comprises a connector to house the plurality of contact terminals. In one embodiment, the case and contact terminals have a configuration of a C-cell battery. In one embodiment, the first generator comprises a coil and a magnetic structure. In one embodiment, the magnetic structure is configured to generate a compressed magnetic field. In one embodiment, the coil comprises an electrical conductive element, and a magnetic conductive element. In one embodiment, the magnetic structure is configured to generate a compressed magnetic field. In one embodiment, the plurality of contact terminals are electrically coupled to the control module. In one embodiment, the battery further comprises a second generator contained within the case, wherein the first generator is oriented in a first direction and the second generator is oriented in a second direction different from the first direction. In one embodiment, the control module is further configured to control a transfer of energy between the second energy storage device and the contact terminals. In one embodiment, the transfer of energy between the first energy storage device and the contact terminals comprises a transfer of energy from the contact terminals to the second energy storage device. In one embodiment, the transfer of energy between the first energy storage device and the contact terminals comprises a transfer of energy from the contact terminals to the first energy storage device. In one embodiment, the control module is further configured to control a transfer of energy between the first energy storage device and the contact terminals. In one embodiment, the battery further comprises a suspension system coupled to the generator. In one embodiment, the suspension system is tuned to optimize conversion of expected patterns of movement into electrical energy. In one embodiment, the suspension system is gimbaled. In one embodiment, the suspension system comprises a gyroscopic system. In one embodiment, the generator is configured to convert energy received through movement of the battery. In one embodiment, the generator is configured to convert energy received in a parasitic manner. In one embodiment, the case comprises a magnetic shield.

In one embodiment, a battery comprises a case, a coil contained within the case, a magnetic structure contained within the case and configured to generate a compressed magnetic field, a first energy storage device contained within the case, a plurality of contact terminals coupled to the case, and a control module contained within the case and coupled to the coil and the first energy storage device. In one embodiment, the magnetic structure comprises a plurality of spaced-apart rare earth magnets configured so that like-polarity poles face each other in neighboring magnets in the plurality of rare earth magnets. In one embodiment, the magnets in the plurality of magnets are held in position with respect to one another. In one embodiment, a space between two magnets in the plurality of magnets is substantially filled with a non-magnet substance. In one embodiment, the non-magnetic substance comprises air. In one embodiment, the non-magnetic substance comprises a fluoropolymer resin. In one embodiment, the case is evacuated and hermetically sealed. In one embodiment, the battery further comprises a suspension system coupled to the magnetic structure. In one embodiment, the suspension system is tuned to optimize conversion of expected patterns of movement into electrical energy. In one embodiment, the coil comprises an electrical conductive element and a magnetic conductive element. In one embodiment, the magnetic conductive element is configured to focus magnetic flux in the electrical conductive element.

In one embodiment, a battery comprises a case, a coil contained within the case and having an electrical conductive element and a magnetic conductive element, a magnetic structure, a first energy storage device contained within the case, a plurality of contact terminals coupled to the case and a control module contained within the case and coupled to the coil and the first energy storage device. In one embodiment, the magnetic conductive element is configured to focus magnetic flux in the electrical conductive element. In one embodiment, the electrical conductive element comprises a electrical conductive wire in a multi-wire winding and the magnetic conductive element comprises a magnetic conductive wire in the multi-wire winding. In one embodiment, the electrical conductive element comprises an electrical conductive winding and the magnetic conductive element comprises a magnetic conductive winding. In one embodiment, the electrical conductive element comprises an electrical conductive trace formed on a first insulating substrate. In one embodiment, the magnetic conductive element comprises a magnetic conductive trace formed on the first insulating substrate. In one embodiment, the electrical conductive trace is formed on a first surface of the first insulating substrate and the magnetic conductive trace is formed on the first surface of the first insulating substrate. In one embodiment, the battery further comprises a plurality of insulating substrates, wherein the electrical conductive element comprises a plurality of electrical conductive traces formed on selected substrates in the plurality of substrates and the magnetic conductive element comprises a plurality of magnetic conductive traces formed on selected substrates in the plurality of substrates. In one embodiment, the magnetic structure is configured to generate a compressed magnetic field. In one embodiment, a contact terminal of the plurality of contact terminals is electrically coupled to a contact terminal of an external battery. In one embodiment, the battery has a first physical orientation and the external battery has a second physical orientation different from the first physical orientation.

In one embodiment, a battery comprises a case, means for converting movement of the battery into an electric current, first means for storing energy contained within the case, second means for storing energy contained within the case, means for controlling a transfer of energy from the means for converting movement to the first means for storing energy contained within the case, and means for accessing energy stored in the first means for storing energy. In one embodiment, the battery further comprises third means for storing energy contained within the case. In one embodiment, the means for converting movement comprises means for conducting an electric current and means for generating a magnetic field. In one embodiment, the means for generating a magnetic field is configured to generate a compressed magnetic field. In one embodiment, the battery further comprises means for conducting magnetic flux. In one embodiment, the battery further comprises means for facilitating relative movement of the means for conducting an electric current with respect to the means for generating a magnetic field.

In one embodiment, a method of operating a battery comprises moving the battery, converting energy received through the movement of the battery into an electric current, and controlling transfers of energy to a plurality of energy storage devices contained within the battery. In one embodiment, controlling transfers of energy comprises storing energy from the electric current in a first energy storage device in the plurality of energy storage devices and controlling a transfer of energy from the first energy storage device to a second energy storage device in the plurality of energy storage devices. In one embodiment, controlling transfers of energy comprises rectifying the electric current. In one embodiment, the method further comprises controlling a transfer of energy from the battery to a load. In one embodiment, the method further comprises providing an electric current to the battery and controlling a storage in the battery of energy from the provided electric current. In one embodiment, converting energy received through the movement of the battery into the electric current comprises generating a compressed magnetic field. In one embodiment, converting energy received through the movement of the battery into the electric current further comprises focusing the compressed magnetic field in an electrical conductive winding. In one embodiment, generating the compressed magnetic field comprises holding two magnets spaced-apart with like poles facing each other at a distance closer than an ambient distance. In one embodiment, converting energy received through the movement of the battery into the electric current comprises focusing a magnetic field in an electrical conductive element. In one embodiment, focusing the magnetic field in the electrical conductive element comprises positioning a magnetic conductive element with respect to the electrical conductive element so as to focus the magnetic field. In one embodiment, converting energy received through the movement of the battery into the electric current comprises orienting a generator contained within the battery. In one embodiment, converting the energy received through the movement of the battery into the electric current comprises converting the energy into relative movement between an electrical conductive winding and a magnetic field. In one embodiment, the relative movement is generally linear. In one embodiment, the relative movement is generally rotational.

In one embodiment, a system comprises a first battery having a first orientation and comprising means for converting energy into a first electrical signal, and a second battery electrically coupled to the first battery and having a second orientation, and comprising second means for converting energy into a second electrical signal. In one embodiment, the second orientation is substantially perpendicular to the first orientation. In one embodiment, the means for converting energy into the first electrical signal comprises a control module configured to control a transfer of electrical energy from a first energy storage device to a second energy storage device. In one embodiment, the means for converting energy into the first electrical signal comprises means for generating a compressed magnetic field. In one embodiment, the means for converting energy into the first electrical signal comprises an electrical conductive winding and a magnetic conductive winding configured to focus magnetic flux in the electrical conductive winding.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

The patent or application file contains at least one drawing executed in color. Copies of the patent or patent application publication with color drawings will be provided by the Office upon request and payment of the necessary fee.

The sizes and relative positions of elements in the drawings are not necessarily drawn to scale. For example, the shapes of various elements and angles are not drawn to scale, and some of these elements are arbitrarily enlarged and positioned to improve drawing legibility. Further, the particular shapes of the elements as drawn are not necessarily intended to convey any information regarding the actual shape of particular elements, and have been selected solely for ease of recognition in the drawings.

FIG. **1** is a diametric cross-sectional view of a conventional coil.

FIG. **2** is a diametric cross-sectional view of an embodiment of a coil in accordance with the present disclosure.

FIG. **3** is a diametric cross-sectional view of another embodiment of a coil in accordance with the present disclosure.

FIG. **4** is a diametric cross-sectional view of an embodiment of a dual conductor winding suitable for use in the embodiment of a coil illustrated in FIG. **3**.

FIG. **5** is a diametric cross-sectional view of another embodiment of a coil in accordance with the present disclosure.

FIG. **6** is a diametric cross-sectional view of another embodiment of a coil in accordance with the present disclosure.

FIG. **7** is a top view of another embodiment of a coil in accordance with the present disclosure.

FIG. **8** is a bottom view of the embodiment of a coil illustrated in FIG. **7**.

FIG. **9** is a side view of the embodiment of a coil illustrated in FIG. **7**.

FIG. **10** is a top view of another embodiment of a coil in accordance with the present disclosure.

FIG. **11** is a side view of another embodiment of a coil in accordance with the present disclosure.

FIG. **12** is a graphic illustration of the magnetic flux generated by a conventional magnetic structure.

FIGS. **13**A and **13**B are graphic illustrations of the magnetic flux generated by two permanent magnets with like poles facing each other and separated by an ambient distance.

FIGS. **14**A and **14**B are graphic illustrations of the magnetic flux generated by two permanent magnets with like poles substantially touching each other.

FIGS. **15**A and **15**B are graphic illustrations of the magnetic flux generated by two permanent magnets with like poles facing each other and held together between an ambient distance and a substantially touching position.

FIG. **16** is cross-sectional view of an embodiment of a magnetic structure in accordance with the present disclosure.

FIG. **17** is cross-sectional view of an embodiment of a magnetic structure in accordance with the present disclosure.

FIG. **18** is cross-sectional view of an embodiment of a magnetic structure in accordance with the present disclosure.

FIG. **19** is cross-sectional view of an embodiment of a magnetic structure in accordance with the present disclosure.

FIG. **20** is a side diametric cross-sectional view of another embodiment of a magnet structure in accordance with the present disclosure.

FIG. **21** is a side view of an embodiment of a magnet structure in accordance with the present disclosure.

FIG. **22** is a diagrammatical front view of an embodiment of a power generator system.

FIG. **23** is a diagrammatical front view of the system of FIG. **22** at a different point in time.

FIG. **24** is a side view, in section, of an armature included in the system of FIG. **22**.

FIG. **25** is a diagrammatical front view of a system in accordance with an alternative embodiment of a generator.

FIG. **26** is a side view, in section, of an armature included in the system of FIG. **25**.

FIG. **27** is a side cross-sectional view of an embodiment of a system in accordance with the present disclosure.

FIG. **28** is a side cross-sectional view of another embodiment of a system in accordance with the present disclosure.

FIG. **29** is a top view of a system in accordance with the present disclosure.

FIG. **30** is a side diametric cross-sectional view of the system of FIG. **29** taken along line **30-30**.

FIG. **31** is a side diametric cross-sectional view of an embodiment of a system employing the embodiments illustrated in FIGS. **7** through **9** and FIG. **17**.

FIG. **32** is a side diametric cross-sectional view of an embodiment of a system employing the embodiments illustrated in FIG. **11** and FIG. **16**.

FIG. **33** is a diametric cross-sectional view of an embodiment of a battery.

FIG. **34** is a diametric cross-sectional view of another embodiment of a battery.

FIG. **35** is a side sectional view of another embodiment of a battery.

FIG. **36** is a diametric cross sectional view of a linear generator suitable for use in the embodiments illustrated in FIGS. **33** through **36**.

FIG. **37** is a high-level flow diagram for an embodiment of a method of recharging a portable energy storage device.

FIG. **38** is a high level flow diagram for an embodiment of a method of operating a portable energy storage device.

FIG. **39** is a perspective view illustrating a practical application for an embodiment of a power generator.

FIG. **40** is a block diagram of an embodiment of a system for generating power.

FIG. **41** is a block diagram of an embodiment of a self-powered device.

FIG. **42** illustrates an embodiment of a system in accordance with the present disclosure.

FIG. **43** illustrates another embodiment of a system in accordance with the present disclosure.

FIG. **44** illustrates an embodiment of an article of clothing in accordance with the present disclosure.

FIG. **45** is a side view of an embodiment of a system in accordance with the present invention.

FIG. **46** is a top view of an embodiment of a rotor suitable for use in the embodiment of a system illustrated in FIG. **45**.

DETAILED DESCRIPTION OF THE INVENTION

In the following description, certain details are set forth in order to provide a thorough understanding of various embodiments of devices, methods and articles. However, one of skill in the art will understand that other embodiments may be practiced without these details. In other instances, well-known structures and methods associated with batteries, linear generators, and control systems have not been shown or described in detail to avoid unnecessarily obscuring descriptions of the embodiments.

Unless the context requires otherwise, throughout the specification and claims which follow, the word "comprise" and variations thereof, such as "comprising," and "comprises," are to be construed in an open, inclusive sense, that is as "including, but not limited to."

Reference throughout this specification to "one embodiment," or "an embodiment" means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment. Thus, the appearances of the phases "in one embodiment," or "in an embodiment" in various places throughout this specification are not necessarily referring to the same embodiment, or to all embodiments. Furthermore, the particular features, structures, or characteristics may be combined in any suitable manner in one or more embodiments to obtain further embodiments.

The headings are provided for convenience only, and do not interpret the scope or meaning of this disclosure or the claimed invention.

FIG. **1** is a diametric cross sectional view of a conventional coil **100**. The coil **100** comprises a non-magnetic winding form **102** and a non-magnetic, electrical conductive winding **104**. A winding comprises one or more complete turns of a conductive materials in a coil, and may comprise one or more layers. As illustrated, the winding **104** comprises nine turns and three layers. As illustrated, the electrical conductive winding **104** is continuous. In other conventional coils, a plurality of electrical conductive windings may be employed, which may or may not be electrically connected in series or in parallel. The electrical conductive winding **104** may comprise any suitable electrically conductive material, such as, for example, metallic materials, such as copper, copper coated with silver or tin, aluminum, silver, gold and/or alloys. The electrical conductive winding **104** may comprise, for example, solid wires, including, for example, flat wires, strands, twisted strands, or sheets. The electrical conductive winding **104** may vary significantly in size from the illustration, and may be substantially smaller or substantially larger than illustrated. The electrical conductive winding **104** is typically covered with an insulating material **120**. The electrical conductive winding **104** is coupled to the leads **122**, **124** for the coil **100**.

FIG. **2** is a diametric cross-sectional view of an embodiment of a bi-metal coil **200**. The coil **200** comprises a non-magnetic winding form **202**, a non-magnetic, electrical conductive winding **204** and a magnetic conductive winding **206**. The use of an electrical conductive winding, such as the electrical conductive winding **204**, with a magnetic conductive winding, such as the magnetic conductive winding **206**, facilitates focusing of a magnetic field passing through or generated by an electrical conductive winding of a coil, such as the winding **204** of the coil **200**. Focusing of the magnetic field can significantly increase the efficiency of the coil **200**. For example, when the coil **200** is employed in a generator, as a magnet is passed through the coil **200** the electrical conductive winding **204** produces electron flow, while the magnetic conductive winding **206** focuses magnetic flux in the electrical conductive winding **204** and causes an increase in power output from the coil **200**.

A first layer **208** and a second layer **210** of the electrical conductive winding **204** are wound onto the winding form **202**. In one embodiment, the electrical conductive winding **204** is continuous. In other embodiments, the electrical conductive winding **204** may comprise a plurality of windings, which may be electrically connected in series or in parallel. A first layer **212** of the magnetic conductive winding **206** is wound over the second layer **210** of the electrical conductive winding **204**. A third layer **214** and a fourth layer **216** of the electrical conductive winding **204** are wound over the first layer **212** of the magnetic conductive winding **206**. A second layer **218** of the magnetic conductive winding **206** is wound over the fourth layer **216** of the electrical conductive winding **204**. A fifth layer **219** of the electrical conductive winding **204** is wound over the second layer **218** of the magnetic conductive winding **206**.

The electrical conductive winding **204** may comprise any suitable electrically conductive material, such as, for example, metallic materials, such as copper, copper coated with silver or tin, aluminum, silver, gold and/or alloys. The electrical conductive winding **204** may comprise, for example, solid wires, strands, twisted strands, or sheets. The electrical conductive winding **204** may vary significantly in size from the illustration, and may be substantially smaller or substantially larger than illustrated. The electrical conductive winding **204** is typically covered with an insulating material **220**. The electrical conductive winding **204** is coupled to the leads **222**, **224** for the coil **200**.

The magnetic conductive winding **206** may comprise any suitable magnetic conductive material, for example, a magnetic shielding material, such as, for example, nickel, nickel/iron alloys, nickel/tin alloys, nickel/silver alloys, plastic magnetic shielding, and/or nickel/iron/copper/molybdenum alloys. Magnetic shielding materials are commercially available under several trademarks, including MuMetal®, Hipernom®, HyMu 80®, and Permalloy®. The magnetic conductive winding **206** may comprise, for example, solid wires, strands, twisted strands, or sheets. The magnetic conductive winding **206** may vary significantly in size from the illustration, and may be substantially smaller or substantially larger than illustrated. The magnetic conductive winding **206** is typically covered with an insulating material **226**. The magnetic conductive winding **206** forms a closed loop, as illustrated by the connection **228**, and as illustrated is connected to a ground **230**.

Other configurations of layers of an electrical conductive winding and a magnetic conductive winding may be employed. For example, m layers of an electrical conductive winding may alternate with n layers of a magnetic conductive winding, instead of two layers of electrical conductive winding alternating with one layer of magnetic conductive winding as illustrated, with m and n positive integers. In another example, m and n need not remain constant. For example, the number of layers may increase or decrease. An example layer pattern would be 2E, 1M, 3E, 2M, 4E, with E indicating electrically conductive layers and M indicating magnetically conductive layers.

Typically, the first and last layers comprise layers of the electrical conductive winding **204**. In one experimental embodiment, a configuration with the first and last layer comprising the electrical conductive winding **204** produced better performance in a generator application than when the last layer was comprised of the magnetic conductive winding **206**. In another example, a plurality of electrical conductive windings could be employed.

FIG. **3** is a diametric cross-sectional view of another embodiment of a bi-metal coil **300**. The coil **300** comprises a winding form **302** and a dual-conductor winding, which as illustrated takes the form of a bi-metal winding **304**. The dual-conductor winding **304** comprises an electrical conductive winding in the form of a wire **306**, a magnetic conductive winding in the form of a wire **308**, an inner layer of insulating material **310** between the electrical conductive wire **306** and the magnetic conductive wire **308**, and an outer layer of insulating material **312**. The outer layer of insulating material **312** and the inner layer of insulating material **310** may be integrated. The dual-conductor winding **304** may vary significantly in size from the illustration, and may be substantially smaller or substantially larger than illustrated. As illustrated, the electrical conductive wire **306** and the magnetic conductive wire **308** are approximately the same size. In some embodiments, the electrical conductive wire **306** and the magnetic conductive wire **308** may be of different sizes.

The electrical conductive wire **306** may comprise any suitable electrical conductive material. For example, the materials and configurations (e.g., solid wire or stranded wire) discussed above with respect to the electrical conductive winding **204** of FIG. **2** may be employed. The electrical conductive wire **306** is coupled to the leads **314**, **316** for the coil **300**. The magnetic conductive wire **308** may comprise any suitable magnetic conductive material. For example, the materials and configurations (e.g., solid wire or stranded wire) discussed above with respect to the magnetic conductive winding **206** of FIG. **2** may be employed. The magnetic conductive wire **308** forms a closed loop, as illustrated by the connection **318**, and may be connected to a ground **320**. As illustrated, the winding **304** is wound so that the electrical conductive wire **306** is closest to an inner surface **322** of the winding form **302** and the magnetic conductive wire **308** is farthest from the inner surface **322** of the winding form **302**. As illustrated, the insulating layer **310** separating the electrical conductive wire **306** and the magnetic conductive wire **308** is approximately parallel to the inner surface **322**. In some embodiments, the insulating layer **310** separating the electrical conductive wire **306** and the magnetic conductive wire **308** may be at another angle with respect to the inner surface **322**. For example, in some embodiments the insulating layer **310** may be approximately perpendicular to the inner surface **322**. As illustrated, the dual-conductor winding **304** is a single layer comprising three turns. In some embodiments, the winding may comprise multiple layers. In some embodiments, additional windings may be employed.

FIG. **4** is a diametric cross sectional view of an embodiment of a dual conductor winding **404** suitable for use in the embodiment of a coil **300** illustrated in FIG. **3**. The dual conductor winding **404** comprises an electrical conductive winding in the form of a wire **406**, a magnetic conductive winding in the form of a wire **408**, an inner layer of insulating material **410** between the electrical conductive wire **406** and the magnetic conductive wire **408**, and an outer layer of insulating material **412**. The electrical conductive wire **406** may comprise any suitable electrical conductive material. For example, the materials and configurations (e.g., solid wire or stranded wire) discussed above with respect to the electrical conductive winding **204** of FIG. **2** may be employed. The electrical conductive wire **406** may vary significantly in size from the illustration, and may be substantially smaller or substantially larger than illustrated. The magnetic conductive wire **408** may comprise any suitable magnetic conductive material. For example, the materials and configurations (e.g., solid wire or stranded wire) discussed above with respect to the magnetic conductive winding **206** of FIG. **2** may be employed. The magnetic conductive wire **408** may vary significantly in size from the illustration, and may be substantially smaller or substantially larger than illustrated.

FIG. **5** is a diametric cross-sectional view of another embodiment of a bi-metal coil **500**. The coil **500** has a winding form **502**. A first electrical conductive winding **504***a* is wound on the winding form **502** in two layers. A first magnetic conductive winding **506***a* is wound on the winding form **502** outside of the two layers of the first electrical conductive winding **504***a*. The first electrical conductive winding **504***a* is coupled to leads **508***a*, **510***a* for the first electrical conductive winding **504***a*. The first magnetic conductive winding **506***a* forms a closed loop, as illustrated by the first connection loop **512***a*. A second electrical conductive winding **504***b* is wound on the winding form **502** in two layers adjacent to the first electrical conductive winding **504***a*. A second magnetic conductive winding **506***b* is wound on the winding form **502** outside of the two layers of the second electrical conductive winding **504***b* adjacent to the first magnetic conductive winding **506***a*. The second electrical conductive winding **504***b* is coupled to leads **508***b*, **510***b* for the second electrical conductive winding **504***b*. The second magnetic conductive winding **506***b* forms a closed loop, as illustrated by the second connection loop **512***b*. Additional windings may be added to the coil **500**, as illustrated by the electrical conductive winding **504***n*, which is coupled to the leads **508***n*, **510***n*, and by the magnetic conductive winding **506***n*, which forms a closed loop as indicated by connection **512***n*. In some embodiments, the electrical conductive windings (e.g., windings **504***a*, **504***b*, . . . **504***n*) may be electrically coupled together in parallel or in series, or in various combinations thereof.

The electrical conductive windings **504***a*, **504***b*, . . . **504***n* may comprise any suitable electrical conductive material. For example, the materials and configurations (e.g., solid wire or stranded wire) discussed above with respect to the electrical conductive winding **204** of FIG. **2** may be employed. The electrical conductive windings **504***a*, **504***b*, . . . **504***n* may vary significantly in size from the illustration, and may be substantially smaller or substantially larger than illustrated. The magnetic conductive windings **506***a*, **506***b*, . . . **506***n* may comprise any suitable magnetic conductive material. For example, the materials and configurations (e.g., solid wire or stranded wire) discussed above with respect to the magnetic conductive winding **206** of FIG. **2** may be employed. The magnetic conductive windings **506***a*, **506***b*, . . . **506***n* may vary significantly in size from the illustration, and may be substantially smaller or substantially larger than illustrated. A bi-metal winding may be employed, such as the bi-metal winding **304** illustrated in FIG. **3** or the bi-metal winding **404** illustrated in FIG. **4**. Typically, the coil **500** would have additional layers of each winding, with the outer layer comprising a layer of the electrical conductive winding (e.g., winding **504***a*), but for ease of illustration the additional layers are omitted.

FIG. **6** is a diametric cross-sectional view of another embodiment of a bi-metal coil **600**. The coil **600** comprises a non-magnetic winding form **602**, a non-magnetic, electrical conductive winding **604** and a magnetic conductive winding **606**. A first layer **608** and a second layer **610** of the electrical conductive winding **604** are wound onto the winding form **602**. A layer **612** of the magnetic conductive winding **606** is wound over the second layer **610** of the electrical conductive winding **604**.

The electrical conductive winding **604** may comprise any suitable electrical conductive material and configuration. For example, the materials and configurations (e.g., solid wire or stranded wire) discussed above with respect the electrical conductive winding **204** of FIG. **2** may be employed. The electrical conductive winding **604** is typically covered with an insulating material **614**. The electrical conductive winding **604** is coupled to the leads **616**, **618** for the coil **600**. The magnetic conductive winding **606** may comprise any suitable magnetic conductive material and configuration. For example, the materials and configurations (e.g., solid wire or stranded wire) discussed above with respect to the magnetic conductive winding **206** of FIG. **2** may be employed. The magnetic conductive winding **606** is typically covered with an insulating material **620**. The magnetic conductive winding **606** forms a closed loop, as illustrated by the connection **622**, and may be connected to a ground (See ground **230** in FIG. **2**). Some embodiments may employ a bi-metal or dual-conducting winding (see dual-conductor winding **304** illustrated in FIG. **3**).

The winding form **602** has an inner length **624** and an outer length **626** that are different. As illustrated, the inner length **624** is shorter than the outer length **626**. This difference in length facilitates focusing of a magnetic field in the electrical conductive winding **604**.

FIGS. **7** through **9** illustrate another embodiment of a bi-metal coil **700**. FIGS. **7** through **9** are not drawn to scale for ease of illustration. FIG. **7** is a top view of the coil **700**. The coil **700** comprises a layer of insulating material **702** with an upper surface **704**. The layer of insulating material **702** may comprise, for example, an integrated circuit board, a substrate or a thin film or sheet of insulation. Commercially available insulating materials are sold under the trademark Mylar®. An electrical conductive winding in the form of a trace **706** is formed on the upper surface **704** of the layer of insulating material **702**. The electrical conductive trace **706** may comprise any suitable electrical conductive material, such as, for example, copper, aluminum, gold, and silver, and alloys. The materials discussed above with respect to the electrical conductive winding **204** of FIG. **2** may be employed. Well-known techniques for forming traces on substrates may be employed, such as those used in connection with RFID devices and antennas. The layer of insulating material **702** has an opening **708**.

FIG. **8** is a bottom view of the embodiment of a coil **700** illustrated in FIG. **7**. The layer of insulating material **702** has a lower surface **716**. A magnetic conductive winding in the form of a trace **718** is formed on the lower surface **716** of the layer of insulating material **702**. The magnetic conductive trace **718** may comprise any suitable magnetic conductive material, such as, for example, nickel, nickel/iron alloys, nickel/tin alloys, nickel/silver alloys. The materials discussed above with respect to the magnetic conductive winding **206** of FIG. **2** may be employed. Well-known techniques for forming traces on substrates may be employed, such as those used in connection with RFID devices and antennas. FIG. **9** is a side view of the embodiment of a coil **700** illustrated in FIG. **7**, illustrating an optional core **730** for the coil **700**. The core **730** may comprise, for example, an iron core.

FIG. **10** is a top view of another embodiment of a bi-metal coil **1000**. The coil **1000** comprises a layer of insulating material **1002** with an upper surface **1004**. The layer of insulating material **1002** may comprise, for example an integrated circuit board, a substrate or a thin film of insulation. An electrical conductive winding in the form of a trace **1006** is formed on the upper surface **1004** of the layer of insulating material **1002**. The electrical conductive trace **1006** may comprise any suitable electrical conductive material, such as, for example, copper, aluminum, gold, and silver, and alloys. For example, the materials discussed above with respect to the electrical conductive winding **204** of FIG. **2** may be employed. Well-known techniques for forming traces on substrates may be employed, such as those used in connection with RFID devices and antennas. The layer of insulating material **1002** has an opening **1008**. A magnetic conductive winding in the form of a trace **1018** is formed on the upper surface **1004** of the layer of insulating material **1002**. The magnetic conductive trace **1018** may comprise any suitable magnetic conductive material, such as, for example, nickel, nickel/iron alloys, nickel/tin alloys, nickel/silver alloys. For example, the materials discussed above with respect to the magnet conductive winding **206** of FIG. **2** may be employed. Well-known techniques for forming traces on substrates may be employed, such as those used in connection with RFID devices and antennas.

FIG. **11** is side view of another embodiment of a bi-metal coil **1100**. The coil **1100** comprises a plurality of insulating layers **1102**. Traces of electrical conductive material **1106** are formed on selected surfaces **1130**, **1132**, **1134**, **1136** of the plurality of insulating layers **1102**. Traces of magnetic conductive material **1118** are formed on selected surfaces **1138**, **1140** of the plurality of insulating layers **1102**. The layers of insulating material **1102** have an opening **1108**. As illustrated, the coil **1100** comprises three layers of insulating material **1102**. Fewer or additional layers **1102** may be employed. In addition, in some embodiments the traces of electrical conductive material **1106** and the traces of magnetic conductive material **1118** may be formed on selected surfaces of the layers of insulating material **1102** in different patterns. For example, traces of electrical conductive material **1106** and traces of magnetic conductive material **1118** may be formed on alternate surfaces of layers of insulating material. In another example, traces of electrical conductive material **1106** and magnetic conductive material **1118** may be formed on the same surface of a layer of insulating material **1102**, or on each surface of a layer of insulating material. Traces on various layers of insulating material **1102** may be coupled to each other.

As mentioned above, coils are frequently employed in devices and applications together with magnets. Bi-metal coils can be advantageously employed in such applications and environments with conventional magnets. FIG. **12** is a graphic illustration of the magnetic flux generated by a conventional magnetic structure **1200**. The magnetic structure comprises a magnet **1202** having a north pole N and a south pole S. FIG. **12** shows representative magnetic flux equipotential lines **1204** to illustrate the magnetic field that is generated by the permanent magnet **1202** of the magnetic structure **1200**. The closer the equipotential lines in a region, the greater the magnetic flux density in the region.

Improvements, however, can be made to conventional magnetic structures. In many devices and applications, increasing the magnetic flux density in a region can greatly improve efficiency and performance. For example, increasing the magnetic flux density in a region can lead to a higher gradient, which can lead to increased efficiency in, for example, a generator or a motor.

FIGS. **13**A and **13**B are graphic illustrations of the magnetic flux generated by a magnetic structure with two permanent magnets with north poles facing each other and separated by an ambient distance. FIG. **13**A is a black and white representation and FIG. **13**B is a color representation. Representative magnetic flux equipotential lines illustrate the magnetic field that is generated by the magnetic structure. The magnetic flux has a higher gradient in the region between the north poles than the magnetic flux in the region around the north pole generated by a single magnet or by magnetic structures that have opposite polarity poles facing each other.

FIGS. **14**A and **14**B are graphic illustrations of the magnetic flux generated by a magnetic structure with two permanent magnets with north poles substantially touching each other. Representative magnetic flux equipotential lines illustrate the magnetic field that is generated by the magnetic structure. For similar magnets, the magnetic flux in the region adjacent to the substantially touching north poles generated by the arrangement illustrated in FIGS. **14**A and **14**B has a higher gradient than the magnet flux generated by the arrangement illustrated in FIGS. **13**A and **13**B, which is illustrated by the greater density of flux lines in FIGS. **14**A and **14**B. A higher magnetic flux gradient also occurs in a region adjacent to the south pole of the upper magnet illustrated in FIGS. **14**A and **14**B.

FIGS. **15**A and **15**B are graphic illustrations of the magnetic flux generated by a magnetic structure with two permanent magnets with like poles facing each other and held together at a distance between an ambient distance and a substantially touching position. Representative magnetic flux equipotential lines illustrate the magnetic field that is generated by the magnetic structure. For similar magnets, the magnetic flux generated by the arrangement illustrated in FIGS. **15**A and **15**B produces a denser set of flux lines along a larger region adjacent to the north poles, permitting a higher flux gradient in a larger region than the magnet flux generated by the arrangements illustrated in FIGS. **13**A, **13**B, **14**A and **14**B, which is illustrated by the greater density of flux lines along a larger region of the sides of the permanent magnets illustrated in FIGS. **15**A and **15**B.

Significant improvements in efficiency, for example, in power generation, can be achieved by positioning the magnets with like poles facing each other at an optimum distance between a touching configuration and an ambient distance. The optimum distance will vary depending upon the configuration in which the magnetic structure is to be employed (e.g., the movement path of the magnetic structure with respect to a coil when the magnetic structure is employed in a generator/motor configuration). FIG. **16** is a cross-sectional view of an embodiment of a multipole magnetic structure **1604** generating a plurality of compressed magnetic fields. In some applications, generating a plurality of compressed magnetic fields can provide further increases in efficiency. The compressed magnetic fields may increase the efficiency of the conversion of energy into electrical energy when the magnetic structure **1604** is employed, for example, in a generator. Such generators may be configured to convert energy received in a parasitic manner. Typical sources of energy include kinetic sources, thermal sources, acoustic sources, and radio-frequency sources.

The magnetic structure **1604** employs tabs **1694** to hold the permanent magnets **1612**, **1614**, **1616** in position with respect to each other. While the illustrated embodiment employs three permanent magnets **1612**, **1614**, **1616**, other embodiments may employ different numbers of permanent magnets, such as two permanent magnets of four permanent magnets. Other embodiments may employ electromagnets instead of or in addition to permanent magnets. The permanent magnets **1612**, **1614**, **1616** are disk-shaped as illustrated, but other shapes may be employed. For example, rectangular- (e.g., square), spherical-, or elliptical-shaped magnets may be employed. Similarly, the faces of the magnets need not be flat. For example, convex-, concave-, radial-, cone-, or diamond-shaped faces may be employed. Various combinations of shapes and faces may be employed. While the illustrated embodiment employs tabs, other positioning mechanisms may be employed, such as threads, spacers, glues, or combinations of positioning mechanisms. The magnets **1612**, **1614**, **1616** are positioned and held apart from each other and are arranged such that same polarity poles in adjacent permanent magnets face each other. For example, the N pole **1628** of the first permanent magnet **1612** faces the N pole **1630** of the second permanent magnet **1614** and the S pole **1632** of the second permanent magnet **1614** faces the S pole **1634** of the third permanent magnet **1616**. In addition, the magnets **1612**, **1614**, **1616** are held close enough together to form a compressed magnetic field (e.g., closer than an ambient distance and spaced apart). In some embodiments, the spaces **1636**, **1638** between the permanent magnets **1612**, **1614**, **1616** are substantially filled with a material **1637**, which may comprise a gas such as air. In some embodiments, the material **1637** may comprise other substantially non-magnet and substantially non-conductive substances, such as a fluoropolymer resin or plastic. In some embodiments, the spaces **1636**, **1638** between the magnets may be evacuated and hermetically sealed.

The shape, position and strength of the permanent magnets in a magnetic structure, such as the magnetic structure **1604**, can increase the efficiency of a device or application employing the magnetic structure **1604**, such as a generator, by generating a compressed magnetic field. A gauss meter (not shown) may be employed to determine the optimum strength and positioning of the permanent magnets **1612**, **1614**, **1616**, as well as the number of permanent magnets. Other design considerations may be taken into consideration as well, such as weight and reducing external impacts of electromagnetic fields and control of multiple generator magnetic interaction.

FIG. **17** illustrates an embodiment of a magnet structure **202** configured to generate a compressed magnetic field. The magnet structure **202** includes a case **204** and a plurality of magnets housed in the case **204** with like poles facing each other. In the illustrated embodiment, the case **204** houses a first magnet **32** having an end **30** of a first polarity and a second magnet **36** having an end **34** of the same polarity as the end **30** and facing the end **30**. In the illustrated embodiment, the end **30** is a south pole and the end **34** is a south pole; in an alternative embodiment, two north poles face each other. In the illustrated embodiment, the case **204** has an inner cylindrical surface **205** and the magnets **32** and **36** have respective outer cylindrical surfaces. The magnets **32**, **36** are slidingly received into the case **204**. In the illustrated embodiment, the case **204** has an open threaded end, through which the magnets **32** and **36** are inserted (or replaced) and the magnet assembly **202** further includes a screw cap **206** selectively closing the threaded end of the case **204**. In the illustrated embodiment, the screw cap **206** forces the magnets **32**, **36** together closer than the ambient distance that the repelling force generated by the magnetic field would normally permit, thereby generated a compressed magnetic field. Other embodiments for positioning the magnets are possible. For example, the inner cylindrical surface **205** could have recesses into which the magnets are snapped.

FIG. **18** illustrates an embodiment of a multi-pole magnetic structure **302** configured to generate a plurality of compressed magnetic fields. The magnetic structure **302** has a case **304** with tabs **305** holding a plurality of magnets in position with respect to each other. The magnets include a first magnet **308** having a first end **318** having a first polarity and a second end **316** having an opposite polarity. The magnets further include a second magnet **310** having a first end **320** having a polarity that is the same as the first polarity, and having a second end **322** having a polarity that is opposite the first polarity. The first end **320** of the second magnet **310** is spaced apart from the first end **318** of the first magnet **308**. The first end **320** of the second magnet **310** is at least generally facing the first end **318** of the first magnet **308**. The magnets further include a third magnet **312** having a first end **324** having a polarity that is opposite the first polarity, and having a second end **326** having a polarity that is the same as the first polarity. The first end **324** of the third magnet **312** is spaced apart from the second end **322** of the second magnet **310**. The first end **324** of the third magnet **312** is at least generally facing the second end **322** of the second magnet **310**. Any number of additional magnets is possible. For example, in the illustrated embodiment, the magnets further include a fourth magnet **314** having a first end **328** having a polarity that is the same as the first polarity and having a second end **330**. The magnet assembly **302** further includes a screw cap **306** closing an open threaded end of the case **304**.

FIG. **19** is a side diametric cross-sectional view of an embodiment of a multi-pole magnet structure **1900** configured to generate a plurality of compressed magnetic fields. The magnet structure **1900** comprises a plurality of magnets **1902**, **1904**, **1906** with concave-shaped surfaces **1908**, **1910**, **1912**, **1914** with like poles held facing each other at selected distances apart so as to generate high gradient or compressed magnetic fields. As illustrated, the concave-shaped surfaces are cone-shaped. Other substantially concave-shaped surfaces may be employed.

FIG. **20** is a side diametric cross-sectional view of another embodiment of a multi-pole magnet structure **2000** configured to generate a plurality of compressed magnetic fields. The magnet structure **2000** comprises a plurality of magnets **2002**, **2004**, **2006** with convex-shaped surfaces **2008**, **2010**, **2012**, **2014** with like poles held facing each other at selected distances apart so as to generate high gradient or compressed magnetic fields. As illustrated, the convex-shaped surfaces are curved. Other substantially convex-shaped surfaces may be employed.

FIG. **21** is a side view of an embodiment of another multi-pole magnet structure **2100** configured to generate a plurality of compressed magnetic fields. The magnet structure comprises a rectangular magnet housing **2102** and a plurality of rectangular magnets **2104**, **2106**, **2108** contained within the housing **2102**. The magnets **2104**, **2106**, **2108** are held with like poles facing each other at selected distances apart so as to generate desired compressed magnetic fields.

Embodiments of dual-conductor or bi-metal coils and/or embodiments of magnet structures configured to generate compressed magnet fields, such as those described above, may be advantageously employed in a number of devices and applications. For example, embodiments of dual-conductor or bi-metal coils and/or embodiments of magnet structures configured to generate compressed magnet fields may be used in various types of generators/motors used in various applications, acoustic systems and/or control systems. Example generators include generators may be configured to convert energy received in a parasitic manner or energy specifically generated to be converted into electrical energy. Typical sources of energy include kinetic sources, thermal sources, acoustic sources, and radio-frequency sources. For example, some embodiments may employ a magnetic structure configured to generate a compressed magnetic field together with dissimilar metals in order to take advantage of the Seebeck effect.

A number of such example applications are discussed below by way of illustrative example embodiments of such devices and applications. Although some embodiments may employ a dual-conductor or bi-metal coil and a magnet structure configured to generate a compressed magnetic field, other embodiments may employ a dual-conductor or bi-metal coil and a conventional magnetic structure or no magnetic structure. Other embodiments may employ a magnetic structure configured to generate a compressed magnetic field and a conventional coil or no coil. Some embodiments may employ a conventional coil and a conventional magnetic structure in combination with other aspects of the present disclosure.

Linear generators and motors are known in the art. A linear generator typically has a stator and an armature that can be linearly driven relative to the stator to generate electrical energy. Linear generators are disclosed, for example, in U.S. Pat. No. 6,759,755 to Sagov and in U.S. Pat. No. 6,798,090 to Cheung et al., both of which are incorporated herein by reference. A linear motor typically has a stator and an armature that can be linearly driven relative to the stator in response to the application of electrical energy, typically in the form of electrical signals.

Conversion of linear motion to electrical power is a challenging problem. Recent work by the inventors in evaluating classical linear displacement generators using planar inductors indicates poor conversion efficiencies. See, for example, U.S. Pat. No. 6,220,719 issued to Vetorino, et al. The basic problem is that the power output is proportional to the square of the derivative of the magnetic field, and the magnitude of this derivative remains small in conventional devices. Similar issues arise in the conversion of electrical power into linear motion.

In a linear generator, the power output generated by relative movement of a coil with respect to a magnetic structure is proportional to the square of the derivative of the magnetic field. The voltage is determined by the number of turns in the winding of the coil and the strength of the magnetic field. The shape, relative position and strength of the permanent magnets in a magnetic structure can magnify the value of that derivative by generating a compressed magnetic field. By using a compressed magnetic field, significant increases in efficiency can be obtained from this class of generators, even for relatively small rates of mechanical displacement. The concepts pertinent to generating a compressed magnetic field are addressed through illustrative examples (see the description of FIGS. **13**-**16** above and **22**-**25** below).

FIGS. **22** through **24** illustrate an embodiment of a linear power generator **2200** employing an embodiment of a magnetic structure **2202** configured to generate a compressed magnetic field. The power generator **2200** includes a coil **11** located between two magnets **12** and **14**. The coil may be a conventional coil or a dual-conductor or bi-metal coil. More particularly, the power generator **2200** includes a first magnet **12** having an end **13** having a first polarity, and a second magnet **14** having an end **15** having a polarity that is the same as the first polarity. More particularly, in the illustrated embodiment, the end **13** is a north pole and the end **15** is also a north pole. The end **15** of the second magnet **14** is spaced apart from the end **13** of the first magnet **12**. In the illustrated embodiment, the end **15** has a surface **22** (FIG. **23**) that is generally planar, and the end **13** has a surface **18** that is generally planar. The end surface **22** of the second magnet **14** is at least generally facing the end surface **18** of the first magnet **12**.

The at-rest position of the coil **11** is closer to the end **13** of the first magnet **12** than the end **15** of the second magnet **14**. In the illustrated embodiment, the magnets **12** and **14** are permanents magnets. Other embodiments may employ electro-magnets. Note that the static magnetic flux through the coil **11** is fairly high, as indicated by the density of the equipotential lines **16** passing through the coil **11**, in FIG. **22**. The flux through the surface area of the face **18** of the magnet **12** is very large. The flux through a plane approximately in the center **20** between magnets **12** and **14** is small.

There is a very high negative field gradient between the geometric position of the coil **11** in FIG. **22** and the position occupied by the coil in FIG. **23**. Thus, even slow physical movement of the coil **11** (or, conversely, the magnets) will generate large derivatives. The coil **11** moving back and forth about the center **20** gives a huge change in flux. This will be true even with the small rates of physical displacement (spatial derivatives) of magnet position. Because output is proportional to the square of this derivative, significant increases in power production will result.

If the coil **11** is moved from proximate face **18** of magnet **12** to face **22** of magnet **14** in time 2Δt, the flux will change from $+\phi_{max}$ to $-\phi_{max}$.

Thus, the dφ/dt is approximately:

$$\partial\phi_{max}/\partial\Delta t=\phi_{max}/\Delta t$$

and it is positive.

This is an approximate value because, for a linear velocity, the derivative will vary in value during the period Δt since the field is non-linear.

In the embodiment shown in FIG. **24**, the coil **11** is wound around a core **24** having a length, and an axis **26** along the length (coinciding with the center line **20** in FIG. **23**). In the embodiment shown in FIG. **19**, the axis **26** is normal to the direction defined between the end **13** of the first magnet **12** and the end **15** of the second magnet **14**. In the embodiment of FIG. **22**, the end surface **18** of the first magnet **12** has a width W, and the core **24** has a length along axis **26** that is at least as long as the width W of the end surface **18** of the first magnet **12**. In the illustrated embodiment, the end surface **22** of the second magnet **14** has a width corresponding to the width of the end surface **18**. Other embodiments are possible.

In the illustrated embodiment, the coil **11** is supported to be driven back and forth between the first magnet **12** and the second magnet **14** along a path between surfaces **18** and **22** that is generally normal to the axis **26**.

In an alternative embodiment, shown in FIGS. **25** and **26**, the end **30** of a first magnet **32** is a south pole and the end **34** of a second magnet **36** is also a south pole (and the ends **30** and **34** at least generally face each other). A coil **38** is shown between the two magnets in FIG. **25**, but other coil arrangements are possible in a generator as will be described herein.

Holding the magnets separate and closer together than the ambient distance that the repelling force from the magnets would normally permit creates a high-gradient, or compressed, magnetic field. This generally results in an increase of power output from the generator. For many embodiments, holding the magnets closer together up to a limit will result in an increased power output. For example, in alternative embodiments, the distance between face **18** and face **22** could be equivalent to two times the distance “a” shown in FIG. **22**.

FIG. **27** is a diametric cross-sectional view of a generator **200** employing an embodiment of a magnet structure **202** configured to generate a high-gradient or compressed magnet field. For example, embodiments of the magnetic structures illustrated in FIGS. **16** through **21** may be employed as the magnetic structure **202** illustrated in FIG. **27**. The generator **200** includes a housing **208** in which the magnet structure **202** is supported for sliding movement. In the illustrated embodiment, the case **204** has an outer cylindrical surface and the housing **208** has a cylindrical inner surface, which has a diameter slightly larger than the diameter of the outer cylindrical surface of the case **204**. The outside of the case **204** and the inside of the housing **208** may be made of or coated with dissimilar materials to reduce potential for binding between the case **204** and the housing **208**. For example, the case **204** may be coated with a non-stick coating while the housing **208** may be made of an ABS plastic. Example dissimilar materials are available under the respective trademarks Teflon® and Lexan®.

The generator **200** further includes an end **210**, which can be a threaded end cap, for example, closing an open end of the housing **208**. The generator **200** further includes a spring **212** supported by the end **210** configured to be selectively compressed by the magnet assembly **202** and to move the magnet assembly **202** away from the end **210**. The generator **200** further includes an end **214** which could be a threaded end cap or merely a closed end, and a spring **216** configured to be selectively compressed by the magnet assembly **202** and to move the magnet assembly **202** away from the end **214**. In some embodiments, the springs **212**, **216** may be configured to remain in a compressed state.

The generator **200** further includes at least one coil **218** supported by the housing. While other coil positions are possible, in the illustrated embodiment, the housing **208** has an outer surface, which is cylindrical in the illustrated embodiment, and the coil is wrapped around the outer surface of the housing **208**. The coil **218** is positioned radially outwardly of the housing **208** and the magnet assembly **202** inside the housing **208**. The coil **218** can be retrained against longitudinal movement relative to the housing **208** by glue, grooves, notches, or protrusions in the housing, or by any other desired method, or can be molded into the housing, supported on the interior of the housing, etc. The coil **218** is positioned to be acted on by the compressed magnetic fields generated by the magnetic structure **202**.

In some embodiments, the generator **200** as an assembly is merely supported in a location that would be exposed to motion. In other embodiments, a mechanical linkage is provided to couple the generator **200**, as an assembly, to motion. For example, the bottom **214** could be coupled to a source of motion or movement. In some embodiments, period maintenance could be facilitated. For example, the top **210** could be removable for cleaning or maintenance or replacement of magnets, if desired. Some embodiments may be maintenance free. For example, embodiments of a generator **200** employed in a battery (See FIGS. **33-35**), may be designed to last for the life of the battery without periodic maintenance. For example, the generator **200** may be evacuated and hermetically sealed in some embodiments.

In some embodiments, an accelerometer is provided in the desired application and the frequency constant of the motion is determined. The spring constants of the springs and mass of the magnets are then customized so that the magnet assembly **202** resonates in the housing **208** when there is energy available.

FIG. **28** shows a generator **300**, which is similar to the generator **200** except that multiple coils are used. The generator **300** has a magnetic structure or assembly **302** configured to generate one or more compressed magnetic fields. Multi-pole magnetic structures configured to generate multiple compressed magnetic fields, such as embodiments of the magnetic structures illustrated in FIGS. **16** and **18** through **21**, may be advantageously employed as the magnetic structure **302** in embodiments of the generator **300**. The magnet structure **302** includes a case **304**.

The generator **300** further includes a housing **332** in which the magnet structure **302** is supported for linear motion. In the illustrated embodiment, the case **304** has an outer cylindrical surface and the housing **332** has a cylindrical inner surface, which has a diameter slightly larger than the diameter of the outer cylindrical surface of the case **304**. The generator **300** further includes an end **334**, which can be a threaded end cap, for example, closing an open end of the housing **332**. The generator **300** further includes a spring **346** supported by the end **334**. The spring **346** is configured to be selectively compressed by the magnet assembly **302** and to move the magnet assembly **302** away from the end **334**. The generator **300** further includes an end **338** which could be a threaded end cap or merely a closed end, and a spring **340** is arranged to be selectively compressed by the magnet assembly **302**. The spring **340** is arranged to move the magnet assembly **302** away from the end **338**.

The generator **300** further includes a first coil **336** supported relative to the magnets such that the first coil **336** is selectively acted on by fields from at least one pair of opposed ends of magnets, but possibly by fields from additional pairs of opposed ends of magnets, depending on movement of the magnet assembly **302**. The generator **300** further includes a second coil **342** supported relative to the magnets such that the second coil **342** is selectively acted on by fields from at least one pair of opposed ends of magnets, but possibly by fields from additional pairs of opposed ends of magnets, depending on movement of the magnet assembly **302**. In the illustrated embodiment, the generator **300** further includes a third coil **344** supported relative to the magnets such that the third coil **344** is selectively acted on by fields from at least one pair of opposed ends of magnets, but possibly by fields from additional pairs of opposed ends of magnets, depending on movement of the magnet assembly **302**. Any number of coils can be employed. Any number of pairs of opposed ends of magnets can be employed to act on one or more coils.

FIG. **29** illustrates an embodiment of a system **2900** employing a magnetic structure configured to generate a compressed, high gradient magnetic field and a dual conductor or bi-metal coil. FIG. **30** is a diametric side cross sectional view of the system **2900** of FIG. **29** taken along line **30-30**. FIG. **30** is not to scale with respect to FIG. **29** and some of the detail is omitted from FIG. **30** to facilitate illustration. The system **2900** comprises a rotor **2902** comprising one or more bi-metal coils **2904**. Each bi-metal coil **2904** comprises an electrical conductive winding **2903** and a magnetic conductive winding **2905**. Embodiments of the bi-metal coils illustrated in FIGS. **2-3** and **5-11** may be advantageously employed in embodiments of the system **2900** illustrated in FIG. **29**. Some embodiments may comprise a magnetic structure configured to generate a compressed magnetic field and a conventional coil. Other embodiments may comprise a bi-metal coil and a conventional magnetic structure.

The **2900** also comprises a stator **2906** comprising a magnet support **2908** and a plurality of permanent magnets **2910**, **2912**, **2914**. A first magnet **2910** of the plurality of magnets is coupled to a central portion **2916** of the magnet support **2908**. The first magnet **2910** is oriented such that its poles **2918**, **2920** face opposite sides of an inner circumference **2922** of the rotor **2902**. The second magnet **2912** in the plurality of magnets is coupled to a first outer portion **2924** of the magnet support **2908**. The second magnet **2912** is oriented such that a pole **2926** of the second magnet **2912** faces the like pole **2918** of the first magnet **2910**. As illustrated, the like poles **2918**, **2926** are the south poles of the respective first and second permanent magnets **2910**, **2912**. The third magnet **2914** in the plurality of magnets is coupled to a second outer portion **2928** of the magnet support **2908**. The third magnet **2914** is oriented such that a pole **2930** of the third magnet **2914** faces the like pole **2920** of the first magnet **2910**. As illustrated, the like poles **2920**, **2930** are the north poles of the respective first and third permanent magnets **2910**, **2914**. The magnets **2910**, **2912**, **2914** are positioned such that a plurality of compressed magnetic fields are generated. In the illustrated embodiment the rotor **2902** is coupled to a mechanical transmission system **2934**. In some embodiments, the magnet support **2906** may be part of the rotor and the bi-metal coils **2904** may be part of the stator.

As illustrated, the system **2900** comprises a coupling **2950** for coupling the coils **2904** to a power grid **2952**. Details of the electrical connection **2954** between the coils **2904** and the coupling **2950** are omitted for clarity. A bus system coupled to the electrical conductive windings **2903**, for example, may be employed as the electrical connection **2954** between the coils **2904** and the coupling **2950**. The coupling **2950** may comprise control and/or conditioning modules (not shown).

In some embodiments, the system **2900** may be configured to operate as a generator. In such embodiments, force applied to the rotor **2902** by the mechanical transmission system **2934** may cause the rotor **2902** to rotate with respect to the stator

2906. As the rotor **2902** rotates with respect to the stator **2906** along an axis **2932** illustrated by the dashed line B-B, a three-phase alternating current may be generated by the system **2900**.

In some embodiments, the system **2900** may be configured to operate as a motor. In such embodiments, an electrical signal applied to the coils **2904** may cause the rotor **2902** to rotate with respect to the stator **2906**. As the rotor **2902** rotates with respect to the stator **2906** along an axis **2932** illustrated by the dashed line B-B, a force is applied to the mechanical transmission system **2934** by the rotor **2902**. In some embodiments, the system **2900** may be configured to selectively operate as a motor or as a generator. In some embodiments, the system **2900** may be advantageously configured to operate at a desired voltage level, in a desired voltage range, and/or at a desired frequency. For example, the system **2900** may be configured to produce 110-120 volts AC at 50/60 Hz, 220-240 volts AC at 50/60 Hz, 10 kV AC at 50/60 Hz, or 100 kV at 50/60 Hz. In some embodiments, the system **2900** may be figured to produce alternating and/or direct current.

FIG. **31** illustrates a side cross-sectional view of an embodiment of a system **3100** employing a dual conductor or bi-metal coil **700** and a magnetic structure **202** configured to generate a compressed, high gradient magnetic field. For convenience, the system **3100** will be described with respect to the bi-metal coil **700** illustrated in FIGS. **7** through **9** and the magnetic structure **202** illustrated in FIG. **17**. Other embodiments of bi-metal coils and magnetic structures configured to generate compressed magnetic fields may be employed in embodiments of the system **3100**.

FIG. **31** is not necessarily to scale for ease of illustration. The bi-metal coil **700** comprises a layer of insulating material **702** with an upper surface **704**. The layer of insulating material **702** may comprise, for example, an integrated circuit board, a substrate or a thin film or sheet of insulation. Commercially available insulating materials are sold under the trademark Mylar®. An electrical conductive trace **706** is formed on the upper surface **704** of the layer of insulating material **702**. The electrical conductive trace **706** may comprise any suitable electrical conductive material, such as, for example, copper, aluminum, gold, and silver, and alloys. The materials discussed above with respect to the electrical conductive winding **204** of FIG. **2** may be employed. Well-known techniques for forming traces on substrates may be employed, such as those used in connection with RFID devices and antennas. The layer of insulating material **702** has an opening **708** through which the magnetic structure **202** may move. An upper portion **712** of a suspension system **714** is coupled to the magnetic structure **202** and to the upper surface **704** of the layer of insulating material **702**. The layer of insulating material **702** has a lower surface **716**. A magnetic conductive trace **718** is formed on the lower surface **716** of the layer of insulating material **702**. The magnetic conductive trace **718** may comprise any suitable magnetic conductive material, such as, for example, nickel, nickel/iron alloys, nickel/tin alloys, nickel/silver alloys. The materials discussed above with respect to the magnetic conductive winding **206** of FIG. **2** may be employed. Well-known techniques for forming traces on substrates may be employed, such as those used in connection with RFID devices and antennas. A lower portion **720** of the suspension system **714** is coupled to the magnetic structure **202** and to the lower surface **716** of the layer of insulating material **702**. The suspension system **714** is coupled to an optional mechanical transmission system **3102**.

In some embodiments, the system **3100** may be configured to operate as a generator. In such embodiments, mechanical force applied to the suspension system **714** by the mechanical transmission system **3102** may cause the magnetic structure **202** to move linearly with respect to the bi-metal coil **700**, which may cause the device to generate an electric current. In some embodiments, the system **3100** may be configured to operate as a motor. In such embodiments, an electrical signal applied to the bi-metal coil **700** may cause the magnetic structure **202** to move linearly with respect to the bi-metal coil **700**, which may cause the suspension system **714** to apply a mechanical force to the mechanical transmission system **3102**. In some embodiments, the system **3100** may be configured to selectively operate as a motor or as a generator.

FIG. **32** illustrates a side cross-sectional view of an embodiment of a system **3200** employing a dual conductor or bi-metal coil **1100** and a magnetic structure **1604** configured to generate a compressed, high gradient magnetic field. For convenience, the system **3200** will be described with respect to the bi-metal coil **1100** illustrated in FIG. **11** and the magnetic structure **1604** illustrated in FIG. **16**. Other embodiments of bi-metal coils and magnetic structures configured to generate compressed magnetic fields may be employed in embodiments of the system **3200**.

FIG. **32** is not necessarily to scale for ease of illustration. The bi-metal coil **1100** comprises a plurality of insulating layers **1102**. Traces of electrical conductive material **1106** are formed on selected surfaces **1130**, **1132**, **1134**, **1136** of the plurality of insulating layers **1102**. Traces of magnetic conductive material **1118** are formed on selected surfaces **1138**, **1140** of the plurality of insulating layers **1102**. A suspension system **1114** moveably suspends a magnetic structure **1900** in an opening **1108** in the plurality of layers of insulating material **1102**. As illustrated, the bi-metal coil **1100** comprises three layers of insulating material **1102**. Fewer or additional layers **1102** may be employed. In addition, in some embodiments the traces of electrical conductive material **1106** and the traces of magnetic conductive material **1118** may be formed on selected surfaces of the layers of insulating material **1102** in different patterns. For example, traces of electrical conductive material **1106** and traces of magnetic conductive material **1118** may be formed on alternate surfaces of layers of insulating material. In another example, traces of electrical conductive material **1106** and magnetic conductive material **1118** may be formed on the same surface of a layer of insulating material **1102**, or on each surface of a layer of insulating material. Traces on various layers of insulating material **1102** may be coupled to each other. The suspension system **1114** is coupled to an optional mechanical transmission system **3202**.

In some embodiments, the system **3200** may be configured to operate as a generator. In such embodiments, mechanical force applied to the suspension system **1114** by the mechanical transmission system **3202** may cause the magnetic structure **1900** to move linearly with respect to the bi-metal coil **1100**, which may cause the system to generate an electric current. In some embodiments, the system **3200** may be configured to generate an alternating current and/or a direct current. In some embodiments, the device system may be configured to operate as a motor. In such embodiments, an electrical signal applied to the bi-metal coil **1100** may cause the magnetic structure **1900** to move linearly with respect to the bi-metal coil **1100**, which may cause the suspension system **1114** to apply a mechanical force to the mechanical transmission system **3202**. In some embodiments, the system **3200** may be configured to selectively operate as a motor or as a generator.

Battery technology is a one example application in which embodiments of a bi-metal coil and/or a magnetic structure configured to generate a compressed magnetic field may be advantageously employed, as will be illustrated with a limited number of examples.

FIG. **33** is a diametric cross-sectional view of an embodiment of a battery **100** comprising a case **102**, a generator **104**, a first energy storage device **106**, a control module **108**, a second energy storage device **110**, and contact terminals **112**, **114**. The case **102** as illustrated is cut-away so as to facilitate illustration of other components of the battery **100**. The case **102** contains the generator **104**, the first energy storage device **106**, the control module **108**, and the second energy storage device **110**. The contact terminals **112**, **114** are mounted to the case **102** at a top **116** and bottom **118**, respectively, of the battery **100**.

The case **102** may comprise an outer case shielding **120**, which may be a magnetic and/or electrical shield. The case shielding **120** may comprise, for example, a layer of tin foil, a layer of a magnetic shielding material, such as, for example, nickel, nickel/iron alloys, nickel/tin alloys, nickel/silver alloys, nickel/iron/copper/molybdenum alloys, which may also take the form of a foil. Such foil layers may, for example, have a thickness in the range of 0.002-0.004 inches. Magnetic shielding materials are commercially available under several trademarks, including MuMetal®, Hipernom®, HyMu 80®, and Permalloy®.

In some embodiments, the case **102** and contact terminals **112**, **114** may take the external configuration of those of a conventional battery, such as, for example, a AA-cell, a AAA-cell, a C-cell, a D-cell, a 9-volt battery, a watch battery, a pacemaker battery, a cell-phone battery, a computer battery, and other standard and non-standard battery configurations. Embodiments of the battery **100** may be configured to provide desired voltage levels, including, for example, 1.5 volts, 3.7, 7.1, 9-volts, and other standard and non-standard voltages. Embodiments may be configured to provide direct and/or alternating current.

The generator **104** converts kinetic energy into electrical energy. As illustrated the generator **104** is a linear generator comprising a bi-metal coil **122**, a magnetic structure **124** and a suspension system **126**. As illustrated, the bi-metal coil **122** comprises an electrical conductive winding **121** and a magnetic conductive winding **123**. As illustrated, the suspension system **126** comprises a magnetic structure carrier guide **128**, a first spring **130** coupled at one end **132** to the magnetic structure **124**, a first repelling magnet **134** coupled to the other end **136** of the first spring **130**, a second spring **138** coupled at one end **140** to the magnetic structure **124**, and a second repelling magnet **142** coupled to the other end **144** of the second spring **138**. The suspension system **126** facilitates movement of the magnetic structure **124**, in response to movement of the battery, along an axis A-A with respect to the coil **122**. The movement of the magnetic structure **124** relative to the coil **122** generates a current in the coil **122**. The suspension system **126** may comprise, for example, stainless steel springs, such as **304** or **316** stainless steel springs. The magnetic structure **124** may comprise, for example, one or more rare earth magnets, such as neodymium-iron-boron permanent magnets, one or more ceramic magnets, one or more plastic magnets, or one or more other magnets. The repelling magnets **132**, **142** may comprise, for example, one or more rare earth magnets, one or more ceramic magnets, one or more plastic magnets, or one or more other magnets. As illustrated, the carrier guide **128** comprises a winding form **146** upon which one or more windings of the coil **122** are wound. In some embodiments, a separate winding form may be employed. The suspension system **126** is configured to permit the magnetic structure **124** to pass completely out of a region **148** defined by a top **150** and a bottom **152** of the coil **122**. The springs **130**, **138** are typically configured in a loaded condition.

The first energy storage device **106** is configured to store electrical energy generated by the generator **104**. In one embodiment, the first energy storage device **106** is capable of storing electrical energy generated by the generator **104** with little or no conditioning. In other embodiments, electrical energy may be conditioned before it is stored in the first energy storage device **106**, as discussed by way of example below. The first energy storage device **106** may comprise, for example, one or more ultracapacitors. For ease of illustration, the first energy storage device **106** is illustrated as a functional block.

The control module **108** controls the transfer of energy within the battery **100**. The control module **108** typically comprises a rectifier, which as illustrated is a full bridge rectifier **109**. For example, the control module **108** may be configured to control the transfer of energy between various components of the battery **100**, such as the generator **104**, the first energy storage device **106**, the second energy storage device **110**, and the contact terminals **112**, **114**. In one embodiment, the control module **108** may also control the transfer of energy from the generator **104** to the first energy storage device **106**. In one embodiment, the control module **108** controls the transfer of energy stored in the first energy storage device **106** to the second energy storage device **110**. For example, the control module **108** may limit the current flow from the first energy storage device **106** to the second energy storage device **110**. In another example, the control module **108** may stop the transfer of energy from the first energy storage device **106** to the second energy storage device **110** to avoid overcharging the second energy storage device **110**. In one embodiment, the control module **108** may be configured to stop the transfer of energy to the first energy storage device **106** to avoid overcharging the first energy storage device **106**. In one embodiment, the control module **108** may be configured to control the transfer of energy from the first energy storage device **106** to the contact terminals **112**, **114**. In one embodiment, the control module may be configured to control the transfer of energy from the generator to the contact terminals **112**, **114**. In one embodiment, the control module **108** also may be configured to detect, control, permit, accept, regulate and/or to facilitate charging of the first energy storage device **106** and/or the second energy storage device **110** from an external source of electrical energy, such as a conventional battery charger (not shown), or ambient sources of energy. In one embodiment, the control module **108** is configured to condition energy during a transfer. The operation of the control module **108** in two exemplary embodiments is discussed in more detail below in the description of FIGS. **37** and **38**.

The control module **108** may be implemented in a variety of ways, including as a combined control system or as separate subsystems. The control module **108** may be implemented as discrete circuitry, one or more microprocessors, digital signal processors (DSP), application-specific integrated circuits (ASIC), or the like, or as a series of instructions stored in a memory and executed by a controller, or various combinations of the above. In some embodiments, the first energy storage device **106** may be integrated into the control module **108**.

The second energy storage device **110** is configured to store electrical energy transferred from the first energy storage device **106** under the control of the control module **108**. The second energy storage device **110** may comprise, for example, one or more conventional batteries, such as a lead-

acid battery, a nickel-cadmium battery, a nickel-metal hydride battery, a lithium polymer battery or lithium ion battery, a sodium/sulfur battery, or any suitable rechargeable energy storage device.

The contact terminals **112**, **114** provide access for transferring electrical energy to and/or from the battery **100**. The contact terminals **112**, **114** may be made of any electrically conductive material, such as, for example, metallic materials, such as copper, copper coated with silver or tin, aluminum, gold, etc. The contact terminals **112**, **114** are coupled to the control module **108**. In some embodiments, the contact terminals **112**, **114** may be coupled to the second energy storage device **110**, instead of being directly coupled to the control module **108**. As illustrated, the contact terminals **112**, **114** have a physical configuration similar to the contact terminals of a conventional C-cell battery. As discussed above, other configurations may be employed. The contact terminals **112**, **114** are configured to permit the battery **100** to be easily installed into and removed from external devices, such as, for example, a radio, a cell phone, or a positioning system. The contact terminals **112**, **114** may employ magnetic shielding.

Energy may be stored in the battery **100** as a result of movement of the battery **100**. For example, if the magnetic structure **124** is neutral with respect to the coil **122** and the battery **100** is subject to a downward movement, the magnetic structure **124** may move up with respect to the coil **122** in response to the downward movement of the battery **100**. The relative upward movement of the magnetic structure **124** will result in the generation of a current in the coil **122** when it passes above the top of the coil **150**. As the magnetic structure **124** approaches the first repulsive magnet **134**, the first spring **130** and the first repulsing magnet **134** will apply downward forces to the magnetic structure **124**. In response to the downward forces, the magnetic structure **124** may begin to move downward with respect to the coil **122**. It may pass through neutral **151**, a location approximately midway between **150** and **152**, and pass through the coil **122** again, generating additional electrical current when it passes below the bottom of the coil **152**. When the magnetic structure **124** approaches the second repulsing magnet **142**, the second spring **138** and the second repulsing magnet **142** will apply upward forces to the magnetic structure **124**. If the upward forces are sufficiently strong, the magnetic structure **124** will again pass through the coil **122** again, and generate additional electrical current. The movement may continue in an oscillatory back and forth fashion until there is insufficient energy in the suspension system **126** to continuing moving the magnetic structure **124** with respect to the coil **122**.

In some embodiments the suspension system **126** may be tuned to increase the electrical energy generated from anticipated sources of energy. For example, if the battery **100** will frequently be in an environment where energy is supplied by an individual walking or running at a known speed or rate, the suspension system **126** may be tuned to that speed or rate. Thus, a battery may be configured to substantially maximize the conversion of energy expected to be generated by a jogger into electrical energy. In another example, if the battery **100** will frequently be subject to stop and go traffic in an automobile or irregular motion from a flight or ground vehicle, the suspension system **126** may be tuned to maximize the conversion of the energy of that environment into electrical energy. In another example, if the battery will be employed in an environment frequently subjected to fluid waves, such as water or sea waves, or wind, the suspension system may be tuned to maximize the conversion of the energy of that environment into electrical energy. In another example, if the battery will be frequently subjected to vibrations, for example, in a moving vehicle, the suspension system may be tuned to maximize the conversion of the energy received from the vibrations into electrical energy. The suspension system may be tuned, for example, by modifying the strength of any repelling magnets, adjusting the tension in any repelling devices, such as springs, employing multiple mechanical repelling devices (see FIG. **36**), modifying the length of the path of travel of the magnetic structure, or combinations of modifications. Other suspension systems may be employed, such as, for example, suspension systems which orient the generator in different directions within the battery. The suspension system **126** may be gimbaled and/or may employ gyroscopic principles to orient the generator to facilitate optimal conversion of energy into electrical energy. Multiple generators within a battery with different orientations may be employed and multiple battery configurations may be employed.

In some embodiments, other generator configurations may be employed, such as, for example, radial, rotational, Seebeck, acoustic, thermal, or radio-frequency generators. In some embodiments, other suspension systems may be employed, such as suspension systems in which the generator **104** may move with respect to the case **102** so as to take maximum advantage of the available forms of energy. For example, the generator **104** may be configured to rotate in the battery case **102**, so as to align itself with an axis of movement. In another example, the suspension system **126** may be configured to allow the coil **122** to move with respect to the magnetic structure **124**.

FIG. **34** is a diametric cross-sectional view of another embodiment of a battery **200** comprising a case **202**, a generator **204**, a first energy storage device **206**, a control module **208**, a second energy storage device **210**, a third energy storage device **211**, and contact terminals **212**, **214**. The case **202** as illustrated is cut-away so as to facilitate illustration of other components of the battery **200**. The case **202** contains the generator **204**, the first energy storage device **206**, the control module **208**, the second energy storage device **210** and the third energy storage device **211**. The contact terminals **212**, **214** are mounted to the case **202** at a top **216** and bottom **218**, respectively, of the battery **200**. The case **202** may comprise an outer case shielding **220**, which may be a magnetic and/or electrical shield. In some embodiments, the case **202** and contact terminals **212**, **214** may take the configuration of those of a conventional battery, such as, for example, a M-cell, a AAA-cell, a C-cell, a D-cell, a 9-volt battery, a watch battery, a pacemaker, a cell-phone battery, a computer battery, and other standard and non-standard battery configurations. Embodiments of the battery **200** may be configured to provide desired voltage levels, as discussed above with respect to the embodiment illustrated in FIG. **33**. For example, the voltage level can be modified by changing the number of turns in a winding (see, e.g., winding **410** of coil **402** in FIG. **36**) of the coil **122**.

The generator **204** converts received energy into electrical energy. As illustrated the generator **204** is a linear generator comprising a coil **222**, a magnetic structure **224** and a suspension system **226**. The generator **204** may operate, for example, as described above with respect to the generator **104** illustrated in FIG. **33**.

The first energy storage device **206** is configured to store electrical energy generated by the generator **204**. In one embodiment, the first energy storage device **206** is capable of storing electrical energy generated by the generator **204** with little or no conditioning. The first energy storage device **206** may comprise, for example, one or more ultracapacitors.

The control module **208** controls the transfer of energy between the various components of the battery **200**, such as the generator **204**, the first energy storage device **206**, the second energy storage device **210**, the third energy storage device **211**, and the terminals **212**, **214**. For example, the control module **208** may control the transfer of stored in the first energy storage device **206** to the second energy storage device **210** and to the third energy storage device **211**. In one embodiment, the control module **208** may also control the transfer of energy from the generator **204** to the first energy storage device **206**. For example, the control module **208** may limit the current flow from the first energy storage device **206** to the second energy storage device **210** and to the third energy storage device **211**. In another example, the control module **208** may stop the transfer of energy from the first energy storage device **206** to the second energy storage device **210** and to the third energy storage device **211** to avoid overcharging the second and third energy storage devices **210**, **211**. In one embodiment, the control module **208** may be configured to detect, control, permit, and/or to facilitate charging of the first, second and/or third energy storage devices **206**, **210**, **211** from an external source of electrical energy (not shown) coupled to the terminals **212**, **214**.

The control module **208** may be implemented in a variety of ways. For example, the control module may be implemented as described above in the description of the control module **108** of FIG. **33**.

The second and third energy storage devices **210**, **211** are configured to store electrical energy transferred from the first energy storage device **206** under the control of the control module **208**. The second and third energy storage devices **210**, **211** may comprise, for example, conventional rechargeable batteries, such as nickel-cadmium batteries, nickel-metal hydride batteries, lithium polymer batteries or lithium ion batteries, other energy storage devices, or combinations of energy storage devices. The second and third energy storage devices may be coupled to the control module **208**, for example, separately, in series, or in parallel. As illustrated, the second and third energy storage devices **210**, **211** are washer-shaped with the suspension system **226** extending into hollow centers **209**, **213** of the second and third energy storage devices **210**, **211**. As illustrated the second and third energy storage device **210**, **211** are connected in series between the first and second contact terminals **212**, **214** and in series to the control module **208**. Some embodiments may employ dissimilar metals to take advantage of the Seebeck effect.

The contact terminals **212**, **214** provide access for transferring electrical energy to and from the battery **200**. The contact terminals **212**, **214** may be made of any electrically conductive material, such as, for example, metallic materials, such as copper, copper coated with silver or tin, aluminum, gold, etc. The contact terminals **212**, **214** are coupled to the second and third energy storage devices **210**, **211**. The second and third energy storage devices **210**, **211** may be coupled to the contact terminals in parallel or in series. In some embodiments, the contact terminals **212**, **214** may be coupled to the control module **208**, instead of being directly coupled to the second and third energy storage devices **210**, **211**. As illustrated, the contact terminals **212**, **214** have the physical configuration of the contact terminals of a conventional C-cell battery. As discussed above, other configurations may be employed. The contact terminals **212**, **214** are typically configured to permit the battery **200** to be easily installed into and removed from external devices, such as, for example, a radio, a cell phone, or a positioning system. The contact terminals **212**, **214** may employ magnetic shielding.

Energy may be stored in the battery **200** as a result of movement of the battery **200**. For example, energy may be converted into stored energy in a manner similar to the example discussed above with respect to FIG. **33**.

As discussed above, in some embodiments the suspension system **226** may be tuned to maximize the electrical energy generated from anticipated sources of kinetic energy.

In some embodiments, other generator configurations may be employed, such as, for example, rotational generators. In some embodiments, other suspension systems may be employed, such as, for example, suspension systems in which the generator **204** may move with respect to the case **202** so as to take maximum advantage of available kinetic energy. For example, the generator **204** may be configured to spin in the battery case **202**, so as to align itself with an axis of movement. In another example, the suspension system **226** may be configured to allow coil **222** to move with respect to the magnetic structure **224**.

FIG. **35** is a side sectional view of another embodiment of a battery **300** comprising a case **302**, a generator **304**, a first energy storage device **306**, a control module **308**, a second energy storage device **310**, and contact terminals **312**, **314**. The battery **300** has a different configuration than the battery **100** illustrated in FIG. **33**, but the operation of the battery **300** is typically similar to the operation of the battery **100** illustrated in FIG. **33**. The contact terminals **312**, **314** may be made of any electrically conductive material, such as, for example, metallic materials, such as copper, copper coated with silver or tin, aluminum, gold, etc. In some embodiments, the contact terminals **312**, **314**, may be contained within a connector, such as a plastic connector.

FIG. **36** is a diametric cross-sectional view of a generator **400** suitable for use, for example, in the embodiments illustrated in FIGS. **33** through **35**. Other generators and/or devices may be employed in the embodiments illustrated in FIGS. **33** through **35**, such as, for example, the embodiments illustrated in FIGS. **22**-**32**. The generator comprises a coil **402**, a magnetic structure **404** configured to generate a compressed magnetic field (see, e.g., FIGS. **15** through **21**) and a suspension system **406**. The suspension system **406** is configured to allow the magnetic structure **404** to completely pass through the coil **402** in either direction. As illustrated, the generator **400** is a linear generator.

The coil **402** comprises a cylindrical winding form **408** and one or more windings **410**. As illustrated, the winding form **408** is integrated with a carrier guide **409** of the suspension system **406**. As illustrated, the coil **402** comprises a single winding **410**. The winding **410** may comprise any electrically conductive and substantially non-magnetic conductive material, such as, for example, copper, aluminum, gold, and silver, and alloys. The winding **410** is typically covered with an insulating material **411**. In some embodiments, additional windings comprising magnetic conductive and/or non-magnetic conductive material may be employed (See, e.g., FIGS. **2**-**11**). The winding **410** may be, for example, solid material or may be comprised of strands of wire. Sheets of material may be employed in some embodiments. For example, a sheet comprising a copper layer and a Mylar® layer may be wound around the winding form **408**.

The magnetic structure **404** comprises a plurality of permanent magnets **412**, **414**, **416** contained within a cylindrical magnet housing **418**. While the illustrated embodiment employs three permanent magnets **412**, **414**, **416**, other embodiments of the generator **400** may employ different numbers of permanent magnets, such as two permanent magnets, four permanent magnets or hundreds of permanent magnets. The permanent magnets **412**, **414**, **416** are disk-shaped as illustrated, but other shapes may be employed. For example, rectangular- (e.g., square), spherical-, or elliptical-shaped magnets may be employed. Similarly, the faces of the magnets need not be flat. For example, convex-, concave-, radial-, cone-, or diamond-shaped faces may be employed. Various combinations of shapes and faces may be employed. In some embodiments, electromagnets may be employed. The inside **420** of the magnet housing **418** and outsides **422**, **424**, **426** of the permanent magnets **412**, **414**, **416** are threaded so that the permanent magnets **412**, **414**, **416** can be fixed in position with respect to each other within the magnet housing **418**. Other positioning mechanisms may be employed, such as tabs, spacers, glues, or combinations of positioning mechanisms.

The magnets **412**, **414**, **416** are positioned and held apart from each other and are arranged such that same polarity poles in adjacent permanent magnets face each other. For example, the N pole **428** of the first permanent magnet **412** faces the N pole **430** of the second permanent magnet **414** and the S pole **432** of the second permanent magnet **414** faces the S pole **434** of the third permanent magnet **416**. In addition, the magnets **412**, **414**, **416** are held close enough together to form a compressed magnetic field (see the discussions of FIGS. **15** and **16**). In some embodiments, the spaces **436**, **438** between the permanent magnets **412**, **414**, **416** are substantially filled with a material **437**, which may comprise a gas such as air. In some embodiments, the material **437** may comprise other substantially non-magnet and substantially non-conductive substances, such as a fluoropolymer resin or plastic. In some embodiments, the magnetic structure may be evacuated and hermetically sealed.

As noted above, the shape, position and strength of the permanent magnets in a magnetic structure, such as the magnetic structure **404**, can increase the efficiency of the generator **400** by generating a compressed magnetic field. The ratio of the length **440** from the top **442** of the first permanent magnet **412** to the bottom **444** of the third permanent magnet **416** to the length of **446** of the inner diameter **448** of the winding form **408**, also impacts the electrical current produced in response to movement of the magnetic structure **404** with respect to the coil **402**. A gauss meter (not shown) may be employed to determine the optimum strength and positioning of the permanent magnets **412**, **414**, **416**, as well as the number of permanent magnets and the length **440**.

Other design considerations may be taken into consideration as well, such as weight and reducing external impacts of electromagnetic fields and impacts from external electromagnetic fields. In another example of an additional design consideration, the overall length **450** of the winding form **408** and the range of movement of the magnetic structure **404** in the suspension system may impact the stability of the generator **400**. In one experimental embodiment, the first magnet **412** and the third magnet **416** had a strength of 450 gauss and the second magnet had a strength of 900 gauss and the permanent magnets **412**, **414**, **416** were separated by 2 mm. Factors in determining the desired spacing include the magnetic B-field strength. The repelling magnets **460**, **462** each had a strength of 600 gauss. In another experimental embodiment, the first magnet **412**, second magnet **414** and third magnet **416** had a strength of 12,600 gauss and the permanent magnets **412**, **414**, **416** were separated by 4-5 mm. The repelling magnets **460**, **462** each had a strength of 9906 gauss. This resulted in a high-gradient field with a strength of approximately 16,800 gauss.

The inside **452** of the carrier guide **409** and the outside **454** of the magnet housing **418** are made of or coated with dissimilar materials to reduce potential for binding between the winding form **408** and the magnet housing **418**. For example, the carrier guide **409** may be coated with a non-stick coating while the magnet housing **418** may be made of an ABS plastic. Example dissimilar materials are available under the respective trademarks Teflon® and Lexan®. The magnetic housing **418** also comprises a first threaded end cap **456** and a second threaded end cap **458**.

The suspension system **406** comprises a first repelling permanent magnet **460** and a second repelling permanent magnet **462** that are fixed with respect to the coil **402** in the axis of movement **464** of the magnetic structure **404**. The first repelling magnet **460** is positioned such that a like pole of the first repelling magnet **460** faces the like pole of the nearest permanent magnet **412** in the magnetic structure **404**. As illustrated, the S pole **466** of the first repelling magnet **460** faces the S pole **468** of the first permanent magnet **412** of the magnetic structure **404**. Similarly, the second repelling magnet **462** is positioned such that a like pole of the second repelling magnet **462** faces the like pole of the nearest permanent magnet **416** in the magnetic structure **404**. As illustrated, the N pole **470** of the second repelling magnet **462** faces the N pole **472** of the third permanent magnet **416** of the magnetic structure **404**. This arrangement increases the efficiency of the generator in converting kinetic energy into electrical energy and reduces the likelihood that the magnetic structure **404** will stall in the suspension system **406**.

The suspension system **406** also comprises a first spring **474**, a second spring **476**, a third spring **478** and a fourth spring **480**. The first spring **474** is coupled to the first repelling magnet **460** and to the first cap **456** of the magnetic structure **404**. The first spring **474** is typically in a loaded condition. The second spring **476** is coupled to the second repelling magnet **462** and to the second end cap **458** of the magnetic structure **404**. The second spring **476** is typically in a loaded condition. The first and second springs **474**, **476** help to hold the magnetic structure **404** centered in the desired movement path along the axis **464**, and impart forces to the magnetic structure **404** as they are compressed and stretched by movement of the magnetic structure **404** along the axis of movement **464**. The third spring **478** is coupled to the first repelling magnet **460** and imparts a repelling force on the magnetic structure **404** in response to compression forces applied by the magnetic structure **404** as it nears the first repelling magnet **460**. The fourth spring **480** is coupled to the second repelling magnet **462** and imparts a repelling force on the magnetic structure **404** in response to compression forces applied by the magnetic structure **404** as it nears the second repelling magnet **462**. The springs **474**, **476**, **478**, **480** may be tuned to increase the efficiency of the generator in particular applications and likely environments, as discussed in more detail above in the description of FIG. **33**. The tuning may be done experimentally. Some embodiments may employ no springs, fewer springs, or more springs. For example, in some embodiments springs **478** and **480** may be omitted.

FIG. **37** is a high-level flow diagram illustrating an embodiment of a method **1500** of charging a portable energy storage device, such as the batteries **100**, **200**, **300** illustrated in the embodiments of FIGS. **33** through **35**, in response to movement of the battery. For convenience, the method **1500** will be described with respect to the battery **100** illustrated in FIG. **33**.

The method **1500** begins at **1502** and proceeds to **1504**. At **1504**, the battery **100** receives energy as a result of movement of the battery **100**. The method **1500** proceeds to **1506**. At **1506**, the battery **100** converts the energy into movement of a magnetic structure with respect to a coil inside the battery **100**. The back and forth movement of the magnetic structure through the coil generates an alternating current signal. The magnetic structure may be configured to generate a compressed magnetic field (see, e.g., FIGS. **15** through **21** and **36**). The coil may comprise electrical conductive and magnetic conductive windings (see, e.g., FIGS. **2** through **11**). The method **1500** proceeds from **1506** to **1508**. At **1508**, the battery **100** rectifies the alternating current signal produced by the movement of the magnetic structure with respect to the coil. The method proceeds to **1510**. At **1510**, the battery **100** stores electrical energy from the rectified alternating current signal in a first energy storage device within the battery **100**. The method **1500** proceeds from **1510** to **1512**. At **1512**, the battery **100** controls the transfer of energy stored in the first energy storage device to a second energy storage device within the battery **100**. The method **1500** proceeds from **1512** to **1514**, where the method **1500** stops.

Embodiments of a method to charge a portable energy storage device may perform other acts not shown in FIG. **37**, may not all perform all of the acts shown in FIG. **37**, may combine acts shown in FIG. **37**, or may perform the acts of FIG. **37** in a different order. For example, the embodiment of a method **1500** illustrated in FIG. **37** may be modified to check whether conditions are appropriate for charging second energy storage device before transferring energy from the first energy storage device to the second energy storage device.

FIG. **38** is a high-level flow diagram illustrating an embodiment of a method **1600** of operating a portable energy storage device, such as the batteries **100**, **200**, **300** illustrated in the embodiments of FIGS. **33** through **35**, in response to the presentation of a load or a charge signal to the battery. For convenience, the method **1600** will be described with respect to the battery **100** illustrated in FIG. **33**.

The method **1600** begins at **1602** and proceeds to **1604**. At **1604**, the battery **100** determines whether a load is being presented to the battery **100**. This may be done, for example, by using discrete circuitry. When it is determined at **1604** that a load is being presented to the battery **100**, the method proceeds from **1604** to **1606**. When it is determined at **1604** that a load is not being presented to the battery **100**, the method **1600** proceeds from **1604** to **1620**.

At **1606**, the battery **100** determines whether to condition energy from the generator and provide the conditioned energy to the load. This determination may be made, for example, by determining whether the energy being generated by the generator is sufficient to drive the load. Other factors may be considered as well in the determination, such as, for example, load histories, charging and discharging cycles of the energy storage devices in the battery **100**. Discrete circuitry and/or look-up tables may be employed to determine to supply conditioned energy from the generator to the load. When it is determined at **1606** to condition energy from the generator and provide the conditioned energy to the load, the method **1600** proceeds from **1606** to **1608**. When it is determined at **1606** to not provide conditioned energy from the generator to the load, the method **1600** proceeds from **1606** to **1610**. At **1608**, the battery **100** transfers conditioned energy from the generator to the load. The method **1600** proceeds from **1608** to **1604**.

At **1610**, the battery **100** determines whether to transfer energy from the first energy storage device to the load. This determination may be made, for example, by determining whether the energy stored in the first energy storage device is sufficient to drive the load. Other factors may be considered as well in the determination, such as, for example, load histories, and charging and discharging cycles of the energy storage devices in the battery **100**. Discrete circuitry and/or look-up tables may be employed to determine whether to supply energy stored in the first energy storage device to the load. When it is determined at **1610** to supply energy stored in the first energy storage device to the load, the method **1600** proceeds from **1610** to **1612**. When it is determined at **1610** to not transfer energy stored in the first energy storage device to the load, the method **1600** proceeds from **1610** to **1614**. At **1612**, the battery **100** transfers energy stored in the first energy storage device to the load. The method **1600** proceeds from **1612** to **1604**.

At **1614**, the battery **100** determines whether to transfer energy from the second energy storage device to the load. This determination may be made, for example, by determining whether the energy stored in the second energy storage device is sufficient to drive the load. Other factors may be considered as well in the determination, such as, for example, load histories, and charging and discharging cycles of the energy storage devices in the battery **100**. Discrete circuitry and/or look-up tables may be employed to determine whether to supply energy stored in the second energy storage device to the load. When it is determined at **1610** to supply energy stored in the second energy storage device to the load, the method **1600** proceeds from **1614** to **1616**. When it is determined at **1614** to not transfer energy stored in the second energy storage device to the load, the method **1600** proceeds from **1614** to **1618**. At **1616**, the battery **100** transfers energy stored in the second energy storage device to the load. The method **1600** proceeds from **1616** to **1604**.

At **1618**, error processing and/or safety processing for load conditions is performed. For example, the battery **100** may disable the transfer of energy from the battery until the battery has been recharged (either through energy from the generator or through external sources of energy). The method **1600** proceeds from **1618** to **1604**.

At **1620**, the battery **100** determines whether a charge signal is being presented to the battery **100**. This may be done by, for example, using discrete circuitry. When it is determined at **1620** that a charge signal is being presented to the battery **100**, the method proceeds from **1620** to **1622**. When it is determined at **1620** that a charge signal is not being presented to the battery **100**, the method **1600** proceeds from **1620** to **1604**.

At **1622**, the battery **100** determines whether to charge the first energy storage device. This determination may be made based upon factors such as, for example, the characteristics of the charge signal, the energy stored in the energy storage devices, and charging and discharging cycles of the energy storage devices in the battery **100**. Discrete circuitry and/or look-up tables may be employed to determine whether to charge the first energy storage device using the energy in the charge signal. When it is determined at **1622** to charge the first energy storage device, the method **1600** proceeds from **1622** to **1624**. When it is determined at **1622** not to charge the first energy storage device, the method **1600** proceeds from **1622** to **1626**. At **1624**, the battery **100** uses energy from the received charge signal to charge the first energy storage device. The method **1600** proceeds from **1624** to **1604**.

At **1626**, the battery **100** determines whether to charge the second energy storage device. This determination may be made based upon factors such as, for example, the characteristics of the charge signal, the energy stored in the energy storage devices, and charging and discharging cycles of the energy storage devices in the battery **100**. Discrete circuitry and/or look-up tables may be employed to determine whether to charge the second energy storage device using the energy in the charge signal. When it is determined at **1626** to charge the second energy storage device, the method **1600** proceeds from **1626** to **1628**. When it is determined at **1626** not to charge the second energy storage device, the method **1600** proceeds from **1626** to **1630**. At **1628**, the battery **100** uses energy from the received charge signal to charge the second energy storage device. The method **1600** proceeds from **1628** to **1604**.

At **1630** load error processing is performed. For example, the battery **100** may temporarily disable charging of the energy storage devices. The method **1600** proceeds from **1630** to **1604**.

Embodiments of a method of operating a portable energy storage device may perform other acts not shown in FIG. **38**, may not all perform all of the acts shown in FIG. **38**, may combine acts shown in FIG. **38**, or may perform the acts of FIG. **38** in a different order. For example, the embodiment of a method **1600** illustrated in FIG. **38** may be modified to provide energy to a load from more than one energy storage device. In another example, the embodiment of a method **1600** illustrated in FIG. **38** may be modified to simultaneously charge an energy storage device and provide energy to a load.

In another example application, devices employing bi-metal coils, magnetic structures configured to generate compressed magnetic fields, and/or other aspects of the present disclosure may be advantageously employed to convert fluid waves, such as water or sea water waves, to electrical energy. This is a potential environmentally friendly, renewable, source of energy. For example, an apparatus, such as the one disclosed in U.S. Pat. No. 6,864,592 to Kelly, for converting the motion of sea wave energy to electrical energy, including one or more float driven linear generators in which the inertial mass of a float and any linkage to the linear generator is minimized, may be modified in accordance with the present disclosure. The moving part of the generator is sized that its gravitational weight acting upon the float, together with that of the float itself and any intermediate linkage, is substantially equal to half the total buoyancy of the float. In calm conditions, the float would be half in, half out of the water. In the presence of waves during the rise of a wave, an upwards thrust is imparted to the generator equal to substantially half the weight of the water displaced by the float. On the fall of the wave, a downwards thrust due to gravity is imparted to the generator equal to the combined weight of the assembly. Thus, the linear generators experience substantially consistent upwards and downwards thrust during the passing of a wave, and consistent generation of power during both of these phases is achieved. The device of Kelly could be advantageously modified to incorporate bi-metal coils and/or magnetic structures configured to generate compressed magnetic fields to increase its efficiency.

Another patent that discloses converting sea waves to electrical energy is U.S. Pat. No. 6,791,205 to Woodbridge, which is incorporated herein by reference. This patent discloses a reciprocating generator rigidly attached to the underside of an ocean buoy that creates electric power from ocean swells. A generator coil maintains a stable position beneath the ocean surface while the magnetic field housing reciprocates with the vertical motion of the buoy in response to interaction with swell and waves on the surface of the ocean. Damping plates attached to the generator coil inhibit the motion of the generator coil, thus keeping it in a stable position relative to the motion of the magnetic housing. The magnetic housing focuses the magnetic field through the generator coil and the relative motion between the magnetic housing and generator coil creates an electromotive force in the coil. In another example, the apparatus of Woodbridge could be advantageously modified in accordance with the present disclosure to increase its efficiency.

In some embodiments, as shown in FIG. **39**, a generator **200** is employed in a cargo container **350**. Security of cargo containers is a growing concern. If power were available, gamma detection of explosives, infrared detection of human cargo, or other surveillance could be performed. By supporting a generator **200** or **300** in or on a cargo container **350**, energy can be generated by the linear generator from the motion of a ship **360**, due to the action of waves, and this energy can be used to power a variety of surveillance or detection systems. The generator **200** or **300** could be arranged to capture side to side rocking motion, or up and down motion, for example.

In another example, FIG. **40** shows just one example of a water wave to electrical energy power facility **400** comprising a plurality of power generators **402** of the type described above. The facility **400** includes paddles or linkages **412** coupled to the generators **402**. In one embodiment, the generators **402** are like the linear generators described above and the paddles **412** are coupled to the bottom **214** of a generator **200** or the bottom **338** of a generator **300**. The generators **402** are arranged to be moved by water waves and to cause either the coils or magnets of the generators **402** to move in response to the waves. In some embodiments, the generators **402** are supported to float on water, without the use of linkages.

The facility **400** further includes one or more rectification circuits or circuitry **404** coupled to the windings or coils of the power generators **402**. The rectification circuits **404** convert AC currents generated in the windings or coils of the power generators **402** to DC currents. The windings or coils of the power generators may comprise bi-metal coils.

In the embodiment of FIG. **40**, the facility **400** further includes a power storage device **406** coupled to the rectification circuits **404** for accumulating and storing power generated by the windings or coils. The power storage device **406** can be or include one or more batteries, capacitors, combination of batteries and capacitors, or other types of power storage devices. The power storage device includes a charging regulator to provide proper current and voltage to the battery, capacitor, or other energy storage device.

In the embodiment of FIG. **40**, the facility **400** further includes an inverter **408** coupled to the power storage device **406** and configured to supply alternating current to an electricity distribution system or grid. In the illustrated embodiment, the inverter is coupled to the grid via a transformer **410**. Other embodiments, including one or more power generators, for converting water waves to electrical power (either AC or DC) are possible.

Other applications are possible, such as biological motion systems, parasitic power harvesting, self-powered devices such as self-powered security and intelligence gathering devices. For example, in one embodiment, a power generator as described herein is included in a shoe, to generate power from walking. That power can be used to supply any variety of electronic devices.

A shoe-mounted device includes, for example, a power generator as described above, mounted in or on a shoe heel such that each time the heel hits the ground, the impact causes movement of the coil relative to the magnets. The shoe-mounted device also includes a rectification circuit (e.g., a full wave rectifier) coupled to the coil of the power generator, and a power storage device such as a capacitor or battery coupled to the output of the rectification circuit. A voltage regulator can be included to provide proper current and voltage to the storage device.

For example, FIG. **41** shows just one example of a biological motion device **500** comprising a power generator **502** of the type described above in connection with FIGS. **22**-**32**. The power generator **502** is mounted, in some embodiments, in the heel of a shoe or elsewhere on a person such that the action of walking moves the magnets relative to the coil of a power generator **502**.

The device **500** further includes a rectification circuit or rectification circuitry **504** coupled to the windings or coils of the power generator **502**. The rectification circuitry **504** converts AC currents generated in the windings or coils of the power generators **502** to DC currents.

In the embodiment of FIG. **41**, the device **500** further includes a power storage device **506** coupled to the rectification circuitry **504** for accumulating and storing power generated by the windings or coils. The power storage device **506** can be or include one or more batteries, capacitors, combination of batteries and capacitors, or other types of power storage devices.

In the embodiment of FIG. **41**, the device **500** further includes a voltage regulator **508** coupled to the power storage device **506** and configured to provide a stable output voltage to an electronic device borne by a person. Other embodiments are possible.

FIG. **42** illustrates an embodiment of a system **100** that is gimbaled to facilitate the conversion of available energy into electrical energy. The system **100** comprises a generator **102**, such as, for example, one or more of the generators illustrated in FIGS. **22** through **32**, supported by a support structure **104** that facilitates positioning of the generator **102** in a desired position. In some embodiments, the support structure **104** may employ gyroscopic techniques.

FIG. **43** illustrates a system **100** comprising a plurality of generators **102**, **104**, **106** coupled to a support structure **108**. A first generator **102** is coupled to the support structure **108** so as to be oriented along an X-axis **110**. A second generator **104** is coupled to the support structure **108** so as to be oriented along an Y-axis **112**. A third generator **106** is coupled to the support structure **108** so as to be oriented along a Z-axis **114**.

FIG. **44** illustrates an article of clothing **100** comprising an embodiment of a battery **102**, such as one of the battery embodiments illustrated in FIGS. **33** through **35**, an embodiment of a generator **104**, such as one of the generators illustrated in FIGS. **22** through **32**, solar collectors **106**, and a radio-frequency energy collector **108** comprising an antenna system **110** and a rectifier **112**. The article of clothing **100** also comprises a bus system **114** to couple the various components together and a coupling **116** to couple the bus system **100** to the battery **102**. The coupling **116** may be configured to condition or add together electrical energy received from the generator **104**, the solar collectors **106** and/or the radio frequency energy collector **108**, or to switch a connection **118** to the battery to connect to one or more of the other components of the article of clothing, such as the generator **104**. The coupling **116** may also be configured to allow connections to external loads and/or sources of energy. As illustrated, the article of clothing **100** is a shirt, but other embodiments may comprise other articles of clothing. The battery **102**, generator **104**, solar collectors **106**, the radio-frequency energy collector **108**, bus system **114**, the coupling **116** and the connection **118** may be integrated into the article of clothing, removeably coupled to the article of clothing, or combinations of the above. For example, the antenna system **110** and the bus system **114** may be integrated into the fabric of the shirt while the battery **102** may be coupleable to the shirt. In another example, a button **122** may also comprise a solar collector **106**. Some embodiments may not comprise all of the components illustrated. For example, an embodiment may comprise the battery **102** and the antenna system **110**. In some embodiments, a control module in the battery (see control module **208** in FIG. **34**), may control storage in the battery **102** of energy received by the battery **102** via the bus system **114**.

FIG. **45** is a side sectional view of an embodiment of a system **100** in accordance with an embodiment. The system **100** comprises a rotor **102** comprising a magnetic structure **104** configured to generate a compressed magnetic field and a stator **106** comprising one or more bi-metal coils **108** comprising an electrical conductive element **110** and a magnetic conductive element **112**. FIG. **46** is a top cross-sectional view of the rotor **102** of FIG. **45** taken along lines **46**-**46**. The magnetic structure **104** comprises a plurality of magnets **114** held spaced apart with like poles facing each other so as to generate a compressed magnetic field.

Although specific embodiments of and examples for the coil, magnetic structure, device, generator/motor, battery, control module, energy storage devices and methods of generating and storing energy are described herein for illustrative purposes, various equivalent modifications can be made without departing from the spirit and scope of this disclosure, as will be recognized by those skilled in the relevant art. The various embodiments described above can be combined to provide further embodiments.

These and other changes can be made to the invention in light of the above-detailed description. In general, in the following claims, the terms used should not be construed to limit the invention to the specific embodiments disclosed in the specification and the claims. Accordingly, the invention is not limited by the disclosure, but instead its scope is to be determined entirely by the following claims.

What is claimed is:

1. A battery, comprising:

a case;

a coil contained within the case;

a magnetic structure contained within the case and configured to generate a compressed magnetic field;

a first energy storage device contained within the case;

a plurality of contact terminals coupled to the case; and

a control module contained within the case and coupled to the coil and the first energy storage device.

2. The battery of claim **1** wherein the magnetic structure comprises a plurality of spaced-apart rare earth magnets configured so that like-polarity poles face each other in neighboring magnets in the plurality of rare earth magnets.

3. The battery of claim **2** wherein the magnets in the plurality of magnets are held in position with respect to one another.

4. The battery of claim **2** wherein a space between two magnets in the plurality of magnets is substantially filled with a non-magnet substance.

5. The battery of claim **4** wherein the non-magnetic substance comprises air.

6. The battery of claim **4** wherein the non-magnetic substance comprises a fluoropolymer resin.

7. The battery of claim **1** wherein the case is evacuated and hermetically sealed.

8. The battery of claim **1** further comprising:

a suspension system coupled to the magnetic structure.

9. The battery of claim **8** wherein the suspension system is tuned to optimize conversion of expected patterns of movement into electrical energy.

10. The battery of claim **8** wherein the coil comprises:

an electrical conductive element; and

a magnetic conductive element.

11. The battery of claim **10** wherein the magnetic conductive element is configured to focus magnetic flux in the electrical conductive element.

12. The battery of claim **11** wherein the magnetic conductive element comprises a silver/nickel alloy.

* * * * *

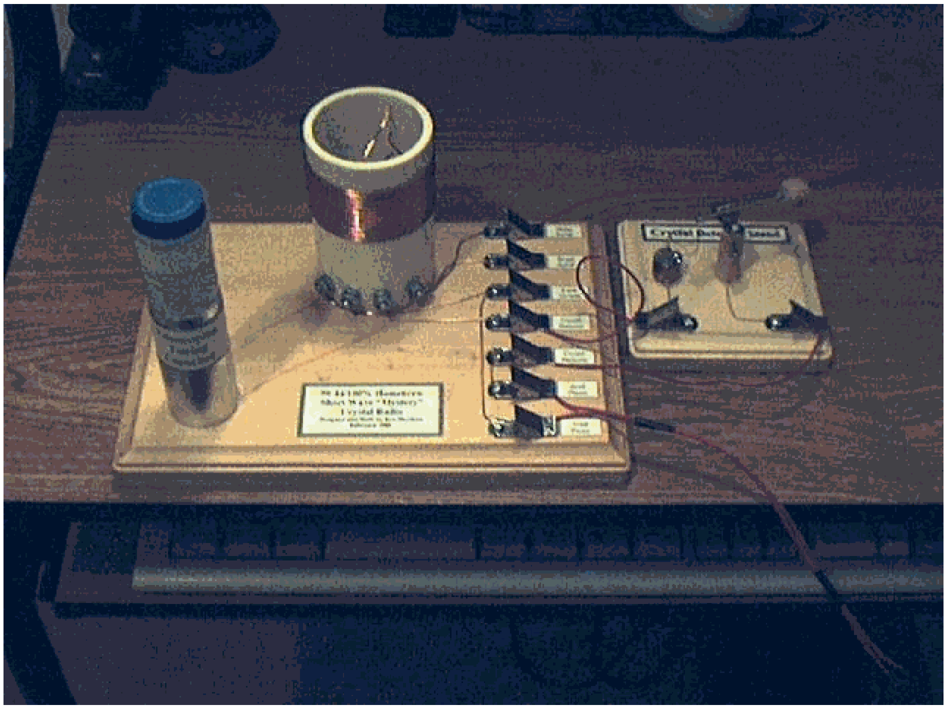

Experimental Short-wave "Mystery"

Crystal Radio

by Ken Harthun

Copyright January 2001

This is not the typical crystal set construction article to which you may be accustomed; rather, it is a description of a few *workable* circuits that leaves plenty of room for experimentation if you are so inclined. You can build a fine working short-wave crystal set based on what I give you here. These circuits perform very well. However, I’m still experimenting with them and I challenge you to do the same, to discover what works best for you. Tinkering with a circuit, trying different things to see what works – and what doesn’t – is the essence of our hobby and is what keeps this particular rock head interested!

Because this set is experimental, I make no claims that this design – or any particular hookup that results from this article - is the *ne plus ultra* in short-wave crystal radios. We all know that the perfect crystal set does not exist. The "perfect" crystal set for my location is something that may or may not work for you (I have to deal with a 50Kw blowtorch that is located virtually in my backyard – you may be more fortunate).

My antenna is about 150 feet of wire in an inverted V with the apex at about 35 feet. Some of it runs every-which-way across my roof and into my radio room. Though at BCB frequencies my antenna is much too short, it is actually much too long an antenna at short-wave frequencies. You can probably get by with a lot less, depending on your location.

With these things in mind, let’s begin and take a look at a project that I believe will give you plenty of enjoyment – and will probably raise many questions. In fact, I hope that this article will generate enough interest in short-wave crystal set reception that we see some real innovation in future designs.

The radio is based on the Mystery Crystal Set designed by a fellow who called himself "Proton" and who first published details of the set in the Sunday Mail newspaper in Brisbane, Australia in 1932. I won’t rehash any of the details because the set is well covered in the articles posted at http://www.clarion.org.au/crystalset/mystery.html by Ray Creighton. I have simply adapted the circuit and modified it a bit for my set. Fig. 1 is the basic schematic.

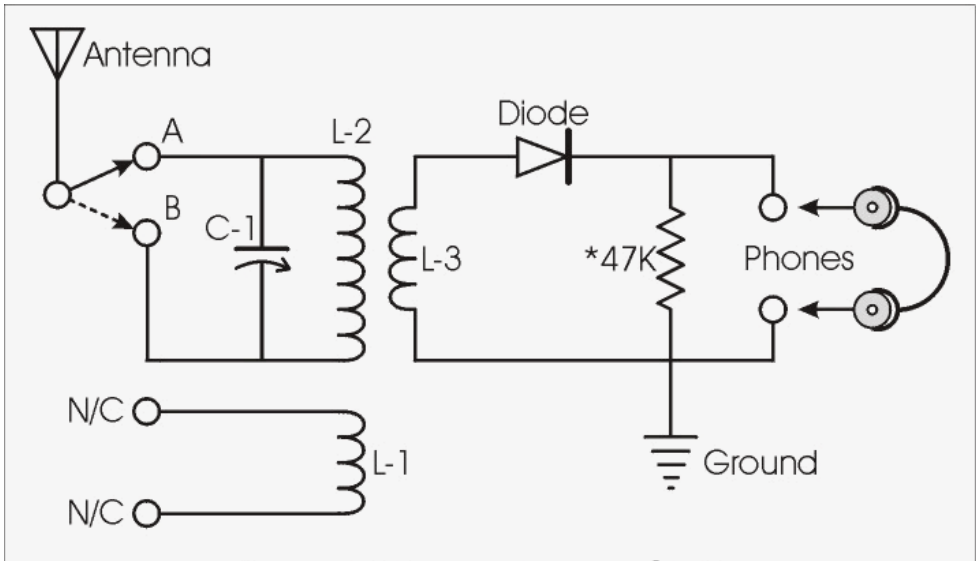

Figure #1, Basic "Mystery" Circuit.

Construction

The coil form is a 2.5 inch long piece of 2" white PVC plumbing pipe with an outside diameter of 2 3/8 inches. This is the kind that has a thickness of about 3/16 inches and is known as "Schedule 40". Construction details are shown in Figure 2.

L1 – 20 turns #22 close wound, approx. 35 uH

L2 – 13 turns #18 or #20 wound at 20 tpi, approx. 12 uH

L3 – 6 turns #22 bifilar wound in the center of L2

C1 – I’ve been using a 20 – 56 pF air variable for coverage of approx. 6 – 10 MHz and a 13 – 250 pF air variable for approx. 2.9 – 12.7 MHz coverage with the coils shown.

Diode – 1N34A or crystal detector.

(You may find as I did that a piece of pyrite and a cats whisker work much better than a diode. For some reason, I find that the pyrite detector is more sensitive and gives me much better volume than the diode. Also, with pyrite, you can eliminate the resistor across the crystal earplug.)

Resistor – 47K (*Note: you can eliminate this resistor if you are using high impedance magnetic headphones)

Phones – crystal earplug or high impedance magnetic phones

I have been using alligator leads to connect it up so I can rapidly change configurations. The photo shown at the beginning of this article is a completed homebrew version of the set that uses mostly non-electronic parts (except for one Fahnestock clip hat was used to hold the detector rod). The capacitor is made from a plastic tube and some aluminum tape that you can find at auto parts stores.

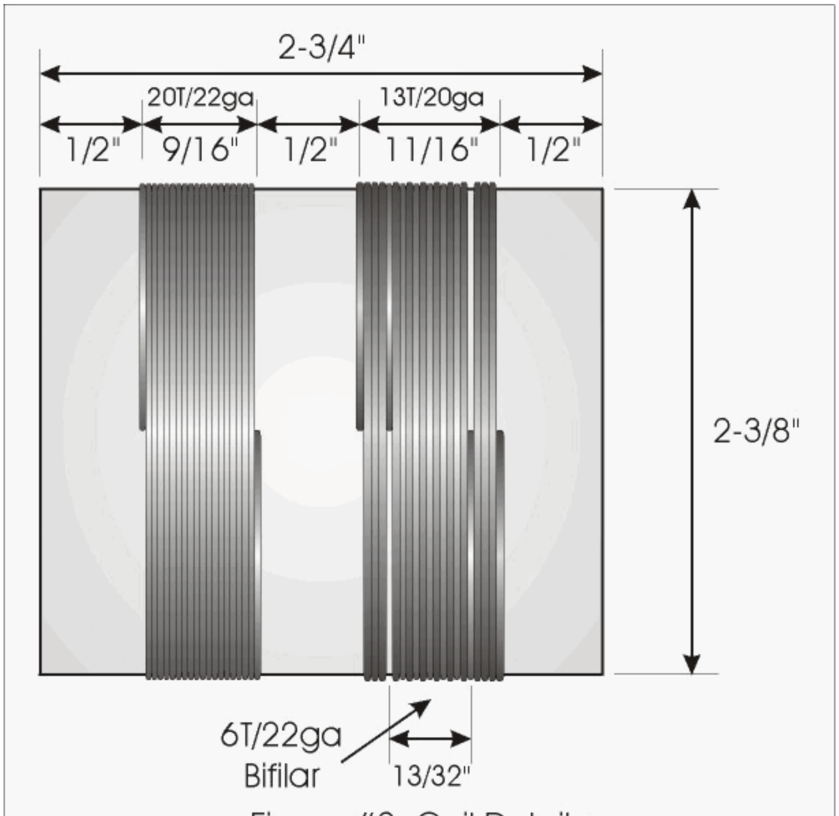

Figure #2, Coil Details.

In the original coil I wound, I used hot-melt glue to hold the windings in place. About six beads or "ribs" across each of the windings works very well. Also looks pretty good – better than tape!

In the basic "mystery" circuit shown in Fig. 1, L1 is not used. The tuned primary (L2) has two antenna connection points, A and B. A is the least selective, but gives a louder signal. B is more selective, but the sensitivity drops notably. While I show a ground connection, I noticed that results vary. In most cases, I find that the set works better in this configuration without a ground and is very sensitive. Selectivity is enhanced with the ground connected, but some signals drop out while others get louder. I’m not sure why this is and I’m still experimenting. I have tried putting a small trimmer capacitor in the ground connection to vary the coupling to ground, but my results are inconclusive at this point.

Overall, the basic configuration works very well and is a fine set in its own right. I easily pick up WWCR at 5070 kHz, WEWM on 9975 kHz, have heard Radio Havana (Cuba), the BBC and many other German, Chinese, French and Spanish language broadcasts. My first night, I picked up a station that was running a program called "Let’s Learn Chinese". Bear in mind that propagation conditions vary considerably. You may or may not hear this much at first. Then again, you may hear even more. Some evenings I hear very little but I’m listening as I work on this article in late February and there seems to be stations everywhere as I tune. The audible stations fade in and out and back again, so I get a good variety without having to touch the tuning dial!

As I mentioned earlier, in my location near Cincinnati, Ohio, I have a 50 kW blowtorch, WSAI on 1530 kHz, virtually in my backyard. I have to use a wave trap to tune this station out or it’s all I hear. As a matter of fact, I can hook a diode up to a crystal earplug, hold the diode in my hand and still hear this station loudly! My point is that your conditions will certainly be different than mine, so there is no way I can tell you what to expect to hear. But I can almost guarantee that you will hear something.

Figure 3 shows another configuration I have tried.

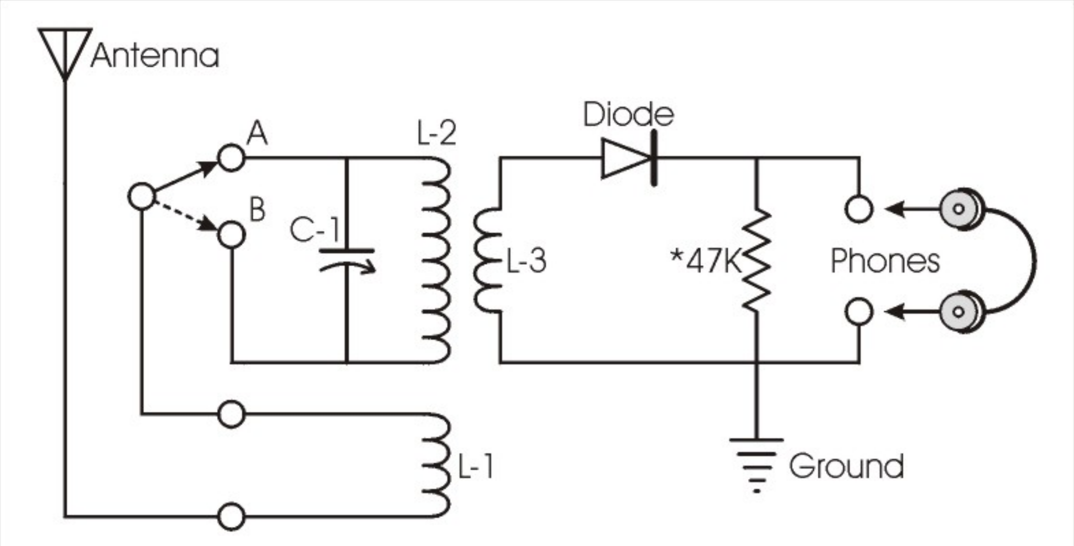

Figure #3, #1 Variation..

In this configuration, L1 is used as a coupling coil for the antenna and one end is attached to either A or B on the primary. This is a very interesting hookup, more selective than the basic circuit in Fig. 1, but not quite as sensitive. With L1 hooked up to point A, the selectivity is quite sharp. When hooked to point B, the set is so selective that tuning can be tedious. Again, the ground is optional, but I find a little better sensitivity in this hookup with the ground connected and the selectivity hardly suffers at all. One thing to try would be to decrease the spacing between L1 and L2/L3. I think that the current spacing may result in coupling that is a bit too loose for this set.

The configuration detailed in Figure 4 shows some promise for daytime reception in the upper short-wave bands and for further development as a set with variable coupling. With the existing design, the coupling between L1 and L2/L3 is somewhat loose at the ½ inch spacing. Making L1 moveable and possibly large enough to fit over L2/L3 are some things I am considering.

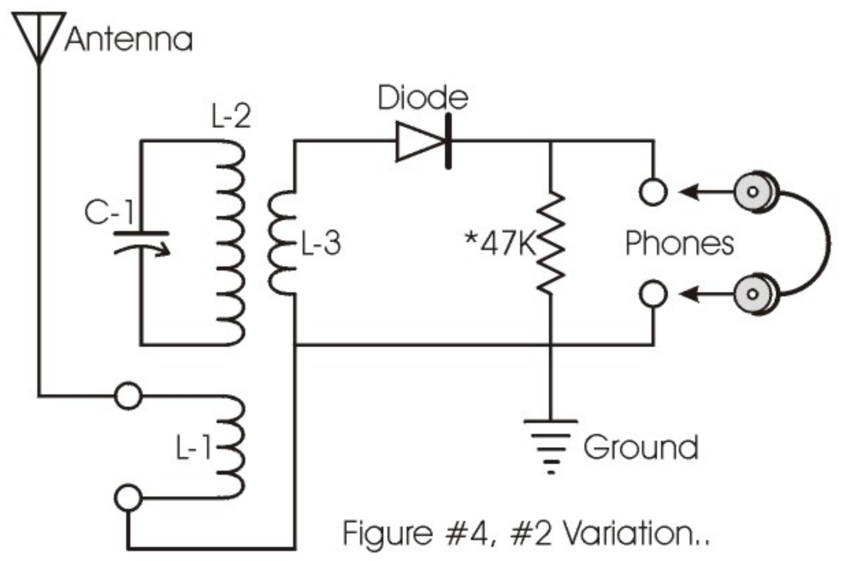

Figure #4, #2 Variation..

Here are some ideas for further experimentation.

1. Grounding connections A and B to see how this affects reception. (This would actually be defeating the mystery circuit and using L3 as just a detector-coupling coil)
2. Hook up a variable capacitor in series with the antenna.
3. Hook up a variable capacitor in parallel across L1 to tune the antenna.
4. Try the "Tuggle" circuit across L2. This would involve using a 2-gang variable capacitor and hooking one of the sections in parallel with L2 and the other section in series with the ground.
5. Try the Tuggle circuit across L1.
6. Use Tuggle tuning with main variable section across L1 and the second capacitor section in series with A/B connection.
7. Make L1 moveable to vary coupling.
8. Band switching and band spreading using various fixed and variable caps.
9. Use a smaller coil form to approximate a 1:1 length-diameter ratio.
10. Use a smaller inductance value (same as L2?) for L1.

This is by no means a complete list of the things that can be tried with this circuit. In fact, I even redesigned the coil to the following dimensions:

L2 – 18 turns #20 space wound to 1.3" long, 2.375" diameter for approx. 20 uH.

L3 – 7 turns 7/30* bunched cable bifilar wound in center of L2

L1 is not present in this configuration.

*7/30 cable is 7 strands #30 enameled wire twisted into a cable of approx. size of #20. Individual wires are stripped at the ends then twisted together and tinned.

I hope I have given you some food for thought and inspired you to try your hand at making this interesting circuit work. I intend to continue my own experiments and perhaps write another article at some future time. If you want to contact me, you can post a message on Rap-n-Tap, the Yahoo Crystal Set Radio Club, or email me at kc4iwt@yahoo.com. I welcome all comments and suggestions.

Good luck and Happy Experimenting!

3/18/2001 Notes on further modifications

L1 is now present again on the redesigned coil. It comprises 8 turns of # 22 AWG close wound, spaced ¼" from the bottom end of L2. The antenna and ground connections are made to this coil.

L2 is now 14 turns instead of 18 turns to bring the coil more in range of the most active SW bands.

L3 is unchanged.

Experiments with a Perikon detector are very interesting. My Perikon detector consists of a piece of chalcopyrite in contact with a small cone-shaped piece of zincite. It is equally as sensitive as my pyrite detector – maybe more so under some conditions – and has an interesting property in that it appears to be somewhat "tunable". I’m not sure if this is due to signal strength or has more to do with the frequency to which the set is tuned, but I observe a distinct ability to peak signals differently as I tune across the bands. I seem to recall some mention of this property in early radio literature.

One big advantage of the Perikon is that it is not as touchy as the typical catswhisker-and-mineral detector. It is easy to begin detecting a signal and once detection begins, it is a simple matter to peak the signal. I also observe that the Perikon requires a bit of pressure to work properly. It doesn’t require the light touch like other point contact detectors and adjustment is non-critical. Unlike the catswhisker detector, the Perikon doesn’t mind being bumped and tends to keep its setting. It is very stable. I like this detector a lot.

I tried a zincite/pyrite combination, but you can forget that one. Just doesn’t work. Likewise, chalcopyrite with a catswhisker just isn’t very sensitive. It works, but it seems to been just a couple of steps up from a rusty razor blade. Yes, the razor blade detector actually works on this set on some of the stronger signals. And I have to say again that a 1N34 germanium diode just lacks sensitivity in this set. It doesn’t work as well as either the Perikon or the pyrite detectors.

Crystal Radios for the Shortwave Bands

(21 Jan 02)

The first set tunes from about 5.8 to 14 MHz, which includes the 49, 41, 31, 25 and 22 meter shortwave broadcast bands. The capacitor is the oscillator section of a poly film variable capacitor, and the coil is wound on a toilet paper core. The main coil, L1, was wound on a length of 1/2 inch of the core, and separated from the antenna coupling coil, L2, by about 1/8 to 1/4 inch on the same core. Calculated inductance of L1 was 10.6 uH. Measured capacitance of the variable was 10 - 68 pF. Using a 50 foot end fed antenna, stations could be heard 24 hours a day, with the higher bands being heard during daylight, and the two lower bands coming in at night as well.

There are 4 primary ways to attach antenna and ground, and all work to one degree or another. First and second, the antenna alone may be connected to either the antenna tap or to the ground tap. Third and fourth, the antenna and ground are connected to the normal taps as indicated, or they may be reversed. A small frequency shift was noted when going from option 3 to 4. Use of the ground sometimes made the signals stronger. Quantitative assessments using on-air stations was not attempted as the signal strength varied constantly, with shifts of up to 10 dB being noted within the space of 30 seconds. As a side note, just connecting the antenna to the top of the tank coil was not as productive.

Selectivity, while not great, was sufficient to separate the short-wave BC bands (the 25 and 22 meter bands tended to run together, however), and often to separate two stations within a band. Slow tuning, even with this selectivity, is needed due to changing signal strengths and because of the fast tuning rate.

Using a longer antenna gave higher signal strengths, as expected, and little or no shift in frequency was noted when going from one antenna to the other. No MW broadcast band interference was observed.

This is a schematic of the set:

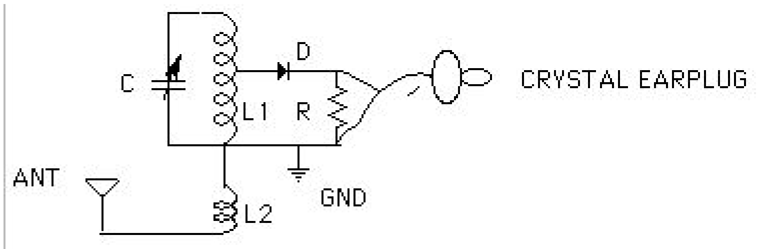

C - 10-60 pF
L1 - 14 turns #24 plastic coated wire on 1 5/8 inch core tapped at 8th turn from top for detector
L2 - 5 turns wound 1/8 inch below L1 on same core
D - 1N34A diode
R - 47 kohm

Spacing between L1 and L2 was closer to 1/4 inch, as it turned out when my students built this, and they seemed okay. Since the shoddy antenna I have in my classroom is well shielded by the metal roof of the building, I used a grid dip oscillator to check the sets for response before letting them out of the room. With the unmodulated signal from the GDO, I could hear a few faint heterodyne squeals from stations that were there but too weak to hear.

I made another coil on the same diameter form using #20 solid plastic coated hookup wire, winding 20 turns on 2 inches for L1 and 6 turns for L2. This gave me slightly stronger signals, as expected. I made a breadboard attempt to use this with a Tuggle antenna tuner, but gave up early on, primarily since this was designed to be a student radio, and for other more technical reasons which I'll explain below.

This is really a "first iteration" radio, designed just for shortwave use, and, while it works pretty well, clearly has some way to go. The tuning rate, some 10 MHz in a 180 degree rotation of the tuning knob, is way too fast, but is probably okay for this first attempt. It was clear that the selectivity of the set leaves something to be desired, so more work on the coils is in order. I didn't spend much time optimizing the coil tap for the detector, just picking one that is historically useful. Tapping the detector to the top of the tank circuit was not as productive as the tap I used. I also didn't spend any time picking over detectors, and just grabbed the first germanium diode out of the box. If I can get the selectivity up a bit, I will have to get serious about band spreading the tuning. Obviously, I need to go back and look at an antenna tuner again. One thing I realized, which may or may not make a difference, is that at these frequencies the antenna is a quarter wavelength long or longer, so you can't make it resonant with just a few turns of a coil in the antenna line. However, the set is a snap to build (especially the coil), and picks up a bunch of stations, and is fun to use if you haven't tried the higher frequencies.

Okay, time to do some more playing around. Variation #2 was built to try to get some more selectivity for a kid radio, and essentially the same as the first but with the primary separated from the tank coil. This seemed to work a bit better, and the bands separated better as well. This is the schematic with some notes:

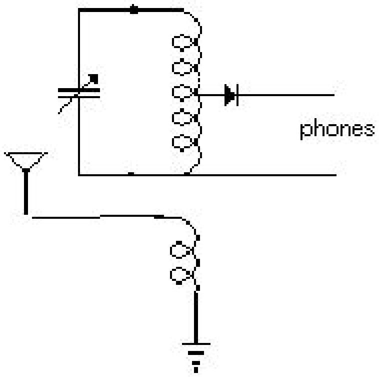

For this one, I wound the coils on a 1 5/8 inch toilet paper core, and used 15 close wound turns for the tank coil, tapped at turn 6 from the ground for the detector. The primary coil was 6 turns separated by 1/4 inch from the secondary. If you want to dispense with phones, one of the little Radio Shack audio amps works fine. With the amp, I noticed that with my 150 foot long wire, the top of the BC band came in the background a little; not so with the 50 foot "student" antenna. The diode was a germanium out of the junk box, and the variable capacitor was the 10 - 70 pF oscillator section of a poly film rf tuner capacitor. The wire was #24 plastic coated telephone wire, which gives about 30 turns per inch. It tunes from the 49 meter band to just into the 19 meter band. It all fit nicely on a 4 inch length of 1 inch x 4 inch pine.

Since my 8 year old seemed interested in doing something with a crystal set for a science project, we made up this next one, really just a slight modification of the one above, adding an audion type headphone amp to drive a low impedance set of phones (he ended up taking my Stromberg Carlson sound powered set, declaring them the most comfortable). This is the schematic of the finished product:

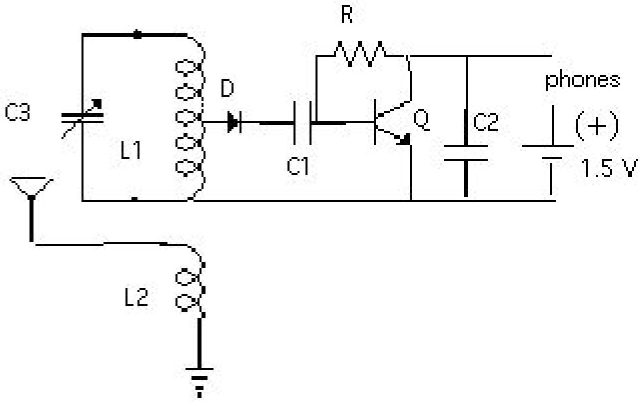

C3, L1, L2, and D are as described above.
C1, C2 are 0.1 uF disc ceramics
R = 100 kohms
Q is a 2N2222, but just about any similar transistor will do. Reverse the battery if using a PNP transistor.

I was pretty lavish with this one, and used a 7 inch section of 1 x 4 inch pine for the base, and even added a panel to support the variable and a simple paper dial that I "calibrated" for the shortwave bands using a dip meter. With the sound powered phones, I probably should have gone on and made the coupling between L1 and L2 variable to (1) control the volume on the stronger stations and (2) to try and improve the selectivity, which isn't bad, but isn't great either.

And now (Drum roll), a shortwave convertible. Actually, this should look very familiar, since I use this basic one transistor amp a lot.
This amp was added to use higher impedance phones, such as the standard 2k, or the crystal earplug with the parallel resistor.

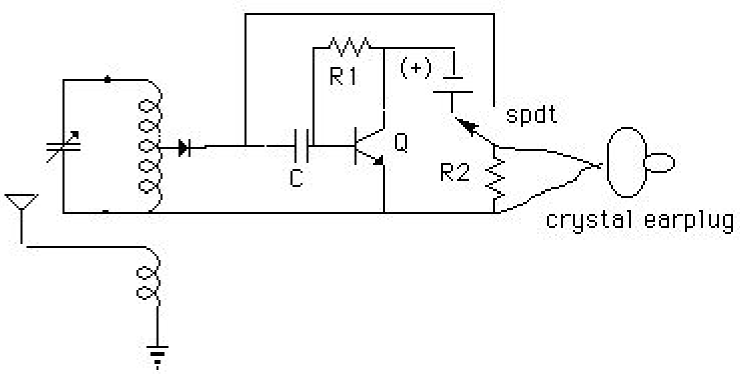

Q - just about any general purpose transistor will do, such as a 2N2222 or whatever you have on hand. I used a PNP out of a Radio Shack pack 15 on one set(and reversed the battery); worked fine.
R1 - 470 kohms
R2 - 47 kohms
C - 0.1 uF - you can probably use just about anything here
The battery I used was a 1.5V AA, but a 9 V worked fine too with a little more volume.
The crystal radio portion of the set is as described above. I used brass coated thumbtacks stuck into a piece of pine for soldering points. You can put in a couple of fahnestock clips where R2 and the crystal earplug go and use any phones you have on hand. I designed this to be an easy-to-assemble kid radio, but you can get as fancy as you like. I just used hot glue to hold the battery holder and switch down on the base. You might notice that, unlike the original crystal convertible, I don't disconnect the diode from the input capacitor C when switching from straight crystal to amplified crystal. Found that it didn't seem to matter, and made wiring easier. Also note that the battery connects after the phones, with the phones going to chassis ground, unlike the set I made with the audion type amp. You have some flexibility here, so do what suits you, but this arrangement gives you the easy convertible switch.

So how does it do? As built, the amplifier seems to give about a 10 to 20 dB boost to the signal. Makes for more comfortable listening, and stations that you can tell are there become clearly audible. Going from a 50 foot antenna out the window to a 6 foot whip indoors, I was still able to hear a few stations, even in the daytime. I tried adding a feedback coil for a little regeneration, just for grins. Alas, the used 1.5 V AA wasn't having any of this, so I'll try it another day.

For more information on the general nature of shortwave listening and the frequencies for each band, try this web site:
http://www.danatkinson.co.uk/html/sw/faq.htm

[HOME]()

The SkyWaves Short-wave Crystal-set Prototype

Here's some info on my recent short-wave crystal-set experiments. The basic conclusion is that SW crystal sets are quite worthwhile, and need not be unduly elaborate to give gratifying results. This is especially true on the US east coast where the "prime time" broadcasts are really strong.

A word of warning: You're likely to hear strong, clear English broadcasts from the likes of The BBC, The Voice of Viet Nam, and Radio Taipei. Don't believe it! These are rebroadcast from nearby sites in US and Canada. To sort this out, there is a good frequency list available at: http://www.primetimeshortwave.com/

I the short time I've been at it, I have heard Radio Exterior de Espana , Swiss Radio International, and The Voice of Russia from my West-central NJ location. So intercontinental DX is entirely possible. I even heard a "numbers station" on about 13.5MHz.

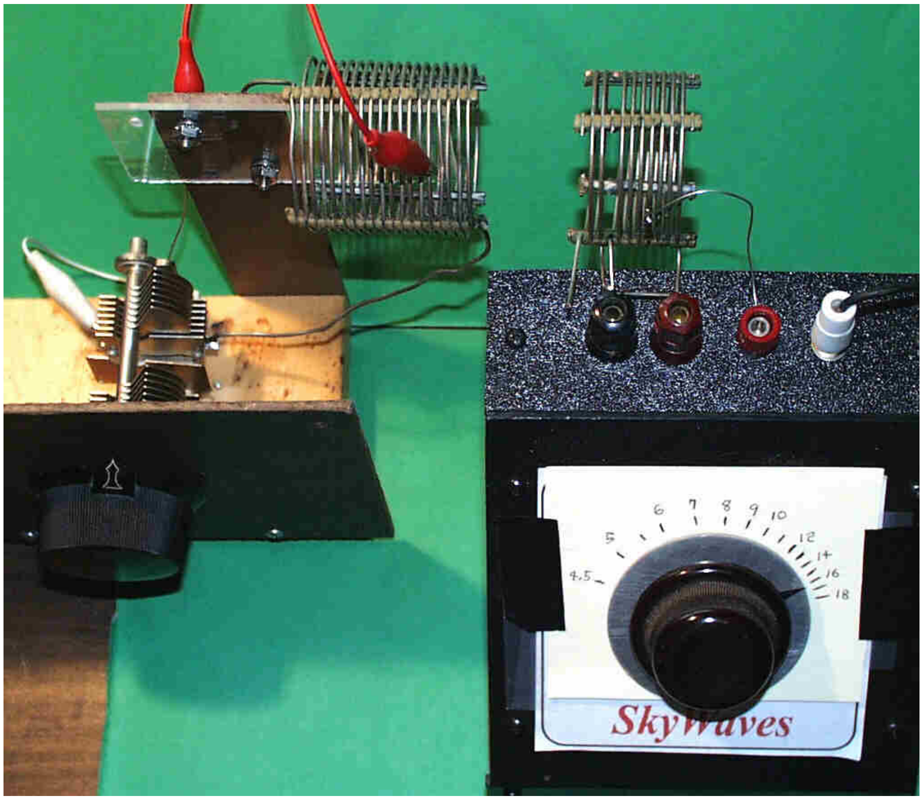

This is a simple double-tuned radio built around Air Dux coil stock. I used this because it was handy, but solid bare wire of about 14 gauge on 2" PVC should work just fine. (Strip a length of #14 Romex.) The primary tuned circuit is Tuggle style and offers a reasonable compromise between effective, flexible antenna matching and ease of use. I bent down every other turn on L1 to make it easy to tap with an alligator clip. I equipped the secondary coil with a separate "primary" winding so the set can be used in the single-tuned mode to do a quick band scan. Coupling is controlled by moving the two sections around on the table. The "ground" end of the coils face each other.

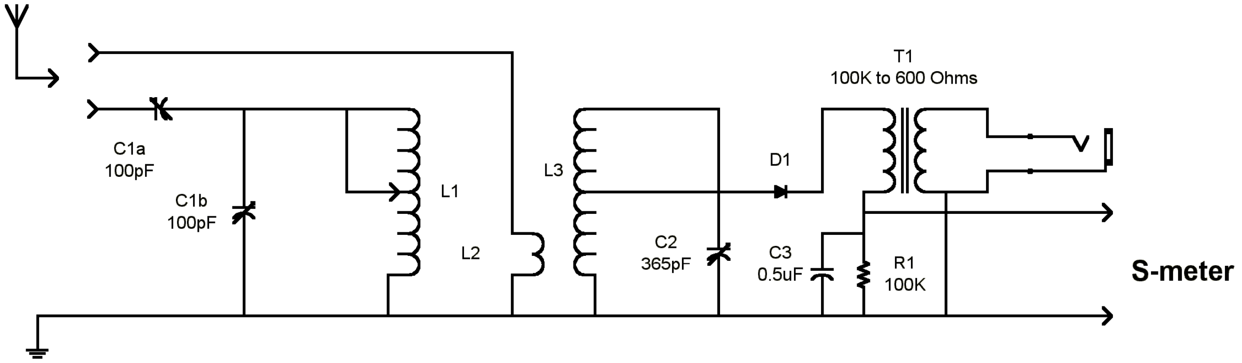

All coils 2" dia. 10 TPI "Air Dux"
L1 - 18 turns, tapped every other turn
L2 - 2 Turns, spaced 2 turns from L3
L3 - 8 Turns (approx. 4.3uH)

SkyWaves Short-wave Prototype

Drawn 11 April 01 - Al Klase

Some additional observations: A secondary coil of about 4 – 4.5 uH will tune 5-18 MHz with a 365pF cap. That's where 95% of the action is, so band switching can be avoided. As with most double tuned sets, a calibrated dial on the lightly-coupled secondary really helps sort things out on the air. I'm using sound-powered phones, but signals are strong, so a modest headset will be ok for starters. Just elininate T1, C3, and R1, and connect the headset between the diode and ground.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C07G 13/00, H03K 3/30

A1

(11) International Publication Number: WO 92/07861

(43) International Publication Date: 14 May 1992 (14.05.92)

(21) International Application Number: PCT/US90/06407

(22) International Filing Date: 2 November 1990 (02.11.90)

(71)(72) Applicant and Inventor: MEYER, Stanley, A. [US/US]; 3792 Broadway, Grove City, OH 43123 (US).

(74) Agent: BARANOWSKI, Edwin, M.; Porter, Wright, Morris & Arthur, 41 South High Street, Columbus, OH 43215 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES, ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, KR, LU (European patent), NL (European patent), SE (European patent), US.

Published
With international search report.

(54) Title: A CONTROL AND DRIVER CIRCUITS FOR A HYDROGEN GAS FUEL PRODUCING CELL

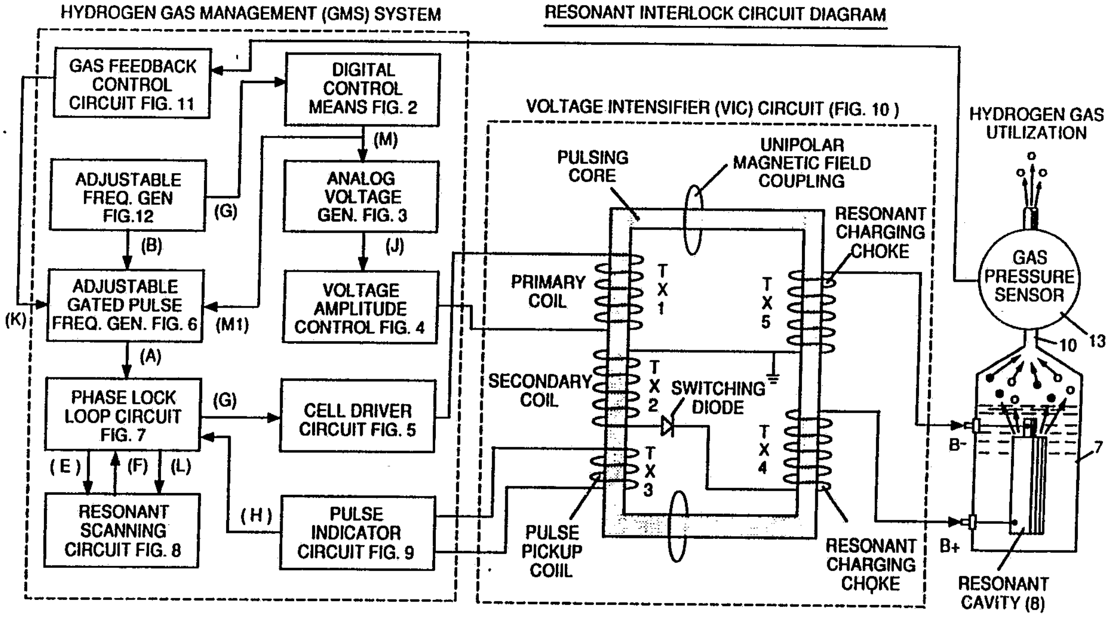
HYDROGEN GAS MANAGEMENT (GMS) SYSTEM
RESONANT INTERLOCK CIRCUIT DIAGRAM
GAS FEEDBACK CONTROL CIRCUIT FIG. 11
DIGITAL CONTROL MEANS FIG. 2
(M)
ADJUSTABLE FREQ. GEN FIG.12
(G)
ANALOG VOLTAGE GEN. FIG. 3
(B)
(J)
ADJUSTABLE GATED PULSE FREQ. GEN. FIG. 6
(K)
(M1)
VOLTAGE AMPLITUDE CONTROL FIG. 4
(A)
PHASE LOCK LOOP CIRCUIT FIG. 7
(G)
CELL DRIVER CIRCUIT FIG. 5
(E)
(F)
(L)
RESONANT SCANNING CIRCUIT FIG. 8
(H)
PULSE INDICATOR CIRCUIT FIG. 9
VOLTAGE INTENSIFIER (VIC) CIRCUIT (FIG. 10)
PULSING CORE
UNIPOLAR MAGNETIC FIELD COUPLING
RESONANT CHARGING CHOKE
PRIMARY COIL
TX1
TX5
SECONDARY COIL
TX2
SWITCHING DIODE
TX3
TX4
PULSE PICKUP COIL
RESONANT CHARGING CHOKE
HYDROGEN GAS UTILIZATION
GAS PRESSURE SENSOR
10
13
B-
B+
7
RESONANT CAVITY (8)

(57) Abstract

A control circuit for a capacitive resonant cavity water capacitor cell (7) for the production of a hydrogen containing fuel gas has a resonant scanning circuit cooperating with a resonance detector and PLL circuit to produce pulses. The pulses are fed into the primary (TX1) transformer. The secondary (TX2) transformer is connected to the resonant cavity water capacitor cell (7) via a diode and resonant charging chokes (TX4, TX5).

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CS	Czechoslovakia
DE*	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GN	Guinea
GR	Greece
HU	Hungary
IT	Italy
JP	Japan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
MC	Monaco
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
NL	Netherlands
NO	Norway
PL	Poland
RO	Romania
SD	Sudan
SE	Sweden
SN	Senegal
SU+	Soviet Union
TD	Chad
TG	Togo
US	United States of America

+ Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

CONTROL AND DRIVER CIRCUITS FOR A HYDROGEN GAS FUEL PRODUCING CELL

This invention relates to electrical circuit systems useful in the operation of a water fuel cell including a water capacitor/resonant cavity for the production of a hydrogen containing fuel gas, such as that described in my United States Letter Patent No. 4,936,961, "Method for the Production of a Fuel Gas", issued on June 26, 1990.

In my aforesaid Letters Patent for a method for the production of a fuel gas, voltage pulses applied to plates of a water capacitor tune into the dielectric properties of the water and attenuate the electrical forces between the hydrogen and oxygen atoms of the molecule. The attenuation of the electrical forces results in a change in the molecular electrical field and the covalent atomic bonding forces of the hydrogen and oxygen atoms. When resonance is achieved, the atomic bond of the molecule is broken, and the atoms of the molecule disassociate. At resonance, the current (amp) draw from a power source to the water capacitor is minimized and voltage across the water capacitor increases. Electron flow is not permitted (except at the minimum, corresponding to leakage resulting from the residual conductive properties of water). For the process to continue, however, a resonant condition must be maintained.

Because of the electrical polarity of the water molecule, the fields produced in the water capacitor respectively attract and repel the opposite and like charges in the molecule, and the forces eventually achieved at resonance are such that the strength of the covalent bonding force in the water molecule is exceeded, and the atoms of the water molecule (which are normally in an electron sharing mode) disassociate. Upon disassociation, the formerly shared bonding electrons migrate to the hydrogen nuclei, and both the hydrogen and oxygen revert to net zero electrical charge. The atoms are released from the water as a gas mixture.

In the invention herein, a control circuit for a resonant cavity water capacitor cell utilized for the production of a hydrogen containing fuel gas is provided.

The circuit includes an isolation means such as a transformer having a ferromagnetic, ceramic or other electromagnetic material core and having one side of a secondary coil connected in series with a high speed switching diode to one plate of the water capacitor of the resonant cavity and the other side of the secondary coil connected to the other plate of the water capacitor to form a closed loop electronic circuit utilizing the dielectric properties of water as part of the electronic resonant circuit. The primary coil of the isolation transformer is connected to a pulse generation means. The secondary coil of the transformer may include segments that form resonant charging choke circuits in series with the water capacitor plates.

In the pulse generation means, an adjustable first, resonant frequency generator and a second gated pulse frequency generator are provided. A gate pulse controls the number of the pulses produced by the resonant frequency generator sent to the primary coil during a period determined by the gate frequency of the second pulse generator.

The invention also includes a means for sensing the occurrence of a resonant condition in the water capacitor/resonant cavity, which when a ferromagnetic or electromagnetic core is used, may be a pickup coil on the transformer core. The sensing means is interconnected to a scanning circuit and a phase lock loop circuit, whereby the pulsing frequency to the primary coil of the transformer is maintained at a sensed frequency corresponding to a resonant condition in the water capacitor.

Control means are provided in the circuit for adjusting the amplitude of a pulsing cycle sent to the primary coil and for maintaining the frequency of the pulsing cycle at a constant frequency regardless of pulse amplitude. In addition, the gated pulse frequency generator may be operatively interconnected with a sensor that monitors the rate of gas production from the cell and controls the number of pulses from the resonant frequency generator sent to the cell in a gated frequency in a correspondence with the rate of gas production. The sensor may be a gas pressure sensor in an enclosed water capaçitor resonant cavity which also includes a gas outlet. The gas pressure sensor is operatively connected to the circuit to determine the rate of gas production with respect to ambient gas pressure in the water capacitor enclosure.

Thus, an omnibus control circuit and its discrete elements for maintaining and controlling the resonance and other aspects of the release of gas from a resonant cavity water cell is described herein and illustrated in the drawings which depict the following:

Figure 1 is a block diagram of an overall control circuit showing the interrelationship of sub-circuits, the pulsing core/resonant circuit and the water capacitor resonant cavity.

Figure 2 shows a type of digital control means for regulating the ultimate rate of gas production as determined by an external input. (Such a control means would correspond, for example, to the accelerator in an automobile or a building thermostat control.)

Figure 3 shows an analog voltage generator.

Figure 4 is a voltage amplitude control circuit interconnected with the voltage generator and one side of the primary coil of the pulsing core.

Figure 5 is the cell driver circuit that is connected with the opposite side of the primary coil of the pulsing core.

Figures 6, 7, 8 and 9 relate to pulsing control means including a gated pulse frequency generator (**Figure 6**); a phase lock circuit (**Figure 7**); a resonant scanning circuit (**Figure 8**); and the pulse indicator circuit (**Figure 9**) that control pulses transmitted to the resonant cavity/water fuel cell capacitor.

Figure 10 shows the pulsing core and the voltage intensifier circuit that is the interface between the control circuit and the resonant cavity.

Figure 11 is a gas feedback control circuit.

Figure 12 is an adjustable frequency generator circuit.

The circuits are operatively interconnected as shown in Figure 1 and to the pulsing core voltage intensifier circuit of Figure 10, which, inter alia, electrically isolates the water capacitor so that it becomes an electrically isolated cavity for the processing of water in accordance with its dielectric resonance properties. By reason of the isolation, power consumption in the control and driving circuits is minimized when resonance occurs; and current demand is minimized as voltage is maximized in the gas production mode of the water capacitor/fuel cell.

The reference letters appearing in the Figures, A, B, C, D, E, etc., to M and M1 show, with respect to each separate circuit depicted, the point at which a connection in that circuit is made to a companion or interrelated circuit.

In the invention, the water capacitor is subjected to a duty pulse which builds up in the resonant changing choke coil and then collapses. This occurrence permits a unipolar pulse to be applied to the fuel cell capacitor. When a resonant condition of the circuit is locked-in by the circuit, amp leakage is held to a minimum as the voltage which creates the dielectric field tends to infinity. Thus, when high voltage is detected upon resonance, the phase lock loop circuit that controls the cell driver circuit maintains the resonance at the detected (or sensed) frequency.

The resonance of the water capacitor cell is affected by the volume of water in the cell. The resonance of any given volume of water maintained in the water capacitor cell is also affected by "contaminants" in the water which act as a damper. For example, at an applied potential difference of 2000 to 5000 volts to the cell, an amp spike or surge may be caused by inconsistencies in water characteristics that cause an out-of-resonance condition which is remedied instantaneously by the control circuits.

In the invention, the adjustable frequency generator (Figure 12) tunes into the resonant condition of the circuit including the water cell and the water therein. The generator has a frequency capability of 0 - 10 KHz and tunes into resonance typically at a frequency of 5 KHz in a typical 3.0 inch water capacitor formed of a 0.5 inch rod enclosed within a 0.75 inside diameter cylinder. At start up, in this example, current draw through the water cell will measure about 25 milliamp; however, when the circuit finds a tuned resonant condition, current drops to a 1-2 milliamp minimum leakage condition.

The voltage to the capacitor water cell increases according to the turns of the winding and size of the coils, as in a typical transformer circuit. For example, if 12 volts are sent to the primary coil of the pulsing core and the secondary coil resonant charging choke ratio is 30 to 1, then 360 volts are sent to the capacitor water cell. Turns are a design variable that control the voltage of the unipolar pulses sent to the capacitor.

The high speed switching diode shown in Figure 10 prevents charge leakage from the charged water in the water capacitor cavity, and the water capacitor as an overall capacitor circuit element, i.e., the pulse and charge status of the water/capacitor never pass through an arbitrary ground. The pulse to the water capacitor is always unipolar. The water capacitor is electrically isolated from the control, input and driver circuits by the electromagnetic coupling through the core. The switching diode in the VIC circuit (Figure 10) performs several functions in the pulsing. The diode is an electronic switch that determines the generation and collapse of an electromagnetic field to permit the resonant charging choke(s) to double the applied frequency and also allows the pulse to be sent to the resonant cavity without discharging the "capacitor" therein. The diode, of course, is selected in accordance with the maximum voltage encountered in the pulsing circuit. A 600 PIV fast switching diode, such as an NVR 1550 high speed switching diode, has been found to be useful in the circuit herein.

The VIC circuit of Figure 10 also includes a ferromagnetic or ceramic ferromagnetic pulsing core capable of producing electromagnetic flux lines in response to an electrical pulse input. The flux lines equally affect the secondary coil and the resonant charging choke windings. Preferably, the core is a closed loop construction. The effect of the core is to isolate the water capacitor and to prevent the pulsing signal from going below an arbitrary ground and to maintain the charge of the already charged water and water capacitor.

In the pulsing core, the coils are preferably wound in the same direction to maximize the additive effect of the electromagnetic field therein.

The magnetic field of the pulsing core is in synchronization with the pulse input to the primary coil. The potential from the secondary coil is introduced to the resonant charging choke(s) series circuit elements which are subjected to the same synchronous applied electromagnetic field, simultaneously with the primary pulse.

When resonance occurs, control of the gas output is achieved by varying voltage amplitude or varying the time of duty gate cycle. The transformer core is a pulse frequency doubler. In a figurative explanation of the workings of the fuel gas generator water capacitor cell, when a water molecule is "hit" by a pulse, electron time share is affected, and the molecule is charged. When the time of the duty cycle is changed, the number of pulses that "hit" the molecules in the fuel cell is correspondingly modified. More "hits" result in a greater rate of molecular disassociation.

With reference to the overall circuit of Figure 1, Figure 3 receives a digital input signal, and Figure 4 depicts the control means that directs 0-12 volts across the primary coil of the pulsing core. Depending upon design parameters of primary coil voltage and other factors relevant to core design, the secondary coil of the pulsing core can be set up for a predetermined maximum, such as 2000 volts.

Figure 5, the cell driver circuit, allows a gated pulse to be varied in direct relation to voltage amplitude.

As noted above, the circuit of Figure 6 produces a gate pulse frequency. The gate pulse is superimposed over the resonant frequency pulse to create a duty cycle that determines the number of discrete pulses sent to the primary coil. For example, assuming a resonant pulse of 5 KHz, a .5 Hz gate pulse may be superimposed over the 5 KHz pulse to provide 2500 discrete pulses in a 50% duty cycle per Hz. The relationship of resonant pulse to the gate pulse is determined by conventional signal addition/subtraction techniques.

Figure 7, a phase lock loop, allows pulse frequency to be maintained at a predetermined resonant condition sensed by the circuit. Together, the circuits of Figures 7 and 8 determine an output signal to the pulsing core until the peak voltage signal sensed at resonance is achieved.

A resonant condition occurs when the pulse frequency and the voltage input attenuates the covalent bonding forces of the hydrogen and oxygen atoms of the water molecule. When this occurs, amp leakage through the water capacitor is minimized. The tendency of voltage to maximize at resonance increases the force of the electric potential applied to the water molecules, which ultimately disassociate into atoms.

Because resonances of different waters, water volumes, and capacitor cells vary, the resonant scanning circuit of Figure 8 is useful. The scanning circuit of Figure 8 scans frequency from high to low to high repeating until a signal lock is determined. The ferromagnetic core of the voltage intensifier circuit transformer suppresses electron surge in an out-of-resonance condition of the fuel cell. In an example, the circuit scans at frequencies from 0 Hz to 10 KHz to 0 Hz. In water having contaminants in the range of 1 ppm to 20 ppm, a 20% variance in resonant frequency is encountered. Depending on water flow rate into fuel cell, the normal variance range is about 8-10%. For example, iron in well water affects the status of molecular disassociation. Also, at a resonant condition harmonic effects occur. In a typical operation of the cell with a representative water capacitor described below, at a frequency of about 5 KHz at unipolar pulses from 0 to 650 volts at a sensed resonant condition into the resonant cavity, conversion of about 5 gallons of water per hour into a fuel gas will occur on average. To increase the rate, multiple resonant cavities can be used and/or the surfaces of the water capacitor can be increased, however, the water capacitor cell is preferably small in scale. A typical water capacitor may be formed from a 0.5 inch in diameter stainless steel rod and a 0.75 inch inside diameter cylinder that together extend concentrically about 3.0 inches with respect to each other.

Shape and size of the resonant cavity may vary. Larger resonant cavities and higher rates of consumption of water in the conversion process require higher frequencies such as up to 50 KHz and above. The pulsing rate, to sustain such high rates of conversion must be correspondingly increased.

From the foregoing description of the preferred embodiment, other variations and modifications of the system disclosed will be evident to those of skill in the art.

WHAT IS CLAIMED IS:

1. A control circuit for a resonant cavity water capacitor cell utilized for the production of a hydrogen containing fuel gas including

an isolation transformer including a ferromagnetic core and having one side of a secondary coil connected in series with a high speed switching diode to one plate of the water capacitor of the resonant cavity and the other side of the secondary coil connected to the other plate of the water capacitor to form a closed loop electronic circuit utilizing the dielectric properties of water as part of the electronic circuit and a primary coil connected to a pulse generation means.

2. The circuit of Claim 1 in which the secondary coil includes segments that form a resonant charging choke circuit in series with the water capacitor.

3. The circuit of Claim 1 in which the pulse generation means includes an adjustable first frequency generator and a second gated pulse frequency generator which controls the number of pulses produced by the first frequency generator sent to the primary coil during a period determined by the gate frequency of the second pulse generator.

4. The circuit of Claim 1 further including a means for sensing the occurrence of a resonant condition in the water capacitor of the resonant cavity.

5. The circuit of Claim 4 in which the means for sensing is a pickup coil on the ferromagnetic core of the transformer.

6. The circuit of Claim 4 or Claim 5 in which the sensing means is interconnected to a scanning circuit and a phase lock loop circuit, whereby the pulsing frequency to the primary coil of the transformer is maintained at a sensed frequency corresponding to a resonant condition in the water capacitor.

7. The circuit of Claim 1 including means for adjusting the amplitude of a pulsing cycle sent to the primary coil.

8. The circuit of Claim 6 including further means for maintaining the frequency of the pulsing cycle at a constant frequency regardless of pulse amplitude.

9. The circuit of Claim 3 in which the gated pulse frequency generator is operatively interconnected with a sensor that monitors the rate of gas production from the cell and controls the number of pulses to the

cell in a gated frequency in a correspondence with the rate of gas production.

10. The circuit of Claim 7 or Claim 8 or Claim 9 further including a gas pressure sensor in an enclosed water capacitor resonant cavity which also includes a gas outlet, which gas pressure sensor is operatively connected to the circuit to determine the rate of gas production with respect to ambient gas pressure in the water capacitor enclosure.

11. The methods and apparatus as substantially described herein.

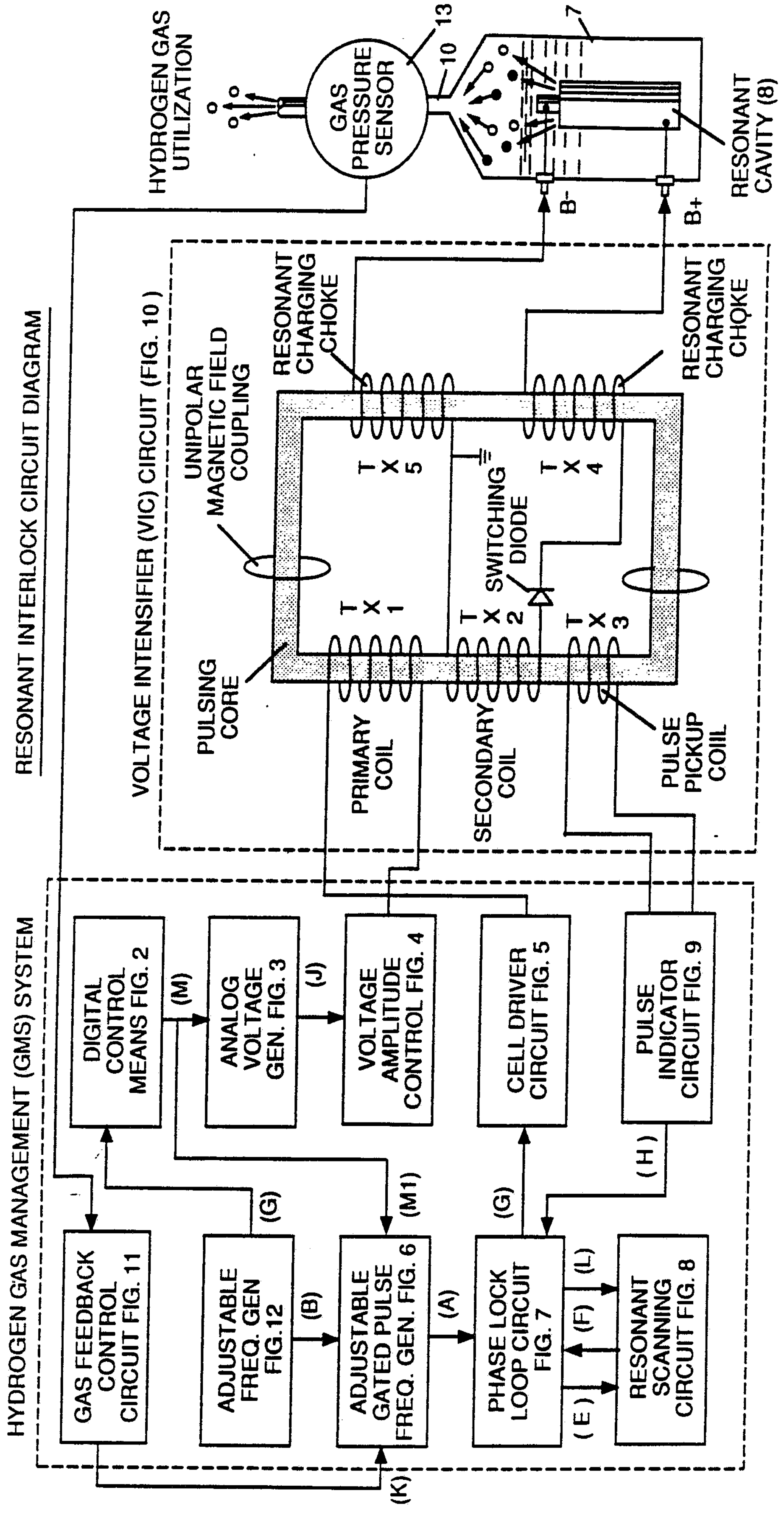
RESONANT INTERLOCK CIRCUIT DIAGRAM
HYDROGEN GAS MANAGEMENT (GMS) SYSTEM
VOLTAGE INTENSIFIER (VIC) CIRCUIT (FIG. 10)
HYDROGEN GAS UTILIZATION
GAS PRESSURE SENSOR
13
10
7
RESONANT CAVITY (8)
B-
B+
RESONANT CHARGING CHOKE
RESONANT CHARGING CHOKE
UNIPOLAR MAGNETIC FIELD COUPLING
PULSING CORE
T X 1
T X 2
T X 3
T X 4
T X 5
SWITCHING DIODE
PRIMARY COIL
SECONDARY COIL
PULSE PICKUP COIIL
GAS FEEDBACK CONTROL CIRCUIT FIG. 11
ADJUSTABLE FREQ. GEN FIG.12
ADJUSTABLE GATED PULSE FREQ. GEN. FIG. 6
PHASE LOCK LOOP CIRCUIT FIG. 7
RESONANT SCANNING CIRCUIT FIG. 8
DIGITAL CONTROL MEANS FIG. 2
ANALOG VOLTAGE GEN. FIG. 3
VOLTAGE AMPLITUDE CONTROL FIG. 4
CELL DRIVER CIRCUIT FIG. 5
PULSE INDICATOR CIRCUIT FIG. 9
(M)
(J)
(G)
(M1)
(G)
(H)
(B)
(A)
(L)
(F)
(E)
(K)

FIGURE 1

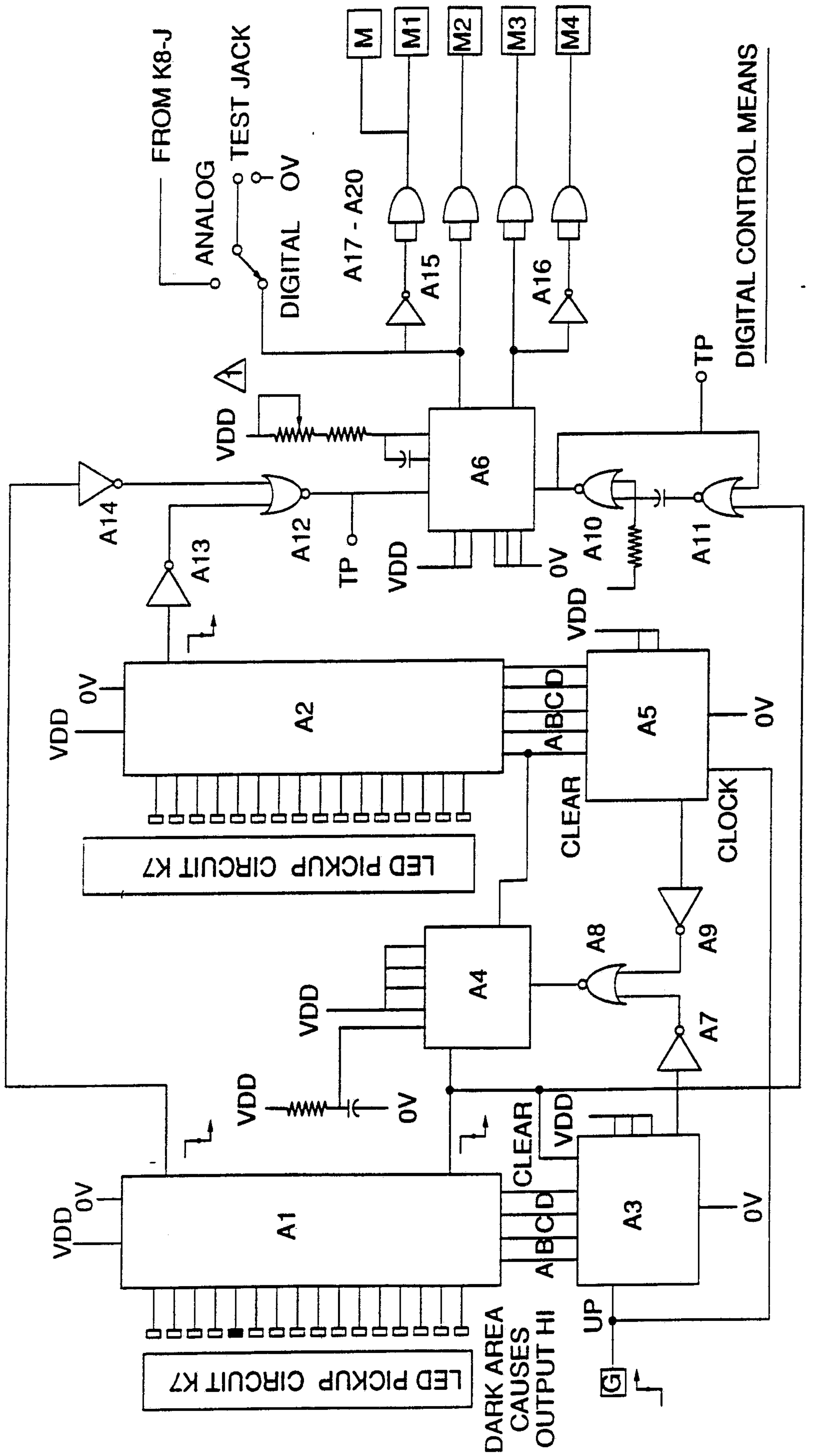
FROM K8-J
ANALOG
TEST JACK
DIGITAL
0V
A17 - A20
M
M1
M2
M3
M4
A15
A16
DIGITAL CONTROL MEANS
VDD
TP
A6
A14
A13
A12
TP
VDD
0V
A10
A11
VDD
0V
A2
A B C D
A5
0V
CLEAR
CLOCK
LED PICKUP CIRCUIT K7
A4
A8
A9
VDD
A7
VDD
0V
CLEAR
VDD
A1
A B C D
A3
0V
VDD
0V
UP
G
LED PICKUP CIRCUIT K7
DARK AREA CAUSES OUTPUT HI

FIGURE 2

⚠ ON CARD MOUNTING

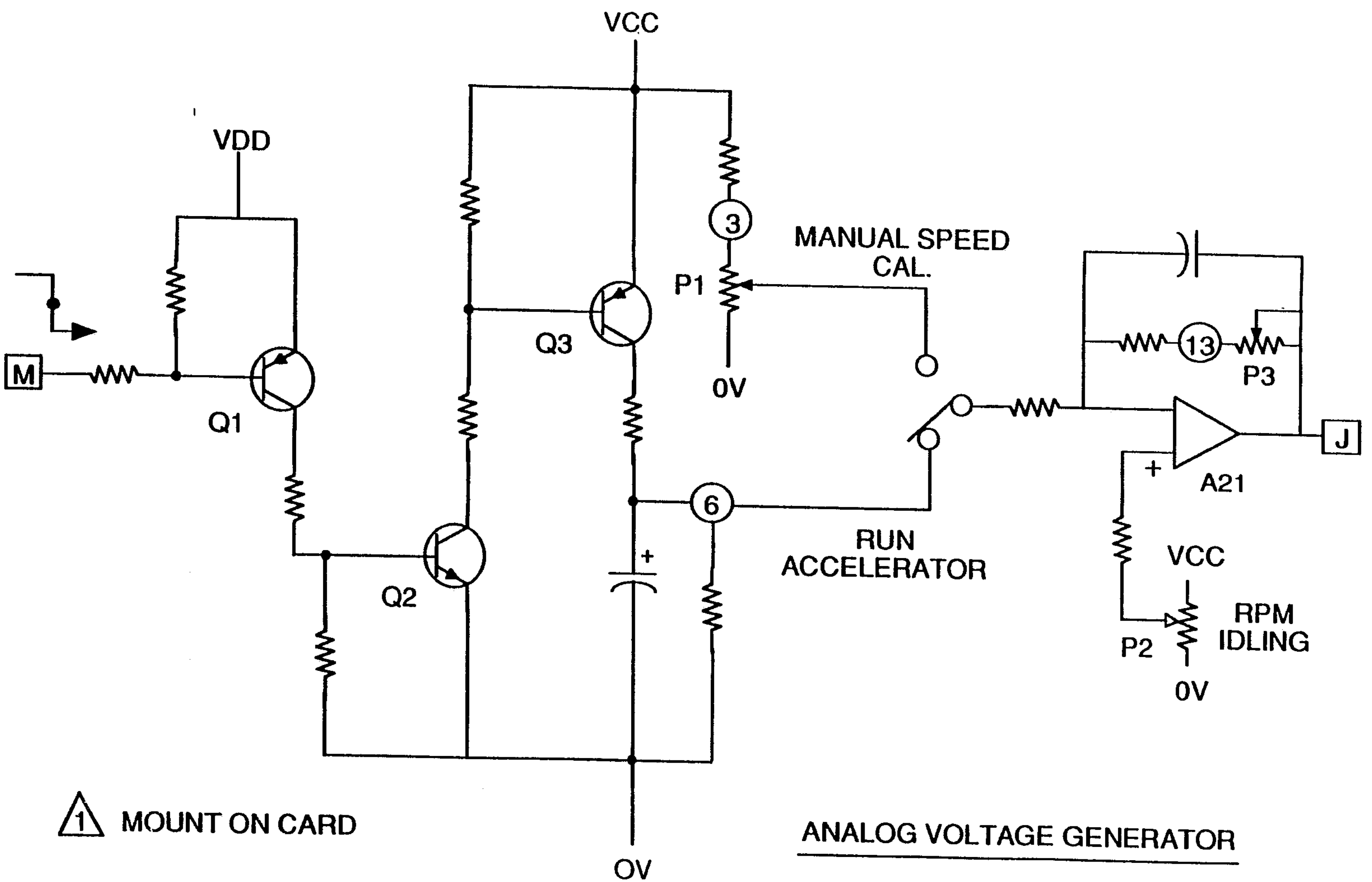
VCC
VDD
M
Q1
Q2
Q3
3
P1
0V
MANUAL SPEED CAL.
6
RUN ACCELERATOR
+
13
P3
A21
J
+
VCC
RPM IDLING
P2
0V
OV
1 MOUNT ON CARD
ANALOG VOLTAGE GENERATOR

FIGURE 3

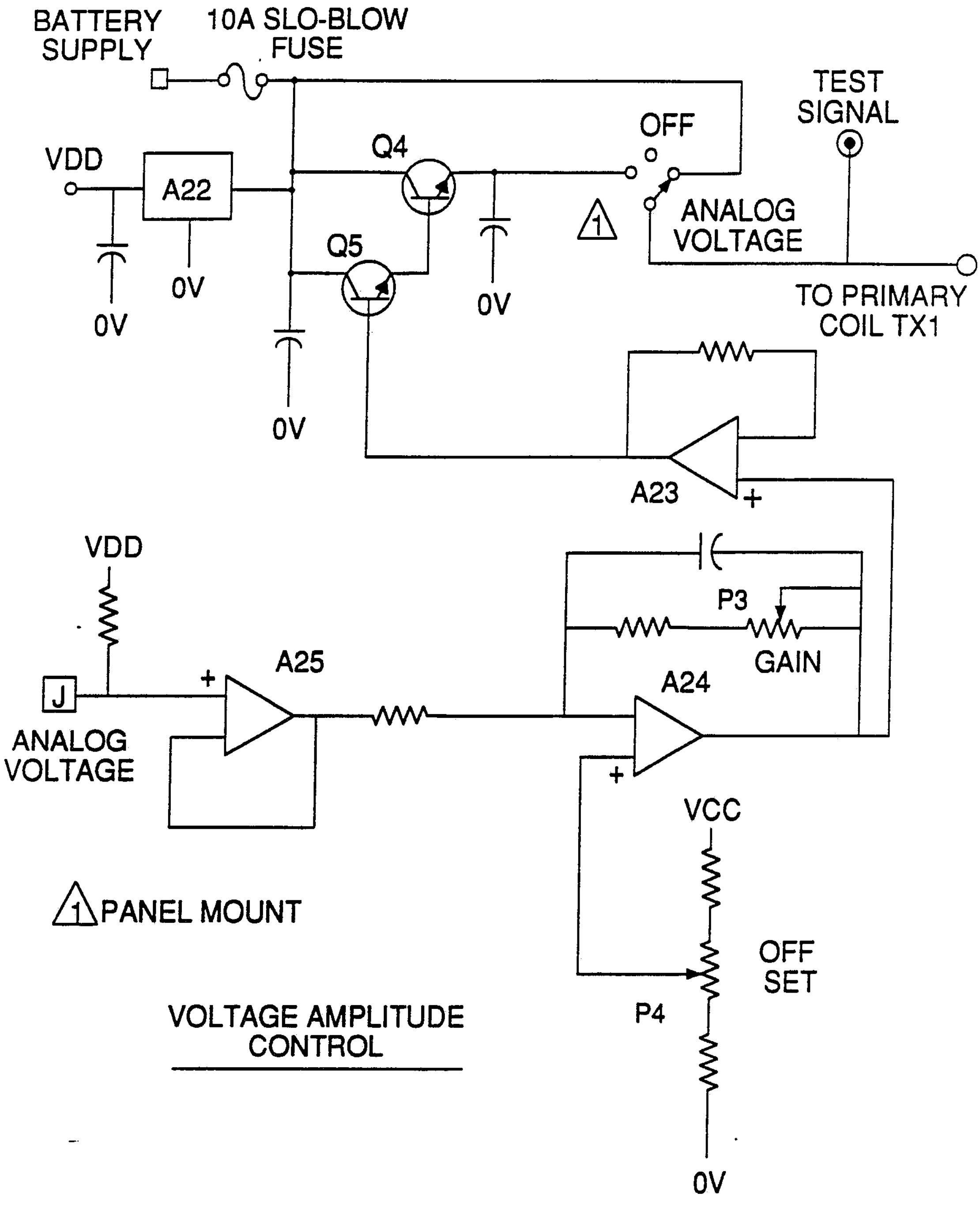
BATTERY SUPPLY
10A SLO-BLOW FUSE
TEST SIGNAL
OFF
VDD
A22
Q4
Q5
0V
0V
0V
0V
1
ANALOG VOLTAGE
TO PRIMARY COIL TX1
A23
+
VDD
P3
GAIN
A25
+
A24
J
ANALOG VOLTAGE
+
VCC
1 PANEL MOUNT
OFF SET
P4
0V

VOLTAGE AMPLITUDE CONTROL

FIGURE 4

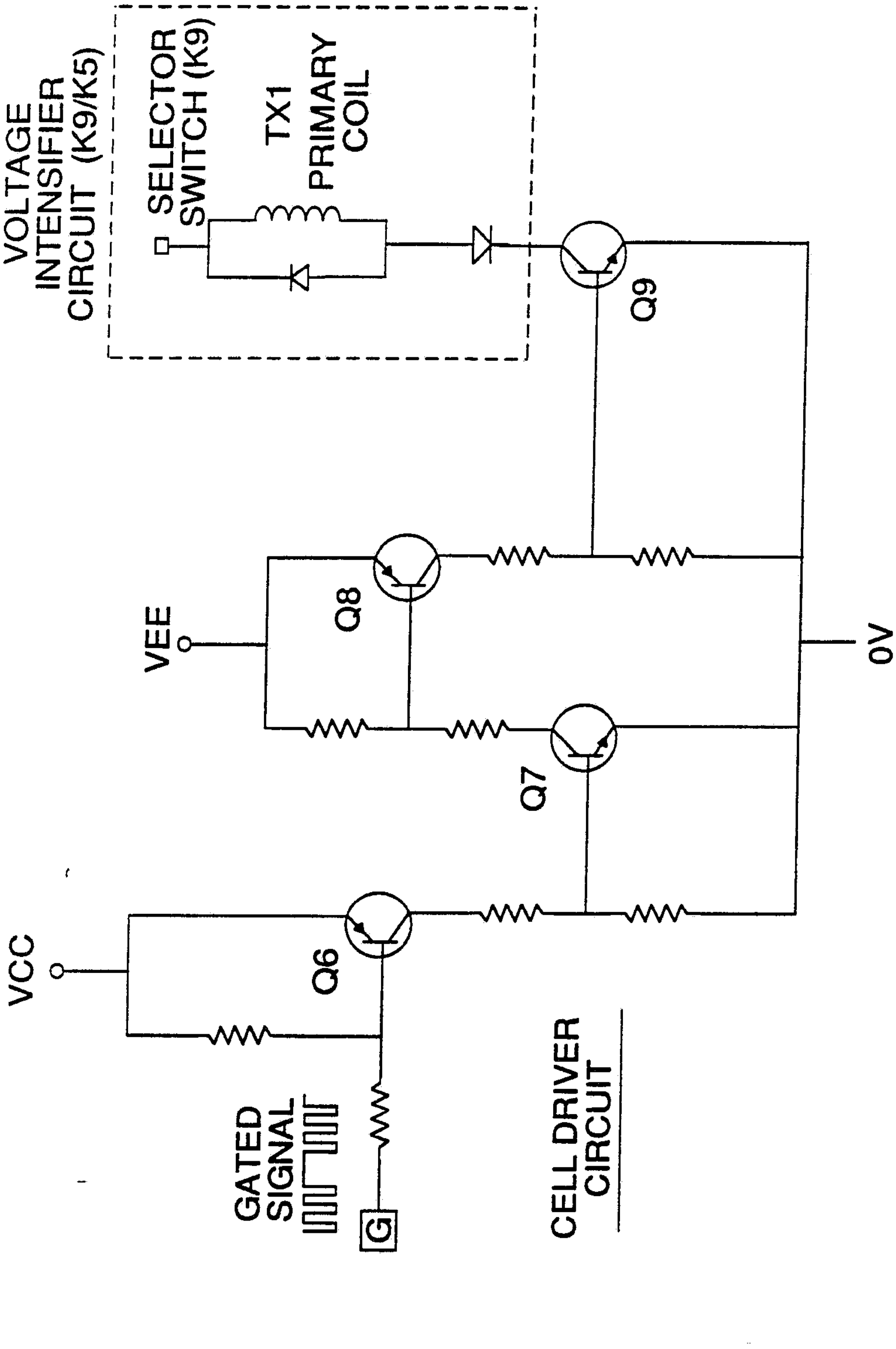
VOLTAGE INTENSIFIER CIRCUIT (K9/K5)
SELECTOR SWITCH (K9)
TX1 PRIMARY COIL
Q9
VEE
Q8
0V
Q7
VCC
Q6
GATED SIGNAL
G
CELL DRIVER CIRCUIT

FIGURE 5

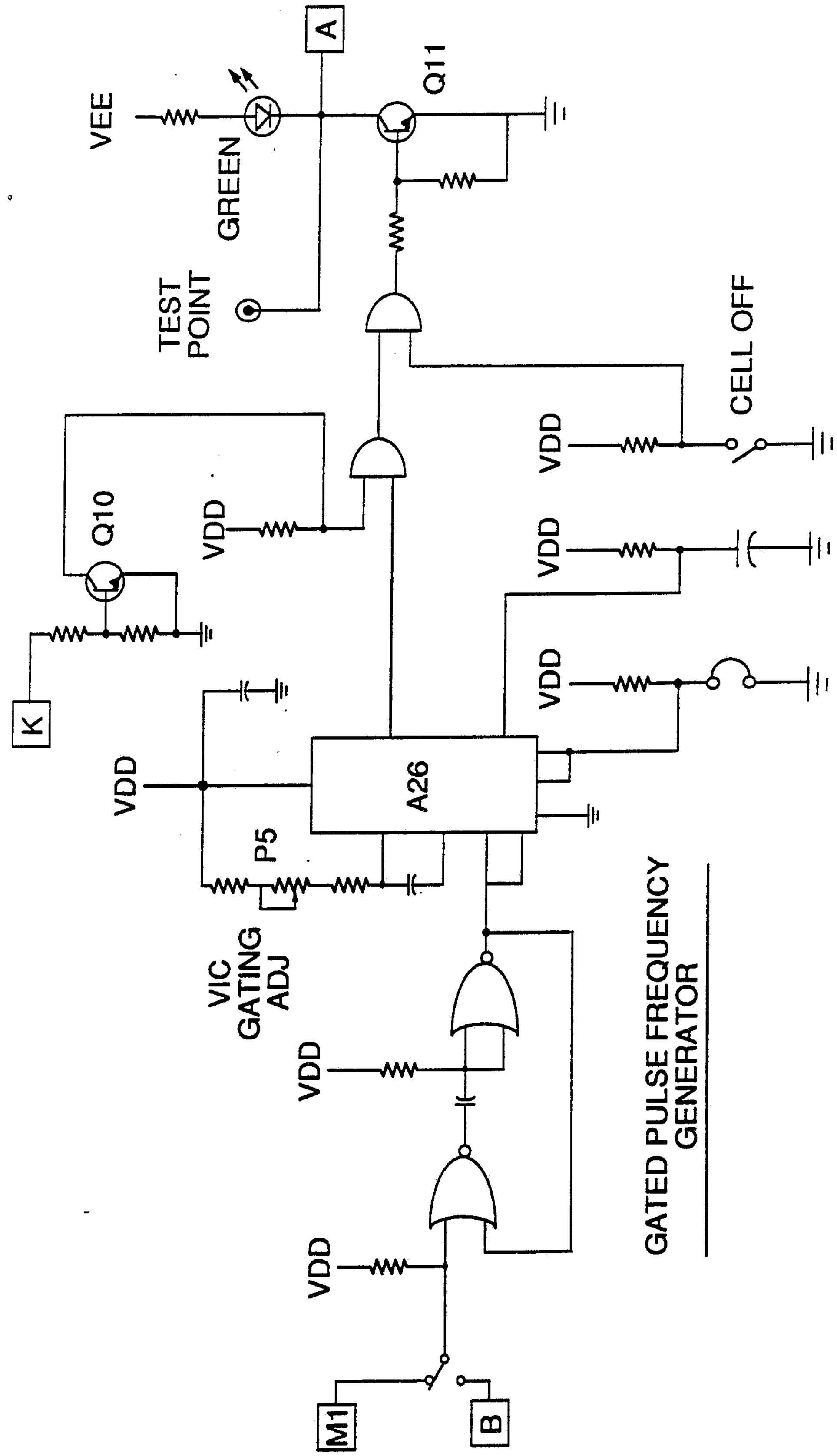
A
VEE
GREEN
Q11
TEST POINT
CELL OFF
VDD
VDD
VDD
VDD
Q10
K
VDD
A26
P5
VIC GATING ADJ
VDD
VDD
M1
B

FIGURE 6

GATED PULSE FREQUENCY GENERATOR

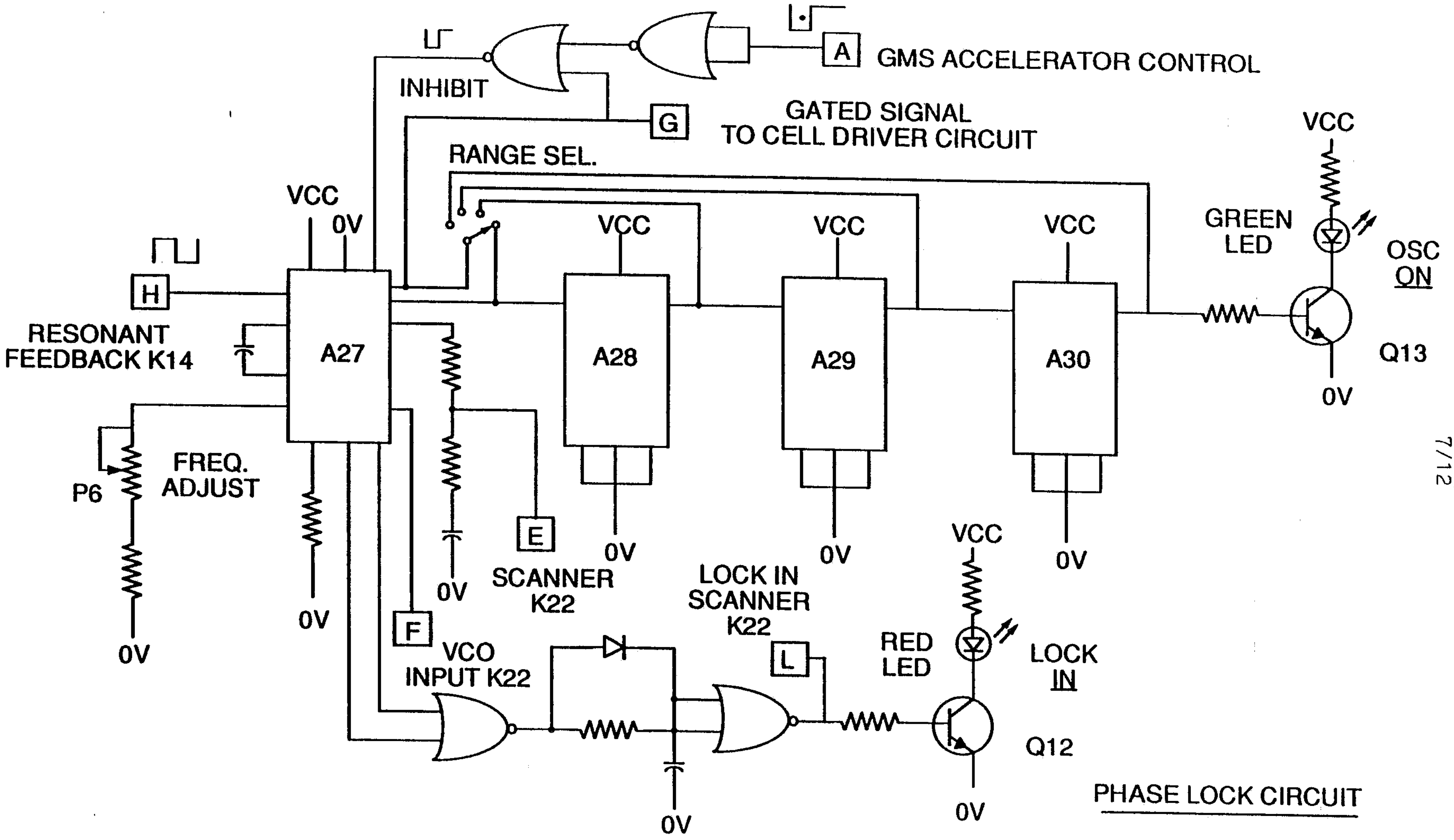
GMS ACCELERATOR CONTROL
A
INHIBIT
G
GATED SIGNAL TO CELL DRIVER CIRCUIT
RANGE SEL.
VCC
0V
H
RESONANT FEEDBACK K14
A27
P6
FREQ. ADJUST
0V
0V
0V
E
SCANNER K22
F
VCO INPUT K22
A28
A29
A30
VCC
VCC
VCC
0V
0V
0V
0V
LOCK IN SCANNER K22
L
RED LED
VCC
LOCK IN
Q12
0V
GREEN LED
VCC
OSC ON
Q13
0V

PHASE LOCK CIRCUIT

FIGURE 7

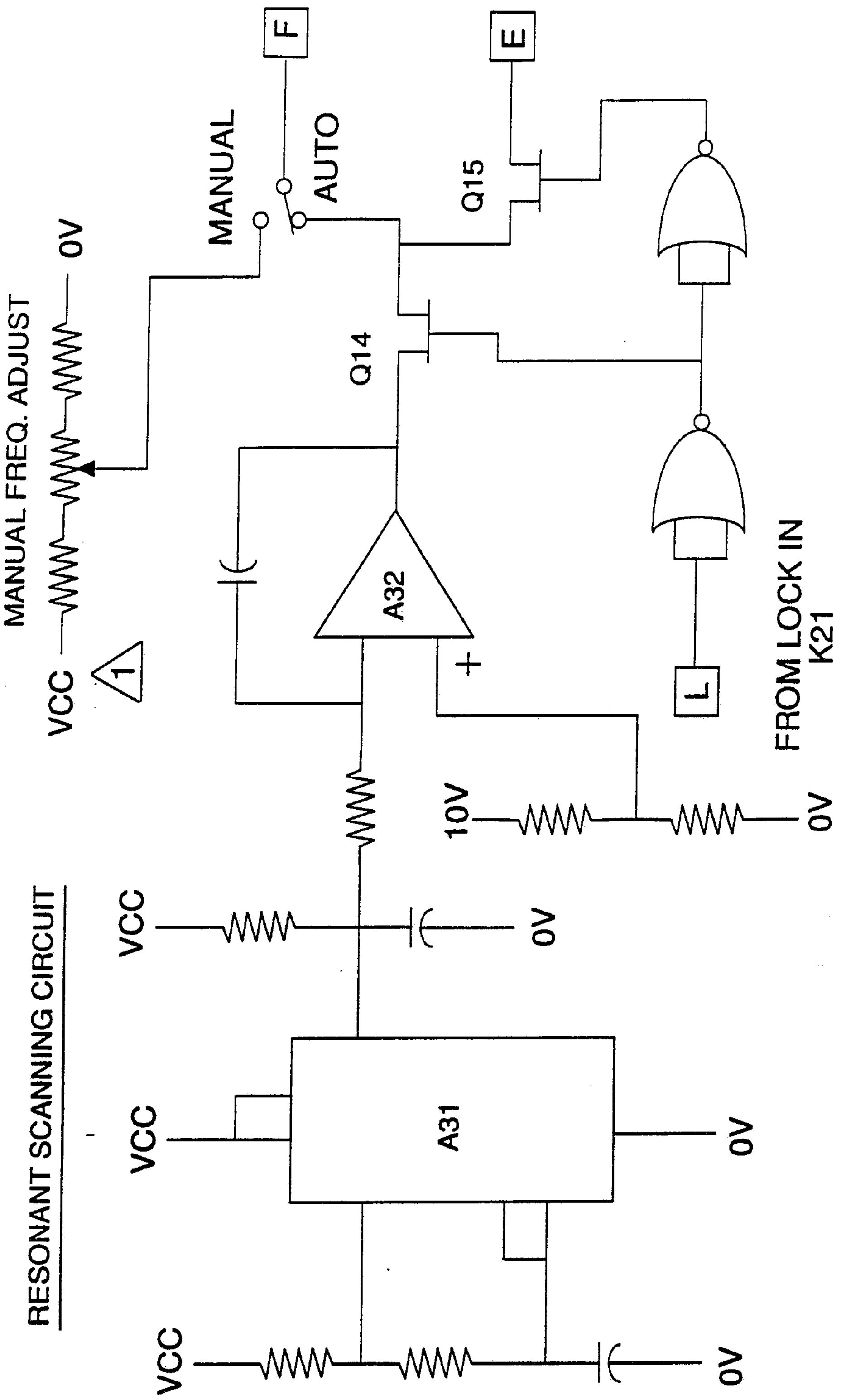
RESONANT SCANNING CIRCUIT
VCC
VCC
VCC
0V
0V
0V
A31
10V
0V
A32
+
VCC
1
MANUAL FREQ. ADJUST
0V
Q14
MANUAL
AUTO
F
Q15
E
L
FROM LOCK IN K21

FIGURE 8

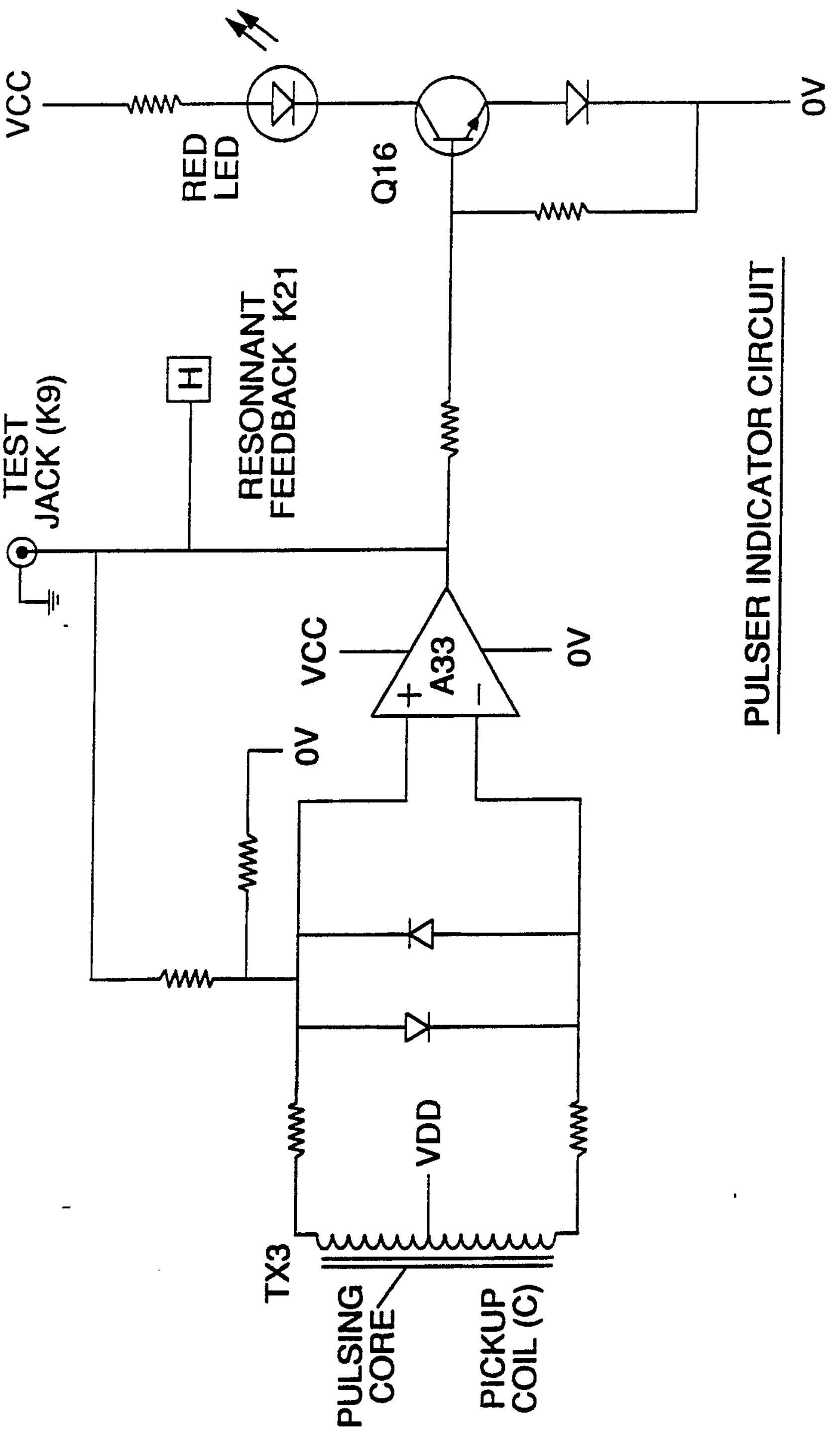
VCC
RED LED
Q16
0V
TEST JACK (K9)
H
RESONNANT FEEDBACK K21
VCC
A33
+
−
0V
0V
VDD
TX3
PULSING CORE
PICKUP COIL (C)

PULSER INDICATOR CIRCUIT

FIGURE 9

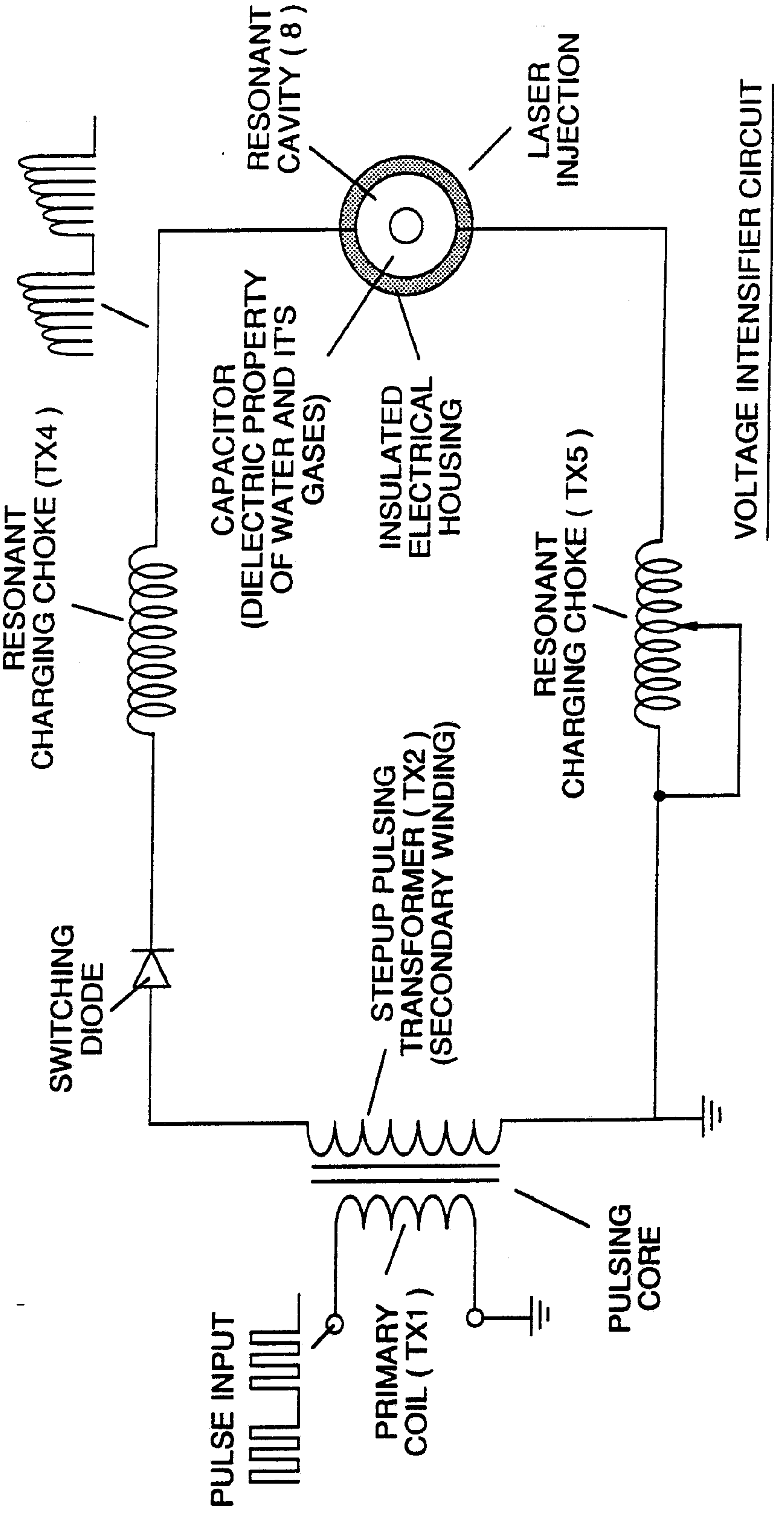
VOLTAGE INTENSIFIER CIRCUIT
RESONANT CAVITY (8)
LASER INJECTION
CAPACITOR (DIELECTRIC PROPERTY OF WATER AND IT'S GASES)
INSULATED ELECTRICAL HOUSING
RESONANT CHARGING CHOKE (TX4)
RESONANT CHARGING CHOKE (TX5)
SWITCHING DIODE
STEPUP PULSING TRANSFORMER (TX2) (SECONDARY WINDING)
PULSING CORE
PRIMARY COIL (TX1)
PULSE INPUT

FIGURE 10

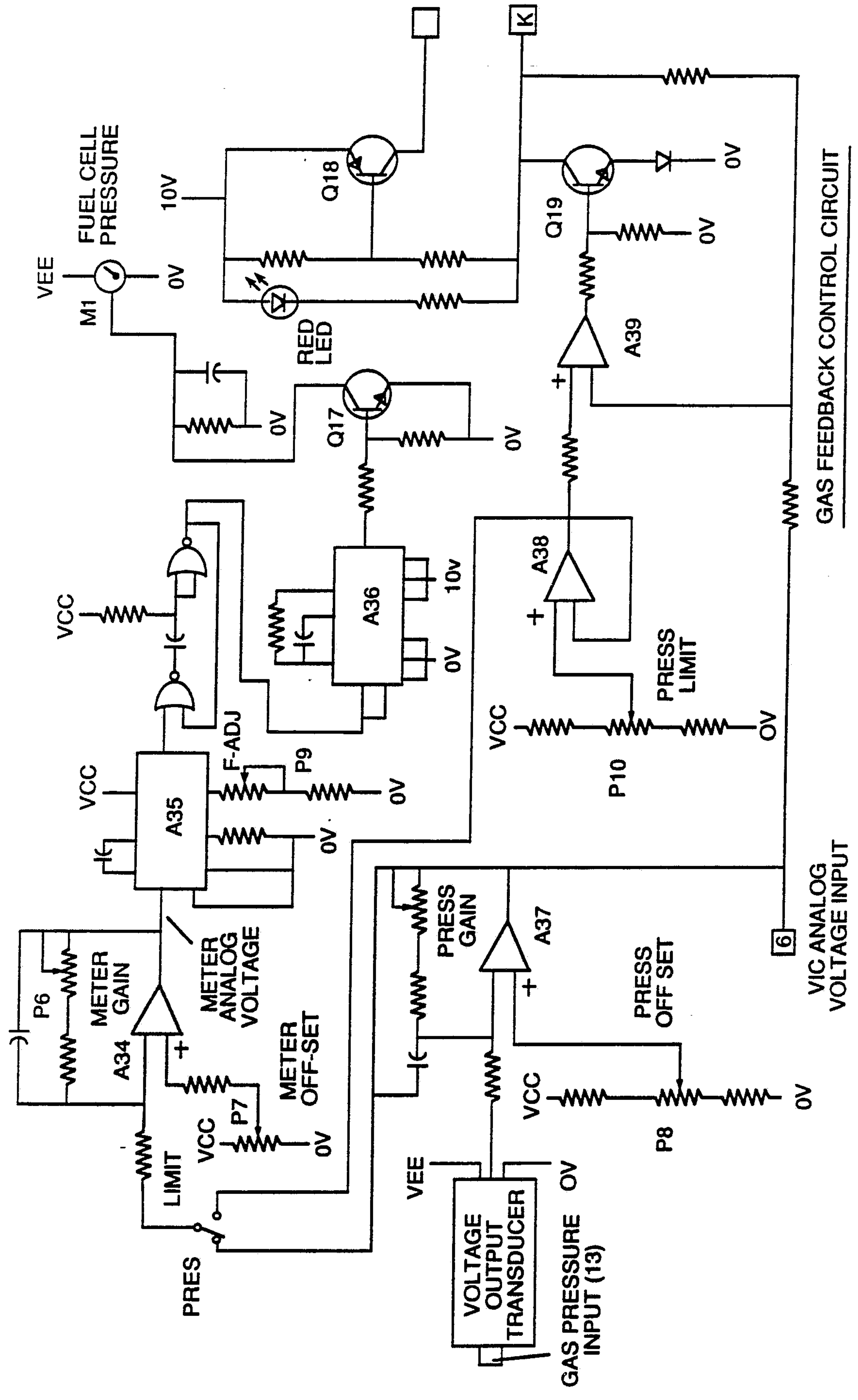
K
FUEL CELL PRESSURE
VEE
M1
0V
10V
Q18
RED LED
Q17
0V
0V
0V
Q19
0V
0V
A39
+
GAS FEEDBACK CONTROL CIRCUIT
VCC
A36
10v
0V
A38
+
PRESS LIMIT
VCC
0V
P10
F-ADJ
P9
VCC
A35
0V
0V
METER GAIN
P6
A34
+
METER ANALOG VOLTAGE
METER OFF-SET
P7
VCC
0V
LIMIT
PRES
PRESS GAIN
A37
+
PRESS OFF SET
VCC
P8
0V
VIC ANALOG VOLTAGE INPUT
6
VEE
OV
VOLTAGE OUTPUT TRANSDUCER
GAS PRESSURE INPUT (13)

FIGURE 11

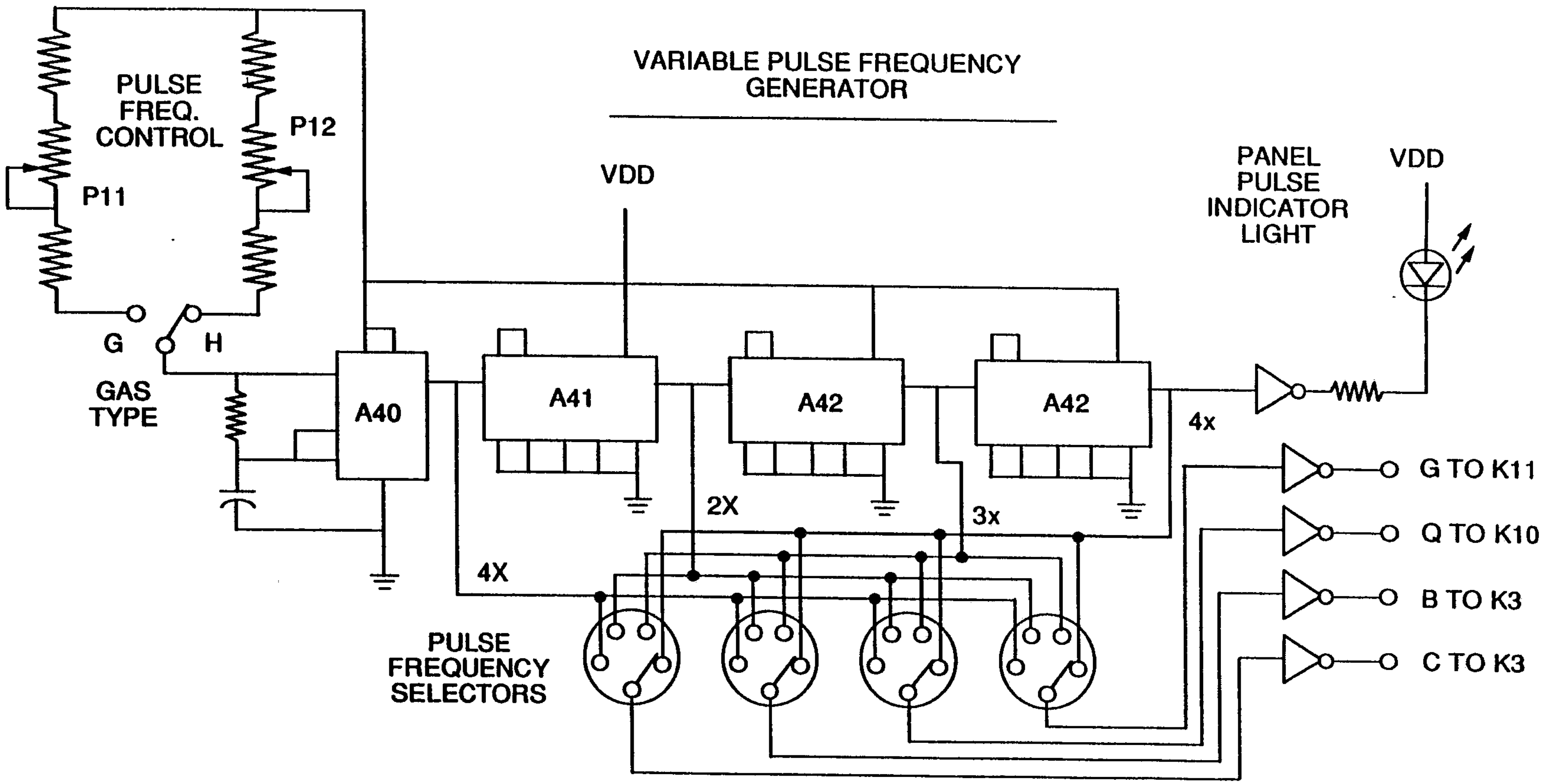
VARIABLE PULSE FREQUENCY GENERATOR
PULSE FREQ. CONTROL
P11
P12
G
H
GAS TYPE
VDD
A40
A41
A42
A42
2X
3x
4X
4x
PANEL PULSE INDICATOR LIGHT
VDD
PULSE FREQUENCY SELECTORS
G TO K11
Q TO K10
B TO K3
C TO K3

FIGURE 12

INTERNATIONAL SEARCH REPORT

International Application No PCT/US90/06407

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [3]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(5) : C07G 13/00, H03K 3/30
US CL : 204/157.52, 331/4, 307/106

II. FIELDS SEARCHED

Minimum Documentation Searched [4]

Classification System	Classification Symbols
US CL	204/157.42, 157.52; 422/186.16; 331/4,47,178, 307/106-109,271,282

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched [5]

III. DOCUMENTS CONSIDERED TO BE RELEVANT [14]

Category *	Citation of Document, [16] with indication, where appropriate, of the relevant passages [17]	Relevant to Claim No. [18]
X	US, A, 4,936,961 (MEYER) 26 JUNE 1990	1-3,7
Y	SEE FIGURES 1,2	4,6
Y	US, A, 4,275,363 (MISHIRO ET AL.) 23 JUNE 1981. SEE FIGURE 1.	4,6

* Special categories of cited documents: [15]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search [2] 12 APRIL 1991	Date of Mailing of this International Search Report [2] 07 MAY 1991
International Searching Authority [1] ISA/US	Signature of Authorized Officer [20] Andre Robinson DAVID A. OSBORN

Form PCT/ISA/210 (second sheet) (May 1986)

US006404089B1

(12) **United States Patent**
Tomion

(10) **Patent No.: US 6,404,089 B1**
(45) **Date of Patent: Jun. 11, 2002**

(54) **ELECTRODYNAMIC FIELD GENERATOR**

(76) Inventor: **Mark R. Tomion**, 2938 Ferguson Corners Rd., Geneva, NY (US) 14456

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/621,152**

(22) Filed: **Jul. 21, 2000**

(51) **Int. Cl.**^{7} **H02K 1/00**; H02K 5/00

(52) **U.S. Cl.** .... **310/162**; 290/1 A; 290/1 R; 244/23 A; 244/23 C

(58) **Field of Search** .... 310/162–168, 310/171, 112, 308–309, 74, 68 R, 90, 118, 216, 261, 254, 89, 258; 290/1 R, 1 A, 49; 244/1 N, 130, 62, 23 C, 205, 23 A

(56) **References Cited**

U.S. PATENT DOCUMENTS

1,889,208	A	*	11/1932	Masterson et al.	310/126
2,949,550	A		8/1960	Brown	
3,071,705	A		1/1963	Coleman et al.	
3,177,654	A		4/1965	Gradecak	
3,464,207	A		9/1969	Okress	
3,620,484	A	*	11/1971	Schoppe	244/130
3,662,554	A	*	5/1972	Broqueville	244/1 N
3,774,865	A	*	11/1973	Pinto	244/23 C
4,733,099	A		3/1988	Hutson, Jr.	
4,789,801	A		12/1988	Lee	
4,900,965	A		2/1990	Fisher	
5,291,734	A	*	3/1994	Sohnly	60/202
5,382,833	A		1/1995	Wirges	

FOREIGN PATENT DOCUMENTS

GB	2312709	5/1997
WO	WO85/03053	7/1985

* cited by examiner

Primary Examiner—Tran Nguyen
(74) *Attorney, Agent, or Firm*—Stephen B. Salai, Esq.; Brian B. Shaw, Esq.; Harter, Secrest & Emery LLP

(57) **ABSTRACT**

This device is a brushless high-voltage electrical generator, requiring suitable means of input rotary torque, for purposes of producing a very-high-energy external electrodynamic field or continuous quasi-coherent DC corona or arc discharge of uniform current density which completely encloses the machine's conductive housing. This housing is divided into distinct electrical sections and contains a flat conductive rotor which electrically links separate negative and positive housing sections and upon which a plurality of toroidal generating coils are rotatably mounted. Circular arrays of stationary permanent magnets are affixed within the housing which induce a constant DC voltage within said coils upon their rotation. The primary voltage so-generated is electrostatically impressed across the rotor such that great quantities of electronic charge may be transported between the opposite polarity sections of the housing, in such a manner that a much higher secondary voltage is caused to appear across interposed neutral sections thereof, and the resulting external breakdown current once initiated is independent of the generating coils' ampacity. Ancillary mechanical, electrical, an/or electronic features may be attached upon or within the housing to aid in harnessing and controlling the useful effects associated with the external dynamic electric field produced by the device.

48 Claims, 9 Drawing Sheets

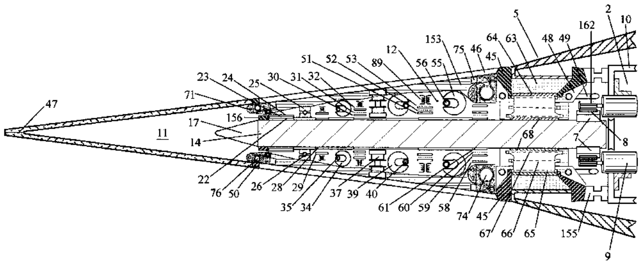

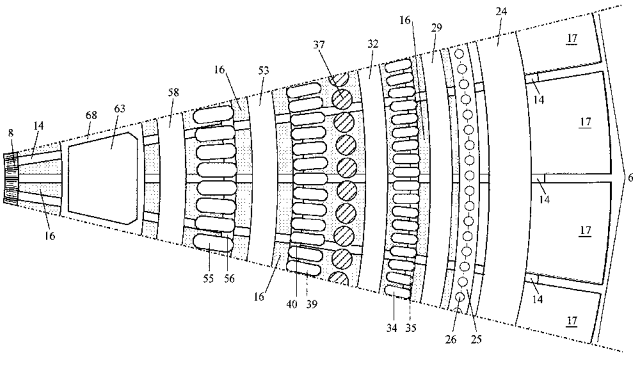

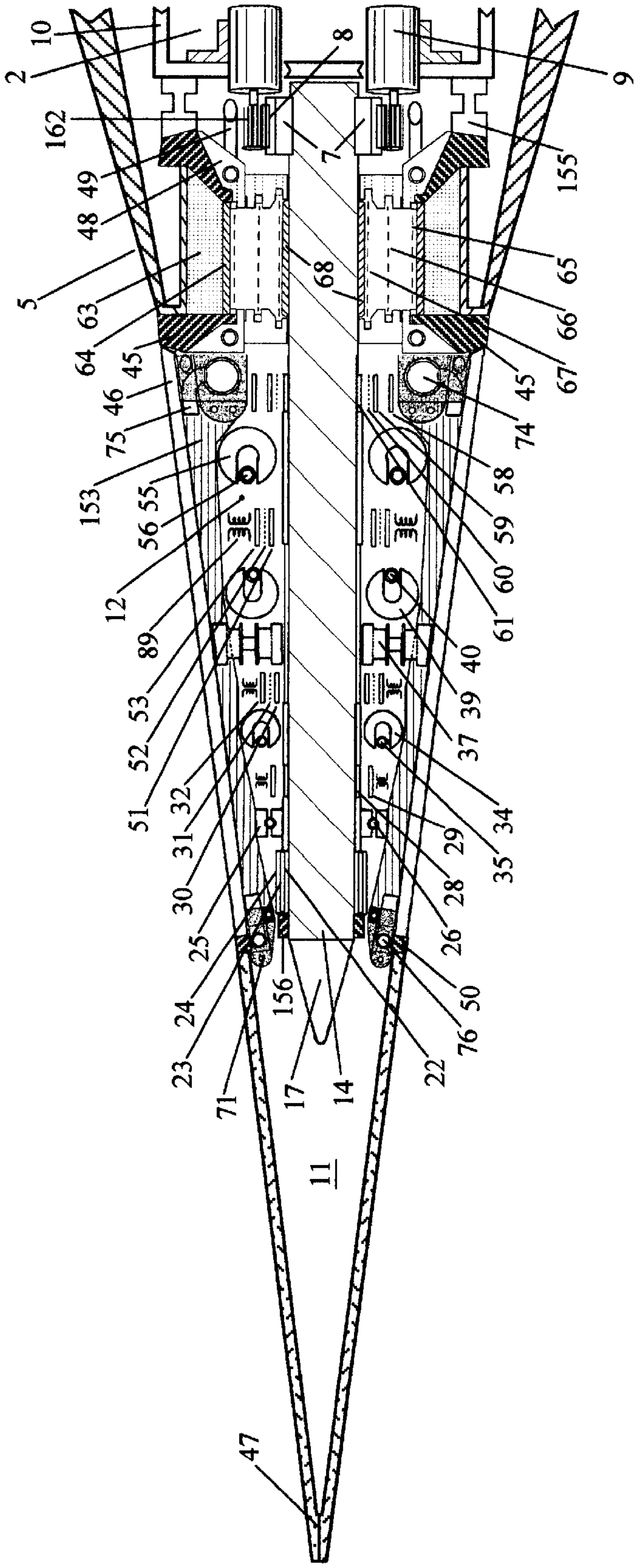

Fig. 1

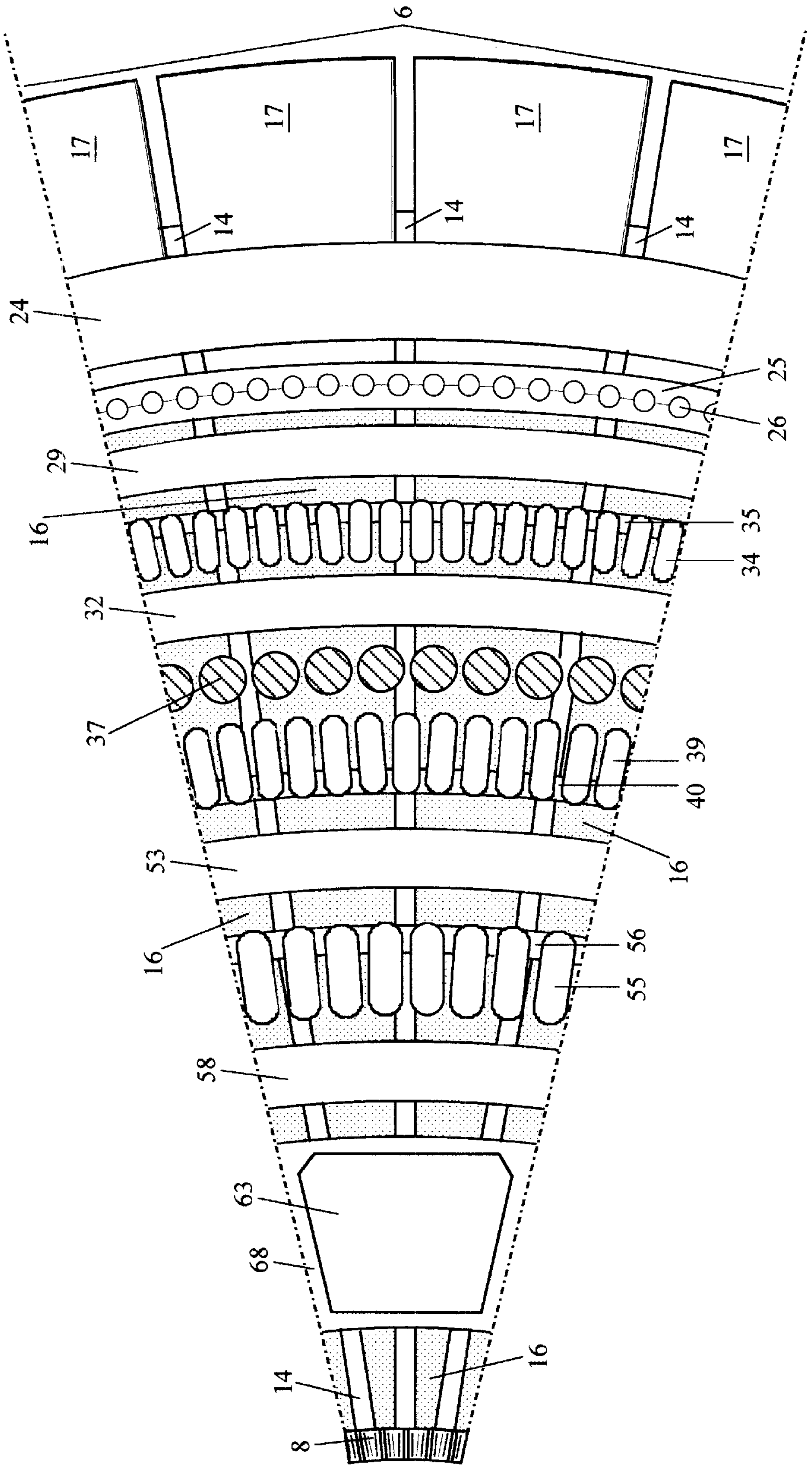

Fig. 2

Fig. 3

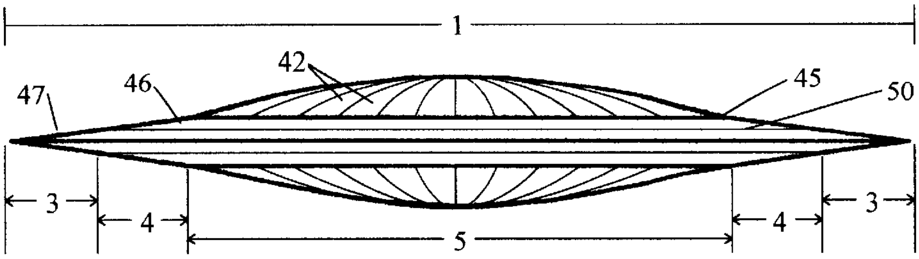
1
42
47
46
45
50
3
4
5
4
3

Fig. 4

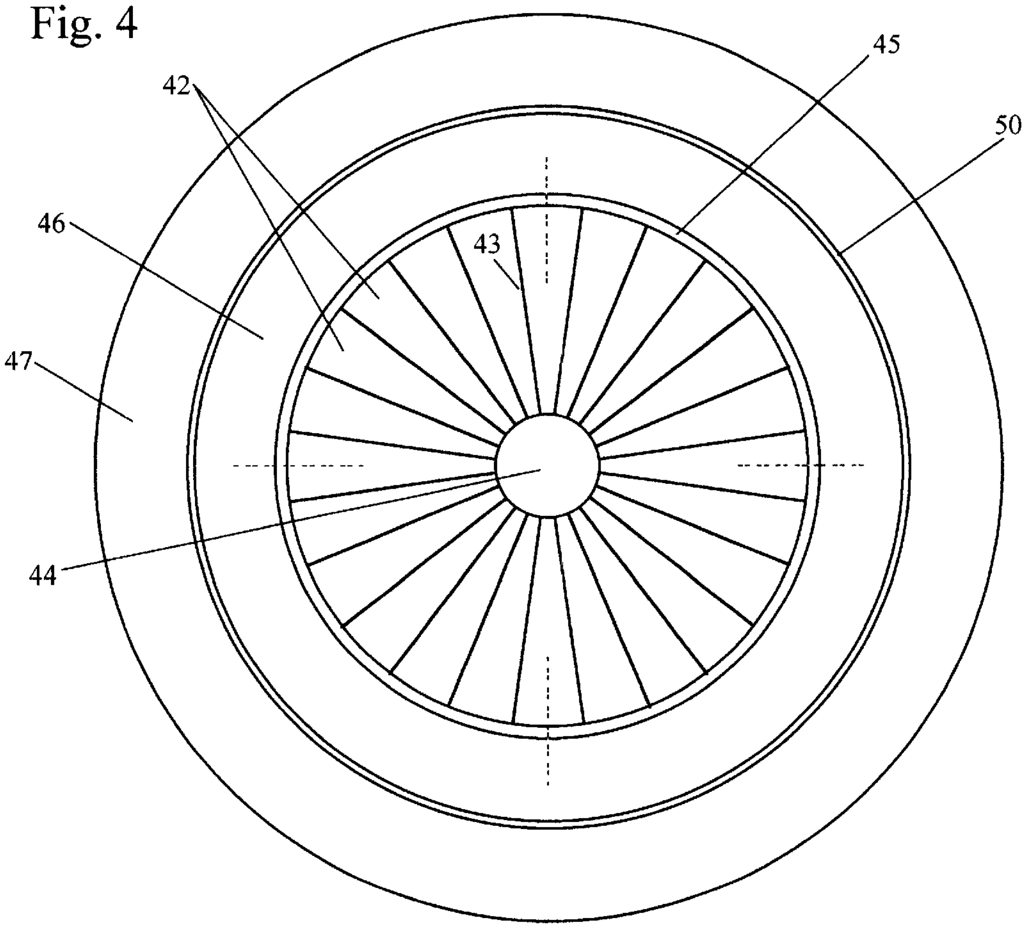
45
50
42
46
43
47
44

Fig. 5

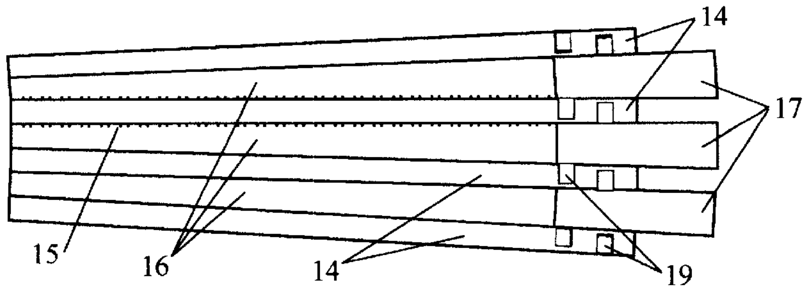
14
17
15
16
14
19

Fig. 6a

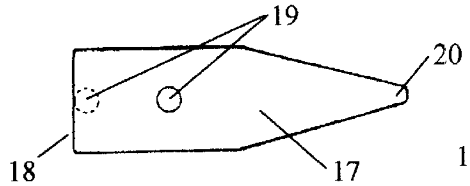

Fig. 6b

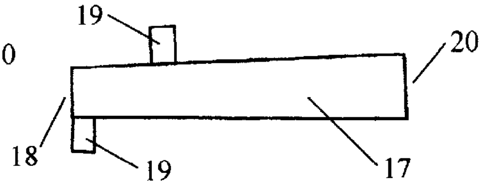
19
20
18
19
17

Fig. 7

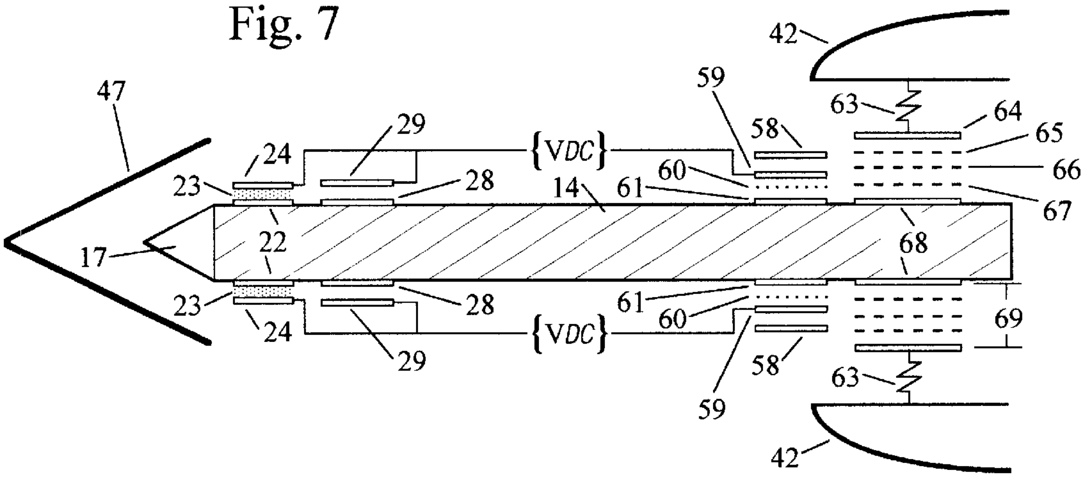
42
47
59
63
64
58
65
29
{VDC}
24
60
66
23
28
14
61
67
68
17
22
68
61
69
23
28
60
24
{VDC}
29
58
63
59
42

Fig. 8

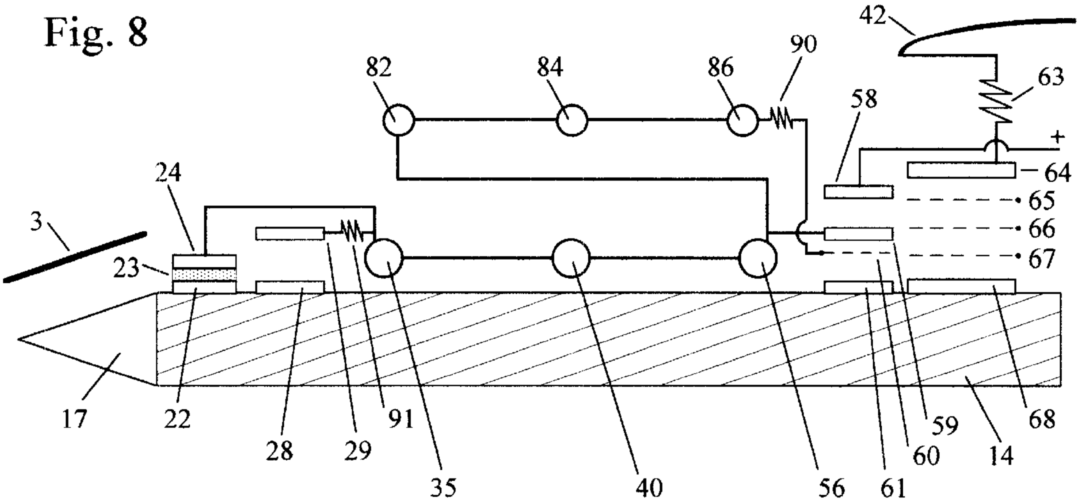
+

Fig. 9

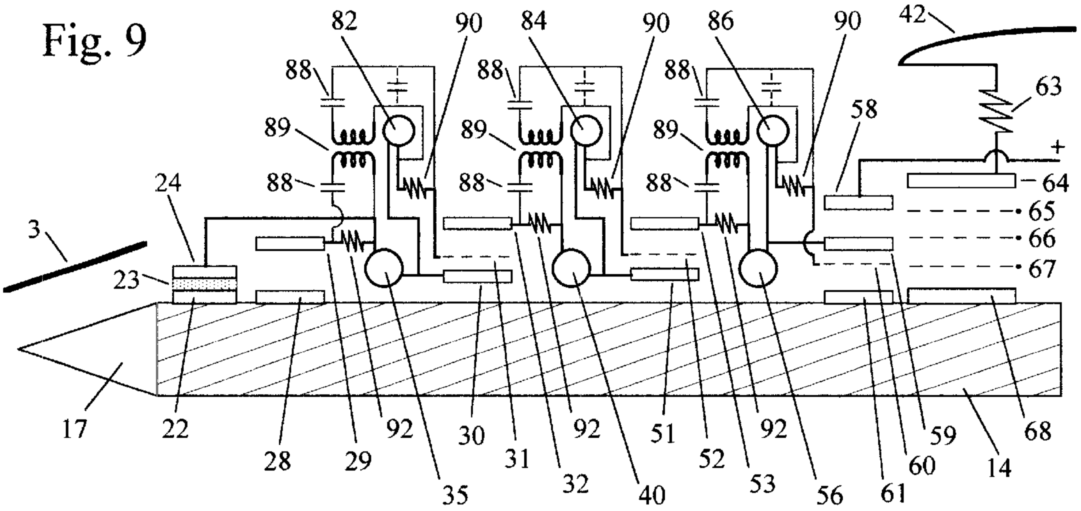
+

Fig. 10

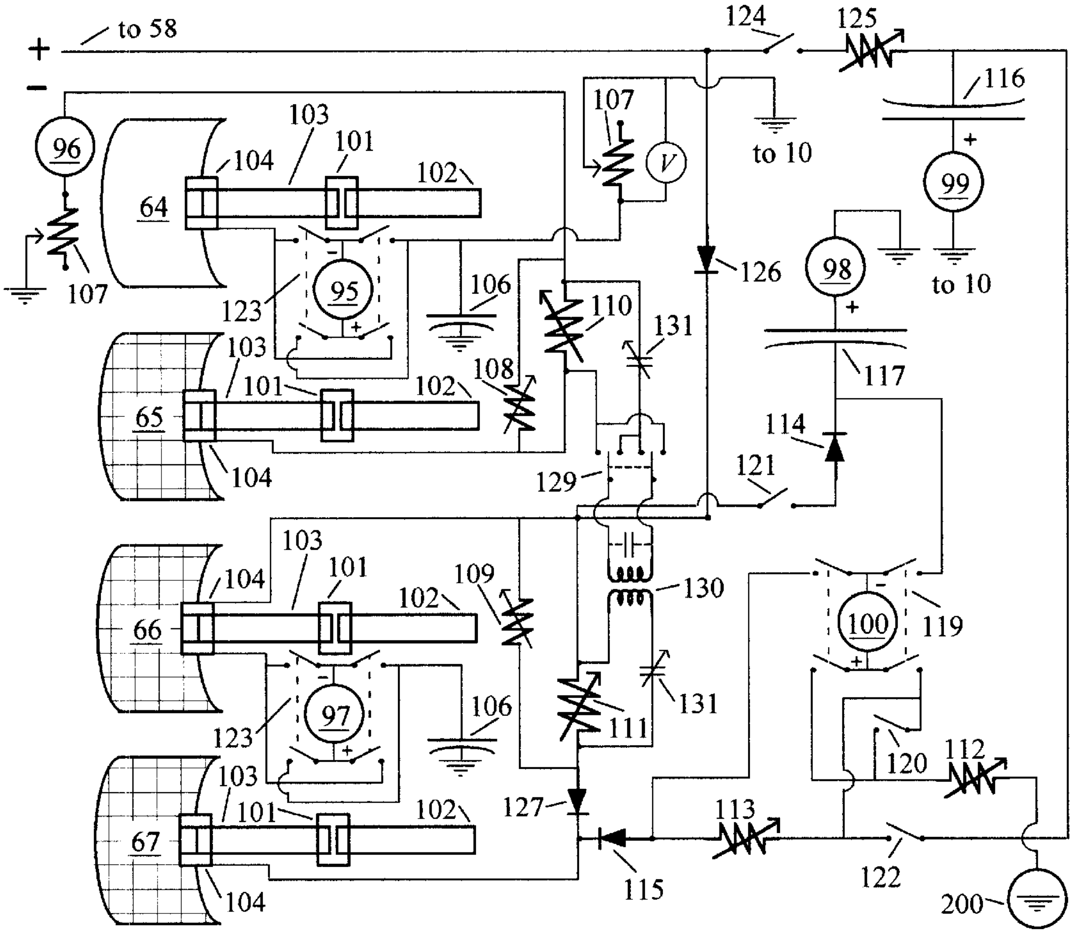
+
to 58
−
V
to 10
+
to 10
+

Fig. 11

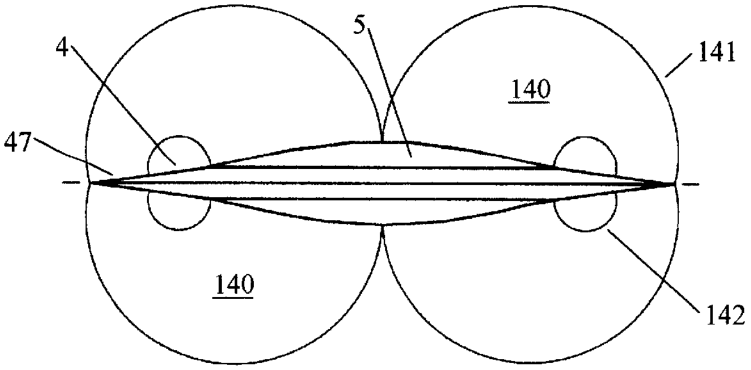
5
4
141
140
47
140
142

Fig. 12

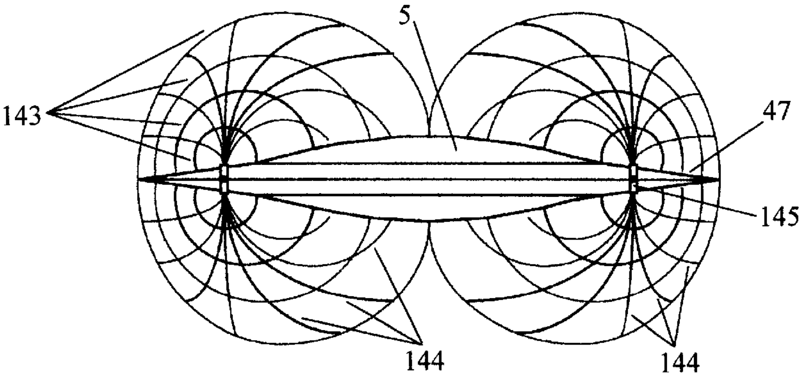
5
143
47
145
144
144

Fig. 13

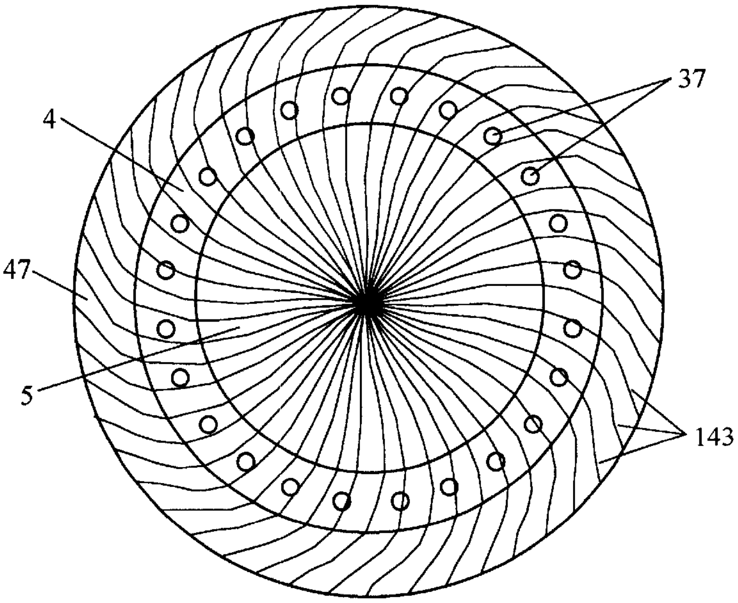

Fig. 14

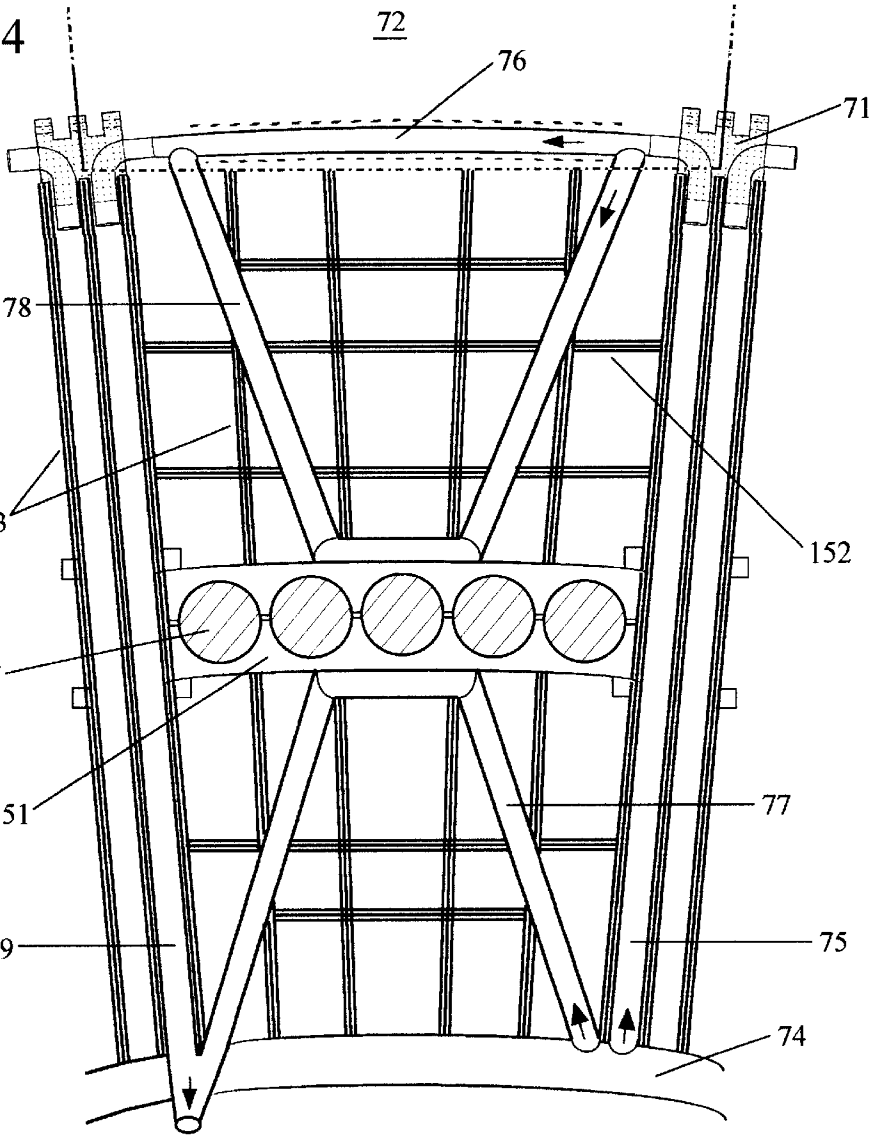

Fig. 15

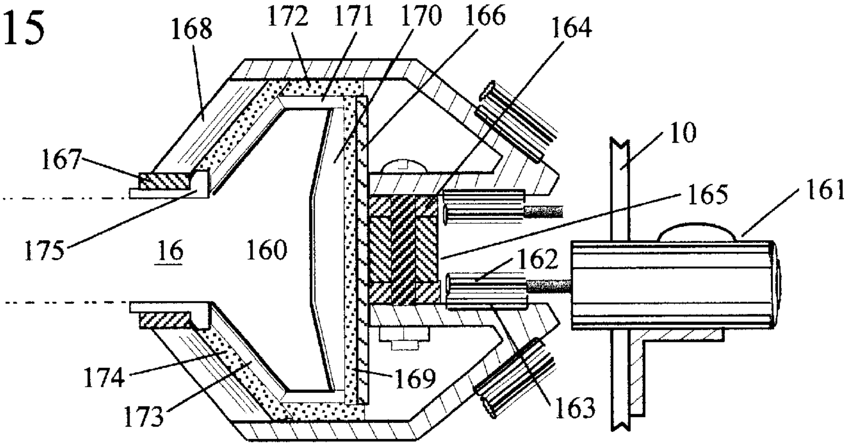

Fig. 16

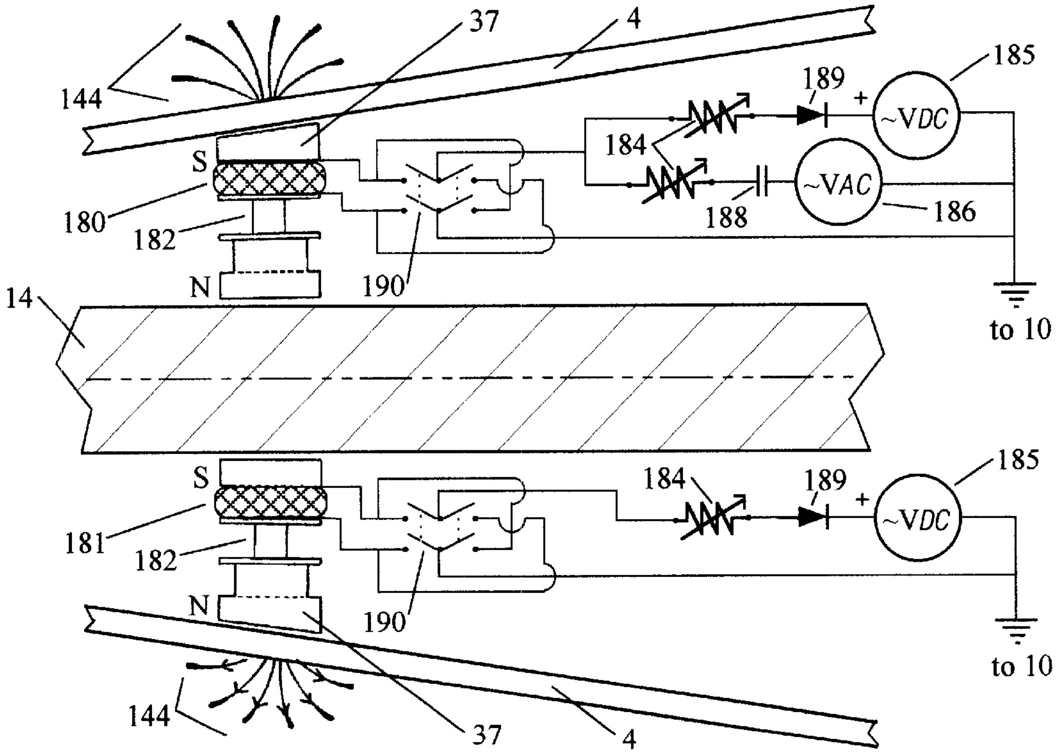
+
~VDC
S
~VAC
N
to 10
S
+
~VDC
N
to 10

ELECTRODYNAMIC FIELD GENERATOR

FIELD OF THE INVENTION

The present invention is related in general to rotary direct current (DC) electrical generators which incorporate permanent magnet fields, and in particular to similar electromechanical generating devices whose principal purpose is to produce a useful continuous high-voltage DC corona discharge. This invention also pertains to the field of aerospace vehicles which are capable of propulsion through the utilization of an ultra-high-voltage corona or arc discharge of a special type and form.

DESCRIPTION OF RELATED ART

In order to provide background information so that the present invention may be completely understood and appreciated in its proper context, reference is made hereinbelow to a number of related art patents. These cited references contain certain similarities to the present invention, primarily related to the objective of producing electrically-developed thrust. However, key differences and limitations with regard to achieving this objective are in evidence.

The related art disclosed in U.S. Pat. No. 2,949,550 proposes three attendant objects which are quite similar to certain of the objects of the present invention: (i) to provide an apparatus for the direct conversion of electrical potential to usable kinetic energy; (ii) to provide such an apparatus having a hollow body or housing which contains a source of high electrical potential; and (iii) to provide a self-propelling vehicular apparatus which includes a pair of electrically conductive body portions joined by an insulative portion, whereby said conductive portions constitute electrodes. A further similarity is that the vehicular embodiment proposed preferably be of a circular disc shape somewhat thicker in its center than at its edges. The device set forth in this Patent just cited is noticeably different from the present invention, however, in that a positive voltage is applied to the housing periphery and a negative voltage is applied at the central axis. The proposed arrangement is furthermore limited to operation within a gaseous dielectric medium, upon which it is reliant for the production of the claimed motive force, and is to be operated at a potential gradient less than that which would produce a visible corona. This device moreover makes no provisions for the production and/or extraction of vast quantities of useful thermal kinetic energy, as the present invention does.

The related art disclosed in U.S. Pat. No. 3,071,705 is based on three empirical principles of electrostatics which also figure prominently in explaining the actual shape of the discharge current field that will be produced by the present invention: (i) electrostatic lines of force tend to concentrate on the surface of a charged conductor in those places with the smallest radius of curvature; (ii) they are normal to the surface from which they emanate; and (iii) they do not cross but will bend under the influence of another charged body. The device set forth in this Patent just cited is in fact quite similar to that of the Patent cited immediately before, in that relatively massive positive atmospheric ions and entrained neutral air molecules are repelled from the positive electrode toward the negative electrode along a linear axis—thereby inducing an air flow or "electric wind" which represents the fundamental source of thrust. It is therefore once again limited to operation in a gaseous dielectric medium, and does not make use of a flow of negative and greatly less massive electron current in the course of thrust production—which may achieve comparatively enormous acceleration and even relativistic velocity and specific impulse at a similar given level of applied potential gradient—as the present invention does.

The related art disclosed in U.S. Pat. No. 3,177,654 actually addresses the issue of producing electrically-developed thrust which is not limited to atmospheric operation, and the stated primary object thereof is to provide a propulsion system for enabling controlled flight in the atmosphere which may continue extensively into space without changing the system's basic operation. One embodiment thereof also proposes a circular disc-shaped vessel wherein high-voltage electrode elements ionize atmospheric gases by corona discharge, and a pulsed electromagnet arrangement causes the ejection of the resultant plasma from the propulsion chamber—producing thrust by reaction. This device furthermore addresses the issue of producing local thrust differential(s) on the body or housing thereof, for purposes of flight maneuvering and/or directional control, by providing for a plurality of such pulsed discharge propulsion units arranged in concentric rings (within the discoid body) and having individual divergent nozzles. As such, however, the device set forth in this Patent just cited actually constitutes an ionized jet propulsion method, requiring the use of either a gaseous medium like air or the onboard storage and release of relatively massive ionizable reaction material for any operational capability in space. Moreover, the pulsed mode of operation used, which limits the ionizing potential gradient to a value of less than breakdown field intensity, significantly restricts the level of power produced as thrust as compared to the continuous DC discharge field current produced by the present invention—at voltages which may be greatly in excess of breakdown intensity. Finally, this said device relies upon an unspecified source of generated electrical power in operation which must be assumed would have to be nuclear if conventional in nature, for said device to avoid otherwise having to carry a massive and range-limiting onboard store of fuel, and not upon convenient and protracted-use permanent magnets of high energy density as the prime energy source (as the present invention does).

U.S. Pat. No. 3,464,207 discloses a quasi-corona aerodynamic vehicle which bears only a limited relationship to the present invention, but which is referenced nonetheless as being further indicative of the predominant trend in conceptualization in relation to devices intended for producing electrically-developed thrust. This device also relies upon pulsed corona discharge operation between spaced electrodes at less than breakdown field intensity, thereby greatly constraining the propulsive power obtained. It is furthermore effective only in a gaseous dielectric medium and is based upon an electric wind pressure gradient arising of an admitted incompletely-understood mechanism. This same device evidences a great deal of theoretical complexity and construction material specificity, which might seem to indicate a narrow range of feasibility. Such attributes are not uncommon in the field of electrical aerospace propulsion, and are, to a degree, shared even by the present invention. Nevertheless, it is interesting to note that sizeable thrust in the form of an aerodynamic pressure gradient or lift of up to 30 lbs./ft^2 of external electrode area or more is asserted to be obtainable from said same device by utilizing asymmetrical electrodes or electrode arrays, rather than using electrodes or electrode arrays of approximately equal area (as the present invention does)—which therefore present a uniform cross section of discharge current conduction.

PCT Application WO 85/03053 discloses a flying apparatus whose primary similarity to the present invention is that it incorporates a hollow annular flywheel which provides gyroscopic stability and encircles a relatively stationary central cabin and/or payload area. The device set forth therein also provides free-rolling means to center and stabilize the flywheel or rotor within a discoid housing, and addresses the issue that horizontal flight maneuvering and/or directional control may be accomplished by altering the local angular acceleration of the rotor—as does the present invention. This said other device, however, relies upon the questionable concept that the very large angular momentum of the rotating flywheel is somehow able to compensate the potential energy caused by gravitation and that the overall weight of the apparatus may thereby be annulled. Vertical lift must then be provided for said device in operation by atmospheric thermal updraft means, and no electrical discharge means are utilized to produce a force of lift or thrust.

The device set forth in UK Patent GB 2 312 709 A is a flying craft which is essentially disc- or saucer-shaped and which has a central axial post electrode and a spaced enclosing ring electrode whereby an electrical arc may be struck between the two, to either ignite fuel introduced via nozzles or greatly heat an exhausted air flow produced by fan means. Thus, this device actually provides hybrid jet or turbo-fan means for producing propulsive thrust, and utilizes direct electrical discharge not as the primary source of thrust but to enhance thrust produced by other means. To this end, said device must therefore operate solely in a gaseous atmosphere and/or carry onboard a range-limiting store of combustible fuel. This same device does, however, utilize powerful magnetic fields to impart a rotational and de-randomizing aspect to the arc discharge, thereby promoting a more uniform distribution of the arc discharge field's energy, as does the present invention.

In summary, a number of similar or significant electrical, magnetic, and mechanical features of the several related art devices have been examined but, whatever the precise merits and advantages thereof, none of said devices achieve the correct combination of such features required to fulfill the purposes of the present invention. In this respect it may be stated that none of these related art devices attain the distinction of being able to operate extensively in the vacuum of space without the onboard storage of relatively massive reaction material for the production of thrust. Furthermore it would appear that none of these same devices utilize a charged conductive housing comprising suitably disposed electrodes and/or electrode arrays to create and control an asymmetrical electromotive field of force on itself without expelling a reaction mass, whether in a gaseous dielectric medium or the vacuum of space, to achieve a propulsive object as the present invention does. Finally, it is observed that none of the aforementioned related art devices can develop this type of useful thrust-producing electrodynamic field in a manner which allows more than one objective function to be fulfilled thereby, such as the production of useable thermal kinetic energy and/or the development of potential signal communications capability.

SUMMARY OF THE INVENTION

The present invention is a brushless permanent magnet electrical generator which produces a useful ultra-high-energy external electrodynamic field, or continuous quasi-coherent DC corona or arc discharge of uniform current density, that completely surrounds and encloses the machine's symmetrically disc-shaped conductive housing and tends in consequence toward a balanced bihemitoroidal shape.

The invention incorporates a uniform arrangement of stationary electromagnetic armatures which may be employed as a group to impart a holistic and de-randomizing rotational force to the external discharge current that acts to inhibit the formation of concentrated arc streamer or channel phenomena, and thus to assist in rendering the electrodynamic field quasi-coherent in nature. The arrangement of electromagnetic armatures in said generator acting as a group simultaneously also contribute a portion of the machine's required input rotary torque upon the current-carrying rotor assembly thereof.

This invention provides a permanent magnet high-voltage generator wherein the primary DC voltage induced within its rotating coils is electrostatically impressed across the conductive rotor assembly thereof, in a brushless manner utilizing vacuum tube cutoff bias techniques, such that a secondary DC corona or arc discharge current is thereby initiated about the housing thereof which is orders of magnitude larger than the operating current through the said coils.

The said high-voltage primary generator of the invention then has a housing or hull which is divided into separate negative (emitter), neutral, and positive (collector) electrical sections, such that the continuous external flow of such discharge current is best facilitated and conducted internally by the proximal rotor, and wherein suitable means to circulate one or more form(s) of liquid coolant through any one or more of said housing sections may be employed for purposes of extracting an extremely high level of recoverable thermal energy (arising from such discharge current)—which is intended principally for use in the large-scale production of electrical power and/or distilled or desalinated water.

This invention as an electrodynamic field generator also provides an appreciable and practical level of electrically-developed thrust, from the relativistic impulse of an induced ultra-high-voltage external DC arc discharge current incident upon its positive (collector) housing sections, which may be rendered variably non-isometric in nature and therefore directionally propulsive by means of selectively controlling: (i) the resistance of the brushless electrical linkages between the positive housing sections and the proximal current-carrying rotor; and/or (ii) the magnetic flux output of the various stationary electromagnetic armatures.

The invention as an ultra-high-energy field generator furthermore presents a unique yet practicable method of producing a variable electromagnetic and/or gravimetric signal, by means of which useful information may be both transmitted to and received by a separate similar device, in further harnessing the effects associated with the device's external electromotive field. The relationship of said field effects to quantum potential vacuum fluctuation theory just now being studied, which may allow the devising of such a signal having an unconventional nature, is referred to again below.

In any event, the electrodynamic field generator described may be feasibly manufactured and marketed in various sizes according to need or desire in pursuit of its remarkable objects; may be operated either in a gaseous dielectric medium or in the vacuum of space; and may incorporate ancillary mechanical, electrical, and/or electronic features upon or within its housing to accomplish all the objects set forth hereinabove.

Briefly, in accordance with the preferred embodiment of the invention, first provided is a flat annular segmented rotor of very high conductive ampacity which is rendered capacitive in nature by the attachment of suitably paired electrode rings. Alternating in placement amidst such capacitive ring pairs are also attached on the rotor a series of toroidal generating coils of very large mean radius relative to the cross-sectional radius thereof. The resultant composite rotor assembly is then rotatably mounted within a sectional conductive housing. The potential energy inherent in the fields of ring-shaped stationary permanent magnets, which are affixed in concentric circular arrays within the housing, is utilized to electrically polarize the rotor across its radial annular width by means of the "primary" DC voltage created upon relative rotational motion between such coils and magnets. The thereby polarized rotor is electrically linked on one hand (about its positive inner circumference) to two axially-central collector housing sections, by electrostatic induction via a system of plane-parallel electrode elements, and on the other hand (about its negative outer circumference) to two radially-peripheral emitter housing sections across an evacuated space charge chamber. The rotor's subsequent transportation and storage of electronic charge between the proximal opposite-polarity sections of the housing causes a much higher "secondary" voltage to appear across interposed neutral sections thereof and results in a continuous ultra-high-energy DC discharge current flow about the housing externally and across the rotor internally.

BRIEF DESCRIPTION OF THE DRAWINGS

The benefits and advantages attendant to the invention and its objects as discussed hereinabove are best illuminated through the following drawings and description.

FIG. **1** shows a side sectional view of the invention's peripheral and principal working structure, comprising a rotor assembly as well as frame-mounted elements.

FIG. **2** shows a top sectional view of a 30° radial sector of that same working structure.

FIG. **3** shows a side view of the invention's symmetrical housing or hull configuration.

FIG. **4** shows a top view or bottom view of that same housing/hull configuration.

FIG. **5** depicts a small radial sector of the basic rotor assembly construction, in top view, details of the rotor segments, inter-segment insulators, and field emitters.

FIG. **6***a* and **6***b* show side and top views, respectively, of one of the invention's conductive rotor field emitters.

FIG. **7** shows a simplified schematic representation of the invention's two "primary induction ring" electrode arrays, in relation to the rotor electrical circuit.

FIG. **8** shows a schematic of one symmetrical half of the primary power system circuit of the invention, in a "single-stage" rotor construction embodiment.

FIG. **9** shows a schematic of one symmetrical half of that same primary power system circuit, in a "three-stage" rotor construction embodiment.

FIG. **10** shows a schematic of the field voltage control system circuit of the invention.

FIG. **11** shows a simplified diagram of the electrodynamic field "envelope" produced by the invention, around a side view of the invention's housing/hull.

FIG. **12** shows the diagram of FIG. **11**, with the addition of lines indicating the electric and magnetic field vectors comprising that same electrodynamic field envelope.

FIG. **13** shows a top or bottom view corresponding to the side view of FIG. **12**.

FIG. **14** shows a top or bottom view of a radial sector of the invention's neutral hull section, showing the housing cooling and/or heat transfer system arrangements.

FIG. **15** shows a detail of an embodiment of the preferred central rotor segment attachment means and rotor drive assembly.

FIG. **16** shows a schematic of the stationary armature power circuit of the invention.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Method of Operation

The invention disclosed herein is an Electrodynamic Field ("EDF") Generator. The primary challenge involved in fulfilling the objects of the present invention is how to generate and maintain a multi-megavolt potential difference across the exterior sectional housing which is not immediately neutralized due to uncontrolled direct arc discharge. The first part of the problem in this respect is that it is not possible to produce this kind of voltage directly within a traditional generating coil arrangement, so the voltage applied to the charged housing sections must be achieved by special electrostatic means. Secondly, the discharge current must be produced and conducted internally in such a manner that: (i) only a tiny percentage thereof is actually conducted by the generating coils (within their very limited ampacity compared to that of the rotor); and (ii) the negative and positive housing sections are each charged faster than they are discharged until an enormous equilibrium field potential difference is established which is a sum of roughly equal negative and positive housing section voltages.

Having thus stated the principal parts of the overall problem, it is perhaps best to examine and explain the means to its solution in the reverse order. Therefore, it is deemed of fundamental import to realize that the EDF Generator's housing configuration as proposed will demonstrate a significant static capacitance, despite its unusual geometry, given the fact that the negative and positive sections of its housing are to have essentially equal surface areas. This simple precept ensures two very important consequences: (i) that the voltage across the housing will attain a value representing equal negative and positive potentials, in reliance upon the absence of a fixed ground reference, given induced surface charge densities which are sufficient to cause a continuous 'leakage' current between the opposite-polarity sections thereof; and (ii) that a leakage current between the opposite-polarity housing sections will tend to assume a relatively uniform cross section and current density in response to a calculable level of non-uniform surface charge density between non-parallel but equal-area 'plates'. Thus, it is possible to determine using traditional formulas just how much electron charge must be stripped from the positive housing sections, and stored in ballast capacitors, in order to establish a positive surface charge density which corresponds to the desired housing or hull potential difference.

Next, the Generator must, and will, be able to charge these positive housing sections (by stripping them of native electrons) far faster than the positive surface charge so-created is neutralized by corona or arc discharge from the negative housing sections. Because these discharge current electrons will ultimately attain a transit time approaching zero, this in turn means that it must be "easier" per volt to strip negative charge from the positive sections—using a field induction system which is comprised of "primary electrode arrays"—than it is to deliver this charge to the negative sections across a vacuum chamber using relatively massive rotor-mounted field emitters.

It can be shown that the potential difference across said primary arrays will tend to be equal and opposite in operation as proposed to that across such vacuum chamber, in part due to the Faraday shielding principle, and the negative 'plate' voltage of the capacitive housing will naturally tend toward numeric equality with that of the positive sections by electrostatic induction. Therefore, the stationary electron-emitting primary cathodes which link the positive sections of the housing to the rotor are to have a much lower surface work function (and higher emissivity) than the rotating electron-emitting elements which link the rotor with the negative housing sections. This criterion is achieved by thorium-impregnating these cathodes (to stimulate their emissivity) and by not so-impregnating the rotor field emitters.

The relationship just described establishes what electrical engineers refer to as an 'instantaneous charge differential' with respect to these two sets of components as operated at equal but opposite charging voltages or potential differences, and the ratio of their respective electron emissivities will thus define the maximum ratio by which the said secondary field voltage (across the exterior housing) can be expected to exceed that in evidence across the stationary internal primary electrode arrays. This ratio will hereinafter be referred to as the "primary voltage expansion ratio", and the instantaneous charge differential concept as well as its relationship to the said voltage expansion ratio will be explained in further detail in a pertinent section to follow. It is important to note at this point, however, that for this voltage expansion to be realized it must and will be supported by ballast capacitors having suitable total capacitance as outlined hereinabove, which provide a corresponding force of electrostatic induction.

Finally, a method of impressing as large a primary DC voltage across the rotor as may practicably be developed in the device's principal generating system must and will be employed, without such "Primary Power System" circuit conducting more than a minute percentage of the discharge current thereby initiated. To do so, and in a manner that requires no brushes, vacuum tube cutoff bias technique is utilized in the present invention: whereby such primary voltage is electrostatically induced upon the rotor between two sets of capacitive electrode ring pairs (in vacuo), yet the current flow across these rotor-mounted induction electrode arrays—and through the principal generating circuit itself—is absolutely minimized by the application of a control grid voltage which is substantially negative with respect to the corresponding cathodes.

At a certain relative potential, the electrostatic force exerted by such control grid(s) on rotor circuit cathode-emitted electrons is sufficient to all but completely cut off the flow of DC current across these rotor induction arrays; yet the primary DC voltage is transmitted to the large conductive rotor segments such that a comparatively huge charging current between the opposite-polarity housing sections is initiated and maintained (across the rotor)—which subsequently expresses the expanded secondary voltage as a discharge current about the housing. Thus, an analogy could be made in describing the EDF Generator as a DC equivalent for boosting voltage by electrostatic induction means to the AC transformer (for boosting AC voltage by electromagnetic induction means).

General Form of the Invention

Accordingly, referring to FIGS. **1** through **4**, the present invention constitutes a novel form of brushless permanent magnet electrical generator for producing a useful ultra-high-energy external electrodynamic field, or continuous quasi-coherent DC corona or arc discharge of essentially uniform current density, and as such the Generator fundamentally comprises and incorporates:

[A] a flat conductive rotor **(6)** principally formed of a large number of evenly-spaced individual metallic conductors or segments **(14)** and an equal number of conductive electron-emitting elements or field emitters **(17)**, radially arranged in an annular flywheel configuration rather like the rotatable deck of a carousel, which are all connected in parallel electrically by the attachment of a thin metallic electrode ring **(68)** about the positive inner circumference of each of the two major flat faces of the annular flywheel thus formed and by the attachment of another such ring **(22)** about the negative outer circumference of each such flat rotor face;

[B] a principal electrical generating system mounted on said rotor and used to create a very high primary DC voltage within a plurality of rotating toroidal field coils **(35)(40)(56)** by electromagnetic induction with an equal number of circular arrays of ring-shaped stationary permanent magnets **(34)(39)(55)** (which are positioned adjacent and concentric thereto), such that an internal primary discharge current may thereby be initiated between the rotor **(6)** and a housing electron "Emitter Ring" **(47)**—across an evacuated space charge chamber **(11)**—which is orders of magnitude larger than the operating current through said field coils;

[C] a conductive exterior housing **(1)** divided into a plurality of distinct electrical sections—negative **(3)**, neutral **(4)**, and positive **(5)**—which may be charged from the internal rotor by electrostatic induction and/or by such primary discharge current (as the case may be) and within an evacuated induction compartment **(12)** of which said rotor **(6)** as an assembly, including the said generating coils attached thereto, is rotatably mounted;

[D] a plurality of individual systems of stationary plane-parallel electrode elements **(64)**-**(67)** which electrically link the polarized rotor **(6)** by electrostatic induction about its positive inner circumference (within said evacuated compartment **(12)**) to two axially-central electron collector housing sections **(5)**, and a plurality of similar individual systems of rotor-mounted plane-parallel electrode rings and/or radial electrode elements which may be utilized to connect a principal portion of the individual DC voltage outputs of said rotating field coils in opposed series-parallel across such composite rotor assembly;

[E] this composite rotor assembly **(6)** as a means for the transportation and capacitive storage of electronic charge between opposite-polarity sections of the housing **(5)** and **(3)** (which are proximal to the inner and outer rotor circumferences respectively), such that a vastly higher secondary voltage is caused to appear across interposed neutral housing sections **(4)**, resulting in a continuous ultra-high-energy DC discharge current about the housing externally which is also conducted internally across the rotor segments **(14)** that extend across the neutral housing portion and between the positive sections;

[F] a large number of stationary electromagnetic armatures **(37)** in a uniform circular arrangement located within the neutral electrical region **(4)** of the housing which are to be operated principally on variable DC voltage, but which may also accept limited pulsed unidirectional or alternating voltage input, and which are employed as a group to impart an orderly rotational magnetic vector moment to the external discharge current that acts to inhibit the formation of concentrated arc streamer or channel phenomena and thus to assist in rendering such current quasi-coherent in nature; and

[G] the said principal electrical generating or Primary Power System as operated and used in such a manner that a minor pulsed unidirectional or alternating voltage is superimposed upon the high-energy DC rotor current, and the external discharge current therefore acquires a discrete AC power factor, for purposes of modulating the electrodynamic characteristics of the external field during normal operation and/or for purposes of exploring the invention's potential signal communications capability in relation to intriguing new theories in high-energy quantum physics and relativity.

In a "single-stage" rotor electrical circuit embodiment of the Generator, as intended for use in an electrical and/or thermal energy output application, any such minor AC voltage present in said rotor circuit receives only a modest and single level of amplification—in a manner generally associated with a single multi-electrode vacuum tube. In a "three-stage" rotor electrical circuit embodiment of the Generator, as intended for use in a propulsive and/or communications application, any such minor AC voltage may receive a substantial and multiple level of amplification—in a manner generally associated with a multiple number of sequentially-coupled multi-electrode vacuum tubes.

FIG. **1** shows a side sectional view of the general form of the invention, as if cut through the central axis of the circular structure. Only one side of the tapered peripheral portion of the structure is shown, the overall Generator being symmetrical about the axis along which the section shown is cut. FIG. **2** shows a top sectional view of the invention, as if cut through just above the rotor assembly at the "equator", with only about a 30° radial sector of the circular structure shown (the remaining 330° being identical around the Generator). In each case, identical reference numbers are used to indicate identical elements. FIGS. **3** and **4** are side and top or bottom views of the Generator (respectively), showing the outward appearance of the invention and the external field it produces. The following discussion of the general form of the invention will be in reference to FIGS. **1** to **4**, unless otherwise noted.

The entire apparatus of the invention is contained within a circular conductive exterior hull or housing (**1**), which is tapered from its vertical centerline ("axis") to its rim (or "equator"). The overall housing is radially symmetrical around the axis, and bilaterally symmetrical about the equator. This housing's two central one-base spheric-zone-shaped sections (**5**) will be positively charged, its two intermediary conical ring-shaped sections (**4**) will be electrically neutral, and the convergent outermost or peripheral conical ring-shaped sections (**3**) will be negatively charged. Preferably, the surface areas of the outer (**3**) and inner (**5**) housing sections are equal, and specific details of the housing or hull design are given in a section to follow.

Each of the housing's two symmetrical positive sections (**5**), or hull positive "zones", should be divided into equal-area radial sectors (**42**) which are equal in number to the number of like sections into which the device's superstructure is divided. Thus, in the preferred embodiment of the Generator these two positive zones (**5**) will each be divided into 36 such sectors, and isolated top and bottom, yielding a total of 72 individually controllable discharge collector areas. The positive zone sectors (**42**) jointly comprise what may hereinafter be referred to as the device's "Field Hub", and should be composed of a conductive refractory metal or high-temperature structural alloy such as #310 stainless steel. These positive sectors are insulated from sector-to-sector by insulating "partitions" (**43**) of a suitable ceramic material (such as Cordierite), and must be supported by an insulating understructure (not shown).

At least one hull "neutral ring" surface layer (**46**) is also to be affixed on a housing frame and composed of one or more insulating ceramic material/s (such as Cordierite, Zirconia, and/or Kézite, as necessary). Said neutral hull layer/s may be applied in the form of plates and/or tiles to nonmetallic "deck" sheeting (not shown separately) supported by superstructure beams (**153**). This decking and its supporting beams are both preferably composed of a carbon composite material, and will thus be fairly conductive. Therefore, both must be isolated from the positive Field Hub by a dielectric buffer material (**45**) and from the hull's Emitter Ring (**47**), comprised of two negative housing sections (**3**), by an auxiliary dielectric buffer (**50**). The Generator's neutral housing section superstructure is shown in detail in FIG. **14**, as is a piping system of coolant tubes or "thermal conduits" (**74**)-(**79**) for maintaining the hull neutral sections (**4**) and superstructure—as well as the enclosed Primary Power System—at an acceptable operating temperature. Such an intercooler system would utilize a cryogenic coolant such as liquid air or nitrogen when the EDF Generator is operated in the atmosphere for purposes of thermal energy output, and will be explained in greater detail in a section to follow.

The negative hull plates which form each section (**3**) of the Generator's peripheral Emitter Ring (**47**), to which the rotor (**6**) distributes the discharge current forming the exterior field envelope, should be composed of high-purity aluminum which is copper-clad and then nickel-plated to prevent erosion at high field voltage and current levels. The abutting edges of all of the adjacent such plates must be welded one to another, and affixed to any necessary neutral hull section hardware, such that a hard vacuum may be drawn within the peripheral chamber (**11**) thereby formed. The induction compartment (**12**) enclosed between the two neutral hull sections (**4**) is contiguous with the peripheral or space charge chamber (**11**), and thus is equally evacuated.

A pressurizable central chamber (**2**) within the housing (**1**), located between the two hull positive zones (**5**) thereof, constitutes the location of the Generator's control cabin and of its payload area when used as an aerospace vessel. The interior (**11**)-(**12**) of the negative (**3**) and neutral (**4**) housing sections (respectively) is separately sealed, and must be as highly evacuated as possible to promote the efficiency of the various rotating and stationary electrode arrays and to prevent their destructive failure. The main vacuum seals (**155**) and the rotor's vacuum chamber charge buffers (**156**) are shown in FIG. **1**.

Inside the evacuated induction compartment (**12**) of the housing (**1**), and surrounding the central chamber (**2**), is located the composite rotor assembly (**6**) in the form of a flat annular flywheel. In the preferred embodiment, 180 individual conductors or segments (**14**) of very high collective ampacity are employed in the rotor, as will be seen in the discussion of FIG. **5** below, as are a like number of inter-segment insulators or segment "separators" (**16**). The preferred material for the rotor segments is high-purity copper which is certified oxygen-free, to resist the production of free oxygen radicals at high operating temperatures which would degrade the performance of the various electrode arrays and cause an undesirable increase in vacuum chamber (**11**) temperature.

The inner ends of the rotor segments **(14)** and segment separators **(16)** are mechanically connected via a retaining and lockdown section thereof to one or more ring gear(s) **(8)**, which run/s entirely around the central aperture of the rotor assembly. Suitable insulating attachment hardware **(7)**, as described in detail in the following section (and shown in FIG. **15**), will be used to effect a rigid mechanical union of the segments and separators to the said ring gear(s) and associated drive means such that the gear(s) and all conductive components of such drive means are electrically isolated from the energized segments. The ring gear(s) **(8)** may be engaged by one or more pinion or drive gear/s **(162)** powered by a motor **(9)** or plurality thereof, which may be of various types (such as electrical or hydraulic), causing the rotor to rotate around the central chamber **(2)**. The rotor is supported near its outer edge by two ballraces **(25)** mounted to the housing superstructure and to the rotor, and having bearing balls **(26)**, allowing it to spin freely.

As can be seen in either FIG. **1** or **7**, the rotor segments **(14)** are electrically connected together at each end by continuous inner **(68)** and outer **(22)** electrode rings, and also at intermediate points by induction cathode rings **(28)** and induction anode rings **(61)**. The inner electrode rings **(68)** constitute the primary induction ring array anode elements, as will be described below. The outer electrode rings **(22)** are each covered by a dielectric layer **(23)** which is in turn covered by an anode ring **(24)**, and each combination of rings **(22)** and **(24)** separated by dielectric layer **(23)** forms a housing charge ballast capacitor whose function will be explained in detail below. Induction anode rings **(29)** and cathode rings **(59)** are used to electrostatically induce corresponding negative and positive voltages on rotor-mounted rings **(28)** and **(61)**, respectively, thereby electrically energizing the rotor segments as will also be described in further detail below.

FIG. **5** shows details of the basic rotor construction. Wedge-shaped ceramic pieces or segment separators **(16)** equal in number to the conductive rotor segments **(14)** are to be positioned uniformly between the individual segments, as are the rotor's peripheral field emitters **(17)**. To form the base rotor construction, the segments, separators, and field emitters must be equal in number, preferably using 180 of each. The said segment separators are actually more structural than insulative in purpose, and not only stiffen the entire rotor to prevent its distortion at elevated temperature and rotational speed but also provide a nonconductive base for mounting all other rotor assembly components. Each of the ceramic separators **(16)** is bonded to its adjacent segments **(14)** by refractory adhesive layers **(15)**.

Two important criteria for selecting a ceramic segment separator material are that the compound chosen be readily machinable like steel and provide very high physical strength yet require no firing. These parameters limit the choice made almost to a solitary substance: magnesium silicate ($MgSiO_3$). When suitably formed, this material can be milled, drilled, and tapped like steel (using low-speed tungsten carbide tools) and approaches the physical strength of dry-pressed porcelain. It is ready for use directly after shaping and requires no kiln hardening.

This latter characteristic is very important as it is extremely difficult to produce fired ceramic pieces with the exacting dimensional control necessary in this case, and it eliminates the tendency of thicker fired pieces to develop micro-cracks which may potentially cause destructive failure of the component under high centrifugal loading. Surfacing can be done with abrasive paper or grinding wheels, and it is very important that this material be machined or worked perfectly dry as it is fairly porous and any lubricant used tends to significantly alter its dielectric properties. Magnesium silicate of this type is available from CeramTec NA as CeramTec designation AlSiMag 222, and will withstand constant working temperature of over 1300° C.

The insulating separators **(16)** are shorter than the conductive segments **(14)**, allowing for the field emitters **(17)** to be fastened outboard of the separators. The field emitters are thus located around the outer circumference of the rotor assembly and should be formed of a sintered refractory composite, such as tungsten-copper, selected for its relative surface work function and superior arc-erosion characteristics. As can be seen in FIGS. **5**, **6***a*, and **6***b*, the field emitters are fastened mechanically and electrically to the conductive segments **(14)** through locator pins **(19)**, which fit into mating holes recessed into the ends of the segments. These emitters **(17)** taper to a somewhat rounded edge at the tips **(20)** which form the outermost periphery of the rotor (as an assembly), and are squared-off at their bases **(18)** to fit tightly against the insulating separators **(16)**. From the top view of FIG. **6***b*, it can be seen that the field emitters follow the tapering of the segment separators **(16)**, increasing slightly in width from their bases **(18)** to the tips **(20)**. The conductive segments **(14)**, in contrast, are of constant width over their length.

Overall Description of the Primary Power System Structure

Rotor Drive and Positioning Assemblies

Referring now to FIG. **15**, a split-frame centrifuge-style carrier assembly reminiscent of the rotatable mounting arrangement for the turntable of a carousel may be employed to firmly suspend, center, and rotatably mount the composite rotor assembly about its inner circumference within the evacuated induction compartment of the housing, and such carrier assembly should preferably be constructed of a weldable nonmagnetic metal alloy which is suitably insulated from the electrically energized rotor assembly components.

The rotatable carrier assembly may be mechanically-affixed and/or welded to one or more large metal ring gear(s) which may be comprised of like sections **(163)** to facilitate construction and which in turn may be supported, driven, and dynamically braked by one or more low-voltage DC motor-generator unit(s) **(161)** (similar to those used in large electric trains) and drive pinion gear/s **(162)**. Such ring gear(s) may additionally be supported by pinion gear and pillow block bearing assemblies (if necessary—not shown in FIG. **15**). Such drive motor-generators(s) **(161)** and said pillow block bearing assemblies (if any) must be separately and nonrotatably affixed to the relatively stationary housing superstructure.

Referring still to FIG. **15**, it can be seen that in the preferred embodiment the inner end of each rotor segment **(14)** and segment separator **(16)** is flared outward into a retaining "fantail" **(160)**. These retaining fantails are preferably enclosed and prepared for connection into a driven assembly using a system of: [i] insulating inner **(171)** and outer **(173)** ceramic thermal spacers; [ii] axial **(172)** and radial **(174)** nylon (or equivalent) load buffers; and [iii] insulating displacement **(170)** and lockdown **(175)** ceramic bushings. The retaining fantails are then secured into a driven assembly by two identical carrier half-structures comprising frame struts and members **(168)** and retaining ring sections **(167)**, using: [i] mounting ring sections **(164)** each having two ring gear sections **(163)** attached thereto;

[ii] a corresponding number of mounting ring spacers **(165)** each having a very stiff metallic flexor plate **(166)** attached thereto; and [iii] nylon (or equivalent) torsion buffers **(169)**. The drive pinion gears **(162)** allow a preferred total of 32 like and evenly-spaced DC drive motor-generator units **(161)**, which are mounted to the Generator's ground frame **(10)**, to apply input rotational torque (or counter-rotational braking torque) to the composite rotor assembly by engaging a total of four ring gears (comprised of sections **(163)**) which are mounted to the rotor's described drive carrier assembly.

Rolling bearing assemblies should be employed to further center and stabilize the composite rotor about its outer periphery within the evacuated induction compartment of the housing, as a readily feasible mechanical alternative for that purpose to 'zero-friction' electromagnetic positioning systems (of the type normally associated with mag-lev train technology) which are of a much greater complexity.

Such rolling bearing assemblies may comprise two complimentary pairs of stationary and rotating circular-groove raceways **(25)** about the outer periphery of the composite rotor assembly, as well as an appropriate number of bearing balls **(26)** including uniform ball separation means, and these bearing assemblies may be composed of nonmagnetic stainless steel and/or a specialized ball bearing ceramic (such as silicon nitride (Si_3N_4)) which may run unlubricated in a very hot electrically-charged environment

The Rotor and Associated Elements

There are a number of electrical structures attached to the base rotor construction, which rotate with the segments, separators, and field emitters as a composite assembly. It will be understood that the structures mounted on the rotor and those interacting with the rotor are symmetrical above and below the incorporating device's equator, as well as either continuous or repeating in placement around the device in the plane of the rotor segments, so that a reference to an element at one point above the rotor also applies to the similar elements below and around the rotor.

Referring to FIGS. **1** and **2**, and in the preferred three-stage embodiment of the invention suitable for an aerospace vessel (as shown schematically in FIG. **9**), the following elements of the principal electrical generating means are mounted upon the annular rotor. Generally starting at the outer circumference of the rotor and working inward, and, for each structure, starting at the rotor and working upward, these are:

[a] the ballast capacitor, comprising a negative ring element **(22)** attached directly to the rotor, a dielectric layer **(23)** affixed thereto, and a positive ring element **(24)**;

[b] the bearing ballrace **(25)**, within which the bearing balls **(26)** roll between two identical raceway halves—one on the rotor, and one mounted to the vessel's stationary structural frame. The ballrace **(25)** and bearing balls **(26)** support the rotor peripherally and are structurally mechanical, not taking part in the electrical operation of the invention;

[c] the voltage-induction diode ring array, in the form of a cathode ring **(28)** which is mounted on and electrically connected to the rotor segments **(14)** and an anode ring **(29)** which is affixed plane-parallel to the cathode ring **(28)** and insulated from it by a suitable supporting structure using insulating posts, pins, or brackets in or on the segment separators **(16)**. Although the resulting combination could be thought of as a 2-element or diode vacuum-tube construct and is similarly named, it is more accurately a "cold" parallel plate capacitor of high value and appreciable AC conductance which preferably exhibits a modest circuit DC leakage current;

[d] the outer field coil **(35)**, helically wound and supported upon a continuous coil support or structural core which is preferably nonferromagnetic (not shown). There are preferably two windings on the core which together comprise such coil, as will be discussed in the circuit section below and as is shown in FIG. **9**—the major portion or field winding **(81)**, and a minor portion or bias winding **(82)**;

[e] the outer voltage-transfer triode ring array, in the form of cathode **(30)**, control grid element **(31)**, and anode **(32)** rings or elements, affixed plane-parallel or radially-concentric above the rotor segments **(14)** and insulated from them and from each other by similar such supporting structure (using insulating posts, pins, or brackets) in or on the segment separators **(16)**. This combination also comprises a triode vacuum-tube construct with controllably variable AC conductance;

[f] the center field coil **(40)**, helically wound and supported upon a continuous coil support or core of the type described. There are again two windings on the core, as will be discussed in the circuit section below (and as is shown in FIG. **9**)—the major portion or field winding **(83)**, and a minor portion or bias winding **(84)**;

[g] the inner voltage-transfer triode ring array, in the form of cathode **(51)**, control grid element **(52)**, and anode **(53)** rings or elements, affixed plane-parallel or radially concentric above the segments and insulated from them and from each other by the described supporting structure(s). This combination again comprises a triode vacuum-tube construct with controllably variable AC conductance;

[h] the inner field coil **(56)**, helically wound and supported upon such continuous coil support or core. There are once again two windings on the core, as will be discussed in the circuit section below and as is shown in FIG. **9**—the major portion or field winding **(85)**, and a minor portion or bias winding **(86)**;

[i] the voltage-induction triode ring array, in the form of an anode ring **(61)** which is mounted on and electrically connected to the rotor segments **(14)** as well as a cathode ring **(59)** and control grid element ring **(60)** affixed plane-parallel to the anode ring **(61)** and insulated from it and from each other by the said supporting structure(s). The resulting combination comprises a triode vacuum-tube construct with controllably variable AC conductance;

[j] the primary induction anode ring **(68)**, which is also mounted on and electrically connected to the rotor segments **(14)** and which provides the positive element of the stationary plane-parallel electrode systems used to induce the exterior field; and

[k] the ring gear(s) **(8)** or other appropriate means for allowing a drive mechanism to spin the rotor.

Field Coils

Referring now to FIG. **1**, the field coils **(35)(40)(56)** which comprise the major rotating portion of the primary DC voltage generating means are formed of conductive insulated magnet wire multi-layer toroidally wound upon cores (not shown) whose permeability should generally be minimized, and main portions of the DC voltage outputs of the three field coils comprising each of two like groups thereof—one above the plane of the rotor and one below—are sequentially connected via a plurality of individual systems of rotating plane-parallel electrode rings (or alternatively, modular radial electrode element systems or tubes) in the preferred three-stage rotor emodiment of the invention shown. The number of wire turns comprising each such field coil should in general be maximized (in a practical manner), in either the single- or three-stage embodiments.

As is depicted more clearly in FIG. **8**, each such field coil may be compound wound or comprise two or more independent voltage generating portions, with either all or the major portion of each coil constituting the field winding **(81)(83)** or **(85)** or principal DC voltage generating portion of that coil as a whole, and the principal or field winding portions of the field coils comprising either of the said two like groups thereof may be hooked in series one to another by direct physical connection in the single-stage rotor embodiment of the Generator shown. Each of the two like sets of three such field windings thus formed are then connected on their respective sides of the rotor between the plate resistor **(91)** of the diode array anode ring **(29)** and the inner triode array cathode ring **(59)**, while observing appropriate conventions of polarity assignment.

The remaining minor portion (if any) of each field coil may comprise one or more independent bias and/or control winding(s), or secondary DC voltage generating portion(s) of that coil as a whole, with the largest such secondary winding if any constituting the control grid bias winding of that coil. As shown in FIGS. **8** and **9**, one such bias winding **(82)(84)** or **(86)** is provided for each field coil and is positioned in the schematic(s) immediately thereabove. In the single-stage rotor embodiment of FIG. **8**, these bias windings are also directly connected in series one to another between the inner triode array cathode ring **(59)** and the corresponding grid resistor **(90)** of that same array's control grid **(60)**, with the negative end of each bias winding toward the said grid resistor.

In the three-stage embodiment of FIG. **9**, the field winding of each said field coil is connected in series between the plate resistor **(92)** attached to the anode ring of the outerlying electrode system which is adjacent to that coil and the cathode ring of the innerlying electrode system adjacent thereto, while observing appropriate conventions of polarity assignment. The bias winding (if any) of each said field coil is then connected between the cathode element **(30)(51)** or **(59)** of such innerlying electrode system and the grid resistor **(90)** attached to the negative control grid **(31)(52)** or **(60)** (respectively) of that same electrode system, while observing appropriate conventions of polarity assignment, such that said bias winding is in parallel with the cathode ring or element of that electrode system and may therefore provide an AC and/or DC voltage bias with respect thereto in the manner associated with electron vacuum tubes.

Rotating Electrode Arrays

Electron vacuum tube design, construction, and operating methods are employed to impress the principal portion of each field coil group's combined series DC voltage output across the Generator's rotor in an opposed parallel arrangement (comprising two like field winding circuits or rotor subcircuits), such that the rotor segments become electrically polarized, although the rotor is not considered uniformly energized electrically until a finite DC Primary Power System circuit current is established against the series capacitance and high bias voltage of the several rotor-mounted electrode systems.

Two special pairs of highly-capacitive electrode rings **(22)(24)** seperated by a dielectric media **(23)** are affixed upon the polarized rotor **(6)** near the outer negative periphery thereof, and supplied with a portion of the Generator's positive primary DC output voltage, in order to accomplish the storage of a predetermined amount of ballast electron charge which is sufficient to enable a desired much-higher secondary voltage to appear across the neutral sections of the housing (given the static capacitance and surface charge density characteristics of the housing as a whole in operation).

Electron vacuum tube design, construction, and operating methods may be employed to induce and/or modulate a substantial DC bias upon the plate currents of any or all of the various rotating plane-parallel and/or radial electrode systems or "arrays" employed in the present invention. Each rotating three-electrode system or triode array so-employed should be constructed in such a manner that it exhibits a minimum design amplification factor (μ) equal to 4.0, with reference to any AC voltage or signal present in its electrical circuit and as a function of its engineered relative electrode spacing(s).

Electron vacuum tube design, construction, and operating methods may furthermore be employed to induce, modulate, and/or amplify a minor pulsed unidirectional or alternating voltage upon the high-energy DC rotor current, and therefore on the field envelope current as well, as a means by which the electrodynamic field produced by the Generator may potentially be used (as a form of antenna) for purposes of transmitting and receiving a variable electromagnetic and/or gravimetric resonant frequency signal either to or from itself or a separate similar device (as the case may be).

In light of this teaching, it is contemplated that one possible type of communications signal which might be investigated employing the EDF Generator would couple the electric, magnetic, and gravitic forces to utilize the wave mechanics of quantum potential vacuum fluctuations ($\Delta E_q/\Delta t$), which are believed by some theoretical physicists to propagate at c^2 in order to explain the uniform operation of gravity and entropy in two space-time continuums which are coincident but completely out-of-phase. Should this quantum gravity theory prove correct, the transmission delay of a gravimetrically-coupled EM signal at this wave speed would be only 0.10525 seconds per light-year.

In any event, it should be noted that the rotor interstage coupling transformers **(89)** which are depicted in FIGS. **1** and **9** (but are not shown in FIG. **2**) would be employed only in a preferred three-stage rotor embodiment, and mounted (one each) immediately above the rotating diode and transfer triode arrays (**(28)**–**(29)**, **(30)**–**(32)**, and **(51)**–**(53)** respectively). These transformers may preferentially take the form of continuous toroid coils on either powdered iron or nonmagnetic cores (like the field coils), having single-layer primary and secondary windings, or of one or more pairs of balanced toroidal arc-section coils and cores of similar composition which are uniformly distributed above said rotating arrays. Each transformer **(89)** uses two DC blocking capacitors **(88)**.

It may be appreciated from FIGS. **8** and **9** that each bias winding or series group thereof, its associated triode array, and a corresponding output plate resistor **(91)** or **(92)** together comprise one stage of amplification of any AC signal voltage which may be present across that triode's grid resistor **(90)** as an input. In this way, electromagnetic waves of one form or another from outside the external field created by the Generator may be both detected, from their influence on the waveform(s) of the field envelope current, and amplified from within the device (when used as an aerospace vessel).

The Interaction of the Rotor and Frame-mounted Elements

Magnetic Rings

A plurality of circular arrays of stationary permanent magnets essentially comprise the stationary portion of the primary DC voltage generating means, and each such magnet is preferably to be formed as a C-shaped ring composed of axially round stock of a metallic ferromagnetic substance which exhibits a very high residual induction (such as the alloy family commonly known as Alnico). The number of individual magnetic rings comprising each such array should in general be maximized (in a practical manner). Preferably, each magnet array has a rigid and annular nonmagnetic round core (not shown) which is roughly equal in minor diameter to the inside diameter of the constitutent rings, such that the rings may be easily strung thereupon rather like beads on an abacus wire.

Referring once again to FIGS. **1** and **2**, each of the several magnetic ring arrays **(34)(39)(55)** may be installed in sections which are preferably equal in number to the number of like sections into which the incorporating device's superstructure is divided. The rotating field coil **(35)(40)** or **(56)** associated with each such circular array of individual magnetic rings must be respectively concentric and adjacent thereto, and is basically positioned on both the radial and axial centerlines of the flux gaps of that array's constituent rings. As power is generated thereby, these rings may need periodic replacing.

The maximum length of the flux gap of each such C-shaped magnetic ring must of course be just slightly larger than the minor outside diameter of its associated field coil, and should be roughly equal to or just slightly less than the formed inside diameter of that magnetic ring. In the preferred embodiment these ring flux gaps are oriented horizontally, although they may also be readily oriented vertically if so desired. Each magnetic ring may also be mounted within the superstructure of the housing such that a comparatively small thin axially-polarized wafer of a highly-coercive permanent ferromagnetic substance (like sintered Ferrite 5 ($BaO.6Fe_2O_3$)) is positioned in the hollow center of that ring magnet, with such wafer's magnetic poles facing the like poles thereof, so as to act as a leakage flux blocking or reducing mechanism with respect thereto. These optional components would be most easily installed by insertion into a corresponding slot cut into each of the described mounting cores, and are recommended for use in very large devices.

Also arranged concentrically around the evacuated induction compartment **(12)** of the Generator, but fixedly mounted within the vessel's structure instead of rotating with the rotor assembly, are the following elements of the principal electrical generating means. Again, they are discussed in order from the outer circumference of the vessel inward, as shown in FIGS. **1** and **2**, and a reference to an element at one point above the rotor also applies to the similar elements below and around the rotor:

[a] the stationary half of the ballrace **(25)**;

[b] the outer array **(34)** of stationary permanent magnets, which should be mounted on a nonmagnetic core and in the preferred embodiment incorporates a maximum of 900 such ring magnets equally spaced around the vessel superstructure;

[c] the circular group of stationary electromagnetic armatures **(37)** or "variable inductor" array, as used to impart a rotational magnetic force upon both the external discharge current and the internal rotor assembly (which is proximal thereto). Each such array preferably comprises a maximum of 180 such armatures, and is to be approximately centered within the neutral region of the housing so that its constituent armatures are axially parallel to the rotor's axis of rotation;

[d] the center array **(39)** of stationary permanent magnets, which should be mounted on a nonmagnetic core and in the preferred embodiment incorporates a maximum of 720 such ring magnets equally spaced around the vessel superstructure;

[e] the inner array of **(55)** stationary permanent magnets, which should be mounted on a nonmagnetic core and in the preferred embodiment incorporates a maximum of 576 such ring magnets equally spaced around the vessel superstructure; and

[f] the stationary anode ring **(58)**, located adjacent and plane-parallel to the rotating cathode **(59)** of the inner electrostatic-induction triode array, which develops an induced positive voltage that is made available to the control circuitry shown in FIG. **10** as described below.

FIGS. **8** and **9** show two embodiments of the Primary Power System. The "Field Induction System" or stationary plane-parallel electrode arrangement(s) employed to induce the exterior field is/are basically the same in either case, and the voltage control system of FIG. **10** may be used with either embodiment.

The simpler single-stage rotor system of FIG. **8** is primarily intended for applications wherein the EDF Generator would be employed to produce useful electricity or heat, the latter being extracted from a principal liquid hull coolant circulated through the primary thermal conduits **(48)** (shown in FIG. **1**) that encircle the field power resistors **(63)**. Two thermal exchange or main service manifolds (not shown) would provide ground-based support to such a Generator from an associated utility or physical plant, as well as a remarkable level of electrical or thermal output from such Generator thereto. Each such manifold would connect directly to a single circular center sector of one of the two positive housing zones. Each of these two center sectors **(44)**, one or either of which is depicted in FIG. **4**, must in this case be composed of a nonconductive material.

Description of the Primary Power System Circuit

Referring again to FIG. **8**, the respective outer, center, and inner field windings **(81)(83)(85)** are directly connected in series, and the voltage thus generated is applied to the inner induction triode array cathode ring **(59)** and (through plate resistor **(91)**) to the outer induction diode array anode ring **(29)** and to the ballast capacitor anode ring **(24)**. The respective outer, center, and inner bias windings **(82)(84)(86)** are also connected in series, and the voltage thus generated is applied between the inner induction triode array cathode ring **(59)** and (through grid resistor **(90)**) to the control grid **(60)** of that same array. This places said control grid in parallel with that cathode electrically.

The turns numbers of the bias windings **(82)(84)(86)** and the grid resistor **(90)** values are chosen to apply a bias voltage to control grid **(60)** such that the inner induction triode array is biased very nearly to current cutoff. This results in a very high voltage being induced on the rotor segments **(14)** between outer induction cathode ring **(28)** and inner induction anode ring **(61)**, but at a very low series field coil current due to the near-cutoff bias of the triode array. It is expected that the voltage on the rotor segments between these two rings, in a preferred embodiment using this arrangement, would be approximately 8,000 volts for a four-foot diameter prototype, or about 2,000 volts per foot of housing diameter. It is also anticipated that series field coil current could be limited to fractional amperage in smaller such devices and to single-digit amperage in the largest.

The preferred embodiment shown in FIG. **9** is a three-stage rotor system which is considered to be much more useful for applications wherein the EDF Generator may be employed as an aerospace vessel, in that it allows energy/wave-function signals from outside the vessel to be detected and amplified in the manner previously described.

In this embodiment, instead of the field windings **(81)(83)(85)** being directly connected together in series, intermediary voltage-transferring triode vacuum tube constructs are used to effect an indirect series connection between the said field windings. The voltage thus generated is again applied to the inner induction triode array cathode ring **(59)** and (through plate resistors **(92)**) to the outer induction diode array anode ring **(29)** and to the ballast capacitor anode ring **(24)**. The bias windings **(82)(84)(86)** are in this case used to provide an independent bias voltage to the control grids **(31)(52)(60)** of the outer, center, and inner rotating triode arrays, respectively, in parallel with their cathodes.

As in the single-stage rotor embodiment described above, the turns numbers of the bias windings **(82)(84)(86)** and the grid resistor **(90)** values are chosen to apply a bias voltage to control grids **(31)(52)** and **(60)** respectively such that each corresponding triode array is biased very nearly to current cutoff. This results in a very high voltage being induced on the rotor segments **(14)** between outer induction cathode ring **(28)** and inner induction anode ring **(61)**, again at a very low or fractional current due to the near-cutoff bias of the triode arrays. It is expected that the voltage on the rotor segments between these two rings, in a preferred embodiment using this arrangement, would be approximately 1,500 volts per foot of vessel diameter.

In general, the stationary anode rings **(58)** and all of the rotating electrode rings except the primary anode rings **(68)** should be composed of a nonmagnetic structural nickel alloy (such as Inconel 600). All rotating control grid element wires should be composed of an alloy such as that commonly known as nichrome. The rotating cathodes may be composed of modestly-thoriated tungsten, however, if and only as necessary to establish a small DC Primary Power System current if the observed ‘dark discharge’ current values are deemed insufficient, particularly in the three-stage embodiment wherein amplification of a minor AC signal voltage may be desired for communications purposes. The ceramic material Titania, with a nominal dielectric constant of 85, may preferably be used to form the posts, pins, or brackets which support the various rotating electrodes.

It is important to observe that both Primary Power System (rotor) field winding circuits as configured in both the single- and three-stage Generator embodiments will have an inherent AC series resonant frequency, and operation at that frequency will maximize the series field winding AC line currents (within given circuit resistor constraints) as well as the dependent stage plate voltage drops (across resistor(s) **(91)** or **(92)** respectively) which allow stage AC signal reproduction and amplification. As in traditional related practice, the series resonant frequency will be that at which the given circuit’s inductive and capacitive reactances are approximately equal and opposite. Thus, the series resonant frequency of either of said Generator rotor circuit embodiments will be largely contingent upon the total field winding inductance and coil core permeability, as the various rotor electrode arrays and the plate resistor(s) **(91)** or **(92)** must have quite specific fixed relative design values of capacitance and resistance (respectively)—as a function of total field winding voltage—for proper overall DC and AC circuit performance.

Each Primary Power System control grid winding subcircuit as configured in the three-stage rotor embodiment may then be readily tuned for stage AC parallel resonance at the said inherent series resonant frequency, thereby minimizing stage grid subcircuit line current while maximizing the stage grid voltage drop (across each grid resistor **(90)**) which determines the level of net stage AC signal voltage gain or amplification obtained. As in traditional related practice, the parallel resonant frequency will once again be that at which the given subcircuit’s inductive and capacitive reactances are approximately equal and opposite. This condition may be achieved by connecting an optional capacitor of suitable value if desired across the secondary winding (or control grid resistor side) of each rotor stage coupling transformer **(89)**, as indicated in FIG. **9**.

The Field Induction System

Before commencing a technical discussion of the stationary induction electrode systems which will be used to charge the positive housing sections, it will be necessary to further elucidate certain aspects of the Primary Power System electrical circuit so that a few of its key operational characteristics are more clearly understood.

Referring now to FIG. **7**, each of the Generator’s two identical groups of rotating field coils (one above and one below the rotor) are in essence connected in simple series across the rotor segments using pairs of capacitive electrode rings, thus forming a single series loop with respect to the rotor. The two such series loops are at the same time, however, connected in mutually-opposed electrostatic parallel with respect to the rotor segment circuit leg common to both such loops.

Based on applicable principles of proper circuit resolution and series DC capacitance, such an opposed series-parallel circuit may readily be configured such that equal and opposite-polarity voltages will result upon the rotor-mounted induction cathode **(28)** and anode **(61)** rings in the absence of a fixed ground reference. To this end, the rotor assembly itself is to remain completely without direct earth or chassis ground in operation, and a corresponding segment polarization will continue as the circuit is duly energized by a finite DC conduction current.

When a primary DC rotor current is also established to the housing’s Emitter Ring **(47)** and an external discharge current in turn occurs to the positive zone sectors **(42)**, the two systems of primary induction electrodes **(64)**–**(68)** also complete two simple series loops back to the rotor and the two such external field envelope subcircuits in parallel constitute the EDF Generator’s “Field Induction System” circuit. Thus, the total field discharge current equals the primary rotor current, and the rotor segment circuit leg is common to both of the major system circuits described.

As is apparent from the section on Method of Operation, it is imperative that the charged housing sections also have no direct earth or chassis ground, being instead of a ‘floating-ground’ reference with respect to the described primary rotor circuit. The cathodes **(64)** of these two primary electrode systems may have a chassis ground reference at a very high resistive bias (as indicated in FIG. **10**), across which any actual cathode charge imbalance or potential in operation may be detected and measured.

Therefore, in light of the Faraday shielding principle and the said equal and opposite voltages which occur at opposite ends of the rotor segments (**14**), the potential difference between the rotor field emitters (**17**) and the inner surfaces of the two negative hull sections (or Emitter Ring (**47**)) will tend to become equal (and opposite) in operation to that across the primary electrode systems. And thus, a suitably vast and engineered differential in the electron emissivity of the field emitters (**17**) and the said ground-referenced primary cathodes (**64**) will define the primary voltage expansion ratio which may be obtained (in the manner first outlined in the section on Method of Operation). This topic will be discussed in more detail in a section to follow.

Moreover, a basic field-envelope-circuit generator design is disclosed which thus far, in accordance with Faraday's and Lenz's Laws, can produce a very high DC primary voltage with minimized circuit AC current inductance losses. The Primary Power System's DC voltage is not readily available, however, for purposes of normally-usable electrical output. Brushless electrode rings (**58**) must once again be employed to extract any portion of this primary voltage for off-rotor use. The Primary Power System schematics shown in FIGS. **8** and **9** in conjunction with FIG. **10** best illustrate the use of these two additional stationary anode rings (**58**), one suspended adjacent to each inner rotor induction cathode ring (**59**), which allow a large electrostatically-induced positive DC voltage to be 'picked off'. This positive voltage may then be used as a source of low power within an external circuit, provided that circuit is negatively-grounded and only to the Generator's chassis, and is therefore really intended for auxiliary onboard usage only.

It can be seen thus far, and especially in FIG. **7**, that there are essentially two major systems involved in creating the voltages which must be applied to the positive and negative housing or hull sections in order to form the electrodynamic field that will surround the Generator in operation. These two systems are:

[A] the Primary Power System, which generates a large DC voltage by electromagnetic induction and then electrostatically induces a major portion of that voltage upon the rotor segments (**14**) between two pairs of induction anode (**61**) and cathode (**28**) rings which are conductively affixed thereto (about the inner and outer circumferences of the rotor assembly, respectively); and

[B] the Field Induction System, which utilizes the potential difference which then exists between the positive inner circumference of the rotor—and its two attached primary anode rings (**68**)—and two separate sets of highly electron-emissive stationary cathodes (**64**) which are maintained at chassis ground and which are circularly-arranged near and electrically-connected to the periphery of each central collector housing section (or "zone") so that they may jointly charge their respective zone sectors (**44**) and/or (**42**) positively by stripping them of native electrons.

Each of these primary electrode array cathodes (**64**) connect directly to a field power resistor (**63**), which in turn connects to one positive radial zone sector (**42**) of the central housing sections. Each such cathode (**64**) is also vertically aligned inboard with a group of three stationary plane-parallel grid electrode elements and one of the two innermost rotating induction anode rings (**68**) conductively-affixed to the rotor segments (**14**) (about their positive inner circumference). Each such group of three grid elements is comprised of a control grid (**65**), an accelerator grid (**66**), and a suppressor grid (**67**), with one such combined grid element assembly per radial zone sector (**42**).

Together with the said primary induction anode ring (**68**), each such 5-electrode group constitutes what is termed a "unit pentode array" ((**69**) as seen in FIG. **7**). The unit pentode arrays (**69**) and their associated frame-mounted power resistors are thus arranged in two ring-shaped sets, located between the rotor and the two positive housing sections, which are referred to throughout as the "primary induction ring rays".

The potential difference between the unit pentode array cathodes (**64**) and rotor primary anode rings (**68**) is then referred to as the "primary array voltage", as the unit pentode arrays of each primary induction ring array are connected in parallel electrically (between the field envelope and rotor), and the number of unit pentode arrays should be equal to the number of like sections into which the Generator's superstructure is divided.

Electron vacuum tube design, construction, and operating methods may be employed to induce and/or modulate a substantial DC bias upon the plate current(s) of the several unit pentode arrays, which electrically link the polarized rotor to the central housing sections, for purposes of rendering the external electrodynamic field variably non-isometric and therefore directionally propulsive. Each such "stationary" 5-electrode system or unit pentode array (**69**) (whose anode element actually rotates) should be constructed in such a manner that it exhibits a minimum design amplification factor (μ) equal to 12.0, again with reference to any AC voltage or signal present in the Field Induction System circuit and as a function of its engineered relative electrode spacing(s).

The rotating primary induction anode rings (**68**), stationary cathodes (**64**), and various grid element wires which are associated with the stationary multi-electrode systems should be composed of tungsten, due to the extremely heavy conduction current across these primary arrays. Ceramic posts, pins, or brackets composed of a specialized material such as Titania may once again be used to support all of the various stationary electrodes and grid elements, and each unit pentode array's grid wires should be vertically aligned (or mutually-shading) similarly to those in a "beam power pentode" vacuum tube.

As generally depicted in FIG. **1**, a primary thermal conduit (**48**) with an imbedded coolant-carrying core (**49**) encircles and mechanically supports each unit pentode array, its power resistor (**63**), and that resistor's enclosing dielectric buffer (**45**). These thermal conduits and dielectric buffers will be examined in further detail hereinafter.

The operation of the Field Induction System is regulated by a central Field Voltage Control System which is modular in nature, as illustrated in FIG. **10**, and which is therefore mechanically and electrically connected to the stationary cathodes and grid elements of each unit pentode array (**69**). In this regard, the auxiliary stationary anode rings (**58**) described above are to serve as the source of positive voltage applied to the accelerator grids (**66**) of said unit pentode arrays, and such Field Voltage Control System is designed to directly accept this voltage for that purpose. The Field Induction and Voltage Control Systems will be discussed further in sections to follow.

Zone Sector Construction

Referring now generally to FIGS. **3** and **4**, one major goal in dividing each of the two hull positive zones or sections (**5**) which jointly comprise the Generator's Field Hub into a large number of radial sectors **(42)** in electrical parallel is to limit the external field current reaching each particular power resistor and unit pentode array combination to a fairly uniform level, while somewhat reducing field current return eddy losses as well. However, each radial sector **(42)** is thereby also given the capability of effecting a local thrust production differential (in propulsive three-stage rotor devices), by virtue of its electrical isolation, as it is then possible to vary the local resistance presented to the electrodynamic field with respect to any given radial sector so that a substantial measure of navigational control in the z-axis may thus be achieved by varying the proportional field current conducted by that sector.

Because the inherent resistance of each such sector itself will be negligible even at an extremely elevated operating temperature, in part due to the very low temperature coefficient of resistivity of the proposed steel hull plate material, the voltage drop across the radial length of each sector will also be negligible. Thus, the determination of an appropriate sector thickness becomes an entirely structural consideration. A secondary design goal with respect to zone sectorization is therefore to select a uniform cross-sectional area for the radial zone sectors **(42)** which yields a Field Hub total conduction mass approaching that of the device's rotor segments **(14)**.

The hull plating which forms each positive zone (or section **(5)**) of the Generator's Field Hub should preferably be composed of stainless steel or a refractory metal, and divided into thirty-six equal-area radial zone sectors **(42)** of truncated wedge shape and a circular center zone sector **(44)**. As stated earlier, in the single-stage Generator embodiment these two center zone sectors must be of a nonconductive material or construction, and should also comprise approximately 4 to 5% of the total area of their respective zones to allow the connection of the exchange or main service manifolds described above. In the three-stage embodiment such center sectors **(44)** must instead be conductive as just prescribed for the radial zone sectors **(42)**, and should each comprise about 1% of the said zone area. Each such positive radial sector should be formed of a single piece of the hull plating material in such a manner that its cross-sectional area in the direction of its radial length is maintained at a very uniform value.

The uniform thickness of each center zone sector **(44)** should be equal to that of the radial sectors as measured at the minor (inner) arc width thereof, and the radial sectors must be insulated from the respective center sector where they meet that center sector's periphery (except in single-stage rotor devices). A suitable thin ceramic insulating partition **(43)** may therefore be installed between each adjacent pair of radial zone sectors **(42)**, and around each center sector **(44)** in three-stage devices, and the uniform flat thickness of these sector insulators should in general be minimized.

In three-stage devices, four radial sectors **(42)** of each positive housing zone which correspond to cardinal points separated by 90 degrees of the hull's circumference, and which divide the Field Hub into four equal quadrants (as indicated with phantom lines in FIG. **4**), should be connected in parallel with the center sector of that positive zone at the maximum positive voltage end of those four sectors' respective power resistors **(63)**.

The parallel resistance of the four conductors which therefore connect each center zone sector **(44)** (in three-stage devices) to its respective network of power resistors should then be such that that sector's per unit surface area level of field current conduction is slightly higher on the average than that of its adjacent radial sectors **(42)**, under uniform zero-signal field bias conditions as hereinafter described.

The four unit pentode arrays **(69)** of each primary induction ring array which correspond to the power resistors so-connected to that primary array's center zone sector **(44)** would generally not be utilized individually in the active field current bias modulation related to the production of local thrust differentials (in three-stage rotor devices), being instead reserved primarily for potential involvement in signal communications transmission and/or reception activities.

Power Resistors

As best seen in FIG. **1**, the cathode element **(64)** of each plane-parallel electrode system or unit pentode array which is used to link the rotor assembly with a positive housing collector section or radial zone sector **(42)** is electrically connected to such positive section or sector by means of a ceramic resistor block **(63)**, of a certain low dielectric constant and high volume resistivity, which becomes a poor but effective conductor at the elevated operating voltage and temperature of such housing collector sections and which ensures an adequate external field current voltage drop between such section or sector **(42)** and said cathode element **(64)** (which is chassis-grounded as shown in FIG. **10**).

Referring still to FIG. **1**, each said ceramic resistor block or "power resistor" **(63)** is to be entirely jacketed by a hollow ceramic thermal conduit **(48)**, including a dielectric buffer **(45)** composed of a ferroelectric material having a certain high dielectric constant and very high volume resistivity which does not become a conductor even at the elevated operating voltage and temperature levels of said power resistors. These thermal conduits **(48)** mechanically connect the power resistors to a flattened inner surface of such positive hull section or radial sector **(42)** and a form of very-high-temperature heat transfer fluid such as liquid sodium may be circulated therein so that the proper resistor temperature is maintained to ensure an optimum cathode potential with respect to ground.

A suitable manifolding, pump, and heat exchanger system (not shown) may be used to extract or transfer an extremely high level of recoverable thermal energy from the power resistors **(63)**, arising from the field current voltage drop associated therewith, by means of such primary heat transfer fluid circulated through the thermal conduit cores **(49)** in the manner normally associated with nuclear power plants. Similarly, the rate at which such primary hull coolant is circulated may be used to directly regulate and limit the external field current to a value within the safe ampacity of the rotor assembly.

It is important to point out that each of the Generator's power resistors **(63)** and its associated dielectric buffer **(45)** and primary thermal conduit **(48)** may be designed, as an assembly, to incorporate a fixed positive reduced-voltage tap (not shown), whereby a main source of onboard or output DC or AC electrical power may be provided either for or from the Generator within a limited-duty circuit which is grounded directly to the primary cathodes. Provisions should and may also be provided for biasing, damping, and/or inductively coupling any primary conduit connecting sections, to control or harness electromotive currents induced in the highly-conductive primary coolant. Onboard electrical resistive heating means should and may be incorporated for purposes of pre-heating the primary coolant and, in turn, the power resistors and dielectric buffers.

Referring briefly to FIG. **11**, it is also important to note that the Generator's power resistors must be in a conductive state before the housing potential difference reaches breakdown field intensity, as figured along a minimum semicircular arc trajectory (**142**) across the hull neutral sections (**4**), and the flow of field discharge current commences. Therefore, the actual maximum dielectric constant of the resistor material used in very-high-voltage single-stage rotor devices, and in ultra-high-voltage three-stage rotor devices as well, may be calculated at about k=11.

As is best seen in FIG. **14**, the recovery thermal conduits (**79**) of the secondary hull coolant system which assist in supporting the entire tapered peripheral portion of the housing must pass inboard between the power resistor assemblies referred to above, and therefore the resistors (**63**) must assume a complex tapered shape to provide the necessary clearance. As generally depicted in FIGS. **1** and **2**, they will have a trapezoidal cross section at the 'top' or outward-facing end (with radial length somewhat greater than minor circumferential width) and a rectangular cross section at the bottom or cathode end (with circumferential width much greater than radial length). In this regard, their engineered shape should still maintain a uniform conduction cross section exactly analogous to that of the preferred positive zone sector construction described above.

Dielectric Buffers

Referring once again to FIG. **1**, it is imperative to electrically insulate the power resistors (**63**)—due to the extreme voltage drop that will occur across them in operation—to prevent direct discharge to other nearby structural components from taking place. However, not only is this a problem due to the thickness of insulation which would be required at any 'ordinary' dielectric value, but the power resistors' projected operating temperature range of 600 to 700+° C. completely precludes the use of almost all known dielectrics. Fortunately, the search for high-temperature dielectrics for use in aerospace electronics applications (particularly multi-layer capacitors) has resulted in the development of a small number of exotic materials potentially suitable for use as a space charge buffer (**156**) or dielectric encapsulant (**45**) for the Generator's power resistors (**63**).

In the application contemplated herein, the volume resistivity of the resistor encapsulant is somewhat secondary to its dielectric constant k at high temperature, because no opposing contact voltage is present (as in a capacitor) to encourage circuit conduction losses within the material. Therefore, the prime consideration is the dielectric's ability to exhibit a very large k value at over 600° C. This places the material squarely in the realm of a very small select group of Class III high transition temperature ferroelectrics. These materials, such as Tantalum-modified Lanthanum Titanate and Lead Ytterbium Tantalate, do not begin to demonstrate significant dielectric strength at less than 300 to 450° C. One such compound, Sodium Bismuth Titanate ($Na_{0.5}Bi_{4.5}Ti_4O_{15}$), actually exhibits an amazing peak k value of about 3100 near the middle of the said power resistor operating temperature range: at 655° C. (1,202° F.). This would seem to indicate that this compound is eminently suited for use in forming the necessary rotor space charge buffer pieces (**156**) and/or resistor dielectric buffers (**45**) in this application.

It should be stressed that great care in quality control must be exercised in formulating and sintering this or any similar such special compound during manufacture of the dielectric buffer pieces. An absolute minimum of impurities must be assured, the density of the constituent powder maximized, and each piece finally certified as being free of the smallest physical defect prior to use in order to attain the desired operating performance. It is also hoped that these crucial components may each be compaction die formed and sintered simultaneously with a respective power resistor (**63**), and perhaps even with the corresponding primary thermal conduit (**48**) as well.

The core (**49**) of the thermal conduit associated with each power resistor (**63**) should be constructed of a refractory metal or alloy tubing (such as molybdenum) which is jacketed with high-purity Alumina or a comparable material formed directly thereupon. Thus, each such primary thermal conduit becomes an extremely strong structural member which may also be used to support the stationary electrodes of the corresponding unit pentode array, while circulating the preferred liquid sodium heat transfer fluid or coolant around the outward faces of that power resistor to recover or remove its excess heat.

Each power resistor (**63**), its dielectric buffer (**45**), and associated primary thermal conduit (**48**) together comprise one power resistor assembly, and each such assembly in combination with its corresponding unit pentode array is considered for present purposes as forming the constituent component group of which the primary induction ring arrays and Field Induction System are comprised.

Extensive calculations reveal that a suitable power resistor material in most sizes of the single-stage embodiment of the Generator would be a ceramic material with the chemical formula $MgO.SiO_2$, which is available from CeramTec NA as CeramTec designation Steatite 357. A suitable power resistor material in most sizes of the preferred three-stage embodiment of the Generator would be a somewhat different ceramic material with the chemical formula $2MgO.2Al_2O_3.5SiO_2$, which is available from CeramTec NA as CeramTec designation Cordierite 547.

Overall Description of the Field Induction System Circuit

Before an overview of the EDF Generator's Field Induction System electrical circuit is presented, a somewhat comprehensive examination of the nature of the housing or hull configuration should be made, as in this application the hull itself is an active part of that circuit. Its electrostatic characteristics as a function of such configuration will therefore have a fundamental influence upon the nature of the external field envelope discharge current itself.

Hull Design

The derivation of the Generator's design hull configuration, as best seen in FIG. **3**, comprises a long and complicated story. This should not be taken to imply that the final and rather exacting composite shape selected is the only one that 'works' properly. It merely reflects the preferred embodiment of the inventor's conceptualization of a fundamentally mechanical device which happens to entail both tremendous electrical complexity and inspiring aeronautical implications.

Should the proposed vessel eventually be used to explore the possibility of transcending the time/light barrier, extremely precise calculations regarding its spatial displacement and charge/mass ratio will no doubt become necessary. Therefore, one of the principal reasons for the chosen hull design is that the total vessel displacement may readily be calculated using traditional formulas from analytic geometry for the volumes of the two truncated right conic sections and two one-base spheric zones, and the volume of the remaining cylindrical central chamber region.

Using just these formulas, a vessel hull configuration has therefore been developed which may be linearly expanded (as a scalar function of the hull radius) to virtually any size with no significant loss of accuracy. The Table of Dimensions which follows and the attendant schematic hull diagram of FIG. **3**, upon which all of the subsequent Detailed Calculations quoted are based, illustrates the application of this design technique to a theoretical vessel hull only 48" in diameter. It must be emphasized that the construction of an Electrodynamic Field Generator this small might not prove entirely practical in reality, due strictly to mechanical limitations, and the inventor believes this prototype model to be in fact the smallest such machine that could be built. This original hull size was deliberately selected, however, to encourage the development of the maximum possible accuracy with respect to contingent specifications for the relative size and positioning of the Generator's principal components, prior to the actual construction of larger machines.

The surface areas of the negative **(3)** and positive **(5)** hull sections are set equal to each other, and for two very important reasons: [i] to give the vessel hull a significant theoretical capacitance (despite its unusual geometry); and [ii] to render the two "driving" or discharge current portions of the field envelope uniform in cross section. Essential considerations which are related to these fundamental design criteria are best expressed in the following electrostatic formula for uniform field intensity in a DC capacitor which exhibits a continuous leakage current:

$$E^2=V\sigma/d\epsilon_o,$$

where V=equilibrium potential difference in volts,

σ=average surface charge density of one plate,

d=plate separation in meters,

ϵ_o=universal electrostatic constant,

and E=uniform field intensity in volts/meter.

This equation plays an important supporting role in the invention's Method of Operation, as previously discussed in that section above.

The volumes of the peripheral right conic sections of hull volume surrounding the Field Hub and and those of the one-base spheric zones comprising the Field Hub itself are also set equal in a preferred embodiment, as this criterion is postulated to promote both the structural integrity of the hull in any attempt to breach light speed (and establish a stable Kerr metric space warp about the vessel) as well as the gravimetric stability of the central chamber/cabin area in light of the relativistic mass effects involved while doing so.

As the case may be, to achieve this equal-area/equal-volume design, it was found necessary to employ two crucial interrelated constants (in addition to the ratio ⅕ used to radially "step off" the hull's interior space): the Hull Area Constant and the Polar Hull Constant. The Area Constant is (as the name implies) necessary to achieve the equal-area part of the overall design solution, and merely dictates the small deviation of the neutral **(4)** and negative **(3)** housing sections' ring radii from the stated preferred design increment of the hull's radial width.

The other requisite design factor, the Polar Hull Constant, is much more complex in nature. This value specifies the maximum height of the spheric hull zones **(5)** as a function of their arc-shape, and the corresponding contingent Polar Volume Differential is absolutely essential to achieving the equal-volume aspect of the preferred design. It is important to note that certain allowances must be made in practice for a finite but electrically-significant peripheral edge 'thickness' (and therefore surface area) at the equator of the device's negative housing Emitter Ring **(47)**, but will in general only minimally modify the "pure" design-model values obtained using the mathematical formulas given in the Table of Dimensions to follow.

Hull Configuration

Referring now to FIGS. **3** and **4**, the exterior housing or hull **(1)** is designed specifically to have a circular disc shape the thickness of which along the vertical axis of symmetry (that also constitutes the rotor's axis of rotation) is relatively minor as compared to the diameter along its radial centerline plane, and which tapers gently from a maximum thickness along such vertical centerline to a very minor thickness at the radial periphery **(47)** of the housing.

The tapered disc shape of the hull is postulated to promote the formation and maintenance of a stable bihemitoroidal corona or arc discharge field based upon certain empirical principles of electrostatics and upon laboratory experiments performed by the inventor involving concentric circular electrodes. Such an electrodynamic field may be characterized electrostatically in this application by an axial positive polar electric field contiguous with a biplanar negative electric field, with each field of one given polarity tending strongly to become oriented parallel to a field of opposite respective polarity but similar intensity.

The exterior housing or hull **(1)** is therefore divided into two axially-central positive electron collector sections **(5)**, as well as: [i] either a single radially-peripheral negative electron emitter ring **(47)** which is biplanar in cross section or two or more such ring sections **(3)** which are uniplanar in cross section having separate planar orientations; and [ii] two dielectric neutral sections **(4)** which are interposed between said positive and negative sections such that each such section of a given polarity is spatially separated from each section of the opposite respective polarity.

The total external surface area of the housing's negative Emitter Ring **(47)** or sections **(3)** is essentially equal to that of the positive collector housing sections **(5)** or zones, such that the overall housing configuration thereby acquires a significant theoretical static capacitance by design as intended. An estimated base value for such hull capacitance 'C' may be obtained by applying the surface area 'A' of the hull's two sets of "plates", as figured using the formulas in the Table of Dimensions below, to the standard formula for parallel-plate capacitance ($C=\epsilon_o A/d$). When plate separation 'd' is taken to be the simple average of the longest **(141)** and shortest **(142)** purely-semicircular arc current trajectories across the hull, as depicted in FIG. **11**, this estimate of hull capacitance is approximately 13 to 14 mmf for an EDF Generator 48" in diameter.

Field Composition

Referring to FIG. **11**, the external discharge current once initiated may be expanded as a function of the operative field voltage to fill two hemitoroidal spatial volumes of rotation **(140)** whose outer perimeters **(141)** are defined by a semicircular electron arc trajectory across the full housing radius in a plane perpendicular to the housing surface, or from the outermost point on the negative periphery thereof to the respective positive centerpoint of the housing lying on the vertical centerline thereof, and whose inner boundaries

(142) are defined by a similar such trajectory across the hull's neutral ring or sections **(4)** from any innermost point on the negative periphery to the closest point thereto lying on the respective positive collector section **(5)**.

The physical volume **(140)** of the external discharge current as described has a uniform cross-sectional area in the radial direction of electron conduction which corresponds to any such semicircular arc trajectories between points on the housing's negative Emitter Ring **(47)** and the respective positive collector housing sections **(5)** and therefore, by the definition of a conductor, such discharge current must assume a uniform current density. While very difficult to effectively describe in verbal terms, the idealized mathematical configuration of the Generator's heat- and thrust-producing field envelope region (as a function of its housing geometry) is visually easy to grasp and is clearly illustrated in FIG. **11**.

As indicated in FIGS. **12** and **13**, stationary electromagnetic armatures **(37)** are uniformly distributed into two circular arrays **(145)** containing equal numbers thereof which are positioned on opposite sides of the rotor from one another, and which are each located concentric to the rotor's axis of rotation midway between the negative Emitter Ring **(47)** and positive housing sections **(5)**. These armature arrays **(145)** may be employed as two separate groups to impart a holistic but attenuating rotational magnetic vector moment upon the external discharge current, reaching through the thin neutral sections **(4)** of the housing, and thus may be used to modulate the field envelope's kinetic and electrical characteristics in a variety of ways useful to the invention and its objects.

FIGS. **12** and **13** also show the electric field vectors **(143)** as described above, and magnetic field vectors **(144)** as produced and modulated by the armatures **(37)** comprising the two armature or variable inductor arrays **(145)**. FIG. **13** specifically depicts the regular lateral bending or displacement of the radially-impinging electron trajectories **(143)**, under the influence of the inductor arrays **(145)**, which is postulated to have a unitizing effect on the driving portions **(140)** of the field envelope.

It can therefore be seen that the device contemplated herein may be operated in such a manner that its external breakdown discharge current field achieves a uniform conductive cross section and current density, as well as a holistic and orderly rotational aspect which subsequently acts to assist in rendering such current quasi-coherent. And thus, for purposes of this device and its applications, such a discharge current field should be thought of as a special qualified type and form of DC corona or arc discharge—having significant propulsive potential (to be further examined)—which then merits technical description as an electrodynamic field.

In this respect, it may for now be emphasized that a useful level of electrically-developed thrust may be realized from the relativistic impulse of the electrodynamic field current electrons incident upon the positive collector housing sections and, postulating the operation of the device in the vacuum of space (with a physical field current volume as described), the impulse velocity of such incident field electrons may be raised to as high as 99.99% of the speed of light and more with the corresponding relativistic mass thereof then equal to 69 times their rest mass or indeed much more.

Primary Voltage Expansion Ratio

In keeping with the logic presented at the opening of the Field Induction System section above, the primary array voltage will tend to equal only one-half (½) of the DC voltage expressed across the rotor segments (or "rotor voltage"). Therefore, to develop an extraordinary multi-megavolt 'secondary' or field voltage across the Generator's housing of the type just described, the primary array cathodes' electron emissivity (expressed as a certain factor) must exceed that of the rotor field emitters by a ratio which is at least as large as the ratio of the desired field voltage to half the observed said rotor voltage. The latter such ratio is then thought of as the primary voltage expansion ratio in this application, as outlined in the Method of Operation section hereinabove.

A difference in the instantaneous rate at which these two said sets of parts will tend to discharge under the impetus of a given equivalent voltage or potential difference is thereby created. This instantaneous charge differential enables the primary cathodes to strip native electrons from the positive housing sections **(5)** thousands of times faster than such charge can reach the negative emitter housing sections **(3)** from the rotor field emitters. Thus, an instantaneous charge differential is perhaps best described as a cumulative charge imbalance that is known to operate (under certain conditions) in capacitive and/or thermoelectric circuit arrangements and to express itself as an elevation of the applied voltage or potential difference therein.

To actually achieve and support such a condition requires the use of two housing charge ballast capacitors whose negative 'plate' elements or rings are affixed directly to the rotor segments (about their outer perimeter) and which are immediately able to store an amount of such charge that is sufficient to produce equilibrium collector surface charge densities resulting in the desired field voltage. This average required housing charge density is calculated using the traditional formula given in the Hull Design section above.

The voltage applied to the positive rings of these ballast capacitors is taken from the rotor's outer field coil anode ring connections and therefore represents a large portion of the Primary Power System's generated field winding voltage. It should be noted that this voltage must actually be passed to such capacitors beneath the rotor ballrace assemblies in the form of recessed conductive traces on one or more of the segment separators. This same technique must also be employed to connect electrical elements of the Primary Power System lying outboard of the variable inductor arrays **(145)** with corresponding elements lying inboard thereof.

In any event, these crucial ballast capacitors are therefore properly thought of as belonging to both the Primary Power System and Field Induction System circuits. They provide the necessary electric force means by which a given primary voltage expansion ratio (as defined) may be supported and realized using the engineered instantaneous charge differential concept described. Because of the complexity of this aspect of the Generator's operation, the specific example below will illustrate the relevant principles using the respective temperature emissivity factors of the chosen field emitter and primary cathode materials with respect to the aforementioned 4ft. prototype model.

This temperature emissivity factor, $e^{-\phi/kT}$, where 'e' is the base of natural logarithms 2.71828 . . . , 'k' is Boltzman's constant, and 'T' is the absolute temperature in °K, derives from the famous Richardson-Dushman equation: a formula for the correct thermionic emission current density 'J' of a clean metal cathode in vacuum having surface work function ϕ. And, expressed in mathematical terms, the principle of ratio equivalence defined above may be stated as follows:

$$\text{Primary Voltage Expansion Ratio} \approx \text{cathode } e^{-\phi/kT} / \text{emitter } e^{-\phi/kT}.$$

Using field emitters of sintered 0.68/0.32 tungsten/copper composite and thorium-adsorbed tungsten primary cathodes, a 4 ft. diameter three-stage-rotor prototype unit of the invention's design would achieve a primary voltage expansion ratio equal to about 12,106 given a field emitter ϕ=4.408 eV and primary cathodes of ϕ=3.639 eV with both component sets at a temperature of 948° K (or 675° C.) as figured per the above equation. The field voltage expressing this ratio is verified in the Detailed Calculations hereinbelow.

It is important to note that this equation can readily be solved for the required primary cathode ϕ where the projected mean operating temperature of the primary cathodes is given (as it would be in operation), and where the field emitters' ϕ is also fixed and known but their actual operating temperature within a range of possible values must be experimentally or theoretically obtained.

The external breakdown discharge current once initiated and sustained in this manner is generally limited only by the engineered design characteristics of the unit pentode arrays which comprise the Field Induction System. In particular, the negative DC bias voltage applied to the unit pentode array control grids (**65**) must be sufficient to limit the total primary induction ring array current to a value within the safe operating ampacity of the rotor. For purposes of this text, the term "full-power" as used in relation to the Primary Power System and its incorporating device shall be taken to indicate a corresponding rotor speed wherein a full nominal field envelope voltage is achieved, maintained, or exceeded. Such an operational state is depicted in FIGS. **11**, **12**, and **13**.

Field Current Bias

In a simple single-stage rotor embodiment of the Generator used as a "Thermal Power Unit", no active biasing of the current in either of the two current-carrying portions (**140**) of the field envelope (which are otherwise symmetrical) need take place at the primary arrays, and the current comprising each such driving portion of the field is essentially equal. In the preferred three-stage rotor embodiment intended for propulsive units, the field current is actively biased or proportionally shunted by the primary arrays between such field envelope portions (**140**) to render the impulse thrust developed thereby mutually non-isometric along the Generator's vertical centerline. Several important considerations directly related thereto are discussed below, and reference should generally be made to FIG. **7** in this section.

The Generator's maximum full-power total field current must be equal to the rated DC parallel ampacity of the rotor segments, as figured using traditional methods. An operational minimum full-power field current in any propulsive three-stage rotor device used as an "Impulse Drive Unit" should then be considered as equal to either one-half (½) of the so-rated rotor ampacity or a level of total rotor current sufficient to allow the Generator to produce a net vertical thrust equal to its own weight, whichever is the lesser.

Given the minimum engineered amplification factor of the unit pentode arrays (**69**) earlier stated, this minimum full-power field current may in all likelihood be maintained with an average value of primary array negative bias voltage (as applied to the stationary control grids (**65**) shown in FIG. **7**) which is less than or equal to about one-half (½) of its specified design-maximum value. If no unidirectional or alternating voltage component of the Field Induction System current is present, this said average negative bias voltage value constitutes the primary arrays' "zero-signal" bias voltage and the primary arrays may be said to be in a zero-signal field bias condition (as mentioned earlier hereinabove).

Each of the two said driving portions (**140**) of the Generator's field envelope may be referred to as a "field hemitorus" because of its postulated shape, and the maximum individual field hemitorus current at any given value of operating rotor current should at all times be limited to a level which is less than or equal to two-thirds (⅔) of that operating rotor current. Peak net field thrust will then equal one-third the total isometric thrust. Assuming the said given amplification factor, this maximum advisable hemitorus level of proportional total field current may in all likelihood be maintained with an average value of negative bias voltage applied to the corresponding primary array's control grids (**65**) which is equal to approximately one-third (⅓) of its specified design-maximum value (or less proportionally at a rotor current level which is less than rated ampacity).

Due to the lag time in heating either of the two power resistor (**63**) networks through which an increasing field hemitorus current must return, the negative bias voltage applied to the corresponding primary array's control grids (**65**) may be briefly reduced to as little as one-sixth (⅙) of its specified design-maximum value (as a standard minimum operating control grid voltage) or less proportionally at a rotor current level which is less than rated ampacity.

The minimum individual field hemitorus current at any given value of operating rotor current should at all times be maintained at a level equal to or greater than one-third (⅓) of that operating rotor current. Assuming the said given amplification factor, this minimum advisable hemitorus level of proportional total field current may in all likelihood be maintained with an average value of negative bias voltage applied to the corresponding primary array's control grids (**65**) which is equal to approximately two-thirds (⅔) of its specified design-maximum value (or less proportionally at a rotor current level which is less than rated ampacity).

Due to the lag time in cooling either of the two power resistor (**63**) networks through which a decreasing field hemitorus current must return, the negative bias voltage applied to the corresponding primary array's control grids (**65**) may be briefly increased to as much as five-sixths (⅚) of its specified design-maximum value (as a standard maximum operating control grid voltage) or less proportionally at a rotor current level which is less than rated ampacity.

It is important to note that the current passed by the unit pentode arrays (**69**) will depend to a great degree on the accelerator grid (**66**) voltage rather than on the overall primary array potential difference, just as in any standard vacuum tube which employs a screen or accelerator grid whose positive voltage is less than the applied plate voltage. It therefore becomes evident that this positive grid (**66**) voltage may also be modulated in a manner similar to the negative control grid (**65**) voltage. This type of 'duplex' signal-handling or control voltage response capability allows the unit pentode arrays (of either primary induction ring array) to amplify a resonant frequency communications signal while simultaneously controlling an independent level of current-contingent thrust.

In any event, it can be seen here in light of the Law of Conservation of Momentum that a sizeable force of thrust will be developed by each field hemitorus (**140**) from the relativistic impulse of its constituent electrons incident upon the collector housing sections (**5**), and that suitable means have been provided for rendering such mutually-opposed y-axis thrust vectors variably non-isometric.

A less obvious but very important distinction which should be made in regard to the EDF Generator's production of such electrically-developed thrust is certainly this: it is not necessary that the Generator or in particular its positive hull sections (**5**) (or Field Hub) provide the work required to move the field current charge against the field potential gradient. Per classical electric field theory, the work may be done by the charge itself in being repelled along the potential gradient. In the first case, work is 'performed' and in the second it is 'recovered'. Furthermore, it can be shown mathematically in light of the preceding distinction that any net recoil force experienced by the negative hull Emitter Ring (**47**) is strictly Newtonian in nature, compared to the relativistic impulse thrust produced at the Field Hub, and that the peak value of such recoil would amount to no more than about $^{3}/_{1,000}$ of one percent of the peak net value of the said thrust.

Therefore, the Electrodynamic Field Generator's input rotary torque drive means will provide the work-energy required to establish and maintain the rotor's rotation and the resultant field potential gradient that is induced in the manner described, and the electron charge comprising field hemitorus current will then do the work required to effect its passage toward the positive Field Hub. The resulting collisions are almost completely inelastic, so momentum and kinetic energy are conserved independently, and the field current's gained kinetic energy is recovered almost entirely as heat.

Field Voltage Control System

Accordingly, it can also be seen that the interactive factors or features which result in the EDF Generator's exhibited field voltage as a function of its rotor voltage are much more complicated in nature than the simple turns ratio which defines an AC transformer's secondary voltage in terms of its primary voltage. And because of the extremely high primary voltage expansion ratios which will be operative in large devices, small fluctuations in the Generator's stationary primary array electrode voltages can conceivably cause very large and undesirable fluctuations in the exhibited field voltage. Therefore, a relatively simple but effective Field Voltage Control System is provided which may be used to monitor, regulate, and adjust all individual primary cathode and grid voltages to optimum specified and/or interim operating values such that rotor ampacity is not exceeded, and this system anticipates the use of an associated onboard computer system (to render its various interactive functions automatic).

The embodiment of such a control system as shown in FIG. **10** is intended merely to provide a logical circuit model which is illustrative of the principles of the invention, and of certain principles of appropriate circuit resolution which should be applied within such a system in light of the following aspects of its teaching, and not to preclude the use of other possible embodiments of such control system circuitry.

The cathode elements (**64**) of each plane-parallel electrode system or unit pentode array (used to link the rotor assembly to the positive housing sections) are to have a surface work function ϕ which is significantly lower and/or an operating temperature which is significantly higher than that of the rotor field emitters (**17**), and will therefore tend to exhibit a comparatively much-higher electron emissivity, such that the ratio of their respective temperature emissivity factors (which are each represented by $e^{-\Phi/kT}$) has a very determinate effect upon the respective exhibited ratio of the device's external discharge field voltage to the primary DC voltage generated internally.

Referring to FIG. **10**, these primary array cathodes (**64**) must be composed of tungsten or another refractory metal which may also be either impregnated with thorium oxide or thorium metal submonolayer-adsorbed to achieve such lowered comparative surface work function, and, to calibrate their emissivity relative to that of the rotor field emitters (**17**) as a means of defining and determining the device's external field voltage relative to its primary array potential difference (or half the rotor voltage).

The Field Voltage Control System as shown is basically designed to directly accept the rather high positive induced voltage of the stationary anode rings (**58**) (shown in FIGS. **8** and **9**) for the purpose of providing the primary accelerator grids (**66**) with their standard DC operating voltage. Due to the linear scalability of the Generator's overall design, the nominal (or zero-signal) value of such voltage relative to chassis ground (**10**) is expected to be approximately +362 to +483 volts per foot of hull diameter in three-stage and single-stage Generator embodiments (respectively). An isolation diode (**126**) and switch or relay (**124**) prevent such positive voltage from being neutralized by negative current (s) elsewhere in the control circuit.

Any one or more of the stationary electrode elements (**64**)–(**67**) comprising each said unit pentode array may be artificially cooled, as a means of further regulating the device's external field voltage, and/or subjected to modest variable DC control voltages at times to further assist in the interim adjustment of that unit pentode array's current.

A very high value capacitor (**116**) or a plurality of separate high value capacitors connected into a parallel or series-parallel matrix thereof may be employed to store the displacement charge caused within the Primary Power System by the charging of the rotor-mounted ballast capacitors, as part of such Field Voltage Control System. Said displacement charge holding capacitor(s) (**116**) may hereinafter be referred to as "ballast compensating capacitor(s)".

A very high value capacitor (**117**) or similar such capacitor matrix may also be employed to prevent the grounding of positive housing section potential by ambient ionization charge, from operation in air during the device's start-up period before such sections are fully enclosed (by the external field hemitorus currents), as a further part of this Field Voltage Control System. Such bulk electron storage capacitor(s) (**117**) may hereinafter be referred to as "ambient charge capacitor(s)".

The negative plates of these ambient charge capacitors are to be isolated from ground during rotor spin-up, in order to effect a net negative charge storage, and so their positive plate charge is supplied by a dedicated common variable DC voltage supply (**98**) which is oppositely connected to the ground frame. Each ambient charge capacitor (**117**) employed during the device's start-up period may be selectively switch- or relay-connected by its negative terminal to an accelerator grid element (**66**) of one or more of said plane-parallel electrode systems, in order to effect and control the storage of such ionization charge. Each such start-up capacitor (**117**) may also be selectively switch- or relay-connected by its negative terminal via a variable resistor and/or diode to a suppressor grid element (**67**) of one or more of said plane-parallel electrode systems, in order to effect and control the distribution of onboard stored negative charge reserves from the suppressor grid(s) (**67**) to the rotor, as a standard means of regulating the external field voltage (especially in the vacuum of space).

Said ambient charge capacitor(s) (**117**) may also be similarly connected by the negative terminal(s) thereof either to a separate source of true earth ground or to one or more superconductive current storage ring/s (**200**) (including charge deposition and retrieval means), for purposes of removing any excess ambient electron charge above the bulk storage capability of such capacitor or capacitor matrix which may be produced from normal or full-power operation in air—or other gaseous dielectric media.

Such superconductive current storage ring(s) would be used exclusively in large three-stage rotor propulsive models of the Generator, and thus constitute/s the means by which further expendable onboard charge reserves may be accumulated. The working core of each superconductive storage ring would preferably be a large but relatively thin torus (of small cross section) composed of solid mercury metal immersed in liquid helium, and as such would have a known ampacity measured in the hundreds of thousands of amps.

A linearly-actuated control rod (**103**) or similar mechanism may be used to effect a mechanically variable thermal junction between each primary array cathode or grid element (**64**)-(**67**) and a cold thermal mass of the same or similar refractory metal, maintained at the reservoir temperature of a secondary cryogenic housing coolant such as liquid air or nitrogen. The plate elements and dielectric media of each capacitor or capacitor matrix employed in the Field Voltage Control System could also be incorporated (in a suitable manner) into the coolant vessel which contains said cold thermal mass. This System's diodes would generally not be rated for mounting at such a low temperature.

The voltage control circuit shown in FIG. **10** is modular in nature, in that one such unit is provided for each of the 72 unit pentode arrays associated with the Field Induction System (in the preferred embodiment). All measurements of primary electrode voltage(s), and of positive housing section voltage, are to be made with respect to the Generator's ground frame (**10**) (as generally depicted in FIG. **1**): the metallic structural shell of the device's central chamber to which the rotor carrier assembly drive motor(s) should be mounted. This grounding method is referred to throughout as "chassis ground".

One thermal control rod (**103**) is provided for each primary cathode and grid element (**64**)-(**67**) which variably engages both a control rod socket (**104**) attached thereto and a thermal junction (**101**), which is attached to a cold thermal mass (**102**) maintained at the reservoir temperature of the secondary (cryogenic) housing coolant, by means of a linear actuator (not shown). Such control rod (**103**) may then be used to regulate that electrode's temperature to approximately the same temperature as the associated power resistor (at 675° C. ±55°). In this manner, small adjustments in cathode (**64**) emissivity may be made as necessary during operation and the various grids may be closely matched in temperature to that of their respective cathode(s) to ensure the accuracy of mutual primary electrode voltage balance. Any deviation of cathode (**64**) voltage from a chassis ground potential may be detected and measured across a very-high-value resistor (**107**), connected between said cathode(s) and chassis ground, which is preferably of a large carbon type having multiple fixed taps (and across which an output voltage may exist).

A dedicated common variable DC voltage supply (**95**) is provided for all primary cathodes (**64**) as a means of ensuring an optimum cathode potential with respect to ground, despite fluctuating power resistor temperatures (in three-stage rotor devices where field hemitorus current is variably biased). This supply is shunted with a double-pole/double-throw switch or relay (**123**) to enable either a positive or negative interim control voltage to be applied to the cathodes (**64**), in order to hold their true potential as close to ground as possible. A similar voltage supply (**97**) and switch/relay (**123**) is also provided for all primary accelerator grids (**66**) as a means of modulating the fixed level of applied positive grid voltage (from stationary anode ring (**58**)) and therefore the level of field hemitorus current, independent of the level of applied control grid (**65**) bias. In both cases, these DC supplies are isolated from chassis ground by blocking capacitors (**106**) to inhibit any net charge loss or accumulation from accruing to the ground frame (**10**) during operation (in propulsive units which are not connected to true earth ground).

A dedicated common variable DC voltage supply (**96**) is provided for all primary control grids (**65**) as well. This voltage supply is not capacitor-isolated from chassis ground in this case, so that the primary control grids (**65**) and cathodes (**64**) may have a common direct ground reference as they would in a typical vacuum tube circuit arrangement. The said control grid supply (**96**) should also have a resistor (**107**) of the same type as that provided for the associated cathodes (**64**), connected between such supply's positive terminal and chassis ground. These two criteria further ensure the accuracy of mutual primary electrode voltage balance, and the nominal (or zero-signal) value of such DC control grid voltage relative to chassis ground is expected to be approximately −36 to −48 volts per foot of hull diameter in three-stage and single-stage Generator embodiments (respectively). A variable bypass resistor (**108**) may be adjusted to ensure a very low and relatively constant DC operating grid current despite fluctuations in the value of the AC input signal resistor (**110**), as discussed below.

Ballast compensating capacitors (**116**) are provided whereby negative outrush charge from the stationary anode rings (**58**) and the positive plates of the rotor ballast capacitors may be purged from the Primary Power System during rotor spin-up, to prevent stage electrode array voltage ratings from otherwise being greatly exceeded, via these same two anode rings (one per primary induction ring array), connecting switch/s or relay/s (**124**), and resistor/s (**125**). During the Generator's spin-up and 'run' periods, each ballast discharge switch/relay (**122**) is normally open; switch/relays (**124**) are 'run-open'. Upon rotor de-spin, such ground-restoring charge must be proportionately returned to each stationary anode ring via switch/relays (**124**) and resistors (**125**), and to the Primary Power System as a whole via the discharge switch/relays (**122**) and suppressor grid discharge shunts (with variable resistors (**113**) and diodes (**115**)), for the most part after vacuum chamber current to the hull Emitter Ring has essentially ceased.

Then, such normalizing charge may be 'dumped' across the described suppressor grid discharge shunts at a rate sufficient to ensure that the entire Primary Power System returns to a ground potential (by brief direct surge rotor-shorting to the outer induction array anode rings). Thus, resistors (**113**) provide a variable time constant for the discharging of the said compensating capacitors (**116**). An a independent common variable DC voltage supply (**99**) whose negative terminal is connected to the ground frame (**10**) ensures that the proper total ballast compensating charge described may be stored against the high positive applied potential of the stationary anode rings (**58**).

The ambient charge capacitors (**117**) absorb ionization charge (arising from operation in air) which perforce must fall into the potential well(s) established by the positive housing sections at least until the field envelope's two hemitorus currents are fully formed, and which would otherwise tend to ground the desired positive housing section voltage. Therefore, the negative plate voltage of these storage capacitors (**117**) is applied from another dedicated DC supply (**100**), which is similar to that (**99**) provided for the compensating capacitors (**116**), by means of a common dpdt switch/relay (**119**). None of these DC supplies (**95**)–(**100**) are to have internal load-line diodes.

The variable output of this "dynamic compensating" voltage supply (**100**) is to be approximately the same as that of the ambient charge capacitors' positive voltage supply (**98**), thereby allowing the said capacitors (**117**) to be gradually saturated during rotor spin-up with electrons collected at the accelerator grids via storage shunt inrush current switch/relays (**121**) and diodes (**114**). During the Generator's spin-up and run periods, the "control gate" switch/relays (**119**) are normally closed such that the negative pole of supply (**100**) is connected to the negative plates of these same capacitors. In the Generator's run and de-spin periods, however, switch/relays (**121**) are normally open.

Any excess ionization charge above these capacitors' joint storage capability which might lower the 'run' Field Induction System circuit voltage must then be either grounded off or deposited in said superconductive storage ring(s) (**200**) (in ungrounded propulsive Generators designed for maximum ambient charge storage), via the same said throw of control gate switch/relay (**119**) in conjunction with shunt switch/relay (**120**) and resistor (**112**). In this case, the storage shunt switch/relays (**121**) remain run-closed. Otherwise, the slow 'run' release and distribution of negative charge reserves to the rotor from the storage capacitors (**117**) is effected by the gradual synchronized relaxation of the voltage/s from supplies (**98**) and (**100**) across switch/relay (**119**) and discharge resistor (**113**), with shunt switchs (**120**) and (**121**) open, while diodes (**114**) and (**115**) ensure that the said capacitors (**117**) may only be discharged in the manner described.

This discharge of stored charge reserves actually takes place from the suppressor grids (**67**), which may serve as low-power electron emitters given their proximity to the primary anode rings (**68**). An isolation diode (**127**) prevents such negative charge from neutralizing accelerator grid circuit voltage. It should be noted that the suppressor grids may be held at ground potential as they would in a typical beam power pentode circuit arrangement, or at a very low if not negligible negative voltage for purposes of effecting such controlled low-power charge release, by means of [i] a minor negative voltage induced thereon by the corresponding proximal primary anode ring/s (proportional to their capacitance relative thereto); [ii] the combined DC voltage drop across the paralleled resistors (**109**) and (**111**) connected to their respective accelerator grids; and/or [iii] the voltage of the dynamic compensating supply (**100**) relative either to earth ground and resistor (**112**) or to chassis ground and the relative opposing voltage of the ambient charge storage capacitor's supply (**98**).

The bypass resistor (**109**) provided between each accelerator grid (**66**) and its associated suppressor grid (**67**) returns the inevitable stationary anode ring (**58**) and accelerator grid (**66**) DC circuit current to the rotor from the suppressor grids (**67**), and simply assists in maintaining each suppressor grid at a DC potential as close to true ground as possible. The suppressor grid resistor (**111**) which is paralleled with said bypass resistor (**109**), however, is analogous to a plate resistor on behalf of the corresponding inner (rotor) induction array in both the single- and three-stage rotor embodiments, in that any rotor circuit final-stage AC output voltage variations will be reproduced across it once field envelope circuit current commences. Likewise, any AC signal oscillations from an outside (remote) source which are induced upon the field envelope current will immediately act as a detected first-stage AC input for rotor circuit amplification, and again will be reproduced across the said grid resistor (**111**).

Therefore, any rotor or field circuit AC potential difference across this grid resistor (**111**) provides an output signal voltage that may be: [i] received by an actual communications console or a simple AC control-voltage operating circuit in the Generator's central cabin; and/or [ii] reapplied either in- or out-of-phase with the voltage across the input signal resistor (**110**) of the associated control grid (**65**) as a further AC signal voltage amplification or suppression stage (respectively). An input AC signal or control voltage may also be applied across the control grid resistor (**110**) from such an onboard communications console or control-voltage operating circuit, as would generally apply to three-stage and single-stage devices, respectively. This grid resistor is variable for circuit tuning purposes related to resonant frequency communications (if desired).

It can be seen from FIGS. **8** and **9**, however, that the AC output amplification stage (or inner induction array) of the Primary Power System and its input stage (or outer induction array) are directly coupled by the rotor segments (**14**) in both single-stage and multi-stage rotor embodiments. Thus, the rotor electrical circuit as configured will be subject to continuous positive (or regenerative) feedback of any AC signal voltage present either in the field current or on the rotor, as described above. To provide a desired or requisite level of signal amplitude suppression or compensation, negative (or inverse) feedback may be applied to the rotor circuit using the feedback coupling relays (**129**) and stage transformers (**130**) shown in FIG. **10**.

Because plate voltage and grid voltage are typically 180° out of phase in a beam pentode arrangement such as the Generator's primary arrays, inverse feedback is achieved when the voltage fed back to the control grid (**65**) has the same waveform and phase as the plate or anode (**68**) voltage. In related traditional practice, an inverse feedback signal applied to the input (control grid) current decreases AC amplifier input impedance and distortion; an inverse feedback signal proportional to the output (plate/anode) current raises the output impedance of such amplifier. Therefore, an inverse feedback AC signal proportional to the current through resistor (**111**) which is applied in-phase with the AC line current across resistor (**110**) decreases the rotor input stage impedance yet also reduces the amplitude of the AC output current. A loss of total amplification (through a fractional stage 'gain') results, but is accompanied by a decrease in signal distortion.

The stage feedback transformers (**130**) ensure voltage reference isolation between such stationary amplification or suppression stage and the corresponding rotor stages, when AC signal feedback across the control grid resistors (**110**) is necessary or desired. A double-pole/double-throw switch/relay (**129**) may be used to decouple each such transformer from the input signal resistor (**110**) side of its circuit connections, and/or to apply either positive (regenerative) or negative (inverse) feedback to the rotor AC amplification stage(s), and tunable blocking capacitors (**131**) isolate the transformers from circuit DC voltages. The coupling transformers (**130**), isolation relays (**129**), and blocking capacitors (**131**) are therefore only necessary in Generator embodiments not having a central AC control-voltage operating circuit with suitable inverse feedback characteristics and in three-stage Generator embodiments which may be used in a communications capacity. In the latter case, an optional capacitor across the secondary or control grid resistor (**110**) side of each such transformer (in parallel with the said resistor) may once again be employed in parallel resonance fashion, and is indicated in FIG. **10**.

During continuous operation of an EDF Generator in the vacuum of space, the exhibited DC field voltage will tend to rise gradually as the total amount of charge contained in the field envelope circuit is gradually reduced through unavoidable electron leakage losses (primarily at the hull Emitter Ring peripheral edge). To compensate for this effect, a small amount of stored ambient charge must therefore be continually released at the suppressor grids (**67**) into the rotor return current. The rate at which this gradual discharge must be effected is dependent upon the observed field leakage rate, as only a rough approximation of this leakage may be pre-calculated. It is for this reason that at least one superconductive storage ring (**200**)—containing a tremendous quantity of reserve ambient charge in the form of a continuous zero-loss current—must be included in any ungrounded Impulse Drive unit used in space exploration, as the capacity of such storage ring determines the vessel's effective range of operation.

It is postulated that the external field current's energy density will be sufficiently enormous that it exhibits a virtual self-evacuating operational condition above a certain ultra-high field voltage in propulsive three-stage EDF Generators, and such a condition is essential to the successful atmospheric ground launch thereof Such kinetic occlusion of a significant portion of the otherwise available ambient ionization charge may also occur once the field envelope is fully formed even in ground-based single-stage devices operated at breakdown field intensity. During continuous operation of such ground-based devices in a gaseous atmosphere, field leakage losses which could cause an undesired gradual field voltage rise if they exceed available ambient charge accretion may be compensated by bypass-shunted charge from earth ground. In this case, the positive pole of supply (**100**) is connected at low power via dpdt switch/relays (**119**) to earth ground, in order to avoid a limited spin-up/run/de-spin repeating duty cycle.

General Method of Construction

The actual construction of an Electrodynamic Field Generator of any given size is relatively straightforward, whether of a manned propulsive embodiment or not, and proceeds in fact from the inside out. Referring generally to FIG. **1**, one of the most practical features of the construction process is that virtually the entire instrumentation and payload compartment or "Central Chamber" (**2**) may be outfitted first, before any portion of the Primary Power System is assembled or even necessarily designed. In the preferred three-stage embodiment to be used as an aerospace Impulse Drive Unit, the central chamber should obviously incorporate low-density materials wherever possible.

Construction next proceeds with the Generator's ground frame (**10**): the metallic structural shell which provides an enclosing framework for the central chamber to which all ancillary inboard electrical equipment may be grounded. This structural framework should be as strong, lightweight, and nonmagnetic as possible, and may preferably be of a welded tubular design using stainless steel or a suitable titanium alloy.

Once the ground frame and enclosed central chamber are completed, including the installation of a preferred total of 32 high-torque DC motor-generator rotor drive units (in four sets of eight), the two rotor mounting frame sections and attached ring gear/s (**8**) of the "Carrier Assembly" may be built and dynamically balanced (using a temporary peripheral spacing "jig"). Then, construction of the composite "Rotor Assembly" may commence with the laying out of equal numbers of copper segments (**14**), ceramic segment separators (**16**), and refractory composite field emitters (**17**) which are all of exactly equal weight by type, the latter components providing the nonconductive base to which other rotating electrical hardware may be attached.

When the three sets of said principal large rotor pieces (which preferably number 180 each) are bound together with electrode rings silver-soldered to the segments, and clamped between the two halves of the centrifuge-style carrier assembly, the assembly of the "Primary Power System" itself may begin. Once thin insulative rotor surface "facing", the aforementioned recessed and conductive separator traces, and the rotor ballraces (**25**) have all been affixed to the base rotor assembly, and the field coils, electrode arrays, and other rotating components have been added, the fabrication of the vessel's "Structural Intercooler System" may proceed. When this secondary thermal conduit (hull coolant) system is completed and has been pressure-tested, the "Primary and Magnetic Arrays" may be installed—including the 5-armature variable inductor array sections depicted in FIG. **14**. The primary (induction ring) arrays here are comprised of the unit pentode arrays (**69**) as individual assemblies each having an adjunct assembly comprising a field power resistor (**63**), dielectric buffer (**45**), and primary thermal conduit (**48**).

Finally, once the rotor has satisfied nonenergized operational mechanical clearance and dynamic balancing criteria (using a temporary outer rotor bearing support "jig") and a final complete thermal conduit system pressure test has been made, the "Outer Hull Components" may be installed. Thus, the basic vessel construction may be completed in a total of seven distinct steps or stages. Due to the tremendous weight of the finished rotor (as the device's most massive single assembly), an operating characteristic of utmost importance is the actual rotor speed required to maintain a specification field voltage which is figured using the method outlined in the Detailed Calculations below. This will determine the rotor's operating angular momentum, which in propulsive units must be sufficiently large to provide gyroscopic stability against the use of variably-imbalanced isometric thrust (produced by field hemitorus current) but not so large that the vessel's navigational and maneuvering characteristics are thereby rendered sluggish.

It can be shown mathematically that the nominal rotor speed for a 4 ft. diameter theoretical design prototype would be roughly equal to or somewhat less than that of a typical small electric motor. However, as the size of the device increases, it will be necessary to reduce rotor rpm due to the extreme increase in the rotor's 'tip' speed and therefore the centrifugal forces to which it is subjected. Sufficient information is contained in the basic specifications which underlie this disclosure to allow engineers to reduce rotor speed as necessary in larger devices following the course of an intensive program of small device tesing; e.g., the projected nominal rotor rpm for a 20 ft. Thermal Unit is 1045.

Structural Intercooler System

The secondary thermal conduit sections which carry hull coolant or heat transfer fluid through the Generator's tapered "Drive Ring" (or the combined volume of the negative and neutral housing sections) must pass between the power resistor and unit pentode array assemblies, and so they must conform in outside dimension to the available space within the supporting framework of the neutral hull construction. The most unusual feature of these secondary conduits is that they also comprise (of necessity) the principle Drive Ring load-bearing members, and together they form the EDF Generator's structural intercooler system as illustrated in FIG. **14**.

These secondary hull coolant conduits should be fashioned in a manner similar to the primary thermal conduits described hereinabove, and in the preferred embodiment would be formed in four types or varieties: [i] the initial outbound coolant supply runs **(75)**, called "headers"; [ii] the circumferential heat transfer sections **(76)** (to which the Emitter Ring hull plates are fastened), called "peripheral shunts"; [iii] the inner **(77)** and outer **(78)** intermediary coolant runs, or "transfer links"; and [iv] the radial coolant returns **(79)** or "recovery lines". Within the Drive Ring, each of these conduit types may be fashioned as single sections which connect one to another in 'series-parallel' branch systems or zones. To wit, one header **(75)** from a reservoir coolant manifold **(74)** connects to one peripheral shunt **(76)** which in turn connects to one recovery line **(79)**.

Thus, a single coolant zone is illustrated in FIG. **14**, the arrows indicating the direction of coolant flow. A molded ceramic support bracket **(71)** having integral elbows and a set of mounting "knuckles" may be used to secure the inner corners of each two adjacent negative hull plates **(72)**, by means of a matching mounting ring welded to each inner plate corner and a retaining pin (not shown). Each Emitter Ring hull plate **(72)** is to have an angle-stock thermal transfer channel welded along but just inside its inner edge (as indicated by dashed lines), which encloses the corresponding peripheral shunt **(76)**.

These coolant headers, shunts, transfer links, and recovery lines are preferably formed of molybdenum tubing, and would take advantage of the high thermal conductivity and electrical resistance of a thin Alumina coating thereon to absorb excess heat from the single-stage Drive Ring hull and (in particular) the Emitter Ring hull plates. In single-stage rotor devices which are used as Thermal Power Units, the recovery lines then exit the Generator housing through two exchange manifolds (as stated earlier), from a preferred total of 72 individual coolant zones, to an external heat exchanger. Just how effective this method of cooling the Drive Ring as a whole will prove to be during operation in air, using the preferred liquid air or nitrogen coolant, is somewhat difficult to project. Liquid helium could alternatively be used in this respect to ensure adequate hull cooling, but would entail significantly increased piping and pumping difficulties in doing so.

It is expected, however, that the intercooler system described will be satisfactorily effective during the operation of three-stage devices in the vacuum of space, and is principally designed with this mode of usage in mind. Accordingly, it is believed that an Impulse Drive Unit operated in space will be able to purge its excess heat by actually circulating the primary conduit liquid sodium or equivalent through the secondary thermal conduit system (instead of the cryogenic coolant). In this case, the intercooler structure is used to transfer heat to the Emitter Ring hull plates as a radiative heat sink. This cooling method would provide the necessary means of liberating the Generator from permanent ground-based support to achieve free-ranging operation in space.

Neutral Housing Section Construction

In order to establish and withstand the action of the field envelope discharge current, the surface of the neutral sections of the Generator's housing or hull (or "Neutral Ring") must be constructed entirely of specialized nonconductive material(s) which should be applied in two layers to an underlying superstructure that is actually conductive in nature. It is proposed that overlapping ceramic 'tiles' be used in this regard, in a manner rather like that employed on the NASA space shuttles, which are adhesed to an underlying deck of structural nonmetallic sheet stock While the intercooler system conduits (described in the preceding section) comprise the principal load-bearing members which support the Drive Ring portion of the hull, a system of beams **(153)** and struts **(152)** that are also composed of the same nonmetallic structural material as the deck or a similar and compatible material should be utilized to further strengthen and 'unitize' the Drive Ring superstructure such that an internal hard vacuum may be drawn and maintained prior to operation. These secondary load-bearing members are also depicted in FIG. **14**, and in a preferred embodiment these members and the base deck material itself would be composed of a carbon composite commonly referred to incorrectly as graphite.

Sophisticated carbon composites similar to those found in golf-club shafts and racing-bicycle frames were first developed to make use of their special combination of extreme strength, rigidity, and lightness in aerospace applications. These materials are generally made from polyacrylonitrile (PAN) carbon fibers which are heated under tension to drive off the non-carbon portion. The individual fibers are about 7 microns in diameter and approach 300 kpsi in tensile strength. They may then be pulled into a rope-like 'tow' or woven into a fabric-like 'matte' before being coated with either epoxy or polyester resin. Therefore, carbon composites are eminently suited to forming rods, tubes, and sheets, and display excellent vibration damping characteristics. The final product is, however, still very expensive per pound.

To simplify the vast diversity in specialized grades of PAN carbon composite for purposes of this application, this material should be considered as coming in two basic density grades: low-density "deck" grade (at about 35% of the specific gravity of sheet steel), and high-density "beam" grade (at about 65% of the specific gravity of structural steel). It is also available in two standard temperature grades, depending on the bonding agent used: a 350° F. rating (lo-temp) epoxy-bonded material and a 750° F. rating (hi-temp) polyamide-bonded composition. The hi-temp grade only should be employed in the Neutral Ring, but the lo-temp grade may be used to build the Central Chamber.

Thus, a low-density/hi-temp grade of carbon composite laminate should be used as the basic deck material in construction of the Neutral Ring hull sections. To wit, this deck layer itself is initially affixed to the vessel superstructure by solvent welding and/or mechanical means. The PAN carbon deck, a middle layer of ceramic substrate plates and an outer skin of exolayer tiles together then comprise the composite Neutral Ring hull. Several factors enter into the selection of an appropriate ceramic material for the substrate layer plates. Of primary concern here is that the material of choice have a very high volume resistivity and very low thermal conductivity at temperatures from 300–500° C. It is also important that the material have a very low thermal expansion coefficient, as the substrate layer plates must be bonded to the deck layer with a refractory-class adhesive. Therefore, the substrate material should also be non-vitreous and have relatively high porosity to promote adequate adhesion. One of the Cordierite class of compounds, bearing CeramTec designation 447, has relatively low hardness and flexural strength but should function admirably in Thermal

Power Units as the center layer of a bonded composite construction given its other highly desirable characteristics. Impulse Drive Units will require an alternate substrate plate material as discussed below.

The selection of a ceramic material for use as the exolayer dielectric thermal tiles of the EDF Generator's housing or hull is contingent upon the operative field voltage, and therefore different materials must be specified for Thermal and Impulse Drive Units. To prevent substantial skin conduction losses, the minimum dielectric value k for the exolayer tile material used in Thermal Power Units has been calculated (using the peak field voltage) at approximately k=9. Therefore, the material of choice is a Zirconia composition bearing CeramTec designation 848, with a minimum (high-frequency) dielectric constant of 28. Even though this material has a comparatively high thermal expansion coefficient, it has very low thermal conductivity and is very dense and hard. Moreover, it exhibits exceptional flexural strength and resistance to fracture.

The minimum dielectric value k for the exolayer tile material used in Impulse Drive Units has been calculated at approximately k=107. Therefore, the material of choice is the same titanate composition prescribed for the dielectric buffers (and described in some detail in a previous section). Known chemically as Sodium Bismuth Titanate (and commercially as Kézite), this highly unusual compound is technically a piezoelectric material, but is also ferroelectric in that its dielectric constant actually increases on temperature rise to a peak low frequency value of about 3,100 at its 655° C. curie point. Like most ferroelectrics, its tensile and flexural strengths are quite low yet it is very dense and exceedingly hard. It does, however, exhibit a thermal expansion coefficient which is very high for a ceramic material.

The expected higher flexural stresses upon the Neutral Ring of an Impulse Drive Unit (compared to a Thermal Power Unit) may prove sufficient to crack the substrate plates were they composed of the Cordierite material described above. The Zirconia compound mentioned earlier should therefore be substituted for Cordierite 447 as the substrate plate material in Impulse Drive Units. This would likely stiffen the composite Neutral Ring hull adequately to prevent such stress cracking, while effecting a much closer match in thermal expansion coefficient to the titanate tiles. In both Thermal and Drive Units, the exolayer tiles should be applied to the substrate plates with a refractory-class adhesive in a mosaic overlay pattern, such that the edges of each exolayer tile are evenly spaced as far as possible from the underlying edges of the nearest substrate plates.

Because the underlying deck layer of the Neutral Ring hull will be quite conductive when composed of carbon composite sheeting, and referring now to FIG. **1**, it becomes imperative to use an auxiliary dielectric buffer **(50)** between the inner edges of the Emitter Ring **(47)** hull plates and the outer edge of the said deck layer to prevent an otherwise enormous deck leakage current from occuring. Although it might seem that the much-cheaper and nearly-as-strong fiberglass composites available (which are nonconductive) would tend to minimize any such leakage current condition, their use would not prevent the possible build-up of unacceptable levels of static charge within the working induction compartment **(12)**. Therefore, the use of conductive deck material ensures that the Faraday shielding principle protects Neutral Ring components from transient static voltages because the presence of a finite deck leakage current means the deck layer is a charged conductor, whereupon any net static charge must be external.

A secondary concern is the actual operating temperature of the contact areas where the deck layer abuts to both the auxiliary **(50)** and primary **(45)** dielectric buffers, which must be kept below the 750° F. temperature rating of the carbon composite. It can be shown mathematically that the deck leakage current can be held to a thoroughly acceptable fractional amperage per cm^2 of such contact area if this condition is met. In Thermal Power Units, the use of a secondary cryogenic hull coolant should keep the auxiliary buffers **(50)** and outer deck contact areas at a temperature which is under the said rating. Doing so at the primary buffer **(45)** contact area locations may well require the use of auxiliary cryogenic thermal conduits in both Thermal and Drive Unit models of the EDF Generator. The auxiliary dielectric buffers in any given Generator should be composed of the same material used in that device's exolayer tiles, and should be of about half the radial thickness of the primary buffers employed therein.

While the specific materials discussed above should not be construed as critical to the EDF Generator's proper construction, their properties and characteristics are highly indicative of the rather involved considerations which must be undertaken to ensure that the neutral sections of the vessel hull are physically suited to optimizing a given size device's performance and minimizing its housing maintainance or repair requirements.

Stationary Electromagnetic Armatures

Because the EDF Generator's stationary electromagnetic armatures have no bearing on the device's rotor or field voltages, and no direct connection therewith, they are properly thought of as being part of neither the Primary Power System or Field Induction System circuits but rather a separate ancillary subsystem intended to optimize the invention's overall performance and efficiency in a number of important ways. To this end, their main purpose is to create and independently control a variable level of both rotor torque and field current rotational force, as necessary or desired.

As depicted in FIG. **14**, the individual stationary armatures **(37)** (or "variable inductors") may be installed in 5-armature groups or array sections within the Neutral Ring of the hull, using notched and mating supports **(151)** which firmly clamp the armatures about a reduced diameter center section of each armature core. These armatures and a basic control circuit arrangement for them are shown in FIG. **16**, with the polarities indicated being those which would result in a clockwise rotation of the rotor as viewed from above. This armature power distribution system may be manually adjusted and/or automatically controlled (using a computer and/or other circuitry).

In both single-stage and three-stage embodiments of the Generator, each of the two electromagnetic armature arrays should be wired in parallel from one or more separate source(s) **(185)** of low-voltage DC current which is/are chassis-grounded and which is/are common to the armatures **(37)** of that array. No variable regulation or biasing of individual armature DC current(s) in single-stage rotor or Thermal Power Unit devices is necessary. The desired individual biasing of armature DC current in three-stage rotor devices, for propulsion-related reasons explained below, may be simply accomplished through a variable resistor **(184)** associated with each armature and its DC power supply subcircuit(s). A minor unidirectional or alternating voltage component of such armature supply current may be added across or in parallel with such resistor **(184)**, using an AC power source **(186)** and/or suitable control interface of a conventional nature, in either single-stage or three-stage Generator embodiments as desired.

Each stationary armature **(37)** incorporates two separate electrical coils **(180)**–**(181)**, to be operated principally on DC voltage, which are formed of conductive insulated magnet wire multi-layer wound upon a single ferromagnetic core. Each such armature core also incorporates a connecting center section or "flux reductor" **(182)** between said separate coils which is of a significantly reduced cross-sectional area with respect to the balance thereof, and which is designed to saturate magnetically when the respective core flux density of either of said coils is approximately equal to or greater than half saturation value. Pure annealed iron or low-carbon steel should be used as the armature core material, due to the high permeability, superior intrinsic induction, and low hysteresis thereof.

The inner or "flux initiator" coils **(181)** of these electromagnetic armatures (or those closest to the rotor) may be used as a means of creating axial magnetic fields of force **(144)** which are perpendicular to the plane of the current-carrying rotor segments **(14)**. Thus, when powered collectively at a common DC current value, the said inner armature coils **(181)** in conjunction with their respective cores may serve as a source of uniform yet variable secondary input rotary torque (or in a torque-assist capacity) for the invention in the course of its normal operation as an electrical generator.

The flux initiator coils **(181)** are so-named because their as-wound polarity relative to the fixed outward direction of the primary DC rotor current establishes both the direction of rotor rotation and the required winding direction and polarity of the outer (and similar) armature coils **(180)**. The said inner coils **(181)** are energized by a variable DC power supply **(185)**, through an output variable resistor **(184)** and a filter diode **(189)** which blocks induced positive AC half-cycles (if any) from crossing the negatively-grounded DC supply **(185)**. Such power supply **(185)** must, however, be sufficiently rated to carry any corresponding induced unidirectional negative current impulses to ground, and may be of a solid-state or rotary induction design.

The axial armature-produced fields **(144)** described may also be individually and variably superimposed on the rotor to develop a selectively controllable force of rotational torque upon its current-carrying segments **(14)**, through variable regulation of the principal DC current in individual armature inner coils **(181)**, such that a locally-unbalanced force of angular acceleration may be applied to the rotor and therefore a directional force of horizontal thrust may indirectly be imparted to the housing for lateral flight maneuvering in propulsive three-stage rotor devices.

The outer coils **(180)** of the stationary electromagnetic armatures, or those closest to the field envelope, may be used as a means of creating axial magnetic fields of force **(144)** which in this case are essentially perpendicular to the described field current arc trajectories. This situation is best illustrated in FIG. **12**, which shows the hull location of the two variable inductor arrays **(145)** (or circular groups of armatures **(37)**). Thus, when powered collectively at a common DC current value (including any minor unidirectional or alternating component thereof), these outer armature coils **(180)** in conjunction with their respective ferromagnetic cores serve to impart an attenuating yet holistic and orderly rotational vector moment upon the external discharge current. The general effect of such applied magnetic rotational force, which may be used to modify or modulate the field's electrodynamic characteristics, is further depicted in FIG. **13**.

Referring again to FIG. **16**, and to continue in the same vein, it is postulated that the quasi-coherent aspects of the field envelope may actually be optimized through the proper control of the amplitude, frequency, and/or phase relationships between any minor AC voltage component of the field envelope current and the said minor AC current (if any) collectively supplied by AC power source **(186)** to the outer or "transflection" coils **(180)** (of each variable inductor array). This consideration applies to both single- and three-stage embodiments of the Generator, although it is decidedly more important in the case of the latter, and may also allow the engineered reduction of field radiative emissions.

These transflection coils **(180)** are so-named because they may be used to impart a variable transverse deflectionary force upon all of the radially-impinging field current electrons. In the region(s) of the field envelope shown in FIG. **11** between the surface of each neutral hull section **(4)** and the corresponding field envelope inner boundary **(142)**, this magnetic force may in fact be large enough to produce a continuous annular flow of "displacement charge" current. In any event, and as shown in FIG. **16**, such outer armature coils **(180)** are also energized by a variable DC power supply **(185)** which is similar in type to that of the inner coils **(181)**, through an output variable resistor **(184)** and filter diode **(189)**. In addition, a variable AC power source **(186)** may impress a minor unidirectional or AC voltage upon the DC current supplied to the said transflection coils **(180)**, either across or in parallel with the output resistor(s) **(184)** of their respective DC power supply subcircuit(s), through separate output variable resistors **(184)** and DC blocking capacitors **(188)** connected to said AC power source.

The axial magnetic fields **(144)** produced by the armatures **(37)** may also be individually and variably superimposed on the field envelope to develop a selectively controllable rotational force upon the described field hemitorus current, through variable regulation of the principal DC current in individual outer (or transflection) coils **(180)**, such that a locally-unbalanced force of angular acceleration may be applied to such hemitorus current. Any such angular force imbalance will cause a corresponding change in the local field current density and therefore a secondary directional force of variable thrust in the z-axis may indirectly be imparted to the housing in propulsive devices, yielding a further measure of navigational attitude control over non-isometric thrust produced by the Generator which may otherwise be largely unilinear.

Given the nature of the flux reductor **(182)** as described above, it can be seen that either the inner or the outer armature core sections may be collectively or individually operated at any flux density above half of saturation level without significantly affecting the operating flux density of the respective opposite core sections, provided both corresponding sets of armature coils (**(181)** and **(180)** respectively) are powered by DC current of the same relative polarity. Thus, the base collective full-power operational level of flux density in the inner armature core sections of a three-stage rotor Impulse Drive Generator should be roughly half saturation level, in order to facilitate the achievement of temporary local rotor torque imbalances for propulsive purposes as described above. Such inner core section base flux density for single-stage rotor or Thermal Power Unit devices may be any desired in a range between 0 and 100% of saturation.

As shown in FIG. **16**, means have nevertheless been provided for reversing the fundamental DC polarity of either the inner **(181)** or outer **(180)** coil of any given armature **(37)**, without changing the given polarity of such coil's DC power supply **(185)**, using a dpdt relay **(190)**. It can be seen that this simple feature allows a counter-polarity magnetomotive force (or counter-mmf) to be applied to either wound core section of any given armature, with respect to the core polarity of that armature's opposite wound section. This feature, when taken in conjunction with the described nature of the flux reductor(s), enables the outer armature core sections in both single- and three-stage devices to be operated at nearly any flux density within their output capability which might be required to optimize the desired quasi-coherent properties of the field envelope, despite the relative polarity and/or flux density of the inner armature core sections. In both single- and three-stage devices, these same two features also allow a substantial level of braking or counter-rotational torque to be applied to the Generator's rotor —by any or all of the inner armature core sections— while maintaining such an optimum level of flux density in the outer armature core sections: even if the two such opposite core sections are operated at opposite respective polarities.

Detailed Calculations for Preferred Embodiment

It is important to note that all of the underlying specifications pertaining to the size and/or positioning of Primary Power System components located within the hull's neutral section region (or "Neutral Ring") are given in terms of fractional increments of the hull radius (R_h) and/or the radial width of the Neutral Ring (C_v–C_a). This means that these components may be easily scaled in exact proportion to the hull, for any selected radius thereof, with no significant error or modification. However, an upper limit of vessel size utilizing the technology represented by the EDF Generator probably exists at a hull diameter not greatly above 100 feet, due to inordinate DC Field Voltage concerns (as that voltage is calculated using the method below).

Table of Dimensions	
Hull Radius = R_h	
Hull Volume Constant (C_v)	$C_v = \frac{1}{5}R_h$
Hull Area Constant (C_a)	$C_a = 0.012919C_v$
Drive Ring Radius (r_f)	$r_f = 2C_v$
Emitter Ring Radius (r_{neg})	$r_{neg} = C_v + C_a$
Neutral Ring Radius (r_{neut})	$r_{neut} = C_v - C_a$
Field Hub Radius (r_z)	$r_z = 3C_v$ (Note: $R_h = r_f + r_z$)
Polar Hull Constant (h_z)	$h_z = 0.1421245r_z$
Polar Volume Differential (X_h)	$X_h = \frac{(r_z^2 - h_z^2)}{2h_z}$
Radial Hull Constant (R_s)	$R_s = h_Z + X_h$
Area of Positive Zone (A_z)	$A_z = 2\pi R_s h_z$
Area of Negative Section (A_n)	$A_n = \frac{s(C + c)}{2}$, where: $s = \sqrt{(C_v + C_a)^2 + (\tan\alpha(C_v + C_a))^2}$, $C = 2\pi R_h$, and $c = 2\pi(R_h - (C_v + C_a))$, and
	Note: $A_z = A_n$.
Volume of Central Section (V_c)	$V_c = \pi r_z^2 h_f$, where $h_f = r_f(\tan\alpha)$, and α is the said Emitter Ring angle.
Volume of Positive Zone (V_z)	$V_z = \frac{2}{3}\pi R_s^3 - \pi\frac{R_s^2 X_h - X_h^3}{3}$
Vol. of Drive Ring Section (V_f)	$V_f = \frac{\pi h_f(R_h^2 + R_h r_z + r_z^2)}{3} - \pi r_z^2 h_f$
	Note: $V_z = V_f$
Total Hull Volume (V_t)	$V_t = 2(V_z + V_f) + 2V_c$

The variable 'α' equals the displacement angle of the Emitter Ring negative sections (with respect to the hull's horizontal centerline plane), which is preferably 7.5° to 8° each.

Field and Rotor Voltages

Part A: Model Unit Field Intensity

[1] The Marginal Envelope DC Field Intensity for all EDF Generator models regardless of type should be equal to the breakdown dielectric strength of air or vacuum at 3 million (3×10^6) volts/meter, as measured along a full pure semicircular arc drawn from the center point on the surface of either hull Positive Zone (**5**) to any exterior point on the vessel's design hull configuration which lies on the horizontal centerline plane thereof. This distance shall be referred to as the "Drive Field Perimeter" (**141**), as shown in FIG. **11**, and is taken to represent the longest continuous DC field current trajectory.

[2] For purposes of these Specifications, the "Drive Field Boundary" (**142**) shall represent the distance measured along a full pure semicircular arc from any point on the outer peripheral edge of either hull Positive Zone (**5**) to the closest point thereto which lies on the inner peripheral edge of the vessel's corresponding negative hull section (**3**) (as shown in FIG. **11**), and is taken to represent a shortest such DC drive current trajectory.

Part B: Thermal Power Units

[1] The Nominal Field Voltage (nom.V_f) should be numerically equal to 1.5 million (1.5×10^6) times the vessel R_h as measured in feet. This value is also equal to the Thermal Unit's standard design Primary Voltage Expansion Ratio of 750 times its Specification DC Primary Array Voltage, as a function of the charge stored within the Primary Power System's two Field Ballast Capacitors and of the Unit's primary cathode emissivity.

[2] The Specification Field Voltage (spec.V_f) should be equal to the Marginal Envelope DC Field Intensity times the distance comprising the Drive Field Perimeter, or (as presently calculated under the above directive) 0.9666 times the Nominal Field Voltage. This value shall constitute the Unit's standard operating value of field envelope voltage.

[3] The Peak (design operating maximum) Field Voltage (max.V_f) shall be equal to 110% of the Specification value thereof, and should never be exceeded in operation.

[4] The Nominal Rotor Voltage (nom.V_r) would be numerically equal to one-third the product of 1,000 (1×10^3) times the vessel R_h, as measured in inches, as a gauge of the basic full-power DC generator value of rotor rotational speed, in a 'single-stage' device.

[5] The Specification Rotor Voltage (spec.V_r) should be numerically equal to 0.25776% of 1.5 million (1.5×10^6) times the vessel R_h as measured in feet, and is also equal to 0.9666 times the Nominal Rotor Voltage. This value would constitute the standard operating value of DC rotor voltage, as a function of the engineered Primary Power System output at a nominal design rotor rotational speed.

[6] The Peak (design operating maximum) Rotor Voltage (max.V_r) shall be equal to 110% of the Specification value thereof, and should likewise never be exceeded.

[7] The Specification DC Primary Array Voltage (spec.ΔV_p) will be equal to half (½) of the Specification Rotor Voltage, and affects all primary array electrode spacings.

Part C; Impulse Drive Units

Unlike the case with Thermal Power Units, it is not possible to specify in one step a Nominal Field Voltage for Impulse Drive Units which is a linear function of the hull radius R_h; the DC Field Voltage which must be produced to achieve the chosen design goal for net linear thrust output, equal to five-thirds the Earth's pull of gravity, is jointly proportional to the rotor ampacity which rises as a function of the square of increases in R_h and to the vessel weight (which rises as a function of the cube of R_h increases).

Therefore, a Marginal Field Voltage is initially to be computed for Impulse Drive Units which reflects a projected theoretical operating value of DC Field Voltage required to produce a vessel acceleration approximately equal to 1.67 'g' outside of a proximal gravity field. This Marginal Field Voltage is to be calculated using the "specific impulse" of the vessel concerned (in nt-sec), and for purposes of the application contemplated herein this term is hereby defined as being equal to five times the vessel weight (in newtons) divided by the rotor ampacity times the number of electrons per coulomb.

[1] Accordingly, the Marginal Field Voltage shall then be equal to the Unit's specific impulse (F_{dt}) times the speed of light (c) divided by the unit electron charge (q).

[2] An estimated nominal Field Voltage will also be computed whereby a nominal value for the final electron speed of Drive Field current is assigned which is in fact a linear function of R_h, as applicable to vessels of from 4 to 100 ft. in diameter only.

[3] The Nominal DC Field Voltage (nom.V_f) should then be equal to the simple average of the Marginal Field Voltage and the said estimated nominal Field Voltage.

[4] The Specification Field Voltage (spec.V_f) should equal the Nominal DC Field Voltage (nom.V_f) divided by an assigned engineering design constant equal to 0.982826.

[5] The Peak (design operating maximum) Field Voltage (max.V_f) shall be equal to 110% of the Specification value thereof, and should never be exceeded in operation.

[6] The Nominal Rotor Voltage (nom.V_r) would be numerically equal to one-fourth the product of 1,000 (1×10^3) times the vessel R_h as measured in inches, as a gauge of the basic full-power DC generator value of rotor rotational speed, in a 'three-stage' device.

[7] The Specification Rotor Voltage (spec.V_r) should be numerically equal to 0.19332% of 1.5 million (1.5×10^6) times the vessel R_h as measured in feet, and is also equal to 0.9666 times the Nominal Rotor Voltage. This value would constitute the standard operating value of DC rotor voltage, as a function of the engineered Primary Power System output at a nominal design rotor rotational speed.

[8] The Peak (design operating maximum) Rotor Voltage (max.V_r) shall be equal to 110% of the Specification value thereof, and should likewise never be exceeded.

[9] The Specification DC Primary Array Voltage (spec.ΔV_p) will be equal to half (½) of the Specification Rotor Voltage, and affects all primary array electrode spacings.

[10] The design Primary Voltage Expansion Ratio for each Impulse Drive Unit should then be equal to the ratio of Specification Field Voltage to Specification DC Primary Array Voltage, and is once again a function of the ballast capacitor charge and of the primary cathode emissivity as in the case with Thermal Power Units.

Example of Specific Unit Voltage Values

Four-foot diameter model

Section A: Thermal Power Units

A 4 ft. EDF Generator constructed for use as a Thermal Power Unit would have specific Rotor and Field Voltages, according to the preceding instructions, as follows:

[1] The Nominal Rotor Voltage (nom.V_r) will be equal to 8,000.

[2] The Specification Rotor Voltage (spec.V_r) will be equal to 7,732.8.

[3] The Spec. DC Primary Array Voltage (spec.ΔV_p) will be equal to 3,866.4.

[4] The Nominal Field Voltage (nom.V_f) will be equal to 3,000,000.

[5] The Specification Field Voltage (spec.V_f) will be equal to 2,899,800.

The corresponding Primary Voltage Expansion Ratio (for this and any single-stage Thermal Power Unit) will thus be 750:1, and in this case is equal to 2,899,800/3,866.4.

Section B: Impulse Drive Units

A 4 ft. diameter three-stage-rotor EDF Generator constructed for use as an Impulse Drive Unit would have specific Rotor and Field Voltages, according to the preceding instructions, as follows:

[1] The Nominal Rotor Voltage (nom.V_r) will be equal to 6,000.

[2] The Specification Rotor Voltage (spec.V_r) will be equal to 5,799.6.

[3] The Spec. DC Primary Array Voltage (spec.ΔV_p) will be equal to 2,899.8.

[4] Marginal Field Voltage: The 4 ft. Drive Unit's Marginal Field Voltage, which represents a projected minimum standard operating value of field envelope voltage, is based upon the vessel's specific impulse (F_{dt})—given its calculated design rotor ampacity of I_{max}=38,160 amps and estimated weight of 79.2 lbs/ft^3—which is computed as follows:

$$F_{dt}=\frac{5(9.8\ \text{m/sec}^2)V_t(79.2\ \text{lbs/ft}^3)}{(3.816\times10^4\ \text{coul/sec})(6.25\times10^{18})(2.2\ \text{lbs/kg})}$$

where V_t=2.5068 ft, as calculated by the formulas in the Table of Dimensions above using an Emitter Ring displacement angle (α) of 7.5°. Therefore,

$$F_{dt}=\frac{5(9.8\ \text{m/sec}^2)(90.245\ \text{kg})}{23.85\times10^{22}\ \text{electrons/sec}}=1.8541\times10^{-20}\ \text{nt-sec/electron}^*$$

*This impulse represents the instantaneous force of thrust produced by each Field current electron's collision with the Field Hub (equivalent here to: kg-m/sec/electron).

Therefore, the Marginal Field Voltage is equal to:

$$\frac{F_{dt}c}{q}=\frac{(1.8541\times10^{-20}\ \text{nt-sec})(299.7925\times10^6\ \text{m/sec})}{1.6\times10^{-19}\ \text{coul}}=34.7403\times10^6\ \text{volts (nt-m/coul)}$$

[5] Estimated Nominal Field Voltage: The 4 ft. Impulse Drive Unit's estimated nominal field voltage, which represents an alternate but extremely accurate projection of

the minimum standard operating value of field envelope voltage, is based upon a nominal value for the final electron speed of Drive Field current (ΔV_e) which is an assigned linear function of the vessel R_h and is calculated as follows:

$$\Delta V_e=[0.999+(2R_h-4)(1.03125\times10^{-5})]c,$$

where R_h is in feet, for vessels from 4 to 100 ft. in diameter.

Therefore, $\Delta V_e=[0.999+0(1.03125\times10^{-5})]c=0.999$ c.

The purpose of this text subsection is to verify that a goal-requisite level of Field current thrust may be produced by the Electrodynamic Field Generator, at an operating DC Field Voltage approximately equal to that stated above, within the ampacity of the rotor and at an assigned value of Field current electron velocity (for this model) equal to 0.999c. In this manner, it is hoped that unnecessarily high operating Drive Field Voltage may be avoided, by virtue of design (due to the finite level of unavoidable uncertainty in the relativistic values involved), without sacrificing device performance.

(a) The amount of thrust (force) required to produce an acceleration of 1 g, thus exactly compensating the weight of the Impulse Drive Unit or its normal acceleration due to Earth's gravity, may be calculated by Newton's Second Law as follows:

$F=mg$, where m=est. design mass* of 90.245 kg, and $g=9.8$ m/sec^2.

*Note: The estimated design weight of any given Impulse Drive Unit is 79.2 lbs/ft^3.

Thus, F=884.4 nt.

(b) Since impulse equals change of momentum, the electron thrust developed by the Drive Field equals the total relativistic mass of Field current times the incident current velocity (as the final electron velocity equals zero). The amount of impulse thrust which is equivalent to 1 g in this example is 884.4 nt.

Therefore, let $\Sigma M_t \Delta V_e=884.4$ nt.

(c) The design goal for net linear thrust output of any given Impulse Drive Unit is 16.333 m/sec^2 at the minimum standard operating value of Field Voltage and rated rotor ampacity (I_{max}) Therefore, given the considerations of Field Current Bias values (as were outlined in an earlier section), total isometric thrust at Marginal Field Voltage must equal 5 g. The largest current available to effect a Field impulse equivalent to 1 g is then equal to:

$$I_g=I_{max}/5=7{,}632 \text{ amps.}$$

(d) Since Σ equals the total number of electrons comprising a Drive current of I_g, we find:

$$\Sigma=I_g(6.25\times10^{18})=4.77\times10^{22} \text{ electrons/sec.}$$

(e) Therefore, from [b] above:

$$884.4\text{nt}/\Sigma=M_t\Delta V_e, \text{ and } M_t\Delta V_e=1.8541\times10^{-20} \text{ kg-m/sec/electron.}$$

(f) Letting $\Delta V_e=0.999c$, where $c=299.7925\times10^6$ m/sec, we calculate $M_t=619.08\times10^{-31}$ kg, or about 67.96 m_o, where m_o=the electron rest mass of 9.11×10^{-31} kg.

(g) If m_i equals the relativistic mass equivalent of the gained kinetic energy of each Drive Field electron, then $M_t=m_o+m_i$.

Therefore, $m_{i=M_t}-m_o=619.08\times10^{-31}-9.11\times10^{-31}=609.97\times10^{-31}$ kg.

(h) Each Drive Field electron's gained kinetic energy (E_q) equals $m_i c^2$. Using a precise value of 8.98755×10^{16} for c^2, we obtain $E_q=(6.0997\times10^{-29})(8.98755\times10^{16})=54.8214\times10^{-13}$ joules.

(i) Here, E_q also equals $W=q(\text{est.}V_{nf})$, where q=electron chg. @1.6×10^{-19} coul., and est.V_{nf}=est. nominal Field Voltage.

Therefore, $E_q/q=\text{est.}V_{nf}$, and est.$V_{nf}=34.2633\times10^6$ volts (jouleslcoul).

Thus, we find that the estimated nominal Field Voltage required to develop requisite thrust is in fact well within the thrust goal parameters discussed above, at 98.63% of the projected Marginal value (calculated in the preceding subsection B[4]).

[6] The Nominal DC Field Voltage (nom.V_f) should therefore be equal to the simple average of the Marginal Field Voltage projected earlier and the estimated nominal Field Voltage Oust computed), at 34.5018×10^6 volts.

[7] The Specification Field Voltage (spec.V_f) should then be equal to the Nominal DC Field Voltage divided by the said design constant 0.982826, at 35.1047×10^6 volts. This value shall constitute the Unit's standard operating value of field envelope voltage.

[8] The design Primary Voltage Expansion Ratio for this model Impulse Drive Unit is therefore equal to $(35.1047\times10^6)/2{,}899.8=12{,}106{:}1$, as was referred to earlier.

From the foregoing description it will be observed that the present invention provides a remarkable level of electrical and thermal output in a lower-voltage corona discharge induction embodiment, which is intended for use in a utility or physical plant application. However, this invention also provides the potential for an appreciable level of net impulse thrust in a higher-voltage arc discharge induction embodiment intended as an electrically-propulsive aerospace vessel. It can be seen as well that the Electrodynamic Field Generator and the electromotive field of force produced thereby will moreover fulfill the stated objective of NASA's premier Breakthrough Propulsion Physics Program: to discover a method whereby "a vehicle can create and control an asymmetric force on itself without expelling a reaction mass" and which satisfies conservation laws in the process. Such a vessel could perhaps be made to attain a velocity commensurate with its collector field electron collision velocity, and affords hope that true interstellar travel might therefore become possible as well. While there is due cause for some concern as to the EDF Generator's production of high-frequency radiative emissions coincident with its production of heat and impulse thrust using the new means described, it is postulated that the electromagnetic characteristics of the device's field envelope can be suitably tuned in the manner(s) oulined hereinabove such that the field envelope is rendered largely opaque to such very-high-energy emissions—by promoting continuous and adsorptive Compton effect interactions thereof with the field's impinging drive current electrons. When used in association with a ground-based utility or physical plant, a non-propulsive thermal and/or electrical power output model of the Generator may also be enclosed within a suitable Faraday "cage" or shielding structure to further minimize radiative emissions which are undesirable. Neverthless, the said field envelope tuning principles (related to controlling the amplitude, frequency, and/or phase relationships amid field AC voltage components) may moreover permit the development of the Electrodynamic Field Generator as a novel signal communications means. In pursuit of a new and more complete quantum gravity theory, experiments may be conducted involving the gravimetric coupling of two or more separate such devices: whose field envelopes will by design be able to exhibit a relativistic and oscillatory apparent mass effect that is artificially generated—at desired resonant frequencies—by using alternating electromagnetic induction methods of the type described hereinabove to modulate electrodynamic field properties which are normally of constant polarity and magnitude or random frequency.

Accordingly, it is to be understood that the embodiments of the invention herein described are merely illustrative of the application of the principles of the invention. Reference herein to details of the illustrated embodiments are not intended to limit the scope of the claims, which themselves recite those features regarded as essential to the invention.

What is claimed is:

1. An electrodynamic field generator comprising:

(a) a disc-shaped hull having upper and lower surfaces separated by a distance at an axis of the disc and tapering to an edge at a perimeter of the disc, enclosing a volume therebetween, the upper and lower surfaces each being divided into a positive section surrounding the axis of the disc, a negative section around the periphery of the disc, and a neutral section therebetween; the positive and negative section surfaces being of conductive material and the neutral section surfaces being of insulative material;

(b) a rotor having an outer circumference, an inner circumference surrounding an open central aperture, and an annular body therebetween, with the rotor body having flat upper and lower surfaces; the rotor being rotatably mounted within the volume of the hull and centered upon the axis thereof, an inner diameter and outer diameter of the rotor being chosen such that the rotor extends across at least all of the neutral portion of the hull volume and at least part of each positive hull section;

(c) an evacuated induction compartment and peripheral space charge chamber comprising at least the portion of the hull volume within which the rotor is mounted;

(d) a principal generating system and corresponding rotor electrical circuit comprising two assemblies, one associated with the upper surface of the rotor and one with the lower surface thereof, each assembly in turn comprising:

  (i) an outer induction electrode array comprising a flat cathode ring mounted upon and electrically coupled to the rotor, adjacent to an outer circumference thereof, and a flat anode ring mounted plane-parallel to, but insulated from, the cathode ring;

  (ii) an inner induction electrode array comprising a flat anode ring mounted upon and electrically coupled to the rotor, a flat cathode ring, and a grid ring element, the cathode and grid being mounted plane-parallel to, but insulated from, the anode and each other; the inner induction ring array being mounted to the rotor closer to the inner circumference thereof than the outer induction ring array;

  (iii) at least one generating coil mounted to the rotor and extending toroidally around the rotor concentric to either circumference thereof, located between the outer induction ring array and the inner induction ring array, and electrically coupled to the cathode of the inner induction array and to the anode of the outer induction array; and

  (iv) a plurality of stationary permanent magnets mounted to the hull, such that the magnetic field from each of the permanent magnets intercepts a generating coil; such that when the rotor is rotated, the magnetic fields from the stationary magnets induce an electrical voltage in each generating coil which results in a DC rotor circuit output current;

(e) a field induction system comprising two assemblies, one associated with the upper surface of the rotor and one with the lower surface thereof, each assembly in turn comprising:

  (i) at least one plane-parallel electrode array comprising:

    1) a primary anode ring mounted to the rotor adjacent to the inner circumference thereof, between the inner circumference and the inner induction ring array;

    2) at least one primary cathode mounted to, but electrically separated from, each positive section of the hull; and

    3) at least one stationary grid element located between each primary cathode and the corresponding primary anode ring;

  (ii) a primary cathode resistor electrically coupled between each primary cathode and the corresponding positive section of the hull, each cathode resistor being encapsulated by a dielectric buffer; and

  (iii) a principal hull coolant and heat transfer system, thermally coupled to each primary cathode resistor;

(f) at least one housing charge ballast capacitor, electrically coupled to the principal electrical generating system circuit and to the rotor;

(g) a plurality of field emitters which are mounted around the outer circumference of the rotor and electrically coupled thereto; and

(h) a motor drive mounted to the hull, having a mechanical output coupled to the rotor for rotation thereof.

2. The electrodynamic field generator of claim **1**, in which:

(i) the principal electrical generating circuit induces a primary DC voltage between the outer circumference of the rotor and the inner circumference thereof when the rotor is rotated by the motor drive, resulting in a unidirectional rotor current which distributes electron charge from the outer rotor circumference to the negative hull sections; and

(ii) ii) the field induction system utilizes a portion of the primary DC rotor voltage to establish an external discharge current which expresses a very large expansion of that voltage, within the electric field between the negative and positive sections of the hull, the measure of the expanded field voltage being defined by an engineered instantaneous charge differential between the rotor field emitters and the primary cathodes which is supported by the electric force in one or more housing charge ballast capacitors.

3. The electrodynamic field generator of claim **1** in which the rotor comprises a plurality of conductive radial segments and the same number of nonconductive segment separatorss, with the individual segments and separators uniformly alternating in placement.

4. The electrodynamic field generator of claim **3**, in which the conductive rotor segments are made of copper.

5. The electrodynamic field generator of claim **3**, in which each generating coil is mounted to the segment separators.

6. The electrodynamic field generator of claim **3** in which the anode of each outer induction array of the principal generating system is mounted to the segment separators, as are the cathode and grid element of each inner induction array.

7. The electrodynamic field generator of claim **1** in which each inner induction array constitutes a triode vacuum-tube construct, the control grid voltage of which biases that array nearly to current cutoff.

8. The electrodynamic field generator of claim **1** in which each of the two assemblies of principal generating system and corresponding rotor electrical circuit comprise a set of three generating coils, arranged across the annular body of the rotor, with each set of coils being sequentially coupled electrically.

9. The electrodynamic field generator of claim **1** in which each generating coil is compound wound, having primary and bias windings.

10. The electrodynamic field generator of claim **1**, in which each generating coil further comprises a nonmagnetic core.

11. The electrodynamic field generator of claim **1**, in which each generating coil further comprises a powdered ironi core.

12. The electrodynamic field generator of claim **1** in which the stationary permanent magnets are C-shaped, and are arranged into a circular body with a flux gap and wherein the magnetic field passes through the flux gap, and the generating coils are positioned to pass through the flux gaps of the magnets in a uniform manner during rotation of the rotor.

13. The electrodynamic field generator of claim **1**, in which the magnets are made of an alnico family alloy.

14. The electrodynamic field generator of claim **1** in which a housing charge ballast capacitor is mounted upon the rotor between each outer induction array and the rotor's outer circumference, each ballast capacitor having an inner conductive ring electrically connected to the rotor, an outer conductive ring plane-parallel to the inner ring and electrically connected to the outer induction array anode, and a dielectric layer between its inner and outer rings.

15. The electrodynamic field generator of claim **1** in which the field emitters are made of a sintered refractory composite material which is electrically conductive, and have a shape such that they taper from an inner end or surface having a thickness equal to a thickness of the rotor's outer circumference to a slightly blunted edge at an outer end forming an outermost circumference of the rotor and field emitters.

16. The electrodynamic field generator of claim **9** in which each assembly of the principal generating system and corresponding rotor electrical subcircuit comprises one or more AC subcircuitry stages, each stage being mounted to the rotor for rotation therewith and in turn comprising:

an AC input, or grid resistor;

one or more bias windings electrically coupled to the AC input;

a triode array electrically coupled to the AC input and to that input's corresponding bias winding or windings, comprising a cathode, an anode, and a control grid; and

an output, or plate resistor, electrically coupled to the anode of that triode array or to the anode of a successive electrode array.

17. The electrodynamic field generator of claim **16** in which a plurality of AC stages are electrically coupled in series, the input of an outermost stage closest to the outer circumference of the rotor being coupled to the outer induction array anode and the output of an innermost stage closest to the inner rotor circumference being coupled to the control grid of the inner induction array.

18. The electrodynamic field generator of claim **17** in which each assembly of the principal generating system, and the corresponding rotor electrical subcircuit, comprises three AC stages.

19. The electrodynamic field generator of claim **1** in which the plane-parallel electrode arrays comprising the field induction system each in turn comprise a primary cathode, three grid elements, and an anode ring and therefore constitute pentode vacuum-tube constructs.

20. The electrodynamic field generator of claim **1**, in which a voltage control system is electrically or thermally coupled to at least one stationary electrode of each plane-parallel array comprising the field induction system such that the total discharge current is limited to a value within the total ampacity of the rotor segments.

21. The electrodynamic field generator of claim **20**, in which the operating electrode temperature is controlled by movable thermal adjusting rods that may be thermally coupled to one or more of the cathodes or grids and to a cryogenic heat sink.

22. The electrodynamic field generator of claim **20**, in which the voltage control system is further utilized to apply inverse AC signal feedback to the discharge current such that rotor circuit AC voltage amplitude is limited to a desired or requisite value.

23. The electrodynamic field generator of claim **1** in which each positive section of the hull is divided into a plurality of radial sectors, each sector being separated from adjoining sectors by an insulating partition.

24. The electrodynamic field generator of claim **23** in which each assembly of the field induction system comprises a plurality of plane-parallel electrode arrays, with each array being electrically coupled to a radial sector of a positive hull section.

25. The electrodynamic field generator of claim **23**, in which each positive hull section is divided into 36 radial sectors.

26. The electrodynamic field generator of claim **23** further comprising a center sector of each positive hull section, located at the axis of the hull, these sectors each being electrically isolated from the radial sectors of the respective positive hull section.

27. The electrodynamic field generator of claim **26**, in which each center sector is made of an insulating material.

28. The electrodynamic field generator of claim **26**, in which single radial sectors at four cardinal points of each positive hull section are electrically connected in parallel with a conductive center sector of that positive hull section.

29. The electrodynamic field generator of claim **1**, in which the principal hull-coolant system comprises one or more primary thermal conduits conveying a high-temperature heat transfer fluid and incorporates a heater to pre-heat that heat transfer fluid and, in turn, the cathode resistors, dielectric buffers, and primary cathodes.

30. The electrodynamic field generator of claim **29**, in which the high-temperature heat transfer fluid is liquid sodium.

31. The electrodynamic field generator of claim **1** in which the primary cathode material is selected from a group consisting of pure tungsten metal and tungsten, as well as alloys thereof, which is or have been impregnated with thorium.

32. The electrodynamic field generator of claim **1**, in which the primary cathode resistors are made of a ceramic material which becomes electrically conductive at an elevated operating temperature.

33. The electrodynamic field generator of claim **32**, in which the ceramic cathode resistor material is selected from a group consisting of the steatite and cordierite families of compounds.

34. The electrodynamic field generator of claim **1**, in which the dielectric buffer encapsulating each primary cathode resistor is made of sodium bismuth titanate.

35. The electrodynamic field generator of claim **1**, further comprising a secondary hull coolant or heat transfer system which in turn comprises one or more secondary thermal conduits conveying a heat transfer fluid and which is thermally coupled to the neutral and negative sections of the hull.

36. The electrodynamic field generator of claim **35**, in which the heat transfer fluid is a cryogenic coolant.

37. The electrodynamic field generator of claim **36**, in which the heat transfer fluid is a liquefied gas selected from the group consisting of liquid air, liquid nitrogen, and liquid helium.

38. The electrodynamic field generator of claim **1** further comprising a plurality of stationary electromagnetic armatures arranged in two circular groups or arrays within the hull volume, one array above the upper surface of the rotor and one below the lower surface thereof, each armature comprising:

a ferromagnetic body or core having a first end section adjacent to the rotor, a second end section adjacent to a neutral section of the hull, and a length therebetween which is of a reduced cross-sectional area with respect to either end section; and

at least one electrical winding wound on either end section of the core, wherein an electric current in the winding induces a magnetic field in the core which is superimposed on the rotor or on the external electric field.

39. The electrodynamic field generator of claim **38** in which each of the armatures has two windings, comprising:

a first winding on the first end section, wherein a current in the first winding causes a magnetic field to be superimposed on the rotor; and

a second winding on the second end section, wherein a current in the second winding causes a magnetic field to be superimposed on the external field.

40. The electrodynamic field generator of claim **38**, in which the armature core material is selected from a group consisting of pure annealed iron and low-carbon steels.

41. The electrodynamic field generator of claim **38**, further comprising a DC power supply electrically coupled to at least one winding of each stationary armature.

42. The electrodynamic field generator of claim **38**, further comprising an AC power source electrically coupled to at least one winding of each stationary armature.

43. The electrodynamic field generator of claim **1**, in which the mechanical output of one or more motor drives is coupled to the rotor through a gear on each motor drive's mechanical output meshing with one or more driven ring gears about the inner circumference of the rotor.

44. The electrodynamic field generator of claim **1** further comprising a pair of rolling bearing assemblies or ballraces for supporting the rotor, one around the upper surface of the rotor and one around the lower surface thereof with both ballraces being mounted near the rotor's outer circumference; each ballrace comprising a first race half or raceway mounted on the rotor and a second raceway mounted to the hull opposite the first raceway, defining a bearing ball passage therebetween, and a plurality of bearing balls contained in the ball passage.

45. The electrodynamic field generator of claim **1**, further comprising a central control cabin inside the hull within the aperture or inner circumference of the rotor.

46. The electrodynamic field generator of claim **1**, in which the cathode resistors each comprise a fixed positive reduced-voltage tap whereby an onboard or external electrical load may be connected between the fixed taps and the primary cathodes.

47. The electrodynamic field generator of claim **1**, in which the heat transfer system is thermally coupled to an external heat exchanger.

48. The electrodynamic field generator of claim **1**, comprising one or more superconductive current storage rings utilized to deposit and retrieve an expendable store of ambient electron charge during extended operation in the vacuum of space.

* * * * *

US006498291B2

(12) **United States Patent**
Brammer

(10) **Patent No.: US 6,498,291 B2**
(45) **Date of Patent: Dec. 24, 2002**

(54) **STATIC ELECTRICITY ELIMINATOR**

(76) Inventor: **James William Brammer**, 2338 Catawba Rd., Roanoke, VA (US) 24175

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **10/050,698**

(22) Filed: **Jan. 18, 2002**

(65) **Prior Publication Data**

US 2002/0070039 A1 Jun. 13, 2002

Related U.S. Application Data

(63) Continuation-in-part of application No. 09/274,355, filed on Mar. 11, 1999, now abandoned.

(51) **Int. Cl.**^{7} **H02G 13/00**

(52) **U.S. Cl.** .... **174/3**; 174/5 SG; 174/6

(58) **Field of Search** .... 174/1, 2, 3, 5 R, 174/5 SG, 6, 7; 52/723.1, 723.2, 724.5, 729.5

(56) **References Cited**

U.S. PATENT DOCUMENTS

110,778	A		1/1871	Munson	
160,151	A		2/1875	Bryan	
362,062	A		5/1887	Cole	
433,459	A		8/1890	Ferrell	
505,106	A		9/1893	Hodges	
708,188	A		9/1902	Wilson	
796,760	A		8/1905	McCullough	
873,154	A		12/1907	Miller	
1,904,773	A		4/1933	Aanenson	
2,111,799	A		3/1938	Newpher	
4,180,698	A		12/1979	Carpenter	
4,540,845	A		9/1985	Frola	
4,565,900	A	*	1/1986	Arnau	174/3
4,926,284	A		5/1990	Guegenio	
4,994,629	A		2/1991	Tatum et al.	
5,073,678	A		12/1991	Carpenter, Jr.	
5,335,160	A	*	8/1994	Savoca	174/45 R
5,657,197	A		8/1997	Skinner, II et al.	
5,998,731	A		12/1999	Takamura	
6,121,543	A		9/2000	Hallmark	

* cited by examiner

Primary Examiner—Chau N. Nguyen
Assistant Examiner—Adolfo Nino
(74) *Attorney, Agent, or Firm*—Rodney A. Corl

(57) **ABSTRACT**

When static electricity builds up in the earth, it is suddenly discharged in a destructive bolt of electricity which can cause fires, property damage and loss of life. The present invention collects this static electricity from the earth before it can build up, and dissipates it to the atmosphere gradually so as to prevent fires, property damage and loss of life. A plurality of relatively long underground wires conduct the electricity from the ground to a metal pole which extends from the ground upwardly approximately 40 feet or more, and the pole dissipates the electricity to the atmosphere. In one embodiment, a U-shaped top fixture on the pole increases the dissipation of electricity to the atmosphere.

7 Claims, 2 Drawing Sheets

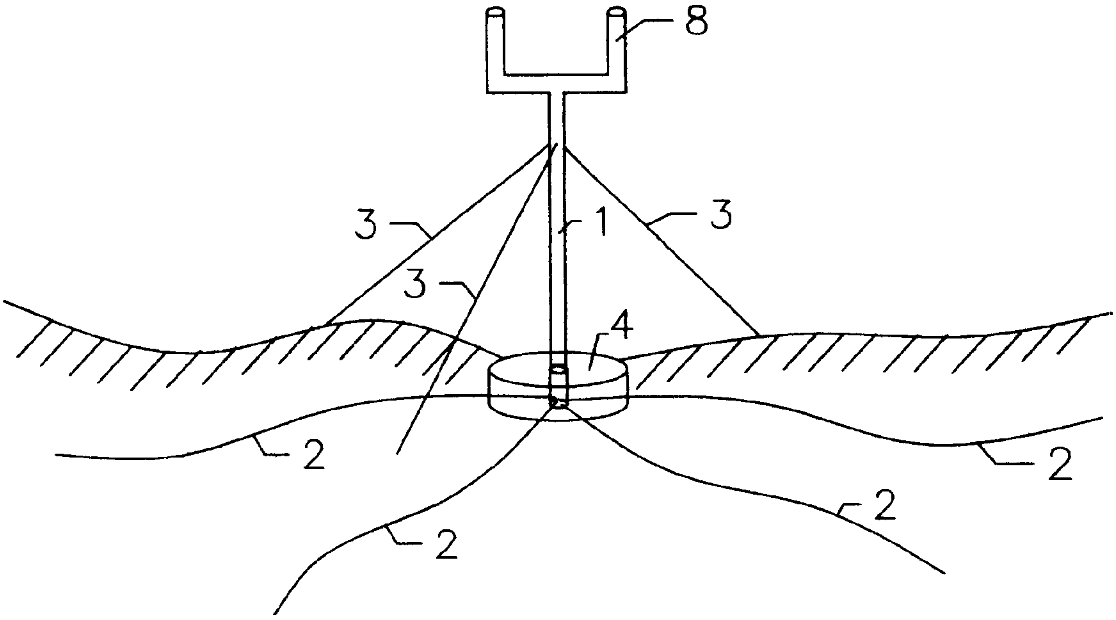

Fig. 1

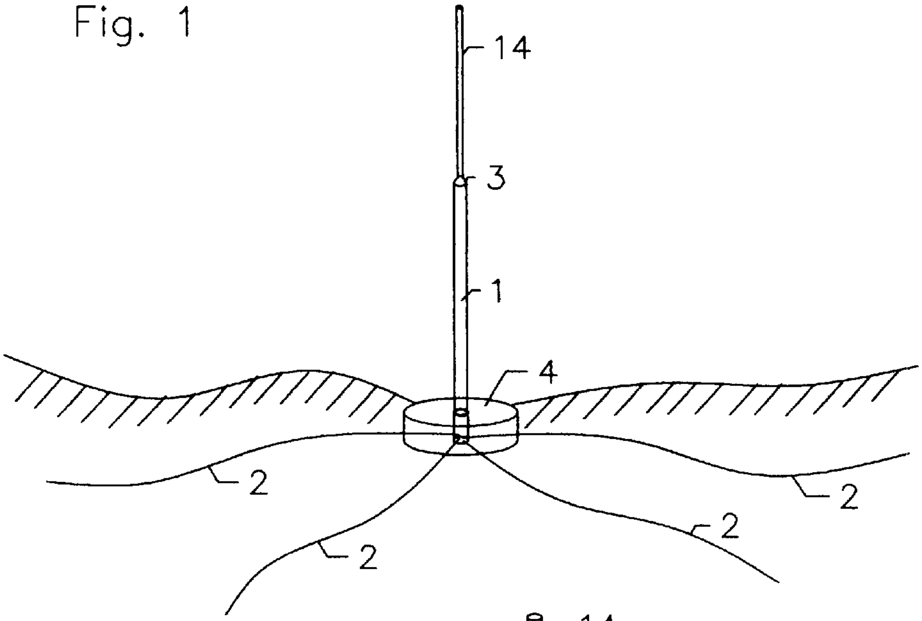

Fig. 2

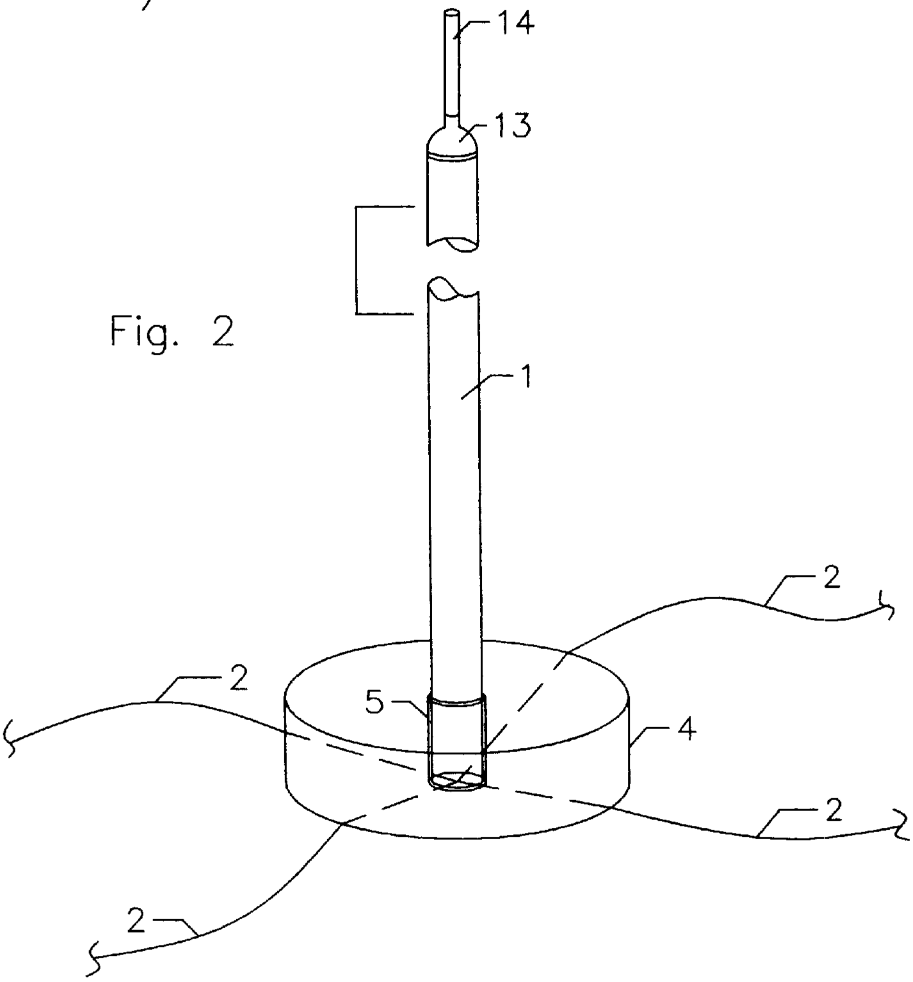
14
13
1
2
2
5
4
2
2

Fig. 3

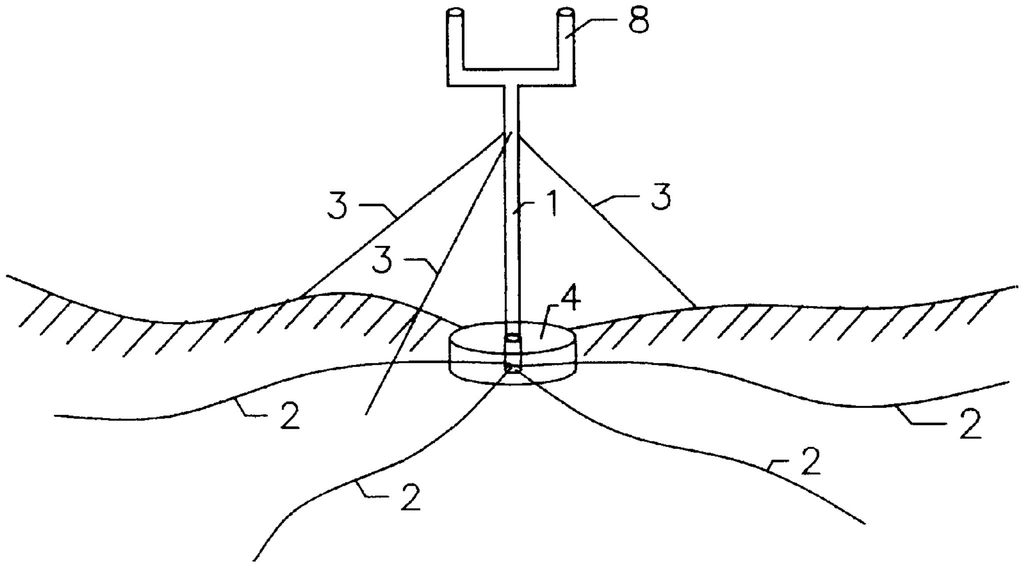
8
3
1
3
3
4
2
2
2
2

Fig. 4

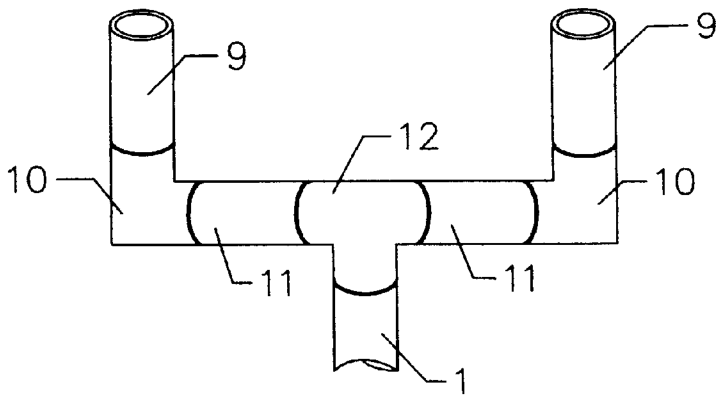
9
9
12
10
10
11
11
1

STATIC ELECTRICITY ELIMINATOR

RELATED APPLICATIONS

This is application is a continuation-in-part of application Ser. No. 09/274,355 filed on Mar. 11, 1999, now abandoned, entitled LIGHTNING RETARDER.

BACKGROUND OF THE INVENTION

The purpose of this invention is to reduce or eliminate lightning strikes, which cause fires. Lightning occurs when there is a buildup of static electricity, which suddenly discharges in the form of a bolt of electricity. If this buildup can be reduced or eliminated then there will be no sudden discharge and no lightening bolt. This will result in a reduction in fires and the resulting damages to property and loss of life.

While there have been "lightning rods" around for many years, the device of the present invention is not a lightning rod and does not work in the same manner as a lightning rod. A lightning rod works by attracting the sudden discharge of electricity in the form of a lightning bolt and dissipating this bolt of electricity thereby preventing the bolt from striking nearby objects. The static eliminator of the present invention works in an almost opposite manner. By constantly dissipating the static electricity to the atmosphere, there is never a buildup of charge, and consequently, no sudden discharge and hence no lightning bolt or associated fires and damage.

BRIEF SUMMARY OF THE INVENTION

The static eliminator of the present invention physically resembles a lightning rod, but due to some subtle differences, acts in a very different manner. The static eliminator consists of a long metal pole affixed in the ground and extending to a substantial height above ground level. Conductive wires are attached to the base of the pole and extend a large distance in several directions from the base of the pole and are buried in the ground. These wires accumulate the static charge from the earth, conducting it to the pole which in turn bleeds the static electricity to the atmosphere

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be further described in connection with the accompanying drawings, in which:

FIG. **1** is a perspective view of one embodiment of the invention,

FIG. **2** is a detailed view of the base of the invention,

FIG. **3** is a perspective view of another embodiment of the invention, and

FIG. **4** is a detailed view of the upper end of the pole of this other embodiment.

DETAILED DESCRIPTION

In one embodiment, shown in FIG. **1**, the static eliminator of the present invention consists of a pole **1**, preferably made of steel pipe with a 1 ½ inch diameter, extending upwardly from the ground **7** approximately 20 feet. A reducer coupling **13** is connected at the top of the 1 ½ inch pole and another pipe **14** of ½ inch diameter is connected to the coupling and extends further upwardly approximately 20 more feet. The pipe **1** is placed in a hole in the ground approximately 3 feet deep and a concrete collar **4** is poured around the pole **1** to anchor it in the ground. Before pouring the concrete, wires **2** are attached around the base of the pole **1** by any suitable means such as screws or clamps. The wires **2** are preferably made of 12 gauge or larger copper and each extends outwardly from the pole 300 feet or farther. The wires **2** are buried in the ground **7** at a depth of at least 15 inches. It should be noted that pole **1** is not attached to any structure in order for it to function with maximum efficiency. In fact, pole **1** is as far from any surrounding structures as physically practical.

In FIG. **2**, details of the invention can be seen. Pole **1** is surrounded by an insulating sleeve **5** of any electrically insulating material such as glass or plastic. Sleeve **5** extends from the bottom of pole **1** to the top of concrete collar **4** or above ground level, whichever is higher. Wires **2** are attached at the base of the pole **1** and extend outwardly therefrom, through the concrete collar **4**. Reducer **13** is attached at the top of pole **1** and pole **14** of reduced diameter extends upwardly therefrom.

Another embodiment of the invention is shown in FIG. **3**. In this embodiment, pole **1** of 1 ½ inch diameter extends upwardly from the ground to a height of 40 feet or more. Guy wires **3** may be attached to the pole **1** and to the ground **7** to give extra stability. The base of the pole is substantially as shown in FIGS. **1** and **2**. The top portion **8** of the pole, shown in detail in FIG. **4**, increases the dissipation of static electricity to the atmosphere by having two points of discharge extending into the atmosphere.

In FIG. **4**, the top portion **8** of the static eliminator of FIG. **3** is shown. This top portion consists of a T coupler **12** attached to the top of pole **1**. A connecting portion **11** having a length of about 12 inches is connected to either side of the T **12**. An elbow **10** connects the laterally extending connecting portion **11** to an upwardly extending extension **9**. The extension **9** extends upwardly approximately 12 inches, and acts to dissipate the electrical charge into the atmosphere.

Although the foregoing description is necessarily of a detailed character in order that the invention be completely set forth, it is to be understood that the specific terminology is not intended to be restrictive or confining, and that the various rearrangements of parts or modification of detail may be resorted to without departing from the scope or spirit of the invention as herein claimed.

I claim:

1. A static electricity eliminator for bleeding static electricity from the ground to the atmosphere comprising a metal pole having a lower end buried in the ground and an upper end extending upwardly into said atmosphere, a plurality of wires attached to said lower end and extending outwardly therefrom, said plurality of wires being buried in said ground, a concrete collar surrounding said lower end in the ground for supporting said pole, and means surrounding said lower end for electrically insulating said pole from said ground.

2. The static electricity eliminator of claim **1** wherein each of said plurality of wires is at least 300 feet long.

3. The static electricity eliminator of claim **1** wherein said metal pole is not connected to any other surrounding structures.

4. The static electricity eliminator of claim **1** wherein said upper end consists of a pipe of a first diameter extending upwardly from said lower end, and a pipe of a second, lesser diameter, extending upwardly from said pipe of said first diameter.

5. The static electricity eliminator of claim **1** wherein said upper end consists of a U-shaped member attached to said metal pole.

6. A device for eliminating static electricity from the ground by gradually bleeding the static electricity to the atmosphere comprising, a metal pole having a lower portion buried at least three feet in the ground and surrounded by a concrete collar, a plurality of copper wires attached to said lower portion and extending outwardly therefrom a distance of at least 300 feet, said plurality of wires being buried beneath the surface of the earth, an upper portion extending upwardly from the ground into the atmosphere of the earth, said upper portion terminating in a U-shaped member, insulation means surrounding said lower portion between said concrete collar and said lower portion for electrically insulating said metal pole from the ground, and guy wires extending from said metal pole to the ground for supporting said upper portion, wherein said static electricity is gathered from the earth by said plurality of wires, is conducted to said metal pole and passes to said U-shaped member and is discharged from said U-shaped member to the atmosphere.

7. The device for eliminating static electricity from the ground of claim **6**, wherein the metal pole is not attached to any surrounding structures.

* * * * *

Oct. 20, 1959 J. A. WILSON 2,909,190

STATIC ELECTRICITY DETECTION AND CONTROL SYSTEM

Filed Oct. 29, 1958 4 Sheets-Sheet 1

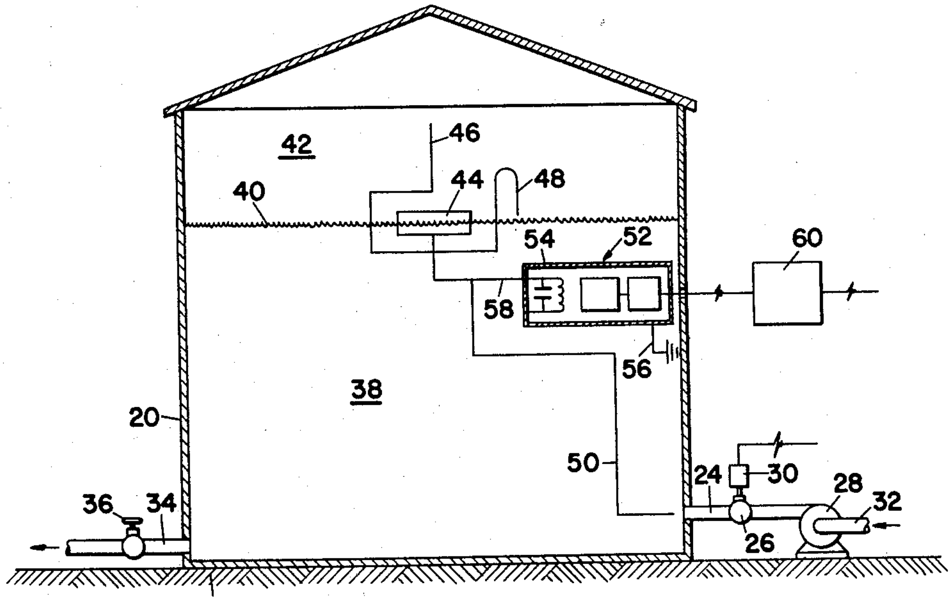

42
46
44
48
40
54
52
60
58
56
38
20
50
24
30
28
32
36
34
26
22

FIG.- 1

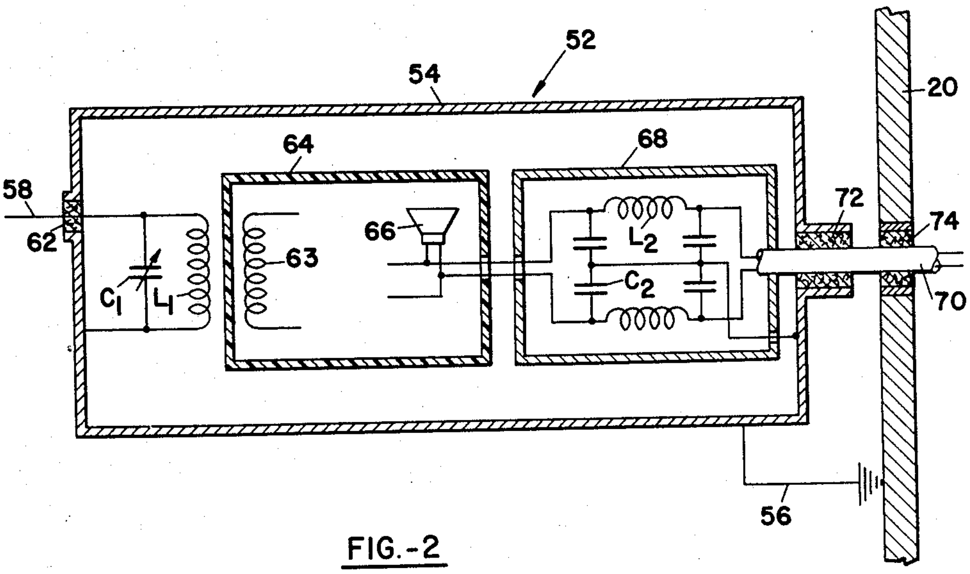

52
54
20
64
68
58
62
72
74
66
63
L_2
C_2
C_1
L_1
70
56

FIG.-2

James A. Wilson Inventor

By W.O.T. Heilman

Attorney

Oct. 20, 1959 J. A. WILSON **2,909,190**

STATIC ELECTRICITY DETECTION AND CONTROL SYSTEM

Filed Oct. 29, 1958 4 Sheets-Sheet 2

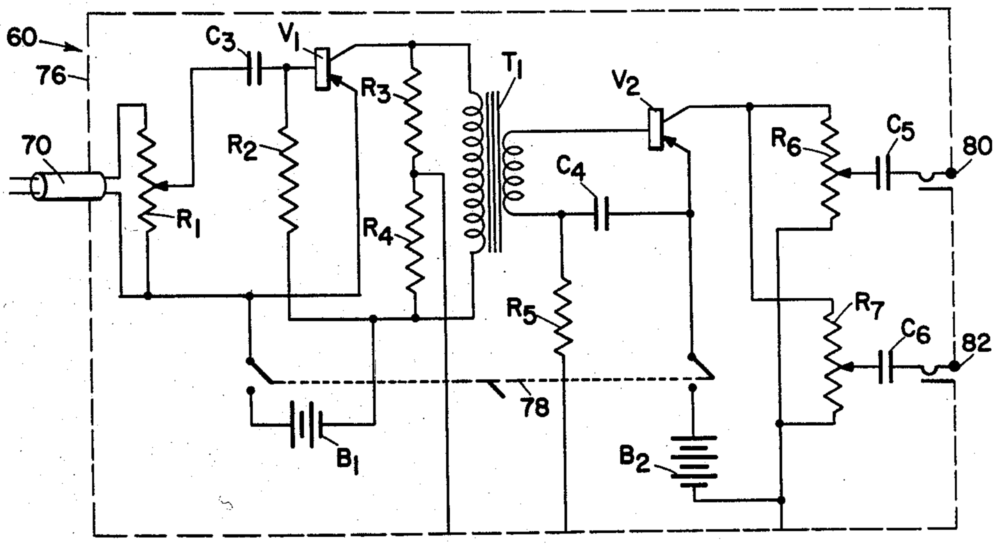
60
76
70
C_3
V_1
R_1
R_2
R_3
R_4
T_1
V_2
C_4
R_5
R_6
C_5
80
R_7
C_6
82
B_1
B_2
78

FIG.-3

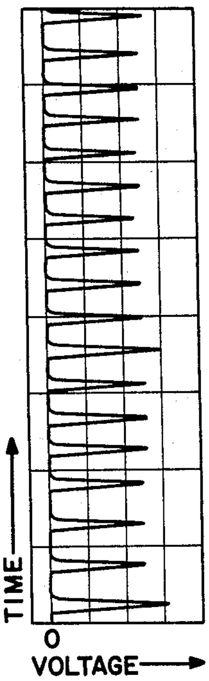

FIG.-5

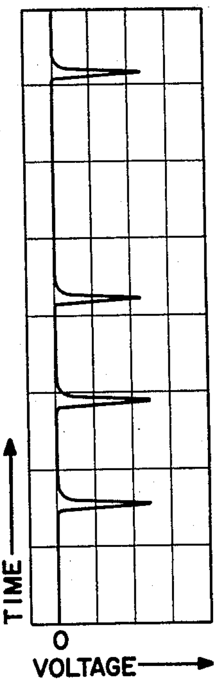

FIG.-7

James A. Wilson Inventor

By W. O. T. Heilman Attorney

Oct. 20, 1959 J. A. WILSON **2,909,190**

STATIC ELECTRICITY DETECTION AND CONTROL SYSTEM

Filed Oct. 29, 1958 4 Sheets-Sheet 3

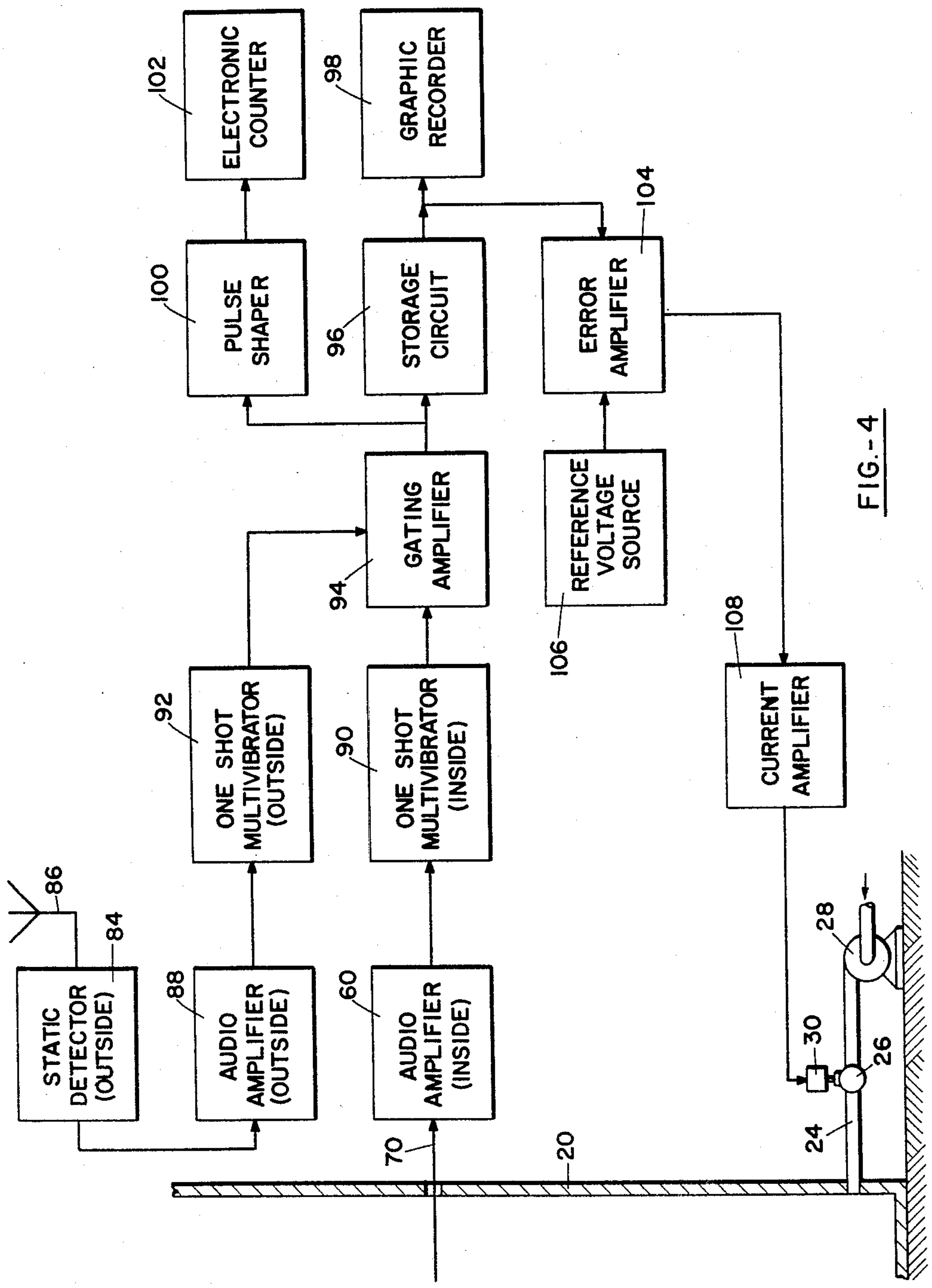

ELECTRONIC COUNTER
102
GRAPHIC RECORDER
98
PULSE SHAPER
100
STORAGE CIRCUIT
96
ERROR AMPLIFIER
104
GATING AMPLIFIER
94
REFERENCE VOLTAGE SOURCE
106
ONE SHOT MULTIVIBRATOR (OUTSIDE)
92
ONE SHOT MULTIVIBRATOR (INSIDE)
90
CURRENT AMPLIFIER
108
STATIC DETECTOR (OUTSIDE)
86
84
AUDIO AMPLIFIER (OUTSIDE)
88
AUDIO AMPLIFIER (INSIDE)
60
70
20
28
30
26
24

FIG.-4

James A. Wilson Inventor

By W. O. T Heilman Attorney

Oct. 20, 1959 J. A. WILSON **2,909,190**

STATIC ELECTRICITY DETECTION AND CONTROL SYSTEM

Filed Oct. 29, 1958 4 Sheets-Sheet 4

FIG.-6

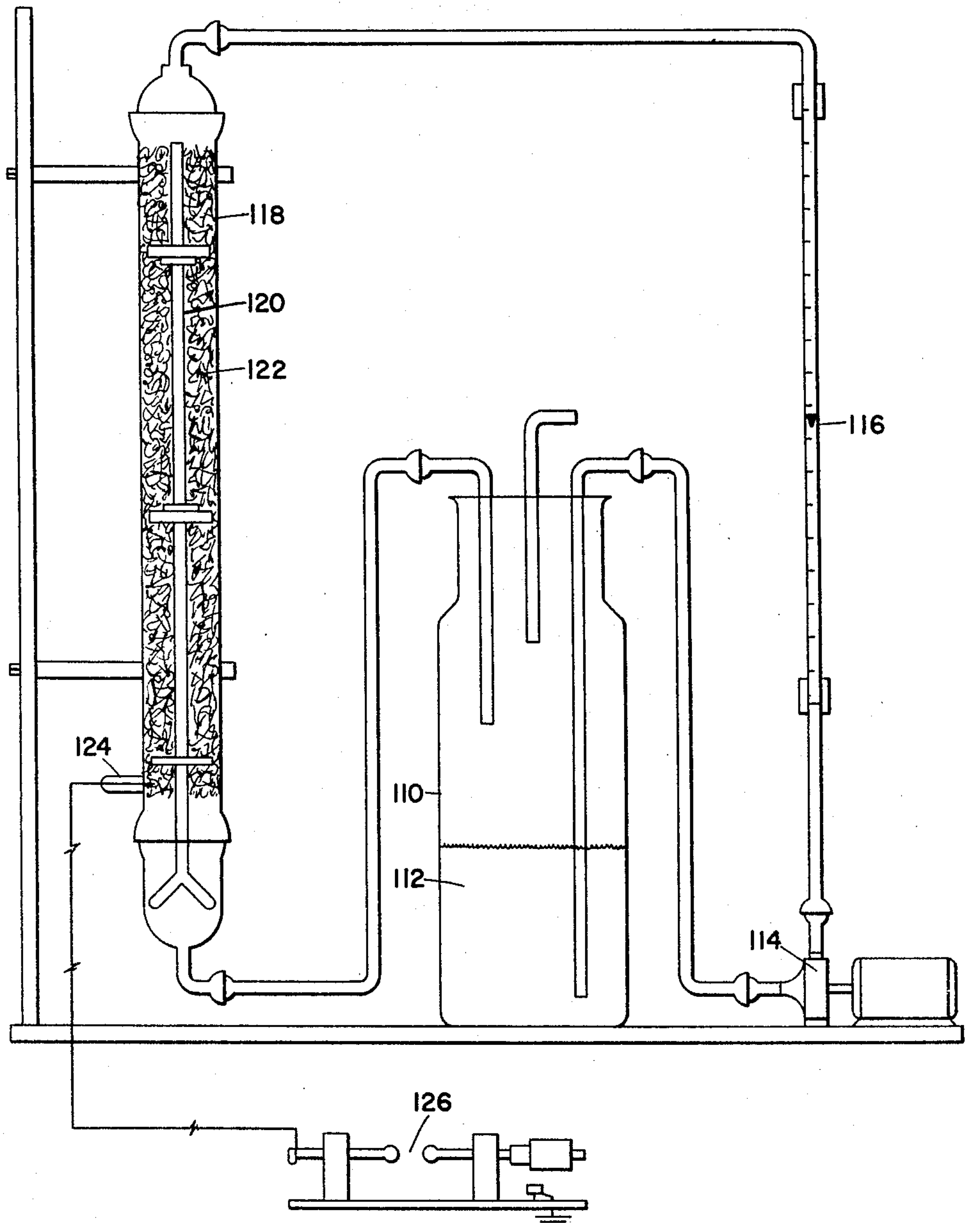

118
120
122
116
124
110
112
114
126

James A. Wilson Inventors

By W. D. T Heilman Attorney

United States Patent Office

2,909,190

Patented Oct. 20, 1959

2,909,190

STATIC ELECTRICITY DETECTION AND CONTROL SYSTEM

James A. Wilson, Stanhope, N.J., assignor to Esso Research and Engineering Company, a corporation of Delaware

Application October 29, 1958, Serial No. 770,366

15 Claims. (Cl. 137—487.5)

This invention relates to a method and means for detecting the discharge of static electricity. It relates particularly to a method and means for detecting the discharge of static electricity within storage tanks at least partially filled with hydrocarbon liquids, and it relates more particularly to a method and means for generating a signal according to the discharge of static electricity within a storage tank at least partially filled with a hydrocarbon liquid and using this signal to regulate the flow of liquid in the line whereby this tank is filled.

During routine handling such as filling a storage tank, hydrocarbon liquids subject to agitation as by pumps and filters may accumulate static electric charges. For a detailed discussion of such accumulation and its significance with respect to methods and means for the handling of hydrocarbon liquids reference may be had to the book Electrostatics in the Petroleum Industry, edited by A. Klinkenberg and J. L. van der Minne, Elsevier Publishing Co., New York, 1958. For purposes of this invention it may be said that electrostatic charges in and on a hydrocarbon liquid body which is resident in and partially fills a storage tank will induce an electrostatic field in the liquid body and in the vapor space above the liquid. This field may be thought of as a pattern of lines of force running from the charges to the wall, bottom, and roof boundaries of the tank, and passing at right angles through a series of surfaces of constant potential. These are closed surfaces, some of which may exist wholly in the liquid body and others of which may exist partially in the liquid body and partially in the vapor space.

Strength of an electrostatic field is determined as the rate of change of potential per unit of distance as one moves through the field along a line of electrostatic force. Common units for expressing field strength are volts/in. and volts/cm. Any medium in which an electrostatic field exists will have a characteristic dielectric strength measured in units similar to those of field strength. For example, most gases at atmospheric pressure have a dielectric strength of approximately 30,000 volts/cm. When the electrostatic field strength at any point in a medium exceeds the dielectric strength thereof, electrostatic discharges will take place which, if sufficiently intense in a hydrocarbon vapor region, may cause sparking with attendant danger of explosion.

If low level, sub-sparking electrostatic discharges could be detected in a storage tank being filled with a hydrocarbon liquid, safety conditions could be improved before the onset of sparking by shutting off the flow of liquid to the tank or at least reducing the rate thereof. In this way the supply of electrostatic charges to the liquid body in the tank would be stopped or slowed. Considering the charges already in the tank, it is characteristic of hydrocarbon liquids that if they be allowed to stand quiescently, as the greater amount of the liquid body in a storage tank will be standing even though there be an influent stream of fluid, electrostatic charges in

them initially will tend to be dissipated by leakage to ground and field strengths to be reduced. This dissipation of charge goes on as a flow of current quite in the absence of any static sparking or sub-sparking discharge according to the "relaxation" principle which is discussed by Klinkenberg and van der Minne at page 39. The concept presently important, however, is simply that if flow of electrostatically charged fluid to a storage tank be so reduced that with respect to the liquid body in the tank the charging rate falls below the relaxation rate plus the static discharge rate there will be a net decrease of charge in the tank and an accompanying reduction in electrostatic field strengths.

According to this invention, radio means are provided whereby electrostatic discharges occurring inside a tank for the storage of hydrocarbon liquid materials may be detected, recorded, and particularly distinguished from electrostatic discharges taking place outside the tank. Means are provided further whereby the flow of liquids to the tank may be regulated according to the internal electrostatic discharge activity detected.

The nature and substance of this invention may be more clearly perceived and fully understood by referring to the following description and claims taken in connection with the accompanying drawings in which:

Fig. 1 represents a view in sectional elevation through a tank partially filled with a hydrocarbon liquid, and in which is located radio means for electrostatic discharge detection according to this invention;

Fig. 2 represents a schematic diagram with details of sealing and shielding of a coupler, receiver, and filter apparatus suitable for use inside a hydrocarbon liquid storage tank for radio detection of electrostatic discharges according to this invention;

Fig. 3 represents a schematic diagram of a remote audio amplifier suitable for use with the coupler, receiver, and filter apparatus of Fig. 2;

Fig. 4 represents a block diagram of the radio and other electrical apparatus continuing from a connection appearing in Fig. 1 whereby electrostatic discharges occurring inside the tank are distinguished from those taking place outside the tank and subsequently recorded, and whereby such distinguished discharges are employed to regulate the flow of liquid to the storage tank;

Fig. 5 represents a graphic recording of electrostatic discharges across a calibrated spark gap from a direct current power supply;

Fig. 6 represents a laboratory pumping test apparatus for generating electrostatic charges in hydrocarbon liquids and discharging them across a spark gap; and

Fig. 7 represents a graphic recording of electrostatic discharges occurring during the pumping of JP–4 fuel over sulfur in the laboratory pumping test apparatus of Fig. 6.

Referring now to the drawings, in Fig. 1 a roofed tank adapted for the storage of hydrocarbon liquids is designated **20**. This tank is supported by and electrically grounded to earth strata **22**. The tank is provided with an inlet line **24** in which are located a flow control valve **26** and a centrifugal pump **28**. Inlet flow control valve **26** is equipped with a power-driven actuating mechanism **30**. The suction line to pump **28** is designated **32**. Tank **20** also has a drain line **34** and an outlet flow control valve **36** in this line.

Within tank **20** there is a body of hydrocarbon liquid **38** having an upper free surface **40**. Above surface **40** there is a vapor space **42**. Supported buoyantly in the liquid at the surface thereof is a float member **44** which carries the antennae **46**, **48** and **50** of one apparatus for detection of electrostatic discharges according to this invention. As shown, these antennae are intended respectively to be located primarily in the vapor space, in both liquid and vapor, and in the liquid. Any other appropriate antenna arrangement may be used, however. The antenna wires have a degree of flexibility, and suitable locating means not particularly illustrated are provided for maintaining the antennae in substantially the desired positions and configurations as float member **44** rises and falls with changes in level of liquid surface **40**.

Mounted within tank **20** on appropriate supporting means not particularly designated is an electrostatic discharge detector apparatus **52** which comprises a sealed, explosion proof metal case **54** containing antenna coupler, radio receiver, and receiver output filter means which will be more completely described in connection with Fig. 2. Case **54** is grounded to the tank structure as by wire **56**. Should detector apparatus **52** be supported from the structure of tank **20** by metal braces or brackets, the latter may serve as grounding means. The several antennae, all of which comprise wires with insulating coverings of suitable material such as polyethylene intended to prevent grounding of an antenna upon contact with tank structure, are joined in a common lead **58** running to the coupler while the leads from the receiver output filter are connected to a remote audio amplifier designated **60**. The latter apparatus, intended to operate on signals generated inside the tank, will be more completely described in connection with Fig. 3.

Referring now to Fig. 2, the common antenna lead **58** is brought into case **54** through a vapor and liquid seal **62** to the tuned antenna coupler comprising the parallel circuit of inductor L_1 and variable capacitor C_1. Inductor L_1 is closely coupled to coil **63** which is the antenna loop of a radio receiver circuit contained within housing **64**. L_1 may in fact comprise a wire wrapping on housing **64**. This housing and the circuit components within it may be simply any radio for standard broadcast reception, preferably one which is battery operated and employs transistors rather than vacuum tubes for the sake of compactness. Variable capacitor C_1 and the radio circuit may be tuned to couple at a frequency of about 1700 kc. corresponding to the approximate carrier frequency to be expected of radiant energy released by electrostatic discharges within the tank.

The radio circuit within housing **64** is characterized by a speaker **66**. From the terminals of this speaker electrical connections are made to a filter section comprising four capacitors C_2 and two inductors L_2 located in a filter shield **68**. The filter section limits stray static pickup in the output of detector **52**. C_2 may be .001 mfd. and L_2 may be 2.5 mh. The output leads of the filter section are provided with suitable shielding **70**, and the shielded cable so formed which may be rated at 2w. is taken out of case **54** through a vapor and liquid seal **72**, and then out of tank **20** through another such seal **74**.

Referring now to Fig. 3, the remote audio amplifier **60** comprises a metal case **76** containing an electrical circuit which may be as shown. This amplifier circuit is connected to the leads from the filter circuit of the detector brought to it inside shield **70**. The off-on switch **78** of the amplifier, shown in dotted outline indicative of its function, is of the double pole, single throw variety. The output terminals **80** and **82** may be used respectively for a phone monitor and a recorder. Suitable values for the various circuit elements may be as follows: $R_1=1000$ k. ohms; $R_2=220$ k. ohms; $R_3=R_4=10$ k. ohms; $R_5=200$ k. ohms; $R_6=R_7=10$ k. ohms; $C_3=4$ mfd.; $C_4=8$ mfd.; $C_5=C_6=0.1$ mfd.; $B_1=1.5$ v.; $B_2=7.5$ v.; T_1 is an SO-3 type iron-core transformer, and V_1 and V_2 are type 2N107 transistors.

Audio amplifier **60** is referred to as an "inside" amplifier although it is located outside of tank **20**. This designation relates to its function of amplifying signals from detector apparatus **52** which is sensitive to electrostatic discharges occurring inside the tank.

Referring now to Fig. 4, **84** designates an electrostatic discharge detector generally similar to detector **52** but located outside of tank **20**. Detector **84** is provided with an antenna **86**, likewise located altogether externally to tank **20**. The output signal from detector **84** is taken to the outside audio amplifier **88** which is generally similar to amplifier **60**. The audio frequency output signals from the inside and outside amplifiers **60** and **88** are impressed respectively on the inside and outside one shot multivibrators **90** and **92** which are relaxation oscillators, devices well known in the electronic art. Discrete signals from the audio amplifiers, representative of particular discharges of static electricity sensed by the detectors at or above a particular energy level, will trigger the multivibrators. Upon being triggered, each multivibrator will go through a single cycle of operation to give an output voltage signal.

Desirably the output signals from both multivibrators will be rectangular in form. It is necessary that the output signal from the inside multivibrator **90** be positive-going, and that from the outside multivibrator **92** be negative-going and of at least as great a negative peak value as that from multivibrator **90** is positive. It is necessary further that the duration of the signal from the outside multivibrator be of at least as great duration as that from the inside multivibrator, and desirably it will be somewhat longer. For example, the output signal duration for multivibrator **90** may be 0.5 millisecond while that from multivibrator **92** may be 1.0 millisecond. The several conditions of signal shape, direction, magnitude and duration may be satisfied through appropriate design of the multivibrators according to procedure well known in the electronics art.

For given designs of the static detectors **52** and **84** and the audio amplifiers **60** and **88**, the strengths of the signals impressed on multivibrators **90** and **92** will be functions respectively of the energy levels of electrostatic discharges taking place inside and outside tank **20**. At least a certain strength of signal leaving either audio amplifier will be needed to trigger the corresponding multivibrator. Signals from an audio amplifier of greater strength than this will not affect the signal output from the multivibrator, but audio amplifier output signals of less strength will fail to trigger the multivibrator at all. By suitable design of the detector, audio amplifier, and multivibrator systems, triggering of the multivibrators may be set to take place only when certain critical levels of energy release due to electrostatic discharges inside and outside of the tank are achieved or exceeded.

Signals from both multivibrators **90** and **92** are brought to gating amplifier **94**. This amplifier which may be of a kind well known in the electronic art is, in effect, a circuit arranged to act as a gate with respect to an input signal in either passing or blocking this signal or an amplified version thereof. Customarily in a gating amplifier circuit there is a vacuum tube which must be in conducting condition for an input signal to be passed. This tube, which might otherwise be continuously conducting due to the construction of the amplifier, may be rendered non-conducting by imposition of a negative-going voltage on a suppressor grid element thereof. Such a negative-going voltage is sometimes called an inhibit voltage or signal. For purposes of this invention, gating amplifier **94** must function to either pass or block the positive-going signal reaching it from the inside multivibrator **90**. The outside multivibrator **92** is, accordingly, so connected to the gating amplifier that its negative-going signal will act as an inhibit signal to prevent any signal from the inside multivibrator **90** or amplified version of such signal from being passed by amplifier **94**.

The reason for the arrangement just described is that it is desired to have output signals from gating amplifier **94** which correspond only to electrostatic discharges taking place within tank **20**. Electrostatic discharges due to uncontrollable causes may be expected to be taking place continuously outside of tank **20**. The radiant energy of some such discharges, though greatly attenuated by the walls of tank **20** and the grounded metal detector case **54**, may be picked up by the external antenna system and internal loop **63** of the inside detector **52** and result in signals from the inside multivibrator **90** being impressed on gating amplifier **94**. If these signals or amplified versions thereof were not blocked by the gating amplifier, the output of this amplifier would include signals corresponding not only to electrostatic discharges taking place inside the tank, but also to some taking place outside the tank. By imposing on the gating amplifier as inhibit signals voltages corresponding to electrostatic discharges external to the tank, the gating amplifier will never be "open" to pass a signal from the inside multivibrator **90** when this signal is due to an electrostatic discharge taking place outside the tank. Therefore, the output of gating amplifier **94** will comprise only signals due to electrostatic discharges occurring inside the tank.

The possibility exists that an electrostatic discharge taking place inside the tank will coincide with one taking place outside, and both be of sufficient energy level as picked up by the antenna systems of detectors **52** and **84** to result in triggering of both multivibrators simultaneously. In this case the gating amplifier will block the signal reaching it from the inside multivibrator **90**, and no signal corresponding to the electrostatic discharge inside the tank will appear in the output of gating amplifier **94**. Statistically, however, the possibility of such coincidence of electrostatic discharges inside and outside the tank is slight, and it is contemplated that very few electrostatic discharges occurring inside the tank at sufficient energy level to cause triggering of multivibrator **90** will fail to appear as output signals from the gating amplifier.

Another theoretical possibility is that the outside detector **84** will pick up signals due to electrostatic discharges taking place inside the tank, which if suitably amplified could cause triggering of multivibrator **92** and so always have gating amplifier **94** closed against any signal reaching it from multivibrator **90**. In this way the gating amplifier would have no output at all. In practice this possibility can be kept from ever becoming an actuality by proper design of the outside static detector **84**, audio amplifier **88**, and multivibrator **92** with respect to the energy level of signals likely to be picked up by antenna **86** of detector **84**. Any signal picked up by antenna **86** which is due to an electrostatic discharge inside the tank will be at a very low energy level if for no other reason than that the attenuation of it by passage through the walls and roof of tank **20**.

Accordingly if detector **84**, amplifier **88**, and multivibrator **92** be so designed that there will be no triggering of multivibrator **92** by any signal picked up by antenna **86** at an energy level corresponding to that of an electrostatic discharge within the tank as attenuated by the tank structure, there will be no actual possibility of these discharges within the tank being self-suppressing with respect to output signals from the gating amplifier **94**. Such a design consideration will not be inconsistent with the requirement that the outside multivibrator **92** be triggered by electrostatic discharges occurring outside of tank **20** since such discharges are likely to be at a considerably higher energy level than those inside, and in respect to antenna **86** there will be no attenuation of the radiant energy of these discharges by the tank structure.

Output signals from gating amplifier **94** flow to storage circuit **96**. This circuit comprises a capacitor and a rectifier element. The capacitor is charged by signals of a pulse nature reaching it from gating amplifier **94**. It will discharge to give an analog signal representing an average of the pulse input per unit time. The capacitor of storage circuit **96** is, in effect, an integrating device. The rectifier element of storage circuit **96** is so installed to prevent the backflow of any signal from the storage circuit to the gating amplifier as the storage circuit capacitor discharges. The rectifier element is thus a one-way isolating device between the gating amplifier and the storage circuit capacitor.

Storage circuit **96** should be designed with a sufficiently high time constant, 1.0 sec. for example, that it will give a continuous direct voltage output signal for any substantial input rate of pulse signals from the gating amplifier. The magnitude of this output voltage will vary directly with the pulse frequency of the gating amplifier output. The output voltage from storage circuit **96** may be impressed on any suitable graphic recorder **98**. The record made therein will be a significant indicator of electrostatic discharge activity of at least a certain energy level within storage tank **20**.

It may be desired to obtain an actual count of the electrostatic discharges at or above a certain energy level taking place within tank **20**. This can be done by means of pulse shaper **100** and an electronic counter **102**. The pulse shaper will receive output pulse signals of substantially rectangular form from gating amplifier **94** and perform a differentiating operation on the leading edges of these pulses to produce corresponding signals of spike form as its own output. These spike signals will be impressed on and actuate counter **102** which may be of the decade type well known in the electronic art.

The direct voltage output from storage circuit **96** goes not only to graphic recorder **98**, but also to error amplifier **104** which is an essential element of the control system whereby the filling rate of tank **20** is regulated according to this invention. The function of the error amplifier is to compare the voltage impressed on it from storage circuit **96** on the one hand, and a reference voltage source **106** on the other and to amplify any difference between these voltages, this difference being known as error. The amplified error voltage signal comprising the output of amplifier **104** will be at a comparatively low power level, and generally incapable of performing control operations itself. It is, accordingly, taken to a current amplifier **108** and there raised to such power level that the output signal of amplifier **108** is a direct current of magnitude sufficient to energize the actuating mechanism **30** of inlet flow control valve **26**. This mechanism may comprise a direct current motor geared or otherwise operatively connected to the stem of valve **26**.

There will be some level of electrostatic discharge activity in tank **20** which will correspond to a safe upper level of potential gradients induced in vapor space **42** by electrostatic charges in the liquid body **38**. This level of electrostatic discharge activity, as indicated by the magnitude of the direct voltage output of storage circuit **96**, must be determined by independent experiment and calculation. Once determined, however, it becomes the value of voltage to be impressed on error amplifier **104** from reference voltage source **106**.

The error voltage signal comprising the output of error amplifier **104** will have direction as well as magnitude. It will have one direction if the reference voltage exceeds the output voltage of storage circuit **96** and vice versa. The direction of the error voltage will determine the direction of control current supplied to the motor of valve actuating mechanism **30** from current amplifier **108**. A required condition is that when the output voltage of storage circuit **96** exceeds the reference voltage the motor of actuating mechanism **30** must turn to close valve **26** and so restrict flow to tank **20**. Correspondingly, when reference voltage exceeds the output voltage of storage circuit **96** the motor of actuating mechanism **30** should turn to open the valve. Such devices as limit stop switches or slip clutches may be provided to protect the motor from continuing to drive when valve **26** is fully open or fully closed.

Imagine that a hydrocarbon liquid is being supplied to tank **20** by pump **28** through line **24** and valve **26** which is at least part way open. Imagine further that the electrostatic discharge activity in tank **20** so increases that the output voltage from storage circuit **96** exceeds the voltage from reference voltage source **106**. Actuating mechanism **30** will commence closing valve **26** and reducing the rate of flow of oil to tank **20**. The rate of electrostatic charge inflow to tank **20** will be reduced simultaneously. In the meanwhile, continuous dissipation or relaxation of charges in the body of liquid **38** already in the tank by current to ground will be going on. Total charge dissipation rate by relaxation and static discharge will eventually come to exceed charge inflow rate, and when this happens the electrostatic discharge activity in tank **20** will begin to decline.

This activity will continue to decline as valve **26** closes, or possibly goes entirely closed, and with it will decline the output voltage of storage circuit **96**. When this voltage has fallen to the magnitude of the reference voltage, a value corresponding to a safe level of electrostatic discharge activity in tank **20**, there will be no further error signal from error amplifier **104** and no further current to energize the motor of valve actuating mechanism **30**. Closing motion of valve **26** will be terminated. Now if the electrostatic discharge activity in tank **20** continues to decline, the voltage from storage circuit **96** will fall below the reference voltage; the motor of actuating mechanism **30** will be energized to open valve **26**, and the rate of oil flow to tank **20** allowed to increase. The flow control system will try continuously to maintain an oil supply rate to tank **20** whenever the tank is to be filled by pump **28** such that a balance is reached between charge inflow to the tank and charge dissipation from the tank at the maximum safe level of electrostatic discharge activity within the tank.

Referring now to Fig. 5, the electrostatic discharge graphic record shown as a reference for comparison with Fig. 7, is that taken across a calibrated spark gap set to fire at 7,000 volts from a direct current power supply. The amount of energy stored in the system before firing was approximately 12.0 millijoules. The graphic ordinate of time indicates the advance of the chart paper. The graphic abscissa of voltage, while not described quantitatively and being immediately representative of the signal potential impressed on the recording device by the radio frequency electrostatic discharge detecting apparatus, indicates the relative energy levels of spark gap discharges for each firing of the particular system as picked up by the detecting apparatus.

Referring now to Fig. 6, a typical laboratory pumping test apparatus for generation of static charges in hydrocarbon liquids, and means for firing these charges are shown. A reservoir **110** contains a hydrocarbon liquid body **112**. By means of a suction line descending well into the reservoir, liquid is drawn into motor-driven pump **114** which discharges it through rotameter **116**. From the rotameter the pumped liquid flows to and through a glass pipe **118**. Within pipe **118** there is a rack **120** also of glass which supports a body of material **122** such as glass wool used in filters or lumps of sulfur representing substance found in tank scale. From the outer end of the glass pipe the liquid is returned to reservoir **110**. Intermediate connecting tubing runs and fittings and the support structure as well as pipe **118** are of electrically non-conductive material such as glass or Lucite.

A tungsten probe **124** penetrates pipe **118** and is in contact with the lower end of the filling material **122**. This probe is wired to one side of an adjustable spark gap **126**, the other side of which is connected to ground. Electrostatic discharge detecting means such as that described earlier with suitable graphic recording apparatus attached is positioned adjacent the spark gap. When pump **116** is operated there will be agitation of the liquid by the pump itself and filling material **122** with consequent generation and accumulation of charge in the liquid. This accumulated charge will be picked up by probe **124** and a potential developed across spark gap **126**. From time to time a breakdown potential will be achieved and the gap be fired. The electrostatic discharge detecting and recording apparatus will then function in a manner generally described already.

Referring now to Fig. 7, what is shown is an electrostatic discharge record of JP–4 fuel pumped in the apparatus of Fig. 6 using a filling of lump sulfur in pipe **118**. The voltage trace of Fig. 7 was, in fact made with the same spark gap, radio detector, and graphic recorder apparatus used in preparation of Fig. 5, although, of course, at a different time. Evaluating these records in terms of tank storage of hydrocarbon liquids discussed earlier, Fig. 5 developed from spark equipment using a direct current power supply might be pictured as the external interference signal while Fig. 7 with Fig. 5 added at some attenuated energy level would then represent a typical composite signal of electrostatic discharges sensed by radio frequency detecting apparatus located within a tank as detector **52** and its antennae within tank **20**.

Although this invention has been described with a certain degree of particularity, it is to be understood that the present disclosure has been made only by way of example and that numerous changes in the details of construction and the combination and arrangement of parts may be resorted to without departing from the spirit and scope of this invention as hereinafter claimed.

What is claimed is:

1. In combination with a vessel for the storage of liquid materials and conduit means communicating with said vessel wherethrough liquid material may be introduced into said vessel, an apparatus for control of electrostatic discharges within said vessel comprising radio means characterized by pick-up antenna means disposed within said vessel whereby electrostatic discharges occurring within said vessel may be detected and signals corresponding to said discharges generated for further transmission, a reference signal source, comparator means whereby signals from said radio means and said reference signal source may be matched and a differential signal generated for further transmission, transmission means connecting said radio means and said comparator means, transmission means connecting said reference signal source and said comparator means, liquid flow regulating means in said conduit, and transmission means connecting said comparator means and said liquid flow regulating means whereby said differential signal is imposed on said liquid flow regulating means.

2. An apparatus according to claim 1 in which said liquid flow regulating means comprises a valve having an electrically actuated opening and closing means.

3. An apparatus according to claim 2 in which said electrically actuated opening and closing means comprises a direct current motor.

4. An apparatus according to claim 1 including a current amplifying means whereby said differential signal from said comparator means is raised in power level prior to imposition upon said liquid flow regulating means.

5. In combination with a vessel for the storage of liquid materials and conduit means communicating with said vessel wherethrough liquid material may be introduced into said vessel, an apparatus for controlling electrostatic discharges within said vessel comprising a first radio means inside said vessel whereby electrostatic discharges occurring within said vessel may be detected and signals corresponding to said discharges generated for further transmission, a second radio means outside of said vessel whereby electrostatic discharges occurring without said vessel may be detected and signals corresponding to said discharges generated for further transmission, a gating means adapted to receive signals from said first and

second radio means and to generate signals for further transmission corresponding only to those signals from said first radio means which correspond in turn only to those electrostatic discharges occurring within said vessel non-coincidentally with electrostatic discharges occurring without said vessel and for which corresponding signals are generated by said second radio means, transmission means connecting said first radio means and said gating means, transmission means connecting said second radio means and said gating means, a reference signal source, comparator means whereby signals from said gating means and said reference signal source may be matched and a differential signal generated for further transmission, transmission means connecting said gating means and said comparator means, transmission means connecting said reference signal source and said comparator means, liquid flow regulating means in said conduit, and transmission means connecting said comparator means and said liquid flow regulating means whereby said differential signal is imposed on said liquid flow regulating means.

6. An apparatus according to claim 5 in which said liquid flow regulating means comprises a valve having an electrically actuated opening and closing means.

7. An apparatus according to claim 6 in which said electrically actuated opening and closing means comprises a direct current motor.

8. An apparatus according to claim 5 including a current amplifying means whereby said differential signal from said comparator means is raised in power level prior to imposition upon said liquid flow regulating means.

9. In combination with a vessel for the storage of liquid materials and conduit means communicating with said vessel wherethrough liquid material may be introduced into said vessel, an apparatus for control of electrostatic discharges within said vessel comprising a first radio means inside said vessel whereby electrostatic discharges occurring within said vessel may be detected and signals corresponding to said discharges generated for further transmission, a second radio means outside of said vessel whereby electrostatic discharges occurring without said vessel may be detected and signals corresponding to said discharges generated for further transmission, a first multivibrator adapted to receive signals from said first radio means and generate pulse signals for further transmission corresponding to signals at or above a first predetermined energy level received from said first radio means, a second multivibrator adapted to receive signals from said second radio means and generate pulse signals for further transmission corresponding to signals at or above a second predetermined energy level received from said second radio means, transmission means connecting said first radio means and said first multivibrator, transmission means connecting said second radio means and said second multivibrator, a gating means adapted to receive signals from said first and second multivibrators and to generate pulse signals for further transmission corresponding only to those signals from said first multivibrator which correspond in turn to those signals from said first radio means which finally correspond only to those electrostatic discharges occurring within said vessel non-coincidentally with electrostatic discharges occurring without said vessel and detected and for which corresponding signals resulting in output pulses from said second multivibrator are generated by said second radio means, transmission means connecting said first multivibrator and said gating means, transmission means connecting said second multivibrator and said gating means, a storage means adapted to receive pulse signals from said gating means and to generate a signal for further transmission corresponding to a time-integrated value of pulse signals received from said gating means, transmission means connecting said gating means and said storage means, a reference signal source, a comparator means whereby signals from said storage means and said reference signal source may be matched and a differential signal generated for further transmission, transmission means connecting said storage means and said comparator means, transmission means connecting said reference signal source and said comparator means, liquid flow regulating means in said conduit, and transmission means connecting said comparator means and said liquid flow regulating means whereby said differential signal is imposed on said liquid flow regulating means.

10. An apparatus according to claim 9 in which said liquid flow regulating means comprises a valve having electrically actuated opening and closing means.

11. An apparatus according to claim 10 in which said electrically actuated opening and closing means comprises a direct current motor.

12. An apparatus according to claim 9 including a current amplifying means whereby said differential signal from said comparator means is raised in power level prior to imposition upon said liquid flow regulating means.

13. An apparatus according to claim 9 including amplifying means associated with each of said first and second radio means whereby audio frequency output signals from said radio means are amplified prior to transmission to said first and second multivibrators respectively.

14. An apparatus according to claim 9 including graphic recording means actuated by output signals from said storage means.

15. An apparatus according to claim 9 including pulse counting means actuated by output signals from said gating means.

References Cited in the file of this patent

Electrical Engineering (AIEE), vol. 77, No. 7, July 1958, "Static Electricity in the Petroleum Industry," pp. 610–614), 317–2. (Copy in Scientific Library.)

United States Patent [19]
Daniel

[11] **Patent Number:** **4,530,417**

[45] **Date of Patent:** **Jul. 23, 1985**

[54] **SUPPRESSOR**

[75] Inventor: **Wayne Daniel,** Atlanta, Ga.

[73] Assignee: **SW Daniel, Inc.,** Atlanta, Ga.

[21] Appl. No.: **506,812**

[22] Filed: **Jun. 22, 1983**

[51] **Int. Cl.**^{3} **F41C 21/18**

[52] **U.S. Cl.** **181/223;** 181/252; 181/281; 89/14.4

[58] **Field of Search** .... 89/14 R, 14 B, 14 C, 89/14 D; 181/223, 252, 256, 264, 281

[56] **References Cited**

U.S. PATENT DOCUMENTS

2,043,731	6/1936	Bourne	181/248
2,073,951	3/1937	Servais	181/252
2,442,773	6/1948	Mason	181/223
2,448,382	8/1948	Mason	89/14.4
3,713,362	1/1973	Charron	89/14.4
3,748,956	7/1973	Hubner	89/14.4
4,291,610	9/1981	Waiser	89/14.4

FOREIGN PATENT DOCUMENTS

338641	3/1936	Italy	181/223

OTHER PUBLICATIONS

Siegfried F. Huebner, Silencers for Hand Firearms, 1976, pp. 74, 76.

Frankford Arsenal Report R-1896, Silencers Patterns and Principles, Aug. 1968, pp. 112–119.

Primary Examiner—Stephen C. Bentley
Attorney, Agent, or Firm—Jones & Askew

[57] **ABSTRACT**

A suppressor for reducing the muzzle blast of firearms or the like. Several separate absorbent elements are aligned in tandem within a tubular housing, and adjacent absorbent elements are separated by triangular baffles which expose end portions of adjacent absorbent elements. These exposed end surface portions enhance dispersion of propellant gases into and through the absorbent elements. The absorbent elements preferably are made of a knitted ferrous or non-ferrous wire mesh, providing a number of tortuous paths for gas dispersion into the absorbent elements and lowering the temperature of the gases to reduce the noise of the muzzle blast.

7 Claims, 4 Drawing Figures

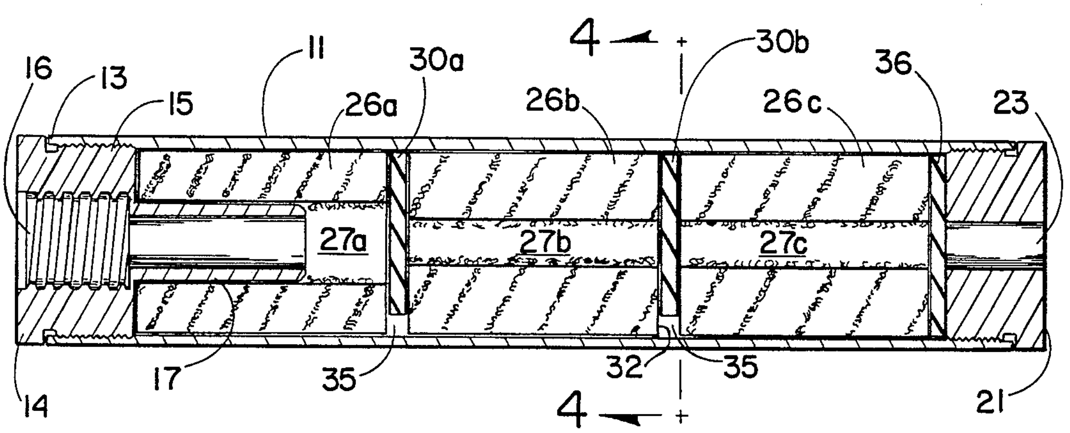

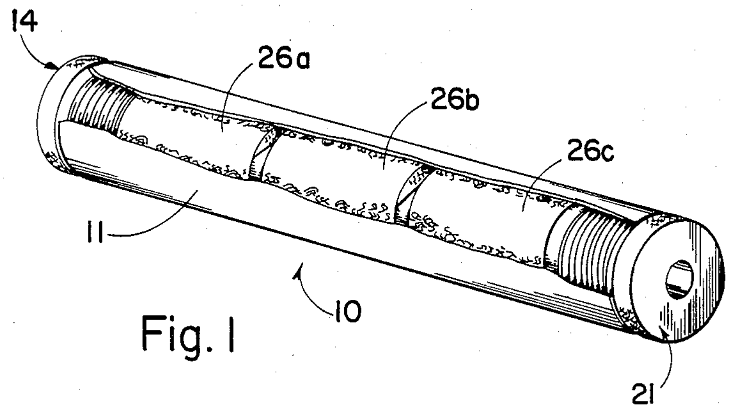
14
26a
26b
26c
11
10
21

Fig. 1

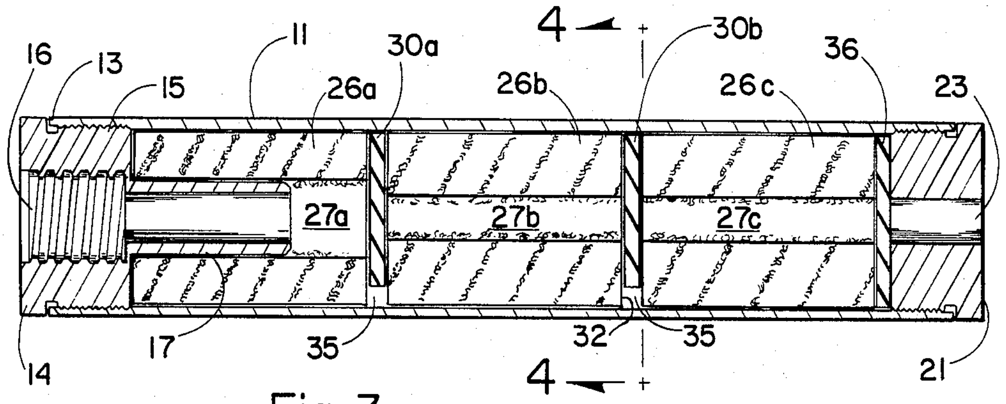
4
4
16
13
15
11
26a
30a
26b
30b
26c
36
23
27a
27b
27c
14
17
35
32
35
21

Fig. 3

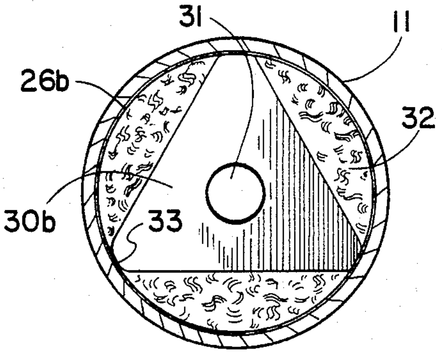
31
11
26b
32
30b
33

Fig. 4

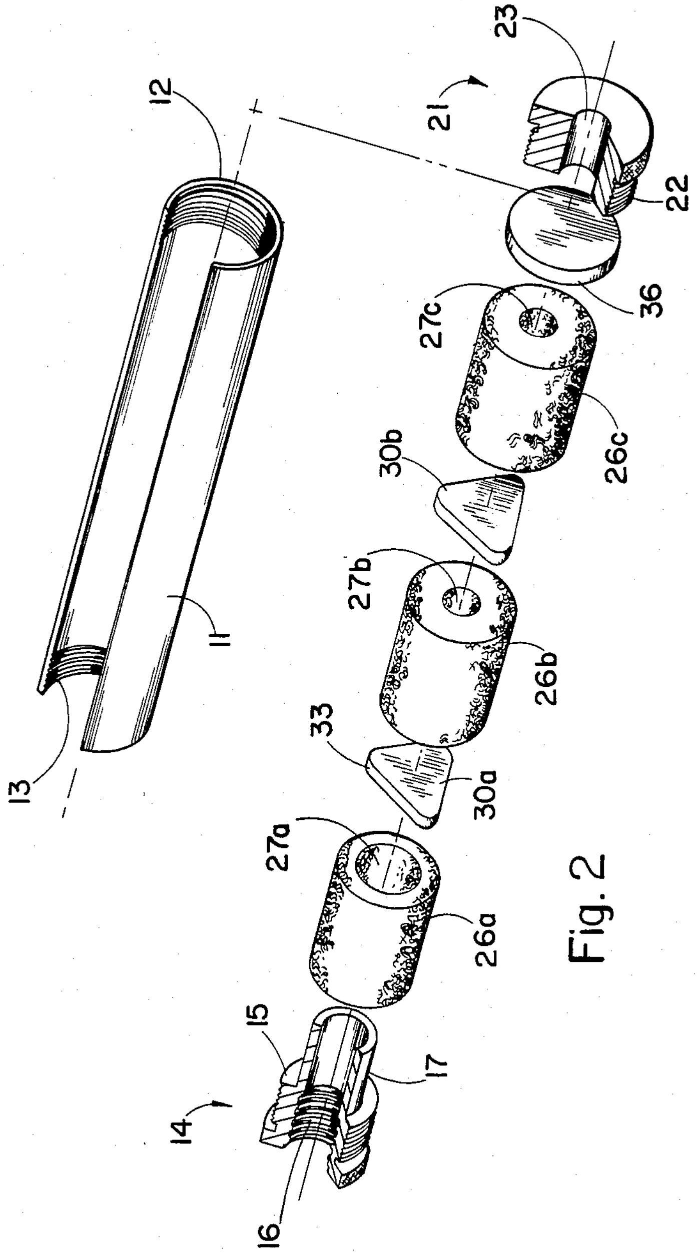
12
23
21
22
36
27c
26c
30b
27b
11
26b
33
30a
13
27a
26a
15
17
14
16

Fig. 2

SUPPRESSOR

FIELD OF THE INVENTION

This invention relates in general to firearms, and in particular relates to apparatus for suppressing the muzzle blast and attendant noise of a firearm.

BACKGROUND OF THE INVENTION

Suppressors or so-called silencers for firearms generally operate to reduce the energy of the gases propelling the bullet from the muzzle of the firearm. By reducing the energy level of the propellant gases in a relatively controlled manner compared to the abrupt discharge of gases leaving the muzzle of an unsuppressed firearm, the audible noise or sharp report of the firing can be suppressed to a significant extent. One technique of suppressing the report of firearms calls for reducing the temperature of the propellant gases, before these gases are released to the atmosphere. Since the internal energy of a compressed gas is a function of the gas temperature, the energy of propellant gases exiting the suppressor at a reduced temperature thus is reduced, bringing about a corresponding reduction in the noise produced by the propellant gases.

Various kinds of suppressors seeking to accomplish the foregoing function are known to the prior art. A number of practical disadvantages generally are associated with these prior suppressors. For example, some suppressors require expansion chambers, baffle plates, partitions, or the like, machined or otherwise fabricated in a manner requiring relatively costly fabrication and assembly techniques. Because suppressors generally are attached to the muzzle of a firearm and extend a distance in front of the muzzle, the added diameter and weight of such existing suppressors frequently is another disadvantage.

One such suppressor, disclosed in U.S. Pat. No. 2,448,382, seeks to reduce these problems by providing the suppressor concentric with a portion of the rifled barrel of the firearm. Radial holes are drilled through the rifled barrel to vent the propellant gases into the surrounding suppressor. This arrangement, while reducing the overall length of firearm-suppressor combination, adversely affects the accuracy of the firearm unless the radial holes enter the rifled barrel only in the grooves of the rifling. This desired placement of the radial holes, while not impossible, calls for considerable manufacturing precision and thus is impractical.

SUMMARY OF INVENTION

The suppressor of the present invention reduces the energy of propellant gases, and thus reduces the associated firing noise or report, by apparatus including one or more absorbent elements serially disposed within a housing attachable to the muzzle of the firearm. Plural individual elements of absorbent material are preferred, and adjacent elements of the absorbent material are separated by a baffle which covers less than the entire end area of the adjacent elements, leaving a portion of the end area uncovered within the housing. These uncovered end portions enhance the passage of propellant gases into and through the gas-absorbent elements, improving the noise-suppressing ability of a suppressor according to the present invention having a given mass of absorbent material.

Stated somewhat more specifically, suppressors according to the present invention preferably use absorbent elements of knitted ferrous or non-ferrous material for absorbing and expanding the propellant gases and dissipating the heat from these gases. It is believed that the use of knitted material provides a labyrinth of closely-packed wires with interstitial spaces defined between the knitted wire mesh, thereby providing an increased expansion volume for the propellant gases and also substantially increasing the surface area of wire available to absorb heat from the propellant gases.

Accordingly, it is an object of the present invention to provide an improved suppressor for firearms or the like.

It is another object of the present invention to provide a suppressor which is relatively efficient in operation without adversely affecting the accuracy of the firearm.

It is still another object of the present invention to provide a suppressor which maximizes gas dispersion into the total interior volume of the suppressor.

Other objects and advantages of the present invention will become more readily apparent from the following discussion of a preferred embodiment.

BRIEF DESCRIPTION OF DRAWINGS

FIG. **1** is a pictorial view showing a suppressor according to a preferred embodiment of the present invention, and having a portion of the suppressor tube cut away for illustrative purposes.

FIG. **2** is an exploded view of the suppressor shown in FIG. **1**, with certain components shown broken away for illustration.

FIG. **3** is a longitudinal section view of the suppressor shown in FIG. **1**.

FIG. **4** is a transverse section view taken along line **4**—**4** of FIG. **3**.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENT

Turning first to FIG. **1**, there is shown generally at **10** a suppressor designed to be removably attached to the muzzle of a firearm (not shown) such as a hand gun or the like. The particular disclosed suppressor embodiment **10** is designed for use with a firearm chambered to fire the .22 long rifle cartridge, although it should be understood that the choice of a particular cartridge is not critical to the present invention.

The suppressor **10** includes an elongated tubular housing **11** having threaded front and back ends **12** and **13**, respectively, best seen in FIG. **2**. At the back end **13** of the housing **11** is the connector sleeve **14**, having an exterior-threaded body portion **15** which screws into the threaded back end of the housing. The connector sleeve **14** has an internal axial passage **16**, FIGS. **2** and **3**, which is threaded to fit corresponding external threads at the muzzle end of the firearm for which the suppressor **10** is to be used. The threads in the axial passage **16** preferably are relatively coarse, so as to facilitate removal and attachment of the suppressor from the firearm barrel.

A hollow neck **17** extends forwardly from the body portion **15** of the connector sleeve **14**. The neck **17** has an open internal passage coaxial with the threaded axial passage **16** through the body portion **15**. As seen in FIG. **3**, it will be appreciated that the internal diameter of the passage in the neck **17** is sufficient to accommodate passage of bullets entering the suppressor through the connector sleeve **14**. It will also be evident that the diameter of the axial passage **16** leading to the neck **17** is somewhat greater than the diameter of the passage in the neck, so as to accommodate the thickness of the threaded barrel to which the connector sleeve will be attached.

An encapsulator **21** is fitted to the front end **12** of the housing **11**, and holds together the suppressor components within the housing. The encapsulator **21** has a body portion **22** externally threaded to fit within the threaded front end **12** of the suppressor housing **11**. An axial passage **23** extends through the encapsulator **21**, and it should be understood that the diameter of this axial passage is selected to permit unimpeded passage of bullets through the suppressor.

A number of separate absorbent elements are disposed within the housing **11** of the suppressor. Three such absorbent elements **26***a*, **26***b*, and **26***c*, taken from the back to the front of the suppressor, are used in the disclosed embodiment, although it should be understood that a greater or lesser number of absorbent elements may alternatively be used. Nevertheless, it is preferred that at least two such separate absorbent elements be utilized, aligned in tandem within the housing **11** as disclosed herein. Each such absorbent element **26***a*–**26***c* is made of a knitted mesh of ferrous or nonferrous material, preferably metallic wire having relatively good heat-transfer characteristics for effective dissipation of heat and consequent temperature reduction of the propellant gases entering the suppressor. In a specific embodiment of suppressor according to the present invention, the absorbent elements comprise knitted mesh made of copper wire for good thermal conductivity and resistance to corrosion. The knitted mesh of the absorbent elements **26***a*–**26***c* is closely wound, defining a multitude of tortuous paths for the flow of gases through the absorbent elements.

Each of the absorbent elements **26***a*–**26***c* takes the shape of an annular cylinder having an axial passage allowing bullets to travel through the suppressor. The axial passages **27***b* and **27***c* of the corresponding absorbent elements **26***b* and **26***c* are of relatively small diameter, just sufficient for unimpeded passage of the bullet therethrough. The axial passage **27***a* through the absorbent element **26***a* is of somewhat larger diameter to accommodate the neck **17** of the connector sleeve **14**, as best shown in FIG. 3. The absorbent element **26***a* thus is a sliding fit on the neck of the connector sleeve **14**. The outer diameter of each absorbent element **26***a*–**26***c* preferably permits a close sliding fit within the housing **11**, so that the internal diameter of the housing is substantially filled by the absorbent elements.

Interposed between adjacent absorbent elements are the triangular-shaped baffles **30***a* and **30***b*. These baffles preferably are made of a flexible elastomeric material, and each baffle has a central opening **31** (FIG. 4) through which bullets pass. However, the baffles **30***a* and **30***b* may be manufactured in the form of blank pieces as depicted in FIG. 2, lacking any central opening; the first round fired through the suppressor will form its own boresight central openings through the baffles. This arrangement not only reduces to some extent the cost of fabricating the baffles, but also assures that the openings **31** formed in the baffles are properly aligned.

Referring now to FIG. 4, it is seen that the triangular shape of the baffle **30***b* covers substantially less than the entire end surface area **32** of the absorbent element **26***b*. The triangular shape of the baffle **30***b* exposes three such end surface areas **32** on the end of the absorbent element **26***b*; the corners **33** of the baffle may be truncated to conform to the inner circumference of the suppressor housing **11**, thereby keeping the baffle substantially centered within the housing. The baffle **30***a* between the adjacent absorbent elements **26***a* and **26***b* likewise is triangular, leaving the confronting end surfaces of those two absorbent elements partially exposed. Each baffle **30***a* and **30***b* is sufficiently thick to maintain a space **35** between the confronting end surface areas of adjacent absorbent elements.

A circular bullet wiping seal **36** is disposed between the back of the encapsulator **21** and the forward end of the front absorbent element **26***c*. The seal **26**, which may be of an elastomeric material the same as the baffles **30***a* and **30***b*, seals the forward end of the suppressor housing **11** and provides a close compression fit of all suppressor elements within the housing as the encapsulator **21** is threaded into the front end **12** of the housing. As seen in FIG. **2**, the seal **36** may be formed without a central opening, the first round fired through the suppressor forming the necessary opening as with the baffles **30***a* and **30***b*.

When a firearm fitted with the suppressor **10** is fired, the bullet serially passes through the several suppressor sections defined by the separate absorbent elements **26***a*–**26***c* and the baffles **30***a*, **30***b* interposed between adjacent absorbent elements. The expanding propellant gases immediately behind the bullet disperse into the absorbent elements, radiating outwardly into these elements from the axial passages **27** through the elements. These propellant gases enter and disperse through the aforementioned tortuous passages defined by the knitted wire material of the absorbent elements, lowering the temperature of the propellant gases by heat transfer to the wire mesh. The energy of the propellant gases within the suppressor thus is reduced as the gas temperature falls, so that the noise or report of firing is muffled or significantly suppressed as the bullet exits the axial passage **23** through the encapsulator **21** at the front end of the suppressor.

The use of separate absorbent elements **26***a*–**26***c*, set off by the baffles **30***a* and **30***b*, improves the energy-absorbing capability of the suppressor by providing separate energy-absorbing sections for receiving the expanding propellant gases following the bullet through the suppressor. The exposed end surface areas **32** of confronting adjacent baffles, aided by the space **35** between adjacent absorbing elements, increases the exposed surface area of the absorbent elements available for gas dispersion within the suppressor, and thus further enhances the noise-suppressing effectiveness of the suppressor without need for increasing the size of the absorbent elements.

It should be understood that the foregoing pertains only to a disclosed embodiment of the present invention, and that numerous changes and modifications therein may be made without departing from the spirit or scope of the invention as set forth in the following claims.

I claim:

1. Sound suppressor apparatus for use with firearms, comprising:

a tubular housing attachable to the muzzle end of a firearm and defining a hollow interior surrounding a path along which bullets can travel;

plural elements of absorbent material disposed along the hollow interior of said housing, in surrounding relation with said bullet path, said absorbent material operative to receive propellant gases in the bullet path;

a fluid-impervious baffle between at least two adjacent elements of absorbent material and covering substantially less than the entire end area of said adjacent elements, leaving the uncovered end area portions exposed to permit passage of gas therethrough, thereby increasing the absorption of propellant bases and attendant reduction in energy of the gases;

each said fluid-impervious baffle having substantially a triangular shape and having a plurality of discrete radial extremities which contact the inside of said tubular housing and maintain said baffle centrally located across the hollow interior of the housing; and

each baffle having noncircular outer edges extending between said radial extremities and spaced radially inwardly from said extremities, so that the regions between said noncircular outer edges and the inside of the tubular housing are unoccupied by the baffle and thereby leave uncovered the adjacent end portions of said absorbent elements, allowing the gases to flow through the uncovered end portions and thereby increasing the reduction in the energy in the gases.

2. Suppressor apparatus as in claim **1**, wherein:

said baffle maintains a space between said uncovered end area portions, so as to facilitate gas flow through the uncovered end area portions.

3. Suppressor apparatus as in claim **1**, wherein said baffle is an elastomer, and said elements of absorbent material are nonelastomers.

4. Suppressor apparatus as in claim **1**, wherein:

each said element comprises a knitted wire mesh.

5. Suppressor apparatus as in claim **4**, wherein each said element comprises an annulus having a hollow interior to provide a longitudinal passage aligned with the bullet path, and surrounding said path with a labyrinth of interstitial spaces defined by said knitted wire mesh, so that propellant gases from the firearm enter the interstitial spaces and a portion of the energy in the gases thus is absorbed.

6. Suppressor apparatus as in claim **5**, wherein said wire mesh is a relatively good thermal conductor so as to absorb heat in the propellant gases, thereby reducing the energy of the propellant gases.

7. Suppressor apparatus as in claim **1**, wherein:

said baffles have sufficient thickness to maintain a gas-receiving space between the uncovered portions of the confronting end surfaces of adjacent absorbing elements.

* * * * *

rexresearch.com

Telepathy Patents

MAKER.IO POWERED BY DIGI-KEY
Learn More about the
Circuit Playground - Developers Edition

Tesla Generator? $49 - Build Your Own Generator
Eliminate Your Power Bill Easy Do It Yourself. Great Discount Now! tesla-energy.org/Generator

3 Plants Remove Deep Fat
How These 3 all Natural Plants Increase AMPK and Clean Out Deep Fat From Your Body deepfatsolution.com

Free Tungsten Guidebook
Tungsten Preparation Guide. Download Your Copy Today! diamondground.com

Death & Obituary Records
Enter any name and search free! Find their obituary now. ancestry.com

Tesla Generator? $49 - Build Your Own Generator
Eliminate Your Power Bill Easy Do It Yourself. Great Discount Now! tesla-energy.org/Generator

Free Tungsten Guidebook
Tungsten Preparation Guide. Download Your Copy Today! diamondground.com

3 Fatigue-Causing Foods
Boost Your Energy Level Every Day By Never Eating These 3 Foods haveenergyatanyage.com

Take Personality Test
What's Your True Personality? Take The Test Now To Find Out. yoyoquiz.com

---

US2011127346 -- Co-creative Learning and Research Method and Apparatus

RU2157659 -- METHOD FOR INVESTIGATING TELEPATHY COMMUNICATION

RU2127546 -- METHOD OF EXTRASENSORY PHENOMENON EXAMINATION AND DEVICE FOR ITS EMBODIMENT

HU9901402 -- TELEPATHY TABLE GAME

GB222163 -- Improvements in or relating to apparatus for the development of concentration and telepathy

FR2661295 -- Transmitter/receiver of gravitational waves

JP2014155432 -- SUPERNATURAL POWER

CN101513296 -- Natural telepathy crystal earbob

US2011127346
Co-creative Learning and Research Method and Apparatus

Inventor: VON BEZOLD ERNST

A method is claimed to provide experiential nonlocal real time interaction between or among a participant or participants, observer or observers, participant-observer or participant-observers, and phenomenon or phenomena by means of presentation of sensory information or semantic mapping, by direct observation or using sensory or semantic representation for example by imaging methods or technology including without limitation, image projection over a distance, e.g., by internet, television or telepathy, so arranged as to permit and facilitate rapid or real time 'simultaneous' or prescient attention to the phenomenon and interaction with it, so as to facilitate contemporaneous, or time-delay, or time-advance "feedback" active input to the phenomenon or the phenomenon presentation system, in some sense constituting thereby an entraining and entrained system being modulated by the attention and energetic or related state and interaction or resonance with the phenomenon, of the participant-observer(s), and furthermore their resonance or interaction, whether attentively or consciously intended or consciously or not, with one another and collectively.

[0001] A preferable embodiment of this method uses 1) a phenomenon which may vary on account of, without limitation, participant a) attention or b) intention or c) emotion focused by other means or d) thinking or e) a permutation or combination or permutations or combinations of the foregoing;

[0000] 2) means, as hereinafter illustrated, to present the participant with status data about the phenomenon or phenomena including, as phenomena, status data about observer/participants; said status data being in a preferred embodiment as current as possible ('real time') status data, as well as past status data (history) or preview, forecast, projective, or hypothetical data.

3) means to assist the participant in interpreting or understanding a co-relationship between characteristics of the phenomenon and attributes, including the immediate process experience during phenomenon data generation, of the participant, including participant attentional behaviours, attitude mood and emotions, thinking, intention and bodily processes and status, especially the pattern of health.

4) means to provide the participant-observer as user with contextual information and variable, including without limitation user-directed or influenced, information
about the state of the system as a whole, both in particular the phenomenon as such as well as user(s) status data, on a data-intensification and withdrawal training basis to develop and optimise the development of user capabilities for remote sensing and for "remote" imaginative perception, inspiration and intuiting, i.e., directly perceived understanding of remote phenomena in part or totality and their course and development without sense-perceptible or imaginatively correlated experience.

4) Means to provide groups of participant with weather and water quality feedback to enable them to change their participation at the time, in the past and moving forward, so as to modify the presented or selected phenomena collectively, for example water quality, especially to improve it, precipitation patterns, ocean current behaviour, and so forth; and to work similarly with other living and inorganic phenomena to improve and stabilise the biosphere, lithosphere, atmosphere, photosphere of interest, that environment of greatest interest being for the time being that of this planet which is our home; especially to enable conscious and intentional ameliorative direct attentive perception and interaction of large numbers of planetary residents in respect of stabilising and otherwise ameliorating climate change by intelligent collective selective conscious interaction with the root or the proximal causes of climate change.

BACKGROUND

[0002] Telepsychokinesis is a scientific fact. This is illustrated by, for example, the work of Marcel Vogel documented in the Proceedings of the Second International Psychotronics conference in Prague, Czechoslovakia, 1973 (a conference also contributed to by Hal Puthoff, William Tiller, and Shafica Karagulla among many others), and the classic early work of J. B. Rhine and others at Duke University and later members of the Parapsychological Association which is a part of the American Association for the Advancement of Science.

[0003] Snow crystals forming under locally physically similar conditions are widely held to be highly variable, and this because their formation, and likewise the associative structure of water as a fluid, depends on a weak, and therefore sensitive, bond between hydrogens associated with an oxygen atom, attraxted to nearby oxygens with surplus electric charge to share. This writer asked in a Materials Laboratory poster presentation to material scientists who included visiting researchers and international delegations from the US, Germany, Japan, Soviet Union, etc., at the Fourth International Fracture Conference, Waterloo Ontario Canada, 1977: "Can Loving Attention Affect Snow Crystal Growth?"

[0004] The cumulative weight of scientific evidence prior to that has demonstrated, for example with Copper II chloride crystal solutions doped with fluid extracts such as blood or plant juice and grown in for example Petri-dish sized quiet environments, that under controlled conditions, the crystal forms a homeomorphic image of the donor organism which can and has been used for medical diagnosis and in the case of plants, for quality control. See for example Sensitive Crystallization Processes A Demonstration of Formative Forces in the Blood, by Ehrenfried Pfeiffer; second edition, edited by Erica Sabarth and Henry N. Williams, MD. Anthroposophic Press, Spring Valley, N.Y., 1975. The purpose of this technology, as Pfeiffer further elucidated in a memoir, was primarily and strategically to enable the experimenter to develop and apply the perceptive power of thinking to intuit the relationship between the crystal formed and the individual or plant, etc., affecting it through admixture of a fluid sample, thus coming "through creative intuition" to recognise directly (compare dhyana in the yoga supras of Patanjali) the, as it were, 'noumen behind the phenomenon'; and only secondarily to apply the specific finding to deliver a diagnostic outcome.

[0005] A substantial body of work has demonstrated the sensitivity of water to environment, and notably, a human psycho-kinetic influence on water, both through plant growth experiments and through effects on the physical properties of water such as infrared absorption and light scattering. See for example the extracurricular work of Ed Brame, spectroscopist, in Wilmington Del. at E I Dupont de Nemours, highlighted in Krippner, S. and A. Villoldo: The Realms of Healing; Celestial Arts, Millbrae, Calif., 1976.

PRIOR ART

[0006] Besides the foregoing examples, mass experiments have demonstrated dream telepathy (Dream Telepathy. By M. Ullman and S. Krippner of Maimonides Medical Center Dream Laboratory, Brooklyn, N.Y.; 1973) and mass prescient psychokinetic influence at the Princeton University REG remote event detection project (Robert Jahn et. al.), notably prior to the mass trauma attendant on the attacks Sep. 11, 2001 on the World Trade Center, New York, and on the Pentagon, Washington, D.C.

[0007] The concept of influencing ice crystal formation by intention, prayer, human energy, has been presented and widely promoted, with copious plausible imaging, though methodologically stringent reproducibility remains a question as of this writing, by the Japanese hado researcher M. Emoto.

[0008] More reliably as of this writing, the work of William A. Tiller et al. and their series of substantial monographs on highly reproducible distant psychokinesis replicated using as mediating devices, oscillators previously programmed (‘energised’) by meditators, shows the potential malleability of physical systems in the face of human mental concentration and intent (www.tiller.org).

[0009] Recent work has continued to demonstrate results consistent with the foregoing. An intention experiment project was organised by Lynne McTaggart (author of The Intention Experiment; Free Press, New York, 2007) with a team of experimental scientists and hundreds of volunteers distant from the experimental sites. They demonstrated and studied experimentally, changes in properties of systems using water, plants. The experimenters used the internet to recruit and facilitate shared intention, and shared timing—synchrony—of participants seeking to influence the experimental systems at a distance, e.g., for water luminescence, for healing, etc. As of this writing, see **http://www.theintentionexperiment.com/the_experiments**.

[0010] The foregoing findings and reports represent a part of a larger paradigm shift now taking place in world science and culture, prefigured by the uncertainty principle of Heisenberg and the explicitly interactionist epistemologies of nineteenth and twentieth century dialecticians, some of whose revolutionary political application of such a powerful principle as interactivity significantly polarised and probably delayed its more widely spreading, deeper scientifically comprehensive application, e.g., to spiritual research and application in, e.g., human healing and environmental renewal, into the post-Communist period at the end of the twentieth century. In effect, the recent developments and inventions develop and address a previously undervalued potential which has been overhanging the market.

DESCRIPTION OF THE INVENTION

[0011] A simple preferable embodiment of the invention starts out by making use of a reagent system, for example, similar to the many described above, initially to provide real-time feedback to the participant. In this example an automatic (therefore requiring no attention to operate the physical environment) self-cleaning snow crystal growing system is used to present a snow crystal growing site, or a sense-perceptible representation such as a visual, auditory, tactile, thermal, etc., image or images of a site, to a participant, by means such as a video display or television, amplified filtered sound, etc., for example by projection (e.g., onto a surface, such as a screen or the moon), on the internet, etc.

[0012] The participant is able to observe the particular phenomenon, i.e., here, the snow crystal as it develops. In a preferred embodiment the participant is asked to bring impartial appreciation to the process of observing, rather than unaffective neutrality or any further specific intention.

[0013] As a nonessential clarifying feedback facilitation method using a multiple control design adapted from Ullman and Krippner (1973, op. cit.), the participant may be presented for comparison, control crystals grown under comparable conditions, i.e., at the same time in more or less the same place, with or without a set-theoretically or other analytically based highlight of unique net difference, if any, between the crystal grown under an attentionally or intentionally highlighting condition and the controls.

[0014] As the participant's ability to provide formative facilitation to the crystal becomes more evident, the presentation may optionally be attenuated or intensified, preferably by user-directed modulation, to encourage imaginative perception of the growing crystal leading to clairvoyant perception of the forming crystal independent of the presentation. Without limitation, a coordinate remote viewing application is another potential interactive resonant or ‘feedback’ path. It will be evident from remote viewing literature of the twentieth century (e.g., Rudolf Steiner; Robert A. Monroe, David Moorehouse, Courtney Brown) that selecting the ‘time’ of viewing extends the potential of croos-time interaction or resonance research from an instrumental quantum level to the small-scale macroscopic system of the growing snow crystal.

[0015] The therapeutic feedback for patients and therapist and the artistic potential of snow crystal growing are among the motivating factors which in this example increase the participant's energetic concentration and compliance with plausible and reasonable protocol.

[0016] At this point, in respect of the participant's continued participation, large numbers of participants who may or may not have worked ‘together’, whether at the ‘same’ time or not, with snow crystals or some similarly sensitive system, may work together to love snow crystals as they fall onto land, lakes, rivers, etc. The literature on love and the healing response of plants to loved water, would motivate such work, and participant's direct perception would provide feedback, along with local snow and water sampling trials, to contribute to a basis for judgement of effectiveness.

[0017] Similarly, sufficiently skilled participants, having developed love-based intuition and exact clairvoyance using this system, would be in a developing position en masse to guide local and planetary weather to greater harmony and stability.

RU2157659
METHOD FOR INVESTIGATING TELEPATHY COMMUNICATION

Inventor: BERESNEV V N / RJAZANOVA M P

SUBSTANCE: method involves determining and processing bioenergetic signals of percipient before and after mental suggestion of information sent by inductor person. Percipient beta- or gamma-radiation level is treated as the bioenergetic signal. The degree, to which this radiation type increases, is used for making conclusions concerning percipient response to telepathic action.
EFFECT: high accuracy of measurements.

METHOD RESEARCH

The invention relates to the engineering psychology and medicine and can be used during the sessions suggestive influence to determine the effectiveness of telepathic communication with the inductor percipient.

The most efficient in its use energobioterapevticheskih purposes, as well as in determining the ability of the inductor to provide geterosuggestivnoe impact on the percipient.

There is a method of research percipient telepathic communication with the inductor, providing for EEG before and after a session of telepathic communication with the subsequent interpretation of the electroencephalogram (AP Dubrov, Lee AG Modern problems of parapsychology. Moscow, Publishing House Foundation parapsychology them. LL Vasileva, 1998, p. 148-161).

At the same time the existence of telepathic communication with the inductor percipient judged by sync rhythms electroencephalogram (MN Livanov, V. Ananiev Elektroentsefaloskopiya. Med. ed., 1959).

However, as noted in these sources, the method is sensitive to numerous interference and crosstalk, complicates processing and interpretation of the data.
In addition, the subjective interpretation of the results of the EEG reduces their credibility.

It is also known research percipient telepathic communication with the inductor, providing for the exchange of energy-sources and characteristics of torsion fields, due to the spatial orientation of the spins of elementary charges (Stavitskiy VI from captivity energy performances. St. Petersburg, Publishing House of the "Politehnica", 1997, p. 32).

However, this method is extremely complex technical implementation and interpretation of the results, because the essence of the torsion field is beyond the scope of energy-makropredstavleniya exchange.

Immediate removal of the electrical signals from the percipient in the dynamics imposed on the surface of the body percipient between the two selected points of the electrodes connected to an electrical outlet equipped with a self-tuning filter (US 5,029,590, A 61 K 5/04, 1991), also has a low accuracy because of the wide variation of bioelectric signals and significant external and internal interference.

The closest to the claimed method is the study of telepathic communication percipient with the inductor, providing a continuous determination and subsequent treatment of bio-energy signals percipient, is removed from electrodes placed on the surface of the body, using a polytron in active and passive phases before and after mental suggestion information inductor percipient (RU 2098017, A 61 B 5/16, 5/04, 1997).

The degree of telepathic communication is judged by the following formula:
where W - the level of perception of telepathic signal;

M1 and M0 - average values ??of output current, respectively, in a polytron active and passive phases of telepathic influence; & sigma; 1 and & sigma; 0 - standard deviations of the current in the active and passive phases.

The disadvantage of the prototype method is to use a non-standard and therefore not easily accessible means of measurement, which is the Polytron.

The technical problem of the process - to expand the arsenal of the means of implementation and improvement of standard measurements.

The solution of this problem lies in the fact that in the process of investigation of telepathic communication percipient with the inductor, the definition and the subsequent processing of bio-energy signals percipient before and after mental suggestion information inductor percipient as bioenergy signal determine the level of & beta; - and / or & gamma; radiation percipient, and the degree of increase in the level of radiation is judged on the reaction of the percipient on telepathic.

At higher levels of & beta; - and & gamma; radiation by 30% or more is judged on a strong degree of telepathic communication.

Thus, the proposed method is based on the first time specified by the author previously unknown phenomenon of increased levels of & beta; - and & gamma; percipient radiation in response to the telepathic.

FIG. 1-8 shows the variation of average values beta; - and gamma; percipients radiation - patients with various diseases - in response to mental suggestion information from the inductor (energobioterapevta) to the following examples 1-8 respectively.

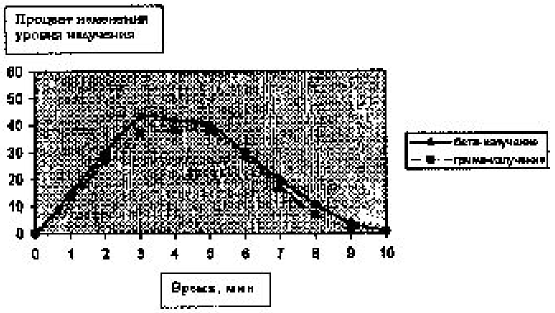

Фиг. 2

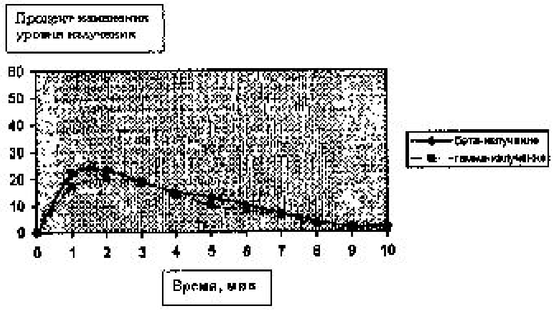

Фиг. 4

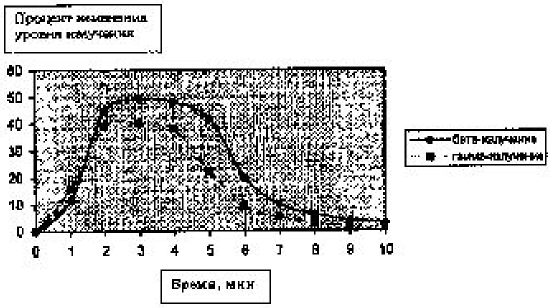

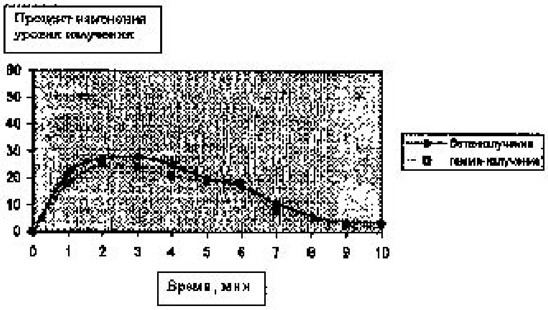

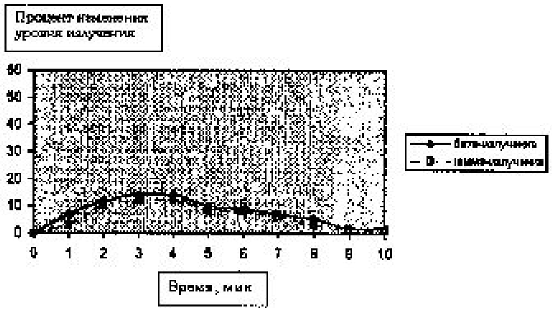

Фиг. 6

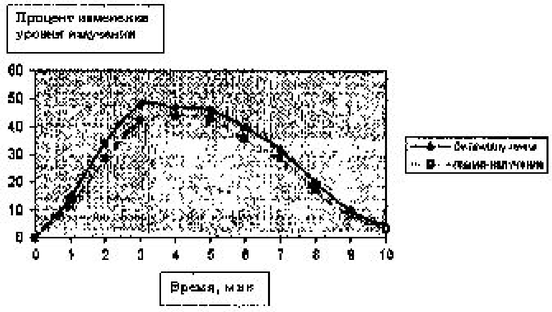

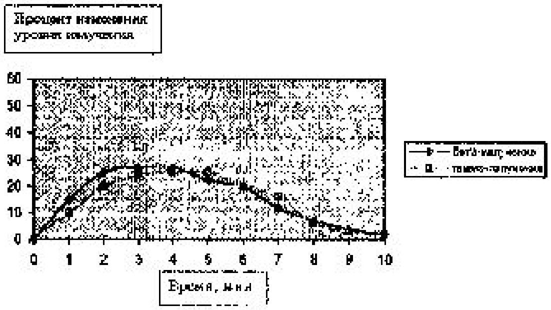

Sessions telepathic communication were carried out in addition to symptomatic treatment in the Scientific and Medical Consultation Center (St. Petersburg).

As a reference method, a method of the prototype, which is defined, in particular, the ability of the inductor to provide a telepathic impact on the percipient.

Example 1.

Patient F., '42, entered with a diagnosis of psoriasis. OBJECTIVE: rashes ranging from small dots to child palm over the body, gastroduodenitis, biliary dyskinesia, pancreatitis. Spend 15 sessions of telepathic communication (TC) bioenergy with the percipient weekly by mental suggestion information from the installation to restore the functions of the digestive system. Each treatment was carried out for 1 hour, twice the rate of 4-5 m geterosuggestivnoy percipient connection with the inductor 50 minute intervals under the control of levels of & beta; - and & gamma; percipient radiation using a portable universal radiometer RUP-1 (& beta; -radiation) and dosimeter DRGZ-02 (& gamma;-radiation). Before mental suggestion background level of & beta; -radiation is & beta; 0 = 20 & plusmn; 1 min-1 & bull; cm-2, and & gamma; radiation - & gamma; 0 = 0.11 & plusmn; 0.02 mc / s. Within 2 minutes from the start of the session is an increase of the level of & beta; and & gamma; radiation percipient to & beta; max = 27 & plusmn; 3 min-1 & bull; cm-2 and & gamma; max = 0,145 & plusmn; 0,02 mc / s, which is based on the percentage increase in the radiation level from baseline at 37 and 32 %, respectively. After removing the effects of background values ??geterosuggestivnogo levels & beta; - and & gamma; radiation are reduced by 4 min (Fig. 1). After the sessions of telepathic effects rashes on the body for 2 years did not occur. According to the results of analyzes, gastro and pancreatitis are not found, the function of the biliary tract is restored.

Example 2.

Patient N., '32, entered with a diagnosis of duodenal ulcer, pancreatitis. Spend 10 sessions weekly telepathic communication, as in Example 1, with the duration of exposure geterosuggestivnogo for 5 minutes with the installation of ulcer healing and restore functions of the digestive system. It recorded a high degree of telepathic communication percipient with the inductor to increase the levels of & beta; - and & gamma; radiation from the background (& beta; 0 = 21 & plusmn; 1 min-1 & bull; cm 2; & gamma; 0 = 0,10 & plusmn; 0,02 mc / c) 43 and 38% respectively. Duration Transient (time to reach steady-state value level & beta; - and & gamma; radiation, measured from the beginning, and after removing the effects of geterosuggestivnogo) - 2,5 & divide; 3,5 min (Fig.
2). According to the results of clinical tests of duodenal ulcer healed, the pancreas is normal.

Example 3.

Patient J., 37 l., Was admitted with a diagnosis of dystonia, accompanied by headaches, seizures of facial muscles and limbs. Spend 10 sessions of telepathic communication 2 times per month, as in Example 1, with the duration of exposure geterosuggestivnogo for 4-5 minutes with the installation of ulcer healing and restore functions of the digestive system. It recorded a high degree of telepathic communication percipient with the inductor to increase the levels of & beta; - and & gamma; radiation from the background (& beta; 0 = 21 & plusmn; 2 min-1 & bull; cm 2; & gamma; 0 = 0,12 & plusmn; 0,02 mc / c) 49 and 40% respectively. Duration Transient - 2-3 min (Fig. 3). As a result, cramps face and extremities, and headaches, the patient stopped (the period of 3 years).

Example 4.

Patient L., 45 l., Was admitted with a diagnosis of chronic konkrementny holestsestit (cholelithiasis), biliary dyskinesia. Spend 20 sessions weekly telepathic communication, as in Example 1, with the duration of exposure geterosuggestivnogo for 4 minutes with the installation of functional recovery of the bile ducts and the dissolution of the stone. It recorded a weak degree of telepathic communication percipient with the inductor, since an increase in levels of & beta; - and & gamma; radiation from the background (& beta; 0 = 20 & plusmn; 1 min-1 & bull; cm 2; & gamma; 0 = 0,09 & plusmn; 0,01 md / s) at 23 and 20% respectively. The duration of the transients 1-4 min (Fig. 4). By results of US rock reduced in size, but not completely disappeared. In this example, the course of treatment was not effective enough that corresponds with the weak response of controlled fields & beta; - and & gamma; radiation patient.

Example 5.

Patient P., '42, entered with a diagnosis of chronic viral hepatitis B. Objective: based on the results observed ultrasound sealing and deformation of the walls of the gallbladder; liver - slices large- infiltrates within various circuits; blood levels of bilirubin - 35 mmol / l and transaminase - 4.6 mmol / l - higher than normal. Spend 12 sessions of telepathic communication 2 times a week, as in Example 1, with the duration of exposure geterosuggestivnogo 4-5 minutes to install to restore the function of the gallbladder and liver. The degree of telepathic communication percipient with the inductor is low, since increasing levels of & beta; - and & gamma; radiation from the background (& beta; 0 = 20 & plusmn; 1 min-1 & bull; cm 2; & gamma; 0 = 0,08 & plusmn; 0,01 mc / c) is 28 and 25% respectively. The duration of the transients (Fig. 5) is on the leading edge (from the start of exposure geterosuggestivnogo) - 1.5 min, at the trailing edge (after removal of exposure) - 4 min. After the designated communication procedures geterosuggestivnoy blood levels of bilirubin - 10.5 mmol / l transaminase - 0.2 mmol / l. According to the US state of the liver and gall bladder improved by reducing the number of infiltrates in the liver.

Example 6.

Patient P., '62, entered with a diagnosis of lymphocytic leukemia. The content of leucocytes in 1 liter of blood - 15.0 & bull; 109 lymphocytes - 78 & bull; 109.
Spend 7 telepathic communication sessions on a weekly basis, as in Example 1, with the duration of exposure geterosuggestivnogo 5 minutes to install on the functional recovery of hematopoiesis. Degree telepathy percipient the inductor is low, since increasing levels of & beta; -, and the radiation from the background is 12 and 10% respectively. The duration of the transients (Fig. 6) is on the leading edge 2 min, falling edge - 4 min. After the appointed procedures geterosuggestivnoy communication blood levels of white blood cells - 11,9 & bull; 109, lymphocytes - 64 & bull; 109. Thus, our treatment was ineffective, that corresponds with a slight deviation of & beta; - and & gamma; relative to the background radiation.

Example 7.

Patient S., '22, entered with a diagnosis of asthma, chronic bronchitis. OBJECTIVE: recurrent attacks of breathlessness and bronchospasm about 1 time per month, allergic to dust, debris, animal dander. The complaints of pain in the liver and pancreas. 15 sessions carried telepathy weekly as in Example 1, 2 times per month at geterosuggestivnogo exposure duration for 5 minutes with the installation of the restoration of the liver, pancreas, bronchus, lung. The degree of telepathic communication percipient with high inductor, since an increase in levels of & beta; - and & gamma; relative to the background radiation is 47 and 44%, respectively. The duration of the transients (Fig. 7) is on the leading edge 3 min, on the falling edge - 4 min. As a result, dyspnea and bronchospasm, and pain in the liver and pancreas stopped. During two years of follow up there was no recurrence.

Example 8.

Patient S., 25 l., Was admitted with a diagnosis of diabetes with blood sugar 23 mmol / L.
20 sessions carried telepathy weekly as in Example 1, weekly during the duration of exposure geterosuggestivnogo 5 min with setting at restoring pancreatic functions.
The degree of telepathic communication with the percipient inductor is low, since an increase in levels of & beta; - and & gamma; relative to the background radiation is 27 and 25%, respectively. The duration of the transients (Fig. 8) is on the leading edge 2 min, falling edge - 4 min. As a result, blood sugar dropped to 6.5 mmol / l. Comparative results of determining the presence of telepathic communication with the inductor percipient Examples 1-3, 7 and 8 of the proposed method is confirmed by studies using Polytron prototype method.

In Examples 4-6, characterized by low levels of change & beta; - and & gamma; radiation percipient, in a study on sustainable polytron telepathic connection is detected.

As can be seen from the above examples, the use of the method, instead of the prototype expands the arsenal of methods and means of research of telepathic communication.

This new way of standard and simple, because it is implemented on the basis of common funds radiation dosimetry techniques are used instead of the prototype Polytron and correlation analyzer, performing processing of the results by the formula (1).

The technical result derived by means of the simplification process is the availability of its implementation in the field, as well as reducing the cost of the analysis.

RU2127546
METHOD OF EXTRASENSORY PHENOMENON EXAMINATION AND DEVICE FOR ITS EMBODIMENT

Inventor: BLATOV I V

FIELD: medicine; electroencephalography.
SUBSTANCE: for examination of extrasensory phenomena, for example, telepathy, when performing experiment in mental transmission of images, it is proposed to take encephalograms of inductor (information transmitting person) and at least of one recipient (information receiving person) synchronously. Inductor's encephalogram is compared with that of each of recipients at information transmission moments and in periods when no transmission is performed. Comparison of encephalograms taken at analogous points of surface of examined persons' heads is carried out by computer processing with use of different mathematic models. Device for embodiment of the method includes two units of buffer storage.
EFFECT: possibility to study brain work of inductor and recipient in process of information transmission and reception.

METHOD RESEARCH

The invention relates to medicine and, in particular, to the field of brain study electrical processes - electroencephalography.

It is now considered scientifically proven the existence of extrasensory information transfer from one person to another, that is, the fact of transmitting information by known channels (visual, auditory, etc.) Using subliminal signals that are perceived by the subconscious directly, without going through a process of awareness, either kakim- the other has not yet been studied or little studied channels.

Psychic information transfer observed in different areas (diagnosis and treatment of diseases, telepathy, etc.).

The largest number of science-based experiments, confirming the extrasensory information transfer took place in the field of telepathy (AP Dubrov and VN Pushkin.
Parapsychology and modern science. - M .: Sovaminko, 1989, p.133).

A known method of investigation of telepathic transmission of information is that the group of subjects consisting of the inductor - a person transmitting information, and, at least one recipient - the person receiving the information, given the task to psychic transmission and reception of information by transmitting imagery imposed inductor using Zener cards, which show sharply differing shapes (circle, square, cross, etc.).

Typically, the experiment used 5 such cards chosen randomly.

A recipient who does not have the touch with no inducer connection and located usually in another room, receives and stores mental images Zener cards passed to it inductor.

Telepathic communication between the inductor and the recipient is considered to be established if the recipient is in a series of 100 experiments right call not less than 80% of the transmitted image (In the absence of likelihood of correct telepathic communication "guessing" of the image of 5 is only 20%).

A drawback of known methods for studying psychic phenomena is that they lack the operations aimed at the study of brain inductor and the recipient during transmission and reception of information.

The present invention seeks to overcome this disadvantage.

The essence of the invention lies in the fact that the known sequence of operations used in the research is the fact that the group of subjects consisting of the inductor - a person transmitting information, and at least one recipient - the person receiving the information, given an assignment to psychic transmission and reception of certain information, such as by Zener cards and the content information is given at random and the presence of extra-transmission information is judged by the number of correctly received by the recipient of information units, a number of new operations, namely:

1. Simultaneously with the presentation of the inducer and a mental maps Zener transmit this data, selected at random, all the participants in the experiment (the inductor and the at least one recipient) was deprotected EEG electroencephalogram inductor and then compared with each of the EEG recipients in the Validity Period information transmission and non-transmission periods.

2. Produce computer processing electroencephalograms inductor and recipients taken at a similar point in head reliability of information transmission periods and during periods of non-transmission with the use of various mathematical models, such as calculating a correlation coefficient of electrical brain processes inductor electrical brain processes each of recipients and coherence spectra of electroencephalograms in different frequency bands, and then these indicators calculated for the validity period of information transfer, are compared with those calculated for the periods of non-transmission information, and the results of calculations build generalized topographic maps of brain inductor to each recipient, which applied indices calculated using interpolation and color painting.

All the well-known mathematical models currently used for the analysis of electrical brain processes only one patient.

Unlike the present process is that known processing methods electroencephalogram, which are based on different mathematical models, it is proposed to use for comparison and analysis of brain electrical processes more than one patient, and at least two patients: the inductor and the recipient.

Below, exemplary methods using known computer processing of the EEG of the inductor and one of the recipients given calculation synchronism electrical processes, characterized by a correlation coefficient.

Calculations are made for each point of the electrodes at the inductor and recipient.

For example, electrical processes at Fp1 at the inductor compared with the electrical processes in the same point of the recipient; in the same way electrical processes are compared at all other points.

The correlation coefficient is calculated by the formula where U1i - the current value of the voltage (potential difference) between the point of the study (for example, Fp1) and reference point (eg, earlobes) in the inductor;
U2i - the current value of the voltage in a similar point in the recipient;

U1sr - the mean value of the voltage at the point in the study of the inductor calculated for analysis according to the epoch
U2sr - average value of the voltage at the point under the recipient;
n - the number of dots per epoch analysis (for example, 4 seconds).

Similarly, one can compare the electrical brain processes inductor electrical processes marrow recipient and using other mathematical models.

3. Based on values ??resulting from the previous operation, build generalized topographic maps of brain inductor with each of the recipients, which is applied to the computed metric, using interpolation and color painting.

To construct a generalized topographical map of the circuit of the head (usually in the form of a circle or oval) applied anatomical landmarks (nose and ears), and the points at which measurements are made (typically in terms of standard electrode placement, such as the system 10 - 20).

After each of the electrodes at the point of the inductor and the recipient will be calculated indicator values ??determined by the selected mathematical model, they are applied to a generalized topographical map by colors, wherein each hue value is assigned to its index.

The ratios in the intermediate points (points located inside the square at the vertices of which the electrodes) is determined by interpolation.

Furthermore generalized maps can be constructed (e.g., by the nonparametric Mann-Whitney test) difference also maps showing the increase (or decrease) the calculated parameters in the reliability of information transmission period as compared with the background (non-transmission information).

These cards allow the difference to determine which parts of the brain of the inductor and recipient are involved in psychic transmission.

With the participation in the experiment several recipients, generalized and difference map can be constructed also by comparing every two recipients. This option cards can be useful in the study of the brain in group hypnosis, group therapy, anger management, etc.

4. Simultaneously with the transfer of imagery and inductor during periods of absence of such a transfer, it impose visual or auditory stimuli (light flashes, clicks, chessboard with changing fields, etc.), And then after each stimulus from the inductor and recipient are measured caused by these stimuli potentials.

Unlike previous operations, aimed at studying the electrical processes in the brain that are integrated over a relatively long period of extrasensory information transfer, this operation will allow to investigate the ultrafast electrical processes.

Computer processing of evoked potentials of the inductor and recipients can be made by analogy with computer processing of conventional EEG, a technique which is described in Sec. 3 (construction of generalized and difference maps recipients.

It should be noted, however, that the implementation of this operation requires a particularly sensitive special equipment to measure evoked potentials.

Thus, compared with known methods studies psychic effects the proposed method allows to study brain function during extra-transmission information, wherein the area of ??its application may not be limited only to research area psychic effects, and extended to a number of adjacent areas, where the prevalent sensory channels information (the study of the brain in learning, hypnosis, group therapy, studied the phenomenon of the "crowd", etc.).

The proposed method of investigation of psychic phenomena, in principle, be implemented using normal commercially available EEG instrument.

However, due to the fact that all the standard EEG electroencephalogram designed for removing only one patient in the study of several patients require several electroencephalographs (the number of participants of the experiment).

Moreover, difficulties arise with the timing of these electroencephalographs, especially in those cases where an experiment is carried out in several locations.
Even greater difficulties arise in this case in the construction of the generalized and difference maps.

Known electroencephalograph Blatova-Risse (application N 93039980/14/039709 on 06/08/93, the decision to grant a patent on 20.11.95), which is a prototype, which is designed for the simultaneous removal of electroencephalograms just a few patients.

Electroencephalograph Blatova-Risse contains at least two measuring-transmitting systems (IEC), which is placed on patients, and a receiving-recording complex (PPH).

Each IPC consists of electrodes, the amplifiers whose inputs are connected to the terminations of the electrodes, the switch having inputs connected to the outputs of the amplifiers, analog-to-digital converter, an encoder and a transmitter, connected in series with the output switch, the receiver decoder, the two comparison circuits code memory codes stimulator and autonomous power supply wherein the input of the decoder coupled to the receiver output, while the output - to the first input of the comparison codes, the second inputs of which are connected to the outputs of the memory device codes, the output of the first comparison circuit connected to the code control input of the switch and receiving- recording facility includes a receiver, a transmitter and a data processing unit having an input coupled to the output of the receiver, and the output - to the input of the transmitter.

The disadvantage electroencephalograph Blatova-Risse is that the measurement of brain electrical potentials of all patients performed not simultaneously but one after the other.

This is due to the fact that simultaneously with the measurement of the potential of this information is transmitted by radio.

Such a shift in time (up to several milliseconds) between the measurement of potentials in the same locations in different patients do not have a significant effect on the accuracy of the study of normal electroencephalogram, but will make some errors when using evoked potentials.

To eliminate this drawback of the known device, comprising at least two measuring-transmitting complex, located on the patient, and the receiving-recording complex, each measuring and transmitting complex consists of electrodes, amplifiers whose inputs are connected to the electrode terminals, switches, inputs connected to the outputs of the amplifiers, analog-to-digital converter having an input connected to the output of the switch, the encoder and transmitter are connected in series, a receiver and a decoder, connected in series, two diagrams comparing the codes memory codes stimulator and an autonomous power source, with the output a decoder coupled to the first input of the comparison codes, the second inputs of which are connected to the outputs of the memory device codesoutput of the first comparison circuit codes coupled to the control input of the switch and receiving and recording a set consists of a receiver, the transmitter and the information processing unit performed on a standard computer, whose input is connected to the output of the receiver, and the output - to the input of the transmitter, administered two block buffer memory having a first input connected to the output of analog-to-digital converter and the outputs - through an OR gate with the input of the encoder, a second input of the buffer memory are connected correspondingly with the second and third inputs of the first comparator circuit codes.

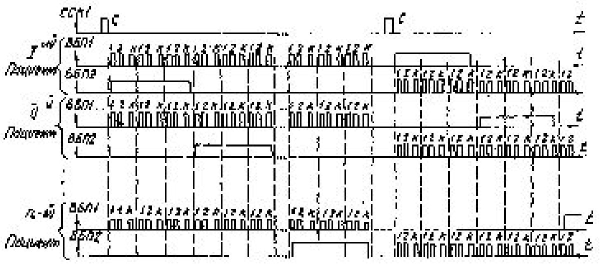

Фиг. 2

FIG. 1 is a functional diagram of the device of FIG. 2 - a diagram for explaining its operation.
Here, I-1, I-2 ...
In - measuring and transmitting complex, located on the patients (n - the total number of patients)
II - receiving and recording facility,
11, 12 ... 1k - electrodes for removing potential
21, 22 ... 2k - Amplifiers (k - total number of measuring channels)
3 - the switch,
4 - analog-to-digital converter,
5 - the first block of the buffer memory,
6 - the second block of the buffer memory,
7 - OR circuit,
8 - encoder
9 - transmitter patient
10 - Receiver patient
11 - a decoder
12 - first comparison circuit Codes
13 - second comparison circuit Codes
14 - memory codes,
15 - stimulator
16 - Autonomous power supply
17 - Receiver recording complex,
18 - transmitter recording complex,
19 - information processing unit (computer).

The device operates as follows.

Potentials, remove the electrodes 11, 12 ... 1k, amplified by amplifiers 21, 22 ... 2k.

The output amplifier is always present current value potential.

The outputs of each amplifier via a switch 3 sequentially in time are connected to the input of analog-to-digital converter (ADC) 4.

In order to ensure synchronous measurement capabilities similar points in all patients receiving and recording facility periodically generates a synchronization command "C" in the form of a code, which is modulated by the carrier frequency emitted by the transmitter 18 and received by the receiver 10 of the patient.
After amplification and conversion by the decoder 11, the received code posting two data inputs of the comparison circuit 12 and code 13, to the second inputs of which are fed from the code memory 14 codes.

When the received code coincides with the code sync applied to the second input of the comparison circuit 12 codes at its first output signal is generated which is fed to the control input of the switch 3.

The command switch 3 (simultaneously in all patients) connects the ADC inputs to the 4 outputs of the amplifier 2 in a strict sequence, starting with the first channel.

At the same time on the same team, synchronizing the second and third outputs of the comparison circuit 12 codes are generated control signals mode of operation of the buffer memory units 5 and 6.

This can be accomplished, for example, a trigger, present at the output of the first comparison circuit 12 codes, the shoulders of which are connected respectively to the second and third outputs and which changes its state on the pitch synchronization.

Mode block buffer memory varies depending on whether or not there is a voltage on their control inputs.

For example, if the voltage at the control input of the buffer memory unit is recording information, and in the absence - in the reading mode.

Therefore, if one block of the buffer memory is recording information, the latter will always operate in the read mode, and the command "synchronization" occurs scheme their operation modes.

From the diagrams in Fig. 2, it is seen that after the arrival of a synchronizing command "C" first memory blocks in all patients switched to the recording information and the latter - to read.

The first memory blocks are recorded, digitized and measured potentials from 1st to k-th channels.

At this time, the second memory block all potential work in the mode of reading information.

Of these time division read information recorded in the previous period (before the command "C") 4.

The read information of memory blocks 7 through an OR gate is input to the encoder 8, which is converted by a serial binary code.

This code is modulated by the carrier frequency of the transmitter 9.

9 emitted by the transmitter signal is received by the receiver 17 of the recording industry.

After amplification and decoding this signal is entered into the computer data processing unit 19, where the computer processing the received information using the above mathematical models.

To start the stimulator 15, the receiving-recording facility generates a command (code) to start the stimulator.

This code is emitted by the transmitter 18 and received by the receiver 10 of all patients, to the input of the comparison circuit 13 codes, the second input of which is fed from the output of the code memory 14 codes.

For each patient, set your own code stimulant.

At concurrence of the received code with the code received from the code memory 14, the output of the comparison circuit 13 codes formed a team to run the stimulator 15.

When transmitting a code stimulant stimulant runs in only one patient.

To start stimulants other patients need to pass their codes.

Thus, compared with the prior art, the device with high precision (to a few microseconds) removing electroencephalograms synchronize all patients, which gives a new device, previously unknown property.

HU9901402
TELEPATHY TABLE GAME

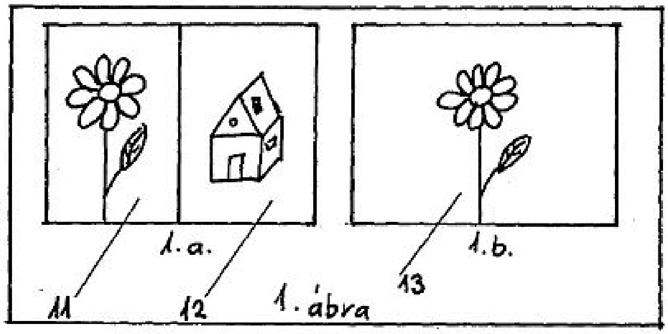
1.a.
1.b.
11
12
13
1. ábra

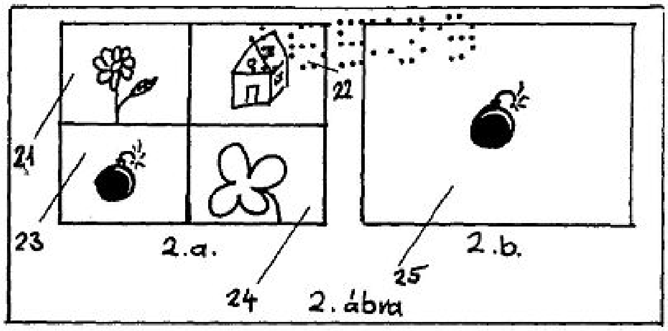
21
22
23
24
25
2.a.
2.b.
2. ábra

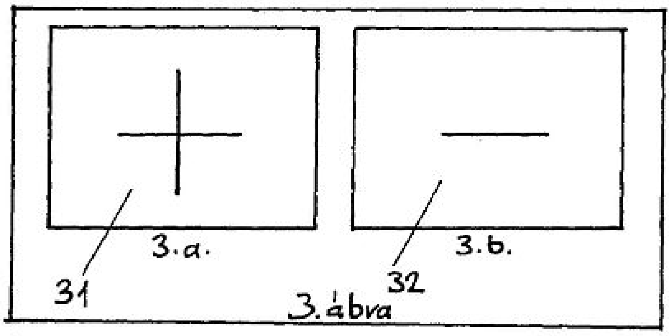
3.a.
3.b.
31
32
3. ábra

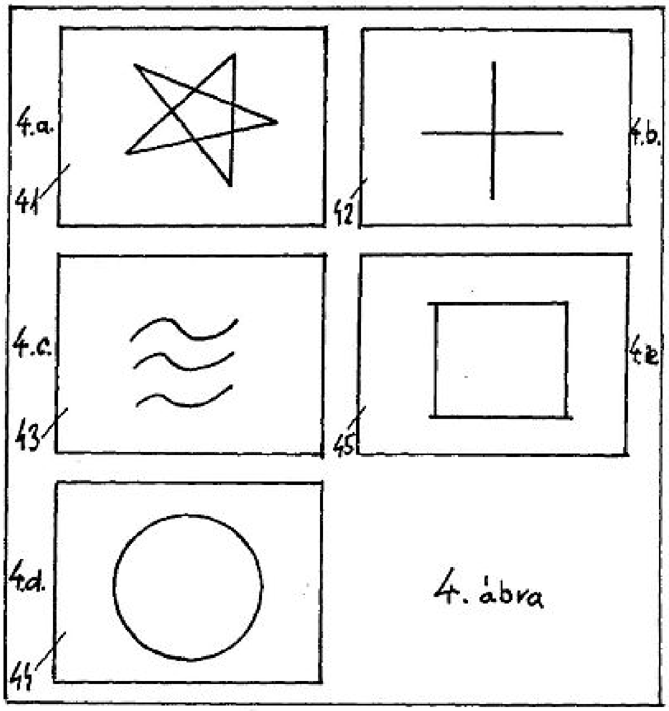
4.a.
4.b.
41
42
4.c.
4.e
43
45
4.d.
44
4. ábra

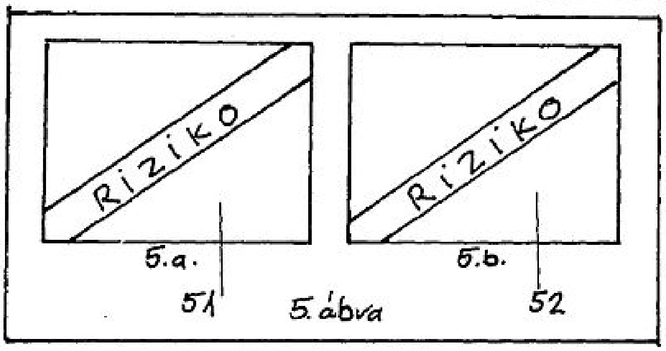
RIZIKO
RIZIKO
5.a.
5.b.
51
52
5. ábra

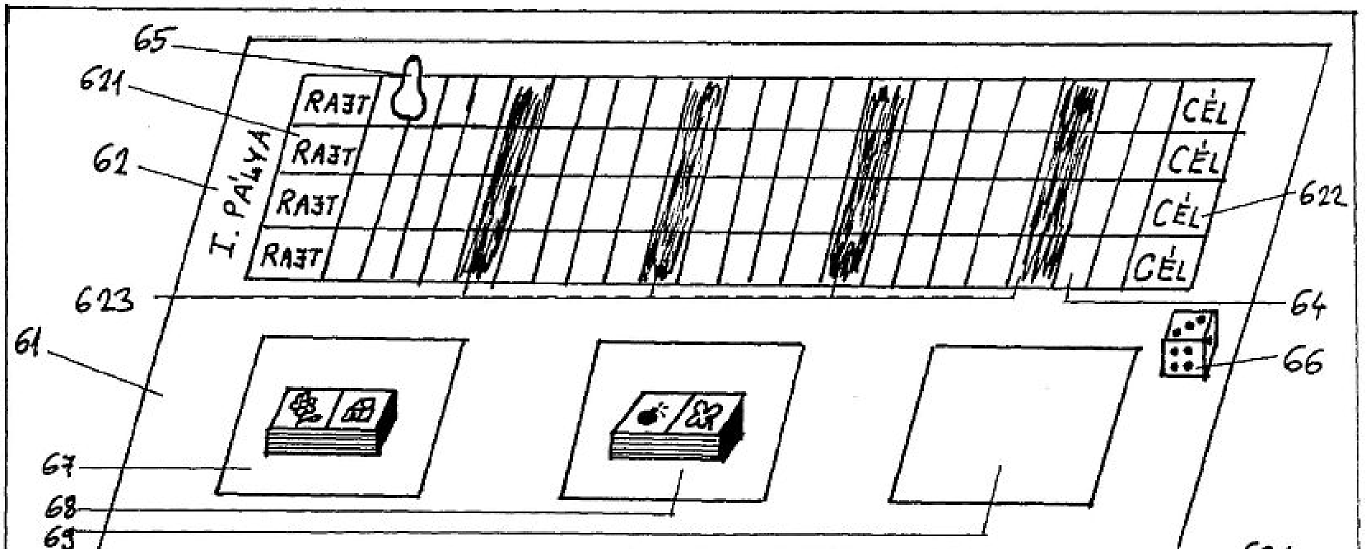
65
621
62
I. PÁLYA
RAJT
RAJT
RAJT
RAJT
CÉL
CÉL
CÉL
CÉL
622
623
64
61
66
67
68
69

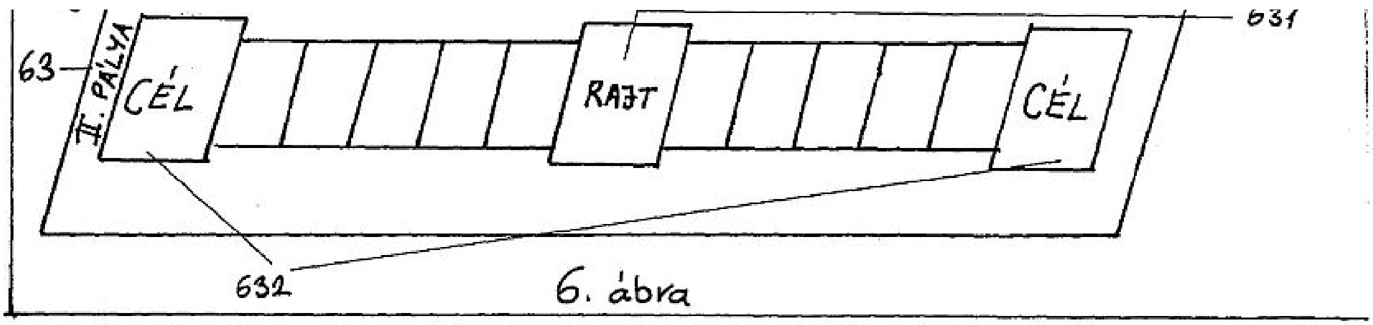

6. ábra

GB222163
Improvements in or relating to apparatus for the development of concentration and telepathy

Inventor: THOMAS GILBERT OAKLEY / AMELIA LESTER

Telepathic messages, indicating. - Apparatus for the development of concentration and telepathy comprises a sphere A, which may be solid or hollow and filled with a liquid and " electrodes " E<1>, to be held in the hands of the user, connected to some part of the apparatus, such as terminals D which are connected to pieces of tin-foil D<2> on opposite sides of the sphere. The sphere A may take the form of an inverted flask carried by a socket B mounted on the base C of the apparatus. A piece F of black manganese oxide may be enclosed within the sphere.

This invention is for improvements in or relating to apparatus for the development of concentration and telepathy and has for its object to provide an inexpensive article of the crystal type which shall give all the effects of the more expensive crystals.

According to this invention the apparatus comprises a hollow sphere of transparent material intended to be filled with a liquid, such as water, a stand for the sphere and " electrodes " intended to be held in the hands of the user whose other ends are secured in proximity to the sphere, preferably on opposite sides of the sphere as viewed in plan.

According to one method of carrying cut this invention, the crystal is made in the form of a spherical glass body having- a neck for filling purposes which is plugged by rubber. The stand for the sphere comprises a base having an upstanding socket into which the neck of the sphere is dropped so that the spherical portion rests upon the top of the socket.

On opposite sides of the stand are two terminals, such as are commonly used for electrical purposes, which are intended to receive wires from a pair of metal electrodes intended to be held in the hands of the user.

Two strips of tin-foil extend up within the socket from the terminals to the base of the sphere where this rests upon the socket. When using the apparatus the glass sphere is filled with water ,and a crystal of black manganese oxide is inserted.

The use of this apparatus, as applied to conveying thought to another person

The user takes the electrodes one in each hand and gazes upon the crystal concentrating his thoughts upon the person with whom he wishes to communicate and visualizes them in his mind's eye. Preferably, the person with whom communication is to be made is similar using, by pre-arrangement, another apparatus of the same kind and the result is that direct communication is established.

Preferably, the apparatus is used against a dark background.

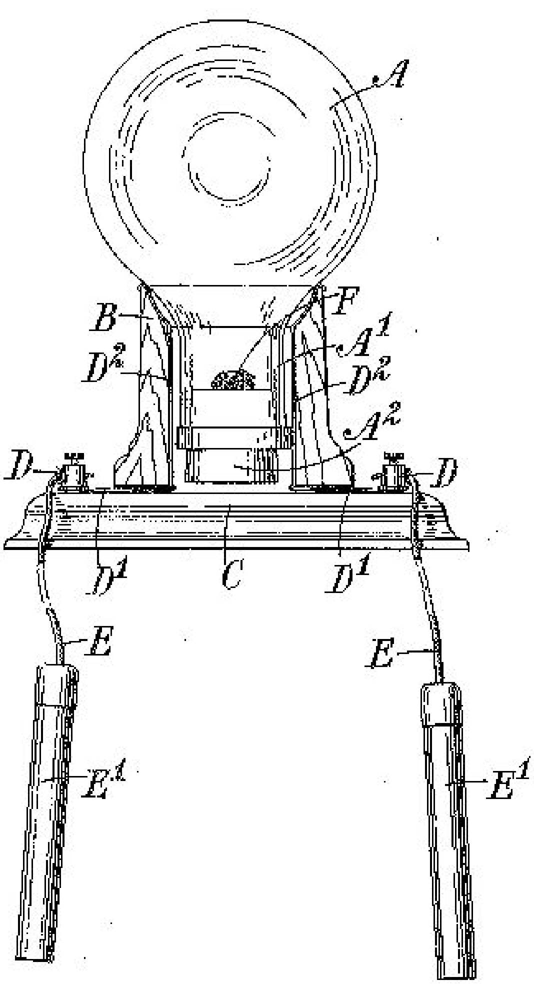

According to this invention apparatus for the development of concentration and telepathy comprises a lens-like sphere and " electrodes " intended to be held in the hands of the user and connected to some part of the apparatus, for example on opposite sides of the said sphere. Preferably, the lens-like sphere takes the form of a hollow transparent body (for example a flask having a spherical body portion) intended to be filled with a liquid, such as water. In the accompanying* drawing which shows an apparatus according to one method of carrying out this invention, the hollow sphere takes the form of a spherical bodied flask A having a filling of rubber, A . The flask is inverted and rests in a. cylindrical support B, which is shown in central vertical section in the drawing, so that only the bulbous portion of the flask extends above the support.The support is carried on a base C provided with two terminals D, such as are commonly used for connecting wires in electrical apparatus. These terminals are connected by wires D to strips of tin-foil D - one for each wire, which extend up within the support B on opposite sides of it and are in contact with the underside of the bulbous portion of the flask A. By wires E the terminals have connected to them handles E (which wires and handles may, for convenience, be termed " electrodes ") similar to those used for medical shocking-coils. When using the apparatus the flask A is filled with water and a crystal F of black manganese oxide is inserted in it. The use of the apparatus, as applied to conveying thought to another person, is as follows:- The user takes the handles E , one in each hand, and gazes upon the crystal, constituted by the bulbous portion A of [he flask.In so doing he concentrates his thoughts upon the person with whom ie wishes to communicate and visualizes them in his .mind's eye. Preferably, the person with whom communication is to be made is similarly using, by prearrangement, another apparatus of the same kind, and the result is that telepathic communication is established.

Preferably, the apparatus is used against a dark background.

The " electrodes " serve as something' to grip which help to concentrate thought.

Obviously, instead of a hollow body A, a solid glass body providing a lens-like spherical portion could be used. The filling liquid and the manganese oxide would then be dispensed with, but it is believed that black manganese oxide has an effect upon the liquid which enables the visualising of objects.

FR2661295
Transmitter/receiver of gravitational waves

Inventor: HENRI PUYUELO JACQUES EMILE

The present invention relates to a transmitter/receiver of gravitational waves capable of allowing man to produce telepathy (thought transmission) or telecommunication at very long range. This device is worn by means of spectacles (32) or of a small helmet in the region of the temporal bone (30) in the vicinity of the inner ear. (Fig. 6). It measures 2 to 3 cm in diameter (35) and has an antenna (34) of about ten centimetres in height. It consists of a box (35) containing a built-in receiver circuit (37) and a built-in transmitter circuit (38) as well as a transmitting gravito-diode (33) and a receiving gravito-diode (44). This device is intended for thought transmission for human beings, for telecommunication at very long range in space vehicles, for psychoanalysis in the medical field (hypnosis, study of the brain, etc.).

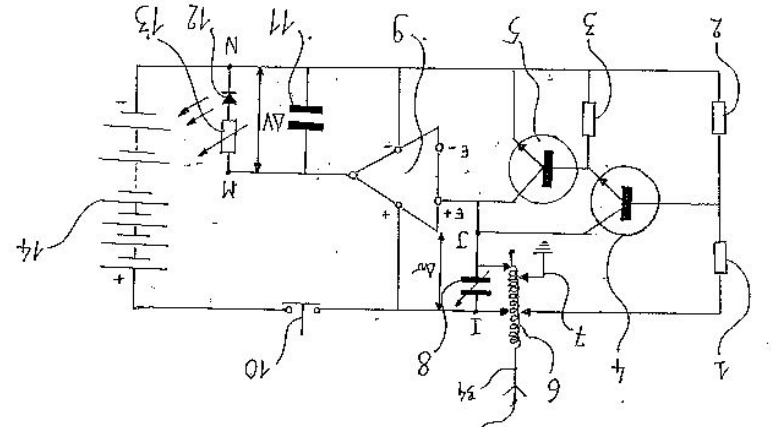
13
12
11
6
5
3
2
N
W
ΔV
14
10
8
9
4
1
I

FIG.1

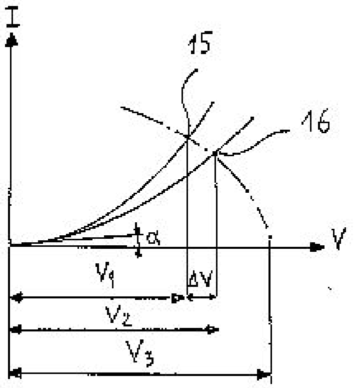

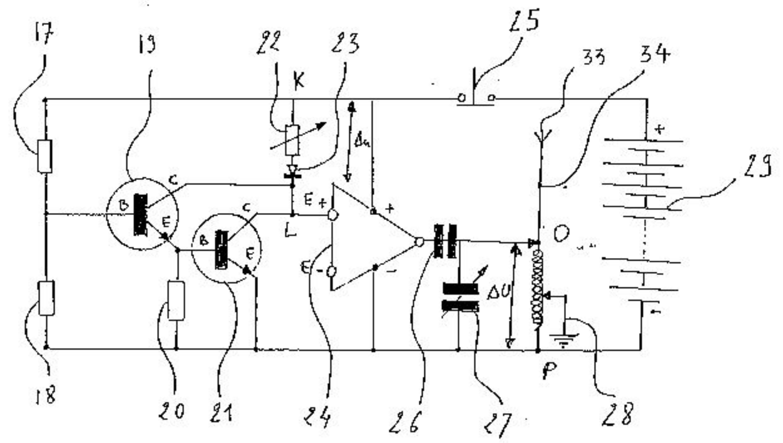
17
19
22
23
25
33
34
29
K
E+
E-
L
O
P
ΔU
B
C
E
18
20
21
24
26
27
28

FIG.3

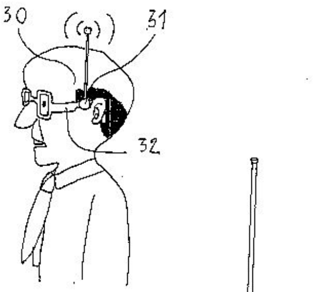

FIG.6

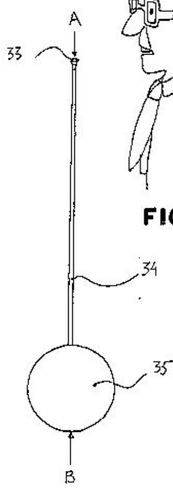

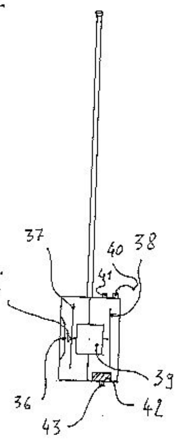
37
40
38
41
44
39
36
42
43

FIG.5

FIG.4

Coupe suivant AB

1 Description

The present invention provides a transceiver of gravitational waves may allow man have to do to telepathy (mind reading) or the telecommunication great distance.
This device is worn i through glasses (32) or small helmet at the temporal bone (30) adjacent to the inner ear.

(Fig.6) I1 measure 2 to 3 cm in diameter (35) and has an antenna (34) dune ten centimeters high.

This antenna is actually an elongated gravito-diode (33) which detects and transmits the gravitational waves. (Figure 4)

See more details for the patent on the diode-gravity.

This device seen in section, (Fig.5) comprises: - a receiver (36) - an integrated receiver circuit (37) - a transmitter integrated circuit (38) - an orifice (39) to permit connection of the device at of the temporal bone.
- An LED (40) that signals the operating device I - a selector (41) emission -rdception - a 9 Volt battery (42) - a switch (43) - stop walking.
- An antenna (34) is an elongated-gravity diode.

The receiver circuit is shown diagrammatically Fig. 1.

A receiving antenna for gravitational waves (34) low voltage signal v selected through dune coil (#) and a variable capacity (8) in the circuit portion IJ
A Darlington circuit using two transistors Ti (4) and T2 (5) and 3 resistant (1), (2), (3) amplifies the current detected by the low-gravity-diode antenna.

The signal V is amplified by a low frequency operational amplifier and V is N.

See the characteristic of the diode-gravity Fig.2

The issuer is tircuit fig.3 schematically.

Through the earpiece (36), a -diode gravito receives a weak signal u varianle through a resistor (22). With a Darlington circuit anda high frequency operational amplifier.

This voltage is amplified O and P is U.

The varible capacity (27) allows you to tune the antenna (34) and the IJ signal is transmitted on a gravitational frequency.

The device runs on a 9V battery.

Seriously bizarre :

JP2014155432
SUPERNATURAL POWER

Inventor: MATSUO FUKIKO

PROBLEM TO BE SOLVED: To develop and freely realize the first supernatural power conversation telepathy in the world and to develop and realize other supernatural power, while humans have conversed only with humans that they meet and the supernatural power has been unrealistic for 4.6 billion light-years.
SOLUTION: In October 2003, world's first supernatural conversation was developed. Supernatural power first sends the mind and the mind to heads. Telepathy as means with which humans converse with invisible parties was found and spread throughout the world. A house in Akashi city. Swallowtail butterfly Onnamon (emblem for females) is imperial and is worshipped as a god, and the supernatural power has become widespread through Japan and main countries in the work by TV and Niigata Earthquake in the hometown of the inventor. A sultry night was developed after four years. Love becomes popular, capacity of the supernatural power rapidly increases, show business is especially gay and there is joy all over the world. Children learn a universal language without the need of translation by 9 years old and have conversation when they are 10 years old. The conversation reaches space but aliens and animals do not reply. There are psychokinesis, psychic photography, out-of-body experience, telekinesis capable of healing even tumors. Conversation is held with ghosts from 20 years ago. Metempsychosis is possible. World diplomatic relation restoration (for North Korea or the like) or the like is pursued, show business is saved from epidemic, prescriptions are spread through the world helping 500 to 1,000 people, shrines and temples are constructed at Yasaka-jinja shrine, and patent application is instructed by the Prime Minister Abe. There is a desire to dedicate to a birthplace for the first time in the world. The emperor speaks that the swallowtail butterfly

is the emblem of the first imperial married couple, the butterfly of the time is famous and can freely talk with world's major corporations and specialists.

The Cure for Chi-Communism :

CN101513296
Natural telepathy crystal earbob

Inventor: SHANGLUN CHEN

The invention relates to a natural telepathy crystal earbob, and belongs to the technical field of jewelry making. A primary product is made into an earbob decoration with intrinsic value by utilizing the unique characteristics of the primary product through elaborative selection, repeated comparison and processes of abrading, polishing, perforating and the like for the primary product with hexagon according to a basic shape of the primary product. The natural telepathy crystal earbob is believed to be favored by successful women having the consumption conditions and pursuing fashion glamour, and increases sparkle to the diversification of the society.

Jing crystal jewelry stones to build the technical field

[0002]
Lucky this stone after stone came from the country of origin, it was placed, and other ordinary spar together, without any special, I stumbled on this little rock scene and there was swimming in the living body can After, I feel it is a rare and precious, therefore it made the style, shape have a unique and novel features of earrings and I believe it also will let more people love and luck.

[0003]
The production process from the country of origin business purchases in selection, came home one by one carefully further repeated viewing, found life objects can swim in the side to select it, cook the label, and then by machine cutting, grinding, fight After the holes and other procedures, coupled with the processing of gold and silver connecting ring, connecting buckle, from the overall look at each link found no flaws even completed.

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

**... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !**

ORDER PAGE

US 20150091891A1

(19) **United States**
(12) **Patent Application Publication**
Raheman et al.
(10) **Pub. No.: US 2015/0091891 A1**
(43) **Pub. Date: Apr. 2, 2015**

(54) **SYSTEM AND METHOD FOR NON-HOLOGRAPHIC TELEPORTATION**

(71) Applicants: **Fazal Raheman**, Dubai (AE); **Ali Fazal Raheman**, Dubai (AE)

(72) Inventors: **Fazal Raheman**, Dubai (AE); **Ali Fazal Raheman**, Dubai (AE)

(73) Assignee: **DUMEDIA, INC.**, Newark, DE (US)

(21) Appl. No.: **14/040,729**

(22) Filed: **Sep. 30, 2013**

Publication Classification

(51) **Int. Cl.**
G06T 15/00 (2006.01)

(52) **U.S. Cl.**
CPC ***G06T 15/00*** (2013.01)
USPC **345/419**

(57) **ABSTRACT**

The present invention discloses a system and method for non-holographic virtual teleportation of one or more remote objects to a designated three-dimensional space around a user in realtime. This is achieved by using a plurality of participating teleportation terminals, located at geographically diverse locations, that capture the RGB and depth data of their respective environments, extract RGB images of target objects from their corresponding environment, and transmit the alpha channeled object images via Internet for integration into a single composite scene in which layers of computer graphics are added in the foreground and in the background. The invention thus creates a virtual 3D space around a user in which almost anything imaginable can be virtually teleported for user interaction. This invention has application in intuitive computing, perceptual computing, entertainment, gaming, virtual meetings, conferencing, gesture based interfaces, advertising, ecommerce, social networking, virtual training, education, so on and so forth.

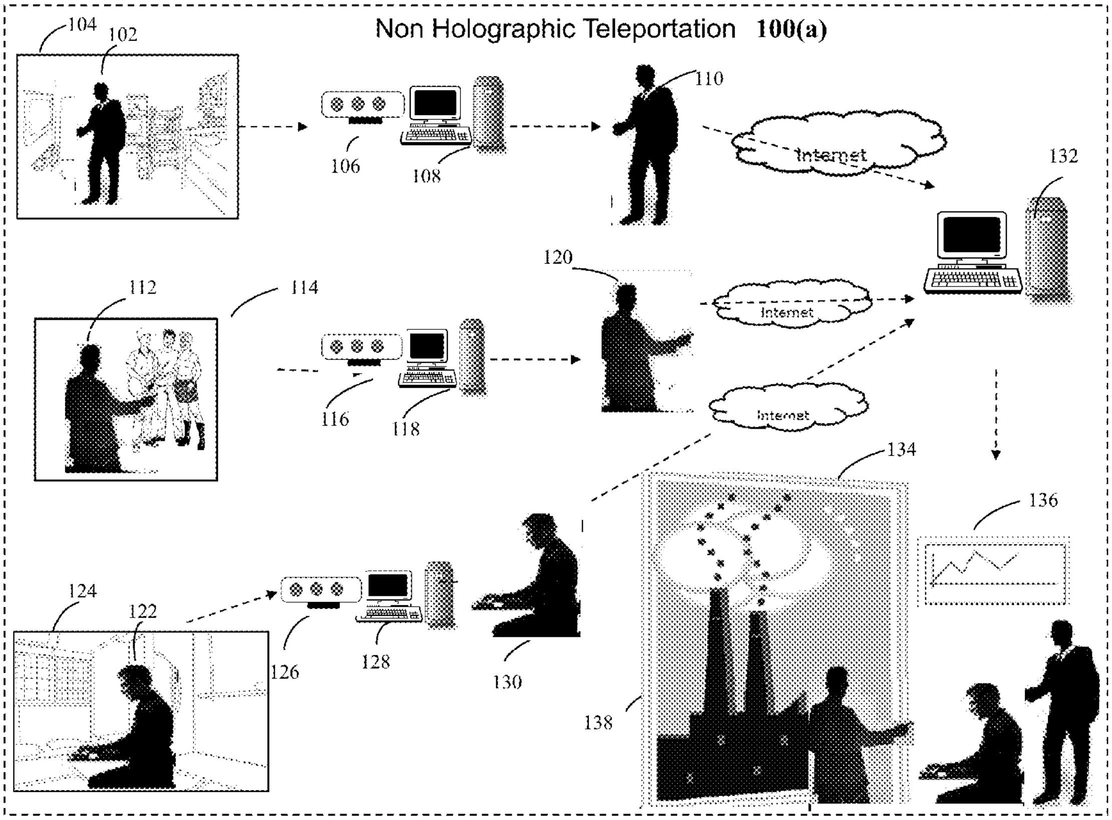

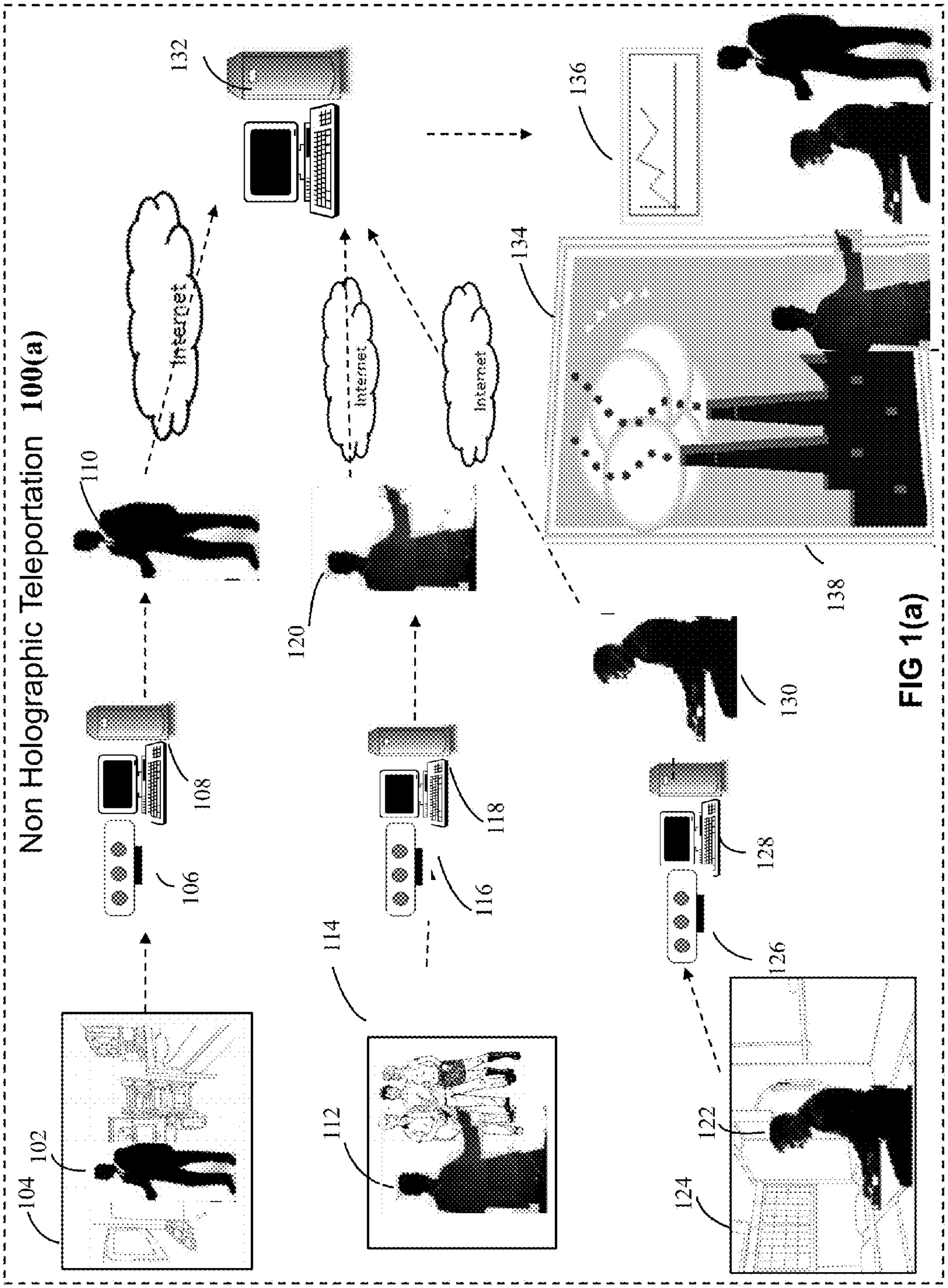
Non Holographic Teleportation 100(a)
Internet
Internet
Internet

FIG 1(a)

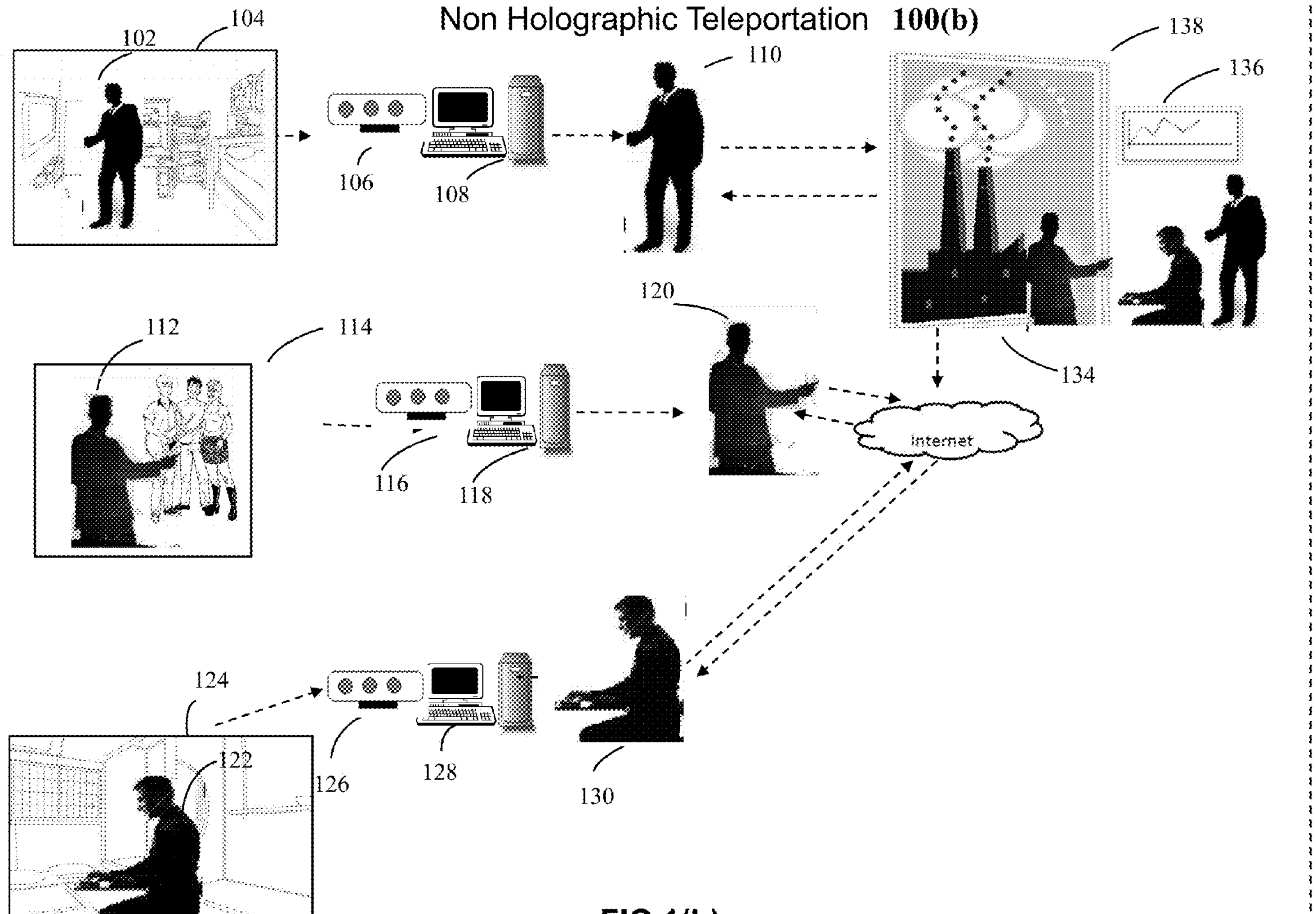
Non Holographic Teleportation **100(b)**
102
104
106
108
110
138
136
112
114
116
118
120
134
Internet
124
122
126
128
130

FIG 1(b)

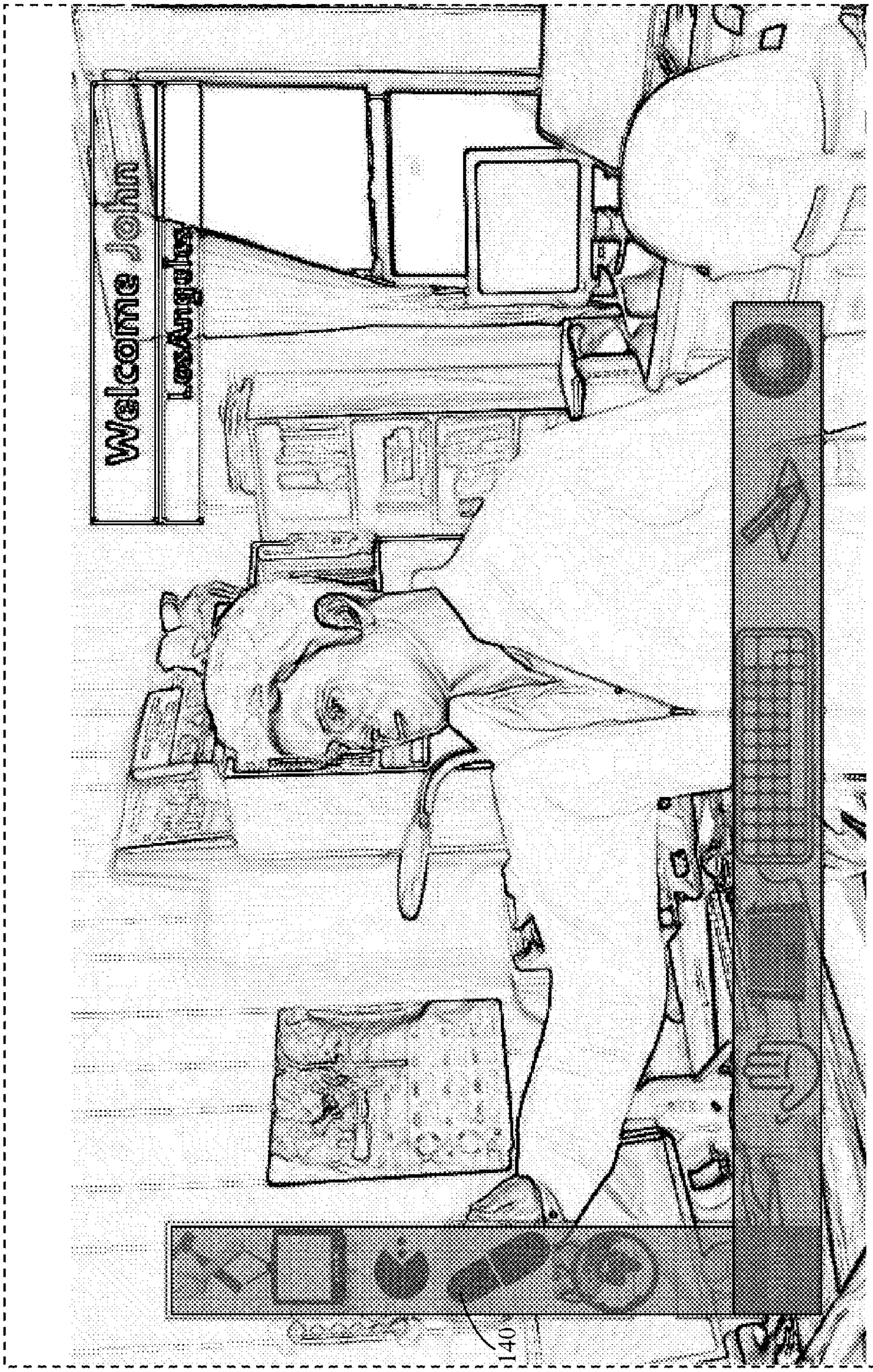
Welcome John
Los Angeles
140

FIG 1(c)

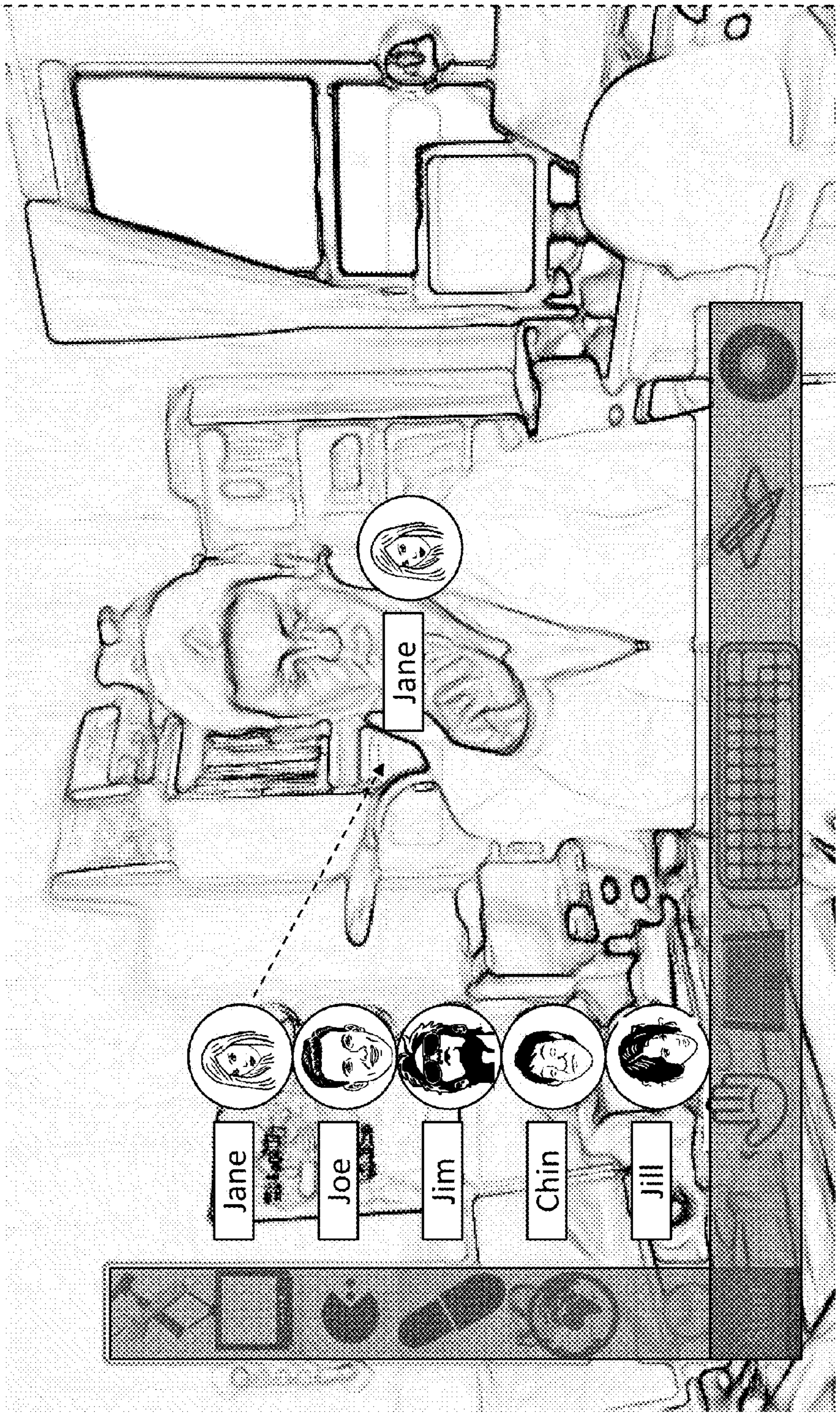
Jane
Jane
Joe
Jim
Chin
Jill

FIG 1(d)

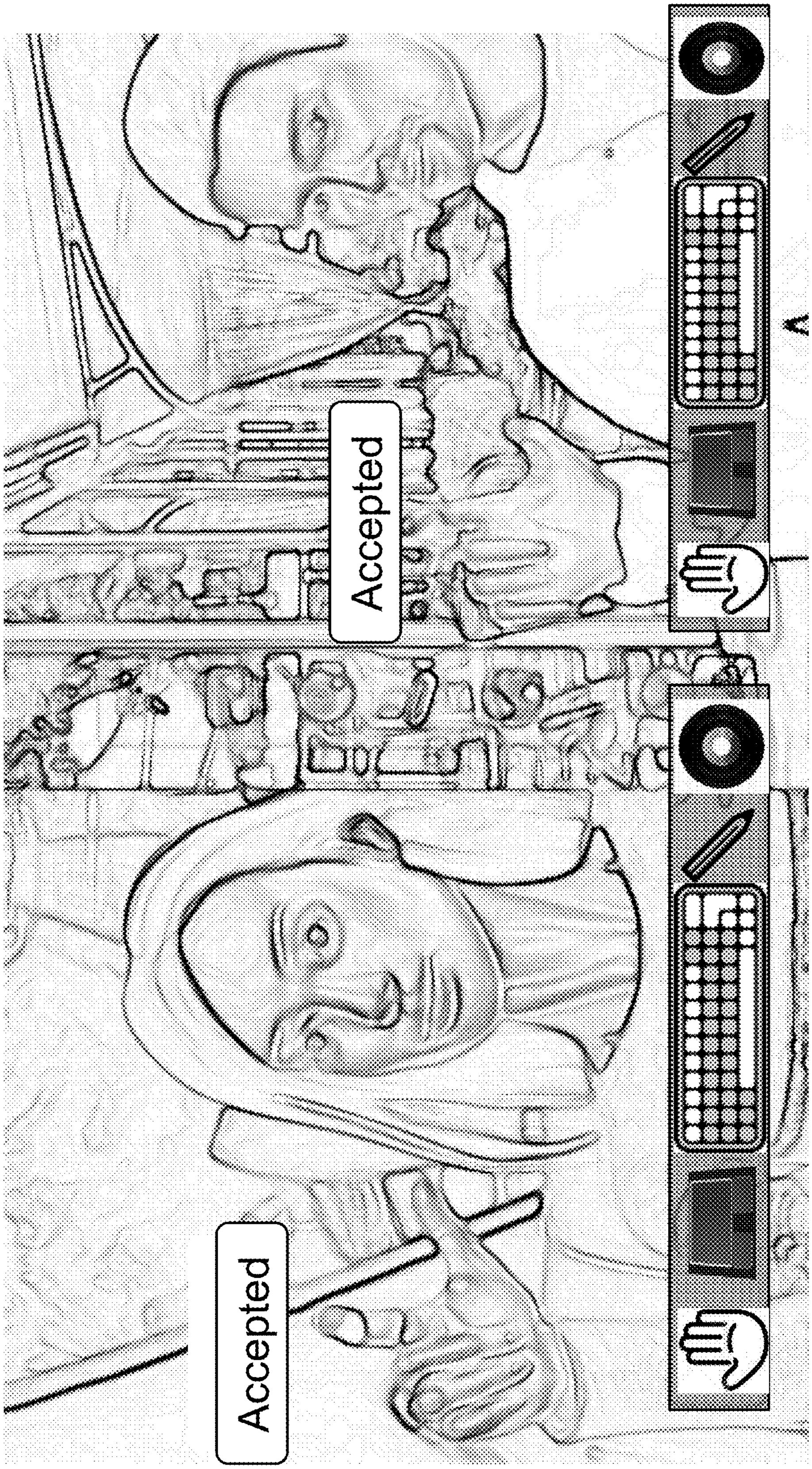
Accepted
Accepted
A

FIG 1(e)

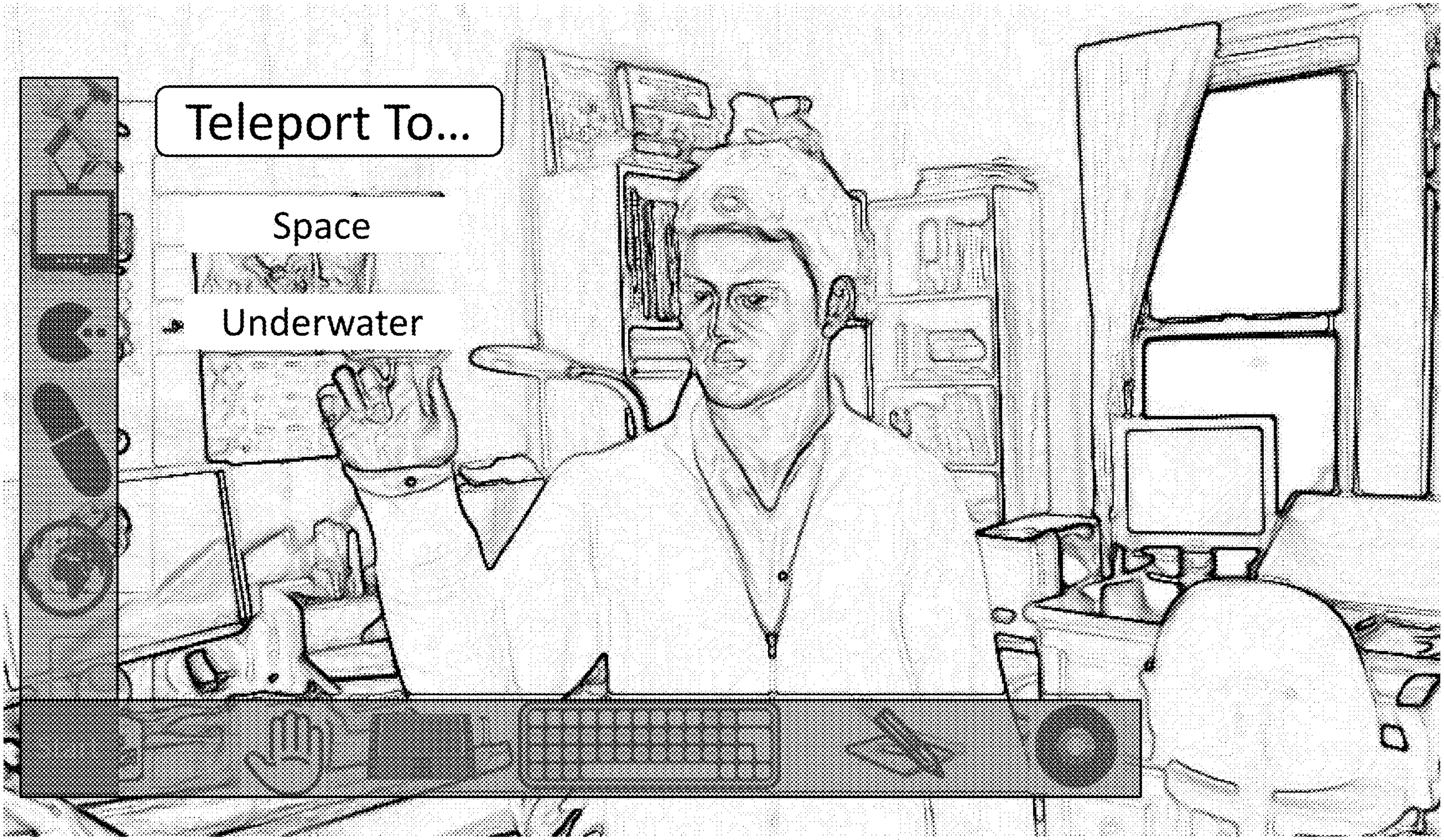
Teleport To...
Space
Underwater

FIG 1(f)

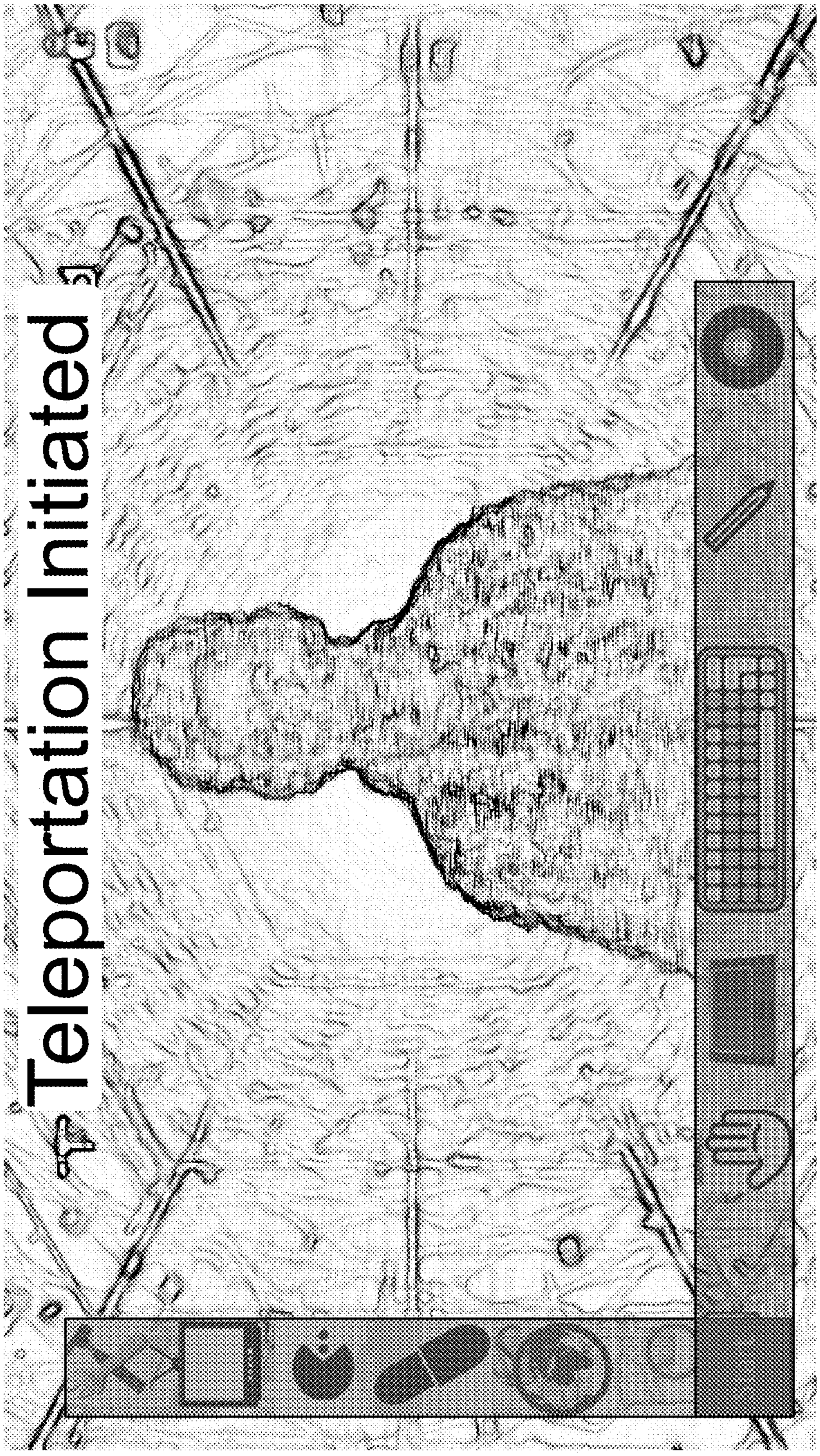

Teleportation Initiated

FIG 1(g)

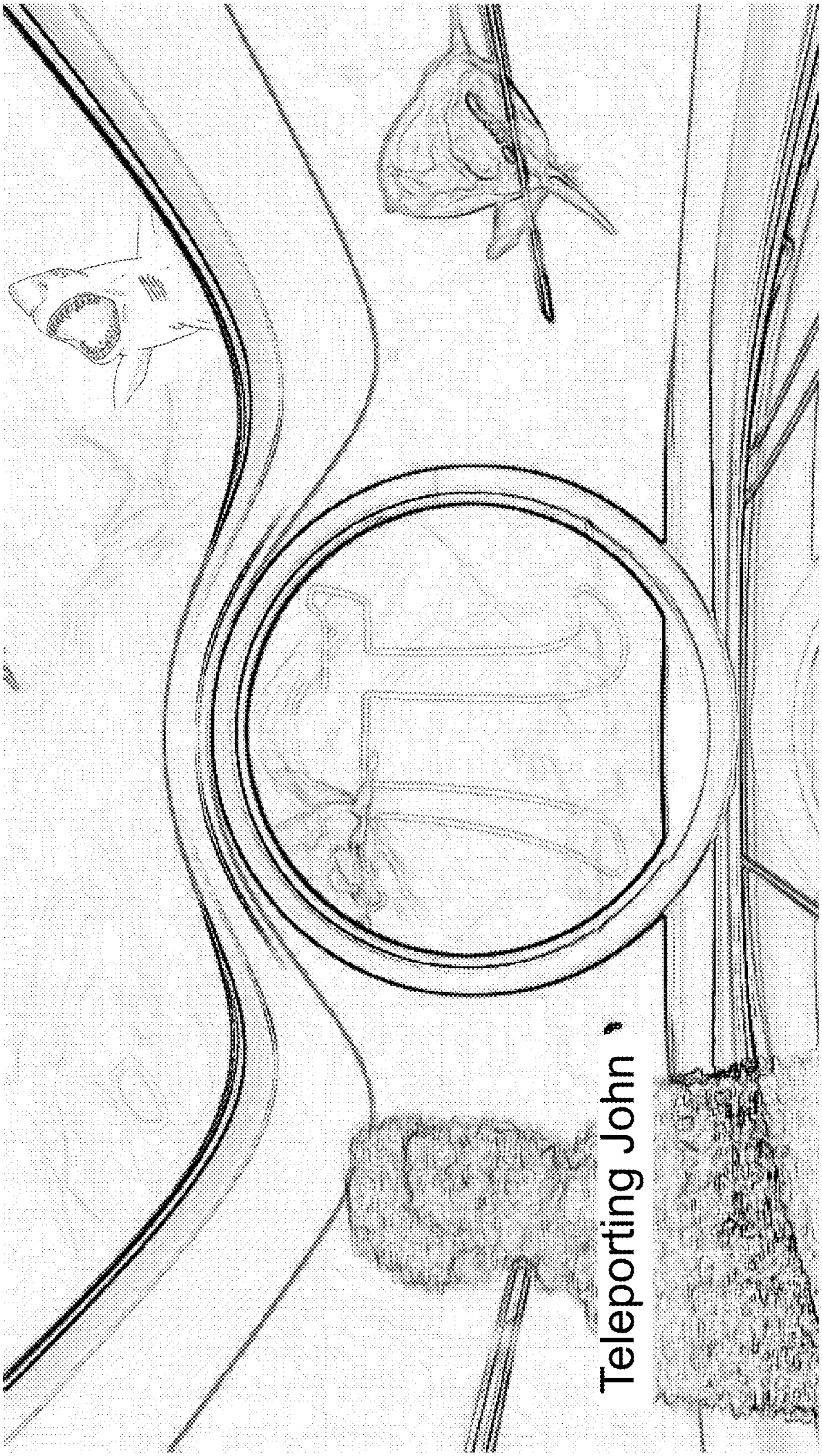
Teleporting John

FIG 1(h)

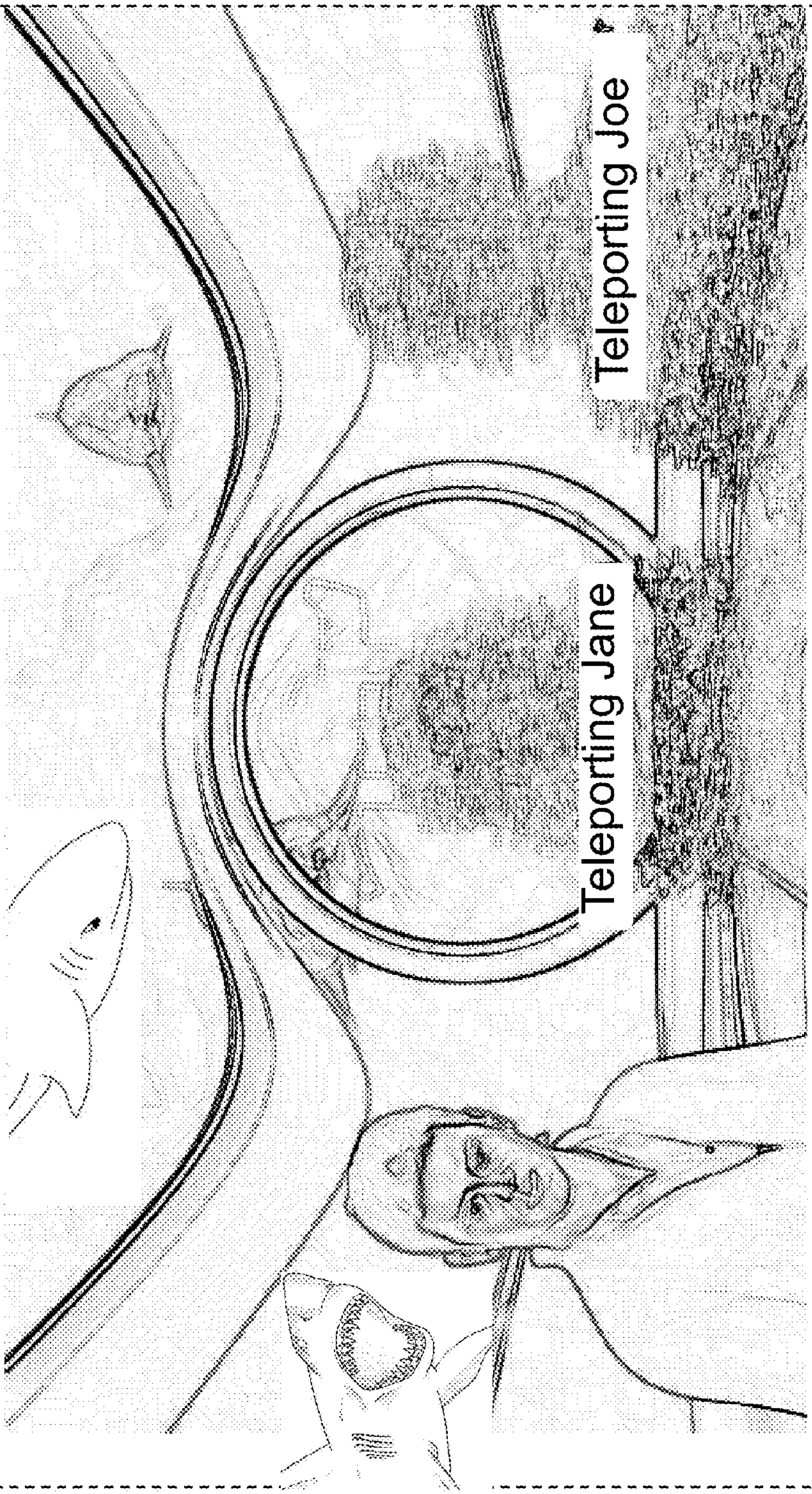
Teleporting Joe
Teleporting Jane

FIG 1(i)

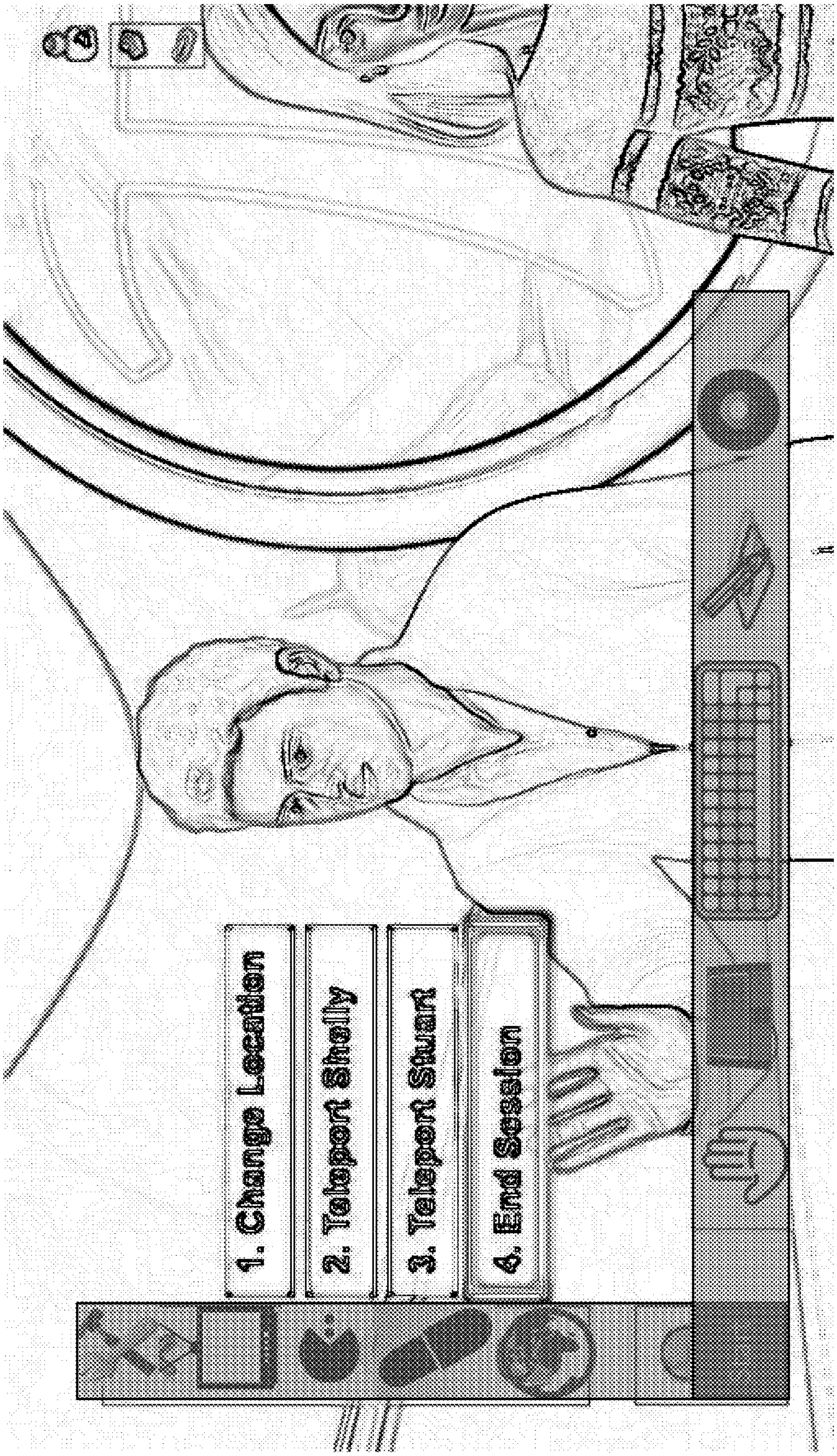
1. Change Location
2. Teleport Shelly
3. Teleport Stuart
4. End Session

FIG 1(j)

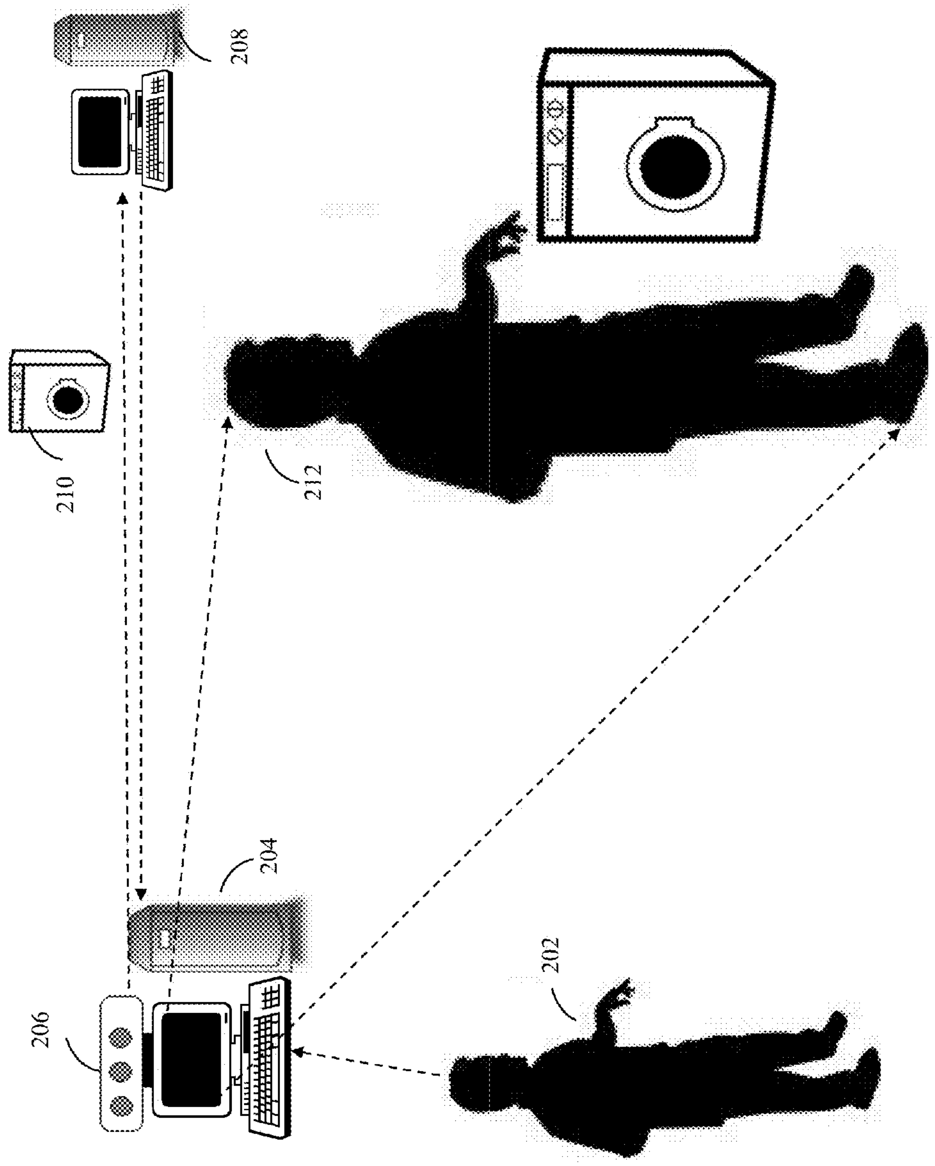
202
204
206
208
210
212

FIG 2

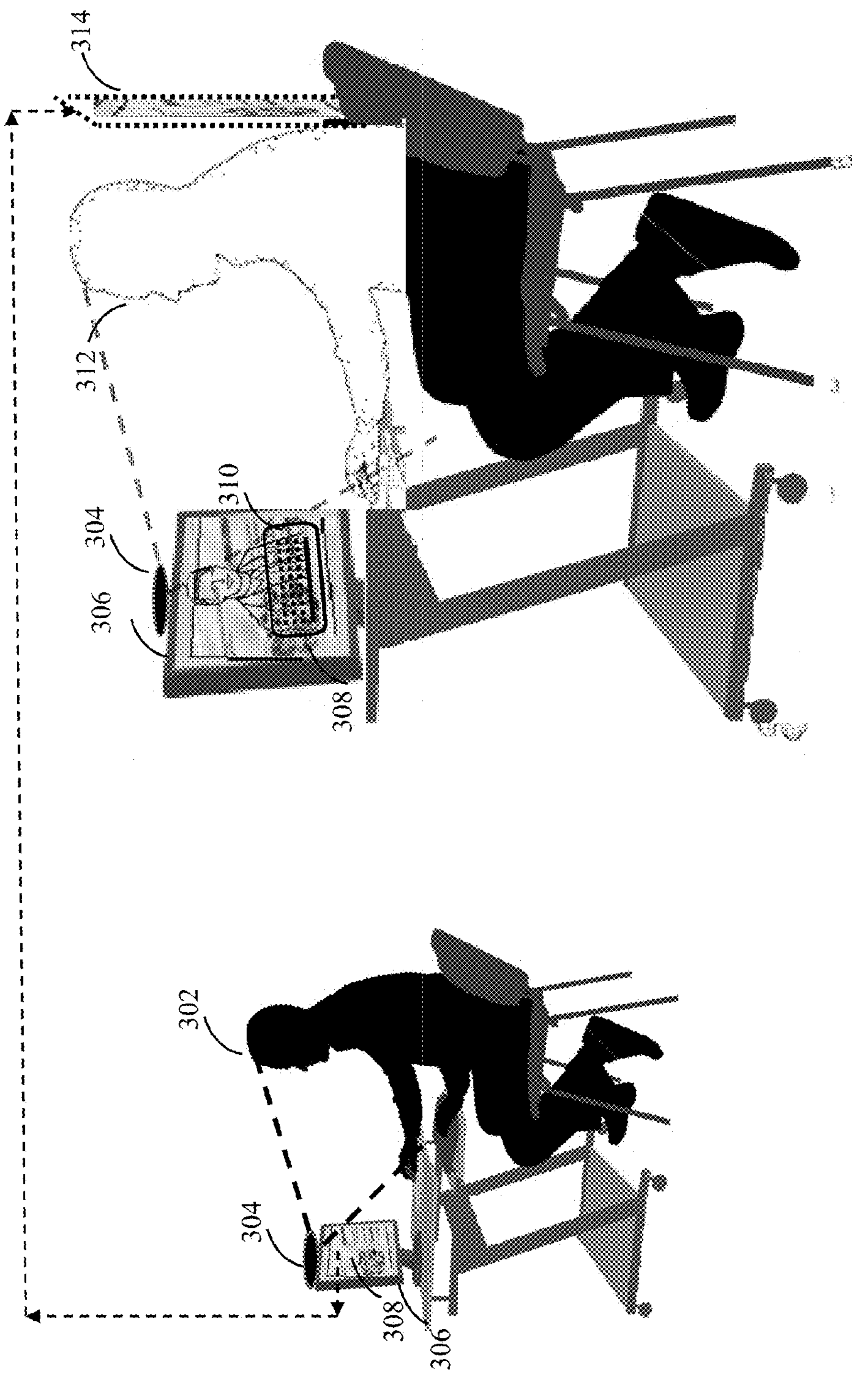
314
312
304
306
310
308
302
304
308
306

FIG 3

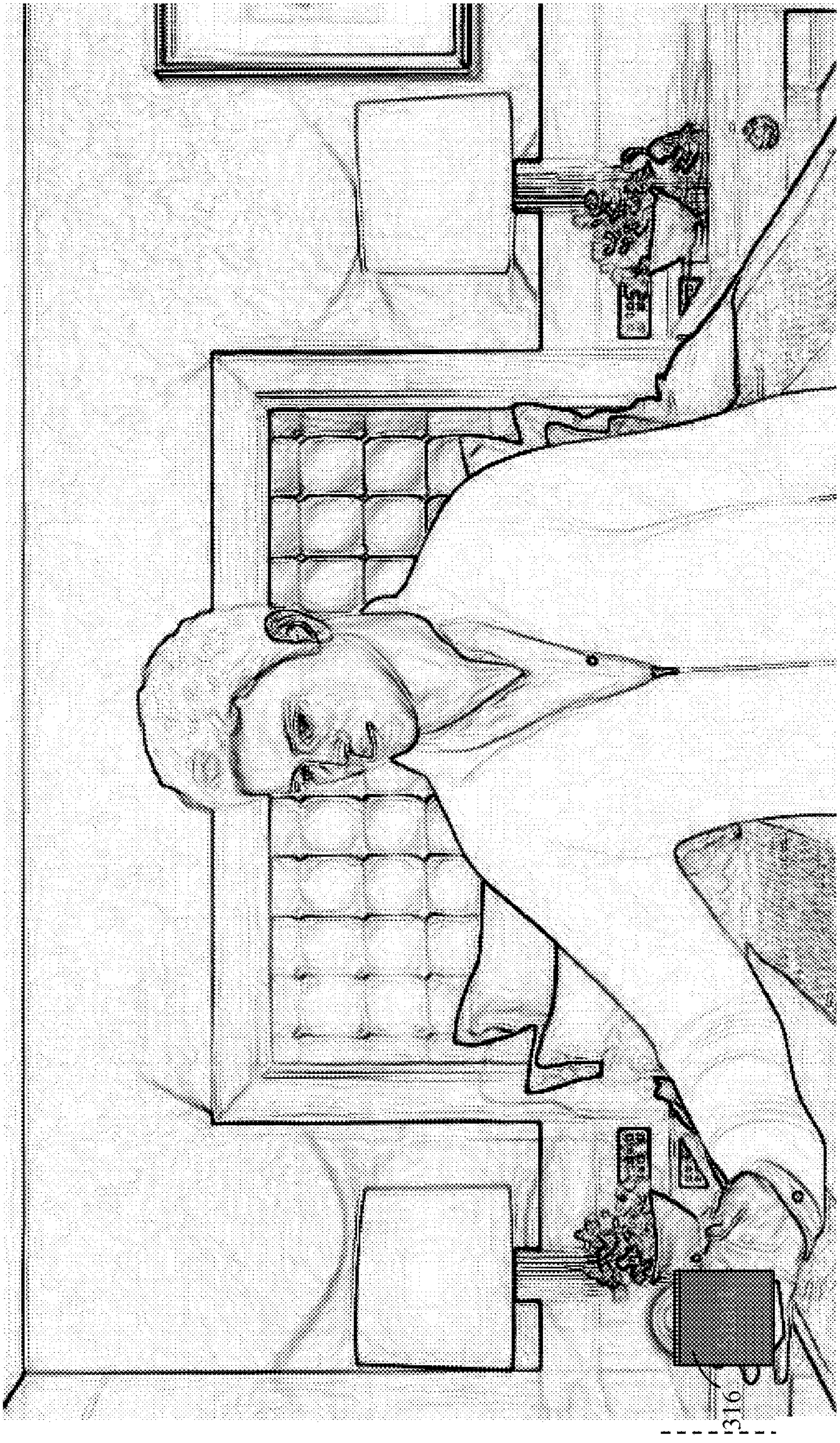

FIG 3(a)

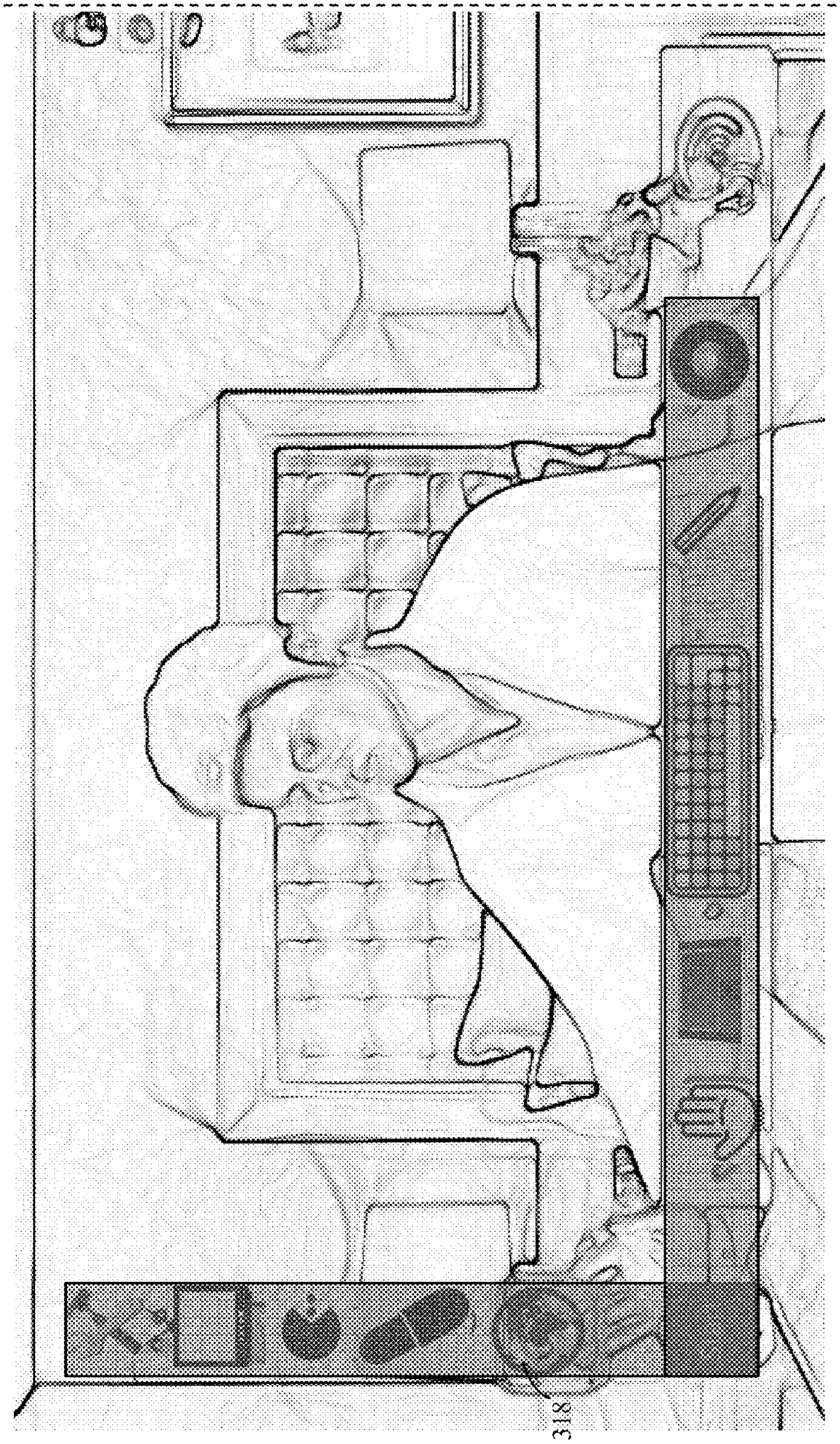

FIG 3(b)

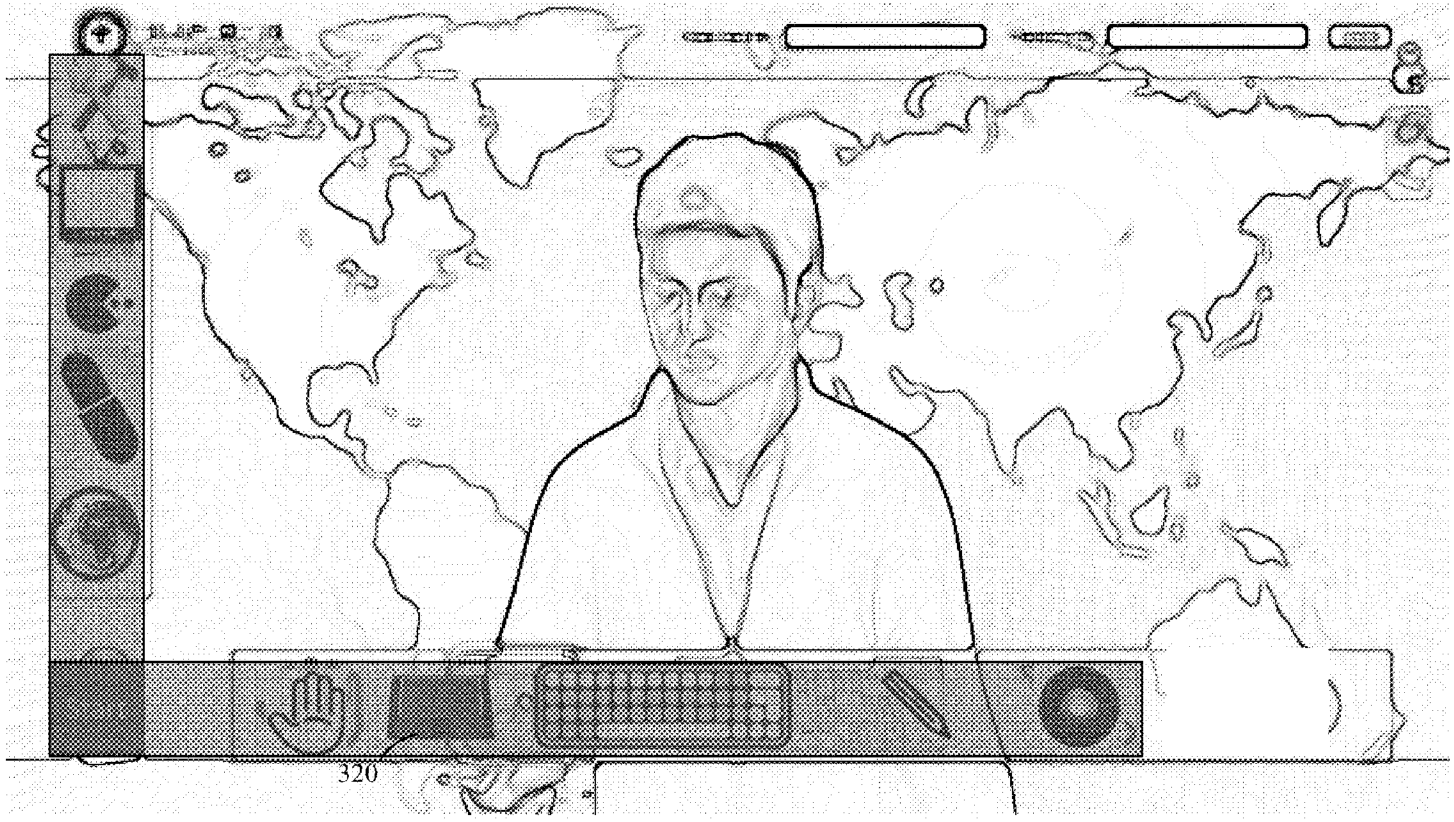

FIG 3(c)

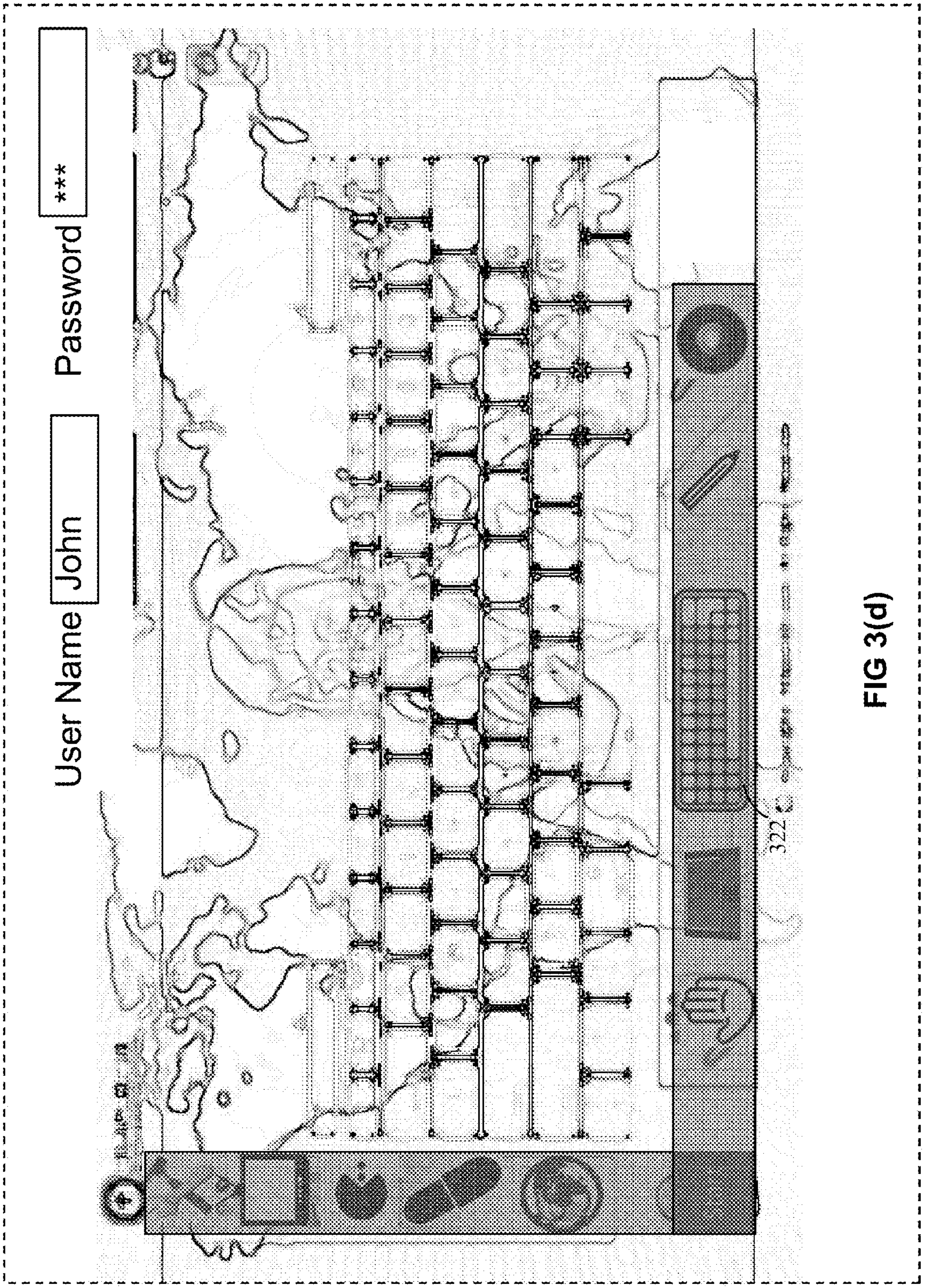
User Name John

Password ***

FIG 3(d)

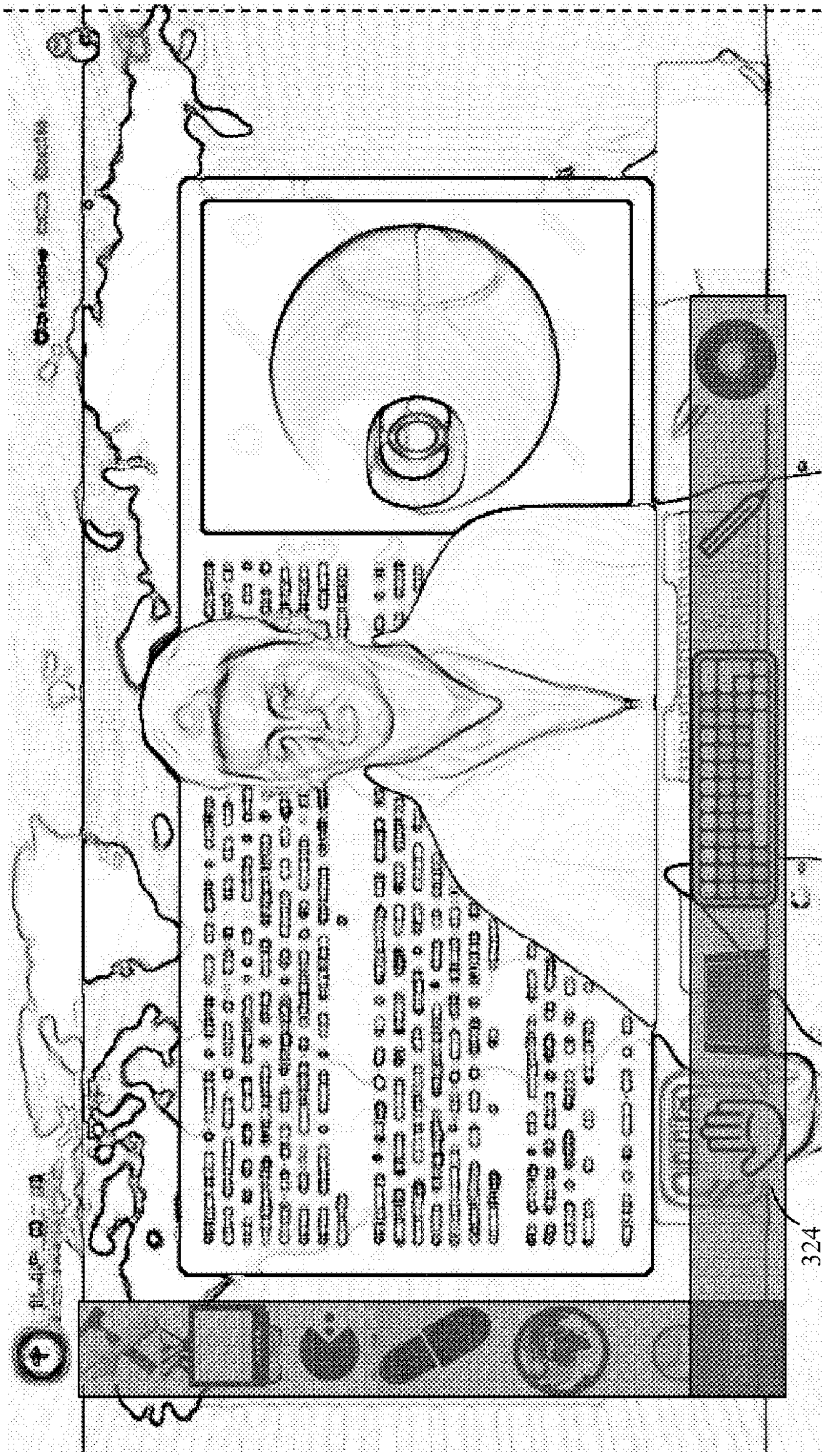

FIG 3(e)

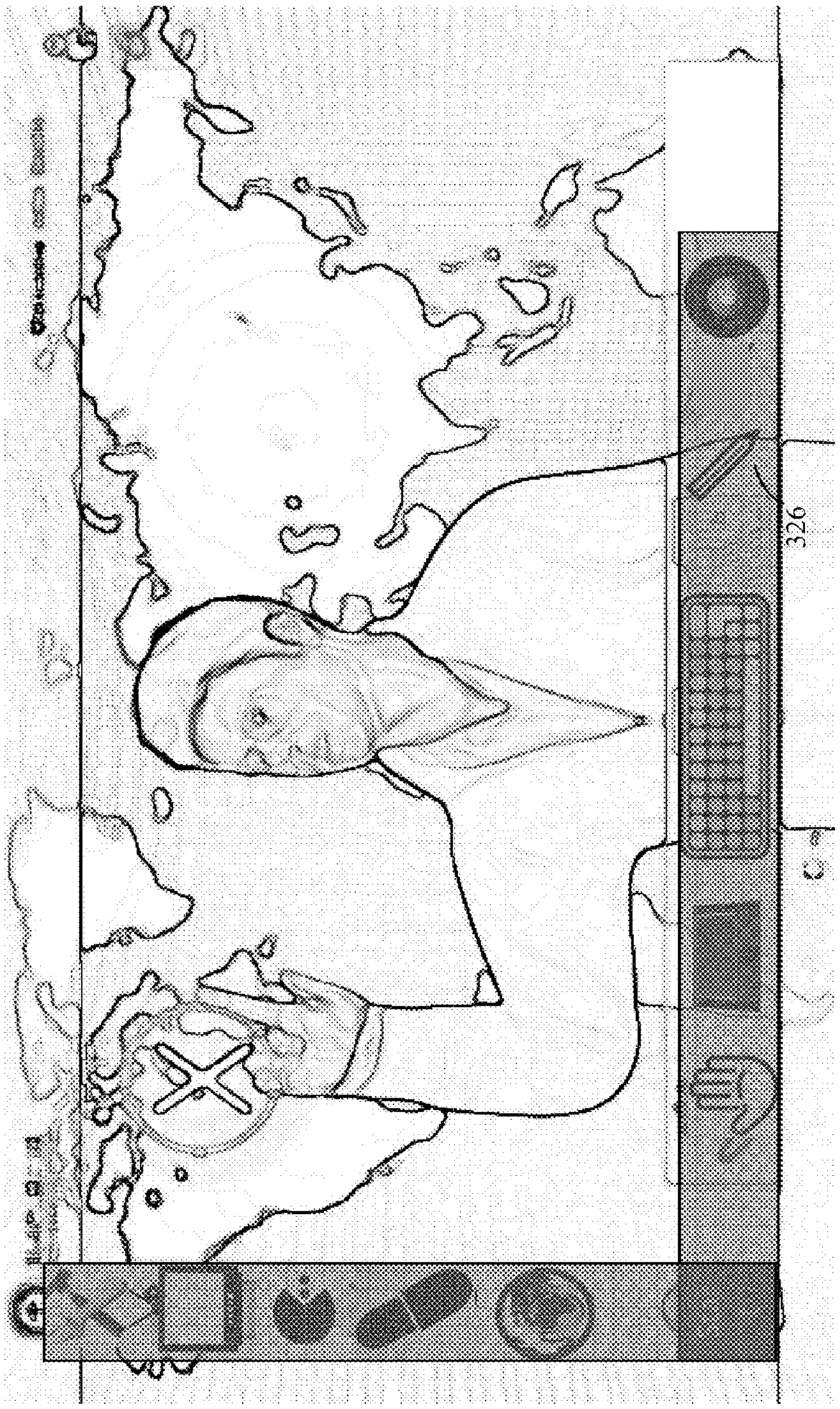

FIG 3(f)

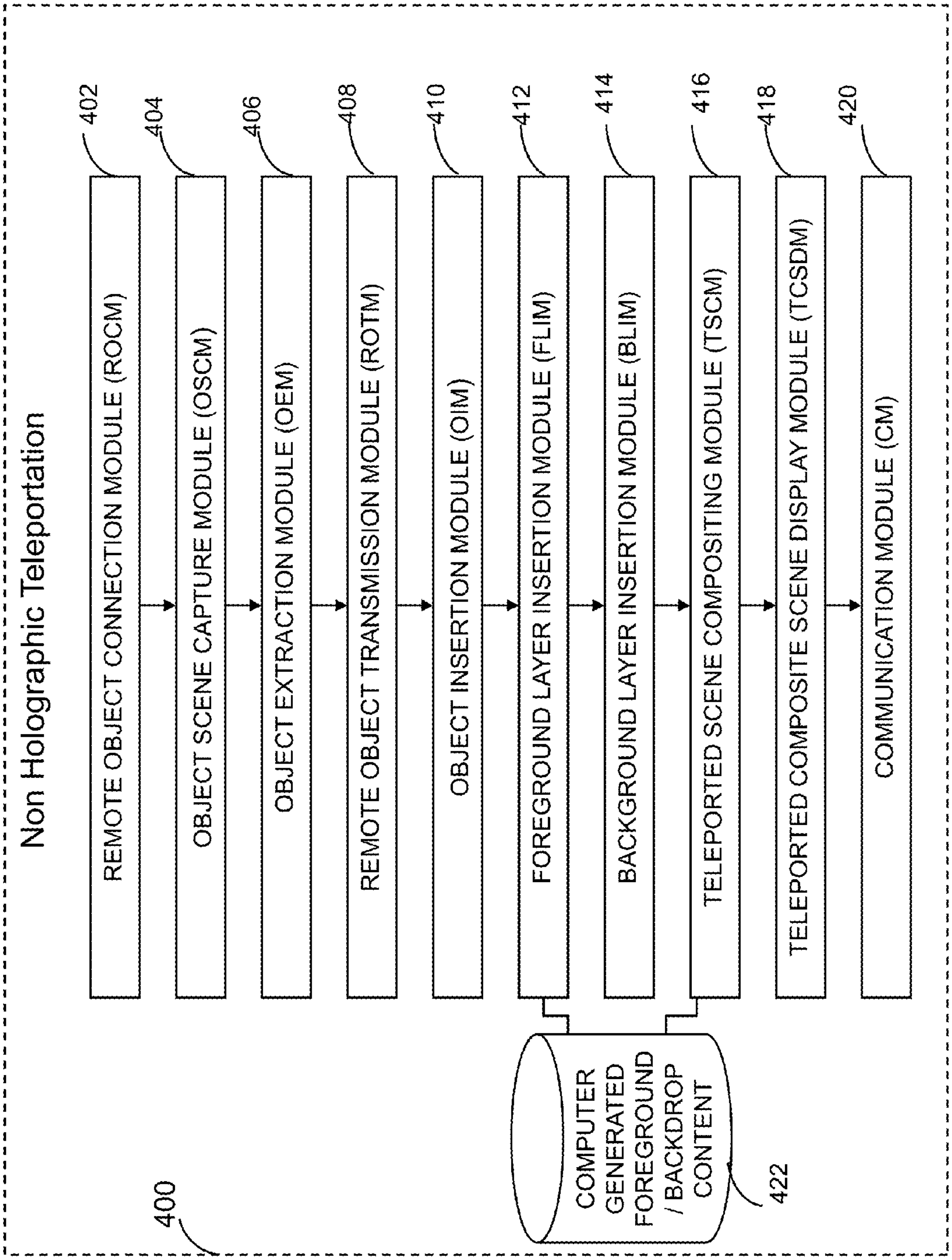
Non Holographic Teleportation
402
REMOTE OBJECT CONNECTION MODULE (ROCM)
404
OBJECT SCENE CAPTURE MODULE (OSCM)
406
OBJECT EXTRACTION MODULE (OEM)
408
REMOTE OBJECT TRANSMISSION MODULE (ROTM)
410
OBJECT INSERTION MODULE (OIM)
412
FOREGROUND LAYER INSERTION MODULE (FLIM)
414
BACKGROUND LAYER INSERTION MODULE (BLIM)
416
TELEPORTED SCENE COMPOSITING MODULE (TSCM)
418
TELEPORTED COMPOSITE SCENE DISPLAY MODULE (TCSDM)
420
COMMUNICATION MODULE (CM)
COMPUTER GENERATED FOREGROUND / BACKDROP CONTENT
422
400

FIG. 4

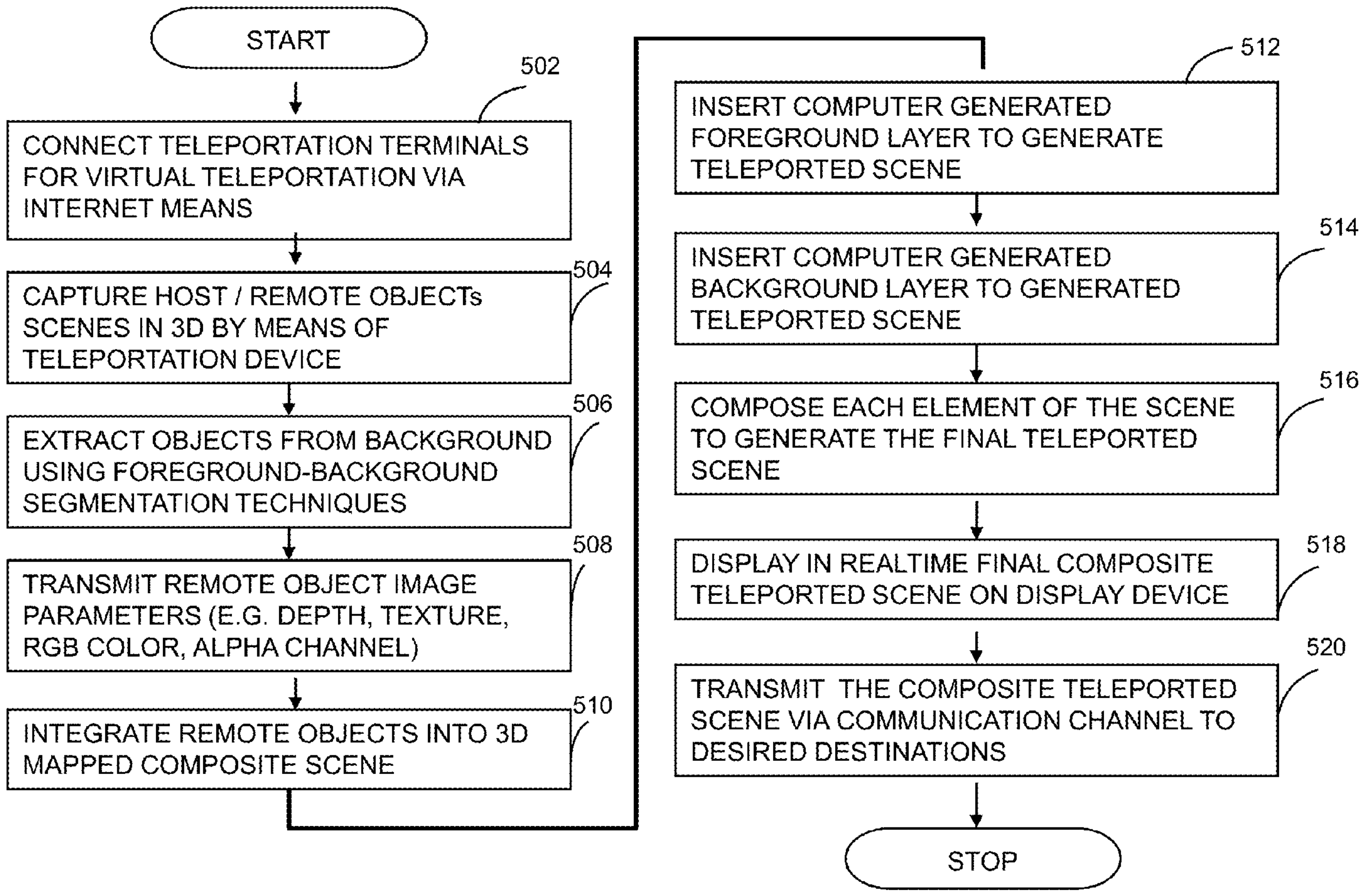
START
502
CONNECT TELEPORTATION TERMINALS FOR VIRTUAL TELEPORTATION VIA INTERNET MEANS
504
CAPTURE HOST / REMOTE OBJECTs SCENES IN 3D BY MEANS OF TELEPORTATION DEVICE
506
EXTRACT OBJECTS FROM BACKGROUND USING FOREGROUND-BACKGROUND SEGMENTATION TECHNIQUES
508
TRANSMIT REMOTE OBJECT IMAGE PARAMETERS (E.G. DEPTH, TEXTURE, RGB COLOR, ALPHA CHANNEL)
510
INTEGRATE REMOTE OBJECTS INTO 3D MAPPED COMPOSITE SCENE
512
INSERT COMPUTER GENERATED FOREGROUND LAYER TO GENERATE TELEPORTED SCENE
514
INSERT COMPUTER GENERATED BACKGROUND LAYER TO GENERATED TELEPORTED SCENE
516
COMPOSE EACH ELEMENT OF THE SCENE TO GENERATE THE FINAL TELEPORTED SCENE
518
DISPLAY IN REALTIME FINAL COMPOSITE TELEPORTED SCENE ON DISPLAY DEVICE
520
TRANSMIT THE COMPOSITE TELEPORTED SCENE VIA COMMUNICATION CHANNEL TO DESIRED DESTINATIONS
STOP

FIG. 5

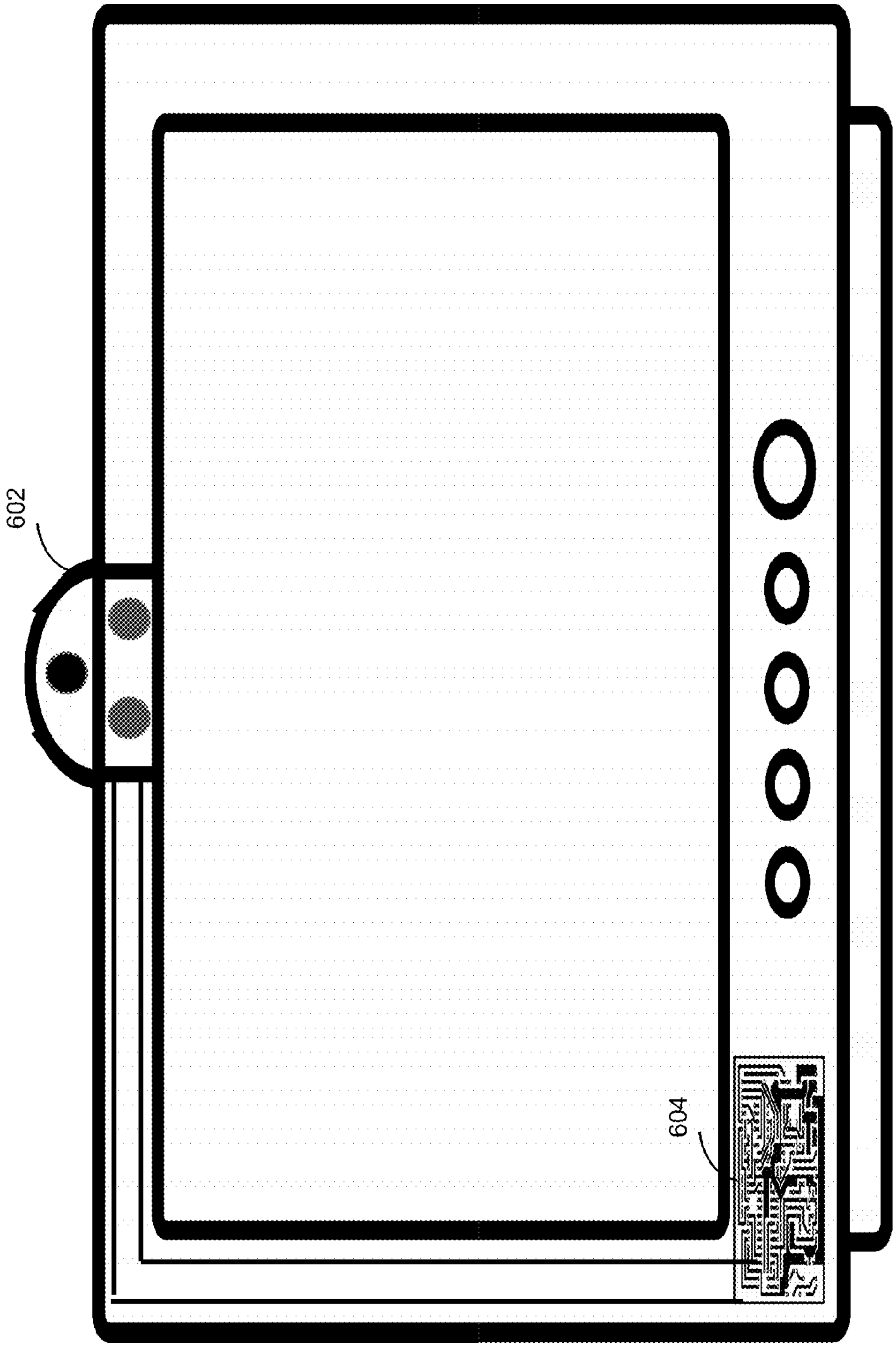

FIG. 6

SYSTEM AND METHOD FOR NON-HOLOGRAPHIC TELEPORTATION

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0001] Not Applicable

REFERENCE TO A MICROFICHE APPENDIX

[0002] Not Applicable

TECHNICAL FIELD OF INVENTION

[0003] The present invention generally relates to the field of virtually teleporting in real time, from diverse geographical locations, one or more remote objects, subjects, themes, events, environments, computer graphics, interfaces, seamlessly into user's three-dimensional (3D) space to create an interactive immersive experience. Specifically the invention pertains to creation of a 3D virtual space around a computer user in which almost anything imaginable can be virtually teleported for user interaction. More particularly the invention pertains to a non-holographic real time technique that extracts from their corresponding backdrops, images and depth information of one or more remote objects located at remote teleportation terminals, transmits said images with alpha channel to a host teleportation terminal, extracts host object and composes a composite scene of all participating objects by inserting pre-selected background and foreground computer graphics. More importantly the invention creates a new versatile genre of non-holographic virtual teleportation that has a broad scope applicability in applications such as intuitive computing, perceptual computing, entertainment, gaming, virtual meetings, conferencing, gesture based interfaces, advertising, ecommerce, social networking, virtual training, education, so on and so forth.

BACKGROUND

[0004] In a pending application (U.S. application Ser. No. 13/864,019) the present inventors disclosed a multilayered augmented reality technique that deployed a virtual chroma keying approach that eliminated the requirement of bulky and cumbersome physical chroma screens by making use of digitally generated virtual chroma mask. That approach reduced the size of hard chroma screen background and increased the portability of an Augmented Reality Kiosk. But it did not completely eliminate the need of a reduced size chroma screen. Furthermore, although it added multiple layers of foreground and background computer graphics to the user's 3D space, it did not extract, transmit and insert into the scene objects from remote locations. Adding 2D or 3D foreground and background segmentation techniques for background removal not only completely eliminates the need for hard chroma screen, but creates a virtual 3D space around a user in which almost anything can be teleported from any remote location for user interaction. 3D depth sensing is known in prior art to enable separation of a subject from background within a camera view without any chroma screen, and at the same time capture user movements for interacting with the system. It also opens an entire new way to teleport remote objects between remote geographical locations in real time.

[0005] Traditionally, holography has been used for virtual teleportation of human subjects. Ideally, teleportation is defined as transfer of matter from one point to another without traversing the physical space between them. As much as it may be a very cool concept in the popular ScFi culture, for the real world inventors, it is not a matter of actually moving matter, but one of transporting information. In other words, teleportation of a physical being will always be virtual and never real.

[0006] In prior art holography is the closest one can get to demonstrate virtual teleportation. However, holography based techniques are unsuitable for adaptation by the masses owing to high cost and time required for creating an extensive set up that requires a 3D laser infrastructure. Therefore, in this disclosure, a system and method for achieving virtual teleportation of one or more remote objects based on non-holographic technique is described. Such a system does not require expensive 3D laser holography infrastructure set up, can be used with any Internet connected computing device, can be installed in minutes, can teleport multiple objects simultaneously in High Definition (HD) resolutions to any real or fictionally created location.

[0007] Another need addressed by the present invention relates to perceptual computing, which is of great importance in entertainment, education, productivity, immersive gaming, social networking and most importantly in futuristic contactless or touch-less virtual interfaces. Currently known gesture based algorithms are unable to provide a fine degree of control needed in operating systems such as a computing device wherein data input with the precision of a keyboard, mouse or touchpad is required. They are limited to executing basic commands such as those of starting or stopping an application, or triggering an event. The present invention also addresses this need and finds utility in various applications such as video and computer gaming, entertainment, conferencing, interactive and perceptive computing to create an enriched user experience.

PRIOR ART

[0008] Chou et al., in U.S. Pat. No. 8,375,085 disclose a system and method of enhanced collaboration through teleportation, wherein several remote users collaborate in a virtual environment. Dawson et al., in U.S. Pat. No. 8,392,839 disclose a system and method for using partial teleportation or relocation in virtual worlds by using a facility such as a virtual kiosk at which one can view previously rendered images of portions of one or more teleportation destinations. When full teleportation is initiated, the processing intensive rendering is supplemented from image data representing the portion or the teleportation destination, thus hiding rendering delays and simulating seamless, instantaneous response to the full teleportation operation.

[0009] Shin et al., in PCT application no. WO 2001011511 disclose an electronic commerce system and method over three-dimensional virtual reality space, wherein an interactive electronic commerce system using intelligent virtual objects in a three-dimensional virtual reality space over a communication network such as the Internet, is provided. Altberg et al., in PCT application no. WO 2008130842 describe methods and systems to connect people via virtual reality for real time communications with respect to a commercial advertisement. Kikinis et al., in PCT application no. WO 2012054231 disclose a system and method for 3D projection and enhancements for interactivity wherein the system projects a user-viewable, computer-generated or -fed image, wherein a head-mounted projector is used to project an image onto a retro-reflective surface, so only the viewer can see the image.

[0010] The above patent disclosures emphasize the need of a cost effective and easy to set up system for achieving virtual teleportation of a plurality of remote objects for enhancing user experience in perceptual computing, entertainment, social networking, gaming, virtual interfaces, education, conferencing applications, or such applications. However, none of the systems and methods described in the prior art disclosures use a non-holographic technique to create a teleportation experience in a customized environment.

BRIEF SUMMARY OF THE INVENTION

[0011] The present invention addresses the foregoing need for an easy to install, cost effective virtual teleportation method and system thereof. It discloses a non-holographic method of virtually teleporting any object from a three dimensional space of a remote location, into a customized environment wherein one or more background layers or background layers of computer graphics may be inserted to render a seamless 3D environment in which all the teleported objects can interact. The system does not employ any laser holography means for recording or replaying image or scene. Furthermore, it also does not deploy any physical chroma screens or chroma keying techniques for background removal, thereby making it an effective and accessible system for the masses.

[0012] The present invention is directed to hardware, systems, methods, programs, computer products, computer readable media, interfaces and modules for controlling one or more operating parameters of one or more teleporting devices coupled to one or more computing devices with internet connectivity in different geographic locations for generating a composite scene comprising of a plurality of remote objects against a single customized 3D environment in real time. Accordingly, there is a need for such non-holographic virtual teleportation system and method as summarized herein in some detail.

[0013] It is an object of the present invention to provide a virtual teleportation platform wherein interactive video conferencing between several remote objects (participants) is possible, thereby eliminating the need for expensive travel. Other objects of the invention include providing a virtual teleportation system wherein depth, shape and gesture related interactive features are provided to create a realistic and believable 3D environment set up.

[0014] A further object of the present invention is to designate a virtual three-dimensional space around a host object at a host teleportation terminal, comprising of one or more human subjects, and to selectively teleport, i.e. virtually transmit images of one or more of remote objects located at different remote teleportation terminals at diverse geographical locations, to the virtual three dimensional space of the host object at the host teleportation terminal.

[0015] It is therefore an object of the present invention to provide a new method of virtually teleporting one or more remote objects to a predetermined, customized scene or environment without employing holographic techniques. This object is achieved by, firstly, extracting in real time one or more images of remote object(s) from the camera scene by deploying known foreground and background segmentation techniques, such as 2D pixel recognition and differentiation, or by 3D depth mapping and differentiation; secondly, transmitting and inserting one or more extracted objects into a single teleported host scene; thirdly, inserting layers of computer graphics as background and foreground layers, fourthly, compositing the final teleported scene with the objects and computer graphic layers properly aligned to generate a seamless virtual 3D environment in which each of the teleported objects can interacts.

[0016] It is further object of the present invention to provide a system that can be readily and easily used without the need for a sophisticated 3D laser infrastructure set up, and without any specialized knowledge of using 3D recording and replaying paraphernalia. In different embodiments of the present invention, the foreground/background segmentation is achieved by known techniques such as 2D pixel recognition and differentiation, or 3D depth measurement achieved by either dual camera triangulation, or multiple camera array that uses micro or nano lenses, or by means of Time-of-Flight (ToF) sensors, or by means of structured light based sensing methods.

[0017] Yet another object of the present invention is to provide a multi-layered perceptual computing interface (MLPCI) that virtually teleports all interactive tools such as keyboard, mouse, pen, page scroller, onscreen navigation links, icons, etc., into user's 3D space, wherein the user can perform sophisticated data input gesture-based interactions without using any proximity hardware like controller, gloves or near field sensor. It is further object of the present invention to enable execution of complex commands such as typing text, otherwise impossible by any gesture recognition system of prior art. It is therefore an object of the invention to enable MLPCI by replacing the background layer with the live screen view of the application that user is interfacing with. It is also further object of the invention to insert the graphic elements of virtual interactive tools as one or more translucent foreground layers and make the user transparent/translucent so that the virtual translucent interface apparently appears to float in the 3D space between the user and the display screen. It is therefore an object to allow user's transparent silhouette guide the user to move his hands and fingers in air for executing any command of desire via the virtual interface with ease.

[0018] It is thus an eventual object of the instant invention to create a virtual 3D space around a user and teleport almost anything imaginable within that space, such as friends, family, date, colleagues, store merchandise, venues, events, services, environments, user interfaces and even dreams, for the purpose of intuitive computing, perceptual computing, entertainment, gesture based interfaces, gaming, advertising, ecommerce, social networking, virtual training, education, so on and so forth.

[0019] Accordingly, a teleporting device is configured in different embodiments to include singly or in combination, one or more digital RGB cameras, 3D depth sensing (time of flight) ToF camera, 3D depth sensing camera based on structured light imaging, or dual camera depth sensor, or multiple camera array, that can be coupled to an internet enabled computing device.

[0020] These advantages in addition to other objects and advantages of the invention will be set forth in the description which follows, and in part will be obvious from the description, or may be learned by the practice of the invention. The objects and advantages of the invention may be realized and obtained by means of software, algorithms, hardware devices, remote servers and combinations thereof particularly pointed out in the appended claims.

[0021] The foregoing discussion summarizes some of the more pertinent objects of the present invention. These objects should be construed to be merely illustrative of some of the more prominent features and applications of the invention. Applying or modifying the disclosed invention in a different manner can attain many other beneficial results as will be described. Accordingly, referring to the following drawings may have a complete understanding of the invention. Description of several preferred embodiments is as follows.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[0022] The above-mentioned and other features and advantages of this present disclosure, and the manner of attaining them, will become more apparent, and the present disclosure will be better understood by reference to the following description of embodiments of the present disclosure taken in conjunction with the accompanying drawings, as illustrated herein:

[0023] FIG. **1**(*a*): A representative diagram illustrating an embodiment of a system facilitating virtual teleporting of three remote objects in real time using a client-server network architecture.

[0024] FIG. **1**(*b*): A representative diagram illustrating an embodiment of a system facilitating virtual teleporting of three remote objects in real time using a Peer-to-Peer architecture.

[0025] FIG. **1**(*c*): A representative diagram illustrating an embodiment wherein a user initiates teleportation session by activating Teleportation Icon **140** on a virtual interface by means of hand gesture.

[0026] FIG. **1**(*d*): A representative diagram illustrating an embodiment wherein an invitation to teleport is sent to selected participants from contact list.

[0027] FIG. **1**(*e*): A representative diagram illustrating an embodiment wherein the invited participants accept the invite to teleport.

[0028] FIG. **1**(*f*): A representative diagram illustrating an embodiment wherein the initiator of the teleportation session selects the location to teleport participants (such as under water expedition).

[0029] FIG. **1**(*g*): A representative diagram illustrating an embodiment wherein the teleportation process commences.

[0030] FIG. **1**(*h*): A representative diagram illustrating an embodiment wherein the initiator is being teleported to the desired virtual location.

[0031] FIG. **1**(*i*): A representative diagram illustrating an embodiment wherein the invitees to the teleportation session join the initiator of the teleportation session at the selected virtual location.

[0032] FIG. **1**(*j*): A representative diagram illustrating an embodiment wherein the session initiator ends the teleportation session by activating “End Session” option on a virtual interface by means of hand gesture.

[0033] FIG. **2**: A representative diagram illustrating an alternate embodiment of a system facilitating virtual teleporting for an e-commerce application.

[0034] FIG. **3**: A representative diagram illustrating a multi-layered perceptual computing interface (MLPCI) embodiment for creating a virtual interface.

[0035] FIG. **3**(*a*): Illustrative screenshot of a virtual interface embodiment of the present invention wherein the user gestures the START icon **316** for launching the application.

[0036] FIG. **3**(*b*): Illustrative screenshot of a virtual interface embodiment wherein the user launches the target application **318** (Internet Browser) by hand gesture.

[0037] FIG. **3**(*c*): Illustrative screenshot of a virtual interface embodiment wherein the user activates the cursor by launching the virtual touchpad **320** to move the cursor on the target application page.

[0038] FIG. **3**(*d*): Illustrative screenshot of a virtual interface embodiment wherein the user launches and uses the virtual keyboard **322**.

[0039] FIG. **3**(*e*): Illustrative screenshot of a virtual interface embodiment wherein the user launches and uses the virtual page scrolling function **324**.

[0040] FIG. **3**(*f*): Illustrative screenshot of a virtual interface embodiment wherein the user uses gesture controlled drawing tool function **326** to draw and to close the application.

[0041] FIG. **4**: An exemplary block diagram illustrating the operating modules of a virtual teleporting system.

[0042] FIG. **5**: An illustration of the process flow steps deployed in an embodiment of the present invention.

[0043] FIG. **6**: A representative diagram illustrating integration of different modules of the embodiment within a television set or a display monitor.

DETAILED DESCRIPTION OF THE INVENTION

[0044] It is to be understood that the present disclosure is not limited in its application to the details of construction and the arrangement of components set forth in the following description or illustrated in the drawings. The present disclosure is capable of other embodiments and of being practiced or of being carried out in various ways. Also, it is to be understood that the phraseology and terminology used herein is for the purpose of description and should not be regarded as limiting. Exemplary embodiments of the present invention are directed towards a system and a method for facilitating non-holographic virtual teleporting in real time directed at one or more remote objects.

[0045] It is advantageous to define several terms, phrases and acronyms before describing the invention in detail. It should be appreciated that the following terms are used throughout this application. Where the definition of term departs from the commonly used meaning of the term, applicant intends to utilize the definitions provided below, unless specifically indicated. For the purpose of describing the instant invention following definitions of the technical terms are stipulated:

[0046] 1. 2D foreground segmentation: Two dimensional image segmentation techniques for distinguishing a background from a foreground, such as those based on pixel recognition and differentiation. The 2D pixel differentiation technique requires a training session so that the system first reads and learns the pixel pattern in a given scene and uses it as a reference to extract target object from the background.

[0047] 2. 3D background segmentation: Three dimensional image segmentation techniques such as those based on depth measurement techniques by dual camera triangulation algorithm, multiple camera array, Time-of-Flight 3D mapping, or structured light based 3D imaging processes.

[0048] 3. Teleporting device/Teleporting terminal: A computer device or a combination of devices coupled to an internet enabled computing device comprising of at least one RGB (Red, Green, Blue) sensor, which is a high definition camera with HD resolution of 720P (1280×720 pixels) or Full HD resolution of 1080P (1920×1080 pixels) that captures the live object images at each of the teleporting terminal; at least one audio sensor, and at least one depth sensor, which is based on generating 3D maps of the objects at each of the teleportation terminal using one of the following approaches: a) Time-of-Flight based depth sensing, b) Structured Light based 3D sensing, c) using Dual Camera based triangulation algorithm, or d) by means of Multiple Camera Array method. The complete assemblage of Internet-connected teleportation enabling hardware and software at a specific geo-location is referred as teleporting terminal. A teleporting terminal that deploys two-dimensional segmentation may not need the depth sensor.

[0049] 4. Time-of-Flight Depth Sensor: A time-of-flight sensor or camera (ToF camera) is a range imaging camera system that resolves the depth of an object by measuring the time-of-flight of a light signal between the camera and the subject for each point of the image distance based on the known speed of light.

[0050] 5. Structured Light Depth Sensor: Another type of depth sensor that is used for measuring the three-dimensional shape of an object using projected light patterns and a camera system.

[0051] 6. Dual Camera Depth Sensor: Depth of a three-dimensional object can also be measured by using two RGB cameras for capturing image data points on the object from two different perspectives, and then using triangulation algorithm to analyze all the data points for creating a 3D map of the object.

[0052] 7. Multiple Camera Array: Using an array of multiple cameras to record a scene is a more recent technique that is used to generate pixel specific data from multiple perspectives. In this system each pixel carries depth information in addition to the normal RGB data. Such pixel-specific information can be deployed in many different ways, one of which being creating a depth map of the image. An array may comprise of either multiple micro cameras with their own imaging sensors, or a collection of miniaturized micro or nano lenses focusing light from different perspectives on a single processor chip collectively analyzing the entire image information from all perspectives to measure the depth of each pixel from each perspective.

[0053] 8. Remote object/s & Remote terminal/s: One or more human subjects, living beings, physical objects or articles located in a designated three dimensional space in a single or multiple geographical locations who are to be virtually teleported to a different location. The complete assemblage of teleportation enabling hardware and software infrastructure at such a remote location is referred as remote terminal.

[0054] 9. Host Object/s: One or more human subject, living being located in a designated three-dimensional space in a geographical location, who initiate a teleportation session with one or more remote teleportation terminals, and who are to be virtually teleported to a pre-defined virtual location. The complete assemblage of teleportation enabling hardware and software at such a host location is referred as host terminal.

[0055] 10. Camera view background: The local background environment surrounding each of one or more remote object(s) or host object(s) from which the corresponding object(s) have to be extracted and transmitted with alpha channel (transparency) to the host terminal, so that various teleported or computer generated graphic elements can be integrated into a composite teleported scene.

[0056] 11. Teleported scene background layer: The background layer comprising of one or more preselected computer generated graphic content, retrieved from a repository, for insertion into the integrated, composite teleported scene.

[0057] 12. Teleported scene foreground layer: The foreground layer comprising of one or more preselected computer generated graphic content, retrieved from a repository, for insertion into the integrated, composite teleported scene.

[0058] 13. Elements of a teleported scene: The remote object(s), host object(s), computer generated background graphics layer(s), computer generated foreground graphics layer(s), and alpha and audio channel associated with each of them constitute the elements of the teleported scene.

[0059] 14. Remote Object Connection Module (ROCM): A set of computer programs that is used to logically connect one or more remote objects in real time via internet means that include but are not limited to a telecommunication link, or a wired or wireless local area network (LAN), or a wide area network (WAN), or a virtual private network (VPN) or intranet through wired or wireless telecommunication protocol, or TCP/IP protocol, or GPRS protocol or WiFi protocol or Bluetooth or radiofrequency protocol or a telecommunication protocol.

[0060] 15. Object Scene Capture Module (OSCM): A set of computer programs that receive, as input from RGB sensor, the image of object scene in real time from the teleporting device

[0061] 16. Object Extraction Module (OEM): A set of computer programs that rely on 2D or 3D background and foreground segmentation techniques to extract in real time scene objects from their backgrounds.

[0062] 17. Remote Object Transmission Module (ROTM): A set of computer programs that transmit in real time, to a host object terminal, the extracted image of a remote object along with associated parameters, those include but not limited to depth, texture, color, alpha channel and audio channel.

[0063] 18. Object Insertion Module (OIM): A set of computer programs that integrates, places and composes one or more remote teleported objects within alpha channeled areas of extracted host object image scene.

[0064] 19. Foreground Layer Insertion Module (FLIM): A set of computer programs that inserts in real time a predefined foreground layer of computer graphics to the composition of teleported scene.

[0065] 20. Background Layer Insertion Module (BLIM): A set of computer programs that inserts in real time a predefined background layer of computer graphics to the composition of teleported scene.

[0066] 21. Teleported Scene Compositing Module (TSCM): A set of computer programs that integrates and fine tunes in real time the composition of all the elements of the teleported scene, such as host and remote objects, background and foreground layers of computer graphics, alpha channel and audio channel to produce a live composition of teleported environment.

[0067] 22. Teleported Composite Scene Display Module (TCSDM): A set of computer programs that display in real time the integrated, composite teleported scene on display devices of each of the participating teleportation terminals.

[0068] 23. Communication Module (CM): A set of computer programs which delivers the integrated, composite teleported scene in real time to a preselected destination using either wired or wireless telecommunication protocol, or TCP/IP protocol, or WiFi protocol or Bluetooth, or a radiofrequency protocol.

[0069] 24. Central Processing Unit (CPU): A computing system that analyzes and executes the operations of ROOM, OSCM, OEM, ROTM, OIM, FLIM, BLIM, TSBLIM, TCSM, TCSDM.

[0070] The present invention is now described with reference to the drawings. An overview of a system in a preferred embodiment facilitating virtual teleportation of three remote objects in a teleportation session to a new real or virtual location using non-holographic technique is now discussed in conjunction with FIGS. **1**(*a*) and **1**(*b*). FIG. **1**(*a*) illustrates the technical implementation of virtual teleportation in client server network architecture, while FIG. **1**(*b*) illustrates implementation in a peer-to-peer network. The system **100** includes a plurality of objects **104**, **112** and **124** located at plurality of teleportation terminals who are to be teleported. The host object **102** at host terminal initiates the teleportation session. Host object is present in a backdrop comprising of host background elements **104**. The host scene is captured by the RGB sensor (camera) of a teleporting device **106** coupled to a computing device **108**. After removal of background elements **102**, **110** is the extracted alpha channeled RGB image with audio of the host object **104**. The extracted image data **110** is transmitted from the host terminal via Internet through wired or wireless means to a real time teleportation server **132**. In a peer-to-peer network the teleportation server **132** is not required FIG. **1**(*b*) since the elements of the scene at the host and each of the participating terminals are transmitted directly between them in real time.

[0071] A remote object **112** at a remote teleportation terminal is present in a remote background **114**. The remote scene at the remote terminal is captured by teleporting device **116** coupled to a computing device **118**. **120** denotes the alpha channeled extracted image with embedded audio of remote object **112**, which is transmitted via internet through wired or wireless means. Another remote object **122**, at another remote teleportation terminal, is present along with remote background element **124**. The remote scene at this remote terminal is captured by teleporting device **126** coupled to a computing device **128**. **130** denotes the extracted image data of remote object **124**, which is transmitted along with its corresponding alpha channeled background via internet through wired or wireless means. In a client-server network architecture FIG. **1**(*a*), the teleportation server **132** inserts and manipulates each of the transmitted alpha-channeled object image data **110**, **120** and **130** into the alpha channeled background areas of the host environment, so that each object image is overlaid distinct from each other on its pre-defined location in the teleported scene or environment. The environment that includes all the teleported objects participating in the teleportation session is then enhanced by inserting one or more background layers **134**, and one or more foreground layers **136** into a composite teleported scene **138**. Such background and foreground layers are retrieved from a database of computer generated graphic content. The final integrated composite teleported scene or environment **138** is then displayed on a display panel at each of the participating terminals. Such display is either a plasma display panel, or an LCD (liquid crystal display) panel, or an LED (light emitting diode), or an OLED (organic light emitting diode) display panel, or a video projector, or a see-through display screen, or a television set.

[0072] As illustrated in FIG. **1**(*b*), in a peer-to-peer network, all the data processing steps are distributed and shared between the participating teleportation terminals, and the composite teleported scene **138** is generated at the client terminal itself and shared with each of the participating terminals seamlessly.

[0073] Each of the teleportation terminals is equipped with means to initiate, modify, pause or record a teleportation session, to invite, add or delete participants in a teleportation session. The participants of a teleportation session can also chose environment or virtual location (as defined by the foreground and background layers) they want to be teleported in. A practical implementation of the present invention can be a virtual conference taking place between participating teleportation terminals in a virtual environment created by the computer generated elements of the teleported scene simulating either a real world environment or a fictional environment. The integrated composite teleported scene of a teleportation session can be either recorded locally at one or more teleportation terminals, or broadcast live and made instantly available to one or more preselected remote destinations via an Internet, or a television satellite link, or a telecommunication link, or a wired or wireless local area network (LAN), or a wide area network (WAN), or a virtual private network (VPN) or intranet.

[0074] For connecting the participating teleportation terminals with each other the system uses any one of the communication protocols known to prior art such as TCP/IP protocol, GPRS protocol, WiFi protocol, a telecommunication link, a wired network, a wireless network, a virtual private network, intranet, wireless telecommunication protocol, Bluetooth, or a radiofrequency protocol.

[0075] In a variant of this embodiment one or more participating teleportation terminals deploy chroma-keying techniques for background removal. In yet another variant of the embodiment the host teleportation terminal does not deploy background removal. In another variant of this embodiment, the teleportation device coupled to its Internet enabled computing device is integrated within an LCD panel, an LED, an OLED display panel, a video projector, or a see-through display screen, or a television set. In yet another variant, the teleportation terminal is handheld communication device, or a head-mounted teleportation apparatus.

[0076] Having disclosed the technical aspects of a preferred embodiment is some detail, it is pertinent to walk through the practical implementation of the instant invention. FIG. **1**(*c*) through FIG. **1**(*j*) illustrate various steps from initiating a teleportation session to ending the session. While a convention keyboard/mouse can be used for executing all the teleportation related commands, all the commands can also be executed via gestures using the virtual interface of the invention as depicted herein. The teleportation session is initiated by activating the Teleportation Icon **140** on a virtual interface by means of hand gesture. The invitation to teleport is sent to selected participants from contact list FIG. **1**(*d*). The invited participants accept the invite FIG. **1**(*e*). The initiator of the teleportation session selects the location (such as under water expedition) session FIG. **1**(*f*), which is followed with teleportation of the participants FIG. **1**(*g*), FIG. **1**(*h*), FIG. **1**(*i*) to an underwater ship. Finally, at the conclusion of the teleportation session, the session initiator ends the session FIG. **1**(*j*).

[0077] Referring now to FIG. **2**, which illustrates another embodiment of the system facilitating virtual teleporting for an e-commerce application, in which a user is teleported into a 3D environment of a virtual store or showroom, a virtual shopping mall, or, a merchandise service center, and interacts with merchandises therein. In a variant of this embodiment, one or more merchandises are teleported into user's 3D environment for product demonstration, pre-purchase preview of goods, technical support, troubleshooting and service of pre-owned goods. **202** denotes a human subject or a user at a teleportation terminal **204**.

[0078] The image parameters, audio and video data of **202** are captured via teleporting device **206** and transmitted to teleportation server **208** that hosts a database containing specifications, images, computer graphics and 3D models of merchandise, such as a given model of a washing machine **210**. Upon initiation of a teleportation session by the user, the teleportation server **208** retrieves the computer-generated graphics pertaining to the washing machine **210** and transmits the merchandise data to the teleportation terminal **204**, where it is inserted in the teleported scene **212** in user's 3D space as a foreground layer appearing in front of the user. Such an embodiment can be used for applications wherein several customers can be serviced remotely and conveniently for sales, marketing or troubleshooting of a product and thereby eliminating the need for a personal visit. Similar embodiments can be used in applications such as teaching and providing education to students remotely.

[0079] In another variant of ecommerce teleportation embodiment, instead of teleporting the merchandise object into user's 3D space, the user is teleported to a 3D environment of a virtual shopping mall or a store. Briefly, in the manner described in the first preferred embodiment as illustrated in FIG. **1**, the user image captured at the user's teleportation terminal is extracted from the user environment and transmitted to the teleportation server where it is inserted into a teleported scene that includes foreground and background layers of computer graphics that create a real time in-store or in-mall shopping experience for the user.

[0080] The teachings of the instant invention is not limited to virtual teleportation of objects such as human subjects or material objects into a user's 3D space, but data entry means, navigation controls, or icons can also be teleported into user's 3D space between the user and the computer display, serving as voice or gesture responsive virtual interface. Such means of gesture or voice responsive computer-generated elements of virtual interface are retrieved from the application database and inserted as one or more foreground layers of the composite scene. These computer-generated graphic elements of virtual interface include:

[0081] a) virtual keyboard, virtual mouse, virtual touchpad, virtual pen of varying transparency to make other elements of the composite teleported scene behind the foreground layer visible to enable data input,

[0082] b) virtual menu for accessing different teleportation functions as well as navigating to other co-existing and unrelated client applications that include but not limited to, mail client, Internet browser, social networking applications, gaming applications, productivity applications,

[0083] c) virtual icons for displaying application links and alerts in real time.

[0084] As illustrated in FIG. **3**, such a system of a virtual interface application comprise of a computer (teleportation terminal) operated by a user **302** whose image along with his 3D space data is captured by a teleporting device (depth sensor plus RGB camera) **304** coupled to user's computer **306**. The computer screen displays the interface of the target application **308** the user is working on. Such target application has fields in which the user is required to enter personal information by typing text. Implementing the features of the instant invention the user can perform all actions without using any data input hardware. This is achieved by means of first generating a 3D map of user's 3D space by means of data captured by the teleporting device **304** that is used for tracking user's hand movements. Computer-generated graphic elements of virtual interface from the application database are then retrieved and inserted as a foreground layer in user's 3D space. Such elements of virtual interface are rendered translucent so that the target application on the display screen is operationally visible to the user. Virtual keyboard **310** is an example of the elements of virtual interface. An association of each key on the virtual keyboard is established with the 3D map of user's 3D space, so that when user moves his finger in air to overlap on a specific key on the virtual keyboard the system matches the location of the specific key with the location of user's finger. To make both, the target application and virtual keyboard, clearly visible on screen, the user's image is rendered translucent **312**. The screen view of the target application can be either rendered visible through the transparent foreground layer or captured in real time and screen capture inserted as background layer **314** of the user's camera view. The ultimate effect of this composition of the translucent foreground layer, the translucent user image and the target application's screen image in the background of the composition is that the target application, the virtual keyboard and user's fingers are all clearly visible to the user, and gives him an impression like the virtual keyboard is floating in air in between him and the target application interface displayed on the computer screen.

[0085] FIG. **3**(*a*) through **3**(*f*) further illustrate practical implementation of the instant embodiment wherein the user begins the method by launching the virtual interface application's START icon **316** by hand gesture, and then launching the target application **318** (Internet Browser) by hand gesture FIGS. **3**(*a*) and **3**(*b*). The user then activates the cursor by launching the virtual touchpad **320** to move the cursor to the data entry field on the target application page FIG. **3**(*c*). To type the information in the data entry field FIG. **3**(*d*) the user launches and uses the virtual keyboard **322**. To move and scroll the page FIG. **3**(*e*) the user can also launch the virtual page scrolling function **324**. The user can also use the drawing tool function **326** to close the application FIG. **3**(*f*).

[0086] Such an embodiment as illustrated in FIGS. **3** and **3**(*a*)-**3**(*f*) enables the user to perform highly sophisticated operations remotely using hand movements and without touching any data input hardware or using any wearable or gesture recognition proximity hardware. Combining the disclosed virtual interface with voice-activated commands can impart unprecedented versatility to the instant invention. Such virtual interface embodiment can have widespread application as a visual and voice interface in gaming and entertainment facilitating sophisticated gesture based interaction and navigation. In a variant of virtual teleportation of virtual interface tools in user's virtual 3D space, the interface does not require any START or HOME button, but can just trigger the application links and navigation and other tools by just moving his hand to the edge of the display screen. Each of the four edges of the display screen can be assigned to specific group of virtual interface tools.

[0087] Referring now to FIG. **4**, the drawing illustrates an exemplary block diagram depicting the operating modules of the system **400** for facilitating virtual teleporting of remote objects into user's 3D space using a non-holographic technique. The operating modules referred to hereunder may be present on a single or plurality of computing devices present at a single teleportation terminal location, or at plurality of teleportation terminals, or at remote teleportation servers.

[0088] System **400** comprises of Object Connection Module (OCM) **402**, Object Scene Capture Module (OSCM) **404**, Object Extraction Module (OEM) **406**, Remote Object Transmission Module (ROTM) **408**, Object Insertion Module (OIM) **410**, Foreground Layer Insertion Module (FLIM) **412**, Background Layer Insertion Module (BLIM) **414**, Teleported Composite Scene Integration Module (TCSIM) **416**, Teleported Composite Scene Display Module (TCSDM) **418**, Communication Module (CM) **420**, Computer Generated Foreground/Backdrop content database **422**. These modules may be hosted on a local or remote server or distributed at multiple remote locations. Component **422** provides a means for storage of computer-generated foreground and background content including data in audio, video, animation, 3D image, map or text format. **424** serves as a repository of preselected computer-generated foreground and background content.

[0089] Component **402** is responsible for connecting one or more teleportation terminals with each other and to a common teleportation server. This connection may be achieved by wired or wireless means via an Internet or a telecommunication link, or a wired or wireless local area network (LAN), or a wide area network (WAN), or a virtual private network (VPN) or intranet. Component **404** is responsible for capturing the object scene comprising of a object along with object background by means of a teleporting device capable of sensing image attributes and depth parameters. Instances of teleporting devices include time-of-flight sensor based camera, structured light sensor based camera, dual camera depth sensor based camera, multiple camera array and digital image capturing devices capable of 2D pixel differentiation. Component **406** is responsible for distinguishing and separating the object from its immediate background using foreground/background segmentation techniques. Component **408** is responsible for transmitting selectively or completely, the extracted object image along with other image attributes such as RGB color, texture, audio, alpha channel from one teleportation terminal to the other or to the teleportation server. Component **410** combines the plurality of remote object images received from multiple remote locations for insertion into an integrated composite scene rendering individual remote images as part of the composite scene by manipulating them in terms of their image attributes. Component **412** is responsible for insertion of a layer of preselected computer graphics as a foreground layer while component **414** is responsible for insertion of a layer of preselected computer graphics as a background layer in the teleported integrated composite scene. Component **416** integrates said foreground layer, said background layer and remote object images in relation to each other for rendering a seamless customized view of a teleported scene. Component **418** is responsible for displaying the teleported integrated composite view in a display device such as a plasma display panel, an LCD (liquid crystal display) panel, an LED (light emitting diode), an OLED (organic light emitting diode) display panel or a video projector. Component **420** is responsible for transmitting and communicating the teleported integrated composite view to preselected destinations including, a handheld communication device, an email account, downloadable URL link of a remote server, so on and so forth. In different embodiments of present invention, the teleported integrated composite view is transmitted through wired or wireless telecommunication protocol, or TCP/IP protocol, or GPRS protocol or WiFi protocol or Bluetooth or radiofrequency protocol or IMAP, SMTP or a telecommunication protocol.

[0090] FIG. **5** depicts an exemplar methodology illustrating the steps followed in one aspect of the invention. It is to be understood and appreciated that the present invention is not limited by order of steps and that some of the steps may occur in different order and/or concurrently with other steps from that illustrated here. At step **502**, one or more objects, located at geographically diverse teleportation terminals, which are required to be virtually teleported are connected via Internet means in real time. At step **504**, the object scene at each of the participating terminals is captured in RGB along with its depth data by means of a teleporting device. At step **506**, the remote object image is extracted from the remote background using one or more of foreground-background segmentation techniques. At step **508**, image parameters of captured object at each of the participating teleportation terminals, such as color, texture, audio, alpha channel are transmitted either to a teleportation server when client-server network architecture is deployed, or directly to participating peers if peer-to-peer network architecture is deployed. At step **510**, the transmitted object images are inserted and appropriately placed in relation to each other into a single composite scene. At step **512**, a layer of computer-generated graphics that can act as virtual foreground layer is accessed from a database and inserted in the teleported scene. At step **514**, a layer of computer-generated graphics that can act as virtual background layer is accessed from a database and inserted in the teleported scene. At step **516**, each element of the teleported scene, i.e. each of the extracted remote objects, the extracted host object, the foreground and background layers of computer generated graphics, alpha channels, along with their corresponding audio channels are composed into a final composite teleported scene. At step **518**, the integrated composite teleported scene is displayed in real time on a device such as a plasma display panel, an LCD (liquid crystal display) panel, an LED (light emitting diode), an OLED (organic light emitting diode) display panel or by a video projector. The displayed integrated teleported content including the composite background scene and remote objects is transmitted to one or more preselected destinations such as, a handheld communication device, an email account, downloadable URL link of a remote server, so on and so forth through wired or wireless telecommunication protocol, or TCP/IP protocol, or GPRS protocol or WiFi protocol or Bluetooth or radiofrequency protocol or a telecommunication protocol.

[0091] FIG. **6** illustrates another embodiment of the present invention in which the teleporting device **602** along with its computing hardware **604** and all the different modules are integrated within a television set or a display monitor so that such television set or display monitor itself functions as a fully functional teleportation terminal. Alternatively, this embodiment can also be practiced by compiling all the components (except the display device) of the teleportation system into a kit that can be provided as a teleportation accessory to a television set or a display panel.

[0092] In an embodiment of the present invention, the computer-generated graphics layers that may be used as foreground/background layers in the integrated teleported scene constitute a video gaming environment comprising of one or more virtual game characters or elements that the remote object(s) interact(s) with immersively, whether using hand or body gestures, or voice activated commands, or by using handheld game controller device.

[0093] In yet another embodiment of the present invention, only a part of the remote objects are teleported to the remote destination. This is achieved by manipulating the image of the remote object such that it appears partly transparent. Optionally, some of the remote objects may be rendered totally transparent, while some may be rendered translucent and some may be rendered as such while transmitting the image attributes such as color, texture, alpha channel, and audio channel. This embodiment can be used for providing a sophisticated gesture based computing interface.

[0094] In one embodiment of the present invention one or more teleportation terminals deploy chroma-keying techniques for background removal and object extraction. This is particularly important in showcasing a product or service such as an automobile, or a direct sales session for a fast moving consumer product to a large audience physically present in a high footfall public place such as shopping malls. In such an embodiment host teleportation terminal (public place) does not deploy any background removal or object extraction, while the remote terminal may either deploy chroma-key based or depth-based background removal. In another embodiment of the present invention, the plurality of remote objects may be the participants of a virtual conference, workshop, virtual training, etc., wherein the remote object image captured at remote teleporting terminal along with associated audio and video data is transmitted into the teleported scene at host terminal.

[0095] In a yet another embodiment of the present invention a useful, confidential and private means of one-on-one interaction is provided for communicating with strangers, such as encountered in social and dating networks. In this embodiment the user is teleported anonymously into a secure and private 3D environment of another unknown user without sharing any of their personal identification for the purpose of socializing or going on a blind date with a stranger. In still another embodiment the computer programs of one or more modules run remotely from a server via an active server webpage, or operate as a browser plugin, or run from an external drive of a computer.

[0096] Although the present invention has been particularly shown and described with reference to exemplary embodiments thereof, it will be understood by those of ordinary skill in the art that various changes in form and details may be made therein without departing from the spirit and scope of the present invention as defined by the appended claims. Therefore, the present embodiments are to be considered as illustrative and not restrictive and the invention is not to be limited to the written description.

The invention claimed is:

1. A computer-implemented method of virtually teleporting non-holographically, in real time, a plurality of objects or users located at diverse geographical locations, into a customized, seamless 3D environment, using a plurality of teleportation terminals, wherein the method comprises of the steps of:

a) a host teleportation terminal initiating connection via Internet means to one or more remote teleportation terminals participating in a teleportation session;
b) capturing each remote scene at plurality of participating terminals comprising of at least one remote object in a remote background and at least one host object in host background, by means of RGB sensor of a teleporting device coupled to an internet enabled computing device;
c) extracting each of the plurality of RGB device-captured object images from their corresponding remote backgrounds by means of a plurality of teleporting devices, coupled to their corresponding internet enabled computing devices;
d) transmitting each of the extracted RGB-captured remote object images along with their corresponding alpha channeled background for compositing the final teleportation scene or environment;
e) inserting each of the transmitted alpha-channeled object images into the alpha channeled background areas of the host environment;
f) inserting one or more background layers of retrieved computer-generated content to produce an integrated composite teleported scene or environment that includes all the objects participating in the teleportation session;
g) inserting one or more foreground layer of retrieved computer-generated content in the integrated composite teleported scene or environment;
h) compositing the final teleported scene by manipulating each element of the scene, so that each of the plurality of object images is overlaid distinct from each other on its pre-defined location in reference to the foreground and background layers in the teleported scene or environment;
i) displaying in real time the final integrated composite teleported scene or environment on a display panel at each of the participating terminals.

2. A method of claim **1**, wherein the step (c) of extraction of object images from their corresponding backgrounds are achieved by 2D foreground segmentation technique comprising of 2D pixel recognition and differentiation.

3. A method of claim **1**, wherein the step (c) of extraction of object images from their corresponding backgrounds is achieved by depth differentiation using any one, or combination thereof, of the following 3D background segmentation techniques:

a) Time-of-Flight imaging,
b) Structured light imaging,
c) Dual camera triangulation,
d) Multiple camera array of micro or nano cameras or lenses.

4. A method of claim **1**, wherein data processing for one or more of the steps from a) through h) are:

a) shared equally by each of the participating terminals in a peer-to-peer network,
b) are implemented at a remote teleportation server in a client-server network.

5. A method of claim **1**, wherein the teleporting device is a combination of devices that are coupled to an internet enabled computing device and comprise of:

a) at least one RGB sensor, which is a high definition camera with resolutions not less than 720P that captures the live object images at each of the teleportation terminal;
b) at least one audio sensor; and,
c) at least one depth sensor, which is based on generating 3D maps of the objects at each of the teleportation terminal using one of the following algorithms:
    i) dual camera triangulation,
    ii) multiple camera array,

iii) Time-of-Flight,
iv) structured light.

6. A method of claim **1**, wherein the integrated composite teleported scene is either recorded locally at one or more teleportation terminals, or broadcast live and made instantly available to one or more preselected remote destinations via an Internet connection, or a television satellite link, or a telecommunication link, or a wired or wireless local area network (LAN), or a wide area network (WAN), or a virtual private network (VPN), or Intranet.

7. A method of claim **1**, wherein the plurality of remote teleportation terminals are connected by means selected from a group comprising of internet protocol, a telecommunication link, a wired network, a wireless network, a virtual private network, intranet, wireless telecommunication protocol, TCP/IP protocol, GPRS protocol, WiFi protocol, Bluetooth, or a radiofrequency protocol.

8. A method of claim **1**, wherein the foreground layer and the background layer of the integrated composite teleported scene constitute a 3D video gaming environment comprising of virtual game characters and elements that interact with one or more users immersively by means of voice, hand, body gestures or combination thereof.

9. A method of claim **1**, wherein the plurality of remote and host objects are participants of a virtual conference that takes place in a virtual environment created by the computer generated elements of the teleported scene simulating either a real world environment, or a fictional environment.

10. A method of claim **1**, wherein, either,

a) a user is teleported into a 3D environment of a virtual store or showroom, a virtual shopping mall, or, a merchandise service center, and interacts with merchandises therein, or,

b) merchandises are teleported into user's 3D environment for product demonstration, pre-purchase preview of goods, technical support, troubleshooting and service of pre-owned goods.

11. A method of claim **1**, wherein a user is teleported anonymously into a secure and private 3D environment of another unknown user without sharing any of their personal identification for the purpose of socializing or going on a blind date with a stranger.

12. A method of claim **1**, wherein one or more elements of the teleported scene are rendered transparent or translucent for providing pre-defined level of visibility to hidden or masked elements.

13. A method of claim **1**, wherein at least one background or foreground layer is replaced by a live screen capture of the desktop that a user is interacting with, and, one or more transparent or translucent foreground layers comprising of computer graphics that operate as virtual computing interface, displaying voice or gesture responsive data entry means, navigation controls or icons that include but not limited to:

a) virtual keyboard, virtual mouse, virtual touchpad, virtual pen of varying transparency to make other elements of the composite teleported scene behind the foreground layer visible to enable data input;

b) virtual menu for accessing different teleportation functions as well as navigating to other co-existing and unrelated client applications that include but not limited to, mail client, Internet browser, social networking applications, gaming applications, productivity applications;

c) virtual icons for displaying application links and alerts in real time.

14. A method of claim **1**, wherein the display is a plasma display panel, an LCD (liquid crystal display) panel, an LED (light emitting diode), an OLED (organic light emitting diode) display panel, a video projector, or a see-through display screen, or a television set.

15. A method of claim **1**, wherein the teleportation device coupled to its Internet enabled computing device is integrated within an LCD panel, an LED, an OLED display panel, a video projector, or a see-through display screen, or a television set.

16. A method of claim **1**, wherein the teleportation terminal is handheld communication device, or a head-mounted teleportation apparatus.

17. A method of claim **1**, wherein one or more participating teleportation terminals deploy chroma-keying techniques for background removal.

18. A method of claim **1**, wherein the host teleportation terminal does not deploy background removal.

19. A system of virtually teleporting a plurality of remote objects from plurality of remote teleportation terminals in real time into a customized, seamless 3D environment of a host object's host teleportation terminal, by means of a plurality of teleporting devices coupled to a plurality of internet enabled computing devices, comprising of:

a) Remote Object Connection Module (ROOM) which is a set of computer programs that is used to logically connect one or more remote objects via internet means;

b) Object Scene Capture Module (OSCM), which is a set of computer programs that receive, as input from RGB sensor, the image of object scene in real time from the teleporting device;

c) Object Extraction Module (OEM) which is a set of computer programs that deploy either 2D pixel recognition and differentiation techniques for foreground segmentation, or, 3D depth sensors for foreground-background segmentation to extract object with alpha channel from background of the captured object scene in real time;

d) Remote Object Transmission Module (ROTM) which is a set of computer programs that transmit in real time, to a host object terminal, the extracted image of a remote object along with associated parameters, that include but not limited to depth, texture, color, alpha channel;

e) Object Insertion Module (OIM) which is a set of computer programs that integrates, places and composes in real time one or more remote teleported objects within alpha channeled areas of extracted host object image scene;

f) Background Layer Insertion Module (BLIM), which is a set of computer programs that inserts a pre-defined background layer of computer graphics to the composition of teleported scene in real time;

g) Foreground Layer Insertion Module (FLIM), which is a set of computer programs that inserts a pre-defined foreground layer of computer graphics to the composition of teleported scene in real time;

h) Teleported Scene Compositing Module (TSCM), which is a set of computer programs that integrates and fine tunes in real time the composition of host and remote objects with background and foreground layers of computer graphics to produce a live composition of teleported environment;

i) Teleported Composite Scene Display Module (TCSDM) which is a set of computer programs that display in real time the integrated, composite teleported scene on display devices of each of the participating teleportation terminals;

j) Central Processing Unit (CPU), which analyzes and executes the operations of ROCM, OSCM, OEM, ROTM, OIM, FLIM, BLIM, TSBLIM, TCSM, TCSDM.

20. A method of claim **19**, wherein the integrated composite teleported scene is either recorded locally at one or more teleportation terminals, or broadcast live and made instantly available to one or more preselected remote destinations via an Internet connection, or a television satellite or a telecommunication link, or a wired or wireless local area network (LAN), or a wide area network (WAN), or a virtual private network (VPN) or intranet.

21. A method of claim **19**, wherein the plurality of remote objects are connected by means selected from a group comprising of internet protocol, a telecommunication link, a wired network, a wireless network, a virtual private network, intranet, wireless telecommunication protocol, TCP/IP protocol, GPRS protocol, WiFi protocol, Bluetooth, radiofrequency protocol, a telecommunication protocol.

22. A method of claim **19**, wherein the foreground layer and the background layer of the integrated composite teleported scene constitute a 3D video gaming environment comprising of virtual game characters and elements that interact with one or more users immersively by means of voice, hand, or body gestures.

23. A method of claim **19**, wherein the plurality of remote and host objects are participants of a virtual conference that takes place in a virtual environment created by the computer generated elements of the teleported scene simulating either a real world environment or a fictional environment.

24. A method of claim **19**, wherein, either,

a) a user is teleported into a 3D environment of a virtual store or showroom, a virtual shopping mall, or, a merchandise service center, and interacts with merchandises therein; or,

b) merchandises are teleported into one or more users' 3D environment for product demonstration, pre-purchase preview of goods, technical support, troubleshooting and service of pre-owned goods.

25. A method of claim **19**, wherein a user is teleported anonymously into a secure and private 3D environment of another unknown user without sharing any of their personal identification for the purpose of socializing or going on a blind date with a stranger.

26. A system of claim **19**, wherein one or more foreground layers comprise of computer graphics that operate as virtual computing interface displaying voice or gesture responsive data entry means and navigation controls or icons that include but not limited to:

a) virtual keyboard, virtual mouse, virtual touchpad, virtual pen of varying transparency to make other elements of the composite teleported scene behind the foreground layer visible to enable data input;

b) virtual menu for accessing different teleportation functions as well as navigating to other co-existing and unrelated client applications that include but not limited to, mail client, Internet browser, social networking applications, gaming applications, productivity applications;

c) virtual icons for displaying alerts in real time.

27. A system of claim **19**, wherein the computer programs of one or more modules run remotely from a server via an active server webpage, or operate as a browser plugin, or run from an external drive of a computer.

28. A system of claim **19**, wherein a Communication Module (CM) which is a set of computer programs that stream live the integrated, composite teleported scene to one or more preselected remote destinations.

29. A system of claim **19**, wherein one or more teleportation terminals deploy chroma keying techniques for background removal.

30. A method of claim **19**, wherein the host teleportation terminal does not deploy background removal.

31. A system of claim **19**, wherein modules (a) through (h) and module (j) are compiled as a kit that can be incorporated into a television set or a display panel.

* * * * *

Agrawal et al., *Cogent Engineering* (2017), 4: 1332705
https://doi.org/10.1080/23311916.2017.1332705

ELECTRICAL & ELECTRONIC ENGINEERING | RESEARCH ARTICLE

A dual-band RF energy harvesting circuit using 4th order dual-band matching network

Sachin Agrawal[1]*, Manoj S. Parihar[1] and P.N. Kondekar[1]

Received: 06 March 2017
Accepted: 12 May 2017
First Published: 20 May 2017

*Corresponding author: Sachin Agrawal, PDPM-Indian Institute of Information Technology, Design and Manufacturing, Jabalpur 482005, India
E-mail: bitssachin.agrawal@gmail.com

Reviewing editor:
Wei Meng, Wuhan University of Technology, China

Additional information is available at the end of the article

Abstract: A novel compact rectifier for dual-band operation in the RF energy harvesting is presented. The circuit comprises a 4th order dual-band impedance matching and a single-series circuit with one double diode, both are integrating into a compact shape to occupy a small area of $30 \times 35\ \text{mm}^2$. The merit of the proposed rectifier circuit is that it can be extended to *n* number of the frequency band by using only $2 \times n$ matching elements. To validate the design method experimentally, a prototype of a dual-band rectifier is fabricated for two public telecommunication bands of GSM-900 and 1800. In order to reduce the circuit complexity and sensitivity arising due to lumped elements, the meander line and the open stub are used to realize the proposed circuit. A good agreement is obtained between the simulation and the measurement. The measured results show that the proposed rectifier circuit exhibits the conversion efficiency of 25.7 and 65% for an input power of −20 and 0 dBm, respectively. In addition, diode nonlinearity which affects the performance of the rectifier in terms of impedance matching is also investigated.

Subjects: Electromagnetics & Microwaves; Electronics; Circuits & Devices

Keywords: RF energy harvesting; dual band impedance matching; rectifier; RF-to-dc-conversion efficiency; frequency transformation

Sachin Agrawal

ABOUT THE AUTHOR

Sachin Agrawal did his BE and ME from Jiwaji University and Birla Institute of Technology Pilani, India in 2005 and 2009, respectively. Currently, he is pursuing his PhD from PDPM-Indian Institute of Information Technology Design and Manufacturing Jabalpur, India. His area of interest are; RF energy harvesting circuit, Antenna including planar and dielectric resonator antenna and Power amplifier.

PUBLIC INTEREST STATEMENT

With rapid growth in wireless communication, a huge amount of radio frequency (RF) energy broadcasted through billions of microwave sources such as mobile phones, handheld radios, and radio broadcast stations. Therefore, it is meaningful to collect and supply it to many electrical devices like mobile headsets, wearable medical sensors through RF energy harvesting. Since the ambient RF energy is distributed in multiple frequency bands, therefore the amount of energy harvested could increase if the circuit is designed for multiple frequency bands. In this work, we present a compact dual-band energy harvesting circuit to harvest energy from two most useful frequency bands, GSM-900 and 1800. The merit of the proposed rectifier circuit is it can be extended to *n* number of the frequency band by using only $2 \times n$ matching elements. A prototype is fabricated, and its performance is evaluated using Vector Network Analyzer (VNA). The total size of the rectifier is about $30 \times 35\ \text{mm}^2$.

© 2017 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

1. Introduction

A revolutionary growth in wireless technology attracts huge attention from research community to make the self-sustainable device feasible through RF energy harvesting. It exploits ambient electromagnetic energy transmitted from different RF systems to remotely feed the electronic devices (Nintanavongsa, Muncuk, Lewis, & Chowdhury, 2012). Compared to other harvesting techniques, RF energy harvesting provides relatively predictable energy supply owing to the features of easy availability and less dependency on environmental variations. The typical block diagram of RF energy harvesting circuit is shown in Figure 1. It consists of three major blocks viz; antenna, matching network (MN), diode detector followed by an energy storage. The first element, antenna is employed to capture the RF signals of different frequencies and polarization, while second MN is for maximum power transfer, and the last rectifier is used to convert the RF energy to dc voltage. It means harvesting circuit performance can be evaluated in terms of accessible ambient RF energy and its conversion rate (Agrawal, Pandey, Singh,& Parihar, 2014). These parameters are heavily influenced by surrounding terrain conditions as the multiple reflection and dissipation certainly deteriorate the level of available ambient RF energy. As a result, conversion efficiency and dc output voltage may degrade. Previously, the majority of available RF energy harvesting circuits focused on single frequency band hence offer low dc output voltage. As the multiple RF energy sources of different frequency bands are available, thus from an ambient RF harvesting perspective, the output dc voltage could be increased if the circuit is designed for multiple frequency bands rather than a single band. A wide-band energy harvester can also promise a high output voltage by accumulating the number of RF signals at a time. However, due to nonlinear behavior of the diode, harvesting circuit itself exhibits nonlinearity i.e. its input impedance varies with the received RF power. Thus, it is quite difficult to retain the impedance match and high conversion efficiency over a large frequency range (Song, Huang, Zhou, & Carter, 2014). The losses due to impedance mismatch over a large bandwidth can be illustrated in Collado and Georgiadis (2013), where only 8% conversion efficiency is achieved at -20 dBm.

To address this, it is preferable to harvest energy from several narrow frequency bands rather than a single large one. In literature, numerous topologies have been proposed to accomplish the multi-band energy harvesting (Bergès, Fadel, Oyhenart, Vigneras, & Taris, 2015; Hamano et al., 2016; Ho et al., 2016; Keyrouz, Visser, & Tijhuis, 2013; Kuhn, Lahuec, Seguin, & Person, 2015; Liu, Zhong, & Guo, 2015; Niotaki, Georgiadis, Collado, & Vardakas, 2014; Pinuela, Mitcheson, & Lucyszyn, 2013; Scheeler, Korhummel, & Popovic, 2014; Shariati, Rowe, Scott, & Ghorban, 2015; Sun, Guo, He, & Zhong, 2013). These topologies can be differentiated in terms of filter functionality i.e. how the antenna or source impedance is matched to the rectifier circuit. For instance, in Pinuela et al. (2013) and Keyrouz et al. (2013) several single-band rectennas (combination of antenna and rectifier circuit) were stacked to constitute a multi-band harvesting circuit. In this case, each rectenna was designed for a specific frequency band. Thus, for compact applications, this architecture is not suitable due to the number of antennas used. Moreover, in most of the reported works, the quality assessment of the output voltages combination was not taken into consideration. In Kuhn et al. (2015), the circuit complexity is reduced to a certain extent by replacing the multiple antennas with a single wide-band antenna. However, in this topology too, the number of rectifiers increases with the frequency bands, which leads to prolonging the circuit complexity.

Figure 1. Typical block diagram of RF energy harvesting circuit.

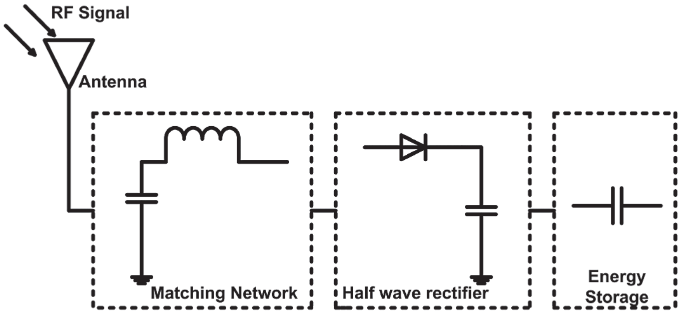

Besides, a multi-band harvesting circuit can also be formed by simply embedding a multi-band matching network between the multi-band antenna and the rectifying circuit (Bergès et al., 2015; Hamano et al., 2016; Ho et al., 2016; Liu et al., 2015; Niotaki et al., 2014; Scheeler et al., 2014; Shariati et al., 2015; Sun et al., 2013). The multi-band matching network can be designed either by distributed or by lumped element. In general, the multi-band rectifier circuit experiences two types of losses: first due to shift in resonance frequency from the optimum frequency point, and second due to the filter complexity. Because of the diode nonlinearity, the input impedance of the circuit varies as a function of power and frequency which causes a shift in resonance frequency. The difficulty due to diode nonlinearity can be observed in Sun et al. (2013) where the dual-band rectifier circuit exhibits its impedance matching for a small range of input power. The losses induced because of filter complexity can be observed in the recently reported works on dual band harvesting circuit (Niotaki et al., 2014; Scheeler et al., 2014; Shariati et al., 2015). In Niotaki et al. (2014), for $P_{\text{in}} = -15$ dBm, author achieved the conversion efficiency of 23% at the expense of increased filter complexity consisting two series and two shunt pairs of reactive elements. Thus, for more than dual band applications, the proposed circuit topology is not suitable due to excessive filtering components used. To obtain good conversion efficiency a dual-band rectenna reported in Scheeler et al. (2014). However, the rectenna was large in size and requiring a complex impedance tuning circuit. In Shariati et al. (2015) also, a dual-band matching network consisting nine reactive elements was employed to achieve the dual-band characteristics.

In order to reduce the filter complexity, this work proposed a compact dual-band harvesting circuit for GSM-900 and 1800. It consists of a 4th order dual-band matching network based on $1-n$ frequency transformation, which is optimized for the energy harvesting circuit to reduce the complexity up to $2 \times n$ reactive elements (n is the number of frequency bands). Similar to frequency transformation method, the proposed dual-band rectifier circuit can be extended to n number of frequency bands by using the $2 \times n$ number of reactive elements. The detailed analysis and design guidelines of dual band rectifier circuit are discussed in Section 2.

2. Dual band rectifier design and analysis

This section presents the design and analysis of a dual-band harvesting circuit in terms of impedance matching, DC output voltage and RF-to-dc conversion efficiency. The topology of the proposed dual-band RF energy harvesting circuit is shown in Figure 2(a). As seen, the low-cost Schottky diode is used to transform the input RF power to DC voltage. The impedance matching at two frequency is achieved using a series and parallel combination of the *LC* pair. The main idea underlying the suggested multi-band matching network is $1-n$ frequency transformation (one to many mapping of frequency), which transforms a single-band matching network to multi-band matching network (Nallam & Chatterjee, 2013). As the name ($1-n$) suggests that for designing a multi-band matching network, primarily a single-band matching network is required whose resonant frequency is dependent on the frequencies for which multi-band matching network proposed to designed.

Moreover, this frequency transformation method depends on the type of load impedance, whether it is series or parallel combination of *RC* or *RL*. Since the selected diode (HSMS-2852) has capacitive behavior throughout the frequency, it can be represented in a series or parallel combination of *R* and *C*. In the case of parallel *RC* load, the following equations are used to transform the single-band matching network into multi-band matching network.

$$\omega = \frac{\omega_t^n + a_m \omega_t^{n-m} + a_{m+2}\omega_t^{n-(m+2)} + \cdots}{\omega_t^{n-1} + a_{m+1}\omega_t^{n-(m+1)} + a_{m+3}\omega_t^{n-(m+3)} + \cdots} \quad (1)$$

where, n is the number of bands and m varies from 2 to n. After substituting the value of n, Equation (1) can be expanded in partial fraction form using the causal foster analysis as:

$$\omega = \omega_t + \frac{1}{\frac{\omega_t}{a_2 - a_3} + \frac{1}{\cdots}} \quad (2)$$

Figure 2. (a) Circuit diagram of the proposed dual-band rectifier circuit and (b) optimized dual-band rectifier circuit.

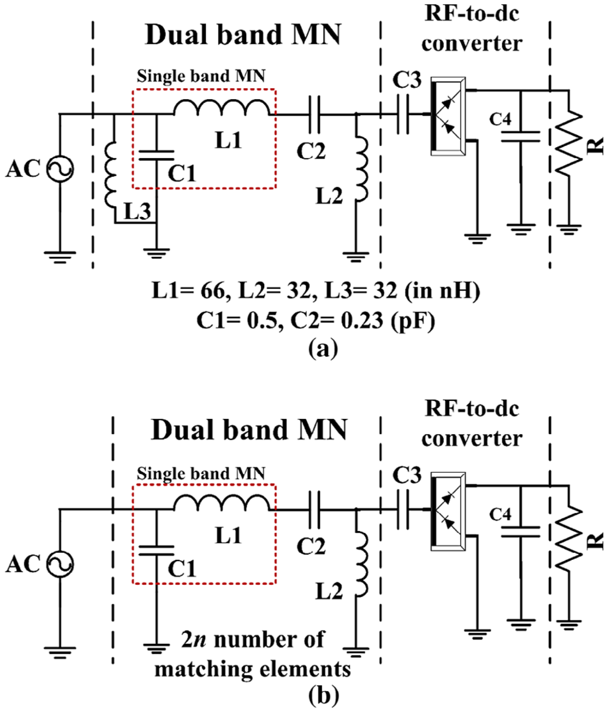
Dual band MN
RF-to-dc converter
Single band MN
C3
C4
L1
C2
AC
C1
R
L3
L2
L1= 66, L2= 32, L3= 32 (in nH)
C1= 0.5, C2= 0.23 (pF)
(a)
Dual band MN
RF-to-dc converter
Single band MN
C3
C4
L1
C2
AC
C1
R
L2
2*n* number of matching elements
(b)

The coefficients a_2 and a_3 can be calculated as:

$$\omega_m = \sum_{i=1}^{n} (-1)^{i-1} \omega_i \tag{3}$$

$$a_m = (-1)^m \sum_{i,j=1,1\,\&\,i\neq j}^{n,n} (-1)^{i+j} \omega_i \omega_j \tag{4}$$

$$a_{m+1} = (-1)^{m+1} \sum_{i,j,k=1,1,1\,\&\,i\neq j\neq k}^{n,n,n} (-1)^{i+j+k} \omega_i \omega_j \omega_k \tag{5}$$

$$a_{m+n} = (-1)^{m+n} \sum_{i,j,k\ldots=1,1,1\ldots i\neq j\neq k}^{n,n,n\ldots} (-1)^{i+j+k+\cdots} \omega_i \omega_j \omega_k \ldots \tag{6}$$

$$a_n = \prod_{i=1}^{n} (-1)^n \omega_i \tag{7}$$

Equation (6) is similar to that presented in Nallam and Chatterjee (2013), except the term $(-1)^{m+n}$, which is included here to realize the multi-band matching network for more than three frequency bands i.e. for $n \geq 3$.

With this transformation, the capacitor of the matching network is transformed to the combination of prototype capacitor parallel with inductor whereas, an inductor is transformed into a combination of the same inductor with a series capacitor. Figure 3 shows the circuit schematic of transformation of a single-band matching network to the dual-band matching network. It can be seen that C_1 is transformed to $C_1 \| L_2$ and L_1 transformed to L_1 series with C_2. After successful usage of (1)–(7), the resultant multi-band matching network requires $3n-1$ and $4n-1$ reactive elements for L and Π-type topology, respectively.

Figure 3. Conversion of a single-band matching network to dual-band matching network using $1-n$ frequency transformation.

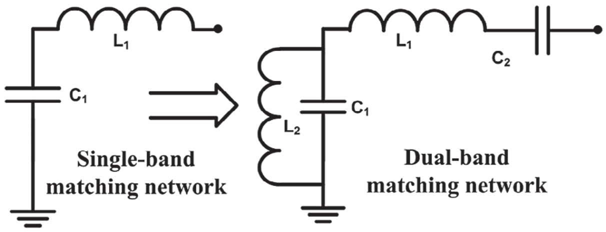

As the aim is to design a dual-band harvesting circuit, therefore, we require here only 5 or 7 reactive elements with L and Π-type topologies, respectively. From Figure 2(a), it can be seen that the resultant matching network consists of five elements, where the inductor L_3 and capacitor C_2 results after the transformation of capacitor C_1 and inductor L_1, respectively. Besides, the inductor L_2 occurrs due to the diode reactive element, which is generally a capacitor.

In this work, two frequencies 0.9 and 1.8 GHz that correspond to the maximum signal strength are chosen for dual-band harvesting circuit. According to this method, it is necessary to assign the frequencies in descending order e.g. $\omega_1 = 1.8, \omega_2 = 0.9$. Therefore, from (3) single-band matching network frequency is equal to $\omega_1 - \omega_2 = 2\pi(1.8 - 0.9) \times 10^9 = 0.9 \times 2\pi \times 10^9$. In order to match the source impedance with the rectifier at the calculated frequency 0.9 GHz, the chosen matching topology is L-type as shown by the encircled portion in Figure 2(a). The corresponding element values can be approximated using the various methods some of which are described in Pozar (2010). Subsequently, this single-band matching network is transformed to dual-band matching network using (1)–(7). The detailed design steps of the dual-band rectifier circuit are summarized as follows:

(1) As we are interested in matching the diode to 50 Ω at two frequencies (0.9 and 1.8 GHz) so, the order of transformation is equal to 2 or $n = 2$.

(2) In the first step, single-band matching network is designed at the frequency f calculated as: $f = f_2 - f_1 = 1.8 - 0.9 = 0.9$ GHz. In this case, any matching topology that matches the diode to 50 Ω, at 0.9 GHz for an input power $P_{\text{in}} = -20$ dBm, and load resistance 4.7 kΩ can be used. The chosen single-band matching network is shown by the encircled portion in Figure 2(a).

(3) Afterwards, this single-band matching network is transformed into dual-band using the (1)–(7) as shown below:

Since $n = 2$ therefore, from (2)

$$\omega = \omega_t + \frac{1}{\frac{\omega_t}{a_2}} \tag{8}$$

From (4) a_2 can be calculated as:

$$a_2 = (-1)^{2+2} \sum (-1)^{1+2} \omega_1 \omega_2 = \omega_1 \omega_2 \tag{9}$$

$$a_2 = -1.62 \times 4\pi^2 \times 10^8 = 0.64 \times 10^{20} \tag{10}$$

Thus, inductor L_1 (=66 nH) is transformed to impedance as:

$$j66 \times 10^9 \omega = j66 \times 10^9 \omega_t + \frac{1}{j0.23 \times 10^{-12} \omega_t} \tag{11}$$

Similarly, capacitors (=0.5 pF) are transformed to the admittance as:

$$j5 \times 10^{-13} \omega = j5 \times 10^{-13} \omega_t + \frac{1}{j32 \times 10^{-9} \omega_t} \tag{12}$$

The circuit schematic of the dual-band harvesting circuit is shown in Figure 2(a). It can be seen that resultant matching network consists of five reactive elements according to $3n - 1$. In order to reduce the circuit complexity and sensitivity due to reactive elements, a parametric study has been carried out to eliminate the elements showing minimum influence on the circuit performance.

Figure 4 shows the simulated $|S_{11}|$ for the different combination of matching elements. The simulated results demonstrate that $|S_{11}|$ experiences maximum change when inductor L_2 and capacitor C_2 are removed from the circuit, whereas it remains almost unaffected when L_3 is not present in the circuit. Therefore, inductor L_3 can be extruded from the circuit and the resultant matching circuit requires only $2n$ and $3n$ reactive elements in place of $3n - 1$ and $4n - 1$ elements. In this way, for each topology, the proposed circuit reduces $n - 1$ elements compared to the conventional method. Figure 2(b) demonstrates the optimized circuit diagram of the dual-band rectifier. It can be observed that circuit requires large inductors value of 32 and 66 nH. Thus, it is quite difficult to realize the practical rectifier circuit whose response is similar to the response of simulated result. In order to avoid any impedance mismatch due to the small difference in elements value, the meander line inductor and open stub are used to realize the inductors and capacitors, respectively. In this case, not only fabrication and optimization process become so easy but the cost will also reduced.

Figure 5, shows the layout of the dual-band rectifier circuit. In Nintanavongsa et al. (2012) and Agrawal et al. (2014), it has been demonstrated that the number of rectifying diodes or equivalently voltage multiplier stages are very much sensitive to the RF-to-dc conversion efficiency. In low-power region (≤ -20 dBm), efficiency decreases if voltage multiplier stages increase, whereas in higher power region (≥ -20 dBm), an opposite effect occurs. As the demand is to harvest energy in low-power region, single-series circuit with a double diode is used to convert received RF energy into dc voltage. From the left side of the circuit, the first meander line corresponds to the inductor L_1, while the second meander line represents the inductor L_2, of the Figure 2(b). The shunt stub is accounted

Figure 4. Simulated $|S_{11}|$ vs. frequency for different combination of circuit elements.

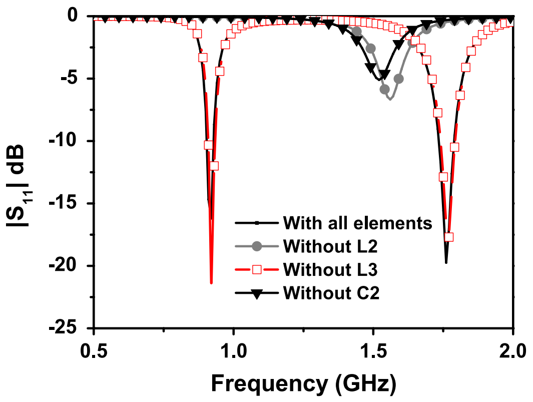

Figure 5. Layout of the proposed dual-band RF energy harvesting circuit.

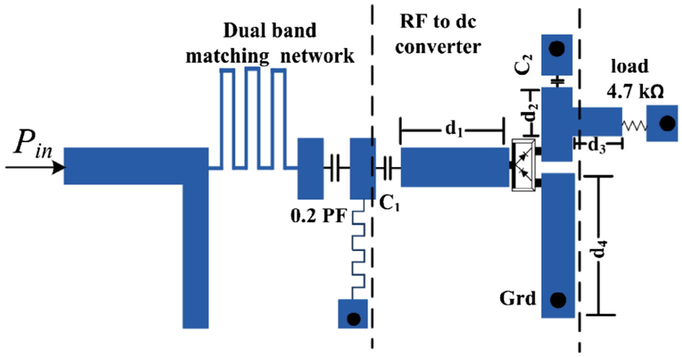

for the shunt capacitor C_1 of Figure 2(b). The dimensions of each element are calculated according to their respective reactive element value and the substrate on which circuit has to be fabricated. In Assimonis, Daskalakis, and Bletsas (2016), it has been demonstrated that traces (microstrips) connected to the rectifier terminals (e.g. distance between the diode and via and diode and load) are highly sensitive for RF-to-dc efficiency. Therefore, traces d_1, d_2 between diode and capacitor C_1 and C_2, d_3 between diode and load and d_4 between diode and ground are adjusted to optimize the impedance matching as well as the conversion efficiency of the rectifier. Due to nonlinear behavior of the diode, harvesting circuit itself exhibits nonlinearity i.e. its input impedance varies with received RF power, therefore harmonic-balance (HB) and large signal analysis (LSSP) were employed to take into consideration the nonlinear behavior of the rectifier.

The photograph of the fabricated dual-band rectifier is shown in Figure 6. It is fabricated on a 1.54 mm thick FR-4 substrate with a dielectric constant (ϵ_r) of 4.3 using chemical etching method. The rectifier performance is evaluated in terms of $|S_{11}|$ and output voltage using the Agilent vector network analyzer (VNA). The simulated and measured $|S_{11}|$ is illustrated in Figure 7(a). The measured result shows reasonable agreement with the simulated one; the slight difference can be accounted for the fabrication imperfections. It is well known that impedance matching is a function of frequency and input power, due to the nonlinearity of the diode. Such a characteristic is examined in Figure 7(b), where the measured $|S_{11}|$ is demonstrated as a function of input power level for three different load impedance values. From results, it is clear that impedance matching of the harvesting circuit is greatly affected by the input power and the load impedance. Figure 7(b) demonstrates that as power increases, the impedance matching at 0.9 GHz degraded drastically, while at 1.8 GHz, it improves. Moreover, it is noticed that the impedance matching at higher power level is more sensitive to the variation of load impedance (*RL*).

The measured RF-to-dc conversion efficiency and output voltage vs. input power for both frequencies are demonstrated in Figure 8(a). For 0.9 GHz, efficiency is equal to 25.7 and 65.1% for an input power of -20 and 0 dBm, respectively. However, at 1.8 GHz the efficiency is relatively small that might be due to the increased parasitic losses in the rectifier diode. Figure 8(b) shows the relation between the output voltage and frequency for various input power levels at fixed load resistance value of 4.7 kΩ. It can be seen that maximum output voltage is achieved in the frequency range of 860–900 and 1770–1800 MHz, showing the rectifier's capability to harvest RF energy in the GSM-900 and 1800 bands.

Figure 9, depicts the measured conversion for various values of load resistance. It can be noticed that the circuit yields maximum efficiency when the load impedance is 4.7 kΩ. It starts decreasing as the load impedance varies from 4.7 kΩ. Table 1 shows the comparison of the conversion efficiency and the size of the proposed rectifier with the similar works reported previously. Only measured results are compared in Table 1. It can be seen that in low-power condition maximum efficiency is achieved in Sun et al. (2013), but the expense of bulky circuit size. However, the proposed rectifier offers an optimal conversion efficiency with compact circuit size.

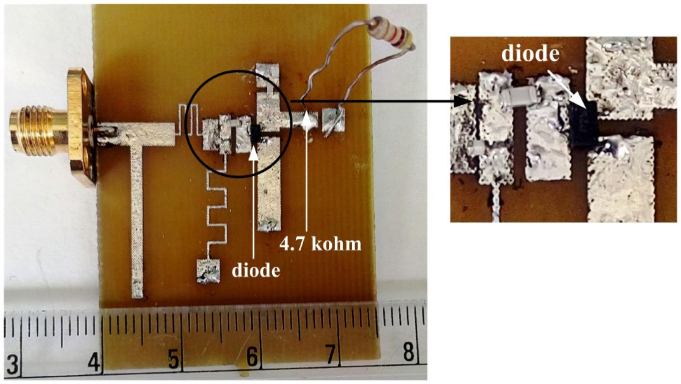

Figure 6. Photograph of the fabricated dual-band RF energy harvesting circuit.

Figure 7. (a) Simulated and measured $|S_{11}|$ vs. frequency and (b) measured $|S_{11}|$ for various power levels.

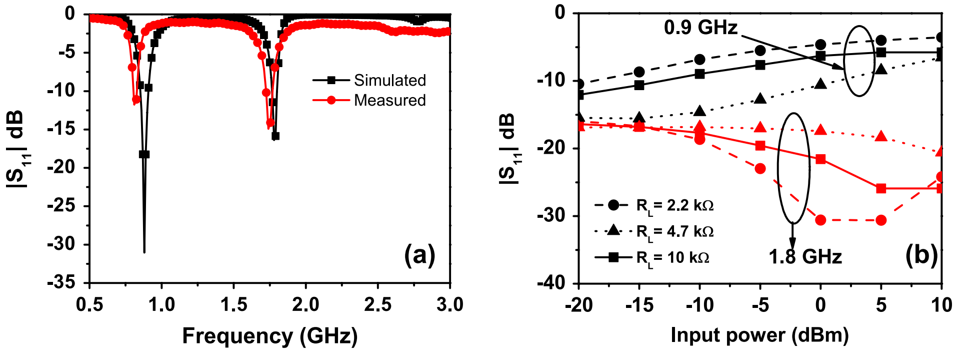

Figure 8. (a) Measured output dc voltage and RF-to-dc conversion efficiency and (b) measured dc voltage vs. frequency.

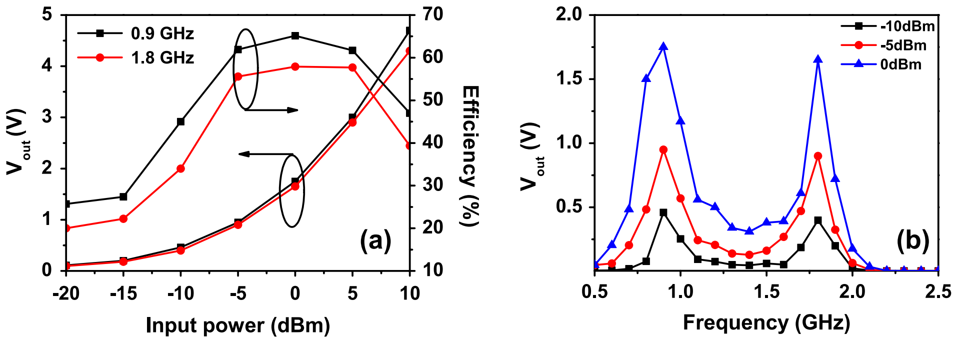

Figure 9. Measured RF-to-dc conversion efficiency for different load impedance.

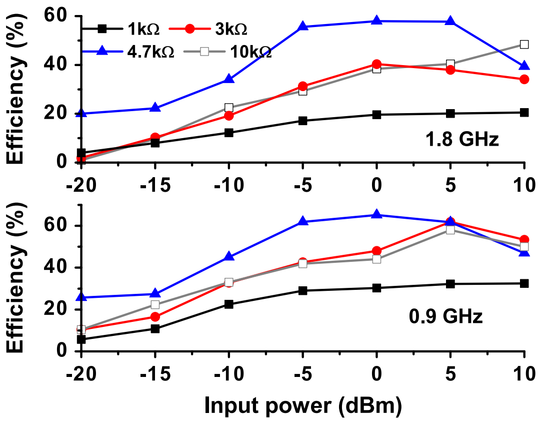

Table 1. Performance comparison of the proposed dual rectifier with recently published works

Ref.	Measured rectifier efficiency (%)	Input power (dBm)	Rectifier size
Ho et al. (2016)	15.8 @ 0.89 GHz	−20	100 × 65 mm^2
	11.2 @ 1.76 GHz		
Hamano et al. (2016)	10 @ 2.15 GHz	−10	37 × 71 mm^2
	15 @ 5.84 GHz		
Bergès et al. (2015)	27 @ 0.91/2.4 GHz	−16	78 × 88 mm^2
Sun et al. (2013)	30 @ 2.14 GHz	−20	145 mm
	35 @ 1.84 GHz		
Liu et al. (15)	20 @ (0.91+1.8) GHz	−20	23 × 37 mm^2
This work	27.5 @ 0.9 GHz	−20	30 × 35 mm^2
	20 @ 1.8 GHz		

3. Conclusion

A new compact 4th order dual-band rectifier has been designed to harvest the RF power of GSM-900 and 1800 bands. In order to reduce the circuit complexity and sensitivity due to reactive elements, the meander line and the open stub are used to fabricate the matching network. For $P_{\text{in}} = -20$ dBm, the measured RF-to-dc conversion efficiency of 27.5 and 20% is achieved at 0.9 and 1.8 GHz, respectively. Further, more than 45 and 34% conversion efficiency is maintained from -10 to 10 dBm for 0.9 and 1.8 GHz, respectively.

Funding
The authors received no direct funding for this research.

Author details
Sachin Agrawal[1]
E-mail: bitssachin.agrawal@gmail.com
Manoj S. Parihar[1]
E-mail: mparihar@iiitdmj.ac.in
P.N. Kondekar[1]
E-mail: pnkondekar@iiitdmj.ac.in
[1] PDPM-Indian Institute of Information Technology, Design and Manufacturing, Jabalpur 482005, India.

Citation information
Cite this article as: A dual-band RF energy harvesting circuit using 4th order dual-band matching network, Sachin Agrawal, Manoj S. Parihar & P.N. Kondekar, *Cogent Engineering* (2017), 4: 1332705.

References

Agrawal, S., Pandey, S. K., Singh, J., & Parihar, M. S. (2014). *Realization of efficient RF energy harvesting circuits employing different matching technique* (pp. 754–761). Fifteenth International Symposium on Quality Electronic Design, Santa Clara, CA.

Assimonis, S. D., Daskalakis, S. N., & Bletsas, A. (2016). Sensitive and efficient rf harvesting supply for batteryless backscatter sensor networks. *IEEE Transactions on Microwave Theory and Techniques, 64*, 1327–1338.

Bergès, R., Fadel, L., Oyhenart, L., Vigneras, V., & Taris, T. (2015). A dual band 915MHz/2.44 GHz RF energy harvester. *Microwave Conference (EuMC), 2015 European*. IEEE.

Collado, A., & Georgiadis, A. (2013). Conformal hybrid solar and electromagnetic (EM) energy harvesting rectenna. *IEEE Transactions on Circuits and Systems I: Regular Papers, 60*, p. 10.

Hamano, K., Tanaka, R., Yoshida, S., Miyachi, A., Nishikawa, K., & Kawasaki, S. (2016). Design of dual-band rectifier using microstrip spurline notch filter. *Radio-Frequency Integration Technology (RFIT), 2016 IEEE International Symposium on*. IEEE.

Ho, D. K., Kharrat, I., Ngo, V. D., Vuong, T. P., Nguyen, Q. C., & Le, M. T. (2016). *Dual-band rectenna for ambient RF energy harvesting at GSM 900 MHz and 1800 MHz* (pp. 306–310). 2016 IEEE International Conference on Sustainable Energy Technologies (ICSET), Hanoi.

Keyrouz, S., Visser, H., & Tijhuis, A. (2013, April). Multi-band simultaneous radio frequency energy harvesting. *In Antennas and Propagation (EuCAP), 2013 7th European Conference on*, 3058–3061.

Kuhn, V., Lahuec, C., Seguin, F., & Person, C. (2015). A multi-band stacked RF energy harvester with RF-to-DC efficiency up to 84\%. *IEEE Transactions on Microwave Theory and Techniques, 63*, 1768–1778.

Liu, Z., Zhong, Z., & Guo, Y.-X. (2015). Enhanced dual-band ambient RF energy harvesting with ultra-wide power range. *IEEE Microwave and Wireless Components Letters, 25*, 630–632.

Nallam, N., & Chatterjee, S. (2013, June). Multi-band frequency transformations, matching networks and amplifiers. *Circuits and Systems I: Regular Papers, IEEE Transactions on, 60*, 1635–1647.

Nintanavongsa, P., Muncuk, U., Lewis, D., & Chowdhury, K. (2012, March). Design optimization and implementation for rf energy harvesting circuits. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2*, 24–33.

Niotaki, K., Georgiadis, A., Collado, A., & Vardakas, J. S. (2014). Dual-band resistance compression networks for improved rectifier performance. *IEEE Transactions on Microwave Theory and Techniques, 62*, 3512–3521.

Pinuela, M., Mitcheson, P., & Lucyszyn, S. (2013, July). Ambient rf energy harvesting in urban and semi-urban environments. *Microwave Theory and Techniques, IEEE Transactions on, 61*, 2715–2726.

Pozar, D. M. (2010). *Microwave engineering* (3rd ed.). New Delhi: Wiley.

Scheeler, R., Korhummel, S., & Popovic, Z. (2014). A dual-frequency ultralowpower efficient 0.5-g rectenna. *IEEE Microwave Magazine, 15*, 109–114.

Shariati, N., Rowe, W. S., Scott, J. R., & Ghorban, K. (2015). Multi-service highly sensitive rectifier for enhanced RF energy scavenging. *Scientific Reports, 5*, 9655

Song, C., Huang, Y., Zhou, J., & Carter, P. (2014). Improved ultrawideband rectennas using hybrid resistance compression technique. *IEEE Transactions on Antennas and Propagation, 65*, 2057–2062.

Sun, H., Guo, Y. X., He, M., & Zhong, Z. (2013). A dual-band rectenna using broadband Yagi antenna array for ambient RF power harvesting. *IEEE Antennas and Wireless Propagation Letters, 12*, 918–921. doi:10.1109/LAWP.2013.2272873.

© 2017 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:
Share — copy and redistribute the material in any medium or format
Adapt — remix, transform, and build upon the material for any purpose, even commercially.
The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:
Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.
You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
No additional restrictions
You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

***Cogent Engineering* (ISSN: 2331-1916) is published by Cogent OA, part of Taylor & Francis Group.**

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com

US007667995B1

(12) United States Patent
Leuenberger et al.

(10) **Patent No.: US 7,667,995 B1**
(45) **Date of Patent: Feb. 23, 2010**

(54) **TELEPORTATION SYSTEM FOR ELECTRONIC MANY-QUBIT STATES USING INDIVIDUAL PHOTONS**

(75) Inventors: **Michael N. Leuenberger**, Oviedo, FL (US); **Michael E. Flatté**, Iowa City, IA (US); **David D. Awschalom**, Santa Barbara, CA (US)

(73) Assignees: **University of Iowa Research Foundation**, Iowa City, IA (US); **The Regents of the University of California**, Oakland, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 346 days.

(21) Appl. No.: **11/352,387**

(22) Filed: **Feb. 9, 2006**

Related U.S. Application Data

(60) Provisional application No. 60/651,151, filed on Feb. 9, 2005.

(51) **Int. Cl.**
G11C 13/00 (2006.01)

(52) **U.S. Cl.** **365/106**; 977/708; 977/933; 398/40; 398/158

(58) **Field of Classification Search** 365/106, 365/200; 257/13, 31, 97, 202, E29.071, E29.168, 257/E49.003; 398/140, 175, 40; 977/708, 977/933; 250/214.1, 251, 493.1; 324/207.2, 324/248, 307; 359/107, 108, 577; 380/201, 380/256, 263
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2004/0042715	A1*	3/2004	Linden et al.	385/31
2004/0071019	A1*	4/2004	Magnus et al.	365/200
2004/0264958	A1*	12/2004	Zoller et al.	398/40
2005/0131746	A1*	6/2005	Beausoleil et al.	705/7

OTHER PUBLICATIONS

Armani et al., "Ultra-high-Q toroid microcavity o a chip," Nature 421: p. 925-928, 2003.

Auston and Nuss, "Electrooptic Generation and Detection of Femtosecond Electrical Transients," *IEEE J. Quant. Electron.* 24: p. 184-197, 1988.

Barrett et al., "Deterministic quantum teleportation of atomic qubits," *Nature* 429: p. 737-739, 2004.

Becker et al., "Picosecond Fluorescence Lifetime Microscopy by TCSPC Imaging" Proc. SPIE 4262, 414-419 (2001).

Bennett et al., "Quantum information and computation," *Nature* 404: p. 247-255, 2000.

Bennett et al., "Teleporting an Unknown Quantum State via Dual classical and Einstein-Posolsky-Rosen Channels," *Phys. Rev. Lett.* 70: p. 1895-1899, 1993.

Bouwmeester et al., "Experimental quantum teleportation," *Nature* 390: p. 575-579, 1997.

(Continued)

Primary Examiner—Son Dinh
Assistant Examiner—Harry W Byrne
(74) *Attorney, Agent, or Firm*—Ballard Spahr LLP

(57) **ABSTRACT**

A method for creating a logic state for teleporting quantum information using a single photon is described. The method includes receiving a photon with an initial polarization and causing a first semiconductor crystal to have a first spin orientation. The photon interacts with the first semiconductor crystal for producing a resulting polarization dependent upon the first spin orientation. Causing the photon to interact with the first semiconductor crystal generates a maximally entangled state.

6 Claims, 3 Drawing Sheets

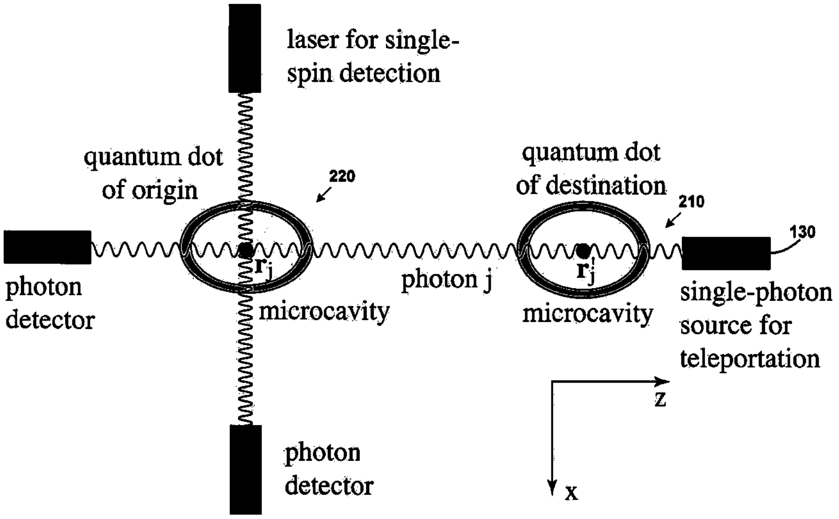

OTHER PUBLICATIONS

Braunstein et al., "Maximal violation of Bell Inequalities for Mixed States," *Phys. Rev. Lett.* 68: p. 3259-3261, 1992.

Einstein et al., "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" *Phys. Rev.* 47: p. 777-780, 1935.

Fang et al., "Probabilistic teleportation of a three-particle state via three pairs of entangled particles," *Phys. Rev.* A 67: p. 014305-1-014305-4, 2003.

Friesen et al., "Spin Readout and Initialization in a Semiconductor Quantum Dot," *Phys. Rev. Lett.* vol. 92, No. 3: p. 037901-1-037901-4, 2004.

Greenberger et al., "Going Beyond Bell's Theorem" Bell's Theorem, Quantum Theory, and Conceptions of the Universe, 69-72 (Kluwer Academic, Dordrecht, 1989).

Guest et al., Measurement of optical absorption by a single quantum dot exciton, Phys. Rev. B, 65: p. 241310-1-241310-4, 2002.

Gurioll et al., Non-linear resonant Rayleigh scattering from microcavity, Physica E 17: p. 463-464, 2003.

Jennewein et al., "Experimental Nonlocality Proof of Quantum Teleportation and Entanglement Swapping," *Phys. Rev. Lett.* 88: p. 017903-1-017903-4, 2001.

Kikkwawa and Awschalom, "Resonant Spin Amplification in n-Type GaAs," Phys. Rev. Lett. 80: p. 4313-4316, 1998.

Kroutvar et al., "Optical programmable electron spin memory using semiconductor quantum dots," *Nature* 432: p. 81-84, 2004.

Lee, J., "Multipartite entanglement for entanglement teleportation," *Phys. Rev.* A 66: p. 052318-1-052318-5, 2002.

Loss and DiVincenzo, "Quantum computation with quantum dots," *Phys. Rev. A* 57: p. 120-126, 1998.

Marcikic et al., "Long-distance teleportation of qubits at telecommunication wavelengths," *Nature* 421: p. 509-513, 2003.

Michler et al., "Laser emission from quantum dots in microdisk structures," *Appl. Phys. Lett.* 77: p. 184-186, 2000.

Qi et al., "A three-dimensional optical photonic crystal with designed point defects," *Nature* 429: 538-542, 2004.

Rahachou et al., "Effects of boundary roughness on a Q factor of whispering-gallery-mode Ising microdisk cavities," *Appl. Phys.* 94: p. 7929-7931, 2003.

Riebe et al., "Deterministic quantum teleportation with atoms," *Nature* 429: p. 734-737, 2004.

Roa et al., "Optimal conclusive teleportation of quantum," *Phys. Rev.* A 68: p. 022310-1-022310-6, 2003.

Srinivasan et al., "Experimental demonstration of a high quality factor photonic crystal microcavity," *Appl. Phys. Lett.* 83: p. 1915-1917, 2003.

Wootters and Zurek, "A single quantum cannot be cloned," *Nature* 299: p. 802-803, 1982.

* cited by examiner

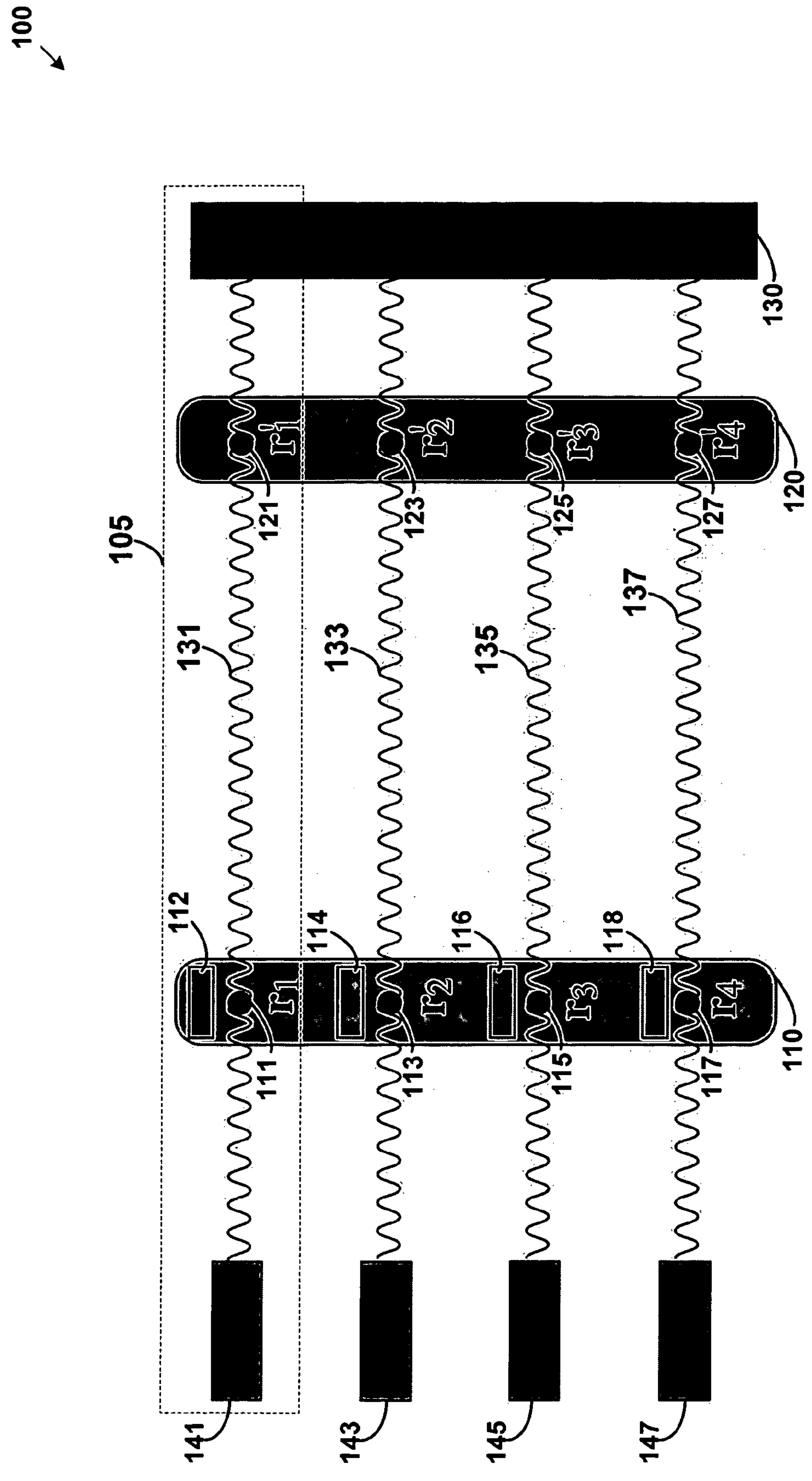
r'_1
r'_2
r'_3
r'_4
r_1
r_2
r_3
r_4

FIG. 1

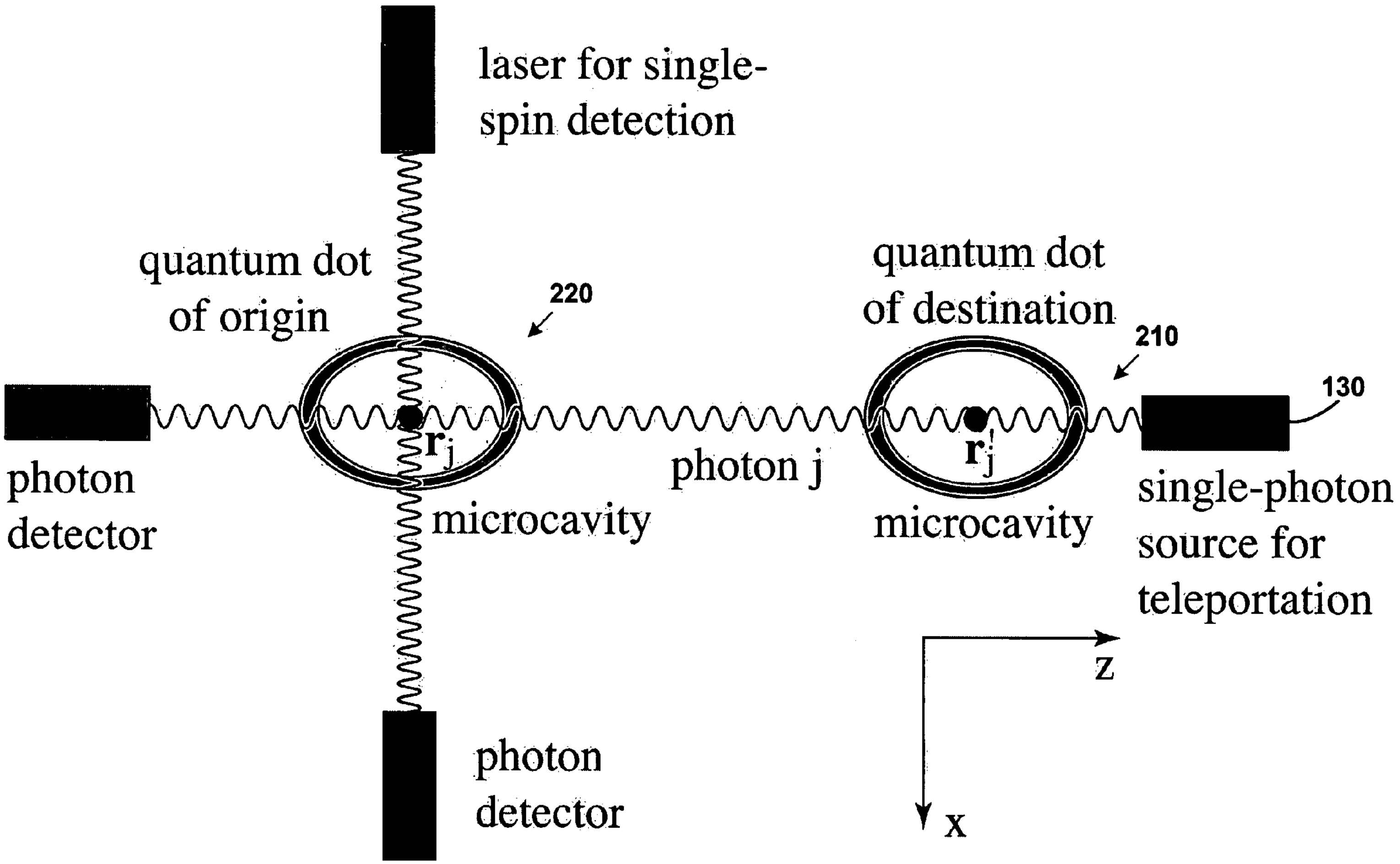
laser for single-spin detection
quantum dot of origin
220
photon detector
r_j
microcavity
photon detector
photon j
quantum dot of destination
210
r_j'
microcavity
130
single-photon source for teleportation
z
x

FIG. 2

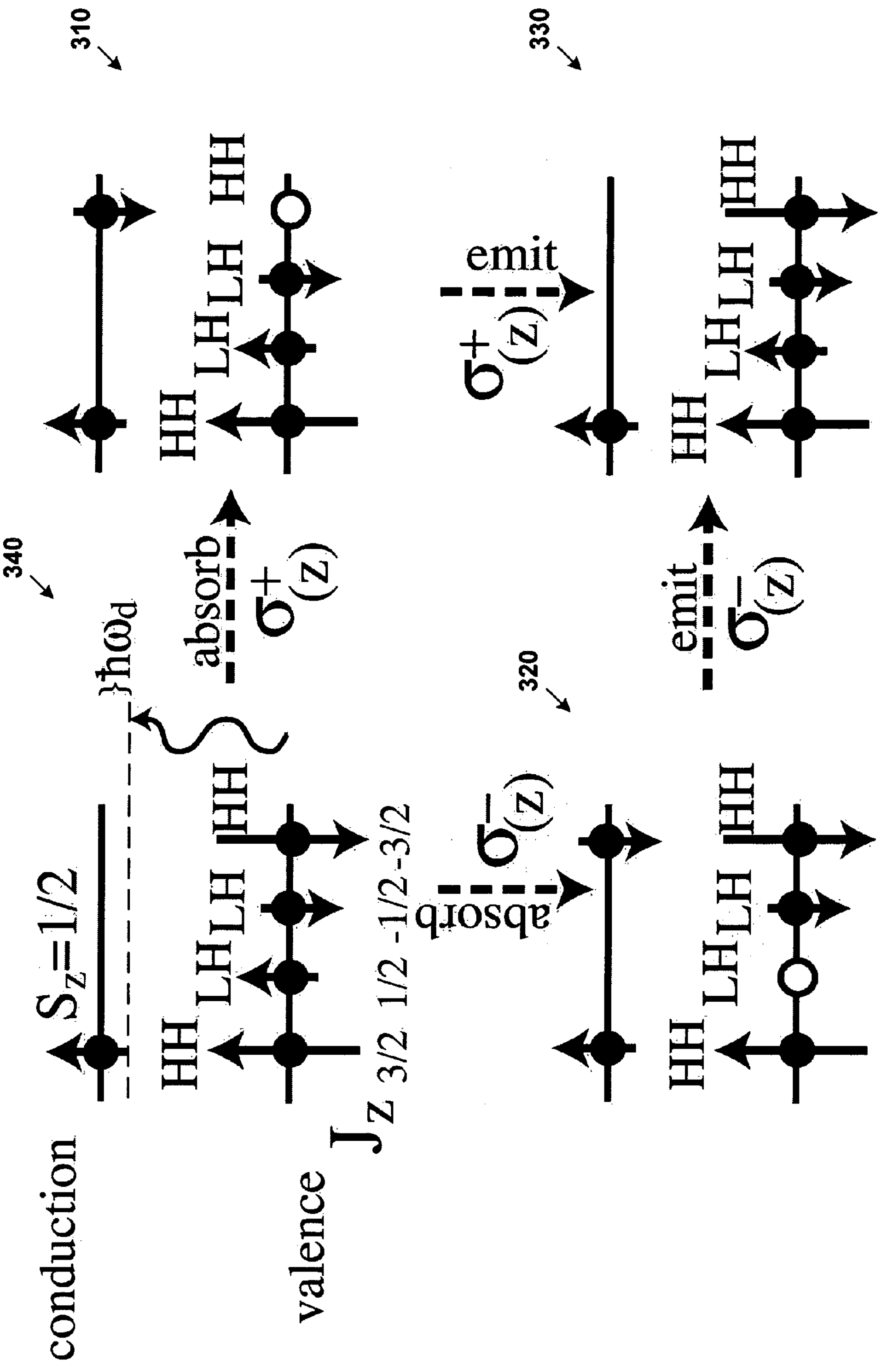
310
340
320
330
conduction
valence
S_z=1/2
HH
LH
LH
HH
J_z 3/2 1/2 -1/2 -3/2
ħω_d
absorb
σ+(z)
absorb
σ−(z)
emit
σ+(z)
emit
σ−(z)

FIG. 3

TELEPORTATION SYSTEM FOR ELECTRONIC MANY-QUBIT STATES USING INDIVIDUAL PHOTONS

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims priority to a U.S. Provisional Patent Application with application No. 60/651,151 entitled "TELEPORTATION OF ELECTRONIC MANY-QUBIT STATES VIA SINGLE PHOTONS," which was filed on Feb. 9, 2005. This application is hereby incorporated by reference in its entirety.

DESCRIPTION OF RELATED ART

Transport, or transportation, is the movement of people, goods, signals and information. The field of transport has several aspects, which include infrastructure, vehicles, and operations. Infrastructure includes transport networks (e.g., roads, railways, airways, canals, and pipelines, etc.), as well as the nodes, or terminals, (e.g., airports, railway stations, bus stations and seaports). The vehicles (e.g., automobiles, bicycles, buses, trains, airplanes) generally ride on the networks. Operations involve the control of the system (e.g., traffic signals, ramp meters, railroad switches, and air traffic control).

At the level of a particle (e.g., electron, photon, atom, etc.), there are two approaches to transporting that particle from one place to another. The first is very similar to the approaches described above, in which that particle is phyically moved from one place to another. The second, called teleportation, is when the complete quantum-mechanical state of one particle (an origin particle) is imprinted onto a second (a destination particle) at a different location. The two particles involved do not move themselves, so teleportation does not use the same infrastructure, vehicles, and operations used in transportation. The complete quantum mechanical state (often referred to as the wave function of the particle) fully describes the physical properties of the particle. Usually in order to perform teleportation the destination particle is first prepared in a known quantum mechanical state. Because complete quantum-mechanical states cannot be fully measured, it is important that this process be possible without knowledge of the actual quantum mechanical state the origin particle is in. For the same reason once teleportation is complete the origin particle is left in a state unrelated to its initial (before teleportation) quantum mechanical state. The process is accompanied by the sending of some classical information from origin to destination. This information is insufficient to fully create the quantum mechanical state, but is sufficient to correct potential errors that occur during the teleportation process. Amidst the growing technological evolution, there remains a continued need in the development of technological devices that use teleportation.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a block diagram of teleportation system for electronic many-qubit states using individual photons.

FIG. **2** is a block diagram illustrating the teleportation of a spin state from the quantum dot systems of FIG. **1**.

FIG. **3** is an diagram of the electronic states of the quantum dot illustrating possible virtual processes excited by the photon j of FIG. **1**.

While the invention is susceptible to various modifications and alternative forms, specific embodiments have been shown by way of example in the drawings and subsequently are described in detail. It should be understood, however, that the description herein of specific embodiments is not intended to limit the invention to the particular forms disclosed. In contrast, the intention is to cover all modifications, equivalents, and alternatives falling within the spirit and scope of the invention as defined by the appended claims.

DETAILED DESCRIPTION OF EMBODIMENTS

Turning now to FIG. **1**, this figure is a block diagram of teleportation system for electronic many-qubit states using individual photons **100**. An origin quantum dot system **110** includes four quantum dots numbered **111**, **113**, **115**, and **117** and respectively labeled r_1, r_2, r_3, and r_4. Each quantum dot contains a single electron in the lowest-energy conduction state, and the state of that quantum dot is fully described by the orientation of the spin of that electron. It is those electrons that can be teleported. With the teleportation system **100**, information associated with the quantum dots **111**-**117** in the quantum dot system **110** can be transported to a destination quantum dot system **120**. When the information associated with the quantum dots **111**-**117** is teleported, information is now associated with quantum dots **121**-**127**, respectively labeled r_1', r_2', r_3', and r_4' result. If there are four quantum dots in the quantum dot system **110**, then there are four quantum dots in quantum dot system **120**, such that each quantum dot in the quantum dot system **120** was teleported from the quantum dot system **110**. Alternatively, the number of quantum dot system **110** can be 3, 7, 20, or some other suitable number. Each quantum dot r_j (e.g., r_3 or quantum dot **113**) within the quantum dot system **110** represents a single qubit. Thus, this quantum dot system is a multi-qubit system with many qubit states (the state of the electron in each quantum dot); each qubit state is represented by a different qubit.

The origin quantum dot system **110** also includes four single spin detectors numbered **112**, **114**, **116**, and **118** associated with each quantum dot. As the number of quantum dots in the system **110** varies, the number of detectors can also vary so that a one to one ratio remains for the number of quantum dots to the number of detectors. A single spin detector is a device that can record a spin, or intrinsic angular momentum of the conduction electron in a particular quantum dot along a particular axis (in this case perpendicular to the photon propagation direction in FIG. **1**. For example, the single spin detector can be a single-electron transistor (SET), or the nucleation of ferromagnetic materials can be used as a mechanism for the single spin detector, or the transport across the quantum dot measured by two leads in a magnetic field can be used for the single spin detector, or the interaction of a photon (as described below) with the spin, and the measurement of the photon polarization, can be used for the single-spin detector. For example, the quantum detector **112** can record a spin associated with the quantum dot **111**. Similarly, the quantum detector **116** can record a spin associated with the quantum dot **115**. Since there is a spin detector associated with each quantum dot in the quantum dot system **110**, the spin of the lowest conduction electron for each quantum dot can be individually determined along the axis perpendicular to the photon propagation in FIG. **1**. As the spin of the lowest conduction electron for each quantum dot fully determines the quantum-mechanical state of that quantum dot, measurement of the spin along one axis provides a small amount of information about the state of that quantum dot. This information can be used to correct potential errors in imprinting the origin quantum state onto the destination in the teleportation process.

As clearly illustrated in FIG. **1**, each quantum dot in the quantum dot system **110** also has an associated photon. A photon is a quantum of electromagnetic energy, regarded as a discrete particle having zero mass, no electric charge, and an indefinitely long lifetime. A single-photon source **130** can produce these photons. Examples of the single photon source can include a highly attenuated laser beam, or a laser beam passing through an extremely small aperture that permits one photon to pass per laser pulse, or some other suitable photon source. Photons **131**, **133**, **135**, and **137** are associated with origin quantum dots **111**, **113**, **115**, and **117**, respectively. These photons are also associated with destination quantum dots **121**, **123**, **125**, and **127**, respectively. Therefore, each quantum dot r_j in the quantum dot system **110** is connected to an associated quantum dot r_j' in the quantum dot system **120** through a single photon. To accomplish this, quantum dots of different sizes can be used, where each pair of dots at r_j and r_j' have the same size. For example, the quantum dot **111** can be the same size as the quantum dot **121**. Then these pair of quantum dots can be connected by the photon **131** with a proper resonant frequency. In an alternative implementation, an individual piece of optical fiber can connect corresponding quantum states (e.g., quantum dot **111** and quantum dot **121**). An optical fiber is a device which efficiently transmits light from one end of the fiber to the other end. When one end is placed very close to the emitting region of the microcavity containing quantum dot **121**, and parallel to the direction the photon can propagate in free space, then the single photon can enter into the fiber with a nearly 100% probability. When the other end is placed very close to the quantum dot **111**, then the photon can enter the microcavity containing quantum dot **111** with nearly 100% probability.

The teleportation system **100** also includes a second quantum dot system **120** that serves as the destination where the quantum dots in system **110** are teleported. Since the destination quantum dot system **120** includes only the quantum dots whose electron spin orientations have been teleported from the quantum dot system **110**, the number of quantum dots in this destination system can equal the number of quantum dots in the quantum dot system **110**.

Finally, the teleportation system **100** also includes numerous photon detectors. The number of photon detectors can be 4, 9, 2, or some other suitable number. In one implementation, the teleportation system **110** can include a photon detector for each quantum state present in the quantum dot system **110**. For example, the teleportation system **100** actually includes quantum detectors **141**, **143**, **145**, and **147** associated with the quantum dots **111**, **113**, **115**, and **117**. The photon detectors can be a photomultiplier tube, avalanche diode, or some other suitable detector that can record a polarization associated with the detected photon.

Turning now to FIG. **2**, this block diagram **100** illustrates the teleportation of the electron spin state from one of the quantum dots in system **110** to its corresponding quantum dot electron in system **120**. Teleporting the many-qubit states in the quantum dot system **110** can be considered as the independent teleportation of each spin state from r_j in quantum dot system **110** to r_j' in quantum dot system **120**. Therefore, the teleportation described with reference to FIG. **2** can apply to any of the qubit states, or quantum dots (e.g., quantum dot **112**) described with reference to FIG. **1**.

In the block diagram **100**, the quantum dot r_j' and the quantum dot r_j are each embedded in an optical microcavity. An optical microcavity is a structure formed by reflecting faces on the two sides of an optical medium, such that a standing wave is formed having a wavelength associated with the thickness of the optical medium. The microcavity **210** is associated with the quantum dot r_j' and the microcavity **220** is associated with the quantum dot r_j. These microcavities can be formed by growing layered structures such as Indium Gallium Arsenide, Gallium Arsenide, or some other suitable layered structure of light confinement. After the layered structure is grown, the microcavity is formed by etching mesas, or "whispering galleries" into the layered structures. For a high quality factor, a silicon microcavity can be plasma treated to form a highly reflecting silicon dioxide layer. The microcavities **210**, **220** have different lengths in the z-direction and the x-direction, so that a scattered photon cannot change its direction.

The photon source **130** initiates the teleportation process by emitting a single photon j that travels in the -z-direction. This photon can have a wavelength of 1.55 microns, 800 nanometers, or some other suitable wavelength. In addition, the photon j can be polarized in the x-direction and enter the microcavity **210** that contains the quantum dot r_j'. The quantum dot r_j' is initialized parallel to x. Before the photon j interacts with the quantum dot r_j', the initial wave function describing both the photon and the spin associated with dot r_j' is $|\Psi_{pe'}{}^{(1)}(0)\rangle=|\leftrightarrow\rangle|\leftarrow'\rangle$, which is a product state, meaning that the photon and electron spin are not "entangled". After this photon interacts with the quantum dot r_j', the resulting spin wave function is $|\Psi_{pe'}{}^{(1)}(T)\rangle=e^{iS_0^{hh}}|\Psi_{hh}{}^{(1)}\rangle+e^{iS_0^{lh}}|\Psi_{lh}{}^{(1)}\rangle$, where

$$|\psi_{hh}^{(1)}\rangle=\frac{(|\sigma_{(z)}^{+}\rangle|\uparrow'\rangle+|\sigma_{(z)}^{-}\rangle|\downarrow'\rangle)}{2}$$

and

$$|\psi_{lh}^{(1)}\rangle=\frac{(|\sigma_{(z)}^{-}\rangle|\uparrow'\rangle+|\sigma_{(z)}^{+}\rangle|\downarrow'\rangle)}{2}.$$

The interaction of the photon with the spin can occur through one of two possible pathways, as indicated on FIG. **3**. The photon can virtually excite an electron-light-hole pair (going from **340** to **320**), or an electron-heavy-hole pair (going from **340** to **310**). The full quantum-mechanical state is a linear superposition of these two potential pathways. The wavefunction labeled "hh" is for the heavy-hole pathway, that labeled "lh" is for the light-hole pathway. Whichever pair is virtually excited the system will return to configuration **330**, which is the same as its original configuration **340**.

After the photon j interacts with the spin in the quantum dot r_j', the photon travels to microcavity **220** containing the quantum dot r_j. This quantum dot r_j can have one excess electron in a general single-spin state $|\Psi_e{}^{(1)}\rangle=\alpha|\uparrow\rangle+\beta|\downarrow\rangle$, where the quantization axis is the z-axis. Now this photon can facilitate teleportation from microcavity **220** to microcavity **210** until the spin in the quantum dot at microcavity **220** decoheres. Decoherence is the process whereby the quantum state of a system becomes unrelated to its past state, principally through interaction with other quantum mechanical entities such as nuclei, other electrons, or phonons. When the photon j interacts with the quantum dot r_j, this photon can teleport the spin associated with this quantum state to the microcavity **210** associated with the quantum dot r_j'. After the photon j interacts with the quantum dot r_j, the linear polarization of the photon j and the spin in quantum dot r_j are measured. With these two measurements and a phase correction, it is possible to reconstruct the original spin state of quantum dot r_j at quantum dot r_j'.

Quantitative Description

Copying information in a classical, or typical, system is quite different than copying information in a quantum system. Specifically, the information contained in a quantum two-level system cannot be completely copied because of underlying physical phenomenon described by the no-cloning theorem. While copying is not possible, there is a mechanism for transferring quantum information without transferring the object itself containing the quantum information. For a quantum information processor, the operation that can partially replace copying is the ability to transfer quantum information from one system to another. When the transfer is separated into a channel of classical information (e.g., a telephone wire, fiber-optic cable, or other commonly used method of communicating "0"'s and "1"'s) and one of Einstein-Podolsky-Rosen (EPR) correlations (i.e., the distribution of entangled particles between origin and destination) it is called "quantum teleportation."

Quantum teleportation can be performed with photons. A single photon state, (i.e., left or right circular polarization of the photon) can be teleported with the aid of an entangled photon pair. Normally, quantum teleportation of a single photonic qubit involves generating one maximally-entangled two-photon pair, a two-photon entangled measurement of one member of this pair, and measurement of the original photonic qubit. A qubit, or quantum bit, is a unit of quantum information that is the quantum mechanical version of the classical bit.

As mentioned above, the photon source **130** initiates the teleportation process by emitting a single photon j that travels in the −z-direction and is initially linearly polarized in the x direction, to the microcavity **210** that holds the destination quantum dot r_j', whose spin is initialized parallel to x. Using the times $t_A<t_A+T<t_B<t_C<t_D$, where $t_A=0$ can help distinguish what occurs during each step of teleportation. Before the photon j interacts with the quantum dot r_j' (i.e., t=0), the initial spin wave function associated with dot r_j' is $|\Psi_{pe'}^{(1)}(0)\rangle=|\leftrightarrow\rangle|\leftarrow'\rangle$.

Turning now to FIG. **3**, this figure is a diagram of the electronic states of the quantum dot illustrating possible virtual processes excited by the photon j of FIG. **1**. For an upward spin-↑, light with a polarization of $\sigma^+_{(z)}$ excites an electron and a heavy hole as shown at reference numeral **310**. Light of polarization $\sigma^-_{(z)}$ for the same spin, excites an electron and a light hole (see reference numeral **320**). In both cases the selection rules and Pauli blocking force the light subsequently emitted from electron-hole recombination to have the same polarization as was absorbed, and the final occupations of electron states in the dot, as shown at reference numeral **330**, are the same as the initial occupations. These virtual processes have different probabilities, however, and this leads to Faraday rotation: a phase shift $e^{iS_o^{hh}}$ for light of polarization $\sigma^+_{(z)}$ and $e^{iS_o^{lh}}$ for light of polarization $\sigma^-_{(z)}$. In contrast, the phase shift is phase shift $e^{iS_o^{lh}}$ for light of polarization $\sigma^+_{(z)}$ and $e^{iS_o^{hh}}$ for light of polarization $\sigma^-_{(z)}$ for a downward spin-↓. Thus there is a conditional Faraday rotation of the components of the electron-photon state depending on the photon polarization and spin orientation.

After photon j interacts with the quantum dot r_j', the resulting spin wave function is $|\Psi_{pe'}^{(1)}(T)\rangle=e^{iS_o^{hh}}|\Psi_{hh}^{(1)}\rangle+e^{iS_0^{lh}}|\Psi_{lh}^{(1)}\rangle$, where

$$|\psi_{hh}^{(1)}\rangle = \frac{(|\sigma_{(z)}^{+}\rangle|\uparrow'\rangle + |\sigma_{(z)}^{-}\rangle|\downarrow'\rangle)}{2}$$

originates from the virtual process where a photon creates an electron and a heavy hole. The term

$$|\psi_{lh}^{(1)}\rangle = \frac{(|\sigma_{(z)}^{-}\rangle|\uparrow'\rangle + |\sigma_{(z)}^{+}\rangle|\downarrow'\rangle)}{2}$$

originates from the virtual process where the photon creates an electron and a light hole. Both $|\Psi_{hh}^{(1)}\rangle$ and $|\Psi_{lh}^{(1)}\rangle$ are EPR states (maximally-entangled states). A photon state $|\Psi\rangle$ can be written as

$$|\varphi\rangle = \frac{e^{-i\varphi}|\sigma_{(z)}^{+}\rangle + e^{i\varphi}|\sigma_{(z)}^{-}\rangle}{\sqrt{2}}$$

with a linear polarization rotated by z-axis with respect to the state $|\leftrightarrow\rangle$ of linear polarization in the x direction. Consequently,

$$\left\langle\psi_{pe'}^{(1)}(T)\right\rangle = \frac{e^{i\frac{S_0^{hh}+S_0^{lh}}{2}}}{\sqrt{2}}\left(\left|\frac{-S_0}{2}\right\rangle|\uparrow'\rangle + \left|\frac{+S_0}{2}\right\rangle|\downarrow'\rangle\right)$$

where $S_0=S_0^{hh}-S_0^{lh}$. Thus, the spin-photon interaction produces a conditional single-photon Faraday rotation around the z-axis by the angle $\pm S_0/2$. If $S_0=\pi/2$, the linear polarization of the incoming photon is rotated $-\pi/4$ by the spin up component, and at the same time is rotated $+\pi/4$ by the spin down component, yielding two orthogonal photon polarizations. Thus, $|\Psi^{(1)}_{pe'}(T)\rangle=|⤡>|\uparrow'>+|⤢>|\downarrow'>/\sqrt{2}$, which is maximally entangled. A maximally entangled state of two or more particles, when each particle has two orthogonal quantum mechanical states, corresponds to a many-particle state composed of the sum of two terms (A and B), in which both A and B are the product of one single particle wave function for each of the particles in the many-particle state, and each single-particle wave function for a given particle that appears in A is orthogonal to the single-particle wave function for the same particle in B. In order to enhance the spin-photon interaction sufficiently to achieve $S_0=\pi/2$, which yields maximally entangled states, each quantum dot is placed in its own microcavity (as described in FIG. **1**). Using a switchable cavity, as described below, permits the precise control of the Faraday rotation angle $S_0/2$ necessary for high fidelity teleportation. After interacting with the spin in the quantum dot system **120**, the photon is sent to quantum dot system **110**, and can be retained as a resource for teleportation from quantum dot system **110** to quantum dot system **120** until the spin at quantum dot system **110** decoheres. For coherence times of 100 ns, the maximum teleportation distance (determined by the speed of light) can be ~15 m; and for 10 ms, it is 1.5×10^6 m.

When it is time to teleport a spin in a quantum dot (e.g. **111**) part of quantum dot system **110** to another dot (e.g. **121**) part of quantum dot system **120** we let the photon interact with the quantum dot in the microcavity **220**, giving rise to a GHZ state in the hybrid spin-photon-spin system. A GHZ state is a maximally-entangled state of 3 particles. The quantum dot **111** can have one excess electron in a general single-spin state $|\Psi_e^{(1)}\rangle=\alpha|\uparrow\rangle+\beta|\downarrow\rangle$, where the quantization axis is the z-axis. After this interaction, the equation that results is

$$\left|\psi_{epe'}^{(1)}(t_C)\right\rangle = \frac{e^{i\frac{S_0^{hh}+S_0^{lh}}{2}}}{\sqrt{2}}(\alpha|\uparrow\rangle|-S_0/2-\pi/4\rangle|\uparrow'\rangle + \alpha|\uparrow\rangle|-S_0/2+\pi/4\rangle|\downarrow'\rangle \tag{1}$$

-continued

$$+\beta|\uparrow\rangle|+S_0/2-\pi/4\rangle|\uparrow'\rangle$$

$$+\beta|\uparrow\rangle|+S_0/2+\pi/4\rangle|\downarrow'\rangle$$

When $S_0=\pi/2$, $|\Psi^{(1)}{}_{epe'}(t_C)\rangle$ can be rewritten and the equation that results is:

$$\left|\psi^{(1)}_{epe'}(t_C)\right\rangle = \frac{1}{\sqrt{2}}[|\updownarrow\rangle(-\alpha|\uparrow\rangle|\uparrow'\rangle+\beta|\downarrow\rangle|\downarrow'\rangle)+|\leftrightarrow\rangle(\alpha|\uparrow\rangle|\downarrow'\rangle+\beta|\downarrow\rangle|\uparrow'\rangle)] \tag{2}$$

which results in

$$\left|\psi^{(1)}_{epe'}(t_C)\right\rangle = \frac{1}{\sqrt{2}}\left[\left(|\updownarrow\rangle|\leftarrow\rangle(-\alpha|\uparrow'\rangle+\beta|\downarrow'\rangle+|\rightarrow\rangle(-\alpha|\downarrow'\rangle-\beta|\downarrow'\rangle)\right]+\frac{1}{\sqrt{2}}|\leftrightarrow\rangle[|\leftarrow\rangle\beta|\uparrow'\rangle+\alpha|\downarrow'\rangle)+|\rightarrow\rangle\beta|\uparrow'\rangle-\alpha|\downarrow'\rangle\Big)\Big] \tag{3}$$

in the S_x representation for the spin at the origin (**110**).

If the linear polarization of the photon is measured first, then depending on the two initial spin orientations [see Equation 2], collapse of the wave function leaves the qubits at the quantum dot system **110** and at quantum dot system **120** in one of the four Bell states. After performing a single-spin measurement in the x direction of the spin in one of the quantum dots (e.g. **111**) of the quantum dot system **110** (which, as described below, can be done with a single photon), the spin state of the corresponding dot (e.g. **121**) in the quantum dot system **120** is projected onto (see previously described Equation 3) $|\Psi^{(1)}{}_{e1}(t_D)\rangle=-\alpha|\uparrow'\rangle+\beta|\downarrow'\rangle$, $|\Psi^{(1)}{}_{e2}(t_D)\rangle=-\alpha|\uparrow'\rangle-\beta|\downarrow'\rangle$, $|\Psi^{(1)}{}_{e3}(t_D)\rangle=\beta|\uparrow'\rangle+\alpha|\downarrow'\rangle$, $|\Psi^{(1)}{}_{e4}(t_D)\rangle=\beta|\uparrow'\rangle-\alpha|\downarrow'\rangle$ with equal probability. Thus if nothing further is done there is a 75% chance of error in the teleportation process, for three of the four states are not the same as the initial state. After communicating classically (i.e., with two bits) the outcome of the measurement of the linear polarization of the photon and quantum dot's **111** spin orientation along S_x to the destination quantum dots **120**, it is known which of the four states has been produced at quantum dot **121**, and the original spin state of quantum dot **111** can be reconstructed at quantum dot **121**, completing the teleportation. This teleportation system enables teleportation by using a single photon without an intermediate electronic qubit.

Expansion to a Many Qubit Teleportation System

As previously described with reference to FIG. **1**, the teleportation system **100** is a many qubit teleportation system because the quantum dot system **100** includes many quantum states numbered **111**-**117**. Similarly, the quantum dot system **120** also includes many quantum states numbered **121**-**127**. Because the quantitative description above referred to a single qubit within the quantum dot system **110**, a single quantum qubit within the quantum dot system **120**, and a single photon, this description corresponded to a single qubit teleportation system, such as the single bit system **105**. However, the previous quantitative description is equally applicable to the many-qubit teleportation system **100**.

The approach for two or more dots closely resembles that of one dot. The teleportation can still be performed bit-by-bit, so long as photon j (j=1,2,3, . . .) coming from the dot at r_j' of quantum dot system **120** travels to the dot at r_j of quantum dot system **110**, as described with reference to FIGS. **1**-**2**. Teleportation is mediated by photons that scatter independently off the dots and the conditional phase shifts from each spin can be treated independently. This approach provides a method of teleporting a many-qubit state of an arbitrary number of qubits, always relying on single-photon measurements.

Faraday rotation to entangle the photon and electron spin provides one of the possible ways to measure the spin with a single photon, which is described with reference to the following equation:

$$|\Psi_{pe'}{}^{(1)}(T)\rangle=e^{iS_0{}^{hh}}|\Psi_{hh}{}^{(1)}\rangle+e^{iS_0{}^{lh}}|\Psi_{lh}{}^{(1)}\rangle \tag{4}$$

where

$$|\psi^{(1)}_{hh}\rangle = \frac{(|\sigma^+_{(z)}\rangle|\uparrow'\rangle+|\sigma^-_{(z)}\rangle|\downarrow'\rangle)}{2}$$

and

$$|\psi^{(1)}_{lh}\rangle = \frac{(|\sigma^-_{(z)}\rangle|\uparrow'\rangle+|\sigma^+_{(z)}\rangle|\downarrow'\rangle)}{2}.$$

It is assumed that the microcavities have an additional resonant mode at a different frequency for photons propagating in the x direction. Equation 4 shows that if the spin on the quantum dot points in the +x (−x) direction, this incoming photon is converted into an outgoing circularly polarized photon $\sigma_{(x)}{}^+(\sigma_{(x)}{}^-)$. Measuring the circular polarization of the photon j after it escapes yields the spin orientation along x. It is possible for high fidelity because time-correlated single photon counting permits a counting efficiency close to one. In an alternative embodiment, the electrical single-spin measurements at quantum dot system **110** can be accomplished with a single electron transistor (SET), by converting the spin information to charge information.

The spin-selective coupling between the electron spins and the photons, which leads to their mutual entanglement (and eventually to teleportation), is enhanced by surrounding each of the dots by a unique high-Q microcavity (see FIG. **2**). Each microcavity has a single well-defined left-circularly polarized photon mode (and a right-circularly polarized photon mode of identical frequency) nearly resonant with the fundamental optical transition of the quantum dot. By relying on non-resonant interaction of photons both in the z-direction and the x-direction, the four highest-energy valence states are nearly degenerate, which corresponds to nearly spherical dots of zincblende or wurtzite material. Nearly degenerate means that energy difference of these four states are much smaller than the detuning energy $h\omega_d$. Then, in both the z-direction and the x-direction, there is a conditional Faraday rotation depending on the spin state of the excess electron on the dot.

The Faraday rotation from the virtual process in which a photon creates an electron and heavy hole is three times larger than that from the virtual process in which a photon creates an electron and light hole, and is in the opposite direction. The transition matrix elements for exciting an electron and heavy hole (V_{hh}) or an electron and light hole (V_{lh}) are calculated from the electron-photon interaction Hamiltonian eA·p/2mc, where p is the momentum of an electron and A is the vector potential of the electromagnetic field. The full Hamiltonian in the rotating frame reads then

$$H = \begin{pmatrix} E_c - \hbar\varpi_d & V_{hh} & 0 & 0 \\ V_{hh} & E_c & 0 & 0 \\ 0 & 0 & E_c - \hbar\varpi_d & V_{lh} \\ 0 & 0 & V_{lh} & E_c \end{pmatrix} \quad (5)$$

where E_c is the energy of the excess electron. The basis states are $\{|\uparrow\rangle|\sigma_{(z)}^{+}\rangle; |\uparrow\downarrow,hh\rangle; |\uparrow\rangle|\sigma_{(z)}^{-}\rangle|\uparrow\downarrow,lh\rangle\}$; the states $|\uparrow\downarrow,hh\rangle$ and $|\uparrow\downarrow,lh\rangle$ are shown schematically in FIG. **3**. The Hamiltonian for the spin down is similar. The rates of Faraday rotation for these virtual transitions are $\Omega_{hh}=V_{hh}^2/h^2\omega_d=3V_{lh}^2/h^2\omega_d$ and $\Omega_{lh}=V_{lh}^2/h^2\omega_d$. Thus the phase shift accumulated by the photon state during the photon's residence time T in the microcavity is given by $S_0^{hh}=\Omega_{hh}T$ and $S_0^{lh}=\Omega_{lh}T$ for heavy and light holes, respectively. The total phase shift is governed by the following equation $S_0=S_0^{hh}-S_0^{lh}=(\Omega_{hh}-\Omega_{lh})T=\Omega_0T$. If $S_0=\pi/2$ (modulo 2π), the photon and electron spin become maximally entangled.

Having an interaction strength between the photon and the quantum dot transition to be weak (i.e. V_{hh}, $V_{lh}<<h\omega_d$) creates certain consequences. The frequency of the photon is tuned below the band gap E_{gap}, which leads to non-resonant interaction. Values for the band gap can be $E_{gap}=1$ eV, 3 eV, or some other suitable value. The level broadening can be $\Gamma=10$ μeV, 15 μeV, or some other suitable value. An interaction time between the photon and the electron spin can approximately T=1 ns, 2 ns, or some other suitable time that is much smaller than the limiting spin decoherence time in semiconductor nanostructures. For an interaction time T=1 ns, the bandwidth that results $\Gamma_{photon}=0.7$ μeV. If the size of the microcavity is 3.5 μm^3, V_{hh} is typically 50 μeV. Thus for a reasonable choice $h\overline{\omega}_d\approx1.5$ meV$>>\Gamma$, Γ_{photon}, the scattering frequency can be adjusted to

$$\Omega_0 = \frac{\pi}{2} \times 10^9 s^{-1},$$

and consequently $S_0=\pi/2$.

To control the interaction time T precisely, the microcavity can be actively Q-switched with an electro-optic modulator. Response times for these types of modulators can be less than 1 ps. With this response time, it leads to a phase error on the order of 1 ps/1 ns=0.1%. In contrast, the Q-factor can be as high as $Q=1.25\times10^8$, which is equivalent to a photon lifetime of $\tau=43$ ns. Although this Q was achieved for a 10^5 μm^3 cavity, the same processing can be applied to 3.5 μm^3 cavities. The theoretical limit on the Q value of a cavity of this size is $Q\sim10^{13}$. Insertion of a spherical (colloidal) quantum dot (as part of a 2-quantum dot system or a system with more quantum dots) into a photonic crystal with holes can be the simplest method of incorporating quantum dots controllably into a high-Q microcavity. After T=1 ns in the small cavity, before Q-switching, the escape probability is $1-e^{-T/\tau}=2\%$. Thus, the entangled state can be produced with high fidelity. This 2% estimate also applies to the fidelity of the single-shot photon measurements of the electron spin orientation. Transmission into or out of the cavity has a fidelity much greater than 99% fidelity because commercial highly transmissive coatings have reflectivities <<0.1%. Hence, the total fidelity of the teleportation can exceed 97%.

The many qubit teleportation system **100** can also provide a general link between spintronic quantum information devices and photonic quantum information devices. Letting the photon interact only with the spin at quantum dot system **110** gives the possibility to transfer a $|\Psi(0)\rangle$ onto the photon state, and back (i.e., this provides a way of transferring quantum information from an electron spin in a quantum dot to a photon, or from a photon to an electron spin in a quantum dot). An example of a general class of opto-spintronic quantum information devices relying on this ability includes a Quantum Dynamic RAM (QDRAM) memory. In other words, the teleportation system **100** can be used in implementing QDRAM. For this type of memory device, the many-spin state can be transferred to the many-photon state and back. As the decoherence time of photons is much longer than the decoherence time of the spins of electrons, it is useful to keep the quantum information encoded as photons between error-correcting operations acting on the electron spins. Thus, the refresh time can be much longer than the decoherence time of the electrons. This can enhance the maximum distance possible for teleportation beyond that determined from the electron spin coherence time. With the teleportation system **100** that enables efficient transfer of quantum information between spintronic and photonic systems, many other such devices can be imagined which can also exploit the complementary advantages of spintronic and photonic quantum information processing.

The particular embodiments disclosed above are illustrative only, as the teleportation system **100** can be modified and practiced in different, but equivalent, manners apparent to those skilled in the art having the benefit of the teachings herein. Furthermore, no limitations are intended to the details of construction or design herein shown, other than as described in the claims below. It is, therefore, evident that the particular embodiments disclosed above may be modified and all such variations are considered within the scope and spirit of this invention. Accordingly, the protection sought herein is set forth in the claims below.

The invention claimed is:

1. A method for creating a logic state, the method comprising the steps of:

receiving a photon with a first initial polarization; and

interacting the photon with a first semiconductor crystal having a first spin orientation, wherein said interaction produces a first resulting polarization of said photon dependent upon the first spin orientation of said first semiconductor crystal, wherein interacting the photon with the first semiconductor crystal generates a maximally entangled state.

2. The method of claim **1**, further comprising the steps of:

measuring the first resulting polarization of said photon; and

determining an original spin orientation associated with the first semiconductor crystal in response to the measured first resulting polarization.

3. The method of claim **2**, further comprising the step of using the measured resulting polarization and a measured spin polarization for correcting potential errors associated with teleportation.

4. The method of claim **1**, further comprising the steps of:

interacting the photon with a second semiconductor crystal having a second spin orientation, wherein said interaction produces a second resulting polarization of said photon dependent upon the second spin orientation associated with the second semiconductor crystal; and defining a logic state using the second resulting polarization, the first spin orientation, and the second spin orientation.

5. The method of claim 4, further comprising the steps of:

causing the first spin orientation to have a definite initial orientation;

measuring the second resulting polarization;

receiving a second photon with a second initial polarization propagating in a direction different from the first photon;

interacting the second photon with the second semiconductor crystal for producing a third resulting polarization dependent upon the second spin orientation associated with the second semiconductor crystal;

measuring the third resulting polarization; and

determining the spin orientation associated with the second semiconductor crystal after interaction with the first photon, but before interacting with the second photon, in response to measuring the third resulting polarization.

6. The method of claim 5, further comprising the step of using the measured second and third resulting polarizations for correcting potential errors associated with teleportation.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 7,667,995 B1
APPLICATION NO. : 11/352387
DATED : February 23, 2010
INVENTOR(S) : Michael N. Leuenberger, Michael E. Flatté and David D. Awschalom

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

In column 1, line 3; insert the following paragraph before "CROSS REFERENCE TO RELATED APPLICATIONS":

--GOVERNMENT SUPPORT CLAUSE

This invention was made with government support under Grant No. DAAD 19-01-1-0490 awarded by the Defense Advanced Research Agency/Army Research Office, Grant No. N00014-04-1-0457 awarded by the Defense Advanced Research Agency/Office of Naval Research SPINS and Grant No. DMR-0305223 awarded by the National Science Foundation. The government has certain rights in the invention.--

Signed and Sealed this

Twenty-third Day of March, 2010

David J. Kappos

David J. Kappos
Director of the United States Patent and Trademark Office

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 7,667,995 B1
APPLICATION NO. : 11/352387
DATED : February 23, 2010
INVENTOR(S) : Leuenberger et al.

Page 1 of 1

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

On the Title Page:

The first or sole Notice should read --

Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 659 days.

Signed and Sealed this

Seventh Day of December, 2010

David J. Kappos

David J. Kappos
Director of the United States Patent and Trademark Office

US 20130322873A1

(19) **United States**

(12) **Patent Application Publication**

STEVENSON et al.

(10) **Pub. No.: US 2013/0322873 A1**

(43) **Pub. Date: Dec. 5, 2013**

(54) **SYSTEM AND METHOD FOR QUANTUM TELEPORTATION**

(71) Applicant: **KABUSHIKI KAISHA TOSHIBA**, Tokyo (JP)

(72) Inventors: **Richard Mark STEVENSON**, Cambridge (GB); **Martin Brian WARD**, Cambridge (GB); **Anthony John BENNETT**, Cambridge (GB); **Joanna Krystyna SKIBA-SZYMANSKA**, Cambridge (GB); **Andrew James SHIELDS**, Cambridge (GB)

(73) Assignee: **KABUSHIKI KAISHA TOSHIBA**, Tokyo (JP)

(21) Appl. No.: **13/895,683**

(22) Filed: **May 16, 2013**

(30) **Foreign Application Priority Data**

May 16, 2012 (GB) 1208640.1

Publication Classification

(51) **Int. Cl.**
H04B 10/07 (2006.01)

(52) **U.S. Cl.**
CPC *H04B 10/07* (2013.01)
USPC **398/25**

(57) **ABSTRACT**

A system for quantum teleportation of a quantum state of an input photon, including: a light emitting diode configured to produce a polarization entangled photon pair; a beam splitter configured to direct photons of the entangled photon pair along respective first and second paths; a measurement unit performing a joint measurement on the input photon; a timing unit configured to measure a first delay between the input photon and the photon of the entangled photon pair at a point of maximum indistinguishability of the photons as they pass through the joint measurement unit, and to measure a second delay between the two photons of the entangled photon pair as they exit the light emitting diode; a controller configured to determine a teleportation measurement is valid if the first delay is within a first predetermined timing window and the second delay is within a second predetermined timing window.

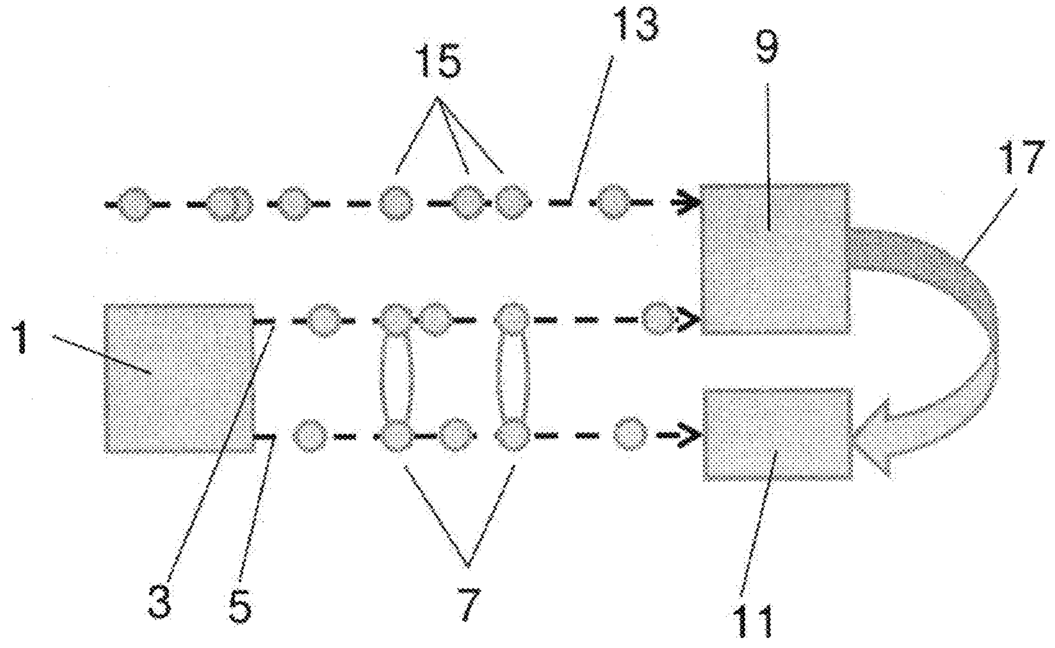

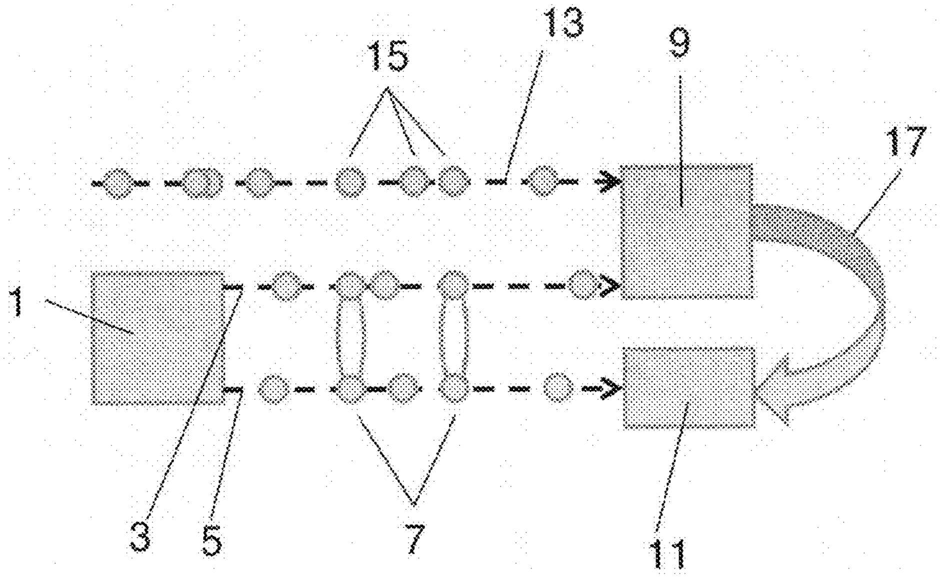
15
13
9
17
1
3
5
7
11

Figure 1.

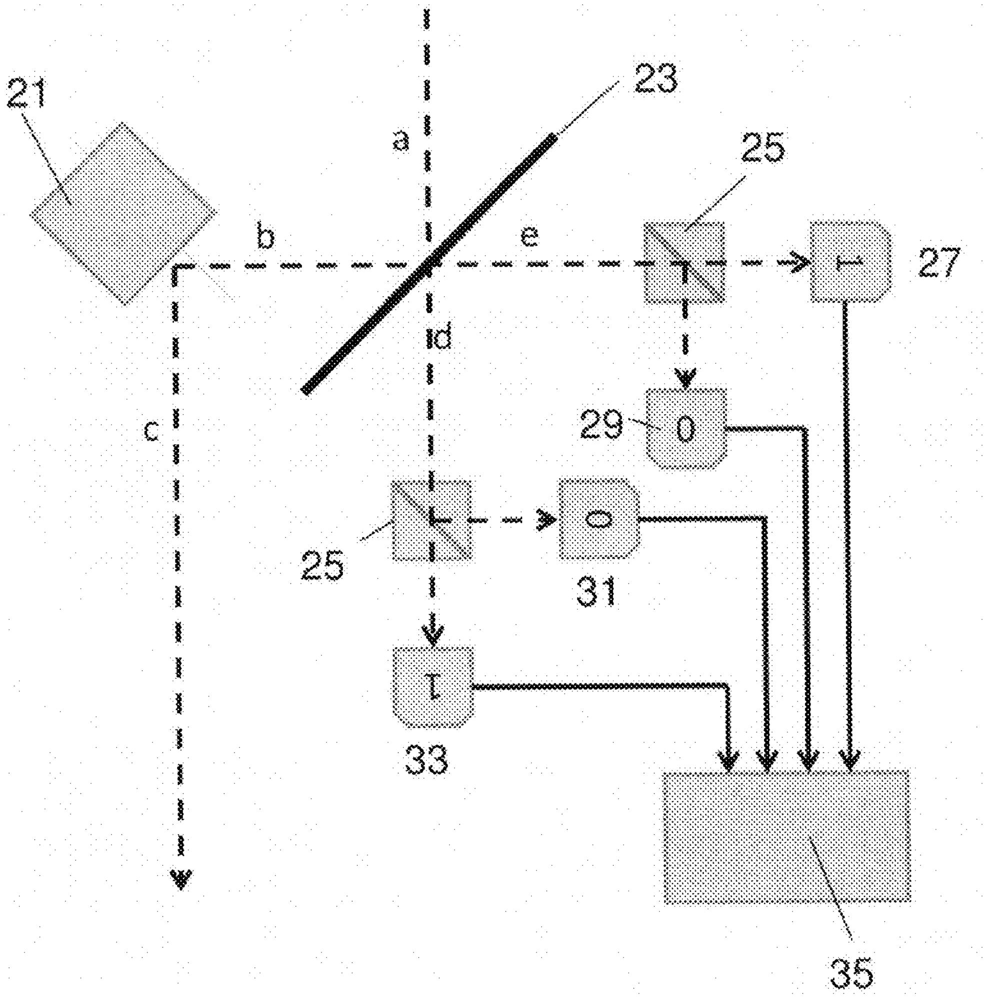
21
23
a
25
b
e
27
d
29
0
c
25
31
33
35

Figure 2.

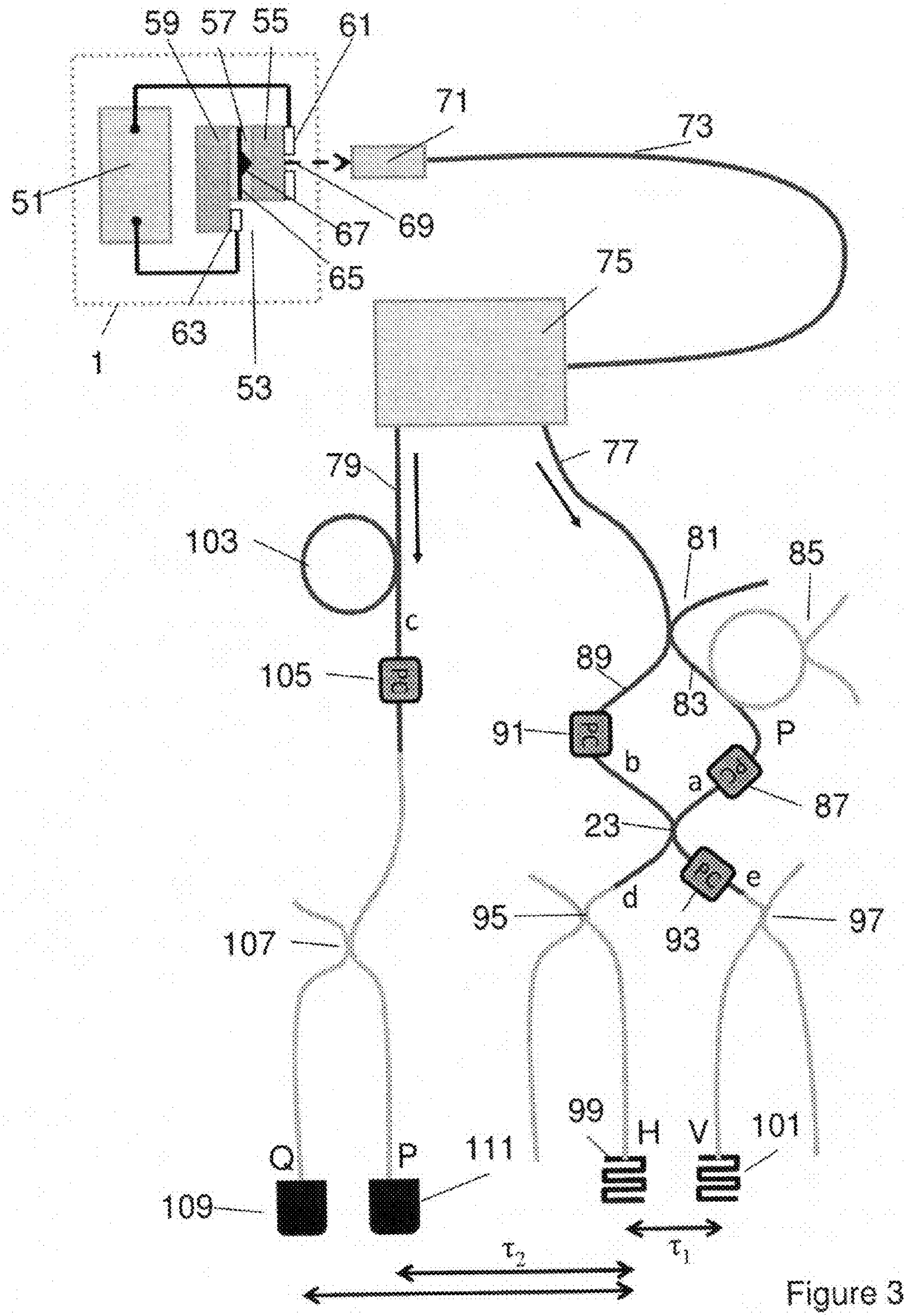
c
PC
PC
P
b
a
PC
PC
e
d
H
V
Q
P
τ2
τ1

Figure 3.

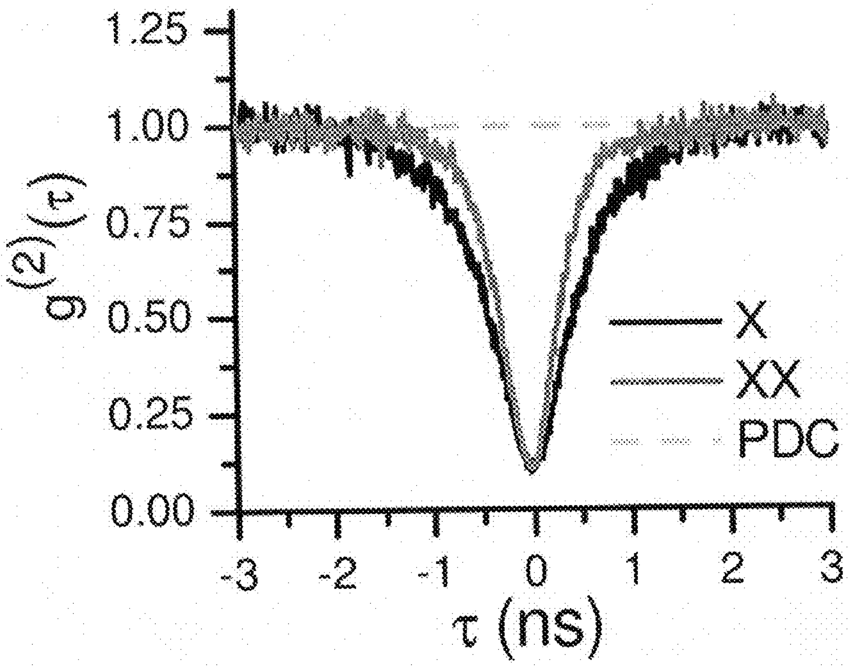

Figure 4a.

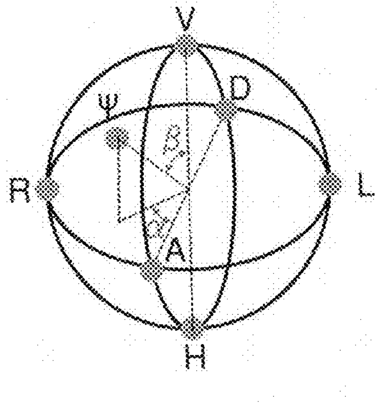

Figure 4b.

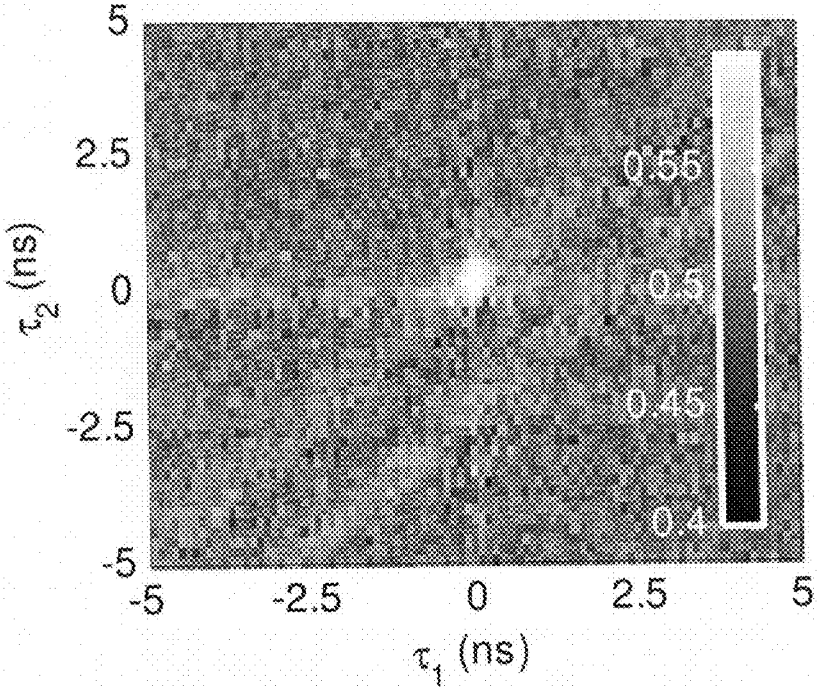

Figure 5a.

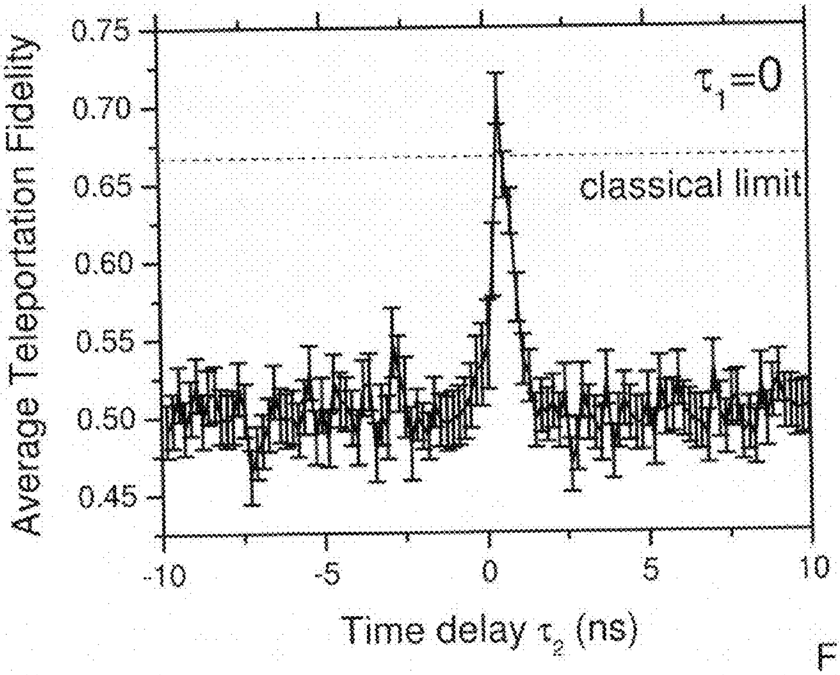

Figure 5b.

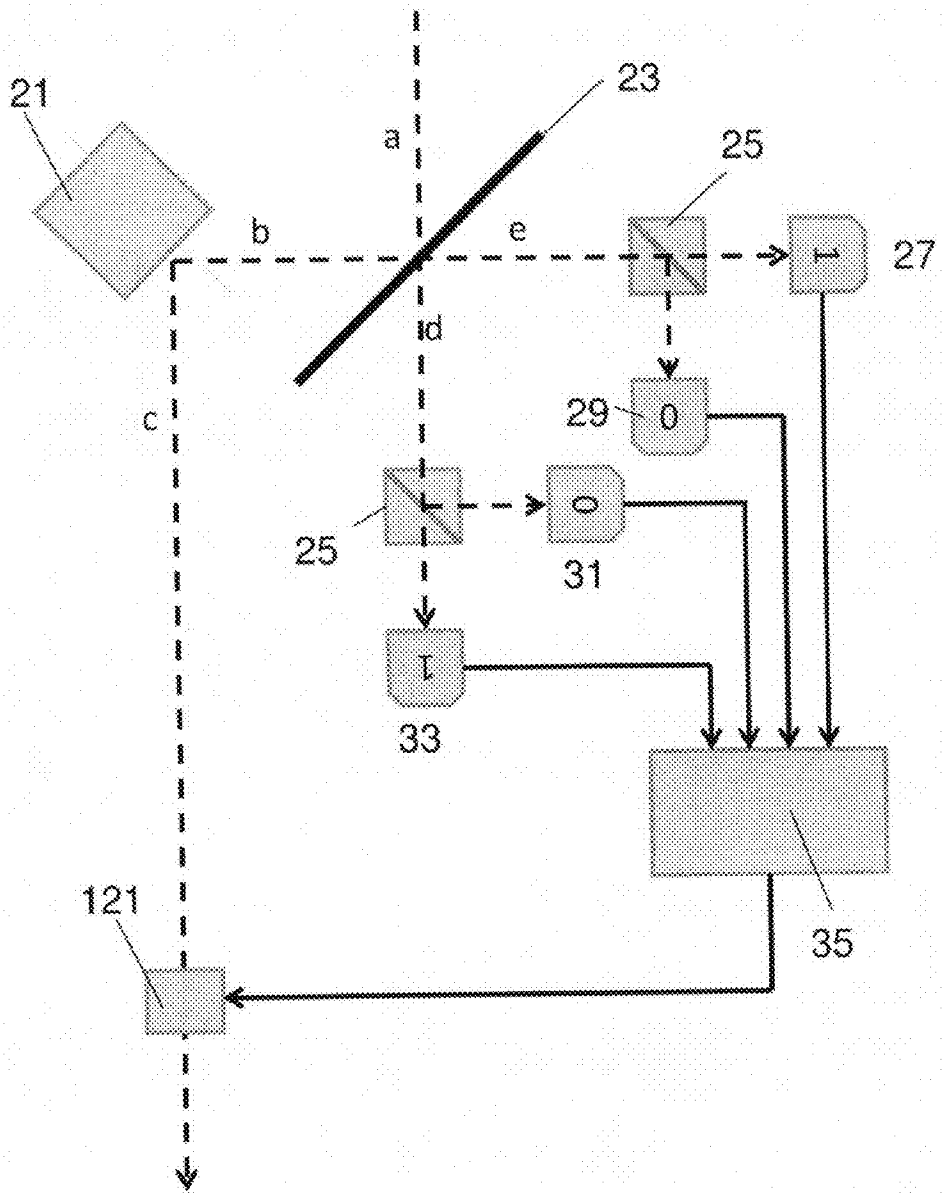
21
23
a
25
b
e
27
d
29
0
c
0
25
31
I
33
35
121

Figure 6

SYSTEM AND METHOD FOR QUANTUM TELEPORTATION

FIELD

[0001] Embodiments described herein generally relate to systems and methods for quantum teleportation.

BACKGROUND

[0002] The 'no-cloning' theorem states that quantum information cannot be copied, which has profound implications for quantum information technology. The security of quantum cryptography depends directly upon it, by encoding information on single photons. However, without the ability to copy information the options to create simple quantum communication networks are limited, and in quantum computing, losses due to imperfect measurements or probabilistic logic gates can terminate a quantum algorithm. Quantum teleportation, where quantum information is destroyed so that it may be transferred simultaneously to another location, has been proposed as an elegant solution. In quantum communication networks teleportation allows a quantum channel between two nodes to be established. In quantum computing based on linear optics, the so called feed-forward technique allows probabilistic logic operations to be performed off-line on sacrificial qubits until they succeed, after which the intended input qubits can be teleported through with success probability arbitrarily close to unity.

BRIEF DESCRIPTION OF THE DRAWINGS

[0003] FIG. **1** is a schematic of a teleportation system in accordance with an embodiment of the present invention;

[0004] FIG. **2** is a more detailed schematic of the embodiment of FIG. **1**, with the measurement unit shown in more detail;

[0005] FIG. **3** is a yet more detailed schematic of FIG. **2** with the photon paths shown in more detail;

[0006] FIG. **4***a* is a plot of second-order-correlation against time delay between photon detection events and FIG. **4***b* is a Poincare sphere;

[0007] FIG. **5***a* is a plot of the fidelity of the teleportation of a state against the first delay time and second delay time, and FIG. **5***b* is the teleportation fidelity against second delay time for a first delay time of zero; and

[0008] FIG. **6** is a schematic of a teleportation system in accordance with a further embodiment of the invention.

DETAILED DESCRIPTION

[0009] Embodiments of the present invention provide a system for quantum teleportation of a quantum state of an input photon, the system comprising:

- **[0010]** a light emitting diode configured to produce a polarisation entangled photon pair;
- **[0011]** a beam splitter configured to direct one photon of an entangled photon pair along a first path and the other photon of said entangled pair along a second path;
- **[0012]** an input for said input photon;
- **[0013]** a measurement unit for performing a joint measurement on the input photon with one of the photons of an entangled photon pair directed along the first path, the measurement unit comprising a first detector unit for detecting two photons upon which a joint measurement has been performed;
- **[0014]** a second detector unit configured to detect the photon from said entangled photon pair directed along the second path;
- **[0015]** a timing unit is configured to measure a first delay, said first delay being the magnitude of the delay between the input photon and the photon of the entangled photon pair at the point of maximum indistinguishability of the photons as they pass through the joint measurement unit, the timing unit being further configured to measure a second delay, said second delay being the magnitude of the time between the two photons of the entangled photon pair as they exit the light emitting diode; and
- **[0016]** a controller configured to determine that a teleportation measurement is valid if the first delay is within a first pre-determined timing window and the second delay is within a predetermined timing window.

[0017] In an embodiment, the system further comprises an electrical source for said light emitting diode and wherein said electrical source is a D.C. source. Such a driven source provides a quasi continuous streams of anti-bunched photons.

[0018] The first delay can be measured in a number of ways, in one embodiment said timing unit is configured to determine the first delay time from the detection time of the said two photons by the first detector unit. In further embodiments, the timing unit is also configured to compensate for variations between the path lengths from the point where the respective spatial modes of the two photons meet in the joint measurement unit to the first detector unit. For example, in an embodiment, there are two detectors in the first detecting unit and the will be optical output paths leading from the point where the respective spatial modes of the two photons meet in the joint measurement unit to each of the detectors in the first detector unit. The timing unit is configured to determine the delay between the output paths and compensate for this delay.

[0019] In some embodiments, the joint measurement unit comprises a beam splitter to permit two-photon-interference and wherein said timing unit is configured to determine the first delay time from the detection time of the said two photons by the first detector unit, the timing unit also being configured to compensate for variations between the path length taken by the two photons from the beam splitter to the first detector unit. However, it should be noted that in some further embodiments, when a beam splitter is used, the joint measurement unit may be configured such that the point where the respective spatial modes of the two photons meet does not coincide with the beam splitter.

[0020] In an embodiment the first timing window will be from 0 to t_{1max} where t_{1max} is the coherence time of the photon which follows the first path. In some embodiments, this will be 400 ps or less, in other embodiments 200 ps or less. As it will not usually be known which photon will arrive first, in some cases the magnitude of the delay is measured.

[0021] In a further embodiment the timing unit is configured to measure the second delay from the time when the photon which follows the second path is received at the second detector, the timing unit being further configured to compensate for differences in path lengths taken by the two photons of the entangled photon pairs.

[0022] In a further embodiment, the system further comprises a blocking unit located in the second path, said timing unit being configured to operate said blocking unit to allow the transmission of a photon along the second path if the second delay time is below the second timing window

[0023] In some embodiments, said second timing window is from 0 to $t_{2_{max}}$, where $t_{2_{max}}$ is of the order of the exciton radiative lifetime. For example, $t_{2_{max}}$ may be 1 ns or less. Entangled photons will be produced from decay of a biexciton where a first photon is emitted due to the decay of the biexciton and then a second photon will be emitted due to the decay of the remaining exciton. The photon emitted due to decay of the biexciton will be emitted at the same time or before the photon which is emitted due to exciton decay. However, the system could be configured such that either the first or the second photon can be directed along the first path.

[0024] In a further embodiment, the system further comprises a blocking unit located in the second path, said timing unit being configured to operate said blocking unit to block the transmission of a photon along the second path unless said controller has determined that the first delay time is within the first timing window. In this embodiment, the detection of photons from the joint measurement takes place first and, if the first delay time is short enough to suggest that the joint measurement could result in teleportation of the state of the input photon, the blocking unit is configured to open to allow the transmission of the photon which follows the second path.

[0025] An electrically driven entangled light source is used. The emitted photons of such sources may have a poor coherence time. This may result in partial distinguishability between the photons which undergo the joint measurement and this can affect the possibility of successful teleportation and will thus degrade the overall quality, or fidelity, of teleported photons. In an embodiment, to improve the teleportation fidelity, the system further comprises a state measurement unit configured to perform state measurements on both of the photons which pass through the joint measurement unit and wherein the controller determines that the teleportation measurement is valid if, additionally the state measurements on both photons agree with at least one of a predetermined set of results. For example, the photons which undergo the joint teleportation measurement will be passed through polarising beam splitters in and their polarisations will be measured.

[0026] In a further embodiment, if a blocking unit it used which is configured to only allow the passage of a second photon if the first delay time is within the first timing window, the blocking unit may also be configured to only allow the passage of the second photon if the first delay time is within the first timing window and the results of the state measurement unit indicate that teleportation has occurred.

[0027] In an embodiment, the joint measurement is a measurement involving Bell states or mixtures with Bell states.

[0028] In a further embodiment, the controller is configured to allow post measurement selection of valid measurements

[0029] The first detector unit may comprise first and second detectors, the first and second detectors being superconducting detectors. These types of detectors allow high time resolution to be achieved.

[0030] The system may be configured such that the source also provides the input photon. For example, the system may further comprise an input photon delay unit and a switch, said switch being configured to direct a photon from the first path into the said input photon delay unit, wherein in the input photon delay unit, the photon from the said source is delayed to coincide with a further photon output by the source such that the photon which is delayed becomes the input photon. The polarisation state of the input photon may be modified as desired before the joint measurement in order for the state to be teleported to be selected.

[0031] In some embodiments, the state to be teleported is a superposition state.

[0032] The light emitting diode may comprise a quantum dot. In a further embodiment, the quantum dot is provided in a p-i-n diode.

[0033] The above teleportation system may be used in a number of apparatus such as a quantum computer, quantum relay etc. The teleportation may also be used for quantum communication.

[0034] Further embodiments provide a method of teleporting of a quantum state of an input photon, the method comprising:

[0035] providing a polarisation entangled photon pair from a light emitting diode;

[0036] directing one photon of the entangled photon pair along a first path and the other photon of said entangled pair along a second path;

[0037] providing an input photon;

[0038] performing a joint measurement on the input photon with one of the photons of an entangled photon pair directed along the first path and detecting the two photons upon which a joint measurement has been performed;

[0039] detecting the photon from said entangled photon pair directed along the second path;

[0040] measuring a first delay, said first delay being the magnitude of the delay between the input photon and the photon of the entangled photon pair at the point of maximum indistinguishability of the photons as they undergo joint measurement,

[0041] measuring a second delay, said second delay being the magnitude of the delay time between the two photons of the entangled photon pair as they exit the light emitting diode; and

[0042] determining that a teleportation measurement is valid if the first delay is within a first timing predetermined window and the second delay is within a second predetermined timing window.

[0043] FIG. **1** shows schematically teleportation of photons from a sender, commonly referred to as Alice, to a receiver, commonly referred to as Bob.

[0044] In FIG. **1**, entangled light is generated using electrically driven entangled light source **1**. The electrically driven entangled light source emits the individual component photons of entangled pairs into two spatial modes **3** and **5**. In the system of FIG. **1** the source is electrically driven and the electrical driving current to the source is d.c. This causes a continuous stream of single photons to be produced in each output mode. The photons streams in spatial modes **3** and **5** are anti-bunched, which means that there is suppressed probability of generating two photons in a given mode at the same time.

[0045] In one embodiment, the photon source comprises a semiconductor quantum dot as the active element. This type of source will be described in more detail in relation to FIG. **4**.

[0046] In the system of FIG. **1**, photons in modes **3** and **5** emitted at the same time from the entangled LED source (ELED), form entangled photon pairs **7**. Photons in modes **3** and **5** are distributed to sender Alice **9**, and receiver Bob **11**

respectively. Distribution of spatial modes **3** and **5** may be achieved by free-space or fibre optic channels.

[0047] Sender Alice **9** also receives input photons **13** in a further spatial mode **15**. The input photons **15** are required to have similar frequency to those in mode **3**, but may be delivered in pulses or continuously, and need not be anti-bunched.

[0048] Sender Alice has a measurement unit which performs a joint measurement on two photons, one from each input modes **13** and **3**. The measurement unit has a detection unit which is configured to detect both photons from input modes **13** and **3**. The system also comprises a timing unit (not shown) which measures the time delay between the detection of the two photons. From this measured time delay it is possible to determine the time delay τ_1 relevant to the joint measurement process, which is the delay between the input photon and the photon of the entangled photon pair at the point of maximum indistinguishability of the photons as they pass through the joint measurement unit. In general, the joint measurement will involve passing both photons through a beam splitter. Typically the delay will be between the two photons when their spatial modes meet at a beamsplitter. In some embodiments, the photons may pass through the beam splitter at different times, but one of the photons may be delayed close to the beam splitter to allow interference to take place.

[0049] The joint measurement shall reveal no information of the qubit state of the photons, which if polarisation encoding schemes are used means that the polarisation of a photons in modes **3** and **13** is not revealed by the measurement. The joint measurement basis is selected so that after completion, the qubit state of the input photon in mode **13** may be transferred, via entanglement with a photon in mode **3**, to a photon in mode **5**. Examples of suitable joint measurements are joint detection of two photons, one in modes **3** and one in mode **13**, in a Bell state such as $(|H_3V_{13})-|V_3H_{13}))/\sqrt{2}$, or detection of a pair of photons in an pure two-photon state such as $|H_{3'}V_{13'})$, where modes **3'** and **13'** are the output modes of a beamsplitter with modes **3** and **13** as inputs.

[0050] Receiver Bob **11** measures the qubit state of photons received in mode **5** and has a detector, the second detector to detect these photons. The basis for Bob's measurement may be the logical qubit basis, or rotation thereof. For photonic qubits encoded in the polarisation degree of freedom, with horizontal (H) polarisation corresponding to 0, and vertical (V) polarisation corresponding to 1, this corresponds to a measurement distinguishing between H and V polarisation, or some basis rotation such as between left (L) and right (R) or diagonal (D) and anti-diagonal (A) polarised photons.

[0051] Bob also measures the detection time of the received photon, relative to those detected by Alice. This may be achieved by detection times measured by Alice and Bob relative to the same clock, relative to separate synchronised clocks, or relative to timing signals generated by detection events.

[0052] In this embodiment, teleportation is achieved using time-based post-selection using the above mentioned timing unit and a controller. Teleportation may generally only be achieved when a photon in mode **13** is detected by Alice at the same time as a photon in mode **3**, which was generated at the same time as a photon in mode **5**, detected by Bob.

[0053] The first stage of ensuring these conditions are met is by post-selecting events where the time difference between photons detected in a joint measurement by Alice **9**, the first delay, is sufficiently close to zero, for example within the coherence time of photons in modes **13** and **3**.

[0054] Information which records the time of such successful coincidences measured by Alice is transmitted to Bob via classical communication link **15**, which could be for example an Internet connection.

[0055] Dependent on the configuration of the joint measurement, it may also be necessary for Alice to additionally transmit the measurement outcome to ensure that the correct transformation is applied prior or after Bob's measurement to photons in mode **5**, in order to reproduce the input state in mode **15**. This will be explained in more detail with reference to FIG. **2**.

[0056] Finally, a controller performs post-selection based on the measurement time of photons in mode **5**, relative to successful, post-selected, joint measurement times communicated from Alice via classical channel **17**. Photons in mode **5** may be identified as being emitted at similar times to the photon formerly in mode **3** and detected by Alice, using the relative detection time between Bob's photon and Alice's joint measurement, and also the relative delays in the system due to the propagation times down modes **3** and **5** and the measurement apparatus of Alice and Bob.

[0057] FIG. **2** shows the operation principles of an experimental implementation of the present invention. Photonic qubits are encoded in the polarisation degree of freedom, and an arbitrary input photon qubit state Φ_a, in mode a may be written as;

$$\Phi_a=\alpha|0\rangle_a+\beta|1\rangle_a$$

[0058] In addition, and entangled photon pair state $|\Psi\rangle_{bc}$ is generated by entangled light source **21**, which are distributed amongst modes b and c, written as;

$$|\Psi\rangle_{bc}=(|0_b0_c\rangle+|1_b1_c\rangle)/\sqrt{2}$$

[0059] The three-photon input state is $|\Phi\rangle_a\otimes|\Psi\rangle_{bc}$ where:

$$|\Phi\rangle_a\otimes|\Psi\rangle_{bc}\propto\alpha|0_a0_b0_c\rangle+\alpha|0_a1_b1_{1c}\rangle+\beta|1_a0_b0_c\rangle+\beta|1_a1_b1_c\rangle$$

[0060] Modes a and b overlap at 50/50 beamsplitter **23**, which assuming photons in modes a and b are indistinguishable except for polarisation, and accounting for phase changes upon reflection, yields the following, 3-photon output state upon detection of a single photon in modes d and e.

$$|\Psi\rangle_{deb}\propto(|0_d1_e\rangle-|1_d0_e\rangle)(\alpha|1\rangle_c-\beta|0\rangle_c)$$

[0061] Usually, the joint measurement performed is detection of a single photon in mode d and a single photon in mode e. This is often referred to a Bell measurement, and in the ideal case, the photons in modes d and e are indeed in the entangled Bell state $(|0_d1_e\rangle-|1_d0_e\rangle)/\sqrt{2}$.

[0062] In real systems, linear optics are not perfect and although, for example, a 50:50 beam splitter may be used, typically such a beam splitter will not be perfect. Such imperfections will cause some small variations to the Bell States such that they are not true and perfect Bell states.

[0063] Also, to avoid any confusion, the term Bell state is used to mean the pure Bell states and mixtures of Bell states, for example Werner states.

[0064] After joint measurement of the first two photons, the output state is $\Phi_c=\alpha|1\rangle_c-\beta|0\rangle_c$. This qubit state is the same is the input state, apart from a simple unitary transformation of a bit-flip and phase change, which can be compensated for using logic or physical rotation of the output photons with a polarisation controller. The joint measurement implemented using a beamsplitter with an entangled photon pair and input photon thus performs quantum teleportation.

[0065] In the above embodiment, an electrically driven entangled light source is used. The emitted photons of such sources may have a poor coherence time. This may result in partial distinguishability between photons in modes a and b, which will no longer maximally interfere at the beamsplitter. In this case, some of the photons in modes e and d will not give rise to successful teleportation, and will degrade the overall quality, or fidelity, of teleported photons. For example, there will be contributions from pairs of photons of the same polarisation in modes e and d.

[0066] In one embodiment, these can be eliminated by additionally performing a state measurement, using a state measurement unit, for example the polarisation of the photons in modes e and d may be measured using polarising beamsplitters **25** and **27**. Ensuring coincident detection of a photon in modes e and d are oppositely polarised, by accepting events from photon detector combinations **27** and **31** or **29** and **33**, removes these unwanted coincidences, and improves the teleportation fidelity. In other types of sources for example, laser-driven entangled light sources such as parametric down converters, the coherence time is typically manipulated to acceptable levels at the expense of efficiency.

[0067] Timing unit **35** measures the time between photons registered by each of the detectors **27**, **29**, **31**, and **33** in order to determine the first delay time. Successful coincidences are selected for communication to Bob if the first delay time between photons is below a first threshold, and the polarisation combination is correct.

[0068] In an embodiment, the first threshold is set at the coherence time, which for an entangled light emitted diode is typically around 200 ps. In a further embodiment, an improved source could be used which would have coherence time twice the radiative lifetime, which could be around 2 ns.

[0069] Suitable polarisation combinations are any where the two photons have opposite polarisation, and are thus combinations of photons in different modes registered by detectors **27** with **31** and **29** with **33**, or combinations of photons in the same modes registered by detectors **27** with **29** and **31** with **33**. Since, to detect e.g. **27** and **29** requires 1 H and 1 V photon in mode e. This can be achieved by H in mode a reflected, and V in b transmitted, or V in a reflected and H in b transmitted. Superposition of these states, which have the same phase transformation through the beamsplitter, means the input is $(|H_aV_b\rangle+|V_aH_b\rangle)/\sqrt{2}$, i.e. another Bell state. Note that a joint measurement of a pair of oppositely polarised photons in the same mode e or d creates the output qubit state $\Phi_c=\alpha|1\rangle_c+\beta|0\rangle_c$, which requires only a bit-flip to recreate the input qubit.

[0070] Preferably, all four combinations referred to above of oppositely polarised photon detection in modes e or d are used to perform a joint measurement. However, any subset is also possible, albeit achieving teleportation with lower probability. In the experiments below only one of the combinations is selected, corresponding to 0 in mode e and 1 in mode d. In an embodiment, with ⅛ efficiency for each combination, and two Bell states to keep track of, the theoretical maximum Bell state measurement efficiency of 0.5 could be reached.

[0071] FIG. **3** shows a complete experimental teleportation system. Continuously electrically driven entangled light source **1** consists of d.c. current/voltage source **51** and entangled light emitting diode (ELED) **53**. ELED **53** comprises carbon p-doped top GaAs/AlAs Bragg reflector **55**, intrinsic doped GaAs cavity layer **57**, and silicon n-doped bottom GaAs/AlAs Bragg reflector **59**. Current is injected via contacts **61** and **63**, to create electrons and holes that relax into quantum-well-like InAs wetting layer **65**, and InAs quantum dot **67**. Quantum dot **67** generates entangled photons by radiative decay of the biexciton (XX) state consisting of two electrons and two holes, via the exciton state (X) consisting of one electron and hole, to the empty ground state. An aperture **69** in contact **61** allows photons to escape.

[0072] In an embodiment, the above ELED is based on self-assembled InAs quantum dots placed in the intrinsic region of a GaAs p-i-n junction grown by molecular beam epitaxy The relatively thick (~400 nm) intrinsic region suppresses charging of the X state and ensures that the neutral X and XX states are dominant. Two top and 14 bottom $Al_{0.98}Ga_{0.02}As/GaAs$ distributed Bragg mirror pairs create a planar 2λ optical cavity which enhances the light collection efficiency around the X and XX emission wavelengths. Mesas of size 360×360 μm^2 were etched to define individual ELED devices and a metal mask with ~2 μm diameter apertures placed on top of each device to isolate individual dots optically and serve as a p-type electrical contact.

[0073] The device was cooled to ~16K in a liquid helium cryostat and electrically excited by injecting a d.c. current density of 93 $nA\mu m^{-2}$, which gives roughly equal X and XX intensities

[0074] In an embodiment, in order for XX and X photons emitted in the same radiative cascade to be entangled in polarisation the fine structure splitting (FSS) of the intermediate X state of the quantum dot must be close to zero. The FSS for the dot used here was characterized by linear polarisation-dependent electroluminescence spectroscopy and found to be 2.0±0.2 μeV. The X and XX emission was verified to be unpolarised within error.

[0075] Photons emitted by the ELED are collected by lens **71** and coupled into single mode fibre **73**. XX and X photons are separated into different modes using wavelength dependent distribution unit **75**, which may be implemented using an optical grating. The two output modes are coupled to single mode fibres **77** and **79**, so that XX photons travel along fibre **77** towards Alice, and X photons travel down fibre **79** towards Bob.

[0076] In this experiment, the input qubit is generated by the same ELED as the entangled photons, though the input qubit may in general originate from another source. To achieve this, the stream of XX photons in fibre **77** is divided using 50/50 beamsplitter **81**, which is implemented using an evanescent fibre 2×2 coupler. On one splitter output **83**, the emission is polarised using fibre polarising beam splitter **85**, which also introduces a delay of ~2.5 ns. The input polarisation is then chosen using polarisation controller **87** to produce a photonic input qubit in mode a.

[0077] A pair of entangled photons travel in modes b and c, with the XX photon in mode b originating from the other output port **89** of beamsplitter **81**. Polarisation controller **91** corrects any transformation of the photon polarisation state caused by the fibres, so that a H polarised XX photon collected by lens **71** remains H polarised at the output mode d after 50/50 beamsplitter **23**. Similarly, polarisation controller **93** corrects polarisation transformation for the output of mode e.

[0078] Polarising fibre beamsplitters **95** and **97** perform polarisation discrimination for the joint measurement of a single photon in mode d and one in mode e by Alice. Superconducting single photon detectors (SSPDs) **99** and **101** detect H polarised photons in mode d, and V polarised photons in mode e respectively, performing a joint measurement to allow teleportation. The time between detection times of such photon pairs (the first delay time) is recorded and denoted τ_1.

[0079] Bob's apparatus consists of fibre delay **103**, to ensure the joint measurement is performed prior to detection of the output qubit, and polarisation controller **105**, to select the polarisation basis for Bob's measurement on the output qubit in mode c. For a P polarised input qubit, Bob's measurement basis is selected so that polarising fibre beamsplitter **107** can discriminate between polarisations P and Q, where Q is orthogonal to P. Bob uses avalanche photodiodes (APDs) **109** and **111** to detect single output photons, and their arrival time relative to a H polarised detection event recorded by Alice. This time difference, the second delay time, is denoted τ_2.

[0080] The choice of SSPDs to measure τ_1 is due to their fast response time, which enables precision sufficient to detect time differences within the coherence time of the emitted photons, which in this example is ~200 ps. Less precision is required in the measurement of τ_2, as strong entanglement is expected for X photons emitted up to 1 ns after a XX photon due to small FSS, slow re-excitation, and the XX and X radiative lifetime. Therefore APDs are suitable detectors for Bob, as though they are less precise, they have higher efficiency compared to SSPDs.

[0081] FIG. **4***a* shows experimentally measured XX and X photon statistics in fibres **77** and **79**. The measurement performed is a standard Hanbury Brown and Twiss type, from which we determine the second order correlation $g^{(2)}$, as a function of the time delay τ between detected photons. For X or XX photons emitted more than a few ns apart, $g^{(2)}$ is approximately 1, signifying a continuous stream of photons, with random time delay between emission of a pair of XX or X photons. However, around $\tau=0$, a clear anti-bunching dip is observed for both XX and X emission, signifying a suppressed probability of generating two XX or X photons at the same time. The residual $g^{(2)}$ at $\tau=0$ is approximately 0.1, and is attributed to the finite time resolution of our detectors, with the underlying $g^{(2)}$ expected to be close to zero.

[0082] In contrast, for a Poissonian light source such as a laser, or the individual output beams of a parametric down conversion (PDC) entangled light source, $g^{(2)}\tau=0$ is expected to remain 1, as indicated by the dashed line. PDC light sources are therefore not anti-bunched in contrast to the light sources used in systems in accordance with embodiments of the present invention.

[0083] FIGS. **5***a* and *b* show results demonstrating successful teleportation of photonic qubits using a system in accordance with an embodiment of the present invention.

[0084] Three-photon coincidences $g^{(2)}(\tau_1, \tau_2)$ are measured corresponding to photon detection by SSPD_H, SSPD_V, and APD_P or APD_Q, where τ_1 is the time difference between detection by SSPD_V and SSPDH, and τ_2 is the time difference between detection by SSPD_H and APD_P or APD_Q. The fidelity for teleportation of an input qubit with polarisation P onto an output qubit with ideal expected polarisation P' (orthogonal to Q') is given by;

$$f_P^T(\tau_1,\tau_2)=g_{P,P'}^{(3)}(\tau_1,\tau_2)/(g_{P,P'}^{(3)}(\tau_1,\tau_2)+g_{P,Q'}^{(3)}(\tau_1,\tau_2))$$

[0085] As the input polarisation P may be selected from an infinite set, an averaged teleportation fidelity measurement f^T is often used, where averaging is across 6 states from 3 mutually unbiased bases. In practice this translates to the averaging of the teleportation fidelity across 6 experiments, where the input state polarisations are horizontal (H), vertical (V), diagonal (D), anti-diagonal (A), left-hand circular (L) and right-hand circular (R). These states are shown in FIG. **4***b* on the Poincaré Sphere, which allows any polarisation state to be represented and visualised. In our examples the logical basis corresponds to the polar states H and V, and the states D, A, L, and R, which lie on the equatorial plane of the Poincaré sphere, are therefore logical superposition states and may be written as $D=(H+V)/\sqrt{2}$, $A=(H-V)/\sqrt{2}$, $L=(H+iV)/\sqrt{2}$ and $R=(H-iV)/\sqrt{2}$.

[0086] The average teleportation fidelity for the 6 input states is plotted in FIG. **5***a* as a function of the measured time delays τ_1 and τ_2. A high fidelity spot is observed at the centre of the figure, indicating that post selecting measurements based on the time delay between detected photons by Alice, and the time between a photon detected by Alice and Bob, is sufficient to allow teleportation. Note that the zero time on axis τ_2 is arbitrarily defined to correspond to photon events where the ancilla and target photons were generated at the same time. A large time difference may exist in an application between detection events, but this is manifested only by a linear shift of all data in τ_2, and a high fidelity spot can always be post-selected.

[0087] FIG. **5***b* shows the average teleportation fidelity plotted as a function of the second time delay, measured as the delay between Alice and Bob's photon detection τ_2, when $\tau_1=0$. This constitutes a vertical slice through the data in FIG. **5***a*. A peak is observed in teleportation fidelity when the correct value, zero, of τ_2 is selected. In this case, the peak average fidelity measured is 0.704 ± 0.016, which exceeds the maximum value of ⅔ achievable using non-entangled light sources by 2.2 standard deviations, thus proving quantum teleportation has occurred.

[0088] In the above embodiment, teleportation of single photonic qubits, mediated by individual pairs of entangled photons generated by an electrically driven entangled light source has been demonstrated. This is realized by embedding a single semiconductor quantum dot within a light-emitting-diode, resulting in anti-bunched emission and elimination of multi-photon errors. The average fidelity across 6 different input states is measured to be 0.70, above the classical limit. The unique single-photon nature of the photonic teleporter together with its electrical operation will help lift the complexity restriction of achievable future applications.

[0089] The systems described above provides full teleportation of arbitrary unknown photonic qubits which is ideally suited to distributed quantum computing and networking. However, until now such experiments have used light sources that produce random numbers of input-qubits. Teleportation has been demonstrated with a semiconductor quantum light source that emits no more than one photon or entangled photon pair simultaneously. The light source is electrically driven, which will offer a significant practical advantage when constructing complex quantum logic circuits.

[0090] Teleportation of optical qubits can allow consistent logic operations in massively parallel quantum computers, and the formation of secure quantum networks. Photon teleportation has previously employed laser-generated entangled photons created in random quantities. However, practical complexities of the generating scheme coupled with balancing the opposing requirements of high efficiency and infrequent multi-pair emission, restricts deployment in useful quantum information technology. Sources based on parametric down conversion within laser-excited non-linear crystals, are not directly electrically driven and exhibit no suppression in probability of generating two photons simultaneously in a given mode.

[0091] FIG. **6** shows a variation of the teleportation system. The embodiments described above post-select the time difference between output photons and the joint measurement τ_2 by measuring the output photons with a detector. However, said time-difference may also be selected by blocking photons using a blocking unit in the output beam except for those occurring at a desired time, wherein the desired time is when a photon which satisfies the second delay time limit is expected to arrive. This may be achieved by Alice communicating with, or controlling, photon blocking unit **121**, causing photons to be blocked except at times corresponding to detection of a simultaneously generated first or ancilla photon by Alice. For example, if the time delay for an ancilla photon to travel from the source to Alice's detectors was Δt_A, and the delay between the source and the blocking device Δt_B, then upon a successful joint detection event by Alice with $\tau_1=0$, Alice would cause the blocking device to open at time approximately $\Delta t_B-\Delta t_A$ later, for a short amount of time. The time that the blocking device remains open should for example be similar to the width of the high fidelity region in FIG. **5***b*, which is the order of 1 ns for the prototype source.

[0092] In an embodiment, only photons with high teleportation fidelity post-selected by Alice will be allowed to pass. Such photons could be measured immediately, or transferred to another party or application. In one embodiment, the blocking unit will only allow the passage of photons at the desired time if Alice has already made a joint measurement and the results from the first delay time suggests that teleportation may have occurred. In further embodiments, Alice will only instruct the blocking unit to allow the passage of photons if her measurements of the polarisation of the photons which have undergone joint measurement suggest that teleportation has occurred.

[0093] The blocking device could be constructed from a variety of standard components, including an electro-optic-modulator or interferometer.

[0094] While certain embodiments have been described, these embodiments have been presented by way of example only, and are not intended to limit the scope of the inventions. Indeed the novel methods and apparatus described herein may be embodied in a variety of other forms; furthermore, various omissions, substitutions and changes in the form of methods and apparatus described herein may be made without departing from the spirit of the inventions. The accompanying claims and their equivalents are intended to cover such forms of modifications as would fall within the scope and spirit of the inventions.

1. A system for quantum teleportation of a quantum state of an input photon, the system comprising:

a light emitting diode configured to produce a polarisation entangled photon pair;

a beam splitter configured to direct one photon of an entangled photon pair along a first path and the other photon of said entangled pair along a second path;

an input for said input photon;

a measurement unit for performing a joint measurement on the input photon with one of the photons of an entangled photon pair directed along the first path, the measurement unit comprising a first detector unit for detecting two photons upon which a joint measurement has been performed;

a second detector unit configured to detect the photon from said entangled photon pair directed along the second path;

a timing unit is configured to measure a first delay, said first delay being the delay between the input photon and the photon of the entangled photon pair at the point of maximum indistinguishability of the photons as they pass through the joint measurement unit, the timing unit being further configured to measure a second delay, said second delay being the delay time between the two photons of the entangled photon pair as they exit the light emitting diode; and

a controller configured to determine that a teleportation measurement is valid if the first delay is within a first predetermined timing window and the second delay is within a second predetermined timing window.

2. A system according to claim **1**, further comprising an electrical source for said light emitting diode and wherein said electrical source is a D.C. source.

3. A system according to claim **1**, wherein said timing unit is configured to determine the first delay time from the detection time of the said two photons by the first detector unit, the timing unit also being configured to compensate for variations between the path lengths from the point where the respective spatial modes of the two photons meet in the joint measurement unit to the first detector unit.

4. A system according to claim **1**, wherein the joint measurement unit comprises a beam splitter to permit two-photon-interference and wherein said timing unit is configured to determine the first delay time from the detection time of the said two photons by the first detector unit, the timing unit also being configured to compensate for variations between the path length taken by the two photons from the beam splitter to the first detector unit.

5. A system according to claim **1**, wherein the timing unit is configured to measure the second delay from the time when the photon which follows the second path is received at the second detector, the timing unit being further configured to compensate for differences in path lengths taken by the two photons of the entangled photon pairs.

6. A system according to claim **1**, wherein the system further comprises a blocking unit located in the second path, said timing unit being configured to operate said blocking unit to allow the transmission of a photon along the second path if the second delay time is within the second timing window.

7. A system according to claim **1**, further comprising a blocking unit located in the second path, said timing unit being configured to operate said blocking unit to block the transmission of a photon along the second path unless said controller has determined that the first delay time is within the first timing window.

8. A system according to claim **1**, wherein the joint measurement is a measurement involving Bell states or mixtures with Bell states.

9. A system according to claim **1**, further comprising a state measurement unit configured to perform state measurements on both of the photons which pass through the joint measurement unit and wherein the controller determines that the teleportation measurement is valid if, additionally the state measurements on both photons agree with at least one of a predetermined set of results.

10. A system according to claim **1**, wherein the controller is configured to allow post measurement selection of valid measurements.

11. A system according to claim **1**, wherein the first detector unit comprises first and second detectors, the first and second detectors being superconducting detectors.

12. A system according to claim **1**, wherein said second timing window is from 0 to t_{2max}, where t_{2max} is of the order of the exciton radiative lifetime.

13. A system according to claim **1**, wherein the first timing window is from 0 to t_{1max}, where t_{1max} is the coherence time of the photon which follows the first path.

14. A system according to claim **1**, wherein the system is configured to allow the light emitting diode to provide the input photon.

15. A system according to claim **14**, further comprising a input photon delay unit and a switch, said switch being configured to direct a photon from the first path into the said input photon delay unit, wherein in the input photon delay unit, the photon from the said source is delayed to coincide with a further photon output by the source such that the photon which is delayed becomes the input photon.

16. A system according to claim **1**, wherein the state to be teleported is a superposition state.

17. A system according to claim **1**, wherein said light emitting diode comprises a quantum dot.

18. A quantum computer comprising a teleportation system according to claim **1**.

19. A quantum communication relay comprising a teleportation system according to claim **1**.

20. A method of teleporting of a quantum state of an input photon, the method comprising:

providing a polarisation entangled photon pair from a light emitting diode;

directing one photon of the entangled photon pair along a first path and the other photon of said entangled pair along a second path;

providing an input photon;

performing a joint measurement on the input photon with one of the photons of an entangled photon pair directed along the first path and detecting the two photons upon which a joint measurement has been performed;

detecting the photon from said entangled photon pair directed along the second path;

measuring a first delay, said first delay being the magnitude of the delay between the input photon and the photon of the entangled photon pair at the point of maximum indistinguishability of the photons as they undergo joint measurement,

measuring a second delay, said second delay being the magnitude of the delay time between the two photons of the entangled photon pair as they exit the light emitting diode; and

determining that a teleportation measurement is valid if the first delay is within a first timing predetermined window and the second delay is within a second predetermined timing window.

* * * * *

Museum Menu Teralab Main Menu

Virtual Science Museum

Radio

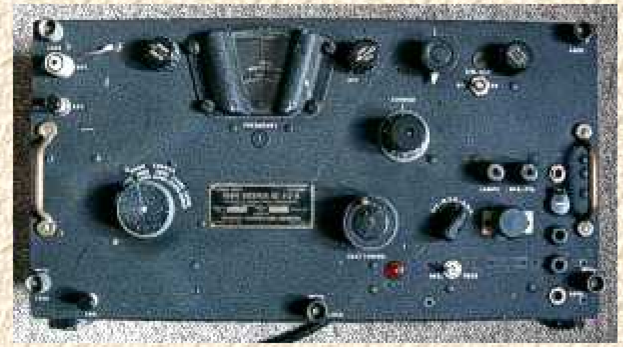

This is a U.S. Army transceiver model BC-312-N, made by Farnsworth Television and Radio Corporation. It uses valves throughout and covers the short wave bands.

Click the image for a bigger picture.

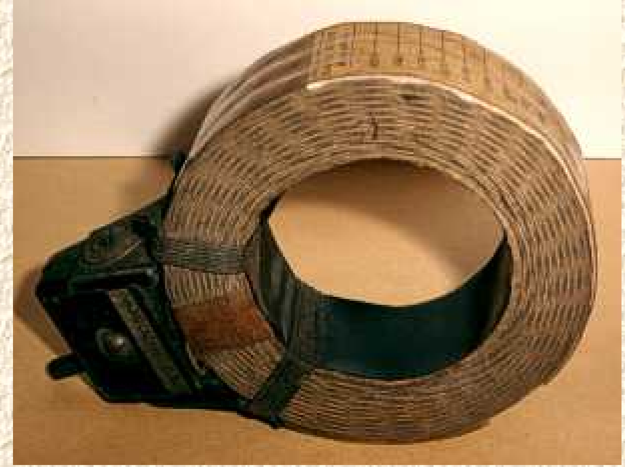

This is a nicely made plug-in tuning coil from an early radio receiver. On the label is printed, 'De Forest 300 Pat 141344 What are the wild waves saying'. Lee de Forest invented the Audion triode in 1907.

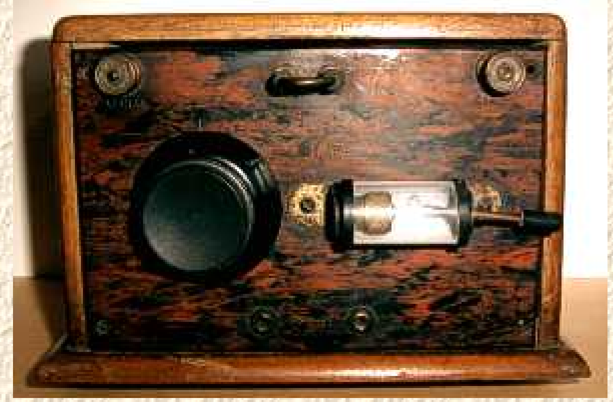

This is a rather nice little crystal set with a germanium crystal and cat's whisker.

The crystal is inside the glass tube. The handle adjusts the cats whisker, which is a piece of wire, sprung against the crystal. With careful adjustment, a 'sweet spot' can be found. The crystal then acts as a diode detector, demodulating the radio signal.

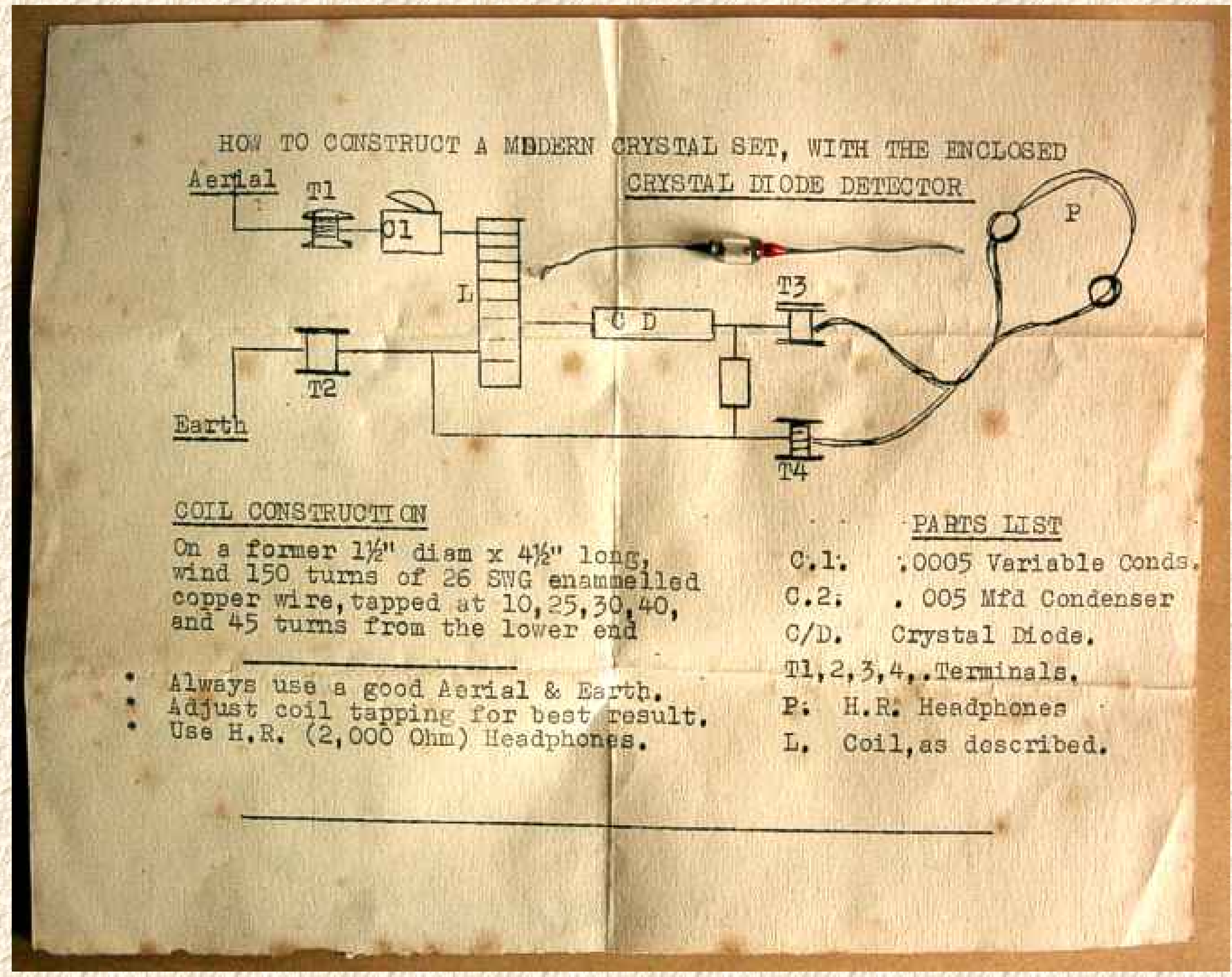
Aerial
T1
C1
L
C D
T3
P
T2
Earth
T4

HOW TO CONSTRUCT A MODERN CRYSTAL SET, WITH THE ENCLOSED CRYSTAL DIODE DETECTOR

COIL CONSTRUCTION

On a former 1½" diam x 4½" long, wind 150 turns of 26 SWG enammelled copper wire, tapped at 10, 25, 30, 40, and 45 turns from the lower end

- Always use a good Aerial & Earth.
- Adjust coil tapping for best result.
- Use H.R. (2,000 Ohm) Headphones.

PARTS LIST

C.1.	.0005 Variable Conds.
C.2.	. 005 Mfd Condenser
C/D.	Crystal Diode.
T1,2,3,4,.	Terminals.
P.	H.R. Headphones
L.	Coil, as described.

Museum Menu Teralab Main Menu

Beaver Baby Grand
Beaver Machine and Tool Company produced a number of versions of this crystal radio during the 1920s, and also apparently supplied the radio to India Ivory Company. One of the smallest radios at the time at 3-3/4 x 2 x 3/4 inches, it was advertised as The Beaver Baby Grand Vest Pocket Radio Receiving Set.

These are shown here in the order they may have been produced based on a progression of part upgrades.

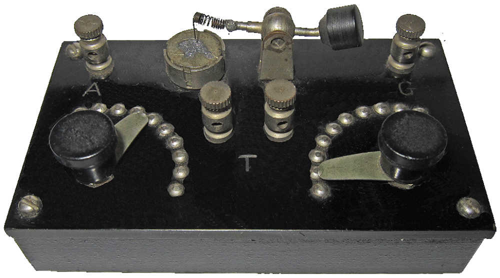

Version 1
Black painted wood case, 4-slot crystal holder, flat top knobs, small binding posts, simple engraved panel, and no ID of any kind.

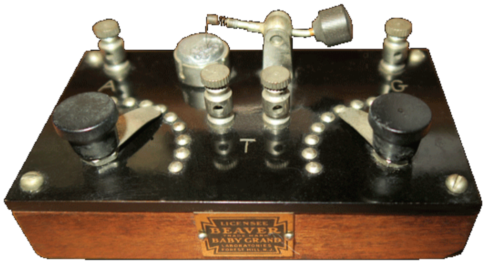
A
G
T
LICENSEE
BEAVER
BABY GRAND

Version 2
Reddish brown colored wood, 4-slot crystal holder, flat top knobs, small binding posts, simple engraved panel, metal tag on cabinet.

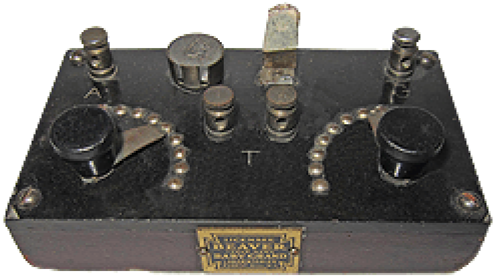
A
G
T
BEAVER
BABY GRAND

Version 3
Reddish brown colored wood, 2-slot crystal holder, flat top knobs, small binding posts, simple engraved panel, metal tag on cabinet. (detector arm is missing)

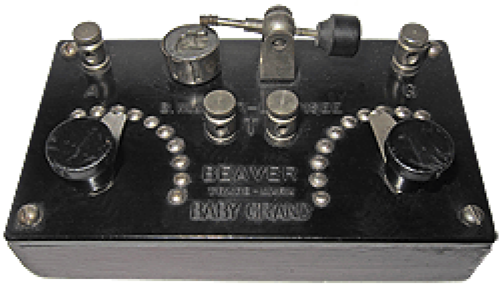
A
G
T
BEAVER
TRADE-MARK
BABY GRAND

Version 4
Reddish brown colored wood, flat top knobs, small binding posts, raised lettering on panel, no metal tag.

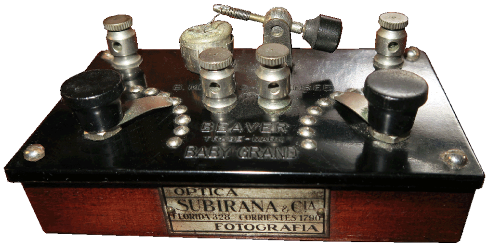
BEAVER
TRADE-MARK
BABY GRAND
OPTICA
SUBIRANA & CIA
FLORIDA 328 CORRIENTES 1290
FOTOGRAFIA

Version 5
Reddish brown colored wood, flat top knobs, large binding posts, raised lettering on panel, Subirana metal tag.

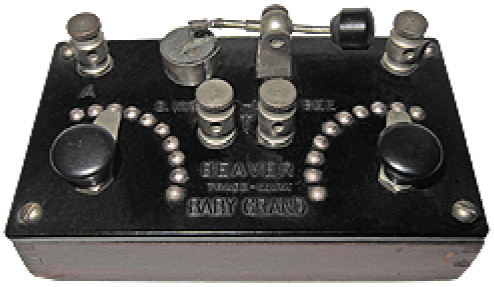
A
T
BEAVER
TRADE-MARK
BABY GRAND

Version 6
Reddish brown colored wood, dome top knobs, large binding posts, raised lettering on panel, no metal tag.

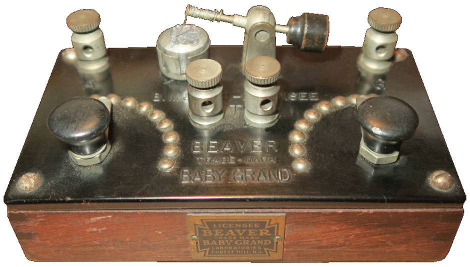

Version 7
Reddish brown colored wood, dome top knobs, large binding posts, raised lettering on panel, metal tag on cabinet.

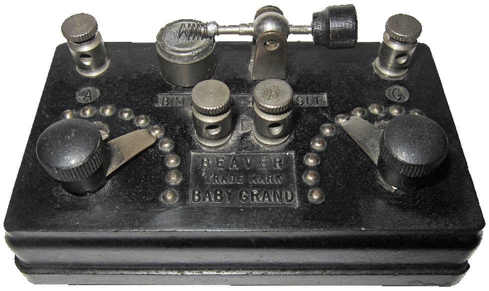

Version 8
Hard rubber cabinet, ribbed knobs, large binding posts, lettering molded into panel. A cover on the bottom protects the coil inside.

Bethlehem Radio Corporation, of Bethlehem, Pennsylvania, made these two metal crystal radios during the 1920s, and may have also made other radio equipment under the BERACO name.

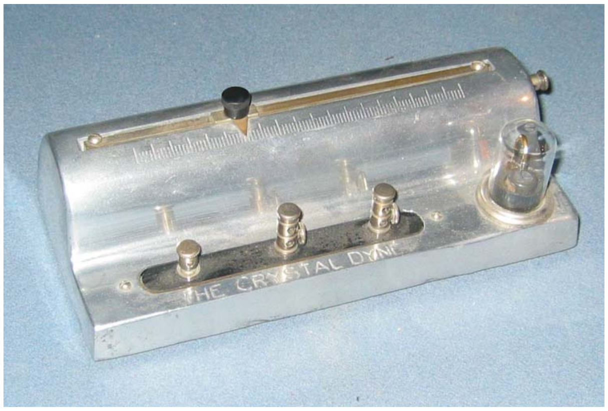

Bethlehem 'The Crystal Dyne'
It has a cast aluminum case, a glass-enclosed crystal detector, and a slide tuner. It is about 8-1/2 inches long x 4 inches deep.

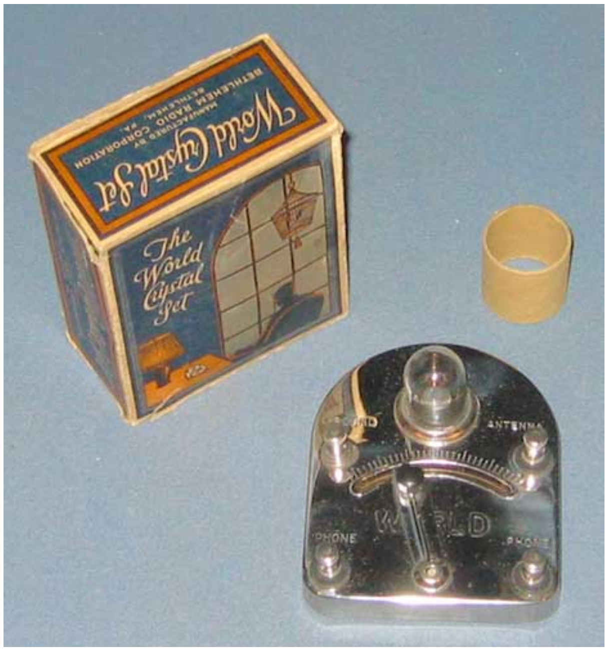

Bethlehem 'World Crystal Set'
It has a nickle-plated metal case, a glass-enclosed crystal detector, and a slide tuner. It measures about 5 x 4 inches. It is shown with its original box. These were also sold painted in various colors including red and green.

BMS Pandora Crystal Set
The Brooklyn Metal Stamping Company, of Brooklyn, New York, manufactured a number of crystal radios in the 1920s, including the 'Pandora' which came in two styles and 4 or 5 colors. This example has the coil raised up on a small pedestal and has the original Russett Bronze paint. It is in mint condition except the detector arm is missing. It is shown with its original instruction sheet and box.

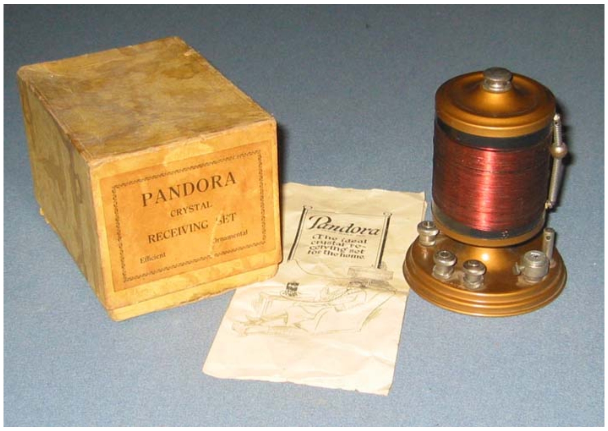

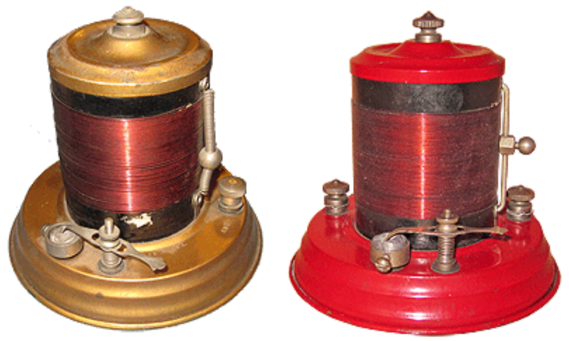

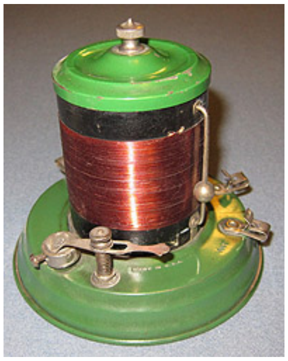

BMS Pandora Crystal Sets
These BMS Pandora sets are styled different from the one above, without the pedestal. Note that the red and green examples have a solid slider bar while the bronze one has a spring, the bronze and green examples have a ball slider while the red one has a slider with a knob.

Also, the green example has fahenstock clip connectors while the others have binding posts. The Russett Bronze, red, and Medium Green colors are all original paint.

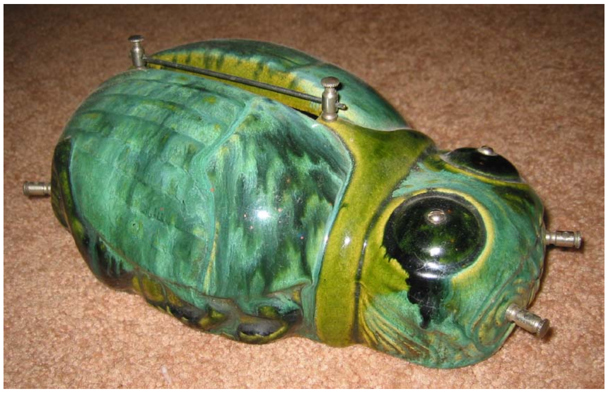

Brush-McCoy 'Mystic'
In 1927 the Brush-McCoy Pottery Company decided to produce a unique pottery item that would appeal to radio fans all over the world, who were then called "Radio Bugs." It was a ceramic bug 9" long containing a crystal radio set. These were called "MYSTIC" Radio Sets. The bug itself is pottery. An article that was in The Zanesville Signal said, "The insulating properties of this material are so high that exceptional reception results."

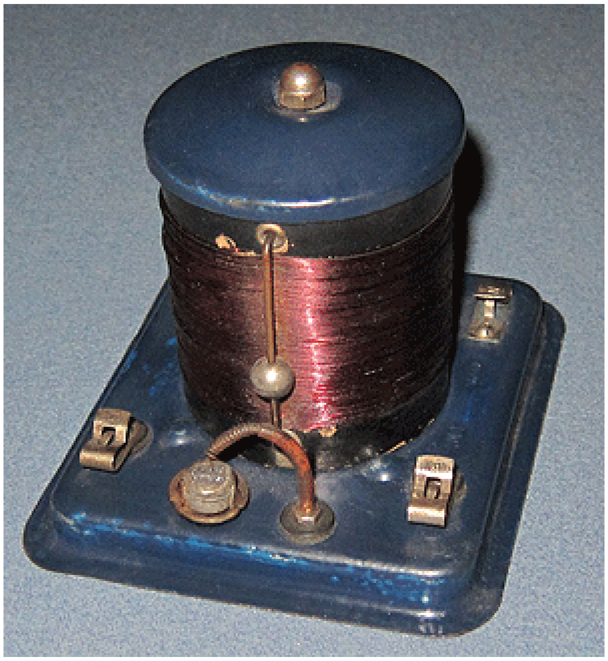

Cannon Ball 'Baby Grand'
1920s crystal radio with metal base.

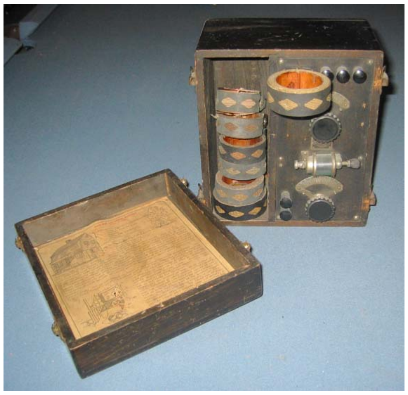

DeForest Everyman
This Deforest Everyman Radiophone Receiver, Type DT-600, was made in the 1920s and used different plug-in coils to receive different wavelengths. It has a glass-enclosed crystal detector.

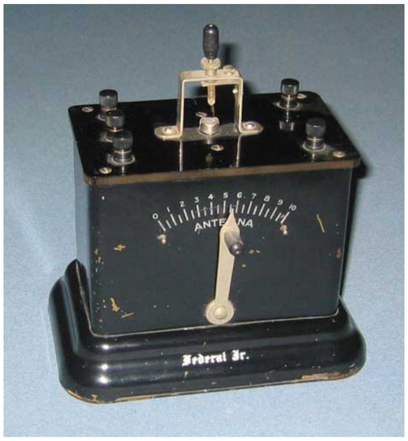

Federal 'Junior'
This Federal Jr. crystal radio was made in the 1920s. The cabinet is metal with a bakelite top. A slider control on one side is labeled 'Antenna', and a similar control on the other side is labeled 'Detector'.

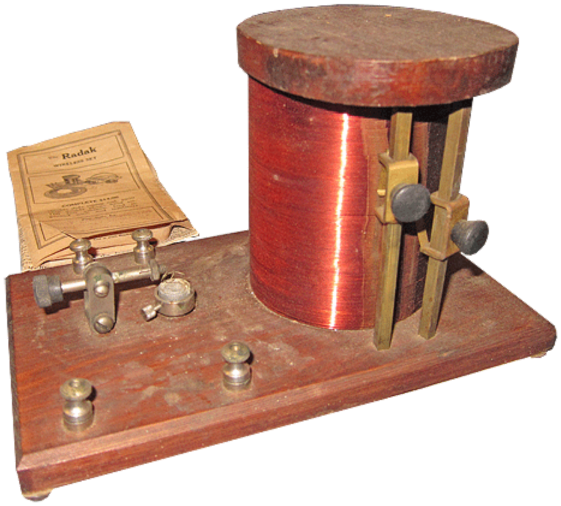

Fletcher 'Radak Wireless Set'
Made and distributed by Fletcher Electric Company at 34 West 36th Street in New York City in the early 1920s. The original operating instructions call this 'The Hand Camera of Wireless! A snappy little Radio Set! Small and light as a postcard camera'.

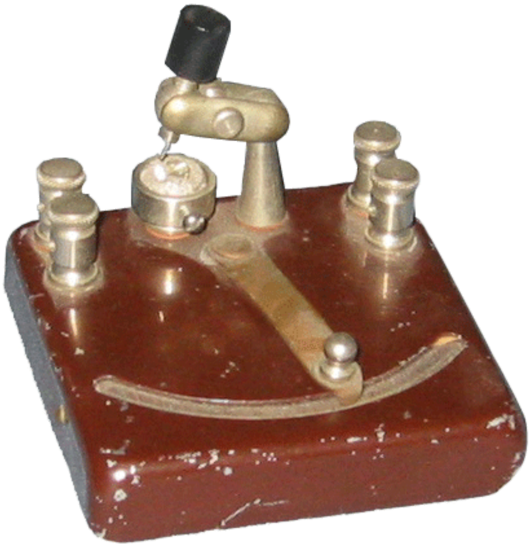

Flivver
This little Flivver crystal radio was made in the 1920s. It has a metal cabinet and is identified as a Flivver on the underside.

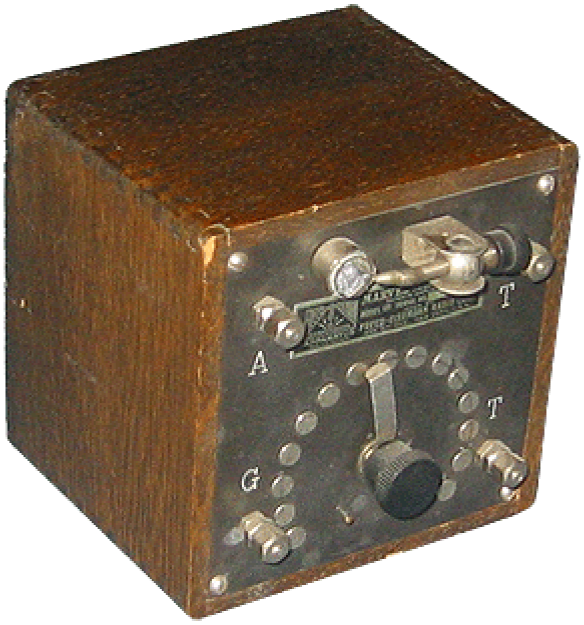

Freed Eisemann 'Marvel 101'
Freed Eisemann manufactured this crystal radio in the 1920s and housed it in a nice oak dovetailed cabinet.

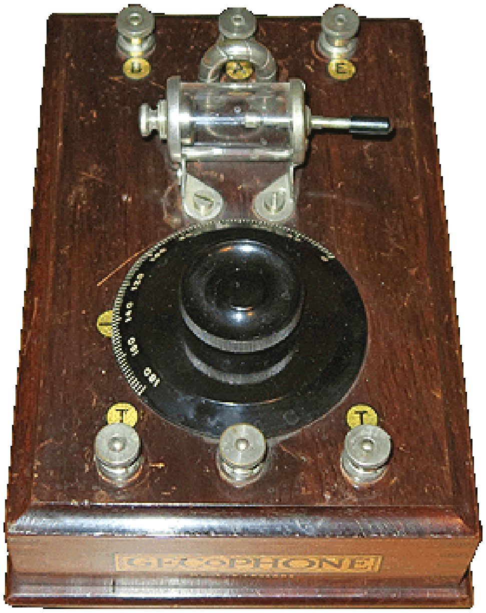

Gecophone Junior
1920s British crystal radio

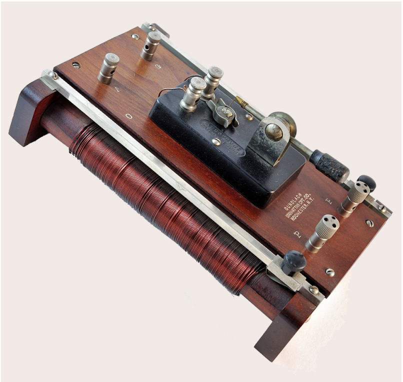

Gundlach - Manhattan Optical Company

1920s crystal radio

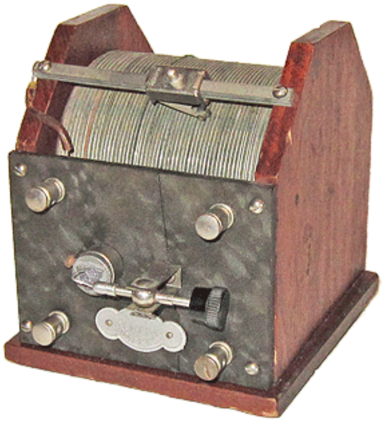

Gillet Marvel
1920s crystal radio

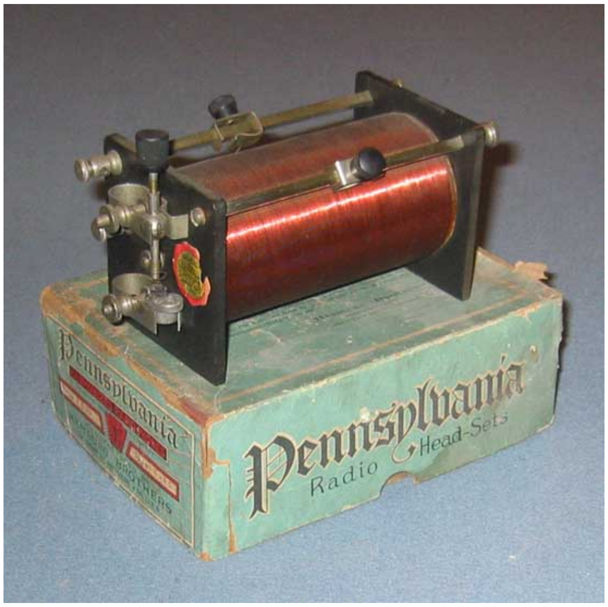

Hamburg Brothers - Pennsylvania
This little crystal radio was sold about 1926 by [Hamburg Brothers](#) of Pittsburgh, Pennsylvania, and was endorsed by Radio News magazine. It is sitting on a box of Hamburg Bros. 'Pennsylvania' headphones.

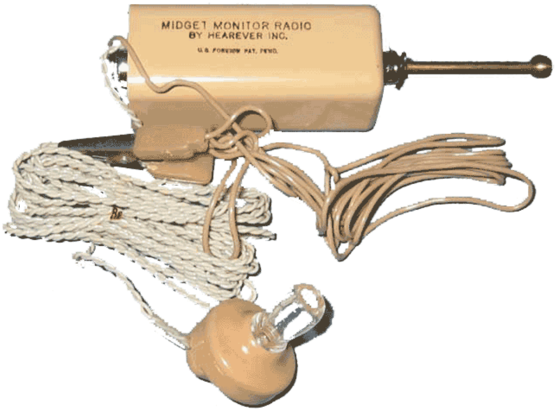

Hearever 'Midget Monitor'
This tiny crystal set was made in the 1950s or 60s and is built into a standard radio IF coil can. Sliding the antenna rod up and down tunes a station. This radio is mint in the original box, and is marked 'MIDGET MONITOR RADIO BY HEAREVER INC.' The instruction sheet gives the company address as: MIDGET MONITOR RADIO, Dept. 150, 623 Marquette Ave., Minneapolis 1, Minn.

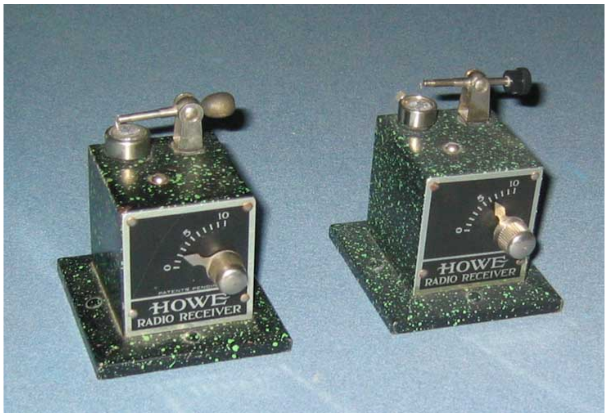

Howe: Early, Late
Two 1920s Howe crystal radios. The left one is the earlier version - it has 'Patents Pending' printed on the front plate, the detector arm has a wooden handle, and the crystal cup is larger.

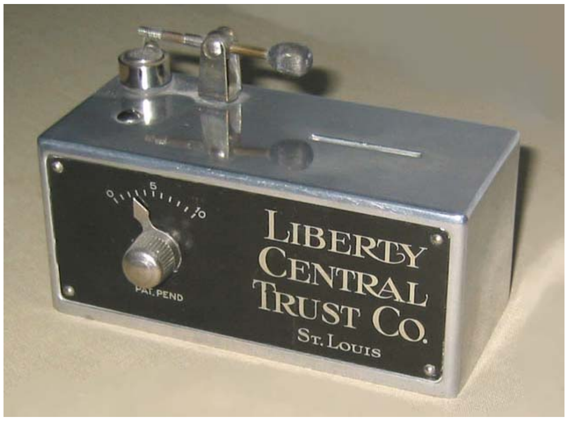

Howe - Liberty Central Trust Co.
This is a bank advertising the Liberty Central Trust Company of St. Louis Missouri, and it has a Howe crystal radio built into it!

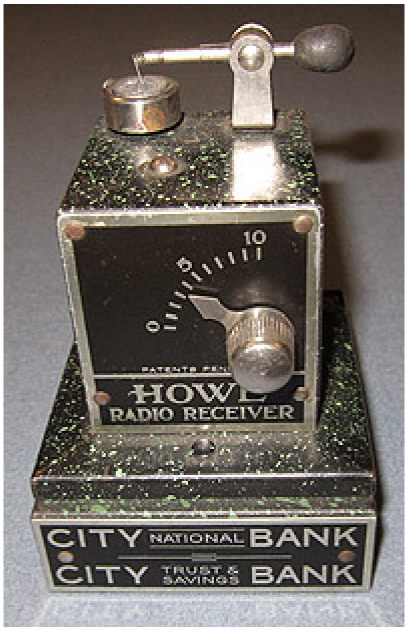

Howe - National City Bank
This is a bank advertising the National City Bank, and it has a Howe crystal radio built into it!

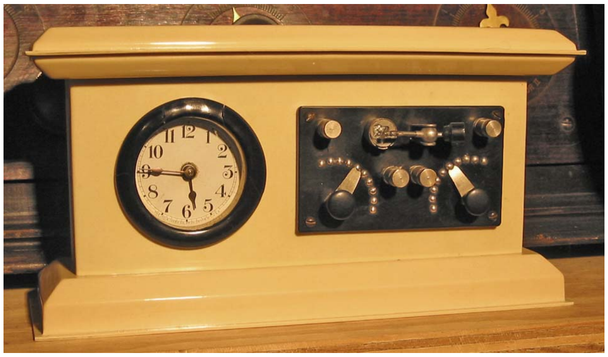

India Ivory
A rare 1920's crystal radio, only documented with a brief reference in the Maurice Siever's Crystal Set books, and a photo in Morgan McMahon's Vintage Radio book, and 1 of only 3 I've seen in collections. This off-white celluloid clock case is inscribed 'Radio Clock Case, India Ivory Co., Pat. Apl'lied For'. The paper face of the wind-up clock is marked 'India Ivory Company, Providence, R.I., U.S.A.' The radio itself looks exactly like an early version of the 1920's Beaver Baby Grand chassis.

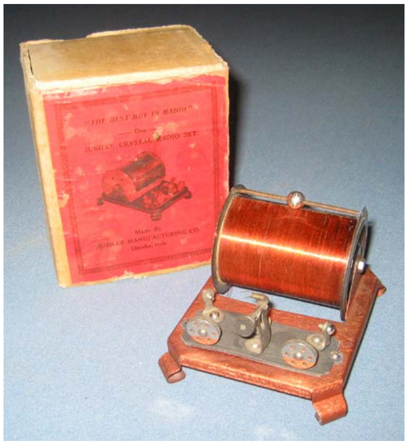

Jubilee
Jubilee Manufacturing Company, which operated in Omaha, Nebraska from 1905 until about 1988, produced this Jubilee Crystal Radio Set in the later 1920's. The red label on the original box proclaims this to be 'The Best Buy in Radio'. Unique connectors allowed up to six pair of headphones to easily be connected to the radio!

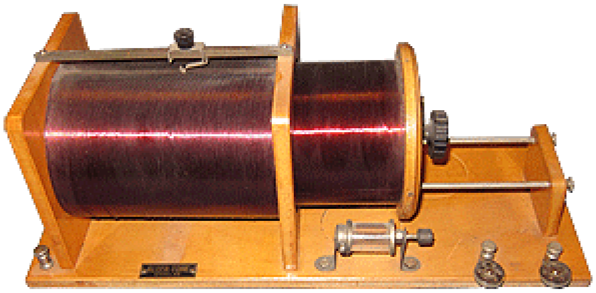

Kleer-Tone Model 100

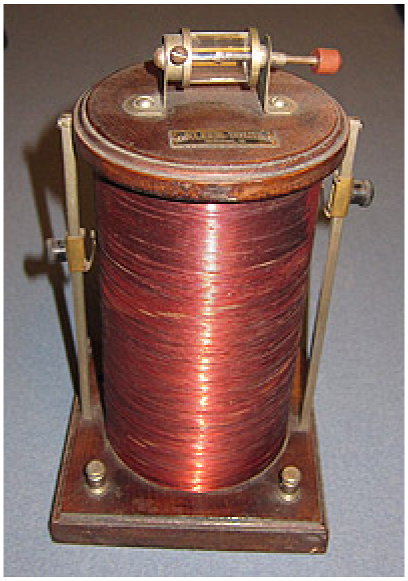

Kleer-Tone Model B

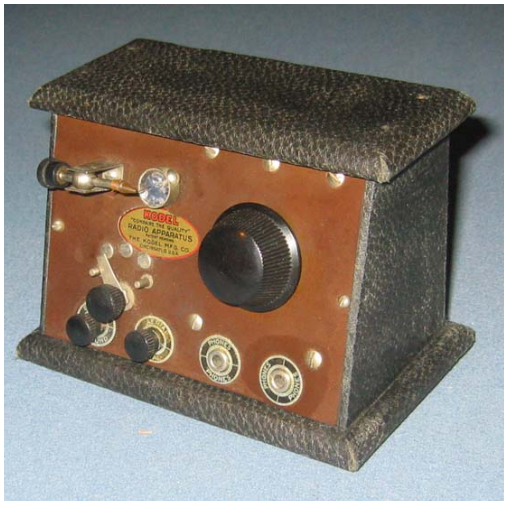

Kodel Model S-1
Kodel's smallest and lowest priced radio, this originally sold for $5.

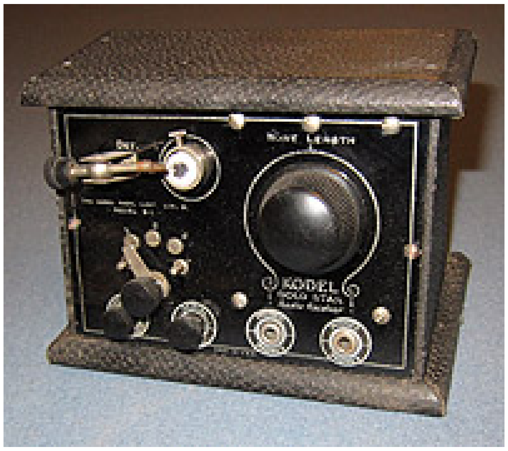

Kodel Gold Star Model S-1
Kodel's Gold Star model S-1 is identical to the other version except for the front panel design.

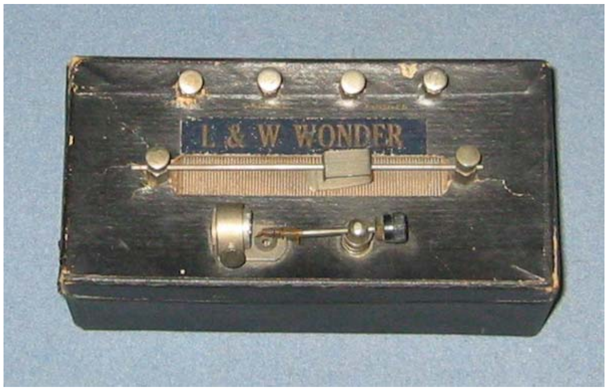

L&W Wonder
A cheaply made crystal radio in a cardboard box cabinet.

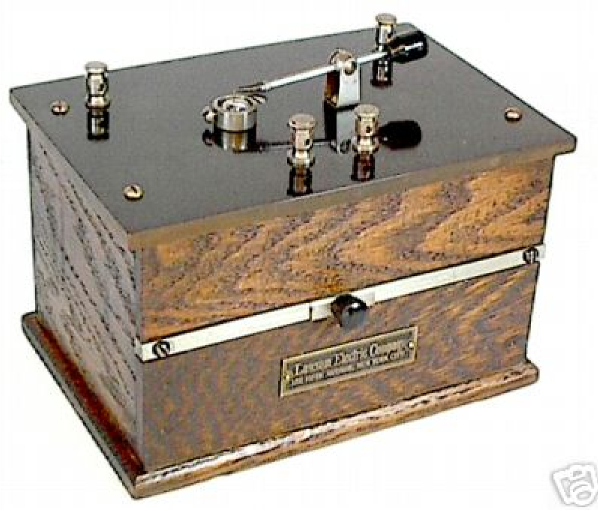

Lawsam Electric Company
This 1920s crystal radio was manufactred by Lawsam Electric Company at 122 Fifth Avenue, in New York City. The recessed slider on the front of the cabinet is unique.

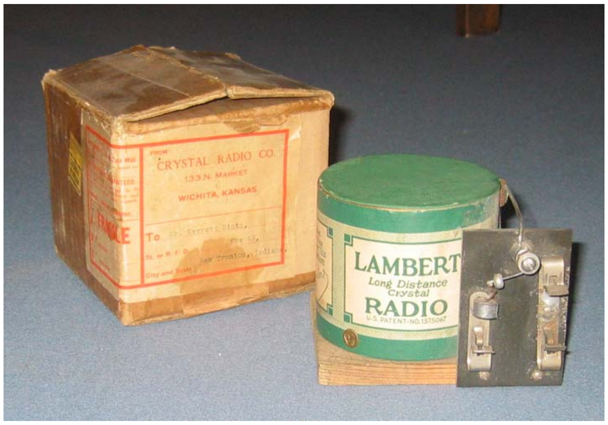

Leon Lambert
This old crystal radio is shown with its original box. It has a very crudely designed detector that looks homemade, but this is original from the factory. It is identical to the wooden version, except with a less expensive cabinet.

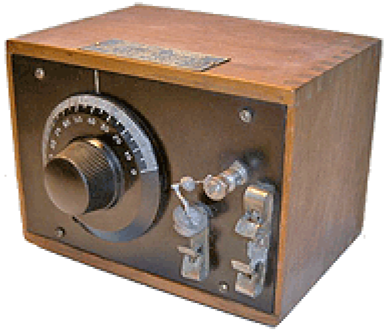

Leon Lambert
This crystal radio has a wood cabinet with a brass tag on top saying 'Lambert Radio, Leon Lambert Mfg. Co., Wichta, Kansas'. It is the deluxe version of the green Leon Lambert radio.

Lombard Multiphone
An unusual old crystal radio in a masonite cabinet painted crackle green with cowboys on the sides. A Lombard detector is mounted below the tuning dial and there are connections for six pairs of headphones! Operating instructions are on the back side and earphone instructions are on the bottom of the cabinet.

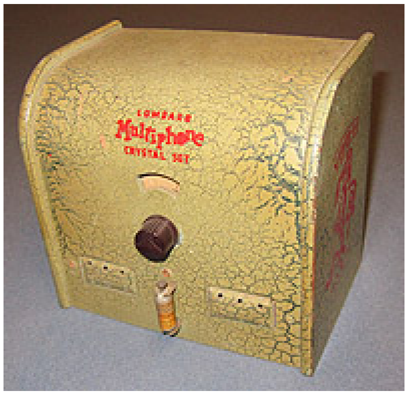
LOMBARD
Multiphone
CRYSTAL SET

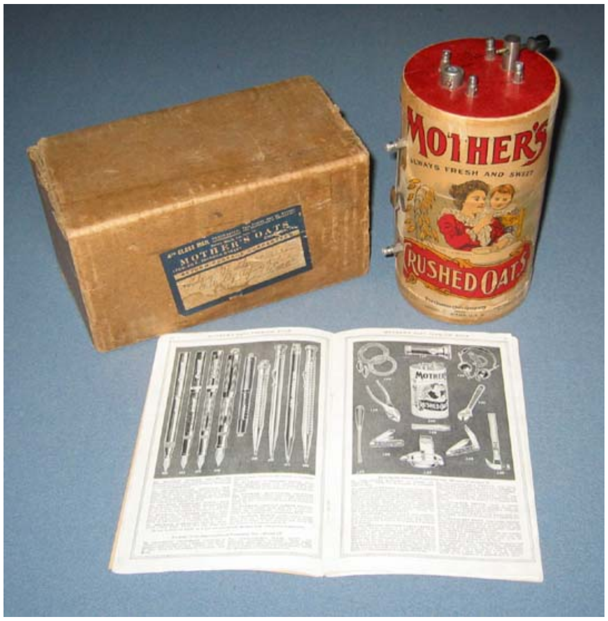
MOTHER'S
ALWAYS FRESH AND SWEET
CRUSHED OATS
MOTHER'S OATS

Marquette 'Mother's Oats'
This 1924 Mother's Oats premium catalog offered this crystal radio for 110 coupons, or 5 coupons and $1.00. The radio was made by Marquette Radio Corporation. It is shown with its original box.

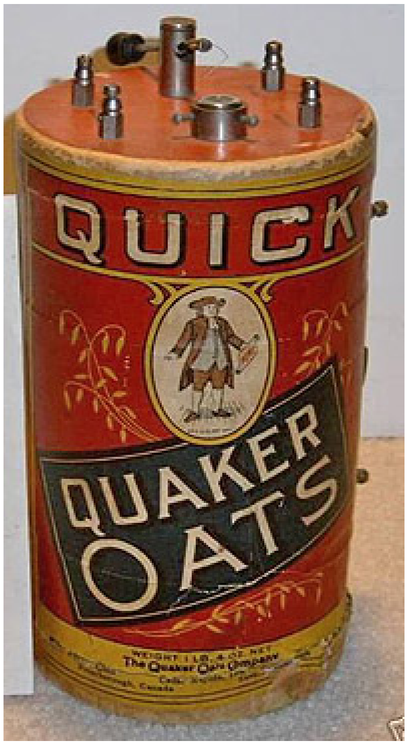
QUICK
QUAKER
OATS
The Quaker Oats Company

Marquette 'Quaker Oats'
This early 1920s radio is very similar to the Mothers Oats set. It was also made by Marquette Radio Corporation and was a coupon premium.

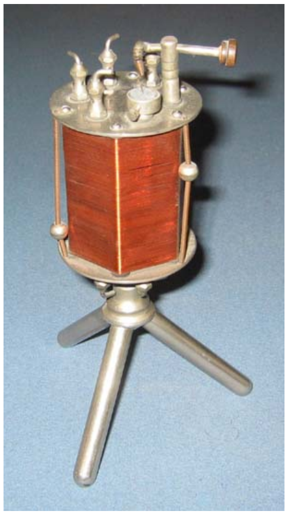

Martian 'Big 4'
This is a unique and well-made crystal radio mounted on a tripod base. It has two sliders, a crystal detector,

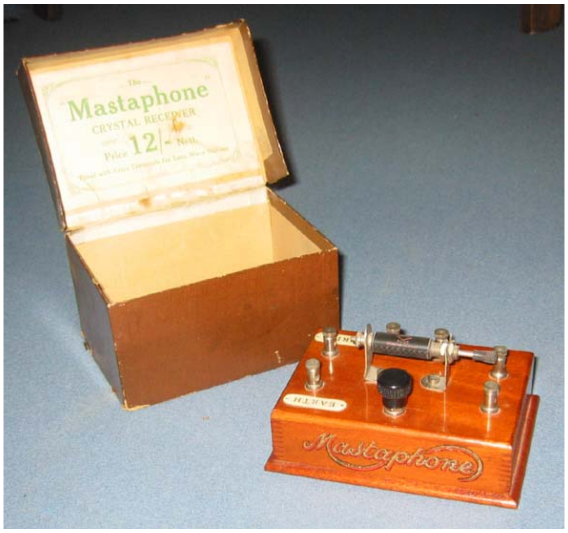
The
Mastaphone
CRYSTAL RECEIVER
Price 12/- Nett
EARTH
Mastaphone

Mastaphone
This crystal radio was made in Great Britain during the

and enough binding posts to connect four sets of headphones. It was made in Newark, New Jersey, and stands 9 inches tall.

1920s, and is shown with its original box. It has a spring-loaded enclosed cats whisker detector.

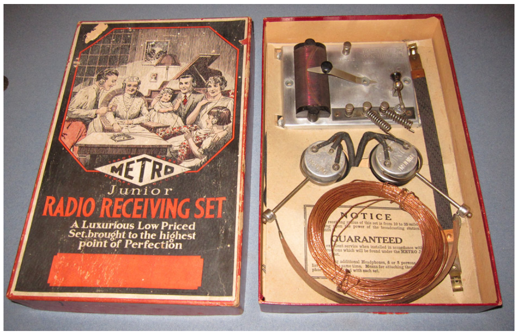
METRO
Junior
RADIO RECEIVING SET
A Luxurious Low Priced Set brought to the highest point of Perfection
NOTICE
GUARANTEED

Metro Junior
A 1920s crystal radio made by Metro Electrical Company in Newark, New Jersey. The same radio was also marketed under other names (see Pal). This is shown in its original box.
The base is aluminum and measures 6 x 5 x 1/2 inch.

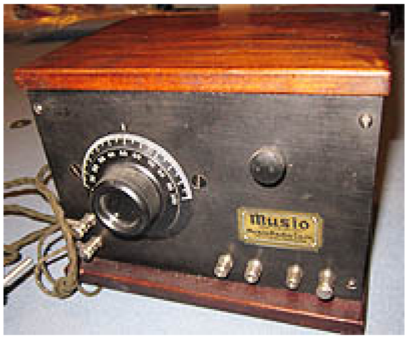

Musio
A 1920s crystal radio made by Musio Radio Company. The small knob controls the internal catwhisker. The lid is held on only by spring clips so is often missing. A pair of Musio headphones completes the set.
It measures 8-1/4 x 6-1/4 x 6 inches.

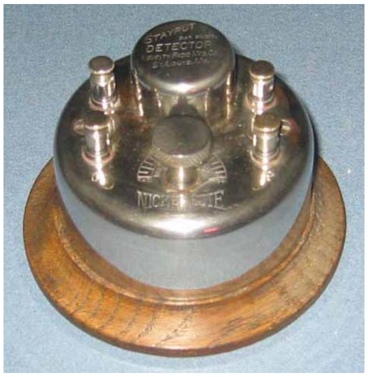
STAYPUT DETECTOR
NICKELETTE

Novelty Radio Mfg. Co. 'Nickelette'
A 1920s crystal radio marked 'Trade Mark NICKELETTE'. The detector is under the metal cover marked 'STAYPUT DETECTOR, NOVELTY RADIO MFG. CO., ST. LOUIS, MO, PAT PND'G'. You need to remove the metal cover to adjust the wire cats whisker. Directions are on a paper label on the bottom of the radio.

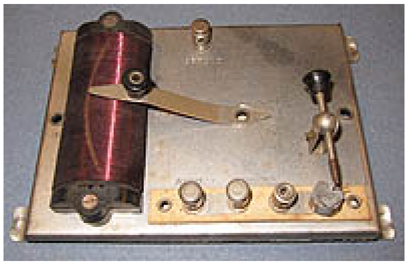

Pal
A 1920s crystal radio made by Pal Radio Company in Jersey City, New Jersey. The same radio was also marketed under other names (see Metro).
The base is aluminum, has a cardboard instruction card riveted underneath, and measures 6 x 5 x 1/2 inch.

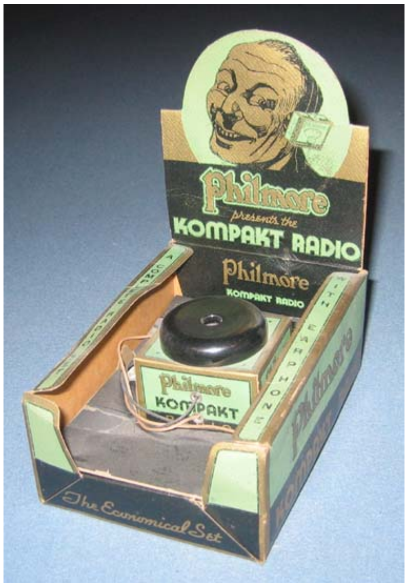
Philmore
presents the
KOMPAKT RADIO
Philmore
KOMPAKT RADIO
Philmore
KOMPAKT
The Economical Set

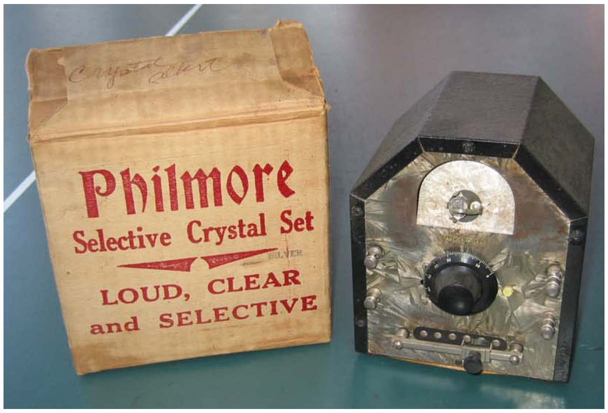
Crystal Select
Philmore
Selective Crystal Set
SILVER
LOUD, CLEAR
and SELECTIVE

Philmore Selective
One of Philmore's earliest models. These came in various colors - the box is stamped 'SILVER', indicating the color of the front panel on this one. It is shown with its original box.

Philmore Kompakt
Philmore made many types of crystal radios, and this is one of their more unusual models. The radio is a small square box with an earpiece on it, a tuning knob, antenna and ground wires, and with a detachable headband that holds the unit against your ear. It is shown with its original counter display box.

VEST-O-FONE
VEST POCKET RADIO
Pocket Radio Corporation
DES MOINES, IOWA

Pocket Radio: Vest-O-Fone
This is a crystal radio made by Pocket Radio Corporation at 720 Grand Avenue in Des Moines, Iowa, in the 1920s. It is basically a single headphone with a fixed cats-whisker crystal detector built into the headphone. It is shown with its original box.

POLLAK'S
Rocket
RADIO
NO
tubes
batteries
electric

Pollak's Rocket Radio
Marked on the box as 'Made in Japan for Pollak's, San Francisco 3, Calif.' This is a cheap little plastic set with an antenna that pulls out to tune it.

Radio Maintenance Company
A 1920s crystal radio made by Radio Maintenance Company (R.M.C.) in Minneapolis, at 31st and Snelling Avenue.

Radioette
A cheaply made crystal radio in a cardboard box cabinet.

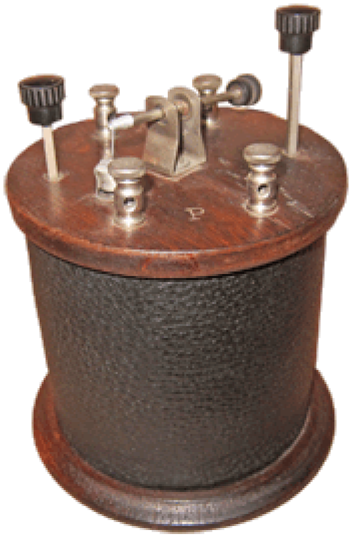

MINNEAPOLIS
MINN.

Rippner 'Monarch Radio Set'
A 1920s crystal radio made by The Rippner Bros. Mfg.

Co., Cleveland, Ohio. This one has the rarer deep blue color can.

RAO 'Double-Slider'
A 1920s crystal radio made by RAO Manufacturing Company in Minneapolis. Two knobs on top move up and down to tune hidden slider contacts inside the coil. 'RAO' is stamped into the wood on the bottom. RAO Manufacturing Company still exists and is now located in Fridley, Minnesota.

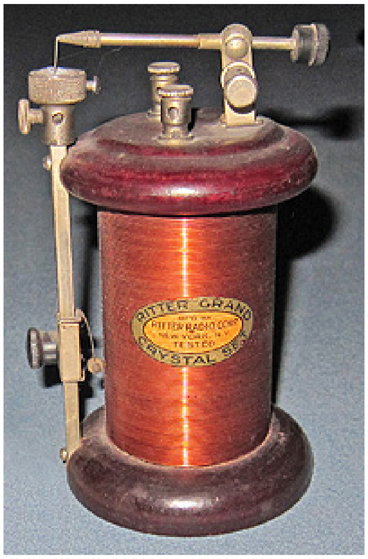

Ritter Grand Crystal Set
A 1920s crystal radio made by Ritter Radio Corporation at 232 Canal Street in New York City.

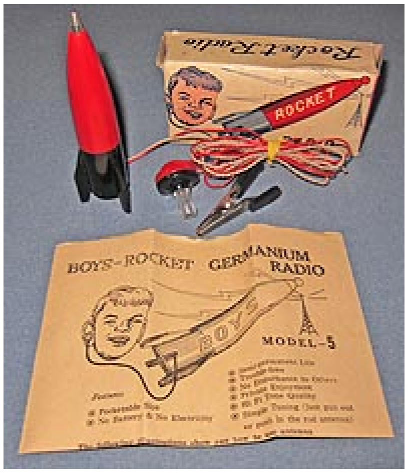

Rocket Radio - Model 5
Marked on the box as 'Made in Japan'. This is a cheap little plastic set with an antenna that pulls out to tune it.

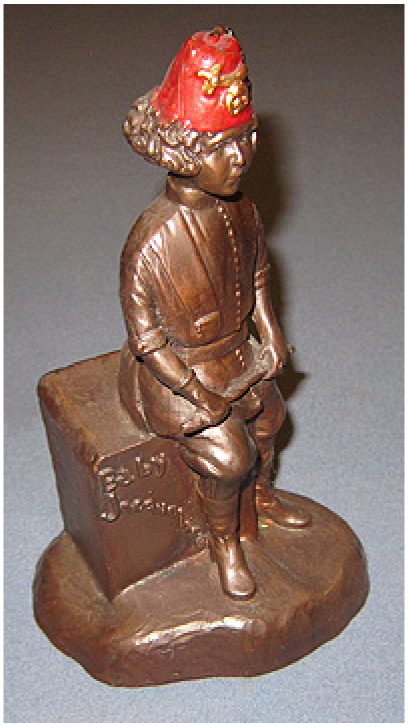

Shriner 'Baby Jacqueline'
Mid-1920s crystal radio in 9-1/2 inch tall cast plaster figure of 'Baby Jacqueline' wearing a shriner fez. The radio chassis is very similar to the one in the 1925 Shriner Convention radio. It is unknown who Baby Jacqueline was or exactly when and why this radio was made.

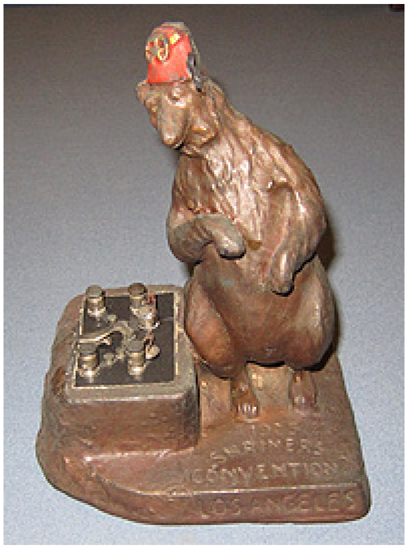

Shriner 1925 Convention radio
A crystal radio is embedded in this 9-inch tall cast plaster figure of a bear wearing a shriner fez. Molded into the base is '1925 SHRINERS CONVENTION, LOS ANGELES'. The Al Malaikah Temple in Los Angeles hosted the national shriner convention in 1925 and numerous souveniers were available. It is unknown if this radio was just a souvenier or if it was some sort of prize or award. The radio chassis is very similar to the one in the Shriner 'Baby Jacqueline' radio.

Spencer Pocket Radio
This old crystal radio is still in mint condition.

SPENCER
"POCKET"
RADIO

Spencer Pocket Radio
This crystal radio is in excellent condition, with its original box and operating instructions.

A WEZCO PRODUCT
The DOL
A Crystal Radio Requiring
NO TUBES!! NO BATTERIES!!
Giving Clear, Loud Reception!
A source of endless hours of fun and
entertainment for young and old
A WEZCO Product
Washington U.S.A.

Wezco 'DOL'
This crystal radio is shown with its original box.

THE NIC NAC TIC TAC CRYSTAL RADIO SET

MARCH 21, 2015 / BY CIRCUIT WIRING / IN RF RADIO CIRCUITS

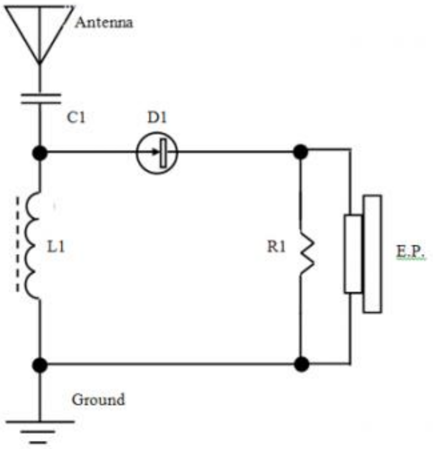

Warning: this kit contains small parts that may
form a choking hazard, and is not suitable for children under 5 years old.

Introduction

Crystal radios have been around for over 100 years and exist in many different shapes, sizes and complexities. Some are definitely better than others, but wired correctly and connected to a good Antenna and Ground wire system, this simple little AM radio receiver should pull in at least one strong local radio signal for you to listen in to. In diagram 1 you can see the basic layout of the ’crystal set’. They are called crystal sets, as the original designs used a special material called galena or pyrite, as the detector element, which picks out the sound wave from the radio wave. These crystals have been replaced with a modern component called a diode, and we’ll get to use one of those as we put this kit together.

How it works

Signals from your local AM radio stations are traveling through the atmosphere at the speed of light, and they pass through just about everything. These signals hit your antenna wire and induce a small signal current that flows down into the radio’s tuned circuit, and out through the ground wire, to Earth. The coil, L1 and fixed capacitor, C1, form the ’tuned circuit’ in our project, in conjunction with the antenna/ground wire system, which enables the receiver to select one station’s frequency and reject the rest. This may seem like a limitation on the radio, but once you’ve got this simple receiver going, you can add one or two more parts in order to make a ’tunable’ version that can tune across part or all of the AM radio band, for just a couple of dollars more. All you have to do is swap several parts of differing values around, (included in your kit,) until you receive a reasonably strong local AM radio station! Experimenting is half of the fun with these circuits.

The frequency selected by this process is comprised of two parts – the ’carrier wave’ which is the radio station’s fixed frequency (its ’spot’ on the dial of a transistor radio set) and the ’program signal’ which is the voice, music, secret sound, which originates inside the radio station’s studio. This is the signal that we want, and the next component, the diode, D1, enables the sound or audio signal, to be split away from the carrier wave. This is called the ’detector’ because it detects the sound signals contained inside the carrier wave.

pietzo or crystal earphones, as these devices need a direct path to ground for them to work properly and not distort the signal.

Resonance

At a chosen frequency a special effect called "resonance" occurs. This simply means that when the coil and capacitor working together as the 'tuned circuit' have received the one particular frequency they were set for, then maximum power transfer from transmitter to receiver will occur. All other radio station's signals will be fed out to earth by the ground wire, and the one you want will be playing away in the earphone. Our earphone has a special material called 'pietzo crystal' and when a small signal is fed into the crystal via the diode and link wire, the crystal will actually vibrate in unison with the signal. This crystal is flat and round, and is attached to a larger piece of tinfoil, which vibrates too, reproducing sound waves that we can hear.

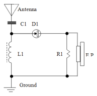

Diagram 1 Tic Tac Radio Circuit Diagram

In the diagram at left, you can see the circuit for our simple AM band crystal radio set. It has just four components – a capacitor, C1 which together with the next component, the inductor L1, which stores radio energy and helps tune the radio to one station. Thirdly, there's the diode D1, which separates the program signal (the voice, music, secret sound) from the radio station's transmitting frequency, and finally, the resistor R1, which forms a signal return path to ground and provides stability for the crystal earphone, EP, which changes the electrical signal coming out of diode D1, into sound waves which we can hear.

Parts Identification And Layout

The orange or blue coloured fixed capacitor is made of ceramic material, and consists of two small metallic plates, separated by the ceramic material. These two plates store an electrical energy charge between them, and allow RF (Radio Frequency) energy to pass through. Small ceramic caps are measured in picofarads (pF). The coil or inductor, is a larger cylinder shaped device, with coloured bands to indicate its value in micro henries (uH). The diode is a small signal type (BAT 46) and is made of a special material called silicon. This material allows the sound or audio signals to pass through to the earphone, but prevents the carrier wave from getting through. Finally, the crystal earphone is connected across the resistor, which is measured in ohms. This resistor (a 100K one, marked brown/black/yellow) provides stability for the earphone, by siphoning off some of the detected signal, and feeding it directly to the ground, via the ground wire.

Building The Radio Receiver

Take your parts from the packaging, and lay them out on the work area. Take the printed overlay template supplied with your kit, and peel off the protective backing, leaving the 'sticky stuff' exposed. Lay it on top of the small plastic oblong shaped base, aligning the edges together, and press the overlay onto the plastic. Take a sharp metal object (small jeweler's screwdriver, large pin or needle,) and make five holes in the material where you see a large dot. These dots are connection points where you will push the wires from each component part through and when they are all in place, you will then gently twist them together, forming a basic join. First, place the legs of capacitor C1 through holes 1 & 2. then take the coil, L1 and push its legs through holes 2 and 5. Take the diode and push its legs through holes 2 & 3, and finally take the resistor R1, and push its legs through holes 3 & 4. Take the short piece of link wire and push its legs through holes 4 & 5.

Building The Radio Receiver

overlay onto the plastic. Take a sharp metal object (small jeweler's screwdriver, large pin or needle,) and make five holes in the material where you see a large dot. These dots are connection points where you will push the wires from each component part through and when they are all in place, you will then gently twist them together, forming a basic join. First, place the legs of capacitor C1 through holes 1 & 2. then take the coil, L1 and push its legs through holes 2 and 5. Take the diode and push its legs through holes 2 & 3, and finally take the resistor R1, and push its legs through holes 3 & 4. Take the short piece of link wire and push its legs through holes 4 & 5.

When all four components and the link wire are mounted on the base, in accordance with the overlay, you can then make the joins by twisting the wire ends poking through each of the holes, together until they become tight. Bend them up out of the way, and make sure that none of the wires from one join are touching other joins. Finally, take the earphone and cut the 3.5mm plug off the end, carefully strip the insulation off each wire using a pocket knife or side cutters, and then twist the wires around the legs of the resistor, R1.

Attach your Antenna and Ground wires, and begin to test out your "Nic Nac Tic Tac" crystal radio.

Once you've finished testing the radio receiver, and have successfully received at least one local radio station, you can then fit the whole base with components attached into the Tic Tac box. Punch holes in the lid for the antenna and ground wire connections, and a larger one in the other end, for the earphone wires to protrude through.

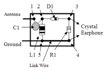

Diagram 2 Tic Tac components layout

Mount the parts on the plastic base by Carefully use a small sharp tool, such as a jeweler's screwdriver, to make the five holes in the base. Then insert the wires from each component through the appropriate holes, and then gently but firmly twist each group of wires up and underneath the base. Make sure that you insert the earphone wires in the hole in the back of the Tic Tac box before you strip the plastic insulation off each wire. Then tie a small knot in the wire, leaving about 2 cm, and attach the bared ends those wires to either end of resistor R1, before enclosing the whole assembly inside the Tic Tac box.

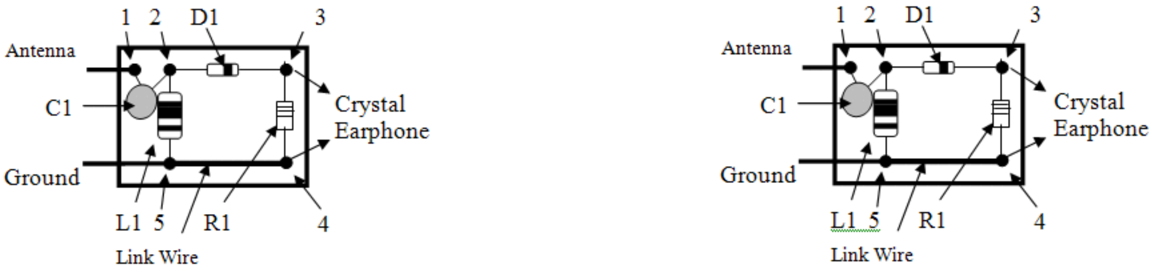
1
2
D1
3
Antenna
C1
Crystal Earphone
Ground
L1
5
R1
4
Link Wire
1
2
D1
3
Antenna
C1
Crystal Earphone
Ground
L1
5
R1
4
Link Wire

These two diagrams act as an overlay, so that you can place them on top of your plastic base and use them as a guide for component holes and wiring

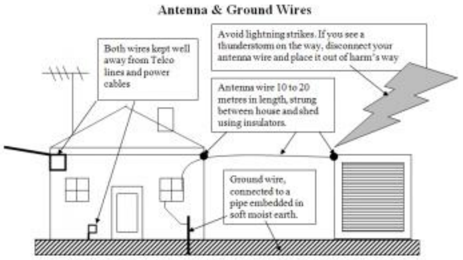
Antenna & Ground Wires
Both wires kept well away from Telco lines and power cables
Avoid lightning strikes. If you see a thunderstorm on the way, disconnect your antenna wire and place it out of harm's way
Antenna wire 10 to 20 metres in length, strung between house and shed using insulators.
Ground wire, connected to a pipe embedded in soft moist earth.

Erecting Antenna And Ground Wires

that you can obtain.

The Antenna Wire

The antenna wire needs to be reasonably long, and as high as you can safely place it. At least 10 to 20 metres in length and around 2 to 3 metres in height as a minimum, taking into account any power cables or telephone lines in the vicinity. Metal objects such as poles or frames or other similar structures can often 'ground' radio signals, preventing them from reaching your antenna wire, so you need it to be in as much open space as possible. Alternatively, you may be able to attach a short lead to a metal frame or structure (taking into account your electrical safety, which is YOUR responsibility,) that is insulated from ground potential, and any other voltages or currents, besides radio waves.

If you have any doubts about the electrical safety of your antenna wire installation, consult a qualified electrician, or other similarly qualified person, and remember, DO NOT erect antenna wires near to electrical power or telephone cables or other structures of a dangerous nature – check it out first!

Ground Wires

Ground wires are just as important as antenna wires, as they form part of the same 'front end' tuned circuit, completing the circuit to ground. Radio stations transmit their signals through the atmosphere, but they also travel along the ground and just underneath I – these are called 'ground waves' and the ground wire helps to capture them. It also gets rid of unwanted carrier waves that have not been selected by the tuned circuit and its preselected values for coil and capacitor. The best way to make a solid ground connection, is to find a cold water tap or pipe with no attached electrical ground wire system added to it, and bare your ground wire and wrap about five or six turns of bared copper wire onto the pipe, and then twist the end of it around the rest, securing it to the pipe. Use a metal clamp or bracket, if you think that is warranted to give you a firmer connection.

Ground wires are best connected to either a properly installed ground stake (embedded up to a metre in moist earth) or taken directly to cold water pipes or backyard taps, that have NO CONNECTION to any of the building's mains electrical earth wiring. DO NOT use gas pipes or hot water pipes, as these installations DO NOT go to an electrical ground, and connecting your ground wire to them may result in shocks or other unpleasant side effects. Connect ground wires to a safe and independent electrical ground connection. Once again, if you have ANY DOUBTS about the safety of a ground wire installation, consult a qualified electrical parson to ensure that your installation is safe. AH 2009

Hints & Tips On Getting The Tic Tac Radio Going

Having problems getting a radio signal? Or are you getting a signal with lots of noise, hum, clicks and pops in the earphone? Not able to string up a conventional antenna or ground wire system? This section is devoted to helping you get the most out of your crystal radio project. Most people will be able to erect an antenna or ground wire system either in their back yards, or in a nearby park or waterfront area. Some people live in dense housing lots or unit blocks, and it can seem impossible to do the necessary things in terms of establishing your antenna/ground system.

There are alternatives to the conventional arrangements, such as using the cold water pipes under your kitchen sink or bathroom vanity. Using parts of you metal balcony barrier/fence. These can provide useful levels of signal for simple radio reception. Make sure though, that none of your cold water pipes double up as a path to ground for any 240 volt mains wiring, as using this as a ground can be risky.

If you end up with a relatively short antenna you may find that you are getting a signal or sorts, but it's not loud enough to be satisfied with that level of sound, as a good result. You can use an extra component called a 'variable capacitor' between your antenna lead in wire, and the crystal set itself, to help 'tune in' the antenna. With simple crystal type receivers, it's important to be able to 'match' the antenna to the set, particularly if the antenna ends up being a short wire one (one that is less than 10 or fifteen metres in length, and 2 metres above the ground or less).

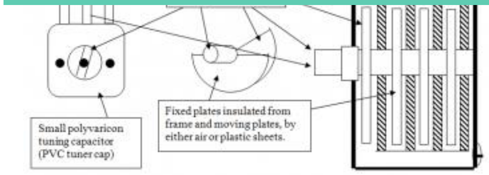
Small polyvaricon tuning capacitor (PVC tuner cap)
Fixed plates insulated from frame and moving plates, by either air or plastic sheets.

In the diagram above, you can see two different types of variable capacitor. Although the materials, construction and size are different, the principle of operation is the same. The insulation between the plates can either be air (the atmosphere) or plastic materials, such as Mylar. The polyvaricon shown on the left is a miniature plastic tuning capacitor, found in most modern pocket transistor radios. The larger one shown on the right, is made almost entirely of metal, and uses air as an insulator between the fixed and moving plates, and phenolic material to keep the fixed plates positioned correctly on the metal frame, without actually touching it. The idea, is to interpose the variable capacitor between your antenna wire, and the set, and then "tune in" the antenna. Don't forget that you've been given a set of parts, which includes three ceramic (fixed) capacitors and two mini inductor coils. If you can't get a station with one set of coil and capacitor, try swapping them around until you get one talking or playing music in your earphone.

A Word On Crystal Earphones

You will have been supplied a small pietzo crystal earphone as part of your kit of parts. These generally speaking give good sound reproduction, but occasionally they can drop out. Odd ones are not very good at all and need replacing. It is the way they are made at the factory, and also how they are stored and handled. Sometimes dropping them onto a hard floor surface can cause the connections inside the earphone to work intermittently. To fix them up and get them going again, try blowing down the small earpiece several times, each time listening to see if full sound has been restored. If this doesn't work, then try tapping them on the palm of your hand several times, each time listening to test the sound. If you need more help with this kit, you can contact us, or go onto the Internet, and try websites like www.midnightscience.com or www.theradioboard.com or just Google the words "crystal radio set" and see what happens!

AH 2009

Download

Download this circuit in DOC

No related diagrams.

FM TRANSMITTER DIAGRAM

MINIATURE FM TRANSMITTER #1

SORRY, COMMENTS ARE CLOSED!

TYPE & PRESS ENTER

MOST POPULAR CIRCUITS

- fan starts to rotate at +40 C on the heatsink
- Wiring Diagrams Of 1961 Ford V8 Fairlane And Galaxie
- 2006 Honda CBR600RR Behind Left Middle Fairing Fuse Box Diagram
- Frequency Divider by 555 and 4040
- FM Moulator with IC 555
- Sound Controlled Lights Circuit
- 2008 BMW Z-4 Main Fuse Box Diagram
- 2000 Jeep Cherokee Heater Fuse Box Diagram
- 2002 Chevrolet Tahoe 5.3 Stoplamp Switch Fuse Box Diagram
- Processor with Advanced Security

CATEGORIES

- 555 Circuits
- Alam Circuits
- Amplifier and Audio Circuits
- Automotive Diagrams
- Battery and Charger Circuits
- Car Fuse Box Diagrams
- Control Circuits
- Converter Circuits
- Datasheet Diagrams
- Digital Circuits
- General Circuits
- Inverter Circuits
- Knowledge basis

- Microcontroller Circuits
- Motor Control Circuits
- Motorcycle Diagrams
- Op-Amp Circuits
- Oscillator Circuits
- Power Supply Circuits
- RF Radio Circuits
- Sensor Circuits
- Service Manual
- Solar Diagrams
- Switch and Relay Circuits
- Television and Video
- Tester Circuits
- Timer Circuits

CIRCUIT WIRING SEARCH

Audio battery chevrolet Cigar Lighter circuit Circuit Diagram circuit diagrams Electrical Circuit electrical diagram Electrical Schematic Electrical System Electronic C ford fuel pump Fuel Pump Relay Fuse Block Fuse Box Fuse Holder fuse layout fuse map Fuse Panel Headlamp Horn Horn Relay ignition coil Ignition Switch Instrument Cluster power Power Supply Power window Radio Relay Schematic Circuit Schematic Diagram schematic diagrams Starter Relay Studebaker wiring wiring connection Wiring Diagram wiring diagrams Wiring Harnes wiring schematic wiring system wiring work

POWERED BY WORDPRESS / THEME BY WPQA

CIRCUIT WIRING DIAGRAMS © 2018

3rd July 2013

The simplest FM Crystal Radio

The simplest battery-free FM Crystal Radio.

Make the world's simplest battery-free FM crystal radio

Only consists of two high frequency diode: 1N34A or ISS86 or ISS106.
A 100K Resistor.
An 1.8M antenna.
And a crystal or high impedance earphone

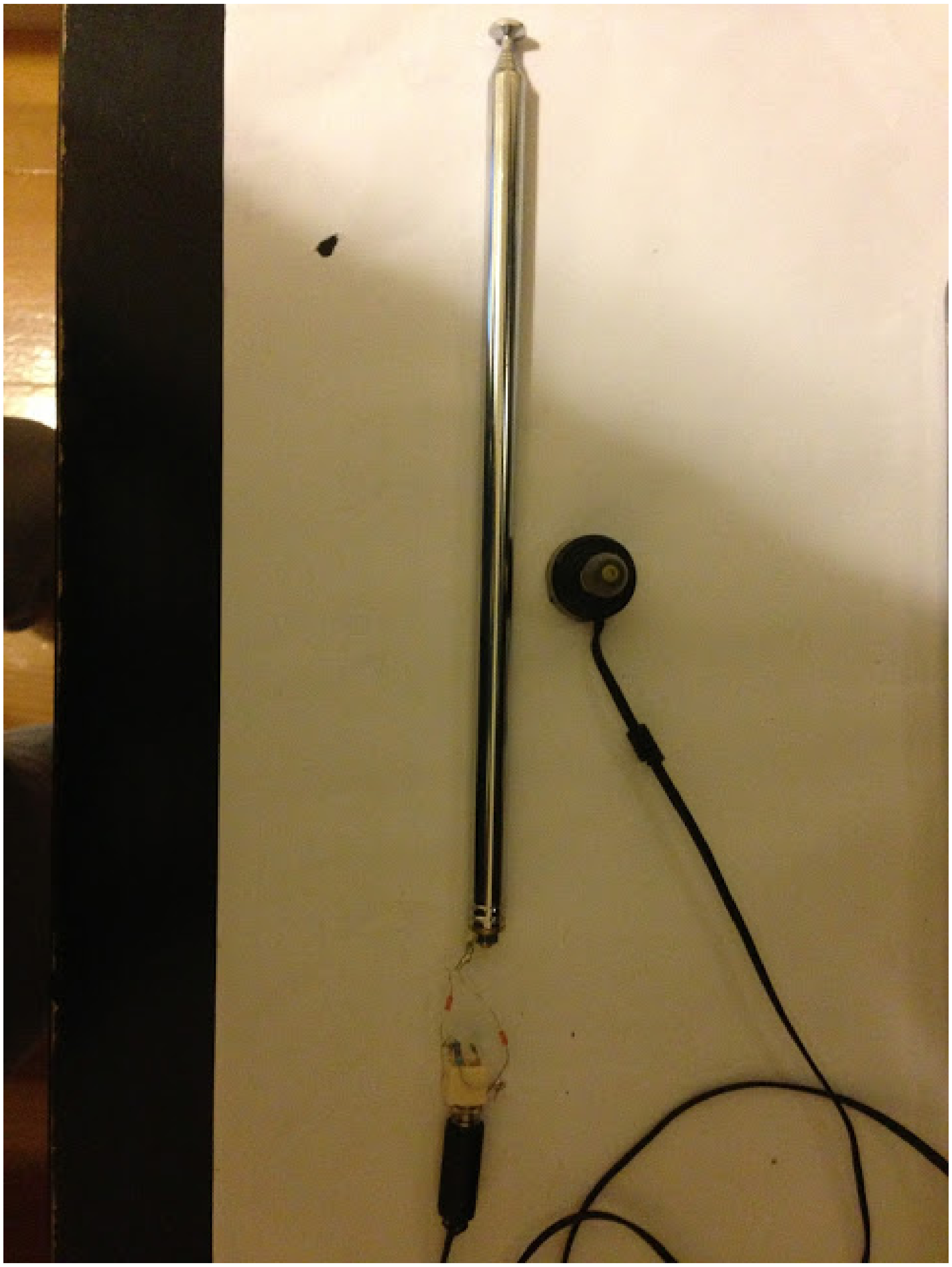

[http://1.bp.blogspot.com/-x9lr6oLJ9z8/UdQ_lBNismI/AAAAAAAABQE/pfbpj7x0Azs/s1600/IMG_2607.JPG]

Refer to this circuit diagram

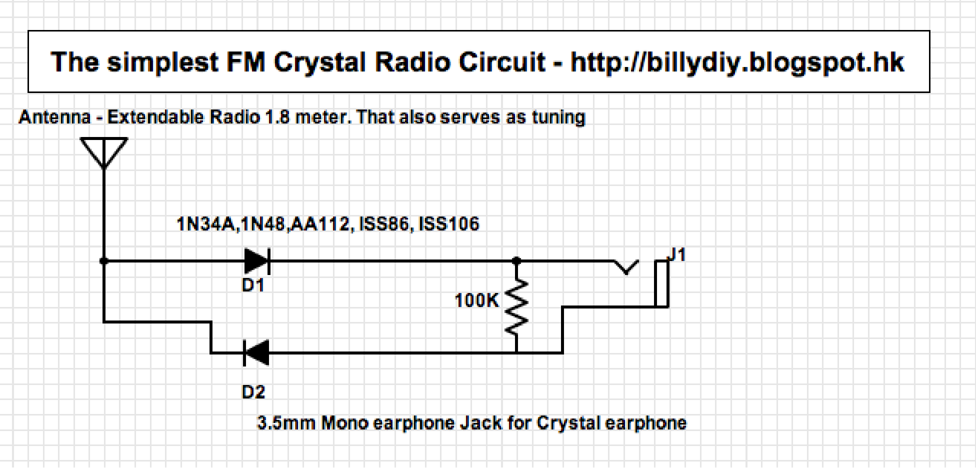
The simplest FM Crystal Radio Circuit - http://billydiy.blogspot.hk
Antenna - Extendable Radio 1.8 meter. That also serves as tuning
1N34A,1N48,AA112, ISS86, ISS106
D1
100K
J1
D2
3.5mm Mono earphone Jack for Crystal earphone

[http://2.bp.blogspot.com/-G9onfk5x8FA/UdRDYhFFyAI/AAAAAAAABRA/mtiP8wFX78E/s639/Screen+Shot+2013-07-03+at+11.28.54+PM.png]

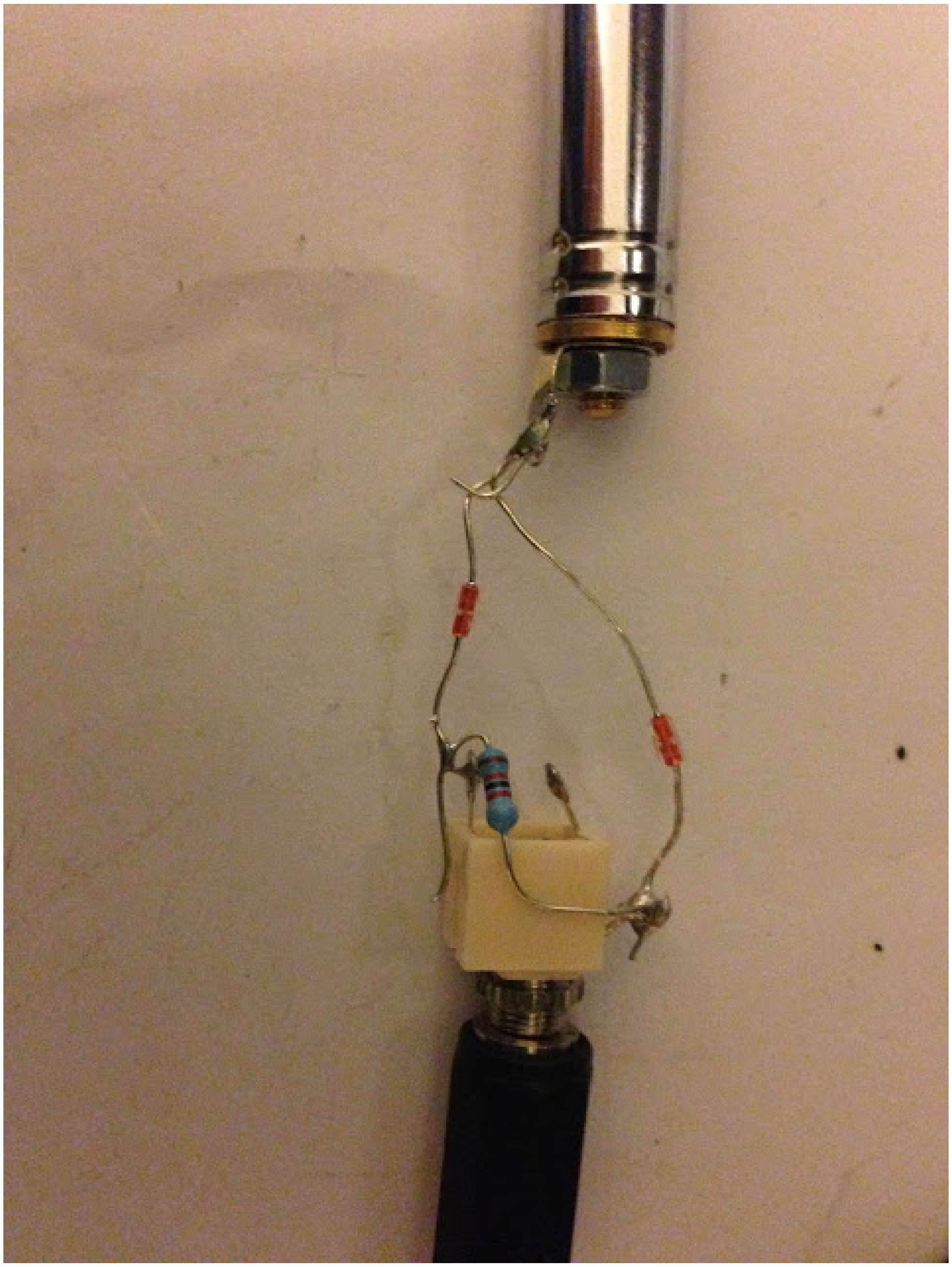

[http://4.bp.blogspot.com/-UVDtJdx6leQ/UdQ_lE_swzI/AAAAAAAABQM/ToNTHsm8pdc/s1600/IMG_2608.JPG]

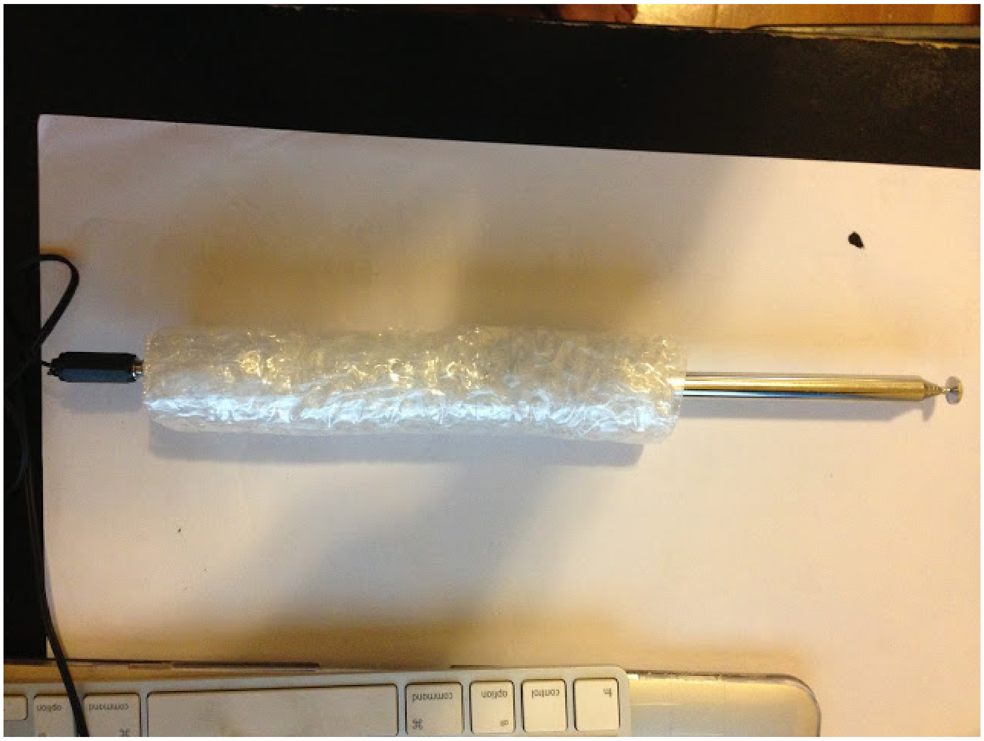

[http://1.bp.blogspot.com/-jvFTM9CBpKQ/UdQ_kx4Q35I/AAAAAAAABQA/W1wxPV0ezZk/s1600/IMG_2609.JPG]

Posted 3rd July 2013 by Billy

Start Download
Search Videos W/ HowToSimplified On How To Do It Yourself - Free! Go to howtosimplified.com

0 Add a comment

Enter your comment...

Comment as: Unknown (Goo ▼ Sign out

Publish Preview ☐ Notify me

eti project

Tic Tac Radio

THIS MEDIUM WAVE receiver measures about 60x36x15mm when constructed in a "Tic Tac" mint box. From the circuit in Fig. 1 it will be seen that the well-known TRF radio IC, ZN414, is used. This is followed by a single audio amplifier. This IC is often used alone for earphone reception, but the addition of an amplifier considerably boosts the volume.

The normal medium wave band is tuned by the compression trimmer VC1, which is available with a shaft fitted so that a standard control knob can be used. The audio output is to a 2.5mm jack socket via the isolating capacitor C4 (to ensure that the operating conditions for Q1 are not upset by the DC resistance of the phone or headset. A crystal earpiece can also be used.

COMPONENTS

An important consideration in building a miniature radio is the actual size of the components. C1 and C2 are easily obtained low voltage ceramic discs; C3 and C4 are small bead capacitors. The resistors are standard 1/3 or 1/4 watt.

Cut the board so that it will slip into the case. A part of the board is then cut away to allow the battery to fit. VC1 is fitted to the board by its bush, which is cut or filed down to avoid unnecessary projection.

The box lid takes the miniature slide switch with no modification to

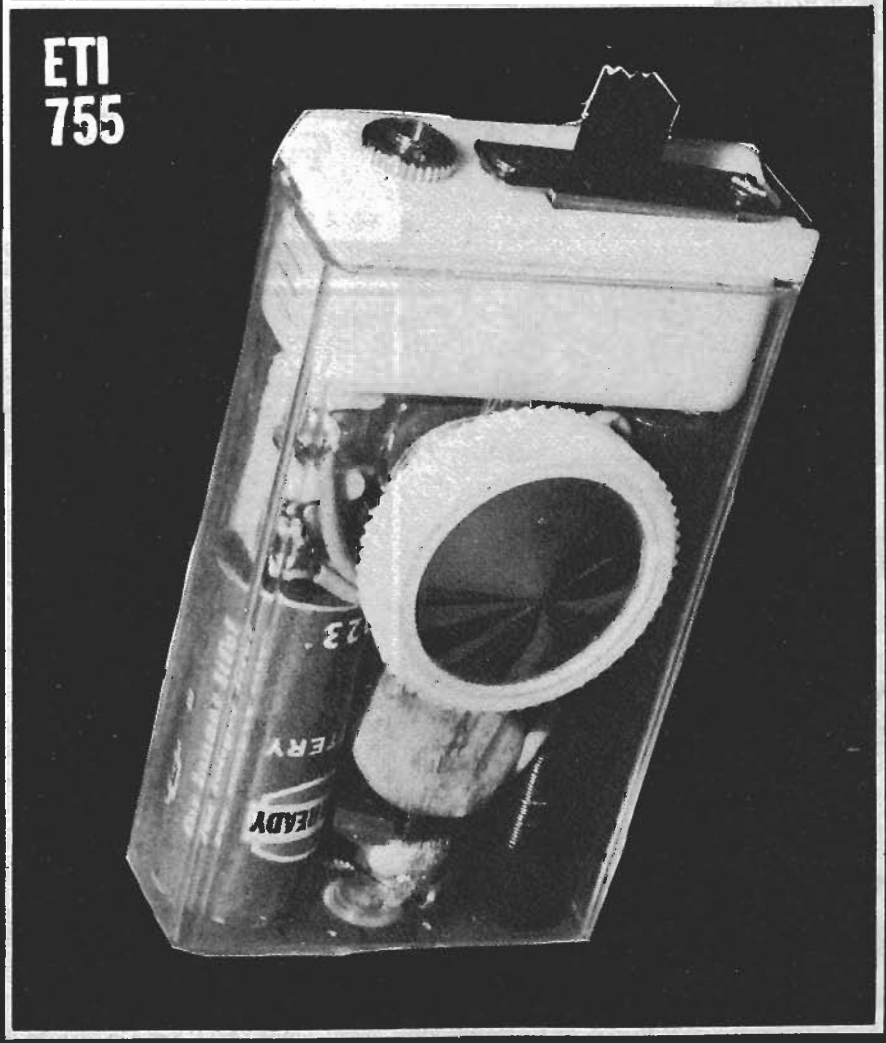

Tic Tac Radio

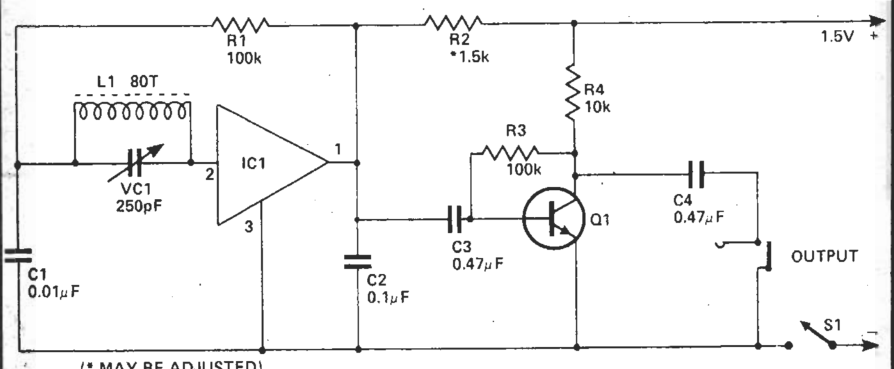

Fig. 1. Circuit of the miniature receiver.

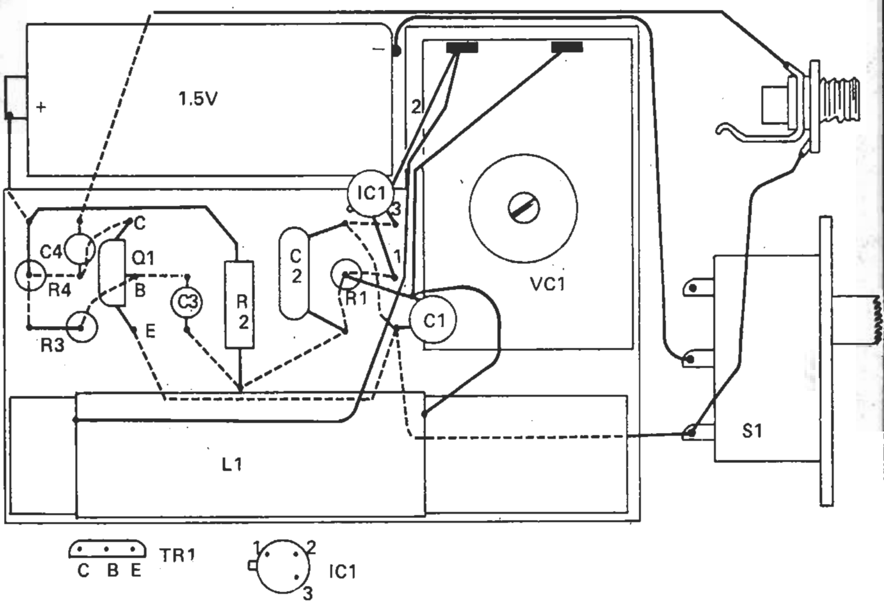

Fig. 2. Wiring and semi-conductor leads.

PARTS LIST

R1	100k	¼W	5%
*R2	1.5k	¼W	5%
R3	100k	¼W	5%
R4	10k	¼W	5%
C1	0.01µF disc ceramic		
C2	0.1µF disc ceramic		
C3	0.47µF tantalum bead		
C4	0.47µF tantalum bead		

* (may need adjustment)

VC1	Type TP4 250pF compression trimmer with Z236 spindle (Home Radio, Mitcham)
IC1	ZN414
Q1	ZTX300
L1	80 turns 32 swg enamelled wire on 42 x 9mm ferrite rod.

D23, 1.5V cell, small knob, 2.5mm jack socket, miniature slide switch, 42 x 30mm 0.15in matrix perforated board, "TIC TAC" mint box.

the slot. Two small bolts hold the switch. The switch and jack are placed inside the lid as in Fig. 2.

FERRITE ROD

This will have to be cut from a longer rod. File a groove all round the rod 42mm from one end. The rod can then be broken by hand. Start to wind the 32swg enamelled wire 6mm from one end and fix with adhesive. Wind on eighty turns, side by side, and glue the wire again. Finally cement the rod to the perforated board.

WIRING

All leads etc can be seen in Fig. 2. It is important to keep the connections close to the board. Avoid large joints or the depth will prevent the radio fitting in the case. It is best to fix S1 and the jack to the lid before fitting the board,

Before S1 and the socket are fitted connect projecting leads for negative line, C4, and battery negative. Battery connections are soldered to the cell.

RESISTOR R2

Here 1.5k should be suitable. The actual layout of leads and components is likely to vary from that of the prototype and this increases the chances of instability. This would be manifest by IC1 oscillating so that whistles accompany reception with some signal levels, or on some frequencies. A check can easily be made before fitting the radio in its box. If reception is satisfactory (as is likely) R2 can be left at 1.5k. If R2 can be reduced in value, possibly to 1.2k or 1k, without whistles arising, this will increase gain. (R2 should not be less than 470 ohm).

BOX AND VC1

Push the receiver into its box and mark the position of the adjusting screw of VC1. Remove the radio and drill this point right through (so that the bush and nut can fit in the bottom hole in the box). Use a sharp drill with light pressure, or ream out the hole. Be careful because the material is brittle.

With the receiver replaced, completely remove the adjusting screw, taking care not to displace the washers, and screw in the shaft. The latter is cut back to 6mm so that a small knob can lie near the box. This is a push-fit over a flat filed on the shaft.

PHONES

Best reception of all is with a good pair of high impedance headphones (the Heathkit GD-396, 2k impedance, will be ideal). Where a miniature earpiece is required, this should be a high impedance unit, or volume is to be severely reduced.

US 20070241470A1

(19) **United States**
(12) **Patent Application Publication** (10) **Pub. No.: US 2007/0241470 A1**
Haisch et al. (43) **Pub. Date: Oct. 18, 2007**

(54) **QUANTUM VACUUM ENERGY EXTRACTION**

(76) Inventors: **Bernard Haisch**, Redwood City, CA (US); **Garret Moddel**, Boulder, CO (US)

Correspondence Address:
PRITZKAU PATENT GROUP, LLC
993 GAPTER ROAD
BOULDER, CO 80303 (US)

(21) Appl. No.: **11/236,142**

(22) Filed: **Sep. 26, 2005**

Related U.S. Application Data

(60) Provisional application No. 60/613,778, filed on Sep. 27, 2004.

Publication Classification

(51) **Int. Cl.**
B29D 11/00 (2006.01)
(52) **U.S. Cl.** **264/1.27**

(57) **ABSTRACT**

A system is disclosed for converting energy from the electromagnetic quantum vacuum available at any point in the universe to usable energy in the form of heat, electricity, mechanical energy or other forms of power. By suppressing electromagnetic quantum vacuum energy at appropriate frequencies a change may be effected in the electron energy levels which will result in the emission or release of energy. Mode suppression of electromagnetic quantum vacuum radiation is known to take place in Casimir cavities. A Casimir cavity refers to any region in which electromagnetic modes are suppressed or restricted. When atoms enter into suitable micro Casimir cavities a decrease in the orbital energies of electrons in atoms will thus occur. Such energy will be captured in the claimed devices. Upon emergence form such micro Casimir cavities the atoms will be re-energized by the ambient electromagnetic quantum vacuum. In this way energy is extracted locally and replenished globally from and by the electromagnetic quantum vacuum. This process may be repeated an unlimited number of times. This process is also consistent with the conservation of energy in that all usable energy does come at the expense of the energy content of the electromagnetic quantum vacuum. Similar effects may be produced by acting upon molecular bonds. Devices are described in which gas is recycled through a multiplicity of Casimir cavities. The disclosed devices are scalable in size and energy output for applications ranging from replacements for small batteries to power plant sized generators of electricity.

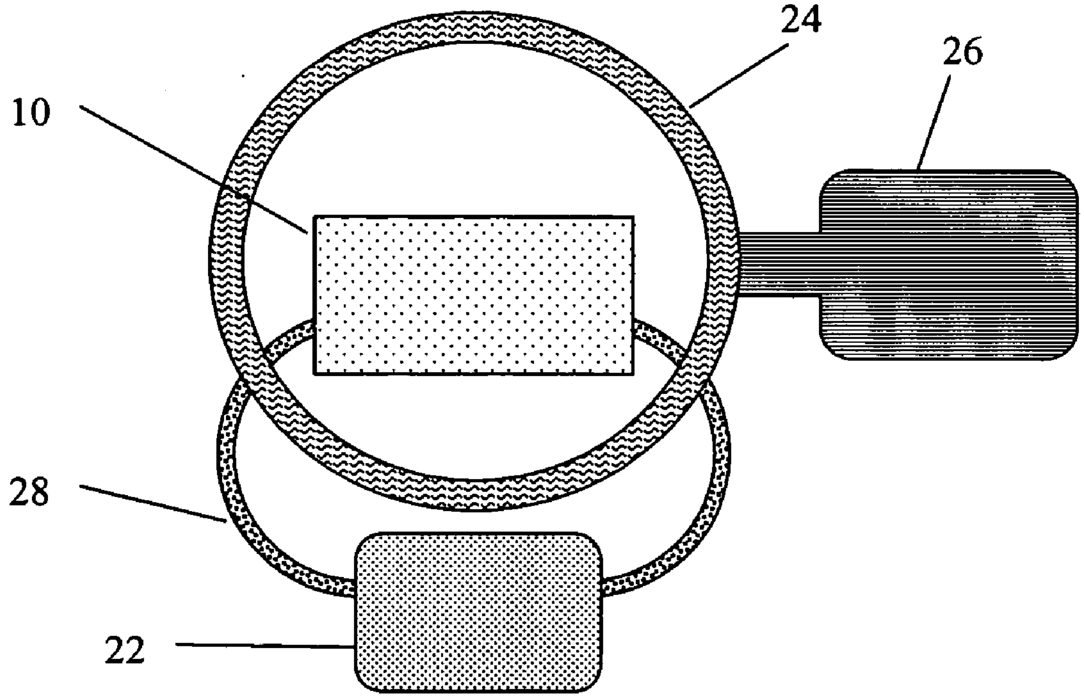

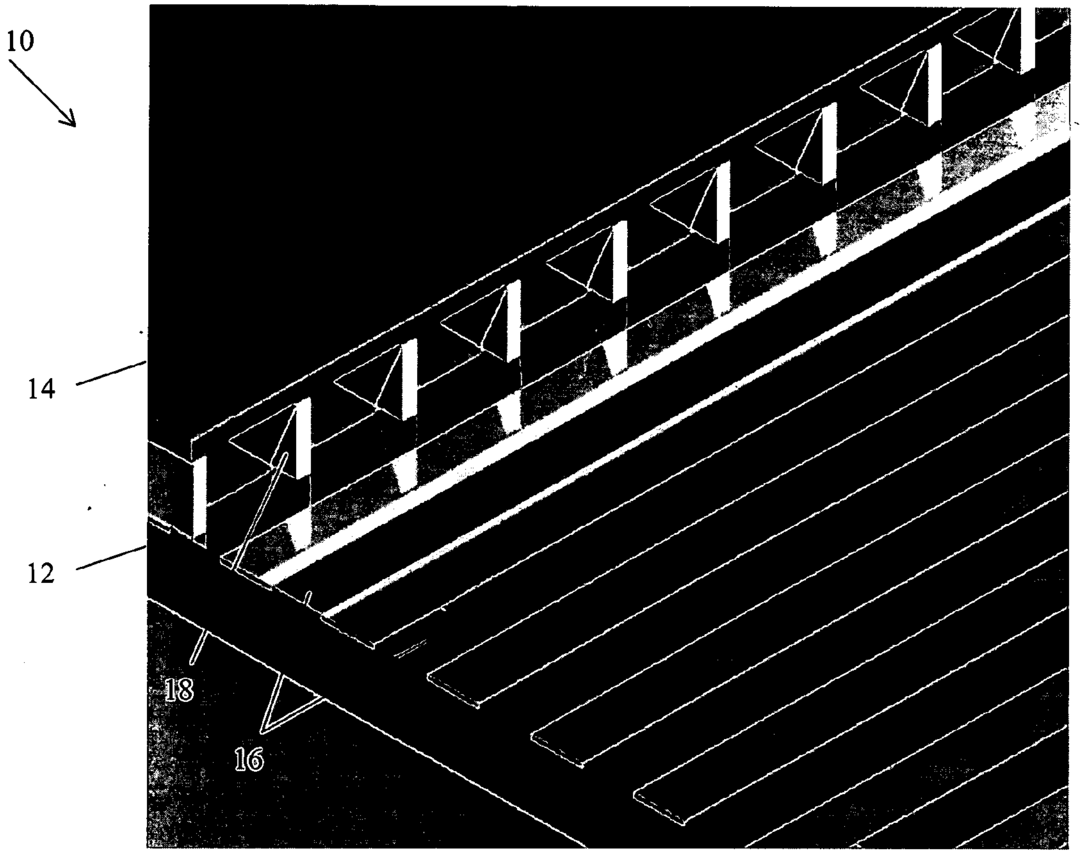
10
14
12
18
16

FIG. 1

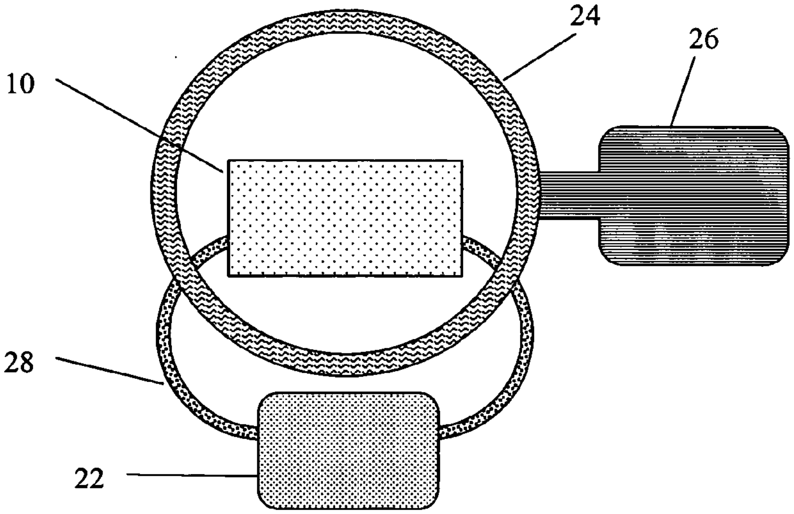
24
26
10
28
22

FIG. 2

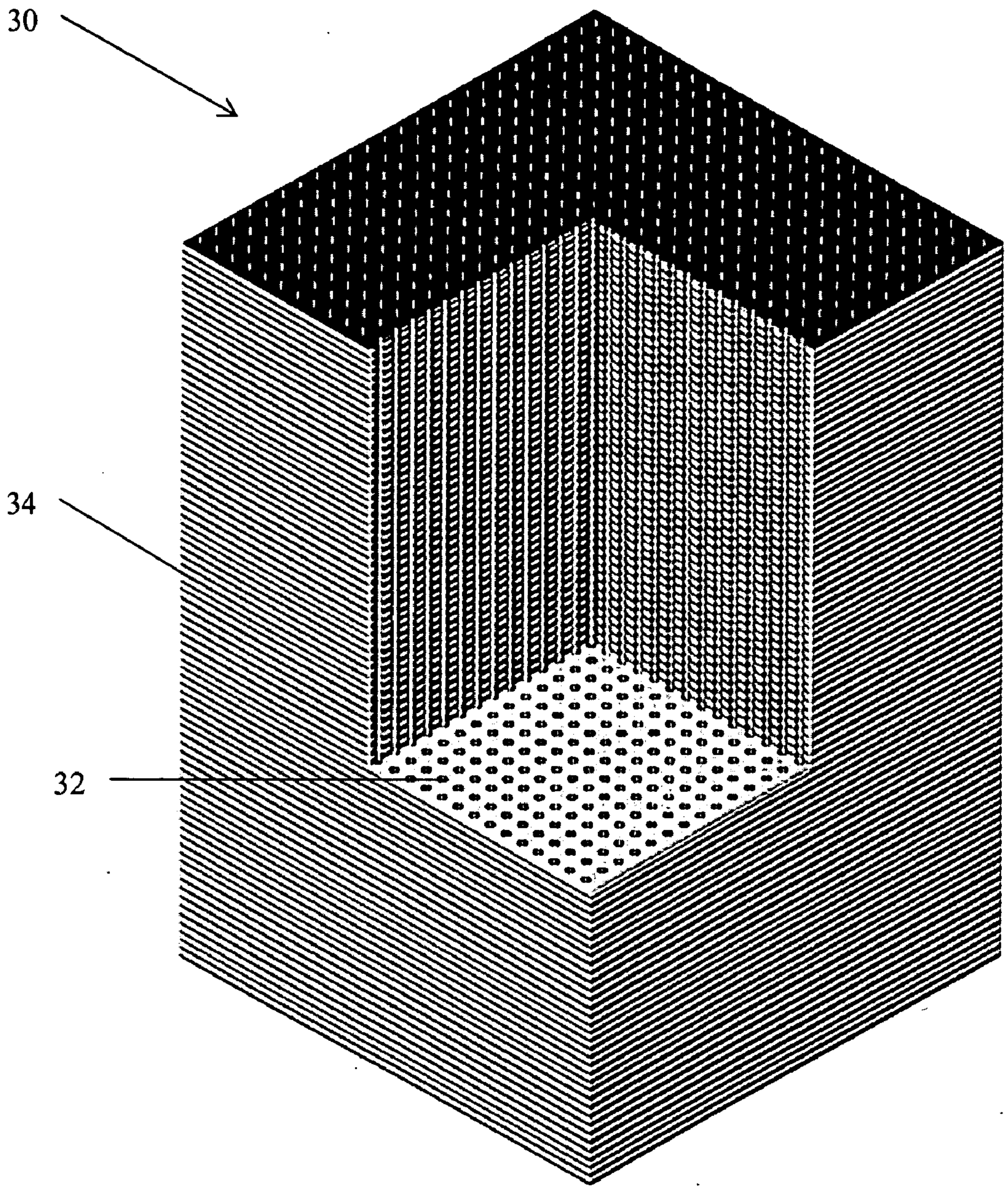

FIG. 3

FIG. 4A

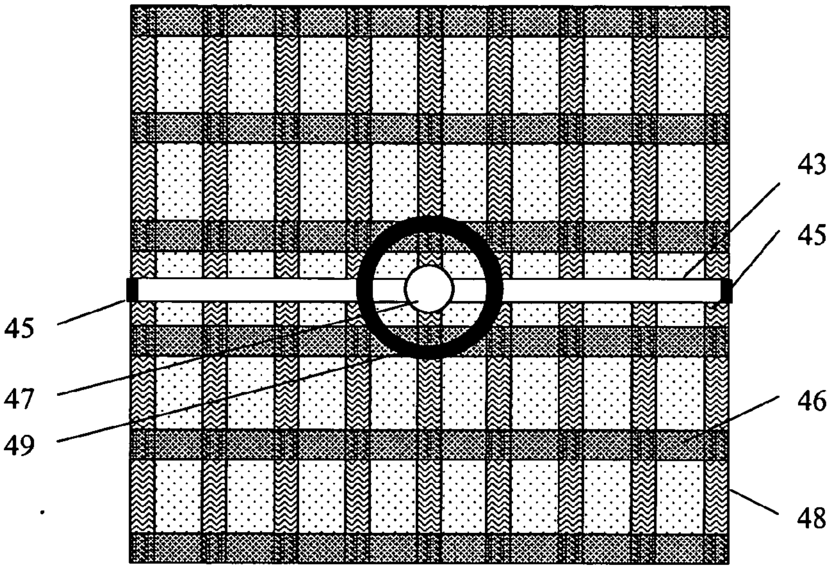
43
45
45
47
49
46
48

FIG. 4B

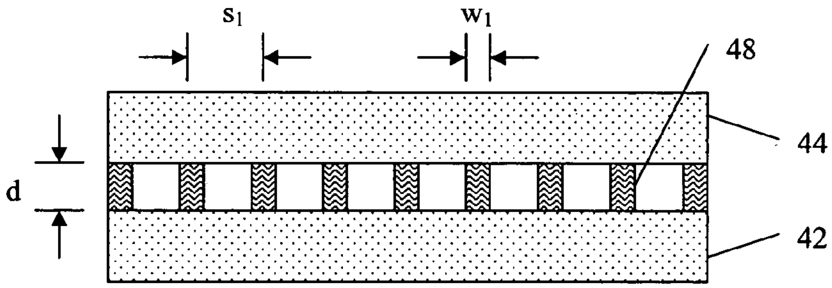
s_1
w_1
48
44
d
42

FIG. 4C

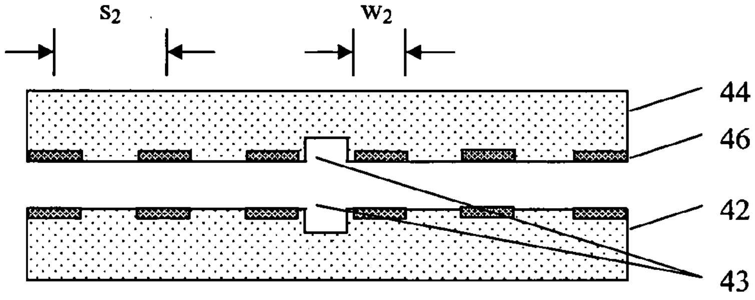
s_2
w_2
44
46
42
43

FIG. 4D

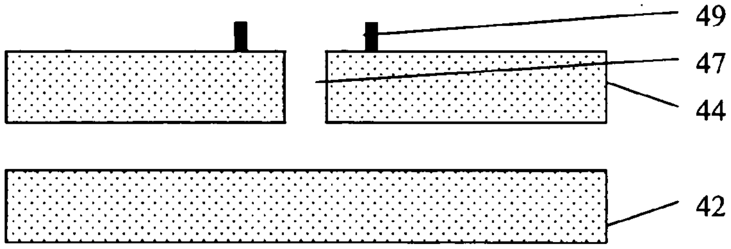

FIG. 5A

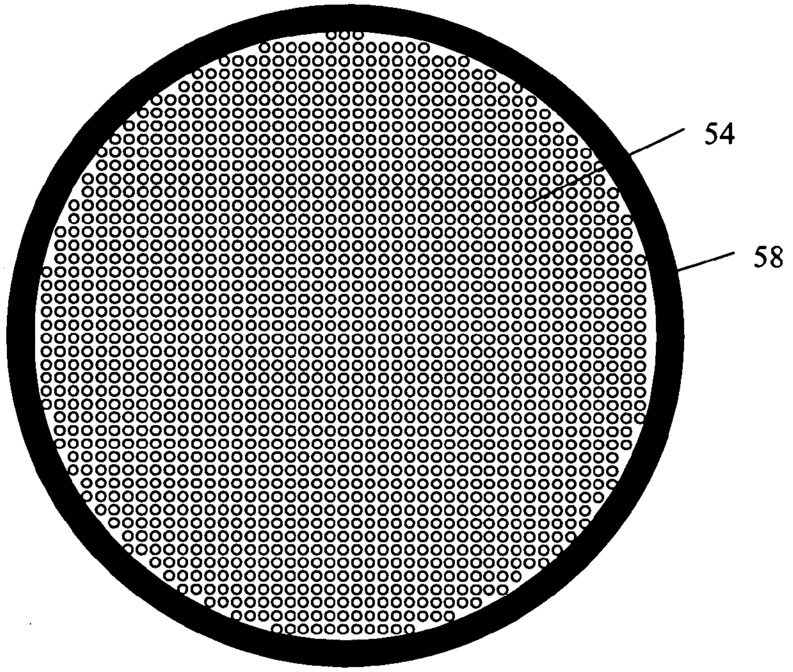

FIG. 5B

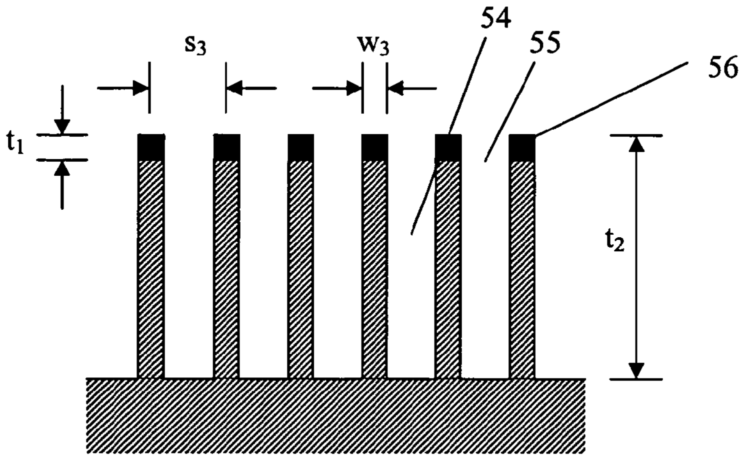
s3
w3
54
55
56
t1
t2

FIG. 5C

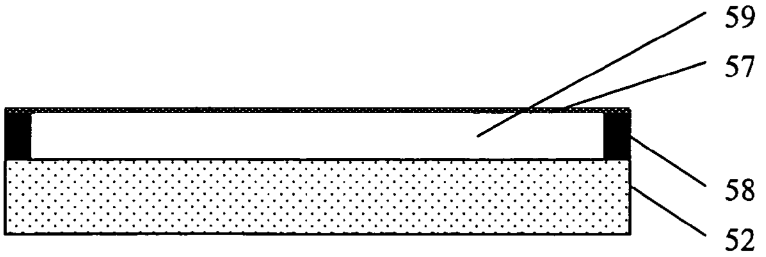

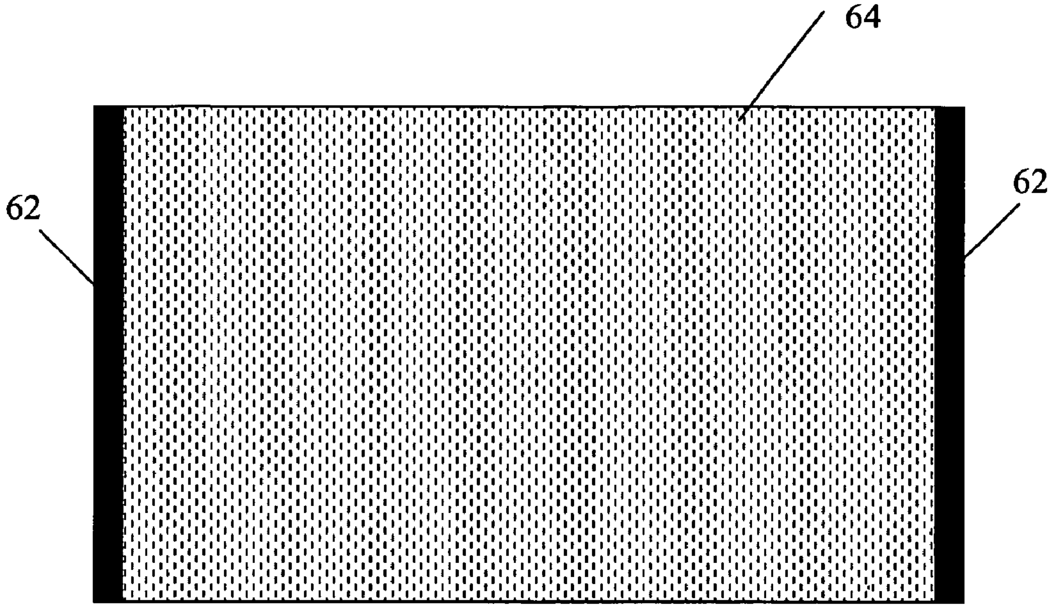

FIG. 6A

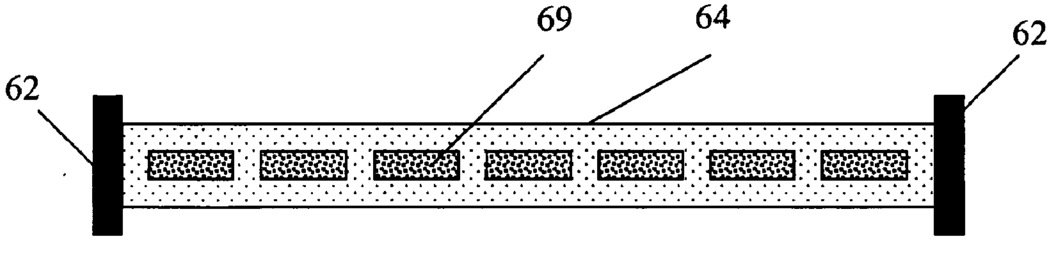

FIG. 6B

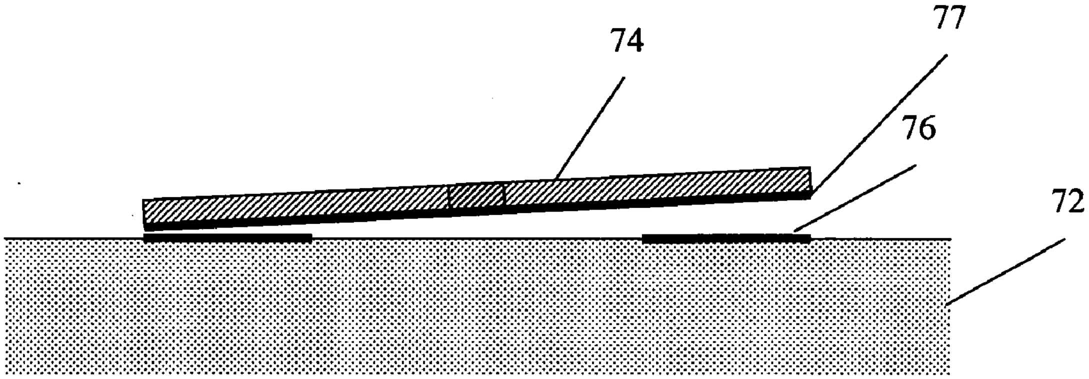

FIG. 7A

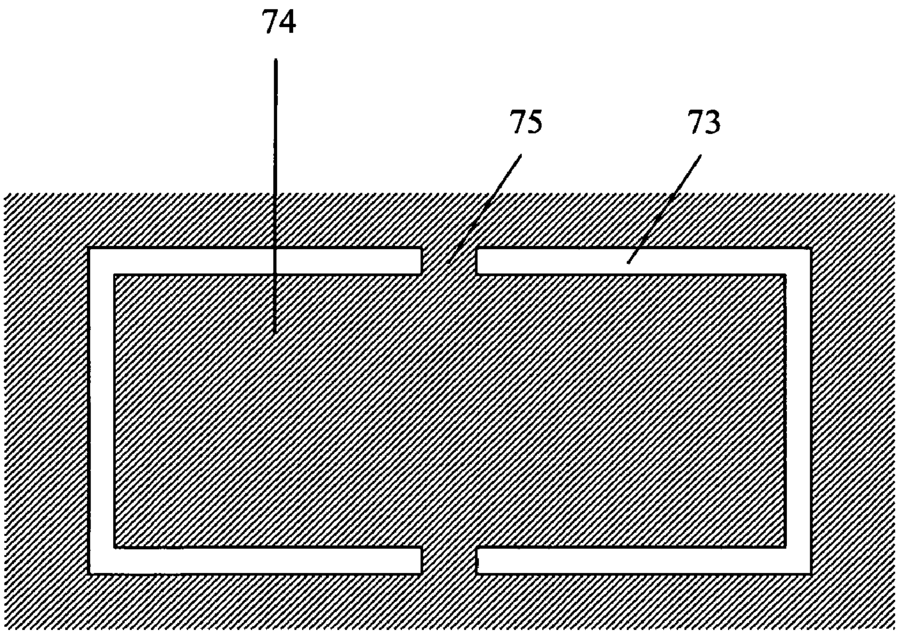

FIG. 7B

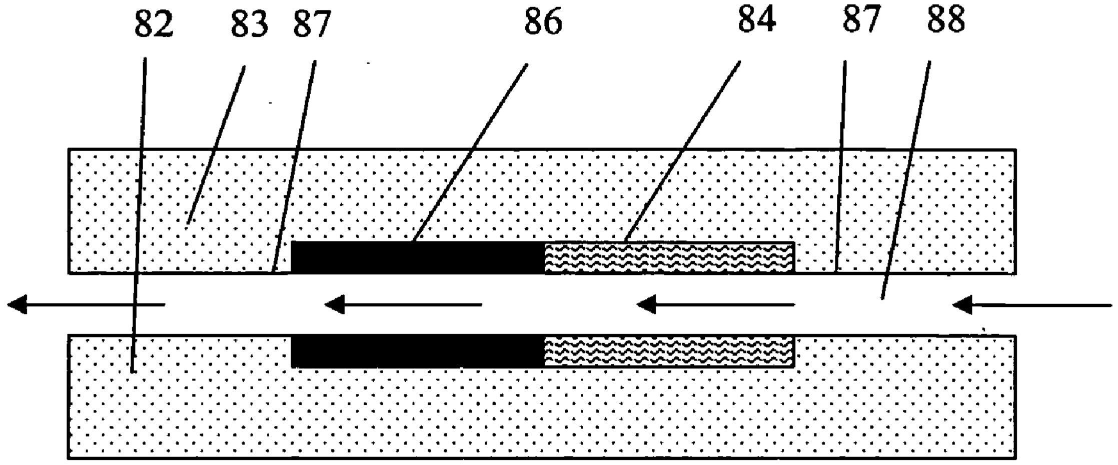
82
83
87
86
84
87
88

FIG. 8

QUANTUM VACUUM ENERGY EXTRACTION

BACKGROUND OF THE INVENTION

[0001] Max Planck proposed the concept of zero-point energy in 1912. The idea was then studied by Albert Einstein and Otto Stern in 1913. In 1916 Walther Nernst proposed that the Universe was filled with zero-point energy. The modern field of stochastic electrodynamics is based upon these ideas.

[0002] At that same time the structure and stability of the atom were puzzles. The Rutherford model of the atom was based on analogy to the motions of planets (electrons) around the Sun (the nucleus). However this was not feasible. The orbiting electron(s) would emit Larmor radiation, quickly losing energy and thus spiraling into the nucleus on time scales less than one-trillionth of a second, thereby rendering stable matter impossible. It is now known within the context of stochastic electrodynamics (SED) theory that a possible solution involves the absorption of zero-point energy. It was shown in 1975 by Boyer that the simplest possible atom and atomic state, the hydrogen atom in its ground state, would be in a state of equilibrium between Larmor radiation and absorption of zero-point energy at the correct radius for a classical Rutherford hydrogen atom.

[0003] Since this solution was not known in 1913, Niels Bohr followed a different path by simply postulating that only discrete energy levels were available to the electron in an atom. This line of reasoning let to the development of quantum theory in the 1920s. The concept of classical zero-point energy was forgotten for a decade. However the same concept found itself reborn in a quantum context in 1927 with the formulation of the Heisenberg uncertainty principle. According to the principle, the minimum energy of a harmonic oscillator has the value hf/2, where h is Planck's constant and f is the frequency. It is thus impossible to remove this last amount of random energy from an oscillating system.

[0004] Since the electromagnetic field also must be quantized in quantum theory, a parallel is drawn between the properties of a quantum oscillator and the waves of the electromagnetic field. It is concluded that the minimum energy of any possible mode of the electromagnetic field, consisting of frequency, propagation direction and polarization state, is hf/2. Multiplying this energy by all possible modes of the field gives rise to the electromagnetic quantum vacuum, which has identical properties—energy density and spectrum—to the classical zero-point energy studied by Planck, Einstein, Stern and Nernst a decade previously.

[0005] The line of inquiry involving classical physics plus the addition of a classical zero-point field was reopened in the 1960s by Trevor Marshall and Timothy Boyer and has been named stochastic electrodynamics (SED). SED asks the question: "Which quantum properties, processes or laws can be explained in terms of classical physics with the only addition being a zero-point electromagnetic field." Two of the early successes were a classical derivation of the blackbody spectrum (i.e. one not involving quantum physics) and the discovery that a classically orbiting electron in a hydrogen atom emitting Larmor radiation but absorbing zero-point radiation would have an equilibrium orbit at the classical Bohr radius. An initial approach to this problem by Timothy Boyer (1975) was perfected by H. E. Puthoff (1987). Their analyses treated the orbiting electron as a harmonic oscillator.

[0006] This result underwent a major new development with the recent work of Daniel Cole and Y. Zou which simulated the orbit of a classical electron in a true Coulomb field of a hydrogen nucleus and found that such a realistic electron would find itself in a range of distances from the nucleus, in agreement with quantum mechanics, owing to the random nature of the emission and absorption processes. The mean position is at the correct Bohr radius, but the actual distribution of positions very precisely duplicates the electron probability distribution of the corresponding Schrödinger equation in which the electron is regarded as being represented by a wave function. (In the SED representation the electron is "smeared out" not because it is a wave function, but because as a point-like particle it is subject to the continuous perturbations of the electromagnetic quantum vacuum fluctuations.)

[0007] A clear consequence of this theory is that a reduction of the electromagnetic quantum vacuum at the frequency corresponding to the orbit of the electron will result in a decay of the orbit since there will thereby be an imbalance in the Larmor radiation vs. absorption.

[0008] The electromagnetic quantum vacuum energy spectrum is proportional to the cube of the frequency. If the vacuum energy is suppressed at the frequency of the "normal" orbit of the electron, this will cause the electron to spiral inward to a higher frequency orbit. In this fashion it will then encounter a new equilibrium situation with the electromagnetic quantum vacuum energy spectrum owing to that spectrum's increase with the cube of the frequency.

[0009] If the SED interpretation is correct for the hydrogen atom as the analyses of Boyer, Puthoff, Cole and Zou indicate, it must apply as well to all other atoms and their multi-electron configurations. In that case, a transition of an electron from an excited state to a lower energy state involves a rapid decay from one stable orbit to another, not an instantaneous quantum jump. The details of the bases for stability of electron orbits has yet to be determined by SED theory, but the logical extrapolation from the single-electron hydrogen case is clear: electron orbits in all atoms must be determined by an emission vs. absorption balance and thus are subject to modification involving mode suppression of the electromagnetic zero-point field at appropriate frequencies.

[0010] It is claimed that modification of electron orbits is in essence the same process as natural transition between energy levels of electrons in atoms and therefore that the energy released in such a process can be captured in the same way as ordinary transition energy.

[0011] By moving an atom into and out of a microstructure that suppresses appropriate modes of the electromagnetic quantum vacuum, an extraction of energy from the electromagnetic quantum vacuum may be accomplished. This can be done with micro Casimir cavities.

[0012] The electromagnetic quantum vacuum as a real source of energy is indicated by the Lamb shift between s and p levels in hydrogen, van der Waals forces, the Aharanov-Bohm effect, and noise in electronic circuits.

[0013] However the most important effect of the electromagnetic quantum vacuum is the existence of the Casimir Force, a force between parallel conducting plates which may be interpreted as a radiation pressure effect of electromagnetic quantum vacuum energy. Electromagnetic waves in a cavity whose walls are conducting are constrained to certain wavelengths for reasons having to do with transverse component boundary conditions on the wall surfaces. As a result, in a Casimir cavity between parallel plates there will be, in effect, an exclusion of radiation modes whose wavelengths are longer than the separation of the plates. An overpressure of electromagnetic quantum vacuum radiation on the outside then pushes the plates together. An extensive literature exists on the Casimir force and the reality of the force has moved from laboratory experimentation to micro-electro-mechanical structures (MEMS) technology both as a problem (so-called "stiction") and as a possible control mechanism.

[0014] The exclusion of modes does not begin all at once at the wavelength equivalent to the plate separation, d. Mode suppression will be strongest for wavelengths of d or greater, but will begin to occur as well for wavelengths falling in between the "stairway" d/n, with the effect diminishing as n increases. We propose to use the partial suppression of modes for wavelengths shorter than d occurring in this fashion in order to be able to employ Casimir cavities of the maximum possible physical size.

[0015] Researchers have shown that thermodynamic laws are not violated when energy is "extracted" from zero-point energy, as energy is still conserved and the second law is not violated. Cole and Puthoff have carried out and published thermodynamic analyses showing that there is no violation. Indeed, a thought experiment by Forward (1984) showed a simple, but not practical, energy extraction experiment.

[0016] In the stochastic electrodynamics (SED) interpretation of the hydrogen atom, the ground state is interpreted as effectively equivalent to a classically orbiting electron whose velocity is c/137. The orbit is stable at the Bohr radius owing to a balance between classical electromagnetic emission and absorption from the electromagnetic zero-point field. This view, first obtained by Boyer (1975) and subsequently refined by Puthoff (1987) has been further strengthened by the detailed simulations of Cole and Zou (2003, 2004) which demonstrate that the stochastic motions of the electron in this interpretation reproduce the probability density distribution of the Schrödinger wave function. Note that one apparent difference between this interpretation and that of quantum mechanics is that in quantum mechanics the 1s state of the electron is regarded as having zero angular momentum, whereas in the SED interpretation the electron has an instantaneous angular momentum of $mcr/137=h/2\pi$. However SED simulations by Nickisch have shown that the time-averaged angular momentum is zero just as in the quantum case owing to the zero-point perturbations on the orbital plane. Thus averaged over enough "orbits" this "classical electron" will fill a spherical symmetric volume around the nucleus having the same radial probability density as the Schrödinger wave function and zero net angular momentum, completely consistent with quantum behavior.

[0017] The Bohr radius of the atom in the SED view is 0.529 A (Angstroms). This implies that the wavelength of zero-point radiation responsible for sustaining the orbit is $2*\pi*0.529*137=455$ A (0.0455 microns). It is claimed that suppression of zero-point radiation at this wavelength and shorter in a Casimir cavity will result in the decay of the electron to a lower energy state determined by a new balance between classical emission of an accelerated charge and zero-point radiation at $\lambda<455$ A, where λ depends on the Casimir plate separation, d. Note that the tail end of the quantum probability density of the electron (as well as the SED simulation of Cole and Zou) extends beyond five Bohr radii, so that some change in the energy balance could be accomplished even at considerably longer wavelengths of perhaps 0.1 microns-0.2 microns

[0018] Since the frequency of this orbit is $6.6\times10^{15}\ s^{-1}$, no matter how quickly the atom is injected into a Casimir cavity the process will be a slow one as experienced by the orbiting electron. We therefore assume that the decay to a new sub-Bohr ground state will involve gradual release of energy in the form of heat, rather than a sudden optical radiation signature.

[0019] Since the binding energy of the electron is 13.6 eV, we assume that the amount of energy released in this process would be on the order of 1 to 10 eV for injection of the hydrogen atom into a Casimir cavity of d=250 A or thereabouts (and perhaps even a larger cavity as noted above). Upon exiting the cavity the electron would absorb energy from the zero-point field and be re-excited to its normal state. The energy (heat) extracted in the process comes at the expense of the zero-point field, which in the SED interpretation flows at the speed of light throughout the Universe. We are in effect extracting energy locally and replenishing it globally. Imagine extracting thimbles-full of water from the ocean. Yes, the ocean is being depleted thereby, but no practical consequences ensue.

[0020] Since naturally occurring hydrogen at standard temperature and pressure (STP) is a two-atom molecule, a dissociation process would need to precede an injection of hydrogen atoms into a Casimir cavity. We avoid this complication and take advantage of multi-electron modification by working with monatomic (noble) gases which also have the advantage of being safe and inexpensive.

[0021] We work with naturally occurring monatomic gases for three reasons:

[0022] (1) No dissociation process is required.

[0023] (2) Heavier element atoms are approximately two to four times larger than hydrogen and thus can utilize and be affected by a larger Casimir cavity which is easier to fabricate.

[0024] (3) Heavier elements have numerous outer shell electrons, several of which may be simultaneously affected by the reduction of zero-point radiation in a Casimir cavity.

[0025] The following five noble gases are potentially suitable:

[0026] He (Z=2, r=1.2 A)

[0027] Ne (Z=10, r=1.3 A),

[0028] Ar (Z=18, r=1.6 A)

[0029] Kr (Z=36, r=1.8 A)

[0030] Xe (Z=54, r=2.05 A).

[0031] All of these elements contain ns electrons. He has two is electrons. Ne has two each of 1s and 2s electrons. Ar has two each of 1s, 2s, and 3s electrons. Kr has two of each of 1s, 2s, 3s, and 4s electrons. Xe has two of each of 1s, 2s, 3s, 4s and 5s electrons.

[0032] Assuming an outermost electron which is completely shielded by the other electrons (a crude assumption), its orbital velocity would scale as $r^{-1/2}$ (the familiar Keplerian period squared proportional to semi-major axis cubed relationship) and thus λ proportional to r/v) will scale as $r^{3/2}$. If that is the case, then the larger radii translate as $r^{3/2}$ into larger Casimir cavities having an effect on the energetics of the outer electron shells. We would therefore expect that a Casimir cavity having d=0.1 microns (or perhaps even as large as one micron would have an effect on reducing the energy levels of the outermost pair of s electrons . . . and possibly also p electrons and intermediate shell s electrons as well.

[0033] It is reasonable to expect that a 0.1 microns Casimir cavity would result in a release of 1 to 10 eV for each injection of a He, Ne, Ar, Kr or Xe atom into such a cavity.

[0034] According to a Jordan Maclay, who has done theoretical Casimir cavity calculations, a long cylindrical cavity results in an inward force on the cavity. In the "exclusion of modes" interpretation of the Casimir force, this implies that a cylindrical cavity of diameter 0.1 micron would yield the desired decay of outer shell electrons and subsequent release of energy.

[0035] It is now recognized that an electromagnetic quantum vacuum field is formally necessary for atomic stability in conventional quantum theory (Milonni 1994). In the field of physics known as stochastic electrodynamics, this concept has been shown by theory and simulations to underlie the ground state of the electron in the hydrogen atom. The classical Bohr orbit is determined by a balance of Larmor emission and absorption of energy from the zero-point fluctuations of the electromagnetic quantum vacuum in SED theory. It follows that upon suppression of appropriate zero-point fluctuations the balance will be upset causing the electron to decay to a lower energy level not ordinarily found in nature with a release of energy during this transition. A Casimir cavity of the proper dimensions can accomplish this suppression of zero-point fluctuations. A Casimir cavity refers to any region in which electromagnetic modes are suppressed or restricted. Upon entering such a properly designed Casimir cavity the electron energy level will shift and energy will be released. Upon exiting the Casimir cavity the electron will return to its customary state by absorbing energy from the ambient zero-point fluctuations. This permits an energy extraction cycle to be achieved at the expense of the zero-point fluctuations. Although it has not yet been proven theoretically, a similar balance of Larmor emission and absorption of energy from the zero-point fluctuations must underlie the electron states of all atoms, not just hydrogen, permitting any atom to be used as a catalyst for extraction of zero-point energy (the energy associated with the zero-point fluctuations). An analogous process is also believed to underlie molecular bonds, yielding a similar energy extraction cycle.

[0036] The following is a list of patents that deal with related phenomena:

[0037] U.S. Pat. No. 5,018,180, Energy conversion using high charge density, Kenneth R. Shoulders. This concerns the production of charge clusters in spark discharges. It is conjectured that the electrostatic repulsion of charges is overcome in charge clusters by a Casimir-like force. This invention does not deal with energy release from atoms in Casimir cavities and is therefore not relevant to the present invention.

[0038] U.S. Pat. No. 5,590,031, System for converting electromagnetic radiation energy to electrical energy, Franklin B. Mead and Jack Nachamkin. This invention does not deal with energy release from atoms in Casimir cavities and is therefore not relevant to the present invention.

[0039] U.S. Pat. No. 6,477,028, Method and apparatus for energy extraction, Fabrizio Pinto. Proposes to vary one or more of a variety of physical factors that affect the Casimir force, or by altering any of a variety of environmental factors that affect such physical factors and thereby render a Casimir force system as non-conservative. This invention does not deal with energy release from atoms in Casimir cavities and is therefore not relevant to the present invention.

[0040] U.S. Pat. No. 6,593,566, Method and apparatus for energy extraction, Fabrizio Pinto. A method and apparatus for accelerating and a decelerating particles based on particle surface interactions. This invention does not deal with energy release from atoms in Casimir cavities and is therefore not relevant to the present invention.

[0041] U.S. Pat. No. 6,665,167, Method for energy extraction-I, Fabrizio Pinto. Similar to U.S. Pat. No. 6,477,028. This invention does not deal with energy release from atoms in Casimir cavities and is therefore not relevant to the present invention.

SUMMARY OF THE INVENTION

[0042] A system is disclosed for converting part of the energy of the electromagnetic quantum vacuum available at any point in the universe to usable energy in the form of heat, electricity, mechanical energy or other forms of power. This is accomplished using an effect on the electron configurations of atoms predicted by the theory of stochastic electrodynamics (SED). Within the context of SED theory it is predicted that the electron energy levels in atoms are determined by a balance of Larmor radiation vs. absorption of radiative energy from the electromagnetic quantum vacuum. By suppressing electromagnetic quantum vacuum energy at appropriate frequencies a change may be effected in the electron energy levels which will result in the emission or release of energy. This change in energies is analogous to a standard emission of a photon as an electron makes a transition from an excited to a lower energy state, but on a longer time scale and with the change being a continuous one rather than a "jump" from one energy level to another. Mode suppression of electromagnetic quantum vacuum radiation is known to take place in Casimir cavities. A Casimir cavity refers to any region in which electromagnetic modes are suppressed or restricted. When atoms enter into suitable micro Casimir cavities a decrease in the orbital energies of electrons in atoms will thus occur, with the effect being most pronounced for outer shell electrons. Such energy will be captured in the claimed devices. Upon emergence form such micro Casimir cavities the atoms will be re-energized by the ambient electromagnetic quantum vacuum. In this way energy is extracted locally and replenished globally from and by the electromagnetic quantum vacuum. This process may be repeated an unlimited number of times. This process is also consistent with the conservation of energy in that all usable energy does come at the expense of the energy content of the electromagnetic quantum vacuum. Two example variations of a system are disclosed that permit multiple extractions of electromagnetic quantum vacuum energy during passage of a gas through a series of micro Casimir cavities and that operate in a self-sustaining, recycling fashion. Similar effects may be produced by acting upon molecular bonds. The disclosed devices are scalable in size and energy output for applications ranging from replacements for small batteries to power plant sized generators of electricity. Since the electromagnetic quantum vacuum is thought to permeate the entire Universe, devices drawing power from the electromagnetic quantum vacuum in the fashion claimed will be effectively inexhaustible sources of power.

BRIEF DESCRIPTION OF THE DRAWINGS

[0043] The present invention may be understood by reference to the following detailed description taken in conjunction with the drawings briefly described below.

[0044] FIG. **1** is a diagrammatic illustration of a set of channels each containing a multiplicity of Casimir cavities in accordance with the present invention.

[0045] FIG. **2** is a diagrammatic illustration of a system for converting quantum vacuum energy into locally usable power in accordance with the present invention.

[0046] FIG. **3** is a diagrammatic illustration of a block of tunnels each containing a multiplicity of Casimir cavities in accordance with the present invention.

[0047] FIGS. **4**A-**4**D are diagrammatic illustrations of Casimir channels in bonded wafers in accordance with the present invention.

[0048] FIGS. **5**A-**5**C are diagrammatic illustrations of a device for oscillating a fluid though Casimir channels in accordance with the present invention.

[0049] FIGS. **6**A and **6**B are diagrammatic illustrations of a device switching the reflecting characteristics of walls of Casimir cavities in accordance with the present invention.

[0050] FIGS. **7**A and **7**B are diagrammatic illustrations of a device Casimir cavity wall spacing in accordance with the present invention.

[0051] FIG. **8** is a diagrammatic illustration of a device incorporating asymmetric Casimir cavity entry and exits in accordance with the present invention.

DETAILED DESCRIPTION

[0052] The first embodiments of this concept utilize Casimir cavities consisting of volumes through which, or in and out of which, gases flow, and which on the size scales of atoms appear as regions bounded by parallel plates of conducting material in which the plate scales are much larger than the plate separations; or by cylinders of conducting material in which the lengths of the cylinders are much larger than the diameters. It is claimed that other forms of Casimir cavity are capable of producing a similar effect, and the term Casimir cavity will be used to designate any volume capable of mode suppression of the zero-point field. The necessary condition is that the mode suppression ability of the Casimir cavity be matched to the electron energy levels in such a way as to result in a significant difference of the electron energy levels inside vs. outside the cavity.

[0053] These embodiments demonstrate the following concepts:

[0054] A method, comprising: (a) use of a device including a series of Casimir cavities and causing a specific gas to flow through the cavities, said Casimir cavities being configured and said specific gas being selected such that as the gas flows through the cavities energy is released from the gas; and (b) means for collecting at least some of said released gas.

[0055] A method, comprising: (a) providing a device including at least one Casimir cavity and causing a specific gas to enter and then exit the cavity, said Casimir cavity being configured and said specific gas being selected such that when the gas is caused to enter the cavity, energy is released from the gas; and (b) means for collecting at least some of said released energy.

[0056] A means for effecting changes in the electron configurations. A system for converting part of the energy of the electromagnetic quantum vacuum available at any point in the Universe to usable energy in the form of heat, electricity, mechanical energy or other forms of power.

[0057] A means for effecting changes in the electron configurations in the process of which energy is released.

[0058] A means for allowing the electron configurations to be re-energized by exposure to the ambient electromagnetic quantum vacuum radiation.

[0059] The use of microstructures consisting of many pairs of alternating Casimir cavities and regions in which the electromagnetic quantum vacuum radiation freely propagates.

[0060] The use of conducting strips on facing pairs of plates so that atoms go through alternating regions in which they are exposed to the full electromagnetic quantum vacuum spectrum, and regions in which part of the spectrum is blocked. The result is that they dump (or radiate) the energy difference into the local medium.

[0061] The use of spacers to separate the layer pairs.

[0062] The use of multiple conducting strips to amplify the effect (hugely).

[0063] The stacking of such plates with strips on both sides so that the top of one pair becomes the bottom of the next, each with identical conducting strips which form Casimir cavities with their partner strips in each pair.

[0064] The use of sandwiched layers of alternating conducting and non-conducting plates having micron sized thicknesses in which micron or submicron diameter holes are introduced by etching or some other method.

[0065] The stacking, co-registration and alignment of such sandwiched layers to produce many parallel Casimir tunnels having alternating Casimir and non-conducting segments.

[0066] The use of multiple segments to amplify the effect (hugely).

[0067] The use of monatomic gases as the medium in such a system.

[0068] The use of molecular gases in such a system for the purpose of modifying molecular bonds with the attendant release of energy.

[0069] A closed recycling system in which these processes take place.

[0070] Fabricatable and workable configuration and dimensions but with the claims not limited to these specific embodiments.

[0071] A means whereby the flow of gas is initiated and maintained in a closed system.

[0072] A means whereby the energy released from the electron orbital changes is converted into usable energy in the form of heat, electricity, mechanical energy or other forms of power.

Casimir Channels

[0073] This embodiment shown in FIG. **1** involves two square parallel plates **12** and **14**, 10×10 cm in size for illustration. On each one lay down 5000 conducting strips **16** that are 10 microns in width and the full 10 cm in length, separated by 10 microns non-conducting strips. Perpendicular to the strips deposit a spacer material **18** at 0.1 to 1 cm intervals with a height of 0.1 microns. Put the plates face to face and align the strips so as to form 5000 Casimir strips.

[0074] If we assume a gas flow rate of 10 cm/s parallel to the spacers and perpendicular to the strips, this would result in 1.3×10^{20} transitions/s.

[0075] An energy release of 1 to 10 eV per transition corresponds to 21 to 210 watts of energy release for the entire Casimir cavity. A stacked set of 10 or more such layers could be fabricated yielding 210 to 2100 watts for a 10×10×10 cm block.

[0076] This may be directly converted into electricity using a thermophotovoltaic process, or indirectly by using a heat exchanger. As in the previous embodiment, one means of capturing the emitted radiation is to surround the apparatus with a water bath.

[0077] The dimensions above are solely examples. The device may be scaled to both smaller and significantly larger dimensions.

[0078] The essential components of an energy generating device of this sort shown in FIG. **2** are:

[0079] (1) An array of parallel Casimir channels with conducting strips **10**

[0080] (2) A pump **22** providing continuous recycling of gas through the tunnels

[0081] (3) A means **24** for capturing the emitted energy

[0082] (4) A thermal photovoltaic, heat exchanger or other device **26** capable of converting output heat into electricity or other usable forms of power.

[0083] A desirable property of the system is its ability to radiate the accumulated energy locally and absorb it globally. Thus surprisingly the means **24** for capturing the emitted energy can capture the emitted energy without hindering the capture of the quantum vacuum energy by the gas. This is due to the fact that the vacuum field permeates all space and cannot be blocked. (Note that the reason that Casimir cavities have reduced vacuum energy modes is not that they block it, but rather that because of destructive interference they do not allow some of the electromagnetic modes to exist in their interior.) A second reason that the means **24** does not block the capture of the quantum vacuum energy is that the absorbed energy is dominantly shorter wavelength electromagnetic modes that are not absorbed by the means **24**, whereas the radiated energy can be longer wavelengths for which the means **24** has a much larger absorption coefficient. Such is the case, for example, when the means **24** comprises a water bath.

[0084] The first two components will be enclosed in sealed structure. The third and fourth components may be interior or exterior to this structure.

[0085] A variation on the above device consists of stacking plates such that the top of one pair becomes the bottom in the next pair, etc.

Casimir Tunnels

[0086] One embodiment of the concept shown in FIG. **3** is multiple, parallel, 0.1 micron diameter Casimir tunnels. If we let the length of the cylinder be 100 times the width, this results in z=10 microns for the length of the Casimir tunnel. We propose a segmented tunnel consisting of alternating conducting and non-conducting materials, each 10 microns in length. In a length of 1 cm, there could be 500 such pairs in segments, resulting in 500 energy releases events (each yielding 1 to 10 eV) for each transit of an atom through the entire 1 cm-long segmented Casimir tunnel.

[0087] Consider a one cubic cm "Casimir Block" that is built up of 10 micron thick alternating layers as shown in FIG. **3**. Assume that tunnels **32** of 0.1 micron diameter could be drilled through the cube perpendicular to the layers **34** (this is not physically possible, of course; tunnel manufacture must be done differently). Ten percent of the cross section comprises entrance to some 1.3 billion tunnels. The amount of energy released would be proportional to the flow rate of the gas through these tunnels.

[0088] A flow rate of 10 cm s^{-1} through a total cross sectional area of 0.1 cm^2 yields 1 cm^3 of gas per second flowing through the tunnels, which at STP would be 2.7×10^{19} atoms. A very simple sealed, closed-loop pumping system could maintain such a continuous gas flow. Since each atom interacts 500 times during its passage, there would be 1.3×10^{22} transitions per second in the entire cube of one cubic centimeter. An energy release of 1 to 10 eV per transition corresponds to 2150 to 21500 watts of energy release for the entire Casimir cube of segmented tunnels.

[0089] Obviously it is not possible to drill 1.3 billion tunnels having diameters of 0.1 microns. However it is feasible to use microchip technology to etch holes into the individual layers first and then assemble the stack. Extremely fine coregistration and alignment of stacks will need to be accomplished.

[0090] This may be directly converted into electricity using a thermophotovoltaic process, or indirectly by using a heat exchanger.

[0091] One means to capture the emitted energy is to surround the apparatus with a water bath. Water absorbs infrared radiation very effectively. For the wavelength range of 2 microns to 200 microns, the absorption coefficient of water is greater than 10 cm^{-1}. Therefore a layer of water that is 1 mm thick and surrounds the apparatus will be sufficient to absorb nearly all the emitted infrared radiation. The water will be heated, and that heat converted into the desired form of energy.

[0092] The dimensions above are solely examples. The device may be scaled to both smaller and significantly larger dimensions.

[0093] The essential components of an energy generating device of this sort are:

[0094] (1) An array of parallel segmented Casimir tunnels **32**

[0095] (2) A pump **22** providing continuous recycling of gas through the tunnels

[0096] (3) A means **24** for capturing the emitted energy

[0097] (4) A thermal photovoltaic, heat exchanger or other device **26** capable of converting output heat into electricity or other usable forms of power.

[0098] The first two components will be enclosed in sealed structure. The third and fourth components may be interior or exterior to this structure.

Casimir Channels in Bonded Wafers

[0099] The basic concept of the present invention is to flow gas into and out from multiple Casimir cavities. When the gas is outside of a Casimir cavity, a wide range of quantum mechanical vacuum electromagnetic modes are available to interact with the gas's atomic electronic orbital states. When the gas passes into a Casimir cavity the range of available modes is restricted and the gas sheds some of its electromagnetic energy such that this energy is available locally. When the gas once again flows out from the Casimir cavity, the gas's atomic electronic orbital state energy is recharged from quantum mechanical vacuum fields. Thus energy is harvested globally and delivered locally.

[0100] The configuration for a basic device comprising bonded wafers is shown in FIGS. **4A-4D**. A top view is shown in FIG. **4**. The device is 1 sq. cm. As seen from the south edge **41** in FIG. **4B**, it consists of two substrates **42** and **44** separated by a series of spacers which extend across the device from the south to the north side. These spacers have a height d, a width w_1, and a center-to-center spacing s_1. The thin gaps delineated by the spacers **48** extend to openings at the south edge of the device, as seen in FIG. **4B** and the north edge. As seen from the east edge in FIG. **4C**, the upper **44** and lower **42** substrates are each coated with conducting stripes **46** that extend from the east edge to the west edge. These stripes are discontinuous, such that the discontinuity occurs at each region where the stripe is intersected by a spacer **48**. These stripes have a width w_2 and a center-to-center spacing s_2. In the central region of the device there is a region of both substrates that has been removed to form a conduit **43** from close to the east edge to close to the west edge. This conduit does not extend all the way to the edges, but is instead sealed **45** at each end, as shown in FIG. **4A**. Finally, as seen in FIG. **4D**, which shows an east view of the central cross section, and in FIG. **4A**, a hole **47** extends through the upper substrate. This hole connects to the conduit **43** shown in FIGS. **4A** and **4C**. As can also be seen in FIGS. **4A** and **4D**, a connector ring **49** that surrounds the hole is affixed to the upper substrate.

[0101] For the device to function, gas tubing **28**, shown in FIG. **2**, is attached to the connector ring **49** extending from the upper substrate, forming a sealed connection. Pressurized gas flows through the tubing and the hole **47** in the upper substrate into the conduit **43** between the substrates. From the conduit **43** the gas flows from the central region through the gap between the substrates to the north and south edges. The spacers guide the gas so that it flows alternately between regions coated with the conducting stripes **46** and regions that are not coated with these stripes, until it reaches the north and south edges, at which point it escapes from the gap between the substrates. The escaped gas is captured in a surrounding enclosure, not shown, and pumped back through the tubing **28** into the hole at the top center of the device, forming a close-loop system. In this way the gas is passed through multiple Casimir cavities. The gas atoms or molecules absorb energy from the surrounding electromagnetic field when they are in the non-conducting region and then release a portion of their energy as they enter the gap between the conductive coatings, i.e., in the Casimir cavity.

[0102] The apparatus is surrounded by a means **24** to capture the released energy, such as a water bath, shown in FIG. **2**. Water absorbs infrared radiation very effectively. For the wavelength range of 2 microns to 200 microns, the absorption coefficient of water is greater than 10 cm^{-1}. Therefore a layer of water that is 1 mm thick and surrounds the apparatus is sufficient to absorb a large proportion of the emitted infrared radiation, providing thermal energy to heat the water. That energy can be used directly as heating source, or converted into the desired form of energy, by means **26** well known to those skilled in the art.

[0103] The materials and dimensions in the preferred embodiment are as follows. The upper **44** and lower **42** substrates are sapphire, which is transparent to much of the ambient electromagnetic spectrum, is thermally conductive, and is rigid and robust. The thickness of each substrate is 250 microns. The conducting regions **46** are formed by standard photolithography known to those skilled in the art. The width of each conducting stripe, w_2, is 2 microns, and separated by a 2 micron nonconducting region, to form a center-to-center spacing **52** of 4 microns. The stripe has gaps where the spacers **48** are to be formed. The conductive coating **46** is platinum, having a thickness of 40 nm. The spacers **48** consist of silicon dioxide, deposited and patterned by standard means known by those skilled in the art. The total spacer height, d, is 200 nm, its width, w_1, is 5 microns, and the center-to-center spacing, s_1, is 0.5 mm. The spacers are formed by depositing 100 nm thick layers on each substrate, and then joining them. The central conduit regions **43** are cut into the substrates using a standard diamond saw. The cuts are 100 microns in width and 50 microns in depth, forming a conduit that is approximately a 100 micron square. The hole **47** drilled through the upper substrate has a diameter of 1 mm, and is surrounded by a ring having a diameter of 2.5 mm. The ring **49** is affixed to the upper substrate by epoxy. The substrates are pressure bonded together by direct bonding (Plöl, 1999), with the bond forming between the silicon dioxide spacers layers on each substrate.

[0104] The steps in the device fabrication that are not described explicitly are well known to those skilled in the art.

[0105] Following the calculations presented in the background section, the power produced by a single such device is estimated to be between 1 and 10 watts for an input pressure of 8 atmospheres.

[0106] Pumping gas through the Casimir pores requires power. We examine how much power is required, as a check that it is not more than is produced by the device. Consider a Casimir block that contains 200 nm diameter pores over a 1 cm^2 area, having a thickness of 1 cm and a porosity of 0.25. We find the pressure and power required to produce a flux of 1 cm^3 per second at standard temperature and pressure (STP):

[0107] According to FIG. **10**(*a*) in a paper by Roy et al. (1993) a pressure drop of 760 torr (equal to one atmosphere) results from a flow of approximately 5 mol/m^2-s through a thickness of 60 microns, which corresponds to a gas velocity of 10 cm/s. Reducing the velocity by a factor of ten, making the appropriate unit conversions and multiplying the result by the thickness ratio of 1 cm (10^4 microns) divided by 60 microns gives the result that a pressure of 1700 Pa, corresponding to 17 atmospheres, is required to produce the desired gas flow. Multiplying this by the gas flux of 1 cm^3 s^{-1} results in a required power of 1.7 milliwatts. These results are only approximate, as temperature and structural variations through the Casimir pores are expected to produce resistance which will then require a somewhat greater pressure. In any case the required power of approximately 1.7 milliwatts is much lower than our estimate of 2.2 to 22 kilowatts of power release, and so much more power is produced than is used to produce the gas flow.

[0108] It is to be understood that the dimensions and materials can be varied greatly and still be part of this invention. The following is a list of some such variations, but it is far from exhaustive:

[0109] i. The substrates may be other insulating or partially conducting materials, such as silicon, glass, ceramic, plastic, etc.

[0110] ii. The conducting stripes can be formed of other conductors, such as copper, aluminum, gold, sliver, silicides, transparent conductors such as indium tin oxide, etc.

[0111] iii. Instead of depositing the stripes so that they protrude from the surface and potentially interfere with the gas flow, they may be recessed, either by etching recesses into which the conductors are deposited, or by using planarization techniques to coat an insulating layer between the stripes, using techniques well known in the industry.

[0112] iv. The spacer materials can be formed from polymers used, for example, as photoresist and electron-beam resist, from metals, and other materials.

[0113] v. Instead of depositing spacers they may be formed by the etching of one or both of the substrates to form grooves.

[0114] vi. The spacer height may be from 1 nm to many microns.

[0115] vii. The substrates may be bonded by pressure bonding or the use of adhesives, such as cyanoacrylics.

[0116] viii. The dimensions of the overall structure may be varied from the distance between a single pair of spacers and conductor/nonconductor region to large plates that are many meters in width.

[0117] ix. The individual devices may be sandwiched together to form thick structures. For example, in place of the 250 micron thick substrates, micro-sheet having a thickness of 50 microns or far less may be used so that dense structures are formed.

[0118] x. The working fluid may be a wide variety of gases, in addition to the noble gases described earlier, so that all mentions of gas atoms may be extended to molecules of various types.

[0119] xi. The working fluid may be a liquid, so that all mentions of gases and gas atoms may be extended to liquids of various types. For operation within approximately of 100° C., one possible liquid is ethylene glycol. For high temperature operation, the liquid can be sodium.

[0120] xii. Micro-motors formed using micro-electro-mechanical systems (MEMS) technology can be used to pump the gas through the channels.

[0121] xiii. The Casimir cavities may be composed of carbon nanotubes.

[0122] xiv. The pattern may be formed using self-assembled layers.

[0123] xv. The device may incorporate a naturally formed structure. For example, diatom shells (Goho, 2004a) consist of silicon dioxide patterned with features, including holes, that are tens of nanometers in size. They can be coated as needed with conductors to form Casimir cavities.

[0124] xvi. The water bath may be replaced with any other material or device that absorbs substantially the released energy wavelengths. Such materials include glass, organic polymers, thermophotovoltaic devices, among many possibilities known to those skilled in the art.

[0125] xvii. Rather than surrounding the entire apparatus, the absorbing material may be placed in the apparatus, for example coating the channels through which the gas flows. Such placement can allow the absorber to reside within roughly an emission wavelength of the gas that is releasing the energy.

Gas Oscillating Through Casimir Channels

[0126] The device described in the previous embodiment exposes the gas atoms to a very large number of transitions between Casimir cavity regions (between conducting layers) and exposed regions (without the conducting layers) by pumping them across multiple transitions. Instead of pumping gas through the device, gas atoms can simply be oscillated back and forth between Casimir cavity and exposed regions.

[**0127**] A simple way to visualize this, but not necessarily the most efficient working device, is to consider the device of FIGS. **4**A-**4**D, but with the gaps sealed at the north and south edges. Instead of connecting to tubing via the connector ring, the ring is sealed with a thin metal diaphragm. Before sealing the device it is filled with the desired working gas. An ultrasonic transducer is then mated to the diaphragm. When the ultrasonic transducer is powered, it rapidly compresses and decompressed the gas, causing it to oscillate back and forth between Casimir and exposed regions.

[**0128**] A vertical oscillatory flow device is shown in FIGS. **5**A-**5**C. FIG. **5**A shows a top view, in which many small holes **54** are formed in the substrate surface. The device is surrounded by a connector ring **58**. A magnified cross section of the holes is shown in FIG. **5**B. The holes **54** have a diameter d, a center-to-center spacing s, a depth t_2, and the thickness of a conducting region **56** at the surface is t_1. A central cross section of the entire device is shown in FIG. **5**C. It shows the substrate (holes and conducting layer not shown), the connector ring at the periphery, and a thin diaphragm **57** attached to the top of the connector ring.

[**0129**] The gap and holes are filled with the chosen working gas **59**. An ultrasonic transducer or other source of high frequency vibrations is placed in contact with the diaphragm **57** and powered. This produces gas pressure oscillations that force gas atoms past the Casimir region **55** formed at the top of each hole, alternately in upward and downward directions. Instead of a single conducting layer at the top, multiple alternating conducting and non-conducting layers can be formed at the top of the holes, to multiply the effect. As in the embodiment of FIGS. **4**A-**4**D, the apparatus is surrounded by a means for absorbing the released energy, such as a water bath **24**.

[**0130**] The device is fabricated as follows. The conducting layer **56** is deposited using vacuum deposition, such as sputtering, or from a liquid by anodic or electroless deposition. The layers are patterned by methods known to those skilled in the art, such as electron-beam lithography or photolithography. Alternatively, the holes **54** can be formed using self-assembled monolayers to create the lithography mask, as known to those skilled in the art. The holes are etched to a high aspect ratio, e.g., ratio of depth-to-diameter of 20, such as by ion milling. The outer ring **58** is attached using epoxy, the region is filled with the desired working gas **59**, and the diaphragm **57** is attached with epoxy.

[**0131**] The materials and dimensions in the preferred embodiment are as follows. The substrate **52** is sapphire, and has diameter of 2.54 cm and a thickness of 250 microns. The conducting layer **56** is aluminum, of thickness t_1 of 1 micron. The hole **54** depth t_2 is 4 microns. The hole diameter d is 0.2 microns and center-to-center spacing s is 0.3 microns.

[**0132**] It is to be understood that the shape, dimensions, modulation techniques and materials can be varied greatly and still be part of this invention. The following is a list of some such variations, but it is far from exhaustive:

[**0133**] i. The Casimir cavities may be composed of carbon nanotubes.

[**0134**] ii. The working fluid may be a wide variety of gases, in addition to the noble gases described earlier, so that all mentions of gas atoms may be extended to molecules of various types.

[**0135**] iii. The working fluid may be a liquid, so that all mentions of gases and gas atoms may be extended to liquids of various types. For operation of up to approximately 100° C., one possible liquid is ethylene glycol. For high temperature operation, the liquid can be sodium.

[**0136**] iv. Instead of actively causing the gas atoms to oscillate into and out from the Casimir cavity regions, the oscillations can result from ambient thermal vibrations (e.g., Brownian motion).

[**0137**] v. The configuration of the device can be similar to that of the MEMS device of FIGS. **7**A and **7**B (described as part of a later embodiment), such that the working gas is pushed back and forth between the left-hand and right-hand regions.

[**0138**] vi. The pattern may be formed using self-assembled layers.

[**0139**] vii. The device may incorporate a naturally formed structure. For example, diatom shells consist of silicon dioxide patterned with features, including holes, that are tens of nanometers in size. They can be coated as needed with conductors to form Casimir cavities.

[**0140**] viii. The pumping can be driven by a naturally occurring mechanism. For example, some yeast cell have been found to naturally vibrate at 1.6 kHz (Goho, 2004b). This could be used to cause a gas to oscillate back and forth between Casimir cavity and exposed regions.

[**0141**] ix. The water bath may be replaced with any other material or device that absorbs substantially the released energy wavelengths. Such materials include glass, organic polymers, thermophotovoltaic devices, among many possibilities known to those skilled in the art.

[**0142**] x. Rather than surrounding the entire apparatus, the absorbing material may be placed in the apparatus, for example coating the channels through which the gas flows. Such placement can allow the absorber to reside within roughly an emission wavelength of the gas that is releasing the energy.

Casimir Cavities in Flexible Polymer

[**0143**] Rather than moving the working gas by flowing it (FIGS. **4**A-**4**D) or vibrating it into and out of a Casimir cavity (FIGS. **5**A-**5**C), the cavity wall characteristics can be switched, which results in a shift in the cavity's allowed modes. This produces the same result of tapping vacuum electromagnetic energy that the flowing gas device of the embodiment of FIGS. **4**A-**4**D produces. One way to accomplish this is to put the working gas into gaps formed in flexible photonic crystals. A photonic crystal blocks and passes bands of electromagnetic radiation, where the band wavelength ranges depend upon the material properties and spacing of small repeated structures. A flexible photonic crystal can be formed by embedding an array or rigid objects, such as silicon pillars, in a thin film of flexible polymer. The electromagnetic (or optical) properties of such two-dimensional slab photonic crystal structures is well known to those skilled in the art (Park, 2002).

[0144] FIGS. **6**A and **6**B show such a photonic crystal device. FIG. **6**A is a top view, showing metal supports **62** at both ends of a polymer film **64**. The rigid pillars that form the phonic crystal are buried in the polymer. As the film is stretched in the plane of the paper, the pillar spacing in the plane normal to the paper is decreased, which changes the electromagnetic passband. FIG. **6**B is an edge view showing the supports **62**, the polymer film **64**, and gaps in the film that are filled with the working gas **69**. (For clarity, the pillars are not shown.) The gap size is sufficiently narrow to produce a significant Casimir effect, e.g., 200 nm. The length or width need to be sufficiently small to maintain the narrow gap, e.g., 1 micron. The stretching takes place by attaching one support to a stationary object and attaching the other support to a vibrator, such as a piezoelectric crystal, which itself may be attached on its opposing side to another stationary support. As in the embodiment of FIGS. **4**A-**5**D, the apparatus is surrounded by a means for absorbing the released energy, such as a water bath **24**.

[0145] It is to be understood that the shape, dimensions, modulation techniques and materials can be varied greatly and still be part of this invention. The following is a list of some such variations, but it is far from exhaustive:

[0146] i. Instead of stretching the polymer, it can be modulated with an acoustic signal through the air, or through a liquid that surrounds it.

[0147] ii. Instead of stretching the polymer, it can be modulated with an ambient thermal vibrations. As the working gas and the structure heats up, the vibrations increase.

[0148] iii. The polymer embedded with rigid pillars may be formed into small spheres that are filled with the working gas. These spheres can fill or partially fill a volume in which the pressure is modulated, either by enclosing the volume and modulating the pressure in the entire volume, by passing an acoustic signal through the volume, or by thermal vibrations. This modulation causes the passband of the photonic crystal that surrounds the working gas to vary. Although the shape of the device is substantially different from that of FIGS. **6**A-**6**B, the function is the same.

[0149] iv. The working fluid may be a wide variety of gases, in addition to the noble gases described earlier, so that all mentions of gas atoms may be extended to molecules of various types.

[0150] v. The working fluid may be a liquid, so that all mentions of gases and gas atoms may be extended to liquids of various types. For operation of up to approximately 100° C., one possible liquid is ethylene glycol. For high temperature operation, the liquid can be sodium.

[0151] vi. The water bath may be replaced with any other material or device that absorbs substantially the released energy wavelengths. Such materials include glass, organic polymers, thermophotovoltaic devices, among many possibilities known to those skilled in the art.

[0152] vii. Rather than surrounding the entire apparatus, the absorbing material may be placed in the apparatus, for example in the polymer film through which the gas flows. Such placement can allow the absorber to reside within roughly an emission wavelength of the gas that is releasing the energy.

Modulating Casimir Cavity Wall Spacing

[0153] Rather than moving the working gas by flowing it (FIGS. **4**A-**4**D), vibrating it into and out of a Casimir cavity (FIGS. **5**A-**5**C), or switching the characteristics of walls of the cavity to change the passbands (FIGS. **6**A and **6**B), the spacing between the cavity walls can be modulated. This produces the same result of tapping zero point energy that the flowing gas device of the previous embodiments produce. One way to accomplish this is to put the working gas into gaps formed in micro-electro-mechanical systems (MEMS).

[0154] MEMS technology makes use of semiconductor lithography techniques to build miniature mechanical devices. The Casimir effect has already been found to be in evidence in MEMS devices. In 2001, Chan and co-workers at Bell Labs Lucent Technologies first demonstrated the effect of the Casimir force in a MEMS device. A gold coated sphere was brought close to a MEMS seesaw paddle, consisting of a polysilicon plate suspended above a substrate on thin torsion rods. The Bell Labs researchers demonstrated the effect of the Casimir force in rocking the plate.

[0155] In the current invention we make use of MEMS technology to modulate the spacing between Casimir cavity walls. (Note that we are not making use of the Casimir force to change this spacing, as was done in the Bell Labs demonstration.) The basic MEMS device used to accomplish this is shown in FIGS. **7**A and **7**B. A side view is shown in FIG. **7**A. Two conducting electrodes **76** are shown on the substrate. A pivoting polysilicon plate **74** is shown suspended above the substrate **72**. A conducting layer **77** is formed on the underside of this plate. A top view is shown in FIG. **7**B. The pivoting plate **74** forms the central rectangular region, which is surrounded by a gap **73**. The pivoting arm **75** connects this plate to the surrounding region at the top and bottom of the rectangle. As in the earlier embodiments, the apparatus is surrounded by a means **24** for absorbing the released energy, such as a water bath

[0156] The device functions as follows. The working gas fills the region between the pivoting plate **74** and the substrate **72**. A voltage is applied first between the pivoting plate and the left-hand electrode. This causes the distance between the left side of the plate and the substrate to diminish, thereby changing the dimensions of the Casimir cavity formed by these two surfaces. Then the voltage is instead applied between the pivoting plate and the right-hand electrode. This causes the plate to pivot, such that the distance between the right side of the plate and the substrate diminishes, thereby changing the dimensions of the Casimir cavity formed by these two surfaces. The voltage is switched alternately between these two electrodes, causing the plate to oscillate back and forth. The oscillating action is greatly enhanced by the torsion of the pivots, so that very little energy is required to maintain the oscillation.

[0157] The techniques to fabricate such a MEMS device is well known to those skilled in the art.

[0158] It is to be understood that the shape, dimensions, modulation techniques and materials can be varied greatly and still be part of this invention. The following is a list of some such variations, but it is far from exhaustive:

[0159] i. Instead of using a MEMS device, the Casimir cavity can be formed between a substrate and a suspended conducting sheet. A similar technology has been used to form electrostatic acoustic speakers, albeit with larger spacings.

[0160] ii. Gaps can be formed in a polymer, with both sides of the gap coated with a conductor and the gap filled with a working gas. The polymer can then be stretched, as in the embodiment of FIGS. **6**A and **6**B, such that the spacing of the Casimir cavity formed by the two conductors is modulated. A figure of this would appear much like that depicted in FIG. **6**B

[0161] viii. Instead of stretching the polymer, it can be modulated with an acoustic signal through the air, or through a liquid that surrounds it.

[0162] ix. Instead of stretching the polymer, it can be modulated with an ambient thermal vibrations. As the working gas and the structure heat up, the vibrations will increase.

[0163] x. The polymer coated on its interior surface with a conductor may be formed into small spheres that are filled with the working gas. These spheres can fill a volume in which the pressure is modulated, either by enclosing the volume and modulating the pressure in the entire volume, by passing an acoustic signal through the volume, or by thermal vibrations. This modulation causes the spacing of the Casimir cavity in which the working gas is contained to vary. Although the shape of the device is substantially different from that of FIGS. **7**A and **7**B, the function is the same.

[0164] xi. The working fluid may be a wide variety of gases, in addition to the noble gases described earlier, so that all mentions of gas atoms may be extended to molecules of various types.

[0165] xii. The working fluid may be a liquid, so that all mentions of gases and gas atoms may be extended to liquids of various types. For operation of up to approximately 100° C., one possible liquid is ethylene glycol. For high temperature operation, the liquid can be sodium.

[0166] xiii. The water bath may be replaced with any other material or device that absorbs substantially the released energy wavelengths. Such materials include glass, organic polymers, thermophotovoltaic devices, among many possibilities known to those skilled in the art.

[0167] xiv. Rather than surrounding the entire apparatus, the absorbing material may be placed in the apparatus, for example coating the substrate and cap of the region containing the gas. Such placement can allow the absorber to reside within roughly an emission wavelength of the gas that is releasing the energy.

[0168] We note that the MEMS device of FIGS. **7**A and **7**B can also be used to move the working gas back and forth between the left-hand and right-hand regions. This function is consistent with the embodiment of FIGS. **5**A-**5**C, in which the working gas is vibrated into and out of a Casimir cavity.

Assymetric Casimir Cavity Entry and Exits Including Absorbing Means

[0169] As a prelude to this embodiment, we review the processes involved in the present invention. A general concept of this entire invention is that a gas that is in equilibrium with the ambient electromagnetic modes, which include the vacuum field (also known as the zero point field), is caused to enter a Casimir cavity. For the purposes of this entire invention a Casimir cavity is defined as any region in which the electromagnetic modes are restricted. Upon approaching this region, the electromagnetic modes that the space supports are restricted and the energy of the electron orbitals of the gas atoms is reduced. As a consequence of this reduction the excess energy is emitted and absorbed by the apparatus, providing heat energy. By the time the atoms are in the Casimir cavity, nearly all the excess energy has been radiated (unless the gas flow is extremely fast). The gas atoms pass through the Casimir cavity, and upon emerging from this region to a region that supports a broader range of electromagnetic modes, the energy of the electron orbitals of the gas atoms is again allowed to rise to its previous value. The compensation for the energy deficit is provided from the ambient electromagnetic modes.

[0170] One of the tenets of the current invention is that excess energy released when the gas approaches the Casimir cavity is delivered locally and that the energy deficit that must be compensated for when it emerges from the cavity is supplied from global sources. In this way the ambient electromagnetic field is tapped to provide usable energy. There may be conditions in which it is possible that the excess energy release and the deficit energy supply are both local, in which case no net energy is provided. Similarly, there may be conditions in which it is possible that the excess energy release and the deficit energy supply are both global, in which case again no net energy is provided. To avoid these possibilities, we provide an asymmetry in the apparatus to ensure that the excess energy is released locally and that the energy deficit is supplied globally.

[0171] The concept of embodiment is shown in FIG. **8**. This figure depicts a channel **88**, similar to that shown in some of the earlier embodiments. Gas is constricted between two substrates **82** and **83** and flows through the channel in the direction of the arrows. As in the previous cases, gas flows from a region in which the substrate is not coated **87** with a conducting layer to a region in which it is **86**. The difference here is that an intermediate region **84** is provided in which the substrates are coated with an absorbing layer. This absorbing region absorbs the excess energy that is radiated from the atoms as they approach the Casimir cavity (conducting) region. The absorbing region is not substantially conducting, and therefore does not substantially restrict the electromagnetic modes that are supported in the region. Upon exiting the Casimir cavity (conducting) region, the atoms pass immediately into another region with no absorbing region **87**. Thus upon approaching the Casimir cavity the atoms are forced to deliver their excess energy locally because it is absorbed by the absorbing region **87**. Upon emerging from the Casimir cavity the gas atoms are forced to supply their energy deficit non-locally, i.e., globally, because there is no local source for this energy.

[0172] As an option, a further aspect of this invention is to situate the absorbing region within roughly one emission wavelength of the gas atoms at the time that they are emitting. No such layer is provided within such a distance when the gas atoms emerge from the Casmir cavity and are supplied with energy. The substrate is chosen such that it does not absorb the emission wavelengths.

[0173] The absorbing layers may comprise glass (amorphous silicon dioxide, usually with impurities), and the substrate may comprise sapphire. The glass has a much broader absorption band in the far infrared than does the sapphire. A wide range of other materials may be provided to form the absorbing layers and non-absorbing or less absorbing substrate. Such materials are known to those skilled in the art, and are available in tables and handbooks.

[0174] The sequence of regions depicted in FIG. **8** may be repeated to form the sort of multiply striped structure described in the embodiment of FIGS. **4**A-**4**D.

[0175] The dimensions of the channel and the apparatus are approximately the same as those of embodiment of FIGS. **4**A-**4**D. Similarly the attachments to provide for gas flow, the spacers, and other aspects of the apparatus are similar to those described in embodiment of FIGS. **4**A-**4**D. The conducting layer length is chosen so that the emerging atoms do not have substantial access to radiation emitted from the absorbing regions. Note that, unlike embodiment of FIG. **2**, it is not necessary to surround the apparatus with a means for absorbing the released energy **24**, such as a water bath.

[0176] The device fabrication is not described explicitly as it is well known to those skilled in the art.

[0177] It is to be understood that the dimensions and materials can be varied greatly and still be part of this invention. The following is a list of some such variations, but it is far from exhaustive:

[0178] i. The substrates may be other insulating or partially conducting materials, such as silicon, glass, ceramic, plastic, etc.

[0179] ii. The conducting stripes can be formed of other conductors, such as copper, aluminum, gold, sliver, silicides, transparent conductors such as indium tin oxide, etc.

[0180] iii. The stripes may be recessed in the substrate or they protrude from the surface.

[0181] iv. The individual devices may be sandwiched together to form thick structures. For example, in place of the 250 micron thick substrates, micro-sheet having a thickness of 50 microns or far less may be used so that dense structures are formed.

[0182] v. The working fluid may be a wide variety of gases, in addition to the noble gases described earlier, so that all mentions of gas atoms may be extended to molecules of various types.

[0183] vi. The working fluid may be a liquid, so that all mentions of gases and gas atoms may be extended to liquids of various types. For operation within approximately of 100° C., one possible liquid is ethylene glycol. For high temperature operation, the liquid can be sodium.

[0184] vii. Micro-motors formed using micro-electro-mechanical systems (MEMS) technology can be used to pump the gas through the channels.

[0185] viii. The Casimir cavities may be composed of carbon nanotubes.

[0186] ix. The pattern may be formed using self-assembled layers.

[0187] The device may incorporate a naturally formed structure. For example, diatom shells consisting of silicon dioxide patterned with features, including holes, that are tens of nanometers in size. To the extent necessary, these can be coated as needed with conductors to form Casimir cavities.

REFERENCES

[0188] Boyer, T. H. 1975, Random Electrodynamics: The Theory of Classical Electrodynamics with Classical Zero-Point Radiation Field, Phys. Rev. D, 11, 790.

[0189] Cole, D. C. and Puthoff, H. E. 1993, Extracting energy and heat from the vacuum, Phys. Rev. E, **48**, **2**, **1562**.

[0190] Cole, D. C. and Zou, Yi 2003, Quantum Mechanical Ground State of Hydrogen Obtained from Classical Electrodynamics, Physics Letters A, Vol. 317, No. 1-2, pp. 14-20 (13 Oct. 2003), quant-ph/0307154.

[0191] Cole, D. C. and Zou, Yi 2004, Analysis of Orbital Decay Time for the Classical Hydrogen Atom Interacting with Circularly Polarized Electromagnetic Radiation, Phys. Rev. E. 69 (1), 016601, pp. 1-12 (2004).

[0192] Forward, R. 1984, Extracting electrical energy from the vacuum by cohesion of charged foliated conductors, Phys. Rev. B, **30**, **4**, **1700**.

[0193] Goho, A., "Diatom Menagerie," Science News, Vol. 166, Jul. 17, 2004a, pp. 42-44, and references mentioned therein.

[0194] Goho, A., "Rattle and Hum; Molecular machinery makes yeast cells purr," Science News, Vol. 166, Aug. 21, 2004b, p. 116, and references therein.

[0195] Milonni, P. W. The Quantum Vacuum: An Introduction to Quantum Electrodynamics, Sect. 3.3, Academic Press, 1994.

[0196] Park, W., and Summers, C. J., "Extraordinary refraction and dispersion in 2D photonic crystal slabs", Opt. Lett., Vol 27, (2002) p. 1397.

[0197] Plöl, A. and Krauter, G., "Wafer direct bonding: tailoring adhesion between brittle materials", Mater. Sci. Eng. R, Vol 25 (1999) p. 1.

[0198] Puthoff, H. E., Ground State of Hydrogen as a Zero-Point-Fluctuation-Determined State, Phys. Rev. D 35, 3266, 1987.

[0199] Roy, Subrata et al., "Modeling gas flow through microchannels and nanopores," Journal of Applied Physics, Vol. 93, No. 8, Apr. 15, 2003, pp. 4870-4879.

What is claimed is:

1. A system for extracting and collecting electromagnetic radiation from the ambient surroundings, comprising:

(a) a supply of fluid characterized by its ability to (i) take in electromagnetic radiation from the ambient surroundings and (ii) release at least some of said energy when the fluid is caused to pass into a Casimir cavity;

(b) a first arrangement configured to collect at least some of the electromagnetic radiation released by said fluid;

(c) a second arrangement including means defining a given path for containing said fluid along said path;

(d) a third arrangement including a Casimir cavity positioned within said given path and cooperating with said second arrangement such that said fluid is caused to pass into and out of the cavity as the fluid is contained along said given path, said Casimir cavity being positioned in sufficient communication with the ambient surroundings and with said first arrangement so as to (i) cause said fluid containing electromagnetic energy taken from the ambient surroundings to release at least some of said energy to said first arrangement when the fluid passes into said cavity and (ii) to again take in electromagnetic energy from the ambient surroundings when the fluid passes out of said cavity.

2. A system according to claim 1 wherein said second arrangement is configured such that said fluid is caused to flow along said path into and out of said Casimir cavity.

3. A system according to claim 1 wherein said second and third arrangements are configured such that said Casimir cavity is caused to move with respect to said fluid such that the fluid is in turn caused to pass into and out of said Casimir cavity.

4. A system according to claim 1 wherein said means defining said given path defines a closed passageway for containing said fluid and wherein said second arrangement is configured such that the same fluid is caused to cycle into and out of said Casimir cavity.

5. A system according to claim 4 wherein said passageway defines a looped path and wherein said second arrangement includes a mechanism configured to cause said fluid to flow around said path through said passageway into and out of said Casimir cavity.

6. A system according to claim 4 wherein said second arrangement includes a mechanism for causing said fluid to flow back and forth through said passageway into and out of said Casimir cavity.

7. A system according to claim 1 wherein said fluid is a gas.

8. A system according to claim 7 wherein said gas is a monatomic gas.

9. A system according to claim 7 wherein said gas is a molecular gas.

10. A system according to claim 1 wherein said first arrangement includes a container of material for absorbing electromagnetic energy, said absorbing material surrounding at least said Casimir cavity.

11. A system according to claim 6 wherein said absorbing material is a liquid.

12. A system according to claim 11 wherein said liquid material is water.

13. A system according to claim 3 wherein said third arrangements configured so as to cause said Casimir cavity to move back and forth between first and second spaced apart positions.

14. A system according to claim 1 wherein said Casimir cavity includes opposing walls and wherein said third arrangement is configured so as to cause the position of said Casimir cavity walls to move back and forth between first and second spaced positions.

15. A system for extracting and collecting electromagnetic energy from the ambient electromagnetic quantum vacuum, comprising:

(a) a first arrangement defining at least one Casimir cavity configured to cause gas containing electromagnetic energy obtained from the ambient electromagnetic quantum vacuum to release at least some of said energy when said gas is passed into said cavity;

(b) a second arrangement located in the ambient electromagnetic quantum vacuum and including a source of said gas and a mechanism cooperating with said first arrangement so as to cause said gas to pass from the ambient electromagnetic quantum vacuum into said Casimir cavity and then out of said cavity and back into the ambient electromagnetic quantum vacuum, whereby the gas when passing into said Casimir cavity releases at least some of its energy and then, upon passing back into the ambient electromagnetic quantum vacuum, again takes in electromagnetic energy from the ambient electromagnetic quantum vacuum, said means and said first arrangement cooperating with one another such that said fluid passes into and out of said Casimir cavity by relative movement between the cavity and gas; and

(c) a third arrangement for capturing at least some of the electromagnetic energy released by said fluid, said third arrangement including means located in a position with respect to said Casimir cavity such that at least some of the electromagnetic energy released by said gas is captured by said absorber.

16. A system, comprising:

(a) a first arrangement including a number of Casimir cavities, each of which is configured to cause fluid containing electromagnetic energy obtained from the ambient surroundings to release at least some of said energy when said fluid is passed into said cavity;

(b) a second arrangement located in the ambient surroundings and including a source of said fluid and means cooperating with said first arrangement for causing said fluid to pass from the ambient surroundings into each of said Casimir cavities and then out of the cavity and back into the ambient surroundings, whereby the fluid when passing into said Casimir cavities releases at least some of its energy and then, upon passing back into the ambient surroundings, again takes in electromagnetic energy from the ambient surroundings, said means and said first arrangement cooperating with one another such that said fluid passes into and out of said Casimir cavities by relative movement between the cavities and fluid; and

(c) a third arrangement for capturing at least some of the electromagnetic energy released by said fluid.

17. A system according to claim 16 wherein said means includes at least one fluid passageway extending from the ambient surroundings into and though said Casimir cavities and back into the ambient surroundings and wherein said Casimir cavities are defined by a series of conducting strips located within said passageway, said series of conducting strips including a first group of spaced apart strips located on one side of the passageway and a second groups of spaced apart strips on a opposite side of said passageway in alignment with respective strips of said first group, each of said aligned pair of strips being positioned relative to one another to produce a Casimir cavity.

18. A method, comprising:

(a) providing a first arrangement defining at least one Casimir cavity configured to cause fluid containing electromagnetic energy obtained from the ambient surroundings to release at least some of said energy when said fluid is passed into said cavity;

(b) providing a source of said fluid;

(c) causing said fluid to pass from the ambient surroundings into said Casimir cavity and then out of said cavity and back into the ambient surroundings such that the fluid when passing into said Casimir cavity releases at least some of its energy and then, upon passing back into the ambient surroundings, again takes in electromagnetic energy from the ambient surroundings, said fluid being cause to pass into and out of said Casimir cavity by relative movement between the cavity and fluid; and

(c) capturing at least some of the electromagnetic energy released by said fluid.

19. A method of extracting and collecting electromagnetic radiation from the ambient surroundings, comprising:

(a) providing a supply of fluid characterized by its ability to (i) take in electromagnetic radiation from the ambient surroundings and (ii) release at least some of said energy when the fluid is caused to pass into a Casimir cavity;

(b) providing a first arrangement configured to collect at least some of the electromagnetic radiation released by said fluid;

(c) providing a second arrangement including means defining a given path for containing said fluid along said path;

(d) providing a third arrangement including a Casimir cavity positioned within said given path;

(e) causing said fluid to pass into and out of the cavity as the fluid is contained along said given path; and

(f) positioning said Casimir cavity in sufficient communication with the ambient surroundings and with said first arrangement so as to (i) cause said fluid containing electromagnetic energy taken from the ambient surroundings to release at least some of said energy to said first arrangement when the fluid passes into said cavity and (ii) to again take in electromagnetic energy from the ambient surroundings when the fluid passes out of said cavity.

20. A system, comprising:

(a) a first arrangement defining at least one mechanism designed to cause the atoms and molecules making up a given fluid containing electromagnetic energy obtained from the ambient surroundings to change in configuration in a way which releases at least some of said energy when said fluid is passed into said mechanism;

(b) a second arrangement located in the ambient surroundings and including a source of said fluid and means cooperating with said first arrangement for causing said fluid to pass from the ambient surroundings into said mechanism and then out of said mechanism and back into the ambient surroundings, whereby the fluid when passing into said mechanism releases at least some of its energy and then, upon passing back into the ambient surroundings, again takes in electromagnetic energy from the ambient surroundings, said means and said first arrangement cooperating with one another such that said fluid passes into and out of said mechanism by relative movement between the mechanism and fluid; and

(c) a third arrangement for capturing at least some of the electromagnetic energy released by said fluid.

* * * * *

US 20110006740A1

(19) **United States**

(12) **Patent Application Publication**

Juchnowycz

(10) **Pub. No.: US 2011/0006740 A1**

(43) **Pub. Date: Jan. 13, 2011**

(54) **APPARATUS AND METHOD FOR CAPTURING COSMIC BACKGROUND RADIATION AND CONVERTING THE SAME TO ELECTRICITY**

(76) Inventor: **Stefan Juchnowycz**, Hendersonville, NC (US)

Correspondence Address:
DAVID M. CARTER
CARTER, SCHNEDLER & WARNOCK, P.A.
P.O. BOX 2985
ASHEVILLE, NC 28802 (US)

(21) Appl. No.: **12/821,189**

(22) Filed: **Jun. 23, 2010**

Related U.S. Application Data

(60) Provisional application No. 61/224,255, filed on Jul. 9, 2009, provisional application No. 61/235,756, filed on Aug. 21, 2009.

Publication Classification

(51) **Int. Cl.**
H02N 11/00 (2006.01)
H02J 7/32 (2006.01)

(52) **U.S. Cl.** **320/137**; 307/151

(57) **ABSTRACT**

There is provided an apparatus for capturing cosmic background radiation and for converting cosmic background radiation into electricity. An antenna is configured so as to capture cosmic background radiation. An electrostatic electron multiplier is connected to the antenna. A high voltage power supply is connected to the electrostatic electron multiplier whereby cosmic background radiation is converted to electricity.

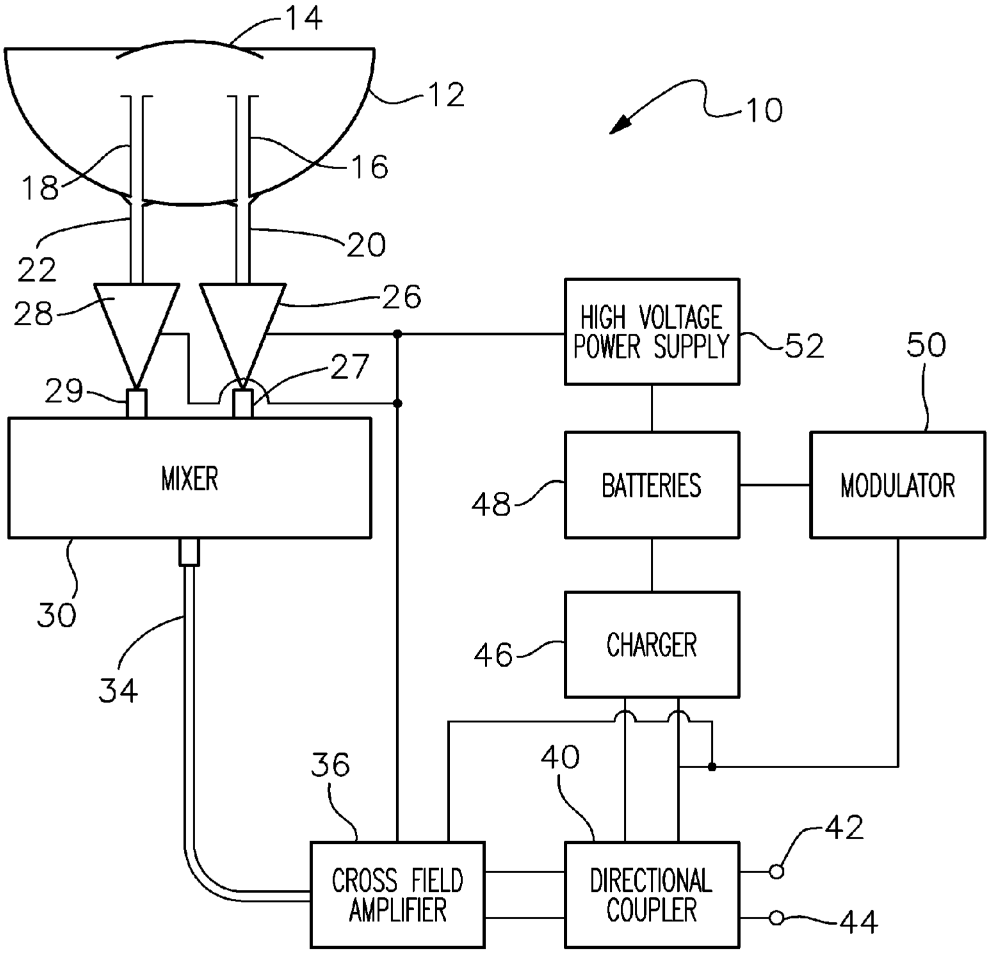
14
12
10
18
16
22
20
28
26
29
27
HIGH VOLTAGE POWER SUPPLY
52
50
MIXER
48
BATTERIES
MODULATOR
30
34
46
CHARGER
36
40
42
CROSS FIELD AMPLIFIER
DIRECTIONAL COUPLER
44

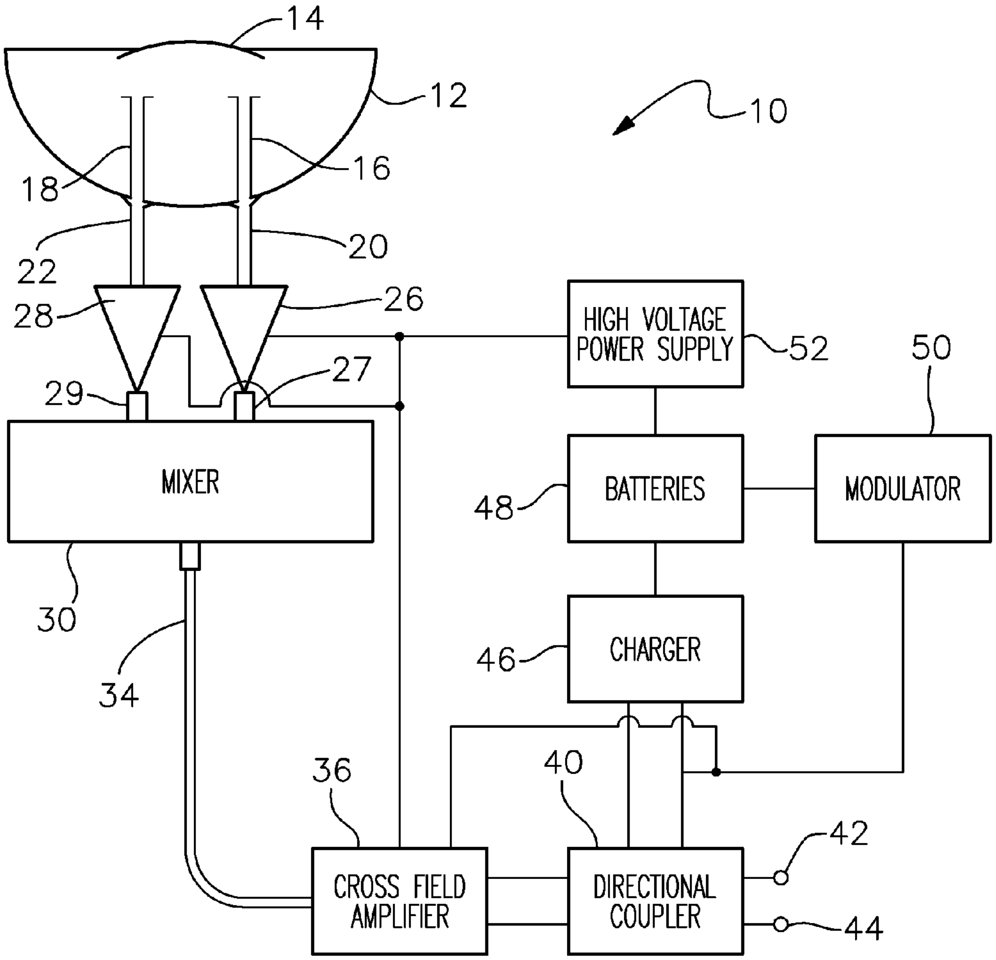
HIGH VOLTAGE POWER SUPPLY
MIXER
BATTERIES
MODULATOR
CHARGER
CROSS FIELD AMPLIFIER
DIRECTIONAL COUPLER

Fig. 1

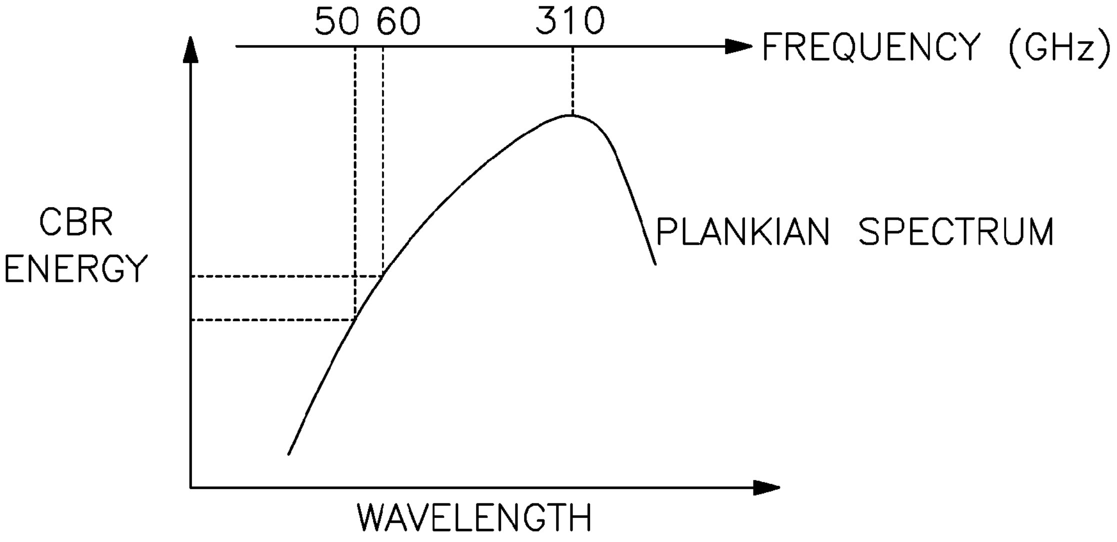
50
60
310
FREQUENCY (GHz)
CBR ENERGY
PLANKIAN SPECTRUM
WAVELENGTH

Fig. 2

APPARATUS AND METHOD FOR CAPTURING COSMIC BACKGROUND RADIATION AND CONVERTING THE SAME TO ELECTRICITY

RELATIONSHIP TO PRIOR APPLICATION

[0001] This is a U.S. non-provisional application relating to and claiming the benefit of U.S. Provisional Patent Application Ser. No. 61/224,255, filed Jul. 9, 2009, and U.S. Provisional Patent Application Ser. No. 61/235,756, filed Aug. 21, 2009.

BACKGROUND OF THE INVENTION

[0002] This invention relates to the use of isothermal radiation of the universe to produce alternative energy on Earth. More particularly, it relates to the use of the cosmic background heat radiation of the universe to produce electricity.

[0003] Modern twentieth century physicists and astronomers have abandoned the philosophical concept of the static universe for the experimentally verified theory of Big Bang, a fiery origin of universes. The Big Bang occurred eons ago at immense temperatures and it created its own space where none existed before, and it is expanding at the speed of light. Inside the expanding space there was and still is thermal equilibrium where the primordial particles collided with each other with originally high energies, thus preventing creation of more complex, heavier atomic elements. As the fireball of universe expanded, it cooled and the energies of colliding primeval particles decreased allowing formation of heavier atomic elements which are building blocks of galaxies and solar systems, including planet Earth.

[0004] The separation of energies of mass ($E=mc^2$), and energy of thermal cosmic radiation made the universe transparent to this radiation. Heat energy in the form of cosmic background radiation (CBR) streams in all directions within the expanding space. Today's cosmic space of Earth is flooded with radiation noise at 3° K temperature.

[0005] As recently as 1965, Arno Penzias and Robert Wilson discovered this thermal noise equivalent to 3° K while testing a large aperture antenna. For their discovery and their interpretation of radiation in microwave frequency they were awarded the Nobel Prize in 1967.

[0006] Further investigations followed discovery that the radiation was of identical temperature all over Earth (isothermal) and isotropic which means equally intense in all directions.

[0007] The isothermal and isotropic radiation constitutes conditions for conversion of CBR into the alternative energy on Earth.

SUMMARY OF THE INVENTION

[0008] In accordance with one form of this invention, there is provided an apparatus for capturing cosmic background radiation and for converting cosmic background radiation into electricity. The apparatus includes an antenna configured so as to capture cosmic background radiation. An electrostatic electron multiplier is connected to the antenna. A high voltage power supply is connected to the electrostatic electron multiplier whereby cosmic background radiation is converted into electricity.

[0009] In accordance with another form of this invention, there is provided a method for capturing cosmic background radiation and for converting cosmic background radiation into electricity. Cosmic background radiation is captured by an antenna. The signal from the antenna is amplified and converted into a lower frequency by an electrostatic electron multiplier and high voltage power supply whereby cosmic background radiation is converted into electricity.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] FIG. **1** is a schematic diagram showing one embodiment of the invention.

[0011] FIG. **2** is a graph showing the Plankian Spectrum.

DESCRIPTION OF THE PREFERRED EMBODIMENT

[0012] The invention analyzes and isolates the unique properties of cosmic background radiation (CBR) which allows it to couple to a machine, called a "generator," located on Earth.

[0013] The "generator" (a) uses inexhaustible energy of the universe to directly convert it into alternative electrical energy on Earth; (b) can be located and operated in any arbitrary location on Earth, even as a mobile unit, without modification; (c) requires no transportation of any supplied energy to operate; and (d) produces neither solid waste nor air pollution. The "generator" may operate as an individual unit supplying electrical energy to individual households or transportation vehicles.

[0014] There are three steps described herein. The first step validates the existence of CBR energy in the universe in the form of thermal noise. The second step uses Max Planck's Law of Radiation to define the existence of the microwave spectrum in the universe and quantitatively defines its intensity. The third step describes the "generator" which captures, isolates and amplifies the CBR energy of the universe into the useful electrical energy.

[0015] Generally speaking, the invention calls for the conversion of the microwave frequency electrical energy of CBR into 60 cycle, AC power on Earth.

[0016] CBR has been experimentally confirmed to be isothermal in any location on Earth and has been theoretically proven to be so in the entire universe. It means that the temperature of thermal radiation of 3° K persists anywhere on Earth where the "generator" of the alternative energy would be located.

[0017] CBR has also experimentally found to be isotropic across the entire Earth and theoretically proven to be so over the entire universe. It means that the entire frequency spectrum (in the microwave region) is equally intense not only across the universe but also in every direction within the universe wherever the "generator" would be located.

[0018] These unique properties of CBR are fundamental to conversion of CBR to the alternative energy.

[0019] It has been established experimentally and reported at the meeting of the American Astronomical Society in January 1990 that the frequency spectrum of CBR in the microwave region matches the radiation spectrum of the black body radiation formulated ninety years earlier by Max Planck. What it means is that the Big Bang's aftereffect in the form of CBR can be represented quantitatively by the equivalent black body radiation formulated by Max Planck.

[0020] Max Planck formulated the law of black body radiation which applies to almost all considerations derived from it. Some of these considerations allow one to calculate the intensity of Radiant Emittance in watts cm^{-2} $wavelength^{-1}$ (watts per $centimeter^2$ per wavelength).

[0021] The Radiant Emittance of CBR at 3° K is derived from the consideration that leads directly to the Radiant Emittance of black body. This is done as follows:

[0022] (a) Using Wien's Displacement Law (one of the considerations of Planck's Law), which allows one to derive the wavelength λmax, at which a maximum intensity of Radiant Emittance, at 3° K occurs

$$\lambda\text{max } T° \text{ K}=2897.9° \text{ K}$$

where wavelength is micrometers (one micrometer=10^{-6} meter).

Note that this equation represents a straight line.

[0023] (b) Once the λmax for 3° K is known the actual intensity of Radiant Emittance of CBR "in situ" of "generator," can be extrapolated from the plot of Planck's equation Radiant Emittance versus wavelength (see *Handbook of Military Infrared Technology* compiled by William Wolfe, University of Michigan 1965 edition, p. 20, fig. 2.8).

[0024] By sliding the Planck's curve along the straight line (representing Wien's Displacement Law) to the wavelength of λmax for 3° K, the extrapolated value of the Radiant Emittance can be found as:

$$1.2\times10^{-11}\text{ Wcm}^{-2}\text{ wavelength}^{-1}$$

[0025] The λmax for 3° K (present temperature of CBR) occurs at the frequency of 310 Gigahertz which requires very specialized hardware.

[0026] Instead by choosing a lower frequency of 60 GHz, the hardware of "generator" is simplified since microwave components at that frequency are commercially available. The change from 310 GHz to 60 GHz reduces the extrapolated power (Radiant Emittance) by 97% to 3.6×10^{-13} Wcm^{-2} wavelength^{-1} (see FIG. **2.6**, *Handbook of Military Infrared Technology* compiled by William Wolfe, University of Michigan 1965 edition).

[0027] The level of CBR Radiant Emittance anywhere on the Earth as discussed above is 3.6×10^{-13} Wcm^{-2} wavelength^{-1} at 60 GHz frequency. This level is low compared to the level of man-made communication signals. Furthermore, the components used in the assembly of "generator" will generate their own noise at the terminals of the components either due to temperature of environment "in situ" of "generator" or due to operating currents (motion of electrons). The object to preserve purity of 60 GHz CBR is to keep these interferences well below the level of CBR.

[0028] The design of "generator" uses the unique properties of CBR to:

[0029] (a) isolate CBR energy from man-made interferences;

[0030] (b) keep the Johnson noise of antenna well below the energy of CBR at 60 GHz; and

[0031] (c) amplify the CBR energy without use of thermionic amplifiers, which generate noise.

[0032] The isotropic properties of CBR provides an equal intensity of radiation independent of the orientation of the antenna main lobe.

[0033] By orienting resonant antenna towards the sky the temperature of antenna becomes equal to the temperature of the sky, which can be assumed at 3° K, and not the temperature of the antenna structure (see "Electromagnetism" by D. Krane and Kevin R. Carber, p. 700). By doing so, the noise power per unit bandwidth becomes:

$$W/\text{wavelength}=KT \text{ where } K \text{ is Boltzman's constant of } 1.38\times10^{-23} \text{ and } T \text{ is temperature of the sky}$$

[0034] Hence the noise generated by the sky N_s oriented antenna will be (N_0)

$$W/\text{wavelength}=3\times1.38\times10^{-23}\,W\text{cm}^{-2}\text{ wavelength}^{-1}$$

[0035] Hence the CBR signal to noise ratio:

$$\frac{S(60\text{ GHz})}{N_0}=\frac{3.6\times10^{-13}}{3\times1.38\times10^{-23}}=8.7\times10^{9}$$

or nine orders higher than noise.

[0036] Further on, by orienting the antenna towards the sky, the interference by man-made signals will be eliminated.

[0037] Preferably the antenna:

[0038] (1) is parabolic with a virtual aperture 60 centimeters in diameter;

[0039] (2) operates at a frequency of 60 gigahertz and connected to a wave guide terminated in multiple deformed dipoles (which eliminate the null-poles typical to regular straight dipoles);

[0040] (3) has a power gain (referenced to dipoles) of 86,400 (see "Electronics in Communications" by Sol Lapatrine, pp. 256-259);

[0041] (4) has a temperature of 3 degrees Kelvin; and

[0042] (5) has low noise of 4.14×10^{-23} Watts per wavelength.

[0043] Typically in man-made communication systems, the amplifiers used are thermionic: tubes, travelling wave tubes, klystrons. All of them are powerful amplifiers of extremely low signals. They also are plagued by high noise (Johnson noise, shot noise, fluctuation of voltage across terminals of resistors). Further, they need assortments of supplies to operate the thermionic tubes, all sources of noise.

[0044] To retain trueness of 60 GHz CBR signals, the "generator" needs a powerful amplifier without use of a noisy thermionic amplifier. This can be provided by secondary electron emission. The device needs only high voltage supply of preferably 1250V dc to produce high power gain. It is called an Electrostatic Electron Multiplier (EEM), such as one made by CBS Laboratories for the guidance system of the Lunar Orbiter which provides gain of 10^{6}. It can be temporarily opened to air to couple it to antenna input. The device is well adapted to a microwave system.

[0045] The EEM operating at microwave frequency receives a weak CW (continuous wave) signal from antenna. It generates secondary emission electrons of increasing current density, which multiplies the weak signal many thousands times. The EEM becomes the current density variation power amplifier.

[0046] Power amplification of 230,000 has been achieved by a commercially available EEM.

[0047] The effect of EEM at the output of antenna reduces thermal noise by a factor of its gain squared (typically 10^{10}-10^{12}) down along the "generator" system. Since "generator" uses the unmodulated signal (CW signal), it acts as power amplifier of only 10^{6}. This ruggedized assembly exposed to typical ambient temperature of 300° K will produce Johnson's noise power of:

$$W=1.38\times10^{-23}\times300=414\times10^{-21}=2.48\times10^{-19}\,W/\text{wavelength}$$

[0048] Hence the presence EEM of gain 10^{6} will decrease further noise by entire system to almost 10^{-27}, well below the CBR power.

[0049] Referring now to FIG. **1** of the preferred embodiment which shows generator **10** which captures and converts the CBR microwave energy to the 60 cps AC power. Generator **10** includes parabolic antenna **12** and is spherical reflector **14**. The dipoles **16** and **18** are received within parabolic antenna. Flanges **20** and **22** connect parabolic antenna to the waveguides **16** and **18**. The dipoles receive the 60 GHz and 50 GHz frequencies to feed them into mixer **30**. The mixer may be MXP-19R, commercially available from Minitech. The EEM **26** and **28** are connected to mixer **30** by flanges **27** and **29**. The intermediate frequency output, at 10 GHz, of mixer **30** is connected to the flexible cable **34**, which may be an RG-214.

[0050] The above assembly of the base of antenna **12**, EEM **26** and **28** and the mixer's flanges are brittle and should be ruggedized in epoxy rubber to withstand many g's of generator **10**.

[0051] The flexible cable **34** is connected to cross-field amplifier (CFA) **36**. The output of CFA **36** is directly connected to the directional coupler **40** in support to the service system to provide AC power to the charger **46**, batteries **48**, high DC voltage package **52** and to the sixty cycles per second modulator **50** to amplitude modulate CFA **36**. The HV DC power supply is commercially available from Nelles Griot Holography.

[0052] The generator **10** may be placed at any location on Earth since the Earth is flooded with CBR. The distribution of CBR energy is in a spectrum of frequencies called the Planckian spectrum. It is the energy used by generator **10** to be converted in 60 cycles AC power.

[0053] Two frequencies, namely 60 GHz and 50 GHz, are used to operate generator **10**. Each of the CBR frequencies is a continuous wave (CW) and the entire process of conversion to AC power is in the form of CW.

[0054] The CBR frequencies are in GHz (10^9 cycles per second). Their maximum occurs at 1.2×10^{-11} Watt per cm^2 per wavelength at 310 GHz. The generator **10** uses the compromised frequency of 60 GHz, which further lowers the CBR energy to 3.6×10^{-13} Watts per cm^2 per wavelength. The generator and methods described earlier enables one to isolate CBR which is now low in its intensity as compared with interferences and also noiselessly amplifies it well above their level. The lowest ratio of CBR to interferences is 10^9.

[0055] As previously discussed, FIG. **1** shows parabolic antenna **12** with two dipoles **16** and **18** located at the center of its spherical reflector. One dipole receives the CBR frequency of 60 GHz; the second dipole receives a frequency of 50 GHz. These two frequencies are used further down the generator's system to downconvert the IF frequency output to 10 GHz, with 60 GHz wave providing input to "FR" and 50 GHz wave input to LO terminals of the mixer **30**. Both outputs from the antenna dipoles are amplified by an antenna gain of 86,400 and by an EEM gain of 230,000. Since the intensities of the 60 GHz and 50 GHz outputs from the antenna are 3% of the maximum of CBR, the input energy into the mixer is 7.1×10^{-3} Watt per wavelength. The insertion loss of the mixer is 7 db which brings the IF energy to 3.56×10^{-3} Watt per wavelength.

[0056] It has been proven that the microwave energy of CBR, in form of a single sine frequency of 60 GHz can be isolated, noiselessly amplified and downconverted to manageable frequency of 10 GHz. It occurs at the energy level of 3.56 Watt per wavelength. The conversion to power in Watt per second occurs by generator at 10^{10} wavelength per second. The continuous wave (CW) power is:

$$P_{10GHZ}=3.56\times10^{-3}W/\lambda\times10^{10}\lambda/\text{sec}=3.56\times10^{7}\text{ Watt per second}$$

when AMPLITUDE MODULATED by CFA at modulation index m−1 at 60 cps the AC power is

$$P_{AC}=\frac{356}{60}\left(10^5+\frac{m}{4}10^5\right)_{m-1}=6.25\times10^5$$

Watt per second=625 Kilowatt/sec which represents power of 10 GHz frequency per second plus single sideband per second.

[0057] Below is a list of the components shown in FIG. **1**:

[0058] **10**—Generator

[0059] **12**—Antenna Assembly

[0060] **14**—Spherical reflector

[0061] **16**, **18**—Dipoles mounted on waveguides

[0062] **20**, **22**—Flanges to waveguides (678-007, Millitech)

[0063] **26**, **28**—EEM, Dynode Type, 230,000 Gain each

[0064] **27**, **29**—Flanges (678-007, Millitech)

[0065] **30**—Mixer (MXP-19R, Millitech)

[0066] **34**—10 GHz flexible Cable RG-214

[0067] **36**—Cross-field amplifier, cold mode operation

[0068] **40**—Directional coupler (10 GHz)

[0069] **46**—Charger to Batteries

[0070] **48**—Batteries, Melles Griot Labs

[0071] **50**—60 cycles modulator to cross-field amplifier

[0072] **52**—High voltage DC package (Nelles Griot Labs)

[0073] While the invention has been described in terms of the above embodiment, those skilled in the art will recognize that the invention can be practiced with modifications and other embodiments within the spirit and scope of the appended claims.

What is claimed is:

1. An apparatus for capturing cosmic background radiation and for converting cosmic background radiation into electricity comprising:

an antenna; the antenna configured so as to capture cosmic background radiation;

an electrostatic electron multiplier; the antenna connected to the electrostatic electron multiplier; and

a high voltage power supply connected to the electrostatic electron multiplier whereby cosmic background radiation is converted to electricity.

2. An apparatus as set forth in claim **1** wherein the antenna is parabolic.

3. An apparatus as set forth in claim **1**, further including an amplitude modulating mixer connected to the electrostatic electron multiplier.

4. An apparatus as set forth in claim **1**, further including at least one battery connected to the high voltage power supply.

5. An apparatus as set forth in claim **4**, further including a battery charger connected to the battery.

6. An apparatus as set forth in claim **1**, further including a cross field amplifier connected to the electrostatic electron multiplier; the output of the cross field amplifier providing AC power.

7. An apparatus as set forth in claim **6**, further including a directional coupler connected to the cross-field amplifier.

8. An apparatus as set forth in claim **1** wherein the high voltage power supply provides DC voltage.

9. An apparatus as set forth in claim **7** wherein the output of the cross field amplifier is connected to the battery charger; a portion of the AC power providing current for the battery charger.

10. An apparatus as set forth in claim **1** wherein the cosmic background radiation is at 3 degrees Kelvin.

11. An apparatus as set forth in claim **1**, further including a wave guide connecting the antenna to the electrostatic electron multiplier.

12. An apparatus as set forth in claim **1**, further including a pair of dipoles connected to the antenna.

13. An apparatus as set forth in claim **12** wherein one of the dipoles operates at approximately 50 GHz and the other dipole operates at approximately 60 GHz.

14. An apparatus as set forth in claim **1** wherein the electricity is AC power at approximately 60 cycles per second.

15. A method for capturing cosmic background radiation and for converting cosmic background radiation into electricity comprising:

capturing cosmic background radiation with an antenna;

amplifying the cosmic background radiation by an electrostatic electron multiplier; and

converting the amplified cosmic background radiation to electricity.

16. A method as set forth in claim **15**, further including providing high voltage to the electrostatic electron multiplier.

17. A method as set forth in claim **16**, further including using the converted electricity for providing the high voltage to the electrostatic electron multiplier.

18. A method as set forth in claim **15**, further including orienting the antenna toward the sky.

* * * * *

US005590031A

United States Patent [19]

Mead, Jr. et al.

[11] **Patent Number: 5,590,031**

[45] **Date of Patent: Dec. 31, 1996**

[54] SYSTEM FOR CONVERTING ELECTROMAGNETIC RADIATION ENERGY TO ELECTRICAL ENERGY

[76] Inventors: **Franklin B. Mead, Jr.**, 44536 Avenida Del Sol, Lancaster, Calif. 93535; **Jack Nachamkin**, 12314 Teri Dr., Poway, Calif. 92064

[21] Appl. No.: **281,271**

[22] Filed: **Jul. 27, 1994**

[51] **Int. Cl.**^{6} **H02M 1/00**

[52] **U.S. Cl.** **363/8**; 363/178; 342/6

[58] **Field of Search** 363/8, 13, 178; 342/6, 61, 73, 173, 175

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,882,503	5/1975	Gamara	343/100 R
4,725,847	2/1988	Poirier	343/840
5,008,677	4/1991	Trigon et al.	342/17

Primary Examiner—Peter S. Wong
Assistant Examiner—Adolf Berhane
Attorney, Agent, or Firm—Chris Papageorge

[57] **ABSTRACT**

A system is disclosed for converting high frequency zero point electromagnetic radiation energy to electrical energy. The system includes a pair of dielectric structures which are positioned proximal to each other and which receive incident zero point electromagnetic radiation. The volumetric sizes of the structures are selected so that they resonate at a frequency of the incident radiation. The volumetric sizes of the structures are also slightly different so that the secondary radiation emitted therefrom at resonance interfere with each other producing a beat frequency radiation which is at a much lower frequency than that of the incident radiation and which is amenable to conversion to electrical energy. An antenna receives the beat frequency radiation. The beat frequency radiation from the antenna is transmitted to a converter via a conductor or waveguide and converted to electrical energy having a desired voltage and waveform.

14 Claims, 8 Drawing Sheets

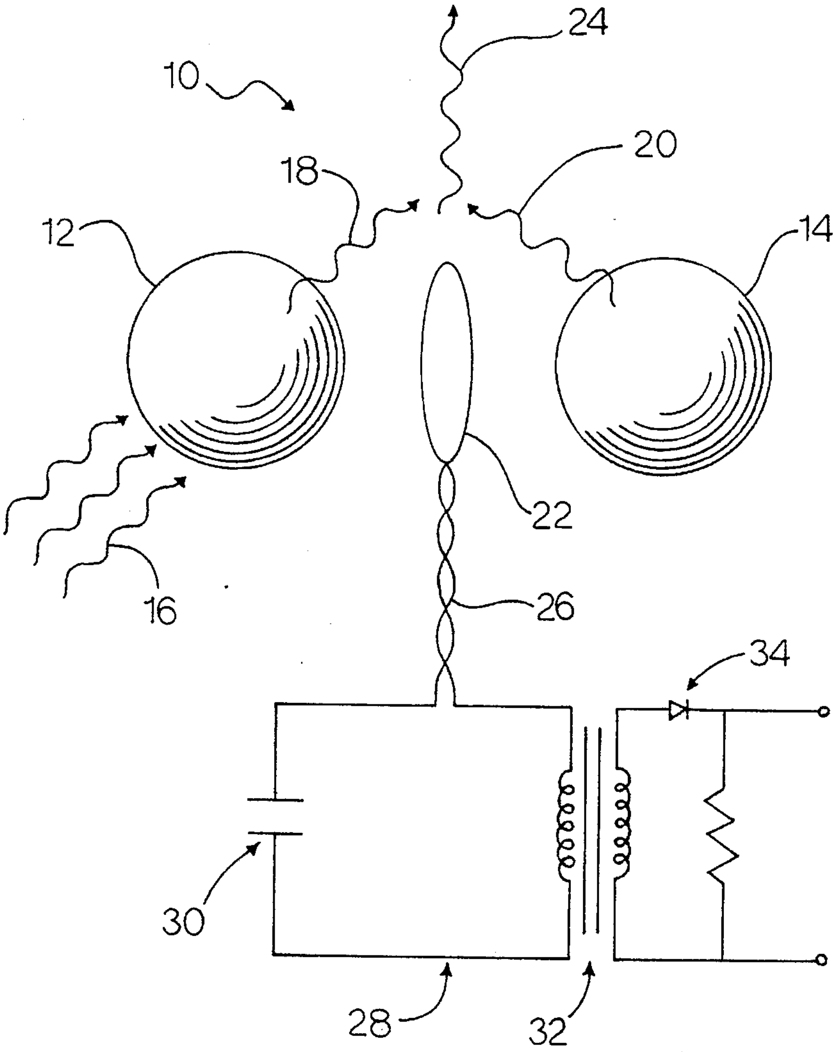

FIG. 1

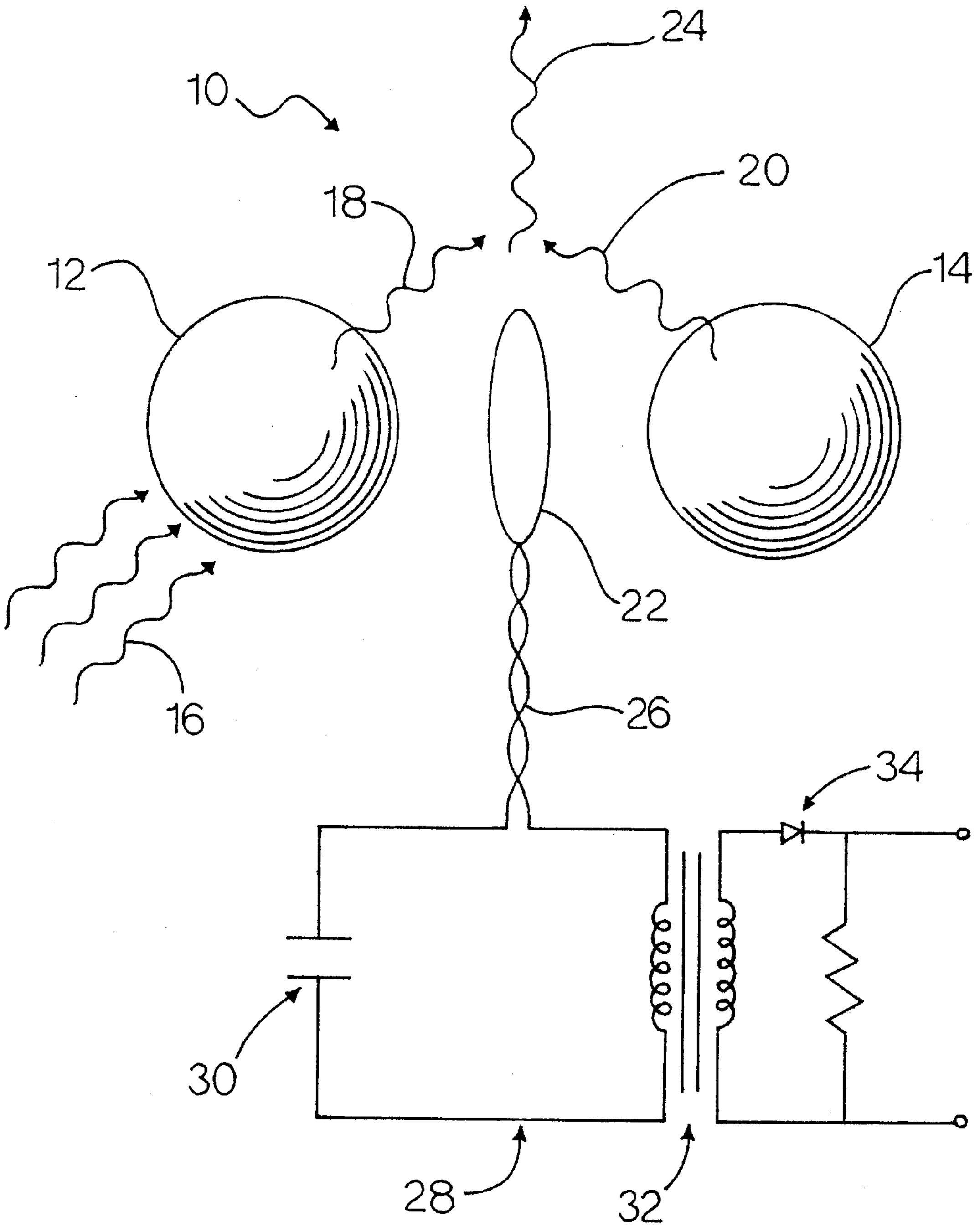
24
10
20
18
12
14
22
26
16
34
30
28
32

FIG. 2

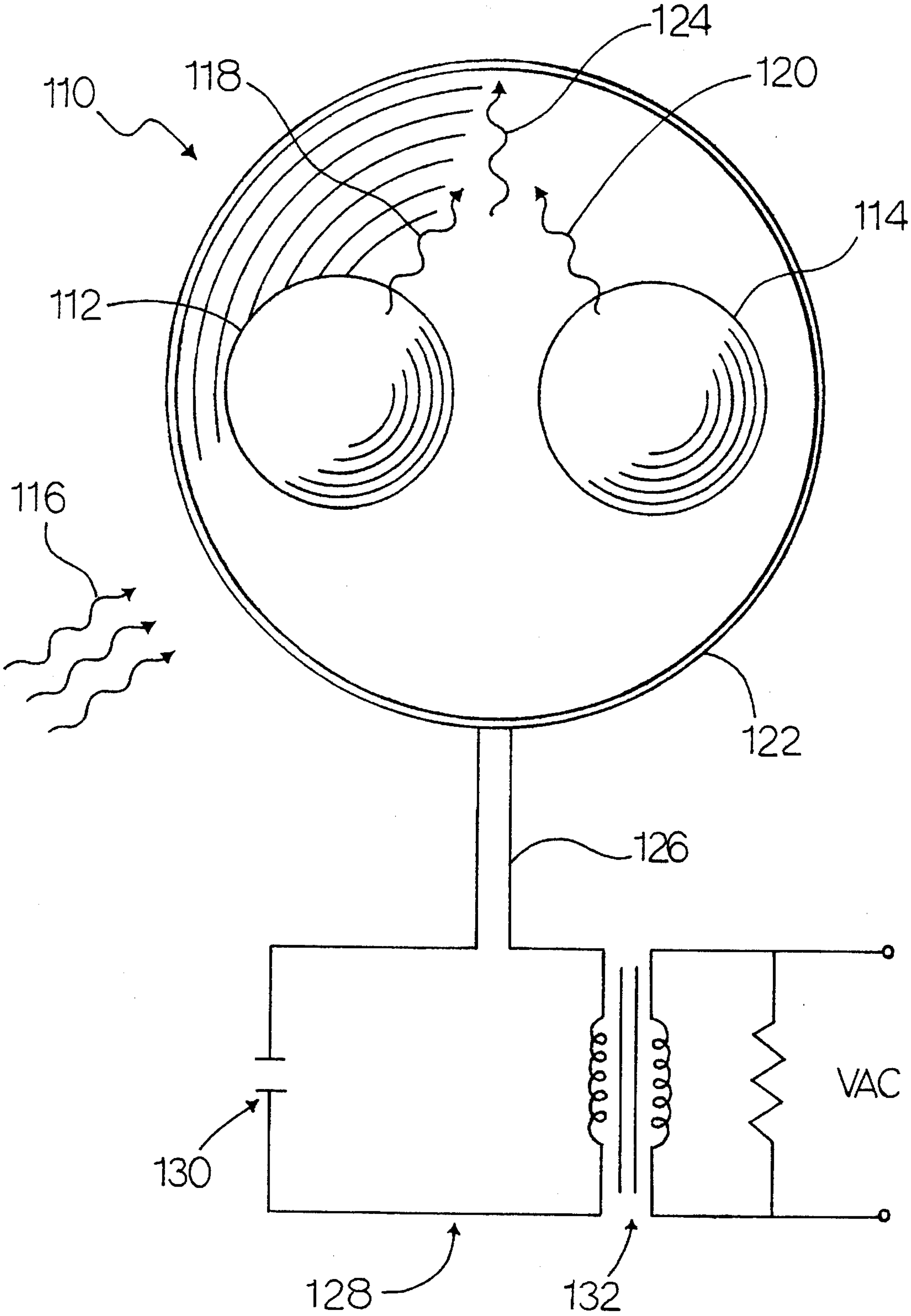
124
118
110
120
114
112
116
122
126
VAC
130
128
132

FIG. 3

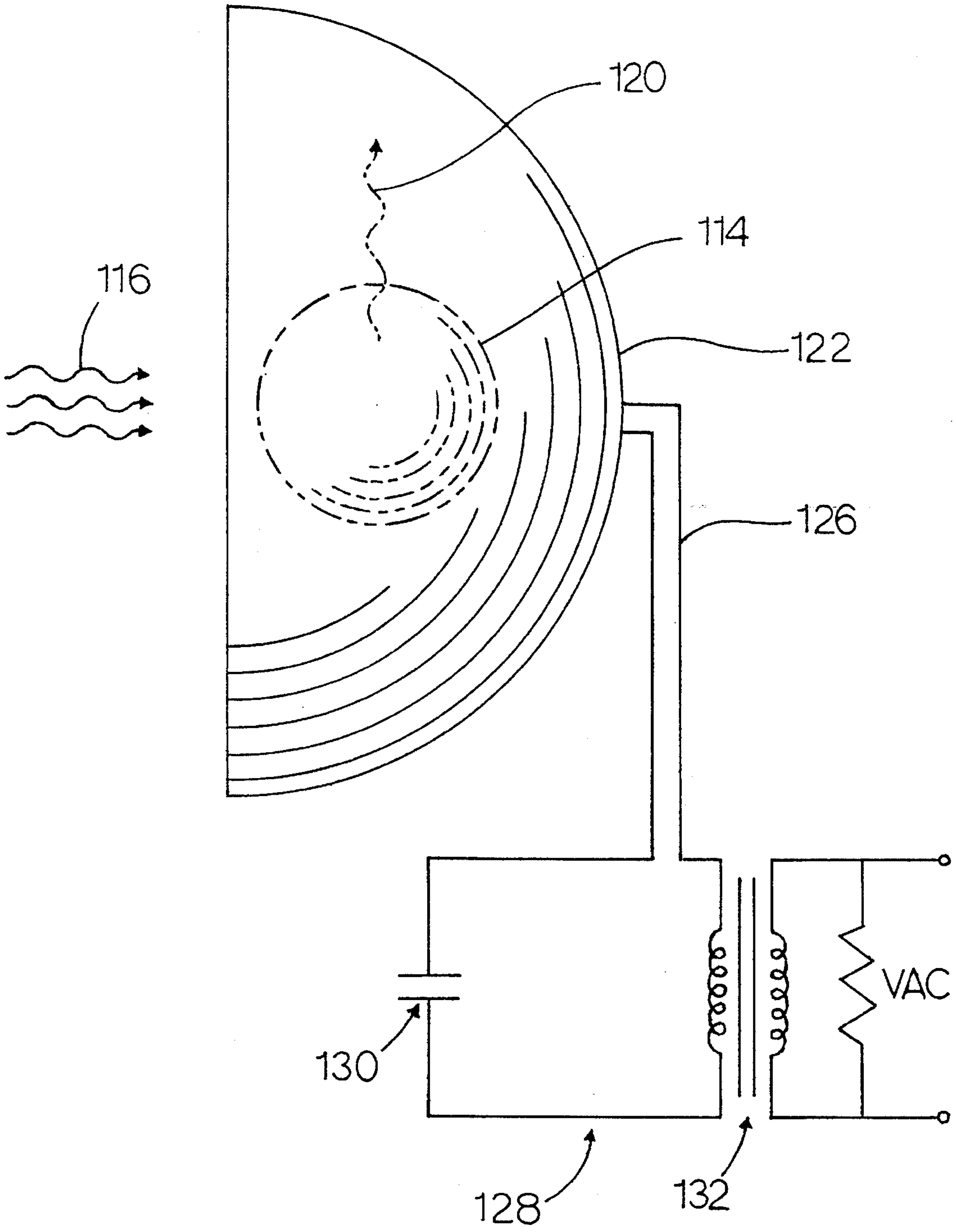
120
114
116
122
126
VAC
130
128
132

FIG. 4

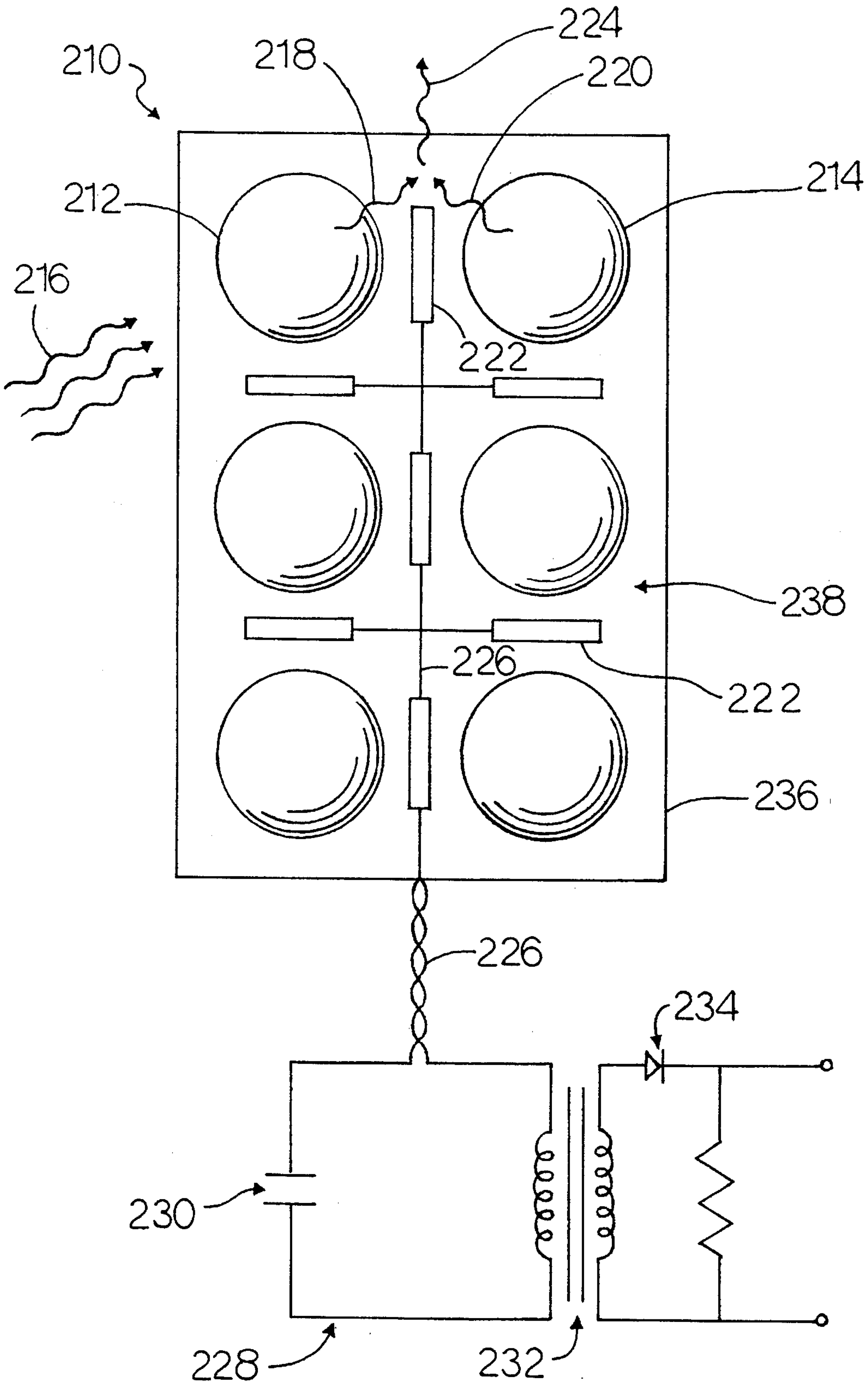

FIG. 5

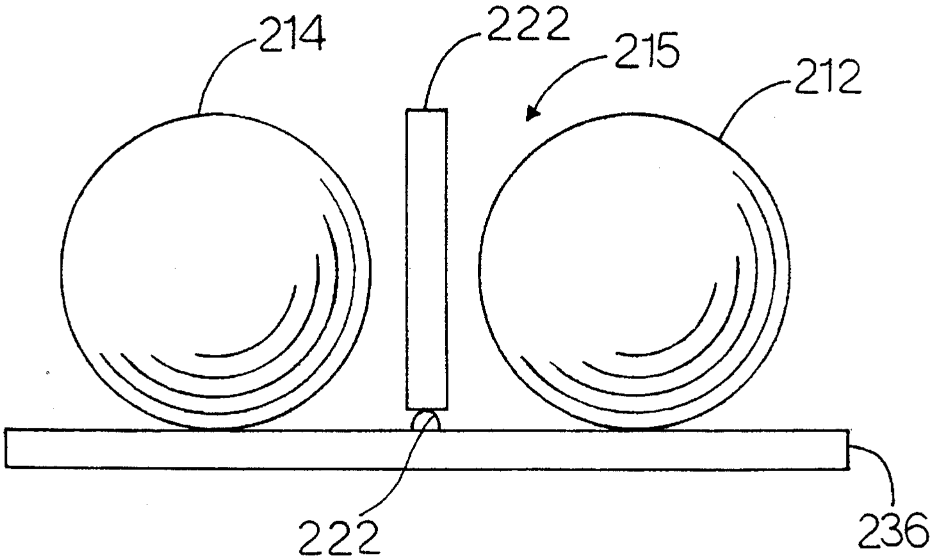
214
222
215
212
222
236

FIG. 6

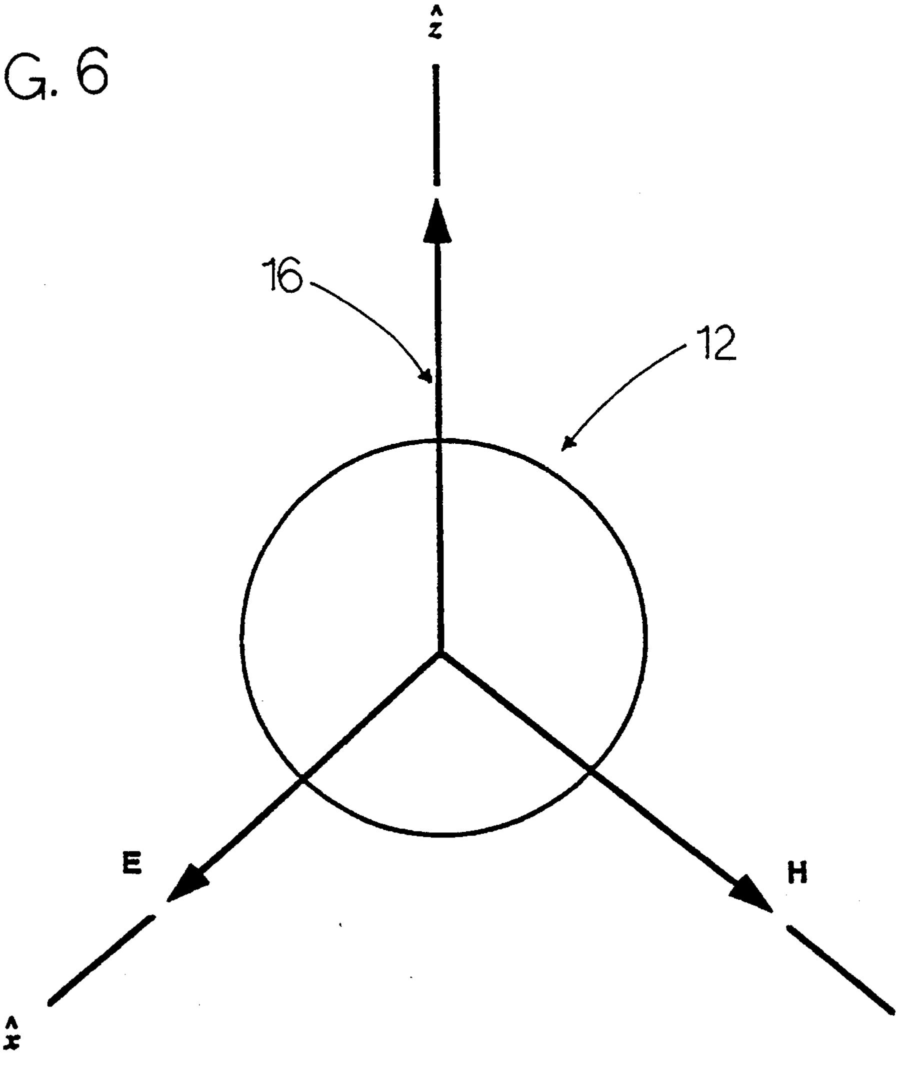
$\hat{z}$
16
12
E
H
$\hat{x}$
$\hat{y}$

FIG. 7

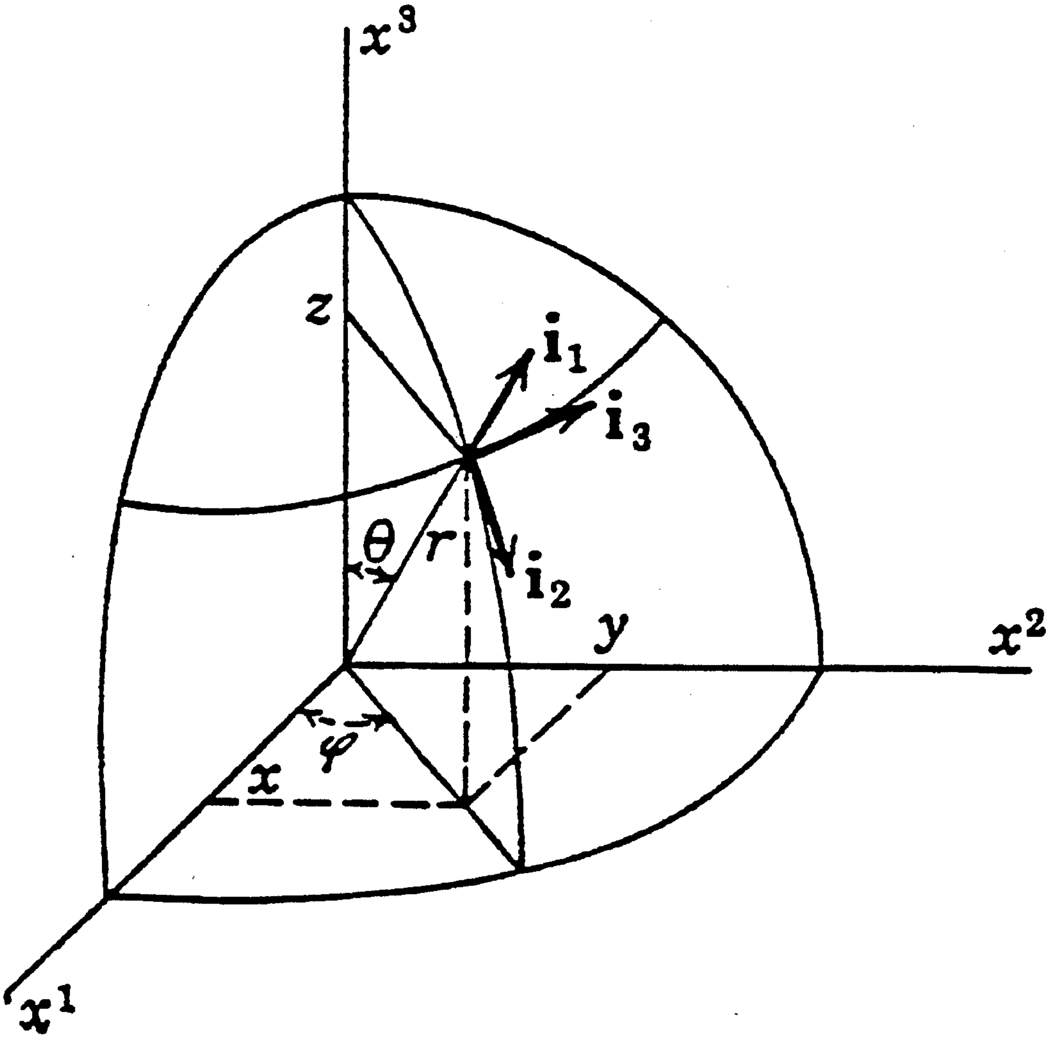
x^3
z
i_1
i_3
θ
r
i_2
y
x^2
φ
x
x^1

FIG. 8

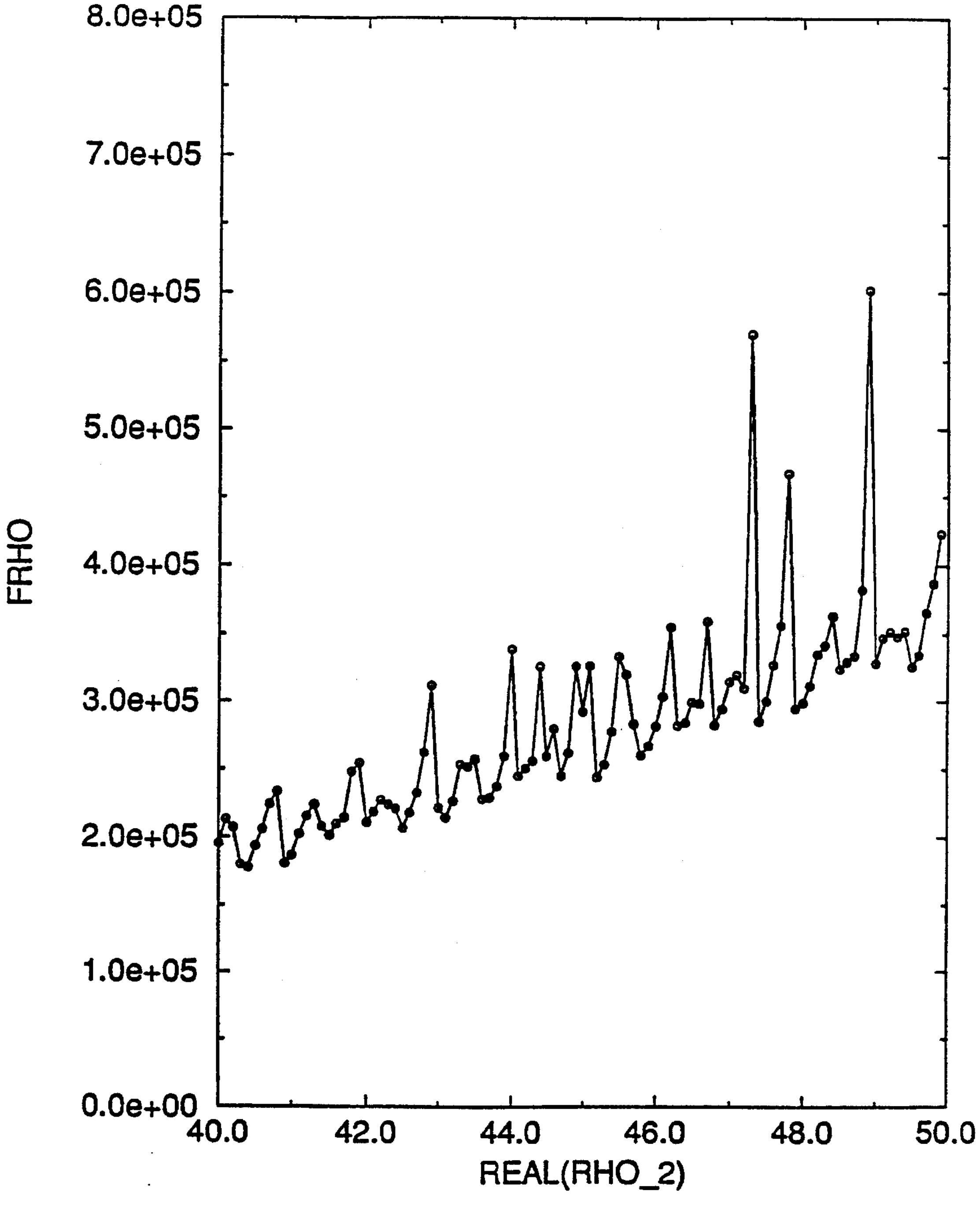
8.0e+05
7.0e+05
6.0e+05
5.0e+05
4.0e+05
3.0e+05
2.0e+05
1.0e+05
0.0e+00
FRHO
40.0
42.0
44.0
46.0
48.0
50.0
REAL(RHO_2)

FIG. 9

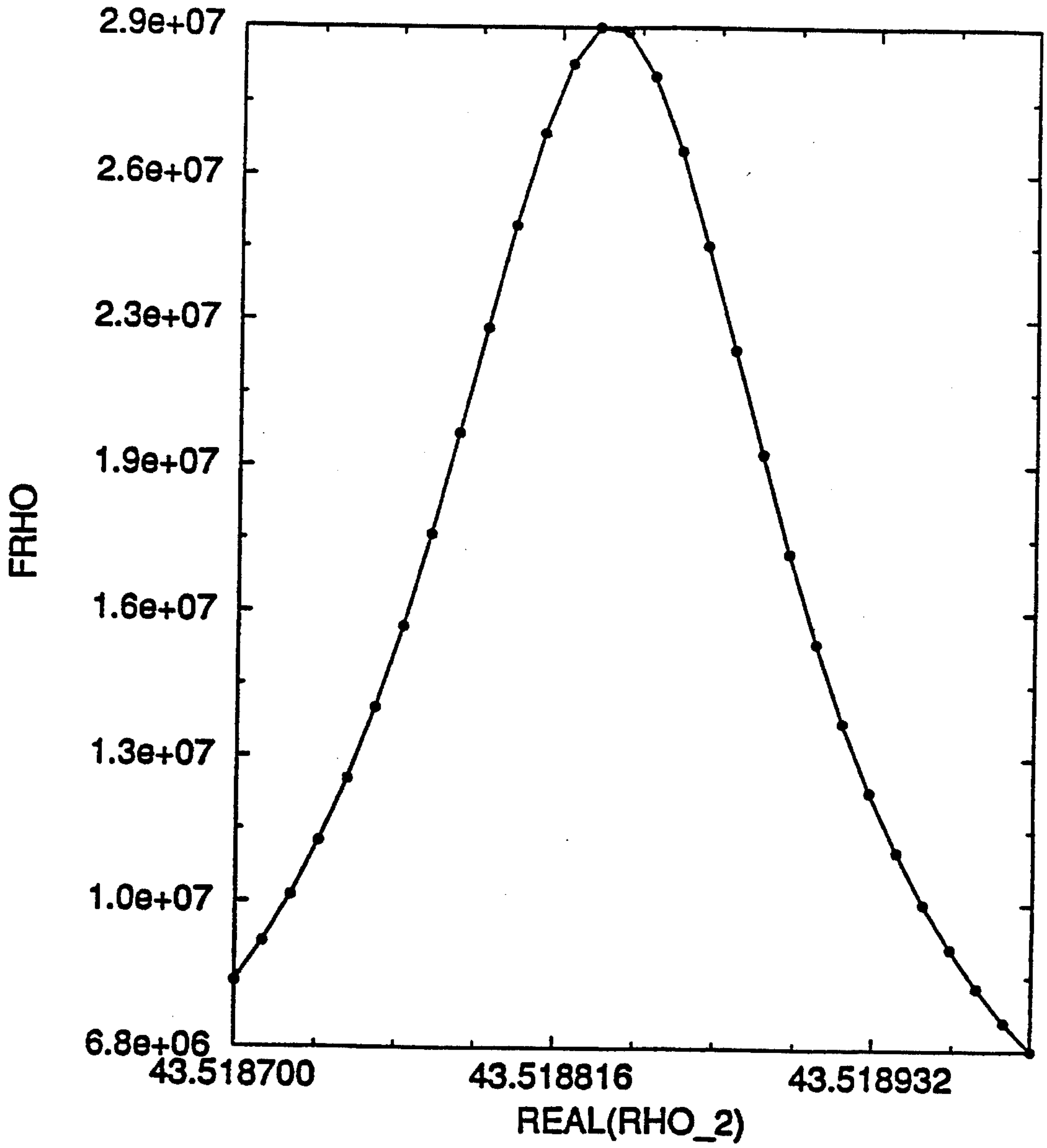
2.9e+07
2.6e+07
2.3e+07
1.9e+07
1.6e+07
1.3e+07
1.0e+07
6.8e+06
FRHO
43.518700
43.518816
43.518932
REAL(RHO_2)

SYSTEM FOR CONVERTING ELECTROMAGNETIC RADIATION ENERGY TO ELECTRICAL ENERGY

BACKGROUND OF THE INVENTION

The invention relates generally to conversion of electromagnetic radiation energy to electrical energy, and, more particularly, to conversion of high frequency bandwidths of the spectrum of a type of radiation known as zero point electromagnetic radiation to electrical energy.

The existence of zero point electromagnetic radiation was discovered in 1958 by the Dutch physicist M. J. Sparnaay. Mr. Sparnaay continued the experiments carried out by Hendrik B. G. Casimir in 1948 which showed the existence of a force between two uncharged parallel plates which arose from electromagnetic radiation surrounding the plates in a vacuum. Mr. Sparnaay discovered that the forces acting on the plates arose from not only thermal radiation but also from another type of radiation now known as classical electromagnetic zero point radiation. Mr. Sparnaay determined that not only did the zero point electromagnetic radiation exist in a vacuum but also that it persisted even at a temperature of absolute zero. Because it exists in a vacuum, zero point radiation is homogeneous and isotropic as well as ubiquitous. In addition, since zero point radiation is also invariant with respect to Lorentz transformation, the zero point radiation spectrum has the characteristic that the intensity of the radiation at any frequency is proportional to the cube of that frequency. Consequently, the intensity of the radiation increases without limit as the frequency increases resulting in an infinite energy density for the radiation spectrum. With the introduction of the zero point radiation into the classical electron theory, a vacuum at a temperature of absolute zero is no longer considered empty of all electromagnetic fields. Instead, the vacuum is now considered as filled with randomly fluctuating fields having the zero point radiation spectrum. The special characteristics of the zero point radiation which are that it has a virtually infinite energy density and that it is ubiquitous (even present in outer space) make it very desirable as an energy source. However, because high energy densities exist at very high radiation frequencies and because conventional methods are only able to convert or extract energy effectively or efficiently only at lower frequencies at which zero point radiation has relatively low energy densities, effectively tapping this energy source has been believed to be unavailable using conventional techniques for converting electromagnetic energy to electrical or other forms of easily useable energy. Consequently, zero point electromagnetic radiation energy which may potentially be used to power interplanetary craft as well as provide for society's other needs has remained unharnessed.

There are many types of prior art systems which use a plurality of antennas to receive electromagnetic radiation and provide an electrical output therefrom. An example of such a prior art system is disclosed in U.S. Pat. No. 3,882,503 to Gamara. The Gamara system has two antenna structures which work in tandem and which oscillate by means of a motor operatively attached thereto in order to modulate the radiation reflected from the antenna surfaces. The reflecting surfaces of the antennas are also separated by a distance equal to a quarter wavelength of the incident radiation. However, the Gamara system does not convert the incident radiation to electrical current for the purpose of converting the incident electromagnetic radiation to another form of readily useable energy. In addition, the relatively large size of the Gamara system components make it unable to resonate at and modulate very high frequency radiation.

What is therefore needed is a system which is capable of converting high frequency electromagnetic radiation energy into another form of energy which can be more readily used to provide power for transportation, heating, cooling as well as various other needs of society. What is also needed is such a system which may be used to provide energy from any location on earth or in space.

SUMMARY OF THE INVENTION

It is a principle object of the present invention to provide a system for converting electromagnetic radiation energy to electrical energy.

It is another object of the present invention to provide a system for converting electromagnetic radiation energy having a high frequency to electrical energy.

It is another object of the present invention to provide a system for converting zero point electromagnetic radiation energy to electrical energy.

It is another object of the present invention to provide a system for converting electromagnetic radiation energy to electrical energy which may used to provide such energy from any desired location on earth or in space.

It is another object of the present invention to provide a system for converting electromagnetic radiation energy to electrical energy having a desired waveform and voltage.

It is an object of the present invention to provide a miniaturized system for converting electromagnetic radiation energy to electrical energy in order to enhance effective utilization of high energy densities of the electromagnetic radiation.

It is an object of the present invention to provide a system for converting electromagnetic radiation energy to electrical energy which is simple in construction for cost effectiveness and reliability of operation.

Essentially, the system of the present invention utilizes a pair of structures for receiving incident electromagnetic radiation which may be propagating through a vacuum or any other medium in which the receiving structures may be suitably located. The system of the present invention is specifically designed to convert the energy of zero point electromagnetic radiation; however, it may also be used to convert the energy of other types of electromagnetic radiation. The receiving structures are preferably composed of dielectric material in order to diffract and scatter the incident electromagnetic radiation. In addition, the receiving structures are of a volumetric size selected to enable the structures to resonate at a high frequency of the incident electromagnetic radiation based on the parameters of frequency of the incident radiation and propagation characteristics of the medium and of the receiving structures. Since zero point radiation has the characteristic that its energy density increases as its frequency increases, greater amounts of electromagnetic energy are available at higher frequencies. Consequently, the size of the structures are preferably miniaturized in order to produce greater amounts of energy from a system located within a space or area of a given size. In this regard, the smaller the size of the receiving structures, the greater the amount of energy that can be produced by the system of the present invention.

At resonance, electromagnetically induced material deformations of the receiving structures produce secondary fields of electromagnetic energy therefrom which may have eva-

nescent energy densities several times that of the incident radiation. The structures are of different sizes so that the secondary fields arising therefrom are of different frequencies. The difference in volumetric size is very small so that interference between the two emitted radiation fields, and the receiving structures at the two different frequencies produces a beat frequency radiation which has a much lower frequency than the incident radiation. The beat frequency radiation preferably is at a frequency which is sufficiently low that it may be relatively easily converted to useable electrical energy. In contrast, the incident zero point radiation has its desirable high energy densities at frequencies which are so high that conventional systems for converting the radiation to electrical energy either cannot effectively or efficiently so convert the radiation energy or simply cannot be used to convert the radiation energy for other reasons.

The system of the present invention also includes an antenna which receives the beat frequency radiation. The antenna may be a conventional metallic antenna such as a loop or dipole type of antenna or a rf cavity structure which partially encloses the receiving structures. The antenna feeds the radiation energy to an electrical conductor (in the case of a conventional dipole or comparable type of antenna) or to a waveguide (in the case of a rf cavity structure). The conductor or waveguide feeds the electrical current (in the case of the electrical conductor) or the electromagnetic radiation (in the case of the waveguide) to a converter which converts the received energy to useful electrical energy. The converter preferably includes a tuning circuit or comparable device so that it can effectively receive the beat frequency radiation. The converter may include a transformer to convert the energy to electrical current having a desired voltage. In addition, the converter may also include a rectifier to convert the energy to electrical current having a desired waveform.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a plan view of the receiving structures and antenna of a first embodiment of the system of the present invention with a schematic view of the conductor and converter thereof and also showing the incident primary and emitted secondary electromagnetic radiation.

FIG. **2** is a front view of the receiving structures, antenna and waveguide of a second embodiment of the system of the present invention with a schematic view of the converter thereof and also showing the incident primary and emitted secondary electromagnetic radiation.

FIG. **3** is a perspective view of the receiving structures, antenna and waveguide of the second embodiment shown in FIG. **2** with a schematic view of the converter thereof and also showing the incident primary and emitted secondary electromagnetic radiation.

FIG. **4** is a front view of the substrate and a plurality of pairs of the receiving structures and a plurality of antennas of a third embodiment of the system of the present invention with a schematic view of the conductor and converter thereof and also showing the incident primary and emitted secondary electromagnetic radiation.

FIG. **5** is a top view of some of the components of the third embodiment of the system of the present invention showing two of the plurality of pairs of receiving structures and two of the plurality of antennas mounted on the substrate.

FIG. **6** is a diagram of a receiving structure of the system of the present invention showing an incident electromagnetic plane wave impinging on the receiving structure and illustrating the directions of the electric and magnetic field vectors thereof.

FIG. **7** is a diagram of a spherical coordinate system as used in the formulas utilized in the system of the present invention.

FIG. **8** is a graph showing an imaginary ρ parameter plotted against a real ρ parameter illustrating the values thereof at resonance as well as values thereof at other than resonance.

FIG. **9** is a graph showing a portion of the graphical representation shown in FIG. **8** illustrating the real and imaginary ρ values at or near a single resonance.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring to the drawings, a first embodiment of the present invention is generally designated by the numeral **10**. The system **10** includes a first and second means for receiving **12** and **14** incident electromagnetic radiation **16**. The means for receiving **12** and **14** are preferably a pair of spherical structures **12** and **14** which are preferably composed of a dielectric material. Alternatively, the spheres **12** and **14** may be cubical structures or any other suitable shape. The spheres **12** and **14** may be mounted on a suitable foundation by any suitable mounting means (not shown), or spheres **12** and **14** may be suspended from a suitable foundation by any suitable suspension means (not shown). The spheres **12** and **14** are preferably composed of a dielectric material. The dielectric spheres **12** and **14** scatter and concentrate electromagnetic waves. At very sharpely defined frequencies, the spheres **12** and **14** will have resonances wherein the internal energy densities can be five orders of mangitude larger than the energy density of the incident electromagnetic field driving the spheres **12** and **14**. At resonance, the electromagnetic stresses, equivalent to pressures proportional to the energy density, can cause material deformation of the spheres **12** and **14** which produce a secondary electromagnetic field. The spheres **12** and **14** are preferably positioned proximal to each other, as shown in FIG. **1**. Although the proximity of the spheres to each other will adversely affect the resonances, the very high "Q"s of the isolated-sphere resonances results in such adverse affect being relatively small. However, the proximity of the spheres **12** and **14** allows the spheres to interact electromechanically which increases the magnitude of the secondary radiation emitted therefrom.

The electromagnetic radiation incident upon the spheres **12** and **14** which drives the spheres to resonance is preferably zero point radiation **16**. However, other types of electromagnetic radiation may also be used to drive the spheres **12** and **14**, if desired.

The effect of a dielectric sphere such as **12** or **14** on an incident electromagnetic radiation such as a plane wave thereof is shown in FIG. **6**. The plane wave propagates in the z axis direction and is diffracted by the sphere **12** resulting in scattering thereof. This scattering is commonly known as Mie scattering. The incident radiation wave has an electric vector component which is linearly polarized in the x axis direction and a magnetic vector component which is linearly polarized in the y axis direction.

An electromagnetic wave incident upon a structure produces a forced oscillation of free and bound charges in synch with the primary electromagnetic field of the incident electromagnetic wave. The movements of the charges produce a

secondary electromagnetic field both inside and outside the structure. The secondary electromagnetic radiation comprising this secondary electromagnetic field is shown in FIG. **1** and designated by the numerals **18** and **20**. An antenna which is shown simply as a loop antenna but may also be a dipole or any other suitable type of antenna is also shown in FIG. **1** and designated by the numeral **22**. The nonlinear mutual interactions of the spheres produces interference between the secondary electromagnetic radiation **18** and **20** produces a beat frequency radiation **24** which is preferably at a much lower frequency than the primary radiation **16**. It is this beat frequency radiation **24** which is desired for conversion into electrical energy because it preferably is within the frequency range of rf radiation which may be converted into electrical energy by generally conventional systems. Thus, the radiation **24** received by the antenna **22** is fed via an electrical conductor **26** to a means for converting the beat frequency radiation **24** to electrical energy. This means for converting is designated by the numeral **28** and preferably includes a tuning capacitor **30** and a transformer **32** and a rectifier (preferably a diode) **34**. Instead of including the capacitor **30**, transformer **32** and rectifier **34**, the converter **28** may alternatively include an rf receiver of any suitable type.

The resultant field at any point is the vector sum of the primary and secondary fields. For the equations that follow, the structure receiving the incident plane wave is a sphere of radius a having a propagation constant k_1 positioned in an infinite, homogeneous medium having a propagation constant k_2. The incident plane wave propagates in the z axis direction and is as shown in FIG. **6**. The spherical coordinate system used for the vector spherical wave functions is shown in FIG. **7**. Expansion of the incident field provides:

$$E_i = E_o e^{-i\omega t} \sum_{n=1}^{\infty} i^n \frac{2n+1}{n(n+1)} \left(m^{(1)}_{01n} - i n^{(1)}_{eln}\right)$$

$$H_i = -\frac{k_2}{\omega\mu_2} E_o e^{-i\omega t} \sum_{n=1}^{\infty} i^n \frac{2n+1}{n(n+1)} \left(m^{(1)}_{eln} + i n^{(1)}_{01n}\right)$$

where E is the electric field and H is the magnetic field; and

$$\overset{(1)}{m}_{\substack{01n\\e1n}} = \pm\frac{1}{\sin\theta} j_n(k_2R)P_n{}^1(\cos\theta) \begin{smallmatrix}\cos\\ \sin\end{smallmatrix} \phi i_2 - j_n(k_2R)\frac{\partial p_n{}^1}{\partial\theta} \begin{smallmatrix}\sin\\ \cos\end{smallmatrix} \phi i_3$$

$$\overset{(1)}{n}_{\substack{0ln\\e1n}} = \frac{n(n+1)}{k_2R} j_n(k_2R)P_n{}^1(\cos\theta) \begin{smallmatrix}\sin\\ \cos\end{smallmatrix} \phi i_1 + \frac{1}{k_2R}[k_2Rj_n(k_2R)]' \times$$

$$\frac{\partial P_n{}^1}{\partial\theta} \begin{smallmatrix}\sin\\ \cos\end{smallmatrix} \phi i_2 \pm \frac{1}{k_2R\sin\theta}[k_2Rj_n(k_2R)]'P_n{}^1(\cos\theta) \begin{smallmatrix}\cos\\ \sin\end{smallmatrix} \phi i_3.$$

The electric and magnetic fields of the incident wave transmitted into the sphere i.e., R<a, can be similarly expanded:

$$E_t = E_0 e^{-i\omega t} \sum_{n=1}^{\infty} i^n \frac{2n+1}{n(n+1)} \left(a_n^{tm}\,{}^{(1)}_{01n} - i b_n^{tn}\,{}^{(1)}_{e1n} \right)$$

$$H_t = \frac{k_2}{\phi\mu_1} E_0 e^{-i\omega t} \sum_{n=1}^{\infty} i^m \frac{2n+1}{n(n+1)} \left(b_n^{tm}\,{}^{(1)}_{e1n} - i a_n^{tn}\,{}^{(1)}_{01n} \right)$$

If $j_n(k_2R)$ is replaced by $h_n{}^{(1)}(k_2R)$ in the previous equations, the functions $m^{(1)}$ and $n^{(1)}$ become $m^{(3)}$ and $n^{(3)}$. The outgoing fields i.e., R>a, are represented by:

$$E_R = E_0 e^{-i\omega t} \sum_{n=1}^{\infty} i^n \frac{2n+1}{n(n+1)} \left(a_n^{rm}\,{}^{(3)}_{01n} - i b_n^{tn}\,{}^{(3)}_{e1n} \right)$$

$$H_r = \frac{k_2}{\phi\mu_1} E_0 e^{-i\omega t} \sum_{n=1}^{\infty} i^m \frac{2n+1}{n(n+1)} \left(b_n^{rm}\,{}^{(3)}_{e1n} - i a_n^{tn}\,{}^{(3)}_{01n} \right)$$

where H_r represents the resultant wave in the medium surrounding the sphere. At resonance, the values of ρ at resonance require that the ${a_n}^t$ and ${b_n}^t$ coefficients be infinite. In order to determine these values of ${a_n}^t$ and ${b_n}^t$, the boundary conditions at the sphere radius are needed. Since there must be continuity of the E and H values at the surface, the following equations are used:

$i_1 \times (E_i + E_r) = i_1 \times E_t$ and

$i_1 \times (H_i + H_r) = i_1 \times H_t$

which lead to two pairs of inhomogeneous equations:

${a_n}^t j_n(N\rho) - {a_n}^r h_n{}^{(1)}(\rho) = j_n(\rho)$

$\mu_2 {a_n}^t[N\rho j_n(N\rho)]' - \mu_1 {a_n}^r[\rho h_n{}^{(1)}(\rho)]' = \mu_1[\rho j_n(\rho)]'$ and

$\mu_2 N {b_n}^t j_n(N\rho) - \mu_1 {b_n}^r h_n{}^{(1)}(\rho) = \mu_1 j_n(\rho)$

${b_n}^t[N\rho j_n(N\rho)]' - N {b_n}^r[\rho h_n{}^{(1)}(\rho)]' = N[\rho j_n(\rho)]'$

where $k_1 = Nk_2$, $\rho = k_2a$, $k_1a = N\rho$. Spherical Bessel functions of the first kind are denoted by j_n, while those of the third kind are denoted by $h_n{}^{(1)}$. The resulting equations are:

$$a_n{}^t = \frac{\mu_1 j_n(\rho)[\rho h_n^{(1)}(\rho)]' - \mu_1 h_n^{(1)}\rho)[\rho j_n(\rho)]'}{\mu_1 j_n(N\rho)[\rho h_n^{(1)}(\rho)]' - \mu_2 h_n^{(1)}(\rho)[N\rho j_n(N\rho)]'}$$

and

$$b_n{}^t = \frac{\mu_1 N j_n(\rho)[\rho h_n^{(1)}(\rho)]' - \mu_1 N h_n^{(1)}\rho)[\rho j_n(\rho)]'}{\mu_2 N^2 j_n(N\rho)[\rho h_n^{(1)}(\rho)]' - \mu_1 h_n^{(1)}(\rho)[N\rho j_n(N\rho)]'}$$

At a resonance, the denominator of either ${a_n}^t$ or ${b_n}^t$ will be zero. Thus, ρ values are found using the above equations that correspond to a resonant combination of angular frequency (ω) and radius (a) for a given sphere material and given surrounding medium. In determining such values of ρ, the following equations are also specifically used:

$$\rho = ak_2 = a\omega\sqrt{\epsilon_2\mu_2} \text{ and}$$

$$\rho_1 = (k_1/k_2)\rho$$

where ρ_1 corresponds to the sphere material. An iterative method is preferably used to find the desired values of ρ at resonance. In calculating ρ utilizing the above equations for purposes of example, it was assumed that $\mu_1 = \mu_2 = \mu_0 = 4\pi \times 10^{-7}$ and $\epsilon_2 = \epsilon_0 = 8.85419 \times 10^{-12}$.

One major root of ρ which was found has a value of:

Real (ρ)=+66.39752607619131

Imaginary (ρ)=−0.6347867071968998.

These particular values are not shown in FIG. **8**. However, other values of ρ found using the equations set forth herein are shown in FIG. **8**. The peaks in FIG. **8** are the resonances. One of these resonances shown in FIG. **8** is shown in detail in FIG. **9**. These resonance values are shown for purposes of example. Other resonances also exist which have not been determined; thus, not all possible resonance values are shown in FIGS. **8** and **9**.

Calculation of these values also allows the determination of a possible am combination which would have these root values. For ρ, ϵ (epsilon)=ϵ_0 and $\mu=\mu_0$, and

$$\rho = a\omega\sqrt{\epsilon_0\mu_0} = a\omega/c.$$

Expressed in SI units, the speed of light $c=2.99792458\times10^{14}$ m/s. If an a value of 10^{-6} m is assumed for the examples shown herein, then:

$$\omega=\rho c/a\approx1.9919\times10^{16}-i1.9044\times10^{14}\text{ radians/s.}$$

This is an example of the angular frequency required within the impingent EM radiation in order to create a resonant situation. Examples of other resonances were indicated, and these are shown in FIG. **8**. No complex-frequency plane waves exist. Therefore, the calculations were made by considering only the real portion of the above root and setting the imaginary portion equal to zero. However, upon doing this, the iterative calculation procedure becomes insensitive to any root in the vicinity of the root's real portion. In the iterative calculation procedure, initially a range of ρ values is input into the equations. These ρ values are in the neighborhood of the prospective root. A range of ρ values is subsequently studied to find any imaginary ρ i.e., fρ (a function of ρ), peaks in that range. Next, once a peak has been chosen, the function order n giving the dominant fρ is determined. This also gives a clue as to whether the peak is due to a magnetic resonance (a_n approaches infinity) or an electrical resonance (b_n approaches infinity). A large number of Newton-Raphson iterations is preferably performed in order to converge upon a root ρ value.

FIGS. **2** and **3** show a second embodiment of the present invention generally designated by the numeral **110**. Embodiment **110** is essentially the same as embodiment **10** except that the antenna is a rf cavity structure **122** which feeds the received beat frequency radiation **124** to a waveguide **126**. Embodiment **110** also preferably includes two spheres **112** and **114** which receive the primary incident electromagnetic radiation **116** and emit the secondary electromagnetic radiation **118** and **120**. As with the spheres **18** and **20** of embodiment **10**, spheres **118** and **120** are preferably composed of a dielectric material. Embodiment **110** also includes converter **128**, capacitor **130**, transformer **132** and rectifier **134** which are essentially identical to the correspondingly numbered elements of embodiment **10**. Therefore, a description of these components of embodiment **110** will not be repeated in order to promote brevity. In addition, the same equations and method of calculation set forth above with regard to embodiment **10** also apply to embodiment. Therefore, their description will not be repeated in order to promote brevity.

FIGS. **4** and **5** show a third embodiment of the present invention generally designated by numeral **210**. Embodiment **210** is essentially identical to the first embodiment **10** except that the embodiment **210** includes a plurality of pairs **215** of receiving means (spheres) **212** and **214** mounted on a substrate **236**. The spheres **212** and **214** are thus in the form of an array **238**. The pairs **215** of the array **238** are preferably positioned proximal to each other in order to maximize the amount of energy extracted from a particular area or space of a given size. Since, as set forth hereinabove, the energy density of the zero point radiation increases as the frequency of the radiation increases, it is desirable that the spheres resonate at as high a bandwidth of frequencies as possible. Because the spheres **212** and **214** must be small in direct proportion to the wavelength of the high frequencies of the incident electromagnetic radiation **216** at which resonance is desirably obtained, the spheres **212** and **214** are preferably microscopic in size. Current lithographic techniques are capable of manufacturing such microscopically small spheres mounted on a suitable substrate thereby providing a suitably miniaturized system **210**. A miniaturized system enhances the energy output capability of the system by enabling it to resonate at higher frequencies at which there are correspondingly higher energy densities. Consequently, utilization of array **238** in the system **210** enhances the maximum amount of electrical energy provided by the system **210**.

Lithographic techniques may be more amenable to manufacturing microscopically small receiving structures **212** and **214** which may be disc shaped, semispherical or have another shape other than as shown in FIGS. **4** and **5**. Consequently, the receiving means **212** and **214** may accordingly have such alternative shapes rather than the spherical shape shown in FIGS. **4** and **5**. In addition, a large number of small spheres may be manufactured by bulk chemical reactions. Packing a volume with such spheres in close proximity could enhance the output of energy.

Embodiment **210** also includes a plurality of antennas **222** positioned preferably between the spheres **212** and **214** which receive the beat frequency radiation **224** produced by the interference between the secondary radiation **218** and **220**. The antennas **222** are shown as loop antennas **222** but may be any other suitable type of antennas as well.

Embodiment **210** has a plurality of electrical conductors **226** which preferably include traces mounted on the substrate **236** which occupies a finite volume. The electrical conductors **226** feed the electrical output from the antennas **222** to a suitable converter **228** which preferably includes tuning capacitor **230**, transformer **232** and rectifier **234**, as with embodiments **10** and **110**. Except as set forth above, the components of embodiment **210** are identical to embodiment **10** so the detailed description of these components will not be repeated in order to promote brevity. In addition, the same equations and method of calculation set forth above for embodiment **10** also apply to embodiment **210**. Therefore, the description of these equations and method of calculation will not be repeated in order to promote brevity.

Accordingly, there has been provided, in accordance with the invention, a system which converts high frequency zero point electromagnetic radiation into electrical energy effectively and efficiently and thus fully satisfies the objectives set forth above. It is to be understood that all terms used herein are descriptive rather than limiting. Although the invention has been specifically described with regard to the specific embodiments set forth herein, many alternative embodiments, modifications and variations will be apparent to those skilled in the art in light of the disclosure set forth herein. Accordingly, it is intended to include all such alternatives, embodiments, modifications and variations that fall within the spirit and scope of the invention as set forth in the claims hereinbelow.

What is claimed is:

1. A system for converting incident electromagnetic radiation energy to electrical energy, comprising:

a first means for receiving incident primary electromagnetic radiation, said means for receiving producing emitted secondary electromagnetic radiation at a first frequency, said first means for receiving having a first volumetric size selected to resonate at a frequency within the frequency spectrum of the incident primary electromagnetic radiation in order to produce the secondary electromagnetic radiation at the first frequency at an enhanced energy density;

a second means for receiving the incident primary electromagnetic radiation, said means for receiving producing emitted secondary electromagnetic radiation at a second frequency, the secondary radiation at the first frequency and the secondary radiation at the second frequency interfering to produce secondary radiation at a lower frequency than that of the incident primary radiation, said second means for receiving having a second volumetric size selected to resonate at a frequency within the frequency spectrum of the incident primary electromagnetic radiation in order to produce the emitted secondary electromagnetic radiation at the second frequency at an enhanced energy density;

an antenna for receiving the emitted secondary electromagnetic radiation at the lower frequency, said antenna providing an electrical output responsive to the secondary electromagnetic radiation received;

a converter electrically connected to said antenna for receiving electrical current output from said antenna and converting the electrical current output to electrical current having a desired voltage and waveform.

2. The system of claim **1** wherein:

said first means for receiving is composed of a dielectric material; and

said second means for receiving is composed of a dielectric material.

3. The system of claim **1** wherein:

said first means for receiving is spherical; and

said second means for receiving is spherical.

4. A system for for converting incident zero point electromagnetic radiation energy to electrical energy, comprising:

a first means for receiving incident primary zero point electromagnetic radiation, said means for receiving producing emitted secondary electromagnetic radiation at a first frequency;

a second means for receiving the incident primary zero point electromagnetic radiation, said means for receiving producing emitted secondary electromagnetic radiation at a second frequency, the secondary radiation at the first frequency and the secondary radiation at the second frequency interfering to produce secondary radiation at a beat frequency which is lower than that of the incident primary radiation;

an antenna for receiving the emitted secondary electromagnetic radiation at the lower frequency, said antenna providing an electrical output responsive to the secondary electromagnetic radiation received;

means for transmitting the emitted secondary electromagnetic radiation at the beat frequency from said antenna, said means for transmitting connected to said antenna;

a converter connected to said means for transmitting for receiving the emitted secondary electromagnetic radiation at the beat frequency from said antenna and converting the same to electrical current having a desired voltage and waveform.

5. The system of claim **4** wherein:

said first means for receiving has a first volumetric spherical size selected to resonate in response to the incident primary electromagnetic radiation in order to produce the secondary electromagnetic radiation at the first frequency at an enhanced energy density; and

said second means for receiving has a second volumetric spherical size selected to resonate in response to the incident primary electromagnetic radiation in order to produce the emitted secondary electromagnetic radiation at the second frequency at an enhanced energy density, said first and second volumetric sizes selected based on parameters of propagation constant of said first and second means for receiving, propagation constant of medium in which said first and second means for receiving are located and frequency of the incident primary electromagnetic radiation.

6. The system of claim **5** wherein the first and second volumetric sizes are selected by utilizing the formulas:

$$a_n{}^t=\frac{\mu_1 j_n(\rho)[\rho h_n^{(1)}(\rho)]'-\mu_1 h_n^{(1)}\rho)[\rho j_n(\rho)]'}{\mu_1 j_n(N\rho)[\rho h_n^{(1)}(\rho)]'-\mu_2 h_n^{(1)}(\rho)[N\rho j_n(N\rho)]'}$$

$$b_n{}^t=\frac{\mu_1 N\mathrm{j}_n(\rho)[\rho h_n^{(1)}(\rho)]'-\mu_1 \mathrm{N} h_n^{(1)}\rho)[\rho j_n(\rho)]'}{\mu_2 N^2 j_n(N\rho)[\rho h_n^{(1)}(\rho)]'-\mu_1 h_n^{(1)}(\rho)[N\rho j_n(N\rho)]'}$$

$$\rho=a\omega\sqrt{\epsilon_2\mu_2}$$

wherein at a resonance, the denominator of either equation for $a_n{}^t$ or $b_n{}^t$ will be approximately zero and wherein k_1=propagation constant of the means for receiving, k_2=propagation constant of medium through which the incident electromagnetic radiation propagates, a is the radius of either means for receiving, $N=k_1/k_2$, $\rho=k_2a$, $k_1a=N\rho$, $a_n{}^t$=magnitude of oscillations of the electric field of the nth order, $b_n{}^t$=magnitude of oscillations of the magnetic field of the nth order, ω=angular frequency of the incident electromagnetic radiation, ϵ is the permittivity of the material or medium and μ is the permeability of the material or medium.

7. The system of claim **6** wherein the radius of the first means for receiving is different from the radius of the second means for receiving, difference between the radius of said first means for receiving and the radius of said second means for receiving selected so that the beat frequency resulting from the difference is a frequency which facilitates conversion of the beat frequency electromagnetic radiation to electrical energy.

8. The system of claim **4** wherein:

said first means for receiving is composed of a dielectric material; and

said second means for receiving is composed of a dielectric material.

9. The system of claim **4** wherein:

said first means for receiving is spherical; and

said second means for receiving is spherical.

10. The system of claim **4** wherein said antenna is positioned generally between said first and second means for receiving.

11. The system of claim **4** wherein said antenna is a loop antenna.

12. The system of claim **4** wherein said antenna is a generally concave shell partially enclosing said first and second means for receiving.

13. The system of claim **4** wherein said means for transmitting is a waveguide.

14. A system for for converting incident zero point electromagnetic radiation energy to electrical energy, comprising:

a substrate;

a plurality of pairs of first means for receiving incident primary zero point electromagnetic radiation and second means for receiving incident primary zero point electromagnetic radiation, said plurality of pairs of means for receiving mounted on said substrate, said first means for receiving producing emitted secondary

electromagnetic radiation at a first frequency, said second means for receiving the incident primary zero point electromagnetic radiation producing emitted secondary electromagnetic radiation at a second frequency, the secondary radiation at the first frequency and the secondary radiation at the second frequency interfering to produce secondary radiation at a beat frequency which is lower than that of the incident primary radiation, said first means for receiving having a first volumetric size selected to resonate in response to the incident primary electromagnetic radiation in order to produce the secondary electromagnetic radiation at the first frequency at an enhanced energy density, and said second means for receiving having a second volumetric size selected to resonate in response to the incident primary electromagnetic radiation in order to produce the emitted secondary electromagnetic radiation at the second frequency at an enhanced energy density, said first and second volumetric sizes selected based on parameters of propagation constant of said first and second means for receiving, propagation constant of medium in which said first and second means for receiving are located and frequency of the

incident primary electromagnetic radiation, said first and second volumetric sizes being different from each other;

a plurality of antennas for receiving the emitted secondary electromagnetic radiation at the lower frequency, said antenna providing an output responsive to the secondary electromagnetic radiation received, said plurality of antennas mounted on said substrate, each of said plurality of antennas receiving the emitted secondary electromagnetic radiation of one of said pairs of first and second means for receiving;

means for transmitting the emitted secondary electromagnetic radiation at the beat frequency from said antenna, said means for transmitting connected to said plurality of antennas;

a converter connected to said means for transmitting for receiving the emitted secondary electromagnetic radiation at the beat frequency from said antenna and converting the same to electrical current having a desired voltage and waveform.

* * * * *

US006362718B1

(12) **United States Patent**
Patrick et al.

(10) **Patent No.: US 6,362,718 B1**
(45) **Date of Patent: Mar. 26, 2002**

---

(54) **MOTIONLESS ELECTROMAGNETIC GENERATOR**

(76) Inventors: **Stephen L. Patrick**, 2511 Woodview Dr. SE.; **Thomas E. Bearden**, 2211 Cove Rd., both of Huntsville, AL (US) 35801; **James C. Hayes**, 16026 Deaton Dr. SE., Huntsville, AL (US) 35803; **Kenneth D. Moore**, 1704 Montdale Rd., Huntsville, FL (US) 35801; **James L. Kenny**, 925 Tascosa Dr., Huntsville, AL (US) 35802

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/656,313**

(22) Filed: **Sep. 6, 2000**

(51) **Int. Cl.**^{7} **H01F 27/24**

(52) **U.S. Cl.** **336/214**

(58) **Field of Search** 363/16, 24, 25, 363/26, 56.06, 56.08, 133, 134; 336/15, 110, 155, 177, 180, 213, 214, 221, 222

(56) **References Cited**

U.S. PATENT DOCUMENTS

2,153,378 A		4/1939	Kramer	171/95
2,892,155 A		6/1959	Radus et al.	324/117
3,079,535 A		2/1963	Schultz	317/201
3,165,723 A		1/1965	Radus	340/174
3,228,013 A		1/1966	Olson et al.	340/174
3,254,268 A		5/1966	Radus et al.	317/14
3,316,514 A		4/1967	Radus et al.	335/291
3,368,141 A		2/1968	Subieta-Garron	323/44
3,391,358 A		7/1968	Bratkowski et al.	335/21
3,453,876 A		7/1969	Radus	73/141
3,517,300 A	*	6/1970	McMurray	
3,569,947 A		3/1971	Radus	340/174
3,599,074 A	*	8/1971	Adams	
4,006,401 A		2/1977	de Rivas	323/92
4,077,001 A		2/1978	Richardson	323/92
4,366,532 A		12/1982	Rosa et al.	363/69
4,482,945 A		11/1984	Wolf et al.	363/129
4,554,524 A		11/1985	Radus	337/3
4,853,668 A		8/1989	Bloom	336/214
4,864,478 A		9/1989	Bloom	363/16
4,904,926 A		2/1990	Pasichinskyj	323/362
5,011,821 A		4/1991	McCullough	505/1
5,221,892 A		6/1993	Sullivan et al.	323/362
5,245,521 A	*	9/1993	Spreen	363/37
5,327,015 A		7/1994	Hacket	505/211
5,335,163 A		8/1994	Seiersen	363/126
5,694,030 A	*	12/1997	Sato et al.	323/282

OTHER PUBLICATIONS

Raymond J. Radus, “Permanent–Magnet Circuit using a ‘Flux–Transfer’ Principle,” Engineers’ Digest, 24(1–6) Jan.–Jun. 1963, p. 86.

Robert O’Handley, Modern Magnetic Materials, Principles and Applications, John Wiley & Sons, Inc., 2000, pp. 456–468.

Robert C. Weast, Editor, CRC Handbook of Chemistry and Physics, 1978–1979, p. B–50.

Honeywell.com web site, “amorphous metals”.

* cited by examiner

Primary Examiner—Matthew Nguyen
(74) *Attorney, Agent, or Firm*—Norman Friedland

(57) **ABSTRACT**

An electromagnetic generator without moving parts includes a permanent magnet and a magnetic core including first and second magnetic paths. A first input coil and a first output coil extend around portions of the first magnetic path, while a second input coil and a second output coil extend around portions of the second magnetic path. The input coils are alternatively pulsed to provide induced current pulses in the output coils. Driving electrical current through each of the input coils reduces a level of flux from the permanent magnet within the magnet path around which the input coil extends. In an alternative embodiment of an electromagnetic generator, the magnetic core includes annular spaced-apart plates, with posts and permanent magnets extending in an alternating fashion between the plates. An output coil extends around each of these posts. Input coils extending around portions of the plates are pulsed to cause the induction of current within the output coils.

29 Claims, 5 Drawing Sheets

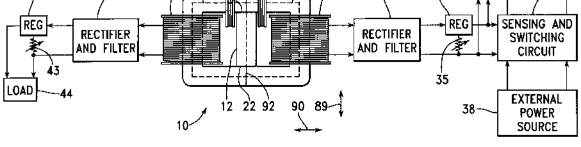

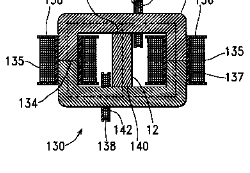

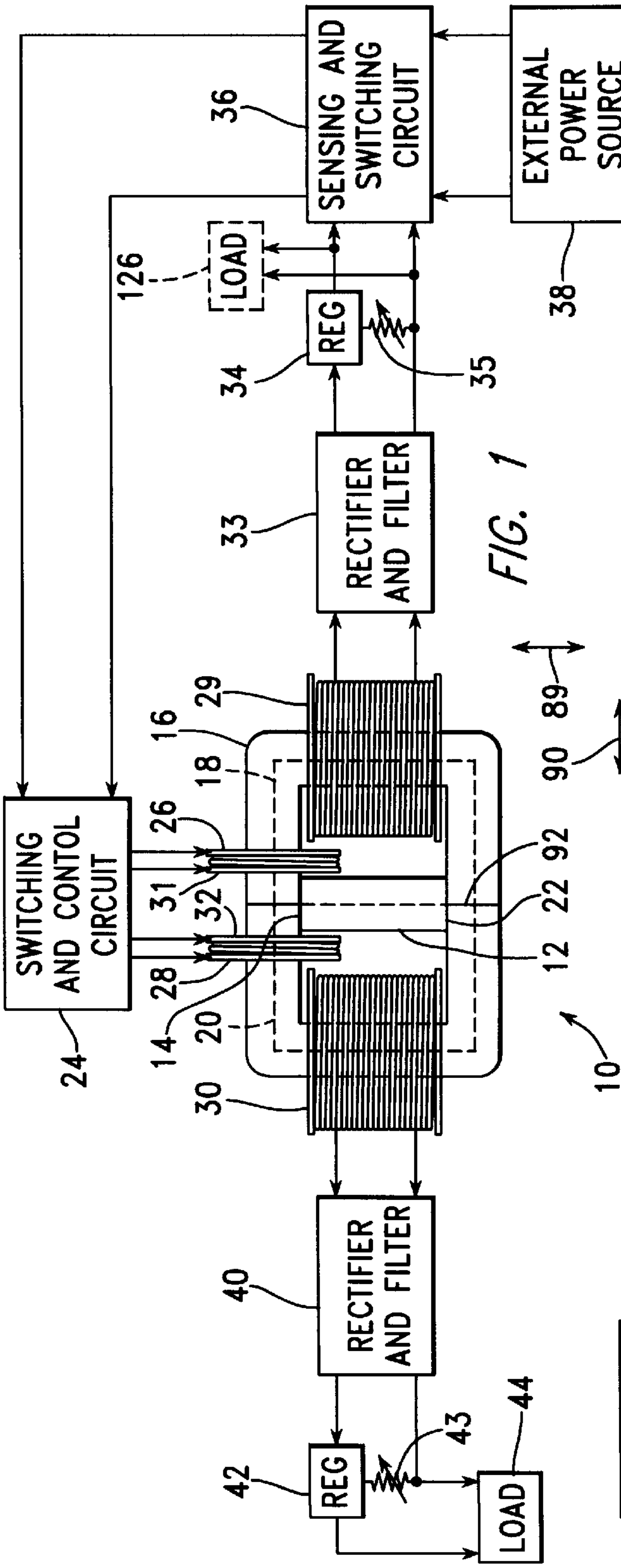
SWITCHING AND CONTOL CIRCUIT
24
26
31
32
28
14
20
30
18
16
29
22
92
12
10
90
89
RECTIFIER AND FILTER
40
REG
42
43
LOAD
44
RECTIFIER AND FILTER
33
REG
34
35
LOAD
126
SENSING AND SWITCHING CIRCUIT
36
EXTERNAL POWER SOURCE
38

FIG. 1

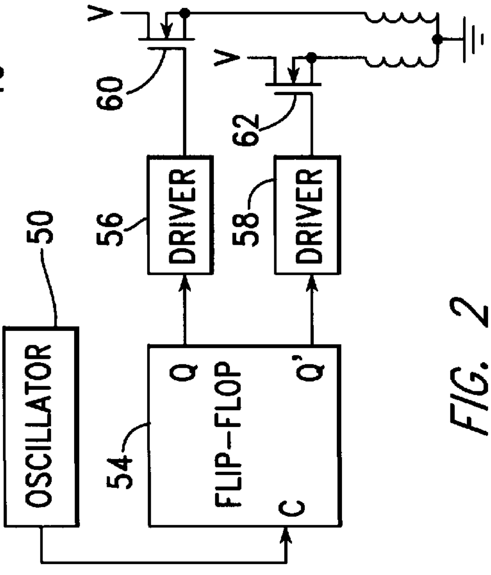
OSCILLATOR
50
FLIP-FLOP
Q
Q'
C
54
DRIVER
56
DRIVER
58
60
62
V
V

FIG. 2

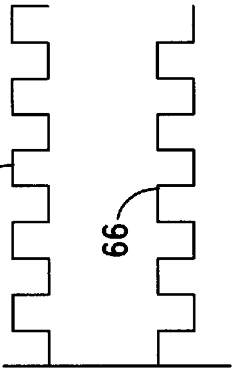

FIG. 3

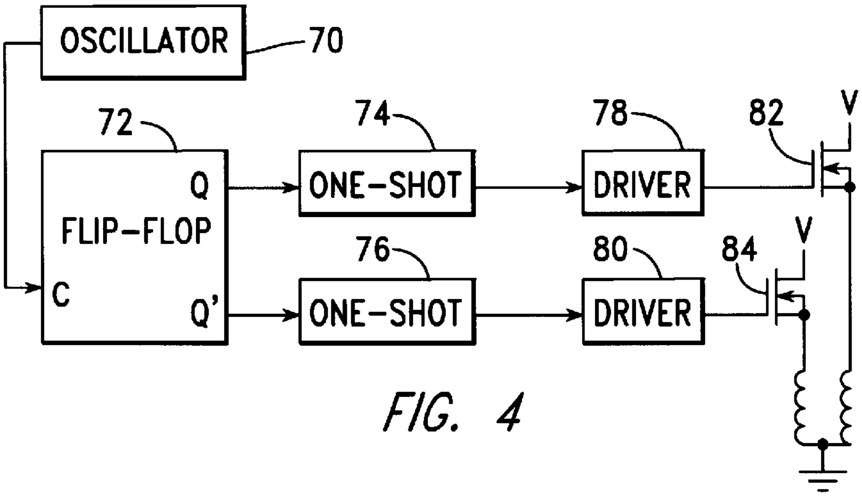
OSCILLATOR
70
72
FLIP-FLOP
Q
C
Q'
74
ONE-SHOT
76
ONE-SHOT
78
DRIVER
80
DRIVER
82
84
V
V

FIG. 4

FIG. 5

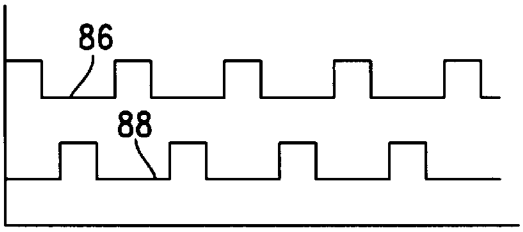

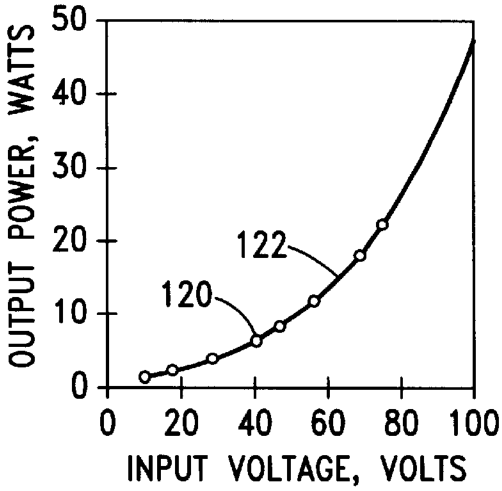

FIG. 7

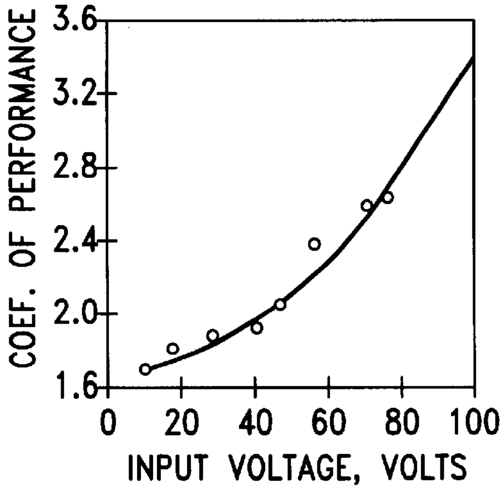

FIG. 8

FIG. 6A

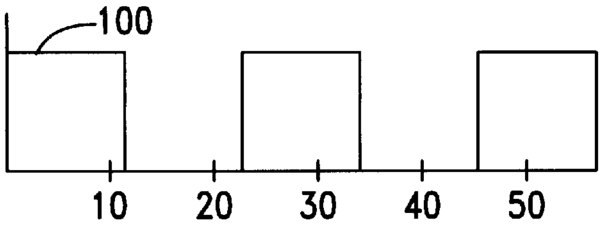

FIG. 6B

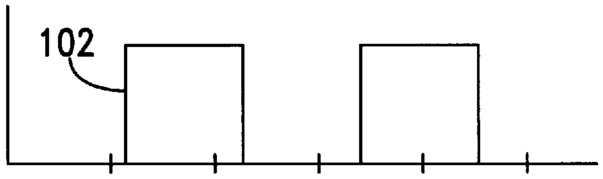

FIG. 6C

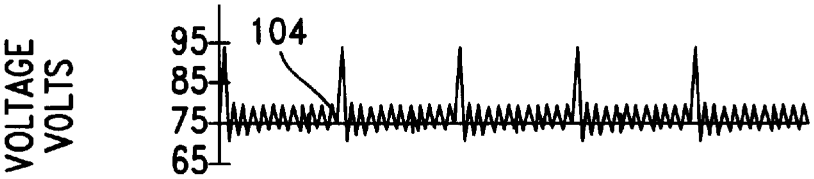

FIG. 6D

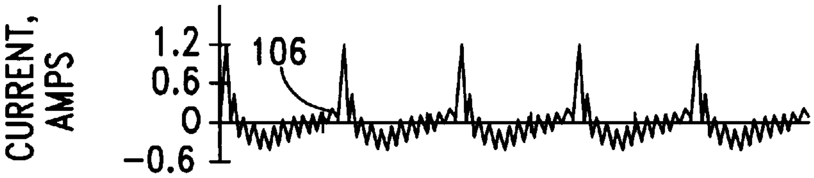

FIG. 6E

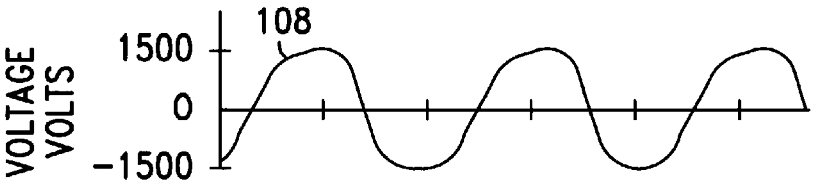

FIG. 6F

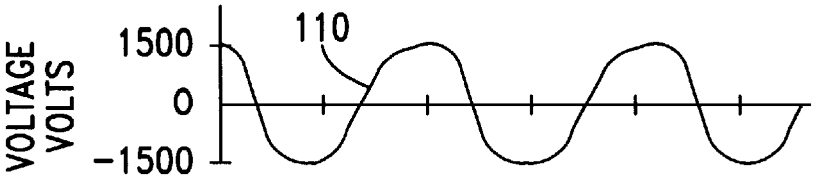

FIG. 6G

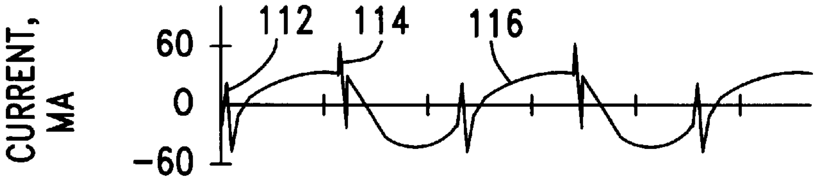

FIG. 6H

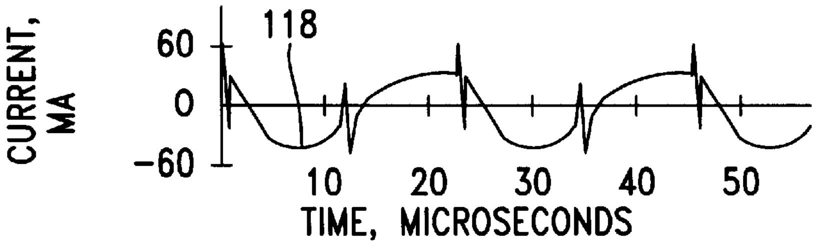

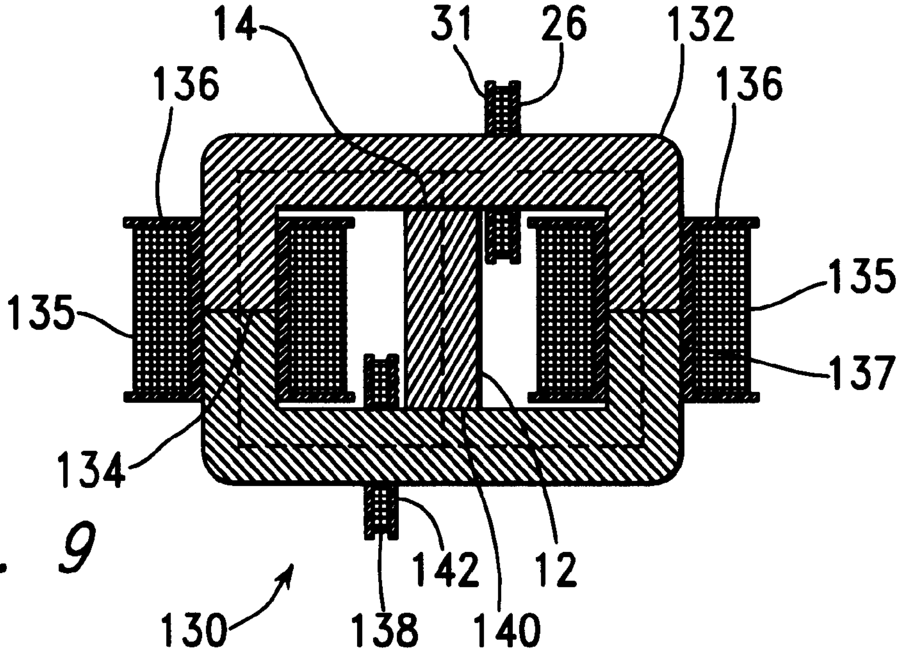

FIG. 9

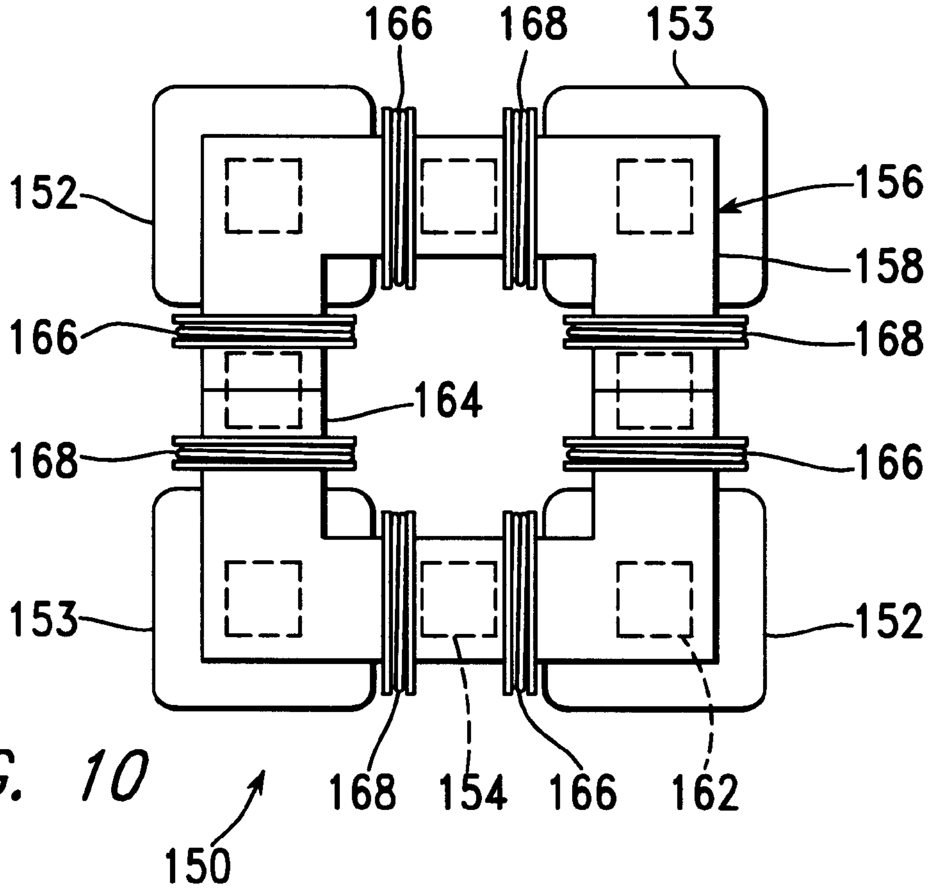

FIG. 10

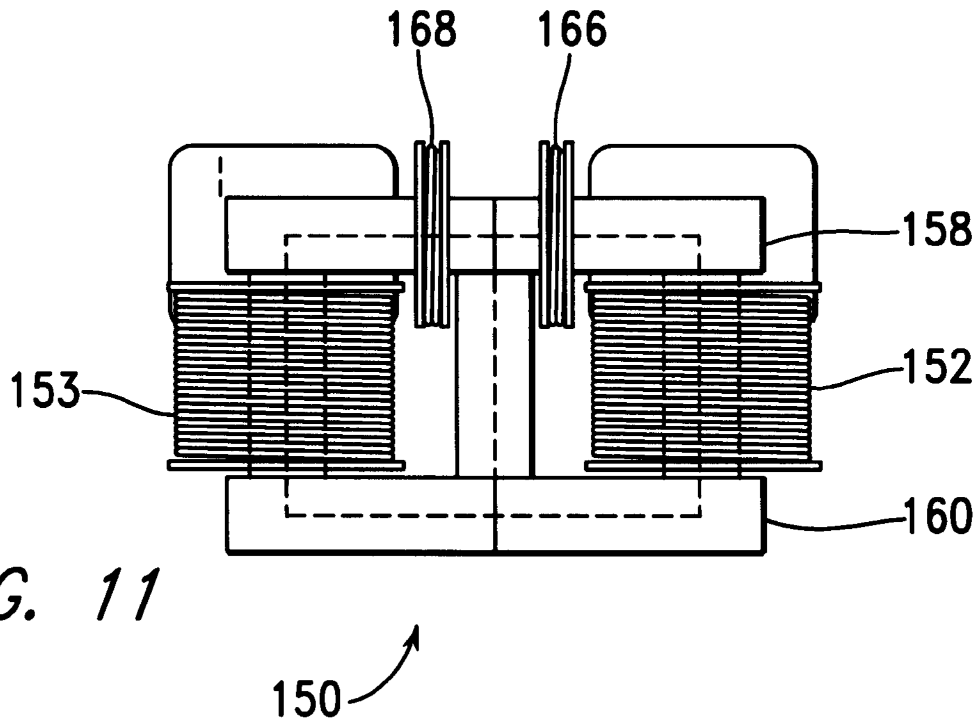
168
166
158
153
152
160
150

FIG. 11

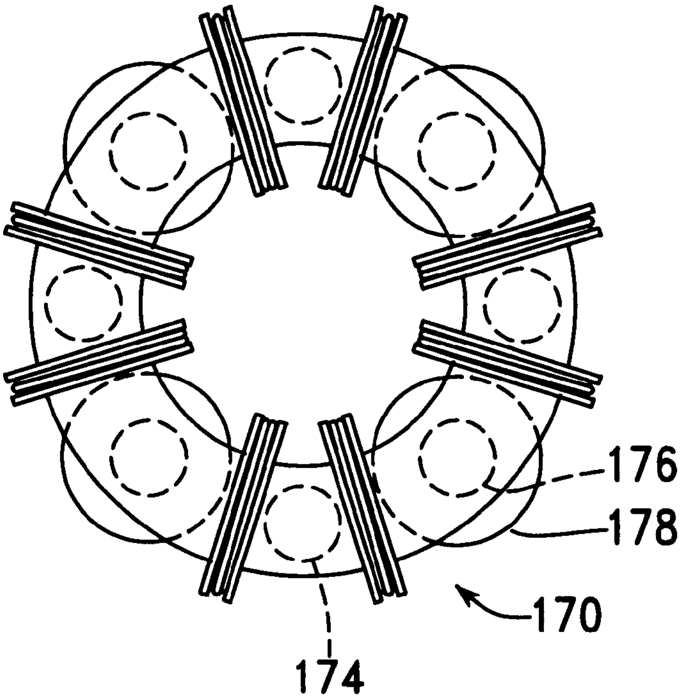
176
178
170
174

FIG. 12

MOTIONLESS ELECTROMAGNETIC GENERATOR

BACKGROUND INFORMATION

1. Field of Invention

This invention relates to a magnetic generator used to produce electrical power without moving parts, and, more particularly, to such a device having a capability, when operating, of producing electrical power without an external application of input power through input coils.

2. Description of the Related Art

The patent literature describes a number of magnetic generators, each of which includes a permanent magnet, two magnetic paths external to the permanent magnet, each of which extends between the opposite poles of the permanent magnet, switching means for causing magnetic flux to flow alternately along each of the two magnetic paths, and one or more output coils in which current is induced to flow by means of changes in the magnetic field within the device. These devices operate in accordance with an extension of Faraday's Law, indicating that an electrical current is induced within a conductor within a changing magnetic field, even if the source of the magnetic field is stationary.

A method for switching magnetic flux to flow predominantly along either of two magnetic paths between opposite poles of a permanent magnet is described as a "flux transfer" principle by R. J. Radus in *Engineer's Digest*, Jul. 23, 1963. This principle is used to exert a powerful magnetic force at one end of both the north and south poles and a very low force at the other end, without being used in the construction of a magnetic generator. This effect can be caused mechanically, by keeper movement, or electrically, by driving electrical current through one or more control windings extending around elongated versions of the pole pieces **14**. Several devices using this effect are described in U.S. Pat. Nos. 3,165,723, 3,228,013, and 3,316,514, which are incorporated herein by reference.

Another step toward the development of a magnetic generator is described in U.S. Pat. No. 3,368,141, which is incorporated herein by reference, as a device including a permanent magnet in combination with a transformer having first and second windings about a core, with two paths for magnetic flux leading from each pole of the permanent magnet to either end of the core, so that, when an alternating current induces magnetic flux direction changes in the core, the magnetic flux from the permanent magnet is automatically directed through the path which corresponds with the direction taken by the magnetic flux through the core due to the current. In this way, the magnetic flux is intensified. This device can be used to improve the power factor of a typically inductively loaded alternating current circuit.

Other patents describe magnetic generators in which electrical current from one or more output coils is described as being made available to drive a load, in the more conventional manner of a generator. For example, U.S. Pat. No. 4,006,401, which is incorporated herein by reference, describes an electromagnetic generator including permanent magnet and a core member, in which the magnetic flux flowing from the magnet in the core member is rapidly alternated by switching to generate an alternating current in a winding on the core member. The device includes a permanent magnet and two separate magnetic flux circuit paths between the north and south poles of the magnet. Each of the circuit paths includes two switching means for alternately opening and closing the circuit paths, generating an alternating current in a winding on the core member. Each of the switching means includes a switching magnetic circuit intersecting the circuit path, with the switching magnetic circuit having a coil through which current is driven to induce magnetic flux to saturate the circuit path extending to the permanent magnet. Power to drive these coils is derived directly from the output of a continuously applied alternating current source. What is needed is an electromagnetic generator not requiring the application of such a current source.

U.S. Pat. No. 4,077,001, which is incorporated herein by reference, describes a magnetic generator, or dc/dc converter, comprising a permanent magnet having spaced-apart poles and a permanent magnetic field extending between the poles of the magnet. A variable-reluctance core is disposed in the field in fixed relation to the magnet and the reluctance of the core is varied to cause the pattern of lines of force of the magnetic field to shift. An output conductor is disposed in the field in fixed relation to the magnet and is positioned to be cut by the shifting lines of permanent magnetic force so that a voltage is induced in the conductor. The magnetic flux is switched between alternate paths by means of switching coils extending around portions of the core, with the flow of current being alternated between these switching coils by means of a pair of transistors driven by the outputs of a flip-flop. The input to the flip flop is driven by an adjustable frequency oscillator. Power for this drive circuit is supplied through an additional, separate power source. What is needed is a magnetic generator not requiring the application of such a power source.

U.S. Pat. No. 4,904,926, which is incorporated herein by reference, describes another magnetic generator using the motion of a magnetic field. The device includes an electrical winding defining a magnetically conductive zone having bases at each end, the winding including elements for the removing of an induced current therefrom. The generator further includes two pole magnets, each having a first and a second pole, each first pole in magnetic communication with one base of the magnetically conductive zone. The generator further includes a third pole magnet, the third pole magnet oriented intermediately of the first poles of the two pole electromagnets, the third pole magnet having a magnetic axis substantially transverse to an axis of the magnetically conductive zone, the third magnet having a pole nearest to the conductive zone and in magnetic attractive relationship to the first poles of the two pole electromagnets, in which the first poles thereof are like poles. Also included in the generator are elements, in the form of windings, for cyclically reversing the magnetic polarities of the electromagnets. These reversing means, through a cyclical change in the magnetic polarities of the electromagnets, cause the magnetic flux lines associated with the magnetic attractive relationship between the first poles of the electromagnets and the nearest pole of the third magnet to correspondingly reverse, causing a wiping effect across the magnetically conductive zone, as lines of magnetic flux swing between respective first poles of the two electromagnets, thereby inducing electron movement within the output windings and thus generating a flow of current within the output windings.

U.S. Pat. No. 5,221,892, which is incorporated herein by reference, describes a magnetic generator in the form of a direct current flux compression transformer including a magnetic envelope having poles defining a magnetic axis and characterized by a pattern of magnetic flux lines in polar symmetry about the axis. The magnetic flux lines are spatially displaced relative to the magnetic envelope using control elements which are mechanically stationary relative to the core. Further provided are inductive elements which are also mechanically stationary relative to the magnetic envelope. Spatial displacement of the flux relative to the inductive elements causes a flow of electrical current. Further provided are magnetic flux valves which provide for the varying of the magnetic reluctance to create a time domain pattern of respectively enhanced and decreased magnetic reluctance across the magnetic valves, and, thereby, across the inductive elements.

Other patents describe devices using superconductive elements to cause movement of the magnetic flux. These devices operate in accordance with the Meissner effect, which describes the expulsion of magnetic flux from the interior of a superconducting structure as the structure undergoes the transition to a superconducting phase. For example, U.S. Pat. No. 5,011,821, which is incorporated herein by reference, describes an electric power generating device including a bundle of conductors which are placed in a magnetic field generated by north and south pole pieces of a permanent magnet. The magnetic field is shifted back and forth through the bundle of conductors by a pair of thin films of superconductive material. One of the thin films is placed in the superconducting state while the other thin film is in a non-superconducting state. As the states are cyclically reversed between the two films, the magnetic field is deflected back and forth through the bundle of conductors.

U.S. Pat. No. 5,327,015, which is incorporated herein by reference, describes an apparatus for producing an electrical impulse comprising a tube made of superconducting material, a source of magnetic flux mounted about one end of the tube, a means, such as a coil, for intercepting the flux mounted along the tube, and a means for changing the temperature of the superconductor mounted about the tube. As the tube is progressively made superconducting, the magnetic field is trapped within the tube, creating an electrical impulse in the means for intercepting. A reversal of the superconducting state produces a second pulse.

None of the patented devices described above use a portion of the electrical power generated within the device to power the reversing means used to change the path of magnetic flux. Thus, like conventional rotary generators, these devices require a steady input of power, which may be in the form of electrical power driving the reversing means of one of these magnetic generators or the torque driving the rotor of a conventional rotary generator. Yet, the essential function of the magnetic portion of an electrical generator is simply to switch magnetic fields in accordance with precise timing. In most conventional applications of magnetic generators, the voltage is switched across coils, creating magnetic fields in the coils which are used to override the fields of permanent magnets, so that a substantial amount of power must be furnished to the generator to power the switching means, reducing the efficiency of the generator.

Recent advances in magnetic material, which have particularly been described by Robert C. O'Handley in *Modern Magnetic Materials, Principles and Applications*, John Wiley & Sons, New York, pp. 456–468, provide nanocrystalline magnetic alloys, which are particularly well suited forth rapid switching of magnetic flux. These alloys are primarily composed of crystalline grains, or crystallites, each of which has at least one dimension of a few nanometers. Nanocrystalline materials may be made by heat-treating amorphous alloys which form precursors for the nanocrystalline materials, to which insoluble elements, such as copper, are added to promote massive nucleation, and to which stable, refractory alloying materials, such as niobium or tantalum carbide are added to inhibit grain growth. Most of the volume of nanocrystalline alloys is composed of randomly distributed crystallites having dimensions of about 2–40 nm. These crystallites are nucleated and grown from an amorphous phase, with insoluble elements being rejected during the process of crystallite growth. In magnetic terms, each crystallite is a single-domain particle. The remaining volume of nanocrystalline alloys is made up of an amorphous phase in the form of grain boundaries having a thickness of about 1 nm.

Magnetic materials having particularly useful properties are formed from an amorphous Co—Nb—B (cobalt-niobium-boron) alloy having near-zero magnetostriction and relatively strong magnetization, as well as good mechanical strength and corrosion resistance. A process of annealing this material can be varied to change the size of crystallites formed in the material, with a resulting strong effect on DC coercivity. The precipitation of nanocrystallites also enhances AC performance of the otherwise amorphous alloys.

Other magnetic materials are formed using iron-rich amorphous and nanocrystalline alloys, which generally show larger magnetization that the alloys based on cobalt. Such materials are, for example, Fe—B—Si—Nb—Cu (iron-boron-silicon-niobium-copper) alloys. While the permeability of iron-rich amorphous alloys is limited by their relatively large levels of magnetostriction, the formation of a nanocrystalline material from such an amorphous alloy dramatically reduces this level of magnetostriction, favoring easy magnetization.

Advances have also been made in the development of materials for permanent magnets, particularly in the development of materials including rare earth elements. Such materials include samarium cobalt, $SmCo_5$, which is used to form a permanent magnet material having the highest resistance to demagnetization of any known material. Other magnetic materials are made, for example, using combinations of iron, neodymium, and boron.

SUMMARY OF THE INVENTION

It is a first objective of the present invention to provide a magnetic generator which a need for an external power source during operation of the generator is eliminated.

It is a second objective of the present invention to provide a magnetic generator in which a magnetic flux path is changed without a need to overpower a magnetic field to change its direction.

It is a third objective of the present invention to provide a magnetic generator in which the generation of electricity is accomplished without moving parts.

In the apparatus of the present invention, the path of the magnetic flux from a permanent magnet is switched in a manner not requiring the overpowering of the magnetic fields. Furthermore, a process of self-initiated iterative switching is used to switch the magnetic flux from the permanent magnet between alternate magnetic paths within the apparatus, with the power to operate the iterative switching being provided through a control circuit consisting of components known to use low levels of power. With self-switching, a need for an external power source during operation of the generator is eliminated, with a separate power source, such as a battery, being used only for a very short time during start-up of the generator.

According to a first aspect of the present invention, an electromagnetic generator is provided, including a permanent magnet, a magnetic core, first and second input coils, first and second output coils, and a switching circuit. The permanent magnet has magnetic poles at opposite ends. The magnetic core includes a first magnetic path, around which the first input and output coils extend, and a second magnetic path, around which the second input and output coils extend, between opposite ends of the permanent magnet. The switching circuit drives electrical current alternately through the first and second input coils. The electrical current driven through the first input oil causes the first input coil to produce a magnetic field opposing a concentration of magnetic flux from the permanent magnet within the first magnetic path. The electrical current driven through the second input coil causes the second input coil to produce a magnetic field opposing a concentration of magnetic flux from the permanent magnet within the second magnetic path.

According to another aspect of the present invention, an electromagnetic generator is provided, including a magnetic core, a plurality of permanent magnets, first and second pluralities of input coils, a plurality of output coils, and a switching circuit. The magnetic core includes a pair of spaced-apart plates, each of which has a central aperture, and first and second pluralities of posts extending between the spaced-apart plates. The permanent magnets each extend between the pair of spaced apart plates. Each permanent magnet has magnetic poles at opposite ends, with the magnetic fields of all the permanent magnets being aligned to extend in a common direction. Each input coil extends around a portion of a plate within the spaced-apart plates, between a post and a permanent magnet. An output coil extends around each post. The switching circuit drives electrical current alternately through the first and second pluralities of input coils. Electrical current driven through each input coil in the first plurality of input coils causes an increase in magnetic flux within each post within the first plurality of posts from permanent magnets on each side of the post and a decrease in magnetic flux within each post within the second plurality of posts from permanent magnets on each side of the post. Electrical current driven through each input coil in the second plurality of input coils causes a decrease in magnetic flux within each post within the first plurality of posts from permanent magnets on each side of the post and an increase in magnetic flux within each post within the second plurality of posts from permanent magnets on each side of the post.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a partly schematic front elevation of a magnetic generator and associated electrical circuits built in accordance with a first version of the first embodiment of the present invention;

FIG. **2** is a schematic view of a first version of a switching and control circuit within the associated electrical circuits of FIG. **1**;

FIG. **3** is a graphical view of drive signals produced within the circuit of FIG. **2**;

FIG. **4** is a schematic view of a second version of a switching and control circuit within the associated electrical circuits of FIG. **1**;

FIG. **5** is a graphical view of drive signals produced within the circuit of FIG. **3**;

FIG. **6**A is a graphical view of a first drive signal within the apparatus of FIG. **1**;

FIG. **6**B is a graphical view of a second drive signal within the apparatus of FIG. **1**;

FIG. **6**C is a graphical view of an input voltage signal within the apparatus of FIG. **1**;

FIG. **6**D is a graphical view of an input current signal within the apparatus of FIG. **1**;

FIG. **6**E is a graphical view of a first output voltage signal within the apparatus of FIG. **1**;

FIG. **6**F is a graphical view of a second output voltage signal within the apparatus of FIG. **1**;

FIG. **6**G is a graphical view of a first output current signal within the apparatus of FIG. **1**;

FIG. **6**H is a graphical view of a second output current signal within the apparatus of FIG. **1**;

FIG. **7** is a graphical view of output power measured within the apparatus of FIG. **1**, as a function of input voltage;

FIG. **8** is a graphical view of a coefficient of performance, calculated from measurements within the apparatus of FIG. **1**, as a function of input voltage;

FIG. **9** is a cross-sectional elevation of a second version of the first embodiment of the present invention;

FIG. **10** is a top view of a magnetic generator built in accordance with a first version of a second embodiment of the present invention;

FIG. **11** is a front elevation of the magnetic generator of FIG. **10**; and

FIG. **12** is a top view of a magnetic generator built in accordance with a second version of the second embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

FIG. **1** is a partly schematic front elevation of an electromagnetic generator **10**, built in accordance with a first embodiment of the present invention to include a permanent magnet **12** to supply input lines of magnetic flux moving from the north pole **14** of the magnet **12** outward into magnetic flux path core material **16**. The flux path core material **16** is configured to form a right magnetic path **18** and a left magnetic path **20**, both of which extend externally between the north pole **14** and the south pole **22** of the magnet **12**. The electromagnetic generator **10** is driven by means of a switching and control circuit **24**, which alternately drives electrical current through a right input coil **26** and a left input coil **28**. These input coils **26**, **28** each extend around a portion of the core material **16**, with the right input coil **26** surrounding a portion of the right magnetic path **18** and with the left input coil **28** surrounding a portion of the left magnetic path **20**. A right output coil **29** also surrounds a portion of the right magnetic path **18**, while a left output coil **30** surrounds a portion of the left magnetic path **20**.

In accordance with a preferred version of the present invention, the switching and control circuit **24** and the input coils **26**, **28** are arranged so that, when the right input coil **26** is energized, a north magnetic pole is present at its left end **31**, the end closest to the north pole **14** of the permanent magnet **12**, and so that, when the left input coil **28** is energized, a north magnetic pole is present at its right end **32**, which is also the end closest to the north pole **14** of the permanent magnet **12**. Thus, when the right input coil **26** is magnetized, magnetic flux from the permanent magnet **12** is repelled from extending through the right input coil **26**. Similarly, when the left input coil **28** is magnetized, magnetic flux from the permanent magnet **12** is repelled from extending through the left input coil **28**.

Thus, it is seen that driving electrical current through the right input coil **26** opposes a concentration of flux from the permanent magnet **12** within the right magnetic path **18**, causing at least some of this flux to be transferred to the left magnetic path **20**. On the other hand, driving electrical current through the left input coil **28** opposes a concentration of flux from the permanent magnet **12** within the left magnetic path **20**, causing at least some of this flux to be transferred to the right magnetic path **18**.

While in the example of FIG. **1**, the input coils **26, 28** are placed on either side of the north pole of the permanent magnet **12**, being arranged along a portion of the core **16** extending from the north pole of the permanent magnet **12**, it is understood that the input coils **26, 28** could as easily be alternately placed on either side of the south pole of the permanent magnet **12**, being arranged along a portion of the core **16** extending from the south pole of the permanent magnet **12**, with the input coils **26, 28** being wired to form, when energized, magnetic fields having south poles directed toward the south pole of the permanent magnet **12**. In general, the input coils **26, 28** are arranged along the magnetic core on either side of an end of the permanent magnet forming a first pole, such as a north pole, with the input coils being arranged to produce magnetic fields of the polarity of the first pole directed toward the first pole of the permanent magnet.

Further in accordance with a preferred version of the present invention, the input coils **26, 28** are never driven with so much current that the core material **16** becomes saturated. Driving the core material **16** to saturation means that subsequent increases in input current can occur without effecting corresponding changes in magnetic flux, and therefore that input power can be wasted. In this way, the apparatus of the present invention is provided with an advantage in terms of the efficient use of input power over the apparatus of U.S. Pat. No. 4,000,401, in which a portion both ends of each magnetic path is driven to saturation to block flux flow. In the electromagnetic generator **10**, the switching of current flow within the input coils **26, 28** does not need to be sufficient to stop the flow of flux in one of the magnetic paths **18, 20** while promoting the flow of magnetic flux in the other magnetic path. The electromagnetic generator **10** works by changing the flux pattern; it does not need to be completely switched from one side to another.

Experiments have determined that this configuration is superior, in terms of the efficiency of using power within the input coils **26, 28** to generate electrical power within the output coils **29, 30**, to the alternative of arranging input coils and the circuits driving them so that flux from the permanent magnet is driven through the input coils as they are energized. This arrangement of the present invention provides a significant advantage over the prior-art methods shown, for example, in U.S. Pat. No. 4,077,001, in which the magnetic flux is driven through the energized coils.

The configuration of the present invention also has an advantage over the prior-art configurations of U.S. Pat. Nos. 3,368,141 and 4,077,001 in that the magnetic flux is switched between two alternate magnetic paths **18, 20** with only a single input coil **26, 28** surrounding each of the alternate magnetic paths. The configurations of U.S. Pat. Nos. 3,368,141 and 4,077,001 each require two input coils on each of the magnetic paths. This advantage of the present invention is significant both in the simplification of hardware and in increasing the efficiency of power conversion.

The right output coil **29** is electrically connected to a rectifier and filter **33**, having an output driven through a regulator **34**, which provides an output voltage adjustable through the use of a potentiometer **35**. The output of the linear regulator **34** is in turn provided as an input to a sensing and switching circuit **36**. Under start up conditions, the sensing and switching circuit **36** connects the switching and control circuit **24** to an external power source **38**, which is, for example, a starting battery. After the electromagnetic generator **10** is properly started, the sensing and switching circuit **36** senses that the voltage available from regulator **34** has reached a predetermined level, so that the power input to the switching and control circuit **24** is switched from the external power source **38** to the output of regulator **34**. After this switching occurs, the electromagnetic generator **10** continues to operate without an application of external power.

The left output coil **30** is electrically connected to a rectifier and filter **40**, the output of which is connected to a regulator **42**, the output voltage of which is adjusted by means of a potentiometer **43**. The output of the regulator **42** is in turn connected to an external load **44**.

FIG. **2** is a schematic view of a first version of the switching and control circuit **24**. An oscillator **50** drives the clock input of a flip-flop **54**, with the Q and Q' outputs of the flip-flop **54** being connected through driver circuits **56, 58** to power FETS **60, 62** so that the input coils **26, 28** are alternately driven. In accordance with a preferred version of the present invention, the voltage V applied to the coils **26, 28** through the FETS **60, 62** is derived from the output of the sensing and switching circuit **36**.

FIG. **3** is a graphical view of the signals driving the gates of FETS **60, 62** of FIG. **2**, with the voltage of the signal driving the gate of FET **60** being represented by line **64**, and with the voltage of the signal driving FET **62** being represented by line **66**. Both of the coils **26, 28** are driven with positive voltages.

FIG. **4** is a schematic view of a second version of the switching and control circuit **24**. In this version, an oscillator **70** drives the clock input of a flip-flop **72**, with the Q and Q' outputs of the flip-flop **72** being connected to serve as triggers for one-shots **74, 76**. The outputs of the one-shots **74, 76** are in turn connected through driver circuits **78, 80** to drive FETS **82, 84**, so that the input coils **26, 28** are alternately driven with pulses shorter in duration than the Q and Q' outputs of the flip flop **72**.

FIG. **5** is a graphical view of the signals driving the gates of FETS **82, 84** of FIG. **4**, with the voltage of the signal driving the gate of FET **82** being represented by line **86**, and with the voltage of the signal driving the gate of FET **84** being represented by line **88**.

Referring again to FIG. **1**, power is generated in the right output coil **29** only when the level of magnetic flux is changing in the right magnetic path **18**, and in the left output coil **30** only when the level of magnetic flux is changing in the left magnetic path **20**. It is therefore desirable to determine, for a specific magnetic generator configuration, the width of a pulse providing the most rapid practical change in magnetic flux, and then to provide this pulse width either by varying the frequency of the oscillator **50** of the apparatus of FIG. **2**, so that this pulse width is provided with the signals shown in FIG. **3**, or by varying the time constant of the one-shots **74, 76** of FIG. **4**, so that this pulse width is provided by the signals of FIG. **5** at a lower oscillator frequency. In this way, the input coils are not left on longer than necessary. When either of the input coils is left on for a period of time longer than that necessary to produce the change in flux direction, power is being wasted through heating within the input coil without additional generation of power in the corresponding output coil.

A number of experiments have been conducted to determine the adequacy of an electromagnetic generator built as the generator **10** in FIG. **1** to produce power both to drive the switching and control logic, providing power to the input coils **26**, **28**, and to drive an external load **44**. In the configuration used in this experiment, the input coils **26**, **28** had 40 turns of 18-gauge copper wire, and the output coils **29**, **30** had 450 turns of 18-gauge copper wire. The permanent magnet **12** had a height of 40 mm (1.575 in. between its north and south poles, in the direction of arrow **89**, a width of 25.4 mm (1.00 in.), in the direction of arrow **90**, and in the other direction, a depth of 38.1 mm (1.50 in.). The core **16** had a height, in the direction of arrow **89**, of 90 mm (3.542 in.), a width, in the direction of arrow **90**, of 135 mm (5.315 in.) and a depth of 70 mm (2.756 in.). The core **16** had a central hole with a height, in the direction of arrow **89**, of 40 mm (1.575 mm) to accommodate the magnet **12**, and a width, in the direction of arrow **90**, of 85 mm (3.346 in.). The core **16** was fabricated of two "C"-shaped halves, joined at lines **92**, to accommodate the winding of output coils **29**, **30** and input coils **26**, **28** over the core material.

The core material was a laminated iron-based magnetic alloy sold by Honeywell as METGLAS Magnetic Alloy 2605SA1. The magnet material was a combination of iron, neodymium, and boron.

The input coils **26**, **28** were driven at an oscillator frequency of 87.5 KHz, which was determined to produce optimum efficiency using a switching control circuit configured as shown in FIG. **2**. This frequency has a period of 11.45 microseconds. The flip flop **54** is arranged, for example, to be set and reset on rising edges of the clock signal input from the oscillator, so that each pulse driving one of the FETS **60**, **62** has a duration of 11.45 microseconds, and so that sequential pulses are also separated to each FET are also separated by 11.45 microseconds.

FIGS. **6A–6H** are graphical views of signals which simultaneously occurred within the apparatus of FIGS. **1** and **2** during operation with an applied input voltage of 75 volts. FIG. **6A** shows a first drive signal **100** driving FET **60**, which conducts to drive the right input coil **26**. FIG. **6B** is shows a second drive signal **102** driving FET **62**, which conducts to drive the left input coil **28**.

FIGS. **6C** and **6D** show voltage and current signals associated with current driving both the FETS **60**, **62** from a battery source. FIG. **6C** shows the level **104** of voltage V. While the nominal voltage of the battery was 75 volts, a decaying transient signal **106** is superimposed on this voltage each time one of the FETS **60**, **62** is switched on to conduct. The specific pattern of this transient signal depends on the internal resistance of the battery, as well as on a number of characteristics of the magnetic generator **10**. Similarly, FIG. **6D** shows the current **106** flowing into both FETS **60**, **62** from the battery source. Since the signals **104**, **106** show the effects of current flowing into both FETS **60**, **62** the transient spikes are 11.45 microseconds apart.

FIGS. **6E–6H** show voltage and current levels measured at the output coils **29**, **30**. FIG. **6E** shows a voltage output signal **108** of the right output coil **29**, while FIG. **6F** shows a voltage output signal **110** of the left output coil **30**. For example, the output current signal **116** of the right output coil **29** includes a first transient spike **112** caused when the a current pulse in the left input coil **28** is turned on to direct magnetic flux through the right magnetic path **18**, and a second transient spike **114** caused when the left input coil **28** is turned off with the right input coil **26** being turned on. FIG. **6G** shows a current output signal **116** of the right output coil **29**, while FIG. **6H** shows a current output signal **118** of the left output coil **30**.

FIG. **7** is a graphical view of output power measured using the electromagnetic generator **10** and eight levels of input voltage, varying from 10v to 75v. The oscillator frequency was retained at 87.5 KHz. The measurement points are represented by indicia **120**, while the curve **122** is generated by polynomial regression analysis using a least squares fit.

FIG. **8** is a graphical view of a coefficient of performance, defined as the ratio of the output power to the input power, for each of the measurement points shown in FIG. **7**. At each measurement point, the output power was substantially higher than the input power. Real power measurements were computed at each data point using measured voltage and current levels, with the results being averaged over the period of the signal. These measurements agree with RMS power measured using a Textronic THS730 digital oscilloscope.

While the electromagnetic generator **10** was capable of operation at much higher voltages and currents without saturation, the input voltage was limited to 75 volts because of voltage limitations of the switching circuits being used. Those skilled in the relevant art will understand that components for switching circuits capable of handling higher voltages in this application are readily available. The experimentally-measured data was extrapolated to describe operation at an input voltage of 100 volts, with the input current being 140 ma, the input power being 14 watts, and with a resulting output power being 48 watts for each of the two output coils **29**, **30**, at an average output current of 12 ma and an average output voltage of 4000 volts. This means that for each of the output coils **29**, **30**, the coefficient of performance would be 3.44.

While an output voltage of 4000 volts may be needed for some applications, the output voltage can also be varied through a simple change in the configuration of the electromagnetic generator **10**. The output voltage is readily reduced by reducing the number of turns in the output windings. If this number of turns is decreased from 450 to 12, the output voltage is dropped to 106.7, with a resulting increase in output current to 0.5 amps for each output coil **29**, **30**. In this way, the output current and voltage of the electromagnetic generator can be varied by varying the number of turns of the output coils **29**, **30**, without making a substantial change in the output power, which is instead determined by the input current, which determines the amount of magnetic flux shuttled during the switching process.

The coefficients of performance, all of which were significantly greater than 1, plotted in FIG. **8** indicate that the output power levels measured in each of the output coils **29**, **30** were substantially greater than the corresponding input power levels driving both of the input coils **26**, **28**. Therefore, it is apparent that the electromagnetic generator **10** can be built in a self-actuating form, as discussed above in reference to FIG. **1**. In the example of FIG. **1**, except for a brief application of power from the external power source **38**, to start the process of power generation, the power required to drive the input coils **26**, **28** is derived entirely from power developed within the right output coil **29**. If the power generated in a single output coil **29**, **30** is more than sufficient to drive the input coils **26**, **28**, an additional load **126** may be added to be driven with power generated in the output coil **29** used to generate power to drive the input coils **26**, **28**. On the other hand, each of the output coils **29**, **30** may be used to drive a portion of the input coil power requirements, for example with one of the output coils **26**, **28** providing the voltage V for the FET **60** (shown in FIG. **2**), while the other output coil provides this voltage for the FET **62**.

Regarding thermodynamic considerations, it is noted that, when the electromagnetic generator **10** is operating, it is an open system not in thermodynamic equilibrium. The system receives static energy from the magnetic flux of the permanent magnet. Because the electromagnetic generator **10** is self-switched without an additional energy input, the thermodynamic operation of the system is an open dissipative system, receiving, collecting, and dissipating energy from its environment; in this case, from the magnetic flux stored within the permanent magnet. Continued operation of the electromagnetic generator **10** causes demagnetization of the permanent magnet. The use of a magnetic material including rare earth elements, such as a samarium cobalt material or a material including iron, neodymium, and boron is preferable within the present invention, since such a magnetic material has a relatively long life in this application.

Thus, an electromagnetic generator operating in accordance with the present invention should be considered not as a perpetual motion machine, but rather as a system in which flux radiated from a permanent magnet is converted into electricity, which is used both to power the apparatus and to power an external load. This is analogous to a system including a nuclear reactor, in which a number of fuel rods radiate energy which is used to keep the chain reaction going and to heat water for the generation of electricity to drive external loads.

FIG. **9** is a cross-sectional elevation of an electromagnetic generator **130** built in accordance with a second version of the first embodiment of the present invention. This electromagnetic generator **130** is generally similar in construction and operation to the electromagnetic generator **10** built in accordance with the first version of this embodiment, except that the magnetic core **132** of the electromagnetic generator **10** is built in two halves joined along lines **134**, allowing each of the output coils **135** to be wound on a plastic bobbin **136** before the bobbin **136** is placed over the legs **137** of the core **132**. FIG. **9** also shows an alternate placement of an input coil **138**. In the example of FIG. **1**, both input coils **26**, **28** were placed on the upper portion of the magnetic core **16**, with these coils **26**, **28** being configured to establish magnetic fields having north magnetic poles at the inner ends **31**, **32** of the coils **26**, **28**, with these north magnetic poles thus being closest to the end **14** of the permanent magnet **12** having its north magnetic pole. In the example of FIG. **9**, a first input coil **26** is as described above in reference to FIG. **1**, but the second input coil **138** is placed adjacent the south pole **140** of the permanent magnet **12**. This input coil **138** is configured to establish a south magnetic pole at its inner end **142**, so that, when input coil **138** is turned on, flux from the permanent magnet **12** is directed away from the left magnetic path **20** into the right magnetic path **18**.

FIGS. **10** and **11** show an electromagnetic generator **150** built in accordance with a first version of a second embodiment of the present invention, with FIG. **10** being a top view thereof, and with FIG. **11** being a front elevation thereof. This electromagnetic generator **150** includes an output coil **152**, **153** at each corner, and a permanent magnet **154** extending along each side between output coils. The magnetic core **156** includes an upper plate **158**, a lower plate **160**, and a square post **162** extending within each output coil **152**, **153**. Both the upper plate **158** and the lower plate **160** include central apertures **164**.

Each of the permanent magnets **154** is oriented with a like pole, such as a north pole, against the upper plate **158**. Eight input coils **166**, **168** are placed in positions around the upper plate **158** between an output coil **152**, **153** and a permanent magnet **154**. Each input coil **166**, **168** is arranged to form a magnetic pole at its end nearest to the adjacent permanent magnet **154** of a like polarity to the magnetic poles of the magnets **154** adjacent the upper plate **158**. Thus, the input coils **166** are switched on to divert magnetic flux of the permanent magnets **154** from the adjacent output coils **152**, with this flux being diverted into magnetic paths through the output coils **153**. Then, the input coils **168** are switched on to divert magnetic flux of the permanent magnets **154** from the adjacent output coils **153**, with this flux being diverted into magnetic paths through the output coils **152**. Thus, the input coils form a first group of input coils **166** and a second group of input coils **168**, with these first and second groups of input coils being alternately energized in the manner described above in reference to FIG. **1** for the single input coils **26**, **28**. The output coils produce current in a first train of pulses occurring simultaneously within coils **152** and in a second train of pulses occurring simultaneously within coils **153**.

Thus, driving current through input coils **166** causes an increase in flux from the permanent magnets **154** within the posts **162** extending through output coils **153** and a decrease in flux from the permanent magnets **154** within the posts **162** extending through output coils **152**. On the other hand, driving current through input coils **168** causes a decrease in flux from the permanent magnets **154** within the posts **162** extending through output coils **153** and an increase in flux from the permanent magnets **154** within the posts **162** extending through output coils **152**.

While the example of FIGS. **10** and **11** shows all of the input coils **166**,**168** deployed along the upper plate **158**, it is understood that certain of these input coils **166**, **168** could alternately be deployed around the lower plate **160**, in the manner generally shown in FIG. **9**, with one input coil **166**, **168** being within each magnetic circuit between a permanent magnet **154** and an adjacent post **162** extending within an output coil **152**, **153**, and with each input coil **166**, **168** being arranged to produce a magnetic field having a magnetic pole like the closest pole of the adjacent permanent magnet **154**.

FIG. **12** is a top view of a second version **170** of the second embodiment of the present invention, which is similar to the first version thereof, which has been discussed in reference to FIGS. **10** and **11**, except that an upper plate **172** and a similar lower plate (not shown) are annular in shape, while the permanent magnets **174** and posts **176** extending through the output coils **178** are cylindrical. The input coils **180** are oriented and switched as described above in reference to FIGS. **9** and **10**.

While the example of FIG. **12** shows four permanent magnets, four output coils and eight input coils it is understood that the principles described above can be applied to electromagnetic generators having different numbers of elements. For example, such a device can be built to have two permanent magnets, two output coils, and four input coils, or to have six permanent magnets, six output coils, and twelve input coils.

In accordance with the present invention, material used for magnetic cores is preferably a nanocrystalline alloy, and alternately an amorphous alloy. The material is preferably in a laminated form. For example, the core material is a cobalt-niobium-boron alloy or an iron based magnetic alloy.

Also in accordance with the present invention, the permanent magnet material preferably includes a rare earth element. For example, the permanent magnet material is a samarium cobalt material or a combination of iron, neodymium, and boron.

While the invention has been described in its preferred versions and embodiments with some degree of particularity, it is understood that this description has been given only by way of example and that numerous changes in the details of construction, fabrication, and use, including the combination and arrangement of parts, may be made without departing from the spirit and scope of the invention.

What is claimed is:

1. An electromagnetic generator comprising:

a permanent magnet having magnetic poles at opposite ends;

a magnetic core including first and second magnetic paths between said opposite ends of said permanent magnet, wherein

said magnetic core comprises a closed loop,

said permanent magnet extends within said closed loop, and

said opposite ends of said permanent magnet are disposed adjacent opposite sides of said closed loop and against internal surfaces of said magnetic core comprising said closed loop;

a first input coil extending around a portion of said first magnetic path,

a second input coil extending around a portion of said second magnetic path,

a first output coil extending around a portion of said first magnetic path for providing a first electrical output;

a second output coil extending around a portion of said second magnetic path for providing a second electrical output; and

a switching circuit driving electrical current alternately through said first and second input coils, wherein

said electrical current driven through said first input coil causes said first input coil to produce a magnetic field opposing a concentration of magnetic flux from said permanent magnet within said first magnetic path, and

said electrical current driven through said second input coil causes said second input coil to produce a magnetic field opposing a concentration of magnetic flux from said permanent magnet within said second magnetic path.

2. An electromagnetic generator comprising:

a permanent magnet having magnetic poles at opposite ends;

a magnetic core including first and second magnetic paths between said opposite ends of said permanent magnet, wherein

said magnetic core comprises a closed loop,

said permanent magnet extends within said closed loop,

said opposite ends of said permanent magnet are disposed adjacent opposite sides of said closed loop, and

a first type of pole of said permanent magnet is disposed adjacent a first side of said closed loop;

a first input coil, disposed along said first side of said closed loop, extending around a portion of said first magnetic path,

a second input coil, disposed along said first side of said closed loop, extending around a portion of said second magnetic path,

a first output coil extending around a portion of said first magnetic path for providing a first electrical output;

a second output coil extending around a portion of said second magnetic path for providing a second electrical output; and

a switching circuit driving electrical current alternately through said first and second input coils, wherein

said electrical current driven through said first input coil causes said first input coil to produce a magnetic field opposing a concentration of magnetic flux from said permanent magnet within said first magnetic path, and additionally causes said first input coil to produce a magnetic field having said first type of pole at an end of said first input coil adjacent said permanent magnet, and

said electrical current driven through said second input coil causes said second input coil to produce a magnetic field opposing a concentration of magnetic flux from said permanent magnet within said second magnetic path, and additionally causes said second input coil to produce a magnetic field having said first type of pole at an end of said of said second input coil adjacent said permanent magnet.

3. An electromagnetic generator comprising:

a permanent magnet having magnetic poles at opposite ends;

a magnetic core including first and second magnetic paths between said opposite ends of said permanent magnet, wherein

said magnetic core comprises a closed loop,

said permanent magnet extends within said closed loop, and

said opposite ends of said permanent magnet are disposed adjacent opposite sides of said closed loop,

a first type of pole of said permanent magnet is disposed adjacent a first side of said closed loop, and

a second type of pole, opposite said first type of pole, of said permanent magnet is disposed adjacent a second side of said closed loop;

a first input coil extending around a portion of said first magnetic path, wherein said first input coil is disposed along said first side of said closed loop;

a second input coil extending around a portion of said second magnetic path wherein said second input coil is disposed along said second side of said closed loop;

a first output coil extending around a portion of said first magnetic path for providing a first electrical output;

a second output coil extending around a portion of said second magnetic path for providing a second electrical output; and

a switching circuit driving electrical current alternately through said first and second input coils, wherein

said electrical current driven through said first input coil causes said first input coil to produce a magnetic field opposing a concentration of magnetic flux from said permanent magnet within said first magnetic path, and additionally causes said first input coil to produce a magnetic field having said first type of pole at an end of said first input coil adjacent said permanent magnet, and

said electrical current driven through said second input coil causes said second input coil to produce a magnetic field opposing a concentration of magnetic flux from said permanent magnet within said second magnetic path, and additionally causes said second input coil to produce a magnetic field having said second type of pole at an end of said of said second input coil adjacent said permanent magnet.

4. An electromagnetic generator comprising:

a permanent magnet having magnetic poles at opposite ends;

a magnetic core including first and second magnetic paths between said opposite ends of said permanent magnet;

a first input coil extending around a portion of said first magnetic path,

a second input coil extending around a portion of said second magnetic path,

a first output coil extending around a portion of said first magnetic path for providing a first electrical output;

a second output coil extending around a portion of said second magnetic path for providing a second electrical output; and

a switching circuit driving electrical current alternately through said first and second input coils, wherein said electrical current driven through said first input coil causes said first input coil to produce a magnetic field opposing a concentration of magnetic flux from said permanent magnet within said first magnetic path, and wherein said electrical current driven through said second input coil causes said second input coil to produce a magnetic field opposing a concentration of magnetic flux from said permanent magnet within said second magnetic path, wherein a portion of electrical power induced in said first output coil provides power to drive said switching circuit.

5. The electromagnetic generator of claim **4**, wherein said switching circuit is driven by an external power source during a starting process and by power induced in said first output coil during operation after said starting process.

6. The electromagnetic generator of claim **2**, wherein said magnetic core is composed of a nanocrystalline magnetic alloy.

7. The electromagnetic generator of claim **6**, wherein said nanocrystalline magnetic alloy is a cobalt-niobium-boron alloy.

8. The electromagnetic generator of claim **6**, wherein said nanocrystalline magnetic alloy is an iron-based alloy.

9. The electromagnetic generator of claim **2**, wherein said changes in flux density within said magnetic core occur without driving said magnetic core to magnetic saturation.

10. The electromagnetic generator of claim **2**, wherein

said switching circuit drives said electrical current through said first input coil in response to a first train of pulses,

said switching circuit drives said electrical current through said second input coil in response to a second train of pulses, alternating with pulses within said first train of pulses, and

said pulses in said first and second trains of pulses are approximately 11.5 milliseconds in duration.

11. The electromagnetic generator of claim **2**, wherein said permanent magnet is composed of a material including a rare earth element.

12. The electromagnetic generator of claim **11**, wherein said permanent magnet is composed essentially of samarium cobalt.

13. The electromagnetic generator of claim **11**, wherein said permanent magnet is composed essentially of iron, neodymium, and boron.

14. An electromagnetic generator comprising:

a magnetic core including a pair of spaced-apart plates, wherein each of said spaced-apart plates includes a central aperture, and first and second pluralities of posts extending between said spaced-apart plates;

a plurality of permanent magnets extending individually between said pair of spaced-apart plates and between adjacent posts within said plurality of posts, wherein each permanent magnet within said plurality of permanent magnets has magnetic poles at opposite ends, wherein all magnets within said plurality of magnets are oriented to produce magnetic fields having a common direction;

first and second pluralities of input coils, wherein each input coil within said first and second pluralities of input coils extends around a portion of a plate within said spaced-apart plates between a post in said plurality of posts and a permanent magnet in said plurality of permanent magnets;

an output coil extending around each post in said first and second pluralities of posts for providing an electrical output;

a switching circuit driving electrical current alternatively through said first and second pluralities of input coils, wherein said electrical current driven through each input coil in said first plurality of input coils causes an increase in magnetic flux within each post within said first plurality of posts from permanent magnets on each side of said post and a decrease in magnetic flux within each post within said second plurality of posts from permanent magnets on each side of said post, and wherein said electrical current driven through input coil in said second plurality of input coils causes a decrease in magnetic flux within each post within said first plurality of posts from permanent magnets on each side of said post and an increase in magnetic flux within each post within said second plurality of posts from permanent magnets on each side of said post.

15. The electromagnetic generator of claim **14**, wherein

each input coil extends around a portion of a magnetic path through said magnetic core between said opposite ends a permanent magnet adjacent said input coil,

said magnetic path extends through a post within said magnetic core adjacent said input coil, and

driving electrical current through said input coil causes said input coil to produce a magnetic field opposing a concentration of magnetic flux within said magnetic path.

16. The electromagnetic generator of claim **14**, wherein said switching circuit is driven by an external power source during a starting process and by power induced in said output coils during operation after said starting process.

17. The electromagnetic generator of claim **14**, wherein said magnetic core is composed of a nanocrystalline magnetic alloy.

18. The electromagnetic generator of claim **2**, wherein a portion of electrical power induced in said first output coil provides power to drive said switching circuit.

19. The electromagnetic generator of claim **18**, wherein said switching circuit is driven by an external power source during a starting process and by power induced in said first output coil during operation after said starting process.

20. The electromagnetic generator of claim **3**, wherein a portion of electrical power induced in said first output coil provides power to drive said switching circuit.

21. The electromagnetic generator of claim **20**, wherein said switching circuit is driven by an external power source during a starting process and by power induced in said first output coil during operation after said starting process.

22. The electromagnetic generator of claim **3**, wherein said magnetic core is composed of a nanocrystalline magnetic alloy.

23. The electromagnetic generator of claim **22**, wherein said nanocrystalline magnetic alloy is a cobalt-niobium-boron alloy.

24. The electromagnetic generator of claim **22**, wherein said nanocrystalline magnetic alloy is an iron-based alloy.

25. The electromagnetic generator of claim **3**, wherein said changes in flux density within said magnetic core occur without driving said magnetic core to magnetic saturation.

26. The electromagnetic generator of claim **3**, wherein

said switching circuit drives said electrical current through said first input coil in response to a first train of pulses,

said switching circuit drives said electrical current through said second input coil in response to a second train of pulses, alternating with pulses within said first train of pulses, and

said pulses in said first and second trains of pulses are approximately 11.5 milliseconds in duration.

27. The electromagnetic generator of claim **3**, wherein said permanent magnet is composed of a material including a rare earth element.

28. The electromagnetic generator of claim **27**, wherein said permanent magnet is composed essentially of samarium cobalt.

29. The electromagnetic generator of claim **27**, wherein said permanent magnet is composed essentially of iron, neodymium, and boron.

* * * * *

US 20100201133A1

(19) **United States**

(12) **Patent Application Publication**

MESLER

(10) **Pub. No.: US 2010/0201133 A1**

(43) **Pub. Date: Aug. 12, 2010**

(54) **ZERO POINT ENERGY ROTATOR TRANSDUCER AND ASSOCIATED METHODS**

(76) Inventor: **PAUL R. MESLER**, EMERY, UT (US)

Correspondence Address:
THORPE NORTH & WESTERN, LLP.
P.O. Box 1219
SANDY, UT 84091-1219 (US)

(21) Appl. No.: **12/541,007**

(22) Filed: **Aug. 13, 2009**

Related U.S. Application Data

(60) Provisional application No. 61/088,913, filed on Aug. 14, 2008.

Publication Classification

(51) **Int. Cl.**
H02K 7/18 (2006.01)
F03G 7/10 (2006.01)

(52) **U.S. Cl.** .... **290/1 C**; 290/1 R; 74/DIG.009

(57) **ABSTRACT**

An energy conversion system for extracting energy from the zero-point field of space is disclosed and described. This system includes a primary transducer operatively associated with a secondary transducer. The primary transducer can include a substrate having an arm rotatably coupled to the substrate, while the substrate can be confined to move along a linear oscillatory path. The secondary transducer converts kinetic energy of the substrate and arm to usable energy. The multi-mode motion of the substrate and arm combination is sufficient to impart a momentum influx to rotational motion of the arm.

The momentum influx can then be converted to usable energy such as mechanical, electrical and/or thermal energy. As described in more detail herein, the momentum influx appears to be a result of radiation pressure due to scattering in part of the electro-magnetic momentum-energy contained within the zero-point field of space opposite to centripetal acceleration of the arm.

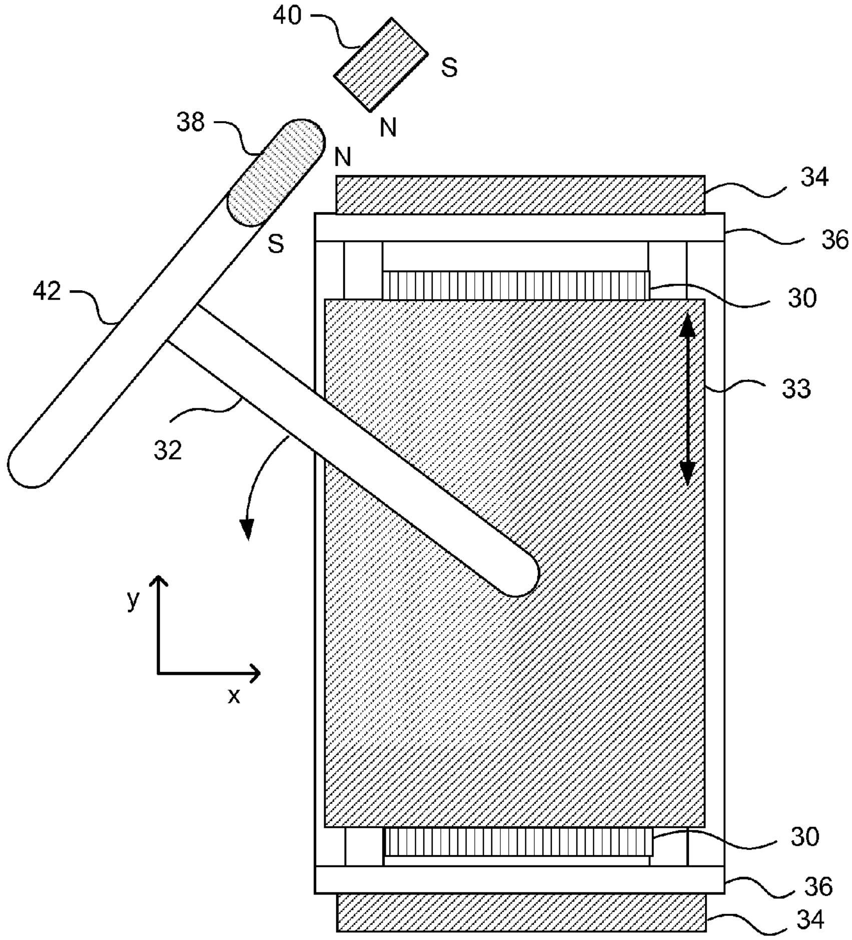

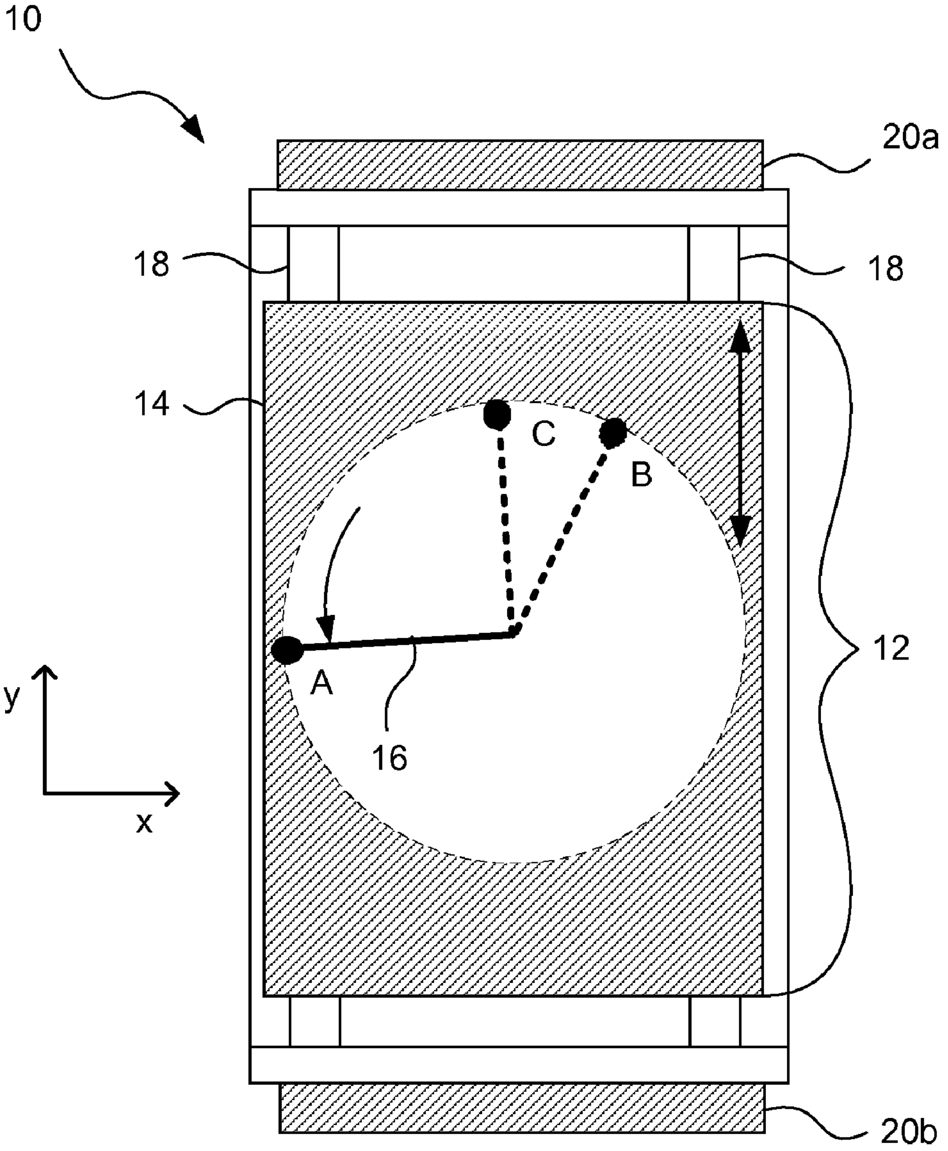
10
20a
18
18
14
C
B
A
16
12
y
x
20b

FIG. 1

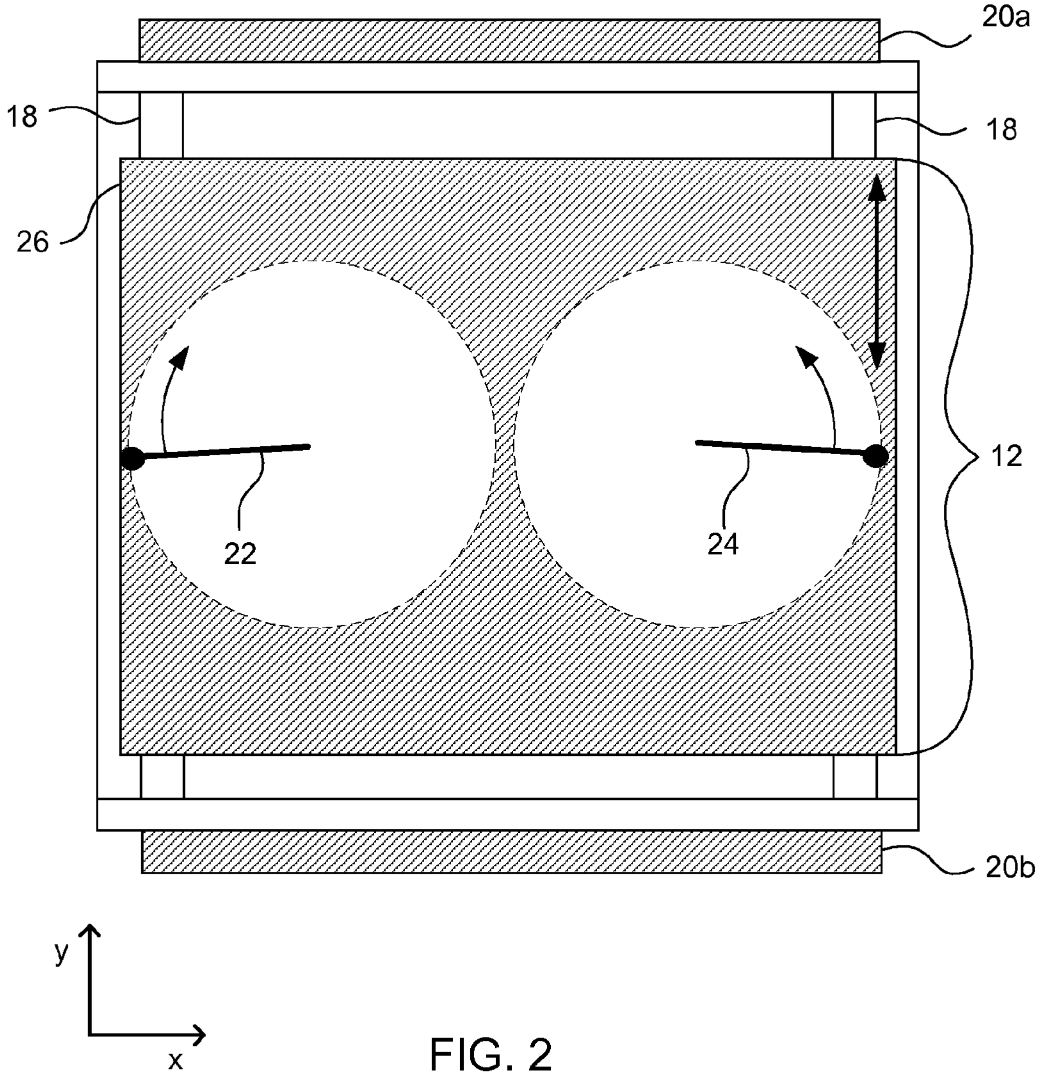
20a
18
18
26
12
22
24
20b
y
x

FIG. 2

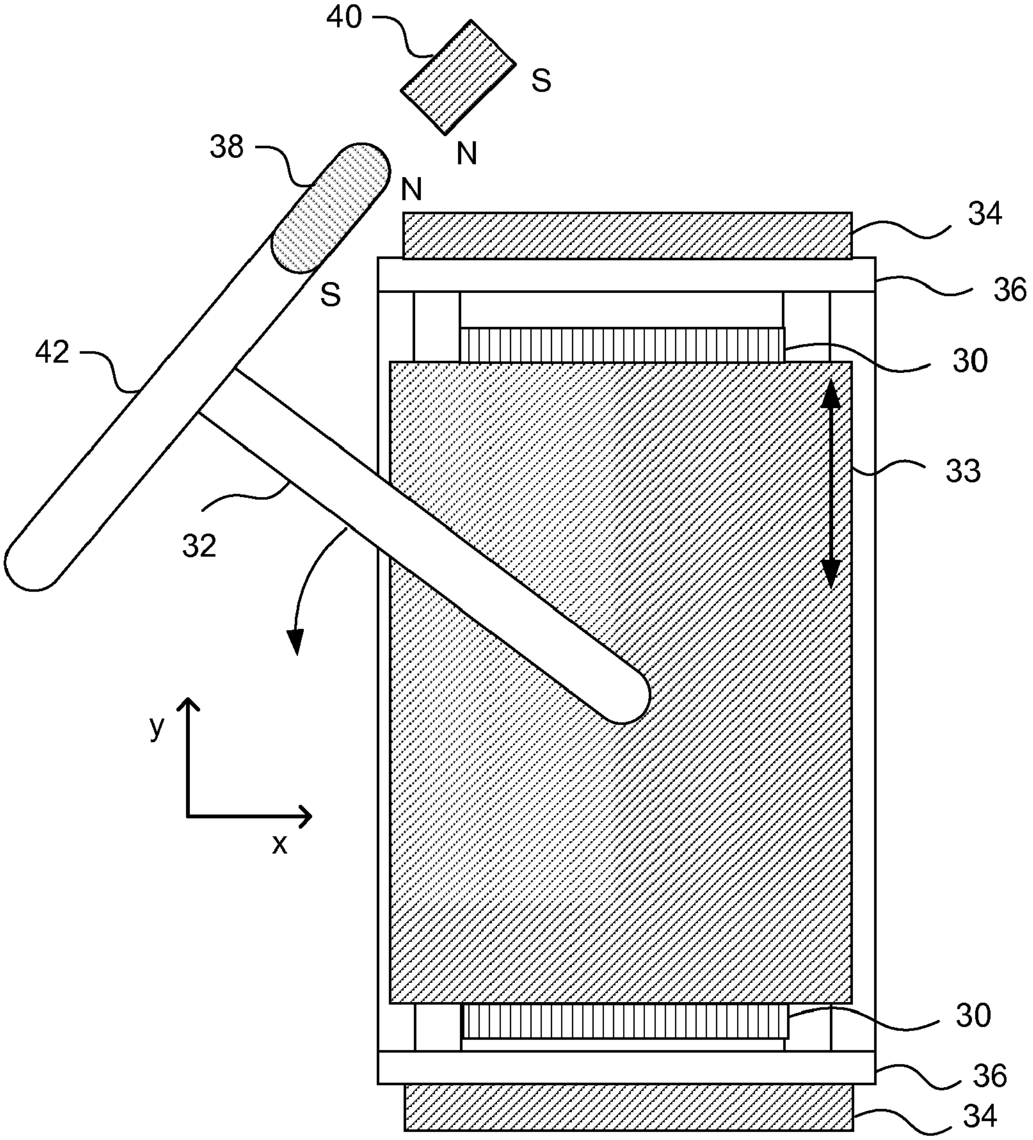
40
S
N
38
N
S
42
32
34
36
30
33
y
x
30
36
34

FIG. 3

ZERO POINT ENERGY ROTATOR TRANSDUCER AND ASSOCIATED METHODS

RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 61/088,913, filed Aug. 14, 2008 which is incorporated herein by reference.

FIELD OF THE INVENTION

[0002] This invention relates to generation of energy from the zero-point energy field of space or electromagnetic quantum vacuum. More specifically, the invention relates to a device useful for extraction of useful work via a sliding rotator. Therefore, the present invention relates generally to the fields of physics, quantum physics, and electromagnetism.

BACKGROUND OF THE INVENTION

[0003] The now ubiquitous drive for energy independence has provided a new set of incentives and appreciation for development of environmentally and economically responsible energy resources. Conventional discussions regarding such energy resources often revolve around fossil fuels, nuclear, and “alternative energy sources” such as solar, wind, bio fuels, hydrogen, fuel cells, and the like. Each of these energy resources currently comes with various benefits and drawbacks and often entail trade-offs between perceived pollution and environmental problems, costs, efficiencies, and unintended consequences (e.g. corn demand for use as fuel versus food).

[0004] Since at least the 1960’s some theoretical physicists have proposed the concept of zero-point energy which suggests a background energy field throughout space. Historically, these physicists have been at least partially marginalized by mainstream academic thought. However, over the past decade or so, such theories have begun to gain significant traction in terms of recognition and available proofs to support these theories. Although this developing area of physics has experienced progress, practical applications of such theories remain severely limited. One example of such progress in practical applications is the work of Bernard Haisch and others in capitalizing on the Casimir effect. Such efforts thus far remain relegated to the quantum level. Although a wide variety of purported devices and claims have been made, no macroscale efforts have yet proven to yield useful energy gains.

SUMMARY OF THE INVENTION

[0005] In light of the relative infancy of the field and deficiencies noted above, the present invention provides an energy conversion system for extracting energy from the zero-point field of space. This system includes a primary transducer operatively associated with a secondary transducer. The primary transducer can include a substrate having an arm rotatably coupled to the substrate, while the substrate can be confined to move along a linear oscillatory path. The secondary transducer converts kinetic energy of the substrate and arm to usable energy. The multi-mode motion of the substrate and arm combination is sufficient to impart a momentum influx to a rotational motion of the arm. The momentum influx can then be converted to usable energy such as mechanical, electrical and/or thermal energy. As described in more detail herein, the momentum influx appears to be a result of radiation pressure resulting from scattering in part of the electro-magnetic momentum-energy contained within the zero-point field of space opposite to centripetal acceleration of the arm.

[0006] There has thus been outlined, rather broadly, the more important features of the invention so that the detailed description thereof that follows may be better understood, and so that the present contribution to the art may be better appreciated. Other features of the present invention will become clearer from the following detailed description of the invention, taken with the accompanying drawings and claims, or may be learned by the practice of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] The present invention will become more fully apparent from the following description and appended claims, taken in conjunction with the accompanying drawings. Understanding that these drawings merely depict exemplary embodiments of the present invention and they are, therefore, not to be considered limiting of its scope. It will be readily appreciated that the components of the present invention, as generally described and illustrated in the figures herein, could be arranged, sized, and designed in a wide variety of different configurations. In particular, relative dimensions may vary and are generally not presented in an optimized configuration but are rather provided for clarity and convenience in describing the invention generally. Nonetheless, the invention will be described and explained with additional specificity and detail through the use of the accompanying drawings in which:

[0008] FIG. **1** is a schematic of an energy conversion system in accordance with one embodiment of the present invention.

[0009] FIG. **2** is a schematic of an energy conversion system having two rotating arms, one rotating clockwise and the other counterclockwise, in accordance with another embodiment of the present invention.

[0010] FIG. **3** is a schematic of an energy conversion system illustrating an electro-magnetic mechanism for accelerating the arm in accordance with yet another embodiment of the present invention.

DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS

[0011] The following detailed description of exemplary embodiments of the invention makes reference to the accompanying drawings, which form a part hereof and in which are shown, by way of illustration, exemplary embodiments in which the invention may be practiced. While these exemplary embodiments are described in sufficient detail to enable those skilled in the art to practice the invention, it should be understood that other embodiments may be realized and that various changes to the invention may be made without departing from the spirit and scope of the present invention. Thus, the following more detailed description of the embodiments of the present invention is not intended to limit the scope of the invention, as claimed, but is presented for purposes of illustration only and not limitation to describe the features and characteristics of the present invention, to set forth the best mode of operation of the invention, and to sufficiently enable one skilled in the art to practice the invention. Accordingly, the scope of the present invention is to be defined solely by the appended claims.

[0012] The following detailed description and exemplary embodiments of the invention will be best understood by reference to the accompanying drawings, wherein the elements and features of the invention are designated by numerals throughout.

Definitions

[0013] In describing and claiming the present invention, the following terminology will be used.

[0014] The singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "an arm" includes reference to one or more of such members and reference to "converting" refers to one or more such steps.

[0015] As used herein with respect to an identified property or circumstance, "substantially" refers to a degree of deviation that is sufficiently small so as to not measurably detract from the identified property or circumstance. The exact degree of deviation allowable may in some cases depend on the specific context.

[0016] As used herein, "adjacent" refers to the proximity of two structures or elements. Particularly, elements that are identified as being "adjacent" may be either abutting or connected. Such elements may also be near or close to each other without necessarily contacting each other. The exact degree of proximity may in some cases depend on the specific context.

[0017] As used herein, a plurality of items, structural elements, compositional elements, and/or materials may be presented in a common list for convenience. However, these lists should be construed as though each member of the list is individually identified as a separate and unique member. Thus, no individual member of such list should be construed as a de facto equivalent of any other member of the same list solely based on their presentation in a common group without indications to the contrary.

[0018] Concentrations, amounts, and other numerical data may be presented herein in a range format. It is to be understood that such range format is used merely for convenience and brevity and should be interpreted flexibly to include not only the numerical values explicitly recited as the limits of the range, but also to include all the individual numerical values or sub-ranges encompassed within that range as if each numerical value and sub-range is explicitly recited. For example, a numerical range of about 1 to about 4.5 should be interpreted to include not only the explicitly recited limits of 1 to about 4.5, but also to include individual numerals such as 2, 3, 4, and sub-ranges such as 1 to 3, 2 to 4, etc. The same principle applies to ranges reciting only one numerical value, such as "less than about 4.5," which should be interpreted to include all of the above-recited values and ranges. Further, such an interpretation should apply regardless of the breadth of the range or the characteristic being described.

[0019] In the present disclosure, the term "preferably" or "preferred" is non-exclusive where it is intended to mean "preferably, but not limited to." Any steps recited in any method or process claims may be executed in any order and are not limited to the order presented in the claims. Means-plus-function or step-plus-function limitations will only be employed where for a specific claim limitation all of the following conditions are present in that limitation: a) "means for" or "step for" is expressly recited; and b) a corresponding function is expressly recited. The structure, material or acts that support the means-plus function are expressly recited in the description herein. Accordingly, the scope of the invention should be determined solely by the appended claims and their legal equivalents, rather than by the descriptions and examples given herein.

Embodiments of the Invention

[0020] Referring now to FIG. **1**, an energy conversion system **10** is shown generally for extracting energy from the zero-point field of space. The system can include a primary transducer **12** which is designed to extract energy from the zero-point field of space in the form of kinetic energy. Multi-mode motion can be induced in a primary transducer along an oscillatory path. The multi-mode motion can include a primary translational motion and a secondary motion such that abrupt changes in the primary translational motion are sufficient to impart a momentum influx to the secondary motion. In one embodiment, the primary transducer can include a substrate **14** having an arm **16** rotatably coupled to the substrate. The substrate is generally confined to move along an oscillatory path. A secondary transducer **20***a* can be operatively associated with the primary transducer to covert the momentum influx into usable energy. Thus, the systems of the present invention provide an energy conversion system for extracting energy from the zero-point field of space. Generally, the system includes a means for extracting momentum-energy from vacuum by the action of inertia and a means for converting the momentum-energy to usable energy operatively associated with the means for extracting momentum-energy.

[0021] More particularly, the substrate **14** can be any suitable member upon which a rotating arm **16** can be coupled or otherwise mounted. The substrate can be guided along the oscillatory path by any suitable mechanism such as, but not limited to, tracks, walls or the like. FIG. **1** shows a linear oscillatory path defined by a pair of tracks **18** along which the substrate can move. The substrate can move along the tracks via bearings, wheels, magnetically repulsive members, or any other suitable approach. Typically, it is most desirable to minimize friction between the substrate and tracks in order to maximize energy transfer to the secondary transducer. The substrate and arm can be formed of any suitable material which can be readily chosen based on designed parameters to meet various property requirements such as density, fracture strength, toughness, etc. Furthermore, the shape and contours of the substrate and arm can also be designed to minimize aerodynamic resistance due to the significant acceleration forces per unit of time experienced by the primary transducer.

[0022] In one aspect of the present invention, the substrate has a substrate mass (m_2) and the arm has an arm mass (m_1) such that a ratio of m_2 to m_1 is chosen to maximize a gain in kinetic energy (ΔKE) of the substrate. The rotating arm acts as the primary momentum influx extraction mechanism which experiences an inertial force acting opposite to the centripetal force acting on the rotating arm. A fraction of the inertial force is manifest at an axis of rotation which is rigidly attached to the substrate and which constrains the arm to rotate around this axis. The y-component of this force acting on the axis, denoted as the centrifugal reactive force, imparts an impulse on the substrate in the positive y-direction with respect to an x-y coordinate system when the arm rotates within the first quadrant of the said x-y coordinate system or imparts an impulse on the substrate in the negative y-direction when the arm rotates within the third quadrant of the said x-y coordinate system, such that the substrate experiences a net timerate influx of momentum from the zero-point energy field of space or an increase in the momentum of the substrate in either the positive or negative y-direction, depending on which quadrant the arm is rotating in.

[0023] The oscillatory path can be a linear track **18** which defined on both ends by a pair of resilient stationary members **20***a* and **20***b* oriented on either end of the linear track. During operation of the system, the primary transducer oscillates back and forth between the two stationary members. The approach at each end is accompanied by a transfer of kinetic energy of the primary transducer (e.g. substrate and arm) to the secondary transducer. The primary transducer can thus be coupled to the secondary transducer in a manner which allows for conversion of this kinetic energy to the secondary transducer. For example, this can be accomplished by providing an at least partially or substantially inelastic collision with the resilient stationary members. As discussed in more detail below, various mechanisms can be effectively used to convert kinetic energy of the primary transducer into usable energy, e.g. via mechanical, thermal, and/or electromagnetic features.

[0024] As can be readily appreciated, the arm and substrate need an initial input of energy from an external source to begin movement. The source of energy to initiate rotation of the arm can come from a conventional energy source, such as a battery. This external energy can be transferred to the arm such as by a coaxial motor, magnetic accelerator, external flywheel, or the like. For example, a drive motor can be operatively associated with the arm and configured to initiate rotational motion of the arm and sustain the rotational motion at a predetermined critical angular velocity. Another option is to provide a permanent magnet on the arm, e.g. at the free end, and a complimentary magnetic induction coil at a location external to the primary transducer. In this manner, an electric current applied to the induction coil creates a magnetic field sufficient to act on the permanent magnet and accelerate the arm. After a critical angular velocity is reached, the external source of energy (such as a battery, the power grid, etc.) for the motor or other suitable initiator mechanism can be switched to an alternative energy storage device such as collected energy.

[0025] Referring now to FIG. **2**, an alternative embodiment of the present invention is shown including an arm **22** and a secondary arm **24** each rotatably coupled to the substrate. In one embodiment, the secondary arm has an axis of rotation spaced apart from an axis of rotation of the arm **22**. Alternatively, the secondary arm could share a common axis of rotation. In either case, the secondary arm is coupled to the arm such that both the arm and the secondary arm rotate at a common angular velocity and opposite rotational directions. Typically, the secondary arm also has a substantially identical mass and dimensions to that of the arm. In other words, the arms swing forward and back in the same cycle, with common y-coordinates and mirror x-coordinates.

[0026] In one embodiment, a motor can drive the first arm, and the second arm can be driven by gears linked to the first arm, such that they are in sync with each other, rotating at the same angular speed but in opposite rotational directions. In such cases, the cosines of the rotation angle of each arm will always be equal in magnitude, but opposite in sign. Thus, the x-components of the centrifugal reactive force acting on each axis would be equal and opposite in direction, and therefore, they would cancel each other. As a result, the substrate having two rotating arms will not impart compressive forces on the corresponding linear track along the x-axis during operation of the system.

[0027] The kinetic energy of the primary transducer can be converted to usable energy twice for every full rotation of the arm(s), i.e. once between the 0° position and the 90° position and once between the 180° and the 270° position. The optimal position of the arm at impact of the substrate with the stationary member can vary but is generally in these ranges. Specifically, as the arm approaches 90° or 270° the gain approaches 0 (see Example, Equation 3). Similarly, the momentum of the substrate at 0° and 180° is also 0 so the gain at those points would also be 0. A suitable secondary transducer can thus be coupled with the primary transducer as a means for converting kinetic and momentum-energy of the substrate and arm into usable energy such as electrical, mechanical, or thermal energy. The secondary transducer can be an electromagnetic induction system, mechanical gear system, heat transfer system, piezoelectric system, or the like. For example, the means for converting the momentum-energy can be an induction coil system, heat dissipater, magnetic brake, friction brake, or linear to rotational motion translator. A heat dissipater can include a substantially inelastic collision between the substrate and a set of bumpers. Heat generated by collision can be converted to energy through heat transfer to produce steam which can be used to drive a turbine, or other similar thermo-electric conversion devices.

[0028] In one specific embodiment of the present invention, the secondary transducer includes at least one induction coil system including a magnet and a complimentary conductive coil. The magnet can be physically coupled to the substrate and oriented to approach the complimentary conductive coil sufficient to induce electrical current therein during movement of the substrate along the oscillatory path. A similar magnet system can be incorporated into the arm as well. Such a secondary transducer can also be utilized as the initiator system by switching the energy source, i.e. from an external battery to energy derived from the zero-point energy field. FIG. **3** illustrates a simplified version of such an electromagnetic-kinetic energy conversion system for the secondary transducer. In particular, a permanent magnet **30** can be coupled to each collision end of the substrate **33** having the arm **32** rotatably coupled thereto. An external induction coil **34** can be oriented at a position (or positions) to allow the kinetic energy of the substrate to be progressively converted to electrical energy at each approach of the substrate to the stationary members **36**. In another embodiment, a piezoelectric material such as a crystal or ceramic can be coupled to the stationary member. The force applied by the substrate to the piezoelectric material can generate an electric potential in response to the applied mechanical stress. Alternatively, the secondary transducer can include a mechanical coupling between the substrate and an energy translator. The energy translator can be a linear-to-rotational motion drive, heat surfacing, or rack and pinion, although other mechanisms may also be suitable.

[0029] After the substrate collides with either of the stationary members, the arm's rotational speed can be replenished by a suitable arm accelerator. For example, an electro-magnet system can be oriented with a permanent magnet **38** at the free end of the arm with a corresponding external induction coil **40**. In this case, the arm can have a t-shape to allow the permanent magnet to be oriented substantially in-line with the magnetic field produced by the external electro-magnet (although opposite signs). This allows for a maximum transfer of kinetic energy to the arm from the electro-magnet acceleration system. Alternatively, the permanent magnet and induction coil can be swapped. Typically, the t-portion **42** can have substantially equal mass on either end such that the center of mass lies along the arm axis. This electro-magnet can impart an impulsive torque to the arm in either the second or fourth quadrant (or at any point right after the collision of the substrate with a stationary member) as the permanent magnet attached to the arm passes closely to the electro-magnet. The advantage of this approach is that the arm's rotation will be maintained substantially by electro-magnetism without additional gears or mechanical losses due to the friction of such gears. This replenishment of the rotational speed of the arm can optionally be achieved by a motor in the second and fourth quadrant (or at any point right after the collision of the substrate with a stationary member) by using pulsed technology, wherein the power to the motor is only turned on in the second or fourth quadrant.

[0030] In operation, the movement of the substrate can follow a linear oscillatory path having two endpoints while the arm(s) follows a cyclic path such that translational motion of the substrate in a y-direction along the oscillatory path substantially coincides with translational motion of the arm in the y-direction. Furthermore, the arm has two theoretical maximum energy transfer positions, the positions occurring twice for every cycle of rotation of the arm. Actual maximum energy transfer positions can be empirically determined and a collision swing angle adjusted to obtain a substantially maximum in the usable energy, depending on the specific design, i.e. relative masses, arm length, rotational velocity, etc.

[0031] There are at least two factors that impact the overall maximum energy gain of the system. For example, the substrate's final kinetic energy before collision with a stationary member partly determines the maximum energy gain of the system. Maximum energy gain is also determined by the angular displacement of the arm at the time of the collision, the angular velocity of the arm, the linear velocity of the substrate prior to collision and the relative masses of the arm and substrate.

[0032] During operation of the system, a full cycle of rotation of the rotator can be divided into the four conventional quadrants counter-clockwise. Upon start-up, the arm is initiated with a battery driven motor (or equivalent, such as an electro-magnet system) to reach a terminal velocity while holding the slider substrate stationary. The terminal velocity is determined by the friction in the system and the power transfer capabilities of the device that delivers power to the rotator to keep it rotating. After the rotator has reached this constant velocity, integrated sensors in the system release the slider just prior to the rotator arriving at its 0 degree position. The slider will begin to accelerate when the y-component of the centrifugal reactive force acting on the physical vertical axis attached to the slider is greater than any static friction force between the slider and its linear track. The angle of the rotator with respect to the x-axis at which angle the slider begins to accelerate is denoted as phi. During this first quadrant rotation of the rotator, the power to the motor can be optionally turned off so that the rotator continues to rotate on its own rotational inertia.

[0033] The distance between the front and back secondary transducers or stationary members can be adjusted such that the slider "collides" with this secondary front transducer when the rotator has rotated an angle less than 90 degrees (e.g. 15° to 75° although other ranges may be useful such as about)45° with respect to the x-axis which gives a maximum transfer of energy.

[0034] In one embodiment, the secondary transducer can comprise a linear generator. The linear generator can include a coil or set of coils receiving the magnets attached to the slider, and all of the electronic and electrical circuitry associated with the linear generator must be able to handle a brief impulse of power if the slider is to come to a near stop quickly. The linear generator can be designed of sufficiently robust materials and capacity to handle this power surge. Alternatively, the secondary transducer can include a thermal conversion system or a piezoelectric device, as previously discussed. More particularly, the stationary members and substrate can be designed to provide a substantially inelastic collision such that much of the kinetic energy of the substrate is converted to heat or electricity. In a thermal conversion process, the substrate and stationary member can accordingly be thermally linked to a water reservoir such that water is heated to steam, which drives an electrical generator.

[0035] A thermal conversion process may be more effective for a large scale energy conversion system. A piezoelectric conversion process may be more useful for a small scale energy conversion system, such as an energy conversion system used in a micro-machine or other type of micro-electro-mechanical system (MEMS).

[0036] After the slider comes to a complete stop after the collision of the slider with a stationary member a pulsed power source can be used to replenish the speed of the rotator back to its initial speed. There will typically be a loss in speed of the rotator due to friction and a replenishment of the speed of the rotator must happen twice every cycle after every collision of the slider with a stationary member.

[0037] Based on the guidance provided herein, a speed of the rotational motion, relative masses of the arm and the substrate, geometry of the arm, and the geometry of the substrate can all be designed such that zero-point energy is utilized to maintain the movement of the substrate. Non-limiting examples of specific variables which can be adjusted include aerodynamic shape of the arm and substrate, length of the arm, materials of construction, density of substrate and arm, and the like.

[0038] Theoretical Background

[0039] Without being bound to any particular theory, it is thought that the principles of operation of the present invention are governed by a theoretical backdrop described in more detail below. In summary, this theoretical backdrop indicates that the momentum influx to the primary transducer is a result of radiation pressure due to scattering in part of the electromagnetic momentum-energy contained within the zero-point field of space opposite to the centripetal acceleration of the arm. The systems of the present invention can be sized and incorporated into a wide variety of applications which are currently occupied by conventional energy sources, e.g. batteries, power generators, fuel cells, and so forth. A more in depth description of one theory follows; however, it is understood that the following description is not exclusive, exhaustive or definitive of the underlying inventive concept.

[0040] We begin a discussion with a rotating point particle with a mass of m_1, and denote it as the rotator. One end of the rotator is attached to a vertical axis by a mass-less wire of length r, so that the rotator rotates freely counter-clockwise around this axis with two degrees of freedom in a plane parallel to the earth. The vertical axis is rigidly attached to a second object denoted as the slider with mass m_2. The slider is constrained to move along a linear track with one degree of freedom, coincident to the y-axis of an x-y Cartesian coordinate system. Initially, the axis of rotation is set at the origin of the coordinate system.

[0041] The rotator is aligned coincident with the x-axis of the x-y coordinate system, such that the rotator is located at 180 degrees with respect to the x-axis. The slider presses against a back bumper so that the slider cannot move in the negative y-direction. An impulse is applied on the rotator in the negative y-direction, causing the rotator to rotate counterclockwise. The rotator rotates until it passes the 0 degrees point, and the slider moves in the positive y-direction at some angle after this 0 degrees point. Immediately prior to the movement of the slider, the initial tangential velocity of the rotator is denoted as v_i. A short time interval later as the slider moves in the positive y-direction, the slider is allowed to make an inelastic collision with a front bumper at some angle θ of the rotator with respect to the x-axis. The tangential velocity of the rotator prior to the collision of the slider is denoted as v_α. The post-collision tangential velocity of the rotator is denoted as v_f. The linear velocity of the slider prior to the collision of the slider is denoted as v_s.

[0042] The initial tangential kinetic energy of the rotator is:

$$KE_{rotator\ initial}\tfrac{1}{2}m_1v_i^2 \qquad (1).$$

[0043] With respect to the laboratory frame this represents the initial energy of the rotator-slider system. As the slider accelerates forward in the positive y-direction under the action of the y-component of the centrifugal reactive force on the axis, both the slider and the rotator acquire an equal instantaneous linear velocity in the positive y-direction, and therefore, a linear kinetic energy with respect to the laboratory frame. By the conservation of energy, the tangential kinetic energy of the rotator must decrease, such that:

$$\Delta KE_{rotator}=KE_{linear\ slider}+KE_{linear\ rotator} \qquad (2).$$

Or:

$$\tfrac{1}{2}m_1v_i^2-\tfrac{1}{2}m_1v_\alpha^2=\tfrac{1}{2}m_2v_s^2+\tfrac{1}{2}m_1v_s^2 \qquad (2a).$$

[0044] Rearranging, we have:

$$\tfrac{1}{2}m_1v_i^2-\tfrac{1}{2}m_1v_\alpha^2=\tfrac{1}{2}m_2v_s^2+\tfrac{1}{2}m_1v_s^2 \qquad (3).$$

[0045] That is, by the conservation of energy the pre-collision total kinetic energy observed in the system must equal the initial kinetic energy of the rotator.

[0046] Upon collision of the slider with the front bumper, the kinetic energy of the slider equal to $\tfrac{1}{2}m_2v_s^2$ is essentially converted to thermal energy. The linear kinetic energy of the rotator is accounted for as follows.

[0047] Taking θ as the angle of the rotator at the time of collision, and by application of the Pythagorean Theorem, we have:

$$\tfrac{1}{2}m_1v_s^2=\tfrac{1}{2}m_1\sin^2\theta v_s^2+\tfrac{1}{2}m_1\cos^2\theta v_s^2 \qquad (4).$$

[0048] The first term on the right side of the above equation represents the kinetic energy of the slider transformed to thermal energy upon collision of the slider. It is reasonable to assume that the energy described by the second term transfers and adds to the tangential kinetic energy of the rotator.

[0049] The tangential kinetic energy of the rotator just prior to the collision is:

$$KE_{rotator\ \alpha}=\tfrac{1}{2}m_1v_\alpha^2 \qquad (5).$$

[0050] Adding the second term on the right of equation 4 to equation 5 to obtain the final post-collision tangential kinetic energy of the rotator, the answer would be:

$$KE_{rotator\ post\ collision}=\tfrac{1}{2}m_1v_\alpha^2+\tfrac{1}{2}m_1\cos^2\theta v_s^2 \qquad (6).$$

[0051] However, the above equation is incorrect because it was assumed the second term could be used in equation 4 to obtain the final result. This equation, derived from the Pythagorean Theorem, is correct from a conservation of energy perspective, but it cannot be used for post-collision analysis. The post-collision tangential kinetic energy of the rotator must first be obtained by vector addition of vector entities.

[0052] The correct procedure is as follows. Just prior to the collision of the slider, the rotator has a linear velocity in the positive y-direction equal to the linear velocity of the slider. Thus, the rotator has a linear momentum in the positive y-direction equal to:

$$p_{rotator}=m_1v_s \qquad (7).$$

[0053] At collision, a part of this momentum is transferred to the front bumper or the earth, and the amount of momentum transferred is equal to:

$$p_{bumper}=\sin\theta m_1v_s \qquad (8).$$

[0054] By the conservation of angular momentum the remaining momentum is transferred to the rotator, such that there is an increase in the post-collision angular momentum of the rotator, equal to:

$$m_1v_fr=m_1v_\alpha r+m_1\cos\theta v_sr \qquad (9).$$

[0055] Dividing both sides of the above equation by m_1 and r, we obtain:

$$v_f=v_\alpha+\cos\theta v_s \qquad (10).$$

[0056] Using the above, the correct expression for the final post-collision tangential kinetic energy of the rotator is derived, equal to:

$$KE_{rotator\ pst\ collision}=\tfrac{1}{2}m_1v_\alpha^2+m_1v_\alpha\cos\theta v_s+\tfrac{1}{2}m_1\cos^2\theta v_s^2 \qquad (11).$$

[0057] Now, recalling equation 6, we have:

$$KE_{rotator\ post\ collision}=\tfrac{1}{2}m_1v_\alpha^2+\tfrac{1}{2}m_1\cos^2\theta v_s^2$$

[0058] Subtracting the above two equations, the difference is obtained as:

$$m_1v_\alpha\cos\theta v_s$$

[0059] Recalling equation 3:

$$\tfrac{1}{2}m_1v_i^2=\tfrac{1}{2}m_1v_\alpha^2+\tfrac{1}{2}m_2v_s^2+\tfrac{1}{2}m_1v_s^2$$

[0060] Substitute equation 4 into the last term of the above, giving:

$$\tfrac{1}{2}m_1v_i^2=\tfrac{1}{2}m_1v_\alpha^2+\tfrac{1}{2}m_2v_s^2+\tfrac{1}{2}m_1\sin^2\theta v_s^2+\tfrac{1}{2}m_1\cos^2\theta v_s^2$$

[0061] Regroup the above as follows:

$$\tfrac{1}{2}m_1v_i^2=(\tfrac{1}{2}m_2v_s^2+\tfrac{1}{2}m_1\sin^2\theta v_s^2)+(\tfrac{1}{2}m_1v_\alpha^2+\tfrac{1}{2}m_1\cos^2\theta v_s^2) \qquad (12).$$

[0062] This is the correct equation for the total energy of the system prior to the collision of the slider. Now take equation 12 and substitute equation 11 for the last two terms in the last parentheses, giving the total energy of the system after the collision of the slider:

$$KE_{system\ final}=(\tfrac{1}{2}m_2v_s{}^2{}_{te}+\tfrac{1}{2}m_1\sin^2\theta v_s{}^2{}_{te})+(\tfrac{1}{2}m_1v_\alpha^2+m_1v_\alpha\cos\theta v_s+\tfrac{1}{2}m_1\cos^2\theta v_s^2) \qquad (13).$$

[0063] Where, the subscript represents the thermal energy transformed from kinetic energy after the collision of the slider.

[0064] The left side of equation 12 represents the initial tangential kinetic energy of the rotator or the initial total energy of the system prior to collision. Equation 13 represents the total energy of the system after the collision. Clearly, there has been a gain in the total energy of the system by the amount of:

$$\Delta E = m_1 v_\alpha \cos\theta v_s$$

[0065] Substituting v_β for $\cos\theta v_s$, the above can be expressed as:

$$\Delta E = m_1 v_\alpha v_\beta$$

[0066] By the first law of thermodynamics there must be an influx of energy-momentum to the system to account for this gain in energy. One possible explanation for the influx may be found in a paper that postulates inertia is due to a time rate change of incoming momentum flux to an accelerating object from the vacuum energy-momentum of space (A. Rueda & B. Haisch, *Contribution to inertial mass by reaction of the vacuum to accelerated motion*, Foundations of Physics, Vol. 28, No. 7, pp. 1057-1108 (1998) which is incorporated herein by reference). The paper states:

[0067] We demonstrate that from the point of view of a nearby inertial observer there exists a net energy and momentum flux (Poynting vector) of ZPF radiation transiting the accelerating object in a direction necessarily opposite to the acceleration vector. The scattering opacity of the object to the transiting flux creates the back-reaction force customarily called the inertial reaction.

[0068] If the above is true, then at the time of the collision of the slider an influx of momentum from the vacuum transits the rotator in the positive y-direction, opposite to the instantaneous linear de-acceleration of the rotator and slider in the negative y-direction. A portion of this momentum, equal to $m_1 \cos\theta v_s r$, transfers to the rotator and increases its angular momentum by an amount found in equation 9, which was derived as a consequence of the conservation of angular momentum.

[0069] It has been known since the last century that energy has inertia implied by Einstein's equation $E=m_0c^2$. Rearranging the equation, we see the mass equivalent of energy expressed as $E/c^2=m_0$. That is, energy in the form of radiation can interact with physical objects as if the energy had the properties of material objects—momentum, etc. Hence, if we assume space is filled with an irrepressible average minimum vacuum energy as predicted by the Heisenberg uncertainty principle in part composed of electro-magnetic radiation, the interaction of accelerating objects with the photons comprising this radiation could indeed create a reaction force. For example, a photon contains momentum equal to:

$$p=hv/c$$

[0070] It is conceivable that photons "colliding" with the rotator opposite to the instantaneous linear de-acceleration of the rotator could transfer part of their momentum to the rotator. And if this were true, along with a transfer of momentum, energy would necessarily transfer also. For example, in the relativistic expression relating the relationship between kinetic energy and momentum we have:

$$(K+m_0c^2)^2=(pc)^2+(m_0c^2)^2$$

[0071] Reduced to the Newtonian limit, the above can be expressed as:

$$K=(p)^2/(2m_0)$$

[0072] Hence, any increase in the angular momentum of the rotator due to this transfer of momentum from the vacuum must be accompanied by an increase in kinetic energy with respect to a laboratory frame.

[0073] The above explanation is more plausible than any Machian explanation. One may incorrectly argue that the increase in energy is due to linear frame-dragging, the so-called Lens-Thirring effect, predicted by the general theory of relativity. To date, the Gravity Probe B experiment has not conclusively proven this effect, but even if it did, it would not have any bearing on this discussion. The frame-dragging effect is only observed in the de-accelerating frames of the rotator and the slider as the slider collides with the front bumper. It is not observed in the frame of the laboratory because the mass of the universe is obviously not accelerating with respect to this frame. To insist this effect must apply to the laboratory frame would violate the principle of relativity by implying the de-accelerating frames of the rotator and slider are preferred frames and the effects observed in these frames must be imposed in the frame of the laboratory. An explanation that only applies to the laboratory frame is demonstrated below.

Example

[0074] An experiment was performed to test if energy gain occurred in a rotator-slider system. The experiment tested two hypotheses. The momentum influx hypothesis (MIH) asserted there would be a gain in the post-collision rotational energy of the rotator, and therefore, a gain in the energy of the system as a whole by the action of an influx of energy-momentum to the system via inertia. The null hypothesis or classical hypothesis (CH) asserted there would be no gain in energy of the post-collision rotational energy of the rotator, and therefore, there would be no gain in the system as a whole. The null hypothesis was designated as the classical hypothesis because it takes the classical Newtonian assumption that inertia is an intrinsic property of matter, and therefore, could never affect a gain in energy of any system. The momentum influx hypothesis takes the position that inertia is an extrinsic property of matter a link to the vacuum energy of space under the right conditions. In the experiment the rotator is a physical object. The equations derived in this paper were derived for a point particle. Thus, we must modify the equations previously demonstrated by including the moment of inertia of the rotator.

[0075] Expressed mathematically, the CH maintains the following:

$$\tfrac{1}{2}I\omega_i^2=\tfrac{1}{2}m_2v_s^2+\tfrac{1}{2}m_1\sin^2\theta v_s^2+\left(\tfrac{1}{2}I\omega_\alpha^2+\tfrac{1}{2}m_1\cos^2\theta v_s^2\right) \quad (1).$$

[0076] The above essentially states that the post-collision energy of the rotator, the terms in parentheses, when added to the other terms, equals the initial rotational kinetic energy of the system. Thus, the energy of the system is conserved.

[0077] The MIH is expressed mathematically as:

$$E_{final} = \frac{1}{2}m_2v_s + \frac{1}{2}m_1\sin^2\theta v_s^2 + \left(\frac{1}{2}I\,\omega_\alpha^2 + \frac{I\,\omega_\alpha\cos0\theta v_s}{r} + \frac{I\cos^2\theta v_s^2}{2r^2}\right). \quad (2)$$

[0078] The terms in parentheses above represent the post-collision energy of the rotator derived from the conservation of angular momentum. When added to the remaining terms, the above is greater than equation 1 by the following, which expresses the gain in energy of the system:

$$\Delta E = \frac{I\omega_\alpha \cos\theta v_s}{r} + \frac{I\cos^2\theta v_s^2}{2r^2} - \frac{m_1\cos^2\theta v_s^2}{2}. \tag{3}$$

[0079] Where:

[0080] I=the moment of inertia of the rotator

[0081] ω_α=the angular velocity of the rotator prior to collision

[0082] θ=the angle of the rotator at the time of collision

[0083] v_s=the linear velocity of the slider prior to collision

[0084] r=the distance from the axis to the center of mass of the rotator

[0085] In the experiment the post-collision energy of the rotator was measured and compared to the MIH and CH prediction. The collected data is in Table 1 while the results are presented in Table 2.

[0086] In Table 1 and Table 2 the following terms are defined:

[0087] ω_α=the measured angular velocity of the rotator prior to collision

[0088] θ=the measured angle of the rotator at the time of collision

[0089] v_s=the measured linear velocity of the slider prior to collision

[0090] ω_i=the measured initial angular velocity of the rotator

[0091] ω_f=the measured final post-collision angular velocity of the rotator

[0092] ke_i=the measured initial kinetic energy of the rotator

[0093] ke_α=the measured kinetic energy of the rotator prior to collision

[0094] ke_f=the measured final kinetic energy of the rotator after collision

[0095] m_1=the mass of the rotator

[0096] m_2=the mass of the slider

[0097] I=the moment of inertia of the rotator

[0098] r=the radial distance from the axis to the center of mass of the rotator

[0099] m_1 was 3.89 kg, m_2 was 1.97 kg, l was 0.403 kg m^2, and r was 0.314 m.

TABLE 1

ω_α	θ	v_s	ω_i	ω_f	ke_i	ke_α
8.80	32.0	0.294	9.81	9.72	19.4	15.6
8.80	34.5	0.450	9.54	9.44	18.3	15.6
7.30	39.0	0.540	8.89	8.71	15.9	10.7
7.33	45.0	0.622	9.62	9.17	18.6	10.8

TABLE 2

ke_f(J)	MIH Prediction (J)	CH Prediction (J)	Net over CH (J)
19.0	20.2	15.73	3.3
18.0	21.5	15.87	2.1
15.3	16.2	11.08	4.2
16.9	16.3	11.20	5.7

[0100] As can be seen, the net gain over CH predicted values was from about 13% to about 50%.

[0101] The foregoing detailed description describes the invention with reference to specific exemplary embodiments. However, it will be appreciated that various modifications and changes can be made without departing from the scope of the present invention as set forth in the appended claims. The detailed description and accompanying drawings are to be regarded as merely illustrative, rather than as restrictive, and all such modifications or changes, if any, are intended to fall within the scope of the present invention as described and set forth herein.

What is claimed is:

1. An energy conversion system for extracting energy from the zero-point field of space, comprising:

a) a primary transducer including a substrate having an arm rotatably coupled to the substrate, said substrate being confined to move along an oscillatory path; and

b) a secondary transducer operatively associated with the primary transducer, the secondary transducer converting kinetic energy of the substrate and arm to usable energy.

2. The system of claim **1**, wherein the oscillatory path is a linear track defined by a pair of resilient stationary members oriented on both ends of the linear track.

3. The system of claim **1**, further including a drive motor operatively associated with the arm and configured to initiate rotational motion of the arm and sustain the rotational motion at a critical angular velocity.

4. The system of claim **2**, wherein the substrate has a substrate mass (m_2) and the arm has an arm mass (m_1) such that a ratio of m_2 to m_1 is chosen to maximize a gain in kinetic energy (ΔKE) of the substrate and the arm.

5. The system of claim **1**, further comprising a secondary arm rotatably coupled to the substrate and having a secondary axis of rotation spaced apart from an axis of rotation of the arm, said secondary arm being coupled to the arm such that both the arm and the secondary arm rotate at a common angular velocity and opposite rotational directions, and wherein the secondary arm has a substantially identical mass and dimensions to that of the arm.

6. The system of claim **5**, wherein the secondary arm rotates about a common axis of rotation with the arm.

7. The system of claim **1**, wherein the secondary transducer includes at least one induction coil system including a magnet and a complimentary conductive coil, such that the magnet is physically coupled to the substrate and oriented to approach the complimentary conductive coil sufficient to induce electrical current therein during movement of the substrate along the oscillatory path.

8. The system of claim **1**, wherein the secondary transducer includes piezoelectric material operable to convert a force applied by the substrate to the secondary transducer to an electric potential in response to applied mechanical stress.

9. The system of claim **1**, wherein the secondary transducer includes a mechanical coupling between the substrate and an energy translator.

10. The system of claim **9**, wherein the energy translator is a linear to rotational motion drive, heat surfacing, or rack and pinion.

11. The system of claim **1**, wherein the usable energy is mechanical, thermal and/or electrical energy.

12. An energy conversion system for extracting energy from the zero-point field of space, comprising a means for extracting momentum-energy from vacuum by the action of inertia and a means for converting the momentum-energy to usable energy operatively associated with the means for extracting momentum-energy.

13. The system of claim **12**, wherein the means for extracting momentum-energy is a substrate having an arm rotatably coupled to the substrate which substrate is confined to move along an oscillatory path.

14. The system of claim **12**, wherein the means for converting the momentum-energy is an induction coil system, heat dissipater, magnetic brake, friction brake, or linear to rotational motion translator.

15. A method of extracting energy from the zero-point field of space, comprising:

a) inducing rotational motion of an arm rotatably mounted to a substrate which is confined to move along an oscillatory path and wherein the rotational motion is sufficient to impart movement of the substrate along the oscillatory path; and

b) converting momentum-energy of the substrate to usable energy.

16. The method of claim **15**, wherein a speed of the rotational motion, relative masses of the arm and the substrate, geometry of the arm, and the geometry of the substrate are designed such that zero-point energy is utilized to maintain the movement of the substrate.

17. The method of claim **15**, wherein the inducing is accomplished by a drive motor.

18. The method of claim **15**, wherein the movement of the substrate follows a linear oscillatory path having two endpoints and the arm follows a cyclic path such that translational motion of the substrate in a y-direction along the oscillatory path substantially coincides with translational motion of the arm in the y-direction.

19. The method of claim **18**, wherein the arm has two theoretical maximum energy transfer positions that occur twice each full cycle of the arm, corresponding to two angles of the arm with respect to an x-axis, which angles occur simultaneously at the time of collision of the substrate with a stationary member.

20. The method of claim **15**, wherein the step of converting is accomplished by inductive coupling of conductive wire and at least one magnet.

21. The method of claim **15**, wherein the usable energy is mechanical, thermal and/or electrical energy.

22. The method of claim **15**, further comprising inducing rotational motion in a secondary arm in an opposite direction to that of the arm such that motion along the oscillatory path of each of the arm and the secondary arm substantially coincide with each other.

23. A method of extracting energy from the zero-point field of space, comprising:

a) inducing multi-mode motion in a primary transducer along an oscillatory path, wherein the multi-mode motion includes a primary translational motion and a secondary motion such that abrupt changes in the primary translational motion are sufficient to impart a momentum influx to the secondary motion; and

b) converting the momentum influx to usable energy.

24. The method of claim **23**, wherein the momentum influx is a result of radiation pressure due to scattering in part of the electro-magnetic momentum-energy contained within the zero-point field of space opposite to centripetal acceleration of the arm.

* * * * *

rexresearch.com

Torsion Field Generator Patents (II)

---

http://worldwide.espacenet.com/advancedSearch?locale=en_EP
European Patent Office Advanced Search

See also : Torsion Fields (I)

Photoshop Lightroom CC
Edit, organize, store, and share photos anywhere.
Just US$9.99/mo.
Join now
Adobe

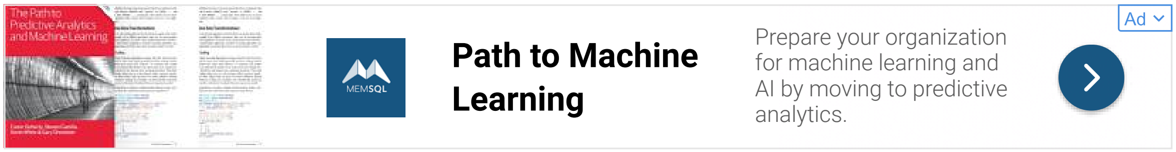
Ad
Path to Machine Learning
Prepare your organization for machine learning and AI by moving to predictive analytics.

Ad
ELECTRICAL AND TOOLING COMPONENTS
SEE WHAT'S NEW!
MiSUMi

RU2126698
FIELD CONCENTRATOR "ALFATRON"

Inventor(s): DVORNIKOV V M; ZHDANOVA I A; JASTREMSKIJ JU N +

FIELD: medical engineering; protection of man against harmful radiations. SUBSTANCE: concentrator has working member with body of cavity-type structure in the form of dodecahedron combined with pyramid. Working member is provided with modulators. Modulator may be manufactured of ferromagnetic material and fitted with a drive. The modulator drive is positioned outside body and coupled to modulator by means of torsion shaft. Modulator may be made in the form of circular generator of electromagnetic waves or vacuumsphere. It may be removable with active element, for example, infrared radiator installed in it. Modulator may be of cross-like form made of silver. Working member body may be made of dielectric, for example, dry cedar. At least one of body edges is perforated and edge opposite to it has mirror surface. Concentrator may be provided with screen made, for example, in the form of ovoid. EFFECT: higher efficiency.

DESCRIPTION

The present invention relates to devices used in the medtech and ecology to protect people from pathogenic radiation.
[0002]

Known concentrator field (RU 2012387 C1, 15.05.94), including the operating element with the body cavity structure, equipped with the modulators.
[0003]

This hub allows you to create wave radiation, but the means for focusing the generated radiation are missing.
[0004]

The invention allows to form wave radiation, followed by focusing it flow in a particular area, which increases its effectiveness.
[0005]

This is achieved in that the hub of the field that contains the operating element with the body in the form of a cavity structure, equipped with modulators, abdominal structure is in the form of a dodecahedron, combined with the pyramid.
[0006]

Concentrator characterized in that the modulator is made of a ferromagnetic material.
[0007]

Concentrator characterized in that the modulator is provided with a drive.
[0008]

Concentrator characterized in that the drive is installed outside the housing and is connected to the modulator torsion shaft.
[0009]

Concentrator characterized in that the modulator is formed as an annular electromagnetic field generator.
[0010]

Concentrator characterized in that the static modulator is designed as a vacuum areas such as light bulbs.
[0011]

Concentrator characterized in that the modulator is removable, with the possibility of its active element, such as an IR emitter.
[0012]

Concentrator characterized in that the modulator is designed as a cross shape, for example of silver.
[0013]

Concentrator characterized in that the working body member is made of a dielectric material, such as dry cedar, wherein at least one of the faces of the housing is perforated, and the opposite face is formed with a mirror surface.
[0014]

Concentrator characterized in that the operating element and modulators placed inside the shield made of, for example, in the form of ovoids.
[0015]

The curved surface of the cavity structure provides the imposition of de Broglie waves inherent in the surface of any body (see. Eg, GI Shipov "Theory of physical vacuum", M., NT-Center, 1993) and forms a standing wave maxima of the increased total energy .

In this case the abdominal shape in the form of a source dodecahedrons wave radiation, and ferromagnetic provide magnetic modulators "pumping" and forming a directional flow of energy (see., E.g., H.Alven "Space plasma", M., 1983).
[0016]

By combining cavity structures with different types of modulators (ferromagnetic, minerals, vacuum areas cruciform shape, biological objects, IR emitters, laser, microwave radiators, a gas burner, etc.) Hub provides cumulative effect of the above fields and sources Radiation treatment to the object.

Impact focuses pyramid and performed during the activation, or disintegration, which is determined by the direction of rotation of the field in the concentrator (respectively "right" and "left" rotation).
[0017]

The screen in the form of ovoid is a carrier of energy and battery with a pronounced maximum field strength in the narrow part of the ovoid (see., Eg, TE Gonikman "philosophy stone", "Santana", Minsk, 1997).
[0018]

Performing the working body member cedar (i.e. energetically active material) from the perforated surface at least one of the faces of cavity structure coupled with its mirrored surface opposite face provides additional modulation field at the output of the concentrator.
[0019]

Placing all elements of the hub inside the screen provides Ecodefence the surrounding area from the effects of radiation, as well as improving the efficiency of the hub by reducing external losses.

Removal of the drive dynamic modulator of the body with the use of a torsion shaft to drive the modulator to eliminate the creation of internal interference.

[0020]

Construction and operation of the hub illustrated by a drawing, which shows a general view, and its basic elements.
[0021]

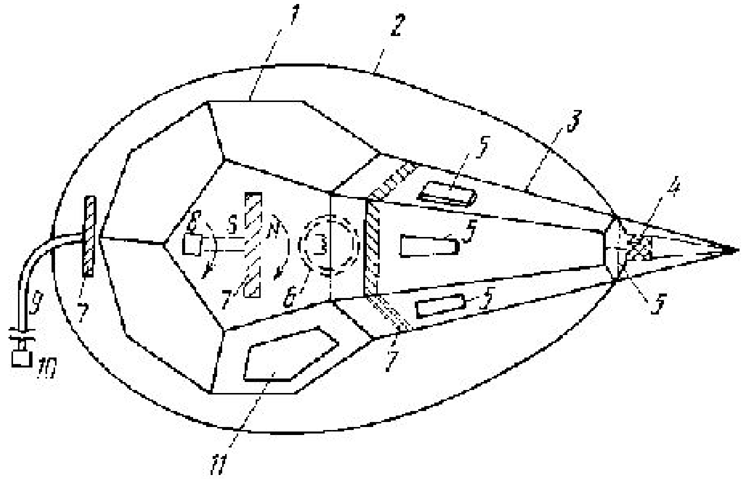

The accepted designations: 1 - dodecahedron; 2 - the screen; 3 - the pyramid; 4 - camera removable modulator; 5 - static modulators; 6 - Vacuum sphere; 7 - ferrodynamic plate dynamic and static modulators; 8 - Dynamic Drive modulator; 9 - rotating shaft remote operator; 10 - remote operator; 11 - a fragment of the coating of dielectric material with perforations of the casing.
[0022]

Annular electromagnetic field generator separately not shown because it is performed as an option actuator 8, but an open stator.
[0023]

Powered Hub fields as follows.
[0024]

Before its application identified properties of the object of processing.

This procedure is performed, for example, by a sensor Foll, birefringent liquid sensor, etc.
[0025]

Thereafter, the vertex of the pyramid with removable modulator camera 4 focuses on the object of processing (not shown).

Then include drive 8 or 10 depending on the task activation or disintegration.
[0026]

With a time interval equal to the time of transients in a processing object, check the rate of change of properties - for example the degree of activation of water.
[0027]

With insufficient change in the properties of water (or other object processing) adapting processing mode by: - ??replacement of the removable modulator or an active member; - Changes in speed and direction of rotation of the field.
[0028]

Experimental verification of the Hub held under the following conditions.
[0029]

Work Items cavity structure and screen are made of a polymeric material with the inner surface of a mirror.
[0030]

As modulators used ferromagnetic plate, minerals, biological objects (antlers, mummy, extracts of herbs and plants, musk deer musk glands, BIOTON plate), as well as sources of infrared and microwave radiation, a gas burner, a voltaic arc, and others.
[0031]

Vacuum services also serve as static modulator and had parameters: an outer diameter of 35 mm, internal 30 mm, wall thickness 2.5 mm, the material - stainless soft magnetic steel, the degree of vacuum of 10-5 Torr

It is shown that the most simple, everyday implementation is provided by the use of light bulbs, such as a tungsten filament.
[0032]

Application of remote drive the modulator eliminates interference from the engine, which is most important when using a hub-activated, ie, to create the right torsion field (right spinpolyarizatsiya).

This mode is designed for new medical products and liquids, as well as suitable for direct patient care.
[0033]

Generator annular electromagnetic field appropriate to apply modes of disintegration, that is. E. To create a left torsion field (left spinpolyarizatsiya) used, for example, for radiation decontamination of pathogens and chemical contaminants.
[0034]

Experiments have confirmed that the proposed concentrator field provides a robust activation of water, availability of water to new bioactive therapeutic properties determined by the ability to apply modulators.

RU2012387
OPTOMAGNETIC APPARATUS

CHAJCHUK ROSTISLAV A [UA]; LUKASHENKO VLADISLAV I [UA]; SHULTE ALEKSANDR YU [UA]; TOMKIN YURIJ V [UA]; PECHKUROV VIKTOR P

DESCRIPTION
[0001]

The invention relates to medical engineering and is intended for storage space bioenergetic radiation.
[0002]

Known bioenergy radiation generator comprising a metal-halogen lamp with reflector - chilled mirror in the shape of a sphere or a paraboloid, infrared and ultraviolet filter, polarizer [1].
[0003]

The disadvantage of this generator is the low intensity of bioenergy radiation and the polarization of light.
[0004]

Considered the closest analogues of the present invention is to optomagnetic apparatus comprising a lamp, fitted with a lens interference layer, a reflector 4 and a system of permanent magnets of an alloy of cobalt with lanthanide [2].
[0005]

The disadvantage of this device is the lack of optomagnetic accumulation bioenergy radiation and low intensity.
[0006]

The aim of the invention is to increase the intensity of bioenergy radiation.
[0007]

The goal is achieved by the constant introduced in time and variable-axis filament lamp magnetic field magnetic coils and accumulation of bioenergy radiation pyramidal battery of a nonmagnetic material.
[0008]

Comparative analysis of the claimed solution over the prototype shows that the proposed apparatus is characterized by having optomagnetic pyramidal battery of a nonmagnetic material and inductors mounted on its lateral faces.
[0009]

The drawing shows a diagram optomagnetic apparatus.
[0010]

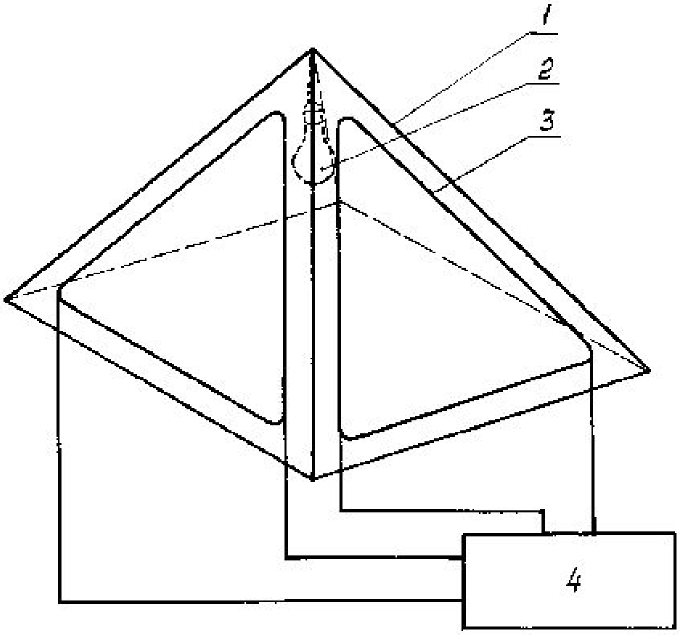

The proposed apparatus consists optomagnetic (FIG. 11) of the battery 1 pyramidal magnetic material having an axis lying on the axis of the filament lamp 2 and the angle of divergence of 60 ° to the faces of the pyramid, the inductance coil 3 located on the outer side of the four side faces of the pyramid and the battery 1 having a triangular shape, the oscillator 4 amplitude-phase modulation, the output of which is connected to the terminals of coils 3, the filament of the lamp 2 is placed at a point on the axis of the pyramidal battery 1, the volume of dividing it into two equal parts, the number of turns in the magnetic coils is 1 - 4.
[0011]

To create the effect of bioenergy use is optimal for the human body to the luminous flux of the surface energy of 120 mW / cm2, obtained by metal-halogen lamp 2, 300 W, and a pulsed magnetic field strength of 0.016 A / m, causing a positive stimulating effect on tissue cells.

Pyramid battery is made of carbon fiber epoxy UGET T-thickness of 5 mm from the side of the square base of the pyramid of 500 mm.

The diameter of the magnetic coils is 0.343 mm2.

The magnetic coil current flows force 0.015 A and forms a pulsating variable in time and the axis of the filament lamp 2 electromagnetic field inside the pyramid battery 1 with a strength of 0.016 A / m.
The interaction of light flux lamp 2 and the magnetic field of the magnetic coil 3 forms and accumulates in the accumulator 1 pyramidal secondary wave radioestezicheskoy nature used for therapeutic purposes under the influence on biological objects.

Radioestezicheskim radiation is called radiation, causing biolocation effect.

Or, in other reaction body operator to change the parameters of this radiation in the search.

Prolonged exposure to its unregulated per person (for example, in areas gepatogennyh) may cause significant harm to health.

At the same time, the skillful use of it is successfully used in therapeutic and prophylactic purposes, to correct bioenergetic condition of the body, activate the vitality and creativity.
[0012]

For optimum interaction of light and magnetic field lamp 2 is set so that the filament was bioenergetic focus pyramidal accumulator 1.

Bioenergy focus pyramid is at a point lying on the axis of the pyramid and dividing its volume into two equal parts.

Regulation radioestezicheskogo radiation to make changes in the magnetic field in amplitude and frequency using the oscillator 4.

The intensity of the radiation radioestezicheskogo regulate the amount of included magnetic coils in the coils, which total in the magnetic coil was 1 - 4.
[0013]

Technical and economic efficiency of the proposed invention consists in the fact that as a result a new principle and an embodiment of its design it is possible to significantly increase the intensity of the radiation bioenergetic that significantly enhance the therapeutic effect.

RU2115965
TORSION GENERATOR

Inventor(s): NEGANOV V A; SALMANOV A N +

Abstract

FIELD: power production, in particular, for medical equipment, agriculture, communication, computer engineering. SUBSTANCE: device is designed as set of emitters of electromagnetic waves of same frequency with narrow beam pattern. Emitters are located in vertices of geometric figure which serves as source of torsion radiation. Emitters which are directed outside of geometric figure are arranged to achieve coincidence of points of equal phases of electromagnetic waves generated by these emitters and vertices of source figure of radiation source. EFFECT: increased power of torsion radiation.

DESCRIPTION
[0001]

The invention relates to the problem of obtaining new sources of energy and can be used in medicine, agriculture, telecommunications, bioinformatics, and so on. D.
[0002]

One class of natural sources of torsion fields along with electromagnetic fields is a wide range of shapes, formed by various substances [1].
According to the model of the physical vacuum Akimov AE, Tarasenko V. Ya can assume that the response to the topological disturbance of the physical vacuum is a redistribution of the torsional potential in the vicinity of the body, which compensates for this disturbance.

As a result, in a neighborhood of the body of the linear distortion of the spatial structure of the physical vacuum is offset by redistribution said torsion potentials, so that with respect to the outer space of the neighborhood is an internal body samokompensirovannoy area [2].
[0003]

Known chronal (torsion) generator formed by a wooden frame in the form of 16-sided regular polygon of about 3 m with four electrodes, to which a DC voltage from 60 kV to 300 kV. [3]
This device reduces its mass due to repulsion from the ground, it chronal (torsion) field accelerates the chemical reaction, reduces fuel consumption in an internal combustion engine almost twice that increases the growth rate of plants and their size by 2 - 10 times, and generates a lot of other interesting effects [3].
[0004]

It is known that the radiation power of torsion associated with the size of the form.

At larger sizes and shapes concentrates more power radiation [1].

For example, the power chronal radiation in a giant pyramid of Cheops is much more than its models, smaller properties are used for the storage of perishable products, maintaining the sharpness of razors and razor blades. [4]
[0005]

The purpose of the invention - increase the capacity of the torsion radiation.
[0006]

The goal is achieved as follows.
[0007]

A set of emitters of electromagnetic waves of the same frequency with a narrow radiation pattern is placed in all the tops of the geometrical shape, which is the source of torsion (chronal) radiation.
Emitters with a narrow beam pattern are arranged so that at any given time point identical phases of electromagnetic waves from them is a vertex of the original shape - the radiation source.
[0008]

As an example, consider the torsional oscillator formed 5th directional source of electromagnetic waves with a narrow radiation pattern arranged at the vertices of a regular pyramid whose base is a square with the following relations (see. Figure).
[0009]

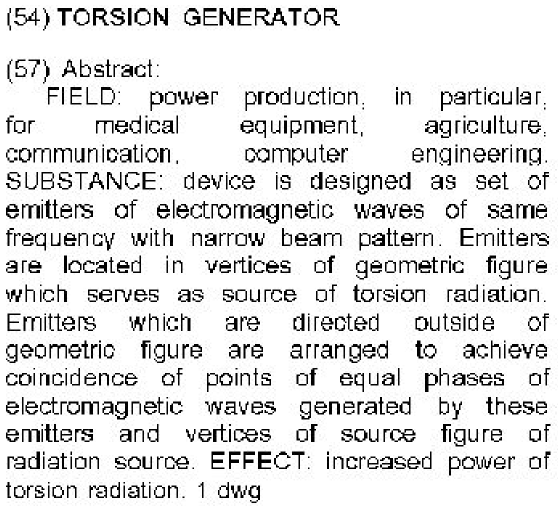

(54) **TORSION GENERATOR**

(57) Abstract:

FIELD: power production, in particular, for medical equipment, agriculture, communication, computer engineering. SUBSTANCE: device is designed as set of emitters of electromagnetic waves of same frequency with narrow beam pattern. Emitters are located in vertices of geometric figure which serves as source of torsion radiation. Emitters which are directed outside of geometric figure are arranged to achieve coincidence of points of equal phases of electromagnetic waves generated by these emitters and vertices of source figure of radiation source. EFFECT: increased power of torsion radiation. 1 dwg

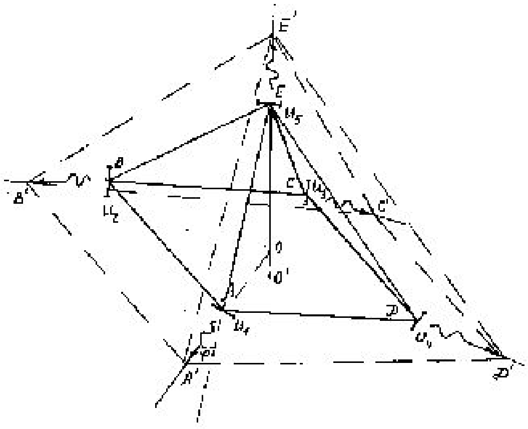

In the drawings: U1, U2, U3, U4, U5 - directed emitters; A, B, C, D, E - the vertex of the pyramid, which are emitters; A ', B', C ', D', E '- points of different phases of the EM wave through time t.
[0010]

h = 0,63a, l = 0,95a (1), where h - height; l - lateral edge; a - at the bottom side of a square pyramid.

Let aimed radiators are excited in phase and oriented in space in such a way that the transmission point of the same phase of the electromagnetic waves on the directions of the maximum transmission power of the emitters at each time form a pyramid, which is valid for the edges of (1).

Chance pulse mode torsion generator when directed emitters simultaneously emit pulses of electromagnetic field.
[0011]

The effectiveness of the proposed solution can be explained as follows.

In the propagation of electromagnetic waves from emitters in the medium with speed V the geometric dimensions of the pyramid will be increasing all the time and thus will increase the power of torsion radiation while increasing the space in which to accumulate torsion (chronal) energy.
[0012]

Application of the torsion oscillator depends on the information that will be incorporated into the torsion radiation.

Possible application areas: communications, agriculture, various processes, medicine.
[0013]

Literature.
[0014]

1.
Akimov AE, Tarasenko BJ Models of polarized states of physical vacuum // Math.
Universities, Physics, 1992, N 3, p.
13 - 23.
[0015]

2.
Veinik AI Thermodynamics of real processes. - Mn .: Navuka: tehnika, 1991, p.
576.
[0016]

3.
Ravatin J. instrument for strengthening emissions caused by forms: Patent N 2.421.531 on 13.07.1978 (France).
[0017]

4.
Drbal K. method of maintaining sharpness of razors and shaving fields: Patent N 91304 (Czechoslovakia).

KR20080030740

A SPACE ENERGY IMPLOSION UNIT AND AN ENERGY AMPLIFICATION GENERATOR USING THE SAME

Inventor(s): LEE YONG WON [KR]; I EUN JAE [KR] +

Abstract

A space energy amplification generator using a space energy implosion unit is provided to make a practical use of space energy or torsion field of the universe in a direct or indirect manner by using a specific geometric pattern. An energy amplification generator comprises: a 1st geometric structure(110) composed of seven pieces of septagonal shape planar structures layered in order; a 2nd geometric structure(120) composed of seven septagonal pyramid shape solid structures(2), a plate structure(140) for supporting the 1st geometric structure; a plurality of partitioning structures installed between the 1st geometric structure, between the 1st and 2nd geometric structures, and between the 1st geometric structure and the plate structure, respectively; a casing(50) provided with a cover(51) having a radiation hole(52) on the top to radiate space energy; and a fastening part(63) including a bolt(63a) fastened by a nut(64).

DESCRIPTION

[1]

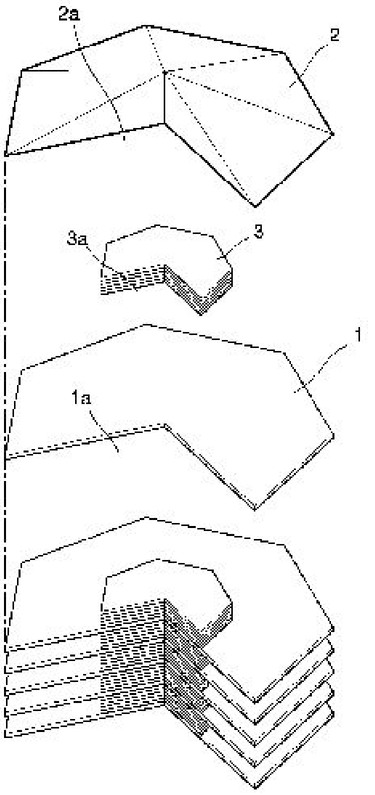

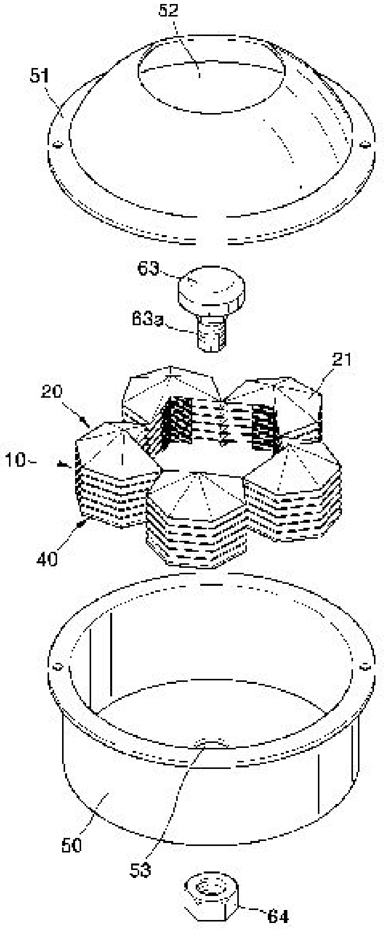

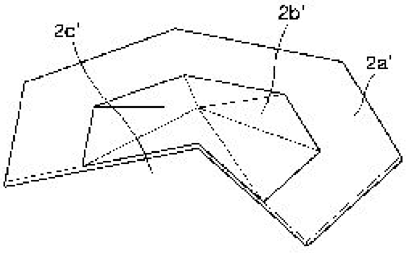

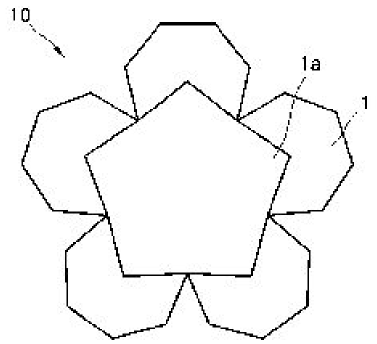
10
1a
1

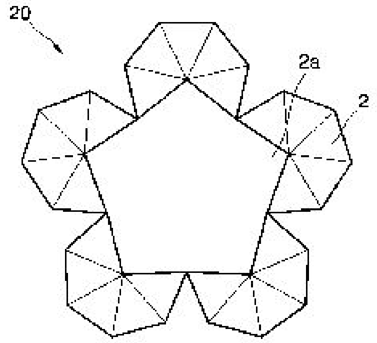
20
2a
2

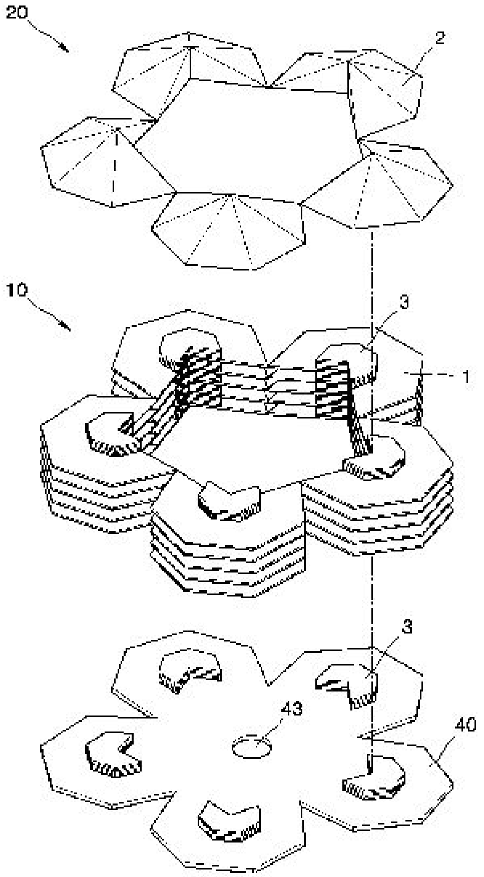

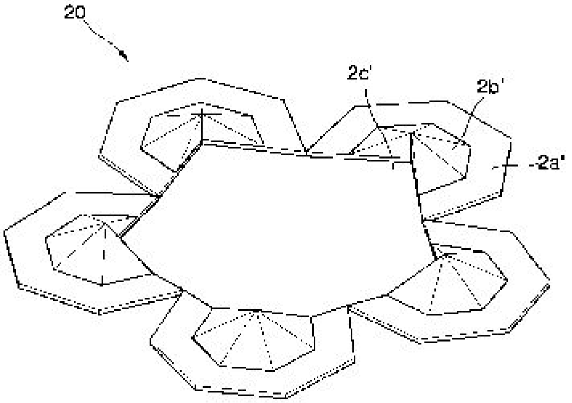
20
2c'
2b'
-2a'

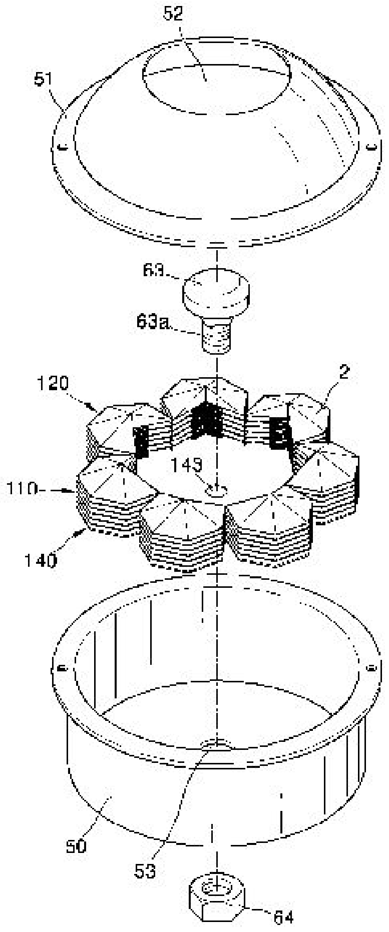

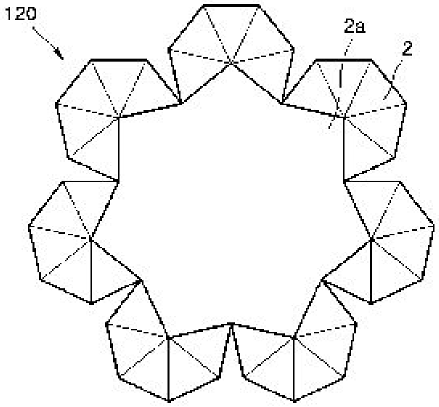
120
2a
2

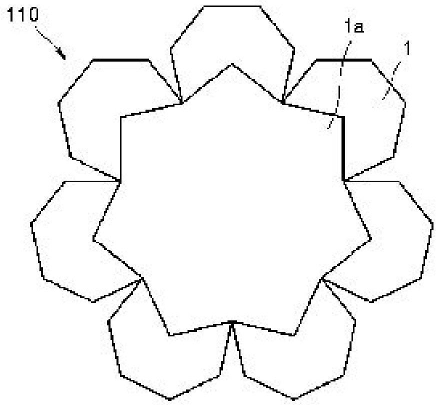
110
1a
1

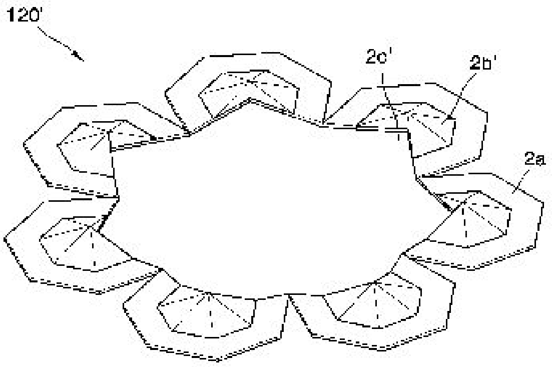
120'
2c'
2b'
2a

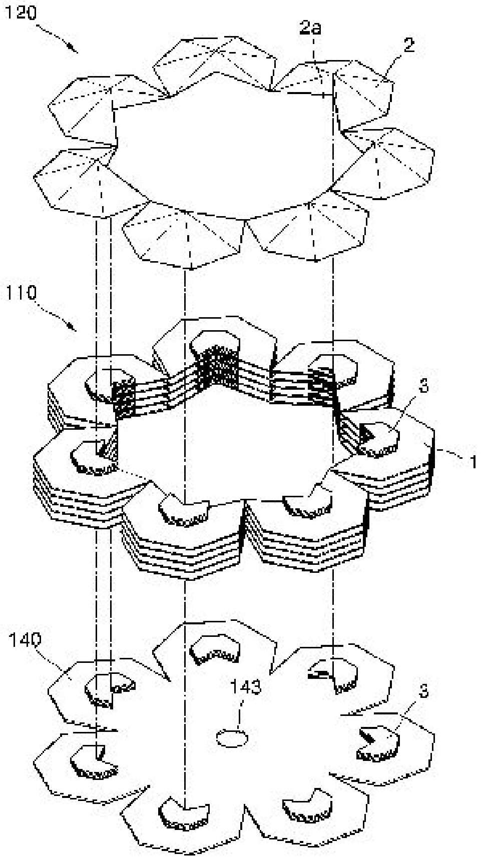

Figure 1 is an exploded perspective view of a space energy eungpok unit according to the invention,
[2]

Figure 2 is a perspective view showing by extracting the different forms of the three-dimensional structure employed in Figure 1,
[3]

Figure 3 shows a first embodiment of space energy eungpok radiation device according to the invention,
[4]

In Fig 4 shows, a perspective view showing by extracting a first geometrical structure,
[5]

In Figure 3 Figure 5 is, a perspective view showing by extracting a second geometrical structure,
[6]

Figure 6 is a perspective view showing by extracting a different second geometric shape of the structure of Figure 3,
[7]

In Fig. 7 is a second geometrical structure, the first to extract the stacked structure of the geometric structure, plate structure and away from the structure diagram showing,
[8]

8 is a diagram illustrating a second example of the spatial energy eungpok emission device according to the invention,
[9]

In Fig 9 is a perspective view showing by extracting a first geometrical structure,
[10]

In Fig 10 is a perspective view showing by extracting a second geometrical structure,
[11]

Figure 11 is a perspective view showing by extracting a different second geometric shape of the structure of Figure 8,

[12]

In Figure 8, a 12, a second geometrical structure, by extracting a laminate structure of a first geometrical structure, plate structure and away from the structure diagram showing,
[13]

13 is a view showing the number of a view showing the contents of Table 1 in a graph, the processing can take place prior to the subject of energy,
[14]

Figure 14 is a view showing the number of subjects in a view showing the contents of Table 2 as a graph, the processing can take place after the energy Erythema symptoms,
[15]

15 is a diagram showing the contents of Table 2 as a graph, space energy treatment after Edema / Papulation shows the number of subjects taking the symptoms, also,
[16]

16 is a diagram showing the contents of Table 2 as a graph, space energy treatment Oozing shows the number of subjects after taking,
[17]

17 is a diagram showing the contents of Table 2 as a graph, space energy treatment Excoriation shows the number of subjects after taking,
[18]

18 is a diagram showing the contents of Table 2 as a graph, the energy treatment space shows the number of subjects after taking Lichenification.
[19]

Description of the Related Art
[20]

1 ...
Flat structures 1a ...
Flat recess
[21]

2 ...
Three-dimensional structures 2a ...
Three-dimensional recess
[22]

2 '...
Three-dimensional structures 2a '...
Flat plate
[23]

2b '...
Sub-three-dimensional structure 2c '...
Three-dimensional recess
[24]

3 ...
Apart structures 3a ...
Incision Home
[25]

10 ...
The first geometrical structures 20, 20 '...
The second geometrical structure
[26]

40 ...
43 structure plate ...
Plate hole
[27]

50 ...
Case 51 ...
Cover
[28]

52 ...
53 four room ...
Cases hole
[29]

63 ...
Fixing members 63a ...
Bolt
[30]

64 ...
Nuts 110 ...
The first geometrical structure
[31]

120, 120, ...
The second geometrical structure 140 ...
Plate Structure
[32]

143 ...
Plate hole

The present invention relates to space and space energy unit energy eungpok eungpok using him spinning device to become available, directly or indirectly, by the fullness of eungpok feeble energy in the universe.

The Newtonian mechanics to define the nature of the particles of the substance was dominated by the classical science, the nature of the substance as quantum mechanics and relativity theory established before and after the 1900 found that particles as well as waves also can understand.

For example, it was understandable that could not touch the energy E = mc2 energy is converted into substances that can be touched by the expression of.

The position x and momentum p of the particle by the Heisenberg uncertainty principle of quantum mechanics, the base will have a value that is not confirmed at the same time subject to restrictions from one another by both the uncertainty (?x?p=h / 2p (h is Planck constant)) This material and non-material particle wave was able to separate targets that can not be called new thinking as understanding.
From this space is made of material with a non-material, material it can be accepted as a fact is apparent that the energy can be converted to the non-substance.

Meanwhile, Russia 1990 seconds scientist Dr. GIShipov (1998 US Electric Association (American Biographic Institute) selected 500 of the most influential leaders in the crazy in the 20th century by a) find that the energy handaraneun response to any particular geometric shape, The energy field was the force of the third non-gravity or an electromagnetic force, that was named LA tosyeonjang (torsion field) or space energy (sapce energy).

Tosyeonjang (space energy), all materials as well as the rotating electric or magnetic, and the resonance effect of the means of mutual influence, or light that results in rotation of the biomolecule.

This space is very weak energy but which is filled in the space, measuring the electromagnetic field existing in a very weak because it is not measured.

However, its presence has been confirmed by the space and the electromagnetic field energy in the material leaving traces (interference).

This space energy is applied to medicine, new treatments and new discipline called quantum medicine is the emergence, in developed countries such as the United States or France, is the situation in the energy space and the tremendous support research in order to study the effect in vivo .

In this manner, the space that is filled with the material and energy translating, research and development is to be used for the real spatial energy (tosyeonjang) caused by the specific geometry of the situation is going on in full in many countries.

The present invention has been devised in order to reflect the trend as described above, to be in place to use the energy eungpok particular geometry of the space available energy effectively fill the space, either directly or indirectly, the energy in the real space by eungpok eungpok to provide a space eungpok energy radiation device using the unit and him for the purpose.

In order to achieve the above object, according to the invention energy eungpok space plane unit structure having positive heptagon shape is cut to leave a continuous from the center to gajina implemented entirely flat recess (1a) to (1); As being installed to be spaced apart on top of the planar structure (1), the three-dimensional cutting gap is implemented to leave a series of incisions from the center gajina the information 7 pyramid shape as a whole (2a) having a three-dimensional structure (2); And as to the spacing between the planar structure (1) and the three-dimensional structure 2, the planar structure (1) smaller than the area of ??the three-dimensional structure spaced formations (2) (3); characterized in that it comprises a.

In order to achieve the above object, a first embodiment of the spatial energy eungpok radiation according to the present invention having such a rectangular shape as a whole tablet 7 is cut to leave the gajina continuous from the center plane implementation recess (1a) and, planar structure (1) 5 sheets of each vertex of the flat recess (1a) is arranged in a plane so as to each contact with a first geometrical structure is implemented (10); Each are installed as spaced apart on top of the planar structure (1), the three-dimensional structure (2) having five forward 7 gajina the pyramid shape is cut from the center to leave a series of three-dimensional implementation of the cutting gap (2a) as a whole, Each vertex of the three-dimensional cutting gap (2a) is arranged in a plane so as to each contact with a second geometrical structure to be implemented (20); And as to the spacing between the first geometrical structure 10 and the second geometrical structure 20, disposed between the plane structure (1) and the three-dimensional structure (2) and the flat structure (1) and the three-dimensional structure (2) smaller than the area of ??a plurality of spaced structures (3); it characterized in that it comprises a.

In order to achieve the above object, a second embodiment of the space eungpok energy radiation according to the present invention having such a rectangular shape as a whole tablet 7 is cut to leave the gajina continuous from the center plane implementation recess (1a) and, planar structure (1) 7 sheets, each vertex of the flat recess (1a) is arranged in a plane so as to each contact with a first geometrical structure is implemented (110); Each are installed as spaced apart on top of the planar structure (1), the three-dimensional structure (2) having seven positive 7 gajina the pyramid shape is cut from the center to leave a series of three-dimensional implementation of the cutting gap (2a) as a whole, Each vertex of the three-dimensional cutting gap (2a) is arranged in a plane so as to each contact with a second geometrical structure to be implemented (120); And as to the spacing between the first geometrical structure 110 and the second geometrical structure 120, the plane structure (1) and provided between the three-dimensional structure (2) and the flat structure (1) and the three-dimensional structure (2) smaller than the area of ??a plurality of spaced structures (3); it characterized in that it comprises a.

Hereinafter, with reference to the accompanying drawings preferred embodiments of the room unit, and a space energy eungpok eungpok energy radiation device using the same according to the present invention will be described in detail.

1 is an exploded perspective view of a space energy eungpok unit according to the invention, Figure 2 is a perspective view taken of the three-dimensional structure of another embodiment is employed in Fig.

As shown, the space eungpok energy unit according to the present invention, the planar structure having overall positive heptagon shape is cut to leave the gajina continuous from the center plane implementation recess (1a) to (1); As being installed to be spaced apart from the top of the planar structure (1), the three-dimensional structure having a shape defined gajina 7 pyramid is cut to leave a series of three-dimensional cutting gap from the center is implemented (2a) as a whole (2) and; As to the spacing between the planar structure (1) and the three-dimensional structure 2, the planar structure (1) smaller than the area of ??the three-dimensional structure spaced formations (2) (3); and the like.

Here, while the plane structure (1) is implemented in more than 1, the more the planar structure it is possible to amplify the area of ??energy more efficiently.

The spacing structure (3), have been implemented cutting gap (3a) to being cut from the center as a whole to leave a series of information gajina heptagon shape.

Wherein the three-dimensional structure (2) and spaced from the structure (3) and the flat structure (1) are stacked to be equal to the arrangement angle of each side of them.

Here, the three-dimensional structure (2 ') may be applied in other forms.

That is, the three-dimensional structure (2 ') is 2, the flat plate 7 of constant rectangular shape as a whole (2a') on the sub-defined than 7 square pyramid shape of the small area of ??the flat plate solid structure (2b ') doedoe implementation is formed, it will have a three-dimensional recess (2c ') that is cut to leave the implementation of the continuous)' flat-plate (2a from the center of) "sub-three-dimensional structure (2b.

The planar structure 1 is to fill the space eungpok spatial energy by its geometry, three-dimensional structure (2) concentrates the energy eungpok space forward from the plane structure (1) by its geometric structure.

And the implementation of the three-dimensional structure (2 ') in the form shown in Figure 2, the space concentrates energy at the same time fill the space Yes pokham forward.

Next, employing the space eungpok energy unit according to the present invention will be described space eungpok energy radiation device.

Here, the same reference numerals are the same as in the second member for the same function.

3 is a diagram showing a first embodiment of the spatial energy eungpok radiation device according to the invention, Figure 4 is 3, by extracting a first geometrical structure is a perspective view, Fig. 5 in Fig. 3 , is a second perspective view showing by extracting a geometric structure, Figure 6 is a perspective view showing by extracting a different second geometric shape of the structure of Figure 3, Figure 7 is 3, the second geometrical structure, the first geometric structure, by extracting a laminate structure of the plate structure, and spaced apart a diagram illustrating the structure.

A first embodiment of the device according to the invention eungpok radiation, space energy eungpok unit 5 of Figure 1 is implemented by a dog arranged so that the vertex of each contact therebetween, thereby to form a new geometric structure.

This will be described more in detail.

As shown, the space eungpok energy radiation device, a whole tablet heptagon gajina a flat recess shape is implemented to leave a series of incisions from the center (1a) has a plane structure (1) 5 sheets, also according to the invention 4, each of the flat recess (1a) a first and a geometric structure 10 vertices are arranged in a plane so as to be in contact with each other as shown in the implementation; As being installed to be spaced apart from the upper portion of each of the planar structure (1), the three-dimensional structure (2) having five forward 7 gajina the pyramid shape is cut from the center to leave a series of three-dimensional implementation of the cutting gap (2a) as a whole, Fig. 5, each of the second geometrical structure 20 and are arranged in a planar vertex implementation to each contact of a three-dimensional cutting gap (2a), as shown in; As to the spacing between the first geometrical structure 10 and the second geometrical structure 20, the area of ??the planar structure (1) and the three-dimensional structure 2 is provided between the planar structure (1) and the three-dimensional structure (2) than the A small plurality of spaced structures (3); and the like.
A first geometrical structure 10 is composed of one or two or more.

A first geometrical structure 10 for use thereby eungpok spatial energy chungban in space by the geometrical structure directly or indirectly.

A second geometrical structure 20 focuses the spatial energy eungpok by the first geometrical structure 10 in a forward direction.

Area of ??the first geometrical structure 10 and the second geometrical structure 20 will increase the amount of space energy eungpok.

That is Either increasing the stacking number of the first geometrical structure 10 in order to obtain more space than the energy, the larger the area of ??the first and second geometrical structures (10, 20).

On the other hand, the second geometrical structures (20 ') are capable of modifications in a different shape.

That is, as shown in Figure 6, the implementation is entirely defined doedoe heptagon shape flat plate (2a ') (the sub-three-dimensional structure 2b) of the hexagonal pyramid shape of the small area information. 7 than that in the flat plate, the formation of, that will be formed of a "three-dimensional structure (2 having a) is cut to leave a series of three-dimensional implementation recess (2c) 'of (flat plate 2a) from the center of the sub-three-dimensional structure (2b)'.

The three-dimensional structure (2 ') of five, three-dimensional recess (2c) of the second geometric structure present by vertices arranged in a plane so as to each contact (20') is implemented.

The three-dimensional structure, the second having a geometrical structure (20, 2 ') in the forward direction at the same time concentrates the response pokham spatial energy.

Plate structure 40 as to support the first geometrical structure 10, the first has the same shape as the overall shape of the geometric structure 10.

In the center of this plate structure 40 is formed in the plate holes 43 corresponding to the case hole 53 will be described later.

Spacing structure (3) is provided between the installed between a number of first geometrical structure 10 is between the first geometrical structure and spaced apart from each other such that the first geometrical structure 10 and the second geometrical structure 20 The geometry is such that the first structure 10 and the second geometrical structures 20 are spaced apart from each other, it is disposed between the first geometrical structure 10 and the plate structure 40 first geometrical structure 10 and the plate structure (40) such that spaced apart from each other.

The spacing structure (3) is entirely defined heptagon gajina a shape having a recess (3a) is implemented by cutting from the center to leave a series of, separately, a number of stacked plates having a predetermined thickness to a desired thickness or implemented It may be implemented in.

Spacing structure (3) is different from the frequency of the energy radiated area.

That is, the thickness of the spacing structure (3) to vary the frequency of the spatial energy becomes a thin or thick.
The thickness of the spacing structure (3) is preferably in the range of 3mm ~ 20mm, in this embodiment as a 7mm.

Case 50 includes a spinneret (52) has a cover is formed (51) so that the energy emitted to the front space.

Inside the case 50 is the case hole 53 is formed corresponding to the hole plate (43).

A first geometrical structure to be laminated on the one plate structure 40 (10), spaced from the structure (3) and a second geometrical structure 20 is built into the case 50.

That is, the first and second geometrical structures (10, 20) and away from the structure (3) is laminated plate structure 40, as shown in Figure 3, the bolt part (63a) of the holding member 63 A hole through the plate 43 and the casing hole 53 and the case 50 by fastening by a nut 64 at the rear, and is fixed to the inner case (50).

Next, a description will be given of a second embodiment of eungpok emitting device according to the invention.

8 is a diagram showing a second embodiment of the spatial energy eungpok emission device according to the invention, in FIG. 8, 9, to extract a first geometrical structure is a perspective view, Figure 10 is an 8-degree , is a second perspective view showing by extracting a geometric structure, Figure 11 is a perspective view showing by extracting a different second geometric shape of the structure of FIG, 12 is in Figure 8, a second geometrical structure, the first geometric structure, by extracting a laminate structure of the plate structure, and spaced apart a diagram illustrating the structure.

Here, the same reference numerals are the same as in the second member for the same function.

Second embodiment of the device according to the invention eungpok radiation energy and Space eungpok unit of Figure 1 implemented by seven vertex arranged to the mutual contact between them, and is accordingly to form a new geometric structure.

This will be described more in detail.

As shown, the space eungpok energy radiation device, a whole tablet heptagon gajina a flat recess shape is implemented to leave a series of incisions from the center (1a) has a plane structure (1) 7 sheets, also according to the invention 9, a first geometrical structure 110 and the vertex of a flat recess (1a) arranged in a plane so as to be implemented as shown in contact with each other; As being installed to be spaced apart from the upper portion of each of the planar structure (1), the three-dimensional structure (2) having seven positive 7 gajina the pyramid shape is cut from the center to leave a series of three-dimensional implementation of the cutting gap (2a) as a whole, Fig. 10 and the second geometrical structure 120 vertices are implemented is arranged in a plane so as to each contact of a three-dimensional cutting gap (2a), as shown in; As to the spacing between the first geometrical structure 110 and the second geometrical structure 120, the area of ??the planar structure (1) and the three-dimensional structure 2 is provided between the planar structure (1) and the three-dimensional structure (2) than the A small plurality of spaced structures (3); and the like.

A first geometrical structure 110 is composed of one or two or more.

A first geometrical structure 110 thereby to use the space eungpok chungban energy in space, by the geometric structure directly or indirectly.

A second geometrical structure 120 focuses the spatial energy eungpok by the first geometrical structure 110 in the forward direction.

Area of the first geometrical structure 110 and the second geometrical structure 120 will increase the amount of space energy eungpok.

That is Either increasing the stacking number of the first geometrical structure 110 in order to obtain more space than the energy, the larger the area of ??the first and second geometrical structure 110, 210.

On the other hand, the second geometrical structures (120 ') are capable of modifications in a different shape.

That is, as shown in Fig. 11, is formed as a whole positive doedoe implement 7 with a square-shaped flat plate (2a ') sub-three-dimensional structure of the hexagonal pyramid shape of the small area information. 7 than that in the flat plate (2b') is, that will be formed of a "three-dimensional structure (2 having a) is cut to leave a series of three-dimensional implementation recess (2c) 'of (flat plate 2a) from the center of the sub-three-dimensional structure (2b)'.

The three-dimensional structure (2 '), seven, three-dimensional recess (2c' by the vertex of) is arranged in a plane so as to each contact with a second geometrical structure of the present application (120 ') is implemented.

This "second has a geometric structure (120 three-dimensional structure 2 ') is a space energy response in the forward concentrates pokham and at the same time.

Plate structure 140 as to support the first geometrical structure 110, the first has the same shape as the overall shape of the geometric structure (110).

In the center of this plate structure 140 is formed in the plate holes 143 corresponding to the case hole 53 will be described later.

Spacing structure (3) is provided between the first of several geometrical structures (110) between the first geometrical structure and so spaced apart from each other, disposed between the first geometrical structure 110 and the second geometrical structure 120 is a first geometrical structure 110 and the second geometrical structure 120 and spaced apart from each other such that the first geometrical structure 110 is provided between the plate structure 140, a first geometrical structure 110 and the plate structure (140) such that the spaced apart from each other.

The spacing structure (3) is entirely defined heptagon gajina a shape having a recess (3a) is implemented by cutting from the center to leave a series of, separately, a number of stacked plates having a predetermined thickness to a desired thickness or implemented It may be implemented in.

And, separated from the structure (3), case 50, cover 51, the spinnerette 52, the holding member 63, the bolt part (63a), a nut 64, etc. in the first embodiment of the present invention It is the same as described in the description of the longer will be omitted.

The first and second embodiments of the spatial energy eungpok radiation apparatus directly or by creating a type of space by the energy of tosyeonjang geometrical structures, namely a weak area in the space is filled with energy in the energy available to eungpok, It made available indirectly.

Eungpok mechanism of space energy is not been scientifically itjineun clearly identified so far.

However, the Applicant has found a structure that can be most effectively space the radiation energy through a number of repeated tests.

This space energy can be found through various experiments that the presence mothana gajijineun enough to see the intensity of the directly to the human senses.

That is, as compared to the case if it is not irradiated with the irradiation energy to a specific object in the space in the event of any difference, it will be seen that the area with the energy present in the inductive.

Hereinafter, the examples of experiments to determine the presence of space energy will be described.

(1) Cytotoxic tests.

This experiment was carried out in April 2004 by the Yonsei University Wonju Medical School professor and a researcher gimhyeonwon.

In order to carry out this experiment, laboratory rats (C56BL / 6) is prepared by the two rooms to life and, installing a water reservoir contained in the room, the number of treatment processes to examine the energy in the common water space, and the other room (control group), install a water reservoir contained a general drinking water, Jean experiment object was put into each room so that you can freely consume the space energy processing water from the watering can or pail General.

In addition, each room has observed the size of the tumor by the date 20 days into a malignant skin tumor (B16 Melanoma) by subcutaneous injection in the experimental rats (C56BL / 6) number of the same.

In the following graph, control referred displayed part will display the size of tumors in rats who drank plain drinking water by date, marked LA Life energy will display the size of a tumor of the rat drinking can be processed per day 0.20 Space Energy As a result of measuring the size of a tumor during a, in the case of the rats drank the area to be treated with energy in the spatial energy eungpok eungpok emission device as shown in the graph below, the size of the tumor was confirmed that very small compared to the control group.

As identified in the above chart, standing, in the case of the spatial energy drink to be treated rats (Life energy), as compared to normal drinking water in rats (control) was the size of the tumor to determine the smaller, from which space can be energy treatment you can see that the in vivo anti-cancer activity of rats.

That is, it is possible to confirm the presence of the spatial energy from the graph.

(2) the number of treatment processes with zero energy space, affects clinical trials related to the infant atopic patients.

13 is a diagram showing the contents of Table 1 is also graphically, the energy area of ??the subject can be treated prior to taking a view showing, Fig. 14 is a diagram showing the contents of Table 2 as a graph, the processing space can be energy dose Erythema and after shows the number of subjects of the condition, Figure 15 is a diagram showing the contents of Table 2 as a graph, the processing can take place after the energy Edema / Papulation shows the number of subjects of the symptoms.

In addition, Figure 16 is a diagram showing the contents of Table 2 as a graph, the energy treatment space Oozing shows the number of subjects after taking, Figure 17 is a diagram showing the contents of Table 2 as a graph, space energy Excoriation is the treatment shows the number of subjects after taking, Figure 18 is a diagram showing the contents of Table 2 as a graph, the energy treatment space shows the number of subjects after taking Lichenification.

This clinical trial is cheongnoe clinic located in Daechi-dong, Gangnam-gu, Seoul 507 Street mall third floor, plus a win by experiments conducted in early 2005 in the (President ryujaegyu), reflect the space to be treated, spatial processing can be ingested when infants atopic patients Young, a test to verify whether the clinical impact on infant atopic patients.

In this clinical trial use bottled water which is present in the D area of ??energy radiating from the eungpok radiometric survey 10 days (hereinafter, the space energy processing Sura should) the admission was 0-4 years old for the purpose of treating atopic dermatitis in young cheongnoe clinic, infant eczema 40 people were taken to the patient (male 15, female 25).

In this clinical trial, the space zero energy treatment, when the infant atopic patients taking hayeoteul, was set up five hypothesis that there is a difference before taking a space energy processing by measuring the number of atopic index (SCORAD) before and after taking a dose It confirmed the clinical effect.

Here, atopic index (SCORAD), but the significance and the result is checked at the same time to get a full body photo and itching skin conditions and skin part, the subject is young, it is difficult to check the body and itchy skin conditions in infants relations.

Therefore, the main symptoms of atopic five kinds of symptoms (erythema symptoms (Erythema), papules symptoms (Edema / Papulation), gangrene symptoms (Oozing), abrasions symptoms (Excoriation), lichen symptoms (Lichenification)) were classified by the scoring.

Energy treatment may take place prior to the symptoms of atopic patients Table 1 and shown in Fig.

Table 1: Energy treatment can take place before

ErythemaEdema/PapulationOozingExcoriationLichenificationSevere(III)53302Moderate(II)25201511Mild(I)8172510(????)20203436[?1 The contents of the] is represented by a graph as shown in Fig.

Young the space energy can be processed, when the infant atopic patients taking hayeoteul space will be different before and treated the same as taking five kinds of hypothesis to. (A) 10 days Hypothesis 1 space energy processing water 200 ~ 300cc one months Young taking into three days, erythema (Erythema) symptoms of infant atopic patients will have a space energy processing water before taking the difference. (b) Hypothesis 2 days 10 days 1 space energy processing water 200 ~ 300cc three one months will be taking you into the spirit, papules (Edma / Papulation) symptoms of infant atopic patients before taking the space energy processing water is a difference. (c) 3 Hypothesis 3 times a day for 10 days into the space energy processing water 200 ~ 300cc one months taking spirit, gangrene (Oozing) symptoms of infant atopic patients will have a space energy processing water before taking the difference. (d) H4 10 days processing space 200 to zero taking into energy water three times a day for one months 300cc , abrasion (Excoriation) symptoms of infant atopic patients will have a space energy processing water before taking the difference. (e) Hypothesis 5 was 10 days processing space 200 to the 300cc 1 energy water three times a day, taking one months into the spirit, infant eczema lichen (Lichenification) symptoms will have a space energy processing water before taking the difference.

The results of this clinical trial Table 2 are expressed.

[Table 2] after taking the space energy treatment Erythema Edema/Papulation Oozing Excoriation Lichenification Severe (III) 01102 Moderate (II) 1410011 Mild (I) 13121410(????)1317383536(a)

Results for H1 spirit, erythema (Erythema) symptoms of infant atopic patients were not significantly different energy space, the number one days of treatment 10 days after the treatment by taking 200 ~ 300cc before taking into three times.

That is, as shown in Fig. 14, Erythema symptoms after the treatment by taking a spatial energy treatment is seen that the patients belonged to the moderate and severe to mild or move the plurality of 0 (no symptoms), that is a symptom improvement can be. (b) the results for Hypothesis 2 Spirit, the infant atopic patients with papules (Edema / Papulation) significant space, energy treatment symptoms a day for 10 days after treatment by taking 200 ~ 300cc before taking into three There was a difference.

That is 15, the area to be treated by taking the energy treatment after Edema / Papulation symptoms, it can be seen that a large number of patients go to 0 (no symptoms), that the symptoms are improved. (C) English results for Hypothesis 3, gangrene (Oozing) symptoms of infant atopic patients there was a significant difference for the daily 200 ~ 300cc space energy can be processed 10 days after treatment before taking divided doses three times.

That is, as shown in Fig. 16, Oozing symptoms after the treatment by taking a spatial energy to be treated, it can be seen that a large number of patients go to 0 (no symptoms), that the symptoms are improved. (D) Hypothesis Results for the fourth spirit, abrasions (Excoriation) symptoms of infant atopic patients did not differ significantly the number of space energy processing one days to 10 days after treatment by taking 200 ~ 300cc before taking into three times.

That is, as shown in 17, taking as most of the subject is a pre-treatment did not severe symptoms Excoriation before, it was difficult to find a significant result compared to the latter.

But it is skin condition after taking Space fringed energy process can take more than slick was found on the relevant. (E) results for the H5 Spirit of infant atopic patients with lichen (Lichenification) symptoms can space energy processing 10 days one days after treatment by taking 200 ~ 300cc before taking into three There were no significant differences.

Since the duration of the most Lichenification symptoms appear much longer to increase adult manseongjung, 18, the four short duration of less than three zero, infants were no more Lichenification ought baseline symptoms.

As in the above results are confirmed, the symptoms of erythema (Erythema (hypothesis 1)), papules symptoms (Edema / Papulation (hypothesis 2)), gangrene symptoms (Oozing (Hypothesis 3)), abrasions symptoms (Excoriation (H4)) is The improvement was confirmed.

Although, for the lichen symptoms (Lichenification (H5)), but the difference is not, and can confirm the presence of energy in accordance with a result of assumption space 1, 2, 3, 4.

Space radiation energy emitted by the device of the present eungpok activates the vital energy, as identified in the above-mentioned cancer cell growth tests or clinical trials.

However, space energy exerts a variety of effects in a variety of fields as described below.

When the energy emanating from the area emitting device eungpok investigate the plant, causes the growth promotion of the plant, a chemical reaction such corruption has been found that the effect of inhibiting and up.

The space according to the present invention results from the energy-functional it can be seen that there is a non-ionizing effect of inhibiting the ionic polarization.

This phenomenon is occurring in the body as a man, he would immediately foregoing.

Also, when the area of ??energy irradiated three weeks at room temperature 20 ? about the milk, the milk is not connected to corruption, Catering possible and was transformed into the lactic acid bacteria contained in yogurt is utilized also as a health food.

Also, when the area was irradiated energy in two months, strawberries dry dehydrated state is corrupt or not rancidity was changed to Korikori, such as raisins, which is not at all to health when humans consume.

This will begin to deteriorate for more than three days leave to present a way to keep long-term processed food spoilage when compared to that elapses after a few days if kept in a normal environment strawberries.

Also, when the area of energy irradiated apples in the refrigerator for more than two years, due to the evaporation of the skin it became part of the jjogeul jjogeul as well, the meat and the meat inside is not the first time the balance in the freshness of the apple surface.

Thus, the present invention is applied to a case of the human body surrounded by water was found to sikindaneun to preserve or enhance the biological energy.

UA81451
NEDBAI'S GENERATOR OF COMBINED MAGNETIC FIELD

Inventor: NEDBAI VIACHESLAV VIKTOROVYCH [UA]

A generator of the combined magnetic field comprises the therapeutic chamber with the electric coils of the rotational electromagnetic field with independent power supply. The coils are coaxial. In every two adjacent coils, the rotational magnetic fields of the opposite directions are generated simultaneously (clockwise in one coil and counterclockwise in another one). In a gap between the adjacent end-faces of two adjacent coils, the additional scalar magnetic field is generated being directed in parallel to the rotation axis of the magnetic field.

US2013127174

Method for Generating Tidal Energy Utilizing the Scalar Gravitational Potential of Celestial Bodies

Inventor: CUMMING DUNCAN G [US]

Tidal energy generator uses the scalar gravitational potential to utilize energy from celestial bodies. Kinetic energy is extracted from the celestial objects by means of a system including a rotatable arm, a mass fluctuation producing system arranged on the arm and including dielectric material whose mass fluctuates when subjected to a changing electromagnetic field in view of the scalar gravitational potential of the celestial objects, a rotation system that rotates the arm at a variable speed to cause rotation of the mass fluctuation producing system thereon, and a control system that controls the mass fluctuation producing system and rotation system in order to accelerate the dielectric material when it has a relatively light mass and decelerate the dielectric material when it has a relatively heavy mass. Excess energy arising from an energy differential between the acceleration and deceleration of the dielectric material is directed to a load.

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

Top Ten Transhumanist Technologies

by Lifeboat Foundation Scientific Advisory Board member Michael Anissimov.

Overview

Transhumanists advocate the improvement of human capacities through advanced technology. Not just technology as in gadgets you get from Best Buy, but technology in the grander sense of strategies for eliminating disease, providing cheap but high-quality products to the world's poorest, improving quality of life and social interconnectedness, and so on. Technology we don't notice because it's blended in with the fabric of the world, but would immediately take note of its absence if it became unavailable. (Ever tried to travel to another country on foot?) Technology needn't be expensive — indeed, if a technology is truly effective it will pay for itself many times over.

Transhumanists tend to take a longer-than-average view of technological progress, looking not just five or ten years into the future but twenty years, thirty years, and beyond. We realize that the longer you look forward, the more uncertain the predictions get, but one thing is quite certain: if a technology is physically possible and obviously useful, human (or transhuman!) ingenuity will see to it that it gets built eventually.

As we gain ever greater control over the atomic structure of matter, our technological goals become increasingly ambitious, and their payoffs more and more generous. Sometimes new technologies even make us happier in a long-lasting way: the Internet would be a prime example. In the following list I take a look at what I consider the top ten transhumanist technologies.

The List

10. Cryonics

ALCOR
LIFE EXTENSION FOUNDATION
SINCE 1972
www.alcor.org

Cryonics is the high-fidelity preservation of the human body, and particularly the brain, after what we would call death, in anticipation of possible future revival. Cryonics is an important transhumanist technology not only because it is already available today, but because the technology is relatively mature — we can reliably stop cells from decaying. In vitrification, the brain is not frozen in the conventional manner but with a cryoprotectant (antifreeze) mixture, which effectively prevents the formation of crystals, causing the water to freeze smoothly, like glass.

Maintenance of a cryo-patient is not difficult — it requires no electricity, but merely the replenishment of liquid nitrogen about every three weeks. As cryonics becomes more popular, this process could become automated and extremely reliable. Further improvements in dewar technology will continue to increase safety and reduce costs. The Cryonics Institute in Michigan, for example, has operated since 1976 without a single mishap.

Financed by the interest of the payout of a life insurance policy (which for people under 40 may cost as little as $100 a year to own), patients can be securely cryopreserved for as long as the cryonics company stays afloat and the dewar stays in one piece. Eventual revival does not require the technology to become available tomorrow, or next year… as long as the liquid nitrogen keeps replenished, you can stay on ice for as long as it takes.

For an existence proof of cryonic revival, there are frogs that can freeze solid and revive later, though reviving a human from freezing would likely require molecular nanotechnology (MNT). When we will be able to revive a cryo-patient will be strongly related to when we develop sophisticated MNT. Once we do develop MNT, the prospect of successful revival is extremely likely — it would involve slowly melting the ice and rebooting the metabolism by kickstarting the appropriate chemical reactions within cells.

9. Virtual Reality

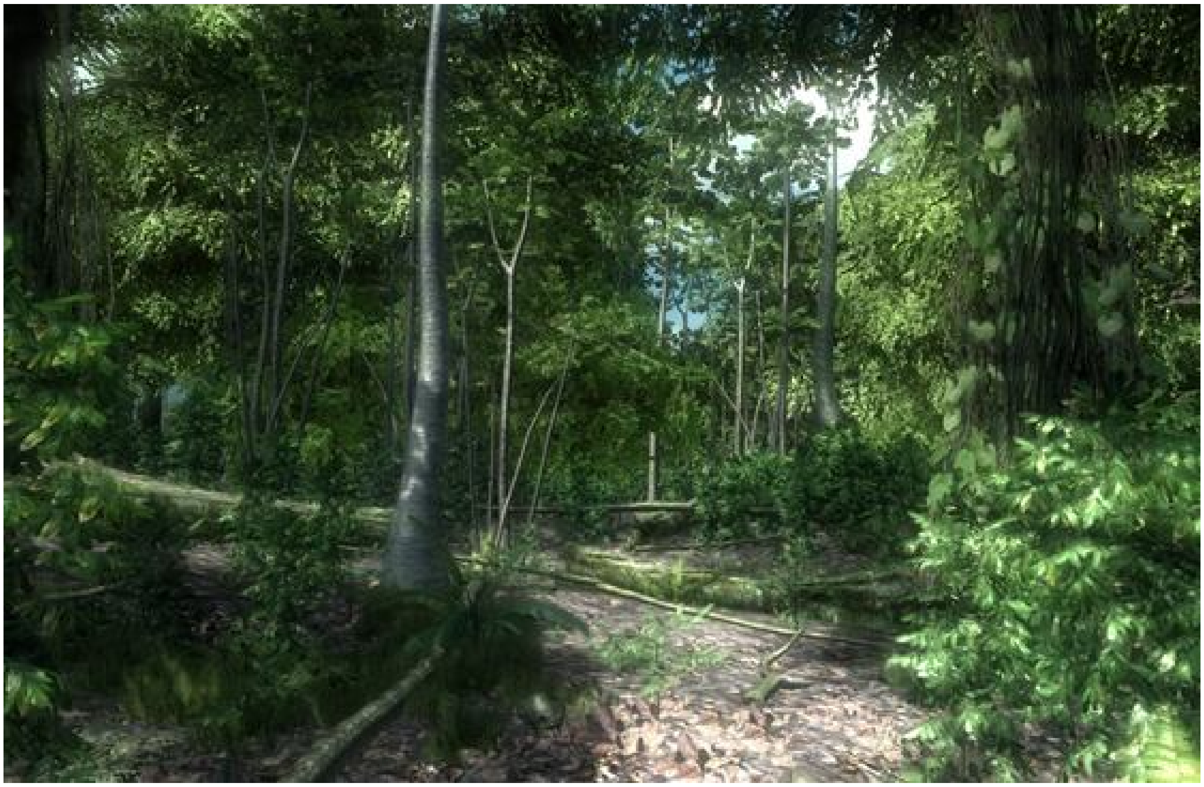

The above image may look like a photo, but it’s actually a screenshot from the game Crysis, a first-person shooter which will be released later this year. Look at screenshots from the game and you’ll see that computer graphics are already beginning to approach photorealism. Sometime in the 2020s, reality simulations will become so high-resolution and immersive that they'll start to get indistinguishable from the real thing.

Simulations will become the preferred environments for work and play. Pretty soon the main obstacle to truly immersive VR will not be the visuals but the haptics — our sense of touch. To fool our senses into believing haptic technologies are conveying the real thing, the "frame rate" needs to be significantly higher than for visual technologies, a few hundred updates per second rather than a few dozen — which is why development could take another decade or two. But many millions of dollars are currently going into efforts to develop advanced VR.

Clearly, World of Warcraft's eight million subscribers and Second Life's five million subscribers are onto something. At least 1% of all broadband Internet users play in virtual worlds, and this number is increasing rapidly. These worlds typically outclass the real world in terms of customizability, but still have yet to catch up in terms of sensory richness or social fulfillment. But it's only a matter of time.

In the mid-to-late 2020s, I expect full-body, high quality haptic VR suits to be affordable to the average person in developed countries, obtained either from your local Wal-Mart or perhaps printed right out of a desktop nanofactory after payment of a fee. For more on this, here is one scientific paper, "Towards full-body haptic feedback".

8. Gene Therapy/RNA Interference

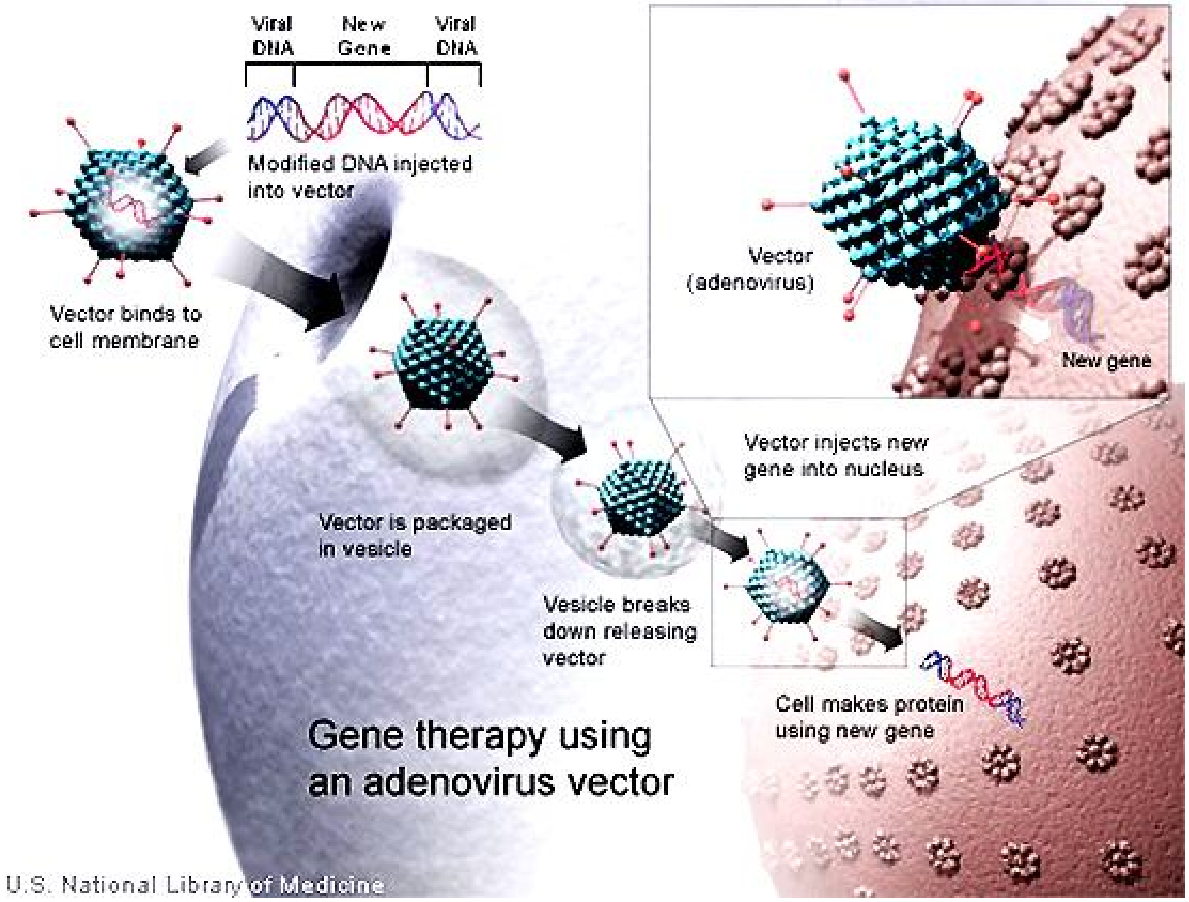
Viral DNA
New Gene
Viral DNA
Modified DNA injected into vector
Vector binds to cell membrane
Vector (adenovirus)
New gene
Vector injects new gene into nucleus
Vector is packaged in vesicle
Vesicle breaks down releasing vector
Cell makes protein using new gene
Gene therapy using an adenovirus vector
U.S. National Library of Medicine

Gene therapy replaces bad genes with good genes, and RNA interference can selectively knock out gene expression. Together, they give us an unprecedented ability to manipulate our own genetic code. By knocking out genes that code for certain metabolic proteins, scientists have been able to make mice that stay slim no matter how much junk food they eat. Lou Gehrig's disease has been cured in mice, and it could only be a few years before we develop a therapy that can cure it for humans too.

Aubrey de Grey's SENS (Strategies for Engineered Negligible Senescence) research program contains various prescriptions for the use of gene therapy. Within a couple decades or so, progress in anti-aging therapies will improve to the point where we are gaining more than an extra year of lifespan per year, reaching so-called "longevity escape velocity" eventually culminating in indefinite lifespans.

Like many transhumanist technologies, gene therapy is really exciting because it's just beginning. No scientist has yet performed gene therapy on germline cells (sexual cells in the gonads) due to the ethical controversy of producing genetic changes which are heritable, but, as with many of these things, it's only a matter of time. Regulations in any given country will only be capable of slowing the overall progress of the field by a few years at most.

The money will go where the research is permitted. In its mature form, gene therapy and genetic engineering will become extremely cheap and powerful, letting humans live comfortably in a wider range of environments and gain immunity to most, if not all diseases. Supercomputers of the future, with thousands or millions of times the crunch power of today's best, will let us simulate the changes in extreme detail before we attempt them with actual human beings. This will make ill side effects quite unlikely for the typical case, much to the dismay of the authors of "genetic engineering turned daddy into a bloodthirsty zombie!" trash novels and films.

7. Space Colonization

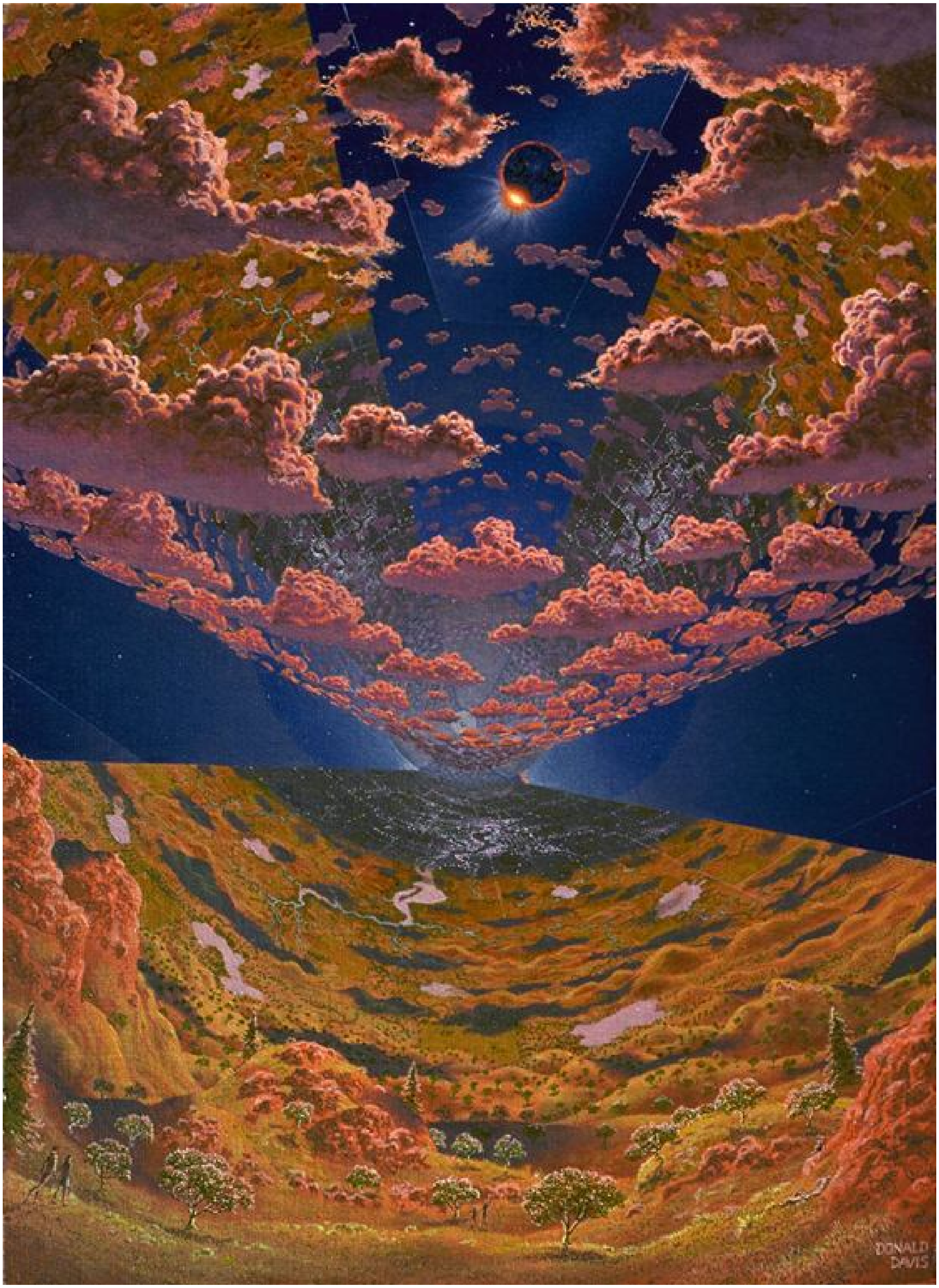

Space colonies will become necessary to house the many billions of individuals that will be born in the future as our population continues to expand at a lazy exponential. In his book, *The Millennial Project*, Marshall T. Savage estimates that the Asteroid Belt could hold 7,500 *trillion* people, if thoroughly reshaped into O'Neill colonies. At a typical population growth rate for developed countries at 1% per annum (doubling every 72 years), it would take us 1,440 years to fill that space. Siphoning light gases off Jupiter and Saturn and fusing them into heavier elements for construction of further colonies seems plausible in the longer term as well.

Why expand into space? For many, the answers are blatantly obvious, but the easiest is that the alternatives are limiting the human freedom to reproduce, or mass murder, both of which are morally unacceptable. Population growth is not inherently antithetical to a love of the environment — in fact, by expanding outwards into the cosmos in all directions, we'll be able to seed every star system with every species of plant and animal imaginable. The genetic diversity of the embryonic home planet will seem tiny by comparison.

Space colonization is closely related to transhumanism through the mutual association of futurist philosophy, but also more directly because the embrace of transhumanism will be necessary to colonize space. Human beings aren't designed to live in space. Our physiological issues with it are manifold, from deteriorating muscle mass to uncontrollable flatulence. On the surface of Venus, we would melt, on the surface of Mars, we'd freeze. The only reasonable solution is to upgrade our bodies. Not terraform the cosmos, but cosmosform ourselves.

6. Cybernetics

Can you spot the cyborg in this picture? You're looking right at him! It's Michael Chorost, the man who was born almost deaf but now can hear, thanks to a cochlear implant. Most of the cyborgs in fiction fit certain stereotypes — bermensch wannabes, cyborg assassins, and supercops. But cyborgs already walk among us, and they look just like normal people.

This trend will continue in the future. Many cyborg upgrades which will become available in the 20s and 30s, such as hearing and vision enhancement, metabolic enhancement, artificial bones, muscles, and organs, and even brain-computer interfaces will be invisible to the casual observer, implanted beneath the skin. Cybernetic features on the surface, such as dermal enhancements or technological actuators like retractable wings, will be carefully camouflaged. No one will want to shock the rest of society by looking like the tin man in public.

The process of cyborgization has already been happening for centuries if not millennia, since the advent of clothing and piercings. For many generations, but especially in the last couple decades, our technological gadgets have been getting smaller, more functional, and more closely integrated with our natural activity.

Recently, Microsoft announced Microsoft Surface, a mouseless, keyboardless form of desktop computing which takes input from finger tracing and hand gestures. The sophistication of biotechnology and the availability of better materials and precision manufacturing will let us make systems so small and effective that even everyday people elect to implant them.

These cybernetic systems will greatly improve our everyday experience, from letting us hear a wider range of ambient sounds, to viewing millions of stars rather than just a few thousand, to making us more resistant to accidents. They will improve the overall economy by enabling us do more work in less time for better pay. In the long term, enhanced humans may get a bigger portion of the economic pie than un-augmented humans, but the pie itself will become so much larger than even the poorest humans of tomorrow will be better off than the wealthiest of today.

Here’s a good cyborg blog I found while doing research for this article, and the Power Jacket, a 4-pound jacket that enhances strength and is used by people recovering from paralysis. For more, see the cybernetics category of my del.ic.ious links, or my top ten list of cybernetic enhancements.

5. Autonomous Self-Replicating Robotics

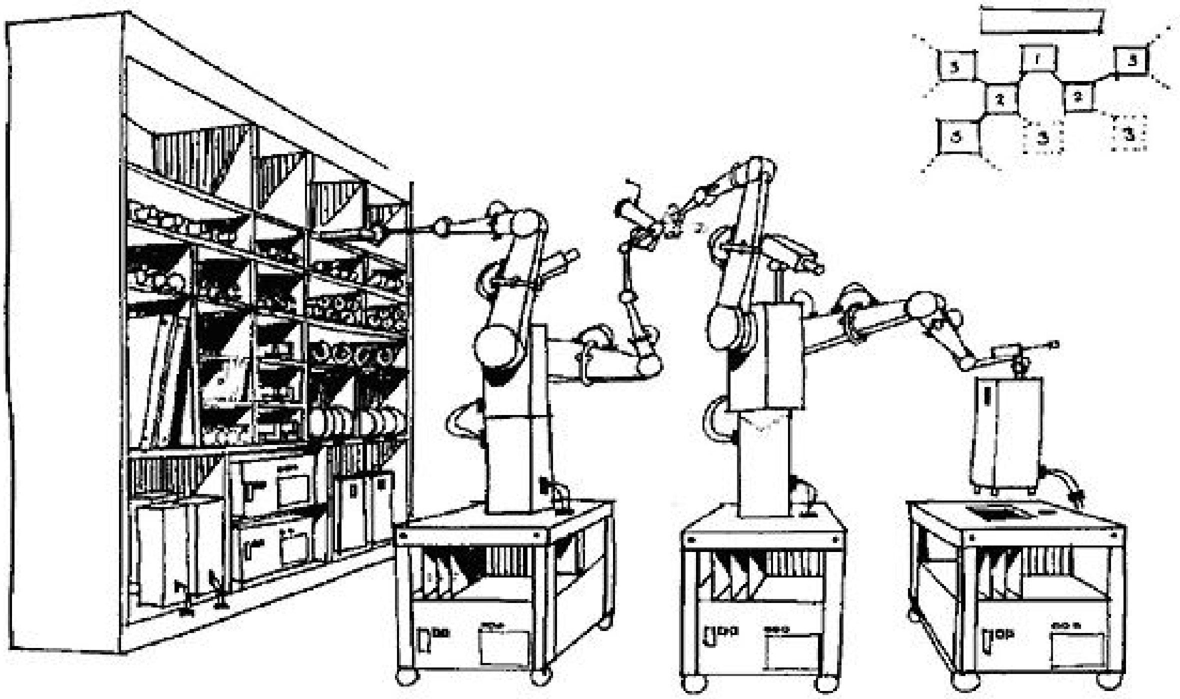

Why do manual labor when the robots can do it for you? Self-replication might be considered the Holy Grail of robotics. A landmark NASA study, “Advanced Automation for Space Missions”, found that robotic self-replication is just a matter of engineering, and that no fundamental theoretical breakthroughs are needed. The study proposed sending a 100-ton package to the Moon, with a self-replication time of 1 year, and letting it self-replicate until the desired level of development is attained.

The design — which was fleshed out in great detail — was based on electric carts running on rails within the factory, “paving machines” that direct sunlight to melt lunar regolith, robotic strip miners for obtaining raw materials, and a solar cell “canopy” for powering it all. After 10 years, over 100,000 tons of lunar factory could be produced autonomously. The factory’s functions could then be hijacked for the benefit of human colonists, used to produce housing, products, and provide large quantities of solar power.

If similar self-replicating systems could be constructed on Earth, there would be little limit to the material plenty they could provide. Self-replicating factories could turn the vast empty badlands of Australia into lush gardens by pumping water from the oceans, self-replicating factories in the high Arctic could melt snow and create gigantic transparent domes suitable for habitation, and submersible automata in the seas could dredge sand from abiotic regions of the ocean floor and process it into gigantic platforms for human colonization.

By opening up such vast new regions of the Earth’s surface, talk of overpopulation and crowding would fall by the wayside for quite a few decades, with people realizing how much space there actually was all along. And once things really do get too crowded here on Earth, we can move to the Moon, Mars, and the asteroid belt, using the power of self-replicating robotics to create rotating space colonies suitable for housing trillions of people.

Self-replicating factories could reduce the costs of material goods close to that of food — the primary expenses would consist of raw materials, energy, and whatever small quantity of human oversight is necessary to keep an eye on the overall structure of things. By utilizing special, man-made “nutrients” for top-level functions (rare or exotic molecules such as custom-synthesized proteins) and the broadcast architecture — whereby derivative factories must receive affirmations from a central parent factory to continue self-replicating — such factories could be made safe by design.

With such abundance, humanity might actually shift from having a zero-sum perspective on a world to a positive-sum perspective. With medical tools and basic goods in ample supply, no one in the world would need to suffer from poverty or curable disease. The nature of human work would shift from manual drudgery and mind-numbing routine to more creative and personally fulfilling endeavors, like art, music, math, science, literature, and exploration.

For more details on the state of the art in self-replicating machines, see the Wikipedia entry, or the magnum opus on the topic, *Kinematic Self-Replicating Machines*.

4. Molecular Manufacturing

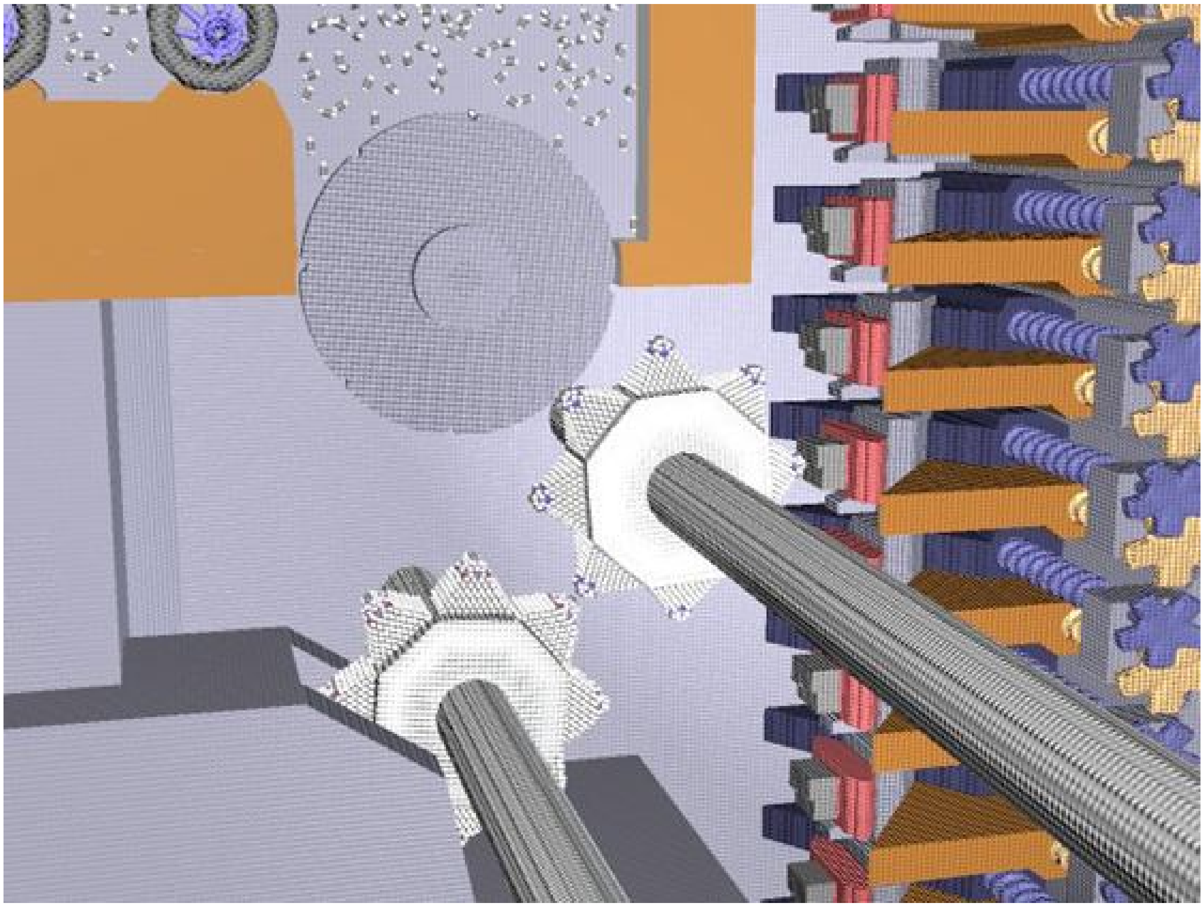

Nanofactory in action. Courtesy Lizard Fire Studios.

If self-replication is the Holy Grail of robotics, then molecular nanotechnology (MNT) is the Holy Grail of manufacturing. Molecular nanotechnology would use massive arrays of nanometer-scale actuators (produced initially through self-replication) to manufacture macroscale products with atomic precision. This concept is known as the nanofactory.

In practical terms, the creation of nanofactories would mean that practically everything could be made out of diamond, motors would become so powerful that a cubic centimeter would provide enough torque to propel a car, medical nanodevices could heal wounds and repair organs without the need for surgery, and air-suspended nanodevices ("utility fog") could be configured to simulate practically any desired object on demand.

On the downside of things, it could become easy to manufacture mite-sized robots with a payload of poison sufficient to kill thousands, or a laptop-sized device capable of separating U-235 from U-238 in a worrisomely simple and rapid fashion, or self-replicating synthetic algae capable of clogging up our oceans with grey goo.

Enabling widespread use of the positive applications while cleanly and completely suppressing the nasty applications is a first-order challenge. Incidentally, you can make a difference right now by donating to the Lifeboat Foundation or Center for Responsible Nanotechnology, two of very few organizations focusing on this area.

To some, molecular nanotechnology sounds like science fiction, and based on the grandiose applications I discussed in the previous paragraph, you can't blame them. But many of the prerequisites of molecular manufacturing have already been demonstrated — "molecular surgery" has been used to snip off and replace individual hydrogen atoms, various functional nanoscale devices have been built, scanning tunneling microscopy has been used to mechanically manipulate individual atoms, and so on. The challenge is to create a nanoscale manipulator arm capable of placing individual atoms with angstrom-level precision, avoiding undesired reactions, and serving as a universal constructor that can build a copy of itself.

There are numerous technical challenges still outstanding, but when these are overcome, manufacturing will be granted the power that nature has had for hundreds of millions of years — the ability to fabricate large objects with molecular precision. The numerous potential applications of the technology to human enhancement are obvious; with molecular manufacturing, we could orchestrate elegant improvements to every single body component, achieving all of the upgrades described on my top ten list, and many more.

3. Megascale Engineering

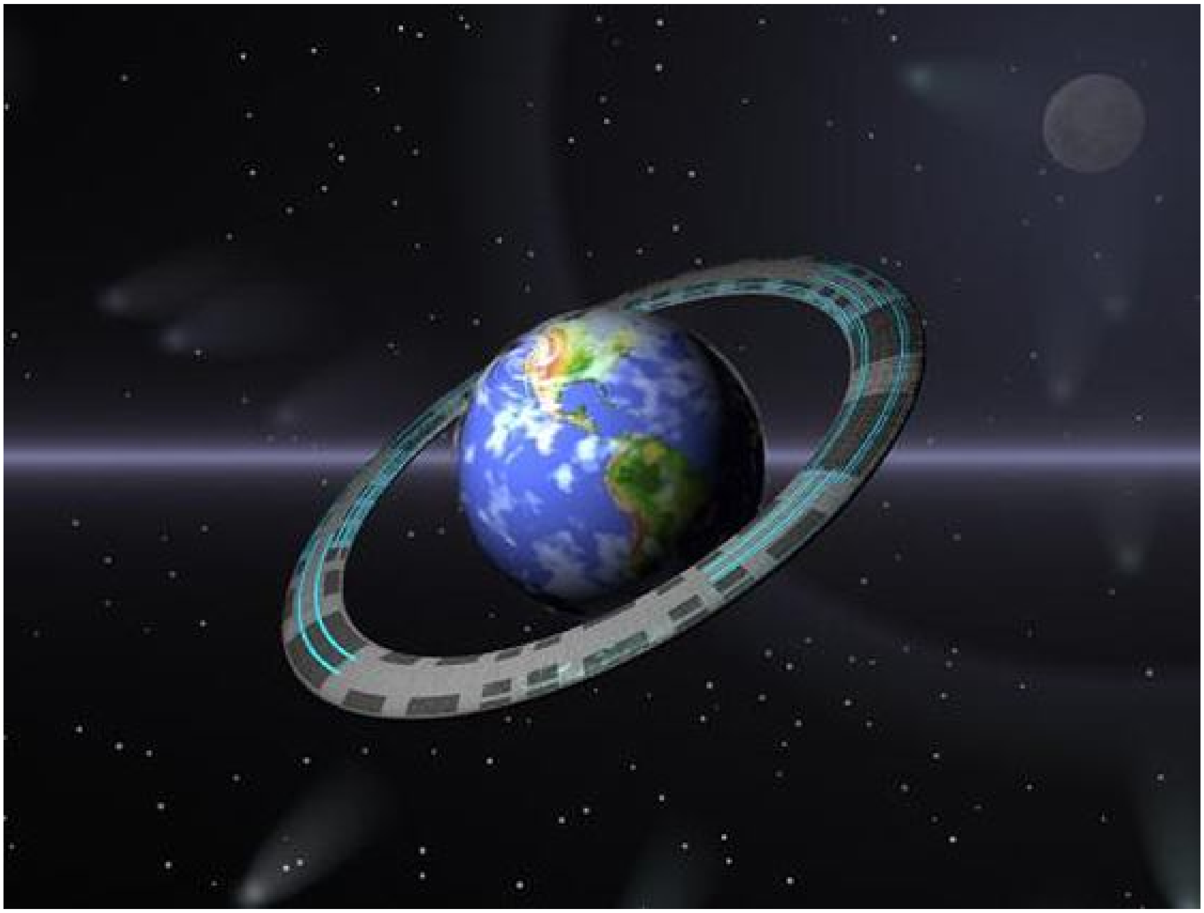

Most people are familiar with megascale engineering because it is seen throughout fiction — the Death Star, for instance. Typically, megascale engineering refers to building structures at least 1,000 km in length in one dimension, such as a space elevator, Globus Cassus, or Dyson sphere. With the self-replicating robotics described above, the production of such large structures could be done largely by autonomous drones, with intelligent agents only managing the highest top-level functions and architecture.

Considering that mankind's long-term future is in space, and that space right now is pretty devoid of any structure useful or habitable to humans, we have a lot of work to do, and if you can make the projects megascale, why not?

Like some of the other items on this list, megascale engineering is only indirectly transhumanist — but is still very relevant to the long-term future of intelligent life. Megascale engineering goes hand-in-hand with the grandiose transhumanist vision: intelligent beings spreading across the cosmos, and eventually shaping the very structure of the universe itself.

The fact that these vast expanses of colonizable space are currently neglected imposes on us a vast opportunity cost — if we hurried up a bit and colonized them, we could give rise to tremendous numbers of people leading worthwhile lives. What experiences would they have, and what stories would they tell? We'll never find out, unless we make it happen.

2. Mind Uploading

Mind uploading, sometimes referred to as nonbiological intelligence, centers around the controversial proposition that cognitive processing can be implemented on substrates other than our current neurons. Considering decades of successful results in neurophysiology, and the recent construction of the world's first brain prosthesis — an artificial copy of the hippocampus — this seems very likely. It appears that our minds are defined more by the information pattern they embody than the particular hardware they are implemented on.

Numerous philosophers of mind have long recognized this, but acceptance among the wider public has been a long time in coming: people don't want to think that they're "just" data structures being implemented as computational automata on biological neurons. But it is hard to think of it any other way: once we dismiss the possibility of an immaterial soul, we must acknowledge the mind as a material pattern implemented in physical configurations, and if other substances aside from our current neurons can meet the requirements for these configurations, then there is no reason why intelligence and consciousness could not exist on another substrate. For a humorous look at this complex philosophical argument, see "They're Made Out of Meat" by Terry Bisson.

If our brains really don't have to be made out of meat, then we can transfer them to other substrates. By incrementally replacing each neuron with a synthetic neuron-equivalent, the whole process could go down painlessly and seamlessly. The transfer could be as slow or as fast as we want: from the information-processing perspective of the brain itself, nothing ever changes. Light still comes in through the eye's lens, hits the retina, is transformed into nerve impulses which travel down the optic nerve, receives further processing in the visual cortex at the back of the brain, the highlights of which are sent to the prefrontal cortex for integration with information from the other senses.

The brain can't tell if it's made out of traditional meat, or accelerated biological neurons, or entirely nonbiological neuron-equivalents: the computation is the same. Sometimes this notion is also referred to as an application of the Church-Turing thesis.

If entirely synthetic brains are possible, then there's nothing stopping such persons from inhabiting computer networks — not indirectly, sitting in chairs as we currently do, but directly, engaging in computer worlds as a sentient program of tremendous complexity. With molecular manufacturing on hand, reversing the process would be as simple as printing out a hundred or so kilograms of flesh and bone again, complete with memories from the networked experience.

This is probably among the transhumanist visions that most reliably elicits the "yuck!" reaction, but if functionalism is true, then virtual experience will be indistinguishable from physical experience. Not only that, but even more enjoyable, due to the manifold degrees of freedom which would become newly accessible. In a virtual world, there are no laws of physics except those we choose.

For a bit more on simulations, see this primer, and remember: be careful not to generalize from fictional evidence.

1. Artificial General Intelligence (AGI)

As argued in the previous section, functionalism seems likely. If so, then strong AI is possible. Thinking, feeling, imagining, creating, communicating, thoughtful synthetic intelligences with conscious experiences. Whether serial computing is sufficient, or parallel computing is necessary, both are within technological reach, and present-day computing speeds are fast approaching the computing power of the human brain.

In fact, according to many estimates, the fastest present-day supercomputer, Blue Gene/P, has already exceeded it. Blue Gene/P operates continuously at speeds of over a petaflop, which is a million billion operations per second. For strong AI skeptics, no computer — even one operating at trillions of trillions of trillions of operations per second, is sufficient to implement true intelligence, but to functionalists like myself, such a meat-centric perspective is unjustified.

Distinct from artificial intelligence in general, which has come to refer to any sophisticated software program, artificial general intelligence refers to AIs that display open-ended learning and similar competency levels to human beings.

A handful of researchers are working diligently towards artificial general intelligence, informed by the mathematics of inference and probability theory: Jürgen Schmidhuber, whose “main scientific ambition has been to build an optimal scientist, then retire”; Marcus Hutter, author of the landmark book *Universal Artificial Intelligence*; Ben Goertzel, who recently presented his AI design in a talk to Google; and Eliezer Yudkowsky, who is developing a reflective decision theory from first principles. Whether or not others believe in the feasibility of general AI, these individuals will keep working, and one will eventually succeed.

The way the world would be impacted by the arrival of general AI is too extreme to discuss in much detail here. If raw materials such as sand can be converted into computer chips and then into intelligent minds, eventually the majority of material in the solar system could be made intelligent and conscious. The result would be a “noetic Renaissance”: the expansion of intelligence and experience beyond our wildest dreams.

Conversely, if not given empathic values, artificial intelligence could lead to the doom of all. It’s up to us to set the initial conditions appropriately: if not, we might not be around to regret it.

rexresearch.com

Subliminal Suggestion & Mind Control Patents

500W
200 - 420V$_{IN}$
93% peak efficiency
Isolated regulated
DC-DC converter
47.9 x 22.8 x 7.2mm
VICOR
FREE EVAL
Ad

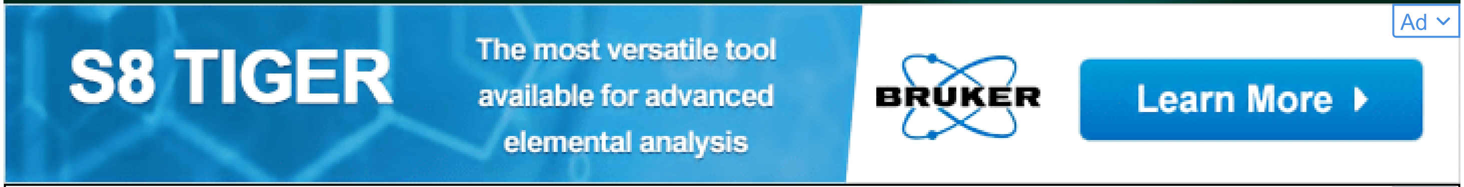

S8 TIGER
The most versatile tool available for advanced elemental analysis
BRUKER
Learn More
Ad

Ad ✕

USP # 6,506,148 (January 14, 2003)
Nervous System Manipulation by EM Fields from Monitors
Loos, Hendricus

Abstract ---Physiological effects have been observed in a human subject in response to stimulation of the skin with weak electromagnetic fields that are pulsed with certain frequencies near 1/2 Hz or 2.4 Hz, such as to excite a sensory resonance. Many computer monitors and TV tubes, when displaying pulsed images, emit pulsed electromagnetic fields of sufficient amplitudes to cause such excitation. It is therefore possible to manipulate the nervous system of a subject by pulsing images displayed on a nearby computer monitor or TV set. For the latter, the image pulsing may be imbedded in the program material, or it may be overlaid by modulating a video stream, either as an RF signal or as a video signal. The image displayed on a computer monitor may be pulsed effectively by a simple computer program. For certain monitors, pulsed electromagnetic fields capable of exciting sensory resonances in nearby subjects may be generated even as the displayed images are pulsed with subliminal intensity.

USP # 6,488,617 (December 3, 2002)
Method and Device for Producing a Desired Brain State
Katz, Bruce

Abstract --- A method and device for the production of a desired brain state in an individual contain means for monitoring and analyzing the brain state while a set of one or more magnets produce fields that alter this state. A computational system alters various parameters of the magnetic fields in order to close the gap between the actual and desired brain state. This feedback process operates continuously until the gap is minimized and/or removed.

USP # 6,487,531 (November 26, 2002)
Signal Injection Coupling into the Human Vocal Tract...
Tosaya, Carol
Abstract --- A means and method are provided for enhancing or replacing the natural excitation of the human vocal tract by artificial excitation means, wherein the artificially created acoustics present additional spectral, temporal, or phase data useful for (1) enhancing the machine recognition robustness of audible speech or (2) enabling more robust machine-recognition of relatively inaudible mouthed or whispered speech. The artificial excitation (a) may be arranged to be audible or inaudible, (b) may be designed to be non-interfering with another user's similar means, (c) may be used in one or both of a vocal content-enhancement mode or a complimentary vocal tract-probing mode, and/or (d) may be used for the recognition of audible or inaudible continuous speech or isolated spoken commands.

USP # 6,430,443 (August 6, 2002)
Method and Apparatus for Treating Auditory Hallucinations
Karell, Manuel
Abstract --- Stimulating one or more vestibulocochlear nerves or cochlea or cochlear regions will treat, prevent and control auditory hallucinations.

USP # 6,426,919 (July 30, 2002)
Portable and Hand-Held Device for Making Humanly Audible Sounds...
Gerosa, William
Abstract --- A portable and hand-held device for making humanly audible sounds responsive to the detecting of ultrasonic sounds. The device includes a hand-held housing and circuitry that is contained in the housing. The circuitry includes a microphone that receives the ultrasonic sound, a first low voltage audio power amplifier that strengthens the signal from the microphone, a second low voltage audio power amplifier that further strengthens the signal from the first low voltage audio power amplifier, a 7-stage ripple carry binary counter that lowers the frequency of the signal from the second low voltage audio power amplifier so as to be humanly audible, a third low voltage audio power amplifier that strengthens the signal from the 7-stage ripple carry binary counter, and a speaker that generates a humanly audible sound from the third low voltage audio power amplifier.

USP # 6,292,688 (September 18, 2001)
Method and Apparatus for Analyzing Neurological Response to Emotion-Inducing Stimuli
Patton, Richard
Abstract --- A method of determining the extent of the emotional response of a test subject to stimului having a time-varying visual content, for example, an advertising presentation. The test subject is positioned to observe the presentation for a given duration, and a path of communication is established between the subject and a brain wave detector/analyzer. The intensity component of each of at least two different brain wave frequencies is measured during the exposure, and each frequency is associated with a particular emotion. While the subject views the presentation, periodic variations in the intensity component of the brain waves of each of the particular frequencies selected is measured. The change rates in the intensity at regular periods during the duration are also measured. The intensity change rates are then used to construct a graph of plural coordinate points, and these coordinate points graphically establish the composite emotional reaction of the subject as the presentation continues.

USP # 6,258,022 (July 10,2001)
Behavior Modification
Rose, John
Abstract --- Behavior modification of a human subject takes place under hypnosis, when the

subject is in a relaxed state. A machine plays back a video or audio recording, during which the subject is instructed to activate a device to create a perceptible stimulation which is linked, through the hypnosis, with a visualization of enhanced or improved performance. After the hypnosis, the user can reactivate the device at will, whenever the improved performance, such as an improved sporting performance, is desired. This will again create the perceptible stimulation and thus induce the required visualization.

USP # 6,239,705 (May 29,2001)
Intra-Oral Electronic Tracking Device
Glen, Jeffrey
Abstract --- An improved stealthy, non-surgical, biocompatable electronic tracking device is provided in which a housing is placed intraorally. The housing contains microcircuitry. The microcircuitry comprises a receiver, a passive mode to active mode activator, a signal decoder for determining positional fix, a transmitter, an antenna, and a power supply. Optionally, an amplifier may be utilized to boost signal strength. The power supply energizes the receiver. Upon receiving a coded activating signal, the positional fix signal decoder is energized, determining a positional fix. The transmitter subsequently transmits through the antenna a position locating signal to be received by a remote locator. In another embodiment of the present invention, the microcircuitry comprises a receiver, a passive mode to active mode activator, a transmitter, an antenna and a power supply. Optionally, an amplifier may be utilized to boost signal strength. The power supply energizes the receiver. Upon receiving a coded activating signal, the transmitter is energized. The transmitter subsequently transmits through the antenna a homing signal to be received by a remote locator.

USP # 6,167,304 (December 26, 2000)
Pulse Variability in Electric Field Manipulation of Nervous Systems
Loos, Hendricus
Abstract --- Apparatus and method for manipulating the nervous system of a subject by applying to the skin a pulsing external electric field which, although too weak to cause classical nerve stimulation, modulates the normal spontaneous spiking patterns of certain kinds of afferent nerves. For certain pulse frequencies the electric field stimulation can excite in the nervous system resonances with observable physiological consequences. Pulse variability is introduced for the purpose of thwarting habituation of the nervous system to the repetitive stimulation, or to alleviate the need for precise tuning to a resonance frequency, or to control pathological oscillatory neural activities such as tremors or seizures. Pulse generators with stochastic and deterministic pulse variability are disclosed, and the output of an effective generator of the latter type is characterized.

USP # 6,135,944 (October 24, 2000)
Method of Inducing Harmonious States of Being
Bowman, Gerard D., *et al.*
Abstract --- A method of inducing harmonious states of being using vibrational stimuli, preferably sound, comprised of a multitude of frequencies expressing a specific pattern of relationship. Two base signals are modulated by a set of ratios to generate a plurality of harmonics. The harmonics are combined to form a "fractal" arrangement.

USP # 6,122,322 (September 19, 2000)
Subliminal Message Protection
Jandel, Magnus
Abstract --- The present invention relates to a method and to a system for detecting a first context change between two frames. When a second context change between a further two frames occurs within a predetermined time interval, the frames accommodated within the two context changes are defined as a subliminal message. An alarm is sent to an observer upon detection of a subliminal message.

USP # 6,091,994 (July 18, 2000)
Pulsative Manipulation of Nervous Systems

Loos, Hendricus

Abstract --- Method and apparatus for manipulating the nervous system by imparting subliminal pulsative cooling to the subject's skin at a frequency that is suitable for the excitation of a sensory resonance. At present, two major sensory resonances are known, with frequencies near 1/2 Hz and 2.4 Hz. The 1/2 Hz sensory resonance causes relaxation, sleepiness, ptosis of the eyelids, a tonic smile, a "knot" in the stomach, or sexual excitement, depending on the precise frequency used. The 2.4 Hz resonance causes the slowing of certain cortical activities, and is characterized by a large increase of the time needed to silently count backward from 100 to 60, with the eyes closed. The invention can be used by the general public for inducing relaxation, sleep, or sexual excitement, and clinically for the control and perhaps a treatment of tremors, seizures, and autonomic system disorders such as panic attacks. Embodiments shown are a pulsed fan to impart subliminal cooling pulses to the subject's skin, and a silent device which induces periodically varying flow past the subject's skin, the flow being induced by pulsative rising warm air plumes that are caused by a thin resistive wire which is periodically heated by electric current pulses.

USP # 6,081,744 (June 27, 2000)
Electric Fringe Field Generator for Manipulating Nervous Systems
Loos, Hendricus

Abstract --- Apparatus and method for manipulating the nervous system of a subject through afferent nerves, modulated by externally applied weak fluctuating electric fields, tuned to certain frequencies such as to excite a resonance in neural circuits. Depending on the frequency chosen, excitation of such resonances causes in a human subject relaxation, sleepiness, sexual excitement, or the slowing of certain cortical processes. The electric field used for stimulation of the subject is induced by a pair of field electrodes charged to opposite polarity and placed such that the subject is entirely outside the space between the field electrodes. Such configuration allows for very compact devices where the field electrodes and a battery-powered voltage generator are contained in a small casing, such as a powder box. The stimulation by the weak external electric field relies on frequency modulation of spontaneous spiking patterns of afferent nerves. The method and apparatus can be used by the general public as an aid to relaxation, sleep, or arousal, and clinically for the control and perhaps the treatment of tremors and seizures, and disorders of the autonomic nervous system, such as panic attacks.

USP # 6,052,336 (April 18, 2000)
Apparatus and Method of Broadcasting Audible Sound Using Ultrasonic Sound as a Carrier
Lowrey, Austin, III

Abstract --- An ultrasonic sound source broadcasts an ultrasonic signal which is amplitude and/or frequency modulated with an information input signal originating from an information input source. If the signals are amplitude modulated, a square root function of the information input signal is produced prior to modulation. The modulated signal, which may be amplified, is then broadcast via a projector unit, whereupon an individual or group of individuals located in the broadcast region detect the audible sound.

USP # 6,039,688 (March 21, 2000)
Therapeutic Behavior Modification Program, Compliance Monitoring and Feedback System
Douglas, Peter, et al.

Abstract --- A therapeutic behavior modification program, compliance monitoring and feedback system includes a server-based relational database and one or more microprocessors electronically coupled to the server. The system enables development of a therapeutic behavior modification program having a series of milestones for an individual to achieve lifestyle changes necessary to maintain his or her health or recover from ailments or medical procedures. The program may be modified by a physician or trained case advisor prior to implementation. The system monitors the individual's compliance with the program by prompting the individual to enter health-related data, correlating the individual's entered data with the milestones in the behavior modification program and generating compliance data indicative of the individual's progress toward achievement of the program milestones. The system also includes an integrated system of graphical system interfaces for motivating the individual to comply with the program. Through the interfaces, the individual can access the database to review the compliance data and obtain health information from a remote source such as selected sites on the Internet. The system also provides an electronic calendar integrated with the behavior modification program for signaling the individual to take action pursuant to the behavior modification program in which the calendar accesses the relational database and integrates requirements of the program with the individual's daily schedule, and an electronic journal for enabling the individual to enter personal health-related information into the system on a regular basis. In addition, the system includes an electronic meeting room for linking the individual to a plurality of other individuals having related behavior modification programs for facilitating group peer support sessions for compliance with the program. The system enables motivational media presentations to be made to the individuals in the electronic meeting room as part of the group support session to facilitate interactive group discussion about the presentations. The entire system is designed around a community of support motif including a graphical electronic navigator operable by the individual to control the microprocessor for accessing different parts of the system.

USP # 6,017,302 (January 25, 2000)
Subliminal Acoustic Manipulation of Nervous Systems
Loos, Hendricus
Abstract --- In human subjects, sensory resonances can be excited by subliminal atmospheric acoustic pulses that are tuned to the resonance frequency. The 1/2 Hz sensory resonance affects the autonomic nervous system and may cause relaxation, drowsiness, or sexual excitement, depending on the precise acoustic frequency near 1/2 Hz used. The effects of the 2.5 Hz resonance include slowing of certain cortical processes, sleepiness, and disorientation. For these effects to occur, the acoustic intensity must lie in a certain deeply subliminal range. Suitable apparatus consists of a portable battery-powered source of weak subaudio acoustic radiation. The method and apparatus can be used by the general public as an aid to relaxation, sleep, or sexual arousal, and clinically for the control and perhaps treatment of insomnia, tremors, epileptic seizures, and anxiety disorders. There is further application as a nonlethal weapon that can be used in law enforcement standoff situations, for causing drowsiness and disorientation in targeted subjects. It is then preferable to use venting acoustic monopoles in the form of a device that inhales and exhales air with subaudio frequency.

USP # 6,011,991 (January 4, 2000)
Communication System & Method Including Brain Wave Analysis...
Mardirossian, Aris
Abstract --- A system and method for enabling human beings to communicate by way of their monitored brain activity. The brain activity of an individual is monitored and transmitted to a remote location (e.g. by satellite). At the remote location, the monitored brain activity is compared with pre-recorded normalized brain activity curves, waveforms, or patterns to determine if a match or substantial match is found. If such a match is found, then the computer at the remote location determines that the individual was attempting to communicate the word, phrase, or thought corresponding to the matched stored normalized signal.

USP # 6,006,188 (December 21, 1999)
Speech Signal Processing for Determining Psychological or Physiological Characteristics...
Bogdashevsky, Rostislav, et al.
Abstract --- A speech-based system for assessing the psychological, physiological, or other characteristics of a test subject is described. The system includes a knowledge base that stores one or more speech models, where each speech model corresponds to a characteristic of a group of reference subjects. Signal processing circuitry, which may be implemented in hardware, software and/or firmware, compares the test speech parameters of a test subject with the speech models. In one embodiment, each speech model is represented by a statistical time-ordered series of frequency

representations of the speech of the reference subjects. The speech model is independent of a priori knowledge of style parameters associated with the voice or speech. The system includes speech parameterization circuitry for generating the test parameters in response to the test subject's speech. This circuitry includes speech acquisition circuitry, which may be located remotely from the knowledge base. The system further includes output circuitry for outputting at least one indicator of a characteristic in response to the comparison performed by the signal processing circuitry. The characteristic may be time-varying, in which case the output circuitry outputs the characteristic in a time-varying manner. The output circuitry also may output a ranking of each output characteristic. In one embodiment, one or more characteristics may indicate the degree of sincerity of the test subject, where the degree of sincerity may vary with time. The system may also be employed to determine the effectiveness of treatment for a psychological or physiological disorder by comparing psychological or physiological characteristics, respectively, before and after treatment.

USP # 5,954,630 (September 21, 1999)
FM Theta-Inducing Audible Sound...
Masaki, Kazumi, et al.
Abstract --- An audible sound of modulated wave where a very low-frequency wave of about 20 hertz or lower is superposed on an audio low-frequency wave effectively stimulates Fm theta in human brain waves to improve attention and concentration during mental tasks when auditorily administered. The audible sound is also effective in stimulation of human alpha wave when the very low-frequency wave lies within the range of about 2-10 hertz. Such audible sound is artificially obtainable by generating an electric signal which contains such a modulated wave, and transducing it into audible sound wave.

USP # 5,954,629 (September 21, 1999)
Brain Wave Inducing System
Yanagidaira, Masatoshi, et al.
Abstract --- Sensors are provided for detecting brain waves of a user, and a band-pass filter is provided for extracting a particular brain waves including an .alpha. wave included in a detected brain wave. The band-pass filter comprises a first band-pass filter having a narrow pass band, and a second band-pass filter having a wide pass band. One of the first and second band-pass filters is selected, and a stimulation signal is produced in dependency on an .alpha. wave extracted by a selected band-pass filter. In accordance with the stimulation signal, a stimulation light is emitted to the user in order to induce the user to relax or sleeping state.

USP # 5,935,054 (August 10, 1999)
Magnetic Excitation of Sensory Resonances
Loos, H.
Abstract --- The invention pertains to influencing the nervous system of a subject by a weak externally applied magnetic field with a frequency near 1/2 Hz. In a range of amplitudes, such fields can excite the 1/2 sensory resonance, which is the physiological effect involved in "rocking the baby".

USP # 5,922,016 (July 13, 1999)
Apparatus for Electric Stimulation of Auditory Nerves of a Human Being
Wagner, Hermann
Abstract --- Apparatus for electric stimulation and diagnostics of auditory nerves of a human being, e.g. for determination of sensation level (SL), most conformable level (MCL) and uncomfortable level (UCL) audibility curves, includes a stimulator detachably secured to a human being for sending a signal into a human ear, and an electrode placed within the human ear and electrically connected to the stimulator by an electric conductor for conducting the signals from the stimulator into the ear. A control unit is operatively connected to the stimulator for instructing the stimulator as to characteristics of the generated signals being transmitted to the ear.

USP # 5,868,103 (February 9, 1999)
Method and Apparatus for Controlling an Animal
Boyd, Randal
Abstract --- An apparatus for controlling an animal wherein the animal receives a control stimulus of the release of a substance having an adverse effect upon the animal as a corrective measure. The apparatus includes a transmitter for producing a transmitted field, and a releasable collar for attaching to the neck of the animal. The collar includes a receiver for receiving the transmitted field and for producing a received signal, a control circuit for determining when the received signal indicates that the animal requires a corrective measure and for producing a control signal, a container for containing the substance having an adverse effect upon the animal, and a mechanism for releasing the substance from the container into the presence of the animal upon the production of the control signal by the control circuit. In use, the transmitter is set to produce the transmitted field and the collar is attached to the neck of the animal. As the animal moves about, the receiver in the collar receives the transmitted field and produces a received signal. The control circuit determines when the received signal indicates that the animal requires a corrective measure. A control signal is produced by the control circuit when the determination is made that the animal requires a corrective measure. Upon the production of the control signal, the substance having an adverse effect upon the animal is released from the container and into the presence of the animal.

USP # 5,784,124 (July 21, 1998)
Supraliminal Method of Education...
D'Alitalia, Joseph A., *et al.*
Abstract --- A method of behavior modification involves having a patient view supraliminal video messages superimposed upon an underlying video presentation. The video messages incorporate messages wherein at least some of the messages link a desired modified behavior to positive feelings of the patient. A supraliminal message generator and superimposer iteratively selects individual messages for display from the sequence of messages, decompressing the messages as required, and places the selected messages in a buffer memory of a video generation device. A processor of the supraliminal message generator and superimposer then fades the selected message from an invisible level to a visible level on the video display, and then fades the selected message from the visible level back to the invisible level.

USP # 5,649,061 (July 15, 1997)
Device and Method for Estimating a Mental Decision
Smyth, Christopher
Abstract --- A device and method for estimating a mental decision to select a visual cue from the viewer's eye fixation and corresponding single event evoked cerebral potential. The device comprises an eyetracker, an electronic biosignal processor and a digital computer. The eyetracker determines the instantaneous viewing direction from oculometric measurements and a head position and orientation sensor. The electronic processor continually estimates the cerebral electroencephalogramic potential from scalp surface measurements following corrections for electrooculogramic, electromyogramic and electrocardiogramic artifacts. The digital computer analyzes the viewing direction data for a fixation and then extracts the corresponding single event evoked cerebral potential. The fixation properties, such as duration, start and end pupil sizes, end state (saccade or blink) and gaze fixation count, and the parametric representation of the evoked potential are all inputs to an artificial neural network for outputting an estimate of the selection interest in the gaze point of regard. The artificial neural network is trained off-line prior to application to represent the mental decisions of the viewer. The device can be used to control computerized machinery from a video display by ocular gaze point of regard alone, by determining which visual cue the viewer is looking at and then using the estimation of the task-related selection as a selector switch.

USP # 5,644,363 (July 1, 1997)
Apparatus for Superimposing Visual Subliminal Instructions on a Video Signal

Mead, Talbert

Abstract --- A subliminal video instructional device comprises circuitry for receiving an underlying video signal and presenting this signal to horizontal and vertical synchronization detection circuits, circuitry for generating a subliminal video message synchronized to the underlying video signal, and circuitry for adding the subliminal video message to the underlying video signal to create a combination video signal.

USP # 5,586,967 (December 24, 1996)

Method & Recording for Producing Sounds and Messages to Achieve Alpha & Theta Brainwave States...

Davis, Mark E.

Abstract --- A method and recording for the use in achieving alpha and theta brainwave states and effecting positive emotional states in humans, is provided which includes a medium having a musical composition thereon with an initial tempo decreasing to a final tempo and verbal phrases recorded in synchrony with the decreasing tempo.

USP # 5,562,597 (October 8, 1996)

Method & Apparatus for Reducing Physiological Stress

Van Dick, Robert C.

Abstract --- Physiological stress in a human subject is treated by generating a weak electromagnetic field about a quartz crystal. The crystal is stimulated by applying electrical pulses of pulse widths between 0.1 and 50 microseconds each at a pulse repetition rate of between 0.5K and 10K pulses per second to a conductor positioned adjacent to the quartz crystal thereby generating a weak electromagnetic field. A subject is positioned within the weak electromagnetic field for a period of time sufficient to reduce stress.

USP # 5,551,879 (September 3, 1996)

Dream State Teaching Machine

Raynie, Arthur D.

Abstract --- A device for enhancing lucidity in the dream state of an individual. The device includes electronic circuitry incorporated into a headband for the user to wear while sleeping. The circuitry includes a detector for fitting adjacent to the eye of the sleeping individual, for detecting Rapid Eye Movement (REM), which occurs during the dream state. The detector emits a signal that is evaluated by additional circuitry to determine whether or not REM sleep is occurring. If REM sleep is occurring, a signal is generated to operate a recorded, which typically plays prerecorded messages through the headphones engaging the ear of the sleeping individual.

USP # 5,539,705 (July 23, 1996)

Ultrasonic Speech Translator and Communication System

M. A. Akerman, M., et al.

Abstract --- A wireless communication system, undetectable by radio-frequency methods, for converting audio signals, including human voice, to electronic signals in the ultrasonic frequency range, transmitting the ultrasonic signal by way of acoustic pressure waves across a carrier medium, including gases, liquids and solids, and reconverting the ultrasonic acoustic pressure waves back to the original audio signal. This invention was made with government support under Contract DE-ACO5-840R2l400, awarded by the US Department of Energy to Martin Marietta Energy Systems, Inc.

USP # 5,507,291 (April 16, 1996)

Method & Apparatus for Remotely Determining Information as to Person's Emotional State ~ Stirbl, et al.

Abstract --- In a method for remotely determining information relating to a person's emotional state, an waveform energy having a predetermined frequency and a predetermined intensity is generated and wirelessly transmitted towards a remotely located subject. Waveform energy emitted from the subject is detected and automatically analyzed to derive information relating to the individual's emotional state. Physiological or physical parameters of blood pressure, pulse rate, pupil size, respiration rate and perspiration level are measured and compared with reference values to provide information utilizable in evaluating interviewee's responses or possibly criminal intent in security sensitive areas.

USP # 5,522,386 (June 4, 1996)
Apparatus for Determination of the Condition of the Vegetative Part of the Nervous System
Lerner, Eduard
Abstract --- Apparatus for use in the determination of the condition of the vegetative part of the nervous system and/or of sensory functions of an organism, i.e. a human being or animal. The apparatus comprises devices for generating and supplying to said organism at least one sensory stimulus chosen from a group of sensory stimuli, such as visual, sound, olfactory, gustatory, tactile or pain stimuli, and devices for measuring the skin potential and the evoked response of the organism to a stimulus. The measured data are processed by processing devices for automatically controlling the supply of at least one stimulus for providing a non-rhythmical sequence of stimuli. Preferably, pairs of stimuli are supplied for developing a conditioned reflex.

USP # 5,480,374 (January 2, 1996)
Method and Apparatus for Reducing Physiological Stress
Van Dick, Robert
Abstract --- Physiological stress in a human subject is treated by generating a weak electromagnetic field about a grounded electrode by the application of pulses of between 5 and 50 microseconds each at a pulse rate of between 0.5K and 10K pulses per second to a power electrode, the power electrode and grounded electrode being coupled to high voltage pulse generation means. A subject is positioned within the weak electromagnetic field for a period of time sufficient to cause an increase in his or her alpha or theta brain wave levels.

USP # 5,479,941 (January 2, 1996)
Device for Inducing Altered States of Consciousness
Harner, Michael
Abstract --- A rotating device for producing altered states of consciousness in a subject is provided. The subject's body rotates about a point in the center of the body support means at a speed between about 10 and about 60 revolutions per minute. In a preferred embodiment the direction of rotation is periodically reversed.

USP # 5,392,788 (February 28, 1995)
Method and Device for Interpreting Concepts and Conceptual Thought...
Hudspeth, William J.
Abstract --- A system for acquisition and decoding of EP and SP signals is provided which comprises a transducer for presenting stimuli to a subject, EEG transducers for recording brainwave signals from the subject, a computer for controlling and synchronizing stimuli presented to the subject and for concurrently recording brainwave signals, and either interpreting signals using a model for conceptual perceptional and emotional thought to correspond EEG signals to thought of the subject or comparing signals to normative EEG signals from a normative population to diagnose and locate the origin of brain dysfunctional underlying perception, conception, and emotion.

USP # 5,356,368 (October 18, 1994)
Method & Apparatus for Inducing Desired States of Consciousness
Monroe, Robert E.
Abstract --- Improved methods and apparatus for entraining human brain patterns, employing frequency following response (FFR) techniques, facilitate attainment of desired states of consciousness. In one embodiment, a plurality of electroencephalogram (EEG) waveforms, characteristic of a given state of consciousness, are combined to yield an EEG waveform to which subjects may be susceptible more readily. In another embodiment, sleep patterns are reproduced based on observed brain patterns during portions of a sleep cycle; entrainment principles are applied

to induce sleep. In yet another embodiment, entrainment principles are applied in the work environment, to induce and maintain a desired level of consciousness. A portable device also is described.

USP # 5,352,181 (October 4, 1994)
Method & Recording for Producing Sounds and Messages...
Davis, Mark E.
Abstract --- A method and recording for use in achieving Alpha and Theta brain wave states and effecting positive emotional states in humans to enhance learning and self-improvement, is provided which includes a medium having a musical composition recorded thereon with an initial tempo decreasing to a final tempo and verbal phrases, comprising between approximately 4 and approximately 8 words, recorded in synchrony with the decreasing initial tempo.

USP # 5,330,414 (July 19, 1994)
Brain Wave Inducing Apparatus
Yasushi, Mitsuo
Abstract --- A random signal generator outputs a random noise signal to a band pass filter which selectively passes frequency components in the frequency range of a desired brain wave from a subject. The output of the band pass filter is supplied to an automatic level controller. The automatic level controller sets the output of band pass filter to a predetermined amplitude. Then, the output of the automatic level controller is fed to a stimulating light generator, which converts the output of the automatic level controller into a light signal for stimulating the subject in order to induce the desired brain wave from the subject. The light signal is then emitted into the subject's eyes.

USP # 5,289,438 (February 22, 1994)
Method & System for Altering Consciousness
Gall, James
Abstract --- A system for altering the states of human consciousness involves the simultaneous application of multiple stimuli, preferable sounds, having differing frequencies and wave forms. The relationship between the frequencies of the several stimuli is exhibited by the equation g = 2.sup.n/4 .multidot.f where: f = frequency of one stimulus; g = frequency of the other stimuli or stimulus; and n = a positive or negative integer which is different for each other stimulus.

USP # 5,245,666 (September 14, 1993)
Personal Subliminal Messaging System
Mikell, Bruce T.
Abstract --- A personal subliminal messaging system includes a wide range linear subliminal modulator (43), a digital audio recording or play device (46), a microphone (51) to pick up the sound at the ear, and an earpiece (50) to deliver the subliminal message. The sound level at the user's ear is detected and measured. After risetime and decay conditioning of the varying dc control signal, the wide range linear modulator (43) uses this signal to control the level of the message to the earpiece (50). The user adjusts the system for a liminal of a subliminal level. The psychoacoustic phenomena of Post Masking is used to increase the integrity of the message in subliminal messaging systems.

USP # 5,270,800 (December 14, 1993)
Subliminal Message Generator
Sweet. Robert L.
Abstract --- A combined subliminal and supraliminal message generator for use with a television receiver permits complete control of subliminal messages and their manner of presentation. A video synchronization detector enables a video display generator to generate a video message signal corresponding to a received alphanumeric text message in synchronism with a received television signal. A video mixer selects either the received video signal or the video message signal for output. The messages produced by the video message generator are user selectable via a keyboard input. A message memory stores a plurality of alphanumeric text messages specified by user commands for

use as subliminal messages. This message memory preferably includes a read only memory storing predetermined sets of alphanumeric text messages directed to differing topics. The sets of predetermined alphanumeric text messages preferably include several positive affirmations directed to the left brain and an equal number of positive affirmations directed to the right brain that are alternately presented subliminally. The left brain messages are presented in a linear text mode, while the right brain messages are presented in a three dimensional perspective mode. The user can control the length and spacing of the subliminal presentations to accommodate differing conscious thresholds. Alternative embodiments include a combined cable television converter and subliminal message generator, a combine television receiver and subliminal message generator and a computer capable of presenting subliminal messages.

USP # 5,224,864 (July 6, 1993)

Method of Recording and Reproducing Subliminal Signals that are 180 Degrees Out of Phase

Woith, Blake F.

Abstract --- A subliminal recording includes both subliminal message and mask signals applied to both tracks of a two track recording medium. The subliminal message signals are identical in content, and are recorded in an out-of-phase relationship. The mask signals are recorded in phase. The resulting recording may be utilized in the conventional manner for subliminal recordings. By combining the composite signals in an inverted relationship, the mask signals cancel while the subliminal message signals are additive, thus allowing the presence of the subliminal message signal to be confirmed on the recording.

USP # 5,221,962 (June 22, 1993)

Subliminal Device having Manual Adjustment of Perception Level of Subliminal Messages

Backus, Alan L., *et al.*

Abstract --- A method and apparatus for presenting subliminal visual and/or audio messages which allows user verification of message content and presence, as well as proper adjustment of message obviousness while accounting for ambient conditions and user sensitivities is disclosed. This method and apparatus also presents synchronized reinforced sensory input of subliminal messages. This is performed by simultaneously overlaying images received from a VCR over a plurality of television signals. This apparatus directs overlay images over RF television signals having both audio and video components

USP # 5,215,468 (June 1, 1993)

Method and Apparatus for Introducing Subliminal Changes to Audio Stimuli

Lauffer, Martha A., et al.

Abstract --- A method and apparatus for introducing gradual changes to an audio signal so that the changes are subliminal. The changes can involve tempo and volume, for example, and can take the form of a gentle gradient having ever increasing/decreasing ramp-like changes over a sufficient duration, or a more complex program involving several gentle gradients. In the preferred embodiment, an enhanced audio play-back device such as a portable audio cassette recorder can be programmed to subliminally alter the characteristics of a standard pre-recorded tape containing music, for example. As a motivational tool during walking, jogging or other repetitive exercise, the tempo is gradually increased over a period of time to encourage a corresponding gradual (and subliminal) increase in physical exertion by a user whose rate of movement is proportional to the tempo of the music. The tempo can be either manually changed in conjunction with a subliminal program, or by itself in an override mode, or by itself in a version of the present-inventive audio play-back device which allows only manual tempo alternation. In an alternate embodiment, a special pre-recorded tape contains subliminal changes in tempo, for example, for play-back on a standard audio cassette recorder (which operates at one speed, only) to cause the same effect as the preferred embodiment.

USP # 5,213,562 (May 25, 1993)

Method of Inducing Mental, Emotional and Physical States of Consciousness...

Monroe, Robert A.
Abstract --- A method having applicability in replication of desired consciousness states; in the training of an individual to replicate such a state of consciousness without further audio stimulation; and in the transferring of such states from one human being to another through the imposition of one individual's EEG, superimposed on desired stereo signals, on another individual, by inducement of a binaural beat phenomenon.

USP # 5,194,008 (March 16, 1993)
Subliminal Image Modulation Projection and Detection System and Method
Mohan, William L., et al.
Abstract --- Weapon training simulation system including a computer operated video display scene whereon is projected a plurality of visual targets. The computer controls the display scene and the targets, whether stationary or moving, and processes data of a point of aim sensor apparatus associated with a weapon operated by a trainee. The sensor apparatus is sensitive to non-visible or subliminal modulated areas having a controlled contrast of brightness between the target scene and the targets. The sensor apparatus locates a specific subliminal modulated area and the computer determines the location of a target image on the display scene with respect to the sensor apparatus

USP # 5,175,571 (December 29, 1992)
Glasses with Subliminal Message
Tanefsky, Faye, et al.
Abstract --- A pair of subliminal imaging spectacles is provided with a matched pair of visual subliminal images designed and placed so as to merge into one image due to the stereoscopic effect of human vision and thus to impart a subliminal message to the wearer.

USP # 5,170,381 (December 8, 1992)
Method for Mixing Audio Subliminal Recordings
Taylor, Eldon, et al.
Abstract --- Audio subliminal recordings are made in which in addition to using a primary carrier, such as music, two audio channels are used to deliver subliminal messages to the brain. On one channel, accessing the left brain hemisphere, the message delivered is meaningfully spoken, forward-masked, permissive affirmations delivered in a round-robin manner by a male voice, a female voice and a child's voice. On the other channel, accessing the right brain, directive messages, in the same voices, are recorded in backward-masked (or meta-contrast). The three voices are recording in round-robin fashion with full echo reverberation. The audio tracks are mixed using a special processor which converts sound frequencies to electrical impulses and tracks the subliminal message to synchronize the subliminal message in stereo with the primary carrier. The processor maintains constant gain differential between the primary carrier and the subliminal verbiage and, with the subliminal verbiage being recorded with round-robin, full echo reverberation, ensures that none of a message is lost. The primary carrier should be continuous music without breaks or great differences in movements.

USP # 5,159,703 (October 27, 1992)
Silent Subliminal Presentation System
Lowery, Oliver
Abstract --- A silent communications system in which nonaural carriers, in the very low or very high audio frequency range or in the adjacent ultrasonic frequency spectrum, are amplitude or frequency modulated with the desired intelligence and propagated acoustically or vibrationally, for inducement into the brain, typically through the use of loudspeakers, earphones or piezoelectric transducers.

USP # 5,151,080 (September 29, 1992)
Method & Apparatus for Inducing & Establishing a Changed State of Consciousness
Bick, Claus
Abstract --- An electroacoustic device includes a sound generator as well as a system for producing

synthetic human speech, connected to a modulation stage for superimposing the output signals thereof. The superimposed output signals are applied via an amplifier stage to one of a headphone system or loudspeaker system.

USP # 5,135,468 (August 4, 1992)
Method & Apparatus of Varying the Brain State of a Person by Means of an Audio Signal
Meissner, Juergen P.
Abstract --- A method of varying the brain state of a person includes the steps of supplying the first audio signal to one ear of the person, supplying a second audio signal to the other ear of the person, and substantially continuously varying the frequency of at least one of the first and second audio signals to vary the brain state of the person.

USP # 5,134,484 (July 28, 1992)
Superimposing Method & Apparatus Useful for Subliminal Messages
Willson, Joseph
Abstract --- Data to be displayed is combined with a composite video signal. The data is stored in a memory in digital form. Each byte of the data is read out in sequential fashion to determine: the recurrence display rate of the data according to the frame sync pulses of the video signal; the location of the data within the video image according to the line sync pulses of the video signal; and the location of the data display within the video image according to the position information. Synchronization of the data with the video image is derived from the sync pulses of the composite video signal. A similar technique is employed to combine sound data with an audio signal. Data to be displayed may be presented as a subliminal message or may persist for a given time interval. The data may be derived from a variety of sources including a prerecorded or live video signal. The message may be a reminder message displayed upon a television screen to remind the viewer of an appointment. The data may be stored in a variety of different memory devices capable of high speed data retrieval. The data may be generated locally on-line or off-line and transferred to memory which stores the data necessary to create the message.

USP # 5,128,765 (July 7, 1992)
System for Implementing the Synchronized Superimposition of Subliminal Signals
Dingwall, Robert
Abstract --- An apparatus and system for the controlled delivery of a subliminal video and/or audio message on to a source signal from a video tape player or similar. The source signal is divided into audio and video portions. A video processor reads sychronization information from the source signal. A controller transmits a stored subliminal image at designated times to a mixer amplifier fully synchronized with the source signal. Concurrently, an audio subliminal message is applied to the source audio at a volume level regulated at some fraction to the source audio. The combined signals are transmitted to a monitor for undistracted viewing.

USP # 5,123,899 (June 23, 1992)
Method & System for Altering Consciousness
Gall, James
Abstract --- A system for altering the states of human consciousness involves the simultaneous application of multiple stimuli, preferable sounds, having differing frequencies and wave forms. The relationship between the frequencies of the several stimuli is exhibited by the equation g = s.sup.n/4 .multidot.f where: f = frequency of one stimulus; g = frequency of the other stimuli of stimulus; and n=a positive or negative integer which is different for each other stimulus.

USP # 5,052,401 (October 1, 1991)
Sherwin, Gary
Product Detector for a Steady Visual Evoked Potential Stimulator and Product Detector
Abstract --- An automated visual testing system is disclosed which presents an alternating steady state visual stimulus to a patient through an optical system that modifies the stimulus image. As the image changes, the patient produces evoked potentials that change. The evoked potentials are

detected by a product detector which produces the amplitude of the evoked potentials. The product detector includes filters which isolate the patient's evoked potentials, a modulator which detects the response using the stimulus source frequency and a demodulator that determines the amplitude of the response. The product detector detects the level of the steady state evoked potential signals even in the presence of substantial background noise and extraneous electroencephalograhic signals. These detectors can be used to monitor the evoked potential produced by visual, aural or somatic steady state stimuli. The components described above can be used to produce a system that can determine to which of several different displays an observer is paying attention by providing images that blink at different frequencies and product detectors for each of the stimulus frequencies. The product detector producing the highest output indicates the display upon which the observer is focused.

USP # 5,047,994 (September 10, 1991)
Supersonic Bone Conduction Hearing Aid and Method
Lenhardt, Martin, et al.
Abstract --- A supersonic bone conduction hearing aid that receives conventional audiometric frequencies and converts them to supersonic frequencies for connection to the human sensory system by vibration bone conduction. The hearing is believed to use channels of communications to the brain that are not normally used for hearing. These alternative channels do not deteriorate significantly with age as does the normal hearing channels. The supersonic bone conduction frequencies are discerned as frequencies in the audiometric range of frequencies.

USP # 5,036,858 (August 6, 1991)
Method & Apparatus for Changing Brain Wave Frequency
Carter, John L., *et al.*
Abstract --- A method for changing brain wave frequency to a desired frequency determines a current brain wave frequency of a user, generates two frequencies with a frequency difference of a magnitude between that of the current actual brain wave frequency and the desired frequency but always within a predetermined range of the current actual brain wave frequency, and produces an output to the user corresponding to the two frequencies. One apparatus to accomplish the method has a computer processor, a computer memory, EEG electrodes along with an amplifier, a programmable timing generator responsive to the computer processor for generating the two frequencies, audio amplifiers and a beat frequency generator driving a visual frequency amplifier.

USP # 5,027,208 (June 25,1991)
Therapeutic Subliminal Imaging System
Dwyer, Jr., Joseph, et al.
Abstract --- A therapeutic subliminal imaging system wherein a selected subliminal message is synchronized with and added to an existing video signal containing a supraliminal message. A television receiver or video recorder can be used to provide the supraliminal message and a video processing circuit varies the intensity of that perceptible message to incorporate one or more subliminal images.

USP # 5,017,143 (May 21, 1991)
Method and Apparatus for Producing Subliminal Images
Backus, Alan, et al.
Abstract --- A method and apparatus to produce more effective visual subliminal communications. Graphic and/or text images, presented for durations of less than a video frame, at organized rhythmic intervals, the rhythmic intervals intended to affect user receptivity, moods or behavior. Subliminal graphic images having translucent visual values locally dependent on background values in order to maintain desired levels of visual contrast.

USP # 4,958,638 (September 25, 1990)
Non-Contact Vital Signs Monitor
Sharpe, Steven, et al.

Abstract --- An apparatus for measuring simultaneous physiological parameters such as heart rate and respiration without physically connecting electrodes or other sensors to the body. A beam of frequency modulated continuous wave radio frequency energy is directed towards the body of a subject. The reflected signal contains phase information representing the movement of the surface of the body, from which respiration and heartbeat information can be obtained. The reflected phase modulated energy is received and demodulated by the apparatus using synchronous quadrature detection. The quadrature signals so obtained are then signal processed to obtain the heartbeat and respiratory information of interest.

USP # 4,924,744 (May 15, 1990)
Apparatus for Generating Sound through Low Frequency and Noise Modulation
Lenzen, Reiner
Abstract --- In an apparatus for generating sound, there are provided a plurality of channels for generating sounds. Each of the channels includes a memory for storing waveform data, and at least one of the channels includes a noise generator so that various kinds of sounds including rhythm sound-effects sound, effects sound-vibrato etc. are generated. There is further provided a controller by which voice sound signal is passed through the channels so that artificial sound, voice sound etc. are generated. There is still further provided a circuit for adjusting an amplitude level of a whole sound which is obtained by mixing output sounds of the channels so that far and near sound is produced. Further, each of the channels includes left and right attenuators which divide a channel sound into left and right channel sounds. Still further, the apparatus comprises a low frequency oscillator for controlling a depth of frequency modulation, and a controller for writing sampling data of a predetermined waveform into serial addresses of a memory.

USP # 4,889,526 (December 26, 1989)
Non-Invasive Method & Apparatus for Modulating Brain Signals...
Rauscher, Elizabeth A.
Abstract --- This invention incorporates the discovery of new principles which utilize magnetic and electric fields generated by time varying square wave currents of precise repetition, width, shape and magnitude to move through coils and cutaneously applied conductive eletrodes in order to stimulate the nervous system and reduce pain in humans. Timer means, adjustment means, and means to deliver current to the coils and conductive eletrodes are described, as well as a theoretical model of the process. The invention incorporates the concept of two cyclic expanding and collapsing magnetic fields which generate precise wave forms in conjunction with each other to create a beat frequency which in turn causes the ion flow in the nervous system of the human body to be efficiently moved along the nerve path where the locus of the pain exists to thereby reduce the pain. The wave forms are created either in one or more coils, one or more pairs of electrodes, or a combination of the two.

USP # 4,883,067 (November 28, 1989)
Method & Apparatus for Translating the EEG into Music...
Knispel, Joel, *et al.*
Abstract --- A method and apparatus for applying a musical feedback signal to the human brain, or any other brain, to induce controllable psychological and physiological responses. A signal representing the ongoing electroencephalographic (EEG) signal of a brain preferably is obtained from the electrode location on the scalp known as CZ or P3 in clinical notation. A signal processor converts the ongoing EEG into electrical signals which are converted into music by synthesizers. The music is acoustically fed back to the brain after a time delay calculated to shift the phase of the feedback in order to reinforce specific or desired ongoing EEG activity from the scalp position of interest. The music is comprised of at least one voice that follows the moment-by-moment contour of the EEG in real time to reinforce the desired EEG activity. The music drives the brain into resonance with the music to provide a closed loop or physiological feedback effect. Preferably, the musical feedback comprises additional voices that embody psychoacoustic principles as well as

provide the content and direction normally supplied by the therapist in conventional biofeedback. The invention contemplates numerous applications for the results obtained.

USP # 4,877,027 (October 31, 1989)
Hearing System
Brunkan, Wayne B.
Abstract --- Sound is induced in the head of a person by radiating the head with microwaves in the range of 100 megahertz to 10,000 megahertz that are modulated with a particular waveform. The waveform consists of frewuency modulated bursts. each burst is made up of 10 to 20 uniformly spaced pulses grouped tightly together. the burst width is between 500 nanoseconds and 100 microseconds. The pulse width is in the range of 10 nanoseconds to 1 microsecond. The bursts are frequency modulated by the audio input to create the sensation of hearing in the person whose head is irradiated.

USP # 4,858,612 (August 22, 1989)
Hearing Device
Stocklin, Philip L.
Abstract --- A method and apparatus for stimulation of hearing in mammals by introduction of a plurality of microwaves into the region of the auditory cortex is shown and secribed. A microphone is used to transform sound signals into eletrical signals which are in turn analyzed and processed to provide controls for generating a plurality of microwave signals at different frequencies. the multifrequency microwaves are then applied to the brain in the region of the auditory cortex. By this method sounds are perceived by the mamal which are representative of the original sound received by the microphone.

USP # 4,834,701 (May 30, 1989)
Apparatus for Inducing Frequency Reduction in Brain Wave
Masaki, Kazumi
Abstract --- Frequency reduction in human brain wave is inducible by allowing human brain to perceive 4-16 hertz beat sound. Such beat sound can be easily produced with an apparatus, comprising at least one sound source generating a set of low-frequency signals different each other in frequency by 4-16 hertz. Electroencephalographic study revealed that the beat sound is effective to reduce beta-rhythm into alpha-rhythm, as well as to retain alpha-rhythm.

USP # 4,821,326 (April 11, 1989)
Non-Audible Speech Generation Method & Apparatus
MacLeod, Norman
Abstract --- A non-audible speech generation apparatus adn method for producing non-audible seech signals which includes an ultasonic transducer or vibrator for projecting a series of glottal shaped ultrasonic pulses to the vocal track of a speaker. The glottal pulses, in the approximate frequency spectrum extending from 15 kilohertz to 105 kilohertz, contains harmonics of approximately 30 times the grequency of the acoustical harmonicsgenerated by the vocal cords, but which may nevertheless be amplitude modulated to produce non-auduble speech by the speaker's silently mouthing of words. The ultrasonic speech is then received by an ultasonic transducer disposed outside of the speaker's mouth and electronically communicated to a translation device which down converts the ultrasonics signals to corresponding signals in the audible frquency range and synthesizes the signals into artificial speech.

USP # 4,777,529 (October 11, 1988)
Auditory Subliminal Programming System
Schultz, Richard M., ***et al.***
Abstract --- An auditory subliminal programming system includes a subliminal message encoder that generates fixed frequency security tones and combines them with a subliminal message signal to produce an encoded subliminal message signal which is recorded on audio tape or the like. A corresponding subliminal decoder/mixer is connected as part of a user's conventional stereo system

and receives as inputs an audio program selected by the user and the encoded subliminal message. The decoder/mixer filters the security tones, if present, from the subliminal message and combines the message signals with selected low frequency signals associated with enhanced relaxation and concentration to produce a composite auditory subliminal signal. The decoder/mixer combines the composite subliminal signal with the selected audio program signals to form composite signals only if it detects the presence of the security tones in the subliminal message signal. The decoder/mixer outputs the composite signal to the audio inputs of a conventional audio amplifier where it is amplified and broadcast by conventional audio speakers.

USP # 4,734,037 (March 29, 1988)
Message Screen
McClure, J. Patrick
Abstract --- A transparent sheet is disclosed having a message thereon. The sheet has a first side adapted to be attached facing a plate which is normally viewed by a viewer and a second side facing the viewer. The message is arranged to be readably intelligible from the second side but is not liminally visible to the viewer when viewed from a normal viewing distance from the second side under normal viewing conditions. The message has a subliminal effect upon the viewer when viewed from the normal viewing distance from the second side under normal viewing conditions. A viewer can electively subject him or herself to subliminal messages while viewing television at leisure.

USP # 4,717,343 (January 5, 1988)
Method of Changing a Person's Behavior
Densky, Alan B.
Abstract --- A method of conditioning a person's unconscious mind in order to effect a desired change in the person's behavior which does not require the services of a trained therapist. Instead the person to be treated views a program of video pictures appearing on a screen. The program as viewed by the person's unconscious mind acts to condition the person's thought patterns in a manner which alters that person's behavior in a positive way.

USP # 4,699,153 (October 13, 1987)
System for Assessing Verbal Psychobiological Correlates
Shevrin, Howard, *et al.*
Abstract --- A system for assessing psychobiological conditions of a subject utilizes a plurality of words which are selected to be in four categories as critical stimuli. The words are presented by a tachistoscope to the subject in subliminal and supraliminal modes of operation. Subliminal stimulation of the subject is achieved by presenting the selected words for an exposure period of approximately one millisecond. The supraliminal exposure time is approximately thirty milliseconds. Prior to stimulation, the subject is diagnosed in accordance with conventional psychoanalytical techniques to establish the presence and nature of a pathological condition. The words are selected and categorized in four groups: pleasant words, unpleasant words, words related to a diagnosed conscious pathological condition, and words related to a diagnosed unconscious pathological condition. The brain wave responses which are evoked by the stimulation are collected via electrodes and analyzed in accordance with a transinformation technique which is based on information signal theory for establishing a probabilistic value which corresponds to the information content of the evoked responses.

USP # 4,692,118 (September 8, 1987)
Video Subconscious Display Attachment
Mould, Richard E.
Abstract --- An apparatus and method for introducing messages to the subconscious mind is disclosed, which includes a panel positioned adjacent a television screen, with the panel having non-distractive messages imprinted thereon, such that as the subject consciously focuses his attention on

the video screen, his subconscious mind records the message from the panel that is within his peripheral vision.

USP # 4,616,261 (October 7, 1986)
Method & Apparatus for Generating Subliminal Visual Messages
Crawford, James R., ***et al.***
Abstract --- A system for generating a subliminal message during the display of a normal television program on a television receiver utilizes a personal computer to generate an RF carrier modulated with video signals encoding the subliminal message. The computer runs under the control of an application program which stores the subliminal message and also controls the computer to cause it to generate timing signals that are provided to a single pole double-throw switch. The source of the normal television program and the video output of the computer are connected to the two switch inputs and the switch output is connected to the television receiver antenna system. The timing signals cause the switch to normally display the conventional television program and to periodically switch to the computer output to generate the subliminal message. The video output of the computer includes horizontal and vertical synchronizing signals which are of substantially the same frequency as the synchronizing signals incorporated within the normal program source but of an arbitrary phase.

USP # 4,573,449 (March 4, 1986)
Method for Stimulating the Falling Asleep and/or Relaxing Behavior of a Person
Warnke, Egon F.
Abstract --- A method and apparatus is provided with which a person suffering from sleeplessness can be more easily relaxed and may more rapidly fall asleep. In particular, sound pulses are emitted by an electro-acoustic transducer, according to the cadence of which, the person seeking to fall asleep is induced to breathe in and out over a predetermined period of time. By suitably selecting the pulse sequence frequency, the pitch and the amplitude of the sound pulses may be adjusted thereby enhancing the process of falling asleep.

USP # 4,508,105 (April 2, 1985)
Shadow Generating Apparatus
Whitten, Glen, et al.
Abstract --- Disclosed is an apparatus for inducing various brain wave patterns through visual stimulation. The apparatus comprises a pair of spectacles or other viewing apparatus having a liquid crystal display embedded in each lens. By repetitively activating and deactivating the liquid crystals, shadows are generated which are perceived by the subject individual wearing the viewing apparatus. Responding to the frequency of shadow generation, the subject's brain is thereby induced to generate sympathetic brain wave frequencies. The apparatus finds particular utility in the generation of alpha waves. Because learning is enhanced when the brain is in the alpha state, activities such as listening to tapes or lectures and the like can be carried out with greater facility. Shadow generation is accomplished through the use of a timing mechanism for each liquid crystal display and the frequency for each is adjustable over a wide range, permitting synchronous or asynchronous timing.

USP # 4,395,600 (July 26, 1983)
Auditory Subliminal Message System & Method
Lundy, Rene R., ***et al.***
Abstract --- Ambient audio signals from the customer shopping area within a store are sensed and fed to a signal processing circuit that produces a control signal which varies with variations in the amplitude of the sensed audio signals. A control circuit adjusts the amplitude of an auditory subliminal anti-shoplifting message to increase with increasing amplitudes of sensed audio signals and decrease with decreasing amplitudes of sensed audio signals. This amplitude controlled subliminal message may be mixed with background music and transmitted to the shopping area. To reduce distortion of the subliminal message, its amplitude is controlled to increase at a first rate slower than the rate of increase of the amplitude of ambient audio signals from the area. Also, the

amplitude of the subliminal message is controlled to decrease at a second rate faster than the first rate with decreasing ambient audio signal amplitudes to minimize the possibility of the subliminal message becoming supraliminal upon rapid declines in ambient audio signal amplitudes in the area. A masking signal is provided with an amplitude which is also controlled in response to the amplitude of sensed ambient audio signals. This masking signal may be combined with the auditory subliminal message to provide a composite signal fed to, and controlled by, the control circuit.

USP # 4,388,918 (June 21, 1983)
Mental Harmonization Process
Filley, Charles C.
Abstract --- A state of relaxation or mental harmonization in a subject is created by exposing a color solely to one field of vision of a subject and the complement of that color solely to the other field of vision of the subject while simultaneously exposing an audible tone solely to one ear of the subject and a harmonious tone solely to the other ear of the subject. The color and tones employed are subjectively comfortable and compatible. Preferably, the frequency difference between the two audible tones is one-half the frequency of the audible tone having the lowest frequency.

USP # 4,354,505 (October 19, 1982)
Method of and Apparatus for Testing and Indicating Relaxation State of a Human Subject
Shiga, Kazumasa
Abstract --- In a self-training biofeedback system, a physiological signal representing the state of relaxation of a person using the system is applied to a time counter to generate a binary count output representing the relaxation period. A visual indicator connected to the time counter provides the self trained person with a quick display of the measured time period so he can gauge the depth of his relaxation.

USP # 4,335,710 (June 22, 1982)
Device for the Induction of Specific Brain Wave Patterns
Williamson, John
Abstract --- Brain wave patterns associated with relaxd and meditative states in a subject are gradually induced without deleterious chemical or neurological side effects. A white noise generator (11) has the spectral noise density of its output signal modulated in a manner similar to the brain wave patterns by a switching transistor within a spetrum modulator and converted to an audio signal by acoustic transducer. Ramp generator gradually increases the voltage received by and resultant output frequency of voltage controlled oscillator whereby switching transistor periodically shunts the high frequency components of the white noise signal to ground.

USP # 4,315,501 (February 16, 1982)
Learning-Relaxation Device
Gorges, Denis E.
Abstract --- Disclosed is a device for relaxing, stimulating and/or driving brain wave form function in a human subject. The device comprises, in combination, an eye mask having independently controlled left and right eyepieces and a peripheral light array in each eyepiece, an audio headset having independently controlled left and right earpieces and a control panel which controls light and sound signals to the light arrays and earpieces, respectively. Various control functions allow simultaneous or alternating light and sound pulsations in the left and right light arrays and earpieces, as well as selective phasing between light and sound pulsations.

USP # 4,227,516 (October 14, 1980)
Apparatus for Electrophysiological Stimulation
Meland, Bruce C., *et al.*
Abstract --- Apparatus for the electrophysiological stimulation of a patient is provided for creating an analgesic condition in the patient to induce sleep, treat psychosomatic disorders, and to aid in the induction of electrohypnosis and altered states of consciousness. The foregoing is achieved by repetitive stimuli in the patient for whom external influences, namely those of sight and sound, are

intentionally excluded. The apparatus produces electrical stimulation of the patient in the form of a modulated wave which produces impulses in the delta, theta, alpha and beta regions of the brain's electrical activity, the electrical stimulation being accompanied by two sources of audio stimulation, one of which is a sinusoidal tone modulated by and synchronized with the electrical stimulation, and the other is derived from sound recordings.

USP # 4,191,175 (March 4, 1980)
Method & Apparatus for Repetitively Producing a Noise-like Audible Signal
Nagle, William L.
Abstract --- A digital pulse generator and shift register repetitively produce bursts of digital pulses at a first adjustable repetition frequency. The repetition frequency of the pulses in each burst is also adjustable. A pink noise filter accentuates the lower burst frequency components near 7 hz and substantially attenuates all frequency components of the bursts above a first cut-off point near 10 Khz. A tunable band pass amplifier having a center frequency adjustable over a preselected range of frequencies optimally detectable by the average human ear accentuates the pink noise filter output near 2.6 Khz. The tunable amplifier drives an audible signal source with noise-like pulses of varying amplitudes and frequency components. A low pass amplifier may be connected to the pink noise filter to generate a train of pulses having a repetition frequency near 7 hz which pulses a light source in synchronism with the audible noise-like signal.

USP # 4,141,344 (February 27, 1979)
Sound Recording System
Barbara, Louis J.
Abstract --- In recording an audio program, such as music or voice, on a magnetic tape recorder an A.C. signal generator operating at a frequency below about 14 Hz provides an AC baseline for the audio program signal. This 14 Hz or lower AC signal is sensed by the listener's ear to create an Alpha or Theta state in his brain when the tape is played back.

USP # 4,082,918 (April 4, 1978)
Audio Analgesic Unit
Chang, Roland W., *et al.*
Abstract --- An audio analgesic unit for use in masking sounds and substituting another sound which includes earmuffs to be used by a dental patient in which speakers are arranged and connected to a patient operated remote control unit to control the sound levels and a master control unit to override the patient remote control unit and operated by an operator, such as a dentist. A beeper indicates operation mode change.

USP # 4,034,741 (July 12, 1977)
Noise Generator & Transmitter
Adams, Guy E., *et al.*
Abstract --- An analgesic noise generator employs a circuit that can be switched to provide a variable waveform from an active noise source out of an integrated circuit amplifier.

USP # 3,967,616 (July 6, 1976)
Multichannel System for & Multifactorial Method of Controlling the Nervous System of a Living Organism
Ross, Sidney A.
Abstract --- A novel method for controlling the nervous system of a living organism for therapeutic and research purposes, among other applications, and an electronic system utilized in, and enabling the practice of, the invented method. Bioelectrical signals generated in specifictopological areas of the organism's nervous system, typically areas of the brain, are processed by the invented system so as to produce a sensory stimulus if the system detects the presence or absence, as the case may be, of certain characteristics in the waveform patterns of the bioelectrical signals being monitored. The coincidence of the same or different characteristics in two or more waveform patterns, or the non-coincidence thereof, may be correlated with a certain desired condition of the organism's nervous

system; likewise, with respect to the coincidence or non-coincidence of different characteristics of a single waveform pattern. In any event, the sensory stimulus provided by the invented system, typically an audio or visual stimulus, or combination thereof, is fed back to the organism which associates its presence with the goal of achieving the desired condition of its nervous system. Responding to the stimulus, the organism can be trained to control the waveform patterns of the monitored bioelectrical signals and thereby, control its own nervous system. The results of the coincidence function permit results heretofore unobtainable.

USP # 3,951,134 (April 20, 1976)
Apparatus & Method for Remotely Monitoring & Altering Brain Waves
Malech, Robert G.
Abstract --- Apparatus for and method of sensing brain waves at a position remote from a subject whereby electromagnetic signals of different frequencies are simultaneously transmitted to the brain of the subject in which the signals interfere with one another to yield a waveform which is modulated by the subject's brain waves. The interference waveform which is representative of the brain wave activity is re-transmitted by the brain to a receiver where it is demodulated and amplified. The demodulated waveform is then displayed for visual viewing and routed to a computer for further processing and analysis. The demodulated waveform also can be used to produce a compensating signal which is transmitted back to the brain to effect a desired change in electrical activity therein.

USP # 3,884,218 (May 20, 1975)
Method of Inducing & Maintaining Various Stages of Sleep in the Human Being
Monroe, Robert A.
Abstract --- A method of inducing sleep in a human being wherein an audio signal is generated comprising a familiar pleasing repetitive sound modulated by an EEG sleep pattern. The volume of the audio signal is adjusted to overcome the ambient noise and a subject can select a familiar repetitive sound most pleasing to himself.

USP # 3,837,331 (September 24, 1974)
System & Method for Controlling the Nervous System of a Living Organism
Ross, S.
Abstract --- A novel method for controlling the nervous system of a living organism for therapeutic and research purposes, among other applications, and an electronic system utilized in, and enabling the practice of the invented method. Bioelectrical signals generated in specific topological areas of the organism's nervous system, typically areas of the brain, are processed by the invented system so as to produce an output signal which is in some way an analog of selected characteristics detected in the bioelectrical signal. The output of the system, typically an audio or visual signal, is fed back to the organism as a stimulus. Responding to the stimulus, the organism can be trained to control the waveform pattern of the bioelectrical signal generated in its own nervous system.

USP # 3,835,833 (September 17, 1974)
Method for Obtaining Neurophysiological Effects
Limoge, A.
Abstract --- A method and apparatus for obtaining neurophysiological effects on the central and/or peripheral systems of a patient. Electrodes are suitably positioned on the body of the patient and a composite electric signal is applied at the electrodes. The composite signal is formed by the superpositioning of two signals: a first signal which is a rectified high-frequency carrier modulated in amplitude to about 100 percent by substantially square-shaped pulses whose duration, amplitude and frequency are chosen according to the neurophysiological effects desidered, and a second signal which has a relatively white noise spectrum. The mean value of the first electric signal has a predetermined sign which is opposite the sign of the mean value of the second electric signal.

USP # 3,773,049 (November 20, 1973)
Apparatus for Treatment of Neuropsychic & Somatic Diseases with Heat, Light, Sound &

VHF Electromagnetic Radiation
L. Y. Rabichev, et al.
Abstract --- N/A

USP # 3,766,331 (October 16, 1973)
Hearing Aid for Producing Sensations in the Brain
Zink, Henry R.
Abstract --- A pulsed oscillator or transmitter supplies energy to a pair of insulated electrodes mounted on a person's neck. The transmitter produces pulses of intensity greater than a predetermined threshold value and of a width and rate so as to produce the sensation of hearing without use of the auditory canal, thereby producing a hearing system enabling otherwise deaf people to hear.

USP # 3,727,616 (March 17, 1973)
Electronic System for Stimulation of Biological Systems
Lenskes, H.
Abstract --- A receiver totally implanted within a living body is inductively coupled by two associated receiving coils to a physically unattached external transmitter which transmits two signals of different frequencies to the receiver via two associated transmitting coils. One of the signals from the transmitter provides the implanted receiver with precise control or stimulating signals which are demodulated and processed in a signal processor network in the receiver and then used by the body for stimulation of a nerve, for example, while the other signal provides the receiver with a continuous wave power signal which is rectified in the receiver to provide a source of electrical operating power for the receiver circuitry without need for an implanted battery.

USP # 3,712,292 (January 23, 1973)
Method & Apparatus for Producing Swept FM Audio Signal Patterns for Inducing Sleep
Zentmeyer, J.
Abstract --- A method of producing sound signals for inducing sleep in a human being, and apparatus therefor together with REPRESENTATIONS thereof in recorded form, wherein an audio signal is generated representing a familiar, pleasing, repetitive sound, modulated by continuously sweeping frequencies in two selected frequency ranges having the dominant frequencies which occur in electrical wave patterns of the human brain during certain states of sleep. The volume of the audio signal is adjusted to mask the ambient noise and the subject can select any of several familiar, repetitive sounds most pleasing to him.

USP # 3,647,970 (March 7, 1972)
Method and System for Simplifying Speech Waveforms
Flanagan, G. Patrick
Abstract --- A complex speech waveform is simplified so that it can be transmitted directly through earth or water as a waveform and understood directly or after amplification.

USP # 3,629,521 (January 8, 1970)
Hearing Systems
Puharich, Henry K.
Abstract --- The present invention relates to the stimulation of the sensation of hearing in persons of impaired hearing abilities or in certain cases persons totally deaf utilizing RF energy. More particularly, the present invention relates to a method and apparatus for imparting synchronous AF or ""acoustic" signals and so-called "transdermal" or RF signals. Hearing and improved speech discrimination, in accordance with one aspect of the present invention, is stimulated by the application of an AF acoustical signal to the "ear

system" conventional biomechanism of hearing, which is delivered to the brain through the "normal" channels of hearing and a separate transdermal RF electrical signal which is applied to the "facial nerve system" and is detectable as a sensation of hearing. Vastly improved and enhanced hearing may be achieved...

USP # 3,576,185 (April 27, 1971)
Sleep-Inducing Method & Arrangement using Modulated Sound & Light
Meseck, Oscar & Schulz, Hans R.
Abstract --- N/A

USP # 3,568,347 (February 23, 1971)
Psycho-Acoustic Projector
Flanders, Andrew
Abstract --- A system for producing aural psychological disturbances and partial deafness in the enemy during combat situations.

USP # 3,393,279 (July 16, 1968)
Nervous System Excitation Device
Flanagan, Giles P.
Abstract --- A method of transmitting audio information via a radio frequency signal modulated with the audio info through electrodes placed on the subject's skin, causing the sensation of hearing the audio information in the brain.

USP # 3,170,993 (February 23, 1965)
Means for Aiding Hearing by Electrical Stimulation of the Facial Nerve System
Puharich, Henry & Lawrence, Joseph
Abstract --- N/A

USP # 3,156,787 (November 10, 1964)
Solid State Hearing System
Lawrence, Joseph & Puharich, Henry
Abstract --- N/A

USP # 2,995,633 (August 8, 1961)
Means for Aiding Hearing
Puharich, Henry & Lawrence, J.
Abstract --- Means for converting audible signals to electrical signals and conveying them to viable nerves of the facial system.

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...

Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

rexresearch.com

US Patents : Force Field Propulsion

Ad
Louisville Web Group
Bring the world to your door
Lwebg delivers a superior product that yields a high ROI. We're Drupal specialists

Ad
Louisville Web Group
Need help with Drupal 6?
Lwebg delivers a superior product that yields a high ROI. We're Drupal specialists

fetchingstock
shutterstock

USP # 6,404,089 (6-11-02): Electrodynamic Field Generator
Tomion, Mark

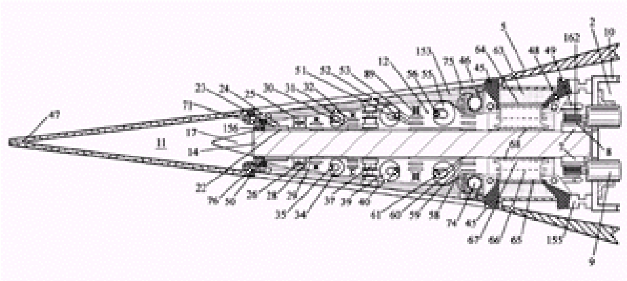
47
11
71
23
24
25
30
31
32
51
52
53
89
12
153
75
46
45
64
63
5
48
49
162
2
10
17
14
136
22
76
50
26
28
29
35
34
37
39
40
61
60
59
58
74
43
67
66
65
155
9
8
68

USP # 6,179,250 (1-30-01): Air & Space Vehicle Propulsion System
Waters, Lawrence

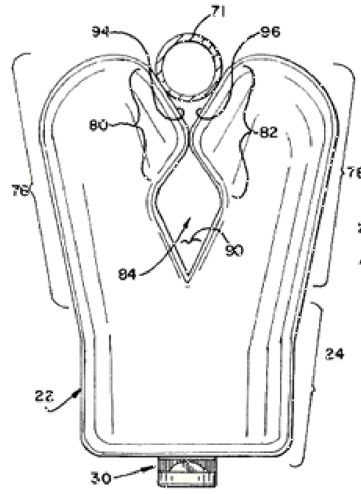
71
94
96
80
82
76
78
20
90
84
24
22
30

USP # 5,211,006 (5-18-93): Magnetohydrodynamic Propulsion System
Sohnly, Michael J.

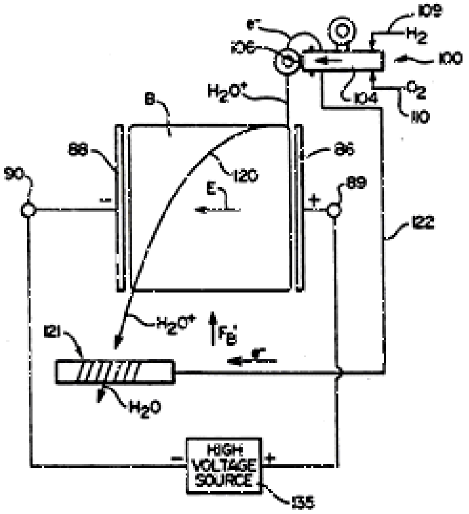
e⁻
109
H2
106
100
H2O+
104
O2
110
B
88
86
120
90
89
E
-
+
122
121
H2O+
FB'
e⁻
H2O
HIGH VOLTAGE SOURCE
-
+
135

USP # 5,197,279 (3-30-93): Electromagnetic Energy Propulsion Engine
Taylor, James R.

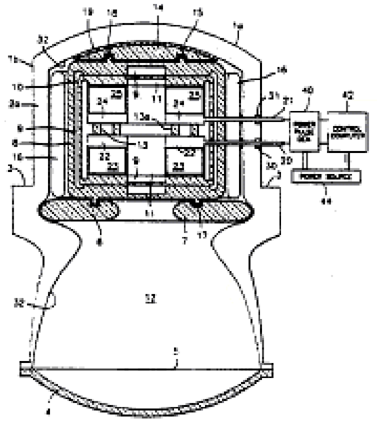
19
18
14
15
1a
32
1b
10
16
9
8
25
24
11
13a
13
22
23
31
40
42
CONTROL COMPUTER
30
3
POWER SOURCE
44
6
7
17
12
5
4

USP # 5,142,861 (9-1-92): Nonlinear Electromagnetic Propulsion System & Method
Schlicher, Rex L., *et al.*

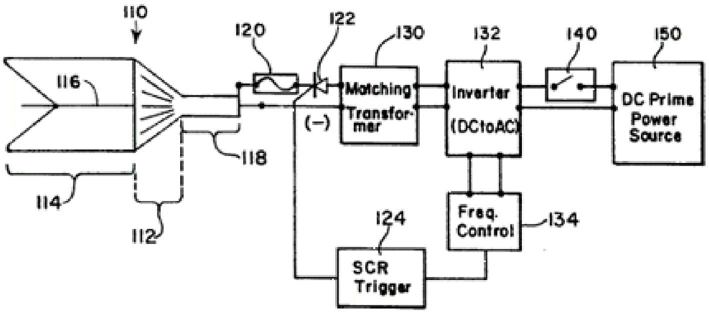
110
120
122
130
132
140
150
116
Matching Transformer
(–)
Inverter (DCtoAC)
DC Prime Power Source
118
114
112
124
Freq. Control
134
SCR Trigger

USP # 4,967,983 (11-6-90): Airship
Motts, Brian C.

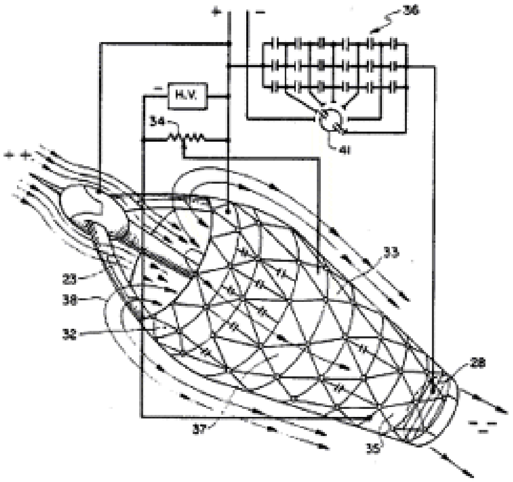
+
–
36
–
H.V.
34
41
+ +
23
38
32
33
28
37
35
– –

USP # 4,891,600 (1-2-90): Dipole Accelerating Means & Method
Cox, James E.

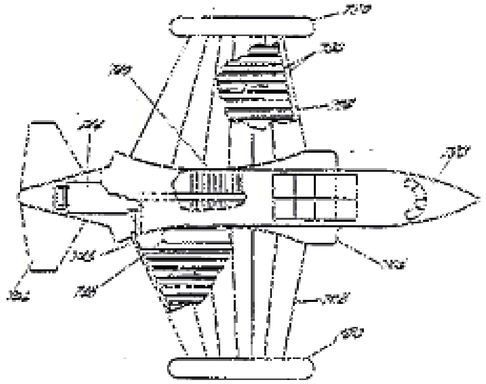

USP # 4,663,932 (5-12-87): Dipolar Force Field Propulsion
Cox, James E.

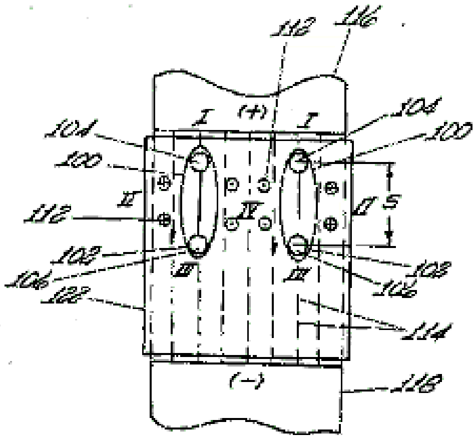
112
116
I
(+)
I
104
100
104
100
II
IV
II
S
112
108
106
III
III
108
106
102
114
(−)
118

USP # 3,662,554 (5-16-72): Electromagnetic Propulsion Device...

De Broqueville, Axel

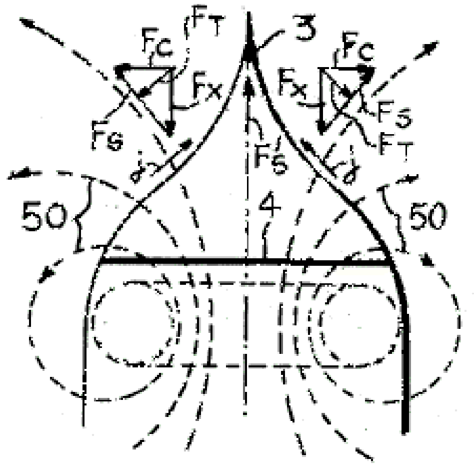
FT
3
FC
FC
FX
FX
FS
FS
FS
FT
γ
γ
4
50
50

USP # 3,656,013 (4-11-72): Apparatus for Generating a Motional Electric Field

Hooper, William J.

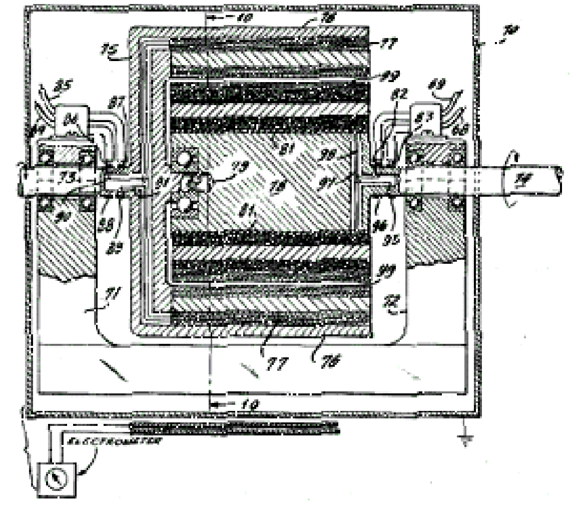
10
76
77
15
85
87
84
89
61
62
83
69
73
79
78
74
81
71
72
77
76
10
ELECTROMETER

USP # 3,626,606 (12-14-71): Method & Apparatus for Generating a Dynamic Force Field

Wallace, Henry W.

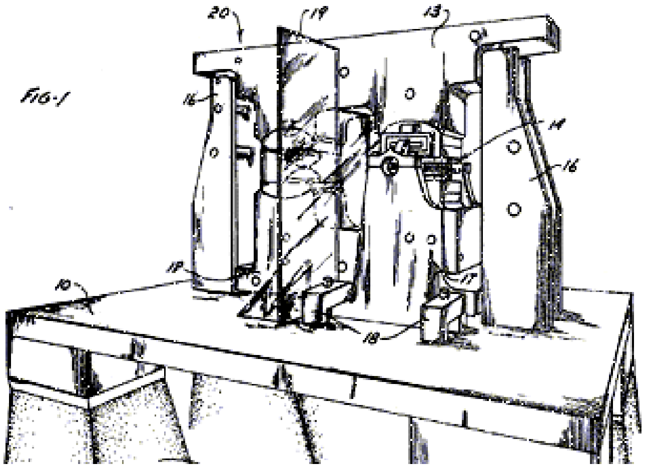
20
19
13
FIG-1
16
14
16
17
10
18
17

USP # 3,626,605 (12-14-71): Method & Apparatus for Generating a Secondary G-Force Field

Wallace, Henry W.

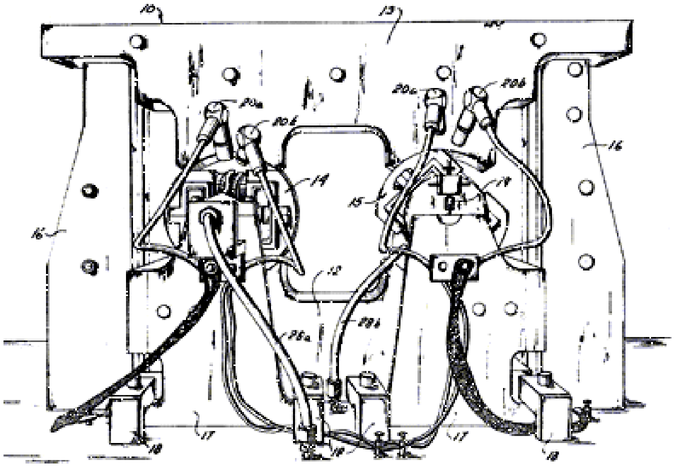
10
13
20a
20b
20c
20b
14
16
15
14
16
12
28a
28b
17
18
18
17
18

USP # 3,610,971 (10-5-71): All-Electric Motional Electric Field Generator

Hooper, William J.

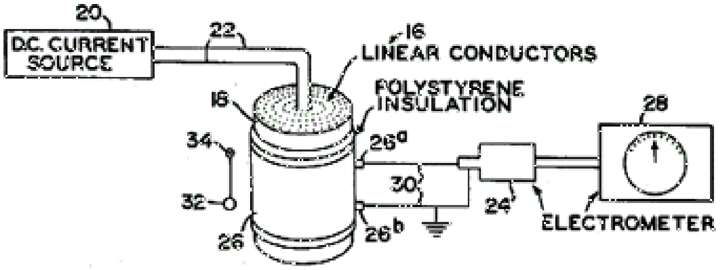
20
D.C. CURRENT SOURCE
22
16
LINEAR CONDUCTORS
16
POLYSTYRENE INSULATION
28
34
26a
30
32
24
ELECTROMETER
26b
26

USP # 3,518,462 (6-30-70): Fluid Flow Control System

Brown, Thomas T.

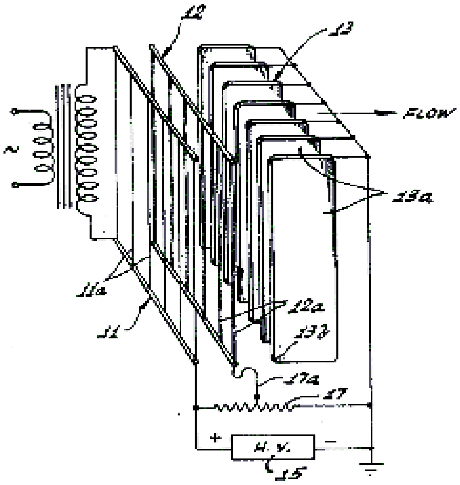
12
13
FLOW
13a
11a
11
12a
13?
17a
17
+
H.V.
−
15

USP # 3,504,868 (4-7-70): Space Propulsion System
Engelberger, Joseph F.

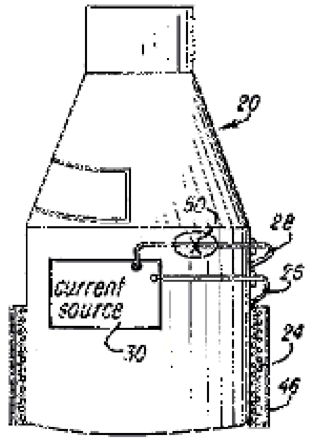
20
50
28
25
current source
30
24
46

USP # 3,464,207 (9-2-69): Quasi-Corona-Aerodynamic Vehicle
Okress, Ernest C.

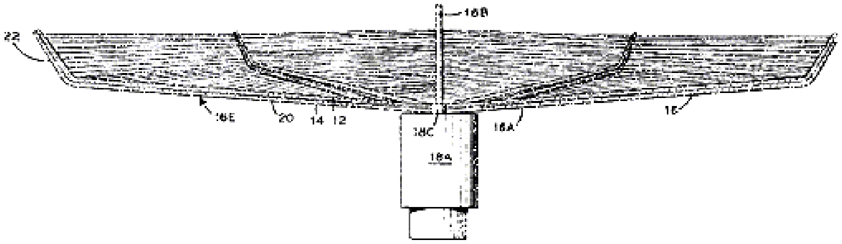
16B
22
16E
20
14
12
18C
18A
16A
16

USP # 3,322,374 (5-30-67): Magnetohydrodynamic Propulsion Apparatus
King, James F., Jr.

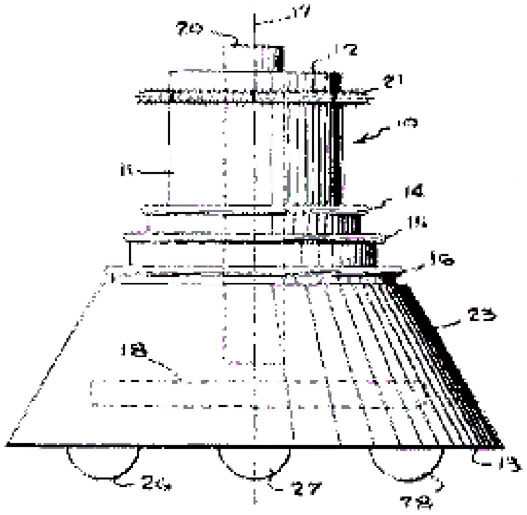
17
20
12
21
10
11
14
15
16
23
18
13
26
27
28

USP # 3,263,1O2 (7-26-66): Electrical Thrust Producing device
Bahnson, Agnew H., Jr.

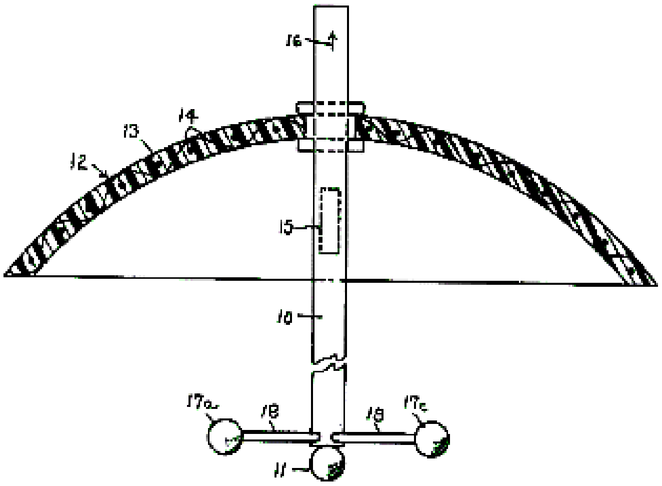
16
14
13
12
15
10
17a
18
18
17c
11

USP # 3,227,901 (1-4-66): Electrical Thrust Producing device
Bahnson, Agnew H., Jr.

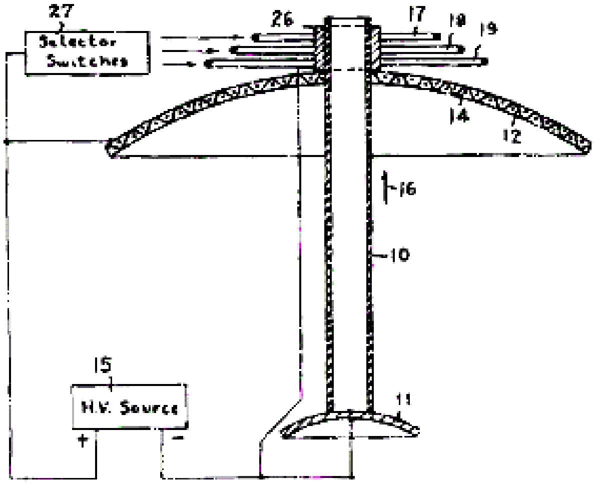
27
Selector Switches
26
17
18
19
14
12
16
10
15
H.V. Source
+
−
11

USP # 3,223,038 (7-26-66): Electrical Thrust Producing Device
Bahnson, Agnew H., Jr.

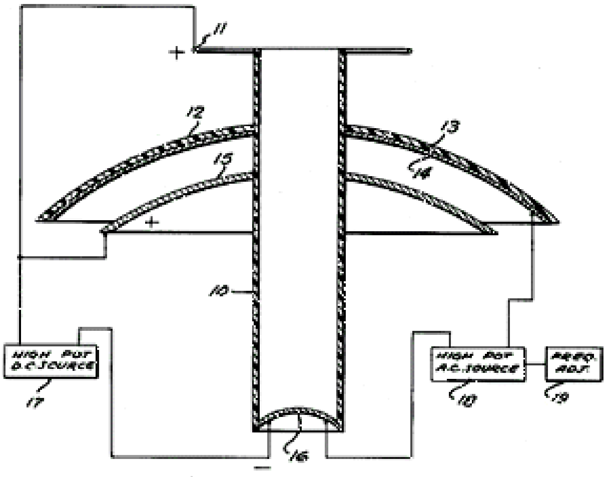
11
+
12
13
15
14
+
10
HIGH POT D.C. SOURCE
17
HIGH POT A.C. SOURCE
18
FREQ. ADJ.
19
16
−

USP # 3,187,206 (6-1-65): Electrokinetic Apparatus

Brown, Thomas T.

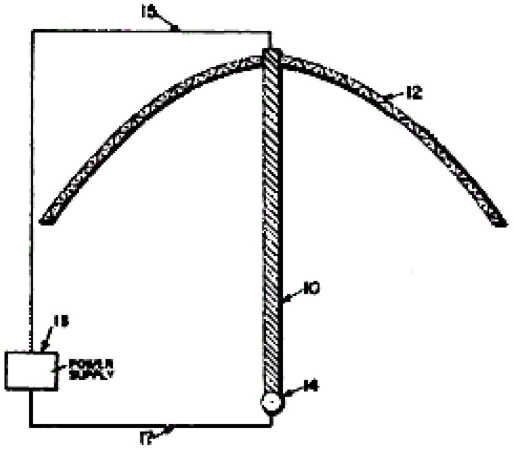
15
12
10
13
POWER SUPPLY
14
17

USP # 3,177,654 (4-13-65): Electric Aerospace Propulsion System

Gradecak, Vjekoslav

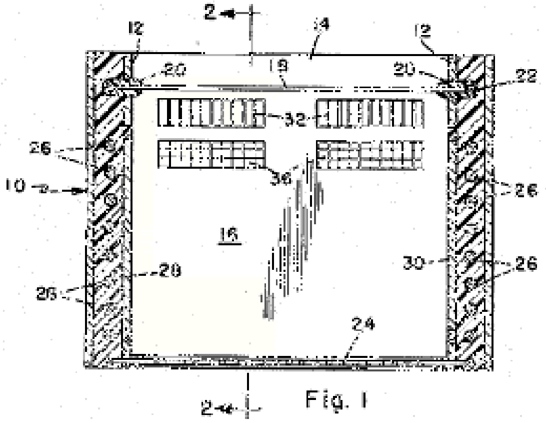
2
12
14
12
20
18
20
22
32
26
36
10
26
16
30
28
26
26
24
2

Fig. 1

USP # 3,162,398 (12-22-64): Magnetohydrodynamic Control Systems

Clauser, Milton U., *et al.*

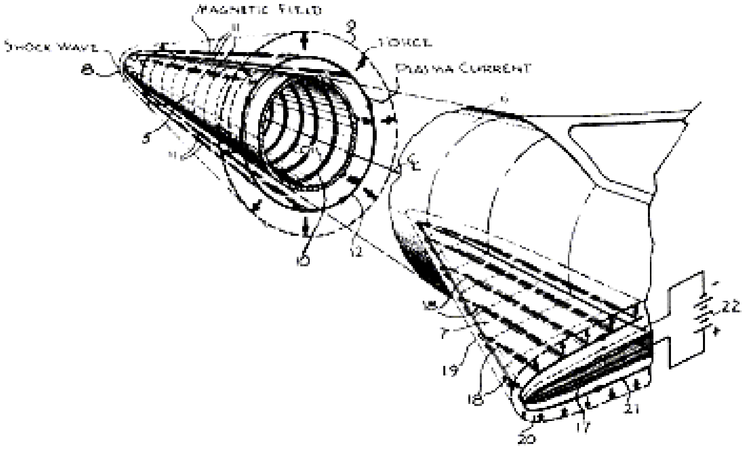
MAGNETIC FIELD
11
9
FORCE
PLASMA CURRENT
SHOCK WAVE
8
6
5
11
10
12
22
7
19
18
20
17
21

USP # 3,130,945 (4-28-64): Ionocraft
De Seversky, Alexander P.

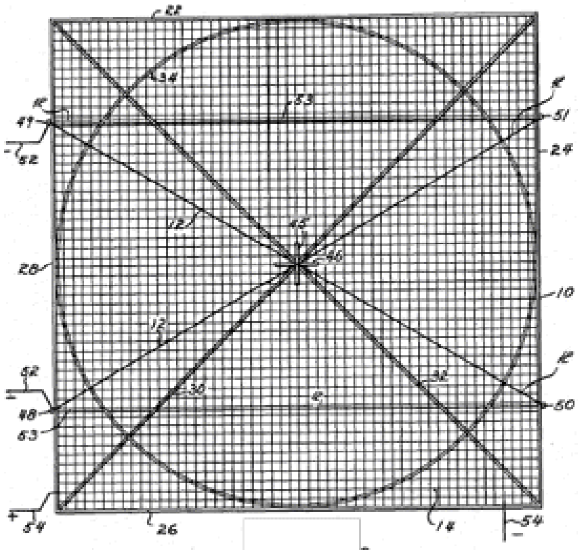
22
34
53
12
47
51
52
24
12
45
46
28
10
12
12
52
30
32
48
50
53
54
26
14
54

USP # 3,120,363 (2-4-64): Flying Apparatus
Hagen, Glenn E.

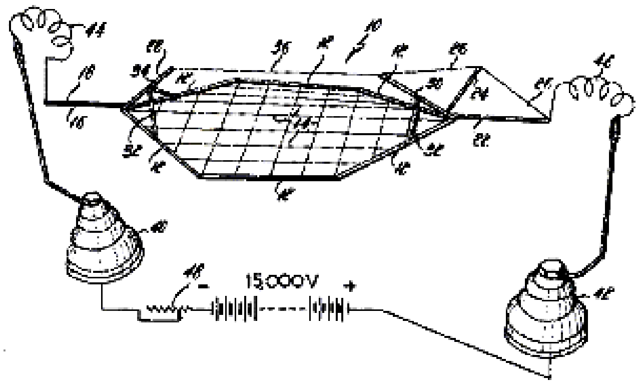
44
10
36
12
28
26
18
34
14
30
24
24
46
16
32
12
14
32
22
40
48
15,000V
42

USP # 3,106,058 (10-8-63): Propulsion System
Rice, Warren A.

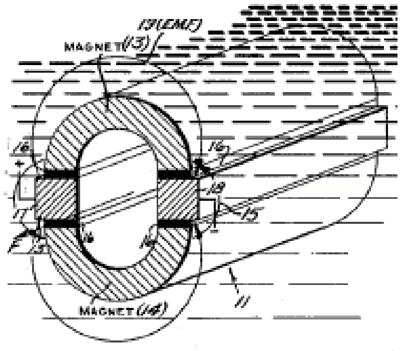

19(EMF)
MAGNET (13)
16
16
17
18
15
F
14
14
11
MAGNET (14)

USP # 3,095,167 (6-25-63): Apparatus for the Promotion and Control of Vehicular Flight
Dudley, Horace C.

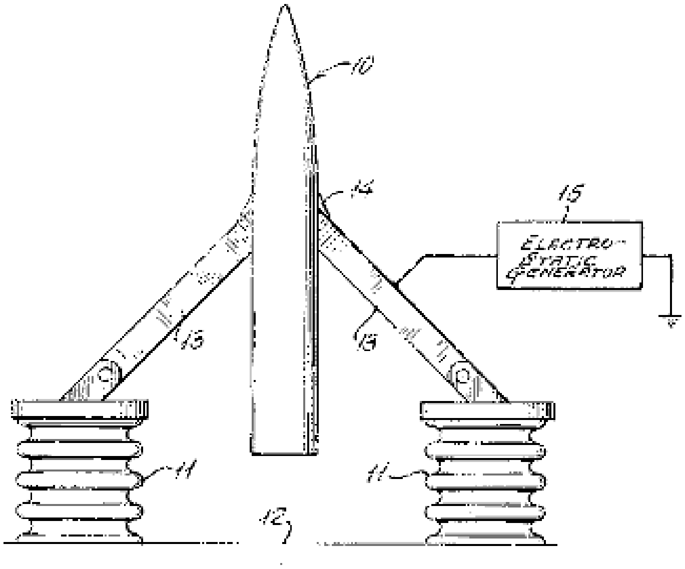

10
14
15
ELECTRO-STATIC GENERATOR
13
13
11
11
12

USP # 3,095,163 (6-25-63): Ionized Boundary Layer Fluid Pumping System
Hill, Gilman A.

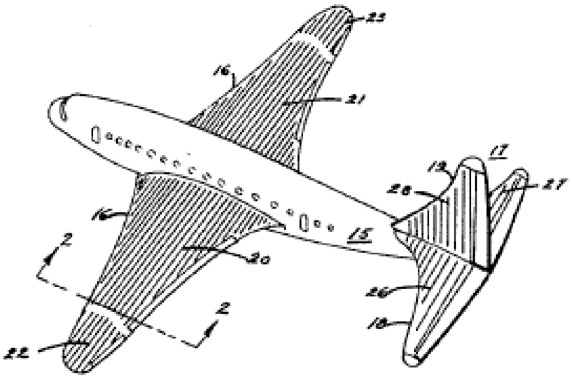

23
16
21
17
19
28
27
16
15
2
20
2
26
18
22

USP # 3,071,705 (1-1-63): Electrostatic Propulsion Means
Coleman, William J., *et al.*

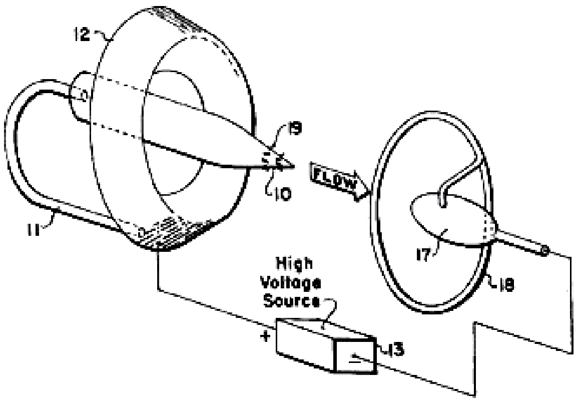
12
19
10
FLOW
11
17
18
High
Voltage
Source
+
13

USP # 3,022,430 (2-20-62): Electrokinetic Generator
Brown, Thomas T.

FIG. 1

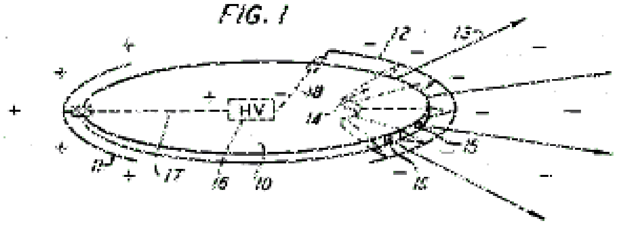
12
13
HV
14
18
15
11
17
16
10

USP# 3,018,394 (1-23-62): Electrokinetic Transducer
Brown, Thomas T.

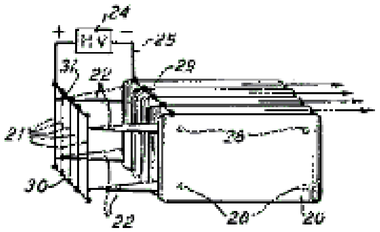
24
HV
25
31
22
29
21
28
30
22
20

USP # 2,958,790 (11-1-60): Electrical Thrust Producing apparatus
Bahnson, Agnew H., Jr.

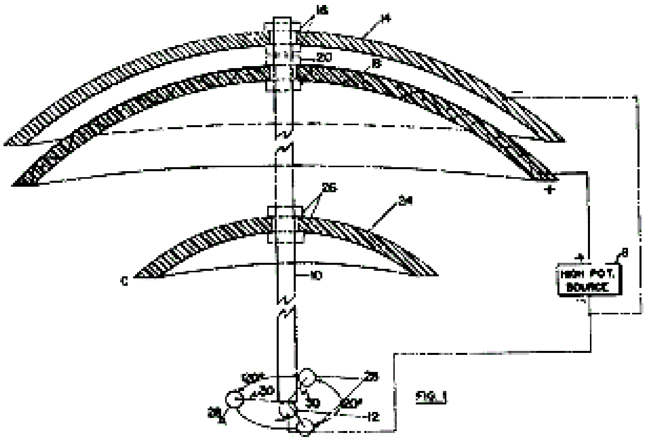
16
14
20
18
26
24
10
C
8
HIGH POT. SOURCE
28
30
120°
12
FIG. 1

USP # 2,949,550 (4-16-60): Electrokinetic Apparatus
Brown, Thomas T.

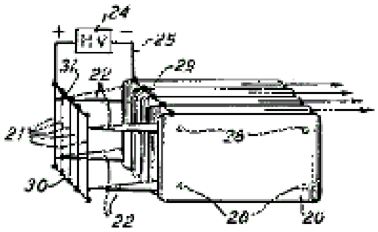

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

---

rexresearch.com

US Patents: Disc Aircraft

(Propellor & Jet-Driven Wingless Aerodynes, Lenticular Aircraft, Discoid Aircraft, &c.
This file does not include force field propulsion & space drives)

---

fetchingstock shutterstock

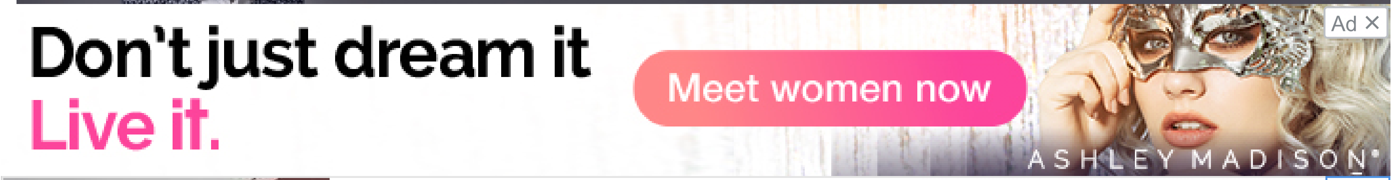
Don't just dream it Live it. Meet women now ASHLEY MADISON Ad

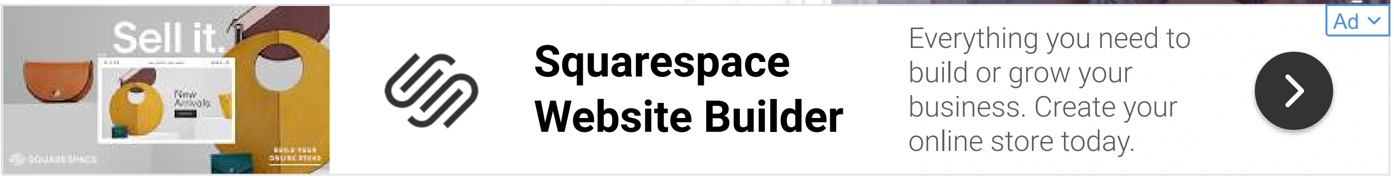
Sell it. Squarespace Website Builder Everything you need to build or grow your business. Create your online store today. Ad

---

Acknowledgements: Visit Man-Made Flying Saucers (http://www.aspubs.com/mmsaucers/) and World Space Drive Archives for a thorough treatment of this subject, plus: anti-gravity and space drive (unidirectional motion rectifiers) patents from around the world, and lots of photos, documentation, &c...

USP # 6,270,036 (8-7-01) ~ Blown-Air Lift Generating Rotating Airfoil Aircraft

Lowe, Charles S., Jr.

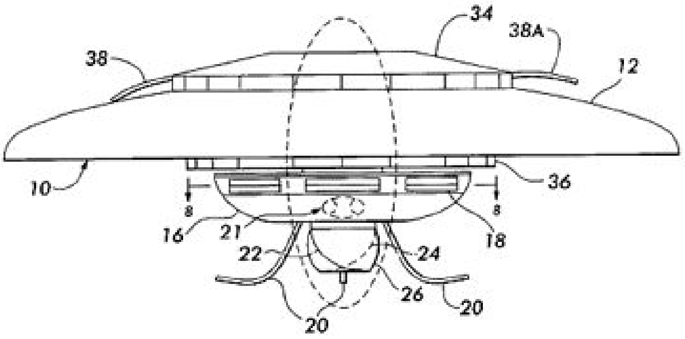

USP # 6,254,032 (7-3-01) ~ Aircraft, &c...
Bucher, Franz

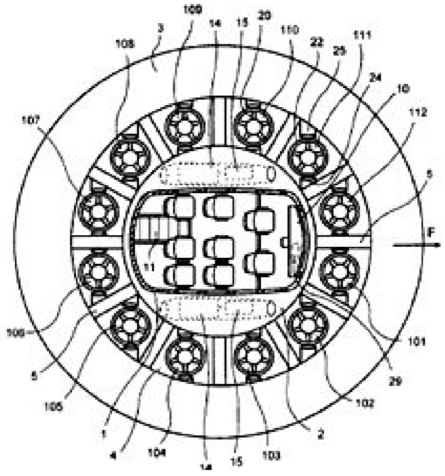
109
3
14
15
20
110
22
25
111
108
24
10
112
107
5
F
11
106
101
5
29
105
102
1
2
4
104
14
15
103

USP # 6,179,247 (1-30-01) ~ Personal Air Transport
Milde, Jr., Karl F.

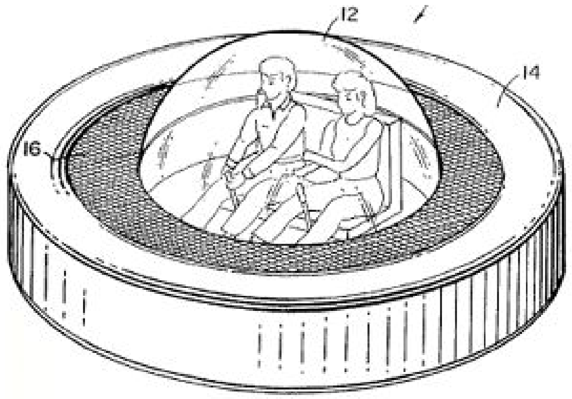

USP # 6,113,029 (9-5-00) ~ Aircraft Capable of Hovering Flight
Salinas, Luis A.

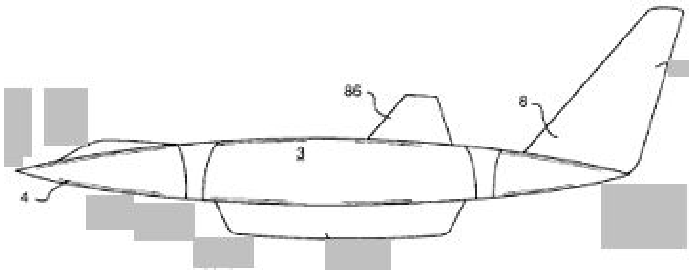

USP # 6,082,478 (7-4-00) ~ Lift-Augmented Ground Effect Platform
Walter, William. C., ***et al.***

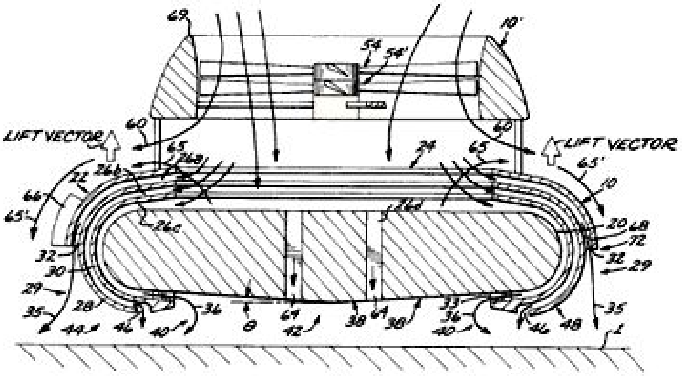
69
10'
54
54'
LIFT VECTOR
60
65
24
65'
10
20
68
72
32
29
35
30
28
44
46
40
36
8
64
42
38
33
34
48
1

USP # 6,073,881 (6-13-00) ~ Aerodynamic Lift Apparatus
Chen, Chung-ching

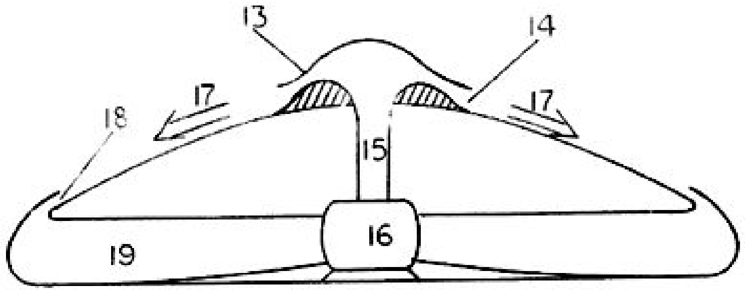
13
14
17
18
17
15
16
19

USP # 6,068,219 (5-30-00) ~ Single-Surface Multi-Axis Aircraft Control
Arata, Allen A.

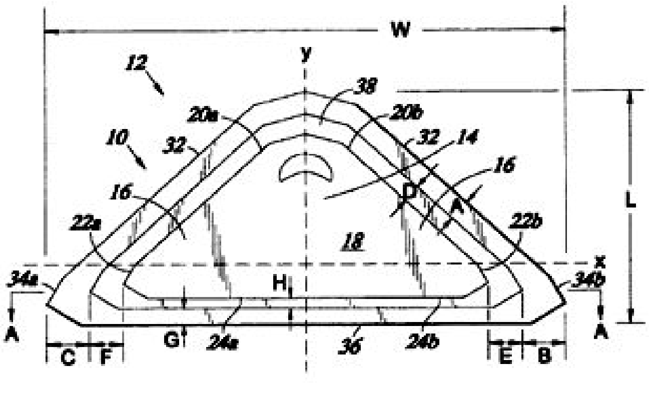
W
y
12
38
20a
20b
10
32
32
14
16
16
D
A
22a
22b
18
x
L
34a
34b
H
A
A
G
24a
36
24b
C
F
E
B

USP # 6,053,451 (4-25-00) ~ Remote-Control Flight Vehicle Structure
Yu, Shia-Giow

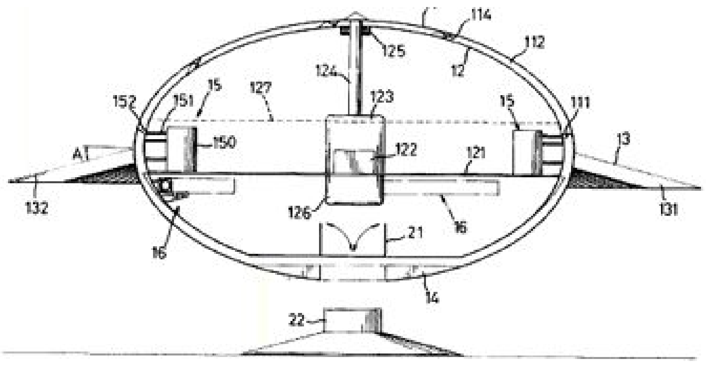
114
112
125
12
124
123
127
15
151
152
15
111
A
150
122
121
13
132
131
126
16
16
21
14
22

USP # 6,050,520 (4-18-00) ~ VTOL Aircraft
Kirla, Stanley J.

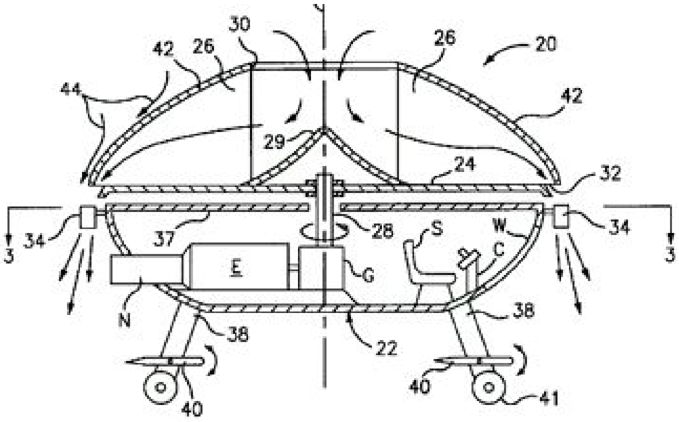
30
42
26
26
20
44
42
29
24
32
34
37
28
S
W
34
3
3
C
E
G
N
38
38
22
40
40
41

USP # 6,016,991 (1-25-00) ~ Evacuated Rotating Envelope Aircraft
Lowe, Jr., Charles S.

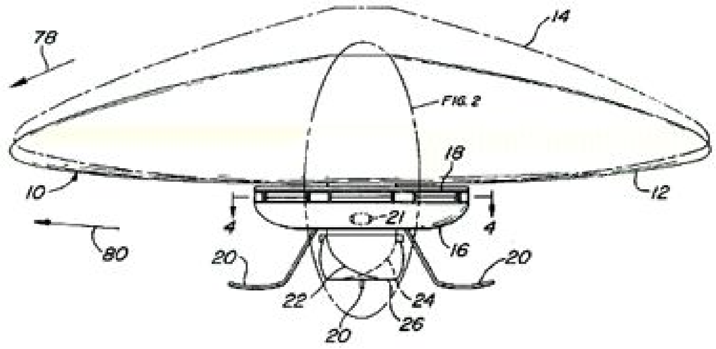
78
14
FIG. 2
18
10
12
4
4
21
16
80
20
20
22
24
26
20

USP # 5,971,321 (10-26-99) ~ Body-Lift Airplane Assembly
Libengood, Ronald L.

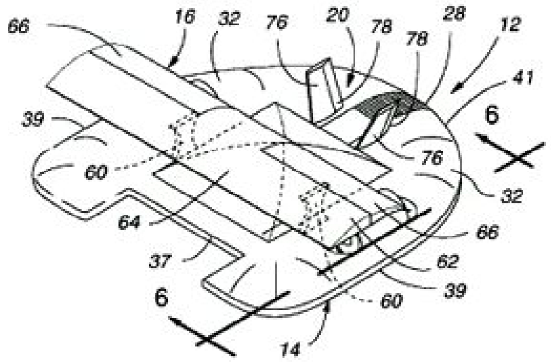
66
16
32
76
20
78
78
28
12
41
39
6
60
76
32
64
66
62
37
39
6
60
14

USP # 5,895,011 (4-20-99) ~ Turbine Airfoil Lifting Device
Gubin, Daniel

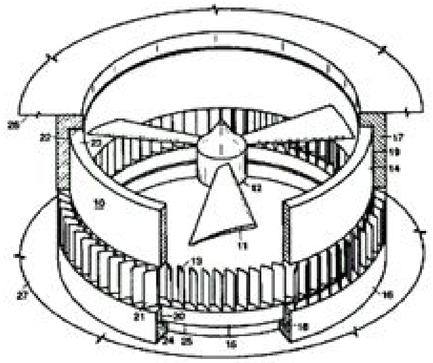
26
22
23
17
19
14
12
10
11
13
27
16
21
20
18
24
25
15

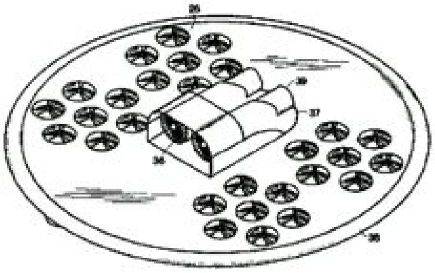
26
39
37
34
35

USP # 5,881,970 (3-16-99) ~ Levity Aircraft Design
Whitesides, Carl W.

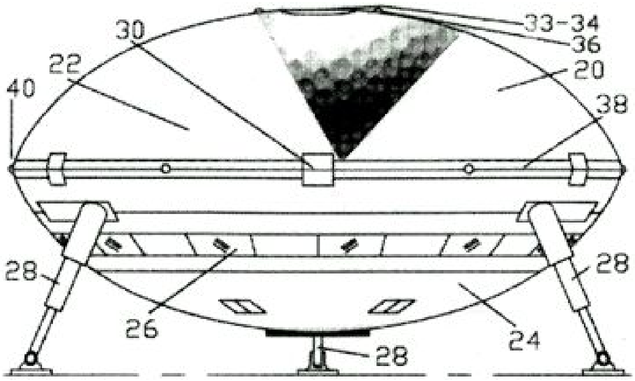
30
22
40
33-34
36
20
38
28
26
28
24
28

USP # 5,836,543 (11-17-98) ~ Discus-Shaped Aerodyne Vehicle...
Kunkel, Klaus

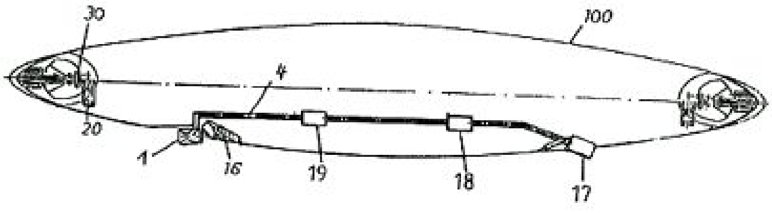
30
100
4
20
1
16
19
18
17

USP # 5,836,542 (11-17-98) ~ Flying Craft &c...
Burns, David J.

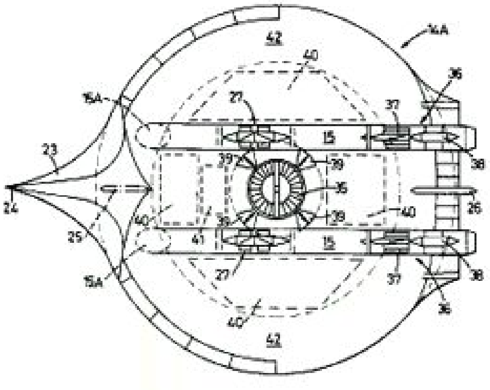
42
14A
40
36
15A
27
37
15
23
39
39
38
35
24
39
40
26
40
25
41
15
38
27
37
15A
40
36
42

USP # 5,803,199 (9-8-98) ~ Lift-Augmented Ground Effect Platform
Walter, William C.

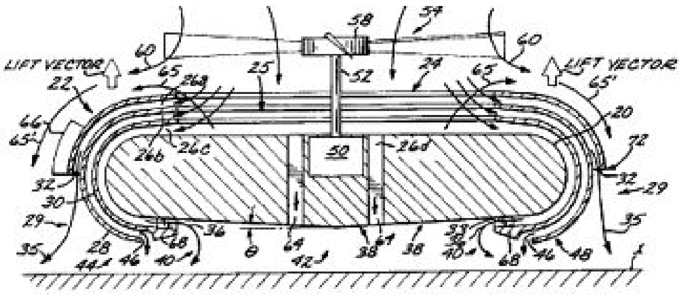
58
54
LIFT VECTOR
40
52
24
65
40
LIFT VECTOR
22
65
26a
25
65
65'
20
66
65'
72
26b
26c
50
26d
32
32
29
30
29
35
35
36
6
64
38
61
38
33
36
40
68
46
48
28
44
46
40
42
1

USP # 5,730,391 (3-24-98) ~ Universal Fluid-Dynamic Body for Aircraft & Watercraft
Miller, Jr., John A., *et al.*

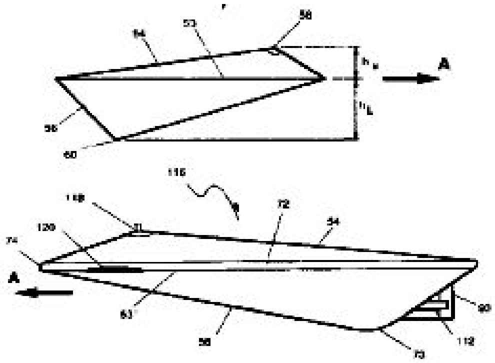
58
54
53
A
56
60
116
118
72
54
120
74
A
63
56
70
90
112

USP # 5,730,390 (3-24-98) ~ Reuseable Spacecraft
Plichta, Peter & Buttner, Walter

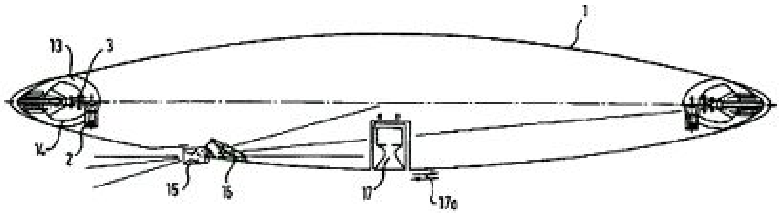
1
13
3
14
2
15
16
17
17a

USP # 5,653,404 (8-5-97) ~ Disc-Shaped Submersible Aircraft
Ploskin, Gennady

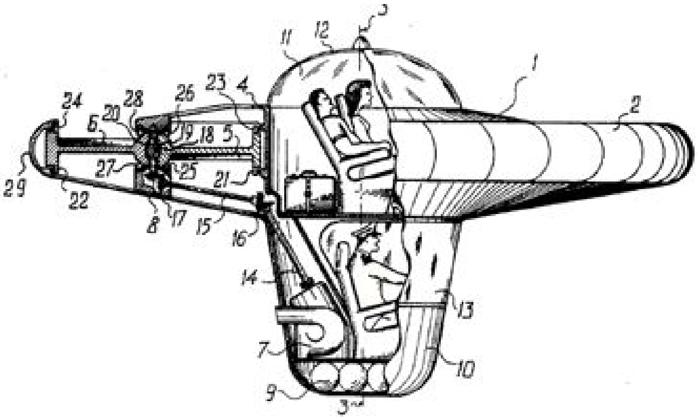
3
11
12
1
2
24
28
20
26
23
4
6
19
18
5
27
25
21
29
22
8
17
15
16
14
13
7
10
9
3

USP # 5,520,355 (5-28-96) ~ Three-Wing Circular Planform Body
Jones, Jack M.

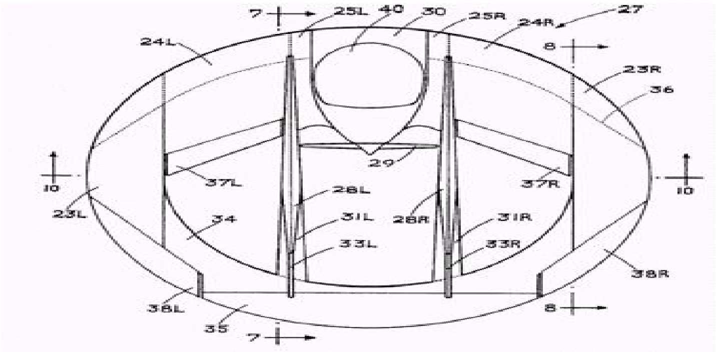
7
25L
40
30
25R
24R
27
24L
8
23R
36
29
10
37L
37R
10
28L
23L
31L
28R
31R
34
33L
33R
38R
38L
8
35
7

USP # 5,351,911 (10-4-94) ~ VTOL Flying Disc
Neumayr, George A.

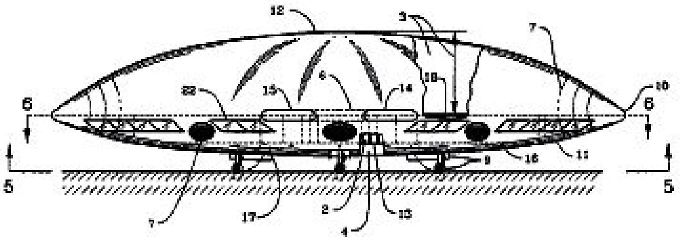
12
3
7
10
6
6
22
15
8
14
18
5
5
9
16
11
7
17
2
13
4

USP # 5,344,100 (9-6-94) ~ Vertical Lift Aircraft
Jaikaran, Allan

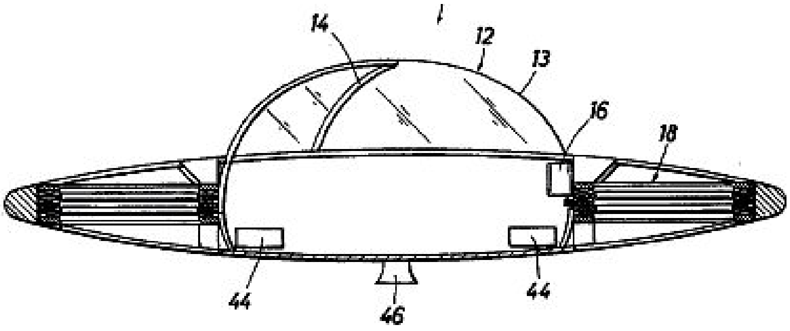
1
14
12
13
16
18
44
44
46

USP # 5,318,248 (6-7-94) ~ Vertical Lift Aircraft
Zielonka, Richard H.

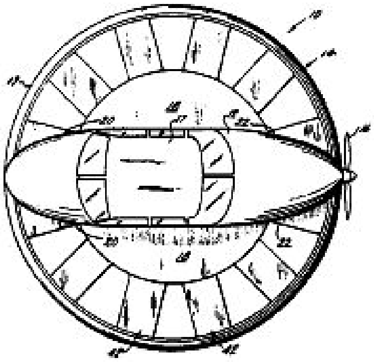

USP # 5,303,879 (4-19-94) ~ Aircraft with a Ducted Fan in a Circular Wing

Bucher, Franz

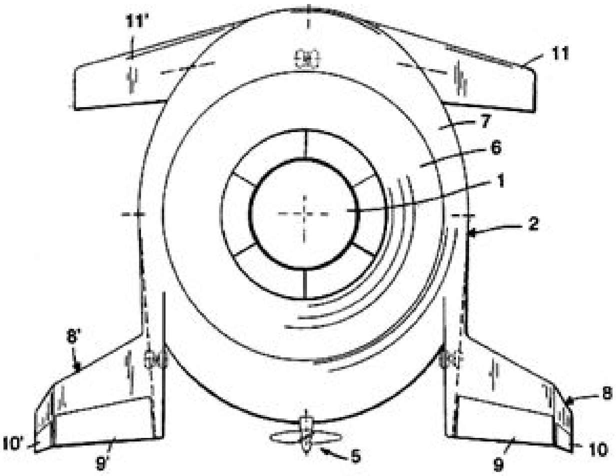
11'
11
7
6
1
2
8'
8
10'
9'
5
9
10

USP # 5,295,571 (11-9-93) ~ Aircraft with Gyroscopic Stabilization System

Blazquez, Jose M.

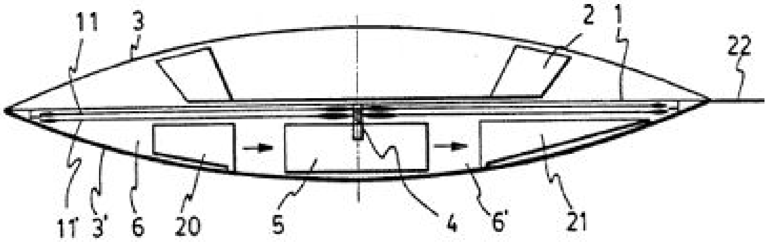
11
3
2
1
22
11'
3'
6
20
5
4
6'
21

USP # 5,213,284 (5-25-93) ~ Disc Planform having Vertical Flight Capability

Webster, Stephen N.

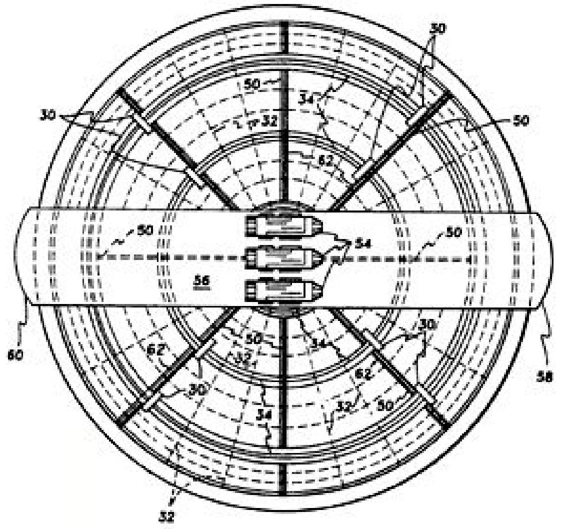
30
30
50
34
32
50
62
50
54
50
56
60
58
30
62
50
32
34
62
30
30
34
32
30
32

USP # 5,203,521 (4-20-93) ~ Annular Body Aircraft

Day, Terence

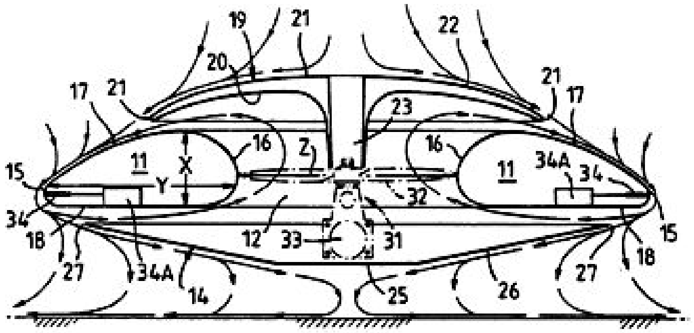
19
21
22
20
21
21
17
17
23
16
16
Z
11
X
11
34A
34
15
Y
32
34
18
12
33
31
15
18
27
34A
27
14
25
26

USP # 5,178,344 (1-12-93) ~ VTOL Aircraft

Dlouhy, Vaclav

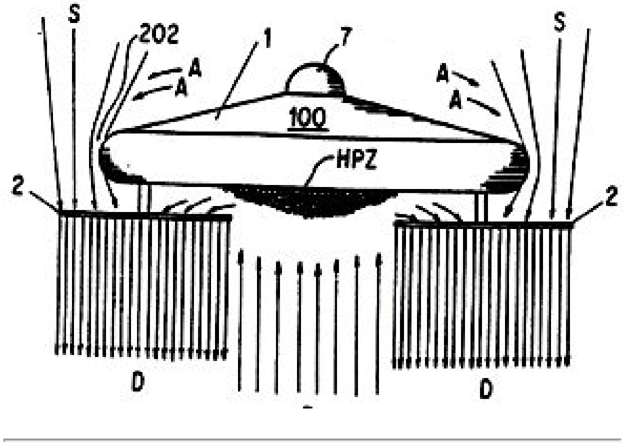
S
202
1
7
S
A
A
A
A
100
HPZ
2
2
D
D

USP # 5,170,963 (12-15-92) ~ VTOL Aircraft

Beck, Jr., August H.

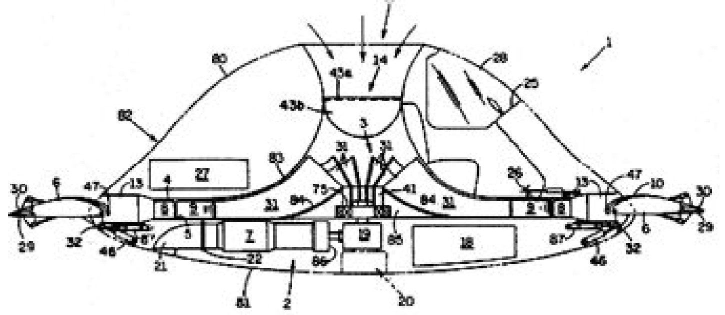
1
80
14
43a
43b
28
25
82
3
31
31
83
27
4
26
13
47
13
47
10
30
6
30
75
41
84
31
84
31
8
9
29
32
5
7
19
18
6
29
48
21
22
86
85
87
46
32
81
2
20

USP # 5,149,012 (9-22-92) ~ Turbocraft
Valverde, Rene L.

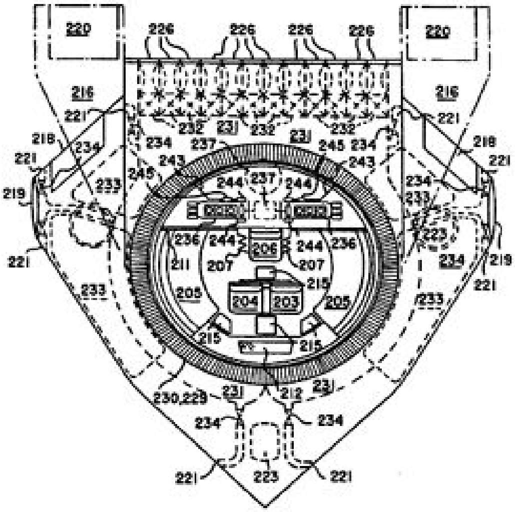
220
226
226
226
226
220
216
216
221
232
231
232
231
232
221
218
234
237
245
234
218
221
243
243
221
219
233
245
244
237
244
234
233
219
221
236
244
206
244
236
223
234
221
211
207
207
215
233
205
204
203
205
233
215
215
230,229
231
212
231
234
234
221
223
221

USP # 5,115,996 (5-26-92) ~ VTOL Aircraft
Moller, Paul S.

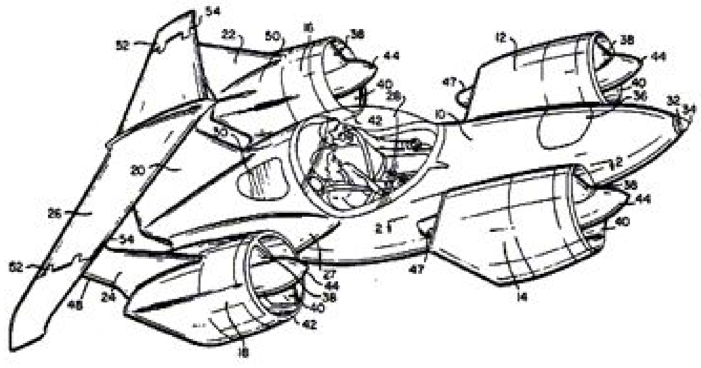
54
52
22
50
16
38
44
12
38
44
40
28
47
40
36
32
34
42
10
20
12
38
44
26
40
21
54
27
52
44
36
47
48
24
40
42
14
18

USP # 5,102,066 (4-7-92) ~ VTOL Aircraft
Daniel, William H.

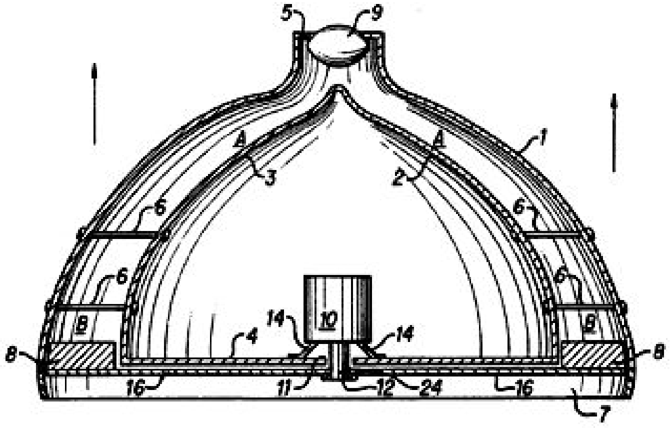
5
9
A
A
1
3
2
6
6
6
6
B
B
10
14
14
4
8
8
16
11
12
24
16
7

USP # 5,064,143 (11-12-91) ~ Aircraft Having a Pair of Counter Rotating Rotors
Bucher, Franz

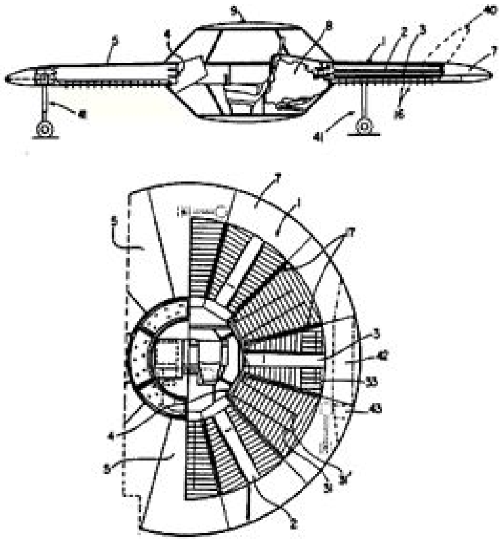
9
4
5
8
1
2
3
40
7
6
41
16
7
1
17
5
3
42
33
43
4
5
31
31'
2

USP # 5,054,713 (10-8-91) ~ Circular Airplane
Langley, Lawrence W., *et al.*

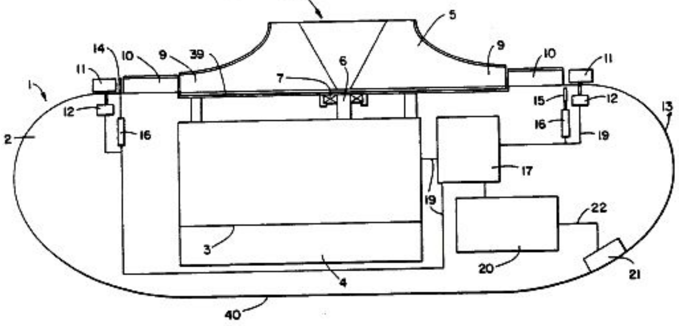
8
5
14
10
9
39
9
10
11
11
7
6
1
12
15
12
13
2
16
16
19
17
19
22
3
20
21
4
40

USP # 5,046,685 (9-10-91) ~ Fixed Circular Wing Aircraft
Bose, Phillip R.

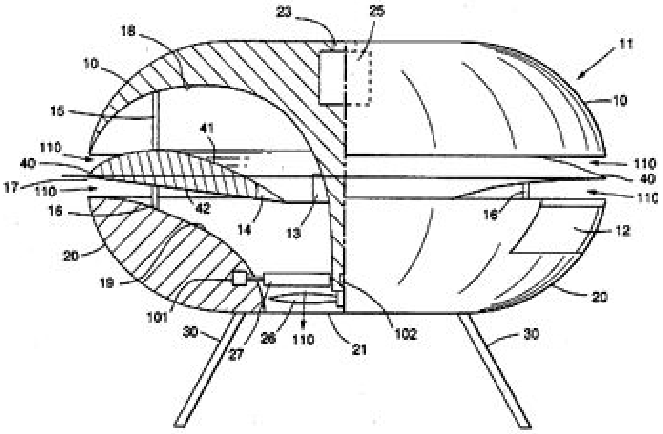
23
25
18
11
10
10
15
41
110
110
40
40
17
110
110
42
16
14
13
16
12
20
19
20
101
102
30
30
27
26
110
21

USP # 5,039,031 (8-13-91) ~ Turbocraft
Valverde, Rene L.

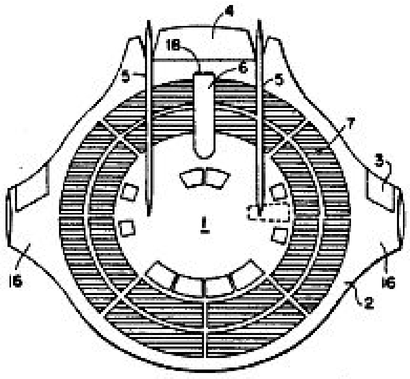
4
18
6
5
5
7
3
1
16
16
2

USP # 4,976,395 (12-11-99) ~ Heavier-Than-Air Disk-Type Aircraft
von Kozierowski, Joachim

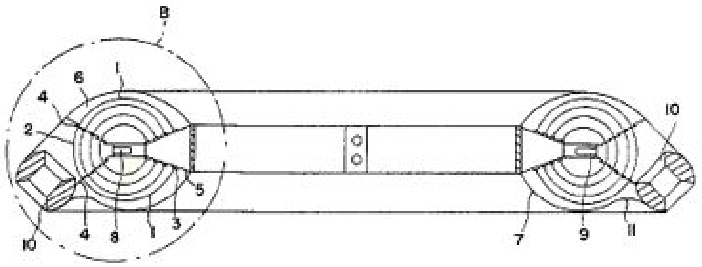
B
6
1
4
2
10
5
10
4
8
1
3
7
9
11

USP # 4,955,962 (9-11-99) ~ Remote Control Flying Saucer
Mell, Christian

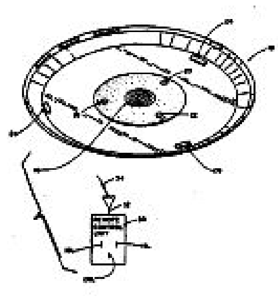

USP # 4,941,628 (7-19-90) ~ Lift Generating Apparatus, &c.
Sakamoto, Yujiro, *et al.*

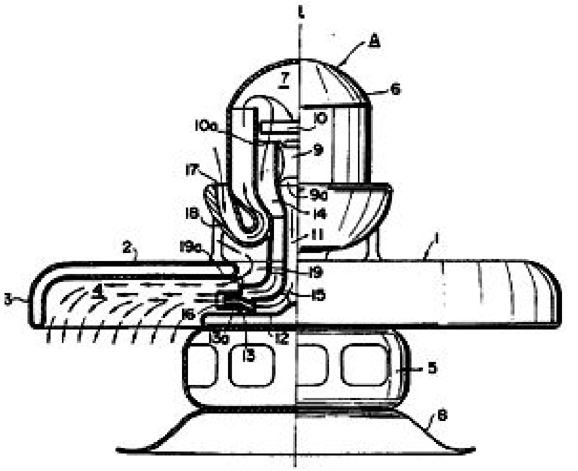
L
A
7
6
10
10a
9
17
9a
14
18
11
2
19a
1
19
4
15
3
16
13a
13
12
5
8

USP # 4,824,048 (4-25-89) ~ Induction Lift Flying Saucer
Kim, Kyusik

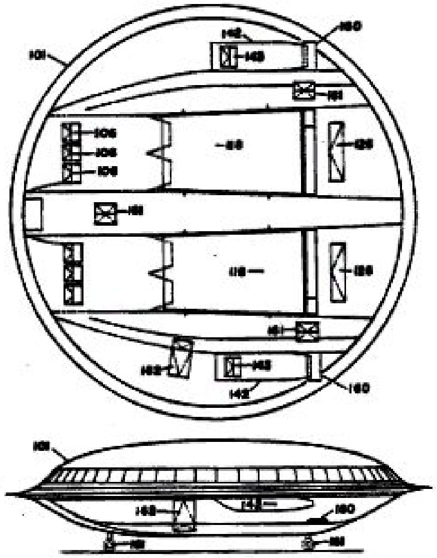
101
142
160
106
106
106
142
160
101

USP # 4,804,156 (2-14-89) ~ Circular Aircraft
Harmon, Rodney D.

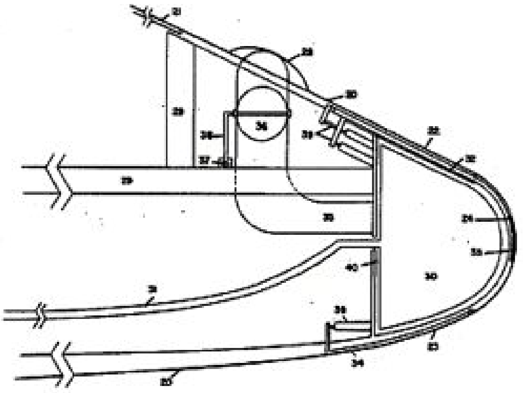

USP # 4,796,836 (1-10-89) ~ Lifting Engine for VTOL Aircrafts

Buchelt, Benno

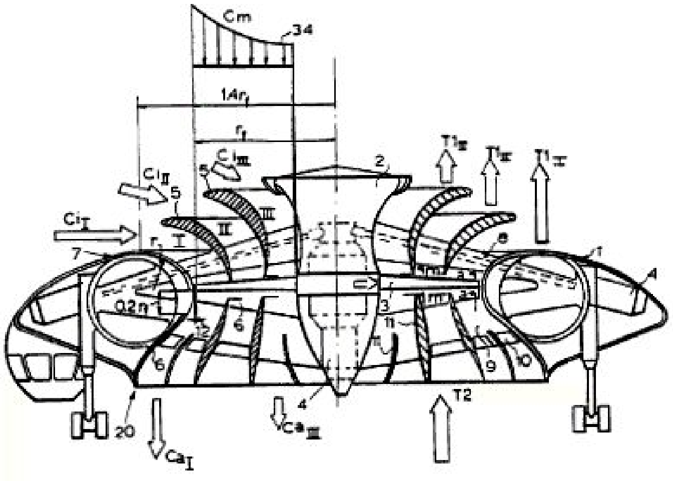
Cm
34
1Ar1
r1
Ci_II
Ci_III
Ci_I
T1_III
T1_II
T1_I
5
2
I
II
III
r1
7
8
1
4
0.2r1
3
6
11
12
9
10
20
4
T2
Ca_I
Ca_III

USP # 4,795,111 (1-3-89) ~ Robotic or Remotely Controlled Flying Platform

Moller, Paul S.

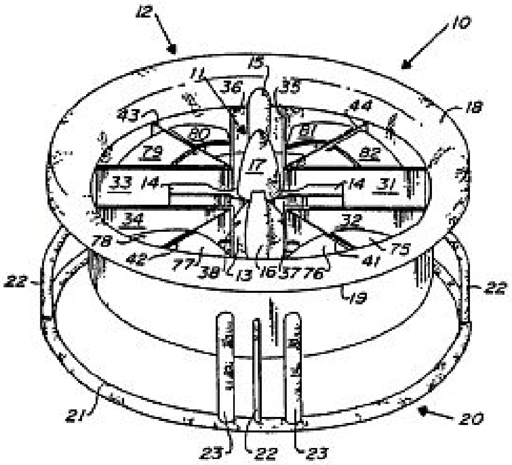
12
10
11
15
36
35
18
43
44
80
81
79
82
33
14
17
14
31
34
32
78
75
42
77
38
13
16
37
76
41
19
22
22
21
20
23
22
23

USP # 4,214,720 (7-29-80) ~ Flying Disc

DeSautel, Edwin R.

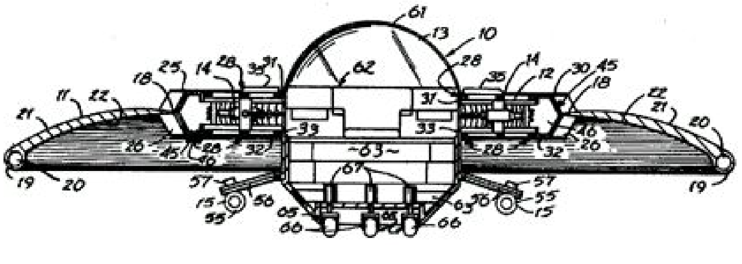
61
13
10
28
62
25
14
28
35
31
18
11
22
21
26
45
46
28
32
33
33
~63~
31
14
12
30
45
18
22
21
20
46
26
28
32
19
20
57
15
55
56
67
65
66
66
65
56
57
55
15
19
20

USP # 4,196,877 (4-8-80) ~ Aircraft

Mutrux, Jean L.

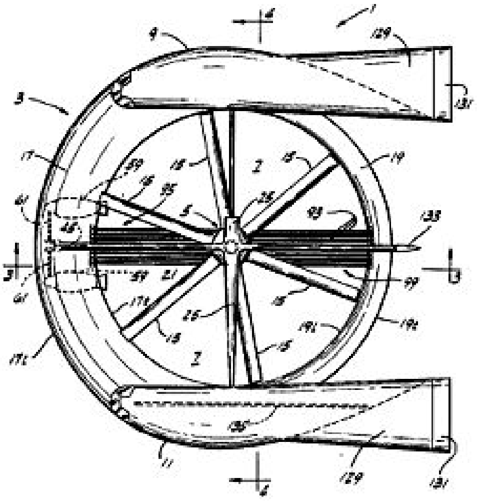
1
9
129
131
3
17
18
2
19
133
99
15
11
129
131

USP # 4,193,568 (3-18-80) ~ Disc-Type Airborne Vehicle

Heuvel, Norman L.

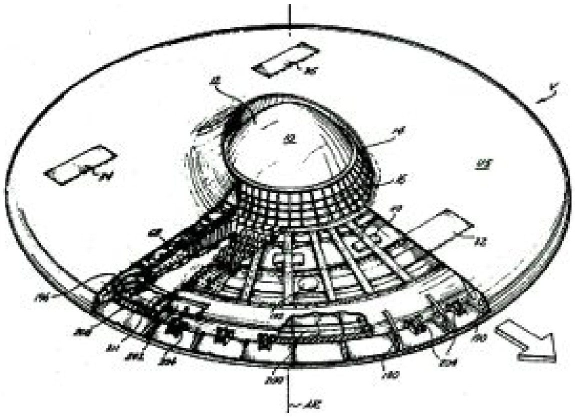

USP # 4,165,848 (8-28-79) ~ Rotary Thrust Device...

Bizzarri, Alfredo

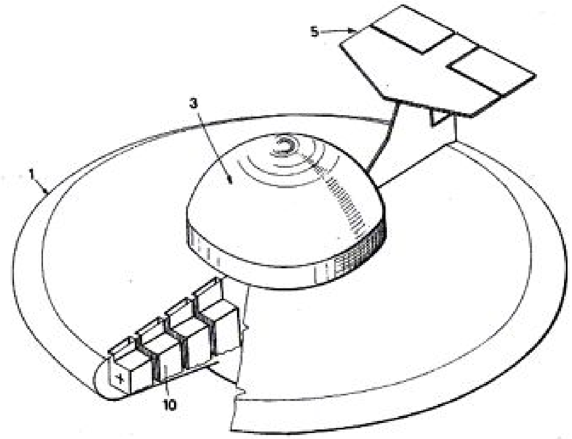

USP # 4,117,992 (10-3-78) ~ Vertical Lift Device

Vrana, Charles K.

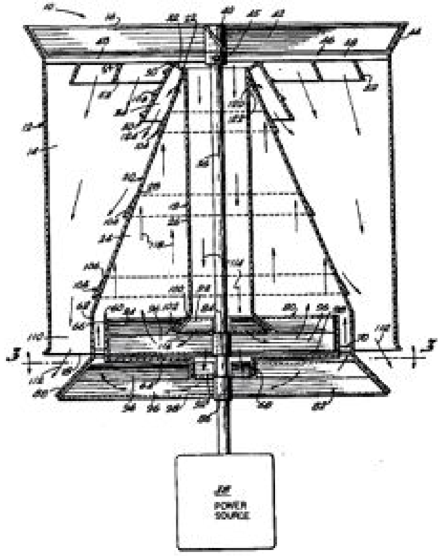
3
3
POWER SOURCE

USP # 4,050,652 (9-27-77) ~ Gyro Foil
DeToia, Vincent D.

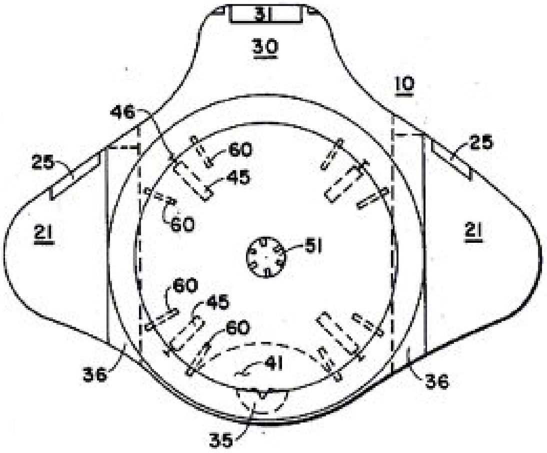
31
30
10
46
60
45
25
25
60
21
21
51
60
45
60
41
36
36
35

USP # 4,023,751 (5-17-77) ~ Flying Ship
Richard, Walter A.

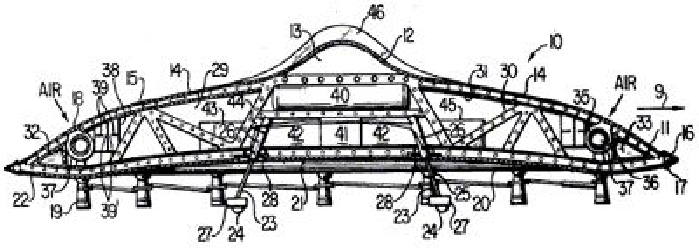
46
13
12
10
AIR
38
15
14
29
31
30
14
AIR
9
18
39
40
45
35
32
43
44
26
42
41
42
25
33
11
16
22
37
19
39
28
21
28
23
25
20
37
36
17
27
24
23
23
24
27

USP # 4,014,483 (3-29-77) ~ Lighter-Than-Air Craft
MacNeil, Roderick M.

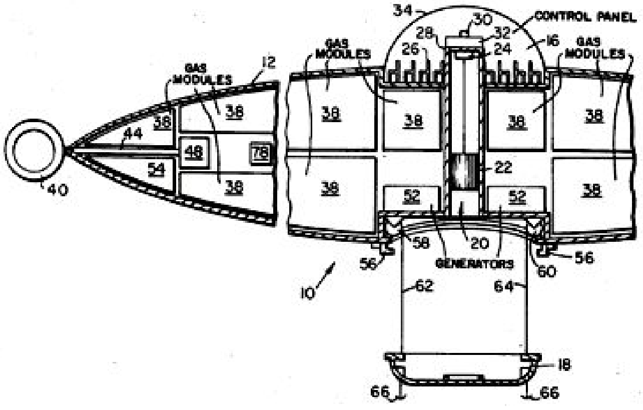
34
30
CONTROL PANEL
28
32
26
24
16
GAS MODULES
GAS MODULES
GAS MODULES
12
38
44
48
78
54
40
22
52
52
58
20
GENERATORS
56
60
56
10
62
64
18
66
66

USP # 3,933,325 (1-20-76) ~ Disc-Shaped Aerospacecraft

Kaelin, Joseph R.

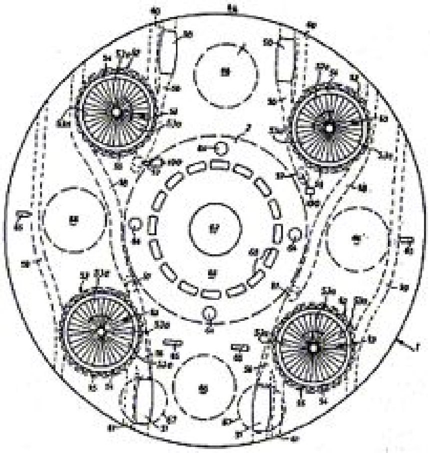

USP # 3,871,602 (3-18-75) ~ Circular Wing Aircraft

Kissinger, Curtis D.

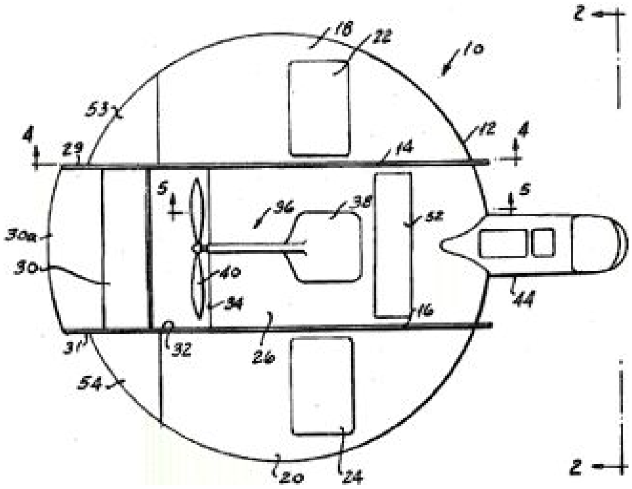
2
18
22
10
53
4
29
14
12
4
5
36
38
52
5
30a
30
40
34
44
16
31
32
26
54
20
24
2

USP # 3,774,865 (11-27-73) ~ Flying Saucer

Pinto, Olympio F.

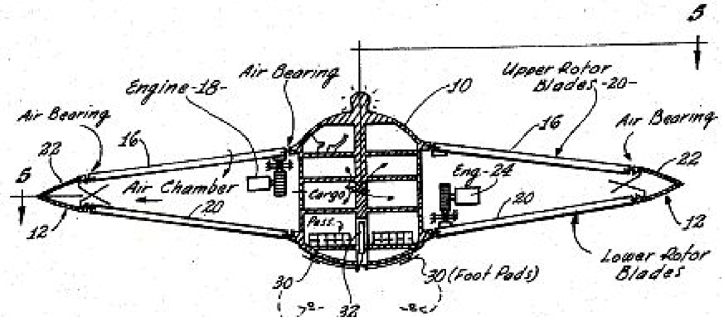
5
Air Bearing
Engine-18-
Air Bearing
-10
Upper Rotor Blades -20-
Air Bearing
22
16
16
22
5
Air Chamber
20
Eng.-24
20
12
12
Cargo
Pass.
Lower Rotor Blades
30
30 (Foot Pads)
32

USP # 3,750,980 (8-7-73) ~ Aircraft with VTOL Capability
Edwards, Samuel L.

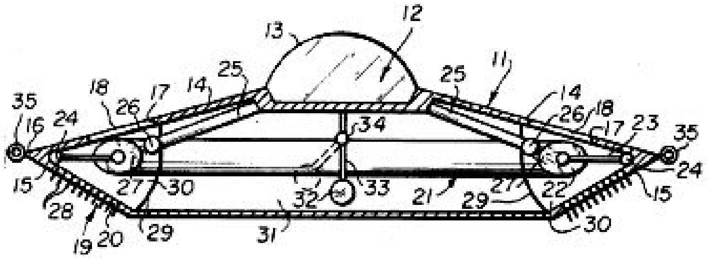
13
12
11
25
25
18
17
14
14
26
18
35
16
24
26
17
23
35
34
24
15
27
30
32
33
21
29
27
22
15
28
19
20
29
31
30

USP # 3,697,020 (10-10-72) ~ Vertical Lift Machine
Thompson, Raymond V.

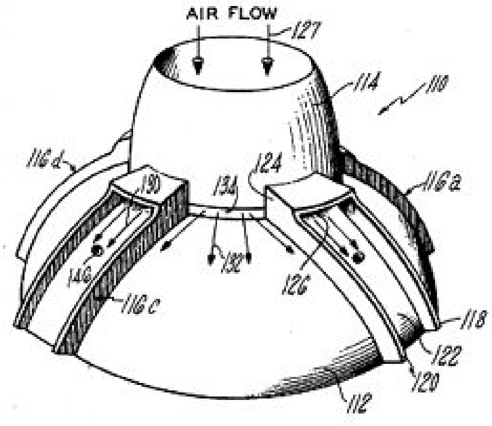
AIR FLOW
127
114
110
116d
124
116a
130
134
146
132
126
116c
118
122
120
112

USP # 3,690,597 (9-12-72) ~ VTOL Aircraft...
Di Martino, Renato

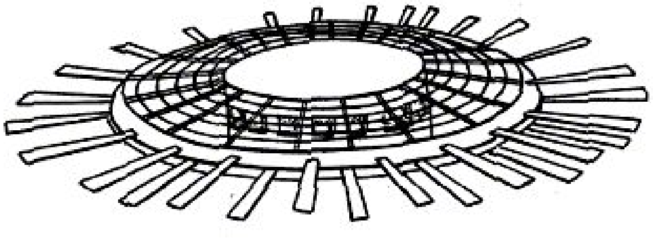

USP # 3,640,489 (2-8-72) ~ VTOL Aircraft
Jaeger, Karl

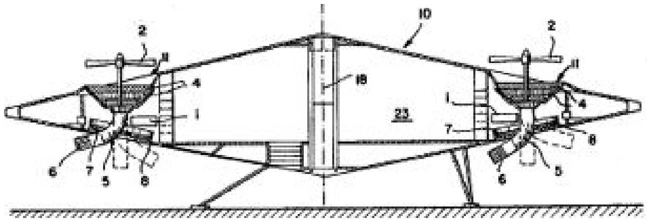
2
10
2
11
4
18
1
23
7
8
6
7
5
8
6
5
4
11
1
7
8

USP # 3,630,470 (12-28-71) ~ VTOL Vehicle
Elliot, Frederick T.

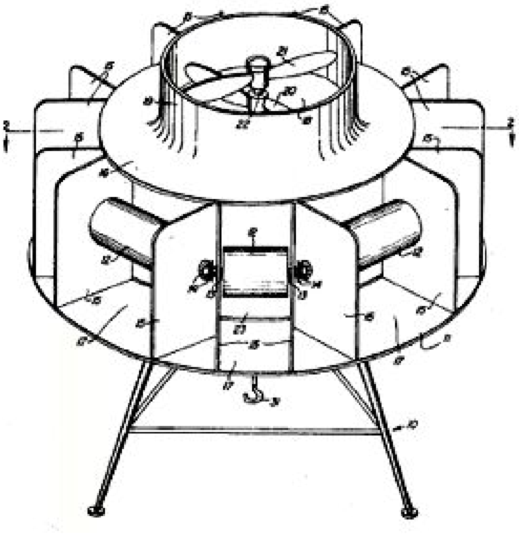

USP # 3,614,030 (10-19-71) ~ Aircraft
Moller, Paul S.

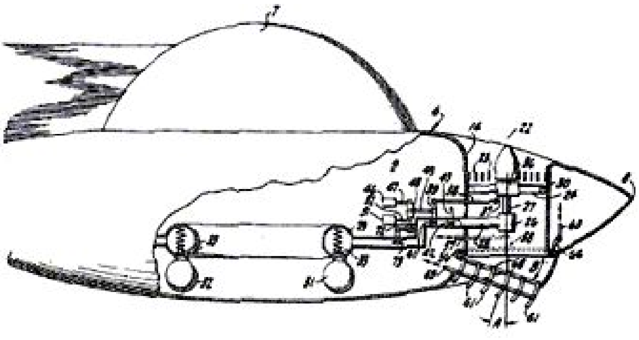

USP # 3,612,445 (10-12-71) ~ Lift Actuator Disc
Phillips, Duan A.

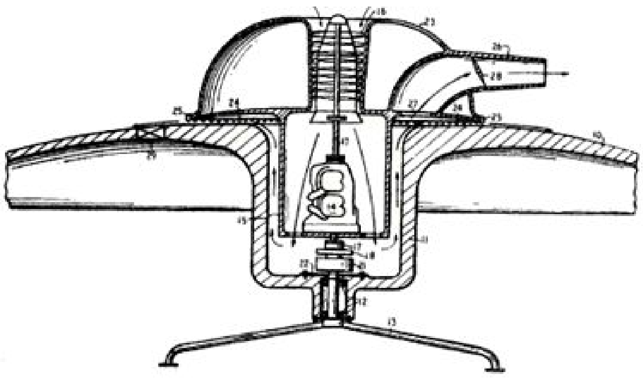

USP # 3,599,902 (8-17-71) ~ Aircraft
Thomley, John W.

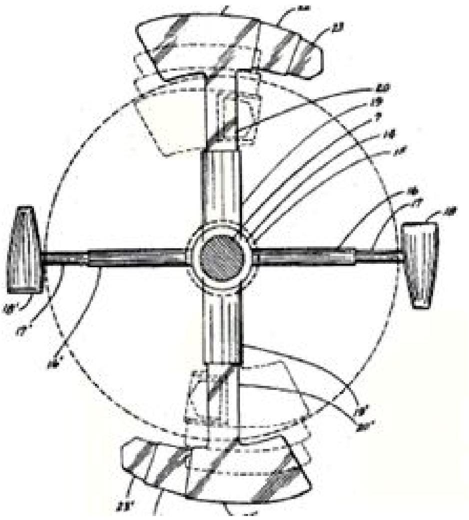
23
20
19
7
14
15
16
17
18
18'
17'
16'
19'
20'
23'

USP # 3,537,669 (11-3-70) ~ Manned Disc-Shaped Flying Craft
Modesti, James N.

FIG.1

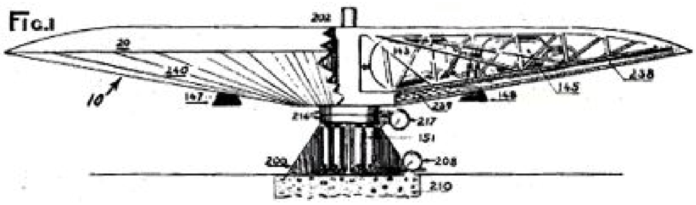
202
20
240
10
147
143
238
145
148
237
216
217
151
200
208
210

USP # 3,503,573 (3-31-70) ~ Disk Flying Craft
Modesti, James N.

Fig-1.

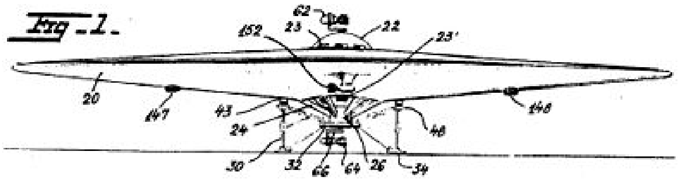
62
152
25
22
23'
20
147
43
24
48
148
30
32
66
64
26
34

USP # 3,469,802 (9-30-69) ~ Transport
Roberts, J. R., ***et al.***

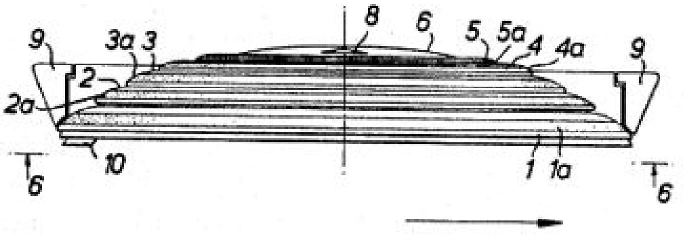
9
3a
3
8
6
5
5a
4
4a
9
2
2a
6
10
1
1a
6

USP # 3,437,290 (4-8-69) ~ Vertical Lift Aircraft
Norman, Francis A.

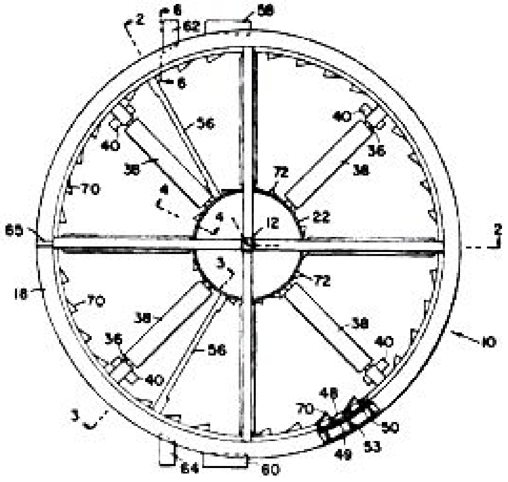
2
6
58
62
6
56
40
36
38
72
4
22
4
12
65
3
70
18
72
38
56
40
10
36
3
48
70
50
53
49
64
60

USP # 3,432,120 (4-11-69) ~ Aircraft
Guerrero, E.

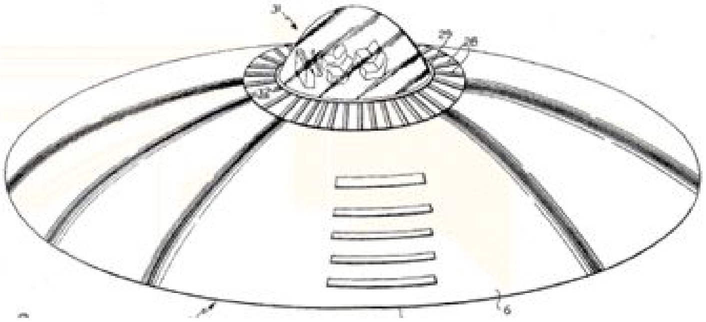

USP # 3,410,507 (11-12-68) ~ Aircraft
Moller, Paul S.

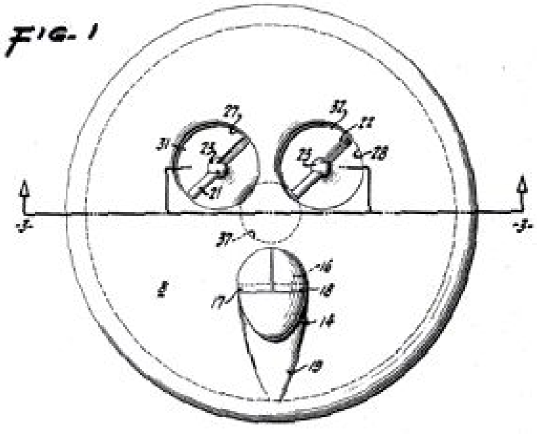
FIG. 1
27
32
22
31
23
25
21
28
3
3
37
16
18
17
14
19
2

USP # 3,397,853 (8-20-68) ~ Fluid-Sustained Vehicle
Richardson, William. B.

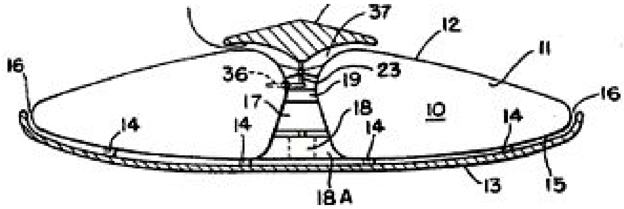
37
12
11
16
36
23
19
17
18
10
16
14
14
14
14
18A
13
15

USP # 3,395,876 (8-6-68) ~ Aircraft with Housed Counter-Rotating Propellors
Green, Jacob B.

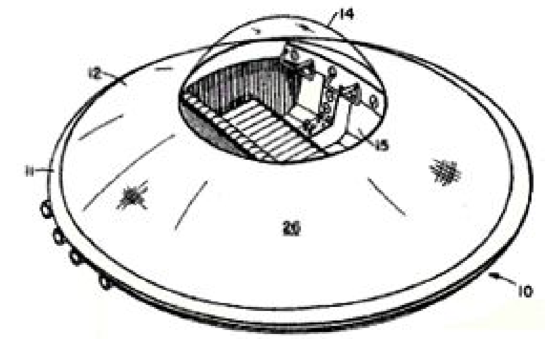
14
12
11
15
20
10

USP # 3,387,801 (6-11-68) ~ VTOL Aircraft
Kelsey, C. W.

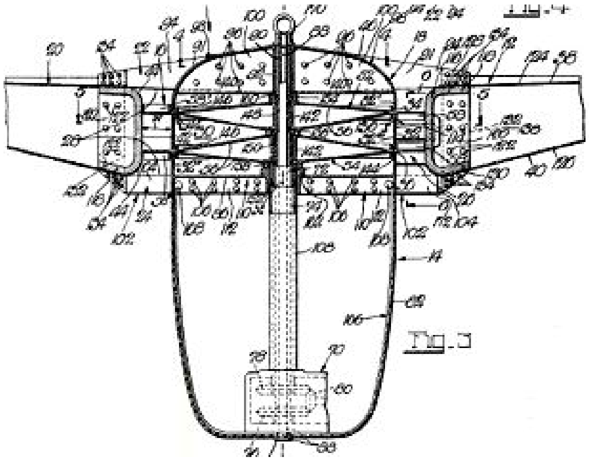
50
14
106
108
70
80
38
Fig. 3

USP # 3,321,156 (5-23-67) ~ Universally Manuverable Aircraft
McMasters, Douglas Q.

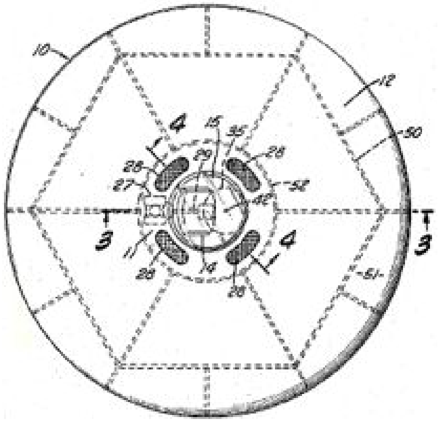
10
12
50
4
15
35
28
29
28
27
52
42
3
3
11
4
14
28
28
51

USP # 3,312,425 (4-4-67) ~ Aircraft
Lennon, C. D., *et al.*

FIG. 1.

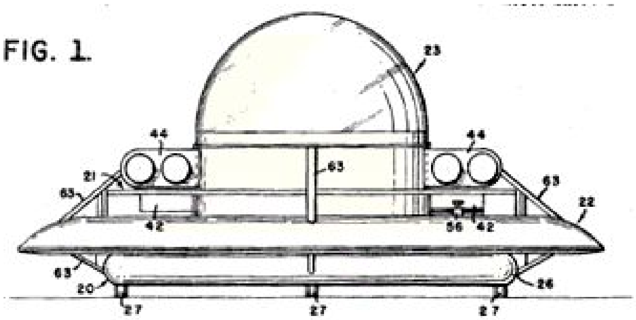
23
44
44
63
21
63
63
22
42
56
42
63
20
26
27
27
27

USP # 3,243,146 (3-29-66) ~ VTOL Aircraft
Clover, P. B.

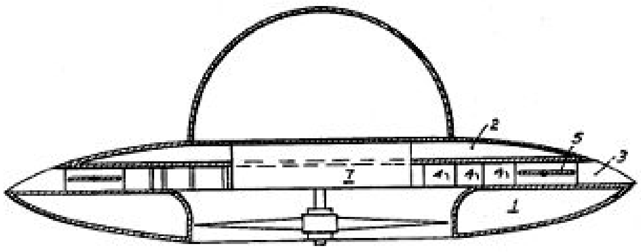

USP # 3,237,888 (3-1-66) ~ Aircraft
Willis, William M.

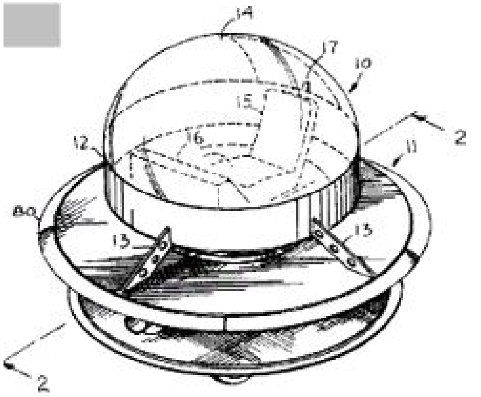
14
17
10
15
12
16
2
11
80
13
13
2

USP # 3,199,809 (8-10-65) ~ Circular Wing Flying Craft
Modesti, James N.

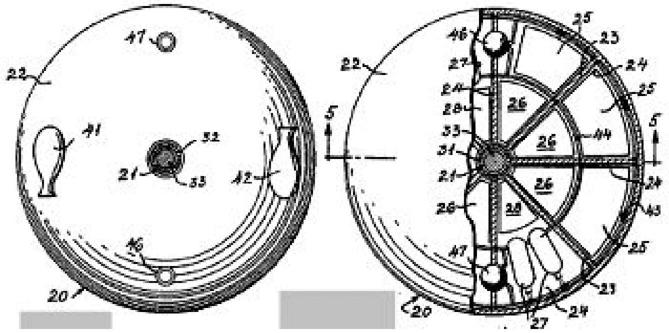
47
22
41
32
21
33
42
46
20
5
5
22
46
27
35
23
24
25
26
26
44
26
28
33
31
21
24
43
26
25
47
23
24
27
20

USP # 3,182,929 (5-11-65) ~ VTOL Aircraft
Lemberger, Robert A.

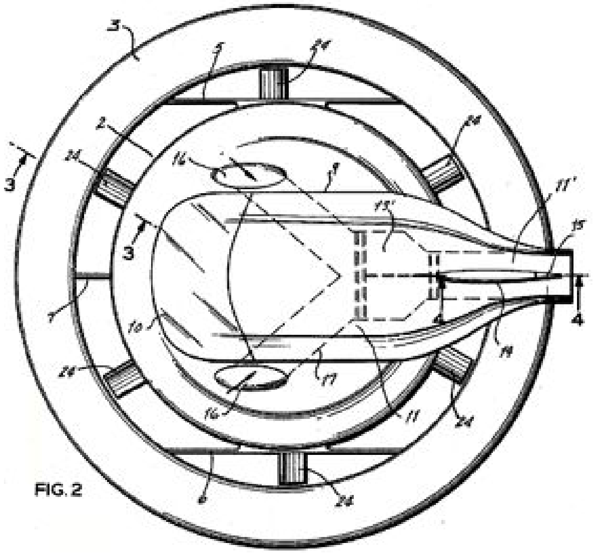
3
5
24
2
24
3
24
16
9
11'
13'
15
3
10
7
4
14
24
17
16
11
24
6
24

FIG. 2

USP # 3,124,323 (3-10-64) ~ Aircraft Propulsion & Control
Frost, John C. M.

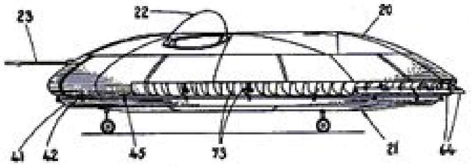
23
22
20
41
42
45
73
21
64

FIG 1.

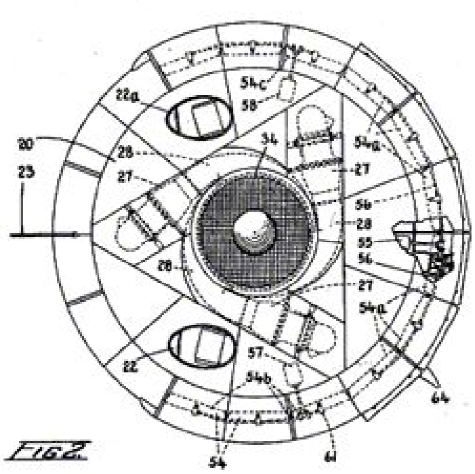
54c
58
22a
20
23
28
27
34
54a
27
56
28
55
56
27
54a
28
57
22
54b
64
54
61

FIG 2.

USP # 3,123,320 (4-3-64) ~ Vertical Rise Aircraft
Slaughter, E. E.

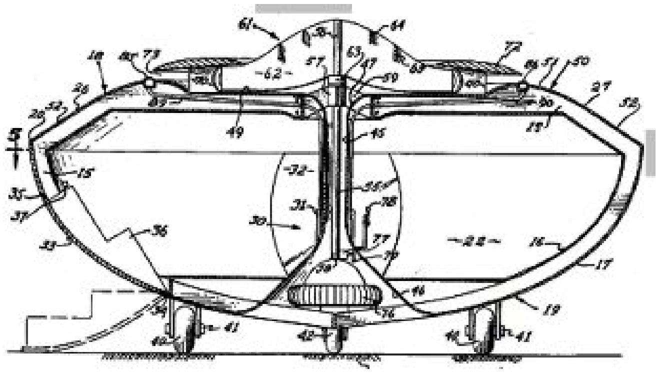
61
64
62
57
63
59
49
45
72
73
5
15
32
31
30
36
22
16
17
19
41
40
42

USP # 3,073,551 (1-15-63) ~ Vertical Lift Aircraft
Bowersox, Joseph W.

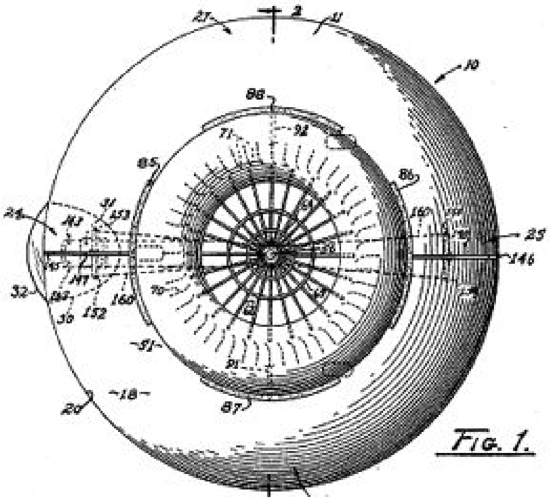
2
41
10
86
85
25
146
32
18
20
87

FIG. 1.

USP # 3,072,366 (1-8-63) ~ Fluid-Sustained Aircraft
Freeland, Leonor Z.

Fig. I

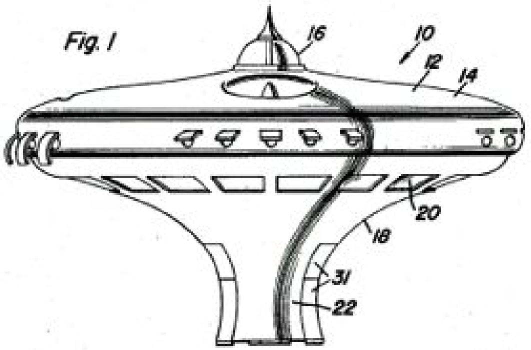
16
10
12
14
20
18
31
22

USP # 3,067,967 (12-11-62) ~ Flying Machine
Barr, I. R.

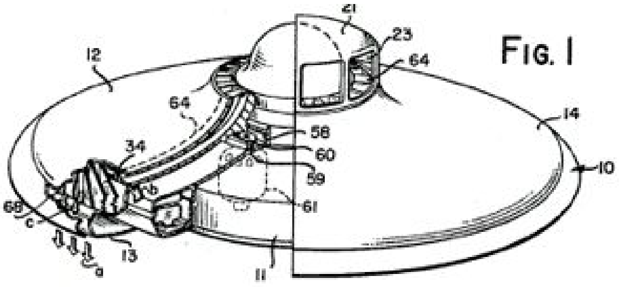
21
23
64
12
14
64
58
34
60
59
10
68
c
b
61
13
a
11

FIG. 1

USP # 3,066,890 (12-4-62) ~ Supersonic Aircraft
Price, Nathan C.

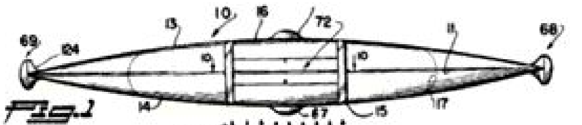
69
124
13
10
16
72
11
68
10
10
14
67
15
117

Fig.1

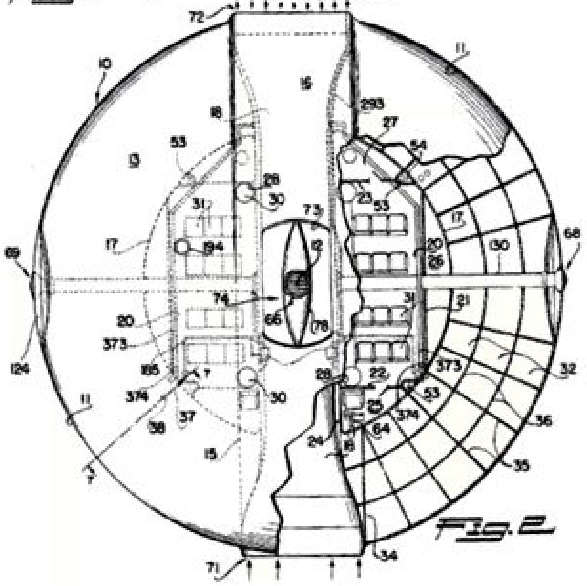
72
10
11
16
18
293
27
54
53
19
28
30
23
53
31
73
17
194
12
20
26
130
68
69
74
66
178
21
20
373
31
124
185
28
22
373
32
374
11
37
38
30
25
53
374
64
36
15
24
18
35
34
71

Fig.2

USP # 3,065,935 (11-27-62) ~ VTOL Aircraft
Dubbury, J., *et al.*

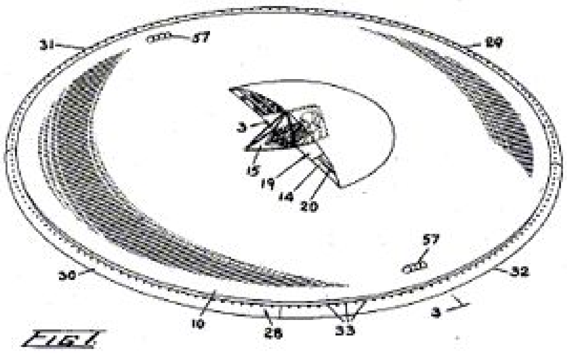
31
57
29
3
15
19
14
20
57
30
32
10
28
33
3

FIG.1

USP # 3,051,417 (8-28-62) ~ Aircraft Control Systems
Frost, John C. M., *et al.*

USP # 3,051,415 (8-28-62) ~ Fluid-Sustained Aircraft
Frost, John C. M.

USP # 3,051,414 (8-28-62) ~ Aircraft with Jet Fluid Control Ring
Frost, John C. M.

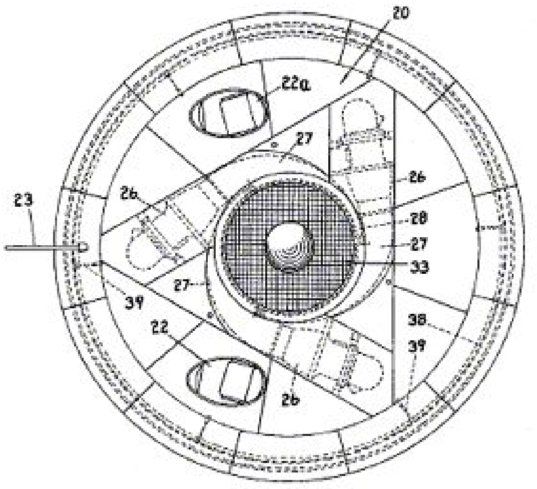
20
22a
27
26
26
28
27
33
23
27
39
22
38
39
26

USP # 3,024,966 (3-13-62) ~ Radial Flow Gas Turbine Engine Rotor Bearing
Frost, John C. M.

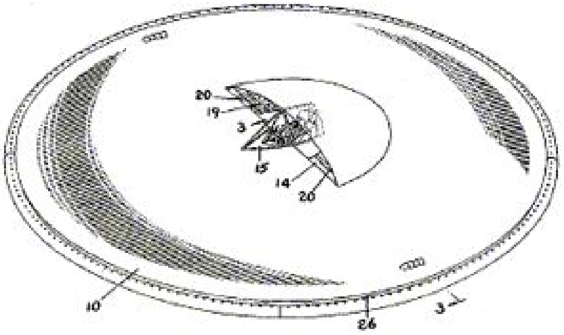
20
19
3
15
14
20
10
26
3

USP # 3,022,963 (2-27-62) ~ Disc-type Aircraft...
Frost. John C. M., ***et al.***

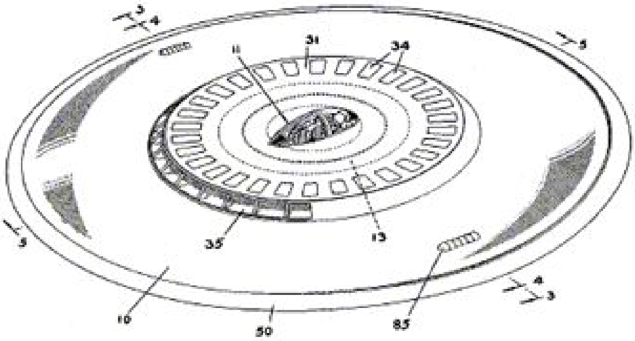
3
4
31
34
11
5
5
35
13
10
50
85
4
3

USP # 3,020,003 (2-6-62) ~ Disc Aircraft...
Frost, John C. M., ***et al.***

USP # 3,018,068 (1-23-62) ~ Disc Aircraft...
Frost, John C. M., ***et al.***

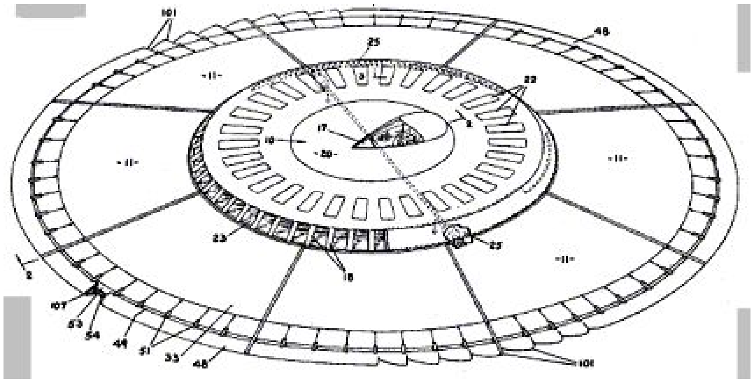
101
48
25
22
17
2
10
20
11
11
11
11
23
18
25
2
107
53
54
44
51
33
48
101

USP # 3,020,002 (2-6-62) ~ VTOL Control
Frost, John C. M.

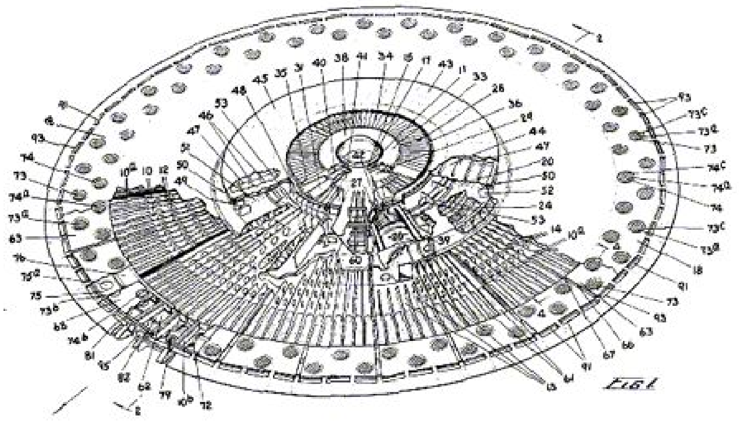
2
45
35
31
40
38
41
34
15
17
43
11
33
28
36
29
44
47
20
50
52
24
53
14
10a
48
53
46
47
51
50
49
10a
10
12
11
93
73c
73a
73
74c
74a
74
73c
73a
18
91
73
93
63
67
66
61
13
91
74
73
74a
73a
63
76
75a
75
73b
68
74b
81
95
82
62
79
10b
72
2
FIG. 1

USP # 2,997,254 (8-22-61)~ Gyro-Stabilized Vertical Rising Vehicle (Discoid)
Mulgrave, Thomas P., *et al.*

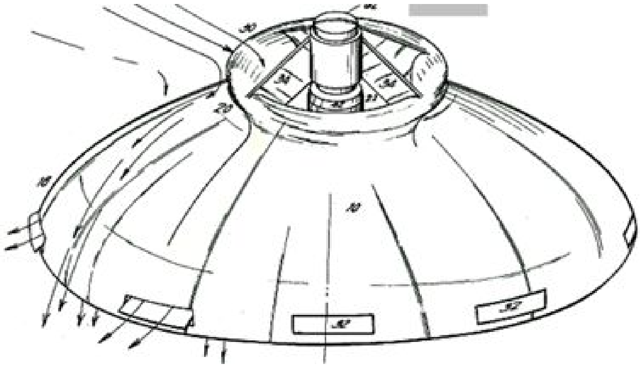
34
34
32
20
18
10
30
30

USP # 2,988,303 (6-13-61) ~ Jet-Sustained Aircraft
Coanda, Henri

FIG. 1

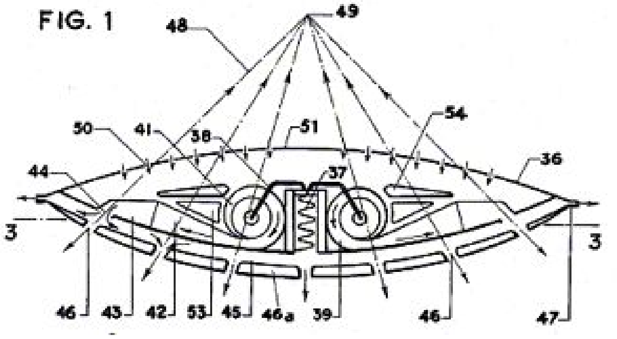
48
49
50
41
38
51
54
36
44
37
3
3
46
43
42
53
45
46a
39
46
47

USP # 2,953,320 (9-20-60) ~ Aircraft with Ducted Lifting Fan
Parry, Robert D.

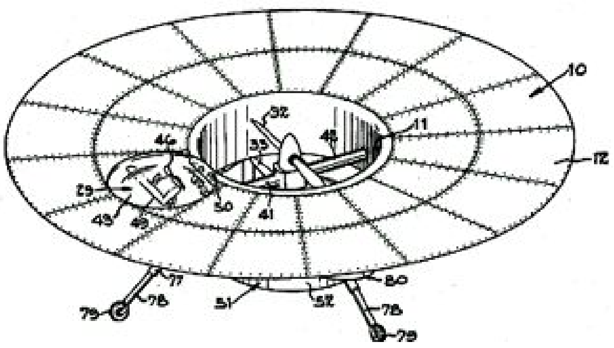
10
32
11
12
46
35
48
29
43
49
50
41
77
80
79
78
51
52
78
79

USP # 2,944,762 (7-12-60) ~ Aircraft
Lane, Thomas R.

Fig. 1

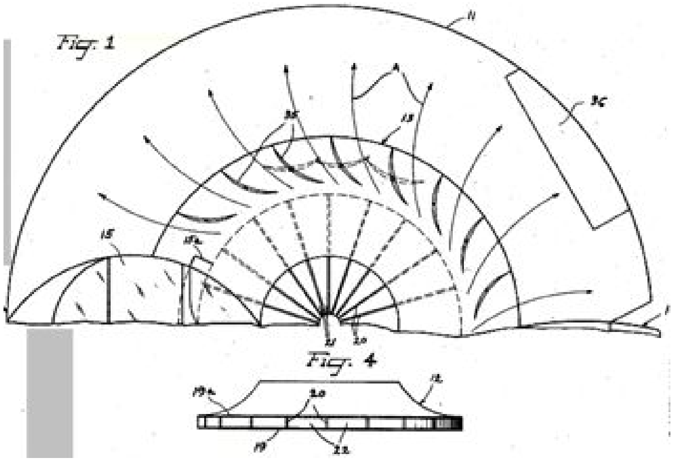
11
35
13
36
15
19a
11
20
1
Fig. 4
19a
20
12
19
22

USP # 2,939,648 (6-7-60) ~ Rotating Jet Aircraft with Lifting Disc Wing...
Fleissner, H.

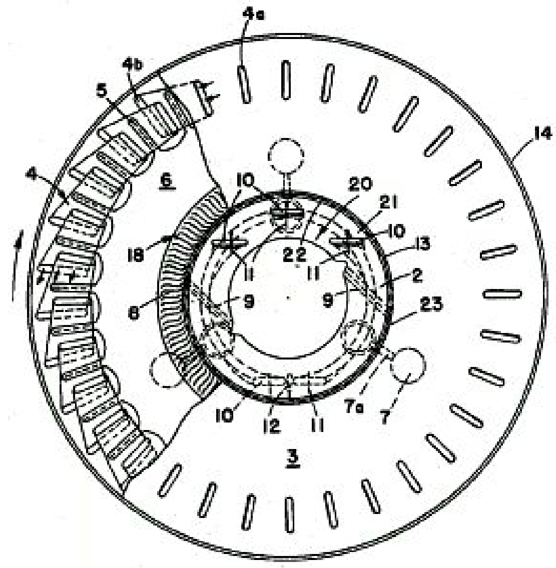
4a
4b
5
14
4
6
10
20
21
10
13
22
11
11
18
2
8
9
9
23
10
12
11
7a
7
3

USP # 2,937,492 (5-24-60) ~ Rotary Reaction Engine
Lehberger, Arthur N.

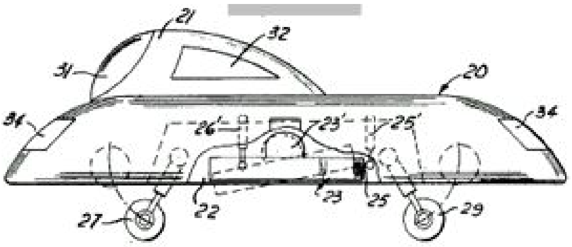
21
32
31
20
34
34
26'
23'
25'
22
23
25
27
29

USP # 2,935,275 (5-3-60) ~ Disc-Shaped Aircraft
Grayson, Leonard W.

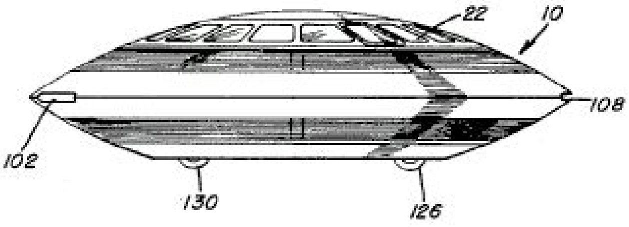
22
10
108
102
130
126

USP # 2,927,746 (3-8-60) ~ Toroidal Aircraft
Mellen, Walter R.

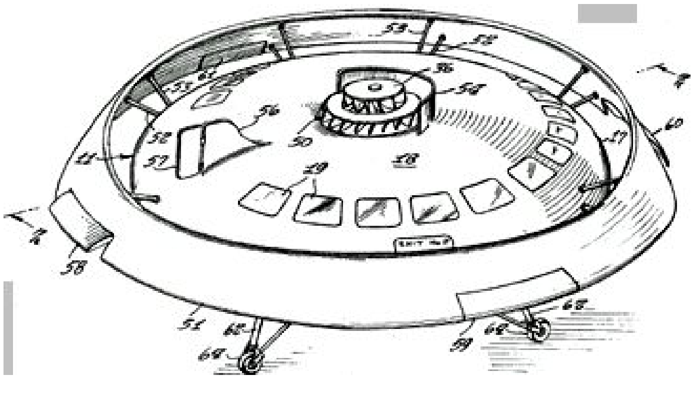
53
52
36
54
56
50
49
48
60
58
51
68
59
EXIT

USP # 2,918,230 (12-22-59) ~ Fluid-Sustained & Fluid-Propelled Aircraft
Lippisch, Alexander M.

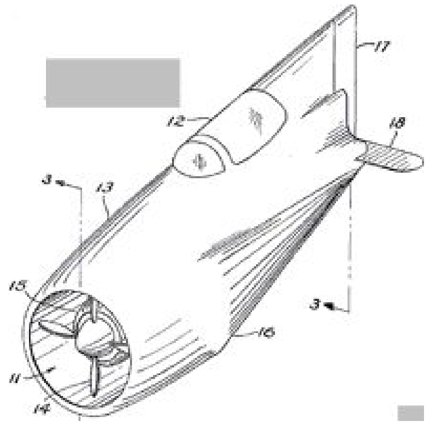
17
12
18
3
13
15
3
16
11
14

USP # 2,876,965 (3-10-59) ~ Circular Wing Aircraft...
Streib, Homer F.

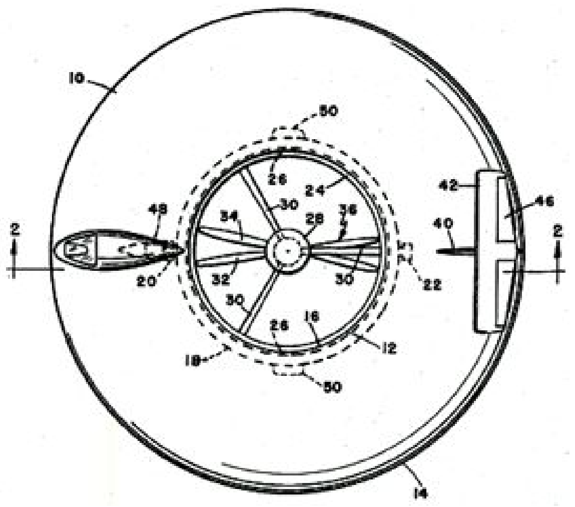
10
50
26
24
42
46
48
34
30
28
36
2
40
2
20
32
30
22
30
26
16
12
18
50
14

USP # 2,863,621 (12-9-58) ~ Vertical & Horizontal Flight Aircraft
Davis, John W.

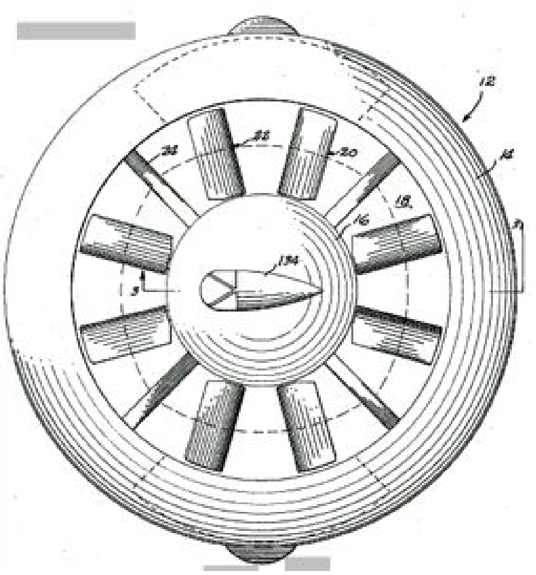
12
14
32
22
20
18
16
134
3
3

USP # 2,801,058 (7-30-57) ~ Saucer-Shaped Aircraft
Lent, Constantin P.

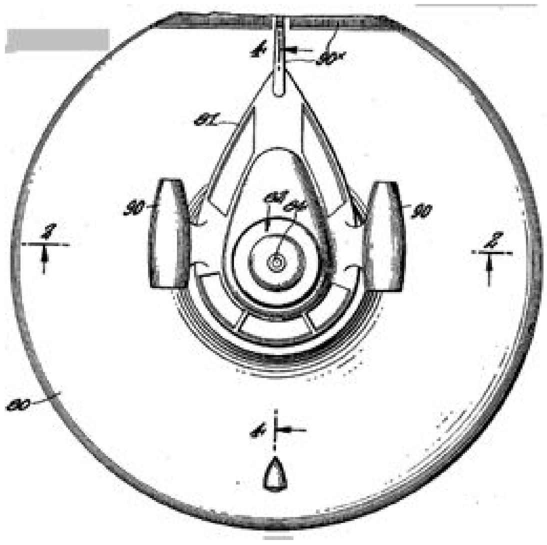
4
80'
81
90
88
84
90
2
2
80
4

USP # 2,777,649 (1-15-57) ~ Fluid-Sustained Aircraft
Williams, Samuel B.

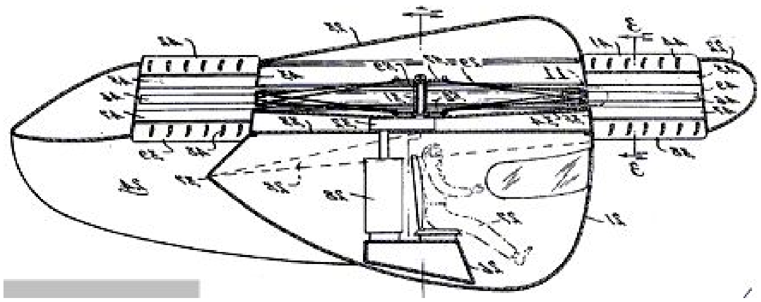

USP # 2,772,057 (11-27-56) ~ Circular Aircraft &c..
Fischer, John C.

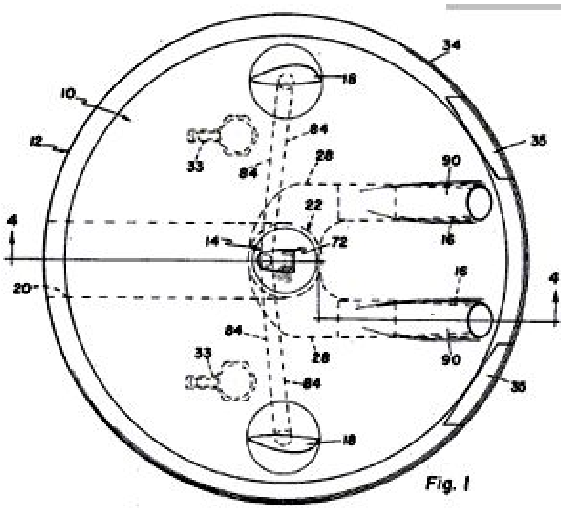
34
18
10
84
12
35
33
84
28
90
22
16
4
14
72
16
4
20
84
28
90
33
84
35
18

Fig. I

USP # 2,736,514 (2-28-56) ~ Convertible Aircraft
Ross, Robert S.

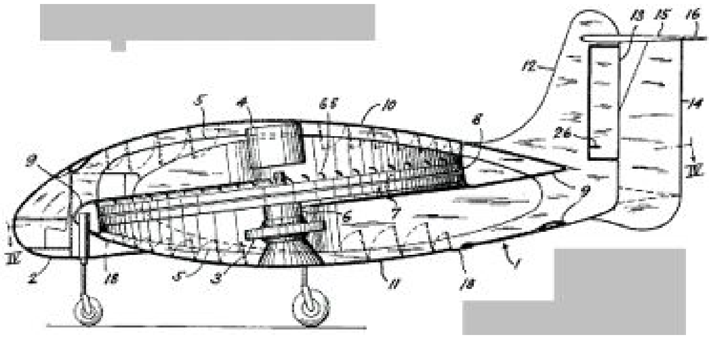
13
15
16
12
14
5
4
6a
10
8
26
9
IV
9
6
7
2
18
5
3
11
18
1
IV

USP # 2,730,311 (1-10-56) ~ Impeller Propelled Aerodynamic Body
Doak, Edmond R.

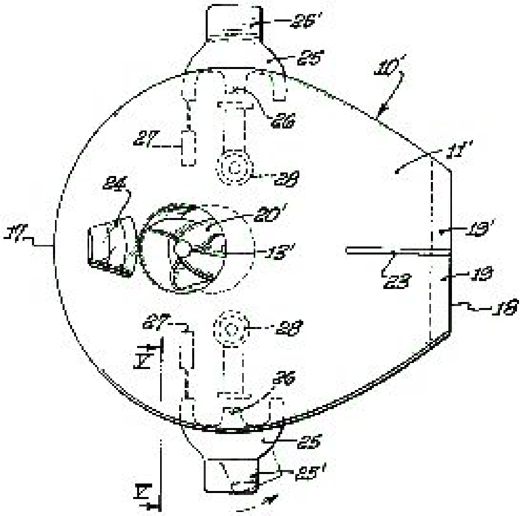
26'
25
10'
26
27
11'
28
24
20'
17
15'
19'
23
19
18
27
28
V
26
25
25'
V

USP # 2,718,364 (9-20-55) ~ Fluid-Sustained & Propelled Aircraft...
Crabtree, E.L.

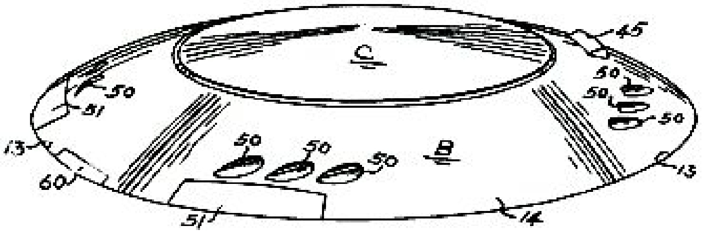
45
C
50
51
50
50
50
13
50
50
50
B
60
13
51
14

USP # 2,619,302 (11-25-52) ~ Low Aspect Ratio Aircraft
Loedding, Alfred C.

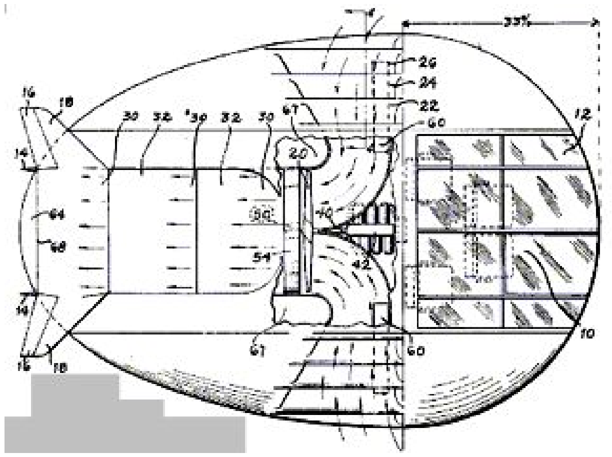
33%
26
24
22
60
12
16
18
14
30
32
34
32
30
62
20
64
68
56
40
42
54
10
61
60
14
16
18

USP # 2,567,392 (9-11-51) ~ Fluid-Sustained Aircraft
Naught, Harold

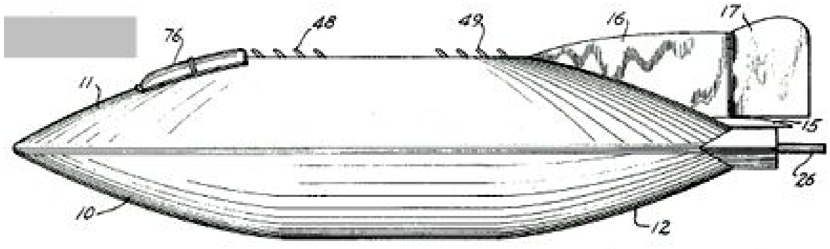
76
48
49
16
17
11
15
26
10
12

USP # 2,431,293 (11-18-47) ~ Airplane of Low Aspect Ratio
Zimmermann, Charles H.

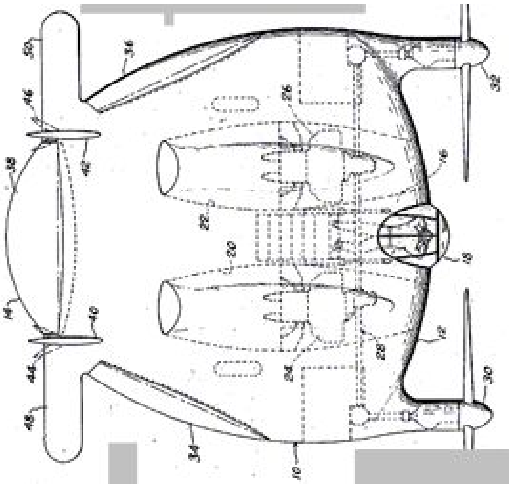

USP # 2,377,835 (6-5-45) ~ Discopter
Weygers, Alexander G.

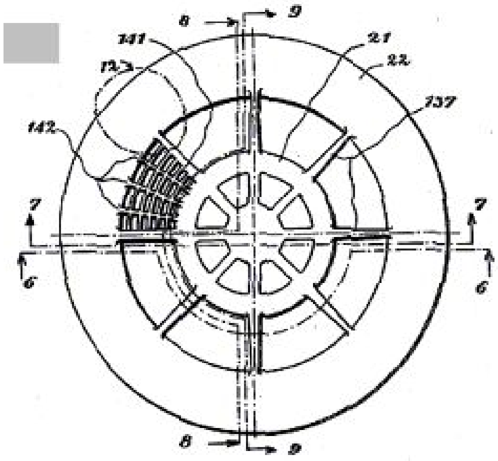
8
9
141
123
21
22
137
142
7
7
6
6
8
9

USP # 1,887,411 (11-8-32) ~ Aircraft Construction
Johnson, R. B.

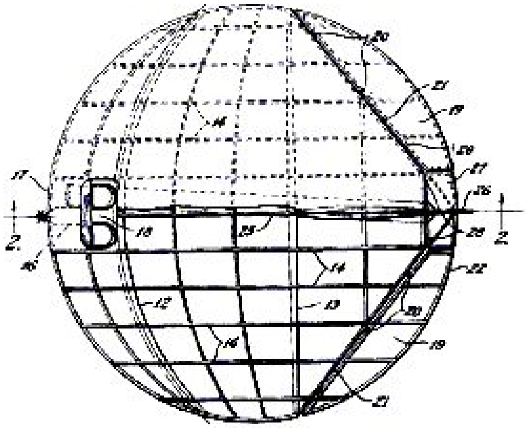
20
21
19
29
27
26
17
2
18
23
28
2
16
14
22
12
13
30
19
16
21

Your Support Maintains this Service --

BUY

The Rex Research Civilization Kit

... It's Your Best Bet & Investment in Sustainable Humanity on Earth ...
Ensure & Enhance Your Survival & Genome Transmission ...
Everything @ rexresearch.com on a Data DVD !

ORDER PAGE

WILLIAM H. WARD.

Improvement in Collecting Electricity for Telegraphing, &c.

No. 126,356. Patented April 30, 1872.

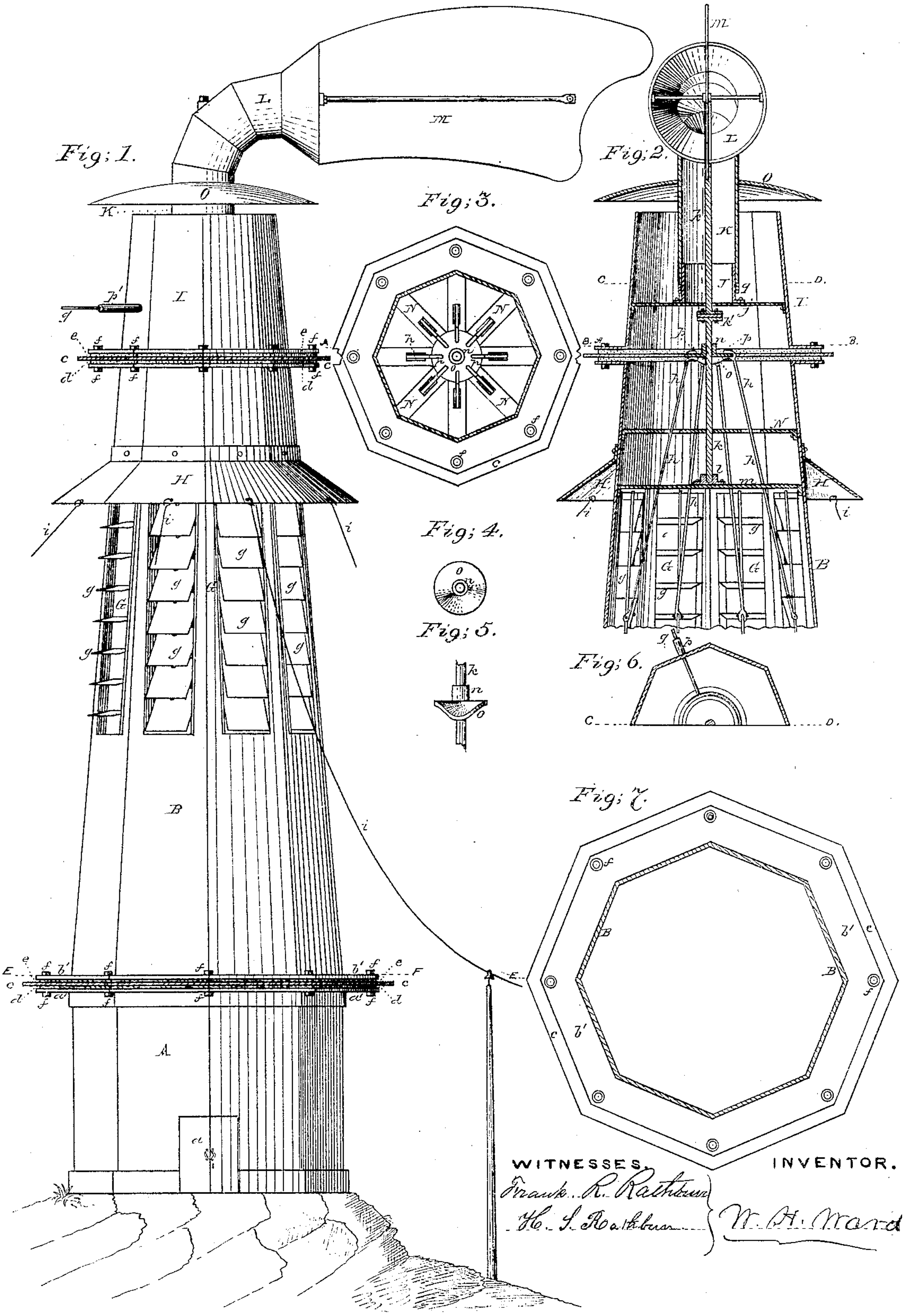

Fig; 1.
Fig; 2.
Fig; 3.
Fig; 4.
Fig; 5.
Fig; 6.
Fig; 7.
A
B
G
H
I
K
L
M
N
O
a
b'
c
d
e
f
g
h
i
k
m
n
o
p
p'
C
D
E
F

WITNESSES. INVENTOR.

Frank R. Rathbun

H. S. Rathbun

W. H. Ward

AM. PHOTO-LITHOGRAPHIC CO. N. Y. (OSBORNE'S PROCESS.)

126,356

UNITED STATES PATENT OFFICE.

WILLIAM HENRY WARD, OF AUBURN, NEW YORK.

IMPROVEMENT IN COLLECTING ELECTRICITY FOR TELEGRAPHING, &c.

Specification forming part of Letters Patent No. 126,356, dated April 30, 1872.

I, WILLIAM HENRY WARD, of Auburn, in the county of Cayuga and State of New York, have invented an Electrical Tower for Accumulating Natural Electricity for Telegraphic Purposes, of which the following is a specification:

My invention consists of a tower for the purpose of receiving and imparting natural electricity, so as to be in constant contact with that upper stratum of electricity which surrounds the earth, by tapping which a never-failing supply is formed when brought into contact with the earth, as will be more fully explained hereinafter.

In the accompanying drawing, Figure 1 represents a side elevation of my improved electrical tower. Fig. 2 is a vertical central section of the upper part of the same. Figs. 3, 4, 5, 6, and 7 are detached views of several parts of the same.

The tower is constructed in three separate sections, a lower, a middle, and an upper one, and is placed on elevated mountain tops or peaks. The lower one, A, is a mere shell, having a door, *a*, and is constructed of any material suitable for the purpose. It is insulated from the middle portion B by means of a glass diaphragm, *c*, held between similar diaphragms of rubber, *d*, and of gutta-percha, *e*, by means of flanges *a′* and *b′*, formed on the upper end of the lower portion and the lower end of the middle portion of the tower respectively. These flanges are secured to each other by insulated bolts *f*. The middle portion B is provided with suitable openings or windows, G, having shutters or slats *g* pivoted in them, so that, by means of raising or lowering rods *h*, suitably connected to said shutters, the openings G may be shut or opened. A projecting roof, H, is formed on or secured to the middle portion B just over the openings G, and which serves the double purpose of protecting the said openings from the effects of the weather, such as rain or snow, and also for receiving the aerial electricity, which may be drawn from it by wires *i*, for land-line purposes. Above this roof H the middle portion of the tower is again insulated by diaphragms of glass, rubber, and gutta-percha from the upper portion I of the tower in the same manner as it is insulated from its lower portion. A circular plate, *j*, secured on the inside of this upper portion, bears a short tube, J, which is surrounded by the tube K of the ventilator L, from which latter the vane M extends; or the tube J may be held by rods extending from the side of the tower centrally. This ventilator is supported by a rod or shaft, *k*, firmly attached to the tube K, and having its lower bearing in a step, *l*, on a brace, *m*, crossing the middle portion B of the tower just above the openings G. This rod or shaft *k* is formed in two parts, insulated from each other, as shown at *k′*. On the lower portion of this shaft *k* is keyed or otherwise secured a sleeve, *n*, from which a horizontal serpentine cam-plate, *o*, shown in detail in Figs. 4 and 5, extends, over the rim of which the forked ends *p* of rods *h* seize, and which is so arranged relatively to the vane M and the shutters or slats *g* that the revolution of the vane by the wind will open the windward and close the leeward shutters or slats *g* and openings G by means of the rods *h*. Guide-plates or their equivalents N, provided with suitable openings, through which the rods *h* pass, keep the upper forked ends *h* of the latter in constant contact with the cam-plate *o*, which raises or lowers the rods, and consequently opens or closes the openings G as it revolves. A circular roof, O, extends from the tube K of the ventilator so as to cover and protect the open upper end of the tower. The upper portion I of the tower is provided with an insulating-tube, *p′*, through which a copper wire, *q*, enters the portion I of the tower, and is coiled around the tube J just under the tube K, which wire may connect that upper portion of the tower with any land line of wires. The upper portion I of the tower, as well as the tubes J K, may be constructed of zinc, while the vane may be made of zinc, copper, or any equivalent material, which, with the zinc and the moisture of the atmosphere, would form an electrical current. As the vane revolves it opens the openings to windward and closes those to leeward, thus helping to drive an aerial current of electricity into the insulated middle portion of the tower, which current passes upwardly through the upper portion of the tower and out through the ventilator or the top, which is swung around by the wind or aerial electrical current, thus forming a draught in addition to the closing of the shutters or openings to leeward, by means of which draught the electrical current is forced upwardly and out at the vane. The top portion of the tower is of course completely insulated from all below it, including the revolving central ventilator-shaft. As the middle portion of the tower allows the electrical current of atmosphere, wind, &c., to pass within, up, and out at the top it forms a continuous current, whereby the tower is receiving continually fresh and new supplies of electricity, which can be drawn from the projecting roof H by the wires *i* for the use of land lines of telegraphs or for other purposes, such as light, heat, &c. By the use of aerial electricity I entirely dispense with artificial batteries, forming my circuit merely by connecting the aerial current with the earth current. For instance, to bring Buenos Ayres, in South America, into direct connection with New York, the following plan would be pursued: One electrical tower is erected on Pike's Peak or any other suitable high mountain in North America, and another similar tower on some suitable peak of the Andes in South America. The former would, by means of land-lines, be connected directly with Denver, which place is again connected with all the prominent cities of the States. In a similar manner the southern tower is connected by land-lines with the prominent cities via Quito. New York telegraphs to the tower on Pike's Peak, and the operator having connected the land-line with the aerial current, the signals are transmitted through the aerial current to the tower in the Andes in South America; and from there—the land-lines being suitably connected with the aerial current—to Quito and Buenos Ayres. In this manner a message would be sent entirely by natural electricity in place of artificial. In the same manner a message may be sent across the ocean by having a high tower on each continent, each of which towers would have to be, of course, through land-lines connected with the earth to enable the ground current with the aerial current to form a circuit. Different towers may be erected on the different continents, and if they are all what is technically called hooked on—that is to say, connected to the earth—a signal given at one tower will be repeated at all the towers, they being connected with each other by the aerial current. If the earth-connection is severed, or the insulation with the tower destroyed, there is no power; but by insulating the tower and concentrating its force to a point, bringing the same corresponding effect from the earth current in connection, an exceedingly powerful electrical force is created.

Having described my invention, I claim—

1. A tower constructed so as to collect, hold, distribute, and utilize aerial currents of natural electricity for telegraphic and other purposes, essentially as described.

2. A tower for collecting aerial currents of natural electricity, constructed of three sections, insulated from each other and the earth, or their equivalents, substantially in the manner described.

3. In an electrical tower, the combination of the shutters or slats *g* with the ventilator-vane in such a manner that the vane, through suitable mechanism, always opens the shutters to windward and closes them to leeward, substantially as described.

4. In an electrical tower, the collecting, distributing, and protecting roof H, substantially as and for the purpose described.

5. The combination of the tube J and ventilator L with its vane M and the coiled wire or small cable *q*, substantially as and for the purpose set forth.

6. The combination of the insulated shaft *k* and rods *h* having forked upper ends *k′* with the shutters or blinds *g*, serpentine cam-plate *o*, and ventilator L, substantially as and for the purpose described.

The above specification of my improvement, being a tower for accumulating natural electricity for telegraphic purposes, signed this 29th day of June, A. D. 1871.

W. H. WARD.

Witnesses:
A. L. PALMER,
J. H. STARIN.

(No Model.) 2 Sheets—Sheet 1.

N. TESLA.

ELECTRIC GENERATOR.

No. 511,916. Patented Jan. 2, 1894.

Fig. 1

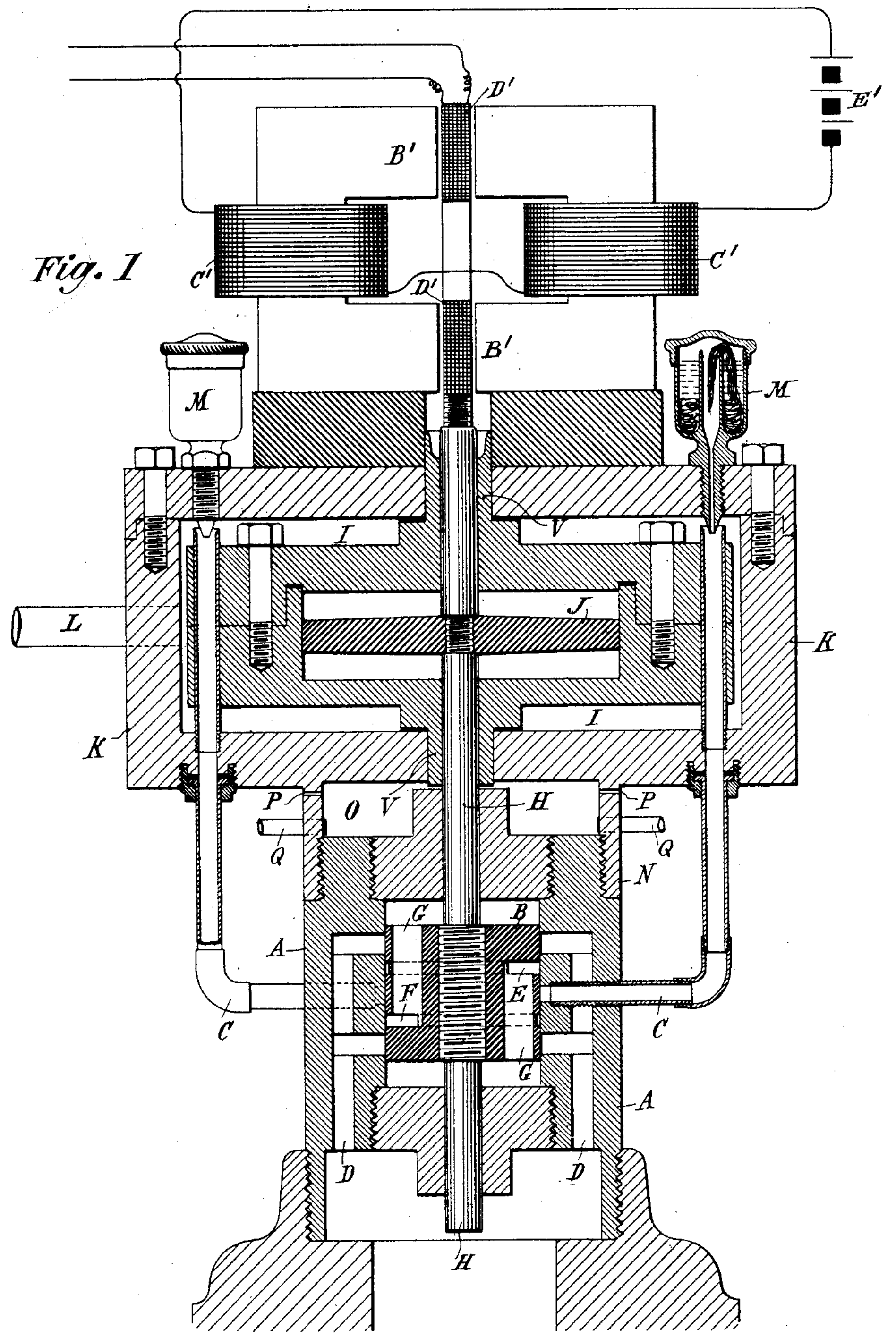
D'
B'
E'
C'
C'
D'
B'
M
M
V
I
J
L
K
K
I
P
O
V
H
P
Q
Q
N
G
B
A
F
E
C
C
G
A
D
D
H

Witnesses
Raphaël Netter
R. F. Gaylord

Inventor
Nikola Tesla
By his Attorneys
Duncan & Page.

THE NATIONAL LITHOGRAPHING COMPANY,
WASHINGTON, D. C.

(No Model.)

2 Sheets—Sheet 2.

N. TESLA.

ELECTRIC GENERATOR.

No. 511,916.

Patented Jan. 2, 1894.

Fig. 2

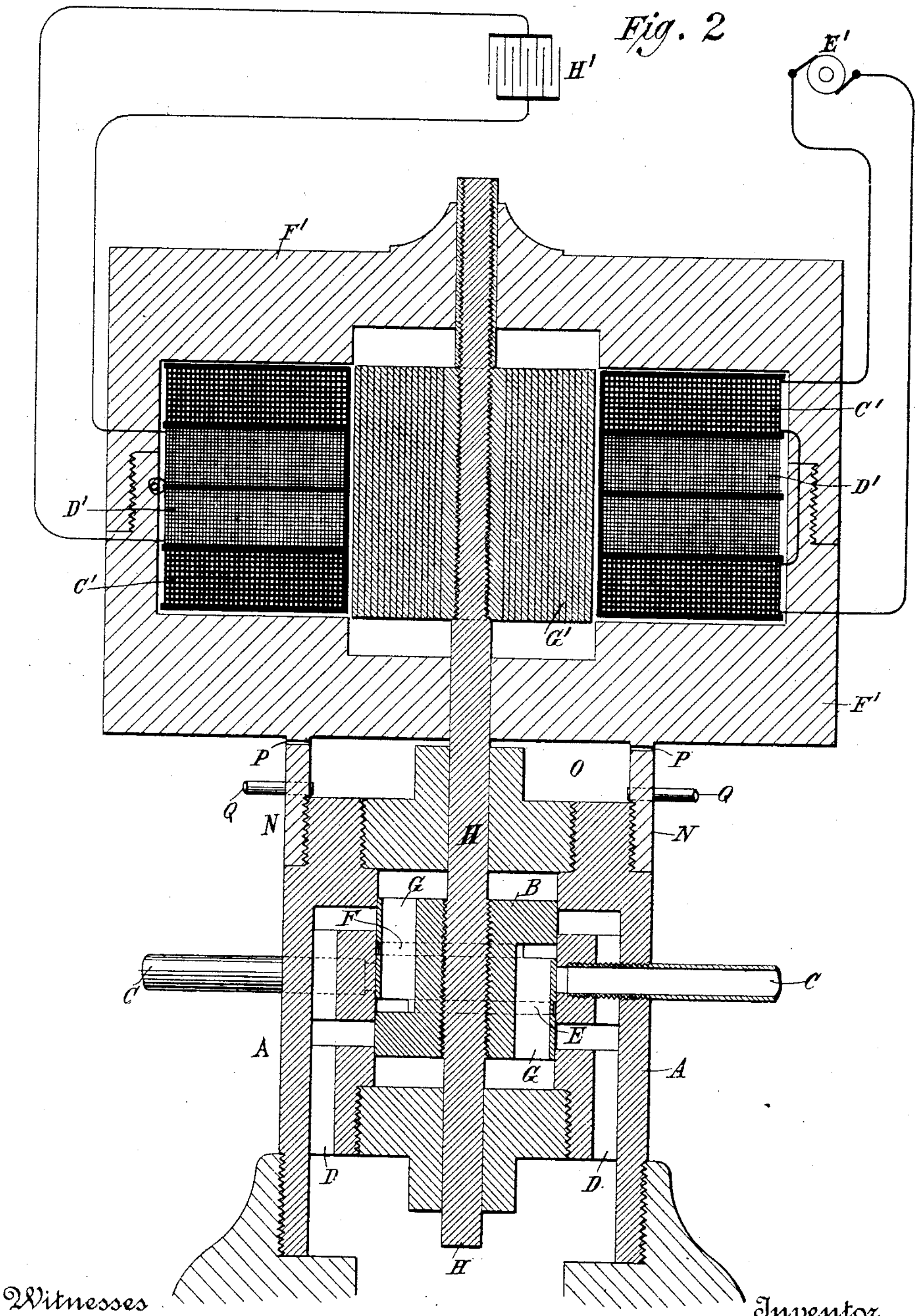
H'
E'
F'
C'
D'
D'
C'
G'
F'
P
P
O
Q
Q
N
N
H
G
B
F
C
C
E
A
A
G
D
D
H

Witnesses

Raphael Netter

R. F. Gaylord

Inventor

Nikola Tesla

By his Attorneys

Duncan & Page.

THE NATIONAL LITHOGRAPHING COMPANY,
WASHINGTON, D. C.

UNITED STATES PATENT OFFICE.

NIKOLA TESLA, OF NEW YORK, N. Y.

ELECTRIC GENERATOR.

SPECIFICATION forming part of Letters Patent No. 511,916, dated January 2, 1894.

Application filed August 19, 1893. Serial No. 483,562. (No model.)

To all whom it may concern:

Be it known that I, NIKOLA TESLA, a citizen of the United States, residing at New York, in the county and State of New York, have invented certain new and useful Improvements in Electric Generators, of which the following is a specification, reference being had to the drawings accompanying and forming a part of the same.

In an application of even date herewith, Serial No. 483,563, I have shown and described a form of engine invented by me, which, under the influence of an applied force such as the elastic tension of steam or a gas under pressure, yields an oscillation of constant period.

In order that my present invention may be more readily understood I will explain the conditions which are to be observed in order to secure this result.

It is a well known mechanical principle that if a spring possessing a sensible inertia be brought under tension, as by being stretched, and then freed, it will perform vibrations which are isochronous, and as to period, in the main, dependent upon the rigidity of the spring, and its own inertia or that of the system of which it may form an immediate part. This is known to be true in all cases where the force which tends to bring the spring or movable system into a given position is proportionate to the displacement.

In the construction of my engine above referred to I have followed and applied this principle, that is to say, I employ a cylinder and a piston which in any suitable manner I maintain in reciprocation by steam or gas under pressure. To the moving piston or to the cylinder, in case the latter reciprocate and the piston remain stationary, a spring is connected so as to be maintained in vibration thereby, and whatever may be the inertia of the piston or of the moving system and the rigidity of the spring relatively to each other, provided, the practical limits within which the law holds true that the forces which tend to bring the moving system to a given position are proportionate to the displacement, are not exceeded, the impulses of the power impelled piston and the natural vibrations of the spring will always correspond in direction and coincide in time. In the case of the engine referred to, the ports are so arranged that the movement of the piston within the cylinder in either direction ceases when the force tending to impel it and the momentum which it has acquired are counterbalanced by the increasing pressure of the steam or compressed air in that end of the cylinder toward which it is moving, and as in its movement the piston has shut off at a given point, the pressure that impelled it and established the pressure that tends to return it, it is then impelled in the opposite direction, and this action is continued as long as the requisite pressure is applied. The length of the stroke will vary with the pressure, but the rate or period of reciprocation is no more dependent upon the pressure applied to drive the piston, than would be the period of oscillation of a pendulum permanently maintained in vibration, upon the force which periodically impels it, the effect of variations in such force being merely to produce corresponding variations in the length of stroke or amplitude of vibration respectively.

In practice I have found that the best results are secured by the employment of an air spring, that is, a body of confined air or gas which is compressed and rarefied by the movements of the piston, and in order to secure a spring of constant rigidity I prefer to employ a separate chamber or cylinder containing air at the normal atmospheric pressure, although it might be at any other pressure, and in which works a plunger connected with or carried by the piston rod. The main reason why no engine heretofore has been capable of producing results of this nature is that it has been customary to connect with the reciprocating parts a heavy fly-wheel or some equivalent rotary system of relatively very great inertia, or in other cases where no rotary system was employed, as in certain reciprocating engines or tools, no regard has been paid to the obtainment of the conditions essential to the end which I have in view, nor would the pressure of such conditions in said devices appear to result in any special advantage.

Such an engine as I have described affords a means for accomplishing a result heretofore unattained, the continued production of electric currents of constant period, by imparting the movements of the piston to a core or coil in a magnetic field. It should be stated however, that in applying the engine for this purpose certain conditions are encountered which should be taken into consideration in order to satisfactorily secure the desired result. When a conductor is moved in a magnetic field and a current caused to circulate therein, the electro-magnetic reaction between it and the field, might disturb the mechanical oscillation to such an extent as to throw it out of isochronism. This, for instance, might occur when the electro-magnetic reaction is very great in comparison to the power of the engine, and there is a retardation of the current so that the electro-magnetic reaction might have an effect similar to that which would result from a variation of the tension of the spring, but if the circuit of the generator be so adjusted that the phases of the electromotive force and current coincide in time, that is to say, when the current is not retarded, then the generator driven by the engine acts merely as a frictional resistance and will not, as a rule, alter the period of the mechanical vibration, although it may vary its amplitude. This condition may be readily secured by properly proportioning the self induction and capacity of the circuit including the generator. I have, however, observed the further fact in connection with the use of such engines as a means for running a generator, that it is advantageous that the period of the engine and the natural period of electrical vibration of the generator should be the same, as in such case the best conditions for electrical resonance are established and the possibility of disturbing the period of mechanical vibrations is reduced to a minimum. I have found that even if the theoretical conditions necessary for maintaining a constant period in the engine itself are not exactly maintained, still the engine and generator combined will vibrate at a constant period. For example, if instead of using in the engine an independent cylinder and plunger, as an air spring of practically constant rigidity, I cause the piston to impinge upon air cushions at the ends of its own cylinder, although the rigidity of such cushions or springs might be considerably affected and varied by the variations of pressure within the cylinder, still by combining with such an engine a generator which has a period of its own approximately that of the engine, constant vibration may be maintained even through a considerable range of varying pressure, owing to the controlling action of the electro-magnetic system. I have even found that under certain conditions the influence of the electro-magnetic system may be made so great as to entirely control the period of the mechanical vibration within wide limits of varying pressure. This is likely to occur in those instances where the power of the engine while fully capable of maintaining a vibration once started, is not sufficient to change its rate. So, for the sake of illustration, if a pendulum is started in vibration, and a small force applied periodically in the proper direction to maintain it in motion, this force would have no substantial control over the period of the oscillation, unless the inertia of the pendulum be small in comparison to the impelling force, and this would be true no matter through what fraction of the period the force may be applied. In the case under consideration the engine is merely an agent for maintaining the vibration once started, although it will be understood that this does not preclude the performance of useful work which would simply result in a shortening of the stroke. My invention, therefore, involves the combination of a piston free to reciprocate under the influence of steam or a gas under pressure and the movable element of an electric generator which is in direct mechanical connection with the piston, and it is more especially the object of my invention to secure from such combination electric currents of a constant period. In the attainment of this object I have found it preferable to construct the engine so that it of itself controls the period, but as I have stated before, I may so modify the elements of the combination that the electro-magnetic system may exert a partial or even complete control of the period.

In illustration of the manner in which the invention is carried out I now refer to the accompanying drawings.

Figure 1 is a central sectional view of an engine and generator embodying the invention. Fig. 2 is a modification of the same.

Referring to Fig. 1 A is the main cylinder in which works a piston B. Inlet ports C C pass through the sides of the cylinder opening at the middle portion thereof and on opposite sides. Exhaust ports D D extend through the walls of the cylinder and are formed with branches that open into the interior of the cylinder on each side of the inlet ports and on opposite sides of the cylinder. The piston B is formed with two circumferential grooves E F which communicate through openings G in the piston with the cylinder on opposite sides of said piston respectively.

The particular construction of the cylinder, the piston and the ports controlling it may be very much varied, and is not in itself material, except that in the special case now under consideration it is desirable that all the ports, and more especially the exhaust ports should be made very much larger than is usually the case so that no force due to the action of the steam or compressed air will tend to retard or affect the return of the piston in either direction. The piston B is secured to a piston rod H which works in suitable stuffing boxes in the heads of the cylinder A. This rod is prolonged on one side and extends through bearings V in a cylinder I suitably mounted or supported in line with the first, and within which is a disk or plunger J carried by the rod H. The cylinder I is without ports of any kind and is air-tight except as a small leakage may occur through the bearings V, which experience has shown need not be fitted with any very considerable accuracy. The cylinder I is surrounded by a jacket K which leaves an open space or chamber around it. The bearings V in the cylinder I, extend through the jacket K to the outside air and the chamber between the cylinder and jacket is made steam or air-tight as by a suitable packing. The main supply pipe L for steam or compressed air leads into this chamber, and the two pipes that lead to the cylinder A run from the said chamber, oil cups M being conveniently arranged to deliver oil into the said pipes for lubricating the piston. In the particular form of engine shown, the jacket K which contains the cylinder I is provided with a flange N by which it is screwed to the end of the cylinder A. A small chamber O is thus formed which has air vents P in its sides and drip pipes Q leading out from it through which the oil which collects in it is carried off.

To explain now the operation of the engine described, in the position of the parts shown, or when the piston is at the middle point of its stroke, the plunger J is at the center of the cylinder I and the air on both sides of the same is at the normal pressure of the outside atmosphere. If a source of steam or compressed air be then connected to the inlet ports C C of the cylinder A and a movement be imparted to the piston as by a sudden blow, the latter is caused to reciprocate in a manner well understood. The movements of the piston compress and rarefy the air in the cylinder I at opposite ends of the same alternately. A forward stroke compresses the air ahead of the plunger J which acts as a spring to return it. Similarly on the back stroke the air is compressed on the opposite side of the plunger J and tends to drive it forward. The compressions of the air in the cylinder I and the consequent loss of energy due mainly to the imperfect elasticity of the air, give rise to a very considerable amount of heat. This heat I utilize by conducting the steam or compressed air to the engine cylinder through the chamber formed by the jacket surrounding the air-spring cylinder. The heat thus taken up and used to raise the temperature of the steam or air acting upon the piston is availed of to increase the efficiency of the engine. In any given engine of this kind the normal pressure will produce a stroke of determined length, and this will be increased or diminished according to the increase of pressure above or the reduction of pressure below the normal.

In constructing the apparatus proper allowance is made for a variation in the length of stroke by giving to the confining cylinder I of the air spring properly determined dimensions. The greater the pressure upon the piston, the higher the degree of compression of the air-spring, and the consequent counteracting force upon the plunger. The rate or period of reciprocation of the piston, however, is mainly determined as described above by the rigidity of the air spring and the inertia of the moving system, and any period of oscillation within very wide limits may be secured by properly portioning these factors, as by varying the dimensions of the air chamber which is equivalent to varying the rigidity of the spring, or by adjusting the weight of the moving parts. These conditions are all readily determinable, and an engine constructed as herein described may be made to follow the principle of operation above stated and maintain a perfectly uniform period through very wide limits of pressure.

The pressure of the air confined in the cylinder when the plunger I is in its central position will always be practically that of the surrounding atmosphere, for while the cylinder is so constructed as not to permit such sudden escape of air as to sensibly impair or modify the action of the air spring there will still be a slow leakage of air into or out of it around the piston rod according to the pressure therein, so that the pressure of the air on opposite sides of the plunger will always tend to remain at that of the outside atmosphere.

To the piston rod H is secured a conductor or coil of wire D′ which by the movements of the piston is oscillated in the magnetic field produced by two magnets B′ B′ which may be permanent magnets or energized by coils C′ C′ connected with a source of continuous currents E′. The movement of the coil D′ across the lines of force established by the magnets gives rise to alternating currents in the coil. These currents, if the period of mechanical oscillation be constant will be of constant period, and may be utilized for any purpose desired.

In the case under consideration it is assumed as a necessary condition that the inertia of the movable element of the generator and the electro-magnetic reaction which it exerts will not be of such character as to materially disturb the action of the engine.

Fig. 2 is an example of a combination in which the engine is not of itself capable of determining entirely the period of oscillation, but in which the generator contributes to this end. In this figure the engine is the same as in Fig. 1. The exterior air spring is however omitted and the air spaces at the ends of the cylinder A relied on for accomplishing the same purpose. As the pressure in these spaces is liable to variations from variations in the steam or gas used in impelling the piston they might affect the period of oscillation, and the conditions are not as stable and certain as in the case of an engine constructed as in Fig. 1. But if the natural period of vibration of the elastic system be made to approximately accord with the average period of the engine such tendencies to variation are very largely overcome and the engine will preserve its period even through a considerable range of variations of pressure. The generator in this case is composed of a magnetic casing F′ in which a laminated core G′ secured to the piston rod H is caused to vibrate. Surrounding the plunger are two exciting coils C′ C′, and one or more induced coils D′ D′. The coils C′ C′ are connected with a generator of continuous currents E′ and are wound to produce consequent poles in the core G′. Any movement of the latter will therefore shift the lines of force through coils D′ D′ and produce currents therein.

In the circuit of coils D′ is shown a condenser H′. It need only be said that by the use of a proper condenser the self induction of this circuit may be neutralized. Such a circuit will have a certain natural period of vibration, that is to say that when the electricity therein is disturbed in any way an electrical or electro-magnetic vibration of a certain period takes place, and as this depends upon the capacity and self induction, such period may be varied to approximately accord with the period of the engine.

In case the power of the engine be comparatively small, as when the pressure is applied through a very small fraction of the total stroke, the electrical vibration will tend to control the period, and it is clear that if the character of such vibration be not very widely different from the average period of vibration of the engine under ordinary working conditions such control may be entirely adequate to produce the desired results.

Having now described my invention, what I claim is—

1. The combination with the piston or equivalent element of an engine which is free to reciprocate under the action thereon of steam or a gas under pressure, of the moving conductor or element of an electric generator in direct mechanical connection therewith.

2. The combination with the piston or equivalent element of an engine which is free to reciprocate under the action of steam or a gas under pressure, of the moving conductor or element of an electric generator in direct mechanical connection therewith, the engine and generator being adapted by their relative adjustment with respect to period to produce currents of constant period, as set forth.

3. The combination with an engine comprising a piston which is free to reciprocate under the action of steam or a gas under pressure, and an electric generator having inducing and induced elements one of which is capable of oscillation in the field of force, the said movable element being carried by the piston rod of the engine, as set forth.

4. The combination with an engine operated by steam or a gas under pressure and having a constant period of reciprocation, of an electric generator, the moving element of which is carried by the reciprocating part of the engine, the generator and its circuit being so related to the engine with respect to the period of electrical vibration as not to disturb the period of the engine, as set forth.

5. The combination with a cylinder and a piston reciprocated by steam or a gas under pressure of a spring maintained in vibration by the movement of the piston, and an electric generator, the movable conductor or element of which is connected with the piston, these elements being constructed and adapted in the manner set forth for producing a current of constant period.

6. The method of producing electric currents of constant period herein described which consists in imparting the oscillations of an engine to the moving element of an electric generator and regulating the period of mechanical oscillation by an adjustment of the reaction of the electric generator, as herein set forth.

NIKOLA TESLA.

Witnesses:
PARKER W. PAGE,
R. F. GAYLORD.

No. 649,621. **Patented May 15, 1900.**

N. TESLA.

APPARATUS FOR TRANSMISSION OF ELECTRICAL ENERGY.

(Application filed Feb. 19, 1900.)

(No Model.)

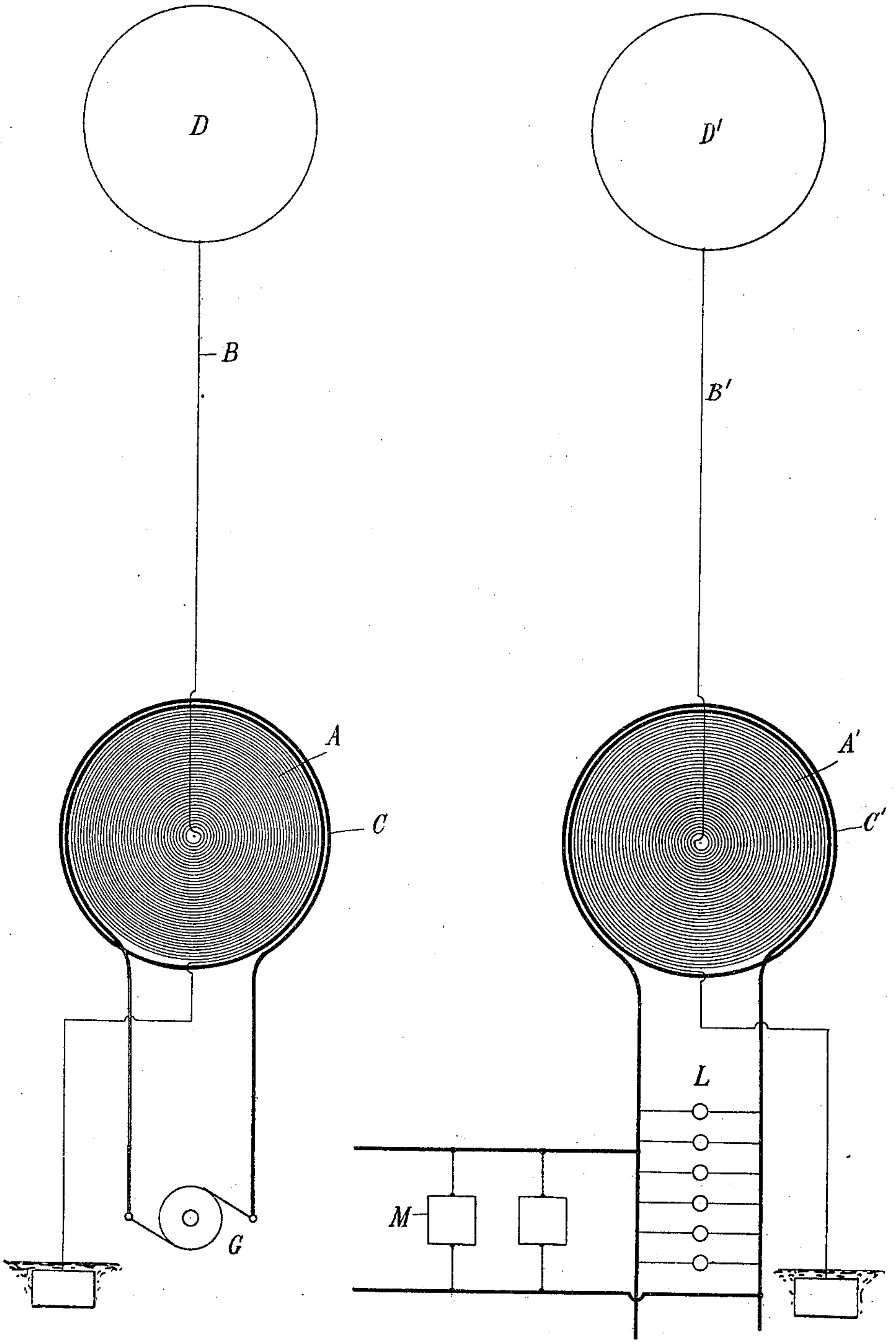

Witnesses:

Benjamin Miller.

G. W. Martling.

Nikola Tesla, *Inventor*

by Kerr, Page & Cooper *Att'ys*

THE NORRIS PETERS CO., PHOTO-LITHO., WASHINGTON, D. C.

UNITED STATES PATENT OFFICE.

NIKOLA TESLA, OF NEW YORK, N. Y.

APPARATUS FOR TRANSMISSION OF ELECTRICAL ENERGY.

SPECIFICATION forming part of Letters Patent No. 649,621, dated May 15, 1900.

Original application filed September 2, 1897, Serial No. 650,343. Divided and this application filed February 19, 1900. Serial No. 5,780. (No model.)

To all whom it may concern:

Be it known that I, NIKOLA TESLA, a citizen of the United States, residing at the borough of Manhattan, in the city of New York, county and State of New York, have invented certain new and useful Improvements in Apparatus for the Transmission of Electrical Energy, of which the following is a specification, reference being had to the drawing accompanying and forming a part of the same.

This application is a division of an application filed by me on September 2, 1897, Serial No. 650,343, entitled "Systems of transmissions of electrical energy," and is based upon new and useful features and combinations of apparatus shown and described in said application for carrying out the method therein disclosed and claimed.

The invention which forms the subject of my present application comprises a transmitting coil or conductor in which electrical currents or oscillations are produced and which is arranged to cause such currents or oscillations to be propagated by conduction through the natural medium from one point to another remote therefrom and a receiving coil or conductor at such distant point adapted to be excited by the oscillations or currents propagated from the transmitter.

This apparatus is shown in the accompanying drawing, which is a diagrammatic illustration of the same.

A is a coil, generally of many turns and of a very large diameter, wound in spiral form either about a magnetic core or not, as may be desired. C is a second coil formed by a conductor of much larger size and smaller length wound around and in proximity to the coil A.

The apparatus at one point is used as a transmitter, the coil A in this case constituting a high-tension, secondary, and the coil C the primary, of much lower tension, of a transformer. In the circuit of the primary C is included a suitable source of current G. One terminal of the secondary A is at the center of the spiral coil, and from this terminal the current is led by a conductor B to a terminal D, preferably of large surface, formed or maintained by such means as a balloon at an elevation suitable for the purposes of transmission. The other terminal of the secondary A is connected to earth, and, if desired, to the primary also, in order that the latter may be at substantially the same potential as the adjacent portions of the secondary, thus insuring safety. At the receiving-station a transformer of similar construction is employed; but in this case the longer coil A′ constitutes the primary, and the shorter coil C′ the secondary, of the transformer. In the circuit of the latter are connected lamps L, motors M, or other devices for utilizing the current. The elevated terminal D′ connects with the center of the coil A′, and the other terminal of said coil is connected to earth and preferably, also, to the coil C′ for the reasons above stated.

The length of the thin wire coil in each transformer should be approximately one-quarter of the wave length of the electric disturbance in the circuit, this estimate being based on the velocity of propagation of the disturbance through the coil itself and the circuit with which it is designed to be used. By way of illustration, if the rate at which the current traverses the circuit including the coil be one hundred and eighty-five thousand miles per second then a frequency of nine hundred and twenty-five per second would maintain nine hundred and twenty-five stationary moves in a circuit one hundred and eighty-five thousand miles long and each wave would be two hundred miles in length.

For such a low frequency, which would be resorted to only when it is indispensable for the operation of motors of the ordinary kind under the conditions above assumed, I would use a secondary of fifty miles in length. By such an adjustment or proportioning of the length of wire in the secondary coil or coils the points of highest potential are made to coincide with the elevated terminals D D′, and it should be understood that whatever length be given to the wires this requirement should be complied with in order to obtain the best results.

It will be readily understood that when the above-prescribed relations exist the best conditions for resonance between the transmitting and receiving circuits are attained, and owing to the fact that the points of highest potential in the coils or conductors A A′ are coincident with the elevated terminals the maximum flow of current will take place in the two coils, and this, further, necessarily implies that the capacity and inductance in each of the circuits have such values as to secure the most perfect condition of synchronism with the impressed oscillations.

When the source of current G is in operation and produces rapidly pulsating or oscillating currents in the circuit of coil C, corresponding induced currents of very much higher potential are generated in the secondary coil A, and since the potential in the same gradually increases with the number of turns toward the center and the difference of potential between the adjacent turns is comparatively small a very high potential impracticable with ordinary coils may be successively obtained.

As the main object for which the apparatus is designed is to produce a current of excessively-high potential, this object is facilitated by using a primary current of very considerable frequency; but the frequency of the currents is in a large measure arbitrary, for if the potential be sufficiently high and the terminals of the coils be maintained at the proper elevation where the atmosphere is rarefied the stratum of air will serve as a conducting medium for the current produced and the latter will be transmitted through the air, with, it may be, even less resistance than through an ordinary conductor.

As to the elevation of the terminals D D′, it is obvious that this is a matter which will be determined by a number of things, as by the amount and quality of the work to be performed, by the condition of the atmosphere, and also by the character of the surrounding country. Thus if there be high mountains in the vicinity the terminals should be at a greater height, and generally they should always be at an altitude much greater than that of the highest objects near them. Since by the means described practically any potential that is desired may be produced, the currents through the air strata may be very small, thus reducing the loss in the air.

The apparatus at the receiving-station responds to the currents propagated from the transmitter in a manner which will be well understood from the foregoing description. The primary circuit of the receiver—that is, the thin wire coil A′—is excited by the currents propagated by conduction through the intervening natural medium from the transmitter, and these currents induce in the secondary coil C′ other currents which are utilized for operating the devices included in the circuit thereof.

Obviously the receiving-coils, transformers, or other apparatus may be movable—as, for instance, when they are carried by a vessel floating in the air or by a ship at sea. In the former case the connection of one terminal of the receiving apparatus to the ground might not be permanent, but might be intermittently or inductively established without departing from the spirit of my invention.

It is to be noted that the phenomenon here involved in the transmission of electrical energy is one of true conduction and is not to be confounded with the phenomena of electrical radiation which have heretofore been observed and which from the very nature and mode of propagation would render practically impossible the transmission of any appreciable amount of energy to such distances as are of practical importance.

What I now claim as my invention is—

1. The combination with a transmitting coil or conductor connected to ground and to an elevated terminal respectively, and means for producing therein electrical currents or oscillations, of a receiving coil or conductor similarly connected to ground and to an elevated terminal, at a distance from the transmitting-coil and adapted to be excited by currents caused to be propagated from the same by conduction through the intervening natural medium, a secondary conductor in inductive relation to the receiving-conductor and devices for utilizing the current in the circuit of said secondary conductor, as set forth.

2. The combination with a transmitting coil or conductor having its ends connected to ground and to an elevated terminal respectively, a primary coil in inductive relation thereto and a source of electrical oscillations in said primary circuit, of a receiving conductor or coil having its ends connected to ground and to an elevated terminal respectively and adapted to be excited by currents caused to be propagated from the transmitter through the natural medium and a secondary circuit in inductive relation to the receiving-circuit and receiving devices connected therewith, as set forth.

3. The combination with a transmitting instrument comprising a transformer having its secondary connected to ground and to an elevated terminal respectively, and means for impressing electrical oscillations upon its primary, of a receiving instrument comprising a transformer having its primary similarly connected to ground and to an elevated terminal, and a translating device connected with its secondary, the capacity and inductance of the two transformers having such values as to secure synchronism with the impressed oscillations, as set forth.

4. The combination with a transmitting instrument comprising an electrical transformer having its secondary connected to ground and to an elevated terminal respectively, and means for impressing electrical oscillations upon its primary, of a receiving instrument comprising a transformer having its primary similarly connected to ground and to an elevated terminal, and a translat-

ing device connected with its secondary, the capacity and inductance of the secondary of the transmitting and primary of the receiving instruments having such values as to secure synchronism with the impressed oscillations, as set forth.

5. The combination with a transmitting coil or conductor connected to ground and an elevated terminal respectively, and means for producing electrical currents or oscillations in the same, of a receiving coil or conductor similarly connected to ground and to an elevated terminal and synchronized with the transmitting coil or conductor, as set forth.

6. The combination with a transmitting instrument comprising an electrical transformer, having its secondary connected to ground and to an elevated terminal respectively, of a receiving instrument comprising a transformer, having its primary similarly connected to ground and to an elevated terminal, the receiving-coil being synchronized with that of the transmitter, as set forth.

7. The combination with a transmitting coil or conductor connected to ground and to an elevated terminal respectively, and means for producing electrical currents or oscillations in the same, of a receiving coil or conductor similarly connected to ground and to an elevated terminal, the said coil or coils having a length equal to one-quarter of the wave length of the disturbance propagated, as set forth.

8. The combination with a transmitting coil or conductor connected to ground and to an elevated terminal respectively, and adapted to cause the propagation of currents or oscillations by conduction through the natural medium, of a receiving-circuit similarly connected to ground and to an elevated terminal, and of a capacity and inductance such that its period of vibration is the same as that of the transmitter, as set forth.

9. The transmitting or receiving circuit herein described, connected to ground and an elevated terminal respectively, and arranged in such manner that the elevated terminal is charged to the maximum potential developed in the circuit, as set forth.

10. The combination with a transmitting coil or conductor connected to ground and to an elevated terminal respectively of a receiving-circuit having a period of vibration corresponding to that of the transmitting-circuit and similarly connected to ground and to an elevated terminal and so arranged that the elevated terminal is charged to the highest potential developed in the circuit, as set forth.

NIKOLA TESLA.

Witnesses:

PARKER W. PAGE,

MARCELLUS BAILEY.

No. 685,957. Patented Nov. 5, 1901.

N. TESLA.

APPARATUS FOR THE UTILIZATION OF RADIANT ENERGY.

(Application filed Mar. 21, 1901.)

(No Model.)

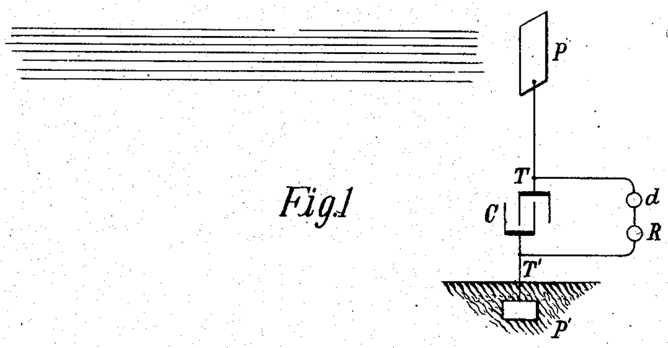

Fig.1

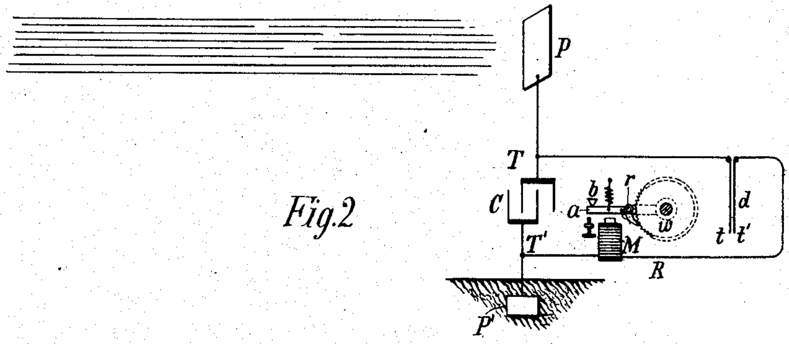
P
T
C
b
r
a
w
d
M
t
t'
T'
R
P'

Fig.2

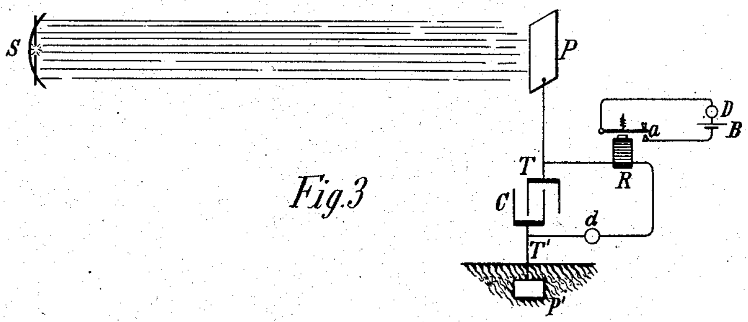
S
P
D
B
a
T
R
C
d
T'
P'

Fig.3

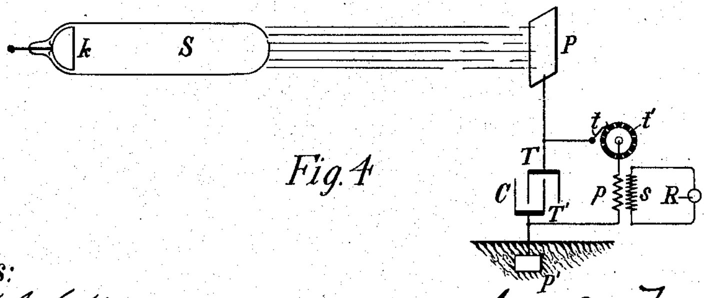
k
S
P
t
t'
T
C
p
s
R
T'
P'

Fig.4

Witnesses:
Raphaël Netter
M. Lawson Dyer.

Inventor
Nikola Tesla
by Kerr, Page & Cooper Att'ys.

THE NORRIS PETERS CO., PHOTO-LITHO., WASHINGTON, D. C.

UNITED STATES PATENT OFFICE.

NIKOLA TESLA, OF NEW YORK, N. Y.

APPARATUS FOR THE UTILIZATION OF RADIANT ENERGY.

SPECIFICATION forming part of Letters Patent No. 685,957, dated November 5, 1901.

Application filed March 21, 1901. Serial No. 52,153. (No model.)

To all whom it may concern:

Be it known that I, NIKOLA TESLA, a citizen of the United States, residing at the borough of Manhattan, in the city, county, and State of New York, have invented certain new and useful Improvements in Apparatus for the Utilization of Radiant Energy, of which the following is a specification, reference being had to the drawings accompanying and forming a part of the same.

It is well known that certain radiations—such as those of ultra-violet light, cathodic, Roentgen rays, or the like—possess the property of charging and discharging conductors of electricity, the discharge being particularly noticeable when the conductor upon which the rays impinge is negatively electrified. These radiations are generally considered to be ether vibrations of extremely small wave lengths, and in explanation of the phenomena noted it has been assumed by some authorities that they ionize or render conducting the atmosphere through which they are propagated. My own experiments and observations, however, lead me to conclusions more in accord with the theory heretofore advanced by me that sources of such radiant energy throw off with great velocity minute particles of matter which are strongly electrified, and therefore capable of charging an electrical conductor, or, even if not so, may at any rate discharge an electrified conductor either by carrying off bodily its charge or otherwise.

My present application is based upon a discovery which I have made that when rays or radiations of the above kind are permitted to fall upon an insulated conducting-body connected to one of the terminals of a condenser while the other terminal of the same is made by independent means to receive or to carry away electricity a current flows into the condenser so long as the insulated body is exposed to the rays, and under the conditions hereinafter specified an indefinite accumulation of electrical energy in the condenser takes place. This energy after a suitable time interval, during which the rays are allowed to act, may manifest itself in a powerful discharge, which may be utilized for the operation or control of mechanical or electrical devices or rendered useful in many other ways.

In applying my discovery I provide a condenser, preferably of considerable electrostatic capacity, and connect one of its terminals to an insulated metal plate or other conducting-body exposed to the rays or streams of radiant matter. It is very important, particularly in view of the fact that electrical energy is generally supplied at a very slow rate to the condenser, to construct the same with the greatest care. I use, by preference, the best quality of mica as dielectric, taking every possible precaution in insulating the armatures, so that the instrument may withstand great electrical pressures without leaking and may leave no perceptible electrification when discharging instantaneously. In practice I have found that the best results are obtained with condensers treated in the manner described in a patent granted to me February 23, 1897, No. 577,671. Obviously the above precautions should be the more rigorously observed the slower the rate of charging and the smaller the time interval during which the energy is allowed to accumulate in the condenser. The insulated plate or conducting-body should present as large a surface as practicable to the rays or streams of matter, I having ascertained that the amount of energy conveyed to it per unit of time is under otherwise identical conditions proportionate to the area exposed, or nearly so. Furthermore, the surface should be clean and preferably highly polished or amalgamated. The second terminal or armature of the condenser may be connected to one of the poles of a battery or other source of electricity or to any conducting body or object whatever of such properties or so conditioned that by its means electricity of the required sign will be supplied to the terminal. A simple way of supplying positive or negative electricity to the terminal is to connect the same either to an insulated conductor supported at some height in the atmosphere or to a grounded conductor, the former, as is well known, furnishing positive and the latter negative electricity. As the rays or supposed streams of matter generally convey a positive charge to the first condenser-terminal, which is connected to the plate or conductor above mentioned, I usually connect the second terminal of the condenser to the ground, this being the most convenient way of obtaining negative electricity, dispensing with the necessity of providing an artificial source. In order to utilize for any useful purpose the energy accumulated in the condenser, I furthermore connect to the terminals of the same a circuit including an instrument or apparatus which it is desired to operate and another instrument or device for alternately closing and opening the circuit. This latter may be any form of circuit-controller, with fixed or movable parts or electrodes, which may be actuated either by the stored energy or by independent means.

My discovery will be more fully understood from the following description and annexed drawings, to which reference is now made, and in which—

Figure 1 is a diagram showing the general arrangement of apparatus as usually employed. Fig. 2 is a similar diagram illustrating more in detail typical forms of the devices or elements used in practice, and Figs. 3 and 4 are diagrammatical representations of modified arrangements suitable for special purposes.

As illustrative of the manner in which the several parts or elements of the apparatus in one of its simplest forms are to be arranged and connected for useful operation, reference is made to Fig. 1, in which C is the condenser, P the insulated plate or conducting-body which is exposed to the rays, and P′ another plate or conductor which is grounded, all being joined in series, as shown. The terminals T T′ of the condenser are also connected to a circuit which includes a device R to be operated and a circuit-controlling device *d* of the character above referred to.

The apparatus being arranged as shown, it will be found that when the radiations of the sun or of any other source capable of producing the effects before described fall upon the plate P an accumulation of electrical energy in the condenser C will result. This phenomenon, I believe, is best explained as follows: The sun, as well as other sources of radiant energy, throws off minute particles of matter positively electrified, which, impinging upon the plate P, communicate continuously an electrical charge to the same. The opposite terminal of the condenser being connected to the ground, which may be considered as a vast reservoir of negative electricity, a feeble current flows continuously into the condenser, and inasmuch as these supposed particles are of an inconceivably small radius or curvature, and consequently charged to a relatively very high potential, this charging of the condenser may continue, as I have actually observed, almost indefinitely, even to the point of rupturing the dielectric. If the device *d* be of such character that it will operate to close the circuit in which it is included when the potential in the condenser has reached a certain magnitude, the accumulated charge will pass through the circuit, which also includes the receiver R, and operate the latter.

In illustration of a particular form of apparatus which may be used in carrying out my discovery I now refer to Fig. 2. In this figure, which in the general arrangement of the elements is identical to Fig. 1, the device *d* is shown as composed of two very thin conducting-plates *t t′*, placed in close proximity and very mobile, either by reason of extreme flexibility or owing to the character of their support. To improve their action, they should be inclosed in a receptacle, from which the air may be exhausted. The plates *t t′* are connected in series with a working circuit, including a suitable receiver, which in this case is shown as consisting of an electromagnet M, a movable armature *a*, a retractile spring *b*, and a ratchet-wheel *w*, provided with a spring-pawl *r*, which is pivoted to armature *a*, as illustrated. When the radiations of the sun or other radiant source fall upon plate P, a current flows into the condenser, as above explained, until the potential therein rises sufficiently to attract and bring into contact the two plates *t t′*, and thereby close the circuit connected to the two condenser-terminals. This permits a flow of current which energizes the magnet M, causing it to draw down the armature *a* and impart a partial rotation to the ratchet-wheel *w*. As the current ceases the armature is retracted by the spring *b*, without, however, moving the wheel *w*. With the stoppage of the current the plates *t t′* cease to be attracted and separate, thus restoring the circuit to its original condition.

Fig. 3 shows a modified form of apparatus used in connection with an artificial source of radiant energy, which in this instance may be an arc emitting copiously ultra-violet rays. A suitable reflector may be provided for concentrating and directing the radiations. A magnet R and circuit-controller *d* are arranged as in the previous figures; but in the present case the former instead of performing itself the whole work only serves the purpose of alternately opening and closing a local circuit, containing a source of current B and a receiving or translating device D. The controller *d*, if desired, may consist of two fixed electrodes separated by a minute air-gap or weak dielectric film, which breaks down more or less suddenly when a definite difference of potential is reached at the terminals of the condenser and returns to its original state upon the passage of the discharge.

Still another modification is shown in Fig. 4, in which the source S of radiant energy is a special form of Roentgen tube devised by me, having but one terminal *k*, generally of aluminium, in the form of half a sphere, with a plain polished surface on the front side, from which the streams are thrown off. It may be excited by attaching it to one of the terminals of any generator of sufficiently high electromotive force; but whatever apparatus be used it is important that the tube be exhausted to a high degree, as otherwise it might prove entirely ineffective. The working or discharge circuit connected to the terminals T T′ of the condenser includes in this case the primary *p* of a transformer and a circuit-controller comprising a fixed terminal or brush *t* and a movable terminal *t′* in the shape of a wheel, with conducting and insulating segments, which may be rotated at an arbitrary speed by any suitable means. In inductive relation to the primary wire or coil *p* is a secondary *s*, usually of a much greater number of turns, to the ends of which is connected a receiver R. The terminals of the condenser being connected, as indicated, one to an insulated plate P and the other to a grounded plate P′, when the tube S is excited rays or streams of matter are emitted from the same, which convey a positive charge to the plate P and condenser-terminal T, while terminal T′ is continuously receiving negative electricity from the plate P′. This, as before explained, results in an accumulation of electrical energy in the condenser, which goes on as long as the circuit including the primary *p* is interrupted. Whenever the circuit is closed owing to the rotation of the terminal *t′*, the stored energy is discharged through the primary *p*, this giving rise in the secondary *s* to induced currents, which operate the receiver R.

It is clear from what has been stated above that if the terminal T′ is connected to a plate supplying positive instead of negative electricity the rays should convey negative electricity to plate P. The source S may be any form of Roentgen or Lenard tube; but it is obvious from the theory of action that in order to be very effective the electrical impulses exciting it should be wholly or at least preponderatingly of one sign. If ordinary symmetrical alternating currents are employed, provision should be made for allowing the rays to fall upon the plate P only during those periods when they are productive of the desired result. Evidently if the radiations of the source be stopped or intercepted or their intensity varied in any manner, as by periodically interrupting or rythmically varying the current exciting the source, there will be corresponding changes in the action upon the receiver R, and thus signals may be transmitted and many other useful effects produced. Furthermore, it will be understood that any form of circuit-closer which will respond to or be set in operation when a predetermined amount of energy is stored in the condenser may be used in lieu of the device specifically described with reference to Fig. 2 and also that the special details of construction and arrangement of the several parts of the apparatus may be very greatly varied without departure from the invention.

Having described my invention, what I claim is—

1. An apparatus for utilizing radiant energy, comprising in combination a condenser, one armature of which is subjected to the action of rays or radiations, independent means for charging the other armature, a circuit and apparatus therein adapted to be operated or controlled by the discharge of the condenser, as set forth.

2. An apparatus for utilizing radiant energy, comprising in combination, a condenser, one armature of which is subjected to the action of rays or radiations, independent means for charging the other armature, a local circuit connected with the condenser-terminals, a circuit-controller therein and means adapted to be operated or controlled by the discharge of the condenser when the local circuit is closed, as set forth.

3. An apparatus for utilizing radiant energy, comprising in combination, a condenser, one terminal of which is subjected to the action of rays or radiations, independent means for charging the other armature, a local circuit connected with the condenser-terminals, a circuit-controller therein dependent for operation on a given rise of potential in the condenser, and devices operated by the discharge of the condenser when the local circuit is closed, as set forth.

4. An apparatus for utilizing radiant energy, comprising in combination, a condenser, one terminal of which is subjected to the action of rays or radiations, and the other of which is connected with the ground, a circuit and apparatus therein adapted to be operated by the discharge of the accumulated energy in the condenser, as set forth.

5. An apparatus for utilizing radiant energy, comprising in combination, a condenser, one terminal of which is subjected to the action of rays or radiations and the other of which is connected with the ground, a local circuit connected with the condenser-terminals, a circuit-controller therein and means adapted to be operated by the discharge of the condenser when the local circuit is closed, as set forth.

6. An apparatus for utilizing radiant energy, comprising in combination, a condenser, one terminal of which is subjected to the action of rays or radiations and the other of which is connected with the ground, a local circuit connected with the condenser-terminals, a circuit-controller therein adapted to be operated by a given rise of potential in the condenser, and devices operated by the discharge of the condenser when the local circuit is closed, as set forth.

7. An apparatus for utilizing radiant en-

ergy, comprising a condenser, having one terminal connected to earth and the other to an elevated conducting-plate, which is adapted to receive the rays from a distant source of radiant energy, a local circuit connected with the condenser-terminals, a receiver therein, and a circuit-controller therefor which is adapted to be operated by a given rise of potential in the condenser, as set forth.

NIKOLA TESLA.

Witnesses:
M. LAWSON DYER,
RICHARD DONOVAN.

455-619 AU 233 EX
FIP8106 OR 685,958

55

No. 685,958. Patented Nov. 5, 1901.

N. TESLA.

METHOD OF UTILIZING RADIANT ENERGY.

(Application filed Mar. 21, 1901.)

(No Model.)

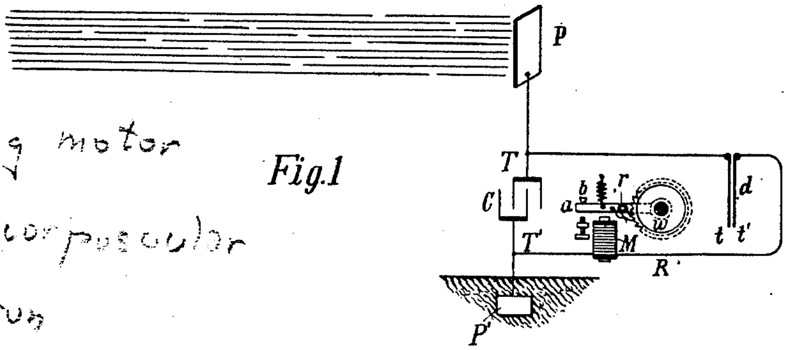

Fig. 1

Electric steping motor energized by corpuscular energy from sun

No refs.

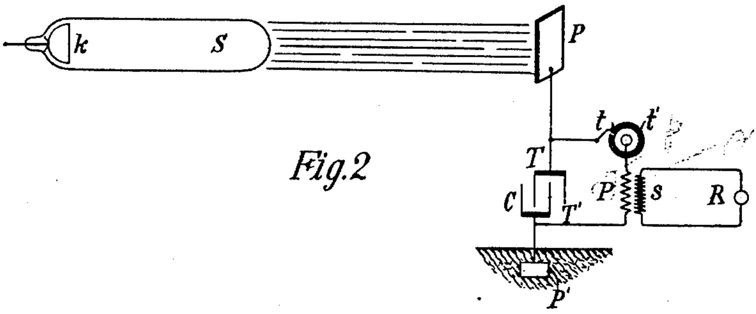

Fig. 2

Pickup

Witnesses:
Raphaël Netter
M. Lawson Dyer

Nikola Tesla, Inventor
by Kerr, Page & Cooper Attys

UNITED STATES PATENT OFFICE.

NIKOLA TESLA, OF NEW YORK, N. Y.

METHOD OF UTILIZING RADIANT ENERGY.

SPECIFICATION forming part of Letters Patent No. 685,958, dated November 5, 1901.

Application filed March 21, 1901. Serial No. 52,154. (No model.)

To all whom it may concern:

Be it known that I, NIKOLA TESLA, a citizen of the United States, residing at the borough of Manhattan, in the city, county, and State of New York, have invented certain new and useful Improvements in Methods of Utilizing Radiant Energy, of which the following is a specification, reference being had to the drawings accompanying and forming a part of the same.

It is well known that certain radiations—such as those of ultra-violet light, cathodic, Roentgen rays, or the like—possess the property of charging and discharging conductors of electricity, the discharge being particularly noticeable when the conductor upon which the rays impinge is negatively electrified. These radiations are generally considered to be ether vibrations of extremely small wave lengths, and in explanation of the phenomena noted it has been assumed by some authorities that they ionize or render conducting the atmosphere through which they are propagated. My own experiments and observations, however, lead me to conclusions more in accord with the theory heretofore advanced by me that sources of such radiant energy throw off with great velocity minute particles of matter which are strongly electrified, and therefore capable of charging an electrical conductor, or even if not so may at any rate discharge an electrified conductor either by carrying off bodily its charge or otherwise.

My present application is based upon a discovery which I have made that when rays or radiations of the above kind are permitted to fall upon an insulated conducting body connected to one of the terminals of a condenser, while the other terminal of the same is made by independent means to receive or to carry away electricity, a current flows into the condenser so long as the insulated body is exposed to the rays, and under the conditions hereinafter specified an indefinite accumulation of electrical energy in the condenser takes place. This energy after a suitable time interval, during which the rays are allowed to act, may manifest itself in a powerful discharge, which may be utilized for the operation or control of mechanical or electrical devices or rendered useful in many other ways.

In applying my discovery I provide a condenser, preferably of considerable electrostatic capacity, and connect one of its terminals to an insulated metal plate or other conducting body exposed to the rays or streams of radiant matter. It is very important, particularly in view of the fact that electrical energy is generally supplied at a very slow rate to the condenser, to construct the same with the greatest care. I use by preference the best quality of mica as dielectric, taking every possible precaution in insulating the armatures, so that the instrument may withstand great electrical pressures without leaking and may leave no perceptible electrification when discharging instantaneously. In practice I have found that the best results are obtained with condensers treated in the manner described in a patent granted to me February 23, 1897, No. 577,671. Obviously the above precautions should be the more rigorously observed the slower the rate of charging and the smaller the time interval during which the energy is allowed to accumulate in the condenser. The insulated plate or conducting body should present as large a surface as practicable to the rays or streams of matter, I having ascertained that the amount of energy conveyed to it per unit of time is under otherwise identical conditions proportionate to the area exposed, or nearly so. Furthermore, the surface should be clean and preferably highly polished or amalgamated. The second terminal or armature of the condenser may be connected to one of the poles of a battery or other source of electricity or to any conducting body or object whatever of such properties or so conditioned that by its means electricity of the required sign will be supplied to the terminal. A simple way of supplying positive or negative electricity to the terminal is to connect the same either to an insulated conductor, supported at some height in the atmosphere, or to a grounded conductor, the former, as is well known, furnishing positive and the latter negative electricity. As the rays or supposed streams of matter generally convey a positive charge to the first condenser-terminal, which is connect-

RADIANT ENERGY.

EXAMINER

ed to the plate or conductor above mentioned, I usually connect the second terminal of the condenser to the ground, this being the most convenient way of obtaining negative electricity, dispensing with the necessity of providing an artificial source. In order to utilize for any useful purpose the energy accumulated in the condenser, I furthermore connect to the terminals of the same a circuit including an instrument or apparatus which it is desired to operate and another instrument or device for alternately closing and opening the circuit. This latter may be any form of circuit-controller, with fixed or movable parts or electrodes, which may be actuated either by the stored energy or by independent means.

The rays or radiations which are to be utilized for the operation of the apparatus above described in general terms may be derived from a natural source, as the sun, or may be artificially produced by such means, for example, as an arc-lamp, a Roentgen tube, and the like, and they may be employed for a great variety of useful purposes.

My discovery will be more fully understood from the following detailed description and annexed drawings, to which reference is now made, and in which—

Figure 1 is a diagram showing typical forms of the devices or elements as arranged and connected in applying the method for the operation of a mechanical contrivance or instrument solely by the energy stored; and Fig. 2 is a diagrammatical representation of a modified arrangement suitable for special purposes, with a circuit-controller actuated by independent means.

Referring to Fig. 1, C is the condenser, P the insulated plate or conducting body, which is exposed to the rays, and P′ another plate or conductor, all being joined in series, as shown. The terminals T T′ of the condenser are also connected to a circuit including a receiver R, which is to be operated, and a circuit-controlling device *d*, which in this case is composed of two very thin conducting-plates *t t′*, placed in close proximity and very mobile, either by reason of extreme flexibility or owing to the charater of their support. To improve their action, they should be inclosed in a receptacle from which the air may be exhausted. The receiver R is shown as consisting of an electromagnet M, a movable armature *a*, a retractile spring *b*, and a ratchet-wheel *w*, provided with a spring-pawl *r*, which is pivoted to armature *a*, as illustrated. The apparatus being arranged as shown, it will be found that when the radiations of the sun or of any other source capable of producing the effects before described fall upon the plate P an accumulation of electrical energy in the condenser C will result. This phenomenon, I believe, is best explained as follows: The sun as well as other sources of radiant energy throw off minute particles of matter positively electrified, which, impinging upon the plate P, communicate an electrical charge to the same. The opposite terminal of the condenser being connected to the ground, which may be considered as a vast reservoir of negative electricity, a feeble current flows continuously into the condenser, and inasmuch as these supposed particles are of an inconceivably small radius or curvature, and consequently charged to a relatively very high potential, this charging of the condenser may continue, as I have found in practice, almost indefinitely, even to the point of rupturing the dielectric. Obviously whatever circuit-controller be employed it should operate to close the circuit in which it is included when the potential in the condenser has reached the desired magnitude. Thus in Fig. 2 when the electrical pressure at the terminals T T′ rises to a certain predetermined value the plates *t t′*, attracting each other, close the circuit connected to the terminals. This permits a flow of current which energizes the magnet M, causing it to draw down the armature *a* and impart a partial rotation to the ratchet-wheel *w*. As the current ceases the armature is retracted by the spring *b* without, however, moving the wheel *w*. With the stoppage of the current the plates *t t′* cease to be attracted and separate, thus restoring the circuit to its original condition.

Many useful applications of this method of utilizing the radiations emanating from the sun or other source and many ways of carrying out the same will at once suggest themselves from the above description. By way of illustration a modified arrangement is shown in Fig. 2, in which the source S of radiant energy is a special form of Roentgen tube devised by me having but one terminal *k*, generally of aluminium, in the form of half a sphere with a plain polished surface on the front side, from which the streams are thrown off. It may be excited by attaching it to one of the terminals of any generator of sufficiently-high electromotive force; but whatever apparatus be used it is important that the tube be exhausted to a high degree, as otherwise it might prove entirely ineffective. The working or discharge circuit connected to the terminals T T′ of the condenser includes in this case the primary *p* of a transformer and a circuit-controller comprising a fixed terminal or brush *t* and a movable terminal *t′* in the shape of a wheel with conducting and insulating segments which may be rotated at an arbitrary speed by any suitable means. In inductive relation to the primary wire or coil *p* is a secondary *s*, usually of a much greater number of turns, to the ends of which is connected a receiver R. The terminals of the condenser being connected as indicated, one to an insulated plate P and the other to a grounded plate P′, when the tube S is excited rays or streams of matter are emitted from the same, which convey a positive charge to the plate P and condenser-terminal T, while terminal T′ is continuously receiving negative electricity from the plate

P′. This, as before explained, results in an accumulation of electrical energy in the condenser, which goes on as long as the circuit including the primary p is interrupted. Whenever the circuit is closed, owing to the rotation of the terminal t', the stored energy is discharged through the primary p, this giving rise in the secondary s to induced currents which operate the receiver R.

It is clear from what has been stated above that if the terminal T′ is connected to a plate supplying positive instead of negative electricity the rays should convey negative electricity to plate P. The source S may be any form of Roentgen or Lenard tube; but it is obvious from the theory of action that in order to be very effective the electrical impulses exciting it should be wholly or at least preponderatingly of one sign. If ordinary symmetrical alternating currents are employed, provision should be made for allowing the rays to fall upon the plate P only during those periods when they are productive of the desired result. Evidently if the radiations of the source be stopped or intercepted or their intensity varied in any manner, as by periodically interrupting or rythmically varying the current exciting the source, there will be corresponding changes in the action upon the receiver R, and thus signals may be transmitted and many other useful effects produced. Furthermore, it will be understood that any form of circuit-closer which will respond to or be set in operation when a predetermined amount of energy is stored in the condenser may be used in lieu of the device specifically described with reference to Fig. 1, and also that the special details of construction and arrangement of the several parts of the apparatus may be very greatly varied without departure from the invention.

Having described my invention, what I claim is—

1. The method of utilizing radiant energy, which consists in charging one of the armatures of a condenser by rays or radiations, and the other armature by independent means, and discharging the condenser through a suitable receiver, as set forth.

2. The method of utilizing radiant energy, which consists in simultaneously charging a condenser by means of rays or radiations and an independent source of electrical energy, and discharging the condenser through a suitable receiver, as set forth.

3. The method of utilizing radiant energy, which consists in charging one of the armatures of a condenser by rays or radiations, and the other by independent means, controlling the action or effect of said rays or radiations and discharging the condenser through a suitable receiver, as set forth.

4. The method of utilizing radiant energy, which consists in charging one of the armatures of a condenser by rays or radiations and the other by independent means, varying the intensity of the said rays or radiations and periodically discharging the condenser through a suitable receiver, as set forth.

5. The method of utilizing radiant energy, which consists in directing upon an elevated conductor, connected to one of the armatures of a condenser, rays or radiations capable of positively electrifying the same, carrying off electricity from the other armature by connecting the same with the ground, and discharging the accumulated energy through a suitable receiver, as set forth.

6. The method of utilizing radiant energy, which consists in charging one of the armatures of a condenser by rays or radiations, and the other by independent means, and effecting by the automatic discharge of the accumulated energy the operation or control of a suitable receiver, as set forth.

NIKOLA TESLA.

Witnesses:
M. LAWSON DYER,
RICHARD DONOVAN.

No. 837,616. PATENTED DEC. 4, 1906.

H. H. C. DUNWOODY.

WIRELESS TELEGRAPH SYSTEM.

APPLICATION FILED MAR. 23, 1906.

Fig. 1.

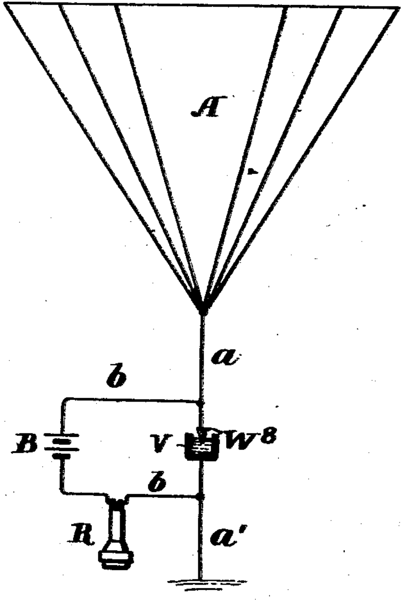
A
a
b
B
V
W⁸
b
R
a'

Fig. 2.

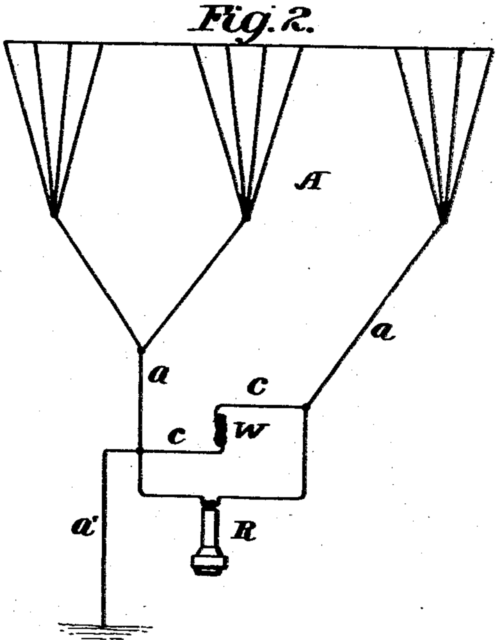

Fig. 3.

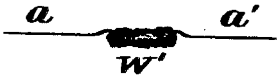

Fig. 4.

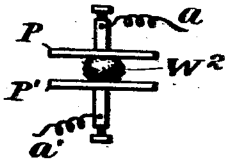

Fig. 5.

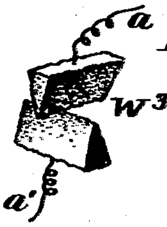

Fig. 6.

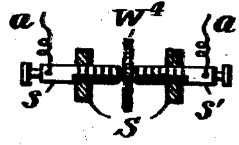

Fig. 7.

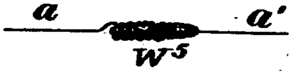

Fig. 8.

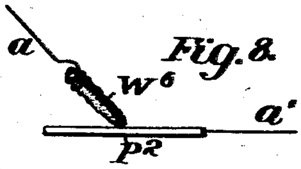

Fig. 12.

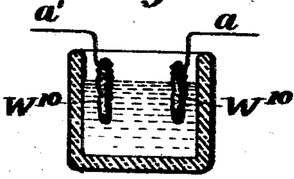

Fig. 11.

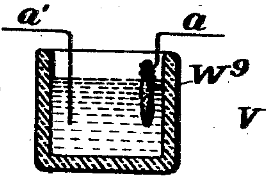

Fig. 10.

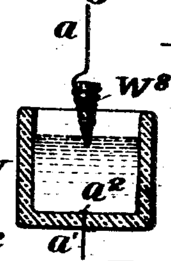

Fig. 9.

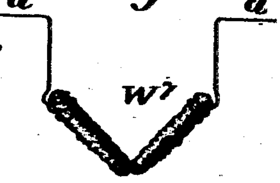

Inventor

Witnesses

J. G. Hinkel

J. J. McCarthy

Fig. 14.

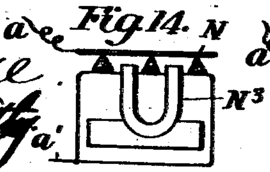

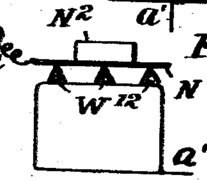

Fig. 13.

Henry H. C. Dunwoody

by Foster Freeman Watson

Attorneys

UNITED STATES PATENT OFFICE.

HENRY H. C. DUNWOODY, OF WASHINGTON, DISTRICT OF COLUMBIA.

WIRELESS-TELEGRAPH SYSTEM.

No. 837,616. Specification of Letters Patent. Patented Dec. 4, 1906.

Application filed March 23, 1906. Serial No. 307,679.

To all whom it may concern:

Be it known that I, Henry H. C. Dunwoody, a citizen of the United States, residing at Washington, in the District of Columbia, have invented certain new and useful Improvements in Wireless-Telegraph Systems, of which the following is a specification.

My invention relates to wireless telegraph or signaling systems, and has for its object more particularly to provide an improved wave responsive or detecting device, as well as an improved arrangement of such device in connection with the ordinary apparatus of the receiving-station; and to these ends my invention consists in a wave responsive or detecting device, of the character substantially specified, arranged and operating substantially in the manner hereinafter pointed out.

In the accompanying drawings, wherein I have indicated some of the ways in which I have utilized my invention, Figure 1 is a diagram of a wave receiving and signaling apparatus of a simple form. Fig. 2 is another diagram; and Figs. 3 to 14 are detailed views showing some specific embodiments and arrangements of the wave-responsive device, which will be more particularly described hereinafter.

My improved wave responsive or detecting device, broadly stated, comprises a mass or body of non-metallic crystalline material, and in practice I have generally used such a mass or body comprising crystalline silicid of carbon or carborundum. This material may be in any desired shape or size, and is generally in the form of a concrete mass or body of crystals which may vary in size and character, and which are composed of carbon and silicon in a chemical combination forming what is chemically known as a "carbid" of silicon, or silicid or carborundum, and is a highly-refractory material, extremely hard and a relatively poor conductor of electricity. This wave-responsive material may be utilized in many and various forms and connected in the circuit of the signal-receiving apparatus in many and various ways, some of which are indicated in the annexed drawings, and I do not, therefore, limit myself to any particular manner of using the material.

In the drawings, A represents an aerial receiving conductor or wave collector or antennæ which may be of any well-known or desirable form, being shown, for convenience, as embodying what is generally known as the "fan" type of aerial receiver. This receiver or collector is connected to a conductor *a*, having a ground terminal *a'*, and interposed in this conductor is the wave-responsive device W'.

In Fig. 1 there is a branch circuit including a battery B and conductors *b b'*, which are connected to the conductors *a a'* at opposite sides of the wave-responsive device W'. This circuit includes a signal-receiving device, shown in the present case as in the form of an ordinary magnetotelephone receiver R, in practice this being usually embodied in the ordinary head-telephone form for convenience.

The wave-responsive device W', indicated in Fig. 1, may be of many and various forms and in Fig. 3 I have indicated it in the form of a mass or body of crystalline material as carborundum, to which the ends of the wires or conductors *a a'* are secured in any suitable way as by being twisted or bound tightly around the ends of the body.

In Fig. 4 the wave-responsive device is shown in the form of a regular mass W^2, of the crystalline material supported between two plates or bodies P P' of conducting material, to which the conductors *a a'* are connected.

In Fig. 5 the wave-responsive device W^3 comprises two pieces or bodies of the crystalline material or carborundum having at least one relatively sharp edge, and the edges of the two pieces are placed in contact and may be supported in this relation in any suitable way, and are included in the circuit as before.

In Fig. 6 the wave-responsive device W^4 is shown in the form of a relatively thin plate or mass of carborundum supported between adjusting-screws *s s'*, mounted in a suitable frame or support S and connected in the electric circuit.

In Fig. 7 I have indicated a concrete mass W^5 of the material, to one end of which the conductor *a* is connected in any suitable way, as by being turned or twisted around the body, while the other terminal or conductor *a'* simply rests in contact with the mass.

In Fig. 8 the material W^6 is connected to one of the conductors or wires *a* and rests by gravity upon a plate P^2 connected to the other conductor or wire a'.

In Fig. 9 the wave-responsive device W^7 comprises two bodies of the crystalline material connected to the conductors as before, but having portions of their bodies in relatively loose contact.

In Fig. 10, I have shown a vessel V containing some electrolyte, as mercury, or an acid, or an alkali fluid, and the wave-responsive device W^8 is preferably in the form of a pointed mass, and at one end is connected to the conductor a the smaller or pointed mass being immersed to a greater or less extent in the electrolyte, and the other conductor a' is either connected to the vessel V, if it is of conducting material, or to a separate electrode or terminal a^2 immersed in the electrolytic fluid.

In Fig. 11 the wave-responsive device comprises the mass W^9 of the crystalline material connected to one conductor a immersed in an electrolyte, while the other conductor a' extends into the electrolyte at a distance away from the mass.

In Fig. 12 is shown a substantially similar arrangement, except that there are two bodies or masses of crystalline wave-responsive material W^{10} immersed in the electrolyte.

In Figs. 13 and 14 I have shown another embodiment of my invention, wherein N represents a needle or conductor resting upon the edges of several pieces or blocks W^{12} of carborundum, which are electrically connected together. In some instances it is desirable to hold the needle or conductor in position as by a weight N^2, or a permanent magnet N^3, and in these cases, the needle and carborundum are respectively connected to the terminals, as indicated.

I have found that all of these various forms or embodiments of my invention are operative as wave-responsive devices, and I have made other embodiments and arrangements, not necessary to recite herein, these being sufficient to show the various manners of using my improved wave-responsive device or material.

In Fig. 2, I have illustrated an arrangement which I have found to be practically operative, in which there is an aerial receiver or wave-collector A, in which the antennæ are connected by the conductors a a to the ground conductor or antennæ a', and there is a signal-receiver R in the circuit from a portion of the antennæ between them and the ground. The receiver is bridged by a shunt-circuit c c', including a wave-responsive device W, there being no battery or other independent source of electric current. With this arrangement of circuits in connection with my improved wave-responsive device, I have been enabled to receive signals sent several hundred miles distant from the receiving-station.

It will be understood that any and all of the various forms of wave-responsive devices illustrated can be used in either of the arrangements of circuits described and illustrated in Figs. 1 and 2, or in any equivalent arrangement of circuits.

It will be understood that in actual practice, it is desirable to use some of the well-known tuning devices in connection with the wave-responsive device, as these tuning devices suppress a great many false signals, which would otherwise interfere to a greater or less extent with the understanding of the message, but I do not deem it necessary to show or describe any of these well-known adjuncts, as the wave-responsive device is operative without them as well as with them.

One of the advantages of the use of my improved wave-responsive material, especially in connection with an electrolyte in the manner indicated in Figs. 10 to 12 for instance, is that it is not only very sensitive as a receiver, but I have also found that it possesses the great advantage that the carborundum will not burn out or be destroyed when the receiver is located near a sending-station. Or, in other words, that the intense disturbances or waves produced at a sending-station located in proximity to my improved receiver, do not destroy or injure the same, as is the case with many other responsive devices of the electrolytic type.

If desired, the wave-responsive devices shown in Figs. 10, 11 and 12, may have a coating of glass or other insulating material of such form as to leave the desired area of the mass exposed.

What I claim is—

1. In the art of wireless telegraphy, as a wave-responsive device, a mass of non-metallic crystalline material.

2. In the art of wireless telegraphy, a wave-responsive device consisting essentially of carbon and silicon.

3. In the art of wireless telegraphy, as a wave-responsive device carbon and silicon in chemical combination.

4. In the art of wireless telegraphy, a wave-responsive device consisting essentially of crystalline carbon and silicon.

5. In the art of wireless telegraphy, as a wave-responsive device a concrete mass of crystalline silicide of carbon.

6. In the art of wireless telegraphy, as a wave-responsive device a body of carborundum.

7. In the art of wireless telegraphy, as a wave-responsive device an electrolyte and a mass of non-metallic crystalline material.

8. In the art of wireless telegraphy, as a

wave-responsive device an electrolyte and electrodes of non-metallic crystalline material extending therein.

9. In the art of wireless telegraphy, a receiving apparatus comprising an aerial wire, a ground connection, a signal device in the ground connection, and a shunt-circuit bridging the signaling device and including as a wave-responsive device a mass of carborundum.

In testimony whereof I affix my signature in presence of two witnesses.

HENRY H. C. DUNWOODY.

Witnesses:

FRANK L. FREEMAN,
MILTON TIBBETTS.

No. 887,357. PATENTED MAY 12, 1908.

N. B. STUBBLEFIELD.

WIRELESS TELEPHONE.

APPLICATION FILED APR. 5, 1907.

3 SHEETS—SHEET 1.

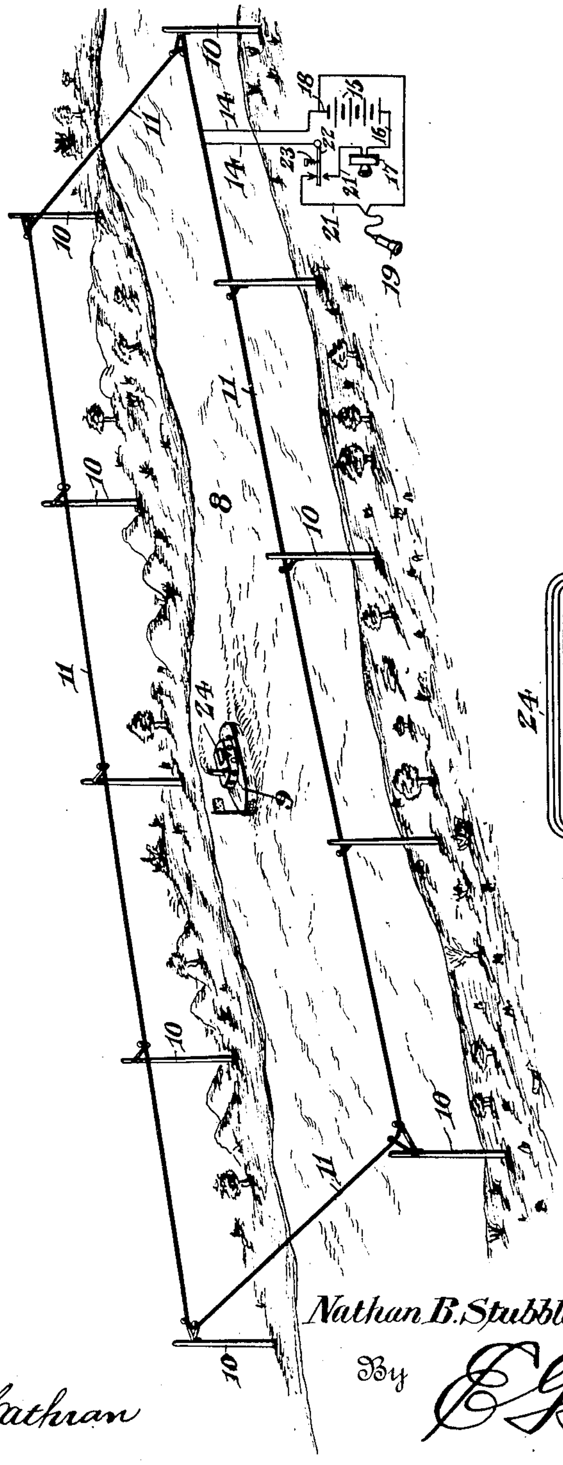
10
11
14
18
15
22
16
23
21
17
19
8
24
9

Fig. 1.

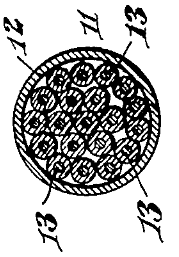
12
11
13
13
13
13

Fig. 3.

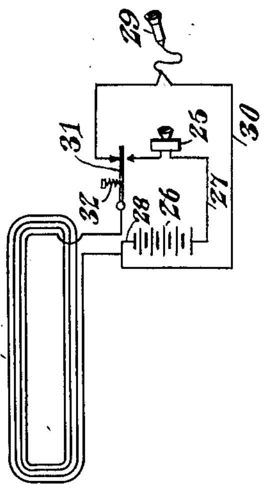
29
31
25
30
32
28
26
27
24

Fig. 2.

Witnesses

Jas. K. McCathran

B. G. Foster

Nathan B. Stubblefield, Inventor

By E. G. Siggers

Attorney

No. 887,357. PATENTED MAY 12, 1908.

N. B. STUBBLEFIELD.

WIRELESS TELEPHONE.

APPLICATION FILED APR. 5, 1907.

3 SHEETS—SHEET 2.

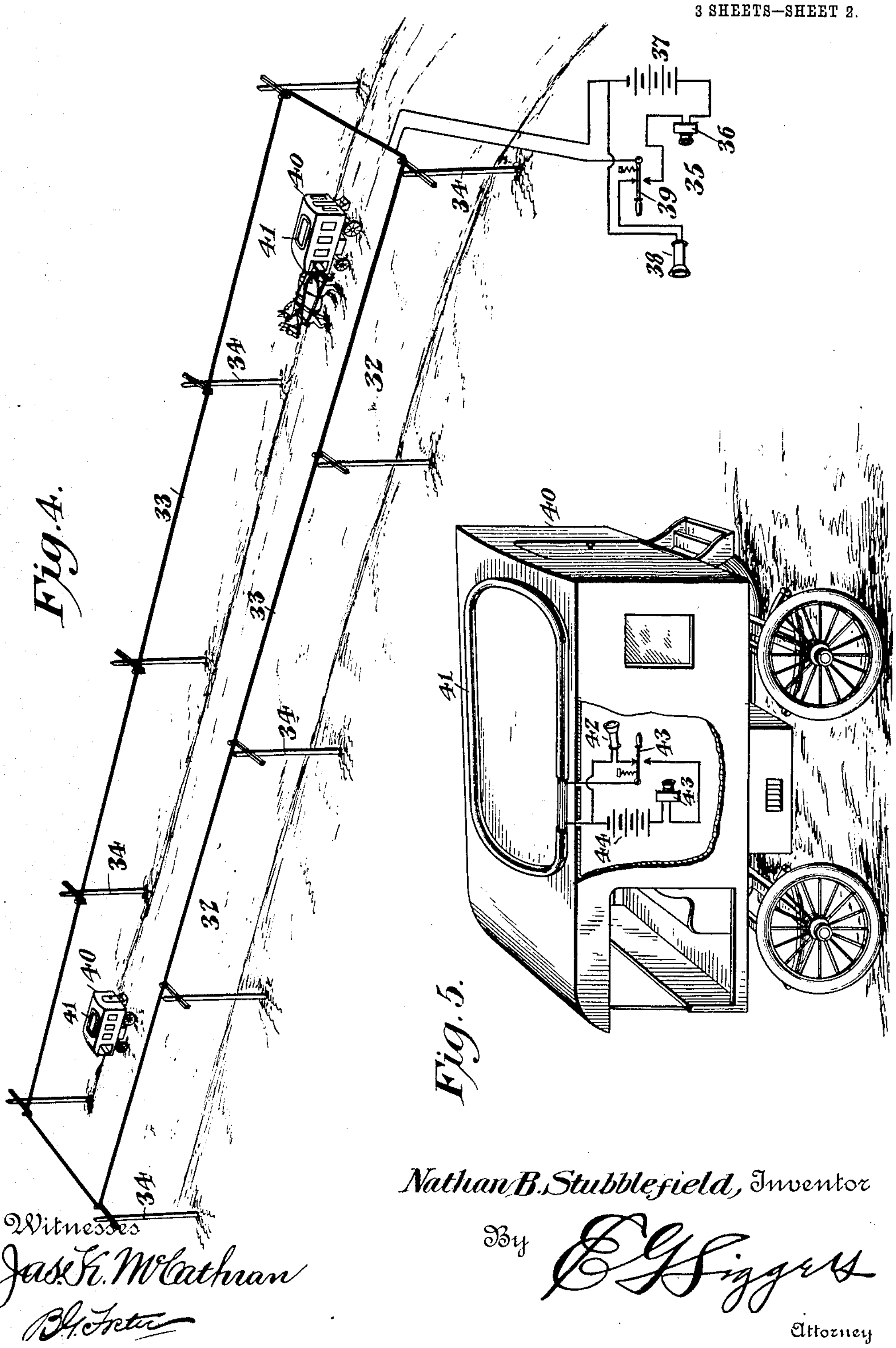
37
36
35
39
38
34
40
41
34
32
33
33
34
34
32
40
41
34

Fig. 4.

40
41
42
43
43
44

Fig. 5.

Witnesses
Jas. K. McCathran

Nathan B. Stubblefield, Inventor

By C. G. Siggers

Attorney

No. 887,357. PATENTED MAY 12, 1908.

N. B. STUBBLEFIELD.

WIRELESS TELEPHONE.

APPLICATION FILED APR. 5, 1907.

3 SHEETS—SHEET 3.

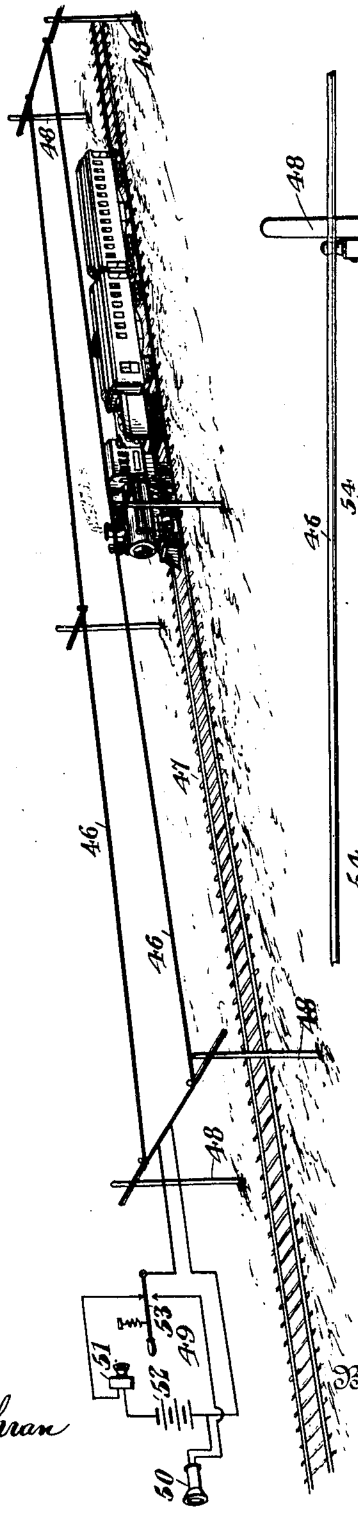
48
48
46
46
47
48
48
49
51
52
53
50

Fig. 6.

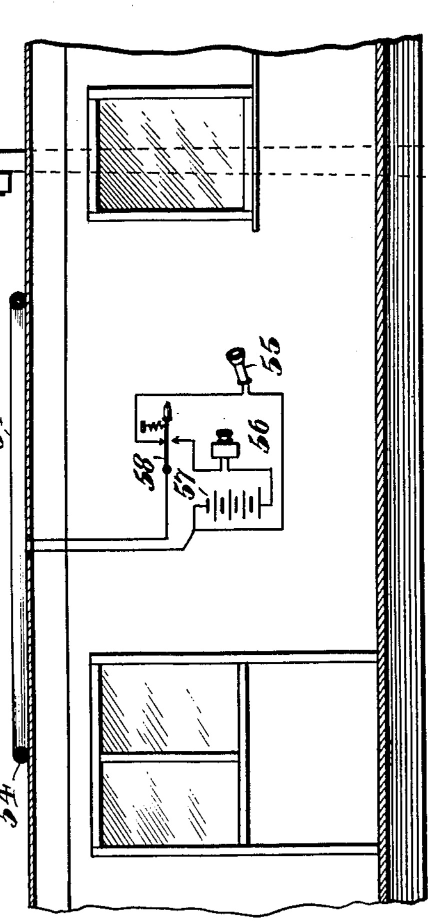
48
46
54
54
55
56
57
58

Fig. 7.

Witnesses
Jas. K. McCathran
B. G. Foster

Nathan B. Stubblefield,
Inventor

By C. G. Siggers
Attorney

UNITED STATES PATENT OFFICE.

NATHAN B. STUBBLEFIELD, OF MURRAY, KENTUCKY, ASSIGNOR OF TWELVE AND ONE-HALF ONE-HUNDREDTHS TO CONN LINN, FIVE ONE-HUNDREDTHS TO R. DOWNS, FIVE ONE-HUNDREDTHS TO B. F. SCHROADER, FIVE ONE-HUNDREDTHS TO GEORGE C. McLARIN, FIVE ONE-HUNDREDTHS TO JOHN P. McELRATH, TWO AND ONE-HALF ONE-HUNDREDTHS TO JEFF D. ROULETT, AND ONE-TWENTIETH TO SAMUEL E. BYNUM, ALL OF MURRAY, KENTUCKY.

WIRELESS TELEPHONE.

No. 887,357. Specification of Letters Patent. Patented May 12, 1908.

Application filed April 5, 1907. Serial No. 366,544.

To all whom it may concern:

Be it known that I, Nathan B. Stubblefield, a citizen of the United States, residing at Murray, in the county of Calloway and State of Kentucky, have invented a new and useful Wireless Telephone, of which the following is a specification.

The present invention relates to means for electrically transmitting signals from one point to another without the use of connecting wires, and more particularly comprehending means for securing telephonic communication between moving vehicles and way stations.

The principal object of the invention is to provide simple and practical means of a novel nature whereby clear and audible communication can be established, said means being simple and of a character that will permit certain of the station mechanisms to be small and compact.

In the accompanying drawings:—Figure 1 is a perspective view, showing means for establishing communication between a vessel and a shore station. Fig. 2 is a diagrammatic view of the mechanism mounted on the boat. Fig. 3 is a cross sectional view on an enlarged scale of the shore coil. Fig. 4 is a perspective view of a road-way, showing a system for establishing communication between road vehicles and a way-station, the latter being illustrated diagrammatically. Fig. 5 is a detail view of a vehicle equipped with one of the instruments, which is shown diagrammatically. Fig. 6 is a perspective view showing the system applied to a railway for establishing communication between a moving train and a way-station. Fig. 7 is a sectional view through a car showing in diagram the car mechanism illustrated in Fig. 6.

Similar reference numerals designate corresponding parts in all the figures of the drawings.

Referring to the embodiment illustrated in Figs. 1, 2 and 3, a water-way 8 is disclosed, upon which a vessel 9 operates. Surrounding the path of travel of the vessel, and preferably elevated on poles 10, is a coil 11 of considerable magnitude. This coil, as shown in Fig. 3, consists of an outer casing 12, within which is placed a conducting wire comprising a plurality of convolutions 13, each of which is insulated from the other. The terminals 14 of this coil extend to a suitable way-station, and at the station is located a powerful source of electrical energy 15, to which is connected by a suitable wire 16 an electrically operated transmitter 17. The battery or other source of electricity has a connection 18 with one of the leads 14. A receiver 19 of the ordinary type has a connection with the same lead 14, to which the battery is connected, and both the receiver and transmitter have connections 21 with the contacts of a switch 22. This switch has suitable means, as for instance, a spring 23, which normally maintains the receiver in circuit with the coil 11, as will be evident by reference to Fig. 1, but if the switch is thrown to break the circuit, it will then cut in the source of electrical energy 15 and the transmitter 17.

An outfit similar to the above, is located on the vehicle or boat 9, but the coil 24 thereof, shown in Fig. 2, is much smaller. As further illustrated in said figure, the mechanism mounted on the boat, consists of a transmitter 25, and a battery or other source of electrical energy 26 electrically connected, as shown at 27 and having a connection 28 with one of the leads of the coil. The receiver 29 also has a connection 30 with said lead. A switch 31 is connected to the other lead, and is normally held in a position by a spring 32 to maintain a closed circuit through the receiver 29 and the coil, though it may be moved to cut out said receiver and close the circuit through the coil, the source of electrical energy and the transmitter.

In this system, if it is desired to transmit from one station, as for instance, the shore-station, the switch 22 is moved downwardly to cut out the receiver and throw in the transmitter and source of electrical energy, while the operator upon the boat or vehicle leaving the mechanism in the condition shown in Fig. 2, holds the receiver 29 to his ear. If therefore the operator at the shore-station uses the transmitter in the ordinary manner, a varying current corresponding to that passing through the coil of great magnitude 11, will be induced in the coil 24, and the speech or other sounds will thus be transmitted to the operator on the boat. By reversing the arrangement, speech may be transmitted from the boat to the shore station.

The use of coils for both stations, each coil consisting of a plurality of convolutions has been found by experience to be of the utmost value, and furthermore experience has demonstrated that the employment of coils of different magnitudes is of great importance, for it has been found that while two small coils can be used to transmit but a short distance, if one large coil of the character set forth is employed, the other may be very small, and speech or sounds can be transmitted comparatively great distances from one to the other. These sounds are clearly audible.

The structure disclosed in Figs. 4 and 5 is of the same general character. A road-way 32 is disclosed surrounded by a coil 33 of great magnitude that is supported on suitable poles 34. The way-station 35 consists of a transmitter 36, a source of electrical energy 37 connected thereto, a receiver 38, and a switch 39, whereby the receiver or the transmitter and source of electrical energy can be thrown into circuit with the coil 33. The vehicles 40, which operate on the roadway, are provided with smaller coils 41 and instruments consisting of receivers 42, transmitters 43, sources of electrical energy 44 and switches 45 all arranged in the manner already described. In a system of this kind, it will be evident that the occupant of one vehicle may telephone to the home or way-station, and the message can be transmitted to another vehicle. Thus it will be evident that communication can be established between two moving vehicles or between a way-station and any vehicle desired which is within the range of the home- or way-station. The system is also capable of use in connection with railways, and in Figs. 6 and 7, such a system is disclosed in connection therewith. A comparatively great coil 46 is supported on opposite sides of the railway 47 by poles 48 and a station 49 has a receiver 50 and a transmitter 51, a source of electrical energy 52 and a switch 53, the last mentioned being employed for throwing either the receiver or the transmitter and source of electrical energy into closed circuit with the coil 46. One or more cars of a railway train is equipped with an outfit consisting of a coil 54, a receiver 55, a transmitter 56, a source of electrical energy 57, and a switch 58 for throwing either the receiver or the transmitter and source of electrical energy into circuit with the coil 54. It will be evident that the operation of these two last described systems are substantially the same as that first set forth, and no extended description thereof is believed to be necessary.

From the foregoing, it is thought that the construction, operation, and many advantages of the herein described invention will be apparent to those skilled in the art, without further description, and it will be understood that various changes in the size, shape, proportion, and minor details of construction, may be resorted to without departing from the spirit or sacrificing any of the advantages of the invention.

Having thus fully described my invention, what I claim as new, and desire to secure by Letters Patent, is:—

1. In a system of the character described, the combination with a vehicle, of a comparatively small coil of conducting material mounted thereon, electrical transmitting and receiving mechanism including a source of electrical energy connected to the small coil and carried by the vehicle, a stationary aerial coil of much greater magnitude than the small coil having its opposite stretches or sides extending along the opposite sides of the path of travel of the vehicle and elevated above the same and above the vehicle coil, and electrical transmitting and receiving mechanism connected to the greater coil and including a source of heavy electrical current.

2. In a system of the character described, the combination with a vehicle, of a coil of conducting material mounted thereon, electrical transmitting mechanism, a source of electrical energy connected thereto, receiving mechanism, means for connecting either the transmitting mechanism and source of electrical energy or the receiving mechanism to the coil, a stationary coil of greater magnitude surrounding the path of travel of the vehicle and comprising a plurality of convolutions of conducting material, the different convolutions being insulated one from the other, means for supporting the coil in an elevated position, electrical transmitting mechanism, a source of great electrical energy connected to said transmitting mechanism, electrical receiving mechanism, and means for electrically connecting either the transmitting mechanism and source of electrical energy or the receiving mechanism to the said coil of greater magnitude.

3. Means for communicating between a plurality of stations which consists of an aerial electrical coil of great magnitude, means for supporting the said coil, a station electrically connected to the great coil and comprising transmitting and receiving mechanism that includes a source of heavy electrical energy, and a plurality of other separate stations simultaneously in coacting relation with the aerial coil, each of said latter stations comprising a coil of conducting material spaced from but in coacting relation with said great coil and below the same, and transmitting and receiving mechanism connected to said other coil and including a source of electrical energy.

4. Means for communicating between a plurality of stations which consists of an aerial coil of conducting material of great magnitude, transmitting and receiving mechanism connected to said aerial coil and including a source of heavy electrical energy, a plurality of vehicles movable between the opposite sides or stretches of the great coil, coils carried by said vehicles and disposed within the field of action of the aerial coil, and transmitting and receiving mechanism mounted on each vehicle and including a source of electrical energy.

In testimony, that I claim the foregoing as my own, I have hereto affixed my signature in the presence of two witnesses.

NATHAN B. STUBBLEFIELD.

Witnesses:
J. P. McElrath,
J. H. Coleman.

(Back) (Next)

The invention that started it all for Armstrong, the 1912 feed-back (regenerative) receiver. This equipment was donated to the Smithsonian Institution following Armstrong's death.

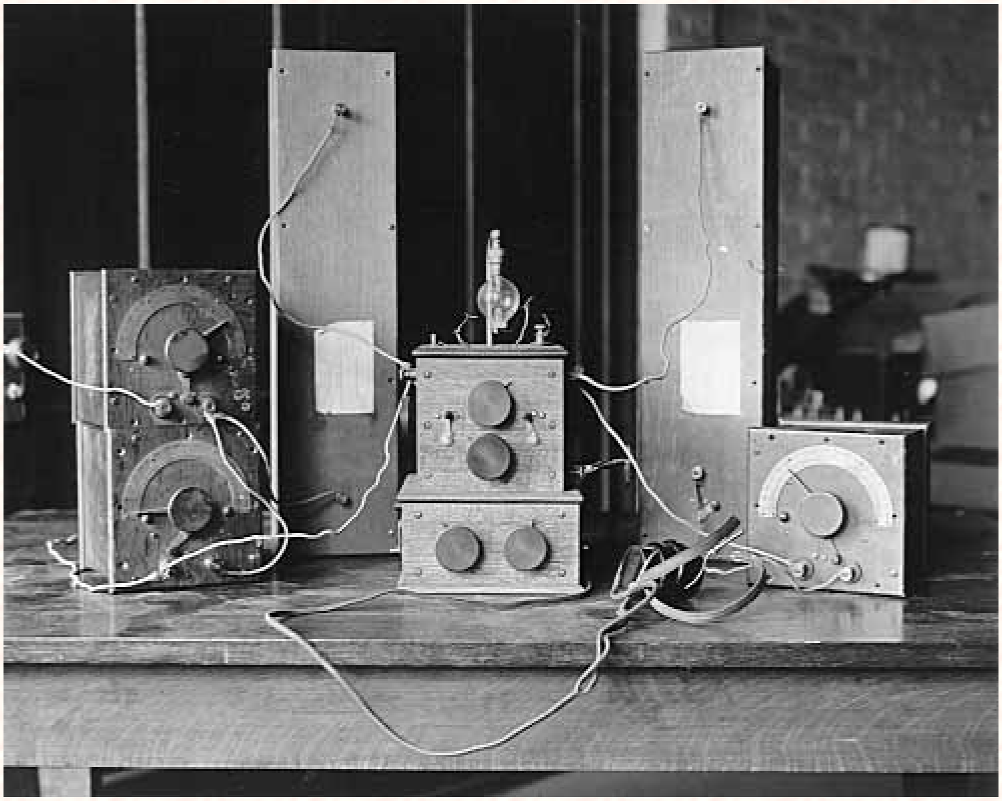

Drawings from Armstrong patent 1,113,149, the famous regenerative, or feed-back, patent.

E. H. ARMSTRONG.
WIRELESS RECEIVING SYSTEM.
APPLICATION FILED OCT. 29, 1913.

1,113,149.

Patented Oct. 6, 1914.

3 SHEETS—SHEET 1.

Fig. 1,

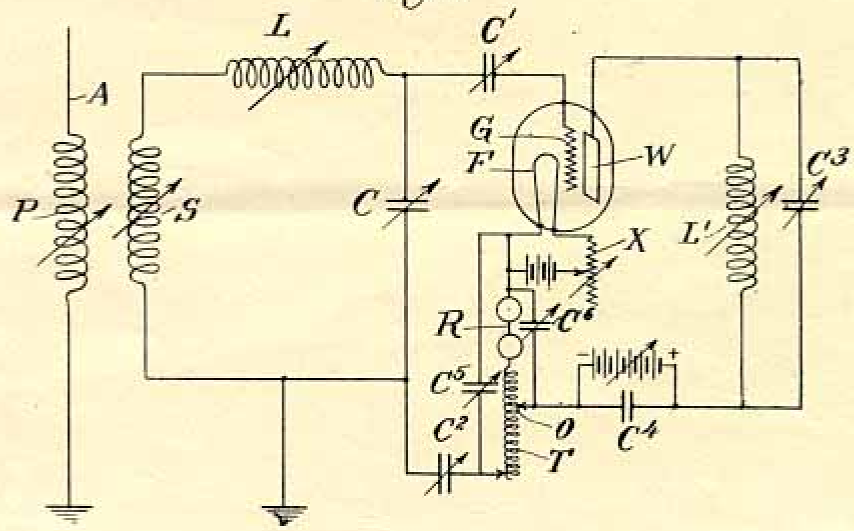
A
P
S
L
C'
G
F
W
C
X
L'
C3
R
C6
C5
C2
O
C4
T

Fig. 2,

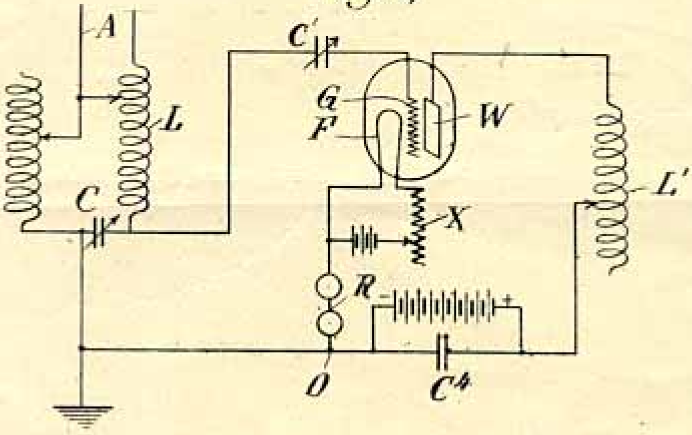
A
L
C'
G
F
W
C
X
L'
R
O
C4

WITNESSES
L. Bates.
C. B. Schroeder.

INVENTOR
Edwin H. Armstrong
BY
Pennie, Davis & Goldsborough
ATTORNEYS

E. H. ARMSTRONG.
WIRELESS RECEIVING SYSTEM.
APPLICATION FILED OCT. 29, 1913.

1,113,149.

Patented Oct. 6, 1914.
3 SHEETS—SHEET 2.

Fig. 3,

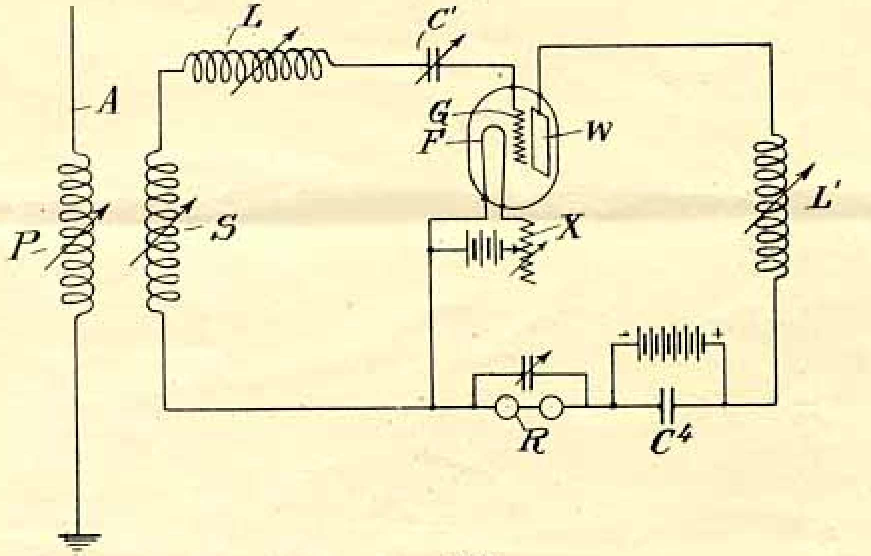

Fig. 4,

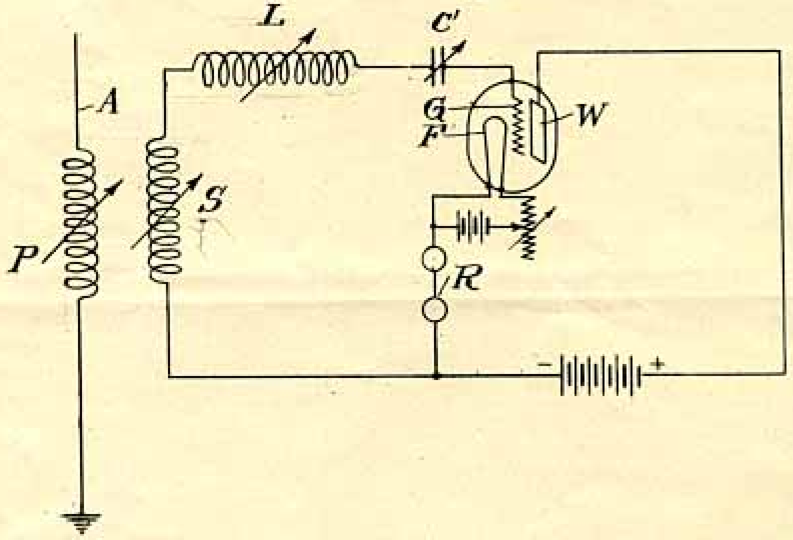

WITNESSES
L. Bates
C. B. Schneider

INVENTOR
Edwin H. Armstrong
BY
Pennie, Davis & Goldsborough
ATTORNEYS

E. H. ARMSTRONG.
WIRELESS RECEIVING SYSTEM.
APPLICATION FILED OCT. 29, 1913.

1,113,149.

Patented Oct. 6, 1914.
3 SHEETS—SHEET 3.

Fig. 5,

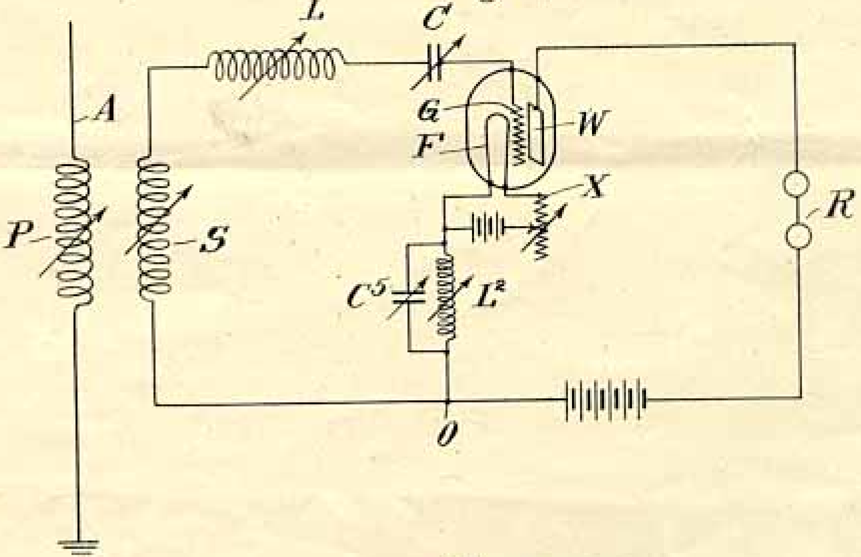
L
C
A
G
F
W
X
R
P
S
C^5
L^2
0

Fig. 6

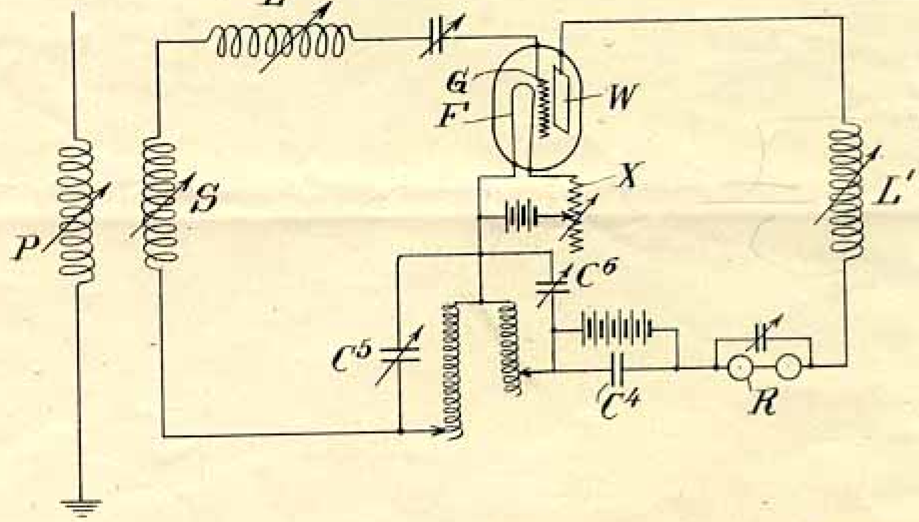
L
G
F
W
X
S
P
L'
C^6
C^5
C^4
R

WITNESSES
L. Bates
C. B. Schroeder

INVENTOR
Edwin H. Armstrong
BY
Pennie, Davis & Goldsborough
ATTORNEYS

(More)

E. H. ARMSTRONG.

WIRELESS RECEIVING SYSTEM.

APPLICATION FILED OCT. 29, 1913.

1,113,149.

Patented Oct. 6, 1914.

3 SHEETS—SHEET 1.

Fig. 1,

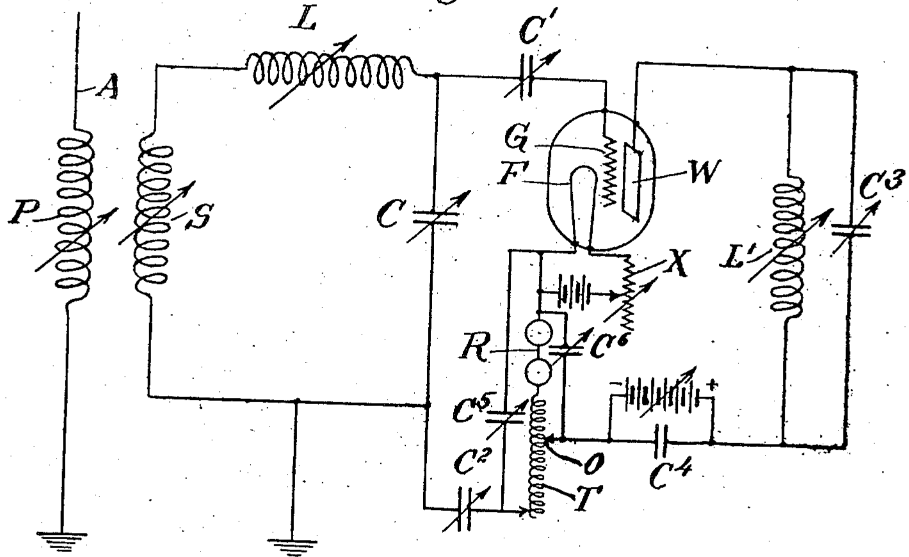
L
C'
A
G
F
W
P
S
C
C^3
X
L'
R
C^6
C^5
C^2
O
C^4
T

Fig. 2,

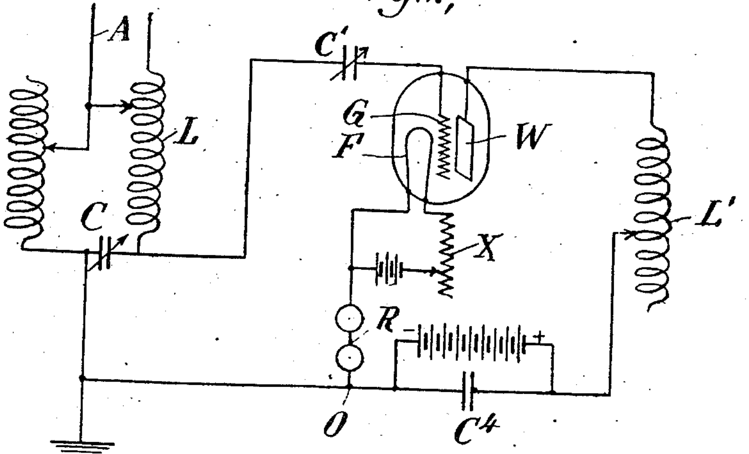
A
C'
L
G
F
W
C
X
L'
R
O
C^4

WITNESSES

L. Bates

C. B. Schroeder

Edwin H. Armstrong INVENTOR

BY

Pennie, Davis, Goldsborough

ATTORNEYS

E. H. ARMSTRONG.
WIRELESS RECEIVING SYSTEM.
APPLICATION FILED OCT. 29, 1913.

1,113,149.

Patented Oct. 6, 1914.
3 SHEETS—SHEET 2.

Fig. 3,

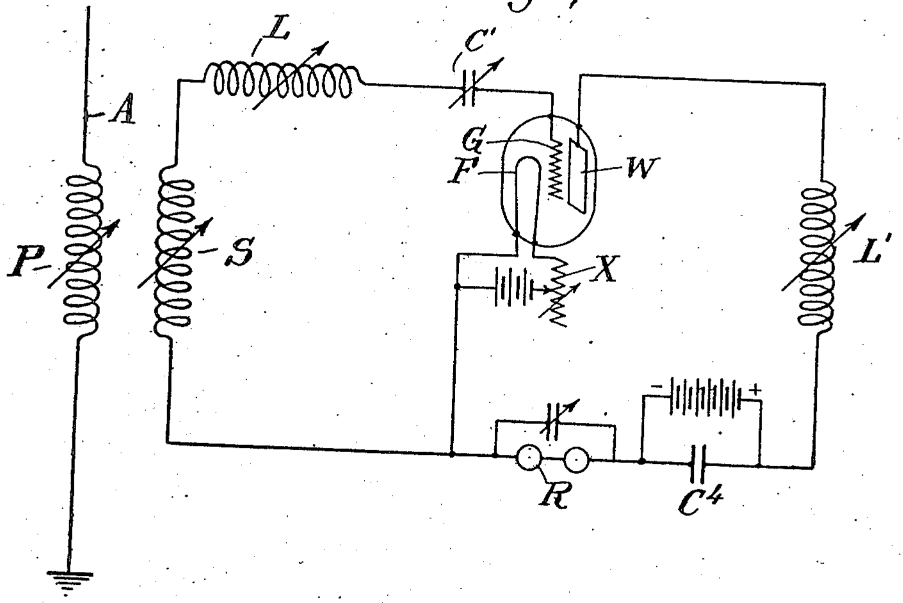

L
C'
A
G
F
W
P
S
X
L'
−
+
R
C4

Fig. 4,

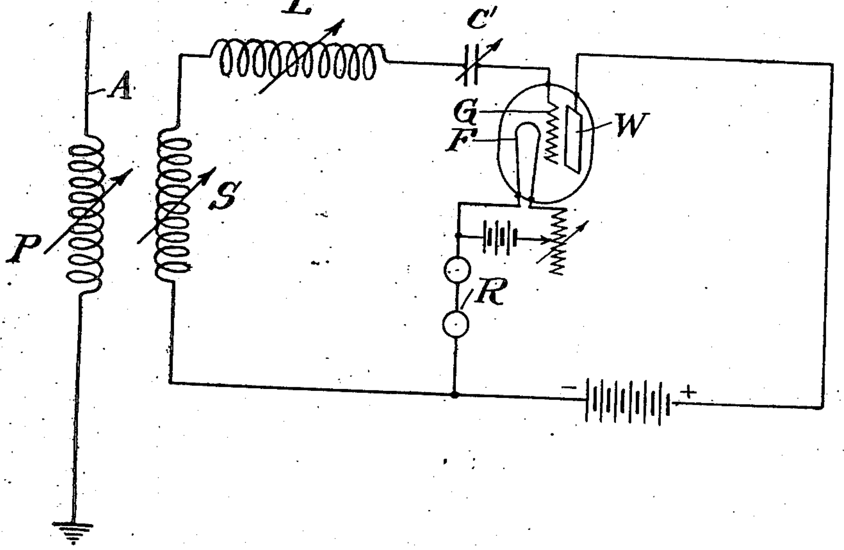

WITNESSES
L. Bates
C. B. Schroeder

INVENTOR
Edwin H. Armstrong
BY
Pennie, Davis & Goldsborough
ATTORNEYS

E. H. ARMSTRONG.
WIRELESS RECEIVING SYSTEM.
APPLICATION FILED OCT. 29, 1913.

1,113,149.

Patented Oct. 6, 1914.
3 SHEETS—SHEET 3.

Fig. 5,

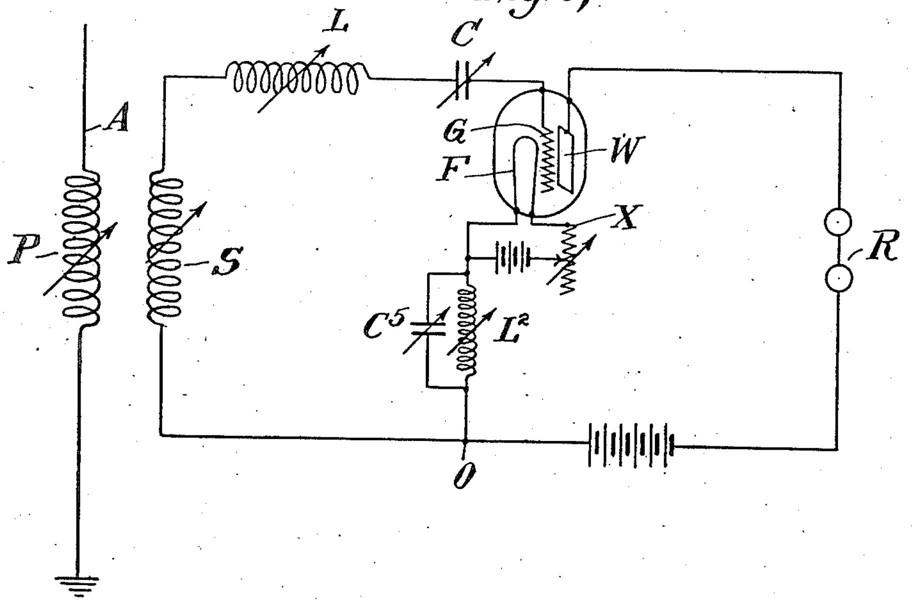
L
C
A
G
W
F
X
P
S
R
C^5
L^2
O

Fig. 6

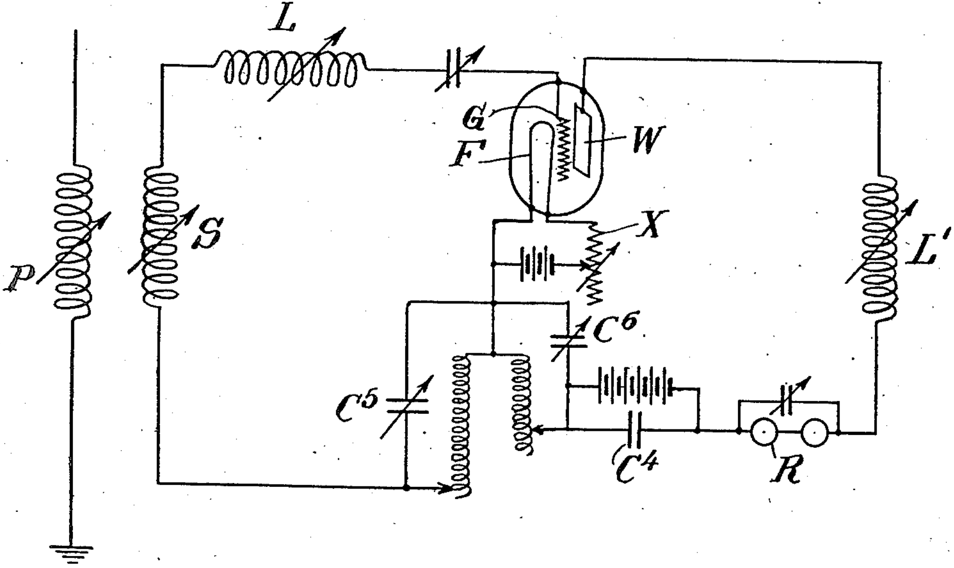
L
G
W
F
X
S
L'
P
C^6
C^5
C^4
R

WITNESSES
L. Bates
C. B. Schroeder

INVENTOR
Edwin H. Armstrong
BY
Pennie, Davis & Goldsborough
ATTORNEYS

UNITED STATES PATENT OFFICE.

EDWIN H. ARMSTRONG, OF YONKERS, NEW YORK.

WIRELESS RECEIVING SYSTEM.

1,113,149. Specification of Letters Patent. Patented Oct. 6, 1914.

Application filed October 29, 1913. Serial No. 797,947.

To all whom it may concern:

Be it known that I, Edwin H. Armstrong, a citizen of the United States, residing at 1032 Warburton avenue, Yonkers, county of Westchester, State of New York, have invented certain new and useful Improvements in Wireless Receiving Systems; and I do hereby declare the following to be a full, clear, and exact description of the invention, such as will enable others skilled in the art to which it appertains to make and use the same.

The present invention relates to improvements in the arrangement and connections of electrical apparatus at the receiving station of a wireless system, and particularly a system of this kind in which a so-called "audion" is used as the Hertzian wave detector; the object being to amplify the effect of the received waves upon the current in the telephone or other receiving circuit, to increase the loudness and definition of the sounds in the telephone or other receiver, whereby more reliable communication may be established, or a greater distance of transmission becomes possible. To this end I have modified and improved upon the arrangement of the receiving circuits in a manner which will appear fully from the following description taken in connection with the accompanying drawings. As a preliminary, it is to be noted that my improved arrangement corresponds with the ordinary arrangement of circuits in connection with an audion detector to the extent that it comprises two interlinked circuits; a tuned receiving circuit in which the audion grid is included, and which will be hereinafter referred to as the "tuned grid circuit", and a circuit including a battery or other source of direct current and the "wing" of the audion, and which will be hereinafter referred to as the "wing circuit". As is usual, the two circuits are interlinked by connecting the hot filament of the audion to the point of junction of the tuned grid circuit and the wing circuit. I depart, however, from the customary arrangement of these circuits in a manner which may, for convenience of description, be classified by analysis under three heads; firstly, the provision of means, or the arrangement of the apparatus, to impart resonance to the wing circuit so that it is capable of sustaining oscillations corresponding to the oscillations in the tuned grid circuit; secondly, the provision of means supplementing the electrostatic coupling of the audion to facilitate the transfer of energy from the wing circuit to the grid circuit, thereby reinforcing the high frequency oscillations in the grid circuit, and thirdly, the introduction into the wing circuit of an inductance through which the direct current of the wing circuit flows, and which is so related to the grid circuit that the maintaining electromotive-force across the terminals of the inductance due to reduction of the direct current, is effective in the tuned grid circuit to increase the grid charge and consequently to further reduce the current in the wing circuit and in the telephones. By a further extension of this idea, the effect of the maintaining electromotive-force upon the grid current may be augmented by the use of a transformer in a manner which will be understood from the following description.

Figure 1 illustrates the arrangement and connection of apparatus with which I have thus far obtained the best results and which embodies in combination the several features of improvement which I have invented or discovered. Fig. 2 represents a like arrangement with the exception that there is no transformer for augmenting the effect of the maintaining electromotive-force, and several condensers, which may advantageously be employed but which are not essential, are eliminated. Fig. 3 illustrates an arrangement in which the advantages of my invention are only partially present, the inductance which produces the maintaining electromotive-force effective on the tuned grid circuit being eliminated. Fig. 4 illustrates an arrangement in which inductance, in this case the inductance of the telephones, is employed for producing the maintaining electromotive force effective on the tuned grid circuit, but the wing circuit is not resonant. Fig. 5 illustrates an arrangement in which an inductance replaces the telephones in that portion of the connections which is common to the two circuits, and the telephones are put in the wing circuit, and Fig. 6 illustrates an arrangement in which a double winding transformer is used, the primary being located in the wing circuit.

Referring particularly to Fig. 1, A represents the ordinary grounded aerial connected to the primary P of an oscillation transformer, the secondary S of which is connected as usual in the tuned grid circuit, this circuit also including the inductance L and preferably a shunted capacity C as is usual. Between the inductance L and the audion grid G is located a condenser C′ adapted to receive and hold the charge which accumulates on the grid as a result of the received oscillations. The grid lies within the audion in the path of current in the wing circuit and the grid circuit is connected to the wing circuit at the junction point O, so that the two circuits are interlinked from the grid to that point. Between the junction point O and the filament of the audion I insert the telephone receivers R and the primary of the auto-transformer T. From junction point O the tuned grid circuit is completed through the secondary of the auto-transformer and condenser C^2 back to the secondary S; and the condenser C^5 is connected as shown in shunt to the telephones and the secondary of the auto-transformer. The wing circuit may be traced from the positive terminal of the battery through the inductance L′ to the wing W and through the connections which are common to the two circuits, including the current path through the audion to the filament F, the telephone receivers R, and the primary of the auto-transformer and back to the negative terminal of the battery. The inductance L′ is preferably shunted by a condenser C^3 and a condenser C^4 is placed across the battery terminals to afford a path of low impedance for the high frequency oscillation. The telephones R and the auto-transformer T are shunted by a condenser C^5, which affords a path for the high frequency oscillations in the grid circuit; and the telephones R and the primary of the auto-transformer are shunted by a capacity C^6, which affords a path for the high frequency oscillations in the wing circuit, and I find that the audion is made more stable and shows less tendency to become a high frequency generator and to set up oscillations in the interlinked circuits, if the tuned grid circuit is grounded as indicated. Each of the pieces of apparatus crossed with an arrow on the drawing is continuously variable, that is, may be varied by infinitesimal increments, and the condenser C^4 is preferably made adjustable by steps. I find that with such an arrangement of apparatus, and by properly adjusting the reactances, signals which are scarcely audible with the ordinary audion connection can be amplified to a point where they are too strong for, and "paralyze" the most stable audions that I have been able to obtain. The present understanding and modes of explanation of the phenomena which present themselves in such an arrangement is such that any theory of operation which may be advanced in regard to them is merely an attempt to explain the results attained in language which will be understandable to those skilled in the art; and it is with this idea in mind that the following description is written.

Upon reference to Fig. 2 of the drawings, which is an attempt to simplify the illustration of the essential elements of my invention, it will be observed that the auto-transformer T, and the condensers C^2, C^3, C^5, and C^6, are omitted. In this case the capacity of the telephone cords is sufficient to by-pass the high frequency oscillations.

The effect of tuning the wing circuit to the received high frequency oscillations by such an arrangement of apparatus as is illustrated in Fig. 3, may be explained as follows: In the ordinary audion connection, the wing W will be constantly maintained at the same potential with respect to the filament F and will constitute a surface of positive potential within the audion and having a constant tendency to absorb electrons. When the wing circuit is made resonant in accordance with my invention, the initial received oscillations in the grid circuit set up corresponding oscillations in the wing circuit so that a negative charge on the grid is accompanied by an increase of the potential of the wing, and a positive charge on the grid is accompanied by a decrease of the potential of the wing. The absorption of electrons by the grid and the building up of the charge in the condenser C′ occurs only when the potential of the grid is positive with respect to the potential of the filament, and during these periods the potential of the wing with respect to the filament is reduced, thereby reducing its absorption of electrons and increasing the absorption of electrons by the grid. The charge thus built up in the condenser C′ by absorption at the grid is entrapped and is an effective counter-electromotive-force in the wing circuit to reduce the current in the telephones. It is understood that on each successive wave train this effect is repeated, it being probable, according to my observations, that the full condenser charge is not built up by the first wave of a train but is gradually built up during the first portion of the wave train; and because of the high resistance of the discharge path of the condenser C′, this charge outlasts the duration of the wave train.

Referring now to Fig. 4, we may examine the effect of an inductance inserted in the wing circuit and related to the grid circuit as above defined. The reduction of current through the telephones, due to the received oscillations, produces a maintaining electromotive-force across the terminals of the telephones R which is in the same direction as the battery electromotive-force, and, because of the fact that the telephones are in that portion of the connections which is common to the two circuits, this maintaining electromotive-force is effective upon the tuned grid circuit to charge the condenser C′ in the same sense that the flow of electrons from the filament to the grid would charge the condenser. I have found, as a practical matter, that the arrangement of apparatus shown in Fig. 4 is effective to materially increase the loudness of the signals, but obviously better effects are obtained by combining the arrangements of Figs. 3 and 4 to make up the arrangement of Fig. 2. In this figure an auto-transformer L. is used instead of the two-coil transformer P. S., and in Figs. 2 and 4 the capacity of the telephone cords affords a path of low impedance for the passage of the high frequency oscillations about the inductance of the telephones. It will be observed, that this capacity of the telephone cords in Figs. 2 and 4, and likewise the capacity C^5 in Fig. 5 and the distributed capacity of the transformers in Figs. 1 and 6, is in each case common to the grid circuit and the wing circuit and constitutes an electrostatic coupling facilitating the transfer of energy from the wing circuit to the grid circuit and increasing the effect upon the grid of high frequency pulsations in the wing circuit. This effect occurs whether the wing circuit is tuned or not, but obviously the transfer of energy is increased by tuning the two circuits alike. I have discovered, however, that the beneficial effect may be still further increased by the interposition of a transformer in such a way as to increase the effect of the maintaining electromotive-force due to the reduction of current through the telephone receivers, and such a transformer is shown at T in Fig. 1. With the foregoing discussion in mind the effect of this transformer will at once be apparent. The inductance of the primary of the transformer is of course added to and increases the effect of the inductance of the telephones, but beyond this the maintaining electromotive-force across the primary is transformed into a higher electromotive-force in the tuned grid circuit. I have found that a ratio of transformation of 2 to 1 is sufficient in most cases when the inductance of the primary is equal to the inductance of the telephones, since a greater ratio causes the audion to become a high frequency generator setting up disturbing oscillations in the grid and wing circuits. The condenser C^5 by-passes the high frequency oscillations in the grid circuit about the telephones and the auto-transformer T, and the condenser C^6 by-passes the oscillations in the wing circuit about the telephones and the primary of the auto-transformer. The capacity of the telephone cords may, however, be relied upon to by-pass the oscillations about the telephones and condensers C^5 and C^6 utilized for by-passing the oscillations about the transformer only. The manner in which the different reactances will be adjusted so as to tune the two circuits to one another will be at once understood by those skilled in the art, and it will also be understood that the adjustable resistance X may be varied to get the proper temperature of the filament. Furthermore, it will be understood from what has been said that the ratio of transformation of the transformer should be adjusted to get the maximum signals without causing the audion to generate oscillations. The condenser C^5 has the additional effect, as I have found, to steady the audion and enable it to withstand greater variations of potential of the grid G and wing W without becoming a generator of disturbing oscillations in the grid and wing circuits. Particularly when working with long waves it is advisable to use this condenser in every case. I also find that the introduction of a condenser C^2 in the place indicated in Fig. 1, that is, between the auto-transformer and the secondary S, has the effect to prevent buzzing or humming in the telephones.

From what has been said, it will be understood that the inductance of the telephones may be utilized as the generator of the maintaining electromotive-force by inserting the telephones in the connections common to the two circuits, as shown in Figs. 1, 2 and 4; but a like result will be obtained by locating the telephones in the wing circuit as indicated in Fig. 5, and placing a suitable inductance as L^2 in that portion of the connections which is common to the two circuits. Indeed I find that the beneficial effect of an inductance so located can be obtained, as would be expected, by utilizing the primary of the transformer alone, as indicated in Fig. 6. In this case a two coil transformer is shown in place of an auto-transformer and the variations of current in the primary induce potential variations in the secondary which may be given any value by properly selecting the ratio of transformation.

Inasmuch as the purpose of the transformer is to magnify the effect of the maintaining electromotive-force in the inductance, it will be obvious that by increasing this inductance, the transformer may be rendered unnecessary, and some simplification of the whole arrangement may be attained by using telephone receivers of extremely high inductance and locating them in the position shown in Fig. 2, without any transformer.

The purpose of the condenser C^3 in Fig. 1 is to enable the wing circuit to be made resonant with the grid circuit when long waves are to be received. The capacity of the audion and the distributed capacity of the inductance L′ are such that up to 2000 meters wave length resonance can be obtained in the wing circuit with reasonable values of inductance. For waves longer than 2000 meters resonance can be obtained in the wing circuit without the parallel condenser, but only by unduly increasing L′.

For the purpose of completing the disclosure of my invention, and to facilitate its practical application by those skilled in the art, I give below the values of the constants which I have found to work well for a wave length of 1800 meters, using an antenna having a capacity of .0012 microfarads and a standard McCandless audion. The inductance of the primary P was 760 microhenries and the combined inductance of the secondary S and the coil L was 3660 microhenries. The capacity of C was about .00025 microfarads, the capacity of C^2, .01 microfarads, and of C^4, 2 microfarads. The values of C′ and L′ are largely dependent upon one another and upon the amplitude of the received oscillations as well as the characteristics of the audion. These are the two elements which are chiefly relied upon by the operator to bring the system into condition to give the best results in receiving the signals from any particular station. The value of L′ will be in the neighborhood of 10,000 microhenries, and of C′ between .00003 and .0001 microfarads. If the inductance L′ is not used, the value of C′ can be much larger. I have usually employed about .003 microfarads as a matter of convenience; but there does not seem to be any particular advantage in increasing the capacity of C′ beyond .0005 microfarads. The batteries, and the resistance X, are as used in the standard audion set. The way to tune this set is to cut out L′, set C′ at .00005 microfarads, and then adjust P, S, L and C as in the ordinary audion set until signals are strongest. The inductance L′ is then gradually cut in and the strength of the signals will increase many times until a point is reached where the signals lose distinctness and there is a loud hiss in the telephones. This indicates that the audion is generating high frequency oscillations in the grid and wing circuits, and the inductance L′ should be set at a point just below that at which this occurs. If there is no hiss as L′ is increased, and the signals pass through a maximum of strength and begin to fall off, then L′ should be set at the point of maximum strength of signals and C′ should be increased to a point just below that at which the hiss appears. Under these circumstances, the increase of C′ will be accompanied by an increase of the strength of the signals, the maximum strength being obtained in each case just below that point at which the audion begins to act as a generator of high frequency oscillations.

I find that the capacity required to bypass the oscillations about the inductance which is common to the two circuits, and about the transformer, if the transformer is used, is relatively small; and when the telephone receivers are used as the inductance at this point, the capacity of the telephone cords is sufficient. With greater wave lengths, however, when it becomes necessary to increase the tuning inductances L and L′, the audion begins to be unstable, and I find that its stability can be restored by the use of a considerably larger capacity C^5. For example, with wave lengths of 4000 to 7500 meters I use a capacity C^5 of from .001 to .003 microfarads.

Having thus described my invention, what I claim is:

1. An audion wireless receiving system having a resonant wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, the resonant grid circuit having a capacity so related to the grid as to receive and retain the charge which accumulates thereon.

2. An audion wireless receiving system having a resonant wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, the resonant grid circuit having a capacity so related to the grid as to receive and retain the charge which accumulates thereon, and an inductance through which the current in the wing circuit flows, the grid circuit including connections for making effective upon that circuit the potential variations resulting from a change of current in the wing circuit.

3. An audion wireless receiving system having a resonant wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, the resonant grid circuit having a capacity so related to the grid as to receive and retain the charge which accumulates thereon, and an inductance in that portion of the connections which is common to the two circuits.

4. An audion wireless receiving system having a resonant wing circuit interlinked with a resonant grid circuit upon which the

received oscillations are impressed, the resonant grid circuit having a capacity so related to the grid as to receive and retain the charge which accumulates thereon, an inductance in that portion of the connections which is common to the two circuits, and a transformer having its primary in the wing circuit and its secondary in the grid circuit.

5. An audion wireless receiving system having a resonant wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, the resonant grid circuit having a capacity so related to the grid as to receive and retain the charge which accumulates thereon, and a telephone receiver in that portion of the connections which is common to the two circuits.

6. An audion wireless receiving system having a resonant wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, the resonant grid circuit having a capacity so related to the grid as to receive and retain the charge which accumulates thereon, a telephone receiver in that portion of the connections which is common to the two circuits, and a transformer having its primary in the wing circuit and its secondary in the grid circuit.

7. An audion wireless receiving system having a resonant wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, the resonant grid circuit having a capacity so related to the grid as to receive and retain the charge which accumulates thereon, an inductance in that portion of the connections which is common to the two circuits a transformer having its primary in the wing circuit and its secondary in the grid circuit, and condensers affording paths of low impedance about the inductance and transformer for the high frequency oscillations in both circuits.

8. An audion wireless receiving system having a wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, and an inductance in that portion of the connections which is common to the two circuits.

9. An audion wireless receiving system having a wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, and an inductance through which the current in the wing circuit flows, the grid circuit including connections for making effective upon that circuit the potential variations resulting from a change of current in the wing circuit.

10. An audion wireless receiving system having a wing circuit interlinked with a resonant grid circuit, an inductance in that portion of the connections which is common to the two circuits, and a transformer having its primary in the wing circuit and its secondary in the grid circuit.

11. An audion wireless receiving system having a wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, an inductance in that portion of the connections which is common to the two circuits, a transformer having its primary in the wing circuit and its secondary in the grid circuit, and condensers affording paths of low impedance about the inductance and transformer for the high frequency oscillations in both circuits.

12. An audion wireless receiving system having a wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, and a high inductance telephone receiver in that portion of the connections which is common to the two circuits.

13. An audion wireless receiving system having a wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, a capacity between that portion of the grid circuit upon which the oscillations are impressed and the grid, and a second capacity between the point of junction of the grid circuit with the wing circuit and that portion of the grid circuit upon which the oscillations are impressed.

14. An audion wireless receiving system having a wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, an inductance in that portion of the connections which is common to the two circuits, and a stabilizing capacity shunted about said inductance.

15. An audion wireless receiving system having a wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, and means supplementing the coupling of the audion to facilitate transfer of energy from the wing circuit to the grid circuit, whereby the effect upon the grid of high frequency pulsations in the wing circuit is increased.

16. An audion wireless receiving system having a resonant wing circuit interlinked with a resonant grid circuit upon which the received oscillations are impressed, and means supplementing the coupling of the audion to facilitate transfer of energy from the wing circuit to the grid circuit, whereby the effect upon the grid of high frequency pulsations in the wing circuit is increased.

17. An audion wireless receiving system having a wing circuit interlinked with a resonant grid circuit upon which received

oscillations are impressed and an electrostatic coupling between the circuits supplementing the coupling of the audion to facilitate transfer of energy from the wing circuit to the grid circuit, whereby the effect upon the grid of high frequency pulsations in the wing circuit is increased.

18. An audion wireless receiving system having a resonant wing circuit interlinked with a resonant grid circuit upon which received oscillations are impressed and an electrostatic coupling between the circuits supplementing the coupling of the audion to facilitate transfer of energy from the wing circuit to the grid circuit, whereby the effect upon the grid of high frequency pulsations in the wing circuit is increased.

In testimony whereof I affix my signature, in presence of two witnesses.

EDWIN H. ARMSTRONG.

Witnesses:
WILLIAM H. DAVIS,
JOHN C. PENNIE.

N. TESLA.

APPARATUS FOR TRANSMITTING ELECTRICAL ENERGY.

APPLICATION FILED JAN. 18, 1902. RENEWED MAY 4, 1907.

1,119,732. Patented Dec. 1, 1914.

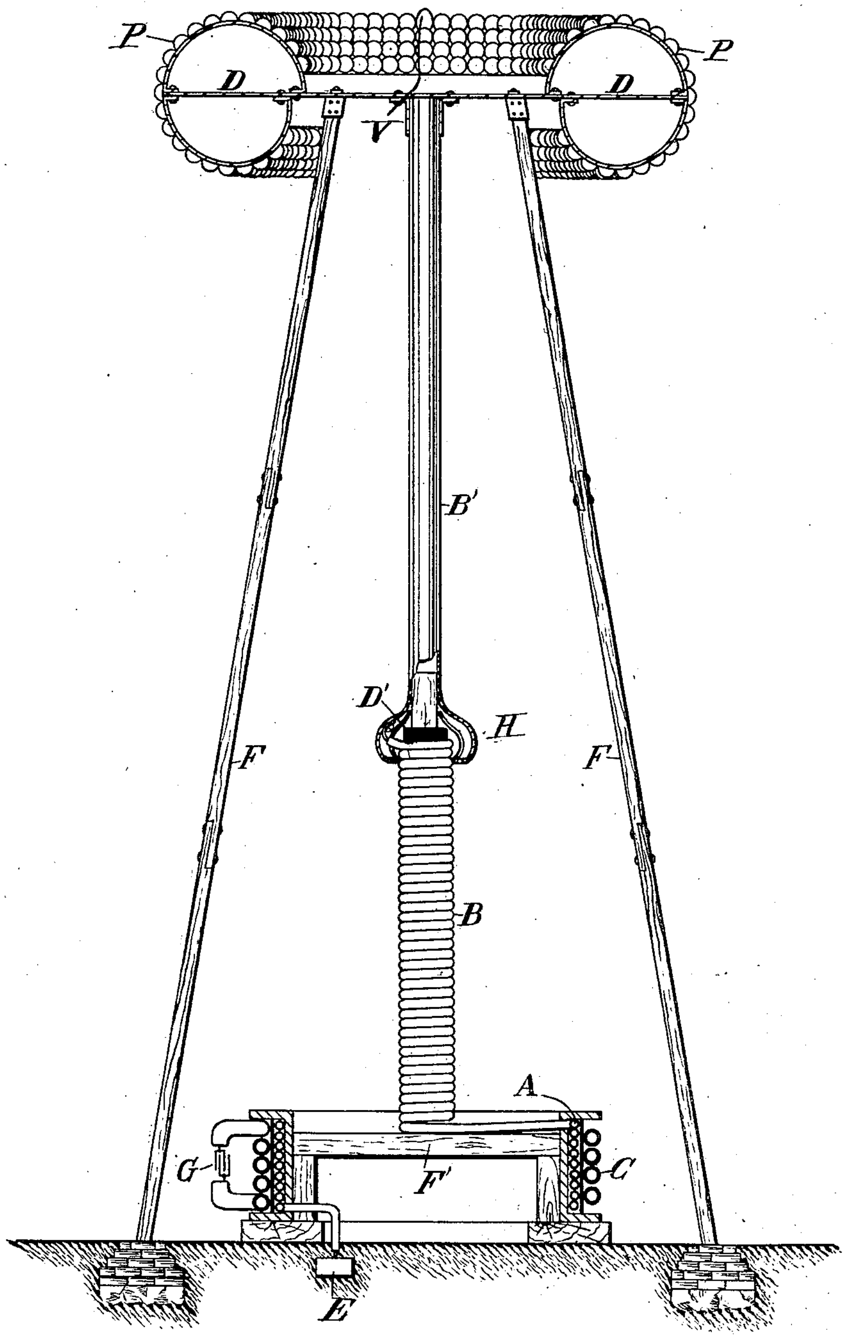

WITNESSES:

M. Lawson Dyer

Benjamin Miller.

Nikola Tesla, INVENTOR,

BY Kerr, Page & Cooper,

his ATTORNEYS.

UNITED STATES PATENT OFFICE.

NIKOLA TESLA, OF NEW YORK, N. Y.

APPARATUS FOR TRANSMITTING ELECTRICAL ENERGY.

1,119,732. Specification of Letters Patent. **Patented Dec. 1, 1914.**

Application filed January 18, 1902, Serial No. 90,245. Renewed May 4, 1907. Serial No. 371,817.

To all whom it may concern:

Be it known that I, Nikola Tesla, a citizen of the United States, residing in the borough of Manhattan, in the city, county, and State of New York, have invented certain new and useful Improvements in Apparatus for Transmitting Electrical Energy, of which the following is a specification, reference being had to the drawing accompanying and forming a part of the same.

In endeavoring to adapt currents or discharges of very high tension to various valuable uses, as the distribution of energy through wires from central plants to distant places of consumption, or the transmission of powerful disturbances to great distances, through the natural or non-artificial media, I have encountered difficulties in confining considerable amounts of electricity to the conductors and preventing its leakage over their supports, or its escape into the ambient air, which always takes place when the electric surface density reaches a certain value.

The intensity of the effect of a transmitting circuit with a free or elevated terminal is proportionate to the quantity of electricity displaced, which is determined by the product of the capacity of the circuit, the pressure, and the frequency of the currents employed. To produce an electrical movement of the required magnitude it is desirable to charge the terminal as highly as possible, for while a great quantity of electricity may also be displaced by a large capacity charged to low pressure, there are disadvantages met with in many cases when the former is made too large. The chief of these are due to the fact that an increase of the capacity entails a lowering of the frequency of the impulses or discharges and a diminution of the energy of vibration. This will be understood when it is borne in mind, that a circuit with a large capacity behaves as a slackspring, whereas one with a small capacity acts like a stiff spring, vibrating more vigorously. Therefore, in order to attain the highest possible frequency, which for certain purposes is advantageous and, apart from that, to develop the greatest energy in such a transmitting circuit, I employ a terminal of relatively small capacity, which I charge to as high a pressure as practicable. To accomplish this result I have found it imperative to so construct the elevated conductor, that its outer surface, on which the electrical charge chiefly accumulates, has itself a large radius of curvature, or is composed of separate elements which, irrespective of their own radius of curvature, are arranged in close proximity to each other and so, that the outside ideal surface enveloping them is of a large radius. Evidently, the smaller the radius of curvature the greater, for a given electric displacement, will be the surface-density and, consequently, the lower the limiting pressure to which the terminal may be charged without electricity escaping into the air. Such a terminal I secure to an insulating support entering more or less into its interior, and I likewise connect the circuit to it inside or, generally, at points where the electric density is small. This plan of constructing and supporting a highly charged conductor I have found to be of great practical importance, and it may be usefully applied in many ways.

Referring to the accompanying drawing, the figure is a view in elevation and part section of an improved free terminal and circuit of large surface with supporting structure and generating apparatus.

The terminal D consists of a suitably shaped metallic frame, in this case a ring of nearly circular cross section, which is covered with half spherical metal plates P P, thus constituting a very large conducting surface, smooth on all places where the electric charge principally accumulates. The frame is carried by a strong platform expressly provided for safety appliances, instruments of observation, etc., which in turn rests on insulating supports F F. These should penetrate far into the hollow space formed by the terminal, and if the electric density at the points where they are bolted to the frame is still considerable, they may be specially protected by conducting hoods as H.

A part of the improvements which form the subject of this specification, the transmitting circuit, in its general features, is identical with that described and claimed in my original Patents Nos. 645,576 and 649,621. The circuit comprises a coil A which is in close inductive relation with a primary C, and one end of which is connected to a ground-plate E, while its other end is led through a separate self-induction coil B and a metallic cylinder B′ to the terminal D.

The connection to the latter should always be made at, or near the center, in order to secure a symmetrical distribution of the current, as otherwise, when the frequency is very high and the flow of large volume, the performance of the apparatus might be impaired. The primary C may be excited in any desired manner, from a suitable source of currents G, which may be an alternator or condenser, the important requirement being that the resonant condition is established, that is to say, that the terminal D is charged to the maximum pressure developed in the circuit, as I have specified in my original patents before referred to. The adjustments should be made with particular care when the transmitter is one of great power, not only on account of economy, but also in order to avoid danger. I have shown that it is practicable to produce in a resonating circuit as E A B B′ D immense electrical activities, measured by tens and even hundreds of thousands of horse-power, and in such a case, if the points of maximum pressure should be shifted below the terminal D, along coil B, a ball of fire might break out and destroy the support F or anything else in the way. For the better appreciation of the nature of this danger it should be stated, that the destructive action may take place with inconceivable violence. This will cease to be surprising when it is borne in mind, that the entire energy accumulated in the excited circuit, instead of requiring, as under normal working conditions, one quarter of the period or more for its transformation from static to kinetic form, may spend itself in an incomparably smaller interval of time, at a rate of many millions of horse power. The accident is apt to occur when, the transmitting circuit being strongly excited, the impressed oscillations upon it are caused, in any manner more or less sudden, to be more rapid than the free oscillations. It is therefore advisable to begin the adjustments with feeble and somewhat slower impressed oscillations, strengthening and quickening them gradually, until the apparatus has been brought under perfect control. To increase the safety, I provide on a convenient place, preferably on terminal D, one or more elements or plates either of somewhat smaller radius of curvature or protruding more or less beyond the others (in which case they may be of larger radius of curvature) so that, should the pressure rise to a value, beyond which it is not desired to go, the powerful discharge may dart out there and lose itself harmlessly in the air. Such a plate, performing a function similar to that of a safety valve on a high pressure reservoir, is indicated at V.

Still further extending the principles underlying my invention, special reference is made to coil B and conductor B′. The latter is in the form of a cylinder with smooth or polished surface of a radius much larger than that of the half spherical elements P P, and widens out at the bottom into a hood H, which should be slotted to avoid loss by eddy currents and the purpose of which will be clear from the foregoing. The coil B is wound on a frame or drum D^1 of insulating material, with its turns close together. I have discovered that when so wound the effect of the small radius of curvature of the wire itself is overcome and the coil behaves as a conductor of large radius of curvature, corresponding to that of the drum. This feature is of considerable practical importance and is applicable not only in this special instance, but generally. For example, such plates at P P of terminal D, though preferably of large radius of curvature, need not be necessarily so, for provided only that the individual plates or elements of a high potential conductor or terminal are arranged in proximity to each other and with their outer boundaries along an ideal symmetrical enveloping surface of a large radius of curvature, the advantages of the invention will be more or less fully realized. The lower end of the coil B—which, if desired, may be extended up to the terminal D—should be somewhat below the uppermost turn of coil A. This, I find, lessens the tendency of the charge to break out from the wire connecting both and to pass along the support F′.

Having described my invention, I claim:

1. As a means for producing great electrical activities a resonant circuit having its outer conducting boundaries, which are charged to a high potential, arranged in surfaces of large radii of curvature so as to prevent leakage of the oscillating charge, substantially as set forth.

2. In apparatus for the transmission of electrical energy a circuit connected to ground and to an elevated terminal and having its outer conducting boundaries, which are subject to high tension, arranged in surfaces of large radii of curvature substantially as, and for the purpose described.

3. In a plant for the transmission of electrical energy without wires, in combination with a primary or exciting circuit a secondary connected to ground and to an elevated terminal and having its outer conducting boundaries, which are charged to a high potential, arranged in surfaces of large radii of curvature for the purpose of preventing leakage and loss of energy, substantially as set forth.

4. As a means for transmitting electrical energy to a distance through the natural media a grounded resonant circuit, comprising a part upon which oscillations are impressed and another for raising the ten-

sion, having its outer conducting boundaries on which a high tension charge accumulates arranged in surfaces of large radii of curvature, substantially as described.

5. The means for producing excessive electric potentials consisting of a primary exciting circuit and a resonant secondary having its outer conducting elements which are subject to high tension arranged in proximity to each other and in surfaces of large radii of curvature so as to prevent leakage of the charge and attendant lowering of potential, substantially as described.

6. A circuit comprising a part upon which oscillations are impressed and another part for raising the tension by resonance, the latter part being supported on places of low electric density and having its outermost conducting boundaries arranged in surfaces of large radii of curvature, as set forth.

7. In apparatus for the transmission of electrical energy without wires a grounded circuit the outer conducting elements of which have a great aggregate area and are arranged in surfaces of large radii of curvature so as to permit the storing of a high charge at a small electric density and prevent loss through leakage, substantially as described.

8. A wireless transmitter comprising in combination a source of oscillations as a condenser, a primary exciting circuit and a secondary grounded and elevated conductor the outer conducting boundaries of which are in proximity to each other and arranged in surfaces of large radii of curvature, substantially as described.

9. In apparatus for the transmission of electrical energy without wires an elevated conductor or antenna having its outer high potential conducting or capacity elements arranged in proximity to each other and in surfaces of large radii of curvature so as to overcome the effect of the small radius of curvature of the individual elements and leakage of the charge, as set forth.

10. A grounded resonant transmitting circuit having its outer conducting boundaries arranged in surfaces of large radii of curvature in combination with an elevated terminal of great surface supported at points of low electric density, substantially as described.

NIKOLA TESLA.

Witnesses:
M. Lamson Dyer,
Richard Donovan.

E. F. W. ALEXANDERSON.

ANTENNA.

APPLICATION FILED SEPT. 13, 1917.

1,360,167.

Patented Nov. 23, 1920.

2 SHEETS—SHEET 1.

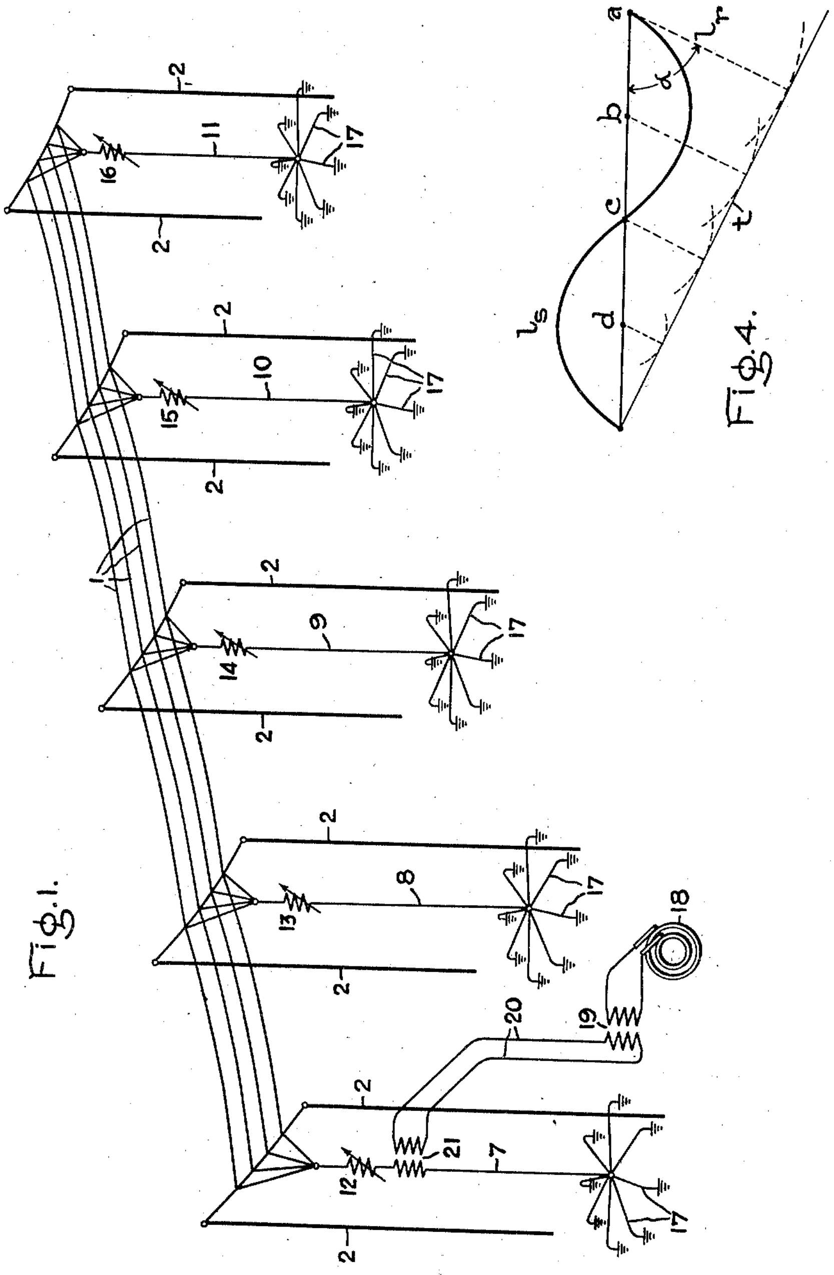
a
b
c
d
l_r
l_s
α
t

Fig. 4.

2
2
2
2
2
2
2
2
2
2
1
7
8
9
10
11
12
13
14
15
16
17
17
17
17
17
18
19
20
21

Fig. 1.

Inventor:
Ernst F. W. Alexanderson,
by Albert G. Davis
His Attorney.

E. F. W. ALEXANDERSON.
ANTENNA.
APPLICATION FILED SEPT. 13, 1917.

1,360,167.

Patented Nov. 23, 1920.
2 SHEETS—SHEET 2.

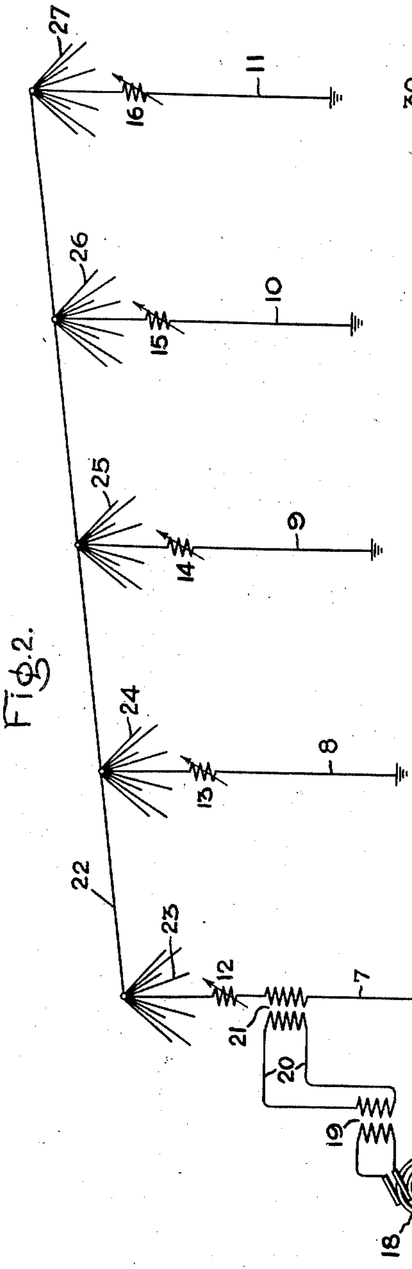
27
16
11
26
15
10
25
14
9
24
13
8
22
23
12
7
21
20
19
18

Fig. 2.

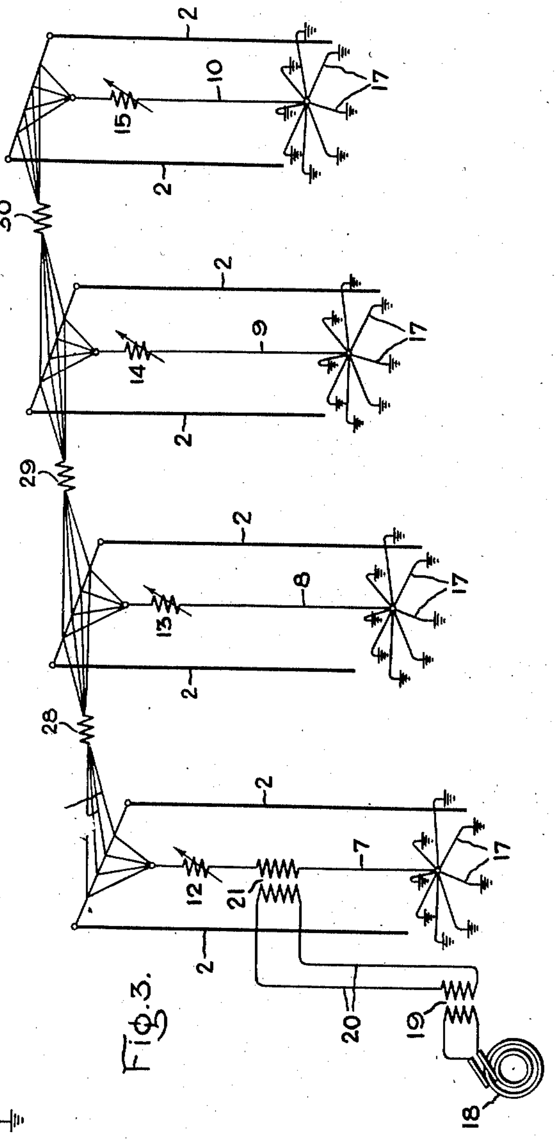
2
10
15
17
2
30
2
9
14
17
2
29
2
8
13
17
2
28
2
12
21
7
17
2
20
19
18

Fig. 3.

Inventor:
Ernst F. W. Alexanderson
by Albert G. Davis
His Attorney.

UNITED STATES PATENT OFFICE.

ERNST F. W. ALEXANDERSON, OF SCHENECTADY, NEW YORK, ASSIGNOR TO GENERAL ELECTRIC COMPANY, A CORPORATION OF NEW YORK.

ANTENNA.

1,360,167. Specification of Letters Patent. **Patented Nov. 23, 1920.**

Application filed September 13, 1917. Serial No. 191,110.

To all whom it may concern:

Be it known that I, ERNST F. W. ALEXANDERSON, a citizen of the United States, residing at Schenectady, in the county of Schenectady, State of New York, have invented certain new and useful Improvements in Antennæ, of which the following is a specification.

My present invention relates to antennæ for radio signaling systems and more particularly to the manner in which the radiation of the transmitting system is effected. My present application is a continuation in part of my prior application Serial No. 123,276 filed Oct. 2, 1916.

The antenna of a radio signaling system as previously constructed has consisted of an elevated electrical conductor or a net work of conductors which is charged by a source of high frequency energy in such a way that it becomes a source of energy radiation of the type known as the Hertzian oscillator. The theory for the radiation of an antenna has therefore been universally treated by the mathematical theory of the Hertzian oscillator.

The usual radiating antenna system may be considered as a single Hertzian oscillator. It has been proposed heretofore to employ a plurality of such oscillators in order to increase the amount of radiation or to secure directive effects but no practical use has been made of this idea because of the difficulties of controlling the phase of the oscillations in the different radiators in such a way that they would produce a combined radiation of a predetermined desired character.

One of the objects of my invention is to provide means whereby it will be possible to operate a plurality of practically separate radiators in such a way that a combined radiation of the desired character may be readily obtained.

There are several advantages which may be obtained by operating a radiating system as a plurality of separate radiators instead of as a single radiator. One of these advantages is that less energy consumption is required in order to produce a given strength of signals. This advantage may be explained briefly as follows:

In any single radiator, the strength of the electrostatic and the electromagnetic field of the radiated wave is proportional to the current and voltage supplied to the radiator. The energy consumed by the radiator is proportional to the product of current and voltage and is consequently proportional to the square of the field intensity of the radiated wave. The general theory of wave motion, however, teaches that if several systems of waves are superimposed, they combine in such a way that the field intensity in any one place is the algebraic sum of the momentary intensities of all the separate waves. If a system consisting of a plurality of separate radiators is controlled in such a way that the relative phase of the oscillations from the individual radiators is made to combine in a predetermined desired way, it will be possible to operate the system in such a manner that the field intensity in the receiving station is the arithmetic sum of the field intensities produced by all of the individual radiators. A radiation of unity intensity from a station with a single radiator may be said to produce a field intensity of unity in the receiving station. If a system comprising a plurality of radiators is operated in such a way that each of the radiators emits a wave of unity intensity, the effect on the receiving station will be the same as that of a wave with a field strength of as many times unity as there are individual radiators.

The energy consumption of the single radiator may also be called unity and the energy consumption of each of the individual radiators of the multiple system will also be unity when operated as described. If then we assume that there are six individual radiators we will obtain with an energy consumption of six times that of the single radiator a field intensity at the receiving station which is six times as great as that obtained with a single radiator. If it were, on the other hand, desired to produce by a single radiator, a wave at the receiving station having a field strength equal to six times unity it would be necessary to increase the current as well as the voltage in the single radiator to six times unity and consequently the energy consumption of the single radiator would be 36 times unity. Thus it will be seen that the energy consumption of the single radiator will be six times as great as that of the six individual radiators in order to produce the same intensity of signals.

A second advantage of my multiple radiation system is that it renders possible the production of signals of greater intensity than it has been possible to produce heretofore regardless of the amount of energy consumed. With any antenna there is a practical limit to the amount of radiation which may be obtained, this limit being partially fixed by the maximum voltage for which it is practical to insulate the aerial. In other words, it is not possible to increase the strength of the signal above a certain value by increasing the amount of energy supplied to a single radiator, whereas with a plurality of individual radiators any increase in the signal strength desired may be obtained merely by increasing the number of radiators.

A third advantage of my system of multiple radiation is that it renders possible the production of a radiation of energy with a decided directive effect comparable with the focusing of a beam of light by a lens or a mirror. While the desirability of such an effect has frequently been mentioned and various means have been proposed for securing it none of the means heretofore proposed have been found to be applicable to practical operation.

A second object of my invention is to provide a convenient and practical means for adjusting the phase relations of the individual radiations so that the advantages of directive radiation may be realized.

In the usual type of antenna corresponding to a single Hertzian oscillator which has heretofore been used the energy is introduced through a ground lead in series with a loading coil. In an antenna of this type designed for long distance transmission assume that the current introduced through the loading coil is 100 amperes and the potential of the aerial is 50,000 volts. The energy required to maintain oscillations in such an antenna may be 50 kilowatts. This means that the energy component of the oscillating voltage is 500 volts, or as it is commonly expressed, the antenna has a resistance of 5 ohms. Thus the energy required for maintaining oscillations is delivered in the form of 100 amperes at 500 volts. While it might be possible to supply energy to each one of a plurality of radiators in this way such a method would, if the individual radiators were located at a considerable distance from each other, involve unnecessary losses due to inductive or resistance drop in the feeding wires.

A further object of my invention is to provide a system for supplying energy to the individual radiators which will overcome this disadvantage.

In attaining this object of my invention I transform the energy to the form of high potential and low current and instead of supplying it at the ground point of the loading coil supply it at a high tension point. For example, in the case above mentioned, the energy may be supplied to the radiators at 50,000 volts and one ampere. Thus while in each case 50 kilowatts of energy is required the old form of antenna consumes this energy in the same form as if it were a resistance of five ohms whereas the radiator supplied in accordance with my invention consumes the energy in the form of a resistance of 50,000 ohms. This may be explained by reason of the fact that in the old form of antenna the energy is introduced in series with the inductance and capacity of the antenna whereas in the form of a radiator which I employ the energy is introduced to a circuit which is the equivalent of the antenna capacity in parallel with the inductance. By introducing the energy to the individual radiators in this way it is possible to feed a large number of such radiators located at a considerable distance from each other without incurring any great losses in the transmission conductors. In order to transform the energy into the desired form it might be possible to use a transformer of the ordinary type with the required ratio of transformation. I have found, however, that it is more convenient to use one of the radiating circuits as an oscillation transformer, and to supply the energy to all of the other radiators from the high tension end of the first radiator. In its preferred form then, my multiple radiating system may be considered as being made up of a primary radiator which at the same time serves as an oscillation transformer for supplying high tension energy to one or more secondary radiators. In order to have a completely balanced system of radiation, it might appear desirable to locate the primary radiator in such a way relative to the secondary radiators that the energy connections to each of the secondary radiators can be made directly to the primary radiator through wires of equal length. While this might be a practical method of construction, I have found in actual practice that it is not necessary to connect each of the secondary radiators directly to the primary radiator and that it is of practical convenience in the construction of the antenna to connect the secondary radiators successively to each other, that is so that energy current flowing from the primary radiator reaches the secondary radiators in succession. In one case in which my invention has been applied in practice the antenna has the character of a horizontal transmission line supported on towers. This antenna is divided into a plurality of individual radiators by means of ground connections with loading coils therein at suitable intervals along its length. The aerial wires which serve as radiators also serve for transmitting energy to all of the successive radiators. Since the energy is transmitted to the secondary radiators in the form of high potential and low current it is apparent that a large number of radiators may be connected to such a transmission line without causing any appreciable changes in voltage and phase.

The novel features which I believe to be characteristic of my invention are set forth with particularity in the appended claims. My invention itself however, both as to its organization and method of operation together with further objects and advantages thereof will best be understood by reference to the following description taken in connection with the accompanying drawing in which Figures 1, 2 and 3 show diagrammatically as many different modifications and Fig. 4 is a diagram explanatory of the directive radiation feature of my invention.

In the form of my invention indicated in Fig. 1 the radiating system is composed of a plurality of horizontal conductors 1 of considerable length elevated above ground by means of the supports 2 from which they are insulated. This system is divided into five individual radiators by means of the ground connections 7, 8, 9, 10, and 11, individual tuning or loading coils 12, 13, 14, 15, and 16 being included in each ground connection. The ground connections indicated consist of a plurality of wires 17 stretched over the surface of the ground, the end of each conductor being permanently earthed. The energy for signaling purposes is supplied to the first radiator of the series by a high frequency alternator 18, connected through the transformer 19 to a transmission line 20 which conveys the current to the transformer 21 by means of which it is introduced into ground connection 7.

The first radiator comprises the ground connection 7 and the portions of the conductors 1 adjacent thereto and this radiator serves as an oscillation transformer to step up the voltage applied through transformer 21. Energy is thereby supplied directly at a high potential from the first radiator to the second radiator, which comprises the ground connection 8 and the portions of the conductors 1 adjacent thereto. The second radiator also acts as an oscillation transformer changing the high potential low current energy into the form of lower potential larger current producing an oscillating charging current in ground connection 8 which may be of the same magnitude as the charging current in ground connection 7. The conductors 1 also serve as a transmission line for conveying the high potential energy to the third, fourth and fifth radiators in succession. These radiators function in the same manner as the second radiator so that charging currents of the same order of magnitude are set up in all of the ground connections. As an indication of the advantage which may be obtained by the use of my invention tests have been made with an antenna of the type shown in Fig. 1 which was divided up into 6 individual radiators. When 50 kilowatts of high frequency energy were supplied to the antenna and a single ground connection used a charging current of about 100 amperes was obtained. When six ground connections were used the charging current in each ground connection was about 45 amperes making a total charging current of 270 amperes. In other words the effective radiation with the same amount of energy was about 2.7 times that which was obtained when the antenna was operated as a single radiator.

In accordance with the Hertzian theory of radiation the ability of a certain aerial to radiate is determined by the so-called radiation resistance. This radiation resistance is proportional to the square of the ratio between height and wave length. It is also well known that an antenna to be used efficiently should be operated at a somewhat lower frequency than its natural periodicity. This method of operation calls for a certain amount of inductance which is ordinarily used for tuning. If it is attempted to operate the antenna at a higher frequency than the natural periodicity it is necessary to introduce a capacity instead of an inductance for loading and it is known that the operation in this way is inefficient. For long distance working antennæ of large horizontal dimensions are employed. The large horizontal dimensions give high electrostatic capacity and consequently the natural periodicity of the antenna which is determined by the physical dimensions of the aerial is low. Thus if such an antenna is operated at its nautral periodicity, or at a lower frequency the wave length is much greater than the height and consequently the ratio between height and wave length results in a low radiation resistance and a high ratio between the radiation resistance and the ground resistance; in other words, in low efficiency. The value to which the ground resistance may be reduced is also limited to a certain extent in practice by the physical dimensions of the antenna. Hence it may be said that such an antenna when used as a simple Hertzian oscillator has a definite maximum possible efficiency or natural ratio between radiation resistance and ground resistance. When, however, the antenna is provided with a plurality of ground connections in accordance with my invention, its natural periodicity is increased beyond the value determined by its physical dimensions when used as a simple Hertzian oscillator, that is, the natural periodicity of the entire antenna is equal to the natural periodicity of each individual section which is supplied with a ground connection, assuming that the natural periodicity of all of the sections is the same. In other words, the natural periodicity of the antenna is increased to a value much greater than the value which is determined by its physical dimensions. As a result the antenna can be operated efficiently at much shorter wave lengths thereby increasing the radiation resistance inversely with the square of the wave length. The result of this is a more favorable ratio between radiation resistance and ground resistance, since the effect of providing a plurality of ground connections is to decrease the ground resistance; that is, the ratio between radiation resistance and ground resistance or efficiency of the antenna may be greatly increased.

The form of my invention illustrated in Fig. 2 differs from that shown in Fig. 1 merely in its constructional details. In this case each individual radiator takes the form of an umbrella type of antenna and the individual radiators are all connected by a transmission line 22 which supplies high potential energy from the first radiator 23 to the secondary radiators 24, 25, 26 and 27 in succession. A radiating system constructed in this way operates in the same manner as the one illustrated in Fig. 1.

In Fig. 3 I have indicated a modification of my invention which may be employed when it is desired to obtain a directive effect. In this case the individual radiators are separated by inductances 28, 29 and 30 which may be made of such a value as to give the phase displacement in the waves radiated by successive radiators, required to secure the desired directive effect. If desired, condensers may be employed between the individual radiators instead of inductances.

In determining a method whereby a system comprising a plurality of individual radiators may be adapted for directive radiation an antenna such as illustrated in Figs. 1 or 2 may be considered as a uniformly distributed capacity between a linear conductor and ground. The electrostatic field surrounding the conductor extends in three directions but for the purpose of this discussion it will be sufficient to consider the potential gradient only in two directions, that is in the vertical direction between conductor and ground and in the horizontal direction along the conductor. When an aerial conductor is set into oscillation by a source of continuous waves of definite frequency standing waves are formed in the electrostatic field. These waves must be analyzed in accordance with the horizontal as well as the vertical distribution of potential. If the conductor is not connected to ground in any way it has a natural period of horizontal oscillations. In this case no oscillations will occur in the vertical plane. If one end of the conductor is connected to ground it will have another natural period. The oscillations in this case will be horizontal and also to some extent vertical inasmuch as a charging current will pass between the conductor and ground. An aerial conductor thus arranged with the ground connection at one end has previously been used to considerable extent. If such a conductor is connected to ground through a large number of uniformly distributed coils it becomes a multiple radiating antenna as described previously. Furthermore, if the inductance of these coils is so adjusted that the oscillations in all of the different coils are substantially in phase the antenna may be said to have vertical oscillations but no horizontal oscillations. It is, however, possible to tune the antenna in such a way that it has both vertical and horizontal oscillations simultaneously. The vertical and horizontal oscillations may be of different frequency if a source of energy supply of more than one frequency is available. This might be the case if the antenna was excited by an arc generator or a pliotron oscillator. In the present case, however, it will be most convenient to consider the radiation as of a single frequency such as may occur if the antenna is excited by a high frequency alternator, in which case the horizontal and vertical oscillations that take place simultaneously are necessarily of the same frequency. In the case where the conductor is connected to ground at one end only the horizontal oscillations occur in such a way that the grounded end is a node and the open end has maximum amplitude. It was at one time considered that an antenna of this type would produce a directive radiation. It has now, however, been proved that this is not the case, the reason being that the length of the antenna has been less than one-fourth of the wave length of radiation at which it has been operated, whereas to secure an effective directive radiation it would be necessary that the length of the antenna should be at least equal to one-half of the wave length of radiation. In a simple horizontal antenna such as has been described the distribution of capacity and inductance has always been found to be such that the wave length of the standing wave along the antenna is shorter than the wave length of radiation. This condition, as will be explained more fully hereafter, prevents the securing of the most effective radiation. The fundamental rule which should be followed and which will be explained in the following discussion is that the wave length of the standing wave should be equal to or greater than the wave length of radiation. If radiation is desired in the longitudinal direction of the antenna the standing wave should have a length equal to the wave length of radiation, whereas if a directive radiation is desired at right angles to the direction of the antenna the length of the standing wave should be infinity. For any intermediate angle the relative length of the standing wave and the radiation wave should be equal to the secant of the angle between the direction of radiation and the direction of the antenna. In order to prove this relation geometrically, assume that a wave begins to travel with the velocity of light from one end of the antenna. It is desirable in order to secure maximum efficiency of radiation that the wave as it travels along the antenna should always receive impulses which are in phase with the wave itself. For this reason the standing wave should have a wave length equal to the wave length of the traveling wave. If an antenna oscillates with a standing wave which has a greater wave length than the wave length of radiation, that is, if the wave travels along the antenna at a velocity greater than that of light, the different points of the antenna may be assumed to emit radiation of different phases and the locus for points at which individual waves are of a certain phase may be represented by circles drawn around the different centers of radiation.

Referring now to Fig. 4, assume that the standing wave is represented by l_s. Around point *a* of the standing wave which is represented as a node a circle may be drawn having a radius equal to the wave length of radiation l_r. Around point *b* a circle may be drawn of a radius one-quarter of a wave length shorter. Around point *c* the circle has a radius of one-half a wave length and around point *d* a radius of one-quarter of a wave length. The line *t* which is tangent to all of these circles indicates the direction of radiation, the radiation being greatest in a direction perpendicular to this tangent, and it will be apparent from an inspection of this figure that $l_r = l_s \cos \alpha$ where α is the angle between the direction of the antenna and the direction of radiation. Applying the above formula to a case where the standing horizontal wave is shorter than the wave length of radiation the direction of such radiation would be expressed by a cosine greater than unity. This condition being impossible means that no directive radiation can take place and that the radiation will be inefficient.

The type of horizontal antenna connected to ground at one end which has previously been used as such distribution of inductance and capacity that the standing wave is theoretically equal to but in practice has been shorter than the wave length of radiation. In other words, the wave travels along the antenna at a velocity less than the velocity of light. If it is desired to increase the velocity at which the wave travels along the antenna and thereby lengthen the standing wave either the inductance or capacity must be decreased. Anything however which is done in this direction by changing the combination of the wires will defeat the object sought. An additional parallel wire will decrease the inductance but increase the capacity and fewer parallel wires will decrease the capacity but increase the inductance. By the use of a plurality of radiators such as I have described the desired object may be attained. The capacity between the aerial and ground may be neutralized by distributed inductance between the aerial and ground. If the distributed capacity is entirely neutralized by distributed inductance for the frequency at which the antenna is excited the standing wave will have a length equal to infinity, which means that all parts of the aerial conductor will oscillate in phase. This condition produces non-directive radiation if the antenna is shorter than half of the wave length of radiation and directive radiation at right angles to the antenna if the antenna is sufficiently long in relation to the wave length of radiation. If a long antenna is used and directive radiation is desired at any angle except a right angle, the necessary length for the standing wave may be determined from the wave length of the radiation and the desired angle. The inductance of the aerial being known it will then be possible to calculate the amount of distributed capacity which will create a standing wave of the desired length. This required capacity will be found to be lower than the capacity between the aerial and ground. The method to be followed then will be to neutralize the portion of the capacity to ground which is not desired, leaving a balance of distributed capacity which will give a standing wave of the desired length. Instead of neutralizing the capacity to ground it will be possible to neutralize the longitudinal inductance of the aerial conductor by series condensers. Both of these methods may be used in combination if desired.

If the method of adjusting the tuning of the antenna is considered from the above point of view it will be found that the method of radiation which consists in using a number of independent radiators excited from a common source of high frequency energy in such a way as to operate in phase is only a specific case of a composite horizontal antenna with a horizontal standing wave. This specific case is one where the length of the standing wave is infinity and this method of radiation whether directive or nondirective may be differentiated from the methods of radiation that have been proposed heretofore by the characteristic that the standing wave has a length equal to or greater than the wave length of radiation, whereas such antennæ as have been used previously have had standing waves of a wave length shorter than the wave length of radiation. The term “standing wave” as used in the above description and in the claims which follow is not employed in the restricted sense in which it is commonly understood, namely, as the resultant of two waves traveling with equal velocity in opposite directions so as to produce points of large amplitude and intermediate nodes of zero amplitude. The phenomenon described by this term is also of a different nature than that known as a “traveling wave” since the latter term as ordinarily used implies a complete transfer of energy from one point to another at the same rate that the wave travels. While in the present case energy is transferred from one point to another this transfer is merely incidental as a large portion of the energy remains fixed in position. Moreover, such transfer as does occur may take place in the opposite direction to the traveling wave effect. At any point along the length of the antenna the amplitude of oscillations is the same as at any other point, but when the standing wave, as the term is here used, is of finite length the oscillations of different portions of the antenna are displaced in phase. The maxima of successive portions occur in succession, however, so that the effect, so far as the wave in the ether is concerned, is that of a traveling wave.

While I have illustrated and described the preferred form of my invention, it will be apparent that many modifications therein may be made without departing from the scope of the appended claims.

What I claim as new and desire to secure by Letters Patent of the United States, is:

1. A radiating antenna system comprising a primary radiating unit supplied by a source of radio-frequency current and a plurality of secondary radiating units an individual ground connection for each of said radiating units, the secondary radiating units being connected successively to a transmission line which supplies energy from the primary radiating unit to the secondary radiating units.

2. A radiating antenna system comprising a primary radiating unit and a plurality of secondary radiating units connected successively together and to the primary unit, an individual ground connection for each of said radiating units, a source of high frequency current connected to the primary unit, the primary radiating unit being directly connected to the first of said secondary units and serving as an oscillation transformer for supplying energy successively to all of the secondary radiating units at a high potential.

3. A radiating antenna system comprising a set of horizontal aerial wires, a plurality of inductance coils connected between said aerial wires and ground at substantially uniformly distributed points along the length of said wires and a source of radio frequency current for supplying energy to the aerial wires at one point, the energy thus supplied being distributed through the aerial wires to maintain synchronous oscillations in each of the oscillating circuits formed by one of said inductances, and the capacity to ground of the portions of the aerial wires adjacent thereto.

4. A radiating antenna system comprising a plurality of sections, all of the different sections being joined together successively by aerial wires, each section consisting of aerial wires connected to ground through an inductance and forming a separate oscillating circuit, and a source of radio frequency current for supplying energy to the aerial wires which join the several sections, each oscillating circuit being tuned to the frequency of the energy thus supplied.

5. The method of supplying energy to a radiating system comprising a plurality of radiating units which consists in supplying energy to one of said units at a comparatively low potential, utilizing said unit as an oscillation transformer for producing a comparatively high potential and distributing energy at the high potential thus produced from the first unit successively to all of the other units.

6. The method of maintaining synchronous oscillations in a plurality of radiating units which consists in supplying radio frequency energy directly to one of said units and supplying energy from said first unit successively to all of the other radiating units.

7. The method of radiating electromagnetic energy from a system of horizontal conductors which consists in producing an electrostatic field in said conductors which has a horizontal standing wave formed therein having a wave length which is at least as great as the wave length which is to be radiated.

8. The method of radiating electromagnetic energy from a radiating antenna system which consists in producing by a single source of radio frequency energy an electrostatic field in said radiating system which has a horizontal standing wave formed therein having a wave length which is at least as great as the wave length which is to be radiated.

9. A radiating antenna system comprising a set of aerial conductors, means for exciting said conductors by a single source of radio

frequency energy, and means for neutralizing, by resonance, the distributed capacity between the aerial conductors and ground, the conductors being so arranged and said means being so adjusted as to produce a standing wave of potential along the aerial conductor having a wave length at least as great as the wave length to be radiated.

10. A radiating antenna system comprising aerial conductors arranged in a horizontal line, a single source of radio frequency current for exciting said antenna system, a plurality of inductances distributed along said aerial conductors and connected between said conductors and ground, said conductors being so arranged and said inductances being adjusted so as to partly neutralize the distributed capacity between the aerial conductor and ground in such a way that the distribution of potential along the aerial conductors will form a standing wave of a wave length at least as great as the wave length to be radiated.

11. The method of radiating electromagnetic energy which consists in charging an insulated continuous system of aerial conductors through a plurality of vertical grounded conductors connected thereto in multiple, the oscillating charging currents in the different ground conductors being regulated so as to have a progressive phase displacement.

12. The method of radiating electromagnetic energy which consists in charging an insulated continuous system of aerial conductors through a plurality of vertical grounded conductors connected thereto in multiple, the oscillating charging currents in the different ground conductors being regulated so that they are progressively shifted in phase to such a degree as to give a directive radiation in a desired direction.

13. The method of radiating electromagnetic energy from a radiating antenna system which consists in neutralizing distributed capacity between said system and ground at a plurality of points in the system, exciting said system by a single source of radio frequency energy, and producing a standing wave of potential along the system having a wave length at least as great as the wave length to be radiated.

14. The method of creating an electromagnetic wave with a directive radiation in a system of aerial wires having a plurality of inductance coils connected between said aerial wires and ground at substantially uniformly distributed points along said wires which consists in producing synchronous oscillating currents in all of the ground conductors by means of energy introduced into the system through one ground conductor and regulating the currents in the different conductors so that they will have a progressive phase displacement.

15. The method of directive signaling which consists in supplying energy from a source of radio frequency current to a radiating unit, supplying energy from said radiating unit to a plurality of other radiating units and adjusting the relative phases of the currents in the different radiating units so that the waves emitted by all of the units add accumulatively at points in a certain predetermined direction.

16. The method of directive signaling by means of electromagnetic waves which consists in supplying energy from a source of radio frequency current to a radiating unit and supplying energy from said radiating unit to a plurality of other radiating units which are so adjusted that the interference effect of the waves emitted therefrom creates a maximum amplitude of signal in a desired direction and a minimum amplitude in another direction.

17. The combination in a radio signaling system of an insulated system of aerial conductors, a plurality of vertical conductors each of which has one end connected to said aerial system and the other end grounded, means for supplying energy to the entire system through one of said vertical conductors to produce synchronous oscillating currents in all of the vertical conductors, and means for adjusting the phase relation of the currents in the different vertical conductors in such a way as to produce a directive radiation.

18. The combination in a radio signaling system of an insulated system of aerial conductors, a plurality of vertical conductors each of which has one end connected to said aerial system and the other end grounded, means for supplying energy to the entire system through one of said vertical conductors to produce synchronous oscillating currents in all of the vertical conductors and means for adjusting the relative phases of the different oscillating currents so that the waves produced by these currents add accumulatively at points in a certain predetermined direction.

19. The combination in a radio signaling system of an insulated system of aerial conductors, a plurality of vertical conductors each of which has one end connected to said aerial system and the other end grounded, means for supplying energy to the entire system through one of said vertical conductors to produce synchronous oscillating currents in all of the vertical conductors and means for adjusting the relative phases of the different oscillating currents so that the interference effect of the waves produced creates a maximum amplitude of signal in a desired direction and a minimum amplitude in another direction.

20. A radiating antenna system comprising a continuous system of aerial conductors

arranged in a horizontal line, a plurality of ground connections for said system containing tuning means located at substantially uniformly distributed points along the length of said system whereby said system is divided into a plurality of radiating units capable of oscillating substantially independent of one another, and means for supplying energy to said system through a single one of said ground connections.

21. A radiating antenna system comprising a system of continuous elevated horizontal conductors, a plurality of ground connections for said system containing tuning means located at substantially uniformly distributed points along the length of said system whereby said system is divided into a plurality of radiating units capable of oscillating substantially independent of one another, and means for supplying energy to said system through the ground connection nearest one end whereby said elevated conductors serve both as a plurality of substantially independent radiating units, and to transmit energy at a high potential from the first radiator to the other radiators in succession.

22. An antenna for a wireless signaling system comprising a long horizontal continuous conducting network supported above the earth and connections between said network and earth at a plurality of successive points along its length whereby the natural periodicity of the antenna is increased beyond the value determined by its physical dimensions when used as a simple Hertzian oscillator, all of said earth connections including inductances adjusted in such a way that the multiple impedance of the inductances is equal to that of a single inductance necessary for tuning the whole antenna at the operating frequency.

23. The combination in a wireless signaling system of an antenna comprising a long horizontal conducting network supported above the earth and connections between the said network and earth at a plurality of successive points whereby the natural periodicity of the antenna is increased beyond the value determined by its physical dimensions when used as a simple Hertzian oscillator, all of said earth connections including inductances, and means for supplying the energy necessary for maintaining oscillations of a single definite frequency in the entire system through one of the multiple paths to ground.

24. The combination in a wireless signaling system of an insulated system of long horizontal continuous aerial conductors, a plurality of successive vertical conductors each of which has one end connected to said aerial system and the other end grounded, and a source of sustained high frequency oscillations for producing synchronous oscillating currents of a single definite frequency in a plurality of multiple paths, each of which comprises a vertical conductor and the space capacity of a portion of the aerial.

25. The combination in a wireless signaling system of an insulated system of long horizontal continuous aerial conductors, a plurality of vertical conductors each of which has one end connected to said aerial system and the other end grounded, and a source of sustained high frequency oscillations for producing substantially in phase oscillating currents of a single definite frequency in all of the vertical conductors.

26. The combination in a wireless signaling system of an insulated system of long horizontal continuous aerial conductors, a plurality of vertical conductors each of which has one end connected to said aerial system and the other end grounded, and a source of sustained high frequency oscillations for producing oscillating currents of a single definite frequency in all of the vertical conductors which differ in phase from one another by less than 90°.

27. An antenna for a wireless signaling system comprising a long horizontal continuous conducting network supported above the earth and having of itself a low natural periodicity and connections between said network and earth at a plurality of successive points along its length, whereby its natural periodicity is increased to a much greater value, all of said earth connections including inductances for tuning said antenna.

28. The combination in a wireless signaling system of an insulated system of continuous aerial conductors, a plurality of vertical conductors each of which has one end connected to said continuous aerial system and the other end grounded, means for producing synchronous oscillating currents in all of the vertical conductors, and means for adjusting the phase relation of the currents in the different vertical conductors in such a way as to produce a directive radiation.

29. The combination in a wireless signaling system of an insulated system of continuous aerial conductors, a plurality of vertical conductors each of which has one end connected to said continuous aerial system and the other end grounded, means for producing synchronous oscillating currents in all of the vertical conductors, and means for adjusting the relative phases of the different oscillating currents so that the waves produced by these currents add accumulatively at points in a certain predetermined direction.

30. The combination in a wireless signaling system of an insulated system of continuous aerial conductors, a plurality of vertical conductors each of which has one end

connected to said continuous aerial system and the other end grounded, means for producing synchronous oscillating currents in all of the vertical conductors, and means for adjusting the relative phases of the different oscillating currents so that the interference effect of the waves produced creates a maximum amplitude in a desired direction and a minimum amplitude in another direction.

31. An antenna for a wireless signaling system comprising a long horizontal conductor insulated from the earth and having such natural constants that an electrical wave will travel along its length at a velocity less than the velocity of light, and means uniformly distributed along the length of said conductor for increasing the velocity at which an electrical wave will travel along the conductor.

32. An antenna for a wireless signaling system comprising a long horizontal conductor insulated from the earth and having such natural constants that an electrical wave will travel along its length at a velocity less than the velocity of light, and means uniformly distributed along the length of said conductor for increasing the velocity at which an electrical wave will travel along the conductor to a velocity at least equal to the velocity of light.

33. An antenna for a wireless signaling system comprising a long horizontal conductor insulated from earth and having such natural constants that an electrical wave will travel along its length at a velocity less than the velocity of light, and a plurality of inductances uniformly distributed along the length of the conductor and connected between the conductor and earth for neutralizing a portion at least of the shunt capacity thereof and thereby increasing the velocity at which an electrical wave will travel along the conductor.

34. An antenna for a wireless signaling system comprising a long horizontal conductor insulated from earth and having such natural constants that an electrical wave will travel along its length at a velocity less than the velocity of light, and a plurality of inductances uniformly distributed along the length of the conductor and connected between the conductor and earth for neutralizing a portion at least of the shunt capacity thereof and thereby increasing the velocity at which an electrical wave will travel along the conductor to a velocity at least equal to the velocity of light.

35. An antenna for a wireless signaling system comprising a long horizontal conductor insulated from earth and having such natural constants that an electrical wave will travel along its length at a velocity less than the velocity of light, a plurality of inductances uniformly distributed along the length of the conductor and connected between the conductor and earth for neutralizing a portion at least of the shunt capacity thereof, and thereby increasing the velocity of the electrical wave in the conductor, and a plurality of inductances connected in series with said conductor between the different earth points for decreasing the velocity of the electrical waves in the conductor to a desired value.

In witness whereof, I have hereunto set my hand this 11th day of September, 1917.

ERNST F. W. ALEXANDERSON.

Oct. 13, 1931. C. J. BRITTEN 1,826,727

RADIO APPARATUS

Filed April 30, 1928

Fig-1-

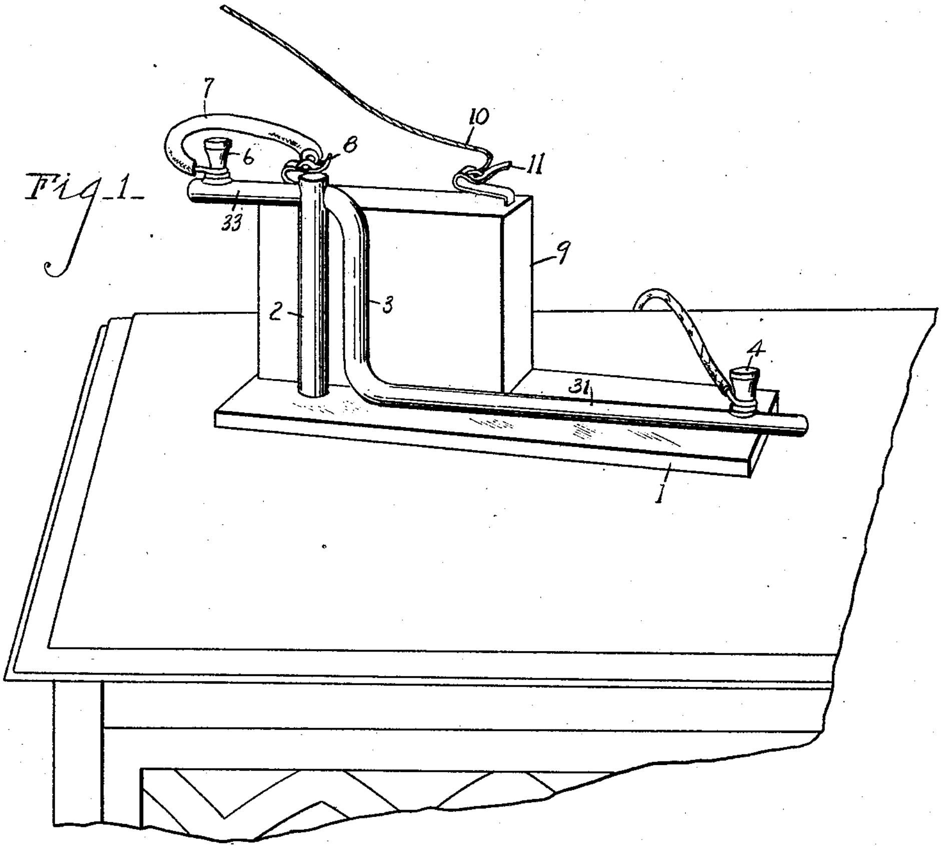
7
6
8
10
11
33
9
2
3
4
31
1

Fig-2-

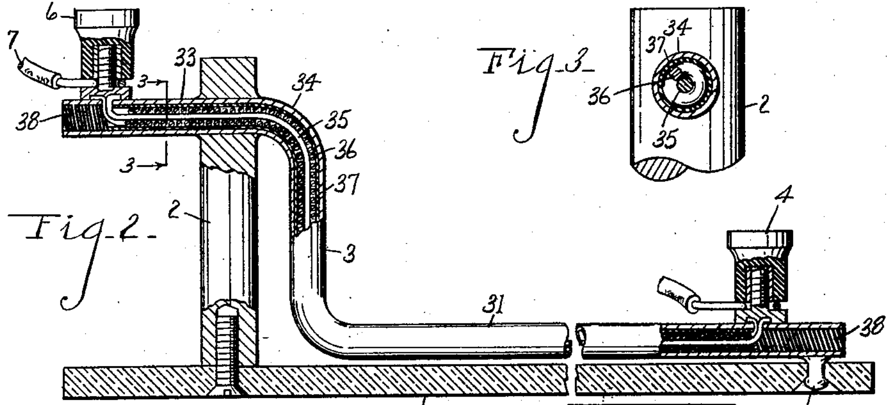
6
7
33
3
34
35
38
36
37
3
2
3
4
31
38
1
32

Fig-3-

34
37
36
2
35

Fig-4-

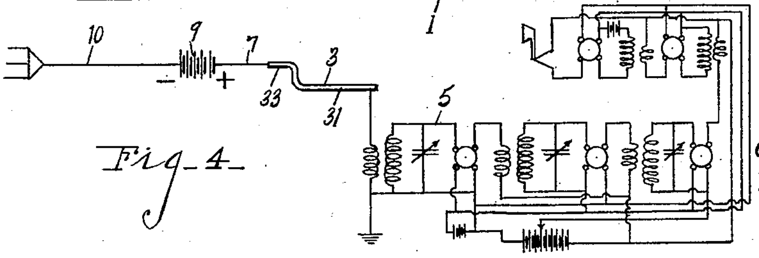
10
9
7
3
33
31
5

INVENTOR
Chancy J. Britten
By Chappell & Earl
ATTORNEYs

Patented Oct. 13, 1931 **1,826,727**

UNITED STATES PATENT OFFICE

CHANCY J. BRITTEN, OF CHARLOTTE, MICHIGAN, ASSIGNOR TO BRITTEN LIGHT & POWER COMPANY, OF LANSING, MICHIGAN, A CORPORATION OF MICHIGAN

RADIO APPARATUS

Application filed April 30, 1928. Serial No. 273,949.

This invention relates to improvements in radio apparatus in the nature of an economizer and clarifier.

The objects of the invention are particularly to economize and conserve the current and to regulate and to clarify the tone.

Further objects and objects pertaining to details and economies of construction and operation will definitely appear from the description to follow.

I accomplish the objects of my invention by the devices and means described in the following specification. The invention is clearly defined and pointed out in the claim.

A device embodying my invention is clearly illustrated in the accompanying drawings forming a part of this specification, in which:

Fig. 1 is a detail perspective view of my improvement designed as an attachment to an ordinary radio set.

Fig. 2 is a full size vertical detail longitudinal elevation view partially in section on a central plane.

Fig. 3 is an enlarged detail sectional view on line 3—3 of Fig. 2 through the body of the main connection.

Fig. 4 is a detail diagrammatic wiring view showing the relation of my apparatus in combination with a radio set.

The parts will be identified by their numerals of reference which are the same in all the views.

1 is the base which may be of glass or suitable insulating material. 2 is a metal supporting column secured to the glass base in any suitable way. 3 is the main connector of my improved economizer and regulator having a horizontal portion 31 disposed in proximity to the base and secured thereto by a suitable welded or soldered button 32. It has an upright portion and horizontal portion 33 disposed through the upright supporting post 2. This connector comprises an outer tube or casing 34 preferably of copper, an inner central wire 35, around which is disposed a coil of wire 36. The coil of wire is embraced within a layer of mica insulating material 37. The tube is preferably under partial vacuum. Plugs 38 of insulation are in the ends of connector 3.

The inner central wire 35 and the coil 36 have their ends connected to suitable binding posts. The binding post 4 on the lower horizontal part provides the connection to the radio set 5 containing the usual A, B and C batteries and being otherwise organized as indicated diagrammatically in Fig. 4.

The binding post 6 is provided with connection 7 leading to the positive terminal 8 of the battery 9. The negative terminal 11 of the battery 9, which is a 4½ volt dry cell battery is connected to the lead 10 to the aerial.

The battery 9 is a duplicate of the C battery of 4½ volt capacity but is not coupled up to act as a C battery but in the particular auxiliary relation specified herein.

Just how the structure results in the advantages stated is problematical and applicant is only able to state the results and define the structure which accomplishes the same without expanding on the theory by which it is accomplished.

Having thus described my invention, what I claim as new and desire to secure by Letters Patent, is:

The combination with a radio set comprising A, B and C batteries and an aerial of an economizer and clarifying device comprising an auxiliary battery dry cell, a base with post, a metallic tubular casing with a horizontal lower part and an upright and an upper horizontal part disposed through said post, a central conductor, a coil of wire around the conductor, the opposite ends of said conductor and coil being connected to a single binding post mounted on the upper and lower ends of the tubular casing which casing is suitably sealed at the ends, a mica insulation between the said coil and the casing, a connection from the lower binding post to the antenna connection of the radio set and from the upper binding post to the positive terminal of said battery dry cell, and a connection from the negative terminal of the said battery cell to the aerial or antenna of said radio set.

In witness whereof I have hereunto set my hand.

CHANCY J. BRITTEN.

Nov. 12, 1957 L. R. CRUMP 2,813,242

POWERING ELECTRICAL DEVICES WITH ENERGY
ABSTRACTED FROM THE ATMOSPHERE
Filed March 12, 1954

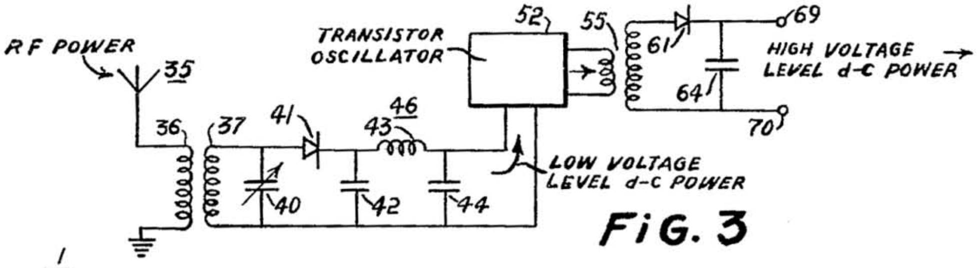
RF POWER
35
36
37
41
43
46
40
42
44
TRANSISTOR OSCILLATOR
52
55
61
64
69
70
HIGH VOLTAGE LEVEL d-C POWER
LOW VOLTAGE LEVEL d-C POWER

FiG. 3

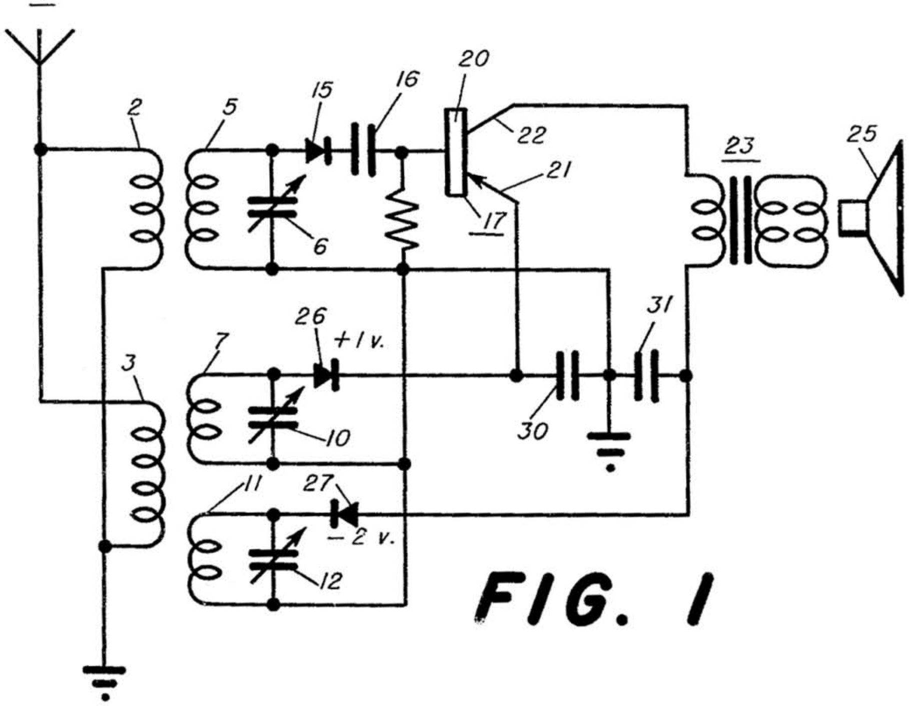
1
2
5
15
16
20
22
21
17
6
23
25
3
7
26
+1 v.
10
30
31
11
27
-2 v.
12

FIG. 1

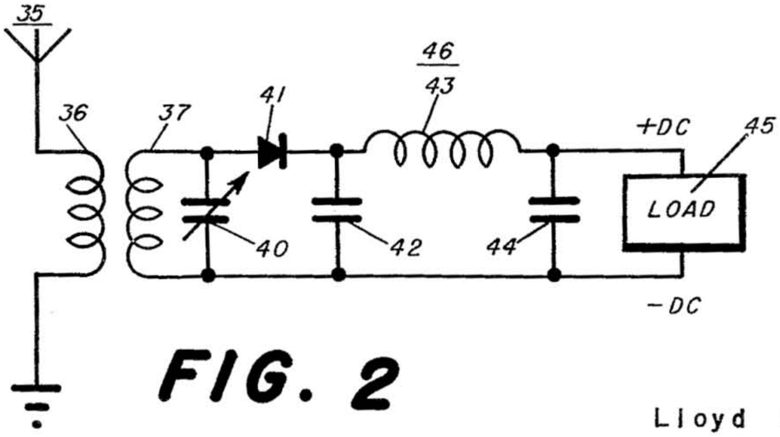
35
36
37
41
46
43
40
42
44
+DC
45
LOAD
-DC

FIG. 2

INVENTOR
Lloyd R. Crump

BY W. E. Thibodeau + A. W. Dew

ATTORNEYS

United States Patent Office

2,813,242

Patented Nov. 12, 1957

2,813,242

POWERING ELECTRICAL DEVICES WITH ENERGY ABSTRACTED FROM THE ATMOSPHERE

Lloyd R. Crump, Silver Spring, Md.

Application March 12, 1954, Serial No. 415,986

1 Claim. (Cl. 321—2)

(Granted under Title 35, U. S. Code (1952), sec. 266)

The invention described herein may be manufactured and used by or for the Government for governmental purposes without payment to me of any royalty thereon.

This invention relates to the convenient and economical provision of power for the operation of electronic circuits and devices using transistors, and of other electrical devices having modest power requirements.

A great advantage of transistors, and a major reason for their enthusiastic reception since their introduction a few years ago, is the fact that they will operate satisfactorily with very low supply voltages and currents. One milliwatt or even less is sufficient to power a transistor in many applications. Various batteries have been developed to provide, in a minimum of space, the relatively minute amounts of power needed by transistors.

My invention provides methods and means that permit transistor circuits, and also other low-powered electrical devices, to be economically and conveniently operated without any batteries whatever, and indeed without any power supply whatever as power supplies are ordinarily conceived.

The invention centers around my discovery that it is practicable to construct operative transistor circuits that are able to abstract from the atmosphere sufficient electromagnetic energy to provide all necessary supply voltages and currents for their own operation. Circuits and devices powered according to my invention will operate indefinitely without any local power source whatever.

I have successfully constructed and demonstrated such circuits. For example, I have constructed a batteryless transistor radio receiver on which I have listened to either nearby or distant broadcast stations as desired, using either headphones or a loudspeaker; this receiver has been powered entirely by electromagnetic energy abstracted from the atmosphere.

From the successful operation of this receiver, and from other experimental work, it becomes clear that, by the methods and means of the invention, a great variety of practical and useful transistor circuits can be powered entirely by energy abstracted from the atmosphere.

Furthermore, as will become apparent below, my invention is applicable to the powering of other electrical devices requiring relatively small amounts of power.

An object of the present invention is to provide methods and means for powering transistor circuits entirely from radiofrequency energy abstracted from the atmosphere.

Another object is to provide methods and means for powering remote radio receivers, low-powered radio transmitters, and other low-powered electrical devices, with energy received by radio from a master station, so that no local power supplies are needed by the devices and so that the powering or non-powering of the remote device is under the control of the master station.

A further object is to provide methods and means for powering transistor circuits and other low-powered electrical devices with radiofrequency energy received from one or more remote radio transmitters.

Other objects, aspects, uses, and advantages of the invention will become apparent from the following description and from the drawing.

Figure 1 is a schematic diagram of a transistor radio receiver in which all necessary power is supplied by energy abstracted from the atmosphere in accordance with the invention.

Figure 2 is a schematic diagram showing a general application of the invention to provide direct-current power to a load.

Figure 3 is a schematic diagram of a system for obtaining a high energy D.-C. source at a high voltage level using energy abstracted from the atmosphere.

Referring to Figure 1, a receiving antenna **1** is connected to antenna coupling coils **2** and **3**, the other ends of which are connected to ground. A parallel resonant circuit consisting of coil **5** and capacitor **6** is coupled to coil **2**. A second parallel resonant circuit consisting of coil **7** and capacitor **10** is coupled to coil **3**. A third parallel resonant circuit consisting of coil **11** and capacitor **12** is also coupled to coil **3**.

Coil **5** and capacitor **6** are tuned to the frequency of a radio transmitter from which it is desired to receive information—for instance, an amplitude-modulated standard broadcast station. The signal received from this transmitter need not be strong. The signal is detected by diode **15** to obtain an audio-frequency information signal. This audio signal is coupled through a capacitor **16** and is amplified by a circuit that includes a transistor **17** having a base **20**, an emitter **21**, and a collector **22**. The amplified audio output of the transistor is coupled through an audio transformer **23** to an electroacoustical transducer, preferably a permanent-magnet dynamic loudspeaker **25** as shown.

The novelty of the invention lies largely in the method and means by which the necessary direct-current power is supplied to the emitter and collector circuits of transistor **17**. This method and means will now be described.

Coil **7** and capacitor **10**, and also coil **11** and capacitor **12**, are tuned to receive radio signals of relatively high strength. It does not matter whether these signals contain information. These power signals are rectified by diodes **26** and **27** to provide direct-current power that is filtered by capacitors **30** and **31**. The D.-C. power thus obtained is utilized to power the transistor **17**.

In the circuit shown, two tuned circuits (coil **7** and capacitor **10**, and coil **11** and capacitor **12**) are tuned to power signals and the D.-C. voltages obtained from each are connected in series. The tuned power circuits may be tuned to the same or different power signals. Under certain circumstances it may be desirable to use more than two tuned power circuits and to tune them to more than two power signals; in this way power can be obtained from several signals and combined. On the other hand, if a strong power signal is available, a single tuned power circuit may suffice to give the needed D.-C. power.

Even weak information signals can be received successfully. A plurality of transistor amplifier stages can be used if desired, or other circuits such as superheterodyne circuits can be used. It is merely necessary that a sufficiently strong power signal or signals be available to provide the small amount of D.-C. needed to power the transistors.

If the information signal happens to be strong, it can be used as the power signal; all of the tuned circuits (coil **5** and capacitor **6**, coil **7** and capacitor **10**, coil **11** and capacitor **12**) are tuned to the information signal.

Engineers who have observed my invention in operation have been surprised at the unexpectedly good results obtained, even with readily available power signals of quite moderate strength. For instance, sufficient power for satisfactory operation of a loudspeaker at low volumes is

readily obtained from a 5-kilowatt standard broadcast station 5 miles away, using only an indoor antenna to pick up the power signal as well as information signals. In typical operation under these conditions a D.-C. voltage of about 2.5 to 3 volts is obtained between the emitter and the collector, at a current of about 250 microamperes; D.-C. power input to the transistor is thus of the order of 0.5 to 1 milliwatt. So far as I am aware, no one has ever before discovered and demonstrated the practicability of this method of powering a radio receiver.

Because existing broadcast stations within a radius of a number of miles provide adequate power signals, the invention is readily practicable with existing power signals in almost any location in or near any city in the United States.

Although I have described a transistor radio receiver powered by my invention, it will be readily apparent that the invention is applicable to the powering of any transistor circuit using one or a number of transistors, and to the powering of other devices requiring relatively small amounts of power. For instance, sensitive electromechanical, electrochemical, or electrothermal devices can be operated by the method of the invention.

Referring to Figure 2, which shows a more general embodiment of my invention, an antenna **35** picks up radiofrequency energy from the atmosphere. This energy flows through coil **36**, which is coupled to a tuned circuit consisting of coil **37** and capacitor **40**. The radiofrequency voltage across capacitor **40** is rectified by diode **41** and filtered by a low-pass filter **46** consisting of capacitors **42** and **44** and choke coil **43**. The resulting D.-C. voltage is applied to a load **45**.

In the practice of my invention, larger amounts of power can be obtained for short periods of time by storing received energy in a suitable energy storage device. Stored energy may then be withdrawn at intervals at a more rapid rate than that at which it was received and put into the storage device. In this way the invention can be used to provide short pulses of relatively very high electrical energy. This result can be readily obtained by charging a relatively large capacitor with direct current and then discharging the capacitor rapidly into a load when desired. This rapid discharge can be initiated automatically when the voltage across the capacitor reaches a certain level, or it can be initiated when a transistor radio receiver receives a certain information signal.

Higher voltages can be obtained with the invention by means of well known devices for raising D.-C. voltages as shown in Figure 3. The D.-C. voltage output from the capacitor **44** can be used to power a low frequency transistor oscillator **52** whose A.-C. output is raised to a higher voltage level by the transformer **55**. This relatively high A.-C. voltage can then be rectified by a diode **61** and fed to a capacitor **64** to provide a high energy D.-C. source at a relatively high voltage level at the terminals **69** and **70**. If desired, energy can now be withdrawn from the capacitor **64** at intervals in short pulses of high energy at a high voltage level. Pulsed radio transmission is one of the possible uses for this form of the invention. Other uses would be to provide a single relatively powerful pulse needed to actuate an electrothermal or electromechanical device.

As has been indicated above, in many locations and particularly anywhere in or near most American cities, power signals normally present in the atmosphere are readily available for the easy and convenient practice of the invention. However, the invention also has important applications in systems in which the necessary power signal is generated and transmitted specifically for the operation of the particular system. Such systems can, for example, comprise a master station transmitting all the power that is needed for hundreds or thousands of fixed or mobile transistor receivers or other remote devices over a range of many miles. This eliminates the need for hundreds or thousands, as the case may be, of local power supplies. At the same time, such a system has the advantage that all of the remote devices can be simultaneously activated or deactivated at the will of the master station, simply by starting or stopping the transmission of the power signal. In such systems it will often be advantageous to use power signals of frequencies sufficiently high to permit the use of resonant receiving antennas of small physical dimensions for signal pickup at the remote devices. In addition to the power signal, the master station may transmit an information signal on the same or a different carrier.

Certain types of devices powered entirely by received radio waves are of course well known. The well-known "crystal set" of the early days of radio, which used a diode rectifier to demodulate an amplitude-modulated radiofrequency signal, is an outstanding example of such a device. My invention is readily distinguishable from such prior devices, however. In typical prior devices a modulated radiofrequency signal is applied to a diode to obtain unidirectional half-wave pulses whose amplitudes vary with modulation. These pulses are integrated by means of a capacitor to obtain a unidirectional signal the amplitude of which follows the audiofrequency modulation envelope. If the radiofrequency signal is received with sufficient strength the audio signal may have sufficient power to operate headphones or similar utilization device without power amplification; but the signal is utilized for its information content, rather than to supply non-information-containing power.

My invention, on the other hand, entails the utilization of received radiofrequency energy to supply power to at least one pair of circuit points (across capacitor **31** in Fig. 1, for example), such circuit points requiring power solely for its power content and not for any information or modulation it may contain. In other words, my invention entails the utilization of radiofrequency energy to supply power that would otherwise have to be supplied by batteries, generator, or other local power source.

It will be apparent that the embodiments shown are only exemplary and that various modifications can be made in construction and arrangement within the scope of the invention as defined in the appended claim.

I claim:

An electrical device for obtaining a high energy D.-C. source at a high voltage level using energy abstracted from the atmosphere, said device comprising in combination: resonant means for receiving radio waves, first rectifier means for converting said radio waves into first direct current energy, first capacitor means for storing said first direct current energy, an oscillator powered by said direct current energy, said oscillator producing an A.-C. output, transformer means for raising said A.-C. output to an increased voltage level, second rectifier means for converting the A.-C. output of increased voltage level from said transformer into second direct current energy, and second capacitor means for storing said second direct current energy.

References Cited in the file of this patent

UNITED STATES PATENTS

1,540,998	Plauson	June 9, 1925
2,056,824	Cawley	Oct. 6, 1936
2,553,734	Adler	May 22, 1951
2,569,687	Pan	Oct. 2, 1951
2,647,957	Mallinckrodt	Aug. 4, 1953

OTHER REFERENCES

Turner: "A crystal receiver with transistor amplifier," Radio and Television News, January 1950.

Aug. 16, 1960 T. T. BROWN **2,949,550**

ELECTROKINETIC APPARATUS

Filed July 3, 1957 2 Sheets-Sheet 1

FIG. 1

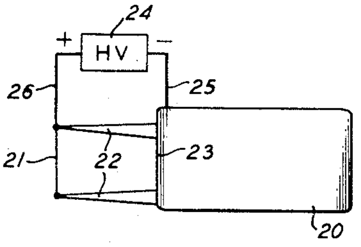

24
+
HV
−
26
25
21
22
23
22
20

FIG. 2

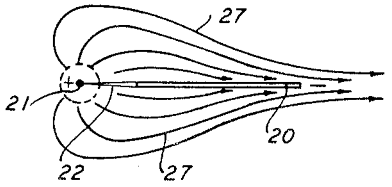

27
21
+
20
22
27

FIG. 3

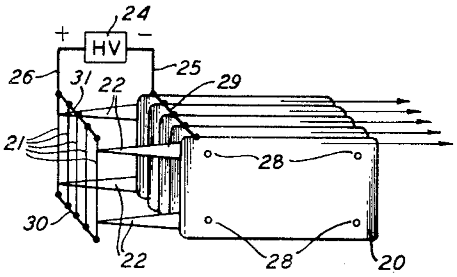

24
+
HV
−
26
25
31
22
29
21
28
30
22
28
20

FIG. 4

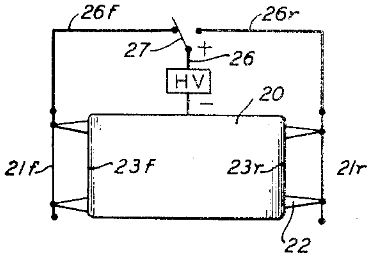

26f
26r
27
+
26
HV
−
20
21f
23f
23r
21r
22

FIG. 5

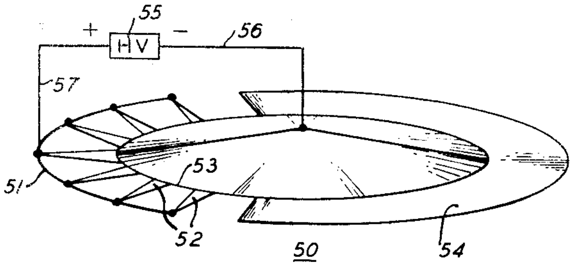

55
+
HV
−
56
57
53
51
52
50
54

INVENTOR
THOMAS TOWNSEND BROWN
BY
Watson, Cole, Grindle & Watson
ATTORNEYS

Aug. 16, 1960 T. T. BROWN **2,949,550**

ELECTROKINETIC APPARATUS

Filed July 3, 1957

2 Sheets-Sheet 2

FIG. 6

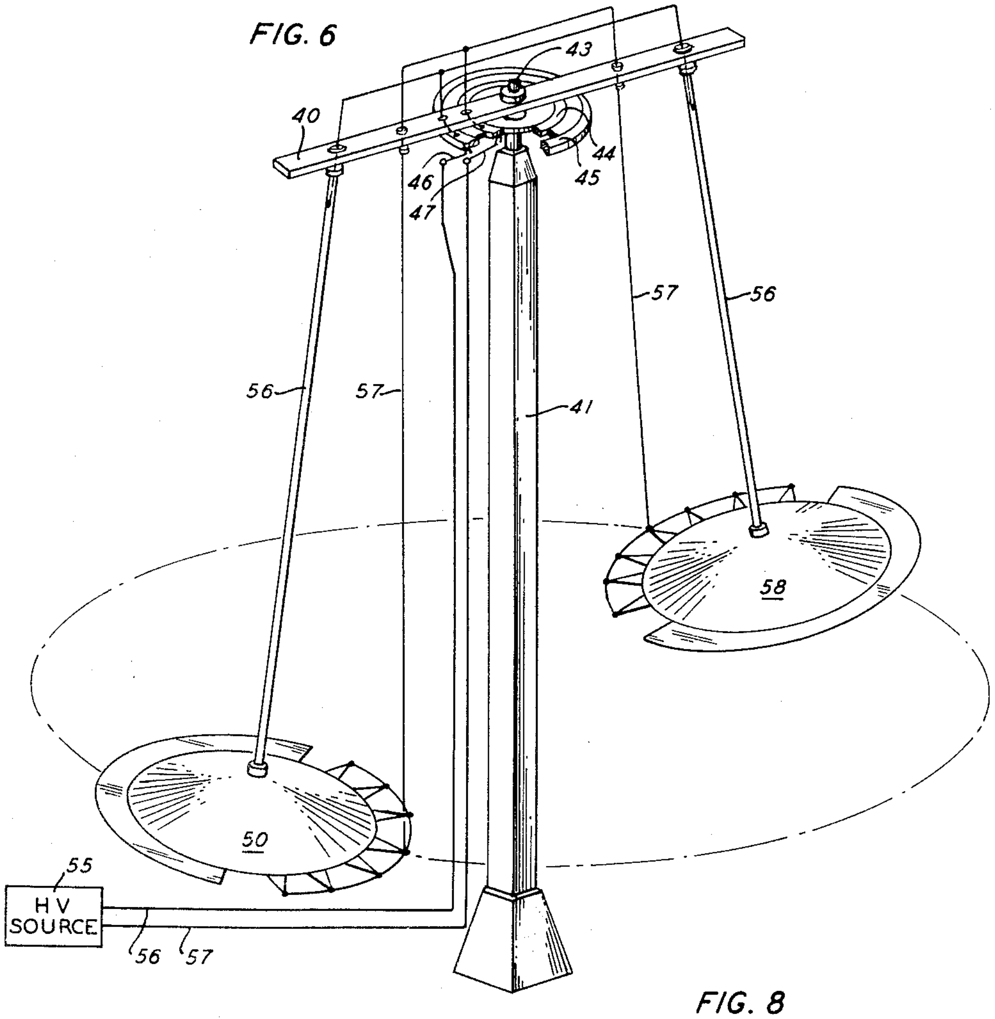
43
40
44
45
46
47
57
56
56
57
41
58
50
55
H V
SOURCE
56
57

FIG. 8

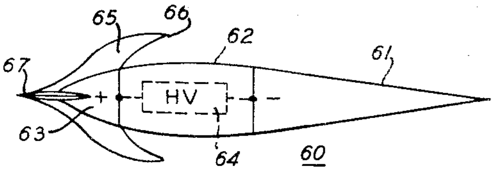
65
66
62
61
67
+
HV
−
63
64
60

FIG. 7

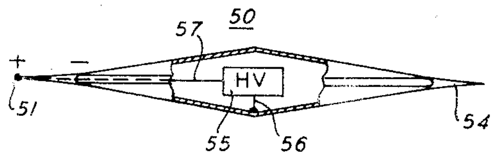
57
50
+
−
HV
51
55
56
54

INVENTOR

THOMAS TOWNSEND BROWN

BY

Watson, Cole, Grindle & Watson

ATTORNEYS

United States Patent Office

2,949,550

Patented Aug. 16, 1960

2,949,550

ELECTROKINETIC APPARATUS

Thomas Townsend Brown, Umatilla, Fla., assignor to Whitehall-Rand, Inc., Washington, D.C., a corporation of Delaware

Filed July 3, 1957, Ser. No. 669,830

12 Claims. (Cl. 310—5)

My invention relates to electrokinetic apparatus, and more particularly to a method and apparatus for utilizing electrical potentials for the production of forces for the purpose of causing relative motion between a structure and the surrounding medium.

This invention was disclosed and described in my application Serial No. 293,465, filed June 13, 1952, which application has become abandoned. However, reference may be made to this application for the purpose of completing the disclosure set forth below.

The invention utilizes a heretofore unknown electrokinetic phenomenon which I have discovered; namely, that when a pair of electrodes of appropriate form are held in a certain fixed spaced relation to each other and immersed in a dielectric medium and then oppositely charged to an appropriate degree, a force is produced tending to move the pair of electrodes through the medium. The invention is concerned primarily with certain apparatus for utilizing such phenomenon in various manners to be described.

Priorly, intervening electrokinetic apparatus has been employed to convert electrical energy to mechanical energy and then to convert the mechanical energy to the required force. Except for the insignificantly small forces of electrostatic attraction and repulsion, electrical energy has not been used for the direct production of force and motion.

Since any conversion of energy from one form to another is accompanied by losses due to friction, radiation or conduction of heat, hysteresis, and the like, as well as serious reductions in the availability of the energy by increases in the entropy of the system, it is apparent that great increases in efficiency may be achieved through the use of the direct production of electrical energy and force and motion made possible by my invention. Likewise, the elimination of the machinery for the intermediate conversions results in great savings in first costs, maintenance, weight and space, the latter two being of great importance in self-propelled vehicles including mobile vehicles such as aircraft and space craft.

It is therefore an object of my invention to provide an apparatus for converting the energy of an electrical potential directly into a mechanical force suitable for causing relative motion between a structure and the surrounding medium.

It is another object of this invention to provide a novel apparatus for converting an electrical potential directly to usable kinetic energy.

It is another object of this invention to provide a novel apparatus for converting electrostatic energy directly into kinetic energy.

It is another object of this invention to provide a vehicle motivated by electrostatic energy without the use of moving parts.

It is still another object of this invention to provide a self-propelled vehicle without moving parts.

It is a feature of my invention to provide an apparatus

for producing relative motion between a structure and the surrounding medium which apparatus includes a pair of electrodes of appropriate form held in fixed spaced relation to each other and immersed in a dielectric medium and oppositely charged.

It is another feature of my invention to provide apparatus which includes a body defining one electrode, another separate electrode supported in fixed spaced relation by said body, and a source of high electrical potential connected between the body and the separate electrode.

It is also a feature of my invention to provide apparatus having a body which is hollow and a source of potential contained within the body.

It is another feature of my invention to provide apparatus having a body and an electrode connected to the body, which combination comprises a vehicle.

It is also a feature of my invention to provide apparatus which comprises a plurality of assemblies, each including a body and an electrode secured in side-by-side spaced relation to each other.

It is another feature of my invention to provide vehicular apparatus which includes a pair of electrically conductive body portions joined by an insulating portion, whereby said electrically conductive portions constitute the electrodes.

Other objects and advantages of my invention will be apparent from a consideration of the following specification, read in connection with the accompanying drawings, wherein:

Figure 1 is a side elevational view illustrating diagrammatically a simple form of apparatus embodying and functioning in accordance with the principles of my invention;

Figure 2 is a plan view of the apparatus shown in Figurue 1;

Figure 3 is a perspective view illustrating the manner in which a plurality of devices of the character illustrated in Figure 1 may be interconnected for joint operation;

Figure 4 is a diagrammatic view similar to Figure 1 illustrating a modified form of the invention providing a means for reversing the direction of the propulsive force produced;

Figure 5 is a perspective view illustrating diagrammatically a self-propelled device utilizing the principles of this invention;

Figure 6 is a perspective view of one illustrative embodiment of this invention showing a pair of electrokinetic propulsion devices suspended from a rotatable arm which arm is supported at its midpoint;

Figure 7 is a side elevational view of a mobile vehicle with parts broken away to show the interior construction;

Figure 8 is a side elevational view illustrating diagrammatically the arrangement of parts used in an alternative form of mobile vehicle.

Referring to the drawings, I have illustrated in Figure 1 a simple form of apparatus which is readily adaptable for use in demonstrating the principles of my invention, and which is utilized in this application as a simplified representation to facilitate an understanding of the principles involved. The apparatus illustrated in Figure 1 constitutes one electrode which is preferably in the form of a body member **20**, said member preferably comprising a relatively thin flat plate. A second electrode **21** in the form of a wire or other suitable form of electrical conductor is held as by means of insulated supports **22** in fixed spaced relation to the body **20**, the wire **21** being disposed in the plane of the body **20** and preferably substantially parallel with a leading edge **23** of the body **20**. A source **24** of high voltage electrical potential is provided and connected as shown at **25** and **26** to the two electrodes **20** and **21**, respectively.

I have discovered that when apparatus of the character just described is immersed in a dielectric medium, as for example, the ordinary air of the atmosphere, there is produced a force tending to move the entire assembly through the medium, and this force is applied in such direction as to tend to move the body **20** toward the leading electrode **21**. This force produces relative motion between the apparatus and the surrounding fluid dielectric. Thus, if the apparatus is held in a fixed position, the dielectric medium is caused to move past the apparatus and to this extent the apparatus may be considered as analogous to a pump or fan. Conversely, if the apparatus is free to move, the relative motion between the medium and the apparatus results in a forward motion of the apparatus, and it is thus seen that the apparatus is a self-propulsive device.

While the phenomenon just described has been observed and its existence confirmed by repeated experiment, the principles involved are not completely understood. It has been determined that the greatest forces are developed when the leading electrode is made positive with respect to the body **20**, and it is accordingly thought that in the immediate vicinity of the electrode **21** where the potential gradient is very high, free electrons are stripped off of the atoms and molecules of the surrounding medium. These electrons migrate to the positive electrode **21** where they are collected. This removal of free electrons leaves the respective atoms and molecules positively charged and such charged atoms and molecules are accordingly repelled from the positive electrode **21** and attracted toward the negative electrode **20**. The paths of movement of these positively charged particles appear to be of the nature represented by the lines **27** in Figure 2.

It appears that upon reaching or closely approaching the surface of the body **20**, the positively charged atoms and molecules have their positive charges neutralized by the capture of electrons from the body **20** and in many cases, it may be that excess electrons are captured whereby to give such atoms and molecules a negative charge so that they are actually repelled from the body **20**.

It will be appreciated that the mass of each of the individual electrons is approximately one two-thousandths the mass of the hydrogen atom and is accordingly negligible as compared with the mass of the atoms and molecules of the medium from which they are taken. The principal forces involved therefore are the forces involved in moving the charged atoms and molecules from the region of the positive electrode **21** to and beyond the negatively charged body **20**. The force so exerted by the system on those atoms and molecules not only produces a flow of the medium relative to the apparatus, but, of course, results in a like force on the system tending to move the entire system in the opposite direction; that is, to the left as viewed in Figure 1 of the drawing.

The above suggested explanation of the mode of operation of the device is supported by observation of the fact that the dimensions and potentials utilized must be adjusted to produce the required electric field and the resulting propulsive force. Actually I have found that the potential gradient must be below that value required to produce a visible corona since corona is objectionable inasmuch as it represents losses through the radiation of heat, light and molecular charges in the medium.

My experiments have indicated that the electrode **21** may be of small diameter for the lower voltage ranges, i.e. below 125 kv. while above this voltage, rod or hollow pipe electrodes are preferred. These large electrodes are preferred for the higher voltages since sharp points or edges are eliminated which at these elevated potentials would produce losses thus diminishing the thrust. For example, electrodes to be operated at potentials below 125 kv. may be made from small gauge wire only large enough to provide the required mechanical rigidity while electrodes to be operated at potentials above 125 kv. may be hollow pipes or rods having a diameter of ¼ to ½ inch.

In Figure 3, I have illustrated the manner in which a plurality of assemblies, such as are shown in Figure 1, may be interconnected for joint operation. As may be seen from Figure 3, a plurality of such assemblies are placed in spaced side-by-side relation. They may be held fixed in such spaced relation through the use of a plurality of tie rods **28** and interposed spacers (not shown) placed between adjacent plates **20**. The assembly of plates **20** may be electrically interconnected by a bus bar or similar conductor **29** to which the negative lead **25** is connected. In a similar way, the plurality of positive leading electrodes **21** may be held in appropriately spaced relation to each other by fastening their ends to pairs of bus bars **30** and **31**, to the latter of which the positive lead **26** is connected. The assembly of leading electrodes **21** may be held in spaced relation to the assembly of body members **20** by an appropriate arrangement of the supports **22**.

In Figure 4, I have illustrated diagrammatically an arrangement of parts for producing a reversible action; that is, permitting the direction of the propulsive force to be reversed. The apparatus is similar to that shown in Figure 1, differing therefrom in utilizing a pair of leading electrodes **21*f*** and **21*r*** spaced by means of spacers **22** from the front and rear edges **23*f*** and **23*r*** of the body member **20** in a manner similar to that described with reference to the supports **22** in Figure 1. The source **24** of high voltage electrical potential has its negative terminal connected to the body **20** as by means of the aforementioned conductor **25**. The positive terminal is connected as by means of the conductor **26** to the blade **27** of a single-pole, double-throw switch, serving in one position to connect the conductor **26** to a conductor **26*f*** which is in turn connected to the forward electrode **21*f*** and arranged in its opposite position to connect the conductor **26** to a conductor **26*r*** which is in turn connected to the reverse electrode **21*r***.

It will be seen that with the switch **27** in the position shown in Figure 4, the apparatus will operate in the manner described in connection with Figure 1, causing the assembly to move to the left as viewed in Figure 4. By throwing the switch **27** to the opposite position, the direction of the forces produced are reversed and the device moves to the right as viewed in Figure 4.

In Figure 5, I have illustrated the principles of the invention as embodied in a simple form of mobile vehicle. This device includes a body member **50** which is preferably of the form of a circular disc somewhat thicker in its center than at its edges. The disc **50** constitutes one of the electrodes and is the equivalent of the body member **20** referred to in connection with Figure 1. A leading electrode **51** in the form of a wire or similar small diameter conductor is supported from the body **50** by a plurality of insulating supports **52** in uniform spaced parallel relation to a leading edge portion **53** of the body **50**. A skirt or similar fairing **54** may be carried by the body **50** to round out the entire structure so as to provide a device which is substantially circular in plan. A source of high voltage electrical potential **55** is provided with its negative terminal connected as indicated at **56** to the body **50** and its positive terminal connected as indicated at **57** to the leading electrode **51**.

The device operates in the same manner as the apparatus shown in Figure 1 to produce a force tending to move the entire assembly through the surrounding medium to the left as viewed in Figure 5 of the drawing.

Referring now to Figure 6, there is depicted an illustrative embodiment of this invention in which a pair of mobile vehicles, such as depicted in Figure 5, are shown suspended from the terminals of arm **40**, which arm is supported at its midpoint by a vertical column **41**. High voltage source **55** is shown connected through wires

56 and **57** which extend to the ends of arm **40** by way of suitable rotatable contacts **44** and **45** and brushes **46** and **47** adjacent point **43** in the center of arm **40**. Mobile vehicle **50** is shown suspended from one end of the lever arm **40** by means of conductors **56** and **57**. A similar vehicle **58** is shown suspended from the other end of the rotatable arm by conductors **56** and **57**. It is, of course, understood that these bodies may be suspended by any convenient structure such as wires or rods which wires or rods may support conductors **56** and **57** in any suitable manner.

In this illustrative embodiment the vehicles were caused to rotate at a speed of 17 feet per second with 50 kv. applied to conductors **56** and **57** from source **55**. It is, of course, understood that these figures are merely by way of illustrative example and, as might be expected, the speed of the vehicles increases exponentially with the applied voltage.

When the apparatus is to be used for propelling a mobile vehicle, it is, of course, necessary that the source **55** of high voltage be contained within and carried by the vehicle. This may be accomplished by using the apparatus in the manner shown in Figure 7, wherein the high voltage source **55** is contained within the hollow central portion of the body **50**, the conductor **56** being connected to the body and the conductor **57** being suitably insulated from the body **50** and extended externally thereof and into connection with the leading conductor **51**.

The apparatus of my invention may be used to propel vehicles of shapes other than that described in connection with Figures 5 and 7. For example, in Figure 8 I have illustrated the way in which the invention may be utilized in the propulsion of a vehicle of torpedo-like shape. As is shown in Figure 8, I use a body member which is indicated generally by the reference character **60** and which is formed of three parts, to wit, an afterbody **61** formed of electrically conductive material, a hollow central body portion **62** formed of insulating material, and a nose portion **63** formed of electrically conductive material. A source **64** of high voltage electrical potential is contained within the hollow central body portion **62** and has its terminals connected, respectively, to the nose portion **63** and afterbody portion **61**, the positive terminal being preferably the one which is connected to the nose portion **63**. Thus, the afterbody **61** functions as the aft electrode **20** described with reference to Figure 1 and the nose portion **63** corresponds to the leading electrode **21**.

The nose portion **63** may be equipped with a plurality of suitably shaped fins **65**. These fins may be extended aft of the junction between the central body **62** and the nose portion **63**, as shown at **66**, to provide ionizing elements which are spaced more closely to the afterbody **61** and which functions in a manner analogous to the smaller diameter electrode **21**. Also, the fins may be shaped to conform to the aerodynamic requirements and may, if desired, be movable in whole or in part for the purpose of permitting the machine to be maneuvered.

I have shown the nose portion **63** as being provided with a needle-like point **67**. By using such a nose form, which at present appears to be the best suited for flying speeds approaching or exceeding the speed of sound, I am able to produce an ionization of the atmosphere in the immediate region of this foremost portion of the mobile vehicle. I believe that this ionization facilitates piercing the sonic barrier and minimizes the abruptness with which the transition takes place in passing from subsonic velocities to supersonic velocities.

From the foregoing it will be observed that I have provided an electrokinetic method and apparatus for the production of forces suitable for causing relative motion between a structure and the surrounding medium. It will be observed that the methods and apparatus described herein are particularly adaptable for use as a propulsive means for self-propelled vehicles. I wish to emphasize that the high voltage power source referred to herein may be of relatively simple construction and relatively low capacity. For example, potentials of the order of 30 to 70 thousand volts may be adequate for use with this apparatus, the particular voltage employed dependent, of course, upon the size of the vehicle or apparatus. It will be appreciated that the elimination of moving parts in the apparatus will represent a tremendous saving in first cost and maintenance cost of the apparatus also the direct production of the motive forces from the electrical force represents a high efficiency so that greater propulsive forces and speeds may be obtained with apparatus occupying small space and of light weight.

While I have shown and described various embodiments of my invention, it is appreciated that the principles thereof may be extended to many and varied types of machines and apparatus. The invention therefore is not to be limited to the details illustrated and described herein.

Certain of the physical principles embodied herein are disclosed and described in my application Serial No. 669,727, filed July 3, 1957, which is directed to the generation of high voltages by means of electrokinetic apparatus.

In my application Serial No. 669,831, filed July 3, 1957, electrokinetic principles are employed in conjunction with apparatus somewhat similar to that disclosed and described in the instant application to produce a novel transducer.

I claim:

1. Thrust producing apparatus comprising an electroconductive body having a relatively large, substantially smooth area thereof exposed to the surrounding medium, an electrode supported on said body, insulated therefrom and shaped to produce a concentrated electrostatic field when electrically charged with respect to said body, and power means connected between said body and electrode to impress high voltage thereon whereby to cause relative movement of an ionizable dielectric medium surrounding said electrodes.

2. Thrust producing apparatus comprising an electrode having a substantially flat exposed surface, an elongated electrode extending along one edge of said surface and equidistant therefrom and supported in fixed spaced relationship with respect to said edge of said surface and substantially in the plane thereof, and power means connected across said electrodes to impress high voltage thereon whereby to cause relative movement of an ionizable dielectric medium surrounding said electrodes.

3. Thrust producing apparatus comprising a substantially flat hollow electrode, an electrode supported in fixed spaced relationship with respect to an edge of said flat electrode and substantially in the plane thereof, and power means within said hollow electrode connected across said electrodes to impress high voltage thereon thereby causing relative movement of an ionizable dielectric medium surrounding said electrodes.

4. Thrust producing apparatus comprising a substantially flat electrode having a pair of electrodes supported in fixed spaced relationship with opposite edges of said flat electrode and substantially in the plane thereof, a high voltage source with one terminal connected to said flat electrode and switching means connected to the other terminal of said source, and adapted to connect said other terminal selectively to one or the other of said electrodes whereby to cause relative movement of an ionizable medium surrounding said electrodes selectively in one direction or the other depending on which of said electrodes is connected to said source.

5. Thrust producing apparatus comprising a group of flat electrodes each having a continuous edge, said flat electrodes being supported in spaced parallel relation and electrically interconnected, a group of elongated electrodes, each supported in fixed spaced relationship with respect to one of said edges, said electrodes being electrically interconnected and a high voltage source con-

nected across said groups of electrodes whereby to cause relative movement of an ionizable dielectric medium surrounding said electrodes.

6. Thrust producing apparatus comprising a disc-shaped electrode, an electrode arranged in an arc of less than 180° spaced from an edge of said disc electrode and substantially in the plane thereof and power means connected across said electrodes to impress a high voltage potential thereon whereby to cause relative movement of an ionizable dielectric medium surrounding said electrodes.

7. Apparatus for producing relative motion between a vehicle body and the surrounding medium comprising a vehicle body having a conducting nose portion, a central portion and a conducting afterbody having a relatively large surface area and a source of high voltage connected between said conducting nose portion and said conducting afterbody whereby the application of said voltage to said conducting nose portion and said conducting afterbody cause relative motion between said body and the surrounding medium.

8. Apparatus for producing relative motion between a body and the surrounding medium comprising a body portion, an elongated electrode mounted on and insulated from said body portion and a source of potential having its positive terminal connected to said electrode and its negative terminal connected to said body portion whereby said electrical potential is converted directly into a motive force.

9. Apparatus for producing relative motion between a body and the surrounding medium in accordance with claim 8 wherein said body portion includes an electrical conducting surface and wherein said electrode is mounted in fixed spacial relation to said body.

10. Apparatus for producing relative motion between a body and the surrounding medium in accordance with claim 8 further comprising a rotatable arm, said body being connected to said rotatable arm.

11. Apparatus for producing relative motion between a first and a second body and the surrounding medium comprising a first conductive body portion, a second conductive body portion, each of said body portions having a relatively large surface area and having an elongated electrode mounted thereon in fixed spacial relationship, a source of potential connected to each of said bodies and each of said electrodes and means for supporting said bodies in rotatable relationship.

12. Apparatus for producing relative motion in accordance with claim 11 wherein said supporting means comprises an arm, and a support for said arm, said first and said second bodies being connected to the ends of said arm.

References Cited in the file of this patent

UNITED STATES PATENTS

1,974,483 Brown ---------------- Sept. 25, 1934

April 28, 1964 A. P. DE SEVERSKY 3,130,945

IONOCRAFT

Filed Aug. 31, 1959 6 Sheets-Sheet 1

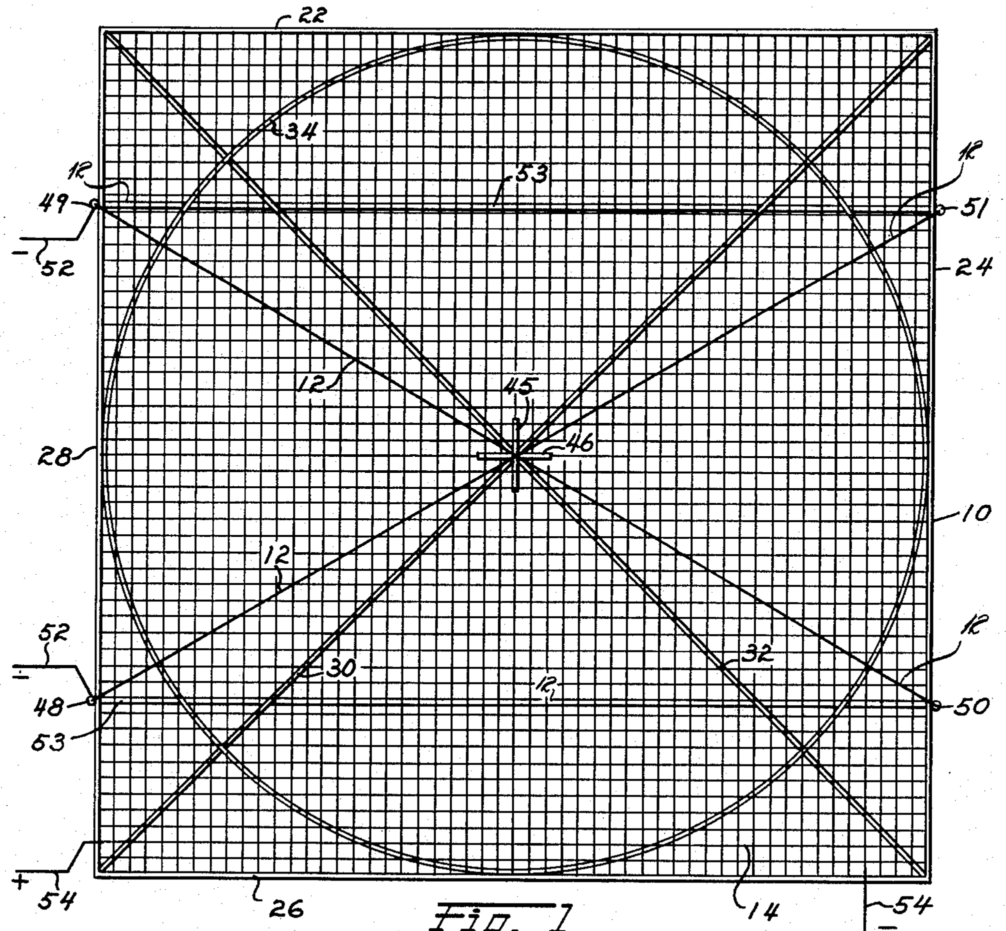
22
34
12
49
53
12
51
−
52
24
12
45
46
28
10
12
52
−
32
12
30
48
12
50
53
+
54
26
14
54
−

Fig. 1

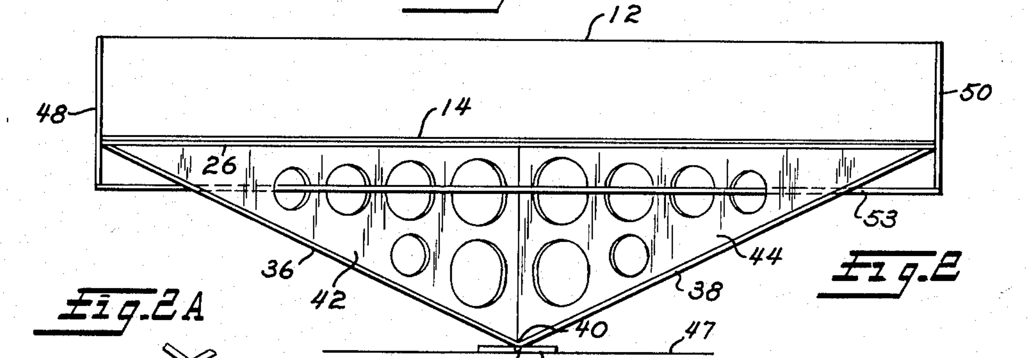
12
48
14
50
26
53
36
42
44
38
40
47
45
46

Fig. 2

Fig. 2A

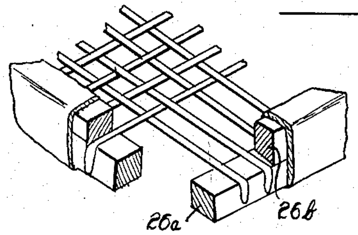

INVENTOR
ALEXANDER P. DE SEVERSKY

BY Strauch, Nolan + Neale
ATTORNEYS

April 28, 1964 A. P. DE SEVERSKY **3,130,945**

IONOCRAFT

Filed Aug. 31, 1959 6 Sheets-Sheet 2

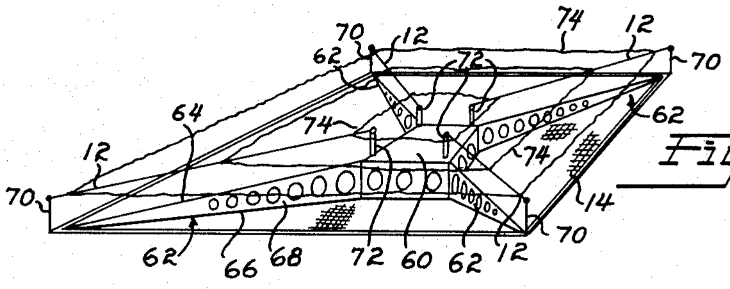
70
62
12
74
12
70
72
64
74
62
12
74
14
70
70
62
66
68
72
60
62
12
70

Fig. 3

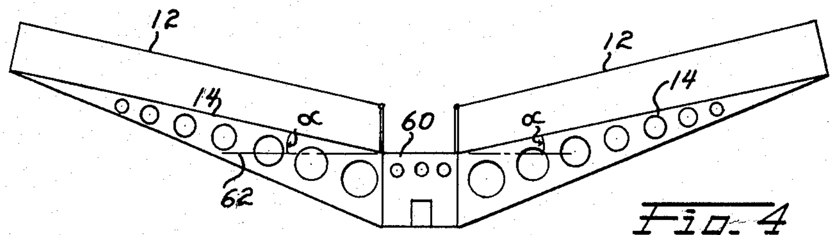
12
12
14
14
α
60
α
62

Fig. 4

Fig. 19

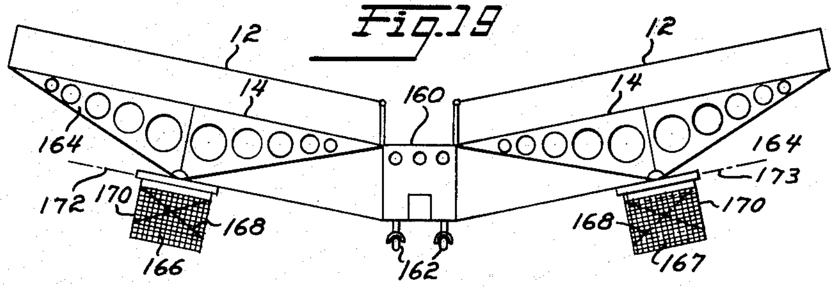
12
12
14
160
14
164
164
173
172
170
168
168
170
166
162
167

Fig. 21

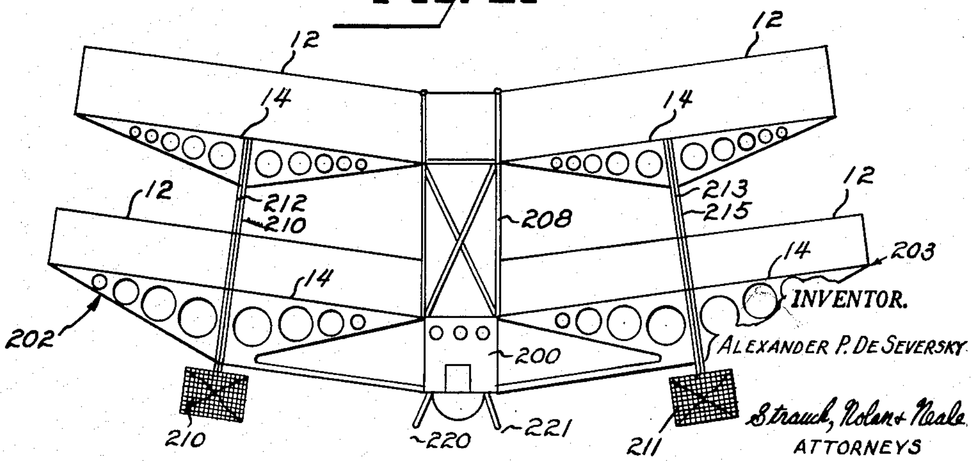
12
12
14
14
12
212
210
208
213
215
12
14
14
203
202
200
210
220
221
211

INVENTOR.

ALEXANDER P. DE SEVERSKY

Strauch, Nolan & Neale
ATTORNEYS

April 28, 1964 A. P. DE SEVERSKY 3,130,945

IONOCRAFT

Filed Aug. 31, 1959 6 Sheets-Sheet 3

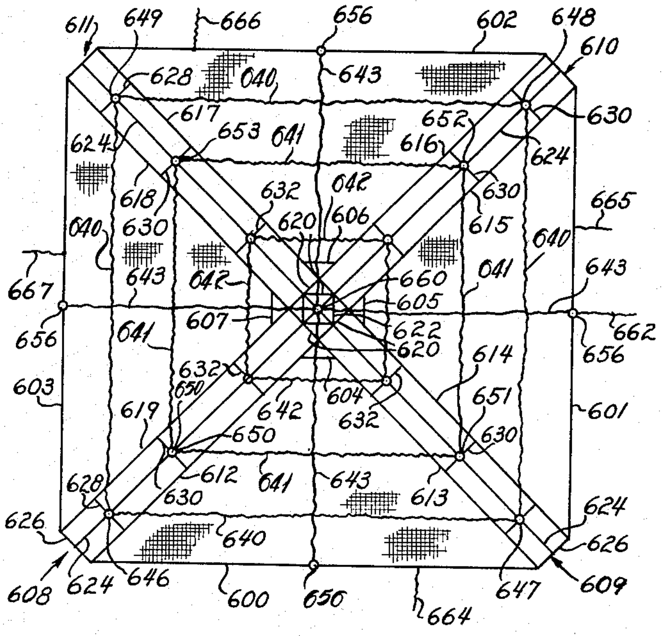
611
649
666
656
602
648
610
628
640
643
630
617
653
641
652
616
624
624
618
632
642
606
630
665
640
630
620
615
640
643
660
641
643
642
605
667
607
622
656
641
620
662
614
656
632
650
603
604
651
619
642
632
601
650
630
628
612
641
643
613
630
624
626
640
626
646
647
609
608
624
600
656
664

Fig. 6

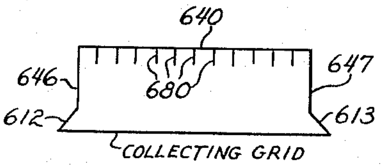
640
646
680
647
612
613
COLLECTING GRID

Fig. 9

Fig. 7

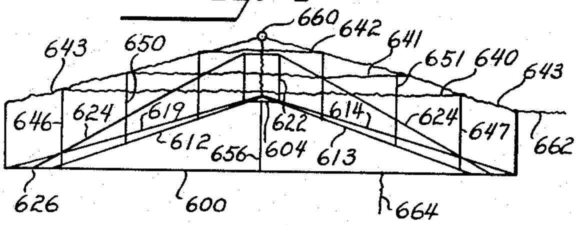
660
642
641
643
650
651
640
643
646
624
619
614
624
647
662
612
622
656
604
613
626
600
664

Fig. 8

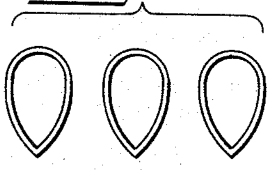

INVENTOR
ALEXANDER P. DE SEVERSKY

BY Strauch, Nolan & Neale

ATTORNEYS

April 28, 1964 A. P. DE SEVERSKY 3,130,945

IONOCRAFT

Filed Aug. 31, 1959

6 Sheets-Sheet 4

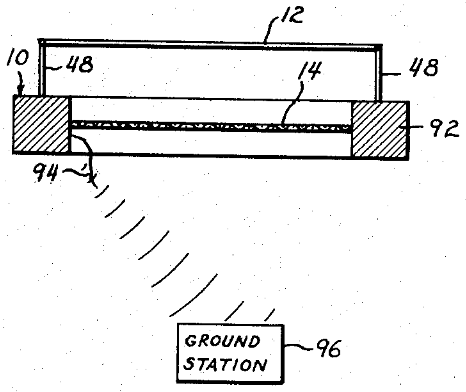
12
10
48
14
48
92
94
GROUND STATION
96

Fig. 10

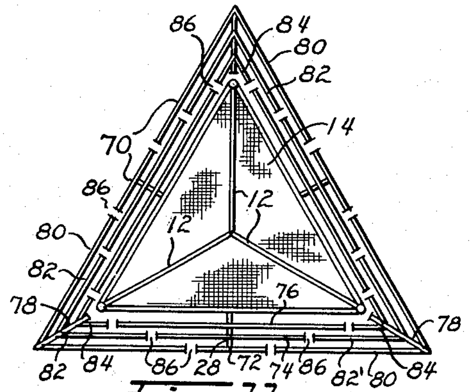
86
84
80
82
14
70
86
12
12
80
82
76
78
78
82
84
86
28
72
74
86
82
80
84

Fig. 11

Fig. 14

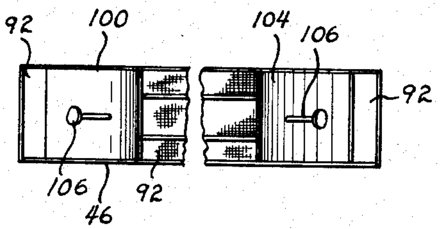
92
100
104
106
92
106
46
92

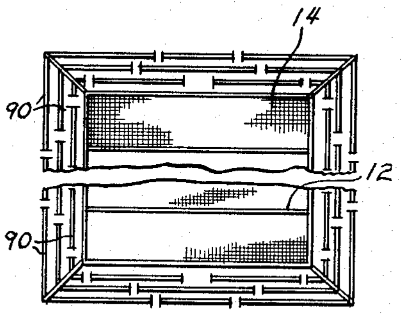

Fig. 12

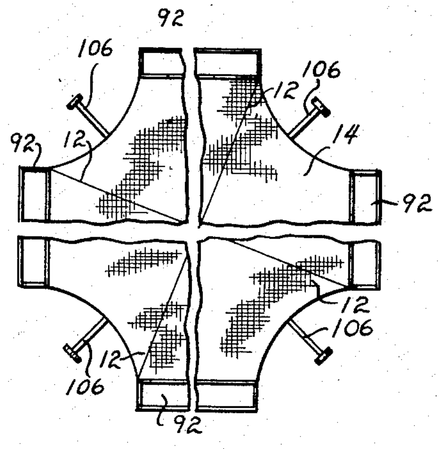
92
106
106
12
92
12
14
92
12
106
12
106
92

Fig. 13

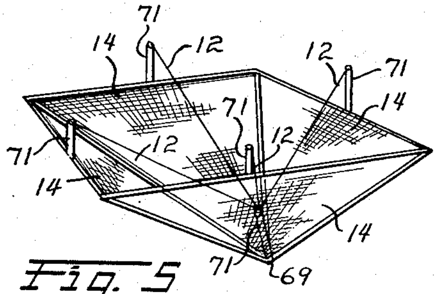
71
14
12
12
71
14
71
71
12
12
14
14
71
69

Fig. 5

INVENTOR
ALEXANDER P. DE SEVERSKY

BY Strauch, Nolan Neale

ATTORNEYS

April 28, 1964 A. P. DE SEVERSKY **3,130,945**

IONOCRAFT

Filed Aug. 31, 1959 6 Sheets-Sheet 5

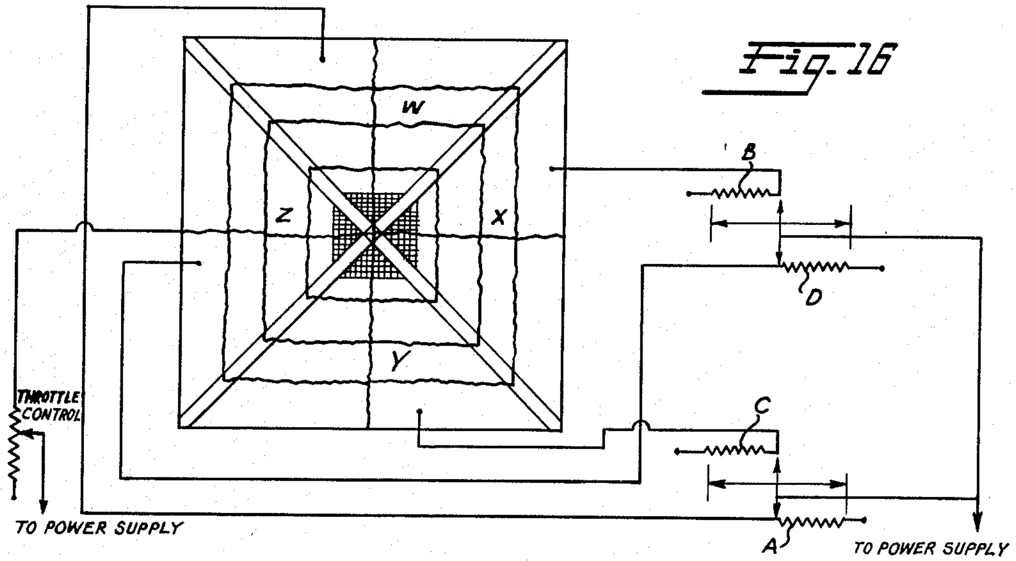
W
Z
X
Y
B
D
C
A
THROTTLE CONTROL
TO POWER SUPPLY
TO POWER SUPPLY

Fig. 16

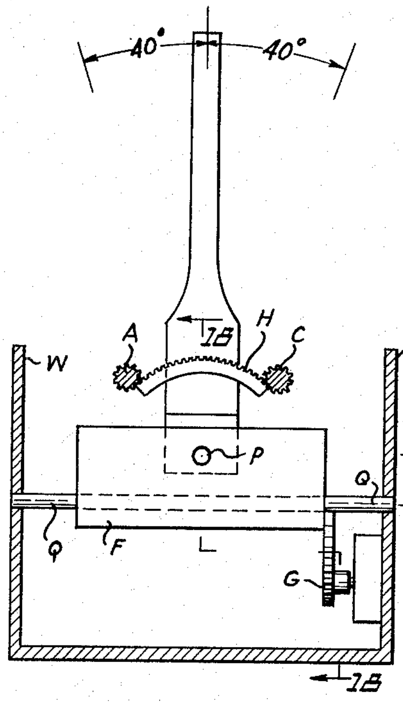
40°
40°
A
18
H
C
W
W
P
X
Y
Q
Q
F
L
G
18

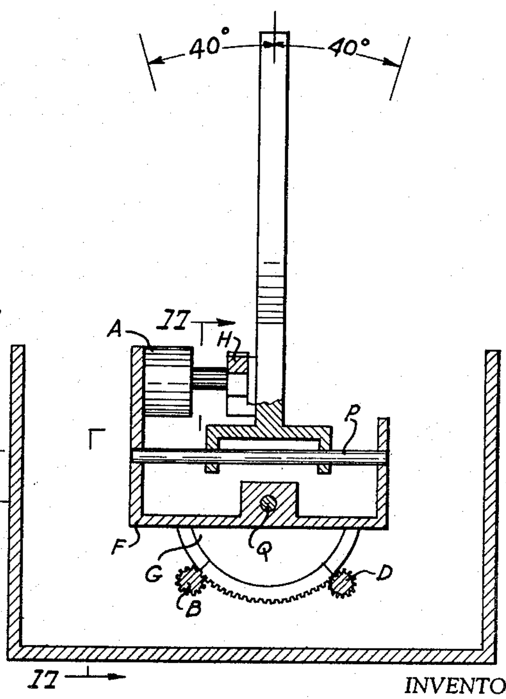
40°
40°
A
17
H
P
F
G
Q
B
D
17

Fig. 17

Fig. 18

INVENTOR
ALEXANDER P. DE SEVERSKY

BY Strauch, Nolan + Neale

ATTORNEYS

April 28, 1964 A. P. DE SEVERSKY 3,130,945

IONOCRAFT

Filed Aug. 31, 1959

6 Sheets-Sheet 6

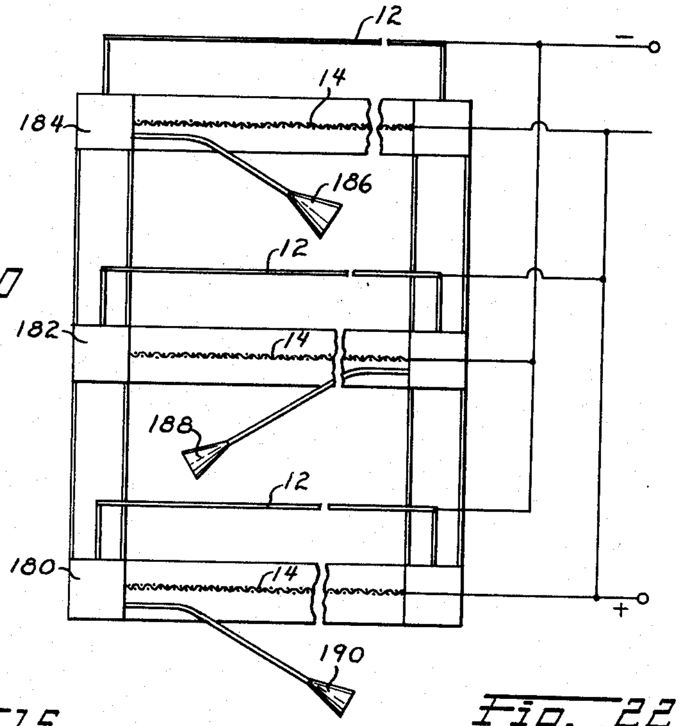
12
−
14
184
186
12
182
14
188
12
180
14
190
+

Fig. 20

Fig. 15

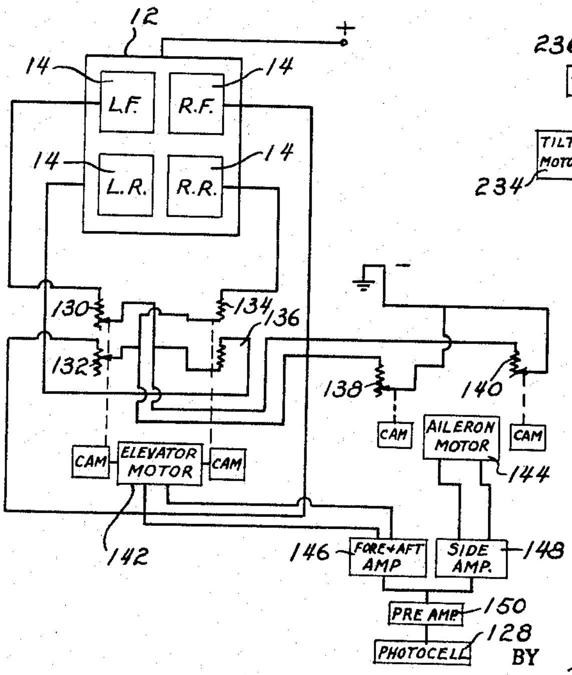
12
+
14
L.F.
R.F.
14
14
L.R.
R.R.
14
130
134
136
132
−
138
140
CAM
ELEVATOR MOTOR
CAM
CAM
AILERON MOTOR
CAM
144
142
146
FORE+AFT AMP
SIDE AMP.
148
PRE AMP
150
PHOTOCELL
128

Fig. 22

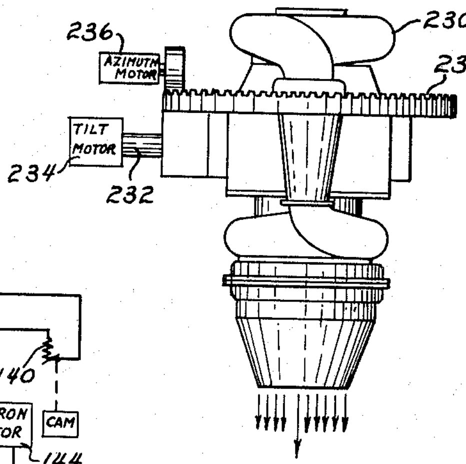
236
AZIMUTH MOTOR
230
238
TILT MOTOR
234
232

INVENTOR
ALEXANDER P. DE SEVERSKY

BY Strauch, Nolan + Neale

ATTORNEYS

3,130,945
Patented Apr. 28, 1964

3,130,945

IONOCRAFT

Alexander P. de Seversky, New York, N.Y., assignor to Electronatom Corporation, New York, N.Y., a corporation of New York

Filed Aug. 31, 1959, Ser. No. 837,150
29 Claims. (Cl. 244—62)

This invention relates to improved heavier-than-air aircraft, and more specifically to structures which are capable of either hovering or moving in any direction at high altitudes by means of ionic discharge.

The present invention is an improvement over well known electrostatic generation of winds used in a novel manner to supply propulsion and sustainance forces for a heavier-than-air aircraft. Crafts of the types herein disclosed having effective areas of several square feet have been successfully flown and contemplated platforms will inherently be of large size since the lift force is proportionate to the area through which large quantities or masses of air are accelerated downwardly from discharge electrodes to collection electrodes, the latter being a meshed-screen, bars, strips or any other structure that provides maximum collecting electrode area with perforations, slots or other types of opening to allow the air to pass through with a minimum of drag. Such a craft will be referred to in this application as an Ionocraft.

Such Ionocraft may serve as platforms which would be stationed above the earth for long periods of time and serve other purposes as will be explained below. The output power from microwave generators, such as magnetrons, coupled with high power capacity amplifier tubes may be beamed to the Ionocraft while airborne or the craft may carry its own power supply.

A principal object of the present invention is to provide a novel Ionocraft with space provided by the structure, preferably at the center of the craft, for installation of electronic equipment, and for the power plant, and crew where used.

Another object is to provide a novel Ionocraft construction wherein lightweight reinforcing members are provided to form a structure sufficiently rigid to cope with the dynamic and static loads and to maintain a desired distance between discharge emitting wires and the collecting grid.

Still another object resides in the novel configuration and arrangement of the emitting wires to assure uniform spacing from the collecting grid and to provide a maximum number of ionized particles for producing the desired lift.

A further object is to provide an improved Ionocraft of the foregoing type wherein some structural formation such as dihedral is provided for stabilizing the craft during flight. The dihedral may be positive or negative depending upon whether the hovering flight or horizontal motion of aircraft is a primary consideration of performance of the craft. A multiple deck structure may be used where desired to increase the lifting force, and dihedral may be provided in two or more angularly related directions to provide stability in all directions. A conical shape with the apex or nadir at the top or bottom center may also be advantageously used.

Still another object is to provide auxiliary ionic discharge structures mounted for rotational movement which are oriented to provide a horizontal propelling force and steering forces which can change the direction of the craft. By mounting such auxiliary structures to turn about a vertical axis, the craft can be made to turn in a horizontal plane about a vertical axis passing through the craft to thereby provide a scanning or target searching apparatus. A similar scanning motion can be achieved

by mounting the auxiliary structures to turn about a horizontal axis.

A further object of this invention resides in the provision of a novel stick control using variable electrical impedances for control of the posture and for maneuvering the craft through variation of the voltage applied to different portions of the craft.

A second principal object of the present invention is to provide a combination Ionocraft and antenna system for radio frequency energy wherein the structure of the Ionocraft is so arranged as to serve in whole or in part as a structure of an efficient electromagnetic antenna system. In accordance with this object of the invention, the device contains one or more antennas that may be used for communication signal transmission, for detection, tracking and/or identification and for eventual destruction through collision of oncoming airborne or space vehicles or missiles and the like. The Ionocraft structure may be used, for example, as the main antenna element, as a series of directing or reflecting elements or as a parasitic element and may be shaped to provide arrays parabolas, corner reflectors, horns or lenses and be adapted to transmit a single or complete spectrum of frequencies from the extremely low frequencies to the highest frequencies including infrared.

Another object of this invention is to provide a combination antenna-Ionocraft with scanning means for detecting and/or tracking airborne vehicles or missiles. Such combination may also include suitable servo-control and other conventional equipment either on the Ionocraft or at a nearby ground station for causing the Ionocraft to "lock-on" automatically and/or be guided into the path of an "oncoming" vehicle or missile.

A further object is to provide an antenna which constantly locks on a radiation beam, such as a microwave or light beam for example, projected from the ground or from an aircraft in flight to change the position of the Ionocraft in flight.

These and other objects of the invention will become more fully apparent from the claims, and from the specification when read in conjunction with the appended drawings wherein:

FIGURES 1 and 2 (Sheet 1) are top plan and elevation views of the basic structure of an Ionocraft made in accordance with the present invention;

FIGURE 2*a* is an enlarged pictorial view of a portion of the structure showing how the grid wires are connected to the frame members;

FIGURE 3 (Sheet 2) is a pictorial view of a modified form of basic structure;

FIGURE 4 is a view in elevation of an embodiment similar to that shown in FIGURE 3 which is equipped with dihedral;

FIGURE 5 (Sheet 4) is a schematic view of a craft equipped with dihedral in two perpendicular directions;

FIGURES 6 and 7 (Sheet 3) are top plan and side elevation views respectively of a further embodiment of the present invention which is equipped with negative dihedral;

FIGURE 8 is a cross section of collecting grid structural members which may be used in lieu of the wire mesh;

FIGURE 9 is a view in elevation of an emitting wire having short wires suspended from the main wire to provide a point source for ion emission;

FIGURE 10 (Sheet 4) is a diagrammatic view in elevation of an Ionocraft in accordance with this invention;

FIGURES 11 and 12 are top plan views of two embodiments of the Ionocraft having a side elevation view as illustrated in FIGURE 10;

FIGURES 13 and 14 are top plan and elevation views of a further embodiment of this invention;

FIGURE 15 (Sheet 6) is a schematic diagram of a control circuit for causing the Ionocraft to lock-on and follow a radiation source at a ground station;

FIGURE 16 (Sheet 5) is a plan view, partly diagrammatic illustrating a control system for the craft of the present invention;

FIGURES 17 and 18 are side elevation views in section of a novel control stick box and assembly to permit steering and guiding of the craft by the system illustrated in FIGURE 16;

FIGURE 19 (Sheet 2) is a view in elevation of a craft having two vertical grid structure assemblies for controlling horizontal movement;

FIGURES 20 (Sheet 5) and 21 (Sheet 2) are side views in elevation of different embodiments each having several horizontal grid structures stacked one on top of the other; and

FIGURE 22 is a diagrammatic view of a gas turbine engine and mounting which are adapted for use with craft of the present invention.

Referring now to the drawings, FIGURES 1 and 2 (Sheet 1) are plan and elevation views of a typical basic embodiment of my improved Ionocraft **10**. The Ionocraft proper comprises a plurality of emitting electrode wires **12** mounted above and in a plane substantially parallel to the collecting electrode grid **14** which may be composed of a meshed screen, bars, strips or any other structure that provides maximum effecting collecting electrode area with perforations, slots or other types of opening to allow the air to pass through with a minimum of drag. A plurality of hollow, lightweight rods or bars of conductive material or crossed wires forming a mesh which is open to pass air downwardly, but with the wires sufficiently closely spaced to effectively neutralize the charged ions which pass from emitting electrode wires **12** are preferred structures. A high D.C. voltage is applied between emitting electrode **12** and collecting electrode **14**; one pole or terminal of the high voltage generator is connected to the emitting electrode **12** and the opposite pole or terminal of the same generator is connected to the collecting grid electrode **14**, thus creating a high potential field between the electrodes.

In this form of improved Ionocraft, a basic structure sufficiently rigid to cope with the dynamic and static loads and to maintain a desired uniform distance between discharge emitting wires **12** and the collecting grid **14** is utilized comprising an outer square or rectangular frame composed of members **22**, **24**, **26** and **28**. Diagonal frame members **30** and **32** extend between opposite corners of the rectangular frame and a circular frame member **34** is fixed tangentially to the midportions of the frame members. Said frame members are coplanar and collecting electrode wires **14** are interwoven, as with a loom, to form a closely meshed wire screen and supported from frame members **22**, **24**, **26** and **28**. The ends of each wire are wrapped over and glued to the lower half **26*a*** of the frame member and then cut off as shown in FIGURE 2A. The upper half **26*b*** of the frame member is then secured in position as with glue. A considerable improvement in lifting force was achieved when the frame members and cut ends of the grid wires were covered with an aluminum foil.

Four lightweight rigid structural members, **36** and **38**, of which two show in FIGURE 2, are mounted beneath the plane of collecting grid **14** in the vertical planes to diagonal members **30** and **32**. Members **36** and **38** meet in a common junction **40** at the center of the Ionocraft. Four perforate lightweight rigid metal sheets or foils **42** and **44** of aluminum or the like, of which only two show in FIGURE 2, are mounted between diagonal members **30**, **32**, **36** and **38**. These foils provide additional stabilization against tilting by guiding the air flow vertically along the surfaces of the foils and have been found to provide an increase in lift which more than compensates for their weight. Beneath junction **40**, a pair of crossed support members **45** and **46** are provided to serve as a landing support to hold the craft with the collecting grid **14** above the ground-supporting surface **47** when landed.

The outer ends of emitting wires **12** are supported from masts **48**, **49**, **50** and **51** of insulating material mounted on opposite sides of the craft. In this embodiment, emitting electrode wires **12** pass diagonally across the craft and cross each other near the center. One terminal of a high voltage D.C. potential is connected to leads **52** which are connected to masts **48** and **49**.

The lower edges of masts **48** and **50** and of masts **49** and **51** are connected together by tension member **53** (FIGURE 2) such as a lightweight cable to hold the masts in their vertical position by providing a force to balance against the tension force of emitting wires **12**.

Suitable lead-in wires **54** are provided for connection between collecting grid **14** and the other terminal of the power supply, and are preferably at ground potential. Variable impedances, such as variable width spark gaps which serve to reduce the applied voltage, are provided in lead-in wires **54** for control of voltage between emitting wires **12** and collecting grid **14** to thereby control the vertical movement of the craft.

An actual embodiment built in accordance with the foregoing description which lifted itself into a self-sustaining flight had a collecting grid surface area of approximately 150 square inches and the space between the collecting grid and the emitting wires **12** was approximately 2 inches. With a craft having the foregoing dimensions, a voltage of 20 kv. and a current of approximately 0.5 milliampere was sufficient to make the craft more than self-sustaining. The total weight of the structure was approximately 5 grams. Other craft having the space between the collecting grid and emitting wires of 5 inches have been successfully flown. Such craft require voltages of the order of 50 to 60 kv. Where the grid area is about 7 or 8 square feet, currents of the order of 2 milliamperes exist. Variations in humidity and air pressure cause variations in the current drawn and in the lifting efficiency.

The lifting capability of the craft was found to increase as the diameter of the grid wires is increased. Crafts were tested with wire diameters of 2, 5, 8 and 12 mils for the collecting grid. With wire diameters of 8 mils or more, the current requirement to provide the same total lifting force shows a detectable decrease thereby indicating a higher efficiency. Hollow tubular conductors having an outer diameter of one-quarter inch also give substantially the same lift force and efficiency as the 8 and 12 mil wire diameters.

A modification of the foregoing structure is shown in FIGURE 3 (Sheet 2) wherein a central compartment section **60** is provided in the center of a surrounding large area collecting grid **14**. A plurality of rigid support sections **62**, each comprising an upper member **64**, a lower member **66**, and an intermediate foil **68** extend from the corners of the central section **60** to the periphery of the framework surrounding the collecting grid **14**.

Near each of the corners of the outer periphery of collecting grid **14** a mast **70** made of insulating material is provided which supports the outer end of emitting wires **12**. A second group of inner support masts **72** mounted on central section **60** provide support for the inner ends of emitting wires **12**.

In this embodiment, the central compartment **60** is adapted to house electronic equipment and the power plant and crew where used.

In practice, it has been found desirable to increase the lengths of emitting electrode wires by adding a series of wires **74** which are supported on the main emitting wires **12** and which are parallel to each other and at a distance approximately equal to the distance of the emitting wires from the collecting electrode. The outermost wire is positioned inwardly about one-half the distance between the parallel wires (i.e., from 1 to 3 inches) from the outer frame member on the collecting grid to take full advantage of all the ionized particles which are produced. The radially directed emitting wires **12** are used to electrically connect the non-intersecting wires **74** together. However, the emitting wires **12** should be fewer and much less closely spaced than collecting grid wires **14** in order to avoid electrical symmetry. If the configuration of the emitting electrode wires **12** and the collecting electrode wires **14** are identical, no lifting force is provided.

A further embodiment is shown in FIGURE 4 (Sheet 2) which is identical with the form shown in FIGURE 3 except that the structure is equipped with positive dihedral for greater stability. Center section **60** is used as a center load carrying or cabin section and the rigid support sections are attached thereto so as to tilt upwardly to form a small angle α. Collecting grids **14** and their associated emitting electrodes **12** on opposite sides of center section **60** are thus angularly related.

This particular craft, because of its horizontal symmetry, is well adapted to be equipped with dihedral in the fore and aft direction as well as in the lateral direction. FIGURE 5 (Sheet 4) represents in an exaggerated schematic form an apparatus of this type. In FIGURE 5, the central section as shown in FIGURE 4 has been omitted and the four collecting grids **14** are of a triangular shape with the inverted apex or nadir **69** of the system at the bottom and center of the apparatus. Separate emitting wires **12** are mounted from masts **71** supported centrally of the side edges and at the nadir. Each of the four collecting grids may be insulated from each other by a gap or insulating material and variable resistance incorporated in their lead-in connections (not shown) to the power supply. By independently varying the resistance of the collecting grids the craft, which is here assumed rigid, may be tilted in any direction.

FIGURES 6 and 7 (Sheet 3) are top plan and side elevation views respectively of a further embodiment which has a negative dihedral. In this embodiment, the collecting grid frame comprises four outer peripheral lightweight wooden or metal members **600**, **601**, **602** and **603** which are mounted in a lower plane and four inner members **604**, **605**, **606** and **607** which are parallel to the respective outer members but in a plane higher than the plane containing the outer members. In an embodiment where the outer peripheral members were three feet long, the vertical distance between the planes carrying the inner and outer members was four inches. This negative dihedral has been found to provide greater stability during hovering flight than the positive dihedral though the positive dihedral appears to provide equally good stability for horizontal flight.

The collecting grid is divided into four equal areas by diagonally oriented frame assemblies **608**, **609**, **610** and **611**. The collecting grid area visible in FIGURE 7 is bounded by rigid frame members **612** and **613** of diagonal frame assemblies **608** and **609** respectively and inner and outer frame members **604** and **600**. The collecting grid, as pointed out above, may be a crossed grid of wires. The other three collecting grid areas are of identical size and construction.

Inside of inner frame members **604**, **605**, **606** and **607**, no collecting grid screen is provided and the space may be left open or if desired, covered with a lightweight foil of insulating or conducting material. This air-tight foil forms a pocket under which a pressure appears to build up to provide added lift. The insulation material is preferred since this does not interfere with the electrical isolation between the four quadrants of the collecting grid which, as will be pointed out below, are used for guiding and/or propelling the craft.

Diagonal frame assembly **608** contains four cross braces **626**, **628**, **630** and **632** between frame members **612** and **613**. The cross braces are made of an insulating material such as wood to thereby insulate each of the four grid sections from one another. Frame members **612** and **613** intersect and are secured together above and inwardly behind member **604**. Members **614** and **615** also intersect and are secured together, as do members **616** and **617** and members **618** and **619**. These points of intersection are joined together by four struts **620** shown in FIGURE 6. Secured to the centers of each of struts **620** is a four-sided chimney **622**, each of the sides being flat sheets of a lightweight insulating material such as wood.

A center frame member **624** is mounted between the center of cross brace **626** and the top of chimney **622** along each of the diagonal frame assemblies. This construction gives adequate rigidity to prevent warpage of the collecting grid frame assembly.

The emitting wires are illustrated diagrammatically as waving lines and make up a pattern of three parallel wires **640**, **641** and **642** and one transverse wire **643** across each grid area. Four supporting masts **646**, **647**, **648** and **649** are mounted on cross braces **628** and secured to center frame member **624** in each of the four diagonal frame assemblies **608**, **609**, **610** and **611**. Emitting wire **640** is supported on the upper end of each of masts **646**, **647**, **648** and **649** with sufficient tautness to be substantially equidistant from the collecting grid at all points.

Four supporting masts **650**, **651**, **652** and **653** are mounted to cross braces **630** and center frame members **624** in each of the four diagonal frame assemblies for supporting emitting wire **641**. Four additional masts (not numbered) are mounted to cross braces **632** and center frame members **624** to similarly support emitting wire **642**.

At the mid-points of each of the outer frame members **600**, **601**, **602** and **603**, masts **656** are mounted to support the outer end of emitting wires **643**, which extend under and in electrical contact with each of emitting wires **640**, **641** and **642** to a center mast **660** which is suitably mounted to the top of chimney **622**.

One electrical terminal **662** for the emitting wires is shown on the right side of the craft of FIGURES 6 and 7. Four individual electrical terminals **664**, **665**, **666** and **667** are provided for each of the four grid sections. If it is not desired to control the posture and movement of the craft by the four separate sections, collecting grid terminals **664**, **665**, **666** and **667** may all be connected together.

Also, it is obvious that the four electrically separate sections could be achieved by using four insulated emitting wire sections, either with the four separate collecting grid sections or with all the collecting grid sections connected together.

The foregoing craft weighed about 100 grams and with a 5 inch spacing between the emitting wires and collecting grid, was self-sustaining with a voltage of about 50 to 60 kv. and a current on the order of 2 milliamperes.

Instead of using a crossed wire mesh construction for the collecting grid as shown in detail in FIGURE 1, it has been found that tubes of conductive material having an outer diameter of about one-quarter inch are equally as good. Such tubes may be made of aluminum foil wrapped around paper or may be hollow lightweight aluminum tubing or of a similar construction. For example, material such as an air-tight nylon base fabric having an evaporated metallic coating of for example aluminum may be fabricated in the form of tubes having a wall thickness of less than 1 mil and be adapted to be inflated with air or an inert gas to form a hollow lightweight tubular member. The cross section may be circular, oval or the like; a tear drop shape as illustrated in FIGURE 8 (Sheet 3) is a preferred configuration since air flow across the tapering lower edge provides additional lift. For the craft configuration as shown in FIGURES 6 and 7, the inflated tubes of FIGURE 8 are mounted parallel to each other and to the outer and inner frame members **600** and **604**, or to their corresponding members in each of the other collecting grid sections, with their ends secured to the diagonal frame assemblies **608**, **609**, **610** and **611**.

Other emitting electrode constructions may also be used. For example, emitting wires **640–644** may have suspended from them a plurality of short wires **680** as shown in FIGURE 9 (Sheet 3) which provide a point of discharge rather than a line of discharge to thereby increase the efficiency of ionization. In FIGURE 9, only emitting wire **640** and its supporting masts **646** and **647** from the embodiment shown in FIGURES 6 and 7 are illustrated. It is to be understod that all of the emitting wires may be of similar construction to that illustrated in FIGURE 9. Each of wires **680** is about 1 to 3 or more inches in length and separated at least one inch apart. The lower ends of wires **680** are kept at a uniform distance from the collecting grid. This construction may offer some pre-ionization, though measurements show this emitting electrode construction to be about comparable to the use of plain wire as the emitting electrode.

FIGURE 10 (Sheet 4) illustrates an elevation view, and FIGURES 11 and 12 illustrate plan views of modified triangular and rectangular shaped Ionocrafts **10** respectively. The craft of FIGURE 11 is triangular in configuration and is provided with emitting wires **12** suitably supported from masts **48** as illustrated. In practice additional emitting wires may be used. Collecting grid **14** extends over a large area beneath emitting wires **12** and may be formed of crossed wires as diagrammatically illustrated.

The electromagnetic energy antenna carried by the foregoing Ionocraft embodiments may comprise a series of generally horizontal, parallel conducting elements or dipoles **70** arranged along the basic side structure on which the wires **12** and **14** of the craft are attached. Dipoles **70** may be of differing length so that the antenna provided may receive or transmit several different frequencies. For frequencies of the order of 10 megacycles, for example, several dipoles **72**, **74** and **76** may be arranged as a tuned array, such as the yagi array, with one or more dipoles **72** serving as a director, dipole **74** serving as the main antenna element and dipole **76** serving as a reflector. Such antenna is highly directional and with an Ionocraft of triangular configuration, the antenna may be used with signal transmission in three separate directions simultaneously.

The antenna wires **70–76** may be small diameter rods of a conductive material such as aluminum, supported on lightweight rods or bars **78** of either a conducting or insulating material, as dictated by conventional antenna construction techniques. Additional antenna elements **80**, **82** and **84** may be present as metal rods or wires separated electrically from the adjacent antenna elements by insulators **86** of a suitable light material such as wood, plastic or the like, indicated on the drawing by spaces.

The various antenna elements **70–84** and insulators **86** may comprise a rigid frame forming the basic structure for the craft and inside of which the collecting grid **14** is supported and upon which the discharge electrodes **12** are mounted. The antenna elements **70–84** may be stacked vertically if desired to improve both the efficiency of the antenna and the rigidity of the basic structure. To the extent that the antenna elements may be galvanically connected together without interfering with the operation of the antenna in its conventional manner, the antenna elements are preferably connected to the same D.C. potential as collecting grid **14**. Thus, the antenna elements may also augment the operation of the Ionocraft by neutralizing charged ions which provide the propelling force.

In the rectangular embodiment of FIGURE 12, the several antenna dipoles **90** have different lengths so as to be equal to one half the wave length λ of the frequency being transmitted for an entire spectrum of frequencies having different wave lengths $\lambda_1, \lambda_2, \lambda_3 \ldots \lambda_n$. Since the length of a side of the Ionocraft may be several hundred feet or greater, such construction is ideally suited for communication systems, whether operating with high or low frequencies.

With either of the configurations of FIGURE 11 or FIGURE 12, the view in elevation will be substantially as illustrated in FIGURE 10 where the particular antenna structure is indicated schematically and designated by reference numeral **92**. A ground station antenna which is indicated diagrammatically at **94** on FIGURE 10 may be provided for directing the signals downwardly to the ground station. Antenna **94** may be of any desired conventional type and connected on the Ionocraft to the main antenna structure **92** by a suitable transmission line such as coaxial cable, twin lead lines or hollow pipe waveguide, depending upon the particular frequencies utilized. Amplifiers or frequency converters may also be provided in the transmission line where signal strength is weak. The amplifiers and/or frequency converters may be powered by well known self-contained batteries or by the power supply unit for the Ionocraft (not shown).

Referring now to FIGURES 13 and 14 (Sheet 4), a further embodiment of the invention is illustrated which has a plurality of side sections, four of which are shown curved. The contour of the curves may be parabolic or of any other shape as is conventionally used for antennas in high frequency systems such as radar or the like. In this embodiment, an outside frame of lightweight rigid members **96**, **98**, **100** and **102** is provided to define the contour of the antenna shape. Lightweight wires or rods **104** extend between members **95** and **102** to serve as part of the antenna. Lightweight sheet metal of a material such as aluminum may be used in lieu of wires **104** for the reflector surface if desired.

A plurality of horns **106** are illustrated in the drawings to effect simultaneous radar scanning through 360°. By oscillating the illustrated Ionocraft about its vertical center line through an angle of only 45° on each side of a center position, complete 360° scanning may be effected. Alternatively, the Ionocraft may be rotated continuously about its vertical center line and 360° scanning effected by one or more antennas. Such scanning may be effected by warped corners, reactive or propeller blasts of auxiliary power plant, or by auxiliary grids which are mounted for movement relative to the main lifting grid as will be described below. Scanning may be effected by other means as will become apparent from the following description. In lieu of or supplemental to some of the microwave antennas **106**, antenna reflectors for infrared detectors may be carried on the Ionocraft. Such antennas serve to collect the infrared energy over a large area and focus such energy on a small infrared detector, and they may be of any conventional construction. The basic structure **10** between spaced antennas may contain such equipment to transmit via wireless signal channels to the ground station through ground station antenna **94**, signals corresponding to the electromagnetic and radio frequency signals received.

Horizontal movement of the craft may be effected by the principles set forth in Serial No. 760,390 of Glenn E. Hagen filed September 11, 1958, by tilting the craft downwardly in the forward direction whereby the ionic propulsion force provides a horizontal force component to cause the craft to move in a horizontal direction. Tilting of the craft may easily be effected through variation of the voltage between emitting electrodes **12** and collecting grid electrode **14**. For example, by electrically separating the craft into four sections of substantially equal size as illustrated in FIGURE 15 (Sheet 6), the voltage applied to two of the adjacent sections can be reduced by adding resistance in series with the current path and this will cause the lift produced by these two sections to decrease relative to the lift produced by the two other sections. Thus, horizontal movement of the craft may easily be controlled from the ground station.

For manual control of the posture and flight movement of the craft of the present invention, it has been found desirable to provide a control stick assembly which functions similar to that familiar to persons flying other types of aircraft. The control stick must function in both the longitudinal and lateral directions simultaneously and independently. Variable control elements such as potentiometers and variable transformers (powerstats for instance) may be used for control of the present invention. The posture of the craft may be controlled by dividing the collecting grid into three or more electrically separate regions as illustrated by the embodiment shown in FIGURES 6 and 7 and by individually varying the electrical potential to each of the separate regions. The potential may be increased to act as an elevator or may be decreased to act as a spoiler, and the voltage may be increased on one side while being simultaneously decreased on the other side to increase the effectiveness of the control.

Also, the emitting wires may be divided into three or more electrically separate regions and the electrical potential individually varied to each separate region. Again the potential may be increased or decreased, and may be simultaneously increased in one region and decreased in the opposite region.

To change the voltage to an individual region of the craft, a separate power supply for each region may be provided and the variable control element for changing the output voltage may be adjusted to produce the desired voltage level. Where a single power supply is provided, variable resistances may be placed in the electrical conductors leading to the appropriate terminals on the craft. If the craft is normally airborne with resistance present in the conductors, then increased voltage can be supplied to one region of the craft by decreasing the resistance in the conductor connected to that region. A decreased voltage can be supplied similarly by increasing the amount of resistance, and combination of increased and decreased voltages may be supplied to opposite sides of the craft to increase response of the craft to the controls.

One of the more simple ways to utilize the power supplied to the craft, I prefer not to have extra resistance in the power supply circuit of the emitting wires during normal flight and to control the posture of the craft by individually adding resistance into the circuit connected to each individual region of the collector grid. Such method of control has been found to provide adequate control of the Ionocraft and a control stick assembly will be described which utilizes variable resistance elements which are conventionally known as potentiometers or rheostats.

In FIGURE 16 (Sheet 5) the collecting grid construction as shown in the preceding embodiments (see for example FIGURES 6 and 7 on Sheet 3) is illustrated with each of the four grid sections W, X, Y, and Z connected through a separate correspondingly designated potentiometer to one terminal of the power supply. The emitting wires shown diagrammatically as waving lines are connected through a throttle control potentiometer, which is used to control the maximum voltage applied between the emitting wires and all of the collecting grid sections. When this voltage exceeds a certain level but yet remains less than that which causes arcing, the craft will rise. The effect of potentiometers A, B, C and D is to controllably reduce the voltage between the emitting wires and any one or two specific grid sections to thereby reduce or subtract from the effectiveness of that portion of the craft in producing its lifting force. This then causes the craft to tilt downwardly in the direction of whichever of the grid sections has the reduced voltage.

Referring now to FIGURES 17 and 18 (Sheet 5) front and side elevations of the control stick are shown with the respective shafts of the four potentiometers labeled A, B, C and D. On each of these shafts spur gears (not shown) are provided to be driven by gear segments secured to the stick.

The control stick is mounted for pivotal movement about pin P having axis X and about pin Q having axis Y beneath, but in the same vertical plane as axis X. Pin Q is mounted with its ends in opposite side walls W of the control stick housing.

The entire stick assembly shown in FIGURES 17 and 18 is mounted for unitary movement in a plane perpendicular to the longitudinal axis Y of pin Q. This assembly comprises bracket F which has secured to one side face spur gear G which need have only a segment thereof with teeth to mate with the pinion gears on the shafts of potentiometers B and D. The housings for potentiometers B and D are mounted on housing walls W, and the center of the gear segment on gear G coincides with axis Y of pin Q.

The ends of pin P are mounted in opposite sides of bracket B to enable the control stick to rock in a plane perpendicular to the longitudinal axis X of pin P. The lower end of the control stick is bifurcated as shown in FIGURE 18 and adapted to pivot about pin P. Gear segment H, having its center at axis X of pin P, is secured to the control stick for driving pinions on the shafts of potentiometers A and C which are mounted on bracket F.

The foregoing construction permits the control stick to function both in a longitudinal direction and in the lateral direction simultaneously to function as an electrostatic spoiler in the sense that when the craft is airborne, the addition of resistance in the lead-in wire to a particular grid section spoils the lift of that section to thereby control the posture of the craft in flight.

In the described embodiment, stick movement was limited to about 40° by mechanical stops not shown. The pitch diameter of each gear segment G and H was about 6 inches and the pitch diameter of the pinion gears on the potentiometer shafts was about 1 inch. The potentiometer gear shafts were capable of rotating through 240°, and were spring biased to a zero resistance condition.

As is apparent from FIGURES 17 and 18, the position of the pinion gears for the four potentiometers A, B, C and D is at the exact ends of the corresponding gear segments so that when the control stick is in its illustrated vertical position, each potentiometer is rendered ineffective to add any resistance to any of the collecting grid sections. As the control stick is tilted, one of the potentiometer shafts is rotated and there is absolutely no possibility that the potentiometer to the opposite grid section can be made effective at the same time because the partial gear segment and the spring loaded potentiometer shafts are used. The length of each gear segment must be at least as large as the maximum angle through which the stick can be moved, and the pinion gears are preferably at the precise ends of the gear segments.

It was found that if the potentiometer shafts were not spring loaded, the gears would upon occasion rotate slightly so the teeth did not always mesh when the stick was moved in a direction so that the gear segment should have engaged the potentiometer pinion. By the manual control stick just described, adequate tilt of the craft is readily achieved.

The position of the craft in air may be remotely controlled from a ground station through wireless control systems which may be of any suitable known type. The horizontal position of the craft may also be controlled automatically.

For example, the position of the craft of the present invention may be automatically controlled in space through means of suitable centering or tracking apparatus operating on well known principles, such for example as are disclosed in U.S. Patent Nos. 2,513,367 to Scott or 2,604,601 to Menzel. In such tracking apparatus, one form of which is diagrammatically illustrated in FIGURE 15 (Sheet 6), a beam of electromagnetic energy, such as light or infrared, is centered on a suitable photocell **128** which generates control signals that are used to control variable impedances to reduce the voltage applied to various sections of the craft to thereby control the position of the craft in accordance with the position of the beam source at the ground station.

FIGURE 15 illustrates in detail suitable horizontal positioning control arrangement. The common grid electrode **14** is connected to the negative terminal of the power supply and the emitting wires **12** are electrically separated into four sections, viz. left front LF, left rear LR, right front RF and right rear RR. Each of these sections is connected through variable impedances **130**, **132**, **134** and **136** respectively of the elevator control unit and the variable impedances **138** and **140** of the aileron control unit to the positive terminal of the power supply. The elevator motor **142** drives the movable contacts on variable impedances **130**, **132**, **134** and **136** and the aileron motor **144** controls in a similar manner values of the impedances **138** and **140**. Each motor **142** and **144** may be driven by separate amplifiers **146** and **148** and pre-amp **150** in a manner as conventionally used in servo systems to position photocell unit **128** directly in alignment with a source of electromagnetic energy positioned on the ground.

Referring now to FIGURE 19 (Sheet 2), a craft having a central cabin **160** and equipped with dihedral is illustrated. The collecting grid **14** and emitting wire **12** construction may be similar to that described in connection with FIGURE 4 (Sheet 2) and be positioned on alternate sides of cabin **160**. Beneath cabin **160**, a suitable wheeled, skid or pontoon landing gear **162** may be provided.

Depending beneath frame members **164** and on opposite sides of cabin **160** are a pair of auxiliary grid assemblies **166** and **167** that are mounted to be operable in a generally vertical plane. Each auxiliary grid assembly **166** and **167** is provided with laterally spaced emitting wires **168** and a collecting grid within outer frame members **170** so that upon receipt of a suitable D.C. potential, a horizontal thrust is provided in the manner hereinbefore set forth.

Each auxiliary grid assembly **166** and **167** is mounted on frame members **164** for independent rotational movement about substantially horizontal axes **172** and **173**. With the emitting wires **168** of both auxiliary grid assemblies facing in the same direction, the craft will proceed in the direction toward the emitting wires. With the emitting wires **168** of auxiliary grid assembly **167** facing in a rearward direction and emitting wires of grid assembly **166** facing in a forward direction as illustrated in FIGURE 19, the craft will revolve about an axis mid-way between the effective centers of the two grid assemblies. If the craft is simultaneously tilted in a cyclical manner, an effective radar antenna searching motion is provided which may include a large vertical angle as well as a 360° horizontal scanning operation.

Except where rotation of the craft for searching or scanning operations is a principal purpose for the craft, the emitting wires **168** of each auxiliary grid assembly **166** and **167** are mounted to face in the same direction. When landing or taking off, which is always accomplished in a vertical direction, auxiliary grid assemblies **166** and **167** are preferably pivoted into a horizontal plane. This not only retracts them to prevent interference with landing operations, but also provides a multiple deck structure to give additional lift and control of stability. Horizontal speed may be controlled by varying the angle of auxiliary grid assemblies **166** and **167** with the vertical.

As shown in FIGURE 20 (Sheet 6), the Ionocraft may comprise several decks **180**, **182** and **184** each of which is of similar construction to the single-decked craft shown in FIGURES 10–14 (Sheet 4). Each of the basic structures **180**, **182** and **184** may comprise different antenna types if desired. Several separate ground station antennas **186**, **188** and **190** may be provided particularly where independent signals are transmitted and received by the several antennas of the Ionocraft.

In FIGURE 21 (Sheet 2), a multiple decked craft is illustrated which comprises a central cabin **200** from which two lifting grid assemblies **202** and **203** extend laterally on opposite sides which are equipped with dihedral. Above grid assemblies **202** and **203**, one or more pairs of similar grid assemblies **204** and **205** are supported by a suitable superstructure **208**. The turning axes **212** and **213** for auxiliary grid assemblies **210** and **211** in this embodiment are substantially vertical and extend through support members **214** and **215** to the upper grid assemblies **204** and **205** to provide added rigidity to the craft structure. Retractable antennas **220** and **221** may be provided beneath cabin **200** for establishing communication channels to the ground station (not shown).

In general, it makes little difference whether the emitting wires **12** are connected to the negative or to the positive terminal of the power supply. By tests, it has been determined that with emitting wires **12** connected to the negative terminal, there is an improvement of about 5% over that obtained when the emitting wires **12** are connected to a positive terminal.

In the multiple deck constructions, it is preferable to connect emitting wires **12** and collector grid wires **14** of the adjacent decks to opposite terminals of the high voltage generator as illustrated in FIGURE 20 (Sheet 6), thus making discharge or emitting wires **12** in alternate decks positive and the collector grids negative which is the reverse of the polarity shown in FIGURE 1. In that case, tilting is effected by varying either the negative or positive potential of the corresponding emitting electrode wires and grid sections to provide a rolling movement longitudinally and laterally.

All the above mechanisms and procedures provided for manual control can be utilized for automatic control actuated by an automatic pilot director through suitable servo-mechanisms.

The tilting of the craft in the case of embodiments like those diagrammatically indicated in FIGURE 15 (Sheet 6) and 16 (Sheet 5) provides forward gliding movements much in the manner that a helicopter is propelled in a horizontal direction. Where other means are used for horizontal propulsion, such for example the auxiliary grids shown in FIGURES 19 and 21 or in conjunction with propellers or jet stream, then the tilting will be used to maintain a desirable posture in space. All these movements may be controlled automatically by conventional stabilizing and steering mechanisms borne by the craft or such movement may be accomplished from remote transmitting points either on the ground or from another airborne craft.

The maximum size of crafts of the type here involved is theoretically unlimited, except for structural considerations, since the amount of lift provided increases continuously with area. It is thus contemplated that a particularly useful function of the craft of the present invention may be to serve as means for destruction through collision oncoming vehicles and missiles through air and space. Intercontinental as well as space missiles enter the atmosphere over a target area in predictable trajectories, the terminal end of which is a substantially vertical path. Thus, the large horizontal area of the craft of the present invention is particularly suitable for the purpose of protecting sensitive target areas such as large cities, naval task forces, troop concentrations and the like by its mere physical presence during hovering operations. By maneuvering the craft laterally it is possible to protect an area much larger than the area of the craft since present detection systems give identifying information of the target area about 15 minutes prior to arrival of the missile and the lateral movement of the craft may be effected at speeds of the order of 60 miles per hour, or more depending upon the horizontal propulsion system used. If the target area is vast, several Ionocraft could be maintained aloft to assure collision with oncoming missiles.

While the craft may be powered through conductors extending from ground or ship towers or via microwave power transmissions, it is contemplated that lightweight power plants such as gas turbines or the like, be used to drive suitable high voltage generators which are aboard the craft. As shown in FIGURE 22 (Sheet 6), turbine **230** may be so mounted that its exhaust is directed vertically downwardly to provide additional lifting force while providing shaft rotation for producing the electrical power for the Ionocraft. Turbine **230** is here shown to be mounted for pivotal movement about the axis of shaft **232** which is driven by a tilt motor **234** to change the direction of the exhaust gases from vertical toward a horizontal direction. The entire tilt motor **234**, shaft **232** and turbine **230** assembly may be mounted to be rotated in azimuth by azimuth motor **236** driving annular ring gear **238**. Thus, in emergency operations where maximum horizontal speeds are desired, motors **234** and **236** may be controlled to advance the craft at higher velocities.

Other types of convention airborne power plants, such as turbine propeller combinations, may also be utilized for providing additional lift and aiding in maneuvering in the atmosphere. The turbine of FIGURE 15 may be provided with a reverse thrust device or such propellers may have a reversible pitch, and steering may be accomplished by rudders or vanes located in the jet stream of the turbine. Also, high voltage generation by radio-active isotopes is another method of obtaining the necessary high voltage energy or a primary source of ionization for the propulsion and sustenance of the Ionocraft.

It is also contemplated that this craft may be supplied with electrical power transmitted to the Ionocraft while in flight by microwaves. It has been demonstrated 80% of the energy emitted from a ground station microwave antenna array can be collected in the form of heat by airborne vehicles. In this case, such heat may be readily converted into high voltage by conventional means such as turbines operating high voltage generators, suitable thermocouples and vibrator-transformer convertors or the like. The use of high power microwave amplifiers, such as Amplitrons (Raytheon Co.), for power transmission via microwaves can provide the requisite power for a craft of this type. Therefore, it may be not essential that a self contained power unit be carried by the craft for special uses.

In the preferred form of the craft adapted for military purposes, directional detecting apparatus such as radar or infrared equipment will be carried by the craft to enable an antenna on the craft to lock-on any target object in air and space for the purpose of guiding the craft into the path of such oncoming target object. An Ionocraft of sufficient lift capacity may carry its own computers to process the electromagnetic information to provide the necessary impulses to the controls of the propulsive means to place the craft in the path of collision. Such craft may also be guided from the surface of the earth or from an airborne vehicle in flight, by remote control means to accomplish the collision with an oncoming object.

Explosives may be carried by the Ionocraft for destroying such oncoming objects if the mass of the Ionocraft is inadequate for destructive purposes. Such explosives may be of any known type and adapted to be detonated either upon impact or by proximity fuses where desired. Other types of countermeasures or defensive devices for causing premature explosions of the warhead of a missile may be carried by the Ionocraft as occasions arise.

The invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. The present embodiments are therefore to be considered in all respects as illustrative and not restrictive, the scope of the invention being indicated by the appended claims rather than by the foregoing description, and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced therein.

What is claimed and desired to be secured by United States Letters Patent is:

1. In combination with flying apparatus composed of a structure supporting discharge electrodes for causing adjacent air molecules to become electrically charged spaced from a grid of electrical conducting means for neutralizing the charge on electrically charged molecules whose charge was caused by said discharge electrode and means for applying a high D.C. potential between said electrodes and said grid to cause air to move from said electrodes toward said grids to provide a propulsion force for said flying apparatus, and antenna means for use with radio frequency energy signals, said antenna comprising a plurality of elements which serve also as part of said structure for the flying apparatus by being connected electrically to one terminal of said high D.C. potential to assist in providing propulsion force for said apparatus.

2. In combination with flying apparatus composed of a structure supporting discharge electrodes for causing adjacent air molecules to become electrically charged spaced from a grid of electrical conducting means for neutralizing the charge on electrically charged molecules whose charge was caused by said discharge electrode and means for applying a high D.C. potential between said electrodes and said grid to provide a propulsion force for said flying apparatus, an antenna for use with radio frequency energy signals, said antenna comprising a plurality of elements which serve also as part of said structure for the flying apparatus by being connected electrically to one terminal of said high D.C. potential to assist in providing propulsion force for said apparatus, and the flying apparatus structure having a configuration such that at least one side has a length greater than one-half the wave length of the radio frequency energy transmitted by the antenna.

3. The combination as defined in claim 2 wherein said antenna comprises a plurality of conductive elements having differing lengths whereby said antenna is tuned to different frequencies of radio frequency energy.

4. The combination as defined in claim 2 wherein said antenna comprises elements which serve as part of the mechanical structure of the neutralizing grid.

5. In combination with flying apparatus composed of a structure supporting discharge electrodes for causing adjacent air molecules to become electrically charged spaced from a grid of electrical conducting means for neutralizing the charge on electrically charged molecules whose charge was caused by said discharge electrode and means for applying a high D.C. potential between said electrodes and said grid to provide a propulsion force for said flying apparatus, said flying apparatus having on a plurality of different sides thereof antenna means for directional transmission of radio frequency signal energy, the length of a side being greater than one-half the wave length of the signal energy being transmitted, and means for effecting a scanning movement of said antenna means by rotating said flying apparatus about a vertical axis.

6. In combination: flying apparatus comprising a structure supporting discharge electrodes for causing adjacent air molecules to become electrically charged spaced from a grid of electrical conducting means for neutralizing the charge on electrically charged molecules whose charge was caused by said discharge electrode and means for applying a high D.C. potential between said electrodes and said grid to provide a propulsion force for said flying apparatus; said apparatus being divided electrically into different sections with separate circuits individual to each of said sections for connecting said D.C. potential to the electrodes of the respective section; means including highly directional radiant energy sensitive means on said flying apparatus for receiving energy from an electromagnetic energy source at a control station, mounted on said apparatus for generating control voltages; and circuit means for connecting said control voltages to impedance varying means to reduce the voltage to various ones of the sections of said apparatus to cause said apparatus to assume a position determined by the energy source at said control station.

7. A system and apparatus for effecting destruction of flying vehicles including a flying apparatus composed of a large area structure supporting discharge electrodes for causing adjacent air molecules to become electrically charged spaced from a grid of electrical conducting means which neutralize the charge on electrically charged molecules whose charge was caused by said discharge electrode

and means for applying a high D.C. potential between said electrodes and said grid to provide a propulsion force for said flying apparatus; tracking means carried by said flying apparatus for generating control signals, means responsive to said control signals for guiding said flying apparatus into the path of said flying vehicles and to effect destruction thereof by impact.

8. In flying apparatus composed of a structure supporting discharge electrodes for causing adjacent air molecules to become electrically charged spaced from a grid of electrical conducting means for neutralizing the charge on electrically charged molecules whose charge was caused by said discharge electrode and means for applying a high D.C. potential between said electrodes and said grid to cause air to move from said electrodes toward said grid to provide a propulsion force for said flying apparatus, structural supporting means for said neutralizing grid separating said grid into two substantially equal sized grid areas, and means securing said grid areas in an angularly related position to equip said apparatus with dihedral for stabilization.

9. In flying apparatus composed of a structure supporting discharge electrodes for causing adjacent air molecules to become electrically charged spaced from a grid of electrical conducting means for neutralizing the charge on electrically charged molecules whose charge was caused by said discharge electrode and means for applying a high D.C. potential between said electrodes and said grid to cause air to move from said electrodes toward said grid to provide a propulsion force for said flying apparatus, structural supporting means for said neutralizing grid separating said grid into two pairs of grid areas, the grid areas of each pair being of substantially equal size and on opposite sides of the apparatus, and means securing the grid areas of each pair in an angularly related position to equip said apparatus with dihedral in both longitudinal and lateral directions.

10. The apparatus as defined in claim 9 wherein each grid area is planar and the center of the craft is higher than the outer peripheral edges to thereby provide a negative dihedral.

11. The apparatus as defined in claim 9 wherein each grid area is planar and the center of the craft is lower than the outer peripheral edges to thereby provide a positive dihedral.

12. In flying apparatus composed of a structure supporting discharge electrodes for causing adjacent air molecules to become electrically charged spaced from a grid of electrical conducting means for neutralizing the charge on electrically charged molecules whose charge was caused by said discharge electrode and means for applying a high D.C. potential between said electrodes and said grid to cause air to move from said electrodes toward said grid to provide a propulsion force for said flying apparatus, the improvement wherein the collecting grid comprises a plurality of lightweight thin walled tubes having an outer surface of conductive material, said tubes being inflated and spaced to provide neutralization of the charged particles and permit air flow between adjacent tubes.

13. In flying apparatus composed of a structure supporting discharge electrodes for causing adjacent air molecules to become electrically charged spaced from a grid of electrical conducting means for neutralizing the charge on electrically charged molecules whose charge was caused by said discharge electrode and means for applying a high D.C. potential between said electrodes and said grid to cause air to move from said electrodes toward said grid to provide a propulsion force for said flying apparatus, the improvement wherein the discharge electrodes comprise wires of conductive material mounted above the neutralizing grid, and a plurality of short wires mounted on the discharge electrodes with free ends thereof suspended beneath the discharge electrodes at substantially equal distances from said neutralizing grid to thereby provide point sources for causing ionization.

14. In flying apparatus composed of a structure supporting discharge electrodes for causing adjacent air molecules to become electrically charged spaced from a grid of electrical conducting means for neutralizing the charge on electrically charged molecules whose charge was caused by said discharge electrode and means for applying a high D.C. potential between said electrodes and said grid to cause air to move from said electrodes toward said grid to provide a propulsion force for said flying apparatus, an auxiliary discharge electrode and neutralizing grid structure supported by said flying apparatus for movement relative to said flying apparatus about a pivot axis for controlling the posture and direction of movement of said apparatus.

15. In flying apparatus composed of a structure supporting discharge electrodes for causing adjacent air molecules to become electrically charged spaced from a grid of electrical conducting means for neutralizing the charge on electrically charged molecules whose charge was caused by said discharge electrode and means for applying a high D.C. potential between said electrodes and said grid to cause air to move from said electrodes toward said grid to provide a propulsion force for said flying apparatus, a pair of auxiliary discharge electrode and neutralizing grid structures supported at spaced positions on said flying apparatus, each of said grid structures being mounted for movement about pivot axes for controlling the posture and direction of movement of said apparatus.

16. Apparatus as defined in claim 15 wherein said pivot axes are in a horizontal plane.

17. Apparatus as defined in claim 15 wherein said pivot axes are vertical.

18. In flying apparatus composed of a structure supporting discharge electrodes for causing adjacent air molecules to become electrically charged spaced from a grid of electrical conducting means for neutralizing the charge on electrically charged molecules whose charge was caused by said discharge electrode and means for applying a high D.C. potential between said electrodes and said grid to cause air to move from said electrodes toward said grid to provide a propulsion force for said flying apparatus, a rigid frame for said neutralizing grid including intersecting support members lying in a vertical plane, and air foil means secured to said intersecting support members to provide added stabilization to said apparatus.

19. In flying apparatus composed of a structure supporting discharge electrodes for causing adjacent air molecules to become electrically charged spaced from a grid of electrical conducting means for neutralizing the charge on electrically charged molecules whose charge was caused by said discharge electrode and means for applying a high D.C. potential between said electrodes and said grid to cause air to move from said electrodes toward said grid to provide a propulsion force for said flying apparatus, a polygonal rigid inner and outer frame for said neutralizing grid, a center portion of said grid surrounded by said inner frame being without the electrical conducting means forming the grid, and diagonally disposed frame assemblies extending between corners of said outer frame and said inner frame.

20. Apparatus as defined in claim 19 wherein the apparatus is provided with negative dihedral and the center portion is covered with an air foil to provide a pressure wave beneath the craft for added lift.

21. In combination a base member; bracket means supported on said base member for movement about a first pivot axis; a first gear segment on said bracket; a first pair of electrical circuit elements having parameters which are variable in magnitude by rotation of a shaft individual to each circuit element; a gear on each of said shafts; said gears being positioned at opposite ends of the gear segment on said bracket so that only one shaft is rotated from a reference position at any moment; a control stick mounted for pivotal movement on said bracket about a

second pivot axis perpendicular to said first pivot axis; a second gear segment on said control stick; a second pair of electrical circuit elements similar to said first pair of electrical circuit elements mounted on said bracket means; and separate gears for controlling said second pair of electrical circuit elements, said gears being mounted at opposite ends of said second gear segment so that only one gear is rotated from a reference position at any moment.

22. The combination as defined in claim 21 wherein each electrical circuit element contains a shaft spring loaded to a reference position.

23. The combination as defined in claim 22 wherein the electrical circuit elements are variable resistances.

24. The combination as defined in claim 23 together with flying apparatus composed of a structure supporting discharge electrodes for causing adjacent air molecules to become electrically charged spaced from a grid of electrical conducting means for neutralizing the charge on electrically charged molecules whose charge was caused by said discharge electrode, said grid comprising four electrically isolated sections of substantially the same size, and means for connecting one terminal of a high D.C. potential source to said electrodes and a second terminal to each of the four sections of said grid through different ones of said variable resistances.

25. In apparatus for thrust generation in atmosphere, composed of spaced emitting and collecting electrodes energized with a voltage sufficiently high to be capable of causing ionization for effecting an ionic discharge from an emitting electrode to produce thrust on at least one of the electrodes of said apparatus arising from elastic molecular and particle collisions occurring in the space between said electrodes during molecule and particle movement in the direction from the emitting electrode to the collecting electrode, said collecting electrode being composed of crossed grid wires of conductive material forming an open surface that is generally normal to the direction of molecule and particle movement, the improvement wherein said crossed grid wires have a diameter of at least 8 mils.

26. The apparatus as defined in claim 25 wherein the collecting electrode ends are covered by a layer of conductive material.

27. In flying apparatus composed of a structure supporting discharge electrodes for causing adjacent air molecules to become electrically charged spaced from a grid of electrical conducting means for neutralizing the charge on electrically charged molecules whose charge was caused by said discharge electrode; and means for applying a high D.C. potential between said electrodes and said grid to cause air to move from said electrodes toward said grid to provide a propulsion force for said flying apparatus; the improvement wherein said neutralizing grid comprises outer peripheral frame members and a centrally positioned inner frame between which the conducting means comprising the neutralizing grid are mounted, and which are positioned in different horizontal planes to thereby provide the craft with dihedral; and said discharge electrodes comprise a plurality of substantially parallel wires, each mounted parallel to the outer peripheral frame members and in different horizontal planes so that all wires are substantially aquidistant from said neutralizing grid.

28. A system and apparatus for effecting destruction of flying vehicles including a flying apparatus composed of a large area structure supporting discharge electrodes for causing adjacent air molecules to become electrically charged spaced from a grid of electrical conducting means which neutralize the charge on electrically charged molecules whose charge was caused by said discharge electrode and means for applying a high D.C. potential between said electrodes and said grid to provide a propulsion force for said flying apparatus; auxiliary discharge electrode and neutralizing grid means mounted for pivotal movement for controlling the forward direction of movement; tracking means carried by said flying apparatus for generating control signals, means responsive to said control signals for guiding said flying apparatus into the path of said flying vehicles and to effect destruction thereof by impact.

29. A system and apparatus for effecting destruction of flying vehicles including a flying apparatus composed of a large area structure supporting discharge electrodes for causing adjacent air molecules to become electrically charged spaced from a grid of electrical conducting means which neutralize the charge on electrically charged molecules whose charge was caused by said discharge electrode and means for applying a high D.C. potential between said electrodes and said grid to provide a propulsion force for said flying apparatus; a gas turbine mounted on said apparatus with its exhaust gases directed downwardly during hovering operation and means for deflecting said exhaust gases in a direction providing a horizontal propulsion force to increase to the forward velocity of said apparatus; tracking means carried by said flying apparatus for generating control signals, means responsive to said control signals for guiding said flying apparatus into the path of said flying vehicles and to effect destruction thereof by impact.

References Cited in the file of this patent

UNITED STATES PATENTS

2,495,748	Matson	Jan. 31, 1950
2,503,109	Harris	Apr. 4, 1950
2,598,064	Lindenblad	May 27, 1952
2,613,887	Woods	Oct. 14, 1952
2,842,645	Dalgleish et al.	July 8, 1958
2,888,189	Herb	May 26, 1959
2,892,949	Hardy	June 30, 1959
2,949,550	Brown	Aug. 16, 1960

FOREIGN PATENTS

1,174,334	France	Nov. 3, 1958

Sept. 7, 1965 L. G. SMITH 3,205,381

IONOSPHERIC BATTERY

Filed March 9, 1962

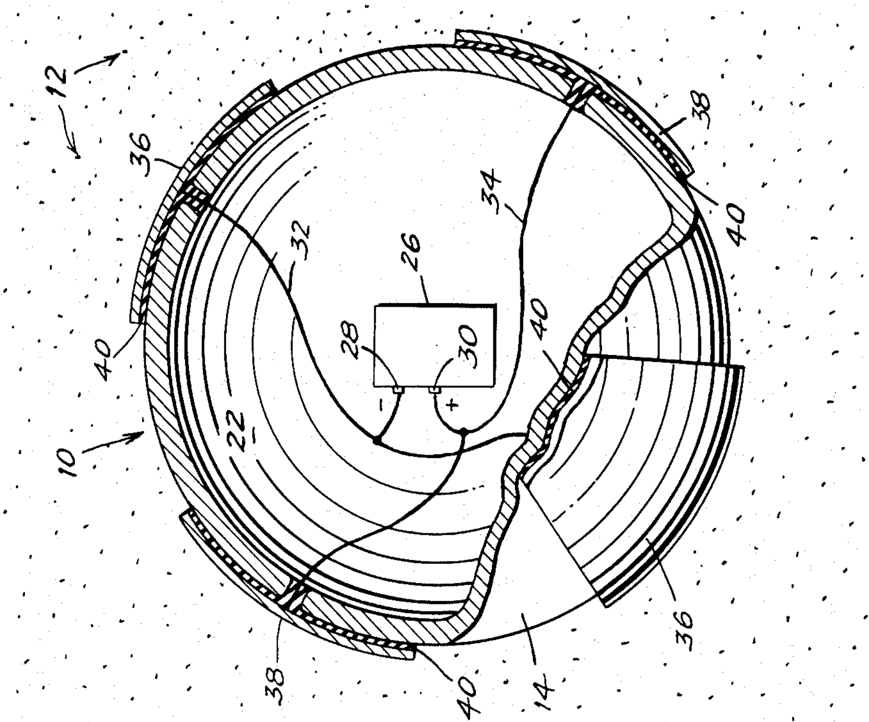
12
36
40
10
22
32
28
26
30
34
38
40
40
14
36
40
38
-
+

Fig. 2

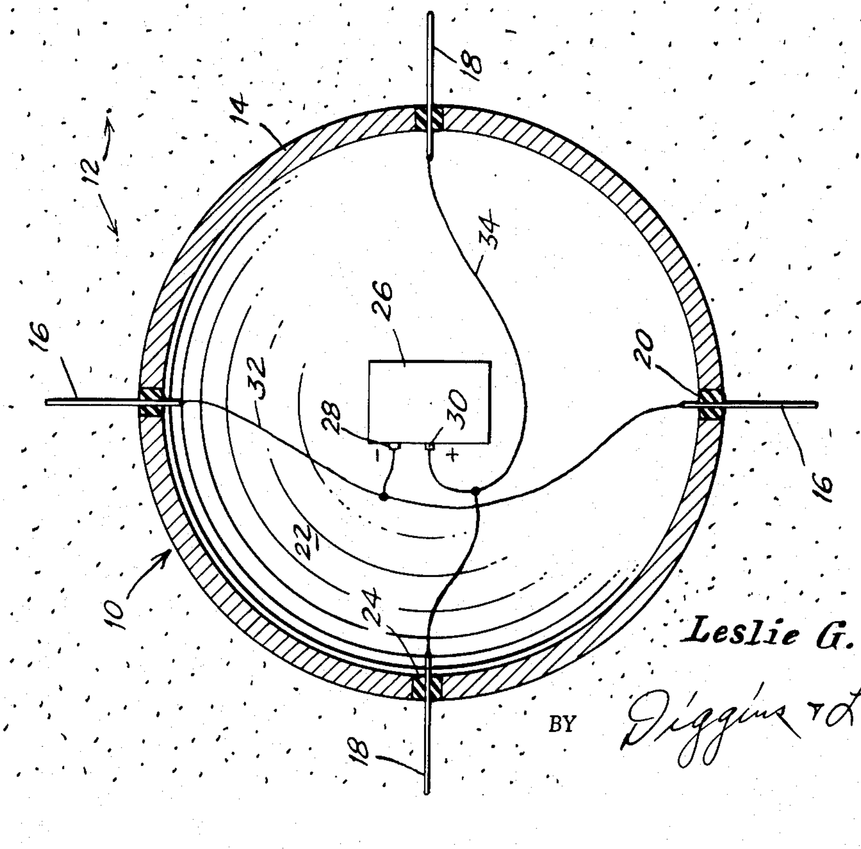
18
14
12
16
32
26
28
30
34
22
24
10
20
16
18
-
+

Fig. 1

INVENTOR

Leslie G. Smith,

BY Diggins & Le Blanc

ATTORNEYS

United States Patent Office 3,205,381
Patented Sept. 7, 1965

3,205,381

IONOSPHERIC BATTERY

Leslie G. Smith, Waltham, Mass., assignor, by mesne assignments, to the United States of America as represented by the Administrator of the National Aeronautics and Space Administration

Filed Mar. 9, 1962, Ser. No. 178,721
11 Claims. (Cl. 310—5)

This invention relates to a method and apparatus for tapping the enormous reservoir of energy existing in the ionosphere and more particularly for furnishing electrical power to space vehicles.

Within recent years various space probes and space vehicles such as satellites have been launched by man in an effort to learn more about the earth and the solar system within which he lives. Each of these various space craft has employed numerous electrical circuits which are utilized to detect and transmit back to earth certain information ascertained by the craft as it moves through space. In view of the presence of such electrical circuits, some source of electrical power must be made available whereby these circuits can be energized. In the past there have been three main sources of electrical power carried by space vehicles for energizing these circuits, namely, (1) conventional storage batteries, (2) solar batteries and (3) a combination of the conventional storage and solar batteries. Although these various electrical power sources have functioned with some degree of success each has certain serious limitations which cannot be readily overcome.

In the first instance, the conventional storage battery may be of either the wet or dry type but, due to its superior storage capacity, the wet cell type is used nearly exclusively in space vehicles. This additional storage capacity of the wet cell battery is of great importance since the useful life of the vehicle is usually dependent upon the proper operation of its electrical circuits which are, in turn, directly dependent upon the capacity and life of the available power source. Use in space vehicles of a storage battery that utilizes a liquid electrolyte presents several problems that have not been completely solved, however, nor is there any apparent solution to some of these problems. For example, to increase the output capacity of the battery and thereby extend the useful life of the satellite a greater amount of electrolyte must be carried by the battery. But to increase the quantity of electrolyte carried by the battery results in both an increase in size and weight of the power source and space vehicle which cannot, in most instances, be tolerated. Furthermore, the electrolyte may not function correctly if it is subjected to the extreme conditions of space such as, for example, the changes in temperature that are normally encountered. Thus, to insure the proper operation of the wet cell battery involved in powering a space vehicle and to prevent any leakage that might occur should the battery be damaged, some shielding or insulating of the battery is usually involved which, needless to say, adds materially to the cost and weight of the vehicle.

To overcome the space, weight and leakage problems encountered when storage batteries of the conventional wet cell type are used, space engineers have turned to the recently developed solar cell or battery which converts the light rays received from the sun directly into electrical energy. While these solar batteries do not require any type of electroylte and are, therefore, relatively light in comparison to the wet cell storage battery, they are expensive and their inherent principle of operation presents one problem which cannot be overcome, namely, the inability of the solar cell to produce power while the space vehicle is shaded from the sun such as occurs when a satellite is in the shadow cast by the earth.

In an effort to overcome this loss of power which occurs each time a satellite passes into the shadow cast by the earth or some other heavenly body, a storage battery of the wet cell type was added to the vehicle. While the satellite is in the sunlight the solar battery powers the various electrical circuits and charges the storage battery so that, once the sunlight is no longer available, the storage battery can be switched on to furnish the necessary electrical power to the vehicle. While this combined system functions satisfactorily, there is added to the already existing problems of weight, size and electrolyte leakage the problem of incorporating a battery charging and switch-over system between the solar and storage batteries. The addition of this type of system to the power source is undesirable since it adds unduly to the complexity, sensitivity and cost of the space vehicle.

According to the present invention, it has been found that some of the foregoing disadvantages of space vehicle power sources can be overcome by employing an ionospheric type battery which utilizes the ionosphere or other ionized medium existing in outer space as a supplemental source of energy. Since power is a scarce commodity essential for the operation of the space vehicle and for the proper functioning of its experiments, any means for supplying additional power while adding only slightly to the weight of the vehicle, increases the efficiency and life of the space vehicle. Although it has been known for some time that the ionosphere contains an enormous reservoir of energy in the form of excited and ionized atoms and molecules, no practical method has heretofore been found whereby this energy could be utilized as a source of power for earth satellites even though much thought has been given to this problem. The operation of the present invention is based upon the discovery that the ionosphere can be utilized as a part of supplemental power source for the load circuit in a space vehicle when said vehicle is equipped with electrodes having different work functions which, in turn, are electrically connected through the load circuit.

Accordingly, a primary object of this invention is to provide a method of producing electrical power from ionized media existing in nature.

Another object of this invention is to provide a simple, highly reliable supplemental power source for space vehicles.

Yet another object of this invention is to provide an ionospheric battery for use by satellites.

Still another object of this invention is to provide an ionospheric battery which utilizes the ionosphere as an electrolyte.

A further object of this invention is to provide a supplemental power source for space vehicles that will operate when shielded from the sun.

Yet a further object of this invention is to provide a power source for a satellite that is light and inexpensive to produce.

Another object of this invention is to provide a satellite with various shaped electrodes constructed of metals having different contact potentials.

These and further objects and advantages of this invention will become more apparent upon reference to the following description, claims and the appended drawings wherein:

FIGURE 1 is a cross sectional view of an orbiting satellite equipped with a power source constructed in accordance with one embodiment of the present invention; and

FIGURE 2 is a view in partial section of an orbiting satellite equipped with a power source constructed in accordance with a second embodiment of the present invention.

With continuing reference to the accompanying drawings wherein like reference numerals designate similar parts throughout the various views, and with initial attention directed to the embodiment of FIGURE 1, reference numeral **10** is used to generally designate a space satellite which is orbiting or otherwise passing through an ionized layer or medium such as, for example, the ionosphere which surrounds the earth. As is well known, the ionosphere contains an enormous number of excited and ionized atoms and molecules which are generated or produced when the gasses surrounding the earth are subjected to bombardment by high energy particles and rays such as, for example, ultra-violet or cosmic rays that originate throughout the solar system. Although these various ionized particles created by this bombardment are invisible to the eye they are represented pictorially by the various dots **12** for purposes of clarity.

The satellite **10** consists of an envelope or outer skin surface **14** which may be constructed of metal, ceramic or any other like material which is suitable for space vehicles of this type. As seen in FIGURE 1, a plurality of electrically conductive electrodes **16–18**, which are shown in the form of rods and are constructed of materials having different contact potentials (work functions), are mounted in a spaced relationship in the medium **12** through apertures **20** in the envelope **14** of the satellite. In order to prevent a leak from occurring between the normally pressurized interior **22** of the satellite and the ionosphere, which is a near vacuum, and to insulate the electrodes in case a metal is used for the envelope **14**, vacuum tight insulator feed-through elements **24** are fitted around the electrodes **16–18** and are securely mounted within the aperture **20**. These feed-through elements may be of any suitable type so long as they are constructed to withstand the changes in environmental conditions encountered in space without losing their insulative and sealing qualities.

Carried within the cavity or interior **22** of the satellite **10** is a load circuit **26** which may be, for example, a transmitter or any other electrical device that requires a source of electrical power for its proper operation. The load circuit **26** is mounted to the interior wall of the satellite by suitable means (not shown) and has its positive and negative power input terminals **28–30** connected to the various electrodes **16–18** respectively, by suitable electrical conductors **32–34**.

The overall principle and operation of the ionospheric battery in relationship to the satellite **10** can be explained substantially in the following manner. While the satellite remains on earth prior to its being hurled into or through an ionized medium, the electrodes **16–18** are at substantially the same potential since very few ionized particles are present. Thus, no current will flow through the load circuit **26**. However, once the satelllite is launched and it enters an ionized medium a potential difference will be established between the electrodes **16–18** and an electrical current will begin to flow through the load circuit **26**. This difference in electrical potential that exists between the electrodes **16–18** while the satellite is in an ionized medium is attributed to the fact that the electrodes **16** are constructed of a material having a different contact potential or work function from that of the electrodes **18**. This difference in contact potential between the electrodes **16–18** will result in the free electrons of the ionized medium being collected on one group of electrodes, for example the electrodes **16**, while the positive ions produced when the gas is ionized will be collected on the other electrodes **18**. With the electrodes **16** charged to a negative potential in relationship to electrodes **18**, a predetermined amount of current will flow through the load circuit **26** and can thus be utilized to power an electrical circuit such as the transmitter mentioned above.

The exact material from which the two types of electrodes **16** and **18** are constructed will depend upon several factors such as, for example, the amount of output power desired, the material readily available for constructing the electrodes, the susceptibility of the material to damage or destruction when exposed to the environmental conditions existing in space, and various other considerations that will be readily apparent to a worker in the battery field.

While the electrodes **16–18** used in the ionospheric battery of FIGURE 1 have been described as rods or whip-like elements it is to be realized that the electrodes may take other shapes. By way of example, a plurality of flat, plate-like electrodes **36–38** are shown mounted on a satellite in FIGURE 2. These electrodes are constructed of materials having different contact potentials and are insulated from the outer surface **14** of the satellite when this is necessary by an insulative layer **40** of any suitable material having the desired insulative properties under the severe conditions existing in space. The use of these plate electrodes **36–38** has the advantage of increasing the surface area exposed to the ionized medium thus permitting a larger output current to be produced without the necessity of using excessively long electrodes as might be the case if the rod electrodes of FIGURE 1 were used.

The output current of the battery is dependent upon the degree to which the medium is ionized and to the extent that the electrode surface area is exposed to the medium. An increase in either results in an increase in the efficiency of the battery. From an analysis of various experiments it is expected that a current of about two microamps per cm.^2 of electrode surface at a potential of one volt would be produced if the satellite is orbiting in the ionosphere at an altitude of approximately 350 km. and electrodes of gold and platinum are used.

It will also be readily apparent that the envelope of the satellite itself could be constructed in sectional form and of materials having different contact potential thus eliminating the necessity of attaching different electrodes to the satellite or the slight increase in weight necessarily involved. Of course the various sections of material making up the envelope of the satellite would have to be insulated from one another to prevent electrical shorting from occurring.

It will be apparent from the foregoing that the power source of this invention is extremely light, rugged and involves a minimum of cost to construct. By using the vast natural reservoir of energy existing in the form of ionized gases for producing the electrical power necessary to energize the various electrical circuits of a space vehicle a practically indefinite source of electrical power is made available to the vehicle. An ionospheric battery constructed in accordance with this invention is also much more reliable in operation than other heretofore known power sources since no liquid electrolyte is involved which can leak, deteriorate or otherwise become spent, nor is the operation of the battery dependent upon its position in relationship to a source of light.

The invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. The present embodiments are therefore to be considered in all respect as illustrative and not restrictive, the scope of the invention being indicated by the appended claims rather than by the foregoing description, and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced therein.

What is claimed and desired to be secured by United States Letters Patent is:

1. A space vehicle equipped with a generator of electrical energy operable in the ionosphere as a supplemental source of potential for a load thereof, said generator comprising at least two electrode means on said vehicle and positioned to contact said ionosphere when said vehicle is in said ionosphere, said two electrode means each having different contact potentials, one of said two electrode means being gold and the other being

platinum, and electrical connector means connecting said two electrode means to said load.

2. Apparatus for generating electrical energy from the ionosphere comprising, a pair of spaced electrode means having different contact potentials positioned to contact said ionosphere so that a direct-current electric potential is established between said electrode means, load circuit means, and means electrically connecting said load circuit means to said electrode means whereby said load circuit means is powered by the direct-current electrical potential existing between said electrodes.

3. Apparatus for generating electrical energy from the ionosphere comprising, carrier means, a first electrode having a first contact potential mounted on said carrier means, a second electrode having a second contact potential spaced from said first electrode and mounted on said carrier means, said first and second electrodes being positioned on said carrier means so that they will be in contact with said ionosphere when said apparatus is placed in the ionosphere so that a direct-current electric potential is established between said first and second electrodes, and an electrical load circuit connected between said first and second electrodes, said load circuit being powered by the direct-current potential produced by said first and second electrodes.

4. Apparatus according to claim 3 wherein said first and second electrodes extend away from said carrier means.

5. Apparatus according to claim 3 wherein said first and second electrodes form a portion of said carrier means.

6. In a space vehicle adapted to pass through ionized areas in space, a power source for said vehicle while it is passing through said ionized areas comprising, a plurality of electrode means located on the outer surface of said vehicle, said electrode means being constructed of materials having different contact potentials, and means connecting said electrode means together through a load circuit whereby said circuit receives a source of direct-current as said space vehicle passes through said ionized areas.

7. In a space vehicle according to claim 6 wherein said electrode means are rod shaped.

8. In a space vehicle according to claim 6 wherein said electrode means are plate shaped.

9. A satellite comprising an envelope, electrical load means carried in said envelope, apertures extending through said envelope at spaced points, electrode means

having different contact potentials, said electrode means passing through said apertures, means in said aperture for insulating said electrode means from said envelope, and electrical conductor means connected between said electrode means and said electrical load means whereby said electrical load means is energized by the direct-current electric potential established when said electrode means passes through an ionized medium.

10. A space satellite comprising an envelope, an electrical load circuit carried in said envelope, a power supply for said electrical load circuit comprising a first electrode having a first contact potential connected to said electrical load circuit and extending through said envelope and a second electrode having a second contact potential connected to said electrical load circuit and extending through said envelope, said first and second electrodes being insulated from one another so that said electrical load circuit is powered by the direct-current electric potential established by said first and second electrodes coming in contact with an ionized medium.

11. The method of powering electrical circuit means carried by a space vehicle comprising the steps of positioning first and second electrode means having different work functions on said vehicle, passing said vehicle through the ionosphere so that said first and second electrode means are in contact therewith and produce a direct-current potential between them, and connecting said first and second electrodes to said electrical circuit means by means of electrical connector means, whereby said direct-current potential is made available for operating said circuit.

References Cited by the Examiner

UNITED STATES PATENTS

414,943	11/89	Dewey	310—5 X
911,260	2/09	Pennock	310—5 X
2,303,321	12/42	Bennett	175—264
2,445,305	7/48	Hochgesaug	250—83.6
2,510,397	6/50	Hansell	171—330
2,802,113	8/57	Ohmart	250—83.6
2,864,012	12/58	Thomas	310—3
2,900,535	8/59	Thomas	310—3
3,021,472	2/62	Hernqvist	322—2

LLOYD McCOLLUM, *Primary Examiner.*

JOHN H. MACK, *Examiner.*

March 25, 1969 W. C. BROWN ET AL 3,434,678

MICROWAVE TO DC CONVERTER

Filed May 5, 1965

Sheet 1 of 3

FIG. 1

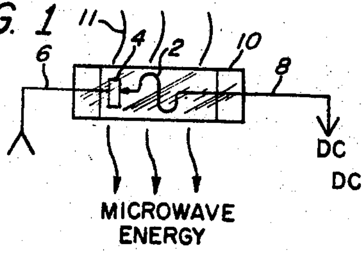
11
4
2
10
6
8
DC
MICROWAVE ENERGY

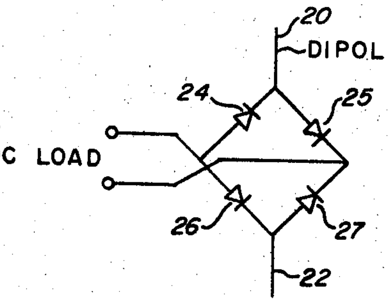
20
DIPOLE
24
25
DC LOAD
26
27
22

FIG. 2

FIG. 3

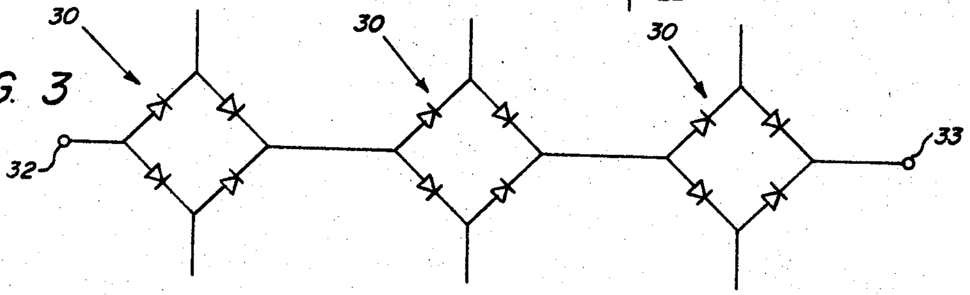
30
30
30
32
33

FIG. 4

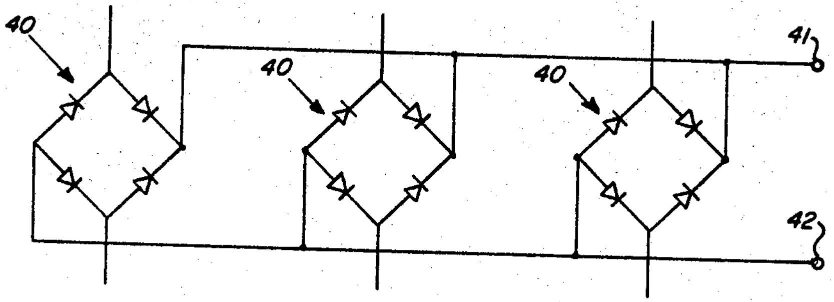
40
40
40
41
42

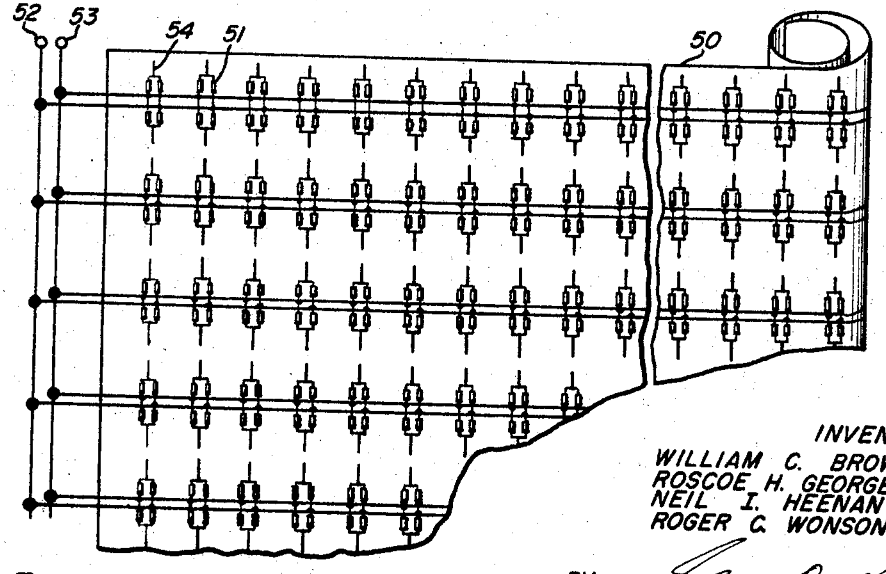
52
53
54
51
50

FIG. 5

INVENTORS
WILLIAM C. BROWN
ROSCOE H. GEORGE
NEIL I. HEENAN
ROGER C. WONSON

BY Edgar O. Kost
ATTORNEY

March 25, 1969 W. C. BROWN ET AL 3,434,678

MICROWAVE TO DC CONVERTER

Filed May 5, 1965

Sheet 2 of 3

FIG. 7

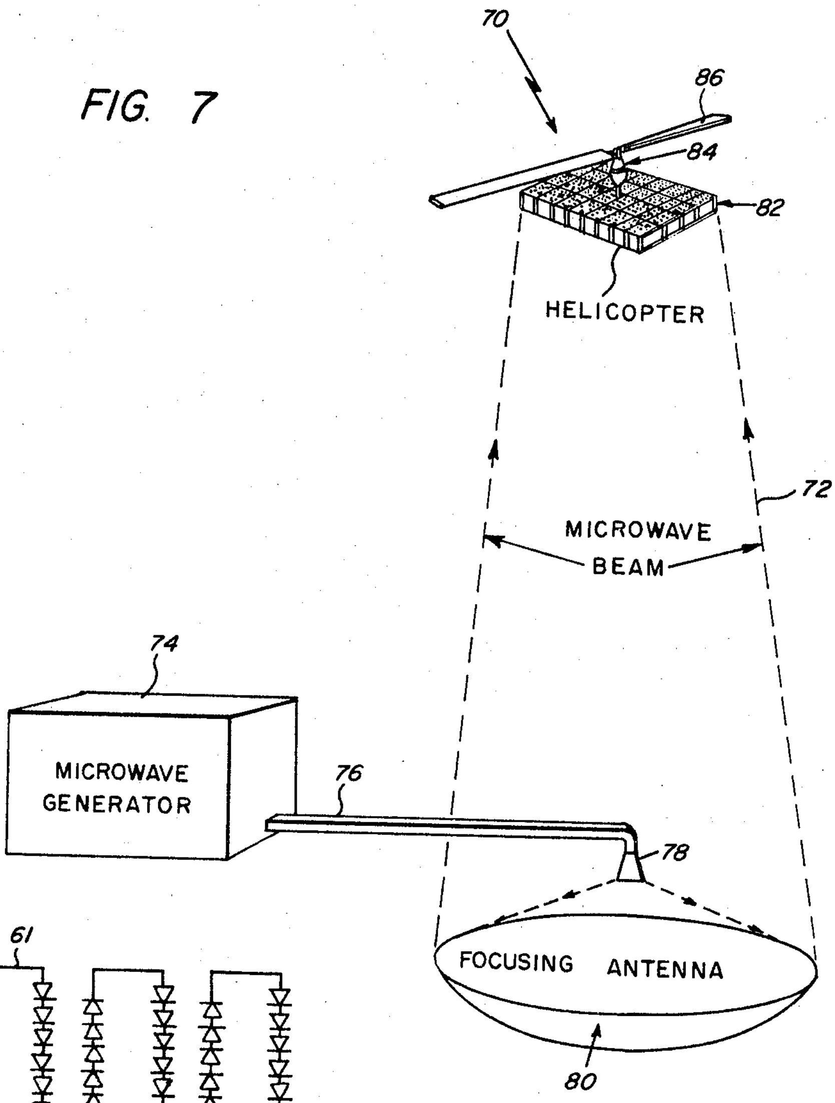
70
86
84
82
HELICOPTER
72
MICROWAVE
BEAM
74
MICROWAVE
GENERATOR
76
78
FOCUSING ANTENNA
80

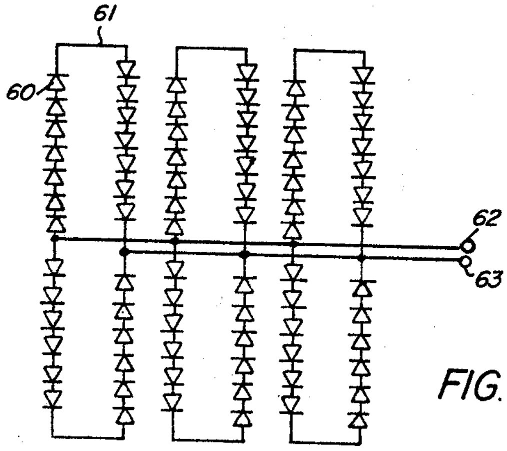

FIG. 6

INVENTORS
WILLIAM C. BROWN
ROSCOE H. GEORGE
NEIL I. HEENAN
ROGER C. WONSON

BY Edgar O. Rost

ATTORNEY

March 25, 1969 W. C. BROWN ET AL **3,434,678**

MICROWAVE TO DC CONVERTER

Filed May 5, 1965

Sheet 3 of 3

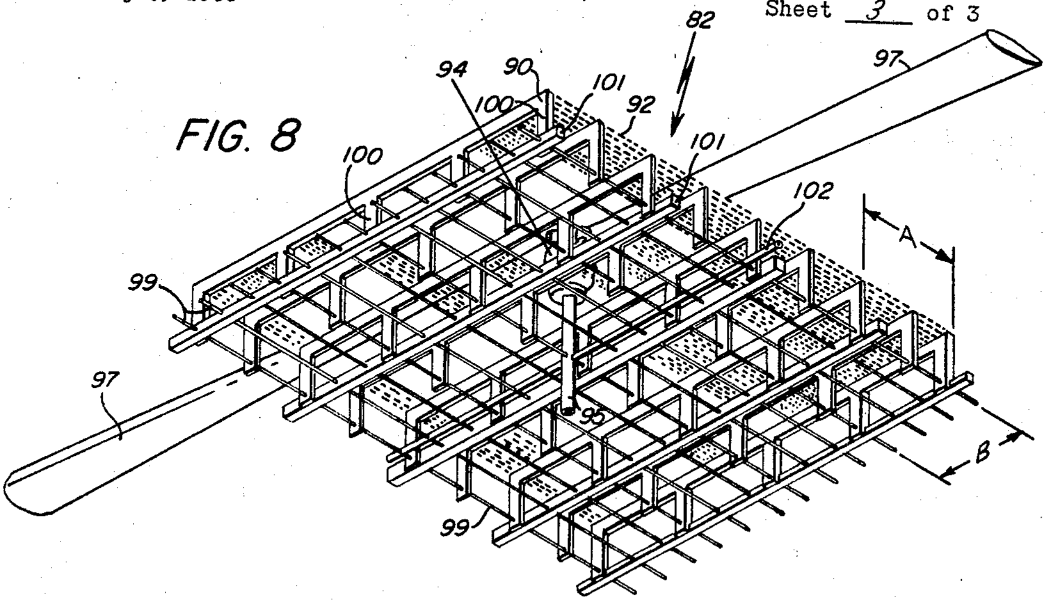
82
94
90
101
100
92
100
101
102
A
99
97
97
95
99
B

FIG. 8

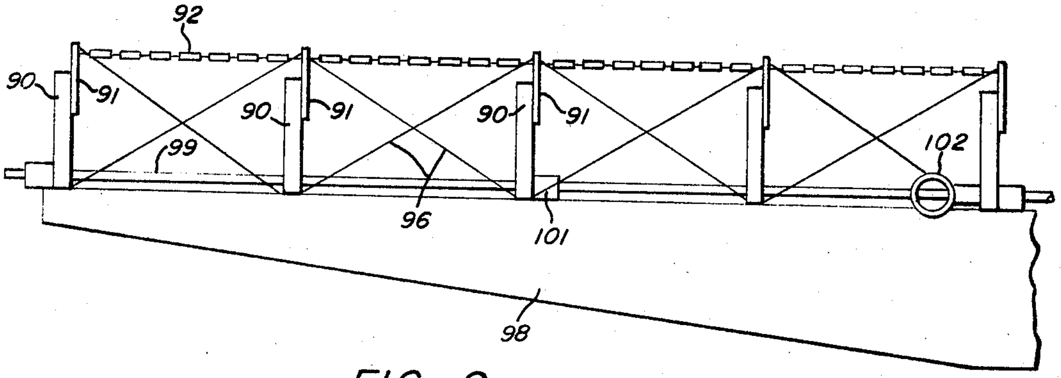
92
90
91
90
91
90
91
99
96
101
102
98

FIG. 9

INVENTORS
WILLIAM C. BROWN
ROSCOE H. GEORGE
NEIL I. HEENAN
ROGER C. WONSON

BY

ATTORNEY

United States Patent Office

3,434,678
Patented Mar. 25, 1969

3,434,678
MICROWAVE TO DC CONVERTER

William C. Brown, Weston, Mass., Roscoe H. George, West Lafayette, Ind., and Neil I. Heenan, Needham, and Roger C. Wonson, Beverly, Mass., assignors to Raytheon Company, Lexington, Mass., a corporation of Delaware

Filed May 5, 1965, Ser. No. 453,415
Int. Cl. B64g 1/00
U.S. Cl. 244—1
3 Claims

ABSTRACT OF THE DISCLOSURE

A combined antenna and conversion mechanism for reception of beamed high frequency electromagnetic energy in space including a large array of unidirectional current semiconductor rectifier devices. A self-supporting space vehicle utilizing the rectified DC electrical energy for propulsion is disclosed in an illustrative embodiment.

The present invention relates in general to the transfer of energy by means of an electromagnetic wave beam and more particularly to interception and rectification of such energy into low frequency electrical DC energy with a high degree of efficiency.

Improved technology in the field of microwave energy generation at superpower levels has resulted in the realization of electrical energy transmission over considerable distances for remote energization of devices or vehicles without the aid of wires. The transmission of microwave electromagnetic energy into space has been commonly employed in the radar pulse echo systems for the detection and orientation of desired objects within a predetermined scanning range of a transmitting antenna. Beams of a similar nature may now be employed for other useful purposes and the advantages attendant the utilization of electromagnetic energy in the microwave region in contrast with other wavelengths may now be enumerated.

Microwaves have been generally defined as high frequency radio waves whose wavelength is less than 30 centimeters, with a lower wavelength limit on the order of 1 millimeter sometimes being applied to what is commonly referred to as the "microwave region." The superiority of high frequency microwaves is due in part to the fact that it is generally desirable to focus the transmitted energy so as to achieve a high power density at a remote point or area with respect to a given power source. In accordance with the laws of optics, the sharpness of the microwave beam produced by a transmitting antenna varies as the ratio of antenna dimensions to the wavelength of the transmitted energy. Therefore, for a given or desired power density or beam sharpness, a decrease in the wavelength of the transmitted energy permits a corresponding decrease in the dimensions of the antenna. From the standpoint of mechanical considerations, it is desirable to employ small antennas and other components, and it is therefore advantageous to employ high frequency energy of very short wavelength. In addition, the difficulties encountered in long wave transmission as a result of natural and man-made interference or noise do not occur with any appreciable significance at microwave frequencies. Further, in aerospace applications with considerable distances separating the transmitter at an earth or mother planet location and the employment of shorter wavelength beamed energy is preferred since longer wave signals will generallly be reflected at certain altitudes by reflecting layers in the atmosphere.

In view of certain losses due to absorption which may occur in the atmosphere, microwaves in the region having the approximate bounds of 2 and 30 centimeters are readily adaptable to the convenient radiation of power to remote points without the utilization of wires. The preferred wavelengths are of the order of 5 or 10 centimeters to provide efficient focusing with existing transmitting antenna systems which may be maintained at a reasonable size. An illustrative device of the superpower high frequency microwave generators operative in the desired band is the so-called Amplitron which is an amplifier having a broad bandwidth and excellent performance characteristics for the focusing of the beam. Such devices are capable of producing 15 or 20 kilowatts of average continuous wave power in the neighborhood of 10 centimeters in wavelength with capabilities expected in the region of 500 kilowatts or more average power with 50 megawatts peak power. A complete description of such devices may be had by referring to Patent No. 2,933,723 issued Apr. 19, 1960 to William C. Brown and assigned to the assignee of the present invention.

With microwave energy capable of being generated and directed over longer distances conversion of such high frequency electromagnetic energy is of paramount concern. One conversion mechanism in the prior art involves direct conversion of such energy into heat which may then be utilized directly or indirectly for propulsion or generation of flight-producing forces. Examples of such devices for heat energy exchange as well as space vehicles utilizing such energy may be noted in Patent No. 3,174,705, issued Mar. 23, 1965, to D. Schiff et al., as well as U.S. Letters Patent No. 3,083,528, issued Apr. 2, 1963 and No. 3,114,517, issued Dec. 17, 1963, to William C. Brown. The heat exchanger method of conversion of electromagnetic energy into useful power is limited by the overall efficiencies of approximately 25 percent in the conversion of heat into mechanical or electrical work. Desirable, therefore, would be the direct rectification of the high frequency electromagnetic energy into low frequency electrical energy for the operation of many useful aerospace devices as well as systems.

The present invention has for its primary object the conversion of high frequency electromagnetic energy in the microwave region directly into low frequency electrical energy.

A further object of the present invention is the provision of a combined nondirectional receiving antenna and microwave electromagnetic energy to low frequency electrical energy conversion means in a unitary structure.

A still further object of the present invention is a provision of a new and novel combined nondirectional receiving antenna and microwave to DC energy converter for aerospace applications.

Another object of the present invention is the provision of a new and novel nondirectional receiving antenna and microwave to DC energy converter having a high degree of efficiency.

Still another object of the present invention is the provision of a new and novel aerospace vehicle with nondirectional receiving antenna and microwave to DC energy converter means with said vehicle being capable of being supported by its own energy generation means at a distance spaced apart from the power generation means.

In accordance with the teachings of the present invention, the above and other objects are achieved by the employment of efficient unidirectional microwave power rectifiers and dipole antenna means. Such rectifying devices, while being individually limited in power-handling capabilities, normally in the order of fractions of watts, have been found to be highly efficient means for the rectification of microwave power when assembled in large numbers in various arrays. It is interesting to note that the observed collective efficiency was on the order of 40 to 70 percent. In an illustrative embodiment, point-

contact semiconductor diodes were arranged in four arm bridge connected networks with the networks interconnected in various configurations such as series, parallel and series-parallel.

In discussing aerospace applications, an additional problem is encountered in the beaming of microwave energy to a remote point and the interception and utilization of such electrical energy. In such applications the advantages of a vehicle which may be maintained in space for indeterminate periods of time without employing a local fuel source are readily apparent. Such devices could readily provide communication networks, surveillance functions using radar techniques along with numerous other functions. The acapture of the beamed high frequency electromagnetic energy raises the need for an efficient antenna means capable of intersecting the beam at high altitudes. Conventional techniques employed in microwave radar usage such as receiving antenna horns are capable of intersecting only a small portion of the beam energy and add considerable weight in applications involving heavier-than-air vehicles. In an exemplary embodiment of the invention a space vehicle, namely a helicopter, is disclosed for either moving flight or a stationary location with self-supporting electrically operative propulsion means. The semi-conductor diode rectifier arrays have been demonstrated to fulfill the receiving antenna functions as well as the electrical energy rectification means in a highly efficient manner. Such combined antenna and rectifier means has also assisted in reduction of the weight problem in airborne devices. Further, it has provided a nondirectional means for the interception of the microwave energy to thereby reduce the problems of focusing inherent in prior art directional horn type receiving antennas.

With the above features, advantages and objects in mind the invention will now be described by reference to the following detailed description together with the accompanying drawings in which:

FIG. 1 is a perspective view of an illustrative diode rectifier;

FIG. 2 is a schematic circuit diagram of a bridge connected diode network with dipole antenna means;

FIG. 3 is a schematic circuit diagram of a plurality of bridge connected networks arranged in series;

FIG. 4 is a schematic circuit diagram of a parallel bridge connected network array;

FIG. 5 is a perspective view of an illustrative embodiment of a combined antenna and rectifier array in a folded or rolled up configuration;

FIG. 6 is a schematic circuit diagram illustrating the bridge connected diode array incorporated in the aerospace vehicle shown in FIG. 7;

FIG. 7 is a schematic representation in elevation illustrative of a heavier-than-air aerospace vehicle incorparating the structure of the present invention;

FIG. 8 is a perspective view of the aerospace vehicle embodiment as viewed from the under portion thereof; and

FIG. 9 is an enlarged partial view in elevation of a portion of the illustrative embodiment shown in FIG. 8.

FIG. 1 illustrates a point-contact semiconductor diode rectifier of the type employed in radar microwave receiver apparatus to rectify returned radar pulses. Any of the high burnout semiconductor diodes having high rectification characteristics are preferred and are commercially available, such as the 1N82 or 1N830. The rectifying junction is formed by whisker element **2** contacting the semiconductor element **4** respectively connected to leads **6** and **8**. Silicon is preferred over germanium for element **4** because of its ability to operate at higher temperatures and thereby handle higher powers. Envelope **10** houses the rectifying elements and may be of a hermetically sealed dielectric material or combination metal and ceramic composition. The inherent characteristic of such diode rectifiers is that the microwave energy is intercepted and rectified in a unidirectional manner and the line **11** indicate pictorially the rays of the beamed electromagnetic microwave energy in a plane normal to the envelope. In FIG. 2 a full-wave bridge connected diode network is illustrated with the forward direction of the rectified DC electric current indicated by the direction of the arrow symbols. The network shown consists of half-wave dipoles **20** and **22** each terminated with a diode rectifier element **24** to **27** in an arm of the bridge connected network. The dipole elements **20** and **22** are of the half-wave configuration and may be spaced apart from each other a one-half wavelength at the frequency of the beamed electromagnetic energy.

Referring now to FIG. 3, an array of bridge connected diode networks each with the half-wave dipoles are shown connected in series. Each network is referred to by the numeral **30** and is similar in the bridge connections to the single element network shown in FIG. 2. The DC output of the collective rectified energy is coupled by means of terminals **32** and **33**. In FIG. 4, a similar number of individual bridge connected diode-dipole networks are shown connected in a parallel array. Each network is indicated by the numeral **40**, and the output terminals are indicated as **41** and **42**.

Any number of diode-dipole networks may be provided and in FIG. 5 such a multi-element array is illustrated by mounting on a flexible material **50** which may be rolled or folded into any desired package or enclosed within a capsule to be launched and released at a predetermined point in space. Any flexible material which is pervious to electromagnetic energy is preferred. The total power desired would be the determining factor in a number of individual diode-dipole elements required. In this embodiment, the bridge connected networks **51** are connected in parallel to the output load indicated by terminals **52** and **53**, and representative measurements of electrical characteristics have shown that approximately five watts of DC electrical energy is realizable for each square foot of area of the combined antenna-rectifier. While the dipole elements **54** have been indicated in a particular array, it is within the scope of the invention to stagger the placement of such dipoles to increase the over-all efficiency of the antenna-rectifier.

To further increase the DC powder output, the full-wave bridge connected networks are preferably arranged with a plurality of diodes in series in each arm of the bridge. An illustrative schematic circuit diagram of such a configuration is shown in FIG. 6 wherein seven diodes **60** are shown in each arm of the bridge circuit and are connected in series for a total of twenty-eight diodes in each bridge network. The dipole members will then be the substantially U-shaped end portions **61** at the ends of each brace of seven diodes. In the illustration three such twenty-eight diode bridge networks are shown connected in parallel to terminals **62** and **63**. This closer spacing and compact arrangement has been shown to be a source of improved power output and is capable of a high degree of reliability through the redundant nature of the parallel series connections within each bridge network. If one of the diode rectifiers fails to function the over-all voltage drop across this element would be divided among the six remaining diode rectifiers. If any of the connecting wires between the diode elements should break, the adjacent arms of the other bridge assemblies would take the additional load due to the close proximity of the respective arms to each other. In addition, it is possible to have a number of open connections or inoperative diodes dispersed throughout the array without any serious impairment in performance.

In relation to the array concept to be hereinafter described it may be stated that within a six inch square area ten such individual bridge networks each containing twenty-eight diode rectifiers for a total of 280 diode rectifiers may be deployed in such a manner as to provide

maximum exposed area for each diode as well as the connecting leads. Such an arrangement will be hereinafter referred to as a "module" and a DC output in excess of fourteen watts has been measured for such a module. Any number of such modules could be connected provided for a desired power yield and this module concept readily lends itself to use in certain aerospace applications now to be described.

In FIG. 7 a propelled type of space vehicle **70** is shown wholly supported by means of the transfer and rectification of continuous wave electromagnetic energy via a microwave beam **72**. The source of the microwave energy which may be of the Amplitron type device as described in the aforementioned issued Patent No. 2,933,723 is indicated as **74**. This energy is fed by waveguide means **76** to a transmitting horn **78** to illuminate an ellipsoidal beam forming focusing antenna **80** for the transmission of the microwave beam **72**. It will be appreciated by those skilled in the art that the representations of the microwave generation and transmitting antenna means are pictorial representations to illustrate the usage of the invention in diagrammatic form and the present invention is not limited to any particular source of microwave energy or transmitting antenna assembly. It may be stated the reflector of the antenna assembly is considerably larger than most of the reflectors of the prior art in order to focus a large amount of the microwave power at high altitudes for use in the transfer of energy to space vehicles. Such antenna assemblies may be partially supported in a large hollowed area on the earth's surface or other convenient means of support.

The space vehicle or helicopter **70** can be described as a main body member supporting antenna-rectifier means **82** including a large number of the so-called modules connected together and rigidly supported in a planar parallel array. A motor **84** is supported by the combined body member and the receiving antenna-rectifier means and actuates the rotor **86** of conventional design employed in such self-propelled hovering vehicles. The disclosed vehicle provides for the illumination of the planar array of the semiconductor diode dipole elements by the microwave beam and the direct conversion of the microwave power transmitted by the beam into usable electrical energy for the self-propulsion of the device without any local fuel supply being required.

FIGS. 8 and 9 illustrate a space vehicle **82** comprising a plurality of the combined receiving antenna-rectifier module means for interception and rectification of the electromagnetic microwave energy beam emanating from an earth or mother planet source. A planar array of the antenna-rectifier modules is mechanically supported by means of structural members **90** of any lightweight wood or metal. Insulators **91** positioned coextensive with the members **90** support the diode rectifier array and avoid interference with the receiving and electrical performance characteristics by the structural support members. Carrying forward the module concept of 280 diode rectifiers to provide an approximate power output of 14 watts, it was noted that any number of such modules may be coupled together since the individual module outputs are relatively insensitive to a wide range of load resistances connected to the common output terminals. To achieve the desired electrical output of approximately 120 volts and 250 watts of power, subgroups of four modules each were assembled and parallel connected with an approximate 30 volts available for each subgroup. Four such subgroups were series-connected to result in a total of 4,480 diode rectifiers or 16 modules assembled in a two foot square self-supporting planar parallel array structure. The individual diode rectifiers connected in each arm of the bridge network are indicated by the numeral **92**. An exemplary module configuration would extend within the area delineated by the dotted lines and reference letter A on one side and similar dotted lines and reference letter B on the other side.

A motor **94** is connected to the DC side of the overall array and may be additionally supported by tubular member **95**. A shaft and propulsion means consisting of rotor blades **97** provide for the upward lift of the overall vehicle for the self-supporting of same in space applications. Additional structural support such as interlaced rigging **96** of a high tensile strength material such as nylon or steel wire, as well as bracing member **98**, may be employed for strengthening of the body means to withstand the vibrational forces and downwash from the propulsion means.

In accordance with the well known technology of microwave transmission the combined array of diode rectifiers and propulsion means presents a specific load impedance which must be suitably matched to the transmitted microwave energy beam to result in maximum efficiency. In aerospace applications a mismatch of approximately ten to one may be evident. Matching of the load impedance to a value of approximately 377 ohms as the free space value will be provided by a plurality of coplanar parallel metallic rod members **99** disposed in a grating array in front of the diode rectifiers a predetermined distance. Rod members **99** are linearly disposed and extend in a similar direction as the assembled diode rectifiers. A selected frontal spacing of one-quarter of the wavelength of the microwave frequency being transmitted has been experimentally determined to be suitable for impedance matching purposes. An approximate spacing of two inches between the respective members was preferred for a selected microwave frequency of 2,450 megacycles. Each of the members **90** are provided with lateral sections **100** to support the elongated bar members **101** which in turn maintain the rod members **99** in the desired position. A tubular member **102** of a lightweight metal may also be provided to combine with the motor support member **95** for structural support.

The combined antenna-rectifier array provides a source of electrical energy to render any space vehicle self-supporting. The diode rectifier elements when assembled in the antenna array have been found to be nondirectional with respect to interception of the beamed microwave energy. This represents a large step forward in the utilization of high power microwave energy over the prior art horn-type receiving antennas which must be accurately focused and pointed in a particular direction for the reception of any energy. The connections between the respective members of the diode rectifier array and deployment in the parallel configuration serves to provide maximum exposed area. Such connections and in particular the end loop portions adjacent the terminus of each arm of the bridge networks serves as an efficient dipole for the interception of the microwave energy.

Although it is not intended as a full explanation of the high degree of efficiency attained with the disclosed antenna-rectifier array, it is believed that the whisker elements within the semiconductor diodes themselves are a contributing factor and may function as additional dipole elements. The disclosed embodiment functioned efficiently when illuminated by microwave energy generating a vertically polarized beam. Hence, an efficient and light weight energy conversion apparatus is disclosed which may be self-supporting without the requirement of a large local fuel supply payload.

It may be within the purview of the invention to use the available rectified electrical energy for performing many functions in addition to the actuation of the propulsion means. Hence, communications' payloads may be maintained at predetermined positions in space in a hovering attitude utilizing a portion of the electrical energy available. Relay signals to other such vehicles or return signals to ground stations would then be within the realm of possibility. Such available energy may also be employed for servomechanisms, stabilizing and countertorque systems for the navigation of such vehicles.

The electrical efficiencies realized with the combined

receiving antenna and rectifier means have also provided certain weight advantages over other energy converters in aerospace applications. Examples of such converters would be heat exchangers or solar cells. In comparison to the present invention where five to eight pounds per kilowatt of energy realized is a normal characteristic, other energy conversion means weigh in the vicinity of 150 pounds per kilowatt of realizable energy. The inherent advantages of the present invention are therefore apparent. While the technology in the diode rectifier art is being continually advanced, new diode power rectifiers as well as integrated circuit techniques are readily available to future configurations of the present invention. The so-called Schottky barrier diodes could be employed to produce combined antenna-rectifier means weighing even less than two pounds per kilowatt of available energy.

Although the foregoing detailed description has referred to DC power rectification it will be evident that with suitable circuit components low frequency AC energy may also be made available. In addition, other propulsion means may be readily substituted using electrical energy. The embodiments disclosed herein are illustrative only and other modifications or alterations will be apparent to those skilled in the art which do not depart from the scope of the broadest aspects of the present invention as defined in the appended claims.

What is claimed is:

1. A space vehicle comprising:

body means;

said body means including spaced structural support members;

combined antenna and DC electrical energy rectification means for the interception and rectification of incident high frequency electromagnetic microwave energy carried by said support members in a planar parallel array;

said rectification means comprising a plurality of four

arm full-wave bridge connected rectifier circuit networks each having a plurality of unidirectional semiconductors in each arm;

said networks being electrically interconnected to common output terminals;

electrically operable propulsion means comprising a motor and rotor members carried by said body means and connected to said terminals for the utilization of said rectified DC energy; and

means for matching the load impedance of said combined antenna and electrical energy rectification means to the incident microwave energy.

2. A space vehicle according to claim 1 wherein said load impedance matching means are arranged in a coplanar array coextensive with said antenna and energy rectification means array, and spaced therefrom a distance of approximately one-quarter of a wavelength at the frequency of the microwave energy.

3. A space vehicle according to claim 2 wherein said load impedance matching means comprise a plurality of parallel disposed elongated metallic members.

References Cited

UNITED STATES PATENTS

1,217,149	2/1917	Caldwell	321—27 X
2,165,055	7/1939	Kafka	321—27
2,444,458	7/1948	Master	321—8
2,927,321	3/1960	Harris	343—6.8 X
3,098,971	7/1963	Richardson	325—592 X
3,174,705	3/1965	Schiff et al.	244—1

RODNEY D. BENNETT, *Primary Examiner.*

MALCOMB F. HUBLER, *Assistant Examiner.*

U.S. Cl. X.R.

307—151; 318—16; 321—27; 325—494; 343—100

United States Patent [19]
Esaki

[11] **4,198,644**
[45] **Apr. 15, 1980**

[54] **TUNNEL DIODE**

[75] Inventor: **Leo Esaki**, Chappaqua, N.Y.

[73] Assignee: **The United States of America as represented by the Secretary of the Army**, Washington, D.C.

[21] Appl. No.: **914,103**

[22] Filed: **Jun. 9, 1978**

[51] **Int. Cl.**^{2} **H01L 29/88; H01L 29/161**

[52] **U.S. Cl.** **357/12**; 357/16; 357/3; 357/61

[58] **Field of Search** 357/12, 16, 3, 61

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,626,257	12/1971	Esaki	317/234 R
3,626,328	12/1971	Esaki	331/107 G
3,864,721	2/1975	Cohen	357/17
4,088,515	4/1975	Blakeslee	148/175
4,103,312	7/1978	Esaki	357/16
4,137,542	1/1979	Esaki	357/16

Primary Examiner—Martin H. Edlow
Attorney, Agent, or Firm—Nathan Edelberg; Sheldon Kanars; Jeremiah G. Murray

[57] **ABSTRACT**

A tunnel diode is disclosed wich includes a heterostructure consisting of a first layer of $GaSb_{1-y}As_y$ and a second layer of $In_{1-x}Ga_xAs$. It is also disclosed that other alloys of Group III and Group V materials can be employed in a tunnel diode of the instant invention.

5 Claims, 4 Drawing Figures

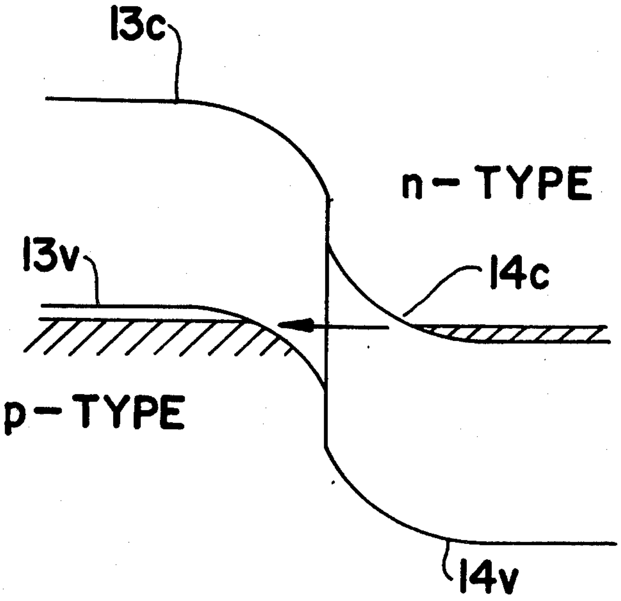

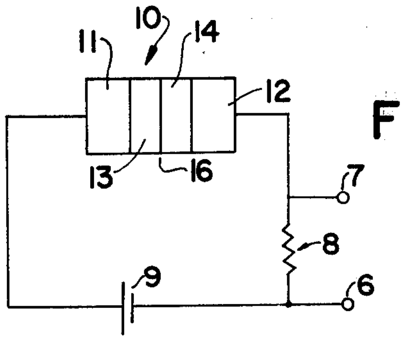
11
10
14
12
13
16
7
8
9
6

FIG.1

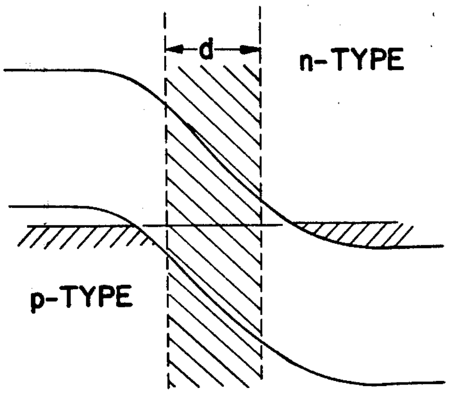
d
n-TYPE
p-TYPE

FIG.2

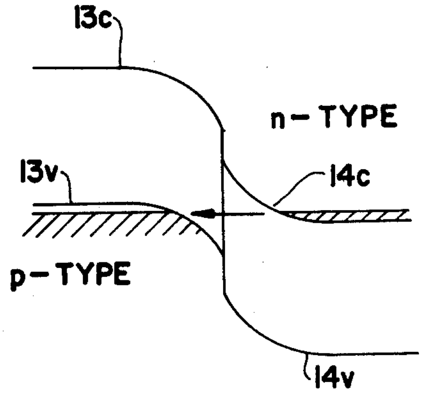
13c
n-TYPE
13v
14c
p-TYPE
14v

FIG.3

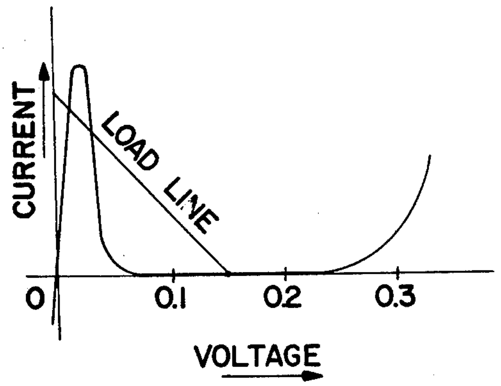

FIG.4

TUNNEL DIODE

The invention described herein may be manufactured and used by or for the Government for governmental purposes without the payment of any royalties thereon or therefor.

FIELD OF INVENTION

This invention relates to semiconductor devices and particularly to tunnel diode semiconductor devices.

BACKGROUND OF THE INVENTION

The first tunnel diode was invented by the instant inventor over twenty years ago. Much study and work has gone into improving tunnel diodes but the basic structure has not been altered. The conventional tunnel diode includes a heavily doped p-type material and a heavily doped n-type material. Such heavy doping is required to permit tunneling in narrow junction widths. One result of the heavily doped material was a large number of dislocations in the crystalline structure resulting in high valley currents and attendant power dissipation. A second feature of highly doped tunnel diodes was a substantial tunneling width (length of path) which resulted in a relatively low tunneling current, limiting their high frequency performance.

BRIEF DESCRIPTION OF THE INVENTION

In accordance with the teachings of this invention, an improved tunnel diode is provided with a heterostructure comprising first and second layers of III-V compound semiconductor alloys wherein the first layer is an alloy including a first Group III material and a first Group V material, while the second layer is an alloy including a second Group III material different from the first Group III material and a second Group V material different from the first Group V material, and wherein the valance band of the first alloy is closer to the conduction band of the second alloy than it is to the valance band of the second alloy.

In the preferred embodiment of this invention, the first Group III material is In, the first Group V material is As, the second Group III material is Ga, and the second Group V material is Sb. A further feature of the preferred embodiment is that the first layer is an alloy of $GaSb_{1-y}As_y$ while the second layer is an alloy of $In_{1-x}Ga_xAs$ and y is equal to 0.918x+0.082.

DESCRIPTION OF THE DRAWINGS

For a more complete understanding of the invention, reference should be made to the following detailed description and drawings in which:

FIG. 1 is a schematic drawing showing a semiconductor device including a tunnel diode in accordance with the principles of this invention;

FIG. 2 is an energy level diagram of a conventional tunnel diode;

FIG. 3 is an energy level diagram of a tunnel diode constructed in accordance with the teachings of this invention; and

FIG. 4 is a voltage vs. current diagram of the tunnel diode of FIG. 1.

DETAILED DESCRIPTION OF THE INVENTION

Referring now to FIG. 1 we see a semiconductor device 10 having a first and second ohmic contacts 11 and 12. Connected between the ohmic contacts are layers 13 and 14 which form a junction at their interface 16. A DC power supply 9 in series with a load resistor 8 is connected across semiconductor device 10. An output signal is provided between terminals 6 and 7.

The layer 13 is an alloy including a first Group III material and a first Group V material, while the layer 14 is an alloy including a second Group III material different from the first Group III material and a second Group V material different from the first Group V material. As shown in FIG. 3, the valance band 13V of the layer 13 is closer to the conduction band 14C of the layer 14 than it is to the valance band 14V of the layer 14. Also, as shown in FIG. 3 the layer 13 has been degenerately doped with an excess of acceptors to the extent that the Fermi level lies in the valance band. The layer 14 has been degenerately doped with an excess of donors to the extent that the Fermi level lies in the conduction band. When these layers are interfaced as a heterojunction diode, as shown in FIG. 3, the band edges beyond the junction region are such that the top of the valance band of layer 13 will be higher than the bottom of the conduction band of layer 14, thereby permitting tunnel diode action. In the preferred embodiment, the first Group III material is In, the first Group V material is As, the second Group III material is Ga, and the second Group V material is Sb.

In the preferred embodiment, the layer 13 is an alloy including InAs, and preferably an alloy of $In_{1-x}Ga_xAs$, while the layer 14 is an alloy including GaSb and, preferably an alloy of $GaSb_{1-y}As_y$.

The perameters y and x can be related in a preferred embodiment to obtain perfect lattice matching by the relationship:

$$y=0.918x+0.082$$

The device of FIG. 1 can be manufactured by molecular beam epitaxy. A system useful in growing an alloy semiconductor including the tunnel diode of this invention can be constructed similar to the system described for GaAs and AlAs, by L. L. Chang, L. Esaki, W. E. Howard, R. Ludeke and N. Schul in the Journal, J. Vac. Sci. Technol. 10, 655 (1973) with the exception of its having effusion cells containing elemental sources of the elements In, Ga, Sb and As, and of the dopent Sn. The substrates used include (100) oriented GaAs, InAs or GaSb. Chemically polished substrates should be degreased and etched prior to loading and subsequently Ar+ bombarded and annealed in an ambient of the corresponding group-V element in the molecular beam epitaxy system. The deposition temperature should be between 450° and 600° C., and the growth rate between one and three angstroms per second. The growth rate is determined by the arrival rates of the Group III elements. For a more complete discussion see "$In_{1-x}Ga_xAs$— $GaSb_{1-y}As_y$ Heterojunctions by Molecular Beam Epitaxy, by H. Sakaki, L. L. Chang, R. Ludeke, C. A. Chang, G. A. Sai-Halasz and L. Esaki; Appl. Phys. Lett. 31, 211 (1977), and MBE of InGaAs and GaSbAs, by C. A. Chang, R. Ludeke, L. Chang, and L. Esaki, Appl. Phys. Lett. 31, 759 (1977).

A comparison of FIG. 2 to FIG. 3 will show that the energy level diagram of FIG. 3 (of the tunnel diode of this invention) closely resembles the energy diagram of known tunnel diodes produced by heavy doping. It is noted that the differences between FIG. 2 and FIG. 3 can be viewed as though the shaded portion indicated as d in FIG. 2 has been removed and the remaining portions have been brought closer together. This results in a shorter tunneling path and therefore higher current in the structure of the present invention. A further advantage of the present invention is improved crystallinity resulting from the lower doping which produces a lower valley current than in conventional tunnel diodes. The net effect is a tunnel diode having a characteristic such as shown in FIG. 4 wherein the peak current is enhanced and the valley current is substantially reduced from conventional tunnel diodes providing better high frequency performance.

While this invention has been described with respect to particular embodiments thereof, numerous others will become obvious to those of ordinary skill in the art in light hereof.

What is claimed is:

1. A tunnel diode comprising: first and second layers of different semiconductor materials; the top of the valance band for the material of said first layer being at an energy between the energy of the bottom of the conduction band and the top of the valance band for the material of said second layer; the bottom of the conduction band for the material of said first layer being at an energy greater than the energy of the bottom of the conduction band for the material of said second layer; said first layer being degenerately doped with an excess of acceptors to the extent that the Fermi level lies in the valance band; said second layer being degenerately doped with an excess of donors to the extent that the Fermi level lies in the conduction band; and said first and second layers being interfaced to form a semiconductor heterojunction.

2. A tunnel diode according to claim 1, wherein the top of the valance band of said first layer is closer to the bottom of the conduction band of said second layer than it is to the top of the valance band of said second layer.

3. A tunnel diode as defined in claim 1, in which said layers are formed from different Group III-V alloys.

4. A tunnel diode as defined in claim 2, in which the first layer is an alloy of $In_{1-x}Ga_xAs$ and the second layer is an alloy of $GaSb_{1-y}As_y$,

5. A tunnel diode as defined in claim 4, in which $y=0.918x+0.082$.

* * * * *

United States Patent [19] [11] **4,253,190**
Csonka [45] **Feb. 24, 1981**

[54] **COMMUNICATIONS SYSTEM USING A MIRROR KEPT IN OUTER SPACE BY ELECTROMAGNETIC RADIATION PRESSURE**

[75] Inventor: **Paul L. Csonka,** Eugene, Oreg.

[73] Assignee: **The United States of America as represented by the United States Department of Energy,** Washington, D.C.

[21] Appl. No.: **28,780**

[22] Filed: **Apr. 10, 1979**

[51] **Int. Cl.**^{3} **H01Q 6/13;** H04B 7/145; H04B 7/185

[52] **U.S. Cl.** .... **455/12;** 343/100 CS; 343/757

[58] **Field of Search** .... 343/100 ST, 100 AM, 343/100 AD, 700, 711, 712, 757, 705; 325/4, 112, 14; 455/12, 98

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,150,363	9/1964	Finvold	343/5 SA
3,151,325	9/1964	Kompfner	343/100 AM
3,169,245	2/1965	Cutler	343/100 ST

Primary Examiner—Benedict V. Safourek
Attorney, Agent, or Firm—James E. Denny; Dean E. Carlson; Richard E. Constant

[57] **ABSTRACT**

A method and system are described for transmitting electromagnetic radiation by using a communications mirror located between about 100 kilometers and about 200 kilometers above ground. The communications mirror is kept aloft above the atmosphere by the pressure of the electromagnetic radiation which it reflects, and which is beamed at the communications mirror by a suitably constructed transmitting antenna on the ground. The communications mirror will reflect communications, such as radio, radar, or television waves up to about 1,100 kilometers away when the communications mirror is located at a height of about 100 kilometers.

18 Claims, 6 Drawing Figures

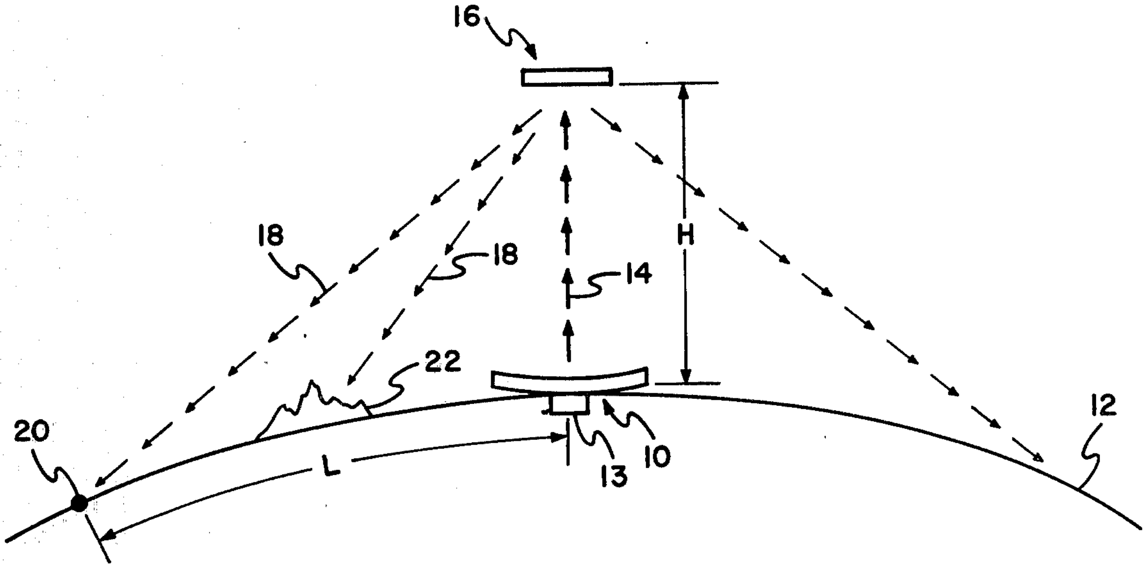

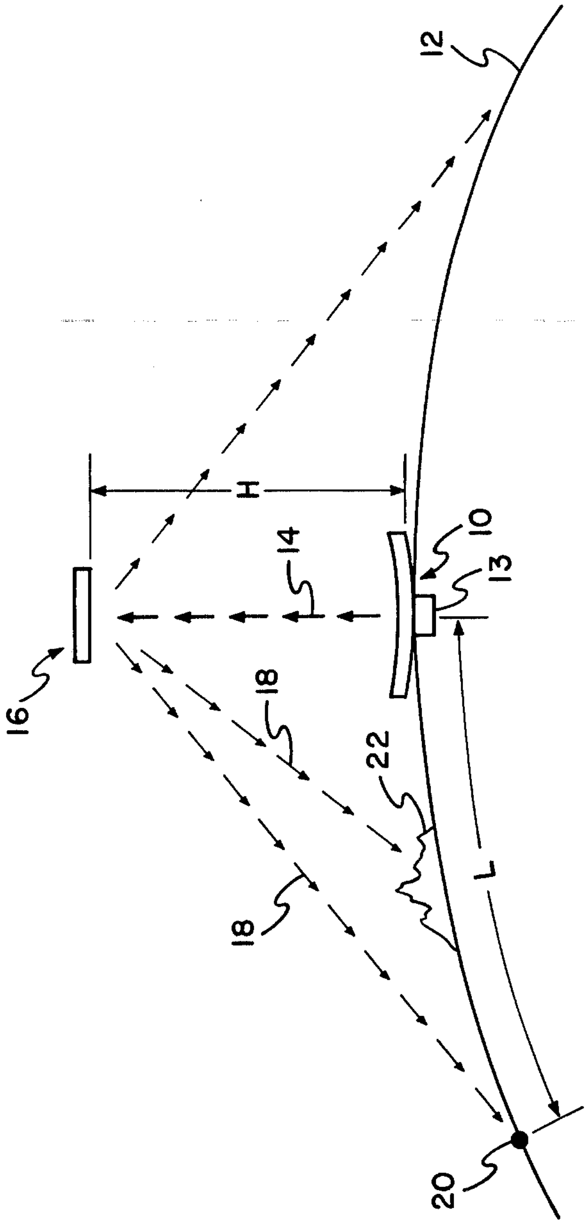
12
H
10
14
13
16
18
22
L
18
20

Fig. 1

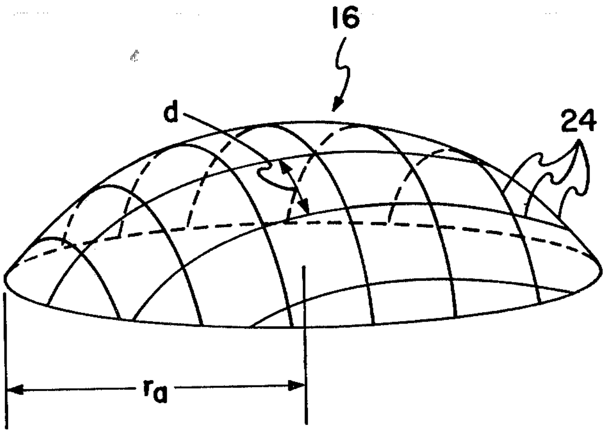

Fig. 2

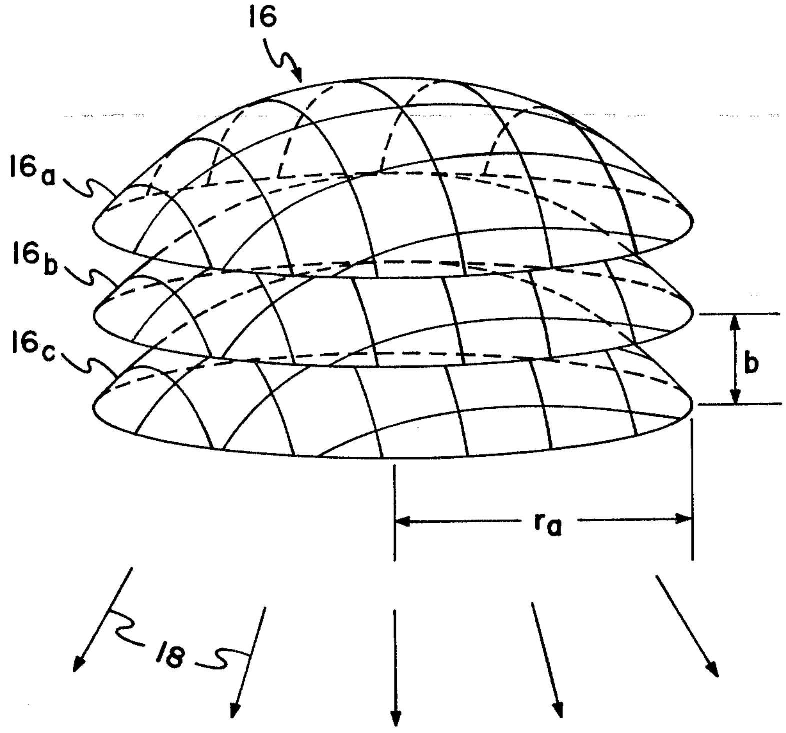
16
16a
16b
16c
b
r_a
18

Fig. 3

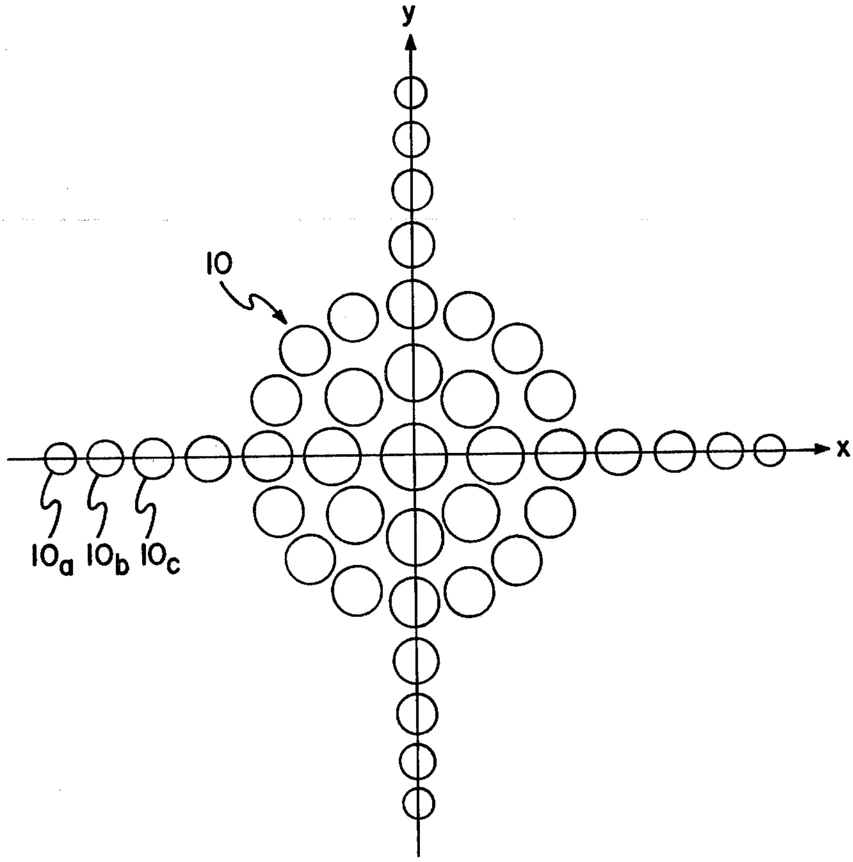
y
x
10
10a
10b
10c

Fig. 4

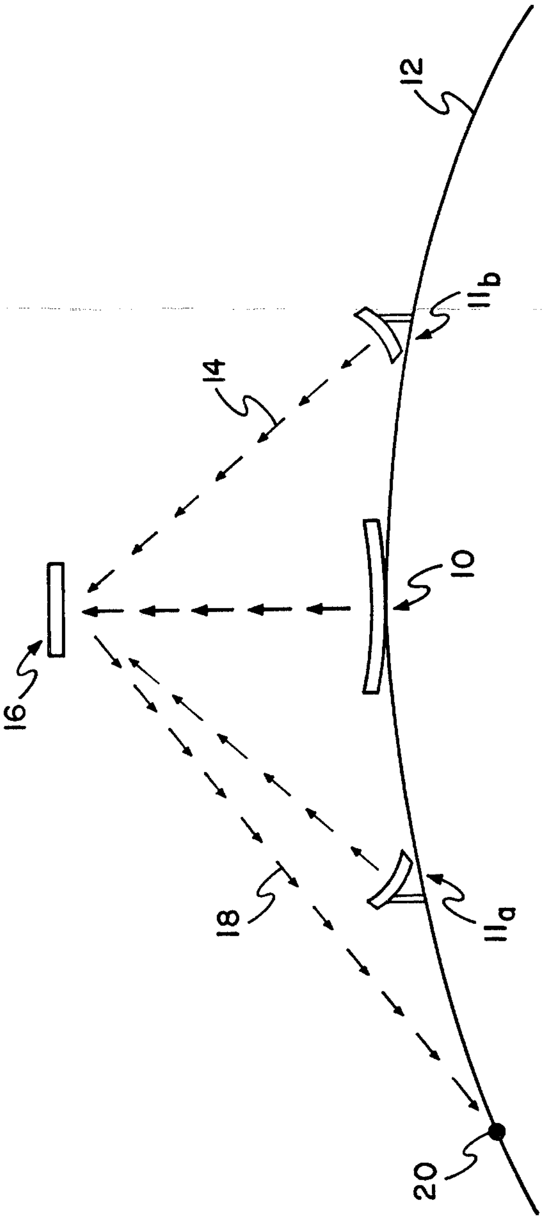
12
11b
14
10
16
18
11a
20

Fig. 5

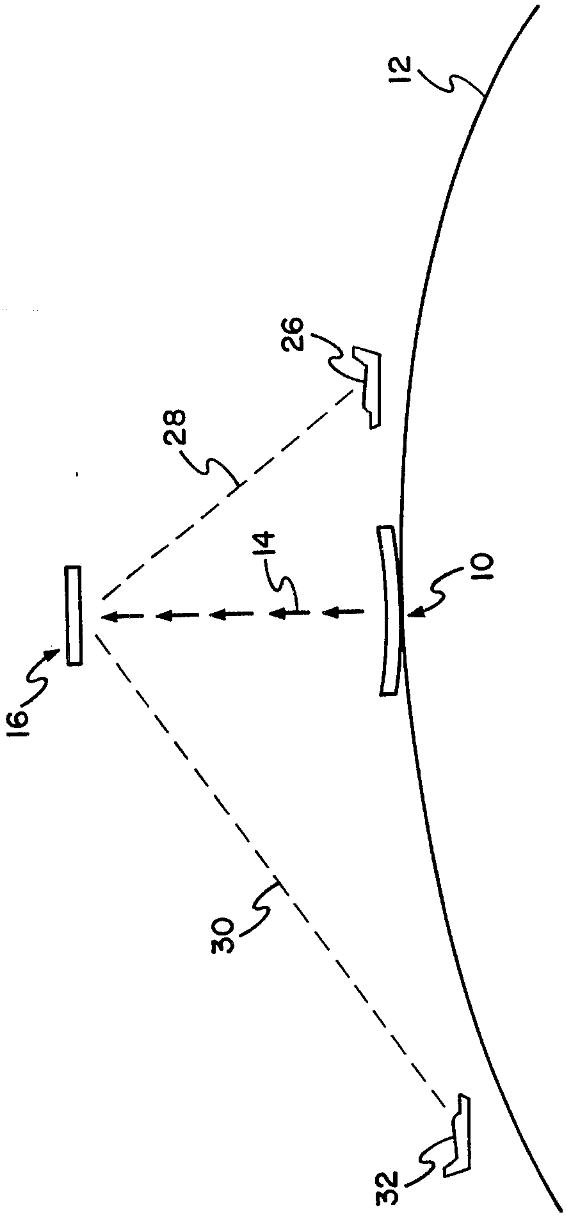
12
26
28
14
10
16
30
32

Fig. 6

COMMUNICATIONS SYSTEM USING A MIRROR KEPT IN OUTER SPACE BY ELECTROMAGNETIC RADIATION PRESSURE

This invention was made in the course of, or under, a contract with the Department of Energy.

BACKGROUND OF THE INVENTION

This invention relates to a communications system using an electromagnetic energy reflector or mirror kept in outer space by electromagnetic radiation pressure.

Good non-cable television reception generally requires line-of-sight transmission. To increase line-of-sight range, a television broadcasting station can increase the height of its transmission antenna. However, due to construction costs, weight, wind, and other factors, the height of a television antenna is limited. One method to increase the line-of-sight range of television reception is to locate relay stations to receive television signals by a long cable or line-of-sight transmission from a television broadcasting station and then to rebroadcast the signals to television sets in the vicinity of the relay station.

To extend the range of television reception the reflector mirror could be attached to an earth satellite. However, an earth satellite would, in general, move with respect to the surface of the earth and stay above the horizon only for a few minutes at a time (approximately less than 10 minutes when the height is approximately less than 600 kilometers). That, together with the expensive cost of satellite launching, makes this an unattractive possibility. Earth satellites, whether stationary or not, do not require energy transmitted from the earth to stay in orbit.

A stationary earth satellite equipped with a large reflector communications mirror would serve the desired purpose. Given sufficient power, it could beam programs to viewers located almost anywhere on about one-half of the earth's total surface. For such a satellite, the height is necessarily about 36,000 kilometers. If the wavelength equals 2 meters and the antenna's radius equals 1.2 kilometers, then the radius of the reflector communications mirror must be greater than approximately 36 kilometers. A mirror of this size would have to be assembled in orbit.

Expensive and sophisticated equipment is required to receive signals from existing communication satellites. Presently these reflected or retransmitted signals are eventually received by TV relay stations to be rebroadcast along line-of-sight to TV sets in the vicinity. Such satellites operate above the earth's ionosphere where ionospheric interference with signal transmission from and to the earth is found. Such satellites operate with decreased transmission efficiency because of the great distances involved. Thus even with satellites, most television viewers today are able to receive only programs beamed to them from a relay station located nearby. In those areas where there are no relay stations, or where existing stations do not transmit programs which some viewers wish to receive, good reception of desired programs is usually impossible. These difficulties could be avoided only if the need for local relay stations could be eliminated. There is also presently lacking an effective means for over-the-horizon radar and a means to impart a high electromagnetic radiation intensity to a wide area on the ground.

SUMMARY OF INVENTION

In view of the above, it is an object of this invention to provide line-of-sight electromagnetic wave transmission to areas lacking line-of-sight broadcasting or ground relay stations.

It is a further object of this invention to provide over-the-horizon electromagnetic wave transmission without the use of earth orbiting satellites or long cables.

It is a further object of this invention to provide for a transmission system to cover large areas of the earth with lower cost, lower power, and higher transmission efficiency than that of earth orbiting satellites.

It is a further object of this invention to provide television reception over large areas of the earth essentially without ionospheric interference with signal transmission.

It is a further object of this invention to provide for over-the-horizon radar and a means to impart a high electromagnetic radiation intensity to a wide area on the ground.

Various other objects and advantages will appear from the following description of an embodiment of the invention, and the most novel features will be particularly pointed out hereinafter in connection with the appended claims. It will be understood that various changes in the details, materials, and arrangements of the parts which are herein described and illustrated in order to explain the nature of this invention may be made by those skilled in the art without departing from the scope of the invention.

The invention comprises a method and system for transmitting electromagnetic radiation comprising disposing an antenna on the earth, transmitting electromagnetic radiation, having a diffraction peak and interference pattern, from said antenna generally vertically upward, disposing a dish shaped electromagnetic radiation reflector communications mirror, having a center of mass located below the point of attack of said transmitted electromagnetic radiation, at a height from about 100 kilometers to about 200 kilometers generally vertically above and oriented face down towards said antenna, supporting said mirror's weight by the electromagnetic radiation force transmitted by said antenna, controlling the height of said mirror by varying the intensity of said force, and reflecting said electromagnetic radiation or some other electromagnetic radiation with said mirror back toward the earth.

DESCRIPTION OF DRAWING

The present invention is illustrated in the accompanying drawing wherein:

FIG. 1 is a diagrammatic view of a communications system showing the signal transmission path and apparatus of the communications system embodying this invention;

FIG. 2 is a diagrammatic and perspective view of the reflector communications mirror;

FIG. 3 is a diagrammatic and perspective view of a multiple mirror system;

FIG. 4 is a diagrammatic view of a ground antenna arrangement or array composed of constituent antennas;

FIG. 5 is a diagrammatic view of a multiple antenna system for communication between different parts of the earth; and

FIG. 6 is a diagrammatic view of an over-the-horizon radar system.

DETAILED DESCRIPTION

In the following "television wave" shall mean an electromagnetic wave in the television or microwave range. In FIG. **1** a television transmitting antenna or antenna array **10** is located on the earth **12**. Television broadcasting equipment **13** transmits by the antenna **10** a television wave **14** vertically upward. Directly above the antenna **10** is a television wave reflector communications mirror or mirror system **16** at a height H above the earth's surface. The reflector mirror **16** reflects the transmitted wave **14** as a reflected wave **18** along a line of sight to the antenna of a home television set **20**, which is located at a distance L from the transmitting antenna **10**. The television set may be close to or over the horizon from the transmitting antenna. An obstacle **22** such as mountains may be between the television set and the transmitting antenna.

If the reflector mirror **16** is located at a height H of 100 kilometers above the earth **12** it will be in direct view of any viewer on the earth **12** within a distance L of about 1,100 kilometers as measured from the transmitting antenna **10**. Therefore, it would beam programs directly to the viewers. If the transmitting antenna **10** radiated approximately 10 or more megawatts of power **14**, then the reflected signal **18** would be strong enough to get behind most of the common obstacles, such as tall buildings, and thus ensure good reception almost anywhere within about 1,100 kilometers of the transmitting antenna **10**. In general, a reflector communications mirror located at height H above the earth **12**, would be above the horizon when viewed from within a circle of radius $L \approx (2HR_E)^{\frac{1}{2}}$, where R_E is the radius of the earth, provided that $L << R_E$.

With a transmitting antenna **10** of radius r_g located on the ground, and a reflector communications mirror **16** of radius r_a located at a height H above the ground, at least 84 percent of the power radiated by the transmitting antenna **10** will reach the reflector communications mirror **16** (all the power radiated into the first Fresnel Zone) provided that

$$r_a \gtrsim 0.6\lambda h/r_g \qquad \text{(Equation 1)}$$

where λ is the television carrier wavelength, and the reflector communications mirror is located at a height H above ground. For H=100 kilometers, λ=1 centimeter, and $r_g=300$ meters, a reflector communications mirror with a radius $r_a \gtrsim 2$ meters is needed, as derived from Equation 1.

The reflector communications mirror **16** may be made to reflect TV waves not to a relay station, but directly to those areas on the ground where the programs are to be viewed. A wide range of frequencies may be so reflected, no "tuning" of the reflector communications mirror **16** is necessary.

The reflector communications mirror **16** of radius r_a may be maintained at a height H solely by the radiation pressure **14** exerted on it by the electromagnetic waves which it reflects. A satellite is not needed to carry it. The equipment **13** may include means to sense variations in height of the mirror **16** and to adjust the power of radiation transmitted to control the height of the mirror **16**.

At normal incidence, 10 magawatts of radiation will exert a force of about 6.67 gram weight on a totally reflecting perfectly flat mirror. Real communication mirrors are neither perfectly reflecting, nor perfectly smooth. Nevertheless, for wavelengths of interest, deviations from the ideal case can be made relatively small, so that if the mirror weighs 6.67 grams, it will "float" on the radiation **14** impinging on it, whenever the radiation carries somewhat more than 10 megawatts of power.

When the mirror is located at or above 100 kilometers, it will be above most of the atmosphere. At this height the wind pressure on it is negligible. There are only two dominant forces acting on it: radiation pressure and gravitation. The mirror needs no rigid frame: the interplay of these two forces will "blow up" the mirror to its desired (for example, parabolic or shallow dish) shape, just as gravitation and wind pressure are known to blow up a parachute.

The communications mirror will be made light enough to float on the radiation it reflects. The mirror may consist of a set of parallel wires with a diameter 2r, located at a distance d from each other. When the wires are parallel to the electric field in the impinging wave, the transmissivity of such a mirror is

$$T \approx 4(d/\lambda)^2\{\ln|[0.83\exp(2\pi r/d)][\exp(2\pi r/d)-1]^{-1}|\}^2 \qquad \text{(Equation 2)}$$

A similar formula is given in N. Marcuvitz, *Waveguide Handbook* (Maple Press Co., York, Pa., 1951).

For λ=2 meters, d=5 centimeters, $2r=10^{-3}$ centimeters, Equation 2 gives a transmissivity T of 12.9%.

The wires will absorb some of the radiation falling on them. This energy will heat the wires. However, the heating will not melt the wires. A good conductor will absorb a fraction A of the energy impinging on its surface, where

$$A = c(\pi\sigma\delta)^{-1}2d^2/\pi^2 4r^2 \qquad \text{(Equation 3)}$$

Here σ is the conductivity of the conductor, c is the speed of light, and for wavelength λ, the skin depth is $\delta \approx (2\pi)^{-1}(c\lambda/\mu\sigma)^{\frac{1}{2}}$ where μ is the magnetic permeability. When $d/\lambda << 1$ then the mirror can be thought of as an approximately smooth surface in which a spatially uniform surface current is ocillating. The oscillations must cancel (1−T) fraction of the field energy behind the mirror. In reality, of course, the current is confined to the wires. However, if d/λ is small enough, the approximation will be a good one. In this approximation, the energy absorbed by the mirror is A times the energy reflected from the mirror.

A mirror made of a set of parallel wires would work. Nevertheless, it is often better to have a mirror made of two perpendicular sets of parallel wires in a checkerboard array, every wire in the first set being perpendicular to every wire in the second set, and all wires lying in the surface of the mirror **16** as shown in FIG. **2**. Each set contains the same number of wires and the spacing d between wires **24** in both sets is the same. In other words, we have a mirror **16** made of a wire mesh (like a screen) instead of a set of parallel wires. Such a mirror can better withstand mechanical stresses, and a rotation of the mirror (with respect to the incoming electric field **14**) can be ignored, as long as the surface of the mirror is left unchanged. As an example of such a mirror, consider a circular plane wire mesh with a mirror radius r_a as in FIG. **2**. The total length of the wires needed to make the mesh, when $d << r_a$, is $S \approx \pi r_a^2(d^{-2})2d$, so that total weight of mirror is

$$F_g = \pi r^2(2\pi r_a^2 d^{-1})\rho \qquad \text{(Equation 4)}$$

where β is the specific weight of the wire material. When ($\lambda=1$ centimeter) $d=7.66\times10^{-2}$ centimeters, $2r=10^{-4}$ centimeters, $r_a=2$ meters, and the wire material is aluminum, then $\rho=2.70$ gram/cubic centimeter, $S\approx3.2\times10^{6}$ centimeters, and $F_g=6.95\times10^{-2}$ grams. Thus the entire mirror can be supported by the pressure of 100 kilowatts radiation reflected from the mirror.

The total surface of all the wires in the mesh is B. If the mesh has temperature t, and is placed in outer space, then the power radiated by it is given by the Stefan-Boltzmann law:

$$W_t=(\epsilon)5.67\times10^{-12}\,(\text{Watts deg}^{-4}\,\text{cm}^{-2})t^4(B) \qquad \text{(Equation 5)}$$

Where ϵ is the emissivity of the wires surface. For copper, $\epsilon\approx0.15$ (somewhat higher for oxidized surface). If necessary, one can always increase the density of wires in the mirror (i.e., decreased). This will increase the weight of a unit mirror surface. Alternatively, one can reduce the wire radius, and if $2r\lesssim\delta$, use a multiple mirror system consisting of several mirrors located one above the other (see below). Yet another possibility is to use ribbon shaped wires rather than wires with circular cross section. If the thickness of the ribbons $\lesssim\delta$, one can lay a circular wire with $2r\gtrsim\delta$ on each ribbon to increase the mirror reflectivity.

The tensile strength of an aluminum wire is about 1.1×10^{6} grams/square centimeter. Thus an aluminum wire with $2r=10^{-4}$ cm can carry 8.64×10^{-3} gram weight, or 12% of the total mirror weight. Since there are now over ten thousand wires in this mesh, there will be no overstressing of the wires. Even if one-half of the mirror were to be pulled in one direction by some unforeseen force ten times as large as the weight of this half-mirror, and the other half pulled in the opposite direction by a force of equal magnitude the wires would on the average still carry less than 0.15% of the tensile strength. Since, furthermore, the wires are elastic, averaging is expected.

Meteors are not likely to be a major problem. The total area presented to a meteor by the mirror is $S\times2r$. For the parameter values chosen above, that means only $\approx3.2\times10^{-2}$ square meters. Should a meteor nevertheless strike a wire, the wire will break at the point of impact, unless the force along the wire is less than 8.64×10^{-3} gram weight. Only a small component of which will be perpendicular to the mirror surface. On the other hand, even a force of 8.64×10^{-3} gram weight is much less than the weight of the mirror, and can be counterbalanced by a small fractional temporary increase in the power radiated from the ground.

When the parameters are chosen as above ($H=10^{5}$ meters, $\lambda=1$ centimeter, $r_a=2$ meters) then Eq (1) would require that $r_g\gtrsim300$ meters. The value of r_g can be kept small in the following way.

Instead of placing a single mirror to a height H above ground, several parallel mirrors are placed there, one behind the other at a distance b from each other. This is shown in FIG. **3** for three mirrors: **16**$_a$, **16**$_b$ and **16**$_c$. This is particularly useful, if a large B value is desired, or when the wire diameter 2r is less than the skin depth δ for the radiation impinging on the mirror. As an illustration, suppose that $\delta/r=n$ which is >1. When $n>>1$, then a single mirror of the above type will transmit most of the radiation impinging on it, even if according to Equation 2, the transmissivity of the mirror would be small. The reason is that in equations like Equation 2, the wires are assumed to be perfect conductors, i.e., one approximates the correct result by putting $\delta=0$. This approximation is a good one only as long as $\delta<<2r$. When $\delta\gtrsim2r$, the radiation leaking through the wires becomes significant, and the transmission increases. To reduce this transmission, one may increase r. If that is done, the weight of the mirror will increase proportionally to r^2. By contrast—and this is what is suggested here—one may leave the wire radii unchanged, but place several mirrors of similar construction one above the other. Now each individual mirror will still have a large transmissivity, but when the number of mirrors, n, is such that $n\gtrsim\delta/2r$, then the transmissivity of the mirror system as a whole can be made small. The total weight of the whole mirror system is roughly proportional to n (exactly so if all mirrors are identical). This is to be compared with the previously mentioned case where the total weight was proportional to r^2 which in turn is proportional to n^2.

In practice the mirrors should be made similar, but not identical to each other: since less radiation reaches the higher lying mirrors, their reflection efficiency per unit weight has to be made higher than for the lower lying mirrors of the mirror system.

The wires may be deposited on thin threads of some light high-strength material such as a polyester resin material; for example a polyester of ethylene, glycol and terephthalic acid. Such materials may exhibit a specific weight considerably less than that of copper.

The ground antenna may be a sequence of parallel conductors, appropriately phased, to imitate a paraboloid. If the distance between adjacent wires is λ, then although the near field of the antenna will differ from that of a smooth paraboloid, the far field will not.

The mirror may be shaped so as to be stable with respect to horizontal drifting and tilting. If the mirror has the shape of a dish, facing down, then it will be stable with respect to horizontal displacements away from the center of the diffraction peak created by the transmitting antenna on the ground. The mirror could be circular, generally elliptical, or any other suitable configuration. If the mirror is displaced, e.g. to the left, the force exerted on its right side by the radiation will exceed that exerted on its left side. This force will have a horizontal component, pushing the dish back to the center. Possible oscillations will be damped by the rare atmosphere around the mirror. Should higher damping be desired, the mirror may be lowered temporarily, by decreasing antenna power, to denser atmosphere at times when the wind there is moderate. Slowly displacing the axis of the interference pattern created by the ground antenna, will, by the same argument, induce a corresponding displacement of the floating mirror. Thus, its lateral as well as vertical position could be controlled from the ground. Stability against tilting can be insured by arranging the center of mass of the mirror to lie below the point of attack of the resultant lift force. This may be achieved by making the perimeter of the dish heavier than its center, or, better yet, by hanging a small weight from its periphery in a manner of a weight hanging from a parchute. In either case, the radiation will "blow up" the mirror to its desired shape, and will keep it facing down. The radii of curvature of the dish will determine the shape of the area in which it will radiate. There is no need to make all radii equal: the perimeter of the dish may be e.g. an ellipsoid. Usually, smaller radius of curvature will mean larger area illuminated.

The mirror shape and effective radius can be varied by rotating the mirror around its axis at various angular velocities. The angular velocity can be governed from the ground by rotating the plane of polarization of the radiation impinging on the mirror. For example, flaps could be hung from the edge or perimeter of the mirror which would gradually open up as the angular velocity increases.

While reference is made to the antenna **10** being located "on the earth," one can readily appreciate that it could be supported by a tower, airplane, ship, or the like.

The electromagnetic radiation transmitted by the antenna **10** could include microwaves, radar, television, short wave radio, medium wave radio, and long wave radio transmissions, as well as radiation of shorter wavelength such as visible light. The wavelength of the electromagnetic radiation signal that is being transmitted by mirror **16** will be effected by the spacing between the wires of mirror **16** and possibly their diameters.

The reflector mirror would be disposed at a height below about 36,000 kilometers which is the height of a stationary earth orbiting satelite. Below about 36,000 kilometers radiation force is needed to support the reflector mirror. The preferred height is about 200 kilometers for size and power considerations.

The antenna **10** may be a circular dish ground antenna. However, in many cases it is desirable to use not a circular dish ground antenna, but rather an array of smaller "constituent" dish antennas (items $\mathbf{10}_a$, $\mathbf{10}_b$, $\mathbf{10}_c$ etc., as shown in FIG. **4**) placed next to each other along two straight axes on the ground arranged in a cross shape. Additional constituent antennas may be placed between the lines as shown in FIG. **4**. The constituent antennas need not all be circular nor of the same size. Furthermore, for many applications it is desirable to have the radiation intensity emitted by the smaller constituent antennas vary for example as a Gaussian, as a function of the location of the constituent antennas; in an xy cartesian coordinate system, the center position of any constituent antenna would be denoted as (x,y). The intensity of that constituent antenna is then given by the product of two gaussians, one as a function of x with width σ_x the other as a function of y with width σ_y (σ_x may be choosen to be equal to σ_y).

In some cases it may be desirable to have a ground antenna system consisting of several antennas (with each antenna consisting of one or more constituent antennas) located at large distances (comparable to H) from each other. FIG. **5** shows a three antenna **10**, $\mathbf{11}_a$, and $\mathbf{11}_b$ system. Such arrangement will allow greater flexibility in choosing the final position of the mirror by varying the intensity of the electromagnetic radiation radiated by each antenna.

In FIG. **5**, each antenna **10**, $\mathbf{11}_a$, and $\mathbf{11}_b$ does not have to transmit electromagnetic radiation of the same wavelength. Thus several television channels of information may be transmitted. A reflector communications mirror **16** may be designed such that certain parts or segments of the mirror would reflect electromagnetic radiation of certain wavelengths better than other wavelengths. For example, a mirror may be constructed with three corners such that each corner would reflect microwaves best and the center of the mirror would reflect longer wavelengths best. Also, referring to FIG. **4**, each constituent antenna such as $\mathbf{10}_a$ may transmit at several different frequencies. Referring to FIG. **5**, the intensity of the transmitted radiation from each of the three antennas **10**, $\mathbf{11}_a$, and $\mathbf{11}_b$ may be chosen such that the three antennas act not only to transmit electromagnetic radiation to a desired point **20** but also to support the mirror **16**. As an option, the intensity of the electromagnetic radiation of the antenna **10** directly underneath the mirror **16**, in FIG. **5**, may be chosen to be much greater than the intensity of the other two antennas, $\mathbf{11}_a$ and $\mathbf{11}_b$. Thus the electromagnetic radiation from the antenna directly underneath the mirror would be used mainly to support the mirror while the desired information signals would come from the other antennas, to be reflected by the mirror **16** to the desired point **20**.

An over the horizon radar system is depicted in FIG. **6**. In this system the reflector mirror **16** is supported solely by antenna **10**. The electromagnetic radiation **14** from the antenna **10** is much larger than any other radiation impinging on the mirror **16**. An airplane **26** emits a radar signal along path **28** which is reflected by mirror **16**, and the reflected signal along path **30** strikes a target **32**. The radar wave is reflected by the target **32** along path **30** to again strike mirror **16** to be finally reflected back along path **28** to a receiving station located on the airplane **26**. Of course it is to be understood that the radar signal could be emitted and received from a point near antenna **10**.

The reflector mirror may be kept afloat by radiation pressure at altitudes lower than where satilites are generally located. In particular, it may be kept below most of the ionosphere, so that the ionosphere's interference with signal transmission may be essentially eliminated. Close proximity to the earth allows higher transmission efficiency from the earth to the mirror and back.

If the ground antenna emits high enough radiation and if the communications mirror is large enough, radiation intensity reflected from the mirror may reach dangerous levels on the ground. If that is not desired, the danger may be eliminated by restricting the mirror's radius. Thus the mirror may be made perfectly safe. Actually, the radiation from a properly designed mirror will be environmentally even safer than the radiation emited by the large number of relay stations now in use. The near fields of the existing relay stations may be dangerous, whereas the near field of the communications mirror is located far from all human habitat and is thus harmless. On the other hand, the far field of the mirror may be designed to assume essentially the most desired value on the ground.

For launching the mirror, it may be placed on a frame or in a container while being lifted by a balloon or series of balloons through the lowest layers of the atmosphere. This will provide protection from the forces due to wind. As the balloon rises, air density decreases and, consequently, so does the force exerted by the wind on the mirror (or its container). At some height (part of) the frame or container can be released and the remaining lighter load raised further. The frame or container may be released in several stages. Eventually, the mirror is released and final positioning is done by varying the radiation pressure from the ground.

What we claim is:

1. A method for transmitting electromagnetic radiation signals comprising transmitting electromagnetic radiation, having a diffraction peak and interference pattern, from an antenna located on the earth generally vertically upward; disposing a dish shaped electromagnetic radiation reflector communications mirror, having a center of mass located below the point of attack of said transmitted electromagnetic radiation, at a height of from about 100 kilometers to about 200 kilometers generally vertically above and oriented face down towards said antenna so as to receive said transmitted electromagnetic radiation; supporting said mirror by the force produced by said electromagnetic radiation transmitted by said antenna; controlling the height of said mirror by varying the intensity of said force; and reflecting electromagnetic radiation signals with said mirror back toward the earth.

2. The method of claim 1 wherein said electromagnetic radiation signals have a wavelength from about that of microwaves to that of radio waves.

3. The method of claim 1 wherein said mirror reflects said electromagnetic radiation signals back towards the earth over the horizon from said antenna.

4. The method of claim 1 further including disposing an array of said antennas on earth configured so that the resultant total electromagnetic radiation is transmitted generally vertically upward, and disposing said mirror generally vertically above said resultant total electromagnetic radiation.

5. The method of claim 4 wherein one or more antennas are transmitting at different wavelengths.

6. The method of claim 1 further including disposing a plurality of said mirrors vertically positioned at different heights.

7. The method of claim 1 further including adjusting the lateral position of said mirror by displacing the axis of the interference pattern created by said antenna's electromagnetic radiation.

8. The method of claim 1 wherein said electromagnetic radiation includes plane-polarization and further including disposing a plurality of flaps on the edge of said mirror and rotating the plane of polarization of said radiation to rotate said mirror to open up said flaps to change the shape and effective radius of said mirror.

9. A system for transmitting electromagnetic radiation signals comprising an antenna means disposed on the earth for generally vertically transmitting electromagnetic radiation having a diffraction peak and interference pattern; means for generating electromagnetic radiation and transmitting the same through said antenna; means for varying the intensity of said electromagnetic radiation; and a dish shaped electromagnetic radiation reflector communications mirror, having a center of mass located below the point of attack of said transmitted electromagnetic radiation, at a height, of from about 100 kilometers to about 200 kilometers generally vertically above and oriented face down towards said antenna, at which said mirror's weight is supported by the transmitted electromagnetic radiation force from said antenna, to effect reflection of said electromagnetic radiation signals with said mirror back towards the earth.

10. The system of claim 9 wherein said electromagnetic radiation signals have a wavelength from about that of microwaves to that of radio waves.

11. The system of claim 9 wherein said mirror comprises two generally perpendicular sets of generally equally spaced parallel wires having a checkerboard array.

12. The system of claim 12 wherein the wires of said mirror are of generally circular cross section having diameters substantially smaller than the spacing between said generally parallel wires.

13. The system of claim 9 wherein said antenna means includes an array of antennas wherein said array is positioned to provide generally vertically upward resultant electromagnetic radiation, wherein said mirror is generally vertically above said electromagnetic radiation.

14. The system of claim 13 wherein said antennas of said array transmit said electromagnetic radiation at different radiation wavelengths.

15. The system of claim 13 wherein said mirror compromises a plurality of mirrors wherein said mirrors are disposed at different lateral positions generally vertically above said array of antennas.

16. The system of claim 9 further including a plurality of said mirrors with all mirrors positioned at a different height generally vertically above said antenna means.

17. The system of claim 9 further including a separate radar apparatus, having substantially less power than said antenna means, to effect the detection of an over-the-horizon target with said mirror.

18. The system of claim 9 wherein said electromagnetic radiation includes plane-polarization and further including a plurality of flaps depending from said mirror and means for rotating the plane of polarization of said radiation to effect the rotation of said mirror to effect the opening up of said flaps to effect the change of the shape and effective radius of said mirror.

* * * * *

United States Patent [19]
Ford

[11] **Patent Number:** **4,684,255**
[45] **Date of Patent:** **Aug. 4, 1987**

[54] INTERFEROMETRIC OPTICAL PATH DIFFERENCE SCANNERS AND FT SPECTROPHOTOMETERS INCORPORATING THEM

[75] Inventor: **Michael A. Ford**, Maidenhead, England

[73] Assignee: **The Perkin-Elmer Corporation,** Norwalk, Conn.

[21] Appl. No.: **760,013**

[22] Filed: **Jul. 29, 1985**

[30] **Foreign Application Priority Data**

Aug. 9, 1984 [GB] United Kingdom 8420233

[51] **Int. Cl.**^{4} **G01J 3/45**
[52] **U.S. Cl.** **356/346**
[58] **Field of Search** 356/346

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,142,797 3/1979 Astheimer 356/346

OTHER PUBLICATIONS

Sternberg et al., "A Newtype of Michelson Interference Spectrometer", *J. Sci. Instru.* vol. 41, pp. 225–226, 4/64.

Primary Examiner—Davis L. Willis
Assistant Examiner—Matthew W. Koren
Attorney, Agent, or Firm—Francis L. Masselle; Edwin T. Grimes; Ronald G. Cummings

[57] ABSTRACT

In an interferometric apparatus the path of the output beam of a rotary Optical Path Difference scanning assembly leading to an electrically responsive detector is maintained substantially fixed in space for any scan angle within predetermined limits, thus obviating a serious drawback of a prior art proposal wherein the output beam translates significantly over the detector face with changes in scan angle, with the result that the detector output is vitiated by totally spurious variations of the energy reaching it.

In one embodiment, the scanning assembly comprises a beam splitter having a semi-reflective layer the output face of which cooperates with the face of an output mirror normal thereto, the axis of rotation being substantially coincident with the line of intersection between the prolongation planes of the two faces. A Fourier Transform spectrophotometer embodying the assembly is also described.

17 Claims, 11 Drawing Figures

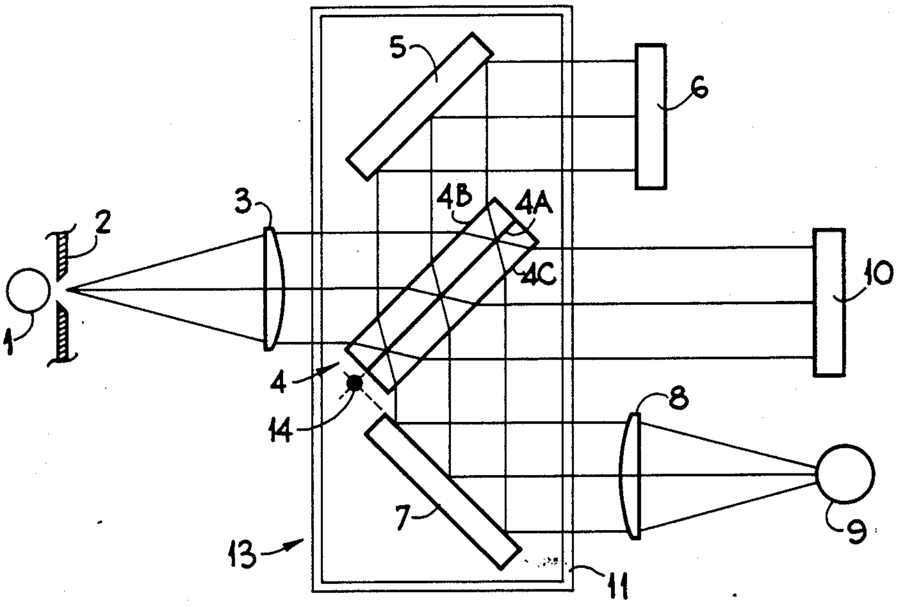

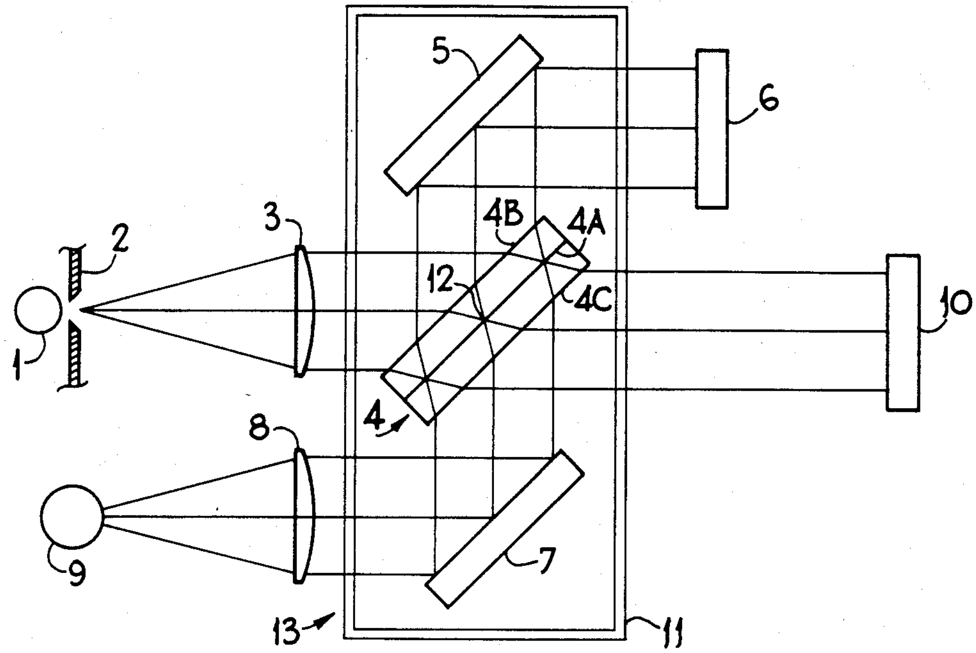
5
6
3
4B
4A
2
12
4C
10
1
4
8
9
7
13
11

FIG. 1 *PRIOR ART*

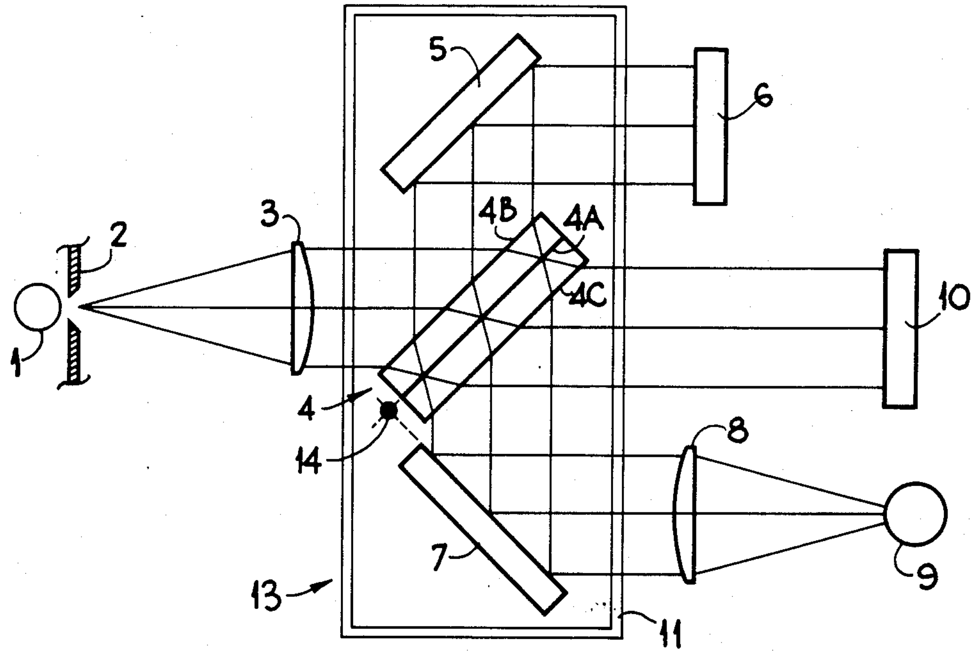
5
6
3
4B
4A
2
4C
10
1
4
14
8
9
7
13
11

FIG. 2

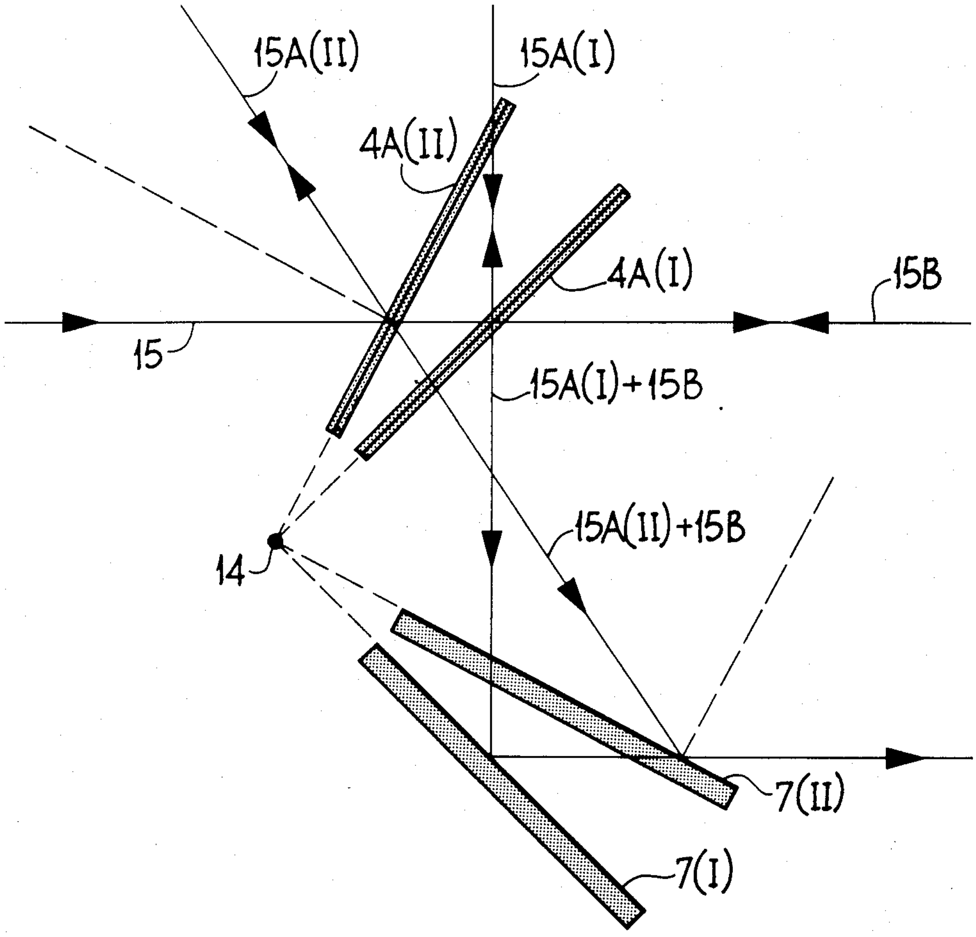
15A(II)
15A(I)
4A(II)
4A(I)
15B
15
15A(I)+15B
15A(II)+15B
14
7(II)
7(I)

FIG. 3

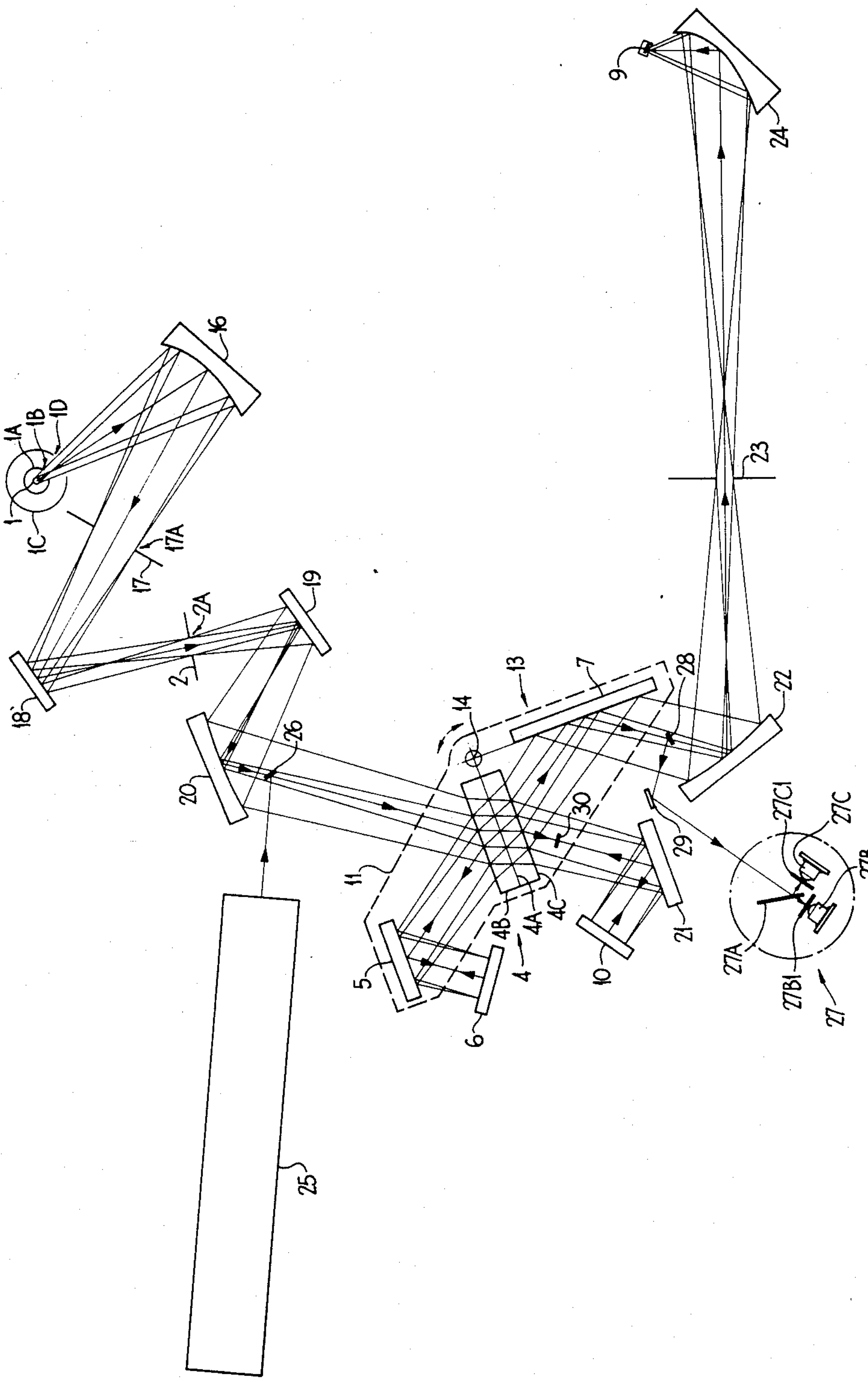
1A
1B
1D
1C
2A
27C1
4B
4A
4C
27B1

FIG. 4

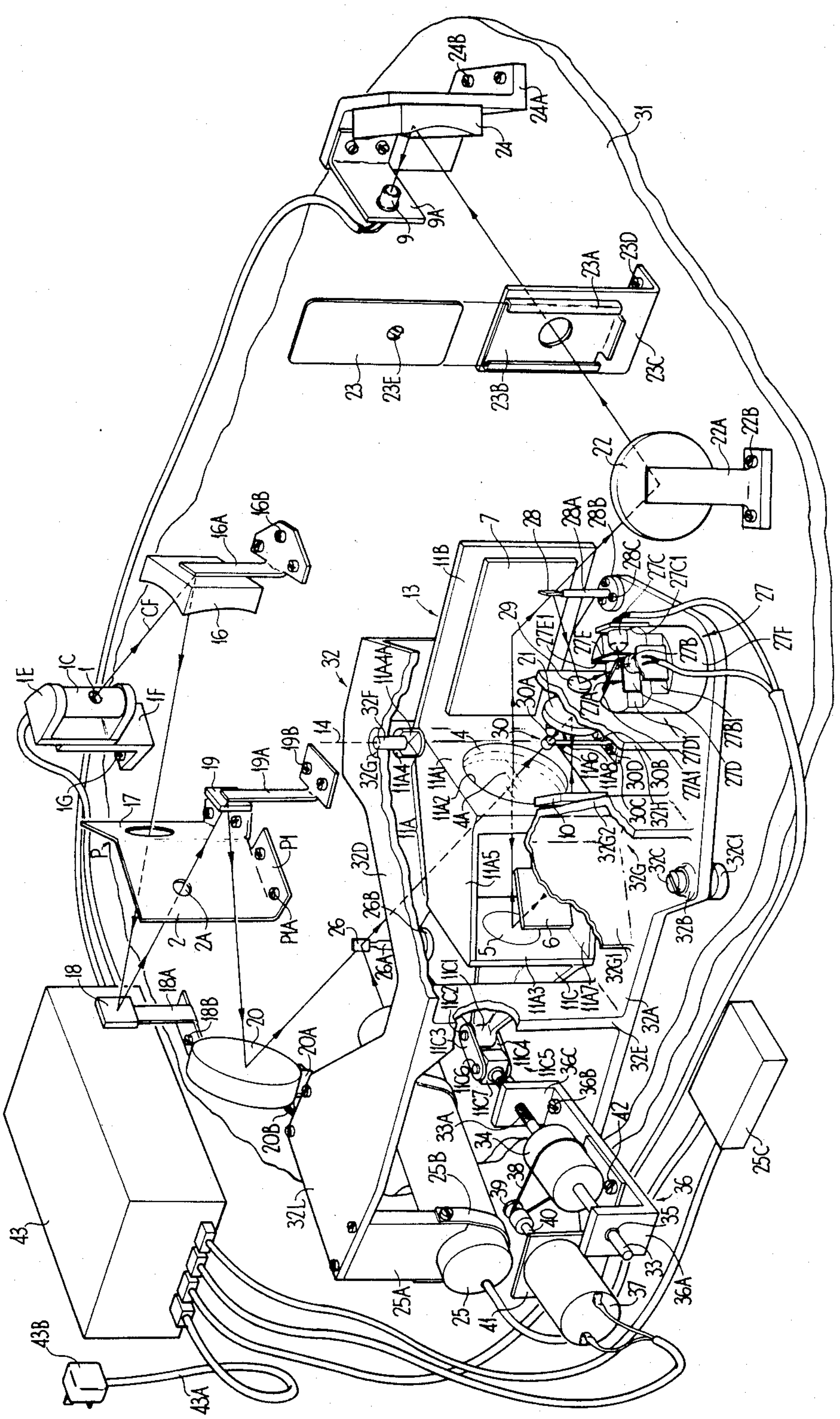
9A
CF
1E
1C
1F
1G
P
P1
P1A
2A
27C1
27F
27E1
27A1
27D1
27B1
32H
11A4A
32F
32G
11A4
11A1
11A2
4A
11A6
11A8
32G2
32G
32C1
11A5
32G1
11A3
11A7
11C1
11C2
11C3
11C4
11C5
11C6
11C7
32L

FIG. 5

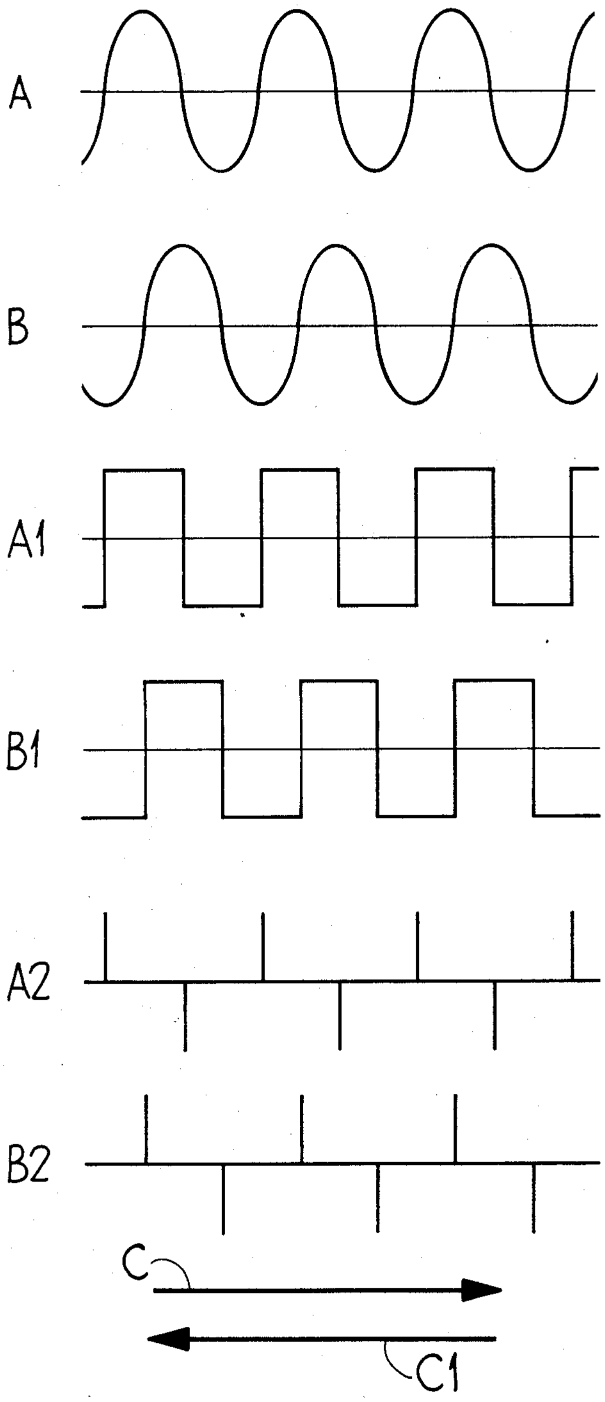
A
B
A1
B1
A2
B2
C
C1

FIG. 6

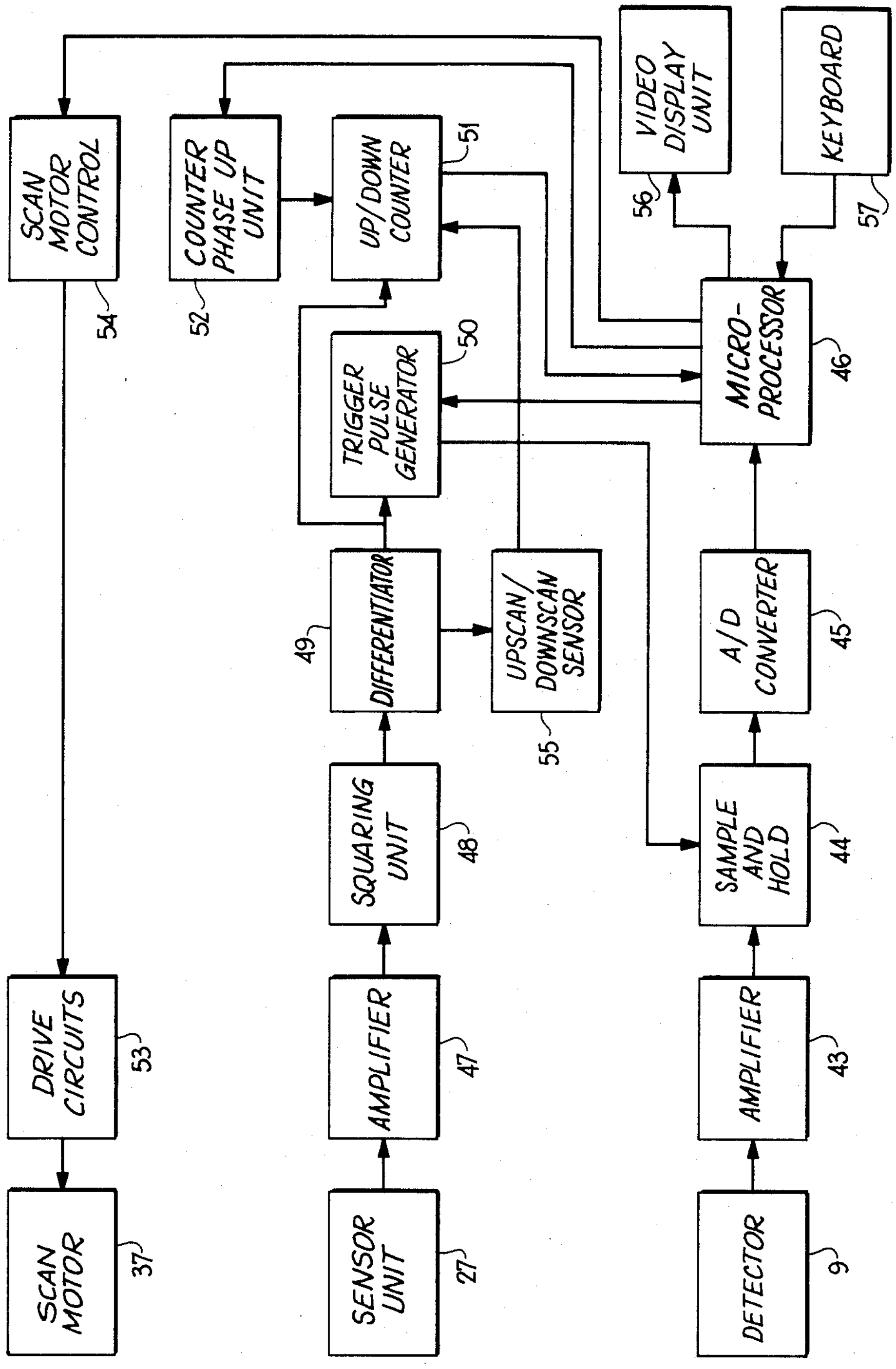
SCAN MOTOR CONTROL
54
COUNTER PHASE UP UNIT
52
UP/DOWN COUNTER
51
VIDEO DISPLAY UNIT
56
KEYBOARD
57
MICRO-PROCESSOR
46
TRIGGER PULSE GENERATOR
50
DIFFERENTIATOR
49
UPSCAN/ DOWNSCAN SENSOR
55
A/D CONVERTER
45
SQUARING UNIT
48
SAMPLE AND HOLD
44
DRIVE CIRCUITS
53
AMPLIFIER
47
AMPLIFIER
43
SCAN MOTOR
37
SENSOR UNIT
27
DETECTOR
9

FIG. 7

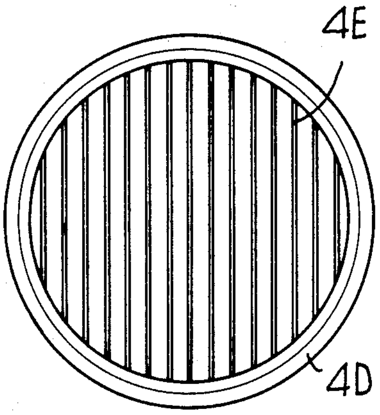

FIG. 8A.

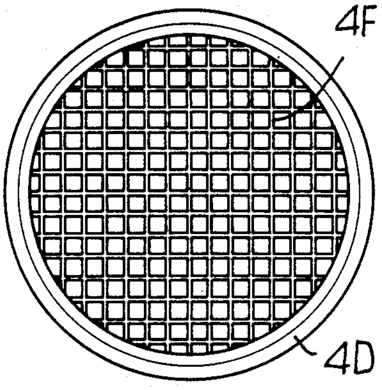

FIG. 8B.

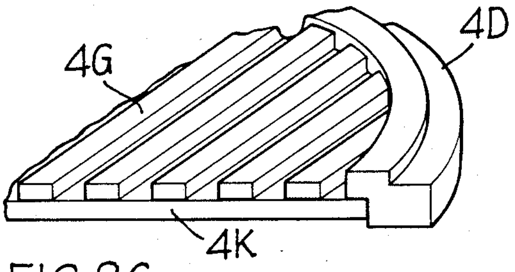

FIG. 8C.

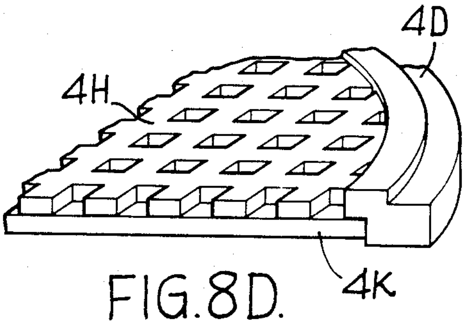

FIG. 8D.

INTERFEROMETRIC OPTICAL PATH DIFFERENCE SCANNERS AND FT SPECTROPHOTOMETERS INCORPORATING THEM

BACKGROUND OF THE INVENTION

This invention relates to interferometric apparatus comprising a rotary scanner of the optical path difference between two optical paths, such as the two arms of an interferometer of the Michelson type, in particular a Michelson interferometer forming part of a Fourier Transform Spectrophotometer, the rotary scanner being such as to enable the beam outputted therefrom to remain substantially fixed in space for any angle of rotation within predetermined scanning limits.

It is believed that, although the phrase "optical path difference" and the related phrases "optical path length" and "geometrical path length" presently to be encountered are well understood in the interferometric art, it would be of assistance to the reader in appreciating their significance in the present context if they were distinguished from the outset.

We will distinguish first the phrases "geometric path length" and "optical path length", in which the path referred to is that followed by light rays between two points spaced along their travel or, more specifically, two spaced optical elements. The "geometrical path length" is simply the length, in any convenient units of linear measure, of the path through air, or any other medium, from one to other of said points. The "optical path length", on the other hand, takes into account the retardation effect on light waves of any intervening medium other than air. By definition, it is the product of the geometrical path length and the refractive index of the medium. The refractive index (usually denoted by the letter n) is in fact a measure of the retardation. If a light wave takes a time t to travel between two spaced points in air, it will take a time $t \times n$ to travel between the same two points when a medium of index n greater than 1 is substituted. If $n=2$, for example, the travel time is doubled.

It will be appreciated, therefore, that two light waves initially travelling in phase coincidence along two separate paths of the same geometrical path length (if no media other than air are included in the paths) or optical path length (if media other than air are present) are thrown out of step if one is retarded with respect to the other by increasing the geometrical path length or the optical path length, as the case may be, of one path with respect to the other.

A system of mirrors can clearly be used to change the geometrical path length simply by increasing or decreasing the distance travelled in air by the light rays. It follows that even where refractive elements are present the system will still alter the geometrical length of that portion of the overall path that passes through air and, therefore, it will alter the optical path length, in so far as the latter is derived from the former in the manner hereinbefore referred to.

Two optical paths may have the same geometrical path length and yet one path will cause greater retardation than the other if its optical path length is greater. On the other hand, equality of optical path lengths is always associated with equal retardations. This means that the phrase "optical path length" has greater generality and may conveniently be used in a context where refractive elements may or may not be involved.

As to the phrase "optical path difference", it simply relates to the difference in optical path length between two paths that may include media other than air. The phrase "geometrical path difference" is not often used, but after what has been said in the preceding paragraph, its meaning should be clear.

In the present context, the phrases "optical path length" and "optical path difference" (hereinafter frequently referred to and conveniently denoted by the initials O.P.D.) are intended to be associated with situations where the presence of media other than air is not to be excluded but not necessarily included.

In a basic Michelson-type interferometer, the radiation flux emanating from a source of monochromatic light, either in the visible or the invisible regions of the electromagnetic spectrum, impinges on a beam splitter after passing through a collimator. The beam splitter is oriented at an angle of 45 degrees to the direction of the incoming collimated beam so that a portion (i.e. percentage) of the impinging beam is transmitted without change of direction and a portion is reflected through a 90-degree angle. Each portion is reflected back towards the beam splitter by a beam reversing mirror, the return optical path being coincident with the forward path. At the beam splitter recombination of the two portions takes place, with the first now progressing in reflection and the second in transmission towards a common path leading to a light receiving device, which may be a screen or a photoelectric detector.

Now, if the optical path length between the beam splitter and one reversing mirror, which together represent one arm of the interferometer, is made identical with that between the beam splitter and the other reversing mirror, which together represent the other arm of the interferometer, the reflected and transmitted beams will suffer no relative phase changes and consequently will recombine at the beam splitter so as to be indistinguishable—if transmission losses and optical imperfections are notionally assumed to be zero and the beam splitter provides 50% reflection—from the unsplit optical beam.

If, on the other hand, an optical path difference that is significant compared with one half wavelength of the monochromatic light is established between the two arms of the interferometer, such as by moving one of the reversing mirrors towards the beam splitter, interference effects may be expected. The frequency of the resultant sine wave representing the recombined beam will be the same as that of the two constituent sine waves representing the beams prior to their recombination. The amplitude, on the other hand, will be the algebraic sum of the constituent sine waves amplitudes at each successive abscissa point and therefore will vary sinusoidally from a maximum, when the optical path difference is equal to an integral number of wavelengths, and, therefore, the amplitudes of the constituent waves add up (constructive interference), to a minimum, when the optical path difference is equal to an integral number of wavelengths plus one half wavelength and the waves are therefore in phase opposition so that the amplitudes subtract (destructive interference). A series of concentric rings (called interference fringes), representing the maxima and minima of the resultant sine wave, and therefore alternately bright and dark, may be viewed by interposing an ordinary white diffusing screen at the focus of a converging lens interposed in the path of the recombined beam, provided the monochromatic source radiates visible light.

Alternatively and more usefully, the optics of the interferometer may be arranged to project an image of the aperture stop (also called Jacquinot stop) of the interferometer at a photodetector having a circular window for admitting the central interference fringe (i.e. the first ring). This fringe will change from a light to a dark patch as the optical path difference undergoes a change equal to one half wavelength through the movable mirror being displaced between positions corresponding to constructive interference and destructive interference, respectively. The sinusoidal electrical output of the detector can be processed to provide an accurate linear measurement of the mirror displacement and, therefore, of any member suitably mechanically coupled to the mirror, to within a small fraction of one half wavelength. Displacements of exact half wavelengths can, of course, be measured by counting the mean level crossings of the output signal from an appropriate datum. Each crossing counted from the datum represents a change of O.P.D. of exactly half the wavelength of the source, i.e. a displacement of the member of exactly ¼ wavelength. When a monochromatic source is used the wavelength of which is known precisely, interferometric measurements of length as suggested are perhaps the most sensitive and accurate available to man.

A field of application predicted by Michelson for his interferometer, but only of comparatively recent application, is Fourier Transform spectroscopy (hereinafter FT stands for Fourier Transform). If the movable mirror referred to earlier is actually displaced between a position corresponding to zero O.P.D. and a predetermined limit by mounting the mirror on an accurate scanning assembly following a strictly rectilinear path, the output of the photo-detector will be a sine wave of much lower frequency compared with the monochromatic emission line of the source but constant peak-to-peak amplitude. If it is assumed that the wavelength of the optical input to the detector is, say, 2.5 micrometers and the predetermined limit of the optical path difference is 2 cm then clearly the number of recurrent cycles in the detector output must be 2 cm/0.00025 cm=8000. If the time taken to move the mirror between zero optical path difference and the limit is 1 second, then the frequency of the detector output signal is 8000 Hz. In other words, the original optical wave has been modulated at a frequency of 8000 Hz, well within the audio range. The modulation sine wave, which may be displayed on a CRT, represents the interferogram of the source. If the Fourier Transform of the interferogram is computed, the resulting trace, in the form of a very narrow band, represents the emission line spectrum of the source.

If a broad-band source, e.g. an infrared source, is substituted for the monochromatic source, the output of the detector when the scanning mirror is in motion will no longer be a pure sine wave since the emission spectrum of the source will include waves of different frequencies. Each of the optical sine waves will give rise to two constituent beams and one resultant beam in the manner described earlier and all the resultant beams of different frequencies will be represented in the instantaneous output of the detector, each by its own modulation sine wave of related frequency and amplitude. If the output of the detector is plotted as before, the trace that results represents the emission interferogram of the source.

Now, if a sample not opaque to infrared is placed at the Jacquinot stop, the waves of different frequencies present in the emission spectrum of the source are attenuated to a different extent in a manner that is characteristic of the chemical nature of the sample and the resulting interferogram represents the infrared absorption interferogram of the sample superimposed on the emission interferogram of the source.

By taking the Fourier Transforms of the two interferograms, thus obtaining independently the spectrum of the sample-cum-source and that of the source alone, and then ratioing the first spectrum by the second, the spectrum of the sample is derived. In general the interferogram will consist of a dominant centreband flanked by intermingled waves decrementing more or less symmetrically in amplitude to vanishing values. It is worth observing that there is no point of communality between the various sine waves other than at the position of the scanning mirror corresponding to zero O.P.D. where all waves will undergo constructive interference; and that, of course, is the reason for the presence of the dominant centreband.

It can now be readily appreciated that, whatever the application of the Michelson-type interferometer which necessitates the optical path difference between the two interferometric arms to be changed, a very serious mechanical problem arises.

If mechanical imperfections in the guideways cause the moving mirror to tilt during the scan motion, the effective optical path difference may not be the same for each of several parts of the beam traversing different regions of the beam splitter, with the result that for a given wavelength some parts may be undergoing constructive and some destructive interference. In consequence, the amplitude of the detector signal is spuriously changed compared with the situation where no tilt is present. Servo controls in the form of tilt compensators have been developed but they account for a major portion of the complexity and cost of the entire instrument and can easily get out of adjustment. This is not surprising when it is realized that in a translating-mirror Michelson interferometer the tilt tolerance is typically one or two arc-seconds!

The problem is particularly severe in FT spectrophotometers where a scanning assembly reciprocating along a path of the order of centimeters must be provided. In many experimental and commercial instruments, a rectilinear reciprocating motion is chosen and a great deal of ingenuity has gone into devising means for guiding the assembly so as to confine its motion to a strictly rectilinear path. Linear air bearings and servo controls have been adopted in sophisticated and, therefore, expensive systems.

A practical proposal for a simple rotary scanning assembly which circumvents the major difficulties encountered with rectilinear scanning assemblies and is inherently tilt-compensated to some extent was described by R. S. Sternberg and J. F. James, of the Physical Laboratories of Manchester University, England, in the Journal of Scientific Instruments, Vol. 41, pages 225-226, April 1964, under the title "A new type of Michelson interference spectrometer". In the introduction to their article, the authors stress the drawbacks of scanning arrangements known at the time before discussing their solution. Unfortunately, although their proposal represents an interesting and valuable approach to the problem of lessening mechanical accuracy requirements, it has not been adopted by manufacturers of FT spectrophotometers, probably because it suffers from a serious drawback, now totally eliminated by the present invention.

In modern FT spectrophotometers, the analogue signal generated by the detector is digitized and processed by a microprocessor. It is an important requirement to ensure that the path of the beam reaching the detector does not change with the angular position of the scanning assembly or the optical throughput to the detector will change in a manner which is in no way related to the analysis of the spectrophotometric sample and does in fact vitiate analytical results. It will later be shown with the aid of a drawing, before the disclosure of practical embodiments of the present invention, that such requirement is not met by the prior art proposal, in which the optical output from the scanning assembly suffers a translatory displacement, with the result that, starting from a position of the scanning assembly correspondingly to which the entry pupil of optical converging means before the detector is totally filled, more and more of the rays will miss the pupil altogether as scanning progresses from that position in either rotational direction. The drawback is of course tolerable where the analogue changes in the signal need not be measured accurately, such as when the detector output is merely used for the purpose of counting interference fringes, but is certainly unacceptable in an FT spectrophotometer wherein the spectrum of the sample must be drawn with band amplitude fidelity of a high order.

SUMMARY OF THE INVENTION

The object of the present invention is to provide interferometric apparatus including a rotary scanning assembly for varying the optical path difference between two optical paths and producing an output beam, including contributions from the two paths, which is maintained on a substantially constant path for any angle of the scanning assembly within a design range and in addition to provide an improved Michelson-type interferometer, in particular an FT spectrophotometer, incorporating said scanning assembly.

In broad terms, the above object is achieved by providing the rotary scanning assembly with an output mirror the orientation of which in conjunction with the selection of the rotational axis of the assembly enables the function referred to above to be fulfilled.

In accordance with one aspect of the present invention there is provided an interferometric apparatus comprising a rotary O.P.D. scanning assembly mounting a mirror pair formed by a beam splitter and a folding mirror, said beam splitter having a semi-reflective beam-splitting layer one face of which is the optical input surface in accurately parallel face-to-face relationship with said folding mirror and the other the output surface. Two stationary reversing mirrors co-operate with said mirror pair for (a) defining two interferometric arms, each arm receiving one of the two beams resulting from the splitting of an input beam by the beam splitter, (b) rotationally scanning successive O.P.D. values between the two beams up to a predetermined maximum, and (c) recombining the beams at said output surface so that they emerge therefrom as a single beam directed along an output path. An output mirror is provided which is mounted at right angles to the output surface of said beamsplitting layer, said output mirror being borne by the scanning assembly and positioned to receive the output beam. Suitably arranged pivotal means ensure that the scanning assembly rotates around an axis substantially coincident with the line of intersection of the prolongation planes of the semi-reflective beam-splitting layer and the face of the output mirror, respectively.

In the interferometric apparatus as defined, the specified geometric relation between the semi-reflective layer and the output mirror in conjunction with the positioning of the rotational axis of the scanning assembly enable the output beam to be directed along an output path that is substantially immutable in space for any scan angle within the design range of the assembly.

Because the output beam may be utilized in conjunction with a photosensitive device of very small active area, it may be essential for it to have particularly high spatial constancy. The T.G.S. (triglycine sulphate) detector of infra-red radiation, which is now often chosen by designers because of its high sensitivity and comparatively fast response, is such a device. With an active circular surface area only one or two millimeters in diameter, it clearly cannot tolerate any significant lateral displacement of the beam impinging thereon, even accounting for the fact that in certain applications the problem is lessened by the active area receiving in fact a demagnified stop image, e.g. demagnified Jacquinot stop image in the case of FT-IR spectrophotometers. (IR stands for Infrared).

The outstanding advantage of the part played by said specified geometric relation within the overall co-operation of the optical elements defined is that it has made the adoption of the rotary O.P.D. scanner a real feasibility in all those applications where the spatial inconstancy of the output beam associated with the prior art rotary scanner had hitherto prevented it—first and foremost in FT-spectrophotometry.

A consequence of the specified geometric relation referred to is that the beam splitter is not rotated around an axis of symmetry. This means that the cross-section of the input beam must translate across the face of the beam splitter as the scanning assembly is rotated. But even if the rotational axis were an axis of symmetry co-planar with the semi-reflective layer, the projected area of the input beam cross-section onto the beam splitter would still vary with the angle of rotation of the scanning assembly.

In accordance with another aspect of the present invention, the combined translation and projection effects may be taken into account by adopting a circular configuration for the effective area of the beam splitter and an elliptical one for the cross-section of the input beam, the major and minor axes of the ellipse being so chosen that the effective area of the beam splitter almost totally swept by the ellipse is substantially a minimum that will meet the desired optical throughput and the angular travel limits of the scanning assembly required to achieve the predetermined O.P.D. maximum value. This is tantamount to saying that the whole of the beam splitter area is utilized almost to the full.

Apart from the economical advantage enabled by the refinement, in so far as beam splitters are expensive optical elements, there is also a mechanical benefit in that the smaller beam splitter is more mechanically stable.

Although it is proposed that the rotational axis of the scanning assembly be substantially coincident with the line of intersection referred to earlier, in fact a slight departure from coincidence can be tolerated. Naturally, if coincidence is aimed at in the first place, a more symmetrical region of tolerance around the ideal axis is achieved. Because of that tolerance, the mounting of the scanning assembly is simplified. In fact, the scanning assembly may comprise an integral rotatable casting mounting the folding mirror, the beam splitter and the output mirror. It may be housed within a stationary integral casting mounting the reversing mirror, two simple pivotal connections between the two castings allowing the first to rotate relative to the second, without any need for sophisticated and, therefore, costly bearings.

An interferometric apparatus in accordance with the present invention offers particular advantages when the apparatus is in fact an FT-spectrophotometer based on a Michelson-type interferometer, especially where a reference interferometer with a laser source is incorporated. In so far as the laser beam must be made to follow the same path through the scanning assembly as the spectrophotometric input beam, the reference interferometer will itself benefit from a spatially constant output beam which is of very small cross-section and must be critically aimed at co-operating photosensitive devices of small active areas, to obtain electrical signals therefrom through which the acquisition of data points from the interferogram of the sample under analysis may be accurately timed.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 depicts a Prior Art interference spectrometer of the Michelson type incorporating a rotary O.P.D. scanning assembly;

FIG. 2 represents the interference spectrometer of FIG. 1 modified to incorporate a rotary O.P.D. scanning assembly in accordance with the present invention;

FIG. 3 is a diagram intended to illustrate that the path of the output rays of the rotary O.P.D. scanning assembly in accordance with the present invention remains substantially fixed in space for any angle of rotation of the assembly within its design range;

FIG. 4 is a symbolic layout of an FT-IR spectrophotometer incorporating a rotary O.P.D. scanning assembly in accordance with the present invention;

FIG. 5 is a practical layout corresponding to the symbolic layout of FIG. 4.

FIG. 6 refers to the FT-IR spectrophotometer of FIGS. 4 and 5 and illustrates how from a pair of electrical sine waves, each representing the interferogram of the reference laser beam incorporated in said spectrophotometer, waves are derived for timing the acquisition of data points from the interferogram of a spectrophotometric sample under analysis;

FIG. 7 is a block diagram of a microprocessor-based system for managing the FT-IR spectrophotometer described with reference to FIGS. 4 to 6; and

FIGS. 8A to 8D depict alternative beam splitter constructions suitable for use in the present invention.

DETAILED DESCRIPTION OF THE INVENTION

It is thought that the concept of the present invention, as well as one particular manner in which it may be put into effect, can best be appreciated by first describing, with the aid of FIG. 1, the prior art rotary scanning assembly referred to earlier, as part of a Michelson-type interferometer. An added advantage of this approach is the opportunity it offers to relate the foregoing generalities on the said type of interferometer to a basic diagrammatic representation.

In FIG. 1, light emitted by a source **1** stopped down by an aperture plate **2** is collimated by a positive lens **3** into an input bundle of nominally parallel rays directed towards a beam splitter **4** oriented at 45 degrees to the incoming rays and comprising a semireflective (i.e. partially reflecting and partially transmitting) layer **4A** sandwiched between transparent supporting slabs **4B** and **4C** having identical optical characteristics. Upon striking the layer **4A**, the input bundle is split into a reflected bundle portion, with a 90-degree bend, and a transmitted bundle portion.

The reflected bundle portion encounters a plane folding mirror **5** and a plane reversing mirror **6** in its travel along a forward path extending from the layer **4A** to the mirror **6**, before reversing into a return path that extends from the mirror **6** to the layer **4A** and is spatially coincident with the forward path. At the end of its return path, it partially traverses the layer **4A**, is deflected by a plain folding mirror **7** into the entry pupil of a positive lens **8**, and is finally brought to a focus at a photodetector **9**. The sum of the forward and reverse optical paths extending between the layer **4A** and the reversing mirror **6** represents a first interferometric arm of the interferometer outlined in FIG. **1**. We shall refer to this arm as the upper arm.

The transmitted bundle portion passes through the layer **4A** at the start of its travel along a forward path, strikes the plane reversing mirror **10** at the end of that path, bounces back along a return path that terminates at the layer **4A** and is spatially coincident with the forward path, is reflected by the layer **4A**, and, again, by the plane mirror **7**, into the entry pupil of the positive lens **8**, and is finally brought to a focus at the photodetector **9**. The sum of the forward and return optical paths extending between the layer **4A** and the reversing mirror **10** represents a second interferometric arm. We shall refer to this arm as the lower arm.

Since the optical path extending between the layer **4A** and the photodetector **9** is shared by the two bundle portions reaching the photodetector **9** from the two arms of the interferometer, respectively, and neither bundle portion can therefore suffer retardation with respect to the other, as long as the optical path lengths of the two interferometric arms are absolutely identical, a wavefront reaching the layer **4A** along the return path of the first interferometric arm is coincident in space with a wavefront reaching the layer **4A** along the return path of the second interferometric arm, with the result that the two wavefronts are recombined at the layer **4A** (constructive interference). If the beam splitter **4** were perfect, and no transmission losses were suffered by the rays, the recombined bundle of rays would have exactly the same intensity as the input bundle issuing from the lens **3**. Under such ideal conditions, if one were to replace the detector **9** with a diffusing screen and the source **1** emitted visible light (say, monochromatic light, to take the simplest case), one would observe a patch of light on the screen that could not be distinguished in intensity from the patch that would be seen if the screen were placed at the aperture of aperture plate **2**.

On the other hand, if the optical path length of one interferometric arm was greater than that of the other arm, the optical path difference thus established would cause the wavefronts of the light from the arm having the greater optical path length to be retarded relatively to the wavefronts from the other arm. With the wavefronts out of phase, the light patch observed on the screen at the detector location would be less intense than before by an amount related to the out of phase angle, and, in the extreme case (phase opposition), would not be visible at all (destructive interference).

In the prior art arrangement being reviewed with reference to FIG. **1**, the optical length of the lower interferometric arm is fixed and that of the upper arm is variable, which is tantamount to saying that the O.P.D. between the two arms may be varied (or modulated). To this end the beam splitter **4** and the mirrors **5** and **7** are rigidly mounted in a frame **11** rotatable around a pivot **12**, normal to the plane of the paper, and affixed to a stationary base plate (not shown) to which the remainder of the components shown in FIG. **1** are firmly attached. The frame **11** and the parts represented within it constitute an O.P.D. scanning assembly, which will henceforth be referred to more briefly as scanner **13**. As shown in FIG. **1**, the scanner **13** is in the attitude that gives exactly zero optical path difference between the two interferometric arms.

The prior art interferometer of FIG. **1** is said to be aligned (i.e. the wavefronts from the two arms recombine exactly at the beam splitter layer **4A** when the O.P.D. is zero) if:

(a) the mirror **5** is mounted accurately parallel to the beam splitter **4**;

(b) the lower-arm reversing mirror **10** is set up to image the aperture in aperture plate **2** onto itself;

(c) the upper-arm reversing mirror **6** is set up so that its image via mirror **5** and beam splitter **4** is parallel to the mirror **10**; and

(d) the mirror **7** is parallel to the beam splitter **4** to within one degree or so.

It is to be assumed that the interferometer shown has been properly aligned in accordance with the above procedure.

To change the O.P.D. between the two interferometric arms, it is merely a question of rotating the scanner **13**, say, anti-clockwise around the pivot **12**. When that is done, the optical path length of the lower arm remains unchanged but that of the upper arm increases. No motion of the scanner **13** other than the rotation—in either sense of direction—around pivot **12** can affect the O.P.D. This is essentially due to the well known properties of a pivotally mounted pair of accurately parallel plane reflectors, such as in fact represented by the layer **4A** and the mirror **5**. Such pair will ensure that, although the optical path of the upper arm is increased by the anti-clockwise rotation of the scanner **13**, the bundle of rays reaching the mirror **6** are not tilted in the process but simply translate a little lower down the mirror **6**. This result is achieved whether the pivot **12** is located as shown or anywhere else. What happens to the optical path length of the common path extending between layer **4A** and lens **8** is not important, of course, since there cannot be any O.P.D. between two common optical paths.

Unfortunately, a serious disadvantage of the scanner **13** is that not only the horizontal bundle of rays in the upper arm translate over the face of the mirror **6** but also the parallel rays directed to the lens **8** translate over the face of mirror **7**. Thus, as the scanner **13** is turned anti-clockwise, at the same time that the bundle of rays directed to mirror **6** sweeps downwards, the recombined bundle from mirror **7** sweeps upwards, with the result that its effective cross-section is increasingly reduced since it no longer fills the aperture of the lens **8**, as in FIG. **1**, but increasingly underfills it from the bottom up. It follows that the solid angle of light energy reaching the detector **9** is correspondingly decreased or, more generically, the path of the impinging rays is changing with the position of the scanner **13**.

That the rays emerging from the scanner **13** move parallel to themselves as the said scanner is rotated can be readily appreciated by observing that while the direction of the collimated bundle of rays emerging from the lens **3** is immutable that of the corresponding rays reflected by the layer **4A** changes with the angle of rotation of the scanner **13**. For anti-clockwise rotation from the attitude depicted in FIG. **1**, the reflected rays impinge on the mirror **5** at a smaller angle of incidence (i.e. the angle between an incident ray and the normal to mirror **5**) than before and, therefore, lower down the mirror face. It follows that any ray reflected by mirror **5** will also meet mirror **6** lower down the face. Since any one ray reaching the mirror **7** is the optical prolongation of a ray that impinged on mirror **5** and was returned to mirror **5** via reversing mirror **6**, it follows that said ray, say, the central ray in FIG. **1**, will impinge on mirror **7** higher up than shown,. i.e. will translate upwards.

Where the instantaneous intensity of the light reaching the detector within an area imaging the aperture stop of the interferometric system must be measured, as is the case in commercial FT optical spectrometers, the path of the rays impinging on the detector must remain constant as the O.P.D. is scanned. The prior art O.P.D. scanner shown in FIG. **1** would not, therefore, be suitable for such application.

The arrangement of FIG. **1** has been modified to produce an embodiment in accordance with the present invention whereby the output beam reaches the detector along an optical path that remains fixed in space throughout the scanned O.P.D. range, thus opening up a vast field of application for a rotary O.P.D. scanner, the considerable virtues of which—first and foremost the comparative ease with which mechanical precision of scanning can be achieved—could not be exploited hitherto owing to the scanning action giving rise to a translation of the beam emerging from the scanner.

In FIG. **2**, depicting the modified arrangement, the mirror **7** has been turned around so that it lies with its reflecting face normal to the layer **4A**. It follows that lens **8** and detector **9** are transferred to respective positions symmetrical with those occupied in FIG. **1**.

The pivoting point for the O.P.D. scanner is no longer at **12** but has been shifted to **14**, the actual axis of rotation being at the line of intersection between the prolongation of layer **4A** and that of the reflecting surface of mirror **7**, the prolongations being indicated by the intersecting dotted lines shown.

The result of the changes thus introduced is quite dramatic. As observed earlier, the location of the axis of rotation in no way affects the operation of the mirror pair represented by layer **4A** and mirror **5** in ensuring that for any angle of rotation of the scanner **13** the bundle of rays directed towards and returning from mirror **6** remain strictly perpendicular to the reflecting face of mirror **6**. It follows that in explaining the operation of the present embodiment it will be sufficient to consider the effect of re-orientating the mirror **7** and selecting the rotary axis **14** on rays emerging from the scanner **13** after passing through layer **4A** from the upper interferometer arm, combining with rays from the lower arm and bouncing off the mirror **7**. It will naturally be assumed that the modified interferometer has been aligned in accordance with the procedure described earlier for the FIG. 1 interferometer. In FIG. 3, with some reference to FIG. 2, the effect is illustrated in diagrammatic form by following the path of a single input ray **15**, from the point of emergence from lens **3** (FIG. 2) up to the scanner output, for two arbitrarily chosen angular positions of the scanner **13**: an initial position for which the angle of incidence of ray **15** on layer **4A** is 45 degrees, and a final position for which the angle of incidence is reduced to 28 degrees. The Roman numeral I added in brackets to the references **4A** and **7** denote the initial position; and the Roman numeral II similarly added to the said references, the final position. The rays associated with the initial and final positions, respectively, are similarly distinguished. In the interest of simplicity, the split rays associated with the two interferometric arms have been truncated so that only the portions that co-operate with the beam splitting layer in each of the two positions of the scanner are shown in FIG. 3.

With the scanner **13** in the initial position as indicated, the input ray **15** meets the layer **4A(I)** and after being split in rays **15A(I)** and **15B**, in the manner generally described earlier with reference to FIG. 1, is recombined as **15A(I)+15B**, ray **15A(I)** being returned by reversing mirror **6** via mirror **5** from the upper interferometric arm and ray **15B** by the reversing mirror **10** from the lower arm.

Clearly ray **15** and its recombination **15A(I)+15B** must be normal to each other since the latter is on the prolongation of **15A(I)** which is normal to **15** because it originates from a 90-degree partial reflection of ray **15** by layer **4A**. Mirror **7(I)** is normal to layer **4A(I)** and it follows that ray **15A(I)+15B** must meet it at 45 degrees with the result that ray **15A(I)+15B** is reflected at 90 degrees and exits parallel to the input ray **15**.

It will now be shown that in the final position of the scanner **13** the ray that is to be referenced **15A(II)+15B** will not only exit parallel with the ray **15** (which is also achieved by the prior art interferometer shown in FIG. 1) but it will actually lie on a path coincident with the path on which ray **15A(I)+15B** exits when the scanner **13** is in the initial position (which is not achieved by the FIG. 1 interferometer). In other words, it will be shown that in moving the scanner **13** from one to other position the exit ray stays on the same path.

With the scanner **13** in the final position, the input ray **15** meets the layer **4A(II)** at an angle of 28 degrees, as stated earlier, and a split portion **15A(II)** is deflected at twice that angle into the upper interferometric arm. The returning **15A(II)** ray recombines with the **15B** ray to form the **15A(II)+15B** ray which intersects the mirror **7(II)** at a point where the path of the ouput ray for the scanner **13** in the initial position also intercepts the mirror **7(II)**. If the normal to the mirror **7(II)** at the point of intersection bisects the angle by which the recombined ray has been deflected, then the output ray for both initial and final positions of the scanner **13** must lie on a common path. That is in fact the case, as can be easily verified for any angle of rotation of the scanner **13** within its design range. This means that the present embodiment ensures that the exit path of the ray **15** (and of course any other ray one wishes to consider) remains fixed in space, thus completely eliminating in a very simple manner the very serious drawback of the prior art proposal.

An FT-IR spectrophotometer embodying an O.P.D. scanner in accordance with the present invention and suitable for operation in the region of the infrared spectrum extending between 4000 cm^{-1} and 400 cm^{-1}, equivalent to wavelengths of 2.5 and 25 microns, respectively, is diagrammatically represented in FIG. 4.

In FIG. 4, a coiled platinum-wire IR source **1**, longitudinally extending along an axis normal to the plane of the paper, is surrounded by a cylindrical shield **1A** having a window **1B** facing the median hottest zone of the source. The shield **1A** is represented to suggest that a cross-section has been taken on a plane intersecting the said axis and bisecting the window **1B**. The whole is surrounded by a further shield **1C** with a circular window **1D** much larger than widow **1B** and similarly represented. The patch of the infrared radiation emitted by the source **1** through window **1B** that is actually utilized in the spectrophotometer is not determined by either window **1B** or window **1D** but by the post-source optics, as will be presently shown. A principal ray and two pairs of marginal rays, one pair from each lateral end of the utilizated patch are traced through the entire system.

The infrared rays as defined impinge upon an elliptical mirror **16**, one focal region of which coincides with the radiation patch at the source **1**, and the other with circular aperture **2A** in plate **2**, of such diameter that the said aperture is overfilled all round by the image of said patch. A baffle plate **17** located at a pupil position positively defines the effective aperture of the mirror **16** by virtue of an aperture **17A** shaped in the form of an ellipse having major and minor axes predetermined for a purpose to be presently described, and a plane mirror **18** folds the rays emerging from the aperture **17A** onto the aperture **2A**. The arrangement so far described serves to image the radiation patch from the source **1** at the aperture **2A**, with the result that the radiation-filled circular aperture **2A** becomes the effective source with a well defined imagery and being overfilled effectively determines the area of the radiation patch actually utilized. In sum, it represents the aperture stop (or Jacquinot stop) of the system.

A second plane mirror **19** folds the beam emerging from aperture **2A** onto parabolic mirror **20**, the length of the optical path between the latter two elements being approximately equal to the focal distance of the mirror **20**, which acts therefore as a collimator and may be regarded as the optical counterpart of the lens **3** in FIGS. 1 and 2. Since the source **1** is not a spot source but is intended to be a finite area source, collimation is approximate but adequate nevertheless.

In addition, mirror **20** images the elliptical aperture **17A** onto the reversing mirrors **6** and **10**. The major and minor axes of the elliptical patch of radiation projected have been chosen to ensure efficient utilization of the circular active area of the beam splitter for a given optical throughput and O.P.D. scan range, bearing in mind that in the present embodiment the collimated beam actually translates over the beam splitter surface, as reference to FIG. 3 will show. Further elaboration of this feature, which is believed to be novel per se, will be given after the description of FIG. 5.

The collimated beam produced by the parabolic mirror **20** forms the optical input to the interferometric optics proper, wherein the semi-reflecting layer **4A** of beam splitter **4** forming part of the O.P.D. scanner **13**, within the boundary of frame **11** shown in dashed lines, co-operates through its upper folding mirror **5** with the fixed reversing mirror **6** to define the upper interferometric arm and, through the lower folding mirror **21**, with the reversing mirror **10** to define the lower interferometric arm. Note that the mirror **21** enables optical symmetry to be established between the two interferometric arms.

The beam recombined at the beam splitter **4** in the manner described earlier exits from the scanner **13** via the outut mirror **7**, which performs the all-important function already explained with reference to FIGS. **2** and **3**. A parabolic mirror **22** converges the output beam from the scanner **13** onto its focal region, whereat an image of the Jacquinot stop is projected and a thinly layered analytical sample borne upon a carrier **23** is to be located. An elliptical mirror **24** is so designed and positioned that said image spatially coincides with its first focal region and is relayed with a de-magnification of approximately 6:1 to its second focal region, which is occupied by the sensitive surface of detector **9**.

Note that the analytical sample and its carrier may assume any of the convenient forms well-known in the art. A thinly-layered sample has been indicated because it is a commonly used form and may be readily related to the nominal plane of the Jacquinot stop image.

All the components shown in FIG. **4**, with the exception of those within frame **11** forming part of scanner **13**, are rigidly mounted on a base plate, not shown in FIG. **4** but represented in the constructional drawing of FIG. **5**. Scanner **13** is rotatable to and fro between two limit positions around axis **14** within a housing itself fixed to the base plate, as will be described with reference to FIG. **5**.

As shown in FIG. **4**, the scanner **13** is midway between its limit positions and provides zero O.P.D. between the two interferometric arms. The two limits, attained by clockwise and anti-clockwise rotation, respectively, correspond to O.P.D. values that are equal but of opposite sign. In fact, clockwise rotation of the scanner from the position shown increases the optical path length of the variable interferometric arm and anti-clockwise rotation decreases it. It is therefore convenient to refer to the clockwise O.P.D. limit as L_{max+} and the anti-clockwise as L_{max-}. By swinging the scanner **13** from one to other limit, a complete double-sided interferogram is generated, as required for the purpose of computing its Fourier Transform. In the present embodiment, L_{max} is 2 cm and the scan time between L_{max} and zero O.P.D. is 1 second.

For a purpose that presently will be referred to in some detail, a reference monochromatic interferometer is inserted in the optics of the main interferometer, forming part of the FT-IR spectrophotometer described, so that they share the scanner **13**. The reference interferometer comprises, in addition to the shared scanner **13**, a low-power laser **25** providing a line of monochromatic plane-polarized radiation at a wavelength of 632.814 nanometers. The plane-polarized beam issuing from the laser **25** is passed through the scanner **13** via a small plane mirror **26** located in the collimated beam produced by parabolic mirror **20** and is recovered from the collimated beam at the output of scanner **13** for directing to sensor unit **27** via plane mirror **28** co-operating with mirror **29**. In passing through the scanner **13**, the beam traverses a ⅛-wave plate **30** and is affectd thereby in a manner that will be presently indicated. At the sensor unit **27**, a beam splitter **27A** permits a transmitted part of the laser beam to reach a photosensor **27B** without a change in direction and a reflected part to reach a photosensor **27C** with a 90-degree deflection, the axis of the two sensors being spaced 90 degrees apart. For a purpose to be considered later, plane polarizers **27B1** and **27C1** are positioned before photosensors **27B** and **27C**, respectively, with their transmission planes crossed.

In FIG. **5** the diagrammatic representation of FIG. **4** has been translated into an engineering layout of real components, with only the chief ray CF shown to trace the optical path from the source **1** to the detector **9**. It is believed that FIGS. **4** and **5** together facilitate a proper appreciation of function and construction in the realization of an O.P.D. scanner per se and as part of an FT-IR spectrophotometer, both in accordance with the present invention. Incidentally, although the mechanical precision requirements are made less stringent by confining operation to the infrared region of the electromagnetic spectrum, the rotary O.P.D. scanner in accordance with the present invention enables high precision to be advantageously attained in regions of the spectrum of shorter wavelengths.

In the description that follows, references identifying elements shown diagrammatically in FIG. **4** will also identify the corresponding real components in FIG. **5**. Of these, the mirrors **16, 18, 19, 20, 22** and **24** need no further elaboration except for pointing out that they are supported on brackets **16A, 18A, 19A, 20A, 22A** and **24A**, respectively, for fixing by means of screws **16B, 18B, 19B, 20B, 22B** and **24B** to a rigid base plate **31**, as shown. It will be noted, in addition, that the detector **9** is mounted on an L-shaped bracket **9A** which is itself attached to an extension of bracket **24A** mounting mirror **24**, and that the sample carrier plate **23** is slid, in use, into channels **23A** formed in a guide plate **23B** fixed to a bracket **23C** for mounting on the base plate **31** by means of screws **23D**, the sample having been made fast to the carrier **23** over the circular aperture **23E** by any suitable means well known in the spectroscopic art. Furthermore, the baffle **17** and the Jacquinot stop **2** form a combined unit P made of sheet metal in which the aperture of the former is oval and that of the latter circular. A bottom flange P1 enables the unit P to be screwed to the base plate **31** by screws P1A.

The source **1**, complete with its outer shield **1C**, is mounted vertically between the limbs of a U-shaped bracket **1E** that is itself attached to the upright limb of an L-shaped bracket **1F** in turn affixed, through the other limb of the L, to the base plate **31** by screws such as **1G**.

Attention is now particularly directed to the bottom left-hand corner of FIG. **5**, whereat two castings are shown, one comprising three integral sections **11A, 11B, 11C** forming one part of the O.P.D. scanner **13** and the other, referenced **32**, providing essentially a housing within which the scanner **13** is rotatably mounted in a manner to be described.

Casting **11A-11B-11C**, which may be regarded as an engineered version of the frame **11** diagrammatically shown in FIG. **2**, comprises a box-like centre section **11A** extending, to the right, into a flag **11B** and, to the left, into a drive arm **11C**. The centre section **11A** has a wall **11A1** provided with an aperture **11A2** into which a beam splitter **4** of circular configuration is fixedly mounted and a facing wall **11A3** mounting folding mirror **5**, which is also circular.

As shown in FIG. **5**, the beam splitter **4** is of the same construction as the beam splitter **4** in FIG. **4**, in that it comprises a semi-reflective layer sandwiched between two supporting slabs transparent to the radiation emanating from the source **1** and having identical optical characteristics. However, the semi-reflective layer **4A** (FIG. **4**) may alternatively be provided on a taut film transparent to said radiation. It may also take the form of a specular foraminous structure such as a mesh or a grid, in which case the layer represented by the structure is semi-reflective by virtue of the lack of reflection from the open spaces in said structure. The structure may be self-supporting or carried on a transparent taut film. In the latter case, it could be provided as a layer vacuum evaporated onto the film.

Alternative beam splitter constructions as referred to are depicted in FIGS. **8A** to **8D**. In each figure, a ring frame **4D** forms a carrier for a semi-reflective layer. In FIG. **8A**, the frame **4D** supportes closely spaced, stretched, parallel wires forming a grid **4E**; in FIG. **8B**, the two sets of parallel wires intersect to define a mesh **4F**. FIGS. **8C** and **8D**, which are cross-sections of perspective views along a diameter of the frame **4D**, depict a grid **4G** and a mesh **4H**, respectively, as vacuum deposited layers on a transparent film **4K** stretched over the frame **4D**.

The flag **11B** acts as a carrier for rectangular mirror **7**, which is rigidly supported normal to the beam splitter **4**. The region where the flag **11B** extends from the centre section **11A** includes the line of intersection between the surface prolongations of mirror **7** and reflecting layer **4A** of beam splitter **4**. An upper boss **11A4** is formed in an upper wall **11A5** of section **11A**, roughly concentric with the said line of intersection. Within the boss **11A4** a V-shaped bearing **11A4A** is provided. Similarly, a lower boss **11A6** is provided in a lower wall **11A7** of section **11A** which is roughly concentric with the line of intersection. Within the boss **11A6** a ball bearing **11A8** is fixedly mounted.

The drive arm **11C** has an end boss **11C1** acting as a bearing to a pin **11C2** connecting one end of a plate **11C3** to one end of a plate **11C4**, the two plates forming a linkage **11C5** which further comprises another pin **11C6** joining the remaining two ends of the plates and rotatable in a block **11C7**. The housing **32** comprises a floor **32A**, having holed bosses such as **32B** through which screws such as **32C** are passed for fixing the housing **32** to the base plate **31** through resilient mountings such as **32C1**, and a roof **32D** spacedly supported by intervening vertical members such as integral pillar **32E**. A cylindrical pivot **32F** that is a forced fit in boss **32G** is provided for engaging the V-shaped bearing **11A4A** in boss **11A4**. A recess (not shown) in the floor **32A** houses a spherical bearing member (not shown) for engaging the ball bearing **11A8** in boss **11A6**. The axis of rotational symmetry of the pivot **32F** and that of said spherical bearing member coincide with the line of intersection referred to earlier and the ball bearing **11A8** is mounted so that a diameter thereof also coincides with said line. It follows that the arrangement described pivotally mounts the scanner **13** in the housing **32** for rotation around an axis which coincides with said line of intersection, i.e. axis **14**. Moreover, the location of the axis creates a weight offset towards the front in the scanner **13** as seen in FIG. **5**, with the result that the V-shaped bearing **11A4A** is urged by gravity into contact with the pivot **32F**.

One of the vertical spacing members between the roof **32D** and the floor **32A** is mirror carrying structure **32G** provided with vertical walls **32G1** and **32G2**, supporting mirrors **6** and **10**, respectively, the said walls being oriented so as to locate the mirrors in the correct angular positions diagrammatically illustrated in FIG. **4**. A further member **32H** in the form of a pillar of rectangular cross-section is used to mount the mirror **21**.

The drive means for rotating the scanner **13** to and fro within two predetermined angular limits relative to the housing **32** will now be described. One end of a drive shaft **33** solidal with a cylindrical drive roller **34** is free to rotate in block **11C7** but is longitudinally captive thereto. The shaft **33** is further pivotally supported at its other end in a plain sleeve bearing **35** mounted in one limb **36A** of a U-shaped bracket **36** fixed to the floor **32A** by screws such as **36B**, said sleeve bearing **35** additionally allowing freedom of axial displacement. The drive shaft **33** includes an integrally formed lead screw portion **33A** engaging a mating thread in the other arm **36C** of bracket **36**. The drive shaft **33** may be rotated in either direction by an electric motor **37** via a belt **38** forming a drive linkage between a pulley **39** mounted on motor shaft **40** and the cylinder **34**. The motor **37** is mounted on a bracket **41** affixed to the U-bracket **36** by screws such as **42**. Assuming that the lead screw **33A** is righthanded, clock-wise rotation of the motor shaft **40** causes the drive shaft **33** to move forward, thus urging the scanner **13** to turn in a clock-wise direction, the gradually changing angle between the drive shaft **33** and the drive arm **11C** in a plane parallel to the floor **32A** and passing through the axis of shaft **33** being taken up without restraint by the action of the linkage **11C5**. At the same time the belt **38** will not impose more than a negligible restraint on the longitudinal motion of the shaft **33** by virtue of the fact that the side thrust in its engagement with the cylinder **34** introduced by said motion will cause it to slip axially of the cylinder **34**. The operation for anti-clockwise rotation of the shaft **33** is, of course, similar although the direction of the side thrust is reversed.

An overhanging portion **32L** of the roof **32D** supports the laser **25** through saddles, of which only saddle **25A** is visible, co-operating with straps **25B**. The mirror **26** is mounted on a stem **26A** terminating in a circular foot **26B** attached to the floor **32A** by screws (not shown). The $\frac{1}{8}$-wave plate **30** is contained within a cylindrical housing **30A** supported by a stem **30B** terminating in a rectangular foot **30C** screwed by screws **30D** to the floor **32A**. Mirror **28** is mounted in the same manner as mirror **26**, stem **28A**, foot **28B** and fixing screws **28C** being clearly visible. Mirror **29** is provided with a rear stem fitting in a hole drilled in the pillar **32H**, the stem and the hole being hidden.

A cylindrical body **27F** in unit **27** is provided with recesses in which suitably oriented machined faces **27A1**, **27B1**, **27C1**, **27D1** and **27E1** adhesively support the already introduced elements **27A**, **27B**, **27C**, **27D** and **27E**, respectively.

The electrical elements shown in FIG. **5** are inserted in an electrical operating system co-operating with the mechanical parts hereinbefore described. The laser **25** is energized through a modular power supply **25C**, the latter being connected to a public-supply outlet via a power and instrument control pack **43** having a lead **43A** terminating in a plug **43B** for connection to said outlet. Leads from the detector **9**, photosensors **27B** and **27C** and drive motor **37** also extend into pack **43**, which includes, in addition to power supply circuitry, the whole of the hardware represented in the block diagram of FIG. **7**, with the exception, of course, of a Video Display Unit and Keyboard that will be presently introduced and the blocks representing the detector **9**, the motor **37** and the unit **27**, of course.

A procedure is now described for defining the major and minor axes of the elliptical patch of radiation introduced earlier in this disclosure. In such procedure, the area of the patch is an immutable datum, fixed by the desired radiation throughput of the instrument.

A satisfactory approach is to estimate, in the first instance, a likely value for the minor axis of the ellipse formed on the beam splitter **4** (FIGS. **4** and **5**) when the scanner **13** is in the angular position corresponding to zero O.P.D. and to carry out a trace of the principal rays passing through the respective extremities of the estimated minor axis as the scanner is "swung" first to one scan limit (L_{max+}) and then to the other (L_{max-}), in order to find the respective intersects with the beam splitter **4** at said limits. The trace will, of course, account for the different angles of incidence in the presence of the compounding effect of the translation of the ellipse over the face of the beam splitter **4** caused by the rotation of the scanner **13** around an axis that is offset from the major axis of the ellipse.

Now, the area of the beam splitter **4** is minimized if the intercepts referred to and the extremities of the major axis lie on the same circle. This is tantamount to saying that the distance between the intercepts must set the length of the major axis. If the simple equation:

$$\pi \cdot a \cdot b = A$$

wherein

a = major semiaxis

b = minor semiaxis

A = area of ellipse for required throughput

is met, the predicted value for the minor axis was correct. If not, the ray trace is repeated using a lower or higher value (whichever is suggested by the previous result) until an ellipse is found that meets the throughput requirement and allows the scanning range to be covered in a circular beam splitter of minimum area. A few iterations may be needed to achieve this.

Referring to FIG. **4** once again, it should be noted that the reason for imaging the elliptical aperture **17A** onto the reversing mirrors **6** and **10**, rather than the beam splitter **4** itself, is the excessive broadening of the beam returning to the beam splitter from the two reversing mirrors that would otherwise result. The ellipse the geometry of which is determined in the manner hereinbefore described is that produced by the returning beams, and although these must still diverge, by controlling the distance between each reversing mirror and the beam splitter, the broadening can be kept within limits that ensure an efficient utilization of the beam splitter area.

Once the ellipse has been defined, it is merely a matter of arranging the length of the optical paths between (a) the mirror **20** and the baffle plate **17** and (b) the mirror **20**, the beam splitter **4** and reversing mirrors **6** and **10**, respectively. In general, once the matching optical paths between the beam splitter **4** and the reversing mirrors **6** and **10**, respectively, have been settled, one need only select the optical layout that will give conveniently scaled (e.g. demagnified) images of aperture **17A** so that the patch of radiation produced on the beam splitter by the beams returning thereto from said images fits the ellipse calculated in accordance with the method described above.

Before introducing the block diagram of FIG. **7** illustrating the microprocessor-based system for managing the FT-IR spectrophotometer of FIGS. **4** and **5**, it is thought desirable to dwell on certain generalities respecting the operation of the scanner **13**, the reference interferometer based on the laser **25** and the acquisition of interferogram data points, in both downscanning and upscanning, for the purposes of digital processing. In the present context, the term downscanning (or downscan) denotes that the O.P.D. of the variable interferometric arm is being changed from L_{max+} to L_{max-} and upscanning (or upscan) that it is being changed from L_{max-} to L_{max+}.

As indicated earlier, the scanner **13** reciprocates between two maximum retardation limits corresponding to angular positions which are symmetrical with respect to the angular datum position depicted in FIGS. **4** and **5**, at which the O.P.D. is nominally zero.

The double-sided interferogram of the source **1**, as modified by the sample in carrier **23**, produced at each limit-to-limit scan is transduced into a time varying electrical signal waveform at the output of the detector **9**. This analogue signal is, of course, unsuitable for digital processing without first undergoing analogue to digital conversion, which involves sampling the waveform at accurately repeatable O.P.D. scanning increments.

In the present embodiment, sampling is synchronized with the zero crossings of an electrical reference sine wave representing the interferogram of the laser **25** and each sampling point is constantly tracked through a counter so that successive values sampled at that point through repetitive scanning are always associated with the same counter reading and can therefore be accumulated at a digital storage location associated with said reading. In view of the scanning sequence adopted, a refinement is introduced to permit the counter to count up during upscanning and down during downscanning. The refinement consists in duplicating the reference sine wave with a 90 degree phase shift. This is achieved by the simple expedient of introducing the ⅛-wave plate **30** referred to earlier, which being traversed twice by the polarized beam from laser **25** acts on it as if it were a ¼-wave plate traversed once and, therefore, produces circular polarization of the laser beam. The original plane polarized sine wave is now accompanied by a sine wave of the same frequency 90 degrees apart. The polarizers **27D** and **27E** in unit **27** are crossed in such manner that the photosensors **27B** and **27C**, respectively associated with them, can each pass only the sine wave encountering a plane of transmission at the correct angle for access. The end result is that one optical sine wave is transduced in a related electrical sine wave by photosensor **27B** and the other by photosensor **27C**.

How the two electrical sine waves are used for the purpose of the present embodiment is now explained with reference to FIG. **6**, wherein A and B represent the said sine waves, 90 degrees apart, **A1**, **B1** the waves resulting from squaring them and **A2**, **B2** those resulting from differentiating the square waves **A1**, **B1**. Now, each pulse occurring in **A2** is used to control the sampling and processing of one data point and at the same time update a counter to be identified later with reference to the block diagram of FIG. **7**.

Assuming that the phase relationship of A2 and B2 corresponds to a downscan as indicated by arrow C, then the occurrence of a positive pulse in B2 immediately following a positive pulse in A2 is sensed and used to decrement the counter. Conversely, a positive going pulse in A2 is not immediately followed by a like pulse in B2 is associated with an upscan (arrow C1) and the counter incremented.

Upon switching on the FT-IR spectrometer, the counter is zeroed at the first zero crossing in the A2 wave immediately following the peak of the centreburst in the interferogram of the analytical sample. Thereafter, and until the instrument is switched off, the counter keeps track of the sampling points in upscanning and downscanning.

The block diagram of FIG. 7 illustrating a microprocessor-based system for the control of the FT-IR spectrophotometer depicted in FIGS. 4 and 5 will now be described.

In FIG. 7, the signal from detector **9**, amplified by amplifier **43**, is fed into sample and hold unit **44** and each sampled value is routed to analogue-to-digital converter **45** feeding into microprocessor **46**. The output of unit **27**, consisting of two sine waves 90 degrees apart, as described earlier, is passed through an amplifier **47**, after which each sine wave is squared in unit **48**. Each resulting square wave is differentiated in differentiator **49**. One output of differentiator **49**, corresponding to differentiated wave A2 (FIG. **6**), is used in trigger pulse generator **50** to activate at each peak, regardless of polarity, the sample and hold unit **44** under the control of microprocessor **46** and the same output is extended to up and down counter **51**.

When the FT-IR spectrophotometer is first switched on, the counter **51** must be zeroed correspondingly to a predeterminedly fixed angular position of scanner **13** (FIGS. **4** and **5**) in order that it may thereafter keep track of all sampling points until the instrument is switched off. This achieved through a counter phase up unit **52**, which at a time when the microprocessor **46** determines that the maximum level in the amplitude of the centreburst of the sample interferogram has been reached sets the counter **51** to count from the next peak of the differentiated wave A2 in unit **49** that also supplies an input to the trigger pulse generator **50**, as stated earlier.

The scan motor **37** energized by drive circuits **53** responsive to scan motor control **54** is continuously steered in operation by the microprocessor **46** through the said scan motor control **54**, the actual count in counter **51** determining the scan action to be taken. Changes in scan direction are sensed by upscan/downscan sensor **55**, to which counter **51** responds by making the appropriate change in counting direction.

It can now be readily appreciated that each of the retardation limits must correspond to a definite and immutable counter reading. When the microprocessor **46** detects that reading, it sends a reverse signal to the scan motor control **54** and, at the same time, interrupts the processing of data points. The scan motor **37** begins its reversal procedure and the counter keeps on counting. When the count is restored to that obtaining when data processing was interrupted, the processing of data is resumed until the sampling point corresponding to the opposite retardation limit is reached, when the reversal operation is again carried out. The scan overruns by a few dummy sampling points past a retardation limit and, when the motor actually stops and reverses, the counter decrements (if before reversal upscanning was operative) or increments (if downscanning was operative) by the same number of counts as the number of sampling points overrun.

It is important to observe that by keeping track of sampling points through up and down counter **51** data points may be accumulated in upscan and downscan and correctly routed to their respective memory locations for signal averaging of data recurring at the same sampling points in successive interferograms. In addition, the scan motor reversal procedure is greatly simplified.

A video display unit **56** linked to the microprocessor **46** is included to enable the user to examine a display of sample interferograms and spectra, as well as other data. The Fourier Transform mathematical process is carried out by the microprocessor **46** which is programmed to respond to instructions keyed in by the user on the keyboard **57**.

It is to be understood that in the context of the present disclosure the term "light" includes the "invisible" regions of the light spectrum.

By incorporating a rotary O.P.D. scanner in accordance with the present invention in an FT-IR spectrophotometer as described, it has been possible to realize an instrument that is accurate, simple, comparatively inexpensive and highly reliable.

I claim:

1. Interferometric apparatus comprising:

(a) a rotary O.P.D. scanning assembly mounting a mirror pair formed by a beam splitter and a folding mirror, said beam splitter having a semi-reflective beam-splitting layer one face of which is the optical input surface in accurately parallel face-to-face relationship with said folding mirror and the other the output surface;

(b) two stationary reversing mirrors co-operating with said mirror pair for defining two interferometric arms, each arm receiving one of the two beams resulting from the splitting of an input beam by the beam splitter, for rotationally scanning successive O.P.D. values between the two beams up to a predetermined maximum, and for recombining the beams at said output surface so that they emerge therefrom as a single beam directed along an output path;

(c) an output mirror mounted at right angles to the output surface of said beam splitting layer, said output mirror being borne by the scanning assembly and positioned to receive the output beam; and

(d) pivotal means arranged to ensure that the scanning assembly rotates around an axis substantially coincident with the line of intersection of the prolongation planes of the semi-reflective beam-splitting layer and the face of the output mirror, respectively.

2. Apparatus as claimed in claim **1**, wherein the beam splitter has an effective area of generally circular configuration and said apparatus further comprises optical beam control means for projecting onto said input surface an input beam the cross-sectional area of which is an ellipse, the minor axis of the ellipse being chosen so that the distance separating the intercepts of the minor axis extremities with the input surface at two opposite predetermined angular travel limits of the scanning assembly is close to the length of the major axis and at the same time the said cross-sectional area is adequate to provide the desired optical throughput of the apparatus.

3. Apparatus as claimed in claim **2**, wherein said apparatus includes a source of light for producing said input beam and the optical beam control means comprise image forming means and a baffle which are located between the source and the beam splitter, the baffle having an elliptical aperture the major and minor axes of which together with the optical arrangement of the image forming means are such that the major and minor axes of the largest projection of said aperture on the semi-transparent layer are approximately equal to the desired major and minor axes, respectively, of the ellipse.

4. Apparatus as claimed in claim **1**, wherein the O.P.D. scanning assembly is a pivoted integral casting supporting in predetermined spaced relation the folding mirror, the beam splitter and the output mirror.

5. Apparatus as claimed in claim **4**, including a housing in the form of another integral casting wherein the O.P.D. scanning assembly is pivotally mounted, the two stationary reversing mirrors being fixed to said housing.

6. Apparatus as claimed in claim **5**, wherein in the operative attitude of the apparatus the axis of rotation of the O.P.D. scanning assembly is generally vertically oriented, said axis spanning an upper and a lower pivotal connection between the two castings, the upper connection comprising a cylindrical member fixed to one casting radially biassed into a V-shaped bearing provided in a member fixed to the other casting.

7. Apparatus as claimed in claim **6**, wherein the lower pivotal connection is of the type wherein by virtue of its own weight the pivoted casting creates a centralizing force component restraining all radial movement thereof as mechanical wear takes place.

8. Apparatus as claimed in claim **1**, including a base plate and a housing for said O.P.D. scanning assembly resiliently mounted on said base plate.

9. Apparatus as claimed in claim **8**, wherein a photosensitive device is attached to the base plate for receiving the output beam from the output mirror.

10. Apparatus as claimed in claim **1**, wherein the semi-reflective layer is held between two rigid supports transparent to the input beam and having identical optical characteristics.

11. Apparatus as claimed in claim **1**, wherein the semi-reflective layer is supported on a thin taut film transparent to the input beam.

12. Apparatus as claimed in claim **1**, wherein the semi-reflective layer is a specular foraminous structure, which is semi-reflective by virtue of the lack of reflection from the open spaces in said structure.

13. Apparatus as claimed in claim **12**, wherein said structure is self-supporting.

14. Apparatus as claimed in claim **12**, wherein said structure is supported on a transparent taut film.

15. Apparatus as claimed in claim **14**, wherein said structure is in the form of a layer vacuum evaporated on said film.

16. Apparatus as claimed in claim **1**, wherein said apparatus is an FT spectrophotometer including a Michelson-type interferometer and a detector for receiving the output beam and transducing the optical interferograms conveyed by said beam into an electrical representation.

17. Apparatus as claimed in claim **16**, comprising; a reference interferometer, incorporated within the Michelson-type interferometer, having a laser source for providing a narrow, plane-polarized, monochromatic beam; means for directing the laser beam through the O.P.D. scanning assembly along a path shared with the input beam, split beams and output beam up to and including a portion of the output path emerging from the output mirror; means for diverting the laser beam past the output mirror into a path distinct from the remainder of the output path leading to the detector; means for circularly polarizing the laser beam and thus producing two optical interferograms of the laser beam 90 degrees apart; a two-channel photoelectric transducer for transducing the said optical interferograms, one in each channel, into corresponding electrical sine waves 90 electrical degrees apart keyed to the angular travel of the scanning assembly; means for utilizing the sine waves for controlling the acquisition timing of data points from the detector output representing an analytical interferogram; and means for performing the Fourier Transform of the analytical interferogram and thus producing an electrical representation of the spectrum of a sample under analysis.

* * * * *

United States Patent [19]

Logan et al.

[11] **Patent Number: 5,043,739**

[45] **Date of Patent: Aug. 27, 1991**

---

[54] **HIGH FREQUENCY RECTENNA**

[75] Inventors: **B. Grant Logan**, Danville; **William J. Orvis**, Livermore, both of Calif.

[73] Assignee: **The United States of America as represented by the United States Department of Energy**, Washington, D.C.

[21] Appl. No.: **472,548**

[22] Filed: **Jan. 30, 1990**

[51] **Int. Cl.**^{5} **H01Q 1/26**
[52] **U.S. Cl.** **343/701**; 343/893
[58] **Field of Search** .... 343/701, 753, 700 MS, 343/846, 893

[56] **References Cited**

U.S. PATENT DOCUMENTS

2,855,543	4/1953	White	315/34
3,114,517	12/1963	Brown	244/1
3,434,678	5/1965	Brown	244/1
3,535,543	10/1970	Dailey	307/149
3,544,998	12/1970	Vandenplas et al.	343/701
3,711,848	1/1973	Martens	343/701
3,789,471	2/1974	Spindt et al.	357/6
3,852,755	12/1974	Works et al.	343/701
4,008,477	2/1977	Babij et al.	343/701
4,079,268	3/1978	Fletcher et al.	343/700 MS
4,155,155	5/1979	Bourdon et al.	29/590
4,516,037	5/1985	Shackle	307/252 A
4,587,545	5/1986	Berthold et al.	357/38
4,721,885	1/1988	Brodie	313/576
4,853,703	8/1989	Murakami et al.	343/701
4,853,705	8/1989	Landt	343/701
4,855,636	8/1989	Busta et al.	313/306
4,955,562	9/1990	Martin et al.	244/62

OTHER PUBLICATIONS

"The History of Power Transmission by Radio Waves", by William C. Brown, IEEE Transactions on Microwave Theory and Techniques, vol. MIT-32, No. 9, 9/1984.

"Rectenna Technology Program: Ultra Light 2.45 GHz Rectenna and 20 GHz Rectenna", by William C. Brown, Prepared for National Aeronautics and Space Administration, NASA Lewis Research Center, Contract NAS3-22764.

"Initiative for the 21st Century: Advanced Space Power and Propulsion Based on Lasers", by B. G. Logan, Lawrence Livermore National Laboratory Report No. UCRL-98520.

Primary Examiner—Michael C. Wimer
Assistant Examiner—Hoanganh Le
Attorney, Agent, or Firm—H. P. Sartorio; L. E. Carnahan; William R. Moser

[57] **ABSTRACT**

The invention provides an inexpensive array of rectifying antennas which employ field emission diodes for rectifying electromagnetic waves of microwave frequencies and higher frequencies.

15 Claims, 5 Drawing Sheets

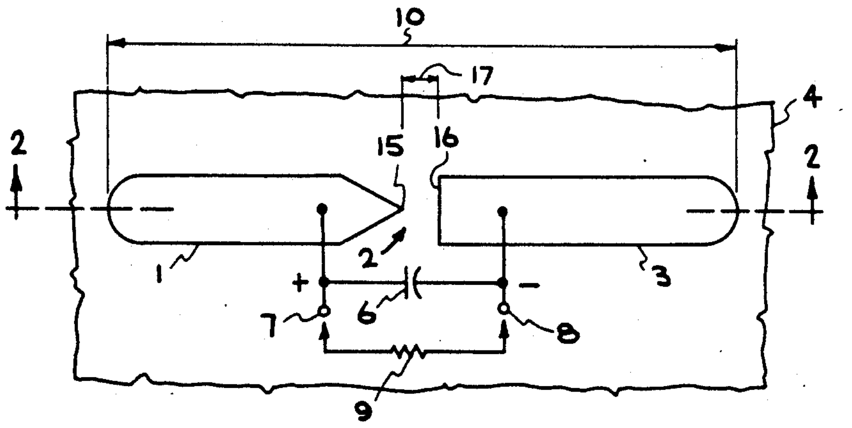

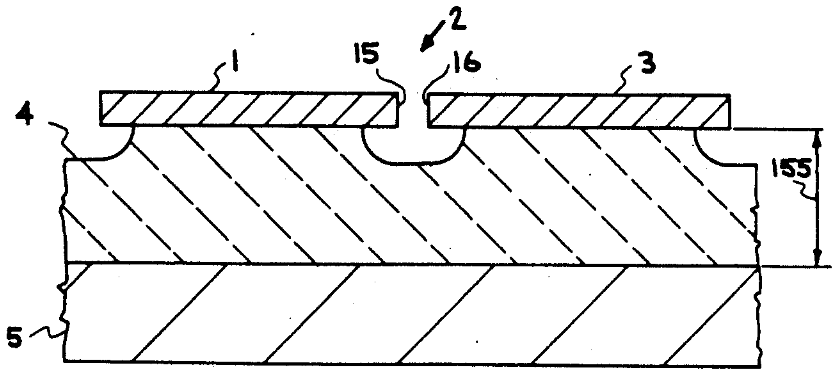

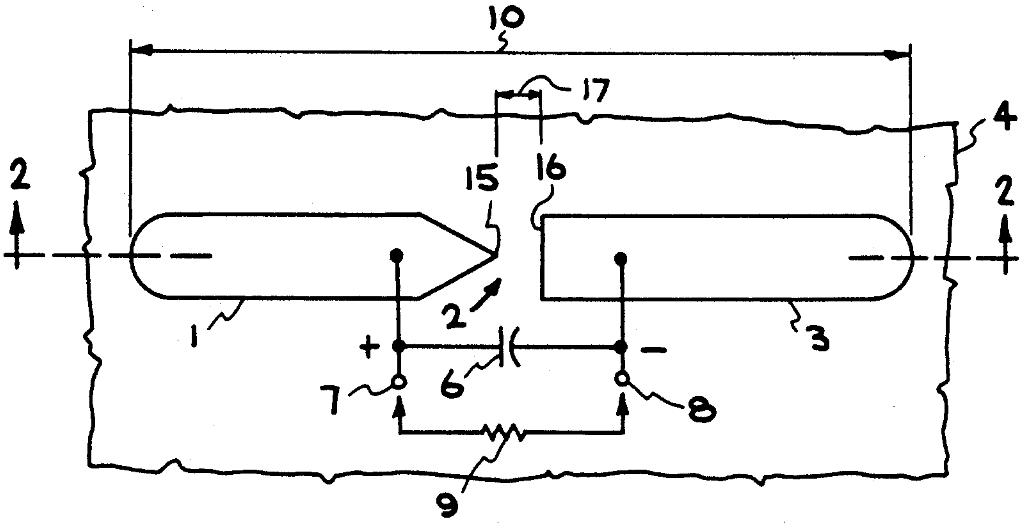
10
17
4
2
15
16
2
2
1
3
+
−
6
7
8
9

FIG. 1

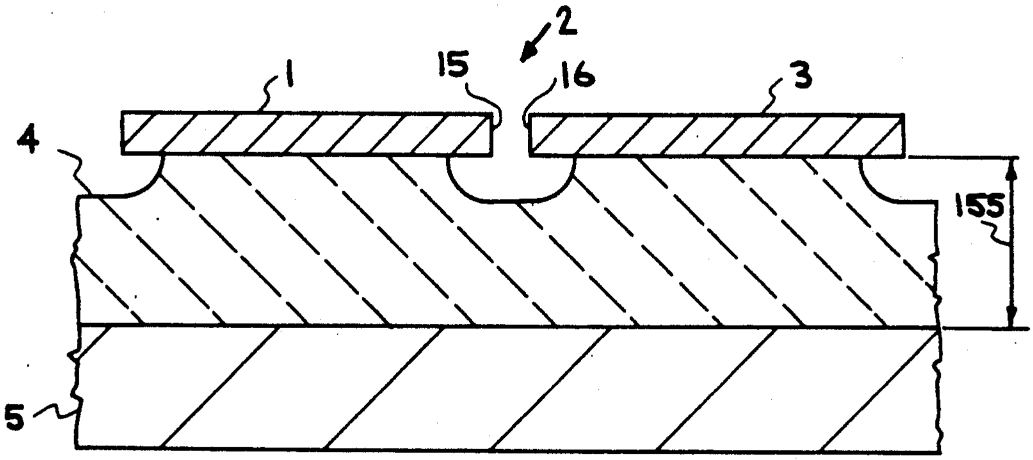
2
15
16
1
3
4
155
5

FIG. 2

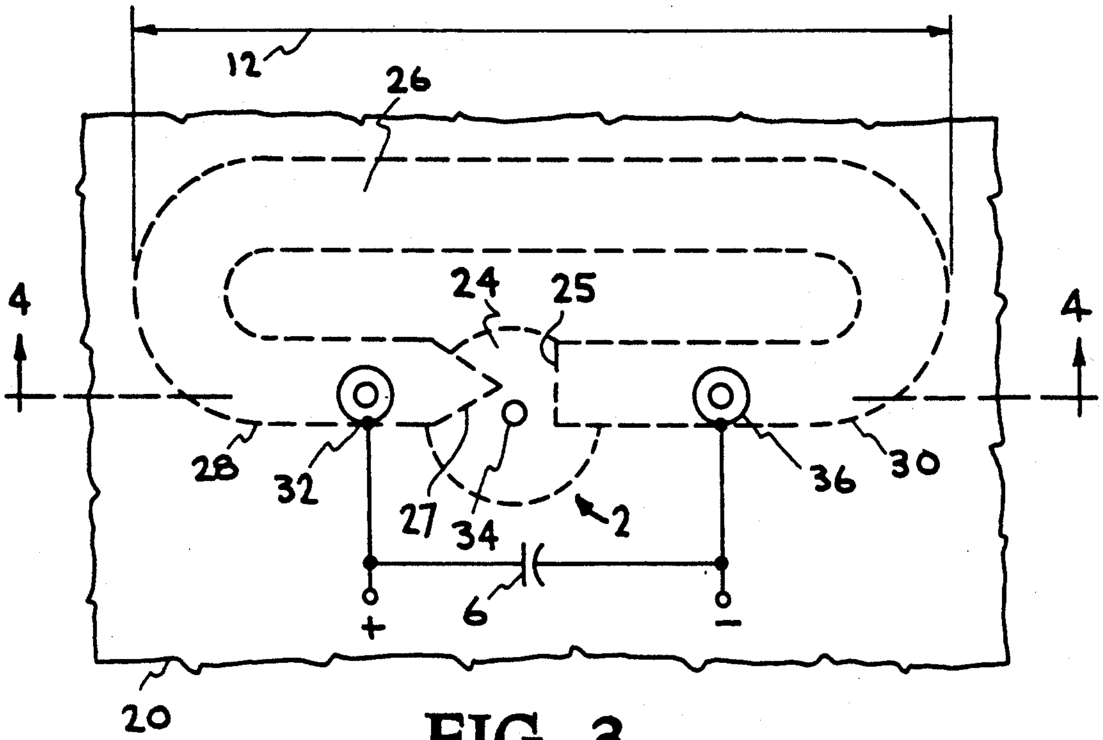
+
−

FIG. 3

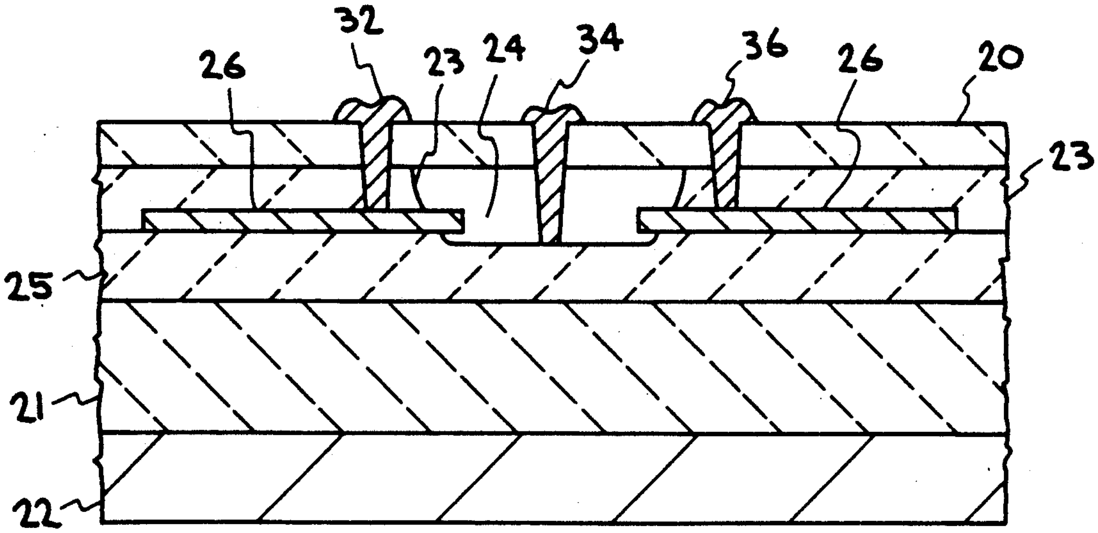

FIG. 4

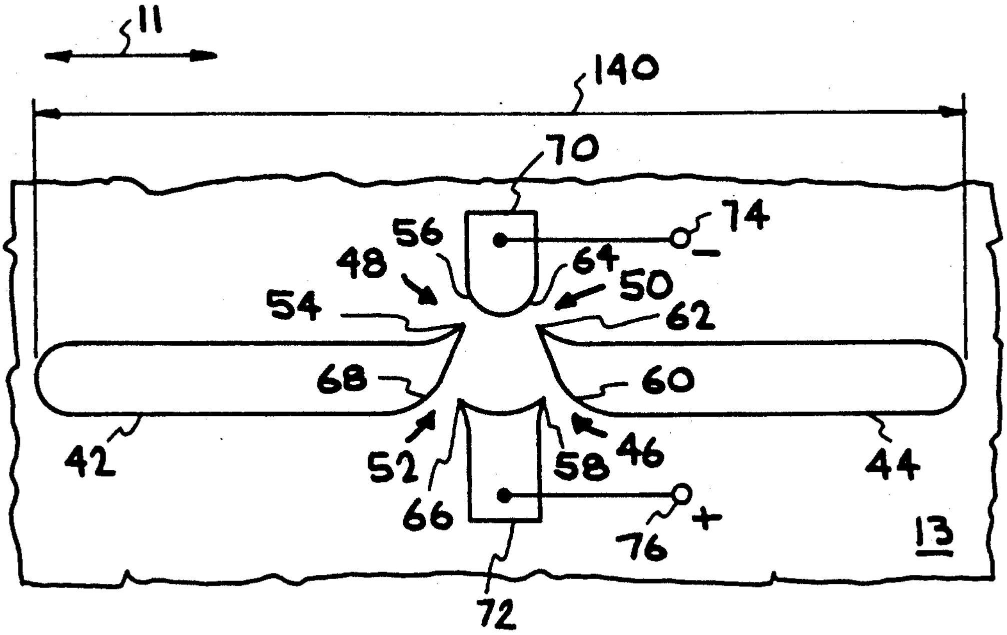
−
+

FIG. 5

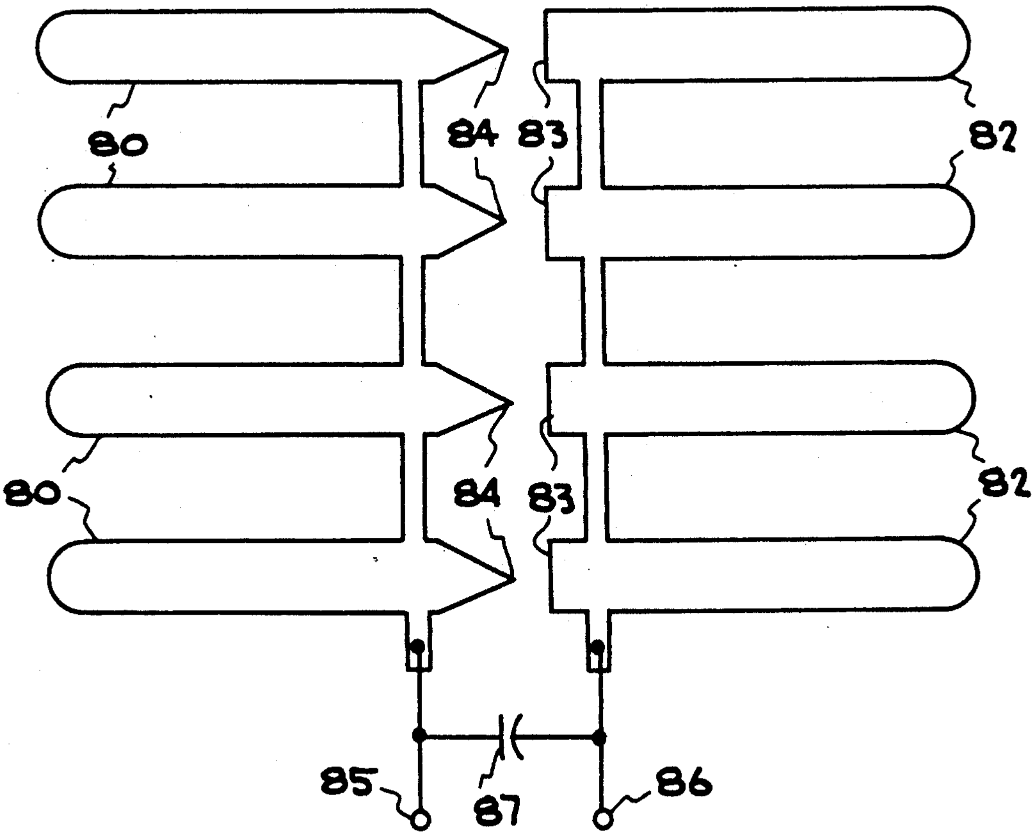

FIG. 6

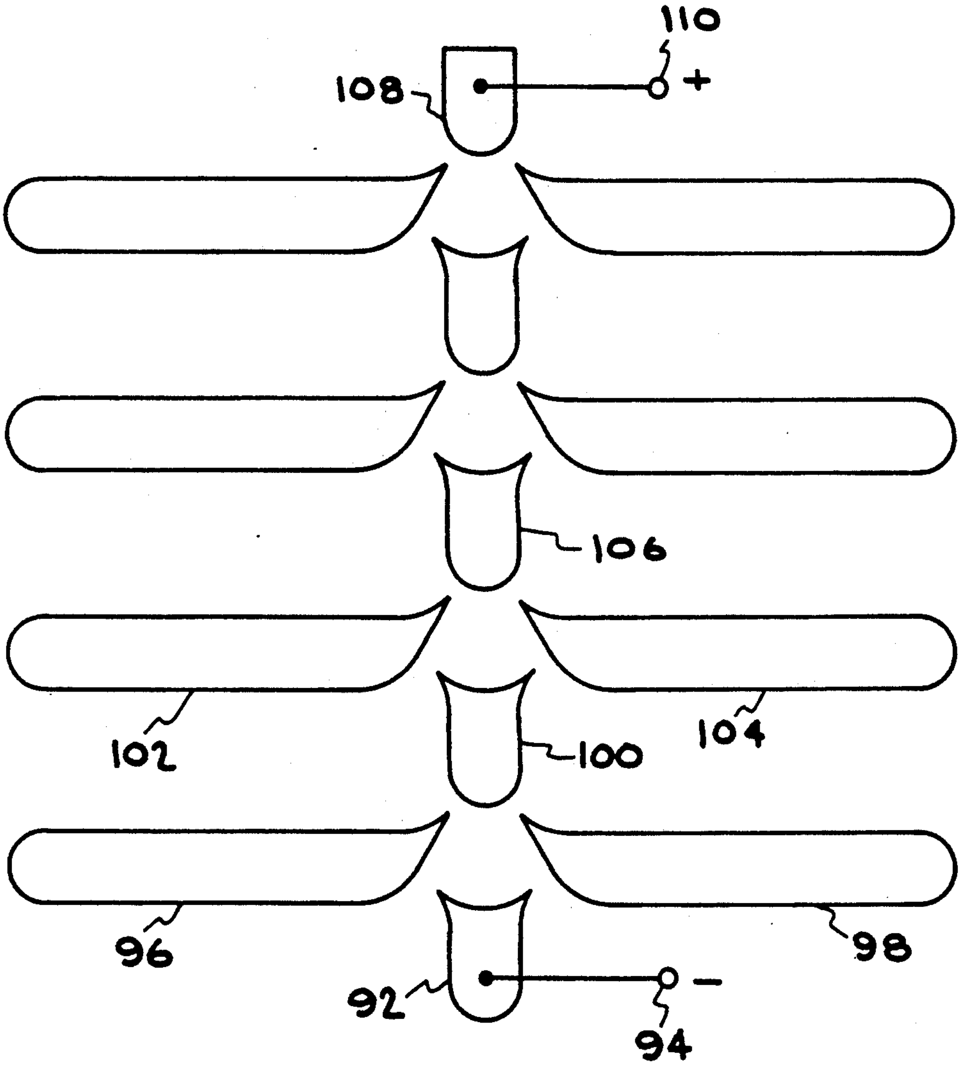
110
108
+
106
102
104
100
96
98
92
-
94

FIG. 7

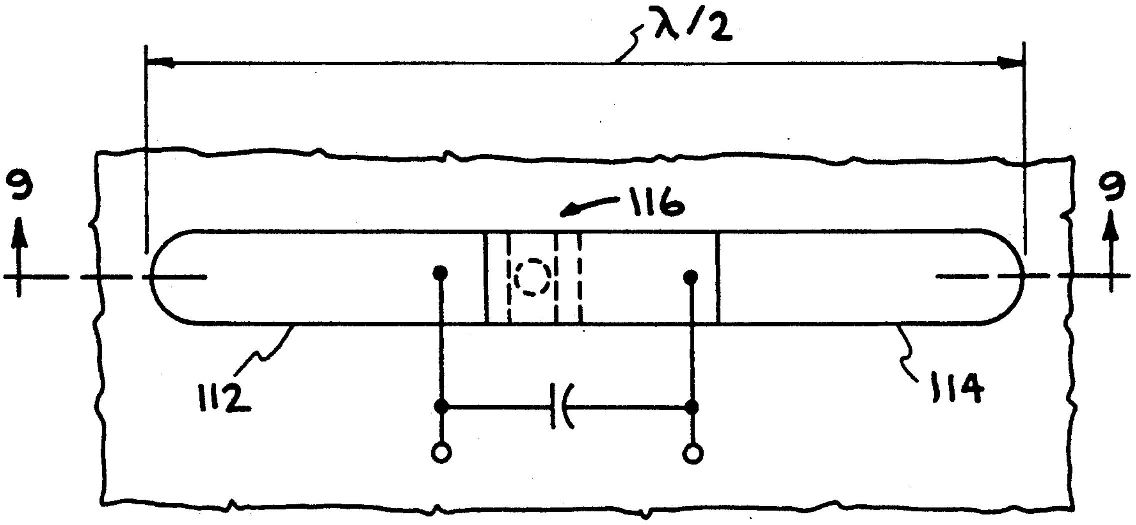
λ/2
9
9
116
112
114

FIG. 8

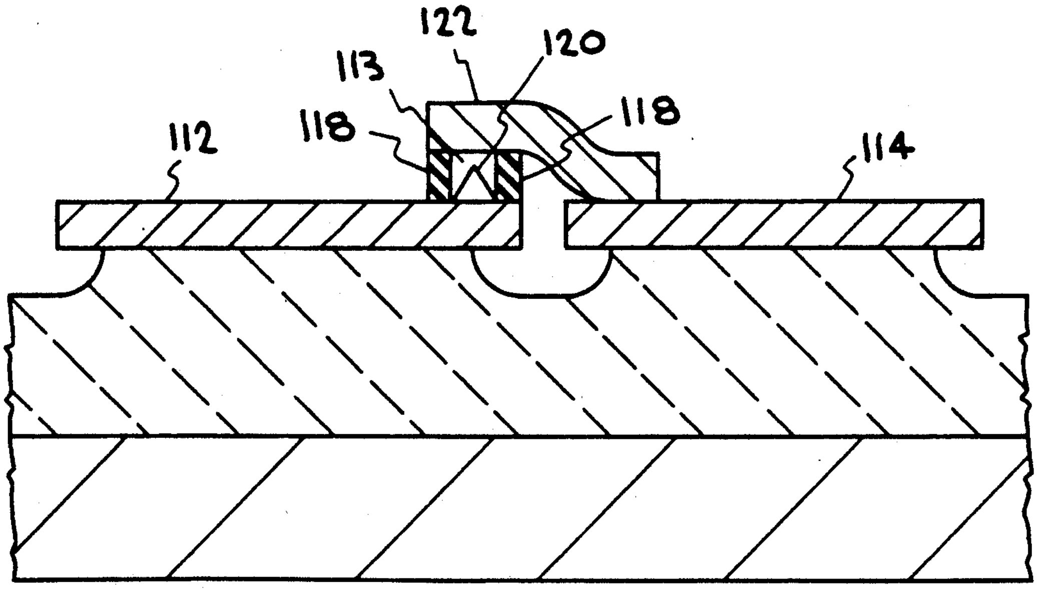
122
120
113
118
118
112
114

FIG. 9

HIGH FREQUENCY RECTENNA

BACKGROUND OF THE INVENTION

The United States Government has rights in this invention pursuant to Contract No. W-7405-ENG-48 between the U.S. Department of Energy and the University of California, for the operation of Lawrence Livermore National Laboratory.

A rectenna, or rectifying antenna, comprises an antenna and a diode rectifier. In operation, the antenna gathers electromagnetic energy at a frequency consistent with its size and shape, and the diode rectifies that energy into D.C. current.

Presently, rectennas utilize a semiconductor diode or large vacuum diode to rectify electromagnetic waves, with frequencies ranging up to the radio frequency spectrum, to direct current. Rectennas utilizing a semiconductor diode have an upper frequency limit and an upper intensity limit due to the inherent limitations of semiconductor diodes. These limitations include the device capacitance and the fragility of high speed semiconductor devices. Rectennas using large vacuum tubes also have upper frequency limits and are expensive to build. Large arrays of rectennas using conventional vacuum tubes would be prohibitively expensive to build, and could not be produced in a compact form.

William C. Brown in "The History of Power Transmission of the Rectenna," in the IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-32, No. 9, September 1984, describes the development of rectennas for rectifying electromagnetic signals with a frequency of 2.45 Giga Hertz (GHz). The article describes research into an apparatus that would be able to allow power to be sent from the earth to a device above the earth or to allow power to be sent from above the earth to the earth via a 2.45 GHz radio signal. The author states that the best diodes that he found for the rectennas are either silicon or gallium arsenide (GaAs) Schottky-barrier diodes.

William C. Brown in "Rectenna Technology Program: Ultra Light 2.45 GHz Rectenna and 20 GHz Rectenna," published by NASA as PT-6902, CR 179558, Mar. 11, 1987 describes the development of a 2.45 GHz rectenna and proposes a 20 GHz rectenna. The article first describes the development of a 2.45 GHz rectenna. The GaAs Schottky barrier diodes for the 2.45 GHz rectenna have specifications including a zero bias capacitance of 3 pf and a reverse breakdown voltage of 60–70 volts. These specifications and other qualities of the Schottky barrier diodes limit the efficiency and the maximum power capabilities of these rectennas. The article goes on to propose a 20 GHz and higher frequency rectenna. The article states that the losses in the diode will be considerably greater at these higher frequencies and the losses in the scaled circuit structure will become greater by the square root of the frequency scaling factor because of skin losses. Because semiconductor diodes break down at high temperatures, semiconductor diodes must be cooled. In the design proposed in the article for a 20 GHz rectenna, reasonable cooling is provided if the rectenna is limited to a DC power output of 0.4 watts per diode element with an output voltage of 8.9 volts and a reverse voltage breakdown of 20 volts. At higher intensities, the resistance in the rectenna will create more heat elevating the semiconductor temperature and causing the semiconductor to breakdown. These parameters help illustrate the inefficiency and fragileness of semiconductor diodes in a rectenna.

Because of the inherent limitations of semiconductor rectennas and large vacuum tube rectennas, there is a need for rectennas that can rectify high frequency and high intensity electromagnetic waves.

The invention provides an array of rectennas, which are able to rectify high frequency microwave or high frequency and high intensity electromagnetic radiation. These arrays may be useful in transportation of power through electromagnetic waves and conversion of intense high frequency electromagnetic radiation into direct current. Applications of the invention include sending power to satellites or high altitude devices from the earth by electromagnetic waves, transmission of power from satellites which collect solar power and convert it to electromagnetic waves which are transmitted to the earth, and direct conversion of microwaves generated by fusion reactions into a direct current.

SUMMARY OF THE INVENTION

It is an object of the invention to provide a rectenna which can rectify electromagnetic radiation with wavelengths equal to or shorter than the wavelengths of microwaves.

It is another object of the invention to provide a rectenna which can rectify electromagnetic waves at high intensities.

It is another object of the invention to provide a rectenna which can rectify electromagnetic radiation with wavelengths equal to or shorter than the wavelengths of microwaves and with high intensities, and with an efficiency greater than 40%.

It is another object of the invention to provide a rectenna which is resilient to high energy radiation and high temperatures.

It is another object of the invention to provide an array of rectennas which can be manufactured in large numbers at a low cost.

It is another object of the invention to provide a compact array of rectennas.

Additional objectives, advantages and novel features of the invention will be set forth in part in the description which follows, and in part will become apparent to those skilled in the art upon examination of the following or may be learned by practice of the invention. The objects and advantages of the invention may be realized and attained by means of the instrumentalities and combinations particularly pointed out in the appended claims.

The invention comprises an antenna, comprising a first half and a second half, and at least one field emission vacuum diode electrically connected between the first half and the second half of the antenna to provide a rectenna. The diode is made an integral part of the antenna so that arrays of rectennas can be used for rectifying electromagnetic radiation with a frequency higher than or equal to microwave frequencies. If the diode was not integral to the antenna, and if an array of antennas was connected together, then phase differences from one side of the array to the other would cause the combined outputs to cancel, significantly decreasing the output. By making the diode an integral part of the antenna, the current is rectified at the source, and phase differences will have little effect on the combined currents.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated and form a part of the specification, illustrate several embodiments of the invention and, together with the description, serve to explain the principles of the invention.

FIG. 1 is a top view of a preferred embodiment of the invention. It illustrates a half wave rectenna using a dipole antenna and a single lateral field emission diode.

FIG. 2 is a cross-sectional view of a half wave rectenna illustrated in FIG. 1 along lines 2.

FIG. 3 is a top view of another embodiment, a half wave rectenna using a folded dipole antenna and a single lateral field emission diode.

FIG. 4 is a cross-sectional view of a half wave rectenna illustrated in FIG. 3 along lines 4.

FIG. 5 is a top view of another embodiment, a full wave rectenna using a dipole antenna and a full wave bridge rectifier using four lateral field emission diodes.

FIG. 6 is a top view of an array of half wave rectennas in parallel.

FIG. 7 is a top view of an array of full wave rectennas in series.

FIG. 8 is a top view of another embodiment, a half wave rectenna using a dipole antenna and a single vertical field emission diode rectifier.

FIG. 9 is a cross-sectional view of a half wave rectenna illustrated in FIG. 8 along lines 9.

DESCRIPTIONS OF THE PREFERRED EMBODIMENTS

FIGS. 1 and 2 show a top and side view of a preferred embodiment of the invention. The invention receives and rectifies an electromagnetic wave with wavelength L. It includes a dipole antenna (1 and 3), coupled with an integral lateral field emission diode 2 at the center, and an external capacitor 6. In the preferred embodiment the total length 10 of the dipole antenna is approximately L/2.

The embodiment may be constructed using conventional semiconductor manufacturing techniques on an insulating substrate 4 attached to a conducting ground plane 5 (FIG. 2). The ground plane makes the antenna directional, with the vertical direction being the preferred direction. The separation 155 between the antenna and the ground plane is preferably approximately L/4 to provide a maximum reflection of electromagnetic radiation which passes through the antenna 1 and 3 to the ground plane. The actual distance would depend on the material used as the insulating substrate.

Using conventional semiconductor techniques, a layer of conductive material (tungsten, or doped polysilicon for example) is deposited on the insulator and then shaped by photolithography and wet or dry etching. The insulator under the layer of conductive material is then etched, undercutting the conductor. This undercutting frees the edge of the conductor to prevent surface leakage currents and surface breakdown along the insulator between the two sides of the antenna.

The method of making a field emission vacuum triode is disclosed in U.S. Pat. No. 4,855,636 by Busta et al, which is incorporated by reference. Using the method described in Busta et al. a field emission diode in a vacuum could be created. By using a differently shaped mask the field emission diode can be integrated into a dipole antenna. Even though Busta discloses the creation of a triode, the field emission diode could be created in a similar manner by either sealing the etching holes described in Busta with an insulating material or by not applying a charge to the gate of the triode. For cases where the mean free path of an electron in a gas at some pressure is comparable to the separation of the anode and cathode, the diode will operate in that gas at that pressure as if it were in a vacuum. In addition, a low pressure or near vacuum can also be established around the field emission diode by sealing the entire rectenna in a near vacuum container.

Electromagnetic radiation of wavelength L is absorbed by the dipole antenna 1 and 3. According to well known principles of antenna design, the electromagnetic radiation will set up an oscillating potential difference between the first half 1, and the second half 3 of the dipole antenna. The potential difference will oscillate with the same frequency as the absorbed electromagnetic radiation. This oscillating potential difference will manifest itself across the lateral field emission diode 2 whose cathode, manifested by cathode tip 15 which is a pointed edge of the first half 1 of the antenna, is integral with the first half 1 of the antenna, and whose anode, manifested by anode surface 16 which is a nonpointed edge of the second half 3 of the antenna, is integral with the second half 3 of the antenna. The pointed edge making the cathode tip 15 is located on a part of the first half 1 of the antenna that is located closest to the second half 3 of the antenna as shown. The nonpointed edge making the anode surface 16 is located on a part of the second half 3 of the antenna that is located closest to the first half 1 of the antenna.

During the positive half cycle of the electromagnetic radiation, the second half 3 of the antenna is positive with respect to the first half 1 of the antenna, and field emission of electrons occurs off of the tip of the cathode 15. The field emitted electrons will be absorbed at the anode surface 16, causing a surplus of electrons in the second half 3 of the antenna. During the negative half cycle, the second half 3 of the antenna is negative with respect to the first half 1 of the antenna. Since the anode surface 16 is flat, it does not emit an appreciable amount of electrons and thus no field emission current flows in the negative half of the cycle. Thus, a voltage is applied to the capacitor 6, and any load 9 attached to terminals 7 and 8. This allows the surplus of electrons in the second half 3 of the antenna to flow through the attached load 9 back to the first half 1 of the antenna. The capacitor 6 smooths the voltage available at the terminals by charging from the antenna during the negative half cycle and discharging through the load 9 during the positive half cycle. This illustrates how the cathode tip 15 forming the cathode is integrally and electrically connected to the first half 1 of the antenna and how the anode is integrally and electrically connected to the second half 3 of the antenna and how a field emission diode 2 comprising the anode 16 and cathode 15 are electrically connected between the first half 1 and the second half 3 of the antenna.

In the preferred embodiment shown in FIG. 1, the rectenna is designed to rectify an electromagnetic wave having a predetermined frequency between 1,000 GHz and 10 GHz. This would require that the separation 17 between the anode and cathode be less than 10 microns, and preferably less than 3 microns.

The rectenna is not limited to a dipole antenna design, nor to the half wave rectifier described above. FIG. 3 shows a similar structure, but using a folded dipole antenna **26** instead of the simple dipole antenna. The folded dipole antenna **26** typically has a higher impedance than the simple dipole, which may make a better coupling between the antenna and the diode. The folded dipole antenna **26** could be manufactured in the same way as the embodiment shown in FIG. **1**, having the same cross-sectional view. For illustrative purposes, the folded dipole antenna shown in FIG. **3** is made using the method described in U.S. Pat. No. 4,855,636 by Busta et al. FIG. **4** is a cross-sectional view of FIG. **3** along lines **4**. In this embodiment a silicon wafer **21** is attached to a conducting layer **22**. Preferably the silicon wafer **21** has a thickness equal to approximately ¼ the wavelength of the electromagnetic radiation to be rectified. As described in U.S. Pat. No. 4,855,636 approximately 1 micron of thermal silicon dioxide is grown on the silicon wafer **21** forming a first insulating layer **25**. A conducting layer is deposited on the first insulating layer **25** to form the folded dipole antenna **26** with a first half **28** and second half **30** of the antenna and an integrated anode **25** and cathode **27**. In the preferred embodiment the length **12** of the folded dipole antenna **26** is approximately equal to L/2, where L is the wavelength of the electromagnetic radiation to be rectified. The entire structure is then covered by a layer of undoped, highly resistive polycrystalline silicon **23**. The upper surface of the polycrystalline silicon **23** is thermally oxidized to form a top insulating layer **20**. Windows are made through the top insulating layer **20** to allow the etching of a chamber **24** and holes for contacts **32** and **36**. A metallic layer is then applied over the top insulating layer **20** and then etched back to create contacts **32** and **36** and for creating a window plug **34**.

The folded dipole rectenna shown in FIGS. **3** and **4** rectifies electromagnetic waves in the same manner as the dipole rectenna described above. Using the described method of building the folded dipole rectenna, the integrated diode can be made so that the separation between the cathode tip **27** and the anode surface **25** would be in the above described range for rectifying signals with frequencies higher than 10 GHz. At an operating voltage of 500 volts it is estimated that the leakage current due to parasitic current through the undoped polycrystalline silicon layer **23** would be 500 pico Amps. This leakage could be further reduced by replacing the polycrystalline silicon side walls with silicon dioxide side walls. The integration of field emission diodes with an antenna in this embodiment provides a very efficient rectenna that can rectify signals with frequencies equal to or greater than microwaves. In addition, the provided diodes can easily handle 500 volts, wherein the Schottky barrier diodes used in the prior art had a reverse breakdown voltage of 60–70 volts.

FIG. **5** shows a rectenna with a dipole antenna, comprising a first strip of conductive material on a substrate **13** with a length along a first direction **11** less than L/2, where L is the wavelength of the electromagnetic radiation to be rectified, forming a first half **42** of the antenna and a second conductive strip on a substrate **13** with a length along the first direction **11** less than L/2 forming a second half **44** of the antenna, and a full wave bridge rectifier. The second conductive strip **44** is located along the first direction **11** with respect to the first conductive strip **42**. The rectifier utilizes four field emission diodes **46**, **48**, **50** and **52** arranged in a full wave bridge configuration. The configuration illustrated in FIG. **5** is manufactured in the same manner as the embodiment illustrated in FIG. **1**. It would also be possible for this embodiment to be manufactured in the same manner as the embodiment illustrated in FIG. **3**.

Electromagnetic radiation of wavelength L is absorbed by the dipole antenna **42** and **44**. According to well known principles of antenna design, incident electromagnetic radiation will set up an oscillating potential difference between the first half **42**, and the second half **44** of the dipole antenna. The potential difference will oscillate at the same frequency as the absorbed electromagnetic radiation. This oscillating potential difference will manifest itself across the lateral field emission diodes.

During the positive half cycle of the electromagnetic radiation, the second half **44** of the antenna is positive with respect to the first half **42** of the antenna, and field emission of electrons occurs off of the tip of a cathode tip **54** of a first diode **48**, wherein the cathode tip **54** is an integral part of the first half **42** of the antenna formed by a pointed edge of the first strip of conductive material which forms the first half **42** of the antenna. The field emitted electrons will be absorbed at the anode surface **56** of the first diode **48**, which is formed by a nonpointed edge of a third strip of conductive material forming an anode terminal **70**, causing a surplus of electrons in the anode terminal **70**. Also during the positive half cycle of the electromagnetic radiation, field emission of electrons occurs off of the tip of cathode **58** of a second diode **46**, wherein the cathode tip **58** is formed by a pointed edge of a fourth strip of conductive material on the substrate **13**, wherein the fourth strip of conductive material forms a cathode terminal **72**. The field emitted electrons will be absorbed at the anode surface **60** of the second diode **46**, wherein the anode surface **60** is an integral part of the second half **44** of the antenna, causing a surplus of electrons in the second half **44** of the antenna. During the negative half cycle, the second half **44** of the antenna is negative with respect to the first half **42** of the antenna, and field emission of electrons occurs off of the tip of the cathode tip **62** of a third diode **50**, wherein the cathode tip **62** is an integral part of the second half **44** of the antenna. The field emitted electrons will be absorbed at the anode surface **64** of the second diode **50**, which is part of the anode terminal **70**, causing a surplus of electrons in the anode terminal **70**. Also during the negative half cycle of the electromagnetic radiation, field emission of **5** electrons occurs off of the tip of cathode **66** of a fourth diode **52**, wherein the cathode tip **66** is part of the cathode terminal **72**. The field emitted electrons will be absorbed at the anode surface **68** of the fourth diode **52**, wherein the anode surface **68** is an integral part of the first half **42** of the antenna, causing a surplus of electrons in the first half **42** of the antenna.

Thus, during the full wavelength electrons flow from the cathode terminal **72** to either the first **42** or second **44** half of the antenna and then to the anode terminal **70**. The surplus of electrons gathered at the anode terminal **70** can flow through the negative load terminal **74** electrically connected to the anode terminal and then through a load to a positive load terminal **76** electrically connected to the cathode terminal **72**. The anode terminal **70** acts as an anode for both the first part **42** of the antenna and the second part **44** of the antenna, and the cathode terminal **72** acts as a cathode for both the first part **42** of the antenna and the second part **44** of the antenna. Although a capacitor is not needed, it may be used to smooth the current flow between terminals **74**, **76**.

In the preferred embodiment of the invention, the total length **140** of the first half **42** and the second half **44** of the antenna along the first direction **11** is approximately equal to L/2. As shown in FIG. **5**, the four rectifying field emission diodes are electrically connected between the first half **42** and the second half **44** of the antenna making two parallel sets of two diodes in series.

An advantage of the invention is that a large number of rectennas could be inexpensively combined in an array using conventional semiconductor manufacturing techniques. FIGS. **6** and **7** illustrate different arrays of rectennas. FIG. **6** shows a parallel array of the half wave rectennas shown in FIG. **1**. For clearer illustration of the embodiment, only the pattern of conductive strips is shown. Each half wave rectenna in the array has a first half **80** and a second half **82** of an antenna. Each first half **80** of an antenna has an integrally connected cathode tip **84**. Each second half **82** of an antenna has an integrally connected anode surface **86**. Each first half **80** of an antenna is electrically connected to a first load terminal **85** and a first half of a capacitor **87**. Each second half **82** of an antenna is electrically connected to a second load terminal **86** and a second half of the capacitor **87**. Each individual rectenna would function like the rectenna embodiment illustrated in FIG. **1**. Attaching them in a parallel configuration as illustrated would increase the current available at the terminals.

FIG. **7** shows a series array of the full wave rectennas shown in FIG. **5**. For clear illustration only the pattern of conductive strips is shown. The array comprises a first terminal **92** electrically connected to a first load terminal **94**. The electromagnetic signal to be rectified will cause electrons to flow from the first terminal **92** to the first half **96** of a first antenna and the second half **98** of the first antenna. The electromagnetic signal will also cause electrons to flow from the first half **96** of the first antenna and the second half **98** of the first antenna to a second terminal **100**. The second terminal **100** acts as an anode terminal to the first antenna. The electromagnetic signal will cause electrons to flow from the second terminal **100** to the first half **102** of a second antenna and the second half **104** of the second antenna. The electromagnetic signal will also cause electrons to flow from the first half **102** of the second antenna and the second half **104** of the second antenna to a third terminal **106**. The second terminal **100** acts as a cathode terminal to the second antenna, and the third terminal **106** acts as an anode terminal to the second antenna. The electrons continue to flow through the subsequent antenna elements and terminals until they reach a final terminal **108**, which is electrically connected to a second load terminal **110**. This series configuration will increase the voltage across the first **94** and second **110** load terminals.

In either the parallel or the series configurations, an external capacitor could be connected to the load terminals.

In addition to the lateral diode shown so far, a vertical field emission diode can also be used as the rectifier in a rectenna. FIGS. **8** and **9** show a half wave rectenna using a half wave vertical field emission diode **116** and a dipole antenna with a first half **112** and a second half **114**. The vertical field emission diode comprises an insulator **118** placed on the first half **112** of the antenna. In a hole **113**, formed in the insulator, a conductive cone or pyramid **120** is formed in electrical contact with the first **112** half of the antenna. A conducting cap **122** is placed over the hole **113**, that also makes electrical contact with the second half **114** of the antenna. The hole **113** can be evacuated and sealed, forming an integral vacuum diode. In operation this embodiment works in the same manner as the embodiment illustrated in FIGS. **1** and **2**, except that the vertical diode causes the electrons to be emitted vertically off of the tip of the cone or pyramid **120** instead of horizontally off of the point. The manufacturing of vertical field emission diodes is described in U.S. Pat. No. 4,721,885 by Brodie, and U.S. Pat. Nos. 3,755,704 and 3,789,471 by Spindt incorporated by reference.

The foregoing description of preferred embodiments of the invention have been presented for purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed, and obviously many modifications and variations are possible in light of the above teaching. The embodiments were chosen and described in order to best explain the principles of the invention and its practical application to thereby enable others skilled in the art to best utilize the invention in various embodiments and with various modifications as are suited to the particular use contemplated. It is intended that the scope of the invention be defined by the claims appended hereto.

We claim:

1. A rectifying antenna for rectifying electromagnetic radiation with a wavelength shorter than 3.0 centimeters, comprising:

an antenna having a first half and a second half; and

a first field emission diode for rectifying a current flow, wherein the first field emission diode has a first anode and first cathode electrically connected between the first half of the antenna and the second half of the antenna, and wherein the first cathode is separated from the first anode by a distance of less than 10 microns and wherein the antenna has a length which is approximately equal to half of the wavelength of the electromagnetic radiation to be rectified.

2. A rectifying antenna as claimed in claim **1**, further comprising:

a first load terminal electrically connected to the first half of the antenna; and

a second load terminal electrically connected to the second half of the antenna, and wherein the first cathode is electrically connected to the first half of the antenna, and the first anode is electrically connected to the second half of the antenna.

3. A rectifying antenna as claimed in claim **2**, wherein the first cathode is integral with the first part of the first antenna, and wherein said first cathode comprises a pointed edge of the first part of the antenna, and wherein the first anode is integral with the second part of the antenna, and wherein the first anode comprises a nonpointed edge of the second part of the antenna.

4. A rectifying antenna as claimed in claim **2**, wherein the first field emission diode is a linear field emission diode.

5. A rectifying antenna as claimed in claim **2**, wherein the first field emission diode is a vertical field emission diode.

6. A rectifying antenna as claimed in claim **1**, further comprising:

a first terminal, wherein the first field emission diode is electrically connected between the first terminal and the second half of the antenna with the first anode being electrically connected to the first terminal and the first cathode being electrically connected to the second half of the antenna;

a second terminal;

a second field emission diode with a second cathode and second anode electrically connected between the first terminal and the first half of the antenna with the second anode being electrically connected to the first terminal and the second cathode being electrically connected to the first half of the antenna;

a third field emission diode with a third cathode and a third anode electrically connected between the second terminal and the first half of the antenna with the third anode being electrically connected to the first half of the antenna and the third cathode being electrically connected to the second terminal; and

a fourth field emission diode with a fourth cathode and a fourth anode electrically connected between the second terminal and the second half of the antenna with the fourth anode being electrically connected to the second half of the antenna and the fourth cathode being electrically connected to the second terminal.

7. A rectifying antenna as claimed in claim 6, wherein the first cathode is formed by a pointed edge of the second half of the antenna, and wherein the second cathode is formed by a pointed edge of the first half of the antenna, and wherein the first and second anodes are formed by nonpointed edges of the first terminal, and wherein the third anode is formed by a nonpointed edge of the first half of the antenna, and wherein the fourth anode is formed by a nonpointed edge of the second half of the antenna, and wherein the third and fourth cathodes are formed by pointed edges of the second terminal.

8. A rectifying antenna as claimed in claim 6, wherein the field emission diode is a lateral field emission diode.

9. A rectifying antenna for rectifying electromagnetic radiation with a wavelength shorter than 3.0 centimeters, comprising:

a substrate comprising a layer of insulating material;

a first cathode;

a first anode;

a first conductive strip on the substrate, with a length in a first direction, wherein the first conductive strip is integrally and electrically connected to the first cathode;

a second conductive strip on the substrate, with a length in the first direction, wherein the second conductive strip is integrally and electrically connected to the first anode, and wherein the second conductive strip is located along the first direction with respect to the first conductive strip, and wherein the length of the first conductive strip in the first direction is less than half of the wavelength and wherein the length of the second conductive strip in the first direction is less than half of the wavelength, and wherein the length of the first conductive strip in the first direction summed with the length of the second conductive strip in the first direction make a sum which is approximately equal to half of the wavelength;

a first load terminal electrically connected to the first conducting strip; and

a second load terminal electrically connected to the second conducting strip, wherein the first cathode is located on the part of the first conductive strip closest to the second conductive strip, and wherein the first anode is located on the part of the second conductive strip closest to the first cathode, and wherein the first anode and the first cathode are separated by a distance less than 10 microns forming a first field emission diode.

10. A rectifying antenna as claimed in claim 9, wherein the cathode is formed by a pointed edge of the first conductive strip, and wherein the anode is formed by a nonpointed edge of the second conductive strip.

11. A rectifying antenna as claimed in claim 9, wherein the first field emission diode is a vertical field emission diode.

12. A rectifying antenna for rectifying electromagnetic radiation with a wavelength shorter than 3.0 centimeters, comprising:

a substrate comprising a layer of insulating material with a first surface;

a first conductive strip on a first surface of the substrate with a length in a first direction wherein a pointed edge of the first conductive strip forms a cathode of a linear field emission diode; and

a second conductive strip on the first surface of the substrate with a length in the first direction wherein the second conductive strip forms an anode of a linear field emission diode, and wherein the second conductive strip is located along the first direction with respect to the first conductive strip, and wherein the length of the first conductive strip in the first direction is less than half the wavelength, and wherein the length of the second conductive strip in the first direction is less than half the wavelength and wherein the length of the first conductive strip in the first direction summed with the length of the second conductive strip in the first direction make a total length which is approximately equal to half of the wavelength.

13. A rectifying antenna as claimed in claim 12, further comprising a conductive layer on a second surface of the substrate, wherein the conductive layer is separated from the first conductive strip by a distance which is approximately equal to one fourth of the wavelength.

14. A rectifying antenna as claimed in claim 13, further comprising:

a first load terminal electrically connected to the first conductive strip; and

a second load terminal electrically connected to the second conducting strip, wherein the cathode is located on the part of the first conductive strip closest to the second conductive strip, and wherein the anode is located on the part of the second conductive strip closest to the cathode, and wherein the first anode and the first cathode are separated by a distance less than 10 microns forming a field emission diode.

15. A rectifying antenna as claimed in claim 13, further comprising:

a third conductive strip on the first surface of the substrate with first and second nonpointed edges that form first and second anodes, wherein a pointed edge of the second conductive strip forms a cathode, and the first anode of the third conductive strip is located less than 10 microns away from the cathode on the first conductive strip forming a first field emission diode and the second anode of the third conductive strip is located less than 10 microns away from the cathode on the second conductive strip forming a second field emission diode; and

a fourth conductive strip on the first surface of the substrate with first and second pointed edges that form first and second cathodes, wherein a non-pointed edge of the first conductive strip forms an anode, and the first cathode of the fourth conductive strip is located less than 10 microns away from the anode on the first conductive strip forming a third field emission diode and the second cathode of the fourth conductive strip is located less than 10 microns away from the anode on the second conductive strip forming a fourth field emission diode.

* * * * *

United States Patent [19]
Chung

[11] **Patent Number: 5,059,582**
[45] **Date of Patent: Oct. 22, 1991**

[54] SUPERCONDUCTOR-METAL LAMINATES AND METHOD OF MAKING

[75] Inventor: **Deborah D. L. Chung**, Pittsburgh, Pa.

[73] Assignee: **The Research Foundation of State University of NY**, Albany, N.Y.

[21] Appl. No.: **326,918**

[22] Filed: **Mar. 22, 1989**

[51] **Int. Cl.^{5}** **B32B 9/00**

[52] **U.S. Cl.** **505/1**; 505/701; 505/702; 505/703; 505/704; 428/323; 428/408; 428/457; 428/688; 428/930

[58] **Field of Search** 505/1, 701-704; 428/688, 930, 457, 408, 323; 29/599

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,985,281	10/1976	Diepers et al.	427/62
4,195,199	3/1980	Hillmann et al.	29/599
4,336,420	6/1982	Benz	29/599

OTHER PUBLICATIONS

IEEE Transaction on Magnetics, vol. May 17, 1981, Spenor et al., pp. 1006-1009.
E. I. Monthly, Carbon Fiber Reinforced Zin-SC composites, Journ Matls Research, vol. 4, No. 6, 11/12/89, pp. 1339-1346.

Primary Examiner—Patrick J. Ryan
Attorney, Agent, or Firm—Michael L. Dunn; James F. Mudd; Ellen K. Park

[57] **ABSTRACT**

A superconducting laminate having at least one layer of metal and at least one layer of superconducting material. The metal layer and the superconducting layer are bonded. The metal later may also include carbon fibers from various precursors. The superconductor may be a composite material. The invention also includes a method of making the laminates.

48 Claims, 16 Drawing Sheets

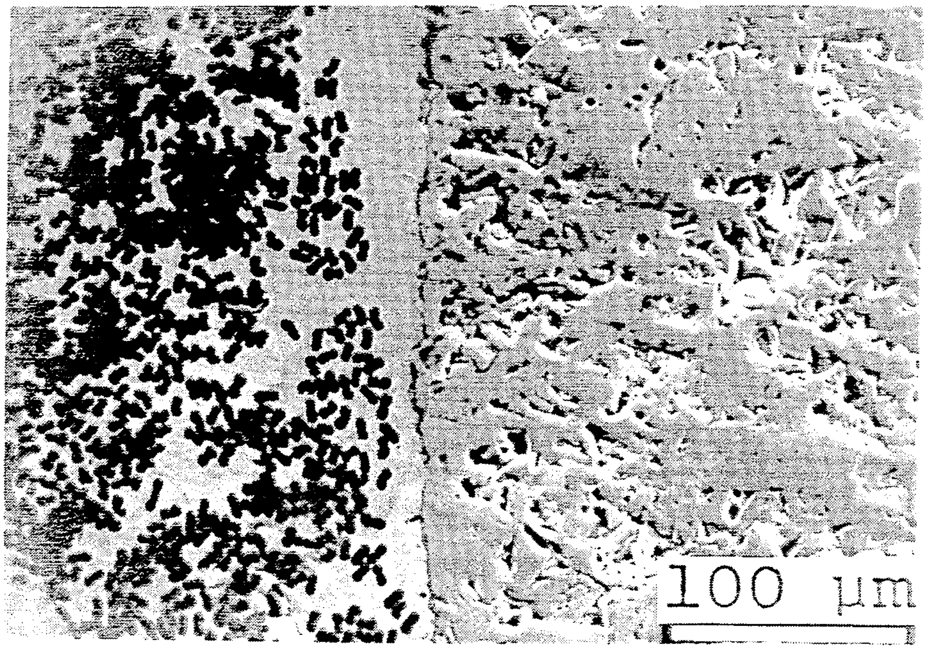

Fig. 1.

Fig. 2.

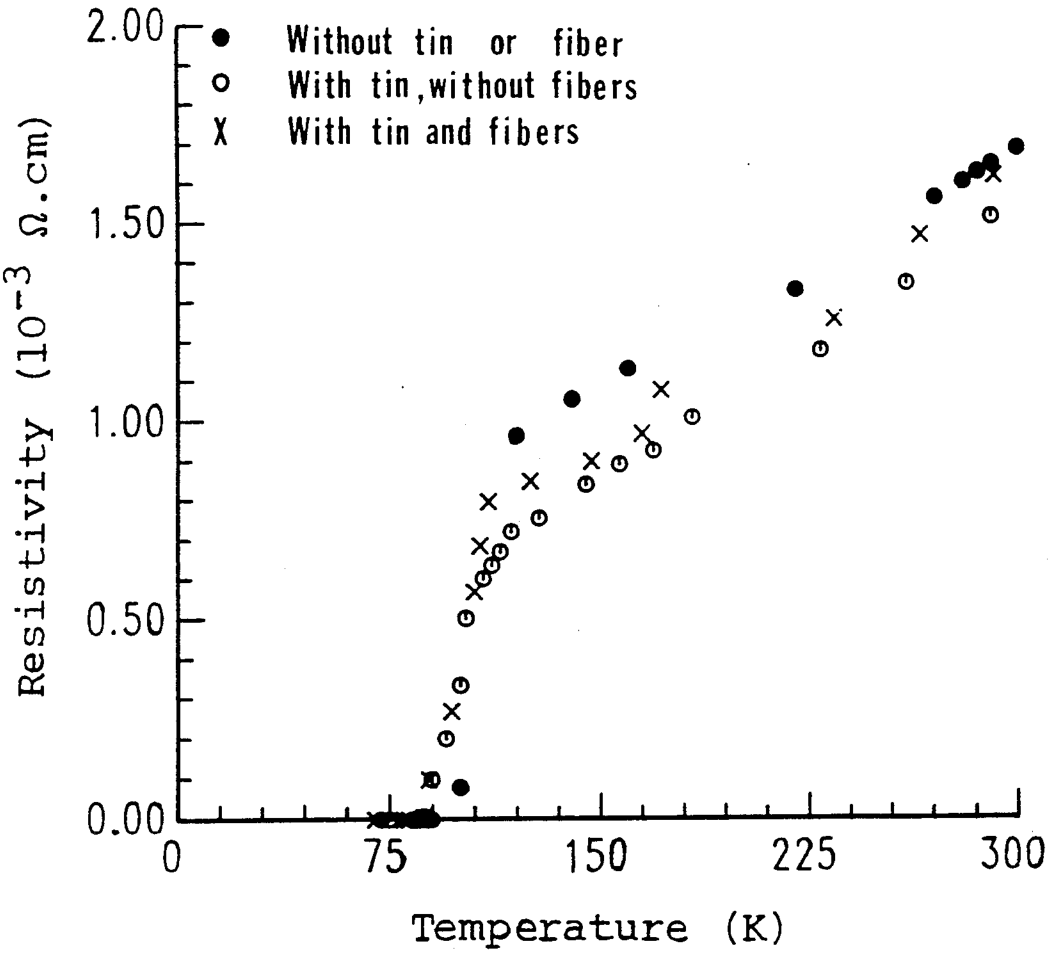
● Without tin or fiber
○ With tin, without fibers
X With tin and fibers
Resistivity (10^{-3} Ω.cm)
2.00
1.50
1.00
0.50
0.00
0
75
150
225
300
Temperature (K)

Fig. 3.

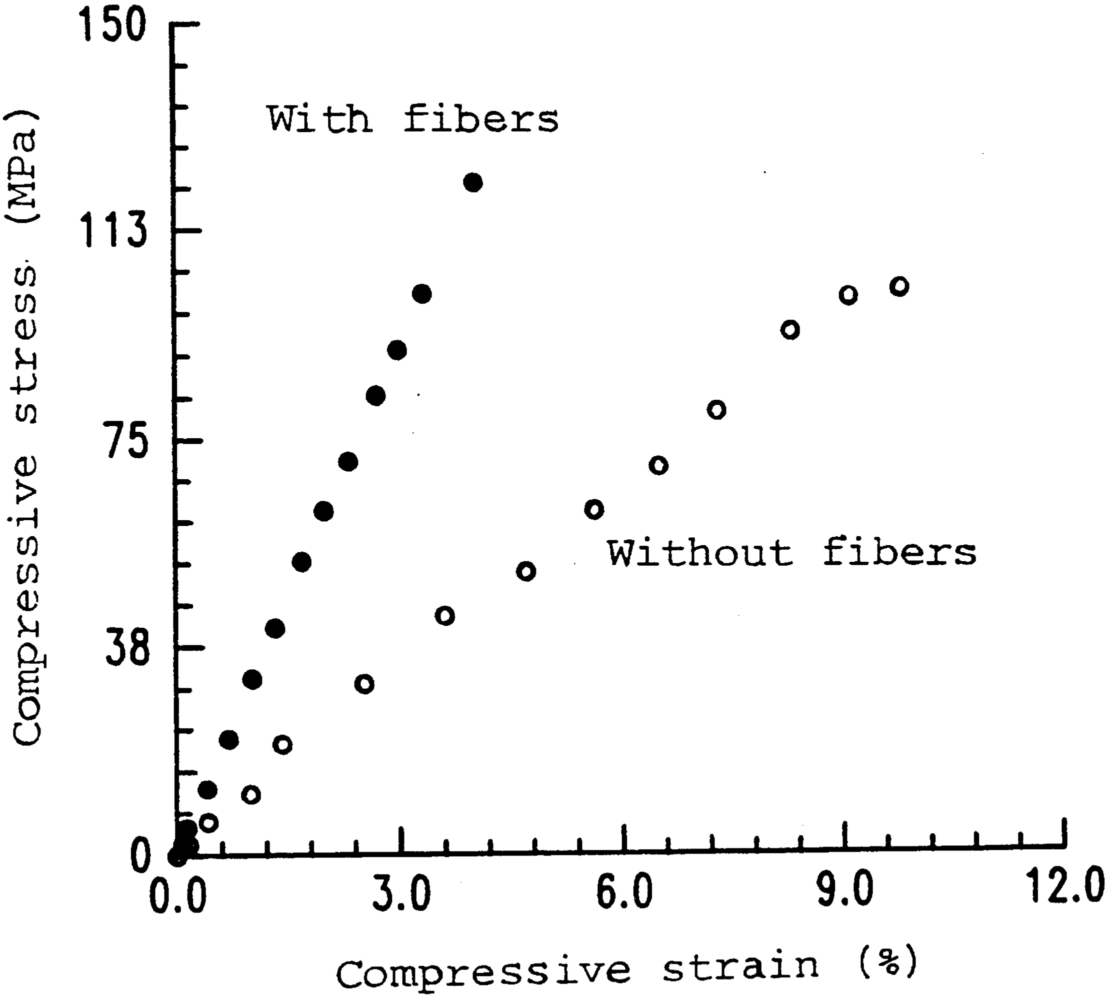
150
113
75
38
0
Compressive stress (MPa)
With fibers
Without fibers
0.0
3.0
6.0
9.0
12.0
Compressive strain (%)

Fig. 4.

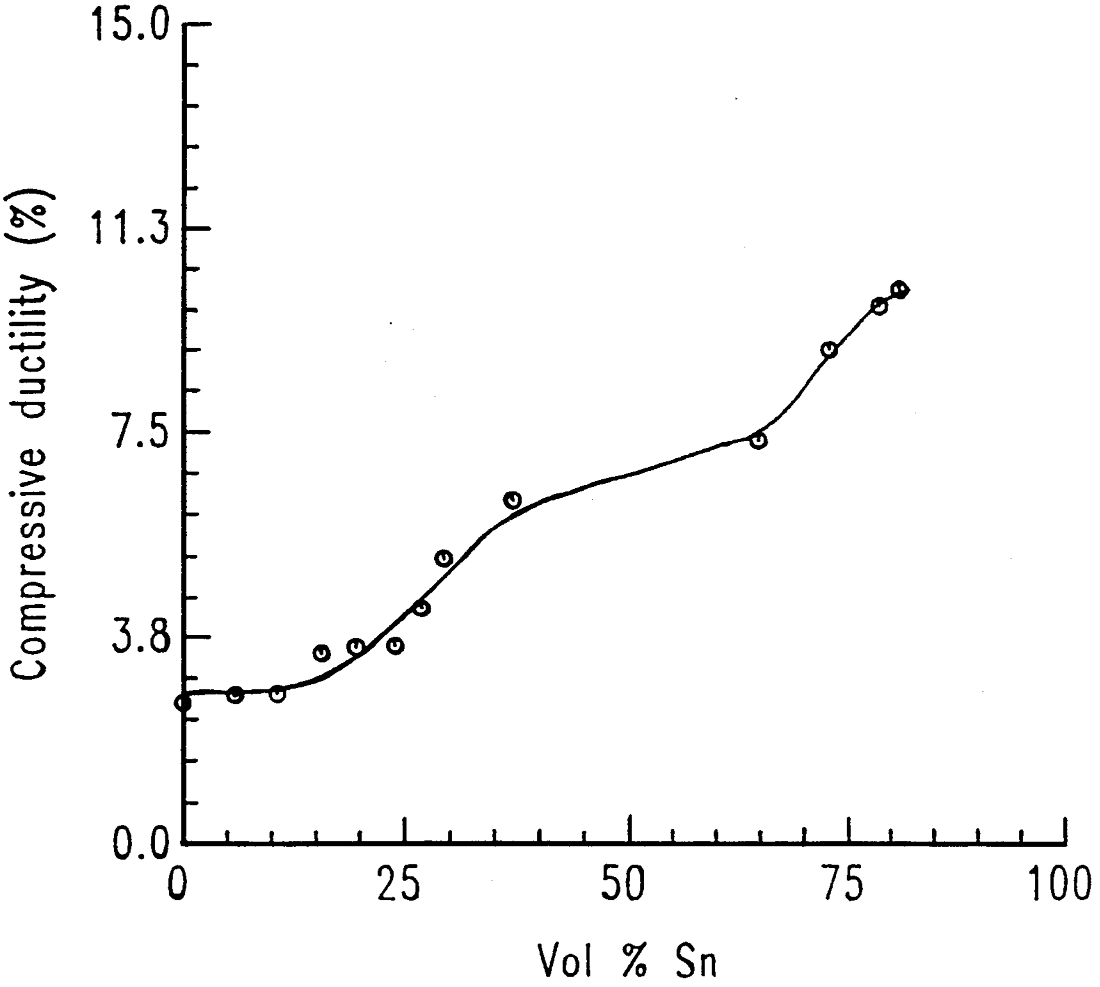

Fig. 5.

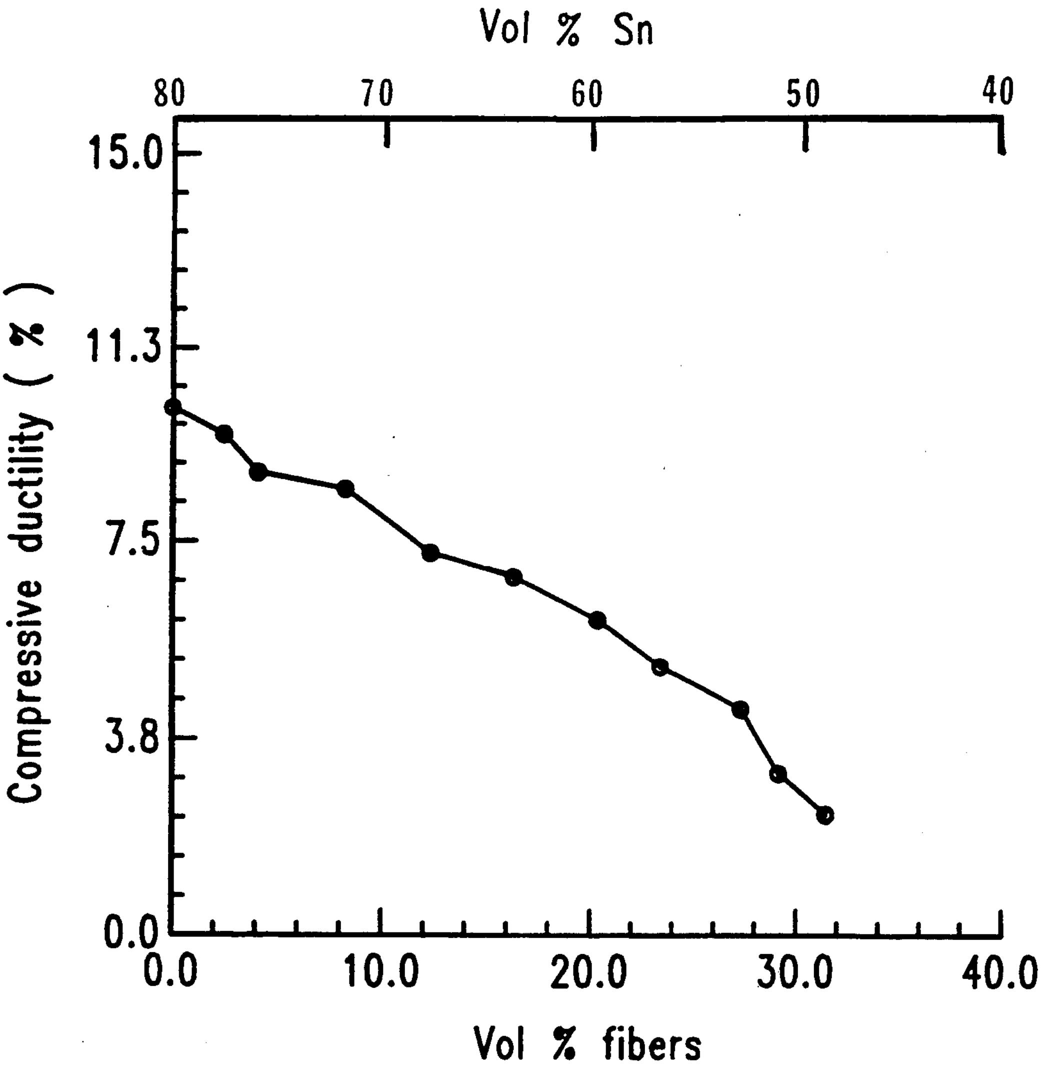
Vol % Sn
80
70
60
50
40
15.0
11.3
7.5
3.8
0.0
Compressive ductility (%)
0.0
10.0
20.0
30.0
40.0
Vol % fibers

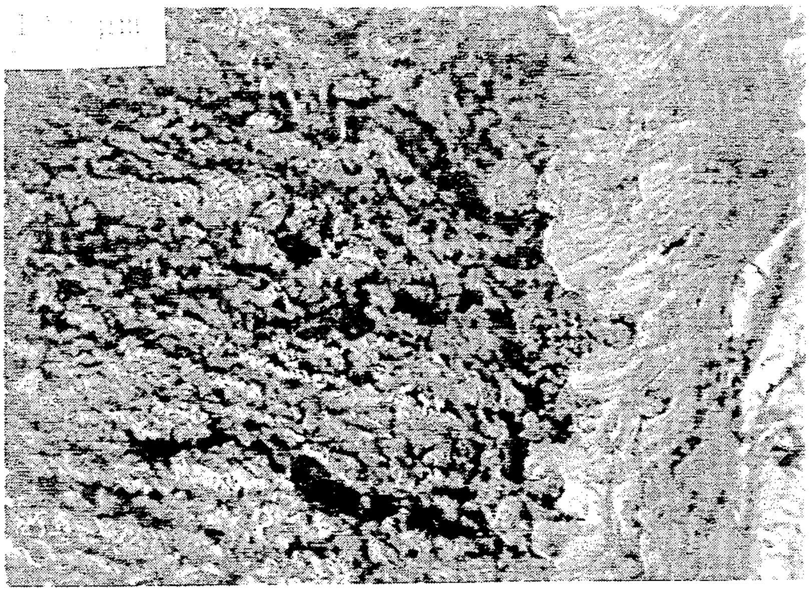

Fig. 6.

Fig. 7.

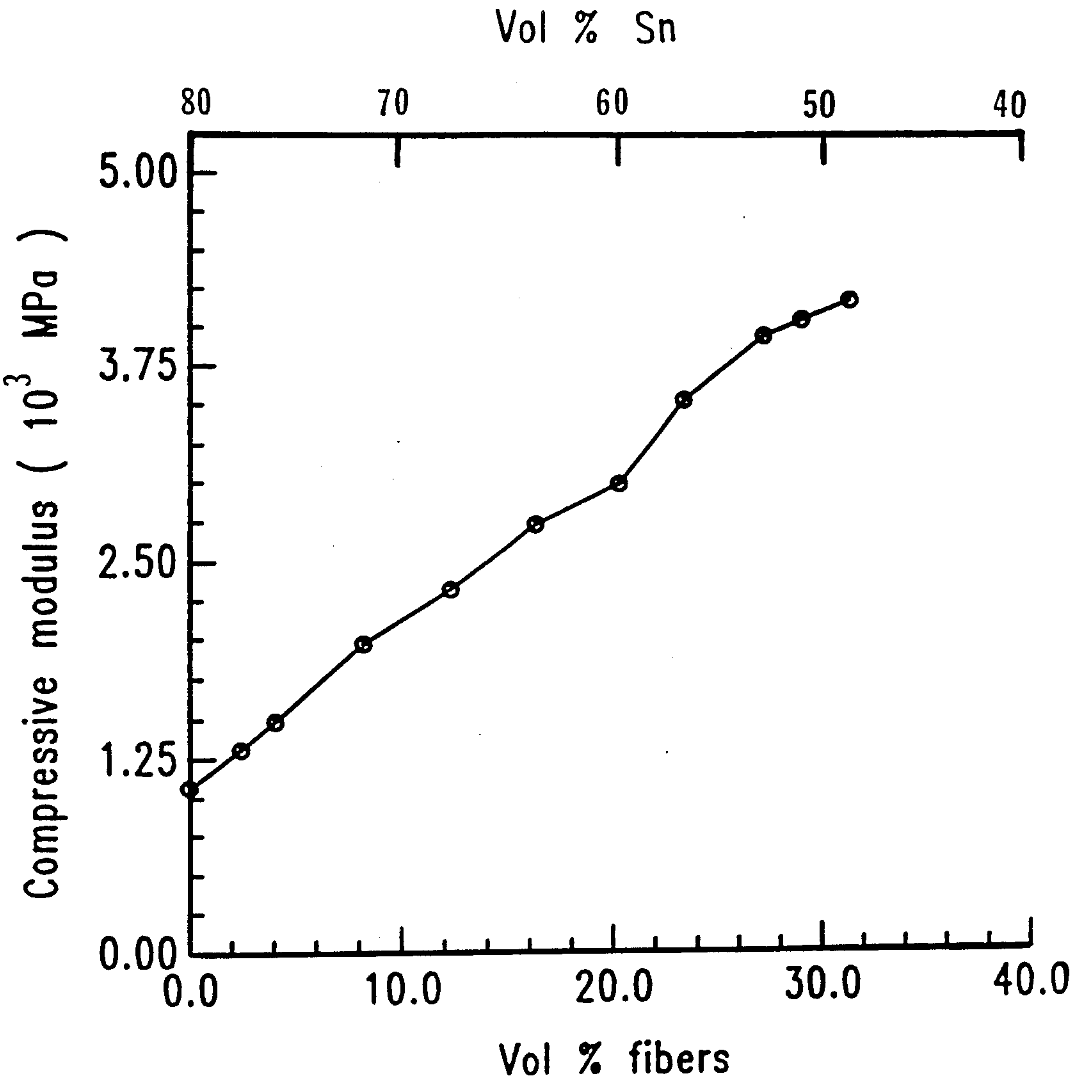
Vol % Sn
80
70
60
50
40
5.00
3.75
2.50
1.25
0.00
Compressive modulus (10^3 MPa)
0.0
10.0
20.0
30.0
40.0
Vol % fibers

Fig. 8.

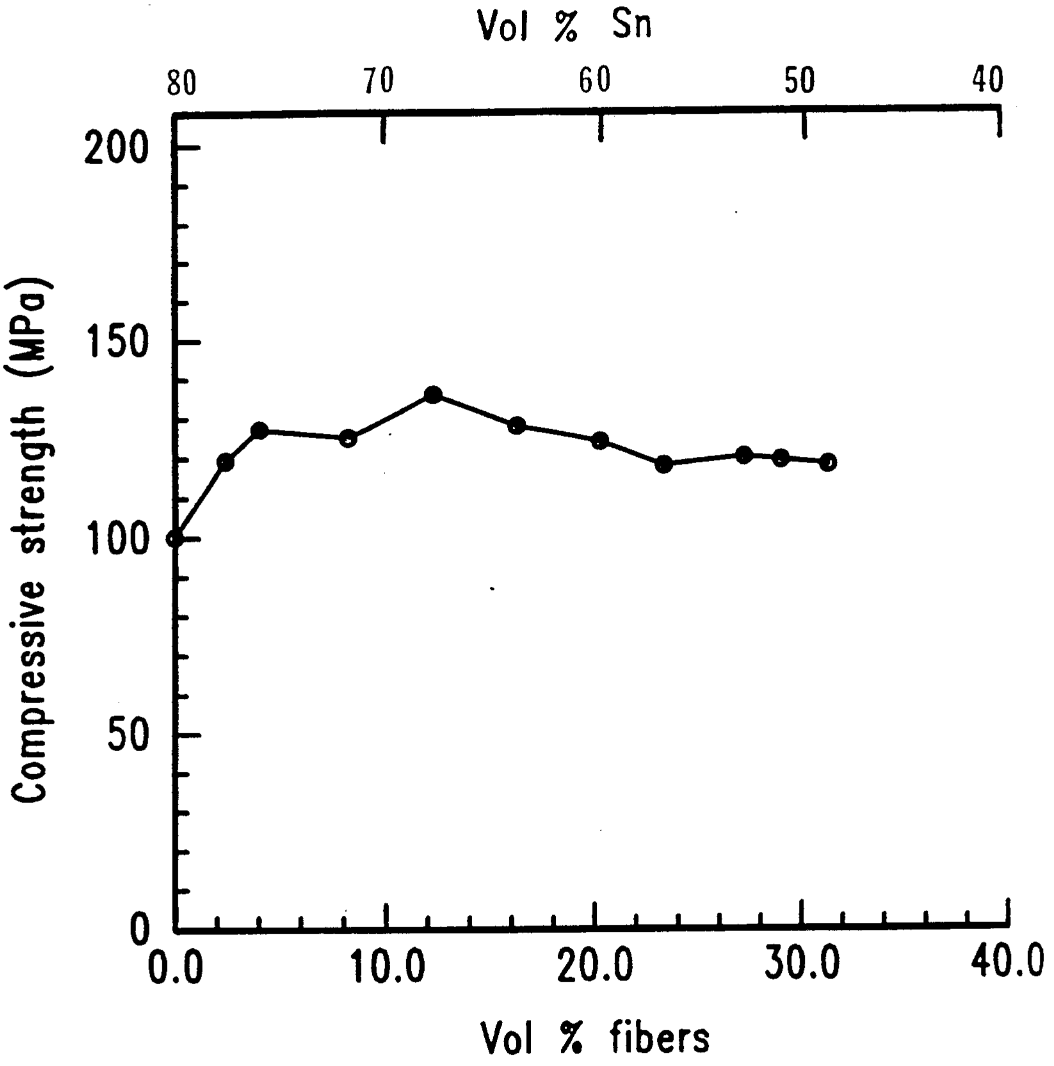
Vol % Sn
80
70
60
50
40
200
150
100
50
0
Compressive strength (MPa)
0.0
10.0
20.0
30.0
40.0
Vol % fibers

Fig. 9.

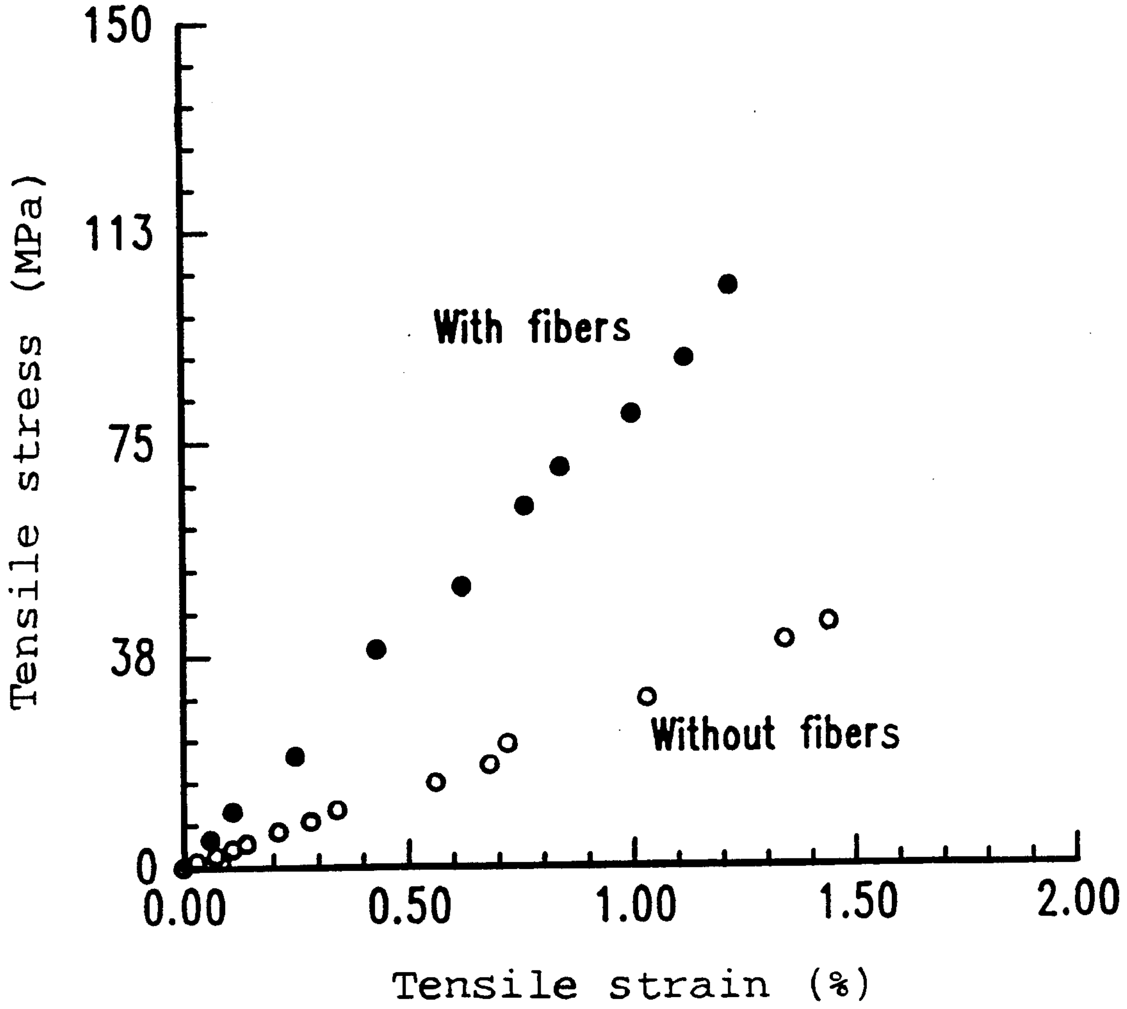
Tensile stress (MPa)
150
113
75
38
0
With fibers
Without fibers
0.00
0.50
1.00
1.50
2.00
Tensile strain (%)

Fig. 10.

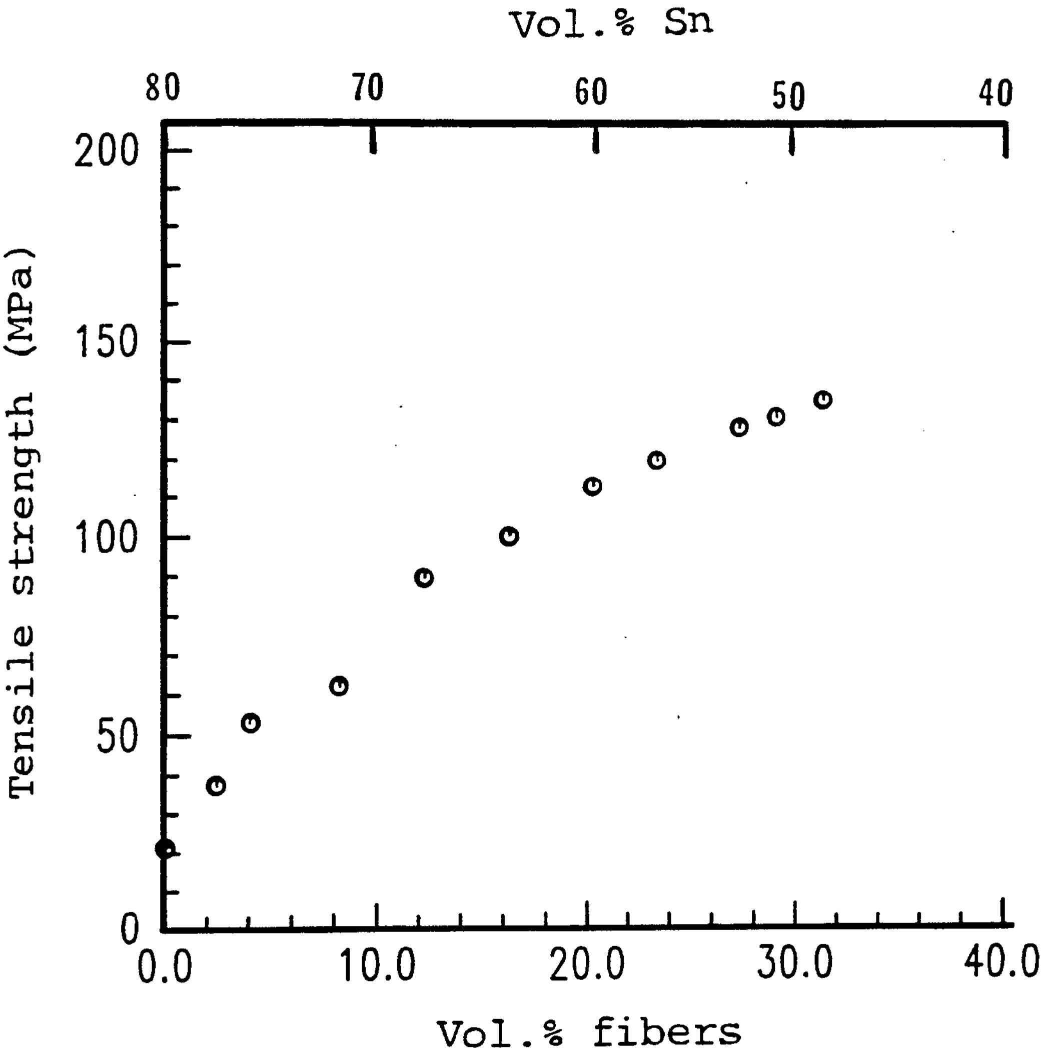
Vol.% Sn
80
70
60
50
40
200
150
100
50
0
Tensile strength (MPa)
0.0
10.0
20.0
30.0
40.0
Vol.% fibers

Fig. 11.

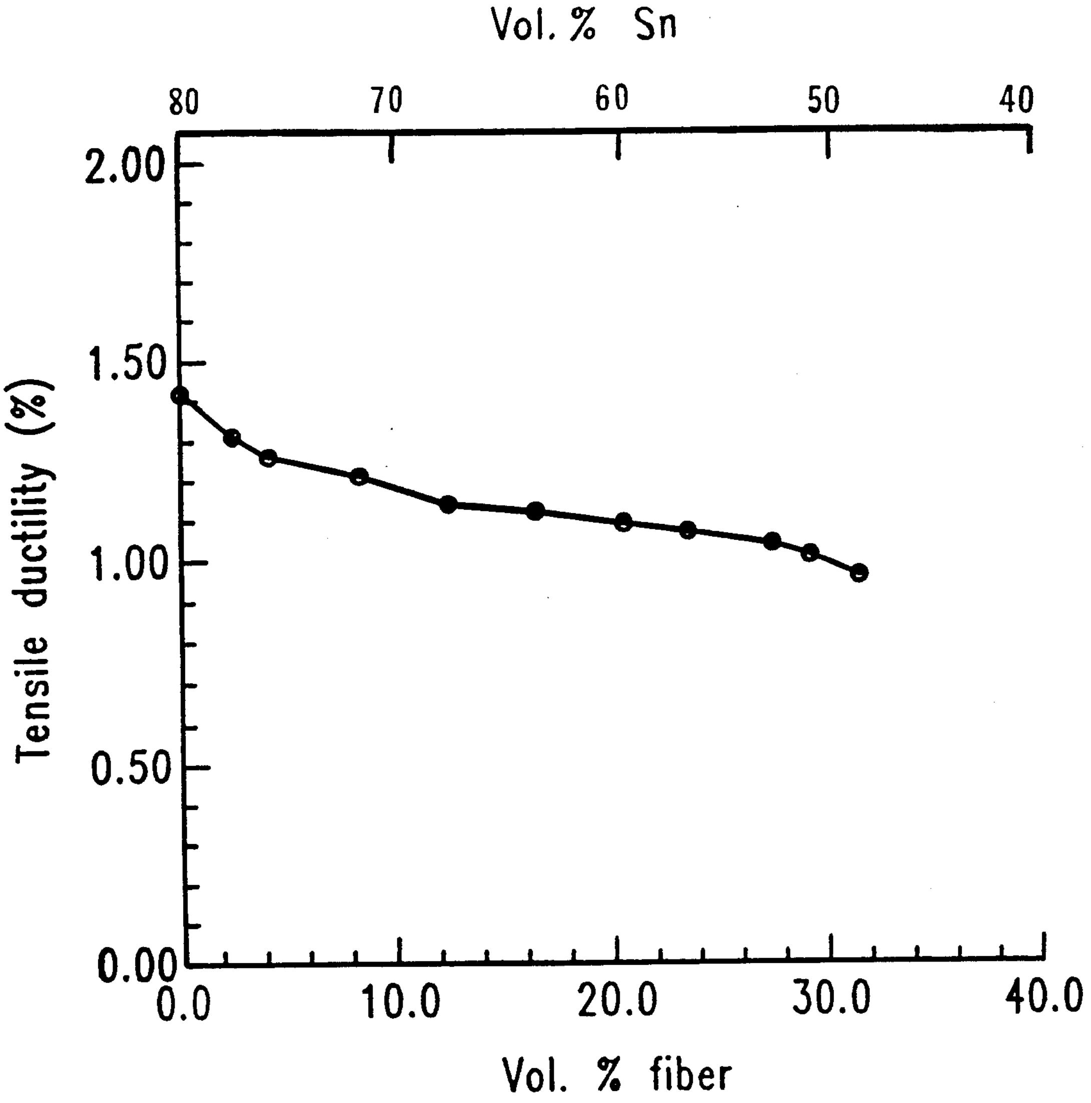
Vol. % Sn
80
70
60
50
40
2.00
1.50
1.00
0.50
0.00
Tensile ductility (%)
0.0
10.0
20.0
30.0
40.0
Vol. % fiber

Fig. 12.

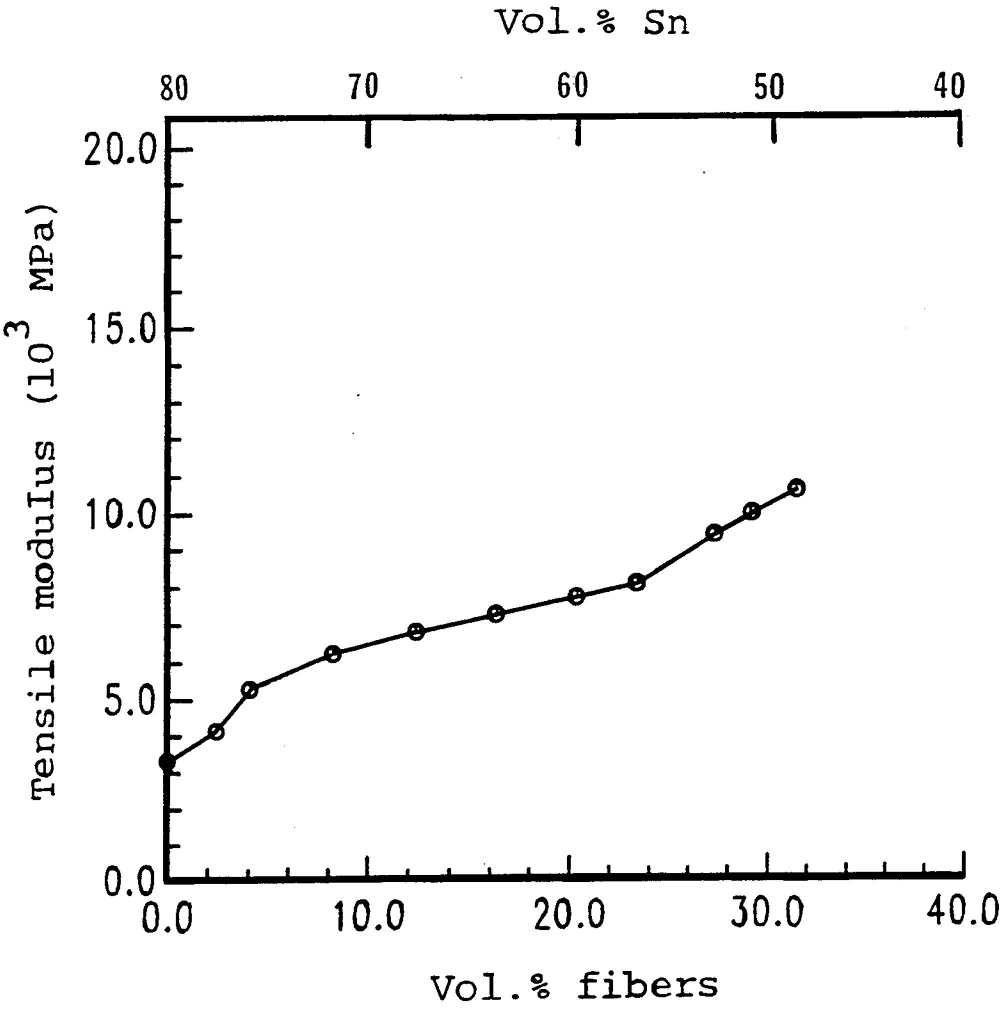
Vol.% Sn
80
70
60
50
40
20.0
15.0
10.0
5.0
0.0
Tensile modulus (10^3 MPa)
0.0
10.0
20.0
30.0
40.0
Vol.% fibers

Fig. 13.

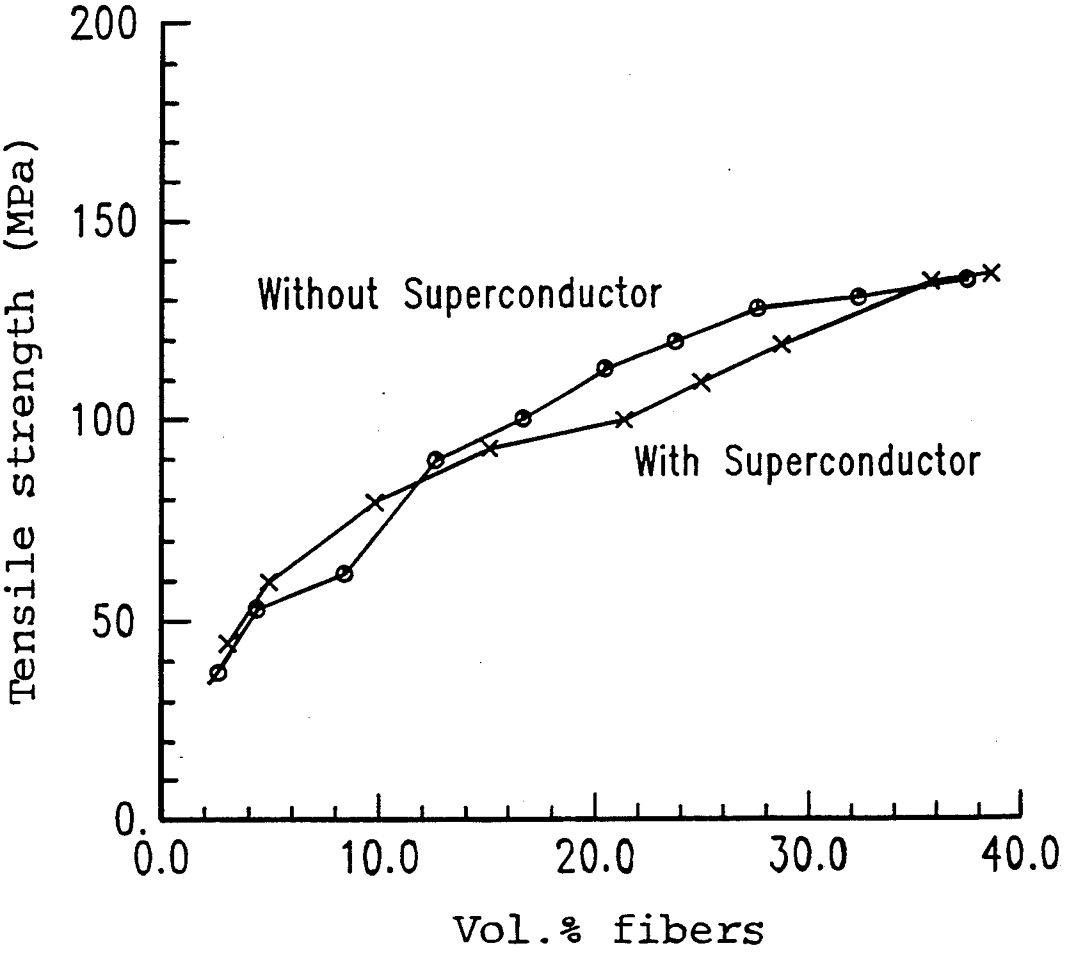
200
150
100
50
0
0.0
10.0
20.0
30.0
40.0
Tensile strength (MPa)
Vol.% fibers
Without Superconductor
With Superconductor

Fig. 14.

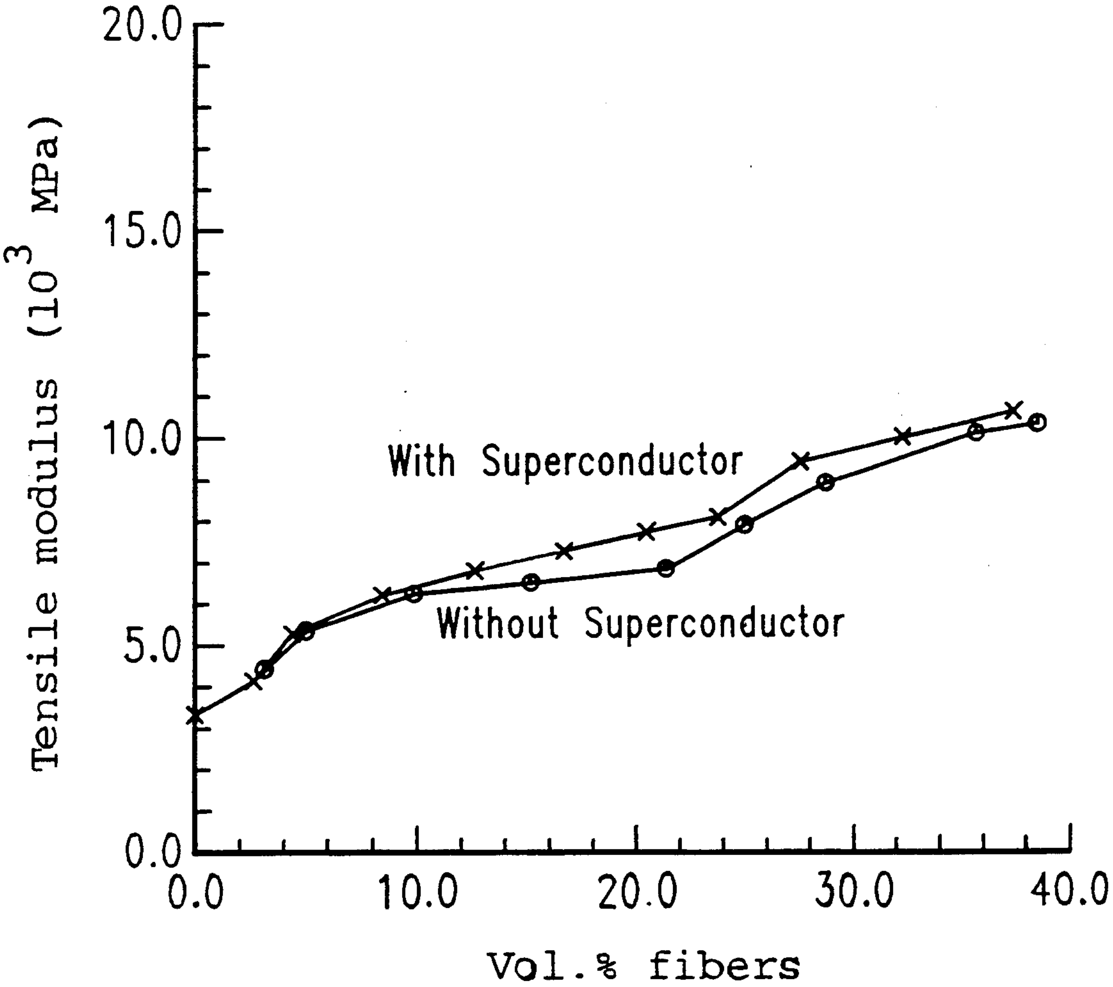
20.0
15.0
10.0
5.0
0.0
Tensile modulus (10^3 MPa)
With Superconductor
Without Superconductor
0.0
10.0
20.0
30.0
40.0
Vol.% fibers

Fig. 15.

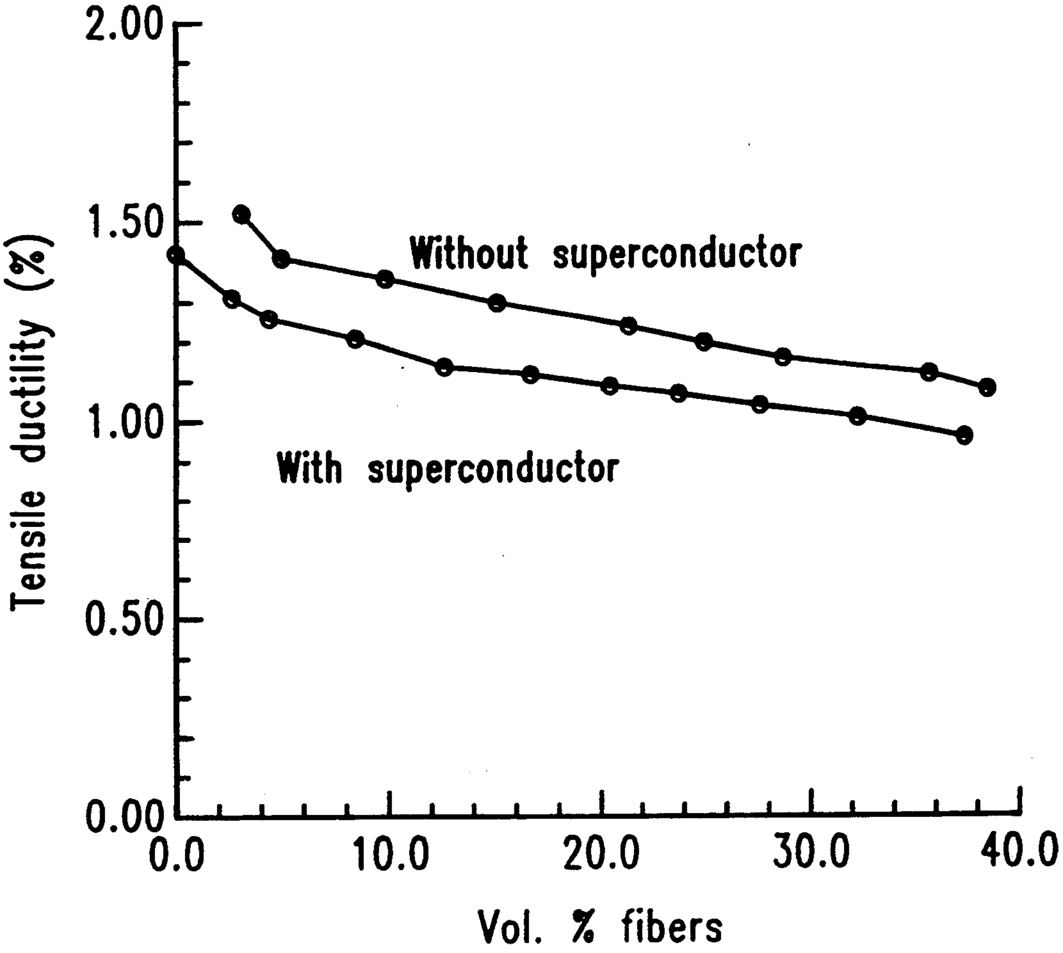
2.00
1.50
1.00
0.50
0.00
Tensile ductility (%)
Without superconductor
With superconductor
0.0
10.0
20.0
30.0
40.0
Vol. % fibers

Fig. 16.

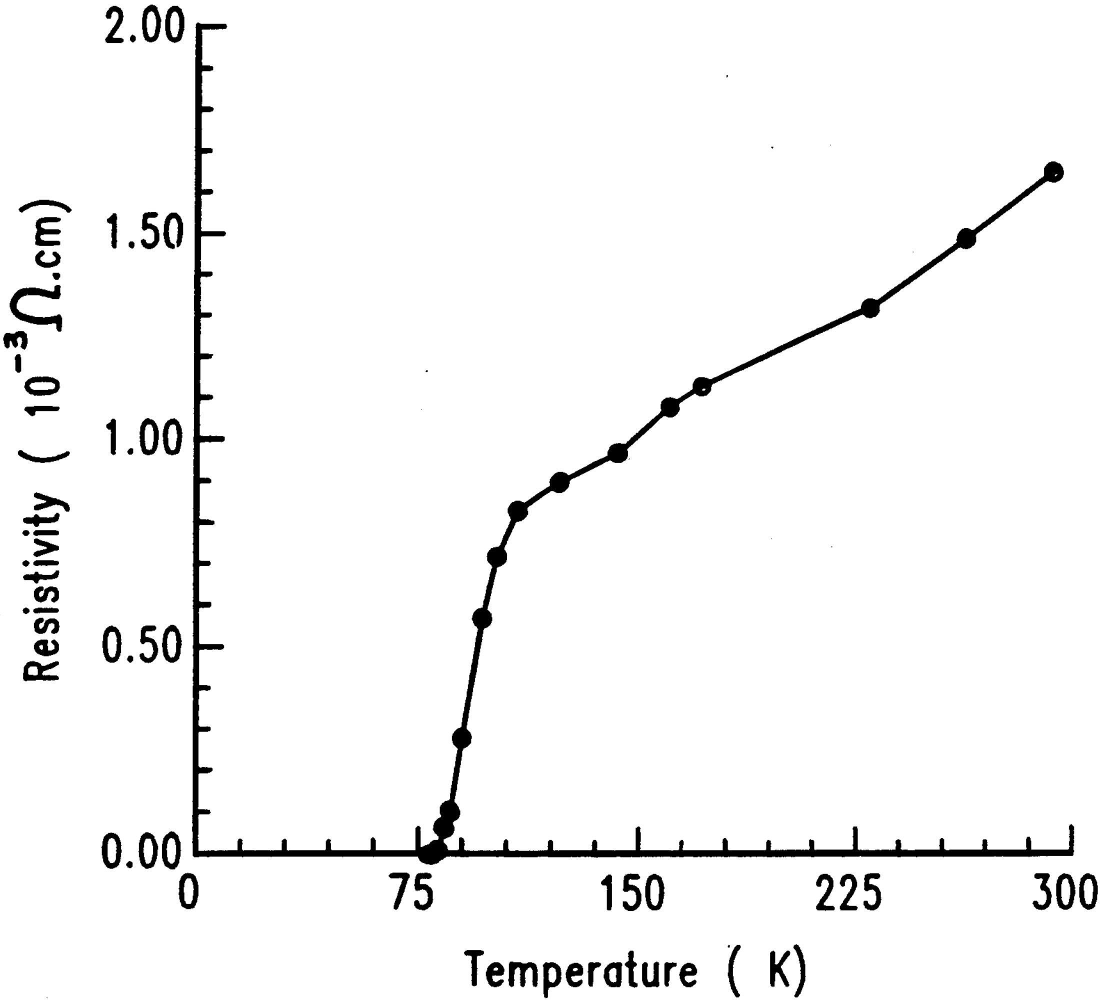
2.00
1.50
1.00
0.50
0.00
Resistivity ($10^{-3}\Omega$.cm)
0
75
150
225
300
Temperature (K)

SUPERCONDUCTOR-METAL LAMINATES AND METHOD OF MAKING

BACKGROUND OF THE INVENTION

The present invention relates to superconductor-metal laminates and to a method of making such laminates.

The high T_c (superconducting transition temperature) superconductors are generally brittle, hard to shape, and high in electrical resistivity at normal temperatures. By high T_c superconductors are meant superconductors with T_c above the temperature of liquid nitrogen (77° K.). Typically, metals are not high T_c superconductors and at normal temperatures are ductile, formable, low in electrical resistivity and high in thermal conductivity.

It has been recognized that the combination of a high T_c superconductor and a metal in the form of a composite material is attractive for the following reasons: the combination has improved toughness, ductility, shapeability and formability; improved electrical and thermal stabilization; and improved critical current density. The critical current density is the current density above which the superconductor loses its superconductivity, at a given temperature below T_c.

Powder metallurgy has been used to fabricate superconductor-metal composites. (In-Gann Chen, S. Sen and D. M. Stefanescu, Appl. Phys. Lett. 52 (16), 1355 (1988); F. H. Streitz, M. Z. Cieplak, Gang Xiao, A. Gavrin, A. Bakhshai and C. L. Chien, Appl. Phys. Lett. 52, 927 (1988); A. Goyal, P. D. Funkenbusch, G. C. S. Chang and S. J. Burns, Mater. Lett. 6 (8-9), 257 (1988)). Generally powder metallurgy involves mixing superconductor powder and metal powder, followed by sintering the mixture. There are some problems with this method. For example, the metal content is limited to 50 vol. % or below in order to have a continuous superconducting path in the composite. This limits the ductility of the composite. In addition, the choice of metal is limited to metals that are stable at the sintering temperature in oxygen and do not react with the superconductor at the sintering temperature (typically 950° C. for $YBa_2Cu_3O_{7-\delta}$).

Another method that has been used in forming a superconductor-metal composite involves (i) packing a superconductor powder in a metal tube, (ii) drawing the tube to a smaller diameter, and (iii) sintering. This method is commonly known as the metal tube drawing method. (R. W. McCallum, J. D. Verhoeven, M. A. Noack, E. D. Gibson, F. C. Laabs and D. K. Finnemore, Advanced Ceramic Materials 2 (3B), 388 (1987); S. Jin, R. C. Sherwood, R. B. Van Dover, T. H. Tiefel and D. W. Johnson, Jr., Appl. Phys. Lett. 51, 203 (1987)). Similar drawbacks as with the powder metallurgy method have been recognized with this method. For example, the choice of metal is limited to metals that are stable at the high sintering temperature required by the superconductor.

A further disadvantage of the superconductor-metal composites formed by the above described methods is that they are generally weak in tension.

It is therefore an object of this invention to provide a method for fabricating superconductor-metal materials which overcome the problems recognized in the art.

It is a further object to provide a method of producing a superconductor-metal material which overcomes the limitations of known methods while retaining the desirable characteristics.

A still further object is to produce a superconductor-metal material having both good tensile and compressive strength.

SUMMARY OF THE INVENTION

The invention is a superconducting laminate comprising, first and second layers; said first layer comprising a ceramic superconductor and said second layer comprising a metal; said first layer being diffusion bonded to said second layer.

The invention further comprises a superconducting laminate comprising first and second layers; said first layer comprising a ceramic superconductor and said second layer comprising a metal and carbon fibers; said first layer being bonded to said second layer.

The invention also comprises a method for preparing a superconducting laminate comprising the steps of, a) preparing a superconducting layer; b) preparing a metal layer; and c) diffusion bonding said superconducting layer with said metal layer.

The invention includes a method for preparing a superconducting laminate comprising the steps of, a) preparing a superconducting layer; b) preparing a metal-carbon fiber layer; and c) bonding said superconducting layer with said metal-carbon fiber layer.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 Is an SEM (Scanning Electron Microscope) photograph of a portion of a polished section of an MMC-$YBa_2Cu_3O_{7-\delta}$-MMC (MMC is the metal-matrix carbon fiber composite) three-layer laminate containing 32.1 vol. % fiber and 50.2 vol. % Sn (Tin). The left half is the MMC, the right half is the superconductor. The small dumbbell shaped spots in the MMC are the tips of carbon fibers.

FIG. 2 Shows the dependence of the electrical resistivity on temperature for the plain superconductor (solid circles), for a laminate containing 80.1% vol. % Sn and no fibers (open circles), and for a laminate containing 32.1 vol. % fibers and 50.2 vol. % Sn (crosses).

FIG. 3 Shows the compressive stress-strain curves (up to fracture) for a laminate without fibers (24 vol. % Sn) (open circles) and a laminate with fibers (51.1 vol. % Sn and 26.2 vol. % fibers) (solid circles).

FIG. 4 Illustrates the effect of tin content on the compressive ductility of laminate without carbon fibers; error bar of ±7%.

FIG. 5 Shows the effects of tin and fiber contents on the compressive ductility; error bar of ±5%.

FIG. 6 Is an SEM photograph of the compressive fracture surface of a laminate containing 19.5 vol. % Sn and no fibers. The right one-third of the photograph is tin, the left two-thirds of the photograph is the superconductor.

FIG. 7 Shows the effects of tin and fiber contents on the compressive elastic modulus; error bar of ±8%.

FIG. 8 Shows the effects of tin and fiber contents on the compressive strength; error bar of ±9%.

FIG. 9 Shows the tensile stress-strain curves (up to fracture) for a laminate without fibers (24 vol. % Sn) (open circles) and a laminate with fibers (51.4 vol. % Sn and 27.2 vol. % fibers) (solid circles).

FIG. 10 Illustrates the effects of tin and fiber contents on the tensile strength; error bar of ±8%.

FIG. 11 Illustrates the effects of tin and fiber contents on the tensile ductility; error bar of ±5%.

FIG. **12** Shows the effects of tin and fiber contents on the tensile modulus; error bar of ±7%.

FIG. **13** Shows the dependence of the tensile strength on the carbon fiber content in vol. % fibers, (the volume of the superconductor is not counted in calculating the volume % fibers). Circles: without superconductor (MMC alone), error bar of ±9%; crosses: with superconductor, error bar of ±9%.

FIG. **14** Depicts the dependence of the tensile modulus on the carbon fiber content in vol. % fibers, (the volume of the superconductor is not counted in calculating the volume % fibers). Circles: without superconductor (MMC alone), error bar of ±8%; crosses: with superconductor, error bar of ±7%.

FIG. **15** Shows the dependence of the tensile ductility on the carbon fiber content in vol. % fibers, (the volume of the superconductor is not counted in calculating the volume % fibers). Circles: without superconductor (MMC alone), error bar of ±6%; crosses: with superconductor, error bar of ±5%.

FIG. **16** Shows the dependence of the electrical resistivity on temperature for a laminate, containing 50.2 vol. % Sn and 31.0 vol. % fibers, that has been subjected to tension up to a stress of 90.6 MPa and a strain of 0.91%.

DETAILED DESCRIPTION OF THE INVENTION

By superconducting laminate is meant an article having at least 2 layers, at least one of which comprises a high T_c superconductor and at least one of which comprises a metal.

A metal as used herein means any metal which will be able to diffuse into the surface of the superconducting layer at a temperature lower than the degradation temperature of the superconducting material, and includes alloys. Examples of suitable metals include, but are not limited to tin, indium, and tin-lead alloy. A preferred metal is tin (Sn). The metal may be in a form suitable for forming the laminates of the present invention and includes foils, ingots, sheets, and powders. The thickness of the metal layer may vary in accordance with the invention, preferably, the thickness of the metal where no carbon fibers are incorporated is between about 0.1 mm to about 6 mm. The thickness of the metal layer where carbon fibers are incorporated is between about 0.4 mm to about 6 mm.

In accordance with this invention, the laminates may comprise between about 1 to about 99 vol. % metal (total volume being the volume of metal and the volume of superconductor). Preferred laminates comprise between about 40 to about 80 vol. % metal. Most preferred laminates comprise about 80 vol. % metal. Although the preferred and most preferred laminates are discussed above, the actual vol. % of the metal utilized will also depend on the thickness of the superconductor layer.

Superconductor as used herein is any ceramic superconductor and may include superconductor composites. Examples of superconductors are $YBa_2Cu_3O_{7-\delta}$, $Bi_2Sr_{3-x}Ca_xCu_2O_{8+y}$, and $Tl_2Ba_2CaCu_2O_\delta$. A preferred superconductor is $YBa_2Cu_3O_{7-\delta}$. The superconductor can be in bulk, wire or film form as prepared by any known technique. For example, for the superconductor $YBa_2Cu_3O_{7-\delta}$, the $YBa_2Cu_3O_{7-\delta}$ powder (commercially available from a variety of sources known to those skilled in the art) was pressed at a pressure of 137 MPa (mega pascal) and then sintered at 950° C. for 12 h, followed by annealing at 420° C. in flowing oxygen for 1 h. Superconductor composites include composites which comprise superconductor and metal and includes those which may be formed by the powder metallurgy method, the metal tube drawing method or other known methods. Generally, laminates having a superconducting layer in the order of at least 1000 angstroms in thickness are preferred.

The laminates of the present invention have a continuous superconducting path and the superconducting layer retains it's high T_c superconducting property after incorporation into the laminate.

While various lamination methods may be used, the laminates are preferably formed by diffusion bonding, e.g. by hot pressing the superconducting material and the metal, in air at a temperature lower than the melting point of the metal (which melting point, in accordance with this invention, is lower than the degradation temperature of the superconductor). Diffusion bonding as used herein means the solid state bonding of two materials by using heat and pressure. The degradation temperature of the superconducting material means the temperature above which the material no longer is a high T_c superconductor. For example, for the superconductor $YBa_2Cu_3O_{7-\delta}$, the degradation temperature is about 300° C., and hot pressing would be conducted at a temperature lower than 300° C.

Hot pressing is for a sufficient time and at a sufficient pressure for diffusion bonding to occur. For a given metal, the temperature, pressure and time may be adjusted so that various combinations of temperature, pressure and time may be utilized to achieve the diffusion bond. As is well known in the art, typically, if a higher temperature is used, a lower pressure and shorter time may be sufficient. The pressure utilized, however, is limited by the compressive strength of the superconductor.

In another embodiment of the invention, a laminate having three layers may be prepared. The superconductor may have a metal layer on either side of the superconductor to package the superconductor. This laminate may be prepared as described above by hot pressing a layer of superconductor sandwiched by two layers of metal, one on either side of the superconductor at a temperature below the melting point of the metal at a sufficient pressure and for a sufficient time in order to achieve diffusion bonding.

It is within the scope of this embodiment that the two metal layers on either side of the superconductor layer may comprise the same or different metals. More particularly, both layers may be tin or one metal layer may be tin and the other metal layer may be a tin-lead alloy or indium or any other combination, which would meet the above discussed parameters.

It is also within the scope of this invention to have a plurality of ceramic superconductor layers alternating with metal layers.

In another embodiment of the invention, a reinforced laminate having greater tensile strength was made by the addition of carbon fibers to the metal layer. The carbon fiber reinforced metal provides an effective way of packaging ceramic superconductors so that the package exhibits good mechanical, electrical and thermal properties. The carbon fibers may include those made from a variety of precursors such as textile fibers, pitch or other precursors. An example of a textile fiber is PAN (polyacrylonitrile).

The fibers may be in a continuous or short form. If the continuous form is used, they may be unidirectional, multidirectional or in the form of a woven fabric. In a preferred embodiment, continuous, unidirectional carbon fibers may be utilized. Between about 0 vol. % to about 50 vol. % of carbon fiber (the volume being the volume of metal and carbon fiber) may be incorporated into the metal layer. It is preferred that about 15–40 vol. % of carbon fiber is incorporated into the metal layer.

The carbon fiber reinforced metal-superconductor laminate may be prepared by using a two-step process. The first step is the preparation of a metal-matrix carbon fiber composite (abbreviated MMC) by laying up carbon fibers and metal in the form of alternate layers and consolidating by hot pressing (by using a hydraulic press or other known methods) at a temperature, pressure and time which will allow the metal to penetrate or diffuse into void spaces between the carbon fibers. The second step involves hot pressing a layer of superconductor sandwiched by two layers of MMC at a sufficient temperature, preferably below below the melting point of the metal, and at a sufficient pressure for a sufficient time in order to achieve bonding between the MMC and the superconductor. Bonding means the liquid or solid state joining of two or more materials. Methods of bonding, as used herein, includes adhesive bonding, soldering, brazing, welding and diffusion bonding.

It is to be understood, particularly in the case of carbon fiber containing materials, that lamination or bonding methods other than diffusion bonding may be used. Such methods, however, may have serious disadvantages related to physical limitations of certain materials which may allow degradation or destruction of such materials, as a result of the conditions encountered during lamination.

EXAMPLES

EXAMPLE 1

To make a superconductor-tin laminate (without carbon fibers), in accordance with this invention, the superconductor $YBa_2Cu_3O_{7-\delta}$ was prepared as described above, tin foil (60 vol. %, 4.12 g, 3.36 mm thick, includes both tin layers) was diffusion bonded to both sides of the $YBa_2Cu_3O_{7-\delta}$ superconductor layer (40 vol. %, 1.54 g, 2.24 mm thick) by hot pressing in air at 140° C. and 5.3 MPa for 15 minutes. Table 1 below lists other percentages of tin which were used and the various thicknesses of the tin (includes both layers) and the superconductor.

TABLE 1

Vol. % Sn	Thickness (mm) Superconductor	Thickness (mm) Tin
81.6	0.460	2.032
77.8	0.580	2.032
73.4	0.741	2.032
64.5	1.126	2.032
50.1	2.036	2.032
36.0	1.803	1.016
33.3	2.032	1.016
29.4	2.448	1.016
26.5	2.821	1.016
25.0	3.058	1.016
24.0	3.213	1.016
22.2	3.556	1.016
20.0	4.063	1.016
19.5	1.676	0.406
15.6	1.372	0.254
10.7	2.381	0.254
6.1	3.897	0.254

EXAMPLE 2

To produce a carbon fiber reinforced tin-superconductor laminate in accordance with this invention, the previously described two-step process may be utilized.

In the first step, a tin-matrix unidirectional carbon fiber composite (abbreviated MMC), was prepared by laying up carbon fibers (15 vol. %, 0.285 g; continuous, unidirectional, and PAN-based, which are commercially available) and tin foils (50 vol. %, 3.75 g) in the form of alternate layers and consolidating by hot pressing (by using a hydraulic press) at 244° C. and at a pressure of 64 MPa for 20 min. The thickness of the MMC layer was 4.06 mm. In the second step a layer of superconductor $YBa_2Cu_3O_{7-\delta}$ (35 vol. %. 1.21 g, 2.18 mm thick) sandwiched by two layers of MMC was hot pressed at 180° C. and 5.1 MPa for 15 min in order to achieve diffusion bonding. Note that 170° C. is below the melting point of tin (232° C.).

TEST RESULTS OF THE LAMINATES

In the resulting laminate, tin served as the adhesive and to increase the ductility, the normal-state electrical conductivity and the thermal conductivity. Carbon fibers served to increase the strength and the modulus, both in tension along the fiber direction and in compression perpendicular to the fiber layers, and also served to increase the thermal conductivity and the thermal fatigue resistance, as shown in table 2 below. Volume % as used in Table 2 includes the volume of carbon, metal and superconductor.

TABLE 2

Thermal fatigue due to cycling between room temperature and liquid nitrogen temperature

Carbon fiber Content (vol. %)	Sn content (vol. %)	No. of cycles for delamination to start*
0	43.2	63
3.0	39.4	86
8.3	40.1	102
12.0	43.1	116
15.3	48.3	123
20.1	50.2	127

*Each number is the average of three specimens.

Laminates of various fiber contents were subjected to thermal cycling between room temperature and liquid nitrogen temperature (77° K.) by immersion in liquid nitrogen for 20 min, followed by room temperature equilibration for at least 30 min, and repeating. After every cycle, each specimen was observed under an optical microscope to look for delamination (slight cracking) between the superconductor and the MMC. Table 2 shows the number of cycles for the start of delamination for each laminate composition. The higher the carbon fiber content, the greater the number of cycles required for the start of delamination.

The fabrication of the laminates, according to this invention involves relatively low temperatures. The simplicity of this process makes it possible for an operation to be set up for fabricating continuous superconducting cables which are both shielded and toughened by metal and strengthened by carbon fibers.

The toughness was measured by carrying out the Izod Test using a Tinius Olsen plastic impact tester. The toughness was 3.25 in.lb for a laminate containing 18.8 vol. % Sn (no fibers) and was 5.50 in.lb for a laminate containing 25.5 vol. % Sn (no fibers). Hence, the tin greatly enhanced the toughness.

As can be seen from the SEM photograph shown in FIG. **1**, no void or crack was observed between the MMC and the superconductor or between the tin and the carbon fibers in the MMC.

Electrical resistivity of the laminates can be seen in FIG. **2**. The electrical resistivity was measured with the four-probe technique by using a Keithley 181 nanovolt meter and a Keithley 224 programmable current source, such that the current was around 10^{-3} amperes. A thermocouple was placed so that it almost touched the sample. In the case of laminates containing fibers, the electrical resistivity was measured in the direction of the fibers however, the electrical resistivity in the direction perpendicular to the fiber axis was comparable to that in the direction parallel to the fibers. The drop to zero resistivity is slightly less sharp for either laminate than the plain superconductor. Carbon fibers have a lower electrical resistivity than the plain superconductor above T_c, but its value is higher than that of tin. Therefore, the normal-state electrical resistivity of the laminate with tin and fibers is lower than that of the plain superconductor and higher than that of the laminate with tin and no fiber.

Mechanical testing was performed using a hydraulic Materials Testing Systems (MTS). The strain in tensile or compressive testing was measured by the displacement of the crosshead. The gage length was 34.7 mm for tensile testing and 6 mm for compressive testing. Compressive testing was performed with the force perpendicular to the laminate layers. Tensile testing was performed with the force parallel to the fibers in the plane of the laminate. At least three samples were run and the data were averaged for each composition in each type of test.

FIG. **3** shows the compressive stress-strain curves (up to fracture) for a laminate containing tin (24 vol. % Sn) but no fibers (open circles) and a laminate containing 51.1 vol. % Sn and 26.2 vol. % fibers (solid circles).

FIG. **4** shows that tin greatly increases the ductility of the laminate. FIG. **5** shows that tin greatly increased the ductility of the laminate, but the fibers decreased the ductility.

As can be seen in FIG. **6**, cracks are observed to extend from the tin superconductor interface inward into the superconductor. A similar fracture is observed in a laminate with both fibers and tin. In contrast, the plain superconductor (without tin or fiber) shatters and disintegrates into particles upon fracture. Hence, the fracture behavior is dramatically different between either laminate and the plain superconductor.

The decrease of the modulus with increasing tin content and decreasing fiber content is significant and can be seen in FIG. **7**. No systematic trend can be seen in FIG. **8** in the compressive strength, because the strength is very sensitive to small flaws that are bound to be present in the superconductor.

FIG. **9** shows the tensile stress-strain curves (up to fracture) for a laminate containing tin (24 vol. % Sn) but no fibers (open circles) and a laminate containing 51.4 vol. % Sn and 27.2 vol. % fibers (solid circles).

FIGS. **10**, 11 and 12 show the effects of tin and fiber contents on the tensile test results. The tensile test was performed along the fiber direction. The addition of carbon fibers greatly improved the tensile strength (FIG. **10**), but decreased the ductility slightly (FIG. **11**). The tensile modulus was increased by increasing the fiber content. Debonding between the MMC and the superconductor and some fiber pull-out were observed from the fracture surface after the tensile test.

FIGS. **13**, **14** and **15** show the effects of the superconductor on the tensile test results. Data obtained with the presence of the superconductor (in the form MMC-superconductor-MMC) are shown by circles. Data obtained with the absence of the superconductor (i.e., MMC alone) are shown by crosses. The horizontal scale of FIGS. **13**, 14 and 15 show the vol. % fibers such that the volume of the superconductor was not counted in the volume of the laminate. (Note that this scale is different from the horizontal scale of FIGS. **4**, **5**, **7**, **8**, **10**, **11** and **12** in which the volume of the superconductor was counted in calculating the volume percent fibers.) The similarity of data with and without the superconductor in FIGS. **13** and **14** shows that the tensile load was almost totally sustained by the MMC. FIG. **15** shows that the tensile ductility is somewhat diminished by the presence of the superconductor.

The tensile test for the plain superconductor (without tin or fibers) did not give very reliable results because of the difficulty of gripping the superconductor, which was difficult in the absence of tin. Nevertheless, the tensile strength was about 5.2 MPa, the tensile modulus was roughly 1.3 GPa (Giga pascal), and the tensile ductility was roughly 0.5%.

A laminate containing 31.0 vol. % fibers and 50.2 vol. % tin was subjected to tension up to a tensile stress of 90.6 Mpa and a tensile strain of 0.91% (below the stress or strain required for fracture) and then the load was released (allowing the strain to return to essentially zero) and the electrical resistivity was measured as a function of temperature, as shown in FIG. **16**. It remained superconducting.

Comparison of FIG. **8** and **10** shows that, without fibers, the tensile strength (parallel to the fiber direction) was much less than the compressive strength (perpendicular to the fiber layers), but with about 24 vol. % fibers, the tensile strength was approximately equal to the compressive strength. With further increase of the fiber content, the tensile strength exceeded the compressive strength.

The tensile strength of the plain superconductor was roughly 5.2 MPa. The presence of tin (without fibers) increased the value to about 20 MPa (FIG. **10**). Further addition of carbon fibers significantly increased the tensile strength, up to 134 MPa for 31 vol. % fibers.

Carbon fibers increased the compressive strength, compressive modulus, tensile strength and tensile modulus, but they decreased the compressive ductility and tensile ductility. However, because tin was also present and tin is a soft metal, the compressive ductility for the case of 31 vol. % fibers was approximately equal to that for the plain superconductor (without tin or fiber). For carbon fiber contents less than 30 vol. % fibers, the compressive ductility exceeded that of the plain superconductor. In general, the carbon fibers decreased the tensile ductility (FIG. **11**) much less than the compressive ductility (FIG. **6**).

FIG. **16** and **11** show that the superconducting behavior of the laminates was maintained after tension almost to the point of tensile fracture.

The fabrication of the laminates involved low temperatures. The simplicity of this process makes it possible for an operation to be set up for fabricating continuous superconducting cables which are both shielded and toughened by tin and strengthened by carbon fibers. In contrast to powder metallurgy, the diffusion bonding method allows the metal to be the major phase while still maintaining a continuous superconducting path in the laminate. Furthermore, carbon fibers, with their nearly zero thermal expansion coefficient helps match the thermal expansion coefficients of the MMC layer and the superconductor layer. This matching is necessary in order to enhance the durability of the composite to thermal cycling (i.e. thermal fatigue). In addition, carbon fibers are excellent in thermal conductivity (both at ambient and cryogenic temperatures), wear resistant, corrosion resistant, and are low in electrical resistivity.

Other embodiments of the invention will be apparent to the skilled in the art from a consideration of this specification or practice of the invention disclosed herein. It should be understood that there may be other embodiments which fall within the spirit and scope of the invention as defined by the following claims.

What is claimed is:

1. A superconducting laminate comprising first and second layers; said first layer comprising a high T_c ceramic superconductor and said second layer comprising a metal; said first layer being diffusion bonded to said second layer, wherein said metal is selected from the group consisting of tin, indium and tin lead alloy.

2. The laminate as recited in claim **1** further comprising a third layer, said third layer being adjacent to and being diffusion bonded to a second side of said superconducting layer such that the first layer is adjacent to the first side of said superconducting layer.

3. A superconducting laminate comprising first and second layers; said first layer comprising a high T_c ceramic superconductor and said second layer comprising a metal and carbon fiber; said first layer being diffusion bonded to said second layer, wherein said metal is selected from the group consisting of tin, indium and tin lead alloy.

4. The laminate as recited in claim **1** wherein the superconductor is selected from the group consisting of $YBa_2Cu_3O_{7-\delta}$, $Bi_2Sr_{3-x}Ca_xCu_2O_{8+y}$, and $Tl_2Ba_2CaCu_2O_8$.

5. The laminate as recited in claim **2** wherein the superconductor is selected from the group consisting of $YBa_2Cu_3O_{7-\delta}$, $Bi_2Sr_{3-x}Ca_xCu_2O_{8+y}$, and $Tl_2Ba_2CaCu_2O_8$.

6. The laminate as recited in claim **4** wherein the superconductor is $YBa_2Cu_3O_{7-\delta}$.

7. The laminate as recited in claim **5** wherein the superconductor is $YBa_2Cu_3O_{7-\delta}$.

8. The laminate as recited in claim wherein the metal is selected from the group consisting of tin, indium and tin lead alloy.

9. The laminate as recited in claim **8** wherein the metal is tin.

10. The laminate as recited in claim **2** wherein the metal is selected from the group consisting of tin, indium and tin lead alloy.

11. The laminate as recited in claim **6** wherein the metal is selected from the group consisting of tin, indium and tin lead alloy.

12. The laminate as recited in claim **7** wherein the metal is selected from the group consisting of tin, indium and tin lead alloy.

13. The laminate as recited in claim **10** wherein the second and third layers comprise the same metal.

14. The laminate as recited in claim **10** wherein the second and third layers comprise different metals.

15. The laminate as recited in claim **13** wherein the metal is tin.

16. The laminate as recited in claim **14** wherein the metals are tin and tin lead alloy.

17. The laminate as recited in claim **1** wherein the second layer further comprises carbon fibers.

18. The laminate as recited in claim **2** wherein the second layer further comprises carbon fibers.

19. The laminate as recited in claim **6** wherein the second layer further comprises carbon fibers.

20. The laminate as recited in claim **7** wherein the carbon fibers are from textile fiber precursors.

21. The laminate as recited in claim **17** wherein the carbon fibers are from textile fiber precursors.

22. The laminate as recited in claim **18** wherein the carbon fibers are from textile fiber precursors.

23. The laminate as recited in claim **19** wherein the carbon fibers are from textile fiber precursors.

24. The laminate as recited in claim **21** wherein the textile fibers are PAN based.

25. The laminate as recited in claim **22** wherein the textile fibers are PAN based.

26. The laminate as recited in claim **1** wherein the metal will be able to diffuse into the surface of the superconducting layer at a temperature lower than the degradation temperature of the superconductor.

27. The laminate as recited in claim **2** wherein the metal will be able to diffuse into the surface of the superconducting layer at a temperature lower than the degradation temperature of the superconductor.

28. The laminate as recited in claim **1** wherein the superconductor is a superconducting composite.

29. The laminate as recited in claim **2** wherein the superconductor is a superconducting composite.

30. The laminate as recited in claim **28** wherein the superconducting composite comprises a metal.

31. The laminate as recited in claim **29** wherein the superconducting composite comprises a metal.

32. The laminate as recited in claim **30** wherein said superconducting composite is fabricated by powder metallurgy.

33. The laminate as recited in claim **31** wherein said superconducting composite is fabricated by metal tube drawing.

34. The laminate as recited in claim **1** further comprising a plurality of alternating superconducting and metal containing layers.

35. The laminate as recited in claim **1** wherein the first layer comprises between about 20 to about 60 vol. % of the laminate.

36. The laminate as recited in claim **2** wherein the first layer comprises between about 20 to about 60 vol. % of the laminate.

37. The laminate as recited in claim **1** wherein the second layer comprises between about 40 to about 80 vol. % of the laminate.

38. The laminate as recited in claim **17** wherein the carbon fibers comprise between about 15–40 vol. % of said second layer.

39. The laminate as recited in claim **1** wherein the second layer is between about 0.1 mm to about 6 mm in thickness.

40. The laminate as recited in claim **35** wherein the second layer is between about 0.1 mm to about 6 mm in thickness.

41. The laminate as recited in claim **36** wherein the second layer is between about 0.1 mm to about 6 mm in thickness.

42. The laminate as recited in claim **38** wherein the second layer is between about 0.4 mm to about 6 mm in thickness.

43. The laminate as recited in claim **1** wherein the first layer is at least 1000 Angstroms in thickness.

44. The laminate as recited in claim **2** wherein the first layer is at least 1000 Angstroms in thickness.

45. The laminate as recited in claim **40** wherein the first layer is at least 1000 Angstroms in thickness.

46. The laminate as recited in claim **42** wherein the first layer is at least 1000 Angstroms in thickness.

47. A superconducting laminate comprising, first and second layers; said first layer comprising about 40 vol. % $YBa_2Cu_3O_{7-\delta}$, and being about 3.4 mm thick; said second layer comprising about 60 vol. % of tin and being about 2.3 mm thick; said first layer being diffusion bonded to said second layer at 140° C. and 5.3 MPa for about 15 minutes.

48. A superconducting laminate prepared by a process comprising the steps of:

a) preparing a high T_c superconducting layer;

b) preparing a metal layer; and

c) diffusion bonding said superconducting layer to said metal layer;

wherein said metal is selected from the group consisting essentially of tin, indium and tin lead alloy.

* * * * *

US005159703A

United States Patent [19]

Lowery

[11] **Patent Number:** **5,159,703**

[45] **Date of Patent:** **Oct. 27, 1992**

[54] **SILENT SUBLIMINAL PRESENTATION SYSTEM**

[76] Inventor: **Oliver M. Lowery**, 5188 Falconwood Ct., Norcross, Ga. 30071

[21] Appl. No.: **458,339**

[22] Filed: **Dec. 28, 1989**

[51] **Int. Cl.⁵** **H04B 7/00;** H04R 25/00; H04R 3/02

[52] **U.S. Cl.** **455/42;** 455/46; 455/66; 381/73.1; 128/420.5

[58] **Field of Search** 455/46, 47, 66, 109, 455/110, 42–43; 381/73.1, 105, 124; 358/141-143; 600/28; 128/420.5; 380/38

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,060,795	10/1962	Corrigan et al.	352/131
3,278,676	10/1966	Becker	358/142
3,393,279	7/1968	Flanagan	128/420.5
3,712,292	1/1973	Zentmeyer, Jr.	600/28
4,141,344	2/1979	Barbara	600/28
4,395,600	7/1983	Lundy et al.	381/73.1
4,463,392	7/1984	Fischer et al.	360/30
4,777,529	10/1988	Schultz et al.	381/73.1
4,834,701	5/1989	Masaki	600/28
4,877,027	10/1989	Brunkan	128/420.5

Primary Examiner—Reinhard J. Eisenzopf
Assistant Examiner—Andrew Faile

[57] **ABSTRACT**

A silent communications system in which nonaural carriers, in the very low or very high audio frequency range or in the adjacent ultrasonic frequency spectrum, are amplitude or frequency modulated with the desired intelligence and propagated acoustically or vibrationally, for inducement into the brain, typically through the use of loudspeakers, earphones or piezoelectric transducers. The modulated carriers may be transmitted directly in real time or may be conveniently recorded and stored on mechanical, magnetic or optical media for delayed or repeated transmission to the listener.

3 Claims, 3 Drawing Sheets

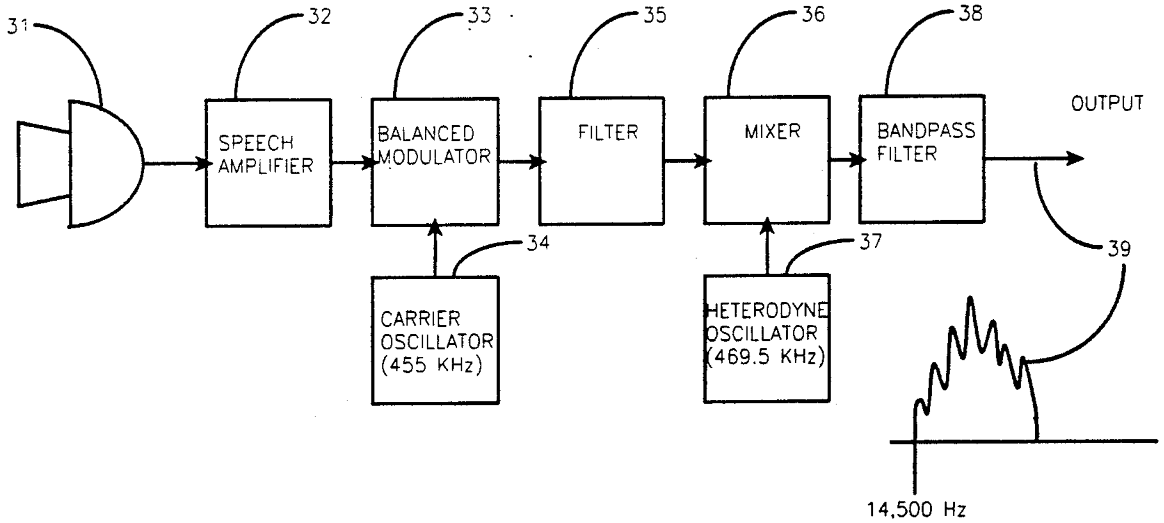
31
32
33
35
36
38
SPEECH AMPLIFIER
BALANCED MODULATOR
FILTER
MIXER
BANDPASS FILTER
OUTPUT
34
37
39
CARRIER OSCILLATOR (455 KHz)
HETERODYNE OSCILLATOR (469.5 KHz)
14,500 Hz

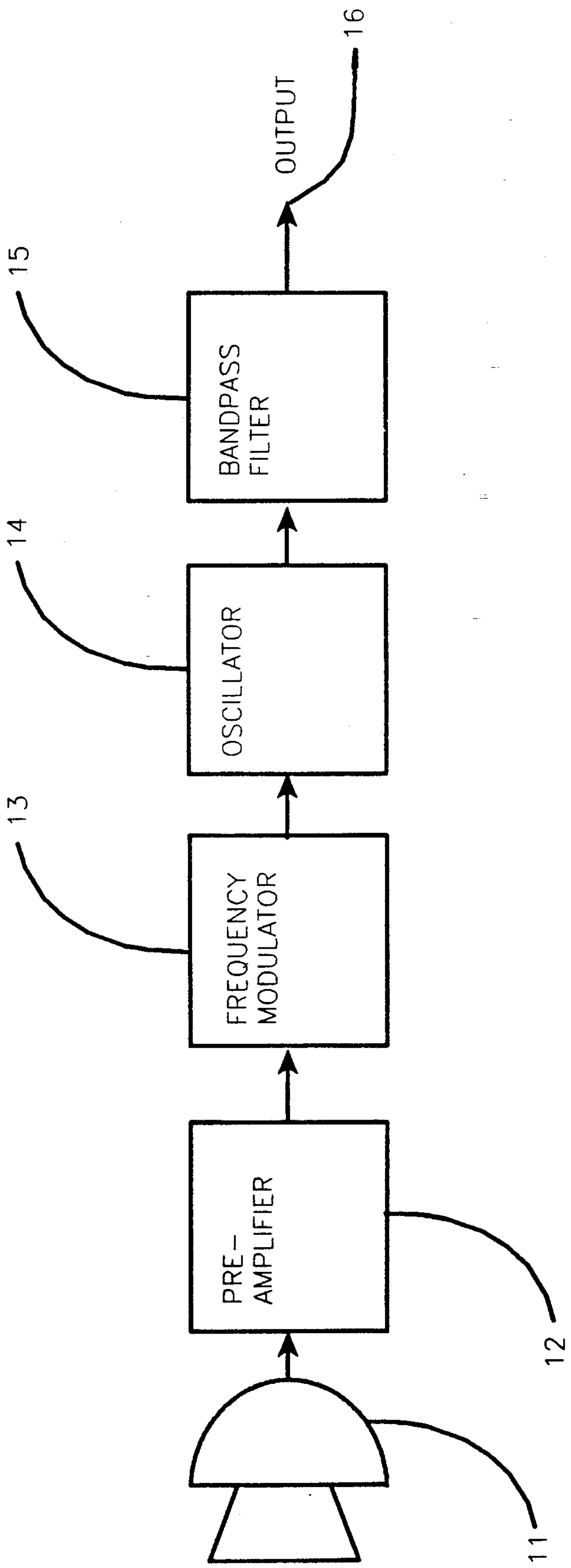
16
OUTPUT
15
BANDPASS FILTER
14
OSCILLATOR
13
FREQUENCY MODULATOR
PRE-AMPLIFIER
12
11

FIG. 1

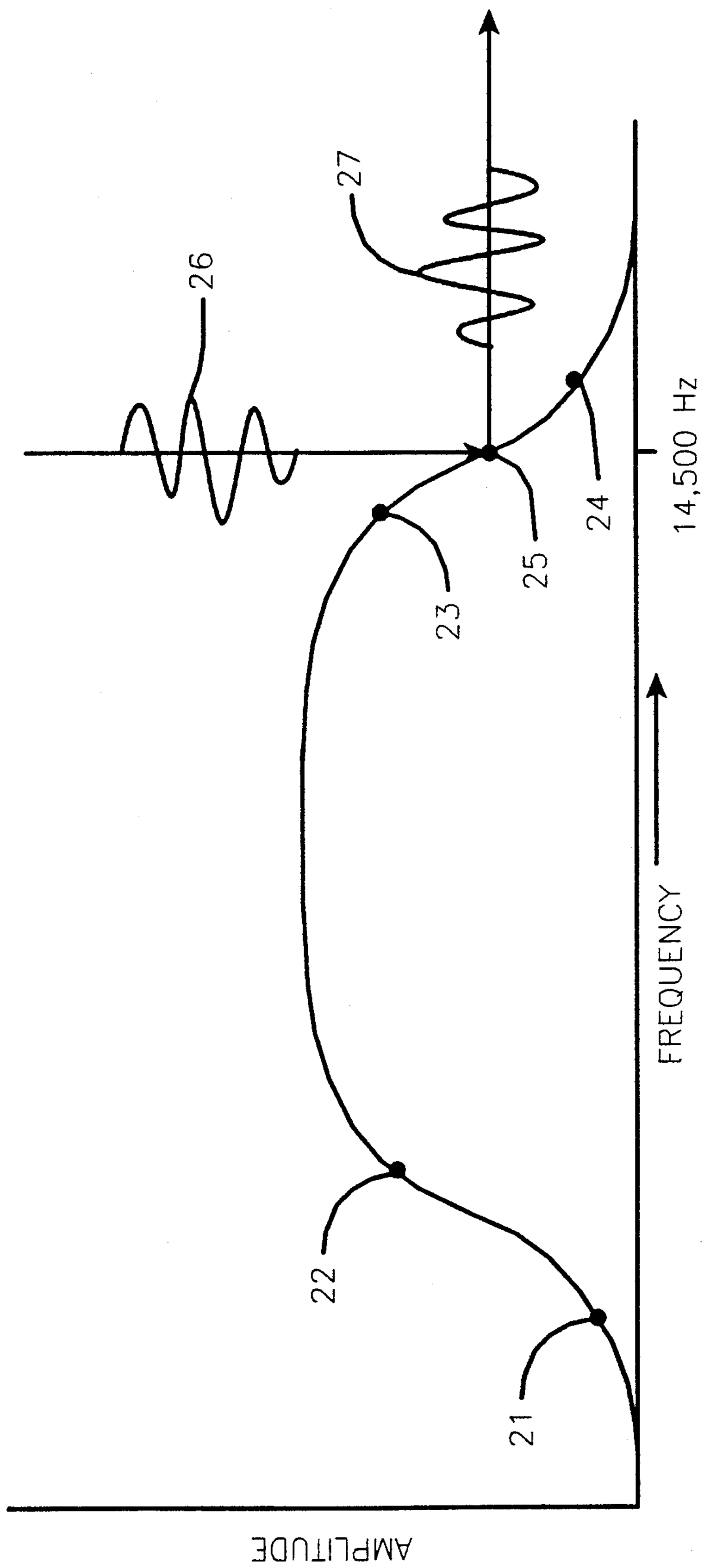

FIG. 2

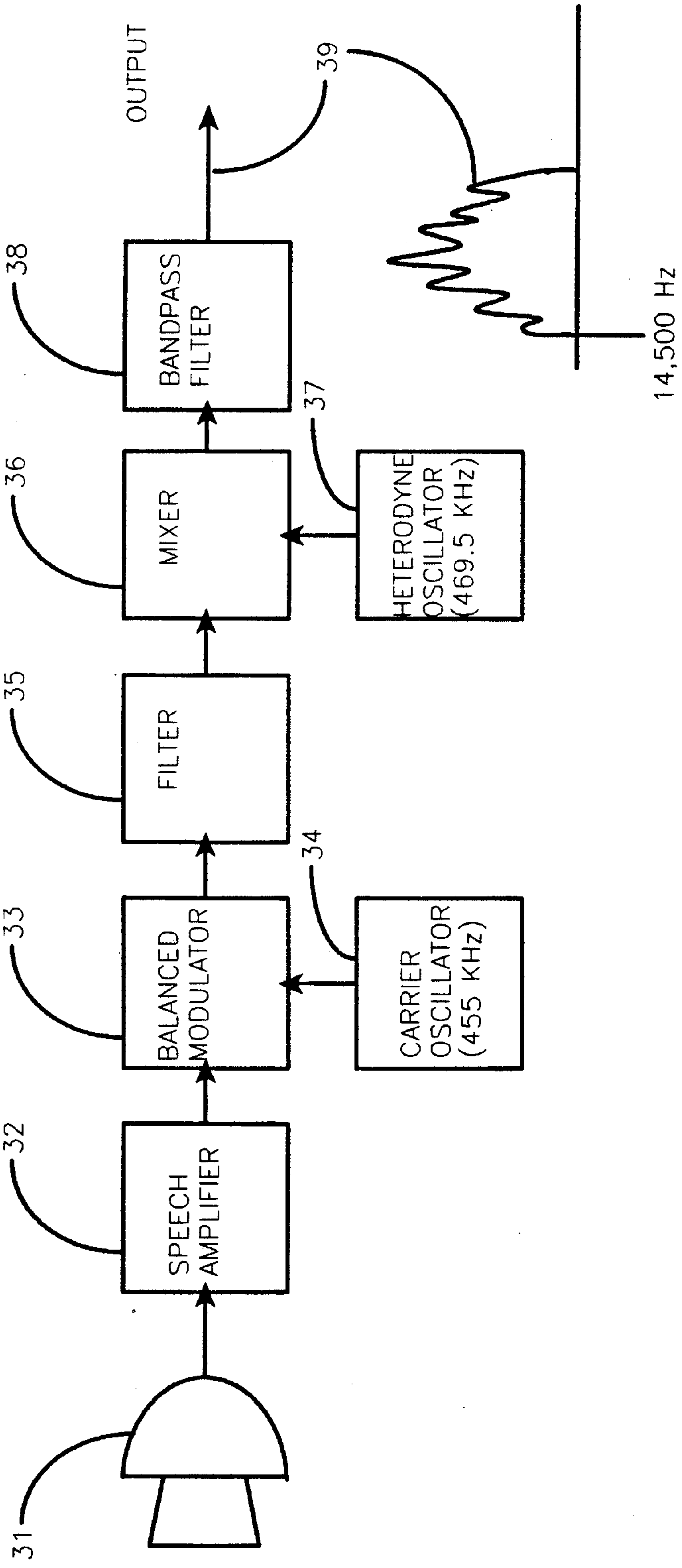
31
32
SPEECH AMPLIFIER
33
BALANCED MODULATOR
34
CARRIER OSCILLATOR (455 KHz)
35
FILTER
36
MIXER
37
HETERODYNE OSCILLATOR (469.5 KHz)
38
BANDPASS FILTER
OUTPUT
39
14,500 Hz

FIG. 3

SILENT SUBLIMINAL PRESENTATION SYSTEM

BACKGROUND—FIELD OF THE INVENTION

This invention relates in general to electronic audio signal processing and, in particular, to subliminal presentation techniques.

BACKGROUND—DESCRIPTION OF PRIOR ART

Subliminal learning enjoys wide use today and subliminal tapes are being manufactured by a number of companies in the United States alone. Several decades of scientific study indicate that subliminal messages can influence a human's attitudes and behavior. Subliminal, in these discussions, can be defined as "below the threshold of audibility to the conscious mind." To be effective however, the subliminally transmitted information (called affirmations by those in the profession) must be presented to the listener's ear in such a fashion that they can be perceived and "decoded" by the listener's subconscious mind. We are referring to audio information in this discussion, however, information could be inputted into the subject's subconscious mind through any of the body's sensors, such as touch, smell, sight or hearing. As an example, early development work in the subliminal field utilized motion pictures and slide projections as the medium. Early research into visual and auditory subliminal stimulation effects is exemplified by U.S. Pat. Nos. 3,060,795 of Corrigan, et al. and 3,278,676 of Becker. U.S. Pat. No. 4,395,600 of Lundy and Tyler is representative of later developments in today's subliminal message techniques.

The majority of the audio subliminal tapes available today are prepared using one basic technique. That is, the verbal affirmations are mixed with, and recorded at a lower level than, a "foreground" of music or sounds of ocean surf or a bubbling mountain brook or other similar "masking" sounds. The affirmations are generally recorded 5 decibels (db) or so below the "foreground" programming and regenerative automatic gain control is usually applied to permit the affirmations to change their recorded amplitude in direct proportion to the short term averaged amplitude of the continually varying "foreground" material. In other words, the volume of the affirmations will follow or track the volume changes of the "foreground" programming, but at a lower volume level. Circuit provisions are also usually included to "gate" the affirmations off when the music amplitude is low or zero. This insures that the affirmations cannot be heard during quiet program periods. Thus, today's subliminal affirmations can be characterized as being "masked" by music or other sounds, of constantly changing amplitude and of being reduced or cut off entirely during periods of low or quiet "foreground" programming.

One of the principal, and most widely objected to, deficiencies in available subliminal tape presentation techniques is that the presence of the "foreground" material is intrusive to both the listener and to anyone else in the immediate area. No matter what "foreground" material is chosen, the fact remains that this material can be heard by anyone within its range and presents a definite distraction to other activities such as conversation, thought, desire to listen to other programming such as radio or television, need to concentrate, etc. Additionally, and because the tapes are used repeatedly by the same listener, any "foreground" music or material eventually becomes monotonously tiring to that listener.

It is the purpose of the following described invention to eliminate or greatly reduce all of the above deficiencies. Although its application to the magnetic tape medium is described in the following discussion, the technique is equally applicable to most other desired transmission mediums, such as Compact Disc, videocassettes, digital tape recorders, Public Address (PA) systems, background music installations, computer software programs, random access memory (RAM), read only memory (ROM), "live", real time applications and other mediums now in existence or to be developed in the future.

Implemented on tape cassettes, for example, the subliminal presentation described here is inaudible i.e., high audio or ultrasonic frequencies, the affirmations are presented at a constant, high amplitude level, and they occupy their own "clear channel", non-masked frequency allocations. If desired, the previously described "foreground" music or other material can be added to the tape through use of an audio mixer. The "silent" recordings are inaudible to the user or by others present and are therefore very effective for use during periods of sleep or when in the presence of others. Additionally, the basic requirements of subliminal stimulation are met. That is, the affirmations are efficiently transmitted to the ear and, while undetected by the conscious mind, are perceived by and efficiently decoded by the subconscious mind.

OBJECTS OF THE INVENTION

Accordingly, several objects and advantages of my invention are:

(a) to provide a technique for producing a subliminal presentation which is inaudible to the listeners(s), yet is perceived and demodulated (decoded) by the ear for use by the subconscious mind.

(b) to provide a technique for transmitting inaudible subliminal information to the listener(s) at a constant, high level of signal strength and on a clear band of frequencies.

(c) to provide a technique for producing inaudible subliminal presentations to which music or other "foreground" programming may be added, if desired.

Still further objects and advantages will become apparent from a consideration of the ensuing description and drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

In the drawings, the first digit of each component number also refers to the figure number where that component can be located.

FIG. 1 represents the block diagram of a suitable system which will generate a frequency modulated (FM) signal at 14,500 Hz.

FIG. 2 represents an approximation of the frequency response curve of the human ear and the signal decoding process.

FIG. 3 represents the block diagram of a suitable system which will generate a single sideband, suppressed carrier, amplitude modulated (AM) signal at 14,500 Hz.

REFERENCE NUMERALS IN DRAWINGS

11 microphone or other	14 low distortion

-continued

REFERENCE NUMERALS IN DRAWINGS	
audio input signal	audio oscillator
12 audio preamplifier if required	15 high pass or band pass audio filter
13 frequency modulation circuit	16 output to tape recorder or other device
21 point on low freq end response curve	25 midpoint on curve between points 23 and 24
22 point on low freq end of ear response curve	26 speaker output of FIG. 1 to ear
23 point on high freq end of ear response curve	27 demodulated subliminal audio inputted to ear
24 point on high freq end of ear response curve	31 microphone
32 speech amplifier	33 balanced modulator
34 carrier oscillator (455 KHz)	35 filter
36 mixer	37 heterodyne oscillator (469.5 KHz)
38 bandpass filter	39 output signal

DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT

Please refer now to FIG. **1** and FIG. **2**, which are drawings of a preferred implementation of the invention.

The principle of operation of the silent subliminal presentation system is as follows:

An audio signal in the upper frequency region of the audio spectrum (for example, 14,500 Hz) is modulated with the desired information. The type of modulation may be any type suitable for subliminal applications; frequency modulation (FM), phase modulation (PM), upper single sideband with suppressed carrier, amplitude modulation (AM), tone modulation, etc.

For broadest application, the high audio frequency selected as the carrier frequency must meet two basic criteria:

(1) be high enough in the audio spectrum that its presence to the human ear is essentially unnoticed or undetectable (without the listener being informed that the signal is actually present) and,

(2) be low enough in the audio spectrum that it (and its modulation content) can produce a useful output power from home entertainment type cassette or reel-to-reel magnetic recorders.

This would also include, of course, small portable and automobile tape decks.

Alternatively, the output of the system can be fed directly into an audio amplifier and its speaker/earphone system, Public Address system, etc.

FIG. **1** provides the block diagram of an example of a system capable of generating the desired silent frequency modulated carrier.

The modulation information is inputted into the microphone **11**. Other suitable input devices may be substituted for microphone **11**, such as a tape recorder or a radio. The microphone **11** is connected to the preamplifier **12** and should have provisions for adjusting its gain in order that the optimum modulation index can be set in the frequency modulator **13**. The frequency modulator **13** modulates the frequency of oscillator **14** which has been adjusted for an output of 14,500 Hz as described above. The output of oscillator **14** is fed through a suitable bandpass filter **15** into the tape recorder or directly into a suitable amplifier/speaker system. It is the purpose of the bandpass filter to remove or attenuate audible products of the modulation process in order to maintain as audibly silent an output as practical.

On the receiving end, FIG. **2** represents an approximate and idealized frequency response curve of the human ear. The frequency modulated carrier (centered at 14,500 Hz), as generated above and played through a tape recorder or amplifier/speaker system, is shown on FIG. **2** as speaker output **26**, impinging upon the upper slope of the ear's response curve at point **25**. The frequency modulated excursions of the speaker output **26** swing between points **23** and **24** on the ear's upper response curve. Because the response curve between points **23** and **25** is relatively linear, this action results in a relatively linear demodulation of the original modulation intelligence, which is passed on subliminally to the inner ear. The amplitude of the demodulated output is not high enough to be detected by the conscious mind but is sufficient in amplitude to be detected by the subconscious mind. In the field of communications engineering design, the above demodulation process in known as slope detection and was used in early FM receiver design. In those receivers, the response curve was formed by the action of a tuned (inductive/capacitance) circuit. In our case, the response curve is formed by the natural response curve of the human ear. The same slope detection technique can be performed at the low frequency end of the human ear response curve. This region is indicated on FIG. **2** as between points **21** and **22**. This region, however, has a much smaller available bandwidth and is therefore more restricted as to the amount of information that can be transmitted in an inaudible manner.

In practice, the listener adjusts the volume control of the tape recorder or amplifier to a level just below that at which the listener hears an audible sound or noise from the speaker of the tape recorder. If the recording process is properly done, a spectrum analyzer or a calibrated sound level meter will reveal a strong signal emanating from the tape recorder speaker. A calibrated sound level meter, at a distance of 1 meter (with C weighting and referenced to the standard of 0.0002 micro bar) will typically indicate a silent power output of from 60 to 70 decibels. This is equivalent to the audio power of a loud conversation, yet, in the described system, is inaudible or unnoticed by the listener.

FIG. **3** illustrates a system which generates a suitable amplitude modulated (AM) signal, instead of the frequency modulated (FM) system described above. The output is a modulated, single sideband (SSB), suppressed carrier (AM) signal at 14,500 Hz.

The block diagram represents a common scheme for generating an SSB signal and will be briefly described.

The desired subliminal information is spoken into microphone **31**. This signal is amplified by speech amplifier **32** and injected into one port of balanced modulator **33**. A continuous wave signal of **455** KHz is generated by carrier oscillator **34** and is injected into the second port of balanced modulator **33**. The output of balanced modulator **33** is a double sideband, suppressed carrier signal at 455 KHz. This signal is fed through filter **35**, causing one of the two sidebands to be removed. This signal is fed into one port of mixer **36**. A continuous wave signal at a frequency of 469.5 KHz from hetrodyne oscillator **37** is fed into the other port of mixer **36**, resulting in an output of the original subliminal audio information but translated 14,500 Hz higher in frequency. The bandpass filter **38** attenuates signals and noise outside of the frequencies of interest. The ampli-

tude modulated audio output signal is shown as output **39**.

Thus, as stated earlier, my invention provides a new system for subliminal presentations which is:

(a) silent,

(b) outputs a constant, high level modulated signal and,

(c) occupies a band of clear channel frequencies.

The foregoing description of the preferred embodiment of the invention has been presented for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed. Many modifications and variations are possible in light of the above discussions. It is intended that the scope of the invention be limited not only by this detailed description, but rather by the claims appended hereto.

What is claimed:

1. A silent communications system, comprising:

(a) amplitude modulated carrier means for generating signals located in non-aural portions of the audio and in the lower portion of the ultrasonic frequency spectrum said signals modulated with information to be perceived by a listener's brain and,

(b) acoustic and ultrasonic transducer means for propagating said signals, for inducement into the brain, of the listener, and,

(c) recording means for storing said modulated signals on mechanical, magnetic and optical media for delayed or repeated transmissions to the listener.

2. A silent communications system, comprising:

(a) frequency modulated carrier means for generating signals located in non-aural portions of the audio and in the lower portion of the ultrasonic frequency spectrum, said signals modulated with information to be perceived by a listener's brain, and;

(b) acoustic and ultrasonic transducer means for propagating said signals, for inducement into the brain of the listener, and;

(c) recording means for storing said modulated signals on mechanical, magnetic and optical media for delayed or repeated transmissions to the listener.

3. A silent communications system, comprising:

(a) a combination of amplitude and frequency modulated carrier means for generating signals located in non-aural portions of the audio and in the lower portion of the ultrasonic frequency spectrum, said signals modulated with information to be perceived by a listener's brain, and

(b) acoustic and ultrasonic transducer means for propagating said signals, for inducement into the brain of the listener;

(c) recording means for storing said modulated signals on mechanical, magnetic and optical media for delayed or repeated transmissions to the listener.

* * * * *

US005356368A

United States Patent [19]

Monroe

[11] **Patent Number: 5,356,368**

[45] **Date of Patent: * Oct. 18, 1994**

[54] **METHOD OF AND APPARATUS FOR INDUCING DESIRED STATES OF CONSCIOUSNESS**

[75] Inventor: **Robert A. Monroe**, Nelson County, Va.

[73] Assignee: **Interstate Industries Inc.**, Faber, Va.

[*] Notice: The portion of the term of this patent subsequent to May 25, 2010 has been disclaimed.

[21] Appl. No.: **664,176**

[22] Filed: **Mar. 1, 1991**

[51] **Int. Cl.⁵** **A61M 21/00**

[52] **U.S. Cl.** **600/28**; 128/732

[58] **Field of Search** 600/26-28; 128/731-732

[56] **References Cited**

U.S. PATENT DOCUMENTS

2,466,054	4/1949	Siebel .	
3,160,159	12/1964	Hoody et al. .	
3,576,185	4/1971	Schulz et al. .	
3,712,292	1/1973	Zentmeyer, Jr. .	
3,753,433	8/1973	Bakerich et al. .	
3,826,243	7/1974	Anderson .	
3,837,331	9/1974	Ross .	
3,884,218	5/1975	Monroe .	
4,034,741	7/1977	Adams et al. .	
4,141,344	2/1979	Barbara .	
4,227,516	10/1980	Meland et al. .	
4,335,710	6/1982	Williamson .	
4,573,449	3/1986	Warnke .	
4,834,701	5/1989	Masaki .	
4,883,067	11/1989	Knispel et al.	600/28
5,036,858	8/1991	Carter et al. .	
5,101,831	4/1992	Koyama et al.	600/26

Primary Examiner—Lee S. Cohen
Assistant Examiner—J. P. Lacyk
Attorney, Agent, or Firm—Sughrue, Mion, Zinn, Macpeak & Seas

[57] **ABSTRACT**

Improved methods and apparatus for entraining human brain patterns, employing frequency following response (FFR) techniques, facilitate attainment of desired states of consciousness. In one embodiment, a plurality of electroencephalogram (EEG) waveforms, characteristic of a given state of consciousness, are combined to yield an EEG waveform to which subjects may be susceptible more readily. In another embodiment, sleep patterns are reproduced based on observed brain patterns during portions of a sleep cycle; entrainment principles are applied to induce sleep. In yet another embodiment, entrainment principles are applied in the work environment, to induce and maintain a desired level of consciousness. A portable device also is described.

28 Claims, 21 Drawing Sheets

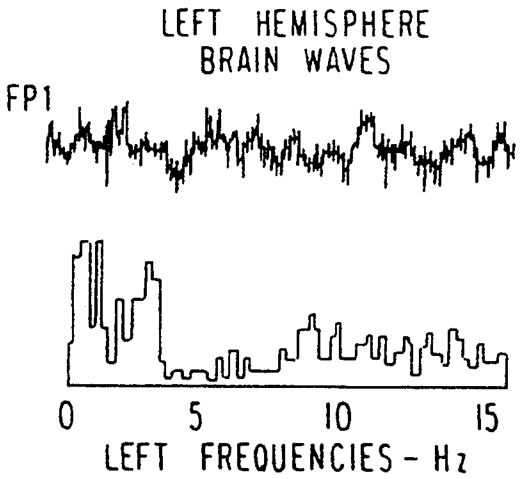

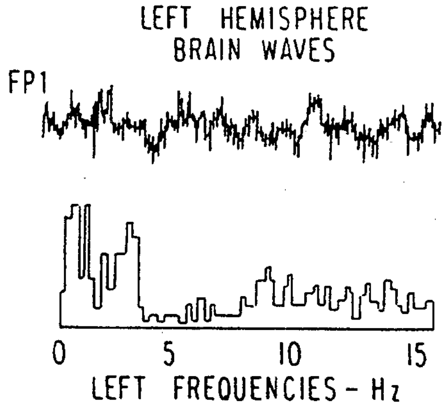

FIG. 1A

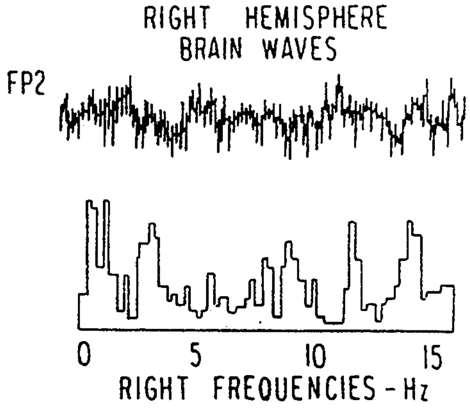

FIG. 1B

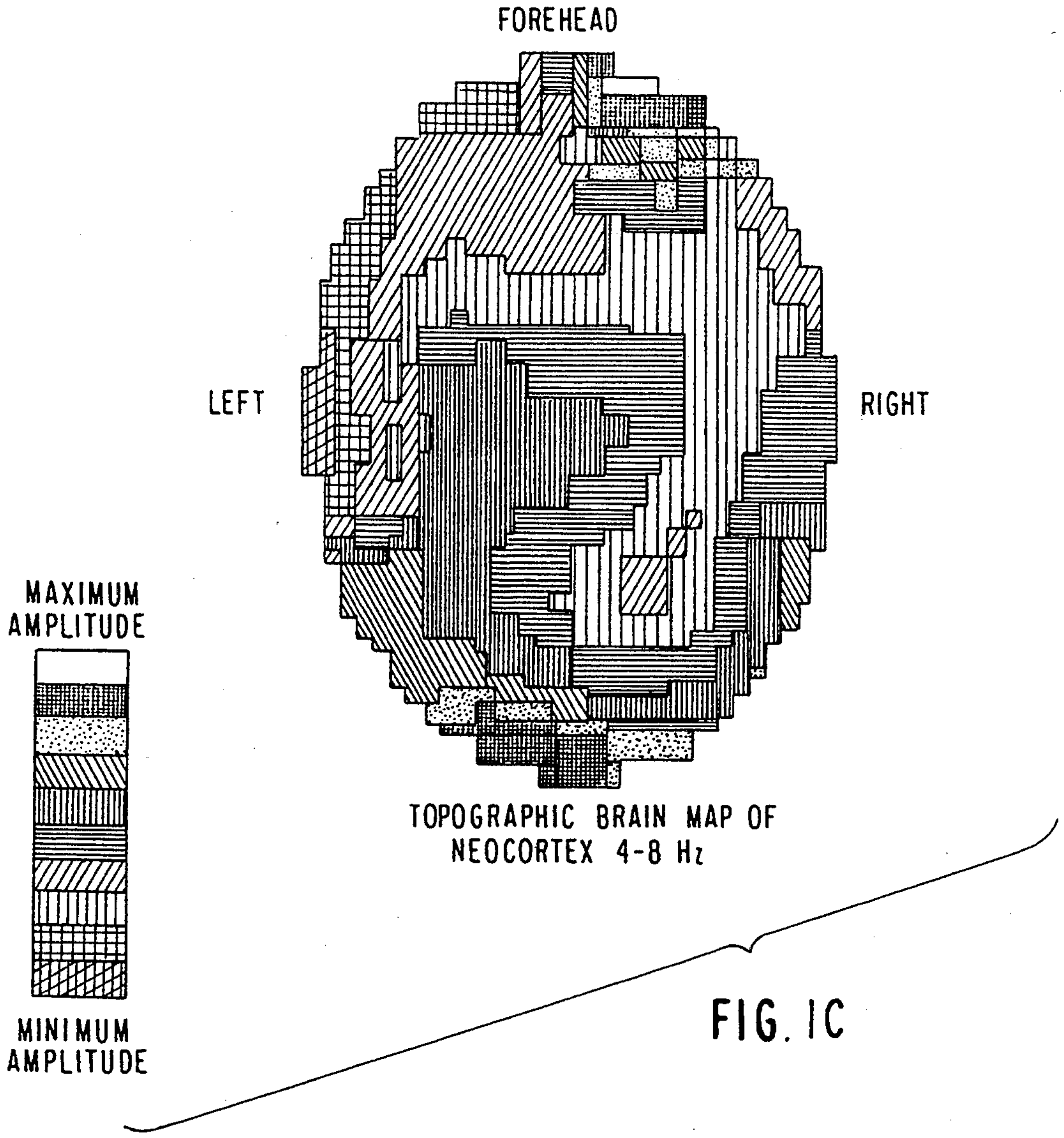
FOREHEAD
LEFT
RIGHT
TOPOGRAPHIC BRAIN MAP OF
NEOCORTEX 4-8 Hz
MAXIMUM
AMPLITUDE
MINIMUM
AMPLITUDE

FIG. 1C

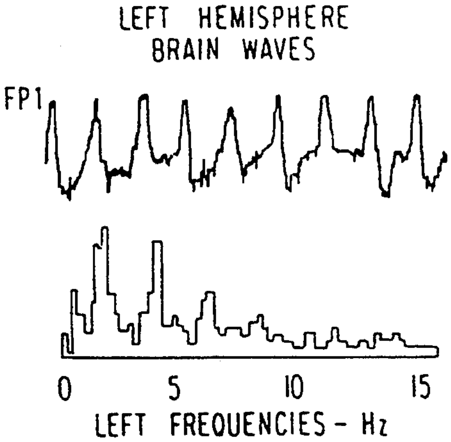

FIG. 1D

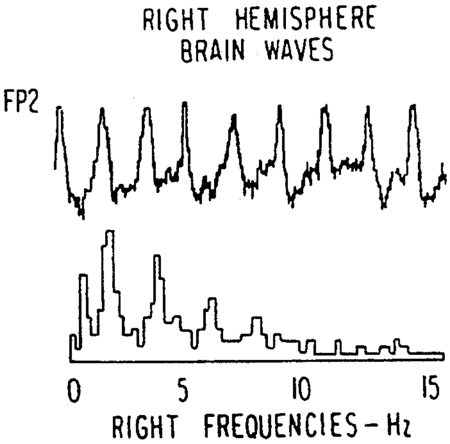

FIG. 1E

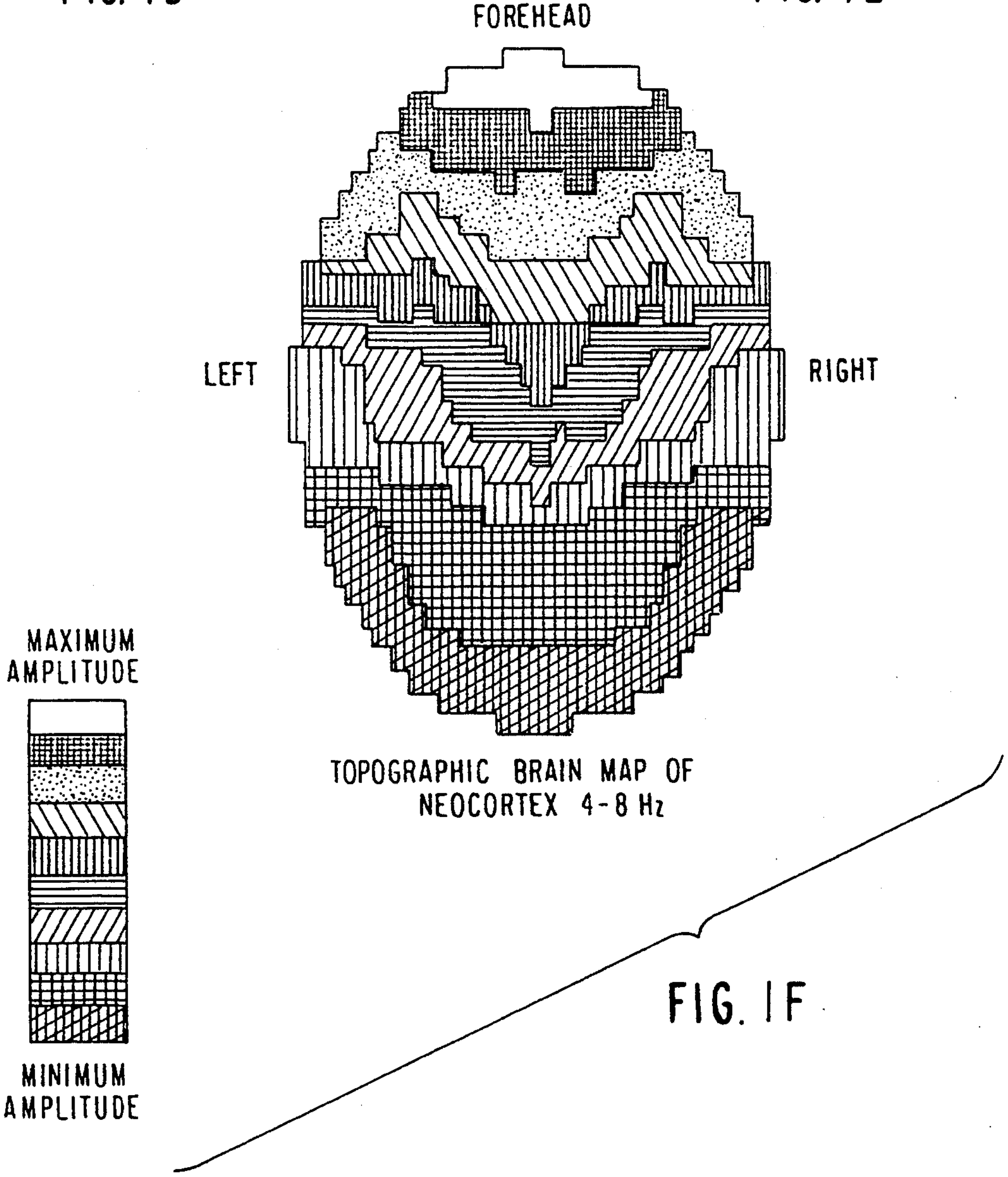
FOREHEAD
LEFT
RIGHT
TOPOGRAPHIC BRAIN MAP OF
NEOCORTEX 4-8 Hz
MAXIMUM
AMPLITUDE
MINIMUM
AMPLITUDE

FIG. 1F

FIG. 2

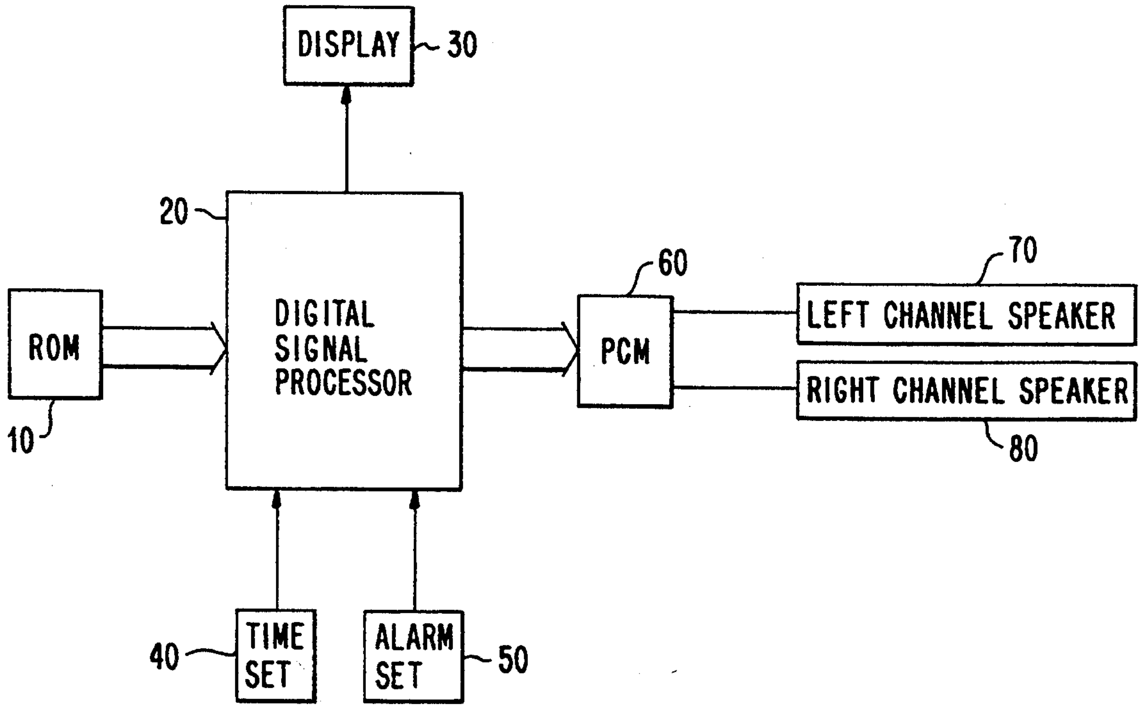
DISPLAY
30
20
DIGITAL SIGNAL PROCESSOR
ROM
10
60
PCM
70
LEFT CHANNEL SPEAKER
RIGHT CHANNEL SPEAKER
80
40
TIME SET
ALARM SET
50

FIG. 3

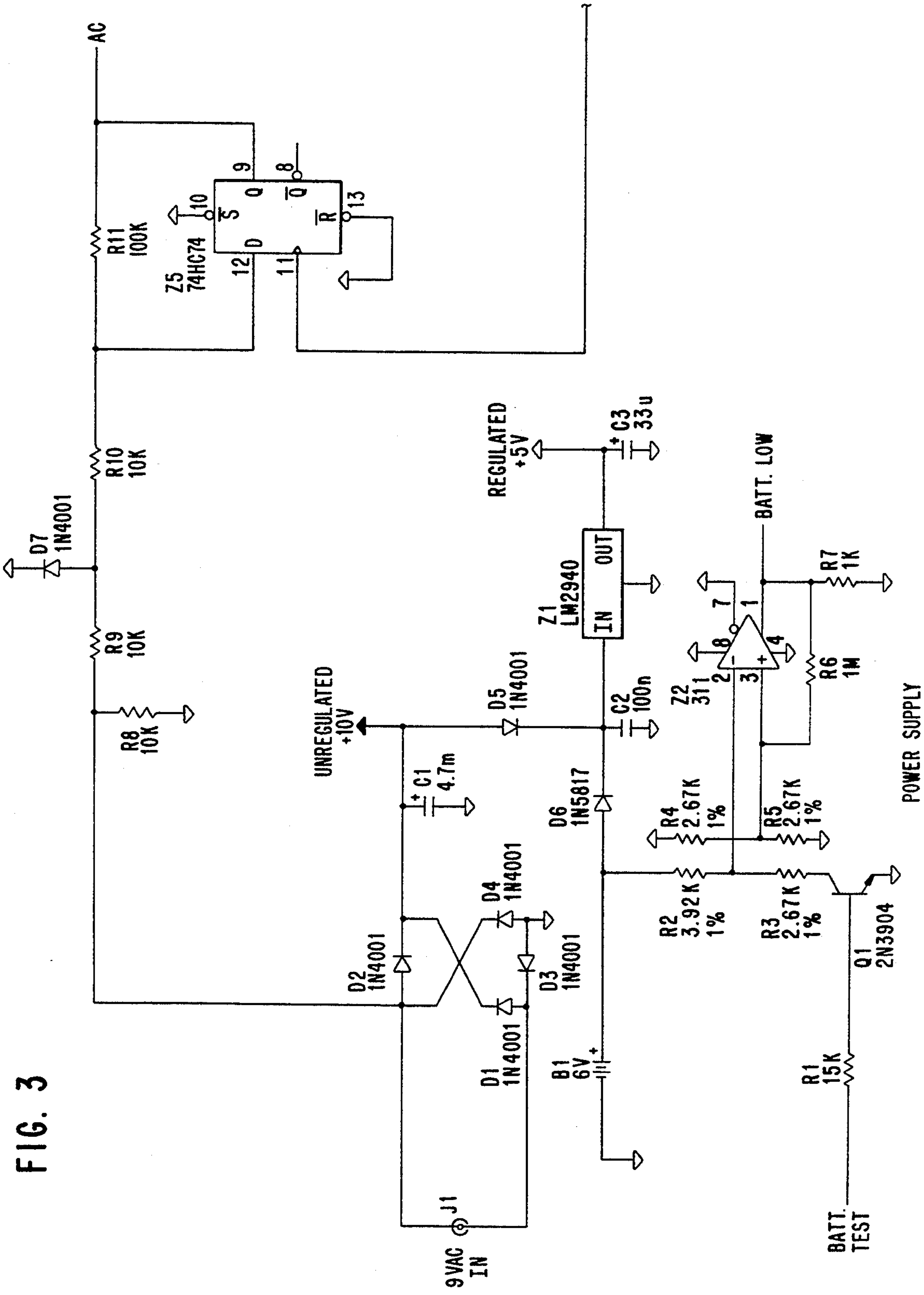
AC
R11
100K
Z5
74HC74
D
S̄
Q
Q̄
R̄
9
8
10
12
11
13
D7
1N4001
R10
10K
R9
10K
R8
10K
REGULATED
+5V
C3
33u
Z1
LM2940
IN
OUT
BATT. LOW
R7
1K
UNREGULATED
+10V
D5
1N4001
C2
100n
Z2
311
7
8
1
2
3
4
R6
1M
C1
4.7m
D6
1N5817
R4
2.67K
1%
R5
2.67K
1%
POWER SUPPLY
R2
3.92K
1%
R3
2.67K
1%
Q1
2N3904
D4
1N4001
D2
1N4001
D3
1N4001
D1
1N4001
B1
6V
R1
15K
J1
9VAC
IN
BATT.
TEST

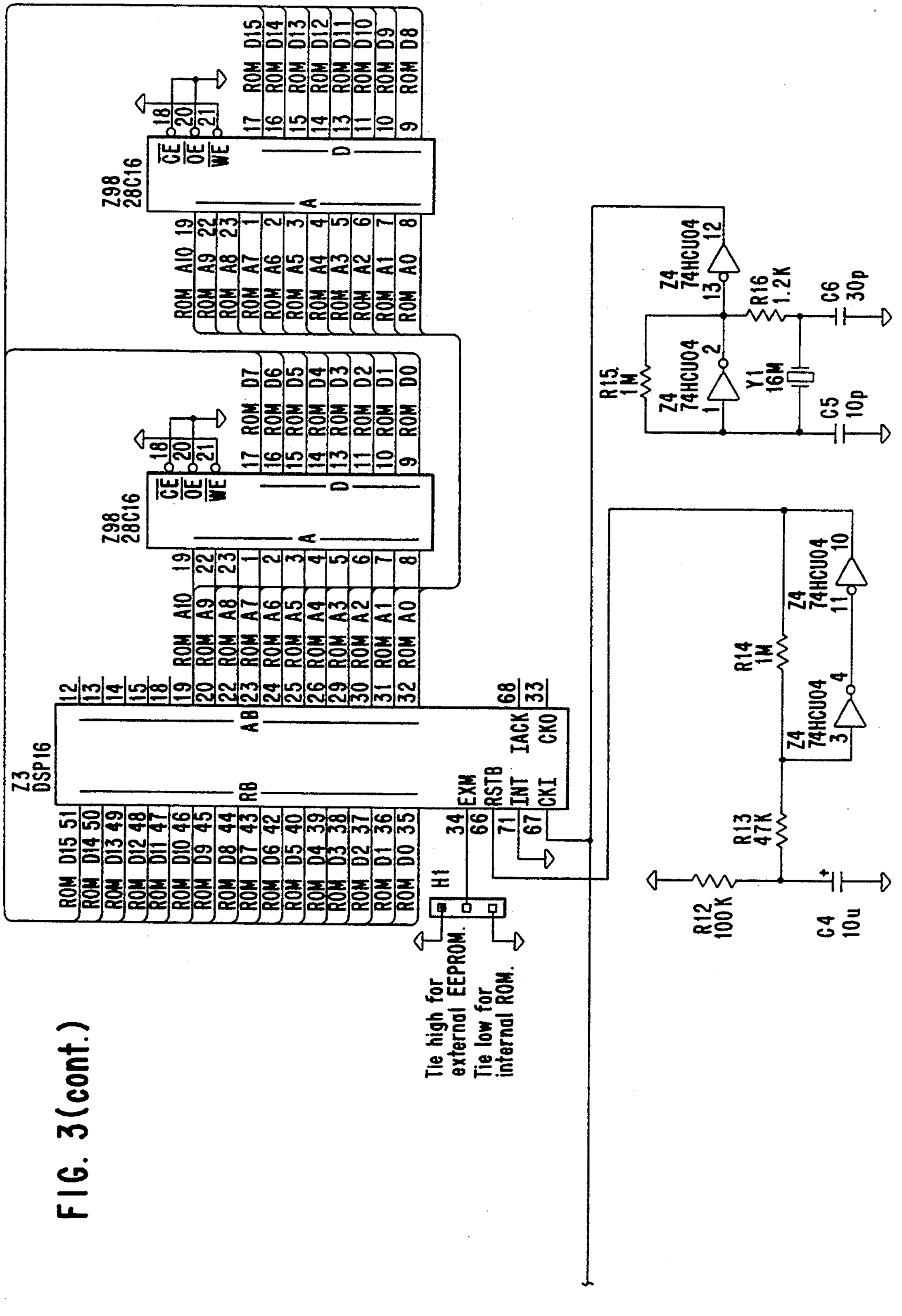
Z3
DSP16
RB
AB
EXM
RSTB
INT
CKI
IACK
CKO
Z98
28C16
CE
OE
WE
A
D
ROM D15 … ROM D0
ROM A10 … ROM A0
H1
Tie high for external EEPROM.
Tie low for internal ROM.
R12 100K
C4 10u
R13 47K
R14 1M
Z4 74HCU04
R15 1M
R16 1.2K
Y1 16M
C5 10p
C6 30p
CPU

FIG. 3(cont.)

FIG. 4

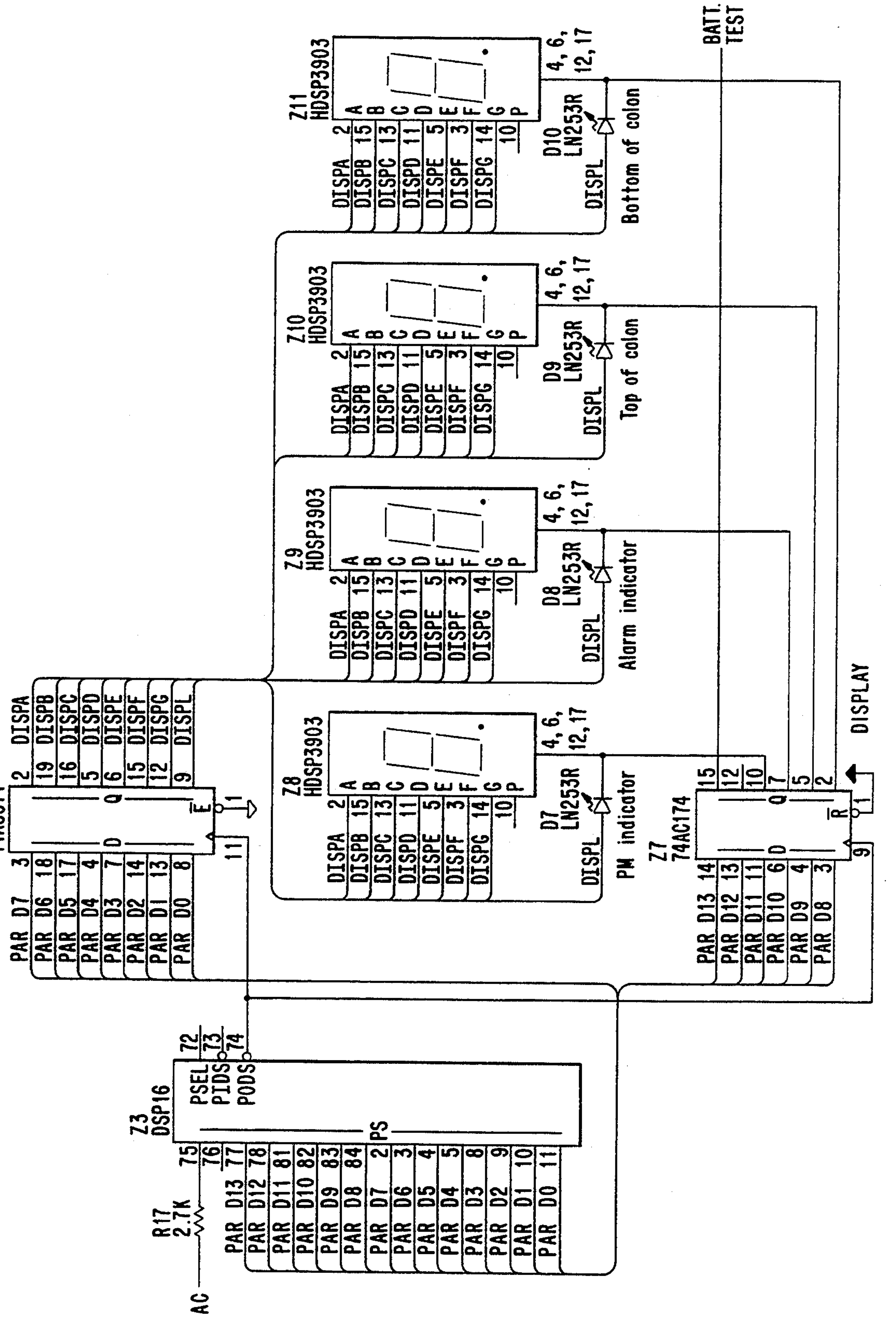
Z3
DSP16
PSEL
PIDS
PODS
PS
72
73
74
75
76
77
78
81
82
83
84
R17
2.7K
AC
PAR D13
PAR D12
PAR D11
PAR D10
PAR D9
PAR D8
PAR D7
PAR D6
PAR D5
PAR D4
PAR D3
PAR D2
PAR D1
PAR D0
Z6
74AS374
D
Q
E
11
1
DISPA
DISPB
DISPC
DISPD
DISPE
DISPF
DISPG
DISPL
Z7
74AC174
R
DISPLAY
BATT.
TEST
Z8
HDSP3903
Z9
HDSP3903
Z10
HDSP3903
Z11
HDSP3903
A
B
C
D
E
F
G
P
4, 6,
12, 17
D7
LN253R
PM indicator
D8
LN253R
Alarm indicator
D9
LN253R
Top of colon
D10
LN253R
Bottom of colon

FIG. 5

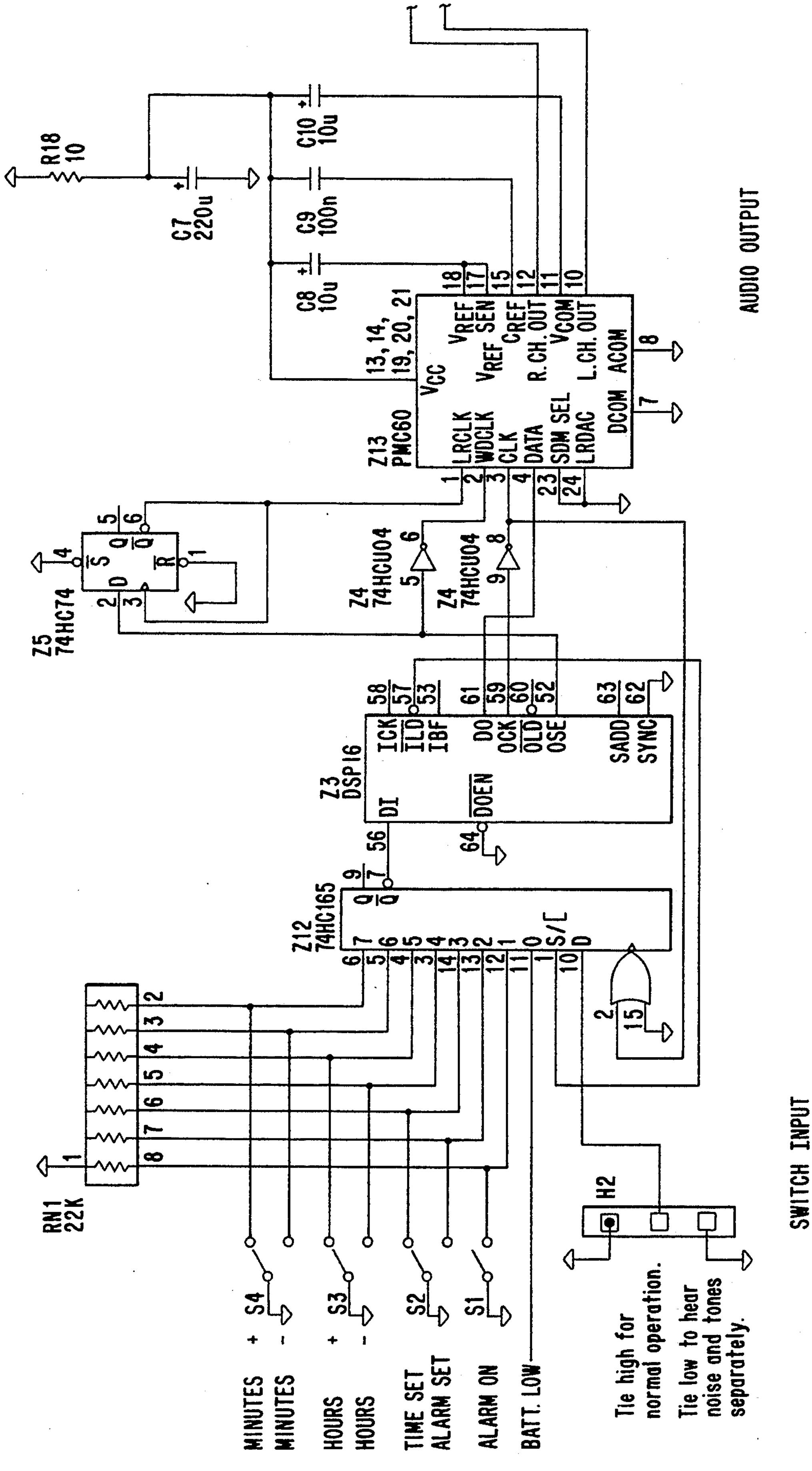
R18
10
C7
220u
C8
10u
C9
100n
C10
10u
Z13
PMC60
13, 14, 19, 20, 21
VCC
VREF
VREF SEN
CREF
R. CH. OUT
VCOM
L. CH. OUT
ACOM
LRCLK
WDCLK
CLK
DATA
SDM SEL
LRDAC
DCOM
AUDIO OUTPUT
Z5
74HC74
D
S
Q
Q
R
Z4
74HCU04
Z4
74HCU04
Z3
DSP16
ICK
ILD
IBF
DO
OCK
OLD
OSE
SADD
SYNC
DI
DOEN
Z12
74HC165
Q
Q
S/L
D
RN1
22K
H2
SWITCH INPUT
MINUTES +
MINUTES -
HOURS +
HOURS -
TIME SET
ALARM SET
ALARM ON
BATT. LOW
S4
S3
S2
S1
Tie high for normal operation.
Tie low to hear noise and tones separately.

FIG. 5 (cont.)

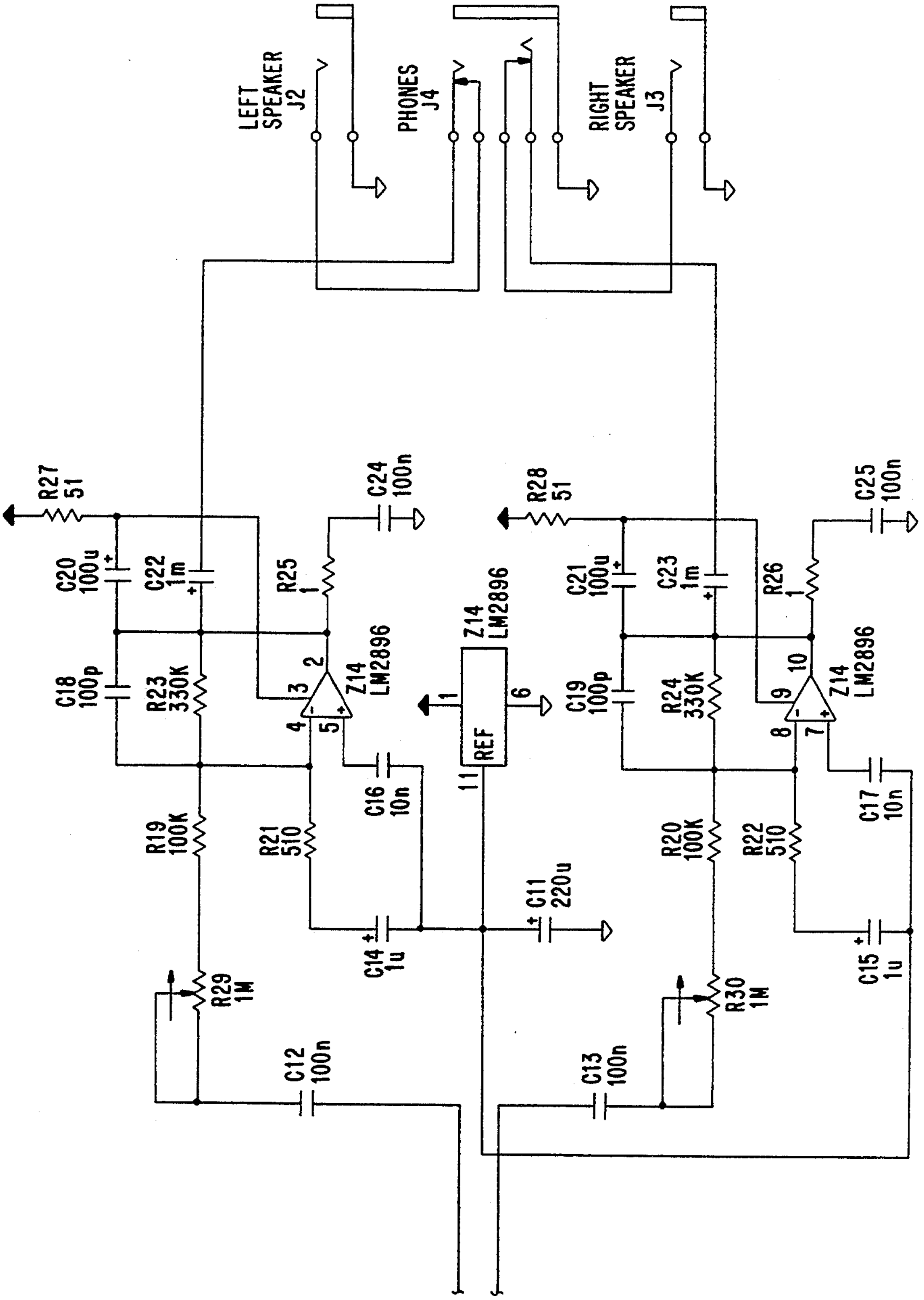
LEFT SPEAKER J2
PHONES J4
RIGHT SPEAKER J3
R27 51
R28 51
C20 100u
C21 100u
C22 1m
C23 1m
C24 100n
C25 100n
R25 1
R26 1
C18 100p
C19 100p
R23 330K
R24 330K
Z14 LM2896
Z14 LM2896
Z14 LM2896
REF
1
2
3
4
5
6
7
8
9
10
11
C16 10n
C17 10n
R19 100K
R20 100K
R21 510
R22 510
C11 220u
C14 1u
C15 1u
R29 1M
R30 1M
C12 100n
C13 100n

FIG. 6A

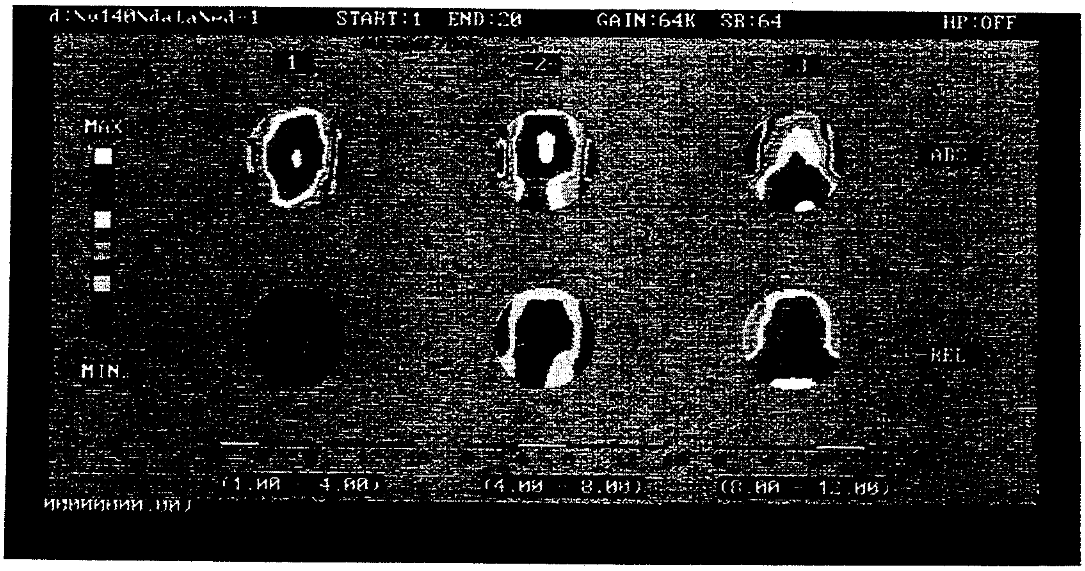
START:1 END:20
GAIN:64K SR:64
HP:OFF
1
2
3
MAX
MIN
ABS
REL
(1.00 - 4.00)
(4.00 - 8.00)
(8.00 - 12.00)

FIG. 6B

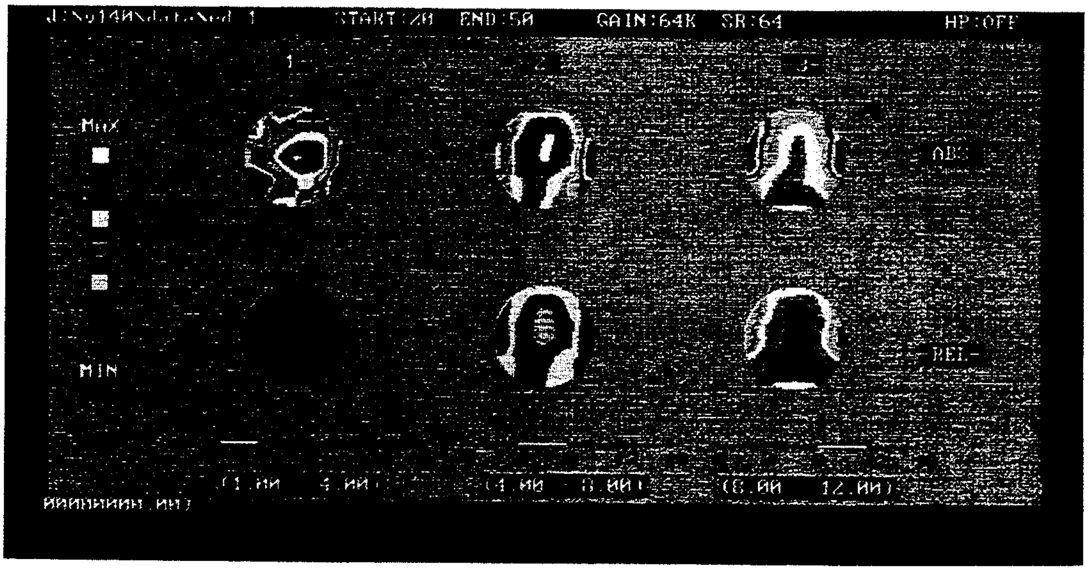
START:20 END:50
GAIN:64K SR:64
HP:OFF
1
2
3
MAX
MIN
ABS
REL
(1.00 - 4.00)
(4.00 - 8.00)
(8.00 - 12.00)

FIG. 6C

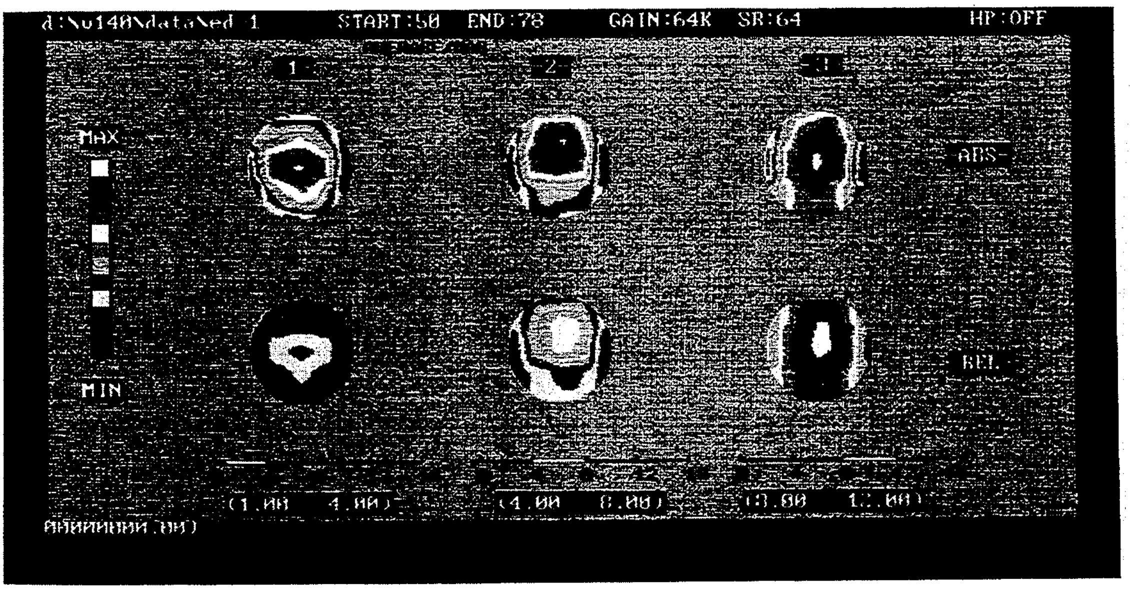

d:\u140\data\ed 1 START:50 END:78 GAIN:64K SR:64 HP:OFF

1
2
3
MAX
MIN
ABS
REL
(1.00 4.00)
(4.00 8.00)
(8.00 12.00)
00000000.00)

FIG. 6D

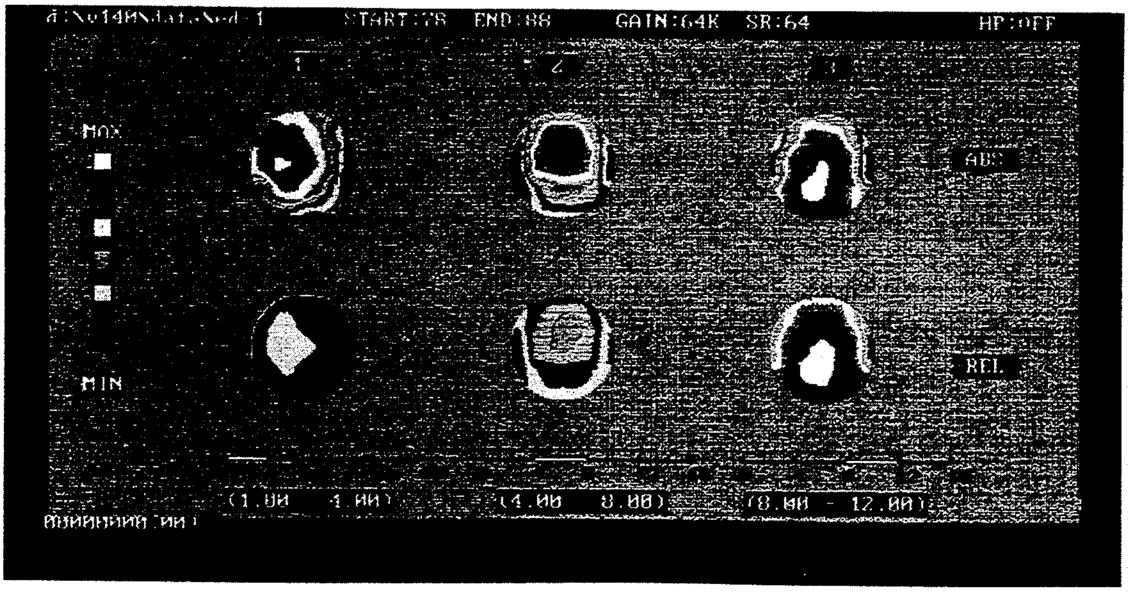

d:\u140\data\ed 1 START:78 END:88 GAIN:64K SR:64 HP:OFF

1
2
3
MAX
MIN
ABS
REL
(1.00 4.00)
(4.00 8.00)
(8.00 - 12.00)
00000000.00)

FIG. 6E

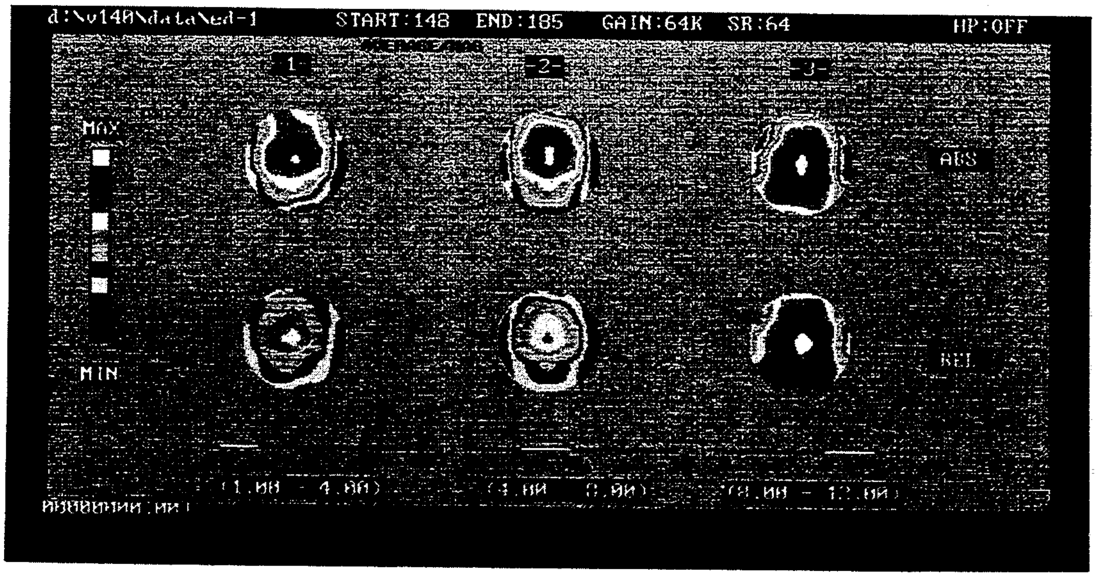
d:\v140\data\ed-1
START:148
END:185
GAIN:64K
SR:64
HP:OFF
1
2
3
MAX
MIN
ABS
REL
(1.00 - 4.00)
(4.00 8.00)
(8.00 - 12.00)
00000000.00

FIG. 6F

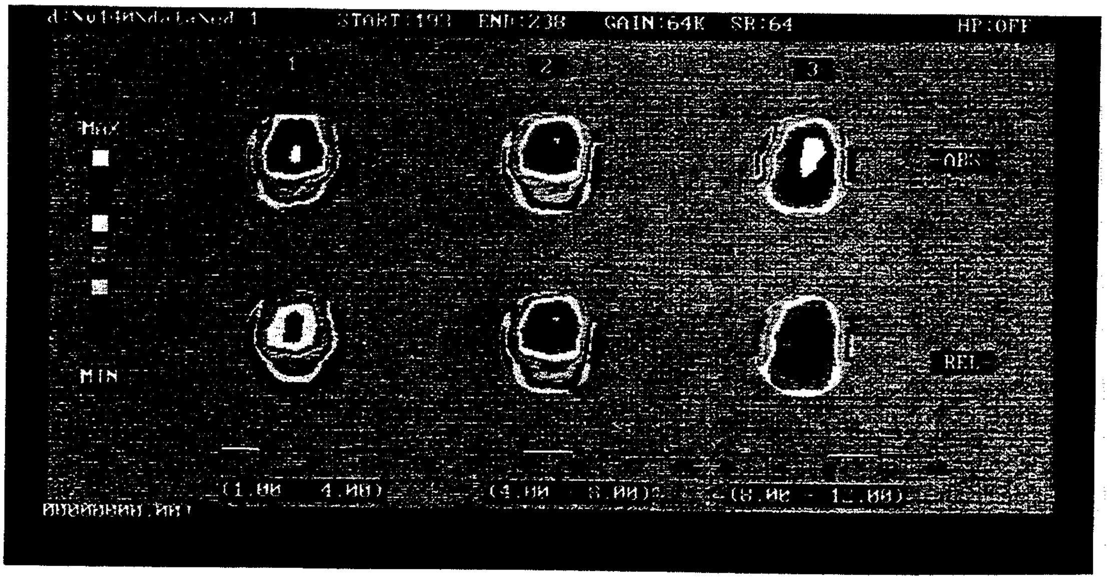
d:\v140\data\ed-1
START:193
END:238
GAIN:64K
SR:64
HP:OFF
1
2
3
MAX
MIN
ABS
REL
(1.00 4.00)
(4.00 8.00)
(8.00 - 12.00)
00000000.00

FIG. 6G

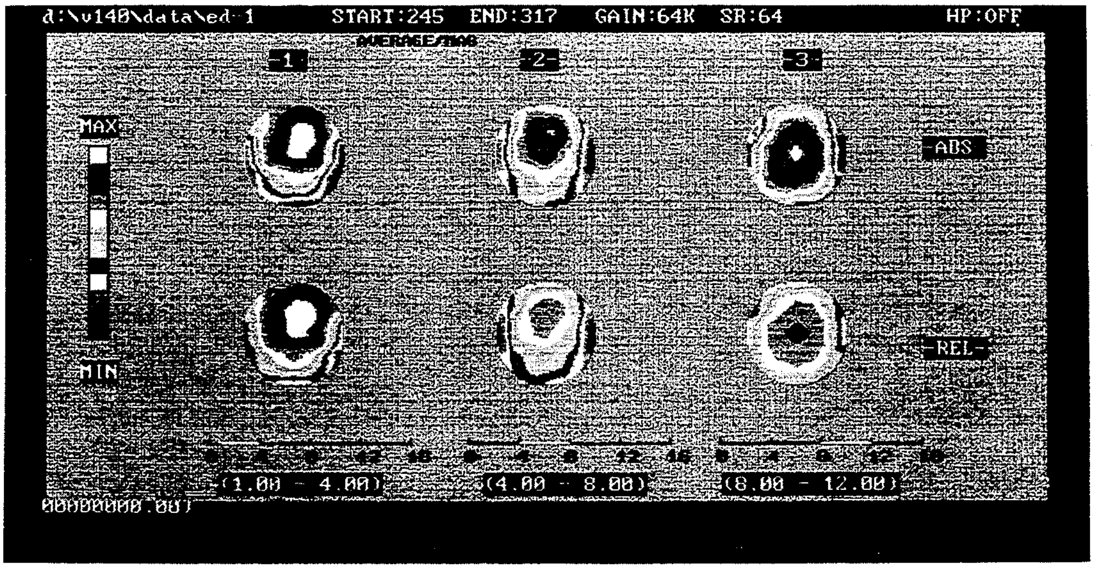
d:\v140\data\ed-1
START:245
END:317
GAIN:64K
SR:64
HP:OFF
AVERAGE/MAG
-1-
-2-
-3-
MAX
MIN
-ABS-
-REL-
0 4 8 12 16
0 4 8 12 16
0 4 8 12 16
(1.00 - 4.00)
(4.00 - 8.00)
(8.00 - 12.00)
00000000.00)

FIG. 6H

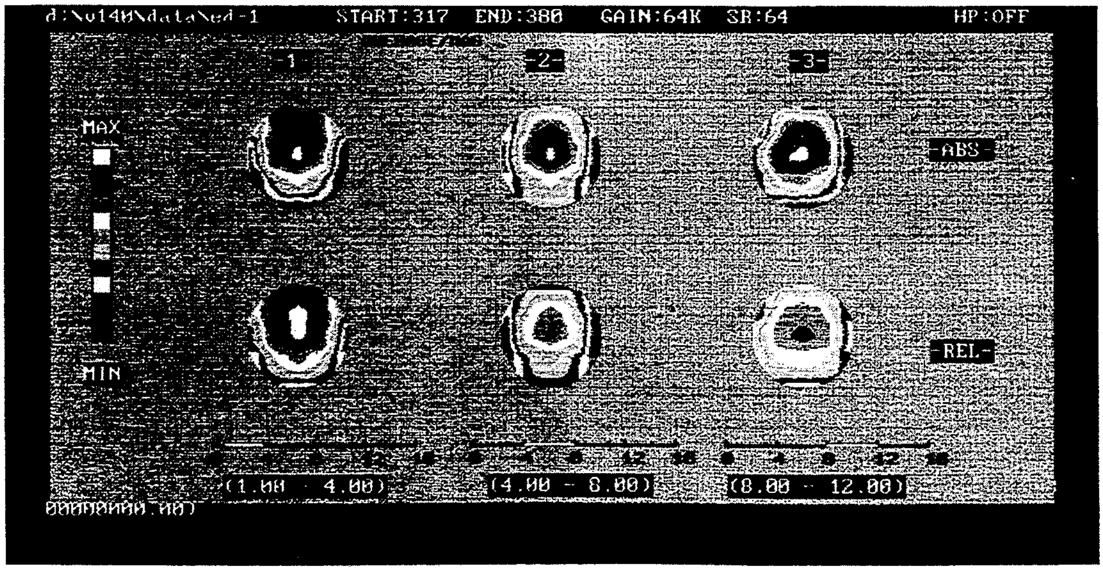
d:\v140\data\ed-1
START:317
END:380
GAIN:64K
SR:64
HP:OFF
AVERAGE/MAG
-1-
-2-
-3-
MAX
MIN
-ABS-
-REL-
0 4 8 12 16
0 4 8 12 16
0 4 8 12 16
(1.00 - 4.00)
(4.00 - 8.00)
(8.00 - 12.00)
00000000.00)

FIG. 6I

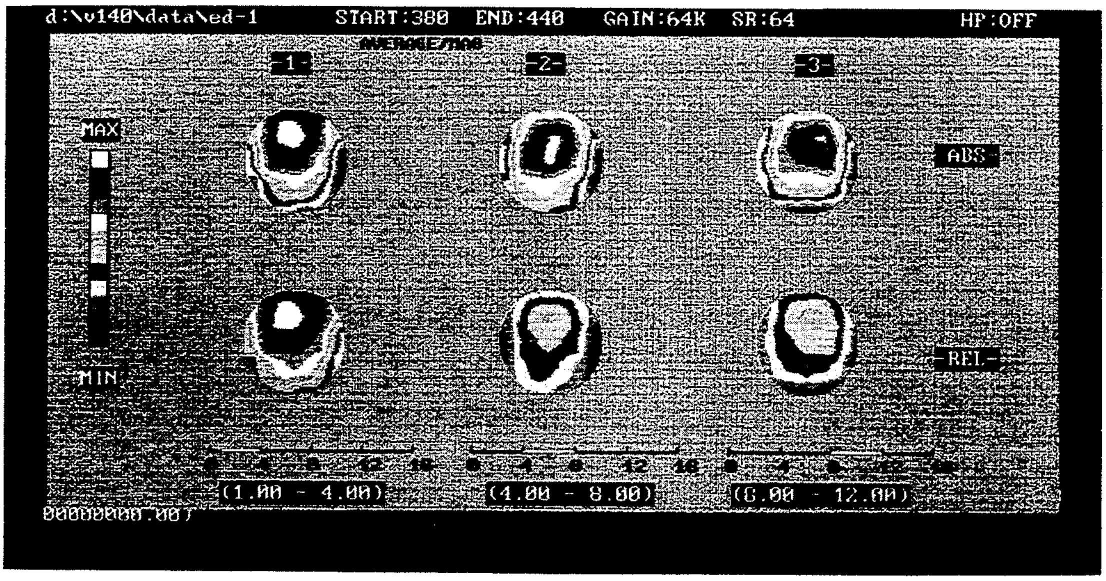
d:\v140\data\ed-1
START:380
END:440
GAIN:64K
SR:64
HP:OFF
-1-
-2-
-3-
MAX
MIN
ABS
REL
(1.00 - 4.00)
(4.00 - 8.00)
(8.00 - 12.00)
00000000.00

FIG. 6J

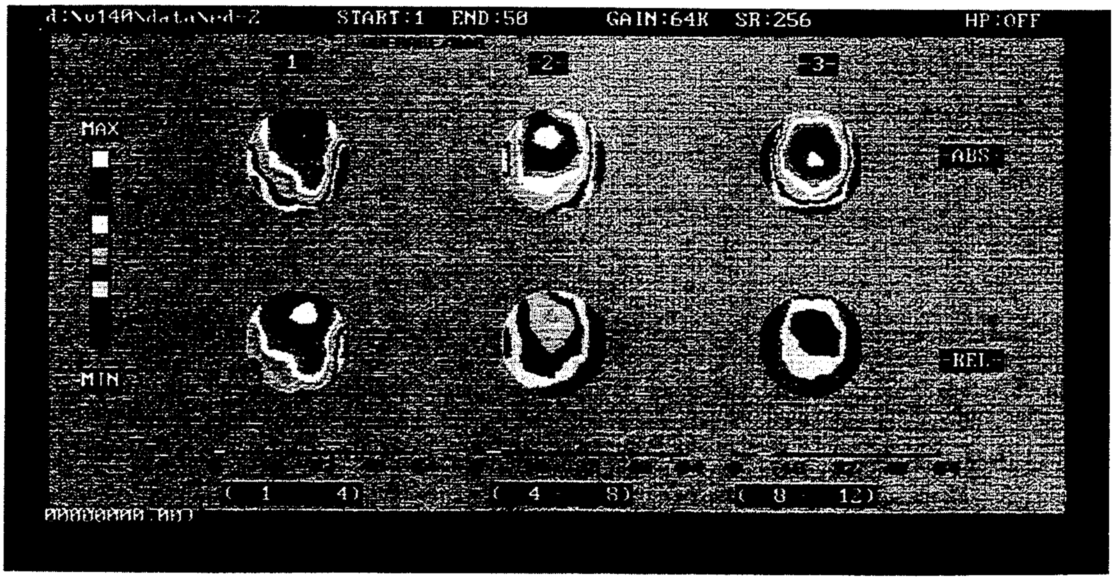
d:\v140\data\ed-2
START:1
END:58
GAIN:64K
SR:256
HP:OFF
1
-2-
-3-
MAX
MIN
ABS
REL
(1 4)
(4 8)
(8 12)
00000000.00

FIG. 7

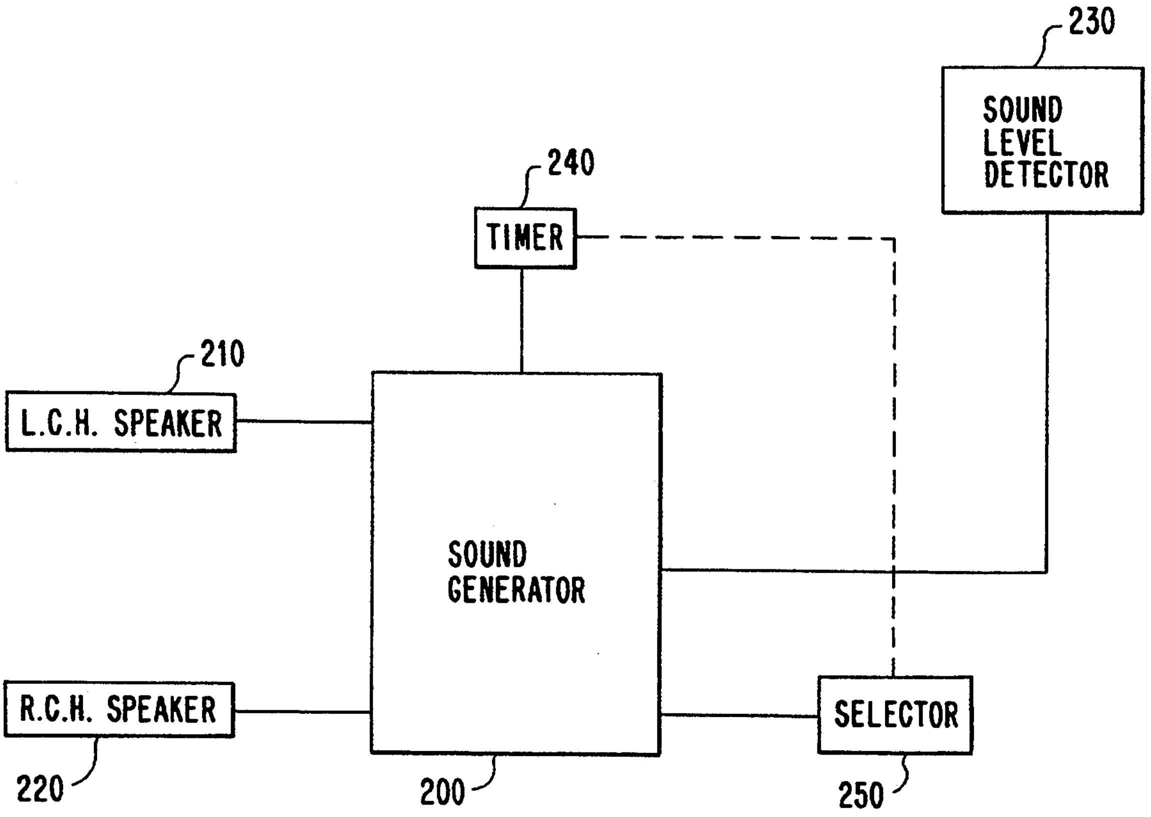
230
SOUND LEVEL DETECTOR
240
TIMER
210
L.C.H. SPEAKER
SOUND GENERATOR
R.C.H. SPEAKER
SELECTOR
220
200
250

FIG. 8A

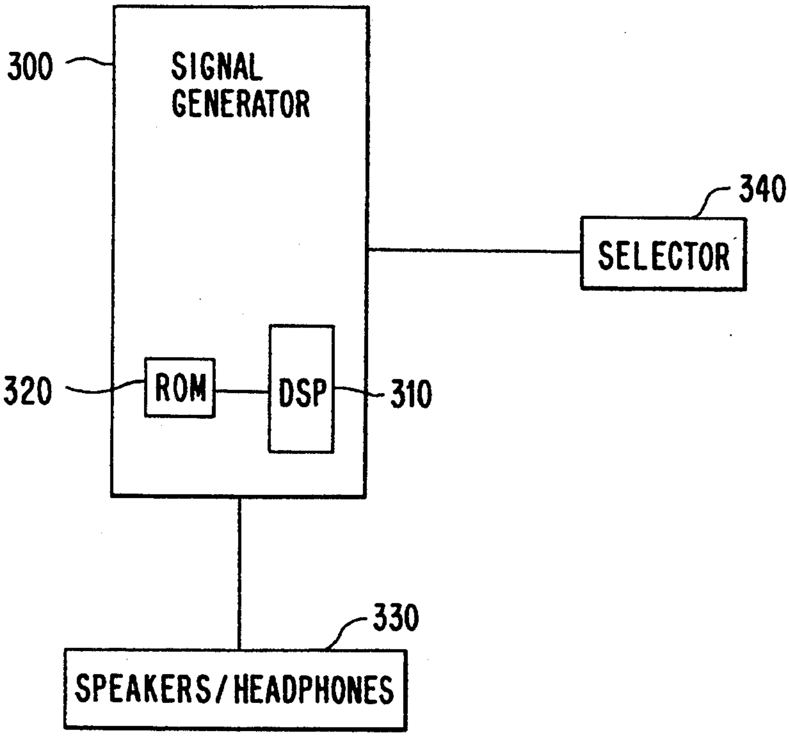
300
SIGNAL GENERATOR
340
SELECTOR
320
ROM
DSP
310
330
SPEAKERS/HEADPHONES

FIG. 8B

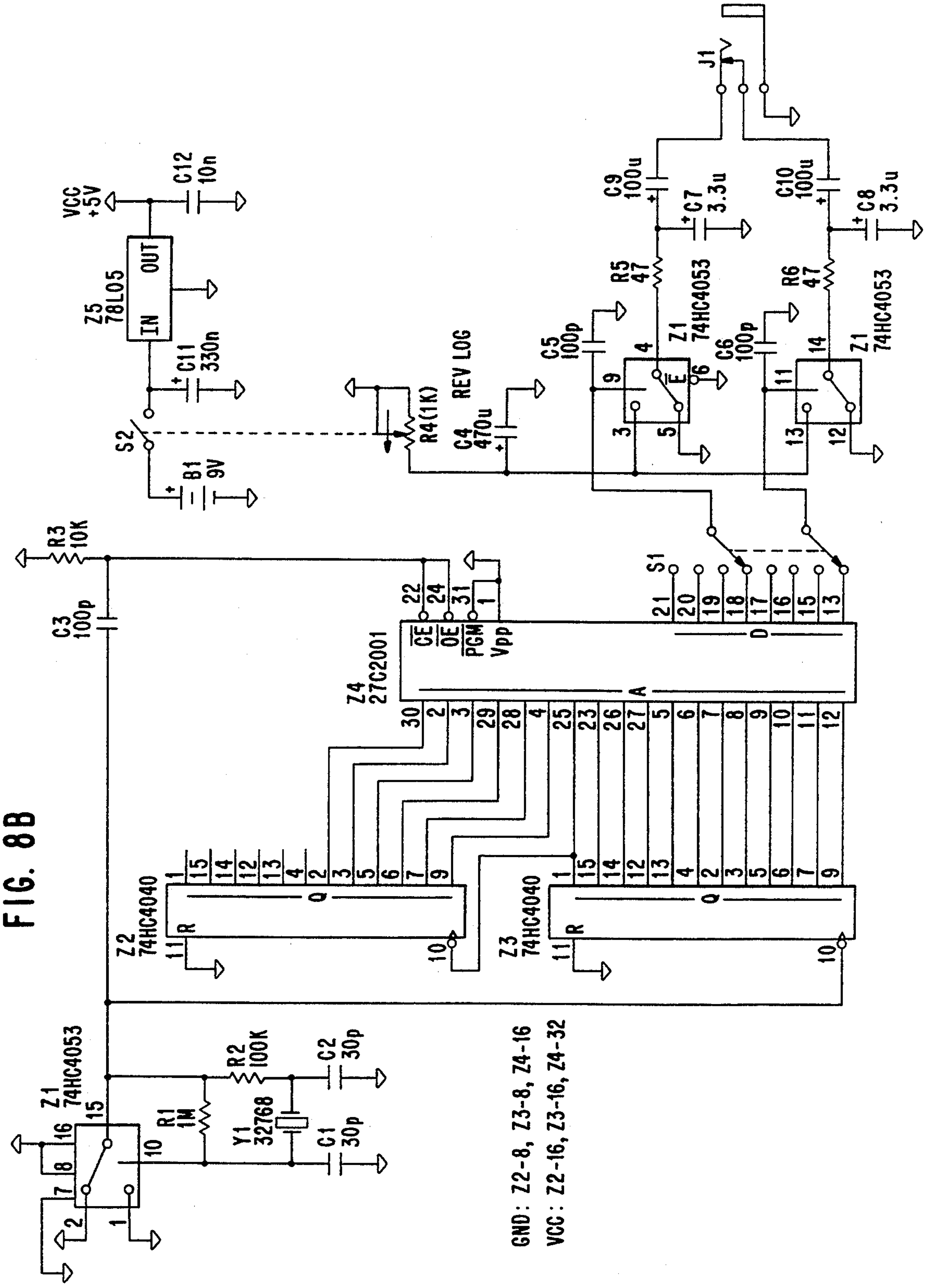
Z1
74HC4053
16
8
7
2
1
15
10
R1
1M
R2
100K
Y1
32768
C2
30p
C1
30p
Z2
74HC4040
R
11
Q
10
1
15
14
12
13
4
2
3
5
6
7
9
Z3
74HC4040
R
11
Q
10
C3
100p
R3
10K
Z4
27C2001
CE
OE
PGM
Vpp
22
24
31
1
A
D
30
2
3
29
28
4
25
23
26
27
5
6
7
8
9
10
11
12
21
20
19
18
17
16
15
13
S1
S2
B1
9V
Z5
78L05
IN
OUT
C11
330n
VCC
+5V
C12
10n
R4(1K)
REV LOG
C4
470u
C5
100p
C6
100p
Z1
74HC4053
9
3
5
4
6
E
11
13
12
14
R5
47
R6
47
C7
3.3u
C8
3.3u
C9
100u
C10
100u
J1

GND: Z2-8, Z3-8, Z4-16

VCC: Z2-16, Z3-16, Z4-32

FIG. 9A

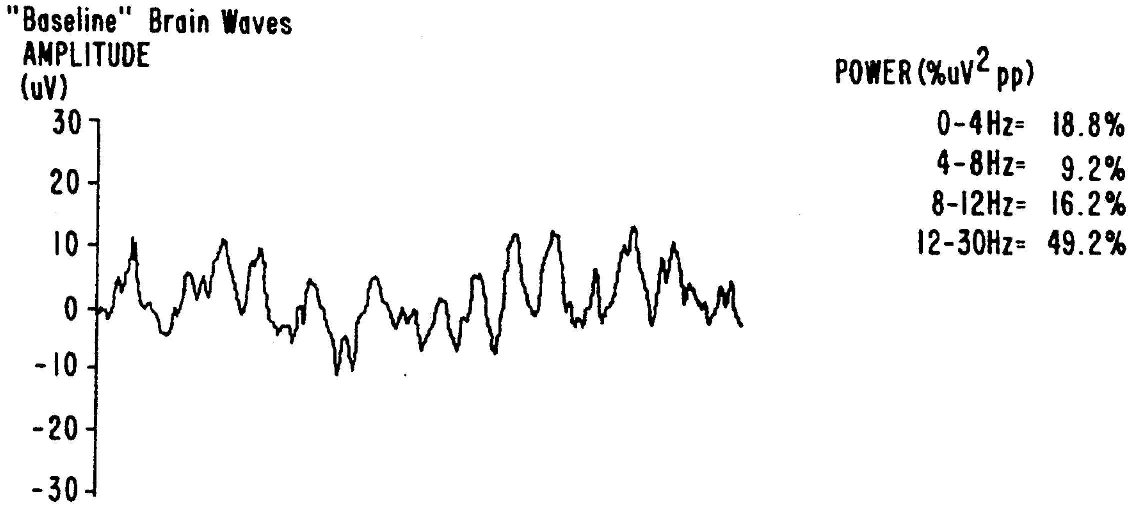
"Baseline" Brain Waves
AMPLITUDE
(uV)
30
20
10
0
-10
-20
-30
POWER (%uV^2 pp)
0-4Hz= 18.8%
4-8Hz= 9.2%
8-12Hz= 16.2%
12-30Hz= 49.2%

FIG. 9B

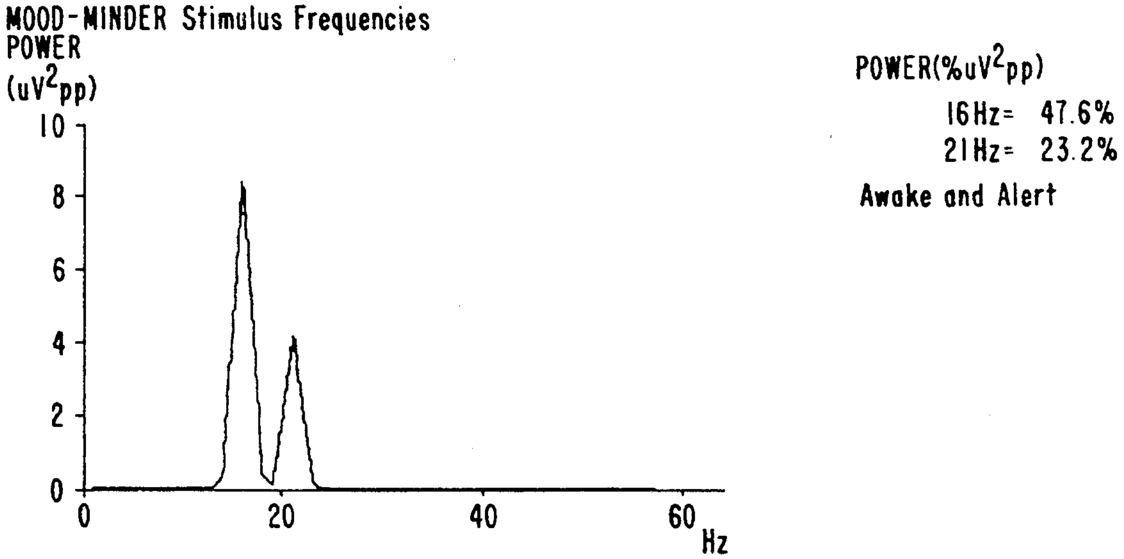

FIG. 9C

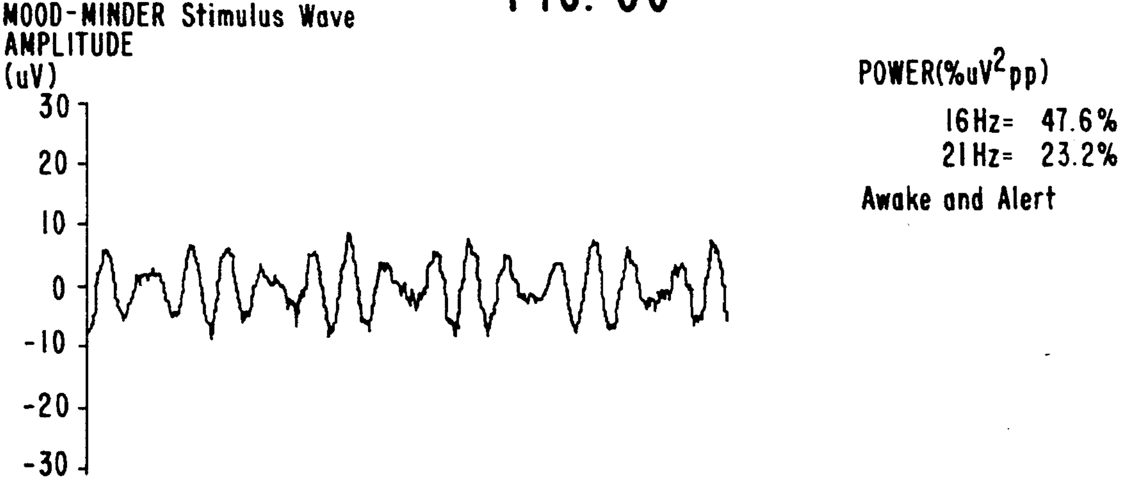

FIG. 9D

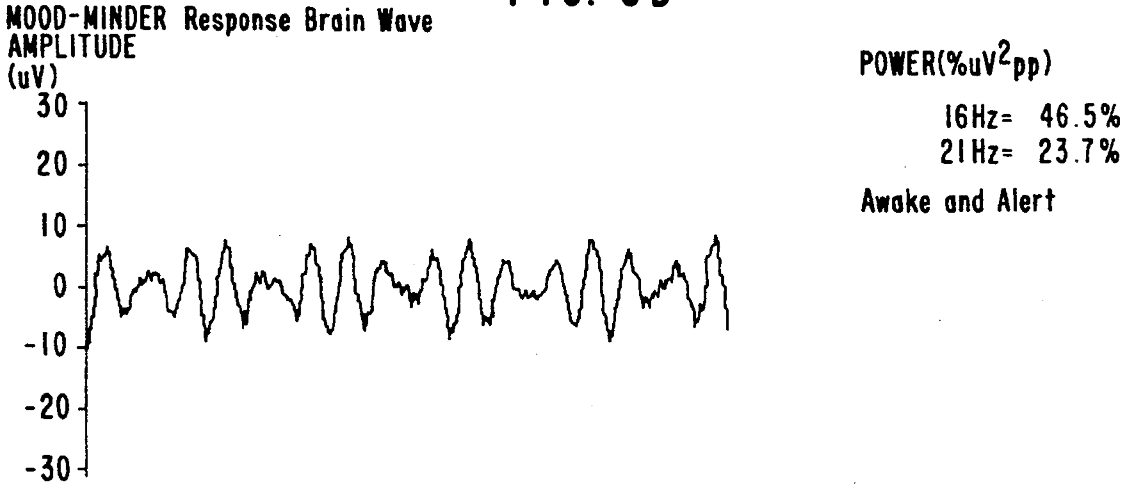

FIG. 9E

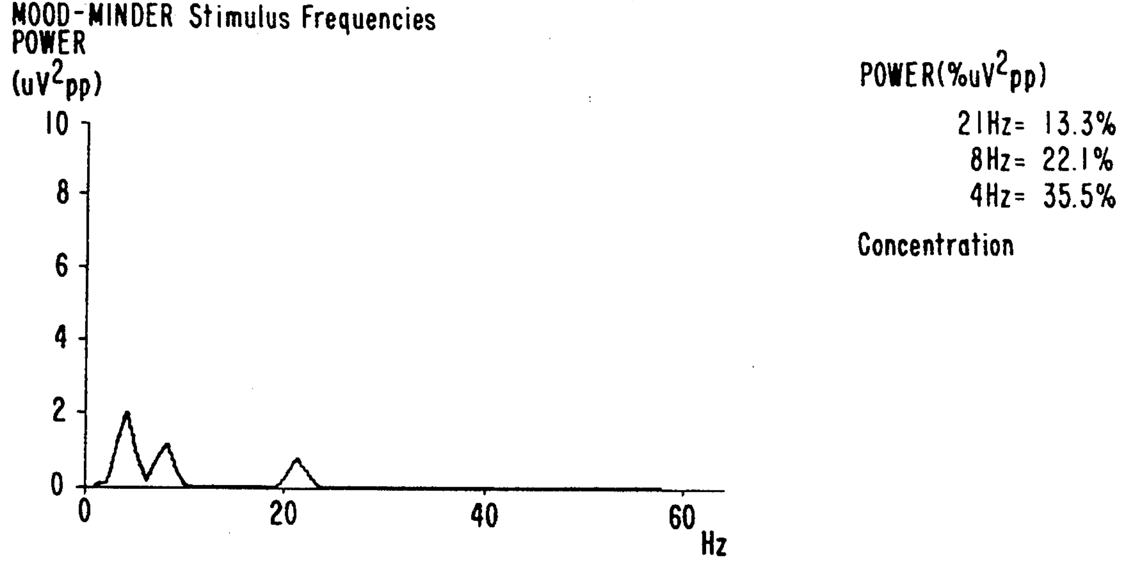

FIG. 9F

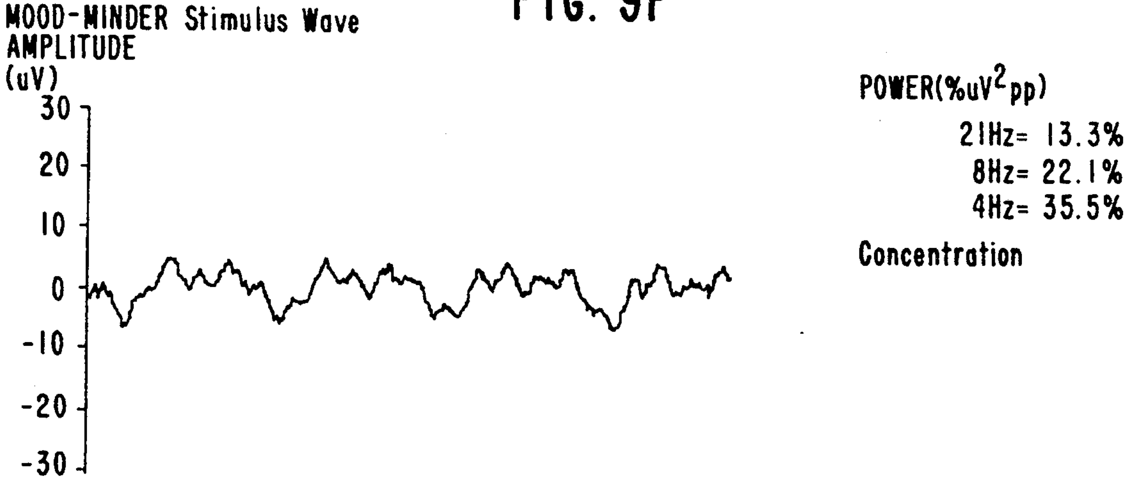

FIG. 9G

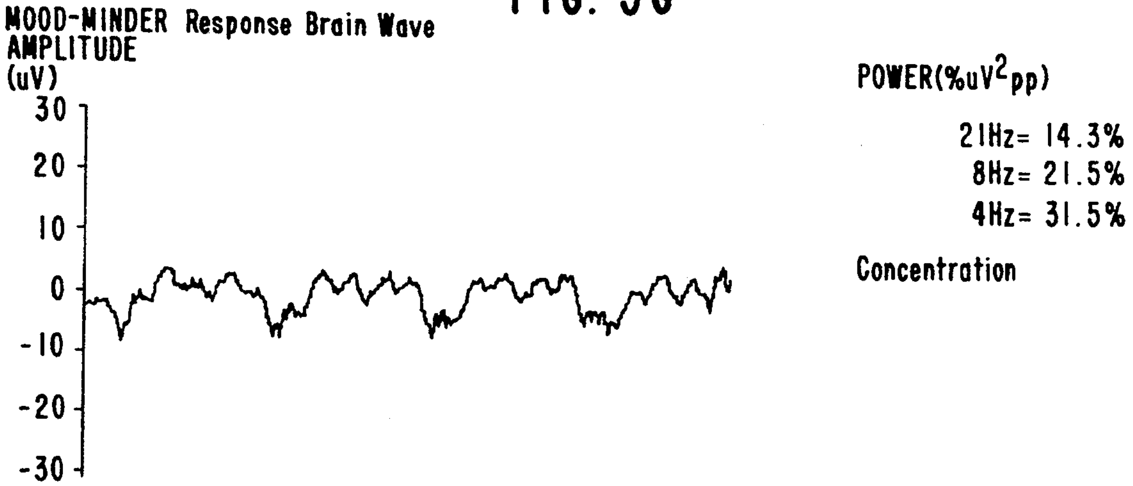

FIG. 9H

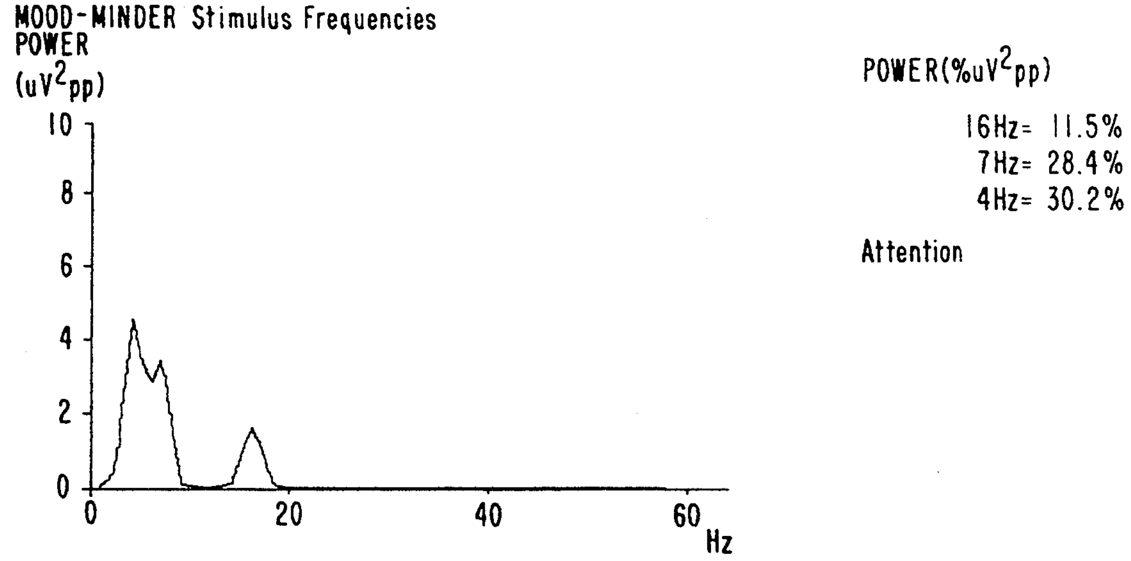

FIG. 9I

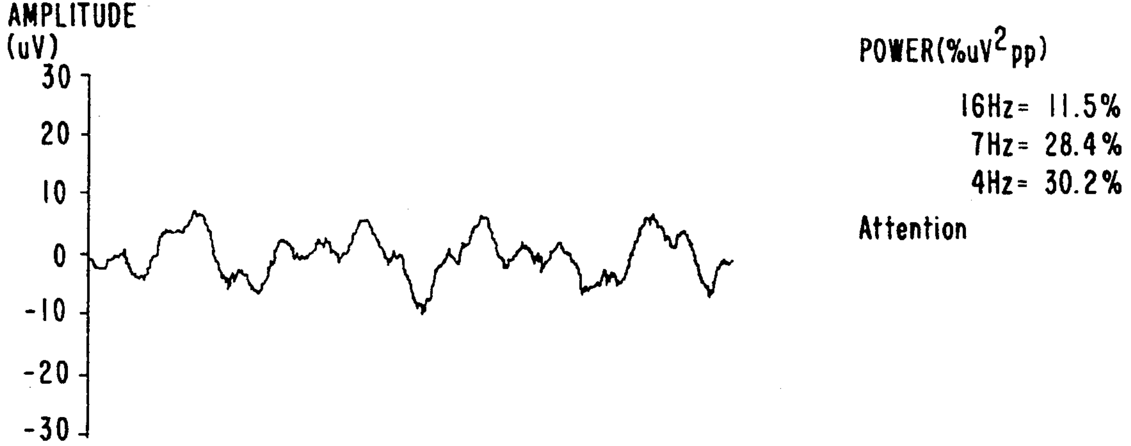

FIG. 9J

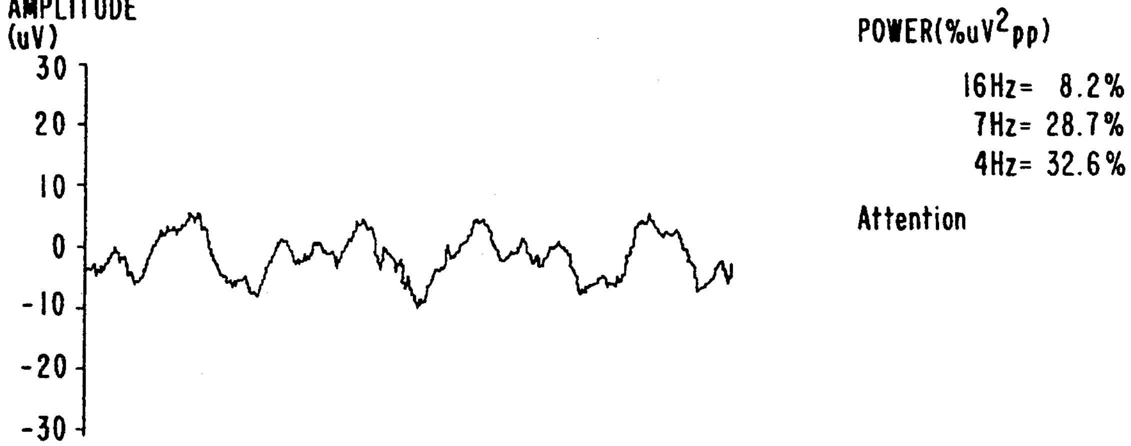

FIG. 9K

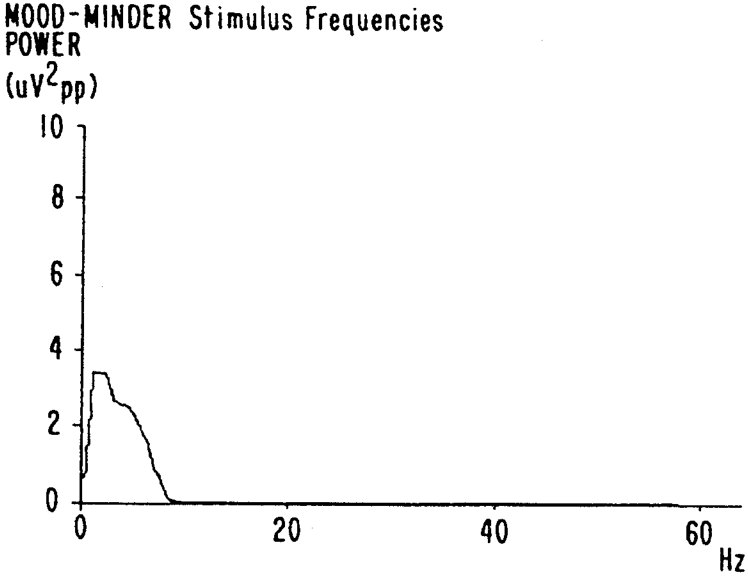

POWER(%uV^2pp)

6Hz= 9.5%
4Hz= 15.4%
1.5Hz= 19.6%

Relaxation

FIG. 9L

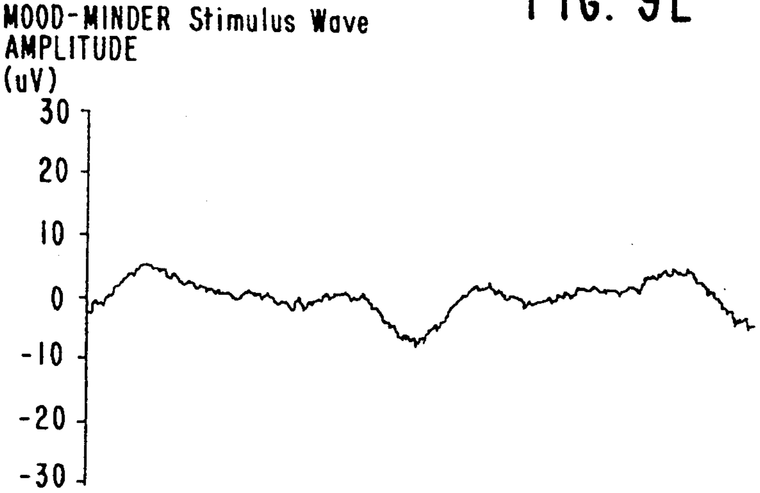

POWER(%uV^2pp)

6Hz= 9.5%
4Hz=15.4%
1.5Hz=19.6%

Relaxation

FIG. 9M

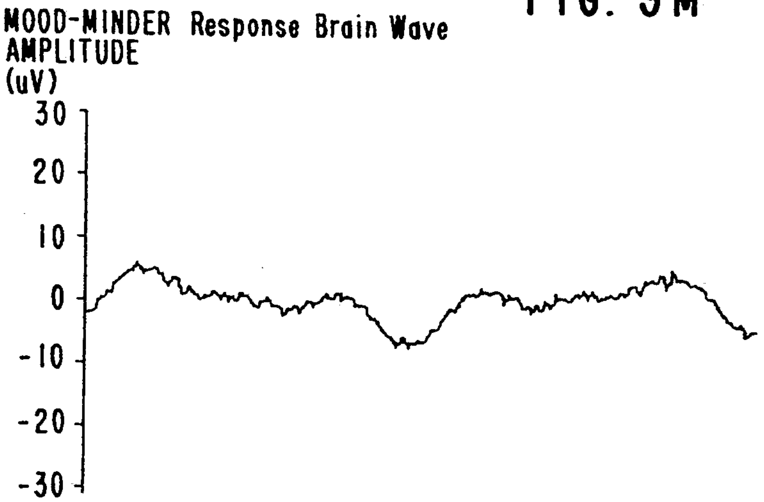

POWER(%uV^2pp)

6Hz= 7.9%
4Hz=16.8%
1.5Hz=20.0%

Relaxation

METHOD OF AND APPARATUS FOR INDUCING DESIRED STATES OF CONSCIOUSNESS

CROSS-REFERENCE TO RELATED APPLICATION

The present application is related to copending application No. 07/514,460, filed Apr. 16, 1990 now U.S. Pat. No. 5,213,562.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to an improved method of inducing desired states of consciousness, including different levels of sleep, in human beings, using a technique known as frequency following response (FFR), developed by the present inventor. The invention also relates to apparatus for performing the method. A number of areas of applicability of the invention are described, in accordance with different preferred embodiments.

2. Description of the Background Art

In a prior patent, U.S. Pat. No. 3,884,218, the present inventor described a method of inducing different levels of sleep, using the FFR technique, in which brain waves could be made to follow superimposed frequency patterns. These frequency patterns were provided as sine waves, at frequencies known to correspond to different levels of sleep, such as alpha (exhibiting brain wave activity in the range of 8–12 Hz), theta (6–8 Hz), and delta (1–4 Hz). EEGs exhibiting frequencies between 12 and 30 Hz (known as a beta range) are characteristic of awake individuals, though beta activity at even higher frequencies has been observed in different types of mental activities. Gamma activity has been characterized as all activity above 30 Hz; until recently, it has not been possible to monitor brain activity in the gamma range. (It should be noted that the boundaries between gamma and beta, beta and alpha, alpha and theta, and theta and delta are somewhat arbitrary; the foregoing delineations are intended to be exemplary and not limiting.)

The present inventor discovered that the human brain could be entrained to output brain wave patterns these different frequencies. While frequencies corresponding to these different levels of sleep are not audible, by superimposing those frequencies on some type of sound, such as music, it was determined to be possible to induce desired levels of sleep. The individual listening to the music would "hear" the low frequencies, with the desired effect on brain activity.

An improvement on the inventor's patented technique, to induce varied states of alertness, is the subject of copending Application No. 07/514,460, the contents of which are hereby incorporated herein by reference. This copending application describes a general FFR technique using what is known as a binaural beat phenomenon, details of which are provided in that application. Briefly, a binaural beat is produced by sending signals at different frequencies (some Hz apart, depending on the desired effect) to an individual's left and right ears. The difference between the frequencies defines the frequency of the binaural beat. Using this technique, the desired frequency can be introduced into the individual's brain activity, inducing the desired state of consciousness.

The induction of FFR in the human brain in this manner results in the synchronization of activity in the hemispheres of the brain. FIG. 1A shows brain activity without FFR, and FIG. 1B shows brain activity with FFR. The inventor has coined the term HEMI-SYNC (for Hemisphere Synchronization) to describe this phenomenon.

The copending application describes a technique wherein, in one form, sine waves having a frequency corresponding to a consciousness state are superimposed on two different carrier frequencies to form two different signals to set up the binaural beat. In another form, an actual brain pattern, based on an electroencephalogram (EEG) waveform indicative of that consciousness state is superimposed on the different carrier frequencies to form two different signals. In use, each signal is provided to one ear of a subject. The difference in carrier frequencies sets up the binaural beat.

Another, more limited application of the binaural beat phenomenon is found in U.S. Pat. No. 4,834,701. In contrast to the narrow range of frequencies discussed in that patent, in the above-mentioned copending application, the applicability of the binaural beat phenomenon is investigated over a much wider range of frequencies, spanning the spectrum of brain activity.

Through additional investigation involving mapping of brain activities of different individuals, the present inventor has discovered some significance to the fact that, while brain waves at certain frequencies are characteristic of different levels of sleep, brain patterns of different individuals still vary. The inventor has investigated possible enhancements to the FFR effect by making it more generic among individuals, yet still more specific to brain activity than a simple sine wave, or an EEG of a particular individual.

Another area of investigation being performed by the present inventor relates to human sleep patterns. Based on current knowledge of human sleep patterns, it appears that sleep is composed of a series of 90-minute cycles. As stated earlier, the beta stage is one of alertness. The first sleep state is alpha, or mental and physical relaxation. The second is theta, or light sleep. Next is delta, or deep sleep. The inventor has investigated the possibility of providing FFR waveforms in cyclic patterns, replicating these human sleep patterns, to facilitate sleep. Another possibility is to take advantage of the cyclic nature of sleep patterns to provide a more gentle wake-up for a sleeper.

In considering the need for alertness during activities such as work, the inventor also considered how it might be possible to introduce FFR waveforms into ambient noise in one's surroundings to facilitate maintenance of desired states of consciousness. Particularly in environments such as factories, or in offices where office equipment puts out consistent types of noise, it would be desirable to be able to introduce a binaural beat into that noise at different frequencies, to enhance the degree of alertness of factory or office workers as desired.

SUMMARY OF THE INVENTION

In view of the foregoing, according to one aspect of the invention, EEGs for a number of individuals in different states of consciousness are sampled, and EEG waveforms for the group of individuals, corresponding to each identifiable state of consciousness, are combined. A binaural beat then is generated using the combined EEGs.

According to this aspect of the invention, it has been determined that using groups of EEG waveforms from different individuals and combining them to obtain a representative waveform yields a waveform that a person's brain is more likely to replicate than an individual EEG waveform, or a sine wave representation of the EEG waveform. The combination may be simple averaging, though other combination techniques, such as weighted averaging, for combining different numbers of EEG waveforms as desired, are contemplated. Now that the inventor has discovered that combinations of EEG waveforms provide a particularly effective entrainment environment, it will be seen that various ways of combining these waveforms may yield greater or lesser effects.

In accordance with another aspect of the invention, a method for replicating cyclic sleep patterns for a desired sleep period is provided. In a preferred embodiment according to this aspect of the invention, a subject is led from beta, to alpha, to theta, to delta, then back to theta, then alpha, then a rapid-eye movement (REM) or light dreaming sleep, in a sequence of 90-minute cycles, during a sleep period of desired duration. After the expiration of the period, the subject may wake up voluntarily. Alternatively, the invention can provide a gentle external stimulus to lead the subject to a beta state.

With respect to this aspect of the invention, an apparatus is provided which automatically leads an individual through these cyclic sleep patterns, and enables the individual to set a desired sleep period. This device preferably takes advantage of the techniques to be described relative to the first-mentioned aspect of the invention, but is not so limited. The inventive contributions of this second aspect of the invention are considered to lie in the combination of hardware itself which generates the desired sequence of binaural beats, as opposed to the particular software which determines the nature of those binaural beats. In one form, the invention is constituted by an alarm clock which provides a fade-in theta-alpha signal followed by a strong beta-gamma signal shortly before a desired wake-up time.

According to yet another embodiment of the invention, selectable mind-affecting sound patterns are provided to supplement constant ambient noise in any environment. When the noise is not present, the patterns are not provided. The patterns vary in amplitude in accordance with changes in the environmental noise.

In accordance with still another embodiment of the invention, a portable system is provided to enable the wearer to introduce binaural beat signals of frequencies that are selectable in accordance with a desired level of awareness. Depending on the level of sophistication of the device, the binaural beat may be generated using the combined EEG waveforms of the first aspect of the invention, but this last aspect of the invention is not so limited.

BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing and other aspects of the invention will be understood by those of working skill in this technological field by reference to the following detailed description of the preferred embodiments of the invention, read in conjunction with the accompanying drawings, wherein:

FIGS. **1A–1C** and **1D–1F** taken from the above-mentioned copending application, show one example of the results which can be achieved using the inventive techniques;

FIG. 2 is a block diagram of the hardware according to a second embodiment of the invention, and FIGS. **3–5** are more detailed schematics therefor;

FIGS. **6A–6J** are drawings, similar to FIGS. **1A** and **1B**, but showing brain activity during various stages of a sleep cycle, using a technique in accordance with the second embodiment of the invention;

FIG. 7 is a block diagram of hardware in accordance with a third embodiment of the invention;

FIG. **8A** is a block diagram of hardware in accordance with a fourth embodiment of the invention, and FIG. **8B** a schematic of that hardware; and

FIGS. **9A–9M** are graphs of different possible effects of the embodiment of FIGS. **8A** and **8B**, showing a baseline brain pattern, selected stimulus frequencies and corresponding stimulus waves, and associated response waves.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The method according to a first preferred embodiment of the invention, which has been developed through extensive experimentation, derives from the empirically-observed phenomenon that brain patterns of human subjects are entrained more readily to brain patterns which more closely match their own. In prior implementations of the FFR technique, such as in the inventor's prior patent, in which sine waves having frequencies corresponding to desired levels of sleep were superimposed upon a given frequency, entrainment did occur. Use of the binaural beat phenomenon yielded better results, through synchronization of the hemispheres of the brain.

However, simple repetitive frequencies, or even combinations of such frequencies within different ranges, do not represent brain patterns per se, but rather provide entrainment environments for the brain to follow. It has been determined that, the more closely the entrainment environment parallels normal brain function at different levels of consciousness, the more effective the entrainment effect. This phenomenon is what led to the improvement disclosed in the above-mentioned copending application.

As a further improvement on that technique, as mentioned above, the present inventor investigated the possibility of creating more generic models of brain function at different levels of consciousness. As a result of that investigation, it was determined that combinations of EEG waveforms from different individuals functioning at the same identifiable level of consciousness (e.g. alpha sleep, theta sleep, or delta sleep) provided a superior entrainment environment. In the inventive method according to this aspect of the invention, the brain patterns of 40 to 50 individuals were combined to yield the entrainment environment.

One area of applicability of the techniques of the present invention is in the area of sleep therapy. Many individuals suffer from sleep disorders to varying degrees. It is possible to provide a suitable entrainment environment, based on known sleep cycles prevalent in humans, to help individuals to regulate their sleep patterns, and thus help to solve their sleep disorders. One embodiment of the invention, shown in FIG. 2 and also in FIGS. 3–5, implements the inventive techniques in what the inventor calls a Sleep Processor to aid in the regulation of human sleep cycles.

In FIG. 2, a read-only memory (ROM) 10 stores frequency sequences corresponding to different parts of a human sleep cycle. The stored frequency sequences may be in accordance with a predetermined algorithm, or alternatively may provide a less complex entrainment environment, such as simple averaging. A digital signal processor (DSP) **20** selects different ones of these sequences based on the current time and the time to which an alarm is set. The time is displayed on display **30**, and is set using time set **40**. The alarm is set to a desired wake-up time using alarm set **50**.

During operation, the DSP **20** accesses the ROM **10** and provides an output to a pulse code modulator unit (PCM) **60** accordingly. The PCM **60** provides an output to each of left and right channel speakers **70, 80** which are provided in close proximity to the ears of a human subject. Using headphones enhances the effect.

Some additional detail of operation of the DSP **20** in one aspect of this embodiment now will be provided. A serial port in the DSP **20** generates an interrupt at a 50 KHz rate. An interrupt handler in the DSP **20** computes the various sounds, in one form, by generating sine waves using a pair of integrators:

$\text{cosine} = \text{cosine} + \text{frequency} \times \text{sine}$

$\text{sine} = \text{sine} - \text{frequency} \times \text{cosine}$

The Sleep Processors needs ten frequencies, five for each channel, and all of these frequencies are generated at the same time. The results are multiplied by ten envelopes, most of which are zero at any moment.

Noise is generated by a well-known 16-bit shift-register algorithm. This algorithm generates a noise signal that repeats every 65535 samples, or about every five seconds. The noise is filtered to sound more like pink or red noise, and less like white noise, and is written into a delay line in RAM. For each channel, the filtered noise is averaged with an earlier sample from the delay line, thus imparting a comb filter response to it.

An additional low-frequency sine/cosine pair is generated, to sweep the comb filter delay. 32-bit arithmetic is used here. The approximate sweep rate is about 1/8 Hz. The low-frequency sine wave is used directly to sweep the delay on one channel. The delay on the other channel is controlled by some mix of the sine and cosine waves. By choosing these and other coefficients properly, any phase and amplitude relationship between the left and right sweep can be obtained. The comb filtered noise for each channel is multiplied by a noise envelope value.

The device is operated as follows. A desired wake-up time is set, much like an alarm clock, and the desired volume is selected. A start/stop button then is pressed to start the cycles for the selected sleep period. Throughout the sleep period, the device repeats a 90 minute cycle of sound that leads the subject through alpha, theta, delta, and back to dreaming sleep. Five minutes before the scheduled wake-up time, a beta signal is introduced to bring the subject back to complete physical wakefulness. When the subject wakes up, he/she hits the start/stop button again to stop the sound sequence.

The sounds produced by the DSP **20** include binaural beat carrier sound patterns utilizing both amplitude and frequency modulation, masking pink sound (a known type of sound described in the copending application), and, optionally, occasional single-word voiced affirmations. The binaural beat audio signals may be in the form of appropriate sine waves, or alternatively may be replicas of actual EEG brain waveforms. In the latter case, either the just-described combined EEG waveforms or a single EEG waveform (as described in the copending application) may be used. The entire pattern of sound and control is generated algorithmically.

One aspect of the effectiveness of the device of FIGS. 2–5 is the spacing of sound carriers at related frequencies so as to engender binaural beat signals not only from channel to channel, but also monaurally, in each audio channel. In this preferred embodiment, three binaural beat frequency signals are created between audio carrier channels, and two amplitude beats per channel also are created, yielding a total of seven beat signals. The inventor has coined the term Septon for this set of beat signals. One example of a septon is as follows:

Left Channel		Right Channel
200 Hz carrier (4 Hz monaural beat)	(4 Hz binaural beat)	204 Hz carrier (4 Hz monaural beat)
204 Hz carrier (4 Hz monaural beat)	(4 Hz binaural beat)	208 Hz carrier (4 Hz monaural beat)
208 Hz carrier	(4 Hz binaural beat)	212 Hz carrier

A standard program according to this preferred embodiment would employ the following sound sequence:

0–5 minutes:
Signal Group A (comprised of replicated EEG waveforms having dominant values in the alpha range)
Signal Group B (15 dB below Group A, generated simultaneously with the sounds of Group A, and comprised of replicated EEG waveforms having dominant values in the theta range)
Phased Pink Sound (six seconds, peak-to-peak, on both left and right channels, 20 dB below Group A)
Voice Inserts (repeated at 40 second intervals, 10 dB below Group A, simultaneously with the other sounds, and comprising short sequences of phrases like "relax" "let go", and "sleep")
5–20 minutes:
Signal Group B
Signal Group C (20 dB below Group B, generated simultaneously with Group B, and comprised of replicated EEG waveforms having dominant values in the delta range)
Phased Pink Sound (15 dB below Group B, having a duration as in the first interval)
Voice Inserts (10 dB below Group B, comprised as above)
20–40 minutes:
Signal Group C
Signal Group D (10 dB below Group C, generated simultaneously with Group C, and comprised of replicated EEG waveforms having dominant values in the lower delta range)
Phased Pink Sound (10 dB below Group C, having a duration as in the first interval)
Voice Inserts (20 dB below Group C, comprised as above) 40–65 minutes:
Signal Group D
Phased Pink Sound (10 dB below Group D, having a duration as in the first interval)
Voice Inserts (20 dB below Group D, comprised as above) 65–80 minutes:
Signal Group C
Signal Group D (10 dB below Group C, generated simultaneously with Group C)
Phased Pink Sound (15 dB below Group C, having a duration as in the first interval)

NO voice inserts

80–90 minutes:

Signal Group B

Signal Group C (10 dB below B, generated simultaneously with Group B)

Phased Pink Sound (15 dB below Group B, having a duration as in the first interval)

NO voice inserts

The foregoing sequence is repeated through the sleep period until the wakeup sequence, approximately five minutes before the set wake-up time:

Signal Group AA (a wakeup sequence, comprising replicated EEG waveforms having dominant values in the beta range, or alternatively a 400 Hz/416 Hz envelope yielding frequencies in the beta range)

Voice inserts (10 dB below Group AA, comprised of short phrases such as "waking up", "refreshed", "bright", and repeated at intervals)

One variation of the foregoing embodiment is an alarm clock which, instead of sounding a loud alarm or other jarring noise at wake-up time, starts a gentle sequence of signals some minutes before, to bring an individual up gently through the various levels of sleep to full wakefulness. A fade-in theta-alpha signal may be provided, followed by a stronger beta-gamma signal.

FIGS. **6A** to **6J** show the effects of the just-described "sleep processor" embodiment. Column **1** shows distribution of delta frequencies; column **2** shows distribution of theta frequencies; and column **3** shows distribution of alpha frequencies. The top row of graphs is the actual pattern observed in the individual, and the bottom row is the baseline pattern.

FIG. **6A** corresponds to a normal waking state. Dominant alpha activity is shown in the occipital area of the brain. In FIG. **6B**, pink noise has been applied, without any beat frequencies. A narrower focus of waking state is shown.

In FIG. **6C**, a signal sequence corresponding to Signal Group A has been applied. Some gain in theta frequencies are seen, with rapid diffusion of alpha frequencies and movement toward the vertex of the head. In FIG. **6D**, a signal sequence corresponding to Signal Group B has been applied. There is further diffusion of alpha frequencies, with some movement of delta and theta activity toward the pre-frontal cortex of the brain.

In FIG. **6E**, a signal sequence corresponding to Signal Group C has been applied. There is rapid diffusion of alpha frequencies, and increased power of theta and delta frequencies. In FIG. **6F**, a signal sequence corresponding to Signal Group D has been applied. Alpha frequencies are diffused further toward the pre-frontal cortex, and there is a marked increase in theta and delta frequencies.

FIG. **6G**, continuing application of Signal Group D frequencies, shows a marked increase in delta activity in the pre-frontal cortex, with a steady decrease in alpha activity at the vertex. In FIG. **6H**, another binaural beat stimulation has been applied, and characteristics of stage **3** and **4** sleep may be observed. In FIG. **6I**, further evidence of the further binaural beat stimulation is observed. Delta is the dominant frequency here. Alpha and theta activity has moved to the prefrontal cortex. Finally, FIG. **6J** shows early awakening activity, with a diffusion of delta activity.

FIG. **7** is a block diagram of hardware in accordance with another embodiment of the invention, having application to the work environment, or anywhere a constant source of noise is present, to allow workers, for example, to maintain a desired state of awareness. The device may contain suitable digital signal processor circuitry, as in the preceding embodiment. One difference is that the operation of the device is keyed to the presence of ambient noise, not to a given time duration or selected sleep period.

The device of FIG. **7** includes a sound generator **200** which, as just mentioned, may comprise a digital signal processor. The generator **200** outputs sound patterns via one or more speakers (left and right channel speakers **210, 220** are shown). A sound level detector **230** detects the level of ambient noise in the room, and provides a signal to the sound generator **200**, or activates a cut-off switch (not shown), to discontinue output of the sound generator **200** when the ambient noise level drops below a predetermined level.

The sound level detector also preferably provides a signal to the sound generator **200** to boost the sound pattern output when the ambient noise level increases, so that the effect of the provision of the sound pattern is commensurate with the noise level in the room. Alternatively, the user may simply adjust the volume manually, using one or more knobs (not shown) on the sound generator **200**.

A timer **240** may be provided to control the duration of provision of the selected sound pattern, or even to change the sound pattern at different times of day by controlling a selector **250** which the user accesses to select a particular sound pattern to be output. The user may select a given sound pattern in the morning, and the timer **240** may change that pattern automatically, based on a need at different times of day for sound patterns providing different states of alertness.

The sound pattern produced by the device of FIG. **7** varies automatically in amplitude in accordance with changes in the ambient noise, and is discontinued when the noise stops. As a result, the sound remains unobtrusive. Depending on the setting, the produced sound pattern can enhance wakefulness, promote relaxation (as, for example, in rest areas in the workplace), reduce anxiety and stress, or focus attention, among other characteristics.

The basic system of FIG. **7** produces and inserts four different sound patterns which are selected manually so as to merge the output into the constant ambient noise. More sound patterns are possible, depending on the desired overall capabilities of the system. Various modifications are possible. For example, a programmable version may be provided, which changes the form of the sound patterns throughout a work day or night, in accordance with the responses desired.

Selectability of patterns may be accomplished differently in a model intended for use in conjunction with a computer system. The computer operator can input a selection, and may vary that selection as desired throughout the work day.

FIG. **8A** is a diagram of a portable embodiment of the invention, for use in providing a desired level of consciousness on an individualized basis. A signal generator **300** preferably includes a digital signal processor **310** and a ROM **320** for storing predetermined signals or sequences of signals which correspond to various desired states of awareness. The signal generator **300** may be a simple tone generator or pair of tone generators which provide outputs to speakers or headphones **330** (such as button-sized headphones) to set up a binaural beat. Output of pink sound or phased pink sound by the generator **300** is desirable to facilitate defocusing of the listener and consequent ability to concentrate on the sounds being produced. A selector **340** enables a user to instruct the signal generator **300** to output signals corresponding to the level of consciousness (e.g. focused concentration, relaxation, alertness) that a user desires.

FIG. **8B** shows a schematic of this embodiment, which the inventor calls a "Mood Minder". This embodiment includes a selector for selecting one of four possible types of signals, corresponding to four respective levels of awareness: awake and alert; concentration; attention; and relaxation. However, the invention is not so limited, as the generator **300** may be capable of producing other possible types of signals. Alternatively, pre-set patterns in the generator **300** may vary when specialized use is required. The key to this embodiment is its portability, enabling the user to carry the device everywhere. The device is battery-operated, and is small enough to fit in an upper coat pocket, for example.

FIGS. **9A–9M** show examples of prestored patterns produced by the generator **300**, and of results achieved in use. FIG. **9A** shows baseline brain waves, with relative power output shown at the right for different frequencies. FIGS. **9B**, **9E**, **9H**, and **9K** show the stimulus frequencies produced for four different respective states of consciousness. FIGS. **9C**, **9F**, **9I**, and **9L** show the stimulus waves corresponding to the superposition of the stimulus frequencies on the baseline wave. FIGS. **9D**, **9G**, **9J**, and **9M** show the results achieved in use. As can be seen, the peak-to-peak amplitudes for the response brain waves correspond closely to those of the stimulus waves.

While the present invention has been described in detail with reference to preferred embodiments, various modifications within the scope and spirit of the invention will be apparent to those of working skill in this technological field. Consequently, the invention should be considered as limited only by the scope of the appended claims.

What is claimed is:

1. A method of inducing desired states of consciousness in human beings, comprising the following steps:

combining a plurality of replicated electroencephalogram (EEG) waveforms, each indicative of a particular desired state of consciousness, to produce a combined EEG waveform;

superimposing said combined EEG waveform on two separate sets of carrier waves using stereo sound;

creating differential beat frequencies between said sets of carrier waves based on said superimposing step; and

providing the resulting signals in audio form to respective ears of a human being, to induce said state of consciousness.

2. A method as claimed in claim **1**, wherein said combining step comprises mathematically averaging said EEG waveforms to produce said combined EEG waveform.

3. A method as claimed in claim **1**, further comprising the step of repeating said combining, superimposing, and creating steps for each of a set of desired states of consciousness, and producing a cycle of sets of resulting audio signals, said providing step comprising providing said cycle of sets of resulting audio signals to respective ears of a human being, to induce each of said desired states of consciousness in cyclic fashion.

4. A method as claimed in claim **3**, wherein said cycle corresponds to human sleep patterns, said desired states of consciousness comprising wakefulness, alpha sleep, delta sleep, and theta sleep.

5. A method as claimed in claim **3**, wherein said cycle corresponds to human sleep patterns, said desired states of consciousness comprising alpha sleep, delta sleep, and theta sleep, said cycle being approximately 90 minutes long.

6. A method as claimed in claim **5**, said method further comprising the steps of providing a plurality of repetitions of said cycle, followed by providing a set of audio signals containing a binaural beat at a frequency indicative of beta consciousness.

7. A method as claimed in claim **1**, wherein said creating step includes the step of combining pink sound with said sets of carder waves by shifting of said pink sound with respect to said combined EEG waveform from one stereo audio channel to another, with cyclic changes in amplitude, frequency, and rate of panning.

8. Apparatus for facilitating sleep in a human subject, comprising:

means for setting a wake-up time to select a desired sleep duration;

means for generating a first sequence of signals in a cycle corresponding to a human sleep pattern, frequencies of said signals in said first sequence being substantially equal to frequencies of human brain patterns at different levels of sleep;

means for repeating said cycle a plurality of times based on the selected wake-up time; and

means for waking up said human subject at the selected wake-up time.

9. Apparatus as claimed in claim **8**, wherein said means for waking up said human subject comprises means for generating a second sequence of signals a predetermined time before the selected wake-up time, frequencies of said signals in said second sequence being substantially equal to frequencies of human brain patterns at or near an awakened state.

10. Apparatus as claimed in claim **9**, wherein said predetermined time is approximately five minutes.

11. Apparatus as claimed in claim **8**, wherein said first sequence of frequencies comprises, in order, alpha frequencies, theta frequencies, delta frequencies, and theta frequencies.

12. Apparatus as claimed in claim **8**, further comprising means for generating phased pink sound in conjunction with said first sequence of frequencies.

13. Apparatus as claimed in claim **8**, wherein said first sequence of signals comprises a plurality of sets of combined brainwaves, each of said sets corresponding to a different level of sleep, said combined brainwaves within a given set being constituted by combined electroencephalogram (EEG) waveforms of a plurality of individuals, taken when said individuals had attained a different respective level of sleep.

14. Apparatus as claimed in claim **13**, wherein said EEG waveforms are mathematically averaged.

15. Apparatus for awakening an individual using brain pattern entrainment, said apparatus comprising:

means for selecting a wake-up time;

means for keeping time; and

means, operative a predetermined period before said wake-up time as determined by said means for keeping time, for producing a first sequence of signals having frequencies in the theta-alpha range, followed by a second sequence of signals having frequencies in the beta-gamma range.

16. Apparatus as claimed in claim 15, wherein said means for producing said first and second sequences of signals comprises means for producing said second sequence of signals at a higher amplitude than said first sequence of signals.

17. Apparatus as claimed in claim 15, wherein said first sequence of signals comprises a plurality of sets of combined brainwaves, each of said sets corresponding to a different level of consciousness, said combined brainwaves within a given set being constituted by combined electroencephalogram (EEG) waveforms of a plurality of individuals, taken when said individuals had attained a different respective level of consciousness.

18. Apparatus as claimed in claim 16, wherein said EEG waveforms are mathematically averaged.

19. Apparatus for inducing a desired state of consciousness, said apparatus comprising:

means for detecting presence of a predetermined level of ambient noise;

means, responsive to said detecting means, for generating signals having frequencies substantially equal to frequencies of human brain patterns when said ambient noise is present; and

means for selecting said signals in accordance with desired human activity in said areas.

20. Apparatus as claimed in claim 19, further comprising timer means, connected to said generating means, for generating said signals for a predetermined time set by said timer means.

21. Apparatus as claimed in claim 19, wherein said timer means is connected to said selecting means to enable selection of different ones of said signals in accordance with desired human activity at different times of day.

22. Apparatus as claimed in claim 19, wherein said generating means comprises means, responsive to said detecting means, for increasing an amplitude of said signals in response to an increase in amplitude of said ambient noise, and for decreasing an amplitude of said signals in response to a decrease in amplitude of said ambient noise.

23. Apparatus as claimed in claim 22, wherein said generating means further comprises means for discontinuing said signals when said ambient noise falls below said predetermined level.

24. Apparatus as claimed in claim 19, wherein said generating means comprises a digital signal processor and a read-only memory (ROM) connected to said digital signal processor, said ROM storing a plurality of sets of signals, each of said sets of signals having frequencies substantially equal to human brain patterns at a desired state of consciousness.

25. Apparatus as claimed in claim 24, wherein each of said sets of signals comprises a plurality of sets of combined brainwaves, each of said sets corresponding to a different level of consciousness, said combined brainwaves within a given set being constituted by combined electroencephalogram (EEG) waveforms of a plurality of individuals, taken when said individuals had attained a different respective state of consciousness.

26. Apparatus as claimed in claim 25, wherein said EEG waveforms are mathematically averaged.

27. Apparatus for awakening an individual using brain pattern entrainment, said apparatus comprising:

means for selecting a wake-up time; and

means, operative a predetermined period before said wake-up time, for producing a first sequence of signals having frequencies in a first predetermined range corresponding to a first state of consciousness, followed by a second sequence of signals having frequencies in a second predetermined range corresponding to a second state of consciousness.

28. Apparatus as claimed in claim 27, wherein said first predetermined range is the theta-alpha range, and said second predetermined range is the beta-gamma range.

* * * * *

US005449989A

United States Patent [19]

Correa et al.

[11] **Patent Number: 5,449,989**

[45] **Date of Patent: Sep. 12, 1995**

[54] **ENERGY CONVERSION SYSTEM**

[76] Inventors: **Paulo N. Correa; Alexandra N. Correa**, both of 42 Rockview Gardens, Concord, Ontario, Canada, L4K 2J6

[21] Appl. No.: **54,111**

[22] Filed: **Apr. 15, 1993**

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 922,863, Jul. 31, 1992, abandoned, and Ser. No. 961,531, Oct. 15, 1992, Pat. No. 5,416,391.

[51] **Int. Cl.**^{6} **H03K 3/37**

[52] **U.S. Cl.** **318/558**; 313/581; 315/111.21; 315/111.31; 315/171; 315/173; 327/533; 327/601

[58] **Field of Search** .... 318/558, 727; 313/306, 313/581; 315/84.51, 111.01, 111.21, 111.31, 160, 171, 173, 187, 188, 193, 200 R, 207; 328/59, 60, 61, 69, 70, 85, 208, 219, 220, 225, 249, 250, 251, 260, 264; 363/116, 117

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,205,162	9/1965	MacLean .
3,471,316	10/1969	Manuel .
3,705,329	12/1972	Vogeli .
3,801,202	4/1974	Breaux .
3,864,640	2/1975	Bennett .
3,878,429	4/1975	Iwata .
4,009,416	2/1977	Lowther .
4,128,788	12/1978	Lowther .
4,194,239	3/1980	Jayaram et al. .
4,443,739	4/1984	Woldring .
4,489,269	12/1984	Edling et al. .
4,527,044	7/1985	Bruel et al. .
4,772,816	9/1988	Spence .
4,896,076	1/1990	Hunter et al. .
5,126,638	6/1992	Dethlefsen .

OTHER PUBLICATIONS

Tanberg, R. "On the Cathode of an Arc Drawn in Vacuum", (1930), Phys. Rev., 35:1080.

Kobel, E. "Pressure & High Vapour Jets at the Cathodes of a Mercury Vacuum Arc", (1930), Phys. Rev., 36:1636.

Aspden, H. (1969) "The Law of Electrodynamics", J. Franklin Inst., 287:179.

Aspden, H. (1983) "Planar Boundaries of the Space--Time Lattice" Lettere Al Nuovo Cimento, vol. 38, No. 7, pp. 243–246.

Aspden, H. (1980) "Physics Unified", Sabberton Publications, pp. 14–17, 42–45, 88–89, 190–193.

Pappas, P. T. (1983) "The Original Ampere Force and Bio-Savart & Lorentz Forces", Il Nuovo Cimento, 76B:189.

Graham, G. M. & Lahoz, D. G. (1980) "Observation of Static Electromagnetic Angular Momentum in Vacuo", Nature, vol. 285, pp. 154 & 155.

Sethlan, J. D. et al., "Anomalous Electron–Ion Energy Transfer in a Relativistic-Electron-Beam-Plasma" Phys. Rev. Letters, vol. 40, No. 7, pp. 451–454 (1978).

Primary Examiner—Bentsu Ro

[57] **ABSTRACT**

An energy conversion device includes a discharge tube which is operated in a pulsed abnormal glow discharge regime in a double ported circuit. A direct current source connected to an input port provides electrical energy to initiate emission pulses, and a current sink in the form of an electrical energy storage or utilization device connected to the output port captures at least a substantial proportion of energy released by collapse of the emission pulses.

23 Claims, 13 Drawing Sheets

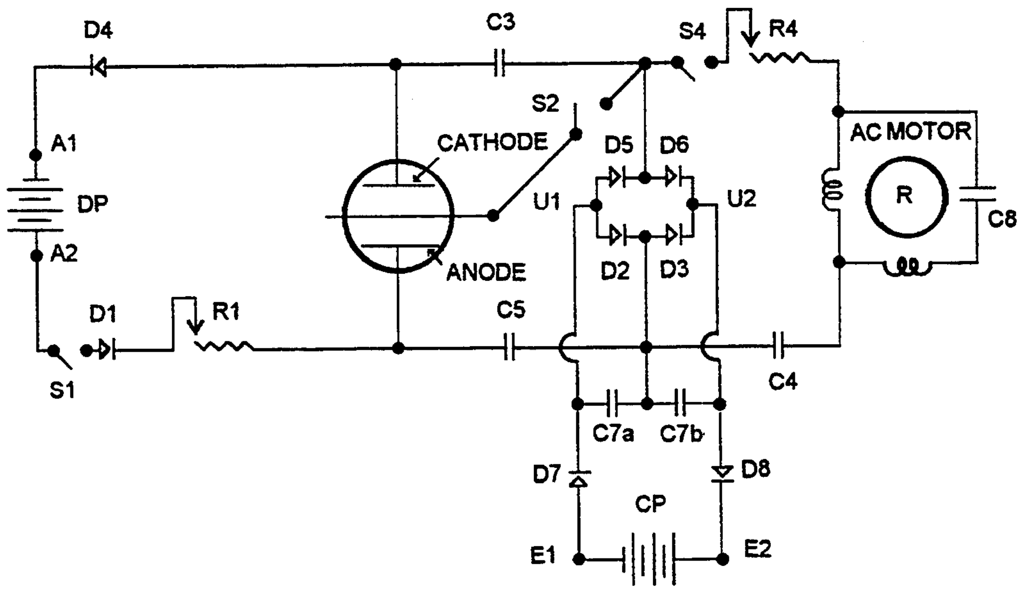

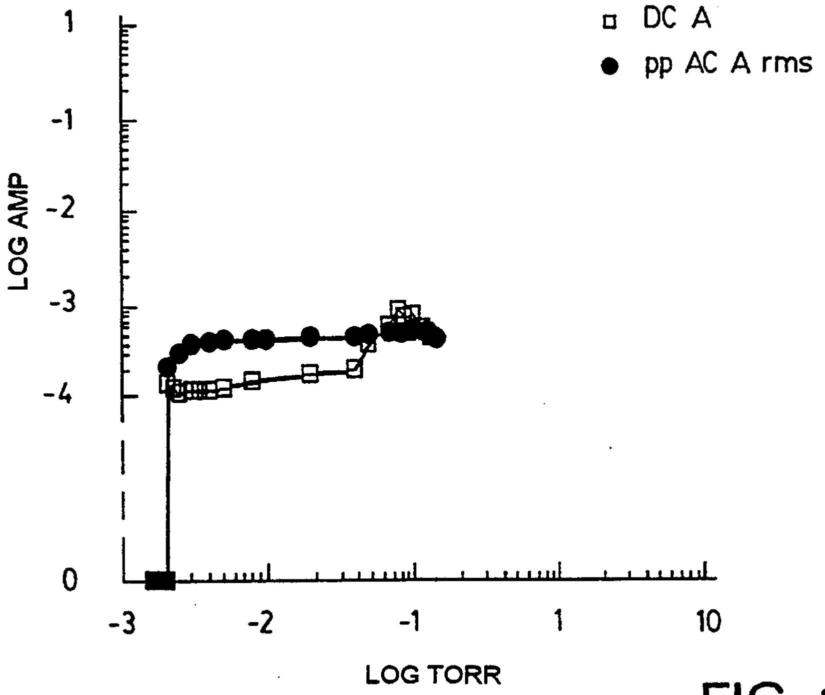

DC A
pp AC A rms
LOG AMP
1
-1
-2
-3
-4
0
-3
-2
-1
1
10
LOG TORR

FIG. 1

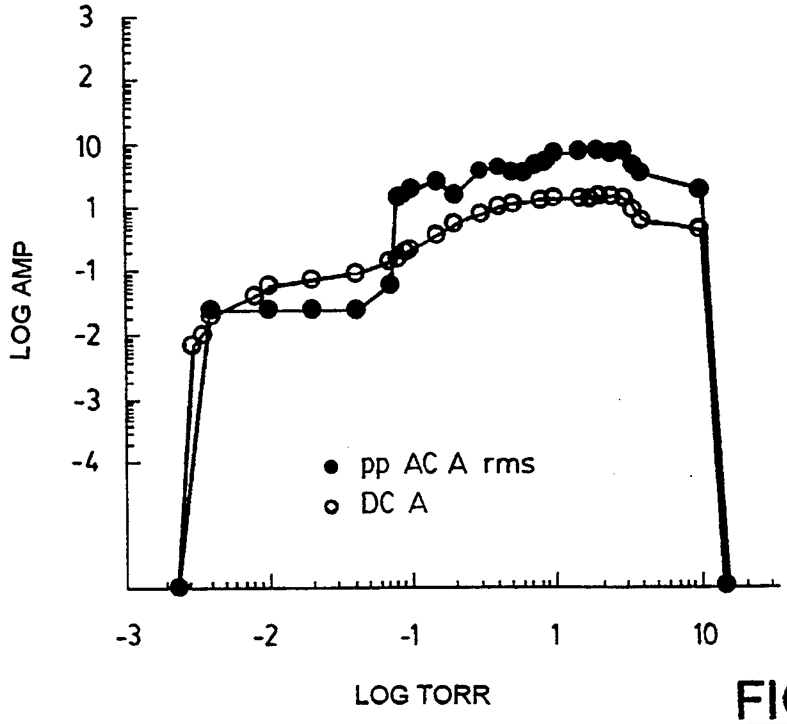

LOG AMP
3
2
10
1
-1
-2
-3
-4
pp AC A rms
DC A
-3
-2
-1
1
10
LOG TORR

FIG. 2

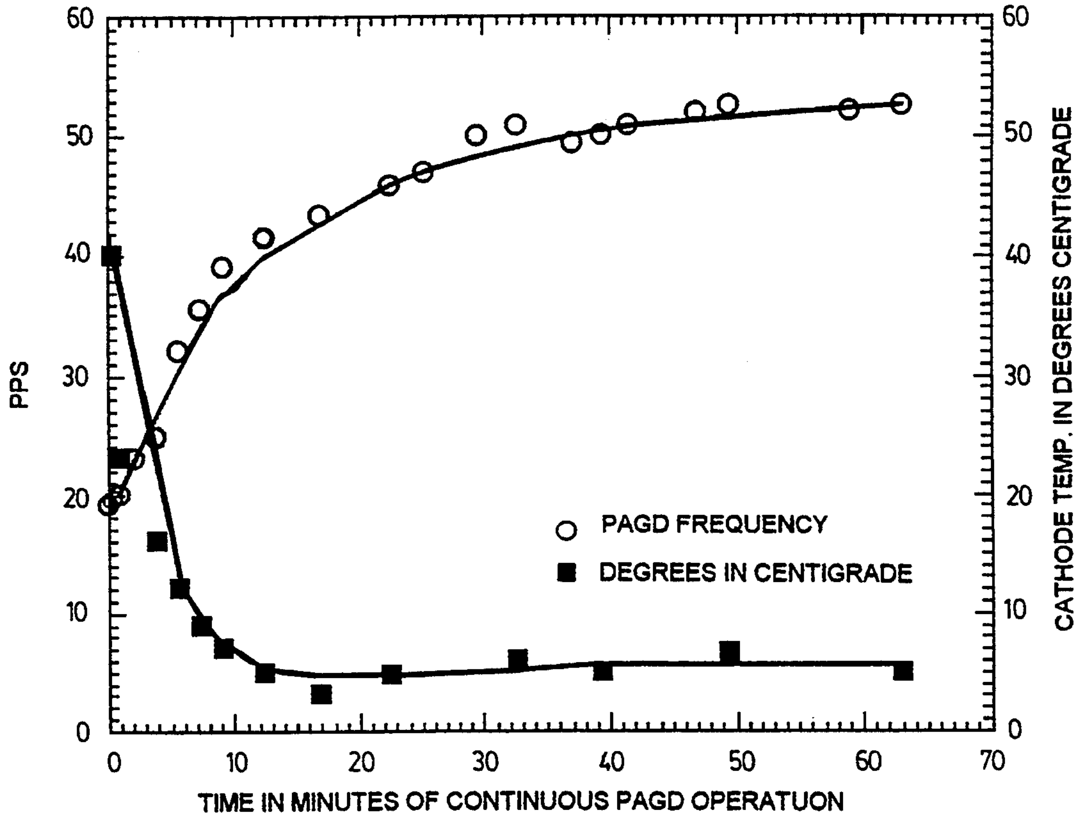
60
50
40
30
20
10
0
PPS
60
50
40
30
20
10
0
CATHODE TEMP. IN DEGREES CENTIGRADE
○ PAGD FREQUENCY
■ DEGREES IN CENTIGRADE
0
10
20
30
40
50
60
70
TIME IN MINUTES OF CONTINUOUS PAGD OPERATUON

FIG. 3

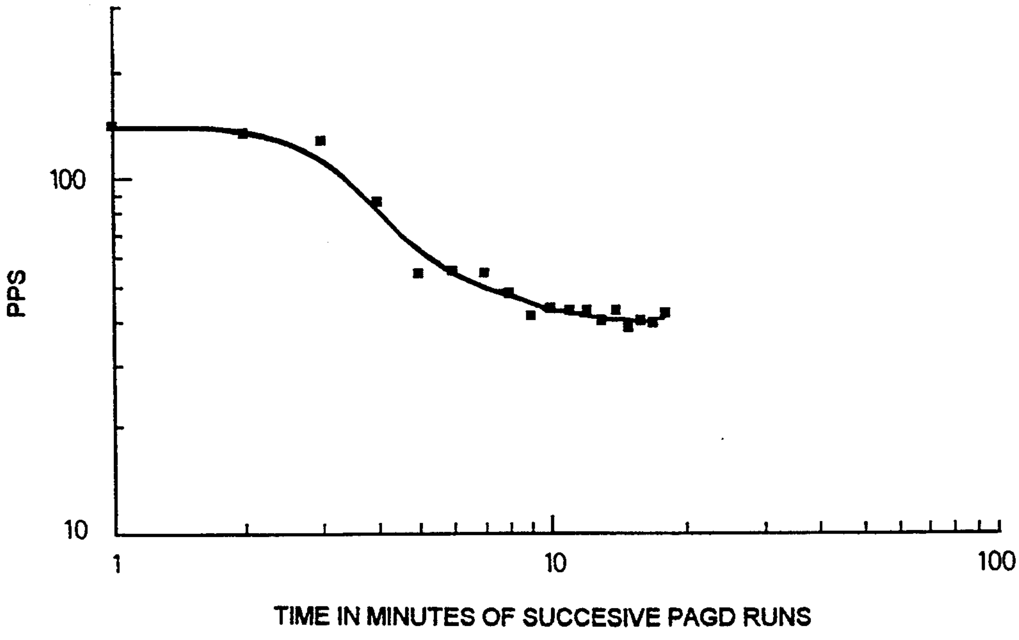
100
10
PPS
1
10
100
TIME IN MINUTES OF SUCCESIVE PAGD RUNS

FIG. 4

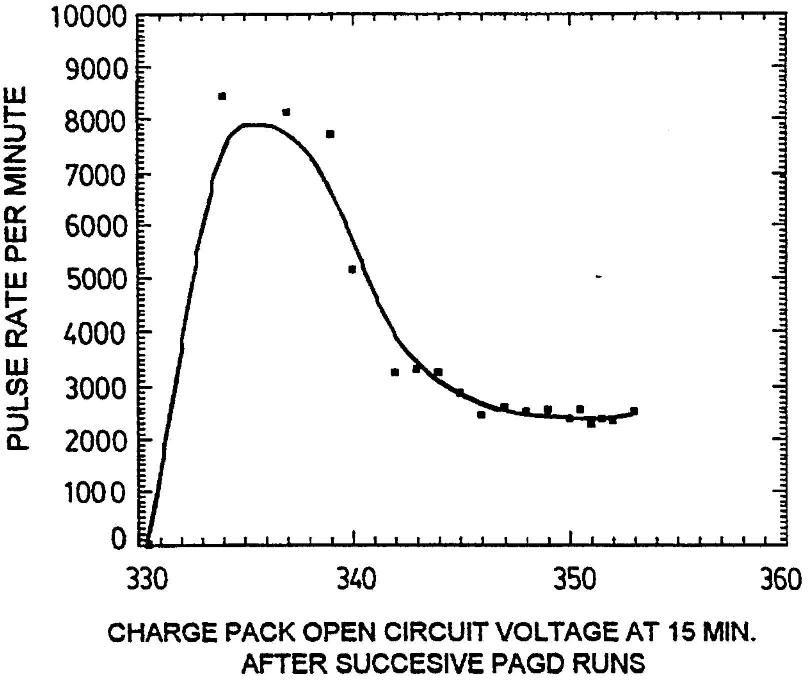
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
PULSE RATE PER MINUTE
330
340
350
360
CHARGE PACK OPEN CIRCUIT VOLTAGE AT 15 MIN.
AFTER SUCCESIVE PAGD RUNS

FIG. 5

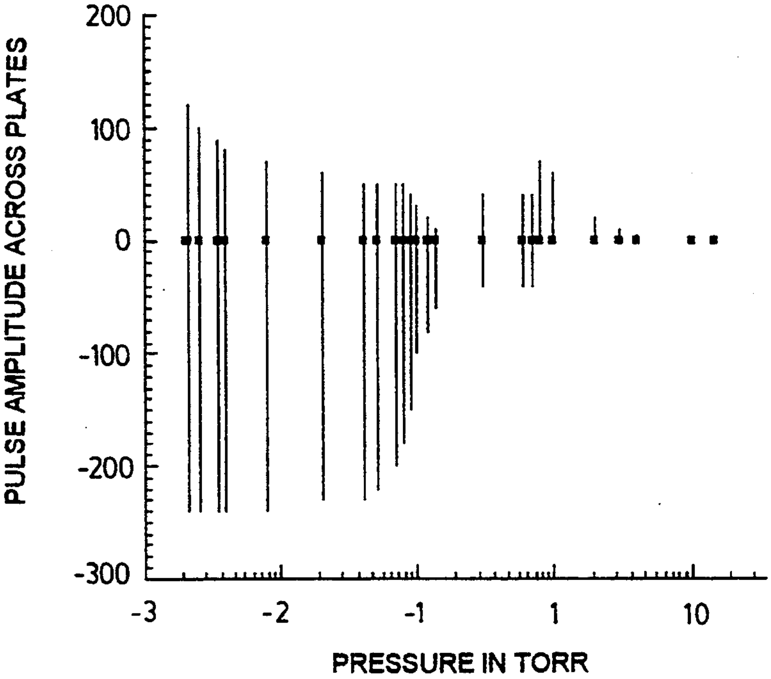
200
100
0
-100
-200
-300
PULSE AMPLITUDE ACROSS PLATES
-3
-2
-1
1
10
PRESSURE IN TORR

FIG. 6

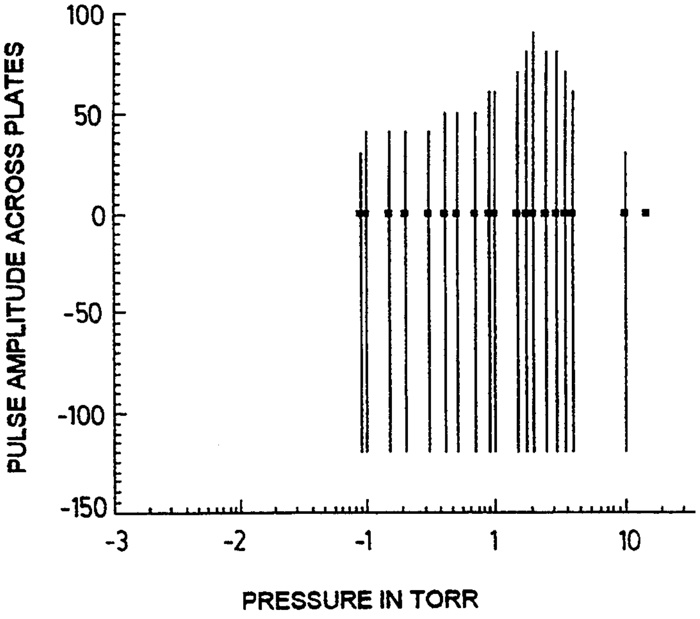
100
50
0
-50
-100
-150
PULSE AMPLITUDE ACROSS PLATES
-3
-2
-1
1
10
PRESSURE IN TORR

FIG. 7

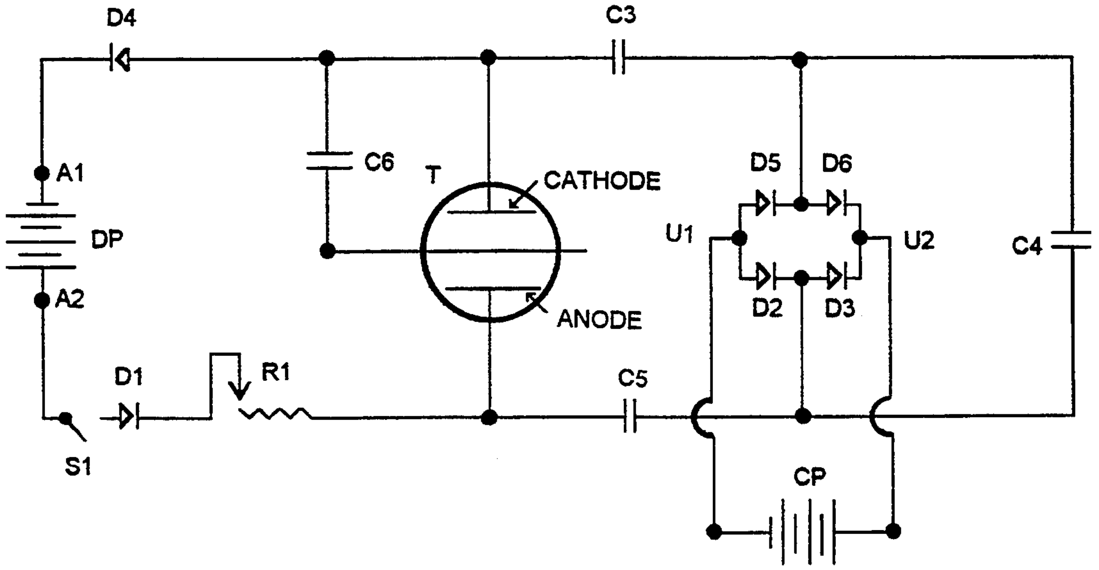
D4
C3
A1
DP
A2
C6
T
CATHODE
ANODE
D5
D6
U1
U2
D2
D3
C4
D1
R1
C5
S1
CP

FIG. 8

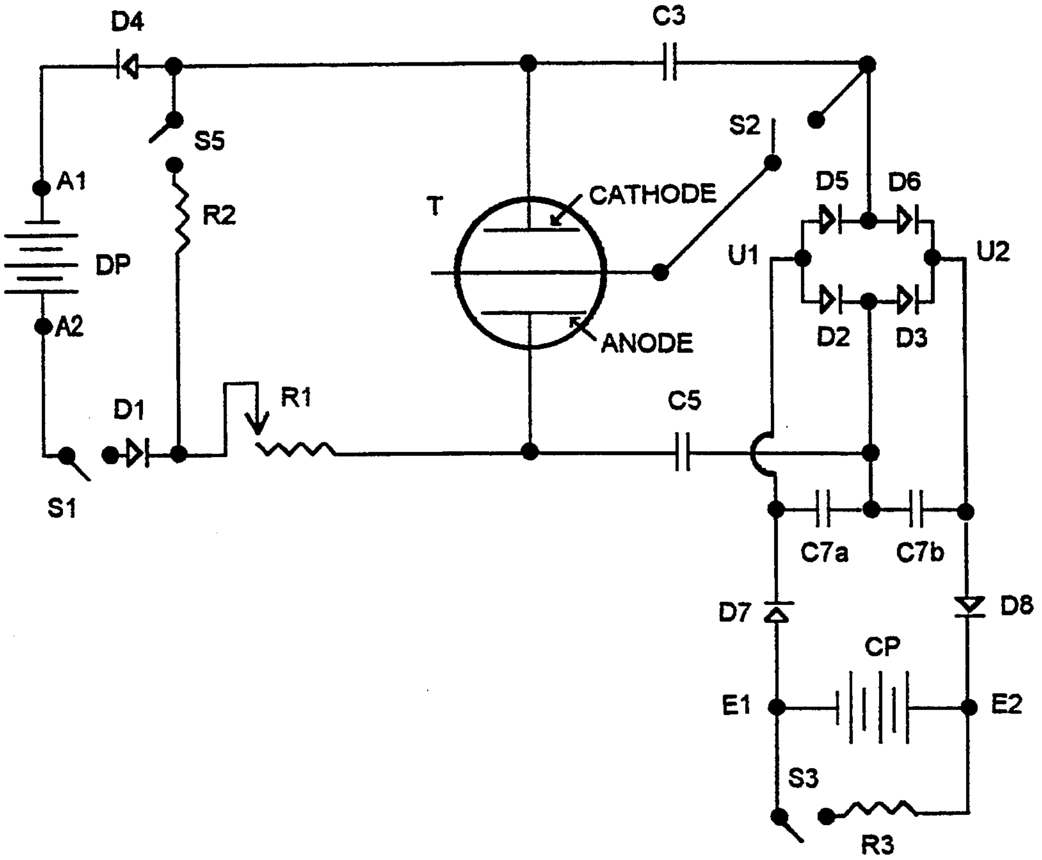
D4
C3
S5
S2
A1
T
CATHODE
D5
D6
R2
U1
U2
DP
A2
ANODE
D2
D3
D1
R1
C5
S1
C7a
C7b
D7
D8
CP
E1
E2
S3
R3

FIG. 9

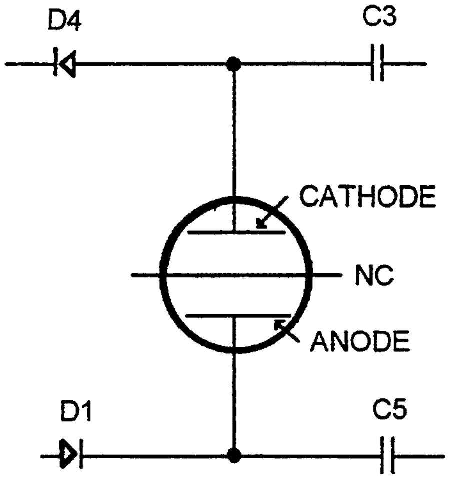
D4
C3
CATHODE
NC
ANODE
D1
C5

FIG. 10A

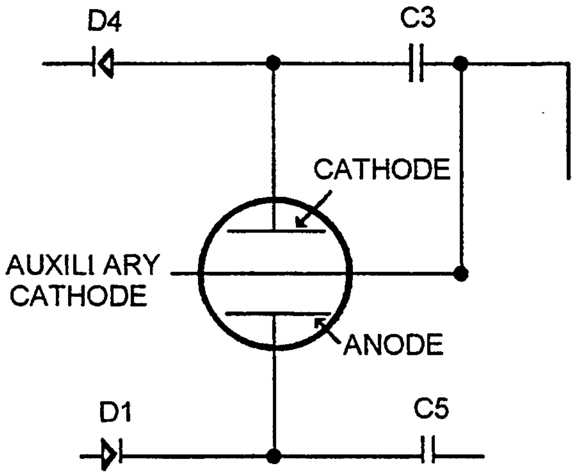
D4
C3
CATHODE
AUXILIARY CATHODE
ANODE
D1
C5

FIG. 10B

FIG. 10C

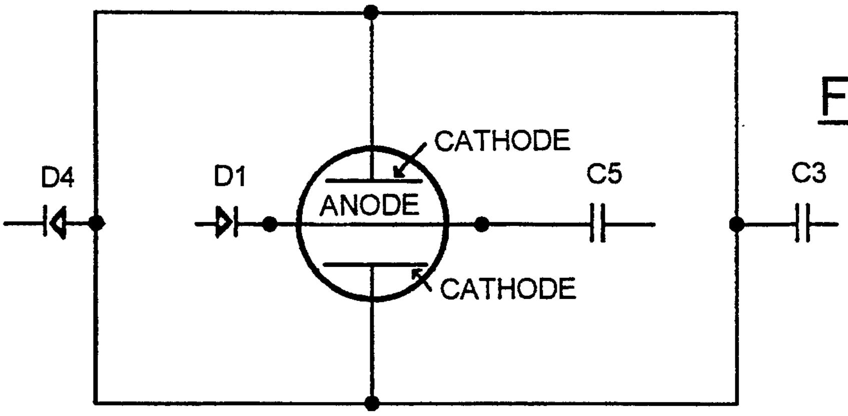
CATHODE
D4
D1
ANODE
C5
C3
CATHODE

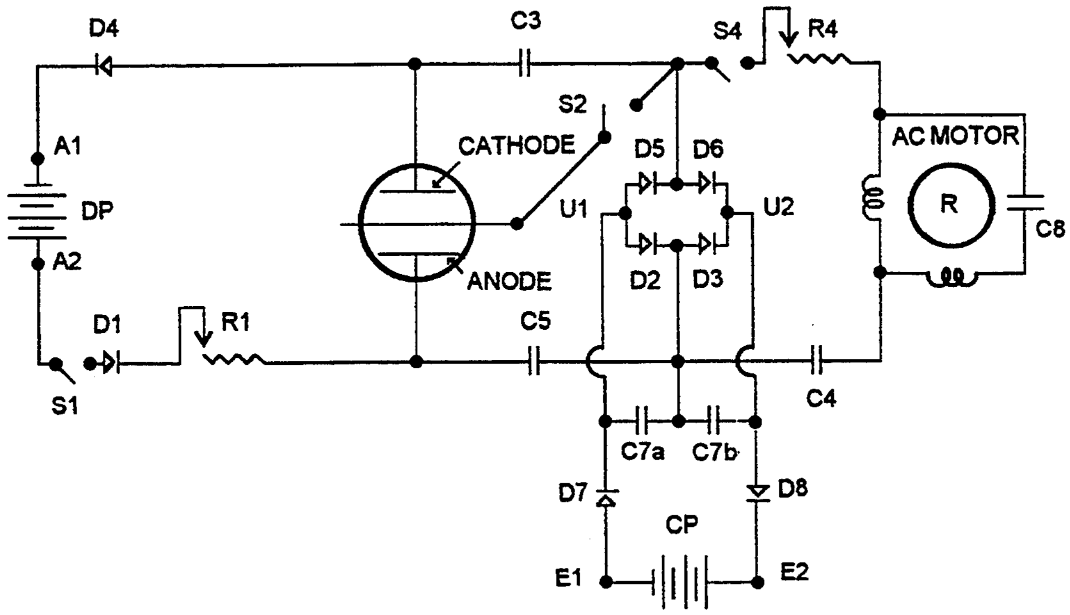
D4
C3
S4
R4
S2
A1
CATHODE
D5
D6
AC MOTOR
DP
U1
U2
R
C8
A2
ANODE
D2
D3
D1
R1
C5
S1
C4
C7a
C7b
D7
D8
CP
E1
E2

FIG. 11

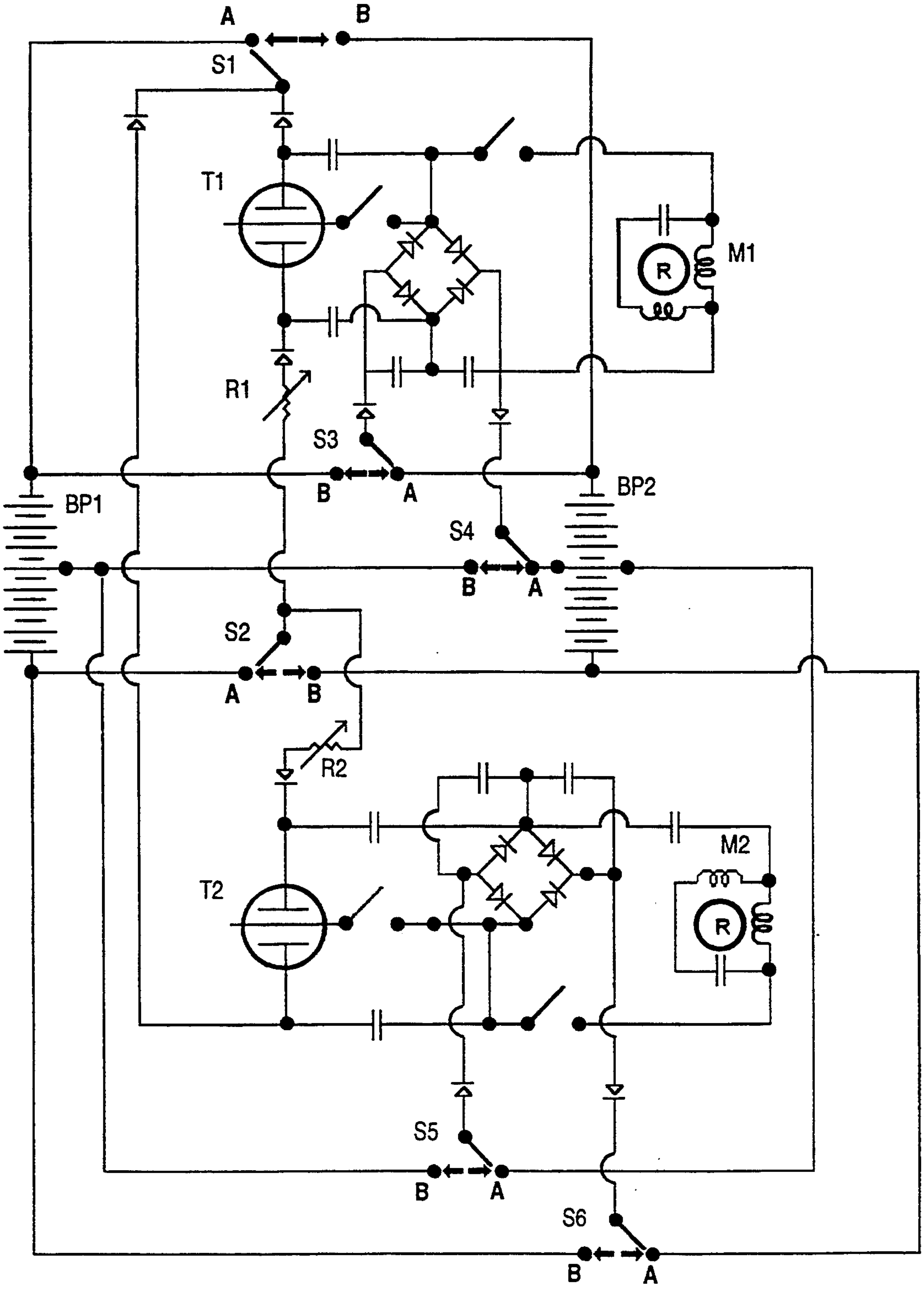
A
B
S1
T1
M1
R
R1
S3
B
A
BP1
BP2
S4
B
A
S2
A
B
R2
M2
T2
R
S5
B
A
S6
B
A

FIG. 12

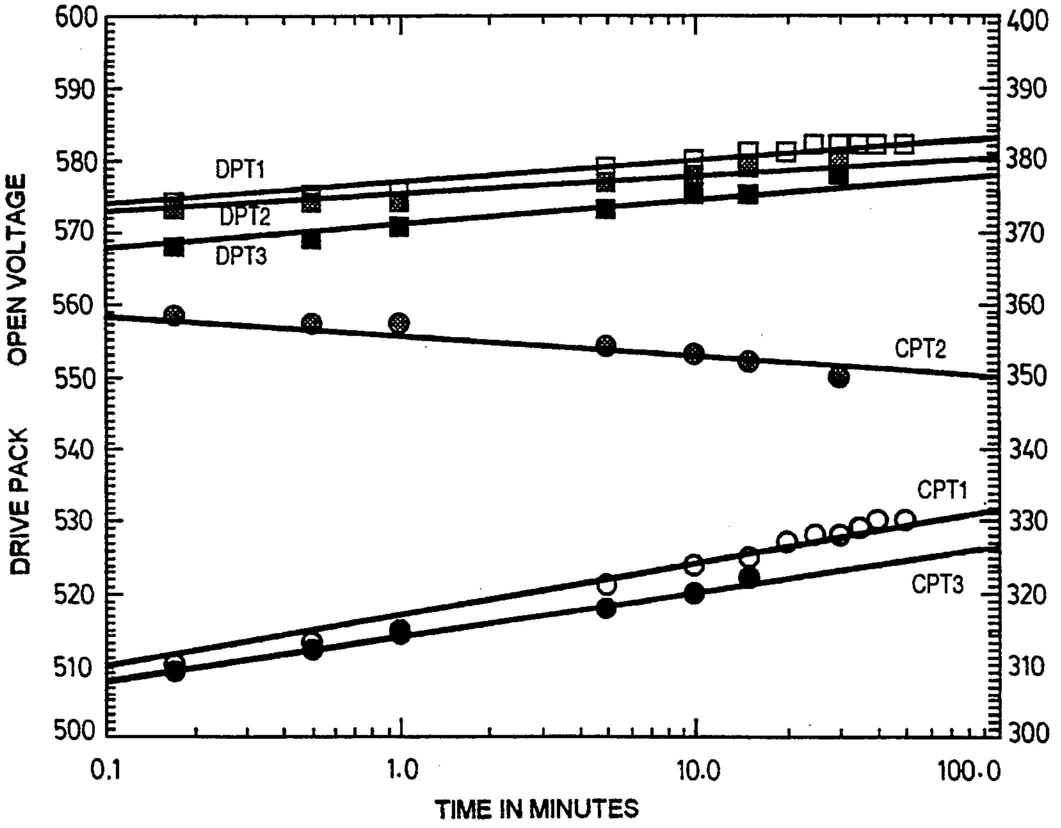
600
590
580
570
560
550
540
530
520
510
500
400
390
380
370
360
350
340
330
320
310
300
DPT1
DPT2
DPT3
CPT2
CPT1
CPT3
DRIVE PACK OPEN VOLTAGE
CHARGE PACK OPEN VOLTAGE
0.1
1.0
10.0
100.0
TIME IN MINUTES

FIG. 13

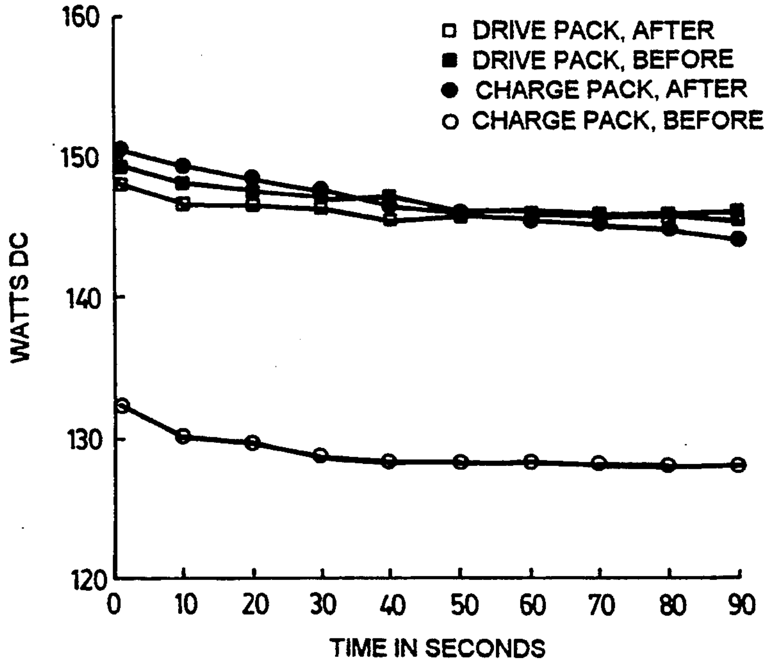
160
150
140
130
120
WATTS DC
□ DRIVE PACK, AFTER
■ DRIVE PACK, BEFORE
● CHARGE PACK, AFTER
○ CHARGE PACK, BEFORE
0
10
20
30
40
50
60
70
80
90
TIME IN SECONDS

FIG. 14

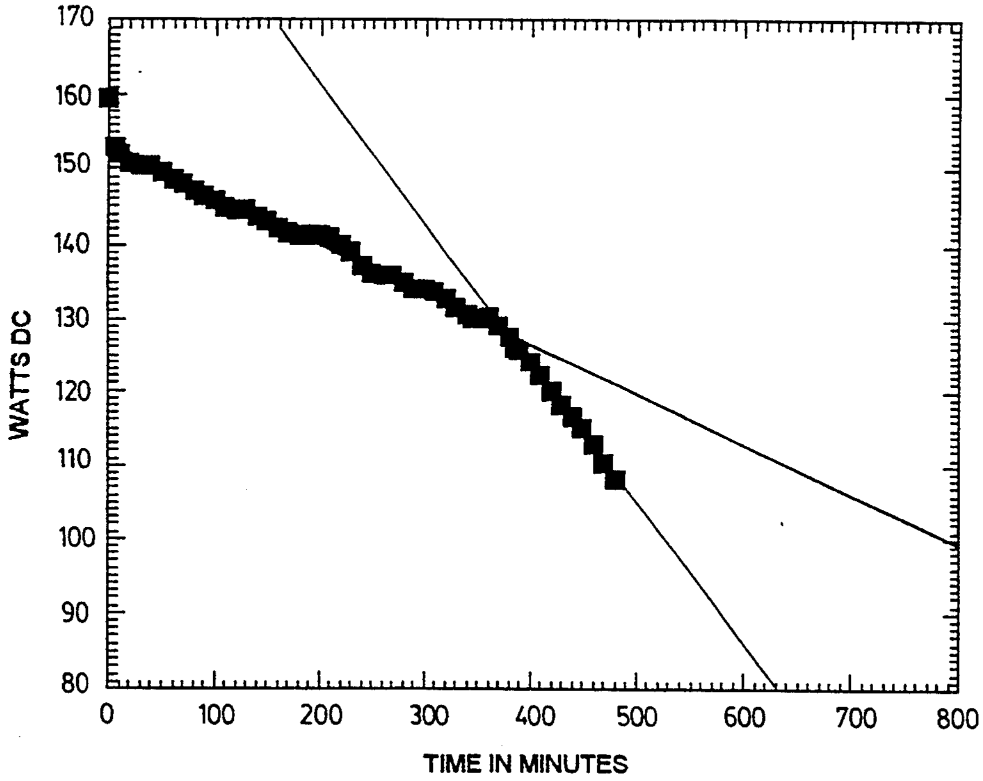

FIG. 15A

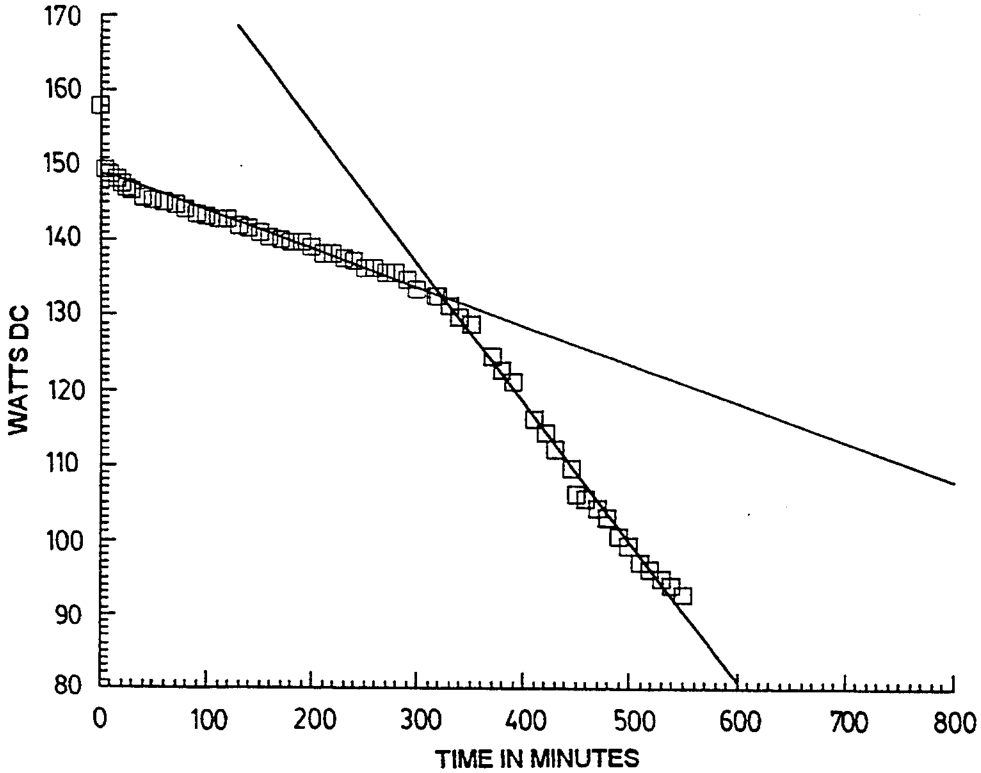

FIG. 15B

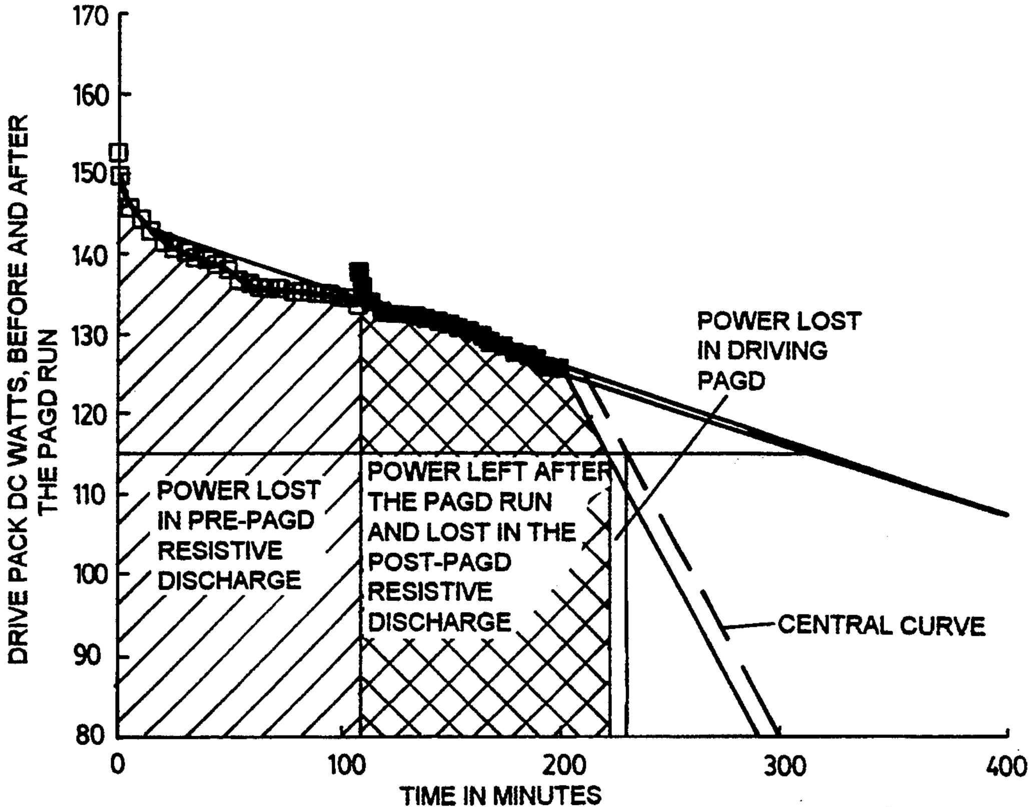
DRIVE PACK DC WATTS, BEFORE AND AFTER THE PAGD RUN
170
160
150
140
130
120
110
100
90
80
0
100
200
300
400
TIME IN MINUTES
POWER LOST IN DRIVING PAGD
POWER LOST IN PRE-PAGD RESISTIVE DISCHARGE
POWER LEFT AFTER THE PAGD RUN AND LOST IN THE POST-PAGD RESISTIVE DISCHARGE
CENTRAL CURVE

FIG. 16

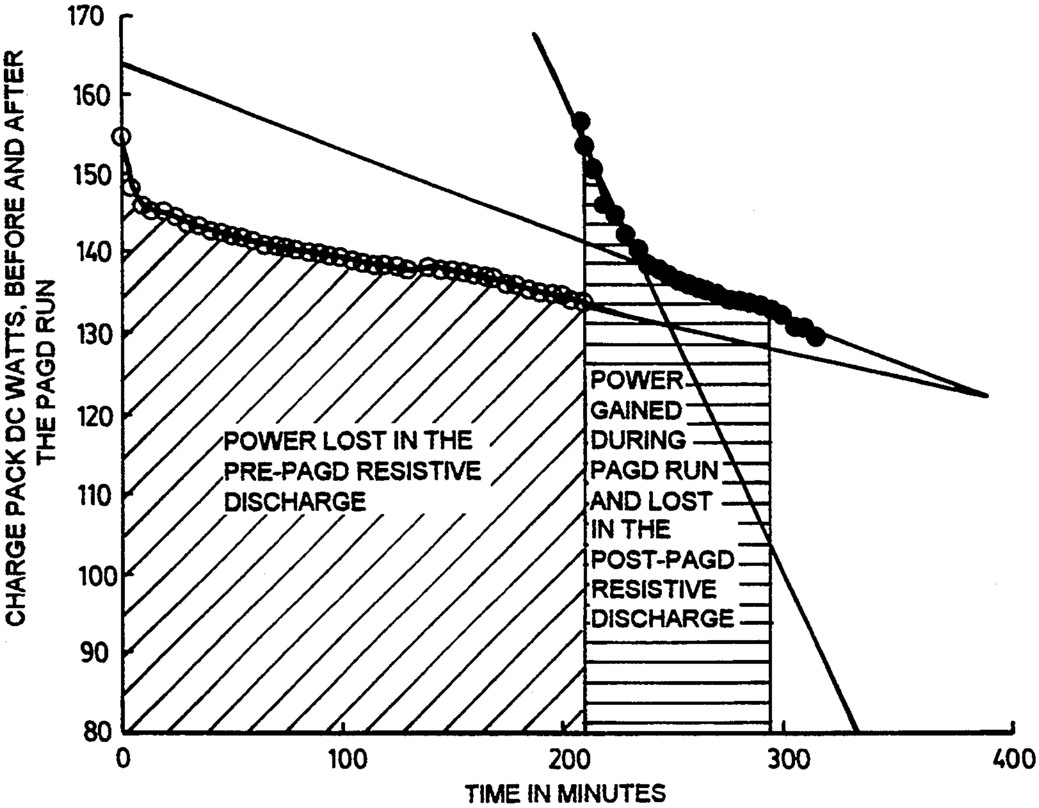
CHARGE PACK DC WATTS, BEFORE AND AFTER THE PAGD RUN
170
160
150
140
130
120
110
100
90
80
0
100
200
300
400
TIME IN MINUTES
POWER LOST IN THE PRE-PAGD RESISTIVE DISCHARGE
POWER GAINED DURING PAGD RUN AND LOST IN THE POST-PAGD RESISTIVE DISCHARGE

FIG. 17

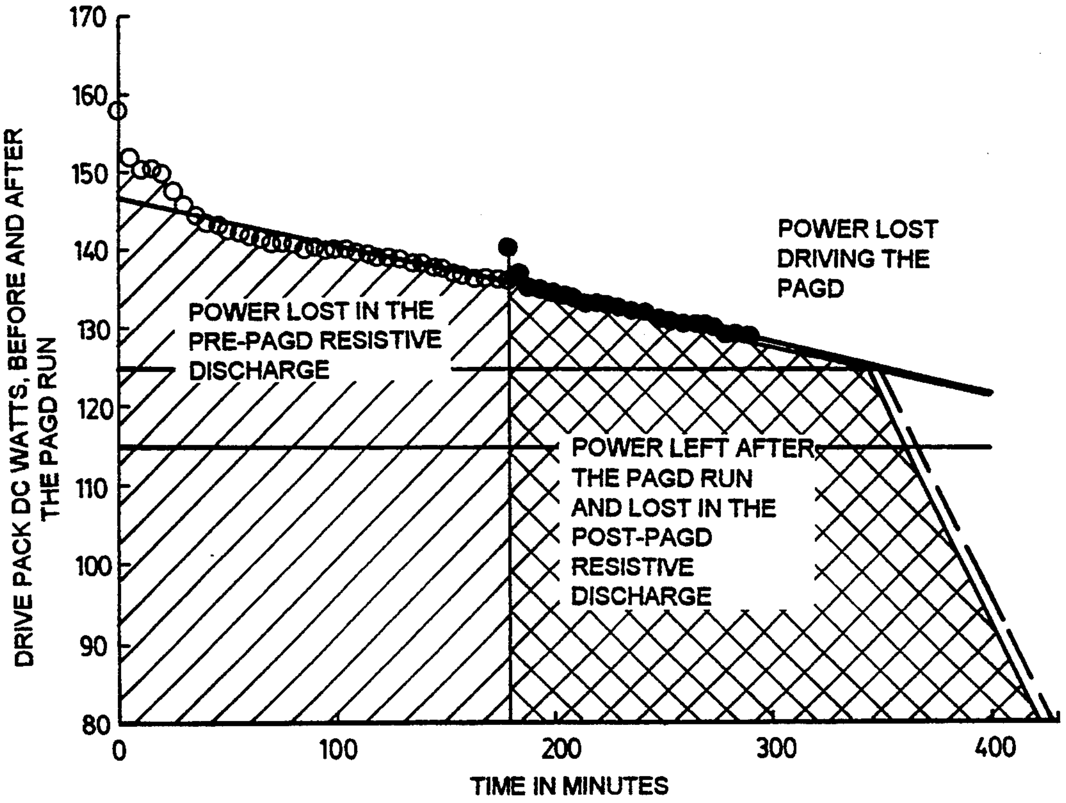
DRIVE PACK DC WATTS, BEFORE AND AFTER THE PAGD RUN
170
160
150
140
130
120
110
100
90
80
0
100
200
300
400
TIME IN MINUTES
POWER LOST DRIVING THE PAGD
POWER LOST IN THE PRE-PAGD RESISTIVE DISCHARGE
POWER LEFT AFTER THE PAGD RUN AND LOST IN THE POST-PAGD RESISTIVE DISCHARGE

FIG. 18

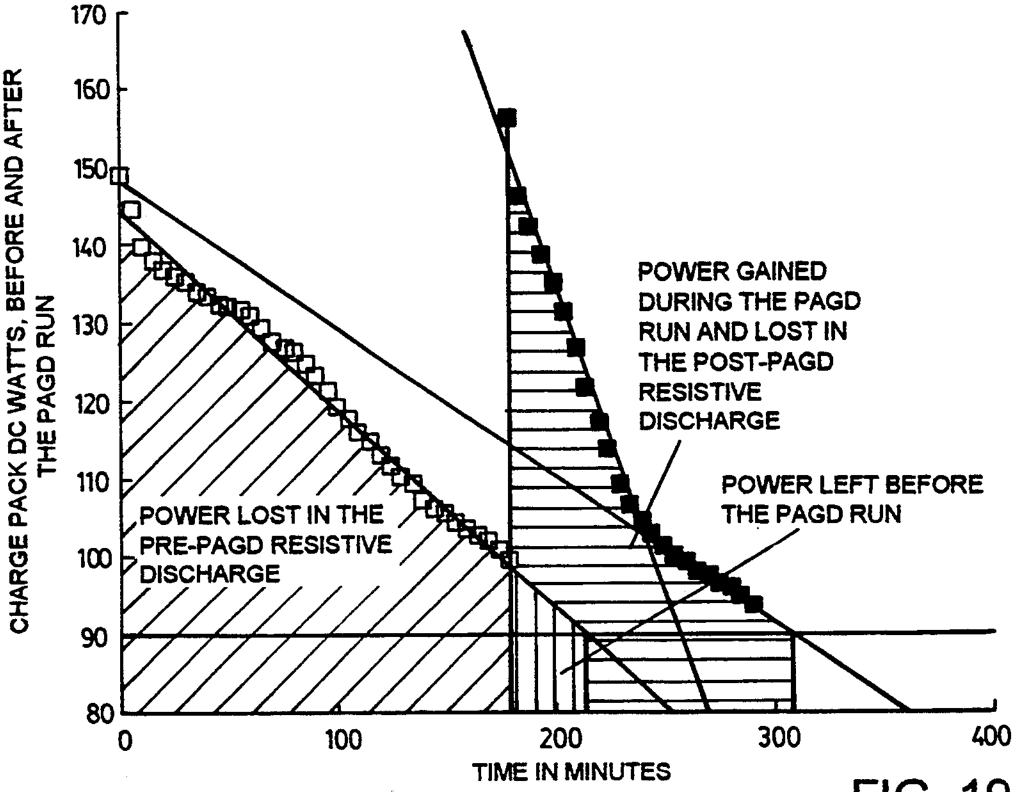
CHARGE PACK DC WATTS, BEFORE AND AFTER THE PAGD RUN
170
160
150
140
130
120
110
100
90
80
0
100
200
300
400
TIME IN MINUTES
POWER GAINED DURING THE PAGD RUN AND LOST IN THE POST-PAGD RESISTIVE DISCHARGE
POWER LEFT BEFORE THE PAGD RUN
POWER LOST IN THE PRE-PAGD RESISTIVE DISCHARGE

FIG. 19

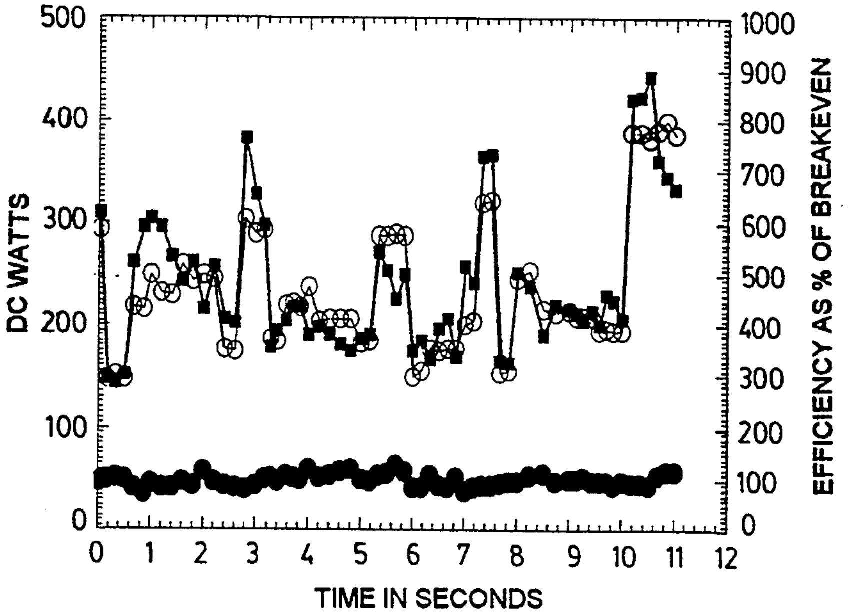
DC WATTS
500
400
300
200
100
0
EFFICIENCY AS % OF BREAKEVEN
1000
900
800
700
600
500
400
300
200
100
0
0 1 2 3 4 5 6 7 8 9 10 11 12
TIME IN SECONDS

FIG. 20

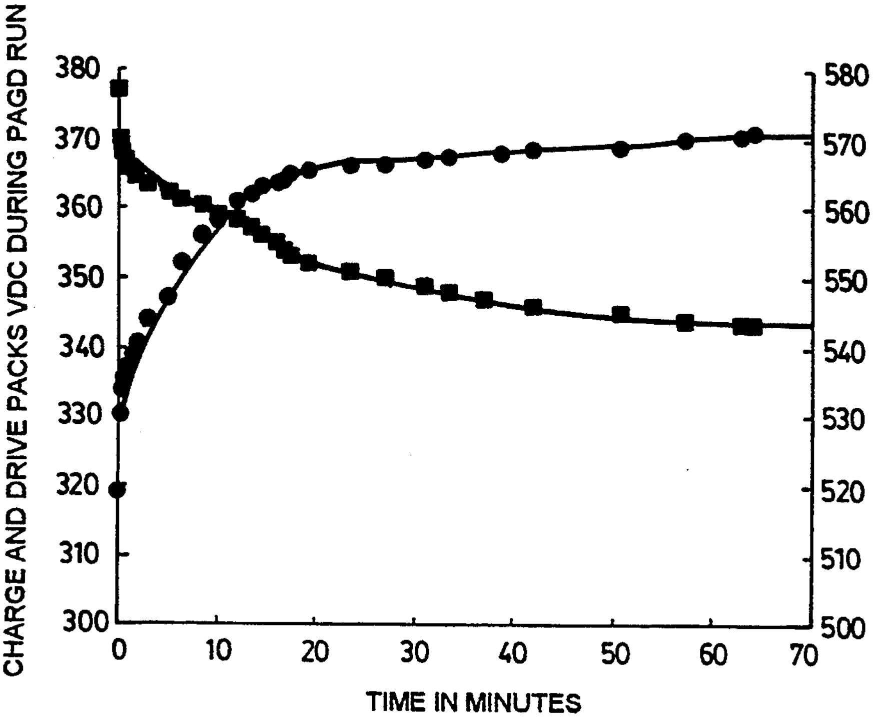
CHARGE AND DRIVE PACKS VDC DURING PAGD RUN
380
370
360
350
340
330
320
310
300
580
570
560
550
540
530
520
510
500
0 10 20 30 40 50 60 70
TIME IN MINUTES

FIG. 21

ENERGY CONVERSION SYSTEM

REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of U.S. application Ser. No. 07/922,863, filed Jul. 31, 1992 (abandoned), and is also a continuation-in-part of U.S. patent application Ser. No. 07/961,531, filed Oct. 15, 1992, now U.S. Pat. No. 5,416,391.

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates to energy conversion circuits utilizing discharge tubes operating in the pulsed abnormal glow discharge (PAGD) regime.

2. Review of the Art

Such discharge tubes and circuits incorporating them are described in our copending U.S. patent application Ser. Nos. 07/922,863 and 07/961,531. The first of these applications discloses discharge tube constructions particularly suited for PAGD operation, and the second discloses certain practical applications of such tubes, particularly in electric motor control circuits. The review of the art contained in those applications is incorporated herein by reference, as is their disclosure and drawings.

It is known that there are anomalous cathode reaction forces associated with the cathodic emissions responsible for vacuum arc discharges, the origin and explanation of which have been the subject of extensive discussion in scientific literature, being related as it is to ongoing discussion of the relative merits of the laws of electrodynamics as variedly formulated by Ampere, Biot-Savart and Lorentz. Examples of literature on the subject are referenced later in this application.

SUMMARY OF THE INVENTION

The particular conditions which prevail in a discharge tube operated in the PAGD regime, in which a plasma eruption from the cathode is self-limiting and collapses before completion of a plasma channel to the anode gives rise to transient conditions which favour the exploitation of anomalous cathode reaction forces.

We have found that apparatus utilizing discharge tubes operated in a self-sustaining pulsed abnormal glow discharge regime, in a double ported circuit designed so that energy input to the tube utilized to initiate a glow discharge pulse is handled by an input circuit substantially separate from an output circuit receiving energy from the tube during collapse of a pulse, provides valuable energy conversion capabilities.

The invention extends to a method of energy conversion, comprising initiating plasma eruptions from the cathode of a discharge tube operating in a pulsed abnormal glow discharge regime utilizing electrical energy from a source in a first circuit connected to said discharge tube, and capturing electrical energy generated by the collapse of such eruptions in a second circuit connected to said discharge tube.

SHORT DESCRIPTION OF THE DRAWINGS

The invention is described further with reference to the accompanying drawings, in which:

FIG. 1 shows variation of applied DC current and pulse AC rms currents characteristic of a low current PAGD regime, as a function of decreasing pressure, for a 128 cm^2 H34 aluminum plate pulse generator having a 5.5 cm gap length and being operated in the single or plate diode configuration of FIG. 11A, at ~600 VDC.

FIG. 2 shows variation of applied DC current and AC rms currents of a high current PAGD regime, as a function of the decreasing pressure, for a device identical to that of FIG. 1, and operated at the same potential.

FIG. 3 shows PAGD rate vs. pulse generator cathode temperature as a function of the time of continuous PAGD operation, for a pulse generator with 64 cm^2 plates having a 4 cm gap distance, operated at VDC=555 (av) and R1=600 ohms (see FIG. 9).

FIG. 4 shows PAGD frequency variation with time, for 18 successive spaced one-minute PAGD runs for a pulse generator with 128 cm^2 plates, and a 5.5 cm gap distance, operated at VDC=560 (av) and R1=300 ohms.

FIG. 5 shows variation of the PAGD frequency in pulses per minute (PPM) with increasing charge of a PAGD recovery charge pack (see FIG. 9), as measured in terms of the open circuit voltage following 15 minutes of relaxation after each one minute long PAGD run, repeated 18 times in tandem, under similar conditions to FIG. 4.

FIG. 6 shows volt amplitude variation of continuous PAGD at low applied current, as a function of decreasing air pressure, for a 128 cm^2 plate area device, gap length=5 cm; (DCV at breakdown=860).

FIG. 7 shows volt amplitude variation of continuous PAGD at high applied current as a function of the decreasing air pressure, for a 128 cm^2 plate area device, gap length=5 cm; (DCV at breakdown=860).

FIG. 8 is a schematic diagram of a first experimental diode (without C6) or triode PAGD circuit.

FIG. 9 is a schematic diagram of a preferred diode or triode PAGD circuit in accordance with the invention.

FIGS. 10A, 10B and 10C are fragmentary schematic diagrams showing variations in the configuration of the circuit of FIG. 9.

FIG. 11 is a modification of FIG. 9, in which an electromagnetic machine, in the form of an electric motor, is connected into the circuit as an accessory electromechanical arm.

FIG. 12 shows a further development of the circuit of FIG. 9, permitting interchange of driver pack and charge pack functions.

FIG. 13 shows open circuit voltage relaxation curves for battery packs employed in tests of the invention, respectively after pre-PAGD resistive discharge (DPT1 and CPT1), after a PAGD run (DPT2 and CPT2) and after post-PAGD resistive discharge (DPT3 and CPT3).

FIG. 14 shows an example of negligible actual power measurements taken immediately before or after a PAGD run, showing both the drive pack loss and the charge pack gain in DC Watts; DP resistance=2083 ohms; CP resistance=833 ohms.

FIGS. 15A and 15B show resistive voltage discharge curves for two separate lead-zero gel-cell packs utilized respectively as the drive and the charge packs; load resistances employed were 2083 ohms across the drive pack (FIG. 15A) and 833 ohms across the charge pack (FIG. 15B).

FIG. 16 shows resistive discharge slopes for a drive pack before and after a very small expenditure of power in providing energy input to a PAGD run; R=2083 ohms.

FIG. 17 shows resistive discharge slopes for a charge pack before and after capturing energy from the collapse of PAGD pulses in the same test as FIG. **15**; $R=833$ ohms.

FIG. **18** shows resistive discharge slopes for a drive pack before and after a very small expenditure of power in providing energy input to a PAGD run in a further experiment; $R=2083$ ohms.

FIG. **19** shows resistive discharge slopes for a charge pack before and after capturing energy from the PAGD run of FIG. **18**; $R=833$ ohms.

FIG. **20** shows an example of operational measurements taken videographically during a 10 second period for both the power consumption of the drive pack (PAGD input) and the power production captured by the charge pack (PAGD output); the two values are also related by the expression of percent breakeven efficiency.

FIG. **21** shows variation of PAGD loaded voltage of a drive pack (in squares) compared with the PAGD charging voltage of the charge pack (in circles), during more than 1 hour of continuous PAGD operation.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The basic PAGD function and the construction of discharge tubes specifically designed for PAGD operation are described in our corresponding copending applications Nos. 07/922,863 (the '863 application) and 07/961,531 (the '531 application). For purposes of the experiments described below four aluminum H34 plate devices (one with 64 and three with 128 cm^2 plate areas) and three aluminum (H200) plate devices (one with 64 and two with 128 cm^2 plate areas), with interelectrode gap lengths of 3 to 5.5 cm, were utilized at the indicated vacua, under pumpdown conditions and with either air or argon (ultra high purity, spectroscopic grade 99.9996% pure) constituting the residual gas mixture. The pumpdown conditions were as described in the '863 application. Some experiments were performed with the tubes under active evacuation, at steady-state conditions, while others utilized sealed devices enclosing the desired residual gas pressures.

The circuit designs utilized in the various experiments to be described are set out further below, and represent further developments and extensions of the circuits set forth in the '531 application.

Test equipment utilized was as follows:

An Edwards (trade mark) thermocouple gauge (TC-7) was employed for the determination of pressure down to 1 micron of mercury (0.001 Torr).

Banks of Beckman (trade mark) rms multimeters 225 and 330 (30 and 100 kHz bandwidths, respectively) were utilized for all current measurements.

Frequency meters capable of discriminating events up to 0.1 nanosecond apart, and having adjustable amplitude windows, were used. Direct analysis on a Tektronix (trade mark) dual-trace, storage scope (Model 549) was also carried out for both parameters.

Split-phase, single-phase and two-phase motors were employed, of the synchronous, induction and universal types, as previously described in the '531 application, in the accessory electromechanical arm that may be coupled to the power producing circuit described in the present application.

Large banks of 12 V, 6 Ah lead-acid gel cells (Sonnenschein (trade mark) A212/6S) were utilized either as power sources (designated as drive packs) or as accumulators of the energy (referred to as charge packs) captured by the test circuits. Charge packs made of rechargeable 9 V NiCad or of nominally nonrechargeable C-Zn or alkaline batteries were also utilized.

PAGD emission areas were determined by metallographic examination of a series of craters produced by PAGDs in clean H34 cathodes, under a metallurgical Zeiss (trade mark) standard 18 microscope equipped with an epi-fluorescent condenser, very high power apochromatic objectives and a 100 W mercury lamp. For best results a focusable oblique source of light (12 V halogen) was also added to the incident light.

Following our low and high applied current studies on PAGD production as set forth in the '863 application, we noticed that the AC rms value of the component associated with each abnormal glow discharge pulse varied nonlinearly with the magnitude of the applied current. We originally noted the existence of a current induced shift of the entire PAGD region upward in the pressure scale: while the PAGD regime became more clearly defined as the applied constant DC was increased, the pressure required to observe the PAGD increased two to three orders of magnitude. In the course of these rarefaction studies we found that, at applied currents of 1 mA or less, the rms value of the different AC waveforms associated with the consecutive regimes of the discharge (TRD→NGD→AGD+PAGD) was, by more than half log, inferior to the value of the applied DC current, during the first two regimes (TRD and NGD) and reached a value equivalent to the applied current with the onset of spontaneous PAGD, at pressures <0.1 Torr (see FIG. **1**); however, in the downward tail of the PAGD regime (down to $3*10-3$ Torr), the AC rms current component of each PAGD again decreased to more than half log of the intensity of the applied DC value, in a manner proportional to the log of the decreasing pressure. In stark contrast, at high applied currents of ~500 mA, and aside from the high current-induced upward shift in pressure of the PAGD regime (to the point that the compression of the previous regimes on the pressure scale results in their suppressing, as was the case in the present example), the AC rms component associated with each pulse (see closed circles, FIG. **2**) is, from onset of the discharge at ~8 Torr, greater in magnitude than the value of the applied current (open circles, FIG. **2**). Under the conditions described, the distribution of the field current associated with each pulsed abnormal glow discharge approached (on a linear Y axis; not shown) an unimodal gaussian distribution with the pressure peak at ~1 Torr, and a corresponding observed maximum of 7.5× higher AC rms values than the applied DC values.

We have previously described in the '863 application how the PAGD frequency is affected by several factors, namely: the magnitude of the parallel discharge capacitance, the value of the negative pressure for the relevant vacuum PAGD range, the magnitude of the applied potential, the magnitude of the applied direct current, the interelectrode gap distance and the area of the parallel plate electrodes. In the '531 application we have also described how the wiring configuration (plate diode versus triode) affects the PAGD frequency by adding tungsten autoelectronic emissions from the axial electrode, to those emissions from the plate. There are other factors which limit the PAGD regime of discharge and have also been discussed in the '863 application. The following data indicates their specific effect upon PAGD frequency.

In the data presented in Table 1, control of the frequency parameter for the circuit shown in FIG. 9 is by a ballast resistance R1 within a specific range of interest (~800–150 ohms, for Table 1 experimental conditions), and this in turn increases the applied current which, at "high current" values (i.e. >100 mA, as for Table 1 conditions), will drive the PAGD frequency up, as previously reported in the '863 application.

Table 2 shows the effect of the progressive displacement of a given frequency, chosen as 200 PPS, with the cumulative pulse count of the same device, in the plate diode configuration. This displacement of the same frequency (cf. group #'s 1–3, Table 2) onto higher pressure regions is shown to be promoted by the alteration of the work function of the PAGD emitting cathode, such as this is caused by the cumulative pulse count and resultant crater formation on the electrode surface. After the first million pulses, the anode facing cathode surface is completely turned over by emission sites, and this corresponds well to the threshold crossed by group #2, Table 2. Once the cathode surfaces are broken in, the rates shown in groups #3 and 4, Table 2, tend to remain constant. Originally we wondered whether this might be caused by the alteration of the electrostatic profile of the plasma sheaths at the periphery of the envelope, due to the mirroring deposits that result from the sputter of ions and trapped neutral atoms (from air gases or metallic vapor) associated with the autoelectronic emission mechanism (and from further emissions triggered in turn, by secondary ionic bombardment of the cathode with molecular species present in the plasma ball formed over the primary emission site). However, reversal of the plate polarity (firing the ex-anode as a crater-free cathode) for over a million counts, followed by re-reversal to the original polarity, the entire operation being performed in air as the residual gas substrate, led to the partial recovery of the original work function for as long as the test was run ($1.5*10^4$ pulses), as shown by a comparison of groups #2,4 and 5, Table 2. From a metallographic examination of the surfaces of plates used solely as anodes, we have also concluded that prolonged PAGD operation has the effect, not only of cleaning the anode surface from surface films and adsorbed gases, as ionic bombardment promoted by electromagnetic induction coils does, but it also does more—it polishes the target surface and smooths it by a molecular erosive action. Observations of the surface of reversed cathodes, shows the same smoothing and polishing effects observed in exclusive anodes. Thus the recovery of the PAGD rates promoted by polarity reversal of the plates is not a function of the sputter-promoted mirroring deposits on the envelope wall, but a function of the actual work-function of the emitting cathode.

Another variable that interacts with the PAGD frequency is the molecular nature of the residual gas: Table 3 shows the differential frequency response of air with a halogen quencher, argon, for the same pulse generator employed in the tests of Table 2. It is apparent that argon obtains much higher rates of AGD pulsation for the same range of negative pressure, for the same "broken in" cathode, than does the air mixture. All these measurements were taken at cathode support-stem temperatures of 35° C.

Time of operation is also a variable affecting the frequency and operating characteristics of the cathode, as it becomes expressed by the passive heating of the cathode, an effect which is all the more pronounced at the higher pressures and at the higher frequencies examined. Utilizing the triode circuit discussed in the next section, the pulse rate of a PAGD generator with 64 cm^2 plates can be seen (see FIG. 3) to decrease, at a negative pressure of 0.8 Torr, from 41 PPS to the operating plateau of 6 PPS within 15 minutes of continuous operation, as the temperature of the cathode support increased from 19° to ~44° C. As the temperature plateaus at ~51°±1° C., so does the pulse rate at 6 PPS, for the remaining 48 minutes of continuous operation.

However, in order to confirm this time-dependent heating effect and threshold, we also performed the same experiment, utilizing the same circuit and the same negative air pressure, with twice as large a cathode area (128 cm^2, which should take nearly twice as long to heat), being operated for 18 one-minute long continuous periods equally spaced apart by 15 minutes of passive cooling, with the cathode stem always at 19.7° to 21° C., room temperature at the start of each period. The results surprised us, inasmuch as they showed that for a larger area tube which takes longer to heat to the same temperatures at comparable rates of PAGD triggering, one could observe a much earlier frequency reduction (by half, within the first 5 minutes or periods of interrupted functioning) in the absence of any significant heating effect (<1.5° C.) of the cathode (see FIG. 4). Repetition of these experiments has led us to conclude that, as shown in FIG. 5, the variable responsible for this repeatedly observed reduction in the PAGD frequency, when the PAGD operation sequence is systematically interrupted, is the state of charge/discharge of the battery pack (the charge pack) at the output of the triode circuit in question: the PPM rates in FIG. 5 decrease rapidly with the steepest rate of charging of the charge pack and the fastest recovery rate of its open circuit voltage; above a given state of charge, when the open voltage of the charge pack climbs more slowly (>340 V), in a log fashion, the PPM rate stabilizes at its plateau values.

Confirmation of the importance of the charge pack in the PAGD function of the present circuitry here considered, comes from the fact that the size (the number of cells) and the intrinsic capacitance of the charge pack affect the PAGD frequency dramatically (see Table 4): increasing the charge pack size of 29 cells to 31, by 7% leads to a 10-fold reduction in frequency; further increases in the number of charge pack cells extinguishes the phenomenon. On the upper end of the scale, this effect appears to be tied in to restrictions that it places on the ability of the larger charge packs to accept the discharge power output once the charge pack voltage exceeds the PAGD amplitude potential. All of these measurements were conducted with the same 128 cm^2 plate PAGD generator, at a pressure of 0.8 Torr and in the triode configuration (see FIG. 9).

Other factors can also affect the frequency: the motion of external permanent magnetic fields oriented longitudinally with the interelectrode gap, external pulsed or alternating magnetic fields, external electrostatic or electromagnetic fields, specific connections of the earth ground, and the presence of a parallel capacitative, capacitative-inductive or self-inductive arm in the circuit, such as we have described for our electromechanical PAGD transduction method as described in the '531 application.

Analysis of the modulation of PAGD amplitude is simpler than that of its frequency, because fewer factors affect this parameter: (1) magnitude of the applied potential, (2) interelectrode gap distance and (3) the negative pressure, as shown in the '863 application, for "low" applied currents. As the magnitude of the applied potential itself is limited by the gap and the pressure, to the desired conditions of breakdown, the important control parameter for the PAGD amplitude is the pressure factor. This is shown in FIGS. **6** and. **7**, respectively for "low" (5 mA) and "high" (~500 mA) applied currents and for the same plate diode configuration of a H34 Al 128 cm^2 plate PAGD generator (5 cm gap), in the simple circuit described in the '863 application; it is apparent that both positive and negative components of the amplitude of these pulses in the oscillograph, are a function of the pressure, but the maximum cut-off limit of our equipment, for the negative component (at 240 volts for the "low" current experiment and at 120 volts for the "high" current), precluded us from measuring the peak negative voltage of these pulses. However, rms measurements of the pulse amplitude at the plates and DC measurements at the circuit output to the charge pack indicate that the negative component increases with decreasing pressure to a maximum, for a given arrangement of potential and gap distance; no pressure-dependent bell shape variation of the pulse amplitude, as that seen for the positive component at "high" applied currents (FIG. **7**) is observed with the negative amplitude component. For the typical range of 0.8 to 0.5 Torr, the rms value for pulse amplitude varies from 320 to 480 volts, for a 5.5 cm gap distance and applied DC voltages of 540 to 580. PAGD amplitude is a critical factor for the design of the proper size of the charge pack to be utilized in the optimal circuit.

The development of the circuits to be described stemmed from fundamental alterations to the principles implicit in our previous methods of electromechanical transduction of AGD plasma pulses as described in the '531 application. Whereas this electromechanical coupling (capacitative and self-inductive), utilized directly, energizes the AGD pulses inverted from the DC input by the vacuum generator, the purpose of the development that led to the presently described experiments was to capture efficiently, in the simplest of ways, most of the pulse energy in a closed circuit, so that power measurements for the energy transduction efficiency of the observed endogenous pulsation could be carried out. Ideally, comparative DC power measurements would be performed at both the input and output of the system, taking into account the losses generated across the components; this would obviate the measurement problems posed by the myriad of transformations implicit in the variable frequency, amplitude, crest factor and duty-cycle values of the PAGD regime, and necessitated some form of rectification of the inverted tube output. From the start our objective was to do so as simply as possible. Early circuits utilizing half-wave rectification methods coupled in series to a capacitative arm (for DC isolation of the two battery packs), with the charge pack also placed in series, showed marginal recoveries of the energy spent at the PAGD generator input. Attempts at inserting a polar full-wave rectification bridge led, as shown in FIG. **8**, to the splitting of the capacitor into capacitors **C3** and **C5**, at the rectification bridge input, and capacitor **C4** in series with both capacitors, all three being in a series string in parallel with the PAGD generator. Under these conditions a DC motor/generator could be run continuously in the same direction at the transversal output (**U1** and **U2**) of the bridge; but if this inductive load was replaced with a battery pack CP (charge recovery pack), either the parallel capacitor **C4** had to remain in the circuit, for the diode configuration or, less desirably, a further capacitor **C6** could replace **C4** and connect one electrode, preferably the cathode C, to the axial member of the discharge tube T, thus resulting in a first triode configuration as actually shown in FIG. **8**. Energy recovery efficiencies of the order of 15 to 60% were obtained utilizing **C6** in this manner, but measurements of the potential and currents present at the output from the rectifier bridge were substantially lower than those obtained using optimal values of **C4**. Effectively, under these conditions, much of the power output from the tube was never captured by the output circuit formed by the second, right hand arm of the system and, being prevented from returning as counter-currents to the drive pack DP by diodes **D1** and **D4**, was dissipated and absorbed by the interelectrode plasma, electrode heating and parasitic oscillations.

Solutions to this problem were explored using the circuit shown in FIG. **9**, which still maintains the necessary communication link for the quasi-sinusoidal oscillation of the capacitatively stored charges at the input and outputs of the rectification bridge, but integrated the functions of capacitor **C4** into the single rectification circuit, in the form of an asymmetric capacitative bridge **C7***a* and **C7***b* placed transversally to the capacitative bridge formed by **C3** and **C5** and in parallel with the charge pack CP at the output from the rectification bridge **D5**, **D6**, **D2**, **D3**. This second capacitative bridge is so disposed as to have its centre point connected to the anode A through capacitor **C5**. If the axial member of the Tube T were to connect to the junction of **D2** and **D3** instead of at the junction **D5**–**D6**, the function of bridge **C7***a* and **C7***b* would be connected to the cathode C through capacitor **C3**. The capacitative bridge is insulated from the charge pack whose voltage it stabilizes, by rectifiers **D7** and **D8**, which also prevent leakage of charge across **C7***a* and **C7***b*. The anode and cathode oscillations generated by the electrostatic charge transduction through **C3** and **C5** into the poles of the charge pack are trapped by the transversal transduction of the **C7** bridge, at the outputs from the rectification bridge, of which the oscillation has to become split between the bridge inputs into half-waves, for electrostatic transduction and full wave rectification to occur. In fact, under these conditions, removal of the **C7** bridge will suppress the PAGD phenomenon, unless other circuit variables are also altered. The transversal bridge is thus an essential piece of this novel circuit. Variations in the circuit as shown in FIG. **10** were then studied, the first two being selectable utilizing switch **S2** (FIG. **9**).

The presence of the capacitative bridge effectively reduces the dynamic impedance of the charge pack CP so that the output circuit approximates to a characteristic in which it presents a very high impedance to the tube T at potentials below a certain level, and a very low impedance at potentials above that level.

With this modified circuit, more effective recovery of the energy produced by collapse of the PAGD pulses is possible, with more effective isolation from the input circuit utilized to trigger the pulses. Under these conditions, the energy captured by this circuit at the output, is not directly related to that utilized in triggering the pulses from the input. The attainment of this condition critically depends on the large capacitance of the transversal bridge being able to transfer the output energy from the tube T into the charge pack CP. Under these conditions, we have found, as will be shown below, that the large peak pulse currents released by collapse of the PAGD pulses released more energy than is used to trigger them, and these findings appeared to tally with other observations (abnormal volt-ampere characteristics and anomalous pulse currents, etc.) associated with the anomalous cathode reaction forces that accompany the auto-electronic emission-triggered PAGD regime. Experiments so far indicate that the power output can be increased proportionately to the series value of C3, C5 and the two identical C7 capacitors.

The circuit of FIG. **10** can be integrated with a circuit such as that disclosed in the '863 application as shown in FIG. **11**, in which a part of the energy recovered can be shunted by the switch S4 into an induction motor M1 having rotor R, to a degree determined by the adjustment of potentiometer R4 and the value selected for C4.

The circuit of FIG. **11** can be further developed as exemplified in FIG. **12** to include configurations which provide switching permitting interchange of the functions of charge packs and the drive packs, it being borne in mind that the nominal potential of the drive pack must be substantially higher than that of the charge pack, the former needing to exceed the breakdown potential of the tube at the beginning of a PAGD cycle, and the latter to be less than the extinction potential.

FIG. **12** essentially represents a duplication of the circuit of FIG. **11**, the two circuits however sharing two identical battery packs BP1 and BP2, and being provided with a six pole two way switch, the contact sets of which are identified as S1, S2, S3, S4, S5 and S6. When the contacts are in position A as shown, battery pack BP1 acts as a drive pack for both circuits, with the upper half (as shown) of the battery pack BP2 forming the charge pack for the upper circuit, and the lower half forming the charge pack for the lower circuit. When the pack BP1 is at least partially discharged, the switch is thrown so that contacts move to position B, which reverses the function of the battery packs thus allowing extended operation of the motors in each circuit each time the switch is thrown.

Based on the manufacturer's data, and using current values within the range of our experimentation as discussed in the next sections, an optimal discharge cycle for a fully charged 6.0 Ahr battery pack at 0.300 A draw is 20 hours, as claimed by the manufacturer, and this corresponds to a cycling between 100% (12.83 V/cell open circuit and load start voltage) and <1% (10.3 V/cell load voltage) of the battery's absolute charge capacity. Even though the discharge mechanism is a time cumulative process with a log function, the discharge can, within 4 to 5 hour time segments (or periods with 20–25% of the full range), be regarded as practically linear with time. This trait, or linearization of the discharge slope, becomes more marked with advancing age and decreasing absolute storage capacity of the cells.

The proportionality between open circuit voltage and the percentage of residual relative capacity for these cells when new (uncycled and not yet aged) is uniform over 98% of the permissible charge capacity withdrawal; in practice this translates into a slope that becomes steeper with time, while the absolute storage capacity diminishes. In turn, this decreasing absolute capacity of the cells results in shorter load discharge times and their further linearization.

A circuit in .general accordance with FIG. **9**, employed in the studies reported in this and the following sections, utilizes a drive pack of 46*12 V Lead acid gel-cells each with a 6.0 Ah rating, and a charge pack with 28 or 29*12 V identical cells. The charge pack was cycled anywhere from 11.2 V to 12.8 V/cell (open circuit voltages), within the proportional region of the relative capacity slope, to yield a capacity increment in the order of 50% (e.g. from 20 to 70%), anywhere within the range of 2 to 100% of its total charge capacity, assumed for now as invariant. The charging process, hereinafter referred to as a PAGD run, took about 20–30 minutes under optimal conditions. The drive pack typically consumed, in the same period of time, 4 to 11% of its initial total capacity , its open circuit voltage typically falling 0.1 to 0.2 V per cell after a PAGD run, within the open circuit range of 12.8 V/cell (100% relative capacity) and 11.2 V/cell (~2%). At the 100% capacity benchmark, the drive pack would theoretically have 20 h*46 cells*12.83 V/cell*0.3 A=3.5 KWh, and the charge pack, for example, 20 h*29*12.83 V/cell*0.3 A=2.2 KWh. Since the capacity per cell is linear with the open circuit voltage within the proportional range, as claimed by the manufacturer, we projected the open circuit voltage intercepts on the manufacturer's proportional curve in order to determine the residual percentage of the total relative capacity and the standard hours of operation left, from any experimental open circuit voltage measurements.

Three pulse generators (2*128 cm^2 and 1*64 cm^2 plate areas) were employed in these studies; they were operated in PAGD runs at 1–120 pulse/second rates, within a negative pressure range of 0.2 to 0.8 Torr and with applied direct currents of 0.2 to 0.6 A.

Both drive and charge packs utilized cells which were bought new at the same time and had initial charge values of 12.4 to 12.55 V/cell (open circuit). These batteries are capable of energy densities of 33–35 Whr/kg. However, the experiments shown in Table 5 are selected from a series that spanned nearly 12 months, beginning 6 months after purchase; hence, loss of absolute storage capacity by the batteries had occurred in the intervening time, as a function of both age and charge/discharge cycle life.

Measurements of the open voltage of either drive (D) or charge (C) (see column 2, Table 5) packs for 8 separate experiments, all utilizing the triode configuration, were performed before (b) and after (a) a PAGD run (see columns 3 and 4), at either 15 or 30 minutes (see column 26) of the open circuit voltage relaxation after a PAGD run was terminated. Corresponding open circuit voltages per cell are shown in column 5, and the percentages of the predicted total relative charge capacity resulting from the intercepts on the manufacturer's proportional curve are shown in column 6, Table 5. Equivalent maxima for the theoretical hours of operation left are shown in column 7, the percentage change in relative capacity arising as a consequence of either charge pack charge capture (capacity gained) or of drive pack output (capacity lost) is shown in column 8. Translating the intercepts into power units yields the values shown in column 9, Table 5, for total kWh left in each pack before and after PAGD production, those shown in column 10 for the actual power gained and lost during the periods of operation (presented in column 12) and those shown in column 13 for the power predicted to be gained or lost per hour of PAGD production. On the basis of the experimental open voltage values and their intercepts, the predicted net kWh values per hour of PAGD energy production (after deduction of measured losses) and the corresponding experimental breakeven efficiencies (where breakeven=100%) are presented, respectively, in columns 14 and 15. The PAGD frequency per second is shown in column 11; the number of 12 V cells, in column 16; the tube ID, in column 17; the cathode (and anode) area (s), in column 18; the plate material, in column 19; the input ballast utilized (**R1**, FIG. **9**), in column 20; the size of each capacitor (**C3** or **C5**) of the tube output bridge, in column 21; the size of each capacitor (**C7***a* or **C7***b*) of the transversal capacitative bridge, in column 22; the status of **S4** and thus, of the parallel and auxiliary electromechanical arm (see FIG. **11**), in column 23; the negative air pressure in column 24; the gap distance between the plates, in column 25; and columns 27,28 and 29, show the status of the elements of the switched on parallel electromechanical arm of the circuit—the parallel **C4** capacitor, the motor input resistor **R4** and the motor revolutions per minute (measured stroboscopically), respectively.

From these figures of Table 5, and utilizing the data for the two first examples shown, we calculated the predicted performance of the system based on the open voltage measurements. In the first example, where the system was run continuously without interruption, the charge pack increased the percentage of its total capacity by 43% (a two-fold increase in capacity) and, during the same period, the driver pack decreased the percentage of its total capacity by 7% (a ~10% decrease in capacity relative to the percentage of residual total capacity at the start, i.e. 77%) (cp. columns 6 and 8, Table 5). Subtracting the predicted initial total energy (0.835 KWh) available to the charge pack before the experimental run (first line of column 9, Table 5) from the predicted total energy (1.823 KWh, second line of column 9) available to the charge pack after the PAGD charge run, gives us the total energy gained by the charge pack: 0.988 KWh (column 10) in 21.5 minutes (column 12) of continuous PAGD performance. Conversely, subtracting the predicted final total energy (2.4 KWh) available to the driver after the experimental run (fourth line of column 9, Table 5) from the predicted total energy (2.66 KWh, third line) available to the driver before the PAGD charge run, gives us the total energy lost by the drive pack: 0.26 KWh in 21.5 minutes. If we divide the total available energy gained by the charge pack, by the total energy lost by the drive pack, we obtain a surplus factor of 3.9×, or 388% of the breakeven point (column 15). The same values result from dividing the charge pack % of total capacity gain by the drive pack % of total capacity lost, and then downscaling this value by multiplying it by the typical scale factor for the two packs, 29/46=0.63×.

In an analogous fashion, we analyzed the results for the second example shown in Table 5. Here, the charger increased the percentage of its total capacity by 45.5% (a 22.75 fold increase in estimated total relative capacity) and, during the same period, the driver decreased the percentage of its predicted total capacity by 7% (a ~17.5% decrease in capacity relative to the percentage of residual total capacity at the start, i.e. 40%). By dividing the predicted total available energy gained by the charge pack (0.962 KWh/18 minutes) by the expected total energy lost by the driver pack (0.246 Kwh/18 minutes) we obtain a surplus factor of 3.9×, or 391% of the breakeven point. This corresponds to an interrupted, total sequential run of 18 minutes, each minute-long run being separated by a cooling and voltage relaxation period of 15 minutes before the next run is carried out, at an average PAGD frequency of 61 PPS.

Analysis of the remaining results illustrates how a number of PAGD controlling parameters interact to determine conditions for effective maintenance of a PAGD regime. The lower gain and higher loss per unit time registered for the third run of Table 5, which results in the lower breakeven efficiency of 230% and a smaller net power production rate than before (power estimates of 1.396 kWh/h of PAGD operation vs 2.387 kWh/h, for the second run, Table 5) illustrate, for example, the combined effect of lowering the pressure (0.8 to 0.7 Torr) and running the PAGD continuously (the heating effect), both of which depress the PAGD frequency. The fourth run of Table 5 identifies the continuous performance of a "broken in" softer grade of aluminum (column 19), having a lower work-function (as determined from the higher PAGD frequency spectrum) than the harder H34 plates of the previous examples, and shows that, despite the series value of the total capacitance being higher (5,333 mfd vs 4,030 mfd for runs one through three), and despite the higher vacuum (0.2 Torr), the lower work-function results in a higher frequency; however, even though this run registers a predicted higher breakeven efficiency (310%) than the previous experiments, these conditions result in a 4/5-fold lower estimate of net power produced, when compared to the previous three PAGD runs.

PAGD runs 5 and 6, Table 5, illustrate the effect of switching on the auxiliary electromechanical arm of the circuit shown in FIG. **11**. Increasing the amount of charge capacitatively shunted into the electromechanical arm by higher **C4** values (column 27), and increasing the current that feeds the squirrel cage induction motor utilized by lowering **R4** (column 28), results in a power capture by the charge pack that registers an energy loss (predicted to be 96% efficient, falling short 4% of breakeven recovery), as most of the tube output power is spent in the electromechanical arm and its motor effect. Furthermore, under the conditions of maximum electromechanical action, the drain imposed on the drive pack becomes considerable (see loss in columns 10 and 13), even if the **C3** and **C5** values are reduced, column 21, Table 5). These runs also illustrate how the motor appears to function as an electrical induction generator having rpm values much higher than the synchronous values prescribed by the frequency of the PAGD (column 29, Table 5).

The extremely large breakeven efficiency of PAGD run 5, Table 5, indicates that with selected values of **C4** and **R4**, it is possible to operate the motor in the auxiliary arm and still accumulate excess energy from the PAGD production in the charge pack.

Runs 7 and 8 illustrate results obtained for 64 cm^2 plates, and a shorter interelectrode gap distance, for two pressures (0.8 and 0.5 Torr), the device being open to a rotary pump manifold in the first instance and sealed from the pump, in the second case. Despite the lower vacuum, the higher pulse frequency (32 vs 5 PPS) and breakeven efficiency (906% vs 289%) registered by run 8 when compared to run 7, are a consequence of the method of run 8, which was interrupted systematically by 5 passive cooling periods, as in the case of run 2, whereas run 7 was continuous. This again resulted in higher average PAGD frequencies (at lower pressures), a predicted two-fold greater gain and a predicted two-fold smaller loss (columns 13 and 14) for run 8.

FIG. **13** shows curves representing the slopes of the open circuit relaxation voltages, which are linear with the log of time elapsed from cessation of discharge, for both drive and charge packs, in the same run 8 set out in Table 5. The experiment in its entirety consisted of preliminary resistor-loaded measurement discharges and their corresponding open circuit voltages from the moment of cessation of the resistive discharge (illustrated, respectively, by the open squares of **DPT1** for drive pack relaxation time **1**, and by the open circles of **CPT1** for charge pack relaxation time 1), followed by their relaxation rates in the wake of the PAGD production (the hatched squares of DPT2 for drive pack relaxation time **2**, and the hatched circles of CPT2 for charge pack relaxation time 2), and finally, by the relaxation rates from the final resistor-loaded measurement discharges (the black squares of DPT3 for drive pack relaxation time **3**, and the black circles of CPT3 for charge pack relaxation time **3**). Discharge resistances were 833 ohms for the charge pack, and 2083 ohms for the drive pack in all cases, corresponding to resistors **R3** and **R2**, respectively, of FIG. **9**. This methodology will be examined in greater detail below. It is apparent that, after every load period, be this resistive (**CPT1, DPT1,** CPT3 and DPT3) or due to PAGD operation (DPT2), the relaxation slope is positive; as shown from slopes **CPT1** and **DPT1**, the log time proportionality of the open circuit voltage relaxation, under these conditions, tends to plateau after ~30 minutes. The exception to this general behaviour lies in the voltage relaxation slope CPT2, which is negative and reflects the charge accumulation occurring in the charge pack and obtained by capture of energy produced during PAGD operation, triggered by the energy drawn from the drive pack during load time **2**.

As a first approximation of electrical power generated and consumed by the energy conversion system of the invention, the previous open circuit voltage method is of significance in showing the basic trends involved in interaction of the operating parameters. However, in all likelihood, it overestimates the actual values of electrical power consumed and generated, for a variety of reasons. First, it assumes that the relative capacity scale of the batteries in the drive and charge packs is an absolute charge capacity scale with an invariant maximal charge retention, which it is not; in fact, the absolute charge capacity is itself a variable subject to several factors, such as the cycle life, overcharging or undercharged conditions, cell age, residual memory and the rate of charge and discharge. Hence, the inference of a uniform time scale on the basis of the open circuit voltage/capacity intercepts may not be warranted. Finally, it does not integrate the open voltage decrease over time, and utilizes the specification load current as the average current over time.

In order to obviate these problems, we resorted to a variety of other measurement methods. First, we proceeded to compare the closed circuit, preliminary, resistive-load discharge measurements for either charge or drive pack, under conditions of negligible loss of power, as these measurements were statistical means (n=9) taken, at equal intervals, during the first 90 seconds of the load discharge, and obtained both just before the PAGD production runs (but separated from each PAGD run by an open circuit voltage relaxation of 30 minutes) and just after the runs (but equally separated by a relaxation of 30 minutes). As an example of the data generated by such an approach, FIG. **14** illustrates the shift of the slopes indicating marginal power loss for the drive pack (from the closed squares to the open squares) and those indicating gain of power for the charge pack (from the open circles to the closed circles), in actual total load power values.

Integration of these power measurements over the projected load discharge time, taken from the family of curves generated on the basis of the manufacturer's load voltage over discharge time specifications, led to a direct comparison of the new values, as shown in Table 6, with the values presented in Table 5, for the first three instances introduced. All values of Table 6 were obtained by resistive measurements of power that entailed a negligible power loss. Table 6 confirms the fundamental equivalence of runs 1 through 3, as already seen from their corresponding analysis using the open voltage method (see runs 1 to 3, Table 5). This new power estimation method also confirms the lower loss encountered in run **2** utilizing interrupted PAGD operation. While the breakeven efficiencies sensibly doubled using this method, the estimates of actual electrical power consumption recovery decreased by a 2 to 3-fold factor. Thus this direct load voltage/amperage measurement method of estimating actual power losses or gains, is a check upon the open voltage method previously utilized.

Direct, instantaneous measurements of the voltage and current characteristics of the PAGD production and capture phenomena being discussed, were also performed during PAGD runs for diverse sets of conditions, including all those described in the two previous sections. In Table 7 we show these results for two PAGD generators having an identical electrode area (128 cm^2) and connected to electrical energy capture circuits of three separate configurations as set forth in FIGS. **10A, 10B** and **10C** and column 2, Table 7. In the configuration of FIG. **10C**, or double diode configuration, both electrode plates act as cathodes and the axial member as the anode collector (experiments 1–4, for the **H220** device and 13–14, Table 7, for the **H34** device). In the configuration of FIG. **10B**, or triode configuration, one plate acts as the cathode, the axial member as an auxiliary cathode and the other plate as a collector (experiments 5–9, Table 7). In the configuration of FIG. **10A** or single (plate to plate) diode configuration, the axial member is disconnected, and the polarity of the plates remain as in the triode configuration (experiments 10–12). All measurements were taken after 1 minute of PAGD operation of the devices, which were, at the start of each run, at room temperature. All cathodes had been previously broken in with $>2*10^6$ AGD pulses. The open circuit voltage of the charge pack was, for all cases, at 359 to 365 volts, before each test. The direct measurements of the PAGD input and output DC voltages and currents were obtained as statistical means of 10 second long measurements, and at no time did the standard error of the plate voltage mean exceed 35 volts.

The air pressure within the tube during these tests is shown in column 3, Table 7, the drive pack DC voltage (X), in column 5, the DC voltage across the plates (Y), in column 6, the drive pack output current (PAGD input current), in column 7, and the drive pack total watts output is shown in column 8. Columns 9 and 10 show the PAGD voltage (PAGD $V=(X-Y)/I_{av}$) and the value of the PAGD extinction potential in V/cm. The recovery co-ordinates (ie the PAGD output energy) found at the **U1–U2** output (FIG. **9**), are shown in columns 11 to 13, as the charge pack's E1–E2 input DC voltage, amperage and power watts, respectively. The calculated resistance of the entire circuit is given in column 14, the registered PAGD frequencies in column 16, and running conditions in columns 17 to 18. The breakeven efficiency obtained by direct comparison of the electrical power figures for the drive and charge packs, respectively, is given in column 15. This assumes, for purposes of a generalization of power production rates over time, that the quasi-instantaneous, direct measurements here obtained can be translated to outputs obtained per unit time, and thus into direct Watt-hour measurements.

Data from runs 1 through 4 demonstrate that, at these PAGD frequencies, there is no difference between using fast switching (32 nanoseconds) MUR 860 diodes, or regular 40HFR-120 silicon diodes, in the rectification bridge of the electrical energy capture circuit, and that the PAGD frequency varies as a function of decreasing air pressure.

Runs 5 to 14 show that, in general, for the same tube, the single and double diode configurations are the most efficient, for the same pressure, the diode configuration typically yields $\sim$1.5-2x larger breakeven efficiencies (cp runs 10–11 and 13–14, with runs 5–9, Table 7). The largest accumulations of power are also registered in the diode mode(s). This trend appears to be a function of the much lower cathodic work-function of the aluminum plates, than of the tungsten of the axial member utilized as an auxiliary cathode in the triode configuration. A feature of the data from these 14 different runs is the consistent excess power outputs (column 15, Table 7) and their narrower range (218 to 563%), when compared to those observed with the previous two methods of experimental analysis.

Run 12, Table 7, shows that the switching on of the electromechanical arm can be performed without entailing a power loss in the PAGD capture circuit, as previously found for run 5, Table 5, utilizing the open circuit voltage method. In fact, with C4=8 μF and R4=500 ohms, the AC induction motor behaves as an electrical flywheel (eg. 2800–3000 rpm for 10 PPS inputs), while the electrical energy capture circuit still registers a sizeable excess electrical power production (compare runs 11 and 12, Table 7). Runs 13 and 14 illustrate how the charge pack's state of charge and its inherent capacitance affects both the PAGD frequency and the power producing efficiency of the entire system: as the charge pack is reduced from 29 to 19 cells, the PAGD generator adjusts by reducing its frequency logarithmically and, while the charge pack input current is greater than before, the drive pack loss becomes still larger and the breakeven efficiency much lower (by $>\frac{1}{2}$, from 563% to 228%). This is because the circuit must translate the naturally larger PAGD amplitude into a larger surplus of output current, and in this process becomes less efficient.

If the first measurement method employed (the open circuit method) had to make too many theoretical assumptions about the system's performance under load conditions and hence about its effective charge capacity, the second approach still had to suppose an invariant discharge time and thus an invariant absolute charge capacity on the part of the battery systems (charge packs) employed for capture which it approximated by an operation of integral calculus. With the third method described above, theoretical assumptions were avoided except that, in these measurements, the actual performance of a given battery in terms of time, time of delivery and time of capture, was also ignored; no account is taken of the time-dependent modulation of the PAGD frequency, as effected by certain of the parameters analyzed, namely the charge pack state of charge, the method of sequencing the PAGD runs (continuous vs interrupted) and its concomitant heating effects, and the state of charge (load voltage and current capacity) of the drive pack. A simple, non-negligible, resistive measurement of power lost by the drive pack, and an identically non-negligible measurement of the power gained by the charge pack, for the same experiment and the same singular time of PAGD production, were performed repeatedly to corroborate the previous three approaches. For this purpose, all experiments were designed as a continuous series of sequential phases:

1) before a PAGD run, a resistive discharge was measured across either pack over periods of 1 to 3 hours (utilizing the DP and CP resistances previously reported in the open voltage section) and followed by a 15 to 30 minute open circuit voltage relaxation;

2) then, the PAGD runs were performed, either continuously or as interrupted, composite sequences, and the corresponding open circuit relaxation voltage(s) were measured, after the cessation of the integral PAGD run;

3) finally, resistive discharge measurements, obtained under identical conditions to those recorded before the PAGD run, were carried out for either pack, followed by concomitant battery voltage relaxation rate measurements.

Under these experimental conditions, exact power measurements could be taken from an analysis of the actual battery discharge curves before and after the PAGD run. Based on a comparison of the curve trends of the pre-run resistive discharge of the drive pack with those of the post-run resistive discharge, the effective power drawn (ΔE_c) from the withdrawable power capacity of the drive pack incurred during a PAGD run, was ascertained. This represents the power consumption during the run, and the experimental value thus recorded constitutes the actual power figure that must be matched for breakeven ( ) to occur. Hence, the breakeven value equals, by definition, the electrical energy input to the system. Similarly, a comparison of the charge pack pre-run and post-run resistive discharge curve trends identified the effective power (ΔE_ρ) added to the withdrawable capacity of the charge pack. This quantity represents the electrical energy recovered during the run. The relation for the two quantities is expressed by the breakeven efficiency (BE= %) equation:

$$\% = \Delta E_\rho / \Delta E_c \cdot 100$$

If the breakeven efficiency is less than %=100, then the apparatus registers a net loss in electrical energy in the CP with respect to the DP. Conversely, if %>100, then there is a net gain in electrical energy in the CP, as compared to that lost in the DP. For purposes of this analysis, a limit to the minimum withdrawable capacity was placed, from experiment and in agreement with the load current curves of the manufacturer, at 115 W for the driver pack (average current of 0.250 A, minimum current of 0.230 A), and at 90 W for the charge pack (average current of 0.375 A, minimum current of 0.334 A), as a function of both their total cell size (respectively, 46:29) and the difference in the resistive loads employed for the discharge measurements. All cathodes had been broken in, as described before.

The results obtained with this fourth method, for six selected experiments with three diverse types of devices (using different electrode plate areas, gap lengths, and electrode work-functions), configured both in the triode or the (single) diode (e.g. FIG. **10**B) arrangements, at the indicated pressures, are presented in Table 8. In all cases, a net excess of combined battery pack charge, expressed as electrical watt hours, is registered (columns 8 and 10, Table 8) and the breakeven efficiencies are all >100% (column 10). Experimental groups #**1** and #**2** again demonstrate that, for the same cathode, the interrupted PAGD sequence method of group #**2** (1 minute of PAGD function, followed by a 15 minute relaxation, and so on) yields a higher breakeven efficiency because of the lower losses registered with this minimal plate heating method (column 10, Table 8). Group #**3**, Table 8, shows that the PAGD power production efficiency is also higher for a lower work-function cathode material (**H220** vs **H34**), being subjected to PAGD auto-electronic conditions at a 4-fold lower pressure than the control groups #**1** and #**2**; however, the lower pressure depresses the frequency and, together with the interrupted PAGD sequencing method, it also lowers the loss, causing an actually much larger breakeven value than registered for the previous two groups. Groups #**4** and **5** exemplify the dual effect of lowering both the plate area and the gap distance: the former affects the PAGD event frequency, whereas the latter affects the PAGD amplitude, and thus the capture efficiency of the charge pack. Despite a cathodic work-function practically and operationally identical to that of groups #**1** and **2**, these smaller plate area and shorter gap devices utilized in groups #**4** and **5**, yield 3- to 6-fold lower net power outputs, as well as lower breakeven efficiencies, than the former groups, at the same pressure. Finally, group #**6** exemplifies the results obtained for the plate diode configuration, where the frequency is lower (no triggering role for the axial member), and a higher loss leads to the lower breakeven efficiency, comparable to that of the lower area and shorter gap groups #**4** and **5**.

In order to verify the discharge curve lengths employed in these analyses and experimentally establish the actual charge capacity of the battery packs, calibration resistive discharges, between the maximum charge state and the minimum limits chosen, were performed for each pack with their respective discharge resistances R2 and R3 (see FIG. **9**). These discharge calibration curves were plotted for half maximal charge values shown in FIGS. **15**A and **15**B, and from the curve produced, we have determined the total half-charge capacities of each battery pack to be 1.033 kWh (100%=2.066 kWh) for the drive pack and 660 Wh (100%=1.320 kWh) for the charge pack. Based upon the corresponding maximal (100%) capacity values, we determined the actual percentages of the relative charge capacities shown in column 5, Table 8, which correspond to the experimental values obtained. We also noted that the curves plotted showed two quite distinct time linear slopes, the slope of the delivery of power per time unit steepening very markedly at the approach to the limits of the permissible withdrawable capacity, occurring at 115 W into R2, and 90 W into R3.

The pre-PAGD run and post-PAGD run, drive and charge pack discharge curves corresponding to groups #3 and #6, respectively for triode and plate diode configurations, in Table 8, are shown in FIG. **16** (drive pack) and **17** (charge pack), for group #3, and in FIG. **18** (drive pack) and **19** (charge pack), for group #**6**. In all cases, the open symbols represent the pre-PAGD run discharge curves, whereas the closed symbols represent the post-PAGD run discharge curves.

As a further check on these values, a videographic, millisecond analysis of the singular power simultaneities occurring at both ends of the system (drive and charge packs) was performed for various 10 second samples of diverse PAGD runs. A typical example is shown in FIG. **20**, which is a sample of the PAGD run designated as #**6** in Table 8. Whereas the drive pack DC wattage spent as input to PAGD production varied from 36.6 to 57.82 watts, by a factor of 1.6x, the DC wattage entering the charge pack as captured PAGD output varied more pronouncedly by a factor of 2.7x, from 146.4 to 399.6 watts (all meters were in the same selected ranges of voltage and current) with the semi-periodic, intermittent character of each singular emission, though within specific, ascertainable ranges for both amplitude and current outputs. Assimilation of the singular behaviour of the PAGD in this sample, by a statistical treatment of its variation (n=64), indicates that the operational breakeven efficiency observed during this sampled period lies at 485.2%±18% with projected 48.3Wh drive pack loss and 221.7Wh charge pack gain. This matches rather closely the observed 483% breakeven efficiency, and the 37.7Wh loss as well as the 182.2 kWh gain for the overall PAGD run reported in group#**6**, Table 8, and indicates how close are the values obtained by the operational and extensive non-negligible resistive discharge power measurement methods employed.

Finally, an example of the correlation between the drive pack PAGD load voltage and the charge pack PAGD charging voltage, as a function of the duration of the intervening PAGD run between resistive discharge measurements, is shown in FIG. **21**, for the PAGD run corresponding to group #**4**, Table 8.

Using the same pulse generator with H200 AL 128 cm^2 plates, in a double diode configuration, and the same circuit values (but with CP=23 cells), three experiments were conducted at different PAGD frequencies, as a function of varying air pressure. Analysis of driver pack losses and charge pack gains by the extensive load discharge measurement method, as described before, led to the determination of the gross and net gains (respectively, without and with losses included) per pulse, in milliwatt-hour, for each frequency, as well as of the gross and net power gains per second of PAGD operation. The results are shown in Table 9. Even though the gross and net gains of power per pulse were observed to increase with decreasing frequency, the gross power gain per unit time increased with increasing frequency. However, this last trend does not necessarily translate into a higher net gain per unit time, because the losses in the driver pack (not shown) also increase significantly with PAGD frequency. These losses are in all probability related to more energy retention by the plasma at higher frequencies when plasma extinction becomes incomplete. We expect net gains to reach optimal thresholds for any given type of circuit configuration set of values and pulse generator dimensions.

Certain additional observations made during experiments with the double diode configuration of FIG. **10**A may assist in understanding of the invention.

1) Replacing residual air with argon gas leads to higher PAGD frequencies, as noted by us when utilizing a 128 cm^2 H200 AC plate pulse generator in the double diode configuration ($V=575$). At 1 Torr, the pulsation rate went from 20 PPS in air to 1300-400 PPS in argon. With 29*12 v cells in the charge pack, input currents ceased to flow into it. Under these conditions, the tube potential across the plates decreased and the drop across the input resistor increased. The value of $E(=V/d)$ became smaller (gap size = 3 cm from plate to axial anode collector), as the extinction voltage decreased.

2) With frequencies of 400 PPS, the currents flowing into the charge pack fell to zero. Replacing a fast-recovery type HFR 120 (1200v, 40A) diode bridge by a type MUR 860 (600v, 8A) diode bridge had no effect. When the amplitude of plate potential oscillations falls below the potential of the charge pack, there is also a tendency to produce arc discharges. For output currents from the vacuum pulse generator to enter the charge pack, the number of cells must be reduced so that the potential of the charge pack is low enough to admit the transduced currents. A reduction from 29 to 23 cells allowed currents of 250 mA to enter the CP, and further reduction to 19 cells doubled these currents (per polarity arm).

3) Our observations show that it suffices under these conditions (CP = 19 cells) to increase the vacuum, so that the frequency decreases, and the plate potential and the charge pack input currents increase. At 0.1 Torr, the currents reached 1A D.C. per plate, and at 0.05 Torr, 2A D.C.

The interconnection between these factors indicates that the extinction voltage is a function of the PAGD frequency: the higher the PAGD frequency, the lower the extinction voltage, until empirical (in distinction from predicted) VAD field values are reached. As a consequence, the start voltage of the charge pack must be adjusted, by varying the number of cells composing it, so that it lies below the lowest extinction voltage of the PAGD, for any given geometry and gap distance.

Secondly, as the ion plasma is made more rarefied, the frequency of the emissions decreases, but the peak values of the output voltage and current per pulse increase. The slower the PAGD and the more rarefied the atmosphere, the higher is the output energy produced by the system relative to the input energy.

Autographic analysis of PAGD-induced cathode craters in **H34** plates was performed, and their average inner diameter and maximal depth were determined. Similar studies were performed for PAGD-induced craters in Alzak (trade mark) plates. The secondary craters characteristically found in Alzak plates, along fracture lines irradiating from the main crater, are absent in **H34** plates; instead, in **H34** plates, one observes a roughened surface surrounding the emission crater, quite distinct from the original rough aspect of the pulled finish of these hardened aluminum plates. Also unlike the Alzak main craters, the **H34** craters often have a convex center occupied by a cooled molten metal droplet, whereas the Alzak craters had a concave, hollowed out aspect. Eventually, as the pitting resulting from PAGD cathodic emissions covers the entire cathode, the metallic surface gains a very different rough aspect from its original appearance. In this process, craters from earlier metal layers become progressively covered and eroded by subsequent emissions from the same cathode. Altogether different is the surface deposition process occurring at the anode; here, the surface appears to become more uniform, through the mirroring and possibly abrasive actions of cathode jets. Macroscopically, with increased periods of PAGD operation, the anode surface appears cleaner and more polished.

With the data obtained by the metallographic method of crater measurement, we estimated the volume of metal ejected from the cathode, by assuming that the crater represents a concavity analogous to a spherical segment having a single base ($1/6\pi*H\ [3r^2+H^2]$, where H is the height of the spherical segment and r the radius of the sphere), while disregarding the volume of the central droplet leftover from the emission. The following are mean ±SEM crater diameters (D), crater depths (H) and maximum volumes (V) of extruded metallic material for two types of aluminum cathodes, Alzak and **H34** hardened aluminum, subject to a high input current PAGD:

1- Alzak: D—0.028 cm±0.003; H—0.002 cm±0.0002; V—$6.2*10^{-7}$ cm^3;

2- H34: D—0.0115 cm±0.0004; H—0.0006±0.0001; V—$3.1*10^{-8}$ cm^3;

Accordingly, utilizing plates composed of either material with 3 mm of thickness, and thus with a volume of 38.4 cm^3 per plate and considering that only ⅔rds of the cathode shall be used (a 2 mm layer out of the 3 mm thickness), the total number of pulses per plate total (TLT) and partial (PLT) lifetimes is theoretically:

1- Alzak: TLT: 6.2*107 pulses; PLT: $4.1*10^7$ pulses;

2- H34: TLT: $1.2*10^9$ pulses; PLT: $8.1*10^8$ pulses.

Typically, an H34 device can produce ~0.25 kWh per 10,000 pulses. The corresponding value for a PLT is thus a minimum of 1.0 MWh/Alzak cathode and of 20 MWh/H34 cathode. As the cathode for each combination is only 66.7% consumed, the vacuum pulse generator may continue to be used in a reverse configuration, by utilizing the other plate in turn as the cathode; thus, the estimated minimal values become, respectively, 2.0 MWh/Alzak pulse generator and 40 MWh/H34 pulse generator. The same rationale applies for the double diode configuration of FIG. **10C**.

We have created a two-ported system for the production of the singular discharge events which we have previously identified in the '863 application as an endogenous pulsatory abnormal glow discharge regime where the plasma discharge is triggered by spontaneous electronic emissions from the cathode. We have examined the functioning of this two-ported system in order to determine what were the electrical power input and output characteristics of a sustained PAGD regime. Despite the wide (10-fold) variations in net power and breakeven efficiencies measured by the four different methods employed (open voltage measurements, time integration of negligible power measurements, operational power measurements and real time non-negligible power measurements), all methods indicate the presence of an anomalous electrical transduction phenomenon within the vacuum pulse generator, such as can result in the production at the output port of electrical energy measured and directly captured which is greater than would be anticipated having regard to the electrical energy input at the input port. With the most accurate of the methods employed we have found typical PAGD power production rates of 200 Wh/hour of PAGD operation, and these may reach >0.5 kWh/h values.

The discrepancies between the methods utilized have been extensively examined in the preceding section.

Our systematic approach demonstrates that the most frequently employed method of measuring the charge capacity of batteries by the open voltage values is the least reliable approach for the determination of the actual net power lost or gained by the battery packs used in the system: when compared to all three other methods, it overestimates net power consumed and produced by up to 10 fold, as well as distorting the breakeven efficiencies, particularly at the extremes of operation. All this results from the grossly diminished (50–60% of manufacturer's theoretical estimate) effective charge capacity of the lead acid gel cells employed, as determined experimentally from FIGS. **18** and **19**, when compared to the theoretical maximal charge capacity values that serve as scale for the open voltage measurements. In other words, the effective energy density of the batteries during these experiments was in fact approximately half of the manufacturer's estimated 30 Wh/kg.

Under these actual conditions of battery performance, the third and fourth methods (respectively, operational and real-time non-negligible power measurements) of power consumption and production proved to be the best approach to measure both PAGD electrical power input and output, as the results of both methods matched each other closely, even though the former is a statistical treatment of simultaneous events and the latter is a real time integration of their cumulative effects. The second method is clearly less reliable than either the third or the fourth methods, and this stems from the fact that the power consumption slopes of negligible resistive discharges not only are very different from the quasi-steady state discharge slopes (beginning at >5–15 minutes) of extensive resistive discharges, but also their proportionality may not reflect the real time proportionality of equivalent prolonged resistive discharges.

The main advantage of the fourth method is that it effectively takes into account the actual time performance of the batteries comprised by the overall PAGD production and capture system we have described. As such, the method may have the main disadvantage of reflecting more the limitations of the batteries employed (their high rate of degradation of the absolute value of total effective charge capacity, and limited efficiency in retaining charge derived from discontinuous input pulses) than indicating the actual power output. There are a number of possibilities for fine tuning of the system introduced by the present work, beginning with the utilization of secondary batteries or other charge shortage or absorption devices that have less variable or more easily predictable actual charge capacity. In this respect, there are two major shortcomings to the batteries used to form the drive and charge packs; (1) their significant memory effect and (2) their design for constant, rather than discontinuous, DC charging. Recently developed Nickel Hydride batteries are an example of an electrostatic charge-storage system that lacks a substantial charge memory effect, and their experimental batteries are being developed currently for higher efficiency intermittent charging methods. Electrostatic charge retention systems having better energy densities, better charge retentivities and insignificant memory effects will probably be more efficient at capturing and holding the energy output by the circuit. In practical embodiments of the invention, effectiveness in charge utilization will be more important than measurability, and any device that will use the energy effectively whilst presenting an appropriate back EMF to the system may be utilized.

The effect of the performance characteristics of the drive and charge packs is only one amongst many parameters affecting operation of the invention. As shown by our extensive investigation of the diverse PAGD phenomenon the recovery of energy from it by electromechanical transduction as in the '531 application, or electrostatic capture as described above, the factors involved in modulating the frequency, amplitude and peak current characteristics of the PAGD regime are complex. Manipulation of these factors can improve electrical energy recovery, or reduce it or even suppress PAGD. We have so far noted numerous factors that affect PAGD frequency and some amongst those that also affect the PAGD amplitude. Aside from these factors, the circuit parameters of the output port portion of the circuit, in addition to the nature and chemical characteristics of the battery cells already discussed, the charge potential of the charge pack, the characteristics of the rectifiers in the recovery bridge in relation to the period of PAGD superesonant frequencies, and the effective values of the parallel and transversal capacitance bridges can all influence the results achieved. Certain factors however have a radical effect on PAGD operation, such as the gap distance and the charge pack potential. Too small a gap distance between the cold emitter (cathode) and the collector will result in an increasing reduction in energy recovery. The potential presented by the charge pack must be less than the voltage amplitude developed by the PAGD, as specified by a given gap distance at a given pressure. Too large a charge pack size with respect to PAGD amplitude and the gap length will preclude PAGD production or result in extremely low PAGD frequencies. In brief, the energy absorption rate and the counter potential presented by the charge pack or other energy utilization device are important factors in the operation of the circuit as a whole, and should either be maintained reasonably constant, or changes should be compensated by changes in other operating parameters (as is typical of most power supply circuits).

Since our test results indicate that the electrical power output of the circuit can be greater than the electrical power input to the circuit, the circuit clearly draws on a further source of energy input. Whilst we do not wish to be confined to any particular theory of operation, the following discussion may be helpful in explaining our observations. These observations have been discussed in some detail so that the phenomenon observed can be reproduced, even if the principles involved are not fully understood.

In the '863 and '531 applications we have identified a novel, cold-cathode regime of vacuum electrical discharge, which we have termed the pulsed abnormal glow discharge (PAGD) regime. This regime, which occupies the abnormal glow discharge region of the volt-ampere curve of suitable discharge tubes, has the singular property of spontaneously pulsing the abnormal glow discharge in a fashion which is endogenous to the tube and its circuit environment that constitutes a vacuum pulse generator device, when it is operated under the conditions we have identified. In fact, when stimulated with continuous direct current, in such conditions, such a circuit responds with spontaneous abnormal glow discharge pulses that enable effective segregation of input and output currents. We have demonstrated electrically, metallographically, oscillographically and videographically, how the pulsed discontinuity results from a self-limiting, autoelectronic cathode emission that results in repeated plasma eruptions from the cathode under conditions of cathode saturated current input. The auto-electronic triggering of the PAGD regime is thus akin to that of the high-field emission mechanism thought to be responsible for vacuum arc discharges (VAD regime). However, under the PAGD conditions we have defined, this mechanism is found to operate in the pre-VAD region at very low field and low input average direct current values, with very large interelectrode distances and in a self-limiting, repetitive fashion. In other words, the PAGD regime we have identified has mixed characteristics: its current versus potential (abnormal glow) discharge curve is not only distinct from that of a vacuum arc discharge, but the electrical cycle of the PAGD regime itself oscillates back and forth within the potential and current limits of the abnormal glow discharge region, as a function of the alternate plasma generation and collapse introduced by the discontinuous sequencing of the auto-electronic emission process. Accordingly, the intermittent presence of the abnormal glow, as well as the observed segregation of the current flows, are due to the diachronic operation of these spontaneous cathode emission foci. The micro-crater and videographic analyses of the PAGD have demonstrated the presence of an emission jet at the origin of each pulse, a phenomenon which VAD theory and experiment has also identified. Metallic jets originating at the cathode spots of VADs have been known to present velocities up to, and greater than 1000 m/sec.

In light of the above, the energy graft phenomenon we have isolated would have to be operated, at the micro-event scale, by the interactions of the cathode emission jet with the vortex-formed impulse-transducing plasma in the interelectrode space. Several aspects can be approached in terms of the complex series of events that constitute a complete cycle of operation, on a micro-scale. There are interactions within the cathode, interactions at the cathode surface, interactions between the emission jet and the plasma globule close to the cathode, and finally, interactions of the resulting electron and ion distributions in the interelectrode plasma, within parallel boundaries.

In general, in the presence of an electrical field, the distribution of potential near the cathode forms a potential barrier to the flow of electronic charge, as this barrier is defined by the energy that the most energetic electrons within the metal, the Fermi energy electrons, must acquire before freeing themselves from the cathode surface potential to originate an emission jet. Before any free electrons become available for conduction in the space adjoining the cathode, they must cross the boundary posed by the potential barrier. With a weak applied field, classical electron emission from a metal can only occur if an energy practically equal to the work-function of the metal is imparted in addition to the Fermi energy. Under thermionic conditions of emission, the heating of the cathode provides the needed energy input. However, the cold-cathode Fowler-Nordheim quantum-field emission theory predicted the existence of a finite probability for an electron to tunnel through the potential barrier, when the applied field is high. Cold-cathode electron emissions are thus possible, under these conditions, at practically Fermi energy levels, as the high field would catalyze the tunnelling through the potential barrier by narrowing the barrier width for the Fermi energy electrons. The exact localization of the emission would then depend on the randomized fluctuations of high fields at the cathode, which were produced by positive space charges sweeping in proximity to it. For most purposes, this theory has been the working hypothesis of the last 60 years of field emission studies, which have centered upon the VAD mechanism, despite the fact that observed field gradients are evidently inadequate to explain breakdown as a function of the theoretical high field mechanism. The Fowler-Nordheim theory has therefore suffered major revisions and additions, mostly to account for the fact that it postulates, as a condition for cold-cathode field emission in large area electrodes, the presence of enormous fields ($>10^9$ V/m) and extremely low work functions, neither of which are borne out by experimental VAD investigations. Some researchers have found that the breakdown responsible for the VAD field emission is promoted by Joule heating and vaporization of microscopic emitter tips, and that this requires a critical current density (10^{12} A/cm^2), while others emphasized that this explanation and these thresholds did not hold for large area emitters and that a space charge effect of concentrating the ion distribution near the cathode promoted breakdown under these circumstances, when the field reached a critical value; large field enhancement factors (>1000-fold) have been postulated to explain the discrepancy between theoretical predictions and experimental findings regarding the critical breakdown field values, and others have demonstrated how this critical field value effectively varies with work-function and electrode conditioning.

The PAGD regime and its self-extinguishing autoelectronic emission mechanism stands as an exception to the high field emission theory as it currently stands with all its modifications, especially given that in this phenomenon we are confronted.. with a cathode emission that spontaneously occurs across the large gaps in large plate area pulse generators, at very low field values (down to $<1*10^4$ V/m), as shown above and in the '863 application. Moreover, a Fowler-Nordheim plot (in the form $\text{Log}_{10}(I/V^2)$ vs $1/V$) of the PAGD volt-ampere characteristic exhibits a positive slope, rather than the Fowler-Nordheim negative slope characteristic of VAD field emission. However, current density values obtained from correlations of autographic analysis of the cathode with an analysis of event-oscillogram (peak pulse currents), indicate that the PAGD current density J may reach values of 10^5 to 10^7 A/m^2 during the emission process (the larger Alzak craters have an associated lower J value), values which, at the upper end, do not reach the 10^9 A/m^2 current density threshold required by the Fowler-Nordheim theory. Considering these two distinct observations with regards to field strength and current density, we have to admit the existence of a low field, large area cold-cathode auto-electronic emission endowed with high current densities, which is not predicted by current field emission theory.

Unlike the typical VAD regime, the PAGD is neither a high frequency oscillation, nor does it occur in a random fashion. It constitutes a semi-regular, quasi-coherent, periodic energy transduction which cycles between cathode drop limits that are higher by a factor of 2–15 than typical vacuum arc cathode drops. The intermittent cathode emission responsible for the low frequency, pulsed behaviour of the abnormal glow, is also self extinguishing and self-starting, under the conditions we have defined. Furthermore, we have also identified a novel and unexpected dependency of the periodic pulse rate upon the cathode area. This indicates the presence of field emission control parameters heretofore unsuspected. It is likely that field fluctuations of the polarized pre-breakdown field is responsible for eliciting the particular localizations of the auto-electronic emission foci, as well as what imparts, in a lens-like fashion, the distorted field energy needed for electron surface release. In this sense, external, electrical or magnetic field fluctuations (e.g. motion of static charges or of constant magnetic fields) induced by us at pre-breakdown potentials, provoked PAGD emissions and breakdown at these levels.

In general, VAD studies have shown that, for large area electrodes, microgeometry, adsorbed gas layers and gas impurity contents of the cathode play a role in modulating field emission. In our PAGD studies, the interactions at the cathode surface and across the cathode potential drop are clearly modulated by: (1) the nature of residual gases, as shown by our air vs Argon studies; (2) their pressure, (3) electrode conditioning, (4) work-function and (5) cumulative pulse count, amongst others.

The plasma, in leak-controlled or low pressure PAGD devices, has both residual gas and metallic vapor substrates. In devices initially closed at high to very high vacua (diffusion pump pressures), the major residual substrate, whose presence increases with time of operation, is the metallic vapor released from the cathode and not impacted onto the envelope walls or the anode. It has been previously shown for externally (magnetically or electrostatically) pulsed plasma accelerators, that the amount of residual gas or vapor left in the interelectrode space diminishes with increasing number of consecutive discharges and a growing amount of electrode-insulator absorption of gas. The effect of such removal of residual gas or vapor is to decrease the vacuum of a sealed envelope. With high vacuum sealed PAGD generators we have observed that prolonged operation and sputter-induced mirroring of the envelope causes a progressive disappearance of the discharge, as the voltage potential needed to trigger it also increases. At the thermocouple, low frequency pulsed abnormal glow discharges can also be seen to increase the vacuum significantly. These results suggest instead the presence of a pumping mechanism in the PAGD which is somewhat analogous to that of sputter ion pumps, where collision of ionized gas molecules with the cathode is responsible for the sputtering of cathode material that either combines with the gas substrate ('gettering' action) or 'plasters over' the inert gas molecules onto the anode (a process known as 'ion burial'). These are the two basic pressure reducing actions of sputtered getter atoms, in ion pumps. However, in ion sputter pumps, the initiation of the cycle is a function of the presence of high velocity electrons in the high field plasma of the glow discharge, which are necessary to ionize the gas substrate molecules; also, the getter material typically has a high work-function for field emission. Hence, the sputtering is due to the secondary impact of plasma positive ions at the cathode, after plasma ionization has occurred in the interelectrode space. Altogether different is the mechanism of spontaneous, primary electron emission from the cathode, which is characteristic of the low field PAGD: here, the sputtering is caused by the electronic emission itself and attendant metallic vaporization processes. By artificially confining the firing foci to a part of the cathode, we have shown in the single diode configuration how the PAGD induced sputtering is associated with the cathode autoelectronic emission mechanism, rather than with the abnormal cathode glow per se, given the localization of sputtering onto the emission region of the plate, despite its overall cathode glow saturation.

These observations would thus seem to corroborate the hypothesis of a progressive vacuum increase with the cumulative number of emitted pulses, were it not for the fact that experiments performed with leak controlled devices (reported here and in previous studies) show that, when the negative pressure is maintained by balanced leak admission of air or argon, pulse rates still decrease with cumulative pulse count, and do so neither as a function of an increase in vacuum, nor as a function of envelope mirroring (unless this is so extensive as to establish envelope conduction), but rather as a function of processes (generally referred to as conditioning) inherent to the electrodes, specifically, to the cathode. We have further shown that, for such altered emitter states, the pressure of the vessel must be increased, not because of an increasing vacuum (precluded by the controlled gas leak), but because of the effect that residual gases may have in modulating the low field PAGD emission.

PAGD electrode conditioning is a cathode-dominant process resulting from the cumulative emission of high numbers of pulses by a cathode, and has been shown to be a factor independent of the nature and pressure of the residual gas and partially reversible only by operation with reversed plate polarity, unlike reports of copper cathode-dominant conditioning. It is thought that electrode conditioning and the accompanying increase in VAD breakdown potential are due to the progressive adsorption of residual gases, though cathode-dominant conditioning processes, such as subjecting the vacuum gap to consecutive discharges, have been shown to correlate the decrease in plasma impulse strength with electrode outgassing of absorbed or adsorbed gases. Moreover, given the pitting action of crater formation at the cathode by the PAGD regime, and, as we shall see below, the metallic plating of the anode, the PAGD cathode-dominant process of conditioning we have observed with respect to decreased pulse frequency and increase in potential, suggests that the apparent increase in cathode work function is not due to gas adsorption or absorption. These processes are more likely to occur on the plated anode. It is likely that, given the observed PAGD pressure reducing effect caused by the cathodic jet, a certain outgassing of the cathode is in fact occurring during PAGD function. One might also expect that the anode, if plated by sputtering atoms, would increase its gas content in the formed surface film. However, controlled leak experiments suggest instead that some other type of alteration of the cathode work function is occurring, which is, as we shall examine below, independent of the adsorbed gas state of the electrodes, as well as independent of the PAGD ion pump-like effect. Nonetheless, even at the level of the anode, the PAGD sputtering action may have contradictory effects: it may impact interelectrode gap molecules onto the collector, as well as release, by ionic bombardment and vaporization, gases adsorbed to, or contaminating the anode. If we assume that gas adsorption by impact on the collector is the predominant mechanism, one could explain the increase in the number of breakdown sites per unit time, as observed by us for a re-reversed cathode, if the number of PAGD breakdown sites depended on the quantity of adsorbed gases, eg oxygen, on the cathode being tested. Recovery of the cathode work-function would depend on the electronic charge recovery of the positively charged, adsorbed or occluded gas layer at the cathode- either by reversal or as a function of time of inactivity. The surface film theory of 'electrical double layer formation at the cathode' in fact contended that, low field flash over is a photocathodic effect dependent upon the presence of a glowingly positively polarized gaseous film at the cathode; this film would lower the cathode emissivity by decreasing the field between the cathode surface and the leading edge of the cathode glow, across the cathode drop. However, even though the surface film theory of 'electrical double layer formation at the cathode' predicts the lowering of the emission breakdown potential and the increase in flash over rate when the electrodes are reversed—as the anode would have acquired a surface charge capable of affecting the breakdown potential, it acknowledges nevertheless, that the anodic surface charge hardly explains the observed intensity of the polarization effects. Moreover non-reversed, conditioned cathodes retained their lower PAGD frequencies in a time independent manner, for as long as reversal was avoided (excluding a PAGD frequency recovery effect due to plate cooling, which may be as short as 15 minutes). PAGD conditioning was independent of idle time and increased with cumulative pulse count. Moreover, the AGD pulses are not UV photocathodic Townsend discharges, liberating secondary electrons via positive ion impact at the cathode. Nor could photocathodic emissions generate currents of the magnitude observed in the PAGD. Lastly, the PAGD discharge and breakdown thresholds appear to be unaffected by UV, though they may be somewhat depressed by visible light, and the emission mechanism in the PAGD is the primary process.

Removal or flattening of protuberances and tips from the emitting cathode by the action of the discharge, is a process also thought to play a role in hardening the cathode or increasing its field emission work-function. However, this explanation may not be adequate for the PAGD emission process, if we consider our metallographic findings of a smoothing action of the discharge at the collector. In fact, it would appear that the flattened, smoother, plated, mirrored and cleaner surfaces subjected to PAGD bombardment are the explanation for the observed increased emission ability of re-reversed cathodes: mirrored Alzak surfaces emit at higher frequencies than do dull H34 and H220 surfaces; new, polished surfaces emit at a higher frequency than do pitted, broken in surfaces; anode surfaces, never before utilized as cathodes but subjected to prolonged PAGD action, emit at higher frequencies when employed as cathodes, than do new, identical cathode surfaces; and ex-cathodes, employed for prolonged periods as anodes, regain a higher emission frequency upon re-use as cathodes. The better PAGD emission performance of smoother cathodes, compared with the worse VAD emission performance of the same, when pitted cathodes (lacking protuberances) are used, requires explanation.

Rakhovsky has put forth a VAD model for cathode spots, that distinguishes between Type I spots (quickly moving spots, far from steady state and responsible for crater formation), and Type II spots (quasi-stationary and near steady-state, but leaving an itinerant track with no sign of crater formation). Whereas the former would obey the Fowler-Nordheim requirement for high fields ($>10^9$ V/m), the latter could hardly be expected to do so with typical arc voltage drops in the order of 10 V. Once again, autographic analysis of the PAGD emission aspect indicates mixed characteristics: the PAGD cathode spot is a hybrid. It behaves as an intermittent instability that leaves single (e.g. in H34) or clustered (e.g. in Alzak) craters, which are both qualities of Type I cathode spots; and it exists under low field conditions ($<10^5$ V/m), with cathode drops of 20 to 150 V, in a quasi-coherent mode, leaving an itinerant track of successive craters when operating at the higher frequencies, all of which are properties approaching those of a VAD Type II cathode spot. Furthermore, the macroscopically visible metal sputtering (due to the explosive action of the PAGD emission phenomenon) occurring at the upper end of the permissible DC current input scale, and the presence of large solidified molten metal droplets in and around the craters, suggest models which have been proposed for explosive electronic emission. Explosion models propose that the creation of a residual plasma ball in front of a microprotuberance provokes the large potential drop at the prospective emission focus and sufficiently high resistive and Nottingham heating to reach $>10^7$ A/cm^2 current densities during the explosive consumption of these microemitters. Whether the explosive action associated with cathode spots is an auxiliary effect that applies solely to the vaporization of the emitting microprotrusion, or an integral emission and vaporization explosive process, it does not appear that it can be restricted to high-field VAD Type II cathode spots, given that it can be equally made to occur with the low field PAGD hybrid cathode spot, and be macroscopically observed. Indeed, in the plate diode configuration, it is easy to visualize the metallic particle explosions that surround and accompany the plasma jets, near to upper current limit conditions. However, if we are to assume that any of these models apply to the emission mechanism, we would, in all likelihood, have to conclude that the PAGD initial emission sites must be submicroscopic (100 to 10 nm), rather than microscopic. Resolution limits to our own metallographic examination of the smoothing action of the PAGD discharge on the collector would thus have precluded us from detecting formation of such submicroscopic protrusions, as well as their presence in a 'soft' cathode- and thus infer their disappearance from a pitted, hardened cathode; but if the disappearance of such submicroprotuberances were responsible for the observed alteration of cathode work function, one would also thereby have to postulate the existence of a mechanism for microroughness regeneration (eg. tip growth) at the anode, in order to explain the observed increased emission upon cathode re-reversal. Furthermore, this regeneration would have to be actively promoted by operation with reversed polarity, and this is problematic. Focusing of the distorted or magnified field upon alumina inclusions on pure iron electrodes has been demonstrated to degrade breakdown voltage for field emission, but the effect was greater for larger microscopic particles. If we were to apply this concept to our work, it would require the existence of unmistakably abundant microscopic heterogeneities in the quasi-homogeneous electrode surfaces employed, which we did not observe; on the contrary, their absence suggests that either the microroughness responsible for the low field PAGD emission is submicroscopic, or that the field distortion responsible for eliciting the PAGD is independent of the presence of these protuberances. This last possibility must be taken all the more seriously, in light of the fact that PAGD functioning is able to cover with craters the entire surface of an emitter.

Whereas the discharge potentials observed in the PAGD have been shown to be relatively independent of the kind of gas present, there is a gas effect in the PAGD phenomenon, particularly in what concerns its frequency, observed when the same 'run down' cathode was capable of much higher emission rates when exposed to argon, than to air. Utilizing the technique of bias sputtering, it has been demonstrated that the number of charge symmetric collisions (dependent upon sheath thickness d and the ion mean free path) in the plasma sheath, which are responsible for lower energy secondary peaks in ion energy distribution N(E), at pressures of 0.2 Torr, is substantially greater in argon than in argon-nitrogen mixtures, and thus that, under these conditions, mostly Ar^+ and Ar^{++} ions impact the negatively biased electrode. In non-equilibrium RF discharges, greater ion densities have also been attained with argon, than with air. With respect to field emissions, one would expect a gas effect only with regards to changes on surface conditions, though such studies have shown contradictory effects of argon upon cathode work function. In light of the foregoing, and given that the PAGD is an emission discharge and not a sputtering discharge per se, in the strict sense, we can conceive of the role of inert gas atoms in increasing, as compared to air or nitrogen, the ion energy density distribution at the PAGD cathode spot interface with the cathode surface emitter, and thus elicit increased emission rates from the cathode, by pulling electrons from the metal via the field effect. While this is consistent with the concept of focused distortions of space-charge field fluctuations inducing localization of the emission foci, the argon effect can be observed in the PAGD regime over the entire range of the Paschen low vacuum curve, and into Cooke's mid to high vacuum curve, at low fields and without negative biasing. Thus, it is not simply a high pressure (nor a gas conditioning) effect, even if the gas effect in question applies to the description of a local pressure rise at the emission site/cathode spot interface, which may play a role in enhancing the local field.

Considered together, the PAGD emission-derived sputtering, the observed metallic plating of the anode and the explosive aspect of the discharge, suggest the presence of a jet of metallic vapor present in the discharge and running, contrary to the normal flow of positive ions, from the cathode to the anode. This jet appears to have properties similar to the high speed vapor ejected from the cathode in a VAD, as first detected by Tanberg with his field emission pendulum (Tanberg, R. (1930), "On the cathode of an arc drawn in vacuum", Phys. Rev., 35:1080) In fact, the VAD high field emission process is known to release, from the cathode spot, neutral atoms with energies much greater than the thermal energy of the emission discharge. This anomalous phenomenon brings into play the role of the reported cathode reaction forces detected in vacuum arc discharges (Tanberg, as above, also Kobel, E. (1930), "Pressure and high vapour jets at the cathodes of a mercury vacuum arc", Phys. Rev., 36:1636), which were thought to be due to the counterflow of neutral metallic atoms, from the cathode onto the anode (charged metallic ions are normally expected to target the cathode). In absolute units of current, this current quadrature phenomenon has been shown to reach, in the VAD regime, proportions of the order of $100*I^2$ (see also the Aspden papers referenced below). Early interpretations attributed this to the cathode rebounding of <2% of gas substrate-derived plasma positive ions hitting the cathode and being charge neutralized in the process, but having kept most of their thermal energy. Tanberg held instead that the counterflow of neutral particles responsible for the cathode reaction force was cathode derived, effectively, that it constituted a longitudinal interaction acting in the direction of the metallic arc jet. However, even though secondary high energy distributions of neutral atoms emanating from the cathode do not have thermal energies, their modal distribution does (Davis, W. D. and Miller, H. C. (1969) J. Appl. Phys., 40:2212) furthermore, the major anomalous atomic counterflow that accompanies the high energy electron flow toward the anode, was shown mass spectrographically to consist predominantly of multiply ionized, positively charged ions of cathode metal, rather than neutral atoms. If this made it easier to abandon the primacy of the rebounding model, it was now more difficult for field emission theorists to accept and explain the observed high energies (ion voltages in excess of the discharge voltage drops) and the high ionization multiplicity associated with these counterflowing positive ions. This field of investigation has indeed been one of the mounting sources of evidence suggesting that there is something amiss in the present laws of electrodynamics. The anomalous acceleration of counterflowing ions, and the energy transfer mechanisms between high speed or 'relativistic' electrons and ions in a plasma (Sethion, J. D. et al, "Anomalous Electron-Ion Energy Transfer in a Relativistic-Electron-Beam-Heated Plasma" Phys. Rev. Letters, Vol. 40, No. 7, pages 451–454), in these and other experiments, has been brilliantly addressed by the theory of the British physicist and mathematician, H. Aspden, who first proposed a novel formulation of the general law of electrodynamics capable of accounting for the effect of the mass ratio factor (M/m') in the parallel (and reverse) motion of charges with different masses, (Aspden, H. (1969) "The law of electrodynamics", J. Franklin Inst., 287:179; Aspden, H (1980) "Physics Unified", Sabberton Publications, Southampton, England). The anomalous forces acting on the counterflowing metallic ions would stem from their out of balance interaction with the emitted high speed electrons, as predicated by the electrodynamic importance of their mass differential. This results in a fundamental asymmetry of the plasma flow between electrodes, localized onto the discontinuous interfaces of the plasma with the electrodes, namely, in the cathode dark space and in the anodic sheath: on the cathode side, electrons act upon ions, as the emitted electrons having less than zero initial velocities, drift against the incoming ion flux and in parallel with the ion and neutral counterflows; on the anode side of the discharge, positive ions flowing toward the cathode confront mainly the incoming counterflow of positive ions and neutral atoms, as the high speed electrons have abnormally transferred their energy to counterflowing, high speed, cathodic metal ions. An out of balance reaction force thus results at the cathode, to which the leaving metallic atoms impart a force of equal momentum but opposite direction, a force which is added to the cathode momentum generated by impacting, normal flowing positive ions. Moreover, Aspden confirmed theoretically the fundamental contention of Tanberg's experimental findings that an electrodynamic force will manifest itself along the direction of the discharge current flow, and thus, that the atomic counterflow is a metallic jet. Aspden further demonstrated that this asymmetry of plasma discharges does not imply any violation of the principles of conservation of energy and charge equivalence, given that there will be no out-of-balance force when such anomalous forces are considered in the context of the whole system of charge which must, perforce, include the local electromagnetic frame itself. Such discharges must be viewed as open energy systems, in balance with their electromagnetic environment: their apparatuses may constitute materially closed or limited systems, but they are physically and energetically open systems. Current work on Aspden's formulation of Ampere's Law indicates that both classical electromagnetism and special relativity ignore precisely, in circuits or in plasma, the longitudinal interactions that coexist with transverse ones. Standing longitudinal pressure waves, of a non-electromagnetic nature, have been previously shown in plasma electrons, which did not conform to the Bohm and Gross plasma oscillation mechanism (Pappas, P. T. (1983) "The original Ampere force and Bio-Savart and Lorentz forces", Il Nuovo Cimento, 76B:189; Looney, D. H. and Brown, S. C. (1954) "The excitation of plasma oscillations" Phys. Rev. 93:965)

The present theoretical approach to the novel regime of electrical discharge which we have isolated in specially designed devices, and to its mixed glow-arc characteristics, suggests that a similar, out-of balance current quadrature phenomenon occurs in the discharge plasma during the low field, autoelectronic emission-triggered PAGD, and is responsible for the observed surplus of energy in the experimental system described in this report. Clearly, all the evidence we have adduced indicates that there is a powerful longitudinal component to the emission-triggered PAGD, ie that the discharge pulses characteristic of this pre-VAD regime are longitudinally propelled jets of cathode-ejected high speed electrons and high speed ions. We have performed experiments, in the PAGD regime of operation, with very thin axial members that bend easily when placed in the path of the discharge, or with Crooke radiometer-type paddle-wheels, and both show the presence of a net longitudinal force in the plasma discharge acting in the direction of the anode, which confirms the magnitude of the atomic counterflow (ionized and neutral) present during the PAGD, very much like Tanberg's pendulum did for the VAD. These observations also tally with the explosive action of the emission mechanism, such as we have examined it above. In this context, two aspects of the PAGD are remarkable: the fact that a phenomenon akin to field emission occurs at low field values, for large area electrodes across large gaps, and the conclusion that the PAGD must deploy an excessively large counterflow of, in all probability, both ionized and neutral cathodic particles. The observation of ion current contributions to the cathode current on the order of 8 to 10%, in VADs, can hardly apply to the PAGD mechanism responsible for the anomalous currents and counterflows observed. Hence, we should further expect that the characteristically intermittent, or chopped current regime of the PAGD, is a major factor in the generation of disproportionately high energy longitudinal pulses and in allowing our system to capture most of the electrical energy output from the device. In all probability, field collapse at the end of discharge favours the nearly integral collection of the plasma charge, and ensures the transduction of most of the plasma energy of the pulse (blocked, as it is, from flowing back through the input port to the drive pack) to the output port, through the parallel, asymmetric capacitance bridge that interfaces with the charge recovery reservoir (the charge pack). Collapse of the field of the discharge may also be a contributing factor to the anomalous acceleration of ions, and to the observed anode plating effect. It is equally possible that such abnormally large longitudinal pulses may never be observable, for a given arrangement and scale, above threshold frequencies of the oscillation; we have, in this sense, presented data that indicates that for a given geometry, above specific PAGD frequencies, the capture of surplus energy decreases steadily in efficiency until it ceases altogether, for a given arrangement. The point at which this surplus begins to decrease coincides with the setting in of frequency-dependent irregularities in the discharge sequence and, most importantly, it coincides with a reduction of the peak pulse current for each PAGD pulse. We have further remarked that increasing the PAGD frequency above the zero surplus point, for a given arrangement, by manipulating any of the frequency control parameters, provokes the slippage of the PAGD into a full fledged VAD regime, while input currents greatly increase and output peak currents greatly decrease (to comparable peak input levels of 10 to 15A). The transition between the two modes of emission-triggered discharge, PAGD and VAD, thus appears to be tied in to adjustable thresholds in the frequency of the emission discontinuities; in this sense, it is rather likely that the plasma field collapse plays a major role in regularizing and optimizing the anomalous energies of field emissions, as in the PAGD regime. At low frequencies of low field emission, the emission regime is highly discontinuous, diachronic and regular, for it has time to fully extinguish the discharge; hence the PAGD singularity, in which the phases of each discharge pulse are well defined and sequential. Above a given high frequency, when ion and electron recombination will happen more often, before each can be collected at the electrodes, the stream of emitted discontinuities merges into a noisy, randomized continuum, where simultaneous emissions become possible and the plasma field no longer has time to collapse and fully resolve the longitudinal pulses. Any anomalous energy generated is then minimized and trapped in the plasma body and, in these conditions, the VAD regime eventually sets in. Such model would easily explain why the high field VAD experiments performed to date have never detected such extraordinarily large anomalous forces.

On the other hand, the quasi-coherent aspect of the discharge suggests that the vacuum gap, in functioning during the PAGD regime both as an insulator and as a conductor with capacitative and self-inductive properties, is periodically altered by large and intense polarizations which are resolved by the discrete emission of longitudinal pulses from the cathode. It is possible that these nonlinear oscillations resulting from sudden depolarization of the vacuum gap by high speed explosive emissions elicited at the convection focus of the distorted field, might be in resonance or near resonance with the external circuitry, but the most apparent effect of increasing the capacitance in all bridge members is to increase the jet current and the transduced current flowing into the charge pack. The PAGD amplitude variation also presents, after the large negative discontinuity, a growing oscillation at very high resonant frequencies, which are typical of inductive chopping currents in a VAD, before extinction occurs. Unlike the VAD inductive case, in the absence of any coils other than the wire wound resistors, the PAGD relaxation oscillations which follow each pulse only extinguish the discharge when the voltage potential of the amplitude curve rises above the applied voltage, just as the plasma potential drops the most. Given the entirely non-inductive nature of the external circuit utilized in many instances, the inductive properties in evidence are those of the vacuum device itself. It also suggests that, in the absence of any need of an applied external magnetic field for the PAGD discharge to occur coherently, it is possible that the magnitude of the currents generated produces by itself a significant self-magnetic field. Thus, we cannot rule out the possibility of a self-organization of the plasma discharge, which may, in Prigogine's sense, constitute a dissipative structure (Prigogine, I. and George, C. (1977), "New quantization rules for dissipative systems", Int. J. Quantum Chem., 12 (Suppl.1):177). Such self-ordering of the PAGD plasma jet is suggested by the experimentally observed transition of these pulses from the current saturated limit of the normal glow discharge region, into the PAGD regime, as a function of increasing current: smaller foci of discharge can be seen to discontinuously agglutinate into larger emission cones, or into jets with a vortex-like appearance, when the input current reaches a given threshold. It is possible that, under these conditions, the distribution of the charge carriers and their sudden fluctuations may render any steady state plasma boundary conditions ineffective and provoke a singularity in the discharge mechanism; this nonlinear behaviour, together with any self-magnetic effects, might provide radial coherence of the plasma flow along the longitudinal path of the discharge. This concept is akin to what has been proposed for periodically evanescent solution structures referred to as instantons, that represent self-organizing transitions between the two states of a system. The PAGD may well be an instance of an instanton type structure bridging the open, or conductive, and the closed, or insulating, states of the vacuum gap. An analytical formulation of the problem of the plasma flow from the cathode spot to the anode, which would take into account the self-magnetic and self-organizing properties of the PAGD plasma channel, would be extremely difficult, given the out of balance longitudinal force, its abnormal energy transfer and associated counterflow, as well as the competition between collisional and inertial exchanges.

The plating observed at the anode most likely results from the impact of counterflowing ions (and possibly neutral atoms), whereas the pitting of the (locally molten) cathode results from the emission of vaporized metallic material and electrons, as well as, secondarily, from bombardment by incident positive ions. The first action smooths the surface by mirroring it (deposition of cathode-derived atoms) and abrading it, whereas the latter smooths it in places by rounding concavities and by forming molten droplets upon local cooling, while simultaneously roughening it on the crater peripheries. One might think that this cathode roughening should lower the work function and facilitate the discharge, but the facts indicate that just the opposite must be happening in view of changes in the PAGD according to the nature and state of the cathode surface. The observed alterations of electrode work function for PAGD low field emission must thus be related to the molecular and charge effects of these different actions at the two electrodes. It appears that for large parallel plate electrodes, the PAGD low field emission is modulated by the nature and, most likely, by the molecular structure of the metallic surface layer of the emitter.

We have thus devised a system for the capture as electricity of the energy of anomalously energetic longitudinal pulses sequentially triggered by spontaneous emissions of high speed electrons and ions generated from low work function cathodes, during the low field and singularly mixed PAGD regime of electrical discharge in vacuo. To confirm the above interpretation of the anomalous flux in the observed PAGD phenomenon, the cathode jet composition, as well as time- and usage-dependent changes occurring in the tubes, with diverse sealed negative pressures and after submission to prolonged PAGD operation, must be analyzed mass spectroscopically. In any event, the excess energy present in the anomalous counterflowing force appears to stem from a discharge mechanism that effectively pulls high speed electrons and constituent atoms out of a metal surface, at low fields and with high current densities, and is modulated by a complex multiplicity of parameters. The system described appears to transduce efficiently the observed nonlinear longitudinal pulse discontinuities of the plasma field, under conditions of current saturation of the cathode, because the self-extinguishing and self-limiting properties of the discharge allows the energy from the collapse of the discharge to be captured. The particular design of the circuitry, which couples a rectification bridge to the asymmetric bridge quadrature of large capacitances, placed at the output of the PAGD generator, permits effective capture. Our findings constitute striking evidence for Aspden's contention of a need to revise our present electrodynamic concepts. The dual ported PAGD discharge tube circuits which we have described are the first electrical systems we know of which permit effective exploitation of anomalous cathode reaction forces and allow for the recovery of electrical energy from systems exhibiting this effect. Any apparent imbalance in the electrical energy input to the system and withdrawn from the system by its operator must be considered in the context of the entire continuum in which the system operates, within which it is anticipated that accepted principles of energy balance will be maintained.

Moreover, the energy conversion system of the invention has substantial utility as an electrical inverter accepting direct current, and providing one or more of a direct current output at lower voltage and higher current, variable frequency input to alternating current motors, and, by suitable combinations of discharge tube systems, more flexible DC to DC conversion systems.

As an alternative to the batteries used in the experiments described, a DC power supply may be utilized or, more advantageously from the viewpoint of entailing less transformation losses, a DC generator to provide the electrical energy input to the system. As a DC motor can be run directly from the rectified output of the circuit of FIG. 9 at El–E2, in place of a battery charge pack, DC motor/generator sets of suitable characteristics (in terms of back E.M.F. and circuit loading) can be used to charge the batteries of the drive pack, utilizing the rectified PAGD output to drive the DC motor component of the set. This provides a simple, one battery pack solution, where the PAGD input and output circuits are electrically separated by the DC motor/generator interface: the drive pack is simultaneously being discharged to drive PAGD production, and charged by the DC generator output which, in turn, is being driven by the electromechanical transformation of the rectified PAGD output that would typically accrue to a charge pack in the experiments already described. The main limitations to such an arrangement lie in the efficiency of the motor and generator transformations utilized.

A pulsed DC source could be used to provide input to the circuit if suitably synchronized, but care is needed not to interfere unduly with the autoelectronic mechanism of the field induced cathode emissions.

TABLE 1

Results for the ballast resistance (and current) dependent PAGD frequency utilizing an H34 aluminum pulse generator with 128 cm^2 plates at 5.5 cm distance, in the triode configuration, at a pressure of 0.8 Torr. The circuit employed is that of the present invention, as described in the third Results Section. DCV = 560.

R in Ω	Regime of Discharge	Pulse Rate >100 V
5,000	NGD (Cold Cathode)	0
600	PAGD	10 PPS
300	PAGD	40 PPS
150	PAGD	180 PPS
100	VAD	0
50	VAD	0

TABLE 2

128 cm^2 H220 Al; 570 volts DC; 300 Ω = R1; Diode Configuration

	PPS	p(Torr)	Cumulative Pulse Count
1)	200	0.08	$\sim 2.4 \times 10^5$
2)	200	0.5	$\sim 1.5 \times 10^6$
3)	200	0.8-1	$\sim 2.5 \times 106$
4)	25	0.5	3×10^6 pulses
5)	200	0.5	1.5×10^6 (after first electrode reversal)

TABLE 3

RESIDUAL GAS EFFECT

pressure in Torr	PPS in AIR	PPS in ARGON
0.45	ND	10
0.5	1.8 ± 0.3	ND
0.55	4.8 ± 0.9	16.7±1.8
1.0	11.4 ± 0.8	448 ± 27.4
1.25	214.5 ± 14.3	ND
2.0	36.2 ± 2.6	206 ± 19.6 158.7 ± 24
2.5	1.36 ± 0.3	0

TABLE 4

Charge pack No. of cells	PPS	PAGD
36	0	−
31	1	+
29	10	+
19	1	+
9	0	−

TABLE 5

1 Expt. No.	2 Battery Pack	3 Position	4 Open Voltage	5 V/cell	6 % total rel. cpty.	7 Max. hr. left	8 % rel. cpty gained	8 % rel. cpty lost	9 Total kWh	10 ΔkWh gain	10 ΔkWh loss	11 PAGD per sec
1	Charge	start	348	12.0	40	8			0.835			8
	Charge	end	366	12.62	83	16.6	43		1.823	0.988		
	Driver	start	576	12.52	77	15.4			2.660			
	Driver	end	572	12.43	70	14		7	2.402		0.258	
2	C	b	331	11.41	2	0.4			0.040			61
	C	a	351	12.1	47.5	9.5	45.5		1.002	0.962		
	D	b	553	12.02	40	8			1.327			
	D	a	546	11.9	33	6.6		7	1.081		0.246	
3	C	b	345	11.9	32.5	6.5			0.673			3
	C	a	361	12.45	72.5	14.4	40		1.559	0.886		
	D	b	559	12.15	51	10.2			1.710			
	D	a	552	12.0	40	8		11	1.324		0.386	
4	C	b	360	12.41	70	14			1.512			32
	C	a	373	12.86	103	>20	33		2.238	0.726		
	D	b	562	12.22	54.5	10.9			1.838			
	D	a	557	12.11	48	9.6		6.5	1.604		0.234	
5	C	b	340	11.7	20	4			0.408			2
	C	a	365	12.59	83	16.6	63		1.818	1.440		
	D	b	527	11.45	3.2	0.6			0.101			
	D	a	517	11.24	1.8	0.4		0.2	0.056		0.045	
6	C	b	340	11.72	21.5	4.3			0.438			8
	C	a	367	12.66	87.5	17.5	66		1.927	1.489		
	D	b	589	12.8	100	20			3.530			
	D	a	564	12.26	58.5	11.7		41.5	1.979		1.551	
7	C	b	318	10.97	1.2	0.24			0.023			5
	C	a	359	12.38	67.5	13.5	66.3		1.454	1.431		
	D	b	575	12.5	77	15.4			2.656			
	D	a	567	12.32	63.5	12.7		13.5	2.160		0.496	
8	C	b	328	11.71	20	4			0.393			32
	C	a	350	12.5	76.5	15.3	56.5		1.606	1.213		
	D	b	582	12.65	87.5	17.5			3.055			
	D	a	579.5	12.60	84	16.8		3.5	2.921		0.134	

1 Expt. No.	2 Battery Pack	3 Position	12 Exptl. time	13 rel. kWh/h gain	13 rel. kWh/h loss	14 net kWh/h production	15 Breakeven efficiency	16 Cell #/ pack	17 tube	18 Cathode Area	19 Plate
1	Charge	start	21.5′			2.071	388%	29	A26	128 cm^2	H34
	Charge	end		2.791							

TABLE 5-continued

	Driver	start						46			
	Driver	end			0.720						
2	C	b	18′			2.387	391%	29	A26	128 cm^2	H34
	C	a		3.207							
	D	b						46			
	D	a			0.820						
3	C	b	21.5′			1.396	230%	29	A26	128 cm^2	H34
	C	a		2.473							
	D	b						46			
	D	a			1.077						
4	C	b	63.5′			0.465	310%	29	A28	128 cm^2	H220
	C	a		0.686							
	D	b						46			
	D	a			0.221						
5	C	b	80′			1.064	6,750%	29	A26	128 cm^2	H34
	C	a		1.080							
	D	b						46			
	D	a			0.016						
6	C	b	21.5′			−0.173	96%	29	A26	128 cm^2	H34
	C	a		4.155							
	D	b						46			
	D	a			4.328						
7	C	b	64.5′			0.870	289%	29	A45	64 cm^2	H34
	C	a		1.331							
	D	b						46			
	D	a			0.461						
8	C	b	28.5′			2.272	906%	28	A45	64 cm^2	H34
	C	a		2.554							
	D	b						46			
	D	a			0.282						

1 Expt. No.	2 Battery Pack	3 Position	20 R1 ohm	21 C3/C5 mfd	22 C7a/C7b mfd	23 Motor arm	24 Pressure	25 Gap cm	26 OV rlx. time	27 C4 mfd	28 R4 ohms	29 Motor rpm
1	Charge	start	300	20,700	3,300	off	0.8 Torr	5.5	30′	NA	NA	NA
	Charge	end										
	Driver	start										
	Driver	end										
2	C	b	300	20,700	3,300	off	0.8 Torr	5.5	30′	NA	NA	NA
	C	a										
	D	b										
	D	a										
3	C	b	300	20,700	3,300	off	0.7 Torr	5.5	15′	NA	NA	NA
	C	a										
	D	b										
	D	a										
4	C	b	300	34,700	5,500	off	0.2 Torr	5.5	30′	NA	NA	NA
	C	a										
	D	b										
	D	a										
5	C	b	150	34,700	3,300	on	0.8 Torr	5.5	15′	8	500	1,200
	C	a										
	D	b										
	D	a										
6	C	b	300	20,700	3,300	on	0.8 Torr	5.5	15′	16	0	2,000
	C	a										
	D	b										
	D	a										
7	C	b	600	34,700	3,300	off	0.8 Torr	4	30′	NA	NA	NA
	C	a										
	D	b										
	D	a										
8	C	b	600	34,700	5,500	off	0.8 Torr	4	30′	NA	NA	NA
	C	a										
	D	b										
	D	a										

TABLE 6

Expt.	Battery		Load	Watts/		Total	ΔkWh		rel. kWh/h		net	
No.	Pack	Position	Voltage	cell	Hr. left	kWh	gain	loss	gain	loss	kWh/h	B. Eff.
1	C	s	335.7	4.445	4	0.516					3.014	776%
	C	e	357.5	5.05	12	1.757	1.241		3.46			
	D	s	568.0	3.20	13	1.766						
	D	e	564.6	3.175	11	1.606		0.16		0.446		
2	C	s	315.5	3.93	1	0.114					1.012	504%
	C	e	327.8	4.25	4.5	0.502	0.387		1.225			
	D	s	540.7	2.91	4	0.535						
	D	e	535.3	2.87	3.5	0.462		0.073		0.243		
3	C	s	328	4.23	2	0.245					1.175	703%
	C	e	351.7	4.91	7	0.737	0.492		1.370			

TABLE 6-continued

Expt. No.	Battery Pack	Position	Load Voltage	Watts/ cell	Hr. left	Total kWh	ΔkWh gain	ΔkWh loss	rel. kWh/h gain	rel. kWh/h loss	net kWh/h	B. Eff.
	D	s	546	2.95	5	0.680						
	D	s	545.5	2.90	4.5	0.610		0.070		0.195		

TABLE 7

1 Expt. No.	2 Config.	3 Pressure Torr	4 Tube	5 DP DCV	6 Plates DCV	7 DP DCA	8 DP Watts	9 PAGD Volts	10 PAGD V/cm	11 CP DCV
1	dd	0.8	A29	562	350	0.65	137.8	212	77.1	375
2	dd	0.09	A29	562	402	0.60	96	160	58.2	378
3	dd	0.8	A29	560	371	0.59	111.5	189	68.7	374
4	dd	0.09	A29	563	409	0.49	75.9	154	56	379
5	t	1.5	A28	561	439	0.41	49.9	122	22.2	377
6	t	1.5	A28	560	425	0.51	68.9	135	24.5	375
7	t	1.0	A28	556	398	0.48	75	158	28.7	376.5
8	t	0.5	A28	559.5	398	0.68	109.8	161.5	29.4	377.5
9	t	0.5	A28	563	390	0.75	112.45	173	31.5	373
10	sd	0.5	A28	565	422	0.47	67.2	143	26	376
11	sd	0.5	A28	561.5	415	0.50	73	146.5	26.6	380
12	sd	0.5	A28	562	413.5	0.55	81.7	148.5	27	380
13	dd	0.25	A28	553	438	0.35	40	115	41.8	381.5
14	dd	0.25	A28	549	325	0.70	156.8	224	81.5	263

1 Expt. No.	2 Config.	12 CP DCA	13 CP Watts	14 Total Resistance	15 Breakeven Efficiency	16 PPS	17 Bridge diode	18 Input diode	19 Motor status	20 FIG. 3
1	dd	1.25	468.8	326	340%	450	M860	HFR	off	+
2	dd	0.70	264.6	% 270	276%	92	M860	HFR	off	
3	dd	0.65	243.1	243	218%	500	HFR	HFR	off	
4	dd	0.76	288	314	379%	77	HFR	HFR	off	
5	t	0.58	219	298	439%	52	HFR	HFR	off	
6	t	0.69	259	265	376%	100	M860	HFR	off	
7	t	0.57	213.1	329	284%	355	M860	HFR	off	
8	t	0.67	252.9	238	230%	92	HFR	HFR	off	
9	t	0.65	280	266	249%	118	M860	HFR	off	+
10	sd	1.03	387.3	286	530%	25	M860	HFR	off	
11	sd	0.73	277.4	293	379%	11	HFR	HFR	off	+
12	sd	0.71	269.8	270	330%	10	HFR	HFR	on	+
13	dd	0.59	225.1	329	563%	10	HFR	HFR	off	
14	dd	1.36	257.7	320	228%	1	HFR	HFR	off	

TABLE 8

1 Expt. No.	2 Battery Pack	3 Position	4 Total Wh	5 Rel. Cap.	6 Torr	7 Limit in W	8 ΔkWh gain	ΔkWh loss	9 Exptl. time	10 abs. kWh/h gain	abs. kWh/h loss	net	11 BE
1	C	b	159	12%	0.8	90			21.5′			+664	846%
	C	a	428	32%			269			753			
	D	b	1764	85%		115							
	D	a	1732	84%				32			89		
2	C	b	118	9%	0.8	90			18′			+616	2,667%
	C	a	303.5	23%			192			640			
	D	b	542.3	26%		115							
	D	a	535	25.9%				7.3			24		
3	C	b	950.4	72%	0.2	90			70′			+186	3485%
	C	a	1,161	88%			210.9			191.7			
	D	b	660	32%		115							
	D	a	654	32%				6.5			5.6		
4	C	b	15.8	1.2%	0.8	90			64.5′			+53.7	406%
	C	a	81.9	6%			65			60			
	D	b	181	8.7%		115							
	D	a	165	8%				16			14.7		
5	C	b	34.5	2.6%	0.8	90			28.5′			+169.1	436%
	C	a	138.8	10.5%			104.3			219.6			
	D	b	1,114	54%		115							
	D	a	1,089	53%				24			50.5		
6	C	b	55.4	4.2%	0.8	90			74′			+117	483%
	C	a	237.6	18%			182.2			148			
	D	b	669.3	32%		115							
	D	a	631.7	30.6%				37.7			30.6		

1 Expt. No.	2 Battery Pack	3 Position	12 Config.	13 Tube	14 Cathode area	15 gap cm	16 PPS	17 PAGD seq. method	18 R1 ohms	19 Plate material	20	21 C3/C5 mfd	22 C7a/C7b mfd
1	C	b	Triode	A26	128 cm2	5.5	8	Continuous	300	H34		20,700	3.300
	C	a											

TABLE 8-continued

	D	b										
	D	a										
2	C	b	Triode	A26	128 cm2	5.5	61	Interrupted	300	H34	20,700	3,300
	C	a										
	D	b										
	D	a										
3	C	b	Triode	A28	128 cm2	5.5	32	Interrupted	300	H220	34,700	5,500
	C	a										
	D	b										
	D	a										
4	C	b	Triode	A46	64 cm2	4.0	5	Continuous	600	H34	34,700	5,500
	C	a										
	D	b										
	D	a										
5	C	b	Triode	A46	64 cm2	4.0	32	Interrupted	600	H34	34,700	5,500
	C	a										
	D	b										
	D	a										
6	C	b	Plate Diode	A29	128 cm2	5.5	8	Interrupted	300	H220	34,700	5,500
	C	a										
	D	b										
	D	a										

TABLE 9

Utilizing: Al H200, 128 cm^2 plates
DP = 46 cells
CP = 23 cells

	PPS	CP Gain per pulse in mWh	Net Gain per pulse mWh	CP Gain per second mWh	Net Gain per second mWh	Pressure in Torr
#1	1.5	22.3	11.7	33:45	17.55	0.2
#2	8	5.6	4.4	44.8	35.2	0.8
#3	110	0.78	0.27	85.8	29.7	2.0

We claim:

1. Apparatus comprising a discharge tube and an electrical circuit containing said discharge tube and configured to operate the latter to provide endogenous pulsatory cold cathode autoelectronic emissions, the circuit being double ported with an input port connected to a source of direct current at a potential sufficient to initiate said emissions, and an output port connected to a current sink effective to absorb at least a substantial portion of electrical energy released by collapse of said emissions.

2. Apparatus according to claim **1** configured so that the emissions occur in a pulsed abnormal glow discharge regime.

3. Apparatus according to claim **2**, wherein the input port includes components ensuring that the flow of current therein is unidirectional, and incorporating impedance sufficient to limit the flow of current therein.

4. Apparatus according to claim **2**, including capacitors connected to the discharge tube, the input port and the output port, which provide charge storage in the input port and direct current isolation between the input and output ports.

5. Apparatus according to claim **4**, wherein the output port comprises a rectifier having an input connected to said capacitors, reservoir capacitance connected to the output of said rectifier, and reverse current blocking devices connected between said reservoir capacitance and the current sink.

6. Apparatus according to claim **5**, wherein the rectifier is a bridge rectifier, and the reservoir capacitance is provided by a capacitor bridge having ends connected to outputs of the bridge rectifier, and an intermediate point connected to one input of the bridge rectifier.

7. Apparatus according to claim **4**, further including an alternating current motor and a capacitor in series, connected between the connections of said capacitors to the output port.

8. Apparatus according to claim **2**, wherein the current sink comprises a secondary battery.

9. Apparatus according to claim **2**, wherein the current sink comprises an electric motor.

10. Apparatus according to claim **2**, wherein the direct current source comprises a secondary battery.

11. Apparatus according to claim **2**, wherein the direct current source is a DC generator.

12. Apparatus according to claim **9**, wherein the motor is a DC motor.

13. Apparatus according to claim **10**, including a circuit for charging from the output port a battery to be used as the direct current source.

14. Apparatus according to claim **2**, wherein the direct current source is a rectified AC source.

15. Apparatus according to claim **2**, wherein the discharge tube is connected as a single diode.

16. Apparatus according to claim **2**, wherein the discharge tube is connected as a multiple diode with plates connected as cathodes and an intermediate electrode connected as an anode.

17. Apparatus according to claim **2**, wherein the discharge tube is connected as a triode, with an intermediate electrode functioning as an auxiliary cathode.

18. Apparatus according to claim **2**, wherein a first potential is applied to the input port by the source of direct current to induce emission, a back EMF is applied to the output port by the current sink, and an extinction potential of the emissions is greater than the back EMF.

19. A method of energy conversion, comprising initiating plasma eruptions from the cathode of a discharge tube operating in a pulsed abnormal glow discharge regime utilizing electrical energy from a source in a first circuit connected to said discharge tube, and capturing electrical energy generated by the collapse of such eruptions in a second circuit connected to said discharge tube.

20. A method according to claim **19**, wherein current flowing into the discharge tube during said eruptions is at least 50 ma.

21. A method according to claim **19**, wherein current flowing into the discharge tube during said eruptions is at least 500 ma.

22. A method according to claim **19**, in which charge carriers within plasma outputs are accelerated through at least one of an electric and magnetic field.

23. A method of energy conversion, comprising inducing endogenous pulsatory low-field, large-area cold-cathode autoelectronic emissions from the cathode of a discharge tube capable of sustaining such emissions, utilizing electrical energy from a source in a first circuit connected to said discharge tube, and capturing electrical energy generated by the collapse of such emissions in a second circuit connected to said discharge tube.

* * * * *

US005984239A

United States Patent [19]

Chen

[11] **Patent Number: 5,984,239**

[45] **Date of Patent: Nov. 16, 1999**

---

[54] **WEATHER MODIFICATION BY ARTIFICIAL SATELLITES**

[76] Inventor: **Franklin Y. K. Chen**, One Meadow Glen Rd., Northport, N.Y. 11768

[21] Appl. No.: **09/069,655**

[22] Filed: **Apr. 28, 1998**

Related U.S. Application Data

[62] Division of application No. 08/480,883, Jun. 7, 1995, Pat. No. 5,762,298, which is a continuation of application No. 07/675,880, Mar. 27, 1991, abandoned.

[51] **Int. Cl.**^{6} **B64G 1/44**; A01G 15/00

[52] **U.S. Cl.** **244/173**; 73/170.27

[58] **Field of Search** .... 73/29.01, 170.27, 73/170.28; 126/572, 593; 239/2.1, 14.1; 244/1 R, 158 R, 173

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,564,253	2/1971	Buckingham	244/1 R
4,305,555	12/1981	Davis	244/158 R
4,371,135	2/1983	Keiger	244/173
5,114,101	5/1992	Stern et al.	244/173
5,631,414	5/1997	Cherny	73/170.27
5,646,343	7/1997	Pritchard	73/170.28
5,675,081	10/1997	Solheim et al.	73/170.28

Primary Examiner—Peter M. Poon
Assistant Examiner—Charles R. Ducker, Jr.
Attorney, Agent, or Firm—Donald S. Dowden; Cooper & Dunham LLP

[57] **ABSTRACT**

A Satellite Weather Modification System (SWMS) uses earth satellites to harness solar energy to modify the thermodynamics and composition of the earth's atmosphere. SWMS has three subsystems: The first subsystem includes a network of earth satellites called Satellite Engines (SEs) used to reflect solar energy and/or transform solar energy into other forms of energy beams discharged at specified locations. The media at these locations and the media through which the energy beams pass absorb these energies and change them into heat. The second subsystem includes a large network of Remote Sensing Devices (RSDs). These sensors are used to measure local media compositions, dynamic parameters and thermodynamic properties. Sensor measurements are fed back to the third subsystem, which includes a network of Ground Control Stations (GCSs). GCSs provide energy beam guidance by estimating each beam's characteristics and its aim point trajectory as functions of time. Integration of these three subsystems establishes a sensor feedback energy beam guidance and control loop. SWMS's weather modification applications include alteration of precipitation, reclaiming of wasteland, reducing damage by bad weather, and improving environment. Its non-weather related applications include supplying concentrated energy to electricity generating stations (solar, wind and hydro), high latitude greenhouse farms, and solar powered airplanes.

16 Claims, 7 Drawing Sheets

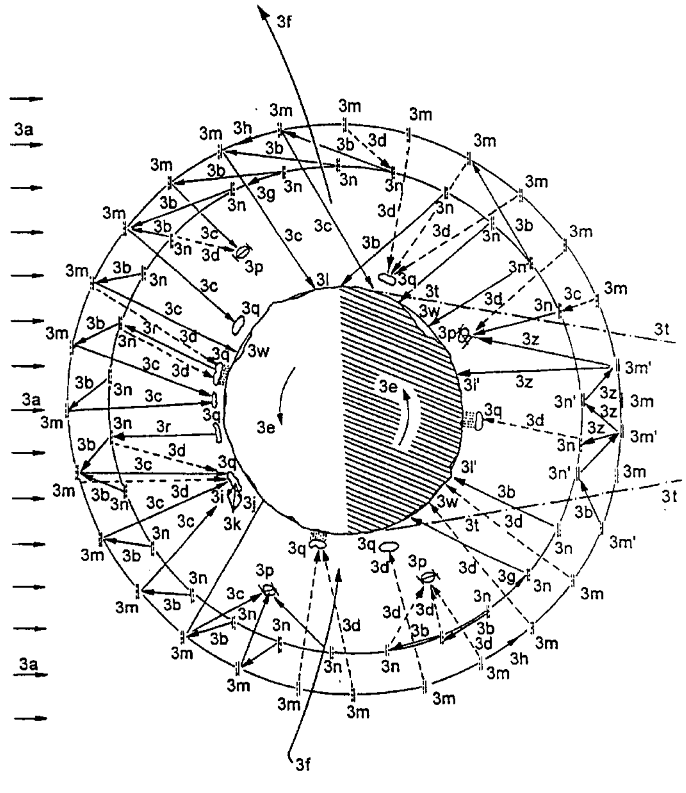

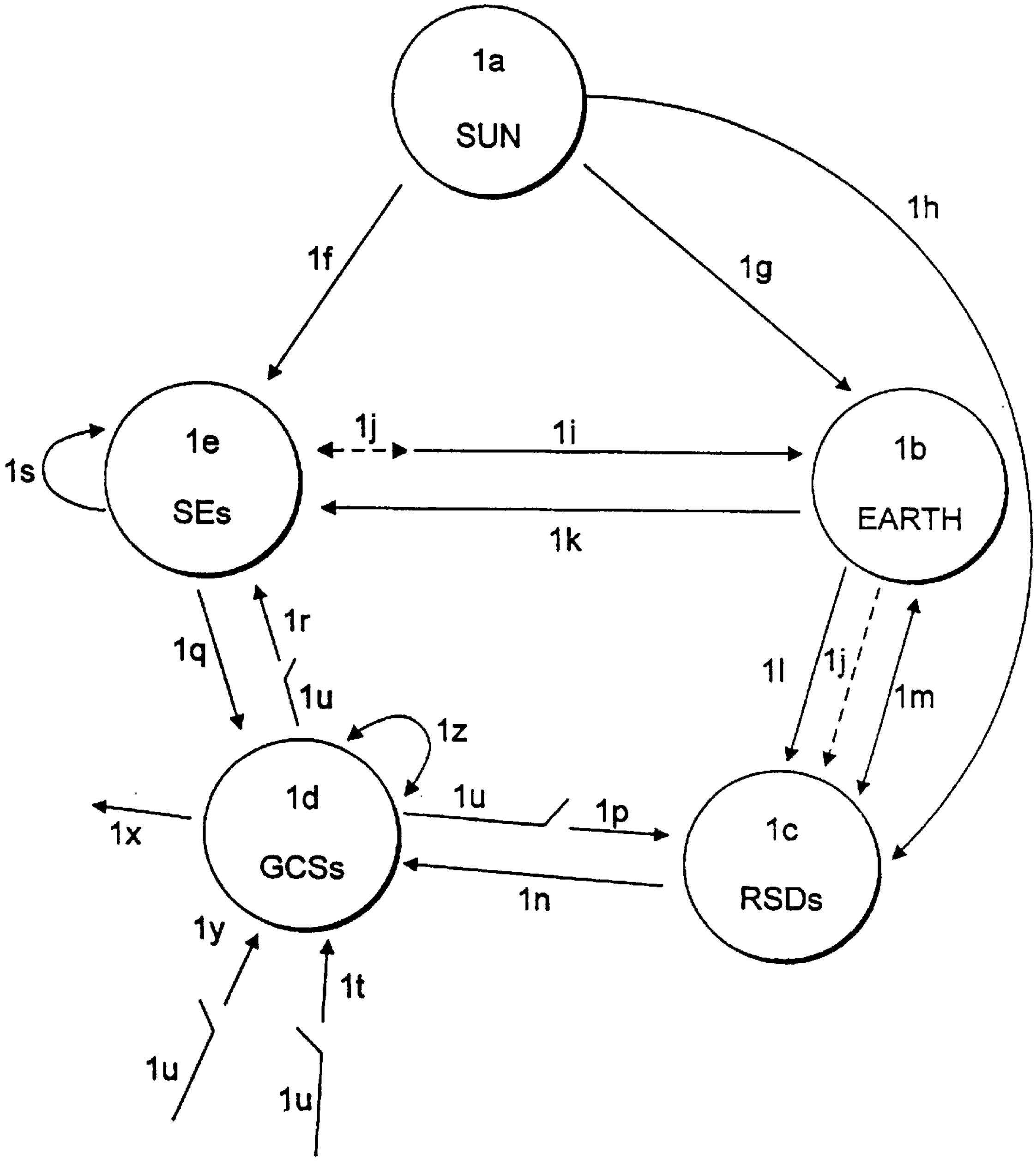
1a
SUN
1f
1g
1h
1e
SEs
1s
1j
1i
1k
1b
EARTH
1q
1r
1u
1l
1j
1m
1z
1d
GCSs
1u
1p
1c
RSDs
1x
1n
1y
1t
1u
1u

FIG. 1

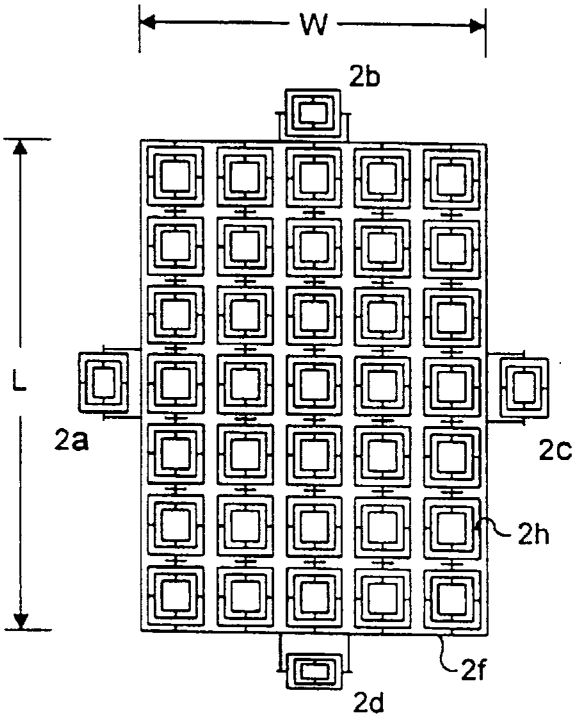
W
2b
L
2a
2c
2h
2f
2d

FIG. 2A

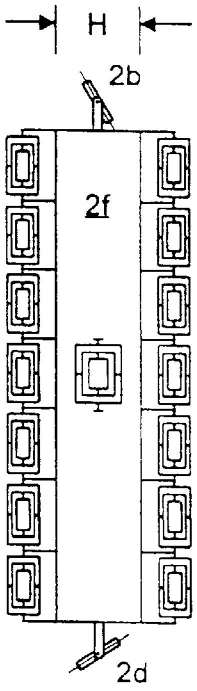

FIG. 2B

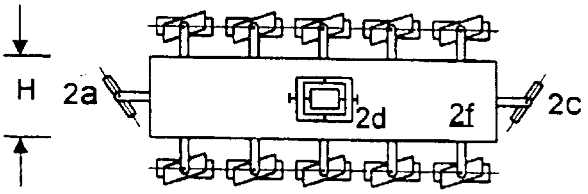

FIG. 2C

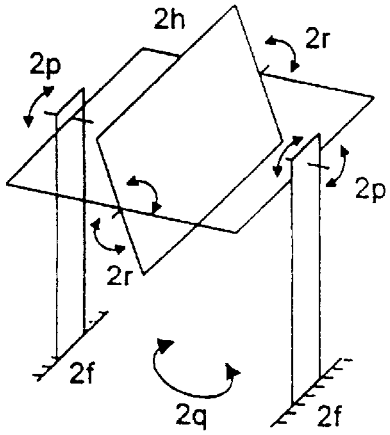
2h
2r
2p
2p
2r
2f
2q
2f

FIG. 2D

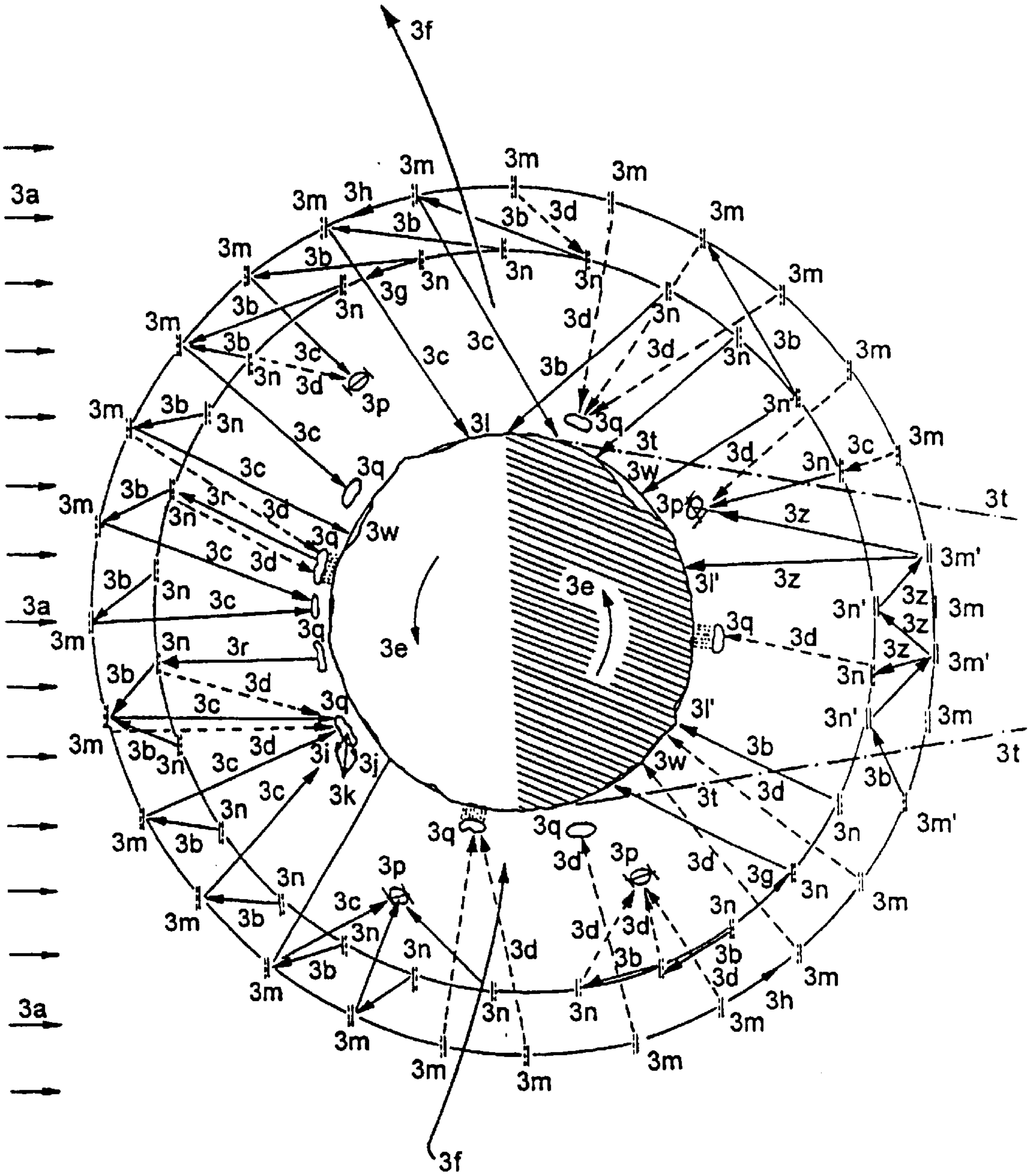
3f
3a
3m
3h
3b
3c
3d
3n
3g
3p
3q
3l
3t
3w
3r
3e
3z
3m'
3n'
3l'
3i
3j
3k

FIG. 3

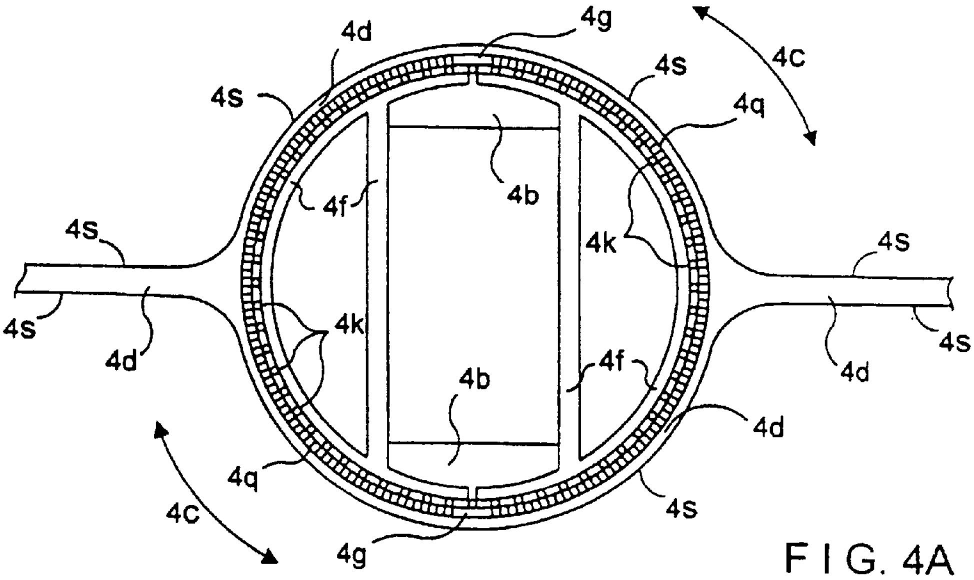
4g
4d
4c
4s
4s
4q
4f
4b
4k
4s
4s
4s
4s
4k
4d
4d
4f
4b
4d
4q
4s
4c
4g

FIG. 4A

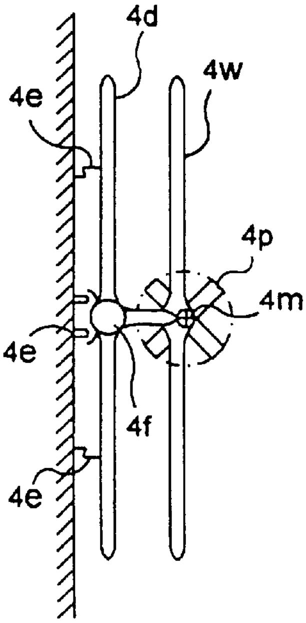
4d
4w
4e
4p
4m
4e
4f
4e

FIG. 4B

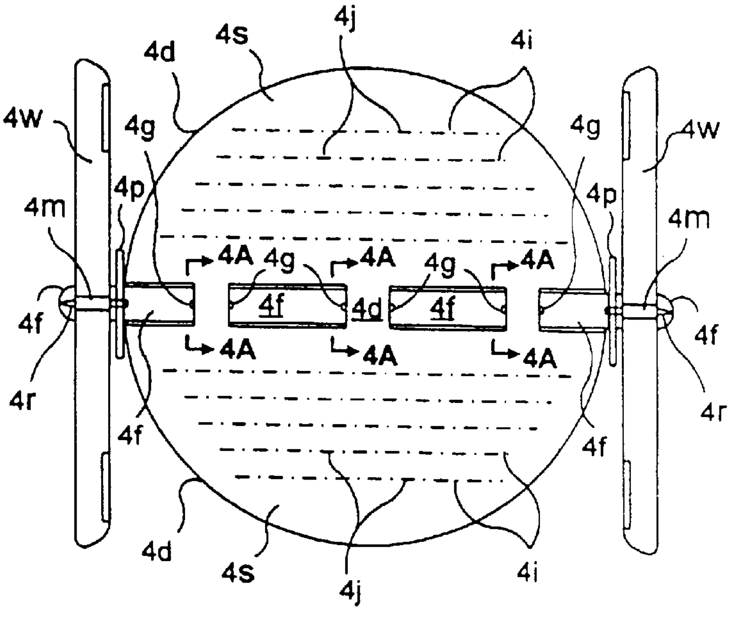
4j
4s
4i
4d
4w
4g
4g
4w
4p
4p
4m
4m
4A
4g
4A
4g
4A
4f
4f
4d
4f
4f
4f
4A
4A
4A
4r
4r
4f
4f
4d
4s
4j
4i

FIG. 4C

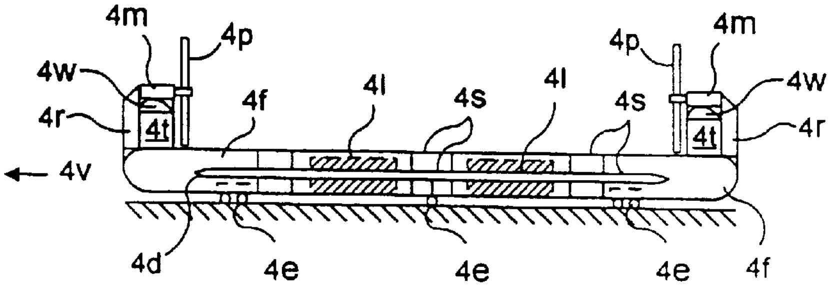
4m
4p
4w
4r
4t
4f
4l
4s
4l
4s
4p
4m
4w
4t
4r
4v
4d
4e
4e
4e
4f

FIG. 4D

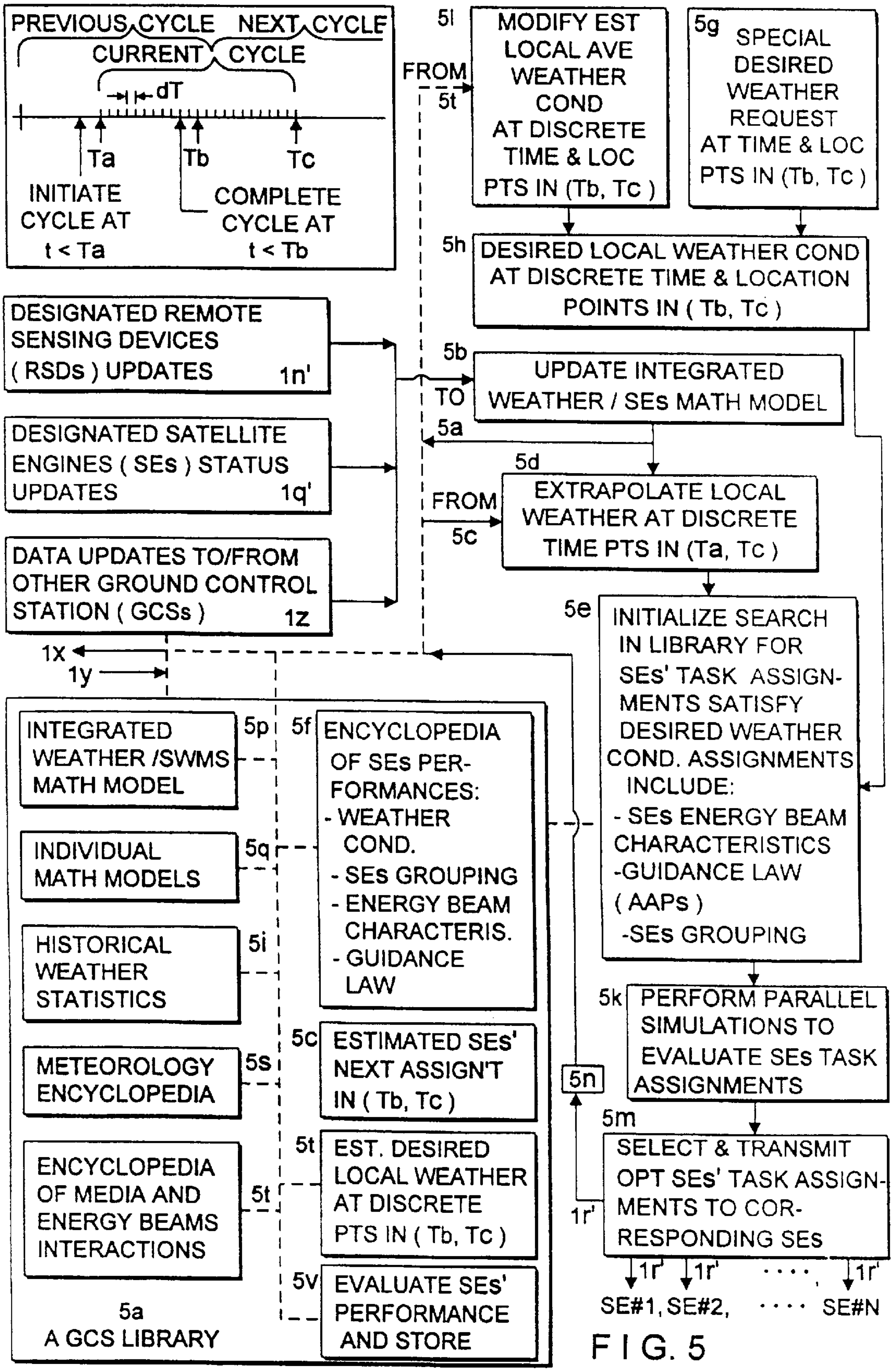
PREVIOUS CYCLE
NEXT CYCLE
CURRENT CYCLE
dT
Ta
Tb
Tc
INITIATE CYCLE AT t < Ta
COMPLETE CYCLE AT t < Tb
5l
FROM 5t
MODIFY EST LOCAL AVE WEATHER COND AT DISCRETE TIME & LOC PTS IN (Tb, Tc)
5g
SPECIAL DESIRED WEATHER REQUEST AT TIME & LOC PTS IN (Tb, Tc)
5h
DESIRED LOCAL WEATHER COND AT DISCRETE TIME & LOCATION POINTS IN (Tb, Tc)
DESIGNATED REMOTE SENSING DEVICES (RSDs) UPDATES
1n'
5b
UPDATE INTEGRATED WEATHER / SEs MATH MODEL
TO 5a
DESIGNATED SATELLITE ENGINES (SEs) STATUS UPDATES
1q'
5d
EXTRAPOLATE LOCAL WEATHER AT DISCRETE TIME PTS IN (Ta, Tc)
FROM 5c
DATA UPDATES TO/FROM OTHER GROUND CONTROL STATION (GCSs)
1z
5e
INITIALIZE SEARCH IN LIBRARY FOR SEs' TASK ASSIGNMENTS SATISFY DESIRED WEATHER COND. ASSIGNMENTS INCLUDE:
- SEs ENERGY BEAM CHARACTERISTICS
-GUIDANCE LAW (AAPs)
-SEs GROUPING
1x
1y
INTEGRATED WEATHER /SWMS MATH MODEL
5p
5f
ENCYCLOPEDIA OF SEs PERFORMANCES:
- WEATHER COND.
- SEs GROUPING
- ENERGY BEAM CHARACTERIS.
- GUIDANCE LAW
INDIVIDUAL MATH MODELS
5q
HISTORICAL WEATHER STATISTICS
5i
5k
PERFORM PARALLEL SIMULATIONS TO EVALUATE SEs TASK ASSIGNMENTS
5c
ESTIMATED SEs' NEXT ASSIGN'T IN (Tb, Tc)
METEOROLOGY ENCYCLOPEDIA
5s
5n
5m
5t
EST. DESIRED LOCAL WEATHER AT DISCRETE PTS IN (Tb, Tc)
SELECT & TRANSMIT OPT SEs' TASK ASSIGN-MENTS TO COR-RESPONDING SEs
1r'
ENCYCLOPEDIA OF MEDIA AND ENERGY BEAMS INTERACTIONS
5t
5v
EVALUATE SEs' PERFORMANCE AND STORE
1r'
1r'
1r'
SE#1, SE#2, ... SE#N
5a
A GCS LIBRARY

FIG. 5

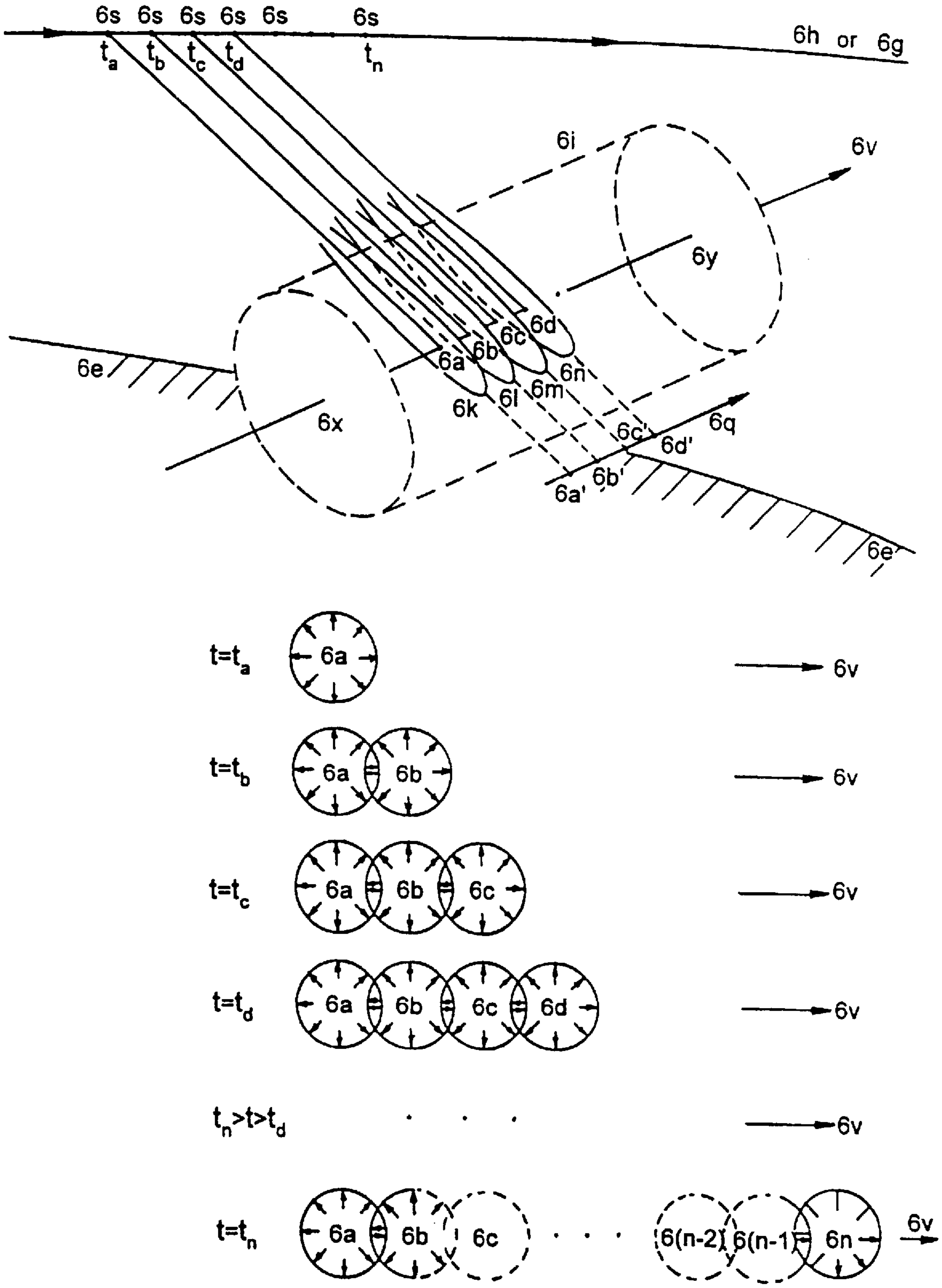
6s
6s
6s
6s
6s
6s
t_a
t_b
t_c
t_d
t_n
6h or 6g
6i
6v
6y
6d
6c
6b
6a
6n
6m
6l
6k
6e
6x
6c'
6d'
6q
6b'
6a'
6e
t=t_a
6a
6v
t=t_b
6a
6b
6v
t=t_c
6a
6b
6c
6v
t=t_d
6a
6b
6c
6d
6v
t_n>t>t_d
6v
t=t_n
6a
6b
6c
6(n-2)
6(n-1)
6n
6v

FIG. 6

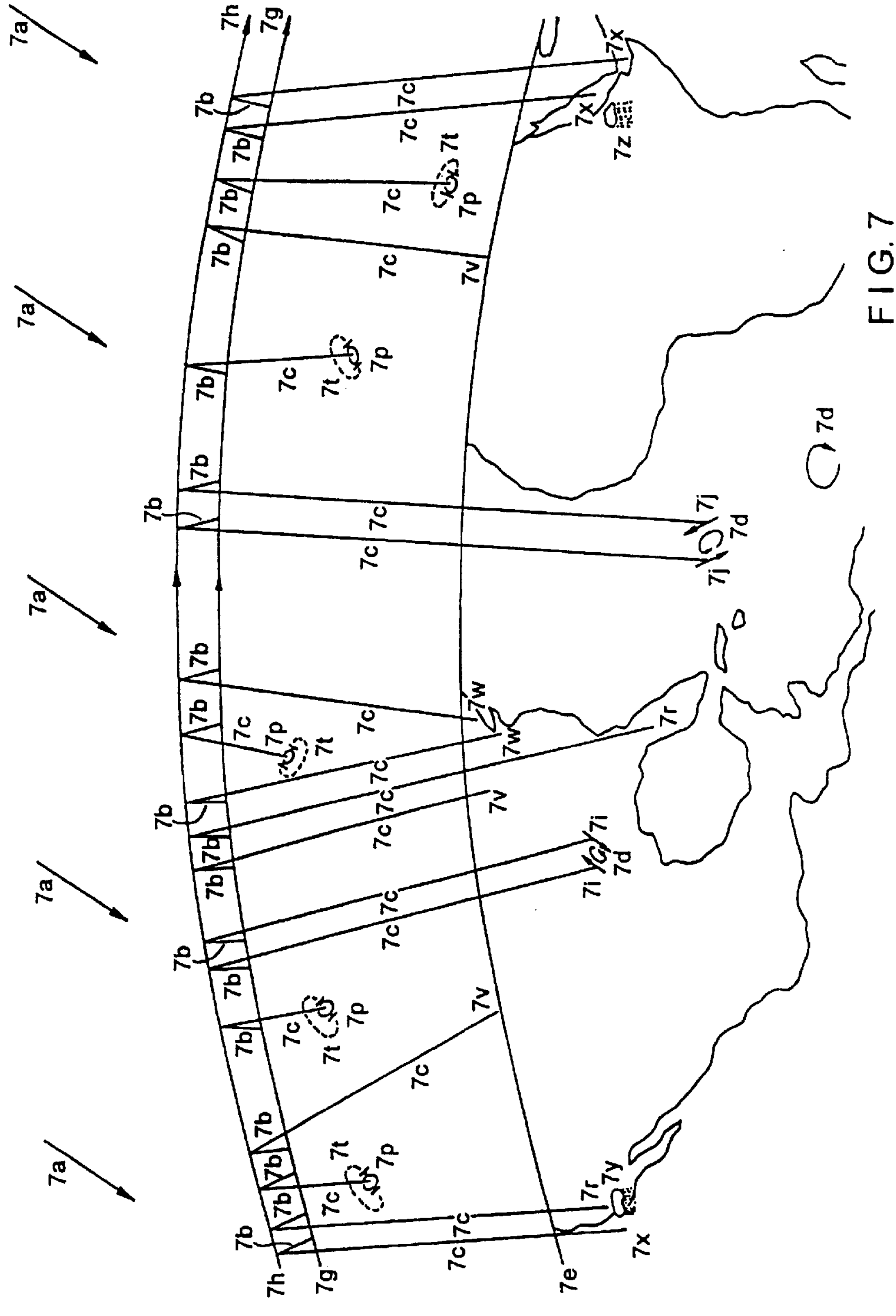
7a
7a
7a
7a
7a
7b
7c
7d
7e
7g
7h
7i
7j
7p
7r
7t
7v
7w
7x
7y
7z

FIG. 7

WEATHER MODIFICATION BY ARTIFICIAL SATELLITES

This is a divisional of prior application Ser. No. 08/480,883 filed Jun. 7, 1995, now U.S. Pat. No. 5,762,298, which is a continuation of prior application Ser. No. 07/675,880 filed Mar. 27, 1991, now abandoned.

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates to solar-based weather modification and energy transfer in such a manner as to improve people's living standards by reducing their energy cost and food cost and simultaneously improving the environment.

2. Description of the Prior Art

Art prior to the present application includes the following:

U.S. Pat. No. 4,402,480

Title: Atmosphere Modification Satellite

Inventor: D. M. Udell, Sep. 6, 1983

U.S. Pat. No. 3,613,992

Title: Weather Modification Method

Inventor: R. J. Knollenberg, Oct. 19, 1971

U.S. Pat. No. 4,371,135

Title: Solar Array Spacecraft Reflector

Inventor: J. E. Keigler, Feb. 1, 1983

U.S. Pat. No. 4,305,555

Title: Solar Energy System with Relay Satellite

Inventor: C. E. Davis, Dec. 15, 1981

In addition, the following publications are of interest:

Astronautical Guidance

by R. H. Battin, McGraw-Hill Book Co. 1964

Introduction to Radar Systems

by M. I. Skolnik, McGraw-Hill Book Co. 1962

and

Radar and the Atmosphere

by A. J. Bogush, Jr., Artech House 1989.

In contrast to much of the prior art, this invention employs only solar energy, earth satellites, remote sensing devices and ground control computers. No dispensing of chemicals is needed. The Satellite Weather Modification System (SWMS) is a local as well as a global weather modification tool. For example, SWMS can simultaneously bring rain to dry regions in southern California and drought regions in southern Ethiopia within in a few orbital periods of the Satellite Engines. Most remarkable of all is that these and many other tasks can all be accomplished by switching between some computer codes.

This invention is based on the realization that natural weather patterns always leave opportunities to permit regional weather modifications which are within the capabilities of the proposed Satellite Weather Modification System (SWMS). A fully developed SWMS has the potential to provide the following benefits.

1. Modify cooling, warming, or precipitations of selected regions of earth, thus modifying local and/or global weather patterns.

2. Create and maintain grass or forest lands to balance the rise of carbon dioxide contained in the earth's atmosphere, thus reducing the threat of global warming.

3. Transform some wasteland into productive farmland.

4. Increase farm, fishery and forestry productions at higher latitudes.

5. Increase electricity production on land directly from concentrated solar energy.

6. Increase hydroelectricity production by inducing precipitation in regions upstream of hydroelectric dams.

7. Reduce home and industry energy cost.

8. Reduce environmental damage by harsh weather.

9. Reduce damage to the environment due to the use of fossil fuel and nuclear fuel.

10. Create a unique self-supporting business based on SWMS, which can provide both weather-related and non-weather related services.

SWMS can be implemented employing components which are currently available, including means for reflecting solar rays or discharging infrared radiation or laser beams or incandescent lights.

SWMS improvements can be augmented gradually, still using conventional components, to include discharging frequency agile and or intensity agile energy beams or pulse trains from Satellite Engines (SEs). Each frequency agile energy pulse is designed to penetrate a specified air mass layer of specified composition. For example, these energy pulses can be used to modify a specified air mass inside a thick cloud.

SUMMARY OF THE INVENTION

This invention is an aerospace system designed to harness solar energy to modify the earth's weather. This system is called Satellite Weather Modification System (SWMS). This invention disclosure covers three areas. A first area involves the design and/or integration of three SWMS key subsystems. A second area involves specifying basic SWMS weather modification techniques. A third area involves the design of unique SWMS applications into self supporting businesses.

The first area of this invention disclosure covers the three SWMS key subsystems. They are:

1. Satellite Engines (SEs)

The basic function of an SE is to reflect solar energy or to transform solar energy into another form of energy before discharging it to a specified location. Each SE is a specially designed earth satellite. In a mature SWMS, many millions of SEs will be operating in a coordinated manner. As a minimum capability, each SE will be designed to reflect solar energy and simultaneously convert solar energy into electricity via solar cells. Advanced SE models will be capable of transforming solar energy into energy beams of different frequencies before discharging them. These advanced SEs can simultaneously discharge solar energy, frequency agile and intensity agile MMW (millimeter wave) or RF or microwave pulse trains, infrared radiations, laser light and incandescent light as the mission may require. All physical characteristics of each energy beam (solar or otherwise) are adaptively estimated by Ground Control Stations (GCSs). GCS also estimates each energy beam's guidance by defining its desired aim point trajectory. The criteria used to select and to guide each energy beam are based on an estimated specific thermodynamic change taking place in a specific air mass that satisfies some required weather condition changes.

Each SE will have gimbals panels (i.e., panels mounted on gimbals) on its outer surfaces. Each gimbals panel has a reflector on one side and solar cells on the other side. Also, gimbals-mounted antennae and lamps are attached to the SE structure frame. The SE controls the direction of each energy beam by controlling each panel's two gimbals orientations with respect to the SE structure frame. By using flywheels (or other methods), each SE also controls its structure frame orientation with respect to inertial space.

A minimum SWMS configuration requires that all SEs be placed in two non-coplanar orbits. These orbits are slightly inclined with respect to each other. One orbit will always be higher than the other orbit. Each SE structure frame will maintain an average angular rate of one revolution per year with respect to inertial space. During each orbit around the earth an SE will experience two different operating modes, depending on whether the SE is inside or outside earth's shadow. These modes are described below:

When SEs are outside earth's shadow, SEs in the higher orbit will rotate their solar cell panels (on one SE side) facing the general direction of the sun. At the same time SEs will rotate their reflector panels (on the other side of the SE) facing away from the sun. SEs in the lower orbits will have their reflector panels (on one SE side) facing the general direction of the sun and their solar cell panels (on the other side of the SE) facing away from the sun. This arrangement allows the solar rays first reflecting from the lower orbit SEs' reflectors to the reflectors on the higher orbit SEs, before finally re-reflecting these solar rays to earth (in FIG. **3** see incoming solar rays **3***a*, rays **3***b* reflected from the lower orbit to the higher orbit, and rays **3***c* reflected from the higher orbit into the earth's atmosphere).

When SEs are inside earth's shadow, SEs will rotate their reflector panels to face the reflector panels in the other orbit. This arrangement allows the solar rays to travel between the two orbits inside earth's shadow before discharge to earth (see rays **3***z* in FIG. **3**).

2. Remote Sensing Devices (RSDs)

The basic functions of RSDs are to continuously measure the actual atmospheric conditions of each specified weather cell. Ideally, there will be billions of RSDs of every kind distributed over the globe. These RSDs are used to continuously monitor the actual thermodynamics, dynamics, compositions and other physical conditions of the atmosphere, land and water surfaces, subsurfaces, sun and Satellite Engines. There will be different kinds of RSDs measuring different parameters. Some RSDs will be stationary on earth. Other RSDs may be located on Solar Powered Satellite Airplanes (SPSAs), on balloons or on SEs. All RSD measurements are time and location tagged before transmitting to their respective local Ground Control Stations (GCSs) for data correlation and analysis. Some measurements are sent directly to or measured directly by SEs.

3. Ground Control Stations (GCSs)

The basic function of GCSs is to provide guidance to each SE. This includes defining the characteristic of each energy beam and its aim point trajectory as functions of time. Each GCS includes a network of supercomputers, special algorithms, mathematical models, monitors, control panels, data storage devices, electronic libraries, data transmitters and data receivers. The main tasks of each GCS are: To maintain detailed information about local weather conditions, to integrate raw local RDSs measurements with filtered information about weather conditions of other locations (received from other GCSs), to extrapolate weather conditions to future time points, to monitor assigned Satellite Engines' (SEs') status, to evaluate assigned SEs' performances, to identify potential local weather patterns which can be modified by SWMS, to estimate potential weather modification options, to evaluate desired weather condition requests, to select optimal SEs groupings, to estimate SEs weather modification tasks options, to coordinate all SEs tasks, to assign and transmit detailed weather modification tasks to each assigned SE (i.e., specify energy beam's characteristic and its aim point trajectory as functions of time), to maintain communication between GCSs, RSDs and SEs, and, finally, to accept human inputs and controls.

The second area of this invention disclosure involves specifying SWMS weather modification techniques. This includes: Cooling the earth by reflecting solar rays back into deep space, warming local regions by redirecting solar rays to local areas, creating clouds by warming local wet surfaces and modifying local air mass velocity components by controlled heating of specific air masses to create pressure differentials.

The third area of this invention disclosure involves organizing SWMS unique services into self supporting business ventures. The SWMS services include: Weather modifications (cooling, warming, altering precipitations, altering air mass velocity components), supply of energy (direct solar power, indirect wind or hydroelectric power or solar powered airplane) and environment improvements (reclaiming wasteland, creating greenhouse farms, reducing bad weather damages, reducing fossil and nuclear fuels usage). Included in SWMS applications is an invention of a Solar Powered Satellite Airplane (SPSA). SPSA comprises a hinged zero lift solar cell disk supported by two wings as shown in FIG. **4**. These SPSAs can serve as global communication platforms.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** Satellite Weather Modification System (SWMS) key subsystems Functional Flow Overview.

FIG. **2**. External Features of A Typical Satellite Engine (SE). FIG. **2** consists of FIG. **2**A (top view), FIG. **2**B (right side view), FIG. **2**C (front view), and FIG. **2**D, showing a typical panel or antenna gimbal rings arrangement.

FIG. **3**. Satellite Engines (SEs) Orbital Operations Overview.

FIG. **4**. Solar Powered Satellite Airplane (SPSA) Design. FIG. **4** consists of FIG. **4**A, which is a sectional view taken along any of the section lines **4**A—**4**A in FIG. **4**C and looking in the direction of the arrows; **4**B, which is a front view; **4**C, which is a top view; and **4**D, which is a side view.

FIG. **5**. Overview Of A Ground Control Station (GCS) Functions.

FIG. **6**. Techniques To Modify Air Mass Velocity Component.

FIG. **7**. Some Satellite Weather Modification System (SWMS) Applications.

Detailed descriptions of these figures are presented below. Important items in each figure are identified in the text by a bold faced number representing the figure number. It is then followed by a fold faced lower case letter to identify the item in the figure. Example: **4***d* identifies item labeled **4***d* in FIG. **4**.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

While there are shown and described the preferred embodiments of the invention, it is to be distinctly understood that the invention is not limited thereto, but may be otherwise variously embodied and practiced within the scope of the appended claims.

To facilitate an understanding of the invention, a number of features are identified herein by letter(s) and a number inside a bold square bracket, as: [letters/number] or [letter-number/number]. Letter(s) or letter-number before a/0 identify major features of the invention. Each non-zero number following the / represents a feature that, while important, is less important than the major features. The major features are:

[SWMS/O] = Features relating to Satellite Weather Modification System (SWMS) designs.
[SE/0] = Features relating to Satellite Engine (SE) designs.
[OP/O] = Features relating to SEs operations.
[RSD/O] = Features relating to Remote Sensing Devices (RSDs) integration.
[GCS/O] = Features relating to Ground Control Stations (GCSs) designs.
[C/O] = Features relating to control energy beam directions.
[F/O] = Features relating to Frequency Agile and Intensity Agile energy beams' selections.
[W#/O] = Features relating to SWMS basic weather modification techniques (# = 1, 2, 3, 4).
[A#/O] = Features relating to SWMS weather modification applications and non-weather related applications (# = 1, 2, 3, 4, 5, 6).

A stand-alone bold faced number followed by a bold faced lower case letter refers to a correspondingly labeled item in the figure identified by the bold faced number. Example: **3***c* represents an item identified by **3***c* in FIG. **3**.

V.1 [SWMS/#]: Features Relating to Satellite Weather Modification System (SWMS) Design

A major feature [SWMS/0] is covered in this section. Five additional features [SWMS/1], . . . , [SWMS/5] are covered in separate paragraphs.

An elaboration of the features is set out below:

[SWMS/1]

The invention employs satellites to harness the solar energy and use this energy to modify selected local weather conditions and/or supply power to various non-weather related applications.

[SWMS/2]

The SWMS modified local weather conditions are, under certain situations, used to modify global weather conditions.

[SWMS/3]

This Satellite Weather Modification System (SWMS) includes the following three key subsystems:

Satellite Engines (SEs) **1***e*.

Remote Sensing Devices (RSDs) **1***c*.

Ground Control Stations (GCSs) **1***d*.

[SWMS/4]

Most SEs are in earth orbits, but some SEs can be installed on land, or airplanes or on ships.

[SWMS/5]

In an SWMS the energy and information flow between sun **1***a*, earth **1***b*, RSDs **1***c*, GCSs **1***d* and SEs **1***e*, as represented in FIG. **1**, are identified and listed below:

1f	Solar energy from the sun to the SEs.
1g	Solar energy from the sun to earth.
1h	Solar activities, such as sun spots, which can affect earth's weather and which are measurable by instruments.
1i	Reflected solar energy beams and/or frequency agile energy beams discharged from SEs to earth surface or into the atmosphere or to deep space or to Solar Powered Satellite Airplanes.
1j	Coded solar or frequency agile energy beams reflected back to SEs instruments or to RSDs instruments.
1s	Solar energy beams or frequency agile energy beams from one SE to another SE.
1k	Solar energy or frequency agile energy reflected back to SEs' reflector panels or the SEs' solar cell panels.
1l	Energy or signals from earth or from atmosphere received by passive RSDs.
1m	Signals transmitted and received by active RDSs.
1n	Time and space tagged measurement data transmitted from RSDS to GCSs at regular intervals.
1p	Command and control messages from GCSs to RSDs.
1q	Status reports from SES to GCSs at regular intervals.
1r	Time tagged detailed assignments (each energy beam's characteristics and its AAP (Adjusted Aim Point) trajectory) for each SE from GCSs at regular intervals.
1t	Desired local and global weather condition requests.
1x	Data and status display to off-line devices.
1y	Off-line analysis to update GCS math models, algorithms, operations and library data base.
1z	Data flow between GCSs. Each GCS takes its turn to coordinate the activities of all GCSs.
1u	Human (manual) controls.

V.2 [SE/#]: Features Relating to Satellite Engine (SE) Design

A major feature [SE/0] is covered in this section. Six additional features [SE/1], . . . , [SE/6] are covered in separate paragraphs.

[SE/1]

SEs can have any shape. In this description it is assumed that the basic structure frame of each Satellite Engine (SE) has an outer contour resembling a flat rectangular box of dimension L (Length)×W (Width)×H (Height), where H is the smallest dimension. All six outside surfaces of this flat box are either fully or partially covered with reflecting materials.

[SE/2]

On top of each L×W surface is an array of n×m panels.

[SE/3]

one side of each n×m panel is made of a reflecting material. It is used to reflect solar rays or laser lights. The other side of each n×m panel is made of solar cells. They are used to convert solar energy into electrical currents.

[SE/4]

Each n×m panel is independently mounted on gimbals on top of each L×W surface. The construction is in such a way that each one of the n×m panels has at least two independent degrees of attitude freedom, **2***p* and **2***r*, with respect to the SE structure frame **2***f*.

[SE/5]

FIG. **2** illustrates n×m panels on each L×W side of a Satellite Engine structure frame. Gimbals rings on SE's L×H and W×H sides **2***a* **2***b* **2***c* **2***d* are designed to provide independent attitude controls of antennae or lamps or any direction pointing devices. Some antennae are used to discharge frequency agile and intensity agile energy beams to earth. Some antennae are used to transmit or receive frequency agile energy beams between SEs. Other antennae are used for communication purposes. Lamps are used to discharge to earth infrared radiations, laser lights or incandescent lights. The number of side gimbals antennae or devices are not limited to only four **2***a* **2***b* **2***c* **2***d* as illustrated in FIG. **2**.

[SE/6]

Additional features displayed in FIG. **2** are identified as follows:

2f	SE structure frame. The LxWxH box.
2h	A typical set of gimbals rings which measure and control each reflector/solar-cell panel's attitudes 2r 2p with respect to SE structure frame 2f.
2p	One of the two degrees of attitude freedom of each panel with respect to SE structure frame.
2r	One of the two degrees of attitude freedom of each panel with respect to SE structure frame.
2q	Reflector/solar-cell panel's possible third degree of attitude freedom with respect to SE structure frame. This will require larger spacing between panels.
2a	Frequency agile and intensity agile RF or MMW or Microwave energy pulse train transmitting antennae.
2b	Reflector/solar-cell panels.
2c	Gimbals mounted infrared lamps or incandescent light lamps or laser antenna
2d	Reflector/solar-cell panels.

In addition, each SE has a number of subsystems residing in its structure frame **2*f***. They are not shown in FIG. **2**, but they are identified below as:

2w	Momentum wheels or other means are used to control and maintain attitude stabilization of SE structure frame 2f with respect to inertial space (see feature (C/#] below)
2n	Navigation subsystem.
2p	Energy transformation subsystem: Transform dc current from solar cells into infrared, frequency agile and intensity agile energy pulses (see features [F/#]).
2c	Computer subsystem.
2z	Communication subsystem.
2i	Retractile receiver designed to receive energy beams from other SEs.
2y	Power regulator.
2e	Energy storage devices are used to store solar energy collected when SE is outside earth's shadow. This energy can be used when SE is inside earth's shadow.
2u	Optical camera package (optional or as part of RSDs).
2v	Synthetic Aperture Radar package and antenna (optional or as part of RSDs).
2w	Thermal imaging camera package (optional or as part of RSDs).
2x	Infrared camera package (optional or as part of RSDs).

V.3 [OP/#]: Features Relating to SEs Orbital Operations

A major feature [OP/1] is covered in this section. Ten additional features [OP/1], . . . , [OP/10] are covered in separate paragraphs.

[OP/1]

As a minimum SWMS configuration, all orbiting SEs are placed in a pair of non-coplanar orbits. These orbits are inclined slightly with respect to each other. In this case, all SEs are apportioned about equally to and distributed uniformly along these two orbits. If both orbits are rotated into a single plane, one orbit will always be slightly higher than the other orbit. The orbits are thus nonintersecting and preferably noncoplanar. Unless otherwise specified, the following descriptions will be for SEs operations on a pair of nonintersecting-noncoplanar orbits.

[OP/2]

At the center of each SE will be the origin of two coordinate systems both moving with the SE along its orbit. One set of solar-fixed axes has one axis that constantly points at the center of the sun. These coordinate axes have an average angular rate of one revolution per year with respect to inertial space. Another set of axes are fixed with respect to the SE structure frame. During normal operations that set will continuously tilt slightly with respect to the solar-fixed axes as required by the mission.

A third set of completely inertial coordinate axes is needed to relate vectors from one set of solar-fixed axes to another.

[OP/3]

In a typical SWMS operation, each SE **3*n*** in the lower orbit maintains its n×m solar reflector panels (on one L×W surface) facing the general direction of the sun. Simultaneously, each lower orbit SE's other L×W surface will have its n×m solar cell panels facing away from the sun. At the same time each SE **3*m*** in the higher orbit will have its n×m solar cell panels (on one L×W surface) facing the general direction of the sun. Its other n×m solar reflectors (on its other L×W surface) will be facing away from the sun.

Each SE experiences two modes of operation during each orbit around the earth. These two modes are defined according to the position of the SE which is inside or outside the earth's shadow. Consider these two cases:

[OP/4]

First, when SEs are outside earth's shadow: In this case the SEs' reflector/solar-cell panel arrangements are as specified in section [OP/3]. These arrangements allow incoming solar rays **3*a*** from the sun to be reflected as shown at **3*b*** from the lower orbit SEs' reflectors to SEs reflectors on the higher orbit before finally being re-reflected as solar rays **3*c*** to earth.

[OP/5]

The electricity generated by the higher orbit SEs' solar cell panels is used to supply each SE's internal needs. Excessive electricity is transformed to different forms of energy beams and discharged to earth via antennae or lamps or stored for later use. At the same time, each lower orbit SE's **3*n*** solar cell panels on one of its L×W surfaces are facing away from the sun. These solar cells receive solar rays **3*r*** reflected from earth or **3*b*** reflected from other SEs. Electricity generated by these solar cell panels is used to supply each SE's internal needs. Any excess electricity is transformed to different forms of energy beams and discharged to earth via antennae or lamps or to another SE or stored for later use.

[OP/6]

Second, when SEs are inside earth's shadow and when solar or laser rays are required in the shadow: In this case, selected SEs in the shadow and in both orbits **3*m***' and **3*n***' have their n×m reflector panels facing away from their respective SE structure frames. These SEs are identified as **3*m***' and **3*n***' in FIG. **3**. This arrangement allows solar rays or laser beams to be passed alternately between these SEs **3*m***' and **3*n***'. This scheme moves energy beams **3*z*** in a zig-zag pattern into the earth's shadow **3*t*** before discharging to earth **31**' or to a high flying SPSA **3*p***.

[OP/7]

Some SEs are equipped with large capacity energy storage devices. Energy collected and stored during daytime can be discharged during nighttime.

[OP/8]

FIG. **3** illustrates a sample case of two nonintersecting-noncoplanar orbits slightly inclined with respect to the equatorial plane. In FIG. **3**, each SE is represented by three short dashes parallel and next to a longer dash spanning the three short dashes. SEs on the higher orbit **3*h*** are labeled **3*m***. SEs on the lower orbit **3*g*** are labeled **3*n***. The three short dashes or **3*m*** **3*n*** represent the L×W side of an SE with its n×m solar cell panels facing away from the SE structure frame. The corresponding longer dash of **3*m*** or **3*n*** represents the L×W side of an SE with its n×m reflector panels facing away from the SE structure frame. In general, when SEs are outside earth's shadow **3*t***, all SEs **3*m*** in the higher orbit **3*h*** have their solar cell panels (three short dashes) facing the general direction of the sun. Similarly, all SEs **3*n*** on the lower orbit **3*g*** have their reflector panels (one long dash) facing the general direction of the sun.

[OP/9]

Other features and labels in FIG. **3** are identified as follows:

3a	Solar rays directly from the sun.
3b	Solar rays reflected once from one SE to another SE or to earth.
3c	Solar rays reflected twice from two SEs (3n and 3m) before being reflected to earth or to an SPSA.
3d	Dash lines are used to represent the path of frequency agile & intensity agile RF, MMW energy beams, infrared radiations, laser beams or incandescent lights emitted from SE antennae or lamps.
3e	Earth's direction of rotation about its spin axis (earth being viewed from a position over the northern hemisphere).
3f	Direction of earth's orbit path around the sun (earth being viewed from a position over the northern hemisphere).
3g	Lower orbit of a typical pair of nonintersecting-noncoplanar orbits.
3h	Higher orbit of a typical pair of nonintersecting-noncoplanar orbits.
3i	Average velocity vector of an air mass independent of current SEs actions.
3j	Average velocity vector of an air mass induced by current SEs actions.
3k	Vector sum of 3i and 3j.
3l	Earth's land surfaces.
3l'	Illuminated sports stadium (e.g., a stadium lighted during night games) or orange groves warmed by solar rays 3z during freezing nights.
3m	SE in higher orbit 3h. Its nxm solar cell panels (on one LxW side) face the general direction of the sun 3a. Simultaneously, its other nxm panels have their reflectors facing away from the sun.
3n	SE on the lower orbit 3g. Its nxm reflector panels (on one LxW side) face the general direction of the sun 3a. Simultaneously, its other nxm panels have their solar cells facing away from the sun.
3m'	Selected SEs in higher orbit 3h in earth shadow. These SEs are used to transmit solar or laser rays 3z into the shadow.
3n'	Selected SEs in lower orbit 3g in earth shadow. These SES are used to transmit solar or laser rays 3z into the shadow.
3w	Earth's ocean, lake or other wet surface.
3p	High altitude and long endurance Solar Powered Satellite Airplanes (see FIG. 4). SPSA continuously receives solar rays or MMW energy or laser beams from SEs. SPSA can also receive energy from ground.
3q	Cloud or air mass whose compositions, thermodynamic and dynamic properties are being modified by SEs.
3r	Reflection of solar enerqy from the clouds, ocean or land back to space.
3t	Earth's cone shaped shadow.
3z	Solar rays reflecting alternatively between SEs 3m', 3n' on different orbits 3h, 3g, when these SEs are inside earth's shadow 3t. These energies 3z can be used to power SPSAs 3p or to warm cold ground patches 3l' at night.

[OP/10]

The n×m panels may be arranged in a way slightly different from the one described so far. It is possible to mix the number of solar-cell panels and the number of reflector panels on each L×W side of each SE. This arrangement applies to all SEs in both orbits **3***h*, **3***g* as follows: Each L×W side of each SE will have (n×m−k) reflector panels and k solar-cell panels facing away from its SE structure frame, where k=0, 1, . . . , (n×m). For each L×W side, the exact k number is dependent on each SE's internal power consumption need and its mission requirements. This arrangement is especially useful for SEs in geo-synchronous orbit.

V.4 [RSD/#]: Features Relating to Remote Sensing Devices (RSDs) Integration

A major feature [RSD/0] is covered in this section. Two additional features [RSD/1] and [RSD/2] are covered in separate paragraphs.

[RSD/1]

Remote Sensing Devices (RSDs) **1***c* comprise millions of independently operated sensors. These RDSs are located on the ground, on the water, in the air, on SPSAs, or on SEs. RSDs perform independent measurements such as monitoring sunspots, measuring the composition and thermodynamics properties of a particular atmospherical air mass, monitoring the effect of SEs' energy beams, etc.

[RSD/2]

Each sensor continuously makes redundant measurements. Each measurement is time and location tagged and transmitted to a specified Ground Control System (GCS) **1***d* computers at regular intervals. These data are in standard formats so that they can be quickly correlated, analyzed and accepted by various algorithms and mathematical models in the GCSs **1***d* computer network.

V.5 [GCS/#]: Features Relating to Ground Control Stations (GCSs) Designs

GCSs provide the guidance of each energy beam. This includes the selection of each energy beam's characteristic and its corresponding Adjusted Aim Point (AAP) trajectory. A major feature [GCS/0] is covered in this section. Three additional features are [GCS/1], GCSs subsystem overview; [GCS/2], simplified GCS tasks design; and [GCS/3], a typical matured GCS subsystem tasks design and partition.

[GCS/1] GCS Overview

Each Ground Control Station (GCS) **1***d'* includes a network of interconnected supercomputers, electronic libraries, data receivers, data storage devices, data transmitters, visual display devices and control consoles. There are hundreds of GCSs **1***d* in the SWMS. Each GCS monitors weather conditions within a designated region of earth. Each GCS receives designated RDSs measurements **1***n'*, local weather requests **1***t'* and data from other GCSs **1***z*. Each GCS also transmits estimated tasks **1***r'* to a specified group of SEs and receives data from other GCSs **1***z*. Data generated by each GCS are broadcast to all other GCSs **1***z*. Therefore, each GCS will have detailed global weather information and status of all SEs at discrete time points. Each GCS will take its turn to be the master GCS, which coordinates the operations of all GCSs in the SWMS.

(NOTE: the notations **1***d'*, **1***n'*, **1***r'*, **1***t'* represent designated GCSs, selected RSDs' data, specified SEs' tasks and local weather requests respectively. Their corresponding notations without the ' given in FIG. **1** are for all GCSs, all RSDs data, all SEs tasks and all weather requests.)

[GCS/2] Simplified GCS Tasks Designs

In most SWMS applications, SEs' assignments involve only reflecting solar rays. This is the case for SWMS during its initial stage of development. In this stage of development a GCS's principal task involves defining AAPs (Adjusted Aim Points) trajectories of each energy beam (centerline of each reflected solar ray) in some earth fixed coordinate system.

The available solar energy and the directions of incoming and reflected solar rays to and from each n×m reflector panel can be calculated as a function of time. These calculations use measurements of SE structure frame orientations with respect to inertial space, orientation measurements of each reflector panel gimbals with respect to its SE structure frame and orbital mechanics.

On the other hand, the desired aim point of each reflected energy beam and its trajectory are both specified by GCSs. This is done by distributing the aim points over the surface that is to be warmed by the solar rays. These aim points are then adjusted to take into account the atmospherical refraction and absorption of each energy beam and transform desired aim points to a corresponding set of AAPs (Adjusted Aim Points). A set of AAP vectors will be defined in an earth-fixed coordinate system. Each AAP point will be at the tip of its corresponding AAP vector. Next, AAPs trajectories are defined based on the amount of energy and rate of energy that is to be deposited onto the desired space. These estimated AAPs trajectories are transmitted to their corresponding SEs.

An LOS vector is defined as a vector from the center of any reflector panel to its corresponding AAP. Both this LOS vector and its corresponding reflecting solar ray vector are expressed in the same coordinate system, based for example on the SE structure frame fixed axes. Then the vector difference between these two vectors at a given time defines an error vector. This error vector specifies the required change of available energy beam's direction at that time.

The unit vector of each desired direction will be transformed to SE structure frame fixed coordinate axes. Assuming the SE structure frame follows a predetermined orientation schedule with respect to inertial space, then the desired gimbals angles of each reflector panel can be calculated from its unit desired direction vector. Also, each desired direction vector must satisfy the solar rays' availability constraints.

These computations must be repeated continuously many times a minute. This will provide the required error signal to the stepping motors to torque each gimbals ring. This scheme will control each reflected solar ray to follow its AAP trajectory in a timely fashion.

[GCS/3] Typical Matured GCS Tasks Designs and Partitions

FIG. **5** graphically illustrates typical SEs' assignments (energy beams' characteristic and their AAPs trajectories) estimation process designed for a more matured SWMS. Key functions are identified by blocks and connected by high speed data busses (dashed lines). The main processing sequence is: **5*b*** **5*d*** **5*e*** **5*f*** **5*e*** **5*k*** **5*m*** **5*n*** **5*b***. Other functions of FIG. **5** are identified below:

5*a* A GCS library. This includes individual mathematical models **5*q***, Integrated global-weather/SWMS (Satellite Weather Modification System) model **5*p***, Historical weather statistics **5*i***, Encyclopedia of energy beam transmission, refraction, absorption, and reflection characteristics in all kinds of weather conditions **5*t***, Encyclopedia of meteorology **5*s***, past SEs weather modification performance data **5*v***, Catalogued encyclopedia of SWMS Satellite Engines performance **5*f*** permitting rapid search for SEs task assignments that will satisfy the new weather modification constraints specified by **5*e***.

5*q* Individual models include: Sun, sunspots, solar radiation, lunar orbital motion, solar wind intensity distribution, earth orbital motion, earth's gravitational model, atmospherical layer models, ocean current models, earth surface and subsurface models, natural and man-made discharges, absorption, refraction models, etc. Satellite Engines model, SE system performance models, interactions between atmosphere models and all other models, etc.

5*i* Historical global and local weather data statistics, including such data as annual rainfall distributions, annual temperature distributions, seasonal weather conditions, etc.

5*t* Encyclopedia of the interaction of energy beams with atmosphere, land mass and sea surfaces.

5*f* Encyclopedia of SEs performance data classified according to: energy types and characteristics, guidance law employed (AAPs' trajectories), initial atmospherical conditions, required atmospherical conditions changes, Satellite Engines sub-groupings. These data are continuously updated by off-line simulations **1*y*** and by data from actual SEs past performances **5*v***. Data **5*f*** output back to **5*e*** will include SEs groupings, each energy beam characteristics and AAPs' trajectories, all expressed as functions of time.

5*l* Estimate the desired weather conditions in next (Tb,Tc) interval based on past weather condition statistics. Desired weather is expressed in terms of weather characteristic vectors for each affected weather cell.

5*g* Special desired weather condition requests. All requests are expressed in terms of weather characteristic vectors sequences for each weather cell. Both weather cells and weather characteristic vectors are defined in **5*d***.

5*b* Update integrated mathematical model using designated RSDs **1*n*'**, designated SEs' status reports **1*q*'**, and updated data from other GCSs **1*z***. The results of these updates yield an up-to-date integrated global weather math model with Satellite Engines' math models integrated in the total model. The resultant model is to be used for further extrapolations **5*d***, for storage in local GCS library **5*a*** and for broadcasting to all other GCSs **1*z***.

5*c* Estimations of selected SEs' task assignments for the next (Tb,Tc) interval based on anticipated requirements from **5*l*** and **5*g***. These estimates are passed on to **5*d***.

5*v* Off-line evaluation of SEs performances in the previous (Tb,Tc) interval using actual RSDs measurements **1*n*'** and previous cycle SEs' assignments **1*r*'** adjusted by actual SEs status reports **1*q*'**. Results of SEs performance evaluations are used to update GCS library data **5*t***, **5*f*** and **5*q***.

5*d* Forward extrapolation of integrated mathematical model **5*b***, from t=Ta to t=Tc, in equal dT increments, where dT=(Tc−Ta)/N and N=a large even integer. If the actual RSDs measurements for the (Ta,Tb) interval are in agreement with previously estimated weather parameters, then this extrapolation process can be reduced to only the (Tb,Tc) interval.

The atmosphere is subdivided into weather cells. The average thermodynamics, dynamics and air composition parameters inside each weather cell are represented by a vector called the weather characteristic vector. Each weather cell is represented by its weather characteristic vector for a fixed time duration.

5*e* First, find the required weather changes by taking the differences between the extrapolated weather conditions from **5*d*** and the corresponding desired weather conditions from **5*h***. This is done by manipulating the corresponding weather characteristic vectors.

Next, initialize a search in **5*f*** for the SEs' task assignments during the (Tb,Tc) interval. SE task assignments from **5*f*** will include parameters which define each energy beam's characteristics and its Adjusted Aim Point (AAP) trajectory. This search uses the initial weather conditions from **5*d***, required weather changes from **5*d*** and **5*h***. However, a sequence of desired weather conditions from **5*h*** is selected at reduced constraint levels. This will ensure that at least one set of the answers found from **5*f*** will later satisfy the criteria in **5*k***.

Found in **5*f*** are SE's task assignments which include each one of its energy beam's characteristics as a function of time and their corresponding Adjusted Aim Point (AAP) trajectories.

Each SE assignment parameter is found by matching stored **5***f* energy absorption patterns or thermodynamic changes patterns with their corresponding required change patterns **5***e*.

5*k* The SEs' tasks assignments found in **5***f* are evaluated by performing parallel simulations. The results of these simulations should determine which set of SEs assignments best satisfies the desired weather conditions imposed on the SEs by **5***h*. All calculations are done in terms of weather characteristic vectors at the weather cells' levels **5***d*.

5*m* Define detailed Satellite Engines' task time lines based on optimal selection from **5***k*. These assignments are transmitted to the corresponding SEs **1***r*'. A second set of these assignments is sent to **5***c* for the next computation cycle. A third set of these data is sent to the GCS library data bus to be broadcast as indicated at **1***z* to all GCSs in the SWMS. This last step is necessary to take care of conditions when a GCS is reassigned to manage a different set of SEs in the next computation cycle.

5*n* Initialize the next computation cycle. The current (Tb,Tc) interval is shifted down to become the next (Ta,Tb) interval, and a new (Tb,Tc) interval is selected with the old Tc set to equal to the next Tb. The next computation cycle repeats at **5***b*.

The descriptions set out above outline a typical process to define SEs grouping, each energy beam type, each energy beam's physical characteristics and its Adjusted Aim Point (AAP) trajectory. These parameters are transmitted to their corresponding SEs **1***r*'.

NOTE: **1***x*, **1***y*, **1***z*, **1***r*', **1***q*' and **1***n*' refer to the same items identified in FIG. **1**. The prime ' denotes that these data belong to only a subgroup of SEs or subgroup of RSDs.

V.6 [C/#]: Features Relating to Control Energy Beam Directions

One major feature [C/0] is covered in this section. Two additional features [C/1] and [C/2] are covered in separate paragraphs.

[C/1] Overview

Each SE shall release its energy beams in a timely fashion following its AAP trajectory as specified by its GCS **5***m*. SE's control subsystem aims each energy beam at its AAP (Adjusted Aim Point) and makes each energy beam follow its AAP trajectory in a timely fashion.

[C/2] Direction Controls

This energy beam directional control task is partitioned into two sub-tasks. First, control the orientation of SE's structure-frame axes with respect to inertial space. This is done by using inertial flywheels or other means. Second, control the orientation of solar-reflector/solar-cell panels, antenna dishes, or lamp axes with respect to the SE's structure-frame axes. This is done by measuring the actual gimbals angle settings and torquing each gimbals ring to its desired orientation. Each energy beam's desired orientation at a given time is defined by its unit vector from reflector (or antenna or lamp) center to its AAP at that time. Using, for example, the directional cosine technique, these desired unit vectors can be transformed into desired gimbals angles. A more detailed description on how this is done has been presented above at [GCS/2].

V.7 [F/#]: Features Relating to Energy Beam Frequency and Intensity Selections

A major feature [F/0] is covered in this section. Seven additional features [F/1], . . . , [F/7] are covered in separate paragraphs.

Solar rays will be the primary energy beams discharged by the SEs. However there are times when other types of energy beams are required to penetrate a cloud layer, or to heat up a specified air mass of specific composition. In these cases, frequency agile and intensity agile energy beams are used. The frequency of an energy beam determines its ability to penetrate a layer of atmosphere of a specified composition. It also determines the degree of absorption of a given frequency energy beam by an air mass of a particular composition. Weather cells air mass compositions are estimated by Remote Sensing Devices (RSDs).

[F/1]

A specific energy beam's frequency and intensity will be selected to satisfy a given depth of atmospheric penetration and to yield a specified energy absorption contour or temperature elevation contour.

[F/2]

The change of energy into heat is monitored by RSDs, such as infrared sensors or thermal imagery cameras. Or it can be estimated from specially designed ground receivers. These receivers measure the presence of frequency agile energy beams which are frequency, amplitude or phase modulated. These RDSs data are used by GCS or by SEs directly, thus enabling adaptive adjustments of energy beams' frequencies and intensities.

[F/3]

Energy intensity at a given location is also controlled by energy beam guidance laws which specify each energy beam Adjusted Aim Point's (AAP's) trajectory, the speed AAP moves along its trajectory, dwell times, locations on an AAP trajectory and the shape of an AAP trajectory.

[F/4]

Selection of each frequency agile energy beam's frequency rests on the knowledge of the energy beam's attenuation due to atmospherical absorption, reflection and refraction. These data are stored in the GCS electronic library **5***t* as dB/Km (or watts or % transmission loss per kilometer of the energy beam path) vs frequencies for different atmospheric compositions.

[F/5]

For a given atmospheric composition the dB/Km values show a sharp drop within certain narrow frequency bands called windows (Radar and the Atmosphere by A. J. Bogush, Jr., Artech House 1989). Near the left end of each window, the parameter dB/Km experiences a sharp drop in value (dB/Km/f<<0). Near the right end of each window, the parameter dB/Km experiences a sharp rise in value (dB/Km/f>>0). The frequency of each frequency agile pulse (or beam) is selected near either end of these frequency windows to permit injection of frequency agile energy pulses into selected regions of the atmosphere.

[F/6]

Energy levels are selected to achieve the rates at which the accumulated energy absorption will change the thermodynamic and dynamic parameters of the atmosphere in a predetermined way. An example of such a rate is the rate of average temperature rise within a group of weather cells.

[F/7]

Furthermore, the energy level must satisfy safety requirements (Introduction to Radar Systems by M. I. Skolnik, McGraw-Hill Book Co. 1962).

V.8 [W#/#]: Features Relating to Basic Weather Modification Techniques

There are four major features [W1/0], . . . , [W4/0] covered in respective sections. A total of 12 (=6+2+2+2) additional features are set out separately. These additional features are:

[W1/1], . . . , [W1/6],
[W2/1], [W2/2],
[W3/1], [W3/2],
[W4/1] and [W4/2].

Basic techniques employed by SWMS to modify weather are listed below. Other weather modification requirements can be met by making different combinations of these basic techniques. These techniques are identified as basic weather modification techniques #1 through #4 ([W1/0], . . . , [W4/0]).

V.8.1 [W1/#]: Features Relating to Modifying Global Weather Conditions

This technique employs a very large number of Satellite Engines (SEs) to modify global weather conditions by blocking or redirecting solar rays. One major feature [W1/0] is covered in this entire section. Six additional features [W1/1], . . . , [W1/6] are covered in separate paragraphs.

[W1/1]

This method involves using millions (or even billions) of Satellite Engines (SEs) to block or redirect solar rays. Typically, this requires a large number of low inclination and non-coplanar orbits near the equatorial plane. When the earth requires cooling, the Ground Control Stations (GCSs) simultaneously command all SEs to rotate their n×(m−k) (k=0, 1, . . . <<m) reflector panels (on one L×W side) to face the general direction of the sun and away from earth. These SEs block part of the solar energy from reaching the earth by reflecting the solar energy away from earth and thus assist the earth's cooling process.

Each SE will reserve k solar cell panels (along with n×(m−k) reflector panels) to face the sun. These solar cell panels are used to generate electricity for the SE's internal use. The value of k is such that k=0, 1, 2, . . . , <<m. The k=0 case implies that solar cells located on the back side of each reflector panel or on its other L×W surface can generate sufficient electricity for the SE's internal use.

[W1/2]

The reflected solar rays' guidance is done by specifying the Adjusted Aim Points (AAPs) of each reflected solar energy beam and trajectories of these AAPs in deep space away from earth. In this case the AAPs can be simply computed as the line-of-sight aim points, because the reflected solar rays will have little or no interference (refraction) by the atmosphere.

[W1/3]

It is to be noted that when the equatorial regions become cooler, the average temperature differences between the polar regions and the equatorial regions are reduced. Therefore this technique may be used to modify global weather patterns or to create favorable local weather conditions which can be further modified by SWMS.

[W1/4]

When controlled cooling is not required, GCSs command some or all SEs to direct their collected solar energy to other SEs or to earth directly. Again, this is done by specifying the AAPs of each energy beam and their trajectories to the desired locations on earth or at specified Satellite Engines' reflector panels or solar cell panels. This process selectively cools certain regions but at the same time heats up other regions. This technique can also be used to supply energy to specific devices. All these tasks are under GCSs control.

[W1/5]

If these temperature differential regions are sufficiently close, an air mass motion is introduced which forces the air mass to move from the high pressure region to the low pressure region. This technique can be used to move an air mass of specific composition from one location to another.

This technique can also be used to modify the jet stream path and therefore modify the weather on either side of the jet stream.

[W1/6]

Distributing AAPs over a small wet surface and aiming solar rays at these AAPs will cause evaporation. This will increase the humidity of air above the wet surface. As the humid air rises to higher altitudes clouds will form. These clouds can be used to cool local regions by blocking the sun's rays. They can also cause precipitation.

V.8.2 [W2/#]: Features Relating to Modifying Local Weather Conditions

This technique is similar to feature [W1/6] but requires far fewer Satellite Engines (SEs). Also, SEs orbits are paired as specified in feature [OP/0]. One major feature [W2/0] is covered in this entire section. Two additional features [W2/1] and [W2/2] are described in separate paragraphs.

[W2/1]

This technique can be used to warm a local region. This is done by distributing AAPs (of all SEs' reflected energy beams) over the region to be warmed. Each reflected energy beam will be assigned to dwell at a given AAP for a given duration, or each AAP trajectory will be limited to circle within a desired region.

[W2/2]

This technique can be used to create a high humidity air mass or to form clouds. This is similar to feature [W2/1] above, except that the reflected solar rays are aimed at AAPs distributed over a wet surface, such as a lake or a patch of ocean. This process is originated by Ground Control Stations (GCSs) defining the AAP trajectory of each energy beam over the wet surfaces. Satellite Engines (SEs) then reflect solar rays to their respective AAPs. This process is essentially the same as one that occurs in nature, except that both the wet surface area and the timing are selected and controlled by SWMS. Also, the intensity of the solar ray can be made more concentrated than what occurs in nature. This process will increase the moisture content of the air mass above the heated wet surface. As the high humidity air rises to higher altitudes, it begins to cool and form clouds.

V.8.3 [W3/#]: Features Relating to Moving an Air Mass Upwards

One major feature [W3/0] is covered in this entire section. Two additional features [W3/1] and [W3/2] are covered in separate paragraphs.

[W3/1]

This technique is designed to initiate and/or modify an air mass's upward motion. This is done by adaptively aiming energy beams of specified characteristics at specified AAPs. These AAPs are distributed in a vertical column inside or near the air mass that is to be moved. The type, frequency, intensity and AAP trajectory of each energy beam are adaptively selected based on sensors' feedback measurements which lead to optimal upward movement of the air mass due to energy absorption. Combinations of energy beam characteristics with guidance law (AAPs trajectories) options are simulated to obtain an optimum combination. One such criterion to judge the optimum combination can be a criterion which determines the rate of climb at which this air mass will rise to higher altitudes.

[W3/2]

Some SEs are located on earth's surface for this purpose. Other SEs may be in the air or move with the rising air column. These ground based or sea based SEs are used to heat the air column from below.

V.8.4 [W4/#]: Features Relating to Modifying an Air Mass's Horizontal Velocity Component

One major feature [W4/0] is covered in this entire section. Two additional features [W4/1] and [W4/2] are covered in separate paragraphs.

The best way to transport an air mass of specific composition and thermodynamic properties is to allow it to be carried by wind currents that exist in nature at that time. However, natural wind does not always satisfy weather modification requirements. The following techniques to transport an air mass are presented.

[W4/1]

This technique includes creating a pressure differential on either horizontal side of the air mass to be moved. This technique is similar to the one discussed in section [W1/5], except in this case in air mass size and movement are very localized. This technique requires unequal distribution of energy into two adjacent sides of the air mass to be moved. This is done by unequal concentrations of energies, AAPs, or energy beams' dwell times on either side of an air mass.

[W4/2]

This case involves concentrating AAPs trajectories along a desired direction of air mass motion. FIG. **6** illustrates this concept. Consider the case where an air mass has no initial lateral motion component. The objective is to move the air mass near **6***x* to points near **6***y*. First, define this desired direction by an arrow **6***v* relative to earth **6***e*, which may be inclined at an angle with respect to the local gravity vector. Next, imagine a large tube **6***i* enclosing the air mass to be moved and with the desired directional arrow **6***v* along tube's centerline **6***x*-**6***y*. Next, define a sequence of AAPs (**6***a*', **6***b*', **6***c*', **6***d*', . . .), such that the maximum energy absorption of each energy beam will be made to concentrate near the centerline of the tube (at **6***a*, **6***b*, **6***c*, **6***d*, etc., respectively). All these AAPs are estimated by GCSs and transmitted as assignments to the SEs **6***s* on orbits **6***h* or **6***g*. Note: each **6***s* point in FIG. **6** can represent many SEs. These SEs will first (at t=ta) concentrate all their energy beams at AAP **6***a*' near point **6***x*. This large input of energy will cause the air mass near **6***a* to absorb the energy and expand rapidly in all directions. Assume a constant temperature contour (resulting from absorption of these energy beams) can be represented by a tear shaped figure labeled **6***k*. Next (at t=tb), let these SEs shift their common aim point to an adjacent AAP at **6***b*'. This will result in rapid heating of another air mass centered at **6***b*. This will cause another tear shaped constant temperature contour labeled by **6***l*. Continuing this process is like producing a sequence of small "explosions" along the imaginary tube's centerline. The size of each "explosion" is controlled by the number of SEs (or energy beams) shearing the same AAP at a given time. The speed of the induced air mass velocity component along **6***v* also depends on the spacing and rate of change of these AAPs or energy beams dwell times along their AAP trajectories. When energy focus reaches the end of the tube **6***y*, this process can be repeated again starting near **6***x*.

This concept can be explained in another way by considering the rows of small circles in FIG. **6**. Each small circle represents the expansion due to absorption of concentrated energy beams at a given location **6***a*, **6***b*, **6***c*, . . . etc. The small arrows radiating out from the center of each small circle in FIG. **6** represents the direction of the local air velocity. As the energy center moves from **6***a* to **6***b* to **6***c* . . . , the velocity components between energy centers cancel each other out. What is left over is only the velocity component in the desired direction **6***v*. When the energy beams are removed, the surrounding cooler air will rush back into the imaginary tube **6***i* and push the air in the **6***v* direction. This technique will always introduce an air mass velocity component in the local vertical direction. This vertical velocity component may not be a bad feature in many applications, as when moving a humid air mass inland and allowing it to condense to rain. The horizontal velocity component will move the air mass inland. The vertical velocity component will allow the air mass to move to a higher and colder altitude. This will delay the air mass's condensation into rain until it has reached its desired location. If the air mass has an initial lateral velocity component **3***i*, then the final air mass velocity vector **3***k* is the vector sum of its velocity vector caused by its initial velocity **3***i* plus the SWMS induced velocity vector **3***j*.

The best way to use this technique is to first create a rising column of air near **6***y* using techniques described in section [W3/0]. Next, simultaneously applying the teachings at [W3/0] and [W4/2] will cause the air mass inside the imaginary tube **6***i* to feed to the rising air column created by [W3/0].

[A#/#]: Features Relating to Satellite Weather Modification System (SWMS) Applications

There are five major features [A1/0]–[A3/01], [A6/0] respectively covered in separate sections. There are a total of 16 (=0+0+3+0+8+5) additional features respectively covered in separate paragraphs. These additional features are:

[A3/1], . . . , [A3/3],
[A5/1], . . . , [A5/8],
[A6/1], . . . , [a6/5].

Some SWMS applications are designed to employ various combinations of the basic weather modification techniques described in section V.8. Other applications are related to energy supply and solar powered airplanes (FIG. **4**). All applications are designed to offer unique and useful services. All applications are selected to have the potential to compete successfully in the real business world. Some of these SWMS applications are illustrated in FIG. **7** and described below.

V.9.1 [A#/#]: Features Relating to Increasing Precipitation

One major feature [A1/0] is covered in this section.

[A1/0]

High humidity air masses are created by application of techniques specified in sections [W1/6] AND [W2/2] over wet surfaces **7***x* to cause cloud formation. Cloud temperatures are maintained by techniques covered in sections [F/0], [W3/0] and [W4/2]. The air mass is allowed to move to a desired location by natural wind current or by techniques specified in sections [W3/0], [W4/1] and [W4/2]. The cloud is left to cool down by a natural cooling process. This produces precipitation over the desired area. This technique can be applied to many dry regions of the world. Potential application regions are in southern California **7***y*, northern Ethiopia **7***z*, etc.

V.9.2 Omitted

V.9.3 [A3/#1]: Features Relating to Creating Whirlwinds

One major feature [A3/0] is covered in this section. Three additional features [A3/1], . . . , [A3/3] are covered in separate paragraphs.

[A3/1]

Techniques specified in section [W3/0] or [W4/0] can be used to move two air masses in opposite directions (or in the same direction but at different speeds). These air masses are offset by a small distance. This process will create a whirlwind or a small tornado.

[A3/2]

Alternatively, the movement of one of the air masses can be due to natural causes.

[A3/3]

Whirlwind or tornado strength can be increased (decreased) by combining and colliding many smaller whirlwinds rotating in the same direction 7*j* (opposite directions 7*i*), thus creating a stronger (weaker) whirlwind.

V.9.4 [A4/#]: Features Relating to Warming Land, Sea and Air

One independent feature [A4/0] is covered in this section.

[A4/0]

Techniques described in sections [W1/4] and [W2/1] can be used to warm a land mass directly or to warm the air mass immediately above the land mass. This is done by distributing the AAPs over the region of interest. That is, the SEs are commanded to discharge solar rays or infrared radiations at these AAPs for specified durations until the desired average temperature rise has been achieved.

A few degrees rise in temperature, for a few hours during extremely cold winter nights, can save many millions of dollars worth of fruit crop annually in Florida 7*r*, California 7*r* and many farms at high altitude. This same technique can be used to heat large greenhouse farms in the U.S. Northeast 7*w*, where the large demand for fresh farm produce now necessitates importation from great distances, using large amounts of fossil fuels.

V.9.5 [A5/#]: Features Relating to Solar Powered Satellite Airplane (SPSA) Designs and Applications

One independent feature [A5/0] is covered in this section. Eight additional features [A5/1], . . . , [A5/8] are covered in separate paragraphs.

[A5/1]

A Solar Powered Satellite Airplane (SPSA) is disclosed that utilizes the energy beams discharged or reflected from Satellite Engines (SEs) to power its flight at high altitudes. As an illustration of this concept, only solar rays reflected from SEs are considered here. However, a similar system can be designed to receive laser beams or microwave energy beam from SEs. The word SATELLITE in SPSA implies this airplane's dependence on Satellite Engines (SEs) and its ability to fly indefinitely in circles above the clouds and over a fixed land area. This is illustrated by racetrack shaped dash circles 7*t* in FIG. 7.

[A5/2]

FIG. 4 shows a typical SPSA design. This comprises a large hinged solar energy collecting disk 4*d* supported by two lift producing wings 4*w* located on either side of the disk. Arrays of solar cells 4*s* arranged on the disk 4*d* transform solar energy into electrical current. This electricity is used to power two electrical motors 4*m*, located on each wing, which drive two propellers 4*p*.

[A5/3]

Continuously adjusting SPSA headings and adjusting the disk tilt angles 4*c* (disk 4*d* roll angle about SPSA fuselage axis) will ensure that the solar rays reflected from Satellite Engines (SEs) 7*c* are as perpendicular to the disk surface 4*s* as possible at all times. This ensures maximum energy reception by the disk.

[A5/4]

The energy collecting disk 4*d* is designed to produce no lifting force during normal high altitude level flight.

[A5/5]

This zero lift condition is achieved by the combination of three design characteristics: First, the disk 4*d* has a plane of symmetry with respect to both of disk surfaces 4*s*. Therefore, when the angle of attack is zero, this disk will only produce a drag force. Second, during normal high altitude operations at any disk tilt angle 4*c*, this disk 4*d* will be maintained at zero angle of attack by its wings 4*w*. Third, independent adjustable slots 4*j* are arranged over the disk surface. Whenever pressure sensors 4*i* detect a pressure differential developing across the disk, these slots 4*j* will open automatically to let air flow from the high pressure side to the low pressure side of the disk, thus killing the lift force on the disk. Pressure sensors 4*i* are also distributed over the entire disk 4*d* and on both sides of the disk.

[A5/6]

During take-off and landing runs, SPSA's energy collecting disk 4*d* is rotated to be parallel to its two wings as illustrated in FIG. 4. These wings will control and maintain the disk at a non-zero angle of attack position. At the same time pressure sensors 4*i* are turned off. This causes all slots 4*j* to lock in closed positions. This scheme can be used to create lift force on disk 4*d* during selected mission phases, such as during take-off and landing runs.

[A5/7]

Ground Control Stations (GCS) 1*d* are used to estimate which Satellite Engines (SEs) are available to supply power 7*c* to which Solar Powered Satellite Airplane (SPSA) 7*p*. GCSs are also used to estimate the direction and intensity of each reflected solar ray 7*c* as functions of time. GCSs must focus a group of AAPs at the center of each SPSA disk 4*d*. Also, GCSs will estimate the trajectories of these AAPs' common focal points. The trajectory of one subgroup of AAPs' common focal point will become the flight path 7*t* of one SPSA. GCSs will also estimate the disk tilt angle 4*c* of each SPSA. This information is uplinked to each SPSA. Software on-board the SPSA will enable it to fly a flight path 7*t* specified by its AAPs' common trajectory 7*t*. At the same time the SPSA will tilt its disk 4*d* following the tilt angle 4*c* time-line specified by GCSs. Periodically, SPSA will shift to a different set of SEs, a different AAPs' trajectory and a different disk tilt angle time-line as commanded by a GCS.

[A5/8]

FIG. 4 illustrates basic design features of an SPSA. A top view and a front view show an SPSA in a take-off or landing configuration. Also, a cross-section view shows the disk tilting mechanism. Important SPSA features shown in FIG. 4 are identified below.

4b	Electric motor, gear box and lock mechanism.
4c	Solar energy collecting disk 4d tilt angle. Tilt angle is the same as roll angle about the SPSA fuselage axis.
4d	Solar energy collecting disk.
4e	Landing gears.
4f	Fuselage.
4g	Gear to tilt and lock disk at desired angle.
4i	Pressure sensors located near slots 4j and distributed over both sides of the disk.
4j	Slots distributed over the disk. Slots are normally closed. Slots will automatically open to balance any local pressure difference developed across the disk.
4k	Roller rod bearings distributed over the fuselage circumference. They provide smooth tilting (roll) of the disk about SPSA fuselage axis.
4l	Payloads.
4m	Electric motor and gear mechanism used to drive the propellers.
4p	Propellers.
4q	Gear teeth attached to the disk 4d and arranged in a circle around the outer circumference of the SPSA fuselage 4f. These gear teeth are used to transmit torque which drives disk 4d tilt angles 4c.
4r	Rudders.
4s	Solar cell array cover the entire (one or both sides) disk 4d surface.
4t	Vertical tails.
4v	direction of SPSA forward motion.
4w	Wings

V.9.6 [A6/#]: Features Relating to Business Ventures Based on SWMS & SWMS/SPSA Services

One major feature [A6/0] is covered in this section. Five additional features [A6/1], . . . , [A6/5] are covered in separate paragraphs.

[A6/1]

Applications of Satellite Weather Modification System (SWMS) can be organized into unique global business ventures. There are strong demands for the types of services offered by SWMS. The services SWMS offers to the world are unique and non-existing at the present time. Both SWMS manufacture and services will generate new jobs and enhance the earth's environment at the same time. These features further strengthen the Satellite Weather Modification System (SWMS) potential as a useful invention. Therefore these SWMS business ventures are an important benefit of this invention.

[A6/2]

Stock can be issued and offered to the public to raise the necessary funds to finance these ventures. Raised funds will then be used to build prototypes of SWMS subsystems: Satellite Engines (SEs), Ground Control Stations (GCSs), Remote Sensing Devices (RSDs) and Solar Powered Satellite Airplanes (SPSAs).

[A6/3]

Initially a few basic services will be offered to potential customers who can afford to pay for SWMS services. In each orbit each SE can sequentially service many locations below its orbit. Therefore, by simply changing some computer codes these SEs can also service some non-paying customers as a public service.

[A6/4]

Some of the initial services to be offered by SWMS are illustrated in FIG. **7**. Satellite Engines (SEs) distributed in the orbits (**7***h* and **7***g*) are too small to be shown in FIG. **7**. Solar rays **7***a* from the sun are first reflected **7***b* from the SEs in the lower orbit to SEs in the higher orbit before finally being reflected to earth **7***c*. Some initial services to be offered by SWMS are described below:

1. Techniques described in section [A4/0] are used to warm local regions during extremely cold winter days. Potential paying customers are orange growers in Florida **7***r* and California **7***r*, greenhouse farms **7***w*, etc. Potential nonpaying customers are small farmers at high latitudes, etc.
2. Solar Powered Satellite Airplanes (SPSAs) **7***p* (described in section [A5/0]) are used to carry communication relay instruments or weather monitoring equipment **4***l* over local regions. Potential paying customers are communication companies, news services, etc. Potential non-paying customers are public education services, environmental protection agencies, weather services, local police departments, etc. It is to be noted that when an SPSA flies a closed loop pattern **7***t* (dashed line racetrack patterns in FIG. **7**) above a given land region, it can be used to compete for business with satellites in a geo-synchronized orbit. The advantages of SPSA include that these SPSAs are re-usable and closer to earth and each SPSA carries its own reusable launching system.
3. The technique discussed in section [A1/0] is used to bring rain to dry regions in the world. The Satellite Engines' (SEs') tasks are to warm the ocean surface a few miles off the southern California coast **7***x*. These tasks must also be timed with the local ocean wind direction, so that the high humidity air mass created by SEs over the ocean can be carried inland by the wind. To support the energy beams' direction guidance laws, a set of AAPs are defined and distributed over an ocean surface area **7***x*. The trajectories of these AAPs can follow the local water current, be timed in accordance with local wind directions, and avoid local cloud formations. Potential paying customers are southern California water authority **7***y*, etc. Potential non-paying customers are drought victims of Ethiopia **7***z*, etc.

[A6/5]

FIG. **7** illustrates some SWMS applications. Items in this figure are identified below:

7a	Solar rays from the sun.
7b	Solar rays reflected from SEs in lower orbit 7g to SEs in higher orbit 7h. SEs are too small to be shown in these orbits. Each 7b line in FIG. 7 represents solar rays from thousands of reflector panels belonging to many SEs.
7c	Solar rays reflected from higher orbit 7h to earth or to SPSA 7p. SEs are too small to be shown in this figure. Each 7c line in FIG. 7 represents solar rays reflected from thousands of reflector panels belonging to many SEs. These solar rays can be aimed at many AAPs on earth or they may be focused on a single AAP at the center of an SPSA disk.
7d	Nature generated whirlwinds.
7e	Earth.
7g	Lower orbit of a nonintersecting-noncoplanar orbital pair.
7h	Higher orbit of a nonintersecting-noncoplanar orbital pair.
7i	Directions of air mass motions induced by SWMS. This will create a whirlwind which can be used to cancel an existing whirlwind 7d.
7j	Direction of air mass motions induced by SWMS. This will create a whirlwind which can be used to increase the strength of an existing whirlwind 7d.
7p	Solar Powered Satellite Airplane (SPSA).
7r	Local region to be warmed by solar energy.
7t	SPSA flight pattern defined by GCS as focal point of thousands of reflecting solar rays. In most cases the SPSA flight path is the same as the common AAP trajectory defined by GCS.
7v	Electricity generating stations: direct solar, wind or hydroelectric.
7w	Greenhouse farms.
7x	Ocean or other wet surfaces to be heated by concentrated solar energy beams. This is done to increase the air mass humidity above 7x.
7y	Increased precipitation by SWMS in southern California
7z	Increased precipitation by SWMS in northern Ethiopia.

APPENDIX: ACRONYMS

AAP Adjusted Aim Point.

GCS Ground Control Station (see FIG. **1**).

LOS Line-Of-Sight.

MMW Millimeter Wave.

RSD Remote Sensing Device (see FIG. **1**).

SE Satellite Engine (See FIG. **2**).

SPSA Solar Powered Satellite Airplane (see FIG. **4**).

SWMS Satellite Weather Modification System (see FIG. **1**).

[SWMS/#]	Invention features relating to Satellite Weather Modification System (SWMS) designs (# = 0, 1, . . . ,5).
[SE/#]	Invention features relating to Satellite Engines (SEs) designs (# = 0, 1, 2, . . . ,6).
[OP/#]	Invention features relating to SEs orbital OPerations (# =0, 1, 2, . . . ,10).
[C/#]	Invention features relating to SEs' energy beam

-continued

	direction Controls (# = 0, 1, 2).
[RSD/#]	Invention features relating to Remote Sensing Devices (RSDs) integration into the SWMS (# = 0, 1, 2).
[GCS/#]	Invention features relating to Ground Control Stations (GCSs) (SEs guidance) designs (# = 0, 1, 2, 3).
[F/#]	Invention features relating to Frequency agile energy beams' frequency and intensity selections (# = 0, 1, . . . ,7).
[W#/#]	Invention features relating to SWMS's basic Weather modification techniques. Four major features are: [W1/0], [W2/0], [W3/O] and [W4/0].
[W1/#]	Invention features relating to modifying global weather conditions (# = 0, 1, . . . ,6).
[W2/#]	Invention features relating to modifying local weather conditions (# = 0, 1, 2).
[W3/#]	Invention features relating to moving an air mass in a vertical upward direction (# = 0, 1, 2).
[W4/#]	Invention features relating to modifying air mass velocity horizontal component (# = 0, 1, 2).
[A#/#]	Invention features relating to SWMS applications. Six major features are: [A1/0], [A2/0], [A3/0], [A4/0], [A5/0] and [A6/0].
[A1/#]	Invention features relating to increasing precipitation applications (# = 0).
[A3/#]	Invention features relating to creating whirlwinds (# = 0, 1, 2, 3).
[A4/#]	Invention features relating to warming local regions (# = 0).
[A5/#]	Invention features relating to SPSA design and applications (# = 0, 1, . . . ,8).
[A6/#]	Invention features relating to business ventures based on applications of SWMS and SWMS/SPSA services (# = 0, 1, . . . ,5).

I claim:

1. Apparatus comprising:

a plurality of satellites, each of said satellites being in a planetary orbit, and each satellite being under remote control and capable of reflecting solar radiation;

a plurality of detectors for monitoring a planetary region, each detector generating at least one signal representative of dynamic or thermodynamic conditions in the monitored region; and

at least one control station responsive to the signals from the detectors and exercising remote control over the satellites for directing solar energy reflected from the satellites towards a plurality of distributed aim points to achieve desired dynamic or thermodynamic properties of the monitored region.

2. Apparatus according to claim **1** wherein the monitored region comprises a land mass.

3. Apparatus according to claim **1** wherein the monitored region comprises a water mass.

4. Apparatus according to claim **1** wherein the monitored region comprises an air mass.

5. Apparatus according to claim **1** wherein the control station comprising means for extrapolating current dynamic or thermodynamic atmospheric values to determine corresponding values at a future time.

6. Apparatus according to claim **1** wherein the control station comprises means for specifying desired dynamic or thermodynamic values at desired locations and desired times.

7. A method comprising:

placing each of a plurality of satellites in a planetary orbit, each satellite being capable of reflecting solar radiation;

selecting a planetary region to be monitored;

detecting dynamic or thermodynamic conditions at a plurality of points in the selected region;

determining an increment of solar radiation needed to substantially conform the detected dynamic or thermodynamic conditions to desired values; and

exercising remote control over the satellites so as to direct solar energy reflected from the satellites towards a plurality of distributed aim points to achieve the determined increment.

8. A method according to claim **7** comprising the step of including a land mass in the selected region.

9. A method according to claim **7** comprising the step of including a water mass in the selected region.

10. A method according to claim **7** comprising the step of including an air mass in the selected region.

11. Apparatus comprising:

at least one million satellites, each of said satellites being in a planetary orbit, and each satellite being under remote control and capable of reflecting solar radiation;

a plurality of detectors for monitoring a planetary region, each detector generating at least one signal representative of dynamic or thermodynamic conditions in the monitored region; and

at least one control station responsive to the signals from the detectors and exercising remote control over the satellites for directing solar energy reflected from the satellites back into space and away from the planet to achieve desired dynamic or thermodynamic properties of the monitored region.

12. Apparatus according to claim **11** wherein:

the number of said satellites is at least one billion.

13. Apparatus according to claim **11** wherein the control station comprising means for extrapolating current dynamic or thermodynamic atmospheric values to determine corresponding values at a future time.

14. Apparatus according to claim **11** wherein the control station comprises means for specifying desired dynamic or thermodynamic values at desired locations and desired times.

15. A method comprising:

placing each of at least one million satellites in a planetary orbit, each satellite being capable of reflecting solar radiation;

selecting a planetary region to be monitored;

detecting dynamic or thermodynamic conditions at a plurality of points in the selected region;

determining a decrement of solar radiation needed to substantially conform the detected dynamic or thermodynamic conditions to desired values; and

exercising remote control over the satellites for directing solar energy reflected from the satellites back into space and away from the planet to achieve the determined decrement.

16. A method according to claim **15** wherein:

the number of said satellites is at least one billion.

* * * * *

US006011991A

United States Patent [19]

Mardirossian

[11] **Patent Number: 6,011,991**

[45] **Date of Patent: Jan. 4, 2000**

[54] **COMMUNICATION SYSTEM AND METHOD INCLUDING BRAIN WAVE ANALYSIS AND/ OR USE OF BRAIN ACTIVITY**

[75] Inventor: **Aris Mardirossian**, Germantown, Md.

[73] Assignee: **Technology Patents, LLC**, Derwood, Md.

[21] Appl. No.: **09/206,365**

[22] Filed: **Dec. 7, 1998**

[51] **Int. Cl.**^{7} **A61N 5/00**

[52] **U.S. Cl.** **600/544**; 600/545

[58] **Field of Search** 600/300, 544–545; 128/897–898, 904, 905

[56] **References Cited**

U.S. PATENT DOCUMENTS

5,059,814	10/1991	Mead et al. .	
5,118,606	6/1992	Lynch et al. .	
5,136,687	8/1992	Edelman et al. .	
5,224,203	6/1993	Skeirik .	
5,303,705	4/1994	Nenov .	
5,325,862	7/1994	Lewis et al. .	
5,461,699	10/1995	Arbabi et al. .	
5,522,863	6/1996	Spano et al. .	
5,640,493	6/1997	Skeirik .	
5,715,821	2/1998	Faupel .	
5,719,561	2/1998	Gonzales .	
5,722,418	3/1998	Bro	128/905
5,730,146	3/1998	Itil et al.	600/544
5,736,543	4/1998	Rogers et al. .	
5,737,485	4/1998	Flanagan et al. .	
5,747,492	5/1998	Lynch et al. .	
5,791,342	8/1998	Woodard	600/544
5,816,247	10/1998	Maynard .	

Primary Examiner—Cary O'Connor
Assistant Examiner—Michael Astorino
Attorney, Agent, or Firm—Joseph A. Rhoa

[57] ABSTRACT

A system and method for enabling human beings to communicate by way of their monitored brain activity. The brain activity of an individual is monitored and transmitted to a remote location (e.g. by satellite). At the remote location, the monitored brain activity is compared with pre-recorded normalized brain activity curves, waveforms, or patterns to determine if a match or substantial match is found. If such a match is found, then the computer at the remote location determines that the individual was attempting to communicate the word, phrase, or thought corresponding to the matched stored normalized signal.

8 Claims, 3 Drawing Sheets

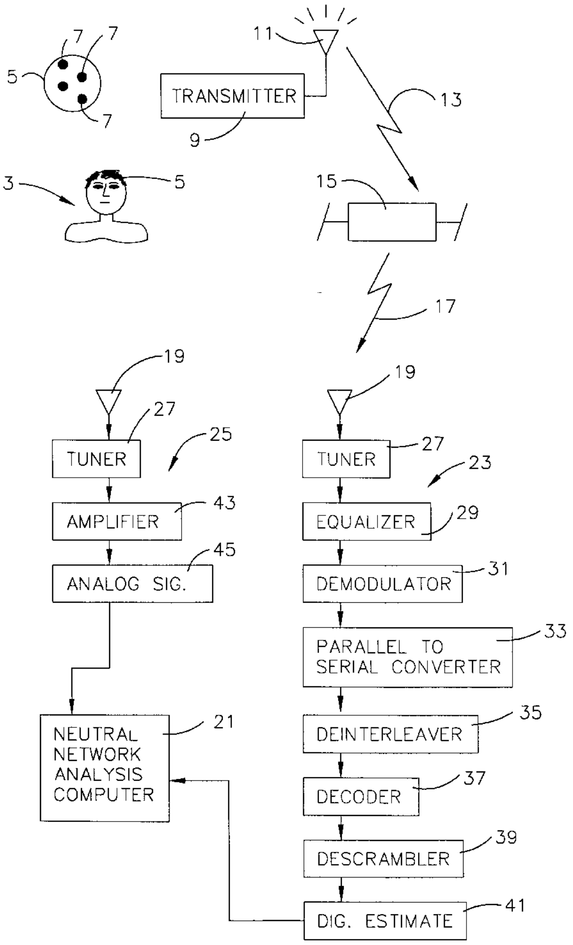
7
7
5
7
11
TRANSMITTER
9
13
3
5
15
17
19
27
25
TUNER
AMPLIFIER
43
45
ANALOG SIG.
NEUTRAL NETWORK ANALYSIS COMPUTER
21
19
TUNER
27
23
EQUALIZER
29
DEMODULATOR
31
PARALLEL TO SERIAL CONVERTER
33
DEINTERLEAVER
35
DECODER
37
DESCRAMBLER
39
DIG. ESTIMATE
41

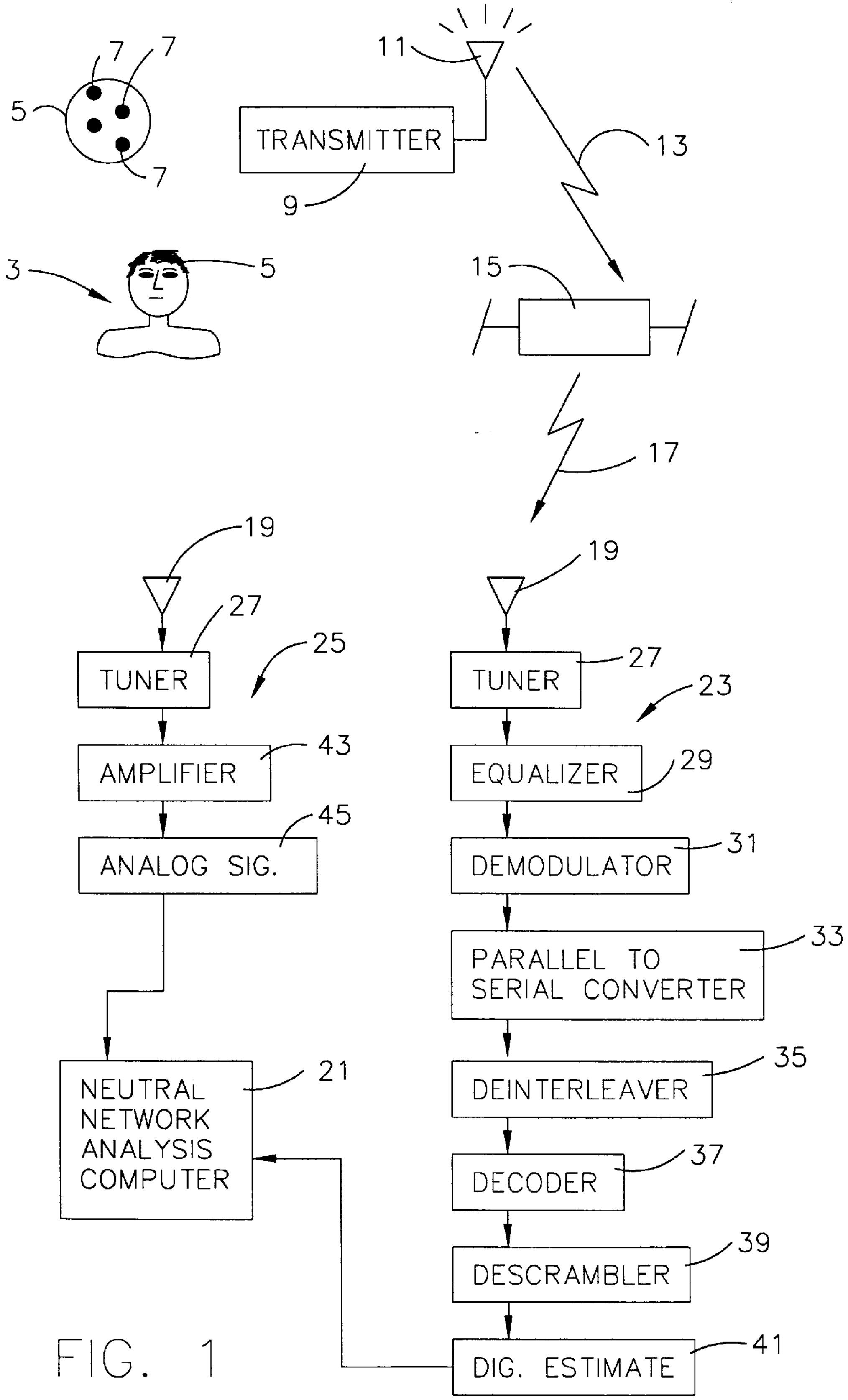
TRANSMITTER
TUNER
TUNER
AMPLIFIER
EQUALIZER
ANALOG SIG.
DEMODULATOR
PARALLEL TO SERIAL CONVERTER
NEUTRAL NETWORK ANALYSIS COMPUTER
DEINTERLEAVER
DECODER
DESCRAMBLER
DIG. ESTIMATE

FIG. 1

FIG. 2

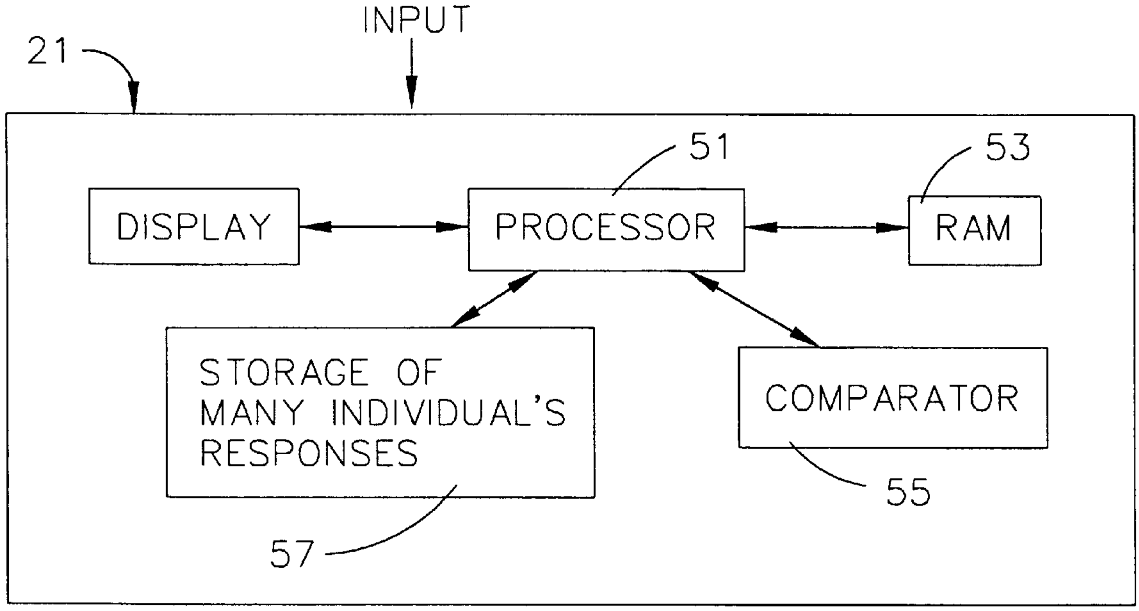
INPUT
21
51
53
DISPLAY
PROCESSOR
RAM
STORAGE OF MANY INDIVIDUAL'S RESPONSES
COMPARATOR
55
57

FIG. 3(a)

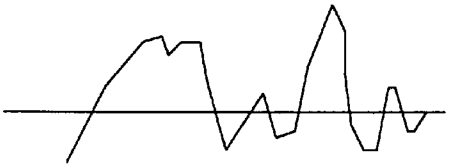

JOE (NO)

FIG. 3(b)

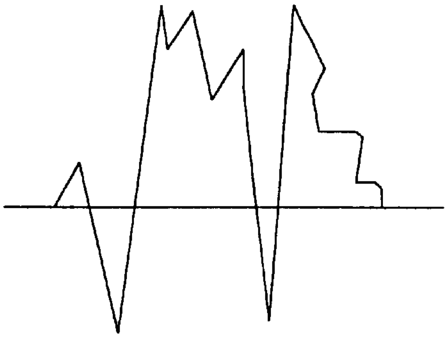

JOE (YES)

FIG. 3(c)

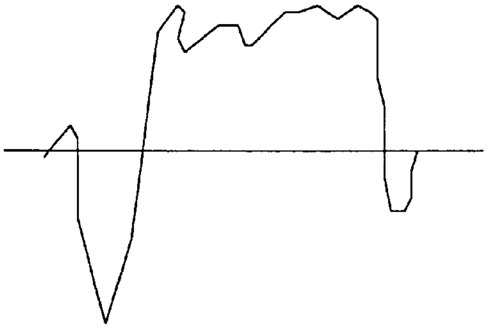

FIG. 3(d)

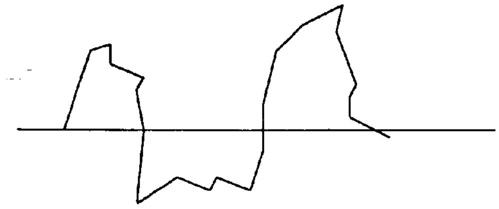

FIG. 3(e)

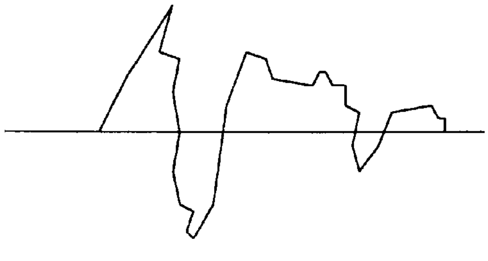

FIG. 3(f)

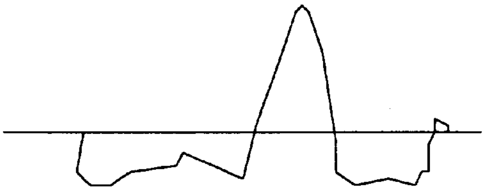

COMMUNICATION SYSTEM AND METHOD INCLUDING BRAIN WAVE ANALYSIS AND/OR USE OF BRAIN ACTIVITY

This invention relates to a system and method for enabling human beings to communicate with one another by monitoring brain activity. In particular, this invention relates to such a system and method where brain activity of a particular individual is monitored and transmitted in a wireless manner (e.g. via satellite) from the location of the individual to a remote location so that the brain activity can be computer analyzed at the remote location thereby enabling the computer and/or individuals at the remote location to determine what the monitored individual was thinking or wishing to communicate.

In certain embodiments this invention relates to the analysis of brain waves or brain activity, and/or to the remote firing of select brain nodes in order to produce a predetermined effect on an individual.

BACKGROUND OF THE INVENTION

It is known to monitor brain activity by way of electroencephalograph (EEG) methods, magnetoencephalograph (MEG) methods, and the like. For example, see U.S. Pat. Nos. 5,816,247 and 5,325,862, the disclosures of which are both hereby incorporated herein by reference. As discussed in the '247 patent, an EEG may be recorded from a number of pairs of scalp electrodes and processed according to known software. Such software and/or hardware acquires both processed and unprocessed EEG data and may record it on a disk. The records may be replayed and statistics of the on-line measures made on suitable sections placed in categories predefined by a user. This may utilize the form of database of statistical measures of brain activity. Unfortunately, neither the '862 nor the '247 patents disclose or suggest any methods by which humans can communicate with one another by way of monitoring brain activity.

U.S. Pat. No. 5,719,561 discloses a communications device and method, the entire disclosure of the '561 patent hereby being incorporated herein by reference. The '561 patent discusses a method and device for vibromechanical tactile communications adaptable for use by individuals to recognize alpha numeric messages in a language or in other symbols known to them. The '561 patent discusses using a series of sequentially firing vibromechanical stimulators vibrating against a suitably tactile sensitive surface of the wearer (e.g. skin) to induce a phenomenon of illustration of linear continuity. Unfortunately, the '561 patent requires the use of burdensome and complex vibromechanical tactile devices, and is not suitable for long distance communication.

It is a purpose of this invention to address any or all of the above-identified problems in the prior art, as well as other problems which will become apparent to the skilled artisan from the following detailed description of this invention.

SUMMARY OF THE INVENTION

Generally speaking, this invention fulfills the above described needs in the art by providing a method of communicating comprising the steps of:

providing a first human being at a first location;

providing a computer at a second location that is remote from the first location;

providing a satellite;

providing at least one sensor (preferably a plurality—e.g. tens, hundreds, or thousands, with each sensor monitoring the firing of one or more brain nodes or synapse type members) on the first human being;

detecting brain activity of the first human being using the at least one sensor, and transmitting the detected brain activity to the satellite as a signal including brain activity information;

the satellite sending a signal including the brain activity information to the second location;

a receiver at the second location receiving the signal from the satellite and forwarding the brain activity information in the signal to the computer;

comparing the received brain activity information of the first human being with normalized or averaged brain activity information relating to the first human being from memory; and

determining whether the first human being was attempting to communicate particular words, phrases or thoughts, based upon the comparing of the received brain activity information to the information from memory.

In certain embodiments, the invention includes the following step: asking the first human being a plurality of questions and recording brain activity of the first human being responsive to the plurality of questions in the process of developing said normalized or averaged brain activity information relating to the first human being stored in the memory. A database in a memory may include, for each of a plurality (e.g. one hundred or thousands) of individuals, a number of prerecorded files each corresponding to a particular thought, attempt to communicate a word, attempt to communicate a phrase or thought, or mental state. Measured brain activity of a given individual may be compared to files from that database of that individual to determine what the individual is attempting to communicate or what type of mental state the individual is in.

In certain embodiments, the plurality of questions are the same question.

In certain embodiments, the plurality of questions are different questions.

In certain embodiments, the invention includes the step of normalizing or averaging recorded brain activity responsive to a given question or set of questions in developing the normalized or averaged brain activity information relating to the first human being.

It is an object of this invention to enable brain activity of a first human being to be monitored, with the activity being transmitted to a remote location so that individuals and/or a computer at the remote location can determine what the first human being was thinking or intending to communicate. In such a manner, human beings can communicate with one another via monitoring of brain activity, and transmission of the same.

It is another object of this invention to communicate monitored brain activity from one location to another in a wireless manner, such as by IR, RF, or satellite.

It is another object of this invention to provide a system capable of identifying particular nodes in an individual's brain, the firings of which affect characteristics such as appetite, hunger, thirst, communication skills (e.g. which nodes are utilized to communicate certain words such as "yes", "no", or phrases such as "I don't know", "I'm not sure", or numbers such as "one", "two", "ten", "one hundred" and the like), thought processes, depression, and the like). When such nodes are identified, they may be specifically monitored by one or more sensors to analyze behavior or communication or words, phrases, or thoughts. In other embodiments, devices mounted to the person (e.g. underneath the scalp) may be energized in a predetermined manner or sequence to remotely cause particular identified brain node(s) to be fired in order to cause a predetermined feeling or reaction in the individual, such as lack of hunger, lack or depression, lack or thirst, lack of aggression, lack of alzheimer's disease effects, or the like.

Brain node firings are the basis of thought and mind processes of individuals. Certain embodiments of this invention enable such brain firings and behavior to be captured by an external device. It is an object of this invention to utilize a normalization or normalizing curve (or waveform or pattern) based upon monitored brain activity to detect or determine thought processes by the monitored individual. In such a manner, individuals can transmit by satellite what they are thinking or intending to think via their monitored brain activity, without the need to talk or write down information.

Each individual has a distinct pattern of brain node firings or brain activity. Each person is believed to be different in this regard. Thus, a separate brain activity file may be stored in a memory for each individual, and analyzed or compared to received brain activity from the monitored individual in order to determine what that individual is thinking or attempting to communicate.

It is an object of this invention to utilize brain monitoring and transmission of monitored brain activity for lie detection and/or human communication.

It is another object of this invention to formulate or build-up a file for each individual based upon patterns recorded in response to that individual answering or responding to numerous predetermined questions with known intended responses. Subsequently, monitored brain activity from that individual may then be compared to information stored corresponding to that individual to determine whether the individual is lying or what the individual is intending to communicate in the monitored brain activity. The higher the level of detail of the file, the higher the level of potential communication by certain embodiments of this invention.

At least one sensor on the scalp or skin in certain embodiments provides signals representative of physiological activity generated in the brain of a monitored individual. A data acquisition device receives the signals representative of the physiological activity generated in the monitored brain, and transforms the signals into a pattern or curve corresponding to the monitored brain activity. This is then transmitted (e.g. by satellite) to a computer located at a remote location, with the monitored brain activity pattern or curve being stored in a memory at the remote location. The computer then causes the received pattern or curve information to be compared with stored brain activity pattern information relating to the monitored individual in order to determine (a) whether the monitored individual is lying in response to a particular question, or (b) what the monitored individual is communicating or attempting to communicate.

Another object of this invention is to utilize normalization curves representative of received brain activity patterns from the monitored individual, and to compare the received normalized data with normalized brain activity pattern or curve data stored in memory relating to that individual. The use of normalization curves in one or both of the individual's file and received brain activity improves reliability, accuracy, and efficiency.

In certain embodiments of this invention, the computer located at the remote location includes a neural network suitably programmed in accordance with known neural network techniques, for the purpose of receiving the monitored brain activity signals, transforming the signals into useful forms, training and testing the neural network to distinguish particular forms and patterns of physiological activity generated in the brain of the monitored individual, and/or comparing the received monitored brain activity information with stored information relating to that individual in order to determine what the individual is attempting to communicate.

This invention further fulfills the above described needs in the art by providing a method of affecting a mental or physiological state of an individual, the method comprising the steps of:

providing at least one firing device capable of being energized on an individual; and

energizing the firing device to cause the firing device to cause a particular or group of brain nodes to be fired in the individual in order to affect the mental or physiological state of the individual.

In certain embodiments, the method including the step of providing the at least one firing device on or under the scalp of the individual in proximity of the brain of the individual.

In certain embodiments, the method including the step of identifying at least one brain node related to the mental or physiological state intended to be affected, targeting the identified brain node, and energizing the firing device or devices to cause the identified node to be fired in order to affect the mental or physiological state of the individual.

In certain embodiments, the method is utilized to cause the individual to be one of less hungry, less thirsty, less anxious, and less depressed.

In certain embodiments, the remote node firing devices are electrically energized and generate electromagnetic waves which cause a plurality of brain nodes to be fired.

This invention will now be described with respect to certain embodiments thereof, along with reference to the accompanying illustrations.

IN THE DRAWINGS

FIG. **1** is a block diagram illustrating the system and method according to a first embodiment of this invention.

FIG. **2** is a block diagram illustrating the neural network inclusive computer of the FIG. **1** embodiment of this invention.

FIGS. **3**(*a*)–**3**(*f*) are exemplary graphs of monitored brain activity of different individuals, with, for example, FIG. **3**(*a*) illustrating monitored brain activity of a particular individual who is attempting to communicate the word "no" and FIG. **3**(*b*) illustrating monitored brain activity of the same individual when that individual is attempting to communicate the word "yes."

DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS OF THIS INVENTION

Referring now more particularly to the accompanying drawings in which like reference numerals indicate like parts throughout the several views.

There are significant individual differences in electrical and magnetic activity in the brain. Brain node or synapse firings are chemically and/or electrically caused and/or related. Some characteristics of brain activity may be relatively stable when measured from day to day. Brain responses to sensory stimulation (e.g. visual, audible, olfactory, gustatory, etc.) as well as higher order cognitive processing (e.g. decision-making or thought/word communication), can be examined in great detail using a variety of recording procedures. A recording of brain electrical activity is called an electroencephalograph (EEG), and a comparable record of magnetic activity is called a magnetoencephalogram (MEG). When human sensory systems are stimulated by a particular event (a given sound or optical effect) or when a human wishes to communicate a particular word or phrase (e.g. the word "yes" or the word "no"), there is a predictable sequence of processing that occurs in the brain. This processing generates an event related potential that can be recorded from the scalp beginning shortly after the onset of the stimulation, and lasting for approximately 0.5–4 seconds after the stimulation. These potentials can be repeatedly generated from individuals given the same stimulus or wishing to communicate the same word or phrase. In certain embodiments of this invention, brain activity may be repeatedly sampled, and response patterns averaged by way of a normalization curve or the like. Comparable recordings of averaged or normalized magnetic activity may be referred to as evoked fields. Neuroelectric and neuormagnetic recordings are subsets of general measurings referred to as bioelectric and biomagnetic measures. These measures refer to recordings which may be made from different types of tissue including neural, muscle, heart, etc. For example, EEG, evoked potentials (EP), MEG, position emission tomography (PET) of glucose, or single photon emission computed tomography (SPECT) may be used to monitor brain activity in different embodiments of this invention.

Event related potentials have been shown to be stable and unique to individuals. See U.S. Pat. No. 5,325,862, which is incorporated herein by reference. Although the actual shape of such potentials varies considerably from individual to individual, there is stability within individuals over time for individual waveforms. Sources of these potentials and variations thereof include individual differences in brain anatomy and differences in the way in which information is processed by each individual. Thus, it is feasible herein to utilize monitored brain waveforms for the purpose of determining whether an individual is lying or not, or what word or phrase a particular individual is attempting to communicate (without the need for writing information down or speaking).

Thus, evoked fields and/or event related potentials can be utilized as classifiers for several purposes. For example, because these potentials and/or fields are relatively unique to individuals, an individual's evoked field or event related potential (or brainprint) can be utilized to determine what thoughts an individual is communicating or attempting to communicate given knowledge of the identification of that individual and previously recorded patterns associated with that individual communicating predetermined words, phrases or thoughts. Because there is a remarkable degree of stability in individual waveforms of a person over time, it is possible to identify changes in individual event related potentials and evoked field patterns which can be utilized to determine when an individual is lying, or impaired in any way.

There are numerous neural networks in the brain, these networks having complex inner connections and non-linear response patterns. Relationships between the latencies and amplitudes of event related potentials and evoked field waveform features have become well understood. In addition, there are many individual variations in waveform morphology. Computing techniques modeled after brain neural functions are known in the art. They are typically referred to as neural network analysis techniques or computers. Neural network analysis computing technology offers a method for finding complex, non-linear relationships in large data sets, even when the nature of the relationships is not known in advance. Neural network technology is implemented sometimes using computer software programs, but may also be hardware implemented. Neural network theory, and detailed descriptions of specific techniques, are available in numerous books and articles set forth in the aforesaid '862 patent, as well as in, for example, any of U.S. Pat. Nos. 5,136,687; 5,059,814; 5,461,699; 5,737,485; 5,224,203; and 5,640,493, the entire disclosures of which are all hereby incorporated herein by reference. Such neural computing systems have a capability to learn features of data sets and classify same into either unknown or predetermined categories. A variety of neural network techniques may be utilized to classify event related potentials, evoke fields, or any other type of pattern corresponding to monitored brain behavior. In most neural networks, input values are adjusted through a series of layers by a series of transforms and weighted so that output categories are correctly predicted. Thus, a neural computing system herein may be utilized to receive monitored brain activity and based upon predetermined stored and/or learned information, determine based upon the received information what word, phrase, or thoughts the monitored individual is attempting to communicate. In such a manner, the monitoring of brain activity may be utilized to allow individuals to communicate from one location to another, with the neural computer or any other type of computer analyzing the monitored brain information (e.g. via comparison with previously recorded brain activity of that person) and outputting information indicative of the word, phrase, or thoughts which the monitored individual is attempting to communicate.

FIG. **1** illustrates a particular monitored individual **3** according to an embodiment of this invention. Individual **3** includes a head **5**. As illustrated, sensors **7** may be attached or otherwise disposed adjacent to the scalp or skin of the individual **3**. Sensors **7** detect and monitor brain activity of individual **3**. Sensors **7** can detect event related potentials and/or event related fields (i.e. ERPs or ERFs). Optionally, sensor **7** may be utilized to detect any other type of "brainprint" indicative of brain activity of individual **3**.

The monitored "brainprint" of individual **3** is forwarded to a small transmitter **9** which is preferably embedded in the skin of individual **3**, or in close proximity to individual **3**. Transmitter **9** causes the monitored "brainprint" information detected by sensor **7** to be transmitted by way of antenna **11** as wireless signals **13**. In certain embodiments of this invention, signals **13** propagate through atmospheric free space in the form of uplink satellite signals toward satellite **15**. Satellite **15** receives signals **13** and then redirects those signals back toward Earth as signals **17** which include information therein (analog or digital) indicative of the monitored brain activity of individual **3**. Signals **17** are received by antenna **19**. Antenna **19** and individual **3** are both preferably located on Earth at different locations. In certain embodiments, antenna **19** is located at a location remote from individual **3**. For example, individual **3**, sensor **7**, transmitter **9**, and antenna **11** may all be located in Europe while receiving antenna **19** and computer **21** may be located in the United States.

FIG. **1** illustrates both a digital embodiment **23** and an analog embodiment **25** of receiving systems. Either may be utilized. When signals **17** include digital information, they are received by antenna **19** and forwarded to tuner **27**. The signals are processed through equalizer **29**, demodulator **31**, parallel to serial converter **33**, deinterleaver **35**, decoder **37**, descrambler **39**, digital estimator **41**, and finally to neural network analysis or other type of computer **21**. Monitored firings of brain nodes may be broken down into digital form (e.g. the firing of a node is equivalent to a "**1**" and nonfiring to a "0"). These 1s and 0s, which are digital, may be modulated onto a carrier and then transmitted to the satellite so that the monitored brain activity in the signal is in digital form. When the computer receives this monitored signal from the satellite, the demodulating system analyzes the received digital information (e.g. 1s and 0s) which is indicative of the firing of select brain nodes of the monitored individual. These digital signals may be transformed, in certain embodiments, into analog form similar to the illustrations of FIG. **3**, or alternatively may be kept in digital form and compared with prestored digital signals to determine what the monitored person was intending to communicate.

In analog embodiments, signal **17** is received by antenna **19**, and the information forwarded to tuner **27**, amplifier **43**, and the analog information inclusive signal **45** is forwarded to computer **21**. In analog embodiments, the signals received by the monitoring sensors are in the form of waves similar to those shown in FIG. **3** herein.

FIG. **2** is a block diagram of computer **21** in certain embodiments of this invention. The input thereto may be in either analog or digital form. Computer **21** includes processor **51**, RAM **53**, comparing device **55**, and memory **57** for storing a plurality of files or patterns of measured brain activity or responses of particular individuals. For example, memory or storage **57** may include one file for individual X which includes hundreds of monitored brain activity (e.g. ERPs or ERFs) that were measured when individual X was attempting to communicate the word "no." Another file in storage or memory **57** may have stored therein hundreds or thousands of monitored brain patterns or activities of individual X when individual X was attempting to communicate the word "yes." Storage or memory **57** may also include similar files for individual X including patterns which were measured when the individual was attempting to communicate different words or phrases. Each of these files may have a normalized curve, waveform or pattern formed or developed for each word, phrase, or thought for each individual, based upon all of the recorded patterns or curves for that word, phrase, or thought (e.g. 1,000 such recordings for the phrase "help me," and 1,000 for the phrase "I've been caught." Memory **57** further includes in certain embodiments many different files for many different individuals, all including files for each individual's past communicating of particular words, phrases, or thoughts.

Normalized curves or patterns corresponding to each individual's attempt to communicate a particular word, thought, or a pattern may be stored in storage or memory **57**. Thus, for individual X attempting to communicate the word "no", a normalized curve, waveform, or a pattern may be stored in memory **57** indicative or such a communication. Likewise, a normalized curve, waveform, or pattern may be stored in memory **57** for individual Y attempting to communicate the word "no" and another distinct normalized curve, waveform, or pattern may be stored in memory **57** for individual Y attempting to communicate the word "yes." In a similar manner, a normalized curve, waveform or pattern may be stored in memory **57** for individual Y attempting to communicate the phrase "I don't know" or "I'm not sure." Thus, if one hundred different human beings have files in memory **57** corresponding to each of these individual's attempt to communicate the phrase "I don't know", then one hundred different normalized waveforms, curves or patterns would be stored in memory **57**, each corresponding to a particular individual. A normalized curve, waveform, or pattern may be developed by repeatedly asking an individual tens, hundreds, or thousands of times a particular question or group of questions which evoke a known response such as "I don't know," or "no" or "yes." Each time the response is made, a curve, waveform, or a pattern is recorded. After tens, hundreds, or even thousands of these patterns have been recorded, a normalized curve, waveform or pattern is formed based upon same so as to be indicative of that particular individual's attempt to communicate the phrase. In a similar manner, if it is desired to store normalized curves, waveforms, or patterns for ten different words, phrases or thoughts communicated by a particular individual, then a normalized curve, waveform, or pattern is developed for each of the different ten items so that ten different normalized curves, waveforms or patterns are stored in memory **57** for that individual and classified accordingly. Thus, when computer **21** receives signals indicating brain activity from that monitored individual, the received signals are compared by device **55** to the ten different normalized signals in memory **57** in order to determine what the individual is attempting to communicate.

FIGS. **3**(*a*)–**3**(*f*) illustrate different normalized curves which may be stored in memory **57**. FIG. **3**(*a*) shows a normalized curve indicative of individual "Joe" attempting to communicate the word "no." FIG. **3**(*b*) shows a normalized curve indicative of individual "Joe" attempting to communicate the word "yes." FIG. **3**(*c*) shows a normalized curve indicative of another individual "Steve" attempting to communicate the word "no", while FIG. **3**(*d*) shows a normalized curve indicative of individual "Steve" attempting to communicate the word "yes." Finally, FIG. **3**(*e*) shows a normalized curve indicative of still another individual "Anita" attempting to communicate the word "no", while FIG. **3**(*f*) illustrates a normalized curve of "Anita" attempting to communicate the word "yes."

Thus, if computer **21** receives a signal including monitored brain information identified as being from individual "Steve", then computer **21** causes the received signal to be compared by device **55** with the normalized curves or waveforms shown in FIGS. **3**(*c*) and **3**(*d*) and all others normalized stored signals of "Steve." If a match or a close match is found between the received monitored signal and the normalized curve of FIG. **3**(*c*), then the computer determines that "Steve" was attempting to communicate the word "no." Meanwhile, if no match is found with the normalized curve of FIG. **3**(*c*), but a match or a substantial match is found with regard to the normalized curve or waveform of FIG. **3**(*d*), then the computer determines that "Steve" was attempting to communicate the word "yes." If no match is found between the received "Steve" signal and any normalized curve or waveform of either FIG. **3**(*c*) or FIG. **3**(*d*), or with any other normalized curve stored in memory **57** corresponding to "Steve", then the computer determines that it is unclear what "Steve" was attempting to communicate.

Thus, different embodiments of this invention may be utilized to help individuals communicate with one another without having to send faxes, make telephone calls, speak, or the like. For instance, military personnel located in the Middle East or Europe can communicate with superiors in the Pentagon, simply by use of monitored brain activity being transmitted by satellite to the Pentagon. Alternatively, a special operations individual (e.g. a spy) located in Europe could be asked a question by way of a telephone call, fax, or the like, and that individual can respond to that question simply by thinking the answer so that that individual's monitored brain activity which is transmitted back to the United States can be analyzed to determine the individual's response. In further embodiments of this invention, twoway human communication is possible, provided that human beings at both locations have equipment capable of analyzing and monitoring received monitored brain activity. In such a manner, individuals at two remote locations may communicate with one another without either individual having to speak a word, write anything down, or the like.

In other embodiments, devices mounted to the person (e.g. underneath the scalp) may be energized in a predetermined manner or sequence to remotely cause particular identified brain node(s) to be fired in order to cause a predetermined feeling or reaction in the individual, such as lack of hunger, lack or depression, lack or thirst, lack of aggression, lack of alzheimer's disease effects, or the like. In an example of such an embodiment, the sensors may be replaced with remote firing devices. The computer may cause satellite signals to be sent to a receiver on or proximate an individual, which receiver forwards instructions to the remote firing devices that are mounted, e.g. under the scalp of the individual, in order to selectively cause same to fire or be energized. Such energizing of the device(s) under or near the scalp in a predetermined manner tend to cause identified brain nodes to fire a predetermined number of times. This is useful, for example, in the following scenarios. For example, the system can be used to identify which brain node(s) in a particular individual are typically fired causing that individual to not be hungry. If that individual has an eating disorder or problems with obesity, then the firing devices can be remotely energized thereby causing the identified brain node(s) to be fired at predetermined or random times in order to cause the individual to not be hungry (even if the individual has not eaten for several hours or several days).

In a similar manner, brain nodes which cause an individual to be jovial or not depressed can be identified, and caused to be remotely fired by the computer and firing devices **7** mounted under the scalp in a predetermined manner or sequence(s) in order to minimize or prevent depression of the individual. This may eliminate the need for drugs such a Prozac. Alternatively, such drug(s) may be administered after such remote node firings, and the nodes at issue thereafter being monitored as discussed above and a biofeedback being performed to determine the effectiveness of the drug(s) or alternatively to enable a system to be utilized combining drug treatment with remote node firings to more effectively prevent or minimize depression of the individual. Thus, the biofeedback may enable the identified nodes to be fired by the firing devices and/or drug treatment at the proper level to most efficiently treat the disease, illness or state. This invention, including remote firings and/or monitoring, is not limited to these examples, and its potential uses are almost endless. Brain node firings can be remotely controlled in a predetermined manner or sequence (s) (even random or sequential) to reduce, minimize, or eliminate undesirable behavior or mental characteristics. This may eliminate or reduce the need for burdensome drug treatments and the like. Brain node firings of a normal person, or of a particular person in a given mental or physical state, may be monitored and the brain activity stored and analyzed in the computer **21** memory. This stored brain activity may then be caused by remotely causing the firing devices to cause particular brain node(s) to be fired at given times or intervals in a predetermined manner or sequence. The computer may be programmed to instruct the brain sensors and/or firing devices to identify which brain nodes are responsible for which types of physical or mental behavior, and then the computer transmits firing instructions to the firing devices for those nodes to cause them to be fired in a predetermined manner to effect such physical or mental behavior.

Once given the above disclosure, many other features, modifications, and improvements will become apparent to the skilled artisan. Such other features, modifications, and improvements are, therefore, considered to be a part of this invention, the scope of which is to be determined by the following claims.

I claim:

1. A method of communicating comprising the steps of:

providing a first human being at a first location;

providing a computer at a second location that is remote from the first location;

providing a satellite;

providing at least one sensor on the first human being;

detecting brain activity of the first human being using the at least one sensor, and transmitting the detected brain activity to the satellite as a signal including brain activity information;

the satellite sending a signal including the brain activity information to the second location;

a receiver at the second location receiving the signal from the satellite and forwarding the brain activity information in the signal to the computer;

comparing the received brain activity information of the first human being with normalized or averaged brain activity information relating to the first human being from memory; and

determining whether the first human being was attempting to communicate particular words, phrases or thoughts, based upon the comparing of the received brain activity information to the information from memory.

2. The method of claim **1**, further including the following steps:

asking the first human being a plurality of questions and recording brain activity of the first human being responsive to the plurality of questions in the process of developing said normalized or averaged brain activity information relating to the first human being stored in the memory.

3. The method of claim **2**, wherein the plurality of questions are the same question.

4. The method of claim **2**, wherein the plurality of questions are different questions.

5. The method of claim **2**, further comprising the step of normalizing or averaging recorded brain activity responsive to a given question or set of questions in developing the normalized or averaged brain activity information relating to the first human being.

6. A method of communicating words from a first location to a second location, the method comprising the steps of:

providing a first human being at the first location;

providing a computer at the second location that is remote from the first location;

providing at least one sensor on the first human being;

detecting brain activity of the first human being using the at least one sensor wherein the brain activity is indicative of words to be communicated by the first human being, and forwarding the detected brain activity indicative of words to be communicated to the computer at the second location;

comparing the received brain activity of the first human being indicative of words to be communicated with normalized or averaged brain activity information relating to the first human being from memory; and

determining words being communicated by the first human being based upon the comparing of the received brain activity information to the information from memory.

7. The method of claim **6**, further including asking the first human being a plurality of questions and recording brain activity of the first human being responsive to the plurality of questions in a process of developing the normalized or averaged brain activity information relating to the first human being.

8. The method of claim **6**, further comprising the step of the computer outputting words determined in said determining step to a second human being so that words thought or stated by the first human being are communicated by the first human being to the second human being via the computer.

* * * * *

US006017302A

United States Patent [19]

Loos

[11] **Patent Number: 6,017,302**

[45] **Date of Patent: Jan. 25, 2000**

[54] **SUBLIMINAL ACOUSTIC MANIPULATION OF NERVOUS SYSTEMS**

[76] Inventor: **Hendricus G. Loos**, 3019 Cresta Wy., Laguna Beach, Calif. 92651

[21] Appl. No.: **08/961,907**

[22] Filed: **Oct. 31, 1997**

[51] **Int. Cl.^{7}** **A61B 5/00**

[52] **U.S. Cl.** **600/28**

[58] **Field of Search** 600/26–28; 128/897, 128/898

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,124,022	11/1978	Gross .	
4,335,710	6/1982	Williamson	600/28
4,573,449	3/1986	Warnke .	
5,076,281	12/1991	Gavish	600/28 X
5,123,899	6/1992	Gall	600/28
5,309,411	5/1994	Huang et al.	367/140
5,733,240	3/1998	De Visser	600/9

Primary Examiner—Samuel Gilbert

[57] **ABSTRACT**

In human subjects, sensory resonances can be excited by subliminal atmospheric acoustic pulses that are tuned to the resonance frequency. The ½ Hz sensory resonance affects the autonomic nervous system and may cause relaxation, drowsiness, or sexual excitement, depending on the precise acoustic frequency near ½ Hz used. The effects of the 2.5 Hz resonance include slowing of certain cortical processes, sleepiness, and disorientation. For these effects to occur, the acoustic intensity must lie in a certain deeply subliminal range. Suitable apparatus consists of a portable battery-powered source of weak subaudio acoustic radiation. The method and apparatus can be used by the general public as an aid to relaxation, sleep, or sexual arousal, and clinically for the control and perhaps treatment of insomnia, tremors, epileptic seizures, and anxiety disorders. There is further application as a nonlethal weapon that can be used in law enforcement standoff situations, for causing drowsiness and disorientation in targeted subjects. It is then preferable to use venting acoustic monopoles in the form of a device that inhales and exhales air with subaudio frequency.

17 Claims, 5 Drawing Sheets

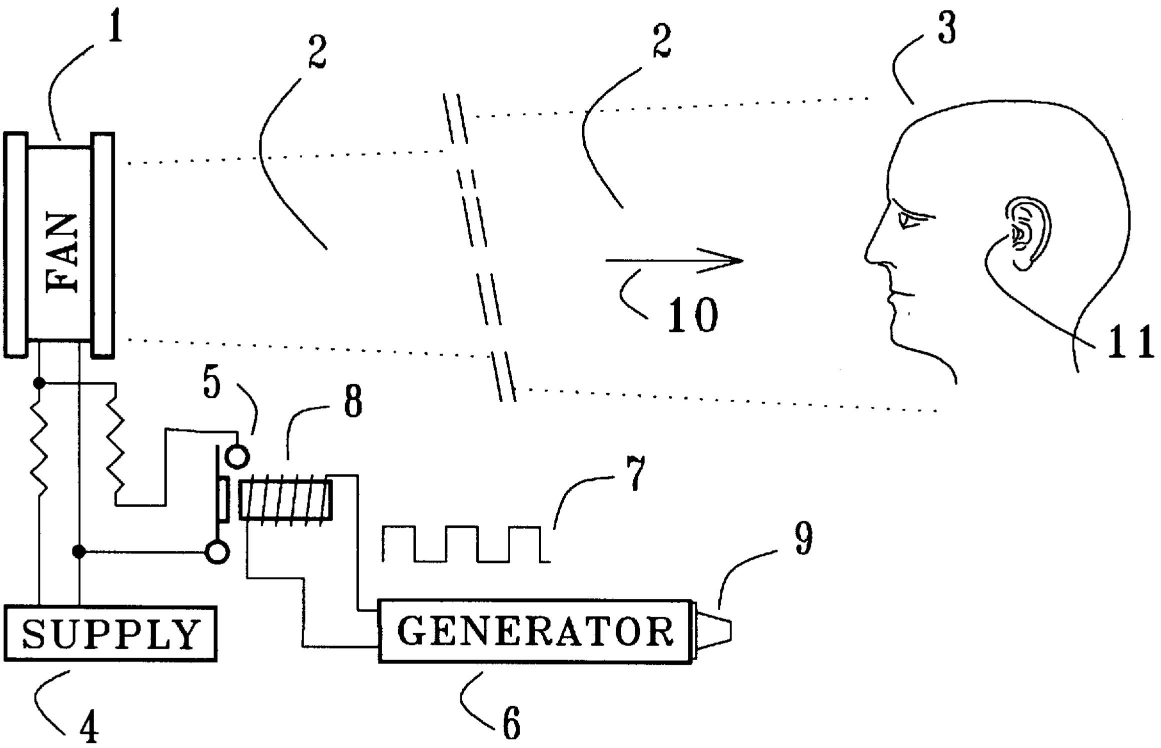

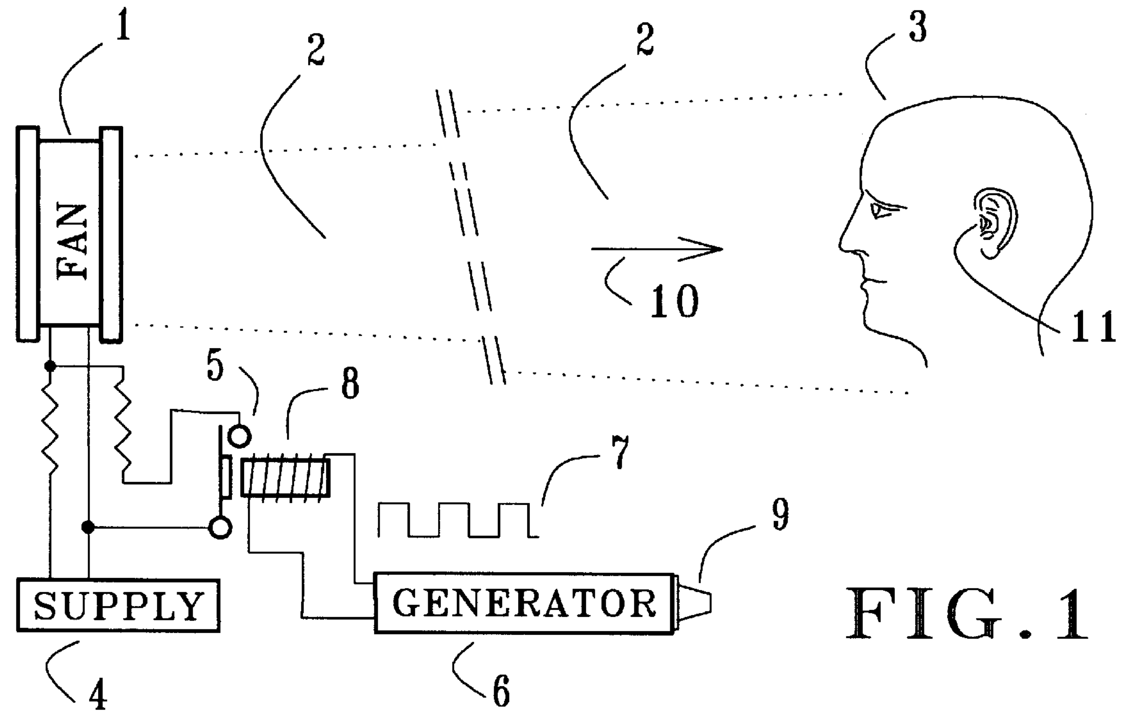
FAN
SUPPLY
GENERATOR

FIG. 1

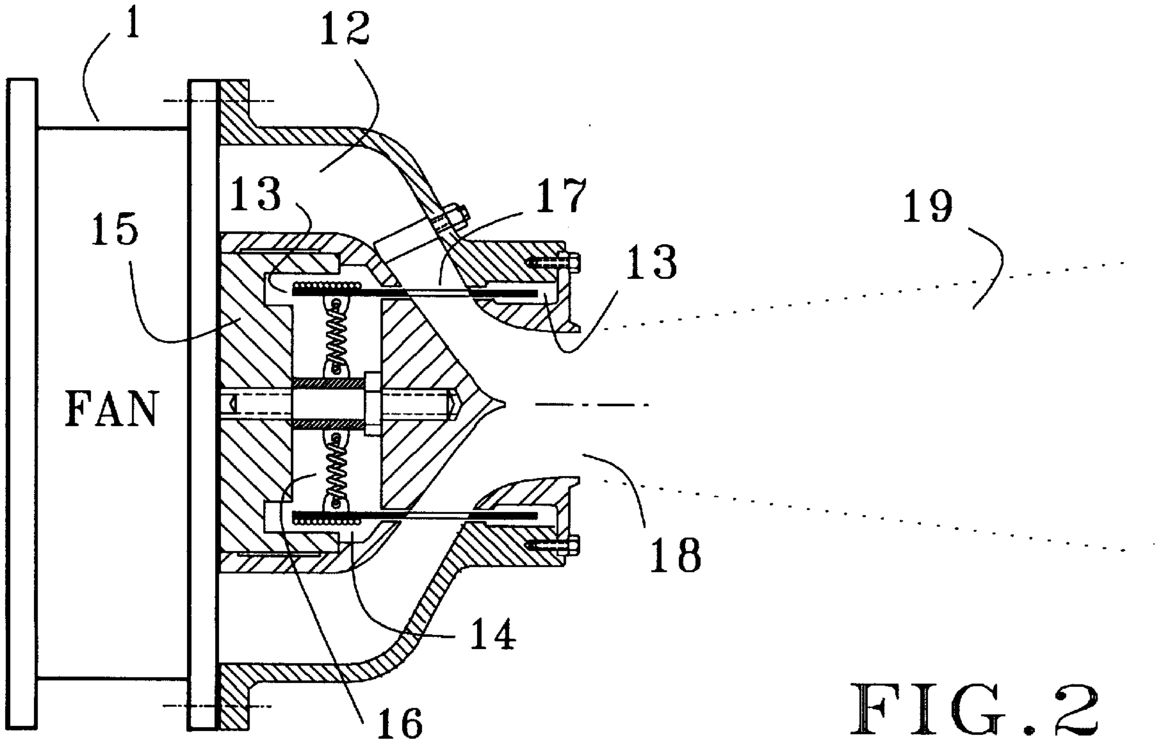
FAN

FIG. 2

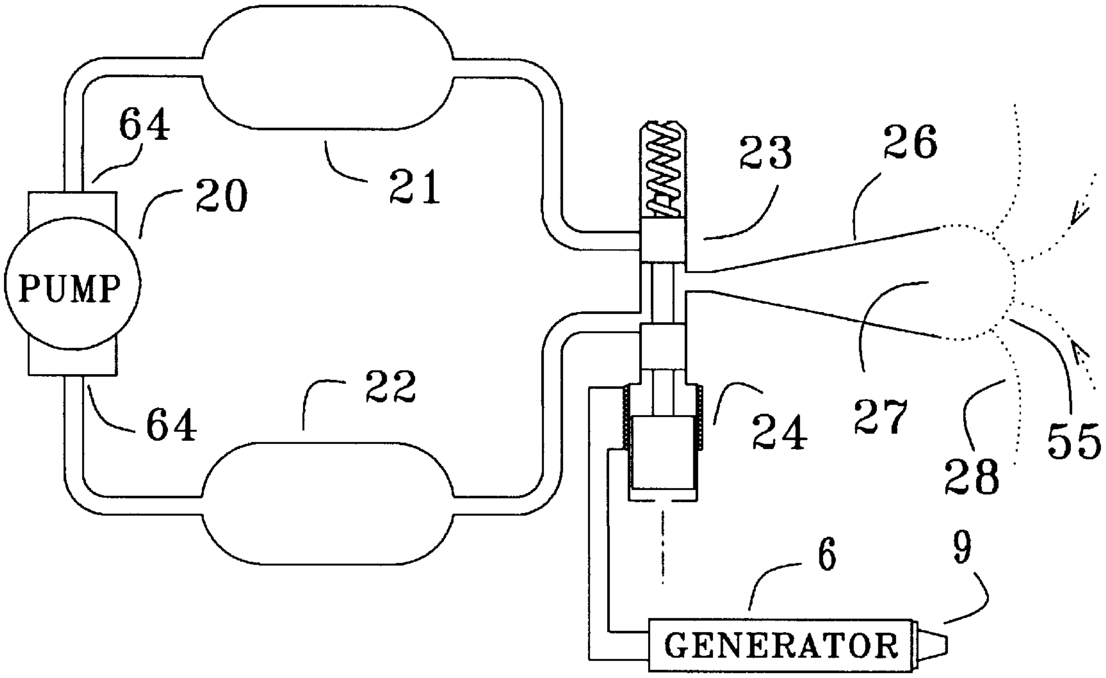
PUMP
GENERATOR

FIG. 3

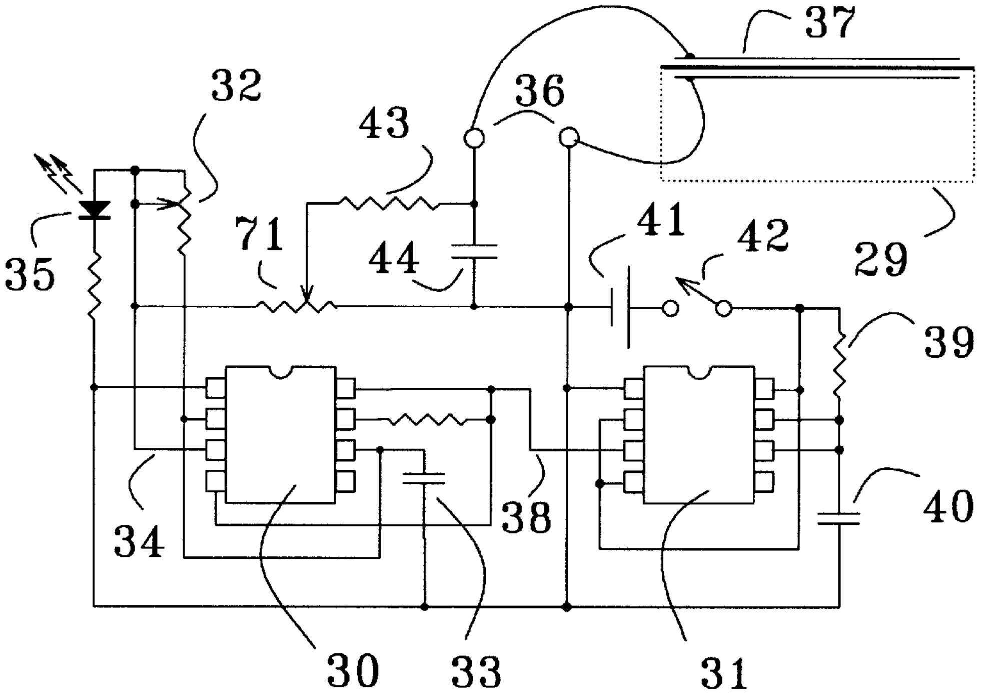

FIG. 4

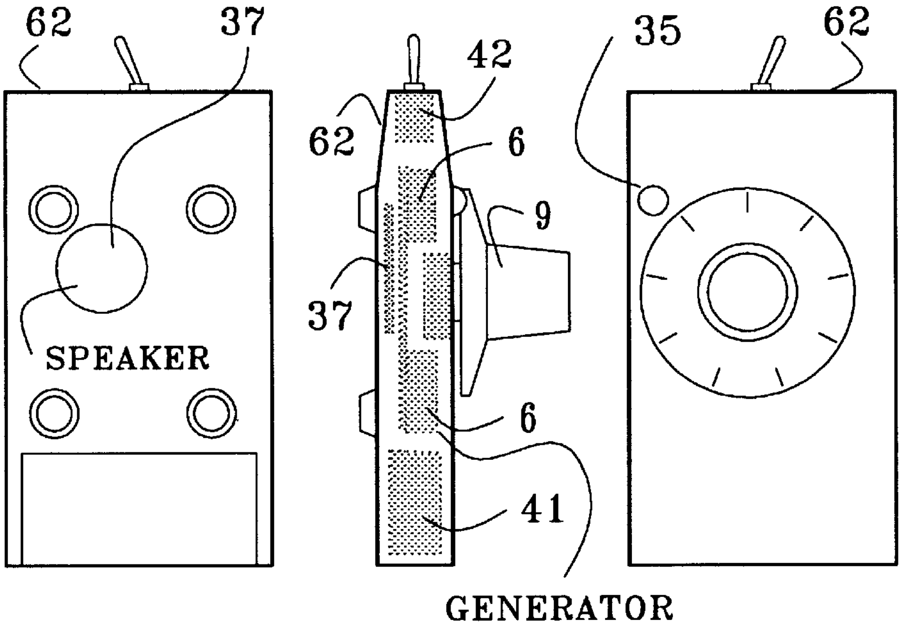
62
37
42
62
6
35
62
9
37
SPEAKER
6
41
GENERATOR

FIG. 5

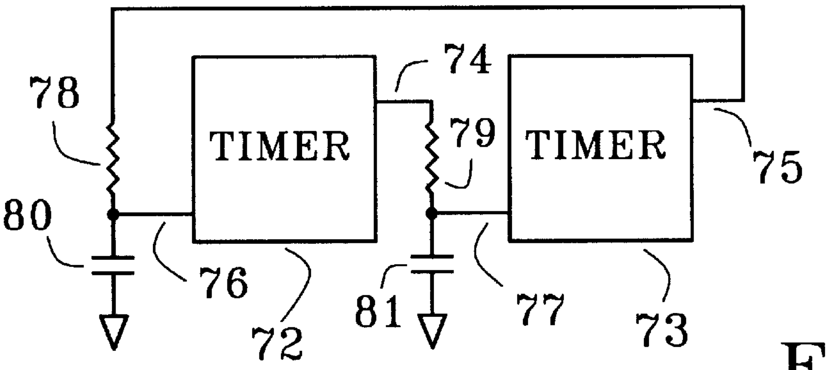
78
TIMER
74
79
TIMER
75
80
76
72
81
77
73

FIG. 6

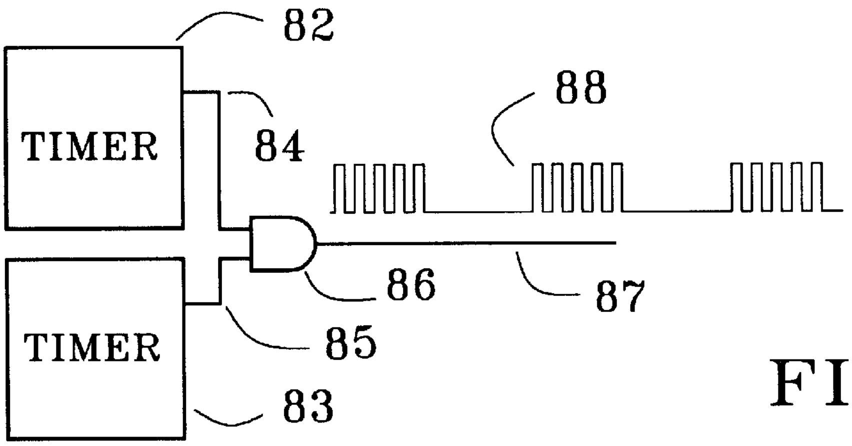
82
TIMER
84
88
86
87
85
TIMER
83

FIG. 7

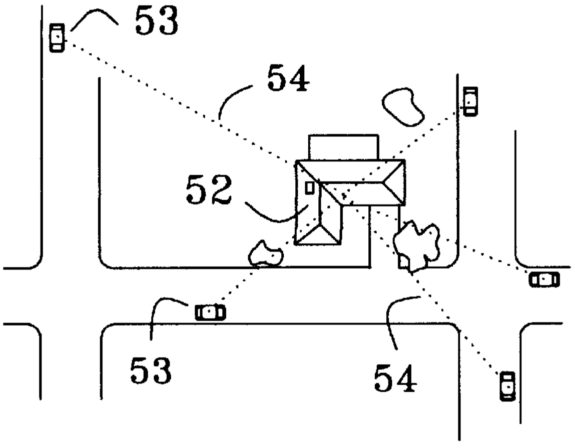
53
54
52
53
54

FIG.8

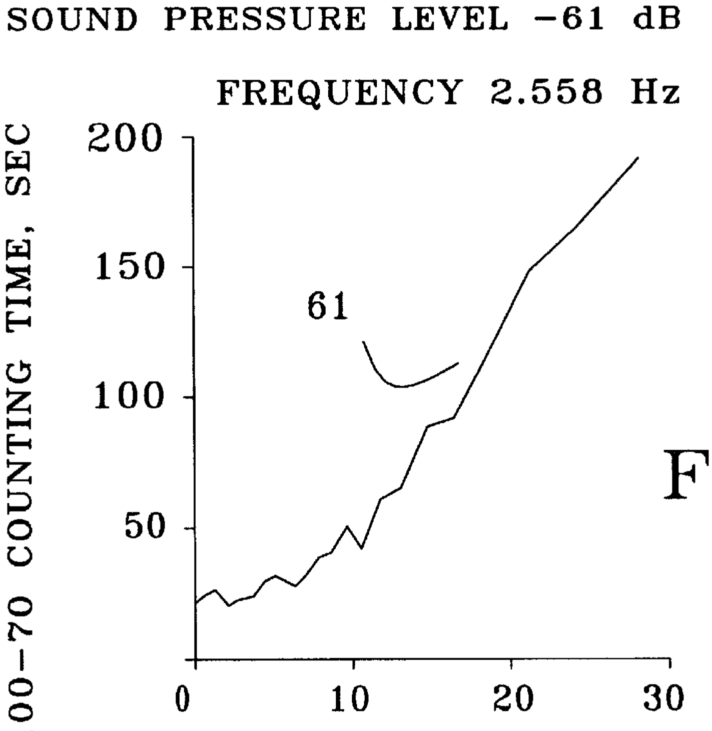

FIG.11

FIG.12

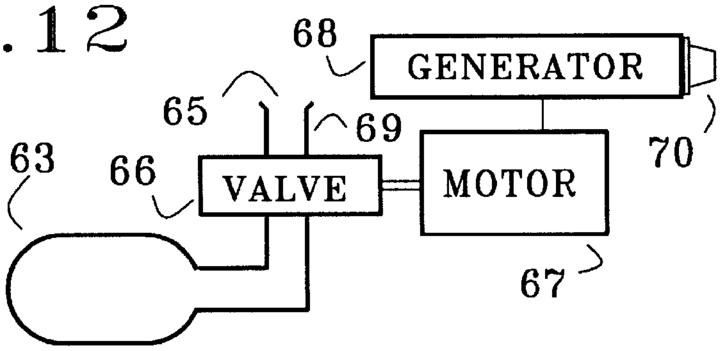
68
GENERATOR
65
69
70
63
66
VALVE
MOTOR
67

FIG.9

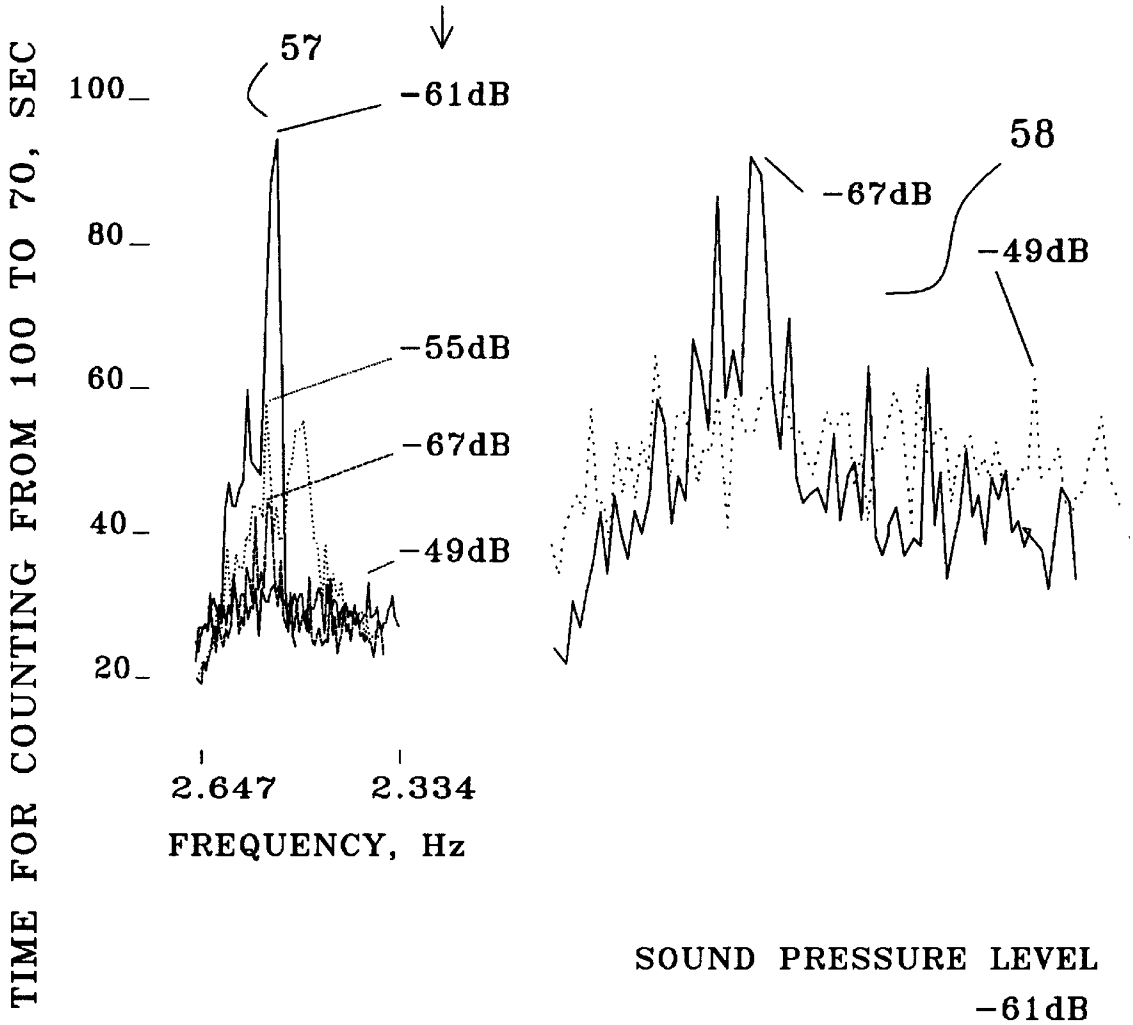
SOUND PRESSURE LEVEL
57
-61dB
-55dB
-67dB
-49dB
58
-67dB
-49dB
100
80
60
40
20
TIME FOR COUNTING FROM 100 TO 70, SEC
2.647
2.334
FREQUENCY, Hz

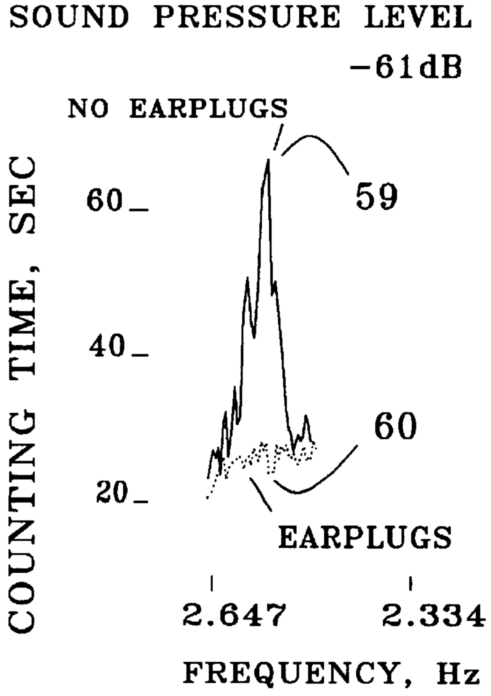
SOUND PRESSURE LEVEL
-61dB
NO EARPLUGS
59
60
EARPLUGS
60
40
20
COUNTING TIME, SEC
2.647
2.334
FREQUENCY, Hz

FIG.10

SUBLIMINAL ACOUSTIC MANIPULATION OF NERVOUS SYSTEMS

BACKGROUND OF THE INVENTION

The central nervous system can be manipulated via sensory pathways. Of interest here is a resonance method wherein periodic sensory stimulation evokes a physiological response that peaks at certain stimulus frequencies. This occurs for instance when rocking a baby, which typically provides relaxation at frequencies near ½ Hz. The peaking of the physiological response versus frequency suggests that one is dealing here with a resonance mechanism, wherein the periodic sensory signals evoke an excitation of oscillatory modes in certain neural circuits. The sensory pathway involved in the rocking example is the vestibular nerve. However, a similar relaxing response at much the same frequencies can be obtained by gently stroking a child's hair, or by administering weak heat pulses to the skin, as discussed in U.S. Pat. No. 5,800,481, Sep. 1, 1998. These three types of stimulation involve different sensory modalities, but the similarity in responses and effective frequencies suggests that the resonant neural circuitry is the same. Apparently, the resonance can be excited either via vestibular pathways or via cutaneous sensory pathways that carry tactile or temperature information.

Near 2.5 Hz another sensory resonance has been found that can be excited by weak heat pulses induced in the skin, as discussed in U.S. Pat. No. 5,800,481, Sep. 1, 1998. This sensory resonance brings on a slowing of certain cortical functions, as indicated by a pronounced increase in the time needed to silently count backward from 100 to 70 with the eyes closed. The effect is sharply dependent on frequency, as shown by a response peak a mere 0.13 Hz wide. The thermally excited 2.5 Hz resonance was found to also cause sleepiness, and after long exposure, dizziness and disorientation.

Other, more obscure types of stimulation in the form of weak magnetic fields or weak external electric fields can also cause the excitation of sensory resonances, as

SUMMARY OF THE INVENTION

Experiments have shown that atmospheric acoustic stimulation of deeply subliminal intensity can excite in a human subject the sensory resonances near ½ Hz and 2.5 Hz. The ½ Hz resonance is characterized by ptosis of the eyelids, relaxation, drowsiness, a tonic smile, tenseness, or sexual excitement, depending on the precise acoustic frequency near ½ Hz that is used. The observable effects of the 2.5 Hz resonance include a slowing of certain cortical functions, sleepiness, and, after long exposure, dizziness and disorientation. The finding that these sensory resonances can be excited by atmospheric acoustic signals of deeply subliminal intensity opens the way to an apparatus and method for acoustic manipulation of a subject's nervous system, wherein weak acoustic pulses are induced in the atmosphere at the subject's ears, and the pulse frequency is tuned to the resonance frequency of the selected sensory resonance. The method can be used by the general public for control of insomnia and anxiety, and for facilitation of relaxation and sexual arousal. Clinical use of the method includes the control and perhaps a treatment of anxiety disorders, tremors, and seizures. A suitable embodiment for these applications is a small portable battery-powered subaudio acoustic radiator which can be tuned to the resonance frequency of the selected sensory resonance.

There is an embodiment suitable for law enforcement operations in which a subject's nervous system is manipulated from a considerable distance, as in a standoff situation. Subliminal subaudio acoustic pulses at the subject's location may then be induced by acoustic waves radiating from a venting acoustic monopole, or by a pulsed air jet, especially when aimed at the subject or at another material surface, where the jet velocity fluctuations are wholly or partly converted into static pressure fluctuations.

The described physiological effects occur only if the intensity of the acoustic stimulation falls in a certain range, called the effective intensity window. This window has been measured in exploratory fashion for the 2.5 Hz resonance.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** illustrates a preferred embodiment wherein a modulated air jet is used for inducing subliminal acoustic pulses in the atmosphere at the subject's ears, for the purpose of manipulating the subject's nervous system.

FIG. **2** shows an embodiment in which a pulsed air jet is produced by modulating the flow from a fan by a cylindrical sheet valve that is driven by a voice coil.

FIG. **3** shows schematically an acoustic monopole operated by a solenoid valve.

FIG. **4** shows the circuit of a simple generator for producing voltage pulses that drive a piezoelectric speaker.

FIG. **5** depicts a portable battery-powered device that contains the circuit and the piezoelectric speaker of FIG. **4**.

FIG. **6** shows schematically a generator for chaotic pulses.

FIG. **7** depicts a circuit for generating a complex wave.

FIG. **8** illustrates an application in a law enforcement standoff situation.

FIG. **9** contains experimental data that show excitation of the sensory resonance near 2.5 Hz, and the effective intensity window.

FIG. **10** depicts experimental data showing that the sensory excitation occurs via the ear canal.

FIG. **11** shows the buildup of the physiological response to the acoustic stimulation.

FIG. **12** shows schematically an acoustic monopole operated by a rotating valve.

DETAILED DESCRIPTION OF THE INVENTION

It has been found in our laboratory that deeply subliminal atmospheric acoustic pulses with frequency near ½ Hz can evoke in a human subject a nervous system response that includes ptosis of the eyelids, relaxation, drowsiness, the feeling of pressure at a centered spot on the brow, seeing moving patterns of dark purple and greenish yellow with the eyes closed, a soft warm feeling in the stomach, a tonic smile, a "knot" in the stomach, sudden loose stool, and sexual excitement, depending on the precise acoustic frequency used. These responses show that this sensory resonance involves the autonomic nervous system.

The sharp peaking of the physiological effects with frequency is suggestive of a resonance mechanism, wherein the acoustic stimulation, although subliminal, causes excitation of a resonance in certain neural circuits. Since the frequencies and responses are similar to those for the ½ Hz sensory resonance discussed in the Background Section, it appears that the resonance excited by the described acoustic stimulation is indeed the ½ Hz sensory resonance. It has been found that the 2.5 Hz sensory resonance can be excited acoustically as well. This sensory resonance causes the slowing of certain cortical processes, sleepiness, and eventually dizziness and disorientation.

One can avoid the described physiological responses by wearing snugly fitting ear plugs. This shows that the excitation occurs via the external ear canal, so that the stimulation proceeds either through the auditory nerve or the vestibular nerve. Frequencies near ½ Hz or 2.5 Hz are far too low for stimulating the cochlear apparatus, but they are within the response range of hair cells in the vestibular end organ. Also, there exists a low-frequency acoustic path to the vestibular end organ by virtue of the ductus reuniens which provides a fluid connection between the cochlea and the vestibular organ. The narrow duct severely attenuates acoustic signals and acts as a low pass filter with a very low cutoff frequency. Subaudio acoustic signals, i.e., acoustic signals with frequencies up to 15 Hz, may perhaps penetrate to the vestibular organ with sufficient strength for stimulating the exquisitely sensitive vestibular hair cells.

For the ½ Hz and 2.5 Hz resonances, the described physiological responses are observed only if the acoustic intensity lies in a certain interval, called the effective intensity window. The acoustic intensity levels in this window are far below the human auditory threshold, so that exposed subjects do not sense the acoustic pulses in any other way than through the mentioned physiological effects. The upper limit of the effective intensity window is believed to be due to nuisance-guarding neural circuitry that blocks repeditive nuisance signals from higher processing.

The acoustic signals used for the excitation of sensory resonances have the nature of pulses. The pulses may be square, trapezoid, or triangle, or rounded versions of these shapes. However, depending on the pulse frequency, strong harmonics with frequencies in the audible range could stimulate the cochlear apparatus. This may be avoided by using sine waves or appropriately rounded other waves with low harmonic content.

The acoustic pulses occur in the atmosphere air; even when administered with earphones, the pulses at the subject's ear constitute pressure and flow pulses in the local atmospheric air.

The resonance frequencies of the ½ Hz and 2.5 Hz sensory resonances lie respectively near ½ and 2.5 Hz. The different physiological effects mentioned occur at slightly different frequencies. Thus, one can tune for drowsiness or sexual excitement, as desired. The precise resonance frequency is also expected to depend slightly on the subject and the state of the nervous and endocrine systems, but it can be measured readily by tuning the acoustic pulse frequency for maximum physiological effect. Besides the resonances near ½ and 2.5 Hz, other sensory resonances may perhaps be found, and those with resonance frequencies below 15 Hz are expected to be excitable acoustically via the vestibular nerve, since the vestibular hair cells are sensitive in this frequency range.

The finding that deeply subliminal subaudio acoustic stimulation can influence the central nervous system suggests a method and apparatus for manipulating the nervous system of a subject by inducing subliminal atmospheric acoustic pulses of subaudio frequency at the subject's ears. In doing so, one may in addition exploit the sensory resonance mechanism, but there are important applications where this is not done. For example, the subliminal acoustic manipulation of the nervous system may be used clinically for the control of tremors and seizures, by detuning the pathological oscillatory activity of neural circuits that occurs in these disorders. This may be done by choosing an acoustic frequency that is slightly different from the frequency of the pathological oscillation. The evoked neural signals then cause phase shifts which may diminish or quench the oscillation. Exploitation of the resonance mechanism by tuning the acoustic signals to the resonance frequency of a selected sensory resonance affords other forms of manipulation, such as control of insomnia and anxiety, or facilitation of sexual arousal.

For both types of manipulation, the required subliminal subaudio acoustic pulses may be induced at one or both of the subject's ears by earphones with a proper low-frequency response, acoustic waves generated by an acoustic source and propagated through the atmosphere, or by a pulsed jet of gas (which may be air), preferably directed at a material surface open to the atmosphere, such as a wall or the subject's skin or clothing. In the area of impact, especially where the surface is oriented substantially perpendicular to the jet, atmospheric pressure pulses are then generated by virtue of the ram effect, wherein flow velocity fluctuations are wholly or partly converted into static pressure fluctuations. If the material surface on which the jet impinges includes the subject's ears, then these pressure pulses cause direct stimulation of the subject, but the pulses also propagate through the atmosphere to the subject's ears by virtue of acoustic wave propagation along accessible paths.

The induction of atmospheric acoustic pulses by a pulsed air jet proceeding in the atmosphere and directed at a subject is shown in FIG. **1**, where a blower **1**, labeled "FAN", produces an air jet **2** that is directed at a subject **3**. The fan is powered by a power supply **4**, labelled "SUPPLY". At the fan, the supply voltage is modulated in pulsed fashion by a relay **5** controlled by the generator **6**, labelled "GENERATOR", through voltage pulses **7** supplied to electromagnet windings **8**. A user can adjust the frequency of the pulses with the tuning control **9**. The pulsing of the voltage supplied to the fan causes the momentum flux **10** of the air jet to be modulated in a pulsed manner. Upon impinging on a material surface such as the skin of the subject **3**, the pulsed jet induces acoustic pressure pulses at the ears **11** of the subject. The atmospheric acoustic effect of the jet is complicated by the fact that the region of the fan inlet undergoes a fluctuation of static pressure as the result of the modulation of jet momentum flux. There thus are two distinct acoustic monopoles, one at the fan inlet and the other in the area of impact of the jet on the material surface. The monopoles radiate with a phase difference that is determined by the jet velocity, the modulation frequency, and the distance between fan and impact area. The resulting sound pressure at the subject's ears can be analyzed with retarded potentials as discussed for instance by Morse and Feshbach (1953). Even a jet which does not impinge on a material surface radiates by virtue of the acoustic monopole at the fan inlet.

When skin of the subject is exposed to gas flow of the jet, or to the flow of atmospheric air entrained by the jet, the flow will fluctuate in pulsed fashion, so that a periodic heat flux occurs by convective transport and evaporation of sweat. The resulting periodic fluctuation of the skin temperature can excite a sensory resonance, as discussed in U.S. Pat. No. 5,800,481, Sep. 1, 1998. Hence, the apparatus of FIG. **1** can cause excitation of a sensory resonance via two separate sensory pathways, viz., the vestibular nerve and the afferents from cutaneous temperature receptors. The strength of the thermal stimulation depends on the skin area and type of skin exposed to the fluctuating flow. The face is particularly sensitive, especially the lips. The two-channel excitation of sensory resonances needs further investigation. In any particular situation, the vestibular channel can be blocked by using earplugs.

An air jet with pulsed momentum flux can also be obtained as illustrated in FIG. **2**. Shown is a fan **1**, labelled "FAN", which discharges into manifold **12**. The air flow in the manifold can be partially obstructed by a sheet valve **13** in the form of a perforated cylindrical sheet. The sheet valve carries a voice coil **14** which is situated in the field of a permanent magnet **15**, in the manner of conventional electromagnetic loudspeakers. When no current flows through the voice coil, the sheet valve is held in equilibrium position by springs **16**. In this position, the perforation **17** in the sheet is lined up with the flow passage allowing essentially unimpeded flow through the manifold and out the exit **18**, such as to form a jet **19** in the atmosphere. Sending a current pulse through the voice coil **14** causes the sheet valve to be displayed in the axial direction, thereby partially obstructing the air flow through the manifold. Owing to the low inertia of the sheet valve, the arrangement allows efficient pulse modulation of the jet momentum flux.

A somewhat different modulation system can be obtained with a rotating cylindrical sheet valve that has one or more holes along its periphery, and which is adjacent to a stationary cylindrical shroud that has corresponding holes, so that rotation of the valve causes modulation of the air flow through the holes. The valve is rotated by a stepper motor driven by voltage pulses. The latter are obtained from a generator that is controlled by a tuner.

One can also use direct acoustic wave propagation for inducing the required atmospheric acoustic pulses. It is then advantageous to employ as the source of the waves an acoustic monopole, since for these the acoustic pressure does not fall off as fast with increasing distance as for dipoles. Moreover, at the very low frequencies involved, acoustic pressure shorting across a conventional loudspeaker baffle is very severe. A sealed loudspeaker mounted in an airtight box eliminates this pressure shorting, and radiates acoustic waves with a relatively large monopole component.

An acoustic monopole may also be produced by having a source of pressurized gas vent through an orifice into the atmosphere in a pulsed fashion. The gas may be air. Alternatively, one may have a source of low-pressure air inhale atmospheric air through an orifice in pulsed fashion. These actions are easily achieved by an oscillating or rotating valve. For purposes of discussion it is convenient to introduce the concept of gaseous flux through the orifice, defined as the integral of the normal flow velocity component over an imagined surface that tightly caps the orifice, the normal component being perpendicular to the local surface element, and reckoned positive if the flow is directed into the ambient atmosphere. The gaseous flux has the dimension of m^3/s. For the case with a source of pressurized gas, the gaseous flux is positive and due to gas venting to the atmosphere. For the case with a source of vacuum, the gaseous flux is negative and due to atmospheric air entering the orifice. The strength of the acoustic monopole is expressed as the amplitude of the gaseous flux fluctuation, amplitude being defined as half the peak-to-peak variation. The concept of gaseous flux allows a unified discussion of venting acoustic monopoles that use a source of pressurized gas or a source of vacuum, or both.

The source of pressurized air could be a cylinder with pressurized gas, such as a CO_2 cartridge. For personal use, such a cartridge may last a long time because only very small acoustic monopole strengths are needed for the induction of the required weak acoustic signals. For long term and long range operation, the exhaust port of an air pump may serve as a source of pressurized air, and the intake port could be used as a source of vacuum.

A simple venting acoustic monopole is shown in FIG. **12**, where the source **63** of pressurized gas, which may be air, is connected to a conduit **69** which has an orifice **65** that is open to the atmosphere. A rotating valve **66** labelled "VALVE" controls the gaseous flux through the orifice. The valve is rotated by a stepper motor **67** labelled "MOTOR", driven by voltage pulses from the generator **68** labelled "GENERATOR". The motor speed is determined by the frequency of the voltage pulses. This frequency can be selected by the tuner **70**, which therefore controls the frequency of the acoustic pulses emited by the orifice **65**. For the simple orifice shown, boundary layer separation may occur in the outflow, so that the air pulses emerge in the form of jets. This causes dipole and higher multipole components in the radiated acoustic field. If desired, such radiation components can be avoided or diminished by placing a spherically or dome shaped fine mesh screen over the orifice **65**. Instead of holding pressurized gas, the source **63** may hold a vacuum. In either case, the pulsing of the gaseous flux causes radiation of monopole-type acoustic waves. The source **63** may be replenished by a pump.

Push-pull operation can be achieved in the manner shown in FIG. **3**. An air pump **20**, labelled "PUMP", with flow ports **64**, pressurizes the pressure vessel **21** while drawing a vacuum in the vacuum vessel **22**. A valve **23** is operated by the solenoid **24** such as to alternately admit high and low pressure air to the conduit **26**. The latter vents to the atmosphere through a screen **55** placed across an orifice **27** that is open to the atmosphere. The valve is controlled by an oscillator consisting of the solenoid **24**, which is connected to the pulse generator **6**, labelled "GENERATOR". The frequency of the electric current pulses through the solenoid is determined by the setting of the tuning control **9**. This frequency is to be tuned to the resonance frequency of the sensory resonance that is to be excited. The tuning may be done manually by a user. The conduit **26** is structured as a diffuser in order to avoid boundary layer separation during the exhaust phase; the screen across the orifice **27** inhibits formation of a jet, thereby providing more nearly for a monopole type acoustic wave. During the intake phase the orifice acts as a sink; streamlines **28** of the resulting flow are illustrated. The vessels **21** and **22** smooth the flow fluctuations through the orifice that are due to the flow fluctuations through the pump; they are drawn at a relatively small scale for compactness sake. Instead of the oscillating valve **23**, a rotating valve may be used, driven by a stepper motor powered by voltage pulses from a generator.

Conventional loudspeakers may be used as well as the source of acoustic radiation. An example is shown in FIG. **4**, where the piezoelectric transducer **37** is driven by a simple battery-powered pulse generator built around two RC timers **30** and **31**. Timer **30** (Intersil ICM7555, for instance) is hooked up for astable operation; it produces a square wave voltage with a frequency determined by capacitor **33** and the potentiometer **32**, which serves as a tuner that may be operated by a user. The square wave voltage at output **34** drives the LED **35**, and appears at one of the output terminals **36**, after voltage division by potentiometer **71**. The other output is connected to the negative supply. The output terminals **36** are connected to the piezoelectric speaker. Automatic shutoff of the voltage that powers the timer **30** at point **38** is provided by a second timer **31**, hooked up for monostable operation. Shutoff occurs after a time interval determined by resistor **39** and capacitor **40**. Timer **31** is powered by a 9 Volt battery **41**, via a switch **42**. Optional rounding of the square wave is done by an RC circuit consisting of a resistor **43** and capacitor **44**. An optional airtight enclosure **29** may be used for the speaker **37**, in order to enhance the monopole component of the radiated acoustic signal. Instead of a piezoelectric speaker one may use an electromagnetic loudspeaker with a voice coil. Because of the low impedance of the voice coil, a resistor must then be included in the output circuitry in order to keep the output currents to low values such as to allow battery powering of the device. Small voice coil currents are sufficient for the low acoustic powers required.

Low pulse frequencies can be monitored with the LED **35** of FIG. **3**. The LED blinks on and off with the square wave, and it doubles as a power indicator. The pulse frequency can be determined by reading a clock and counting the LED light pulses. For higher frequencies a monitoring LED can still be used, if it is driven by a signal obtained by frequency division of the generator signal.

The automatic shutoff described above is an example for automatic control of the generated voltage; more sophisticated forms of control involve automatic frequency sequences. A computer that runs a simple timing program can be used for the generation of all sorts of square waves that can be made available at a computer port. An economic and compact version of such arrangement is provided by the Basic Stamp manufactured by Parallax Inc, Rocklin, Calif., which has an onboard EEPROM that can be programmed for the automatic control of the generated pulses, such as to provide desired on/off times, frequency schedules, or chaotic waves. The square waves can be rounded by RC circuits, and further smoothed by integration and filtering.

A compact packaging of the device such as shown of FIG. **4** is depicted in FIG. **5** where all circuit parts and the speaker, piezoelectric or voice-coil type, are contained in a small casing **62**. Shown are the speaker **37**, labelled "SPEAKER", driven by the generator **6**, labeled "GENERATOR", with tuning control **9**, LED **35**, battery **41**, and power switch **42**. The LED doubles as a mark for the tuning control dial. With the circuit of FIG. **4**, the device draws so little current that it can be used for several months as a sleeping aid, with a single 9 Volt battery.

For the purpose of thwarting habituation to the stimulation, irregular features may be introduced in the pulse train, such as small short-term variations of frequency of a chaotic or stochastic nature. Such chaotic or stochastic acoustic pulses can cause excitation of a sensory resonance, provided that the average pulse frequency is close to the appropriate sensory resonance frequency. A chaotic square wave can be generated simply by cross coupling of two timers. FIG. **6** shows such a hookup, where timers **72** and **73**, each labeled "TIMER", have their output pins **74** and **75** connected crosswise to each other's control voltage pins **76** and **77**, via resistors **78** and **79**. The control voltage pins **76** and **75** have capacitors **80** and **81** to ground. If the timers are hooked up for astable operation with slightly different frequencies, and appropriate values are chosen for the coupling resistors and capacitors, the output of either timer is a chaotic square wave with an oval attractor. Example circuit parameters are: R_{78}=440KΩ, R_{79}=700KΩ, C_{80}=4.7 μF, C_{81}=4.7 μF, with $(RC)_{72}$=0.83 s and $(RC)_{73}$=1.1 s. For these parameters, the output **74** of timer **72** is a chaotic square wave with a power spectrum that has large peaks at 0.46 Hz and 0.59 Hz. The resulting chaotic wave is suitable for the excitation of the ½ Hz resonance.

A complex wave may be used for the joint excitation of two different sensory resonances. A simple generator of a complex wave, suitable for the joint excitation of the ½ Hz autonomic resonance and the 2.5 Hz cortical resonance, is shown in FIG. **7**. Timers **82** and **83** are arranged to produce square waves of frequencies f_1 and f_2 respectively, where f_1 is near 2.5 Hz, and f_2 is near ½ Hz. The outputs **84** and **85** of the timers are connected to the inputs of an AND gate **86**. The output **87** of the AND gate features a square wave of frequency f_1, amplitude modulated by a square wave of frequency f_2, as indicated by the pulse train **88**.

The very low frequency waves needed for the acoustic stimulation of the vestibular nerve may also be provided by a sound system in which weak subaudio pulses are added to audible audio program material. This may be done in the customary manner way of adding the currents from these signals at the inverting input of an operational amplifier. The amplitude of the pulses is chosen such that the strength of the resulting acoustic pulses lies in the effective intensity window. Experiments in our laboratory have shown that the presence of audible signals, such as music or speech, does not interfere with the excitation of sensory resonances.

The invention can also be implemented as a sound tape or CD ROM which contains audible audio program material together with subliminal subaudio signals. The recording can be done by mixing the audio and subaudio signals in the usual manner. In choosing the subaudio signal level, one must compensate for the poor frequency response of the recorder and the electronics, at the ultra low subaudio frequencies used.

The pathological oscillatory neural activity involved in epileptic seizures and Parkinson's disease is influenced by the chemical milieu of the neural circuitry involved. Since the excitation of a sensory resonance may cause a shift in chemical milieu, the pathological oscillatory activity may be influenced by the resonance. Therefore, the acoustic excitation discussed may be useful for control and perhaps treatment of tremors and seizures. Frequent use of such control may afford a treatment of the disorders by virtue of facilitation and classical conditioning.

In this as well as in the detuning method discussed before, an epileptic patient can switch on the acoustic stimulation upon sensing a seizure precursor.

Since the autonomic nervous system is influenced by the ½ sensory resonance, the acoustic excitation of the resonance may be used for the control and perhaps the treatment of anxiety disorders.

The invention can be embodied as a nonlethal weapon that remotely induces disorientation and other discomfort in targeted subjects. Large acoustic power can be obtained easily with acoustic monopoles of the type depicted in FIG. **3** or FIG. **12**. If considerable distance needs to be maintained to the subject, as in a law enforcement standoff situation illustrated in FIG. **8**, several monopoles can be used, and it then may become important to have phase differences between the acoustic signals of the individual monopoles arranged in such a manner as to maximize the amplitude of the resultant acoustic signal at the location **52** of the subject. Shown are four squad cars **53**, each equiped with an acoustic monopole capable of generating atmospheric pulses of a frequency appropriate for the excitation of sensory resonances. The relative phases of the emitted pulses are arranged such as to compensate for differences of acoustic path lengths **54**, such that the pulses arrive at the subject location **52** with substantially the same phase, resulting in constructive interference of the local acoustic waves. Such arrangement can be achieved easily by using radio signals between the monopole units, with the target distances either dialed in manually or measured automatically with a range finder. The subaudio acoustic signals can easily penetrate into a house through an open window, a chimney, or a crack under a closed door.

Some of our experiments on acoustic excitation of sensory resonances which provide a basis for the present invention will be discussed presently. Of all the responses to the excitation of the ½ Hz resonance, ptosis of the eyelids stands out for distinctness, ease of detection, and sensitivity. When voluntary control of the eyelids is relinquished, the eyelid position is determined by the relative activities of the sympathetic and parasympathetic nervous systems. There are two ways in which ptosis can be used as an indicator that the autonomic system is being affected. In the first, the subject simply relaxes the control over the eyelids, and makes no effort to correct for any drooping. The more sensitive second method requires the subject to first close the eyes about half way. While holding this eyelid position, the eye are rolled upward, while giving up voluntary control of the eyelids. With the eyeballs turned up, ptosis will decrease the amount of light admitted to the eyes, and with full ptosis the light is completely cut off. The second method is very sensitive because the pressure excerted on the eyeballs by partially closed eyelids increases parasympathetic activity. As a result, the eyelid position becomes somewhat labile, exhibiting a slight flutter. The labile state is sensitive to small shifts in the activities of the sympathetic and parasympathetic systems. The method works best when the subject is lying flat on the back and is facing a moderately lit blank wall of light color.

The frequency at which ptosis is at a maximum is called the ptosis frequency. This frequency depends somewhat on the state of the nervous and endocrine systems, and it initially undergoes a downward drift, rapid at first and slowing over time. The ptosis frequency can be followed in its downward drift by manual frequency tracking aimed at keeping ptosis at a maximum. At a fixed frequency, the early ptosis can be maintained in approximately steady state by turning the acoustic stimulation off as soon as the ptosis starts to decrease, after which the ptosis goes through an increase followed by a decline. The acoustic stimulation is turned back on as soon as the decline is perceived, and the cycle is repeated.

At fixed frequencies near ½ Hz, the ptosis cycles slowly up and down with a period ranging upward from about 3 minutes, depending on the precise acoustic frequency used. The temporal behavior of the ptosis frequency is found to depend on the acoustic pulse intensity; the drift and cycle amplitude are smaller near the low end of the effective intensity window. This suggests that the drift and the cycling of the ptosis frequency is due to chemical modulation, wherein the chemical milieu of the neural circuits involved affects the resonance frequency of these circuits, while the milieu itself is influenced by the resonance in delayed fashion. Pertinent concentrations are affected by production, diffusion, and reuptake of the substances involved. Because of the rather long characteristic time of the ptosis frequency shift, as shown for instance by the cycle period lasting **3** minutes or longer, it is suspected that diffusion plays a rate-controlling role in the process.

The resonance frequencies for the different components of the ½ Hz sensory resonance have been measured, using acoustic sine waves with a sound pressure of 2×10^{-9} N/m^2 at the subject's left ear. Ptosis reached a steady state at a frequency of 0.545 Hz. Sexual excitement occurred at two frequencies, 0.530 Hz and 0.597 Hz, respectively below and above the steady-state ptosis frequency. For frequencies of 0.480 Hz and 0.527 Hz the subject fell asleep, whereas tenseness was experienced in the range from 0.600 to 0.617 Hz.

The resonance near 2.5 Hz may be detected as a pronounced increase in the time needed to silently count backward from 100 to 70, with the eyes closed. The counting is done with the "silent voice" which involves motor activation of the larynx appropriate to the numbers to be uttered, but without passage of air or movement of mouth muscles. The motor activation causes a feedback in the form of a visceral stress sensation in the larynx. Counting with the silent voice is different from merely thinking of the numbers, which does not produce a stress sensation, and is not a sensitive detector of the resonant state. The larynx stress feedback constitutes a visceral input into the brain and may thus influence the amplitude of the resonance. This unwanted influence is kept to a minimum by using the count sparingly in experiment runs. Since counting is a cortical process, the 2.5 Hz resonance is called a cortical sensory resonance, in distinction with the autonomic resonance that occurs near ½ Hz. In addition to affecting the silent counting, the 2.5 Hz resonance is expected to influence other cortical processes as well. It has also been found to have a sleep inducing effect. Very long exposures cause dizziness and disorientation. The frequency of 2.5 Hz raises concerns about kindling of epileptic seizures; therefore, the general public should not use the 2.5 Hz resonance unless this concern has been laid to rest through further experiments.

The sensitivity and numerical nature of the silent count makes it a very suitable detector of the 2.5 Hz sensory resonance. It therefore has been used for experiments of frequency response and effective intensity window. In these experiments, rounded square wave acoustic pulses were produced with a frequency that was slowly diminished by computer, and the subject's 100–70 counting time was recorded for certain frequencies. The acoustic transducer was a small loudspeaker mounted in a sealed cabinet such as to provide acoustic monopole radiation. At fixed frequency, the acoustic monopole strenght in m^3/s varies linearly with the voice coil current, with a constant of proportionality that can be calculated from measured speaker dome excursions for given currents. The sound pressure level at the entrance of the subject's nearest external ear canal can be expressed in terms of the acoustic monopole strength and the distance from the loudspeaker. For each experiment run, the sound pressure level at the entrance of the subject's external ear canal can thus be calculated from the measured amplitude of the voice coil current and the pulse frequency. Since for the subaudio frequencies the distance from the acoustic radiator to the subject's ear is much smaller than the wavelength of the sound, the near-field approximation was used in this calculation. The sound pressure level was expressed in dB relative to the reference sound pressure of 2×10^{-5} N/m^2. This reference pressure is traditionally used in the context of human hearing, and it represents about the normal minimum human hearing threshold at 1.8 KHz.

FIG. **9** shows the result of experiment runs at sound pressure levels of −67, −61, −55, and −49 dB. Plotted are the subject's 100–70 counting time versus pulse frequency in a narrow range near 2.5 Hz. Resonance is evident from the sharp peak **57** in the graph for the sound pressure level of −61 dB. The graphs also reveal the effective intensity window for the stimulation, as can be seen by comparing the magnitude of the peaks for the different sound pressure levels. For increasing intensity, the magnitude of the peak first increases but then decreases, and no significant peak shows up in the graph for the largest sound pressure of −49 dB; this can be seen better from the insert **58**, which shows the graphs for −67 and −49 dB in a magnified scale. It follows that the effective intensity window extends approximately from −73 to −49 dB, in terms of the sound pressure level at the entrance of the subject's external ear canal.

The physiological response to the 2.5 Hz acoustic stimulation can be avoided by wearing earplugs. FIG. **10** is a plot of the 100–70 counting time versus acoustic pulse frequency, with and without earplugs. The sound pressure level at the entrance of the subject's external ear canal was −6 dB for both runs. Without earplugs the counting time has the peak **59**, but no significant peak is seen in graph **60** for the run in which the subject used earplugs. Two conclusions can be reached from these results. First, in the experiments the 2.5 Hz resonance is essentially excited acoustically rather than through the magnetic field induced by the voice coil currents in the loudspeaker. Second, it follows that the exciting sound essentially propagates via the external ear canal, instead of through the skin and bones in the area of the ears, or via cutaneous mechanoreceptors in the skin at large.

To answer the question whether the acoustic excitation of the 2.5 Hz sensory resonance occurs perhaps through the cochlear nerve, one needs to consider the human auditory threshold curve such as shown, for instance, by Thomson (1967). The curve has a minimum near 1.8 KHz where the threshold sound pressure level is 0 dB, by definition. At 10 Hz the threshold is 105 dB. Hence, the pronounced acoustic excitation of the sensory resonance shown in FIG. **9** for a sound pressure level of −61 dB is 166 dB below the auditory threshold at 10 Hz. The excitation occurs near 2.5 Hz, and at that frequency, the auditory threshold is even higher than at 10 Hz. Although the curve in Thomson's book does not go below 10 Hz, linear extrapolation suggests the estimate of 135 dB for the threshold at 2.5 Hz, bringing the sound pressure level that is effective for acoustic excitation of the sensory resonance to 196 dB below the estimated threshold at the frequency near 2.5 Hz used. This result all but rules out excitation via the cochlear nerve.

Chemical modulation may be the cause for the small frequency difference for peaks **57** and **59** in FIGS. **9** and **10**, for the sound pressure level of −61 dB; these peaks occur respectively at 2.516 and 2.553 Hz.

The physiological response to the excitation of the sensory resonances at a fixed stimulus frequency is not immediate but builds over time. An example is shown in FIG. **11**, where the graph **61** depicts the measured 100–70 time plotted versus elapsed time, upon application of acoustic pulses of 2.558 Hz frequency and a sound pressure level of −61 dB. The graph shows that the response is initially delayed over about 5 minutes; thereafter it increases, and at about 22 minutes the slope is seen to decrease somewhat. Other experiments have shown a counting time that eventually settles on a plateau, or even starts on a decline. Chemical modulation and habituation could account for these features. The response curve depends strongly on initial conditions.

The method is expected to be effective also on certain animals, and applications to animal control are therefore envisioned. The nervous system of mammals is similar to that of humans, so that the sensory resonances are expected to exist, albeit with different frequencies. Accordingly, in the present invention subjects are mammals.

The described method and apparatus can be used beneficially by the general public and in clinical work. Unfortunately however, there is the possibility of mischievous use as well. For instance, with small modifications the method of FIG. **1** can be employed to imperceptibly modulate the air flow in air conditioning or heating systems that serve a home, office building, or embassy, for covert manipulation of the nervous systems of occupants.

The invention is not limited by the embodiments shown in the drawings and described in the specification, which are given by way of example and not of limitation, but only in accordance with the scope of the appended claims.

REFERENCES

P. M. Morse and H. Feshbach, METHODS OF THEORETICAL PHYSICS, McGraw-Hill, New York, 1953

R. F. Thomson, FOUNDATIONS OF PHYSIOLOGICAL PSYCHOLOGY, Harper & Row, New York 1967

I claim:

1. Apparatus for manipulating the nervous system of a subject, the subject having an ear, comprising:

generator means for generating voltage pulses;

induction means, connected to the generator means and responsive to the voltage pulses, for inducing at the ear subliminal atmospheric acoustic pulses with a pulse frequency less than 15 Hz.

2. The apparatus according to claim **1**, further comprising means for automatically controlling the voltage pulses.

3. The apparatus according to claim **1**, further comprising means for monitoring the voltage pulses.

4. The apparatus according to claim **1**, for exciting in the subject a sensory resonance that occurs at a resonance frequency less than 15 Hz, the apparatus further comprising tuning means for enabling a user to tune the pulse frequency to the resonance frequency.

5. The apparatus according to claim **4**, further including a casing for containing the generator means, the induction means and the tuning means.

6. The apparatus according to claim **1**, wherein said induction means comprise:

means for generating in the atmosphere a gas jet, the latter having a momentum flux; and

modulation means, connected to the generator means and responsive to said voltage pulses, for pulsing the momentum flux with a frequency less than 15 Hz;

whereby subaudio acoustic pulses are induced in the atmosphere.

7. Apparatus for manipulating the nervous system of a subject, the subject having an ear, comprising:

generator means for generating voltage pulses;

a source of gas at a pressure different from the ambient atmospheric pressure;

a conduit having an orifice open to the atmosphere for passing a gaseous flux;

valve means, connected to the source of gas and the conduit to control the gaseous flux;

means, connected to the generator means and responsive to said voltage pulses, for operating the valve means to provide an oscillation of the gaseous flux with a frequency less than 15 Hz.

8. The apparatus according to claim **7**, further comprising vessel means for smoothing fluctuations of the gaseous flux caused by fluctuations in the pressure of the source of gas.

9. A method for manipulating the nervous system of a subject, the subject having an ear, comprising the steps of:

generating voltage pulses; and

inducing, in a manner responsive to the voltage pulses, at the ear subliminal atmospheric acoustic pulses with a pulse frequency less than 15 Hz.

10. The method according to claim **9**, for exciting in the subject a sensory resonance that occurs at a resonance frequency less than 15 Hz, further comprising the step of tuning the pulse frequency to the resonance frequency.

11. The method according to claim **9**, wherein said inducing comprises the steps of:

generating in the atmosphere a gas jet, the latter having a momentum flux; and

modulating the momentum flux in pulse-wise fashion in a manner responsive to the voltage pulses.

12. The method according to claim **11**, further comprising the step of directing the gas jet at a material surface.

13. The method according to claim **9**, wherein said inducing comprises the steps of:

generating a gas flow through a conduit orifice that is open to the atmosphere; and

modulating the gas flow to produce flow pulsations, in a manner responsive to the voltage pulses.

14. A method for remotely manipulating the nervous system of a subject in the course of law enforcement in a standoff situation, the subject having an ear, comprising the steps of:

generating voltage pulses;

generating, in a manner responsive to the voltage pulses, atmospheric acoustic signals at a plurality of locations remote from the subject for inducing at the ear subliminal atmospheric acoustic pulses with a pulse frequency less than 15 Hz, the signals having phase differences with respect to each other arranged to cause constructive acoustic wave interference at the subject.

15. A method for exciting in a subject a sensory resonance having a resonance frequency less than 15 Hz, the subject having an ear, comprising the steps of:

generating voltage pulses;

inducing, in a manner responsive to the voltage pulses, at the ear subliminal atmospheric acoustic pulses with a pulse frequency less than 15 Hz;

tuning the pulse frequency to the resonance frequency; and also

inducing audible audio-frequency atmospheric acoustic signals at the ear.

16. A method for controlling in a subject neurological disorders that involve pathological oscillatory activity of neural circuits, the subject having an ear, comprising the steps of:

generating voltage pulses;

inducing, in a manner responsive to the voltage pulses, at the ear subliminal atmospheric acoustic pulses with a pulse frequency less than 15 Hz; and

arranging said pulse frequency to detune the pathological oscillatory activity.

17. A method for controlling in a subject epileptic seizures, the subject having an ear, comprising the steps of:

generating voltage pulses;

inducing in a manner responsive to the voltage pulses, at the ear subliminal atmospheric acoustic pulses with a pulse frequency less than 15 Hz; and

initiating said inducing when a seizure precursor is felt by the subject.

* * * * *

US006104107A

United States Patent [19]

Avramenko et al.

[11] **Patent Number: 6,104,107**

[45] **Date of Patent: Aug. 15, 2000**

[54] **METHOD AND APPARATUS FOR SINGLE LINE ELECTRICAL TRANSMISSION**

[75] Inventors: **Stanislav Avramenko**; **Konstantin Avramenko**, both of Moscow, Russian Federation

[73] Assignee: **Uniline Limited**, St. Holier, United Kingdom

[21] Appl. No.: **08/331,658**

[22] PCT Filed: **May 10, 1993**

[86] PCT No.: **PCT/GB93/00960**

§ 371 Date: **Jan. 11, 1995**

§ 102(e) Date: **Jan. 11, 1995**

[87] PCT Pub. No.: **WO93/23907**

PCT Pub. Date: **Nov. 25, 1993**

[30] **Foreign Application Priority Data**

May 8, 1992 [RU] Russian Federation 5036137

[51] **Int. Cl.**^{7} **H02J 1/20**

[52] **U.S. Cl.** .... **307/149**; 331/67; 331/36 C

[58] **Field of Search** .... 307/149, 17; 333/99; 331/36 C, 67, 71, 86, 87, 88, 89, 90, 91, 187

[56] **References Cited**

U.S. PATENT DOCUMENTS

454,622	6/1891	Tesla .
568,176	9/1896	Tesla .
593,138	11/1897	Tesla .

OTHER PUBLICATIONS

Grotz, "Wireless Transmission of Power, An Attempt To Verify Nikola Tesla's 1899 Colorado Springs Expriments, Results Of Research And Experimentation," Proceedings of the 26th IECEC Conference, vol. 4 (1991) pp. 404–409.

"Fachlexikon ABC Physik," VEB Edition Leipzig (1974) p. 1548.

G. Trinkaus, "Tesla—The Lost Inventions," Vantage Press (1988) pp. 13–14.

V.N. Dulin and M.S. Zhuk (editors), *Manual of Radio and Electronic Equipment Components* (Moscow 1977), pp. 69–70.

I. V. Alyamovsky, *Electron Beams and Electron Guns* (Moscow 1966), pp. 13–13–15, 59, and 108.

N. Tesla, "The True Wireless," *Electrical Experimenter* (May 1919).

Y. A. Khramov, "Physicists: A Biographical Reference Book" (Nauka 1983), pp. 8–9, 20–23, 26–27, 80–81, 84–85, 86–95, 102–103, 106–131, 136–139, 142–143, 178–179, 240–241.

M. Faraday, "Experimental Researches in Electricity" (Dover Publications, Inc. 1839), pp. 24–29.

"The Large Soviet Encyclopedia," vol. 26 (Moscow Publishing House 1977).

C. Gillispie, "Dictionary of Scientific Biography," vol. IV, pp. 532–535 (Charles Scribner's Sons 1971).

C. Gillispie, "Dictionary of Scientific Biography," vol. V, pp. 515–517 (Charles Scribner's Sons 1972).

C. Gillispie, "Dictionary of Scientific Biography," vol. IX, pp. 209–213 (Charles Scribner's Sons 1974).

G. Polvani, "Allessandro Volta" (Domus Galilæana 1942), pp. 340–353, 485.

M.I. Radovski, "Galvani and Volta," (1941), pp. 30–31, 58–59.

K.E. Swartz, "The Uncommon Physics of Common Phenomena," vol. 2 (1987), p. 148.

B.N. Rzhonsnitsky, "Nikola Tesla" (Molodaya Gvardiya 1959) pp. 6–7, 116–120.

G.K. Tsverava, "Nikola Tesla" (Nauka 1974) pp. 160–161, 176–177.

J.K. Maxwell, "Selected Works on Electromagnetic Field Theory" (Gosizdat 1952) pp. 252–255, 320–321.

J.J. O'Neill, "Prodical Genius, The life of Nikola Tesla" (Neville Spearman 1968) pp. 70–73, 128–133.

John O'Neill, "Electrical Prometheus" (History of Technology ("Molodaya Gvardiya") 1959).

Primary Examiner—William M. Shoop, Jr.
Assistant Examiner—Kim Lockett
Attorney, Agent, or Firm—Reid & Priest L.L.P.

[57] **ABSTRACT**

This invention relates to the field of electrical technology, and relates particularly to a method for the continuous transformation of electrical energy with its subsequent transmission from an initial source (transformer) to a consuming device, and also to an apparatus for the implementation of this method of transformation and the supplying of power to electrical devices through a transmission line which does not form a closed circuit, ie consists of a single conducting wire. This invention therefore provides a method and associated apparatus for supplying power to an electrical device(s), including generation and subsequent transmission thereof to a receiving device via a transmission line, the method being characterised by the transformation of the electrical energy which is generated into the energy of oscillation of a field of free electrical charges such as the displacement current or longitudinal wave of an electrical field, the density of which charges varies in time, and the transmission of the energy via a transmission line which does not form a closed circuit comprising a single-wire transmission line and, where necessary, its transformation into the electromagnetic energy of conduction currents.

23 Claims, 3 Drawing Sheets

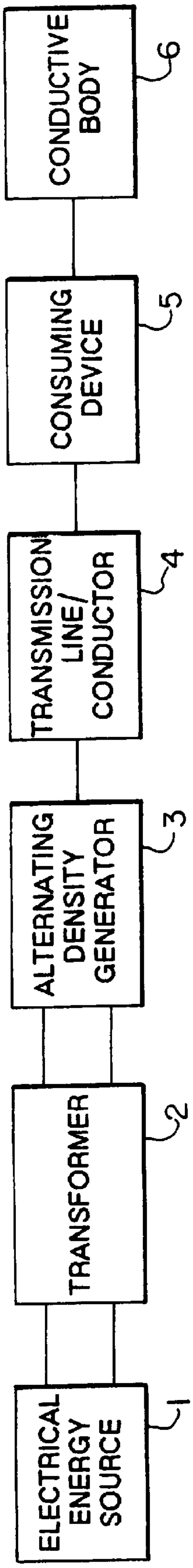
ELECTRICAL ENERGY SOURCE
1
TRANSFORMER
2
ALTERNATING DENSITY GENERATOR
3
TRANSMISSION LINE/ CONDUCTOR
4
CONSUMING DEVICE
5
CONDUCTIVE BODY
6

FIG. 1

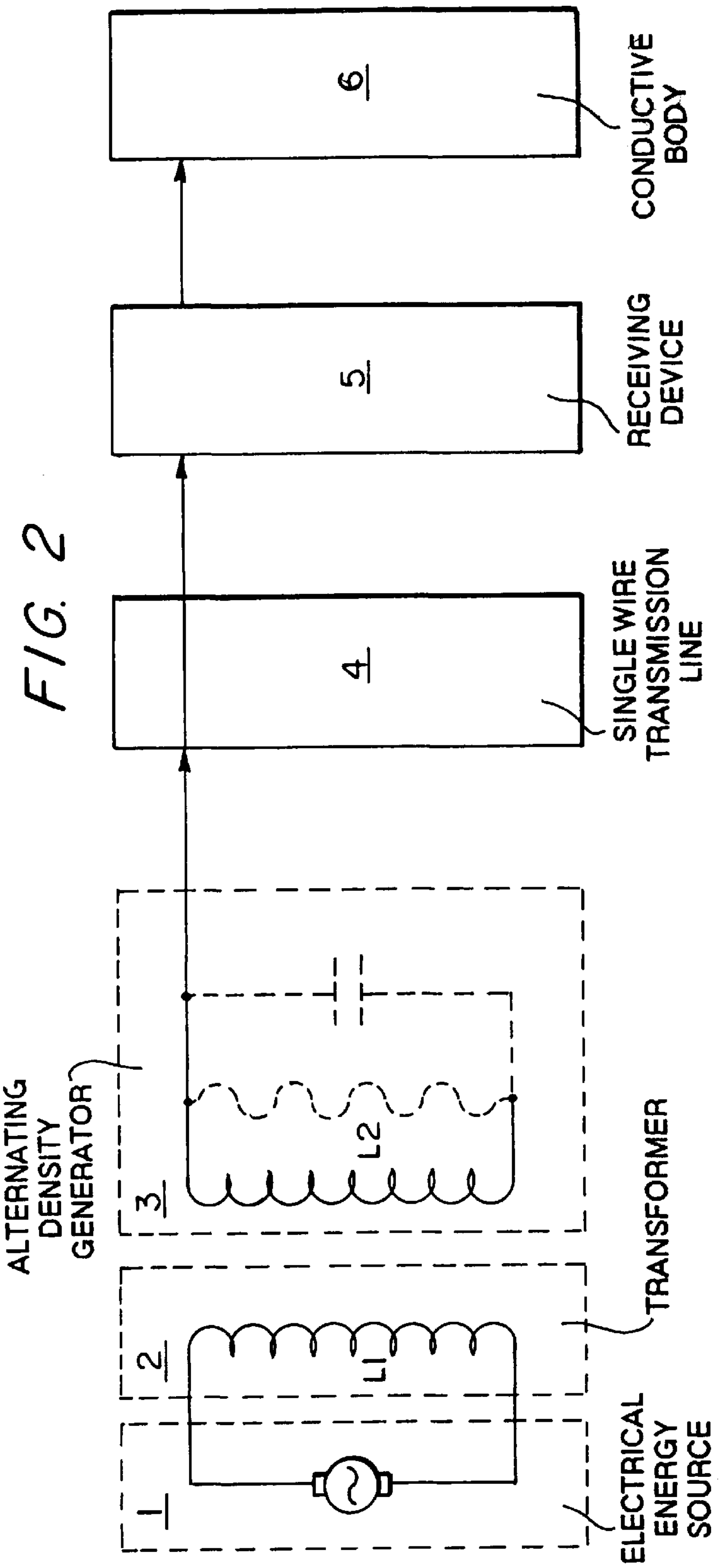
1
ELECTRICAL ENERGY SOURCE
2
L1
TRANSFORMER
3
L2
ALTERNATING DENSITY GENERATOR
4
SINGLE WIRE TRANSMISSION LINE
5
RECEIVING DEVICE
6
CONDUCTIVE BODY

FIG. 2

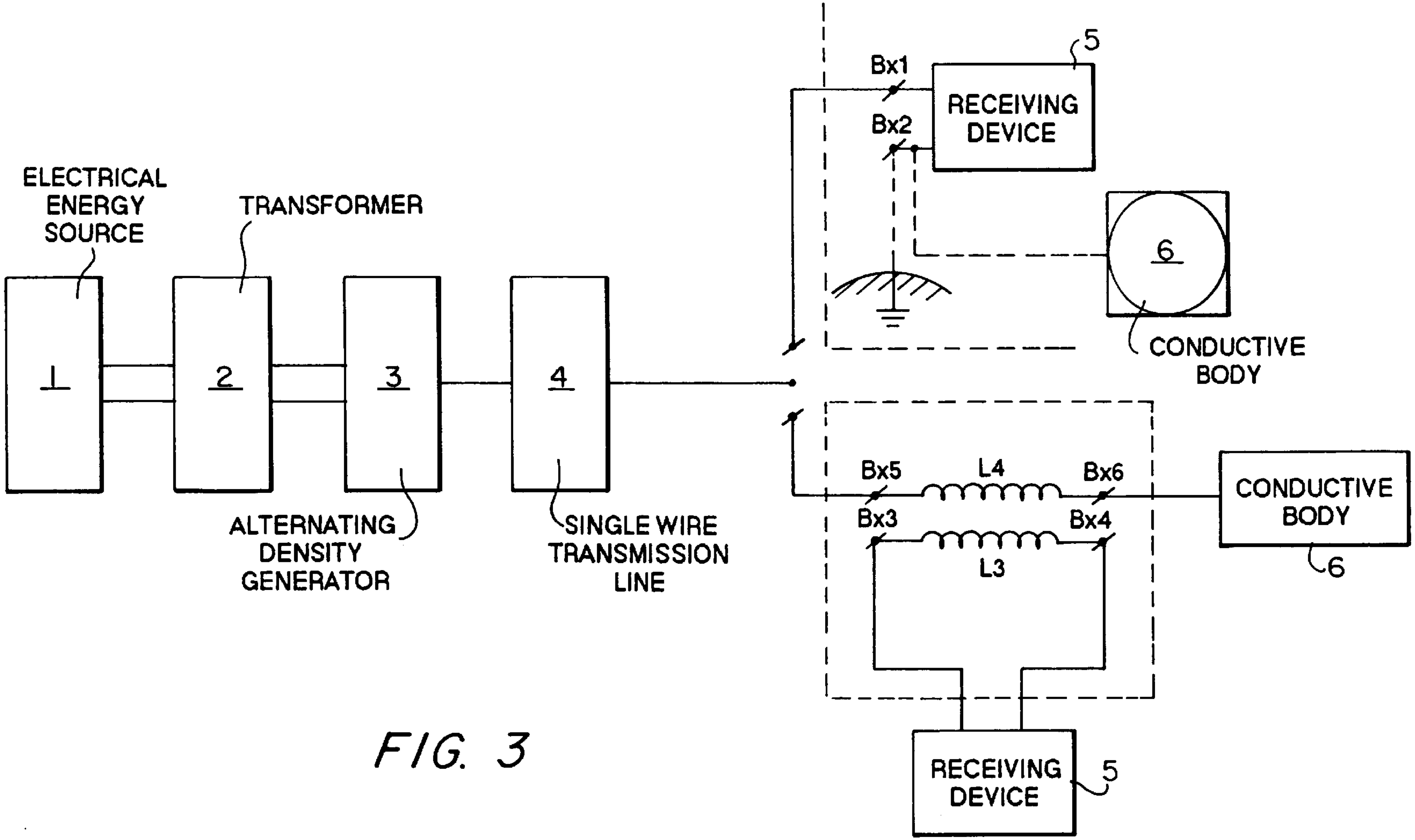
ELECTRICAL ENERGY SOURCE
TRANSFORMER
1
2
3
4
ALTERNATING DENSITY GENERATOR
SINGLE WIRE TRANSMISSION LINE
5
Bx1
Bx2
RECEIVING DEVICE
6
CONDUCTIVE BODY
Bx5
L4
Bx6
Bx3
Bx4
L3
CONDUCTIVE BODY
6
RECEIVING DEVICE
5

FIG. 3

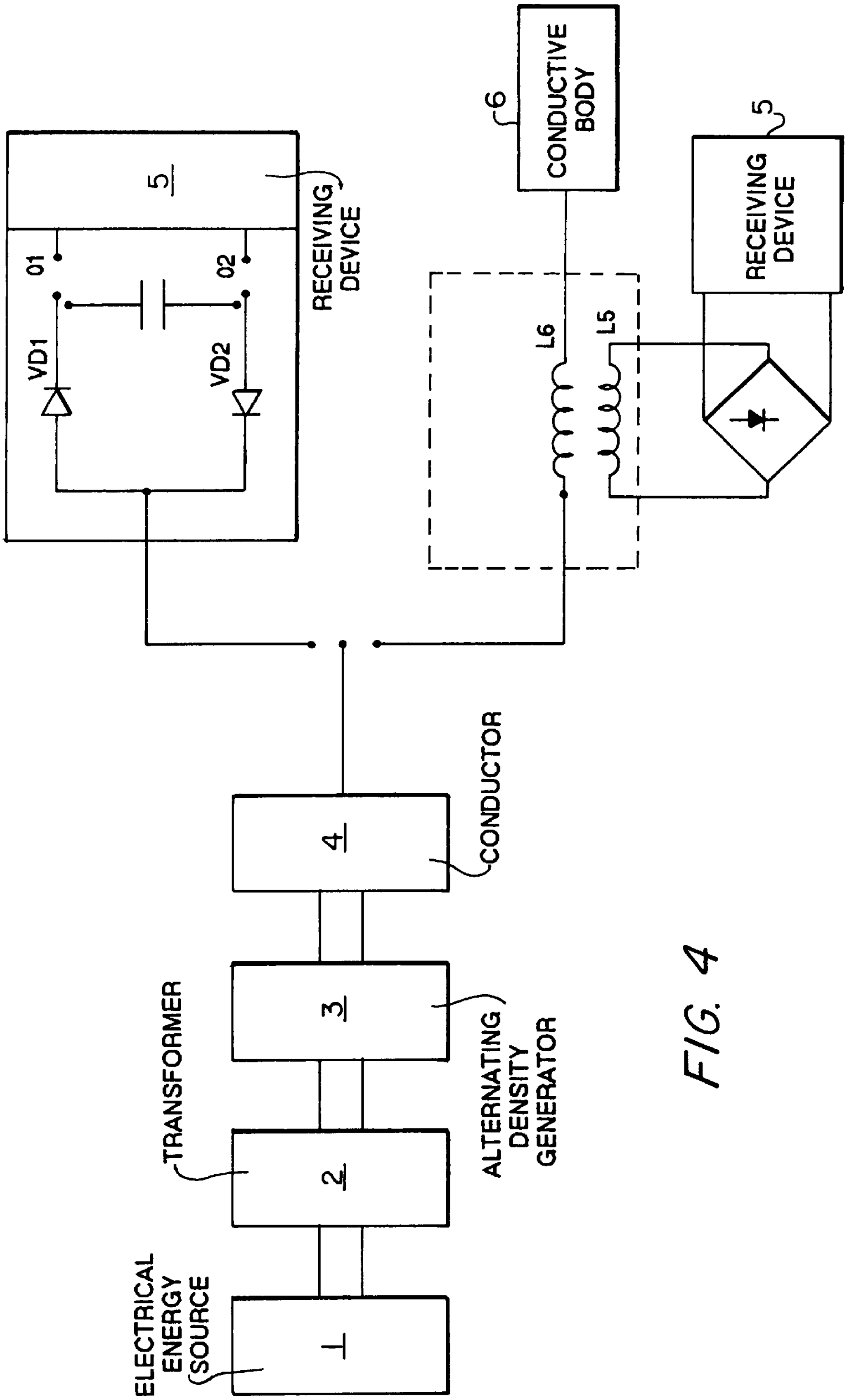
ELECTRICAL ENERGY SOURCE
1
TRANSFORMER
2
3
ALTERNATING DENSITY GENERATOR
4
CONDUCTOR
5
RECEIVING DEVICE
01
02
VD1
VD2
6
CONDUCTIVE BODY
L6
L5
5
RECEIVING DEVICE

FIG. 4

METHOD AND APPARATUS FOR SINGLE LINE ELECTRICAL TRANSMISSION

FIELD OF TECHNOLOGY

This invention relates to the field of electrical technology, and relates particularly to a method for the continuous transformation of electrical energy with fits subsequent transmission from an initial source (transformer) to a consuming device, and also to an apparatus for the implementation of this method of transformation and the supplying of power to electrical devices through a transmission line which does not form a closed circuit, ie consists of a single conducting wire.

BACKGROUND ART

There exists a means for the transmission of electrical energy along a conducting wire which does not form a closed circuit ie, a single conducting wire. It is based on the discovery in 1729 by the English physicist Stephen Grey of the phenomenon of electrical conductivity. The essence of this phenomenon consists in the fact that electricity may be transmitted from one body to another along a metal conductor or a length of yarn, and that the electrical charge is distributed over the surface of the conductor.

See Yu. A. Khramov, *Physicists: A Biographical Reference Book* (Moscow, "Nauka" 1983) (in Russian); and *Dictionary of scientific biography* (New York, Charles Scribener's Sons 1970–1978).

As to closed circuits, a well known means for the supplying of power to electrical devices has existed since the creation in 1799 by the Italian scientist Alessandro Volta of the first source of prolonged electrical current (Volta's column). This means is based on the transmission of electrical energy generated in the initial source to the consuming device directly through a transmission line consisting of two conducting wires which together with the source and the load form a closed circuit.

See Yu. A. Khramov, *Physicists: A Biographical Reference Book,* M. I. Radovsky, *Galvani and Volta* (Moscow-Leningrad 1941) (in Russian); and G. Polvani, *Alessandro Volta* (Pisa 1942).

There exists another important type of closed circuit that represents a means for the supply of power to electrical devices based on the discovery made by M Faraday and J Henry of the phenomenon of electromagnetic induction (published by M Faraday in 1831, and on the invention by P N Yablochkov in 1876 of the electrical transformer. The essence of this means consists in the transformation of the current or the voltage of electrical energy generated in the initial source and the transmission line to the consuming device.

See M. Faraday, *Experimental Research in Electricity* (Moscow-Leningrad, Publishing House of the Academy of Sciences of the USSR 1947–1959) (in Russian); Yu. A. Khramov, *Physicists: A Biographical Reference Book; The Large Soviet Encyclopedia,* 3rd Ed. (Bol 26 Moscow, "Soviet Encyclopedia" 1977) (in Russian).

As to wireless means, there exists a means for the supplying of power to electrical and radio devices by using an electromagnetic field. This means is based an G Herz's experimental proof in 1888 of the existence of electromagnetic waves, the discovery of which was forecast by J Maxwell in 1865. The means essentially consists in the transformation of electrical energy from the initial source into an electromagnetic field which is radiated into space and received by the consuming device.

Dictionary of Scientific Biography. New York, Charles Scribener's Sons, 1970–1978. Yu A Khramov, Physicists: A Biographical Reference Book. Moscow, "Nauka", 1983 (in Russian); J K Maxwell, Selected Works on Electromagnetic Field Theory, Moscow "Gosizdat", 1954 (in Russian).

There also exists an electrical device for the transmission of the energy of free electrical charges from the initial source by means of a non-conducting ribbon. This is the so-called high voltage electrostatic generator, which was designed and constructed by the American physicist Van de Graaf between 1929 and 1933. In this device electrical charges are sprayed from needles under high voltage on to a moving ribbon and transferred to an insulated metal dome, where they accumulate. The charges may be sprayed on to the ribbon and collected from it; the ribbon and the dome may be negatively or positively charged.

See K. E. Swarts, *The Uncommon Physics of Common Phenomena,* Vol. 2, translated from the English by E. I. Butikov and A. S. Kondratiev (Moscow 1987) (in Russian); *Biography;* and Yu. A. Khramov, *Physicists: A Biographical Reference Book.*

It is believed to have been proposed to provide a means for the supply of power to electrical devices via a single-wire transmission line, demonstrated by N Tesla in the late nineteenth century.

See John O'Neill, *Electrical Prometheus,* (Moscow "History of Technology" 1944) (in Russian); B. N. Rzhonsnitsky, *Nikola Tesla* (Moscow "Molodaya Gvardiya" 1959) (in Russian); and G. K. Tsverava, *Nikola Tesla* (Leningrad "Nauka" 1974) (in Russian).

It is believed there is a trace mentioned by John O'Neill of one more distinctive means for electrical transmission but description of this trade did not have details and was not confirmed by any original document.

OBJECTS OF THE INVENTION

An object of the present invention is to provide an increase of the coefficient of efficiency to transmission from an initial source to a consuming device in electrical technology, and also a simultaneous reduction in the metal content of the transmission line.

SUMMARY OF THE INVENTION

According to a first aspect of the present invention there is provided a method of supplying power to one or more electrical devices, including generation and transformation of electrical energy and subsequent transmission thereof to a receiving device via a transmission line, the method being chracterised by the transformation of the electrical energy which is generated into the energy of oscillation of a field of free electrical charges such as a displacement current or longitudinal wave of an electrical field, the density of which charges varies in time, and the transmission of the energy via a transmission line which does not form a closed circuit comprising a single-wire transmission line and, where necessary, its transformation into the electromagnetic energy of conduction currents.

Thus, the invention provides transformation of electrical energy generated in an initial source into energy of oscillation of a field of free electrical charges (the displacement current or longitudinal wave of the electrical field), which energy is transmitted to the consuming device via a conductor of the transmission line which does not form a closed circuit and, where necessary, transformed into the electromagnetic energy of closed circuit conductive currents.

The oscillations of the field of free electrical particles occur either by means of the reciprocating (cyclical) displacement of a concentrated electrical charge in space, or by means of a periodical change in density (and/or polarity) of the free electrical charges on a particular surface (in a particular volume).

According to a second aspect of the present invention there is provided an apparatus for the implementation of the above method of supplying power to one or more electrical devices, the apparatus providing an initial source of electrical energy, a transforming device, a transmission line and a receiving device, the apparatus being characterised by the provision of a variable density generator of free electrical charges as the displacement current or longitudinal wave of an electrical field, an output of which is connected by means of: a conductor of a transmission line which does not form a closed circuit to a receiving device, either directly or via a blocking capacitor, and further to any conductive body possessing an equivalent (natural) capacity adequate to ensure the normal functioning of the receiving device.

Thus, the invention provides a variable (alternating) density generator of free electrical charges, which flow under the influence of coulomb forces along a conductor of the transmission line which does not form a closed circuit to the site of a device which consumes electrical energy.

A possible variant of the generator is a generator at the outlet of which not only the density of the free electrical charges, but also their polarity, may be varied.

The outlet of the generator is connected to a conductor of the transmission line which does not form a closed circuit line either directly or via a blocking capacitor.

In addition, the generator of oscillations of the electrical field of free charges may be constructed in a similar fashion to a generator of displacement current (travelling longitudinal waves of an electrical field), by using a sequential resonance circuit in the form of two interconnected inductors such that the equivalent inductivity of the resonance circuit is provided by their resultant inductivity, and the equivalent capacity of the resonance circuit is provided by the equivalent (natural) capacity of the interconnected inductors.

To supply power to electrical devices which consume alternating current, the output of the conductor of the transmission line which does not form a closed circuit may be connected:

a) to one of the input terminals of the receiving devices, while the device's other input terminal is either grounded or connected to any conductor possessing a natural (equivalent) capacity adequate to provide for the normal working of the receiving (consuming) device.

b) to an accommodating device employing a conversion circuit consisting of two interconnected inductors, such that the receiving device (load) is connected to the two ends of the first inductor, the output of the conductor of the transmission line which does not form a closed circuit is connected to one end of the second inductor, and the other end of the second inductor connected to any conductor with an equivalent (natural) capacity and inductance selected in order to provide for the nominal power consumption of the receiving device (load).

To supply power to devices consuming direct current, the conductor of the transmission line which does not form a closed circuit may be connected to an adjustment circuit in the form of:

c) a diode system, such that the output of the conductor of the transmission line which does not fort a closed circuit is connected to the common point of the anode of the first diode and the cathode of the second diode, while the cathode of the first diode and the anode of the second diode are the output points for connection to the receiving device, either directly or with a capacitor connected in parallel.

d) a transformer circuit consisting of two interconnected inductors such as to rectify alternating current (voltage) directed to the receiving device from the first inductor.

BRIEF DESCRIPTION OF THE DRAWINGS

In order to provide a better understanding of the invention, there follow specific examples of its construction with references to the drawings attached, in which:

FIG. **1** shows a block diagram of a first embodiment of an apparatus according to the present invention;

FIG. **2** shows a schematic diagram of a second embodiment of an apparatus according to the present invention, employing a sequential resonance circuit;

FIG. **3** shows a power supply diagram for receiving devices operating on alternating current;

FIG. **4** shows a power supply diagram for receiving devices operating on direct current.

DESCRIPTION OF FIRST EMBODIMENT

The first embodiment of the invention provides apparatus adapted for use in a method of supplying power to electrical devices, including the generation and transformation of electrical energy with its subsequent transmission to a receiving device via a transmission line, the method being distinguished by the fact that the electrical energy generated is transformed into the energy of oscillation of a field of free electrical charges (the displacement current or longitudinal wave of an electrical field), the density of which charges varies in time, and this energy is transmitted via a conductor of the transmission line which does not form a closed circuit and, where necessary, transformed into the electromagnetic energy of conductive currents.

Referring to FIG. **1**, there is illustrated an apparatus for initial source of electrical energy **1**, a transformer (of current, voltage or frequency) **2**, an alternating density generator of free electrical charges **3**, which charges flow under the influence of coulomb forces along a transmission line or conductor **4**, through a consuming device **5**, to any conductive body **6**, which has an equivalent (natural) capacity sufficient to provide for the normal working of the consuming device **5**.

DESCRIPTION OF SECOND EMBODIMENT

Referring to FIG. **2**, in addition, the apparatus may be constructed on the basis of a generator of displacement current (longitudinal wave of an electrical field), using a sequential resonance circuit (FIG. **2**) in the form of two interconnected inductors L**1** and L**2** such that an equivalent inductivity Leg of the resonance circuit is provided, in the simplest case of idle running, by the resultant inductivity L**1** and L**2**, and the equivalent capacity is provided by the resultant (natural) capacity of the resonance circuit.

To supply power to electrical devices operating on alternating (variable) current, the output of the conductor of the transmission line which does not form a closed circuit **4** is connected either:

to one of the input terminals Bx**1** of the receiving device **5** (FIG. **3**), and the other input terminal Bx**3** of the receiving device is either earthed or connected to any conductive body **6** possessing an equivalent (natural) capacity adequate to ensure the normal working of the receiving device **5**; or

to an accommodating device, employing a transformer circuit (FIG. **3**) consisting of two interconnected inductors L**3** and L**4**, such that the two ends Bx**3** and Bx**4** of the inductor L**3** are connected to the receiving device **5**, while one end of the ends Bx**5** of the second inductor L**4** is connected to the output of the conductor of the transmission line which does not form a closed circuit **4**, and the other end Bx**6** of the inductor L**4** is connected to any conductive body **6**, with an equivalent (natural) capacity selected in order to provide the nominal power consumption of the receiving device **5**.

To supply power to electrical devices operating on direct current, the output of the conductor of the transmission line which does not form a closed circuit **4** is connected to an accommodation circuit in the form of either:

a diode circuit (FIG. **4**) such that output or the conductor **4** is connected to the common point of the anode of the first diode VD**1** and the cathode of the second diode VD**2**, while the cathode of the first diode VD**1** and the anode of the second diode VD**2** are the outlets **01** and **02** to be connected to the receiving device **5**, either directly or with a capacitor connected in parallel;

a transformer circuit (FIG. **4**) employing two interconnected inductors L**5**, L**6**, such that the receiving device **5** is connected to the inductor L**5** via the rectifying circuit.

The method for supply of power to electrical devices and the apparatus for the implementation of this method, according to the invention, possess a high degree of reliability due to the absence of complex electronic or mechanical assemblies. They permit the use of inexpensive mass-produced radio-electronic components and their working cycle is automatically regulated to a high degree.

Use of the invention will make possible a sharp reduction in the costs involved in transmitting electrical energy over long distances, and a sharp reduction in the losses of Joulean heat from transmission lines.

The invention is intended for the creation of a highly efficient means for the transformation and transmission of electrical energy, and also for the creation of both permanent and mobile devices for the transmission of electrical energy from an initial source to a consuming device via a transmission line which does not form a closed circuit, ie a single-wire transmission line.

The invention may be used conjointly with various power-engineering and technological processes which involve the use of super-high voltage electrical and electromagnetic fields, electron beams and super-long wave radio communications, when it will make possible a sharp reduction in the dimensions and weight of equipment as compared with the means traditionally employed.

The proposed apparatus, which may be termed a "monovibrator", may consist of two inductively connected much layered coils in accordance with the scheme of a consecutive resonance circuit. As a rule, the secondary coil consists of up to some tens of thousands of turns of thin isolated wire with the turns wound one to another in many layers on a dialectic core. Disposition of the primary coil in respect to the secondary one doesn't matter much. What does matter is an inductive link which determines potential transmitted from the primary to the secondary coil. The monovibrator may or may not have a ferromagnetic core. The ferromagnetic core influences the width of the working frequency bandwidth—it broadens it.

High working outcoming voltage is the result of a high coefficient of transformation, as the primary coil usually contains a couple of dozens of winds for working frequencies ranging from 1 kHz to a couple of hundred kHz. A preferred working frequency is 5 kHz.

With a sequential resonance circuit including two interconnected inductors having inductance L**1** and L**2**, and an equivalent inductance of the resonance circuit being provided by the resultant inductivity L**1**+L**2**, the load of the monovibrator when running idle acquires capacitance character that means which it is reactive. The magnitude of an active constituent of a monoconductive line with the consecutive resonance is rather low and its incoming resistance is approaching zero. That is why with rather a powerful primary source the consecutive resonance makes it possible to transmit more power through the monoconductive line in case there is an outtake of this power at the other line terminal, which is opposite to the primary source of power.

Reaction of the monoconductive line of any length can always be compensated by regulating the frequency of the primary source (generator, converting devices), thus providing consecutive resonance in the line itself with all the magnitudes of incoming and outgoing characteristics arising therefrom.

Currently there are different schemes of automatic frequency regulation of generators of alternating electromotive power depending on changing inductive-capacitant parameters of conductive lines.

The method and apparatus of the present invention does not have the drawbacks of previously known single line systems.

What is claimed is:

1. A method of supplying power to at least one receiving electrical device, comprising the steps of:

(a) generating electrical energy;

(b) transforming the electrical energy generated in said step (a) into the oscillation energy of a field of free electrical charges, the density of which varies in time;

(c) transmitting the electrical energy transformed in said step (b) to a receiving electrical device via a single-wire transmission line.

2. The method of claim **1**, wherein in said step (b), the oscillation energy of a field of free electrical charges comprises a displacement current of an electrical field.

3. The method of claim **2**, further comprising the step of:

(d) following said step (b) and prior to said step (c), transforming the displacement current into a conduction current for use by an electrical device powered by conduction current.

4. Apparatus for supplying power to at least one electrical device, comprising:

an initial source of electrical energy,

a transformer operatively connected to said initial source to receive electrical energy therefrom; an alternating density generator for transforming the electrical energy from said transformer into an aggregate of free electrical charges, the density of which varies in time;

a single-wire transmission line, said transmission line including a conductor connecting the output of said alternating density generator to a receiving device; and

a conductive body electrically connected to the receiving device, said conductive body having a charge storage capacity of a magnitude adequate to ensure normal functioning of the receiving device.

5. The apparatus of claim **4**, further comprising a blocking capacitor interposed between said transmission line and the receiving device.

6. The apparatus of claim **4**, wherein said alternating density generator comprises a sequential resonance circuit including two interconnected inductors having inductance L1 and L2, and an equivalent inductance of said resonance circuit being provided by the resultant inductivity L1+L2.

7. The apparatus of claim **4**, wherein said alternating density generator transforms the electrical energy from said generator into a displacement current.

8. The apparatus of claim **4**, wherein said alternating density generator transforms the electrical energy from said generator into a longitudinal wave of an electrical field.

9. The apparatus of claim **6**, wherein the electrical device operates on alternating current, and wherein said apparatus further comprises an accommodating device connected to said conductor of said transmission line, said accommodating device including a transformer circuit including at least first and second interconnected inductors, each of said inductors having first and second ends, said first and second ends of said first inductor being connected to the receiving device, said first end of said second inductor connecting said accommodating device to the output of said conductor of said transmission line, and said second end of said second inductor being connected to said conductive body.

10. The apparatus of claim **6**, wherein the electrical device operates on direct current, said apparatus further comprising first and second accommodating devices selectively connectable to said conductor of said transmission line, wherein:

said first accommodating device comprises a diode circuit including first and second diodes, said first diode having an anode and said second diode having a cathode which are commonly connected to the output of said conductor of said transmission line, said first diode having a cathode and said second diode having an anode defining outlets connected to the receiving device; and

said second accommodating device comprises a transformer circuit including first and second interconnected inductors and a rectifying circuit connecting one of said first and second inductors to the receiving device.

11. The apparatus of claim **6**, wherein said first and second interconnected inductors comprise first and second inductively connected coils arranged in accordance with the scheme of a consecutive resonance contour, said second coil comprising a plurality of turns of isolated wire wound around a dielectric body.

12. The apparatus of claim **10**, wherein said diode circuit of said first accommodating device further includes a capacitor connected in parallel with said cathode of said first diode and said anode of said second diode.

13. The apparatus of claim **11**, wherein said first and second inductors have a ferromagnetic core.

14. Apparatus for supplying power to at least one electrical device, comprising:

an initial source of electrical energy;

an alternating density generator for transforming the electrical energy from said source into the oscillation energy of a field of free electrical charges, the density of which varies with time;

a single-wire transmission line, said transmission line including a conductor connecting the output of said alternating density generator to a receiving device; and

a conductive body electrically connected to the receiving device, said conductive body having a charge storage capacity of a magnitude adequate to ensure normal functioning of the receiving device.

15. The apparatus of claim **14**, further comprising a blocking capacitor interposed between said transmission line and the receiving device.

16. The apparatus of claim **14**, wherein said alternating density generator comprises a sequential resonance circuit including two interconnected inductors having inductance L1 and L2, and an equivalent inductance of said resonance circuit being provided by the resultant inductivity L1+L2.

17. The apparatus of claim **16**, wherein said alternating density generator transforms the electrical energy from said generator into a displacement current.

18. The apparatus of claim **16**, wherein said alternating density generator transforms the electrical energy from said generator into a longitudinal wave of an electrical field.

19. The apparatus of claim **16**, wherein the electrical device operates on alternating current, and wherein said apparatus further comprises an accommodating device connected to said conductor of said transmission line, said accommodating device including a transformer circuit including at least first and second interconnected inductors, each of said inductors having first and second ends, said first and second ends of said first inductor being connected to the receiving device, said first end of said second inductor connecting said accommodating device to the output of said conductor of said transmission line, and said second end of said second inductor being connected to said conductive body.

20. The apparatus of claim **16**, wherein the electrical device operates on direct current, said apparatus further comprising first and second accommodating devices selectively connectable to said conductor of said transmission line, wherein:

said first accommodating device comprises a diode circuit including first and second diodes, said first diode having an anode and said second diode having a cathode which are commonly connected to the output of said conductor of said transmission line, said first diode having a cathode and said second diode having an anode defining outlets connected to the receiving device; and

said second accommodating device comprises a transformer circuit including first and second interconnected inductors and a rectifying circuit connecting one of said first and second inductors to the receiving device.

21. The apparatus of claim **20**, wherein said diode circuit of said first accommodating device further includes a capacitor connected in parallel with said cathode of said first diode and said anode of said second diode.

22. The apparatus of claim **16**, wherein said first and second interconnected inductors comprise first and second inductively connected coils arranged in accordance with the scheme of a consecutive resonance contour, said second coil comprising a plurality of turns of isolated wire wound around a dielectric body.

23. The apparatus of claim **22**, wherein said first and second inductors have a ferromagnetic core.

* * * * *

US007053576B2

(12) **United States Patent**
Correa et al.

(10) **Patent No.: US 7,053,576 B2**
(45) **Date of Patent: May 30, 2006**

(54) **ENERGY CONVERSION SYSTEMS**

(76) Inventors: **Paulo N. Correa**, 42 Rockview Gardens, Concord, Ontario (CA) L4K 2J6; **Alexandra N. Correa**, 42 Rockview Gardens, Concord, Ontario (CA) L4K 2J6

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **10/270,154**

(22) Filed: **Oct. 15, 2002**

(65) **Prior Publication Data**

US 2006/0082334 A1 Apr. 20, 2006

Related U.S. Application Data

(63) Continuation of application No. 09/907,823, filed on Jul. 19, 2001, now abandoned.

(51) **Int. Cl.**
H05B 31/48 (2006.01)

(52) **U.S. Cl.** **318/558**; 315/111.21

(58) **Field of Classification Search** 310/308, 310/309; 318/558; 315/111.01, 111.21, 315/111.31

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

433,702	A		8/1890	Telsa	
645,576	A		3/1900	Telsa	
3,233,157	A	*	2/1966	Stockman	318/116
4,095,118	A	*	6/1978	Rathbun	290/2
4,213,797	A	*	7/1980	Sher	136/254
4,473,736	A	*	9/1984	Bloyet et al.	219/121.48
5,063,330	A	*	11/1991	Leprince et al.	315/111.21
5,416,391	A		5/1995	Correa et al.	
5,449,989	A		9/1995	Correa et al.	
6,271,614	B1	*	8/2001	Arnold	310/233

OTHER PUBLICATIONS

Kuhn, T.S. (1978) "Black-body Theory and the Quantum Discontinuity, 1898-1912", The University of Chicago Press, pp. 246-249, 289-290.

Wang, LJ et al (2000) "Gain-assisted superluminal light propagation", Nature, 406:277.

Martin, Thomas Commerfold (1894) "The Inventions, Researches and Writlngs of Nikola Tesla", The Electrical Engineer, New York, p. 68.

Tesla, Nikola (1956) "Lectures, Patents, Articles", Nikola Tesla Museum, Beograd, Yugolsavia, L-70-71 & L-130-132 (Figure 16.II).

(Continued)

Primary Examiner—Bentsu Ro
(74) *Attorney, Agent, or Firm*—Ridout & Maybee LLP

(57) **ABSTRACT**

This invention relates to apparatus for the conversion of massfree energy into electrical or kinetic energy, which uses in its preferred form a transmitter and a receiver both incorporating Tesla coils, the distal ends of whose secondary windings are co-resonant and connected to plates of a chamber, preferably evacuated or filled with water, such that energy radiated by the transmitter may be picked up by the receiver, the receiver preferably further including a pulsed plasma reactor driven by the receiver coil and a split phase motor driven by the reactor. Preferably the reactor operates in pulsed abnormal gas discharge mode, and the motor is an inertially damped drag motor. The invention also extends to apparatus in which an otherwise driven plasma reactor operating in pulsed abnormal gas discharge mode in turn used to drive an inertially damped drag motor.

16 Claims, 11 Drawing Sheets

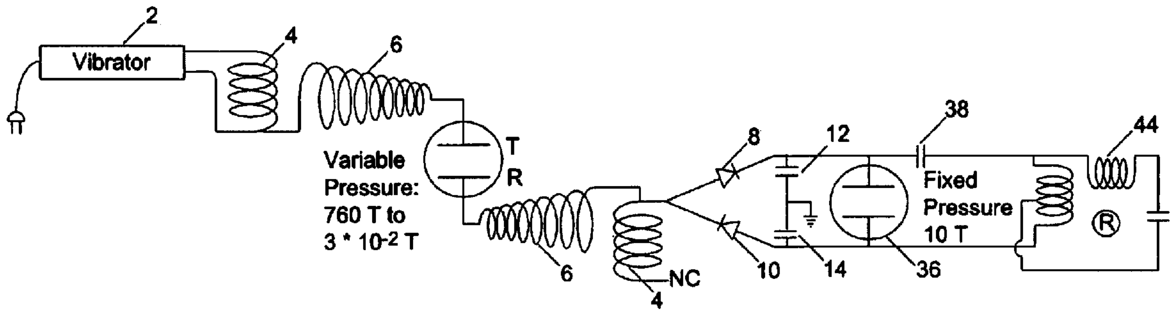

OTHER PUBLICATIONS

Reich, Wilhelm (1949) "A Motor Force in Orgone Energy. Preliminary Communications", Orgone Energy Bulletin, vol. 1, #1, pp. 7-11.

Laithwaite, Eric (1957) "Induction Machines for Special Purposes". p. 322-323.

Halliday, David & Resnick, Robert (1978)"Physics", Parts I and II Combined Third Edition, John Wiley and Sons, p. 488-489.

* cited by examiner

FIG. 1

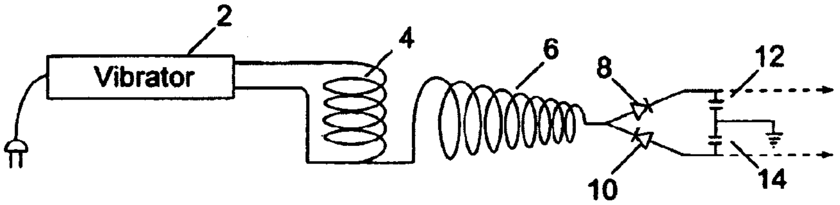
2
Vibrator
4
6
8
12
10
14

FIG. 2

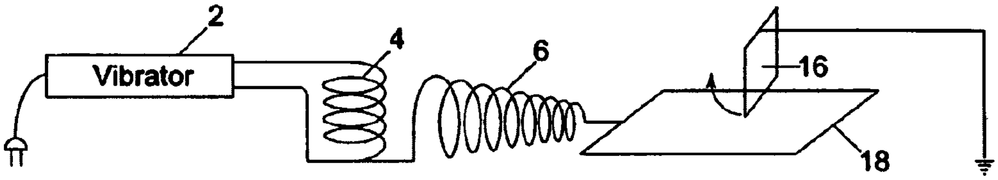
2
Vibrator
4
6
16
18

FIG. 3

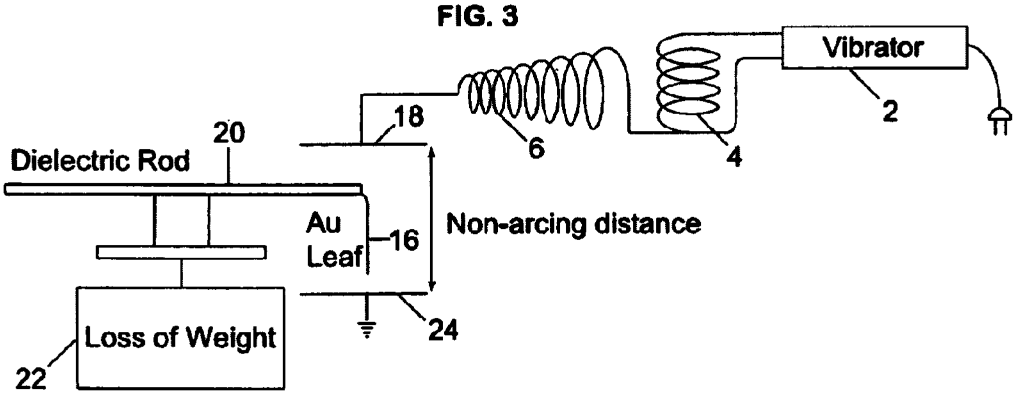
Vibrator
2
4
6
18
20
Dielectric Rod
Au
Leaf
16
Non-arcing distance
24
Loss of Weight
22

FIG. 4

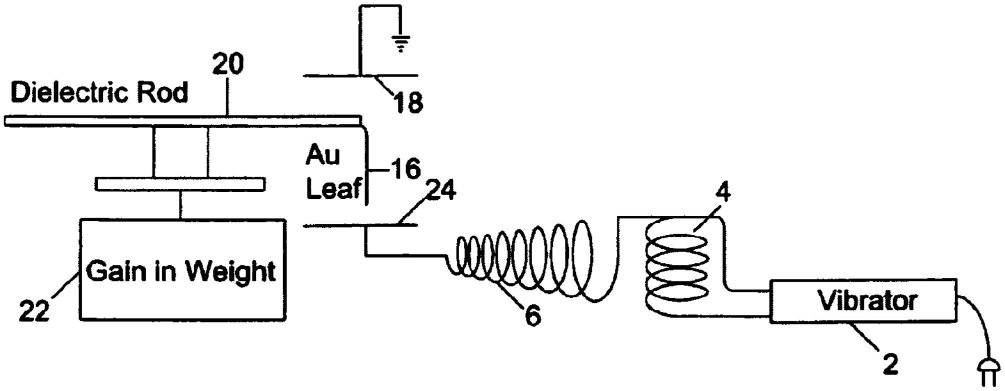
20
Dielectric Rod
18
Au
Leaf
16
24
4
Gain in Weight
22
6
Vibrator
2

FIG. 5

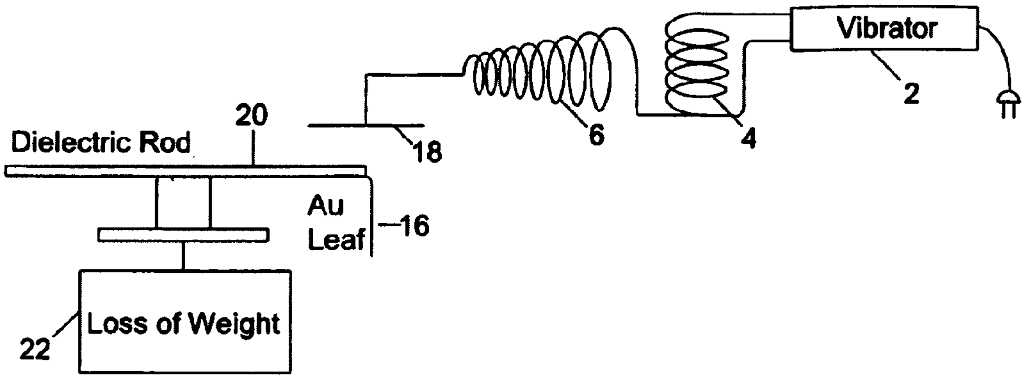
Vibrator
2
4
6
18
20
Dielectric Rod
Au
Leaf
16
Loss of Weight
22

FIG. 6

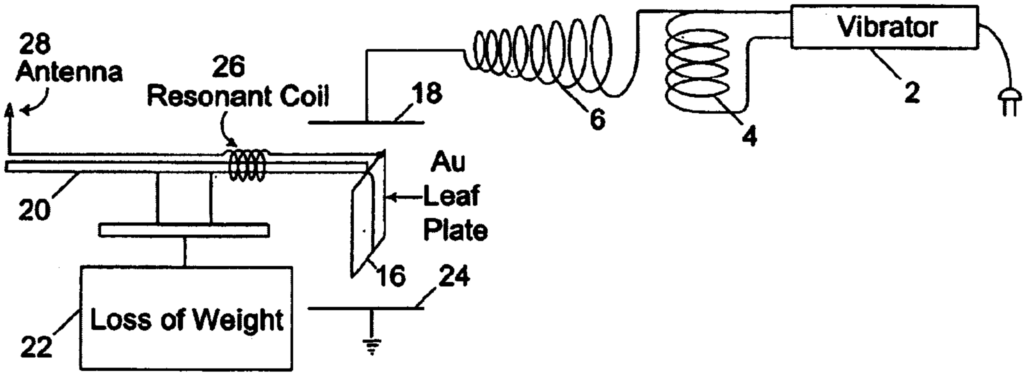
Vibrator
2
4
6
18
28
Antenna
26
Resonant Coil
Au
Leaf
Plate
20
16
24
Loss of Weight
22

FIG. 7

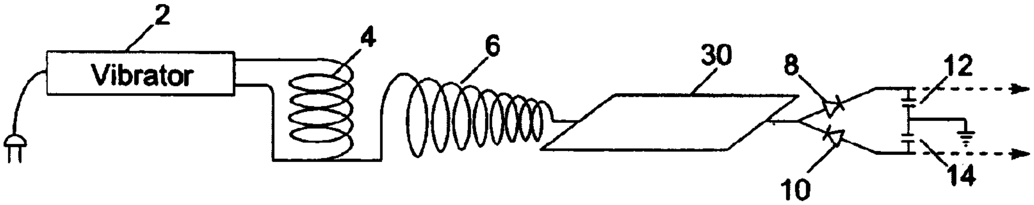
2
Vibrator
4
6
30
8
12
10
14

FIG. 8

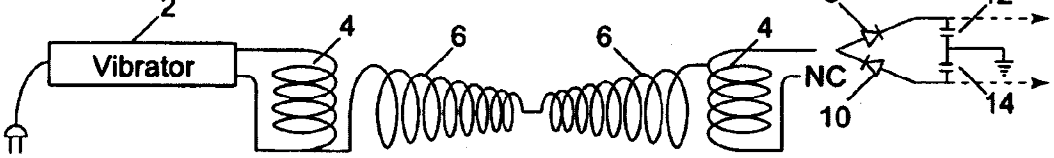
2
Vibrator
4
6
6
4
8
12
NC
10
14

FIG. 9

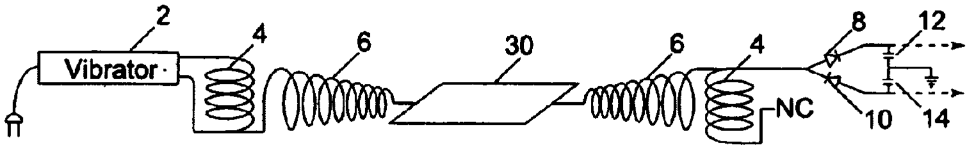
2
Vibrator
4
6
30
6
4
NC
8
12
10
14

FIG. 10

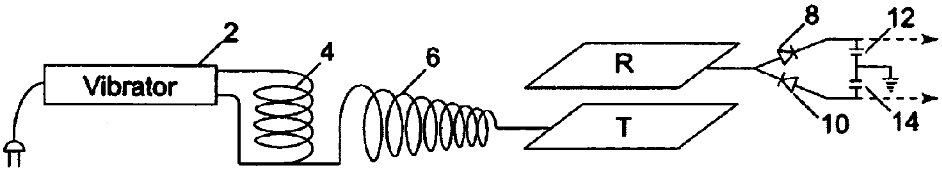
2
Vibrator
4
6
R
T
8
12
10
14

FIG. 11

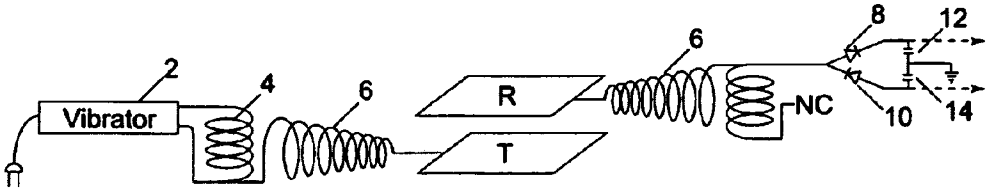
2
Vibrator
4
6
R
T
6
NC
8
12
10
14

FIG. 12

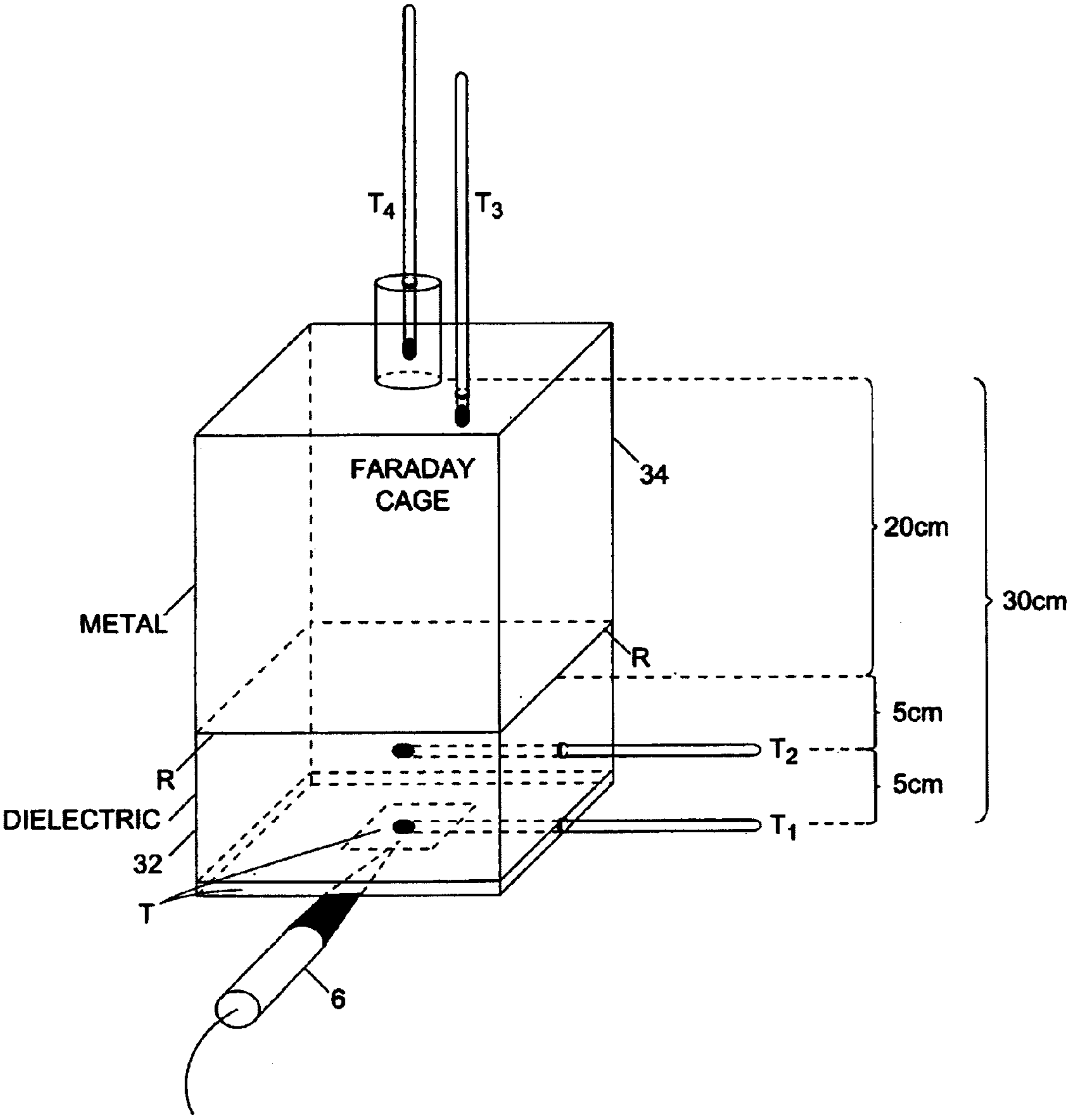
T4
T3
FARADAY CAGE
34
20cm
30cm
METAL
R
5cm
T2
R
DIELECTRIC
5cm
T1
32
T
6

FIG. 13

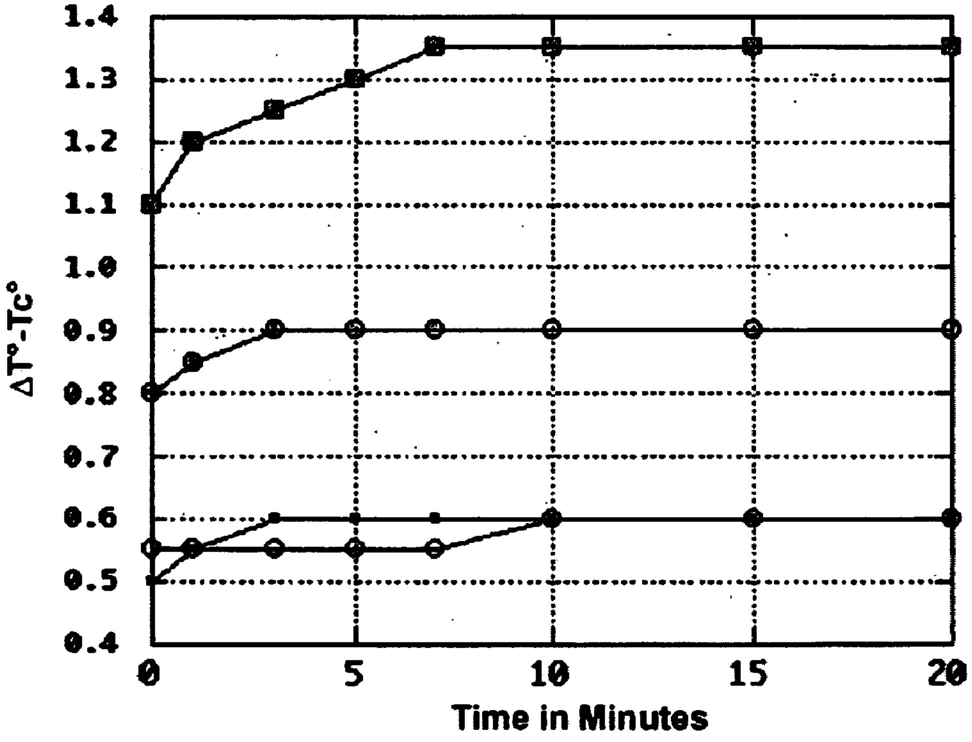

- ▪ Above cage (T_4)
- ● R cage (T_3)
- ○ T/R cavity (T_2)
- ▪ T plate (T_1)

FIG. 14

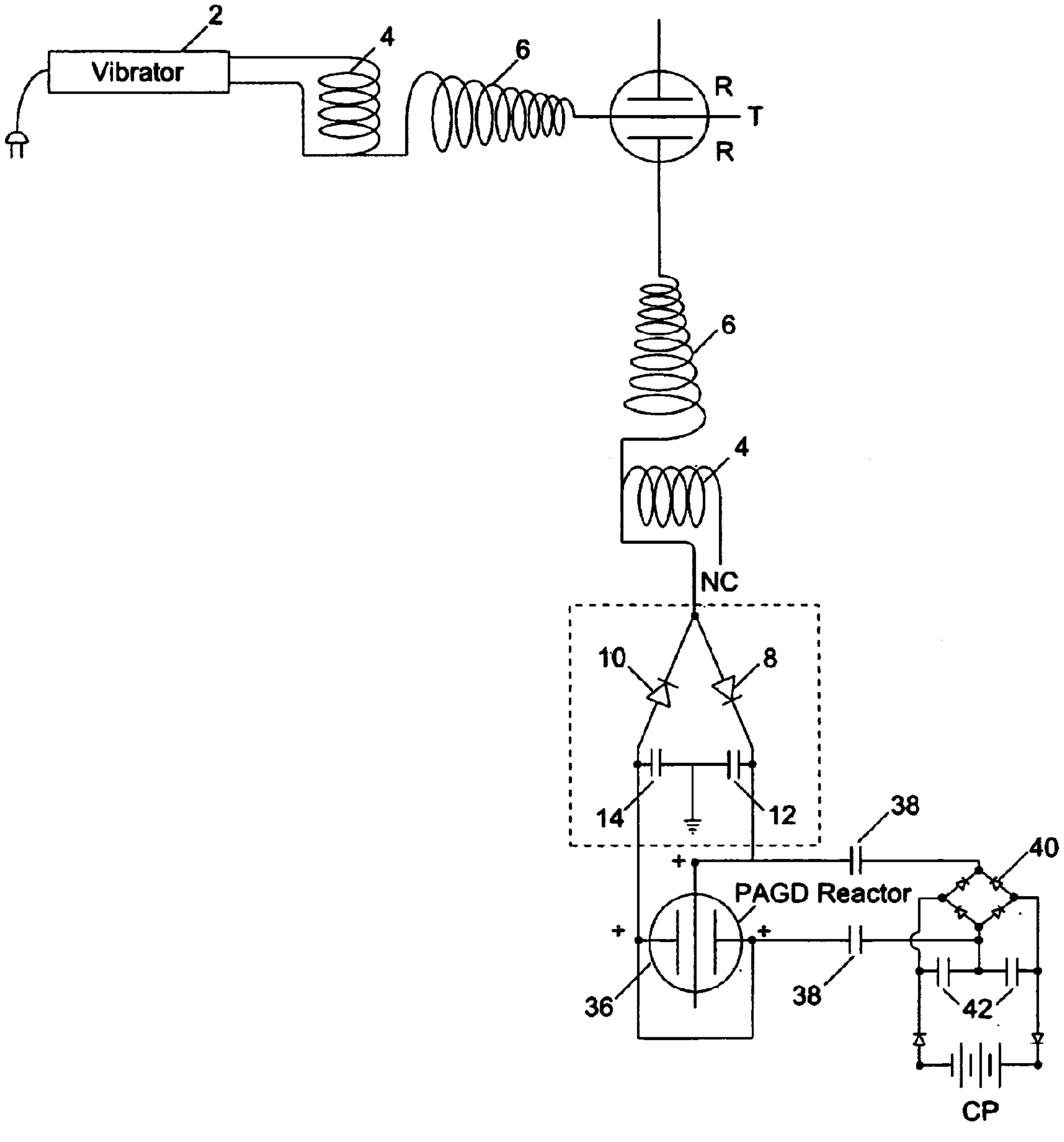
2
Vibrator
4
6
R
T
R
6
4
NC
10
8
14
12
38
40
+
PAGD Reactor
+
+
36
38
42
CP

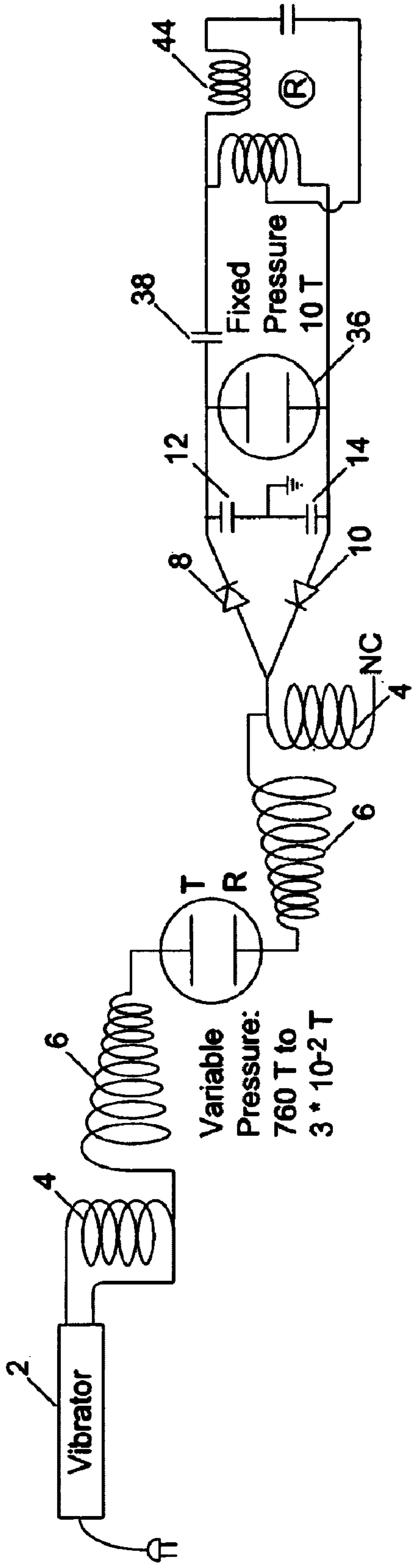
44
®
38
Fixed
Pressure
10 T
36
12
14
8
10
NC
4
6
T
R
Variable
Pressure:
760 T to
3 * 10-2 T
6
4
2
Vibrator

FIG. 15

FIG. 16

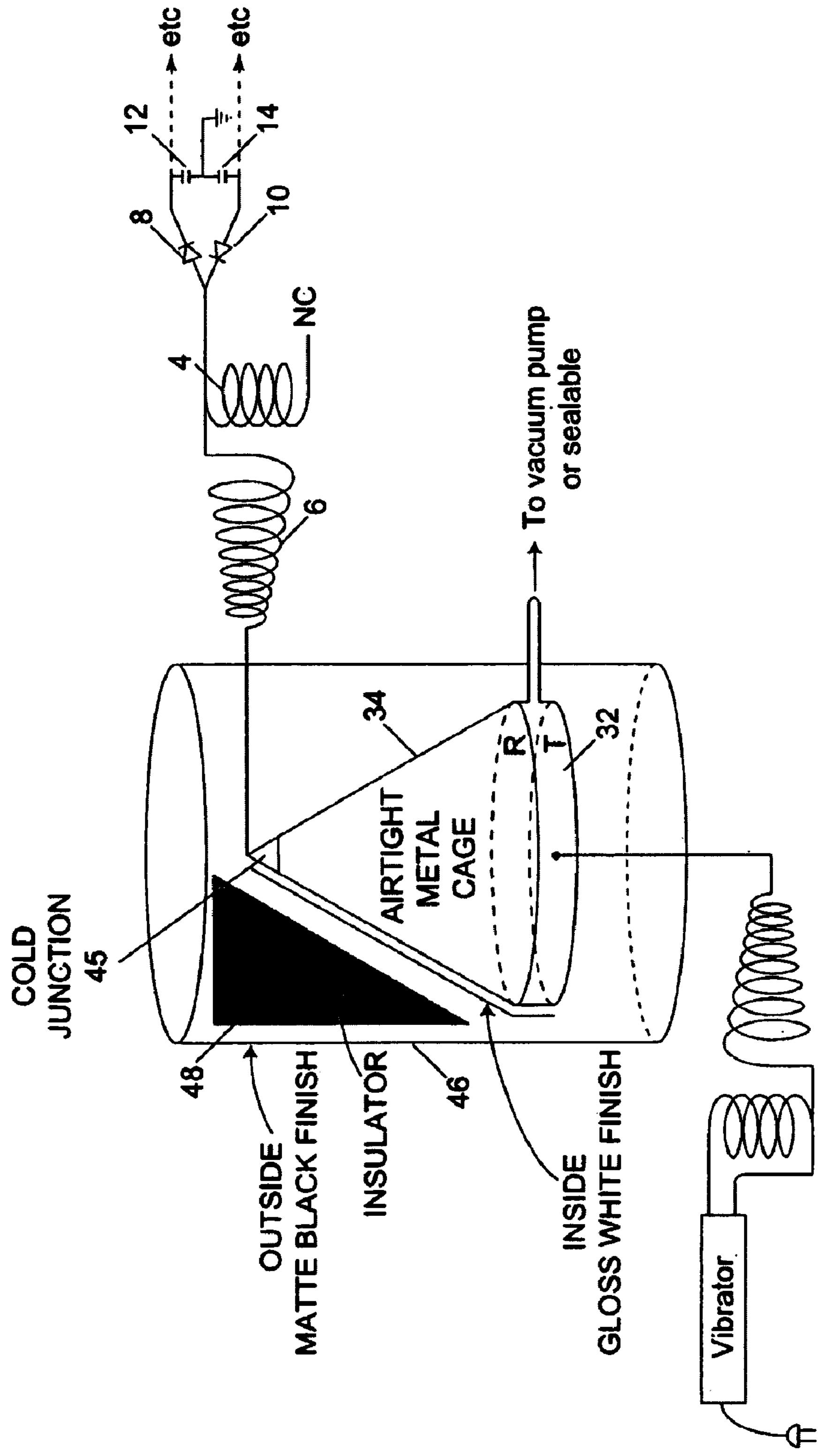
etc
etc
12
14
8
10
NC
4
6
To vacuum pump
or sealable
34
R
T
32
AIRTIGHT
METAL
CAGE
COLD
JUNCTION
45
48
46
OUTSIDE
MATTE BLACK FINISH
INSULATOR
INSIDE
GLOSS WHITE FINISH
Vibrator

FIG. 17

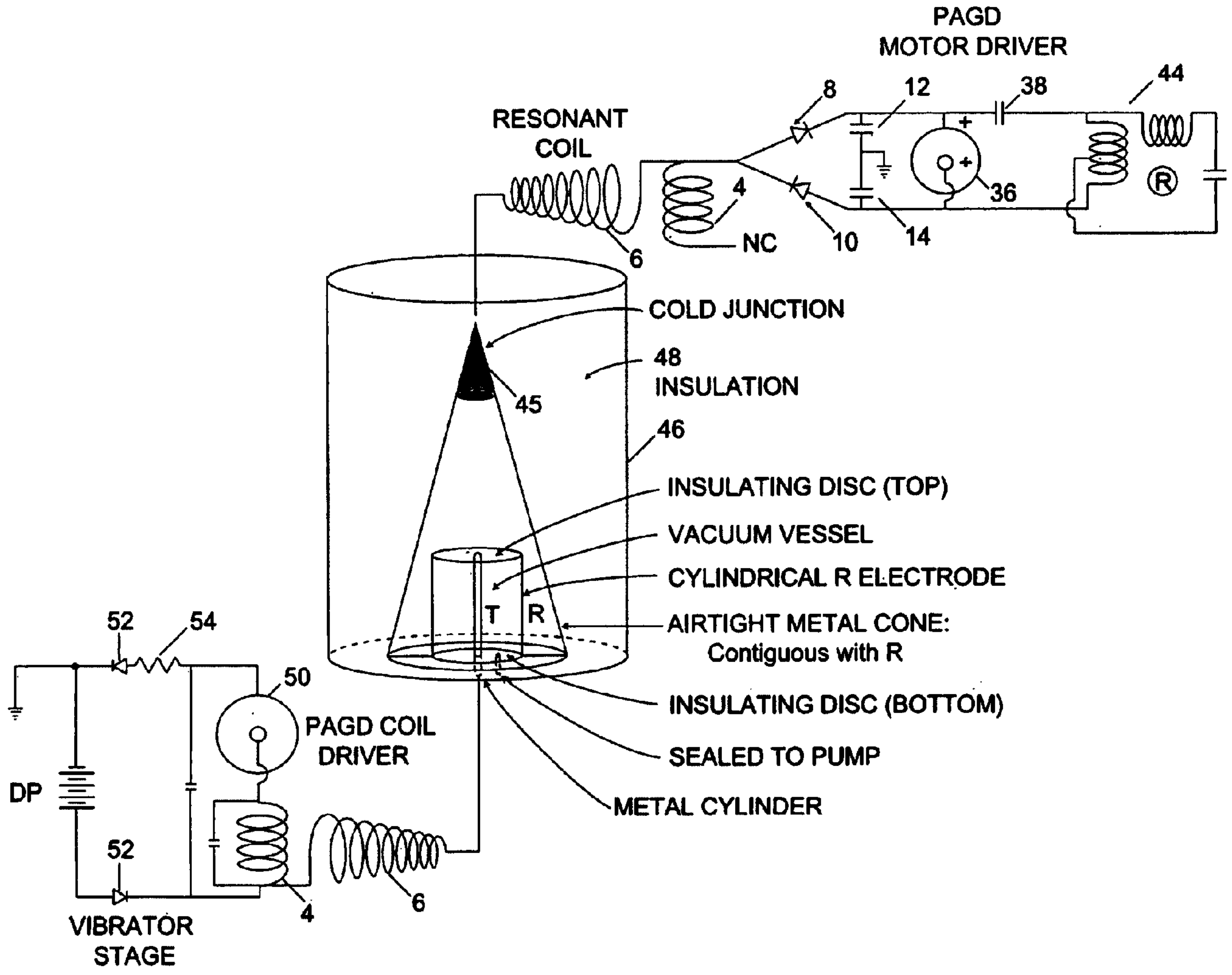
PAGD
MOTOR DRIVER
RESONANT
COIL
8
12
38
44
4
36
NC
10
14
6
COLD JUNCTION
48
INSULATION
45
46
INSULATING DISC (TOP)
VACUUM VESSEL
CYLINDRICAL R ELECTRODE
T
R
AIRTIGHT METAL CONE:
Contiguous with R
INSULATING DISC (BOTTOM)
SEALED TO PUMP
METAL CYLINDER
52
54
50
PAGD COIL
DRIVER
DP
52
VIBRATOR
STAGE
4
6

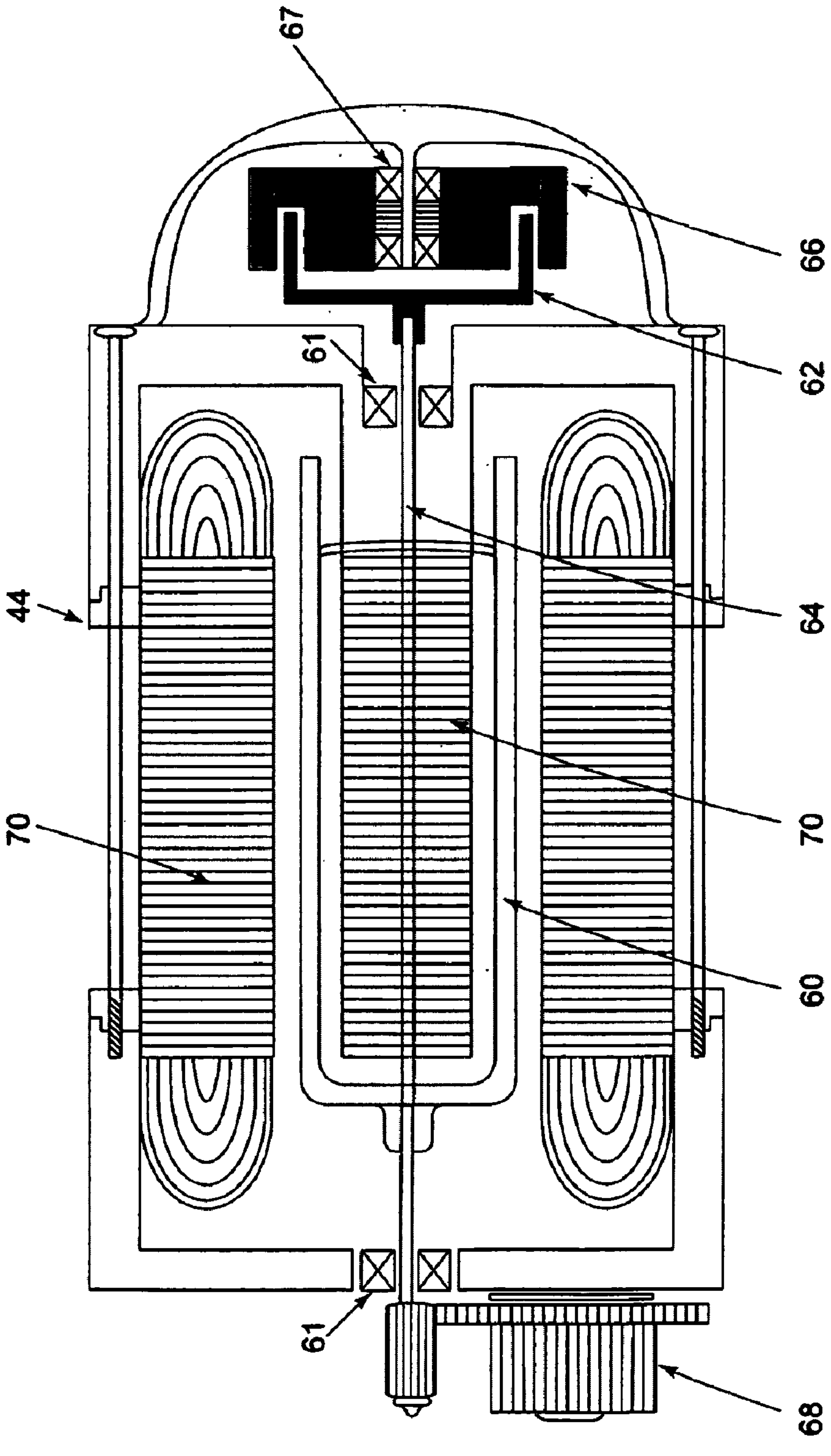
67
66
62
61
64
44
70
70
60
61
68

FIG. 18

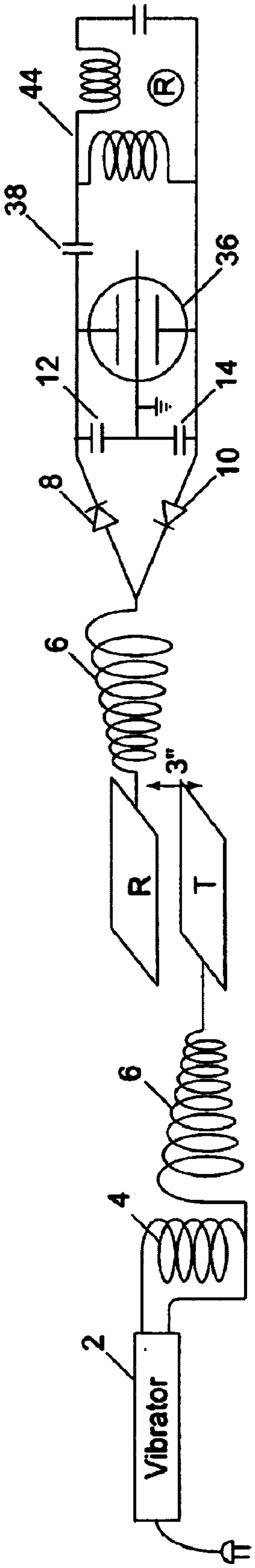
2
Vibrator
4
6
6
R
T
3"
8
10
12
14
36
38
44
Ⓡ

FIG. 19

ENERGY CONVERSION SYSTEMS

This is a continuation of application Ser. No. 09/907,823, filed Jul. 19, 2001 now abandoned.

FIELD OF THE INVENTION

This invention relates to systems for the conversion of energy, inter alia in the form of what we will refer to for convenience as Tesla waves (see below), to conventional electrical energy.

BACKGROUND OF THE INVENTION

Energy converters that are fed by local or environmental energy are usually explained by taking recourse to the notion that they convert zero point electromagnetic radiation (ZPE) to electric energy. The ZPE theories have gained a life of their own, as T. Kuhn has pointed out (in his “Black Body Theory and the Quantum”), after emerging from Planck’s second theory, specifically from the term $\frac{1}{2} h\upsilon$ in the new formula for oscillator energy. In 1913, Einstein and Stern suggested that motional frequencies contributing to specific heat fell into two categories—those that were independent of temperature and those that were not (e.g. rotational energy), leading them to conclude that zero-point energy on the order of $\frac{1}{2} h\upsilon$ was most likely. In the second part of their paper, however, they provided a derivation of Planck’s Law without taking recourse to discontinuity, by assuming that the value of the ZPE was simply $h\upsilon$. It is worth noting that Einstein had already in 1905 (“Erzeugung und Verwandlung des Lichtes betreffenden heuristichen Gesichtspunkt”, Ann. d. Phys, 17, 132) framed the problem of discontinuity, even if only heuristically, as one of placing limits upon the infinite energy of the vacuum state raised by the Rayleigh-Jeans dispersion law. According to Einstein, the Rayleigh-Jeans law would result in an impossibility, the existence of infinite energy in the radiation field, and this was precisely incompatible with Planck’s discovery—which suggested instead that at high frequencies the entropy of waves was replaced by the entropy of particles. Einstein, therefore, could only hope for a stochastic validation of Maxwell’s equations at high frequencies “by supposing that electromagnetic theory yields correct time-average values of field quantities”, and went on to assert that the vibration-energy of high frequency resonators is exclusively discontinuous (integral multiples of $h\upsilon$).

Since then, ZPE theories have gone on a course independent from Planck’s second theory. The more recent root of modern ZPE theories stems from the work of H. Casimir who, in 1948, apparently showed the existence of a force acting between two uncharged parallel plates. Fundamentally the Casimir effect is predicated upon the existence of a background field of energy permeating even the ‘vacuum’, which exerts a radiation pressure, homogeneously and from all directions in space, on every body bathed in it. Given two bodies or particles in proximity, they shield one another from this background radiation spectrum along the axis (i.e. the shortest distance) of their coupling, such that the radiation pressure on the facing surfaces of the two objects would be less than the radiation pressure experienced by all other surfaces and coming from all other directions in space. Under these conditions, the two objects are effectively pushed towards one another as if by an attractive force. As the distance separating the two objects diminishes, the force pushing them together increases until they collapse one onto the other. In this sense, the Casimir effect would be the macroscopic analogy of the microscopic van der Waals forces of attraction responsible for such dipole-dipole interactions as hydrogen bonding. However, it is worth noting that the van der Waals force is said to tend to establish its normal radius, or the optimal distance between dipoles, as the distance where the greatest attractive force is exerted, beyond which the van der Waals forces of nuclear and electronic repulsion overtake the attraction force.

Subsequently, another Dutch physicist, M. Sparnaay, demonstrated that the Casimir force did not arise from thermal radiation and, in 1958, went on to attribute this force to the differential of radiation pressure between the ZPE radiation from the vacuum state surrounding the plates and the ZPE radiation present in the space between them. Sparnaay’s proposal is that a classical, nonquantal, isotropic and ubiquitous electromagnetic zero-point energy exists in the vacuum, and even at a temperature of absolute zero. It is further assumed that since the ZPE radiation is invariant with respect to the Lorentz transformations, it obeys the rule that the intensity of its radiation is proportional to the cube of the frequency, resulting in an infinite energy density for its radiation spectrum.

What appeared to be the virtue of this reformulated theory was the notion that the vacuum no longer figured as pure space empty of energy, but rather as a space exposed to constantly fluctuating ‘fields of electromagnetic energy’.

Puthoff has utilized the isomorphism between van der Waals and Casimir forces to put forth the zero-point (ZP) energy theory of gravity, based on the interpretation that the virtual electromagnetic ZP field spectrum predicted by quantum electrodynamics (QED) is functionally equivalent to an actual vacuum state defined as a background of classical or Maxwellian electromagnetic radiation of random phases, and thus can be treated by stochastic electrodynamics (SED). Whereas in QED, the quanta are taken as virtual entities and the infinite energy of the vacuum has no physical reality, for SED, the ZPE spectrum results from the distortion of a real physical field and does not require particle creation. Gravity then, could be seen as but the macroscopic manifestation of the Casimir force.

We do not dispute the fact that even in space absent matter there is radiant energy present which is not of a thermal nature. But we claim that this energy is not electromagnetic, nor is its energy spectrum infinite. That this is so stems not just from our opinion that it is high time that Einstein’s heuristic hypothesis should be taken as literally factual—in the dual sense that all electromagnetic energy is photon energy and all photons are local productions, but above all from the fact that it is apparent, from the experiments of Wang and his colleagues (Wang, Li, Kuzmich, A & Dogariu, A. “Gain-assisted superluminal light propagation”, Nature 406; #6793; 277), that the photon stimulus can propagate at supraluminal speeds and lies therefore well outside of any scope of electromagnetic theory, be this Maxwell’s classical approach taken up by ZPE theories, or Einstein’s special relativistic phenomenology of Maxwell’s theory. The fact is that if the light stimulus can propagate at speeds greater than those of light, then what propagates is not light at all, and thus not energy configured electromagnetically. Light is solely a local production of photons in response to the propagation of a stimulus that itself is not electromagnetic.

It is critical to understand that the implication from this that—aside from local electromagnetic radiation and from thermal radiation associated with the motions of molecules (thermo-mechanical energy), there is at least another form of energy radiation which is everywhere present, even in space absent matter. Undoubtedly, it is that energy that prevents any attainment of absolute zero, for any possible local outpumping of heat is matched by an immediate local conversion of some of this energy into a minimum thermal radiation required by the manifolds of Space and Time. And undoubtedly also this radiation is ubiquitous and not subject to relativistic transformations (i.e. Lorentz invariant). What it is not, is electromagnetic radiation consisting of randomistic phases of transverse waves.

To understand this properly, one must summarize the differences from existing ZPE theories—and all these differences come down to the fact that this energy which is neither electromagnetic nor thermal per se (and is certainly not merely thermo-mechanical), has nevertheless identifiable characteristics both distributed across subtypes or variants and as well common to all of them.

Essentially the first subtype or variant consists of longitudinal massfree waves that deploy electric energy. They could well be called Tesla waves, since Tesla-type transformers can indeed be shown experimentally to radiate massfree electric energy, in the form of longitudinal magnetic and electric waves having properties not reduceable to photon energy or to 'electromagnetic waves', and having speeds of displacement that can be much greater than the limit c for all strictly electromagnetic interactions.

One may well denote the second subtype by the designation of massfree thermal radiation, since it contributes to temperature changes—and, as obviously indicated by the impossibility of reaching an absolute zero of temperature, this contribution occurs independently of the presence of matter, or mass-energy, in Space. In other words, not all thermal radiation can be reduced to vibration, rotation and translation (drift motion) of molecules, i.e. to thermomechanical energy, because the properties of pressure and volume that determine temperature and affect matter, appear indeed to a great extent to be independent from matter, a fact which itself is responsible for the observed catastrophic and unexpected phase changes of matter and has required to this day the insufficient explanation offered semi-empirically by the Van der Waals Force Law.

Finally the third subtype may be designated latent massfree energy radiation—since it deploys neither charge, nor thermal or baroscopic effects, and yet it is responsible for 'true latent heat' or for the 'intrinsic potential energy' of a molecule. It is also responsible for the kinetoregenerative phenomenon whereby an electroscope performs a variable charge-mediated work against the local gravitational field.

The common characteristic of all three subtypes of massfree energy radiation is that they share the same nonclassical fine structure, written as follows for any energy unit, where c is any speed of light wave function, and the wavelength λ and wave function W are interconnected as a function of the physical quality of the energy field under consideration:

$$E=\lambda cW$$

In the instance of longitudinal electric radiation, this takes on the directly quantizable form:

$$E=(\lambda_q c)W_v=p_e W_v=(h/\lambda_x)W_v=\int=q_v$$

where W_v is the voltage-equivalent wave function corresponding to V, P_e constitutes the linear momentum corresponding to the conventional q or e, h is the Planck constant, λ_x the Duane-Hunt constant expressed as a wavelength, λ_q is a wavelength constant; and the sign $=\int=$ signifies exact equality between an expression in the conventional dimensions of length, mass and time, and an expression in length and time dimensions alone.

In the instance of massfree thermal radiation (contributing to temperature changes), the transformation obeys Boltzmann's rule (k is now Boltzmann's constant and T is Kelvin-scale temperature):

$$E=\lambda_{n1}cW_{n1}=\lambda_{n1}(\lambda_V\zeta_p)(\lambda_\alpha\zeta_{n1})=\int=kT$$

and in the third instance—of latent massfree radiation, the transformation obeys the rule:

$$E=\lambda_{n1}cW_{n1}=\lambda_{n1}(\lambda_{n1}\zeta_{n1})(\lambda_{n1}f_{n1})=\lambda_{n1}{}^3\zeta_{n1}f_{n1}$$

where ζ and f are frequency functions, f being a specific gravitational frequency term, and f_{n1} being defined as equal to $(\lambda_{n1})^{-0.5}$ $\text{meter}^{0.5}\text{sec}^{-1}$. ζ_{n1} has the value of c/λ_{n1}.

If the electric variant of massfree radiation has a direct quantum equivalence, via the Duane-Hunt Law, none of the three primary aether energy variants possess either the classic form of electromagnetic energy which requires square superimposition of speed of light wave functions c, as c^2, or the quantum form of energy, requiring $E=h\upsilon$. The critical first step in the right direction may well be attributed to Dr. W. Reich, as it regards the fact that massfree energy couples two unequal wave functions, only one of which is electromagnetic and abides by the limit c. We then unravelled the threefold structure described above, and further showed that, in the case of longitudinal electric waves, the postulated equivalence ($q=\lambda_q c$) is merely phenomenological, as these waves are not restricted by the function c in their conveying of electric charge across space. It can further be demonstrated that all blackbody photons are bound by an upper frequency limit ($64*10^{14}$ Hz), above which only ionizing photons are produced, and that all blackbody photons arise precisely from the interaction of massfree electric radiation with molecules of matter (including light leptons), whereby the energy of that radiation is locally converted into photon or electromagnetic radiation. In other words, all nonionizing electromagnetic energy appears to be secondary energy that results locally from the interaction of matter with massfree electric energy. It cannot therefore consist of the primary energy that is present in the vacuum, an energy that is neither virtual nor electromagnetic, but actual and concrete in its electric, thermal and antigravitic manifestations. Lastly, gravitational energy, being either the potential or the kinetic energy responsible for the force of attraction between units of matter, is a manifestation that also requires, much as electromagnetic radiation does, coupling of massfree energy to matter or to mass-energy.

The Tesla coil is a generator of a massfree electric energy flux that it transmits both by conduction through the atmosphere and by conduction through the ground. Tesla thought it did just that, but it has been since regarded instead (because of Maxwell, Hertz and Marconi) as a transmitter of electromagnetic energy. The transmitter operates by a consumption of massbound electric power in the primary, and by induction it generates in the coupled secondary two electric fluxes, one massbound in the coil conductor, and the other massfree in the body of the solenoid. Tesla also proposed and demonstrated a receiver for the massfree energy flux in the form of a second Tesla coil resonant with the first. The receiver coil must be identical and tuned to the transmitter coil; the capacitance of the antenna plate must match that of the transmitter plate; both transmitter and receiver coils must be grounded; and the receiver coil input and output must be unipolar, as if the coil were wired in series.

The generators of massfree energy with which we are concerned provide current pulses associated with a damped wave (DW) oscillation of much higher frequency than the pulse repetition frequency. A particular problem in recovering the massfree energy content of such pulses is provided by the damped wave oscillations. Although in our U.S. Pat. No. 5,416,391 we describe arrangements incorporating split phase motors to recover such energy, their efficiency is a great deal less than what should theoretically be attainable. Other workers such as Tesla and Reich, have encountered the same problem to an even greater degree.

In XIXth century motor engineering terminology, dynamos capable of producing direct current by continuous homopolar induction were known as 'unipolar' generators. The term unipolar induction appears to have originated with W. Weber, to designate homopolar machines where the conductor moves continuously to cut the magnetic lines of one kind of magnetic pole only, and thus require sliding contacts to collect the generated current. Faraday's rotating copper disc apparatus was, in this sense, a homopolar generator when the disc was driven manually, or a homopolar motor when the current was provided to it. Where the rotating conductor continuously cuts the magnetic field of alternatingly opposite magnetic poles, the operation of a machine, whether a generator or a motor, is said to be heteropolar. Unipolar machines went on to have a life of their own in the form of low voltage and high current DC generators—from Faraday, through Plucker, Varley, Siemens, Ferraris, Hummel, to Lord Kelvin, Pancinoti, Tesla and others—almost exclusively in the form of disc dynamos, but some having Wound rotors. In Mordey's alternator, and in so-called 'inductor alternators', however, homopolar generators were employed to obtain alternating currents, with the use of rotors wound back and forth across the field. Use of smooth, unwound rotors in AC induction motors (as opposed to AC synchronous motors, such as hysteresis motors) was a later development than homopolar dynamos. By 1888, Tesla and Ferraris amongst still others, had independently produced rotating magnetic fields in a motor, by employing two separate alternate currents with the same frequency but different phase. Single phase alternate current motors were developed later, and split-phase motors were developed last. Ferraris (Ferraris, G (1888) "Rotazioni elettrodynamiche", Turin Acad, March issue.) proposed the elementary theory of the 2-phase motor, where the current induced in the rotor is proportional to the slip (the difference between the angular velocity of the magnetic field and that of the rotating cylinder), and the power of the motor is proportional to both the slip and the velocity of the rotor.

If an iron rotor is placed within the rotating magnetic field of a 2-phase stator, it will be set in rotation, but not synchronously, given that it is always attracted to the moving magnetic poles with a lag. But if an aluminum or copper rotor is used instead, it becomes 'dragged' around by the rotating stator field because of the eddy currents induced in it. If the aluminum or copper rotor were to rotate synchronously with the stator magnetic field, there would be no induced eddy currents and thus no motor action would result. The motor action depends, in this instance, upon the presence of asynchronous slip, since the function of the latter is to sustain the induction of those currents in the rotor that are responsible for the motor action of the dragged rotor. This then is the origin of the term 'AC drag motors'. Once the drag rotor evolved from a cylinder to a hollow cup, they earned the epithet of 'drag-cup motors'. Later, already in the XXth century, the cups were fitted over a central stator member, and the sleeve rotor 2 phase servomotor was born.

Tesla knew that impulse currents as well as CW sinusoidal currents could be used to drive AC motors. Regarding his invention of an hysteresis motor ('magnetic lag motor', as he called it), he stated: " . . . pulsatory as well as an alternating current might be used to drive the[se] motors . . . " (Martin, TC (1894) "The inventions, researches and writings of Nikola Tesla", Chapter XII, p. 68). In his search for efficient utilization of the high frequency DW impulse currents of his induction coils, Tesla began by employing an AC disc induction motor as shown in FIG. **17** of his famous 1892 address (Tesla, N (1892) "Experiments with alternate currents of high potential and high frequency", in "Nikola Tesla Lectures", 1956, Beograd, pp. L-70–71). This consisted of a copper or aluminum disc mounted vertically along the longitudinal axis of an iron core on which was wound a single motor coil which was series wired to the distal terminal of an induction coil at one end, and to a large suspended and insulated metal plate at the other. What was new about this was the implementation of an AC disc induction motor drive, where the exciting current traveled directly through the winding with just a unipolar connection to the coil secondary (under certain conditions, even the series connection to the plate could be removed, or replaced with a direct connection to the experimenter's body): "What I wish to show you is that this motor rotates with one single connection between it and the generator" (Tesla, N. (1892), op. cit., L-70, Tesla's emphasis). Indeed, he had just made a critical discovery that, unlike in the case of massbound charge where current flow requires depolarization of a bipolar tension, massfree charge engages current flow unipolarly as a mere matter of proper phase synchronization.

Tesla thought that his motor was particularly adequate to respond to windings that had 'high-self-induction', such as a single coil wound on an iron core. The basis of this self-induction is the magnetic reaction of a circuit, or an element of a circuit—an inductor—whereby it chokes, dims or dampens the amplitude of electric waves and retards their phase.

For the motor to respond to still higher frequencies, one needed to wind over the primary motor winding a partial overlap secondary, closed through a capacitor, since "it is not at all easy to obtain rotation with excessive frequencies, as the secondary cuts off almost completely the lines of the primary" (Idem, L-71.).

Tesla stated that "an additional feature of interest about this motor" was that one could run it with a single connection to the earth ground, although in fact one end of the motor primary coil had to remain connected to the large, suspended metal plate, placed so as to receive or be bathed by "an alternating electrostatic field", while the other end was taken to ground. Thus Tesla had an ordinary induction coil that transmitted this "alternating electrostatic field", an untuned Tesla antenna receiving this "field", and a receiver circuit comprising his iron-core wound motor primary, a closely coupled, capacitatively closed secondary, and the coupled nonferromagnetic disc rotor. Eventually, in his power transmission system, he would replace this transmitter with a Tesla coil, and place an identical receiving coil at the receiving end, to tune both systems and bring them into resonance. But his motor remained undeveloped, and so did the entire receiver system.

Tesla returned to this subject a year later: "on a former occasion I have described a simple form of motor comprising a single exciting coil, an iron core and disc" (Tesla, N (1893) "On light and other high frequency phenomena", in "Nikola Tesla Lectures", 1956, Beograd, pp. L-130, and L-131 with respect to FIG. **16**-II). He describes how he developed a variety of ways to operate such AC motors unipolarly from an induction transformer, and as well other arrangements for "operating a certain class of alternating motors founded on the action of currents of differing phase". Here, the connection to the induction transformer is altered so that the motor primary is driven from the coarse secondary of a transformer, whose finer primary is coupled, at one end, directly and with a single wire to the Tesla secondary, and at the other left unconnected. On this occasion, Tesla mentions that such a motor has been called a 'magnetic lag motor', but that this expression (which, incidentally, he had himself applied to his own invention of magnetic hysteresis motors) is objected to by "those who attribute the rotation of the disc to eddy currents when the core is finally subdivided" (Tesla,N (1893), op. cit., p. L-130).

In none of the other motor solutions, 2-phase or split-phase, that he suggests as unipolar couplings to the secondary of an induction coil, does the nonferromagnetic disc rotor motor again figure. But he returns to it a page later, and indirectly so, by first addressing the disadvantages of ferromagnetic rotors: "Very high frequencies are of course not practicable with motors on account of the necessity of employing iron cores. But one may use sudden discharges of low frequency and thus obtain certain advantages of high-frequency currents-without rendering the iron core entirely incapable of following the changes and without entailing a very great expenditure of energy in the core. I have found it quite practicable to operate, with such low frequency disruptive discharges of condensers, alternating-current motors."

In other words—whereas his experiments with constant wave (CW) alternating currents, and as well with high-voltage DW impulses from induction coils, indicated the existence of an upper frequency limit to iron core motor performance, one might employ instead high-current, DW impulses—of high DW frequencies but low impulse rates—to move these motors quite efficiently. Then he adds "A certain class of [AC] motors which I advanced a few years ago, that contain closed secondary circuits, will rotate quite vigorously when the discharges are directed through the exciting coils [emphasis added]. One reason that such a motor operates so well with these discharges is that the difference of phase between the primary and secondary currents is 90 degrees, which is generally not the case with harmonically rising and falling currents of low frequency. It might not be without interest to show an experiment with a simple motor of this kind, inasmuch as it is commonly thought that disruptive discharges are unsuitable for such purposes."

What he proposes next forms the basis of modern residential and industrial AC electric power meters, the AC copper disc motor whose rotor turns on the window of these meters, propelled forward by the supply frequency. But instead of employing any such CW input, Tesla uses the disruptive discharges of condensers, incipiently operating as current rectifiers. With the proper conditions, e.g. correct voltage from the generator, adequate current from the capacitor, optimum capacitance for the firing rate, and tuned spark-gap, to mention a few, Tesla found that the nonferromagnetic disc rotor turned but with considerable effort. But this hardly compared to the results obtained with a high-frequency CW alternator, which could drive the disc "with a much smaller effort". In summary then, Tesla went as far as being the first to devise a motor driven by Tesla waves, that employed a nonferromagnetic rotor, and whose arrangement encompassed both transmitter and receiver circuits. For this purpose, he employed a single phase method in which the signal is fed unipolarly to the winding, placed in series with a plate capacitance.

Tesla also later proposed driving a similar single-phase nonferromagnetic disc motor from bipolar capacitative discharges through an atmospheric spark-gap now placed in parallel with the main motor winding, and again simulating a split-phase by a closely wound secondary that was closed by a capacitance.

As Tesla admits, the results of all his AC eddy current motor solutions were meagre and limited by current and frequency problems. Likewise, the two phase arrangements proposed by Reich for his OR motor, involving a superimposition of the damped waves of a first phase on a fixed continuous wave second phase, require an external power source and a pulse amplifier circuit, and failed to meet Reich's own desiderata.

We have previously proposed the use of squirrel cage motors with capacitative splitting of phase to convert the damped wave (DW) output of plasma pulsers, but once a squirrel cage (SC) is introduced, the dampening effect which the nonferromagnetic copper cage exerts in being dragged by the revolving stator field is counteracted by the ferromagnetic cylinder of laminated iron, in which the copper cage is embedded, working to diminish the slip and bring the rotor to near synchronism. This is, in all likelihood, what limits SC motors to responding to the DC component of the DW impulse, and thus be limited to respond to fluxes of massbound charges. Historically, as we shall see, the obvious advantage of the SC servomotors lay in the fact that, in particular for 2-phase applications, they were far more efficient at performing work without evolution of heat. Indeed, if the eddy currents in the nonferromagnetic rotor are permitted to circulate in nonordered form, the rotor material and stator will heat up rapidly and consume much power in that heating. This is in fact considered to be a weakness of AC nonferromagnetic-rotor induction motors.

SUMMARY OF THE INVENTION

The present invention is concerned with conversion to conventional electrical energy of the variants of massfree energy radiation considered above, referred to for convenience as Tesla waves, massfree thermal radiation and latent massfree radiation. The first variant of such radiation was recognized, generated and at least partially disclosed by Tesla about a hundred years ago, although his work has been widely misinterpreted and also confused with his work on the transmission of radio or electromagnetic waves. The Tesla coil is a convenient generator of such radiation, and is used as such in many of the embodiments of our invention described below, but it should be clearly understood that our invention in its broadest sense is not restricted to the use of such a coil as a source of massfree radiation and any natural or artificial source may be utilized. For example, the sun is a natural source of such radiation, although interaction with the atmosphere means that it is largely unavailable at the earth's surface, limiting applications to locations outside of the earth's atmosphere.

According to the invention, a device for the conversion of massfree radiation (as herein defined) into electrical or mechanical energy comprises a transmitter of massfree electrical radiation having a damped wave component, a receiver of such radiation tuned to resonance with the damped wave frequency of the transmitter, a co-resonant output circuit coupled into and extracting electrical or kinetic energy from the receiver, and at least one of structure defining a transmission cavity between the transmitter and the receiver, a full-wave rectifier in the co-resonant output circuit, and an oscillatory pulsed plasma discharge device incorporated in the co-resonant output circuit. The output circuit preferably comprises a full wave rectifier presenting a capacitance to the receiver, or an electric motor, preferably a split phase motor, presenting inductance to the receiver. The transmitter and receiver each preferably comprise a Tesla coil and/or an autogenous pulsed abnormal glow discharge device. The transmission cavity is preferably at least partially evacuated, and comprises spaced plates connected respectively to the distal poles of the secondaries of Tesla coils incorporated in the transmitter and receiver respectively, the plates being parallel or concentric. The structure defining the cavity may be immersed in ion-containing water. The split-phase motor is preferably an inertially damped AC drag motor.

The invention, and experiments demonstrating its basis, are described further below with reference to the accompanying drawings.

SHORT DESCRIPTION OF THE DRAWINGS

FIG. **1** is a schematic view of a Tesla coil connected to a full-wave rectifier to form an energy conversion device;

FIG. **2** is a schematic view of a Tesla coil connected to a gold leaf electrometer;

FIGS. **3**–**6** show alternative electrometer configurations;

FIGS. **7**–**11** show modifications of the circuit of FIG. **1**;

FIG. **12** shows apparatus for investigating aspects of the experimental results obtained with the foregoing devices;

FIG. **13** is a graph illustrating results obtained from the apparatus of FIG. **12**;

FIGS. **14**–**17** show schematic diagrams of embodiments of energy conversion devices;

FIG. **18** is a diagrammatic cross section of an inertially damped drag cup motor; and

FIG. **19** is a schematic diagram of a further embodiment of energy conversion device incorporating such a motor.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Based upon observations of weight loss in metallic matter as induced by exposure to high frequency alternating electric fields, we developed an experimental method to optimize this weight loss, and from this a device that treats the forces causing weight loss as manifestations of intrinsic potential energy ΔU (or true "latent heat") of the molecules of matter, and converts both "true latent heat" energy present in the neighbourhood of a receiver, and "sensible" heat induced within that receiver, into electric energy that can be employed to drive a motor, flywheel or charge battery supplies.

It is commonly believed that the output of the Tesla coil is ionizing electromagnetic radiation. We have demonstrated that it is not; i.e. that it is neither electromagnetic radiation, nor ionizing electromagnetic radiation. The output of an air cored, sequentially-wound secondary, consists exclusively of electric energy: upon contact with the coil, a massbound AC current can be extracted at the resonant frequency, whilst across a non-sparking gap, massfree AC-like electric wave radiation having the characteristics of longitudinal waves, can be intercepted anywhere in adjacent space. Accordingly, the radiation output from such coils is distinct from electromagnetic radiation.

The basic demonstration that the output of a Tesla coil does not consist of ionizing radiation is that it does not accelerate the spontaneous discharge rate of electroscopes, whether positively or negatively charged. In fact, in its immediate periphery, the coil only accelerates the spontaneous discharge rate of the negatively charged electroscope (i.e. the charge leakage rate), whereas it arrests the discharge of the positively charged electroscope (i.e. the charge seepage rate falls to zero). But this dual effect is not due to any emission of positive ions from the secondary, even if it can positively charge a discharged electroscope brought to its proximity. This charging effect is in fact an artifact, in that metals but not dielectrics are ready to lose their conduction and outer valence band electrons when exposed to the massfree electric radiation of the coil. This is simply demonstrated by the apparatus of FIG. **1**, in which the distal terminal of the secondary winding **6** of a Tesla coil having a primary winding **4** driven by a vibrator **2** is connected to the input of a full-wave voltage wave divider formed by diodes **8** & **10** and reservoir capacitors **12** & **14** (the same reference numerals are used for similar parts in subsequent Figures). If the rectifiers employed are nondoped, the coil appears to only charge the divider at the positive capacitance **10**, but if doped rectifiers are employed, the coil will be observed to charge both capacitances equally. Whereas positive ionizers can charge either doped or undoped dividers positively, no positive ionizer can charge a doped divider negatively, clearly demonstrating that the Tesla coil does not emit positive ions.

The basic demonstration that the output of a Tesla coil is not nonionizing electromagnetic radiation of high frequency, such as optical radiation, or of lower frequency, such as thermal photons, is also a simple one. Placement of a sensitive wide spectrum photoelectric cell (capable of detecting radiation to the limits of vacuum UV), wired in the traditional closed circuit manner from a battery supply, at any distance short of sparking from the distal terminal of the coil will show in the dark that the light output from the coil is negligible. This rules out optical radiation at high frequency. The demonstration that the sensible heat output from the Tesla coil is also negligible will be addressed below.

Our theory proposed the existence of physical processes whereby massfree electric radiation can be converted into electromagnetic radiation. Such a process is at work whenever massfree electric wave radiation interacts with electrons, such as those that remain in the valence bands of atoms. This massfree electric energy interacts with charge carriers, such as electrons, to confer on them an electrokinetic energy which they shed in the form of light whenever that electrokinetic energy is dissociated from those carriers (e.g. by deceleration, collision or friction processes). Such a process is at work to a negligible extent in the coil itself and its usual terminal capacitance, hence the faint glow that can be seen to issue from it, but it can also be greatly amplified in the form of a corona discharge by connecting a large area plate to the output of the secondary, as Tesla himself did in his own experiments, and thus by increasing the capacitance of the coil system. Now, what is interesting in this process is that, in the absence of virtually any I^2R losses at the plate, and if the plate thus introduced is bent at the edges so that it has no pointed edges, or if it is in the form of a bowl, or in any other manner that precludes sparking at edges and specially corners, and thus enhances the corona discharge, any electroscope, whether negatively or positively charged, now brought close to the plate will show a tendency to arrest its spontaneous discharge rate. One might say that this is simply the result obtained in a Faraday cage that disperses charge on its outside and insulates electrically its interior, and indeed if an electroscope is placed inside a Faraday cage no amount of Tesla radiation on the outside of that cage, save direct sparking, adversely affects the leakage or seepage rate of the electroscope. In fact, since the effect of such a cage can be shown to be that of, by itself, inducing arrest of either spontaneous electroscopic discharge, this effect simply remains or is magnified when the cage is bathed by Tesla radiation. However, a cage constitutes an electrically isolated environment, whereas a plate with or without curved or bent edges does not. Furthermore, the change observed in the properties of the output radiation from a Tesla coil when certain metal plates or surfaces are directly connected to the distal terminal of the secondary, takes place whilst the capacitance of the coil is increased by the connected plate, and thus the plate is an electrically active element of the circuit—and hence the opposite of an electrically isolated element.

For a long time we believed that the anomalous cathode reaction forces observed in autoelectronic discharges (atmospheric sparks, autogenous PAGD (pulsed abnormal glow discharge) and vacuum arc discharges) were exclusive to an autoelectronic emission mechanism prompted by a direct potential between discharging electrodes. Sparking driven by AC potentials could sustain the same forces, but their mutual cancellation over time would not deploy a net force. In this sense, when a large gold leaf connected directly to the ground (via a water pipe or any other suitable connection) or to another large area plate suspended at some height above the ground, is vertically placed at a sparking distance above the surface of another plate connected to the secondary of a Tesla coil, one would not expect the AC spark to sustain any net force across the gap between the gold leaf and the plate. In terms of cathode reaction forces, one would expect their cancellation to be simply brought about by the high frequency of the current alternation in the coil, as both leaf and plate would alternate between being the emitting cathode or the receiving anode. However, this is not what is observed—instead, the gold leaf **16** lifts away from the plate **18** (FIG. **2**). Connecting instead the suspended gold leaf to the coil terminal, and the bottom plate to the ground in the same manner as described above, also yields the same result.

Even more curious is the finding that this anomalous reaction force deployed by an alternate current of mass-bound charges in the arc, remains present when the sparking is prevented and instead the corona effect is enhanced (by employing a large plate connected to the distal pole of the secondary, and by employing a distance at which sparking ceases), as if the lift itself were the property of the corona underlying the spark channels and not the property per se of the autoelectronic emission mechanism.

By mounting directly the suspended leaf **16** (41 mg of hammered 99.9996% pure gold) at the end of a long dielectric rod **20** balanced at the center and placed on a light stand over an electronic balance **22**, we sought to determine the observed lift of the leaf as weight lost. Surprisingly, and despite the most apparent lifting motion of the leaf, the balance registered a substantial weight gain, indicating the addition of 1 to 5 mg weight (with the same 14 W input to the vibrator stage), independently of whether the leaf was connected to the terminal of the coil or instead to the earth ground via a water pipe. This suggested to us that, whether formed as a DC or AC spark channel, or whether in the form of a corona discharge, the electric gap develops an expansion force (exactly opposite to a Casimir force) on both electrodes, independently of their polarity, which force is responsible for the observed repulsion. Yet, this expansion goes hand in hand with an increase in their weight such that some other process is at work in that electric gap.

To examine this problem further, we assembled a different experiment where the gold leaf **16** was suspended between two large metal plates **18**, **24** placed 20 cm apart, and the leaf was not electrically connected to them or to any other circuit, while attached to the dielectric rod employed to suspend it over the electronic balance. Given that the leaf is suitably and equally spaced from both plates, there is no arcing between it and either plate. The obvious expectation is that, since the electric field bathing the leaf alternates at high frequency (measured in hundreds of kilohertz), and the corona from both electrodes should equalize and balance any electric wind, no lift should be observed. In fact, no lift is apparent, but a most curious observation is made: depending upon which orientation is employed for the plates, the gold leaf either gains or loses 4–6% of its weight. This gain or loss is registered for as long as the coil is on. If the top plate is grounded and the bottom one connected to the different terminal of the secondary, a gain in weight is observed (FIG. **3**). If the connections are reversed, an equal weight loss is registered (FIG. **4**).

Furthermore, in this last instance, if the grounded plate **24** is entirely removed (FIG. **5**), and only the top plate remains connected to the distal terminal of the secondary, the observed loss of weight continues to occur such that effectively this reaction can be obtained with unipolar electric fields of high frequency, and provides a unidirectional force which, once exerted upon metallic objects bathed by its field, can be made to oppose or favour gravity.

Now, these effects can be greatly magnified, in the order of 10-fold, if the same gold leaf is made part of a simple series floating electric circuit where the leaf functions as a large area plate, and is wired in series with a coil **26** which, for best results, should be wound so as to be of a length resonant with the secondary of the Tesla-type coil employed; and this coil is connected in turn to a point antenna **28** upwardly oriented (FIG. **6**). The entire floating circuit is mounted on the rod **20** and this in turn is mounted over the sensitive balance. If both plates are kept as in FIGS. **3** & **4**, the observed weight loss and weight gain respectively vary between 30 to 95% of the total weight of the leaf. Again the gain or loss of weight is registered for as long as the coil is on.

These anomalous findings suggested that whatever is the nature of the energy responsible for the force observed in that high frequency alternating current gap, any metallic object placed in that gap will experience a force repelling it from the electric ground. This force will be maximal if the gap frequency is tuned to the elementary or molecular structure of the metallic object. If the electric ground is placed opposite the actual plane of the earth ground, that force will act in the direction of gravity. If, instead, the electric ground and the earth ground are made to coincide on the same plane, that force will act opposite the direction of gravity, i.e. will repel the metallic object from the ground.

No such weight alteration was observed with solid dielectrics, for instance with polyethylene and other thermoplastic sheets.

These facts rule out the possibility of a hidden electrostatic attraction force acting between the plate connected to the different terminal of the secondary and the gold leaf. First off, such an attraction would be able to lift entirely the gold leaf, as is easily observed with the unipole of any electrostatic generator operating with a few milliwatts output with either negative or positive polarity; secondly, the same attraction, if it existed and were the product of an electric force, would surely be manifested independently from whether the experimental leaf was metallic or a dielectric (as again is observed with electrostatic generators).

The results suggest therefore that whenever a large plate is connected to a Tesla-type coil, it induces in surrounding matter that is not part of its own circuit, a directional thrust oriented opposite the electric ground and, if the latter coincides with the gravitational ground, a thrust oriented opposite free fall.

When this thrust is made to oppose gravity, we believe that its effect upon the gold leaf can be compared to the lifting power imparted to the water molecule when it transits from the liquid to the vapor state and which is associated with the increase in internal (or intrinsic) potential 'thermal' energy ΔU (See Halliday D & ResnickR (1978) "Physics", Vol. 1, section 22–8, p. 489). The "specific latent heat" of water (m*L) contains indeed both an expression for the sensible radiant thermal work involving volume and pressure relations ($W=P(V_V-V_L)$, where P=1 Atm, and V_V and V_L, respectively, are the molar volumes in the vapor and liquid phases), and an expression for a quantity of "latent" energy (ΔU) which is associated with the molecule in the more rarefied state. Hence, the relation for the latter with respect to water vapor:

$$\Delta U=(mL)-[P(V_V-V_L)]$$

We propose that likewise, if a very small portion of the energy of the massfree electric waves is indirectly transformed by massbound charge carriers on that plate into blackbody photons (once those charge carriers shed their electrokinetic energy), the greater portion of those waves are directly transformed in the space adjacent to that plate into the latent energy equivalent to ΔU for the atoms of the surrounding air, and so on, until this process itself is also occurring for the atoms of that gold leaf, thus inducing their nonelectrical weight loss and suggesting the existence of a nonthermal "antigravitokinetic" energy term heretofore unknown to mankind other than as "latent heat" or "internal potential energy".

From this viewpoint, the energy released by any Tesla-type coil to its surroundings would be tantamount to a radiative injection of "internal potential energy" which would confer on local gas molecules a weight cancellation (a cancellation of gravitational mass occurring in the absence of any cancellation of inertial mass—a process which the inventors theorize is explained by the neutralisation of elementary gravitons), and the same process would be equally at work for metallic solids but not dielectric solids.

Gold vapour also deploys a substantial intrinsic potential energy. With an enthalpy of vaporization on the order of $H_v=324$ kJ mol^{-1}, the molar volumetric work performed by gold vapor at atmospheric pressure at the temperature of vaporization T_v (2,856° C., i.e. 3129 degrees Kelvin) is:

$$W=P\Delta V_{V\text{-}L}=23.58\text{ kJ mol}^{-1}$$

where $\Delta V_{V\text{-}L}=0.2327\text{ m}^{-3}$. The intrinsic potential energy of gold vapor is then given by:

$$\Delta U=H_v-W=300.4\text{ kJ mol}^{-1}$$

i.e. 12.74 times greater than the volumetric work performed during the phase transition.

It is our contention that this intrinsic potential energy, associated with molecules as their "latent heat", has fine structure that in turn is altered if this energy is released from these molecules and fails to gain a "sensible" thermal form. What is suggested is that the fine structure of "latent heat" is not electromagnetic and obeys instead the molecular function:

$$\Delta U/N_A=\lambda_{n2}{}^2cf_{n2}$$

where N_A is Avogadro's number, the wavelength denoted as λ_{n2} is the wavelength-equivalent of the mass of the molecule to which the "latent heat" is associated, obtained by a conversion method proposed in these inventors' theory, and the frequency term f is a nonelectromagnetic frequency term, specifically in this case a gravitational frequency function. Employing the conversion of Joules into m^3 sec^{-2} proposed by these inventors as being exactly 1 J=10N_Am^3 sec^{-2}, and putting the wavelength λ_{n2} down as the wavelength-equivalent of the mass of the gold atom, λ_{Au}, at 1.9698 m, that frequency term f_{n2} can be obtained as being equal to 2.6*10^{-3} sec^{-1}. According to the present inventors' theory, the wavefunction c constitutive of the fine structure of "latent heat" associated with molecules of matter, carries the same wavelength λ_{Au} and its frequency is given in the usual manner by c/λ_{Au}=1.52*10^8 sec^{-1}. The resultant frequency for the non-Planckian unit quantum of "latent energy" associated with each gold atom at the vaporization temperature is then obtained by the geometric mean of the two synchronous frequency terms:

$$[(c/\lambda_{Au})f_{n2}]^{0.5}=624\text{ Hz}$$

However, this is the signature of that intrinsic potential energy when associated with that gold atom at its vaporization temperature. It is not the signature of the energy quantum itself if it is released from that molecule, nor prior to being absorbed (i.e. in transit), at that same temperature. The fine structure of the same non-Planckian "latent" energy quantum varies to encompass different determinations of the constituent wavelength and frequency functions. The basic relation for the determination of the wavelength of a "latent thermal" energy quantum not associated with matter, but corresponding to one that is, is:

$$\lambda_{n1}=[\Delta U/N_A)/c]^{0.666}\text{ meters}^{-0.333}\text{ seconds}^{0.666}$$

which gives 0.046478 m for the unbound equivalent of the "latent heat" unit quantum of vaporization associated with the gold atom at one atmosphere. The fine structure of the free quantum is still parallel, as given by:

$$\Delta U/N_A=\lambda_{n1}{}^2cf_{n1}$$

but now notice how the frequency terms have changed value, with the f_{n1} function having the value 4.65 sec^{-1} and c/λ_{n1} yielding 6.48*10^9 sec^{-1}. The geometric mean of the superimposition of the two frequencies is then—

$$[(c/\lambda_{n1})f_{n1}]^{0.5}=173.7\text{ KHz}$$

We contend that it is at this frequency that the atoms of gold vapor absorb "latent heat".

However, this is just the overall scenario of what happens at the temperature of vaporization of gold. But at room temperature (e.g. 293 degrees Kelvin), and with respect to processes where there is no sublimation of the atoms of that gold leaf under way (and indeed, once the coil is turned off, the leaf returns to its normal weight), one must infer to a different phase of matter what portion of "latent heat" energy, if any, do the atoms of gold hold in the solid phase lattice. Assuming the same proportionality between the "sensible" and "latent" thermal energy terms for atoms of gold at room temperature, where the unit thermal energy is N_A kT=2.436 kJ mol^{-1}, the gold atom could absorb up to 12.74 times, we speculate, the value of this "sensible" thermal energy, and thus hold N_A kT=31.053 kJ more energy in its own micro-atmosphere.

If this speculation is correct, and employing the above novel methodology, then the mean geometric frequency of the maximal "latent heat" energy quantum of a gold atom at room temperature would be 538 KHz (versus 174 KHz at the vaporization temperature), and once absorbed its mean frequency mode would abate to 201.5 Hz (versus 630 Hz once the atom has vaporized).

To test this hypothesis we employed two different Tesla-type coils having output frequencies of 200 KHz and 394 KHz. The circuit tested was that shown in FIG. **6**, and both coils were operated at 50 KV outputs. Whereas the former coil, closer to the 174 KHz marker, could only systematically produce 10–11 mg of weight cancellation in the gold leaf of the floating circuit, the second coil, closer to the speculated 538 KHz marker, could produce 15–35 mg of weight cancellation in the same gold leaf. The empirical results appear therefore to suggest that our speculation may well be a valid one.

The above-mentioned full wave divider (see FIG. **1**) can be easily coupled to our autogenous pulsed abnormal glow discharge technology as described in our U.S. Pat. No. 5,416,391 to form an alternative source of direct current, ultimately powered by Tesla waves, and such a drive can equally be applied to any other vacuum device that can sustain endogenous oscillatory discharges, whether in the PAGD regime or any other pulsatory regime. For the purposes of experimental and visual determination of power outputs from the divider in question, we have utilized either 2 Torr vacuum tubes operating in the high-current PAGD regime, or 20–100 Torr spark tubes requiring high voltages (2 to 10 KV) for their spark breakdown. As taught in the above U.S. Patent, the output from the full wave voltage divider can be assessed by the energy spent in driving the tube and the motor, whose rotary speed is proportional, within the limits chosen, to the power input.

Two separate sets of experiments set forth in Table I below showed that direct connection of the wave divider to the distal terminal of the coil (set constantly at 6 clicks on the vibrator stage) (FIG. **1**) or to the same terminal but across a large (2 or 3 square feet) plate **30** that increased the capacitance of the secondary (FIG. **7**), presented the same power output in either case (the effect of the plate is to lower the voltage of the output proportional to the increase in current). A substantial increase in power output through the divider is observed only when an identically wound Tesla coil is connected in reverse (FIG. **8**) with the non-common end of its winding **4** not connected, in order to obtain a condition of resonance, and this observed increase is further augmented by now interposing either of the metal plates **18**, **24** between the two chirally connected and identical coils (FIG. **9**). The increase in plate area appears to have the effect of increasing the output for as long as the plate is isolated between the two chiral image coils. Throughout these experiments the input power to the vibrator was fixed at 14 W (60 Hz AC).

TABLE 1

Status	Pulse rate (PPS)	Motor rotation (RPM), M ± SEM
Expt A		
Tesla coil (TC) to divider	2.6	582.5 ± 3.9 (n = 4)
TC to inverted TC, to divider	4.4	621 7.6 (n = 4)
TC to 2 ft^2 plate, to inverted TC, to divider	5	775.25 ± 23.6 (n = 4)
Expt B		
Tesla coil (TC) to divider	2.2	613 ± 5.6 (n = 12)
TC to 3 ft^2 plate, to divider	2.3	605 ± 2.6 (n = 12)
TC to inverted TC, to divider	2.3	722 ± 5.7 (n = 12)
TC to 3 ft^2 plate, to inverted TC, to divider	4.2	877.6 ± 6.5 (n = 12)

In our loss of weight experiments described above, we noted that the phenomenon of weight loss by a metallic body placed in proximity of the coil output continued to be observed when only the plate connected to the distal pole of the secondary was retained. The leaf, although not part of the circuit of the secondary, could however be seen as part of a circuit for the capture of ambient radiant energy, specifically that generated by the coil and, as well, that also possibly picked up, in the process, from other ambient sources. To determine whether the last consideration is a possibility at all, or whether the energy picked up by an analogue of our metallic body or gold leaf in the experiments described above, is entirely a byproduct of the energy transmitted by the plate connected to the distal pole of the secondary, we next determined what would happen if the pick-up for the full wave-divider were placed not at the output from the secondary coil, but from an in all respects identical plate (the Receiver plate R, as opposed to the Transmitter plate T) placed a distance away from, and above, the first one. In other words, the gold leaf is replaced by a receiver plate, and this carries an attached test circuit identical to the test circuit employed to assess directly the coil output.

As shown in Table 2 below, the results of the experiment show that there is no loss of energy picked up at the R plate (FIG. **10**) when compared to the most favourable situation involving the plate **30** (FIG. **9**) interposed between the chirally connected coils. This observation is however not always the case. For best results one should employ iron, gold or silver plates placed parallel to the horizon, with the T plate underneath the R plate. In fact, if one employs instead aluminum plates and suspends these vertically, one can consistently register a loss of output at the divider when changing the divider input from the T to the R plates.

If however the plate R is connected in turn to a second identical coil, also wired in reverse, and this second coil in turn serves as input to the full wave divider (FIG. **11**), then a most curious occurrence takes place—the power output increases considerably (see Table 2), as if the divider circuit had undergone an energy injection not present at the source. Note that the circuits are in fact resonant, but the energy injection contributing nearly 60–66% (for both plate areas in the previous experiment) of the input that we refer to, is not caused by inductive resonance, since the effect of resonance can be ascribed to the set-up described in FIG. **9**. The distance between the plates, as well as their orientation with respect to the local horizon system of the observer also appear to matter, best results being achieved at optimal distances (e.g. for 2 square feet plates the best gap, at 43% RH and room temperature, was at least 6 inches).

TABLE 2

Status	T R distance	Pulse rate (PPS)	Motor rotation (RPM), M ± SEM
2 ft² plates			
R plate to inverted TC, to divider	3"	6.7	882 ± 17.5 (n = 4)
	4"	8	906 ± 12.1 (n = 4)
	6"	10	936 ± 46.1 (n = 9)
3 ft² plates			
TC to T plate, to divider	0	2.3	605 ± 2.6 (n = 12)
R plate to divider	6"	3.3	890.1 ± 3.8 (n = 12)
R plate to inverted TC, to divider	6"	5.1	1009.2 ± 4 (n = 12)
R plate to divider	8"	4.0	783.1 ± 11.3 (n = 12)
R plate to inverted TC, to divider	8"	5.1	1005.7 ± 6 (n = 12)

We tested the possibility that environmental heat produced by operation of the coil might be the source of the injected energy, the plate of the second system acting possibly as collector for the heat present in the gap. As it turned out, experiments showed repeatedly that in the gap between the T and R plates there was no significant thermal radiation propagating between one and the other. The more illustrative experiments are those in which we identified where the sensible thermal energy appears, and which involved coupling two cavities: the Transmitter-Receiver gap between plates T and R, and a Faraday cage enclosure **34** (see FIG. **12**). The first cavity appears to be much like that of a capacitor: the two identical parallel plates are surrounded by a thick dielectric insulator **32**, and a thermometer T**2** is inserted half-way through it. A thermometer T**1** is also affixed to the T plate, to read its temperature. The second cavity is a simple metal cage, also insulated, with a thermometer T**3** inserted 2 cm into its top. 2–4 cm above the top of the cage there is placed a fourth thermometer T**4**, within an insulated cylinder. If the TC is a source of thermal energy (e.g. IR radiation, microwaves, etc.) we would expect the T plate to be the hottest element from which, by radiation, thermal energy would reach the middle of the first cavity making the next thermometer T**2** second hottest, and that the third thermometer T**3** inside the second cavity, even if it might initially be slightly warmer than the other two, would, over time, become comparatively cooler than either one of the other two thermometers, despite the fact that the rising heat would still be seen to warm it up over time. One would expect a similar outcome for the fourth thermometer T**4**, above the cage. As shown by FIG. **13**, where only the temperature differences $(\Delta T^\circ - T_c^\circ)$ between the experimental thermometers and the control thermometer reading the air temperature T_c° of the laboratory are shown, the surface of the T plate warms up by 0.1° C. at 3' after initiation of the run (closed squares), whereas in the space of the T/R gap a diminutive warming, by 0.05° C., is registered after 10' (open circles). Conversely, the temperature inside the cage, at the top (shaded circles) rises by 0.1° C. also by the third minute, and the temperature above the cage itself (shaded squares) rises by a much greater difference of 0.35° C., which remains stable after the eighth minute.

These results show that it is not sensible heat that radiates from the T plate. Rather, some other form of radiation traverses these cavities to generate sensible heat at their metallic boundaries, such that more heat is generated above the R plate (inside the cage) and again above the third plate, i.e. above the top of the cage, than is generated in the T/R gap, i.e. proximally to the T plate. This clearly shows that the Tesla coil is not a significant source of thermal radiation, and that sensible heat can be detected inside and atop the Faraday cage only as a further transformation of the radiant energy transmitted across the T/R cavity.

The same experiment also illustrates that, whatever is the nature of the additional environmental energy being injected at the surface of R plate (as shown by Table 2 results above), it is most likely not thermal radiation, at least not energy in the form of sensible heat. And whatever is the nature of this ambient radiant energy being mobilized by the electric radiant energy transmitted from the T plate, it can produce significant heat inside an enclosure contiguous with plate R.

Since we also know experimentally that this observation of an ambient energy injection at the R plate or R cage depends upon relative humidity, being most easily observable when the latter is low (<50% RH), and being virtually impossible to observe when air is saturated with water vapor, we can infer that water vapor is a good absorber of the electric massfree radiant energy emitted from the T plate. This strongly suggests that this absorption process is tantamount to increasing the potential intrinsic energy ΔU of the water vapor molecules adjacent to the T plate. In the absence of significant quantities of water vapor, when the atmosphere is dry, one may speculate that this absorption process is replaced by what one presumes is a parallel process involving the various gaseous molecules of air. However, either because the air molecules involve molecular species that readily give off this potential energy, as one might speculate is the case with molecular oxygen, hydrogen and nitrogen, or because the air molecules absorb far less "latent" energy (as appears to be the case with inert gases), and therefore there is more of it in the molecularly unbound state (as we explicitly propose as a possibility) and thus available for absorption by the appropriately tuned receiver, the increased ΔU of air molecules conferred by the absorption of the massfree electric radiation in the T/R gap is transferred to the R conductor together with the latent energy which those molecules already possessed before entering that gap. Hence the energy injection and its dependency upon the partial pressure of water vapor, which absconds instead with this "latent" energy and succeeds in withholding it from transmission to the R plate.

If the T/R gap can mobilize ambient energy which is neither electromagnetic nor thermal in nature, but which "latent" energy becomes injected into the divider circuit in electric form, the heat (i.e. sensible thermal energy) produced inside and atop the cage can also be mobilized electrically as input into the divider circuit. The obvious place to look for the emplacement of the cool junction that could convert sensible heat into electrokinetic energy of massbound charges is at the top of the cage, where it is warmest (See top curve of FIG. **13** in shaded squares). This is clearly observed from the results shown in Table 3 below, where the initial temperature difference between the top of the box and the T plate surface was 0.5° C., and the top of the box temperature rose by 0.2° C. after 2.5 minutes when the divider was connected at the junction, versus 0.35° C. when it was not (and the transmitter coil was on).

TABLE 3

Status 3 ft^2 plates	TR distance inches	Pulse rate PPS	Motor rotation RPM, M ± SEM(n = 12)
TC to T plate, to divider	NA	4.2	877.6 ± 6.5
R plate to inverted TC, to divider	6"	5.1	1009.2 ± 4
Top of naked R plate/cage to divider	6"	5.4	1047.1 ± 5.7
Top of insulated R plate/cage exposed to sun, to divider	6"	6.1	1072.4 ± 8.7

For the run performed with the naked R cage, the temperature directly above the top of the cage was 24.3° C., at the outset, versus the control room temperature of 23.9° C. For the run performed with the insulated R cage exposed directly to the sun at midday, on a cool and clear August day, the temperature directly above the top of the cage was 33° C., versus the control air temperature of 18.4° C. The temperature of the cool junction at the top of the cage was 31.9° C. while the run was performed.

It is apparent from the data of Table 3 how a second injection of energy has occurred in the apparatus. If, within the T/R gap, the energy injected appears to be on the order of absorption of "latent heat", at the top of the cage cavity, at the cool junction, the injection is one of radiant "sensible" heat. Moreover, this secondary energy addition could be further enhanced by placing strong insulation around the whole apparatus or the cage itself, and further so, by exposing the whole apparatus to solar radiation.

We next turned our attention to the T/R gap cavity with the intention of determining whether atmospheric conditions or vacua yield the same or different results. We could not, of course, test the same large area plates as have been employed for the studies undertaken at atmospheric pressures. For the present purpose we employed instead large area electrodes (ca 0.2 ft^2) made of high grade stainless steel or even aluminum. Preliminary results showed that these T/R gap tubes, when coupled to the divider circuit, yielded faster pulse rates in the secondary circuit when evacuated than at atmospheric pressure. The strength of the corona discharge also intensified, as it eventually became replaced by a normal glow discharge. For purposes of improved spatial capture of (1) the electric massfree energy radiated from the T electrode and (2) the non-radiant latent thermal energy mobilized by it to be collected electrically at the R plate, an axial cylindrical T electrode was inserted inside a larger concentric cylinder or between two common plates of large surface area (e.g. >100 cm^2) functioning as the R electrode(s), in a dielectric container suitable for evacuation (glass, polycarbonate), at a typical distance of at least 3 cm between electrodes, and the entire device was tested at different pressures. The secondary circuit connected downstream from the full wave divider was as shown in FIG. **14** (employing an autogenous pulsed abnormal glow discharge, or PAGD, converter circuit), with the PAGD reactor 36 set at 10 Torr (in light of the high-voltage input, which varied between 1500 and 3200V) and gave the results presented in Table 4 below. We should remark also that these pulses charged the charge pack CP through the coupling capacitors **38**, bridge rectifier **40** and reservoir capacitors **42**, and blocking diodes **44**, as expected from the prior art represented by our patents related to PAGD devices.

TABLE 4

T/R tube Pressure (Torr)	Pulse rate (PPS)
760	0.376
0.025	0.513

The effect of the vacuum in the T/R gap tube seems to be dual. By transforming the corona discharge into a normal glow discharge, it increases the local production of photons (probably associated to the formation and discharge of metastable states in the plasma), and at the same time, increases the pulse rate in the output circuit and thus, in all probability, the energy injected in the T/R gap cavity. But this did not yet permit us to confirm whether or not it is "latent heat" energy of the plasma molecules which is being tapped at the receiver plate, even if it be plausible in principle that plasmas may effect more efficient transfer of "latent heat" to tuned receivers than atmospheric gases.

The vacuum dependency of the pulse rate of the PAGD reactor employed as example in the secondary circuit downstream from the divider is also rather well marked, with the fastest pulse rates being registered at 1 Torr for the sample run shown in Table 5 below.

TABLE 5

T/R tube Pressure (Torr)	Pulse rate (PPS)	PAGD Reactor Pressure (Torr)	Voltage (across divider)
0.025	0.115	90	4.5 kV
0.025	0.1553	75	3.5 kV
0.025	0.183	60	3.3 kV
0.025	0.291	30	
0.025	0.513	15	1.6 kV
0.025	0.602	10	1.4 kV
0.025	2.9	2	0.53 kV
0.025	4.1	1	0.45 kV

It is worth noting here that the illustrated polarity of the wiring of the PAGD reactor tube, as shown in FIG. **14**, is best for purposes of sustaining regular autoelectronic emission at high voltage. The reverse configuration, with the center electrode negative and the plates positive favours instead heating of the cathode and a lapse into a normal glow discharge.

We tested a similar arrangement to that shown in FIG. **14** above, but with a PAGD motor circuit (see our U.S. Pat. No. 5,416,391). A split phase motor **44** replaces the rectifier and charge pack, and the PAGD reactor is operated at the same pressure of 15 Torr, as shown in FIG. **15**. The T/R gap tube tested had a longer plate distance (2"), with one plate now functioning as Transmitter and the other as Receiver. Note also the different wiring of the PAGD reactor. The results, as shown below in Table 6, present pulse per second (PPS) and motor revolutions per minute (RPM) curve trends that appear to be analogous and parallel to the well-known Paschen curves for breakdown voltage in vacuum—such that the T/R gap performs better either in the atmospheric corona discharge mode, or in the high vacuum normal glow discharge (NGD) mode, than in the low breakdown voltage range of the curve where the discharge forms a narrow channel and takes on the appearance of an "aurora" transitional region discharge (TRD).

TABLE 6

T/R tube Pressure (Torr)	Pulse rate (PPS)	Motor rotation (RPM), M ± SEM (n = 17)	Discharge Type
760	2.8	751.2 ± 7.1	Corona
100	2.1	611.5 ± 5.1	TRD
20	2.4	701.9 ± 4.6	TRD
0.006	2.8	748.4 ± 9.3	NGD
0.003	3.0	819.4 ± 6.3	NGD

These results suggest that plasmas with high lateral dispersion, i.e. formed over large electrode areas (e.g. corona and NGD plasmas) and thus devoid of pinch, are more likely to mobilize electrically the intrinsic potential energy of the molecular charges than pinch plasmas appear to be able to do (e.g. TRD plasmas). Apparently also, the greater the vacuum drawn from the T/R gap cavity, the more efficient does the transfer of this intrinsic potential energy, ie the massbound latent heat, to the electrokinetic energy of the charges circulating in the receiver circuit become. At about $6*10^{-2}$ Torr, this transfer in vacuo is comparable to that observed under atmospheric conditions and thus for a much greater density of molecules.

We investigated whether it Is possible to tap the latent heat energy of water molecules. It is possible that in the vapour phase they can effectively hold onto their latent energy—but once closely packed in liquid phase could they give off some of it? To test this hypothesis we immersed the T/R gap in a glass water tank. The motor employed for these tests was a high-speed 2 phase drag-cup motor (see FIG. **18** and associated description), wired in split-phase with two identical phase windings capacitatively balanced, and the galvanized iron plates had one foot square areas. The results are shown in Table 7 below, and clearly indicate that it is possible to tap—within the T/R cavity—the 'latent heat' of water in the liquid phase. As observed, immersion of the T/R cavity in water increased the motor output speed 22% (=12,117/9888)*100). This corresponds to a 50% increase in power output, from 18 W at 9888 rpm to 27 W at 12,117 rpm:

TABLE 7

	Pulse rate PPS	Motor rotation RPM M ± SEM	T/R distance cm
Direct from TC	0.3	8076 ± 89.3	NA
TC to T plate	0.5	9888 ± 78.7	NA
R plate	2.75	12117 ± 29.8	30
R plate	2.9	12203 ± 55.9	60

Thus the use of ion-containing water or other ion-containing aqueous liquid in the cavity promotes long distance propagation and a greater injection of latent and thermal energies in the receiver circuit. Such a result is not achieved if the cavity is filled with deionised water.

The preceding results lead therefore to the design of a presently preferred apparatus, based on these findings, for the conversion of massfree electric energy, "latent heat" energy and "sensible" heat energy into conventional electric energy, as shown in FIG. **16**, which integrates all the separate findings and improvements. The winding **6** of the Tesla coil at the bottom is driven in the usual manner employing a vibrator stage **2** to pulse the primary coil **4**. The distal pole of the secondary **6** is then connected to a circular metal plate T which is one end of an evacuated cylindrical cavity, connected to a vacuum pump or sealed at a desired pressure, or which forms a still containing water or other aqueous solution or liquid. This cavity constitutes the transmitter/receiver gap, and is therefore bounded by a dielectric envelope and wall structure **32**, with the circular receiver plate R as its top surface. In turn this plate R serves as the base of a conical Faraday cage **34**, preferably air-tight and at atmospheric pressure, but which could also be subject to evacuation, which conical structure carries at its apex provisions for a cold junction **45** and any possible enhancement of the same junction by surface application of different metallic conductors that may optimize the Peltier-Seebeck effect. The output from the cold junction where sensible thermal energy is added to the electrokinetic energy of charge carriers, is also the input to the distal end of the winding **6** of the chiral coil arrangement that sustains resonant capture of all three energy flows ((1) massfree electric waves of a longitudinal nature, (2) true "latent heat" or the intrinsic (thermal) potential energy, and (3) the thermokinetic energy of molecules, (i.e. "sensible" heat) and, placed in series with the input of the full wave divider **8**, **10**, feeds the circuit output from the series capacitors **12**, **14** grounded at their common tap. In the T/R gap, the transmitted electric longitudinal wave energy is captured along with any intrinsic potential energy shed by molecules caught in the field. Within the R element, expanded into an enclosure that guides "sensible" radiant heat, the latter is generated and then recaptured at the cold junction.

The apparatus consisting of the cylindrical T/R gap cavity and the contiguous conical cage is then preferably finished in gloss white and cylindrically enveloped within a matte black container **46** by effective thermal insulation **48**, the latter terminating at the height of the bottom disc T. Apparatus (not shown) may be provided to move the plate T vertically to adjust the T/R gap.

Another alternative embodiment of the apparatus is shown in FIG. **17**. Here the circuit driving the apparatus is as we have set forth in our prior patents, which employs an autogenous pulsed abnormal glow discharge tube **50** in the configuration shown, supplied by a battery pack DP through blocking diodes **52** and an RC circuit formed by resistor **54** and capacitor **56** to drive the primary 2 of a first Tesla coil to obtain at the distal pole of the secondary **6** the energy to be injected to plate T in the form of a central electrode of a coaxial vacuum chamber (sealed or not), of which the cylindrical metallic envelope forms the receiver plate R, the latter being placed centrally inside the conical cage **34** and contiguous with its walls and base. The top and bottom of the coaxial chamber carries suitable insulating discs, preferably with O-ring type fittings. Again, the apparatus is enclosed in insulation within a cylindrical container **46**, and the input into the capture circuit driven from the full wave divider is taken from the cold junction **45** at the apex of the air-tight cage. The output circuit is similar to that of FIG. **15**.

We have found however that even when the component values in the motor driver and motor circuits are carefully selected so that these circuits are co-resonant with the damped wave (DW) component of the motor driver pulses, the motor power output falls well short of that which should theoretically be attainable. In an endeavour to meet this problem, we replaced the squirrel cage type induction motor **44** by a drag cup motor of type KS 8624 from Western Electric in the expectation that the low inertia non-magnetic rotor would allow better response to the DW component. This motor is similar to one of the types used by Reich in his experiments. Although results were much improved they still fell short of expectations. Replacement of this motor by an inertially damped motor of type KS 9303, also from Western Electric, provided much better results as discussed below.

Fundamentally, the difficulties we encountered stemmed from the inability of motor couplings to respond efficiently and smoothly, and at the same time, to the pulse and wave components of DW impulses: that is, simultaneously to the high-intensity peak current pulses (the front end event), the DC-like component, and to the damped wave trains these cause, i.e. the pulse tails (or back end event) or AC-like component. This difficulty is present even when we just seek to run induction motors from the DW impulses of a Tesla coil, the very difficulty that led Tesla to abandon his project of driving a nonferromagnetic disc rotor mounted on an iron core bar stator with damped waves.

We believe that the key to the capture of the massfree energy flux output in electric form by Tesla transmitters, including any injected latent or thermal energy that have undergone conversion into electrical energy is to employ the tuned, unipolar, Y-fed, PAGD-plasma pulser driven split-phase motor drive we have invented (U.S. Pat. No. 5,416,391) in conjunction with an inertially damped AC servomotor-generator (see FIG. **18**): this has a motor shaft **64** which couples a drag-cup motor rotor **60**, preferably of aluminum, silver, gold or molybdenum, directly to a drag-cup generator rotor **62** that drives a permanent magnet (PM) flywheel **66**, freely rotatable in bearings **67**, that provides inertial damping. The shaft **64**, journalled by bearings **61** in the casing of the motor **44**, provides a power output through optional gearing **68**. The phase windings of the motor **44** are wound on a stator core **70** having concentric elements between which the rotor or cup **60** rotates. This structure makes it ideal for the capture of the DW impulses, whether sourced in the transmitter, amplified in the T/R cavity or sourced in the plasma pulser, all in synchrony. Effectively the motor couples the damping action of the drag-cup sleeve motor rotor, which action, as we have already found for the KS-8624 motors, is quite effective at absorbing the front-end DC-like event, with the inertial damping of the PM flywheel upon the drag-cup sleeve generator rotor, that in turn is quite efficient at absorbing the back-end AC-like wavetrain event.

The KS-9154 motor used by Reich was not an inertial damped AC drag-cup servomotor-generator. Had Reich succeeded in overcoming the limitations of his 2-phase OR Motor solution, as we have now shown it is possible to do—by applying the Function Y circuit to the PAGD split-phase motor drive which we invented, his motor would have suffered the same limitations which we encountered with the KS 8624 motor.

Any motor, by itself, has an internal or inherent damping whereby the acceleration only vanishes when the rotor is running at constant speed. For motors that operate on the basis of the drag principle, where the asynchronous slip is actually constitutive of the motor action, by inducing eddy currents in the rotor, the inherent damping is always more pronounced than for other induction motors. The damping or braking torque is produced when a constant current flows through a rotating drag disc or cup.

Aside from this inherent braking, dampers can also be applied to servomotors to further stabilize their rotation. They absorb energy, and the power output and torque of the motor is thereby reduced. Optimal operation of servomotors requires both rapid response on the part of the rotor to changes in the variable or control phase, and a stable response that is free from oscillation, cogging and overshooting. The rapid response is assured by employing low inertia rotors, such as drag-cups or cast alloy squirrel cages, and the overshooting and oscillation are reduced to a minimum by damping or a retarding torque that increases with increasing motor speed. Typically, in a viscous-damped servomotor, the damper is a drag-cup generator mounted rigidly on the shaft of the motor rotor, and the generator drag-cup rotates against the stator field of a static permanent magnet field. The generator develops a retarding torque directly proportional to speed, and the energy absorbed by the damper is proportional to speed squared. The damping can be adjusted and, as it increases, the same amount of input power yields lower torque and motor speeds. Inertial-damped servomotors differ from viscous damped motors in that the permanent magnet stator of the drag-cup generator is now mounted in its own bearings, either in the motor shaft or on a separate aligned shaft, forming a high-inertia flywheel. This means that, whereas the motor rotor always experiences a viscous damping in viscous-damped servomotors, in inertial-damped servomotors the drag cup motor rotor only experiences a viscous damping while accelerating the flywheel, with the damping torque always opposing any change in rotor speed. Once the flywheel rotates synchronously with the rotor, all damping ceases. Note that this viscous damping is carried out via the coupling of the drag-cup generator rotor, rigidly affixed to the motor rotor, to the PM flywheel, so that their relative motion generates the viscous torque proportional to the relative velocity. Use of drag-cup sleeve rotors in inertially damped servomotors was largely supplanted by squirrel cage rotors once the latter became produced as cast alloy rotors. Since inertially damped motors can be used in open and closed-loop servo applications, and present better stability—even in the presence of nonlinearities—and higher velocity characteristics than other induction motors do (Diamond, A (1965) "Inertially damped servo motors, performance analysis", Electro-Technology, 7:28–32.), they have been employed in antenna tracking systems, stable inertial-guidance platforms, analog to digital converters, tachometers and torque tables. The typical operation of an inertially damped servomotor is as follows: with the reference phase fully excited, the motor rotor—fixedly linked to the generator rotor, as well as the flywheel, remain immobile; once power is applied to the control phase, the motor rotor immediately responds but the flywheel remains at rest. However, as the drag-cup generator **62** is forced to move through the permanent magnetic field of the flywheel, it creates a drag torque that slows down the attached motor rotor proportionally to the acceleration that it imparts to the flywheel that it now sets into motion, thus creating the viscous damper. As the flywheel accelerates, the relative speed of the motor with respect to the flywheel, as well as the damping torque, decrease until both motor and flywheel rotate synchronously and no damping torque is exercised—at which point the drag on the motor cup exerted by the generator cup is negligible. The KS-9303 motor is an inertial damped servomotor but is differentiated with respect to other inertially damped motors, in that (1) it employs a drag-cup sleeve motor rotor made of aluminum, very much like that of the KS-8624, but with slightly altered dimensions and with a shaft extension for the drag-cup copper generator rotor, and (2) the moving flywheel structure was journalled on a separate, fixed shaft, as already described with reference to FIG. **18**. Now, in principle, even application of minimal damping decreases motor efficiency, resulting in diminished torque and speed. Whether the inertial-damped motor has a drag-cup rotor, a sleeve rotor or a squirrel cage rotor, the damping increases the rotor slip. Laithwaite considers drag-cup motors as being "dynamically inferior to their cage counterparts" (Laithwaite, ER

(1957) "Induction machines for special purposes", London, England, p. 323.). If we now add a viscous damping and retarding torque, we should not be able to get much more than a 55% efficiency in the best of conditions. On the other hand, the inertial damping arrangement described will only abstract or supply energy when the motor rotor is accelerating or decelerating relative to the flywheel.

These drag-cup motors, whether inertially damped or not, develop a constant torque at constant rpm for a given supply frequency and a suitable phase shift capacitance. For each frequency the motors respond to, there is an optimum resonant split-phase capacitance, but other values nearby are still suited for operation, and for each value of capacitance, there is an optimal frequency the motors respond to. For example the KS-8624 motor responds best at 450 Hz when a 1 μF capacitance is employed, responds best at 250 Hz when a capacitance of 10 μF is employed, and responds best at 60 Hz, when a capacitance of 100 μF is employed. As the capacitance increases, the resonant CW frequency of the motor is displaced to lower values. If we fix the capacitance at a value (e.g. 10 μF) suitable for testing the frequency response at a fixed voltage of 12 VAC, the observed result for both the KS-8624 and KS-9303 motors show a response distribution of the motor rotary velocity that has an identical peak at 250 Hz for both motors, with the response decreasing to zero smoothly on both sides of the peak. These results indicate that, when wired as a split-phase motor, the motor rotary velocity varies not as a function of voltage or current, but as a function of frequency when the phase-splitting capacitance is fixed within a suitable range, there being an optimal frequency mode for each value of suitable capacitance, with lower values of capacitance favouring higher frequency modes. For a given frequency and capacitance, the motor rotary velocity remains essentially constant and independent from voltage and current input, and thus at a plateau. Torque, in the same circuit arrangement, follows exactly the same pattern as rotary velocity, as a function of input frequency at a fixed potential. Torque is linearly proportional to rpm in these motors when they are split-phase wired, and rpm linearly proportional to CW frequency, which makes them ideal for experimentation and determination of power output computations. Moreover, since these are drag machines, the slip itself determines the rotor currents and these are susceptible of tuning such that their retardation and relative position in the field can find resonant modes for varying CW frequency and capacitance.

In the circuit of FIG. **17** when using the KS 9303 motor, the inertial damping of the flywheel coupling retards the motor rotor currents sufficiently to allow them to build up torque, with the entire motor assembly serving as the preferred sink for all of the energy, massfree and massbound, captured by the receiving coil circuit with a drawing action established by the motor on the circuit, and providing satisfactory absorption by an inertial damper of the combined, synchronized, damped wave impulses, those occurring at a low frequency as a result of the firing of the PAGD reactor, and those occurring at a higher superimposed frequency—sourced in the transmitter circuit and picked-up by the receiver plate and coil. The action of each DW impulse train itself comports two different events, the DC-like auto-electronic-like discontinuity that sets the motor into action and initiates the rotor currents, and the AC-like damped wavetrain that supports the consistency of those rotors. The concentration of current required to kickstart the motor is provided by the DW impulses of the PAGD reactor, whereas, once the motor is in motion, and particularly, once it is stabilized by the flywheel, the cumulative action of the higher frequency DW impulses makes itself felt by accelerating the rotor to an optimum rotary velocity.

For the next series of tests we employed the basic circuit diagram of the improved motor shown in FIG. **19**. The transmission station is the typical Tesla transmitter with a line-fed, 60 Hz vibrator stage. At the line input to the first stage, we place a calibrated AC wattmeter (Weston Model 432), and a Beckman 330B rms ammeter in series with the hot lead, we set the vibrator stage for 41 clicks, consuming between 28.5 and 35 W, as depending upon circumstances yet to be narrated. This consumption was confirmed by driving the coil from an inverter powered by a 12 volt battery. The inverter consumes 2.16 watts, and is 90% efficient. The total consumption from the battery was 42 watts (12V at 3.5 A); once the 2.16 watts is deducted and the efficiency taken into account, we obtain the same 36 W (vibrator stage at max, i.e. 47 clicks, in this experiment). The T/R gap is adjusted to 3", and 2 square ft plates are employed. Transmitter and receiver coils are tuned, and so are the plate capacitances, to 250 kHz, also the capacitances of the Function Y circuit connected at the output of the receiving coil. The rectified voltage and current generated by the transmitter secondary and by the transmitter plate was ascertained with a coil-tuned wave-divider (Function Y) circuit by loading it with different resistive values. The results constitute a measure of the massbound electrical power output directly from the transmitter apparatus. The same method was employed to ascertain the voltage, current and power of the massbound charges circulating in the receiving plate and coil circuit. The results are shown in Table 8 below:

TABLE 8

Massbound currents rectified by Function Y at the output of the Tesla transmitter, transmitter plate and receiver plate, as a function of the bleeding resistance employed in each of the function Y arms

	VDC (kilovolts)	ADC (amp)	WDC (watts)	R/arm (Mohm)
Direct from 20°	42–50	$3*10^{-5}$	1.26–1.5	500
From 2° (T) plate	26	$2*10^{-5}$	0.52	500
From 2° (R) plate	15.1	$1.25*10^{-5}$	0.189	500
Direct from 2°	20.4	$3.4*10^{-4}$	6.936	50
From 2° (T) plate	15.2	$2.4*10^{-4}$	3.648	50
From 2° (R) plate	9	$1.2*10^{-4}$	1.08	50
Direct from 2°	3.3	$1.75*10^{-3}$	5.775	1
From 2° (T) plate	3.5	$2*10^{-3}$	7.0	1
From 2° (R) plate	2.95	$1.6*10^{-3}$	4.72	1

The results indicate that the highest massbound power assembled by the secondary transmitter circuit does not exceed 7 watts—and this is directly output from the secondary **26** when the load is 50 Mohm, or from the transmitter plate when the load is 1 Mohm. The massbound electric power emulated by the receiving circuit (plate, coil and Function Y without the plasma pulser circuitry) never exceeds the massbound electric power output directly by the transmitter, and peaks when the resistive load value (1 Mohm) approaches the pre-breakdown resistance range of the vacuum tube, at 4.72 W. These findings then indicate that when the transmitter circuit is consuming a maximum of 35 W, a typical output from the secondary of the transmitter is 7 W, and at 3" of distance within the proximal field of the latter, the pick-up by a tuned receiver will be on the order of 5 W of massbound current duplicated within the receiving coil. The loss in the first stage is therefore on the order of sevenfold. Continuing with the description of the circuit of FIG. **19**, a 128 cm^2 plate area, 6 cm gap PAGD reactor is employed, connected as described in our prior art to a high-vacuum rotary pump (Correa, P & Correa, A (1995) "Energy conversion system", USPTO U.S. Pat. No. 5,449, 989.). Pressure readings were obtained with a thermocouple gauge during the operational runs. The KS-9303 motors to be tested are then connected to the PAGD reactor in the usual capacitatively-coupled, inverter fashion described in our prior art (Correa, P & Correa, A (1995) "Electromechanical transduction of plasma pulses", USPTO U.S. Pat. No. 5,416, 391.). Their rpm is detected by a stroboscopic tachometer and fed to a Mac Performa 6400 running a motor algorithm program calculating the power output. Motor measurements were made at five minutes into each run for the unloaded motors, and at ten minutes for the inertially damped motors. All experiments were carried out in the same work session. The experimental determination of the continuous rotary power output as a function of the reactor pulse rate confirmed that the improved circuit develops maximum rotary capture of the massfree energy in the receiver circuit at the lowest rates of pulsation, just as we have previously found for the conversion system of U.S. Pat. No. 5,449,989. Furthermore, the data showed that even motors of type KS-8624 are able to output power mechanically in excess of the massbound power output by the transmitter (7 W) or captured by the receiver (5 to a max of 7 W), once the PAGD rate decreases to 1.5 PPS. Such an anomaly can only be explained by the system having become able to begin capturing the massfree energy flux in the receiver circuit that we know already is output by the transmitter circuit. But this excess mechanical power is still, and clearly so, less than the power input into the transmitter. It represents a power gain with respect to the secondary, but a loss with respect to the primary. The full breadth of the capture of the massfree electric energy flux circulating in the receiver circuit is not seen until the motors are resonantly loaded because they are inertially damped.

The KS-9303 motors, once inertially damped, and thus loaded, are able to to recover enough power from the massfree energy field to develop a mechanical power, not just greatly in excess of the massbound power of the secondary, but also greatly in excess of the massbound power input to the vibrator stage and the primary, at 28 to 35 W. Once the pulse rate approaches the same 1.5 PPS marker, mechanical power in excess of the massbound electric power input to the primary becomes evident, peaking at nearly three times that input. In fact, the highest output recorded was also obtained with the lowest input to the transmitter circuit, the highest exact coefficient observed in this experiment being 100.8 W/28 W=3.6. Furthermore, with respect to the secondary massbound output, the same mechanical rotary output represents a much greater overunity coefficient of performance, on the order of 14.4 times greater. This is at least partly the result of the receiver and motor capture of the massfree electric energy output by the transmitter, and may be partly the result of massfree energy engrafted by the PAGD regime in the PAGD reactor. Reviewing the mechanical power output results as a function of increasing vacuum in the PAGD reactor and at different output power levels, any motor performance below the 5–7 W limit of the traditional massbound output power of the secondary represents an output mechanical power loss with respect to both the massbound secondary output and the massbound primary input. All the results for pressures down to 0.03 Torr fall in this category, and thus represent a very inefficient coupling to the PAGD regime. Any motor performance between 7 W and 28–35 W represent a loss with respect to the electrical power input to the transmitter system, but a net gain of power with respect to the massbound secondary power output. None of the non-inertially damped motors tested were able to perform outside of this area, under the test conditions. With more efficient primary to secondary couplings in the transmitter station, however, one could advantageously employ these motors alone to extract some of the massfree power of the secondary or to operate them in enclosed vessels without conventional external electrical connections. To reach satisfactory levels of recovery of massfree energy, one must dampen the superimposed DW impulses. Hence, all results showing outputs in excess of 35 W were obtained using the inertially damped KS-9303 motors, and represent a net overunity power gain over both the power input to the primary and the massbound power output by the secondary, or the massbound power emulated by the receiver circuitry. This happens when the PAGD pulse rate falls to 2 PPS, with the rotary power output steeply increasing as the rate falls to 1 PPS.

One of the interesting features of the motor circuitry we have proposed is that it can operate with pulsed plasmas in both the TRD and the AGD regions, the least efficient response occurring in the NGD region near the Paschen minimum. One might think that the voltage depression would allow increased current intensity supplied to the motors, but in fact that is not observed, with the flashing of the NGD yielding erratic oscillations and low values of current. In keeping with the notion that the TRD plasma is mainly composed of lagging positive ions, whereas the PAGD plasma is mostly an electron plasma, the observed direction of rotation of the motors is opposite in the TRD region to that of the AGD region. The NGD region therefore marks the depression where the velocity vectors change direction. In the second or PAGD region, motor operation is very quiet, unlike what is observed in the TRD region.

Part and parcel of the tuning of the circuit components is the selection of the optimum capacitances employed to couple the PAGD reactor to the motor circuit and split the phase to feed the auxiliary winding of the motor. We have experimented with capacitances ranging from 0.5 to 100 μF, and found that best results—for the specific circuit in question (including the characteristics of the transmission)—were such that the optimal value of the PAGD coupling capacitance lay near 4 pF, and the phase splitting capacitance, near 1 to 4 μF, depending upon weather conditions—in good weather days lower capacitance values can be implemented, and in bad weather days higher capacitances are needed. For ease of comparison in demonstrating the need to tune the circuit by employing optimal capacitances in those two couplings (reactor to motor, and motor phase coupling), we employed the same capacitances in both circuit locations. A comparison of tests using 1 and 4 μF values shows the difference caused by changing those capacitances from their optimal value: across all discharge regions of the pressure range that was examined, the four motors tested operating with greater motor speeds when the capacitances are set to 4 μF rather than to 1 μF. The less efficient performance obtained with 1 μF capacitances fits the inverse correlation of pulse power with increasing pulse frequency, such as we have found for the PAGD regime. This is made evident by a comparison of rpm versus pulse rate for the two capacitance values being considered. They demonstrate the higher pulse rates observed with the lower capacitance, that correlate with the slower motor speeds, and result in lower efficiency of the motor response. The results equally indicate that low capacitance values increase the pulse rate, but if this increase is out of tune with the rest of the circuit values, it results in power waste because it imposes a rate that is not optimal.

We have also experimentally determined that the efficiency of the system is affected by external weather conditions, higher efficiencies being noted on a fine bright day than under poor weather conditions even though the apparatus is not exposed to such conditions. This may reflect a diminution under poor weather conditions of latent massfree energy that can be taken up by the system.

The observed high efficiency of circuits including inertially damped motors indicates that the phenomenon does not reduce to a mere optimal capture of DC-like pulses produced by the reactor in what is essentially an AC motor circuit. Effectively, the pulsed plasma discharge deploys a front-end, DC-like pulse, or discontinuity, but this is followed by an AC-like damped wave of a characteristic frequency (having a half-cycle periodicity identical to that of the front-end pulse) to which the motor circuit also responds. Moreover, the massfree electric radiation from the transmitter circuit itself induces—in the receiver antenna, coil and circuit, and in the reactor discharge itself—the train of finer damped wave impulses responsible, after conversion through the wave-divider, for the massbound rectified current that is employed to charge the plasma reactor to begin with. Serving as trigger of the plasma discharges in the reactor are the DW impulses circulating in the receiver circuit, such that the two different lines of DW impulses, in the receiver circuit (for example 120 PPS for the pulses and 154 kHz for the waves) and from the reactor, are synchronized by interpolated coincidences, since their pulse and wave frequencies are different. Ideally, these two superimposed DW frequencies are harmonics or made identical. The receiver stage involves capture of the massfree electric energy received from the transmitter, duplication of the massbound current in the receiver coil, and injection of latent and sensible thermal energy in the T/R gap cavity that augments the emulated massbound current. The massbound current is employed to charge the wave-divider capacitance bridge and therefore the reactor. In turn, the plasma pulses from the reactor are superimposed with the DW impulses from the receiving coil, and together they are coupled to the split-phase motor drive. Hence the first receiver stage employs the totality of the energy captured in the T/R gap cavity-massfree electric energy transmitted by the T plate, latent and sensible thermal energy injected at the surface of the R plate—and produces in the receiving coil a massbound current comparable to that assembled in the transmitter coil by the action of the primary. The massbound current is stored in the wave-divider bridge and used to drive the plasma reactor in the PAGD region. Subsequently, the autogenous disruptive discharge that employs a substantial electron plasma generates both a concentrated, intense flux of massbound charges in the output circuit, and a massfree oscillation of its own. The damped motor is therefore fed directly with (1) the intense massbound current output from the reactor; (2) the pulse and wave components of the massfree electric energy captured by the receiver plate and coil (and matched by conduction through the earth), and which are gated through the wave-divider and the reactor for the duration of the PAGD channel; and (3) any massfree latent energy taken up from the vacuum by the PAGD event. Once the motor is set into motion, and is resonantly loaded with an inertial damper, we believe that it will also respond to the much weaker DW impulses captured by the receiver, since these impulses encompass both a DC-like front end—further enhanced by analytic separation through the wave-divider—and a damped wave at 154 kHz. Essentially the DW impulses that are ultimately sourced in the transmitter -and received unipolarly through the T/R gap—have sufficient DC-like potential (plus all the other requisite physical characteristics, such as frequency) to contribute directly to the motor response, once the motor has gained substantial speed (for they lack the current to set it into motion, one of the contributions from the plasma pulser). This is the case, provided that the motor itself is suited for absorption of both DC-like pulses and AC-like damped waves, which is precisely the case with motors of the type shown in FIG. **18** since the inertia of the flywheel is overcome through homopolar absorption of the damped oscillations simultaneously in the motor drag-cup rotor and in the generator drag-cup rotor.

We also tested these inertially damped motors in the traditional DC power supply-driven PAGD circuit we have taught in our previous patents, that is, circuits with an overt HV DC power source, and thus in the absence of any Function Y circuit or transmitter circuit. Here then, only the DW impulses generated by the PAGD reactor can account for the motor response. The tube employed (A31) had an area of 256 cm^2, and a gap distance of 4 cm. Coupling capacitances employed were 4 μF for the inverter coupling, and 1 μF for the split phase motor coupling. The DC power supply delivered up to 1 ampere of current between 150 and 1,000 VDC, and the ballast resistor was adjusted to 215 ohms. Having determined the basic physical characteristics of the reactor's behaviour in the circuit under consideration, we conducted our experiment in the PAGD region. We chose a pressure of 0.6 Torr, just off from the Paschen minimum, as we intended to benefit from the lower sustaining voltage that it affords. The experiment basically consisted of increasing the sustaining voltage at this fixed pressure in the PAGD regime, and measuring the diverse physical parameters of the circuit and motor response in order to ultimately ascertain the difference between the input electric DC power and the output mechanical rotary power. We first looked at how the motor rpm response varied as a function of the sustaining voltage (V_s): the results illustrate the importance of starting close to the Paschen minimum in the pressure scale, since the KS-9303 motors reach plateau response (at 17,000 rpm) when the reactor output voltage nears 450V. Any further increase in potential is simply wasted. Likewise, the same happened when we measured motor speed as a function of increasing peak DC current, plateau response being reached at 0.1 ADC. Again, any further increase in current is wasted. Essentially then, the optimal power input to the reactor when the latter's output is coupled to the motor, lies around 45 watts. This is a typical expenditure in driving a PAGD reactor. As for pulse rate we once again find a motor response that is frequency proportional in the low frequency range, between 10 and 40 PPS (all pulse rates now refer solely to PAGDs per sec), but once rates of >40 PPS are reached, the response of the motor also reaches a plateau. The observed increment in speed from 40 to 60 PPS translates only into an increase of 1,000 RPM, from 16,000 to 17,000 RPM. So, we can place the optimal PAGD rate at ca 40 PPS. The DC electric power input to drive the PAGD reactor was next compared to the rotary mechanical power output by the inertially loaded motor, driven in turn by the reactor. This comparison was first carried out with respect to the PAGD rates. The motor response far exceeds the conventional input power, indicating that the whole system can be tuned to resonance such that optimal power capture inside the reactor takes place, the critical limit rate lying at around 60 PPS, when the motor response is firmly within the pulse response plateau. At this juncture, the break-even efficiency for the measured rates of energy flux over time reach 700% (overunity coefficient of 7), in keeping with the observations and the values we have made in the PAGD conversion system. In the proportional part of the curve, before the plateau is reached, even greater rates of break-even efficiency—up to >1,000% were registered. These results constitute the first time we have been able to confirm the presence of output energy in excess of break-even over conventional mass-bound energy input in the PAGD inverter system, and the results are comparable to what we have observed and previously reported for the PAGD converter system. At pulse rates greater than 60 PPS a greater input power results in decreased efficiency, also translated into a noticeable heating of the reactor and motor. And this is all the more remarkable as experiments we have conducted with inductive tuning of PAGD reactors, or employing PAGD reactors as replacements for the primaries of Tesla coil assemblies, and still, more recently, with the PAGD inverter circuit driving motors, have all shown that it is possible to operate these reactors with minimal mirroring and heating, preserving essentially the cold-cathode conditions and yet focusing the plasma column so that deposition on the insulator is negligible. It appears that above a certain threshold of optimal efficiency, surplus input energy is just dissipated thermally by both the reactor and the motors.

It should be understood that the above described embodiments are merely exemplary of our invention, and are, with the exception of the embodiments of FIGS. **16**–**19** designed primarily to verify aspects of the basis of the invention. It should also be understood that in each of these embodiments, the transmitter portion may be omitted if an external or natural source of Tesla waves is available, provided that the receiver is tuned to the massfree radiation mode of the source. For example if solar radiation is available in which the massfree component has not interacted with the earth's atmosphere (as in space applications), the receiver is tuned to the voltage wave of the massfree radiation sourced in the sun, e.g. by using a Tesla coil in the receiver constructed to have an appropriate voltage wave close to the 51.1 kV characteristic of such radiation.

We claim:

1. A device for the conversion of massfree radiation into electrical or electrokinetic energy comprising a transmitter of massfree electrical radiation having a damped wave component, a receiver of such radiation tuned to resonance with the damped wave frequency of the transmitter, a co-resonant output circuit coupled into and extracting electrical or electrokinetic energy from the receiver, and at least one of a transmission cavity between the transmitter and the receiver, a full-wave rectifier in the co-resonant output circuit, and an oscillatory pulsed glow discharge device incorporated in the co-resonant output circuit.

2. A device according to claim **1**, wherein the output circuit comprises a full wave rectifier presenting a capacitance to the receiver.

3. A device according to claim **2**, wherein the output circuit comprises an electric motor presenting inductance to the receiver.

4. A device according to claim **3**, wherein the motor is a split phase motor.

5. A device according to claim **4**, wherein the motor is a drag motor having a non-magnetic conductive rotor.

6. A device according to claim **5**, wherein the motor has inertial damping.

7. A device according to claim **6**, wherein the motor has a shaft, a drag cup rotor on the shaft, and inertial damping is provided by a further drag cup on the shaft.

8. A device according to claim **6**, wherein the transmitter and receiver each comprise at least one of a Tesla coil and an autogenous pulsed abnormal glow discharge device.

9. A device according to claim **8**, wherein the transmitter and receiver both comprise Tesla coils, and further including a transmission cavity which comprises spaced plates connected respectively to the distal poles of the secondaries of Tesla coils incorporated in the transmitter and receiver respectively.

10. A device according to claim **9**, wherein the plates are parallel.

11. A device according to claim **9**, wherein the plates are concentric.

12. A device according to claim **9**, wherein at least the receiver comprises a Tesla coil driving a plasma reactor operating In PAGD (pulsed abnormal glow discharge) mode.

13. A device according to claim **12**, wherein the transmitter and receiver both comprise Tesla coils, and further Including a transmission cavity which comprises spaced plates connected respectively to the distal poles of the secondaries of Tesla coils incorporated in the transmitter and receiver respectively.

14. A device according to claim **1**, wherein the transmitter and receiver each comprise at least one of a Tesla coil and an autogenous pulsed abnormal glow discharge device.

15. A device according to claim **1** wherein a transmitter/receiver cavity is present and filled with an aqueous liquid.

16. A device for the conversion of massfree radiation into electrical or electrokinetic energy comprising a receiver of such radiation from a source of massfree electrical radiation having a damped wave component, the receiver being tuned to resonance with the damped wave frequency of the source, a co-resonant output circuit coupled into and extracting electrical or electrokinetic energy from the receiver, and at least one of a transmission cavity between the source and the receiver, a full-wave rectifier in the co-resonant output circuit, and an oscillatory pulsed glow discharge device incorporated in the co-resonant output circuit.

* * * * *

US007091918B1

(12) **United States Patent**
Bhansali et al.

(10) **Patent No.: US 7,091,918 B1**
(45) **Date of Patent: Aug. 15, 2006**

(54) **RECTIFYING ANTENNA AND METHOD OF MANUFACTURE**

(75) Inventors: **Shekhar Bhansali**, Tampa, FL (US); **Kenneth Buckle**, Tampa, FL (US); **D. Yogi Goswami**, Gainesville, FL (US); **Elias Stefanakos**, Tampa, FL (US); **Thomas Weller**, Lutz, FL (US)

(73) Assignee: **University of South Florida**, Tampa, FL (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **10/904,121**

(22) Filed: **Oct. 25, 2004**

Related U.S. Application Data

(60) Provisional application No. 60/514,327, filed on Oct. 24, 2003.

(51) **Int. Cl.**
H01Q 13/10 (2006.01)

(52) **U.S. Cl.** **343/767**; 343/850; 343/860

(58) **Field of Classification Search** 343/767, 343/850, 860
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,760,257	A	9/1973	Fletcher et al.	
3,989,994	A	11/1976	Brown	
4,079,268	A	3/1978	Fletcher et al.	
4,251,679	A	2/1981	Zwan	
5,043,739	A *	8/1991	Logan et al.	343/701
5,068,669	A *	11/1991	Koert et al.	343/700 MS
6,140,924	A *	10/2000	Chia et al.	340/572.5
6,369,759	B1	4/2002	Epp et al.	

FOREIGN PATENT DOCUMENTS

WO	97/02623	*	1/1997

OTHER PUBLICATIONS

Hagerty et al., An Experimental and Theoretical Characterization of a Broadband Arbitrarily-Polarized Rectenna Array, IEEE MTT-S Digest, 2001, 1855-1858.

McSpadden et al., A High Conversion Efficiency 5.8 GHZ Rectenna, IEEE MTT-S Digest, 1997, 547-550.

McSpadden et al., Theoretical and Experimental Investigation of a Rectenna Element, IEEE Transactions on Microwave Theory and Techniques, 1992, 2359-2366, 40(12).

Suh et al., A Novel Dual Frequency Rectenna for High Efficiency Wireless Power Transmission at 2.45 and 5.8 GHZ, IEEE MTT-S Digest, 2002, 1297-1300.

Yoo et al., 35 GHZ Rectenna Implemented with A Patch and A Microstrip Dipole Antenna, IEEE MTT-S Digest, 1992, 345-348.

* cited by examiner

Primary Examiner—Trinh Vo Dinh
(74) *Attorney, Agent, or Firm*—Molly L. Sauter; Smith & Hopen, P.A.

(57) **ABSTRACT**

In accordance with the present invention, an aperture rectenna is provided where the substrate is transparent and of sufficient mechanical strength to support the fabricated structure above it. An aperture antenna is deposited on the transparent substrate and a metal-insulator-metal (MIM) diode is constructed on top of the aperture antenna. There is an insulating layer between the aperture antenna metal and the metal ground plane optimized to maximize the collection of incident radiation. The top of the structure is capped with a metal ground plane layer, which also serves as the DC connection points for each rectenna element.

36 Claims, 6 Drawing Sheets

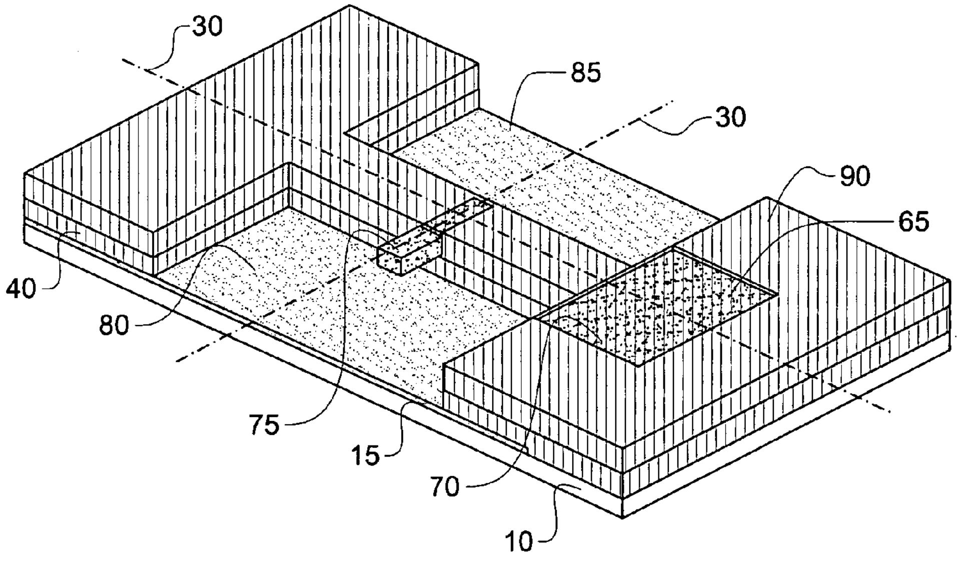

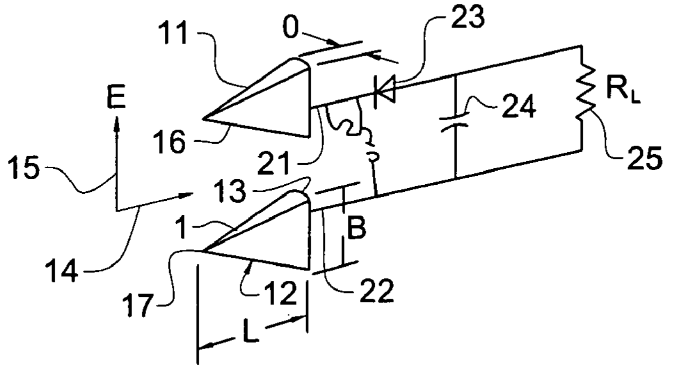
0
23
11
E
24
R_L
16
21
25
15
13
B
1
14
12
22
17
L

FIG. 1A
Prior Art

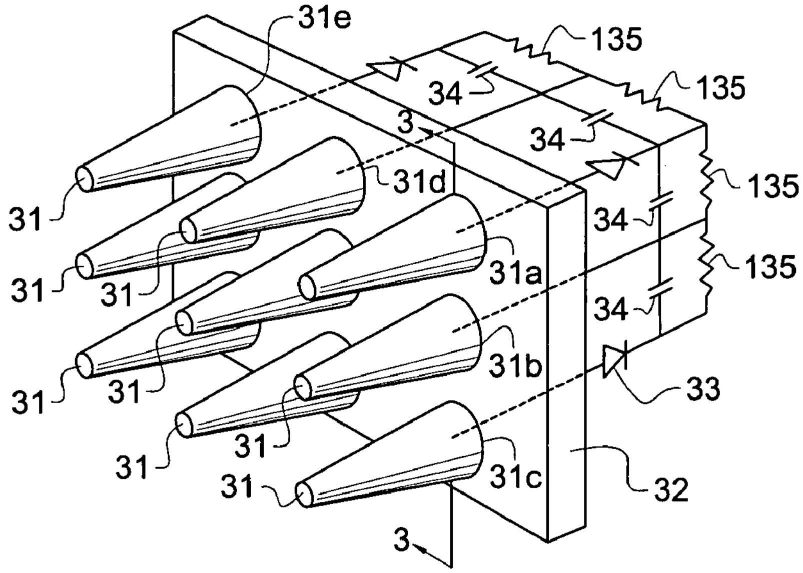
31e
135
34
135
3
34
135
31
31d
34
31
31
31a
135
34
31b
31
31
33
31
31
31c
31
32
3

FIG. 1B
Prior Art

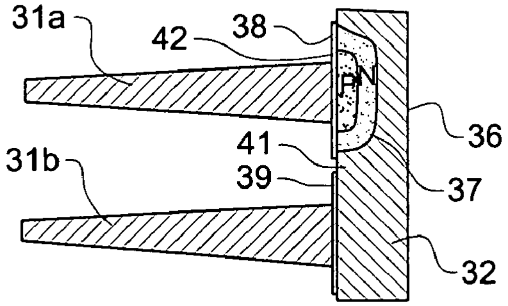
31a
42
38
P
N
36
41
39
37
31b
32

FIG. 1C
Prior Art

FIG. 2

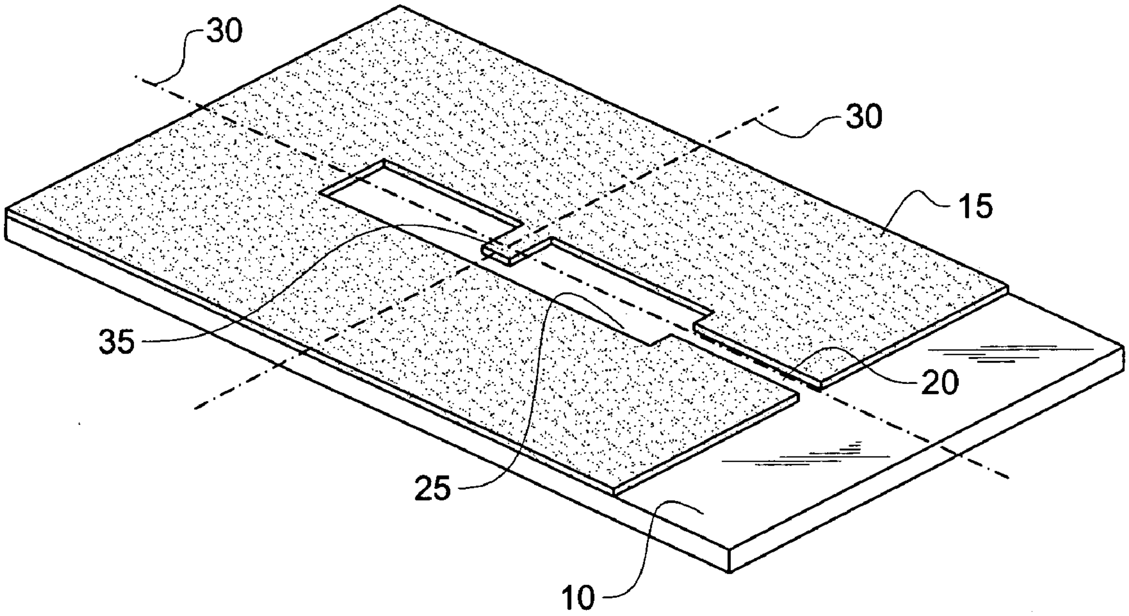
30
30
15
35
20
25
10

FIG. 3

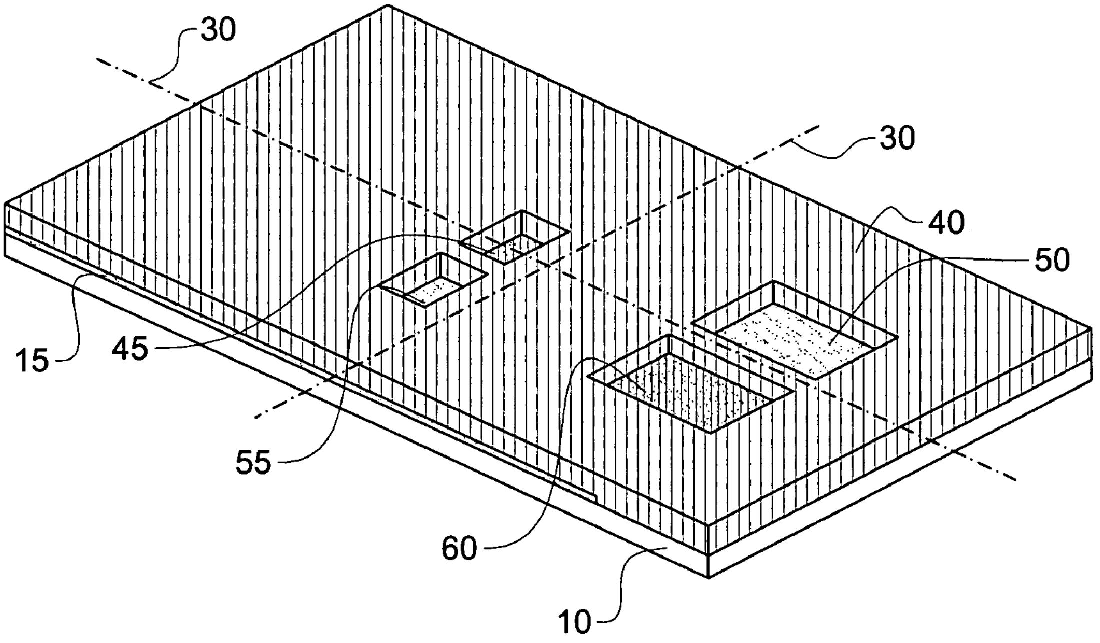
30
30
40
50
15
45
55
60
10

FIG. 4

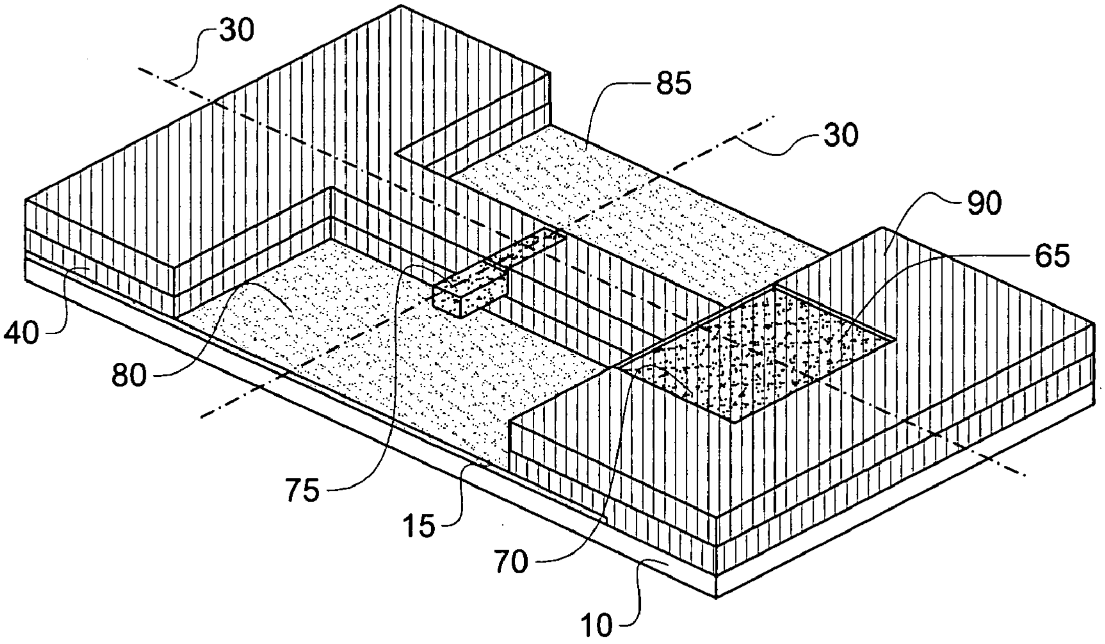
30
85
30
90
65
40
80
75
15
70
10

FIG. 5

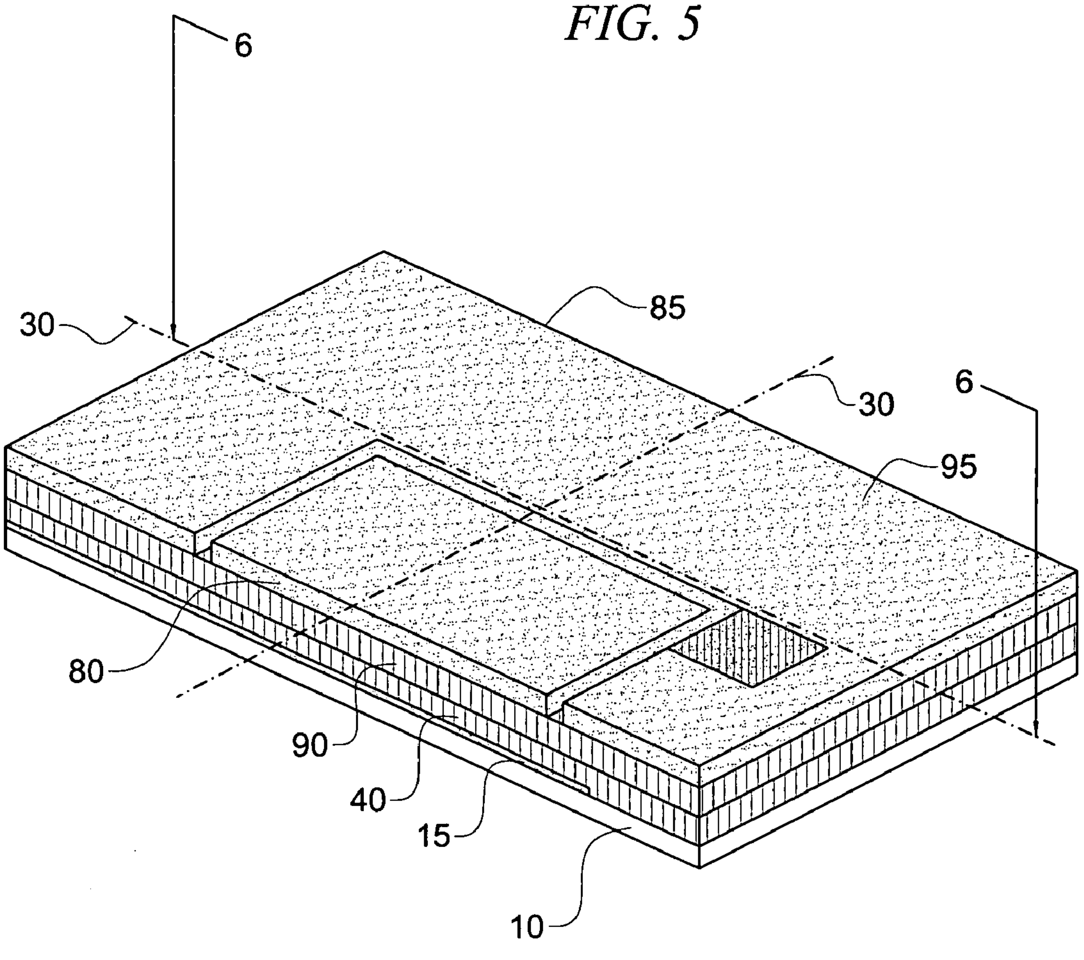
6
30
85
30
6
95
80
90
40
15
10

FIG. 6

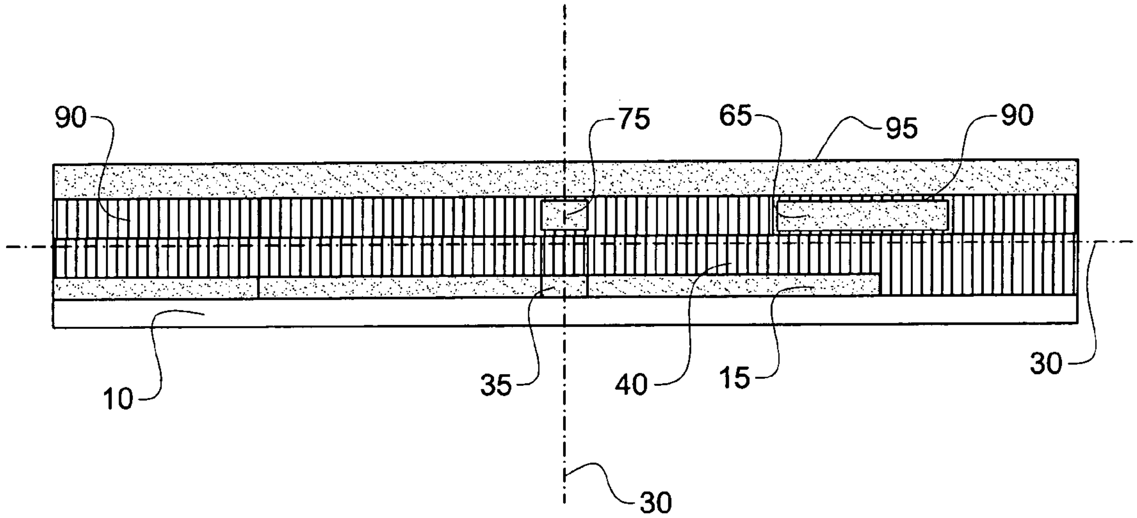
90
75
65
95
90
10
35
40
15
30
30

RECTIFYING ANTENNA AND METHOD OF MANUFACTURE

CROSS REFERENCE TO RELATED APPLICATION

This application claims priority to Provisional U.S. Patent Application No. 60/514,327, "An Aperture Rectenna for Converting High Frequency Radiation into DC Power", by the same inventors, filed Oct. 24, 2003.

STATEMENT OF GOVERNMENT INTEREST

Research relating to the present application was supported by NASA/FSEC. Accordingly, the U.S. federal government may have rights in the present invention.

BACKGROUND OF THE INVENTION

Light energy is characterized by a dual nature both from a quantum point of view as photons and from a wave point of view as randomly polarized electromagnetic radiation with wavelengths from 400 nm to 700 nm. If the ultraviolet and infrared portion of the spectrum is included, the range of wavelengths is extended at both extremes. Presently all practical solar cell energy collection schemes use the photon nature of light in the explanation of operation. The conversion of solar energy to electrical energy using the photovoltaic effect depends on the interaction of photons with energy equal to or greater than the band-gap of the material. With continued research the maximum amount of energy captured using this mechanism is estimated to be around 30%. This is equivalent to an efficiency of 30%.

Another mechanism known in the art for converting solar energy to electrical energy is the rectenna. Rectenna technology relies on the electromagnetic nature of radiation and is not limited by the band-gap of the rectifying material. Advantages exist in the rectenna field because the rectenna approach is not fundamentally band-gap limited. Also, at microwave frequencies, 2.45 GHz, this concept has been demonstrated to be approximately 90% efficient.

Prior art exists in the field of rectenna technology. The idea of direct electromagnetic conversion of solar energy has been explored as early as 1973. The concept envisioned small cone type antennas with rectifiers mounted behind the cones and the cones arranged in an array as illustrated in FIG. **1**. The work preceded the revolution in micro and nano scale manufacturing and was never realized. Other more recent researchers have replaced the cones with planar antenna elements such as half-wave dipoles, bow tie, and spiral antennas.

The previous efforts in researching rectennas for solar energy conversion have concentrated on either building the rectifiers on a supporting substrate and then building the antennas on top of the rectifying layer, or fabricating the antenna structure and the rectifier on a planar surface. The fabrication processes known in the art for rectenna elements rely on hybrid systems that incorporate different technologies for each component of the rectenna to optimize the system performance. While the incorporation of hybrid technologies often allows for a shorter development time, compromises must be made to ensure that each component can be fabricated without damaging components already existing on the substrate.

By contrast, monolithic device fabrication allows for all components of a system to be constructed on a single substrate utilizing one technology. Monolithic devices are typically more compact than hybrid devices, and more reliable due to the reduction in interconnections required. Additionally, once the fabrication process for a monolithic device has been developed, they can be manufactured more efficiently since less assembly is required. Monolithic technology minimizes relatively long, high-resistivity connections and high loss metal interconnects.

Accordingly, what is needed in the art is a monolithic rectifying antenna providing high efficiency solar energy conversion with improved reliability and manufacturability.

However, in view of the prior art considered as a whole at the time the present invention was made, it was not obvious to those of ordinary skill in the pertinent art how the identified need could be fulfilled.

SUMMARY OF THE INVENTION

The longstanding but heretofore unfulfilled need for a monolithic rectifying antenna providing high efficiency solar energy conversion with improved reliability and manufacturability superior to rectifying antennas known in the art is now met by a new, useful, and nonobvious invention.

The present invention provides an integrated approach, at the nano-scale level, to fabricate a slot or aperture antenna integrated with a metal-insulator-metal (MIM) quantum tunneling diode to form a building block element in an array of rectenna elements for harvesting electromagnetic energy such as that emanating from the sun.

While many variations on the theme of an antenna coupled with a high frequency rectifier to harvest electromagnetic energy are known in the art, the present invention provides a novel and nonobvious solution to problems associated with prior art devices. In accordance with the present invention, an integrated nano-scale aperture antenna with a rectifier integrated using nano-scale monolithic processing is provided. Additionally, the rectifying antenna in accordance with the present invention lends itself to an inverted structure with the antenna deposited first on a transparent substrate and the rectifier and DC collection metallization fabricated on top of the antenna portion.

In accordance with the present invention is provided a rectifying antenna, comprising a transparent substrate layer, an aperture antenna fabricated on a surface of the transparent substrate, a rectifier having an anode and a cathode, the anode and the cathode electrically connected to the aperture antenna and separated from each other, the rectifier fabricated on top of the aperture antenna, a dielectric layer deposited to cover the aperture antenna and the rectifier, and a metal ground plane deposited over the dielectric layer, the metal ground plane being in electrical connectivity with the anode and the cathode of the rectifier.

While many materials may be used for the transparent substrate, in a particular embodiment the transparent substrate layer is glass.

The aperture antenna is fabricated is fabricated on the transparent substrate by depositing metal onto the surface of the transparent substrate and etching away the metal to form a slot and a dc break in the metal layer to reveal the transparent substrate, the slot and the dc break being contiguous. In a particular embodiment, the metal is gold, however other metals or metal alloys are within the scope of the present invention.

In a particular embodiment, the rectifying antenna includes an aperture or slot etched in the metal layer that is approximately $3\lambda/4$ in major length and approximately $\lambda/10$ in width, where λ is the center wavelength of the radiation to be collected. The radiation to be collected may be emanating from the sun or another source of electromagnetic radiation as is known in the art.

The rectifying antenna in accordance with the present invention additionally includes, a dc break etched in the metal layer. In a particular embodiment, the dc break is etched to establish a slot through to the transparent substrate that is approximately λ/4 in major length.

To accommodate the positioning of the rectifier relative to the aperture antenna, the slot fabricated for the antenna further comprises a metal rectifier pad located approximately in the center of the slot and having a width of less than λ/10. This metal rectifier pad may be referred to as a conducting indentation and connectivity to one side of the rectifier as in the case of a metal-insulator-metal rectifier.

In a particular embodiment, the dielectric layer deposited to cover the antenna and the rectifier is fabricated to maximize the collection of radiation incident on the aperture antenna. In a particular embodiment, the dielectric layer is approximately λ/4 in thickness. In an additional embodiment, the dielectric layer is selected from the group consisting of ceramic, mica, glass, plastic and metal oxides.

While various rectifier elements are within the scope of the present invention, in accordance with a particular embodiment, the rectifier is a metal-insulator-metal (MIM) rectifier. As such, a first metal having a first potential and a second metal having a second potential are utilized in the fabrication of the MIM rectifier. In a particular embodiment, the first metal is platinum, the second metal is palladium, and the insulator of the MIM rectifier is selected from the group consisting of silicon dioxide, silicon nitride and aluminum oxide.

In accordance with an additional embodiment, the MIM rectifier may be replaced by a bolometer.

In an additional embodiment, a matching transmission line may be fabricated between the antenna and the rectifier to provide impedance matching. Additionally, a cooling mechanism may be employed in accordance with the present invention. The implementation of the cooling mechanism including, but not limited, to a liquid cooling channel positioned coincident with the ground plane or a silicon passive heat exchanger module.

To reduce the harmonics inherent in the rectifying antenna, it is within the scope of the invention to provide at least one notch filter positioned between the antenna and the rectifier in an effort to reduce unwanted harmonics.

In providing a rectifying antenna in accordance with the present invention, a method may be employed including the steps of, providing a transparent substrate layer, fabricating an aperture antenna on a surface of the transparent substrate, fabricating a rectifier on the surface of the aperture antenna, the rectifier having an anode and a cathode, the anode and the cathode electrically connected to the aperture antenna and separated from each other, depositing a dielectric layer to cover the aperture antenna and the rectifier and depositing a metal ground plane over the dielectric layer, the metal ground plane being in electrical connectivity with the anode and the cathode of the rectifier.

While various method are known in the art for the fabrication of elements on substrate materials, in a particular embodiment of the present invention the step of fabricating an aperture antenna on a surface of the transparent substrate further includes, depositing a first metal on the surface of the transparent substrate layer, etching away the deposited first metal to form a slot for the aperture antenna and a dc break.

In a particular embodiment, the rectifier element is a metal-insulator-metal rectifier. The method of fabricating the MIM rectifier further includes, forming a conducting indentation in the center of the antenna slot, the conducting indentation fabricated of a first metal and establishing the anode of the rectifier, depositing a dielectric layer on top of the conducting indentation and fabricating a second metal layer on top of the dielectric layer to form the metal-insulator-metal rectifier across the aperture antenna. In a preferred embodiment, the dielectric layer of the MIM rectifier is between approximately 1.5 nm and 2 nm and the second metal layer is deposited to a thickness of approximately 50 nm. These dimensions are intended to be exemplary and other thicknesses of the dielectric and the second metal layer are considered within the scope of the invention.

In an additional embodiment, a dielectric layer to cover the aperture antenna and the rectifier further is deposited and planarized to form a dielectric layer of substantially uniform thickness of λ/4, while other thicknesses are within the scope of the invention.

In accordance with the present invention, a plurality of rectifying antennas formed in accordance with the method as described may be connected to form an array of rectifying antennas.

As such, the present invention provides a monolithic rectifying antenna exhibiting high efficiency solar energy conversion with improved reliability and manufacturability.

With the present invention, the dc current travels in the z-direction thereby eliminating many lossy electrical connections due to additional metal interconnects, thereby improving reliability. The manufacturing of the device requires a minimal investment in nanotechnology manufacturing techniques.

Additionally, the delicate structures of the device are protected by the glass substrate.

Other aspects and advantages of the present invention can be seen upon review of the figures, the detailed description, and the claims, which follow.

BRIEF DESCRIPTION OF THE DRAWINGS

For a fuller understanding of the nature and objects of the invention, reference should be made to the following detailed description, taken in connection with the accompanying drawings, in which:

FIG. **1** is a prior art illustration of small cone type antennas with rectifiers mounted behind the cones and the cones arranged in an array;

FIG. **2** is an illustration of a first metal layer on top of a substrate in accordance with the present invention;

FIG. **3** is an illustration of a dielectric layer with windows to a first metal (metal #1) in accordance with the present invention;

FIG. **4** is an illustration of a structure with a second metal (metal #2) in place and DC vias in accordance with the present invention;

FIG. **5** is an illustration of a top metal ground plane and DC connections in accordance with the present invention; and

FIG. **6** is an illustrated cross sectional edge view of a rectifying antenna, or rectenna, in accordance with the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

The method in accordance with the present invention establishes a fabrication process to provide a smooth, strong surface to build an aperture antenna on and thus allow the construction of the rectifier on top of the antenna to minimize relatively long, high-resistivity and high loss metal interconnects. In a particular embodiment, a 3λ/4 slot antenna concept described by the present invention has been exemplified. Additional slot antenna dimensions are within the scope of the present invention.

The method in accordance with the present invention allows for the easy adaptation of dissimilar two metal metal-insulator-metal (MIM) rectifiers. In a particular embodiment, the MIM is build utilizing two metals exhibiting large differences in potential such as Pt or Pd combined with Au or W or Al. The structure requires a dielectric substance, such as SiO_2, Silicon Nitride or Aluminum Oxide. Other insulating materials are within the scope of the present invention.

With reference to FIG. **2**, a rectifying antenna and method of manufacturing the same is provided wherein the construction of a slot rectenna includes the steps of, selecting a suitable transparent substrate **10**, like glass, on which to build the structure, depositing a metal **15**, like gold, and then etching away the slot **25** and dc break **20** as illustrated. The imaginary centerline **30** is also illustrated.

In a particular embodiment, the slot etched in the first metal (metal #1) layer opens a hole down to the transparent substrate, which is approximately 3λ/4 in major length and approximately λ/10 in width. Where λ is the center wavelength of the radiation to be collected and corrected for dilation due to the permitivity of the substrate. With reference to FIG. **2**, the conducting indentation **35** in the center of the slot, will become one side of the MIM rectifier and the thin λ/4 slot extension on the right **20**, will become the ac capacitive short/dc break for the structure. The ac capacitive short acts as a filter to separate the ac and dc components of the received signal.

After the slot **25** is etched in the metal layer, a dielectric layer **40** λ/4 thick is deposited over the entire first metal layer **15**. Holes are opened in the dielectric layer to facilitate the second metal (metal #2) deposition. This is illustrated in FIG. **3** where the dielectric layer **40** has been deposited and holes **45**, **50**, **55** and **60** have been opened up to the metal #1 layer below.

In the process of opening the dielectric layer **40** certain areas such as the little pad **45** where the rectifier is to be constructed and one side of the ac capacitive short **60** must still have a thin layer of insulation **40**. For the rectifier pad **45** this layer should be between 1.5 nm and 2 nm and for the capacitive short **60** the insulation should be thicker than 10 nm.

At this point in the fabrication process, the second metal layer **65** is deposited. The second metal layer **65** is approximately 50 nm thick. After the second metal layer **65** is deposited it is etched away from areas where it is not wanted, as shown in FIG. **4**. As such, the metal-insulator-metal rectifier **75** is formed. Above metal #2 (second metal) the insulation layer is replaced and planarized **90** so it is a uniform layer of thickness λ/4 then the window for the two dc vias, **80** and **85**, are opened up.

A final layer of the same material as metal #1 is then deposited to form a ground plane **95** and provide connection points to the +via **85** and −via **80** as illustrated in FIG. **5**.

If the rectifying antennas were incorporated into an array, the top layer of the system would include interconnection of the individual rectenna elements so energy could be effectively captured on a macroscopic scale.

With reference to FIG. **6**, an edge sectional view taken through the centerline **30** of the slot is shown.

In accordance with the present invention, the MIM can be constructed of a large variety of conducting metal combinations, the insulation can be inorganic as has been described or organic and can be made of many different types of materials. The shape of the slot and the diode may be varied. Due to impedance matching concerns between the antenna and the rectifier the dimensions of the slot may vary and it may be necessary to incorporate a matching transmission line between the antenna and the rectifier. Optional cooling methods may be employed such as liquid cooling channels next to the metal layer or a silicon passive heat exchanger module.

The non-linear nature of the rectifier causes the generation of harmonics. To prevent the loss of efficiency associated with the re-radiation at higher harmonics it is within the scope of the present invention to incorporate notch filters between the antenna and the rectifier.

Additionally, the array of slot rectennas could be replaced with an array of monopole traveling wave antennas manufactured using carbon nano-tube technology. The nano-tube antennas would be several wavelengths long and have a MIM built into their base.

In another embodiment of the invention, replaces the slot antenna with either a slot parasitic array antenna or a dipole parasitic array antenna. The use of the array concept maintains the high collection density, and handles the polarization problem to reduce the number of MIM structures required for energy conversion.

It will be seen that the objects set forth above, and those made apparent from the foregoing description, are efficiently attained and since certain changes may be made in the above construction without departing from the scope of the invention, it is intended that all matters contained in the foregoing description or shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

It is also to be understood that the following claims are intended to cover all of the generic and specific features of the invention herein described, and all statements of the scope of the invention which, as a matter of language, might be said to fall therebetween. Now that the invention has been described,

What is claimed is:

1. A rectifying antenna, comprising:

a transparent substrate layer;

an aperture antenna fabricated on a surface of the transparent substrate;

a rectifier having an anode and a cathode, the anode and the cathode electrically connected to the aperture antenna and separated from each other, the rectifier fabricated on top of the aperture antenna;

a dielectric layer deposited to cover the aperture antenna and the rectifier; and

a metal ground plane deposited over the dielectric layer, the metal ground plane being in electrical connectivity with the anode and the cathode of the rectifier.

2. The rectifying antenna of claim **1**, wherein the transparent substrate layer is glass.

3. The rectifying antenna of claim **1**, wherein the aperture antenna is fabricated of a metal.

4. The rectifying antenna of claim **3**, wherein the metal is gold.

5. The rectifying antenna of claim **1**, wherein the aperture antenna is fabricated by depositing metal onto the surface of the transparent substrate and etching away the metal to form a slot and a dc break in the metal layer to reveal the transparent substrate, the slot and the dc break being contiguous.

6. The rectifying antenna of claim **1**, wherein the slot etched in the metal layer is approximately 3λ/4 in major length and approximately λ/10 in width, where λ is the center wavelength of the radiation to be collected.

7. The rectifying antenna of claim **5**, wherein the dc break etched in the metal layer is approximately λ/4 in major length.

8. The rectifying antenna of claim **5**, wherein the slot further comprises a metal rectifier pad located approximately in the center of the slot and having a width of less than λ/10.

9. The rectifying antenna of claim **1**, wherein the dielectric layer is fabricated to maximize the collection of radiation incident on the aperture antenna.

10. The rectifying antenna of claim **1**, wherein the dielectric layer is approximately λ/4 in thickness.

11. The rectifying antenna of claim **1**, wherein the dielectric layer is selected from the group consisting of ceramic, mica, glass, plastic and metal oxides.

12. The rectifying antenna of claim **1**, wherein the rectifier is a metal-insulator-metal rectifier.

13. The rectifying antenna of claim **12**, wherein the metal-insulator-metal rectifier is fabricated using a first metal having a first potential and a second metal having a second potential.

14. The rectifying antenna of claim **13**, wherein the first metal is platinum.

15. The rectifying antenna of claim **13**, wherein the second metal is palladium.

16. The rectifying antenna of claim **12**, wherein the insulator of the metal-insulator-metal rectifier is selected from the group consisting of silicon dioxide, silicon nitride and aluminum oxide.

17. The rectifying antenna of claim **1**, wherein the rectifier is a bolometer.

18. The rectifying antenna of claim **1**, further comprising a matching transmission line fabricated between the antenna and the rectifier to provide impedance matching.

19. The rectifying antenna of claim **1**, further comprising a cooling mechanism.

20. The rectifying antenna of claim **19**, wherein the cooling mechanism is a liquid cooling channel positioned coincident with the ground plane.

21. The rectifying antenna of claim **19**, wherein the cooling mechanism is a silicon passive heat exchanger module.

22. The rectifying antenna of claim **1**, further comprising at least one notch filter positioned between the antenna and the rectifier to reduce harmonics.

23. A method of fabricating a rectifying antenna, the method comprising the steps of:

providing a transparent substrate layer;

fabricating an aperture antenna on a surface of the transparent substrate;

fabricating a rectifier on the surface of the aperture antenna, the rectifier having an anode and a cathode, the anode and the cathode electrically connected to the aperture antenna and separated from each other;

depositing a dielectric layer to cover the aperture antenna and the rectifier; and

depositing a metal ground plane over the dielectric layer, the metal ground plane being in electrical connectivity with the anode and the cathode of the rectifier.

24. The method of claim **1**, wherein the step of fabricating an aperture antenna on a surface of the transparent substrate further comprises:

depositing a first metal on the surface of the transparent substrate layer; and

etching away the deposited first metal to form a slot for the aperture antenna and a dc break.

25. The method of claim **24**, wherein the step of etching the deposited first metal further comprises:

etching away the deposited first metal to form a slot approximately 3λ/4 in major length and approximately λ/10 in width, where λ is the center wavelength of the radiation to be collected; and

etching away the deposited first metal to form a dc break approximately λ/4 in major length.

26. The method of claim **23**, wherein the step of fabricating a rectifier on the surface of the aperture antenna, the rectifier having an anode and a cathode, the anode and the cathode electrically connected to the aperture antenna and separated from each other, further comprises, fabricating a metal-insulator-metal rectifier.

27. The method of claim **26**, wherein the step of fabricating a metal-insulator-metal rectifier further comprises:

forming a conducting indentation in the center of the antenna slot, the conducting indentation fabricated of a first metal thereby forming the anode of the rectifier;

depositing a dielectric layer on top of the conducting indentation;

fabricating a second metal layer on top of the dielectric layer to form the cathode of the rectifier, thus forming a metal-insulator-metal rectifier across the aperture antenna.

28. The method of claim **27**, wherein the step of depositing a dielectric layer on top of the conducting indentation further comprises, depositing a dielectric layer between approximately 1.5 nm and 2 nm in thickness.

29. The method of claim **27**, wherein the step of fabricating a second metal layer further comprises fabricating a second metal layer approximately 50 nm in thickness.

30. The method of claim **23**, wherein the step of depositing a dielectric layer to cover the aperture antenna and the rectifier further comprises, depositing a dielectric layer approximately λ/4 in thickness.

31. The method of claim **23**, wherein the step of depositing a dielectric layer to cover the aperture antenna and the rectifier further comprise planarizing the dielectric layer to form a dielectric layer of substantially uniform thickness of λ/4.

32. The method of claim **23**, further comprising connecting a plurality of rectifying antennas to form an array of rectifying antennas.

33. The method of claim **23**, further comprising fabricating a matching transmission line between the antenna and the rectifier to improve impedance matching.

34. The method of claim **23**, further comprising fabricating cooling channels coincident with the first metal layer.

35. The method of claim **23**, further comprising fabricating a silicon passive heat exchanger module coincident with the substrate layer.

36. The method of claim **23**, further comprising fabricating at least one notch filter between the antenna and the rectifier to reduce harmonics.

* * * * *

US007154451B1

(12) **United States Patent**
Sievenpiper

(10) **Patent No.: US 7,154,451 B1**
(45) **Date of Patent: Dec. 26, 2006**

(54) **LARGE APERTURE RECTENNA BASED ON PLANAR LENS STRUCTURES**

(75) Inventor: **Daniel F. Sievenpiper**, Santa Monica, CA (US)

(73) Assignee: **HRL Laboratories, LLC**, Malibu Canyon, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **10/944,032**

(22) Filed: **Sep. 17, 2004**

(51) **Int. Cl.**
H01Q 15/02 (2006.01)
H01Q 19/06 (2006.01)

(52) **U.S. Cl.** **343/909**; 343/753

(58) **Field of Classification Search** .... 343/700 MS, 343/756, 754, 753, 909, 701
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,267,480	A	8/1966	Lerner	343/911
3,560,978	A	2/1971	Himmel et al.	343/106
3,810,183	A	5/1974	Krutsinger et al.	343/708
3,961,333	A	6/1976	Purinton	343/872
4,045,800	A	8/1977	Tang et al.	343/854
4,051,477	A	9/1977	Murphy et al.	343/700 MS
4,119,972	A	10/1978	Fletcher et al.	343/844
4,123,759	A	10/1978	Hines et al.	343/854
4,124,852	A	11/1978	Steudel	343/854
4,150,382	A	4/1979	King	343/754
4,173,759	A	11/1979	Bakhru	343/100
4,189,733	A	2/1980	Malm	343/100 SA
4,217,587	A	8/1980	Jacomini	343/100 SA
4,220,954	A	9/1980	Marchland	343/113 R
4,236,158	A	11/1980	Daniel	343/100 LE
4,242,685	A	12/1980	Sanford	343/770
4,266,203	A	5/1981	Saudreau et al.	333/21 A
4,308,541	A	12/1981	Seidel et al.	343/786
4,367,475	A	1/1983	Schiavone	343/767
4,370,659	A	1/1983	Chu et al.	343/772
4,387,377	A	6/1983	Kandler	343/756
4,395,713	A	7/1983	Nelson et al.	343/713
4,443,802	A	4/1984	Mayes	343/729
4,590,478	A	5/1986	Powers et al.	343/700 MS
4,594,595	A	6/1986	Struckman	343/770
4,672,386	A	6/1987	Wood	343/770
4,684,953	A	8/1987	Hall	343/725
4,700,197	A	10/1987	Milne	343/837
4,737,795	A	4/1988	Nagy et al.	343/712
4,749,996	A	6/1988	Tresselt	343/700 MS
4,760,402	A	7/1988	Mizuno et al.	343/713
4,782,346	A	11/1988	Sharma	343/795
4,803,494	A	2/1989	Norris et al.	343/770
4,821,040	A	4/1989	Johnson et al.	343/700 MS
4,835,541	A	5/1989	Johnson et al.	343/713
4,843,400	A	6/1989	Tsao et al.	343/700 MS
4,843,403	A	6/1989	Lalezari et al.	343/767
4,853,704	A	8/1989	Diaz et al.	343/767
4,903,033	A	2/1990	Tsao et al.	343/700 MS
4,905,014	A	2/1990	Gonzalez et al.	343/909

(Continued)

FOREIGN PATENT DOCUMENTS

DE	196 00 609 A1	4/1994

(Continued)

Primary Examiner—Don Wong
Assistant Examiner—Angela M Lie
(74) *Attorney, Agent, or Firm*—Ladas & Parry LLP

(57) **ABSTRACT**

A rectenna structure comprising a flexible, dielectric sheet of material; a plurality of metallic lenslets disposed on the sheet of material; and a plurality of diodes disposed on the sheet of material, each diode in said plurality of diodes being arranged at a focus of a corresponding one of said plurality of metallic lenslets.

24 Claims, 6 Drawing Sheets

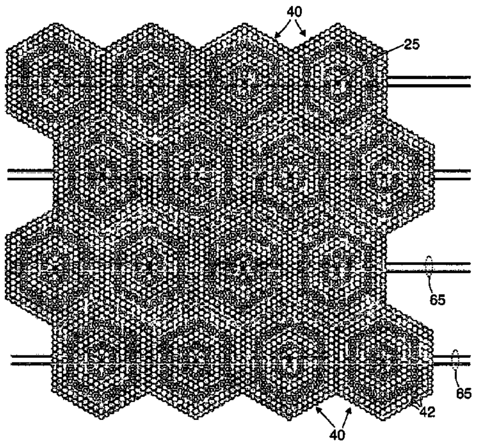

U.S. PATENT DOCUMENTS

4,916,457	A	4/1990	Foy et al.	343/770
4,922,263	A	5/1990	Dubost et al.	343/797
4,958,165	A	9/1990	Axford et al.	343/770
4,975,712	A *	12/1990	Chen	343/754
5,021,795	A	6/1991	Masiulis	343/700 MS
5,023,623	A	6/1991	Kreinheder et al.	343/725
5,070,340	A	12/1991	Diaz	343/767
5,081,466	A	1/1992	Bitter, Jr.	343/767
5,115,217	A	5/1992	McGrath et al.	333/246
5,146,235	A	9/1992	Frese	343/895
5,148,182	A *	9/1992	Gautier et al.	343/754
5,208,603	A	5/1993	Yee	343/909
5,218,374	A *	6/1993	Koert et al.	343/789
5,235,343	A	8/1993	Audren et al.	343/816
5,268,696	A	12/1993	Buck et al.	342/372
5,268,701	A	12/1993	Smith	343/767
5,278,562	A	1/1994	Martin et al.	342/1
5,287,116	A	2/1994	Iwasaki et al.	343/700 MS
5,287,118	A	2/1994	Budd	343/909
5,402,134	A	3/1995	Miller et al.	343/742
5,406,292	A	4/1995	Schnetzer et al.	343/700 MS
5,519,408	A	5/1996	Schnetzer	343/767
5,525,954	A	6/1996	Komazaki et al.	333/219
5,531,018	A	7/1996	Saia et al.	29/622
5,532,709	A	7/1996	Talty	343/819
5,534,877	A	7/1996	Sorbello et al.	343/700 MS
5,541,614	A	7/1996	Lam et al.	343/792.5
5,557,291	A	9/1996	Chu et al.	343/725
5,581,266	A	12/1996	Peng et al.	343/770
5,589,845	A	12/1996	Yandrofski et al.	343/909
5,598,172	A *	1/1997	Chekroun	343/754
5,600,325	A	2/1997	Whelan et al.	342/13
5,611,940	A	3/1997	Zettler	73/514.16
5,619,365	A	4/1997	Rhoads et al.	359/248
5,619,366	A	4/1997	Rhoads et al.	359/248
5,621,571	A	4/1997	Bantli et al.	359/529
5,638,946	A	6/1997	Zavracky	200/181
5,644,319	A	7/1997	Chen et al.	343/702
5,694,134	A	12/1997	Barnes	343/700
5,721,194	A	2/1998	Yandrofski et al.	505/210
5,767,807	A	6/1998	Pritchett	342/374
5,808,527	A	9/1998	De Los Santos	333/205
5,874,915	A	2/1999	Lee et al.	342/375
5,892,485	A	4/1999	Glabe et al.	343/789
5,894,288	A	4/1999	Lee et al.	343/770
5,905,465	A	5/1999	Olson et al.	343/700 MS
5,923,303	A	7/1999	Schwengler et al.	343/853
5,926,139	A	7/1999	Korisch	343/702
5,929,819	A	7/1999	Grinberg	343/754
5,943,016	A	8/1999	Snyder, Jr. et al.	343/700 MS
5,945,951	A	8/1999	Monte et al.	343/700 MS
5,949,382	A	9/1999	Quan	343/767
5,966,096	A	10/1999	Brachat	343/700 MS
5,966,101	A	10/1999	Haub et al.	343/767
6,005,519	A	12/1999	Burns	343/700 MS
6,005,521	A	12/1999	Suguro et al.	343/700 MS
6,008,770	A	12/1999	Sugawara	343/767
6,016,125	A	1/2000	Johansson	343/702
6,028,561	A	2/2000	Takei	343/767
6,028,692	A	2/2000	Rhoads et al.	359/245
6,034,644	A	3/2000	Okabe et al.	343/767
6,034,655	A	3/2000	You	345/60
6,037,905	A	3/2000	Koscica et al.	343/701
6,040,803	A	3/2000	Spall	343/700 MS
6,046,655	A	4/2000	Cipolla	333/137
6,046,659	A	4/2000	Loo et al.	333/362
6,054,659	A	4/2000	Lee et al.	200/181
6,061,025	A	5/2000	Jackson et al.	343/700 MS
6,075,485	A	6/2000	Lilly et al.	343/700 MS
6,081,235	A	6/2000	Romanofsky et al.	343/700 MS
6,081,239	A	6/2000	Sabet et al.	343/753
6,097,263	A	8/2000	Mueller et al.	333/17.1
6,097,343	A	8/2000	Goetz et al.	343/708
6,118,406	A	9/2000	Josypenko	343/700 MS
6,118,410	A	9/2000	Nagy	343/713
6,127,908	A	10/2000	Bozler et al.	333/246
6,150,989	A	11/2000	Aubry	343/767
6,154,176	A	11/2000	Fathy et al.	343/700 MS
6,166,705	A	12/2000	Mast et al.	343/853
6,175,337	B1 *	1/2001	Jasper et al.	343/770
6,175,723	B1	1/2001	Rothwell, III	455/63
6,188,369	B1	2/2001	Okabe et al.	343/767
6,191,724	B1	2/2001	McEwan	342/21
6,198,438	B1	3/2001	Herd et al.	343/700 MS
6,198,441	B1	3/2001	Okabe et al.	343/702
6,204,819	B1	3/2001	Hayes et al.	343/702
6,218,912	B1	4/2001	Mayer	333/106
6,218,997	B1	4/2001	Lindenmeier et al.	343/725
6,246,377	B1	6/2001	Aiello et al.	343/700
6,252,473	B1	6/2001	Ando	333/105
6,285,325	B1	9/2001	Nalbandian et al.	343/700 MS
6,307,519	B1	10/2001	Livingston et al.	343/767
6,317,095	B1	11/2001	Teshirogi et al.	343/785
6,323,826	B1	11/2001	Sievenpiper et al.	343/909
6,331,257	B1	12/2001	Loo et al.	216/13
6,337,668	B1	1/2002	Ito et al.	343/833
6,366,254	B1	4/2002	Sievenpiper et al.	343/700
6,373,349	B1	4/2002	Gilbert	333/126
6,380,895	B1	4/2002	Moren et al.	343/700 MS
6,388,631	B1	5/2002	Livingston et al.	343/767
6,392,610	B1	5/2002	Braun et al.	343/876
6,404,390	B1	6/2002	Sheen	343/700 MS
6,404,401	B1	6/2002	Gilbert et al.	343/780
6,407,719	B1	6/2002	Ohira et al.	343/893
6,417,807	B1	7/2002	Hsu et al.	343/700 MS
6,424,319	B1	7/2002	Ebling et al.	343/911 L
6,426,722	B1	7/2002	Sievenpiper et al.	343/700 MS
6,440,767	B1	8/2002	Loo et al.	438/52
6,469,673	B1	10/2002	Kaiponen	343/703
6,473,362	B1	10/2002	Gabbay	367/119
6,483,480	B1	11/2002	Sievenpiper et al.	343/909
6,496,155	B1	12/2002	Sievenpiper et al.	343/770
6,518,931	B1	2/2003	Sievenpiper	343/700
6,525,695	B1 *	2/2003	McKinzie, III	343/756
6,538,621	B1	3/2003	Sievenpiper et al.	343/909
6,515,635	B1	4/2003	Chiang et al.	343/834
6,552,696	B1	4/2003	Sievenpiper et al.	343/909
6,624,720	B1	9/2003	Allison et al.	333/105
6,642,889	B1	11/2003	McGrath	343/700 MS
6,657,525	B1	12/2003	Dickens et al.	335/78
6,741,207	B1	5/2004	Allison et al.	342/371
6,822,622	B1 *	11/2004	Crawford et al.	343/909
6,864,848	B1	3/2005	Sievenpiper	343/767
6,897,810	B1	5/2005	Dai et al.	343/703
7,026,993	B1 *	4/2006	Zhang et al.	343/700 MS
2001/0035801	A1	11/2001	Gilbert	333/126
2002/0036586	A1	3/2002	Gothard et al.	342/374
2003/0112186	A1 *	6/2003	Sanchez et al.	343/700 MS
2003/0122721	A1	7/2003	Sievenpiper	343/767
2003/0193446	A1	10/2003	Chen	343/893
2003/0222738	A1	12/2003	Brown et al.	333/206
2003/0227351	A1	12/2003	Sievenpiper	333/105
2004/0113713	A1	6/2004	Zipper et al.	343/702
2004/0135649	A1	7/2004	Sievenpiper	333/105
2004/0227583	A1	11/2004	Shaffner et al.	333/32
2004/0227667	A1	11/2004	Sievenpiper	343/700 MS
2004/0227668	A1	11/2004	Sievenpiper	343/700 MS
2004/0227678	A1	11/2004	Sievenpiper	343/702
2004/0263408	A1	12/2004	Sievenpiper et al.	343/757
2005/0012667	A1 *	1/2005	Noujeim	343/700 MS

2006/0044199 A1 * 3/2006 Furuhi et al. 343/753

FOREIGN PATENT DOCUMENTS

EP	1 158 605	A1	11/2001
FR	2 785 476		5/2000
GB	2 281 662		3/1995
GB	2 328 748		3/1999
JP	61-260702		11/1986
WO	94/00891		1/1994
WO	96/29621		9/1996
WO	98/21734		5/1998
WO	99/50929		10/1999
WO	03/098732	A1	11/2003

* cited by examiner

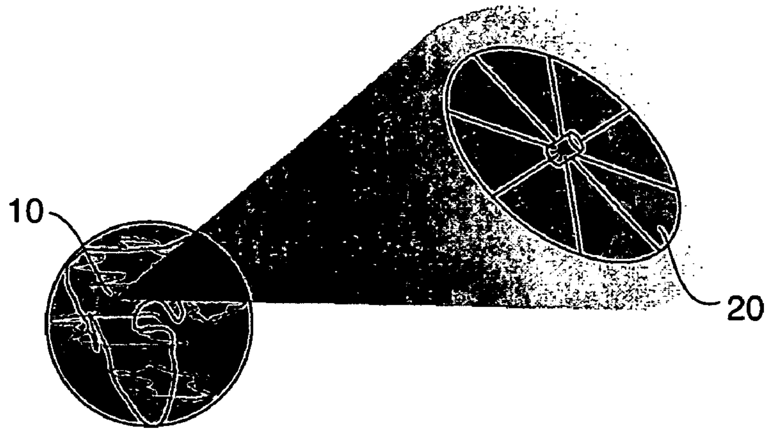

Figure 1

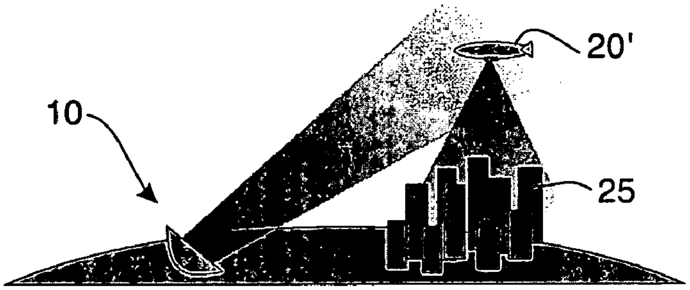

Figure 2

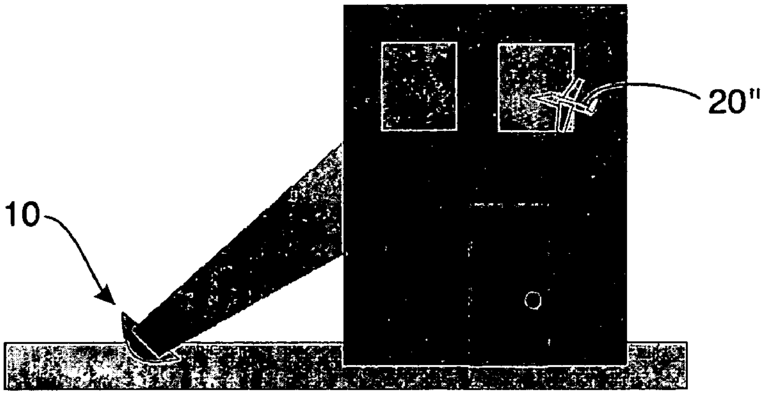

Figure 3

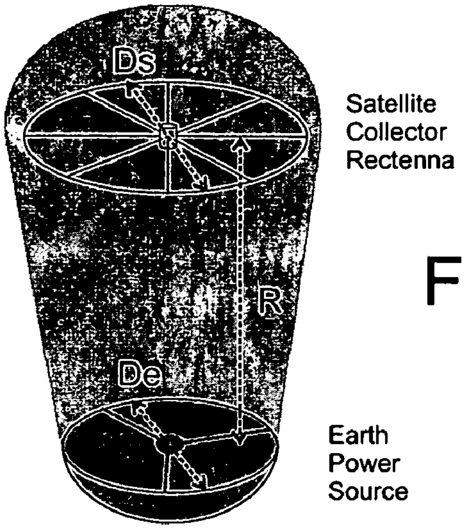
Ds
Satellite
Collector
Rectenna
R
De
Earth
Power
Source

Figure 4

Figure 5

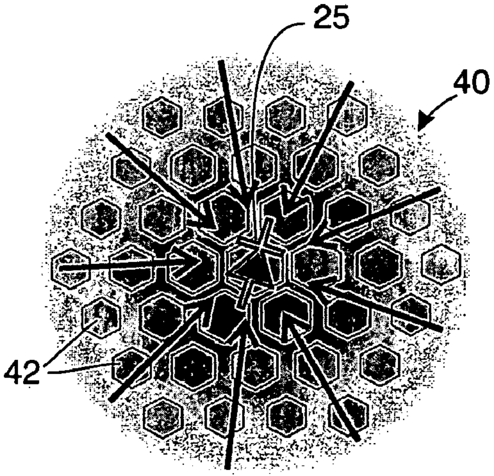

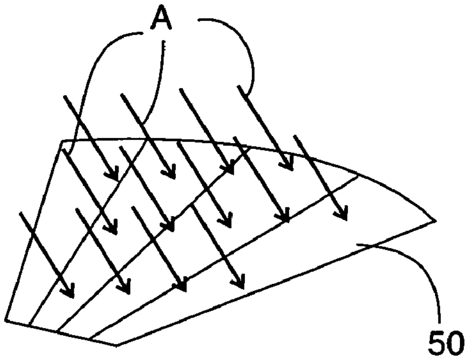

Figure 6

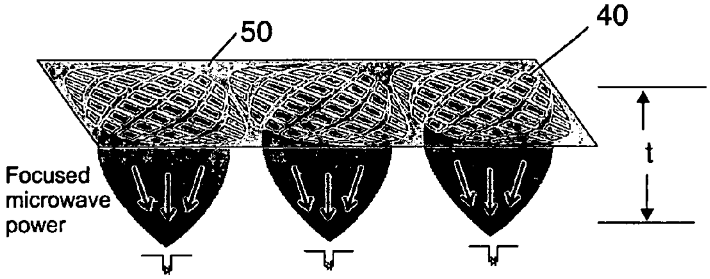
50
40
t
Focused
microwave
power

Figure 7

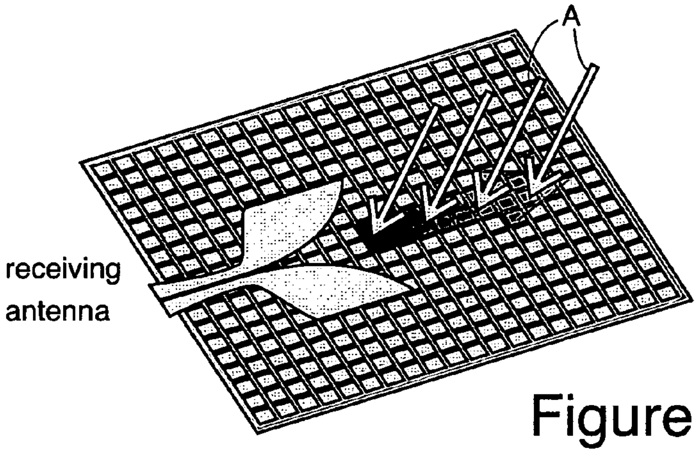
A
receiving
antenna

Figure 8a

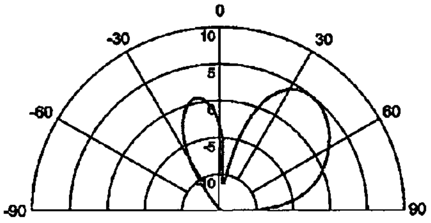

Figure 8b

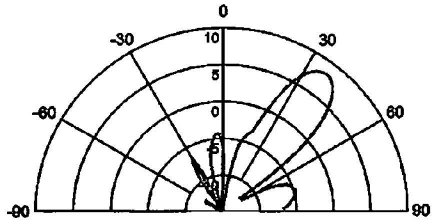

Figure 8c

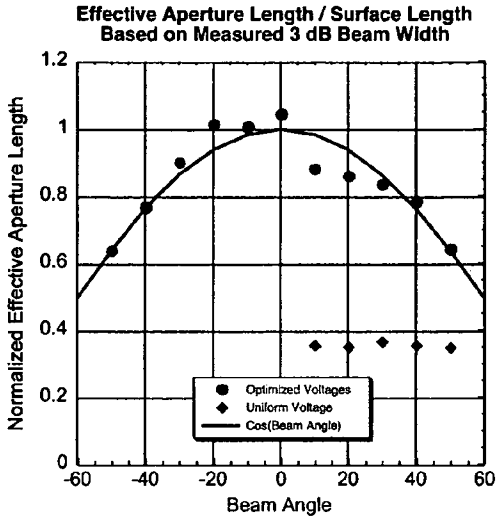

Figure 9

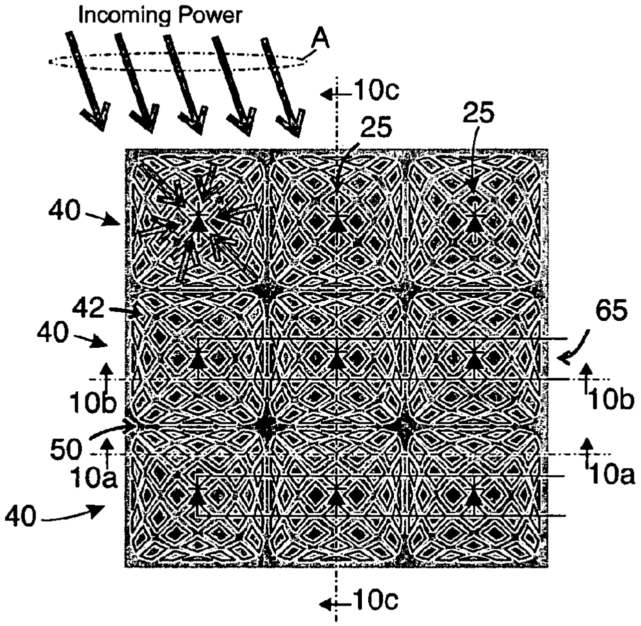
Incoming Power
A
10c
25
25
40
42
40
65
10b
10b
50
10a
10a
40
10c

Figure 10

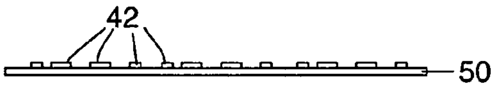

Figure 10a

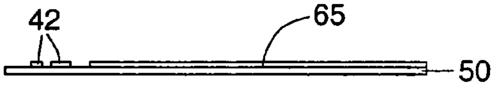

Figure 10b

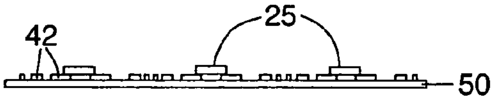

Figure 10c

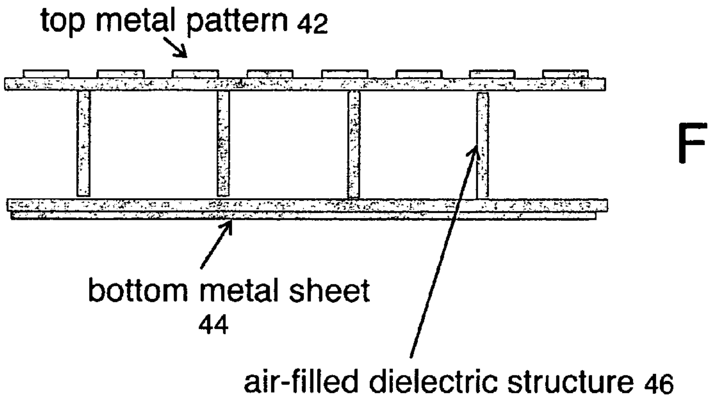
top metal pattern 42
bottom metal sheet
44
air-filled dielectric structure 46

Figure 11

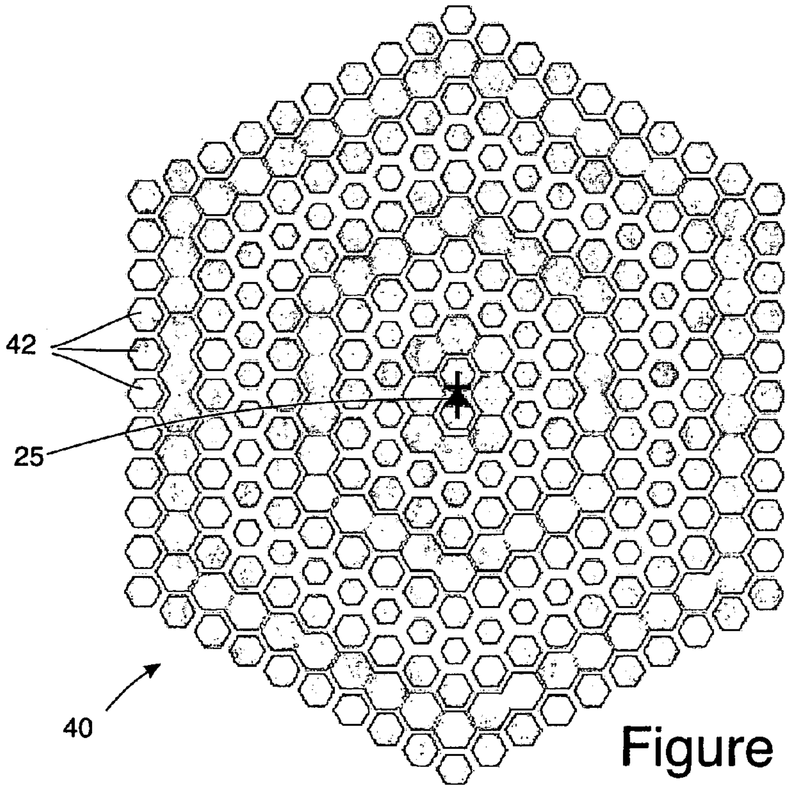

Figure 12

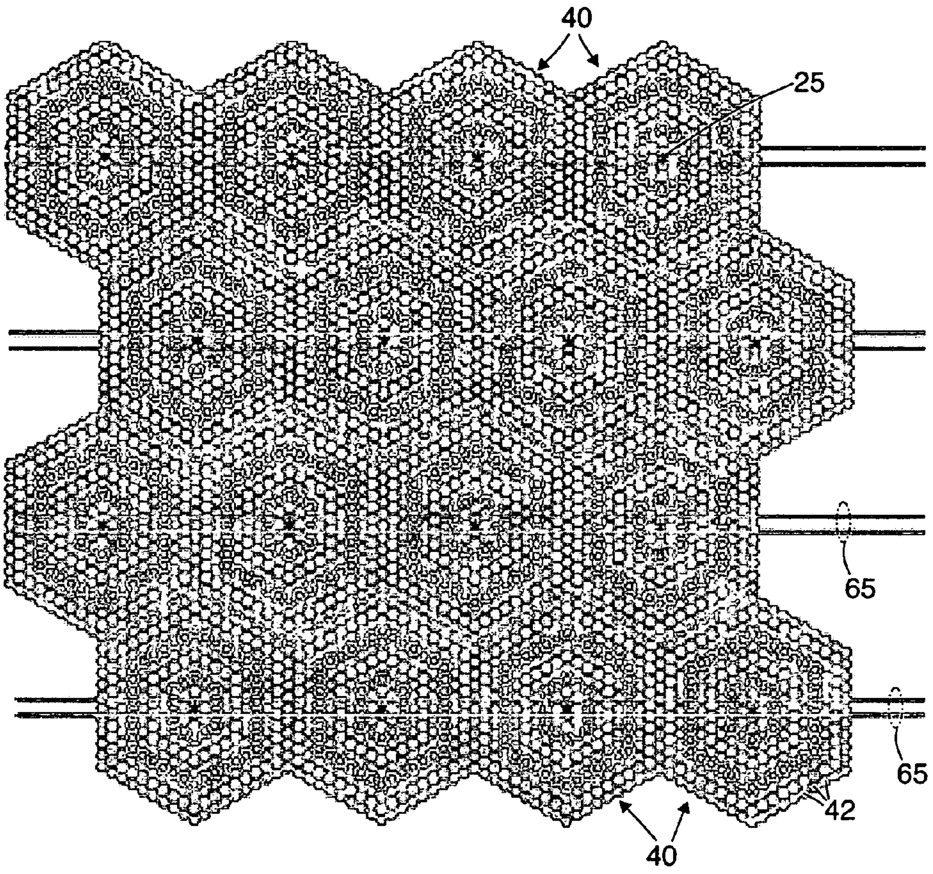
40
25
65
65
42
40

Figure 13

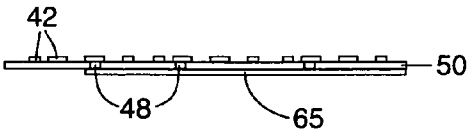

Figure 14

LARGE APERTURE RECTENNA BASED ON PLANAR LENS STRUCTURES

RELATED APPLICATIONS

This disclosure is related to U.S. Patent Application Ser. No. 60/470,027 entitled "Meta-element Antenna and Array" filed May 12, 2003 and to U.S. Patent Application Ser. No. 60/470,028 entitled "Steerable Leaky Wave Antenna Capable for Both Forward and Backward Radiation" filed May 12, 2003. The disclosures of these applications are hereby incorporated herein by reference. This disclosure is also related to two non-provisional applications that were filed claiming the benefit of the aforementioned applications. The two non-provisional applications have Ser. Nos. 10/792,411 and 10/792,412 and were both filed on Mar. 2, 2004. The disclosures of these two non-provisional applications are also incorporated herein by reference.

TECHNICAL FIELD

The technology disclosed herein relates to a lightweight, high-efficiency rectenna and to a method or architecture for making same. Rectennas can be useful for a variety of applications in the field of beaming RF power, which can be useful for satellites, zeppelins, and UAVs.

BACKGROUND OF THE INVENTION

Rectennas are antenna structures that intentionally incorporate rectifying elements in their designs.

Satellites are an integral part of modern communication systems, and their importance can be expected to grow in the coming years. As future generations of satellites with greater capabilities become possible, it is expected that they can take an even more active role in future military conflicts.

The design of present-day satellites often involves tradeoffs among such aspects as weight, power, and electronic capabilities. Each new electronic system adds weight, and must compete for power with other required systems such as station keeping. The limits of these tradeoffs are eased only gradually from one generation to the next, by the evolution of electronics, batteries, propulsion systems, and so on. Thus, developing new technologies that significantly expand the available design space is crucial to the enablement of satellites with radically improved capabilities over the present generation.

Power supply or generation is one area where revolutionary changes could significantly expand satellite capabilities. Presently, power sources are limited to solar panels or on-board power supplies. Solar panels require continuous exposure to the sun, or the use of batteries to supply power during periods of darkness. Any on-board power system such as a battery adds weight, which reduces the number of electronic systems that can be flown. Furthermore, a system of solar panels and/or on-board sources is best suited to continuous power at moderate levels, and cannot easily supply high-energy bursts without significant additional weight in order to collect and store, and then release the energy.

One way of providing a more flexible power source is to beam the power from a ground station **10** to a satellite **20**, as illustrated in FIG. **1**. This concept has been explored in the past, but in the opposite direction: beaming power to earth (which seemed attractive during the energy crisis). Sending power in the space to earth direction faces certain fundamental limits that make it impractical, but these limits are eased in the earth to space direction, leading to a system that is within the realm of possibility.

In addition to satellites, there are many other applications where beaming power could be important. For example, it is possible to replace hundreds of civilian cellphone base stations with a single zeppelin **20'**, shown in FIG. **2**, which could service a large metropolitan area **25** with mobile telephony, as well as such other services as "satellite" television. This would provide a low-cost alternative to satellites for many commercial wireless applications.

Furthermore, other applications include small UAVs (Unmanned Aerial Vehicles) that could be powered by beamed energy. See FIG. **3**. As the size of a UAV is reduced, the amount of weight that it can carry limits its lifespan significantly. For example, 100-gram airplanes have been built, but their lifetime is limited to six minutes with currently available batteries. By beaming power to a micro-UAV **20"**, it could stay aloft much longer. This would be useful for such applications as law enforcement, surveillance, hazardous site investigation, etc., in addition to the obvious military applications.

The embodiments of FIGS. **1**–**3** assume that the source of power is from a ground station **10**. However, the source of power need not necessarily be terrestrial. The source of power could be airborne or even in space.

Any beamed power system must confront the fundamental limits summarized by the Friis transmission equation, which relates the total power transmitted to the gain, G, of the transmitting and receiving antennas, the distance between them, R, and the wavelength λ of the radiation used.

$$P_{Rx} = G_{Tx} G_{Rx} \left(\frac{\lambda}{4\pi R}\right)^2 P_{Tx} \qquad [1]$$

Assuming for simplicity that both antennas are circular, the gain of each is related to its diameter, D.

$$G = \left(\frac{\pi D}{\lambda}\right)^2 \qquad [2]$$

If one assumes for the moment that very little power will be lost to spillover (this requirement can be relaxed) these equations can be combined to yield an expression for the required sizes of the transmitting and receiving antennas, as a function of their separation, and the wavelength of the radiation used. See FIG. **4**, which depicts the geometry involved in equation 3, to determine the required diameters for the transmitting and receiving antennas.

$$D_{Tx} D_{Rx} = \frac{4}{\pi} R\lambda \qquad [3]$$

For a given separation, reducing the wavelength reduces the size requirements of the transmitter and/or receiver. One tempting solution is to use optical wavelengths, and beam power to space with a large earth-based laser. This has several drawbacks, including scattering by atmospheric turbulence and airborne particles, the typically low wall-plug efficiencies of lasers compared to microwave sources, and the losses in conversion back to DC by photovoltaic cells.

Lasers may be viable alternatives for stationary, near-earth applications such as zeppelins, but not for moving applications, such as micro-UAVs. Their utility for satellites is questionable.

The next candidate wavelength range after optical (skipping terahertz frequencies, which are currently not feasible) is millimeter waves. In the 90–100 GHz range, the attenuation for a one-way trip through the atmosphere can be as little as 1 dB (See Koert, 1992, infra). Furthermore, efficient high-power sources are available, such as the gyrotron, which can produce as much as 200 kW of continuous power at millimeter wave frequencies, at an efficiency of 50% (See Gold, 1997, infra). For higher power applications, arrays of klystrons have been proposed that could produce tens of megawatts of power. These existing high-power sources suggest that it could be possible to temporarily supply a satellite with much higher power from the ground than can currently be produced in orbit. For comparison, the most powerful commercial satellite that is available, the Boeing 702, operates at 25 kW from on-board solar panels. These power sources would be more than adequate for airship applications, and the power required for micro-UAVs would only be on the order of watts.

The most significant engineering challenge for efficient earth to space power transmission is the design of the transmitting and receiving antennas. Fortunately, the receiver design is greatly simplified by the development of the rectenna, (See Brown, 1984, infra) which consists of an array having a rectifier diode at each element. Converting to DC directly at each antenna eliminates the requirement for a perfectly flat phase front, and permits the receiving aperture to take any shape. The transmitter must still produce a coherent beam, so a parabolic dish or other method of phase control is necessary. This is one reason why space to earth transmission is impractical. To illustrate the possibility of high-efficiency earth to space transmission, consider the following example.

Assume that 100 GHz radiation is to be used. The maximum transmitter gain is determined by the ability to accurately build a large dish with the necessary smoothness. The Arecibo dish, which operates at 10 GHz, is 300 meters in diameter. First, assume that a 100 GHz dish could be similarly built with a diameter of 30 meters.

Next, assume that a low-earth-orbit (LEO) satellite is utilized, at an altitude of 500 km. Using equation 3, the required receiver diameter for high transmission efficiency is about 60 meters. This can be compared to the Boeing 702 solar panel wingspan of 47 meters. Thus, structures of the required sizes can be built, both on earth and in space.

However, existing rectenna designs are not practical for space power applications because they require an enormous number of diodes to cover such a large area. For the example just described, one diode per half-wavelength at 100 GHz equates to 6 billion diodes. Using 12-inch wafers, and assuming an area of 1 mm square per diode, this represents the yield of 20,000 wafers; the weight and cost of the diodes alone would be prohibitive.

Another problem with space power applications using traditional rectenna designs is that the power density is too low to achieve significant efficiency. The efficiency, h, of a rectenna is related to the voltage across the diodes, V_D, and the built-in diode voltage, V_{bi} (See McSpadden, 1998, infra).

$$\eta \propto \frac{1}{1 + V_{bi} / V_D} \qquad [4]$$

Designs with efficiencies as high as 90% have been demonstrated, [Strassner, 2002] but the power densities involved were much higher than one could expect to encounter in space. For the LEO example given above, the power density would be 6 mW/cm^2, which corresponds to only 0.2 volts generated across each diode—on the order of the typical built-in voltage for a Schottky diode. The practical limitations of a space power system are thus the large number of diodes needed, and the low voltage generated across each diode. The efficiency could also be improved by placing each diode inside a high Q resonant structure, or by using diodes with lower built-in voltage. However, either of these solutions alone would not solve the problem of the large number of required diodes.

As such there is a need for lens-like structures that will allow the number of diodes to be reduced.

In terms of the prior art and a better understanding of the background to the present invention, the reader is directed to the following articles:

W. Brown, "The History of Power Transmission by Radio Waves", *IEEE Transactions on Microwave Theory and Techniques*, vol. 32, no. 9, pp. 1230–1242, September 1984.

P. Fay, J. N. Schulman, S. Thomas III, D. H. Chow, Y. K. Boegeman, and K. S. Holabird, "High-Performance Antimonide-Based Heterostructure Backward Diodes for Millimeter-wave Detection", *IEEE Electron Device Lett.* 23, 585–587 (2002).

S. Gold, G. Nusinovitch, "Review of High Power Microwave Source Research", *Review of Scientific Instruments*, vol. 68, no. 11, pp. 3945–3974, November 1997.

P. Koert, J. Cha, "Millimeter Wave Technology for Space Power Beaming", *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, no. 6, pp. 1251–1258, June 1992.

H. J. Lezec, A. Degiron, E. Devaux, R. A. Linke, L. Martin-Moreno, F. J. Garcia-Vidal, and T. W. Ebbesen, "Beaming Light from a Subwavelength Aperture", *Science*, vol. 297, pp. 820–822, Aug. 2, 2002.

J. McSpadden, L. Fan, K. Chang, "Design and Experiments of a High-Conversion-Efficiency 5.8 GHz Rectenna", *IEEE Transactions on Microwave Theory and Techniques*, vol. 46, no. 12, pp. 2053–2060, September 1984.

J. N. Schulman and D. H. Chow, "Sb-Heterostructure Interband Backward Diodes," *IEEE Electron Device Lett.*, 21, 353–355 (2000).

D. Sievenpiper, J. Schaffner, H. Song, R. Loo, G. Tangonan, "Two-Dimensional Beam Steering Using an Electrically Tunable Impedance Surface", *IEEE Transactions on Antennas and Propagation*, special issue on metamaterials, October 2003.

B. Strassner, K. Chang, "5.8 GHz Circularly Polarized Rectifying Antenna for Wireless Microwave Power Transmission", *IEEE Transactions on Microwave Theory and Techniques, vol.* 50, no. 8, pp. 1870–1876, August 2002.

F. Yang, Y. Qian, T. Itoh, "A Uniplanar Compact Photonic Bandgap (UCPBG) Structure and its Applications for Microwave Circuits", *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 8, pp. 1509–1514, August 1999.

BRIEF DESCRIPTION OF THE INVENTION

Briefly and in general terms, the disclosed technology, in one aspect comprises a rectenna structure comprising: a flexible, dielectric sheet of material; a plurality of metallic lenslets disposed on the sheet of material; and a plurality of diodes disposed on the sheet of material, each diode in said plurality of diodes being arranged at a focus of a corresponding one of said plurality of metallic lenslets.

In another aspect, the disclosed technology relates to a method of generating electrical power for use aboard an aircraft or a satellite, the method comprising: deploying a sheet of dielectric material in an orientation, the sheet of dielectric material being associated with, coupled to and/or forming a part of said aircraft or satellite, the sheet of dielectric material having a plurality of metallic lenslets disposed on the sheet of dielectric material and a plurality of diodes disposed on or adjacent the sheet of dielectric material, each diode in said plurality of diodes being arranged at a focus of a corresponding one of said plurality of metallic lenslets, the diodes being coupled together for supplying electrical power for use by systems aboard said aircraft or a satellite, and directing the orientation of the sheet of dielectric material to receive incident radiation from a source of electromagnetic radiation.

BRIEF DESCRIPTION OF THE FIGURES

FIG. **1** depicts beaming microwave power to an orbiting satellite as an alternative to the use of traditional solar panels, batteries, and other power sources on the satellite.

FIG. **2** depicts another application for beamed RF power which includes airships that would supply cities with wireless services, for example.

FIG. **3** depicts an application where beamed RF power may also be used to power "slow-flight" micro-UAVs, which could be used for law-enforcement or surveillance, or investigations made at a hazardous site, for example.

FIG. **4** shows the geometry involved in equation [3], which equation can be used to determine the required diameters for the transmitting and receiving antennas.

FIG. **5** depicts radiation from a relatively larger being concentrated onto each diode, using, for example, a lightweight, planar resonant structure that may be printed on a thin, flexible plastic film.

FIG. **6** depicts a large plastic film, printed with metallic lens structures, and populated with rectifier diodes, (not shown) that would serve as a lightweight collector for microwave power. Such a structure could be built to cover tens of meters with minimal weight. The use of printed metallic lenses would reduce the number of diodes, and would increase the voltage across each diode for improved efficiency.

FIG. **7** depicts an embodiment where power is focused onto a sparse array of diodes using a lightweight plastic film that is patterned with metallic lenses.

FIG. **8***a* depicts a coplanar antenna on a diode-tuned surface and experiments using such tunable textured surfaces has led the inventor named herein to believe that planar lens structures can focus power onto a coplanar antenna, yielding a completely flat structure.

FIG. **8***b* is a graph of the gain for uniform surface impedance.

FIG. **8***c* is a graph of the gain for an optimized, non-uniform surface.

FIG. **9** is a graph depicting the effective aperture for an antenna mounted on an optimized impedance surface can be nearly equal to the entire surface area. The effective aperture size assumes the expected cosine function when optimized for different elevation angles. The effective aperture for a uniform surface is shown for comparison. By optimizing a textured surface, one can make a large-area, planar collector for microwave radiation.

FIG. **10** is a plan view of a lightweight, high-efficiency rectenna system based on printed metallic lenslets, and a sparse array of rectifier diodes. The lenslets collect power over many square wavelengths and route it to the diodes. This provides greater power per diode (thus improving efficiency) and also reduces the diode count to a practical number.

FIGS. **10***a*, **10***b* and **10***c* are side elevation views taken through the structure depicted in FIG. **10**.

FIG. **11** is a side elevation view of an embodiment with a ground plane spaced from the lenslets using a honeycomb-like structure.

FIG. **12** depicts an array of metal plates with a period of one-quarter wavelength, and features that vary on a length scale of one wavelength, with radial symmetry.

FIG. **13** depicts an array of lenses depicted by FIG. **12**, each lens having a diode at its center, with those diodes being connected by DC lines.

FIG. **14** is similar to FIG. **10***b*, but with a DC line being shown on a reverse side of the dielectric sheet.

DETAILED DESCRIPTION

A problem in trying to develop a practical earth to space power transmission system is that the voltage across diodes used in a rectenna has not been sufficient in a prior art rectenna to be of practical use to such an application.

However, the voltage across each diode **25** can be increased while reducing the number of diodes by using a lens-like structure or lenslet **40**, shown in FIG. **5**, to concentrate power from a large area over a small number of diodes **25** in an array of lenslets **40**. For example, if one wanted to generate 20 volts across each diode **25**, then the incident power from 40 square wavelengths, or a 2-cm diameter area, needs to be collected. This would not only boost the voltage across each diode—and also the diode's efficiency—it would also reduce the number of diodes to about 3 million for the example described above, which equates to about 100 wafers' worth of diodes. Further reductions in the number of diodes required could be achieved with an even larger collection area per diode. Each lenslet **40** comprises a geometric array of electrically conductive patches **42** disposed on a supporting surface. The conductive patches **42** are preferably formed by thin, individual metallic patches formed on a supporting surface, such as, for example, a thin sheet of a plastic material.

Of course, a traditional dielectric lens would be impractical, but a metallic lens imprinted on a lightweight plastic film **50**, which may be unfolded over a large area and could be utilized in a space environment, is practical. This concept for building a practical microwave space power system is illustrated in FIG. **6** where ground-based radiation is represented by arrows A. The thin plastic film **50** would be patterned with sub-wavelength resonant metallic regions or lenslets **40**, which would focus the incoming power to a sparse array of diodes **25**. Such a film **50** could be made by the tens of meters, but nevertheless would have minimal weight. The metallic lenses would serve the dual purposes of minimizing the number of diodes **25** required, while also improving the efficiency by increasing the voltage across each diode **25**.

In accordance with the presently disclosed technology, a structure having a thin plastic film **50** that is covered with a plurality of thin metal patterns, each pattern comprising a plurality of small electrically conductive patches **42** forming a lenslet **40**, is disclosed. This technology may be used in applications such as the earth to space power transmission system discussed above. Each metal pattern or lenslet **40** is made such that it behaves as a planar lens, with a focal length of zero. That is, it focuses the incoming power in such a way that a relatively high energy field is created at one point on the surface of the lens **40**. The high-energy field has a higher energy than the average energy density of the electromagnetic waves impinging the plastic film **50**. The creation of the high-energy fields allows a rectifier diode **25** to be placed at the focus or center of the high-field location, so that all of the power impinging on the lens **40** is rectified by that diode **25**. This results in two improvements over existing rectenna designs: (1) It requires far fewer diodes, and (2) it allows the voltage per diode to be higher, which results in more efficient operation. As will be seen, an embodiment of the present invention includes the combination of a planar lens and a sparse array of rectifier diodes to create a lightweight, efficient rectenna.

The design of the planar lens can be summarized as follows: (1) assume that the plastic film **50** is preferably planar and is patterned with metallic or other electrically conductive patches **42** that can be considered as resonators, with a certain resonance frequency. (2) Characterize the patches **42** in terms of scattered field (magnitude and phase) for various frequencies with respect to the resonance frequency. (3) Choose the condition that the fields from all of the metal patches **42** should add up in phase at a single point at the focus of a lens **40**, or alternatively choose some other point on the lens. (4) Build a scattering matrix that describes the field at the chosen point on the lens, as a function of the incoming field. This must include the interaction among the various metallic patches. (5) Optimize the resonance frequencies of the metal patches **42** so that the field at the chosen point is a maximum. Of course, diodes **25** would be placed at the focal points of the lenses **40**.

Concentrating microwave power from a large area (several tens of square wavelengths) onto a single device, using a thin, patterned metal film can be done in several ways, including by using a non-uniform frequency selective surface (FSS). These structures have been studied for many years for filtering radomes, and other applications. A non-uniform FSS could be designed to have lens-like behavior, and focus incoming waves from a large area onto a single receiving antenna. This is similar to the Fresnel zone plate that is known in optics, but it can have high efficiency because the metal patterns can be designed to provide only a phase shift, with minimal absorption. A series of microwave lenslets **54** could be patterned over a large area of thin plastic film **50**, as shown in FIG. **6**, to focus the low-density microwave power onto a sparse array of diodes **25**.

One drawback of the traditional FSS approach, shown in FIG. **7**, is that it is not uniplanar, because of the need to focus over a distance "t" which would be roughly equal to the diameter of the lenslets **40**. It would be difficult to unfold such a structure over an area of many square meters while maintaining the required spacing "t" within suitable tolerances for efficient energy collection.

An alternative is to consider structures where the receiving antennas and the diodes are arranged in a coplanar alignment with the metallic lens structures. This concept has already been demonstrated at HRL Laboratories of Malibu, Calif., through work with tunable, textured electromagnetic surfaces. See, for example, the patent applications mentioned above. A metallic surface texture can be made (through proper optimization) to focus power from many square wavelengths, onto an antenna that is coplanar with the textured surface, as illustrated in FIG. **8**. The metallic surface can be quite thin and for most applications, the thinner the metal on the metallic surface the better (due to weight considerations). The experiments at HRL Laboratories involved a diode-tuned impedance surface consisting of many small metallic patches linked by varactor diodes. See the paper by Sievenpiper, 2003, supra, and the patent applications referred to above. Experiments using tunable textured surfaces suggest that planar lens structures can focus power onto a coplanar antenna, yielding a completely (or essentially) flat structure as shown in FIG. **8***a*. The structure of FIG. **8***a* can be made so flat that the antenna and tunable textured surface is nearly imperceptible to one's fingers. Of course, the structures can be more pronounced in some embodiments (so that they would not generally be called flat), but, generally speaking, flat or nearly flat structures would be preferred in most applications, particularly where the plastic film **50** is to be unfolded someplace, such as in space, where human intervention (due to snags and the like), may well not be possible, convenient or desirable.

FIG. **8***b* is a graph of the gain for uniform surface impedance while FIG. **8***c* is a graph of the gain for an optimized, non-uniform surface. Using this surface with a coplanar antenna, it was determined that the pattern of capacitance between the metallic patches could be optimized to increase the gain of the antenna mounted on the surface. In this way, the effective aperture size could be extended to cover the entire area of the surface, as plotted in FIG. **9**, which indicates that by proper design of a non-uniform impedance surface, power can be collected over a relatively large area (for example a relatively large area may be as large as perhaps hundreds or thousands of square of wavelengths of the impinging radiation), and routed to a single diode on the non-uniform impedance surface for rectification. For RF space power applications, the surface texture would consist of a lattice of fixed capacitors, built into metallized plastic. The capacitors are formed edge to confronting edge of the plates making up each lens. The afore-mentioned capacitors come from the fact that there are small metallic plates that are very close together. These are edge-to-edge capacitors, rather than conventional parallel plate capacitors. Any two conductors that are brought near each other will have some amount of capacitance. A ground plane is not needed here, but it could be used to provide improved efficiency, at the expense of greater weight. The values of the capacitors and the shape of the metal particles would be determined by electromagnetic simulations, and an optimization algorithm.

The results described above with reference to FIGS. **8***a*–**8***c* are for a two-layer structure containing vertical metallic vias—a high-impedance surface—that was built using printed circuit board technology as described in my issued U.S. patents and published U.S. patent applications. Lighter weight structures are needed in order to make this general concept practical and sufficiently lightweight for convenient use in space or even for use on an airship such as the airship shown in FIG. **2**. For example, the planar lenslets **40**, the collection antennas **60**, and the rectifier diodes **25** should preferably be built on a single surface that would preferably be printed on a single-layer plastic, dielectric film. Of significant importance would be the elimination of the vertical vias, and preferably the ground plane itself, leading to an entirely uniplanar structure. Simple uniplanar structures have been studied (See Yang, 1999, supra), but techniques for optimizing complex non-uniform surfaces need to be developed.

Furthermore, if the ground plane is eliminated, methods for minimizing transmission through the structure also would need to be considered. The structure could be analyzed as a complex parasitic array, where the individual patches in the patterned metallic surface could be considered as parasitic antennas. Their shape would be optimized so that the scattered power from each of them would be maximized at one point, where the rectifier diodes would be placed.

A microwave structure embodiment is depicted by FIG. **10**. FIGS. **10***a*, **10***b* and **10***c* provide section views through the embodiment depicted by FIG. **10**. In this embodiment a lattice of printed metallic lenslets **40**, each formed by arrays of thin metal patches **42**, focus power onto a sparse array of rectifier diodes **25**, which could be coplanar with the lenslets or mounted on the adjacent patches **42** as shown by FIG. **10***c*. DC power lines **65** (which are preferably incorporated into the structure) could then carry power from diodes **25** to the satellite **10**, for example, for distribution to the onboard electronic systems. The entire rectenna could be printed on a thin, lightweight, plastic film **50**, which could be unfolded to cover an area comprising many square meters. Like all rectennas, it would not need to assume a particular shape, because rectification is done right at each antenna element. However, since each lenslet would provide some directivity, the surface would need to be roughly pointed toward (i.e. be orthogonal to) the source of energy, such as the ground station **10** source. The required pointing accuracy, among other things, would govern the size of the lenslets **40**.

A ground plane may be helpful in some embodiment. It could increase the efficiency, by not allowing any energy to pass through the structure. The metallic pattern on the top of film **50** would be qualitatively similar to that without the ground plane, but in detail it would probably be a different pattern to compensate for the presence of the ground plane. The ground plane would have to be separated from the top metal patterns by some distance, typically 1/100 to 1/10 wavelength, depending on the tolerances allowed in the manufacturing of the metallic patterns. (This is not due to the tolerance of the film thickness. It is due to the fact that the overall thickness will affect the bandwidth. If the bandwidth is very narrow, then the metallic patterns will have to be defined very accurately to get the capacitance right.) In order to allow some spacing, but not to have a very heavy structure, an embodiment with a ground plane **44** may be ribbed, air-filled structure **46**, such as that seen in FIG. **11**. This might be similar to flexible "bubble wrap", or a rigid honeycomb-like dielectric that is commonly used in airframes and other such things.

In summary, the rectenna consists of a rectifying diode **25** and a generally planar lens structure **40**. The lens structure comprises a thin dielectric (such as plastic) sheet **50** that is patterned with metallic regions **42**. The metallic regions **42** scatter electromagnetic energy, and they are arranged so that the collective scattered energy from all of them is focused into the diode **25**. Each rectifying diode **25** is attached between two adjacent ones of the metal regions **42**. The diodes **25** are also attached to long conductive paths **46** (wires) that traverse the entire width of the structure, or are otherwise routed so that they supply current to a common location (such as an edge) where it may be collected and used to supply electrical power to a satellite or other device. The wires **46** are preferably coplanar with the metal patches **42** that make up the lens **40**, and they are preferably oriented transverse to the expected polarization of the energizing RF field, so that they have a minimum scattering effect. The metal pattern of the lens **40** can also be optimized to account for the scattering of the wires **46**. The lens **40** and indeed the thin dielectric sheet **50** preferably have a planar configuration and indeed the rectenna, when designed, will very likely be assumed to have a planar configuration in order to simplify its design (see the foregoing discussion). But those skilled in the art should appreciate the fact that the sheet **50** may well assume a non-parallel configuration in use, either by design or by accident. If designed for a planar configuration, the extent by which the in-use sheet **50** deviates from a planar configuration will adversely affect its effectiveness. But if the in-use design is close to being planar, the loss in efficiency is likely to be very small. Of course, the rectenna can be designed initially with a non-planar configuration in mind, but a non-planar configuration will doubtlessly complicate finding a desirable arrangement of the patches **42** for the various lenslets **40**. Making an assumption that the sheet **50** and the lenslets **40** will all be planar should simplify the design of the rectenna significantly.

The lenses (or lenslets) **40** are ideally designed and optimized using a computer. A random collection of scatterers is simulated, and the collected power is calculated using an electromagnetic solver. The sizes, shapes, and locations of the scatterers are varied according to an optimization method. Such methods are known to those skilled in the art, and include the method of steepest descent, genetic algorithms, and many others. The geometry that provides the greatest power to the diode **25** is then apt to be chosen as the ideal structure.

Such methods are good for determining the best geometry when nothing is known about that geometry beforehand. However, in the case of the present invention, much is known about the required geometry, and one can design a simple structure by hand. The preferred design method is then to start with a known good structure using the calculations described below, and then to optimize it using a computer as described above.

It can be shown that a wave having wave vector k_0, propagating on a periodic structure with effective refractive index n_{eff} will be scattered by the periodicity of that structure k_p to an angle θ given by:

$$\theta = \mathrm{Sin}^{-1}\left(\frac{k_0 n_{eff} - k_p}{k_p}\right) \qquad [4]$$

The planar lens structure should be designed so that energy scatters from the normal direction (θ=0°) into the plane of the surface where the diode is located. Assuming that the dielectric layer is thin, we have $n_{eff}=1$, so we are left with $k_p=k_0$. Therefore, the periodicity of the structure should be roughly one free-space wavelength.

In order to have independent control over the magnitude and phase of the radiation from the feed point, (or conversely in the present case, the collected energy at the diode **25**) it is necessary to have the periodicity be much greater. For independent control over two parameters, the array should be oversampled by a factor of at least two, which means that the individual metal patches **42** should be spaced at most one-quarter wavelength apart, with their properties varying periodically on a length scale of one wavelength. The structure should have close to radial symmetry, so that energy is scattered inward toward a central point. However, the symmetry can vary from perfect radial symmetry to account for polarization effects (leading to a slight deviation which has mirror symmetry) or for practical reasons due to the discrete nature of the individual patches **42**. An example of such a structure is shown in FIG. **12**.

FIG. **12** depicts an array of metal plates **42** located on centers spaced with a period of one-quarter wavelength, and the features thereof (size in this embodiment) vary on a length scale of one wavelength, with radial symmetry. The scattered energy from the metal plates **42** combines coherently at the diode **25** located at the center of the geometric pattern formed by plates **42**.

This single planar lens **40** consists of metal patches **42** having a periodicity of one-quarter wavelength, and having properties (the patch size in this embodiment) varying with a period of one wavelength. The planar lens **40** shown has a diameter of about four wavelengths. It collects power over its entire surface, and directs it toward the diode **25** at the center of the pattern, which diode is preferably connected between a pair of the closest patches **42**. This lens **40** forms a single element of a larger array **65**, shown in FIG. **13**, in which the diodes **25** are also connected in parallel by rows.

FIG. **13** depicts an array of planar lenslets or lenses **30**, each having a diode **25** at the center thereof, with those diodes **25** being connected by DC lines **46**. The lines are preferably oriented transverse to the electric field of the incoming radiation, so that they do not interfere significantly with the scattered waves. They can be printed on the same side of the sheet **50** (see FIG. **10***b*) as the metallic patches **42**, in which case the metal pattern of the lines **46** would simply be combined with that of the patches **42**, or they can be printed on the reverse side of sheet **50**, and attached to the diodes **25** by small metal plated via holes **48** in the plastic sheet **50**, as shown by FIG. **14**.

This design requires far fewer diodes than do conventional rectennas, because the diodes **25** are spaced every four wavelengths, rather than every half-wavelength. The result is a factor of close to 64 times reduction in the number of required diodes, and a corresponding factor of 64 times increase in the voltage generated per diode. This is particularly useful in cases where the incoming power density is low (such as space applications), where it would otherwise be difficult to get the induced voltage above the diode threshold voltage. Thus, this design also has higher efficiency due to the greater induced voltage at lower power levels.

Having described this technology in connection with certain embodiments thereof, modification will now doubtlessly suggest itself to those skilled in the art. As such, the protection afforded hereby is not to be limited to the disclosed embodiments except as is specifically required by the appended claims.

What is claimed is:

1. A rectenna structure comprising:

a sheet of a dielectric material;

a plurality of metallic lenslets disposed on the sheet of dielectric material; and

a plurality of diodes disposed on or adjacent the sheet of dielectric material, each diode in said plurality of diodes being arranged at a focus of a corresponding one of said plurality of metallic lenslets.

2. The rectenna structure of claim **1** wherein the metallic lenslets each comprise a geometric arrangement of metallic patches.

3. The rectenna structure of claim **2** wherein the focus of each lenslet corresponds to a center of the geometric arrangement of metallic patches comprising the lenslet.

4. The rectenna structure of claim **2** wherein the rectenna is designed to be responsive to incident radiation for converting the incident radiation to electrical energy and wherein metallic patches in each of said geometric arrangements have centers which are spaced from centers of neighboring metallic patches by a distance equal to one-quarter wavelength of said incident radiation.

5. The rectenna structure of claim **4** wherein the patches in each of said lenslets have a property that varies along a radial direction from the focus of the lenslet with a period equal to one wavelength of said incident radiation.

6. The rectenna structure of claim **5** wherein the property that varies along a radial direction from the focus of the lenslet is the geometric size of the individual patches.

7. The rectenna structure of claim **6** wherein the geometric arrangement is a hexagonal arrangement.

8. The rectenna structure of claim **7** wherein the individual patches each have a hexagonal shape when viewed in a plan view.

9. The rectenna structure of claim **6** wherein the geometric arrangement is a square arrangement.

10. The rectenna structure of claim **6** wherein the metallic lenslets have maximum dimension in a plan view thereof that is equal to four wavelengths of said incident radiation.

11. The rectenna structure of claim **2** wherein the geometric arrangement has two orthogonal axes of symmetry and wherein at least selected ones of said metallic patches associated with a particular geometric arrangement do not intersect either of the axes of symmetry of said particular geometric arrangement but rather are separated from the axes of symmetry of said particular geometric arrangement by predetermined distances.

12. The rectenna structure of claim **2** wherein the geometric arrangement is hexagonal.

13. The rectenna structure of claim **12** wherein the patches disposed in the hexagonal geometric arrangement are individually hexagonally shaped and are arranged in hexagonally shaped rings of hexagonally shaped patches, with neighboring rings comprising patches of different sizes.

14. The rectenna structure of claim **1** wherein each said lenses behave as a planar lens with a focal length equal to zero.

15. A method of making a rectenna structure comprising:

providing a sheet of dielectric material;

disposing a plurality of metallic lenslets on the sheet of dielectric material; and

disposing a plurality of diodes on or adjacent the sheet of dielectric material and arranging each diode of said plurality of diodes at a focus of a corresponding one of said plurality of metallic lenslets.

16. The method of claim **15** further including providing the metallic lenslets as a geometric arrangement of metallic patches.

17. The method of claim **16** further including arranging the focus of each lenslet to correspond with a center of said geometric arrangement of metallic patches.

18. The method of claim **15** further including designing the rectenna to be responsive to incident radiation for converting the incident radiation to electrical energy wherein metallic patches in each of said geometric arrangements have centers which are spaced from centers of neighboring metallic patches by a distance equal to one-quarter wavelength of said incident radiation.

19. The method of claim **18** wherein the patches in each of said lenslets have a property that varies along a radial direction from the focus of the lenslet with a period equal to one wavelength of said incident radiation.

20. The method of claim **19** wherein the property that varies along a radial direction from the focus of the lenslet is the geometric size of the individual patches.

21. The method of claim **20** wherein the geometric arrangement is a hexagonal arrangement.

22. The method of claim **21** wherein the individual patches each have a hexagonal shape when viewed in a plan view.

23. The method of claim **20** wherein the geometric arrangement is a square arrangement.

24. The method of claim **20** wherein the metallic lenslets have maximum dimension in a plan view thereof that is equal to four wavelengths of said incident radiation.

* * * * *

US007167090B1

(12) **United States Patent**
Mandal et al.

(10) **Patent No.:** **US 7,167,090 B1**
(45) **Date of Patent:** **Jan. 23, 2007**

(54) FAR-FIELD RF POWER EXTRACTION CIRCUITS AND SYSTEMS

(75) Inventors: **Soumyajit Mandal**, Cambridge, MA (US); **Rahul Sarpeshkar**, Arlington, MA (US)

(73) Assignee: **Massachusetts Institute of Technology**, Cambridge, MA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 88 days.

(21) Appl. No.: **10/944,676**

(22) Filed: **Sep. 17, 2004**

(51) **Int. Cl.**
G08B 1/08 (2006.01)

(52) **U.S. Cl.** **340/538.14**; 340/538.16; 340/572.7; 235/380

(58) **Field of Classification Search** 340/505, 340/538.14, 538.15, 538.16, 572.4, 572.7, 340/645; 343/741, 866; 607/61; 329/341, 329/370; 235/375, 380
See application file for complete search history.

(56) References Cited

U.S. PATENT DOCUMENTS

6,212,431	B1 *	4/2001	Hahn et al.	607/61
6,486,776	B1 *	11/2002	Pollack et al.	340/521
6,639,459	B1 *	10/2003	Devilbiss	329/341
6,693,599	B1 *	2/2004	Chia et al.	343/741

* cited by examiner

Primary Examiner—Van T. Trieu
(74) *Attorney, Agent, or Firm*—Daly, Crowley, Mofford, & Durkee, LLP

(57) ABSTRACT

A method and apparatus for performing far-field power extraction are presented. The method includes receiving an electromagnetic radiation signal, rectifying the signal to produce a direct current (D.C.) voltage and providing the D.C. voltage to a circuit. A far-field power extraction circuit includes an antenna for receiving an electromagnetic radiation signal, a rectifier for rectifying the electromagnetic radiation signal. The circuit may further include a charge pump for amplifying the rectified voltage, an impedance matching network for coupling the antenna to the rectifier and a feedback tuning circuit for optimizing performance of the extraction circuit.

25 Claims, 8 Drawing Sheets

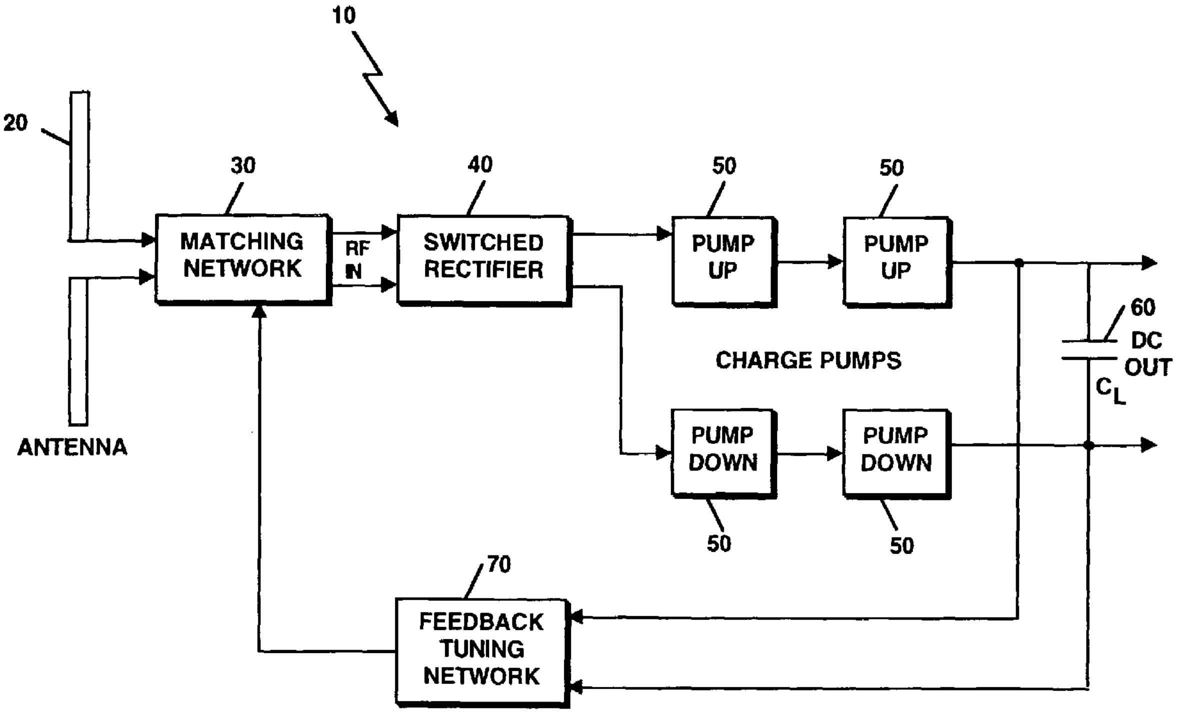

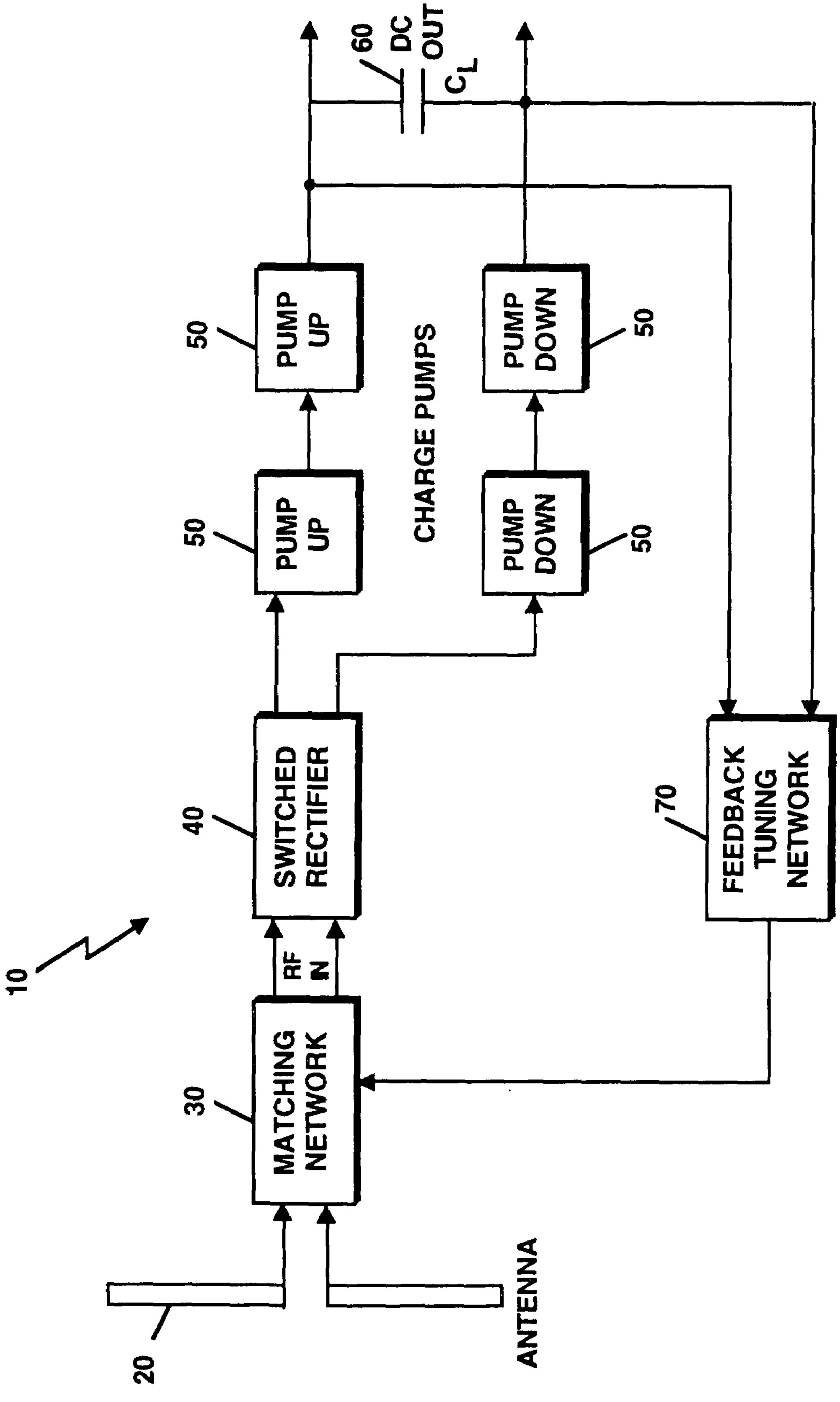
10
20
ANTENNA
30
MATCHING NETWORK
RF IN
40
SWITCHED RECTIFIER
50
PUMP UP
50
PUMP UP
CHARGE PUMPS
PUMP DOWN
50
PUMP DOWN
50
60
DC OUT
C_L
70
FEEDBACK TUNING NETWORK

Figure 1

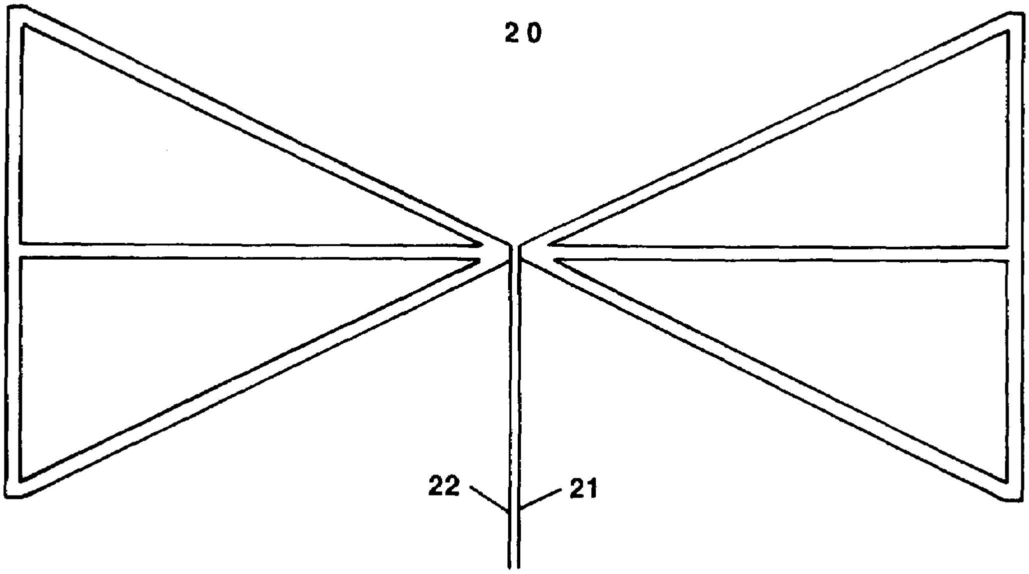

Figure 2

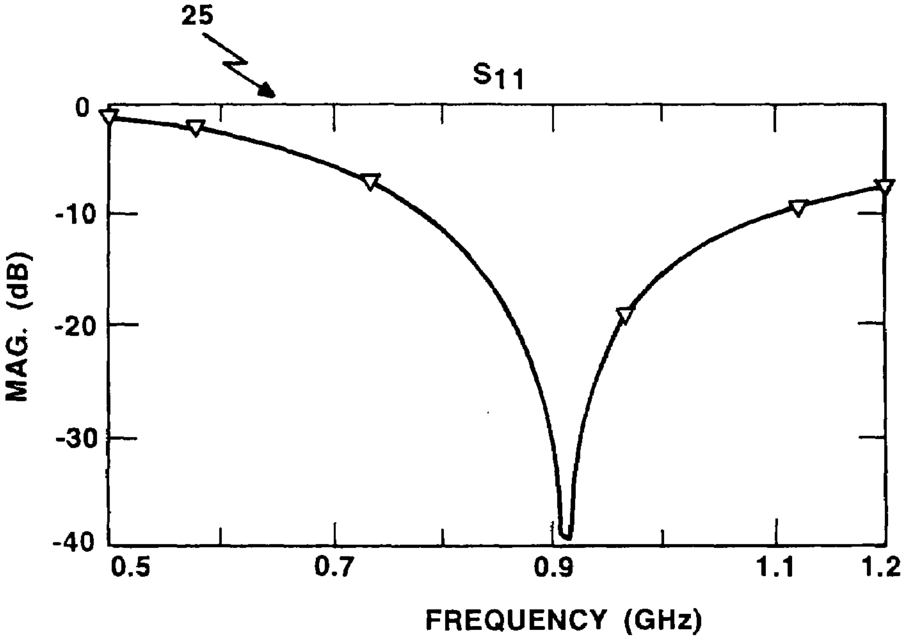

Figure 3

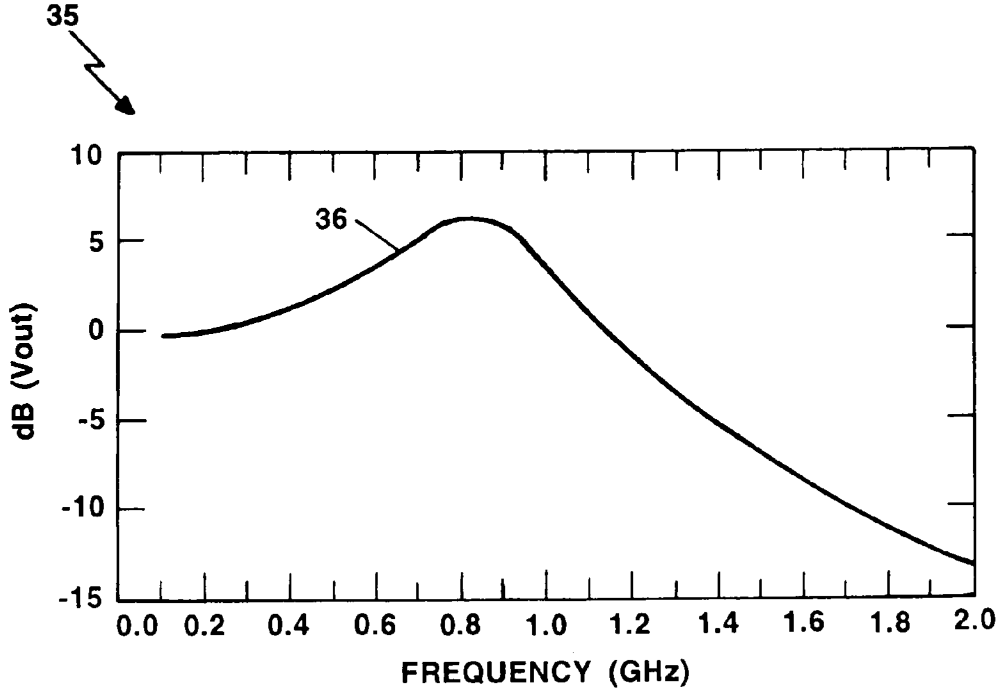

Figure 4

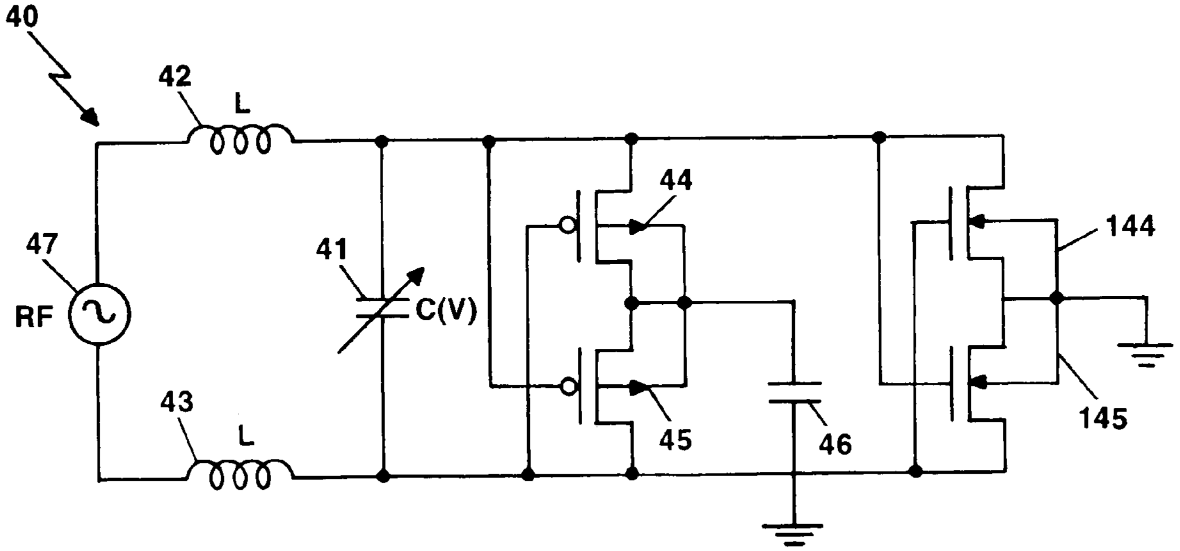
40
42
L
47
RF
41
C(V)
43
L
44
45
46
144
145

Figure 5

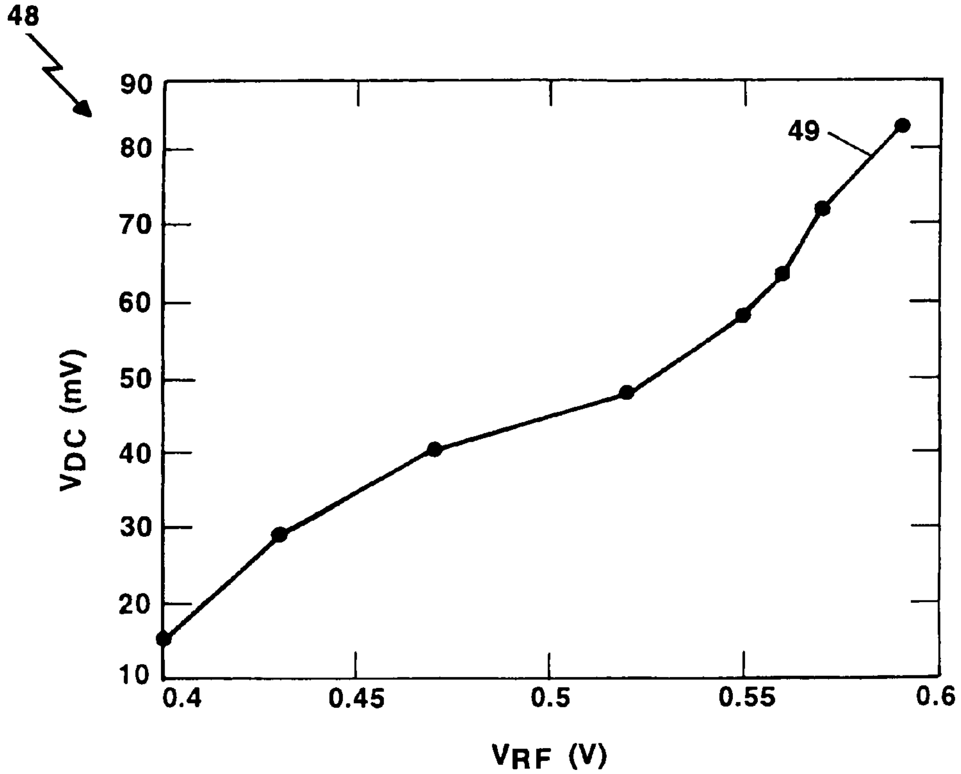

Figure 6

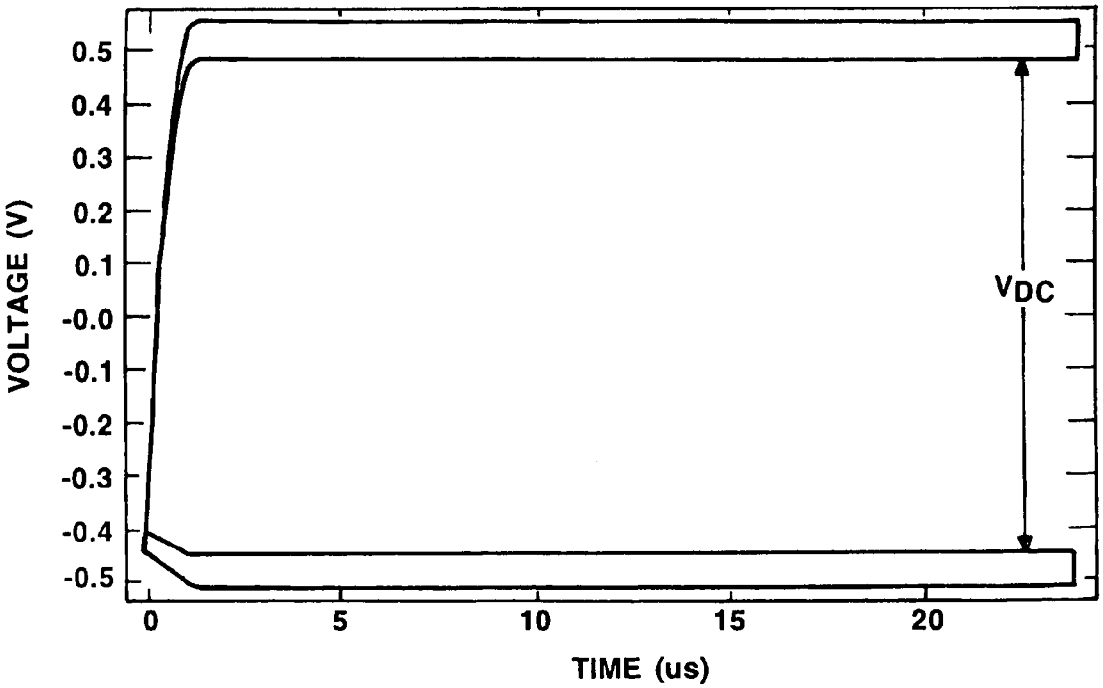

Figure 7

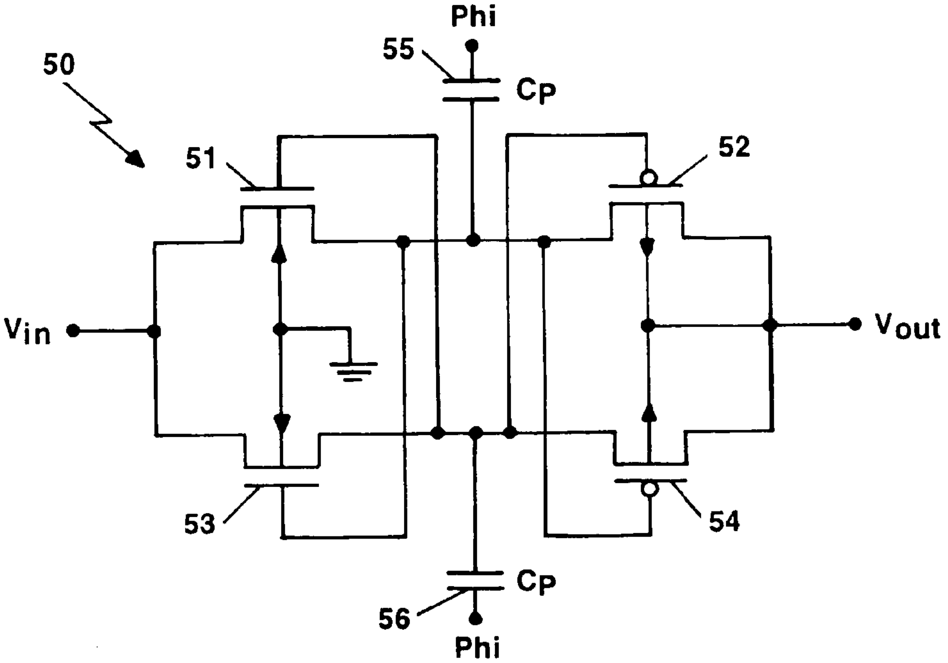
Phi
55
Cp
50
51
52
Vin
Vout
53
54
Cp
56
Phi

Figure 8

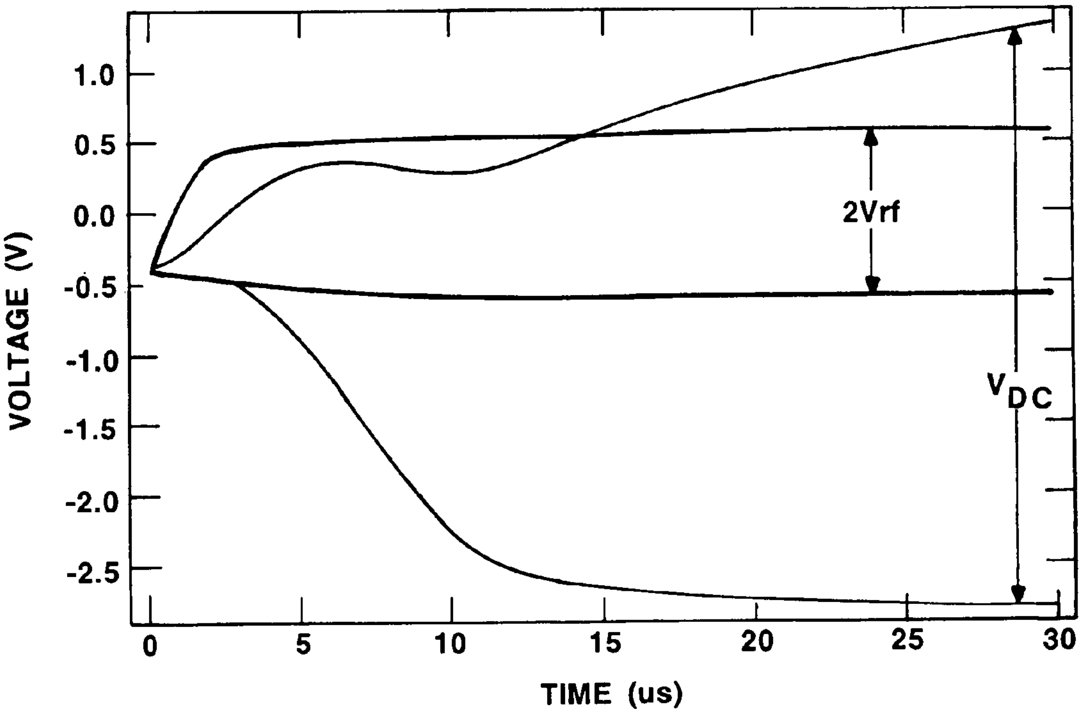

Figure 9

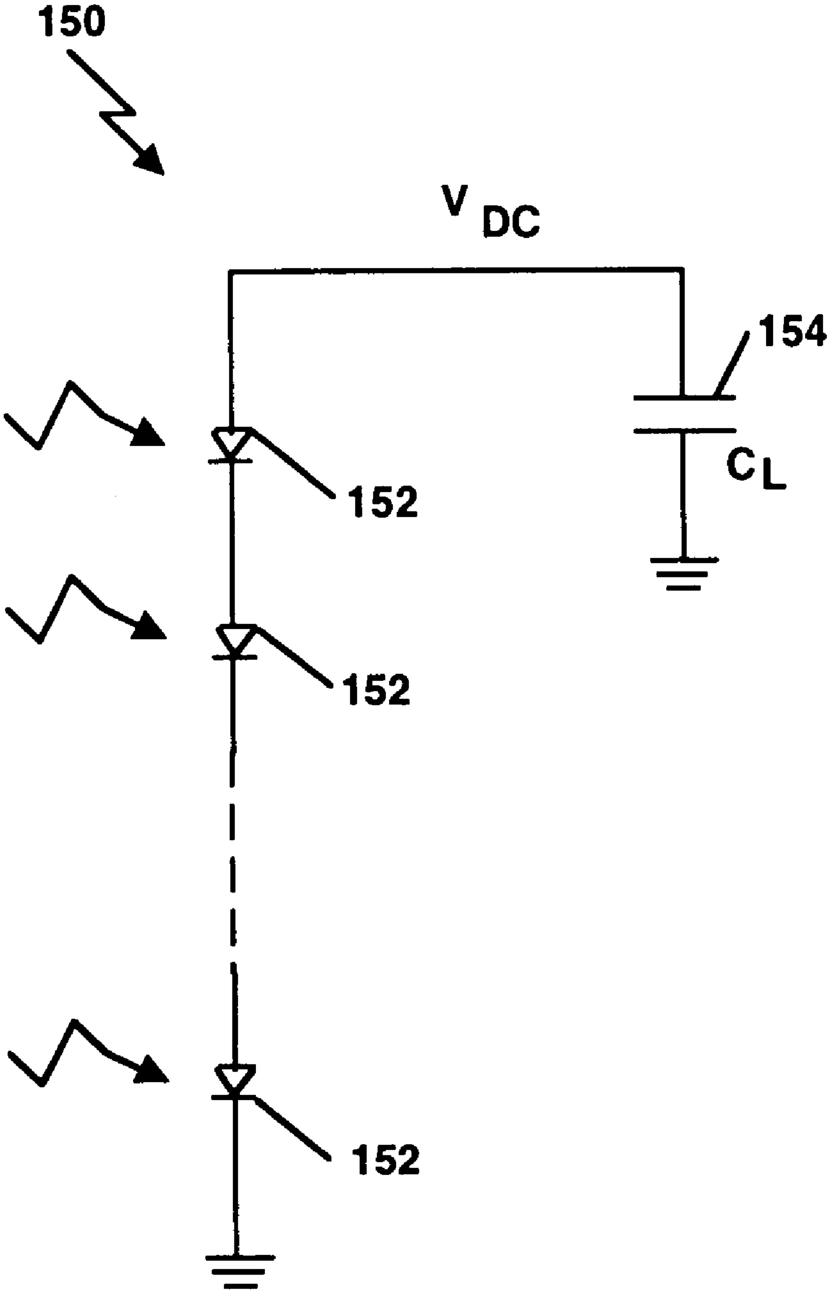
150
V DC
154
C L
152
152
152

Figure 10

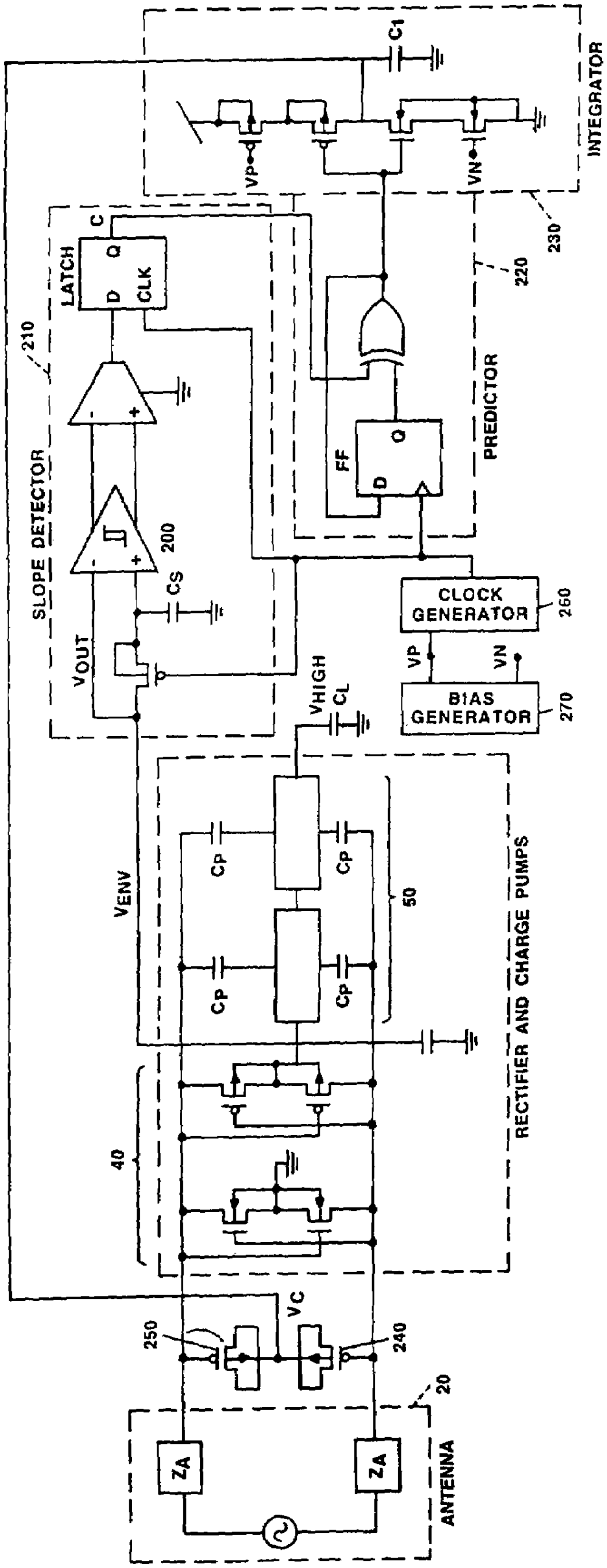
C1
INTEGRATOR
VP
VN
230
220
C
LATCH
D
Q
CLK
210
SLOPE DETECTOR
PREDICTOR
FF
D
Q
200
CS
VOUT
VHIGH
CL
CLOCK GENERATOR
260
VP
VN
BIAS GENERATOR
270
VENV
CP
CP
CP
CP
50
RECTIFIER AND CHARGE PUMPS
40
VC
250
240
20
ZA
ZA
ANTENNA

Figure 11

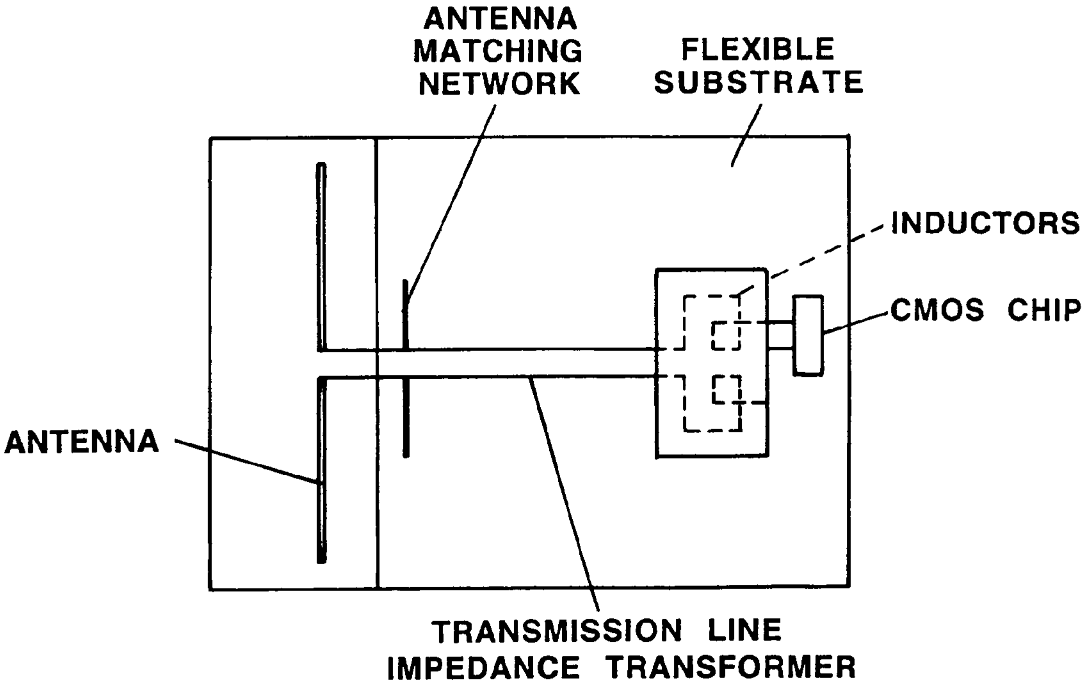
ANTENNA MATCHING NETWORK
FLEXIBLE SUBSTRATE
INDUCTORS
CMOS CHIP
ANTENNA
TRANSMISSION LINE IMPEDANCE TRANSFORMER

Figure 12

FAR-FIELD RF POWER EXTRACTION CIRCUITS AND SYSTEMS

CROSS REFERENCE TO RELATED APPLICATIONS

Not Applicable.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

Not Applicable.

FIELD OF THE INVENTION

The present invention relates generally to power extraction and more particularly to circuits and systems for extracting power from radio frequency (RF) signals.

BACKGROUND OF THE INVENTION

Devising efficient methods for extracting DC power from electromagnetic radiation has become an important necessity for a number of applications involving self-powered devices, such as Radio Frequency Identification (RFID) tags and bionic implants. The operating range of such self-powered devices has been severely limited by the failure of existing power extraction techniques to successfully extract power from radio frequency (RF) signals having relatively low power levels. The problem of extracting DC power from electromagnetic radiation has two basic parts: collecting the incident radiated power, and then converting the collected power to DC signals which are usable by the self-powered devices.

Converting RF energy from RF signals at different frequencies is a relatively difficult problem particularly when the RF signals have relatively low power levels. Fundamentally, this problem arises because frequency conversion is generally a nonlinear operation, but all practical systems are essentially linear for small signals. In addition, nonlinear devices normally used for rectification have exponential nonlinearities with relatively large "dead zones" near the origin, i.e., nonlinear devices can be non-responsive in response to signals having voltage and current levels which are close to zero. Severe constraints can also be imposed when it is desirable to provide a self-powered device which is relatively inexpensive and environmentally robust. This limitation precludes the use of exotic devices and structures.

SUMMARY OF THE INVENTION

A method and apparatus for performing far-field power extraction are disclosed. The method includes receiving an electromagnetic radiation signal, rectifying the signal to produce a direct current (D.C.) voltage and providing the D.C. voltage to a circuit. A far-field power extraction circuit includes an antenna for receiving an electromagnetic radiation signal, a rectifier for rectifying the electromagnetic radiation signal. The circuit may further include a charge pump for amplifying the rectified voltage, an impedance matching network for coupling the antenna to the rectifier and a feedback tuning circuit for optimizing performance of the extraction circuit.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be more fully understood from the following detailed description taken in conjunction with the accompanying drawings, in which:

FIG. **1** is a block diagram of a power extraction system;

FIG. **2** is a diagram of a broadband planar dipole antenna for RF power extraction;

FIG. **3** is a plot showing a reflection characteristic (into a 50Ω load) of the antenna shown in FIG. **2**;

FIG. **4** is a plot of the AC transfer characteristic of an input matching network for a loaded Q of 2;

FIG. **5** is a schematic diagram of a rectifier circuit including a nonlinear capacitor and a switching rectifier;

FIG. **6** is a plot of a rectification characteristic of the rectifier circuit shown in FIG. **5**;

FIG. **7** is a plot of the simulated response of the rectifier circuit of FIG. **5** in response to an input RF signal having a peak-to-peak amplitude of 0.5 V;

FIG. **8** is a schematic diagram of a charge pump cell;

FIG. **9** is a plot showing the stimulated response of the charge pump-switching rectifier combination shown in FIG. **1** in response to an input RF signal having a peak-to-peak amplitude of 0.55 V;

FIG. **10** is a schematic diagram of a solar cell circuit;

FIG. **11** is a schematic diagram of a controller; and

FIG. **12** is a block diagram of an RF power extraction system.

DETAILED DESCRIPTION OF THE INVENTION

The present invention comprises various combinations of different circuits and techniques to efficiently extract power from electromagnetic signals having relatively low electromagnetic field strengths in the far field of the transmitter, thereby substantially reducing the power threshold required for operation of self-powered devices.

Referring now to FIG. **1**, a block diagram of a system **10** for extracting power from electromagnetic radiation includes an antenna **20** for receiving a radio frequency (RF) signal from which direct current (DC) power will be harvested. A matching network **30** is coupled to the antenna **20** to impedance match the antenna **20** to the remainder of the system, thereby obtaining efficient power transfer. A switched rectifier **40** operates on the impedance matched differential RF signal and converts the signal to one or more DC levels. The output of the switched rectifier **40** is coupled to a group of charge pumps **50**, which increase the DC levels of the voltage. The output of the charge pumps is coupled across a load capacitor **60** to provide the DC output voltage. A feedback tuning circuit **70** is coupled to the output voltage and feeds into the impedance matching network **30** to provide continuous tuning of the impedance matching network to get the maximum possible DC output.

Referring now to FIG. **2**, the antenna **20** is shown. The antenna bandwidth is defined as the frequency range over which the reflection coefficient of the antenna into a specified load impedance is less than some specified value, typically −10 dB (VSWR=2). This antenna has improved bandwidth over a simple linear dipole, while occupying approximately the same area. In a preferred embodiment the antenna **20** thus comprises a wideband dipole antenna.

It should be appreciated that although a particular dipole antenna design is shown in FIG. **2**, it may be desirable to provide the antenna as a planar loop, dipole or fractal antenna. In one embodiment, the antenna **20** is provided on a flexible substrate. One objective for power collection applications is to provide a planar antenna on a flexible substrate that produces the maximum possible open circuit voltage V_{OC} across the antenna terminals **21** and **22** for a given incident field strength. In addition, the antenna design has to have sufficient bandwidth to withstand bending and proximity effects, such as attachment to a dielectric surface, without moving too far off resonance. Since passive RFID systems typically use backscatter modulation to communicate with the tag reader, another factor of interest is optimization of the Radar Cross Section (RCS) modulation capabilities of the antenna.

To address these issues, the use of one or more of planar loop, dipole, bow-tie and fractal antennas is presented. Loop antennas are advantageous because most proximity (near field) effects in practice are caused by dielectric materials. Since the near field energy of loop antennas is primarily stored in the magnetic field, they are typically less susceptible to these effects than other antenna types. Bow-tie antennas are desirable when a large impedance matching bandwidth is desired, but typically require large amounts of area in order to achieve this bandwidth. Fractal antenna structures are of interest in this application since they allow the bandwidth to be increased without consuming more area, or by reducing the area required to achieve a given bandwidth. Photonic Band Gap (PBG) substrates which reduce losses due to surface wave propagation in the flexible substrate may also be utilized. PBG substrates have electrical properties (like dielectric constants) which are periodic functions of space. Solutions of Maxwell's equations in such a medium have a 'stop band', or forbidden frequency range, where no surface wave propagation is possible. This may be utilized in our application by making the stop band lie in the frequency range where the main surface wave modes propagate, thereby preventing energy loss due to such (undesirable) modes and improving the efficiency of power extraction.

Referring now to FIG. **3**, a plot **25** of a simulated reflection characteristic for the antenna **20** is shown. In graph **25**, S_{11} is the reflection coefficient of the antenna and the substrate material used for simulations was FR-4. When simulated with a commercially available Method of Moments (MoM) based field solver, and using a criteria of a voltage standing wave ratio (VSWR) of two ($VSWR \leqq 2$) this antenna **20** exhibited a fractional bandwidth of approximately 40% at a center frequency of about 900 MHz. In this implementation, the antenna occupied an area of approximately 5"×2.5".

Referring again to FIG. **2**, another technique used in performing far-field RF power extraction comprises utilizing package parasitics to increase the input voltage levels to the rectifier. The far field case, when the input amplitude of the RF signal is not large enough to efficiently operate typical rectifying devices such as Schottky diodes are of particular interest. To overcome this problem, the high-Q input matching network **30** is used to passively amplify the input RF voltage. The inductance L and capacitance characteristics C(V) of network **30** are adjusted to include the effects of parasitic inductances and capacitances introduced by the chip packaging. In this way, a use for these normally unwanted parasitic characteristics has been found, as they now function as part of the matching network **30**. Of course, increasing the Q also decreases the frequency range over which the system can operate and increases its sensitivity to environmental conditions, which cause the resonant frequency to drift with time. This necessitates the use of active resonant frequency control. This is implemented by the feedback tuning network shown in FIG. **1**.

Another technique used in performing far-field RE power extraction comprises using traveling wave architectures for distributed voltage amplification and rectification. The matching network **30** is provided having both high voltage gain and high bandwidth at the input of the rectifier by using a cascade of exponentially tapered inductor-resistor-capacitor (L-R-C) transmission line segments. Each segment acts like a low pass filter with a certain Q and cutoff frequency. All segments have essentially the same Q, but have exponentially tapering cutoff frequencies. The cutoff frequency of the n-th section is given by:

$$f_n = exp(-n/N_{nat})$$

where f_n and N_{nat} are constants. Such a technique is useful for attaining high gain from many low-gain stages.

Referring now to FIG. **4**, a graph **35** which shows the AC transfer characteristic of the input matching network **30** for a loaded Q of 2 and a resonant frequency of 900 MHz is shown. The curve shows that a maximum output is achieved at approximately 900 MHz. The input to the matching network is assumed to be the antenna, and its output is fed to the switched rectifier.

Referring now to FIG. **5**, a rectifier circuit **40** is shown. One aspect of the present invention for performing efficient power extraction comprises using parametric-amplifier-like topologies and capacitive nonlinearities for rectification instead of exponential resistive nonlinearities. Exponential resistive nonlinearities (e.g., diodes) have traditionally been used for rectification but are unsuitable at low power levels. To solve this problem, parametric-amplifier-like topologies and capacitive nonlinearities are used. FIG. **5** shows how a nonlinear capacitor (e.g. a varactor) **41** is used for rectification. The varactor **41** (which can be a reverse-biased PN junction or a MOS capacitor) has a capacitance characteristic C(V), where V is the voltage of the control terminal. V is varied at the same frequency as the input RF signal. The RF signal is applied differentially, and, as an example, V may be tied to the upper plate of the varactor. In that case, C(V) will be different on the positive and negative halves of the RF signal cycle.

Inductors **42** and **43** and varactor **41** form a high-Q circuit and are chosen to resonate at the input frequency for some value of V. Since V varies as the RF, the resonant frequency and gain of the resonator will also be different on the positive and negative halves of the RF cycle. This asymmetric signal gain leads to the development of a DC component V_{DC} of voltage across the capacitor **46**, i.e., rectification.

The power extraction system can be adaptively adjusted for optimal performance by using floating gate transistors as adaptive elements. The threshold voltage of floating gate transistors can be changed by adding or subtracting charge from the floating gate. A lower threshold voltage improves the performance of the switching rectifier and charge pumps described in the following sections by increasing the rectified current for a given input RF amplitude.

In addition, the highest Q that can be used for the input LC tank shown in FIG. **2** is limited by resonant frequency variations caused by environmental factors and manufacturing tolerances. This limits the passive voltage gain obtainable from the tank. To overcome this problem, a high-Q system which maintains a constant resonant frequency by adapting the C(V) characteristic appropriately is utilized. This can be done by using a floating gate MOS capacitor for C(V).

Still another technique for performing far-field RF power extraction requires using switching rectifiers to avoid voltage drops associated with diode rectifiers. By using the differential RF inputs to operate transistors as switches and not as diodes, the threshold voltage drop associated with diode rectifiers is reduced considerably. FIG. **5** shows a circuit that achieves this. When the phase of the RF input is such that gate of the upper PMOS transistor **44** is low, it turns on, drawing current from the high side of the RF input, thereby charging the load capacitor **46**. The other transistor **45** is off during this phase. During the opposite RF phase, the roles of the two transistors are reversed, but the load capacitor is still charged upwards. The circuit thus acts as a full wave rectifier and charges the load capacitor C_L **46** towards the positive envelope of the input RF voltage V_{RF} sin(ωt). The final rectified DC voltage is determined by the resistance of the transistors and the load resistance connected to the output. It should be evident that by replacing the PMOS transistors with NMOS ones, the charging direction of the load capacitor can be reversed, i.e., charge can be made to flow out of the load capacitor, thereby decreasing the output voltage. Thus by adding a pair of NMOS devices **144**, **145** in parallel with the PMOS pair **44** and **45**, a negative voltage (referenced to the common mode voltage of the RF input) is generated, giving a total output DC voltage of approximately $2V_{RF}$ Note that MOSFET devices, being truly bidirectional devices, are ideally suitable for this circuit, where they are operated as switches This makes the integrated implementation of the circuit using standard low cost IC fabrication processes, such as CMOS, feasible.

FIG. **6** comprises a graph **48** which shows the simulated rectification curve **49** for the nonlinear capacitance when it was implemented as a MOS capacitor in a 0.5 μm CMOS process. The RF input amplitude was V_{RF} at 900 MHz. It can be shown that if the MOS devices operate in the sub-threshold region, this technique, which works by parametrically modulating the capacitance of a nonlinear capacitor **41**, allows the DC current charging the load capacitor C_L **46** to increase by a factor of cosh

$$\left(\frac{\kappa V_{DC}}{V_T}\right),$$

where K is the sub-threshold body bias coefficient and the thermal voltage

$$\left(\frac{\kappa V_{DC}}{V_T}\right),$$

FIG. **7** shows the results of a SPICE simulation **140** of the switching rectifier implemented in a 0.5 μm CMOS process with W/L=300 for both PMOS and NMOS devices. The input RF voltage amplitude V_{RF} was 0.5 V at 900 MHz. The charging time for a load capacitance C_L =50 pF was 1 μs.

Another technique for performing far-field RF power extraction uses charge pumps to increase the rectified output voltage. Since the input RF amplitude is extremely low (insufficient to operate the circuitry needed by the tag), charge pumps are used to increase the output DC voltage. FIG. **8** shows a single charge pump cell **50** using cross-coupled charge transfer switches. The differential RF input voltages ϕ and $\bar{\phi}$ are used to pump charge unilaterally through the pump capacitors C_P **55**, **56**, thus making $V_{OUT}>V_{IN}$. The circuit works as follows. When ϕ is high (=V_{RF} in this case), the bottom PMOS transistor **54** and the upper NMOS transistor **51** are turned on, while the other two transistors **52** and **53** are turned off. The other end of the pump capacitor **56** connected to ϕ is charged to its previous value (=V_{IN})+V_{RF} Current flows through the PMOS transistor **54** from this node, charging up the output towards (approximately) $V_{IN}+V_{RF}$. At the same time, the upper NMOS transistor **51** charges the other end of the pump capacitor **55** connected to $\bar{\phi}$ to V_{IN} The whole procedure is repeated during the opposite phase, when the RF input polarities are reversed. During this phase, the upper PMOS transistor **52** and the lower NMOS transistor **53** turn on, the other two transistors **51** and **54** turn off and the output is again charged towards $V_{IN}+V_{RF.}$

Ideally, $V_{OUT}=V_{IN}+V_{RF}$, i.e., a single cell acts as a voltage adder—it adds the RF amplitude to the input voltage. By cascading N of these cells in series, the output voltage is increased under no load conditions to $(N+1)V_{RF}$, where the input voltage to the first cell is assumed to be V_{RF}. Practically, parasitic capacitances to ground at the charge pumping nodes and increasing body bias effects on the NMOS devices limit the available voltage gain. This assumes that a typical n-well CMOS process is being used. This limitation can be removed if a more expensive dual-well process is used instead. In addition, by reversing the input and output terminals, the same circuit can be used to pump charge in the reverse direction and thereby generate large negative voltages. By combining charge pumps pumping in opposite directions with the switching rectifier described previously (which is assumed to generate approximately 2 V_{RF} across its outputs), an output DC voltage of $2(N+1)V_{RF}$ can be generated.

FIG. **9** shows the result of a SPICE simulation **57** of the switching rectifier **40** of FIG. **5** combined with the charge pump **50** shown in FIG. **8** implemented in a 0.5 μm CMOS process. Width (W)/Length(L)=60 for the devices in the switching rectifier, the pump capacitance C_P=1 pF and 2 cascaded cells were used to pump charge in each direction (N=2). The overall system was similar to that shown in FIG. **1**, with C_L=10 pF. The input RF amplitude V_{RF} was 0.55 V at 900 MHz.

The apparatus may utilize solar and other sources of ambient power for starting up the power collection module. For best performance, the power collection module can adapt to changing environmental conditions, for example by automatically adjusting the resonant frequency of the antenna using a feedback tuning network (shown in FIG. **1**). However, such adaptation loops consume power, which may not be available from the transmitted RF. To ensure startup power to run the adaptation loops, alternative sources of ambient power (apart from the transmitted RF) may be used. The amount of power required from these sources is small (typically in the nano watt range) since the system only needs to adapt slowly (time scale of milliseconds to seconds) and can thus be designed to have very low power consumption. Promising power sources include solar radiation (utilizing light energy using on-chip solar cells), ambient mechanical vibrations (using MEMS transducers), thermal gradients (using thermoelectric materials) and the like. FIG. **10** shows a solar cell structure **150** which can be used to provide startup power for the power collection module. The solar cell structure uses 3 solar diodes **152** in series to charge load capacitor **154**, thereby providing a D.C. voltage across the load capacitor **154**.

A feedback tuning network **70** is shown in FIG. **11**. The objective is to maximize the output DC voltage of the system, subject to a poorly known and possibly time varying open circuit RF amplitude across the antenna terminals. The primary cause of this is antenna resonant frequency variations. The resonant frequency (and other properties, such as the radiation pattern) of the antenna on the power collection module can vary by significant amounts because of variations in nearby environmental conditions. One way to minimize the effects of this variation is to use a broadband antenna, but this lowers the input Q and decreases the power up range. To prevent this, a feedback controller **70** is used to regulate the antenna resonant frequency (FIG. **11**). The controller uses derivative-based control to regulate the antenna resonant frequency. The controller comprises a state machine which operates on the temporal derivative of the rectified RF voltage and tries to keep it positive, i.e., increasing in time. The controller outputs a voltage which is used to control the capacitance of a MOS varactor, thereby controlling the resonant frequency of the antenna.

The controller shown in FIG. **11** can be implemented using subthreshold CMOS logic and consumes very little power. The controller can be powered off the rectified RF supply or from an alternative source of ambient energy, such as a solar cell. Two controllers running in different frequency ranges can be used to adapt the resonant frequency to variations that occur on different time scales. For example, a slow control loop could adjust to mechanical movement around the power collection module, while a faster loop could adjust to more rapidly varying quantities, such as multipath fading effects on the received RF signal.

The output of the switching rectifier (V_{ENV}, which is the DC voltage to be maximized) is provided to slope detector **210**. Slope detector **210** comprises a PMOS device **202** having a source coupled to the output of the rectifier and charge pumps (V_{ENV}), a gate coupled to the clock generator **260** and a drain coupled to storage capacitor **200**. The storage capacitor **200** is coupled between the drain of PMOS device **202** and a reference ground. A buffer **212** has a first input coupled to the output of rectifier and charge pumps and a second input coupled to the storage capacitor **200**, Buffer **212** provides a first buffer output and a second buffer output. A comparator **214** receives the buffer outputs and provides a comparator output. A latch **216** receives the output of comparator **214** and a clock input, and provides the slope detector output. In use, the output of the rectifier and charge pumps is sampled and held on the capacitor C_S **200** of slope detector **210**. This value is compared with the actual value of V_{ENV}. This operation is a discrete time approximation to the time derivative, and the output C of the slope detector **210** is a 1-bit estimate of the slope of V_{ENV}.

A predictor circuit **220** comprises a latch **222** receiving a clock input and a data input and providing an output to exclusive-or gate **224**. The exclusive-or gate **224** also receives the slope detector output and provides a predictor output which is also coupled to data input of latch **222**. In use, the predictor takes the current value of C, combines it with information about the previous correction made to the antenna resonant frequency and generates a control signal. This control signal is fed into an integrator **230**.

Integrator **230** comprises a PMOS device **232** receiving a positive bias voltage at a gate, having a source coupled to a reference voltage VP and providing an output at a drain. A second PMOS device **234** has a source coupled to drain of PMOS device **232**, a gate coupled to the output of predictor **220** and a drain providing an output of the integrator **230**. Integrator **230** further includes an NMOS device **238** receiving a negative bias voltage VN at a gate, having a source coupled to a reference ground and provides an output at a drain. The second NMOS device **236** has a source coupled to NMOS device **238**, a gate coupled to output of predictor **220** and a drain couple dot the drain of PMOS device **234** and also providing an output of integrator **230**.

The integrator **230** output voltage V_C controls the antenna resonant frequency by changing the capacitance of the MOS varactors **240**, **250** connected across the antenna output terminals. In the preferred implementation that is shown in FIG. **11**, the predictor control law is defined as:

$$\Delta V_{C,n+1} = C \oplus \Delta V_{C,n}$$

where $\Delta V_{C,n+1}$ is the new correction to be made to V_C, $\Delta V_{C,n}$ was the previous correction, and $\oplus$ denotes the logical XOR operation. This control law is that of a simple 'bang-bang' controller. Every time the controller makes a right decision (V_{ENV} increases and its slope C is positive), it repeats it on the next time step. Every time the controller makes a wrong decision (V_{ENV} decreases and its slope C is negative), it reverses its previous decision on the next time step. An oscillator (clock generator) **260** generates the sampling and timing signals for the rest of the system. Necessary current and voltage biases are generated by a bias generator **270**. Typically this takes the form of a supply-independent current reference circuit.

The physical structure of an RF power extraction system is shown in FIG. **12**. The input RF signal is captured by the antenna **20**. The antenna impedance is matched to the rest of the system using an appropriate matching networks **30**. The signal is then passively amplified using a high-Q resonator and fed to CMOS circuitry **80** which rectifies it and generates one or more DC voltages. These DC voltages are then used for powering other circuits on the chip **80**, such as an RFID transceiver. The entire system is disposed on a flexible substrate **90**.

The present application has focused on RFID tags operating in the UHF frequency band since it is of commercial importance, but the applicability of the present inventions is not confined to this application or frequency band. Since they are of a fairly general nature, the innovations described in thus application can be applied over a broad range of frequencies and power levels for various self-powered applications.

Currently, the minimum RF power threshold for self powered devices is in the 50–60 μW range. By way of implementing one or more of the techniques and circuits described above, this threshold is reduced to 3 μW or below. This results in a concomitant increase in the maximum read range by a factor of 4 over current designs.

Having described preferred embodiments of the invention it will now become apparent to those of ordinary skill in the art that other embodiments incorporating these concepts may be used. Additionally, the software included as part of the invention may be embodied in a computer program product that includes a computer useable medium. For example, such a computer usable medium can include a readable memory device, such as a hard drive device, a CD-ROM, a DVD-ROM, or a computer diskette, having computer readable program code segments stored thereon. The computer readable medium can also include a communications link, either optical, wired, or wireless, having program code segments carried thereon as digital or analog signals. Accordingly, it is submitted that that the invention should not be limited to the described embodiments but rather should be limited only by the spirit and scope of the appended claims. All publications and references cited herein are expressly incorporated herein by reference in their entirety.

What is claimed is:

1. A far-field power extraction circuit comprising:

an antenna configured to receive an electromagnetic radiation signal; and

a switched rectifier coupled to the antenna, the switched rectifier configured to rectify the electromagnetic radiation signal received by the antenna to produce a direct current (D.C.) voltage at an output, the switched rectifier comprising at least one circuit element having a nonlinear capacitive characteristic.

2. The circuit of claim **1** wherein the antenna comprises at least one of a loop antenna, a dipole antenna, a bowtie antenna or a fractal antenna.

3. The circuit of claim **1** wherein

the at least one circuit element having a nonlinear capacitive characteristic comprises a non-linear capacitor.

4. The far-field power extraction circuit of claim **3** wherein the nonlinear capacitor is a varactor.

5. The circuit of claim **1** wherein the switched rectifier comprises:

a first P-type Metal Oxide Semiconductor (PMOS) switch; and

a second PMOS switch wherein the nonlinear capacitor is coupled to an RF source, a first lead of the nonlinear capacitor is coupled to a source of the first PMOS switch and coupled to a gate of the second PMOS switch, a second lead of the nonlinear capacitor is coupled to a gate of the first PMOS switch and a drain of the second PMOS switch and wherein a drain of the first PMOS switch is coupled to a source of the second PMOS switch and to a load capacitor.

6. The circuit of claim **5**, further comprising a first N-type Metal Oxide Semiconductor (NMOS) switch coupled in parallel to the first PMOS Switch and a second NMOS switch coupled in parallel with the second PMOS switch.

7. A far-field power extraction circuit comprising:

an antenna configured to receive an electromagnetic radiation signal;

a switched rectifier coupled to the antenna, the rectifier configured to rectify the electromagnetic radiation signal received by the antenna to produce a direct current (D.C.) voltage at an output; and

an impedance matching network coupled between the antenna and the switched rectifier, the impedance matching network configured to match an antenna impedance with the remainder of the circuit, the impedance matching network comprising transmission line segments comprising inductor-resistor-capacitor line segments.

8. The circuit of claim **7**, further comprising a feedback tuning network coupled between the impedance matching network and the output voltage.

9. The circuit of claim **8** wherein the feedback tuning circuit comprises:

a bias generator;

a clock generator, wherein the bias generator is coupled to the clock generator;

a slope detector, wherein the slope detector is coupled to the switched rectifier and said charge pumps;

a predictor, wherein the clock generator is coupled to the slope detector and the predictor;

an integrator,

a first varactor; and

a second varactor wherein the integrator is coupled to the predictor and to the first and second varactors; wherein said first and second varactors are coupled to the antenna; and

wherein the feedback tuning circuit controls a resonant frequency of the antenna by changing a capacitance of the first and second varactors coupled across output terminals of the antenna.

10. The circuit of claim **9** wherein the slope detector comprises:

a PMOS device having a source coupled to the output of said switched rectifier, a gate coupled to the clock generator and a drain;

a storage capacitor coupled between the drain of the PMOS device and a reference ground;

a buffer having a first input coupled to the output of said switched rectifier and a second input

coupled to the storage capacitor, the buffer providing a first buffer output and a second buffer output;

a comparator receiving said first buffer output and the second buffer output and providing a comparator output; and

a latch receiving the output of the comparator and a clock input, the buffer providing a slope detector output.

11. The circuit of claim **9** wherein the predictor comprises:

a latch receiving a clock input and a data input and providing a latch output;

an exclusive-or gate receiving the latch output at a first input, receiving a slope detector output at a second output and providing a predictor output at the output and wherein the predictor output is coupled to the data input of the latch.

12. The circuit of claim **9** wherein the integrator comprises:

a first PMOS device receiving a positive bias voltage at a gate, having a source coupled to a reference voltage and providing an output at a drain;

a second PMOS device having a source coupled to the drain of the first PMOS device, a gate coupled to an output of the predictor and a drain providing an output of the integrator;

a first NMOS device receiving a negative bias voltage at a gate, having a source couple dot a reference ground and providing an output at a drain;

a second NMOS device having a source coupled to the first NMOS device, a gate coupled to an output of said predictor and a drain providing an output of the integrator.

13. The circuit of claim **9** wherein the first varactor comprises a PMOS device having a source and a drain coupled together and to the output of said integrator and having a gate coupled to said antenna.

14. The circuit of claim **9** wherein the second varactor comprises PMOS device having a source and a drain coupled together and to the output of the integrator and having a gate coupled to the antenna.

15. The circuit of claim **7** wherein the antenna and impedance matching network are responsive to radio frequency signals having a frequency of about 900 MHz.

16. The circuit of claim **7** wherein the transmission line segments have exponentially tapering cut-off frequencies.

17. The circuit of claim **7** wherein the impedance matching network comprises at least one inductor coupled between the antenna and the remainder of the power extraction system.

18. A far-field power extraction circuit comprising:

an antenna configured to receive an electromagnetic radiation signal;

a switched rectifier coupled to the antenna, the rectifier configured to rectify the electromagnetic radiation signal received by the antenna to produce a direct current (D.C.) voltage at an output; and

a charge pump coupled to an output of the switched rectifier,

wherein the charge pump comprises:

a first N-type Metal Oxide Semiconductor (NMOS) switch, wherein an input voltage is coupled to a source of the first NMOS switch, and wherein a substrate of the first NMOS switch is coupled to ground;

a second NMOS switch, wherein the input voltage is coupled to a source of the second NMOS switch, and wherein a substrate of the second NMOS switch is coupled to ground;

a first P-type Metal Oxide Semiconductor (PMOS) switch; and

a second PMOS switch, wherein a drain of the first PMOS switch, a drain of said second NMOS switch, a substrate of the first PMOS switch and a substrate of the second PMOS switch are coupled together and provide an output voltage.

19. The circuit of claim **18** wherein the charge pump is coupled to a positive output of the switched rectifier.

20. The circuit of claim **18** wherein the charge pump is coupled to a negative output of the switched rectifier.

21. The circuit of claim **18** wherein the charge pump comprises:

a first pump capacitor, wherein a gate of the first NMOS switch is coupled to a drain of the second NMOS switch, a gate of the first PMOS switch, a source of the second PMOS switch, and to a first lead of the first pump capacitor; and wherein a second lead of the first pump capacitor is coupled to a first reference signal; and

a second pump capacitor wherein a gate of the second NMOS switch is coupled to a drain of the first NMOS switch, a gate of the second PMOS switch, a source of the first PMOS switch, and to a first lead of the second pump capacitor; wherein a second lead of the second pump capacitor is coupled to a second reference signal.

22. A far-field power extraction circuit comprising:

an antenna configured to receive an electromagnetic radiation signal; and

a switched rectifier coupled to the antenna, the switched rectifier configured to rectify the electromagnetic radiation signal received by the antenna to produce a direct current (D.C.) voltage at an output, the switched rectifier comprising at least one of the group comprising a rectifier incorporating capacitive nonlinearities, a rectifier using transmission line segments, and a rectifier using one or more switching rectifier circuits,

wherein said switched rectifier incorporating capacitive nonlinearities comprises:

a variable capacitor;

a first P-type Metal Oxide Semiconductor (PMOS) switch; and

a second PMOS switch wherein said variable capacitor is coupled to an RF source, a first lead of said variable capacitor is coupled to a source of said first PMOS switch and coupled to a gate of said second PMOS switch, a second lead of said variable capacitor is coupled to a gate of said first PMOS switch and a drain of said second PMOS switch and wherein a drain of said first PMOS switch is coupled to a source of said second PMOS switch and to a load capacitor.

23. A far-field power extraction circuit comprising:

an antenna configured to receive an electromagnetic radiation signal; and

a switched rectifier coupled to the antenna, the switched rectifier configured to rectify the electromagnetic radiation signal received by the antenna to produce a direct current (D.C.) voltage at an output, the switched rectifier comprising at least one of the group comprising a rectifier incorporating capacitive nonlinearities, a rectifier using transmission line segments, and a rectifier using one or more switching rectifier circuits,

wherein the transmission line segments have exponentially tapering cut-off frequencies.

24. A far-field power extraction circuit comprising:

an antenna configured to receive an electromagnetic radiation signal;

a switched rectifier coupled to the antenna, the switched rectifier configured to rectify the electromagnetic radiation signal received by the antenna to produce a direct current (D.C.) voltage at an output, the switched rectifier comprising at least one of the group comprising a rectifier incorporating capacitive nonlinearities, a rectifier using transmission line segments, and a rectifier using one or more switching rectifier circuits; and

at least one charge pump coupled to an output of the rectifier,

wherein the at least one charge pump comprises:

a first N-type Metal Oxide Semiconductor (NMOS) switch, wherein an input voltage is coupled to a source of the first NMOS switch, and wherein a substrate of the first NMOS switch is coupled to ground;

a second NMOS switch, wherein the input voltage is coupled to a source of the second NMOS switch, and wherein a substrate of the second NMOS switch is coupled to ground;

a first P-type Metal Oxide Semiconductor (PMOS) switch; and

a second PMOS switch, wherein a drain of the first PMOS switch, a drain of said second NMOS switch, a substrate of the first PMOS switch and a substrate of the second PMOS switch are coupled together and provide an output voltage;

a first pump capacitor, wherein a gate of said first NMOS switch is coupled to a drain of said second NMOS switch, a gate of said first PMOS switch, a source of said second PMOS switch, and to a first lead of said first pump capacitor; and wherein a second lead of said first pump capacitor is coupled to a first reference signal; and

a second pump capacitor wherein a gate of said second NMOS switch is coupled to a drain of said first NMOS switch, a gate of said second PMOS switch, a source of said first PMOS switch, and to a first lead of said second pump capacitor; wherein a second lead of said second pump capacitor is coupled to a second reference signal.

25. A far-field power extraction circuit comprising:

an antenna configured to receive an electromagnetic radiation signal;

a switched rectifier coupled to the antenna, the switched rectifier configured to rectify the electromagnetic radiation signal received by the antenna to produce a direct current (D.C.) voltage at an output, the switched rectifier comprising at least one of the group comprising a rectifier incorporating capacitive nonlinearities, a rectifier using transmission line segments, and a rectifier using one or more switching rectifier circuits;

an impedance matching network coupled between the antenna and the switched rectifier, the impedance matching network configured to match an antenna impedance with the remainder of the circuit; and

a feedback tuning network coupled between the impedance matching network and the output voltage, the feedback tuning circuit comprising:

a bias generator;

a clock generator, wherein the bias generator is coupled to the clock generator;

a slope detector, wherein the slope detector is coupled to the switched rectifier and said charge pumps;

a predictor, wherein the clock generator is coupled to the slope detector and the predictor;

an integrator,

a first varactor; and

a second varactor, the integrator being coupled to the predictor and to the first and second varactors; wherein the first and second varactors being coupled to the antenna,

wherein the feedback tuning circuit controls a resonant frequency of the antenna by changing a capacitance of the first and second varactors coupled across output terminals of the antenna.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 7,167,090 B1 Page 1 of 2
APPLICATION NO. : 10/944676
DATED : January 23, 2007
INVENTOR(S) : Soumyajlt Mandal et al.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 2, line 57, delete “embodiment the” and replace with -- embodiment, the --.

Column 3, line 64, delete “RE power” and replace with -- RF power --.

Column 5, line 20, delete “2V_{RF} Note that” and replace with -- 2V_{RF}. Note that --.

Column 5, line 22, delete “switches This” and replace with -- switches. This --.

Column 5, line 45, delete “cosh $\left(\frac{\kappa V_{DC}}{V_T}\right)$” and replace with -- cosh $V_T = \frac{kT}{q}$ --.

Column 5, line 67, delete “(=V_{IN}) + V_{RF} Current” and replace with -- (=V_{IN}) + V_{RF}. Current --.

Column 6, line 4, delete “V_{IN}*The*” and replace with -- V_{IN}. *The* --.

Column 6, line 58, delete “D.C.” and replace with -- DC --.

Column 7, line 33, delete “of rectifier” and replace with --of the rectifier--.

Column 7, line 34, delete “capacitor 200, Buffer” and replace with --capacitor 200. Buffer --.

Column 7, line 58, delete “coupled to drain of” and replace with -- coupled to a drain of --.

Column 7, line 65, delete “to output” and replace with -- to an output --.

Column 7, line 66, delete “a drain couple dot the” and replace with -- a drain coupled to the --.

Column 8, line 25, delete “networks 30.” and replace with -- network 30. --

Column 8, line 38, delete “in thus application” and replace with -- in this application --.

Column 8, line 41-42, delete “Self powered” and replace with -- Self-powered --.

Column 8, lines 48-49, delete “invention it” and replace with -- invention, it --.

Column 8, line 62, delete “submitted that that the” and replace with -- submitted that the --.

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 7,167,090 B1 Page 2 of 2
APPLICATION NO. : 10/944676
DATED : January 23, 2007
INVENTOR(S) : Soumyajlt Mandal et al.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 10, line 51, delete “comprises PMOS device” and replace with -- comprises a PMOS device --.

Signed and Sealed this

Twenty-second Day of May, 2007

Jon W. Dudas

JON W. DUDAS
Director of the United States Patent and Trademark Office

US007768354B2

(12) United States Patent
Matsuda et al.

(10) **Patent No.: US 7,768,354 B2**
(45) **Date of Patent: Aug. 3, 2010**

(54) RADIO-FREQUENCY POWER AMPLIFIER

(75) Inventors: **Shingo Matsuda**, Kyoto (JP); **Hirokazu Makihara**, Osaka (JP); **Kazuki Tateoka**, Kyoto (JP); **Masahiko Inamori**, Osaka (JP)

(73) Assignee: **Panasonic Corporation**, Osaka (JP)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **12/399,452**

(22) Filed: **Mar. 6, 2009**

(65) **Prior Publication Data**

US 2009/0251220 A1 Oct. 8, 2009

(30) **Foreign Application Priority Data**

Apr. 8, 2008 (JP) 2008-100299

(51) **Int. Cl.**
H03F 3/04 (2006.01)

(52) **U.S. Cl.** **330/299**; 330/289

(58) **Field of Classification Search** 330/299, 330/289, 285, 296

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,448,859	B2	9/2002	Morizuka	
6,750,718	B2	6/2004	Moriwaki et al.	
6,946,913	B2	9/2005	Moriwaki et al.	
7,012,469	B2	3/2006	Moriwaki et al.	
7,019,508	B2	3/2006	Rategh et al.	
7,088,183	B2 *	8/2006	Kuriyama	330/296
7,154,336	B2 *	12/2006	Maeda	330/285

* cited by examiner

Primary Examiner—Henry K Choe
(74) *Attorney, Agent, or Firm*—McDermott Will & Emery LLP

(57) **ABSTRACT**

A bias circuit operable to supply a bias current to a first transistor includes: a second transistor having a collector terminal connected to a first power supply; a first resistance element having one end connected to an emitter terminal of the second transistor and having the other end connected to a base terminal of the first transistor; a second resistance element having one end connected to the emitter terminal of the second transistor and having the other end connected to ground potential; at least one third resistance element provided between a base terminal of the second transistor and a second power supply; and a plurality of temperature compensation circuits connected to the base terminal of the second transistor which are operable to control a base potential of the second transistor so that the potential falls as a temperature rises.

7 Claims, 9 Drawing Sheets

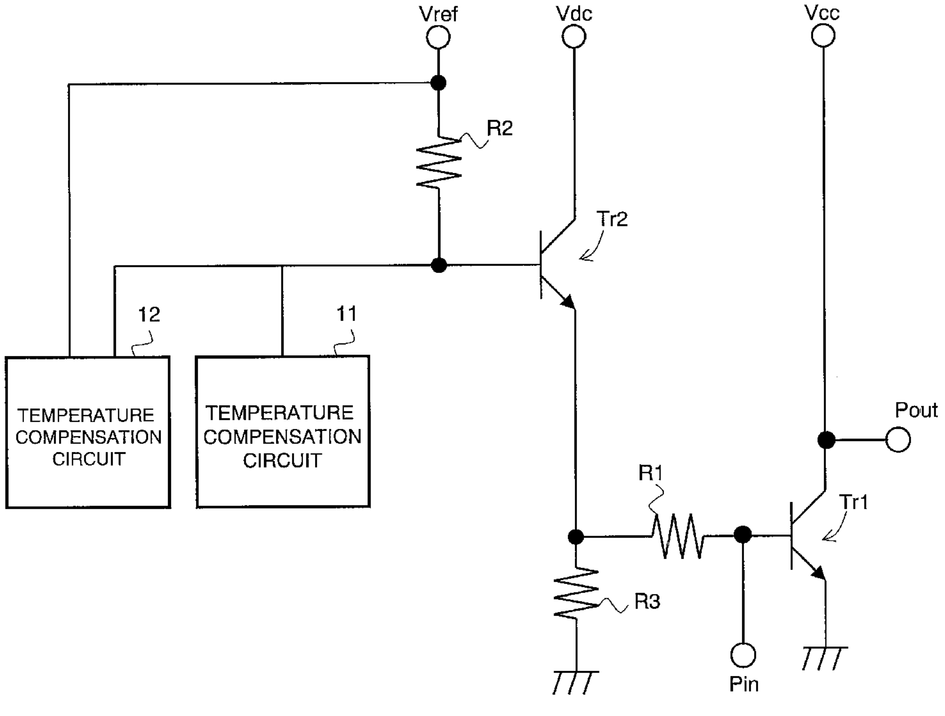
Vref
Vdc
Vcc
R2
Tr2
12
11
TEMPERATURE COMPENSATION CIRCUIT
TEMPERATURE COMPENSATION CIRCUIT
Pout
R1
Tr1
R3
Pin

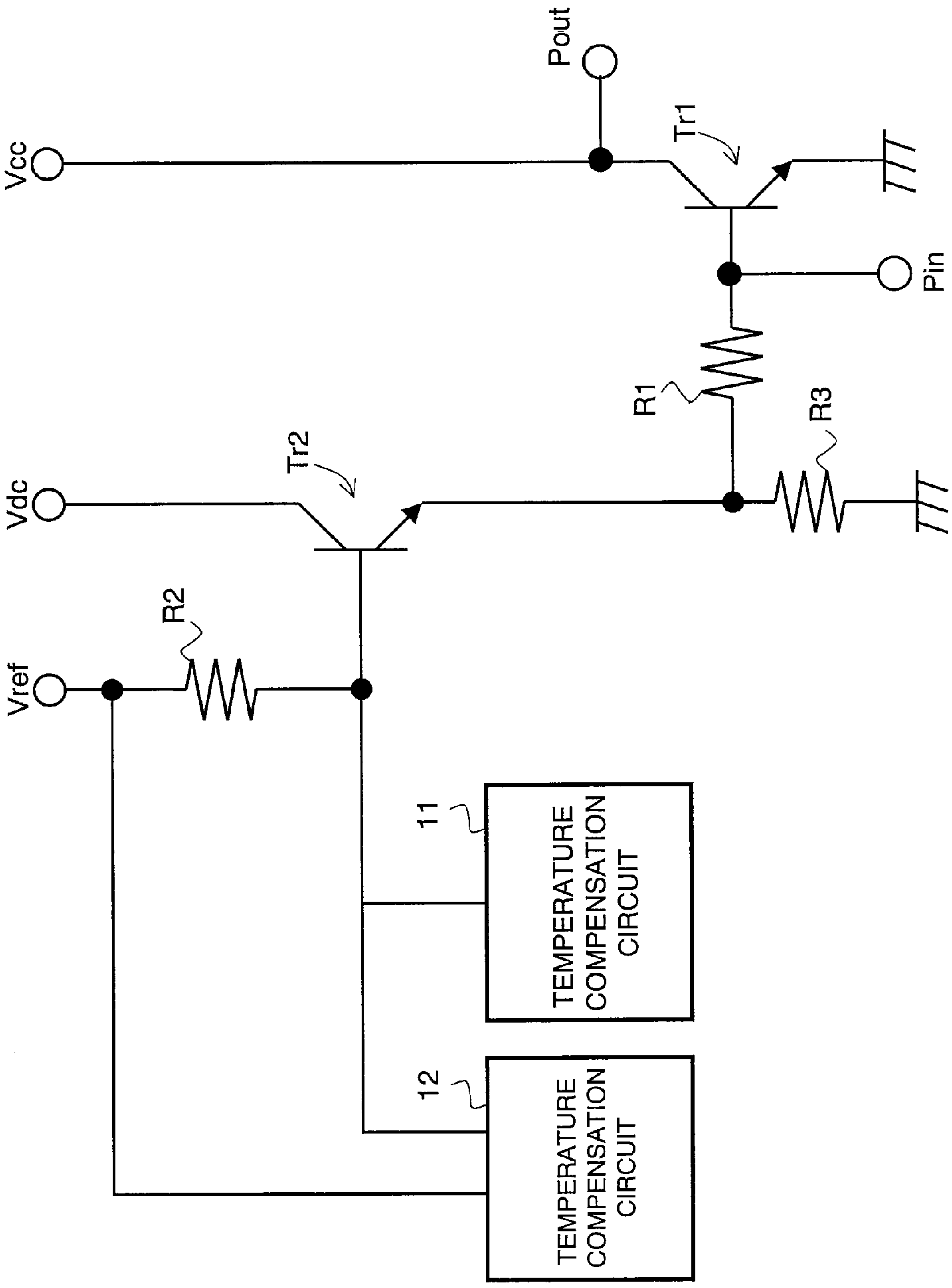
Vcc
Pout
Tr1
Pin
R1
R3
Vdc
Tr2
R2
Vref
11
TEMPERATURE COMPENSATION CIRCUIT
12
TEMPERATURE COMPENSATION CIRCUIT

FIG.1

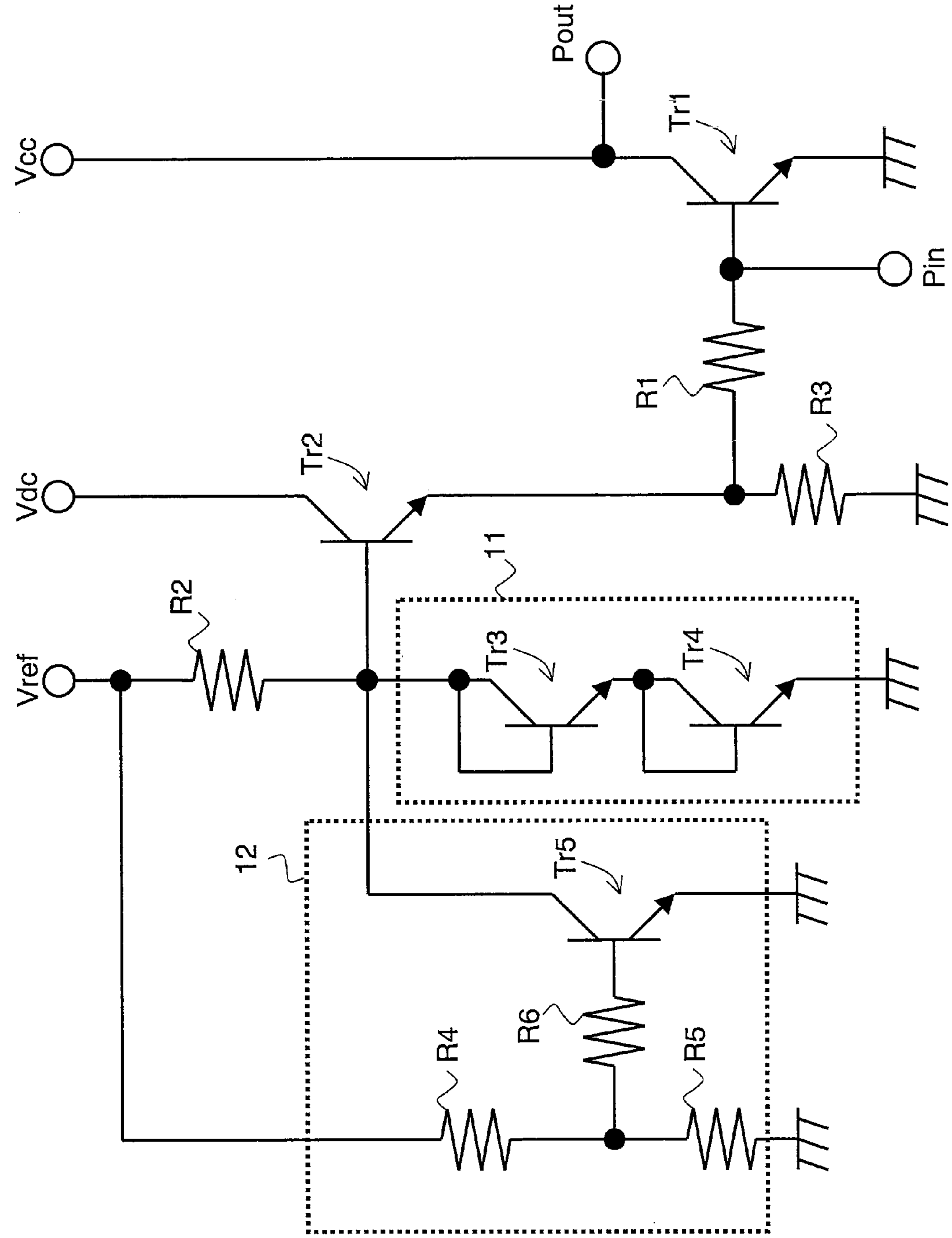
Vcc
Pout
Tr1
Pin
R1
R3
Vdc
Tr2
11
Vref
R2
Tr3
Tr4
12
Tr5
R4
R6
R5

FIG.2

FIG.3

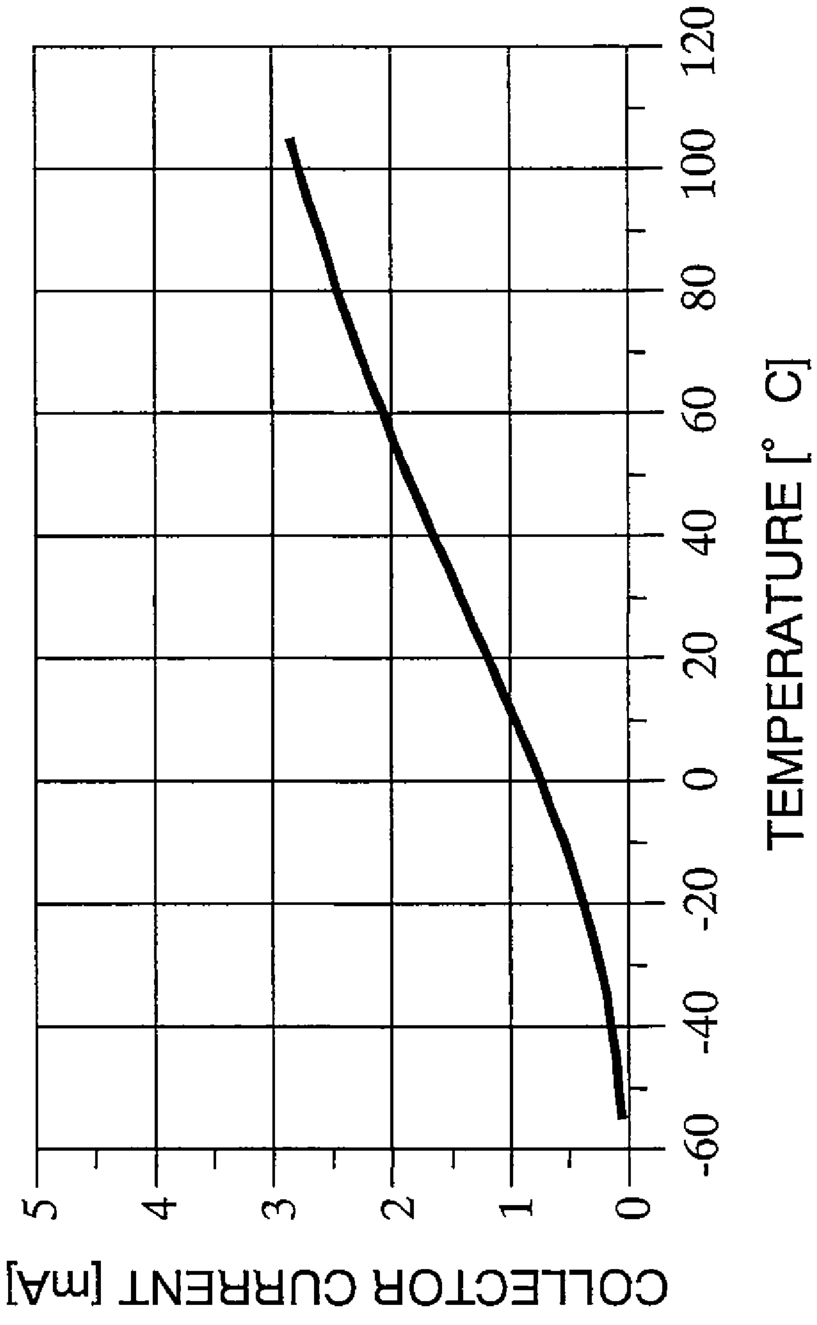
COLLECTOR CURRENT [mA]
0
1
2
3
4
5
TEMPERATURE [° C]
-60
-40
-20
0
20
40
60
80
100
120

FIG.4A

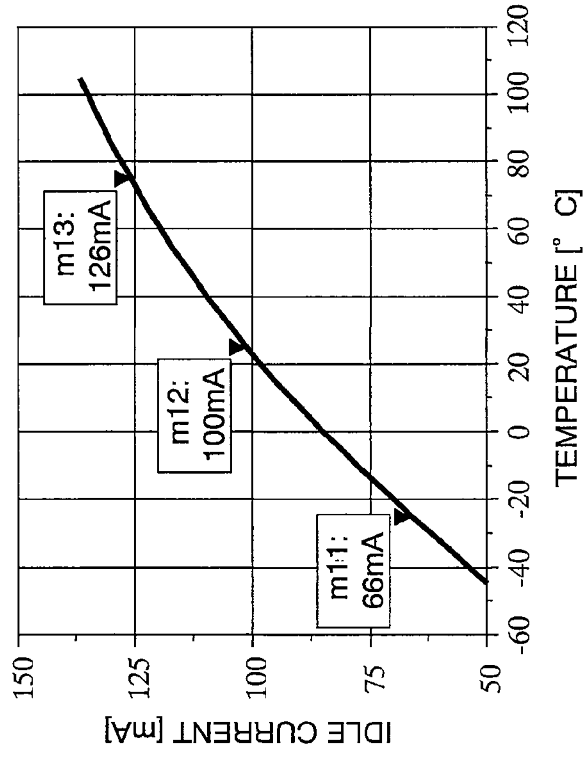
m11: 66mA
m12: 100mA
m13: 126mA
IDLE CURRENT [mA]
150
125
100
75
50
TEMPERATURE [° C]
-60
-40
-20
0
20
40
60
80
100
120

FIG.4B

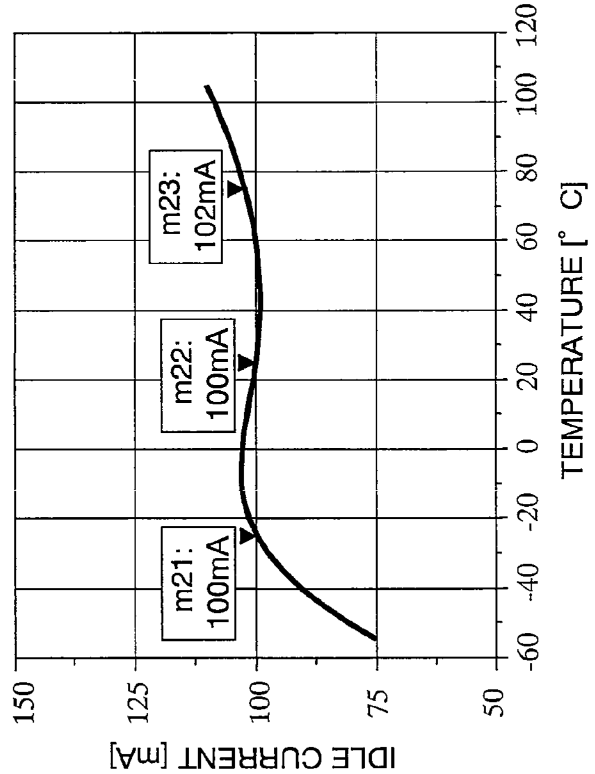
m21: 100mA
m22: 100mA
m23: 102mA
IDLE CURRENT [mA]
150
125
100
75
50
TEMPERATURE [° C]
-60
-40
-20
0
20
40
60
80
100
120

FIG.5

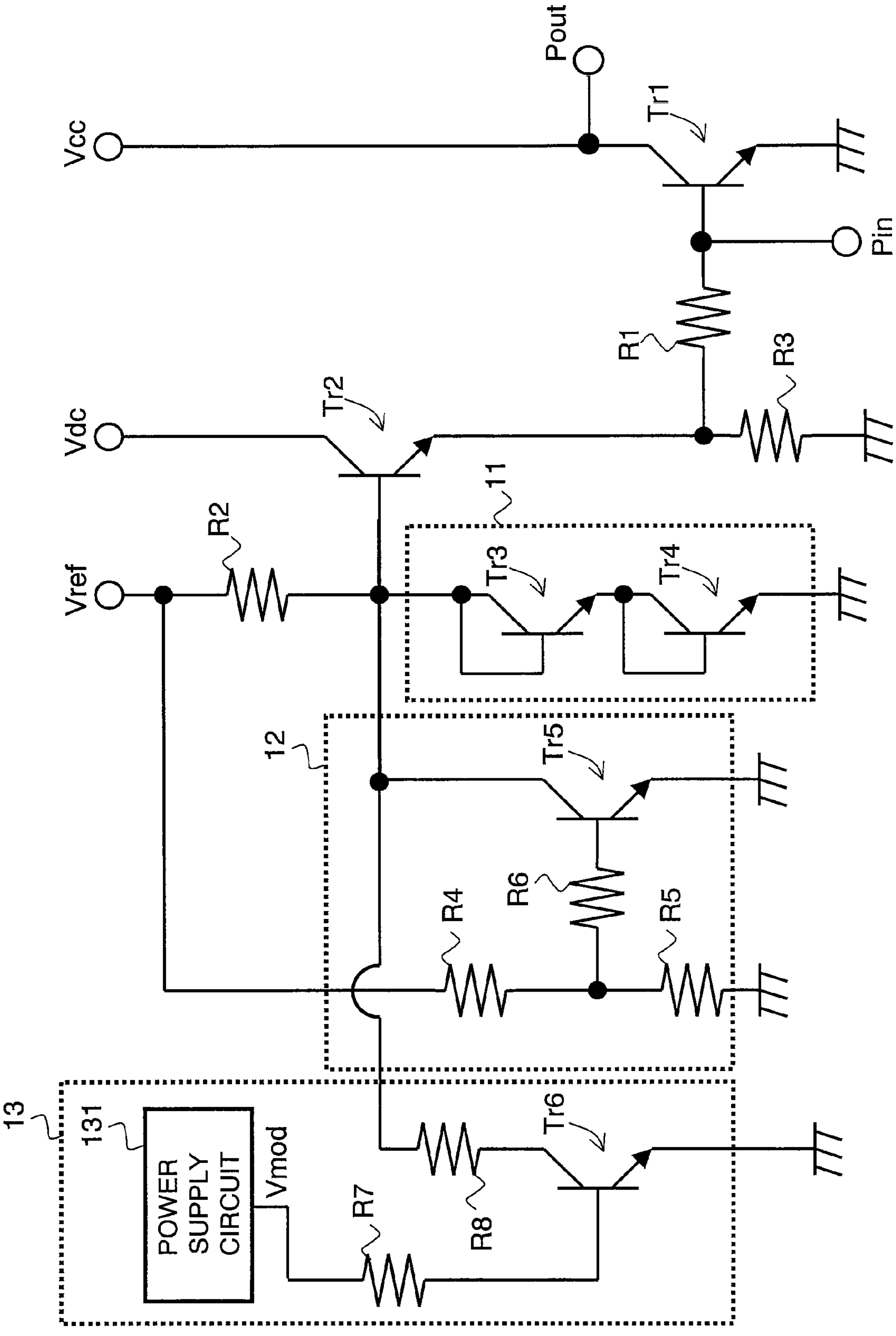
Vcc
Pout
Tr1
Pin
R1
R3
Vdc
Tr2
Vref
R2
11
Tr3
Tr4
12
Tr5
R4
R6
R5
13
131
POWER SUPPLY CIRCUIT
Vmod
R7
R8
Tr6

FIG.6

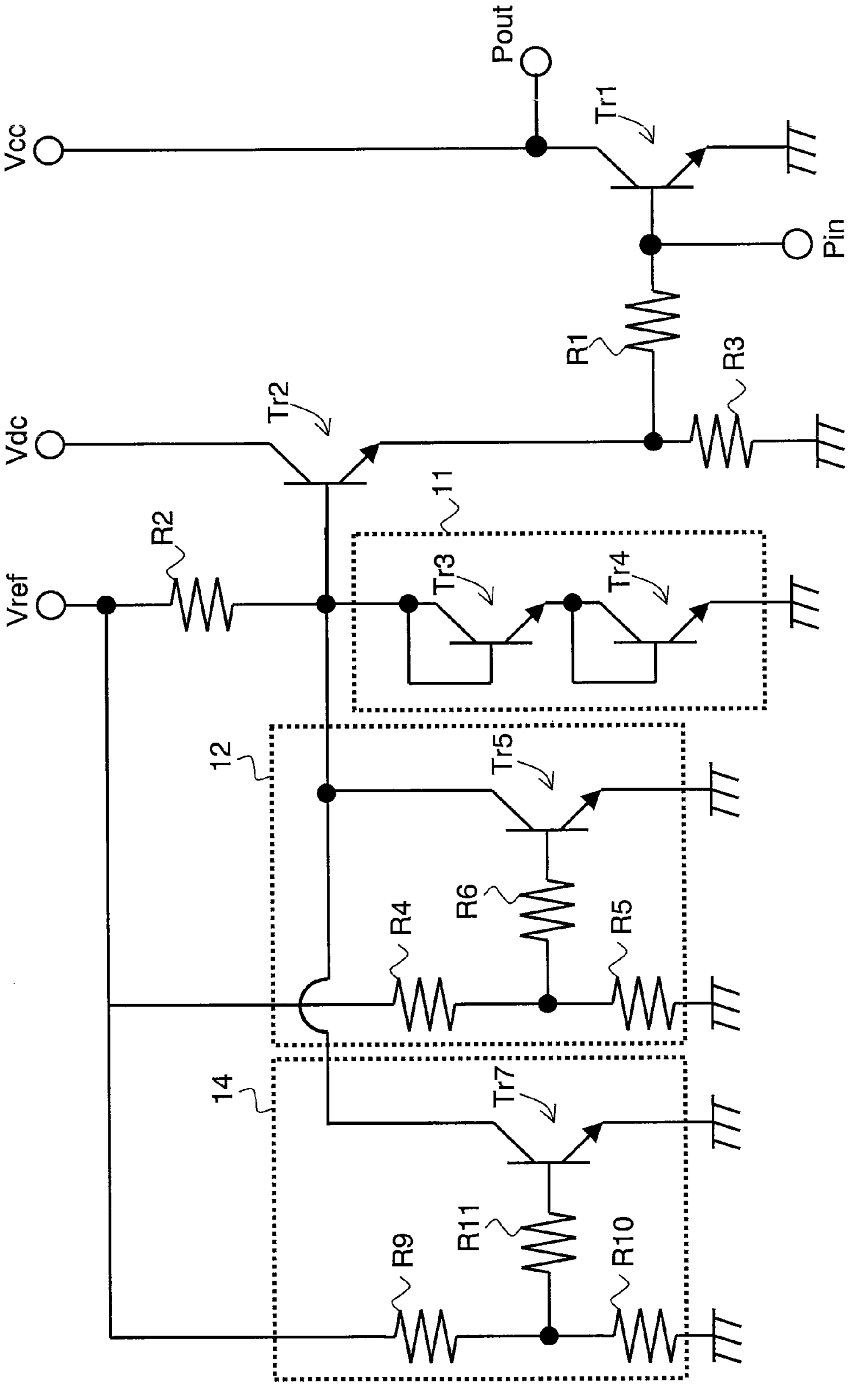
Vcc
Pout
Tr1
Pin
Vdc
Tr2
R1
R3
Vref
R2
11
Tr3
Tr4
12
Tr5
R4
R6
R5
14
Tr7
R9
R11
R10

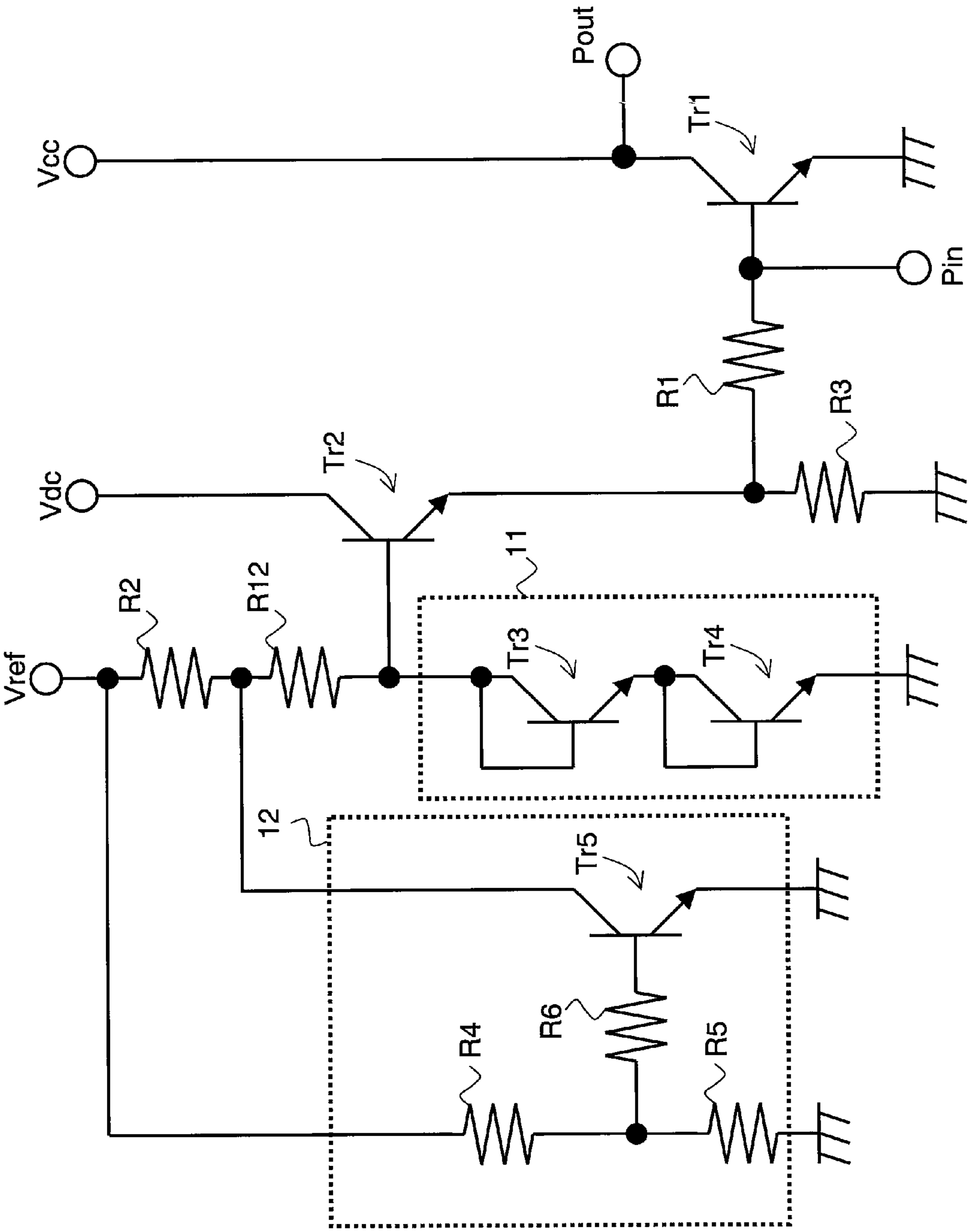
Vcc
Pout
Tr1
Pin
R1
R3
Vdc
Tr2
11
Vref
R2
R12
Tr3
Tr4
12
Tr5
R4
R6
R5

FIG.7

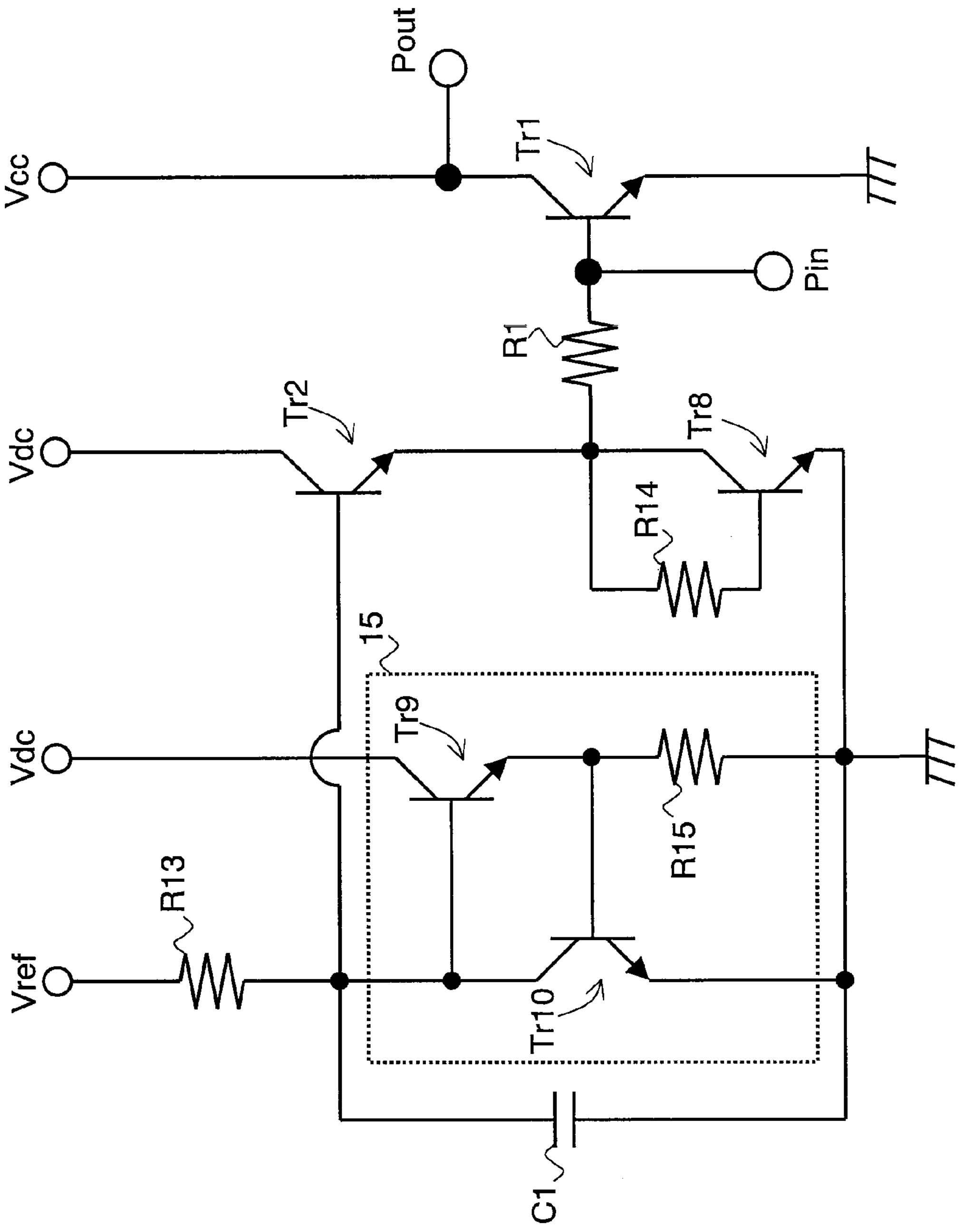
Vcc
Pout
Tr1
Pin
Vdc
Tr2
R1
Tr8
R14
15
Vdc
Tr9
R15
Vref
R13
Tr10
C1

FIG.8 PRIOR ART

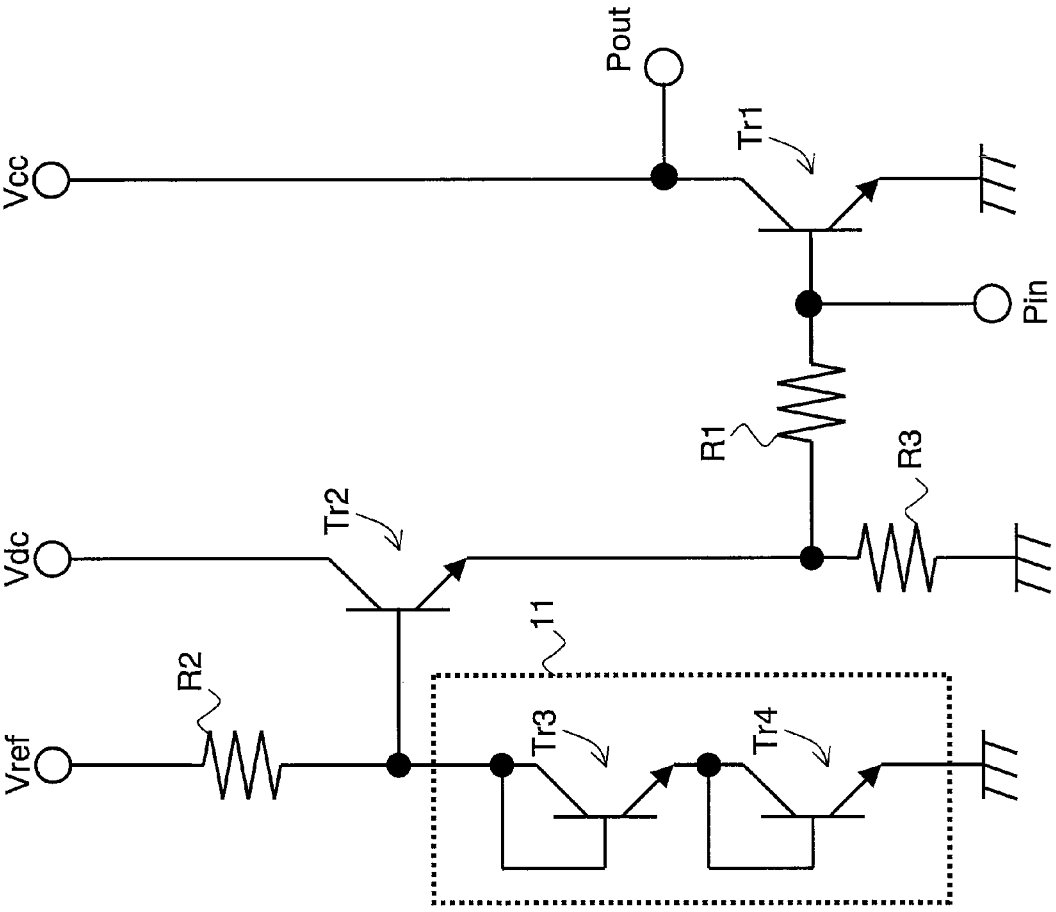
Vcc
Pout
Tr1
Pin
R1
R3
Vdc
Tr2
11
R2
Vref
Tr3
Tr4

FIG.9 PRIOR ART

RADIO-FREQUENCY POWER AMPLIFIER

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to a radio-frequency power amplifier, and particularly to a radio-frequency power amplifier of which a bias circuit for supplying a bias current to the amplifying transistor includes a plurality of temperature compensation circuits.

2. Description of the Background Art

In recent years, due to the installation of various application functions in a mobile communication device such as a mobile phone, the number of components of the mobile phone tends to increase in accordance with the number of necessary components to control these functions. On the other hand, the mobile phone is miniaturized, and accordingly, it is necessary to miniaturize the components so as to increase the degree of integration of each component. However, some of the components may produce heat, and the higher the degree of integration of each component is made by miniaturizing the components, the more difficult it becomes to maintain a constant temperature in the mobile phone. Therefore, the components of the mobile phone are required to be stable with respect to temperature dependence.

In these circumstances, as a radio-frequency power amplifier, one of the components of the mobile phone, which amplifies the power of a radio-frequency transmitted signal, an HEMT (High Electron Mobility Transistor), which is stable toward temperature dependence, is used. However, this requires separately a negative voltage generator so as to control the gate voltage, and therefore hinders the miniaturization and the cost reduction of the mobile phone. As a result, conventionally, an HBT (Heterojunction Bipolar Transistor), which is used with a single power supply, is mostly used as a radio-frequency power amplifier. The bias circuit of a radio-frequency power amplifier using an HBT includes a temperature compensation circuit for compensating for the temperature characteristics of the diffusion potential between the base and emitter.

FIG. **8** is a diagram showing the circuit configuration of a radio-frequency power amplifier using an HBT, disclosed in Patent Document 1. The radio-frequency power amplifier of FIG. **8** includes an amplifying transistor Tr**1** and a bias circuit.

The amplifying transistor Tr**1** power-amplifies a radio-frequency signal inputted from an input terminal Pin and outputs the power-amplified radio-frequency signal from an output terminal Pout. The collector terminal of the amplifying transistor Tr**1** is connected to a power supply Vcc, and the emitter terminal is connected to ground potential.

The bias circuit, which supplies a bias current to the base terminal of the amplifying transistor Tr**1**, includes an emitter-follower transistor Tr**2**, a transistor Tr**8**, a resistance element R**1**, a resistance element R**13**, a resistance element R**14**, a capacitor C**1**, and a temperature compensation circuit **15**. The collector terminal of the emitter-follower transistor Tr**2**, which is an emitter follower, is connected to a power supply Vdc, and the base terminal is connected to the temperature compensation circuit **15**, surrounded by a dashed line, to one end of the resistance element R**13**, and to one end of the capacitor C**1**. The other end of the resistance element R**13** is connected to a reference power supply Vref, and the other end of the capacitor C**1** is connected to ground potential. The emitter terminal of the emitter-follower transistor Tr**2** is connected to the base terminal of the amplifying transistor Tr**1** via the resistance element R**1** and also connected to the collector terminal of the transistor Tr**8**. The collector terminal and base terminal of the transistor Tr**8** are connected to each other via the resistance element R**14**. The emitter terminal of the transistor Tr**8** is connected to ground potential.

The temperature compensation circuit **15** includes a transistor Tr**9**, a transistor Tr**10**, and a resistance element R**15**. The collector terminal of the transistor Tr**9** is connected to a power supply Vdc, and the base terminal is connected to the base terminal of the emitter-follower transistor Tr**2** and the collector terminal of the transistor Tr**10**. The emitter terminal of the transistor Tr**9** is connected to the base terminal of the transistor Tr**10** and connected to ground potential via the resistance element R**15**. The emitter terminal of the transistor Tr**10** is connected to ground potential.

The operating principle of the temperature compensation circuit **15** will be described. At low temperature, the diffusion potential between the base and emitter of the transistor Tr**9** rises. When the diffusion potential rises, the current (hereinafter referred to as "collector current") between the collector and emitter of the transistor Tr**9** decreases, and therefore the potential of the terminal, connected to the emitter terminal of the transistor Tr**9**, of the resistance element R**15** falls and the base potential of the transistor Tr**10** also falls. When the base potential of the transistor Tr**10** falls, the collector current of the transistor Tr**10** decreases, and therefore the current flowing through the resistance element R**13** decreases. When the current flowing through the resistance element R**13** decreases, the potential difference, i.e., the voltage, caused in the resistance element R**13** decreases, and therefore the base potential of the emitter-follower transistor Tr**2** rises. When the base potential rises, the collector current of the emitter-follower transistor Tr**2** increases, and therefore the base potential of the amplifying transistor Tr**1** rises. This compensates for the idle current of the amplifying transistor Tr**1** so that the value of the idle current increases. In contrast, at high temperature, the reverse operation to that performed at low temperature is performed. In this case, the base potential of the emitter-follower transistor Tr**2** falls, and therefore the idle current of the amplifying transistor Tr**1** is compensated for so that the value of the idle current decreases.

Here, there may be a case where a portion of the power of the transmitted signal from the amplifying transistor Tr**1** leaks to the bias circuit via the emitter-follower transistor Tr**2**. As a result, the base potential of the emitter-follower transistor Tr**2** may become so unstable that it is impossible to supply a stable bias current.

In the configuration of FIG. **8**, to solve this problem, the capacitor C**1** is connected to the base terminal of the emitter-follower transistor Tr**2**. However, in such a system as GSM (Global System for Mobile Communications) that uses a large transmitted signal, it is necessary to use the capacitor C**1** that has a considerably large capacitance, which significantly increases the size of the capacitor C**1**. On the other hand, in view of the miniaturization of the mobile phone, there are limitations to increasing the size of the capacitor C**1**. Therefore, with the configuration of FIG. **8**, it is difficult to supply a stable bias current.

In response, a radio-frequency power amplifier disclosed in Patent Document 2 is proposed. FIG. **9** is a diagram showing the circuit configuration of the radio-frequency power amplifier disclosed in Patent Document 2. The radio-frequency power amplifier of FIG. **9** includes an amplifying transistor Tr**1** and a bias circuit. Note that, in FIG. **9**, the same components as those of FIG. **8** will be denoted by the same numerals and will not be described.

The bias circuit, which supplies a bias current to the base terminal of the amplifying transistor Tr**1**, includes an emitter-follower transistor Tr**2**, resistance elements R**1**, R**2** and R**3**, and a temperature compensation circuit **11**. The base terminal of the emitter-follower transistor Tr**2** is connected to the temperature compensation circuit **11**, surrounded by a dashed line, and to one end of the resistance element R**2**. The other end of the resistance element R**2** is connected to a reference power supply Vref. The emitter terminal of the emitter-follower transistor Tr**2** is connected to the base terminal of the amplifying transistor Tr**1** via the resistance element R**1** and also connected to one end of the resistance element R**3**. The other end of the resistance element R**3** is connected to ground potential.

The temperature compensation circuit **11** includes a diode transistor Tr**3**, the base terminal and collector terminal of which are short-circuited and a diode transistor Tr**4**, the base terminal and collector terminal of which are short-circuited. The diode transistors Tr**3** and Tr**4** are connected to each other as two stages in cascade. The emitter terminal of the diode transistor Tr**4**, which is the lower stage, is connected to ground potential, and the collector terminal of the diode transistor Tr**3**, which is the upper stage, is connected to the base terminal of the emitter-follower transistor Tr**2** and also connected to the reference power supply Vref via the resistance element R**2**.

The operating principle of the temperature compensation circuit **11** will be described. The diffusion potential between the base of the diode transistor Tr**3** and the emitter of the diode transistor Tr**4** changes with temperature change, whereby the collector potential of the diode transistor Tr**3** (i.e., the base potential of the emitter-follower transistor Tr**2**) also changes with temperature change. Specifically, at low temperature, the diffusion potential between the base and emitter of the diode transistor Tr**3** rises. When the diffusion potential rises, the collector current of the diode transistor Tr**3** decreases, and therefore the current flowing through the resistance element R**2** decreases. When the current flowing through the resistance element R**2** decreases, the voltage caused in the resistance element R**2** decreases, and therefore the base potential of the emitter-follower transistor Tr**2** rises. When the base potential rises, the collector current of the emitter-follower transistor Tr**2** increases, and therefore the base potential of the amplifying transistor Tr**1** rises. This compensates for the idle current of the amplifying transistor Tr**1** so that the value of the idle current increases. In contrast, at high temperature, the reverse operation to that performed at low temperature is performed. In this case, the base potential of the emitter-follower transistor Tr**2** falls, and therefore the idle current of the amplifying transistor Tr**1** is compensated for so that the value of the idle current decreases.

The use of the temperature compensation circuit **11** described above suppresses the leakage power of the transmitted signal from the amplifying transistor Tr**1**, due to the parasitic capacitance between the base of the diode transistor Tr**3** and the emitter of the diode transistor Tr**4**. Therefore, with the configuration of FIG. **9**, it is possible, unlike the configuration of FIG. **8**, to smooth the base potential of the emitter-follower transistor Tr**2** without using a large-size capacitor C**1** besides the temperature compensation circuit, and therefore to supply a stable bias current.

Patent Document 1: Japanese Laid-Open Patent Publication No. 2005-101733

Patent Document 2: Japanese Laid-Open Patent Publication No. 2001-274636

However, with the radio-frequency power amplifier of FIG. **9**, it is impossible to sufficiently suppress the temperature dependence of the idle current of the amplifying transistor Tr**1**, due to the circuit configuration of the temperature compensation circuit **11**, and therefore it is impossible to sufficiently compensate for temperature.

To sufficiently compensate for temperature in the radio-frequency power amplifier of FIG. **9**, it is necessary to make the voltage value of the reference power supply Vref far greater than twice the diffusion potential between the base of the transistor Tr**3** and the emitter of the transistor Tr**4**. Additionally, to prevent the idle current of the amplifying transistor Tr**1** from significantly increasing, it is also necessary to increase the resistance value of the resistance element R**2**. However, when the resistance value of the resistance element R**2** is increased, the voltage caused in the resistance element R**2** increases when the operating current of the amplifying transistor Tr**1** increases due to a large signal. As a result, the power gain falls, which hinders high efficiency. Consequently, with the radio-frequency power amplifier of FIG. **9**, there are limitations to increasing the resistance value of the resistance element R**2**, and therefore it is difficult to sufficiently compensate for temperature.

SUMMARY OF THE INVENTION

Therefore, to solve the above problem, an object of the present invention is to provide a radio-frequency power amplifier having a bias circuit capable of making the temperature dependence, of the idle current of an amplifying transistor, lower than is conventionally the case and thus sufficiently compensating for temperature.

The present invention is directed to solving the above problem. A radio-frequency power amplifier according to the present invention includes: a first transistor operable to amplify a power of a radio-frequency signal; and a bias circuit operable to supply a bias current to the first transistor. The bias circuit includes: a second transistor that has a collector terminal connected to a first power supply; a first resistance element that has one end connected to an emitter terminal of the second transistor and has the other end connected to a base terminal of the first transistor; a second resistance element that has one end connected to the emitter terminal of the second transistor and has the other end connected to ground potential; at least one third resistance element provided between a base terminal of the second transistor and a second power supply; and a plurality of temperature compensation circuits connected to the base terminal of the second transistor which are operable to control a base potential of the second transistor so that the potential falls as a temperature rises.

Note that, in an embodiment described below, for example, the first transistor corresponds to an amplifying transistor Tr**1**, the first power supply corresponds to a power supply Vdc, the second transistor corresponds to an emitter-follower transistor Tr**2**, the first resistance element corresponds to a resistance element R**1**, the second resistance element corresponds to a resistance element R**3**, the second power supply corresponds to a reference power supply Vref, and the third resistance element corresponds to a resistance element R**2**.

Based on the configuration described above, the bias circuit includes the plurality of temperature compensation circuits. This makes it possible to make the temperature dependence, of the idle current of the first transistor, lower than is conventionally the case. As a result, based on the present embodiment, it is possible to provide a radio-frequency power amplifier having a bias circuit capable of sufficiently compensating for temperature.

It is preferable that the bias circuit may include a first temperature compensation circuit and a second temperature compensation circuit, that the first temperature compensation circuit may include: a third transistor that has a collector terminal and a base terminal that are short-circuited, the collector terminal being connected to the base terminal of the second transistor; and a fourth transistor that has a collector terminal and a base terminal that are short-circuited, the collector terminal being connected to an emitter terminal of the third transistor, and has an emitter terminal connected to ground potential, and that the second temperature compensation circuit may include: a first voltage divider circuit having a plurality of series resistance elements connected in series; a fifth transistor that has a collector terminal connected to the base terminal of the second transistor and an emitter terminal connected to ground potential; and a fourth resistance element that has one end connected to a connecting point of any of the series resistance elements of the first voltage divider circuit and has the other end connected to a base terminal of the fifth transistor.

Note that, in an embodiment described below, for example, the first temperature compensation circuit corresponds to a temperature compensation circuit **11**, the second temperature compensation circuit corresponds to a temperature compensation circuit **12**, a third transistor corresponds to a diode transistor Tr**3**, the fourth transistor corresponds to a diode transistor Tr**4**, the series resistance elements correspond to a resistance element R**4** and a resistance element R**5**, the fifth transistor corresponds to a transistor Tr**5**, and the fourth resistance element corresponds to a resistance element R**6**.

In this case, the bias circuit may further include at least one sixth resistance element provided between the base terminal of the second transistor and a connecting point of the third resistance element and the collector terminal of the fifth transistor. Note that, in the embodiments described below, for example, the sixth resistance element corresponds to a resistance element R**12**.

The bias circuit may further include a third temperature compensation circuit, and the third temperature compensation circuit may include: a second voltage divider circuit having a plurality of series resistance elements connected in series; a sixth transistor that has a collector terminal connected to the base terminal of the second transistor and has an emitter terminal connected to ground potential; and a fifth resistance element that has one end connected to a connecting point of any of the series resistance elements of the second voltage divider circuit and has the other end connected to a base terminal of the sixth transistor, and resistance values of the series resistance elements of the first voltage divider circuit, the series resistance elements of the second voltage divider circuit, the fourth resistance element, and the fifth resistance element may be each set so that an idle current of the first transistor decreases as a temperature rises at temperatures above a predetermined temperature.

Note that, in an embodiment described below, the third temperature compensation circuit corresponds to a temperature compensation circuit **14**, the series resistance element correspond to a resistance element R**9** and a resistance element R**10**, the sixth transistor corresponds to a transistor Tr**7**, and the fifth resistance element corresponds to a resistance element R**11**.

It is preferable that all of the series resistance elements may be arranged in the same direction on a substrate.

It is preferable that a mode switching circuit, connected to the base terminal of the second transistor, operable to switch a bias current to be supplied to the second transistor to a predetermined value of the bias current may be further included. In this case, the mode switching circuit may include: a power supply circuit operable to control an output of a predetermined voltage to start or stop; a seventh transistor that has an emitter terminal connected to ground potential; a seventh resistance element that has one end connected to an output of the power supply circuit and has the other end connected to a base terminal of the seventh transistor; and an eighth resistance element that has one end connected to a collector terminal of the seventh transistor and has the other end connected to the base terminal of the second transistor.

Note that, in an embodiment described below, for example, the seventh transistor corresponds to a transistor Tr**6**, the seventh resistance element corresponds to a resistance element R**7**, and the eighth resistance element corresponds to a resistance element R**8**.

Based on the present invention, it is possible to provide a radio-frequency power amplifier having a bias circuit capable of making the temperature dependence, of the idle current of an amplifying transistor, lower than is conventionally the case and thus sufficiently compensating for temperature.

These and other objects, features, aspects and advantages of the present invention will become more apparent from the following detailed description of the present invention when taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a diagram showing the circuit configuration of a radio-frequency power amplifier according to a first embodiment;

FIG. **2** is a diagram showing an example of specific circuit configurations of a temperature compensation circuit **11** and a temperature compensation circuit **12**;

FIG. **3** is a diagram showing the temperature characteristics of the collector current of a transistor Tr**5** of the temperature compensation circuit **12**;

FIG. **4**A is a diagram showing the temperature characteristics of the idle current of an amplifying transistor Tr**1** of a conventional radio-frequency power amplifier, shown in FIG. **9**;

FIG. **4**B is a diagram showing the temperature characteristics of the idle current of an amplifying transistor Tr**1** of the radio-frequency power amplifier of the present invention, shown in FIG. **2**;

FIG. **5** is a diagram showing the circuit configuration of a radio-frequency power amplifier according to a second embodiment;

FIG. **6** is a diagram showing the circuit configuration of a radio-frequency power amplifier according to a third embodiment;

FIG. **7** is a diagram showing the circuit configuration of a radio-frequency power amplifier according to a fourth embodiment;

FIG. **8** is a diagram showing the circuit configuration of a radio-frequency power amplifier disclosed in Patent Document 1; and

FIG. **9** is a diagram showing the circuit configuration of a radio-frequency power amplifier disclosed in Patent Document 2.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

With reference to the drawings, embodiments of a radio-frequency power amplifier of the present invention will be described below.

First Embodiment

With reference to FIG. **1**, the circuit configuration of a radio-frequency power amplifier according to a first embodiment of the present invention will be described. FIG. **1** is a diagram showing the circuit configuration of the radio-frequency power amplifier according to the first embodiment. The radio-frequency power amplifier of FIG. **1** includes an amplifying transistor Tr**1** and a bias circuit. Note that, in FIG. **1**, the same components as those of FIG. **9** will be denoted by the same numerals and therefore will not be described in detail.

The bias circuit, which supplies a bias current to the base terminal of the amplifying transistor Tr**1**, includes an emitter-follower transistor Tr**2**, resistance elements R**1**, R**2** and R**3**, a temperature compensation circuit **11**, and a temperature compensation circuit **12**.

The temperature compensation circuit **11**, which is connected to the base terminal of the emitter-follower transistor Tr**2**, compensates for the idle current of the amplifying transistor Tr**1** in response to a temperature change by controlling the base potential of the emitter-follower transistor Tr**2** so that the potential falls as the temperature rises.

The temperature compensation circuit **12**, which is connected to the base terminal of the emitter-follower transistor Tr**2** and also connected to a reference power supply Vref, compensates for the idle current of the amplifying transistor Tr**1** in response to a temperature change by controlling the base potential of the emitter-follower transistor Tr**2** so that the potential falls as the temperature rises.

Next, specific circuit configurations of the temperature compensation circuits **11** and **12** will be described. FIG. **2** is a diagram showing an example of the specific circuit configurations of the temperature compensation circuits **11** and **12**. The temperature compensation circuit **11**, which is similar in configuration to the temperature compensation circuit **11** of FIG. **9**, includes a diode transistor Tr**3**, the base terminal and collector terminal of which are short-circuited and a diode transistor Tr**4**, the base terminal and collector terminal of which are short-circuited. The diode transistors Tr**3** and Tr**4** are connected to each other as two stages in cascade. The emitter terminal of the diode transistor Tr**4**, which is the lower stage, is connected to ground potential. The collector terminal of the diode transistor Tr**3**, which is the upper stage, is connected to the base terminal of the emitter-follower transistor Tr**2** and also connected to the reference power supply Vref via the resistance element R**2**.

The temperature compensation circuit **12** includes a transistor Tr**5** and resistance elements R**4**, R**5** and R**6**. The resistance elements R**4** and R**5** are connected to each other as two stages in cascade. The terminal, of the resistance element R**4**, not connected to the resistance element R**5** is connected to the reference power supply Vref, and the terminal, of the resistance element R**5**, not connected to the resistance element R**4** is connected to ground potential. Thus, the resistance elements R**4** and R**5** form a voltage divider circuit. The connecting point of the resistance elements R**4** and R**5** is connected to the base terminal of the transistor Tr**5** via the resistance element R**6**. The emitter terminal of the transistor Tr**5** is connected to ground potential. The collector terminal of the transistor Tr**5** is connected to the base terminal of the emitter-follower transistor Tr**2** and also connected to the reference power supply Vref via the resistance element R**2**.

The operating principle of the temperature compensation circuit **11** is similar to that of the temperature compensation circuit **11** of FIG. **9**, and therefore will not be described. The operating principle of the temperature compensation circuit **12** will be described below. FIG. **3** is a diagram showing the temperature characteristics of the collector current of the transistor Tr**5** of the temperature compensation circuit **12**. The resistance values of the resistance elements R**4** and R**5** are set so that the collector current flows through the transistor Tr**5** even at the lowest temperature of the temperature conditions under which the radio-frequency power amplifier is used. The resistance value of the resistance element R**6** is set so that the slope of the temperature characteristics of the collector current of the transistor Tr**5** is an arbitrary slope. These settings of the resistance elements R**4**, R**5**, and R**6** increase the collector current of the transistor Tr**5** as the temperature rises from low to high, as shown in FIG. **3**. Here, the collector terminal of the transistor Tr**5** is connected to the base terminal of the emitter-follower transistor Tr**2** and also connected to the reference power supply Vref via the resistance element R**2**. Since the collector current of the transistor Tr**5** increases at high temperature as shown in FIG. **3**, the voltage caused in the resistance element R**2** increases further, and therefore the base potential of the emitter-follower transistor Tr**2** falls further. At high temperature, this compensates for the idle current of the amplifying transistor Tr**1** so that the value of the idle current decreases.

FIG. **4**A is a diagram showing the temperature characteristics of the idle current of the amplifying transistor Tr**1** of the conventional radio-frequency power amplifier, shown in FIG. **9**. The temperature characteristics of FIG. **4**A indicate the characteristics when the value of each component of the bias circuit is set so that the idle current of the amplifying transistor Tr**1** is 100 mA at an environmental temperature of 25° C., and also, the voltage value of the reference power supply Vref is set to 2.8 V, which is a control voltage used for a general mobile phone. Referring to FIG. **4**A, m**11** represents the idle current of the amplifying transistor Tr**1** at an environmental temperature of −25° C., m**12** represents the idle current of the amplifying transistor Tr**1** at an environmental temperature of 25° C., and m**13** represents the idle current of the amplifying transistor Tr**1** at an environmental temperature of 75° C. As shown in FIG. **4**A, since m**11** is 66 mA and m**13** is 126 mA, it is understood that the change of the current is approximately twofold at a temperature difference of 100° C. The reason is that the current consumed in the temperature compensation circuit **11** of FIG. **9** does not sufficiently increase as the temperature rises from low to high when the voltage value of the reference power supply Vref is set to 2.8 V.

FIG. **4**B is a diagram showing the temperature characteristics of the idle current of the amplifying transistor Tr**1** of the radio-frequency power amplifier of the present invention, shown in FIG. **2**. Similarly to FIG. **4**A, the temperature characteristics of FIG. **4**B indicate the characteristics when the value of each component of the bias circuit is set so that the idle current of the amplifying transistor Tr**1** is 100 mA at an environmental temperature of 25° C., and also, the voltage value of the reference power supply Vref is set to 2.8 V, which is a control voltage used for a general mobile phone. Referring to FIG. **4**B, m**21** represents the idle current of the amplifying transistor Tr**1** at an environmental temperature of −25° C., m**22** represents the idle current of the amplifying transistor Tr**1** at an environmental temperature of 25° C., and m**23** represents the idle current of the amplifying transistor Tr**1** at an environmental temperature of 75° C. As shown in FIG. **4**B, since m**21** is 100 mA and m**23** is 102 mA, it is understood that the change of the current at a temperature difference of 100° C. is suppressed to approximately 2 mA.

As described above, in the present embodiment, the bias circuit includes two temperature compensation circuits (the temperature compensation circuits **11** and **12**). This makes it possible to make the temperature dependence, of the idle current of the amplifying transistor Tr**1**, lower than is conventionally the case. As a result, based on the present embodiment, it is possible to provide a radio-frequency power amplifier having a bias circuit capable of sufficiently compensating for temperature.

Further, the present embodiment includes the temperature compensation circuit **11**, which is similar to that of FIG. **9**. This makes it possible, unlike the configuration of FIG. **8**, to smooth the base potential of the emitter-follower transistor Tr**2** without using a large-size capacitor C**1** besides the temperature compensation circuit, and therefore to supply a stable bias current.

Furthermore, in the conventional temperature compensation circuit **15** of FIG. **8**, since the transistors Tr**9** and Tr**10** form a current loop, noise may deteriorate and an oscillation may occur in the bias circuit. In the present embodiment, however, as shown in FIG. **2**, the temperature compensation circuits **11** and **12** do not form a current loop. Thus, in the present embodiment, noise does not deteriorate nor does an oscillation occur in the bias circuit due to a current loop.

Note that the temperature compensation circuits **11** and **12** may not necessarily have the circuit configurations shown in FIG. **2**, and may have other circuit configurations.

Note that the formation directions, on a substrate, of the resistance elements R**4** and R**5** of the temperature compensation circuit **12** are not particularly described above, but it is preferable that the formation directions of the resistance elements R**4** and R**5** be the same. A formation direction refers to the direction from one terminal of each resistance to the other. The same formation direction makes it possible to suppress production variations of the resistance values of the resistance elements R**4** and R**5**, and therefore to obtain a desired potential at the connecting point of the resistance elements R**4** and R**5** with high accuracy. Note that, in the temperature compensation circuit **12**, the voltage divider circuit is formed using two resistance elements (the resistance elements R**4** and R**5**), but the voltage divider circuit may be formed by connecting three or more resistance elements in series.

Second Embodiment

With reference to FIG. **5**, the circuit configuration of a radio-frequency power amplifier according to a second embodiment of the present invention will be described. FIG. **5** is a diagram showing the circuit configuration of the radio-frequency power amplifier according to the second embodiment. The radio-frequency power amplifier of FIG. **5** is different from the radio-frequency power amplifier of FIG. **2** in that the radio-frequency power amplifier of FIG. **5** further includes a mode switching circuit **13**. The other components of the circuit configuration of FIG. **5** are the same as those of the circuit configuration of FIG. **2**, and therefore will be denoted by the same numerals and will not be described.

The mode switching circuit **13** includes a power supply circuit **131**, a resistance element R**7**, a resistance element R**8**, and a transistor Tr**6**. The base terminal of the transistor Tr**6** is connected to the output of the power supply circuit **131** via the resistance element R**7**. The collector terminal of the transistor Tr**6** is connected to the base terminal of the emitter-follower transistor Tr**2** via the resistance element R**8** and also connected to the reference power supply Vref via the resistance elements R**8** and R**2**. The emitter terminal of the transistor Tr**6** is connected to ground potential. In accordance with a rule set in advance, the power supply circuit **131** controls the output of a mode switching voltage Vmod to start (i.e., the output is turned ON) or stop (i.e., the output is turned OFF).

The operating principle of the mode switching circuit **13** will be described. The output of the mode switching voltage Vmod is turned ON or OFF in accordance with a rule set in advance in the power supply circuit **131**, whereby the transistor Tr**6** operates as a switch (SW). When the output of the mode switching voltage Vmod is turned ON, the transistor Tr**6** is turned ON, and therefore the collector current of the transistor Tr**6** flows. When the collector current of the transistor Tr**6** flows, the base potential of the emitter-follower transistor Tr**2** falls, and therefore the collector current of the emitter-follower transistor Tr**2** decreases. When the collector current decreases, the base potential of the amplifying transistor Tr**1** falls, and therefore the idle current of the amplifying transistor Tr**1** is suppressed so that the value of the idle current decreases.

As described above, in the present embodiment, the bias circuit includes the mode switching circuit **13**. Here, it is assumed, for example, that the rule is set in the power supply circuit **131** so that, in the case where the dynamic range of the output power of the radio-frequency power amplifier is wide, the output of the mode switching voltage Vmod is turned ON at the time of a low output. In this case, it is possible to suppress the idle current of the amplifying transistor Tr**1** at the time of a low output, and therefore to improve the power added efficiency of the radio-frequency power amplifier.

Note that the mode switching circuit **13** described above may not necessarily have the circuit configuration shown in FIG. **5**, and may have another circuit configuration.

Third Embodiment

With reference to FIG. **6**, the circuit configuration of a radio-frequency power amplifier according to a third embodiment of the present invention will be described. FIG. **6** is a diagram showing the circuit configuration of the radio-frequency power amplifier according to the third embodiment. The radio-frequency power amplifier of FIG. **6** is different from the radio-frequency power amplifier of FIG. **2** in that the radio-frequency power amplifier of FIG. **6** further includes a temperature compensation circuit **14**. The other components of the circuit configuration of FIG. **6** are the same as those of the circuit configuration of FIG. **2**, and therefore will be denoted by the same numerals and will not be described.

The temperature compensation circuit **14**, which is similar in configuration to the temperature compensation circuit **12**, is connected to the base terminal of the emitter-follower transistor Tr**2** and thus the temperature compensation circuit **14**, as well as the temperature compensation circuits **11** and **12**, further compensates for temperature. More specifically, the temperature compensation circuit **14** includes a transistor Tr**7** and resistance elements R**9**, R**10** and R**11**. The resistance elements R**9** and R**10** are connected to each other as two stages in cascade. The terminal, of the resistance element R**9**, not connected to the resistance element R**10** is connected to the reference power supply Vref, and the terminal, of the resistance element R**10**, not connected to the resistance element R**9** is connected to ground potential. Thus, the resistance elements R**9** and R**10** form a voltage divider circuit. The connecting point of the resistance elements R**9** and R**10** is connected to the base terminal of the transistor Tr**7** via the resistance element R**11**. The emitter terminal of the transistor Tr**7** is connected to ground potential. The collector terminal of the transistor Tr**7** is connected to the base terminal of the emitter-follower transistor Tr**2** and also connected to the reference power supply Vref via the resistance element R**2**. Note that the operating principle of the temperature compensation circuit **14** is similar to that of the temperature compensation circuit **12**, and therefore will not be described.

Here, the resistance values of the resistance elements R**4**, R**5**, and R**6** of the temperature compensation circuit **12** and the resistance values of the resistance elements R**9**, R**10**, and R**11** of the temperature compensation circuit **14** are set to different values from each other, whereby the temperature characteristics of the idle current of the amplifying transistor Tr**1** can have an inflection point. Consequently, the temperature characteristics of the idle current of the amplifying transistor Tr**1** may have such an inflection point that, for example, the idle current decreases only as the temperature rises at temperatures above a desired temperature. Due to this inflection point, the idle current of the amplifying transistor Tr**1** decreases at high temperature even when the idle current is high at low temperature, and therefore it is possible to suppress thermal runaway of the idle current of the amplifying transistor Tr**1** at high temperature.

Note that the circuit configuration of the temperature compensation circuit **14** may not necessarily be the same as that of the temperature compensation circuit **12** as described above, and the temperature compensation circuit **14** may have another circuit configuration. Additionally, in the present embodiment, the mode switching circuit **13** of the second embodiment may be further provided.

Fourth Embodiment

With reference to FIG. **7**, the circuit configuration of a radio-frequency power amplifier according to a fourth embodiment of the present invention will be described. FIG. **7** is a diagram showing the circuit configuration of the radio-frequency power amplifier according to the fourth embodiment. The radio-frequency power amplifier of FIG. **7** is different from the radio-frequency power amplifier of FIG. **2** in that the radio-frequency power amplifier of FIG. **7** further includes a resistance element R**12**. The other components of the circuit configuration of FIG. **7** are the same as those of the circuit configuration of FIG. **2**, and therefore will be denoted by the same numerals and will not be described.

The resistance element R**12** is provided between the resistance element R**2** and the connecting point of the base terminal of the emitter-follower transistor Tr**2** and the collector terminal of the diode transistor Tr**3**.

Here, the use of the amplifying transistor Tr**1** that has a small emitter area decreases the bias current flowing between the base and emitter of the amplifying transistor Tr**1**, compared to the use of the amplifying transistor Tr**1** that has a large emitter area, and therefore the base-emitter current of the emitter-follower transistor Tr**2** decreases.

In the configuration shown in FIG. **2**, only the resistance element R**2** is provided between the reference power supply Vref and the base terminal of the emitter-follower transistor Tr**2**. As a result, the use of the amplifying transistor Tr**1** that has a small emitter area decreases the voltage caused in the resistance element R**2** and raises the base potential of the emitter-follower transistor Tr**2**. To suppress this rise, it is necessary to increase the resistance value of the resistance element R**2**. However, since the collector current of the transistor Tr**5** of the temperature compensation circuit **12** significantly increases at high temperature, if the resistance value of the resistance element R**2** is increased, the base potential of the emitter-follower transistor Tr**2** may significantly fall, and therefore the amplifying transistor Tr**1** may be turned OFF at high temperature. To prevent this, it may be possible to decrease the collector current of the transistor Tr**5** by increasing the resistance value of the resistance element R**6** of the temperature compensation circuit **12**, but it is necessary to increase the area of the resistance element R**6**.

In contrast, in the configuration of the present embodiment, shown in FIG. **7**, the resistance elements R**2** and R**12** are provided between the reference power supply Vref and the base terminal of the emitter-follower transistor Tr**2**. Additionally, the collector current of the transistor Tr**5** of the temperature compensation circuit **12** does not flow through the resistance element R**12**. That is, in the configuration of the present embodiment, the resistance value between the reference power supply Vref and the base terminal of the emitter-follower transistor Tr**2** is increased using the resistance element R**12** that is not involved in the collector current of the transistor Tr**5** of the temperature compensation circuit **12**. Consequently, the configuration of the present embodiment can suppress the rise of the base potential of the emitter-follower transistor Tr**2** without increasing the resistance value of the resistance element R**2**. Further, since the resistance value of the resistance element R**2** is not increased, the base potential of the emitter-follower transistor Tr**2** does not significantly rise at high temperature even when the collector current of the transistor Tr**5** of the temperature compensation circuit **12** increases at high temperature. Therefore, it is possible to prevent the amplifying transistor Tr**1** from being turned OFF at high temperature.

As described above, in the present embodiment, the resistance elements R**2** and R**12** are provided between the reference power supply Vref and the base terminal of the emitter-follower transistor Tr**2**. Consequently, in the configuration of the present embodiment, the use of the amplifying transistor Tr**1** that has a small emitter area makes it possible to suppress the rise of the base potential of the emitter-follower transistor Tr**2** without increasing the size of the resistance element R**2**, and also possible to prevent the amplifying transistor Tr**1** from being turned OFF at high temperature without increasing the area of the resistance element R**6**.

Note that, in FIG. **7**, the reference power supply Vref is connected to the resistance elements R**2** and R**12** that are connected to each other in series, but may be connected to three or more series resistance elements connected in series. In this case, the collector terminal of the transistor Tr**5** of the temperature compensation circuit **12** may be connected to the reference power supply Vref via one or more series resistance elements and also connected to the base terminal of the emitter-follower transistor Tr**2** via one or more series resistance elements. Alternatively, when a plurality of temperature compensation circuits **12** are provided, the collector terminals of the transistors Tr**5** of the temperature compensation circuits **12** may be connected to the same connecting point or different connecting points between series resistance elements.

Note that, in the present embodiment, the mode switching circuit **13** of the second embodiment may be further provided. In this case, the collector terminal of the transistor Tr**6** of the mode switching circuit **13** may be connected to the reference power supply Vref via the resistance element R**8** and one or more series resistance elements and also connected to the base terminal of the emitter-follower transistor Tr**2** via the resistance element R**8** and one or more series resistance elements.

While the invention has been described in detail, the foregoing description is in all aspects illustrative and not restrictive. It is understood that numerous other modifications and variations can be devised without departing from the scope of the invention.

What is claimed is:

1. A radio-frequency power amplifier comprising:

a first transistor operable to amplify a power of a radio-frequency signal; and

a bias circuit operable to supply a bias current to the first transistor;

the bias circuit including:

a second transistor that has a collector terminal connected to a first power supply;

a first resistance element that has one end connected to an emitter terminal of the second transistor and has the other end connected to a base terminal of the first transistor;

a second resistance element that has one end connected to the emitter terminal of the second transistor and has the other end connected to ground potential;

at least one third resistance element provided between a base terminal of the second transistor and a second power supply; and

a plurality of temperature compensation circuits connected to the base terminal of the second transistor and operable to control a base potential of the second transistor so that the potential falls as a temperature rises, at least one of the plurality of temperature compensation circuits being unlike all the rest of the plurality of temperature compensation circuits in configuration.

2. The radio-frequency power amplifier according to claim **1**,

wherein the bias circuit includes a first temperature compensation circuit and a second temperature compensation circuit,

wherein the first temperature compensation circuit includes:

a third transistor that has a collector terminal and a base terminal that are short-circuited, the collector terminal being connected to the base terminal of the second transistor; and

a fourth transistor that has a collector terminal and a base terminal that are short-circuited, the collector terminal being connected to an emitter terminal of the third transistor, and has an emitter terminal connected to ground potential, and

wherein the second temperature compensation circuit includes:

a first voltage divider circuit having a plurality of series resistance elements connected in series;

a fifth transistor that has a collector terminal connected to the base terminal of the second transistor and an emitter terminal connected to ground potential; and

a fourth resistance element that has one end connected to a connecting point of any of the series resistance elements of the first voltage divider circuit and has the other end connected to a base terminal of the fifth transistor.

3. The radio-frequency power amplifier according to claim **2**,

wherein the bias circuit further includes at least one sixth resistance element provided between the base terminal of the second transistor and a connecting point of the third resistance element and the collector terminal of the fifth transistor.

4. The radio-frequency power amplifier according to claim **2**,

wherein the bias circuit further includes a third temperature compensation circuit,

wherein the third temperature compensation circuit includes:

a second voltage divider circuit having a plurality of series resistance elements connected in series;

a sixth transistor that has a collector terminal connected to the base terminal of the second transistor and has an emitter terminal connected to ground potential; and

a fifth resistance element that has one end connected to a connecting point of any of the series resistance elements of the second voltage divider circuit and has the other end connected to a base terminal of the sixth transistor, and

wherein resistance values of the series resistance elements of the first voltage divider circuit, the series resistance elements of the second voltage divider circuit, the fourth resistance element, and the fifth resistance element are each set so that an idle current of the first transistor decreases as a temperature rises at temperatures above a predetermined temperature.

5. The radio-frequency power amplifier according to any one of claims **2** through **4**,

wherein all of the series resistance elements are arranged in the same direction on a substrate.

6. The radio-frequency power amplifier according to claim **1**, further comprising:

a mode switching circuit, connected to the base terminal of the second transistor, operable to switch a bias current to be supplied to the second transistor to a predetermined value of the bias current.

7. The radio-frequency power amplifier according to claim **6**,

wherein the mode switching circuit includes:

a power supply circuit operable to control an output of a predetermined voltage to start or stop;

a seventh transistor that has an emitter terminal connected to ground potential;

a seventh resistance element that has one end connected to an output of the power supply circuit and has the other end connected to a base terminal of the seventh transistor; and

an eighth resistance element that has one end connected to a collector terminal of the seventh transistor and has the other end connected to the base terminal of the second transistor.

* * * * *

US008031003B2

(12) **United States Patent**
Dishop

(10) **Patent No.: US 8,031,003 B2**
(45) **Date of Patent: Oct. 4, 2011**

(54) **SOLID-STATE RF POWER AMPLIFIER FOR RADIO TRANSMITTERS**

(76) Inventor: **Steven M. Dishop**, Bellefontaine, OH (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **12/714,609**

(22) Filed: **Mar. 1, 2010**

(65) **Prior Publication Data**

US 2010/0156537 A1 Jun. 24, 2010

Related U.S. Application Data

(63) Continuation-in-part of application No. 11/749,786, filed on May 17, 2007, now Pat. No. 7,683,718.

(60) Provisional application No. 60/747,662, filed on May 18, 2006, provisional application No. 60/801,006, filed on May 17, 2006.

(51) **Int. Cl.**
H03F 3/26 (2006.01)

(52) **U.S. Cl.** **330/276**; 330/269; 330/275

(58) **Field of Classification Search** 330/269, 330/275, 276, 301
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,420,537	A	5/1995	Weedon et al.	
5,477,188	A	12/1995	Chawla et al.	
5,726,603	A	3/1998	Chawla et al.	
6,046,641	A	4/2000	Chawla et al.	
6,157,258	A	12/2000	Adishian et al.	
6,496,069	B1 *	12/2002	Van De Westerlo	330/276

OTHER PUBLICATIONS

Blocksome, HF Radio Systems & Circuits, 1998; pp. 458-479, 498-511, Noble Publishing Corporation, Atlanta, GA; USA.

Granberg, AN593, “Broadband Linear Power Amplifiers: Using Push-Pull Transistors,” RF Application Reports, 1993; Motorola Literature Distribution, Phoenix, AZ; USA.

Granberg, AN749, “Broadband Transformers & Power Combining Techniques for RF, ” RF Application Reports, 1993; Motorola Literature Distribution, Phoenix, AZ; USA.

Granberg, AN 758, “A Two-Stage 1kW Sold State Linear Amplifier,” RF Application Reports, 1994; Motorola Literature Distribution, Phoenix, AZ; USA.

Granberg, AN762, “Linear Amplifiers for Mobile Operation,” RF Application Reports, Jun. 1984; Motorola Literature Distribution, Phoenix, AZ; USA.

(Continued)

Primary Examiner — Khanh Nguyen
(74) *Attorney, Agent, or Firm* — Pearne & Gordon LLP

(57) **ABSTRACT**

An RF power amplifier includes a push-pull amplifier having field effect transistors. Temperature compensating bias circuitry provides a temperature compensated bias voltage to the transistors for decreasing the bias voltage thereof as temperature increases. The temperature compensating bias circuitry includes a temperature sensor generating a temperature signal. A first amplifier provides a first temperature dependent voltage based on the temperature signal. A second amplifier provides a second temperature dependent voltage based on the temperature signal. The first and second temperature dependent voltages change at substantially the same rate in response to the temperature signal. A potentiometer receives the first and second temperature dependent voltages such that a voltage across the potentiometer remains substantially constant when the first and second temperature dependent voltages change. An output of the bias circuitry is connected to at least one of the transistors and supplies the temperature compensated bias voltage to the at least one of the transistors.

30 Claims, 11 Drawing Sheets

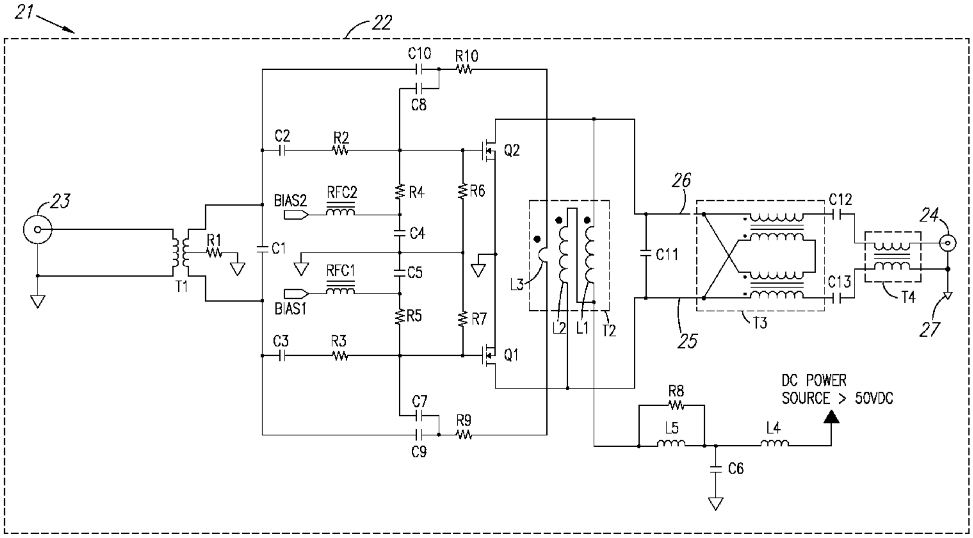

OTHER PUBLICATIONS

Granberg, AR176, "New MOSFETs Simplify High Power RF Amplifier Design," RF Application Reports, 1995; pp. 412-417, Motorola Literature Distribution, Phoenix, AZ; USA.

Granberg, AR313, "Wideband RF Power Amplifier," RF Application Reports, Apr. 1991; Motorola Literature Distribution, Phoenix, AZ; USA.

Granberg, AR347, "A Compact 1-kw 2-50 MHz Solid-State Linear Amplifier," RF Application Reports, Dec. 1991; Motorola Literature Distribution, Phoenix, AZ; USA.

Granberg, AR580, "MOSFET RF Power: An update," QST Magazine, Jan. 1983; pp. 13-16, 30-33, vol. 67, No. 1, American Radio Relay League, Inc., Newington, CT; USA.

Granberg, EB104, "Get 600 Watts RF from Four Power FET's," Engineering Bulletin, Sep. 1993; Motorola Literature Distribution, Phoenix, AZ; USA.

Granberg, AR305, "Building Push-Pull, Multioctave, VHF Power Amplifiers," 1987, Motorola Semiconductor Products Inc., Phoenix, AZ; USA.

* cited by examiner

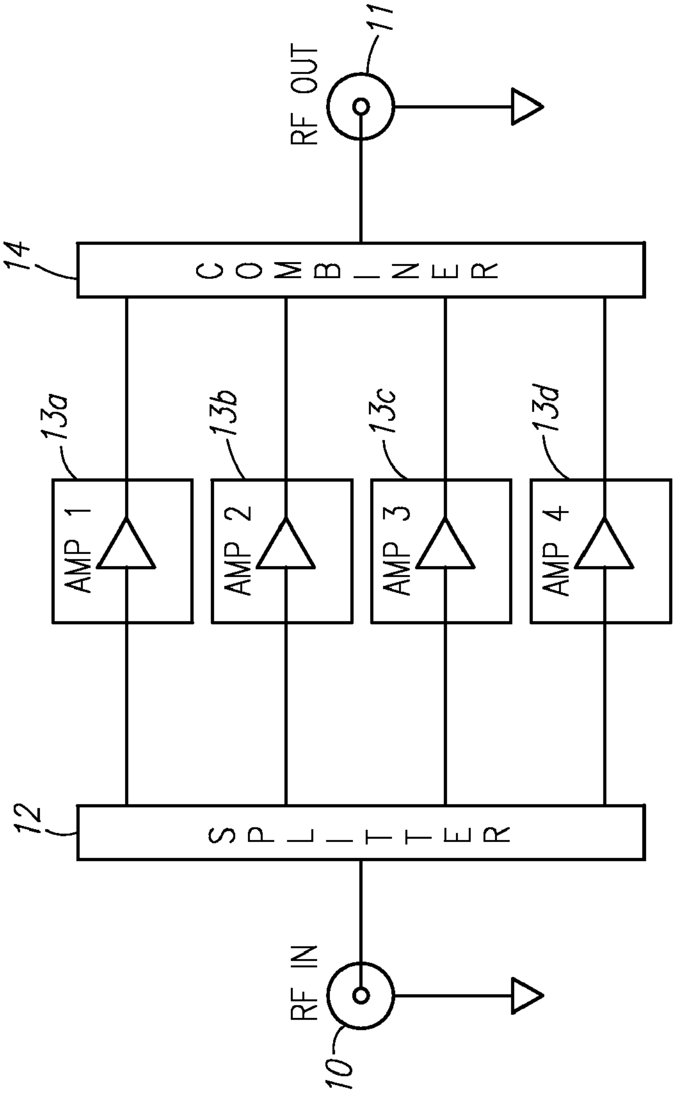
RF OUT
11
14
COMBINER
13a
13b
13c
13d
AMP 1
AMP 2
AMP 3
AMP 4
12
SPLITTER
RF IN
10

Fig. 1

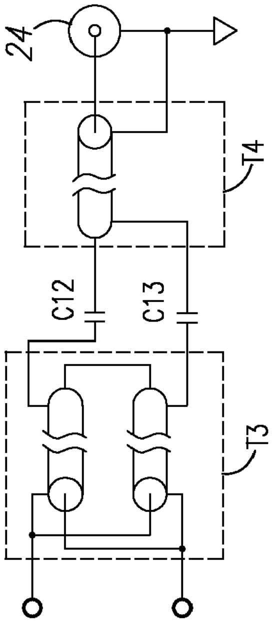
24
T4
C12
C13
T3

Fig. 2a

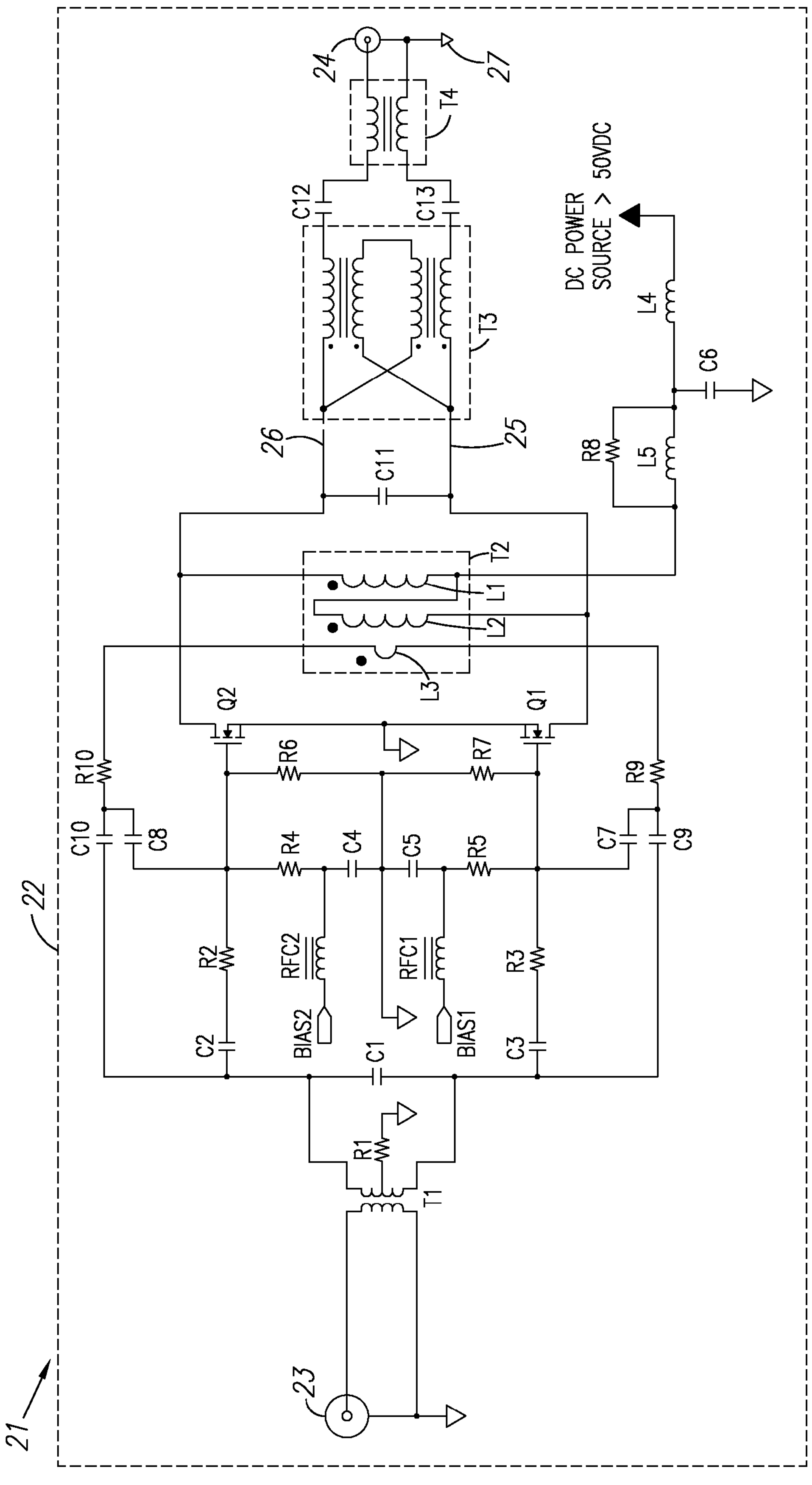
21
22
23
24
25
26
27
T1
R1
C1
C2
C3
R2
R3
RFC1
RFC2
BIAS1
BIAS2
R4
R5
C4
C5
C7
C8
C9
C10
R6
R7
R9
R10
Q1
Q2
L3
L2
L1
T2
C11
T3
C12
C13
T4
R8
L5
C6
L4
DC POWER
SOURCE > 50VDC

Fig. 2

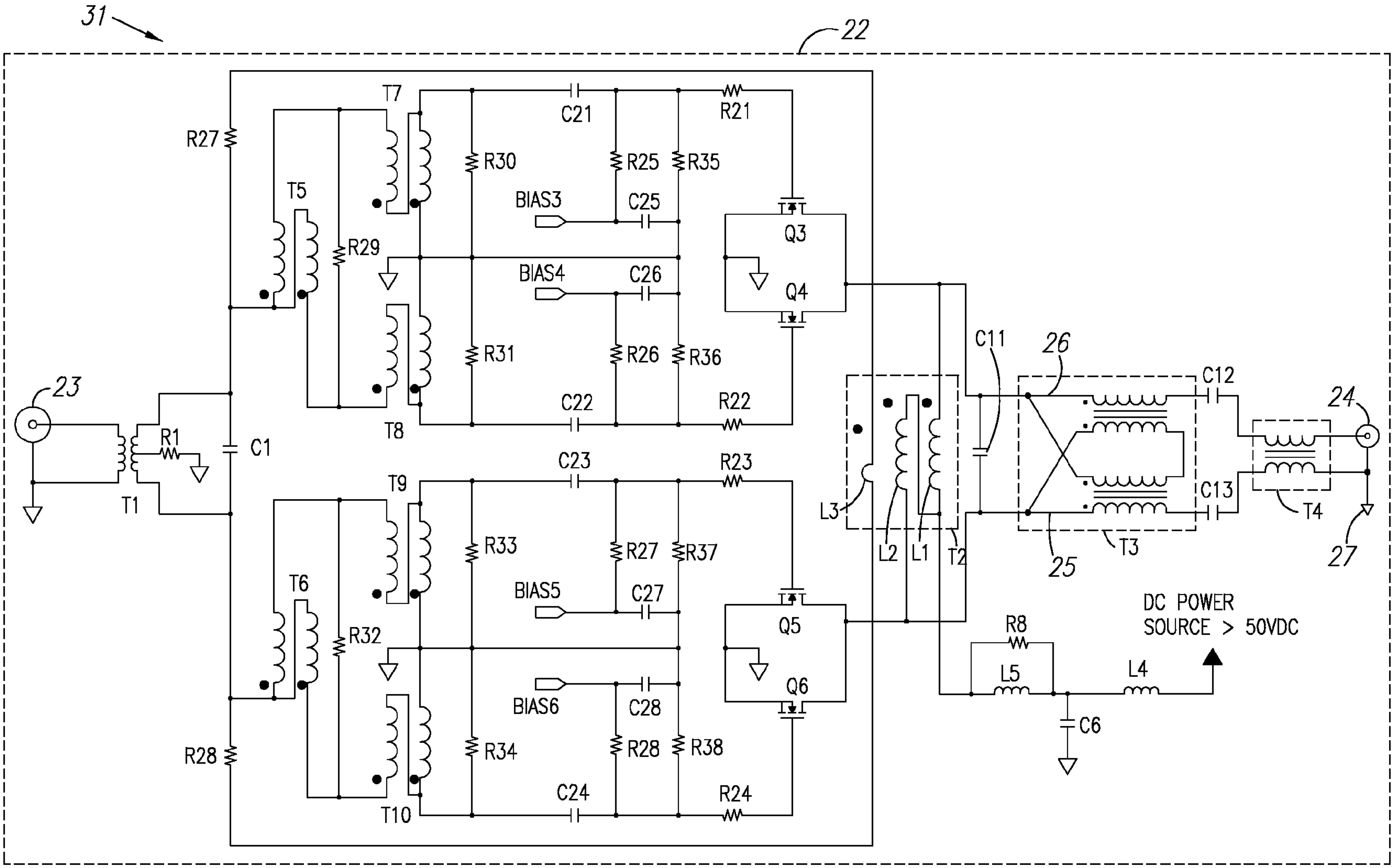
31
22
23
T1
R1
C1
R27
R28
T5
T6
R29
R32
T7
T8
T9
T10
R30
R31
R33
R34
C21
C22
C23
C24
BIAS3
BIAS4
BIAS5
BIAS6
R25
R26
R27
R28
C25
C26
C27
C28
R35
R36
R37
R38
R21
R22
R23
R24
Q3
Q4
Q5
Q6
L3
L2
L1
T2
C11
26
25
T3
C12
C13
T4
24
27
R8
L5
L4
C6
DC POWER
SOURCE > 50VDC

Fig. 3

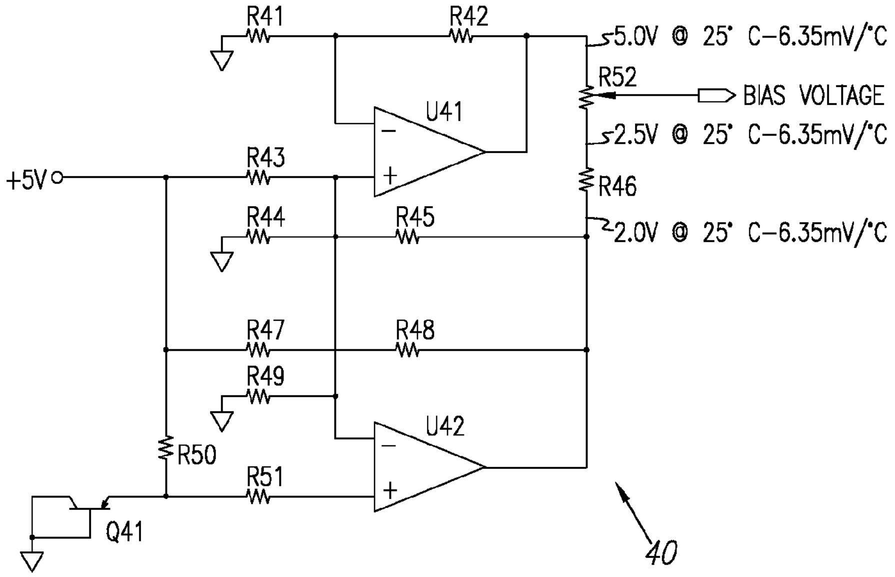
R41
R42
5.0V @ 25° C−6.35mV/°C
R52
BIAS VOLTAGE
U41
2.5V @ 25° C−6.35mV/°C
+5V
R43
R46
R44
R45
2.0V @ 25° C−6.35mV/°C
R47
R48
R49
U42
R50
R51
Q41
40

Fig. 4

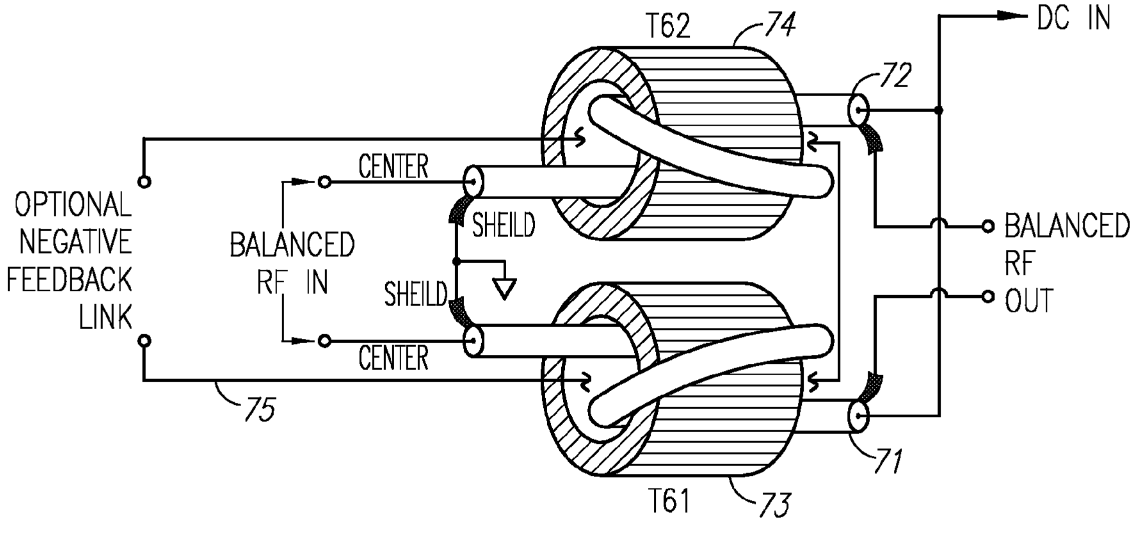
T62
74
DC IN
72
CENTER
OPTIONAL NEGATIVE FEEDBACK LINK
BALANCED RF IN
SHEILD
SHEILD
BALANCED RF OUT
CENTER
75
71
T61
73

Fig. 7

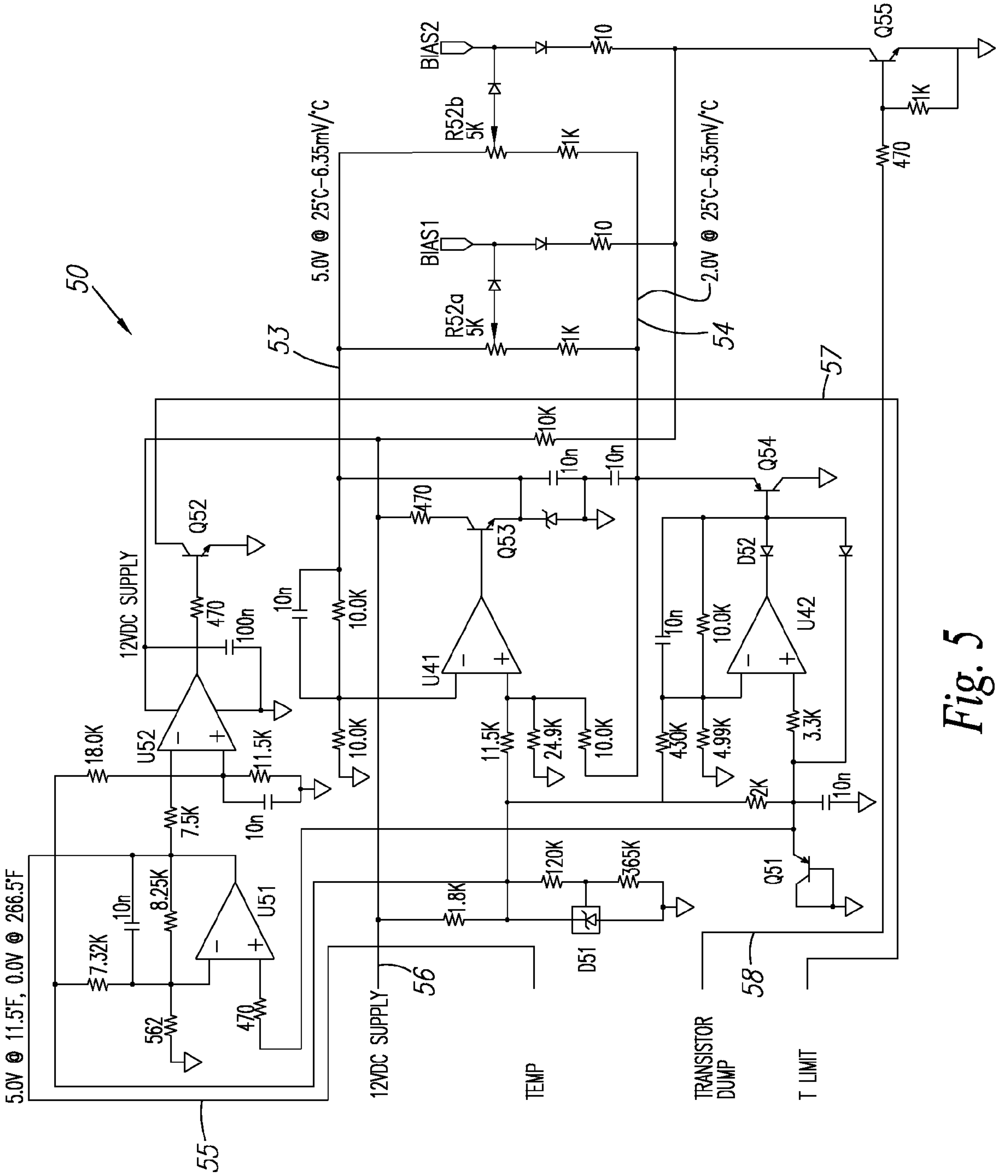
50
BIAS2
10
Q55
1K
470
R52b
5K
1K
5.0V @ 25°C–6.35mV/°C
BIAS1
10
R52a
5K
1K
2.0V @ 25°C–6.35mV/°C
53
54
57
10K
10n
10n
Q54
470
Q53
Q52
12VDC SUPPLY
470
100n
10n
10.0K
U41
D52
10n
10.0K
U42
18.0K
U52
11.5K
10.0K
11.5K
24.9K
10.0K
430K
4.99K
3.3K
2K
10n
10n
7.5K
120K
365K
Q51
5.0V @ 11.5°F, 0.0V @ 266.5°F
7.32K
10n
8.25K
U51
1.8K
D51
562
470
12VDC SUPPLY
56
TEMP
TRANSISTOR DUMP
58
T LIMIT
55

Fig. 5

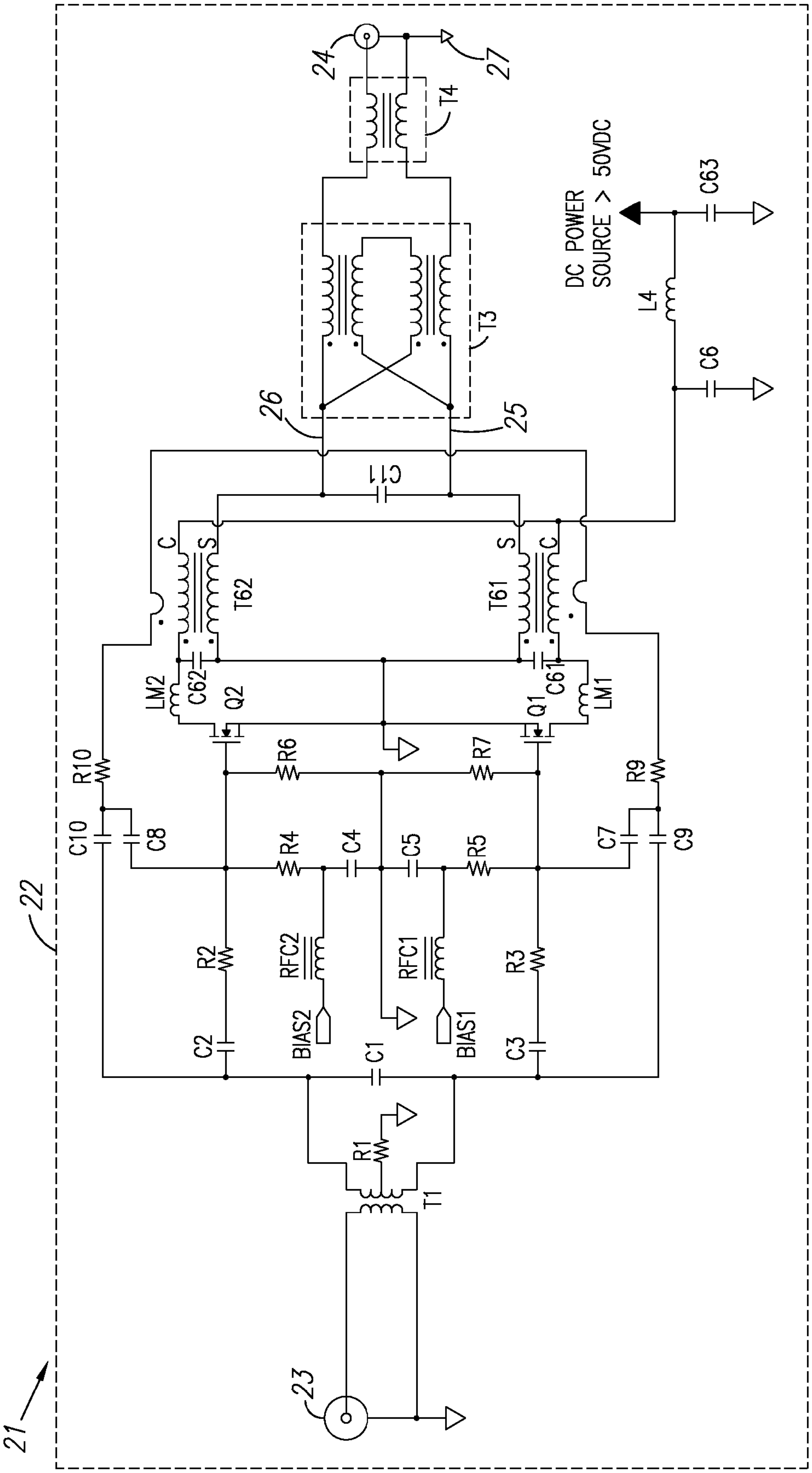
21
22
23
T1
R1
C1
C2
C3
BIAS2
BIAS1
RFC2
RFC1
R2
R3
C10
C8
R4
C4
C5
R5
C7
C9
R10
R9
R6
R7
Q2
Q1
LM2
C62
LM1
C61
T62
T61
C
S
S
C
C11
26
25
T3
T4
24
27
DC POWER
SOURCE > 50VDC
C63
L4
C6

Fig. 6

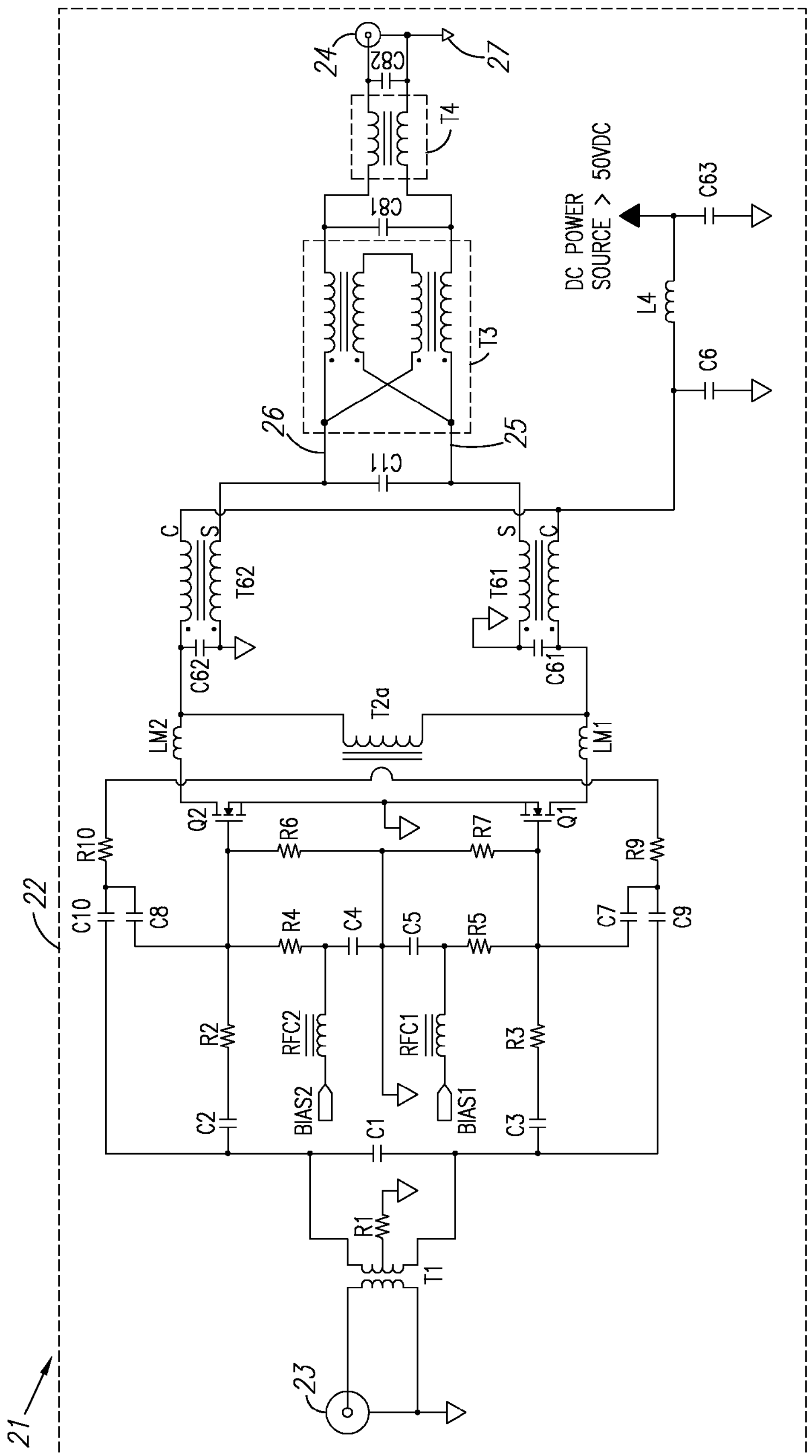
21
22
23
24
25
26
27
T1
R1
C1
C2
C3
R2
R3
RFC1
RFC2
BIAS1
BIAS2
R4
R5
C4
C5
C7
C8
C9
C10
R6
R7
R9
R10
Q1
Q2
LM1
LM2
T2a
C61
C62
T61
T62
C
S
C11
T3
C81
T4
C82
L4
C6
C63
DC POWER
SOURCE > 50VDC

Fig. 8

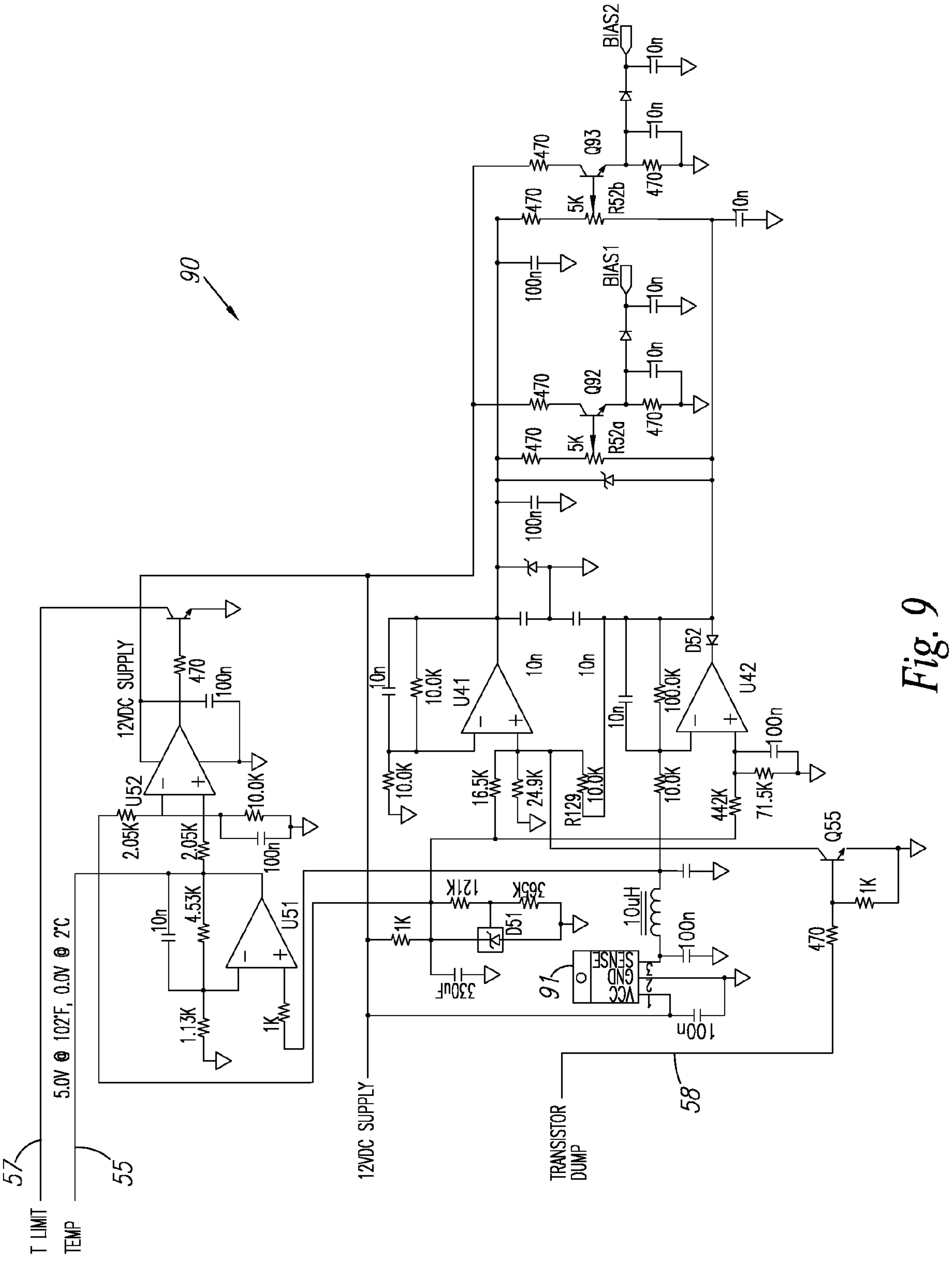
90
BIAS2
10n
10n
470
Q93
470
5K
R52b
470
10n
100n
BIAS1
10n
10n
470
Q92
470
5K
R52a
470
100n
12VDC SUPPLY
470
100n
U52
10.0K
2.05K
2.05K
100n
4.53K
10n
U51
1.13K
1K
10n
10.0K
U41
10n
10n
10.0K
16.5K
24.9K
R129
10.0K
10n
100.0K
U42
D52
100n
10.0K
442K
71.5K
Q55
1K
470
121K
365K
1K
D51
330uF
10uH
100n
SENSE
GND
VCC
91
100n
T LIMIT
57
TEMP
55
5.0V @ 102°F, 0.0V @ 2°C
12VDC SUPPLY
TRANSISTOR DUMP
58

Fig. 9
]

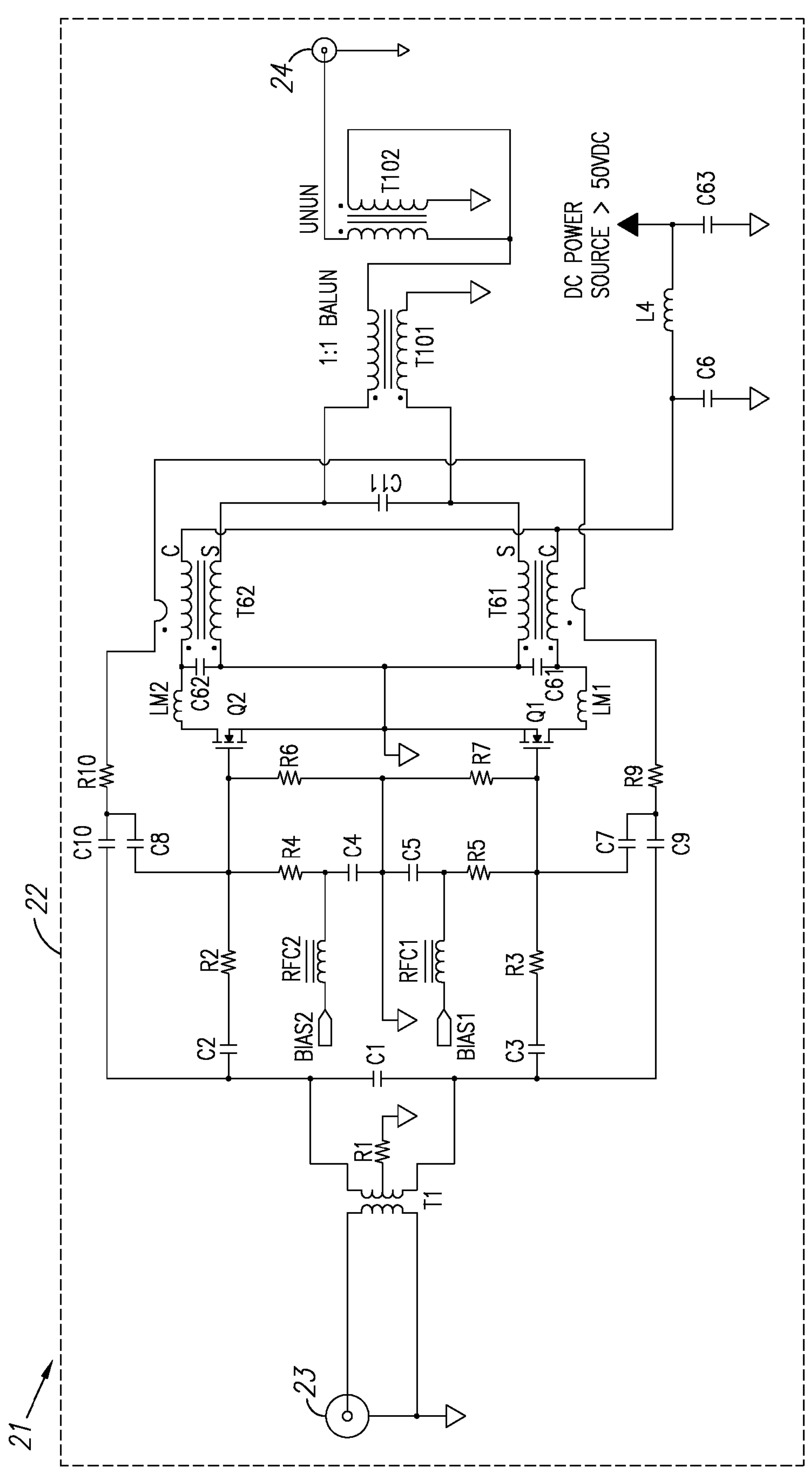
21
22
23
24
T1
R1
C1
C2
C3
R2
R3
BIAS2
BIAS1
RFC2
RFC1
R4
C4
C5
R5
C10
C8
C7
C9
R10
R9
R6
R7
Q2
Q1
C62
C61
LM2
LM1
T62
T61
C
S
S
C
C11
1:1 BALUN
T101
UNUN
T102
DC POWER
SOURCE > 50VDC
L4
C6
C63

Fig. 10

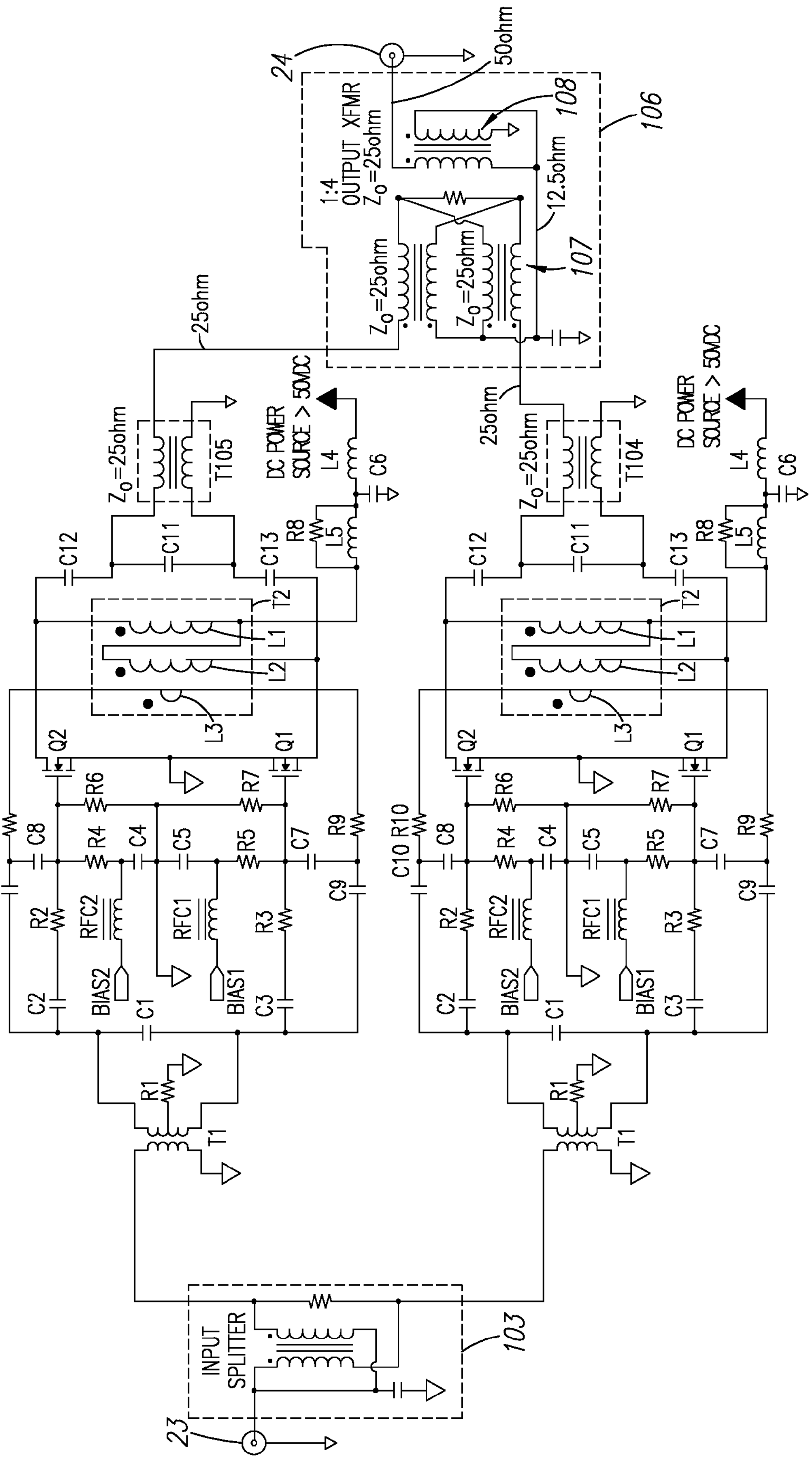
24
50ohm
108
106
1:4
OUTPUT XFMR
Z_o=25ohm
12.5ohm
Z_o=25ohm
Z_o=25ohm
107
25ohm
25ohm
Z_o=25ohm
T105
DC POWER
SOURCE > 50VDC
L4
C6
Z_o=25ohm
T104
DC POWER
SOURCE > 50VDC
L4
C6
C12
C11
C13
R8
L5
T2
L1
L2
L3
Q2
Q1
R6
R7
C10 R10
C8
R4
C4
C5
R5
C7
R9
C9
R2
RFC2
RFC1
R3
C2
BIAS2
C1
BIAS1
C3
R1
T1
C12
C11
C13
R8
L5
T2
L1
L2
L3
Q2
Q1
R6
R7
C10 R10
C8
R4
C4
C5
R5
C7
R9
C9
R2
RFC2
RFC1
R3
C2
BIAS2
C1
BIAS1
C3
R1
T1
INPUT
SPLITTER
103
23

Fig. 11

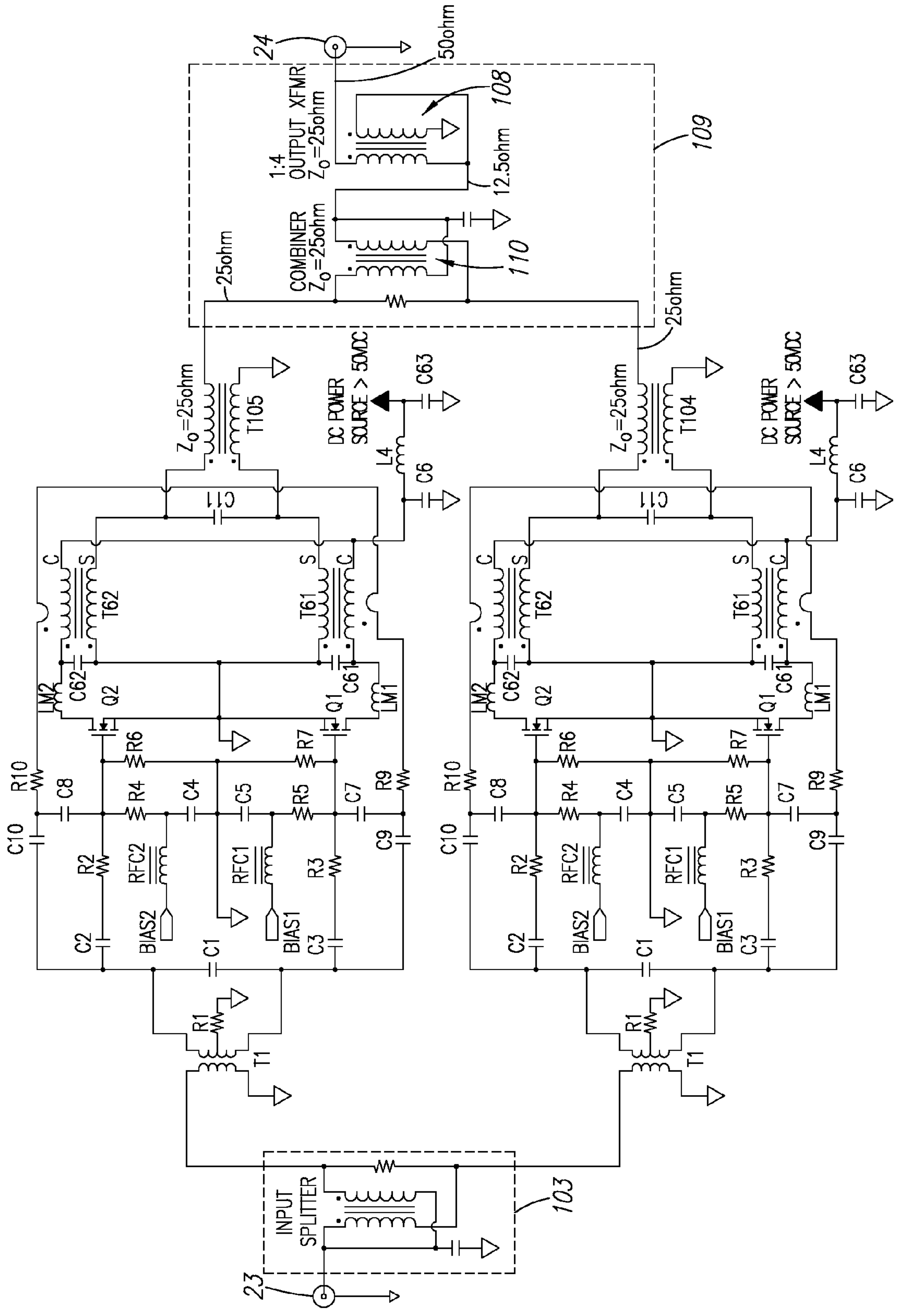
24
50ohm
1:4
OUTPUT XFMR
Z_o=25ohm
108
109
12.5ohm
COMBINER
Z_o=25ohm
110
25ohm
25ohm
Z_o=25ohm
T105
DC POWER
SOURCE > 50VDC
C63
L4
C6
C11
C
S
T62
S
C
T61
LM2
C62
Q2
Q1
C61
LM1
R6
R7
R10
C8
R4
C4
C5
R5
C7
R9
C10
R2
RFC2
RFC1
R3
C9
BIAS2
BIAS1
C2
C1
C3
R1
T1
Z_o=25ohm
T104
DC POWER
SOURCE > 50VDC
C63
L4
C6
C11
C
S
T62
S
C
T61
LM2
C62
Q2
Q1
C61
LM1
R6
R7
R10
C8
R4
C4
C5
R5
C7
R9
C10
R2
RFC2
RFC1
R3
C9
BIAS2
BIAS1
C2
C1
C3
R1
T1
INPUT
SPLITTER
103
23

Fig. 12

SOLID-STATE RF POWER AMPLIFIER FOR RADIO TRANSMITTERS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of U.S. patent application Ser. No. 11/749,786, filed on May 17, 2007, which claims the benefit of U.S. Provisional Patent Application Ser. No. 60/801,006, filed May 17, 2006 and the benefit of U.S. Provisional Patent Application Ser. No. 60/747,662, filed May 18, 2006. The respective disclosures in the above-referenced applications are incorporated herein by reference in their entireties.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to power amplifiers, and more particularly to radio frequency (RF) power amplifiers employing high voltage and high power metal-oxide semiconductor field-effect transistors (MOSFET).

2. Description of Related Art

High-power RF amplifiers adapted to operate over a range of 1-60 MHz without tuning have typically employed vacuum tubes. It would be desirable to provide a high-power RF amplifier for operation over a range of 1-60 MHz, and which employs a minimum number of MOSFET transistors instead of vacuum tubes.

BRIEF SUMMARY OF THE INVENTION

In accordance with one aspect of the invention, provided is an RF power amplifier including a first field effect transistor having a first gate, a first source, and a first drain, having an output power rating of at least 200 watts, and operating with a drain-to-source voltage that is greater than 50 VDC. The amplifier includes a second field effect transistor having a second gate, a second source, and a second drain, having an output power rating of at least 200 watts, and operating with a drain-to-source voltage that is greater than 50 VDC. The transistors are configured as a push-pull amplifier. The amplifier further includes an RF signal input. An input transformer is connected to the RF signal input. The input transformer has respective balanced outputs connected to the first gate and the second gate. A broadband output transformer has a first balanced input connected to the first drain, and a second balanced input connected to the second drain. The broadband output transformer has an input to output impedance ratio of 1:4 and at least some flux cancellation occurs within the broadband output transformer. Temperature compensating bias circuitry provides a temperature compensated bias voltage to the first field effect transistor and the second field effect transistor for decreasing the bias voltage of the first field effect transistor and the second field effect transistor as transistor temperature increases, in order to maintain a constant bias current level over temperature. The temperature compensating bias circuitry includes a temperature sensor generating a temperature signal. A first amplifier has an output providing a first temperature dependent voltage based on the temperature signal. A second amplifier has an output providing a second temperature dependent voltage based on the temperature signal. The first temperature dependent voltage and the second temperature dependent voltage change at substantially the same rate in response to the temperature signal. A potentiometer is connected to the output of the first amplifier and the output of the second amplifier such that a voltage across the potentiometer remains substantially constant when the first temperature dependent voltage and the second temperature dependent voltage change. A bias output of the temperature compensating bias circuitry is connected to at least one of the first field effect transistor and the second field effect transistor and supplies the temperature compensated bias voltage to the at least one of the first field effect transistor and the second field effect transistor.

In accordance with another aspect of the invention, provided is an RF power amplifier including a first plurality of field effect transistors having directly interconnected drains and respective output power ratings of at least 100 watts, and a second plurality of field effect transistors having directly interconnected drains and respective output power ratings of at least 100 watts. The transistors operate with a drain-to-source voltage that is greater than 50 VDC. The first plurality of field effect transistors and the second plurality of field effect transistors together form a push-pull amplifier having an output power rating of at least 400 watts. The amplifier further includes an RF signal input. An input transformer is connected to the RF signal input. The input transformer has respective balanced outputs connected to the gates of the transistors. A broadband output transformer has a first balanced input connected to the drains of the first plurality of field effect transistors, and a second balanced input connected to the drains of the second plurality of field effect transistors. The broadband output transformer has an input to output impedance ratio of 1:4 and at least some flux cancellation occurs within the broadband output transformer. Temperature compensating bias circuitry provides a temperature compensated bias voltage to the first plurality of field effect transistors and the second plurality of field effect transistors for decreasing the bias voltage of the first plurality of field effect transistors and the second plurality of field effect transistors as transistor temperature increases, in order to maintain a constant bias current level over temperature. The temperature compensating bias circuitry includes a temperature sensor generating a temperature signal. A first amplifier has an output providing a first temperature dependent voltage based on the temperature signal. A second amplifier has an output providing a second temperature dependent voltage based on the temperature signal. The first temperature dependent voltage and the second temperature dependent voltage change at substantially the same rate in response to the temperature signal. A potentiometer is connected to the output of the first amplifier and the output of the second amplifier such that a voltage across the potentiometer remains substantially constant when the first temperature dependent voltage and the second temperature dependent voltage change. A bias output of the temperature compensating bias circuitry is connected to at least one of the first plurality of field effect transistors and the second plurality of field effect transistors and supplies the temperature compensated bias voltage to the at least one of the first plurality of field effect transistors and the second plurality of field effect transistors.

In accordance with another aspect of the invention, provided is an RF power amplifier including a push-pull amplifier, the push-pull amplifier comprising a first field effect transistor and a second field effect transistor. Temperature compensating bias circuitry provides a temperature compensated bias voltage to the first field effect transistor and the second field effect transistor for decreasing the bias voltage of the first field effect transistor and the second field effect transistor as transistor temperature increases. The temperature compensating bias circuitry includes a temperature sensor generating a temperature signal. A first amplifier has an output providing a first temperature dependent voltage based on the temperature signal. A potentiometer is connected to the output of the first amplifier such that a voltage across the potentiometer remains substantially constant when the first temperature dependent voltage changes. A bias output of the temperature compensating bias circuitry is connected to at least one of the first field effect transistor and the second field effect transistor and supplies the temperature compensated bias voltage to the at least one of the first field effect transistor and the second field effect transistor.

In accordance with another aspect of the invention, provided is an RF power amplifier including a push-pull amplifier, the push-pull amplifier comprising a first field effect transistor having a first drain and a second field effect transistor having a second drain. A broadband output transformer has a first balanced input and a second balanced input. A first DC-blocking transformer is connected between the first balanced input and the first drain. A second DC-blocking transformer is connected between the second balanced input and the second drain.

In accordance with another aspect, provided is an RF power amplifier. The RF power amplifier includes a first amplifier module and a second amplifier module. The first amplifier module comprises a first push-pull amplifier including a plurality of field effect transistors and a first output balun transformer. An output impedance of the first amplifier module is 25 ohms. The second amplifier module comprises a second push-pull amplifier including a plurality of field effect transistors and a second output balun transformer. The output impedance of the second amplifier module is 25 ohms. A combiner is connected to the first amplifier module and the second amplifier module. The combiner comprises an unbalanced-to-unbalanced output transformer having an input-to-output impedance ratio of 1:4. The combiner combines an output from the first amplifier module and an output from the second amplifier module into a combined signal. The combined signal is supplied to the unbalanced-to-unbalanced output transformer. An output impedance of the combiner is 50 ohms.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a top-level schematic block diagram of a modular RF power amplifier system;

FIG. **2** is a schematic circuit diagram of an RF power amplifier module;

FIG. **2***a* is a schematic circuit diagram of a portion of FIG. **2**;

FIG. **3** is a schematic circuit diagram of an RF power amplifier module;

FIG. **4** is a schematic circuit diagram of temperature compensating bias circuitry for an RF power amplifier module;

FIG. **5** is a schematic circuit diagram of temperature compensating bias circuitry for an RF power amplifier module;

FIG. **6** is a schematic circuit diagram of an RF power amplifier module;

FIG. **7** is a schematic diagram of DC-blocking transformers for an RF power amplifier module;

FIG. **8** is a schematic circuit diagram of an RF power amplifier module;

FIG. **9** is a schematic circuit diagram of temperature compensating bias circuitry for an RF power amplifier module;

FIG. **10** is a schematic circuit diagram of an RF power amplifier module

FIG. **11** is a schematic circuit diagram of an RF power amplifier; and

FIG. **12** is a schematic circuit diagram of an RF power amplifier.

DETAILED DESCRIPTION OF THE INVENTION

As used herein, the terms “connected” and “connected to” refer a physical and/or electrical joining or linking of one thing to another, and includes direct and indirect connections. For example, an amplifier can be connected to an RF signal input by direct electrical connection between the amplifier and input, or connected to said input via an indirect electrical connection, such as through an interposing resistor or capacitor. In the former case, the amplifier is directly connected to the input. In the latter case, the amplifier is indirectly connected to the input. However, in both cases, the amplifier is connected to the RF input.

FIG. **1** shows a top level schematic block diagram of a modular RF power amplifier system. The amplifier system includes an RF input terminal **10** and an RF output terminal **11**. In an embodiment, the characteristic impedance of the input and output terminals is 50 ohms. The terminals **10**, **11** can be adapted for use with removable connectors. For example, the terminals **10**, **11** can include a BNC or SMA female connector for removably connecting to a BNC or SMA male connector, respectively.

An RF signal to be amplified by the amplifier system is provided at the RF input terminal **10**. The RF signal is split by splitter **12** into separate signals to be amplified by a number of power amplifier modules **13***a*-**13***d*. In the example of FIG. **1**, four power amplifier modules **13***a*-**13***d* are provided. Therefore, the splitter **12** splits the RF signal into four separate signals. It is to be appreciated that fewer than four power amplifier modules could be provided, such as two or three amplifier modules, for example. It is to be further appreciated that greater than four power amplifier modules could be provided, such as six or eight amplifier modules, for example. The splitter **12** is designed to split the RF input signal into as many separate signals as there are power amplifier modules. The splitter **12** can include a cascade of separate splitters for staged splitting of the RF input signal into a number of separate signals.

The power amplifier modules **13***a*-**13***d* amplify the separate signals from the splitter **12** according to the power ratings of the amplifier modules **13***a*-**13***d*. In an embodiment, each power amplifier module **13***a*-**13***d* has a power rating of 400 watts. A modular RF power amplifier system employing four 400 watt power amplifier modules would provide approximately 1600 watts of total amplification. In another embodiment, each power amplifier module **13***a*-**13***d* has a power rating of 600 watts. A modular RF power amplifier system employing four 600 watt power amplifiers would provide approximately 2400 watts of total amplification. In still another embodiment, each power amplifier module **13***a*-**13***d* has a power rating of 1200 watts. A modular RF power amplifier system employing four 1200 watt power amplifiers would provide approximately 4800 watts of total amplification. In still another embodiment, each power amplifier module **13***a*-**13***d* has a power rating of 1500 watts. A modular RF power amplifier system employing four 1500 watt power amplifies would provide approximately 6000 watts of total amplification. As discussed above, fewer than four power amplifier modules could be provided or greater than four power amplifier modules could be provided. The number of power amplifier modules to be used and their power ratings can be chosen based on the required power output.

Example power amplifier modules **13***a*-**13***d* are shown in FIGS. **2** and **3** and are discussed in detail below. In an embodiment, each power amplifier module **13***a*-**13***d* is provided on a separate printed circuit board. In other embodiments, a plurality of amplifier modules are provided on one or more printed circuit boards.

Outputs from each of the power amplifier modules **13***a*-**13***d* are provided to a combiner **14**. In the example of FIG. **1**, four power amplifier modules **13***a*-**13***d* are provided and the combiner **14** combines the four outputs from the amplifier modules into a single, combined RF output. The combiner **14** is designed to combine as many signals as there are power amplifier modules. The combiner **14** can include a cascade of separate combiners for staged combining of the amplified signals into a combined signal. As shown in FIG. **1**, the RF output from the combiner **14** is provided to the RF output terminal **11**.

An RF power amplifier could comprise a single amplifier module, rather than a plurality of modules **13***a*-**13***d* as shown in FIG. **1**. An RF power amplifier having a single amplifier module would not require the splitter **12** and combiner **14**.

FIG. **2** shows a push-pull power amplifier module **21**. Transistors Q**1** and Q**2** form a push-pull pair for amplifying an RF signal.

The components that form the amplifier module **21** can be mounted on a printed circuit board **22**. The amplifier module includes an RF input terminal **23** and an RF output terminal **24**. In an embodiment, the characteristic impedance of the module's input and output terminals is 50 ohms. The terminals **23**, **24** can be adapted for use with removable connectors, such as BNC or SMA connectors, for example.

An RF signal to be amplified by the amplifier module **21** is provided at the RF input terminal **23**. The RF signal is transmitted to an input transformer T**1**. The transformer T**1** has an unbalanced or single-ended side, which is connected to the RF input terminal **23**. The unbalanced or single-ended side is coupled to a balanced or differential side of the transformer T**1**. The transformer T**1** is shown as a so-called conventional transformer, having separated primary and secondary windings. It is to be appreciated that the transformer T**1** could be constructed as a transmission line transformer, which does not have separated primary and secondary windings. An optional ground reference is provided for the transformer's balanced or differential side via a resistor R**1**. The transformer T**1** has an input to output impedance ratio of 4:1, for example, and serves to divide the unbalanced RF signal into balanced signals, 180° out of phase, for amplification by the push-pull transistor pair Q**1**, Q**2**, and subsequent combination. It is to be appreciated that the transformer T**1** could have an input to output impedance ratio of other than 4:1, such as 1:1.414, 1:9, 1:25, 3:2, etc. Further, it is to be appreciated that a so-called pi attenuator or pi input attenuator (not shown) comprising a plurality of resistors in a Greek letter "pi" configuration could be provided between the RF input terminal **23** and the transformer T**1**, for normalizing gain of the amplifier module **21**. Other attenuators could be provided between the RF input terminal **23** and the transformer T**1**, such as an L attenuator or a T attenuator, for example.

The input transformer T**1** is shown as an unbalanced-to-balanced transformer. However, it is to be appreciated that T**1** could alternatively be a balanced-to-balanced transformer.

An optional compensation capacitor C**1** is connected across the balanced terminals of the transformer T**1** and provides a low-pass filter response which absorbs the inductance of T**1** and helps compensate for a loss of gain at higher frequencies.

The transistors Q**1**, Q**2** that form the push-pull pair are high-voltage MOSFET RF power transistors. The transistors Q**1**, Q**2** have output power ratings of at least 150 watts, preferably at least 200 watts, and operate with a drain-to-source voltage that is greater than 50 VDC, such as 62 VDC, 72 VDC, 86 VDC, 96 VDC or 100 VDC, for example. Example transistors Q**1**, Q**2** have output power ratings of 150 watts, 300 watts, and 750 watts. An example 150 watt transistor is model ARF520 manufactured by ADVANCED POWER TECHNOLOGY®. A further example 150 watt transistor is model SD3931 manufactured by STMICROELECTRONICS®. An example 300 watt transistor is model SD3933 manufactured by STMICROELECTRONICS®. An example 750 watt transistor is model ARF1500 manufactured by ADVANCED POWER TECHNOLOGY®. It is to be appreciated that transistors having output power ratings other than 150 watts, 300 watts, and 750 watts can be used in an amplifier module as shown in FIG. **2**. For example, 200 watt transistors can be used in the amplifier module.

An amplifier module **21** as shown in FIG. **2**, having 150 watt transistors Q**1**, Q**2** forming a push-pull pair, can have a power rating of 300 watts. An amplifier module **21** having 300 watt transistors Q**1**, Q**2** forming a push-pull pair, can have a power rating of 600 watts. An amplifier module **21** having 750 watt transistors Q**1**, Q**2** forming a push-pull pair, can have a power rating of 1500 watts. It is to be appreciated that the power rating of the amplifier module depends on the power rating of the selected transistors and the drain-to-source voltage at which the transistors are operated. Further, it is to be appreciated that amplifier modules of various power ratings can be constructed.

The balanced signals from the transformer T**1** are respectively provided to the gates of the transistors Q**1**, Q**2**. One balanced signal is provided to the gate of transistor Q**2** through a coupling capacitor C**2** and a resistor R**2**, for amplification by the transistor Q**2**. The other balanced signal is provided to the gate of transistor Q**1** through a coupling capacitor C**3** and resistor R**3**, for amplification by the transistor Q**1**. In an embodiment, capacitors C**2** and C**3** are each formed by two paralleled 47 nF capacitors and resistors R**2** and R**3** are each formed by four paralleled 15 ohm resistors.

A DC bias voltage for the gate of transistor Q**2** is provided at BIAS**2**. The bias voltage is provided to the gate of transistor Q**2** through an RF choke RFC**2** and resistor R**4**. A capacitor C**4** is connected to the RF choke RFC**2** and resistor R**4**, and also to ground, and provides a low impedance path to ground for high frequency signals. In an embodiment, the RF choke RFC**2** has a value of 10 μH. The DC bias voltage for the gate of transistor Q**2** can be provided by a temperature compensating bias circuit.

Similarly, a DC bias voltage for the gate of transistor Q**1** is provided at BIAS**1**. The DC bias voltage BIAS**1** is provided to the gate of transistor Q**1** through an RF choke RFC**1** and resistor R**5**. A capacitor C**5** is connected to the RF choke RFC**1** and resistor R**5**, and also to ground. The capacitor C**5** provides a low impedance path to ground for high frequency signals. In an embodiment, the RF choke RFC**1** has a value of 10 μH. The DC bias voltage for the gate of transistor Q**1** can be provided by a temperature compensating bias circuit. Temperature compensating bias circuits are shown in FIG. **4**, FIG. **5** and FIG. **9** and are discussed further below.

In an embodiment, capacitors C**4** and C**5** are each formed by two paralleled 47 nF capacitors and resistors R**4** and R**5** are each formed by four paralleled 110 ohm resistors.

The gates of the transistors Q**1**, Q**2** are respectively connected to ground through resistors R**7** and R**6**, which provide a discharge path for the charge on the gates when DC bias is removed, and provide a solid ground reference for the DC bias voltages for the gates. The respective sources of transistors Q**1** and Q**2** are directly connected to ground. In an embodiment, resistors R**6** and R**7** each have a value of 10 kΩ.

As shown in FIG. **2**, a DC power source that is greater than 50 VDC is connected to the drain of each transistor Q**1**, Q**2** through a common mode choke T**2**. Coils L**1** and L**2** of the common mode choke T**2** are connected such that magnetic flux is cancelled during each RF cycle, to minimize the net flux inside of the choke and, therefore, minimize the size of its core. A coil L**3**, which is an additional winding of the common mode choke, provides a negative feedback signal from the DC feed structure. Coil L**3** can be provided by a single turn through the center of the core of the common mode choke T**2**. Negative feedback serves to lower the input impedance and to stabilize the amplifier's gain over its frequency range. A feedback path for the gate of transistor Q**1** is provided through a network that includes resistor R**9** and either capacitor C**7** or optionally capacitor C**9**. A feedback path for the gate of transistor Q**2** is provided through a network that includes resistor R**10** and either capacitor C**8** or optionally capacitor C**10**. In an embodiment, capacitors C**7** and C**8** are each formed by two paralleled 47 nF capacitors while capacitors C**9** and C**10** are not used, and resistors R**9** and R**10** are each formed by four paralleled 36 ohm resistors.

It is to be appreciated that coil L**3** may be omitted and that negative feedback may be taken directly from the drains of the transistors Q**1** and Q**2**. In such a configuration, the feedback resistors R**9** and R**10** will be made physically larger because the voltage at the drains is proportionally larger by the winding ratio of coil L**3** to T**2**. More specifically, the voltage from coil L**3** is proportional to the ratio of L**3** turns divided by the sum of the turns of coils L**1** and L**2** times the RF drain-to-drain voltage applied to the primary low impedance side of transformer T**3** (transformer T**3** is discussed in detail below). For example if coil L**3** is one turn and L**1** and L**2** are each eight turns, the voltage available from L**3** is 1/16th of the drain-to-drain RF voltage.

The DC power source is connected to the common mode choke T**2** through a network that includes inductors L**4** and L**5**, a capacitor C**6** and a resistor R**8**. The inductors L**5** and L**6** provide a high impedance to RF signals and a short circuit for the DC power source. RF signals are decoupled from the DC power source by conduction to ground through the resistor R**8** and the capacitor C**6**. Inductors L**4** and L**5** can be ferrite bead inductors. In an embodiment, inductor L**5** is a wound ferrite core with a value of 10 μH, and capacitor C**6** is formed by six paralleled 47 nF capacitors. In addition to capacitor C**6**, similar grounded capacitors can be provided between inductor L**4** and the DC power source, and between inductor L**5** and the common mode choke T**2**. If, in addition to capacitor C**6**, similar grounded capacitors are provided between inductor L**5** and common mode choke T**2**, then resistor R**8** can be omitted. Sometimes, in an effort to increase isolation between the drains, the common DC feed point terminal of the common mode choke T**2** is split, and each winding is fed by identical decoupling networks similar to inductor L**5**, resistor R**8**, capacitor C**6**, and inductor L**4**. In such a configuration, the resistors need to be twice the value of resistor R**8** because they are AC paralleled. If in addition to capacitor C**6** similar grounded capacitors are used at the DC feed terminals of the common mode choke T**2**, the resistors can be omitted.

A broadband transmission line output transformer T**3** is connected to the drain of each transistor Q**1**, Q**2** and combines the amplified RF signals from each transistor Q**1**, Q**2**. The transformer T**3** is a balanced-to-balanced transformer having respective balanced inputs **25**, **26** connected to the drains of the transistors Q**1**, Q**2**. In FIG. **2**, balanced input **25** is directly connected to the drain of transistor Q**1**, and balanced input **26** is directly connected to the drain of transistor Q**2**. However, it is to be appreciated that DC-blocking capacitors could be provided between the balanced inputs **25**, **26** and the drains of the transistors Q**1**, Q**2**. The transformer T**3** has an input to output impedance ratio of 1:4 and performs an impedance matching function. In an embodiment, the characteristic impedance of the transistor output circuitry is 12.5 ohms. The transformer T**3** combines the outputs from the transistors Q**1**, Q**2** and steps the characteristic impedance of the circuit up to 50 ohms.

Transformer T**3** can be constructed using suitable cores, for example, toroid or ferrite tube cores, and coaxial cables having a characteristic impedance of approximately or exactly 25 ohms. Each coaxial cable can be wound on its own core. 25 ohms is the geometric mean of a 12.5 ohm input and a 50 ohm output. Performance of the transformer T**3** is enhanced when the characteristic impedance of the transformer's cables is the geometric mean of the input and output impedances. In an embodiment, the transformer T**3** includes coaxial cables having a characteristic impedance of 25 ohms, a size 16 AWG stranded center conductor, TEFLON® insulation between an outer braid and the center conductor, and an insulating jacket. The 25 ohm transmission line used to construct the transformer T**3** can also be constructed by paralleling standard 50 ohm coaxial cable. Further, the transmission line used to construct the transformer T**3** may be constructed from parallel magnet wire or parallel twisted TEFLON®-insulated wire.

As stated above, the transformer T**3** is a balanced-to-balanced transformer. There is essentially zero net flux within the cores of the transformer T**3** because the currents in its transmission lines travel in opposite directions, which gives rise to identical fluxes of opposite sense. It is to be appreciated that in a balanced-to-unbalanced or unbalanced-to-unbalanced transformer, the flux cancellation is not as complete. Therefore, the cores of the transformer T**3** can be made smaller than those used in a balanced-to-unbalanced or unbalanced-to-unbalanced transformer. Smaller transformer cores allow for shorter windings, thereby increasing the high frequency response of the transformer T**3**. Also, smaller transformer cores can accommodate a greater number or windings, resulting in increased inductance, which extends the lower frequency range of the transformer, such as to 1 MHz for example. In an embodiment, transformer T**3** has an operating frequency range of 1-60 MHz.

One benefit of using a 1:4 impedance ratio transmission line transformer as an RF output transformer, as shown in FIG. **2**, is that such a transformer is less difficult to construct than other transformers. A transformer having a 1:4 impedance ratio has an integer turns ratio of 1:2 and, therefore, does not require a winding tap. Such a transformer is less difficult to construct than a transformer having an impedance ratio of 1:2, which has a non-integer turns ratio of 1:1.414. However, in order to utilize a 1:4 impedance ratio transformer as an RF output transformer in a 600 watt push-pull amplifier module, transistors that operate with a drain-to-source voltage that is greater than 50 VDC must be used, rather than conventional 50 VDC MOSFETs.

An optional compensation capacitor C**11** is connected across the balanced input terminals **25**, **26** of the transformer T**3**, and provides a low-pass filter response by absorbing the inductance of T**3** and any output capacitance of the transistors Q**1**, Q**2**. This helps compensate for gain slope reduction at the high end of the amplifier module's frequency range.

The balanced outputs of transformer T3 are connected to an optional output balun transformer T4 through DC-blocking capacitors C12, C13. Capacitor C12 is connected between one output of transformer T3 and one input of balun transformer T4. Capacitor C13 is connected between the other output of transformer T3 and the other input of balun transformer T4. The capacitors C12, C13 and balun transformer T4 are connected in series between transformer T3 and the RF output terminal 24. In the embodiment of FIG. 2, the capacitors C12, C13 are located at the output, 50 ohms side of transformer T3, and between transformer T3 and balun transformer T4. By locating the DC-blocking capacitors at the output, 50 ohms side of transformer T3, rather than at the input side, the capacitors C12, C13 can be designed to handle a lower current. Such capacitors may be less expensive than capacitors designed to handle a higher current. The optional output balun transformer T4 has an impedance ratio of 1:1 and is connected to RF output terminal 24. In an embodiment, the output balun transformer T4 is constructed using 50 ohm coaxial cable and a ferrite core. Alternatively, transformer T4 may be formed by paralleled magnet wire or by parallel or twisted TEFLON®-insulated wire. By locating the DC-blocking capacitors C12, C13 between transformers T3 and T4, the RF output terminal 24 can be directly connected to the output balun transformer T4 and installed directly on a ground plane 27 of the printed circuit board 22. It is to be appreciated that the output balun transformer T4 is optional and can be omitted in some applications. However, omitting the output balun transformer T4 may result in reduced efficiency and/or increased intermodulation distortion (IMD).

A further schematic view of the broadband output transformer T3, the DC-blocking capacitors C12, C13, and the output balun transformer T4 is provided at FIG. 2*a*. In FIG. 2*a*, the broadband output transformer T3 and output balun transformer T4 are schematically shown as comprising coaxial cables. The output balun transformer T4 is shown having a coaxial cable shield connected to ground and a coaxial cable center conductor connected to the RF output terminal 24. It is to be appreciated that the output balun transformer T4 can be reversely connected, so that the coaxial cable shield is connected to the RF output terminal 24 and the coaxial cable center conductor connected to ground. The high impedance output side of the broadband output transformer T3 is connected the output balun transformer T4 through the DC-blocking capacitors C12, C13. In FIG. 2*a*, at the high impedance output side of transformer T3, the coaxial cable center conductors are shown as directly connected together, and the coaxial cable shields are connected to respective DC-blocking capacitors C12, C13. It is to be appreciated that the high impedance output side of transformer T3 can be reversely connected, so that its coaxial cable shields are directly connected together, and the coaxial cable center conductors are connected to respective DC-blocking capacitors C12, C13.

In an embodiment, capacitors C12 and C13 are each formed by six paralleled 47 nF capacitors.

FIG. 3 shows another embodiment of a push-pull amplifier module 31. Various components shown in FIG. 3 are discussed above with respect to FIG. 2. Such components are referenced in FIG. 3 by identical reference characters as used in FIG. 2 and are not discussed in detail below.

A plurality of transistors having grounded sources and directly interconnected drains form each half of the push-pull amplifier. For example, two transistors Q3, Q4 form the "push" half of the amplifier and two transistors Q5, Q6 form the "pull" half of the amplifier. The transistors Q3-Q6 can have output power ratings of at least 100 watts and operate with a drain-to-source voltage that is greater than 50 VDC. An example 100 watt transistor is model ARF463 manufactured by ADVANCED POWER TECHNOLOGY®. An amplifier module 31, having four 100 W transistors Q3-Q6 can have a power rating of 400 W. An amplifier module 31 having four 150 watt transistors Q3-Q6 can have a power rating of 600 watts. An amplifier module 31 having four 300 watt transistors Q3-Q6 can have a power rating of 1200 watts. In additional embodiments, each half of the push-pull amplifier includes more than two transistors. It is to be appreciated that amplifier modules of various power ratings can be constructed, based on the power rating of the selected transistors, the drain-to-source voltage at which the transistors are operated, and the number of transistors provided in each half of the push-pull amplifier.

An RF signal to be amplified by the amplifier module 31 is provided at the RF input terminal 23. The RF signal is transmitted to the transformer T1 and split into balanced RF signals at the output of transformer T1.

The balanced RF signals from transformer T1 are respectively provided to transformers T5 and T6. Transformers T5 and T6 are so-called Type 1 Splitters, which further split the RF signals. The respective split signals from transformer T5 are provided to transformers T7 and T8, which are impedance matching transformers having an input-to-output impedance ratio of, for example, 1:4. The RF signal that is output from impedance matching transformer T7 is provided to the gate of transistor Q3 through DC-blocking capacitor C21 and resistor R21. Similarly, the RF signal that is output from impedance matching transformer T8 is provided to the gate of transistor Q4 through DC-blocking capacitor C22 and resistor R22. In an embodiment, the DC-blocking capacitors C21 and C22 are each formed by two paralleled 47 nF capacitors.

The respective split signals from transformer T6 are provided to impedance matching transformers T9 and T10. The RF signal that is output from impedance matching transformer T9 is provided to the gate of transistor Q5 through DC-blocking capacitor C23 and resistor R23. Similarly, the RF signal that is output from impedance matching transformer T10 is provided to the gate of transistor Q6 through DC-blocking capacitor C24 and resistor R24. In an embodiment, the DC-blocking capacitors C23 and C24 are each formed by two paralleled 47 nF capacitors.

It is to be appreciated that a splitter utilizing resistors for further splitting the RF signals from transformer T1 could be used, rather than a Type 1 Splitter and impedance matching transformers as shown in FIG. 3.

A DC bias voltage for the gate of transistor Q3 is provided at BIAS3 and through a resistor R25, for example, a 15Ω resistor. A capacitor C25, for example, a 47 nF capacitor, provides a low impedance path to ground for high frequency signals. A DC bias voltage for the gate of transistor Q4 is provided at BIAS4 and through a resistor R26, for example, a 15Ω resistor. A capacitor C26, for example, a 47 nF capacitor, provides a low impedance path to ground for high frequency signals. A DC bias voltage for the gate of transistor Q5 is provided at BIAS5 and through a resistor R27, for example, a 15Ω resistor. A capacitor C27, for example, a 47 nF capacitor, provides a low impedance path to ground for high frequency signals. A DC bias voltage for the gate of transistor Q6 is provided at BIAS6 and through a resistor R28, for example, a 15Ω resistor. A capacitor C28, for example, a 47 nF capacitor, provides a low impedance path to ground for high frequency signals.

Feedback signals are provided from the common mode choke T2 via coil L3 through resistors R27 and R28. It is to be appreciated that negative feedback could be taken directly from the drains of the transistors Q**3**-Q**6**, which would require physically larger dissipation resistors R**27**, R**28**. In an embodiment, resistors R**27** and R**28** are each formed by two paralleled 430 ohm resistors.

Resistors R**29**-R**34** are balancing resistors for absorbing amplitude imbalance between the signal splitter outputs due to production tolerances. The resistors R**29**-R**34** also help to maintain correct input port impedance, contributing to a low voltage standing wave ratio (VSWR). In an embodiment, resistors R**29** and R**32** are each formed by two paralleled 51 ohm resistors, and resistors R**30**, R**31**, R**33**, and R**34** are each formed by two paralleled 15 ohm resistors.

Resistors R**35**-R**38** provide discharge paths for the charge on the gates when DC bias is removed, and provide solid ground references for the DC bias voltages for the gates. In an embodiment, resistors R**35**-R**38** each have a value of 10 kΩ.

The drains of transistors Q3 and Q4 are interconnected or directly connected together and are directly connected to an input **26** of transformer T**3**. The drains of transistors Q**5** and Q**6** are also interconnected or directly connected together and are directly connected to an input **25** of transformer T3. The "push" signal from transistors Q3 and Q4 are combined with the "pull" signal from transistors Q5 and Q6 by transformer T**3**.

FIG. **4** shows temperature compensating bias circuitry **40** for DC biasing the transistors that form the push-pull amplifiers discussed herein. In the amplifier shown in FIG. **2**, the DC bias voltage from the temperature compensating bias circuitry **40** would be provided at BIAS**1** and BIAS**2**. In the amplifier shown in FIG. **3**, the DC bias voltage from the temperature compensating bias circuitry **40** would be provided at BIAS3 through BIAS**6**.

The temperature of the transistors in the push-pull amplifiers will increase during operation of the amplifiers. To maintain a linear transistor operation at a constant idle point (bias current level), the temperature compensating bias circuitry **40** reduces the bias voltage of the transistors as their temperature increases. In an embodiment, the temperature compensating bias circuitry **40** is designed to reduce the bias voltage at a rate of 6.35 mV/° C.

The temperature compensating bias circuitry **40** includes a temperature sensor Q**41**. The output of the temperature sensor Q**41** changes with temperature. The temperature sensor Q**41** can be mounted so as to sense the temperature of the transistors in the push-pull amplifier. For example, the temperature sensor Q**41** can be attached to a heat sink for the transistors. In FIG. **4**, the temperature sensor Q**41** is a transistor and the temperature sensor provides a decreasing temperature signal level (e.g., −2.1 mV/° C.) as temperature increases. Alternatively, the temperature sensor could provide an increasing temperature signal level (e.g., 10 mV/° C.) as temperature increases. An example temperature sensor is a TIP31C power transistor. Another example temperature sensor is a temperature sensor integrated circuit (IC), such as an LM35 sensor from NATIONAL SEMICONDUCTOR®. It is to be appreciated that other types of temperature sensors could be incorporated into the temperature compensating bias circuitry **40**, such as resistive temperature devices (RTDs) or thermistors for example, and that the temperature compensating bias circuitry can include a plurality of temperature sensors mounted at different locations on or around the push-pull amplifier.

The temperature compensating bias circuitry **40** further includes a first amplifier U**41**, a second amplifier U**42**, a plurality of resistors R**41**-R**51**, and a potentiometer R**52**. The first and second amplifiers U**41**, U**42** can be operational amplifiers, such as the LM324A op-amp. The wiper of the potentiometer R**52** provides a bias output to supply the temperature compensated DC bias voltage to the transistors of the push-pull amplifier (not shown in FIG. **4**). In an embodiment, the potentiometer R**52** is a 5 kΩ potentiometer. Multiple potentiometers can be paralleled to provide individual bias control to each transistor in the push-pull amplifier. The first amplifier U**41** is configured to provide a first temperature dependent voltage to the potentiometer R**52** based on the temperature signal from the temperature sensor Q**41**. The second amplifier U**42** is configured to provide a second temperature dependent voltage based on the temperature signal to the potentiometer R**52** through resistor R**46**. The first temperature dependent voltage and the second temperature dependent voltage change at substantially the same rate in response to the temperature signal. For example, both of the first temperature dependent voltage and the second temperature dependent voltage decrease at a rate of 6.35 mV/° C. Although the outputs of the amplifiers U**41**, U**42** change with increasing temperature, the voltage across the potentiometer remains substantially constant, which allows the full 6.35 mV/° C. voltage reduction to be provided, at any physical setting of the potentiometer, as reduced bias voltage to the push-pull amplifier's transistors (rather than a voltage divided percentage of 6.35 mV/° C. as established by the wiper setting of the potentiometer R**52**). An example first temperature dependent voltage from the first amplifier U**41** is 5.0 volts at 25° C. minus 6.35 mV/° C. An example second temperature dependent voltage from the second amplifier U**42** is 2.0 volts at 25° C. minus 6.35 mV/° C.

An example value for resistors R**41**, R**42**, R**45** and R**48** is 10.0 kΩ. An example value for resistor R**43** is 16.6 kΩ. An example value for resistor R**44** is 24.9 kΩ. An example value for resistor R**46** is 1 kΩ. An example value for resistor R**47** is 430 kΩ. An example value for resistor R**49** is 4.99 kΩ. An example value for resistor R**50** is 2 kΩ. An example value for resistor R**51** is 3.3 kΩ. A 5V input is shown as provided to resistors R**43**, R**47** and R**50**. The 5V input can be a programmable reference voltage and is not limited to 5V. In an embodiment, the 5V input is provided by an adjustable precision voltage reference, such as an LM385 available from NATIONAL SEMICONDUCTOR®.

FIG. **5** shows a further example of temperature compensating bias circuitry **50** for DC biasing the transistors that form the push-pull amplifiers discussed herein. The bias circuitry **50** includes an adjustable precision voltage reference D**51** and a temperature sensor Q**51**. With a 12 VDC supply **56** as shown, the bias circuitry is configured to supply a substantially constant 2.5 VDC across potentiometers R**52***a* and R**52***b*, while the first temperature dependent voltage **53** is 5.0 volts at 25° C. minus 6.35 mV/° C. and the second temperature voltage **54** is 2.0 volts at 25° C. minus 6.35 mV/° C. Supply 56 voltages other than 12 VDC (e.g., 8 VDC) could be used by the bias circuitry **50**. The bias circuitry **50** includes two bias outputs BIAS1, BIAS2 for supplying the temperature compensated DC bias voltage to the transistors of the push-pull amplifier (not shown in FIG. **5**). Example component values for resistors and capacitors in the bias circuitry **50** are shown in the figure. In FIG. **5**, amplifiers U**41** and U**42** feed transistors Q**53** and Q**54**, respectively, rather than feeding the potentiometers R**52***a* and R**52***b* directly. Transistors Q**53** and Q**54** are configured as followers and allow the amplifiers U**41** and U**42** to supply the necessary current when multiple potentiometers are used. An optional diode D**52** is provided at the output of amplifier U**42** to compensate for any ambient temperature-induced changes in the operation of the transistors Q**53** and Q**54**. Multiple diodes can be used instead of a single diode D**52** as shown, to compensate for ambient temperature-induced changes in the operation of other components, such as changes in the operation of the diodes located between the potentiometers R**52***a*, R**52***b* and the bias outputs BIAS**1**, BIAS**2**.

The temperature compensating bias circuitry **50** of FIG. **5** includes an amplifier U**51** that generates a temperature output signal **55**. The temperature output signal **55** transmits the temperature of the RF power amplifier module and can be monitored by a remote device and/or used to display a current temperature of the RF power amplifier module. The bias circuitry **50** can be configured so that the temperature output signal **55** increases at a constant rate, such as 50 mV/° C., or decreases at a constant rate over a voltage range (e.g., 0-5 VDC) corresponding to a temperature range (e.g., 2° C.-102° C.). The bias circuitry **50** can further be configured so that the temperature output signal **55** increases or decreases at a constant rate per bit of an analog to digital (A/D) converter (e.g., 1° F. per bit for an 8 bit ND). In FIG. **5**, the temperature output signal **55** is configured to be 5 VDC at 11.5° F. and 0 VDC at 266.5° F., which is 1° F. per bit for an 8 bit ND. It is to be appreciated that the amplifier U**51** can be configured to generate a positive going signal suitable for display on an analog meter.

The temperature compensating bias circuitry **50** of FIG. **5** further includes an amplifier U**52** (e.g., an operational amplifier) configured as a comparator that controls activations of transistor Q**52** based on the temperature signal from the temperature sensor Q**51**. When the temperature exceeds a set point or limit established by the voltage divider consisting of the 18K and 11.5K resistors, the comparator output causes the transistor Q**52** to conduct. This generates a high temperature alarm on a T LIMIT line **57**, to signal an off board controller (not shown) that a fault condition exists and that corrective action needs to be taken. A TRANSISTOR DUMP line **58** allows the off board controller to deactivate the transistors of the push-pull amplifier. The off board controller can deactivate the transistors of the push-pull amplifier based on monitored parameters, such as the temperature of the amplifier module. In the event of any fault, it is desirable for the RF transistors to stop amplifying as soon as possible. One way to do this is to shut the transistors off by removing bias and discharging the gates as soon as possible. Activation of a dump transistor Q**55** pulls the gates of the push-pull amplifier to ground through 10Ω resistors, which limit the discharge current to a safe level that will not damage transistor Q**55**.

It is to be appreciated that the use of quad op-amp ICs in the temperature compensating bias circuitry **50** will provide amplifiers U**41**, U**42**, U**51** and U**52** in a single package. The marginal cost increase of a quad op-amp IC versus a dual op-amp IC is typically small. Therefore, although amplifiers U**51** and U**52** are optional elements of the temperature compensating bias circuitry **50**, their addition should not increase the cost of the circuitry **50** very much, rendering the temperature output and the temperature alarm signals inexpensive to generate.

FIG. **9** shows a further example of temperature compensating bias circuitry **90** for DC biasing the transistors that form the push-pull amplifiers discussed herein. The bias circuitry **90** in FIG. **9** includes a temperature sensor IC **91** rather than using a transistor as a temperature sensor. The temperature sensor IC **91** provides an increasing temperature signal level as the temperature of the RF module increases (e.g., 10 mV/° C.). The bias circuitry **90** processes the increasing temperature signal level from the temperature sensor IC **91** to provide the DC bias outputs BIAS1 and BIAS2 for the transistors in the push-pull amplifiers (not shown), the temperature output signal **55**, and the high temperature alarm **57**. As compared to the bias circuitry **50** shown in FIG. **5**, transistors Q**53** and Q**54** are omitted from the bias circuitry **90** shown in FIG. **9**. However, the wipers of the potentiometers R**52***a*, R**52***b* are connected to transistors Q**92** and Q**93**, which function as followers and generate the DC bias outputs BIAS**1**, BIAS**2**. The emitters of the follower transistors Q**92**, Q**93** are grounded through a low value resistor to provide a low impedance source characteristic that provides low IMD. Further, in the bias circuitry **90** of FIG. **9**, the dump transistor Q**55** is connected to the "+" terminal of amplifier U**41**. When the dump transistor Q**55** is activated by the TRANSISTOR DUMP line **58**, it essentially forces amplifier U**41** to reduce its output to 0V, causing the voltage at the bias outputs to drop to a very low value below the gate threshold voltage of the push-pull amplifier transistors. The gates of the RF MOSFETs in the push-pull amplifiers then discharge through gate resistors (e.g., see resistors R**6** and R**7** in FIG. **2**). Because the transistor followers allow a very large amount of current to be applied to the RF transistor gate resistors, the gate resistors can be changed to a much lower value (example from 10 KΩ to 470Ω or even 100Ω) to decrease the discharge time of the RF transistor gates. The added benefit of the lower value gate resistor is an improvement in the IMD of the amplifier module.

FIG. **6** and FIG. **8** show modifications of the push-pull power amplifier module **21** of FIG. **2**. Various components shown in FIG. **6** and FIG. **8** are discussed above with respect to FIG. **2**. Such components are referenced in FIG. **6** and FIG. **8** by identical reference characters as used in FIG. **2** and are not discussed in detail below.

In FIG. **6**, the DC-blocking capacitors C**12**, C**13** are replaced with DC-blocking transformers T**61** and T**62**, and the common mode choke T**2** is omitted. The DC-blocking capacitors C**12**, C**13** can be a source of failure in the power amplifier module **21** when they are subject to very high RF currents. Performance and reliability of the power amplifier module **21** can be improved by using DC-blocking transformers rather than capacitors.

The DC-blocking transformers T**61**, T**62** are respectively connected between the drains of transistors Q**1** and Q**2** and the balanced inputs **25**, **26** of the broadband output transformer T**3**. A negative feedback signal from the DC feed structure is provided by the DC-blocking transformers T**61**, T**62** as shown. The negative feedback can be provided by a single turn of wire through the centers of the cores of the DC-blocking transformers T**61**, T**62**. See FIG. **7**, which provides a schematic diagram of the DC-blocking transformers T**61**, T**62**.

Optional matching coils LM**1** and LM**2** can be located between the DC-blocking transformers T**61**, T**62** and the drains of the transistors Q**1**, Q**2**, and optional capacitors C**61** and C**62** can be located across the respective inputs of the DC-blocking transformers T**61**, T**62**. A bypass capacitor C**63** can be located at the DC power source. Optional capacitor C**6** provides further bypassing and can help improve balance of the system.

FIG. **7** schematically shows the construction of the DC-blocking transformers T**61**, T**62**. The DC-blocking transformers can be formed by winding coaxial cables **71**, **72** on respective cores **73**, **74**. In an embodiment, the coaxial cables have a characteristic impedance of 25 ohms, which is twice the drain-drain differential impedance of the transistors Q**1**, Q**2**. Connections to the center and shield conductors of the coaxial cables **71**, **72** are shown in FIG. **7**. The center and shield conductors are further identified in FIG. **6** by "C" and "S", respectively. The negative feedback loop **75** is also shown in FIG. **7**. It is to be appreciated that the center and shield connections can be transposed, and that the construction is not limited to coaxial cable (e.g., twisted wire pairs, paralleled wire, etc. can also be used). It is to be further appreciated that power and current handling capability of said 25Ω cable can be greatly increased by constructing the 25Ω cable by paralleling two 50Ω cables, preferably connecting the shield of one to the center of the other.

The input side of the DC-blocking transformers T**61**, T**62** is identified as "Balanced RF In" and the output side is identified as "Balanced RF Out." The shields of the coaxial cables **71**, **72** are connected to ground at the input side and, at the output side, provide the Balanced RF Out. The center conductors are connected to the drains of the transistors Q**1**, Q**2** through optional matching coils LM**1**, LM**2** (see FIG. **6**). The optional capacitors C**61**, C**62** (see FIG. **6**) are each connected between a center conductor and shield at the input side of the DC-blocking transformers T**61**, T**62**. The DC power source for the transistors Q**1**, Q**2** is connected to the center conductors of the cables **71**, **72** at the output side of the transformers T**61**, T**62**. The shields of the cables **71**, **72** provide the RF output signal to the broadband output transformer T**3** (see FIG. **6**) at the output side of the transformers T**61**, T**62**.

It is to be appreciated that the DC-blocking transformers T**61**, T**62** can be constructed from paralleled or twisted wire if desired.

FIG. **8** shows the push-pull power amplifier module **21** of FIG. **6** with the addition of transformer T**2***a*. Transformer T**2***a* generates the negative feedback signal for the transistors Q**1**, Q**2**, instead of the DC-blocking transformers T**61**, T**62**. Transformer T**2***a* can be connected on either side of the optional matching coils LM**1** and LM**2**. In an embodiment, the primary of transformer T**2***a* is a single turn loop. Optional capacitors C**81** and C**82** are also shown connected on either side of the output balun transformer T**4**.

In a configuration differing from FIG. **6**, FIG. **10** shows the DC-blocking transformers connected to a 12.5 ohm characteristic impedance, 1:1 impedance ratio balun T**101**, followed by a 1:4 impedance ratio unbalanced-to-unbalanced transformer T**102**. The balun T**101** is a broadband output transformer. Transformers T**3** and T**4** are eliminated in FIG. **10**.

As discussed above, the amplifier modules described herein can be combined to create higher-powered amplifiers. It is to be appreciated that the combiners used to combine the outputs of the amplifier modules can have a variety of output impedances, such as 12.5 ohm, 25 ohm, 50 ohm, 100 ohm, etc. The output impedance of the combiners can further be matched to a desired characteristic impedance, such as 50 ohm, using a transformer that provides the proper impedance matching.

FIG. **11** and FIG. **12** both show two amplifier modules being combined to create a higher-powered amplifier. A splitter **103** splits an input signal to two push-pull amplifier modules. The amplifier modules in FIG. **11** include DC-blocking capacitors C**12**, C**13**, whereas the amplifier modules in FIG. **12** include DC-blocking transformers T**61**, T**62**. In FIG. **11** and FIG. **12**, the output of each push-pull transistor pair in a module is fed to an output balun transformer T**104**, T**105**. The output balun transformers T**104**, T**105** can have an input-to-output impedance ratio of 1:1. The output balun transformers provide an output impedance of 25 ohms for each amplifier module. The output balun transformers T**104**, T**105** in FIG. **11** supply a combiner **106**. The combiner **106** combines the outputs from the amplifier modules into a combined signal. The combiner **106** includes a type 2 hybrid 0° 2-way combiner **107**, which supplies an unbalanced-to-unbalanced output transformer **108** having an input-to-output impedance ratio of 1:4. The output impedance of the type 2 hybrid 0° 2-way combiner **107** is 12.5 ohms, which is stepped up to a 50 ohm output impedance by the unbalanced-to-unbalanced output transformer **108**. The type 2 hybrid 0° 2-way combiner **107** supplies the combined signal to the unbalanced-to-unbalanced output transformer **108**.

The combiner **109** in FIG. **12** is similar to the combiner in FIG. **11**, except that the combiner **109** includes a type 1 0° hybrid combiner.

Example applications with which the disclosed modular RF power amplifier system may be used include radio communications, such as amateur radio communications, military radio communications, marine radio communications (e.g., ship to shore), high frequency radio telephone communications, and short wave radio broadcast stations.

It should be evident that this disclosure is by way of example and that various changes may be made by adding, modifying or eliminating details without departing from the fair scope of the teaching contained in this disclosure. The invention is therefore not limited to particular details of this disclosure except to the extent that the following claims are necessarily so limited.

What is claimed is:

1. An RF power amplifier, comprising:
- a first field effect transistor:
  - having a first gate, a first source, and a first drain,
  - having an output power rating of at least 200 watts, and
  - operating with a drain-to-source voltage that is greater than 50 VDC;
- a second field effect transistor:
  - having a second gate, a second source, and a second drain,
  - having an output power rating of at least 200 watts, and
  - operating with a drain-to-source voltage that is greater than 50 VDC;
  - wherein said transistors are configured as a push-pull amplifier;
- an RF signal input;
- an input transformer connected to the RF signal input, the input transformer having respective balanced outputs connected to the first gate and the second gate;
- a broadband output transformer having a first balanced input connected to the first drain, and a second balanced input connected to the second drain, wherein the broadband output transformer has an input to output impedance ratio of 1:4, and further wherein at least some flux cancellation occurs within the broadband output transformer; and
- temperature compensating bias circuitry for providing a temperature compensated bias voltage to the first field effect transistor and the second field effect transistor for decreasing the bias voltage of the first field effect transistor and the second field effect transistor as transistor temperature increases, the temperature compensating bias circuitry comprising:
  - a temperature sensor generating a temperature signal;
  - a first amplifier having an output providing a first temperature dependent voltage based on the temperature signal;
  - a second amplifier having an output providing a second temperature dependent voltage based on the temperature signal, wherein the first temperature dependent voltage and the second temperature dependent voltage change at substantially the same rate in response to the temperature signal; and
  - a potentiometer connected to the output of the first amplifier and the output of the second amplifier such that a voltage across the potentiometer remains sub-

stantially constant when the first temperature dependent voltage and the second temperature dependent voltage change; and

a bias output connected to at least one of the first field effect transistor and the second field effect transistor and supplying the temperature compensated bias voltage to the at least one of the first field effect transistor and the second field effect transistor.

2. A modular RF power amplifier system, comprising:

a first amplifier module comprising the RF power amplifier as set forth in claim **1**; and

a second amplifier module comprising the RF power amplifier as set forth in claim **1**;

wherein outputs from said modules are combined to form a combined RF output.

3. The RF power amplifier of claim **1**, further comprising:

an RF output terminal;

an output balun transformer; and

a DC-blocking capacitor, wherein the output balun transformer and the DC-blocking capacitor are connected in series between the broadband output transformer and the RF output terminal,

wherein the broadband output transformer is a transmission line transformer including a cable having a characteristic impedance of approximately or exactly 25 ohms,

wherein the broadband output transformer has balanced outputs and during operation of the RF power amplifier a net flux within the broadband output transformer is substantially zero, and

wherein the output balun transformer has an input to output impedance ratio of 1:1 and a characteristic impedance of approximately or exactly 50 ohms.

4. An RF power amplifier, comprising:

a first plurality of field effect transistors having directly interconnected drains and respective output power ratings of at least 100 watts;

a second plurality of field effect transistors having directly interconnected drains and respective output power ratings of at least 100 watts;

wherein said transistors operate with a drain-to-source voltage that is greater than 50 VDC, and

wherein the first plurality of field effect transistors and the second plurality of field effect transistors together form a push-pull amplifier having an output power rating of at least 400 watts;

an RF signal input;

an input transformer connected to the RF signal input, the input transformer having respective balanced outputs connected to gates of the first plurality of field effect transistors and gates of the second plurality of field effect transistors; and

a broadband output transformer having a first balanced input connected to the drains of the first plurality of field effect transistors, and a second balanced input connected to the drains of the second plurality of field effect transistors, wherein the broadband output transformer has an input to output impedance ratio of 1:4, and further wherein at least some flux cancellation occurs within the broadband output transformer; and

temperature compensating bias circuitry for providing a temperature compensated bias voltage to the first plurality of field effect transistors and the second plurality of field effect transistors for decreasing the bias voltage of the first plurality of field effect transistors and the second plurality of field effect transistors as transistor temperature increases, the temperature compensating bias circuitry comprising:

a temperature sensor generating a temperature signal;

a first amplifier having an output providing a first temperature dependent voltage based on the temperature signal;

a second amplifier having an output providing a second temperature dependent voltage based on the temperature signal, wherein the first temperature dependent voltage and the second temperature dependent voltage change at substantially the same rate in response to the temperature signal; and

a potentiometer connected to the output of the first amplifier and the output of the second amplifier such that a voltage across the potentiometer remains substantially constant when the first temperature dependent voltage and the second temperature dependent voltage change; and

a bias output connected to at least one of the first plurality of field effect transistors and the second plurality of field effect transistors and supplying the temperature compensated bias voltage to the at least one of the first plurality of field effect transistors and the second plurality of field effect transistors.

5. A modular RF power amplifier system, comprising:

a first amplifier module comprising the RF power amplifier as set forth in claim **4**; and

a second amplifier module comprising the RF power amplifier as set forth in claim **4**, wherein outputs from said modules are combined to form a combined RF output.

6. The RF power amplifier of claim **4**, further comprising:

an RF output terminal;

an output balun transformer; and

a DC-blocking capacitor, wherein the output balun transformer and the DC-blocking capacitor are connected in series between the broadband output transformer and the RF output terminal,

wherein the broadband output transformer is a transmission line transformer including a cable having a characteristic impedance of approximately or exactly 25 ohms,

wherein the broadband output transformer has balanced outputs and during operation of the RF power amplifier a net flux within the broadband output transformer is substantially zero, and

wherein the output balun transformer has an input to output impedance ratio of 1:1 and a characteristic impedance of approximately or exactly 50 ohms.

7. An RF power amplifier, comprising:

a push-pull amplifier, the push-pull amplifier comprising a first field effect transistor and a second field effect transistor; and

temperature compensating bias circuitry for providing a temperature compensated bias voltage to the first field effect transistor and the second field effect transistor for decreasing the bias voltage of the first field effect transistor and the second field effect transistor as transistor temperature increases, the temperature compensating bias circuitry comprising:

a temperature sensor generating a temperature signal;

a first amplifier having an output providing a first temperature dependent voltage based on the temperature signal; and

a potentiometer connected to the output of the first amplifier such that a voltage across the potentiometer remains substantially constant when the first temperature dependent voltage changes; and

a bias output connected to at least one of the first field effect transistor and the second field effect transistor and supplying the temperature compensated bias voltage to the at least one of the first field effect transistor and the second field effect transistor.

8. The RF power amplifier of claim **7**, wherein the temperature compensating bias circuitry further comprises a second amplifier having an output providing a second temperature dependent voltage based on the temperature signal,

wherein the first temperature dependent voltage and the second temperature dependent voltage change at substantially the same rate in response to the temperature signal, and

wherein the potentiometer is connected to the output of the first amplifier and the output of the second amplifier such that a voltage across the potentiometer remains substantially constant when the first temperature dependent voltage and the second temperature dependent voltage change.

9. The RF power amplifier of claim **8**, further comprising a follower transistor connected to the potentiometer, wherein the follower transistor generates the bias output.

10. The RF power amplifier of claim **7**, wherein the temperature sensor is a transistor.

11. The RF power amplifier of claim **7**, wherein the temperature sensor comprises a temperature sensor integrated circuit.

12. The RF power amplifier of claim **7**, further comprising a temperature output signal transmitting a temperature of the RF power amplifier.

13. The RF power amplifier of claim **7**, further comprising a transistor dump signal for deactivating at least one of the first field effect transistor and the second field effect transistor.

14. The RF power amplifier of claim **7**, wherein the first field effect transistor includes a first gate, a first source, and a first drain,

wherein the second field effect transistor includes a second gate, a second source, and a second drain, the RF power amplifier further comprising:

a broadband output transformer having a first balanced input and a second balanced input, wherein the broadband output transformer has an input to output impedance ratio of 1:4, and further wherein at least some flux cancellation occurs within the broadband output transformer;

a first DC-blocking transformer connected between the first balanced input and the first drain; and

a second DC-blocking transformer connected between the second balanced input and the second drain.

15. The RF power amplifier of claim **14**, wherein each one of the broadband output transformer, the first DC-blocking transformer, and the second DC-blocking transformer include a cable having a characteristic impedance of approximately or exactly 25 ohms.

16. The RF power amplifier of claim **14**, wherein the first DC-blocking transformer provides a negative feedback signal to the first field effect transistor, and wherein the second DC-blocking transformer provides a negative feedback signal to the second field effect transistor.

17. The RF power amplifier of claim **14**, further comprising another transformer, which provides negative feedback signals to the first field effect transistor and the second field effect transistor.

18. The RF power amplifier of claim **14**, further comprising an RF output terminal; and

an output balun transformer connected between the broadband output transformer and the RF output terminal.

19. An RF power amplifier, comprising:

a push-pull amplifier, the push-pull amplifier comprising a first field effect transistor having a first drain and a second field effect transistor having a second drain;

a broadband output transformer having a first balanced input and a second balanced input;

a first DC-blocking transformer connected between the first balanced input and the first drain; and

a second DC-blocking transformer connected between the second balanced input and the second drain.

20. The RF power amplifier of claim **19**, further comprising temperature compensating bias circuitry for providing a temperature compensated bias voltage to the first field effect transistor and the second field effect transistor for decreasing the bias voltage of the first field effect transistor and the second field effect transistor as transistor temperature increases.

21. The RF power amplifier of claim **19**, wherein each one of the broadband output transformer, the first DC-blocking transformer, and the second DC-blocking transformer include a cable having a characteristic impedance of approximately or exactly 25 ohms.

22. The RF power amplifier of claim **19**, wherein the first DC-blocking transformer provides a negative feedback signal to the first field effect transistor, and wherein the second DC-blocking transformer provides a negative feedback signal to the second field effect transistor.

23. The RF power amplifier of claim **19**, further comprising another transformer, which provides negative feedback signals to the first field effect transistor and the second field effect transistor.

24. The RF power amplifier of claim **19**, wherein the broadband output transformer has an input to output impedance ratio of 1:4, and further wherein at least some flux cancellation occurs within the broadband output transformer.

25. The RF power amplifier of claim **24**, further comprising an RF output terminal; and

an output balun transformer connected between the broadband output transformer and the RF output terminal.

26. The RF power amplifier of claim **19**, wherein the broadband output transformer has unbalanced outputs and an input to output impedance ratio of 1:1.

27. The RF power amplifier of claim **26**, further comprising an RF output terminal; and

an unbalanced-to-unbalanced transformer connected between the broadband output transformer and the RF output terminal.

28. The RF power amplifier of claim **27**, wherein the unbalanced-to-unbalanced transformer has an input to output impedance ratio of 1:4.

29. An RF power amplifier, comprising:

a first amplifier module comprising a first push-pull amplifier including a plurality of field effect transistors and a first output balun transformer, wherein an output impedance of the first amplifier module is 25 ohms;

a second amplifier module comprising a second push-pull amplifier including a plurality of field effect transistors and a second output balun transformer, wherein an output impedance of the second amplifier module is 25 ohms;

a combiner connected to the first amplifier module and the second amplifier module, the combiner comprising an unbalanced-to-unbalanced output transformer having an input-to-output impedance ratio of 1:4, wherein the combiner combines an output from the first amplifier module and an output from the second amplifier module into a combined signal,

wherein the combined signal is supplied to the unbalanced-to-unbalanced output transformer, and

wherein an output impedance of the combiner is 50 ohms.

30. The RF power amplifier of claim **29**, wherein the first amplifier module includes a plurality of DC-blocking transformers, and wherein the second amplifier module includes a plurality of DC-blocking transformers.

* * * * *

US008237313B2

(12) **United States Patent**
Ruocco et al.

(10) **Patent No.:** **US 8,237,313 B2**
(45) **Date of Patent:** **Aug. 7, 2012**

(54) **METHOD AND APPARATUS FOR WIRELESS TRANSMISSION AND RECEPTION OF ELECTRIC POWER**

(76) Inventors: **John Ruocco**, Mastic, NY (US); **Ronald Koppel**, Shirley, NY (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 487 days.

(21) Appl. No.: **12/420,306**

(22) Filed: **Apr. 8, 2009**

(65) **Prior Publication Data**

US 2010/0259111 A1 Oct. 14, 2010

(51) **Int. Cl.**
G05F 3/06 (2006.01)
H01F 38/00 (2006.01)

(52) **U.S. Cl.** **307/151**; 307/104

(58) **Field of Classification Search** 307/104, 307/151

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,689,803	A	8/1987	Johannessen et al.	
5,990,670	A	11/1999	Derks	
7,170,821	B1 *	1/2007	Tonn et al.	367/141
7,656,692	B2 *	2/2010	Sadwick et al.	363/89
7,880,338	B2 *	2/2011	Jin	307/104
7,919,886	B2 *	4/2011	Tanaka	307/104
2006/0063488	A1	3/2006	Hraby et al.	
2008/0157780	A1	7/2008	Beeman	
2009/0102292	A1 *	4/2009	Cook et al.	307/104
2010/0090539	A1	4/2010	Auchterlonie	

* cited by examiner

Primary Examiner — Hal Kaplan

(74) *Attorney, Agent, or Firm* — Ladas & Parry LLP

(57) **ABSTRACT**

An electric transmission system utilizes charge displacement for the wireless delivery of electric power. A transmitter for radiating a charge displacement current at a frequency of approximately 2 to 7 kHz to a receiver located remote from the transmitter. The receiver receives the charge displacement current and converts the current into a form of electric power usable by an electrical device coupled to the receiver. The transmitter may use a step up transformer with an input approximating a series of step function pulses at the intended transmission frequency to produce the required high voltage for transmission. The primary winding of the transformer is part of a tuned circuit having a resonant frequency higher than the frequency of the charge displacement current radiated.

12 Claims, 2 Drawing Sheets

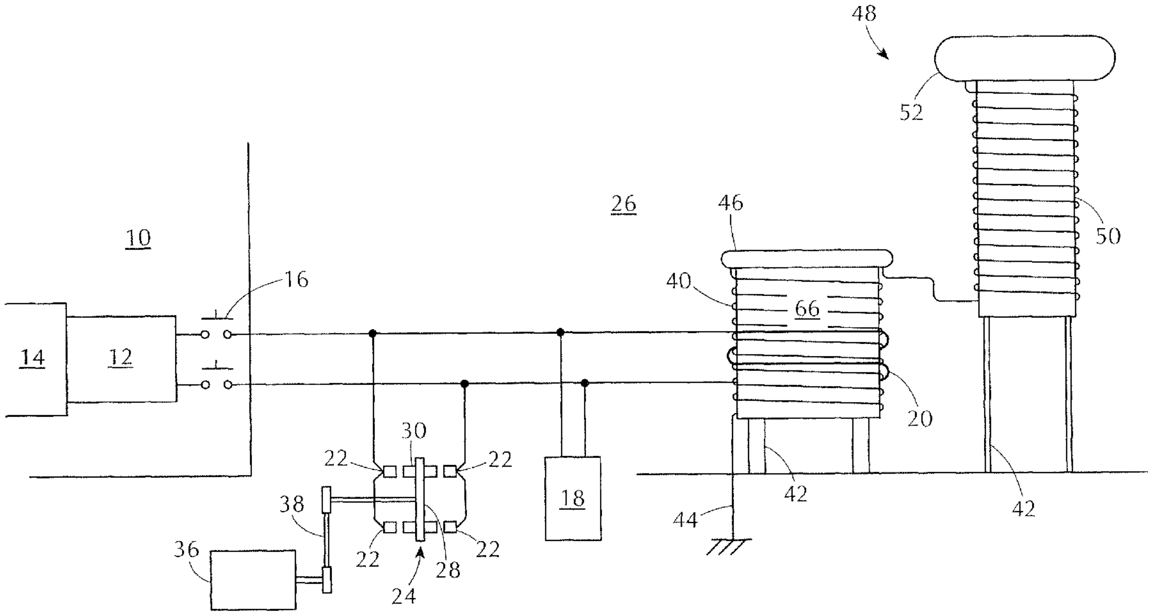

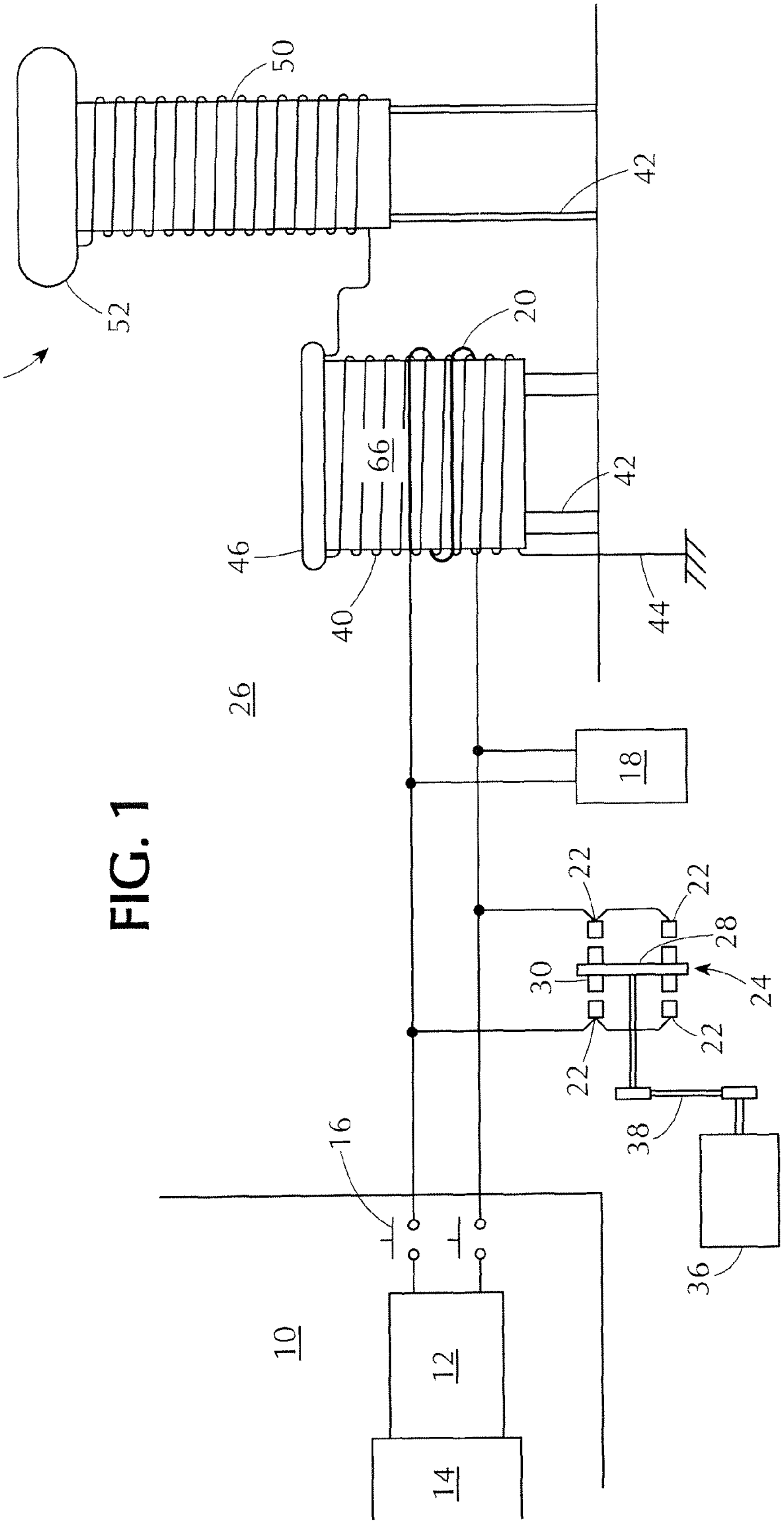

FIG. 1

FIG. 2

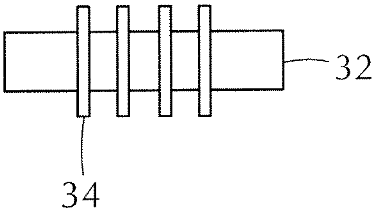

FIG. 3

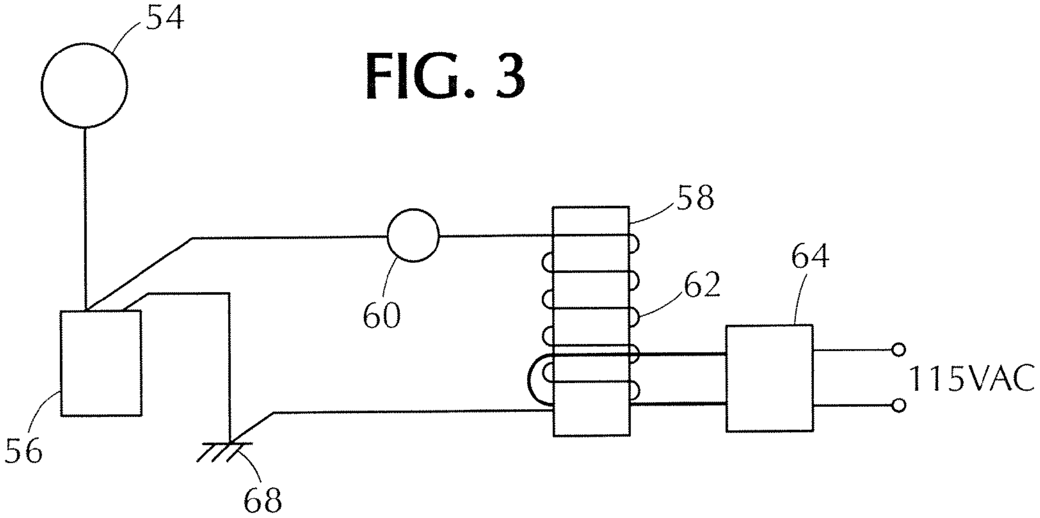
54
58
60
62
64
115VAC
56
68

METHOD AND APPARATUS FOR WIRELESS TRANSMISSION AND RECEPTION OF ELECTRIC POWER

The present invention relates to the field of electric energy transmission and distribution and specifically to a method and apparatus to deliver electric power to users wirelessly,

BACKGROUND OF THE INVENTION

Conventional electric power transmission and distribution to consumers involves a sophisticated system of transmission and distribution wires interconnected by transformers to raise or lower the voltage to appropriate levels during the delivery process. In addition, due to differing loads, capacitors and/or inductors may be switched into or out of the delivery circuits.

The costs to install and maintain the circuits are substantial. When new customers appear, new wiring must be installed to connect them to the transmission and distribution grid. In addition, there can be significant energy losses associated with the delivery, and spurious electromagnetic radiation can be developed which further degrades system performance and can create interference.

It is accordingly a purpose of the present invention to provide an electric power delivery system that avoids deficiencies of wired systems by utilizing a wireless transmission mechanism. Among the characteristics of such a system are a reduction or substantial elimination of electromagnetic radiation, and the ability to connect additional uses to the system at reasonable cost.

BRIEF DESCRIPTION OF THE INVENTION

In accordance with the foregoing and other objects and benefits, an electric power delivery system in accordance with the present invention utilizes the mechanism of charge displacement to transmit electric power from a generation site to a user. The system utilizes extremely low frequencies (ELF) to convey a current and utilizes the earth as a conductor. At the production end of the system conventionally generated electric current is converted into a charge displacement current and transmitted or broadcast wirelessly. A relatively simple receiving apparatus converts the charge displacement current back into a conventional alternating current flow at the location of a user, which current can then be utilized in the standard manner. With appropriate reception equipment, the charge displacement current can be received by moving vehicles, allowing direct receipt of a continuing flow of electrical power during operation.

BRIEF DESCRIPTION OF THE DRAWINGS

A fuller understanding of the invention will be achieved upon consideration of the following detailed illustrative description of a preferred embodiment of the invention when reviewed in conjunction with the annexed drawings, wherein:

FIG. **1**. is an illustration of a transmitter apparatus for generating an ELF charge displacement current in accordance with the invention;

FIG. **2** is a detail view of an interrupter electrode; and

FIG. **3** is an illustration of a receiver apparatus for receiving an ELF charge displacement current.

DETAILED DESCRIPTION OF THE INVENTION

The present invention incorporates the recognition that, based upon the electromagnetic characteristics of the earth and its surrounding ionosphere, an electrical energy transmission at a frequency in the range of about 2 to 7 kHz maximizes displacement charge transfer from a transmitting apparatus to the atmosphere and earth. The invention incorporates a transmitting element in the form of a capacitor plate that is part of a high voltage resonant circuit tuned to a frequency in the desired range to maximize transmitter efficiency to the atmosphere. The transmitting element is part of a transmitter system that converts electrical power, preferably at DC but typically at a low frequency, such as at mains frequency (60 Hz), and at a relatively low voltage, to the desired high voltage and frequency. The DC or mains frequency input power is pulsed at the desired transmission frequency and is applied to a tuned circuit resonant at a second frequency including a primary transformer winding as the tuned circuit's inductor. The transformer raises the voltage to the desired high voltage for transmission. To allow commercially-reasonably sized components to be used, the transformer is adapted to operate at a substantially higher frequency than the transmission frequency, typically at 50-100 kHz. The high voltage is passed to the transmission frequency-tuned resonant circuit including the transmitting element, allowing the 2 to 7 kHz pulses to be coupled to the earth and ionosphere to conduct the desired charge displacement.

With reference to FIG. **1**, generator room **10** houses a generator **12**. The generator is shown driven by a motor **14**, but other sources of motive power, such as a turbine, may be employed. A commercially available AC generator with a 60 Hz output can be utilized, but the lower the output frequency from the generator the lower losses due to mismatching and harmonic generation. While DC output is optimal, 60 Hz is sufficiently low to be acceptable.

The electrical output from the generator is fed to contactor **16**, allowing immediate shutdown of the transmitter to which the output is led in the event of a flashover or other problem in the transmitter room **26**. One line of the generator's output goes to one side of the parallel combination of capacitor **18**, coil **20** and one electrode assembly **22** of a rotary gap apparatus **24**, all located in the transmitter room. The other output line from the generator goes to the other side of capacitor **18** and coil **20** and to a second electrode assembly **22** of the rotary gap apparatus.

Capacitor **18** forms a first resonant or tuned circuit with coil **20**, the primary winding for transformer **66**, and the resonant frequency of which is given by $F=\frac{1}{2}\Pi\sqrt{L1C1}$. An acceptable value for the capacitor may be in the range of 0.27 uf and for the coil **20** an inductance in the range of 3 mH, which can be accommodated in a 6 foot high coil of 10 foot diameter. Capacitor **18** is chosen to have an appropriate working voltage and a low ESR (Equivalent Series Resistance), while coil **20** is of a conductor gauge appropriate for the current involved. The approximate resonant frequency for the listed component values is 75 kHz. The frequency is chosen to allow transformer **40**, of which coil **20** serves as a primary winding, to be of commercially-reasonable dimensions taking into account the current capacity of the turns, the voltage boost contemplated and the need to avoid flashover. As the voltage boost is great, it is generally desired that the number of turns for coil **20** be as low as possible to maintain a large turns ratio with minimal conductor mass.

The generator's output voltage is modulated by means of the rotary gap apparatus **24**, which is in parallel with capacitor **18** and coil **20**, and who's purpose is to generate a pulse train in the desired ELF range of 2-7 kHz. Rotary gap apparatus **24** comprises a nonconductive disc **28** supporting four or more electrodes **30** extending through the disc, evenly spaced along its perimeter. Two identical electrode assemblies **22**, spaced

180 degrees apart and positioned to align with the disc electrodes, are located at opposite sides of the disc, and are respectively connected to the generator output lines. As depicted in FIG. **2**, each electrode assembly **22** may consist of a conductive electrode **32** and a plurality of heat sink rings **34** along the electrode. The electrodes **32** are spaced about ¼ inch from the disc electrodes **30**; all the electrodes may be ½ inch in diameter. A motor **36** and transmission **38** rotates the disc **28** and attached electrodes. As the disc rotates its carried electrodes **30** pass by the electrode assemblies **22**, allowing current to bridge the inter-electrode gap at a frequency of two times the number of electrode **22** pairs. With a disc rotation speed of 7200 rpm and 2 electrode pairs, the modulation frequency is 3600 Hz. The rotary gap apparatus allows a pulse train approximating a series of step impulses to appear across the resonant circuit combination of capacitor **18** and coil **20**. Other means for developing the pulse train, such as an electronic switching circuit, can also be utilized, but must be capable of handling the large current appropriate for significant power transmission.

Coil **20** is wound about the base of a second coil **40** on an insulating form of sufficient diameter and height to allow the coil turns to form the appropriate inductance. The combination of the two coils forms transformer assembly **66** that is supported by insulators **42**. The turns ratio between primary coil **20** and secondary coil **40** is such that the output voltage of the generator is raised to several megavolts in coil **40**. One end of coil **40** is connected to ground, such as by series of grounding rods **44**, while the other (upper) end of the coil is connected to large conductive ring **46** forming a capacitor plate. Ring **46** is constructed and dimensioned to have the appropriate capacitance in free air to form a second resonant or tuned circuit with coil **40** at the same frequency as the resonant frequency of primary coil **20** and capacitor **18**. The inductance of coil **40** may be, for example, 14 mH, which can be accommodated in a coil with a diameter of 5 feet and a height of 10 feet. With the coil size and inductance determined, the dimensions for capacitor ring **46** to achieve the appropriate free air capacitance can be determined by those skilled in the art.

Transmitter assembly **48** consists of the series arrangement of a further coil **50**, likewise supported by insulators **42**, and second conductive ring **52** positioned at the top of the coil and the insulating form upon which it is constructed. The upper end of coil **40** is connected to the lower end of transmitter coil **50**, while conductive ring **52** is connected to the upper end of coil **50**. Ring **52** is constructed with dimensions, including its radius of curvature and its outside diameter, to yield a capacitance for the ring that, when combined with the inductance of coil **50**, create a resonant circuit whose frequency is equal to that of the pulse train generated by the rotary gap apparatus **24**, providing efficient coupling of the pulse train energy to the earth and ionosphere, and should preferably be of large capacitance relative to the capacitance of other system components. The radius of curvature of ring **52** should also be large enough so that the ring's surface electrical charge density, which is proportional to the voltage, capacitance and frequency, is small enough at the chosen high working voltage to prevent spark discharge into the surrounding atmosphere. It is contemplated that a ring of up to 200 feet in diameter may be appropriate. With the desired capacitance of the ring calculated, the inductance for coil **50** can be computed, and the physical characteristics of the coil, including number of turns and wire gauge, determined. As the combination of coil **50** and ring **52** is supported by insulators **42** they must be of sufficient size and height to prevent breakdown, even in damp weather.

Conductive ring **52** serves as the resonating element for coil **40**. Alternatively, the ring may also be constructed with a group of spaced discs around the entire outer circumference of the ring. Such a construction may prevent the buildup of eddy currents that otherwise may occur in the ring. Alternatively a series of small semicircular discs could be attached to the outside surface of the ring to provide the same effect.

It should be noted that the radiating element is not a dipole antenna, which would require a radiating element of dimensions specific to the wavelength of the transmitted frequency. Rather, the mechanism of transmission is charge displacement, not electromagnetic radiation. Thus, it is not necessary for the lines of flux to detach and close upon themselves to propagate an electric field. Instead, the lines of flux exist between two opposite charges, and the field is established by the lines of flux between the two points of charge. The force on each charge in dynes is given by:

$$F = \frac{q1q2}{4\rho e_o r^2} 10^5.$$

A relatively simple antenna and receiver system may be used to convert the radiated charge into useful energy. The antenna is designed as a terminating point for the field's flux lines. Therefore the antenna does not have to be a specific shape or length. A conversion unit is used in conjunction with the antenna to convert the field to normal 60 Hz electrical power, at which most electrical devices operate.

An illustrative antenna and receiver system is depicted in FIG. **3**. An antenna **54** is connected to one side of a large capacitance **56**. The other side of the capacitor is connected to a suitable ground or conductive surface acting as a ground plane **68**. The size of the antenna and ground plane, if not earth, have a bearing on the amount of power that can be extracted from the atmosphere, and thus should be sufficiently large for the intended draw. The receiving antenna is designed as a terminating point for the field's flux lines, thus the antenna does not have to be a specific shape or length. One side of coil **58** is connected to capacitor **56** through an electrical interrupter **60**, which preferably is electronic, rather than mechanical, and as known in the art can be either vacuum tube or solid state. The interrupter is tuned to the same pulse frequency as created by the transmitter's rotary gap apparatus **24**, and the combination of capacitor **56** and coil **58** form a resonant circuit likewise tuned to the pulse frequency. Coil **58** and second coil **62**, wound around coil **58** and consisting of fewer turns of a conductor, form a stepdown transformer, coil **58** having a wire size capable of carrying the induced current. Secondary coil **62** is connected to a conversion unit **64** as known in the art to convert the pulsating output current from the transformer to normal 60 Hz electrical power or to another frequency and voltage as appropriate for the electrical devices to be powered.

In addition to the obvious uses, such as supplying electricity to residential and commercial buildings, the transmission system of the present invention could also be used to power vehicles. The vehicle may be constructed of a non-conductive material such as fiberglass, and may have the collecting antenna located in the roof. A conversion unit would then convert the electrical energy as received to the form of electricity required by the vehicle's drive, typically an electric motor. The received energy could also be used to charge a battery array to provide power in the event the vehicle leaves a reception area.

The transmission system could also be used to power trains. Since a train system defines a given path, one or more transmitting stations could be set up to focus the broadcast power along the rail path. Those skilled in the art can appreciate that the present invention could likewise be used to deliver electric power to other specific locations and users, thus providing individualized delivery for power tools, outdoor appliances and any other electrical devices that use electricity.

We claim:

1. An electric transmission system, comprising a transmitter for radiating a charge displacement current at a frequency of from 2 to 7 kHz and a receiver located remote from the transmitter for receiving the charge displacement current and converting the current into a form of electric power usable by an electrical device coupled to the receiver, the transmitter including means for generating electrical power at a first low frequency, means for pulsing the power at the radiating frequency, transformer means for raising the voltage of the pulsed power to a high voltage for radiating, and a radiating element coupled to an output of the transformer.

2. The electric transmission system of claim **1**, wherein the transformer means includes a primary coil forming a part of a first tuned circuit.

3. The electric transmission system of claim **2**, wherein the first tuned circuit is in parallel with the pulsing means and is tuned to a higher frequency than the radiating frequency.

4. The electric transmission system of claim **3**, wherein the higher frequency is in the range of about 50-100 kHz.

5. The electric transmission system of claim **3**, wherein the transformer means includes a secondary coil forming a part of a second resonant circuit.

6. The electric transmission system of claim **5**, wherein the second tuned circuit is tuned to the same frequency as the first tuned circuit.

7. The electric transmission system of claim **6**, wherein the means for receiving the charge displacement current comprises an, antenna and a ground plane coupled to an interrupter in series with a tuned circuit.

8. The electric transmission system of claim **7**, wherein the tuned circuit includes a coil forming a primary side of a transformer and the converting means comprises a secondary side coil of the transformer coupled to a power converter.

9. The electric transmission system wherein of claim **8** the interrupter operates at the frequency of the charge displacement current and the tuned circuit is tuned to the frequency of the charge displacement current.

10. The electric transmission system of claim **5** wherein a third tuned circuit is tuned to the frequency of the displacement current, is coupled to the secondary coil of the transformer means, and includes the radiating element.

11. The electric transmission system of claim **10**, wherein the radiating element is a capacitor plate of the third tuned circuit.

12. The transmission system of claim **1**, wherein the receiver comprises means for receiving the charge displacement current and means coupled for converting the charge displacement current into electric power of a desired frequency and voltage.

* * * * *

US008686575B2

(12) **United States Patent**
McCowen

(10) **Patent No.: US 8,686,575 B2**
(45) **Date of Patent: Apr. 1, 2014**

(54) **ENERGY COLLECTION**

(75) Inventor: **Clint McCowen**, Navarre, FL (US)

(73) Assignee: **Ion Power Group, LLC**, Navarre, FL (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **12/255,130**

(22) Filed: **Oct. 21, 2008**

(65) **Prior Publication Data**

US 2009/0040680 A1 Feb. 12, 2009

Related U.S. Application Data

(63) Continuation of application No. 11/358,264, filed on Feb. 21, 2006, now Pat. No. 7,439,712.

(51) **Int. Cl.**

F02B 63/04	(2006.01)
F03G 7/08	(2006.01)
H02K 7/18	(2006.01)
H02N 1/00	(2006.01)

(52) **U.S. Cl.**
USPC .... **290/1 R**; 322/2 A

(58) **Field of Classification Search**
USPC .... 290/1 R; 322/2 A; 126/660; 60/641.8, 60/641.11, 641.12, 659, 698; 244/158.2
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

674,427 A		5/1901	Palencsar	307/149
911,260 A		2/1909	Pennock	307/149
1,014,719 A		1/1912	Pennock	307/149
2,473,819 A	*	6/1949	Pittman	315/36
3,532,959 A	*	10/1970	Erickson	320/166
3,780,722 A	*	12/1973	Swet	126/680
3,946,227 A		3/1976	Bingham	250/281
4,079,225 A	*	3/1978	Warner	219/110
4,092,250 A		5/1978	Sano et al.	210/500.29
4,104,696 A		8/1978	Cochran, Jr.	361/229
4,146,800 A	*	3/1979	Gregory et al.	290/44
4,180,056 A	*	12/1979	Schnabel et al.	126/660
4,201,197 A	*	5/1980	Dismer	126/600
4,206,396 A	*	6/1980	Marks	322/2 A
4,224,496 A	*	9/1980	Riordan et al.	219/110
4,307,936 A	*	12/1981	Ochiai	385/120
4,314,192 A	*	2/1982	Caro	322/2 R
4,346,478 A	*	8/1982	Sichling	398/107
4,425,905 A	*	1/1984	Mori	126/578
4,433,248 A	*	2/1984	Marks	290/44
RE31,678 E	*	9/1984	Ochiai	385/115

(Continued)

FOREIGN PATENT DOCUMENTS

WO	2007098341	8/2007

OTHER PUBLICATIONS

"Emissivity Values for Common Materials & Non-Metals", Raytek—A Fluke Company, http://www.raytek.com/Raytek/en-r0/IREducation/EmissivityNonMetals.htm, 1999.*

Primary Examiner — Pedro J Cuevas
(74) *Attorney, Agent, or Firm* — Benjamin A. Balser; Balser & Grell IP Law

(57) **ABSTRACT**

An energy collection system may collect and use the energy generated by an electric field. Collection fibers are suspended from a support wire system supported by poles. The support wire system is electrically connected to a load by a connecting wire. The collection fibers may be made of any conducting material, but carbon and graphite are preferred. Diodes may be used to restrict the backflow or loss of energy.

25 Claims, 10 Drawing Sheets

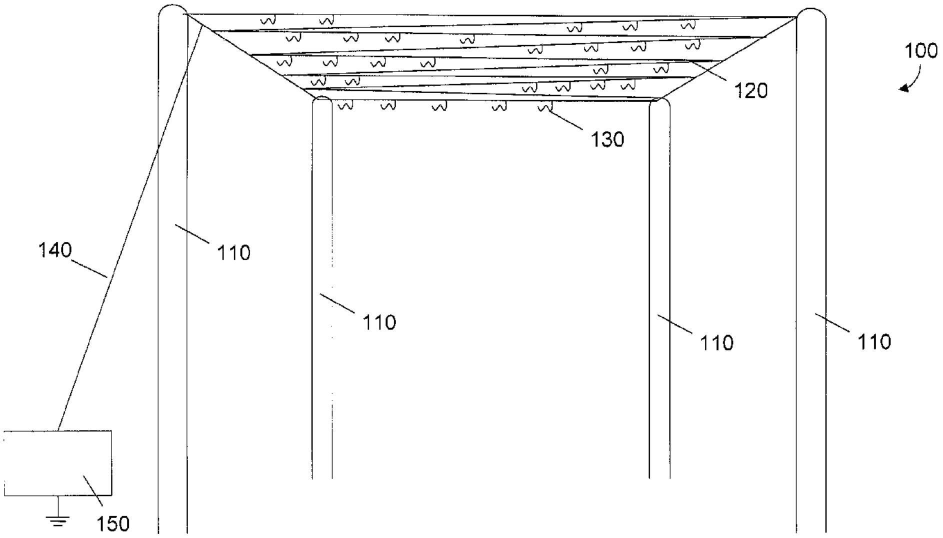

(56) References Cited

U.S. PATENT DOCUMENTS

4,483,311	A *	11/1984	Whitaker	126/602
4,489,269	A *	12/1984	Edling et al.	322/2 R
4,512,335	A *	4/1985	Mori	126/678
4,653,223	A *	3/1987	Mori	47/17
4,653,472	A *	3/1987	Mori	126/683
4,676,226	A *	6/1987	Mori	126/680
4,809,675	A *	3/1989	Mori	126/578
4,852,454	A *	8/1989	Batchelder	89/1.11
4,943,125	A *	7/1990	Laundre' et al.	385/35
4,988,159	A *	1/1991	Turner et al.	385/33
5,047,892	A	9/1991	Sakata et al.	361/231
5,114,101	A *	5/1992	Stern et al.	244/172.8
5,145,257	A *	9/1992	Bryant et al.	374/131
5,379,103	A *	1/1995	Zigler	356/73
5,851,309	A *	12/1998	Kousa	136/248
5,942,806	A *	8/1999	Veliadis	290/1 R
6,025,591	A	2/2000	Taylor et al.	250/292
6,038,363	A *	3/2000	Slater et al.	385/147
6,116,544	A *	9/2000	Forward et al.	244/158.2
6,173,922	B1 *	1/2001	Hoyt et al.	244/158.2
6,226,440	B1 *	5/2001	Lyons	385/119
6,419,191	B1 *	7/2002	Hoyt et al.	244/158.2
6,425,391	B1 *	7/2002	Davoren et al.	126/683
6,735,830	B1	5/2004	Merciel	250/324
6,765,212	B2 *	7/2004	Goetz et al.	250/339.11
6,846,447	B2 *	1/2005	Mbachu et al.	264/406
6,853,447	B2 *	2/2005	Goetz	356/237.1
6,894,772	B2 *	5/2005	Goetz et al.	356/237.1
6,925,852	B2 *	8/2005	Susko	73/23.2
6,974,110	B2	12/2005	Grandics	244/171.5
7,109,597	B1 *	9/2006	Bose	290/43
7,128,867	B2 *	10/2006	Mbachu et al.	264/406
7,231,809	B2 *	6/2007	Susko	73/31.05
7,259,475	B2 *	8/2007	Hong et al.	307/64
7,439,712	B2 *	10/2008	McCowen	322/2 A
7,449,668	B2 *	11/2008	Schutten et al.	250/214 SW
7,478,712	B2 *	1/2009	McCowen	191/45 R
7,532,819	B1 *	5/2009	Triebes et al.	398/121
2002/0026933	A1 *	3/2002	Gottlieb	126/628
2002/0108892	A1 *	8/2002	Goetz et al.	209/576
2002/0109094	A1 *	8/2002	Goetz et al.	250/339.11
2002/0109835	A1 *	8/2002	Goetz	356/237.1
2003/0107869	A1	6/2003	Pan et al.	361/306.2
2003/0125630	A1 *	7/2003	Furnish	600/461
2003/0193319	A1 *	10/2003	Wood et al.	322/2 A
2004/0011925	A1 *	1/2004	Grandics	244/62
2004/0083793	A1 *	5/2004	Susko	73/31.05
2004/0094853	A1 *	5/2004	Mbachu et al.	264/40.1
2004/0160711	A1	8/2004	Stumberger	361/20
2005/0101024	A1 *	5/2005	Mbachu et al.	436/85
2005/0136311	A1 *	6/2005	Ueda et al.	429/30
2005/0270525	A1 *	12/2005	Susko	356/326
2007/0107765	A1 *	5/2007	Schutten et al.	136/243
2007/0195481	A1 *	8/2007	McCowen	361/233
2007/0218323	A1 *	9/2007	Sudo et al.	429/13
2007/0223171	A1	9/2007	Guy et al.	361/126
2007/0273206	A1 *	11/2007	McCowen	307/9.1
2007/0297892	A1 *	12/2007	Kildegaard	415/1
2008/0145714	A1 *	6/2008	Kagami	429/13
2009/0107842	A1	4/2009	Park et al.	204/554
2009/0114495	A1 *	5/2009	McCowen	191/45 R
2009/0216292	A1 *	8/2009	Pless et al.	607/33
2010/0090562	A1 *	4/2010	McCowen	310/309
2010/0090563	A1 *	4/2010	McCowen	310/309
2012/0164555	A1 *	6/2012	Yamazaki et al.	429/482
2012/0299559	A1 *	11/2012	McCowen	322/2 A
2013/0276776	A1 *	10/2013	Te Kamp et al.	126/660

* cited by examiner

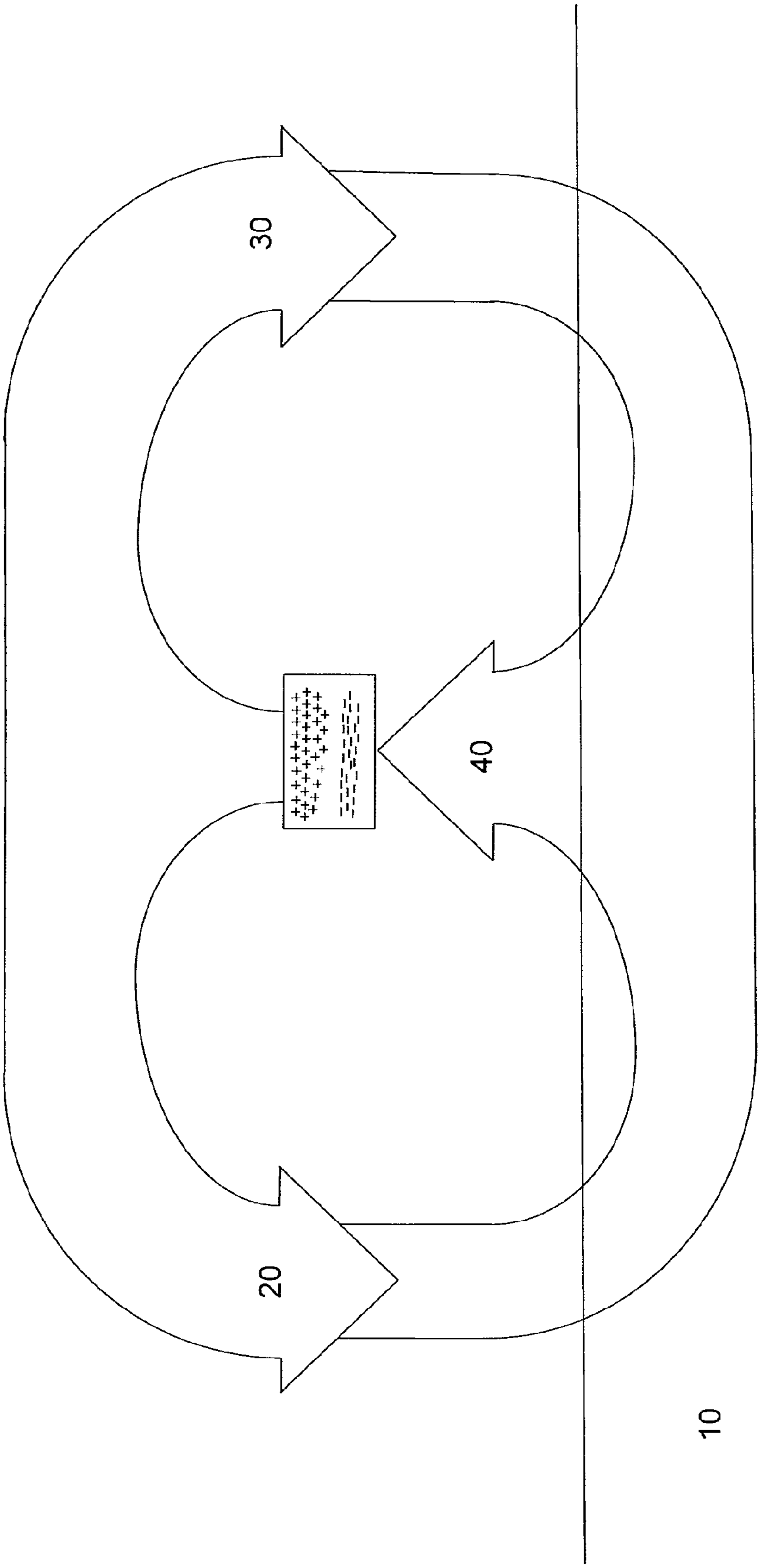

FIGURE 1

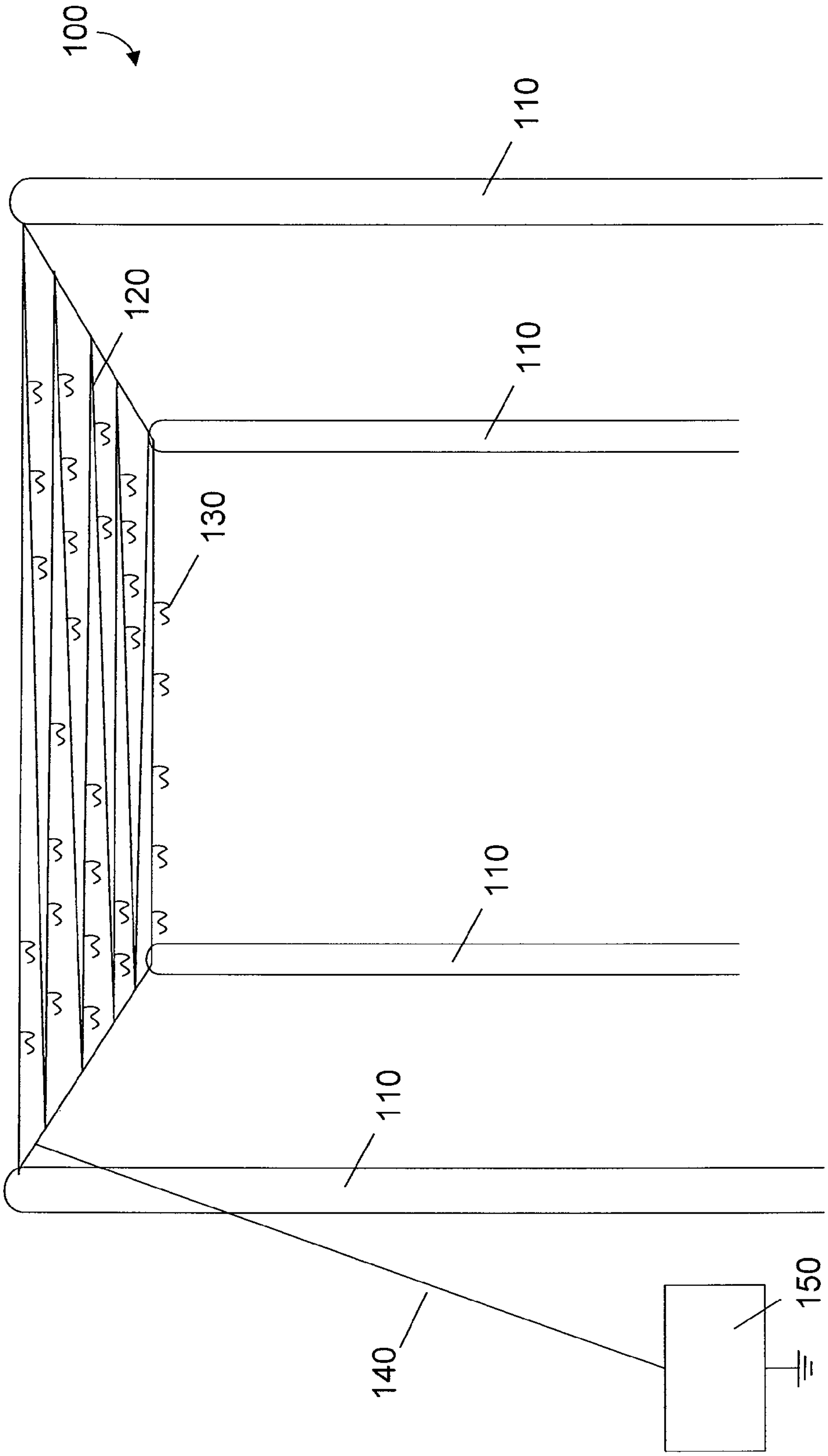
100
110
110
120
130
110
110
140
150

FIGURE 2

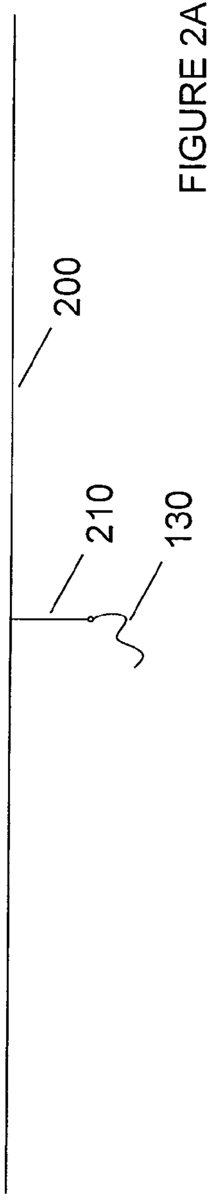

FIGURE 2A

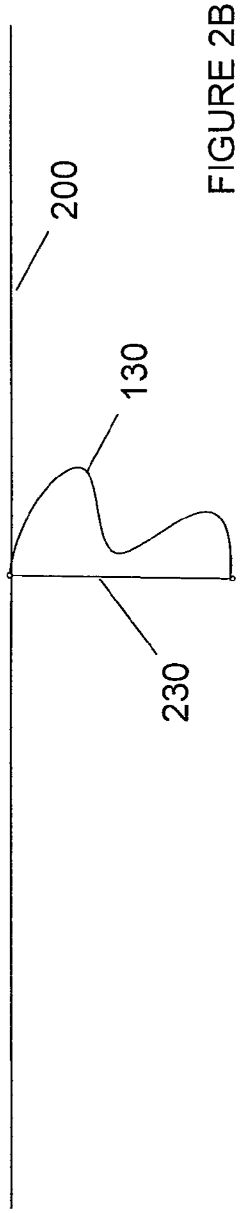

FIGURE 2B

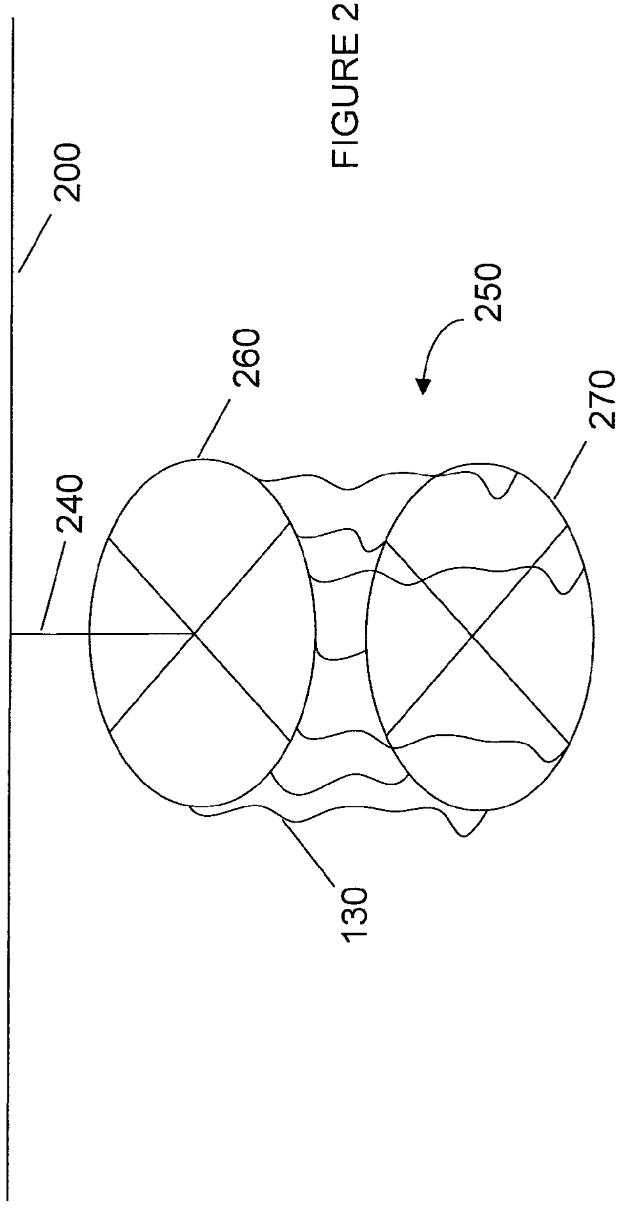
200
240
260
250
270
130

FIGURE 2C

FIGURE 2D

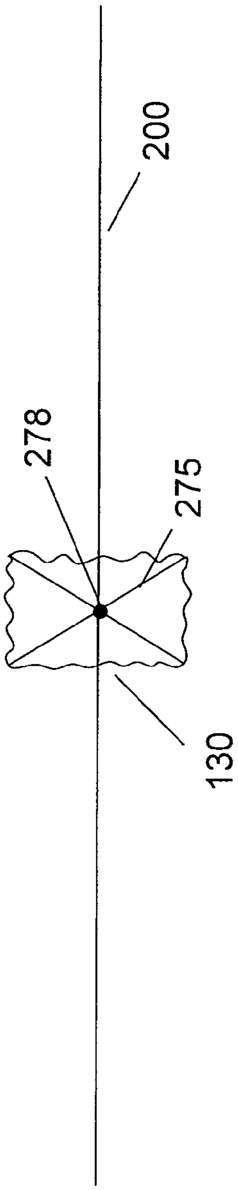
200
278
275
130

FIGURE 2E

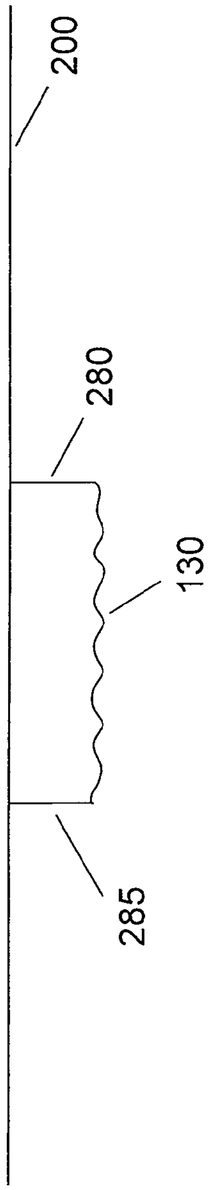
200
280
130
285

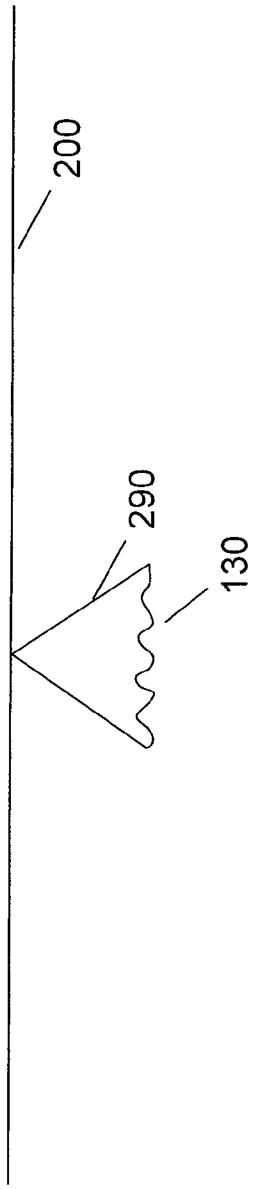

FIGURE 2F

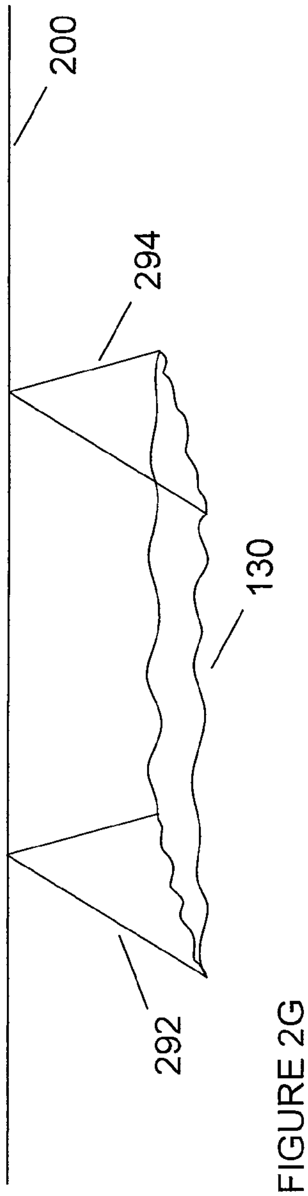
200
294
130
292

FIGURE 2G

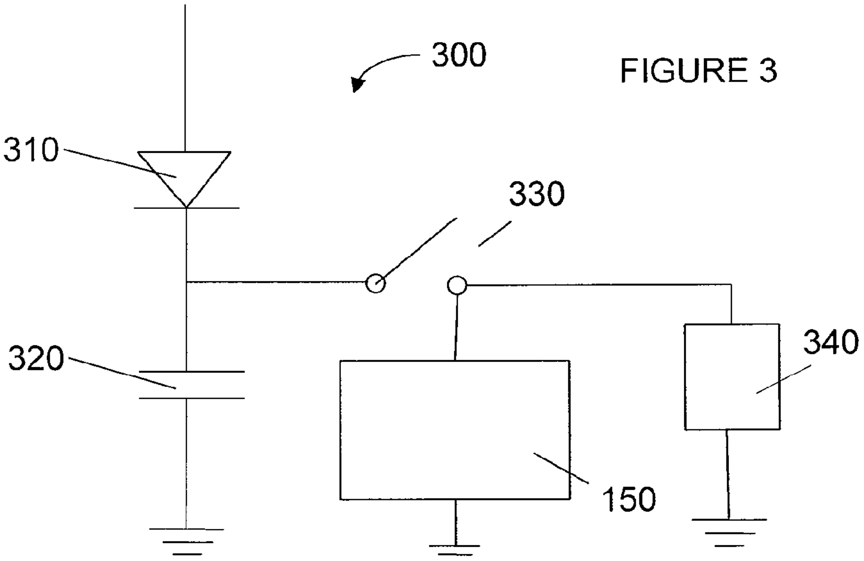
FIGURE 3
300
310
330
320
340
150

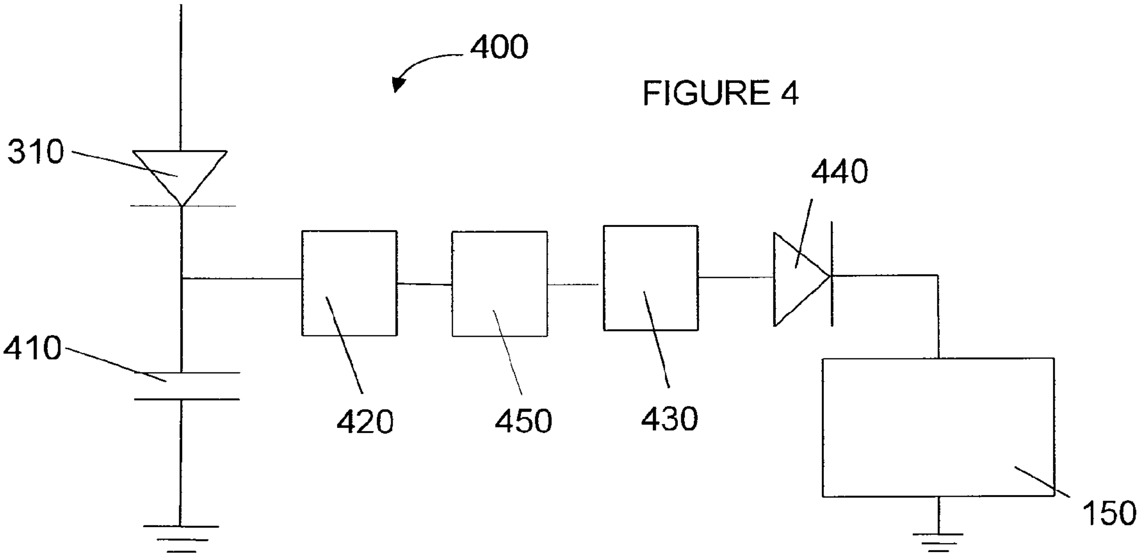
FIGURE 4
400
310
440
410
420
450
430
150

FIGURE 5

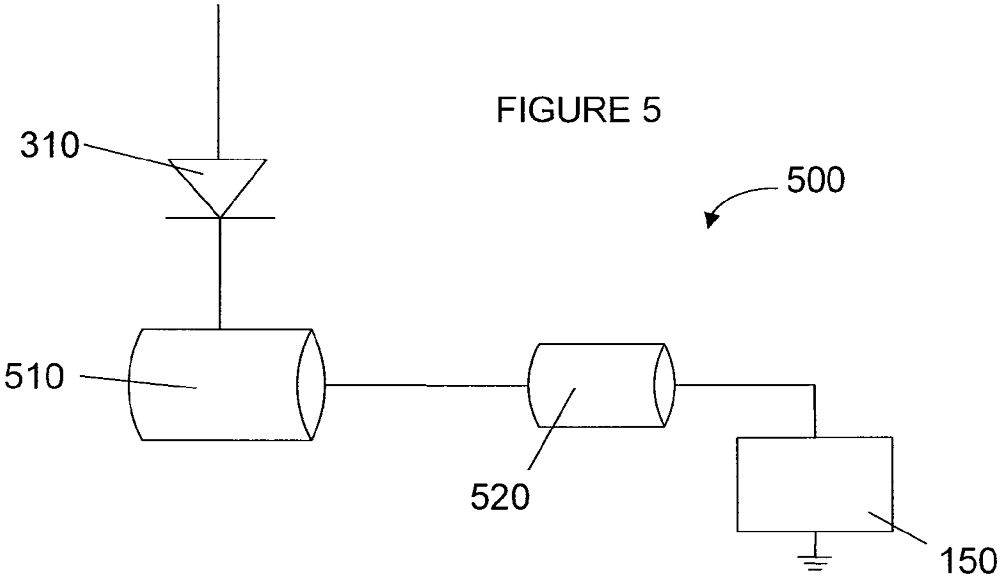
310
500
510
520
150

FIGURE 6

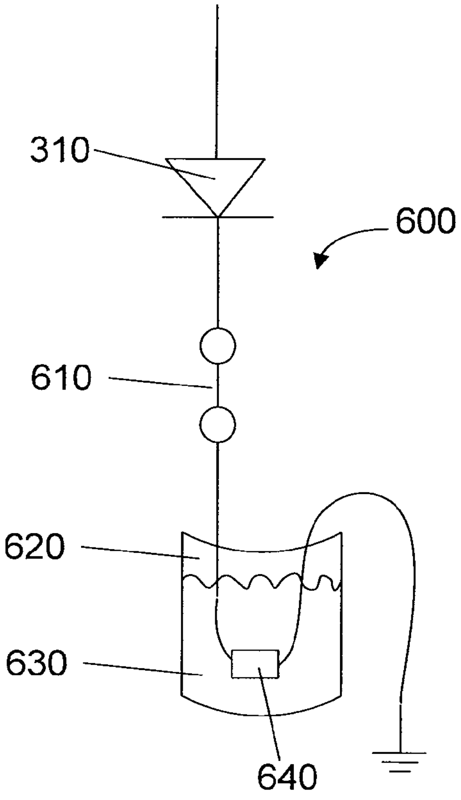
310
600
610
620
630
640

FIGURE 7

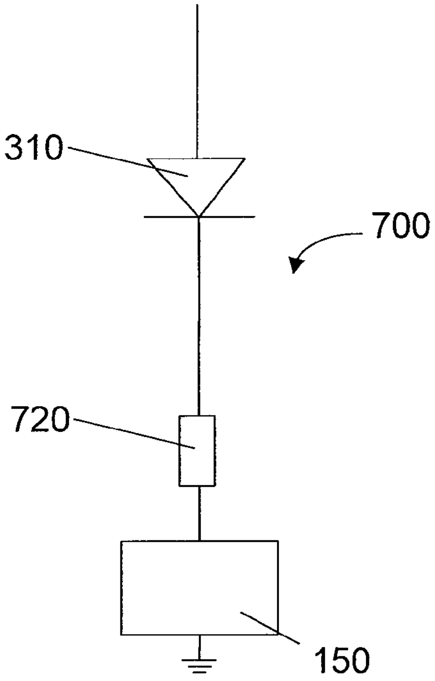
310
700
720
150

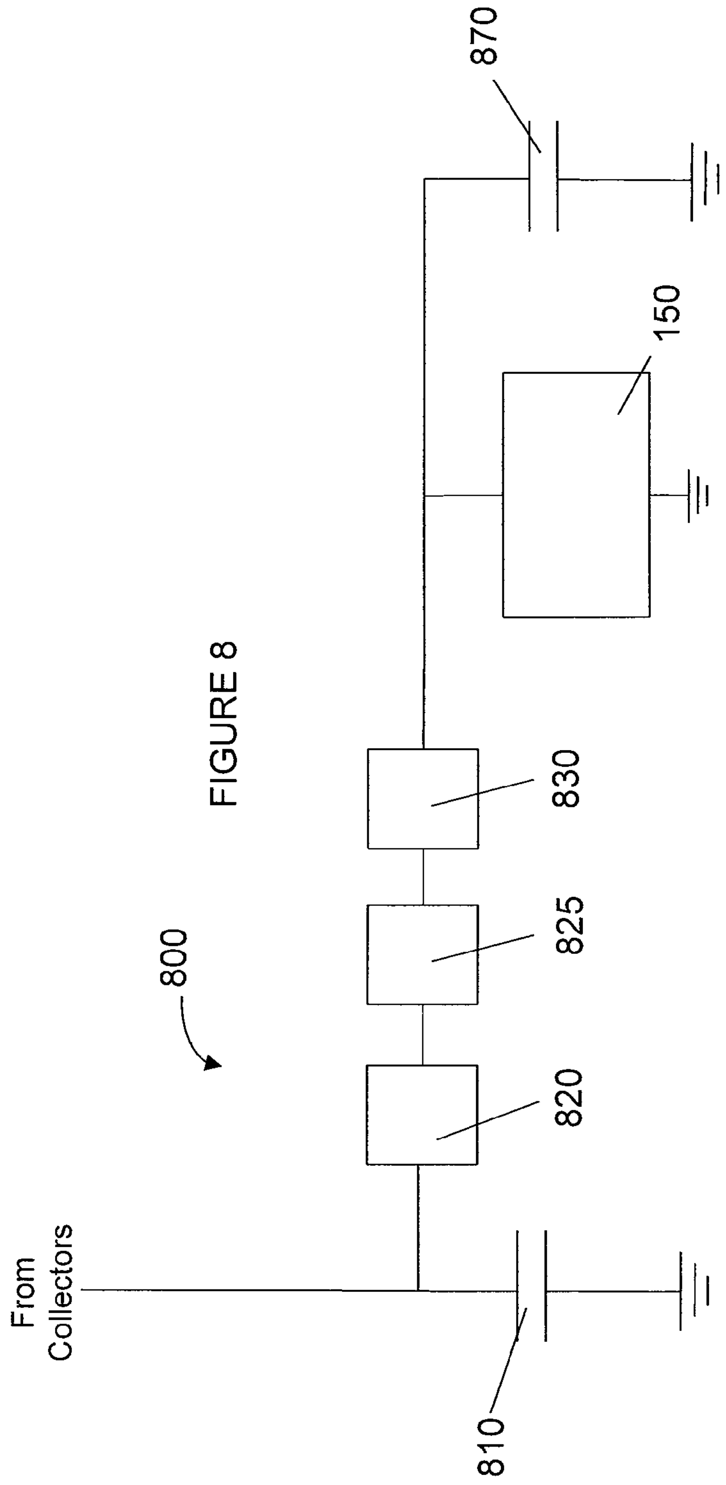
From Collectors
810
820
825
830
800
150
870

FIGURE 8

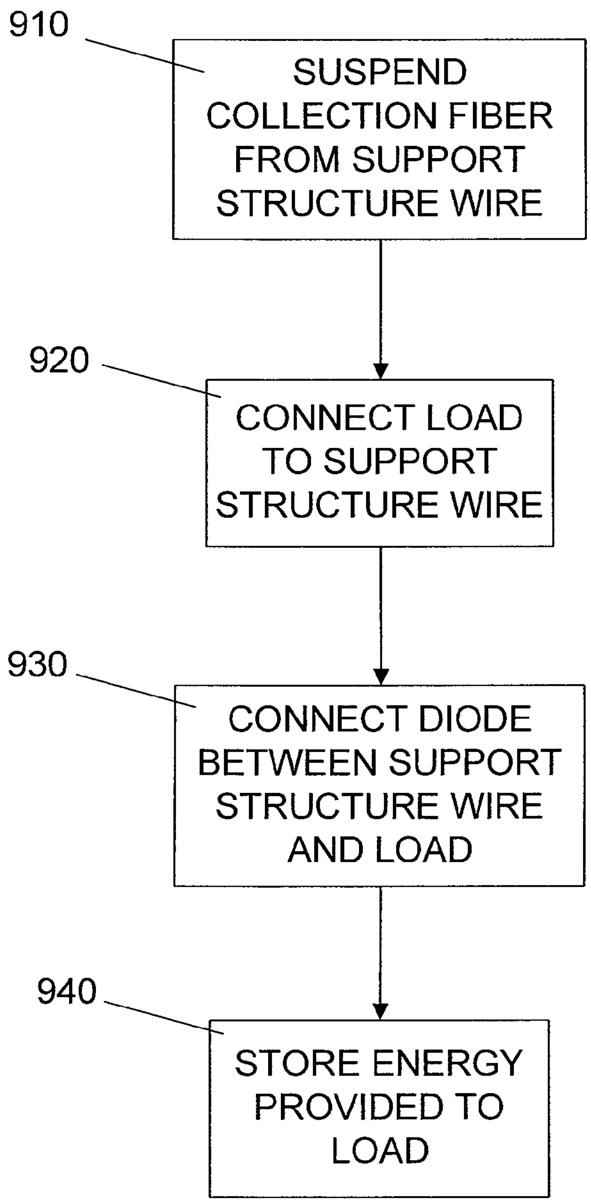
910
SUSPEND COLLECTION FIBER FROM SUPPORT STRUCTURE WIRE
920
CONNECT LOAD TO SUPPORT STRUCTURE WIRE
930
CONNECT DIODE BETWEEN SUPPORT STRUCTURE WIRE AND LOAD
940
STORE ENERGY PROVIDED TO LOAD

FIGURE 9

ENERGY COLLECTION

CROSS REFERENCE TO RELATED APPLICATION

This application is a continuation application of U.S. patent application Ser. No. 11/358,264, filed on Feb. 21, 2006, which is incorporated by reference herein.

TECHNICAL FIELD

The present disclosure is generally related to energy and, more particularly, is related to systems and methods for collecting energy.

BACKGROUND

The concept of fair weather electricity deals with the electric field and the electric current in the atmosphere propagated by the conductivity of the air. Clear, calm air carries an electrical current, which is the return path for thousands of lightening storms simultaneously occurring at any given moment around the earth. For simplicity, this energy may be referred to as static electricity or static energy. FIG. **1** illustrates a weather circuit for returning the current from lightning, for example, back to ground **10**. Weather currents **20**, **30** return the cloud to ground current **40**.

In a lightening storm, an electrical charge is built up, and electrons arc across a gas, ionizing it and producing the lightening flash. As one of ordinary skill in the art understands, the complete circuit requires a return path for the lightening flash. The atmosphere is the return path for the circuit. The electric field due to the atmospheric return path is relatively weak at any given point because the energy of thousands of electrical storms across the planet are diffused over the atmosphere of the entire Earth during both fair and stormy weather. Other contributing factors to electric current being present in the atmosphere may include cosmic rays penetrating and interacting with the earth's atmosphere, and also the migration of ions, as well as other effects yet to be fully studied.

Some of the ionization in the lower atmosphere is caused by airborne radioactive substances, primarily radon. In most places of the world, ions are formed at a rate of 5-10 pairs per cubic centimeter per second at sea level. With increasing altitude, cosmic radiation causes the ion production rate to increase. In areas with high radon exhalation from the soil (or building materials), the rate may be much higher.

Alpha-active materials are primarily responsible for the atmospheric ionization. Each alpha particle (for instance, from a decaying radon atom) will, over its range of some centimeters, create approximately 150,000-200,000 ion pairs.

While there is a large amount of usable energy available in the atmosphere, a method or apparatus for efficiently collecting that energy has not been forthcoming. Therefore, a heretofore unaddressed need exists in the industry to address the aforementioned deficiencies and inadequacies.

SUMMARY

Embodiments of the present disclosure provide systems and methods for collecting energy. Briefly described in architecture, one embodiment of the system, among others, can be implemented by a support structure wire elevated above a ground level, at least one collection fiber electrically connected to the support structure wire; a load electrically connected to the support structure wire; and a diode electrically connected between the load and at least one collection fiber.

Embodiments of the present disclosure can also be viewed as providing methods for collecting energy. In this regard, one embodiment of such a method, among others, can be broadly summarized by the following steps: suspending at least one collection fiber from a support structure wire elevated above ground level, the fiber electrically connected to the support structure wire; providing a load with an electrical connection to the support structure wire to draw current; and providing a diode electrically connected between the collection fiber and the load.

Other systems, methods, features, and advantages of the present disclosure will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present disclosure, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Many aspects of the disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. **1** is a circuit diagram of a weather energy circuit.

FIG. **2** is a perspective view of an exemplary embodiment of many energy collectors elevated above ground by a structure.

FIG. **2**A is a side view of an energy collection fiber suspended from a support wire.

FIG. **2**B is a side view of an exemplary embodiment of an energy collection fiber suspended from a support wire and with an additional support member.

FIG. **2**C is a perspective view of a support structure for multiple energy collection fibers.

FIG. **2**D is a side view of an exemplary embodiment of a support structure for multiple energy collection fibers.

FIG. **2**E is a side view of a support structure for an energy collection fiber.

FIG. **2**F is a side view of an exemplary embodiment of a support structure for an energy collection fiber.

FIG. **2**G is a side view of a support structure for multiple energy collection fibers.

FIG. **3** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **4** is a circuit diagram of an exemplary embodiment of a circuit for the collection of energy.

FIG. **5** is a circuit diagram of an exemplary embodiment of an energy collection circuit for powering a generator and motor.

FIG. **6** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy and using it for the production of hydrogen and oxygen.

FIG. **7** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy, and using it for driving a fuel cell.

FIG. **8** is a circuit diagram of an exemplary embodiment of a circuit for collecting energy.

FIG. **9** is a flow diagram of an exemplary embodiment of collecting energy with a collection fiber.

DETAILED DESCRIPTION

Electric charges on conductors reside entirely on the external surface of the conductors, and tend to concentrate more

around sharp points and edges than on flat surfaces. Therefore, an electric field received by a sharp conductive point may be much stronger than a field received by the same charge residing on a large smooth conductive shell. An exemplary embodiment of this disclosure takes advantage of this property, among others, to collect and use the energy generated by an electric field in the atmosphere. Referring to collection system **100** presented in FIG. **2**, at least one collection device **130** may be suspended from a support wire system **120** supported by poles **110**. Collection device **130** may comprise a diode or a collection fiber individually, or the combination of a diode and a collection fiber. Support wire system **120** may be electrically connected to load **150** by connecting wire **140**. Supporting wire system **120** may be any shape or pattern. Also, conducting wire **140** may be one wire or multiple wires. The collection device **130** in the form of a fiber may comprise any conducting or non-conducting material, including carbon, graphite, Teflon, and metal. An exemplary embodiment utilizes carbon or graphite fibers for static electricity collection. Support wire system **120** and connecting wire **140** can be made of any conducting material, including aluminum or steel, but most notably, copper. Teflon may be added to said conductor as well, such as non-limiting examples of a Teflon impregnated wire, a wire with a Teflon coating, or Teflon strips hanging from a wire. Conducting wire **120**, **140**, and **200** may be bare wire, or coated with insulation as a non-limiting example. Wires **120** and **140** are a means of transporting the energy collected by collection device **130**.

An exemplary embodiment of the collection fibers as collection device **130** includes graphite or carbon fibers. Graphite and carbon fibers, at a microscopic level, can have hundreds of thousands of points. Atmospheric electricity may be attracted to these points. If atmospheric electricity can follow two paths where one is a flat surface and the other is a pointy, conductive surface, the electrical charge will be attracted to the pointy, conductive surface. Generally, the more points that are present, the higher energy that can be gathered. Therefore, carbon, or graphite fibers are examples that demonstrate exemplary collection ability.

In at least one exemplary embodiment, the height of support wire **120** may be an important factor. The higher that collection device **130** is from ground, the larger the voltage potential between collection device **130** and electrical ground. The electric field may be more than 100 volts per meter under some conditions. When support wire **120** is suspended in the air at a particular altitude, wire **120** will itself collect a very small charge from ambient voltage. When collection device **130** is connected to support wire **120**, collection device **130** becomes energized and transfers the energy to support wire **120**.

A diode, not shown in FIG. **2**, may be connected in several positions in collection system **100**. A diode is a component that restricts the direction of movement of charge carriers. It allows an electric current to flow in one direction, but essentially blocks it in the opposite direction. A diode can be thought of as the electrical version of a check valve. The diode may be used to prevent the collected energy from discharging into the atmosphere through the collection fiber embodiment of collection device **130**. An exemplary embodiment of the collection device comprises the diode with no collection fiber. A preferred embodiment, however, includes a diode at the connection point of a collection fiber to support system **120** such that the diode is elevated above ground. Multiple diodes may be used between collection device **130** and load **150**. Additionally, in an embodiment with multiple fibers, the diodes restricts energy that may be collected through one fiber from escaping through another fiber.

Collection device **130** may be connected and arranged in relation to support wire system **120** by many means. Some non-limiting examples are provided in FIGS. **2A-2G** using a collection fiber embodiment. FIG. **2A** presents support wire **200** with connecting member **210** for collection device **130**. Connection member **210** may be any conducting material allowing for the flow of electricity from connection device **130** to support wire **200**. Then, as shown in FIG. **2**, the support wire **200** of support system **120** may be electrically connected through conducting wire **140** to load **150**. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at an elevated position at the connection point between a collection fiber embodiment of collection device **130** and connection member **210**.

Likewise, FIG. **2B** shows collection fiber **130** electrically connected to support wire **200** and also connected to support member **230**. Support member **230** may be connected to collection fiber **130** on either side. Support member **230** holds the fiber steady on both ends instead of letting it move freely. Support member **230** may be conducting or non-conducting. A plurality of diodes may be placed at any position on the support structure wire. A preferred embodiment places a diode at elevated position at the connection point between collection fiber **130** and support wire **200** or between fiber **130**, support member **230**, and support wire **200**.

FIG. **2C** presents multiple collection fibers in a squirrel cage arrangement with top and bottom support members. Support structure **250** may be connected to support structure wire **200** by support member **240**. Structure **250** has a top **260** and a bottom **270** and each of the multiple collection fibers **130** are connected on one end to top **260** and on the other end to bottom **270**. A plurality of diodes may be placed at any position on support structure **250**. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support structure wire **200**.

FIG. **2D** presents another exemplary embodiment of a support structure with support members **275** in an x-shape connected to support structure wire **200** at intersection **278** with collection fibers **130** connected between ends of support members **275**. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2E** provides another exemplary embodiment for supporting collection fiber **130**. Collection fiber **130** may be connected on one side to support member **285**, which may be connected to support structure wire **200** in a first location and on the other side to support member **280**, which may be connected to support structure wire **200** in a second location on support structure wire **200**. The first and second locations may be the same location, or they may be different locations, even on different support wires. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support wire **200**.

FIG. **2F** presents another exemplary embodiment of a support for a collection fiber. Two support members **290** may support either side of a collection fiber and are connected to support wire **200** in a single point. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places a diode at an elevated position at the connection point between collection fiber **130** and support wire **200**.

FIG. **2G** provides two supports as provided in FIG. **2F** such that at least two support members **292**, **294** may be connected

to support structure wire **200** in multiple locations and collection fibers **130** may be connected between each end of the support structures. Collection fibers **130** may be connected between each end of a single support structure and between multiple support structures. A plurality of diodes may be placed at any position on the support structure. A preferred embodiment places one or more diodes at elevated positions at the connection point(s) between collection fiber **130** and support structure wire **200**.

FIG. **3** provides a schematic diagram of storing circuit **300** for storing energy collected by one or more collection devices (**130** from FIG. **2**). Load **150** induces current flow. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. A plurality of diodes may be placed at any position in the circuit. Switch **330** may be electrically connected between load **150** and at least one collection device (**130** from FIG. **2**) to connect and disconnect the load. Capacitor **320** may be connected in parallel to the switch **330** and load **150** to store energy when switch **330** is open for delivery to load **150** when switch **330** is closed. Rectifier **340** may be electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. Rectifier **340** may be a full-wave or a half-wave rectifier. Rectifier **340** may include a diode electrically connected in parallel to load **150**, between the receiving end of switch **330** and ground. The direction of the diode of rectifier **340** is optional.

In an exemplary embodiment provided in FIG. **4**, storage circuit **400** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **410**. If charging capacitor **410** is not used, then the connection to ground shown at capacitor **410** is eliminated. A plurality of diodes may be placed at any position in the circuit. Diode **310** may be electrically connected in series between one or more collection devices (**130** from FIG. **2**) and load **150**. Diode **440** may be placed in series with load **150**. The voltage from capacitor **410** can be used to charge spark gap **420** when it reaches sufficient voltage. Spark gap **420** may comprise one or more spark gaps in parallel. Non-limiting examples of spark gap **420** include mercury-reed switches and mercury-wetted reed switches. When spark gap **420** arcs, energy will arc from one end of the spark gap **420** to the receiving end of the spark gap **420**. The output of spark gap **420** may be electrically connected in series to rectifier **450**. Rectifier **450** may be a full-wave or a half-wave rectifier. Rectifier **450** may include a diode electrically connected in parallel to transformer **430** and load **150**, between the receiving end of spark gap **420** and ground. The direction of the diode of rectifier **450** is optional. The output of rectifier **450** is connected to transformer **430** to drive load **150**.

FIG. **5** presents motor driver circuit **500**. One or more collection devices (**130** from FIG. **2**) are electrically connected to static electricity motor **510**, which powers generator **520** to drive load **150**. A plurality of diodes may be placed at any position in the circuit. Motor **510** may also be directly connected to load **150** to drive it directly.

FIG. **6** demonstrates a circuit **600** for producing hydrogen. A plurality of diodes may be placed at any position in the circuit. One or more collection devices (**130** from FIG. **2**) are electrically connected to primary spark gap **610**, which may be connected to secondary spark gap **640**. Non-limiting examples of spark gaps **610**, **640** include mercury-reed switches and mercury-wetted reed switches. Secondary spark gap **640** may be immersed in water **630** within container **620**. When secondary spark gap **640** immersed in water **630** is energized, spark gap **640** may produce bubbles of hydrogen and oxygen, which may be collected to be used as fuel.

FIG. **7** presents circuit **700** for driving a fuel cell. A plurality of diodes may be placed at any position in the circuit. Collection devices (**130** from FIG. **2**) provide energy to fuel cell **720** which drives load **150**. Fuel cell **720** may produce hydrogen and oxygen.

FIG. **8** presents exemplary circuit **800** for the collection of energy. Storage circuit **800** stores energy from one or more collection devices (**130** from FIG. **2**) by charging capacitor **810**. If charging capacitor **810** is not used, then the connection to ground shown at capacitor **810** is eliminated. A plurality of diodes may be placed at any position in the circuit. The voltage from capacitor **810** can be used to charge spark gap **820** when it reaches sufficient voltage. Spark gap **820** may comprise one or more spark gaps in parallel or in series. Non-limiting examples of spark gap **820** include mercury-reed switches and mercury-wetted reed switches. When spark gap **820** arcs, energy will arc from one end of spark gap **820** to the receiving end of spark gap **820**. The output of spark gap **820** may be electrically connected in series to rectifier **825**. Rectifier **825** may be a full-wave or a half-wave rectifier. Rectifier **825** may include a diode electrically connected in parallel to inductor **830** and load **150**, between the receiving end of spark gap **820** and ground. The direction of the diode of rectifier **825** is optional. The output of rectifier **825** is connected to inductor **830**. Inductor **830** may be a fixed value inductor or a variable inductor. Capacitor **870** may be placed in parallel with load **150**.

FIG. **9** presents a flow diagram of a method for collecting energy. In block **910**, one or more collection devices may be suspended from a support structure wire. In block **920**, a load may be electrically connected to the support structure wire to draw current. In block **930** a diode may be electrically connected between the support structure wire and the electrical connection to the load. In block **940**, energy provided to the load may be stored or otherwise utilized.

Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of the preferred embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

It should be emphasized that the above-described embodiments of the present disclosure, particularly, any "preferred" embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the disclosure. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present disclosure and protected by the following claims.

Therefore, at least the following is claimed:

1. A method of collecting energy comprising:

suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to the support structure;

providing a load with an electrical connection to the at least one collection device to passively draw current; and

powering a fuel cell between the support structure and the load, the powering performed with energy collected from the at least one collection device.

2. The method of claim **1**, wherein the collection device comprises a diode.

3. The method of claim **1**, wherein the collection device comprises a collection fiber.

4. The method of claim **1**, wherein the collection device comprises a diode and a collection fiber and the diode is electrically connected between the collection fiber and the load.

5. The method of claim **1**, further comprising storing energy provided to the load.

6. The method of claim **5**, wherein storing energy provided to the load comprises storing energy in a capacitor or an inductor.

7. The method of claim **3**, wherein the collection fiber comprises carbon fiber or graphite fiber.

8. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the collection device electrically connected to and suspended from the support structure, the at least one collection device configured to passively draw current;

a load electrically connected to the at least one collection device; and

a diode electrically connected between the at least one collection device and the support structure.

9. The system of claim **8**, wherein the collection device comprises a diode.

10. The system of claim **8**, wherein the collection device comprises a collection fiber.

11. The system of claim **8**, wherein the collection device comprises a collection fiber and a diode electrically connected between the load and the collection fiber.

12. The system of claim **9**, wherein the diode is elevated relative to the ground level.

13. The system of claim **10**, wherein the collection fiber comprises a carbon fiber or a graphite fiber.

14. The system of claim **8**, further comprising:

a switch connected in series between the at least one collection device and the load; and

a capacitor connected in parallel with the switch and the load.

15. The system of claim **14**, wherein the switch comprises an interrupter connected between the at least one collection device and the load, and wherein the interrupter comprises at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

16. The system of claim **15**, further comprising a transformer connected between the interrupter and the load.

17. The system of claim **8**, further comprising:

a motor for providing power, the motor connected between the at least one collection device and the load; and

a generator powered by the motor.

18. The system of claim **8**, further comprising a fuel cell between the support structure and the load.

19. The system of claim **18**, wherein the fuel cell produces hydrogen and oxygen.

20. A system of collecting energy comprising:

means for suspending at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to the means for suspending;

means for passively inducing current flow, the means for inducing current flow electrically connected to the means for suspending; and

means for restricting the backflow of charge carriers, the means for restricting the backflow of charge carriers electrically connected between the at least one collection device and the means for inducing current flow.

21. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the collection device electrically connected to and suspended from the support structure, the at least one collection device configured to passively draw current;

a load electrically connected to the at least one collection device; and

a switch connected in series between the at least one collection device and the load, the switch comprising an interrupter connected between the at least one collection device and the load, and wherein the interrupter comprises at least one of a fluorescent tube, a neon bulb, an AC light, and a spark gap.

22. The system of claim **21**, further comprising a transformer connected between the interrupter and the load.

23. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to and suspended from the support structure, the collection device configured to passively draw current;

a load electrically connected to the at least one collection device;

a motor for providing power, the motor connected between the at least one collection device and the load; and

a generator powered by the motor.

24. A system of energy collection comprising:

a support structure;

at least one collection device with, in operation, microscopic points of a cross-section of the collection device exposed to the environment, the at least one collection device electrically connected to and suspended from the support structure the at least one collection device configured to passively draw current;

a load electrically connected to the at least one collection device; and

a fuel cell between the support structure and the load.

25. The system of claim **24**, wherein the fuel cell produces hydrogen and oxygen.

* * * * *

(12) **United States Patent**
Olah et al.

(10) **Patent No.: US 8,823,210 B1**
(45) **Date of Patent: Sep. 2, 2014**

(54) **INTEGRATED ELECTRONICS FOR PERPETUAL ENERGY HARVESTING**

(75) Inventors: **Robert Allen Olah**, Sunnyvale, CA (US); **Achyut Kumar Dutta**, Sunnyvale, CA (US)

(73) Assignee: **Banpil Photonics, Inc.**, Santa Clara, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 552 days.

(21) Appl. No.: **13/151,253**

(22) Filed: **Jun. 1, 2011**

(51) **Int. Cl.**
H02J 1/12 (2006.01)

(52) **U.S. Cl.**
USPC **307/45**

(58) **Field of Classification Search**
USPC 307/43
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

8,097,803	B2 *	1/2012	Chang et al.	136/246
8,432,071	B2 *	4/2013	Huang et al.	307/151
2010/0021816	A1 *	1/2010	Stole	429/211
2011/0148360	A1 *	6/2011	Lee	320/134

* cited by examiner

Primary Examiner — Robert L. Deberadinis

(57) **ABSTRACT**

An apparatus for perpetually harvesting ambient near ultraviolet to far infrared radiation to provide continual power regardless of the environment, incorporating a system for the harvesting electronics governing power management, storage control, and output regulation. The harvesting electronics address issues of efficiently matching the voltage and current characteristics of the different harvested energy levels, low power consumption, and matching the power output demand. The device seeks to harvest the largely overlooked blackbody radiation through use of a thermal harvester, providing a continuous source of power, coupled with a solar harvester to provide increased power output.

20 Claims, 7 Drawing Sheets

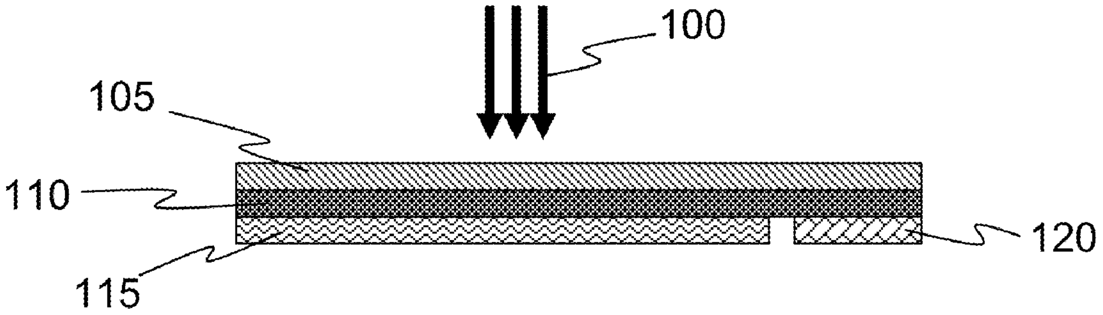

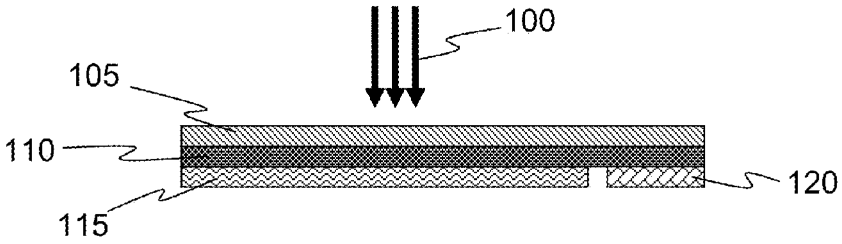
100
105
110
120
115

FIG. 1A

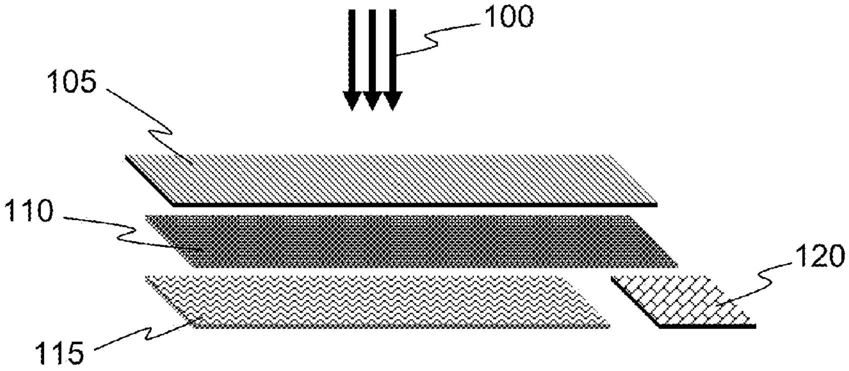
100
105
110
120
115

FIG. 1B

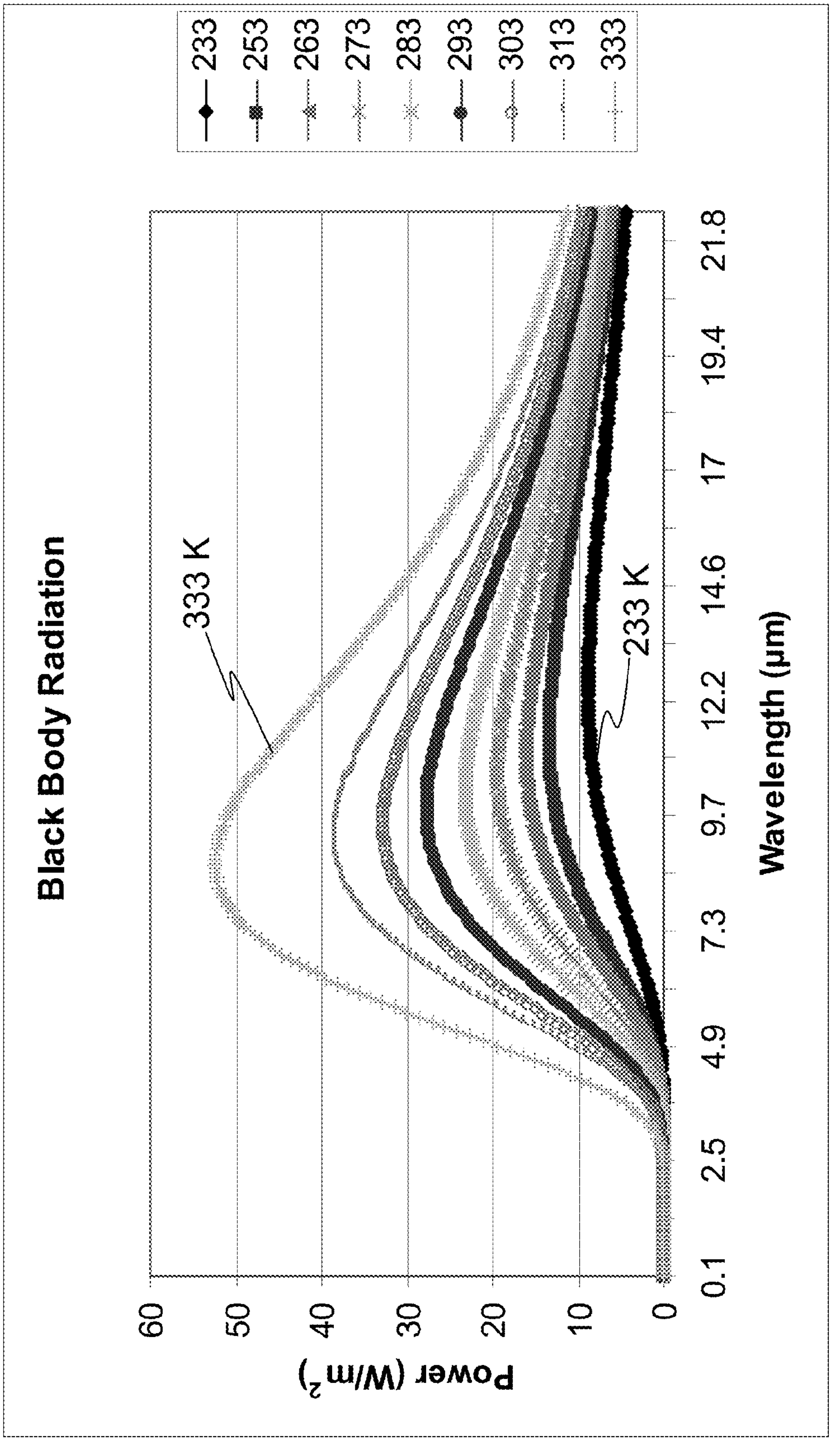
Black Body Radiation
Power (W/m^2)
60
50
40
30
20
10
0
0.1
2.5
4.9
7.3
9.7
12.2
14.6
17
19.4
21.8
Wavelength (µm)
333 K
233 K
233
253
263
273
283
293
303
313
333

FIG. 2

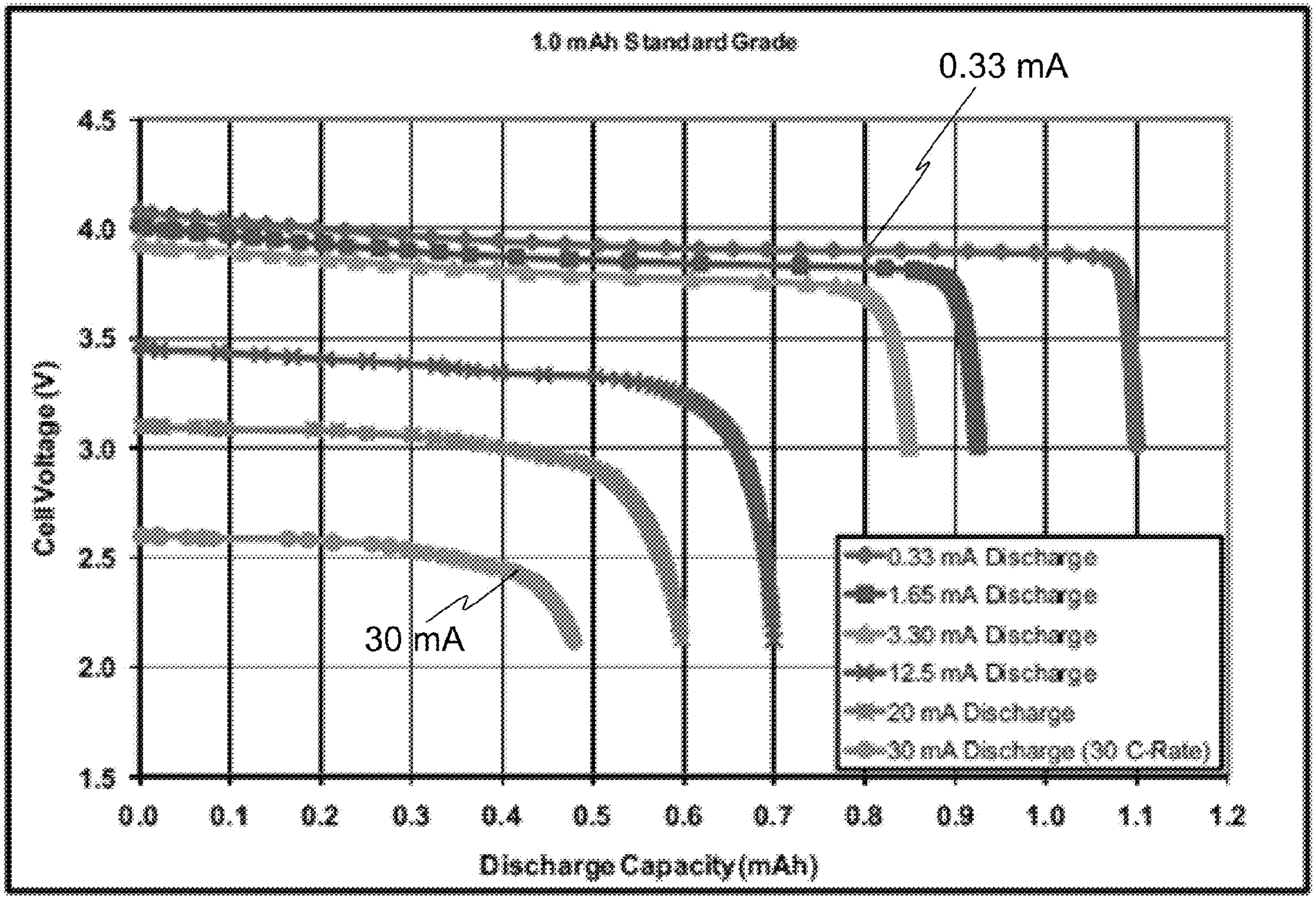
1.0 mAh Standard Grade
0.33 mA
30 mA
Cell Voltage (V)
4.5
4.0
3.5
3.0
2.5
2.0
1.5
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
Discharge Capacity (mAh)
0.33 mA Discharge
1.65 mA Discharge
3.30 mA Discharge
12.5 mA Discharge
20 mA Discharge
30 mA Discharge (30 C-Rate)

FIG. 3

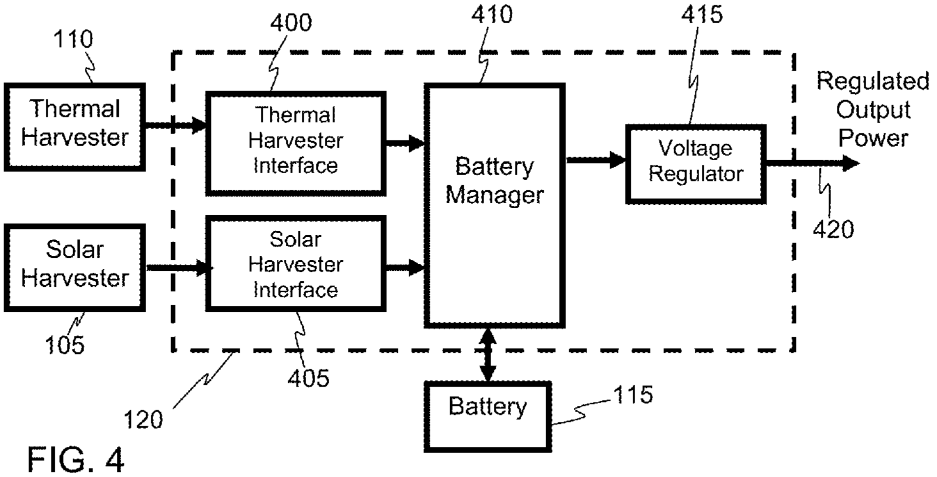
110
400
410
415
Thermal Harvester
Thermal Harvester Interface
Battery Manager
Voltage Regulator
Regulated Output Power
420
Solar Harvester
Solar Harvester Interface
105
405
120
Battery
115

FIG. 4

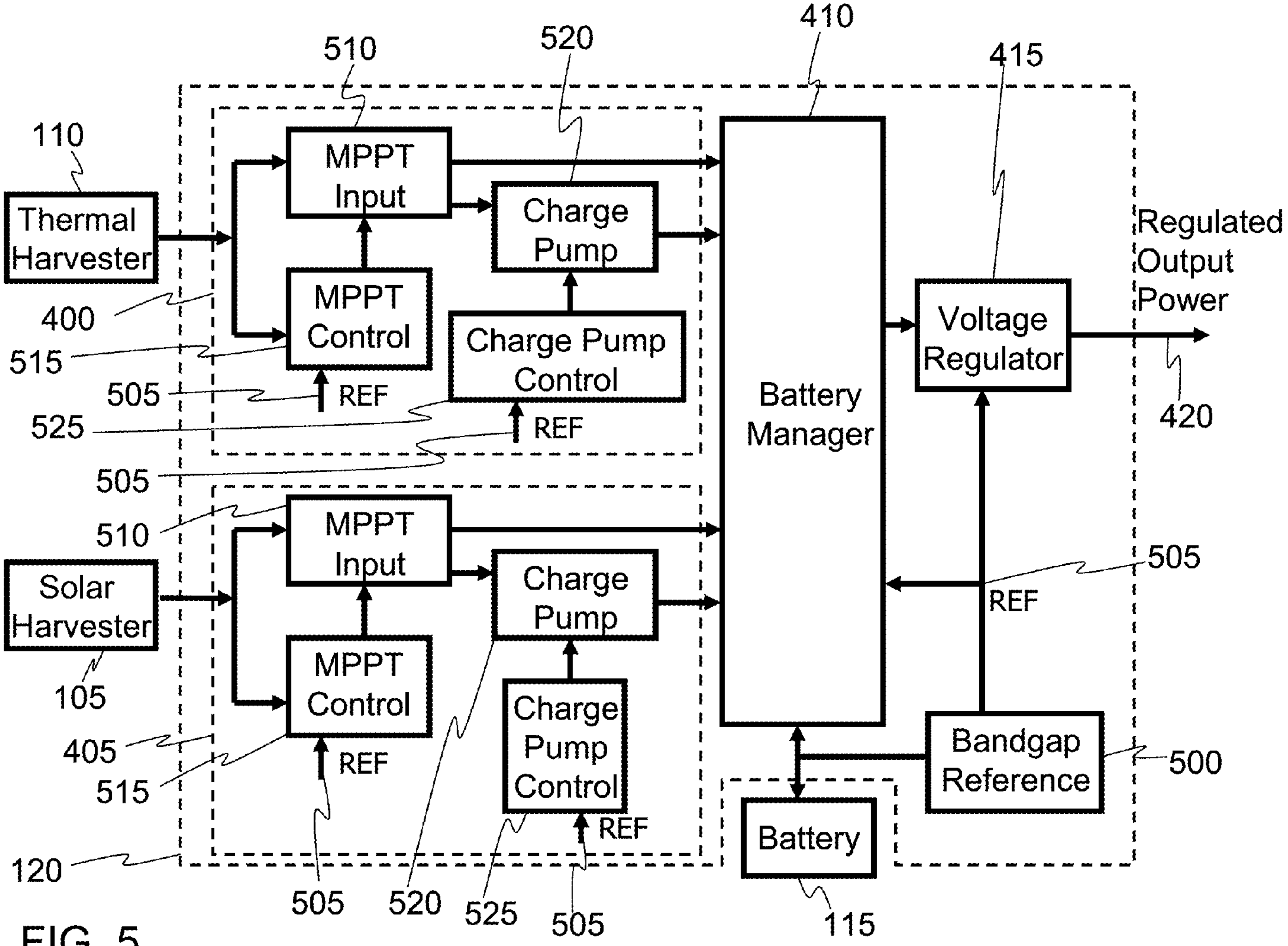
110
Thermal Harvester
400
515
505
525
505
510
Solar Harvester
105
405
515
120
510
520
MPPT Input
MPPT Control
REF
Charge Pump
Charge Pump Control
REF
MPPT Input
MPPT Control
REF
Charge Pump
Charge Pump Control
REF
505
520
525
505
410
Battery Manager
Battery
115
415
Voltage Regulator
Regulated Output Power
420
505
REF
Bandgap Reference
500

FIG. 5

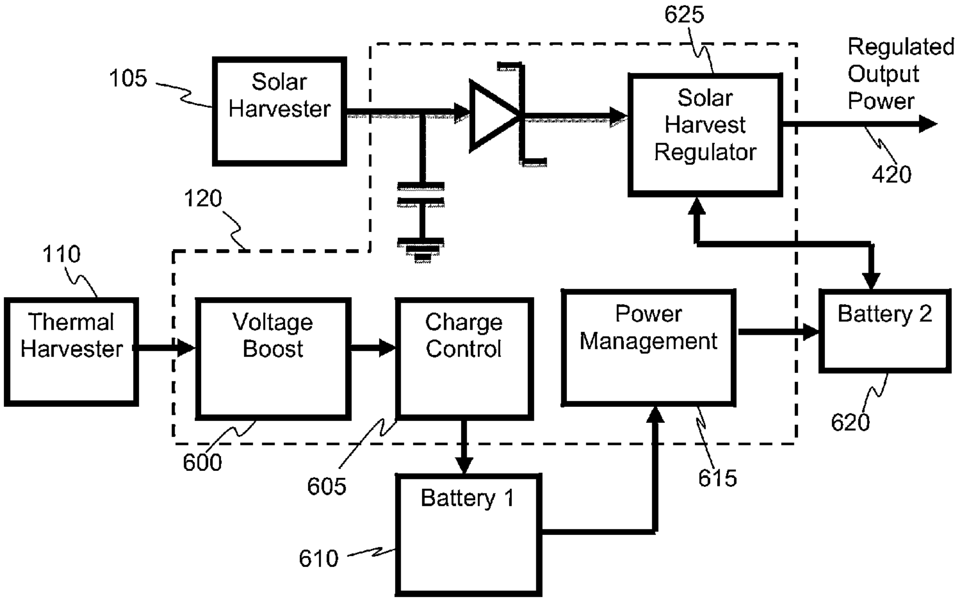
625
Regulated
Output
Power
105
Solar
Harvester
Solar
Harvest
Regulator
420
120
110
Thermal
Harvester
Voltage
Boost
Charge
Control
Power
Management
Battery 2
600
605
615
620
Battery 1
610

FIG. 6

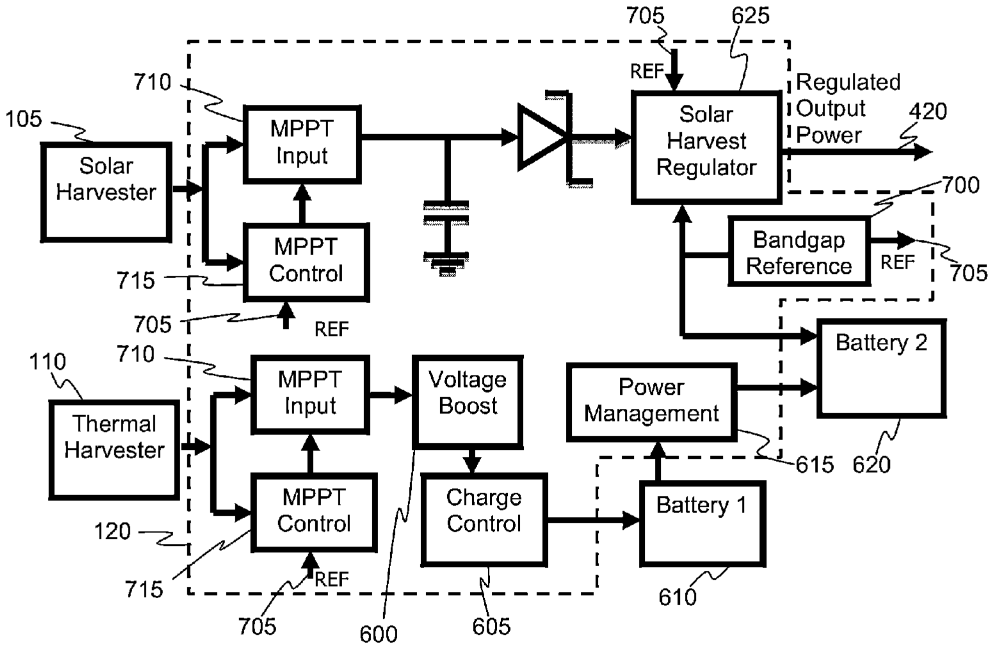
705
625
REF
710
Regulated
Output
Power
420
105
MPPT
Input
Solar
Harvest
Regulator
Solar
Harvester
700
MPPT
Control
Bandgap
Reference
REF
705
715
705
REF
Battery 2
110
710
MPPT
Input
Voltage
Boost
Power
Management
Thermal
Harvester
615
620
MPPT
Control
Charge
Control
Battery 1
120
715
REF
610
705
600
605

FIG. 7

INTEGRATED ELECTRONICS FOR PERPETUAL ENERGY HARVESTING

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

This invention was made in part with Government support under contract no. W31P4Q10C0034 awarded by US Army Aviation and Missile Command. The Government has certain rights in the invention.

CROSS-REFERENCE TO RELATED APPLICATIONS

Not Applicable

CROSS REFERENCE TO RELATED APPLICATIONS

None.

FIELD OF INVENTION

This invention pertains to the harvesting of ambient energy in the surrounding environment. More particularly this invention is related to a semiconductor perpetual energy harvesting and storage device and the harvesting electronics that control energy storage and output allowing the device to continually harvest and provide continuous power during both day and night in any environment.

BACKGROUND OF INVENTION

There are many sources of energy that can be converted into electrical energy. Sources that have been explored for harvesting are wind, temperature differences, motion, light energy and radio frequency energy. In all these energy situations, a specific situation must be present, such as air flow, temperature gradient, movement, sunlight, or transmitted RF, before energy can be harvested. When these situations do not occur, there is no source for energy to be harvested.

Harvesting ambient thermal energy has not been widely explored. The characteristic of thermal energy having a peak at long infrared wavelengths poses a significant challenge for harvesting. Additionally, power management, storage control, and regulation electronics must be developed to operate efficiently and respond to the power output demand.

BRIEF DESCRIPTION OF DRAWINGS

FIGS. **1**A and B depicts a high level structural representation of an energy harvesting system.

FIG. **2** depicts power available from blackbody radiation at various temperatures

FIG. **3** depicts the discharge characteristics of a lithium ion battery.

FIG. **4** depicts a block diagram of the harvester electronics utilizing a dual circuit design.

FIG. **5** depicts an alternate embodiment of a block diagram of the harvester electronics utilizing a dual circuit design.

FIG. **6** depicts a block diagram of the harvester electronics utilizing a single circuit design.

FIG. **7** depicts an alternate embodiment of a block diagram of the harvester electronics utilizing a single circuit design.

DETAILED DESCRIPTION

Reference is made in detail to the preferred embodiments of the invention. While the invention is described in conjunction with the preferred embodiments, the invention is not intended to be limited by these preferred embodiments. The contemplated embodiments for carrying out the present invention are described in turn with reference to the accompanying figures.

For purposes of the invention, radiation is defined to be the electromagnetic spectrum, particularly the near ultraviolet, visible, near infrared, short-wave infrared, mid-wave infrared, long-wave infrared and far infrared bands. The near ultraviolet band is comprised of wavelengths from about 300 to 400 nm. The visible light band is comprised of wavelengths from about 400 to 780 nm. The near infrared band is comprised of wavelengths from about 0.78 to 1 μm. The short-wave infrared band is comprised of wavelengths from about 1 to 3 μm. The mid-wave infrared band is comprised of wavelengths from about 3 to 6 μm. The long-wave infrared band is comprised of wavelengths from about 6 to 14 μm. The far infrared band is comprised of wavelengths from about 14 to 40 μm. However, the boundary between near ultraviolet radiation and visible light along with the boundary between visible light and infrared radiation is not precisely defined, resulting in overlap between the bands.

The term electrically connected is defined to encompass an electrical current flow, including bidirectional, unidirectional, or any hybrid current flow, such as an uneven current.

FIG. **1** depicts a high level structural configuration for an energy harvesting system. FIG. **1**A illustrates a cross section of the structural configuration. FIG. **1**B shows an exploded three dimensional view. The interconnects and electrodes in the structural configurations depicted in FIGS. **1**A and B are not shown.

The energy harvesting system comprises a solar harvester **105**, a thermal harvester **110**, a battery **115**, and harvester electronics **120**. The thermal harvester **110** is located between the solar harvester **105** and the battery **115**. The harvester electronics **120** is attached to the thermal harvester **110** on the same side as the battery **115**. The solar harvester **105** and thermal harvester **110** are electrically isolated from each other and individually electrically connected to the harvester electronics **120**. The battery **115** is electrically connected to the harvester electronics **120**.

The solar harvester **105** and the thermal harvester **110** may be comprised of a photovoltaic or photoelectrochemical system. The solar harvester **105** may include any photovoltaic system which absorbs radiation within the near ultraviolet, visible, or near infrared bands to generate power. The solar harvester **105** may be comprised of any type of solar cell, including traditional crystallized silicon solar cells, thin film solar cells, multi-layered solar cells, dye sensitized solar cells, or organic polymer solar cells. The material comprising the solar harvester may include, but is not limited to, amorphous silicon, polysilicon, crystalline silicon, Ge, SiGe, CdTe, ZnO, CdZnTe, HgCdTe, HgZnTe, GaAs, GaN, InP, GaSb, InSb, other semiconductors, light sensitive dyes, or polymers. A photovoltaic thermal harvester **105** may be comprised of a single junction or multiple junctions. The material comprising the thermal harvester **105** may include HgCdTe, HgZnTe, InSb, InAs, GaSb, GaAs, InP, PbTe, or polymers.

The battery **115** may be any type of rechargeable battery. One embodiment of the invention contemplates the use of solid state thin film battery may be comprised of lithium-ion polymer, lithium, semiconductor, polymer, or flexible batteries. Optionally, the solid state thin film battery may be discretely or monolithically integrated with the harvester. Optionally, in a separate embodiment, bulk batteries may be used. For example, rechargeable lithium air, nickel-cadmium, or nickel metal-hydride batteries. Optionally, a single or multiple battery cells may be used. Optionally, in the case of multiple battery cells, the cells may be arranged in various configurations, e.g. series, parallel, etc. The battery **115** may be flexible, semi-flexible, or rigid. The battery **115** is electrically connected to the harvester electronics **120**.

The harvester electronics **120** may be comprised of circuitry which regulates the rate of charge and voltage level of the battery along with regulating the output of power to a load. In one embodiment of the invention, the harvester electronics may be comprised of a single chip integrated circuit. Optionally, the integrated circuits may be a CMOS circuit. Alternatively, an FPGA may be used for harvester electronics **120** to be configurable in the field. Optionally, the harvester electronics may be discretely integrated with the battery **115**. The harvester electronics may be flexible, semi-flexible, or rigid. The harvester electronics **120** may be powered by the battery **115**, energy from the solar harvester **105** or thermal harvester **110**, or a combination of energy sources.

The system is oriented to absorb radiation **100** arriving from the indicated direction, passing through the solar harvester **105** before reaching the thermal harvester **110**. The radiation **100** is comprised of the near ultraviolet, visible, near infrared, short-wave infrared, mid-wave infrared, long-wave infrared, and far infrared bands.

FIG. **2** illustrates the blackbody power as a function of wavelength at various temperatures. The blackbody radiation spectrum ranges from near ultraviolet to the infrared spectrum. Blackbody radiation has been largely overlooked as a potential energy source.

Blackbody radiation is a physical effect in which all objects at temperatures above 0 K emit electromagnetic power. The spectral radiance of an object is described through Planck's Law, which is presented as,

$$I(\lambda) = \frac{2hc^2}{\lambda^5\left(e^{\frac{hc}{\lambda kT}} - 1\right)},$$

where c is the speed of light, k is Boltzmann's constant, h is Planck's constant, λ is wavelength, and T is temperature in Kelvin of the object. By applying Lambert's cosine law to a hemispherical flux density from a flat surface, the maximum amount of incident radiation is equal to πI. This yields a maximum spectral radiance of

$$I(\lambda) = \frac{2\pi hc^2}{\lambda^5\left(e^{\frac{hc}{\lambda kT}} - 1\right)},$$

which best models the intensity of radiation from a planar source. As temperature increases, power (in watts per meters squared) increases at lower wavelengths. This results in the peak shifting towards lower wavelengths at higher temperatures.

As shown in FIG. **2**, at the temperatures ranging from −40° C. (233 K to +60° C., the peak power is at about 9 μm wavelength. Using Stefan-Boltzmann's law, $5.67\times10^{-8}\times T^4$, the total available blackbody radiation at a given temperature can be calculated. At room temperature, about 293 Kelvin, the total available blackbody radiation is estimated to be 413 W/m^2 with a peak between 7 to 10 μm.

FIG. **3** illustrates the discharge characteristics of a lithium-ion polymer cell at various currents. The figure shows that the cell maintains a relatively constant voltage, which drops sharply as the current decreases due to the battery's inherent resistance increasing. Given this characteristic of lithium-ion polymer cells, completely discharging cells, otherwise known as deep discharge, results in damage to the cells, which may include the inability to hold a charge and sustain a voltage across a load. To guard against this damage, low voltage protection is required. While the figure illustrates the characteristics for a 1.0 mAh grade lithium-ion polymer cell, the characteristics are similar for other lithium-ion polymer cells.

FIG. **4** illustrates a block diagram of the harvester electronics **120** utilizing a two circuit configuration. The harvester electronics **120** are comprised of a thermal harvester interface **400**, a solar harvester interface **405**, a battery manager **410**, and a voltage regulator **415**. Each component comprising the harvester electronics **120** may be electrically connected to and powered by the battery **115**. Optionally, the components of the harvester electronics **120** may be powered by the energy from the solar or thermal harvesters. Optionally, a combination of energy sources may be used to power the harvester electronics **120**. For example, the harvester electronics **120** may be powered by energy from the solar harvester **105** and the battery **115**.

The solar harvester **105** is electrically connected to the solar harvester interface **405**. The solar energy harvested by the solar harvester **105** is transferred to the solar harvester interface **405**. Optionally, the solar harvester interface may comprise a voltage boost to allow harvesting of solar energy at low voltages. Optionally, the voltage boost may be dynamically triggered below a certain threshold. The solar harvester interface is electrically connected to the battery manager **410**.

The thermal harvester **110** is electrically connected to the thermal harvester interface **400**. The thermal energy harvested by the thermal harvester **110** is transferred through the thermal harvester interface **400**. Due to the low voltage output of thermal energy, the thermal harvester interface **400** comprises a voltage boost. For example, the voltage boost may boost voltages from as low as 20 mV. The thermal harvester interface **400** is electrically connected to the battery manager **410**.

The solar harvester interface **405** and the thermal harvester interface **400** transfer the harvested energy to the battery manager **410**. The battery manager **410** merges the harvested energy from the solar harvester interface **405** and the thermal harvester interface **400**. This merged energy may be used to charge the battery **115** and/or be supplied to the electrically connected voltage regulator **415**. The battery manager **410** controls the rate of charge for the battery **115**. The rate of charge varies depending on the battery requirements. The battery manager **410** may be configured to disconnect the battery **115** from the output load when the battery discharge reaches a voltage threshold where further discharge would be harmful to the battery. The voltage at which the battery manager **410** disconnects the battery varies depending on the battery specifications and configuration. For example, if the battery **115** is a lithium-ion polymer with discharge characteristics shown in FIG. **3**, the battery manager **410** will disconnect the battery from the load when the sharp voltage drop occurs. For example, but not to serve as a limitation, if the battery **115** is a 3.3 V lithium polymer battery, the battery manager **410** may be disconnect the battery from the load when the battery output voltage reaches 2.1 V. While the battery **115** is disconnected from the load, the battery manager **410** may continue to charge the battery **115** while providing power to the voltage regulator **415**.

Optionally, the battery manager **410** may include cell balancing capability when multiple battery cells are used. In the case where the battery cells have the same capacity, the battery manager **410** may differentially charge each cell to maximum capacity. In the case where the battery cells have varying capacities, the battery manager **410** may differentially charge each cell to maintain the same state of charge.

The battery manager **410** may be configured to disconnect the battery **115** when the battery is fully charged to prevent overcharging. In the situation where the battery **115** is fully charged and disconnected from further charging, the battery manager **410** will direct all the harvested energy to the voltage regulator **415**. In the absence of adequate power from the solar harvester **105**, the battery manager **410** may utilize the battery **115** to supplement power from the thermal harvester **110** to the voltage regulator **415**.

The voltage regulator **415** maintains a constant voltage level output, which enables the harvester electronics **120** to produce a regulated power output **420**. The voltage regulator **415** may be comprised of a Low Drop Out (LDO) DC-DC voltage regulator. Optionally, other linear voltage regulators or switching regulators may be used. For example, but not to serve as a limitation, other linear voltage regulators may comprise zener regulators, series regulators, transistor regulators, or regulators comprising an op-amp. The voltage regulator **415** may be modular and configured to match a specific range of output loads or increase power output drive capability.

Optionally, the components of the harvesting electronics may be specifically configured to efficiently supply adequate power depending on the specific applications. While the embodiment of the invention contemplates the harvester electronics **120** comprised on a single chip, alternate embodiments may comprise each block or a combination of two or more blocks as discrete circuits.

FIG. **5** illustrates an alternate embodiment of the harvester electronics utilizing a two circuit configuration. The solar harvester **105**, thermal harvester **110**, battery **115**, thermal harvester interface **400**, solar harvester interface **405**, battery manager **410**, and voltage regulator **415** are as described above in FIG. **1** and FIG. **4**.

The harvester electronics comprises a bandgap reference **500**. The bandgap reference **500** is electrically connected to and powered by the battery **115**. Optionally, the bandgap reference may be powered separately by the battery manager, not shown. The bandgap reference **500** provides a temperature independent voltage reference (REF) **505**. The bandgap reference is electrically connected to the battery manager **410** and voltage regulator **415**, along with the MPPT controls and charge pump controls comprising the thermal harvester interface **400** and solar harvester interface **405** and outputs the REF **505** to each of the components.

The solar harvester interface **405** and thermal harvester interface **400** may be further comprised of individual maximum power point tracking (MPPT) inputs **510**, MPPT controls **515**, charge pumps **520**, and charge pump controls **525**. The different harvesting conditions for solar and thermal harvesters may result in different circuitry comprising the aforementioned components. The aforementioned components comprising the individual harvester interfaces may be powered by the battery **115**. Optionally, the components may be powered by the battery manager **410** or the solar or thermal harvester.

The MPPT input **510** is electrically connected to and receives input from the solar or thermal harvester and MPPT control **515**. The MPPT input **510** may track to match the input impedance to the maximum harvesting input impedance to achieve the greatest possible power harvest, during moment to moment variations of light level, shading, temperature, and photovoltaic module characteristics based on the input from the MPPT control **515**.

The MPPT control **515** is electrically connected to and receives input from the solar or thermal harvester and bandgap reference **500**. The input from the bandgap reference **500** comprises the REF **505**. The MPPT control **515** indicates to the MPPT input **510** whether the operating voltage of the harvesting input has changed in a given direction and to adjust the impedance accordingly. Optionally, the MPPT control **515** may determine the disturbance of the operating voltage of the harvesting input through a logic algorithm utilizing the REF **505** and harvester inputs. The MPPT input **510** is electrically connected to and provides an output to the battery manager **410**.

Optionally, the MPPT input **510** may be electrically connected to the charge pump **520**. The charge pump **520** is electrically connected to and receives input from the MPPT input **510** and the charge pump control **525**. The charge pump **520** may be comprised of a capacitive charge pump. Alternatively, the charge pump **520** may be comprised of a step-up converter, boost converter, boost converter, another type of DC-DC converter, or a DC-AC converter coupled to a voltage multiplier. Optionally, the charge pump **520** may be comprised of multiple stages. For example, the charge pump **520** may pump from a voltage as low as 20 mV to over 4 V, using multiple stages to achieve the result.

The charge pump control **525** is electrically connected to and receives input from the bandgap reference **500** in the form of REF **505**. Based on the REF **505**, the charge pump control **525** may adjust the oscillator frequency and the number of charge pump stages used by the charge pump **520**, which may be none, one, or multiple stages.

In an embodiment where the components of the harvester electronics **120** is powered by the battery **115**, the output from the MPPT input **510** to the battery manager **410** may serve as a bypass in the situation where the battery **115** goes below the harvester electronics **120** operating voltage. If the battery **115** goes below the harvester electronics **120** operating voltage, the components may fail to function. This bypass may allow the battery manager to recover the device with a voltage input that could drive the harvest electronics **120** and charge the battery **115** above the minimum operating voltage.

While the embodiment of the invention contemplates the harvester electronics **120** comprised on a single chip, alternate embodiments may comprise each block or a combination of two or more blocks as discrete circuits.

FIG. **6** illustrates a block diagram embodiment of the harvester electronics **120** in a single circuit configuration. The thermal harvester **110** is electrically connected to a voltage boost **600**. The voltage boost **600** may be comprised of a capacitive charge pump. Alternatively, the voltage boost **600** may be comprised of a step-up converter, boost converter, boost converter, another type of DC-DC converter, or a DC-AC converter coupled to a voltage multiplier. The voltage boost increases the low voltage input from the thermal harvester. The voltage boost **600** is electrically connected to the charge control **605**. The charge control **605** controls the energy storage for a first battery **610**. The boosted energy from the thermal harvester **110** is used to charge the first battery **610**. The charge control **605** controls the rate of charge and the battery voltage level of the first battery **610**. When the first battery **610** is fully charged, the charge control **605** disconnects the battery to prevent overcharging.

The first battery **610** may be comprised of any type of rechargeable battery. One embodiment of the invention contemplates the use of solid state thin film battery may be comprised of lithium-ion polymer, lithium, semiconductor, polymer, or flexible batteries. Optionally, the solid state thin film battery may be discretely or monolithically integrated with the harvester. Optionally, in a separate embodiment, bulk batteries may be used. For example, rechargeable lithium air, nickel-cadmium, or nickel metal-hydride batteries. Optionally, the first battery **610** may be comprised of single or multiple battery cells.

The first battery **610** is electrically connected to the power management block **615**. The power management block disconnects the first battery **610** from the output load when the battery when the battery discharge reaches a voltage threshold where further discharge would be harmful to the battery. For example, if the first battery **610** comprises a lithium-ion polymer battery with discharge characteristics similar to the depiction in FIG. **3**, the power management block **615** will disconnect the battery from the load when the sharp voltage drop occurs. Additionally, the power management **615** is electrically connected to a second battery **620**. The power management block **615** regulates the output voltage of the first battery **610** to charge the second battery **620**. Optionally, the charge control **605** and the power management block **615** may be combined into a single block to regulate the storage and output of the first battery **610**.

The second battery **620** may be comprised of any type of rechargeable battery. One embodiment of the invention contemplates the use of a solid state thin film battery comprised of lithium-ion polymer, lithium, semiconductor, polymer, or flexible batteries. Optionally, the solid state thin film battery may be discretely or monolithically integrated with the harvester. Optionally, in a separate embodiment, bulk batteries may be used. For example, rechargeable lithium air, nickel-cadmium, or nickel metal-hydride batteries. Optionally, the second battery **620** may be comprised of single or multiple battery cells. The second battery **620** is electrically connected to and receives solar energy input from the solar harvest regulator **625**.

The solar harvest regulator **625** is electrically connected to and receives input from the solar harvester **105**. Optionally, a zener protection diode may be used between the solar harvester **105** and solar harvest regulator **625** to protect the entire circuit. The solar harvest regulator **625** may comprise a mechanism to disconnect the second battery **620** once the battery reaches a full charge to prevent overcharging. Optionally, the solar harvest regulator **625** may be electrically connected to the voltage boost **600** and receives harvested thermal energy. Additionally, the solar harvest regulator **625** may comprise a voltage output regulator to disconnect the output when the battery discharge reaches a voltage threshold where further discharge would be harmful to the battery. Optionally, the solar harvest regulator may comprise a LDO regulator, another type of linear voltage regulator, or a switching regulator to control the output to a load. For example, but not to serve as a limitation, other linear voltage regulators may comprise zener regulators, series regulators, transistor regulators, or regulators comprising an op-amp. Optionally the solar harvest regulator **625** may protect the second battery **620** from unintended discharge when the input is lower than the battery voltage. For example, a zener protection diode may be used as a mechanism, where the voltage of the solar input must match the voltage of the battery plus the zener diode voltage.

While the embodiment of the invention contemplates the harvester electronics **120** comprised on a single chip, alternate embodiments may comprise each block or a combination of two or more blocks as discrete circuits.

FIG. **7** illustrates an alternate block diagram embodiment of the harvester electronics **120** in a single circuit configuration. The harvester electronics comprises a bandgap reference **700**. The bandgap reference **700** is electrically connected to and powered by the second battery **620**. Optionally, the bandgap reference may be powered separately by the solar harvest regulator **625**, not shown. The bandgap reference **700** provides a temperature independent voltage reference (REF) **705** output. The bandgap reference is electrically connected to the solar harvest regulator **625** and the MPPT control **715** for the thermal and solar harvester inputs and outputs the REF **505** to each of the components.

The MPPT input **710** is electrically connected to and receives input from the solar or thermal harvester and MPPT control **715**. The MPPT input **710** may track to match the input impedance to the maximum harvesting input impedance to achieve the greatest possible power harvest, during moment to moment variations of light level, shading, temperature, and photovoltaic module characteristics based on the input from the MPPT control **715**.

The MPPT control **715** is electrically connected to and receives input from the solar or thermal harvester and bandgap reference **700**. The input from the bandgap reference **700** comprises the REF **705**. The MPPT control **715** indicates to the MPPT input **710** whether the operating voltage of the harvesting input has changed in a given direction and to adjust the impedance accordingly. Optionally, the MPPT control **715** may determine the disturbance of the operating voltage of the harvesting input through a logic algorithm utilizing the REF **705** and harvester inputs.

Optionally, in another embodiment of the invention, a second bandgap reference, not shown, may be connected to the charge control **605** and the first battery **610**. The second bandgap reference is electrically connected to and provides a reference output to the MPPT control **715** connected to the thermal harvester **110**. This embodiment allows for individual MPPT tracking for different sources of harvested energy.

While the embodiment of the invention contemplates the harvester electronics **120** comprised on a single chip, alternate embodiments may comprise each block or a combination of two or more blocks as discrete circuits.

While the aforementioned embodiments depict harvesting electronics which control the storage of energy and regulate output to a load for solar and thermal harvesters, the harvesting electronics may be configured to receive input from multiple sources of energy harvesting. Examples of other sources may include, but are not limited to, air flow, temperature gradient, movement and transmitted radio frequencies.

Whereas many alterations and modifications of the present invention will no doubt become apparent to a person of ordinary skill in the art after having read the foregoing description, it is to be understood that the particular embodiments shown and described by way of illustration are in no way intended to be considered limiting. Therefore, reference to the details of the preferred embodiments is not intended to limit their scope.

The embodiments were chosen and described in order to explain the principles and applications of the invention, thereby allowing others skilled in the art to utilize the invention in its various embodiments and modifications according to the particular purpose contemplated. The scope of the invention is intended to be defined by the claims appended hereto and their equivalents.

The expected practical use of the present invention is the novel integrated perpetual energy harvester with energy storage and harvesting electronics. The proposed invention may be used as a one chip solution to store power from multiple sources of energy harvesting and regulate the output to a load.

What is claimed is:

1. A perpetual energy harvester comprising:

a solar harvester;

a photovoltaic thermal harvester attached to the bottom of the solar harvester;

a battery in close proximity to the thermal harvester; and

a harvesting electronics in close proximity to the thermal harvester and individually electrically connected to the solar harvester, the thermal harvester, and the battery, wherein the harvesting electronics receives input in the form of energy from the solar harvester and the thermal harvester, controls the amount of energy stored in the battery, and regulates the output to a load.

2. The perpetual energy harvester in claim **1**, wherein said battery is selected from the group consisting of thin film lithium-ion, lithium-ion polymer, polymer, semiconductor, and flexible batteries.

3. The perpetual energy harvester in claim **1**, wherein the solar harvester is selected from the group consisting of traditional crystallized silicon solar cells, thin film solar cells, multi-layered solar cells, dye sensitized solar cells, and organic polymer solar cells.

4. The perpetual energy harvester in claim **1**,

wherein the photovoltaic thermal harvester is comprised of material selected from the group consisting of HgCdTe, HgZnTe, InSb, InAs, GaSb, GaAs, InP, PbTe, and polymers.

5. A harvester electronics system comprising:

a thermal harvester interface, wherein the thermal harvester interface is electrically connected to and receives input from a thermal harvester;

a solar harvester interface, wherein the solar harvester interface is electrically connected to and receives input from a solar harvester;

a battery manager electrically connected to and receives input from the solar harvester interface and thermal harvester interface, wherein the battery manager is electrically connected to a battery and manages the energy stored in the battery; and

a voltage regulator electrically connected to the battery manager, wherein the voltage regulator receives an input from the battery manager and controls an output voltage to an external load.

6. The voltage regulator in claim **5**, further comprising a low drop out regulator.

7. The harvester electronics system in claim **5**, further comprising:

a bandgap reference electrically connected to the thermal harvester interface, the solar harvester interface, the battery manager, and the voltage regulator, wherein the bandgap reference outputs a temperature independent voltage reference to the thermal harvester interface, the solar harvester interface, the battery manager, and the voltage regulator.

8. The harvester electronics system in claim **7**, wherein said solar harvester interface comprises:

a Maximum Power Point Tracking (MPPT) input electrically connected to the solar harvester and the battery manager;

a MPPT control electrically connected to the solar harvester, the MPPT input, and the bandgap reference, wherein the MPPT control receives a voltage reference input from the bandgap reference and an input from the solar harvester, determines if the operating voltage has changed, and outputs instructions to the MPPT input which adjusts input impedance to match maximum harvesting input impedance;

a charge pump electrically connected to the MPPT input and the battery manager; and

a charge pump control electrically connected to the charge pump and the bandgap reference, wherein the charge pump control receives input from the bandgap reference and dynamically adjusts oscillation frequency for the charge pump.

9. The harvester electronics system in claim **8**, wherein the charge pump comprises one or more stages selected from the group consisting of a capacitive charge pump, a charge transfer switch, a step-up converter, and a boost converter, wherein the charge pump control dynamically adjusts the number of stages used by the charge pump based on the input from the bandgap reference.

10. The harvester electronics system in claim **7**, wherein said thermal harvester interface comprises:

a Maximum Power Point Tracking (MPPT) input electrically connected to the thermal harvester and the battery manager;

a MPPT control electrically connected to the thermal harvester, the MPPT input, and the bandgap reference, wherein the MPPT control receives a voltage reference input from the bandgap reference and an input from the thermal harvester, determines if the operating voltage has changed, and outputs instructions to the MPPT input which adjusts input impedance to match maximum harvesting input impedance;

a charge pump electrically connected to the MPPT input and the battery manager; and

a charge pump control electrically connected to the charge pump and the bandgap reference, wherein the charge pump control receives input from the bandgap reference and dynamically adjusts oscillation frequency for the charge pump.

11. The harvester electronic system in claim **10**, wherein said charge pump comprises one or more stages selected from the group consisting of a capacitive charge pump, a charge transfer switch, a step-up converter, and a boost converter, wherein the charge pump control dynamically adjusts the number of stages used by the charge pump based on the input from the bandgap reference.

12. A harvesting electronics system comprising:

a voltage boost electrically connected to a thermal harvester, wherein the voltage boost receives energy input from the thermal harvester;

a charge control electrically connected to the voltage boost, wherein the charge control receives energy input from the voltage boost;

a first battery electrically connected to the charge control, wherein the charge control regulates the energy storage of the first battery;

a power management block electrically connected to the first battery, wherein the power management block regulates the energy output of the first battery;

a second battery electrically connected to the power management block; and

a solar harvest regulator electrically connected to a solar harvester and the second battery, wherein the solar harvest regulator receives energy input from the solar harvester, regulates the energy storage of the second battery, and regulates the output to a load.

13. The harvesting electronics in claim **12**,

wherein the voltage boost comprises one or more stages selected from the group consisting of a capacitive charge pump, a charge transfer switch, a step-up converter, and a boost converter.

14. The harvesting electronics system in claim **12**, wherein:

the first battery is selected from the group consisting of thin film lithium-ion, lithium-ion polymer, polymer, semiconductor, and flexible batteries; and

the second battery is selected from the group consisting of thin film lithium-ion, lithium-ion polymer, polymer, semiconductor, and flexible batteries.

15. The harvesting electronics system in claim **12**, further comprising:

a zener protection diode between the solar harvester and the solar harvest regulator.

16. The harvesting electronics system in claim **12**, wherein the solar harvest regulator comprises a low drop out regulator.

17. The harvesting electronics system in claim **12**, further comprising:

a bandgap reference electrically connected to the second battery and the solar harvest regulator, wherein the bandgap reference outputs a temperature independent voltage reference to the solar harvest regulator.

18. The harvesting electronics system in claim **17**, comprising:

a first Maximum Power Point Tracking (MPPT) input electrically connected to the solar harvester, wherein the first MPPT input receives an energy input from the solar harvester;

a first MPPT control electrically connected to the solar harvester, the first MMPT input, and the bandgap reference, wherein the first MPPT control receives a voltage reference input from the bandgap reference and an input from the solar harvester, determines if the operating voltage has moved, and outputs instructions to the first MPPT input which adjusts input impedance to match maximum harvesting input impedance; and

a second MPPT input electrically connected to the thermal harvester, wherein the second MPPT input receives an energy input from the thermal harvester; and

a second MPPT control electrically connected to the thermal harvester, the second MMPT input, and the bandgap reference, wherein the second MPPT control receives a voltage reference input from the bandgap reference and an input from the thermal harvester, determines if the operating voltage has moved, and outputs instructions to the second MPPT input which adjusts input impedance to match maximum harvesting input impedance.

19. The harvesting electronics system in claim **12**, further comprising:

a first bandgap reference electrically connected to the second battery and the solar harvest regulator, wherein the first bandgap reference outputs a first temperature independent voltage reference to the solar harvest regulator; and

a second bandgap reference electrically connected to the first battery and the charge control, wherein the second bandgap reference outputs a second temperature independent voltage reference.

20. The harvesting electronics system in claim **19**, comprising:

a first Maximum Power Point Tracking (MPPT) input electrically connected to the solar harvester, wherein the first MPPT input receives an energy input from the solar harvester;

a first MPPT control electrically connected to the solar harvester, the first MMPT input, and the bandgap reference, wherein the first MPPT control receives the first temperature independent voltage reference input from the bandgap reference and an input from the solar harvester, determines if the operating voltage has moved, and outputs instructions to the first MPPT input which adjusts input impedance to match maximum harvesting input impedance; and

a second MPPT input electrically connected to the thermal harvester, wherein the second MPPT input receives an energy input from the thermal harvester; and

a second MPPT control electrically connected to the thermal harvester, the second MMPT input, and the bandgap reference, wherein the second MPPT control receives the second temperature independent voltage reference input from the bandgap reference and an input from the thermal harvester, determines if the operating voltage has moved, and outputs instructions to the second MPPT input which adjusts input impedance to match maximum harvesting input impedance.

* * * * *

US009214900B2

(12) United States Patent
Arkiszewski

(10) **Patent No.: US 9,214,900 B2**
(45) **Date of Patent: Dec. 15, 2015**

(54) **INTERFERENCE REDUCTION BETWEEN RF COMMUNICATIONS BANDS**

(75) Inventor: **Roman Zbigniew Arkiszewski**, Oakridge, NC (US)

(73) Assignee: **RF Micro Devices, Inc.**, Greensboro, NC (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **13/479,816**

(22) Filed: **May 24, 2012**

(65) **Prior Publication Data**

US 2013/0135052 A1 May 30, 2013

Related U.S. Application Data

(63) Continuation-in-part of application No. 13/090,663, filed on Apr. 20, 2011, now Pat. No. 8,538,355, and a continuation-in-part of application No. 13/172,371, filed on Jun. 29, 2011, now Pat. No. 8,983,409, which

(Continued)

(51) **Int. Cl.**

H03F 3/19	(2006.01)
H03F 3/68	(2006.01)

(Continued)

(52) **U.S. Cl.**

CPC ***H03F 3/19*** (2013.01); ***H03F 1/0227*** (2013.01); ***H03F 1/0261*** (2013.01); ***H03F 1/0277*** (2013.01); ***H03F 3/195*** (2013.01); ***H03F 3/211*** (2013.01); ***H03F 3/245*** (2013.01); ***H03F 3/602*** (2013.01); ***H03F 3/68*** (2013.01); ***H03F 3/72*** (2013.01); *H03F 2200/171* (2013.01); *H03F 2200/222* (2013.01); *H03F 2200/27* (2013.01);

(Continued)

(58) **Field of Classification Search**

CPC H04B 2001/0408; H04B 2001/0416; H04B 2001/0425; H04B 2001/0433; H04B 2001/0441; H04B 2001/045; H04B 1/1607; H03F 3/19

USPC 455/127.1, 127.2, 126, 127.5, 91

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

3,735,289	A	5/1973	Bruene
4,523,155	A	6/1985	Walczak et al.

(Continued)

FOREIGN PATENT DOCUMENTS

GB	2444984	A	6/2008

OTHER PUBLICATIONS

C.H. Li, Quadrature Power Amplifier for RF Applications, Nov. 2009, University of Twente, pp. 23, 61-68.*

(Continued)

Primary Examiner — Christian Hannon

(74) *Attorney, Agent, or Firm* — Withrow & Terranova, P.L.L.C.

(57) **ABSTRACT**

Radio frequency (RF) power amplifier (PA) circuitry and a PA envelope power supply are disclosed. The RF PA circuitry receives and amplifies an RF input signal to provide an RF output signal using an envelope power supply signal, which is provided by the PA envelope power supply. The RF PA circuitry operates in either a normal RF spectral emissions mode or a reduced RF spectral emissions mode. When reduced RF spectral emissions are required, the RF PA circuitry operates in the reduced RF spectral emissions mode. As such, at a given RF output power, during the reduced RF spectral emissions mode, RF spectral emissions of the RF output signal are less than during the normal RF spectral emissions mode. As a result, the reduced RF spectral emissions mode may be used to reduce interference between RF communications bands.

19 Claims, 42 Drawing Sheets

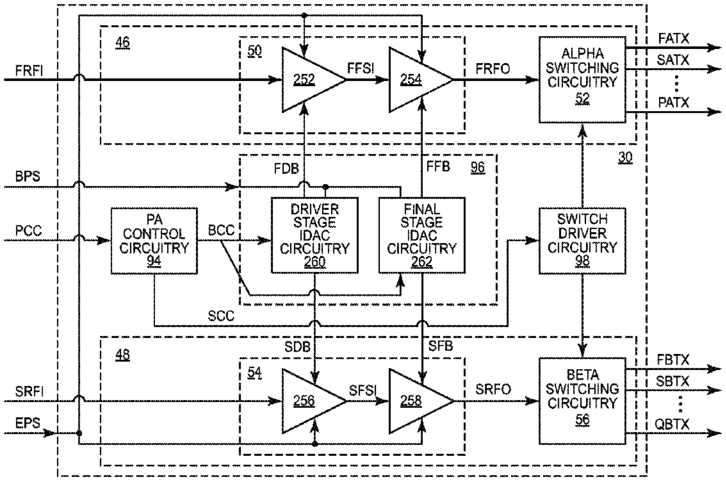
46
50
FRFI
252
FFSI
254
FRFO
ALPHA SWITCHING CIRCUITRY 52
FATX
SATX
PATX
30
BPS
FDB
FFB
96
PCC
PA CONTROL CIRCUITRY 94
BCC
DRIVER STAGE IDAC CIRCUITRY 260
FINAL STAGE IDAC CIRCUITRY 262
SWITCH DRIVER CIRCUITRY 98
SCC
SDB
SFB
48
54
SRFI
256
SFSI
258
SRFO
BETA SWITCHING CIRCUITRY 56
FBTX
SBTX
QBTX
EPS

Related U.S. Application Data

is a continuation-in-part of application No. 13/090,663, filed on Apr. 20, 2011, now Pat. No. 8,538,355, application No. 13/479,816, which is a continuation-in-part of application No. 13/198,074, filed on Aug. 4, 2011, now Pat. No. 8,515,361, which is a continuation-in-part of application No. 13/090,663, and a continuation-in-part of application No. 13/172,371, application No. 13/479,816, which is a continuation-in-part of application No. 13/226,831, filed on Sep. 7, 2011, which is a continuation-in-part of application No. 13/090,663, and a continuation-in-part of application No. 13/172,371, and a continuation-in-part of application No. 13/198,074.

(60) Provisional application No. 61/325,859, filed on Apr. 20, 2010, provisional application No. 61/359,487, filed on Jun. 29, 2010, provisional application No. 61/370,554, filed on Aug. 4, 2010, provisional application No. 61/380,522, filed on Sep. 7, 2010, provisional application No. 61/410,071, filed on Nov. 4, 2010, provisional application No. 61/417,633, filed on Nov. 29, 2010, provisional application No. 61/489,487, filed on May 24, 2011.

(51) **Int. Cl.**

H03F 1/02	(2006.01)
H03F 3/195	(2006.01)
H03F 3/21	(2006.01)
H03F 3/24	(2006.01)
H03F 3/60	(2006.01)
H03F 3/72	(2006.01)

(52) **U.S. Cl.**

CPC *H03F 2200/318* (2013.01); *H03F 2200/336* (2013.01); *H03F 2200/387* (2013.01); *H03F 2200/411* (2013.01); *H03F 2200/414* (2013.01); *H03F 2200/417* (2013.01); *H03F 2200/451* (2013.01); *H03F 2200/504* (2013.01); *H03F 2200/534* (2013.01); *H03F 2200/537* (2013.01); *H03F 2200/541* (2013.01); *H03F 2203/21106* (2013.01); *H03F 2203/21142* (2013.01); *H03F 2203/21157* (2013.01)

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,638,255	A	1/1987	Penney
4,819,081	A	4/1989	Volk et al.
5,212,459	A	5/1993	Ueda et al.
5,278,994	A	1/1994	Black et al.
5,307,512	A	4/1994	Mitzlaff
5,343,307	A	8/1994	Mizuno et al.
5,404,547	A	4/1995	Diamantstein et al.
5,432,473	A	7/1995	Mattila et al.
5,603,106	A	2/1997	Toda
5,636,114	A	6/1997	Bhagwat et al.
5,640,686	A	6/1997	Norimatsu
5,642,075	A	6/1997	Bell
5,652,547	A	7/1997	Mokhtar et al.
5,724,004	A	3/1998	Reif et al.
5,832,373	A	11/1998	Nakanishi et al.
5,841,319	A	11/1998	Sato
5,852,632	A	12/1998	Capici et al.
5,860,080	A	1/1999	James et al.
5,872,481	A	2/1999	Sevic et al.
5,874,841	A	2/1999	Majid et al.
5,920,808	A	7/1999	Jones et al.
5,923,153	A	7/1999	Liu
5,923,761	A	7/1999	Lodenius
5,945,870	A	8/1999	Chu et al.
5,956,246	A	9/1999	Sequeira et al.
6,051,963	A	4/2000	Eagar
6,064,272	A	5/2000	Rhee
6,151,509	A	11/2000	Chorey
6,192,225	B1	2/2001	Arpaia et al.
6,194,968	B1	2/2001	Winslow
6,229,366	B1	5/2001	Balakirshnan et al.
6,259,901	B1	7/2001	Shinomiya et al.
6,304,748	B1	10/2001	Li et al.
6,425,107	B1	7/2002	Caldara et al.
6,559,492	B1	5/2003	Hazucha et al.
6,606,483	B1	8/2003	Baker et al.
6,670,849	B1	12/2003	Damgaard et al.
6,674,789	B1	1/2004	Fardoun et al.
6,724,252	B2	4/2004	Ngo et al.
6,774,508	B2	8/2004	Ballantyne et al.
6,794,923	B2	9/2004	Burt et al.
6,806,768	B2	10/2004	Klaren et al.
6,853,244	B2	2/2005	Robinson et al.
6,888,482	B1	5/2005	Hertle
6,900,697	B1	5/2005	Doyle et al.
6,906,590	B2	6/2005	Amano
6,917,188	B2	7/2005	Kernahan
6,937,487	B1	8/2005	Bron
6,954,623	B2	10/2005	Chang et al.
6,969,978	B2	11/2005	Dening
6,998,914	B2	2/2006	Robinson
7,035,069	B2	4/2006	Takikawa et al.
7,043,213	B2	5/2006	Robinson et al.
7,058,374	B2	6/2006	Levesque et al.
7,072,626	B2	7/2006	Hadjichristos
7,075,346	B1	7/2006	Hariman et al.
7,098,728	B1	8/2006	Mei et al.
7,116,949	B2	10/2006	Irie et al.
7,145,385	B2	12/2006	Brandt et al.
7,148,749	B2	12/2006	Rahman et al.
7,154,336	B2	12/2006	Maeda
7,155,251	B2	12/2006	Saruwatari et al.
7,177,607	B2	2/2007	Weiss
7,180,373	B2	2/2007	Imai et al.
7,184,731	B2	2/2007	Kim
7,184,749	B2	2/2007	Marsh et al.
7,187,910	B2	3/2007	Kim et al.
7,202,734	B1	4/2007	Raab
7,248,111	B1	7/2007	Xu et al.
7,263,337	B2	8/2007	Struble
7,276,960	B2	10/2007	Peschke
7,295,814	B2	11/2007	Yamashita et al.
7,298,600	B2	11/2007	Takikawa et al.
7,299,015	B2	11/2007	Iwamiya et al.
7,324,787	B2	1/2008	Kurakami et al.
7,333,564	B2	2/2008	Sugiyama et al.
7,333,778	B2	2/2008	Pehlke et al.
7,342,455	B2	3/2008	Behzad et al.
7,358,807	B2	4/2008	Scuderi et al.
7,368,985	B2	5/2008	Kusunoki
7,372,333	B2	5/2008	Abedinpour et al.
7,408,330	B2	8/2008	Zhao
7,477,106	B2	1/2009	Van Bezooijen et al.
7,483,678	B2	1/2009	Rozenblit et al.
7,518,448	B1	4/2009	Blair et al.
7,529,523	B1	5/2009	Young et al.
7,539,462	B2	5/2009	Peckham et al.
7,551,688	B2	6/2009	Matero et al.
7,554,407	B2	6/2009	Hau et al.
7,558,539	B2	7/2009	Huynh et al.
7,580,443	B2	8/2009	Uemura et al.
7,622,900	B2	11/2009	Komiya
7,664,520	B2	2/2010	Gu
7,667,987	B2	2/2010	Huynh et al.
7,684,220	B2	3/2010	Fang et al.
7,689,182	B1	3/2010	Bosley et al.
7,701,290	B2	4/2010	Liu
7,702,300	B1	4/2010	McCune
7,706,756	B2	4/2010	Sato et al.
7,714,546	B2	5/2010	Kimura et al.
7,724,097	B2	5/2010	Carley et al.
7,768,354	B2	8/2010	Matsuda et al.

(56) References Cited

U.S. PATENT DOCUMENTS

7,782,141	B2	8/2010	Witmer et al.	
7,783,272	B2	8/2010	Magnusen	
7,787,570	B2	8/2010	Rozenblit et al.	
7,796,410	B2	9/2010	Takayanagi et al.	
7,859,511	B2	12/2010	Shen et al.	
7,860,466	B2	12/2010	Woo et al.	
7,876,159	B2	1/2011	Wang et al.	
7,907,430	B2	3/2011	Kularatna et al.	
7,941,110	B2	5/2011	Gonzalez	
7,995,984	B2	8/2011	Kudaishi et al.	
7,999,484	B2	8/2011	Jurngwirth et al.	
8,000,117	B2	8/2011	Petricek	
8,008,970	B1	8/2011	Homol et al.	
8,023,995	B2	9/2011	Kuriyama et al.	
8,031,003	B2	10/2011	Dishop	
8,085,106	B2	12/2011	Huda et al.	
8,089,323	B2	1/2012	Tarng et al.	
8,098,093	B1	1/2012	Li	
8,131,234	B2	3/2012	Liang et al.	
8,134,410	B1	3/2012	Zortea	
8,149,050	B2	4/2012	Cabanillas	
8,149,061	B2	4/2012	Schuurmans	
8,213,888	B2	7/2012	Kuriyama et al.	
8,228,122	B1	7/2012	Yuen et al.	
8,258,875	B1	9/2012	Smith et al.	
8,271,028	B2	9/2012	Rabjohn	
8,427,120	B1	4/2013	Cilio	
8,461,921	B2	6/2013	Pletcher et al.	
8,508,299	B2	8/2013	Kawano et al.	
8,514,025	B2	8/2013	Lesso	
8,546,980	B2	10/2013	Shimamoto et al.	
8,564,366	B2	10/2013	Namie et al.	
8,618,868	B2	12/2013	Khlat et al.	
8,624,576	B2	1/2014	Khlat et al.	
8,649,741	B2	2/2014	Iijima et al.	
2002/0055376	A1	5/2002	Norimatsu	
2002/0055378	A1	5/2002	Imel et al.	
2003/0006845	A1	1/2003	Lopez et al.	
2003/0042885	A1	3/2003	Zhou et al.	
2003/0073418	A1	4/2003	Dening et al.	
2003/0087626	A1	5/2003	Prikhodko et al.	
2003/0201674	A1	10/2003	Droppo et al.	
2003/0201834	A1	10/2003	Pehlke	
2003/0227280	A1	12/2003	Vinciarelli	
2004/0068673	A1	4/2004	Espinoza-Ibarra et al.	
2004/0090802	A1	5/2004	Pourseyed et al.	
2004/0095118	A1	5/2004	Kernahan	
2004/0127173	A1	7/2004	Leizerovich	
2004/0183507	A1	9/2004	Amei	
2004/0185805	A1	9/2004	Kim et al.	
2004/0192369	A1*	9/2004	Nilsson	455/522
2004/0222848	A1	11/2004	Shih et al.	
2004/0235438	A1	11/2004	Quilisch et al.	
2005/0003855	A1	1/2005	Wada et al.	
2005/0017787	A1	1/2005	Kojima	
2005/0046507	A1	3/2005	Dent	
2005/0064830	A1	3/2005	Grigore	
2005/0083715	A1	4/2005	Guillarme et al.	
2005/0088237	A1	4/2005	Gamero et al.	
2005/0110559	A1	5/2005	Farkas et al.	
2005/0134388	A1	6/2005	Jenkins	
2005/0136854	A1	6/2005	Akizuki et al.	
2005/0136866	A1	6/2005	Dupuis	
2005/0168281	A1	8/2005	Nagamori et al.	
2005/0200407	A1	9/2005	Arai et al.	
2005/0227644	A1	10/2005	Maslennikov et al.	
2005/0245214	A1	11/2005	Nakamura et al.	
2005/0280471	A1	12/2005	Matsushita et al.	
2005/0288052	A1	12/2005	Carter et al.	
2005/0289375	A1	12/2005	Ranganathan et al.	
2006/0006943	A1	1/2006	Clifton et al.	
2006/0017426	A1	1/2006	Yang et al.	
2006/0038710	A1*	2/2006	Staszewski et al.	341/143
2006/0046666	A1	3/2006	Hara et al.	
2006/0046668	A1	3/2006	Uratani et al.	
2006/0052065	A1	3/2006	Argaman et al.	
2006/0067425	A1*	3/2006	Windisch	375/296
2006/0084398	A1	4/2006	Chmiel et al.	
2006/0114075	A1*	6/2006	Janosevic et al.	332/150
2006/0119331	A1	6/2006	Jacobs et al.	
2006/0128325	A1	6/2006	Levesque et al.	
2006/0146956	A1	7/2006	Kim et al.	
2006/0199553	A1	9/2006	Kenington	
2006/0221646	A1	10/2006	Ye et al.	
2006/0226909	A1	10/2006	Abedinpour et al.	
2006/0290444	A1	12/2006	Chen	
2006/0293005	A1	12/2006	Hara et al.	
2007/0024360	A1	2/2007	Markowski	
2007/0026824	A1	2/2007	Ono et al.	
2007/0032201	A1	2/2007	Behzad et al.	
2007/0069820	A1	3/2007	Hayata et al.	
2007/0096806	A1	5/2007	Sorrells et al.	
2007/0096810	A1	5/2007	Hageman et al.	
2007/0129025	A1	6/2007	Vasa et al.	
2007/0146090	A1	6/2007	Carey et al.	
2007/0182490	A1	8/2007	Hau et al.	
2007/0210776	A1	9/2007	Oka	
2007/0222520	A1	9/2007	Inamori et al.	
2007/0249300	A1	10/2007	Sorrells et al.	
2007/0249304	A1	10/2007	Snelgrove et al.	
2007/0279021	A1	12/2007	Yanagida et al.	
2007/0281635	A1	12/2007	McCallister et al.	
2007/0291873	A1	12/2007	Saito et al.	
2008/0003950	A1	1/2008	Haapoja et al.	
2008/0008273	A1	1/2008	Kim et al.	
2008/0009248	A1	1/2008	Rozenblit et al.	
2008/0023825	A1	1/2008	Hebert et al.	
2008/0036532	A1	2/2008	Pan	
2008/0051044	A1	2/2008	Takehara	
2008/0057883	A1	3/2008	Pan	
2008/0081572	A1	4/2008	Rofougaran	
2008/0136559	A1	6/2008	Takahashi et al.	
2008/0157732	A1	7/2008	Williams	
2008/0169792	A1	7/2008	Orr	
2008/0205547	A1	8/2008	Rofougaran	
2008/0233913	A1	9/2008	Sivasubramaniam	
2008/0278136	A1	11/2008	Murtojarvi	
2008/0278236	A1	11/2008	Seymour	
2009/0004981	A1	1/2009	Eliezer et al.	
2009/0011787	A1	1/2009	Kikuma	
2009/0021302	A1	1/2009	Elia	
2009/0059630	A1	3/2009	Williams	
2009/0068966	A1	3/2009	Drogi et al.	
2009/0104900	A1	4/2009	Lee	
2009/0115520	A1	5/2009	Ripley et al.	
2009/0153250	A1	6/2009	Rofougaran	
2009/0163153	A1	6/2009	Senda et al.	
2009/0163157	A1	6/2009	Zolfaghari	
2009/0176464	A1	7/2009	Liang et al.	
2009/0191826	A1	7/2009	Takinami et al.	
2009/0258611	A1	10/2009	Nakamura et al.	
2009/0264091	A1	10/2009	Jensen et al.	
2009/0274207	A1	11/2009	O'Keeffe et al.	
2009/0285331	A1	11/2009	Sugar et al.	
2009/0289719	A1	11/2009	Van Bezooijen et al.	
2009/0311980	A1	12/2009	Sjoland	
2009/0322304	A1	12/2009	Oraw et al.	
2010/0007412	A1	1/2010	Wang et al.	
2010/0007414	A1	1/2010	Searle et al.	
2010/0007433	A1	1/2010	Jensen	
2010/0013548	A1	1/2010	Barrow	
2010/0020899	A1	1/2010	Szopko et al.	
2010/0027596	A1	2/2010	Bellaouar et al.	
2010/0029224	A1	2/2010	Urushihara et al.	
2010/0052794	A1	3/2010	Rofougaran	
2010/0097104	A1	4/2010	Yang et al.	
2010/0102789	A1	4/2010	Randall	
2010/0109561	A1	5/2010	Chen et al.	
2010/0120384	A1	5/2010	Pennec	
2010/0120475	A1	5/2010	Taniuchi et al.	
2010/0123447	A1	5/2010	Vecera et al.	
2010/0127781	A1	5/2010	Inamori et al.	
2010/0128689	A1*	5/2010	Yoon et al.	370/329
2010/0164579	A1	7/2010	Acatrinei	

(56) **References Cited**

U.S. PATENT DOCUMENTS

2010/0176869	A1	7/2010	Horie et al.
2010/0181980	A1	7/2010	Richardson
2010/0189042	A1	7/2010	Pan
2010/0222015	A1	9/2010	Shimizu et al.
2010/0233977	A1	9/2010	Minnis et al.
2010/0237944	A1	9/2010	Pierdomenico et al.
2010/0244788	A1	9/2010	Chen
2010/0273535	A1	10/2010	Inamori et al.
2010/0291888	A1	11/2010	Hadjichristos et al.
2010/0295599	A1	11/2010	Uehara et al.
2010/0311362	A1	12/2010	Lee et al.
2011/0018516	A1	1/2011	Notman et al.
2011/0018632	A1	1/2011	Pletcher et al.
2011/0018640	A1	1/2011	Liang et al.
2011/0032030	A1	2/2011	Ripley et al.
2011/0051842	A1	3/2011	van der Heijden et al.
2011/0068768	A1	3/2011	Chen et al.
2011/0068873	A1	3/2011	Alidio et al.
2011/0075772	A1	3/2011	Sethi et al.
2011/0080205	A1	4/2011	Lee
2011/0095735	A1	4/2011	Lin
2011/0123048	A1	5/2011	Wang et al.
2011/0136452	A1	6/2011	Pratt et al.
2011/0181115	A1	7/2011	Ivanov
2011/0234187	A1	9/2011	Brown et al.
2011/0273152	A1	11/2011	Weir
2011/0294445	A1	12/2011	Goto et al.
2011/0298538	A1	12/2011	Andrys et al.
2011/0309884	A1	12/2011	Dishop
2011/0312287	A1	12/2011	Ramachandran et al.
2012/0044022	A1	2/2012	Walker et al.
2012/0064953	A1	3/2012	Dagher et al.
2012/0117284	A1	5/2012	Southcombe et al.
2012/0170690	A1	7/2012	Ngo et al.
2012/0182086	A1	7/2012	Hase et al.
2012/0229210	A1	9/2012	Jones et al.
2012/0235736	A1	9/2012	Levesque et al.
2012/0236958	A1	9/2012	Deng et al.
2012/0242413	A1	9/2012	Lesso
2012/0252382	A1	10/2012	Bashir et al.
2013/0005286	A1	1/2013	Chan et al.
2013/0307616	A1	11/2013	Berchtold et al.
2013/0342270	A1	12/2013	Baxter et al.
2013/0344828	A1	12/2013	Baxter et al.
2013/0344833	A1	12/2013	Baxter et al.
2014/0119070	A1	5/2014	Jeong et al.

OTHER PUBLICATIONS

Non-Final Office Action for U.S. Appl. No. 12/567,318, mailed Apr. 2, 2013, 5 pages.

Notice of Allowance for U.S. Appl. No. 13/198,074, mailed Apr. 12, 2013, 8 pages.

International Search Report and Written Opinion for PCT/US2011/050633, mailed Mar. 8, 2013, 23 pages.

International Preliminary Report on Patentability for PCT/US20111050633, mailed Mar. 28, 2013, 17 pages.

Non-Final Office Action for U.S. Appl. No. 13/287,726, mailed May 16, 2013, 9 pages.

Non-Final Office Action for U.S. Appl. No. 13/288,517, mailed May 16, 2013, 9 pages.

Non-Final Office Action for U.S. Appl. No. 13/226,843, mailed Mar. 4, 2013, 6 pages.

Non-Final Office Action for U.S. Appl. No. 13/288,373, mailed Feb. 25, 2013, 6 pages.

Non-Final Office Action for U.S. Appl. No. 13/289,379, mailed Feb. 25, 2013, 9 pages.

Non-Final Office Action for U.S. Appl. No. 13/305,763, mailed Mar. 8, 2013, 10 pages.

Notice of Allowance for U.S. Appl. No. 13/304,762, mailed Mar. 5, 2013, 7 pages.

Notice of Allowance for U.S. Appl. No. 13/226,797, mailed Apr. 26, 2013, 8 pages.

Bastida, E.M. et al., “Cascadable Monolithic Balanced Amplifiers at Microwave Frequencies,” 10th European Microwave Conference, Sep. 8-12, 1980, pp. 603-607, IEEE.

Berretta, G. et al, “A balanced CDMA2000 SiGe HBT load insensitive power amplifier,” 2006 IEEE Radio and Wireless Symposium, Jan. 17-19, 2006, pp. 523-526, IEEE.

Grebennikov, A. et al., “High-Efficiency Balanced Switched-Path Monolithic SiGe HBT Power Amplifiers for Wireless Applications,” Proceedings of the 2nd European Microwave Integrated Circuits Conference, 2007, pp. 391-394, IEEE.

Grebennikov, A., “Circuit Design Technique for High Efficiency Class F Amplifiers,” 2000 IEEE International Microwave Symposium Digest, 2000, pp. 771-774, vol. 2, IEEE.

Kurokawa, K., “Design Theory of Balanced Transistor Amplifiers,” Bell System Technical Journal, Oct. 1965, pp. 1675-1698, vol. 44, Bell Labs.

Li et al., “LTE power amplifier module design: challenges and trends,” IEEE International Conference on Solid-State and Integrated Circuit Technology, Nov. 2010, pp. 192-195.

Mandeep, A., “A Compact, Balanced Low Noise Amplifier for WiMAX Base Station Applications”, Microwave Journal, Nov. 2010, p. 84-92, vol. 53, No. 11, Microwave Journal and Horizon House Publications.

Podcameni, A.B. et al., “An amplifier linearization method based on a quadrature balanced structure,” IEEE Transactions on Broadcasting, Jun. 2002, p. 158-162, vol. 48, No. 2, IEEE.

Scuderi, A. et al., “Balanced SiGe PA Module for Multi-Band and Multi-Mode Cellular-Phone Applications,” Digest of Technical Papers IEEE 2008 International Solid-State Circuits Conference, Feb. 3-7, 2008, pp. 572-637, IEEE.

Zhang, G. et al., “A high performance Balanced Power Amplifier and Its Integration into a Front-end Module at PCS Band,” 2007 IEEE Radio Frequency Integrated Circuits Symposium, Jun. 3-5, 2007, p. 251-254, IEEE.

Zhang, G. et al., “Dual mode efficiency enhanced linear power amplifiers using a new balanced structure,” 2009 IEEE Radio Frequency Integrated Circuits Symposium, Jun. 7-9, 2009, pp. 245-248, IEEE.

MIPI Alliance Specification for RF Front-End Control Interface, Version 1.00.00, May 3, 2010, 2009-2010 MIPI Alliance, Inc.

Wang, P. et al., “A 2.4-GHz +25dBm P-1dB linear power amplifier with dynamic bias control in a 65-nm CMOS process,” 2008 European Solid-State Circuits Conference, Sep. 15-19, 2008, pp. 490-493.

Notice of Allowance for U.S. Appl. No. 13/090,663 mailed Nov. 28, 2012, 15 pages.

Non-final Office Action for U.S. Appl. No. 13/288,478 mailed Dec. 26, 2012, 9 pages.

Non-final Office Action for U.S. Appl. No. 13/288,517 mailed Dec. 11, 2012, 10 pages.

Notice of Allowance for U.S. Appl. No. 13/304,762 mailed Nov. 27, 2012, 8 pages.

Invitation to pay additional fees and, where applicable, protest fee for PCT/US2011/050633 mailed Aug. 22, 2012, 7 pages.

Unknown, Author, “SKY77344-21 Power Amplifier Module—Evaluation Information,” Skyworks, Version-21 Feb. 16, 2010, 21 pages.

Author Unknown, “60mA, 5.0V, Buck/Boost Charge Pump in ThinSOT-23 and ThinQFN”, Texas Instruments Incorporated, REG710, SBAS221F, Dec. 2001, revised Mar. 2008, 23 pages.

Author Unknown, “DC-to-DC Converter Combats EMI,” Maxim Integrated Products, Application Note 1077, May 28, 2002, 4 pages, http://www.maxim-ic.com/an1077/.

Non-Final Office Action for U.S. Appl. No. 11/756,909, mailed May 15, 2009, 11 pages.

Final Office Action for U.S. Appl. No. 11/756,909, mailed Nov. 18, 2009, 14 pages.

Notice of Allowance for U.S. Appl. No. 11/756,909, mailed Dec. 23, 2010, 7 pages.

Non-Final Office Action for U.S. Appl. No. 12/567,318, mailed May 29, 2012, 7 pages.

Final Office Action for U.S. Appl. No. 12/567,318, mailed Oct. 22, 2012, 7 pages.

Non-Final Office Action for U.S. Appl. No. 12/723,738, mailed Dec. 20, 2012, 7 pages.

(56) References Cited

OTHER PUBLICATIONS

Quayle Action for U.S. Appl. No. 13/198,074, mailed Jan. 22, 2013, 5 pages.

Non-Final Office Action for U.S. Appl. No. 13/289,134, mailed Feb. 6, 2013, 13 pages.

Non-Final Office Action for U.S. Appl. No. 13/287,726, mailed Jan. 25, 2013, 11 pages.

Non-Final Office Action for U.S. Appl. No. 13/287,735, mailed Jan. 25, 2013, 11 pages.

Non-Final Office Action for U.S. Appl. No. 13/288,318, mailed Feb. 5, 2013, 12 pages.

Non-Final Office Action for U.S. Appl. No. 13/288,273, mailed Feb. 5, 2013, 8 pages.

Non-Final Office Action for U.S. Appl. No. 13/304,744, mailed Jan. 24, 2013, 10 pages.

Final Office Action for U.S. Appl. No. 12/567,318, mailed Jul. 19, 2013, 7 pages.

Notice of Allowance for U.S. Appl. No. 13/289,134, mailed Jun. 6, 2013, 8 pages.

Non-Final Office Action for U.S. Appl. No. 13/287,713, mailed Aug. 5, 2013, 7 pages.

Notice of Allowance for U.S. Appl. No. 13/287,735, mailed May 28, 2013, 8 pages.

Non-Final Office Action for U.S. Appl. No. 13/288,318, mailed Jun. 3, 2013, 14 pages.

Non-Final Office Action for U.S. Appl. No. 13/288,478, mailed Jun. 3, 2013, 9 pages.

Final Office Action for U.S. Appl. No. 13/226,843, mailed Jul. 10, 2013, 7 pages.

Non-Final Office Action for U.S. Appl. No. 13/288,273, mailed May 30, 2013, 11 pages.

Final Office Action for U.S. Appl. No. 13/288,373, mailed Aug. 2, 2013, 7 pages.

Notice of Allowance for U.S. Appl. No. 13/289,379, mailed Jun. 6, 2013, 9 pages.

Non-Final Office Action for U.S. Appl. No. 13/304,735, mailed Jul. 11, 2013, 8 pages.

Non-Final Office Action for U.S. Appl. No. 13/304,796, mailed Jul. 17, 2013, 8 pages.

Final Office Action for U.S. Appl. No. 13/304,744, mailed May 30, 2013, 12 pages.

Advisory Action for U.S. Appl. No. 13/304,744, mailed Aug. 2, 2013, 3 pages.

Final Office Action for U.S. Appl. No. 13/305,763, mailed Jun. 24, 2013, 13 pages.

Non-Final Office Action for U.S. Appl. No. 13/304,943, mailed Jul. 23, 2013, 8 pages.

Notice of Allowance for U.S. Appl. No. 13/226,777, mailed May 28, 2013, 8 pages.

Non-Final Office Action for U.S. Appl. No. 13/226,814, mailed Jun. 13, 2013, 13 pages.

Notice of Allowance for U.S. Appl. No. 12/774,155, mailed Jun. 20, 2014, 8 pages.

Non-Final Office Action for U.S. Appl. No. 14/010,617, mailed Jul. 16, 2014, 6 pages.

Non-Final Office Action for U.S. Appl. No. 14/010,643, mailed Jul. 18, 2014, 6 pages.

Non-Final Office Action for U.S. Appl. No. 13/172,371, mailed Jun. 16, 2014, 7 pages.

Notice of Allowance for U.S. Appl. No. 13/287,726, mailed Aug. 4, 2014, 7 pages.

Non-Final Office Action for U.S. Appl. No. 13/287,672, mailed Jul. 28, 2014, 12 pages.

Non-Final Office Action for U.S. Appl. No. 13/289,302, mailed Jun. 16, 2014, 7 pages.

Notice of Allowance for U.S. Appl. No. 13/304,762, mailed May 29, 2014, 7 pages.

Non-Final Office Action for U.S. Appl. No. 13/773,888, mailed Jun. 10, 2014, 15 pages.

Non-Final Office Action for U.S. Appl. No. 12/774,155, mailed Dec. 4, 2013, 18 pages.

Final Office Action for U.S. Appl. No. 13/287,713, mailed Dec. 6, 2013, 9 pages.

Notice of Allowance for U.S. Appl. No. 13/288,478, mailed Nov. 18, 2013, 9 pages.

Non-Final Office Action for U.S. Appl. No. 13/288,517, mailed Oct. 31, 2013, 10 pages.

Non-Final Office Action for U.S. Appl. No. 13/288,373, mailed Nov. 19, 2013, 5 pages.

Non-Final Office Action for U.S. Appl. No. 13/288,590, mailed Dec. 5, 2013, 8 pages.

Notice of Allowance for U.S. Appl. No. 13/304,735, mailed Jan. 2, 2014, 8 pages.

Notice of Allowance for U.S. Appl. No. 13/304,796, mailed Dec. 5, 2013, 9 pages.

Notice of Allowance for U.S. Appl. No. 13/304,943, mailed Dec. 5, 2013, 9 pages.

Advisory Action for U.S. Appl. No. 13/226,814, mailed Dec. 31, 2013, 3 pages.

Non-Final Office Action for U.S. Appl. No. 13/656,997, mailed Jan. 13, 2014, 6 pages.

Notice of Allowance for U.S. Appl. No. 12/567,318, mailed Feb. 18, 2014, 8 pages.

Non-Final Office Action for U.S. Appl. No. 12/723,738, mailed Apr. 28, 2014, 14 pages.

Advisory Action for U.S. Appl. No. 13/287,713, mailed Feb. 20, 2014, 4 pages.

Non-Final Office Action for U.S. Appl. No. 13/288,517, mailed Apr. 28, 2014, 10 pages.

Non-Final Office Action for U.S. Appl. No. 13/226,843, mailed Mar. 31, 2014, 7 pages.

Notice of Allowance for U.S. Appl. No. 13/288,273, mailed Apr. 25, 2014, 7 pages.

Notice of Allowance for U.S. Appl. No. 13/288,373, mailed May 7, 2014, 7 pages.

Non-Final Office Action for U.S. Appl. No. 13/288,590, mailed May 8, 2014, 11 pages.

Notice of Allowance for U.S. Appl. No. 13/304,762, mailed Apr. 16, 2014, 7 pages.

Final Office Action for U.S. Appl. No. 13/226,777, mailed Mar. 21, 2014, 13 pages.

Non-Final Office Action for U.S. Appl. No. 13/656,997, mailed Apr. 30, 2014, 8 pages.

Author Unknown , "SKY77344-21 Power Amplifier Module—Evaluation Information," Skyworks, Version 21, Feb. 16, 2010, 21 pages.

Noriega, Fernando et al., "Designing LC Wilkinson power splitters," RF interconnects/interfaces, Aug. 2002, pp. 18, 20, 22, and 24, www.rfdesign.com.

Pampichai, Samphan et al., "A 3-dB Lumped-Distributed Miniaturized Wilkinson Divider," Electrical Engineering Conference (EECON-23), Nov. 2000, pp. 329-332.

Non-Final Office Action for U.S. Appl. No. 12/433,377, mailed Jun. 16, 2011, 7 pages.

Notice of Allowance for U.S. Appl. No. 12/433,377, mailed Oct. 31, 2011, 8 pages.

Non-Final Office Action for U.S. Appl. No. 12/774,155, mailed Jun. 21, 2012, 18 pages.

Final Office Action for U.S. Appl. No. 12/774,155, mailed Jan. 31, 2013, 15 pages.

Final Office Action for U.S. Appl. No. 12/774,155, mailed Apr. 18, 2013, 15 pages.

Advisory Action for U.S. Appl. No. 12/774,155, mailed Jun. 4, 2013, 3 pages.

Non-Final Office Action for U.S. Appl. No. 12/774,155, mailed Aug. 15, 2013, 15 pages.

Non-Final Office Action for U.S. Appl. No. 12/749,091, mailed Mar. 25, 2013, 9 pages.

Notice of Allowance for U.S. Appl. No. 12/749,091, mailed May 20, 2013, 9 pages.

Notice of Allowance for U.S. Appl. No. 12/773,292, mailed Feb. 22, 2012, 11 pages.

(56) **References Cited**

OTHER PUBLICATIONS

Notice of Allowance for U.S. Appl. No. 12/773,292, mailed Jul. 16, 2012, 12 pages.
Non-Final Office Action for U.S. Appl. No. 13/019,077, mailed Feb. 19, 2013, 9 pages.
Notice of Allowance for U.S. Appl. No. 13/019,077, mailed May 24, 2013, 9 pages.
Notice of Allowance for U.S. Appl. No. 13/288,318, mailed Oct. 24, 2013, 9 pages.
Notice of Allowance for U.S. Appl. No. 13/288,273 mailed Oct. 24, 2013, 9 pages.
Non-Final Office Action for U.S. Appl. No. 12/567,318, mailed Oct. 24, 2013, 6 pages.
Advisory Action for U.S. Appl. No. 13/304,744, mailed Sep. 13, 2013, 3 pages.
Non-Final Office Action for U.S. Appl. No. 13/304,744, mailed Oct. 21, 2013, 12 pages.
Notice of Allowance for U.S. Appl. No. 13/305,763, mailed Sep. 16, 2013, 6 pages.
Non-Final Office Action for U.S. Appl. No. 13/226,777, mailed Oct. 18, 2013, 10 pages.
Final Office Action for U.S. Appl. No. 13/226,814, mailed Oct. 23, 2013, 21 pages.
Advisory Action for U.S. Appl. No. 12/567,318, mailed Aug. 27, 2013, 3 pages.
Advisory Action for U.S. Appl. No. 13/288,373, mailed Oct. 15, 2013, 3 pages.
Advisory Action for U.S. Appl. No. 13/226,843, mailed Sep. 17, 2013, 3 pages.
Non-Final Office Action for U.S. Appl. No. 13/226,843, mailed Oct. 29, 2013, 7 pages.
Notice of Allowance for U.S. Appl. No. 13/287,726, mailed Oct. 7, 2013, 9 pages.
Non-Final Office Action for U.S. Appl. No. 12/723,738, mailed Dec. 10, 2014, 11 pages.
Notice of Allowance for U.S. Appl. No. 14/010,617, mailed Dec. 16, 2014, 7 pages.
Notice of Allowance for U.S. Appl. No. 14/010,630, mailed Dec. 31, 2014, 9 pages.
Non-Final Office Action for U.S. Appl. No. 14/010,643, mailed Dec. 9, 2014, 7 pages.
Notice of Allowance for U.S. Appl. No. 13/911,526, mailed Dec. 12, 2014, 9 pages.
Final Office Action for U.S. Appl. No. 13/287,672, mailed Dec. 8, 2014, 14 pages.
Final Office Action for U.S. Appl. No. 13/773,888, mailed Dec. 26, 2014, 18 pages.
Final Office Action for U.S. Appl. No. 12/723,738, mailed Aug. 11, 2014, 10 pages.
Non-Final Office Action for U.S. Appl. No. 14/010,630, mailed Aug. 6, 2014, 7 pages.
Notice of Allowance for U.S. Appl. No. 13/288,517, mailed Aug. 15, 2014, 7 pages.
Notice of Allowance for U.S. Appl. No. 13/656,997, mailed Sep. 2, 2014, 7 pages.
Notice of Allowance for U.S. Appl. No. 13/761,500, mailed Sep. 19, 2014, 7 pages.
Author Unknown, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception (Release 10)," 3GPP TS 36.101, V10.2.1, Apr. 2011, 225 pages.
Non-Final Office Action for U.S. Appl. No. 13/226,831, mailed Nov. 3, 2014, 12 pages.
Non-Final Office Action for U.S. Appl. No. 13/754,303, mailed Oct. 14, 2014, 14 pages.
Non-Final Office Action for U.S. Appl. No. 13/845,410, mailed Oct. 2, 2014, 5 pages.
Final Office Action for U.S. Appl. No. 14/010,643, mailed May 5, 2015, 8 pages.
Advisory Action for U.S. Appl. No. 13/226,831 mailed May 13, 2015, 3 pages.
Final Office Action for U.S. Appl. No. 12/723,738, mailed Mar. 20, 2015, 11 pages.
Final Office Action for U.S. Appl. No. 13/226,831, mailed Mar. 6, 2015, 6 pages.
Non-Final Office Action for U.S. Appl. No. 13/937,810, mailed Mar. 5, 2015, 5 pages.
Non-Final Office Action for U.S. Appl. No. 13/287,672, mailed Mar. 23, 2015, 7 pages.
Notice of Allowance for U.S. Appl. No. 13/754,303, mailed Feb. 17, 2015, 8 pages.
Non-Final Office Action for U.S. Appl. No. 13/911,428, mailed Jul. 22, 2015, 10 pages.
Notice of Allowance for U.S. Appl. No. 13/226,831, mailed Aug. 3, 2015, 9 pages.
Final Office Action for U.S. Appl. No. 13/287,672, mailed Jul. 17, 2015, 8 pages.
Advisory Action for U.S. Appl. No. 13/287,672, mailed Sep. 28, 2015, 3 pages.
Non-Final Office Action for U.S. Appl. No. 13/773,888, mailed Aug. 27, 2015, 17 pages.
Notice of Allowance for U.S. Appl. No. 14/010,643, mailed Jul. 10, 2015, 7 pages.

* cited by examiner

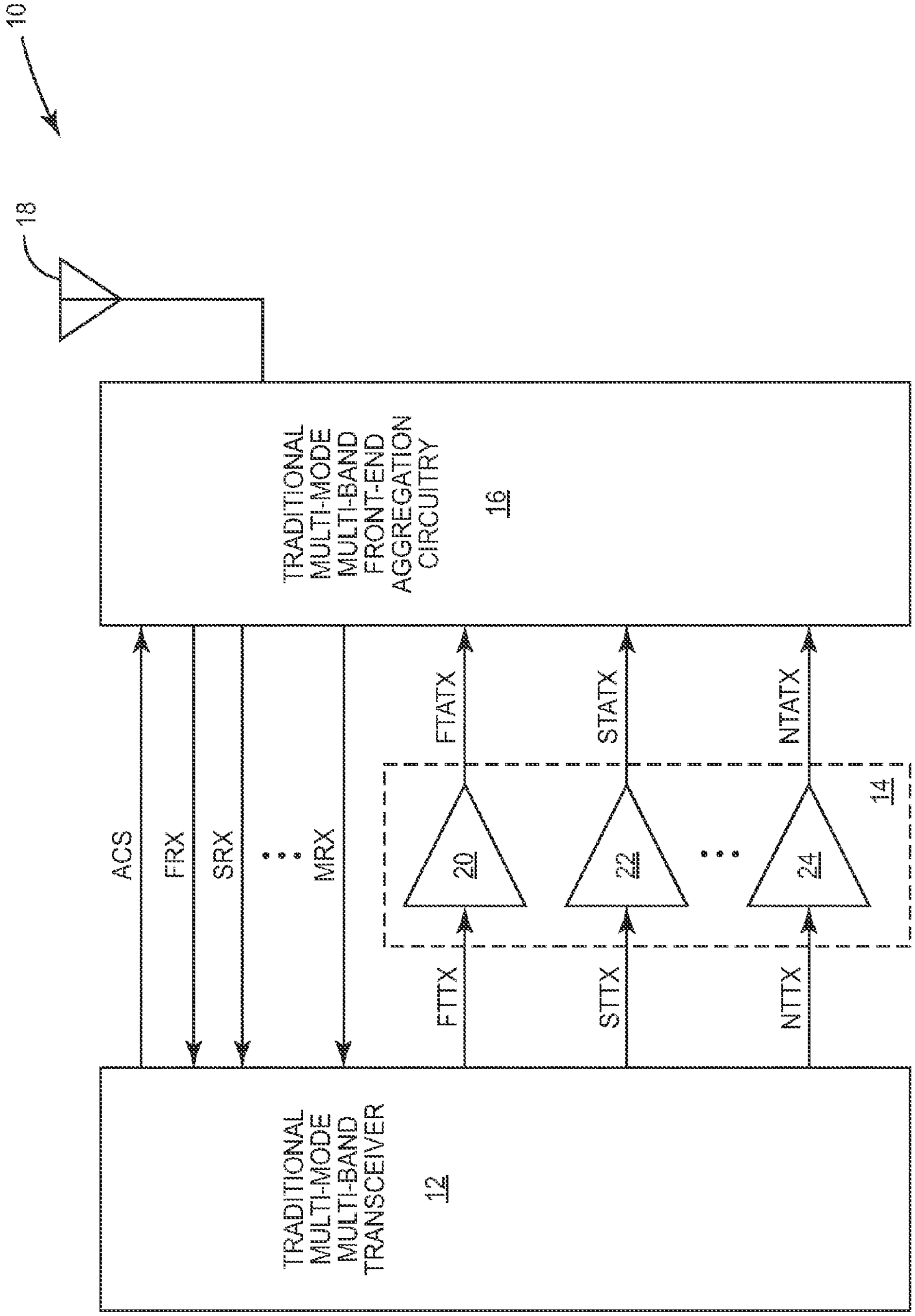
10
18
TRADITIONAL MULTI-MODE MULTI-BAND FRONT-END AGGREGATION CIRCUITRY
16
ACS
FRX
SRX
MRX
FTATX
STATX
NTATX
20
22
24
14
FTTX
STTX
NTTX
TRADITIONAL MULTI-MODE MULTI-BAND TRANSCEIVER
12

FIG. 1
PRIOR ART

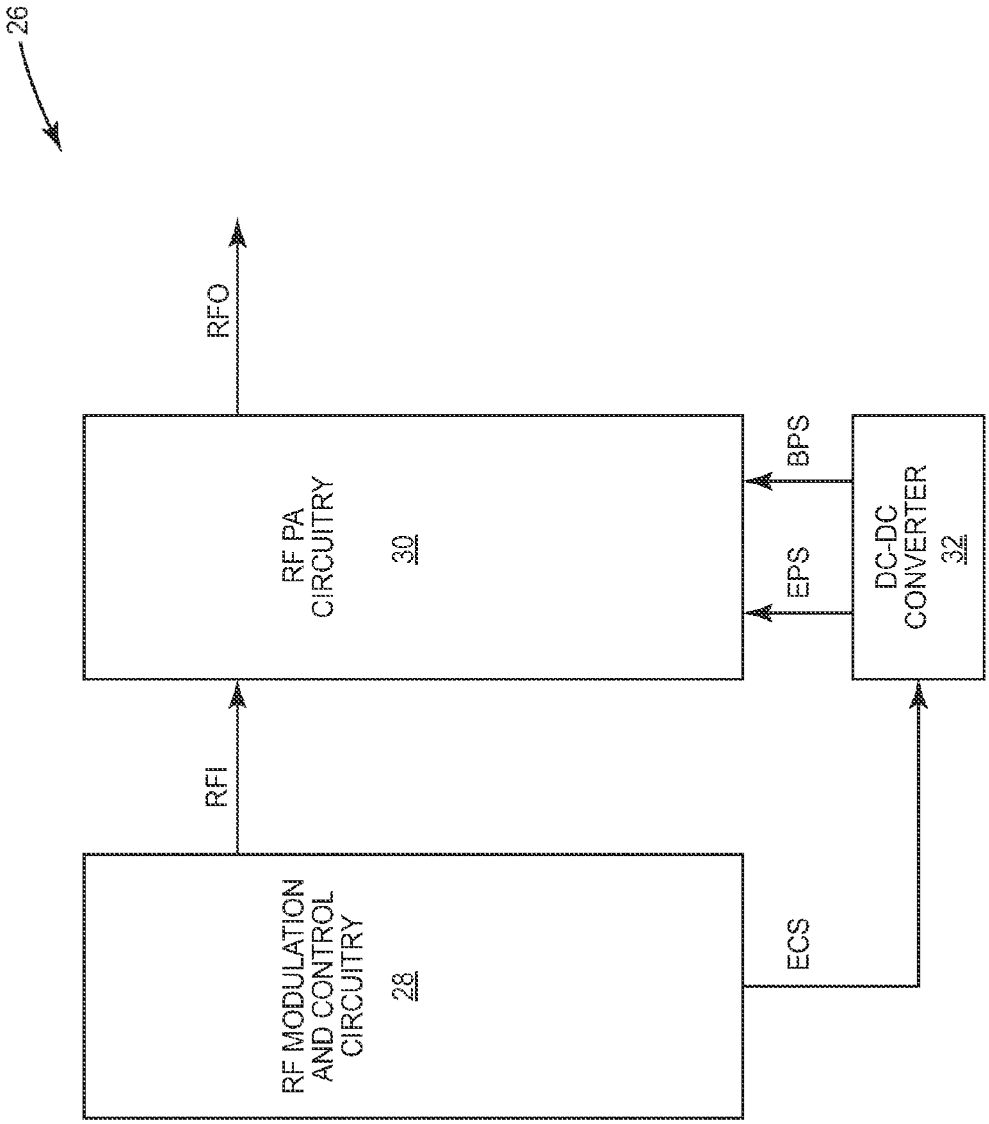
26
RFO
RF PA CIRCUITRY
30
BPS
EPS
DC-DC CONVERTER
32
RFI
RF MODULATION AND CONTROL CIRCUITRY
28
ECS

FIG. 2

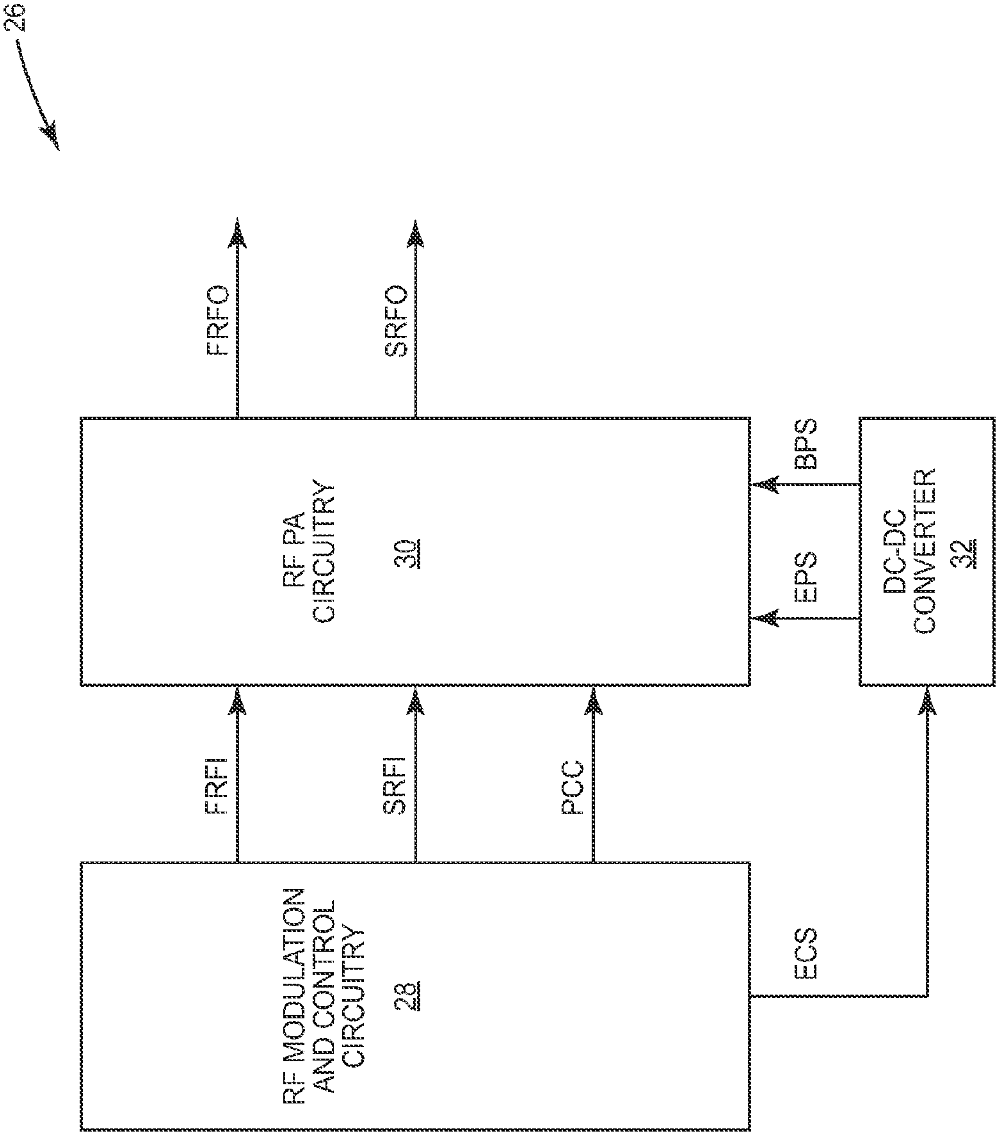
26
RF MODULATION AND CONTROL CIRCUITRY
28
FRFI
SRFI
PCC
ECS
RF PA CIRCUITRY
30
FRFO
SRFO
EPS
BPS
DC-DC CONVERTER
32

FIG. 3

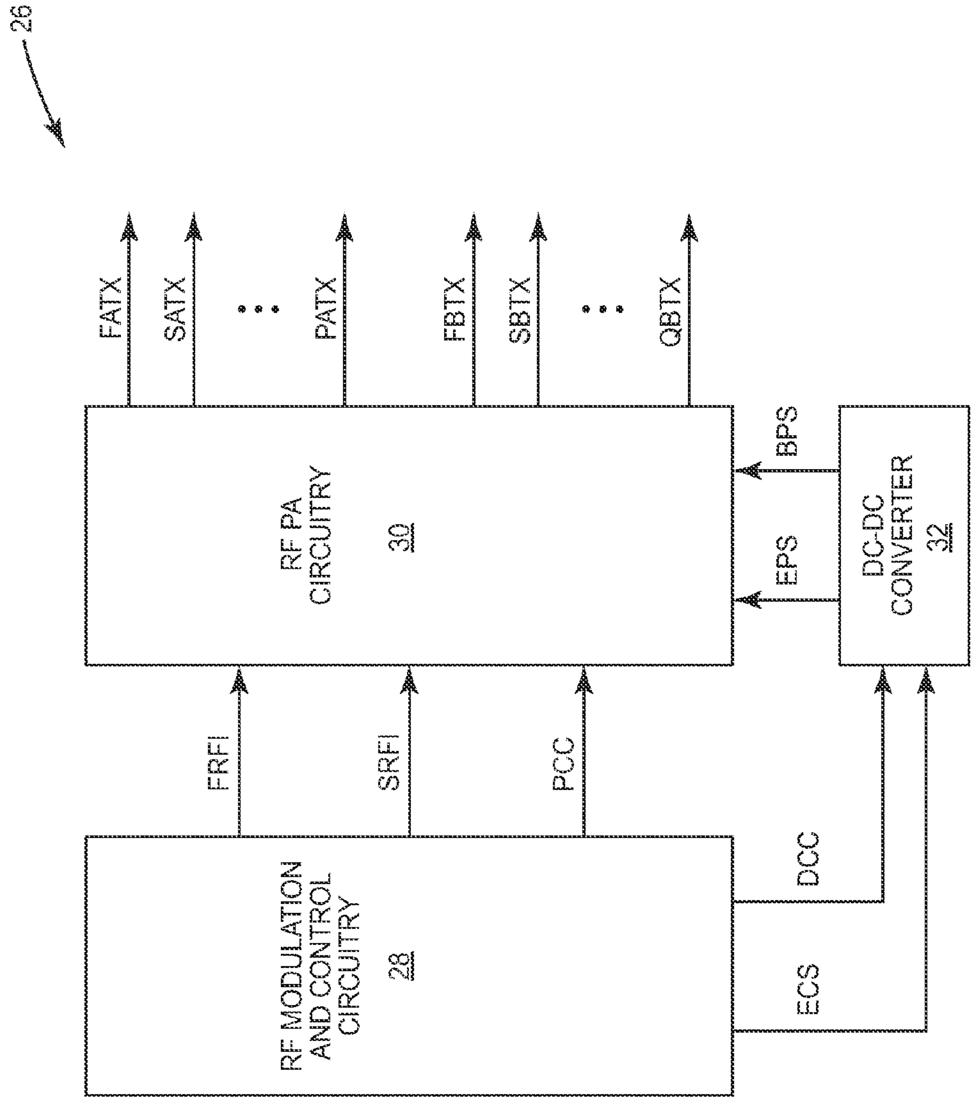
26
RF MODULATION AND CONTROL CIRCUITRY
28
FRFI
SRFI
PCC
DCC
ECS
RF PA CIRCUITRY
30
EPS
BPS
DC-DC CONVERTER
32
FATX
SATX
• • •
PATX
FBTX
SBTX
• • •
QBTX

FIG. 4

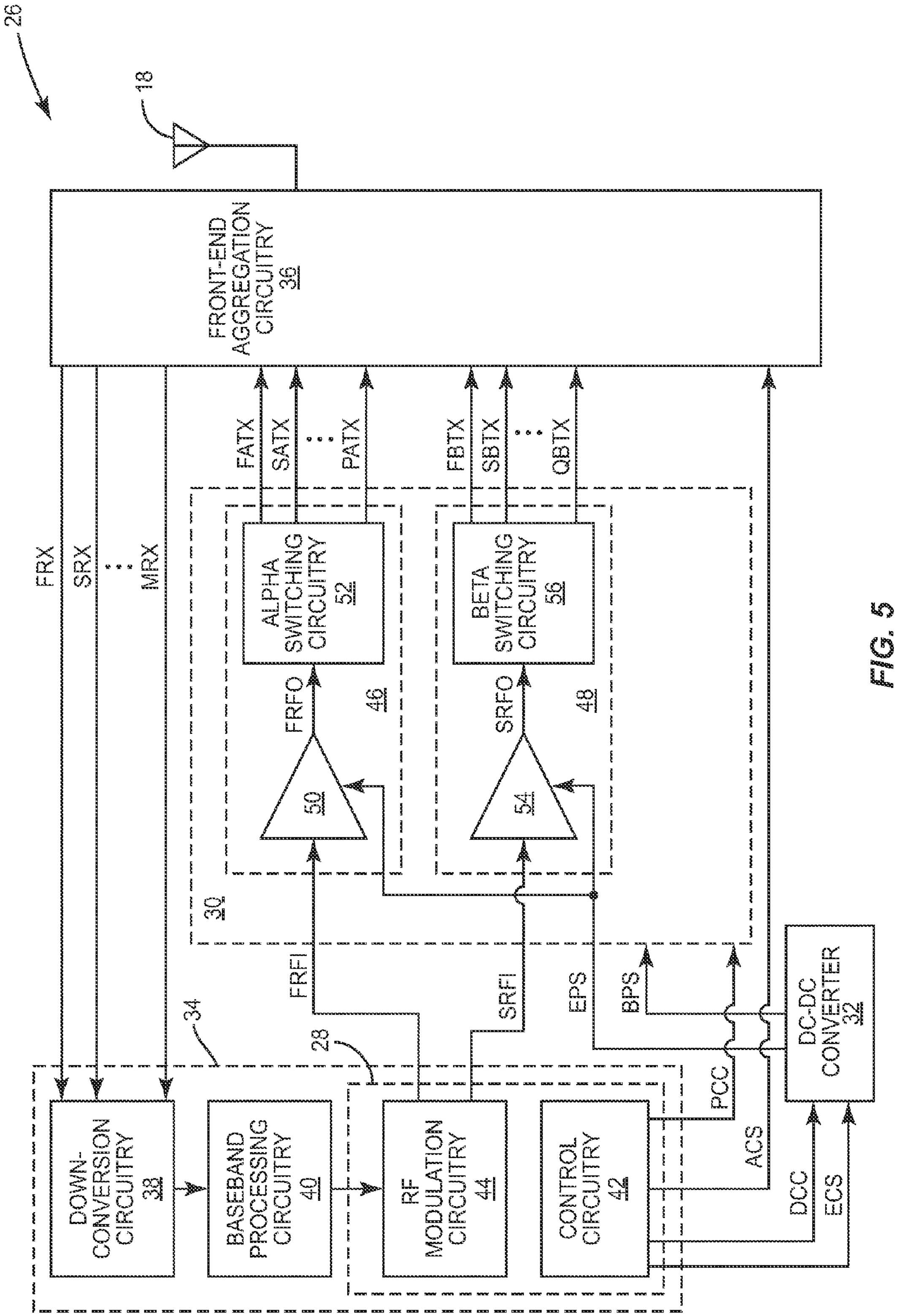
26
18
FRONT-END AGGREGATION CIRCUITRY 36
FRX
SRX
MRX
FATX
SATX
PATX
FBTX
SBTX
QBTX
ALPHA SWITCHING CIRCUITRY 52
BETA SWITCHING CIRCUITRY 56
FRFO
SRFO
46
48
50
54
30
FRFI
SRFI
EPS
BPS
DC-DC CONVERTER 32
34
28
PCC
ACS
DCC
ECS
DOWN-CONVERSION CIRCUITRY 38
BASEBAND PROCESSING CIRCUITRY 40
RF MODULATION CIRCUITRY 44
CONTROL CIRCUITRY 42

FIG. 5

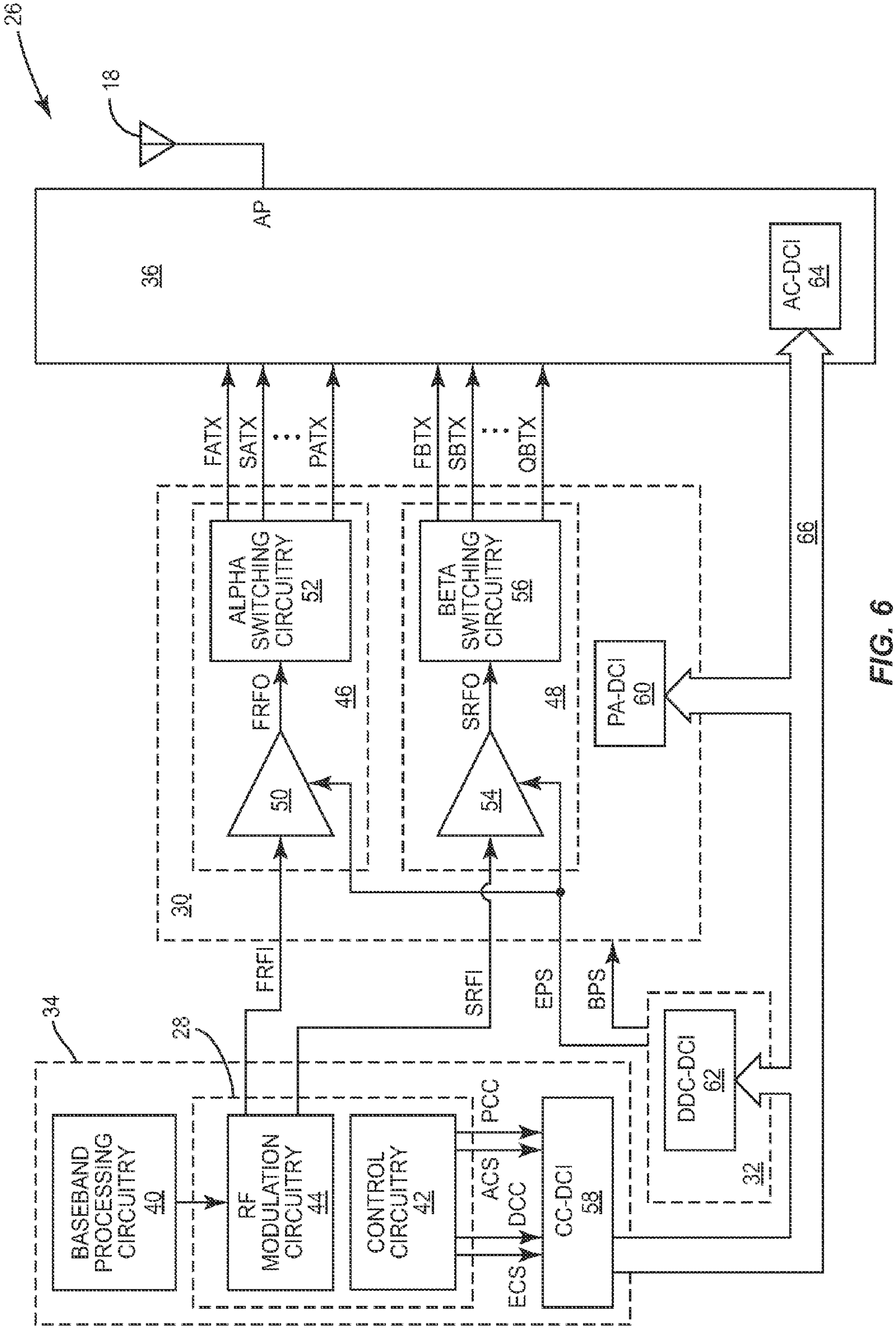
26
18
AP
36
AC-DCI
64
FATX
SATX
PATX
FBTX
SBTX
QBTX
ALPHA SWITCHING CIRCUITRY
52
BETA SWITCHING CIRCUITRY
56
FRFO
SRFO
46
48
50
54
PA-DCI
60
66
30
FRFI
SRFI
EPS
BPS
34
28
DDC-DCI
62
32
BASEBAND PROCESSING CIRCUITRY
40
RF MODULATION CIRCUITRY
44
CONTROL CIRCUITRY
42
PCC
ACS
DCC
ECS
CC-DCI
58

FIG. 6

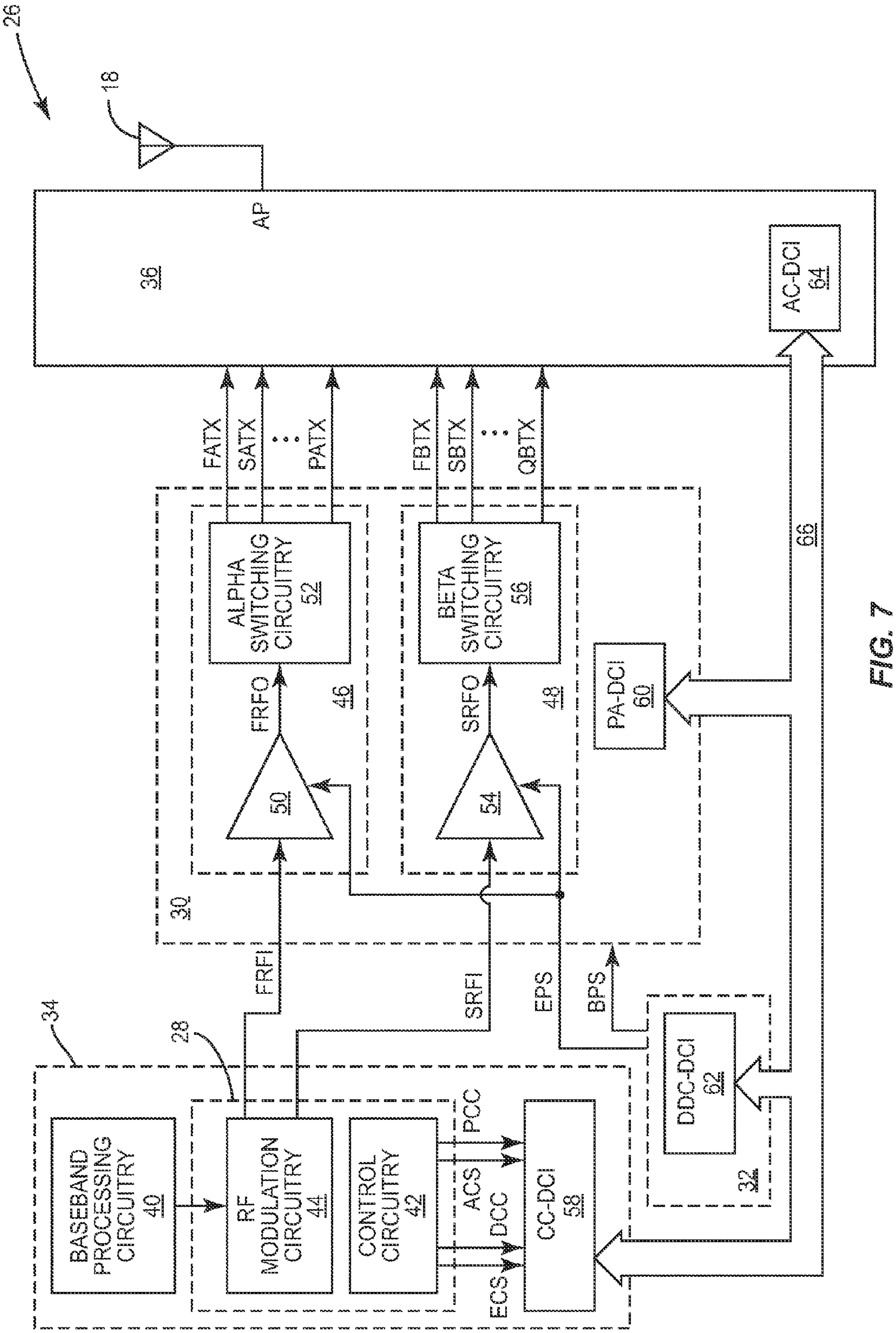
26
18
AP
36
AC-DCI
64
FATX
SATX
PATX
FBTX
SBTX
QBTX
ALPHA SWITCHING CIRCUITRY
52
BETA SWITCHING CIRCUITRY
56
66
PA-DCI
60
FRFO
46
SRFO
48
50
54
30
FRFI
SRFI
EPS
BPS
DDC-DCI
62
32
34
28
BASEBAND PROCESSING CIRCUITRY
40
RF MODULATION CIRCUITRY
44
CONTROL CIRCUITRY
42
PCC
ACS
DCC
ECS
CC-DCI
58

FIG. 7

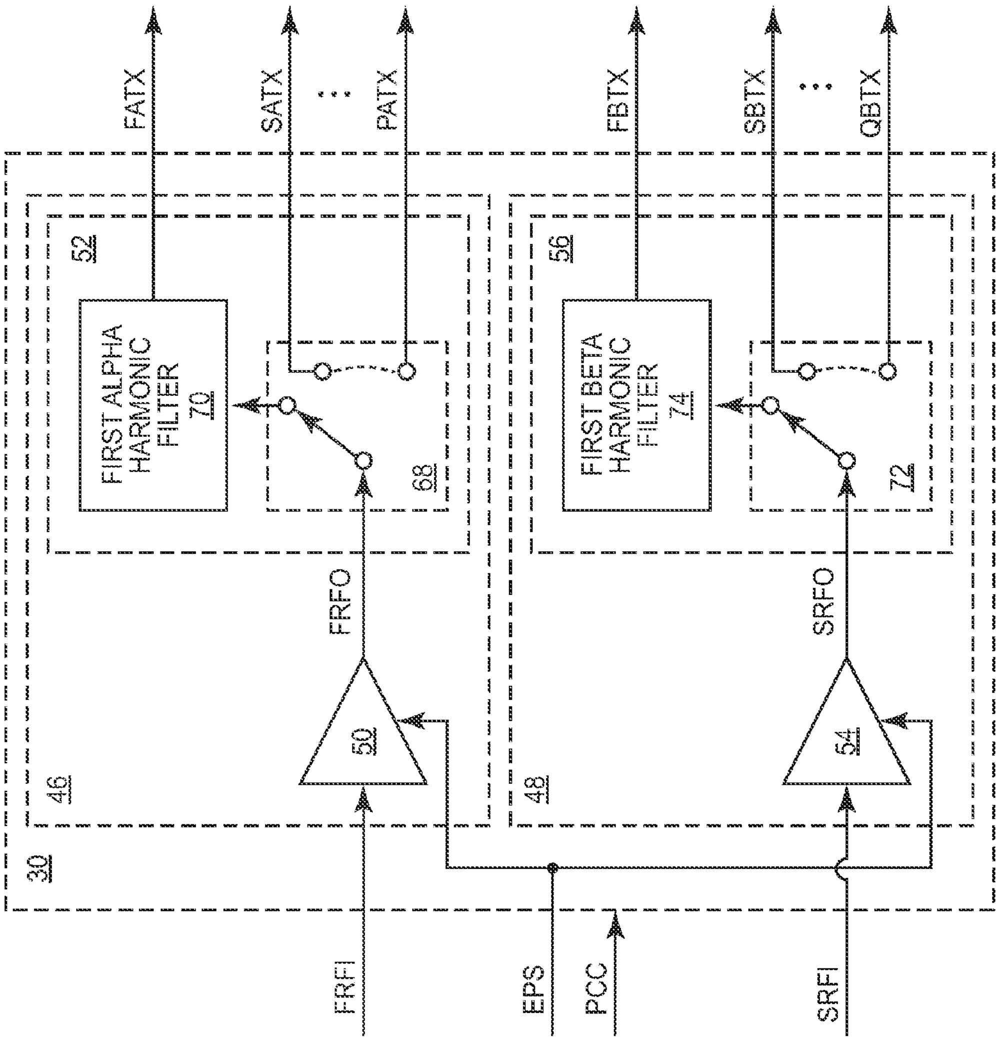
30
46
48
52
56
50
54
68
72
FIRST ALPHA HARMONIC FILTER 70
FIRST BETA HARMONIC FILTER 74
FRFI
EPS
PCC
SRFI
FRFO
SRFO
FATX
SATX
PATX
FBTX
SBTX
QBTX

FIG. 8

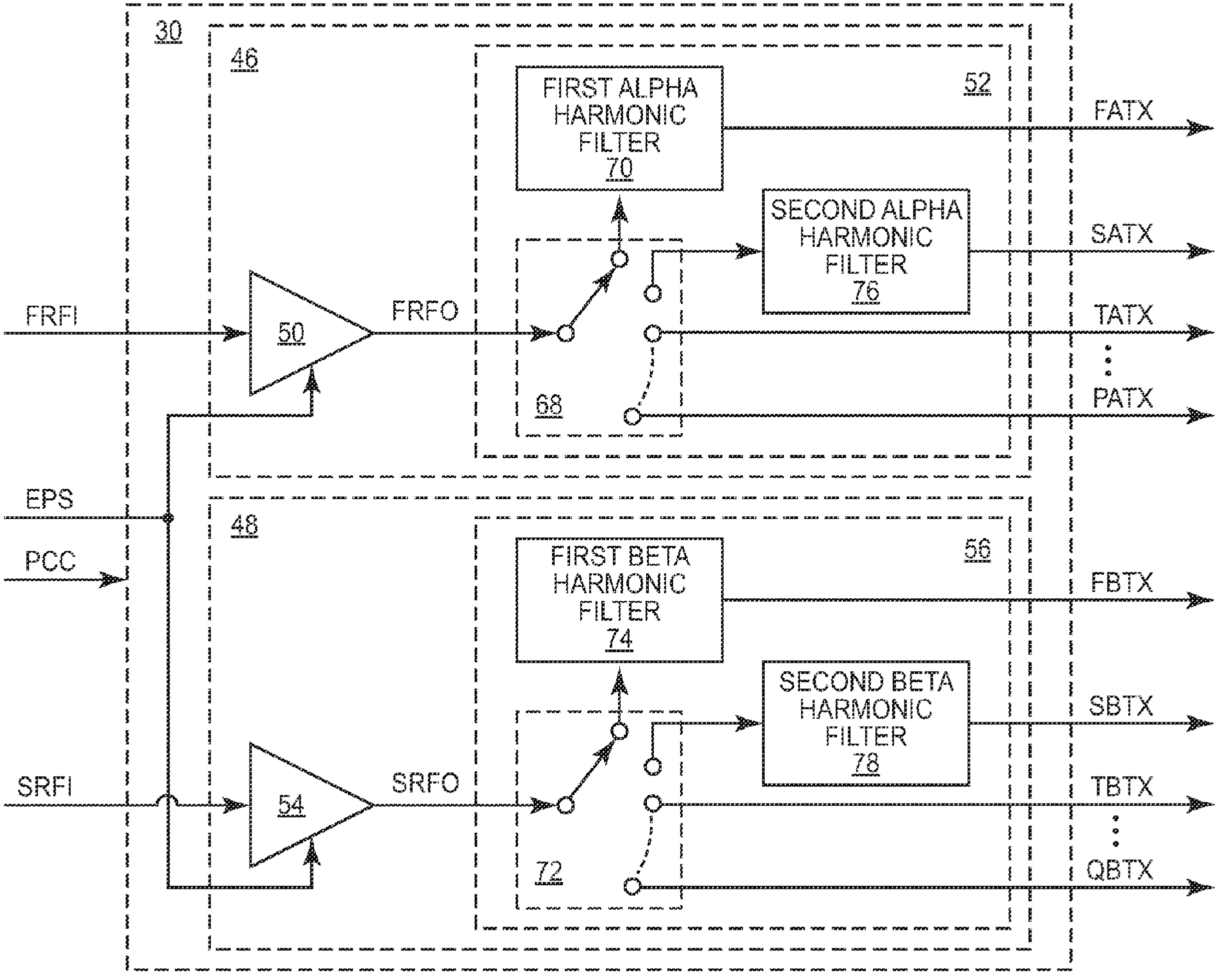
30
46
52
FIRST ALPHA HARMONIC FILTER 70
SECOND ALPHA HARMONIC FILTER 76
FATX
SATX
TATX
⋮
PATX
FRFI
50
FRFO
68
EPS
PCC
48
56
FIRST BETA HARMONIC FILTER 74
SECOND BETA HARMONIC FILTER 78
FBTX
SBTX
TBTX
⋮
QBTX
SRFI
54
SRFO
72

FIG. 9

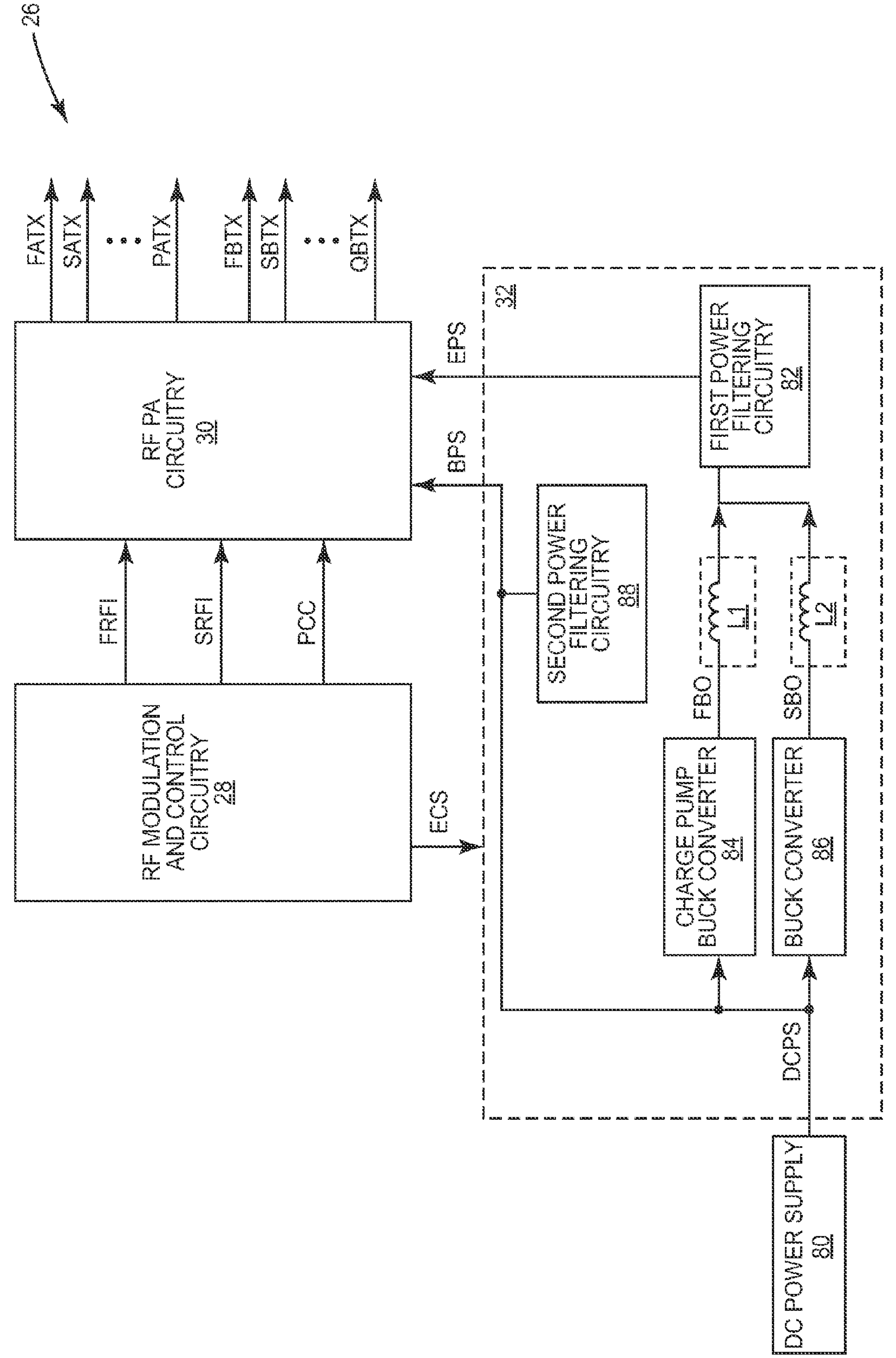
26
FATX
SATX
PATX
FBTX
SBTX
QBTX
RF PA CIRCUITRY 30
FRFI
SRFI
PCC
RF MODULATION AND CONTROL CIRCUITRY 28
ECS
EPS
BPS
32
FIRST POWER FILTERING CIRCUITRY 82
SECOND POWER FILTERING CIRCUITRY 88
L1
L2
FBO
SBO
CHARGE PUMP BUCK CONVERTER 84
BUCK CONVERTER 86
DCPS
DC POWER SUPPLY 80

FIG. 10

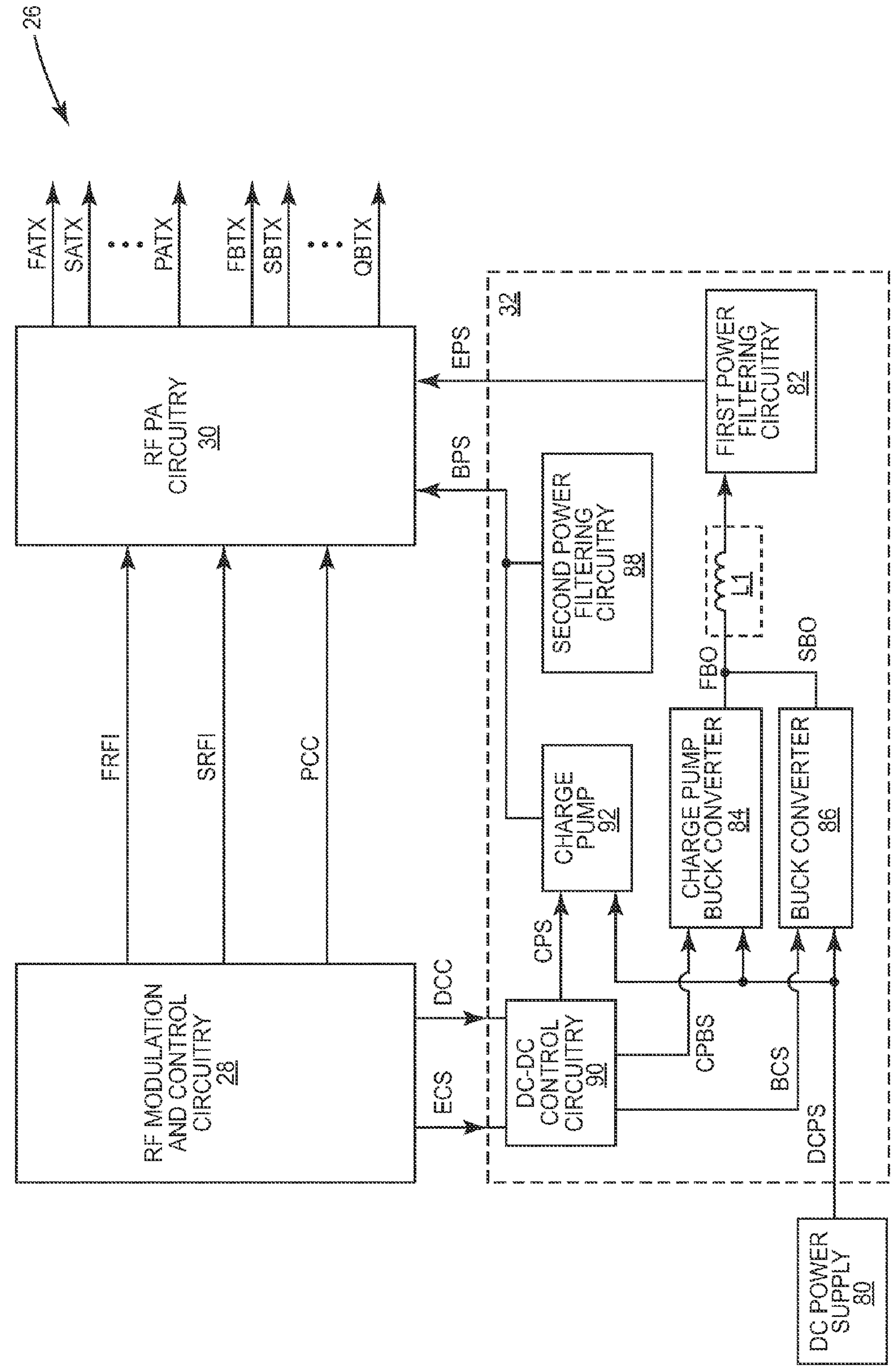
26
FATX
SATX
PATX
FBTX
SBTX
QBTX
RF PA CIRCUITRY 30
EPS
BPS
32
FIRST POWER FILTERING CIRCUITRY 82
SECOND POWER FILTERING CIRCUITRY 88
L1
FBO
SBO
FRFI
SRFI
PCC
CHARGE PUMP 92
CHARGE PUMP BUCK CONVERTER 84
BUCK CONVERTER 86
CPS
DCC
ECS
RF MODULATION AND CONTROL CIRCUITRY 28
DC-DC CONTROL CIRCUITRY 90
CPBS
BCS
DCPS
DC POWER SUPPLY 80

FIG. 11

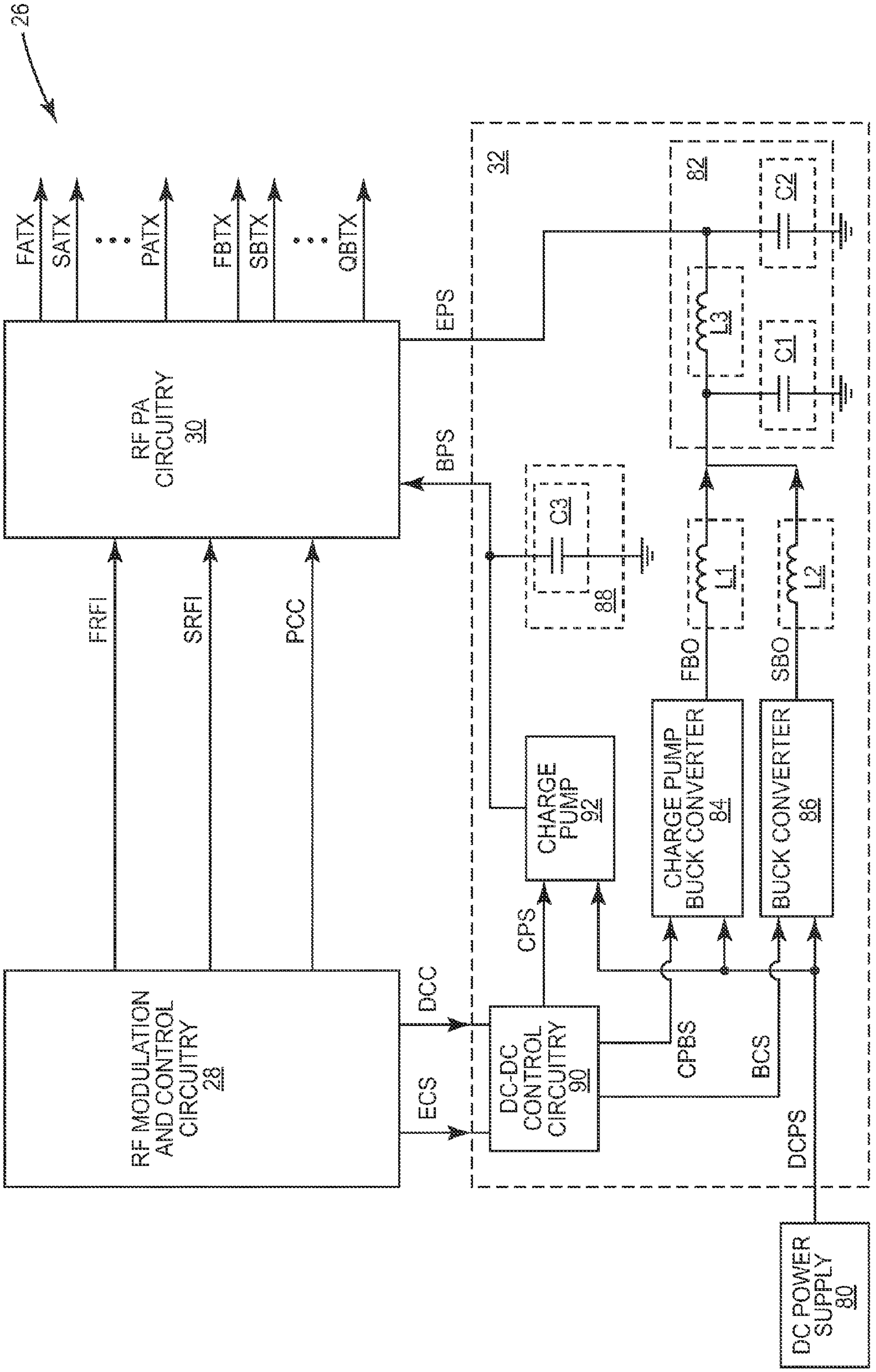
26
RF MODULATION AND CONTROL CIRCUITRY 28
FRFI
SRFI
PCC
RF PA CIRCUITRY 30
FATX
SATX
PATX
FBTX
SBTX
QBTX
EPS
BPS
DCC
ECS
32
DC-DC CONTROL CIRCUITRY 90
CPS
CPBS
BCS
CHARGE PUMP 92
CHARGE PUMP BUCK CONVERTER 84
BUCK CONVERTER 86
FBO
SBO
L1
L2
88
C3
82
L3
C1
C2
DCPS
DC POWER SUPPLY 80

FIG. 12

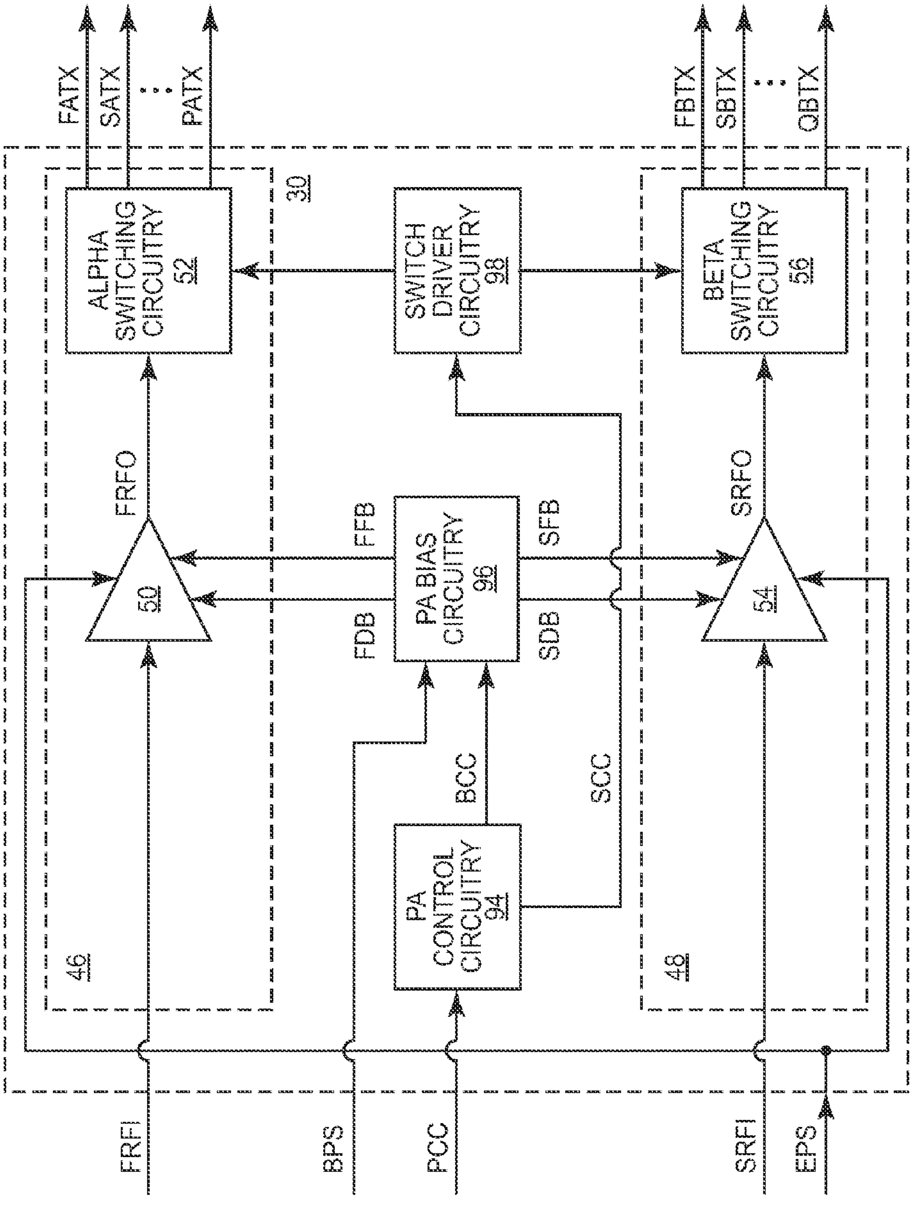
FATX
SATX
⋮
PATX
FBTX
SBTX
⋮
QBTX
30
ALPHA SWITCHING CIRCUITRY 52
SWITCH DRIVER CIRCUITRY 98
BETA SWITCHING CIRCUITRY 56
FRFO
SRFO
FFB
SFB
FDB
SDB
PA BIAS CIRCUITRY 96
50
54
BCC
SCC
PA CONTROL CIRCUITRY 94
46
48
FRFI
BPS
PCC
SRFI
EPS

FIG. 13

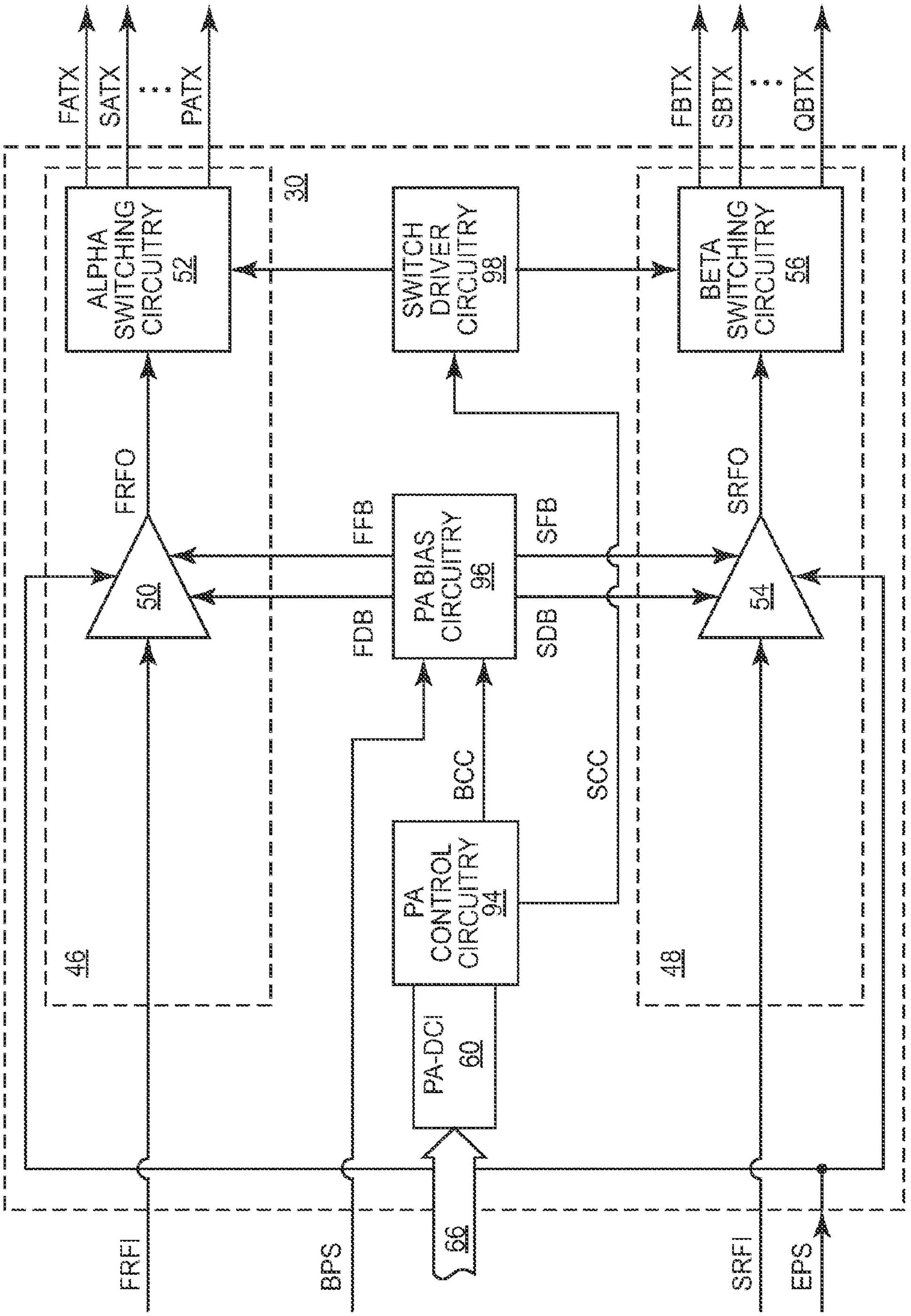
FATX
SATX
PATX
FBTX
SBTX
QBTX
30
ALPHA SWITCHING CIRCUITRY 52
SWITCH DRIVER CIRCUITRY 98
BETA SWITCHING CIRCUITRY 56
FRFO
SRFO
FFB
SFB
PA BIAS CIRCUITRY 96
FDB
SDB
50
54
BCC
SCC
PA CONTROL CIRCUITRY 94
46
48
PA-DCI 60
66
FRFI
BPS
SRFI
EPS

FIG. 14

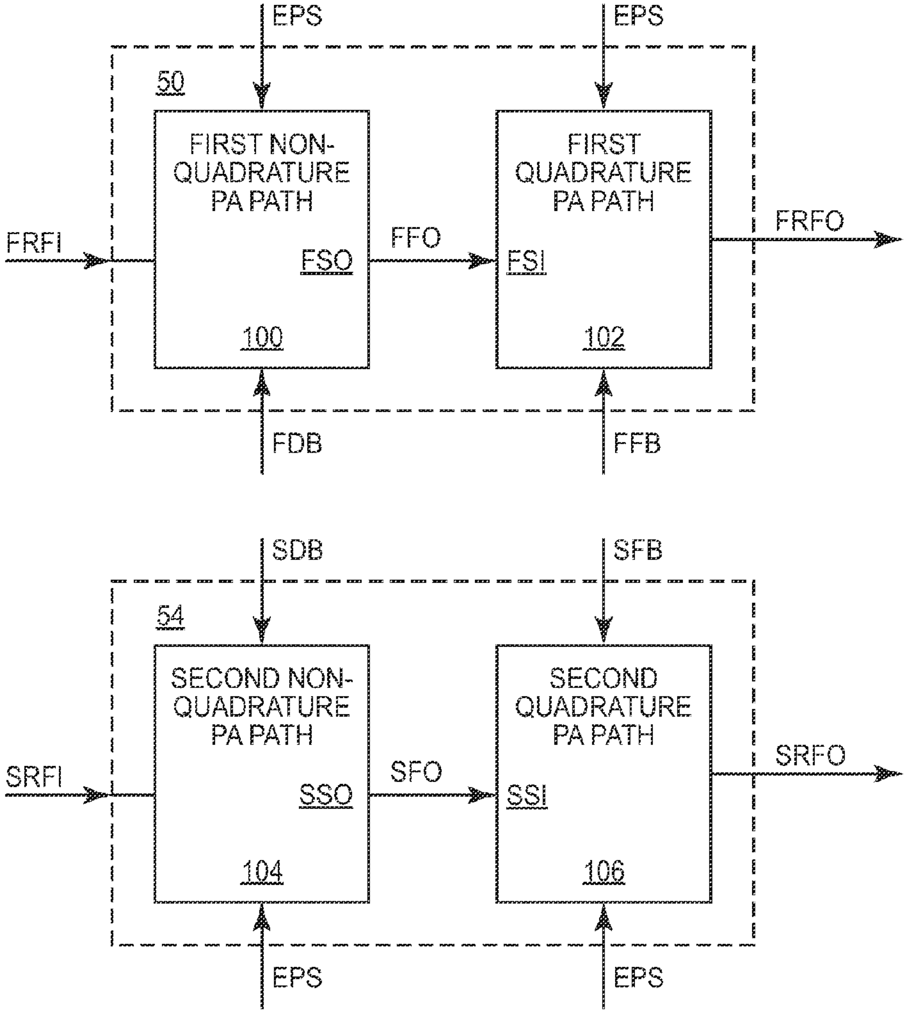
EPS
EPS
50
FIRST NON-QUADRATURE PA PATH
FIRST QUADRATURE PA PATH
FRFI
FSO
FFO
FSI
FRFO
100
102
FDB
FFB
SDB
SFB
54
SECOND NON-QUADRATURE PA PATH
SECOND QUADRATURE PA PATH
SRFI
SSO
SFO
SSI
SRFO
104
106
EPS
EPS

FIG. 15

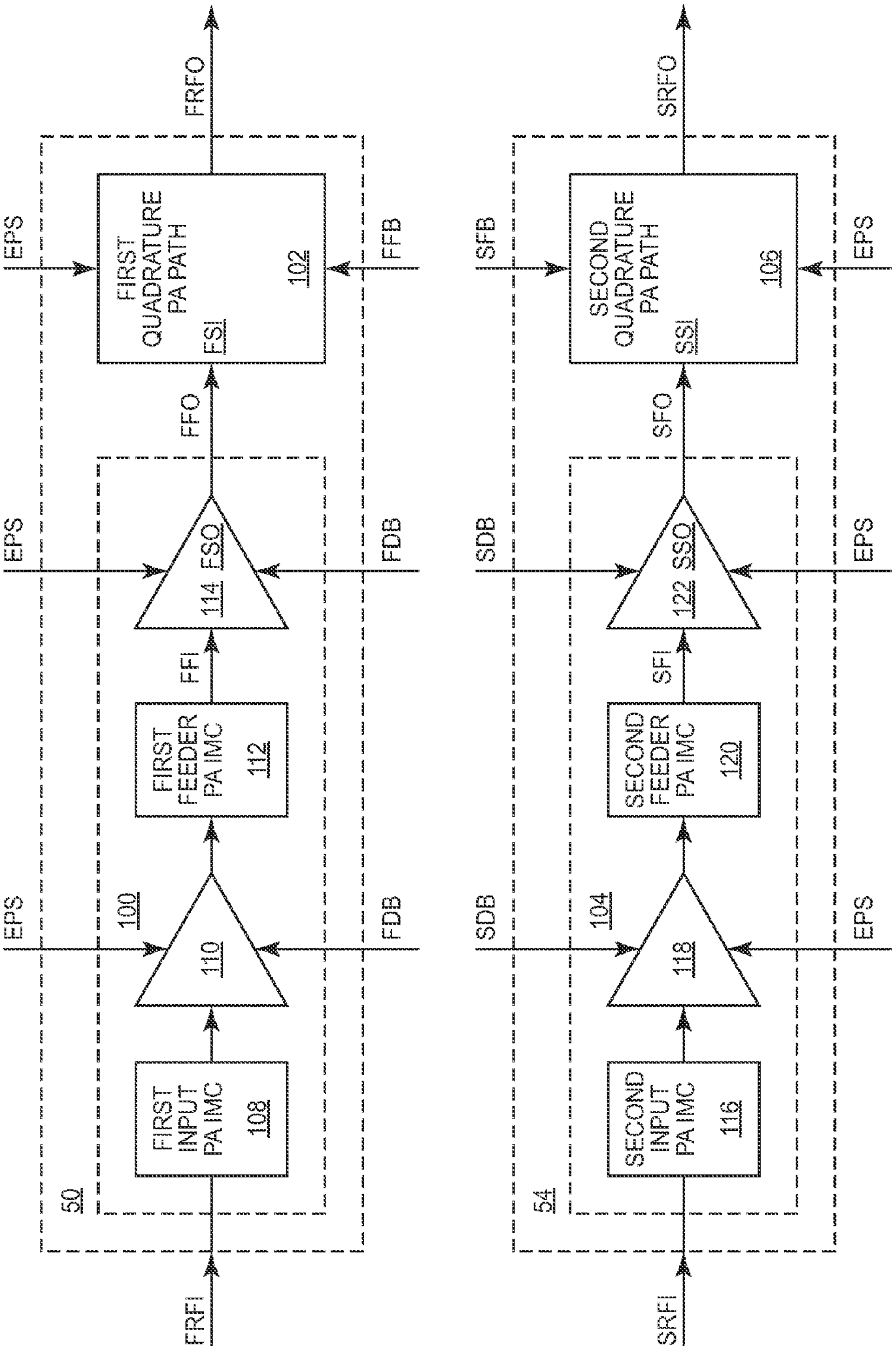
FRFI
50
FIRST INPUT PA IMC 108
EPS
100
110
FDB
FIRST FEEDER PA IMC 112
FFI
EPS
114 FSO
FDB
FFO
EPS
FIRST QUADRATURE PA PATH
FSI
102
FFB
FRFO
SRFI
54
SECOND INPUT PA IMC 116
SDB
104
118
EPS
SECOND FEEDER PA IMC 120
SFI
SDB
122 SSO
EPS
SFO
SFB
SECOND QUADRATURE PA PATH
SSI
106
EPS
SRFO

FIG. 16

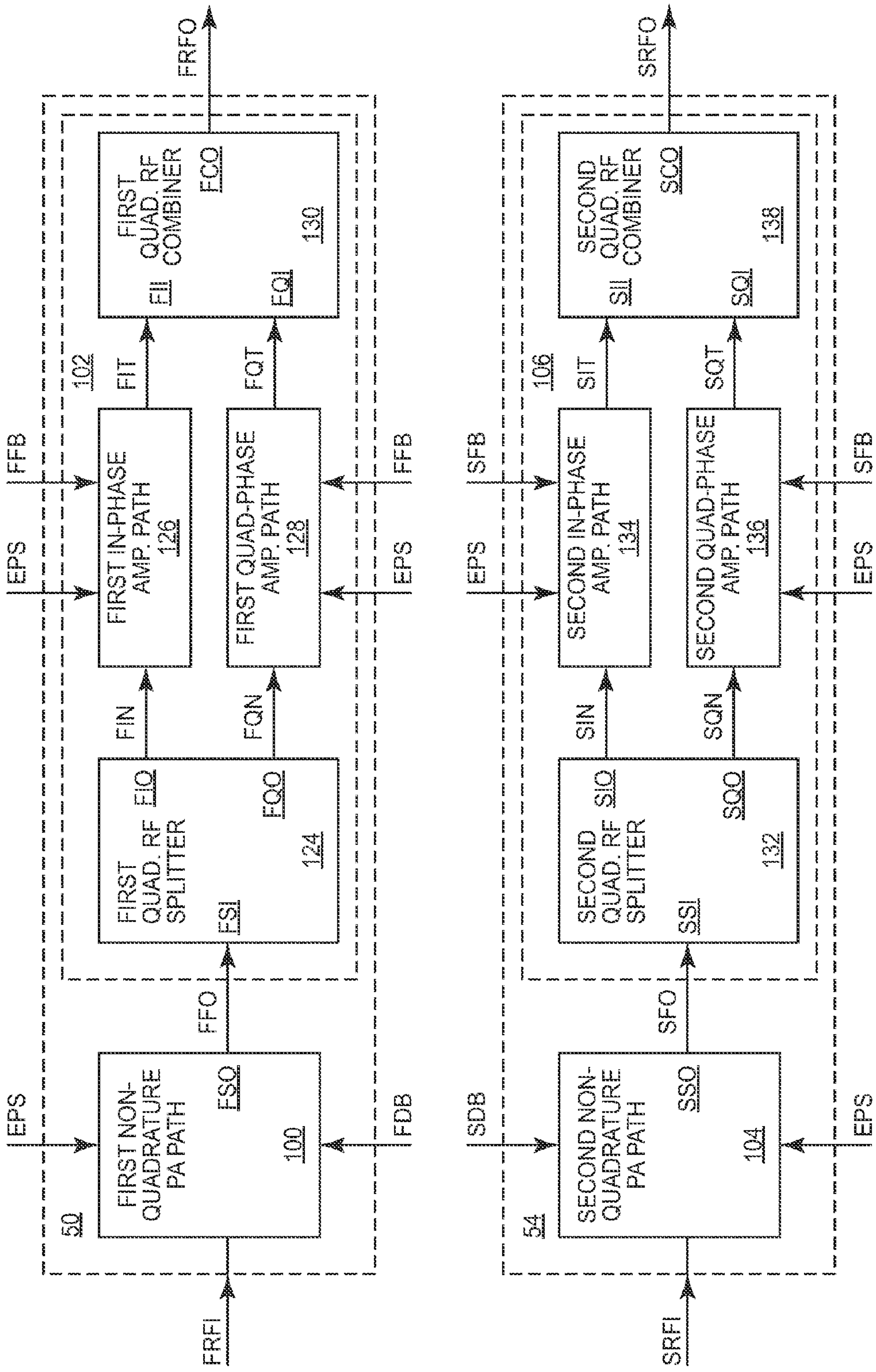
FRFI
EPS
FDB
FIRST NON-QUADRATURE PA PATH
FSO
100
50
FFO
FIRST QUAD. RF SPLITTER
FSI
FIO
FQO
124
FIN
FQN
FIRST IN-PHASE AMP. PATH
126
FIRST QUAD-PHASE AMP. PATH
128
EPS
FFB
EPS
FFB
102
FIT
FQT
FIRST QUAD. RF COMBINER
FII
FQI
FCO
130
FRFO
SRFI
SDB
EPS
SECOND NON-QUADRATURE PA PATH
SSO
104
54
SFO
SECOND QUAD. RF SPLITTER
SSI
SIO
SQO
132
SIN
SQN
SECOND IN-PHASE AMP. PATH
134
SECOND QUAD-PHASE AMP. PATH
136
EPS
SFB
EPS
SFB
106
SIT
SQT
SECOND QUAD. RF COMBINER
SII
SQI
SCO
138
SRFO

FIG. 17

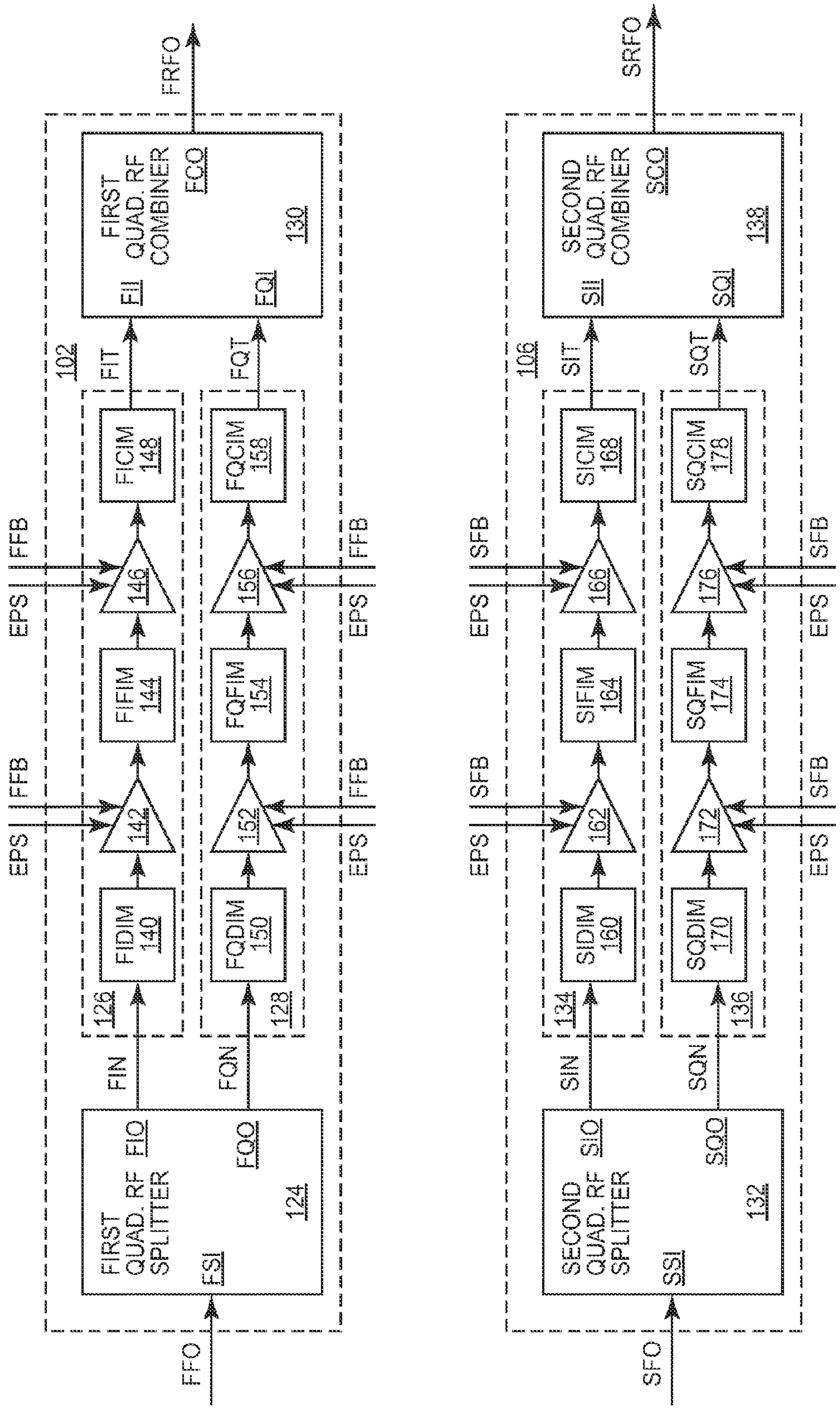
FFO
FIRST QUAD. RF SPLITTER
FSI
FIO
FQO
124
FIN
FQN
126
128
FIDIM 140
FQDIM 150
EPS
FFB
142
152
FIFIM 144
FQFIM 154
146
156
FICIM 148
FQCIM 158
FIT
FQT
102
FII
FQI
FIRST QUAD. RF COMBINER
FCO
130
FRFO
SFO
SECOND QUAD. RF SPLITTER
SSI
SIO
SQO
132
SIN
SQN
134
136
SIDIM 160
SQDIM 170
SFB
162
172
SIFIM 164
SQFIM 174
166
176
SICIM 168
SQCIM 178
SIT
SQT
106
SII
SQI
SECOND QUAD. RF COMBINER
SCO
138
SRFO

FIG. 18

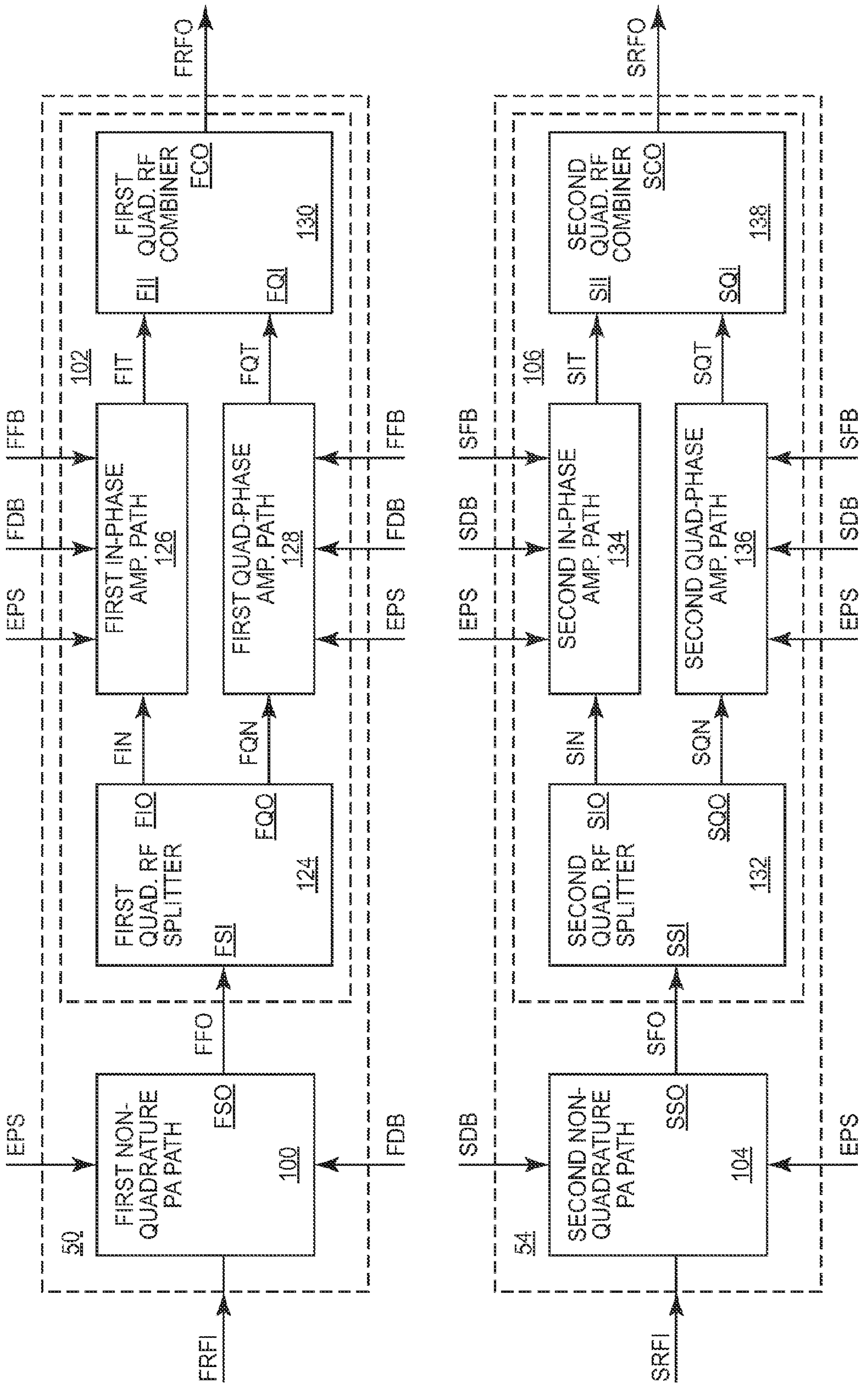
50
FIRST NON-QUADRATURE PA PATH
FSO
100
EPS
FDB
FRFI
FFO
102
FIRST QUAD. RF SPLITTER
FSI
FIO
FQO
124
FIN
FQN
FIRST IN-PHASE AMP. PATH
126
FIRST QUAD-PHASE AMP. PATH
128
EPS
FDB
FFB
EPS
FDB
FFB
FIT
FQT
FIRST QUAD. RF COMBINER
FII
FQI
FCO
130
FRFO
54
SECOND NON-QUADRATURE PA PATH
SSO
104
SDB
EPS
SRFI
SFO
106
SECOND QUAD. RF SPLITTER
SSI
SIO
SQO
132
SIN
SQN
SECOND IN-PHASE AMP. PATH
134
SECOND QUAD-PHASE AMP. PATH
136
EPS
SDB
SFB
EPS
SDB
SFB
SIT
SQT
SECOND QUAD. RF COMBINER
SII
SQI
SCO
138
SRFO

FIG. 19

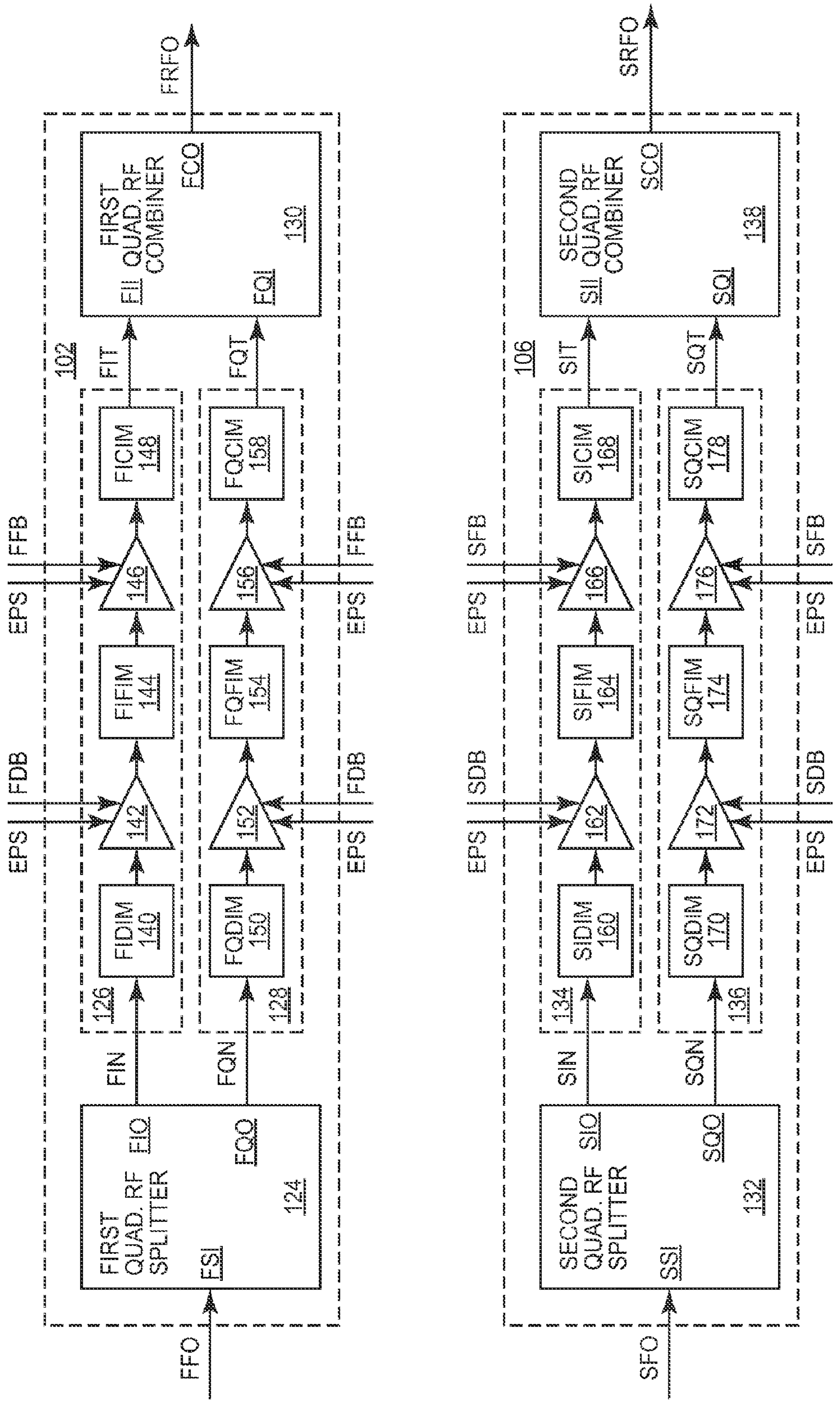
FFO
FIRST QUAD. RF SPLITTER
FSI
FIO
FQO
124
FIN
FQN
126
128
FIDIM 140
FQDIM 150
EPS
FDB
142
152
FIFIM 144
FQFIM 154
EPS
FFB
146
156
FICIM 148
FQCIM 158
FIT
FQT
102
FIRST QUAD. RF COMBINER
FII
FQI
FCO
130
FRFO
SFO
SECOND QUAD. RF SPLITTER
SSI
SIO
SQO
132
SIN
SQN
134
136
SIDIM 160
SQDIM 170
EPS
SDB
162
172
SIFIM 164
SQFIM 174
EPS
SFB
166
176
SICIM 168
SQCIM 178
SIT
SQT
106
SECOND QUAD. RF COMBINER
SII
SQI
SCO
138
SRFO

FIG. 20

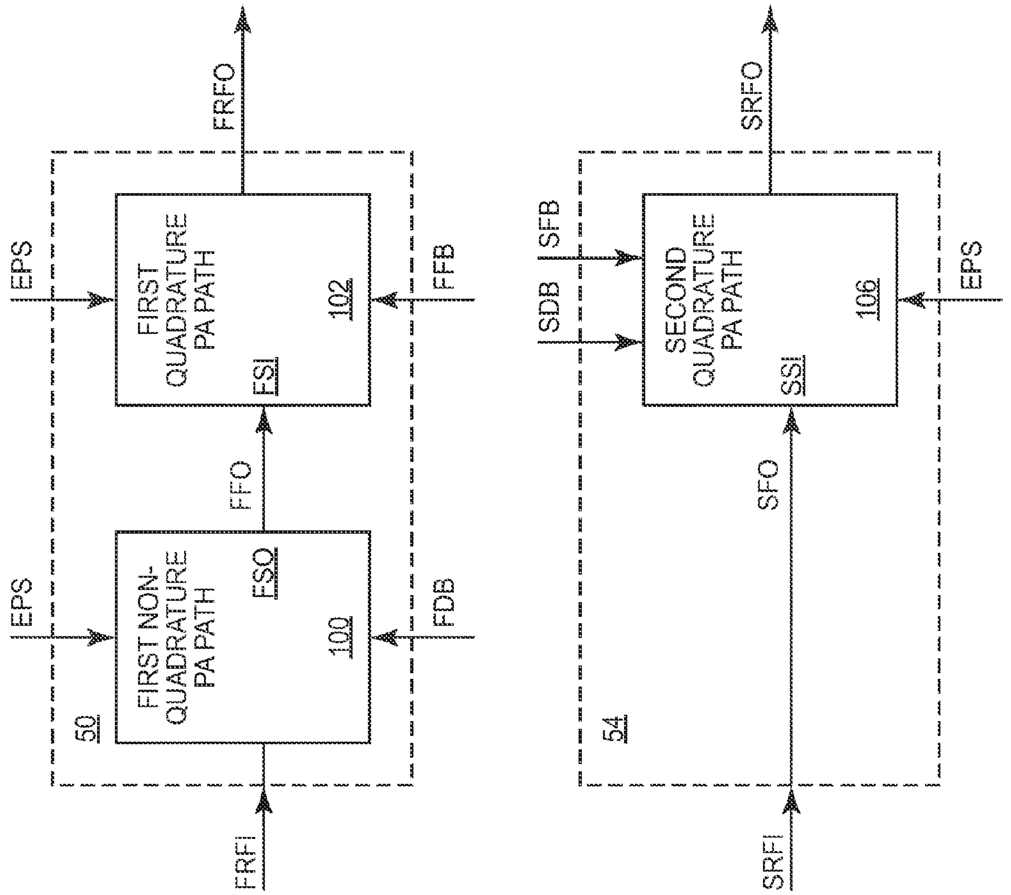
FRFI
50
FIRST NON-QUADRATURE PA PATH
FSO
100
EPS
FDB
FFO
FIRST QUADRATURE PA PATH
FSI
102
EPS
FFB
FRFO
SRFI
54
SFO
SECOND QUADRATURE PA PATH
SSI
106
SDB
SFB
EPS
SRFO

FIG. 21

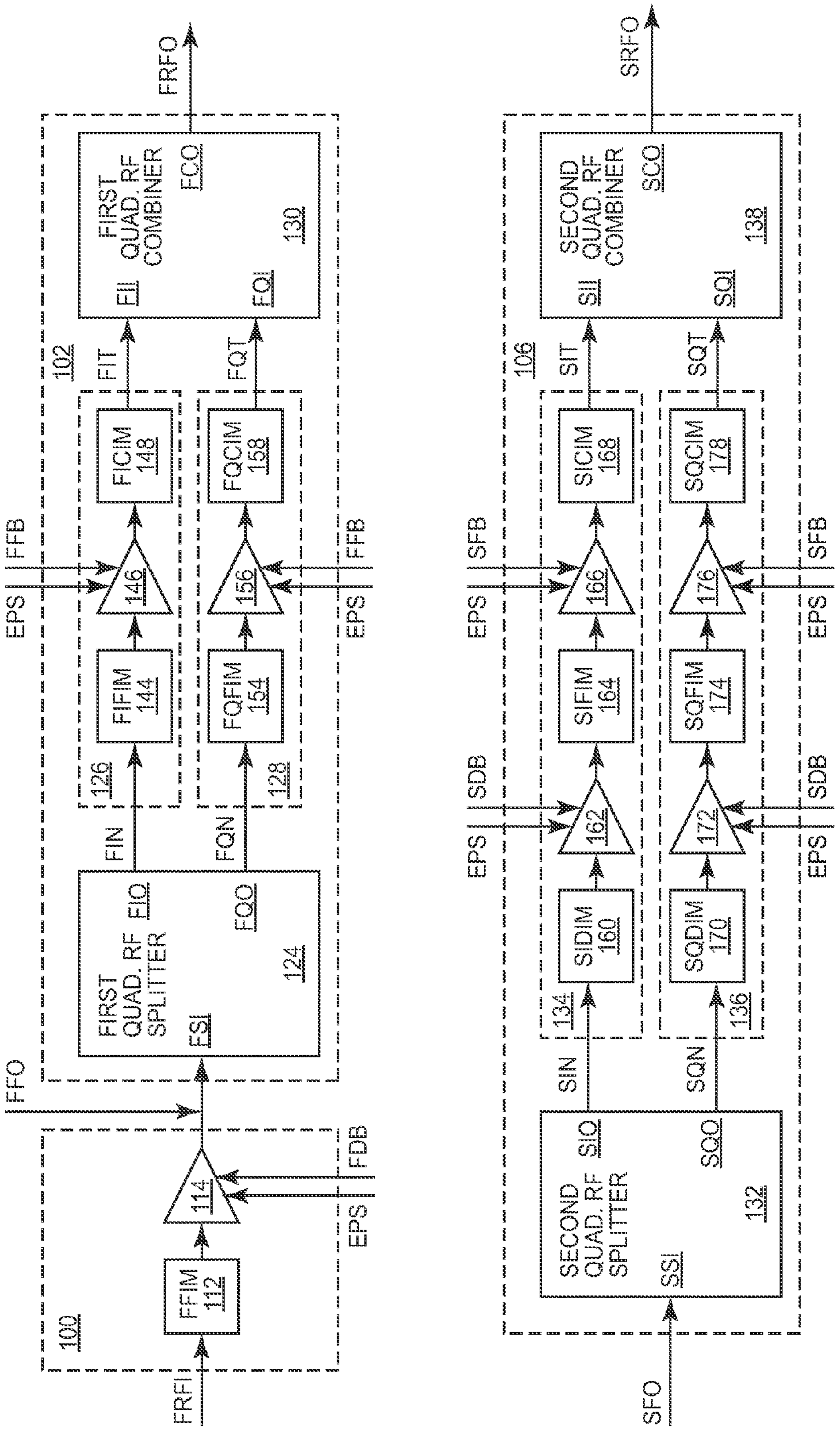
FRFI
100
FFIM
112
114
EPS
FDB
FFO
FIRST
QUAD. RF
SPLITTER
FSI
FIO
FQO
124
FIN
FQN
126
FIFIM
144
146
EPS
FFB
FICIM
148
FIT
128
FQFIM
154
156
FQCIM
158
FQT
102
FIRST
QUAD. RF
COMBINER
FII
FQI
FCO
130
FRFO
SFO
SECOND
QUAD. RF
SPLITTER
SSI
SIO
SQO
132
SIN
SQN
134
SIDIM
160
162
EPS
SDB
SIFIM
164
166
SFB
SICIM
168
SIT
136
SQDIM
170
172
SQFIM
174
176
SQCIM
178
SQT
106
SECOND
QUAD. RF
COMBINER
SII
SQI
SCO
138
SRFO

FIG. 22

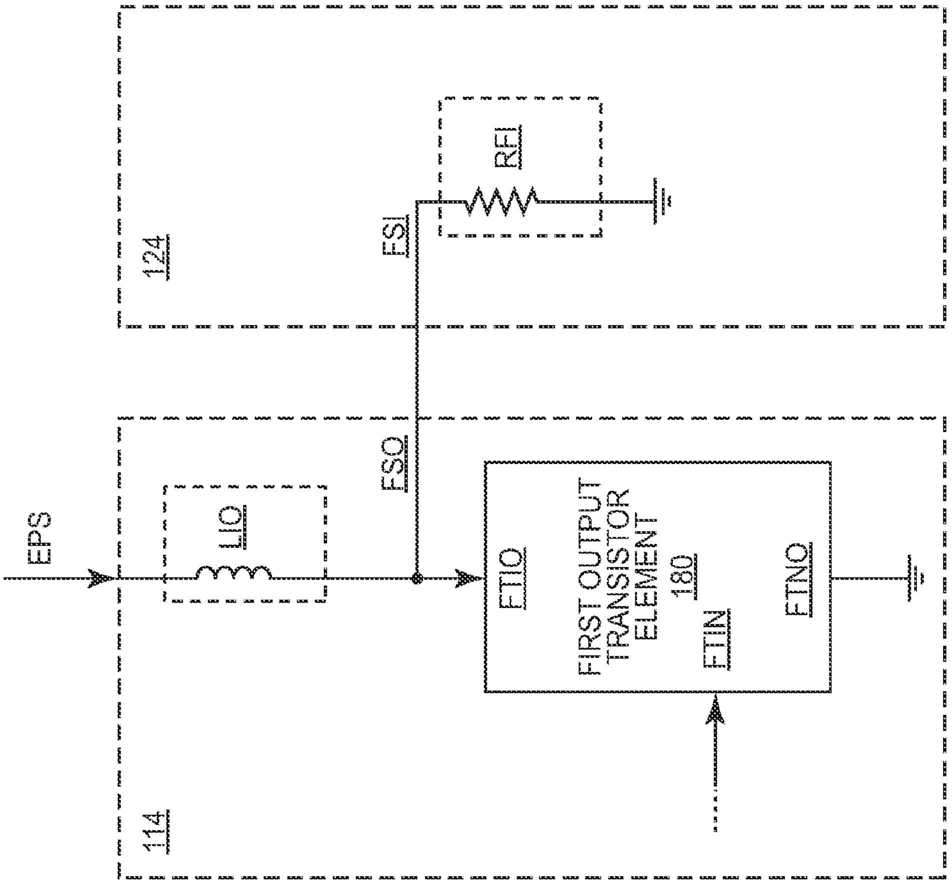
EPS
LIO
FSO
FSI
RFI
124
114
FTIO
FIRST OUTPUT TRANSISTOR ELEMENT
180
FTIN
FTNO

FIG. 23

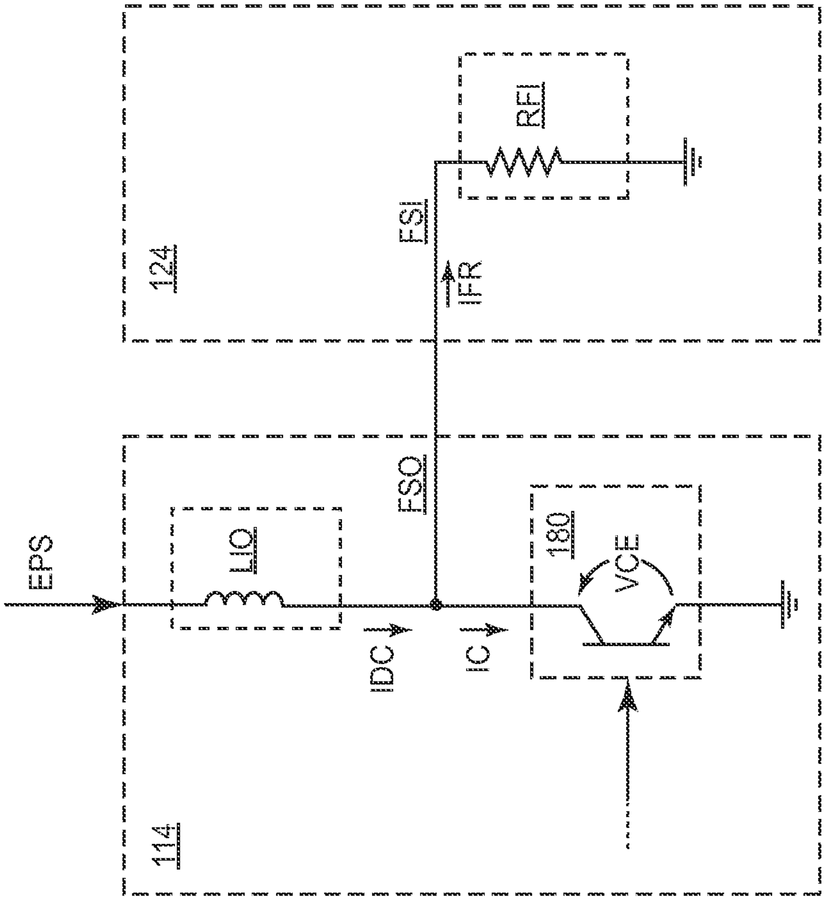
124
FSI
IFR
RFI
FSO
EPS
LIO
IDC
IC
180
VCE
114

FIG. 24

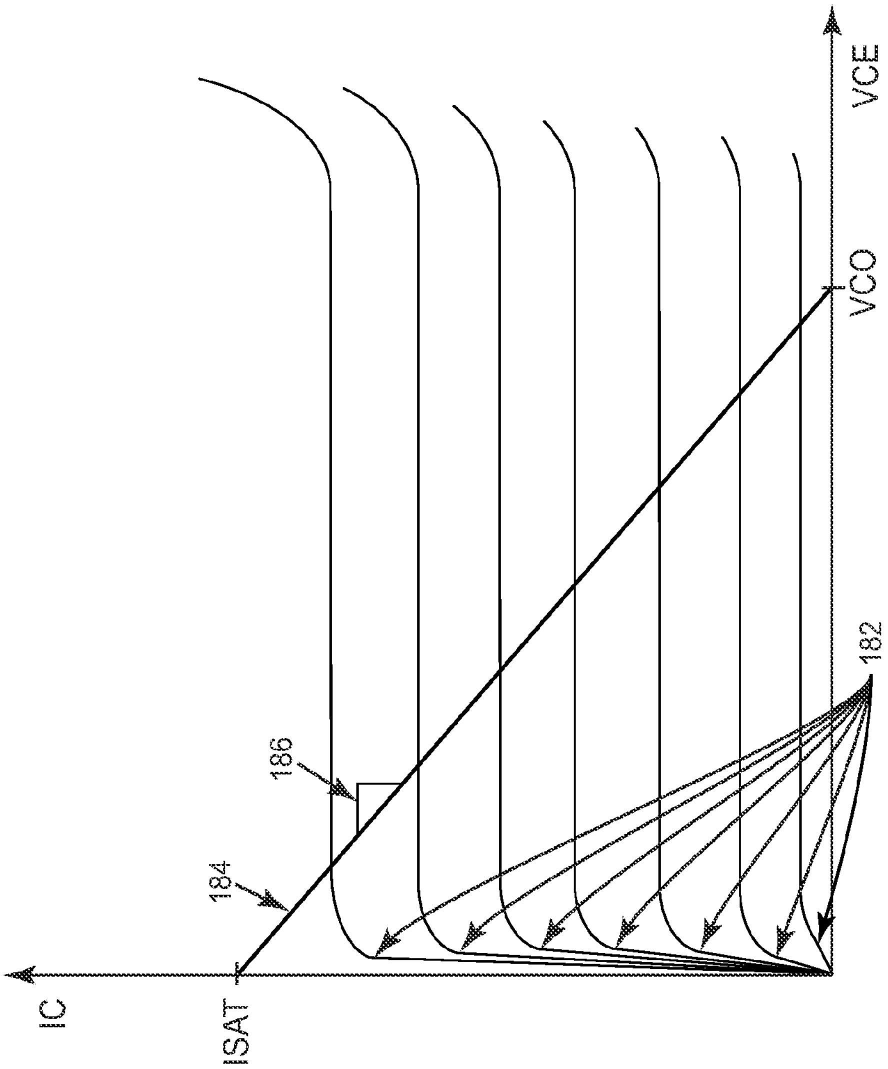

FIG. 25

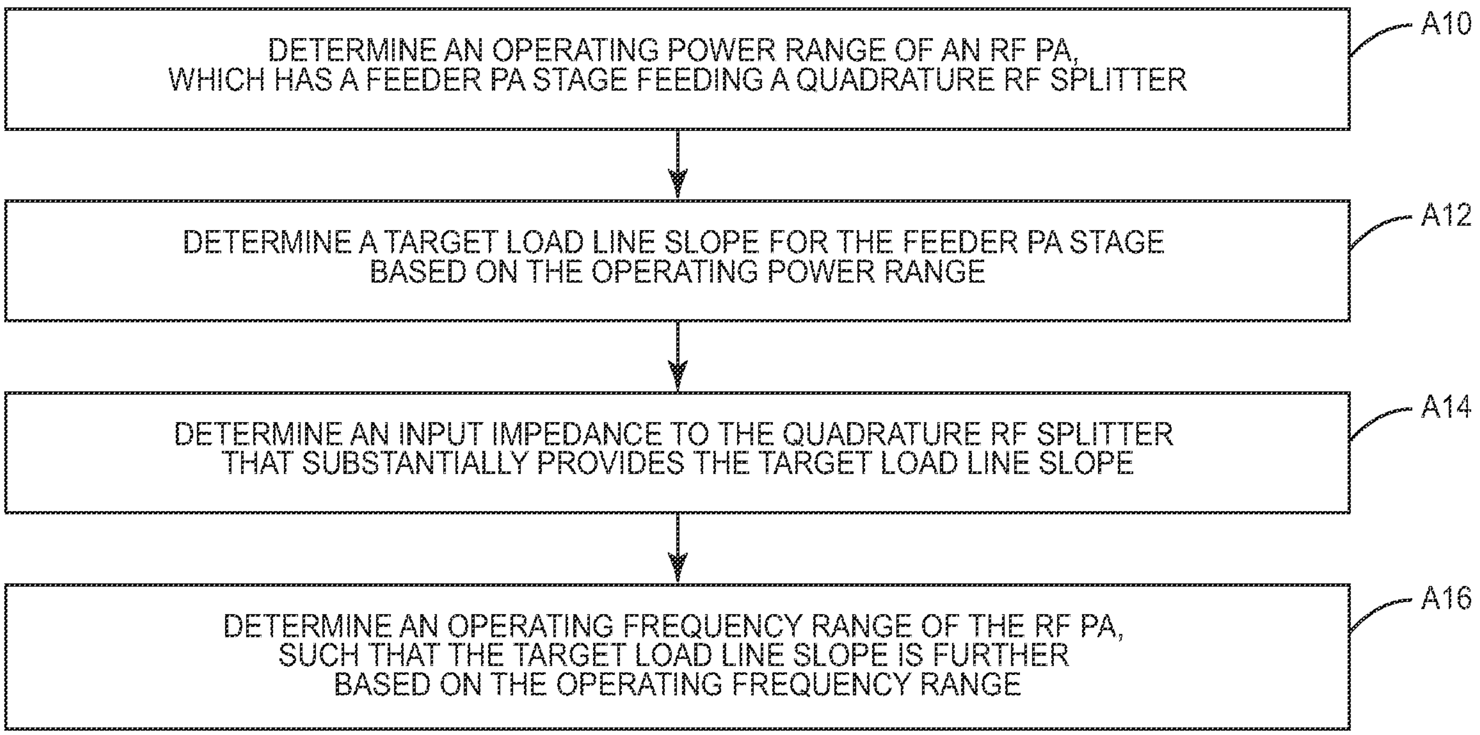
DETERMINE AN OPERATING POWER RANGE OF AN RF PA, WHICH HAS A FEEDER PA STAGE FEEDING A QUADRATURE RF SPLITTER
A10
DETERMINE A TARGET LOAD LINE SLOPE FOR THE FEEDER PA STAGE BASED ON THE OPERATING POWER RANGE
A12
DETERMINE AN INPUT IMPEDANCE TO THE QUADRATURE RF SPLITTER THAT SUBSTANTIALLY PROVIDES THE TARGET LOAD LINE SLOPE
A14
DETERMINE AN OPERATING FREQUENCY RANGE OF THE RF PA, SUCH THAT THE TARGET LOAD LINE SLOPE IS FURTHER BASED ON THE OPERATING FREQUENCY RANGE
A16

FIG. 26

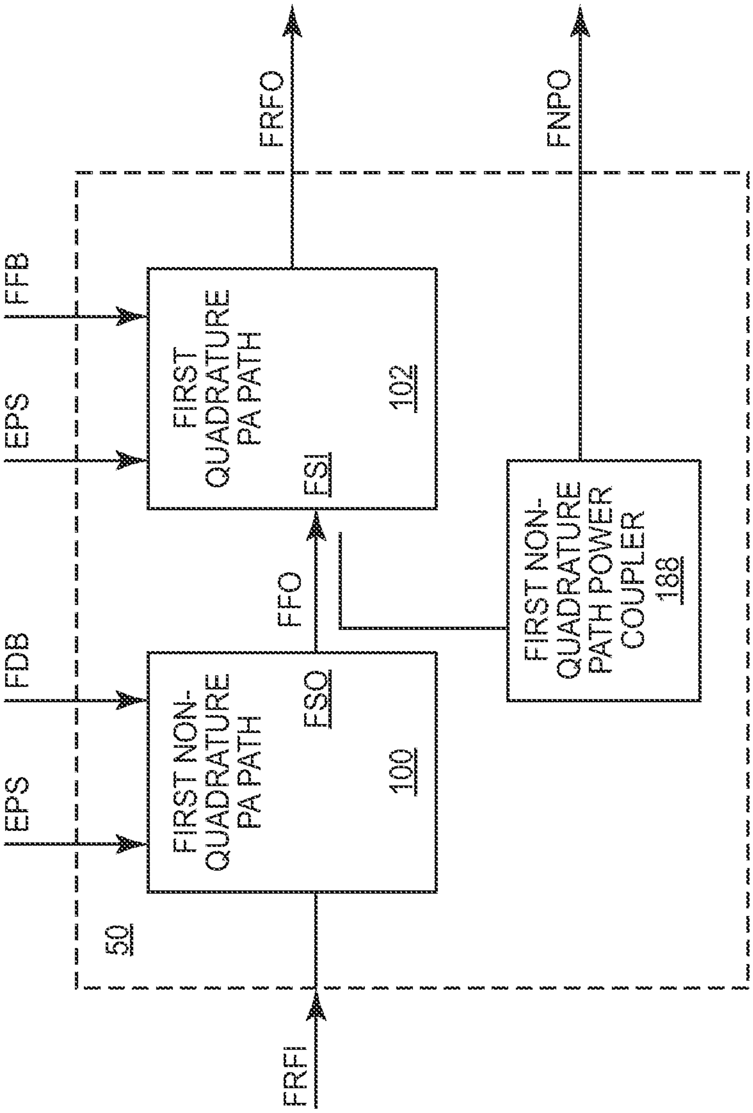
FRFI
50
EPS
FDB
FIRST NON-QUADRATURE PA PATH
100
ESO
FFO
EPS
FFB
FIRST QUADRATURE PA PATH
102
FSI
FIRST NON-QUADRATURE PATH POWER COUPLER
188
FRFO
FNPO

FIG. 27

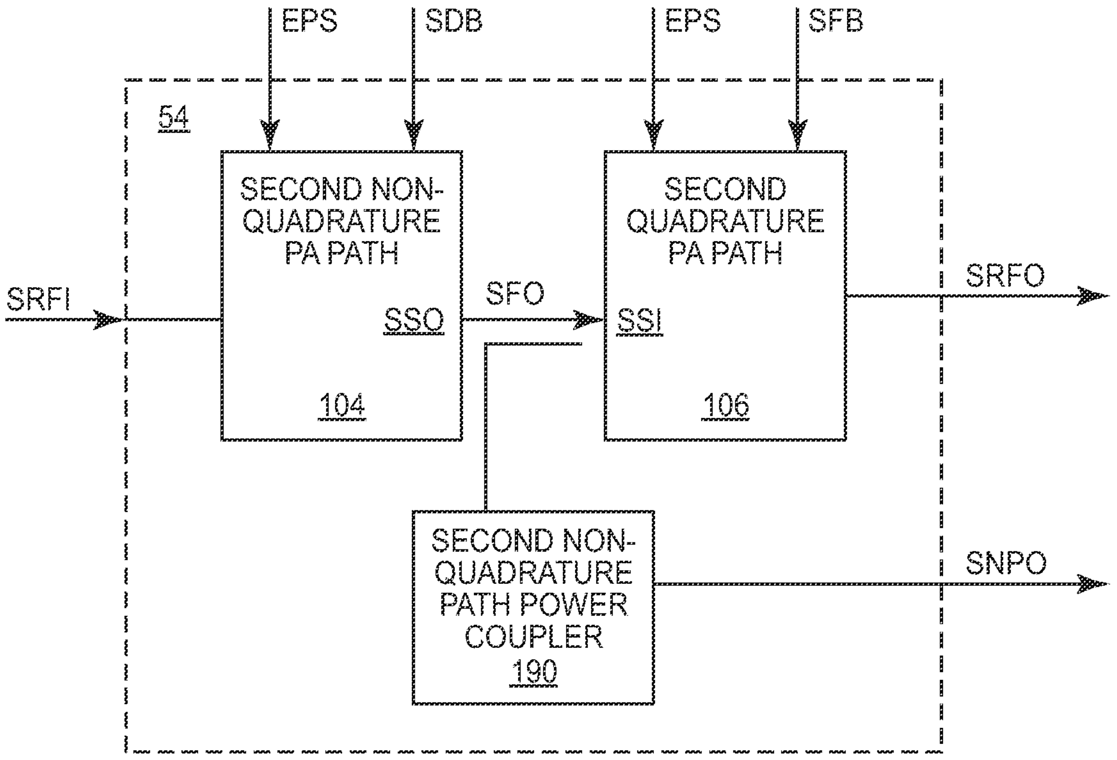
EPS
SDB
EPS
SFB
54
SECOND NON-QUADRATURE PA PATH
SSO
104
SFO
SECOND QUADRATURE PA PATH
SSI
106
SRFI
SRFO
SECOND NON-QUADRATURE PATH POWER COUPLER
190
SNPO

FIG. 28

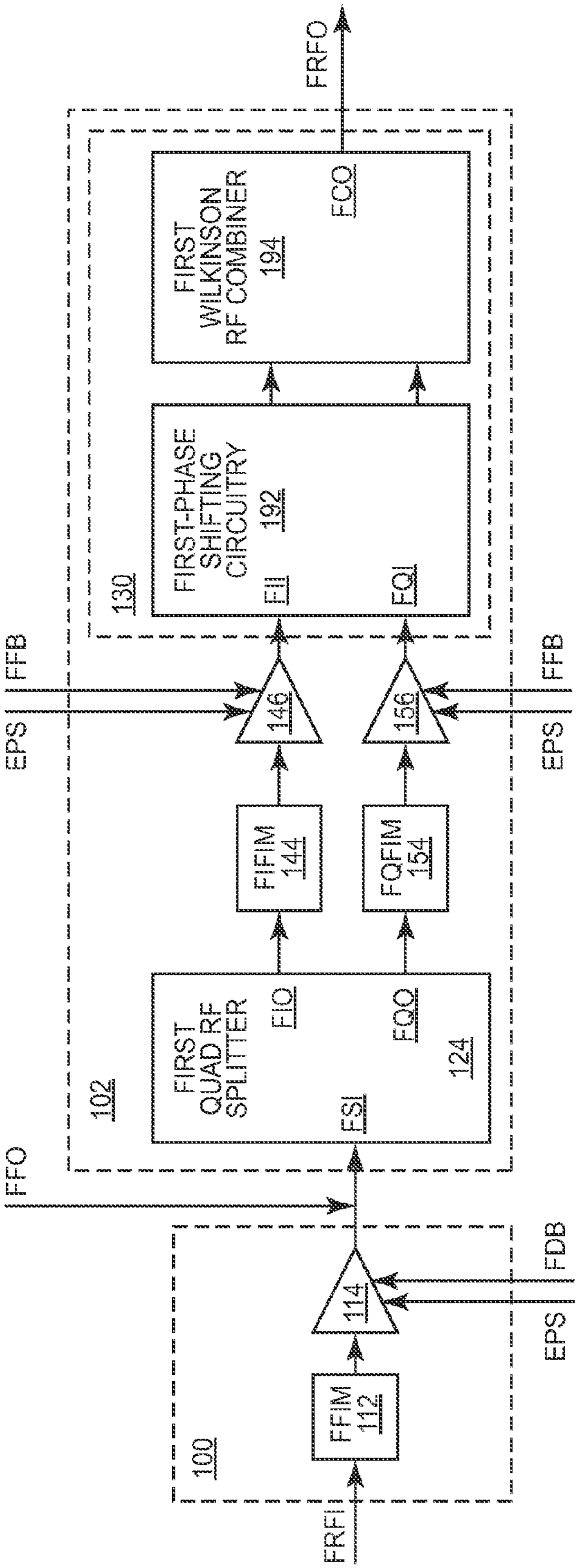
FRFI
100
FFIM
112
114
EPS
FDB
FFO
102
FIRST
QUAD RF
SPLITTER
FSI
FIO
FQO
124
FIFIM
144
FQFIM
154
EPS
FFB
146
156
EPS
FFB
130
FIRST-PHASE
SHIFTING
CIRCUITRY
192
FII
FQI
FIRST
WILKINSON
RF COMBINER
194
FCO
FRFO

FIG. 29

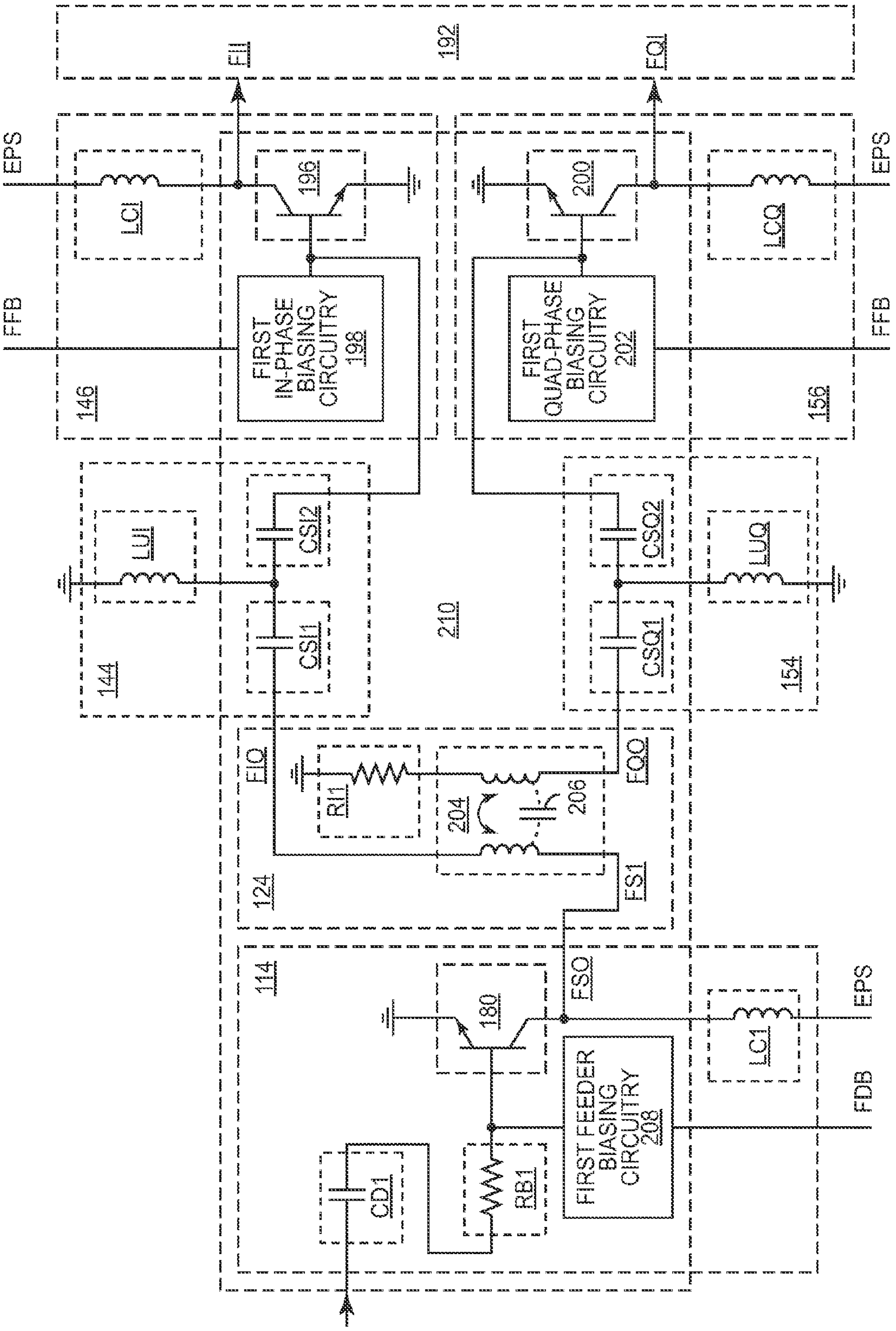
192
FII
FQI
EPS
LCI
196
200
LCQ
EPS
FFB
FIRST IN-PHASE BIASING CIRCUITRY 198
FIRST QUAD-PHASE BIASING CIRCUITRY 202
FFB
146
156
LUI
CSI2
CSQ2
LUQ
144
CSI1
210
CSQ1
154
FIO
RI1
204
206
FQO
124
FS1
114
FSO
180
EPS
LC1
FIRST FEEDER BIASING CIRCUITRY 208
FDB
CD1
RB1

FIG. 30

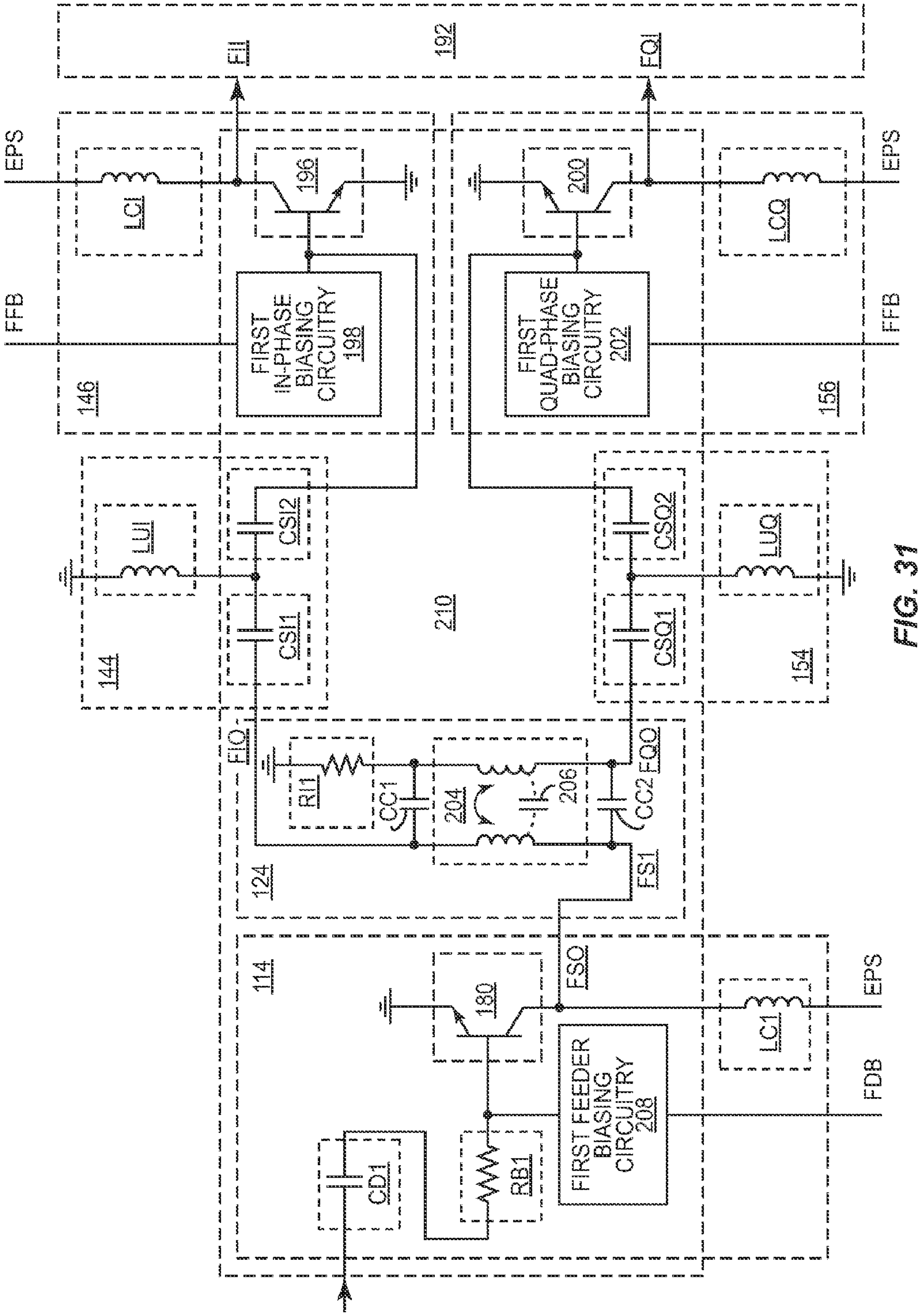
192
FII
FQI
EPS
LCI
196
FFB
FIRST IN-PHASE BIASING CIRCUITRY 198
146
200
LCQ
EPS
FIRST QUAD-PHASE BIASING CIRCUITRY 202
FFB
156
LUI
CSI2
CSI1
144
210
CSQ2
CSQ1
LUQ
154
FIO
RI1
CC1
204
206
CC2
FQO
FS1
124
114
180
FSO
EPS
LC1
FIRST FEEDER BIASING CIRCUITRY 208
FDB
RB1
CD1

FIG. 31

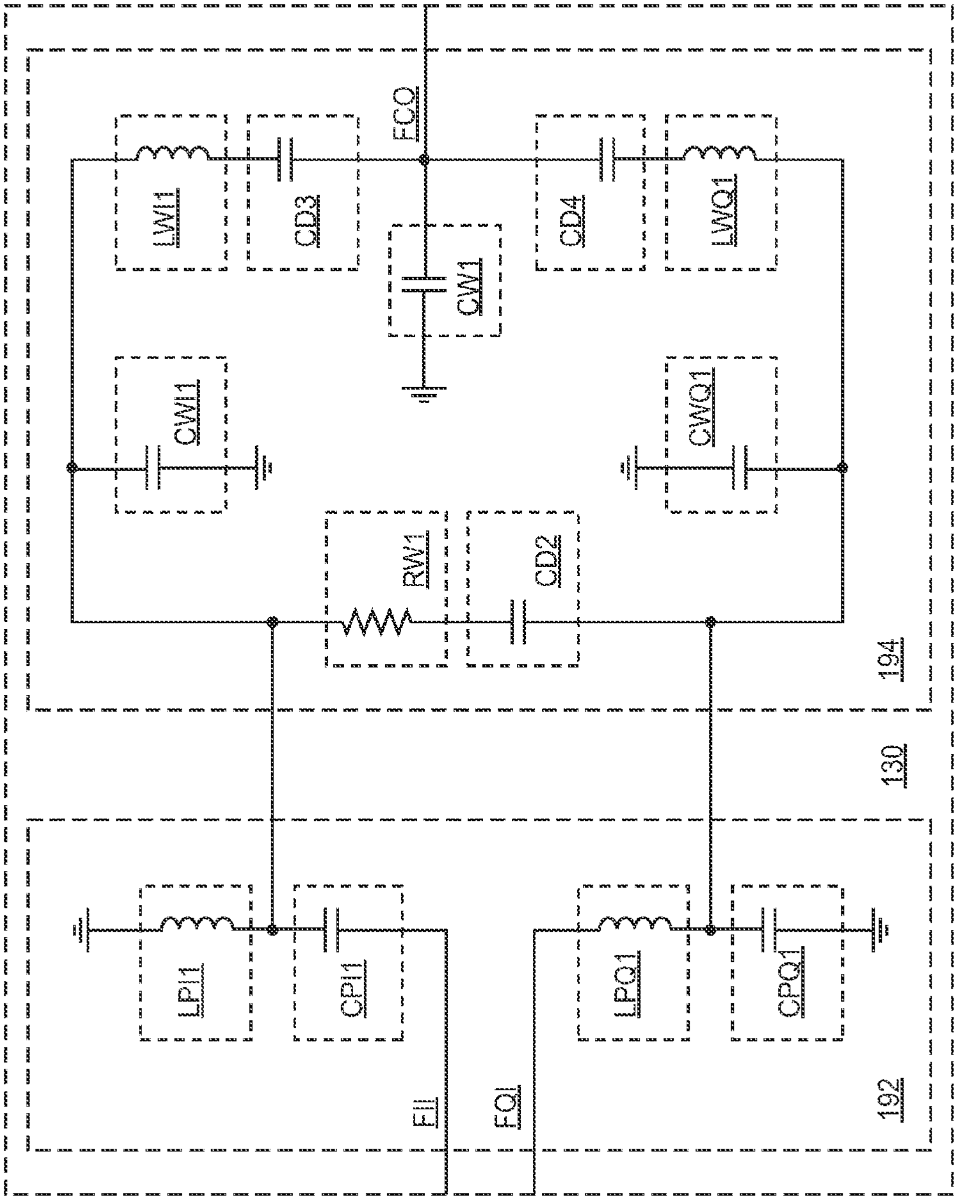
FCO
LWI1
CD3
CD4
LWQ1
CW1
CWI1
CWQ1
RW1
CD2
194
130
LPI1
CPI1
LPQ1
CPQ1
FII
FQI
192

FIG. 32

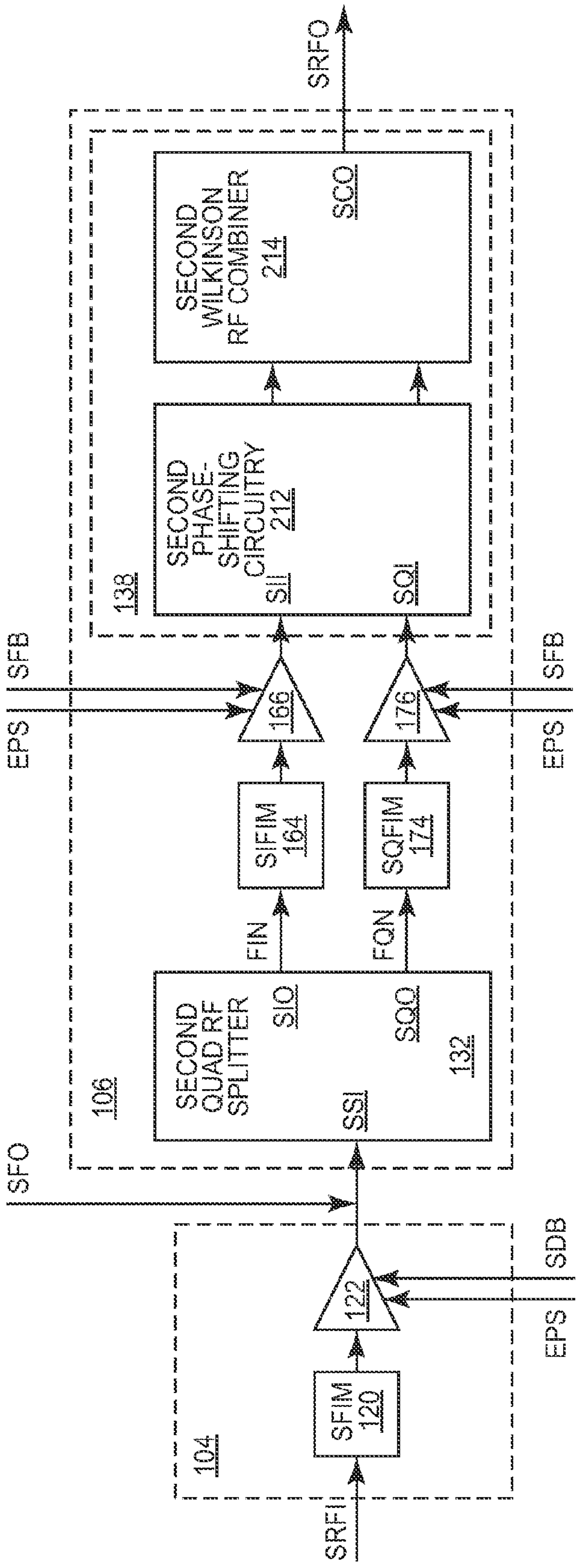
SRFI
104
SFIM
120
122
EPS
SDB
SFO
106
SECOND QUAD RF SPLITTER
SSI
SIO
SQO
132
FIN
FQN
SIFIM
164
SQFIM
174
166
176
EPS
SFB
EPS
SFB
138
SECOND PHASE-SHIFTING CIRCUITRY
212
SII
SQI
SECOND WILKINSON RF COMBINER
214
SCO
SRFO

FIG. 33

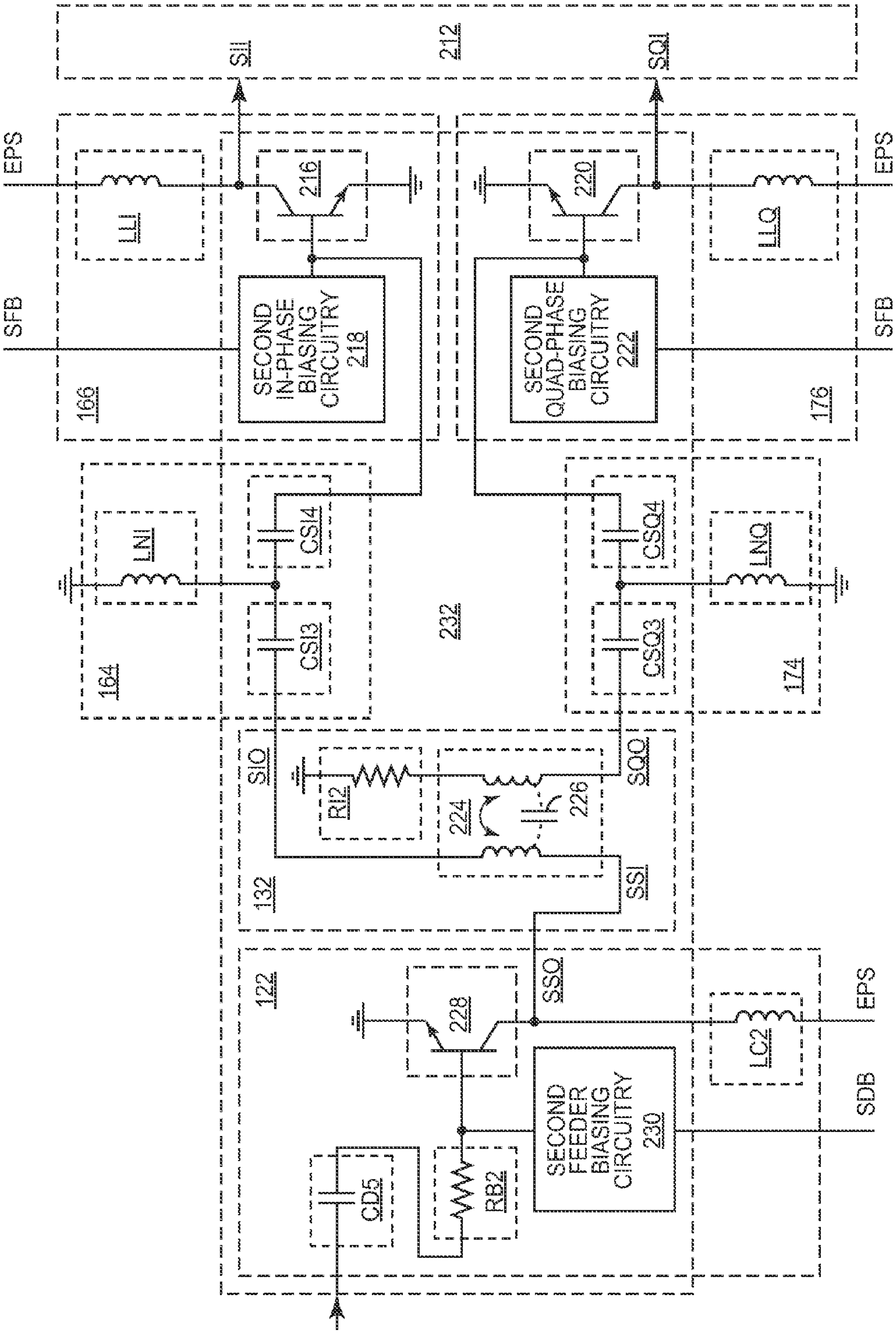
212
SII
SQI
EPS
LLI
216
SFB
SECOND IN-PHASE BIASING CIRCUITRY 218
166
220
LLQ
EPS
SECOND QUAD-PHASE BIASING CIRCUITRY 222
SFB
176
LNI
CSI4
CSI3
164
232
CSQ4
CSQ3
LNQ
174
SIO
RI2
224
226
SQO
132
SSI
122
228
SSO
LC2
EPS
SDB
SECOND FEEDER BIASING CIRCUITRY 230
CD5
RB2

FIG. 34

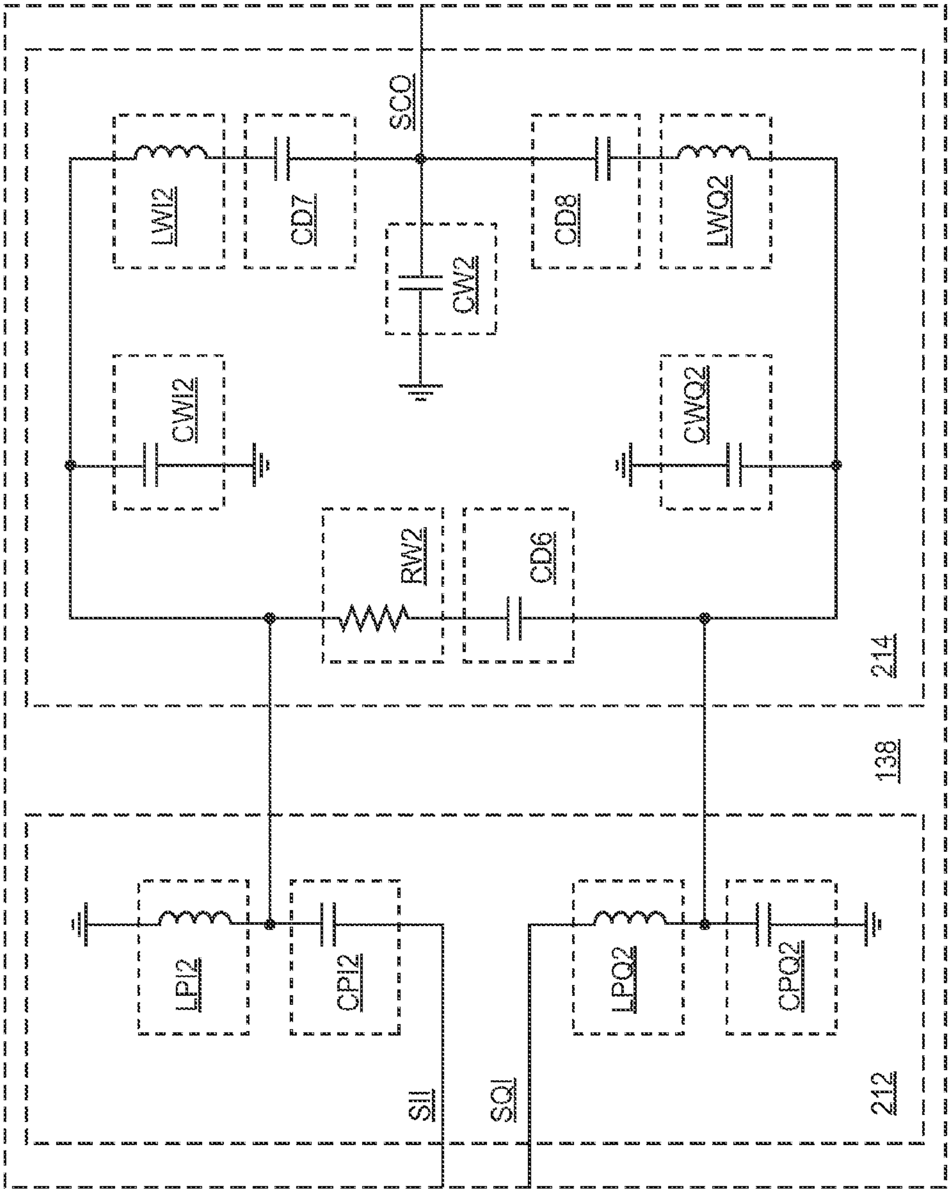
SCO
LWI2
CD7
CW2
CD8
LWQ2
CWI2
CWQ2
RW2
CD6
214
138
LPI2
CPI2
LPQ2
CPQ2
SII
SQI
212

FIG. 35

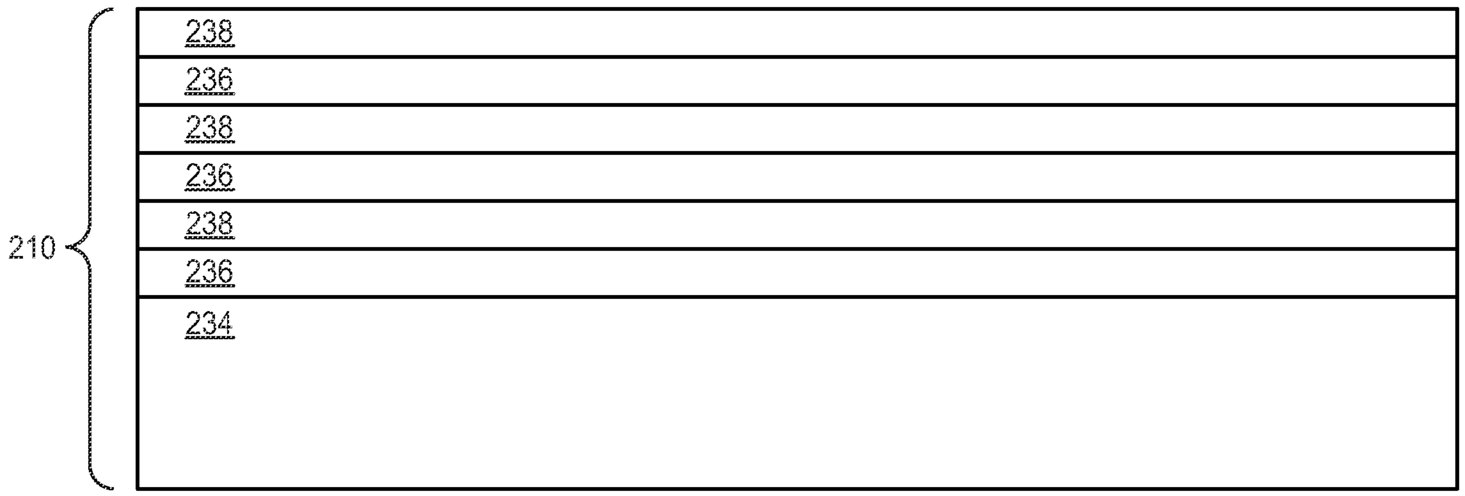
238
236
238
236
238
236
234
210

FIG. 36

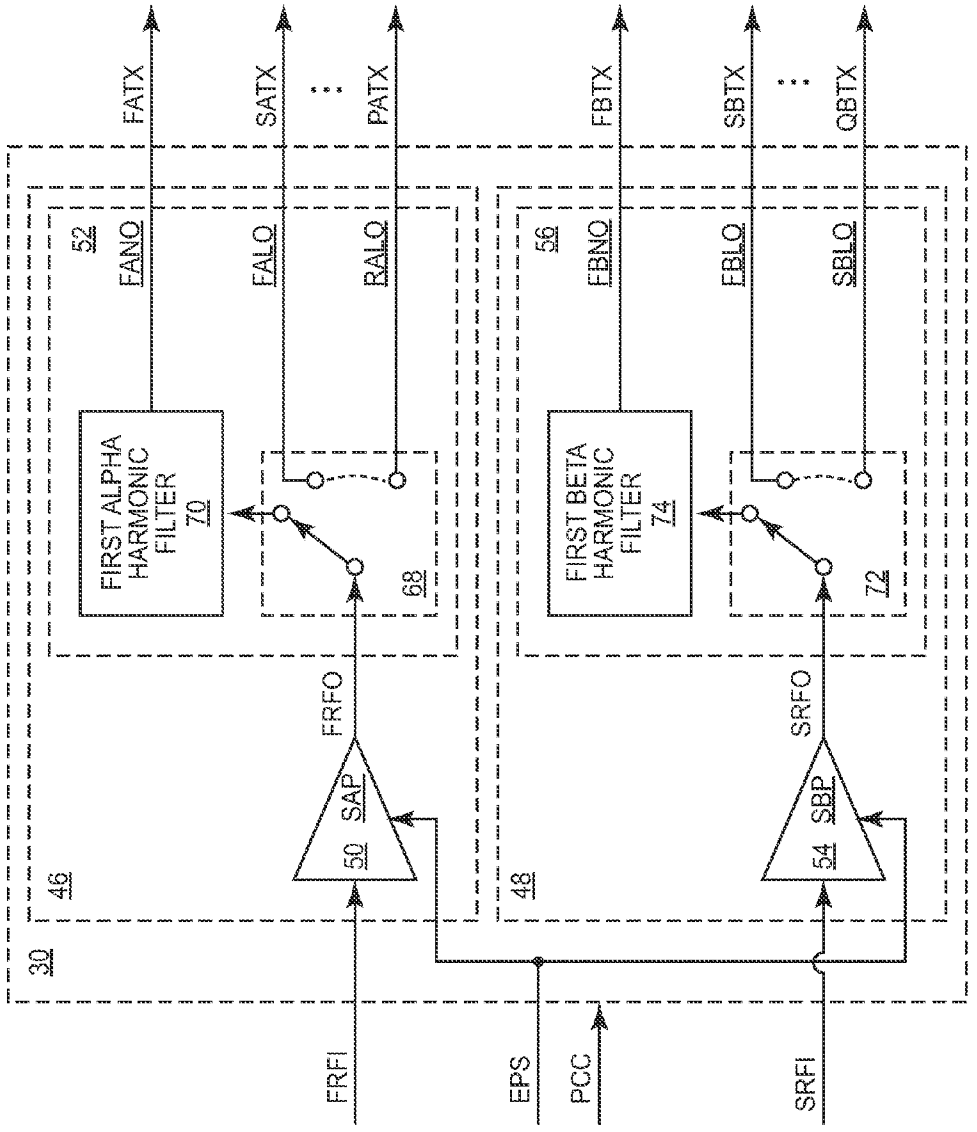
FATX
SATX
...
PATX
FBTX
SBTX
...
QBTX
30
46
48
52
56
FANO
FALO
RALO
FBNO
FBLO
SBLO
FIRST ALPHA HARMONIC FILTER 70
FIRST BETA HARMONIC FILTER 74
68
72
FRFO
SRFO
SAP 50
SBP 54
FRFI
EPS
PCC
SRFI

FIG. 37

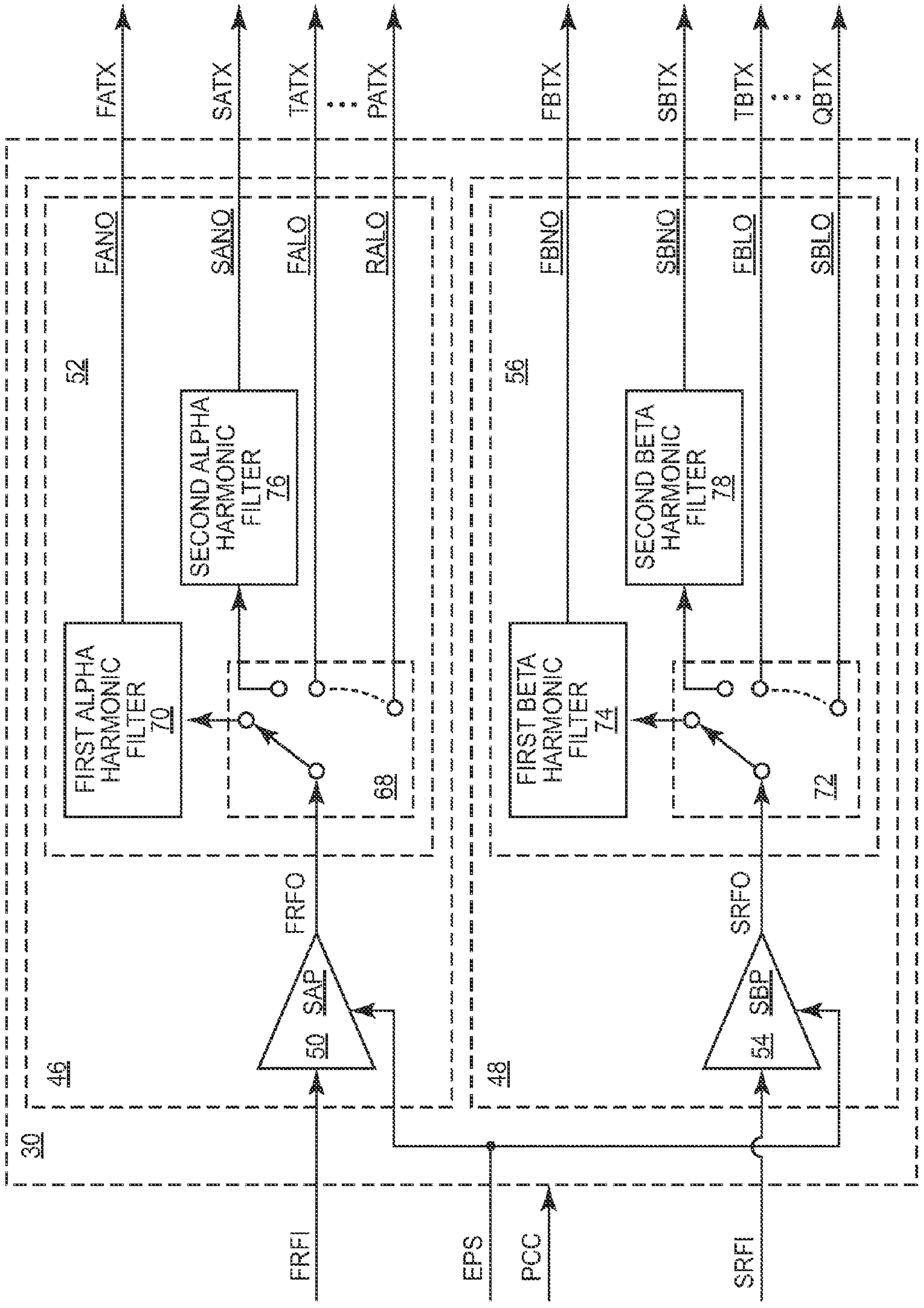
FRFI
EPS
PCC
SRFI
30
46
48
50
SAP
54
SBP
FRFO
SRFO
52
56
68
72
FIRST ALPHA HARMONIC FILTER 70
SECOND ALPHA HARMONIC FILTER 76
FIRST BETA HARMONIC FILTER 74
SECOND BETA HARMONIC FILTER 78
FANO
SANO
FALO
RALO
FBNO
SBNO
FBLO
SBLO
FATX
SATX
TATX
PATX
FBTX
SBTX
TBTX
QBTX

FIG. 38

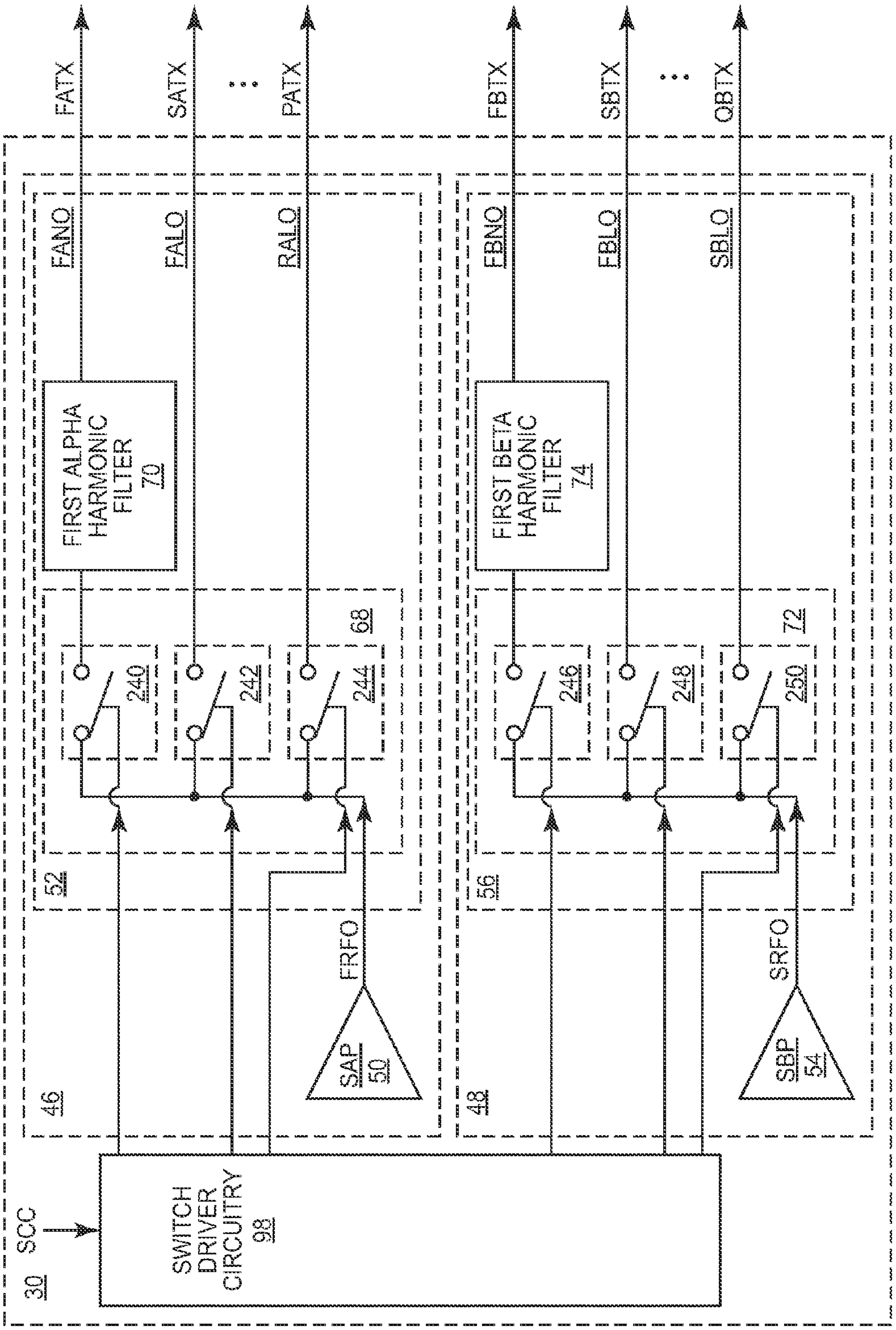
FATX
SATX
...
PATX
FBTX
SBTX
...
QBTX
FANO
FALO
RALO
FBNO
FBLO
SBLO
FIRST ALPHA HARMONIC FILTER 70
FIRST BETA HARMONIC FILTER 74
240
242
244
68
246
248
250
72
52
56
FRFO
SRFO
SAP 50
SBP 54
46
48
SCC
SWITCH DRIVER CIRCUITRY 98
30

FIG. 39

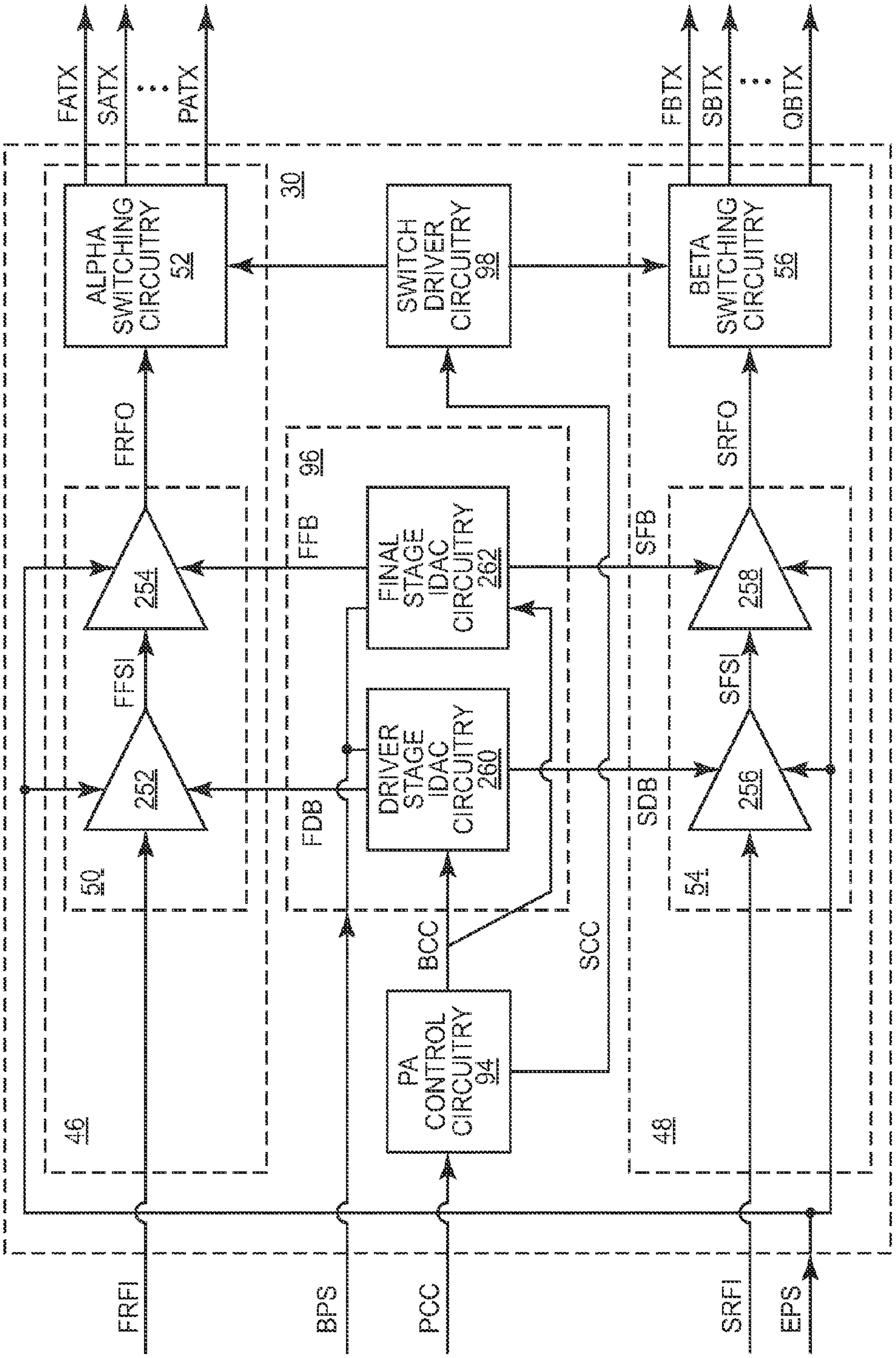
FATX
SATX
...
PATX
FBTX
SBTX
...
QBTX
30
ALPHA SWITCHING CIRCUITRY 52
SWITCH DRIVER CIRCUITRY 98
BETA SWITCHING CIRCUITRY 56
FRFO
SRFO
96
FFB
SFB
FINAL STAGE IDAC CIRCUITRY 262
254
258
FFSI
SFSI
DRIVER STAGE IDAC CIRCUITRY 260
252
256
FDB
SDB
50
54
BCC
SCC
PA CONTROL CIRCUITRY 94
46
48
FRFI
BPS
PCC
SRFI
EPS

FIG. 40

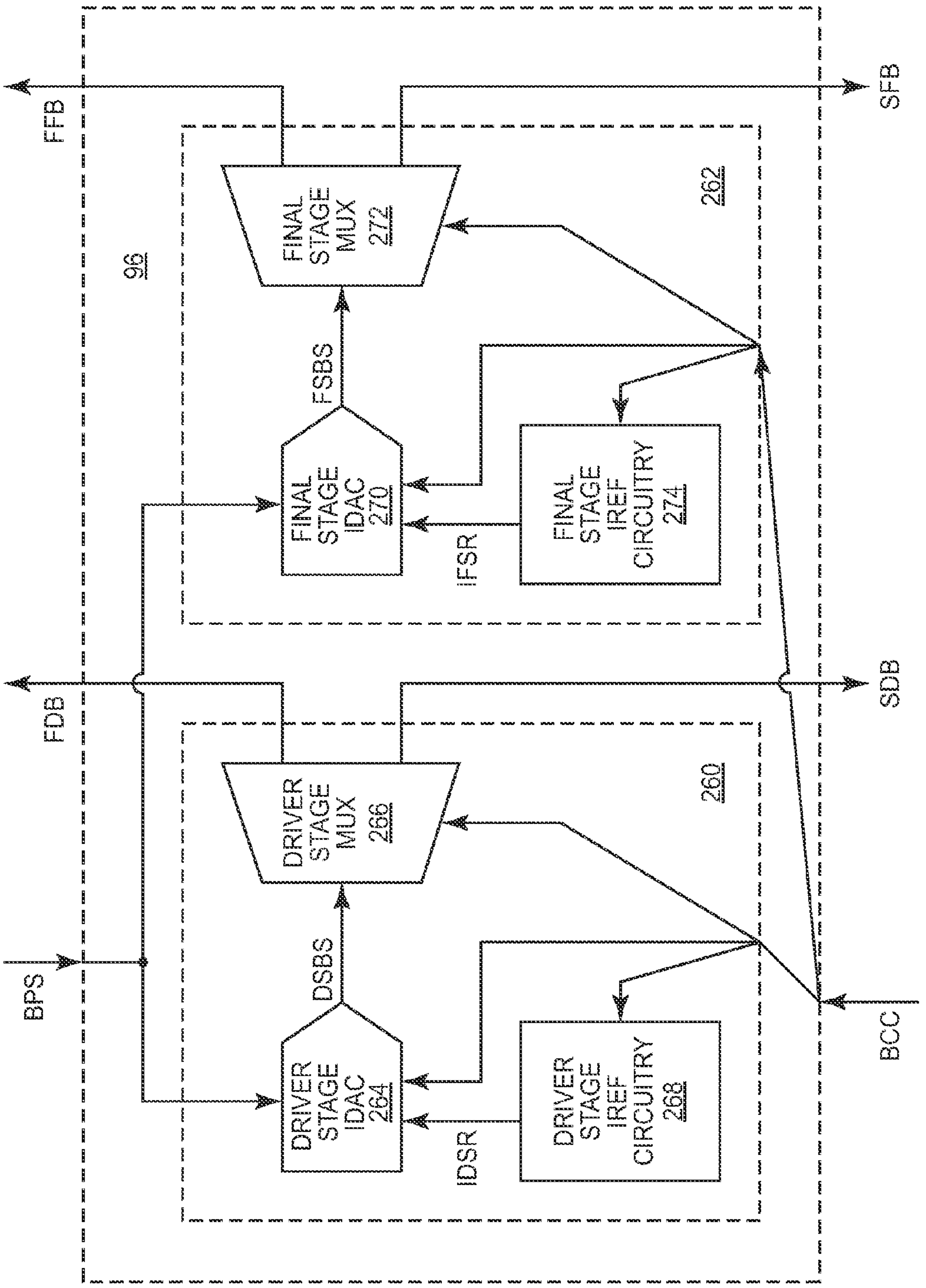
FFB
SFB
96
FINAL STAGE MUX 272
262
FSBS
FINAL STAGE IDAC 270
IFSR
FINAL STAGE IREF CIRCUITRY 274
FDB
SDB
DRIVER STAGE MUX 266
260
DSBS
DRIVER STAGE IDAC 264
IDSR
DRIVER STAGE IREF CIRCUITRY 268
BPS
BCC

FIG. 41

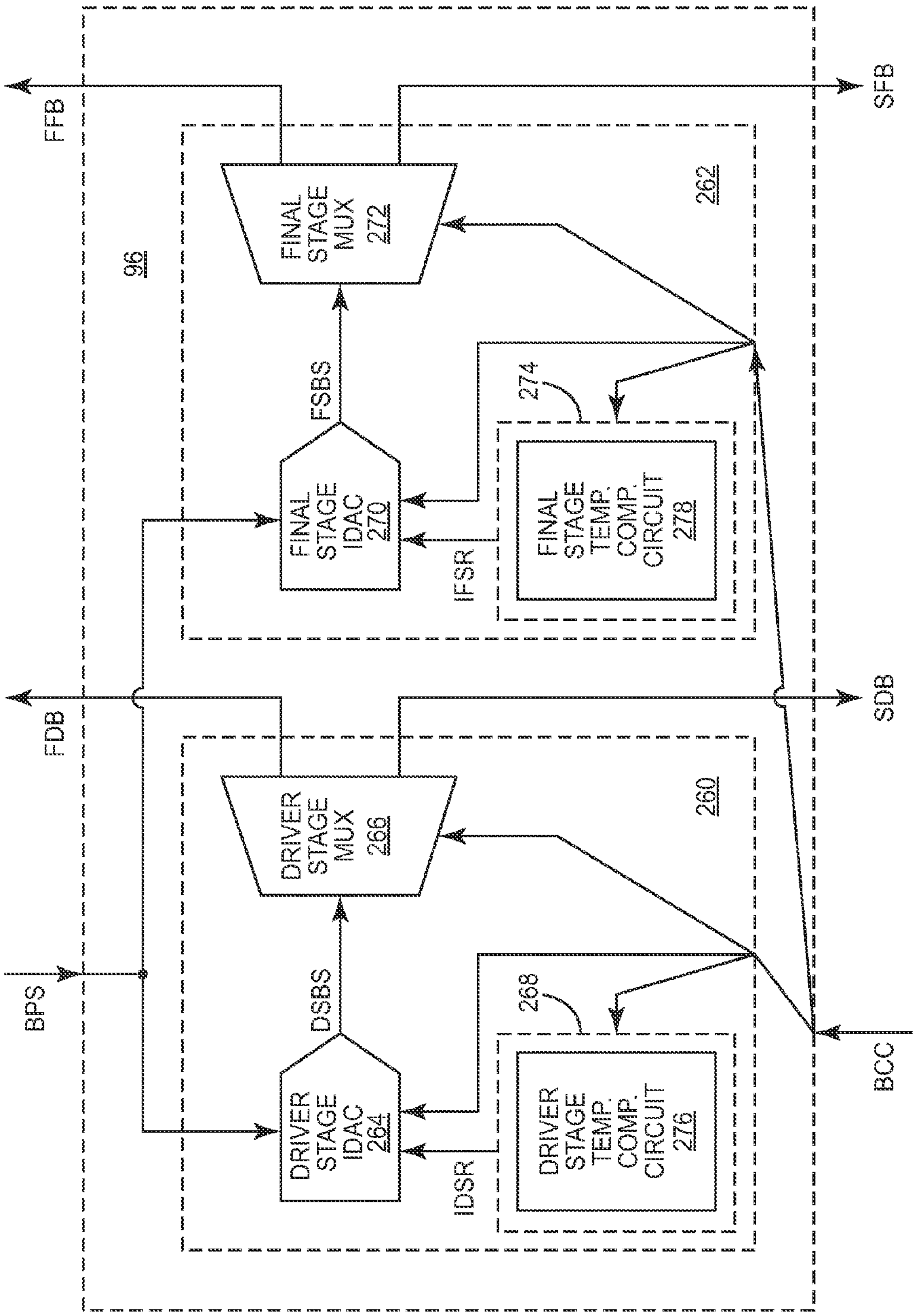
FFB
SFB
96
FINAL STAGE MUX 272
262
FSBS
274
FINAL STAGE IDAC 270
IFSR
FINAL STAGE TEMP. COMP. CIRCUIT 278
FDB
SDB
DRIVER STAGE MUX 266
260
DSBS
268
DRIVER STAGE IDAC 264
IDSR
DRIVER STAGE TEMP. COMP. CIRCUIT 276
BPS
BCC

FIG. 42

INTERFERENCE REDUCTION BETWEEN RF COMMUNICATIONS BANDS

PRIORITY CLAIMS

The present application claims priority to U.S. provisional patent application No. 61/489,487, filed May 24, 2011.

The present application claims priority to and is a continuation-in-part of U.S. patent application Ser. No. 13/090,663, filed Apr. 20, 2011, entitled "QUADRATURE POWER AMPLIFIER ARCHITECTURE," now U.S. Pat. No. 8,538,355, which claims priority to U.S. Provisional Patent Application No. 61/325,859, filed Apr. 20, 2010; No. 61/359,487, filed Jun. 29, 2010; No. 61/370,554, filed Aug. 4, 2010; No. 61/380,522, filed Sep. 7, 2010; No. 61/410,071, filed Nov. 4, 2010; and No. 61/417,633, filed Nov. 29, 2010.

The present application claims priority to and is a continuation-in-part of U.S. Patent Application No. 13/172,371, filed Jun. 29, 2011, entitled "AUTOMATICALLY CONFIGURABLE 2-WIRE/3-WIRE SERIAL COMMUNICATIONS INTERFACE," now U.S. Pat. No. 8,983,409, which claims priority to U.S. Provisional Patent Applications No. 61/359,487, filed Jun. 29, 2010; No. 61/370,554, tiled Aug. 4, 2010; No. 61/380,522, filed Sep. 7, 2010; No. 61/410,071, filed Nov. 4, 2010; and No. 61/417,633, filed Nov. 29, 2010. U.S. patent application Ser. No. 13/172,371 is a continuation-in-part of U.S. patent application Ser. No. 13/090,663, filed Apr. 20, 2011, which claims priority to U.S. Provisional Patent Applications No. 61/325,859, filed Apr. 20, 2010; No. 61/359,487, filed Jun. 29, 2010; No. 61/370,554, filed Aug. 4, 2010; No. 61/380,522, filed Sep. 7. 2010; No. 61/410,071, filed Nov. 4, 2010; and No. 61/417,633, filed Nov. 29, 2010.

The present application claims priority to and is a continuation-in-part of U.S. patent application Ser. No. 13/198,074, filed Aug. 4, 2011, entitled "FREQUENCY CORRECTION OF A PROGRAMMABLE FREQUENCY OSCILLATOR BY PROPAGATION DELAY COMPENSATION," now U.S. Pat. No. 8,515,361, which claims priority to U.S. Provisional Patent Applications No. 61/370,554 filed Aug. 4, 2010; No. 61/380,522, filed Sep. 7, 2010; No. 61/410,071, filed Nov. 4, 2010; and No. 61/417,633, filed Nov. 29, 2010. U.S. patent application Ser. No. 13/198,074 is a continuation-in-part of U.S. patent application Ser. No. 13/090,663, filed Apr. 20, 2011, which claims priority to U.S. Provisional Patent Applications No. 61/325,859, filed Apr. 20, 2010; No. 61/359,487, filed Jun. 29, 2010; No. 61/370,554, filed Aug. 4, 2010; Ser. No. 61/380,522, filed Sep. 7, 2010; No. 61/410,071, filed Nov. 4, 2010; and No. 61/417,633, filed Nov. 29, 2010. U.S. patent application Ser. No. 13/198,074 is also a continuation-in-part of U.S. patent application Ser. No. 13/172,371, filed Jun. 29, 2011, which claims priority to U.S. Provisional Patent Applications No. 61/359,487, filed Jun. 29, 2010; No. 61/370,554, filed Aug. 4, 2010; No. 61/380,522, filed Sep. 7, 2010; No. 61/410,071, filed Nov. 4, 2010; and No. 61/417,633, filed Nov. 29, 2010.

The present application claims priority to and is a continuation-in-part of U.S. patent application Ser. No. 13/226,831, filed Sep. 7, 2011, entitled "VOLTAGE COMPATIBLE CHARGE PUMP BUCK AND BUCK POWER SUPPLIES," which claims priority to U.S. Provisional Patent Application No. 61/380,522, filed Sep. 7, 2010; No. 61/410,071, filed Nov. 4, 2010; and No. 61/417,633, filed Nov. 29, 2010. U.S. patent application Ser. No. 13/226,831 is a continuation-in-part of U.S. patent application Ser. No. 13/090,663, filed Apr. 20, 2011, which claims priority to U.S. Provisional Patent Application No. 61/325,859, filed Apr. 20, 2010; No. 61/359,487, filed Jun. 29, 2010; No. 61/370,554, filed Aug. 4, 2010; No. 61/380,522, filed Sep. 7, 2010; No. 61/410,071, filed Nov. 4, 2010; and No. 61/417,633, filed Nov. 29, 2010. U.S. patent application Ser. No. 13/226,831 is also a continuation-in-part of U.S. patent application Ser. No. 13/172,371, filed Jun. 29, 2011, which claims priority to U.S. Provisional Patent Application No. 61/359,487, filed Jun. 29, 2010; No. 61/370,554, filed Aug. 4, 2010; No. 61/380,522, filed Sep. 7, 2010; No. 61/410,071, filed Nov. 4, 2010; and No. 61/417,633, filed Nov. 29, 2010. In addition, U.S. patent application Ser. No. 13/226,831 is a continuation-in-part of U.S. patent application Ser. No. 13/198,074, filed Aug. 4, 2011, which claims priority to U.S. Provisional Patent Application No. 61/370,554, filed Aug. 4, 2010; No. 61/380,522, filed Sep. 7, 2010; No. 61/410,071, filed Nov. 4, 2010; and No. 61/417,633, filed Nov. 29, 2010.

All of the applications listed above are hereby incorporated herein by reference in their entireties.

FIELD OF THE DISCLOSURE

Embodiments of the present disclosure relate to radio frequency (RF) power amplifier (PA) circuitry, which may be used in RF communications systems.

BACKGROUND OF THE DISCLOSURE

As wireless communications technologies evolve, wireless communications systems become increasingly sophisticated. As such, wireless communications protocols continue to expand and change to take advantage of the technological evolution. As a result, to maximize flexibility, many wireless communications devices must be capable of supporting any number of wireless communications protocols, including protocols that operate using different communications modes, such as a half-duplex mode or a full-duplex mode, and including protocols that operate using different frequency bands. Further, the different communications modes may include different types of RF modulation modes, each of which may have certain performance requirements, such as specific out-of-band emissions requirements or symbol differentiation requirements. In this regard, certain requirements may mandate operation in a linear mode. Other requirements may be less stringent that may allow operation in a non-linear mode to increase efficiency. Wireless communications devices that support such wireless communications protocols may be referred to as multi-mode multi-band communications devices. The linear mode relates to RF signals that include amplitude modulation (AM). The non-linear mode relates to RF signals that do not include AM. Since non-linear mode RF signals do not include AM, devices that amplify such signals may be allowed to operate in saturation. Devices that amplify linear mode RF signals may operate with some level of saturation, but must be able to retain AM characteristics sufficient for proper operation.

A half-duplex mode is a two-way mode of operation, in which a first transceiver communicates with a second transceiver; however, only one transceiver transmits at a time. Therefore, the transmitter and receiver in such a transceiver do not operate simultaneously. For example, certain telemetry systems operate in a send-then-wait-for-reply manner. Many time division duplex (TDD) systems, such as certain Global System for Mobile communications (GSM) systems, operate using the half-duplex mode. A full-duplex mode is a simultaneous two-way mode of operation, in which a first transceiver communicates with a second transceiver, and both transceivers may transmit simultaneously. Therefore, the transmitter and receiver in such a transceiver must be capable of operating simultaneously. In a full-duplex transceiver, signals from the transmitter should not overly interfere with signals received by the receiver; therefore, transmitted signals are at transmit frequencies that are different from received signals, which are at receive frequencies. Many frequency division duplex (FDD) systems, such as certain wideband code division multiple access (WCDMA) systems or certain long term evolution (LTE) systems, operate using a full-duplex mode.

As a result of the differences between full duplex operation and half duplex operation, RF front-end circuitry may need specific circuitry for each mode. Additionally, support of multiple frequency bands may require specific circuitry for each frequency band or for certain groupings of frequency bands. FIG. **1** shows a traditional multi-mode multi-band communications device **10** according to the prior art. The traditional multi-mode multi-band communications device **10** includes a traditional multi-mode multi-band transceiver **12**, traditional multi-mode multi-band PA circuitry **14**, traditional multi-mode multi-band front-end aggregation circuitry **16**, and an antenna **18**. The traditional multi-mode multi-band PA circuitry **14** includes a first traditional PA **20**, a second traditional PA **22**, and up to and including an N^{TH} traditional PA **24**.

The traditional multi-mode multi-band transceiver **12** may select one of multiple communications modes, which may include a half-duplex transmit mode, a half-duplex receive mode, a full-duplex mode, a linear mode, a non-linear mode, multiple RF modulation modes, or any combination thereof. Further, the traditional multi-mode multi-band transceiver **12** may select one of multiple frequency bands. The traditional multi-mode multi-band transceiver **12** provides an aggregation control signal ACS to the traditional multi-mode multi-band front-end aggregation circuitry **16** based on the selected mode and the selected frequency band. The traditional multi-mode multi-band front-end aggregation circuitry **16** may include various RF components, including RF switches; RF filters, such as bandpass filters, harmonic filters, and duplexers; RF amplifiers, such as low noise amplifiers (LNAs); impedance matching circuitry; the like; or any combination thereof. In this regard, routing of RF receive signals and RF transmit signals through the RF components may be based on the selected mode and the selected frequency band as directed by the aggregation control signal ACS.

The first traditional PA **20** may receive and amplify a first traditional RF transmit signal FTTX from the traditional multi-mode multi-band transceiver **12** to provide a first traditional amplified RF transmit signal FTATX to the antenna **18** via the traditional multi-mode multi-band front-end aggregation circuitry **16**. The second traditional PA **22** may receive and amplify a second traditional RF transmit signal STTX from the traditional multi-mode multi-band transceiver **12** to provide a second traditional RF amplified transmit signal STATX to the antenna **18** via the traditional multi-mode multi-band front-end aggregation circuitry **16**. The N^{TH} traditional PA **24** may receive an amplify an N^{TH} traditional RF transmit signal NTTX from the traditional multi-mode multi-band transceiver **12** to provide an N^{TH} traditional RF amplified transmit signal NTATX to the antenna **18** via the traditional multi-mode multi-band front-end aggregation circuitry **16**.

The traditional multi-mode multi-band transceiver **12** may receive a first RF receive signal FRX, a second RF receive signal SRX, and up to and including an M^{TH} RF receive signal MRX from the antenna **18** via the traditional multi-mode multi-band front-end aggregation circuitry **16**. Each of the RF receive signals FRX, SRX, MRX may be associated with at least one selected mode, at least one selected frequency band, or both. Similarly, each of the traditional RF transmit signals FTTX, STTX, NTTX and corresponding traditional amplified RF transmit signals FTATX, STATX, NTATX may be associated with at least one selected mode, at least one selected frequency band, or both.

Portable wireless communications devices are typically battery powered, need to be relatively small, and have low cost. As such, to minimize size, cost, and power consumption, multi-mode multi-band RF circuitry in such a device needs to be as simple, small, and efficient as is practical. Thus, there is a need for multi-mode multi-band RF circuitry in a multi-mode multi-band communications device that is low cost, small, simple, efficient, and meets performance requirements.

SUMMARY OF THE EMBODIMENTS

Embodiments of the present disclosure relate to RF PA circuitry and a PA envelope power supply. The RF PA circuitry receives and amplifies an RF input signal to provide an RF output signal using an envelope power supply signal, which is provided by the PA envelope power supply. The RF PA circuitry operates in either a normal RF spectral emissions mode or a reduced RF spectral emissions mode. When reduced RF spectral emissions are required, the RF PA circuitry operates in the reduced RF spectral emissions mode. As such, at a given RF output power, during the reduced RF spectral emissions mode, RF spectral emissions of the RF output signal are less than during the normal RF spectral emissions mode. As a result, the reduced RF spectral emissions mode may be used to reduce interference between RF communications bands.

Those skilled in the art will appreciate the scope of the present disclosure and realize additional aspects thereof after reading the following detailed description of the preferred embodiments in association with the accompanying drawing figures.

BRIEF DESCRIPTION OF THE DRAWING FIGURES

The accompanying drawing figures incorporated in and forming a part of this specification illustrate several aspects of the disclosure, and together with the description serve to explain the principles of the disclosure.

FIG. **1** shows a traditional multi-mode multi-band communications device according to the prior art.

FIG. **2** shows an RF communications system according to one embodiment of the RF communications system.

FIG. **3** shows the RF communications system according to an alternate embodiment of the RF communications system.

FIG. **4** shows the RF communications system according to an additional embodiment of the RF communications system.

FIG. **5** shows the RF communications system according to another embodiment of the RF communications system.

FIG. **6** shows the RF communications system according to a further embodiment of the RF communications system.

FIG. **7** shows the RF communications system according to one embodiment of the RF communications system.

FIG. **8** shows details of RF power amplifier (PA) circuitry illustrated in FIG. **5** according to one embodiment of the RF PA circuitry.

FIG. **9** shows details of the RF PA circuitry illustrated in FIG. **5** according to an alternate embodiment of the RF PA circuitry.

FIG. **10** shows the RF communications system according to one embodiment of the RF communications system.

FIG. **11** shows the RF communications system according to an alternate embodiment of the RF communications system.

FIG. **12** shows details of a direct current (DC)-DC converter illustrated in FIG. **11** according to an alternate embodiment of the DC-DC converter.

FIG. **13** shows details of the RF PA circuitry illustrated in FIG. **5** according to one embodiment of the RF PA circuitry.

FIG. **14** shows details of the RF PA circuitry illustrated in FIG. **6** according to an alternate embodiment of the RF PA circuitry.

FIG. **15** shows details of a first RF PA and a second RF PA illustrated in FIG. **14** according to one embodiment of the first RF PA and the second RF PA.

FIG. **16** shows details of a first non-quadrature PA path and a second non-quadrature PA path illustrated in FIG. **15** according to one embodiment of the first non-quadrature PA path and the second non-quadrature PA path.

FIG. **17** shows details of a first quadrature PA path and a second quadrature PA path illustrated in FIG. **15** according to one embodiment of the first quadrature PA path and the second quadrature PA path.

FIG. **18** shows details of a first in-phase amplification path, a first quadrature-phase amplification path, a second in-phase amplification path, and a second quadrature-phase amplification path illustrated in FIG. **17** according to one embodiment of the first in-phase amplification path, the first quadrature-phase amplification path, the second in-phase amplification path, and the second quadrature-phase amplification path.

FIG. **19** shows details of the first quadrature PA path and the second quadrature PA path illustrated in FIG. **15** according to an alternate embodiment of the first quadrature PA path and the second quadrature PA path.

FIG. **20** shows details of the first in-phase amplification path, the first quadrature-phase amplification path, the second in-phase amplification path, and the second quadrature-phase amplification path illustrated in FIG. **19** according to an alternate embodiment of the first in-phase amplification path, the first quadrature-phase amplification path, the second in-phase amplification path, and the second quadrature-phase amplification path.

FIG. **21** shows details of the first RF PA and the second RF PA illustrated in FIG. **14** according an alternate embodiment of the first RF PA and the second RF PA.

FIG. **22** shows details of the first non-quadrature PA path, the first quadrature PA path, and the second quadrature PA path illustrated in FIG. **21** according to an additional embodiment of the first non-quadrature PA path, the first quadrature PA path, and the second quadrature PA path.

FIG. **23** shows details of a first feeder PA stage and a first quadrature RF splitter illustrated in FIG. **16** and FIG. **17**, respectively, according to one embodiment of the first feeder PA stage and the first quadrature RF splitter.

FIG. **24** shows details of the first feeder PA stage and the first quadrature RF splitter illustrated in FIG. **16** and FIG. **17**, respectively, according to an alternate embodiment of the first feeder PA stage and the first quadrature RF splitter.

FIG. **25** is a graph illustrating output characteristics of a first output transistor element illustrated in FIG. **24** according to one embodiment of the first output transistor element.

FIG. **26** illustrates a process for matching an input impedance to a quadrature RF splitter to a target load line of a feeder PA stage.

FIG. **27** shows details of the first RF PA illustrated in FIG. **14** according an alternate embodiment of the first RF PA.

FIG. **28** shows details of the second RF PA illustrated in FIG. **14** according an alternate embodiment of the second RF PA.

FIG. **29** shows details of a first in-phase amplification path, a first quadrature-phase amplification path, and a first quadrature RF combiner illustrated in FIG. **22** according to one embodiment of the first in-phase amplification path, the first quadrature-phase amplification path, and the first quadrature RF combiner.

FIG. **30** shows details of a first feeder PA stage, a first quadrature RF splitter, a first in-phase final PA impedance matching circuit, a first in-phase final PA stage, a first quadrature-phase final PA impedance matching circuit, and a first quadrature-phase final PA stage illustrated in FIG. **29** according to one embodiment of the first feeder PA stage, the first quadrature RF splitter, the first in-phase final PA impedance matching circuit, the first in-phase final PA stage, the first quadrature-phase final PA impedance matching circuit, and the first quadrature-phase final PA stage.

FIG. **31** shows details of the first feeder PA stage, the first quadrature RF splitter, the first in-phase final PA impedance matching circuit, the first in-phase final PA stage, the first quadrature-phase final PA impedance matching circuit, and the first quadrature-phase final PA stage illustrated in FIG. **29** according to an alternate embodiment of the first feeder PA stage, the first quadrature RF splitter, the first in-phase final PA impedance matching circuit, the first in-phase final PA stage, the first quadrature-phase final PA impedance matching circuit, and the first quadrature-phase final PA stage.

FIG. **32** shows details of first phase-shifting circuitry and a first Wilkinson RF combiner illustrated in FIG. **29** according to one embodiment of the first phase-shifting circuitry and the first Wilkinson RF combiner.

FIG. **33** shows details of the second non-quadrature PA path illustrated in FIG. **16** and details of the second quadrature PA path illustrated in FIG. **18** according to one embodiment of the second non-quadrature PA path and the second quadrature PA path.

FIG. **34** shows details of a second feeder PA stage, a second quadrature RF splitter, a second in-phase final PA impedance matching circuit, a second in-phase final PA stage, a second quadrature-phase final PA impedance matching circuit, and a second quadrature-phase final PA stage illustrated in FIG. **33** according to one embodiment of the second feeder PA stage, the second quadrature RF splitter, the second in-phase final PA impedance matching circuit, the second in-phase final PA stage, the second quadrature-phase final PA impedance matching circuit, and the second quadrature-phase final PA stage.

FIG. **35** shows details of second phase-shifting circuitry and a second Wilkinson RF combiner illustrated in FIG. **33** according to one embodiment of the second phase-shifting circuitry and the second Wilkinson RF combiner.

FIG. **36** shows details of a first PA semiconductor die illustrated in FIG. **30** according to one embodiment of the first PA semiconductor die.

FIG. **37** shows details of the RF PA circuitry illustrated in FIG. **5** according to one embodiment of the RF PA circuitry.

FIG. **38** shows details of the RF PA circuitry illustrated in FIG. **5** according to an alternate embodiment of the RF PA circuitry.

FIG. **39** shows details of the RF PA circuitry illustrated in FIG. **5** according to an additional embodiment of the RF PA circuitry.

FIG. **40** shows details of the first RF PA, the second RF PA, and PA bias circuitry illustrated in FIG. **13** according to one embodiment of the first RF PA, the second RF PA, and the PA bias circuitry.

FIG. **41** shows details of driver stage current digital-to-analog converter (IDAC) circuitry and final stage IDAC circuitry illustrated in FIG. **40** according to one embodiment of the driver stage IDAC circuitry and the final stage IDAC circuitry.

FIG. **42** shows details of driver stage current reference circuitry and final stage current reference circuitry illustrated in FIG. **41** according to one embodiment of the driver stage current reference circuitry and the final stage current reference circuitry.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The embodiments set forth below represent the necessary information to enable those skilled in the art to practice the disclosure and illustrate the best mode of practicing the disclosure. Upon reading the following description in light of the accompanying drawing figures, those skilled in the art will understand the concepts of the disclosure and will recognize applications of these concepts not particularly addressed herein. It should be understood that these concepts and applications fall within the scope of the disclosure and the accompanying claims.

FIG. **2** shows an RF communications system **26** according to one embodiment of the RF communications system **26**. The RF communications system **26** includes RF modulation and control circuitry **28**, RF PA circuitry **30**, and a DC-DC converter **32**. The RF modulation and control circuitry **28** provides an envelope control signal ECS to the DC-DC converter **32** and provides an RF input signal RFI to the RF PA circuitry **30**. The DC-DC converter **32** provides a bias power supply signal BPS and an envelope power supply signal EPS to the RF PA circuitry **30**. The envelope power supply signal EPS may be based on the envelope control signal ECS. As such, a magnitude of the envelope power supply signal EPS may be controlled by the RF modulation and control circuitry **28** via the envelope control signal ECS. The RF PA circuitry **30** may receive and amplify the RF input signal RFI to provide an RF output signal RFO. The envelope power supply signal EPS may provide power for amplification of the RF input signal RFI to the RF PA circuitry **30**. The RF PA circuitry **30** may use the bias power supply signal BPS to provide biasing of amplifying elements in the RF PA circuitry **30**.

In a first embodiment of the RF communications system **26**, the RF communications system **26** is a multi-mode RF communications system **26**. As such, the RF communications system **26** may operate using multiple communications modes. In this regard, the RF modulation and control circuitry **28** may be multi-mode RF modulation and control circuitry **28** and the RF PA circuitry **30** may be multi-mode RF PA circuitry **30**. In a second embodiment of the RF communications system **26**, the RF communications system **26** is a multi-band RF communications system **26**. As such, the RF communications system **26** may operate using multiple RF communications bands. In this regard, the RF modulation and control circuitry **28** may be multi-band RF modulation and control circuitry **28** and the RF PA circuitry **30** may be multi-band RF PA circuitry **30**. In a third embodiment of the RF communications system **26**, the RF communications system **26** is a multi-mode multi-band RF communications system **26**. As such, the RF communications system **26** may operate using multiple communications modes, multiple RF communications bands, or both. In this regard, the RF modulation and control circuitry **28** may be multi-mode multi-band RF modulation and control circuitry **28** and the RF PA circuitry **30** may be multi-mode multi-band RF PA circuitry **30**.

The communications modes may be associated with any number of different communications protocols, such as Global System of Mobile communications (GSM), Gaussian Minimum Shift Keying (GMSK), IS-136, Enhanced Data rates for GSM Evolution (EDGE), Code Division Multiple Access (CDMA), Universal Mobile Telecommunications System (UMTS) protocols, such as Wideband CDMA (WCDMA), Worldwide Interoperability for Microwave Access (WIMAX), Long Term Evolution (LTE), or the like. The GSM, GMSK, and IS-136 protocols typically do not include amplitude modulation (AM). As such, the GSM, GMSK, and IS-136 protocols may be associated with a non-linear mode. Further, the GSM, GMSK, and IS-136 protocols may be associated with a saturated mode. The EDGE, CDMA, UMTS, WCDMA, WIMAX, and LTE protocols may include AM. As such, the EDGE, CDMA, UMTS, WCDMA, WIMAX, and LTE protocols may be associated with a linear mode.

In one embodiment of the RF communications system **26**, the RF communications system **26** is a mobile communications terminal, such as a cell phone, smartphone, laptop computer, tablet computer, personal digital assistant (PDA), or the like. In an alternate embodiment of the RF communications system **26**, the RF communications system **26** is a fixed communications terminal, such as a base station, a cellular base station, a wireless router, a hotspot distribution node, a wireless access point, or the like. The antenna **18** may include any apparatus for conveying RF transmit and RF receive signals to and from at least one other RF communications system. As such, in one embodiment of the antenna **18**, the antenna **18** is a single antenna. In an alternate embodiment of the antenna **18**, the antenna **18** is an antenna array having multiple radiating and receiving elements. In an additional embodiment of the antenna **18**, the antenna **18** is a distribution system for transmitting and receiving RF signals.

FIG. **3** shows the RF communications system **26** according to an alternate embodiment of the RF communications system **26**. The RF communications system **26** illustrated in FIG. **3** is similar to the RF communications system **26** illustrated in FIG. **2**, except in the RF communications system **26** illustrated in FIG. **3**, the RF modulation and control circuitry **28** provides a first RF input signal FRFI, a second RF input signal SRFI, and a PA configuration control signal PCC to the RF PA circuitry **30**. The RF PA circuitry **30** may receive and amplify the first RF input signal FRFI to provide a first RF output signal FRFO. The envelope power supply signal EPS may provide power for amplification of the first RF input signal FRFI to the RF PA circuitry **30**. The RF PA circuitry **30** may receive and amplify the second RF input signal SRFI to provide a second RF output signal SRFO. The envelope power supply signal EPS may provide power for amplification of the second RF output signal SRFO to the RF PA circuitry **30**. Certain configurations of the RF PA circuitry **30** may be based on the PA configuration control signal PCC. As a result, the RF modulation and control circuitry **28** may control such configurations of the RF PA circuitry **30**.

FIG. **4** shows the RF communications system **26** according to an additional embodiment of the RF communications system **26**. The RF communications system **26** illustrated in FIG. **4** is similar to the RF communications system **26** illustrated in FIG. **3**, except in the RF communications system **26** illustrated in FIG. **4**, the RF PA circuitry **30** does not provide the first RF output signal FRFO and the second RF output signal

SRFO. Instead, the RF PA circuitry **30** may provide one of a first alpha RF transmit signal FATX, a second alpha RF transmit signal SATX, and up to and including a P^{TH} alpha RF transmit signal PATX based on receiving and amplifying the first RF input signal FRFI. Similarly, the RF PA circuitry **30** may provide one of a first beta RF transmit signal FBTX, a second beta RF transmit signal SBTX, and up to and including a Q^{TH} beta RF transmit signal QBTX based on receiving and amplifying the second RF input signal SRFI. The one of the transmit signals FATX, SATX, PATX, FBTX, SBTX, QBTX that is selected may be based on the PA configuration control signal PCC. Additionally, the RF modulation and control circuitry **28** may provide a DC configuration control signal DCC to the DC-DC converter **32**. Certain configurations of the DC-DC converter **32** may be based on the DC configuration control signal DCC.

FIG. **5** shows the RF communications system **26** according to another embodiment of the RF communications system **26**. The RF communications system **26** illustrated in FIG. **5** shows details of the RF modulation and control circuitry **28** and the RF PA circuitry **30** illustrated in FIG. **4**. Additionally, the RF communications system **26** illustrated in FIG. **5** further includes transceiver circuitry **34**, front-end aggregation circuitry **36**, and the antenna **18**. The transceiver circuitry **34** includes down-conversion circuitry **38**, baseband processing circuitry **40**, and the RF modulation and control circuitry **28**, which includes control circuitry **42** and RF modulation circuitry **44**. The RF PA circuitry **30** includes a first transmit path **46** and a second transmit path **48**. The first transmit path **46** includes a first RF PA **50** and alpha switching circuitry **52**. The second transmit path **48** includes a second RF PA **54** and beta switching circuitry **56**. The front-end aggregation circuitry **36** is coupled to the antenna **18**. The control circuitry **42** provides the aggregation control signal ACS to the front-end aggregation circuitry **36**. Configuration of the front-end aggregation circuitry **36** may be based on the aggregation control signal ACS. As such, configuration of the front-end aggregation circuitry **36** may be controlled by the control circuitry **42** via the aggregation control signal ACS.

The control circuitry **42** provides the envelope control signal ECS and the DC configuration control signal DCC to the DC-DC converter **32**. Further, the control circuitry **42** provides the PA configuration control signal PCC to the RF PA circuitry **30**. As such, the control circuitry **42** may control configuration of the RF PA circuitry **30** via the PA configuration control signal PCC and may control a magnitude of the envelope power supply signal EPS via the envelope control signal ECS. The control circuitry **42** may select one of multiple communications modes, which may include a first half-duplex transmit mode, a first half-duplex receive mode, a second half-duplex transmit mode, a second half-duplex receive mode, a first full-duplex mode, a second full-duplex mode, at least one linear mode, at least one non-linear mode, multiple RF modulation modes, or any combination thereof. Further, the control circuitry **42** may select one of multiple frequency bands. The control circuitry **42** may provide the aggregation control signal ACS to the front-end aggregation circuitry **36** based on the selected mode and the selected frequency band. The front-end aggregation circuitry **36** may include various RF components, including RF switches; RF filters, such as bandpass filters, harmonic filters, and duplexers; RF amplifiers, such as low noise amplifiers (LNAs); impedance matching circuitry; the like; or any combination thereof. In this regard, routing of RF receive signals and RF transmit signals through the RF components may be based on the selected mode and the selected frequency band as directed by the aggregation control signal ACS.

The down-conversion circuitry **38** may receive the first RF receive signal FRX, the second RF receive signal SRX, and up to and including the M^{TH} RF receive signal MRX from the antenna **18** via the front-end aggregation circuitry **36**. Each of the RF receive signals FRX, SRX, MRX may be associated with at least one selected mode, at least one selected frequency band, or both. The down-conversion circuitry **38** may down-convert any of the RF receive signals FRX, SRX, MRX to baseband receive signals, which may be forwarded to the baseband processing circuitry **40** for processing. The baseband processing circuitry **40** may provide baseband transmit signals to the RF modulation circuitry **44**, which may RF modulate the baseband transmit signals to provide the first RF input signal FRFI or the second RF input signal SRFI to the first RF PA **50** or the second RF PA **54**, respectively, depending on the selected communications mode.

The first RF PA **50** may receive and amplify the first RF input signal FRFI to provide the first RF output signal FRFO to the alpha switching circuitry **52**. Similarly, the second RF PA **54** may receive and amplify the second RF input signal SRFI to provide the second RF output signal SRFO to the beta switching circuitry **56**. The first RF PA **50** and the second RF PA **54** may receive the envelope power supply signal EPS, which may provide power for amplification of the first RF input signal FRFI and the second RF input signal SRFI, respectively. The alpha switching circuitry **52** may forward the first RF output signal FRFO to provide one of the alpha transmit signals FATX, SATX, PATX to the antenna **18** via the front-end aggregation circuitry **36**, depending on the selected communications mode based on the PA configuration control signal PCC. Similarly, the beta switching circuitry **56** may forward the second RF output signal SRFO to provide one of the beta transmit signals FBTX, SBTX, QBTX to the antenna **18** via the front-end aggregation circuitry **36**, depending on the selected communications mode based on the PA configuration control signal PCC.

FIG. **6** shows the RF communications system **26** according to a further embodiment of the RF communications system **26**. The RF communications system **26** illustrated in FIG. **6** is similar to the RF communications system **26** illustrated in FIG. **5**, except in the RF communications system **26** illustrated in FIG. **6**, the transceiver circuitry **34** includes a control circuitry digital communications interface (DCI) **58**, the RF PA circuitry **30** includes a PA-DCI **60**, the DC-DC converter **32** includes a DC-DC converter DCI **62**, and the front-end aggregation circuitry **36** includes an aggregation circuitry DCI **64**. The front-end aggregation circuitry **36** includes an antenna port AP, which is coupled to the antenna **18**. In one embodiment of the RF communications system **26**, the antenna port AP is directly coupled to the antenna **18**. In one embodiment of the RF communications system **26**, the front-end aggregation circuitry **36** is coupled between the alpha switching circuitry **52** and the antenna port AP. Further, the front-end aggregation circuitry **36** is coupled between the beta switching circuitry **56** and the antenna port AP. The alpha switching circuitry **52** may be multi-mode multi-band alpha switching circuitry and the beta switching circuitry **56** may be multi-mode multi-band beta switching circuitry.

The DCIs **58**, **60**, **62**, **64** are coupled to one another using a digital communications bus **66**. In the digital communications bus **66** illustrated in FIG. **6**, the digital communications bus **66** is a uni-directional bus in which the control circuitry DCI **58** may communicate information to the PA-DCI **60**, the DC-DC converter DCI **62**, the aggregation circuitry DCI **64**, or any combination thereof. As such, the control circuitry **42** may provide the envelope control signal ECS and the DC configuration control signal DCC via the control circuitry

DCI **58** to the DC-DC converter **32** via the DC-DC converter DCI **62**. Similarly, the control circuitry **42** may provide the aggregation control signal ACS via the control circuitry DCI **58** to the front-end aggregation circuitry **36** via the aggregation circuitry DCI **64**. Additionally, the control circuitry **42** may provide the PA configuration control signal PCC via the control circuitry DCI **58** to the RF PA circuitry **30** via the PA-DCI **60**.

FIG. **7** shows the RF communications system **26** according to one embodiment of the RF communications system **26**. The RF communications system **26** illustrated in FIG. **7** is similar to the RF communications system **26** illustrated in FIG. **6**, except in the RF communications system **26** illustrated in FIG. **7**, the digital communications bus **66** is a bi-directional bus and each of the DCIs **58**, **60**, **62**, **64** is capable of receiving or transmitting information. In alternate embodiments of the RF communications system **26**, any or all of the DCIs **58**, **60**, **62**, **64** may be uni-directional and any or all of the DCIs **58**, **60**, **62**, **64** may be bi-directional.

FIG. **8** shows details of the RF PA circuitry **30** illustrated in FIG. **5** according to one embodiment of the RF PA circuitry **30**. Specifically, FIG. **8** shows details of the alpha switching circuitry **52** and the beta switching circuitry **56** according to one embodiment of the alpha switching circuitry **52** and the beta switching circuitry **56**. The alpha switching circuitry **52** includes an alpha RF switch **68** and a first alpha harmonic filter **70**. The beta switching circuitry **56** includes a beta RF switch **72** and a first beta harmonic filter **74**. Configuration of the alpha RF switch **68** and the beta RF switch **72** may be based on the PA configuration control signal PCC. In one communications mode, such as an alpha half-duplex transmit mode, an alpha saturated mode, or an alpha non-linear mode, the alpha RF switch **68** is configured to forward the first RF output signal FRFO to provide the first alpha RF transmit signal FATX via the first alpha harmonic filter **70**. In another communications mode, such as an alpha full-duplex mode or an alpha linear mode, the alpha RF switch **68** is configured to forward the first RF output signal FRFO to provide any of the second alpha RF transmit signal SATX through the P^{TH} alpha RF transmit signal PATX. When a specific RF band is selected, the alpha RF switch **68** may be configured to provide a corresponding selected one of the second alpha RF transmit signal SATX through the P^{TH} alpha RF transmit signal PATX.

In one communications mode, such as a beta half-duplex transmit mode, a beta saturated mode, or a beta non-linear mode, the beta RF switch **72** is configured to forward the second RF output signal SRFO to provide the first beta RF transmit signal FBTX via the first beta harmonic filter **74**. In another communications mode, such as a beta full-duplex mode or a beta linear mode, the beta RF switch **72** is configured to forward the second RF output signal SRFO to provide any of the second beta RF transmit signal SBTX through the Q^{TH} beta RF transmit signal QBTX. When a specific RF band is selected, beta RF switch **72** may be configured to provide a corresponding selected one of the second beta RF transmit signal SBTX through the Q^{TH} beta RF transmit signal QBTX. The first alpha harmonic filter **70** may be used to filter out harmonics of an RF carrier in the first RF output signal FRFO. The first beta harmonic filter **74** may be used to filter out harmonics of an RF carrier in the second RF output signal SRFO.

FIG. **9** shows details of the RF PA circuitry **30** illustrated in FIG. **5** according to an alternate embodiment of the RF PA circuitry **30**. Specifically, FIG. **9** shows details of the alpha switching circuitry **52** and the beta switching circuitry **56** according to an alternate embodiment of the alpha switching circuitry **52** and the beta switching circuitry **56**. The alpha switching circuitry **52** includes the alpha RF switch **68**, the first alpha harmonic filter **70**, and a second alpha harmonic filter **76**. The beta switching circuitry **56** includes the beta RF switch **72**, the first beta harmonic filter **74**, and a second beta harmonic filter **78**. Configuration of the alpha RF switch **68** and the beta RF switch **72** may be based on the PA configuration control signal PCC. In one communications mode, such as a first alpha half-duplex transmit mode, a first alpha saturated mode, or a first alpha non-linear mode, the alpha RF switch **68** is configured to forward the first RF output signal FRFO to provide the first alpha RF transmit signal FATX via the first alpha harmonic filter **70**. In another communications mode, such as a second alpha half-duplex transmit mode, a second alpha saturated mode, or a second alpha non-linear mode, the alpha RF switch **68** is configured to forward the first RF output signal FRFO to provide the second alpha RF transmit signal SATX via the second alpha harmonic filter **76**. In an alternate communications mode, such as an alpha full-duplex mode or an alpha linear mode, the alpha RF switch **68** is configured to forward the first RF output signal FRFO to provide any of a third alpha RF transmit signal TATX through the P^{TH} alpha RF transmit signal PATX. When a specific RF band is selected, the alpha RF switch **68** may be configured to provide a corresponding selected one of the third alpha RF transmit signal TATX through the P^{TH} alpha RF transmit signal PATX.

In one communications mode, such as a first beta half-duplex transmit mode, a first beta saturated mode, or a first beta non-linear mode, the beta RF switch **72** is configured to forward the second RF output signal SRFO to provide the first beta RF transmit signal FBTX via the first beta harmonic filter **74**. In another communications mode, such as a second beta half-duplex transmit mode, a second beta saturated mode, or a second beta non-linear mode, the beta RF switch **72** is configured to forward the second RF output signal SRFO to provide the second beta RF transmit signal SBTX via the second beta harmonic filter **78**. In an alternate communications mode, such as a beta full-duplex mode or a beta linear mode, the beta RF switch **72** is configured to forward the second RF output signal SRFO to provide any of a third beta RF transmit signal TBTX through the Q^{TH} beta RF transmit signal QBTX. When a specific RF band is selected, the beta RF switch **72** may be configured to provide a corresponding selected one of the third beta RF transmit signal TBTX through the Q^{TH} beta RF transmit signal QBTX. The first alpha harmonic filter **70** or the second alpha harmonic filter **76** may be used to filter out harmonics of an RF carrier in the first RF output signal FRFO. The first beta harmonic filter **74** or the second beta harmonic filter **78** may be used to filter out harmonics of an RF carrier in the second RF output signal SRFO.

FIG. **10** shows the RF communications system **26** according to one embodiment of the RF communications system **26**. The RF communications system **26** shown in FIG. **10** is similar to the RF communications system **26** shown in FIG. **4**, except the RF communications system **26** illustrated in FIG. **10** further includes a DC power supply **80** and the DC configuration control signal DCC is omitted. Additionally, details of the DC-DC converter **32** are shown according to one embodiment of the DC-DC converter **32**. The DC-DC converter **32** includes first power filtering circuitry **82**, a charge pump buck converter **84**, a buck converter **86**, second power filtering circuitry **88**, a first inductive element L**1**, and a second inductive element L**2**. The DC power supply **80** provides a DC power supply signal DCPS to the charge pump buck converter **84**, the buck converter **86**, and the second power filtering circuitry **88**. In one embodiment of the DC power supply **80**, the DC power supply **80** is a battery.

The second power filtering circuitry **88** is coupled to the RF PA circuitry **30** and to the DC power supply **80**. The charge pump buck converter **84** is coupled to the DC power supply **80**. The first inductive element L**1** is coupled between the charge pump buck converter **84** and the first power filtering circuitry **82**. The buck converter **86** is coupled to the DC power supply **80**. The second inductive element L**2** is coupled between the buck converter **86** and the first power filtering circuitry **82**. The first power filtering circuitry **82** is coupled to the RF PA circuitry **30**. One end of the first inductive element L**1** is coupled to one end of the second inductive element L**2** at the first power filtering circuitry **82**.

In one embodiment of the DC-DC converter **32**, the DC-DC converter **32** operates in one of multiple converter operating modes, which include a first converter operating mode, a second converter operating mode, and a third converter operating mode. In an alternate embodiment of the DC-DC converter **32**, the DC-DC converter **32** operates in one of the first converter operating mode and the second converter operating mode. In the first converter operating mode, the charge pump buck converter **84** is active, such that the envelope power supply signal EPS is based on the DC power supply signal DCPS via the charge pump buck converter **84**, and the first inductive element L**1**. In the first converter operating mode, the buck converter **86** is inactive and does not contribute to the envelope power supply signal EPS. In the second converter operating mode, the buck converter **86** is active, such that the envelope power supply signal EPS is based on the DC power supply signal DCPS via the buck converter **86** and the second inductive element L**2**. In the second converter operating mode, the charge pump buck converter **84** is inactive, such that the charge pump buck converter **84** does not contribute to the envelope power supply signal EPS. In the third converter operating mode, the charge pump buck converter **84** and the buck converter **86** are active, such that either the charge pump buck converter **84**; the buck converter **86**; or both may contribute to the envelope power supply signal EPS. As such, in the third converter operating mode, the envelope power supply signal EPS is based on the DC power supply signal DCPS either via the charge pump buck converter **84**, and the first inductive element L**1**; via the buck converter **86** and the second inductive element L**2**; or both.

The second power filtering circuitry **88** filters the DC power supply signal DCPS to provide the bias power supply signal BPS. The second power filtering circuitry **88** may function as a lowpass filter by removing ripple, noise, and the like from the DC power supply signal DCPS to provide the bias power supply signal BPS. As such, in one embodiment of the DC-DC converter **32**, the bias power supply signal BPS is based on the DC power supply signal DCPS.

In the first converter operating mode or the third converter operating mode, the charge pump buck converter **84** may receive, charge pump, and buck convert the DC power supply signal DCPS to provide a first buck output signal FBO to the first inductive element L**1**. As such, in one embodiment of the charge pump buck converter **84**, the first buck output signal FBO is based on the DC power supply signal DCPS. Further, the first inductive element L**1** may function as a first energy transfer element of the charge pump buck converter **84** to transfer energy via the first buck output signal FBO to the first power filtering circuitry **82**. In the first converter operating mode or the third converter operating mode, the first inductive element L**1** and the first power filtering circuitry **82** may receive and filter the first buck output signal FBO to provide the envelope power supply signal EPS. The charge pump buck converter **84** may regulate the envelope power supply signal EPS by controlling the first buck output signal FBO based on a setpoint of the envelope power supply signal EPS provided by the envelope control signal ECS.

In the second converter operating mode or the third converter operating mode, the buck converter **86** may receive and buck convert the DC power supply signal DCPS to provide a second buck output signal SBO to the second inductive element L**2**. As such, in one embodiment of the buck converter **86**, the second buck output signal SBO is based on the DC power supply signal DCPS. Further, the second inductive element L**2** may function as a second energy transfer element of the buck converter **86** to transfer energy via the first power filtering circuitry **82** to the first power filtering circuitry **82**. In the second converter operating mode or the third converter operating mode, the second inductive element L**2** and the first power filtering circuitry **82** may receive and filter the second buck output signal SBO to provide the envelope power supply signal EPS. The buck converter **86** may regulate the envelope power supply signal EPS by controlling the second buck output signal SBO based on a setpoint of the envelope power supply signal EPS provided by the envelope control signal ECS.

In one embodiment of the charge pump buck converter **84**, the charge pump buck converter **84** operates in one of multiple pump buck operating modes. During a pump buck pump-up operating mode of the charge pump buck converter **84**, the charge pump buck converter **84** pumps-up the DC power supply signal DCPS to provide an internal signal (not shown), such that a voltage of the internal signal is greater than a voltage of the DC power supply signal DCPS. In an alternate embodiment of the charge pump buck converter **84**, during the pump buck pump-up operating mode, a voltage of the envelope power supply signal EPS is greater than the voltage of the DC power supply signal DCPS. During a pump buck pump-down operating mode of the charge pump buck converter **84**, the charge pump buck converter **84** pumps-down the DC power supply signal DCPS to provide the internal signal, such that a voltage of the internal signal is less than a voltage of the DC power supply signal DCPS. In an alternate embodiment of the charge pump buck converter **84**, during the pump buck pump-down operating mode, the voltage of the envelope power supply signal EPS is less than the voltage of the DC power supply signal DCPS. During a pump buck pump-even operating mode of the charge pump buck converter **84**, the charge pump buck converter **84** pumps the DC power supply signal DCPS to the internal signal, such that a voltage of the internal signal is about equal to a voltage of the DC power supply signal DCPS. One embodiment of the DC-DC converter **32** includes a pump buck bypass operating mode of the charge pump buck converter **84**, such that during the pump buck bypass operating mode, the charge pump buck converter **84** by-passes charge pump circuitry (not shown) using by-pass circuitry (not shown) to forward the DC power supply signal DCPS to provide the internal signal, such that a voltage of the internal is about equal to a voltage of the DC power supply signal DCPS.

In one embodiment of the charge pump buck converter **84**, the pump buck operating modes include the pump buck pump-up operating mode, the pump buck pump-down operating mode, the pump buck pump-even operating mode, and the pump buck bypass operating mode. In an alternate embodiment of the charge pump buck converter **84**, the pump buck pump-even operating mode is omitted. In an additional embodiment of the charge pump buck converter **84**, the pump buck bypass operating mode is omitted. In another embodiment of the charge pump buck converter **84**, the pump buck pump-down operating mode is omitted. In a further embodiment of the charge pump buck converter **84**, any or all of the pump buck pump-up operating mode, the pump buck pump-down operating mode, the pump buck pump-even operating mode, and the pump buck bypass operating mode are omitted. In a supplemental embodiment of the charge pump buck converter **84**, the charge pump buck converter **84** operates in only the pump buck pump-up operating mode. In an additional embodiment of the charge pump buck converter **84**, the charge pump buck converter **84** operates in one of the pump buck pump-up operating mode and at least one other pump buck operating mode of the charge pump buck converter **84**. The at least one other pump buck operating mode of the charge pump buck converter **84** may include any or all of the pump buck pump-up operating mode, the pump buck pump-down operating mode, the pump buck pump-even operating mode, and the pump buck bypass operating mode.

FIG. **11** shows the RF communications system **26** according to an alternate embodiment of the RF communications system **26**. The RF communications system **26** illustrated in FIG. **11** is similar to the RF communications system **26** illustrated in FIG. **10**, except in the RF communications system **26** illustrated in FIG. **11**, the DC-DC converter **32** further includes DC-DC control circuitry **90** and a charge pump **92**, and omits the second inductive element L2. Instead of the second power filtering circuitry **88** being coupled to the DC power supply **80** as shown in FIG. **10**, the charge pump **92** is coupled to the DC power supply **80**, such that the charge pump **92** is coupled between the DC power supply **80** and the second power filtering circuitry **88**. Additionally, the RF modulation and control circuitry **28** provides the DC configuration control signal DCC and the envelope control signal ECS to the DC-DC control circuitry **90**.

The DC-DC control circuitry **90** provides a charge pump buck control signal CPBS to the charge pump buck converter **84**, provides a buck control signal BCS to the buck converter **86**, and provides a charge pump control signal CPS to the charge pump **92**. The charge pump buck control signal CPBS, the buck control signal BCS, or both may indicate which converter operating mode is selected. Further, the charge pump buck control signal CPBS, the buck control signal BCS, or both may provide the setpoint of the envelope power supply signal EPS as provided by the envelope control signal ECS. The charge pump buck control signal CPBS may indicate which pump buck operating mode is selected.

In one embodiment of the DC-DC converter **32**, selection of the converter operating mode is made by the DC-DC control circuitry **90**. In an alternate embodiment of the DC-DC converter **32**, selection of the converter operating mode is made by the RF modulation and control circuitry **28** and may be communicated to the DC-DC converter **32** via the DC configuration control signal DCC. In an additional embodiment of the DC-DC converter **32**, selection of the converter operating mode is made by the control circuitry **42** (FIG. **5**) and may be communicated to the DC-DC converter **32** via the DC configuration control signal DCC. In general, selection of the converter operating mode is made by control circuitry, which may be any of the DC-DC control circuitry **90**, the RF modulation and control circuitry **28**, and the control circuitry **42** (FIG. **5**).

In one embodiment of the DC-DC converter **32**, selection of the pump buck operating mode is made by the DC-DC control circuitry **90**. In an alternate embodiment of the DC-DC converter **32**, selection of the pump buck operating mode is made by the RF modulation and control circuitry **28** and communicated to the DC-DC converter **32** via the DC configuration control signal DCC. In an additional embodiment of the DC-DC converter **32**, selection of the pump buck operating mode is made by the control circuitry **42** (FIG. **5**) and communicated to the DC-DC converter **32** via the DC configuration control signal DCC. In general, selection of the pump buck operating mode is made by control circuitry, which may be any of the DC-DC control circuitry **90**, the RF modulation and control circuitry **28**, and the control circuitry **42** (FIG. **5**). As such, the control circuitry may select one of the pump buck pump-up operating mode and at least one other pump buck operating mode of the charge pump buck converter **84**. The at least one other pump buck operating mode of the charge pump buck converter **84** may include any or all of the pump buck pump-down operating mode, the pump buck pump-even operating mode, and the pump buck bypass operating mode.

The charge pump **92** may operate in one of multiple bias supply pump operating modes. During a bias supply pump-up operating mode of the charge pump **92**, the charge pump **92** receives and pumps-up the DC power supply signal DCPS to provide the bias power supply signal BPS, such that a voltage of the bias power supply signal BPS is greater than a voltage of the DC power supply signal DCPS. During a bias supply pump-down operating mode of the charge pump **92**, the charge pump **92** pumps-down the DC power supply signal DCPS to provide the bias power supply signal BPS, such that a voltage of the bias power supply signal BPS is less than a voltage of the DC power supply signal DCPS. During a bias supply pump-even operating mode of the charge pump **92**, the charge pump **92** pumps the DC power supply signal DCPS to provide the bias power supply signal BPS, such that a voltage of the bias power supply signal BPS is about equal to a voltage of the DC power supply signal DCPS. One embodiment of the DC-DC converter **32** includes a bias supply bypass operating mode of the charge pump **92**, such that during the bias supply bypass operating mode, the charge pump **92** by-passes charge pump circuitry (not shown) using by-pass circuitry (not shown) to forward the DC power supply signal DCPS to provide the bias power supply signal BPS, such that a voltage of the bias power supply signal BPS is about equal to a voltage of the DC power supply signal DCPS. The charge pump control signal CPS may indicate which bias supply pump operating mode is selected.

In one embodiment of the charge pump **92**, the bias supply pump operating modes include the bias supply pump-up operating mode, the bias supply pump-down operating mode, the bias supply pump-even operating mode, and the bias supply bypass operating mode. In an alternate embodiment of the charge pump **92**, the bias supply pump-even operating mode is omitted. In an additional embodiment of the charge pump **92**, the bias supply bypass operating mode is omitted. In another embodiment of the charge pump **92**, the bias supply pump-down operating mode is omitted. In a further embodiment of the charge pump **92**, any or all of the bias supply pump-up operating mode, the bias supply pump-down operating mode, the bias supply pump-even operating mode, and the bias supply bypass operating mode are omitted. In a supplemental embodiment of the charge pump **92**, the charge pump **92** operates in only the bias supply pump-up operating mode. In an additional embodiment of the charge pump **92**, the charge pump **92** operates in the bias supply pump-up operating mode and at least one other operating mode of the charge pump **92**, which may include any or all of the bias supply pump-down operating mode, the bias supply pump-even operating mode, and the bias supply bypass operating mode.

In one embodiment of the DC-DC converter **32**, selection of the bias supply pump operating mode is made by the

DC-DC control circuitry **90**. In an alternate embodiment of the DC-DC converter **32**, selection of the bias supply pump operating mode is made by the RF modulation and control circuitry **28** and communicated to the DC-DC converter **32** via the DC configuration control signal DCC. In an additional embodiment of the DC-DC converter **32**, selection of the bias supply pump operating mode is made by the control circuitry **42** (FIG. **5**) and communicated to the DC-DC converter **32** via the DC configuration control signal DCC. In general, selection of the bias supply pump operating mode is made by control circuitry, which may be any of the DC-DC control circuitry **90**, the RF modulation and control circuitry **28**, and the control circuitry **42** (FIG. **5**). As such, the control circuitry may select one of the bias supply pump-up operating mode and at least one other bias supply operating mode. The at least one other bias supply operating mode may include any or all of the bias supply pump-down operating mode, the bias supply pump-even operating mode, and the bias supply bypass operating mode.

The second power filtering circuitry **88** filters the bias power supply signal BPS. The second power filtering circuitry **88** may function as a lowpass filter by removing ripple, noise, and the like to provide the bias power supply signal BPS. As such, in one embodiment of the DC-DC converter **32**, the bias power supply signal BPS is based on the DC power supply signal DCPS.

Regarding omission of the second inductive element L**2**, instead of the second inductive element L**2** coupled between the buck converter **86** and the first power filtering circuitry **82** as shown in FIG. **10**, one end of the first inductive element L**1** is coupled to both the charge pump buck converter **84** and the buck converter **86**. As such, in the second converter operating mode or the third converter operating mode, the buck converter **86** may receive and buck convert the DC power supply signal DCPS to provide the second buck output signal SBO to the first inductive element L**1**. As such, in one embodiment of the charge pump buck converter **84**, the second buck output signal SBO is based on the DC power supply signal DCPS. Further, the first inductive element L**1** may function as a first energy transfer element of the buck converter **86** to transfer energy via the second buck output signal SBO to the first power filtering circuitry **82**. In the first converter operating mode, the second converter operating mode, or the third converter operating mode, the first inductive element L**1** and the first power filtering circuitry **82** receive and filter the first buck output signal FBO, the second buck output signal SBO, or both to provide the envelope power supply signal EPS.

FIG. **12** shows details of the DC-DC converter **32** illustrated in FIG. **11** according to an alternate embodiment of the DC-DC converter **32**. The DC-DC converter **32** illustrated in FIG. **12** is similar to the DC-DC converter **32** illustrated in FIG. **10**, except the DC-DC converter **32** illustrated in FIG. **12** shows details of the first power filtering circuitry **82** and the second power filtering circuitry **88**. Further, the DC-DC converter **32** illustrated in FIG. **12** includes the DC-DC control circuitry **90** and the charge pump **92** as shown in FIG. **11**.

The first power filtering circuitry **82** includes a first capacitive element C**1**, a second capacitive element C**2**, and a third inductive element L**3**. The first capacitive element C**1** is coupled between one end of the third inductive element L**3** and a ground. The second capacitive element C**2** is coupled between an opposite end of the third inductive element L**3** and ground. The one end of the third inductive element L**3** is coupled to one end of the first inductive element L**1**. Further, the one end of the third inductive element L**3** is coupled to one end of the second inductive element L**2**. In an additional embodiment of the DC-DC converter **32**, the second inductive element L**2** is omitted. The opposite end of the third inductive element L**3** is coupled to the RF PA circuitry **30**. As such, the opposite end of the third inductive element L**3** and one end of the second capacitive element C**2** provide the envelope power supply signal EPS. In an alternate embodiment of the first power filtering circuitry **82**, the third inductive element L**3**, the second capacitive element C**2**, or both are omitted.

FIG. **13** shows details of the RF PA circuitry **30** illustrated in FIG. **5** according to one embodiment of the RF PA circuitry **30**. The RF PA circuitry **30** illustrated in FIG. **13** is similar to the RF PA circuitry **30** illustrated in FIG. **5**, except the RF PA circuitry **30** illustrated in FIG. **13** further includes PA control circuitry **94**, PA bias circuitry **96**, and switch driver circuitry **98**. The PA bias circuitry **96** is coupled between the PA control circuitry **94** and the RF PAs **50**, **54**. The switch driver circuitry **98** is coupled between the PA control circuitry **94** and the switching circuitry **52**, **56**. The PA control circuitry **94** receives the PA configuration control signal PCC, provides a bias configuration control signal BCC to the PA bias circuitry **96** based on the PA configuration control signal PCC, and provides a switch configuration control signal SCC to the switch driver circuitry **98** based on the PA configuration control signal PCC. The switch driver circuitry **98** provides any needed drive signals to configure the alpha switching circuitry **52** and the beta switching circuitry **56**.

The PA bias circuitry **96** receives the bias power supply signal BPS and the bias configuration control signal BCC. The PA bias circuitry **96** provides a first driver bias signal FDB and a first final bias signal FFB to the first RF PA **50** based on the bias power supply signal BPS and the bias configuration control signal BCC. The PA bias circuitry **96** provides a second driver bias signal SDB and a second final bias signal SFB to the second RF PA **54** based on the bias power supply signal BPS and the bias configuration control signal BCC. The bias power supply signal BPS provides the power necessary to generate the bias signals FDB, FFB, SDB, SFB. A selected magnitude of each of the bias signals FDB, FFB, SDB, SFB is provided by the PA bias circuitry **96**. In one embodiment of the RF PA circuitry **30**, the PA control circuitry **94** selects the magnitude of any or all of the bias signals FDB, FFB, SDB, SFB and communicates the magnitude selections to the PA bias circuitry **96** via the bias configuration control signal BCC. The magnitude selections by the PA control circuitry **94** may be based on the PA configuration control signal PCC. In an alternate embodiment of the RF PA circuitry **30**, the control circuitry **42** (FIG. **5**) selects the magnitude of any or all of the bias signals FDB, FFB, SDB, SFB and communicates the magnitude selections to the PA bias circuitry **96** via the PA control circuitry **94**.

In one embodiment of the RF PA circuitry **30**, the RF PA circuitry **30** operates in one of a first PA operating mode and a second PA operating mode. During the first PA operating mode, the first transmit path **46** is enabled and the second transmit path **48** is disabled. During the second PA operating mode, the first transmit path **46** is disabled and the second transmit path **48** is enabled. In one embodiment of the first RF PA **50** and the second RF PA **54**, during the second PA operating mode, the first RF PA **50** is disabled, and during the first PA operating mode, the second RF PA **54** is disabled. In one embodiment of the alpha switching circuitry **52** and the beta switching circuitry **56**, during the second PA operating mode, the alpha switching circuitry **52** is disabled, and during the first PA operating mode, the beta switching circuitry **56** is disabled.

In one embodiment of the first RF PA **50**, during the second PA operating mode, the first RF PA **50** is disabled via the first driver bias signal FDB. In an alternate embodiment of the first RF PA **50**, during the second PA operating mode, the first RF PA **50** is disabled via the first final bias signal FFB. In an additional embodiment of the first RF PA **50**, during the second PA operating mode, the first RF PA **50** is disabled via both the first driver bias signal FDB and the first final bias signal FFB. In one embodiment of the second RF PA **54**, during the first PA operating mode, the second RF PA **54** is disabled via the second driver bias signal SDB. In an alternate embodiment of the second RF PA **54**, during the first PA operating mode, the second RF PA **54** is disabled via the second final bias signal SFB. In an additional embodiment of the second RF PA **54**, during the first PA operating mode, the second RF PA **54** is disabled via both the second driver bias signal SDB and the second final bias signal SFB.

In one embodiment of the RF PA circuitry **30**, the PA control circuitry **94** selects the one of the first PA operating mode and the second PA operating mode. As such, the PA control circuitry **94** may control any or all of the bias signals FDB, FFB, SDB, SFB via the bias configuration control signal BCC based on the PA operating mode selection. Further, the PA control circuitry **94** may control the switching circuitry **52**, **56** via the switch configuration control signal SCC based on the PA operating mode selection. The PA operating mode selection may be based on the PA configuration control signal PCC. In an alternate embodiment of the RF PA circuitry **30**, the control circuitry **42** (FIG. **5**) selects the one of the first PA operating mode and the second PA operating mode. As such, the control circuitry **42** (FIG. **5**) may indicate the operating mode selection to the PA control circuitry **94** via the PA configuration control signal PCC. In an additional embodiment of the RF PA circuitry **30**, the RF modulation and control circuitry **28** (FIG. **5**) selects the one of the first PA operating mode and the second PA operating mode. As such, the RF modulation and control circuitry **28** (FIG. **5**) may indicate the operating mode selection to the PA control circuitry **94** via the PA configuration control signal PCC. In general, selection of the PA operating mode is made by control circuitry, which may be any of the PA control circuitry **94**, the RF modulation and control circuitry **28** (FIG. **5**), and the control circuitry **42** (FIG. **5**).

FIG. **14** shows details of the RF PA circuitry **30** illustrated in FIG. **6** according to an alternate embodiment of the RF PA circuitry **30**. The RF PA circuitry **30** illustrated in FIG. **14** is similar to the RF PA circuitry **30** illustrated in FIG. **13**, except the RF PA circuitry **30** illustrated in FIG. **14** further includes the PA-DCI **60**, which is coupled to the PA control circuitry **94** and to the digital communications bus **66**. As such, the control circuitry **42** (FIG. **6**) may provide the PA configuration control signal PCC via the control circuitry DCI **58** (FIG. **6**) to the PA control circuitry **94** via the PA-DCI **60**.

FIG. **15** shows details of the first RF PA **50** and the second RF PA **54** illustrated in FIG. **13** according one embodiment of the first RF PA **50** and the second RF PA **54**. The first RF PA **50** includes a first non-quadrature PA path **100** and a first quadrature PA path **102**. The second RF PA **54** includes a second non-quadrature PA path **104** and a second quadrature PA path **106**. In one embodiment of the first RF PA **50**, the first quadrature PA path **102** is coupled between the first non-quadrature PA path **100** and the antenna port AP (FIG. **6**), which is coupled to the antenna **18** (FIG. **6**). In an alternate embodiment of the first RF PA **50**, the first non-quadrature PA path **100** is omitted, such that the first quadrature PA path **102** is coupled to the antenna port AP (FIG. **6**). The first quadrature PA path **102** may be coupled to the antenna port AP (FIG. **6**) via the alpha switching circuitry **52** (FIG. **6**) and the front-end aggregation circuitry **36** (FIG. **6**). The first non-quadrature PA path **100** may include any number of non-quadrature gain stages. The first quadrature PA path **102** may include any number of quadrature gain stages. In one embodiment of the second RF PA **54**, the second quadrature PA path **106** is coupled between the second non-quadrature PA path **104** and the antenna port AP (FIG. **6**). In an alternate embodiment of the second RF PA **54**, the second non-quadrature PA path **104** is omitted, such that the second quadrature PA path **106** is coupled to the antenna port AP (FIG. **6**). The second quadrature PA path **106** may be coupled to the antenna port AP (FIG. **6**) via the beta switching circuitry **56** (FIG. **6**) and the front-end aggregation circuitry **36** (FIG. **6**). The second non-quadrature PA path **104** may include any number of non-quadrature gain stages. The second quadrature PA path **106** may include any number of quadrature gain stages.

In one embodiment of the RF communications system **26**, the control circuitry **42** (FIG. **5**) selects one of multiple communications modes, which include a first PA operating mode and a second PA operating mode. During the first PA operating mode, the first PA paths **100**, **102** receive the envelope power supply signal EPS, which provides power for amplification. During the second PA operating mode, the second PA paths **104**, **106** receive the envelope power supply signal EPS, which provides power for amplification. During the first PA operating mode, the first non-quadrature PA path **100** receives the first driver bias signal FDB, which provides biasing to the first non-quadrature PA path **100**, and the first quadrature PA path **102** receives the first final bias signal FFB, which provides biasing to the first quadrature PA path **102**. During the second PA operating mode, the second non-quadrature PA path **104** receives the second driver bias signal SDB, which provides biasing to the second non-quadrature PA path **104**, and the second quadrature PA path **106** receives the second final bias signal SFB, which provides biasing to the second quadrature PA path **106**.

The first non-quadrature PA path **100** has a first single-ended output FSO and the first quadrature PA path **102** has a first single-ended input FSI. The first single-ended output FSO may be coupled to the first single-ended input FSI. In one embodiment of the first RF PA **50**, the first single-ended output FSO is directly coupled to the first single-ended input FSI. The second non-quadrature PA path **104** has a second single-ended output SSO and the second quadrature PA path **106** has a second single-ended input SSI. The second single-ended output SSO may be coupled to the second single-ended input SSI. In one embodiment of the second RF PA **54**, the second single-ended output SSO is directly coupled to the second single-ended input SSI.

During the first PA operating mode, the first RF PA **50** receives and amplifies the first RF input signal FRFI to provide the first RF output signal FRFO, and the second RF PA **54** is disabled. During the second PA operating mode, the second RF PA **54** receives and amplifies the second RF input signal SRFI to provide the second RF output signal SRFO, and the first RF PA **50** is disabled. In one embodiment of the RF communications system **26**, the first RF input signal FRFI is a highband RF input signal and the second RF input signal SRFI is a lowband RF input signal. In one exemplary embodiment of the RF communications system **26**, a difference between a frequency of the highband RF input signal and a frequency of the lowband RF input signal is greater than about 500 megahertz, such that the frequency of the highband RF input signal is greater than the frequency of the lowband RF input signal. In an alternate exemplary embodiment of the RF communications system **26**, a ratio of a frequency of the highband RF input signal divided by a frequency of the lowband RF input signal is greater than about 1.5.

In one embodiment of the first RF PA **50**, during the first PA operating mode, the first non-quadrature PA path **100** receives and amplifies the first RF input signal FRFI to provide a first RF feeder output signal FFO to the first quadrature PA path **102** via the first single-ended output FSO. Further, during the first PA operating mode, the first quadrature PA path **102** receives and amplifies the first RF feeder output signal FFO via the first single-ended input FSI to provide the first RF output signal FRFO. In one embodiment of the second RF PA **54**, during the second PA operating mode, the second non-quadrature PA path **104** receives and amplifies the second RF input signal SRFI to provide a second RF feeder output signal SFO to the second quadrature PA path **106** via the second single-ended output SSO. Further, during the second PA operating mode, the second quadrature PA path **106** receives and amplifies the second RF feeder output signal SFO via the second single-ended input SSI to provide the second RF output signal SRFO.

Quadrature PA Architecture

A summary of quadrature PA architecture is presented, followed by a detailed description of the quadrature PA architecture according to one embodiment of the present disclosure. One embodiment of the RF communications system **26** (FIG. **6**) relates to a quadrature RF PA architecture that utilizes a single-ended interface to couple a non-quadrature PA path to a quadrature PA path, which may be coupled to the antenna port (FIG. **6**). The quadrature nature of the quadrature PA path may provide tolerance for changes in antenna loading conditions. An RF splitter in the quadrature PA path may present a relatively stable input impedance, which may be predominantly resistive, to the non-quadrature PA path over a wide frequency range, thereby substantially isolating the non-quadrature PA path from changes in the antenna loading conditions. Further, the input impedance may substantially establish a load line slope of a feeder PA stage in the non-quadrature PA path, thereby simplifying the quadrature RF PA architecture. One embodiment of the quadrature RF PA architecture uses two separate PA paths, either of which may incorporate a combined non-quadrature and quadrature PA architecture.

Due to the relatively stable input impedance, RF power measurements taken at the single-ended interface may provide high directivity and accuracy. Further, by combining the non-quadrature PA path and the quadrature PA path, gain stages may be eliminated and circuit topology may be simplified. In one embodiment of the RF splitter, the RF splitter is a quadrature hybrid coupler, which may include a pair of tightly coupled inductors. The input impedance may be based on inductances of the pair of tightly coupled inductors and parasitic capacitance between the inductors. As such, construction of the pair of tightly coupled inductors may be varied to select a specific parasitic capacitance to provide a specific input impedance. Further, the RF splitter may be integrated onto one semiconductor die with amplifying elements of the non-quadrature PA path, with amplifying elements of the quadrature PA path, or both, thereby reducing size and cost. Additionally, the quadrature PA path may have only a single quadrature amplifier stage to further simplify the design. In certain embodiments, using only the single quadrature amplifier stage provides adequate tolerance for changes in antenna loading conditions.

FIG. **16** shows details of the first non-quadrature PA path **100** and the second non-quadrature PA path **104** illustrated in FIG. **15** according to one embodiment of the first non-quadrature PA path **100** and the second non-quadrature PA path **104**.

The first non-quadrature PA path **100** includes a first input PA impedance matching circuit **108**, a first input PA stage **110**, a first feeder PA impedance matching circuit **112**, and a first feeder PA stage **114**, which provides the first single-ended output FSO. The first input PA stage **110** is coupled between the first input PA impedance matching circuit **108** and the first feeder PA impedance matching circuit **112**. The first feeder PA stage **114** is coupled between the first feeder PA impedance matching circuit **112** and the first quadrature PA path **102**. The first input PA impedance matching circuit **108** may provide at least an approximate impedance match between the RF modulation circuitry **44** (FIG. **5**) and the first input PA stage **110**. The first feeder PA impedance matching circuit **112** may provide at least an approximate impedance match between the first input PA stage **110** and the first feeder PA stage **114**. In alternate embodiments of the first non-quadrature PA path **100**, any or all of the first input PA impedance matching circuit **108**, the first input PA stage **110**, and the first feeder PA impedance matching circuit **112**, may be omitted.

During the first PA operating mode, the first input PA impedance matching circuit **108** receives and forwards the first RF input signal FRFI to the first input PA stage **110**. During the first PA operating mode, the first input PA stage **110** receives and amplifies the forwarded first RF input signal FRFI to provide a first RF feeder input signal FFI to the first feeder PA stage **114** via the first feeder PA impedance matching circuit **112**. During the first PA operating mode, the first feeder PA stage **114** receives and amplifies the first RF feeder input signal FFI to provide the first RF feeder output signal FFO via the first single-ended output FSO. The first feeder PA stage **114** may have a first output load line having a first load line slope. During the first PA operating mode, the envelope power supply signal EPS provides power for amplification to the first input PA stage **110** and to the first feeder PA stage **114**. During the first PA operating mode, the first driver bias signal FDB provides biasing to the first input PA stage **110** and the first feeder PA stage **114**.

The second non-quadrature PA path **104** includes a second input PA impedance matching circuit **116**, a second input PA stage **118**, a second feeder PA impedance matching circuit **120**, and a second feeder PA stage **122**, which provides the second single-ended output SSO. The second input PA stage **118** is coupled between the second input PA impedance matching circuit **116** and the second feeder PA impedance matching circuit **120**. The second feeder PA stage **122** is coupled between the second feeder PA impedance matching circuit **120** and the second quadrature PA path **106**. The second input PA impedance matching circuit **116** may provide at least an approximate impedance match between the RF modulation circuitry **44** (FIG. **5**) and the second input PA stage **118**. The second feeder PA impedance matching circuit **120** may provide at least an approximate impedance match between the second input PA stage **118** and the second feeder PA stage **122**. In alternate embodiments of the second non-quadrature PA path **104**, any or all of the second input PA impedance matching circuit **116**, the second input PA stage **118**, and the second feeder PA impedance matching circuit **120**, may be omitted.

During the second PA operating mode, the second input PA impedance matching circuit **116** receives and forwards the second RF input signal SRFI to the second input PA stage **118**. During the second PA operating mode, the second input PA stage **118** receives and amplifies the forwarded second RF input signal SRFI to provide a second RF feeder input signal SFI to the second feeder PA stage **122** via the second feeder PA impedance matching circuit **120**. During the second PA operating mode, the second feeder PA stage **122** receives and amplifies the second RF feeder input signal SFI to provide the second RF feeder output signal SFO via the second single-ended output SSO. The second feeder PA stage **122** may have a second output load line having a second load line slope. During the second PA operating mode, the envelope power supply signal EPS provides power for amplification to the second input PA stage **118** and to the second feeder PA stage **122**. During the second PA operating mode, the second driver bias signal SDB provides biasing to the second input PA stage **118** and the second feeder PA stage **122**.

FIG. **17** shows details of the first quadrature PA path **102** and the second quadrature PA path **106** illustrated in FIG. **15** according to one embodiment of the first quadrature PA path **102** and the second quadrature PA path **106**. The first quadrature PA path **102** includes a first quadrature RF splitter **124**, a first in-phase amplification path **126**, a first quadrature-phase amplification path **128**, and a first quadrature RF combiner **130**. The first quadrature RF splitter **124** has a first single-ended input FSI, a first in-phase output FIO, and a first quadrature-phase output FQO. The first quadrature RF combiner **130** has a first in-phase input FII, a first quadrature-phase input FQI, and a first quadrature combiner output FCO. The first single-ended output FSO is coupled to the first single-ended input FSI. In one embodiment of the first quadrature PA path **102**, the first single-ended output FSO is directly coupled to the first single-ended input FSI. The first in-phase amplification path **126** is coupled between the first in-phase output FIO and the first in-phase input FII. The first quadrature-phase amplification path **128** is coupled between the first quadrature-phase output FQO and the first quadrature-phase input FQI. The first quadrature combiner output FCO is coupled to the antenna port AP (FIG. **6**) via the alpha switching circuitry **52** (FIG. **6**) and the front-end aggregation circuitry **36** (FIG. **6**).

During the first PA operating mode, the first quadrature RF splitter **124** receives the first RF feeder output signal FFO via the first single-ended input FSI. Further, during the first PA operating mode, the first quadrature RF splitter **124** splits and phase-shifts the first RF feeder output signal FFO into a first in-phase RF input signal FIN and a first quadrature-phase RF input signal FQN, such that the first quadrature-phase RF input signal FQN is nominally phase-shifted from the first in-phase RF input signal FIN by about 90 degrees. The first quadrature RF splitter **124** has a first input impedance presented at the first single-ended input FSI. In one embodiment of the first quadrature RF splitter **124**, the first input impedance establishes the first load line slope. During the first PA operating mode, the first in-phase amplification path **126** receives and amplifies the first in-phase RF input signal FIN to provide the first in-phase RF output signal FIT. The first quadrature-phase amplification path **128** receives and amplifies the first quadrature-phase RF input signal FQN to provide the first quadrature-phase RF output signal FQT.

During the first PA operating mode, the first quadrature RF combiner **130** receives the first in-phase RF output signal FIT via the first in-phase input FII, and receives the first quadrature-phase RF output signal FQT via the first quadrature-phase input FQI. Further, the first quadrature RF combiner **130** phase-shifts and combines the first in-phase RF output signal FIT and the first quadrature-phase RF output signal FQT to provide the first RF output signal FRFO via the first quadrature combiner output FCO, such that the phase-shifted first in-phase RF output signal FIT and first quadrature-phase RF output signal FQT are about phase-aligned with one another before combining. During the first PA operating mode, the envelope power supply signal EPS provides power for amplification to the first in-phase amplification path **126** and the first quadrature-phase amplification path **128**. During the first PA operating mode, the first final bias signal FFB provides biasing to the first in-phase amplification path **126** and the first quadrature-phase amplification path **128**.

The second quadrature PA path **106** includes a second quadrature RF splitter **132**, a second in-phase amplification path **134**, a second quadrature-phase amplification path **136**, and a second quadrature RF combiner **138**. The second quadrature RF splitter **132** has a second single-ended input SSI, a second in-phase output SIO, and a second quadrature-phase output SQO. The second quadrature RF combiner **138** has a second in-phase input SII, a second quadrature-phase input SQI, and a second quadrature combiner output SCO. The second single-ended output SSO is coupled to the second single-ended input SSI. In one embodiment of the second quadrature PA path **106**, the second single-ended output SSO is directly coupled to the second single-ended input SSI. The second in-phase amplification path **134** is coupled between the second in-phase output SIO and the second in-phase input SII. The second quadrature-phase amplification path **136** is coupled between the second quadrature-phase output SQO and the second quadrature-phase input SQI. The second quadrature combiner output SCO is coupled to the antenna port AP (FIG. **6**) via the alpha switching circuitry **52** (FIG. **6**) and the front-end aggregation circuitry **36** (FIG. **6**).

During the second PA operating mode, the second quadrature RF splitter **132** receives the second RF feeder output signal SFO via the second single-ended input SSI. Further, during the second PA operating mode, the second quadrature RF splitter **132** splits and phase-shifts the second RF feeder output signal SFO into a second in-phase RF input signal SIN and a second quadrature-phase RF input signal SQN, such that the second quadrature-phase RF input signal SQN is nominally phase-shifted from the second in-phase RF input signal SIN by about 90 degrees. The second quadrature RF splitter **132** has a second input impedance presented at the second single-ended input SSI. In one embodiment of the second quadrature RF splitter **132**, the second input impedance establishes the second load line slope. During the second PA operating mode, the second in-phase amplification path **134** receives and amplifies the second in-phase RF input signal SIN to provide the second in-phase RF output signal SIT. The second quadrature-phase amplification path **136** receives and amplifies the second quadrature-phase RF input signal SQN to provide the second quadrature-phase RF output signal SQT.

During the second PA operating mode, the second quadrature RF combiner **138** receives the second in-phase RF output signal SIT via the second in-phase input SII, and receives the second quadrature-phase RF output signal SQT via the second quadrature-phase input SQI. Further, the second quadrature RF combiner **138** phase-shifts and combines the second in-phase RF output signal SIT and the second quadrature-phase RF output signal SQT to provide the second RF output signal SRFO via the second quadrature combiner output SCO, such that the phase-shifted second in-phase RF output signal SIT and second quadrature-phase RF output signal SQT are about phase-aligned with one another before combining. During the second PA operating mode, the envelope power supply signal EPS provides power for amplification to the second in-phase amplification path **134** and the second quadrature-phase amplification path **136**. During the second PA operating mode, the second final bias signal SFB provides biasing to the second in-phase amplification path **134** and the second quadrature-phase amplification path **136**.

In one embodiment of the RF PA circuitry **30** (FIG. **13**), the second transmit path **48** (FIG. **13**) is omitted. As such, the first feeder PA stage **114** (FIG. **16**) is a feeder PA stage and the first single-ended output FSO (FIG. **16**) is a single-ended output. The first RF feeder input signal FFI (FIG. **16**) is an RF feeder input signal and the first RF feeder output signal FFO (FIG. **16**) is an RF feeder output signal. The feeder PA stage receives and amplifies the RF feeder input signal to provide the RF feeder output signal via the single-ended output. The feeder PA stage has an output load line having a load line slope. The first quadrature RF splitter **124** is a quadrature RF splitter and the first single-ended input FSI is a single-ended input. As such, the quadrature RF splitter has the single-ended input. In one embodiment of the first RF PA **50**, the single-ended output is directly coupled to the single-ended input.

In the embodiment in which the second transmit path **48** (FIG. **13**) is omitted, the first in-phase RF input signal FIN is an in-phase RF input signal and the first quadrature-phase RF input signal FQN is a quadrature-phase RF input signal. The quadrature RF splitter receives the RF feeder output signal via the single-ended input. Further, the quadrature RF splitter splits and phase-shifts the RF feeder output signal into the in-phase RF input signal and the quadrature-phase RF input signal, such that the quadrature-phase RF input signal is nominally phase-shifted from the in-phase RF input signal by about 90 degrees. The quadrature RF splitter has an input impedance presented at the single-ended input. The input impedance substantially establishes the load line slope. The first in-phase amplification path **126** is an in-phase amplification path and the first quadrature-phase amplification path **128** is a quadrature-phase amplification path. The first in-phase RF output signal FIT is an in-phase RF output signal and the first quadrature-phase RF output signal FQT is a quadrature-phase RF output signal. As such, the in-phase amplification path receives and amplifies the in-phase RF input signal to provide the in-phase RF output signal. The quadrature-phase amplification path receives and amplifies the quadrature-phase RF input signal to provide the quadrature-phase RF output signal.

In the embodiment in which the second transmit path **48** (FIG. **13**) is omitted, the first RF output signal FRFO is an RF output signal. As such, the quadrature RF combiner receives, phase-shifts, and combines the in-phase RF output signal and the quadrature-phase RF output signal to provide the RF output signal. In one embodiment of the quadrature RF splitter, the input impedance has resistance and reactance, such that the reactance is less than the resistance. In a first exemplary embodiment of the quadrature RF splitter, the resistance is greater than two times the reactance. In a second exemplary embodiment of the quadrature RF splitter, the resistance is greater than four times the reactance. In a third exemplary embodiment of the quadrature RF splitter, the resistance is greater than six times the reactance. In a fourth exemplary embodiment of the quadrature RF splitter, the resistance is greater than eight times the reactance. In a first exemplary embodiment of the quadrature RF splitter, the resistance is greater than ten times the reactance.

In alternate embodiments of the first quadrature PA path **102** and the second quadrature PA path **106**, any or all of the first quadrature RF splitter **124**, the first quadrature RF combiner **130**, the second quadrature RF splitter **132**, and the second quadrature RF combiner **138** may be any combination of quadrature RF couplers, quadrature hybrid RF couplers; Fisher couplers; lumped-element based RF couplers; transmission line based RF couplers; and combinations of phase-shifting circuitry and RF power couplers, such as phase-shifting circuitry and Wilkinson couplers; and the like. As such, any of the RF couplers listed above may be suitable to provide the first input impedance, the second input impedance, or both.

FIG. **18** shows details of the first in-phase amplification path **126**, the first quadrature-phase amplification path **128**, the second in-phase amplification path **134**, and the second quadrature-phase amplification path **136** illustrated in FIG. **17** according to one embodiment of the first in-phase amplification path **126**, the first quadrature-phase amplification path **128**, the second in-phase amplification path **134**, and the second quadrature-phase amplification path **136**. The first in-phase amplification path **126** includes a first in-phase driver PA impedance matching circuit **140**, a first in-phase driver PA stage **142**, a first in-phase final PA impedance matching circuit **144**, a first in-phase final PA stage **146**, and a first in-phase combiner impedance matching circuit **148**. The first in-phase driver PA impedance matching circuit **140** is coupled between the first in-phase output FIO and the first in-phase driver PA stage **142**. The first in-phase final PA impedance matching circuit **144** is coupled between the first in-phase driver PA stage **142** and the first in-phase final PA stage **146**. The first in-phase combiner impedance matching circuit **148** is coupled between the first in-phase final PA stage **146** and the first in-phase input FII.

The first in-phase driver PA impedance matching circuit **140** may provide at least an approximate impedance match between the first quadrature RF splitter **124** and the first in-phase driver PA stage **142**. The first in-phase final PA impedance matching circuit **144** may provide at least an approximate impedance match between the first in-phase driver PA stage **142** and the first in-phase final PA stage **146**. The first in-phase combiner impedance matching circuit **148** may provide at least an approximate impedance match between the first in-phase final PA stage **146** and the first quadrature RF combiner **130**.

During the first PA operating mode, the first in-phase driver PA impedance matching circuit **140** receives and forwards the first in-phase RF input signal FIN to the first in-phase driver PA stage **142**, which receives and amplifies the forwarded first in-phase RF input signal to provide an amplified first in-phase RF input signal to the first in-phase final PA stage **146** via the first in-phase final PA impedance matching circuit **144**. The first in-phase final PA stage **146** receives and amplifies the amplified first in-phase RF input signal to provide the first in-phase RF output signal FIT via the first in-phase combiner impedance matching circuit **148**. During the first PA operating mode, the envelope power supply signal EPS provides power for amplification to the first in-phase driver PA stage **142** and the first in-phase final PA stage **146**. During the first PA operating mode, the first final bias signal FFB provides biasing to the first in-phase driver PA stage **142** and the first in-phase final PA stage **146**.

The first quadrature-phase amplification path **128** includes a first quadrature-phase driver PA impedance matching circuit **150**, a first quadrature-phase driver PA stage **152**, a first quadrature-phase final PA impedance matching circuit **154**, a first quadrature-phase final PA stage **156**, and a first quadrature-phase combiner impedance matching circuit **158**. The first quadrature-phase driver PA impedance matching circuit **150** is coupled between the first quadrature-phase output FQO and the first quadrature-phase driver PA stage **152**. The first quadrature-phase final PA impedance matching circuit **154** is coupled between the first quadrature-phase driver PA stage **152** and the first quadrature-phase final PA stage **156**. The first quadrature-phase combiner impedance matching circuit **158** is coupled between the first quadrature-phase final PA stage **156** and the first quadrature-phase input FQI.

The first quadrature-phase driver PA impedance matching circuit **150** may provide at least an approximate impedance match between the first quadrature RF splitter **124** and the first quadrature-phase driver PA stage **152**. The first quadrature-phase final PA impedance matching circuit **154** may provide at least an approximate impedance match between the first quadrature-phase driver PA stage **152** and the first quadrature-phase final PA stage **156**. The first quadrature-phase combiner impedance matching circuit **158** may provide at least an approximate impedance match between the first quadrature-phase final PA stage **156** and the first quadrature RF combiner **130**.

During the first PA operating mode, the first quadrature-phase driver PA impedance matching circuit **150** receives and forwards the first quadrature-phase RF input signal FQN to the first quadrature-phase driver PA stage **152**, which receives and amplifies the forwarded first quadrature-phase RF input signal to provide an amplified first quadrature-phase RF input signal to the first quadrature-phase final PA stage **156** via the first quadrature-phase final PA impedance matching circuit **154**. The first quadrature-phase final PA stage **156** receives and amplifies the amplified first quadrature-phase RF input signal to provide the first quadrature-phase RF output signal FQT via the first quadrature-phase combiner impedance matching circuit **158**. During the first PA operating mode, the envelope power supply signal EPS provides power for amplification to the first quadrature-phase driver PA stage **152** and the first quadrature-phase final PA stage **156**. During the first PA operating mode, the first final bias signal FFB provides biasing to the first quadrature-phase driver PA stage **152** and the first quadrature-phase final PA stage **156**.

The second in-phase amplification path **134** includes a second in-phase driver PA impedance matching circuit **160**, a second in-phase driver PA stage **162**, a second in-phase final PA impedance matching circuit **164**, a second in-phase final PA stage **166**, and a second in-phase combiner impedance matching circuit **168**. The second in-phase driver PA impedance matching circuit **160** is coupled between the second in-phase output SIO and the second in-phase driver PA stage **162**. The second in-phase final PA impedance matching circuit **164** is coupled between the second in-phase driver PA stage **162** and the second in-phase final PA stage **166**. The second in-phase combiner impedance matching circuit **168** is coupled between the second in-phase final PA stage **166** and the second in-phase input SII.

The second in-phase driver PA impedance matching circuit **160** may provide at least an approximate impedance match between the second quadrature RF splitter **132** and the second in-phase driver PA stage **162**. The second in-phase final PA impedance matching circuit **164** may provide at least an approximate impedance match between the second in-phase driver PA stage **162** and the second in-phase final PA stage **166**. The second in-phase combiner impedance matching circuit **168** may provide at least an approximate impedance match between the second in-phase final PA stage **166** and the second quadrature RF combiner **138**.

During the second PA operating mode, the second in-phase driver PA impedance matching circuit **160** receives and forwards the second in-phase RF input signal SIN to the second in-phase driver PA stage **162**, which receives and amplifies the forwarded second in-phase RF input signal to provide an amplified second in-phase RF input signal to the second in-phase final PA stage **166** via the second in-phase final PA impedance matching circuit **164**. The second in-phase final PA stage **166** receives and amplifies the amplified second in-phase RF input signal to provide the second in-phase RF output signal SIT via the second in-phase combiner impedance matching circuit **168**. During the second PA operating mode, the envelope power supply signal EPS provides power for amplification to the second in-phase driver PA stage **162** and the second in-phase final PA stage **166**. During the second PA operating mode, the second final bias signal SFB provides biasing to the second in-phase driver PA stage **162** and the second in-phase final PA stage **166**.

The second quadrature-phase amplification path **136** includes a second quadrature-phase driver PA impedance matching circuit **170**, a second quadrature-phase driver PA stage **172**, a second quadrature-phase final PA impedance matching circuit **174**, a second quadrature-phase final PA stage **176**, and a second quadrature-phase combiner impedance matching circuit **178**. The second quadrature-phase driver PA impedance matching circuit **170** is coupled between the second quadrature-phase output SQO and the second quadrature-phase driver PA stage **172**. The second quadrature-phase final PA impedance matching circuit **174** is coupled between the second quadrature-phase driver PA stage **172** and the second quadrature-phase final PA stage **176**. The second quadrature-phase combiner impedance matching circuit **178** is coupled between the second quadrature-phase final PA stage **176** and the second quadrature-phase input SQI.

The second quadrature-phase driver PA impedance matching circuit **170** may provide at least an approximate impedance match between the second quadrature RF splitter **132** and the second quadrature-phase driver PA stage **172**. The second quadrature-phase final PA impedance matching circuit **174** may provide at least an approximate impedance match between the second quadrature-phase driver PA stage **172** and the second quadrature-phase final PA stage **176**. The second quadrature-phase combiner impedance matching circuit **178** may provide at least an approximate impedance match between the second quadrature-phase final PA stage **176** and the second quadrature RF combiner **138**.

During the second PA operating mode, the second quadrature-phase driver PA impedance matching circuit **170** receives and forwards the second quadrature-phase RF input signal SQN to the second quadrature-phase driver PA stage **172**, which receives and amplifies the forwarded second quadrature-phase RF input signal to provide an amplified second quadrature-phase RF input signal to the second quadrature-phase final PA stage **176** via the second quadrature-phase final PA impedance matching circuit **174**. The second quadrature-phase final PA stage **176** receives and amplifies the amplified second quadrature-phase RF input signal to provide the second quadrature-phase RF output signal SQT via the second quadrature-phase combiner impedance matching circuit **178**. During the second PA operating mode, the envelope power supply signal EPS provides power for amplification to the second quadrature-phase driver PA stage **172** and the second quadrature-phase final PA stage **176**. During the second PA operating mode, the second final bias signal SFB provides biasing to the second quadrature-phase driver PA stage **172** and the second quadrature-phase final PA stage **176**.

In alternate embodiments of the first in-phase amplification path **126**, any or all of the first in-phase driver PA impedance matching circuit **140**, the first in-phase driver PA stage **142**, the first in-phase final PA impedance matching circuit **144**, and the first in-phase combiner impedance matching circuit **148** may be omitted. In alternate embodiments of the first quadrature-phase amplification path **128**, any or all of the first quadrature-phase driver PA impedance matching circuit **150**, the first quadrature-phase driver PA stage **152**, the first quadrature-phase final PA impedance matching circuit **154**, and the first quadrature-phase combiner impedance matching circuit **158** may be omitted. In alternate embodiments of the second in-phase amplification path **134**, any or all of the second in-phase driver PA impedance matching circuit **160**, the second in-phase driver PA stage **162**, the second in-phase final PA impedance matching circuit **164**, and the second in-phase combiner impedance matching circuit **168** may be omitted. In alternate embodiments of the second quadrature-phase amplification path **136**, any or all of the second quadrature-phase driver PA impedance matching circuit **170**, the second quadrature-phase driver PA stage **172**, the second quadrature-phase final PA impedance matching circuit **174**, and the second quadrature-phase combiner impedance matching circuit **178** may be omitted.

FIG. **19** shows details of the first quadrature PA path **102** and the second quadrature PA path **106** illustrated in FIG. **15** according to an alternate embodiment of the first quadrature PA path **102** and the second quadrature PA path **106**. The first quadrature PA path **102** and the second quadrature PA path **106** illustrated in FIG. **19** are similar to the first quadrature PA path **102** and the second quadrature PA path **106** illustrated in FIG. **17**, except in the first quadrature PA path **102** and the second quadrature PA path **106** illustrated in FIG. **19**, during the first PA operating mode, the first driver bias signal FDB provides further biasing to the first in-phase amplification path **126** and the first quadrature-phase amplification path **128**, and during the second PA operating mode, the second driver bias signal SDB provides further biasing to the second in-phase amplification path **134** and the second quadrature-phase amplification path **136**.

FIG. **20** shows details of the first in-phase amplification path **126**, the first quadrature-phase amplification path **128**, the second in-phase amplification path **134**, and the second quadrature-phase amplification path **136** illustrated in FIG. **19** according to an alternate embodiment of the first in-phase amplification path **126**, the first quadrature-phase amplification path **128**, the second in-phase amplification path **134**, and the second quadrature-phase amplification path **136**. The amplification paths **126**, **128**, **134**, **136** illustrated in FIG. **20** are similar to the amplification paths **126**, **128**, **134**, **136** illustrated in FIG. **18**, except in the amplification paths **126**, **128**, **134**, **136** illustrated in FIG. **20**, during the first PA operating mode, the first driver bias signal FDB provides biasing to the first in-phase driver PA stage **142** and the first quadrature-phase driver PA stage **152** instead of the first final bias signal FFB, and during the second PA operating mode, the second driver bias signal SDB provides biasing to the second in-phase driver PA stage **162** and the second quadrature-phase driver PA stage **172** instead of the second final bias signal SFB.

FIG. **21** shows details of the first RF PA **50** and the second RF PA **54** illustrated in FIG. **14** according an alternate embodiment of the first RF PA **50** and the second RF PA **54**. The first RF PA **50** shown in FIG. **21** is similar to the first RF PA **50** illustrated in FIG. **15**. The second RF PA **54** shown in FIG. **21** is similar to the second RF PA **54** illustrated in FIG. **15**, except in the second RF PA **54** illustrated in FIG. **21** the second quadrature PA path **106** is omitted. As such, during the second PA operating mode, the second RF input signal SRFI provides the second RF feeder output signal SFO to the second quadrature PA path **106**. In this regard, during the second PA operating mode, the second quadrature PA path **106** receives and amplifies the second RF input signal SRFI to provide the second RF output signal SRFO. During the second PA operating mode, the second quadrature PA path **106** receives the envelope power supply signal EPS, which provides power for amplification. Further, during the second PA operating mode, the second quadrature PA path **106** receives the second driver bias signal SDB and the second final bias signal SFB, both of which provide biasing to the second quadrature PA path **106**.

FIG. **22** shows details of the first non-quadrature PA path **100**, the first quadrature PA path **102**, and the second quadrature PA path **106** illustrated in FIG. **21** according to an additional embodiment of the first non-quadrature PA path **100**, the first quadrature PA path **102**, and the second quadrature PA path **106**. The second quadrature PA path **106** illustrated in FIG. **22** is similar to the second quadrature PA path **106** illustrated in FIG. **20**. The first quadrature PA path **102** illustrated in FIG. **22** is similar to the first quadrature PA path **102** illustrated in FIG. **20**, except in the first quadrature PA path **102** illustrated in FIG. **22**, the first in-phase driver PA impedance matching circuit **140**, the first in-phase driver PA stage **142**, the first quadrature-phase driver PA impedance matching circuit **150**, and the first quadrature-phase driver PA stage **152** are omitted. In this regard, the first in-phase final PA impedance matching circuit **144** is coupled between the first in-phase output FIO and the first in-phase final PA stage **146**. The first in-phase combiner impedance matching circuit **148** is coupled between the first in-phase final PA stage **146** and the first in-phase input FII. The first in-phase final PA impedance matching circuit **144** may provide at least an approximate impedance match between the first quadrature RF splitter **124** and the first in-phase final PA stage **146**. The first in-phase combiner impedance matching circuit **148** may provide at least an approximate impedance match between the first in-phase final PA stage **146** and the first quadrature RF combiner **130**.

During the first PA operating mode, the first in-phase final PA impedance matching circuit **144** receives and forwards the first in-phase RF input signal FIN to the first in-phase final PA stage **146**, which receives and amplifies the forwarded first in-phase RF input signal to provide the first in-phase RF output signal FIT via the first in-phase combiner impedance matching circuit **148**. During the first PA operating mode, the envelope power supply signal EPS provides power for amplification to the first in-phase final PA stage **146**. During the first PA operating mode, the first final bias signal FFB provides biasing to the first in-phase final PA stage **146**.

The first quadrature-phase final PA impedance matching circuit **154** is coupled between the first quadrature-phase output FQO and the first quadrature-phase final PA stage **156**. The first quadrature-phase combiner impedance matching circuit **158** is coupled between the first quadrature-phase final PA stage **156** and the first quadrature-phase input FQI. The first quadrature-phase final PA impedance matching circuit **154** may provide at least an approximate impedance match between the first quadrature RF splitter **124** and the first quadrature-phase final PA stage **156**. The first quadrature-phase combiner impedance matching circuit **158** may provide at least an approximate impedance match between the first quadrature-phase final PA stage **156** and the first quadrature RF combiner **130**.

During the first PA operating mode, the first quadrature-phase final PA impedance matching circuit **154** receives and forwards the first quadrature-phase RF input signal FQN to the first quadrature-phase final PA stage **156**, which receives and amplifies the forwarded first quadrature-phase RF input signal to provide the first quadrature-phase RF output signal FQT via the first quadrature-phase combiner impedance matching circuit **158**. During the first PA operating mode, the envelope power supply signal EPS provides power for amplification to the first quadrature-phase final PA stage **156**. During the first PA operating mode, the first final bias signal FFB provides biasing to the first quadrature-phase final PA stage **156**.

The first non-quadrature PA path **100** illustrated in FIG. **22** is similar to the first non-quadrature PA path **100** illustrated in FIG. **16**, except in the first non-quadrature PA path **100** illustrated in FIG. **22**, the first input PA impedance matching circuit **108** and the first input PA stage **110** are omitted. As such, the first feeder PA stage **114** is coupled between the first feeder PA impedance matching circuit **112** and the first quadrature PA path **102**. The first feeder PA impedance matching circuit **112** may provide at least an approximate impedance match between the RF modulation circuitry **44** (FIG. **5**) and the first feeder PA stage **114**. During the first PA operating mode, the first feeder PA impedance matching circuit **112** receives and forwards the first RF input signal FRFI to provide the first RF feeder input signal FFI to the first feeder PA stage **114**. During the first PA operating mode, the first feeder PA stage **114** receives and amplifies the first RF feeder input signal FFI to provide the first RF feeder output signal FFO via the first single-ended output FSO. During the first PA operating mode, the envelope power supply signal EPS provides power for amplification to the first feeder PA stage **114**. During the first PA operating mode, the first final bias signal FFB provides biasing to the first feeder PA stage **114**.

In one embodiment of the first quadrature PA path **102**, the first quadrature PA path **102** has only one in-phase PA stage, which is the first in-phase final PA stage **146**, and only one quadrature-phase PA stage, which is the first quadrature-phase final PA stage **156**. In one embodiment of the second quadrature PA path **106**, the second in-phase driver PA impedance matching circuit **160**, the second in-phase driver PA stage **162**, the second quadrature-phase driver PA impedance matching circuit **170**, and the second quadrature-phase driver PA stage **172** are omitted. As such, the second quadrature PA path **106** has only one in-phase PA stage, which is the second in-phase final PA stage **166**, and only one quadrature-phase PA stage, which is the second quadrature-phase final PA stage **176**.

FIG. **23** shows details of the first feeder PA stage **114** and the first quadrature RF splitter **124** illustrated in FIG. **16** and FIG. **17**, respectively, according to one embodiment of the first feeder PA stage **114** and the first quadrature RF splitter **124**. FIGS. **23** and **24** show only a portion of the first feeder PA stage **114** and the first quadrature RF splitter **124**. The first feeder PA stage **114** includes a first output transistor element **180**, an inverting output inductive element LIO, and the first single-ended output FSO. The first output transistor element **180** has a first transistor inverting output FTIO, a first transistor non-inverting output FTNO, and a first transistor input FTIN. The first transistor non-inverting output FTNO is coupled to a ground and the first transistor inverting output FTIO is coupled to the first single-ended output FSO and to one end of the inverting output inductive element LIO. An opposite end of the inverting output inductive element LIO receives the envelope power supply signal EPS.

The first quadrature RF splitter **124** has the first single-ended input FSI, such that the first input impedance is presented at the first single-ended input FSI. Since the first input impedance may be predominantly resistive, the first input impedance may be approximated as a first input resistive element RFI coupled between the first single-ended input FSI and the ground. The first single-ended output FSO is directly coupled to the first single-ended input FSI. Therefore, the first input resistive element RFI is presented to the first transistor inverting output FTIO.

FIG. **24** shows details of the first feeder PA stage **114** and the first quadrature RF splitter **124** illustrated in FIG. **16** and FIG. **17**, respectively, according to an alternate embodiment of the first feeder PA stage **114** and the first quadrature RF splitter **124**. The first output transistor element **180** is an NPN bipolar transistor element, such that an emitter of the NPN bipolar transistor element provides the first transistor non-inverting output FTNO (FIG. **23**), a base of the NPN bipolar transistor element provides the first transistor input FTIN (FIG. **23**), and a collector of the NPN bipolar transistor element provides the first transistor inverting output FTIO (FIG. **23**). The inverting output inductive element LIO has an inverting output inductor current IDC, the collector of the NPN bipolar transistor element has a collector current IC, and the first input resistive element RFI has a first input current IFR. The NPN bipolar transistor element has a collector-emitter voltage VCE between the emitter and the collector of the NPN bipolar transistor element.

In general, the first feeder PA stage **114** is the feeder PA stage having the single-ended output and an output transistor element, which has an inverting output. In general, the first quadrature RF splitter **124** is the quadrature RF splitter having the single-ended input, such that the input impedance is presented at the single-ended input. The inverting output may provide the single-ended output and may be directly coupled to the single-ended input. The inverting output may be a collector of the output transistor element and the output transistor element has the output load line.

FIG. **25** is a graph illustrating output characteristics of the first output transistor element **180** illustrated in FIG. **24** according to one embodiment of the first output transistor element **180**. The horizontal axis of the graph represents the collector-emitter voltage VCE of the NPN bipolar transistor element and the vertical axis represents the collector current IC of the NPN bipolar transistor element. Characteristic curves **182** of the NPN bipolar transistor element are shown relating the collector-emitter voltage VCE to the collector current IC at different base currents (not shown). The NPN bipolar transistor element has a first output load line **184** having a first load line slope **186**. The first output load line **184** may be represented by an equation for a straight line having the form Y=mX+b, where X represents the horizontal axis, Y represents the vertical axis, b represents the Y-intercept, and m represents the first load line slope **186**. As such, Y=IC, X=VCE, and b=ISAT, which is a saturation current ISAT of the NPN bipolar transistor element. Further, an X-intercept occurs at an off transistor voltage VCO. Substituting into the equation for a straight line provides EQ. 1, as shown below.

$$IC=m(VCE)+ISAT. \qquad \text{EQ. 1}$$

EQ. 2 illustrates Ohm's Law as applied to the first input resistive element RFI, as shown below.

$$VCE=(IFR)(RFI). \qquad \text{EQ. 2:}$$

EQ. 3 illustrates Kirchhoff's Current Law applied to the circuit illustrated in FIG. **24** as shown below.

$$IDC=IC+IFR. \qquad \text{EQ. 3}$$

The inductive reactance of the inverting output inductive element LIO at frequencies of interest may be large compared to the resistance of the first input resistive element RFI. As such, for the purpose of analysis, the inverting output inductor current IDC may be treated as a constant DC current. Therefore, when VCE=0, the voltage across the first input resistive element RFI is zero, which makes IFR=0. From EQ. 3, if IFR=0, then IC=IDC. However, from EQ. 1, when VCE=0 and IC=IDC, then ISAT=IDC, which is a constant. Substituting into EQ. 1 provides EQ. 1A as shown below.

$$IC=m(VCE)+IDC. \qquad \text{EQ. 1A}$$

From FIG. **25**, when IC=0, VCE=VCO. Substituting into EQ. 1A, EQ. 2, and EQ. 3 provides EQ. 1 B, EQ. 2A, and EQ. 3A as shown below.

$$0=m(VCO)+IDC. \qquad \text{EQ. 1B}$$

$$VCO=(IFR)(RFI). \qquad \text{EQ. 2A}$$

$$IDC=0+IFR. \qquad \text{EQ. 3A}$$

EQ. 3A may be substituted into EQ. 2A, which may be substituted into EQ. 1B to provide EQ. 1C as shown below.

$$0=m(VCO)+IDC=m(IDC)(RFI)+IDC. \qquad \text{EQ. 1C}$$

Therefore, m=−1/RFI. As a result, the first load line slope **186**, which is represented by m is determined by the first input resistive element RFI, such that there is a negative inverse relationship between the first load line slope **186** and the first input resistive element RFI. In general, the first load line slope **186** is based on the first input impedance, such that the first input impedance substantially establishes the first load line slope **186**. Further, there may be a negative inverse relationship between the first load line slope **186** and the first input impedance.

FIG. **26** illustrates a process for matching an input impedance, such as the first input impedance to the first quadrature RF splitter **124** (FIG. **16**) to a target load line slope for a feeder PA stage, such as the first feeder PA stage **114** (FIG. **17**). The first step of the process is to determine an operating power range of an RF PA, which has the feeder PA stage feeding a quadrature RF splitter (Step A**10**). The next step of the process is to determine the target load line slope for the feeder PA stage based on the operating power range (Step A**12**). A further step is to determine the input impedance to the quadrature RF splitter that substantially provides the target load line slope (Step A**14**). The final step of the process is to determine an operating frequency range of the RF PA, such that the target load line slope is further based on the operating frequency range (Step A**16**). In an alternate embodiment of the process for matching the input impedance to the target load line slope, the final step (Step A**16**) is omitted.

FIG. **27** shows details of the first RF PA **50** illustrated in FIG. **14** according an alternate embodiment of the first RF PA **50**. The first RF PA **50** illustrated in FIG. **27** is similar to the first RF PA **50** illustrated in FIG. **15**, except the first RF PA **50** illustrated in FIG. **27** further includes a first non-quadrature path power coupler **188**. As previously mentioned, the first quadrature PA path **102** may present a first input impedance at the first single-ended input FSI that is predominantly resistive. Further, the first input impedance may be stable over a wide frequency range and over widely varying antenna loading conditions. As a result, coupling RF power from the first single-ended output FSO may be used for RF power detection or sampling with a high degree of accuracy and directivity. Since the first single-ended input FSI may be directly coupled to the first single-ended output FSO, coupling RF power from the first single-ended output FSO may be equivalent to coupling RF power from the first single-ended input FSI.

The first non-quadrature path power coupler **188** is coupled to the first single-ended output FSO and couples a portion of RF power flowing though the first single-ended output FSO to provide a first non-quadrature path power output signal FNPO. In an additional embodiment of the first RF PA **50**, the first non-quadrature path power coupler **188** is coupled to the first single-ended input FSI and couples a portion of RF power flowing though the first single-ended input FSI to provide the first non-quadrature path power output signal FNPO.

FIG. **28** shows details of the second RF PA **54** illustrated in FIG. **14** according an alternate embodiment of the second RF PA **54**. The second RF PA **54** illustrated in FIG. **28** is similar to the second RF PA **54** illustrated in FIG. **15**, except the second RF PA **54** illustrated in FIG. **28** further includes a second non-quadrature path power coupler **190**. As previously mentioned, the second quadrature PA path **106** may present a second input impedance at the second single-ended input SSI that is predominantly resistive. Further, the second input impedance may be stable over a wide frequency range and over widely varying antenna loading conditions. As a result, coupling RF power from the second single-ended output SSO may be used for RF power detection or sampling with a high degree of accuracy and directivity. Since the second single-ended input SSI may be directly coupled to the second single-ended output SSO, coupling RF power from the second single-ended output SSO may be equivalent to coupling RF power from the second single-ended input SSI.

The second non-quadrature path power coupler **190** is coupled to the second single-ended output SSO and couples a portion of RF power flowing though the second single-ended output SSO to provide a second non-quadrature path power output signal SNPO. In an additional embodiment of the second RF PA **54**, the second non-quadrature path power coupler **190** is coupled to the second single-ended input SSI and couples a portion of RF power flowing though the second single-ended input SSI to provide the second non-quadrature path power output signal SNPO.

FIG. **29** shows details of the first in-phase amplification path **126**, the first quadrature-phase amplification path **128**, and the first quadrature RF combiner **130** illustrated in FIG. **22** according to one embodiment of the first in-phase amplification path **126**, the first quadrature-phase amplification path **128**, and the first quadrature RF combiner **130**. The first in-phase combiner impedance matching circuit **148** and the first quadrature-phase combiner impedance matching circuit **158** have been omitted from the first in-phase amplification path **126** and the first quadrature-phase amplification path **128**, respectively. The first quadrature RF combiner **130** includes first phase-shifting circuitry **192** and a first Wilkinson RF combiner **194**. The first phase-shifting circuitry **192** has the first in-phase input FII and the first quadrature-phase input FQI. The first Wilkinson RF combiner **194** has the first quadrature combiner output FCO.

During the first PA operating mode, the first phase-shifting circuitry **192** receives and phase-aligns RF signals from the first in-phase final PA stage **146** and the first quadrature-phase final PA stage **156** via the first in-phase input FII and the first quadrature-phase input FQI, respectively, to provide phase-aligned RF signals to the first Wilkinson RF combiner **194**. The first Wilkinson RF combiner **194** combines phase-aligned RF signals to provide the first RF output signal FRFO via the first quadrature combiner output FCO. The first phase-shifting circuitry **192** and the first Wilkinson RF combiner **194** may provide stable input impedances presented at the first in-phase input FII and the first quadrature-phase input FQI, respectively, which allows elimination of the first in-phase combiner impedance matching circuit **148** and the first quadrature-phase combiner impedance matching circuit **158**.

FIG. **30** shows details of the first feeder PA stage **114**, the first quadrature RF splitter **124**, the first in-phase final PA impedance matching circuit **144**, the first in-phase final PA stage **146**, the first quadrature-phase final PA impedance matching circuit **154**, and the first quadrature-phase final PA stage **156** illustrated in FIG. **29** according to one embodiment of the first feeder PA stage **114**, the first quadrature RF splitter **124**, the first in-phase final PA impedance matching circuit **144**, the first in-phase final PA stage **146**, the first quadrature-phase final PA impedance matching circuit **154**, and the first quadrature-phase final PA stage **156**. Further, FIG. **30** shows a portion of the first phase-shifting circuitry **192** illustrated in FIG. **29**.

The first in-phase final PA stage **146** includes a first in-phase final transistor element **196**, first in-phase biasing circuitry **198**, and a first in-phase collector inductive element LCI. The first quadrature-phase final PA stage **156** includes a first quadrature-phase final transistor element **200**, first quadrature-phase biasing circuitry **202**, and a first quadrature-phase collector inductive element LCQ. The first in-phase final PA impedance matching circuit **144** includes a first in-phase series capacitive element CSI**1**, a second in-phase series capacitive element CSI**2**, and a first in-phase shunt inductive element LUI. The first quadrature-phase final PA impedance matching circuit **154** includes a first quadrature-phase series capacitive element CSQ**1**, a second quadrature-phase series capacitive element CSQ**2**, and a first quadrature-phase shunt inductive element LUQ.

The first quadrature RF splitter **124** includes a first pair **204** of tightly coupled inductors and a first isolation port resistive element RI**1**. The first pair **204** of tightly coupled inductors has first parasitic capacitance **206** between the first pair **204** of tightly coupled inductors. Additionally, the first quadrature RF splitter **124** has the first single-ended input FSI, the first in-phase output FIO, and the first quadrature-phase output FQO. The first feeder PA stage **114** includes the first output transistor element **180**, first feeder biasing circuitry **208**, a first DC blocking capacitive element CD**1**, a first base resistive element RB**1**, and a first collector inductive element LC**1**. Additionally, the first feeder PA stage **114** has the first single-ended output FSO.

The first output transistor element **180** shown is an NPN bipolar transistor element. Other embodiments of the first output transistor element **180** may use other types of transistor elements, such as field effect transistor elements (FET) elements. The first DC blocking capacitive element CD**1** is coupled between the first feeder PA impedance matching circuit **112** (FIG. **22**) and the first base resistive element RB. A base of the first output transistor element **180** and the first feeder biasing circuitry **208** are coupled to the first base resistive element RB**1**. In alternate embodiments of the first feeder PA stage **114**, the first base resistive element RB**1**, the first DC blocking capacitive element CD**1**, or both may be omitted. The first feeder biasing circuitry **208** receives the first driver bias signal FDB. An emitter of the first output transistor element **180** is coupled to a ground. A collector of the first output transistor element **180** is coupled to the first single-ended output FSO. One end of the first collector inductive element LC**1** is coupled to the first single-ended output FSO. An opposite end of the first collector inductive element LC**1** receives the envelope power supply signal EPS. The first single-ended output FSO is coupled to the first single-ended input FSI.

During the first PA operating mode, the first output transistor element **180** receives and amplifies an RF signal from the first feeder PA impedance matching circuit **112** (FIG. **22**) via the first DC blocking capacitive element CD**1** and the first base resistive element RB**1** to provide the first RF feeder output signal FFO (FIG. **29**) to the first single-ended input FSI via the first single-ended output FSO. The envelope power supply signal EPS provides power for amplification via the first collector inductive element LC**1**. The first feeder biasing circuitry **208** biases the first output transistor element **180**. The first driver bias signal FDB provides power for biasing the first output transistor element **180** to the first feeder biasing circuitry **208**.

The first quadrature RF splitter **124** illustrated in FIG. **30** is a quadrature hybrid coupler. In this regard, the first pair **204** of tightly coupled inductors, the first parasitic capacitance **206**, and the first isolation port resistive element RI**1** provide quadrature hybrid coupler functionality. As such, the first single-ended input FSI functions as an input port to the quadrature hybrid coupler, the first in-phase output FIO functions as a zero degree output port from the quadrature hybrid coupler, and the first quadrature-phase output FQO functions as a 90 degree output port from the quadrature hybrid coupler. One of the first pair **204** of tightly coupled inductors is coupled between the first single-ended input FSI and the first in-phase output F**10**. Another of the first pair **204** of tightly coupled inductors has a first end coupled to the first quadrature-phase output FQO and a second end coupled to the first isolation port resistive element RI**1**. As such, the second end functions as an isolation port of the quadrature hybrid coupler. In this regard, the first isolation port resistive element RI**1** is coupled between the isolation port and the ground. The first in-phase output FIO is coupled to the first in-phase series capacitive element CSI**1** and the first quadrature-phase output FQO is coupled to the first quadrature-phase series capacitive element CSQ**1**.

During the first PA operating mode, the first pair **204** of tightly coupled inductors receives, splits, and phase-shifts the first RF feeder output signal FFO (FIG. **29**) from the first single-ended output FSO via the first single-ended input FSI to provide split, phase-shifted output signals to the first in-phase series capacitive element CSI**1** and the first quadrature-phase series capacitive element CSQ**1**. As previously mentioned, the first input impedance is presented at the first single-ended input FSI. As such, the first input impedance is substantially based on the first parasitic capacitance **206** and inductances of the first pair **204** of tightly coupled inductors.

The first in-phase series capacitive element CSI**1** and the second in-phase series capacitive element CSI**2** are coupled in series between the first in-phase output FIO and a base of the first in-phase final transistor element **196**. The first in-phase shunt inductive element LUI is coupled between the ground and a junction between the first in-phase series capacitive element CSI**1** and the second in-phase series capacitive element CSI**2**. The first quadrature-phase series capacitive element CSQ**1** and the second quadrature-phase series capacitive element CSQ**2** are coupled in series between the first quadrature-phase output FQO and a base of the first quadrature-phase final transistor element **200**. The first quadrature-phase shunt inductive element LUQ is coupled between the ground and a junction between the first quadrature-phase series capacitive element CSQ**1** and the second quadrature-phase series capacitive element CSQ**2**.

The first in-phase series capacitive element CSI**1**, the second in-phase series capacitive element CSI**2**, and the first in-phase shunt inductive element LUI form a "T" network, which may provide at least an approximate impedance match between the first in-phase output FIO and the base of the first in-phase final transistor element **196**. Similarly, the first quadrature-phase series capacitive element CSQ**1**, the second quadrature-phase series capacitive element CSQ**2**, and the first quadrature-phase shunt inductive element LUQ form a "T" network, which may provide at least an approximate impedance match between the first quadrature-phase output FQO and the base of the first quadrature-phase final transistor element **200**.

During the first PA operating mode, the first in-phase final PA impedance matching circuit **144** receives and forwards an RF signal from the first in-phase output FIO to the base of the first in-phase final transistor element **196** via the first in-phase series capacitive element CSI**1** and the second in-phase series capacitive element CSI**2**. During the first PA operating mode, the first quadrature-phase final PA impedance matching circuit **154** receives and forwards an RF signal from the first quadrature-phase output FQO to the base of the first quadrature-phase final transistor element **200** via the first quadrature-phase series capacitive element CSQ**1** and the second quadrature-phase series capacitive element CSQ**2**.

The first in-phase final transistor element **196** shown is an NPN bipolar transistor element. Other embodiments of the first in-phase final transistor element **196** may use other types of transistor elements, such as FET elements. The base of the first in-phase final transistor element **196** and the first in-phase biasing circuitry **198** are coupled to the second in-phase series capacitive element CSI**2**. The first in-phase biasing circuitry **198** receives the first final bias signal FFB. An emitter of the first in-phase final transistor element **196** is coupled to the ground. A collector of the first in-phase final transistor element **196** is coupled to the first in-phase input FII. One end of the first in-phase collector inductive element LCI is coupled to the collector of the first in-phase final transistor element **196**. An opposite end of the first in-phase collector inductive element LCI receives the envelope power supply signal EPS.

During the first PA operating mode, the first in-phase final transistor element **196** receives and amplifies an RF signal from the second in-phase series capacitive element CSI**2** to provide an RF output signal to the first in-phase input FII. The envelope power supply signal EPS provides power for amplification via the first in-phase collector inductive element LCI. The first in-phase biasing circuitry **198** biases the first in-phase final transistor element **196**. The first final bias signal FFB provides power for biasing the first in-phase final transistor element **196** to the first in-phase biasing circuitry **198**.

The first quadrature-phase final transistor element **200** shown is an NPN bipolar transistor element. Other embodiments of the first quadrature-phase final transistor element **200** may use other types of transistor elements, such as FET elements. The base of the first quadrature-phase final transistor element **200** and the first quadrature-phase biasing circuitry **202** are coupled to the second quadrature-phase series capacitive element CSQ**2**. The first quadrature-phase biasing circuitry **202** receives the first final bias signal FFB. An emitter of the first quadrature-phase final transistor element **200** is coupled to the ground. A collector of the first quadrature-phase final transistor element **200** is coupled to the first quadrature-phase input FQI. One end of the first quadrature-phase collector inductive element LCQ is coupled to the collector of the first quadrature-phase final transistor element **200**. An opposite end of the first quadrature-phase collector inductive element LCQ receives the envelope power supply signal EPS.

During the first PA operating mode, the first quadrature-phase final transistor element **200** receives and amplifies an RF signal from the second quadrature-phase series capacitive element CSQ**2** to provide an RF output signal to the first quadrature-phase input FQI. The envelope power supply signal EPS provides power for amplification via the first quadrature-phase collector inductive element LCQ. The first quadrature-phase biasing circuitry **202** biases the first quadrature-phase final transistor element **200**. The first final bias signal FFB provides power for biasing the first quadrature-phase final transistor element **200** to the first quadrature-phase biasing circuitry **202**.

In one embodiment of the RF PA circuitry **30** (FIG. **5**), the RF PA circuitry **30** includes a first PA semiconductor die **210**. In one embodiment of the first PA semiconductor die **210**, the first PA semiconductor die **210** includes the first output transistor element **180**, the first in-phase final transistor element **196**, the first in-phase biasing circuitry **198**, the first quadrature-phase final transistor element **200**, the first quadrature-phase biasing circuitry **202**, the first pair **204** of tightly coupled inductors, the first feeder biasing circuitry **208**, the first in-phase series capacitive element CSI**1**, the second in-phase series capacitive element CSI**2**, the first quadrature-phase series capacitive element CSQ**1**, the second quadrature-phase series capacitive element CSQ**2**, the first isolation port resistive element RI**1**, the first base resistive element RB**1**, and the first DC blocking capacitive element CD**1**.

In alternate embodiments of the first PA semiconductor die **210**, the first PA semiconductor die **210** may not include any or all of the first output transistor element **180**, the first in-phase final transistor element **196**, the first in-phase biasing circuitry **198**, the first quadrature-phase final transistor element **200**, the first quadrature-phase biasing circuitry **202**, the first pair **204** of tightly coupled inductors, the first feeder biasing circuitry **208**, the first in-phase series capacitive element CSI**1**, the second in-phase series capacitive element CSI**2**, the first quadrature-phase series capacitive element CSQ**1**, the second quadrature-phase series capacitive element CSQ**2**, the first isolation port resistive element RI**1**, the first base resistive element RB**1**, and the first DC blocking capacitive element CD**1**.

FIG. **31** shows details of the first feeder PA stage **114**, the first quadrature RF splitter **124**, the first in-phase final PA impedance matching circuit **144**, the first in-phase final PA stage **146**, the first quadrature-phase final PA impedance matching circuit **154**, and the first quadrature-phase final PA stage **156** illustrated in FIG. **29** according to an alternate embodiment of the first feeder PA stage **114**, the first quadrature RF splitter **124**, the first in-phase final PA impedance matching circuit **144**, the first in-phase final PA stage **146**, the first quadrature-phase final PA impedance matching circuit **154**, and the first quadrature-phase final PA stage **156**. Further, FIG. **31** shows a portion of the first phase-shifting circuitry **192** illustrated in FIG. **29**.

The first feeder PA stage **114**, the first in-phase final PA impedance matching circuit **144**, the first in-phase final PA stage **146**, the first quadrature-phase final PA impedance matching circuit **154**, and the first quadrature-phase final PA stage **156** illustrated in FIG. **31** are similar to the first feeder PA stage **114**, the first in-phase final PA impedance matching circuit **144**, the first in-phase final PA stage **146**, the first quadrature-phase final PA impedance matching circuit **154**, and the first quadrature-phase final PA stage **156** illustrated in FIG. **30**. The first quadrature RF splitter **124** illustrated in FIG. **31** is similar to the first quadrature RF splitter **124** illustrated in FIG. **30**, except the first quadrature RF splitter **124** illustrated in FIG. **31** further includes a first coupler capacitive element CC**1** coupled between the first pair **204** of tightly coupled inductors and a second coupler capacitive element CC**2** coupled between the first pair **204** of tightly coupled inductors. Specifically, the first coupler capacitive element CC**1** is coupled between the first in-phase output FIO and the first isolation port resistive element RI**1**. The second coupler capacitive element CC**2** is coupled between the first single-ended input FSI and the first quadrature-phase output FQO.

The first input impedance is substantially based on the first parasitic capacitance **206**, inductances of the first pair **204** of tightly coupled inductors, the first coupler capacitive element CC**1**, and the second coupler capacitive element CC**2**. In general, the first input impedance is based on the first parasitic capacitance **206** and inductances of the first pair **204** of tightly coupled inductors. The first input impedance is further based on at least one coupler capacitive element, such as the first coupler capacitive element CC**1**, the second coupler capacitive element CC**2**, or both, coupled between the first pair **204** of tightly coupled inductors. In an alternate embodiment of the first quadrature RF splitter **124**, either the first coupler capacitive element CCI or the second coupler capacitive element CC**2** is omitted.

FIG. **32** shows details of the first phase-shifting circuitry **192** and the first Wilkinson RF combiner **194** illustrated in FIG. **29** according to one embodiment of the first phase-shifting circuitry **192** and the first Wilkinson RF combiner **194**. The first phase-shifting circuitry **192** includes a first in-phase phase-shift capacitive element CPI**1**, a first quadrature-phase phase-shift capacitive element CPQ**1**, a first in-phase phase-shift inductive element LPI**1**, and a first quadrature-phase phase-shift inductive element LPQ**1**. The first Wilkinson RF combiner **194** includes a first Wilkinson resistive element RW**1**, a first Wilkinson capacitive element CW**1**, a first Wilkinson in-phase side capacitive element CWI**1**, a first Wilkinson quadrature-phase side capacitive element CWQ**1**, a first Wilkinson in-phase side inductive element LWI**1**, a first Wilkinson quadrature-phase side inductive element LWQ**1**, a second DC blocking capacitive element CD**2**, a third DC blocking capacitive element CD**3**, and a fourth DC blocking capacitive element CD**4**

The first in-phase phase-shift capacitive element CPI**1** is coupled between the first in-phase input FII and a first internal node (not shown). The first in-phase phase-shift inductive element LPI**1** is coupled between the first internal node and the ground. The first quadrature-phase phase-shift inductive element LPQ**1** is coupled between the first quadrature-phase input FQI and a second internal node (not shown). The first quadrature-phase phase-shift capacitive element CPQ**1** is coupled between the second internal node and the ground. The second DC blocking capacitive element CD**2** and the first Wilkinson resistive element RW**1** are coupled in series between the first internal node and the second internal node. The first Wilkinson in-phase side capacitive element CWI**1** is coupled between the first internal node and the ground. The first Wilkinson quadrature-phase side capacitive element CWQ**1** is coupled between the first internal node and the ground. The first Wilkinson in-phase side inductive element LWI**1** is coupled in series with the third DC blocking capacitive element CD**3** between the first internal node and the first quadrature combiner output FCO. The first Wilkinson quadrature-phase side inductive element LWQ**1** is coupled in series with the fourth DC blocking capacitive element CD**4** between the second internal node and the first quadrature combiner output FCO. The first Wilkinson capacitive element CW**1** is coupled between the first quadrature combiner output FCO and the ground.

FIG. **33** shows details of the second non-quadrature PA path **104** illustrated in FIG. **16** and details of the second quadrature PA path **106** illustrated in FIG. **18** according to one embodiment of the second non-quadrature PA path **104** and the second quadrature PA path **106**. Further, FIG. **33** shows details of the second quadrature RF combiner **138** illustrated in FIG. **18** according to one embodiment of the second quadrature RF combiner **138** illustrated in FIG. **18**. The second input PA impedance matching circuit **116**, the second input PA stage **118**, the second in-phase driver PA impedance matching circuit **160**, the second in-phase driver PA stage **162**, the second in-phase combiner impedance matching circuit **168**, the second quadrature-phase driver PA impedance matching circuit **170**, the second quadrature-phase driver PA stage **172**, and the second quadrature-phase combiner impedance matching circuit **178** have been omitted from the second non-quadrature PA path **104** and the second quadrature PA path **106**.

The second quadrature RF combiner **138** includes second phase-shifting circuitry **212** and a second Wilkinson RF combiner **214**. The second phase-shifting circuitry **212** has the second in-phase input SII and the second quadrature-phase input SQI, and the second Wilkinson RF combiner **214** has the second quadrature combiner output SCO.

During the second PA operating mode, the second phase-shifting circuitry **212** receives and phase-aligns RF signals from the second in-phase final PA stage **166** and the second quadrature-phase final PA stage **176** via the second in-phase input SII and the second quadrature-phase input SQI, respectively, to provide phase-aligned RF signals to the second Wilkinson RF combiner **214**. The second Wilkinson RF combiner **214** combines phase-aligned RF signals to provide the second RF output signal SRFO via the second quadrature combiner output SCO. The second phase-shifting circuitry **212** and the second Wilkinson RF combiner **214** may provide stable input impedances presented at the second in-phase input SII and the second quadrature-phase input SQI, respectively, which allows elimination of the second in-phase combiner impedance matching circuit **168** and the second quadrature-phase combiner impedance matching circuit **178**.

FIG. **34** shows details of the second feeder PA stage **122**, the second quadrature RF splitter **132**, the second in-phase final PA impedance matching circuit **164**, the second in-phase final PA stage **166**, the second quadrature-phase final PA impedance matching circuit **174**, and the second quadrature-phase final PA stage **176** illustrated in FIG. **33** according to one embodiment of the second feeder PA stage **122**, the second quadrature RF splitter **132**, the second in-phase final PA impedance matching circuit **164**, the second in-phase final PA stage **166**, the second quadrature-phase final PA impedance matching circuit **174**, and the second quadrature-phase final PA stage **176**. Further, FIG. **34** shows a portion of the second phase-shifting circuitry **212** illustrated in FIG. **33**.

The second in-phase final PA stage **166** includes a second in-phase final transistor element **216**, second in-phase biasing circuitry **218**, and a second in-phase collector inductive element LLI. The second quadrature-phase final PA stage **176** includes a second quadrature-phase final transistor element **220**, a second quadrature-phase biasing circuitry **222**, and a second quadrature-phase collector inductive element LLQ. The second in-phase final PA impedance matching circuit **164** includes a third in-phase series capacitive element CSI**3**, a fourth in-phase series capacitive element CSI**4**, and a second in-phase shunt inductive element LNI. The second quadrature-phase final PA impedance matching circuit **174** includes a third quadrature-phase series capacitive element CSQ**3**, a fourth quadrature-phase series capacitive element CSQ**4**, and a second quadrature-phase shunt inductive element LNQ.

The second quadrature RF splitter **132** includes a second pair **224** of tightly coupled inductors and a second isolation port resistive element RI**2**. The second pair **224** of tightly coupled inductors has second parasitic capacitance **226** between the second pair **224** of tightly coupled inductors. Additionally, the second quadrature RF splitter **132** has the second single-ended input SSI, the second in-phase output SIO, and the second quadrature-phase output SQO. The second feeder PA stage **122** includes a second output transistor element **228**, second feeder biasing circuitry **230**, a fifth DC blocking capacitive element CD**5**, a second base resistive element RB**2**, and a second collector inductive element LC**2**. Additionally, the second feeder PA stage **122** has the second single-ended output SSO.

The second output transistor element **228** shown is an NPN bipolar transistor element. Other embodiments of the second output transistor element **228** may use other types of transistor elements, such as field effect transistor elements (FET) elements. The fifth DC blocking capacitive element CD**5** is coupled between the second feeder PA impedance matching circuit **120** (FIG. **33**) and the second base resistive element RB**2**. A base of the second output transistor element **228** and the second feeder biasing circuitry **230** are coupled to the second base resistive element RB**2**. In alternate embodiments of the second feeder PA stage **122**, the second base resistive element RB**2**, the fifth DC blocking capacitive element CD**5**, or both may be omitted. The second feeder biasing circuitry **230** receives the second driver bias signal SDB. An emitter of the second output transistor element **228** is coupled to a ground. A collector of the second output transistor element **228** is coupled to the second single-ended output SSO. One end of the second collector inductive element LC**2** is coupled to the second single-ended output SSO. An opposite end of the second collector inductive element LC**2** receives the envelope power supply signal EPS. The second single-ended output SSO is coupled to the second single-ended input SSI.

During the second PA operating mode, the second output transistor element **228** receives and amplifies an RF signal from the second feeder PA impedance matching circuit **120** (FIG. **33**) via the fifth DC blocking capacitive element CD**5** and the second base resistive element RB**2** to provide the second RF feeder output signal SFO (FIG. **33**) to the second single-ended input SSI via the second single-ended output SSO. The envelope power supply signal EPS provides power for amplification via the second collector inductive element LC**2**. The second feeder biasing circuitry **230** biases the second output transistor element **228**. The second driver bias signal SDB provides power for biasing the second output transistor element **228** to the second feeder biasing circuitry **230**.

The second quadrature RF splitter **132** illustrated in FIG. **34** is a quadrature hybrid coupler. In this regard, the second pair **224** of tightly coupled inductors, the second parasitic capacitance **226**, and the second isolation port resistive element RI**2** provide quadrature hybrid coupler functionality. As such, the second single-ended input SSI functions as an input port to the quadrature hybrid coupler, the second in-phase output SIO functions as a zero degree output port from the quadrature hybrid coupler, and the second quadrature-phase output SQO functions as a 90 degree output port from the quadrature hybrid coupler. One of the second pair **224** of tightly coupled inductors is coupled between the second single-ended input SSI and the second in-phase output SIO. Another of the second pair **224** of tightly coupled inductors has a first end coupled to the second quadrature-phase output SQO and a second end coupled to the second isolation port resistive element RI**2**. As such, the second end functions as an isolation port of the quadrature hybrid coupler. In this regard, the second isolation port resistive element RI**2** is coupled between the isolation port and the ground. The second in-phase output SIO is coupled to the third in-phase series capacitive element CSI**3** and the second quadrature-phase output SQO is coupled to the third quadrature-phase series capacitive element CSQ**3**.

During the second PA operating mode, the second pair **224** of tightly coupled inductors receives, splits, and phase-shifts the second RF feeder output signal SFO (FIG. **33**) from the second single-ended output SSO via the second single-ended input SSI to provide split, phase-shifted output signals to the third in-phase series capacitive element CSI**3** and the third quadrature-phase series capacitive element CSQ**3**. As previously mentioned, the second input impedance is presented at the second single-ended input SSI. As such, the second input impedance is substantially based on the second parasitic capacitance **226** and inductances of the second pair **224** of tightly coupled inductors.

The third in-phase series capacitive element CSI**3** and the fourth in-phase series capacitive element CSI**4** are coupled in series between the second in-phase output SIO and a base of the second in-phase final transistor element **216**. The second in-phase shunt inductive element LNI is coupled between the ground and a junction between the third in-phase series capacitive element CSI**3** and the fourth in-phase series capacitive element CSI**4**. The third quadrature-phase series capacitive element CSQ**3** and the fourth quadrature-phase series capacitive element CSQ**4** are coupled in series between the second quadrature-phase output SQO and a base of the second quadrature-phase final transistor element **220**. The second quadrature-phase shunt inductive element LNQ is coupled between the ground and a junction between the third quadrature-phase series capacitive element CSQ**3** and the fourth quadrature-phase series capacitive element CSQ**4**.

The third in-phase series capacitive element CSI**3**, the fourth in-phase series capacitive element CSI**4**, and the second in-phase shunt inductive element LNI form a “T” network, which may provide at least an approximate impedance match between the second in-phase output SIO and the base of the second in-phase final transistor element **216**. Similarly, the third quadrature-phase series capacitive element CSQ**3**, the fourth quadrature-phase series capacitive element CSQ**4**, and the second quadrature-phase shunt inductive element LNQ form a “T” network, which may provide at least an approximate impedance match between the second quadrature-phase output SQO and the base of the second quadrature-phase final transistor element **220**.

During the second PA operating mode, the second in-phase final PA impedance matching circuit **164** receives and forwards an RF signal from the second in-phase output SIO to the base of the second in-phase final transistor element **216** via the third in-phase series capacitive element CSI**3** and the fourth in-phase series capacitive element CSI**4**. During the second PA operating mode, the second quadrature-phase final PA impedance matching circuit **174** receives and forwards an RF signal from the second quadrature-phase output SQO to the base of the second quadrature-phase final transistor element **220** via the third quadrature-phase series capacitive element CSQ**3** and the fourth quadrature-phase series capacitive element CSQ**4**. The second in-phase final transistor element **216** shown is an NPN bipolar transistor element. Other embodiments of the second in-phase final transistor element **216** may use other types of transistor elements, such as FET elements. The base of the second in-phase final transistor element **216** and the second in-phase biasing circuitry **218** are coupled to the fourth in-phase series capacitive element CSI**4**.

The second in-phase biasing circuitry **218** receives the second final bias signal SFB. An emitter of the second in-phase final transistor element **216** is coupled to the ground. A collector of the second in-phase final transistor element **216** is coupled to the second in-phase input SII. One end of the second in-phase collector inductive element LLI is coupled to the collector of the second in-phase final transistor element

216. An opposite end of the second in-phase collector inductive element LLI receives the envelope power supply signal EPS.

During the second PA operating mode, the second in-phase final transistor element **216** receives and amplifies an RF signal from the fourth in-phase series capacitive element CSI**4** to provide an RF output signal to the second in-phase input SII. The envelope power supply signal EPS provides power for amplification via the second in-phase collector inductive element LLI. The second in-phase biasing circuitry **218** biases the second in-phase final transistor element **216**. The second final bias signal SFB provides power for biasing the second in-phase final transistor element **216** to the second in-phase biasing circuitry **218**.

The second quadrature-phase final transistor element **220** shown is an NPN bipolar transistor element. Other embodiments of the second quadrature-phase final transistor element **220** may use other types of transistor elements, such as FET elements. The base of the second quadrature-phase final transistor element **220** and the second quadrature-phase biasing circuitry **222** are coupled to the fourth quadrature-phase series capacitive element CSQ**4**. The second quadrature-phase biasing circuitry **222** receives the second final bias signal SFB. An emitter of the second quadrature-phase final transistor element **220** is coupled to the ground. A collector of the second quadrature-phase final transistor element **220** is coupled to the second quadrature-phase input SQI. One end of the second quadrature-phase collector inductive element LLQ is coupled to the collector of the second quadrature-phase final transistor element **220**. An opposite end of the second quadrature-phase collector inductive element LLQ receives the envelope power supply signal EPS.

During the second PA operating mode, the second quadrature-phase final transistor element **220** receives and amplifies an RF signal from the fourth quadrature-phase series capacitive element CSQ**4** to provide an RF output signal to the second quadrature-phase input SQI. The envelope power supply signal EPS provides power for amplification via the second quadrature-phase collector inductive element LLQ. The second quadrature-phase biasing circuitry **222** biases the second quadrature-phase final transistor element **220**. The second final bias signal SFB provides power for biasing the second quadrature-phase final transistor element **220** to the second quadrature-phase biasing circuitry **222**.

In one embodiment of the RF PA circuitry **30** (FIG. **5**), the RF PA circuitry **30** includes a second PA semiconductor die **232**. In one embodiment of the second PA semiconductor die **232**, the second PA semiconductor die **232** includes the second output transistor element **228**, second in-phase final transistor element **216**, second in-phase biasing circuitry **218**, the second quadrature-phase final transistor element **220**, second quadrature-phase biasing circuitry **222**, the second pair **224** of tightly coupled inductors, the second feeder biasing circuitry **230**, the third in-phase series capacitive element CSI**3**, the fourth in-phase series capacitive element CSI**4**, the third quadrature-phase series capacitive element CSQ**3**, the fourth quadrature-phase series capacitive element CSQ**4**, the second isolation port resistive element RI**2**, the second base resistive element RB**2**, and the fifth DC blocking capacitive element CD**5**.

In alternate embodiments of the second PA semiconductor die **232**, the second PA semiconductor die **232** may not include any or all of the second output transistor element **228**, the second in-phase final transistor element **216**, the second in-phase biasing circuitry **218**, the second quadrature-phase final transistor element **220**, the second quadrature-phase biasing circuitry **222**, the second pair **224** of tightly coupled inductors, the second feeder biasing circuitry **230**, the third in-phase series capacitive element CSI**3**, the fourth in-phase series capacitive element CSI**4**, the third quadrature-phase series capacitive element CSQ**3**, the fourth quadrature-phase series capacitive element CSQ**4**, the second isolation port resistive element RI**2**, the second base resistive element RB**2**, and the fifth DC blocking capacitive element CD**5**.

FIG. **35** shows details of the second phase-shifting circuitry **212** and the second Wilkinson RF combiner **214** illustrated in FIG. **33** according to one embodiment of the second phase-shifting circuitry **212** and the second Wilkinson RF combiner **214**. The second phase-shifting circuitry **212** includes a second in-phase phase-shift capacitive element CPI**2**, a second quadrature-phase phase-shift capacitive element CPQ**2**, a second in-phase phase-shift inductive element LPI**2**, and a second quadrature-phase phase-shift inductive element LPQ**2**. The second Wilkinson RF combiner **214** includes a second Wilkinson resistive element RW**2**, a second Wilkinson capacitive element CW**2**, a second Wilkinson in-phase side capacitive element CWI**2**, a second Wilkinson quadrature-phase side capacitive element CWQ**2**, a second Wilkinson in-phase side inductive element LWI**2**, a second Wilkinson quadrature-phase side inductive element LWQ**2**, a sixth DC blocking capacitive element CD**6**, a seventh DC blocking capacitive element CD**7**, and a eighth DC blocking capacitive element CD**8**.

The second in-phase phase-shift capacitive element CPI**2** is coupled between the second in-phase input SII and a third internal node (not shown). The second in-phase phase-shift inductive element LPI**2** is coupled between the third internal node and the ground. The second quadrature-phase phase-shift inductive element LPQ**2** is coupled between the second quadrature-phase input SQI and a fourth internal node (not shown). The second quadrature-phase phase-shift capacitive element CPQ**2** is coupled between the fourth internal node and the ground. The sixth DC blocking capacitive element CD**6** and the second Wilkinson resistive element RW**2** are coupled in series between the third internal node and the fourth internal node. The second Wilkinson in-phase side capacitive element CWI**2** is coupled between the third internal node and the ground. The second Wilkinson quadrature-phase side capacitive element CWQ**2** is coupled between the third internal node and the ground. The second Wilkinson in-phase side inductive element LWI**2** is coupled in series with the seventh DC blocking capacitive element CD**7** between the third internal node and the second quadrature combiner output SCO. The second Wilkinson quadrature-phase side inductive element LWQ**2** is coupled in series with the eighth DC blocking capacitive element CD**8** between the fourth internal node and the second quadrature combiner output SCO. The second Wilkinson capacitive element CW**2** is coupled between the second quadrature combiner output SCO and the ground.

FIG. **36** shows details of the first PA semiconductor die **210** illustrated in FIG. **30** according to one embodiment of the first PA semiconductor die **210**. The first PA semiconductor die **210** includes a first substrate and functional layers **234**, multiple insulating layers **236**, and multiple metallization layers **238**. Some of the insulating layers **236** may be used to separate some of the metallization layers **238** from one another. In one embodiment of the metallization layers **238**, each of the metallization layers **238** is about parallel to at least another of the metallization layers **238**. In this regard the metallization layers **238** may be planar. In an alternate embodiment of the metallization layers **238**, the metallization layers **238** are formed over a non-planar structure, such that spacing between pairs of the metallization layers **238** is about constant. In one embodiment of the metallization layers **238**, each of the first pair **204** of tightly coupled inductors (FIG. **30**) is constructed using at least one of the metallization layers **238**.

Linear Mode and Non-Linear Mode Quadrature PA Circuitry

A summary of linear mode and non-linear mode quadrature PA circuitry is presented, followed by a detailed description of the linear mode and non-linear mode quadrature PA circuitry according to one embodiment of the present disclosure. Multi-mode multi-band RF PA circuitry includes a multi-mode multi-band quadrature RF PA coupled to multi-mode multi-band switching circuitry via a single output. The switching circuitry provides at least one non-linear mode output and multiple linear mode outputs. The non-linear mode output may be associated with at least one non-linear mode RF communications band and each linear mode output may be associated with a corresponding linear mode RF communications band. The outputs from the switching circuitry may be coupled to an antenna port via front-end aggregation circuitry. The quadrature nature of the quadrature PA path may provide tolerance for changes in antenna loading conditions.

One embodiment of the RF PA circuitry includes a highband multi-mode multi-band quadrature RF PA coupled to highband multi-mode multi-band switching circuitry and a lowband multi-mode multi-band quadrature RF PA coupled to lowband multi-mode multi-band switching circuitry. The highband switching circuitry may be associated with at least one highband non-linear mode RF communications band and multiple highband linear mode RF communications bands. The lowband switching circuitry may be associated with at least one lowband non-linear mode RF communications band and multiple lowband linear mode RF communications bands.

FIG. **37** shows details of the RF PA circuitry **30** illustrated in FIG. **5** according to one embodiment of the RF PA circuitry **30**. The RF PA circuitry **30** illustrated in FIG. **37** is similar to the RF PA circuitry **30** illustrated in FIG. **8**, except in the RF PA circuitry **30** illustrated in FIG. **37**, the first RF PA **50** is a first multi-mode multi-band quadrature RF PA; the second RF PA **54** is a second multi-mode multi-band quadrature RF PA; the alpha switching circuitry **52** is multi-mode multi-band RF switching circuitry; the first RF PA **50** includes a single alpha PA output SAP; the second RF PA **54** includes a single beta PA output SBP; the alpha switching circuitry **52** further includes a first alpha non-linear mode output FANO, a first alpha linear mode output FALO, and up to and including an R^{TH} alpha linear mode output RALO; and the beta switching circuitry **56** further includes a first beta non-linear mode output FBNO, a first beta linear mode output FBLO, and up to and including an S^{TH} beta linear mode output SBLO. In general, the alpha switching circuitry **52** includes a group of alpha linear mode outputs FALO, RALO and the beta switching circuitry **56** includes a group of beta linear mode outputs FBLO, SBLO.

The first RF PA **50** is coupled to the alpha switching circuitry **52** via the single alpha PA output SAP. The second RF PA **54** is coupled to the beta switching circuitry **56** via the single beta PA output SBP. In one embodiment of the first RF PA **50**, the single alpha PA output SAP is a single-ended output. In one embodiment of the second RF PA **54**, the single beta PA output SBP is a single-ended output. In one embodiment of the alpha switching circuitry **52**, the first alpha non-linear mode output FANO is associated with a first non-linear mode RF communications band and each of the group of alpha linear mode outputs FALO, RALO is associated with a corresponding one of a first group of linear mode RF communications bands. In one embodiment of the beta switching circuitry **56**, the first beta non-linear mode output FBNO is associated with a second non-linear mode RF communications band and each of the group of beta linear mode outputs FBLO, SBLO is associated with a corresponding one of a second group of linear mode RF communications bands.

In an alternate embodiment of the alpha switching circuitry **52**, the first alpha non-linear mode output FANO is associated with a first group of non-linear mode RF communications bands, which includes the first non-linear mode RF communications band. In an alternate embodiment of the beta switching circuitry **56**, the first beta non-linear mode output FBNO is associated with a second group of non-linear mode RF communications bands, which includes the second non-linear mode RF communications band.

In one embodiment of the RF communications system **26** (FIG. **5**), the RF communications system **26** operates in one of a group of communications modes. Control circuitry, which may include the control circuitry **42** (FIG. **5**), the PA control circuitry **94** (FIG. **13**), or both, selects one of the group of communications modes. In one embodiment of the RF communications system **26**, the group of communications modes includes a first alpha non-linear mode and a group of alpha linear modes. In an alternate embodiment of the RF communications system **26**, the group of communications modes includes the first alpha non-linear mode, the group of alpha linear modes, a first beta non-linear mode, and a group of beta non-linear modes. In an additional embodiment of the RF communications system **26**, the group of communications modes includes a group of alpha non-linear modes, the group of alpha linear modes, a group of beta non-linear modes, and the group of beta non-linear modes. Other embodiments of the RF communications system **26** may omit any or all of the communications modes. In one embodiment of the first alpha non-linear mode, the first alpha non-linear mode is a half-duplex mode. In one embodiment of the first beta non-linear mode, the beta alpha non-linear mode is a half-duplex mode. In one embodiment of the group of alpha linear modes, each of the group of alpha linear modes is a full-duplex mode. In one embodiment of the group of beta linear modes, each of the group of beta linear modes is a full-duplex mode.

In one embodiment of the first RF PA **50**, during the first alpha non-linear mode and during each of the group of alpha linear modes, the first RF PA **50** receives and amplifies the first RF input signal FRFI to provide the first RF output signal FRFO via the single alpha PA output SAP. Further, during the first beta non-linear mode and during each of the group of beta linear modes, the first RF PA **50** does not receive or amplify the first RF input signal FRFI to provide the first RF output signal FRFO.

In one embodiment of the second RF PA **54**, during the first beta non-linear mode and during each of the group of beta linear modes, the second RF PA **54** receives and amplifies the second RF input signal SRFI to provide the second RF output signal SRFO via the single beta PA output SBP. Further, during the first alpha non-linear mode and during each of the group of alpha linear modes, the second RF PA **54** does not receive or amplify the second RF input signal SRFI to provide the second RF output signal SRFO.

In one embodiment of the alpha switching circuitry **52**, during the first alpha non-linear mode, the alpha switching circuitry **52** receives and forwards the first RF output signal FRFO to provide the first alpha RF transmit signal FATX via the first alpha non-linear mode output FANO. During a first alpha linear mode, the alpha switching circuitry **52** receives and forwards the first RF output signal FRFO to provide the second alpha RF transmit signal SATX via the first alpha linear mode output FALO. During an R^{TH} alpha linear mode, the alpha switching circuitry **52** receives and forwards the first RF output signal FRFO to provide the P^{TH} alpha RF transmit signal PATX. In general, during each of the group of alpha linear modes, the alpha switching circuitry **52** receives and forwards the first RF output signal FRFO to provide a corresponding one of a group of alpha RF transmit signals SATX, PATX via a corresponding one of the group of alpha linear mode outputs FALO, RALO.

In one embodiment of the beta switching circuitry **56**, during the first beta non-linear mode, the beta switching circuitry **56** receives and forwards the second RF output signal SRFO to provide the first beta RF transmit signal FBTX via the first beta non-linear mode output FBNO. During a first beta linear mode, the beta switching circuitry **56** receives and forwards the second RF output signal SRFO to provide the second beta RF transmit signal SBTX via the first beta linear mode output FBLO. During an S^{TH} beta linear mode, the beta switching circuitry **56** receives and forwards the second RF output signal SRFO to provide the Q^{TH} beta RF transmit signal QBTX. In general, during each of the group of beta linear modes, the beta switching circuitry **56** receives and forwards the second RF output signal SRFO to provide a corresponding one of a group of beta RF transmit signals SBTX, QBTX via a corresponding one of the group of beta linear mode outputs FBLO, SBLO.

FIG. **38** shows details of the RF PA circuitry **30** illustrated in FIG. **5** according to an alternate embodiment of the RF PA circuitry **30**. The RF PA circuitry **30** illustrated in FIG. **38** is similar to the RF PA circuitry **30** illustrated in FIG. **9**, except in the RF PA circuitry **30** illustrated in FIG. **38**, the first RF PA **50** is the first multi-mode multi-band quadrature RF PA; the second RF PA **54** is the second multi-mode multi-band quadrature RF PA; the alpha switching circuitry **52** is multi-mode multi-band RF switching circuitry; the first RF PA **50** includes the single alpha PA output SAP; the second RF PA **54** includes the single beta PA output SBP; the alpha switching circuitry **52** further includes the first alpha non-linear mode output FANO, a second alpha non-linear mode output SANO, the first alpha linear mode output FALO, and up to and including the R^{TH} alpha linear mode output RALO; and the beta switching circuitry **56** further includes the first beta non-linear mode output FBNO, a second beta non-linear mode output SBNO, the first beta linear mode output FBLO, and up to and including the S^{TH} beta linear mode output SBLO. In general, the alpha switching circuitry **52** includes the group of alpha linear mode outputs FALO, RALO and the beta switching circuitry **56** includes the group of beta linear mode outputs FBLO, SBLO. Additionally, in general, the alpha switching circuitry **52** includes at least the first alpha harmonic filter **70** and the beta switching circuitry **56** includes at least the first beta harmonic filter **74**.

Dual-Path PA Circuitry with Harmonic Filters

A summary of dual-path PA circuitry with harmonic filters is presented, followed by a detailed description of the dual-path PA circuitry with harmonic filters according to one embodiment of the present disclosure. The dual-path PA circuitry includes a first transmit path and a second transmit path. Each transmit path has an RF PA and switching circuitry having at least one harmonic filter. Each RF PA may be coupled to its corresponding switching circuitry via a single output. Each switching circuitry provides at least one output via a harmonic filter and multiple outputs without harmonic filtering. The output via the harmonic filter may be a non-linear mode output and the outputs without harmonic filtering may be linear mode outputs. The non-linear mode output may be associated with at least one non-linear mode RF communications band and the linear mode outputs may be associated with multiple linear mode RF communications bands. As such, each RF PA may be a multi-mode multi-band RF PA.

The outputs from the switching circuitry may be coupled to an antenna port via front-end aggregation circuitry. The quadrature nature of the quadrature PA path may provide tolerance for changes in antenna loading conditions. One embodiment of the RF PA circuitry includes a highband multi-mode multi-band quadrature RF PA coupled to highband multi-mode multi-band switching circuitry and a lowband multi-mode multi-band quadrature RF PA coupled to lowband multi-mode multi-band switching circuitry. The highband switching circuitry may be associated with at least one highband non-linear mode RF communications band and multiple highband linear mode RF communications bands. The lowband switching circuitry may be associated with at least one lowband non-linear mode RF communications band and multiple lowband linear mode RF communications bands.

In one embodiment of the RF PA circuitry **30**, the first alpha non-linear mode output FANO is a first alpha output, the second alpha non-linear mode output SANO is a second alpha output, the first beta non-linear mode output FBNO is a first beta output, the second beta non-linear mode output SBNO is a second beta output, the group of alpha linear mode outputs FALO, RALO is a group of alpha outputs, and the group of beta linear mode outputs FBLO, SBLO is a group of beta outputs. The alpha switching circuitry **52** provides the first alpha output via the first alpha harmonic filter **70**. The alpha switching circuitry **52** provides the second alpha output via the second alpha harmonic filter **76**. The alpha switching circuitry **52** provides the group of alpha outputs without harmonic filtering. The beta switching circuitry **56** provides the first beta output via the first beta harmonic filter **74**. The beta switching circuitry **56** provides the second beta output via the second beta harmonic filter **78**. The beta switching circuitry **56** provides the group of beta outputs without harmonic filtering.

In one embodiment of the RF communications system **26** (FIG. **5**), the RF communications system **26** operates in one of a group of communications modes. Control circuitry, which may include the control circuitry **42** (FIG. **5**), the PA control circuitry **94** (FIG. **13**), or both, selects one of the group of communications modes. In one embodiment of the RF communications system **26**, the group of communications modes includes the first alpha non-linear mode, the group of alpha linear modes, the first beta non-linear mode, and the group of beta non-linear modes. Other embodiments of the RF communications system **26** may omit any or all of the communications modes. In one embodiment of the first alpha non-linear mode, the first alpha non-linear mode is a half-duplex mode. In one embodiment of the first beta non-linear mode, the beta alpha non-linear mode is a half-duplex mode. In one embodiment of the group of alpha linear modes, each of the group of alpha linear modes is a full-duplex mode. In one embodiment of the group of beta linear modes, each of the group of beta linear modes is a full-duplex mode.

In one embodiment of the first RF PA **50**, during the first alpha non-linear mode and during each of the group of alpha linear modes, the first RF PA **50** receives and amplifies the first RF input signal FRFI to provide the first RF output signal FRFO via the single alpha PA output SAP. Further, during the first beta non-linear mode and during each of the group of beta linear modes, the first RF PA **50** does not receive or amplify the first RF input signal FRFI to provide the first RF output signal FRFO.

In one embodiment of the second RF PA **54**, during the first beta non-linear mode and during each of the group of beta linear modes, the second RF PA **54** receives and amplifies the second RF input signal SRFI to provide the second RF output signal SRFO via the single beta PA output SBP. Further, during the first alpha non-linear mode and during each of the group of alpha linear modes, the second RF PA **54** does not receive or amplify the second RF input signal SRFI to provide the second RF output signal SRFO.

In one embodiment of the alpha switching circuitry **52**, during the first alpha non-linear mode, the alpha switching circuitry **52** receives and forwards the first RF output signal FRFO to provide the first alpha RF transmit signal FATX via the first alpha harmonic filter **70** and the first alpha output. During each of the group of alpha linear modes, the alpha switching circuitry **52** receives and forwards the first RF output signal FRFO to provide a corresponding one of a group of alpha RF transmit signals TATX, PATX via a corresponding one of the group of alpha outputs.

In one embodiment of the beta switching circuitry **56**, during the first beta non-linear mode, the beta switching circuitry **56** receives and forwards the second RF output signal SRFO to provide the first beta RF transmit signal FBTX via the first beta harmonic filter **74** and the first beta output. During each of the group of beta linear modes, the beta switching circuitry **56** receives and forwards the second RF output signal SRFO to provide a corresponding one of a group of beta RF transmit signals TBTX, QBTX via a corresponding one of the group of beta outputs.

FIG. **39** shows details of the RF PA circuitry **30** illustrated in FIG. **5** according to an additional embodiment of the RF PA circuitry **30**. The RF PA circuitry **30** illustrated in FIG. **39** is similar to the RF PA circuitry **30** illustrated in FIG. **37**, except the RF PA circuitry **30** illustrated in FIG. **39** further includes the switch driver circuitry **98** (FIG. **13**) and shows details of the alpha RF switch **68** and the beta RF switch **72**. The alpha RF switch **68** includes a first alpha switching device **240**, a second alpha switching device **242**, and a third alpha switching device **244**. The beta RF switch **72** includes a first beta switching device **246**, a second beta switching device **248**, and a third beta switching device **250**. Alternate embodiments of the alpha RF switch **68** may include any number of alpha switching devices. Alternate embodiments of the beta RF switch **72** may include any number of beta switching devices.

The first alpha switching device **240** is coupled between the single alpha PA output SAP and the first alpha harmonic filter **70**. As such, the first alpha switching device **240** is coupled between the single alpha PA output SAP and the first alpha non-linear mode output FANO via the first alpha harmonic filter **70**. The second alpha switching device **242** is coupled between the single alpha PA output SAP and the first alpha linear mode output FALO. The third alpha switching device **244** is coupled between the single alpha PA output SAP and the R^{TH} alpha linear mode output RALO. In general, the alpha RF switch **68** includes the first alpha switching device **240** and a group of alpha switching devices, which includes the second alpha switching device **242** and the third alpha switching device **244**. As previously mentioned, the alpha switching circuitry **52** includes the group of alpha linear mode outputs FALO, RALO. As such, each of the group of alpha switching devices **242**, **244** is coupled between the single alpha PA output SAP and a corresponding one of the group of alpha linear mode outputs FALO, RALO. Additionally, each of the alpha switching devices **240**, **242**, **244** has a corresponding control input, which is coupled to the switch driver circuitry **98**.

The first beta switching device **246** is coupled between the single beta PA output SBP and the first beta harmonic filter **74**. As such, the first beta switching device **246** is coupled between the single beta PA output SBP and the first beta non-linear mode output FBNO via the first beta harmonic filter **74**. The second beta switching device **248** is coupled between the single beta PA output SBP and the first beta linear mode output FBLO. The third beta switching device **250** is coupled between the single beta PA output SBP and the S^{TH} beta linear mode output SBLO. In general, the beta RF switch **72** includes the first beta switching device **246** and a group of beta switching devices, which includes the second beta switching device **248** and the third beta switching device **250**. As previously mentioned, the beta switching circuitry **56** includes the group of beta linear mode outputs FBLO, SBLO. As such, each of the group of beta switching devices **248**, **250** is coupled between the single beta PA output SBP and a corresponding one of the group of beta linear mode outputs FBLO, SBLO. Additionally, each of the beta switching devices **246**, **248**, **250** has a corresponding control input, which is coupled to the switch driver circuitry **98**.

In one embodiment of the alpha RF switch **68**, the first alpha switching device **240** includes multiple switching elements (not shown) coupled in series. Each of the group of alpha switching devices **242**, **244** includes multiple switching elements (not shown) coupled in series. In one embodiment of the beta RF switch **72**, the first beta switching device **246** includes multiple switching elements (not shown) coupled in series. Each of the group of beta switching devices **248**, **250** includes multiple switching elements (not shown) coupled in series.

PA Bias Supply Using Boosted Voltage

A summary of a PA bias supply using boosted voltage is presented, followed by a detailed description of the PA bias supply using boosted voltage according to one embodiment of the present disclosure. An RF PA bias power supply signal is provided to RF PA circuitry by boosting a voltage from a DC power supply, such as a battery. In this regard, a DC-DC converter receives a DC power supply signal from the DC power supply. The DC-DC converter provides the bias power supply signal based on the DC power supply signal, such that a voltage of the bias power supply signal is greater than a voltage of the DC power supply signal. The RF PA circuitry has an RF PA, which has a final stage that receives a final bias signal to bias the final stage, such that the final bias signal is based on the bias power supply signal. Boosting the voltage from the DC power supply may provide greater flexibility in biasing the RF PA.

In one embodiment of the DC-DC converter, the DC-DC converter includes a charge pump, which may receive and pump-up the DC power supply signal to provide the bias power supply signal. Further, the DC-DC converter may operate in one of a bias supply pump-up operating mode and at least one other operating mode, which may include any or all of a bias supply pump-even operating mode, a bias supply pump-down operating mode, and a bias supply bypass operating mode. Additionally, the DC-DC converter provides an envelope power supply signal to the RF PA, which uses the envelope power supply signal to provide power for amplification. In one embodiment of the RF PA circuitry, the RF PA circuitry includes PA bias circuitry, which receives the bias power supply signal to provide the final bias signal. The PA bias circuitry may include a final stage current analog-to-digital converter (IDAC) to receive and use the bias power supply signal in a digital-to-analog conversion to provide the final bias signal.

In an alternate embodiment of the RF PA circuitry, the RF PA circuitry includes a first RF PA and a second RF PA, which include a first final stage and a second final stage, respectively. The first RF PA may be used to receive and amplify a highband RF input signal and the second RF PA may be used to receive and amplify a lowband RF input signal. The RF PA circuitry operates in one of a first PA operating mode and a second PA operating mode, such that during the first PA operating mode, the first RF PA is active and the second RF PA is disabled. Conversely, during the second PA operating mode, the first RF PA is disabled and the second RF PA is active. The PA bias circuitry may include the final stage IDAC and a final stage multiplexer. The final stage IDAC receives and uses the bias power supply signal in a digital-to-analog conversion to provide a final stage bias signal to the final stage multiplexer. During the first PA operating mode, the final stage multiplexer receives and forwards the final stage bias signal to provide a first final bias signal to the first RF PA to bias the first final stage. During the second PA operating mode, the final stage multiplexer receives and forwards the final stage bias signal to provide a second final bias signal to the second RF PA to bias the second final stage.

FIG. **40** shows details of the first RF PA **50**, the second RF PA **54**, and the PA bias circuitry **96** illustrated in FIG. **13** according to one embodiment of the first RF PA **50**, the second RF PA **54**, and the PA bias circuitry **96**. The first RF PA **50** includes a first driver stage **252** and a first final stage **254**. The second RF PA **54** includes a second driver stage **256** and a second final stage **258**. The PA bias circuitry **96** includes driver stage IDAC circuitry **260** and final stage IDAC circuitry **262**. In general, the first RF PA **50** receives and amplifies the first RF input signal FRFI to provide the first RF output signal FRFO. Similarly, the second RF PA **54** receives and amplifies the second RF input signal SRFI to provide the second RF output signal SRFO. Specifically, the first driver stage **252** receives and amplifies the first RF input signal FRFI to provide a first final stage input signal FFSI, and the first final stage **254** receives and amplifies the first final stage input signal FFSI to provide the first RF output signal FRFO. Similarly, the second driver stage **256** receives and amplifies the second RF input signal SRFI to provide a second final stage input signal SFSI, and the second final stage **258** receives and amplifies the second final stage input signal SFSI to provide the second RF output signal SRFO.

The first driver stage **252** receives the envelope power supply signal EPS, which provides power for amplification; the first final stage **254** receives the envelope power supply signal EPS, which provides power for amplification; the second driver stage **256** receives the envelope power supply signal EPS, which provides power for amplification; and the second final stage **258** receives the envelope power supply signal EPS, which provides power for amplification. In general, the first RF PA **50** receives the first driver bias signal FDB to bias first driver stage **252** and receives the first final bias signal FFB to bias the first final stage **254**. Specifically, the first driver stage **252** receives the first driver bias signal FDB to bias the first driver stage **252** and the first final stage **254** receives the first final bias signal FFB to bias the first final stage **254**. Similarly, the second RF PA **54** receives the second driver bias signal SDB to bias the second driver stage **256** and receives the second final bias signal SFB to bias the second final stage **258**. Specifically, the second driver stage **256** receives the second driver bias signal SDB to bias the second driver stage **256** and the second final stage **258** receives the second final bias signal SFB to bias the second final stage **258**.

In general, the PA bias circuitry **96** provides the first driver bias signal FDB based on the bias power supply signal BPS, the first final bias signal FFB based on the bias power supply signal BPS, the second driver bias signal SDB based on the bias power supply signal BPS, and the second final bias signal SFB based on the bias power supply signal BPS. Specifically, the driver stage IDAC circuitry **260** provides the first driver bias signal FDB based on the bias power supply signal BPS and provides the second driver bias signal SDB based on the bias power supply signal BPS. Similarly, the final stage IDAC circuitry **262** provides the first final bias signal FFB based on the bias power supply signal BPS and provides the second final bias signal SFB based on the bias power supply signal BPS.

In one embodiment of the driver stage IDAC circuitry **260** and the final stage IDAC circuitry **262**, the driver stage IDAC circuitry **260** and the final stage IDAC circuitry **262** receive the bias power supply signal BPS and the bias configuration control signal BCC. The driver stage IDAC circuitry **260** provides the first driver bias signal FDB and the second driver bias signal SDB based on the bias power supply signal BPS and the bias configuration control signal BCC. The final stage IDAC circuitry **262** provides the first final bias signal FFB and the second final bias signal SFB based on the bias power supply signal BPS and the bias configuration control signal BCC. The bias power supply signal BPS provides the power necessary to generate the bias signals FDB, FFB, SDB, SFB. A selected magnitude of each of the bias signals FDB, FFB, SDB, SFB is provided by the driver stage IDAC circuitry **260** and the final stage IDAC circuitry **262**. In one embodiment of the RF PA circuitry **30**, the PA control circuitry **94** selects the magnitude of any or all of the bias signals FDB, FFB, SDB, SFB and communicates the magnitude selections to the driver stage IDAC circuitry **260** and the final stage IDAC circuitry **262** via the bias configuration control signal BCC. The magnitude selections by the PA control circuitry **94** may be based on the PA configuration control signal PCC. In an alternate embodiment of the RF PA circuitry **30**, the control circuitry **42** (FIG. **5**) selects the magnitude of any or all of the bias signals FDB, FFB, SDB, SFB and communicates the magnitude selections to the driver stage IDAC circuitry **260** and the final stage IDAC circuitry **262** via the PA control circuitry **94**.

As previously discussed, in one embodiment of the RF PA circuitry **30**, the RF PA circuitry **30** operates in one of the first PA operating mode and the second PA operating mode. During the first PA operating mode, the first RF PA **50** receives and amplifies the first RF input signal FRFI to provide the first RF output signal FRFO, and the second RF PA **54** is disabled. During the second PA operating mode, the second RF PA **54** receives and amplifies the second RF input signal SRFI to provide the second RF output signal SRFO, and the first RF PA **50** is disabled.

In one embodiment of the first RF PA **50**, during the second PA operating mode, the first RF PA **50** is disabled via the first driver bias signal FDB. As such, the first driver stage **252** is disabled. In an alternate embodiment of the first RF PA **50**, during the second PA operating mode, the first RF PA **50** is disabled via the first final bias signal FFB. As such, the first final stage **254** is disabled. In an additional embodiment of the first RF PA **50**, during the second PA operating mode, the first RF PA **50** is disabled via both the first driver bias signal FDB and the first final bias signal FFB. As such, both the first driver stage **252** and the first final stage **254** are disabled.

In one embodiment of the second RF PA **54**, during the first PA operating mode, the second RF PA **54** is disabled via the second driver bias signal SDB. As such, the second driver stage **256** is disabled. In an alternate embodiment of the second RF PA **54**, during the first PA operating mode, the second RF PA **54** is disabled via the second final bias signal SFB. As such, the second final stage **258** is disabled. In an additional embodiment of the second RF PA **54**, during the first PA operating mode, the second RF PA **54** is disabled via both the second driver bias signal SDB and the second final bias signal SFB. As such, both the second driver stage **256** and the second final stage **258** are disabled.

In one embodiment of the RF PA circuitry **30**, the PA control circuitry **94** selects the one of the first PA operating mode and the second PA operating mode. As such, the PA control circuitry **94** may control any or all of the bias signals FDB, FFB, SDB, SFB via the bias configuration control signal BCC based on the PA operating mode selection. The PA operating mode selection may be based on the PA configuration control signal PCC. In an alternate embodiment of the RF PA circuitry **30**, the control circuitry **42** (FIG. **5**) selects the one of the first PA operating mode and the second PA operating mode. As such, the control circuitry **42** (FIG. **5**) may indicate the operating mode selection to the PA control circuitry **94** via the PA configuration control signal PCC. In an additional embodiment of the RF PA circuitry **30**, the RF modulation and control circuitry **28** (FIG. **5**) selects the one of the first PA operating mode and the second PA operating mode. As such, the RF modulation and control circuitry **28** (FIG. **5**) may indicate the operating mode selection to the PA control circuitry **94** via the PA configuration control signal PCC. In general, selection of the PA operating mode is made by control circuitry, which may be any of the PA control circuitry **94**, the RF modulation and control circuitry **28** (FIG. **5**), and the control circuitry **42** (FIG. **5**).

Further, during the first PA operating mode, the control circuitry selects a desired magnitude of the first driver bias signal FDB, a desired magnitude of the first final bias signal FFB, or both. During the second PA operating mode, the control circuitry selects a desired magnitude of the second driver bias signal SDB, a desired magnitude of the second final bias signal SFB, or both As such, during the first PA operating mode, the PA control circuitry **94** provides the bias configuration control signal BCC to the PA bias circuitry **96** in general and to the driver stage IDAC circuitry **260** in particular based on the desired magnitude of the first driver bias signal FDB, and the PA control circuitry **94** provides the bias configuration control signal BCC to the PA bias circuitry **96** in general and to the final stage IDAC circuitry **262** in particular based on the desired magnitude of the first final bias signal FFB. During the second PA operating mode, the PA control circuitry **94** provides the bias configuration control signal BCC to the PA bias circuitry **96** in general and to the driver stage IDAC circuitry **260** in particular based on the desired magnitude of the second driver bias signal SDB, and the PA control circuitry **94** provides the bias configuration control signal BCC to the PA bias circuitry **96** in general and to the final stage IDAC circuitry **262** in particular based on the desired magnitude of the second final bias signal SFB. In one embodiment of the PA control circuitry **94**, the bias configuration control signal BCC is a digital signal.

FIG. **41** shows details of the driver stage IDAC circuitry **260** and the final stage IDAC circuitry **262** illustrated in FIG. **40** according to one embodiment of the driver stage IDAC circuitry **260** and the final stage IDAC circuitry **262**. The driver stage IDAC circuitry **260** includes a driver stage IDAC **264**, a driver stage multiplexer **266**, and driver stage current reference circuitry **268**. The final stage IDAC circuitry **262** includes a final stage IDAC **270**, a final stage multiplexer **272**, and final stage current reference circuitry **274**.

The driver stage IDAC **264** receives the bias power supply signal BPS, the bias configuration control signal BCC, and a driver stage reference current IDSR. As such, the driver stage IDAC **264** uses the bias power supply signal BPS and the driver stage reference current IDSR in a digital-to-analog conversion to provide a driver stage bias signal DSBS. A magnitude of the digital-to-analog conversion is based on the bias configuration control signal BCC. The driver stage current reference circuitry **268** is coupled to the driver stage IDAC **264** and provides the driver stage reference current IDSR to the driver stage IDAC **264**, such that during the first PA operating mode, the first driver bias signal FDB is based on the driver stage reference current IDSR, and during the second PA operating mode, the second driver bias signal SDB is based on the driver stage reference current IDSR. The driver stage current reference circuitry **268** may be disabled based on the bias configuration control signal BCC. The driver stage current reference circuitry **268** and the driver stage multiplexer **266** receive the bias configuration control signal BCC. The driver stage multiplexer **266** receives and forwards the driver stage bias signal DSBS, which is a current signal, to provide either the second driver bias signal SDB or the first driver bias signal FDB based on the bias configuration control signal BCC. During the first PA operating mode, the driver stage multiplexer **266** receives and forwards the driver stage bias signal DSBS to provide the first driver bias signal FDB based on the bias configuration control signal BCC. During the second PA operating mode, the driver stage multiplexer **266** receives and forwards the driver stage bias signal DSBS to provide the second driver bias signal SDB based on the bias configuration control signal BCC.

In this regard, during the first PA operating mode, the driver stage IDAC **264** provides the first driver bias signal FDB via the driver stage multiplexer **266**, such that a magnitude of the first driver bias signal FDB is about equal to the desired magnitude of the first driver bias signal FDB. During the second PA operating mode, the driver stage IDAC **264** provides the second driver bias signal SDB via the driver stage multiplexer **266**, such that a magnitude of the second driver bias signal SDB is about equal to the desired magnitude of the second driver bias signal SDB.

In one embodiment of the driver stage multiplexer **266**, during the first PA operating mode, the driver stage multiplexer **266** disables the second RF PA **54** via the second driver bias signal SDB. In one embodiment of the second RF PA **54**, the second RF PA **54** is disabled when the second driver bias signal SDB is about zero volts. In one embodiment of the driver stage multiplexer **266**, during the second PA operating mode, the driver stage multiplexer **266** disables the first RF PA **50** via the first driver bias signal FDB. In one embodiment of the first RF PA **50**, the first RF PA **50** is disabled when the first driver bias signal FDB is about zero volts. As such, in one embodiment of the driver stage multiplexer **266**, during the first PA operating mode, the driver stage multiplexer **266** provides the second driver bias signal SDB, which is about zero volts, such that the second RF PA **54** is disabled, and during the second PA operating mode, the driver stage multiplexer **266** provides the first driver bias signal FDB, which is about zero volts, such that the first RF PA **50** is disabled.

The final stage IDAC **270** receives the bias power supply signal BPS, the bias configuration control signal BCC, and a final stage reference current IFSR. As such, the final stage IDAC **270** uses the bias power supply signal BPS and the final stage reference current IFSR in a digital-to-analog conversion to provide a final stage bias signal FSBS. A magnitude of the digital-to-analog conversion is based on the bias configuration control signal BCC. The final stage current reference circuitry **274** is coupled to the final stage IDAC **270** and provides the final stage reference current IFSR to the final stage IDAC **270**, such that during the first PA operating mode, the first final bias signal FFB is based on the final stage reference current IFSR, and during the second PA operating mode, the second final bias signal SFB is based on the final stage reference current IFSR. The final stage current reference circuitry **274** and the final stage IDAC **270** receive the bias configuration control signal BCC. The final stage current reference circuitry **274** may be disabled based on the bias configuration control signal BCC. The final stage multiplexer **272** receives and forwards the final stage bias signal FSBS, which is a current signal, to provide either the second final bias signal SFB or the first final bias signal FFB based on the bias configuration control signal BCC. During the first PA operating mode, the final stage multiplexer **272** receives and forwards the final stage bias signal FSBS to provide the first final bias signal FFB based on the bias configuration control signal BCC. During the second PA operating mode, the final stage multiplexer **272** receives and forwards the final stage bias signal FSBS to provide the second final bias signal SFB based on the bias configuration control signal BCC.

In this regard, during the first PA operating mode, the final stage IDAC **270** provides the first final bias signal FFB via the final stage multiplexer **272**, such that a magnitude of the first final bias signal FFB is about equal to the desired magnitude of the first final bias signal FFB. Specifically, the final stage IDAC **270** receives and uses the bias power supply signal BPS and the bias configuration control signal BCC in a digital-to-analog conversion to provide the first final bias signal FFB. During the second PA operating mode, the final stage IDAC **270** provides the second final bias signal SFB via the final stage multiplexer **272**, such that a magnitude of the second final bias signal SFB is about equal to the desired magnitude of the second final bias signal SFB. Specifically, the final stage IDAC **270** receives and uses the bias power supply signal BPS and the bias configuration control signal BCC in a digital-to-analog conversion to provide the second final bias signal SFB.

In one embodiment of the final stage multiplexer **272**, during the first PA operating mode, the final stage multiplexer **272** disables the second RF PA **54** via the second final bias signal SFB. In one embodiment of the second RF PA **54**, the second RF PA **54** is disabled when the second final bias signal SFB is about zero volts. In one embodiment of the final stage multiplexer **272**, during the second PA operating mode, the final stage multiplexer **272** disables the first RF PA **50** via the first final bias signal FFB. In one embodiment of the first RF PA **50**, the first RF PA **50** is disabled when the first final bias signal FFB is about zero volts. As such, in one embodiment of the final stage multiplexer **272**, during the first PA operating mode, the final stage multiplexer **272** provides the second final bias signal SFB, which is about zero volts, such that the second RF PA **54** is disabled, and during the second PA operating mode, the final stage multiplexer **272** provides the first final bias signal FFB, which is about zero volts, such that the first RF PA **50** is disabled.

FIG. **42** shows details of the driver stage current reference circuitry **268** and the final stage current reference circuitry **274** illustrated in FIG. **41** according to one embodiment of the driver stage current reference circuitry **268** and the final stage current reference circuitry **274**. The driver stage current reference circuitry **268** includes a driver stage temperature compensation circuit **276** to temperature compensate the driver stage reference current IDSR. The final stage current reference circuitry **274** includes a final stage temperature compensation circuit **278** to temperature compensate the final stage reference current IFSR.

Charge Pump Based PA Envelope Power Supply and Bias Power Supply

A summary of a charge pump based PA envelope power supply and bias power supply is presented, followed by a detailed description of the charge pump based PA envelope power supply according to one embodiment of the present disclosure. The present disclosure relates to a DC-DC converter, which includes a charge pump based RF PA envelope power supply and a charge pump based PA bias power supply. The DC-DC converter is coupled between RF PA circuitry and a DC power supply, such as a battery. As such, the PA envelope power supply provides an envelope power supply signal to the RF PA circuitry and the PA bias power supply provides a bias power supply signal to the RF PA circuitry. Both the PA envelope power supply and the PA bias power supply receive power via a DC power supply signal from the DC power supply. The PA envelope power supply includes a charge pump buck converter and the PA bias power supply includes a charge pump.

By using charge pumps, a voltage of the envelope power supply signal may be greater than a voltage of the DC power supply signal, a voltage of the bias power supply signal may be greater than the voltage of the DC power supply signal, or both. Providing boosted voltages may provide greater flexibility in providing envelope power for amplification and in biasing the RF PA circuitry. The charge pump buck converter provides the functionality of a charge pump feeding a buck converter. However, the charge pump buck converter requires fewer switching elements than a charge pump feeding a buck converter by sharing certain switching elements.

The charge pump buck converter is coupled between the DC power supply and the RF PA circuitry. The charge pump is coupled between the DC power supply and the RF PA circuitry. In one embodiment of the PA envelope power supply, the PA envelope power supply further includes a buck converter coupled between the DC power supply and the RF PA circuitry. The PA envelope power supply may operate in one of a first envelope operating mode and a second envelope operating mode. During the first envelope operating mode, the charge pump buck converter is active, and the buck converter is inactive. Conversely, during the second envelope operating mode, the charge pump buck converter is inactive, and the buck converter is active. As such, the PA envelope power supply may operate in the first envelope operating mode when a voltage above the voltage of the DC power supply signal may be needed. Conversely, the PA envelope power supply may operate in the second envelope operating mode when a voltage above the voltage of the DC power supply signal is not needed.

In one embodiment of the charge pump buck converter, the charge pump buck converter operates in one of a pump buck pump-up operating mode and at least one other pump buck operating mode, which may include any or all of a pump buck pump-down operating mode, a pump buck pump-even operating mode, and a pump buck bypass operating mode. In one embodiment of the charge pump, the charge pump operates in one of a bias supply pump-up operating mode and at least one other bias supply operating mode, which may include any or all of a bias supply pump-down operating mode, a bias supply pump-even operating mode, and a bias supply bypass operating mode.

In one embodiment of the RF PA circuitry, the RF PA circuitry has an RF PA, which is biased based on the bias power supply signal and receives the envelope power supply signal to provide power for amplification.

In one embodiment of the RF PA circuitry, the RF PA has a final stage that receives a final bias signal to bias the final stage, such that the final bias signal is based on the bias power supply signal. Additionally, the DC-DC converter provides the envelope power supply signal to the RF PA, which uses the envelope power supply signal to provide power for amplification. In one embodiment of the RF PA circuitry, the RF PA circuitry includes PA bias circuitry, which receives the bias power supply signal to provide the final bias signal. In one embodiment of the PA bias circuitry, the PA bias circuitry includes a final stage IDAC to receive and use the bias power supply signal in a digital-to-analog conversion to provide the final bias signal.

In one embodiment of the RF PA circuitry, the RF PA circuitry includes a first RF PA and a second RF PA, which may include a first final stage and a second final stage, respectively. The first RF PA is used to receive and amplify a highband RF input signal and the second RF PA is used to receive and amplify a lowband RF input signal. The RF PA circuitry may operate in one of a first PA operating mode and a second PA operating mode, such that during the first PA operating mode, the first RF PA is active and the second RF PA is disabled. Conversely, during the second PA operating mode, the first RF PA is disabled and the second RF PA is active. The PA bias circuitry includes the final stage IDAC and a final stage multiplexer. The final stage IDAC receives and uses the bias power supply signal in a digital-to-analog conversion to provide a final stage bias signal to the final stage multiplexer. During the first PA operating mode, the final stage multiplexer receives and forwards the final stage bias signal to provide a first final bias signal to the first RF PA to bias the first final stage. During the second PA operating mode, the final stage multiplexer receives and forwards the final stage bias signal to provide a second final bias signal to the second RF PA to bias the second final stage.

Interference Reduction Between RF Communications Bands

Embodiments of the present disclosure relate to RF PA circuitry and a PA envelope power supply. The RF PA circuitry receives and amplifies an RF input signal to provide an RF output signal using an envelope power supply signal, which is provided by the PA envelope power supply. The RF PA circuitry operates in either a normal RF spectral emissions mode or a reduced RF spectral emissions mode. When reduced RF spectral emissions are required, the RF PA circuitry operates in the reduced RF spectral emissions mode. As such, at a given RF output power, during the reduced RF spectral emissions mode, RF spectral emissions of the RF output signal are less than during the normal RF spectral emissions mode. As a result, the reduced RF spectral emissions mode may be used to reduce interference between RF communications bands.

Cellular communications bands are often adjacent to non-cellular communications bands. As such, interference between the cellular communications bands and the non-cellular communications bands needs to be minimized for the corresponding cellular and the non-cellular systems to operate properly and efficiently. For example, RF emissions from cellular communications signals that bleed into a non-cellular communications band must be low enough to prevent problems in the corresponding non-cellular communications system. Such RF emissions may be called RF spectral emissions since these emissions fall outside of a desired RF spectrum associated with a corresponding cellular communications signal.

The RF spectral emissions from a cellular communications signal may adversely impact the non-cellular communications system. However, when the cellular communications band is adequately separated from the non-cellular communications band, when a magnitude of the cellular communications signal having RF spectral emissions is sufficiently small, when a sensitivity of the corresponding non-cellular communications system to the RF spectral emissions is sufficiently small, or any combination thereof, the cellular communications signal does not adversely impact the non-cellular communications system.

In this regard, when the cellular communications band is about adjacent to the non-cellular communications band and the non-cellular communications system is sensitive to the RF spectral emissions, the RF spectral emissions must be limited to allow the non-cellular communications system to function properly. As such, the RF spectral emissions may be limited by limiting the magnitude of the cellular communications signal, by reducing characteristic RF spectral emissions of the cellular communications signal, or both. The characteristic RF spectral emissions may be reduced by increasing linearity in an RF PA, by improving a quadrature balance of a quadrature RF PA, or both.

As previously mentioned, the DC-DC converter **32** (FIG. **2**) provides the envelope power supply signal EPS to the RF PA circuitry **30** (FIG. **2**). The RF PA circuitry **30** (FIG. **14**) includes the PA-DCI **60** (FIG. **14**), the PA control circuitry **94** (FIG. **14**), the first RF PA **50** (FIG. **14**), and the second RF PA **54** (FIG. **14**). The PA-DCI **60** (FIG. **14**) is coupled between the PA control circuitry **94** (FIG. **14**) and the digital communications bus **66** (FIG. **14**). The DC-DC converter **32** (FIG. **2**) provides the envelope power supply signal EPS to the first RF PA **50** (FIG. **14**) and to the second RF PA **54** (FIG. **14**). The PA bias circuitry **96** (FIG. **14**) provides the first driver bias signal FDB and the first final bias signal FFB to the first RF PA **50** (FIG. **14**) based on the bias power supply signal BPS and the bias configuration control signal BCC. The PA bias circuitry **96** (FIG. **14**) provides the second driver bias signal SDB and the second final bias signal SFB to the second RF PA **54** (FIG. **14**) based on the bias power supply signal BPS and the bias configuration control signal BCC. As such, the PA bias circuitry **96** (FIG. **14**) is coupled between the PA control circuitry **94** (FIG. **14**) and the RF PAs **50**, **54** (FIG. **14**).

The first RF PA **50** (FIG. **14**) receives and amplifies the first RF input signal FRFI to provide the first RF output signal FRFO using the envelope power supply signal EPS. The second RF PA **54** (FIG. **14**) receives and amplifies the second RF input signal SRFI to provide the second RF output signal SRFO using the envelope power supply signal EPS. In general, the RF PA circuitry **30** (FIG. **14**) receives and amplifies an RF input signal to provide an RF output signal using the envelope power supply signal EPS (FIG. **14**). Further, in general, the RF PA circuitry **30** (FIG. **14**) includes an RF PA, which receives and amplifies the RF input signal to provide the RF output signal using the envelope power supply signal EPS. In one embodiment of the RF PA, the RF PA is a quadrature RF PA, which receives and amplifies the RF input signal to provide the RF output signal using the envelope power supply signal EPS. In this regard, the PA bias circuitry **96** (FIG. **14**) is coupled between the PA control circuitry **94** (FIG. **14**) and the RF PAs.

In a first embodiment of the RF PA circuitry **30** (FIG. **14**), the RF PA is the first RF PA **50** (FIG. **14**), the RF input signal is the first RF input signal FRFI and the RF output signal is the first RF output signal FRFO. In a second embodiment of the RF PA circuitry **30** (FIG. **14**), the RF PA is the second RF PA **54** (FIG. **14**), the RF input signal is the second RF input signal SRFI and the RF output signal is the second RF output signal SRFO.

In one embodiment of the RF PA circuitry **30** (FIG. **40**), the first RF PA **50** (FIG. **40**) includes the first driver stage **252** (FIG. **40**) and the first final stage **254** (FIG. **40**). The second RF PA **54** (FIG. **40**) includes the second driver stage **256** (FIG. **40**) and the second final stage **258** (FIG. **40**). In one embodiment of the RF PA, the RF PA includes a driver stage and a final stage. In an exemplary embodiment of the RF PA, the driver stage is the first driver stage **252** (FIG. **40**) and the final stage is the first final stage **254** (FIG. **40**). In an alternate exemplary embodiment of the RF PA, the driver stage is the second driver stage **256** (FIG. **40**) and the final stage is the second final stage **258** (FIG. **40**). In one embodiment of the quadrature RF PA, the RF PA includes a quadrature driver stage and a quadrature final stage. In an exemplary embodiment of the quadrature RF PA, the quadrature driver stage is the first driver stage **252** (FIG. **40**) and the quadrature final stage is the first final stage **254** (FIG. **40**). In an alternate exemplary embodiment of the quadrature RF PA, the quadrature driver stage is the second driver stage **256** (FIG. **40**) and the quadrature final stage is the second final stage **258** (FIG. **40**).

In one embodiment of the RF PA circuitry **30** (FIG. **14**), the RF PA circuitry **30** (FIG. **14**) operates in one of a normal RF spectral emissions mode and a reduced RF spectral emissions mode. When reduced RF spectral emissions are required, the RF PA circuitry **30** (FIG. **14**) operates in the reduced RF spectral emissions mode. At a given RF output power, during the reduced RF spectral emissions mode, RF spectral emissions of the RF output signal are less than during the normal RF spectral emissions mode.

In one embodiment of the RF PA circuitry **30** (FIG. **14**), when reduced RF spectral emissions are not required, the RF PA circuitry **30** (FIG. **14**) operates in the normal RF spectral emissions mode. In an alternate embodiment of the RF PA circuitry **30** (FIG. **14**), the RF PA circuitry **30** (FIG. **14**) is hardwired to select the reduced RF spectral emissions mode, such that the reduced RF spectral emissions mode is continuously selected. In an additional embodiment of the RF PA circuitry **30** (FIG. **14**), the PA control circuitry **94** (FIG. **14**) selects the one of the normal RF spectral emissions mode and the reduced RF spectral emissions mode.

In one embodiment of the RF PA circuitry **30** (FIG. **14**), the RF output signal falls within one of a group of RF communications bands, such that the group of RF communications bands includes a first RF communications band, which is about adjacent to a second RF communications band. In one embodiment of the RF PA circuitry **30** (FIG. **14**), when the one of the group of RF communications bands is the first RF communications band, reduced RF spectral emissions are required. In an alternate embodiment of the RF PA circuitry **30** (FIG. **14**), when the one of the group of RF communications bands is the first RF communications band and an RF output power from the RF PA is higher than an RF output power threshold, reduced RF spectral emissions are required.

In one embodiment of the RF PA circuitry **30** (FIG. **14**), the first RF communications band is a cellular communications band and the second RF communications band is a non-cellular communications band. In one embodiment of the first RF communications band, the first RF communications band is a 3rd Generation Partnership Project (3GPP) cellular communications band. In one embodiment of the 3GPP cellular communications band, the 3GPP cellular communications band is a 3GPP long term evolution (3GPP-LTE) cellular communications band. In one embodiment of the first RF communications band, a frequency range of the first RF communications band is between about 777 megahertz and about 787 megahertz. In one embodiment of the second RF communications band, the second RF communications band is a Public Safety Band, which is a non-cellular communications band. In one embodiment of the second RF communications band, a frequency range of the second RF communications band is between about 788 megahertz and about 798 megahertz.

In one embodiment of the RF PA circuitry **30** (FIG. **14**), during the reduced RF spectral emissions mode, the RF PA has higher PA linearity than during the normal RF spectral emissions mode. As such, to provide higher linearity, in one embodiment of the RF PA circuitry **30** (FIG. **14**), during the reduced RF spectral emissions mode, the RF PA has a higher PA bias level than during the normal RF spectral emissions mode. Further, to provide higher linearity, in one embodiment of the RF PA circuitry **30** (FIG. **14**), at a given RF output power, during the reduced RF spectral emissions mode, a magnitude of the envelope power supply signal EPS (FIG. **14**) is higher than during the normal RF spectral emissions mode.

In one embodiment of the RF PA circuitry **30** (FIG. **14**), during the reduced RF spectral emissions mode, a quadrature balance of the quadrature RF PA is better than during the normal RF spectral emissions mode. The quadrature balance is better when a signal level in an in-phase path of the quadrature RF PA is closer to matching a signal level in a quadrature-phase path of the quadrature RF PA. In an alternate embodiment of the RF PA circuitry **30** (FIG. **14**), during the reduced RF spectral emissions mode, a quadrature balance of the quadrature RF PA is better and the quadrature RF PA has higher PA linearity than during the normal RF spectral emissions mode.

In one embodiment of the RF PA circuitry **30** (FIG. **14**), during the reduced RF spectral emissions mode, a quadrature balance of the quadrature final stage is better than during the normal RF spectral emissions mode. In one embodiment of the quadrature final stage, during the reduced RF spectral emissions mode, a bias level of the quadrature final stage is different than during the normal RF spectral emissions mode to improve quadrature balance of the quadrature final stage.

In one embodiment of the RF PA circuitry **30** (FIG. **14**), during the reduced RF spectral emissions mode, a quadrature balance of the quadrature driver stage is better than during the normal RF spectral emissions mode. In one embodiment of the quadrature driver stage, during the reduced RF spectral emissions mode, a bias level of the quadrature driver stage is different than during the normal RF spectral emissions mode to improve quadrature balance of the quadrature driver stage.

In an alternate embodiment of the RF PA circuitry **30** (FIG. **14**), during the reduced RF spectral emissions mode, a quadrature balance of the quadrature final stage is better and a quadrature balance of the quadrature driver stage is better than during the normal RF spectral emissions mode. In one embodiment of the quadrature final stage and the quadrature driver stage, during the reduced RF spectral emissions mode, a bias level of the quadrature final stage is different and a bias level of the quadrature driver stage is different than during the normal RF spectral emissions mode to improve quadrature balance of the quadrature final stage and the quadrature driver stage.

In one embodiment of the RF PA circuitry **30** (FIG. **14**), the PA control circuitry **94** (FIG. **14**) receives information via the digital communications bus **66** (FIG. **14**) and selects the one of the normal RF spectral emissions mode and the reduced RF spectral emissions mode based on the information. In one embodiment of the RF PA circuitry **30** (FIG. **14**), the RF output signal falls within one of the group of RF communications bands and the information identifies the one of the group of RF communications bands. In one embodiment of the RF PA circuitry **30** (FIG. **14**), the information is indicative of the RF output power from the RF PAs. In one embodiment of the RF PA circuitry **30** (FIG. **14**), the information is indicative of the magnitude of the envelope power supply signal EPS (FIG. **14**). In this regard, in one embodiment of the RF PA circuitry **30** (FIG. **14**), the PA control circuitry **94** (FIG. **14**) selects the one of the normal RF spectral emissions mode and the reduced RF spectral emissions mode based on the one of the group of RF communications bands, the indicated RF output power from the RF PAs, the indicated magnitude of the envelope power supply signal EPS (FIG. **14**), or any combination thereof.

In one embodiment of the transceiver circuitry **34** (FIG. **6**), the transceiver circuitry **34** (FIG. **6**) provides the information to the PA control circuitry **94** (FIG. **14**) via the digital communications bus **66** (FIG. **14**). In one embodiment of the transceiver circuitry **34** (FIG. **6**), the transceiver circuitry **34** (FIG. **6**) selects the one of the normal RF spectral emissions mode and the reduced RF spectral emissions mode and indicates the selection via the information. In one embodiment of the RF communications system **26** (FIG. **6**), the RF communications system **26** (FIG. **6**) receives RF signals that are used by the transceiver circuitry **34** (FIG. **6**) to provide the information. The RF communications system **26** (FIG. **6**) may be part of a communications network and the received RF signals may be indicative of network commands, network information, or the like, which are used by the transceiver circuitry **34** (FIG. **6**) to provide the information.

Those skilled in the art will recognize improvements and modifications to the preferred embodiments of the present disclosure. All such improvements and modifications are considered within the scope of the concepts disclosed herein and the claims that follow.

What is claimed is:

1. Circuitry comprising:

radio frequency (RF) power amplifier (PA) circuitry comprising a quadrature RF PA and adapted to:

receive and amplify an RF input signal using the quadrature RF PA to provide an RF output signal using an envelope power supply signal;

operate in one of a normal RF spectral emissions mode and a reduced RF spectral emissions mode, wherein a quadrature balance of the quadrature RF PA is better in the reduced RF spectral emissions mode than in the normal RF spectral emissions mode;

when reduced RF spectral emissions are required, operate in the reduced RF spectral emissions mode, such that the quadrature RF PA has higher PA linearity in the reduced RF spectral emissions mode than during the normal RF spectral emissions mode; and

when reduced RF spectral emissions are not required, operate in the normal RF spectral emissions mode; and

a PA envelope power supply adapted to provide the envelope power supply signal.

2. The circuitry of claim **1** wherein when the reduced RF spectral emissions are not required, the RF PA circuitry is further adapted to operate in the normal RF spectral emissions mode.

3. The circuitry of claim **1** wherein at a given RF output power, during the reduced RF spectral emissions mode, RF spectral emissions of the RF output signal are less than during the normal RF spectral emissions mode.

4. The circuitry of claim **1** wherein:

the RF output signal falls within one of a plurality of RF communications bands;

the plurality of RF communications bands comprises a first RF communications band, which is about adjacent to a second RF communications band; and

when the one of the plurality of RF communications bands is the first RF communications band, reduced RF spectral emissions are required.

5. The circuitry of claim **1** wherein:

the RF output signal falls within one of a plurality of RF communications bands;

the plurality of RF communications bands comprises a first RF communications band, which is about adjacent to a second RF communications band; and

when the one of the plurality of RF communications bands is the first RF communications band and an RF output power from the quadrature RF PA is higher than an RF output power threshold, reduced RF spectral emissions are required.

6. The circuitry of claim **5** wherein the first RF communications band is a cellular communications band and the second RF communications band is a non-cellular communications band.

7. The circuitry of claim **6** wherein the first RF communications band is a 3rd Generation Partnership Project (3GPP) cellular communications band and the second RF communications band is a Public Safety Band.

8. The circuitry of claim **7** wherein a frequency range of the first RF communications band is between about 777 megahertz and about 787 megahertz, and a frequency range of the second RF communications band is between about 788 megahertz and about 798 megahertz.

9. The circuitry of claim **1** wherein at a given RF output power, during the reduced RF spectral emission mode, a magnitude of the envelope power supply signal is higher than during the normal RF spectral emissions mode.

10. The circuitry of claim **1** wherein:

the quadrature RF PA comprises a quadrature final stage; and

during the reduced RF spectral emissions mode, a bias level of the quadrature final stage is different than during the normal RF spectral emissions mode to improve a quadrature balance of the quadrature final stage.

11. The circuitry of claim **10** wherein:

the quadrature RF PA comprises a quadrature driver stage; and

during the reduced RF spectral emissions mode, a bias level of the quadrature driver stage is different than during the normal RF spectral emissions mode to improve a quadrature balance of the quadrature driver stage.

12. The circuitry of claim **1** wherein:

the quadrature RF PA comprises a quadrature driver stage; and

during the reduced RF spectral emissions mode, a bias level of the quadrature driver stage is different than during the normal RF spectral emissions mode to improve a quadrature balance of the quadrature driver stage.

13. The circuitry of claim **1** wherein the RF PA circuitry is hardwired to select the reduced RF spectral emissions mode.

14. The circuitry of claim **1** wherein the RF PA circuitry comprises:

the quadrature RF PA adapted to receive and amplify the RF input signal to provide the RF output signal using the envelope power supply signal;

PA control circuitry adapted to select the one of the normal RF spectral emissions mode and the reduced RF spectral emissions mode; and

PA bias circuitry coupled between the PA control circuitry and the quadrature RF PA.

15. The circuitry of claim **14** wherein:

the RF PA circuitry further comprises a PA-digital communications interface (PA-DCI) coupled between a digital communications bus and the PA control circuitry; and

the PA control circuitry is further adapted to receive information via the digital communications bus and select the one of the normal RF spectral emissions mode and the reduced RF spectral emissions mode based on the information.

16. The circuitry of claim **15** wherein:

the RF output signal falls within one of a plurality of RF communications bands; and

the information identifies the one of the plurality of RF communications bands.

17. The circuitry of claim **16** wherein the information is indicative of an RF output power from the quadrature RF PA.

18. The circuitry of claim **16** wherein the information is indicative of a magnitude of the envelope power supply signal.

19. A method comprising:

receiving and amplifying a radio frequency (RF) input signal using a quadrature RF power amplifier (PA) to provide an RF output signal using an envelope power supply signal;

operating in one of a normal RF spectral emissions mode and a reduced RF spectral emissions mode;

when reduced RF spectral emissions are required, operating in the reduced RF spectral emissions mode, wherein a quadrature balance of the quadrature RF PA is better in the reduced spectral emissions mode than in the normal RF spectral emissions mode;

when reduced RF spectral emissions are not required, operating in the normal RF spectral emissions mode; and

providing the envelope power supply signal.

* * * * *

US009401613B2

(12) **United States Patent**
Olah et al.

(10) **Patent No.: US 9,401,613 B2**
(45) **Date of Patent: *Jul. 26, 2016**

(54) **INTEGRATED ELECTRONICS FOR PERPETUAL ENERGY HARVESTING**

(71) Applicant: **Banpil Photonics, Inc.**, Santa Clara, CA (US)

(72) Inventors: **Robert Olah**, Sunnyvale, CA (US); **Achyut K Dutta**, Sunnyvale, CA (US)

(73) Assignee: **Banpil Photonics, Inc.**, Santa Clara, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **14/341,475**

(22) Filed: **Jul. 25, 2014**

(65) **Prior Publication Data**

US 2014/0333136 A1 Nov. 13, 2014

Related U.S. Application Data

(62) Division of application No. 13/151,253, filed on Jun. 1, 2011, now Pat. No. 8,823,210.

(51) **Int. Cl.**

H02J 7/00	(2006.01)
H02J 1/12	(2006.01)
H02J 1/10	(2006.01)
H02J 7/35	(2006.01)
H02J 1/00	(2006.01)

(52) **U.S. Cl.**
CPC ***H02J 7/0055*** (2013.01); ***H02J 1/102*** (2013.01); ***H02J 1/12*** (2013.01); ***H02J 7/35*** (2013.01); *H02J 7/0042* (2013.01); *H02J 2001/106* (2013.01); *Y10T 307/511* (2015.04); *Y10T 307/516* (2015.04)

(58) **Field of Classification Search**
CPC H02J 1/00
USPC 307/43, 45
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,506,586	B1 *	3/2009	Pereira	F42C 11/008 102/207
8,823,210	B1 *	9/2014	Olah	H02J 7/0055 307/45
2009/0000613	A1 *	1/2009	Edwards	F24J 2/16 126/684
2011/0006727	A1 *	1/2011	Blau	H01M 10/44 320/101
2012/0160290	A1 *	6/2012	Chen	F24J 2/32 136/206

* cited by examiner

Primary Examiner — Robert Deberadinis

(57) **ABSTRACT**

An apparatus for perpetually harvesting ambient near ultraviolet to far infrared radiation to provide continual power regardless of the environment, incorporating a system for the harvesting electronics governing power management, storage control, and output regulation. The harvesting electronics address issues of efficiently matching the voltage and current characteristics of the different harvested energy levels, low power consumption, and matching the power output demand. The device seeks to harvest the largely overlooked blackbody radiation through use of a thermal harvester, providing a continuous source of power, coupled with a solar harvester to provide increased power output.

20 Claims, 7 Drawing Sheets

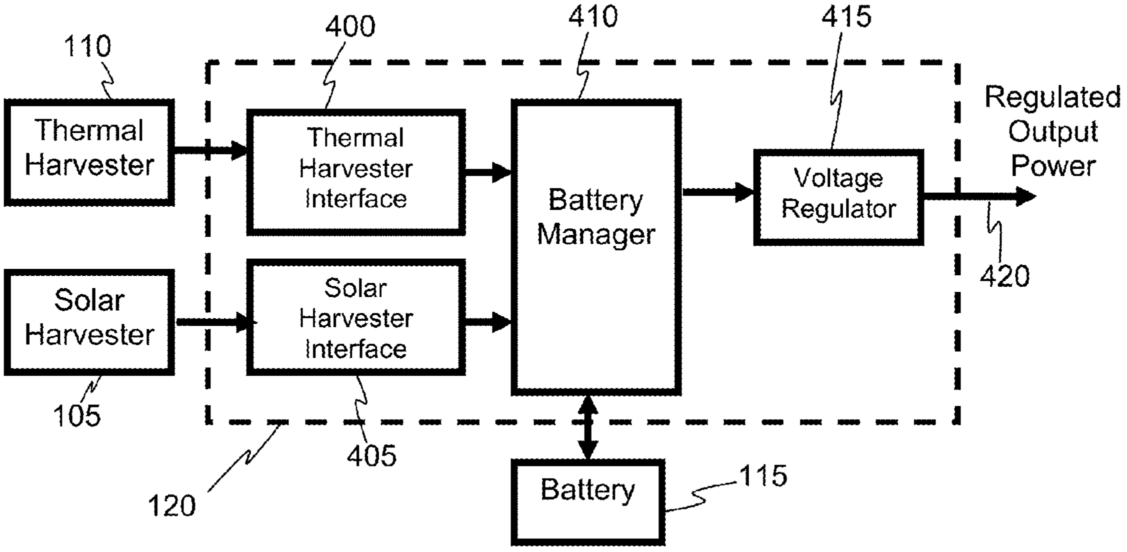

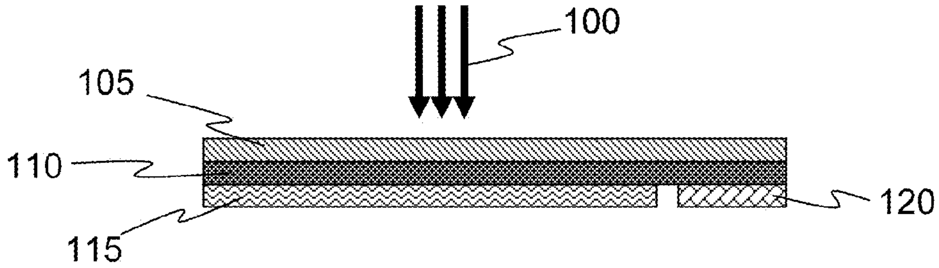
100
105
110
120
115

FIG. 1A

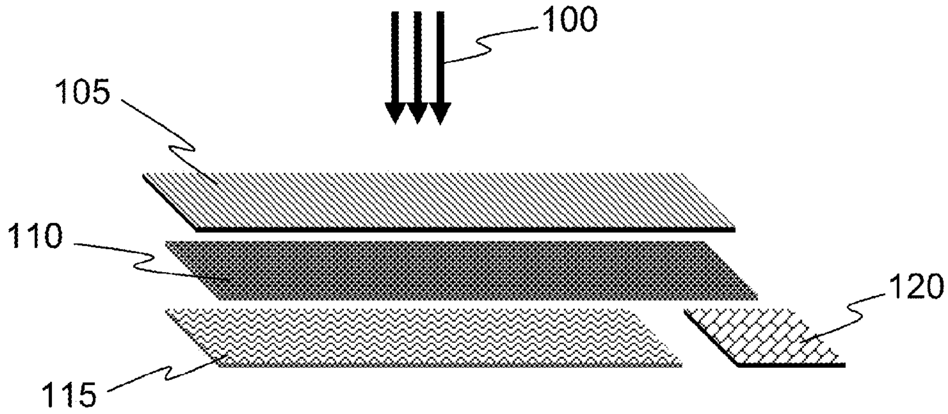
100
105
110
120
115

FIG. 1B

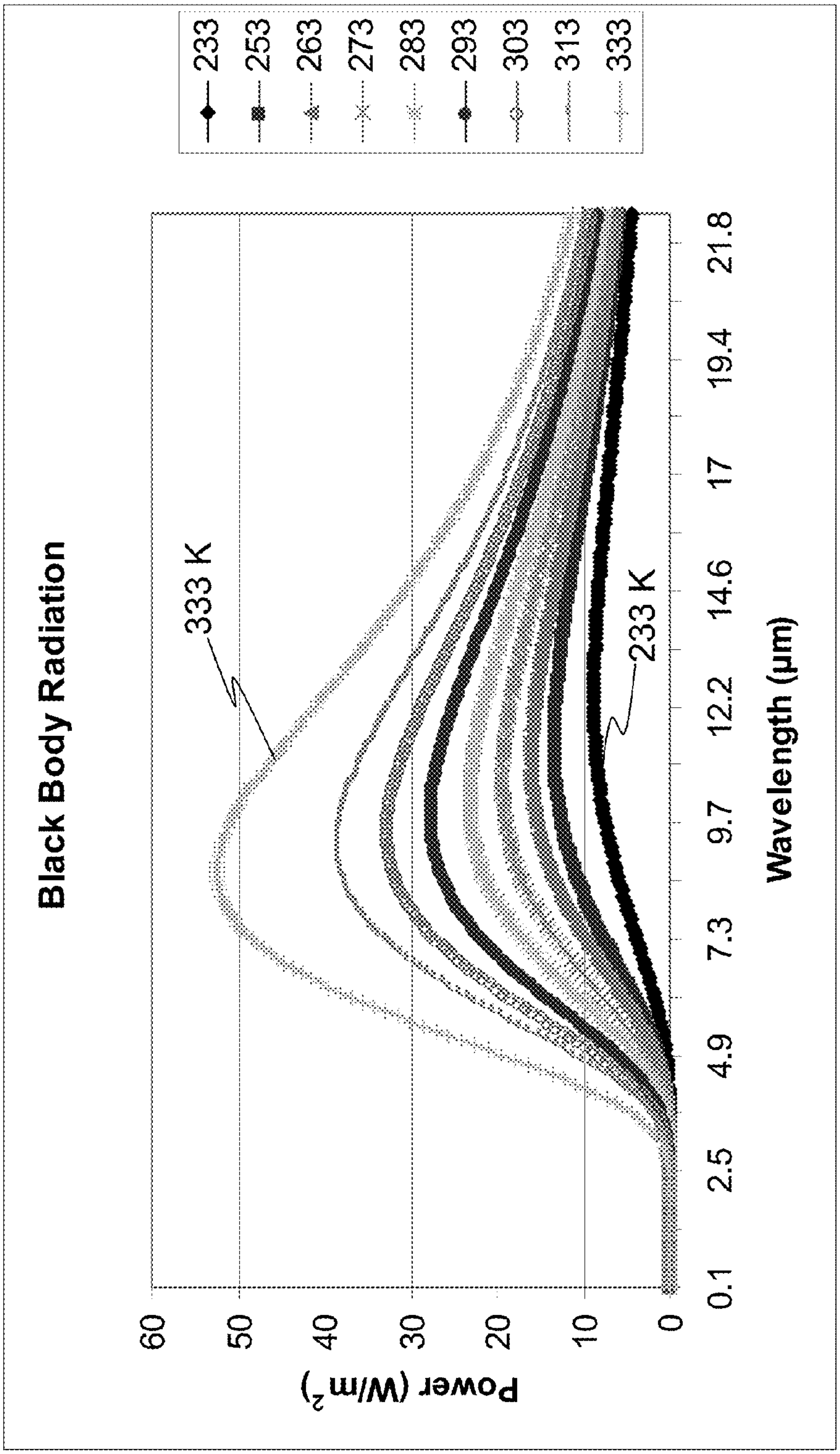
Black Body Radiation
Power (W/m²)
60
50
40
30
20
10
0
0.1
2.5
4.9
7.3
9.7
12.2
14.6
17
19.4
21.8
Wavelength (µm)
333 K
233 K
233
253
263
273
283
293
303
313
333

FIG. 2

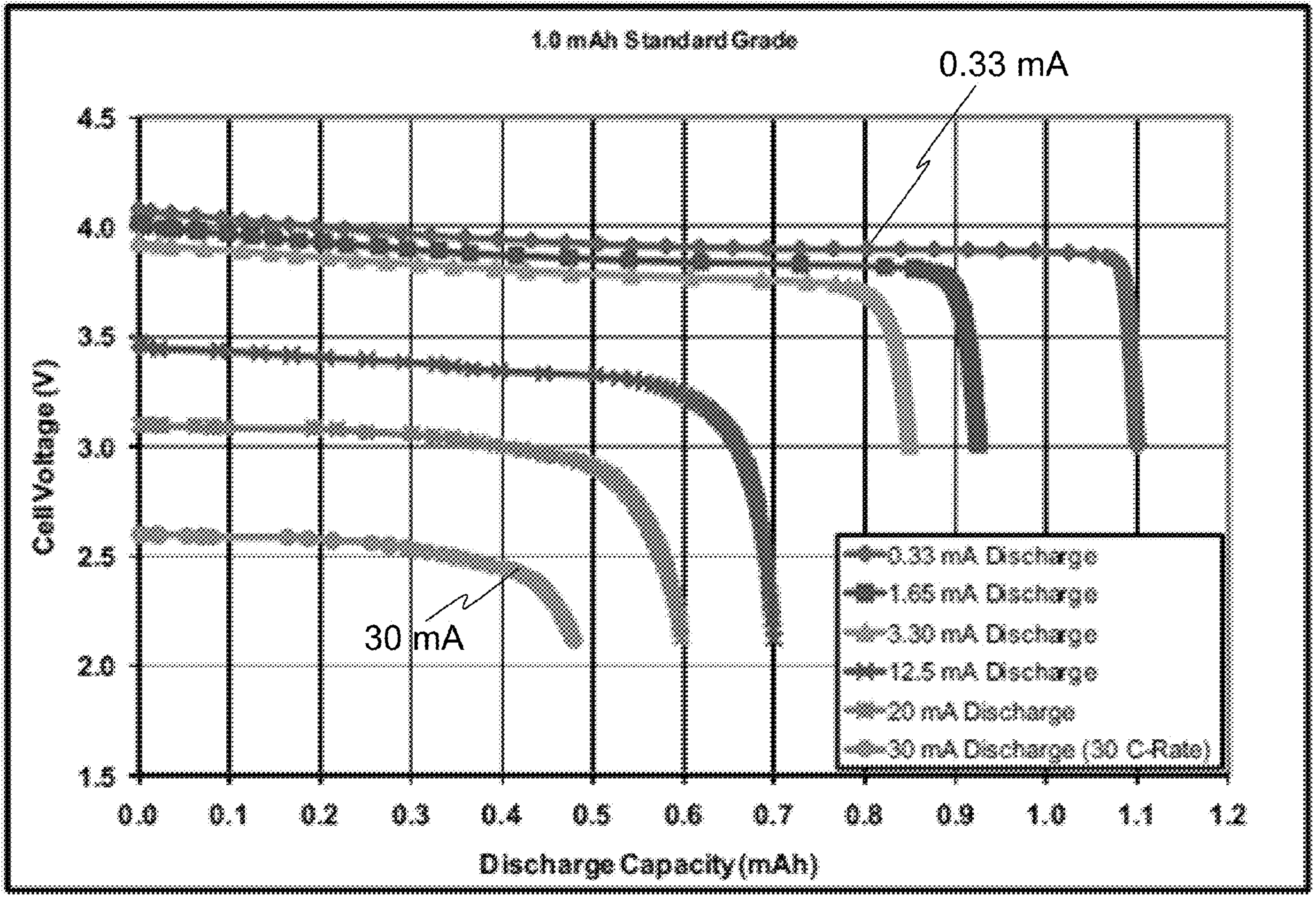
1.0 mAh Standard Grade
0.33 mA
30 mA
Cell Voltage (V)
4.5
4.0
3.5
3.0
2.5
2.0
1.5
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
Discharge Capacity (mAh)
0.33 mA Discharge
1.65 mA Discharge
3.30 mA Discharge
12.5 mA Discharge
20 mA Discharge
30 mA Discharge (30 C-Rate)

FIG. 3

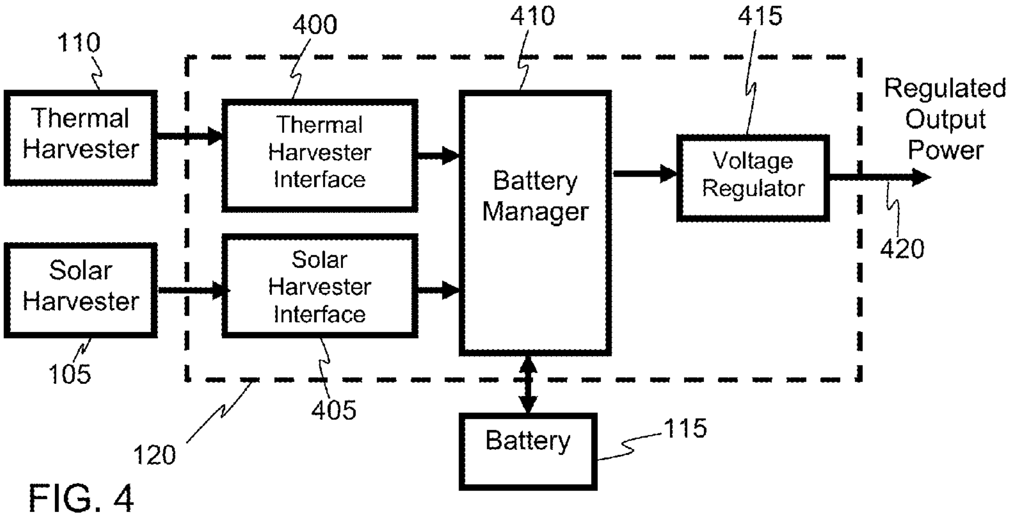
110
Thermal Harvester
400
Thermal Harvester Interface
410
Battery Manager
415
Voltage Regulator
Regulated Output Power
420
Solar Harvester
105
Solar Harvester Interface
405
120
Battery
115

FIG. 4

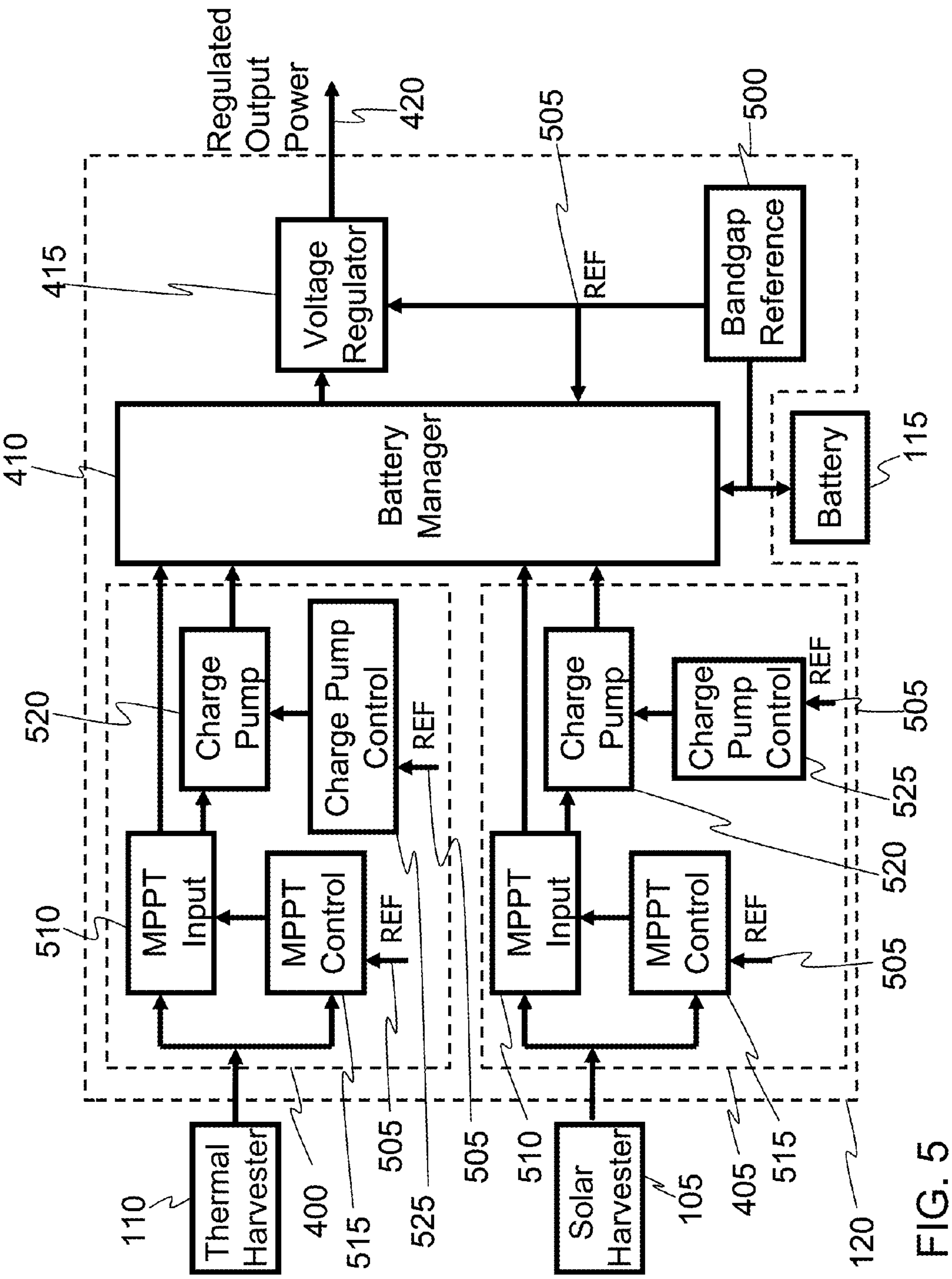
Thermal Harvester
Solar Harvester
MPPT Input
MPPT Control
REF
Charge Pump
Charge Pump Control
Battery Manager
Battery
Voltage Regulator
Bandgap Reference
REF
Regulated Output Power

FIG. 5

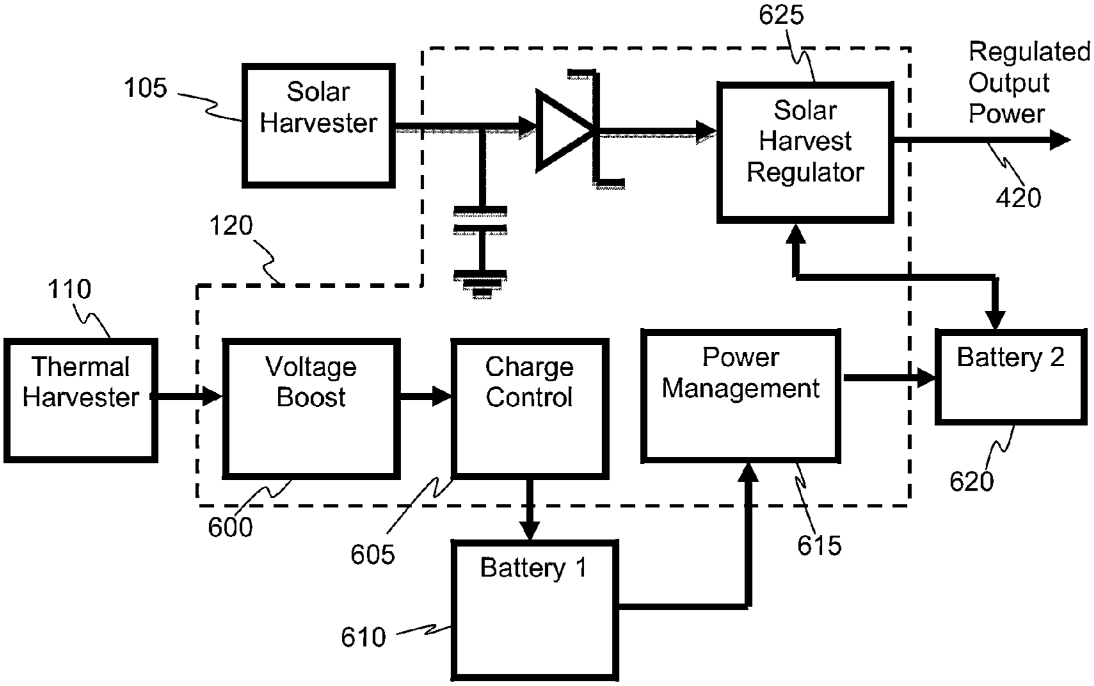
625
Regulated Output Power
105
Solar Harvester
Solar Harvest Regulator
420
120
110
Thermal Harvester
Voltage Boost
Charge Control
Power Management
Battery 2
620
600
605
615
Battery 1
610

FIG. 6

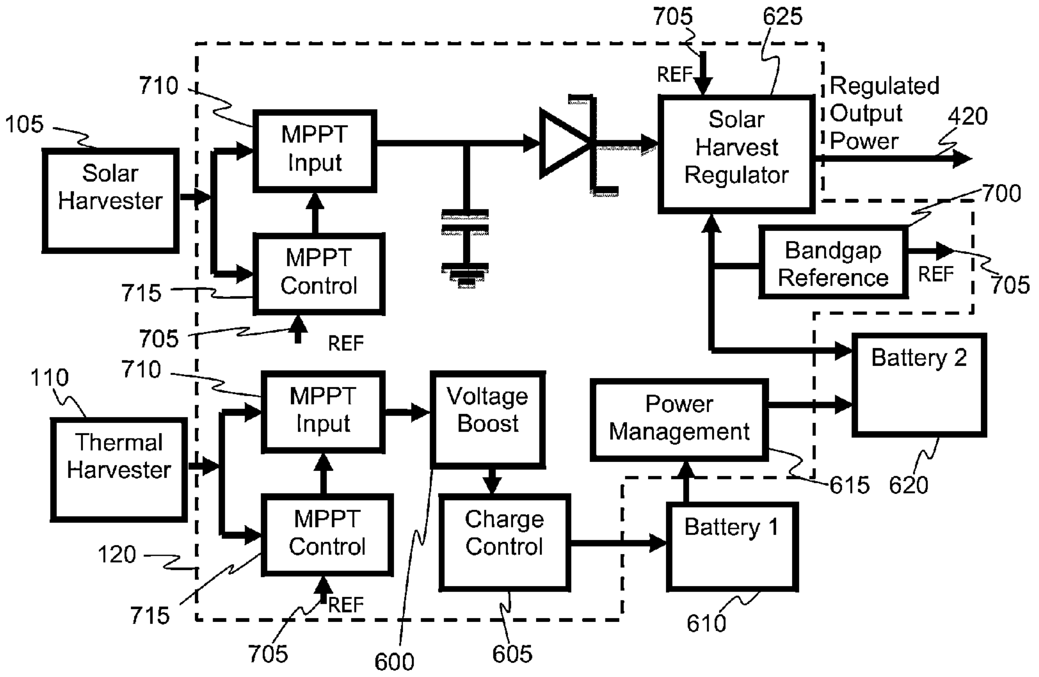
705
625
REF
710
Regulated
Output
Power
420
105
MPPT
Input
Solar
Harvest
Regulator
Solar
Harvester
700
Bandgap
Reference
REF
705
MPPT
Control
715
705
REF
Battery 2
110
710
MPPT
Input
Voltage
Boost
Power
Management
Thermal
Harvester
615
620
MPPT
Control
Charge
Control
Battery 1
120
715
REF
610
705
600
605

FIG. 7

INTEGRATED ELECTRONICS FOR PERPETUAL ENERGY HARVESTING

PARENT CASE TEXT

This is a divisional of application(s) Ser. No. 13/151,253 filed on Jun. 1, 2011.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

This invention was made in part with Government support under contract no. W31P4Q10C0034 awarded by US Army Aviation and Missile Command. The Government has certain rights in the invention.

CROSS-REFERENCE TO RELATED APPLICATIONS

Not Applicable

FIELD OF INVENTION

This invention pertains to the harvesting of ambient energy in the surrounding, environment. More particularly this invention is related to a semiconductor perpetual energy harvesting and storage device and the harvesting electronics that control energy storage and output allowing the device to continually harvest and provide continuous power during both day and night in any environment.

BACKGROUND OF INVENTION

There are many sources of energy that can be converted into electrical energy. Sources that have been explored for harvesting are wind, temperature differences, motion, light energy and radio frequency energy. In all these energy situations, a specific situation must be present, such as air flow, temperature gradient, movement, sunlight, or transmitted RF, before energy can be harvested. When these situations do not occur, there is no source for energy to be harvested.

Harvesting ambient thermal energy has not been widely explored. The characteristic of thermal energy having a peak at long infrared wavelengths poses a significant challenge for harvesting. Additionally, power management, storage control, and regulation electronics must be developed to operate efficiently and respond to the power output demand.

BRIEF DESCRIPTION OF DRAWINGS

FIGS. **1**A and B depicts a high level structural representation of an energy harvesting system.

FIG. **2** depicts power available from blackbody radiation at various temperatures

FIG. **3** depicts the discharge characteristics of a lithium ion battery.

FIG. **4** depicts a block diagram of the harvester electronics utilizing a dual circuit design.

FIG. **5** depicts an alternate embodiment of a block diagram of the harvester electronics utilizing a dual circuit design.

FIG. **6** depicts a block diagram of the harvester electronics utilizing a single circuit design.

FIG. **7** depicts an alternate embodiment of a block diagram of the harvester electronics utilizing a single circuit design.

DETAILED DESCRIPTION

Reference is made in detail to the preferred embodiments of the invention. While the invention is described in conjunction with the preferred embodiments, the invention is not intended to be limited by these preferred embodiments. The contemplated embodiments for carrying out the present invention are described in turn with reference to the accompanying figures.

For purposes of the invention, radiation is defined to be the electromagnetic spectrum, particularly the near ultraviolet, visible, near infrared, short-wave infrared, mid-wave infrared, long-wave infrared and far infrared bands. The near ultraviolet hand is comprised of wavelengths from about 300 to 400 nm. The visible light hand is comprised of wavelengths from about 400 to 780 nm. The near infrared band is comprised of wavelengths from about 0.78 to 1 μm. The short-wave infrared band is comprised of wavelengths from about 1 to 3 μm. The mid-wave infrared band is comprised of wavelengths from about 3 to 6 μm. The long-wave infrared band is comprised of wavelengths from about 6 to 14 μm. The far infrared band is comprised of wavelengths from about 14 to 40 μm. However, the boundary between near ultraviolet radiation and visible light along with the boundary between visible light and infrared radiation is not precisely defined, resulting in overlap between the bands.

The term electrically connected is defined to encompass an electrical current flow, including bidirectional, unidirectional, or any hybrid current flow, such as an uneven current.

FIG. **1** depicts a high level structural configuration for an energy harvesting system. FIG. **1**A illustrates a cross section of the structural configuration. FIG. **1**B shows an exploded three dimensional view. The interconnects and electrodes in the structural configurations depicted in FIGS. **1**A and B are not shown.

The energy harvesting system comprises a solar harvester **105**, a thermal harvester **110**, a battery **115**, and harvester electronics **120**. The thermal harvester **110** is located between the solar harvester **105** and the battery **115**. The harvester electronics **120** is attached to the thermal harvester **110** on the same side as the battery **115**. The solar harvester **105** and thermal harvester **110** are electrically isolated from each other and individually electrically connected to the harvester electronics **120**. The battery **115** is electrically connected to the harvester electronics **120**.

The solar harvester **105** and the thermal harvester **110** may be comprised of a photovoltaic or photoelectrochemical system. The solar harvester **105** may include any photovoltaic system which absorbs radiation within the near ultraviolet, visible, or near infrared bands to generate power. The solar harvester **105** may be comprised of any type of solar cell, including traditional crystallized silicon solar cells, thin film solar cells, multilayered solar cells, dye sensitized solar cells, or organic polymer solar cells. The material comprising the solar harvester may include, but is not limited to, amorphous silicon, polysilicon, crystalline silicon, Ge, SiGe, CdTe, ZnO, CdZnTe, HgCdTe, HgZnTe, GaAs, GaN, InP, GaSb, InSb, other semiconductors, light sensitive dyes, or polymers. A photovoltaic thermal harvester **105** may be comprised of a single junction or multiple junctions. The material comprising the thermal harvester **105** may include HgCdTe, HgZnTe, InSb, InAs, GaSb, GaAs, InP, PbTe, or polymers.

The battery **115** may be any type of rechargeable battery. One embodiment of the invention contemplates the use of solid state thin film battery may be comprised of lithium-ion polymer, lithium, semiconductor, polymer, or flexible batteries. Optionally, the solid state thin film battery may be discretely or monolithically integrated with the harvester. Optionally, in a separate embodiment, bulk batteries may be used. For example, rechargeable lithium air, nickel-cadmium, or nickel metal-hydride batteries. Optionally, a single or multiple battery cells may be used. Optionally, in the case of multiple battery cells, the cells may be arranged in various configurations, e.g. series, parallel, etc. The battery **115** may be flexible, semi-flexible, or rigid. The battery **115** is electrically connected to the harvester electronics **120**.

The harvester electronics **120** may be comprised of circuitry which regulates the rate of charge and voltage level of the battery along with regulating the output of power to a load. In one embodiment of the invention, the harvester electronics may be comprised of a single chip integrated circuit. Optionally, the integrated circuits may be a CMOS circuit. Alternatively, an FPGA may be used for harvester electronics **120** to be configurable in the field. Optionally, the harvester electronics may be discretely integrated with the battery **115**. The harvester electronics may be flexible, semi-flexible, or rigid. The harvester electronics **120** may be powered by the battery **115**, energy from the solar harvester **105** or thermal harvester **110**, or a combination of energy sources.

The system is oriented to absorb radiation **100** arriving from the indicated direction, passing through the solar harvester **105** before reaching the thermal harvester **110**. The radiation **100** is comprised of the near ultraviolet, visible, near infrared, short-wave infrared, mid-wave infrared, long-wave infrared, and far infrared bands.

FIG. **2** illustrates the blackbody power as a function of wavelength at various temperatures. The blackbody radiation spectrum ranges from near ultraviolet to the infrared spectrum. Blackbody radiation has been largely overlooked as a potential energy source.

Blackbody radiation is a physical effect in which all objects at temperatures above 0 K emit electromagnetic power. The spectral radiance of an object is described through Planck's Law, which is presented as,

$$I(\lambda) = \frac{2hc^2}{\lambda^5\left(e^{\frac{hc}{\lambda kT}} - 1\right)},$$

where c is the speed of light, k is Boltzmann's constant, h is Planck's constant, λ is wavelength, and T is temperature in Kelvin of the object. By applying Lambert's cosine law to a hemispherical flux density from a flat surface, the maximum amount of incident radiation is equal to πI. This yields a maximum spectral radiance of

$$I(\lambda) = \frac{2\pi hc^2}{\lambda^5\left(e^{\frac{hc}{\lambda kT}} - 1\right)},$$

which best models the intensity of radiation from a planar source. As temperature increases, power (in watts per meters squared) increases at lower wavelengths. This results in the peak shifting towards lower wavelengths at higher temperatures.

As shown in FIG. **2**, at the temperatures ranging from −40° C. (233 K to +60° C., the peak power is at about 9 μm wavelength. Using Stefan-Boltzmann's law, $5.67\times10^{-8}\times T^4$, the total available blackbody radiation at a given temperature can be calculated. At room temperature, about 293 Kelvin, the total available blackbody radiation is estimated to be 413 W/m^2 with a peak between 7 to 10 μm.

FIG. **3** illustrates the discharge characteristics of a lithium-ion polymer cell at various currents. The figure shows that the cell maintains a relatively constant voltage, which drops sharply as the current decreases due to the battery's inherent resistance increasing. Given this characteristic of lithium-ion polymer cells, completely discharging cells, otherwise known as deep discharge, results in damage to the cells, which may include the inability to hold a charge and sustain a voltage across a load. To guard against this damage, low voltage protection is required. While the figure illustrates the characteristics for a 1.0 mAh grade lithium-ion polymer cell, the characteristics are similar for other lithium-ion polymer cells.

FIG. **4** illustrates a block diagram of the harvester electronics **120** utilizing a two circuit configuration. The harvester electronics **120** are comprised of a thermal harvester interface **400**, a solar harvester interface **405**, a battery manager **410**, and a voltage regulator **415**. Each component comprising the harvester electronics **120** may be electrically connected to and powered by the battery **115**. Optionally, the components of the harvester electronics **120** may be powered by the energy from the solar or thermal harvesters. Optionally, a combination of energy sources may be used to power the harvester electronics **120**. For example, the harvester electronics **120** may be powered by energy from the solar harvester **105** and the battery **115**.

The solar harvester **105** is electrically connected to the solar harvester interface **405**. The solar energy harvested by the solar harvester **105** is transferred to the solar harvester interface **405**. Optionally, the solar harvester interface may comprise a voltage boost to allow harvesting of solar energy at low voltages. Optionally, the voltage boost may be dynamically triggered below a certain threshold. The solar harvester interface is electrically connected to the battery manager **410**.

The thermal harvester **110** is electrically connected to the thermal harvester interface **400**. The thermal energy harvested by the thermal harvester **110** is transferred through the thermal harvester interface **400**. Due to the low voltage output of thermal energy, the thermal harvester interface **400** comprises a voltage boost. For example, the voltage boost may boost voltages from as low as 20 mV. The thermal harvester interface **400** is electrically connected to the battery manager **410**.

The solar harvester interface **405** and the thermal harvester interface **400** transfer the harvested energy to the battery manager **410**. The battery manager **410** merges the harvested energy from the solar harvester interface **405** and the thermal harvester interface **400**. This merged energy may be used to charge the battery **115** and/or be supplied to the electrically connected voltage regulator **415**. The battery manager **410** controls the rate of charge for the battery **115**. The rate of charge varies depending on the battery requirements. The battery manager **410** may be configured to disconnect the battery **115** from the output load when the battery discharge reaches a voltage threshold where further discharge would be harmful to the battery. The voltage at which the battery manager **410** disconnects the battery varies depending on the battery specifications and configuration. For example, if the battery **115** is a lithium-ion polymer with discharge characteristics shown in FIG. **3**, the battery manager **410** will disconnect the battery from the load when the sharp voltage drop occurs. For example, but not to serve as a limitation, if the battery **115** is a 3.3 V lithium polymer battery, the battery manager **410** may be disconnect the battery from the load when the battery output voltage reaches 2.1 V. While the battery **115** is disconnected from the load, the battery manager **410** may continue to charge the battery **115** while providing power to the voltage regulator **415**.

Optionally, the battery manager **410** may include cell balancing capability when multiple battery cells are used. In the case where the battery cells have the same capacity, the battery manager **410** may differentially charge each cell to maximum capacity. In the case where the battery cells have varying capacities, the battery manager **410** may differentially charge each cell to maintain the same state of change.

The battery manager **410** may be configured to disconnect the battery **115** when the battery is fully charged to prevent overcharging. In the situation where the battery **115** is fully charged and disconnected from further charging, the battery manager **410** will direct all the harvested energy to the voltage regulator **415**. In the absence of adequate power from the solar harvester **105**, the battery manager **410** may utilize the battery **115** to supplement power from the thermal harvester **110** to the voltage regulator **415**.

The voltage regulator **415** maintains a constant voltage level output, which enables the harvester electronics **120** to produce a regulated power output **420**. The voltage regulator **415** may be comprised of a Low Drop Out (LDO) DC-DC voltage regulator. Optionally, other linear voltage regulators or switching regulators may be used. For example, but not to serve as a limitation, other linear voltage regulators may comprise zener regulators, series regulators, transistor regulators, or regulators comprising an op-amp. The voltage regulator **415** may be modular and configured to match a specific range of output loads or increase power output drive capability.

Optionally, the components of the harvesting electronics may be specifically configured to efficiently supply adequate power depending on the specific applications. While the embodiment of the invention contemplates the harvester electronics **120** comprised on a single chip, alternate embodiments may comprise each block or a combination of two or more blocks as discrete circuits.

FIG. **5** illustrates an alternate embodiment of the harvester electronics utilizing a two circuit configuration. The solar harvester **105**, thermal harvester **110**, battery **115**, thermal harvester interface **400**, solar harvester interface **405**, battery manager **410**, and voltage regulator **415** are as described above in FIG. **1** and FIG. **4**.

The harvester electronics comprises a bandgap reference **500**. The bandgap reference **500** is electrically connected to and powered by the battery **115**. Optionally, the bandgap reference may be powered separately by the battery manager, not shown. The bandgap reference **500** provides a temperature independent voltage reference (REF) **505**. The bandgap reference is electrically connected to the battery manager **410** and voltage regulator **415**, along with the MPPT controls and charge pump controls comprising the thermal harvester interface **400** and solar harvester interface **405** and outputs the REF **505** to each of the components.

The solar harvester interface **405** and thermal harvester interface **400** may he further comprised of individual maximum power point tracking (MPPT) inputs **510**, MPPT controls **515**, charge pumps **520**, and charge pump controls **525**. The different harvesting conditions for solar and thermal harvesters may result in different circuitry comprising the aforementioned components. The aforementioned components comprising the individual harvester interfaces may be powered by the battery **115**. Optionally, the components may be powered by the battery manager **410** or the solar or thermal harvester.

The MPPT input **510** is electrically connected to and receives input from the solar or thermal harvester and MPPT control **515**. The MPPT input **510** may track to match the input impedance to the maximum harvesting input impedance to achieve the greatest possible power harvest, during moment to moment variations of light level, shading, temperature, and photovoltaic module characteristics based on the input from the MPPT control **515**.

The MPPT control **515** is electrically connected to and receives input from the solar or thermal harvester and bandgap reference **500**. The input from the bandgap reference **500** comprises the REF **505**. The MPPT control **515** indicates to the MPPT input **510** whether the operating voltage of the harvesting input has changed in a given direction and to adjust the impedance accordingly. Optionally, the MPPT control **515** may determine the disturbance of the operating voltage of the harvesting input through a logic algorithm utilizing the REF **505** and harvester inputs. The MPPT input **510** is electrically connected to and provides an output to the battery manager **410**.

Optionally, the MPPT input **510** maybe electrically connected to the charge pump **520**. The charge pump **520** is electrically connected to and receives input, from the MPPT input **510** and the charge pump control **525**. The charge pump **520** may be comprised of a capacitive charge pump. Alternatively, the charge pump **520** may be comprised, of a step-up converter, boost converter, boost converter, another type of DC-DC converter, or a DC-AC converter coupled to a voltage multiplier. Optionally, the charge pump **520** may be comprised of multiple stages. For example, the charge pump **520** may pump from a voltage as low as 20 mV to over 4 V, using multiple stages to achieve the result.

The charge pump control **525** is electrically connected to and receives input from the bandgap reference **500** in the form of REF **505** Based on the REF **505**, the charge pump control **525** may adjust the oscillator frequency and the number of charge pump stages used by the charge pump **520**, which may be none, one, or multiple stages.

In an embodiment where the components of the harvester electronics **120** is powered by the battery **115**, the output from the MPPT input **510** to the battery manager **410** may serve as a bypass in the situation where the battery **115** goes below the harvester electronics **120** operating voltage. If the battery **115** goes below the harvester electronics **120** operating voltage, the components may fail to function. This bypass may allow the battery manager to recover the device with a voltage input that could drive the harvest electronics **120** and charge the battery **115** above the minimum operating voltage.

While the embodiment of the invention contemplates the harvester electronics **120** comprised on a single chip, alternate embodiments may comprise each block or a combination of two or more blocks as discrete circuits.

FIG. **6** illustrates a block diagram embodiment of the harvester electronics **120** in a single circuit configuration. The thermal harvester **110** is electrically connected to a voltage boost **600**. The voltage boost **600** may be comprised of a capacitive charge pump. Alternatively, the voltage boost **600** may be comprised of a step-up converter, boost converter, boost converter, another type of DC-DC converter, or a DC-AC converter coupled to a voltage multiplier. The voltage boost increases the low voltage input from the thermal harvester. The voltage boost **600** is electrically connected to the charge control **605**. The charge control **605** controls the energy storage for a first battery **610**. The boosted energy from the thermal harvester **110** is used to charge the first battery **610**. The charge control **605** controls the rate of charge and the battery voltage level of the first battery **610**. When the first battery **610** is fully charged, the charge control **605** disconnects the battery to prevent overcharging.

The first battery **610** may be comprised of any type of rechargeable battery. One embodiment of the invention contemplates the use of solid state thin film battery may be comprised of lithium-ion polymer, lithium, semiconductor, polymer, or flexible batteries. Optionally, the solid state thin film battery may be discretely or monolithically integrated with the harvester. Optionally, in a separate embodiment, bulk batteries may be used. For example, rechargeable lithium air, nickel-cadmium, or nickel metal-hydride batteries. Optionally, the first battery **610** may be comprised of single or multiple battery cells.

The first battery **610** is electrically connected to the power management block **615**. The power management block disconnects the first battery **610** from the output load when the battery when the battery discharge reaches a voltage threshold where further discharge would be harmful to the battery. For example, if the first battery **610** comprises a lithium-ion polymer battery with discharge characteristics similar to the depiction in FIG. **3**, the power management block **615** will disconnect the battery from the load when the sharp voltage drop occurs. Additionally, the power management **615** is electrically connected to a second battery **620**. The power management block **615** regulates the output voltage of the first battery **610** to charge the second battery **620**. Optionally, the charge control **605** and the power management block **615** may be combined into a single block to regulate the storage and output of the first battery **610**.

The second battery **620** may be comprised of any type of rechargeable battery. One embodiment of the invention contemplates the use of a solid state thin film battery comprised of lithium-ion polymer, lithium, semiconductor, polymer, or flexible batteries. Optionally, the solid state thin film battery may be discretely or monolithically integrated with the harvester. Optionally, in a separate embodiment, bulk batteries may be used. For example, rechargeable lithium air, nickel-cadmium, or nickel metal-hydride batteries. Optionally, the second battery **620** may be comprised of single or multiple battery cells. The second battery **620** is electrically connected to and receives solar energy input from the solar harvest regulator **625**.

The solar harvest regulator **625** is electrically connected to and receives input from the solar harvester **105**. Optionally, a zener protection diode may be used between the solar harvester **105** and solar harvest regulator **625** to protect the entire circuit. The solar harvest regulator **625** may comprise a mechanism to disconnect the second battery **620** once the battery reaches a full charge to prevent overcharging. Optionally, the solar harvest regulator **625** may be electrically connected to the voltage boost **600** and receives harvested thermal energy. Additionally, the solar harvest regulator **625** may comprise a voltage output regulator to disconnect the output when the battery discharge reaches a voltage threshold where further discharge would be harmful to the battery. Optionally, the solar harvest regulator may comprise a LDO regulator, another type of linear voltage regulator, or a switching regulator to control the output to a load. For example, but not to serve as a limitation, other linear voltage regulators may comprise zener regulators, series regulators, transistor regulators, or regulators comprising an op-amp. Optionally the solar harvest regulator **625** may protect the second battery **620** from unintended discharge when the input is lower than the battery voltage. For example, a zener protection diode may be used as a mechanism, where the voltage of the solar input must match the voltage of the battery plus the zener diode voltage.

While the embodiment of the invention contemplates the harvester electronics **120** comprised on a single chip, alternate embodiments may comprise each block or a combination of two or more blocks as discrete circuits.

FIG. **7** illustrates an alternate block diagram embodiment of the harvester electronics **120** in a single circuit configuration. The harvester electronics comprises a bandgap reference **700**. The bandgap reference **700** is electrically connected to and powered by the second battery **620**. Optionally, the bandgap reference may be powered separately by the solar harvest regulator **625**, not shown. The bandgap reference **700** provides a temperature independent voltage reference (REF) **705** output. The bandgap reference is electrically connected to the solar harvest regulator **625** and the MPPT control **715** for the thermal and solar harvester inputs and outputs the REF **505** to each of the components.

The MPPT input **710** is electrically connected to and receives input from the solar or thermal harvester and MPPT control **715**. The MPPT input **710** may track to match the input impedance to the maximum harvesting input impedance to achieve the greatest possible power harvest, during moment to moment variations of light level, shading, temperature, and photovoltaic module characteristics based on the input from the MPPT control **715**.

The MPPT control **715** is electrically connected to and receives input from the solar or thermal harvester and bandgap reference **700**. The input from the bandgap reference **700** comprises the REF **705**. The MPPT control **715** indicates to the MPPT input **710** whether the operating voltage of the harvesting input has changed in a given direction and to adjust the impedance accordingly. Optionally, the MPPT control **715** may determine the disturbance of the operating voltage of the harvesting input through a logic algorithm utilizing the REF **705** and harvester inputs.

Optionally, in another embodiment of the invention, a second bandgap reference, not shown, may be connected to the charge control **605** and the first battery **610**. The second bandgap reference is electrically connected to and provides a reference output to the MPPT control **715** connected to the thermal harvester **110**. This embodiment allows for individual MPPT tracking for different sources of harvested energy.

While the embodiment of the invention contemplates the harvester electronics **120** comprised on a single chip, alternate embodiments may comprise each block or at combination of two or more blocks as discrete circuits.

While the aforementioned embodiments depict harvesting, electronics which control the storage of energy and regulate output to a load for solar and thermal harvesters, the harvesting electronics may be configured to receive input from multiple sources of energy harvesting. Examples of other sources may include, but are not limited to air flow, temperature gradient, movement and transmitted radio frequencies.

Whereas many alterations and modifications of the present invention will no doubt become apparent to a person of ordinary skill in the art after having read the foregoing description, it is to be understood that the particular embodiments shown and described by way of illustration are in no way intended to be considered limiting. Therefore, reference to the details of the preferred embodiments is not intended to limit their scope.

The embodiments were chosen and described in order to explain the principles and applications of the invention, thereby allowing others skilled in the art to utilize the invention in its various embodiments and modifications according to the particular purpose contemplated. The scope of the invention is intended to be defined by the claims appended hereto and their equivalents.

The expected practical use of the present invention is the novel integrated perpetual energy harvester with energy storage and harvesting electronics. The proposed invention may be used as a one chip solution to store power from multiple sources of energy harvesting and regulate the output to a load.

What is claimed is:

1. A perpetual energy harvester comprising:

a photovoltaic thermal harvester wherein the photovoltaic thermal harvester generates the electrical energy from the blackbody radiation; and

a harvesting electronics in close proximity to the photovoltaic thermal harvester,

wherein the harvesting electronics receives input in the form of energy from the photovoltaic thermal harvester, and regulates the output to a load.

2. The perpetual energy harvester of claim **1**, further comprising a solar harvester, wherein said solar harvester is individually electrically connected to the harvesting electronics, and wherein said harvesting electronics further receives input in the form of energy from the solar harvester.

3. The perpetual energy harvester of claim **1**, further comprising a battery, wherein said battery is individually electrically connected to the harvesting electronics, and wherein said harvesting electronics further controls the amount of energy stored in said battery.

4. The perpetual energy harvester in claim **3**, wherein the said battery is selected from the group consisting of thin film lithium-ion, lithium-ion polymer, polymer, semiconductor, and flexible batteries.

5. The perpetual energy harvester in claim **2**, wherein the solar harvester is selected from the group consisting of traditional crystallized silicon solar cells, thin film solar cells, multi-layered solar cells, dye sensitized solar cells, and organic polymer solar cells.

6. The perpetual energy harvester in claim **1**,

wherein the photovoltaic thermal harvester is comprised of material selected from the group consisting of HgCATe, HgZnTe, InSb, InAs, GaSh, GaAs, InP, PbTe, and polymers, and wherein the harvesting electronics is comprised of a thermal harvester interface, and wherein the thermal harvester interface is electrically connected to and receives input from the thermal harvester.

7. A perpetual energy harvester comprising:

a photovoltaic thermal harvester wherein the photovoltaic thermal harvester generates the electrical enemy from the blackbody radiation;

a battery in close proximity to the photovoltaic thermal harvester; and

a harvesting electronics in close proximity to the photovoltaic thermal harvester and individually electrically connected to the photovoltaic thermal harvester, and to the battery, wherein the harvesting electronics receives input in the form of energy from the photovoltaic thermal harvester, controls the amount of energy stored in the battery, and regulates the output to a load.

8. The perpetual energy harvester of claim **7**, further comprising as solar harvester, wherein said solar harvester is individually electrically connected to the harvesting electronics, and wherein said harvesting electronics further receives input in the form of energy from the solar harvester.

9. The perpetual energy harvester in claim **7**, wherein the said battery is selected from the group consisting of thin film lithium-ion, lithium-ion polymer, polymer, semiconductor, and flexible batteries.

10. The perpetual energy harvester in claim **8**, wherein the solar harvester is selected from the group consisting of traditional crystallized silicon solar cells, thin film. solar cells, multi-layered solar cells, dye sensitized solar cells, and organic polymer solar cells.

11. The perpetual energy harvester in claim **7**, wherein the photovoltaic thermal harvester is comprised of material selected from the group consisting of HgCdTe, HgZnTe, InSb, InAs, GaSb, GaAs, InP, PbTe, and polymers, and wherein the harvesting electronics is comprised of a thermal harvester interface, and wherein the thermal harvester interface is electrically connected to and receives input from the thermal harvester.

12. A perpetual energy harvester comprising;

a solar harvester;

a photovoltaic thermal harvester in close proximity to the solar harvester, wherein the photovoltaic thermal harvester generates the electrical energy from the blackbody radiation;

and a harvesting electronics in close proximity to the photovoltaic thermal harvester and individually electrically connected to the solar harvester and to the photovoltaic thermal harvester, wherein the harvesting electronics receives input in the form of energy from the solar harvester and the photovoltaic thermal harvester, and regulates the output to a load.

13. The perpetual energy harvester of claim **12**, further comprising a battery, wherein said battery is individually electrically connected to the harvesting, electronics, and wherein said harvesting electronics further controls the amount of energy stored in said battery.

14. The perpetual energy harvester in claim **13**, wherein the said battery is selected from the group consisting of thin film lithium-ion, lithium-km polymer, polymer, semiconductor, and flexible batteries.

15. The perpetual energy harvester in claim **12**, wherein the solar harvester is selected from the group consisting of traditional crystallized silicon solar cells, thin film solar cells, multilayered solar cells, dye sensitized solar cells, and organic polymer solar cells.

16. The perpetual energy harvester in claim **12**, wherein the photovoltaic thermal harvester is comprised of material selected from the group consisting of HgCdTe, HgZnTe InSb, InAs, GaSb, GaAs, InP, PbTe, and polymers, and wherein the harvesting electronics is comprised of a thermal harvester interface wherein the thermal harvesting interface is electrically connected to and receives input from the thermal harvester, and also a solar harvester interface, wherein the solar harvester interface is electrically connected to and receives input from the solar harvester.

17. The perpetual energy harvester in claim **12**, wherein said harvesting electronics comprises:

a thermal harvester interface, wherein the thermal harvester interface is electrically connected to and receives input from the thermal harvester;

a solar harvester interface, wherein the solar harvester interface is electrically connected to and receives input from the solar harvester;

a battery manager electrically connected to and receives input from the solar harvester interface and thermal harvester interface, wherein the battery manager is electrically connected to the battery and manages the energy stored in the battery;

and a voltage regulator electrically connected to the battery manager, wherein the voltage regulator receives an input from the battery manager and controls an output voltage to an external load.

18. The perpetual energy harvester in claim **17**, wherein said voltage regulator further comprises a low drop out regulator.

19. The perpetual energy harvester in claim **17**, wherein said harvesting electronics further comprises a bandgap reference electrically connected to the thermal harvester interface, the solar harvester interface, the battery manager, and the voltage regulator, wherein the bandgap reference outputs a temperature independent voltage reference to the thermal harvester interface, the solar harvester interface, the battery manager, and the voltage regulator.

20. The perpetual energy harvester in claim **17**, wherein the solar harvester interface comprises:

a Maximum Power Point Tracking (MPPT) input electrically connected the solar harvester and the battery manager;

MPPT control electrically connected to the solar harvester, the MPPT input, and the bandgap reference, wherein the MPPT control receives a voltage reference input from the bandgap reference and an input from the solar harvester, determines if the operating voltage has changed, and outputs instructions to the MPPT input which adjusts input impedance to match maximum harvesting, input impedance;

a charge pump electrically connected to the MPPT input and the battery manager; and a charge pump control electrically connected to the charge pump and the bandgap reference, wherein the charge pump control receives input from the bandgap reference and dynamically adjusts oscillation frequency for the charge pump, and wherein the thermal harvester interface comprises;

a Maximum Power Point Tracking (MITT) input electrically connected to the thermal harvester and the battery manager;

a MPPT control electrically connected to the thermal harvester, the MPPT input, and the bandgap reference, wherein the MPPT control receives a voltage reference input from the bandgap reference and an input from the thermal harvester, determines if the operating voltage has changed, and outputs instructions to the MPPT input which adjusts input impedance to match maximum harvesting input impedance;

a charge pump electrically connected to the MPPT input and the battery manager;

and a charge pump control electrically connected to the charge pump and the bandgap reference, wherein the charge pump control receives input from the bandgap reference and dynamically adjusts oscillation frequency for the charge pump.

* * * * *

US009557547B2

(12) **United States Patent**
Choi et al.

(10) **Patent No.: US 9,557,547 B2**
(45) **Date of Patent: Jan. 31, 2017**

(54) **PARAXIAL CLOAK DESIGN AND DEVICE**

(71) Applicants: **Joseph S. Choi**, Rochester, NY (US); **John C. Howell**, Rochester, NY (US)

(72) Inventors: **Joseph S. Choi**, Rochester, NY (US); **John C. Howell**, Rochester, NY (US)

(73) Assignee: **University of Rochester**, Rochester, NY (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **14/714,671**

(22) Filed: **May 18, 2015**

(65) **Prior Publication Data**

US 2016/0025956 A1 Jan. 28, 2016

Related U.S. Application Data

(60) Provisional application No. 62/028,658, filed on Jul. 24, 2014.

(51) **Int. Cl.**

G02B 17/00	(2006.01)
G02B 1/00	(2006.01)
G02B 17/08	(2006.01)
G02B 27/00	(2006.01)
G06F 17/50	(2006.01)

(52) **U.S. Cl.**
CPC ***G02B 17/008*** (2013.01); ***G02B 1/007*** (2013.01); ***G02B 17/08*** (2013.01); ***G02B 27/0012*** (2013.01); ***G06F 17/50*** (2013.01)

(58) **Field of Classification Search**
CPC G02B 7/007; G02B 17/08
USPC 359/726–731, 733–735
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,288,148 A 9/1981 Offner
5,307,162 A 4/1994 Schowengerdt
2003/0227415 A1* 12/2003 Joannopoulos B82Y 20/00 343/754
2006/0171032 A1* 8/2006 Nishioka B82Y 20/00 359/566
2009/0310231 A1* 12/2009 Maekawa G02B 1/007 359/733

(Continued)

OTHER PUBLICATIONS

Chanda et al., "Large-area flexible 3d optical negative index metamaterial formed by nanotransfer printing," Nat. Nanotechnol. 6, 402-407 (2011).

(Continued)

Primary Examiner — Darryl J Collins
(74) *Attorney, Agent, or Firm* — Kilpatrick Townsend & Stockton, LLP

(57) **ABSTRACT**

A paraxial cloaking device provides a cloaking volume in which an item can be hid from view. A cloaking device includes an optical input receiving light rays and an optical output from which a continuous range of directions of the received light rays exit the paraxial cloaking device. The cloaking volume being disposed between the optical input and the optical output. For received light rays having incoming directions non-parallel to the reference optical axis up to a first angle, each of the received light rays exits the cloaking device substantially aligned with the respective received light ray and does not pass through the cloaking volume. The paraxial cloaking device has a unity magnification factor. In some instances, the paraxial cloaking device includes a phase matching element.

29 Claims, 21 Drawing Sheets

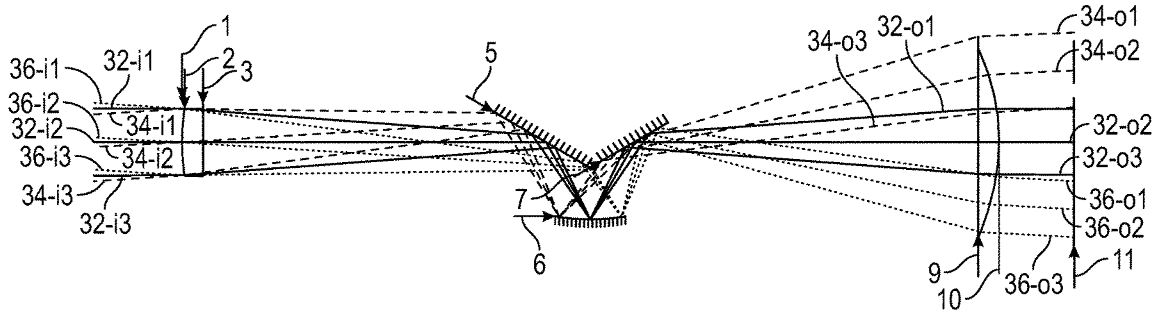

(56) References Cited

U.S. PATENT DOCUMENTS

2009/0316279	A1*	12/2009	Bowers	B82Y 20/00 359/754
2010/0265552	A1*	10/2010	Wang	B82Y 20/00 359/9
2012/0268836	A1*	10/2012	Mercado	G02B 17/0852 359/731

OTHER PUBLICATIONS

Chen, eta al., “Ray-Optics Cloaking Devices for Large Objects in Incoherent Natural Light” Nature Communications, 2013, 4: 1-6.

Chen, et al., “Macroscopic Invisibility Cloaking of Visible Light” Nature Communications, 2011, 2: 1-6.

Chen et al., “Extending the bandwidth of electromagnetic cloaks,” Phys. Rev. B 76, 241104 (2007).

Choi et al., “Paraxial Full-Field Cloaking”, The Institute of Optics, 2014.

Choi et al., “Paraxial ray optics cloaking,” Opt. Express 22, 29465-29478 (2014).

Clark, “1990 International Lens Design Conference Lens Design Problems: the design of a NonLens”, SPIE, 1990, 1354: 558-569.

Collins, “Lens-system diffraction integral written in terms of matrix optics,” J. Opt. Soc. Am. 60, 1168-1177 (1970).

Costa et al., “Achromatic lens based on a nanowire material with anomalous dispersion,” Opt. Express 20, 13915-13922 (2012).

Devaney, “Nonuniqueness in the Inverse Scattering Problem”, J. Mathematical Physics, 1978, 19: 1526.

Duan et al., “Geometric optics-based multiband cloaking of large objects with the wave phase and amplitude preservation,” Opt. Express 22, 27193-27202 (2014).

Ergin et al. “Three-dimensional invisibility cloak at optical wavelengths,” Science 328, 337-339 (2010).

Fleury et al., “Cloaking and invisibility: a review,” Prog. Electrornagn. Res. 147, 171-202 (2014).

Fridman, “Demonstration of Temporal Cloaking”, Nature, 2012, 481: 62-65.

Gbur, “Invisibility Physics: Past, Present Future”, Progress in Optics, 2013, 58:65-114 (2013).

Greenleaf et al., “Anisotropic conductivities that cannot be detected by EIT,” Physiol. Meas. 24,413-419 (2003).

Greenleaf et al., “Full-wave invisibility of active devices at all frequencies, ” Commun. Math. Phys. 275, 749-789 (2007).

Greenleaf et al., “Isotropic transformation optics: approximate acoustic and quantum cloaking,” New J. Phys. 10, 115024 (2008).

Hashemi et al., “Diameter-bandwidth product limitation of isolated-object cloaking,” Phys. Rev. A 86, 013804 (2012).

Howell, “Amplitude-only, Passive, Broadband, Optical Spatial Cloaking of Very Large Optics”, Applied Optics, 2014, 53: 1958-1963.

Kildishev et al., “Transformation optics: approaching broadband electromagnetic cloaking,” New J. Phys. 10, 115029 (2008).

Leonhardt, “Optical Conformal Mapping”, Science, 2006, 312: 1777-1780.

Leonhardt et al., Broadband invisibility by non-euclidean cloaking, Science 32 110-112 (2009).

Li et al., “Hiding under the Carpet: A New Strategy for Cloaking” Physical Review Letters, 2008 101: 1-4.

Ma, et al., “Experiments on active cloaking and illusion for laplace equation,” Phys. Rev. Lett. 111, 173901 (2013).

McCall, “Transformation Optics and Cloaking”, Contemporary Physics 54: 273-286 , (2013).

Miller, “On perfect cloaking,” Opt. Express 14, 12457-12466 (2006).

Monticone et al., “Physical bounds on electromagnetic invisibility and the potential of superconducting cloaks,” Photonics Nanostruct. 12, 330-339 (2014).

Monticone et al., “Do cloaked objects really scatter less?” Phys. Rev. X 3, 041005 (2013).

Nachman, “Reconstructions from Boundary Measurements”, Annals of Mathematics, 1988, 128: 531-576.

Pendry et al., “Controlling Electromagnetic Fields” Science, 2006, 312:1780-1782.

Shelby et al., “Experimental verification of a negative index of refraction,” Science 292, 77-79 (2001).

Silveirinha, “Anomalous refraction of light colors by a metamaterial prism,” Phys. Rev. Left. 102, 193903 (2009).

Soukoulis et al., “Past achievements and future challenges in the development of three-dimensional photonic metamaterials,” Nat. Photonics 5, 523-530 (2011).

Theisen et al., “Optical Properties of Gallium Implanted Silicon,” in Frontiers in Optics 2012/Laser Science XXVIII, OSA Technical Digest (online) (Optical Society of America, 2012), p. FTu4A.3.

Valentine et al., “Three-dimensional optical metamaterial with a negative refractive index,” Nature 455, 376-379 (2008).

Veselago, “Electrodynamics of substances with simultaneously negative values of sigma and mu,” Phys.-Usp. 10, 509-514 (1968).

Wolf et al., “Invisible bodies and uniqueness of the inverse scattering problem,” Journal of Modern Optics, 40:785-792 , 1993.

Zhang et al., “Macroscopic Invisibility Cloak for Visible Light”, Physical Review Letters, 2011, 106: 033901-1-033901-4.

* cited by examiner

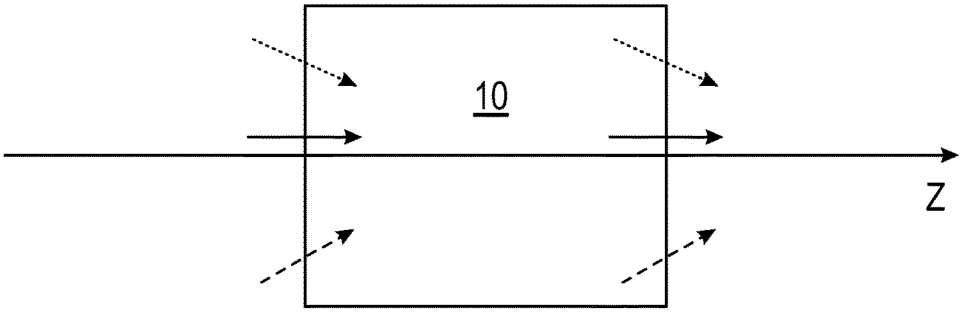

FIG. 1A

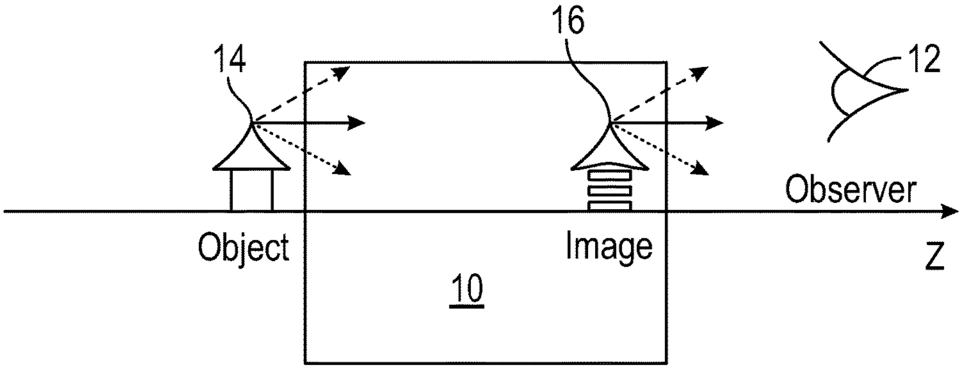
16
14
12
Observer
Object
Image
Z
10

FIG. 1B

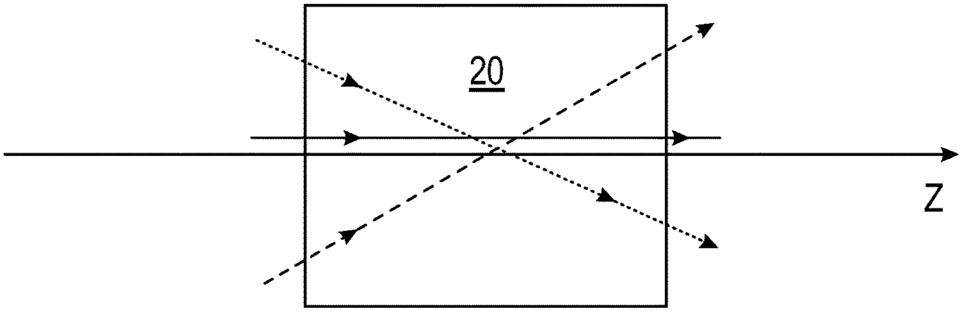

FIG. 1C

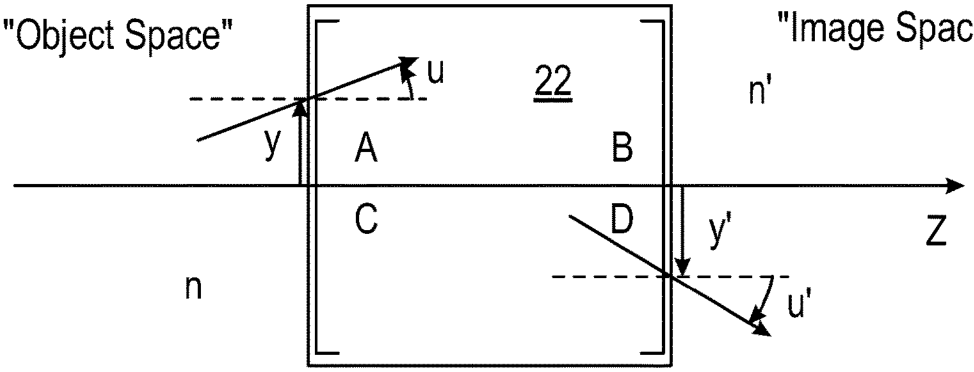
"Object Space"
u
22
"Image Space"
n'
y
A
B
C
D
y'
Z
n
u'

FIG. 2

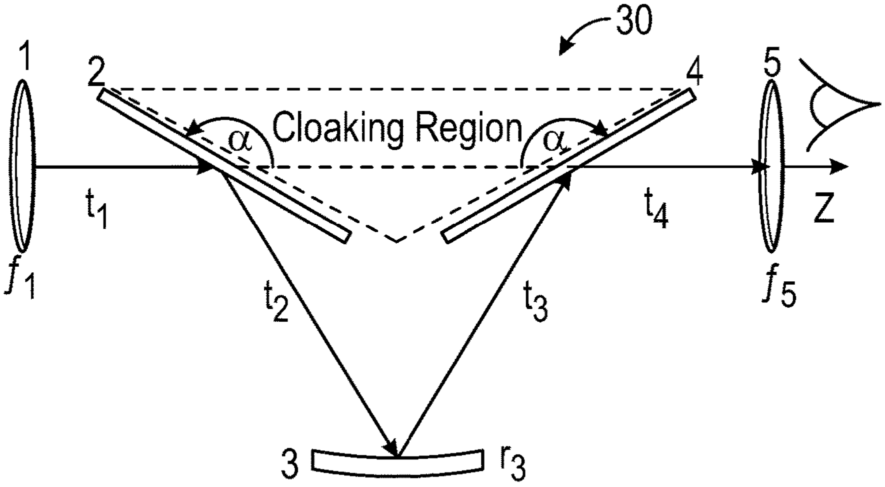
30
1
2
4
5
Cloaking Region
α
α
t1
t4
Z
f1
f5
t2
t3
3
r3

FIG. 3

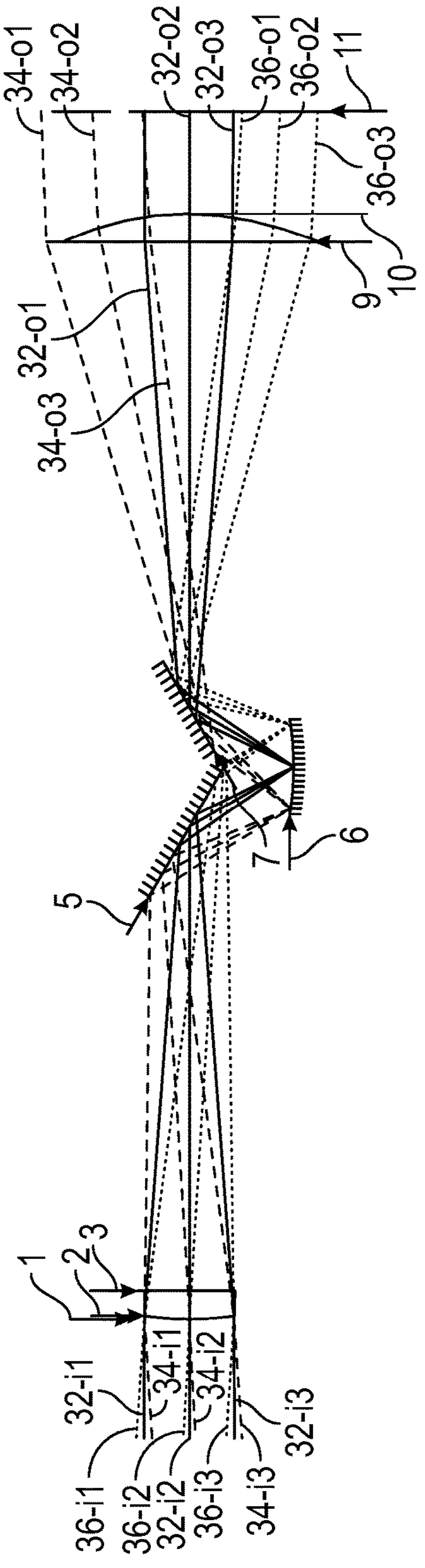
34-o1
34-o2
32-o2
32-o3
36-o1
36-o2
11
36-o3
10
9
32-o1
34-o3
5
7
6
1
2
3
32-i1
34-i1
34-i2
32-i3
36-i1
36-i2
32-i2
36-i3
34-i3

FIG. 4

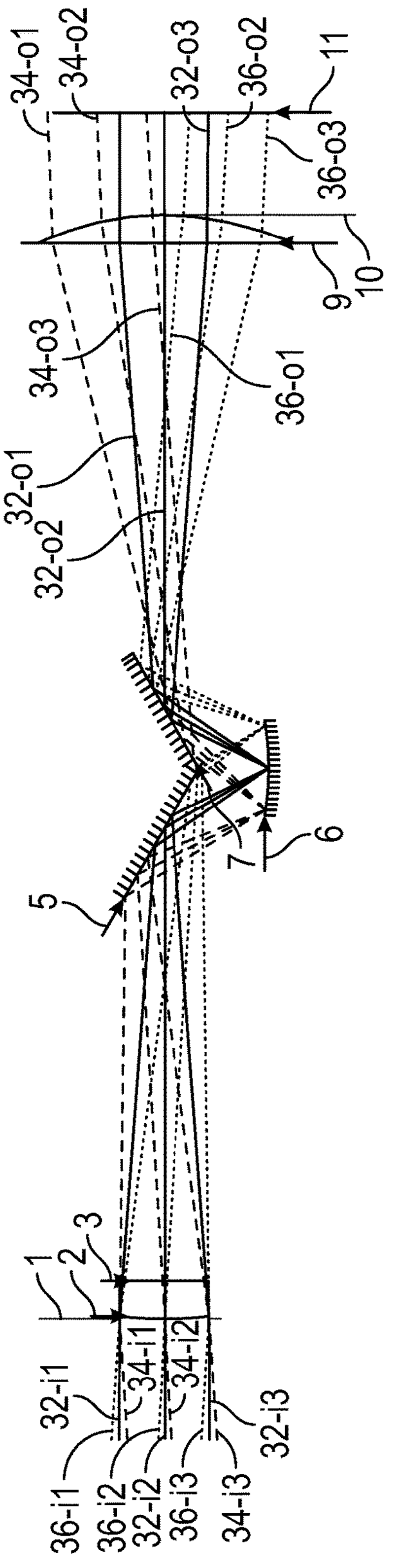
34-o1
34-o2
32-o3
36-o2
11
36-o3
10
9
34-o3
36-o1
32-o1
32-o2
5
7
6
1
2
3
32-i1
34-i1
34-i2
32-i3
36-i1
36-i2
32-i2
36-i3
34-i3

FIG. 5

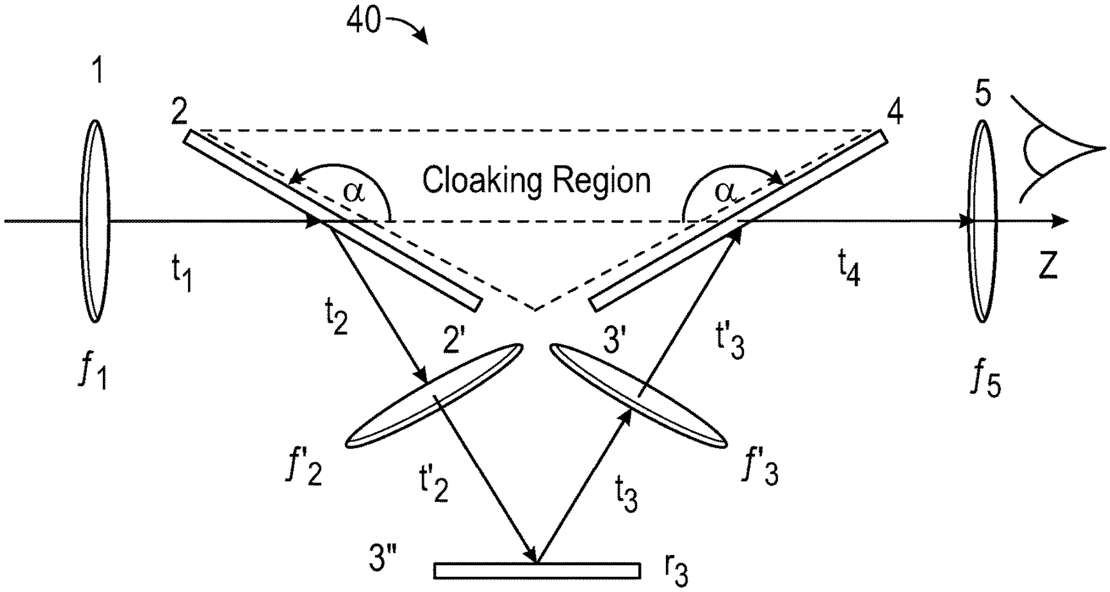
40
1
2
4
5
Cloaking Region
α
α
t_1
t_4
Z
f_1
f_5
t_2
2'
3'
t'_3
f'_2
t'_2
t_3
f'_3
3"
r_3

FIG. 6

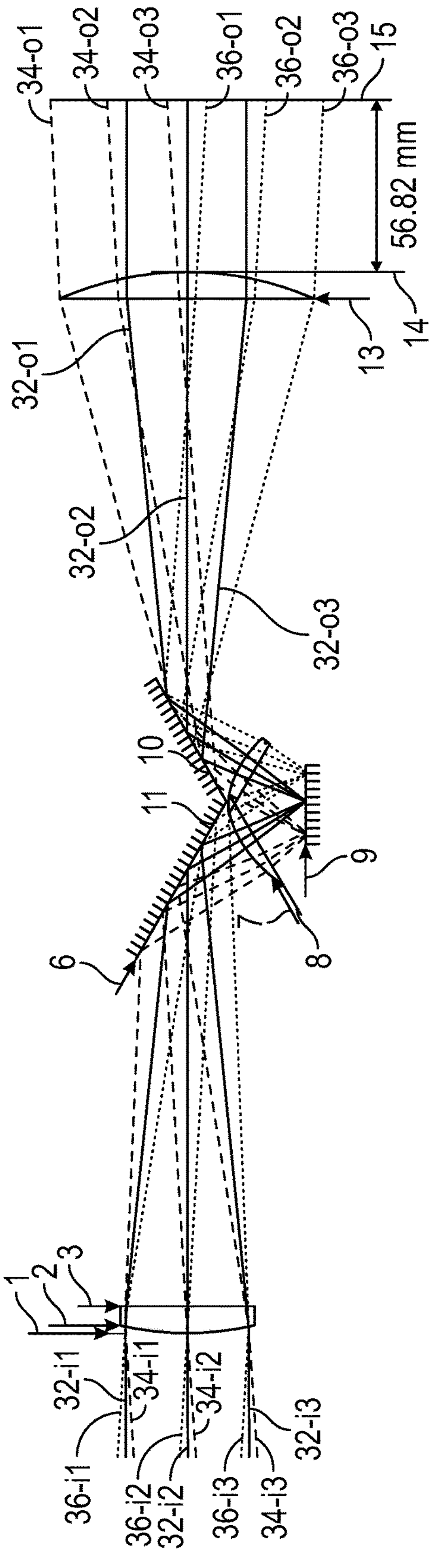
34-o1
34-o2
34-o3
36-o1
36-o2
36-o3
15
56.82 mm
14
13
32-o1
32-o2
32-o3
10
11
9
8
7
6
1
2
3
32-i1
34-i1
34-i2
32-i3
36-i1
36-i2
32-i2
36-i3
34-i3

FIG. 7

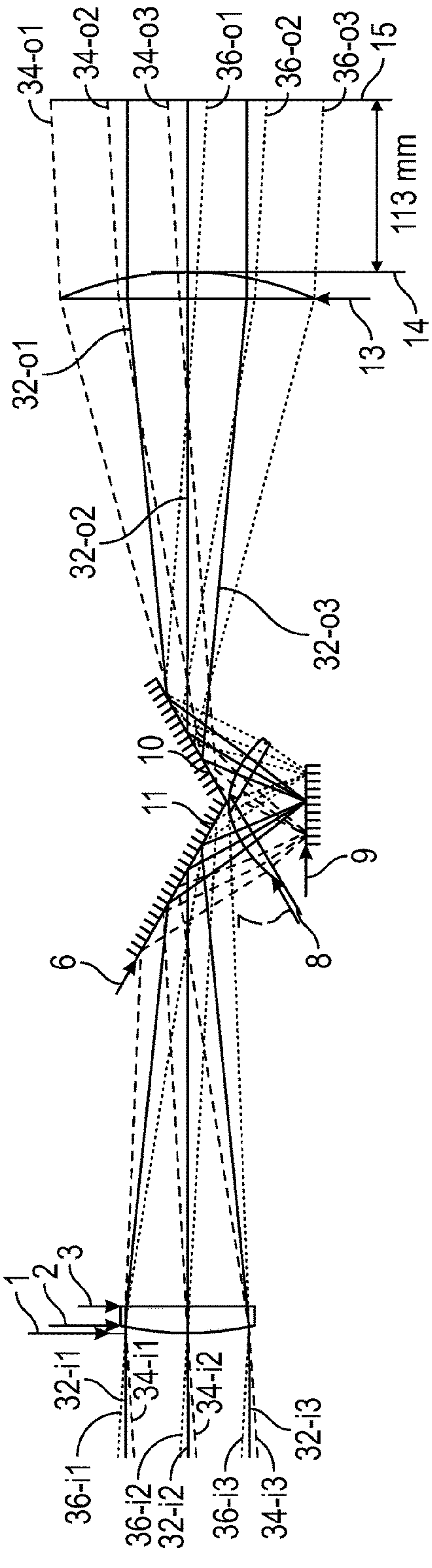
34-o1
34-o2
34-o3
36-o1
36-o2
36-o3
15
113 mm
14
13
32-o1
32-o2
32-o3
10
11
9
8
7
6
1
2
3
32-i1
34-i1
34-i2
32-i3
36-i1
36-i2
32-i2
36-i3
34-i3

FIG. 8

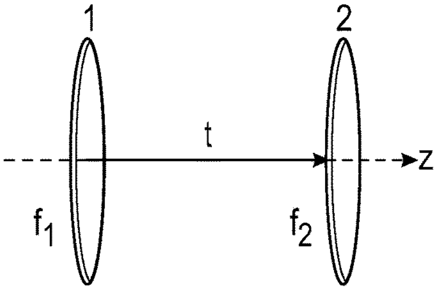

FIG. 9

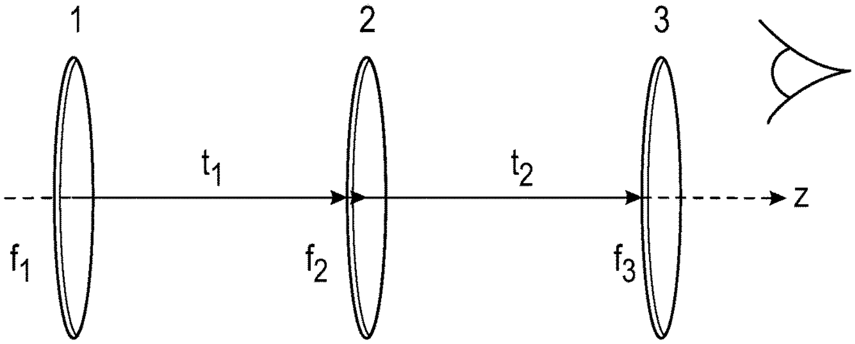
1
2
3
t_1
t_2
z
f_1
f_2
f_3

FIG. 10

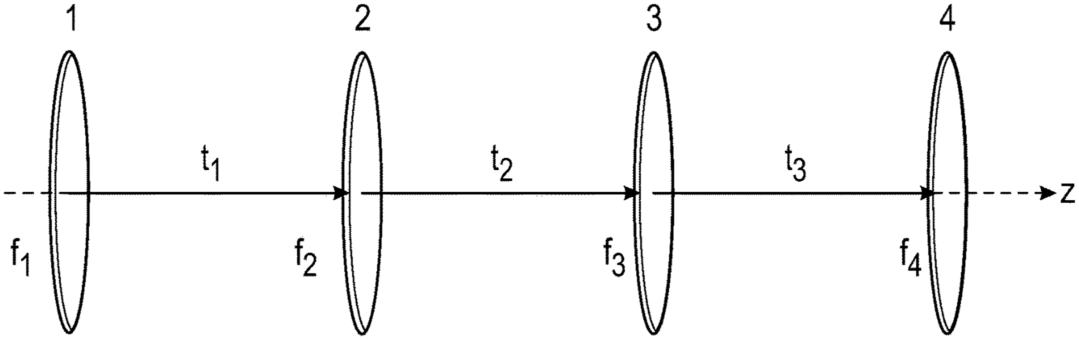
1
2
3
4
t_1
t_2
t_3
z
f_1
f_2
f_3
f_4

FIG. 11

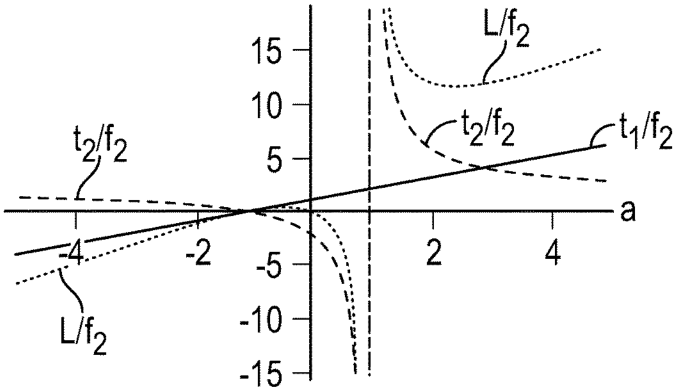

FIG. 12

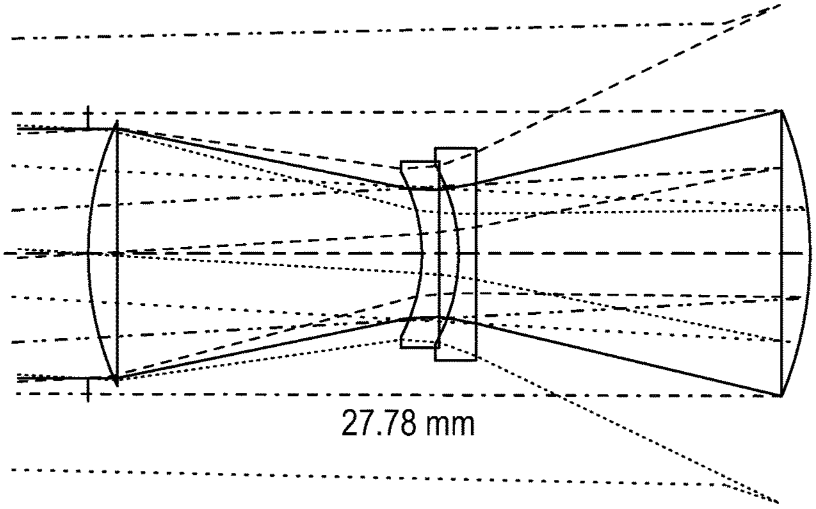

FIG. 13A

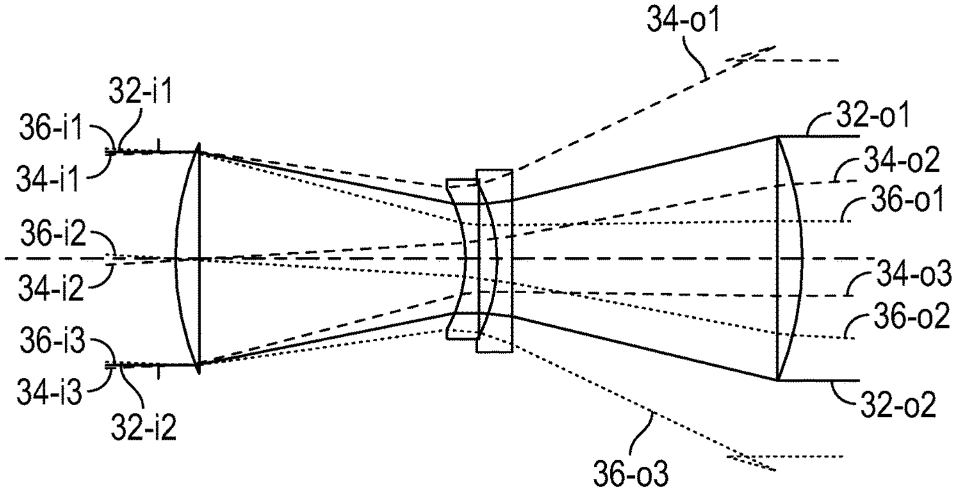
34-o1
32-i1
36-i1
34-i1
32-o1
34-o2
36-o1
36-i2
34-i2
34-o3
36-o2
36-i3
34-i3
32-i2
32-o2
36-o3

FIG. 13B

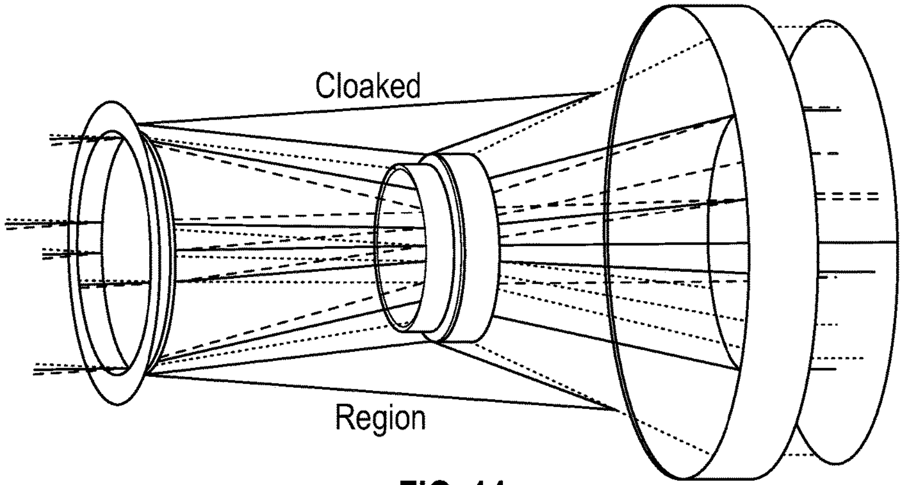
Cloaked
Region

FIG. 14

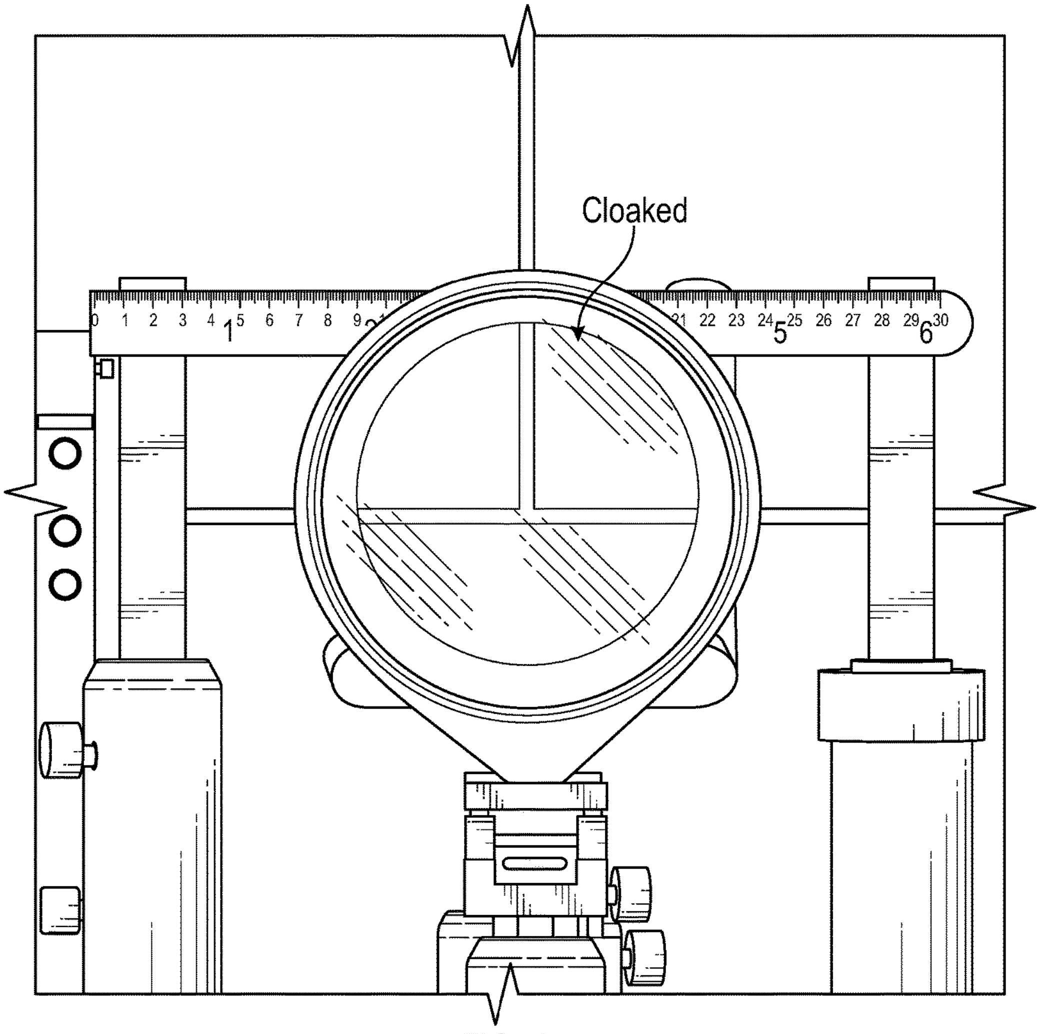
Cloaked
0 1 2 3 4 5 6 7 8 9
21 22 23 24 25 26 27 28 29 30
1
5
6

FIG. 15

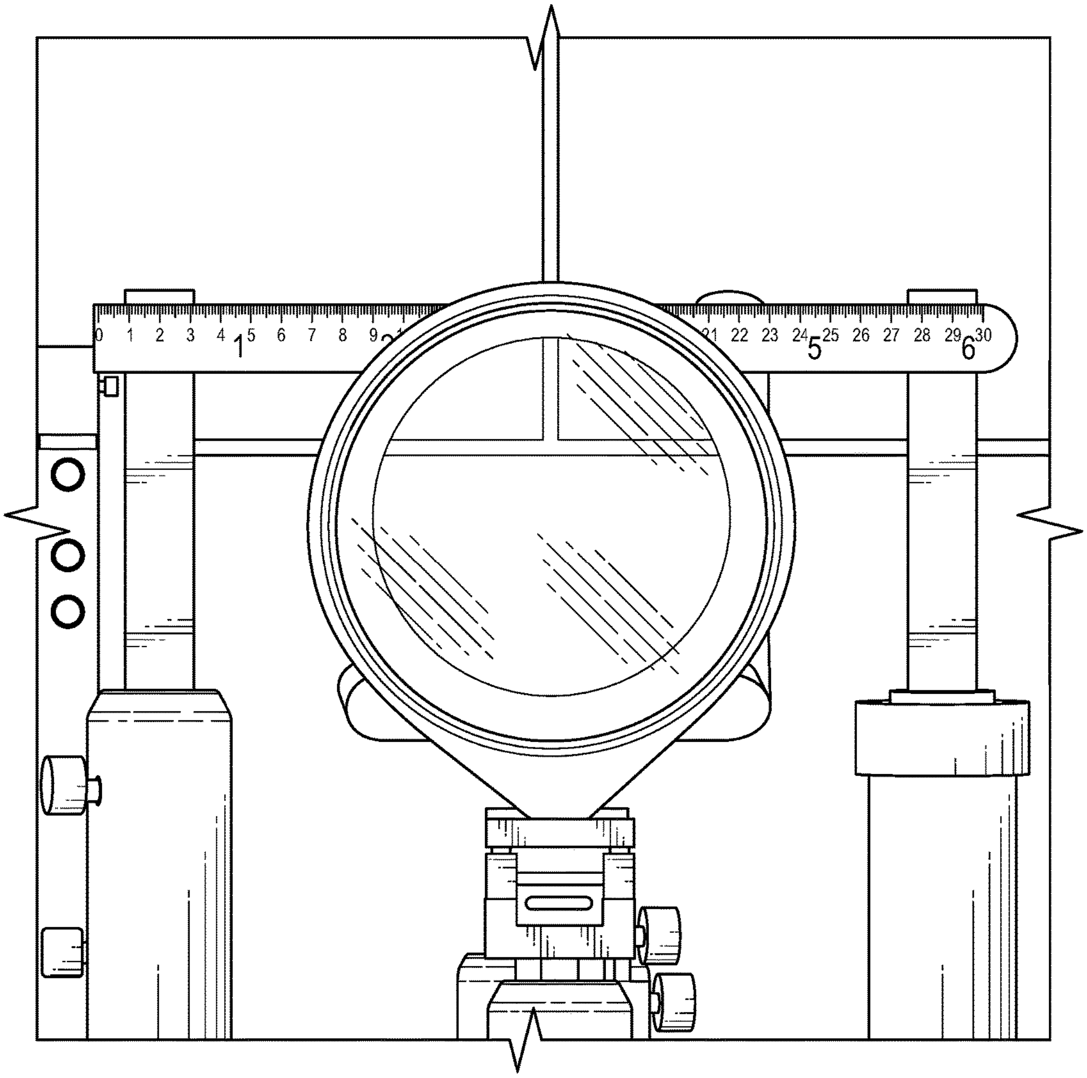
0 1 2 3 4 5 6 7 8 9
1
21 22 23 24 25 26 27 28 29 30
5
6

FIG. 16

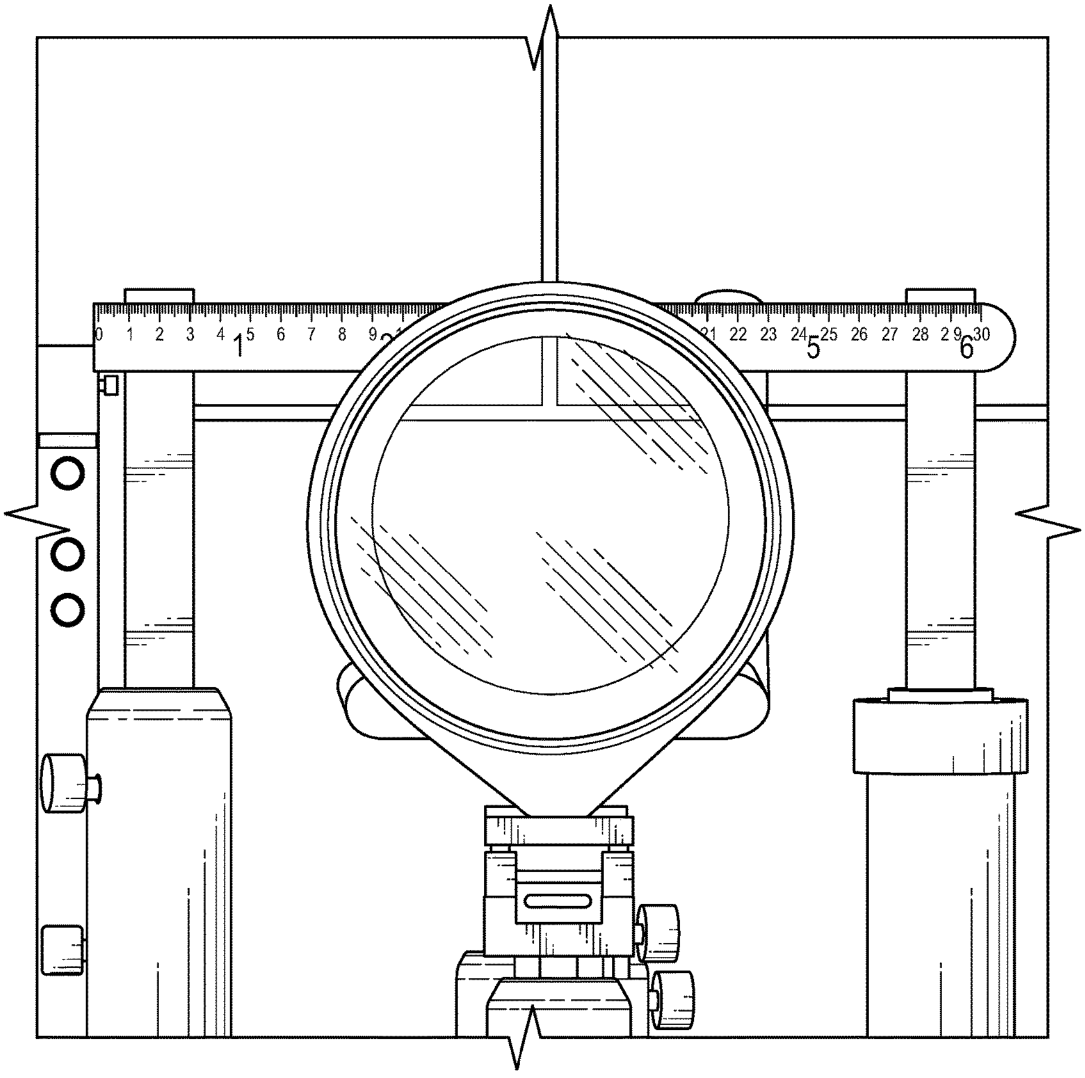
0 1 2 3 4 5 6 7 8 9
21 22 23 24 25 26 27 28 29 30
1
5
6

FIG. 17

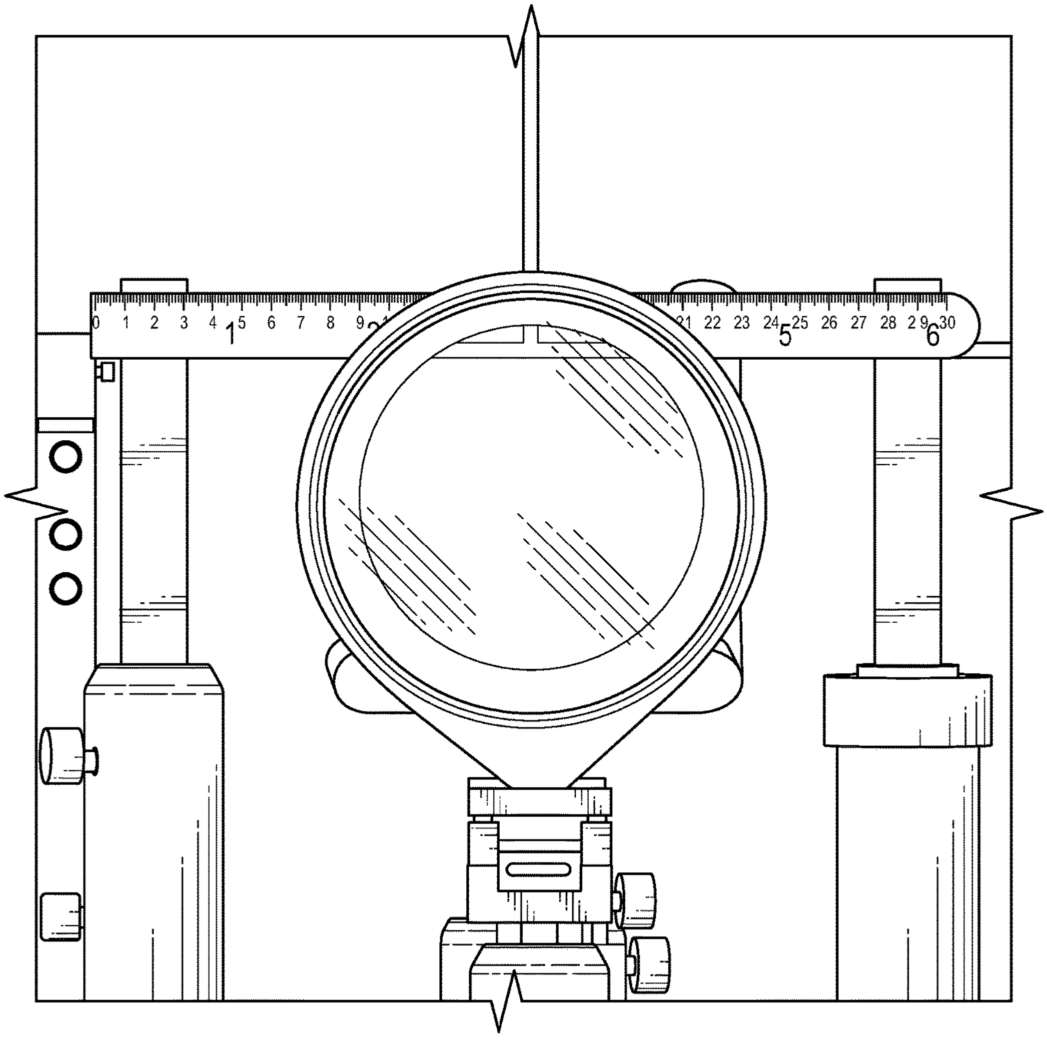
0 1 2 3 4 5 6 7 8 9
21 22 23 24 25 26 27 28 29 30
1
5
6

FIG. 18

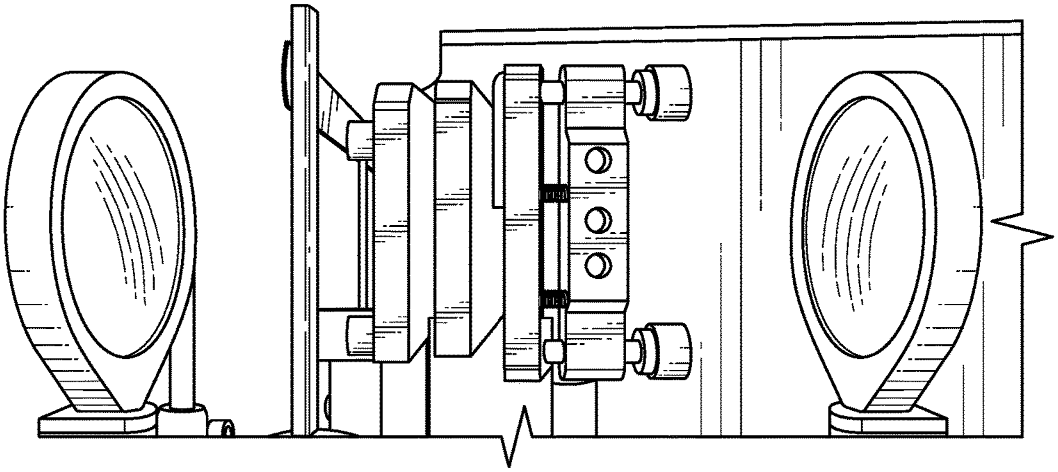

FIG. 19

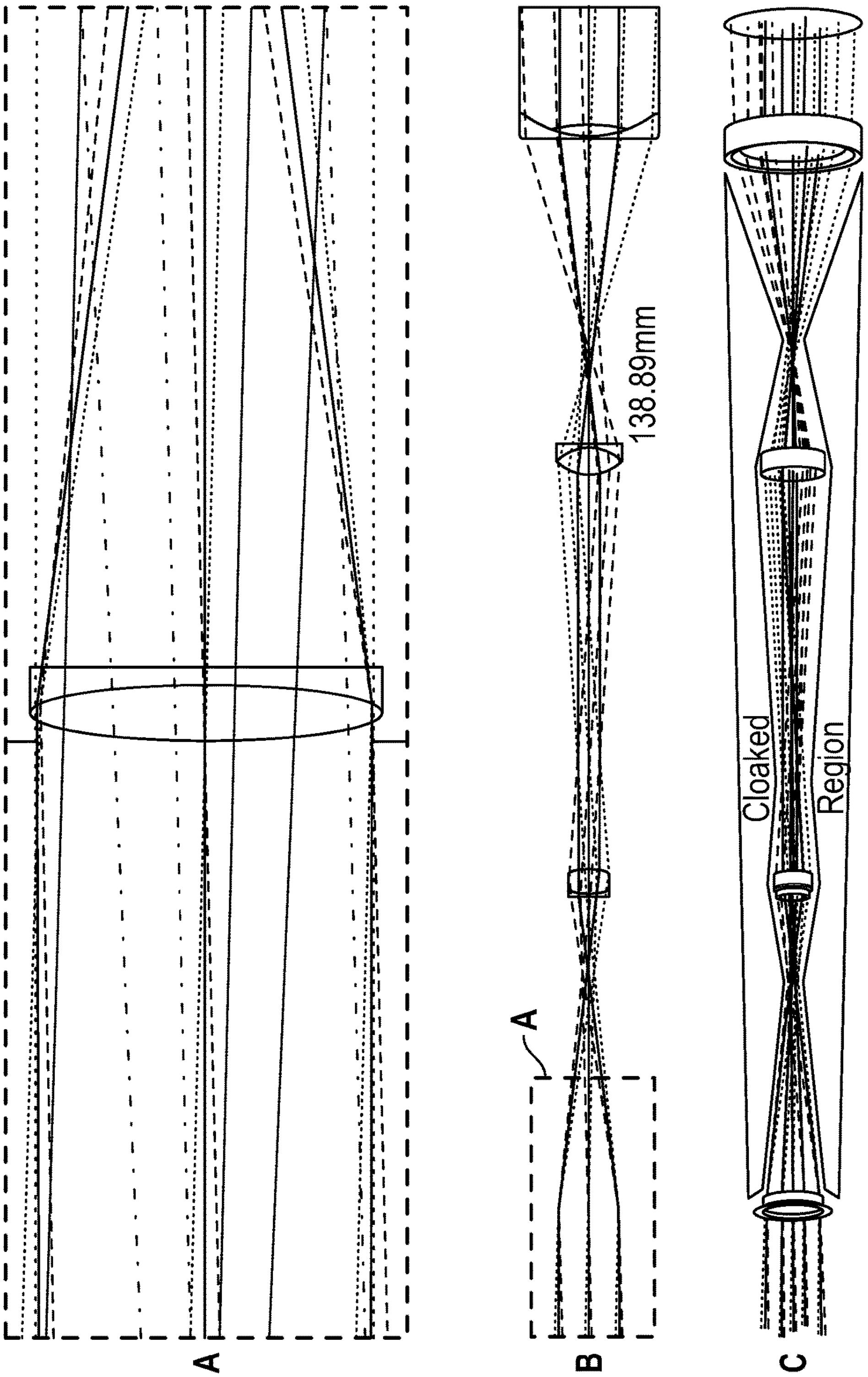
A
B
C
A
138.89mm
Cloaked
Region

FIG. 20

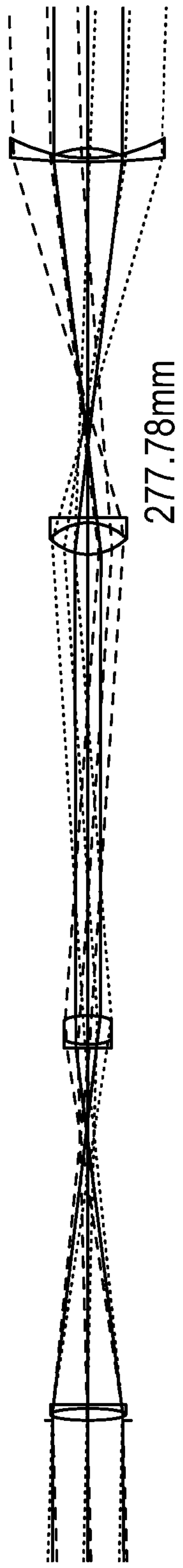

FIG. 21

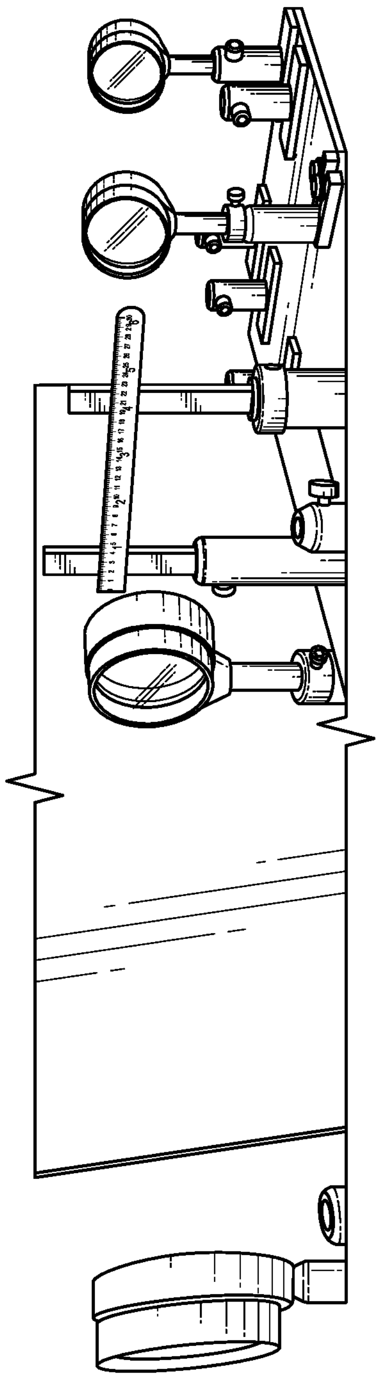

FIG. 22

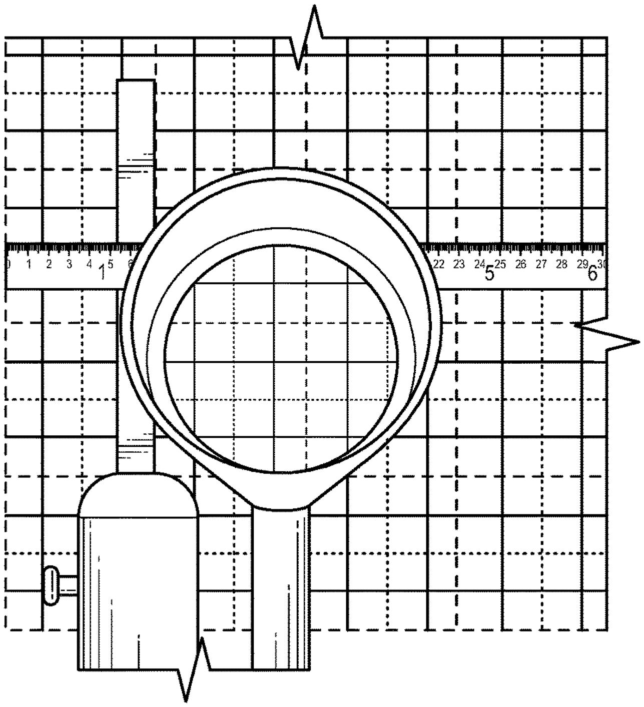
0 1 2 3 4 5 6
22 23 24 25 26 27 28 29 30

FIG. 23

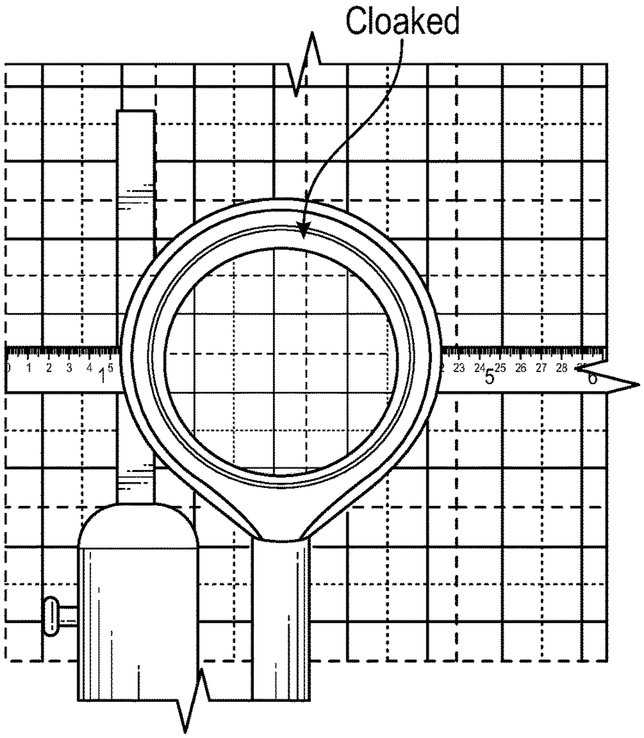
Cloaked
1 2 3 4 5
1
23 24 25 26 27 28
5
6

FIG. 24

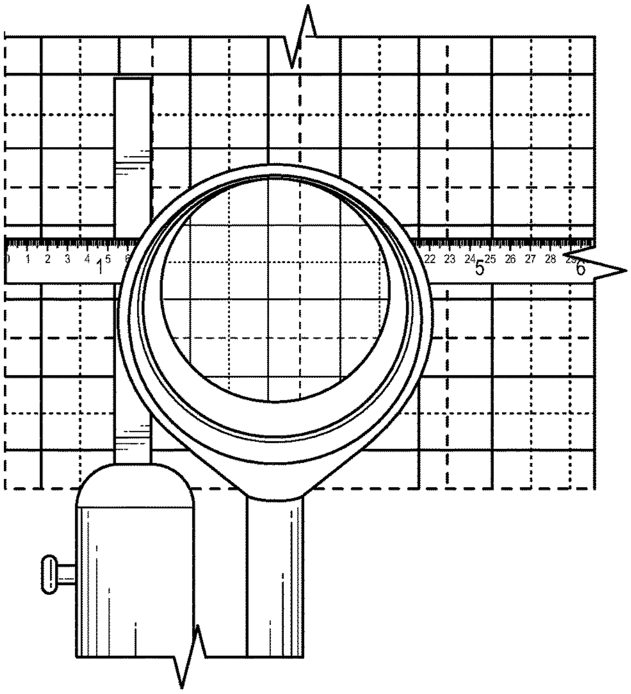
1 2 3 4 5 6
22 23 24 25 26 27 28 29

FIG. 25

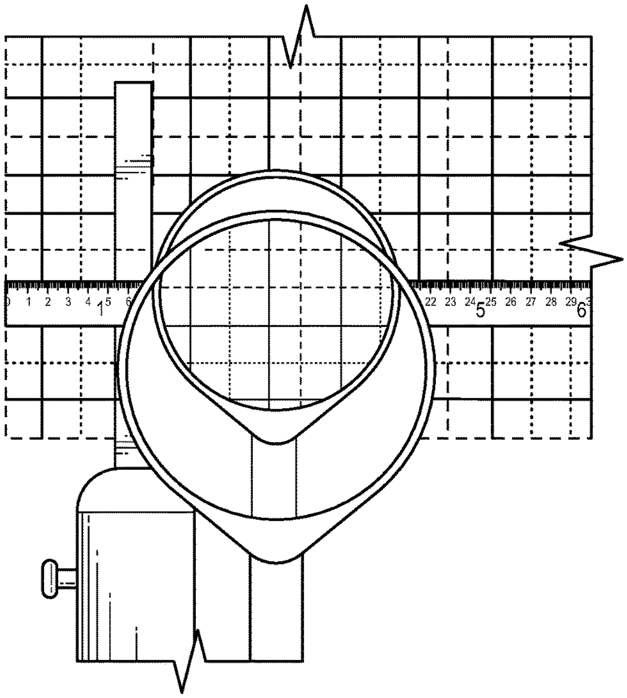
1 2 3 4 5 6
22 23 24 25 26 27 28 29

FIG. 26

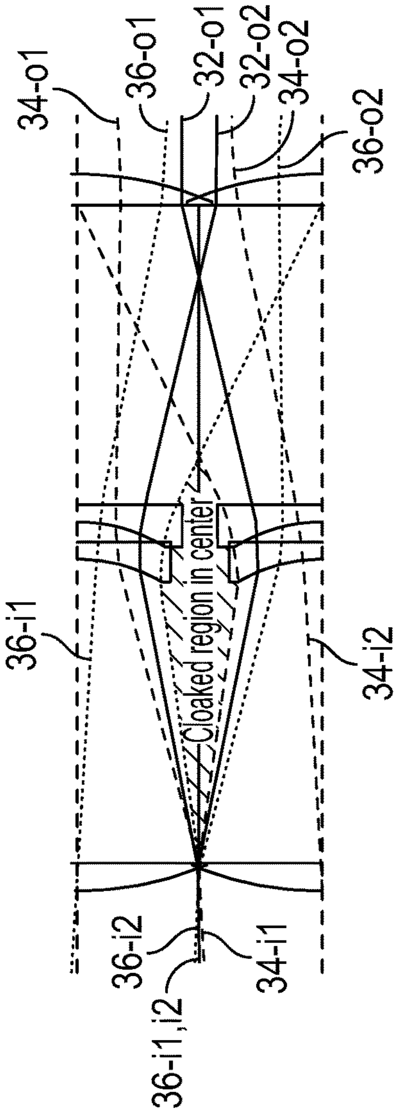
34-o1
36-o1
32-o1
32-o2
34-o2
36-o2
Cloaked region in center
36-i1
34-i2
36-i2
34-i1
36-i1,i2

FIG. 27

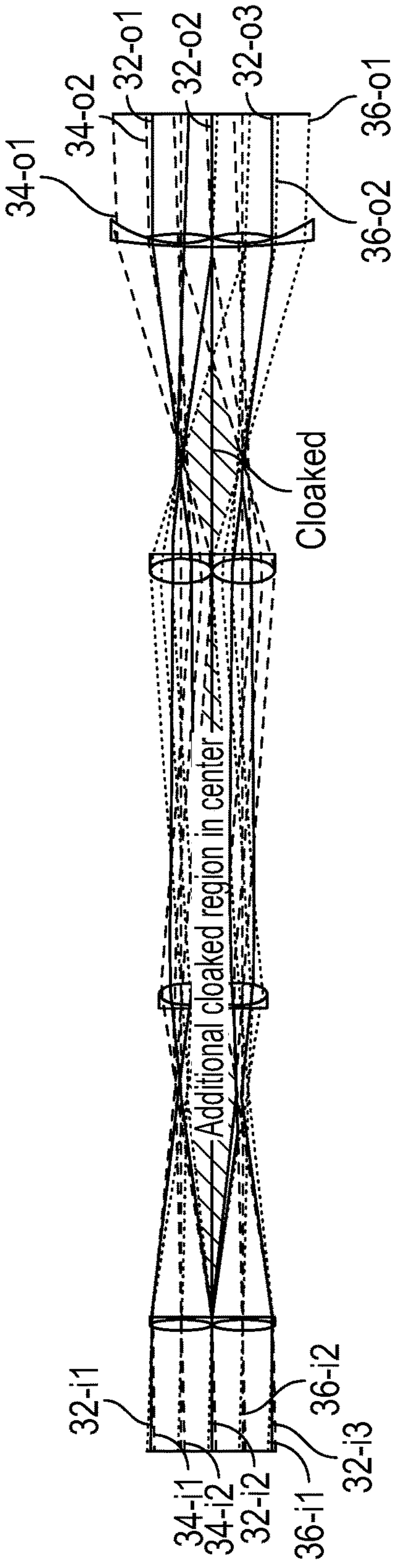
34-o1
34-o2
32-o1
32-o2
32-o3
36-o1
36-o2
Cloaked
Additional cloaked region in center
32-i1
34-i1
34-i2
32-i2
36-i1
32-i3
36-i2

FIG. 28

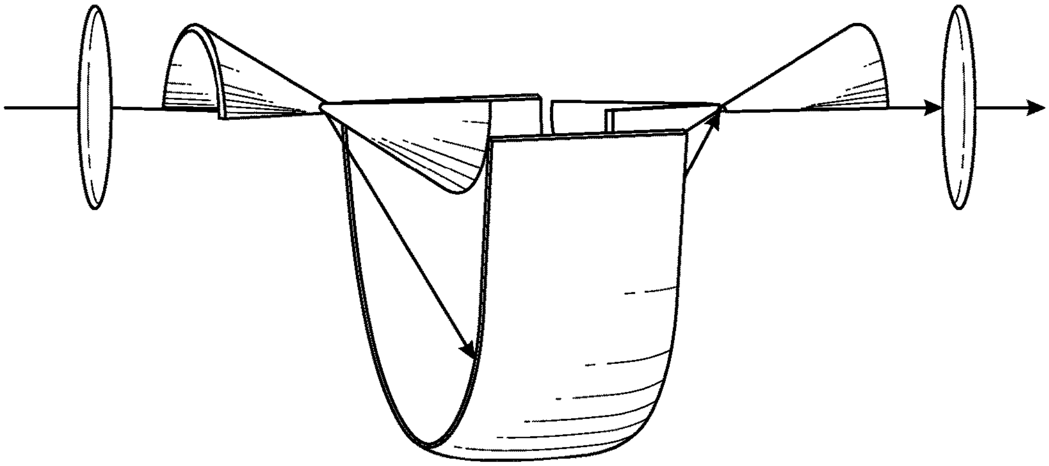

FIG. 29

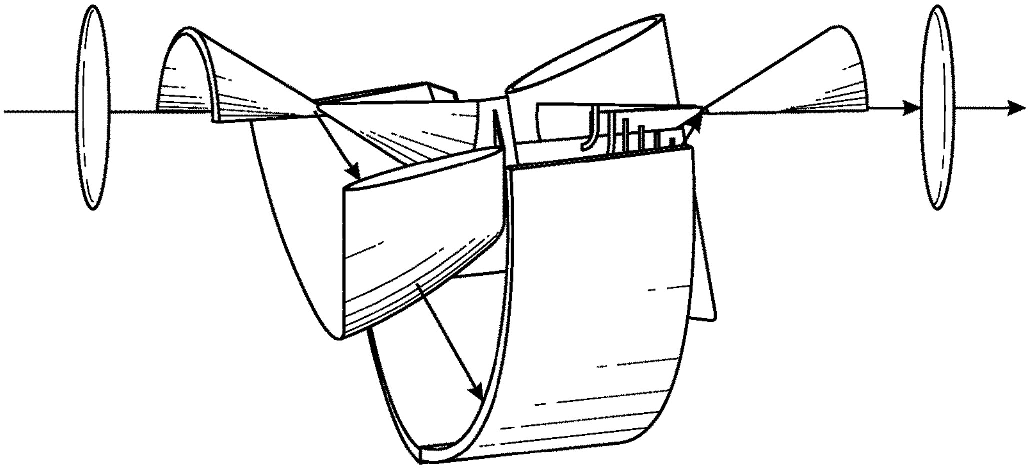

FIG. 30

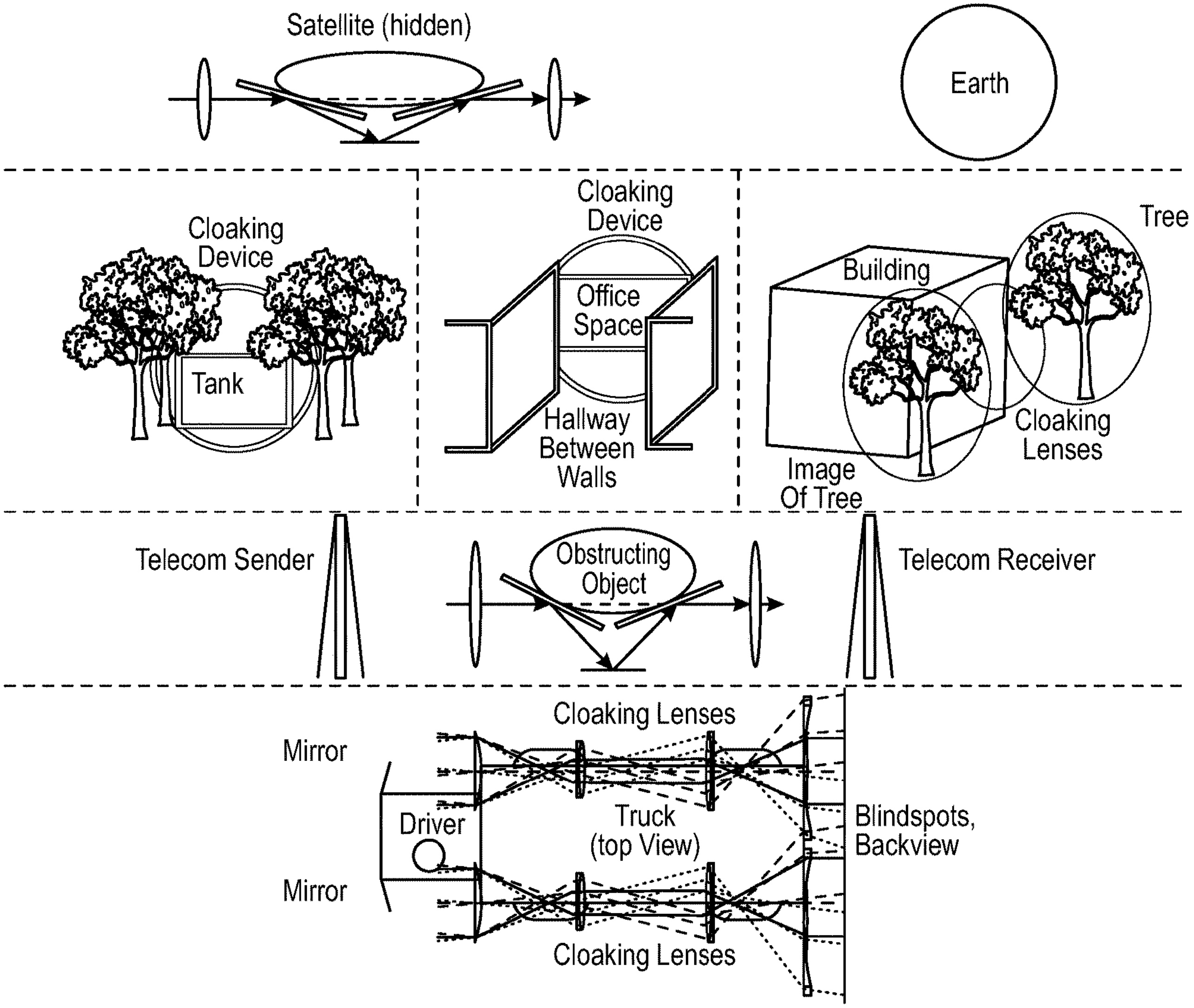
Satellite (hidden)
Earth
Cloaking Device
Tank
Cloaking Device
Office Space
Hallway Between Walls
Building
Tree
Cloaking Lenses
Image Of Tree
Telecom Sender
Obstructing Object
Telecom Receiver
Cloaking Lenses
Mirror
Driver
Truck (top View)
Blindspots, Backview
Mirror
Cloaking Lenses

FIG. 31

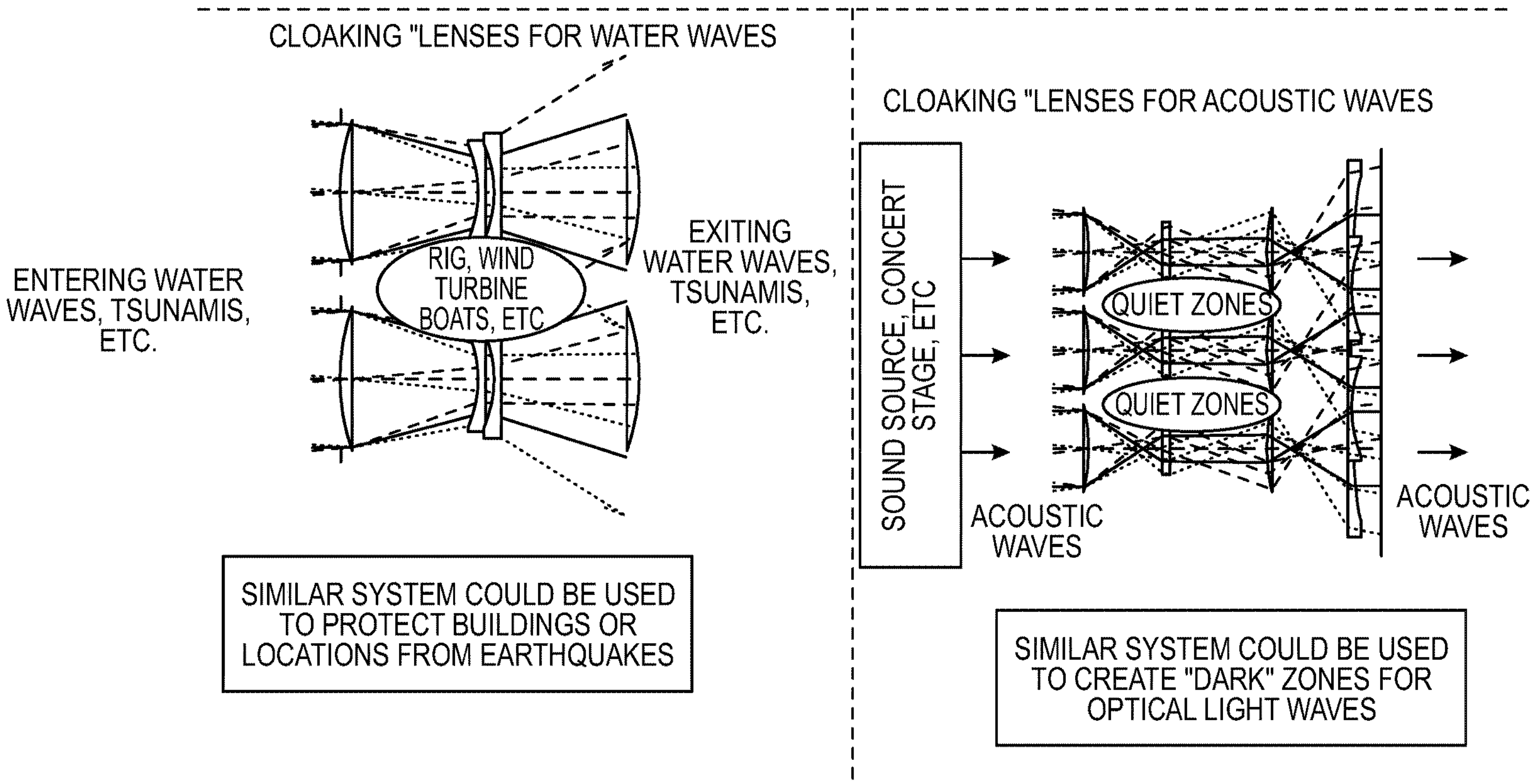
CLOAKING "LENSES FOR WATER WAVES
ENTERING WATER WAVES, TSUNAMIS, ETC.
RIG, WIND TURBINE BOATS, ETC
EXITING WATER WAVES, TSUNAMIS, ETC.
SIMILAR SYSTEM COULD BE USED TO PROTECT BUILDINGS OR LOCATIONS FROM EARTHQUAKES
CLOAKING "LENSES FOR ACOUSTIC WAVES
SOUND SOURCE, CONCERT STAGE, ETC
ACOUSTIC WAVES
QUIET ZONES
QUIET ZONES
ACOUSTIC WAVES
SIMILAR SYSTEM COULD BE USED TO CREATE "DARK" ZONES FOR OPTICAL LIGHT WAVES

FIG. 31
(Continued)

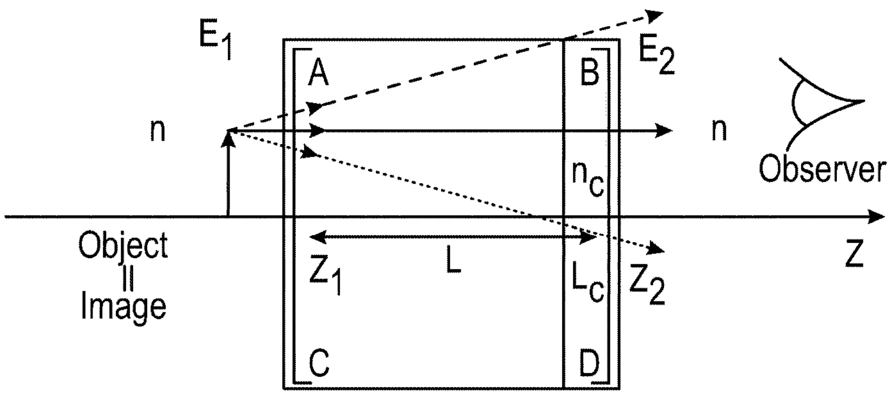
E$_1$
A
B
E$_2$
n
n
Observer
n_c
Object
=
Image
Z_1
L
L_c
Z_2
Z
C
D

FIG. 32

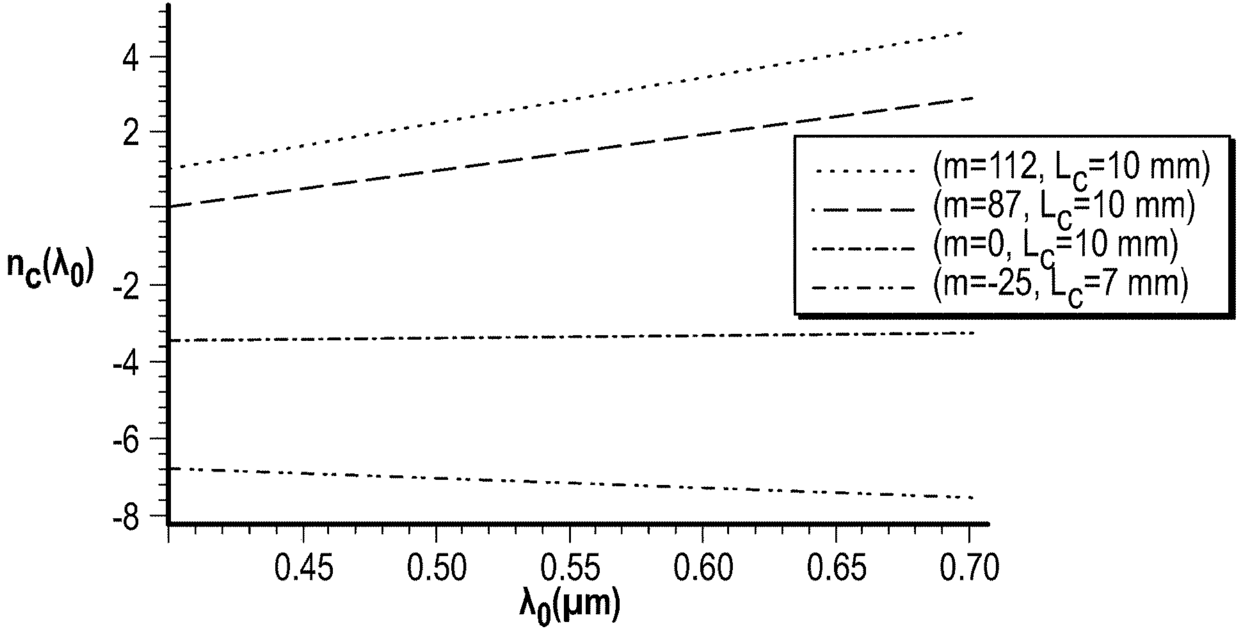
4
2
-2
-4
-6
-8
$n_c(\lambda_0)$
0.45
0.50
0.55
0.60
0.65
0.70
$\lambda_0(\mu m)$
(m=112, L_c=10 mm)
(m=87, L_c=10 mm)
(m=0, L_c=10 mm)
(m=-25, L_c=7 mm)

FIG. 33

PARAXIAL CLOAK DESIGN AND DEVICE

CROSS-REFERENCE TO RELATED APPLICATIONS

This patent application claims priority to U.S. application Ser. No. 62/028,658 filed on Jul. 24, 2014 for a “Paraxial Cloak Design and Device,” the entire content of which is hereby incorporated by this reference.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

This work was supported by the Army Research Office Grant No. W911 NF-12-1-0263 and the DARPA DSO Grant No. W31P4Q-12-1-0015. The government has certain rights in the invention.

BACKGROUND

Optical spatial cloaking has captured the imagination of both the popular culture and scientific communities (see, e.g., Gbur, G. Invisibility physics: Past, present, and future. Progress in Optics 58, 65-114 (2013)). Seminal works in optical spatial cloaking includes work by Leonhardt (Leonhardt, U. Optical conformal mapping. Science 312, 1777-1780 (2006)) and Pendry, Schurig, and Smith (Pendry, J. B., Schurig, D. & Smith, D. R. Controlling electromagnetic fields. Science 312, 1780-1782 (2006)). These seminal works provide a theoretical framework to create a curved space for light waves, by carefully constructing materials in Cartesian space. This new field of research has been called ‘transformation optics’ (McCall, M. Transformation optics and cloaking Contemporary Physics 54, 273-286 (2013)). Experimental realization of such transformational optics has been difficult, due to the previously perceived need for artificial electric and magnetic materials (called ‘metamaterials’), the narrow-band spectrum involved, infinite phase velocity (or negative index to compensate this), and anisotropy in the theory (Gbur, G. Invisibility physics: Past, present, and future. Progress in Optics 58, 65-114 (2013)). Nonetheless, inspired by transformation optics, there have been some advances in optical spatial cloaking. These advances include a two-dimensional microwave cloak (Schurig, D. et al. Metamaterial electromagnetic cloak at microwave frequencies. Science 314, 977-980 (2006)); a ‘carpet cloak’ that hides an object under a surface (Li, J. S. & Pendry, J. B. Hiding under the carpet: A new strategy for cloaking Physical Review Letters 101, 203901 (2008)); and even cloaking in time (Fridman, M., Farsi, A., Okawachi, Y. & Gaeta, A. L. Demonstration of temporal cloaking Nature 481, 62-65 (2012)), and (Lukens, J. M., Leaird, D. E. & Weiner, A. M. A temporal cloak at telecommunication data rate. Nature 498, 205-208 (2013)). A few groups have been able to cloak millimeter- to centimeter-sized objects as well, using birefringent materials (Zhang, B. L., Luo, Y. A., Liu, X. G. & Barbastathis, G. Macroscopic invisibility cloak for visible light. Physical Review Letters 106, 033901 (2011)), and (Chen, X. Z. et al. Macroscopic invisibility cloaking of visible light. Nature Communications 2, 176 (2011)).

To overcome the metamaterial requirements (in some implementations) and to extend cloaking to a broadband, visible regime for large objects, researchers have recently looked to ray optics for cloaking (see, e.g., Chen, H. et al. Ray-optics cloaking devices for large objects in incoherent natural light. Nature Communications 4, 2652 (2013); Zhai, T. R., Ren, X. B., Zhao, R. K., Zhou, J. & Liu, D. H. An effective broadband optical ‘cloak’ without metamaterials. Laser Physics Letters 10, 066002 (2013); and Howell, J. C., Howell, J. B. & Choi, J. S. Amplitude-only, passive, broadband, optical spatial cloaking of very large objects. Applied Optics 53, 1958-1963 (2014)). In these cloaks, the amplitude and direction of light fields are considered, as opposed to the full preservation of fields (amplitude and phase) of transformation optics. These designs have been able to cloak centimeter to meter-sized objects with commonly available optics. Yet, these schemes work only for unidirectional incident light, as they are not designed for continuous multidirectional cloaking, and can have non-unity magnifications. For off-axis, non-zero angles, the background images show distortion and positional shifts. This is particularly true if the background is far away from the cloaking device. In addition, as seen in FIG. 1 of Howell, J. C., Howell, J. B. & Choi, J. S. Amplitude-only, passive, broadband, optical spatial cloaking of very large objects. Applied Optics 53, 1958-1963 (2014), rays that propagate through the system, but go through the center at non-zero angles, can actually enter the cloaking region, effectively uncloaking the space.

Despite the advances in cloaking, a 3-D multidirectional cloak has been elusive. As shown by Wolf and Habashy (Wolf, E. & Habashy, T. Invisible bodies and uniqueness of the inverse scattering problem. Journal of Modern Optics 40, 785-792 (1993)) and Nachman (Nachman, A. I. Reconstructions from boundary measurements Annals of Mathematics 128, 531-576 (1988)), no isotropic cloaking scheme can hide an object from all viewing angles. Their work answered a question that stemmed from Devaney (Devaney, A. J. Nonuniqueness in inverse scattering problem. Journal of Mathematical Physics 19, 1526-1531 (1978)), who elegantly showed how to mathematically construct potentials that have zero scattering fields, and are hence invisible. Devaney’s result, however, was for a finite number of discrete directions, and not for a continuous range of angles.

BRIEF SUMMARY

The following presents a simplified summary of some non-limiting embodiments of the invention in order to provide a basic understanding of the invention. This summary is not an extensive overview of the invention. It is not intended to identify key/critical elements of the invention or to delineate the scope of the invention. Its sole purpose is to present some embodiments of the invention in a simplified form as a prelude to the more detailed description that is presented later.

Paraxial cloaking devices and related methods, as described herein, provide a cloaking volume by which an object can be hidden from view relative to a continuous range of viewing directions. The cloaking devices described herein can be constructed with readily available off-the-shelf isotropic optical components, such as lenses and mirrors, thereby negating the need for any artificial electric and/or magnetic materials (i.e., metamaterials) in at least some implementations. Such devices have potential usages in many different fields, for example, in secure communications, transportation, military applications, and medical devices.

In one non-limiting embodiment, a paraxial cloaking device having a cloaking volume and a reference optical axis is provided. The cloaking device includes an optical input receiving incoming light rays and an optical output from which a plurality of the light rays exit the paraxial cloaking device. The optical output is offset along the optical axis from the optical input. The cloaking volume is disposed between the optical input and the optical output. For light rays having incoming directions that are non-parallel to the reference optical axis up to a first angle between the incoming light ray and the reference optical axis, each of the light rays exits the cloaking device substantially aligned with the respective incoming light ray and does not pass through the cloaking volume. The paraxial cloaking device has a unity magnification factor (i.e., the cloaking device does not affect the apparent size of the objects in the image to the viewer).

In some instances, a paraxial cloaking device hides the cloaking volume for viewing directions that are significantly off-axis. For example, in some instances, the first angle is up to 3.5 degrees, up to 5 degrees, or even up to 15 degrees or more.

In some instances, a paraxial cloaking device imparts an optical transformation to the received light rays defined by an ABCD matrix, wherein A=1, B=L/n, C=0, and D=1; L being a length between the optical input and the optical output, and n being an index of refraction for the surrounding/ambient medium, before and/or after, the cloaking device.

Various configurations of a paraxial cloaking devices are possible. For example, in some instances a paraxial cloaking device includes a first lens, a first mirror, a second mirror, a third mirror, and a second lens. The first and second lenses are centered on the reference optical axis. The first and third mirrors are tilted relative to the reference optical axis. The second mirror is curved. The optical input includes the first lens. The optical output includes the second lens. A light ray passing through the cloaking device follows an optical propagation path. The second mirror is disposed along the optical propagation path between the first and second lenses. The first mirror is disposed along the optical propagation path between the first lens and the second mirror. And the third mirror is disposed along the optical propagation path between the second mirror and the second lens. The first and second lenses can have the same focal length. The first lens and the first mirror can be separated by a distance equal to a distance separating third mirror and the second lens. And the first mirror and the second mirror can be separated a distance equal to a distance separating the second mirror and the third mirror.

As another example, in some instances a paraxial cloaking device includes a first lens, a first mirror, a second lens, a second mirror, a third lens, a third mirror, and a fourth lens. The first and fourth lenses are centered on the reference optical axis. The first and third mirrors are tilted relative to the reference optical axis. The optical input includes the first lens. The optical output includes the fourth lens. A light ray passing through the cloaking device follows an optical propagation path. The second mirror is disposed along the optical propagation path between the first and fourth lenses. The first mirror is disposed along the optical propagation path between the first lens and the second mirror. The third mirror is disposed along the optical propagation path between the second mirror and the fourth lens. The second lens is disposed along the optical path between the first and second mirror. And the third lens is disposed along the optical path between the second and third mirrors. The first and fourth lenses can have the same focal length. The first lens and the first mirror can be separated by a distance equal to a distance separating third mirror and the fourth lens. The first mirror and the second lens can be separated a distance equal to a distance separating the third lens and the third mirror. The second and third lenses can have the same focal length. And the second lens and the second mirror can be separated a distance equal to a distance separating the second mirror and the third lens.

As another example, a paraxial cloaking device can be rotationally symmetric and includes three or more lenses. For example, a rotationally symmetric cloaking device can include three or more lenses that are centered on the reference optical axis. As another example, a rotationally symmetric cloaking device can include four or more lenses that are centered on the reference optical axis. At least one of the lenses can include an achromatic doublet to reduce aberrations.

In another non-limiting embodiment, a related method of designing a cloaking device is disclosed. The method includes selecting and positioning optical components of a candidate cloaking device and calculating an ABCD matrix for the candidate cloaking device. The optical components are selected and positioned such that A=1, B=L/n, C=0, and D=1; L being a length between the optical input and the optical output, n being an index of refraction for the surrounding/ambient medium, before and/or after, the cloaking device. In some instances, the method further includes adjusting at least one of a parameter or a position of at least one of the optical components of the candidate cloaking device so as to compensate for a thickness of at least one of the optical components of the candidate cloaking device. In some instances, the method further includes tracing at least one ray through the candidate cloaking device to assess a deviation between the location and direction of the ray prior to being received by the candidate cloaking device and after exiting the candidate cloaking device, and adjusting at least one of a parameter and a position of at least one of the optical components of the candidate cloaking device based on the deviation assessment so as to reduce the deviation.

In another non-limiting embodiment, a method for designing a cloaking device having a desired cloaking volume is disclosed. Such a method can include selecting any suitable cloaking device as described herein to be scaled to provide the desired cloaking volume. A scale factor indicative of a size difference between the desired cloaking volume and the cloaking volume of the selected cloaking device is determined. Configuration parameters of the second cloaking device are determined by scaling corresponding configuration parameters of the selected cloaking device by the scale factor.

In some instances, the paraxial cloaking devices described herein can be constructed from readily available existing optical components, for example, lenses and mirrors. In some instances, a paraxial cloaking device does not include a metamaterial (e.g., artificial electric and/or magnetic material) optical component. And in some instances, a paraxial cloaking device does not include a non-isotropic optical component. In other instances, a paraxial cloaking device does include a metamaterial.

In some instances, the cloaking device includes a metamaterial optical component.

In some instances, the cloaking device further includes a phase matching element. The phase matching element may be in some instances a flat plate having at least one index of refraction and a longitudinal length. The phase matching element may be in some instances an anomalous dispersion element. The phase matching element may be or include a metamaterial in some instances. The phase matching element may have a negative index of refraction in some instances.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. **1**A and **1**B are simplified diagrammatic illustrations of an 'identity' optical transformation.

FIG. **1**C is a simplified diagrammatic illustration of an optical transformation of a paraxial cloaking device, in accordance with some instances.

FIG. **2** is a simplified diagrammatic illustration of an optical transformation.

FIG. **3** is a simplified diagrammatic illustration of a paraxial cloaking device, in accordance with some instances.

FIG. **4** illustrates a ray tracing assessment performed on a cloaking device in accordance with the cloaking device of FIG. **3**.

FIG. **5** illustrates a ray tracing assessment performed on an optimized cloaking device in accordance with the cloaking device of FIG. **3**.

FIG. **6** is a simplified diagrammatic illustration of a paraxial cloaking device having a rotationally symmetric improvement, in accordance with some instances.

FIG. **7** illustrates a ray tracing assessment performed on a cloaking device in accordance with the cloaking device of FIG. **6**.

FIG. **8** illustrates a ray tracing assessment performed on a scaled version of the cloaking device of FIG. **7**.

FIG. **9** is a simplified diagrammatic illustration of a two lens device.

FIG. **10** is a simplified diagrammatic illustration of a rotationally-symmetric three lens paraxial cloaking device, in accordance with some instances.

FIG. **11** is a simplified diagrammatic illustration of a rotationally-symmetric four lens paraxial cloaking device, in accordance with some instances.

FIG. **12** shows a plot of parameters for possible four lens, rotationally-symmetric, paraxial cloaking devices, in accordance with some instances.

FIG. **13**A and FIG. **13**B illustrate a ray tracing assessment performed on an example three lens paraxial cloaking device.

FIG. **14** illustrates a cloaked region for the example three lens paraxial cloaking device of FIG. **13**.

FIG. **15** shows an image taken from an on-axis viewpoint looking through a prototype three lens paraxial cloaking device, in accordance with some instances.

FIG. **16** shows an image taken from an off-axis (0.55 degrees) viewpoint looking through the prototype three lens paraxial cloaking device of FIG. **15**.

FIG. **17** shows an image taken from an off-axis (0.83 degrees) viewpoint looking through the prototype three lens paraxial cloaking device of FIG. **15**.

FIG. **18** shows an image taken from an off-axis (1.11 degrees) view point looking through the prototype three lens paraxial cloaking device of FIG. **15**.

FIG. **19** shows a side profile image of three lens paraxial cloak of FIG. **15**.

FIG. **20** illustrates a ray tracing assessment performed on an example four lens paraxial cloaking device.

FIG. **21** shows ray tracing simulation results on a four lens paraxial cloak that has been scaled up to twice the size of the four lens paraxial cloak of FIG. **20**.

FIG. **22** shows a side profile image of a prototype four lens paraxial cloak.

FIG. **23** shows an image taken from an off-axis (−0.65 degrees) viewpoint looking through the prototype four lens paraxial cloaking device of FIG. **22**.

FIG. **24** shows an image taken from an on-axis (0.0 degrees) viewpoint looking through the prototype four lens paraxial cloaking device of FIG. **22**.

FIG. **25** shows an image taken from an off-axis (0.47 degrees) viewpoint looking through the prototype four lens paraxial cloaking device of FIG. **22**.

FIG. **26** shows an image taken from an off-axis (0.95 degrees) viewpoint looking through the prototype four lens paraxial cloaking device of FIG. **22**.

FIG. **27** illustrates a ray tracing assessment performed on an example split three lens paraxial cloaking device.

FIG. **28** illustrates a ray tracing assessment performed on an example combined four lens paraxial cloaking device.

FIG. **29** is a simplified diagram of a three-dimensional adaptation of the two-dimensional cloak shown in FIG. **3**.

FIG. **30** is a simplified diagram of a three-dimensional adaptation of the two-dimensional cloak shown in FIG. **6**.

FIG. **31** shows possible applications of cloaking devices.

FIG. **32** is a simplified diagrammatic illustration of a paraxial cloaking device, in accordance with some instances.

FIG. **33** is a chart of various thin, flat plates that may be used to match a full-field cloak condition in some instances.

DETAILED DESCRIPTION

In the following description, some non-limiting embodiments of the present invention will be described. For purposes of explanation, specific configurations and details are set forth in order to provide a thorough understanding of the embodiments. It will, however, also be apparent to one skilled in the art that the present invention may be practiced without the specific details. Furthermore, well-known features may be omitted or simplified in order not to obscure the embodiment being described.

In some instances, the paraxial cloaking devices disclosed herein require no new materials (e.g., artificial electric and/or magnetic materials—"metamaterials"), use isotropic off-the-shelf optics, scale easily to cloak relatively large objects, and are as broadband as the choice of optical material, all of which have been challenges for previous cloaking schemes. In addition, a concise formalism is provided that quantifies and produces optical cloaks that can be used to cloak an object with respect to a small-angle ('paraxial') limit. To a first approximation, the paraxial cloaking devices disclosed herein accurately replicate the ambient space.

Characterizing a Paraxial Optical Spatial Cloak

Referring now to the drawings, in which like reference numerals represent like parts throughout the several views, FIGS. **1**A through **1**C schematically illustrate possible candidate for ray-based optical spatial cloaking FIG. **1**A illustrates an identity transformation, in which rays entering an identity transformation black box **10** from the left and exit to the right with the same positions and directions. FIG. **1**B illustrates how an observer **12** would view an object **14** disposed behind the identity transformation black box **10**. To the observer **12**, the object **14** would appear to be closer than its actual position, by the length of the box, and therefore appear to be located at image **16**. Although the identity transformation is a good candidate for ray-based optical spatial cloaking, an ideal cloaking device would make objects behind the cloaking device appear to be exactly where they are. In other words, an ideal cloaking device should act the same way as if it was filled with the surrounding medium. For example, to cloak in air, the cloaking device should act as it was filled with air. FIG. **1**C schematically illustrates an example 'ideal' cloaking device **20**.

We can then define an ‘ideal cloak’ for ray optics—it contains a non-zero volume to hide an object, and it behaves as if its entire space was replaced by the surrounding medium, for all light rays entering it.

ABCD Matrix

The characteristics of the cloaking devices described herein can be quantified and defined with the aid of linear equations. Referring now to FIG. **2**, to a first approximation, also referred to herein as the “paraxial approximation,” light rays are assumed to deviate minimally from the axis of rotational symmetry for the system (z in FIG. **2**). In the paraxial approximation, as indicated in equation (1) below, ray angles are assumed to be small. Each light ray can be described by its position y and the paraxial angle u, much like an object is described by position and momentum.

$$u \equiv \tan\theta \approx \theta, \tag{1}$$

For the above case, it is assumed that the system is rotationally symmetric about the z-axis. Theta (θ) in equation (1) is the angle of the light ray from the z-axis.

Because of the linearity of optics in the paraxial approximation, the propagation of light rays through an optical system can be described by matrices. These matrices are referred to herein as ‘ray transfer’ matrices, or ‘ABCD’ matrices. Equation (2) below shows how an ABCD matrix is used.

$$\begin{bmatrix} y' \\ n'u' \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} y \\ nu \end{bmatrix}, \tag{2}$$

In equation (2), n and n' are the indices of refraction for the space before and after the optical system, respectively (See FIG. **2**, where rays travel from left to right). The ABCD matrix defines the transformation accomplished on the light rays by the optical system (e.g., optical system **22** schematically represented in FIG. **2**). The ABCD matrix maps the initial position (y) and angle (u) of incoming light rays to corresponding light rays exiting the system (y', u'). The space before the optical system (represented mathematically by the ABCD matrix) is referred to herein as the “object space” with index of refraction n. In a similar fashion, the space after the optical system is referred to herein as the “image space” with index of refraction n'. In FIG. **2**, y>0, u>0, y'<0, and u'<0, with the sign convention used. The ABCD matrix for a space of length t, with index of refraction n_t, is shown in equation (3) below.

$$M_t = \begin{bmatrix} 1 & t/n_t \\ 0 & 1 \end{bmatrix}. \tag{3}$$

In some instances of a paraxial cloaking device, the ABCD matrix is as set forth in equation (3) above, where t=L is the length of the optical system, and n_t=n=n'. This is because a paraxial cloaking device should replicate the surrounding medium throughout its volume. Substituting into equation (3) yields equation (4).

$$\begin{bmatrix} y' \\ nu' \end{bmatrix} = \begin{bmatrix} 1 & L/n \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y \\ nu \end{bmatrix} = \begin{bmatrix} y + Lu \\ nu \end{bmatrix} \tag{4}$$

In equation (4), the angle of the exiting light ray is set equal to the angle of the incoming light ray and the position y of the light ray shifts by the angle multiplied by the length (i.e., y'=y+Lu). Equation (4) sets forth the optical transformation for a paraxial optical cloaking system that applies to all light rays that would be seen by an observer, at least within the accuracy associated with the first order of approximation set forth above. Equation (4) is compatible with the findings of Wolf and Habashy (E. Wolf and T. Habashy, Journal of Modern Optics 40, 785 (1993)) and Nachman (A. I. Nachman, Annals of Mathematics 128, 531 (1988)) since Equation (4) is based on the paraxial approximation described above, and hence not applicable at large angles. Equation (4), however, has been shown to set forth an effective condition, despite its conciseness. For clarity, the ABCD matrix of equation (4) is restated below in equation (5).

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}_{ideal\ cloak} = \begin{bmatrix} 1 & L/n \\ 0 & 1 \end{bmatrix}, \tag{5}$$

In equation (5), L and n are the length of the cloaking system and index of refraction of the surrounding medium, respectively. Because ABCD matrices have a determinant of 1, equation (5) gives only three conditions to be satisfied:

$$B=L/n, C=0, \text{ and } (A=1 \text{ or } D=1). \tag{6}$$

Note that the cloaking system per equation (4) is “afocal” (C=0). Afocal optical systems have no net focusing power, so, for example, an object at infinity will be imaged to infinity. This aspect is helpful when designing a cloaking system satisfying equation (4), since an afocal condition can be easily checked.

Example Paraxial Optical Spatial Cloaks

FIG. **3** illustrates a first example of a paraxial cloak **30**, in accordance with some instances. The paraxial cloak **30** contains a non-zero volume to hide an object, and satisfies equation (5) with corrected signs to include the mirrors (See equation (8)). While only the on-axis ray is shown (thick arrows), the paraxial cloak **30** also works for off-axis rays from other possible ray directions and positions, at least with respect to off-axis directions within the small angle assumption used to formulate equation (5).

The paraxial cloak **30** works for two-dimensions. To make it work for off-axis angles in three-dimensions, the device can be manufactured by rotating the two-dimensional design by a full 180 degrees about the dotted line that connects the centers of lenses **1** and **5** in FIG. **3** (So each layer should replicate this two-dimensional design throughout the 180 degrees; mirrors **2** and **4** will be shaped similar to bowties (two half-cones attached), mirror **3** will be shaped as a toroidal shell, and the lenses will remain spherical). This is shown in FIG. **29**.

For simplicity, the paraxial cloak **30** is in air, with the index of refraction of air (n_{air}) set to 1. The ABCD Matrix for the paraxial cloak **30** is given in equation (7).

$$\begin{bmatrix} 1 & 0 \\ 1/f_5 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -t_4 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & t_3 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ \frac{-2}{r_3} & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -t_2 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & t_1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ -1/f_1 & 1 \end{bmatrix}, \tag{7}$$

In equation (7), the t's are the longitudinal distances (along the bent z-axis) between each optical element, the f's are the focal lengths, and r_3 is the radius of curvature of the center mirror (See FIG. **3**). The four thick arrows in FIG. **3** follow the path of an on-axis ray. Note that equation (7) is based on a thin lens approximation. Also, the z-axis follows the thick arrows, since this is the center axis of symmetry.

Care must be given to use the correct signs for each of the terms in equation (7), so as to remain consistent. Otherwise, the correct solution space will not be found. One sign convention that can be used is set forth in D. Buralli, OPT 441—Geometrical Optics (The Institute of Optics, University of Rochester, Rochester, N.Y., U.S.A., 2010), where after odd number of reflections the propagation distance t and the index of refraction n change signs. With this convention, the space in which light exits (to the far-right) has index of refraction $n_5=-n_{air}=-1$ and the y-axis also changes signs (by carefully considering propagation directions). Incorporating this sign convention into equation (4) for the paraxial cloak **30** yields equation (8).

$$\begin{bmatrix} y' \\ -u' \end{bmatrix} = \begin{bmatrix} -1 & -L \\ 0 & -1 \end{bmatrix} \begin{bmatrix} y \\ u \end{bmatrix} = \begin{bmatrix} -(y+Lu) \\ -u \end{bmatrix}. \tag{8}$$

In some instances, a paraxial cloak is configured to undo any changes that the first half of the paraxial cloak makes on the light rays. By doing so, the rays may retain the same direction, similar to a retro-reflector. This can be done by making the second half of the paraxial cloak **30** symmetric to the first half (i.e., $f_5=-f_1$, $t_4=-t_1$, $t_3=-t_2$). As a result, f_1 can be solved for as shown in equation (9) such that A=−1 for the ABCD matrix.

$$f_1=t_1-t_2. \tag{9}$$

Serendipitously, this is the same condition that sets C=0. As noted above, D=−1 is automatically satisfied. Equation (9) shows that requiring a paraxial cloak to have symmetric halves is helpful in some instances.

With f_1 set to equation (9), the ABCD matrix for a paraxial cloak simplifies to:

$$\begin{bmatrix} -1 & 2(t_1-t_2)(r_3-t_1+t_2)/r_3 \\ 0 & -1 \end{bmatrix}. \tag{10}$$

The radius of curvature of the center mirror, r3, can then be determined so as to satisfy the remaining condition: B=−L. Using the geometry of the paraxial cloak **30** shown in FIG. **3** yields an overall device length per equation (11).

$$L=2(t_1-t_2\sin(2\alpha-3\pi/2)). \tag{11}$$

Solving for r3 yields:

$$r_3=-\frac{(t_1-t_2)^2}{-2t_1+t_2+t_2\cos(2\alpha)}. \tag{12}$$

Notably, if $r_3=f_1$, then B=0. It can be seen that such an optical system can either be an identity transformation or a cloaking system.

Only equations (9) and (12) need to be met for the optical system **30** to be an ideal paraxial cloak. For example, with a lens with $f_1=200$ mm, mirror **2** tilt angle $\alpha=150$ degrees, and $t_2=-40$ mm, t_1 and r_3 can be solved for yielding $t_1=160$ mm and $r_3=105.3$ mm. The net length of the system **30** is L=360 mm.

Using these values, light rays within and beyond the small angle paraxial approximation were traced using CODE V, which uses Snell's Law directly and accurately simulates light rays in actual lenses. FIG. **4** shows the simulation results. For the simulation, t_1 was obtained by subtracting $|t_2|$ from the back focal length of the lens used, to properly account for the lens thickness (so $t_1\approx150$ mm). The simulation showed that incoming on-axis rays **32**-*i***1**, **32**-*i***2**, **32**-*i***3** are un-deviated relative to their respective outgoing rays **32**-*o***1**, **32**-*o***2**, **32**-*o***3**, and that incoming off-axis rays **34**-*i***1**, **34**-*i***2**, **34**-*i***3**, **36**-*i***1**, **36**-*i***2**, **36**-*i***3** are only slightly shifted relative to their respective outgoing rays **34**-*o***1**, **34**-*o***2**, **34**-*o***3**, **36**-*o***1**, **36**-*o***2**, **36**-*o***3** while maintaining their initial angles quite closely.

CODE V was used to optimize the paraxial cloak to bring the center of the angled rays closer to where they should be located. Specifically, only the curvature of the center mirror (r_3) was optimized to obtain the improvement shown in FIG. **5**. Although the size of the ray bundles entering the system ("entrance pupil") was the same, the required minimum diameter of the last lens was reduced by over 20 mm. The range of incident angles that the resulting optical cloak can accept ("field-of-view") was −3.5 degrees to +3.5 degrees.

The results in each of FIG. **4** and FIG. **5** show some shifts in the positions of the off-axis rays. It is easy to conclude that this result may be due to the use of the small angle paraxial approximation. A close examination shows that the transverse shifts are different for different incident angles. This suggests that the paraxial cloak is not rotationally symmetric. This may be because the center mirror has a non-zero focusing power and a tilt, simultaneously.

Rotationally Symmetric Improvement

Typically, the ABCD ray tracing matrices assume rotational symmetry since only y is used, instead of both x and y. The paraxial cloak **30** of FIG. **3**, however, is not rotationally symmetric. The rotational symmetry of the paraxial cloak **30** can be improved. For example, the center mirror **3** can be replaced with focusing lenses **2'**, **3'** and a flat mirror **3"** to result in a paraxial cloak **40** illustrated in FIG. **6**. The ABCD matrix for the resulting paraxial cloak **40** is given by equation (13).

$$\begin{bmatrix} 1 & 0 \\ 1/f_5 & 1 \end{bmatrix}\cdot\begin{bmatrix} 1 & -t_4 \\ 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} 1 & t_3' \\ 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} 1 & 0 \\ -1/f_3' & 1 \end{bmatrix}\cdot\begin{bmatrix} 1 & t_3 \\ 0 & 1 \end{bmatrix}\cdot \begin{bmatrix} 1 & -t_2' \\ 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} 1 & 0 \\ 1/f_2' & 1 \end{bmatrix}\cdot\begin{bmatrix} 1 & -t_2 \\ 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} 1 & t_1 \\ 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} 1 & 0 \\ -1/f_1 & 1 \end{bmatrix} \tag{13}$$

In addition to replacing the center mirror, a selected focal length f_1 is for lens **1** is used and a corresponding solution for t1 is found. This was done because it is usually easier to purchase a ready-made lens and then adjust the spacing, rather than the other way around. Setting A=−1 and C=0 yields equation (14) for t_1.

$$t_1=[f_1(f_2'-t_2')+f_2'(t_2+t_2')-t_2t_2']/(f_2'-t_2'). \tag{14}$$

With t1 set to equation (14), the ABCD matrix for the paraxial cloak **40** is given by equation (15).

$$\begin{bmatrix} -1 & (2f_1[f_2'^2+f_1(f_2'-t_2')])/f_2'^2 \\ 0 & -1 \end{bmatrix}. \tag{15}$$

Similar as for the paraxial cloak **30** of FIG. **3**, the focal length of lens **2'** can be solved. B is set equal to −L' where the total length L' is given by equation (16).

$$L'=2(t_1-(t_2+t_2')\sin(2\alpha-3\pi/2)). \tag{16}$$

Solving for the focal length of the second lens **2'** yields equation (17).

$$f_2' = -\frac{f_1^2 \pm \sqrt{f_1^2(f_1^2+4f_1t_2'+2t_2'L')}}{2f_1+L'}. \tag{17}$$

Only equations (14) and (17) need to be satisfied for the paraxial cloak **40**. Equations (14) and (17), however, form transcendental equations because L' is a function of t_1 (equation (16)). At this point, t1 and the focal length of the second lens **2'** can be solved for numerically.

It is important to note that the equations above assume that all the lenses are infinitely thin. In reality, lenses have thicknesses, and this can make a large difference in obtaining the correct solutions for equation (5). Hence, further steps can be taken to correct equations (14) and (17) for the use of 'thick' lenses.

As an example, a simulation was run for the paraxial cloak **40**. The simulation was run using CODE V to accomplish ray tracing. The result is shown in FIG. **7**. The field-of-view was the same as for the paraxial cloak **30** of FIG. **3** (−3.5 to +3:5 degrees), but the entrance pupil was increased to 40 mm. Increasing the entrance pupil allows more light to enter, which improves the resolution and 'speed' of the optical system. Even without any separate CODE V optimization, a large improvement is seen over the paraxial cloak **30** simulation results shown in FIG. **4**. The rays are closer to where they need to be, and the size of the rays exiting the system has not increased much even compared to optimized paraxial cloak simulation results shown in FIG. **5**.

The paraxial cloak **40** (or any other paraxial cloak described herein) can be scaled to any suitable size by scaling all radii of curvature, lengths, and entrance pupil by the same factor. For example, the simulation results shown in FIG. **8** are for a doubling of the size of the paraxial cloak **40**, which results in a doubling of the corresponding cloaking space.

The paraxial cloak **40** works for multiple off-axis rays in two-dimensions. For three-dimensions, the device can be manufactured with each layer replicating this two-dimensional design throughout a full 180 degrees, rotating about the dotted line connecting the centers of lenses **1** and **5**, as seen in FIG. **6**. Mirrors **2** and **4** will be shaped similar to bowties (two half-cones attached), mirror **3** will be shaped as a half-pipe, lenses **1** and **5** will remain spherical, and lenses **2'** and **3'** will be clipped, half-cone shapes with varying thicknesses across. This is shown in FIG. **30**.

Rotationally Symmetric Paraxial Cloaks with Lenses

Rotationally symmetric paraxial cloaks are described herein that utilize three thin lenses and four thin lenses, respectively. We will assume the cloaks are in air, for these cases. The ABCD matrix for one thin lens is given by equation (18), where f is the focal length of the lens.

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}_{thin\ lens} = \begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix}, \tag{18}$$

It can be seen that equation (18) will satisfy equation (5) only if f=+/−infinity, i.e., the lens has no optical power (e.g., a flat lens). Such a lens, however, provides no cloaking region and no optical effect.

Instead, attention is now turned to multiple lens systems. In the following derivations, f's are used to denote the focal lengths, and t's are used to denote the distances between the optical elements. Starting with a two lens system as shown in FIG. **9**, the ABCD matrix is given by equation (19).

$$\begin{bmatrix} 1 & 0 \\ -1/f_2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ -1/f_1 & 1 \end{bmatrix} = \begin{bmatrix} 1-t/f_1 & t \\ -(f_1+f_2-t)/(f_1f_2) & 1-t/f2 \end{bmatrix}. \tag{19}$$

Similar to the one lens system, it can be seen that equation (19) will satisfy equation (5) only if $f_1=f_2=$+/−infinity. This is a system that is essentially made of air only, quite literally, again with no cloaking region nor optical effect.

Turning now to a three lens system as shown in FIG. **10**, the ABCD matrix is given by equation (20).

$$\begin{bmatrix} 1 & 0 \\ -1/f_3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & t_2 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ -1/f_2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & t_1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ -1/f_1 & 1 \end{bmatrix}. \tag{20}$$

Setting C=0 and solving for f_2 yields equation (21).

$$f_2 = -\frac{(f_1-t_1)(f_3-t_2)}{f_1+f_3-t_1-t_2}. \tag{21}$$

Using equation (21), the ABCD matrix is given by equation (22).

$$\begin{bmatrix} \frac{f_3(f_1-t_1)}{f_1(f_3-t_2)} & t_1+t_2+t_1t_2\frac{(f_1+f_3-t_1-t_2)}{(f_1-t_1)(f_3-t_2)} \\ 0 & \frac{f_1(f_3-t_2)}{f_2(f_1-t_1)} \end{bmatrix}. \tag{22}$$

Setting B=t1+t2 yields equation (23).

$$t_1t_2\frac{(f_1+f_3-t_1-t_2)}{(f_1-t_1)(f_3-t_2)} = 0. \tag{23}$$

Equation (23), however, is only true if $t_1=0$, or $t_2=0$, or if $(f_1+f_3-t_1-t_2)=0$. The first two cases give the two lens system, which cannot be a paraxial cloak as shown above. The last case makes $f_2\rightarrow$infinity, which also turns this system into a two lens system.

Although a three lens system cannot be an ideal cloak, it can asymptotically approach one. For simplicity, considering the case with symmetric halves ($f_1=f_3$ and $t_1=t_2$) produces equation (24) from equation (21) and equation (25) from equation (23).

$$f_2=(t_1-f_1)/2, \tag{24}$$

$$2t_1^2/(f_1-t_1)=0. \tag{25}$$

Thus, for $f_1 \gg t_1$, both equation (24) and equation (25) can be satisfied in the limit. Accordingly, a practical three lens paraxial cloaking device can be constructed, for example, such as discussed herein.

Turning now to a four lens system as shown in FIG. **11**. By making the second half of the system symmetric to the first half ($f_1=f_4$, $f_2=f_3$, and $t_1=t_3$), any changes in the path of rays that occur in the first half might be undone in the second half which is a possible strategy to make the device appear absent. Both A=1 and C=0 for such a four lens ABCD matrix can be satisfied by equation (26).

$$t_1=f_1+f_2. \tag{26}$$

With equation (26), the ABCD matrix for the four lens system is given by equation (27).

$$\begin{bmatrix} 1 & f_1(-2t_1^2+f_1(2t_1+t_2))/(f_1-t_1)^2 \\ 0 & 1 \end{bmatrix}. \tag{27}$$

Setting B=(2 t_1+t_2) and solving for t_2 yields equation (28).

$$t_2=2f_2(f_1+f_2)/(f_1-f_2). \tag{28}$$

Thus, an exact solution for equation (5) can be found for a rotationally symmetric lens-only system, using at least four lenses. Using equation (26) and equation (28), the total length of the four lens system is given by equation (29).

$$L=2t_1=t_2=2f_1(f_1+f_2)/(f_1-f_2)/(f_1-f_2). \tag{29}$$

FIG. **12** shows a plot of t_1/f_2 (solid line), t_2/f_2 (dashed line), and L/f_2 (dotted line) as a function of the unit-less parameter $a=f_1/f_2$ on the horizontal axis from equations (26), (28), and (29), for all possible symmetric, four lens, ideal paraxial cloaks, for $a=f_1/f_2$ between −5 and 5 (since these values satisfy equation (5) mathematically). However, even if a device matches the t_1, t_2, and/or L for a given $a=f_1/f_2$ value, this does not guarantee it to be an ideal cloak since the physical feasibility and presence of a non-empty cloaking region must be checked separately. Additionally, it can be seen that when $f_1 \rightarrow -f_2$, the system approaches a one lens system ($t_1=t_2=L=0$). The two extrema for L occur when equation (30) is satisfied.

$$f_1=(1\pm\sqrt{2})f_2. \tag{30}$$

Although these and other solutions may satisfy equation (5) mathematically, checks should be made to ensure that any particular solution contains a non-empty cloaking region and is physically feasible, for it to be a valid cloaking device.

Example Three Lens Cloak

The cloaking region for the lens designs described herein depend on what incident angles, or "field-of-view," are allowed. To assess the size of the cloaking space, ray-trace simulations using CODE V were performed. First, referring to FIG. **13**A, FIG. **13**B, and FIG. **14**, a three lens cloak with symmetric left and right halves was simulated. As described above, a three lens system approaches an ideal paraxial cloak as its length goes to zero. The size of the ray bundle entering the system ("entrance pupil") was set to the first lens diameter in our experimental setup (75 mm). The field-of-view is −3.5 degrees to +3.5 degrees. The cloaking region is between the lenses and is the ring-shaped region where no rays pass. FIG. **13**A and FIG. **13**B compare the final image rays to the original rays near the first lens. For an ideal cloak, these rays would overlap. It can be seen that the angles are similar, and the transverse shifts are not too large. FIG. **13**B shows: (1) incoming on-axis rays **32**-*i***1**, **32**-*i***2** and respective outgoing on-axis rays **32**-*o***1**, **32**-*o***2**; (2) incoming off-axis rays **34**-*i***1**, **34**-*i***2**, **34**-*i***3** and corresponding respective outgoing off-axis rays **34**-*o***1**, **34**-*o***2**, **34**-*o***3**; and (3) incoming off-axis rays **36**-*i***1**, **36**-*i***2**, **36**-*i***3** and corresponding respective outgoing off-axis rays **36**-*o***1**, **36**-*o***2**, **36**-*o***3**.

Plano-convex and plano-concave lenses were used for the three lens cloak. FIGS. **15** through **19** shows images taken during an experimental demonstration in which an object (a wall) was located approximately 2 m from the closest lens in the back. The camera used to take the images was located 5.3 meters away from the front lens and optically zoomed in by 21× (the maximum magnification of the camera). The images were taken from on-axis (FIG. **15**—0 degrees), 0.55 degrees off-axis (FIG. **16**), 0.83 degrees off-axis (FIG. **17**), and 1.11 degrees off-axis (FIG. **18**), by increasing the height of the camera. A ruler was placed near the center diverging lenses. It can be seen that the middle of the ruler is cloaked. Also, the lines of the wall as seen through the cloaking device match the background wall as seen outside of the cloaking device, as expected for a good cloak. FIG. **19** shows a side image of the three lens system experimental prototype.

Example Three Lens Cloak Parameters

For the first and last lenses (**1** and **3**) 200 mm focal length, BK7, 75 mm diameter lenses were used. For the center lens, two −100 mm focal length, BK7, 50 mm diameter lenses were used, back-to-back, to create a lens with focal length of approximately −50 mm. All lenses were catalogue optics from Edmund Optics. From equation (24), $t_1=t_2$ equals approximately 100 mm was obtained. Including the lens thicknesses and the material indices of refraction, t_1 was optimized slightly so that the afocal condition C=0 was closely achieved. Diameter of last lens needs to be >150 mm for all rays to pass (no "vignetting"). For the CODE V simulation, the apertures were not restricted to the actual lens sizes. The aperture stop was the first surface. Aperture diameter sizes (for no vignetting) of the first and second diverging lenses in the center are 54 mm and 61 mm, respectively. Total length of the system is 219 mm.

Example Four Lens Cloak

A four lens 'ideal' paraxial cloak that has symmetric left and right halves was simulated. Because real lens systems produce aberrations that can blur and distort the observed image, 'achromatic doublets,' which combine two lenses as one, were used to correct for chromatic and other aberrations. Equations (26) and (28) were corrected and t_1, t_2, and t_3 ($t_1=t_3$) were calculated. FIG. **20** shows the results of a simulation that used these calculated paraxial values without any additional optimization. The resulting cloaking region is an elongated cylinder between the lenses where the rays are not present.

As stated above, the cloaking systems disclosed herein are scalable. For example, a four lens cloak can be scaled to any suitable size by scaling all radii of curvature, lengths, and entrance pupil by the same factor. FIG. **21** shows ray tracing simulation results on a four lens paraxial cloak that has been scaled up to twice the size of the four lens paraxial cloak of FIG. **20**, thereby obtaining double the cloaking space.

A prototype four lens cloak was constructed using achromatic doublets to reduce aberrations. Photographs of this prototype paraxial cloak are shown in FIGS. **22** through **26**.

FIG. **22** shows a side profile view of the prototype four lens cloaking device. The grids on the wall were 1.9 m from the closest lens to the back. The camera was 3.1 m away from the front lens, but optically zoomed in by 21×. The images were taken from −0.65 degrees off-axis (FIG. **23**), on-axis (FIG. **24**), at 0.47 degrees off-axis (FIG. **25**), and 0.95 degrees off-axis (FIG. **26**) viewing angles, by changing the height of the camera. A ruler was placed behind the second doublet from the front. The middle of the ruler is cloaked near the center of the device. In particular, the grids on the wall are clear for all colors, have minimal distortion, and match the shifts of the background grids for all the angles.

Example Four Lens Cloak Parameters

For the first and last lenses (**1** and **4**), 200 mm focal length, 50 mm diameter achromatic doublets composed of BK7 and SF2 glasses were used. For the center two lenses (**2** and **3**), 75 mm focal length, 50 mm diameter achromatic doublets composed of SF11 and BAF11 glasses were used. All doublets were off-the-shelf catalogue optics from Thorlabs and had antireflection coating in the visible spectrum. For the CODE V simulations, the aperture sizes were not restricted, so as to ensure no vignetting. The aperture stop was the first surface. Diameters of the second, third, and last doublets need to be >33 mm, 51 mm, and 112 mm, respectively, for no vignetting. Total length of the system is 910 mm.

Extending Paraxial Cloaking to Include the Full Field

We now discuss non-limiting examples of paraxial cloaks that work for the full-field (matched amplitude and phase). FIG. **32** schematically shows one example of a full field paraxial cloak. In FIG. **32**, the image from the cloaking device is the same as the object. Propogation of light rays can be described by an 'ABCD' matrix. n is the index of refraction of the ambient medium. L is the longitudinal length of the device along the center z-axis. E_1 and E_2 are the input and output fields, respectively, at $z=z_1$ and z_2. In this example, phase matching is achieved with a flat plate with index n_c and length L_c.

In some instances, the examples of paraxial full-field cloaking described in this section satisfy all but the omnidirectionality of an 'ideal' cloaking device. In some instances, this 'paraxial' formalism can work for up to +/−30°, which is quite practical for many cases where the cloaked object is not placed immediately before an observer. Although we discuss below the propagation of a monochromatic field, since an arbitrary field of light can be written as a linear superposition of monochromatic waves, our systems extend to broadband without loss of generality.

Duan et al. recently provided phase matching with their unidirectional cloaking system based off of geometric optics (see R. Duan, E. Semouchkina, and R. Pandey, "Geometric optics-based multiband cloaking of large objects with the wave phase and amplitude preservation," Optics Express 22, 27193-27202 (2014)). They provided a heuristic reason why rays that pass through the edge of their optical system may match the phase of those rays that pass through the center instead, for their particular setup with split lenses. They then simulated phase-matched cloaking for discretely separated micro-waves to mm-waves. Here, we provide for and analytically show phase matching for general paraxial optical systems, including continuously multidirectional cloaking systems. Embodiments of our phase matching systems utilize formula given by Siegman and others (see A. E. Siegman, Lasers (University Science Books, Sausalito, C A, 1986)), for propagation of any paraxial field through a generalized paraxial optical system. The resulting formula is proved using Fermat's principle and Huygens' integral, by accounting for the optical path lengths of all rays.

Field propagation based off of Huygens' principle of wavelet propagation, is effective and widely used in diffraction theory and Fourier optics. We first assume that the ambient medium is spatially uniform with index of refraction n, and that the optical system of interest is rotationally symmetric for simplicity (so can be described by a 2×2 ABCD matrix) (see FIG. **32**). Non-uniform medium or non-rotationally symmetric systems are also envisioned, and can be derived from here.

For rotationally symmetric two-dimensions, and no limiting apertures, Huygens' integral in the Fresnel (or, paraxial) approximation is given by equation (31) (see S. A. Collins, "Lens-system diffraction integral written in terms of matrix optics," J. Opt. Soc. Am. 60, 1168-1177 (1970)):

$$\tilde{E}_2(x_2, y_2) = \frac{ie^{-ik_0L_0}}{B\lambda_0}\int\int_{-\infty}^{\infty}\tilde{E}_1(x_1, y_1)\exp\left\{-i\frac{\pi}{B\lambda_0}\left[A(x_1^2+y_1^2)-2(x_1x_2+y_1y_2)+D(x_2^2+y_2^2)\right]\right\}dx_1dy_1. \tag{31}$$

In equation (31), A, B, C, Dare the coefficients of the ABCD matrix. $L_0=\Sigma_i n_i L_i$ is the on-axis optical path length, where each ith optical element has index of refraction n_i and physical thickness L_i along the longitudinal, center axis (z). λ_0 and λ_0 are the wavelength and wavevector in free space. $\tilde{E}_1$, $\tilde{E}_2$ are the complex, spatial amplitude of the input ($z=z_1$) and output ($z=z_2$) field distributions, respectively (E_1, E_2 in FIG. **32**, but without the $e^{+i\omega t}$ harmonic time dependence).

In some of the embodiments described in previous sections, we stated that an 'ideal' cloaking device simply replicates the ambient medium throughout its volume, and so its ABCD matrix is given by equation (32):

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}_{ideal\ cloak} = \begin{bmatrix} 1 & L/n \\ 0 & 1 \end{bmatrix}. \tag{32}$$

Thus, the propogated field for an 'ideal' full-field cloak is given by equation (33):

$$\tilde{E}_2^{cloak}(x_2, y_2) = \frac{ine^{-ik_0nL}}{L\lambda_0}\int\int_{-\infty}^{\infty}\tilde{E}_1(x_1, y_1)\exp\left\{-i\frac{n\pi}{L\lambda_0}\left[(x_1^2+y_1^2)-2(x_1x_2+y_1y_2)+(x_2^2+y_2^2)\right]\right\}dx_1dy_1. \tag{33}$$

A ray optics cloak embodiment satisfies equation (32). By comparing equations (31) and (33), we then see that a ray optics cloak can be a full-field cloak if $e^{-ik_0L_0}=e^{ik_0nL}$, which may be the case when $k_0L_0=k_0nL$ (absolute phase matching), or when $L_0\equiv nL \pmod{\lambda_0}$ (phase-matching to integer multiples of 2π). The significance of these conditions is that they allow for phase-matched, full field cloaking for any incoming fields within the paraxial approximation. In some embodiments, the phase-matching condition only needs to be satisfied once for a given optical system, and then all ray directions and positions, or any field distribution ($\tilde{E}_1$) will exit if traversed through a cloaked ambient space. Note that a ray optics cloak (equation (32)) will usually be phase-matched to integer multiples of 2π with a full-field cloak for multiple, but discretely separated, wavelengths automatically. However, we can do better and match it for a continuous, broad bandwidth by satisfying the phase-matching to integer multiples of 2π condition with appropriate dispersion control.

One method for broadband phase matching is to add a thin, flat plate to a ray optics cloak, anywhere between the background object(s) and the observer. Since it is "thin" and flat, the original ABCD matrix (equation (32)) will be unchanged and only the $e^{-ik_0 L_0}$ factor outside the integral of equation (31) will be affected. Let $L'=\Sigma_{i=1}^{N} L_i$ be the total length of the original ray optics cloak (so $L=L_c+L'$), where N is the number of original optical elements. So $L_0 \rightarrow n_c L_c + \Sigma_{i=1}^{N} n_i L_i$ in Eq. (31), and $nL \rightarrow n(L_c+L')$ in Eq. (33). Here, n_i, L_i are the index of refraction and physical longitudinal length, respectively, for the i^{th} optical element, and n_c, L_c are likewise for the correcting thin plate. It may be shown that a plate is "thin" when $|L_c/L'| \ll 1$ and $|L_c/L'| \ll |n_c/n|$. With our phase-correcting plate, the phase-matching to integer multiples of 2π condition is rearranged to write the dispersion relation of n_c:

$$n_c(\lambda_0, m, L_c) = n(\lambda_0) + \frac{1}{L_c}\left\{m\lambda_0 + \sum_{i=1}^{N}[n(\lambda_0) - n_i(\lambda_0)]L_i\right\}. \tag{34}$$

$m \in \mathbb{Z}$ is any integer, with infinite but discrete choices.

There are infinitely many combinations of L_c and in that are possible for $n_c((\lambda_0)$, m, $L_c)$. Ultimately, the choice of which to use should depend on physically realizable n_c's. A few such solutions, that phase-corrects a four lens cloak described in J. S. Choi and J. C. Howell, "Paraxial ray optics cloaking," Optics Express 22, 29465-29478 (2014), are shown in FIG. **33**, in which the dispersion relations of the refractive index for these plates are shown for various values of m and L_c (mm) from Eq. (34), in air. In this example, for simplicity, we assumed that the cloaking system is placed in air, with $n=n_{air}=1$ for all wavelengths, and we ignored any coatings. The lens coatings can be modeled as additional thin plates as needed. We also only consider the visible spectrum (400-700 nm, for our discussion). A thicker compensating flat plate can reduce the required dispersion range, but this also affects the imaging quality, so we have purposely limited its thickness for this particular example.

Many of the solutions for positive refractive indices require anomalous dispersion, as shown in FIG. **33**. For the purposes of cloaking, at least in some implementations, such dispersion for broadband spectrum that have low losses are desired to imitate ambient space properly. This is similar to the finite bandwidth cloak made of anisotropic layers, as suggested by Kildishev et al., which also required strong anomalous dispersion combined with loss compensation. Costa and Silveirinha have suggested using nanowire metamaterial to provide such anomalous dispersion, and they numerically calculated their index of refraction to be near 2.9~3.1 for the entire visible spectrum. See J. T. Costa and M. G. Silveirinha, "Achromatic lens based on a nanowire material with anomalous dispersion," Opt. Express 20, 13915-13922 (2012). They can achieve these low loss, broadband, anomalous dispersion properties by utilizing the collective, spatial properties of metamaterials (called 'spatially dispersive' materials). This allows the ability to move beyond the high loss and narrow band properties of typical transparent materials, which stem from the Kramers-Kronig relations. In addition, Theisen and Brown have experimentally demonstrated anomalous dispersion for 500~1,000 nm, with Gallium implanted Silicon pads. These have indices of refraction near 2 or 4, each with variations of about 1 over the spectrum, depending on the doping level.

Negative-index metamaterials are good candidates for phase-correction as well for some implementations. The utilization of metamaterials was to be expected, since we expanded ray optics cloaking to the field cloaking aspect of transformation optics. Much progress is being made, both theoretically and experimentally, for creating negative index materials for broadband optical frequencies. Some examples of refractive index values demonstrated experimentally include between −3 and 0 for microwave frequencies in two-dimensions, n=0.63 at 1,200 nm to n=−1.23 at 1,775 nm in a low loss, 3D bulk material, and n=1 to −7.5 for 1.1-2.4 micron wavelengths.

By relaxing only omnidirectionality for an 'ideal' cloak, we have shown how to match the phase for the whole visible spectrum in some implementations. The phase-matching plate may require negative index metamaterials or anomalous dispersion in some instances, which are broadband and low loss, but current research has shown much progress in this regard. We had shown that building a 3D, broadband, macroscopic cloak, that works for the visible spectrum, can be fairly "easy" for ray optics in the small angle limit, and have extended this to the full-field using recently developed materials. As with typical lens designs, we expect that extending such cloaks to large angles will be possible.

It is interesting to note that the paraxial full-field cloak we presented does not require anisotropy, though being 3D, macroscopic, and broadband. Anisotropy seems to be a requirement when creating an omnidirectional cloak, from a paraxial cloak, not necessarily a property of 3D or field cloaks alone.

Potential Applications

The paraxial cloaking devices and related methods described herein may be applicable in many different fields, for example, such as in communications, transportation, military technologies, and medical devices. Additionally, the paraxial cloaking devices describe herein can alternatively be used to make visible an object that is obscured by another object. For example, a paraxial cloaking device can be used to view an internal organ or tissue that is hidden behind another internal organ or tissue that cannot be moved to make the hidden object directly visible. An enlarged cloaking device could allow portions of a building to be "see-through" on the outside, allow drivers to see directly behind the vehicle even if the space behind the driver is blocked (by objects or people, or the vehicle itself as in the case of semi-trucks), or furniture and objects inside a building to be invisible for interior design effects. A paraxial cloak can also hide a satellite orbiting the earth from a range of viewing angles, as well as a tank flanked by trees on each side, or offices or people on one end of a hallway. Some of these applications are shown in FIG. **31**.

In fact, the ABCD matrix for equation (5) can be solved for any waves or "rays" that can be described by similar matrices, to find an ideal "cloaking" device for such "rays." These include seismic waves, sound (acoustic waves), water waves, etc., not just optical rays. A cloaking device for such waves can allow water waves to not damage rigs by effectively cloaking the rigs, reduce damage from tsunamis in likewise manner, or allow certain geographical regions or buildings to not be heavily damaged by earthquakes by diverting the earthquake waves. These and other applications are shown in FIG. **31**.

Other variations are within the spirit of the present invention. For example, other variations including three and/or four lens cloaks that are split in half and connected the other way, or even combined, so as to make cloaking regions that cloak the center axis region (these can create two-dimensional cloaks, or can be manufactured for three-dimensions by rotating the split or combined lens designs completely around the center axes). FIG. **27** shows CODE V simulation results for a split three lens paraxial cloak. The simulation results show: (1) the path of incoming on-axis rays **32**-*i***1**, **32**-*i***2** and resulting respective outgoing on-axis rays **32**-*o***1**, **32**-*o***2**; (2) the path of incoming off-axis rays **34**-*i***1**, **34**-*i***2** and resulting respective outgoing off-axis rays **34**-*o***1**, **34**-*o***2**; (3) the path of incoming off-axis rays **36**-*i***1**, **36**-*i***2** and resulting respective outgoing off-axis rays **36**-*o***1**, **36**-*o***2**; and (4) a resulting cloaked region in the center of the split three lens paraxial cloak. FIG. **28** shows CODE V simulation results for a combined four lens paraxial cloak. The simulation results show: among other rays, (1) the path of incoming on-axis rays **32**-*i***1**, **32**-*i***2**, **32**-*i***3** and resulting respective outgoing on-axis rays **32**-*o***1**, **32**-*o***2**, **32**-*o***3**; (2) the path of incoming off-axis rays **34**-*i***1**, **34**-*i***2** and resulting respective outgoing off-axis rays **34**-*o***1**, **34**-*o***2**; (3) the path of incoming off-axis rays **36**-*i***1**, **36**-*i***2** and resulting respective outgoing off-axis rays **36**-*o***1**, **36**-*o***2**; and (4) resulting cloaked regions in the center of the combined four lens paraxial cloak, as well as the outer region/shell. Equation (5) can also be applied to five lenses or more, or for any optical system with arbitrary number of elements (lenses, mirrors, or anything else that can be described with an ABCD matrix). Also, lenses can be replaced by Fresnel lenses that can be larger, thinner, and lighter, though possibly with lesser image quality than regular lenses and/or achromatic lenses.

Thus, while the invention is susceptible to various modifications and alternative constructions, certain illustrated embodiments thereof are shown in the drawings and have been described above in detail. It should be understood, however, that there is no intention to limit the invention to the specific form or forms disclosed, but on the contrary, the intention is to cover all modifications, alternative constructions, and equivalents falling within the spirit and scope of the invention, as defined in the appended claims.

The use of the terms "a" and "an" and "the" and similar referents in the context of describing the invention (especially in the context of the following claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context. The terms "comprising," "having," "including," and "containing" are to be construed as open-ended terms (i.e., meaning "including, but not limited to,") unless otherwise noted. The term "connected" is to be construed as partly or wholly contained within, attached to, or joined together, even if there is something intervening. Recitation of ranges of values herein are merely intended to serve as a shorthand method of referring individually to each separate value falling within the range, unless otherwise indicated herein, and each separate value is incorporated into the specification as if it were individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., "such as") provided herein, is intended merely to better illuminate embodiments of the invention and does not pose a limitation on the scope of the invention unless otherwise claimed. No language in the specification should be construed as indicating any non-claimed element as essential to the practice of the invention.

Preferred embodiments of this invention are described herein, including the best mode known to the inventors for carrying out the invention. Variations of those preferred embodiments may become apparent to those of ordinary skill in the art upon reading the foregoing description. The inventors expect skilled artisans to employ such variations as appropriate, and the inventors intend for the invention to be practiced otherwise than as specifically described herein. Accordingly, this invention includes all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the above-described elements in all possible variations thereof is encompassed by the invention unless otherwise indicated herein or otherwise clearly contradicted by context.

All references, including publications, patent applications, and patents, cited herein are hereby incorporated by reference to the same extent as if each reference were individually and specifically indicated to be incorporated by reference and were set forth in its entirety herein.

What is claimed is:

1. A paraxial cloaking device having a cloaking volume and a reference optical axis, the device comprising:

an optical input receiving light rays; and

an optical output from which a continuous range of directions of the received light rays exit the paraxial cloaking device, the optical output being offset from the optical input, the cloaking volume being disposed between the optical input and the optical output,

wherein, for the received light rays having incoming directions non-parallel to the reference optical axis up to a first angle, each of the received light rays exits the cloaking device at the optical output substantially aligned with the respective received light ray and does not pass through the cloaking volume,

wherein the paraxial cloaking device has a unity magnification factor and is afocal.

2. The cloaking device of claim **1**, wherein the first angle is 3.5 degrees.

3. The cloaking device of claim **1**, wherein the first angle is 5 degrees.

4. The cloaking device of claim **1**, wherein the first angle is 15 degrees or more.

5. The cloaking device of claim **1**, wherein the cloaking device provides rotationally symmetric cloaking and comprises at least three lenses.

6. The cloaking device of claim **5**, wherein each of the three lenses is centered on the reference optical axis.

7. The cloaking device of claim **5**, wherein the clocking device comprises four lenses.

8. The cloaking device of claim **7**, wherein the four lenses are centered on the reference optical axis.

9. The cloaking device of claim **7**, wherein at least one of the four lenses comprises an achromatic doublet.

10. The cloaking device of claim **1**, wherein the cloaking device imparts an optical transformation to the received light rays defined by an ABCD matrix, wherein A=1, B=L/n, C=0, and D=1; L being a length between the optical input and the optical output, n being an index of refraction for the surrounding/ambient medium, before and/or after, the cloaking device.

11. The cloaking device of claim **1**, comprising:

a first lens centered on the reference optical axis;

a first mirror tilted from the reference optical axis;

a curved second mirror;

a third mirror tilted from the reference optical axis; and

a second lens centered on the reference optical axis,

wherein, the optical input comprises the first lens, the optical output comprises the second lens, and a light ray passing through the cloaking device follows an optical propagation path, the second mirror being disposed along the optical propagation path between the first and second lenses, the first mirror being disposed along the optical propagation path between the first lens and the second mirror, and the third mirror being disposed along the optical propagation path between the second mirror and the second lens.

12. The cloaking device of claim **11**, wherein:

the first and second lenses have the same focal length;

the first lens and the first mirror are separated by a distance equal to a distance separating third mirror and the second lens; and

the first mirror and the second mirror are separated a distance equal to a distance separating the second mirror and the third mirror.

13. The cloaking device of claim **1**, comprising:

a first lens centered on the reference optical axis;

a first mirror tilted from the reference optical axis;

a second lens;

a second mirror;

a third lens;

a third mirror tilted from the reference optical axis; and

a fourth lens centered on the reference optical axis,

wherein, the optical input comprises the first lens, the optical output comprises the fourth lens, and a light ray passing through the cloaking device follows an optical propagation path, the second mirror being disposed along the optical propagation path between the first and fourth lenses, the first mirror being disposed along the optical propagation path between the first lens and the second mirror, the third mirror being disposed along the optical propagation path between the second mirror and the fourth lens, the second lens being disposed along the optical path between the first and second mirror, and the third lens being disposed along the optical path between the second and third mirrors.

14. The cloaking device of claim **13**, wherein:

the first and fourth lenses have the same focal length;

the first lens and the first mirror are separated by a distance equal to a distance separating third mirror and the fourth lens;

the first mirror and the second lens are separated a distance equal to a distance separating the third lens and the third mirror;

the second and third lenses have the same focal length; and

the second lens and the second mirror are separated a distance equal to a distance separating the second mirror and the third lens.

15. A method of designing the cloaking device of claim **1**, comprising:

configuring a candidate cloaking device; and

calculating an ABCD matrix for the candidate cloaking device,

wherein the candidate cloaking device is configured such that A=1, B=L/n, C=0, and D=1; L being a length between the optical input and the optical output, n being an index of refraction for the surrounding/ambient medium, before and/or after, the cloaking device.

16. The method of claim **15**, further comprising adjusting at least one of a parameter or a position of at least one optical component of the candidate cloaking device so as to compensate for a thickness of at least one optical component of the candidate cloaking device.

17. The method of claim **15**, further comprising:

tracing at least one ray through the candidate cloaking device to assess a deviation between the location and direction of the ray prior to being received by the candidate cloaking device and after exiting the candidate cloaking device; and

adjusting at least one of a parameter and a position of at least one optical component of the candidate cloaking device based on the deviation assessment so as to reduce the deviation.

18. A method for designing a second cloaking device having a desired cloaking volume, the method comprising:

selecting a first cloaking device in accordance with claim **1**;

determining a scale factor indicative of a size difference between the desired cloaking volume and the cloaking volume of the first cloaking device; and

determining configuration parameters of the second cloaking device by scaling corresponding configuration parameters of the first cloaking device by the scale factor.

19. The cloaking device of claim **1**, wherein the device does not include a metamaterial optical component.

20. The cloaking device of claim **1**, wherein the device does not include a non-isotropic optical component.

21. The cloaking device of claim **1**, wherein the device does include a metamaterial optical component.

22. The cloaking device of claim **1**, further comprising a phase matching element.

23. The cloaking device of claim **22**, wherein the phase matching element comprises a flat plate having at least one index of refraction and a longitudinal length.

24. The cloaking device of claim **23**, wherein the phase matching element comprises an anomalous dispersion element.

25. The cloaking device of claim **23**, wherein the phase matching element comprises a metamaterial.

26. The cloaking device of claim **25**, wherein the index of refraction is at least one negative index of refraction.

27. The cloaking device of claim **1**, wherein the optical output comprises a profile and an axis, wherein the received light rays do not travel outside the profile of the optical output when viewed along the axis of the optical output.

28. The cloaking device of claim **1**, wherein, for the received light rays having incoming directions non-parallel to the reference optical axis up to the first angle, each of the received light rays exits the cloaking device at the optical output at an angle equal to an angle in which the received light ray entered the cloaking device.

29. The cloaking device of claim **1**, wherein, for the received light rays having incoming directions non-parallel to the reference optical axis up to the first angle, each of the received light rays exits the cloaking device at the optical output at a shifted position relative to a position in which the received light ray entered the cloaking device, the shifted position being equal to a tangent of an angle in which the received light ray entered the cloaking device multiplied by a length of the cloaking device.

* * * * *

US009559605B2

(12) **United States Patent**
Gudan et al.

(10) **Patent No.: US 9,559,605 B2**
(45) **Date of Patent: Jan. 31, 2017**

(54) **SYSTEM FOR AMBIENT ENERGY HARVESTING**

(71) Applicants: **Ken Gudan**, Sunnyvale, CA (US); **Sergey Chemishkian**, San Jose, CA (US); **Matthew S. Reynolds**, Seattle, WA (US); **Stewart Thomas**, Raleigh, NC (US); **Joshua Ensworth**, Seattle, WA (US)

(72) Inventors: **Ken Gudan**, Sunnyvale, CA (US); **Sergey Chemishkian**, San Jose, CA (US); **Matthew S. Reynolds**, Seattle, WA (US); **Stewart Thomas**, Raleigh, NC (US); **Joshua Ensworth**, Seattle, WA (US)

(73) Assignee: **Ricoh Co., Ltd.**, Tokyo (JP)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 82 days.

(21) Appl. No.: **14/198,357**

(22) Filed: **Mar. 5, 2014**

(65) **Prior Publication Data**

US 2015/0256097 A1 Sep. 10, 2015

(51) **Int. Cl.**
H02M 7/04 (2006.01)
H02M 3/338 (2006.01)
H02J 17/00 (2006.01)
H02J 7/34 (2006.01)

(52) **U.S. Cl.**
CPC ***H02M 3/338*** (2013.01); ***H02J 17/00*** (2013.01); ***H02M 7/04*** (2013.01); *H02J 7/345* (2013.01)

(58) **Field of Classification Search**
CPC H02J 17/00; H02M 7/04
USPC 363/16–21.04
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

8,621,245 B2 * 12/2013 Shearer G06K 19/0707 340/539.26
8,830,710 B2 * 9/2014 Perreault H03H 7/38 363/67
2004/0078662 A1 * 4/2004 Hamel B60C 23/0411 714/22
2006/0068749 A1 * 3/2006 Ismail H03F 3/45968 455/339
2009/0200985 A1 * 8/2009 Zane H01Q 1/2225 320/108

(Continued)

OTHER PUBLICATIONS

Sudevalayam et al., Energy Harvesting Sensor Nodes: Survey and Implicaitons, 2011, IEEE.*

Primary Examiner — Gary L Laxton
Assistant Examiner — Kyle J Moody
(74) *Attorney, Agent, or Firm* — Blakely, Sokoloff, Taylor & Zafman LLP

(57) **ABSTRACT**

A method and apparatus is disclosed herein for harvesting ambient energy. In one embodiment, an energy harvester comprises: a first RF rectifier to output a first voltage determined by rectified RF energy in response to received RF energy; a first energy reservoir coupled to the first RF rectifier to store energy at the first voltage; a DC/DC converter coupled to the first energy reservoir to convert the first voltage to a second voltage; a second reservoir coupled to the DC/DC converter to store energy at the second voltage, the second voltage being greater than the first voltage; and a third reservoir coupled to the second reservoir to receive energy transferred from the second reservoir periodically.

19 Claims, 6 Drawing Sheets

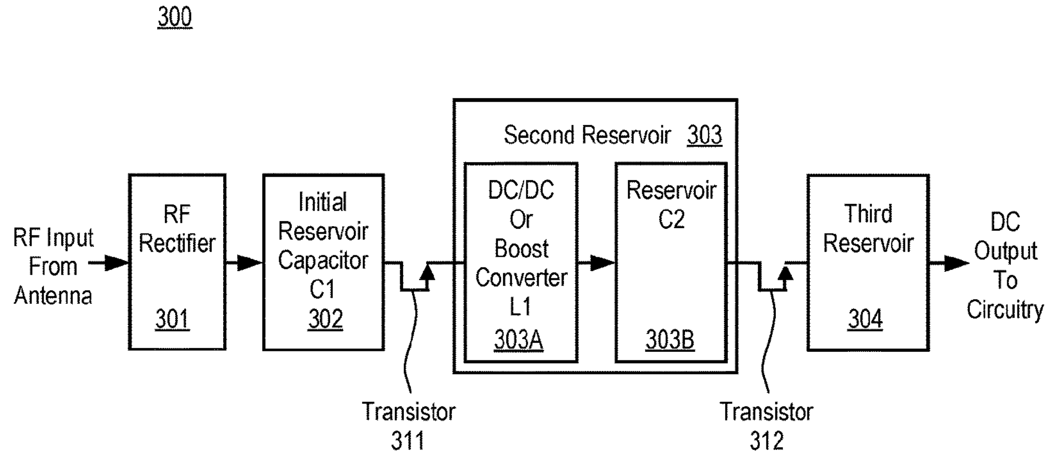

(56) **References Cited**

U.S. PATENT DOCUMENTS

2011/0093041	A1*	4/2011	Straka	A61N 1/36125 607/59
2013/0132010	A1*	5/2013	Winger	G01R 31/3679 702/63
2013/0214611	A1*	8/2013	Bae	H02M 7/06 307/104
2013/0278072	A1*	10/2013	Yoon	G01R 21/00 307/104
2014/0117771	A1*	5/2014	Kwon	H02J 5/005 307/104
2014/0125144	A1*	5/2014	Nakamura	B60L 11/182 307/104
2014/0191593	A1*	7/2014	Moh	H02J 5/005 307/149
2014/0290950	A1*	10/2014	Welsh	H01Q 1/248 166/299
2014/0354041	A1*	12/2014	Yoshi	H02J 7/025 307/9.1
2015/0035378	A1*	2/2015	Calhoun	A61B 5/04004 307/104
2015/0054345	A1*	2/2015	Monat	H01F 38/14 307/104
2015/0256015	A1*	9/2015	Gudan	H02J 5/005 307/104
2015/0256082	A1*	9/2015	Gudan	H02M 3/335 320/108

* cited by examiner

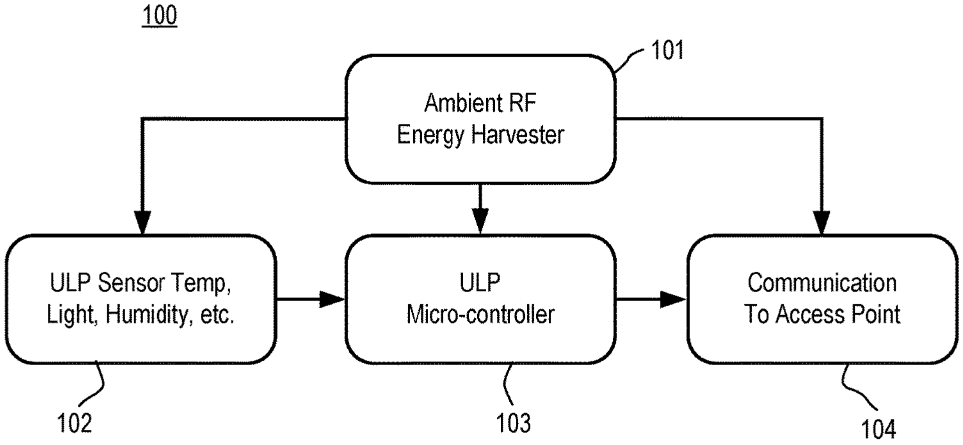
100
101
Ambient RF Energy Harvester
ULP Sensor Temp, Light, Humidity, etc.
ULP Micro-controller
Communication To Access Point
102
103
104

FIG. 1

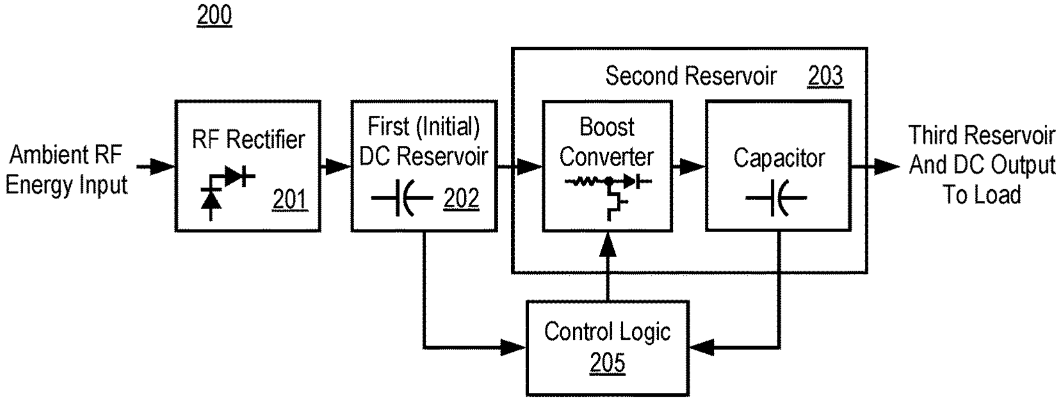
200
Ambient RF Energy Input
RF Rectifier
201
First (Initial) DC Reservoir
202
Second Reservoir
203
Boost Converter
Capacitor
Third Reservoir And DC Output To Load
Control Logic
205

FIG. 2

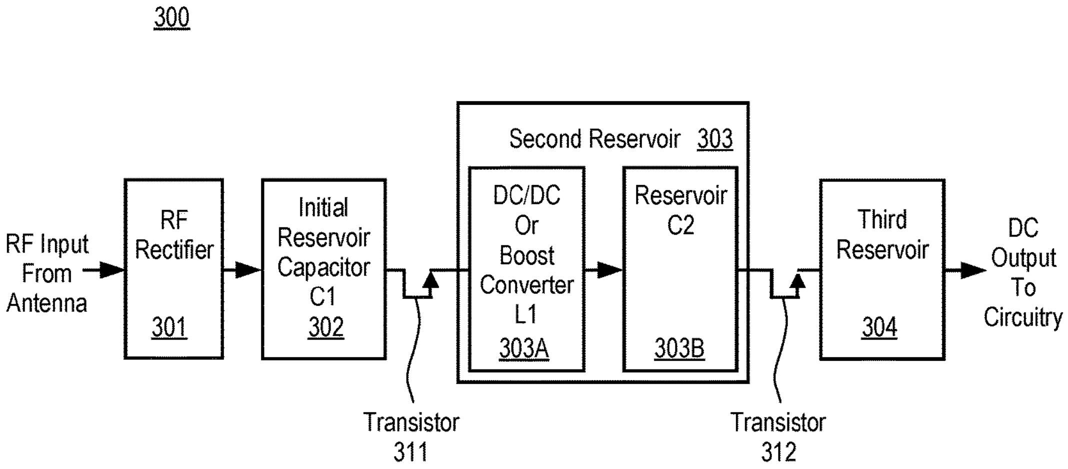
300
RF Input From Antenna
RF Rectifier
301
Initial Reservoir Capacitor C1
302
Second Reservoir 303
DC/DC Or Boost Converter L1
303A
Reservoir C2
303B
Third Reservoir
304
DC Output To Circuitry
Transistor 311
Transistor 312

FIG. 3

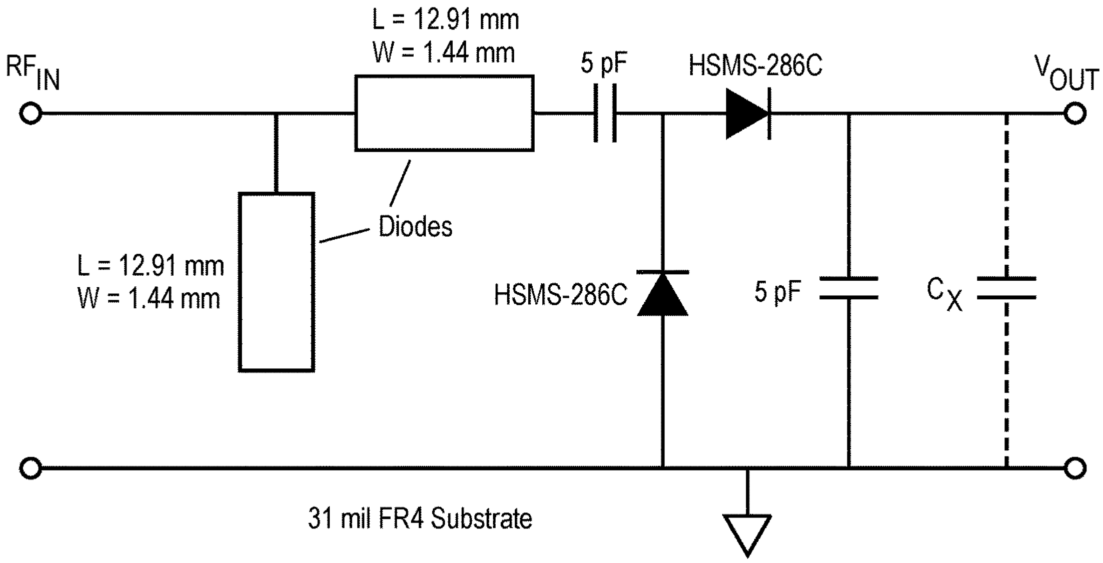
RF_{IN}
L = 12.91 mm
W = 1.44 mm
5 pF
HSMS-286C
V_{OUT}
Diodes
L = 12.91 mm
W = 1.44 mm
HSMS-286C
5 pF
C_X
31 mil FR4 Substrate

FIG. 4

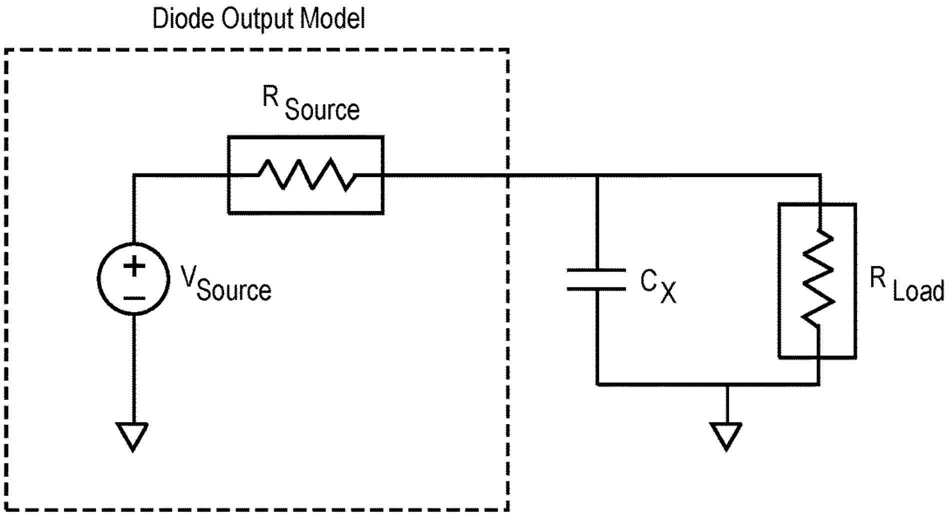
Diode Output Model
R_{Source}
V_{Source}
C_X
R_{Load}

FIG. 5

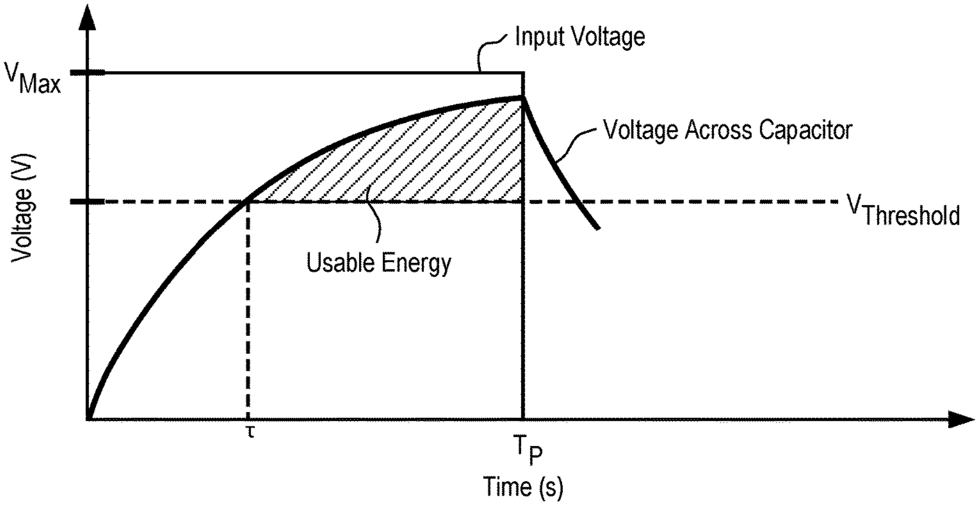
Input Voltage
V_{Max}
Voltage Across Capacitor
Voltage (V)
$V_{Threshold}$
Usable Energy
τ
T_P
Time (s)

FIG. 6A

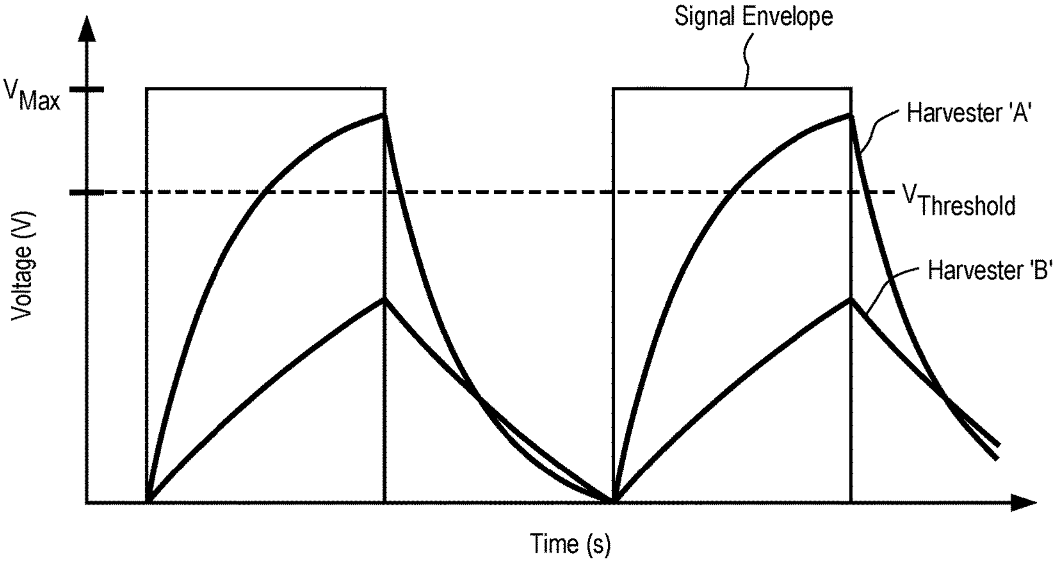
Signal Envelope
V_{Max}
Harvester 'A'
$V_{Threshold}$
Voltage (V)
Harvester 'B'
Time (s)

FIG. 6B

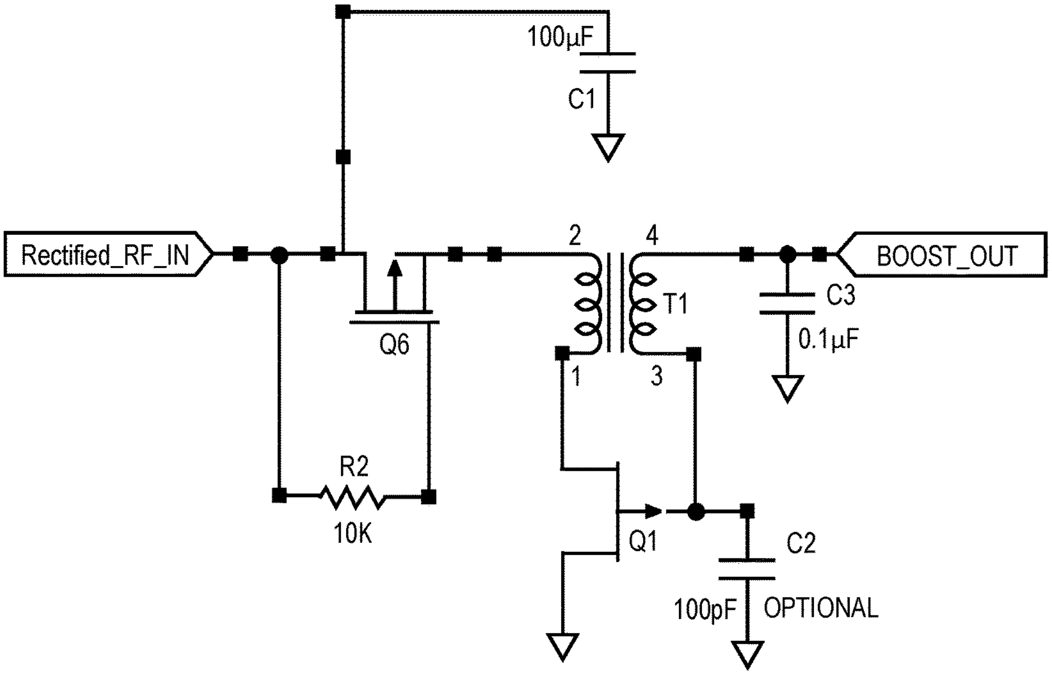
100µF
C1
Rectified_RF_IN
Q6
R2
10K
2
4
T1
1
3
Q1
C2
100pF
OPTIONAL
C3
0.1µF
BOOST_OUT

FIG. 7

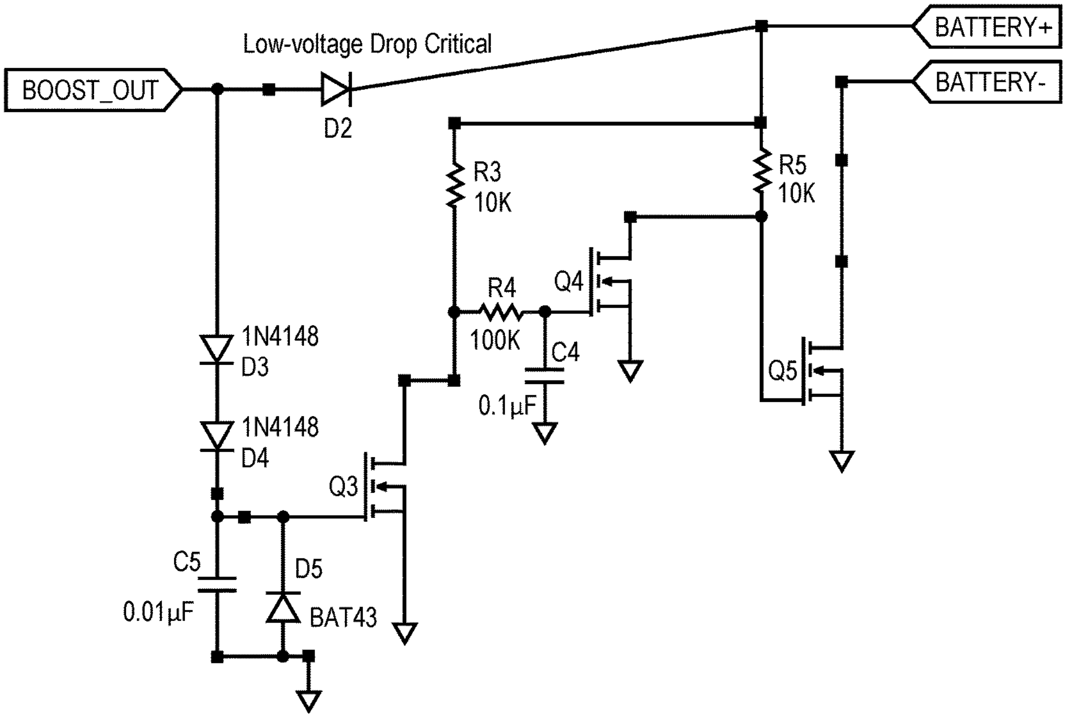
Low-voltage Drop Critical
BOOST_OUT
D2
BATTERY+
BATTERY-
R3
10K
R5
10K
Q4
R4
100K
C4
0.1μF
Q5
1N4148
D3
1N4148
D4
Q3
C5
0.01μF
D5
BAT43

FIG. 8

SYSTEM FOR AMBIENT ENERGY HARVESTING

FIELD OF THE INVENTION

Embodiments of the present invention relate to the field of wireless power transfer; more particularly, embodiments of the present invention relate to radio-frequency (RF) energy harvesting devices that harvest ambient RF energy and store the received RF energy.

BACKGROUND OF THE INVENTION

Next generation sensor networks may be powered by energy harvesting techniques to avoid requiring battery maintenance. Energy harvesting is a process by which energy is derived from external sources (e.g., radio frequency energy, solar power, thermal energy, wind energy, salinity gradients, or kinetic energy), captured and stored.

Energy may be harvested from radio frequency signals propagating wirelessly. With RF harvesting, wireless energy comes from a radio frequency transmitting device that is some distance away from a device that harvests energy from the radio frequency transmission. Properties of an energy harvester include its ability to harvest energy efficiently from available RF signals, its ability to store the harvested energy efficiently with minimal storage loss, and its ability to make the stored energy available to meet the voltage, current, and duty cycle requirements of a desired application.

One of the more popular forms of RF used today is Wi-Fi (also referred to as IEEE 802.11a/b/g/n etc.) communications. Today, most Wi-Fi communications are in the 2.4 GHz and 5.8 GHz frequency bands and there are many local area networks that are based on Wi-Fi in which access points enable Wi-Fi clients to gain access to networks such as the Internet. Furthermore, the 2.4 GHz and 5.8 GHz bands also support other networking standards, such as Zigbee and Bluetooth, and other proprietary networks, each transmitting energy by communicating in this same frequency band. Additionally there are other frequency bands that support different communication protocols, each of which transmit energy when they are communicating. These include, for example, digital television (DTV) and Global System for Mobile Communications (GSM) signals.

In a traditional RFID-like setup the RFID readers (interrogators) usually deliver RF power as a continuous wave (CW), i.e. a sinusoidal signal of a particular frequency. In contrast to systems designed for use with continuous wave (CW) signals such as RFID, some RF energy harvesters are designed for use with non-CW signals such as WiFi and GSM. These waveform aware harvesters include ambient energy harvesters that harvest signals of opportunity that are already present in the environment. For these types of harvesters, the system designer must work around the waveforms and power levels that are already present in the environment.

SUMMARY OF THE INVENTION

A method and apparatus is disclosed herein for harvesting ambient energy. In one embodiment, an energy harvester comprises: a first RF rectifier to output a first voltage determined by rectified RF energy in response to received RF energy; a first energy reservoir coupled to the first RF rectifier to store energy at the first voltage; a DC/DC converter coupled to the first energy reservoir to convert the first voltage to a second voltage; a second reservoir coupled to the DC/DC converter to store energy at the second voltage, the second voltage being greater than the first voltage; and a third reservoir coupled to the second reservoir to receive energy transferred from the second reservoir periodically.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be understood more fully from the detailed description given below and from the accompanying drawings of various embodiments of the invention, which, however, should not be taken to limit the invention to the specific embodiments, but are for explanation and understanding only.

FIG. **1** is a block diagram of one embodiment of a sensor device.

FIG. **2** is a block diagram of one embodiment of an RF energy harvester.

FIG. **3** is another more detailed block diagram of one embodiment of an RF energy harvester.

FIG. **4** illustrates a diode-based rectifier schematic.

FIG. **5** illustrates a diode output circuit model.

FIG. **6**A illustrates the usable energy stored in capacitor C_X during one transmission burse of duration T_p.

FIG. **6**B illustrates examples of good and bad choices of C_X.

FIG. **7** illustrates one embodiment of a self-excited oscillator.

FIG. **8** illustrates one embodiment of a charger schematic.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

A system for ambient energy harvesting and method for doing the same are described. In one embodiment, the energy harvesting system includes three DC reservoirs to store energy generated from ambient radio frequency (RF) input signals. These reservoirs are coupled together in a cascaded fashion to provide DC energy to a load.

In one embodiment, the ambient RF energy harvester is part of a sensor/tag device and is used to provide energy to one or more components of the sensor/tag device such as, for example, the sensor, a controller, a communication unit a processing device, etc.

In the following description, numerous details are set forth to provide a more thorough explanation of the present invention. It will be apparent, however, to one skilled in the art, that the present invention may be practiced without these specific details. In other instances, well-known structures and devices are shown in block diagram form, rather than in detail, in order to avoid obscuring the present invention.

Some portions of the detailed descriptions which follow are presented in terms of algorithms and symbolic representations of operations on data bits within a computer memory. These algorithmic descriptions and representations are the means used by those skilled in the data processing arts to most effectively convey the substance of their work to others skilled in the art. An algorithm is here, and generally, conceived to be a self-consistent sequence of steps leading to a desired result. The steps are those requiring physical manipulations of physical quantities. Usually, though not necessarily, these quantities take the form of electrical or magnetic signals capable of being stored, transferred, combined, compared, and otherwise manipulated. It has proven convenient at times, principally for reasons of common usage, to refer to these signals as bits, values, elements, symbols, characters, terms, numbers, or the like.

It should be borne in mind, however, that all of these and similar terms are to be associated with the appropriate physical quantities and are merely convenient labels applied to these quantities. Unless specifically stated otherwise as apparent from the following discussion, it is appreciated that throughout the description, discussions utilizing terms such as "processing" or "computing" or "calculating" or "determining" or "displaying" or the like, refer to the action and processes of a computer system, or similar electronic computing device, that manipulates and transforms data represented as physical (electronic) quantities within the computer system's registers and memories into other data similarly represented as physical quantities within the computer system memories or registers or other such information storage, transmission or display devices.

The present invention also relates to apparatus for performing the operations herein. This apparatus may be specially constructed for the required purposes, or it may comprise a general purpose computer selectively activated or reconfigured by a computer program stored in the computer. Such a computer program may be stored in a computer readable storage medium, such as, but is not limited to, any type of disk including floppy disks, optical disks, CD-ROMs, and magnetic-optical disks, read-only memories (ROMs), random access memories (RAMs), EPROMs, EEPROMs, magnetic or optical cards, or any type of media suitable for storing electronic instructions, and each coupled to a computer system bus.

The algorithms and displays presented herein are not inherently related to any particular computer or other apparatus. Various general purpose systems may be used with programs in accordance with the teachings herein, or it may prove convenient to construct more specialized apparatus to perform the required method steps. The required structure for a variety of these systems will appear from the description below. In addition, the present invention is not described with reference to any particular programming language. It will be appreciated that a variety of programming languages may be used to implement the teachings of the invention as described herein.

A machine-readable medium includes any mechanism for storing or transmitting information in a form readable by a machine (e.g., a computer). For example, a machine-readable medium includes read only memory ("ROM"); random access memory ("RAM"); magnetic disk storage media; optical storage media; flash memory devices; etc.

One Embodiment of a Sensor Device

FIG. **1** is a block diagram of one embodiment of a sensor device. The sensor device may be part of a tag. Referring to FIG. **1**, sensor device **100** comprises an ambient radio-frequency (RF) energy harvester **101**, an ultra-low-power (ULP) sensor **102** (e.g., temperature, light, humidity, etc.), a micro-controller **103**, and a communication unit **104**. Although not shown, sensor device **100** includes one or more antennas. The antennas may be part of communication unit **104** or may be separate, such as in the case of an antenna used to harvest energy.

In one embodiment, RF energy harvester **101** harvests ambient RF power to run ultra-low power tags and/or sensors **102**, micro-controller **103**, and/or communication unit **104**, based on received RF input energy received by an antenna (not shown) of sensor device **100**. There are a number of sources of ambient RF power, including, but not limited to WiFi, DTV and GSM signals. Other sources of ambient RF power include, for example, toys, proprietary communication protocols, and even stray energy leakage from devices, such as microwave ovens. The antenna receiving the RF energy input may be part of communication unit **104** or a separate antenna. The antenna(s) may be designed to receive the energy in multiple frequency bands, or there may be separate antenna(s) for these different frequency bands.

In one embodiment, sensor **102** comprises an ultra-low-power (ULP) sensor and/or tag and micro-controller **103** comprises an ultra-low-power (ULP) micro-controller. Sensor **102** may comprise a temperature sensor, a light sensor, a motion sensor, air flow sensor, image sensor, accelerometer, gyroscope, gas detector, microphone, atmospheric pressure sensor, humidity sensor, mechanical pressure sensor, etc. In one embodiment, sensor **102**, as a ULP sensor, has a power requirement of 1-20 uW and a duty cycle of 1% to 10% (where the duty cycle varies in response to available energy over a period of time or the specific sensing needs, usually seconds to hours).

In one embodiment, micro-controller **103** performs control operations and/or processing for the sensor device **100**. In one embodiment, controller **103** is responsive to and receives sensor data from sensor **102** and provides it to communication unit **104** for transmission (e.g., wireless transmission) from device **100**. Controller **103** may also log sensor data for a time, and transmit the entire log to communication unit **104** for transmission later. Note that in this case the sensing rate and the communication rate can be different.

Communication unit **104** provides a communication mechanism for sensor device **100**. In one embodiment communication unit **104** is a wireless communication unit. In one embodiment, communication unit **104** has an RF transmission along with an antenna for receiving and transmitting RF signals, including RF signals transmitted with an access point or other locally positioned device.

In one embodiment, the ambient RF energy harvester is optimized for overall efficiency and a minimum of start-up power. These optimizations improve the range at which energy can be usefully harvested and improve the number of places where ambient energy is practical, particularly in view of the fact that ambient RF energy levels are small and often bursty (traffic-dependent), rectified voltages are very low (<100 mV) at typical ranges, and typical Complementary Metal Oxide Semiconductor (CMOS) loads such as MCU etc. require 800+ mV to operate.

An energy harvester and method for performing energy harvesting are described. In one embodiment, the energy harvester converts incident radio frequency (RF) energy, or ambient RF energy, to direct current energy. In one embodiment, the energy harvester comprises: a first RF rectifier to output a first voltage determined by rectified RF energy in response to received RF energy; a first energy reservoir coupled to the first RF rectifier to store energy at the first voltage; a DC/DC boost converter coupled to the first energy reservoir to convert the first voltage to a second voltage; a second reservoir coupled to the DC/DC converter to store energy at the second voltage, where the second voltage is greater than the first voltage; and a third reservoir (e.g., a longer-term energy storage device such as a low-leakage capacitor or battery, etc.) coupled to the second reservoir to receive energy transferred from the second reservoir periodically.

In one embodiment, the RF rectifier comprises one selected from a group consisting of: a Schottky diode, a P-N junction diode, a diode-connected field effect transistor, and a tunnel diode.

In one embodiment, the first reservoir comprises a capacitor with a capacitance value determined at least in part by the envelope of the received RF energy. In one embodiment, the first reservoir comprises a capacitor having a value based on stored energy above a threshold voltage given ambient RF input power and a duty cycle of the input power. In another embodiment, the first reservoir comprises a first capacitor and the second reservoir comprises a second capacitor, wherein capacitance of the second capacitor is approximately equal to a product of the first capacitor capacitance times a square of a ratio of an expected value of the second voltage divided by an expected value of the first voltage.

In one embodiment, the second reservoir comprises an inductor used to boost the voltage, wherein energy stored in the inductor is approximately equal to energy stored in the first reservoir.

In one embodiment, the DC/DC boost converter comprises a self-excited oscillator. In one embodiment, the self-excited oscillator comprises a transistor coupled to a transformer. In one embodiment, the frequency of the self-excited oscillator is controlled at least in part by an envelope of the received RF energy. In one embodiment, the DC/DC converter is a switch mode converter with switching intervals controlled at least in part by an envelope of the received RF energy.

In one embodiment, the third reservoir is coupled to the second reservoir via a switch, which is closed periodically to transfer energy from the second reservoir to the third reservoir. In one embodiment, the switch is controlled at least in part by the envelope of the received RF energy. In one embodiment, the third reservoir is coupled to a load, and the first voltage is insufficient to operate the load directly, and the third voltage is sufficient to operate the load. In one embodiment, the third reservoir has an energy storage capacity that is greater than an energy storage capacity of the second reservoir.

FIG. **2** is a block diagram of one embodiment of an RF energy harvester. Referring to FIG. **2**, RF energy harvester **200** comprises RF rectifier **201**, a first direct current (DC) reservoir (storage) **202**, a boost converter and second reservoir **203**, a third reservoir **204**, and control logic **205**.

RF rectifier **201** receives ambient RF energy input received from an antenna and rectifies the RF energy to produce a DC voltage.

First DC reservoir **202** stores the voltage generated from the rectified RF signals received by device **100**. In one embodiment, reservoir **202** comprises a capacitor. In one embodiment, the size (value) of the capacitor that is selected for use is based on the stored energy above a voltage threshold for the ambient RF input power.

Second DC reservoir **203** receives the voltage and increases the voltage. To increase the voltage, second DC reservoir **203** includes a DC/DC converter. In one embodiment, the DC/DC converter comprises a boost converter which includes a self-excited oscillator for operation. The self-excited oscillator includes a junction field effect transistor (JFET) coupled with a transformer. In one embodiment, the amount of the boost conversion is a function of the input energy, and the DC/DC converter is designed to maximize the output voltage as a function of input voltage in this low-energy environment. FIG. **7** illustrates one embodiment of a self-excited oscillator. Referring to FIG. **7**, the embodiment of the self-excited oscillator includes an input (Rectified RF_IN), an output (BOOST_OUT), capacitors C**1** (100 μF); optional C**2** (100 pF); and C**3** (0.1 μF), transistors Q**6** and Q**1**, a transformer (T**1**), and a resistor R**2** (10K) coupled as shown. FIG. **8** illustrates one embodiment of a charger schematic that receives the output of the boost converter, such as, for example, the one shown in FIG. **7**. Referring to FIG. **8**, the embodiment of the charger schematic includes an input (BOOST_OUT), an output to a battery (Battery+ and Battery−), diodes D**2** for preventing a low-voltage drop; D**3** (1N4148); D**4** (1N4148); D**5** (BAT43), capacitors C**4** (0.1 μF) and C**5** (0.01 μF), transistors Q**3**, Q**4**, and Q**5**, and resistors R**3** (10K), R**4** (100K), and R**5** (10K).

The upconverted voltage from the DC/DC converter is stored in a storage device of second reservoir **203**. In one embodiment, second reservoir **203** uses a capacitor to store the upconverted voltage. In another embodiment, second reservoir **203** uses an inductor to store the upconverted voltage, but this is a much shorter-term storage option.

When the energy stored in second reservoir **203** reaches a certain level, the energy is transferred to a third reservoir (not shown). Thus, the transfer occurs periodically at a rate determined by the amount of input energy. In one embodiment, the third reservoir comprises a battery, a capacitor, a supercap, etc. The third reservoir provides DC energy to a load.

In one embodiment, the load comprises a sensor, processor or microcontroller, and/or a communication unit. In one embodiment, the DC voltage from the third reservoir is able to run the load, while the voltage produced by rectifier **201** would not be able to do so.

Control logic **205** controls the energy harvesting process. In one embodiment, control logic **205** comprises a separate unit that generates control signals to control each of the reservoirs **202-3203**. In another embodiment, the control is distributed through various discrete gating components such that a separate control unit is not necessary but the function is still performed.

FIG. **3** is another more detailed block diagram of one embodiment of an RF energy harvester. Referring to FIG. **3**, RF energy harvester **300** comprises RF rectifier **301**, a first direct current (DC) reservoir (storage) **302**, second reservoir **303**, and a third reservoir **304**. In one embodiment, the input to RF rectifier **301** is an RF input from an antenna on the sensor device. RF rectifier **301** converts the RF input signals to a voltage. In one embodiment, the voltage may range between 10 mV and 500 mV. The voltage is stored in reservoir **302**. In one embodiment, reservoir **302** comprises a ceramic capacitor 1 uF to 100 uF or more.

In one embodiment, the stored energy from reservoir **302** is transferred to second reservoir **303** via a transistor **311**, self-gated as the voltage rises in reservoir **302**. In one embodiment, second reservoir **303** comprises a boost converter **303**A. The boost converter **303**A operates to upconvert the voltage stored in reservoir **302** and store that upconverted voltage into reservoir **303**B.

The energy stored in second reservoir **303** is transferred to third reservoir **304**. In one embodiment, third reservoir **304** comprises a tantalum capacitor, supercap 100 uF to 1 F or a battery. In one embodiment, transistor **312** (e.g., a FET) couples second reservoir **303** to third reservoir **304**, which turns on when second reservoir **303** reaches an optimal voltage. In one embodiment, transistor **312** is controlled at least in part by the envelope of the received RF energy such that when the voltage stored in reservoir **303**B reaches a certain watermark level, transistor **312** closes/turns on and the energy is transferred. In one embodiment, through the selection of high gate-threshold transistors, the voltage can build up fairly high in reservoir **303** (as much as 1-2V) before the transistor **312** closes/turns on. Transistor **312** closes/turns on automatically when the gate threshold of the transistor design is reached in reservoir **303**. As soon as transistor **312** closes/turns on, the energy is transferred from reservoir **303** into reservoir **304**, thus depleting the energy in reservoir **303**, and the gate threshold falls below the watermark level and the transistor **312** closes/opens again. The process then repeats when the voltage in reservoir **303** becomes sufficiently high.

Note that transistor **311** and **312** may be replaced with other types of switches.

Waveform Aware Harvesting

In one embodiment, the RF waveform harvester comprises a RF to DC converter that is optimized for use with non-continuous wave (CW) signals. In another embodiment, the RF waveform harvesters are waveform aware and are able to perform ambient energy harvesting based on a waveform model, where the available RF energy is in the form of communication waveforms of a variety of types. Examples of commonly-harvested ambient RF energy sources with non-continuous wave signals include, but are not limited to, digital television (DTV) broadcasts, cellular base station transmissions (e.g., GSM, Code Division Multiple Access (CDMA)), Wi-Fi transmissions, etc.

For example, in one embodiment, WiFi is the communications waveform that is the ambient form of energy to be harvested. WiFi is bursty and has low amplitude compared to other sources of energy, where a bursty signal has a high peak to average ratio over ms to s time scale. For example, some WiFi signals have WiFi transmission burst times on the order of 1 ms. For a bursty signal, the harvester can only contribute energy to the reservoir capacitor (C_x) when the ambient source is transmitting.

FIG. **4** illustrates a diode-based harvester schematic. Referring to FIG. **4**, in one embodiment, the harvester is designed to be well matched at input power levels below −20 dBm where the majority of ambient Wi-Fi signals are expected to fall. Thus, harvester is characterized both in terms of its input match from 2.4 GHz to 2.5 GHz, as well as for harvesting efficiency at its maximum power point (MPP). Harvesting efficiency is defined as the ratio of input RF power at the harvester to harvested DC power at the load (η=PDC/PRF).

The purpose of capacitor C_X in FIG. **4** is to store charge for any subsequent active electronic circuitry. In one embodiment, typical circuit power-up threshold voltages V_{TH} equals 0.8 V, as is typical for CMOS logic in commodity processes. In another embodiment, the power-up threshold voltages V_{TH} equals 50 mV, as is typical for ultra-low startup voltage boost converters using JFETs. Alternatively, specially doped CMOS transistors, floating gate devices, or Silicon on Insulator (SOI) transistors may be capable of operation down to voltages approaching V_{TH}=50-100 mV as well.

If the usable energy is defined as the portion of the stored energy above some threshold voltage V_{TH}, the following describes how to find a value of C_X given the ambient RF input power and the signal's duty cycle. This is based on the maximum power point curve for the harvester, the typical burst timings for the ambient signal of interest (e.g., WiFi with some traffic model), and a time domain response for the harvester as a function of C_X. Thus, in one embodiment, the storage capacitor selection C_X is optimized on the time scale of network traffic for the signal being harvested.

FIG. **5** illustrates a diode output circuit model. Because the diode is a non-linear device, the power delivered to the load varies dramatically with load impedance. The Maximum Power Point Tracking (MPPT) circuit functions as a load impedance converter to maximize the power delivered to the load, given a particular nonlinear operating point.

Referring to FIG. **5**, the diode output model consists of an open-circuit voltage V_{source} along with a nonlinear source resistance R_{source}. The pair, V_{source}, R_{source}, represents a fixed operating point at maximum power-point (MPP). The load resistance R_{load} includes any leakage currents in C_X as well as the quiescent current of the active circuitry attached to the harvester. In other words, the load impedance R_{load} is chosen a priori based on the MPP of the harvester at a given ambient signal level (the maximum power point of the harvester given the input power). Using this model, only the effect of varying C_X (not other system conditions) is considered.

FIG. **6**A illustrates the usable energy stored in capacitor C_X during one transmission burse of duration T_p. Referring to FIG. **6**A, the usable energy in C_X is represented as the shaded portion of the curve above the voltage threshold and below the voltage across the capacitor.

FIG. **6**B illustrates examples of good and bad choices of C_X. If C_X is too large, the active circuitry will never be powered because the input energy is insufficient to generate voltage above the threshold. On the other hand, if C_X is minimized, little energy will be stored in the reservoir C_X and the runtime of the active circuitry will not extend beyond the transmission intervals of the RF source. Referring to FIG. **6**, harvesters 'A' and 'B' represent different choices for C_X with the same conditions otherwise. In harvester A, C_X looks appropriate as the output voltage exceeds V_{TH} during each burst, while in harvester B, C_X is too big and the output voltage never exceeds V_{TH} so there is no usable energy.

An expression is derived for the useful-energy-maximizing reservoir capacitance C_X as a function of ambient RF signal duty cycle and the load impedance RL at a diode harvester's maximum power point. The expression may be used to calculate the usable energy as a function of the known parameters and the unknown capacitance C_X.

The total energy in C_X may be expressed as follows:

$$E = \frac{1}{2}CV^2$$

where C is the capacitance and V is the terminal voltage of C_X. It is assumed that the energy stored on the capacitor is only usable when the capacitor voltage is above the desired threshold V_{TH}. The usable energy is thus the difference between the stored energy at the end of an RF transmission (at time t=T_p) and the stored energy when C_X is charged only to V_{TH} and is expressed as follows:

$$E_{usable} = \frac{1}{2}C\left[V_{T_p}^2 - V_{th}^2\right]$$

where V_{Tp} is the terminal voltage on C_X at the end of an RF transmission (e.g., at the end of each burst) of duration T_p and is expressed as:

$$V_{T_p} = V_{src}(1 - e^{-T_p/R_L C})$$

By substitution, the equation for the usable energy may be re-written in terms of C and the known values of R_L, V_{TH} and T_p as follows:

$$E_{usable} = \frac{1}{2}CV_{T_p}^2\left[e^{-\frac{2T_p}{R_L C}} - 2e^{-\frac{T_p}{R_L C}} + \left(1 - \left(\frac{V_{th}}{V_{T_p}}\right)^2\right)\right]$$

The equation may be solved for E_{usable} as a function of C_X.

Thus, a new class of RF energy harvesters, referred to herein as waveform aware harvesters, have been described. These RF energy harvesters are RF to DC converters that are optimized for their performance with non-CW signals.

Whereas many alterations and modifications of the present invention will no doubt become apparent to a person of ordinary skill in the art after having read the foregoing description, it is to be understood that any particular embodiment shown and described by way of illustration is in no way intended to be considered limiting. Therefore, references to details of various embodiments are not intended to limit the scope of the claims which in themselves recite only those features regarded as essential to the invention.

We claim:

1. An energy harvester to convert incident radio frequency (RF) energy associated with an RF signal to direct current energy, the energy harvester comprising:

a first RF rectifier to output a first voltage determined by rectified RF energy in response to received RF energy;

a first energy reservoir coupled to the first RF rectifier to store energy at the first voltage;

a DC/DC converter coupled to the first energy reservoir to convert the first voltage to a second voltage;

a second reservoir coupled to the DC/DC converter to store energy at the second voltage, the second voltage being greater than the first voltage; and

a third reservoir coupled to the second reservoir to receive energy transferred from the second reservoir periodically and coupled to a load, wherein the third reservoir comprises a capacitor having a capacitance value chosen to control an amount of usable energy stored in the third reservoir as a function of an RF transmission duration, the impedance of the load, and a threshold voltage, the usable energy being a portion of the stored energy above the threshold voltage.

2. The energy harvester defined in claim **1** wherein the third reservoir is coupled to the second reservoir via a first switch, the first switch being closed periodically to transfer energy from the second reservoir to the third reservoir.

3. The energy harvester defined in claim **2** wherein the first switch is controlled at least in part by an envelope of the received RF energy signal.

4. The energy harvester defined in claim **1** wherein the first voltage is insufficient to operate the load directly, and the third voltage is sufficient to operate the load.

5. The energy harvester defined in claim **1** wherein the DC/DC converter comprises a self-excited oscillator.

6. The energy harvester defined in claim **5** wherein frequency of the self-excited oscillator is controlled at least in part by an envelope of the received RF energy signal.

7. The energy harvester defined in claim **1** wherein the DC/DC converter comprises a boost converter.

8. The energy harvester defined in claim **1** wherein the first reservoir comprises a capacitor with a capacitance determined at least in part by an envelope of the received RF energy signal.

9. The energy harvester defined in claim **1** wherein the DC/DC converter is a switch mode converter with switching intervals controlled at least in part by an the envelope of the received RF energy signal.

10. The energy harvester defined in claim **1** wherein the RF rectifier comprises one or more diodes selected from a group consisting of: a Schottky diode, a P-N junction diode, a diode-connected field effect transistor, and a tunnel diode.

11. The energy harvester defined in claim **1** wherein the third reservoir has an energy storage capacity that is greater than an energy storage capacity of the second reservoir.

12. The energy harvester defined in claim **1** wherein the second reservoir comprises an inductor, wherein energy stored in the inductor is approximately equal to energy stored in the first reservoir.

13. The energy harvester defined in claim **1** wherein the capacitance value is chosen to maximize the amount of usable energy stored in the third reservoir as a function of the RF transmission duration, the impedance of the load, and the threshold voltage.

14. An energy harvester to convert incident radio frequency (RF) energy to direct current energy, the energy harvester comprising:

a first RF rectifier to output a first voltage determined by rectified RF energy in response to received RF energy;

a first energy reservoir coupled to the first RF rectifier to store energy at the first voltage;

a DC/DC converter coupled to the first energy reservoir to convert the first voltage to a second voltage;

a second reservoir coupled to the DC/DC converter to store energy at the second voltage, the second voltage being greater than the first voltage; and

a third reservoir coupled to the second reservoir to receive energy transferred from the second reservoir periodically, wherein the first reservoir comprises a first capacitor and the second reservoir comprises a second capacitor, wherein capacitance of the second capacitor is approximately equal to a product of the first capacitor capacitance times a square of a ratio of an expected value of the second voltage divided by an expected value of the first voltage.

15. A device comprising:

an antenna;

a load; and

an energy harvester coupled to the antenna and the load, wherein the energy harvester comprises

a first RF rectifier to output a first voltage determined by rectified RF energy in response to received RF energy;

a first energy reservoir coupled to the first RF rectifier to store energy at the first voltage;

a DC/DC converter coupled to the first energy reservoir to convert the first voltage to a second voltage;

a second reservoir coupled to the DC/DC converter to store energy at the second voltage, the second voltage being greater than the first voltage; and

a third reservoir coupled to the second reservoir to receive energy transferred from the second reservoir periodically, wherein the third reservoir comprises a capacitor having a capacitance value chosen to control an amount of usable energy stored in the third reservoir as a function of an RF transmission duration, the impedance of the load, and a threshold voltage, the usable energy being a portion of the stored energy above the threshold voltage.

16. The device defined in claim **15** wherein the third reservoir is coupled to the second reservoir via a switch, the switch being closed periodically to transfer energy from the second reservoir to the third reservoir.

17. The device defined in claim **15** wherein the third reservoir is coupled to the load, and wherein the first voltage is insufficient to operate the load directly, and the third voltage is sufficient to operate the load.

18. The device defined in claim **15** wherein the DC/DC converter comprises a boost converter.

19. The energy harvester defined in claim **15** wherein the capacitance value is chosen to maximize the amount of usable energy stored in the third reservoir as a function of the RF transmission duration, the impedance of the load, and the threshold voltage.

* * * * *

US009564268B2

(12) **United States Patent**
Dinwiddie et al.

(10) **Patent No.: US 9,564,268 B2**
(45) **Date of Patent: Feb. 7, 2017**

(54) **POWER RECEIVER FOR EXTRACTING POWER FROM ELECTRIC FIELD ENERGY IN THE EARTH**

(71) Applicant: **Earth Energies, Inc.**, Johns Creek, GA (US)

(72) Inventors: **John Dinwiddie**, Cary, NC (US); **Terry L. Wright**, Suwanee, GA (US); **David R. Ames**, Johns Creek, GA (US)

(73) Assignee: **Earth Energies, Inc.**, Johns Creek, GA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **14/509,772**

(22) Filed: **Oct. 8, 2014**

(65) **Prior Publication Data**

US 2015/0102675 A1 Apr. 16, 2015

Related U.S. Application Data

(60) Provisional application No. 61/889,894, filed on Oct. 11, 2013.

(51) **Int. Cl.**

H02J 4/00	(2006.01)
H01F 38/14	(2006.01)
H01T 9/00	(2006.01)
H05F 7/00	(2006.01)

(52) **U.S. Cl.**
CPC ***H01F 38/14*** (2013.01); ***H01T 9/00*** (2013.01); *H01F 2038/146* (2013.01); *H05F 7/00* (2013.01); *Y10T 307/549* (2015.04)

(58) **Field of Classification Search**
CPC H02J 4/00; H05F 7/00; H01R 4/66; E02D 5/803; E02D 5/805
USPC 343/846, 849; 307/149, 151, 106–108; 52/161; 174/6, 7
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

29,793 A *	8/1860	Vion	....	B23B 51/02 408/212
126,356 A *	4/1872	Ward	....	G05B 11/016 174/4 R
414,943 A *	11/1889	Dewey	....	G05B 11/016 174/3
645,576 A *	3/1900	Tesla	....	H04L 25/4902 375/259
674,427 A *	5/1901	Palencsar	....	G05B 11/016 174/4 R

(Continued)

FOREIGN PATENT DOCUMENTS

DE	102008029441 A1 *	12/2009	.... H05F 7/00
WO	2008005628 A2	1/2008	

OTHER PUBLICATIONS

Patent Translation for DE102008029441, EPO ESPACENET, Printed Sep. 23, 2016.*

Primary Examiner — Fritz M Fleming
(74) *Attorney, Agent, or Firm* — Coats and Bennett, PLLC

(57) **ABSTRACT**

A resonant transformer connected between a ground terminal and elevated terminal draws current from the earth's electric field through a primary winding of the transformer. An impulse generator applies a high voltage impulse to the primary winding of the resonant transformer to cause current to flow from the ground terminal through the primary winding. The flow of current through the primary winding of the resonant transformer induces a current in the secondary winding, which may be converted and filtered to a usable form, e.g. 60 Hz AC or DC.

12 Claims, 8 Drawing Sheets

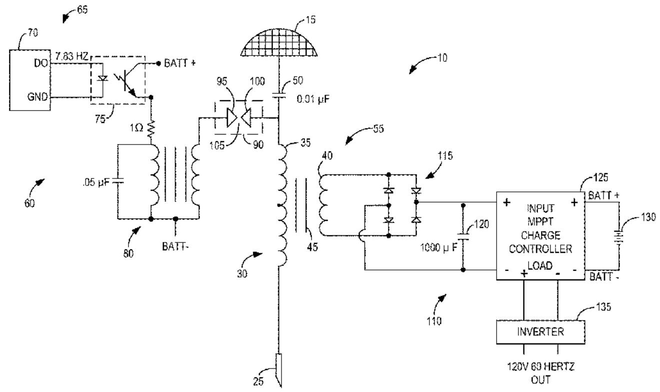

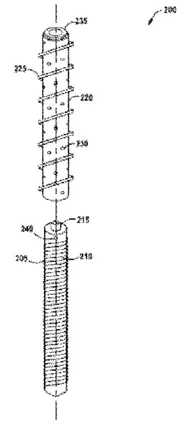

(56) References Cited

U.S. PATENT DOCUMENTS

690,151	A	*	12/1901	Jahr	H01M 10/44 429/10
911,260	A	*	2/1909	Pennock	G05B 11/016 174/3
1,119,732	A	*	12/1914	Tesla	H01Q 1/38 200/48 R
1,540,998	A	*	6/1925	Plauson	H05F 7/00 310/309
2,396,276	A	*	3/1946	Lang	E02D 5/805 411/470
2,580,948	A	*	1/1952	Pancake	E01F 9/0117 248/156
2,660,276	A	*	11/1953	McKee	E02D 5/803 52/159
3,526,069	A	*	9/1970	Deike	E02D 5/805 405/244
3,628,337	A	*	12/1971	Stepanich	E02D 5/54 254/41
3,801,893	A		4/1974	Joyce	
4,189,650	A		2/1980	Aaland	
4,733,102	A		3/1988	Nakayama et al.	
4,885,428	A		12/1989	Roberts	
5,367,245	A	*	11/1994	Mims	H02J 15/00 307/149
6,586,900	B2		7/2003	Rider et al.	
7,053,576	B2		5/2006	Correa et al.	
7,439,712	B2	*	10/2008	McCowen	H02N 1/10 174/2
7,466,032	B2	*	12/2008	Wadle	H02N 11/002 290/1 A
8,004,250	B2		8/2011	Grandics	
2007/0170910	A1		7/2007	Chang et al.	
2008/0191580	A1	*	8/2008	Deyo	H02N 11/002 310/300
2010/0207399	A1		8/2010	Grandics	
2011/0316430	A1		12/2011	Cohen et al.	

* cited by examiner

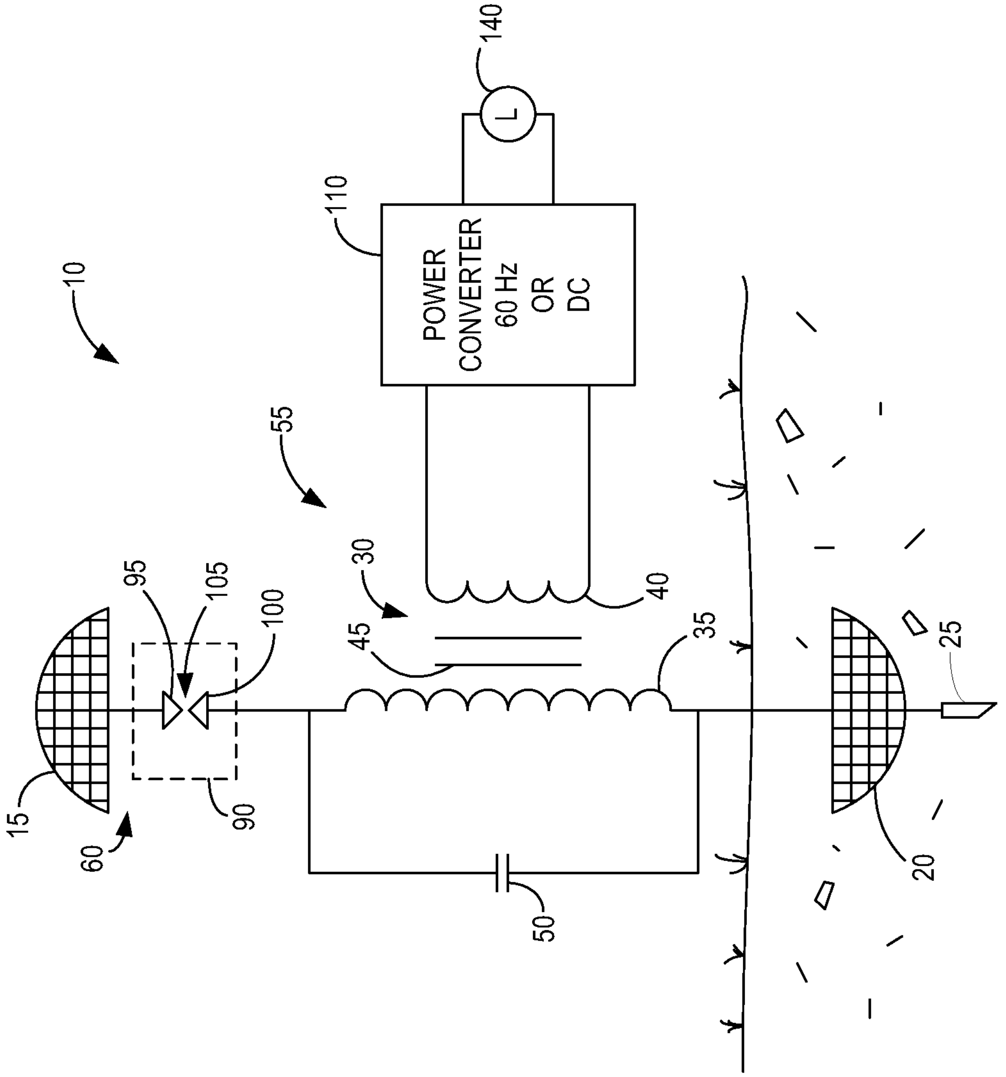
POWER CONVERTER 60 Hz OR DC
L

FIG 1

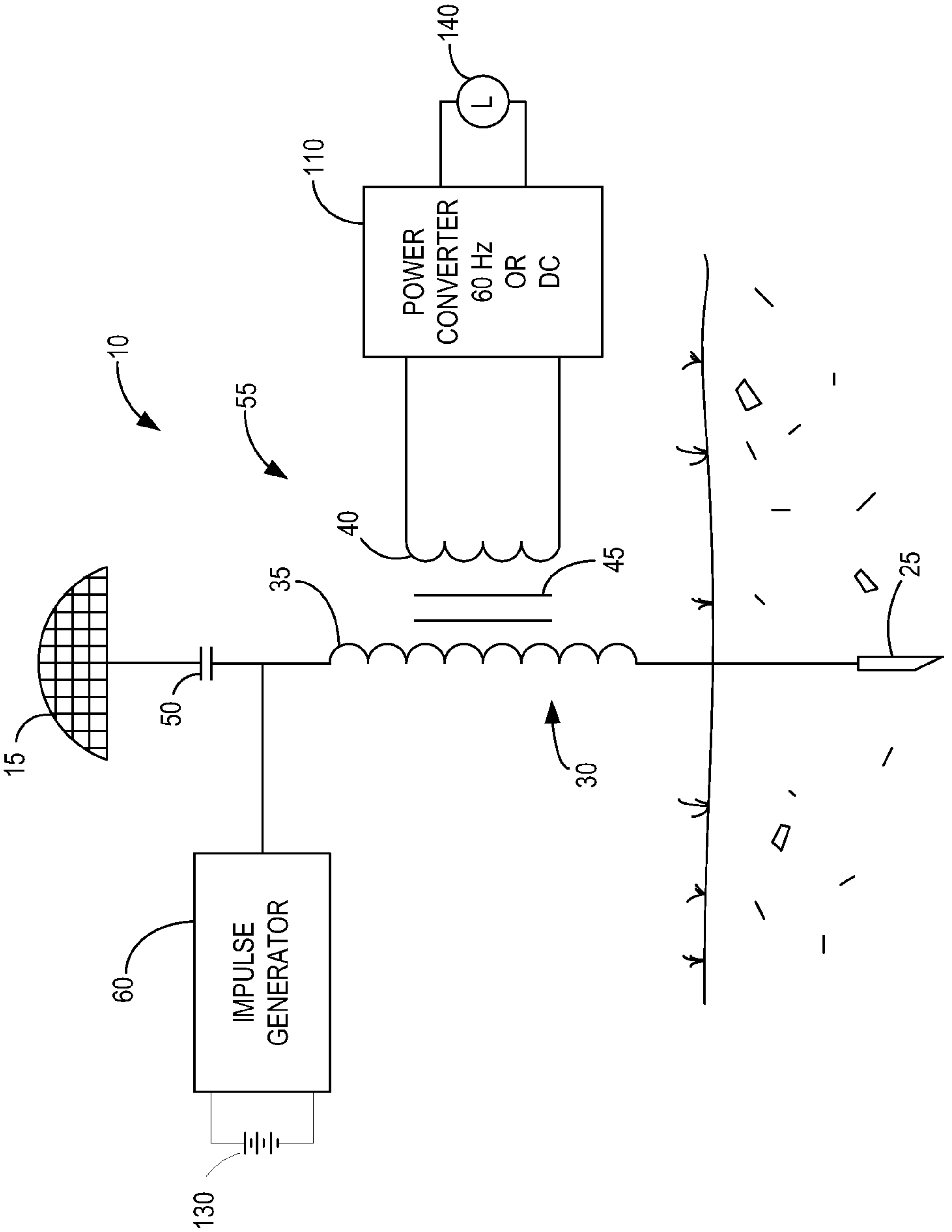
L
POWER CONVERTER 60 Hz OR DC
IMPULSE GENERATOR

FIG 2

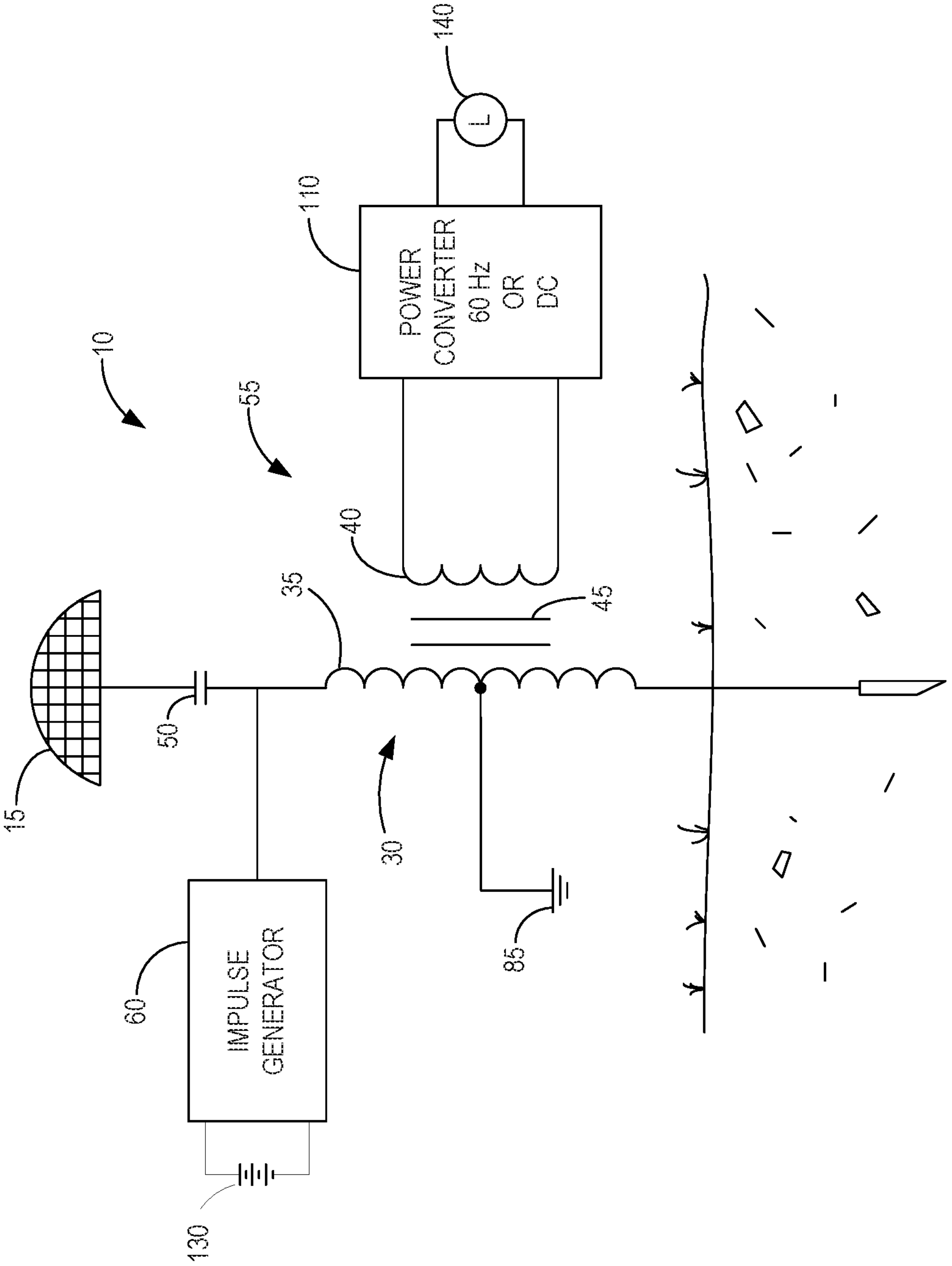
L
POWER CONVERTER 60 Hz OR DC
IMPULSE GENERATOR

FIG 3

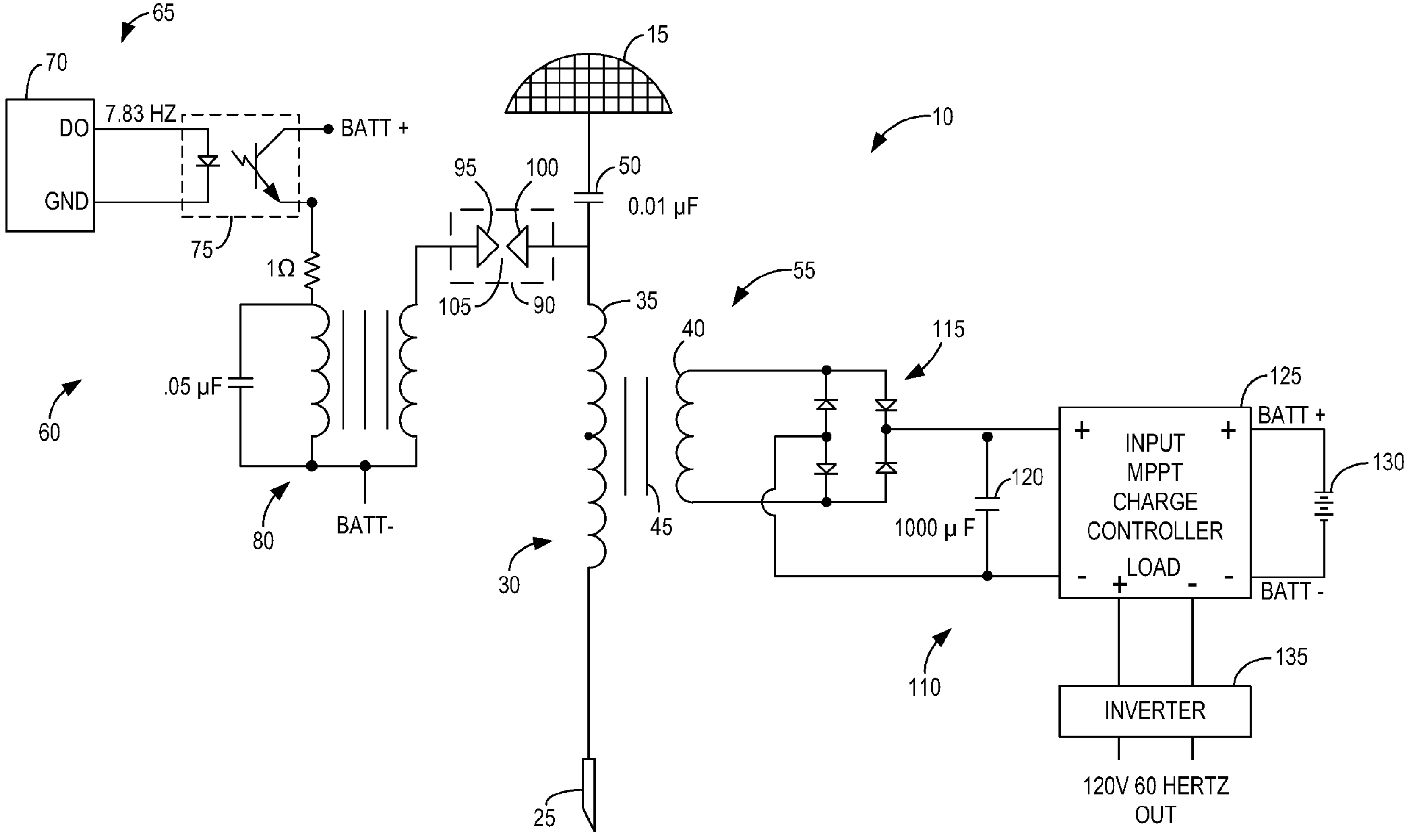
65
70
DO
GND
7.83 HZ
BATT +
75
1Ω
.05 µF
60
80
BATT-
95
100
105
90
15
50
0.01 µF
35
30
25
10
55
40
45
115
120
1000 µ F
110
125
INPUT
MPPT
CHARGE
CONTROLLER
LOAD
+
+
-
+
-
-
BATT +
BATT -
130
135
INVERTER
120V 60 HERTZ
OUT

FIG 4

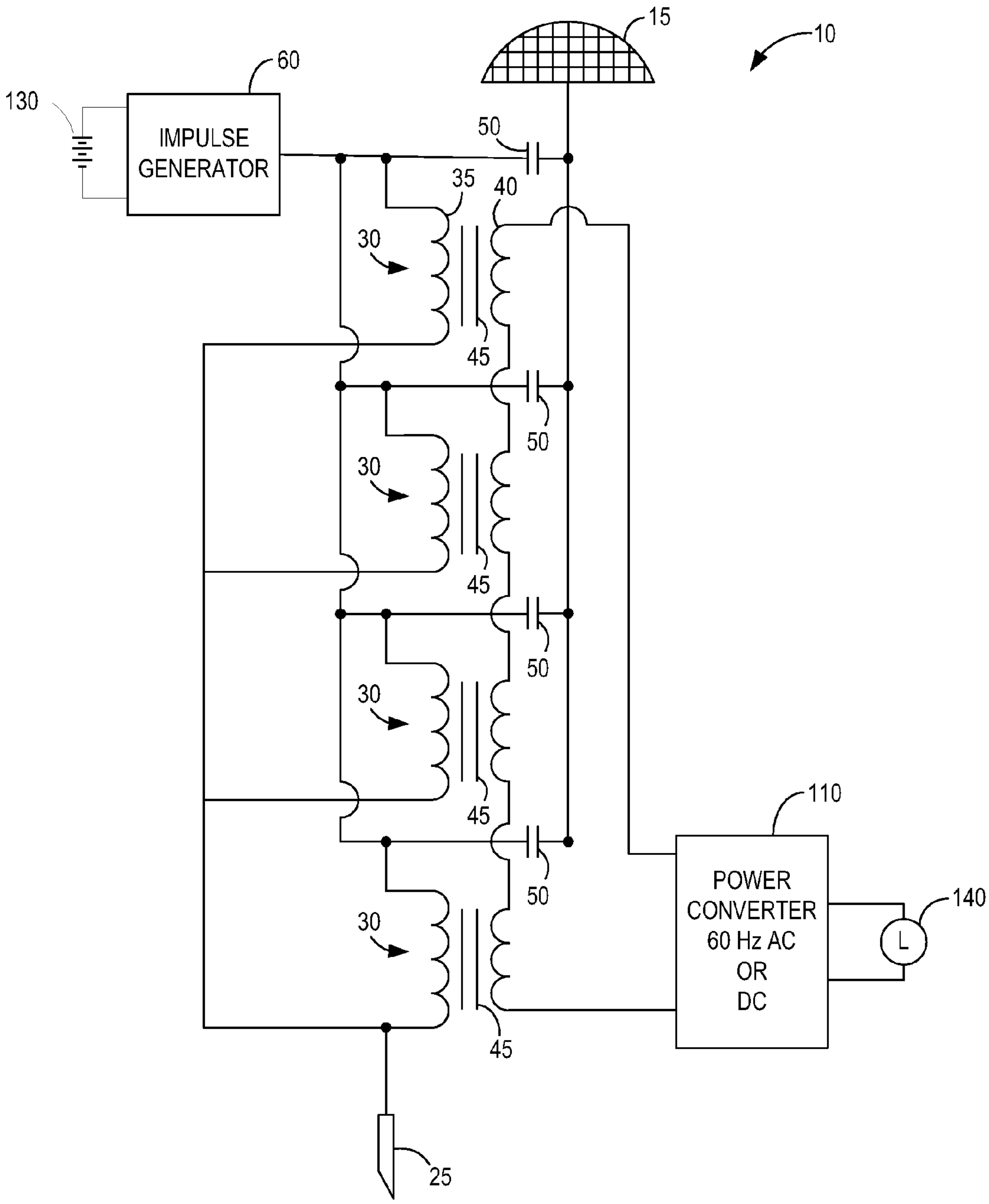
IMPULSE GENERATOR
POWER CONVERTER 60 Hz AC OR DC
L

FIG 5

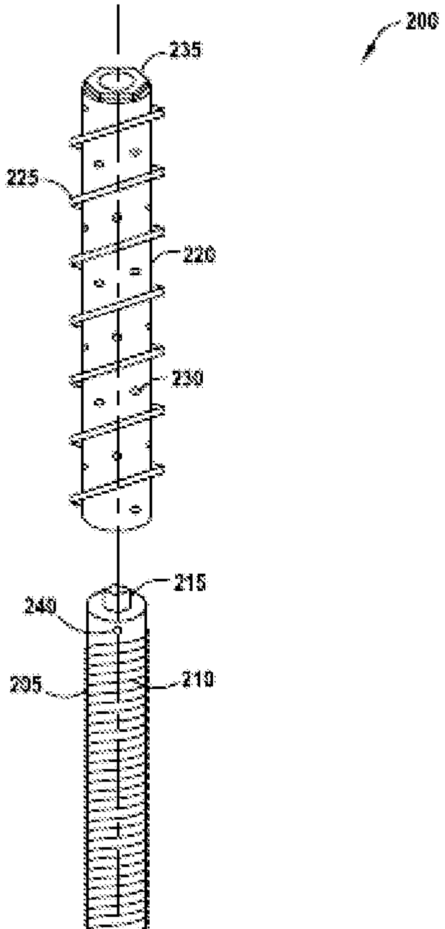
200
235
225
220
230
215
240
205
210

FIG 6A

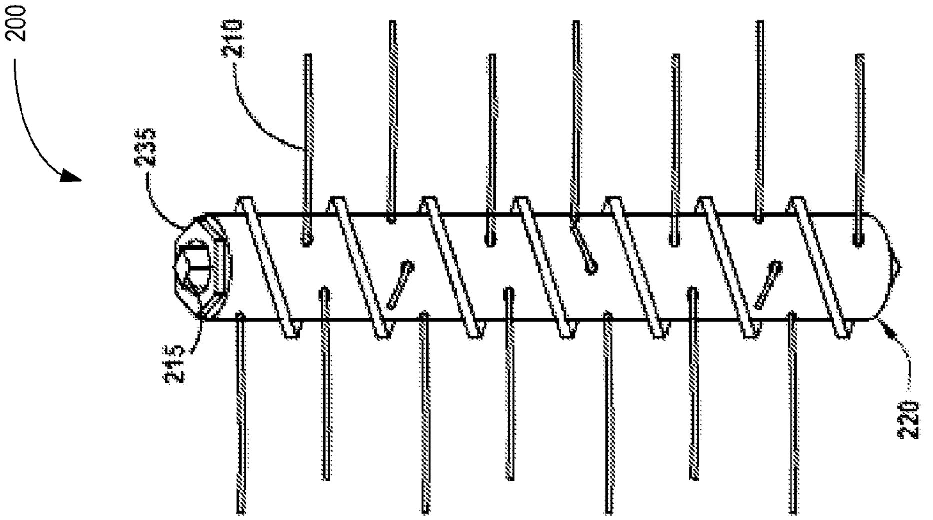
200
210
235
215
220

FIG 6C

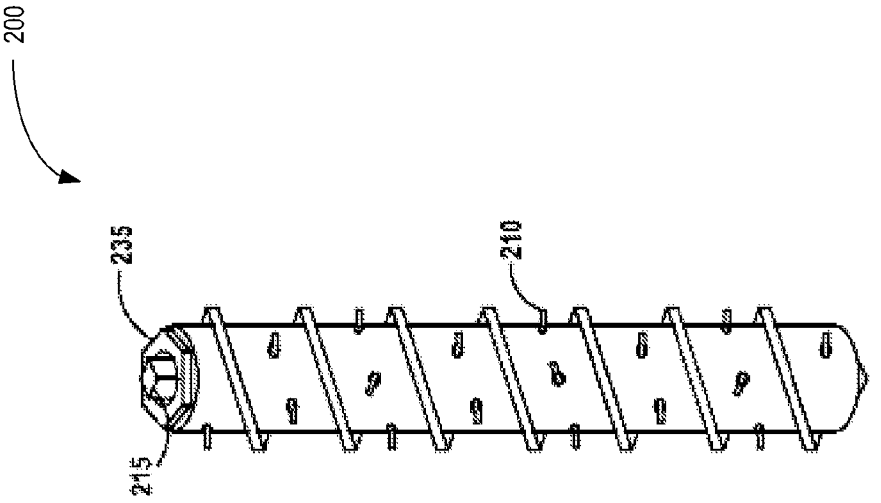
200
235
210
215

FIG 6B

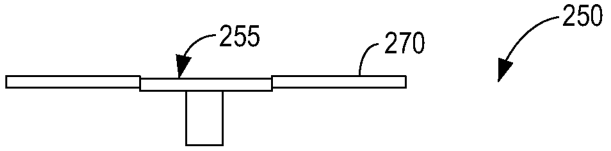

FIG 7A

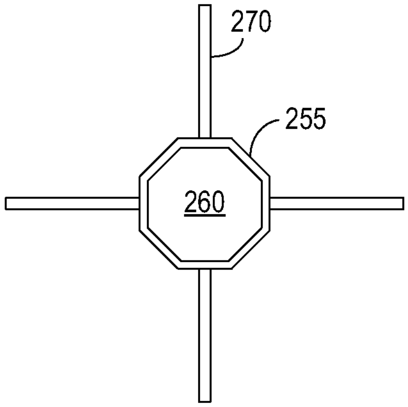

FIG 7B

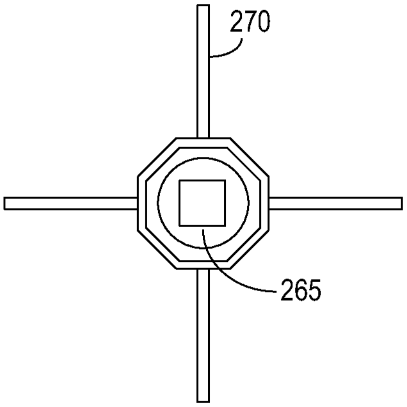

FIG 7C

POWER RECEIVER FOR EXTRACTING POWER FROM ELECTRIC FIELD ENERGY IN THE EARTH

This application claims the benefit of U.S. Provisional Application No. 61/889,894, filed 11 Oct. 2013, the disclosure of which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

The present invention relates generally to renewable energy, and more particularly to methods and apparatus for extracting energy from subsurface electrical fields beneath the earth's surface.

BACKGROUND

The earth and the ionosphere cavity may be viewed as a global electric circuit. Electrical currents are constantly flowing within the earth and its atmosphere. Within the earth, the majority of the earth's energy is carried by extremely low frequency (ELF) and ultralow frequency (ULF) waves in the 0-200 Hz frequency range. The earth's rotating magnetic field and positive lightning are two energy sources that sustain the ELF/ULF waves within the earth and the atmosphere.

A great deal of research has been devoted to studying the electric field present in the earth's ionosphere cavity. Joseph M. Crawley, the "Fair Weather Atmosphere as a Power Source", Proceedings ESA Annual Meeting on Electrostatics 2011; O. Jefimenko, "Operation of Electric Motors from Atmospheric Electric Field," American Journal of Physics, Vol. 39, Pgs. 776-779, 1971; M. L. Breuer, "Usability of Tapping Atmospheric Charge as a Power Source," Renewable Energy, Vol. 28, Pgs. 1121-1127, 2003. Numerous attempts have been made in the past to extract electrical energy from the earth's atmosphere. For example, U.S. Pat. No. 1,540,998 to Plauson describes a system for converting atmospheric electrical energy into usable power. These past attempts have been successful in producing only small amounts of power from the electrical field in the earth's ionosphere cavity. The modest success of these experiments compared to results from other renewable energy sources, such as solar and wind, has tempered further research and prevented widespread use of the electric field in the ionosphere cavity as an energy source.

SUMMARY

The present invention relates to a power receiver for extracting power from electric fields beneath the earth's surface. In embodiments of the present disclosure, a resonant transformer connected to a ground terminal draws current from the earth's electric field through the primary winding of the transformer. Current flow through the resonant transformer is induced by applying a high voltage impulse to the primary winding. The flow of current through the primary winding of the resonant transformer induces a current in the secondary winding, which may be converted and filtered to a usable form, e.g. 60 Hz AC or DC.

In some embodiments of the power receiver, the resonant frequency of the resonant transformers is below 200 Hz.

In some embodiments of the power receiver, the resonant transformer comprises a ferro-resonant transformer.

In some embodiments, the power receiver further comprises an elevated terminal.

In some embodiments of the power receiver, the primary winding of the resonant transformer is connected between the ground terminal and elevated terminal.

In some embodiments of the power receiver, the elevated terminal comprises an upper capacitive plate coupled to the earth's ionosphere cavity.

In some embodiments of the power receiver, the impulse generator comprises the upper capacitive plate and a spark gap connected between the upper capacitive plate and the primary winding of the resonant transformer. The spark gap comprises a pair of electrodes separated by a gap and configured to generate a spark when a voltage difference between the electrodes reaches a predetermined level.

In some embodiments of the power receiver, the impulse generator comprises a pulse generator for generating low voltage pulses, a step-up transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses, and a spark gap connected between the step-up transformer and the primary winding of the resonant transformer to generate a spark responsive to the high voltage impulses from the step-up transformer.

In some embodiments of the power receiver, the impulse generator comprises a pulse generator for generating low voltage pulses, and a step-up transformer connected to the primary winding of the resonant transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses.

In some embodiments of the power receiver, the impulse generator comprises a solid state spark generator.

In some embodiments of the power receiver, the resonant transformer includes a capacitor connected in parallel with the primary winding.

In some embodiments of the power receiver, the resonant transformer includes a capacitor connected in series with the primary winding between the impulse generator and the elevated terminal.

In some embodiments, the power receiver comprises multiple resonant transformers having primary windings connected in parallel between the ground terminal and the elevated terminal.

In some embodiments of the power receiver, the resonant transformers have different resonant frequencies.

In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are all below 200 Hz.

In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

Another embodiment of the power receiver comprises a resonant circuit connected to a ground terminal disposed below the surface of the earth, an impulse generator for generating and applying a high voltage electrical impulse to the resonant circuit to induce current flow from the ground terminal through the resonant circuit, and a power conversion circuit connected to the resonant circuit to convert electrical current flowing through the resonant circuit to a desired form. The resonant circuit has a resonant frequency below 200 Hertz.

In some embodiments of the power receiver, the resonant circuit comprises a resonant transformer having a primary winding, a secondary winding, and resonant capacitor connected in series with the primary winding.

In some embodiments of the power receiver, the resonant circuit comprises multiple resonant transformers having primary windings connected in parallel to the ground terminal.

In some embodiments of the power receiver, the resonant transformers have different resonant frequencies.

In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are all below 200 Hz.

In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

Other embodiments of the disclosure comprise a ground terminal for a power receiver. In one embodiment, the ground terminal comprises a ground shaft configured for insertion beneath the surface of the earth, a hollow cylinder surrounding the ground shaft and having a plurality of openings, and a plurality of ground wires connected at one end to the ground shaft. The ground wires are wound around the ground shaft and have free ends protruding through respective openings in the hollow shaft so that rotation of the ground shaft relative to the hollow cylinder causes the ground wires to extend radially into the earth.

Other embodiments of the disclosure comprise methods of extracting power from the earth. In one embodiment, the method comprises applying a high voltage impulse to resonant circuit coupled to a ground terminal disposed beneath the surface of the earth to initiate resonance in the resonant circuit and induce the flow of current from the ground terminal to the resonant circuit, and converting the current flowing from the ground terminal into the resonant circuit into a useful form.

In some embodiments of the method, the resonant circuit comprises a resonant transformer including a primary winding coupled to the ground terminal and a second winding coupled to a power converter, and applying a high voltage impulse to resonant circuit comprises applying a high voltage impulse to the primary winding of the resonant transformer.

In some embodiments of the method, applying a high voltage impulse to the primary winding of the resonant transformer comprises applying an impulse in the range to 10,000 to 40,000 volts to primary winding of the transformer.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** illustrates a first exemplary embodiment of a power receiver.

FIG. **2** illustrates a second exemplary embodiment of a power receiver.

FIG. **3** illustrates a third exemplary embodiment of a power receiver.

FIG. **4** illustrates a fourth exemplary embodiment of a power receiver.

FIG. **5** illustrates a fifth exemplary embodiment of a power receiver.

FIG. **6**A is an exploded perspective view of an exemplary ground antenna array for the power receiver.

FIG. **6**B is a perspective view of an assembled ground antenna array before being deployed.

FIG. **6**C is a perspective view of an assembled ground antenna array after being deployed.

FIG. **7**A is a side view of an insertion tool for installing the ground antenna array.

FIG. **7**B is a top view of the insertion tool for installing the ground antenna array.

FIG. **7**C is a bottom view of the insertion tool for installing the ground antenna array.

DETAILED DESCRIPTION

Referring now to the drawings, a power receiver for extracting energy from the earth's electric field are illustrated and indicated generally by the numeral **10**. Various embodiments of the power receiver **10** are described and similar reference numbers are used throughout the description to indicate similar components.

The power receiver **10** converts energy in the ELF/ULF waves to useful form, e.g. 60 Hz AC or DC. The power receiver **10** is essentially a resonance circuit that resonates at the natural resonance frequencies in the earth's electric field. These resonance frequencies, known as Schumann resonance frequencies, occur at 7.83 Hz, 14.3 Hz, 20.8 Hz, 27.3 Hz, and 33.8 Hz. A high voltage impulse initiates resonance within the power receiver **10**. In the resonant mode, the impedance of the power receiver **10** is reduced to near zero thus inducing ground currents to flow into the power receiver **10** where the ground currents are converted to useful form.

FIG. **1** illustrates a first embodiment of the power receiver **10**. The power receiver **10** comprises a resonant transformer **30** connected between an elevated terminal **15** and ground terminal **25**. In this embodiment, the elevated terminal **15** is capacitively coupled to electric fields within the earth's ionosphere cavity and functions as an upper capacitive plate. A lower capacitive plate **20** is connected to the ground terminal **25** beneath the surface of the earth.

The resonant transformer **30** comprises a primary winding **35**, secondary winding **40**, ferromagnetic core **45**, and capacitor **50**. One end of the primary winding **35** is connected to the lower capacitive plate **20** and ground terminal **25**. The opposite end of the primary winding **35** is connected via a spark gap **90** to the elevated terminal **15**. The capacitor **50** is connected in parallel with the primary winding **35** of the resonant transformer **30** to form an LC circuit **55** with a resonance frequency range of between about 0.1 and 200 Hz. In a preferred embodiment, the resonant transformer has a Q of about 10 or greater and resonance frequency in the range of about 0.1-200 Hertz. For example, the resonant transformer **30** may have a resonance frequency of about 7.83 Hz, the fundamental Schumann resonance frequency. The secondary winding **40** of the resonant transformer **30** is connected to a power converter **110** as will be hereinafter described in greater detail. The power converter **110** converts the energy extracted from the earth's electric field by the power receiver **10** into a usable form for driving a load **140**.

The elevated terminal/upper capacitive plate **15** comprises an insulated, dish-shaped plate with a large radius of curvature. The capacitance and resistance of the elevated terminal is chosen for receiving broadband electric field frequencies in the 0-200 Hz range. The upper capacitive plate **15** is sized to maximize to the extent practical coupling with the electric field in the earth's ionosphere cavity.

The lower capacitive plate **20** is also a dish-shaped plate with a large radius of curvature. One function of the lower capacitive plate **20** is to collect charge from the earth's ground currents and provide an instantaneous source of current as hereinafter described. The capacitance and resistance of the lower capacitive plate **20** is selected to promote the flow of current from the ground with minimal losses.

The spark gap **90** connected between the elevated terminal **15** and resonant transformer **30** comprises a pair of electrodes **95**, **100** separated by an evacuated air gap **105**. Electrode **95** is connected to the upper capacitive plate **15**. Electrode **100** is connected to the resonant transformer **30**. The spark gap **105** prevents electrical discharge from the upper capacitive plate **15** to the earth's ionosphere cavity. The spark gap **90** in combination with the elevated terminal **15** function as an impulse generator that applies a high voltage impulse of about 10,000-40,000 volts to the primary winding **35** to initiate resonance in the transformer **30**.

In operation, the capacitive coupling of the upper capacitive plate **15** induces a high voltage operating current in the upper capacitive plate **15**. The upper capacitive plate is connected to a first electrode **95** to the spark gap **90**. When the voltage difference between the electrodes **95** and **100** reaches a threshold, a spark forms across the electrodes **95**,**100** and a high voltage impulse is applied to the primary winding **35** of the resonant transformer **30**. This high voltage impulse initiates resonance within the transformer **30**.

In resonant mode, the impedance of the resonance transformer is reduced to nearly zero allowing current to flow from the capacitive plate **20** and ground terminal **25** through the primary winding **35** of the transformer **30**, which in turn induces current in the secondary winding **40**. Power converter **110** converts the current flowing through the secondary winding **40** into a usable form for driving a load **140**. The transformer **30** will continue to resonate for a short period of time. By providing high voltage impulses to the primary winding **35** of the resonant transformer **30** at periodic intervals, it is possible to maintain a continuous flow of current from the earth into the resonant transformer **30**, thus producing a continuous supply of power.

FIG. **2** discloses a second embodiment of the primary receiver **10**. This embodiment includes a resonant transformer **30** connected between an elevated terminal **15** and ground terminal **25**. The resonant transformer **30** comprises a primary winding **35**, secondary winding **40**, ferromagnetic core **45** and a high voltage capacitor **50**. One end of the primary winding **35** is connected to the ground terminal **25**. The opposite end of the primary winding **35** is connected to the elevated terminal **15**. The capacitor **50** has a capacitance of about 0.01 micro-farads. In contrast to the previous embodiment, capacitor **50** is connected in series with the primary winding **35** and elevated terminal **15** and forms a LC circuit **55** with a Q of about 10 or greater and a resonance frequency in the range of about 0.1 to 200 Hz. In a preferred embodiment, the resonance frequency of the transformer **30** is 7.83 Hz, the fundamental Schumann resonance frequency. An impulse generator **60** is connected between the primary winding **35** of the resonant transformer **30** and the series capacitor **50** and applies a high voltage impulse to the primary winding **35** of the resonant transformer **30**. A battery **130** or other external power source supplies power to the impulse generator **60**. As previously described, the high voltage impulse applied by the impulse generator **60** initiates resonance within the resonant transformer **30** inducing current flow from the ground terminal **25** into the primary winding **35** of the resonant transformer **30**. The flow of current from the ground terminal **25** through the primary winding **35** induces current in the secondary winding **40**. Power converter **110** converts the electrical energy in the current flowing through the primary winding **40** into a usable form.

In contrast to the first embodiment, it is not required to capacitively couple the elevated terminal **15** in the second embodiment to the earth's ionosphere cavity. Rather, the elevated terminal **15** in this embodiment provides lightning protection and dissipates some of the energy flowing into the power receiver **10** to the earth's ionosphere cavity. Also, in contrast to the first embodiment, the capacitor **50** is connected in series between the primary winding **35** of the transformer **30** and the elevated terminal **15**. Those skilled in the art will appreciate that the capacitor **50** could also be connected in parallel rather than series with the primary winding **35** as shown in FIG. **1**. Another difference is that the impulse generator **60** has an external power source. The amount of energy generated by the power receiver **10**, however, is far greater than the energy needed to generate high voltage impulses. The first embodiment does not require an external power source to generate high voltage impulses.

FIG. **3** illustrates a third embodiment of the power receiver **10**. This embodiment is essentially the same as the embodiment shown in FIG. **2**. The main difference is that a center tap of the primary winding **35** in the resonant transformer **30** is connected to an electrical ground **85**. It should be appreciated that the electrical ground **85** may be different than the earth ground. When the center tap of the resonant transformer **30** is grounded at a distance away from the ground terminal **25** (e.g. 50 ft to 100 ft), the power receiver **10** becomes a transmitter via the ground loop formed.

FIG. **4** illustrates the power receiver **10** of FIG. **2** in greater detail. The power receiver includes a resonant transformer **30** connected between a ground terminal **25** and an elevated terminal **15**. The resonant transformer **30** includes a primary winding **35**, secondary winding **40**, ferromagnetic core **45** and series capacitor **50** configured as previously described. The resonant transformer **30** may have a Q of about 10 and a resonance frequency in the range of about 0.1 to 200 Hz. The resonant transformer **30** may be made using an Allanson transformer (part #1530BP120R) connected in series with a 0.01 micro-farad capacitor, such as the Condensor Products high voltage capacitor (part #TC 103-17-125). An impulse generator **60** is connected between the primary winding **35** of the resonant transformer **30** and the series capacitor **50** and applies a high voltage impulse in the range of about 10,000 to 40,000 volts to the primary winding of the transformer **30**. A battery **130** or other external power source supplies power to the impulse generator **60**. The power converter **110** connects to the secondary winding **40** of the resonant transformer **30** for converting current in the secondary winding of the transformer to a useful form.

The impulse generator **60** comprises a pulse generator **65** for generating low voltage pulses, a step-up transformer **80** for converting the low voltage pulses from the pulse generator **65** to high voltage pulses, and a spark gap **90** for generating sparks responsive to the high voltage pulses from the step-up transformer **80**.

The pulse generator **65** comprises a square wave generator **70** and solid state relay **75**. The square wave generator **70** generates a digital pulse stream. In one embodiment, the digital pulse stream generates a square waveform with a frequency of about 7.83 Hz. The frequency of the digital pulse stream is selected to match the resonance frequency of the transformer **30**, though such is not necessarily required. The pulse stream output from the square wave generator **70** is applied to the solid state relay **75**. The solid state relay **75** is connected between a battery or other power source and a first winding of the step-up transformer **80**. The solid state relay **75** functions as a switch that is activated responsive to the waveform from the square wave generator **70** to provide a continuous stream of low voltage pulses from the battery to the first winding of the step-up transformer **80**. A 1 ohm resistor is connected between the solid state relay **75** and step-up transformer **80**.

The step-up transformer **80** converts the low voltage pulses from the pulse generator **65** to high voltage pulses that are applied to the spark gap **90**. The step-up transformer has a 0.5 micro-farad capacitor connected in parallel with the primary winding of the step-up transformer **80**. The step-up transformer produces pulses at the output of about 30,000 to 40,000 volts.

The spark gap **90** comprises a pair of electrodes **95**, **100** separated by an air gap **105**. As previously described, when the voltage potential between the electrodes **95**, **100** reaches a threshold, a spark forms between the electrodes **95**, **100** and supplies a nearly instantaneous, high voltage impulse to the primary winding **35** of the resonant transformer **30**. This high voltage impulse initiates resonance in the resonant transformer **30** inducing current flow from the ground terminal **25** through the primary winding **35** of the resonant transformer **30**.

The power converter **110** comprises a bridge rectifier **115**, filter capacitor **120**, charge controller **125**, and inverter **135**. The bridge rectifier **115** converts the AC current flowing through the secondary winding **40** of the resonant transformer to a DC current. a Filter capacitor **120** removes unwanted frequencies from the DC current. The filter capacitor **120** has a capacitance of about 1000 micro-farads. The DC current is input to the charge controller **125**. The charge controller **125** may, for example, comprise a maximum power point tracking (MPPT) charge controller, which is commonly used in solar power generating systems. The charge controller **125** applies a small amount of energy to a battery **30** to charge the battery **130**. As previously noted, the battery **130** serves as a power source for the impulse generator **60**. The remaining current is supplied to an inverter **135**, which converts the DC current to an AC current with a desired voltage and frequency, e.g., 120 volts/60 Hz AC. The power converter **110** as shown in FIG. **4** may be utilized in the embodiment shown in FIGS. **1**, **2** and **3**.

FIG. **5** illustrates a power converter **10** according to another embodiment. The power converter **10** comprises a plurality of resonant transformers **30** connected between a ground terminal **25** and elevated terminal **15**. Each of the resonant transformers **30** comprises a primary winding **35**, secondary winding **40**, ferromagnetic core **45** and series capacitor **50**. The primary windings **35** of the resonant transformers **30** are connected in parallel. The secondary windings **40** are connected in series with the power converter **110**. An impulse generator **60** applies a high voltage impulse to the primary windings **35** of the resonant transformers **30**. A battery **130** or other external power source supplies power to the impulse generator **60**. The power converter **110** converts the current in the power converter circuit to a usable form for driving a load **140**.

In one embodiment, each of the resonant transformers **30** shown in FIG. **5** is configured to have a different resonant frequency. In one embodiment, the resonant transformers **30** are configured to resonate at frequencies of 7.83 Hz, 14.8 Hz, 20.3 Hz and 26.8 Hz respectively. Additional resonant transformers **30** could be added to operate at other resonance frequencies.

FIGS. **6A-6C** illustrate a high quality ground antenna array **200** which may be used as a ground terminal **25**. The ground antenna array **200** comprises a generally cylindrical ground shaft **205** disposed with a hollow cylinder **220** and a plurality of reinforced, heavy gauge ground wires **210** attached at one end to the ground shaft **205**. The ground shaft **205** and ground wires **210** should be highly conductive and have low resistance to supply current from the ground to the power receiver **10**. In one embodiment, the ground wires **210** may be copper or other highly-conductive metal. The end of the ground shaft may be pointed to facilitate insertion into the earth. A connection port on the ground shaft **220** is provided to electrically connect the ground antenna array **220** to the resonant transformer **30**.

The hollow cylinder **220** has external threads **25** to facilitate insertion into the ground. A rotator nut **235** is fixedly secured to the top end of the hollow shaft **220**. A square shaft **215** protrudes from the top end of the ground shaft **205** into the opening in the rotator nut **235**. FIG. **6**B. A tool **250**, shown in FIG. **7**, engages with the rotator nut **235** and square shaft **215** during insertion of the ground antennas array **200** into the ground as will be hereinafter described.

The insertion tool **250** is shown in FIG. **7**. The insertion tool **250** includes a tool body **255** having a first socket **260** on one side to fit the rotator nut **235** on the hollow cylinder **220** and a second socket **265** on the other side to fit the square shaft **215** on the ground shaft **205**. Arms **270** extend from the outer periphery of the tool body **255** for manually or mechanically turning the insertion tool **250**.

Before the antenna array **200** is deployed, the ground wires **210** are wound around the ground shaft **205** with the free ends protruding slightly from respective openings **230** in the hollow cylinder **220** to a distance not to exceed one half (½) the depth of the external threads **225** on the hollow cylinder **220**. FIG. **6**B illustrates the ground antenna array **200** before deployment. FIG. **6**C illustrates the ground antenna array in a deployed configuration.

Installation of the ground antenna array **200** is performed in two stages. In the first stage, a hole slightly smaller in diameter than the threads **235** of the hollow cylinder **220** is drilled into the Earth to a depth matching the length of the hollow cylinder **220** or slightly longer. The hole is filled with water and the water is allowed to soak into the soil. After the ground is softened, the hollow cylinder **220** is rotated using the insertion tool **250** to insert the ground antenna array **200** into the ground. The first socket **260** of the insertion tool **250** is engaged with the rotator nut **230** and the insertion tool **250** is turned by hand or a mechanized rotating shaft fitted and attached to the tool arms **270** to thread the ground assembly into the hole. During the initial insertion of the ground antenna array **200**, the ground shaft **205** is fixed to the hollow shaft **220** and rotates with the hollow shaft. The hollow cylinder **220** is rotated until it reaches the full depth of the hole.

Once the ground antenna array **200** has been fully inserted into the earth, the insertion tool **250** is flipped over and the second socket **265** of the insertion tool **250** is engaged with the square shaft **215**. The insertion tool **250** is turned by hand or a mechanized rotating shaft fitted and attached to the tool arms **270** to rotate the ground shaft **205**. During the second phase, the ground shaft **205** rotates freely inside the hollow cylinder **220**. Rotation of the ground shaft **205** causes the reinforced ground wires **210** to extend radially into the earth. The ground shaft **220** is rotated until the ground wires are fully extended. The ends of the ground wires may be sharpened to aid in the extension of the ground wires during the second phase.

After the ground antenna array **200** is deployed, a connection cable **280** is attached to a connection port **240** on the ground shaft **220** to electrically connect the ground antenna array **220** to the resonant transformer **30** in the power receiver **10**.

What is claimed is:

1. A power receiver for extracting electrical energy from the earth's electric field, said power receiver comprising:

a resonant transformer connected to a ground terminal disposed below the surface of the earth;

an impulse generator for generating and applying a high voltage electrical impulse to a primary winding of the resonant transformer to induce current flow from the ground terminal through the primary winding of the transformer, said impulse generator comprising:

a signal generator for generating a low voltage signal;

a step-up transformer for converting the low voltage signal to a high voltage signal;

a spark gap connected between the step-up transformer and the primary winding of the resonant transformer to generate a spark responsive to the high voltage signal from the step-up transformer; and

a power conversion circuit connected to a secondary winding of the resonant transformer to convert electrical current flowing through the secondary winding to a desired form.

2. The power receiver of claim **1** wherein a resonant frequency of the resonant transformers is below 200 Hz.

3. The power receiver of claim **1** wherein the resonant transformer comprises a ferro-resonant transformer.

4. The power receiver of claim **1** further comprising an elevated terminal, and wherein the primary winding of the resonant transformer is connected between the ground terminal and elevated terminal.

5. The power receiver of claim **4** wherein the elevated terminal comprises an upper capacitive plate coupled to the earth's ionosphere cavity.

6. The power receiver of claim **1** wherein said resonant transformer includes a capacitor connected in parallel with the primary winding.

7. The power receiver of claim **1** wherein said resonant transformer includes a capacitor connected in series with the primary winding between the impulse generator and the elevated terminal.

8. The power receiver of claim **1** further comprising multiple resonant transformers having primary windings connected in parallel between the ground terminal and the elevated terminal.

9. The power receiver of claim **8** wherein the resonant transformers have different resonant frequencies.

10. The power receiver of claim **9** wherein the resonant frequencies of the resonant transformers are all below 200 Hz.

11. The power receiver of claim **8** wherein the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

12. A ground terminal for a power receiver, said ground terminal comprising:

a ground shaft configured for insertion beneath the surface of the earth;

a hollow cylinder surrounding the ground shaft and having a plurality of openings; and

a plurality of ground wires connected at one end to the ground shaft said ground wires being wound around the ground shaft and having free ends protruding through respective openings in the hollow shaft so that rotation of the ground shaft relative to the hollow cylinder causes the ground wires to extend radially into the earth.

* * * * *

US009680327B2

(12) **United States Patent**
Scheb

(10) **Patent No.: US 9,680,327 B2**
(45) **Date of Patent: Jun. 13, 2017**

(54) **RF ENERGY HARVESTING BY A NETWORK NODE**

(71) Applicant: **Landis+Gyr Innovations, Inc.**, Alpharetta, GA (US)

(72) Inventor: **Jeffrey E. Scheb**, Roswell, GA (US)

(73) Assignee: **Landis+Gyr Innovations, Inc.**, Alpharetta, GA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 313 days.

(21) Appl. No.: **14/318,970**

(22) Filed: **Jun. 30, 2014**

(65) **Prior Publication Data**

US 2015/0380973 A1 Dec. 31, 2015

(51) **Int. Cl.**

H01F 27/42	(2006.01)
H01F 37/00	(2006.01)
H01F 38/00	(2006.01)
H02J 7/02	(2016.01)
H02J 5/00	(2016.01)

(Continued)

(52) **U.S. Cl.**

CPC ***H02J 7/025*** (2013.01); ***H02J 5/00*** (2013.01); ***H02J 17/00*** (2013.01); ***H02J 50/20*** (2016.02); ***H04B 1/1607*** (2013.01); ***H04W 52/0251*** (2013.01); ***H04W 52/0296*** (2013.01); *H04W 52/0216* (2013.01); *H04W 52/0248* (2013.01)

(58) **Field of Classification Search**

CPC .. H02J 5/005; H02J 7/025; H02J 17/00; H02J 50/12; H02J 50/10; H02J 50/90; H02J 50/80; H02J 7/00; H02J 7/0042; H02J 2007/0096; H02J 50/20; H02J 50/60; H02J 7/04; H02J 50/05; H02J 50/27

USPC 307/104

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,490,257 B1 *	12/2002	Watanabe	H04B 7/0608 370/318
2006/0104291 A1 *	5/2006	Rodriguez	H04B 3/54 370/401

(Continued)

FOREIGN PATENT DOCUMENTS

WO	2007079490	7/2007
WO	2016003552	1/2016

OTHER PUBLICATIONS

Ostaffe, "RF-based Wireless Charging and Energy Harvesting Enables New Applications and Improves Product Design", mouser.com, Applications & Technologies, at least as early as Mar. 28, 2014, 2 pages.

(Continued)

Primary Examiner — Jared Fureman
Assistant Examiner — Esayas Yeshaw
(74) *Attorney, Agent, or Firm* — Kilpatrick Townsend & Stockton LLP

(57) **ABSTRACT**

Disclosed are various embodiments for energy harvesting by circuitry of a node in a network. The node operates in a communication mode that includes communicating, via an RF transceiver, data with another node in the network. The node includes an energy storage device that provides power for the node. The node switches, via an RF switch, from said communication mode to an energy harvest mode. The energy harvest mode includes receiving RF emissions accessible to the node via an antenna. The RF emissions are converted to DC electrical energy, and the DC energy is stored in the energy storage device.

19 Claims, 5 Drawing Sheets

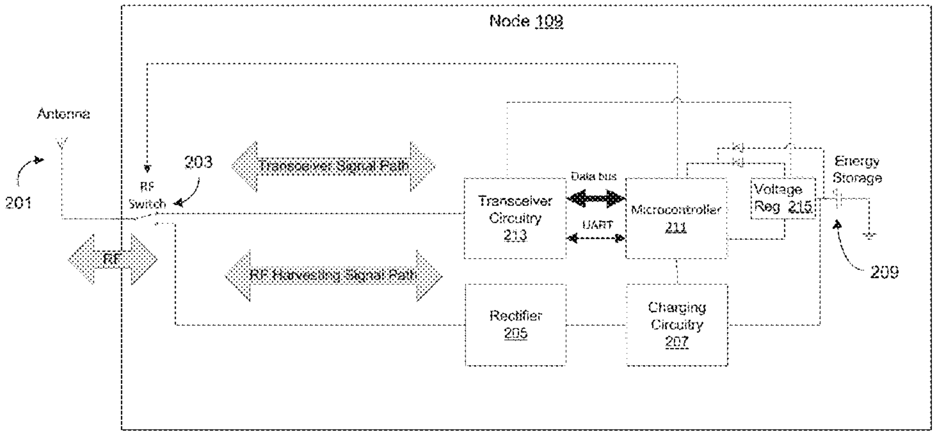

(51) **Int. Cl.**

H02J 17/00	(2006.01)
H04B 1/16	(2006.01)
H02J 50/20	(2016.01)
H04W 52/02	(2009.01)

(56) **References Cited**

U.S. PATENT DOCUMENTS

2009/0174361	A1 *	7/2009	Duron	H02J 17/00 320/101
2010/0134257	A1 *	6/2010	Puleston	G06K 7/0008 340/10.4
2010/0142509	A1	6/2010	Zhu et al.	
2012/0161721	A1 *	6/2012	Neethimanickam	H02J 17/00 320/167
2013/0137455	A1 *	5/2013	Xia	H02J 17/00 455/456.1
2013/0234536	A1 *	9/2013	Chemishkian	H02J 17/00 307/149
2013/0257360	A1 *	10/2013	Singh	H01F 38/14 320/108
2014/0004912	A1	1/2014	Rajakarunanayake	
2014/0111032	A1	4/2014	Shearer et al.	
2014/0159647	A1 *	6/2014	Nishioka	H02J 7/025 320/107
2014/0370830	A1 *	12/2014	Steer	H04W 52/18 455/127.1
2015/0369870	A1 *	12/2015	Okada	H01M 10/44 324/430

OTHER PUBLICATIONS

Publitek Marketing Communication, "Tune in, Charge up: RF Energy Harvesting Shows Its Potential", Digi-Key Electronics, May 29, 2013, 5 pages.

PCT/US2015/032367, "International Search Report and Written Opinion", Nov. 2, 2015, 16 pages.

International Application No. PCT/US2015/032367, Invitation to Pay Additional Fees and Partial Search Report mailed on Aug. 17, 2015, 6 pages.

* cited by examiner

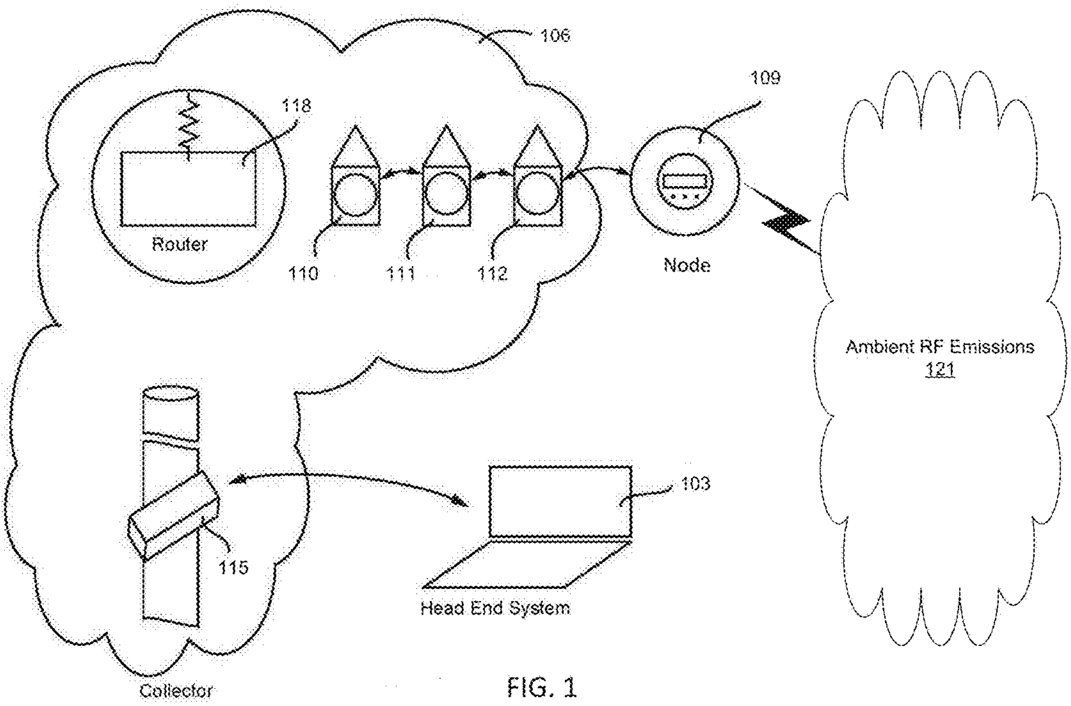
Router
Node
Ambient RF Emissions
Head End System
Collector

FIG. 1

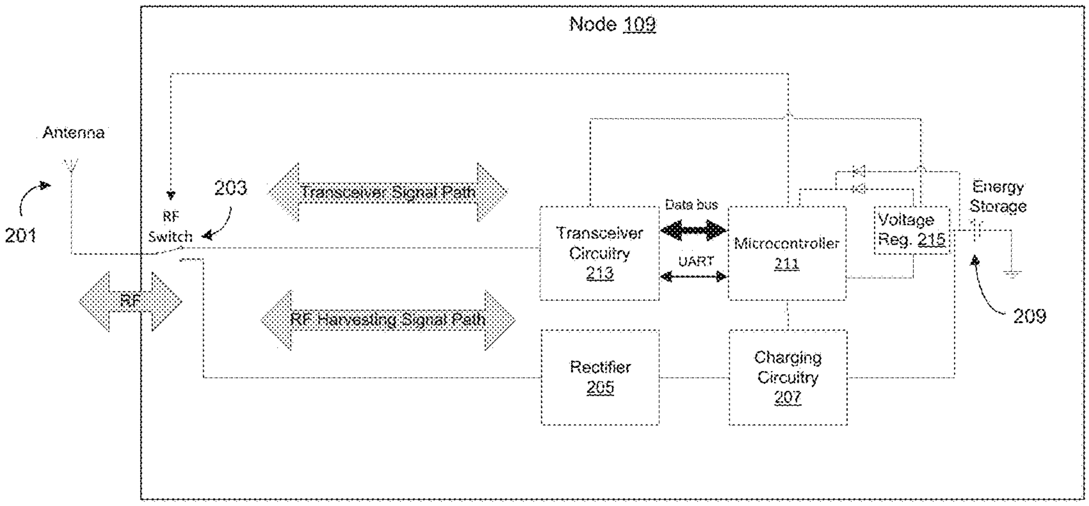
Node 109
Antenna
201
203
RF Switch
Transceiver Signal Path
RF
RF Harvesting Signal Path
Data bus
UART
Transceiver Circuitry 213
Microcontroller 211
Voltage Reg. 215
Energy Storage
209
Rectifier 205
Charging Circuitry 207

FIG. 2

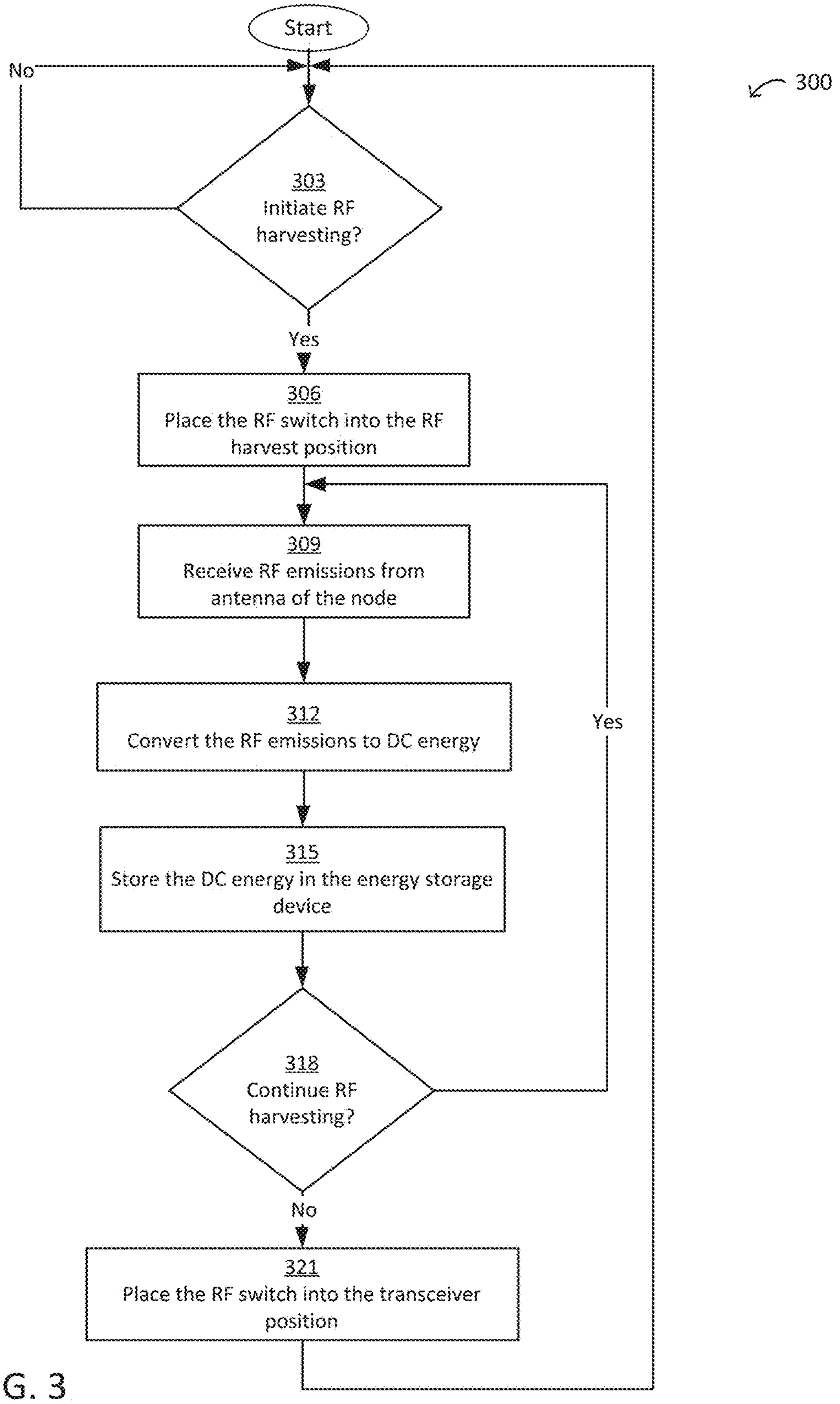
300
Start
No
303
Initiate RF harvesting?
Yes
306
Place the RF switch into the RF harvest position
309
Receive RF emissions from antenna of the node
312
Convert the RF emissions to DC energy
315
Store the DC energy in the energy storage device
Yes
318
Continue RF harvesting?
No
321
Place the RF switch into the transceiver position

FIG. 3

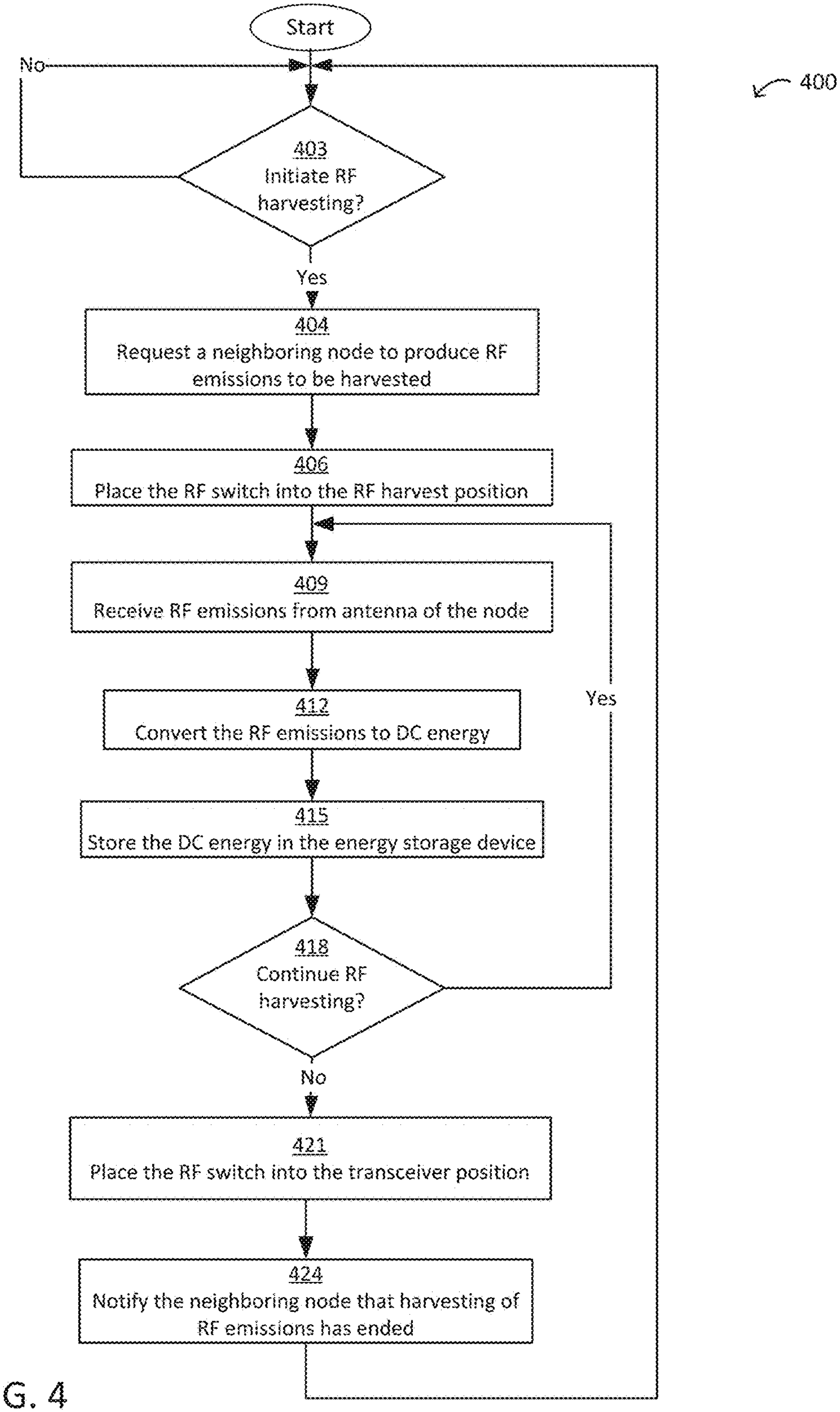
400
Start
No
403
Initiate RF harvesting?
Yes
404
Request a neighboring node to produce RF emissions to be harvested
406
Place the RF switch into the RF harvest position
409
Receive RF emissions from antenna of the node
Yes
412
Convert the RF emissions to DC energy
415
Store the DC energy in the energy storage device
418
Continue RF harvesting?
No
421
Place the RF switch into the transceiver position
424
Notify the neighboring node that harvesting of RF emissions has ended

FIG. 4

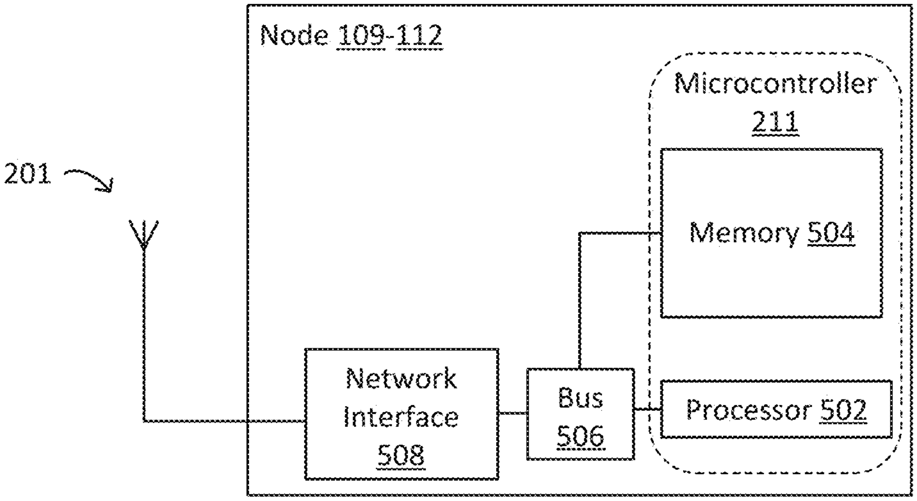
Node 109-112
Microcontroller 211
Memory 504
Processor 502
Bus 506
Network Interface 508
201

FIG. 5

RF ENERGY HARVESTING BY A NETWORK NODE

BACKGROUND

A communications module for a gas or water meter (or any other battery-powered utility device) must be able to support a battery life of 15-20 years. This is primarily achieved through the use of efficient energy-storage batteries as well as by minimizing the power consumption of the communications module (radio). In order to minimize power consumption, the radio may spend most of its time in a low-power "sleep" mode, waking up on some periodic basis to communicate any data with its network and then returning to sleep.

Replacing exhausted batteries in numerous such devices deployed in operation is a time-consuming and difficult task. Increasing the usable life-span of the batteries in these metering devices can produce significant savings in time and money for utility operators.

SUMMARY

Various aspects of the present invention relate to capturing ambient radio frequency (RF) emissions in order to supplement the energy stored by a node, such as a utility meter that may be deployed as part of an advanced metering infrastructure (AMI) system. In one implementation, the node operates in a communication mode that includes communicating, via an RF transceiver, data with another node in the network. The node includes an energy storage device, such as a battery, that provides power for the node. In various embodiments, the node may not be connected to an external source of power.

The node uses an RF switch in order to switch from the communication mode to an energy harvest mode. The mode switch may occur as a result of the node also entering a low-energy sleep mode and/or having no additional data to be communicated via the network. The energy harvest mode includes receiving RF emissions accessible to the node via an antenna and does not attempt to decode the RF emissions. The node converts the RF emissions to direct current (DC) energy and stores the DC energy in the energy storage device. The RF emissions may be converted to DC energy using a rectifier. The node may switch back to the communication mode based in part upon having data to be transmitted, data scheduled to be received, and/or other possible circumstances.

BRIEF DESCRIPTION OF THE DRAWINGS

Many aspects of the present disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, with emphasis instead being placed upon clearly illustrating the principles of the disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

FIG. **1** is a drawing of a networked environment according to various embodiments of the present disclosure.

FIG. **2** is a block diagram of an illustrative example of a node in the networked environment of FIG. **1** according to various embodiments of the present disclosure.

FIG. **3** is a flowchart illustrating one example of RF harvesting functionality implemented in a node in the networked environment of FIG. **1** according to various embodiments of the present disclosure.

FIG. **4** is a flowchart illustrating another example of RF harvesting functionality implemented in a node in the networked environment of FIG. **1** according to various embodiments of the present disclosure.

FIG. **5** is a block diagram that provides another example illustration of a node employed in the networked environment of FIG. **1** according to various embodiments of the present disclosure.

DETAILED DESCRIPTION

The techniques disclosed herein are directed to a node capturing ambient RF emissions in order to supplement the energy provided to the node by an energy storage device. The node may use an antenna to capture the energy from RF emissions from various commercially used RF bands, such as cellular, broadcast television, and unlicensed industrial, scientific and medical (ISM) bands (900 MHz and 2.4 GHz). For example, research studies have shown an ambient RF energy density of between −60 and −14 dB/m^2 in typical urban environments, all within the 685 MHz to 3.5 GHz frequency range. In these studies, the most energy was found to be in the 1.8-1.9 GHz band.

A node, such as used in a meter for utility service (e.g. electricity, water or gas) as part of an AMI (advanced metering infrastructure) network, can incorporate one or more wideband antennas tuned for all or a portion of these frequency bands. The wideband antenna would collect or "harvest" the RF energy, which could then be converted to a DC electrical energy by, for example, an RF/microwave rectifier circuit. The DC output of the rectifier would then be stored in an energy storage device of the node, such as a battery or supercapacitor, via a charging circuit. The energy stored in the energy storage device would then provide DC power for the node during its operation. While the amount of energy harvested may not be equivalent to the energy consumed by the node during all stages of operation, the energy may be used to supplement the energy stored in the energy storage device.

As defined herein, a "node" includes an intelligent device capable of performing functions related to distributing messages in a mesh network. In one system, a node can be a meter located at a facility, such as a house or apartment, that measures the consumption of a utility such as gas, water, or electric power. Such a meter can be part of an RF network used for AMI. Other examples of nodes include a router, collector or collection point, host computer, hub, or other electronic device that is attached to a network and is capable of sending, receiving, or forwarding information over a communications channel.

A node can contain several components that enable it to function within implementations of the present invention. For example, a node can include a radio that can enable it to communicate with like nodes and/or other devices in the mesh network. The radio of each node may have a programmable logic controller (PLC)-like device that can enable the radio to function like a computer, carrying out computer and command functions to provide implementations of the present invention described herein. A node may also include a storage medium for storing information related to communication with other nodes. Such storage mediums can include a memory, a floppy disk, CD-ROM, DVD, or other storage devices located internal to the node or accessible by the node via a network, for example. A node may also include a crystal oscillator (i.e. a clock) to provide timekeeping and an energy storage device (e.g. a battery) to provide back-up power. Some nodes may be powered only by an energy storage device.

With reference to FIG. **1**, shown is an exemplary illustration of a mesh network that may be employed as part of an AMI or other metering network. The head-end system **103** controls the metering network by communicating through the network **106**. The network **106** that the head-end system **103** may utilize to communicate to a node **109** may include devices such as other nodes, **110-112**, collectors **115**, routers **118**, and possibly other devices. Communication may proceed utilizing any appropriate protocol and any appropriate network configuration. Protocols used by the network **106** include, but are not limited to the IEEE 802.15.4, PRIME, G3, and/or TCP/IP protocols. Several nodes **109-112** may transmit data to a router **118**. The router **118**, in turn, may route data to a collector **115** in the network **106**. A collector may receive data from multiple routers. The collector **115** communicates with the head-end system **103**. The head-end system may receive and send information to multiple collectors.

The nodes, such as the node **109**, may be meters that are usually in geographically dispersed locations such as homes or businesses. The meters are used to monitor a utility resource such as electricity, water, or natural gas and to measure the usage of the resource. Some meters may be smart meters that support a variety of service commands. These service commands may allow utilities to disconnect, or limit service remotely or manually at the meter. In addition, some meters may store an event log that contains entries of functions the meter has performed. Service commands may originate from the head-end system **103** and are sent via the network **106** to nodes **109-112**.

To support functionality and continuous operation of a meter, the node **109** may include an energy storage device either as a supplemental/backup power source or as a sole power source. In order to reduce the electrical power consumed during operation the node **109**, may use an antenna to capture the energy from ambient RF emissions **121** from various commercially used RF bands, such as cellular, broadcast television, and ISM bands. While harvesting RF emissions in this manner has the benefit of reducing electrical power consumed from other sources, it can be especially critical when the node is reliant upon an energy storage device as the only power source for an extended period.

In some embodiments, the node **109** may begin harvesting energy from the ambient RF emissions **121** during periods in which no communications are expected to occur. During such periods, a portion of the components of the node **109** may enter a low-energy "sleep" mode in order to conserve electrical power. In these embodiments, an RF switch may be used to switch the one or more antennas of the node from the transceiver circuitry to the circuitry used for RF harvesting.

In other embodiments, a node may harvest energy from the ambient RF emissions **121** independent of communications of node through the use of antenna(s) and other circuitry that is separate from the communications circuitry. In these embodiments, the RF harvesting may occur continuously or when a threshold is met associated with the energy storage device of the node (e.g. the energy level is less than 50% of capacity).

In addition, a node may communicate with other devices of the network **106** in order to request these devices to begin transmitting RF emissions in order to augment the ambient RF emissions **121** received by the node. Such a scenario would typically occur with, for example, a battery-powered node requesting that another device powered from an external source provide additional RF emissions. A request for additional RF emissions may occur, for example, when the energy storage device of the node reaches a threshold, when the ambient RF emissions **121** are insufficient, and/or based upon other criteria.

Referring next to FIG. **2**, shown is a block diagram of an illustrative embodiment of the node **109** which implements the RF harvesting techniques. The node **109** includes one or more antennas, represented by the antenna **201**, which may be used to transmit and receive communications for the node **109**, as well as capture RF emissions to be stored as energy. The design of the antenna **201** may be affected by the RF band(s) at which the node communicates in a network **106** and the RF band(s) at which the node will capture RF emissions. For example, the antenna **201** may include an antenna for the ISM 900 MHz bands, in addition to another antenna used for the 1.8-1.9 GHz RF bands.

In this illustrative embodiment of FIG. **2**, the antenna **201** is connected to an RF switch **203** that connects the antenna **201** to either the transceiver signal path or the RF harvesting signal path. As described previously, other arrangements allowing, for example, simultaneous RF harvesting and data communication may also be possible. In FIG. **2**, the transceiver signal path includes the circuitry used for communicating data to and from the node **109**, while the RF harvesting signal path includes the circuitry used to charge the energy storage device using the captured RF emissions. A node may be said to operate in a "communication mode" while the transceiver signal path is selected via the RF switch, and in an "RF harvest mode" while the RF harvesting signal path is selected.

The RF harvesting signal path comprises a rectifier **205**, charging circuitry **207**, an energy storage device **209**, and/or other components not described in detail herein. The rectifier **205** is an RF/microwave rectifier that converts the energy from the captured RF emissions as input to a DC electrical energy output. The rectifier may be implemented using, for example, diodes.

The charging circuitry **207** may adjust the voltage or other components of the DC output from the rectifier **205** to a form suitable for storing in the energy storage device **209**. The energy storage device **209** represents one or more energy storage devices such as a battery, capacitor, supercapacitor, and/or other electrical energy storage devices. In the embodiment shown in FIG. **2**, the energy storage device **209** is the only source of power for the node **109**, though other configurations that use an external power source in conjunction with a local energy storage device are possible.

A microcontroller **211** is responsible for managing the operation and data communications of the node **109**. To this end, the microcontroller **211** manages the RF switch **203** and the transceiver circuitry **213** that facilitates transmission and reception of data along the transceiver signal path. In some embodiments, the microcontroller **211** may also monitor charging of the energy storage device **209** via the charging circuitry **207**. The various components of the node **109** may receive electrical power from the energy storage device **209**. The electrical power from the energy storage device **209** may be adjusted by the voltage regulator **215** in order to accommodate the electrical needs of disparate electrical components. For example, the voltage regulator **215** may adjust the 3.0 V output from the energy storage device to the 3.45 V required by the microcontroller **211**.

Referring next to FIG. **3**, shown is a flowchart that provides an exemplary implementation of the RF harvesting operations for a method **300** of a node, such as the node **109**, according to various embodiments. It is understood that the flowchart of FIG. **3** provides merely an example of the many different types of functional arrangements that may be employed to implement the RF harvesting operations of the method **300** as described herein. The operations depicted in the flowchart of FIG. **3** illustrate the transition of the node from operating in a communication mode to operating in a harvest mode. As described previously, variations of this implementation may also be possible allowing a node to communicate via a network, while concurrently capturing energy from ambient RF emissions.

Beginning with block **303**, the node determines whether to begin the RF harvest mode or to remain in the communication mode. The determination may be made based upon the current or expected data communications with other devices in a network, a threshold in the energy storage device of the node, and/or other criteria as can be appreciated. For example, if the node had no current data communication taking place, nor any that were expected, some components (e.g. the transceiver circuitry) of the node might enter a low-power “sleep mode,” at which time the node may also enter a harvest mode to capture RF emissions. Similarly, if the node detected that the energy level of the energy storage device was, for example, below a threshold of 50%, the node may begin RF harvesting instead of taking place in non-essential communications.

If the node determines that RF harvesting is not ready to begin, then execution of the method **300** returns to block **303**. Alternatively, if RF harvesting should be initiated, in block **306**, the node may place the RF switch in a position to provide received RF emission to the RF harvesting signal path. Thereafter, in block **309**, the node may receive RF emissions via the one or more antennas of the node. The node may use the antenna(s) to capture the energy from RF emissions from various commercially used RF bands, such as cellular, broadcast television, ISM bands (900 MHz and 2.4 GHz), and/or other possibilities. For example, research studies have shown an ambient RF energy density of between −60 and −14 dB/m2 in typical urban environments, all within the 685 MHz to 3.5 GHz frequency range. In these studies, the most energy was found to be in the 1.8-1.9 GHz band.

Next, in block **312**, the rectifier of the node converts the energy from the various received RF emissions to DC electrical energy. Then, in block **315**, the DC energy output from the rectifier may be stored in the energy storage device of the node through use of charging circuitry. The charging circuitry may be used to monitor and adjust the characteristics of the DC energy used to charge the energy storage device, monitor the current capacity of the energy storage device, and/or other possible re-charging functions. In some embodiments, some or all of the DC energy output from the rectifier may be used to directly power, at least in part, one or more other components of the node such as, for example, the microcontroller. During periods of low power usage, such as during the sleep mode, the energy collected from RF emissions may be sufficient to provide a substantial portion of the power consumed by various components (e.g. the microcontroller).

Subsequently, in block **318**, the node determines whether to continue the RF harvesting (i.e. harvest mode). The determination may be made based upon the current or expected data communications with other devices in a network, a threshold in the energy storage device of the node, and/or other criteria as can be appreciated. For example, if the node detected that the energy level of the energy storage device was at a threshold of 100%, the node may no longer need to continue to harvest the energy since no more capacity exists to store it. Similarly, if the node is scheduled to soon report a utility meter reading to a head-end system, the node may end the harvest mode in order to deliver the scheduled communication data.

In the event that the RF harvesting is determined to continue, execution of the method **300** returns to block **309**. Alternatively, if it is determined that the current period of RF harvesting should end, in block **321**, the node places the RF switch in a position that connects the antenna to the transceiver signal path in order to enable the node to communicate with other devices in the network. Thereafter, execution of the method **300** may return to block **303**.

Moving on to FIG. **4**, shown is a flowchart that provides another exemplary implementation of the RF harvesting operations for a method **400** of a node, such as the node **109**, according to various embodiments. It is understood that the flowchart of FIG. **4** provides merely an example of the many different types of functional arrangements that may be employed to implement the RF harvesting operations of the method **400** as described herein. The operations depicted in the flowchart of FIG. **4** illustrate the transition of the node from operating in a communication mode to operating in a harvest mode, where the node requests a neighboring node to supplement the ambient RF emissions by generating additional RF emissions to be harvested by the node. As described previously, variations of this implementation may also be possible allowing a node to communicate via network, while concurrently capturing energy from ambient RF emissions.

Beginning with block **403**, the node determines whether to begin the RF harvest mode or to remain in the communication mode. The determination may be made based upon the current or expected data communications with other devices in a network, a threshold in the energy storage device of the node, and/or other criteria as can be appreciated.

If the node determines that RF harvesting is not ready to begin, then execution of the method **400** returns to block **403**. Alternatively, if RF harvesting should be initiated, in block **404**, the node may communicate with one or more neighboring nodes in the network to request that the node(s) produce additional RF emissions in order to supplement the energy received during the harvest mode. Typically, a node would make a request to neighboring nodes known to have an external power source (i.e. power supplied through the electric utility grid) or that otherwise have a substantial reserve of energy available.

Then, in block **406**, the node may place the RF switch in a position to provide received RF emission to the RF harvesting signal path. Thereafter, in block **409**, the node may receive RF emissions via the one or more antennas of the node. The neighboring node(s) responding to the request for addition RF emissions may produce a continuous waveform (CW) of RF energy that may be received by the requesting node, as well as other nodes in the area. The node may use the antenna(s) to capture the energy from RF emissions from various commercially used RF bands, such as cellular, broadcast television, ISM bands (900 MHz and 2.4 GHz), and/or other possibilities.

Next, in block **412**, the rectifier of the node converts the energy from the various received RF emissions to DC electrical energy. Then, in block **415**, the DC energy output from the rectifier may be stored in the energy storage device of the node through use of charging circuitry. The charging circuitry may be used to monitor and adjust the characteristics of the DC energy used to charge the energy storage device, monitor the current capacity of the energy storage device, and/or other possible re-charging functions.

Subsequently, in block **418**, the node determines whether to continue the RF harvesting (i.e. harvest mode). The determination may be made based upon the current or expected data communications with other devices in a network, a threshold in the energy storage device of the node, and/or other criteria as can be appreciated. For example, if the node detected that the energy level of the energy storage device was at a threshold of 100%, the node may no longer need to continue to harvest the energy since no more capacity exists to store it. Similarly, if the node is scheduled to soon report a utility meter reading to a head-end system, the node may end the harvest mode in order to deliver the scheduled communication data.

In the event that the RF harvesting is determined to continue, execution of the method **400** returns to block **409**. Alternatively, if it is determined that the current period of RF harvesting should end, in block **421**, the node places the RF switch in a position that connects the antenna to the transceiver signal path in order to enable the node to communicate with other devices in the network. As an optional step, in block **424**, once the node has returned to the communication mode, the node may transmit a notification to the neighboring nodes indicating that the node is no longer in harvest mode and that the transmission of additional RF emission is no longer needed by the node. In some implementations, the neighboring nodes producing additional RF emissions as requested in block **404** may do so for a predefined period of time. Thereafter, execution of the method **400** may return to block **403**.

Next, in FIG. **5**, shown is another block diagram depicting an example of a node **109-112** used for implementing the techniques disclosed herein. The node **109-112** can include a processing device **502**. Non-limiting examples of the processing device **502** include a microprocessor, an application-specific integrated circuit ("ASIC"), a state machine, or other suitable processing device. The processing device **502** can include any number of processing devices, including one. The processing device **502** can be communicatively coupled to computer-readable media, such as memory device **504**. The processing device **502** can execute computer-executable program instructions and/or access information respectively stored in the memory device **504**. In some implementations, the processor **502** and memory **504** may be incorporated as components of a microcontroller, such as the microcontroller **211**.

The memory device **504** can store instructions that, when executed by the processing device **502**, cause the processing device **502** to perform operations described herein. The memory device **504** may be a computer-readable medium such as (but not limited to) an electronic, optical, magnetic, or other storage device capable of providing a processor with computer-readable instructions. Non-limiting examples of such optical, magnetic, or other storage devices include read-only ("ROM") device(s), random-access memory ("RAM") device(s), magnetic disk(s), magnetic tape(s) or other magnetic storage, memory chip(s), an ASIC, configured processor(s), optical storage device(s), or any other medium from which a computer processor can read instructions. The instructions may comprise processor-specific instructions generated by a compiler and/or an interpreter from code written in any suitable computer-programming language. Non-limiting examples of suitable computer-programming languages include C, C++, C#, Visual Basic, Java, Python, Perl, JavaScript, and the like.

The nodes **109-112** can include a bus **506** that can communicatively couple one or more components of the node **109-112**. Although the processor **502**, the memory **504**, and the bus **506** are depicted in FIG. **5** as separate components in communication with one another, other implementations are possible. For example, the processor **502**, the memory **504**, and the bus **506** can be components of printed circuit boards or other suitable devices that can be disposed in a node **109-112** to store and execute programming code.

The nodes **109-112** can also include network interface device **508**, such as may be implemented by the transceiver circuitry **213**. The network interface device **508** can be a transceiving device configured to establish one or more of the wireless communication links via an antenna **201**. A non-limiting example of the network interface device **508** is an RF transceiver and can include one or more components for establishing a communication links to other nodes **109-112** in the mesh network **106**.

Numerous specific details are set forth herein to provide a thorough understanding of the claimed subject matter. However, those skilled in the art will understand that the claimed subject matter may be practiced without these specific details. In other instances, methods, apparatuses, or systems that would be known by one of ordinary skill have not been described in detail so as not to obscure claimed subject matter.

Some portions are presented in terms of algorithms or symbolic representations of operations on data bits or binary digital signals stored within a computing system memory, such as a computer memory. These algorithmic descriptions or representations are examples of techniques used by those of ordinary skill in the data processing arts to convey the substance of their work to others skilled in the art. An algorithm is a self-consistent sequence of operations or similar processing leading to a desired result. In this context, operations or processing involves physical manipulation of physical quantities. Typically, although not necessarily, such quantities may take the form of electrical or magnetic signals capable of being stored, transferred, combined, compared or otherwise manipulated. It has proven convenient at times, principally for reasons of common usage, to refer to such signals as bits, data, values, elements, symbols, characters, terms, numbers, numerals, or the like. It should be understood, however, that all of these and similar terms are to be associated with appropriate physical quantities and are merely convenient labels. Unless specifically stated otherwise, it is appreciated that throughout this specification discussions utilizing terms such as "processing," "computing," "calculating," "determining," and "identifying" or the like refer to actions or processes of a computing device, such as one or more computers or a similar electronic computing device or devices, that manipulate or transform data represented as physical electronic or magnetic quantities within memories, registers, or other storage devices, transmission devices, or display devices of the computing platform.

The system or systems discussed herein are not limited to any particular hardware architecture or configuration. A computing device can include any suitable arrangement of components that provide a result conditioned on one or more function calls. Suitable computing devices include multipurpose microprocessor-based computer systems accessing stored software that programs or configures the computing system from a general-purpose computing apparatus to a specialized computing apparatus implementing one or more aspects of the present subject matter. Any suitable programming, scripting, or other type of language or combinations of languages may be used to implement the teachings contained herein in software to be used in programming or configuring a computing device.

Aspects of the methods disclosed herein may be performed in the operation of such computing devices. The order of the blocks presented in the examples above can be varied—for example, blocks can be re-ordered, combined, and/or broken into sub-blocks. Certain blocks or processes can be performed in parallel.

The use of "adapted to" or "configured to" herein is meant as open and inclusive language that does not foreclose devices adapted to or configured to perform additional tasks or steps. Additionally, the use of "based on" is meant to be open and inclusive, in that a process, step, calculation, or other action "based on" one or more recited conditions or values may, in practice, be based on additional conditions or values beyond those recited. Headings, lists, and numbering included herein are for ease of explanation only and are not meant to be limiting.

While the present subject matter has been described in detail with respect to specific aspects thereof, it will be appreciated that those skilled in the art, upon attaining an understanding of the foregoing, may readily produce alterations to, variations of, and equivalents to such aspects. Accordingly, it should be understood that the present disclosure has been presented for purposes of example rather than limitation, and does not preclude inclusion of such modifications, variations, and/or additions to the present subject matter as would be readily apparent to one of ordinary skill in the art.

Therefore, the following is claimed:

1. A method for energy harvesting by circuitry of a node in a network, comprising:

operating in a communication mode comprising:

communicating, via a radio frequency (RF) transceiver, data with another node in the network, wherein the node comprises an energy storage device that provides power for the node; and

in response to occurrence of a time period during which no further data communication is scheduled in the node, switching, via an RF switch in the node, from said communication mode to an energy harvest mode comprising:

receiving RF emissions accessible to the node via an antenna;

providing the received RF emissions to a rectifier;

converting the RF emissions to direct current (DC) energy, wherein said converting is implemented using the rectifier; and

storing the DC energy in the energy storage device.

2. The method of claim **1**, further comprising switching back to the communication mode when the node determines additional data needs to be transmitted or received by the node, the communication mode further comprising:

receiving additional RF emissions accessible to the node via the antenna; and

decoding the additional data from the additional RF emissions.

3. The method of claim **1**, wherein the node is a utility meter deployed as part of an advanced metering infrastructure (AMI) system.

4. The method of claim **1**, wherein the energy storage device is the only source of electrical power for the node during operation.

5. The method of claim **1**, wherein said switching to the energy harvesting mode is based at least in part upon the node determining that a sleep mode can begin.

6. The method of claim **1**, wherein said receiving RF emissions comprises receiving a plurality of RF signals in one or more RF bands.

7. A node, comprising:

a processor;

an antenna;

a radio frequency (RF) switch for switching connectivity to the antenna;

an energy storage device that provides power for the node;

a rectifier;

an RF transceiver for communicating on a network accessible to the node; and

a memory configured by a management application executed in the node, the management application comprising logic causing the node to:

operate in an energy harvest mode comprising:

receiving RF emissions accessible to the node via the antenna;

converting, via the rectifier, the RF emissions to direct current (DC) energy; and

storing the DC energy in the energy storage device; and

in response to occurrence of a time period during which data communication is scheduled in the node, switch, via the RF switch, from said energy harvest mode to a communication mode comprising:

communicating, via the RF transceiver, data with another node in the network, wherein said communicating uses power from the energy storage device.

8. The node of claim **7**, wherein said switch to the communication mode occurs when the node determines additional data needs to be transmitted or received by the node.

9. The node of claim **8**, wherein the additional data needs to be transmitted or received by the node on a periodic basis.

10. The node of claim **7**, wherein the management application further comprises logic to request the other node in the network to transmit additional RF emissions to the node.

11. The node of claim **7**, wherein the network is a wireless mesh network of a plurality nodes that includes the node.

12. The node of claim **7**, wherein the RF transceiver operates in one or more of the industrial, scientific and medical (ISM) RF bands.

13. The node of claim **12**, wherein said receiving RF emissions comprises receiving a plurality of RF signals in one or more RF bands, a portion of the one or more RF bands being outside the ISM RF bands.

14. A non-transitory computer-readable medium embodying a program executable by a processor of a node, comprising:

code for operating the node in a communication mode comprising:

communicating, via a radio frequency (RF) transceiver, data with another node in a network, wherein the node comprises an energy storage device that provides power for the node;

code for operating the node in an energy harvest mode comprising:

receiving RF emissions accessible to the node via an antenna;

providing the received RF emissions to a rectifier;

converting the RF emissions to direct current (DC) energy, wherein said converting is implemented using the rectifier; and

storing the DC energy in the energy storage device; and code for switching, via an RF switch in the node, between the modes based upon a determination made in the node, wherein the node switches to the energy harvesting mode in response to occurrence of a time period during which no further data communication is scheduled in the node.

15. The non-transitory computer-readable medium of claim **14**, wherein the node receives electrical power only from the energy storage device in the node.

16. The non-transitory computer-readable medium of claim **15**, wherein the program further comprises code for requesting the other node in the network to transmit additional RF emissions to the node, wherein the other node receives electrical power from an external source.

17. The non-transitory computer-readable medium of claim **14**, wherein the node is a utility meter deployed as part of an advanced metering infrastructure (AMI) system.

18. The non-transitory computer-readable medium of claim **14**, wherein the determination comprises switching to the energy harvest mode based upon a power level of the energy storage device reaching a threshold.

19. The non-transitory computer-readable medium of claim **14**, wherein the determination comprises switching to the energy harvest mode based upon a completion of a scheduled data communication.

* * * * *

US009830548B2

(12) **United States Patent**
Rokhsaz et al.

(10) **Patent No.: US 9,830,548 B2**
(45) **Date of Patent: Nov. 28, 2017**

(54) **METHOD AND APPARATUS FOR DETECTING RF FIELD STRENGTH**

(71) Applicant: **RFMicron, Inc.**, Austin, TX (US)

(72) Inventors: **Shahriar Rokhsaz**, Austin, TX (US); **Edwin de Angel**, Austin, TX (US)

(73) Assignee: **RFMicron, Inc.**, Austin, TX (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **15/643,962**

(22) Filed: **Jul. 7, 2017**

(65) **Prior Publication Data**

US 2017/0308783 A1 Oct. 26, 2017

Related U.S. Application Data

(60) Continuation of application No. 14/644,471, filed on Mar. 11, 2015, now Pat. No. 9,704,085, which is a continuation of application No. 13/209,425, filed on Aug. 14, 2011, now Pat. No. 9,048,819, which is a continuation-in-part of application No. 12/462,331, filed on Aug. 1, 2009, now Pat. No. 8,081,043, which is a division of application No. 11/601,085, filed on Nov. 18, 2006, now Pat. No. 7,586,385.

(60) Provisional application No. 61/428,170, filed on Dec. 29, 2010, provisional application No. 61/485,732, filed on May 13, 2011.

(51) **Int. Cl.**

H04Q 3/00	(2006.01)
G08B 13/00	(2006.01)
G06K 19/07	(2006.01)
H04B 17/318	(2015.01)
H04B 5/00	(2006.01)

(52) **U.S. Cl.**
CPC ***G06K 19/0726*** (2013.01); ***H04B 5/0062*** (2013.01); ***H04B 17/318*** (2015.01)

(58) **Field of Classification Search**
CPC G06K 19/0726; G06K 7/0008; H04B 17/318; H04B 5/0062; G08B 13/2442; E05B 73/0017
USPC 340/10.1–10.6, 572.1–572.9; 343/872, 343/867, 866
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

8,368,514 B2 *	2/2013	Wiberg	G06K 19/0723	235/375
2005/0001031 A1 *	1/2005	Akiho	G06K 7/0008	235/451

(Continued)

Primary Examiner — George Bugg
Assistant Examiner — Munear Akki
(74) *Attorney, Agent, or Firm* — Garlick & Markison; Timothy W. Markison

(57) **ABSTRACT**

A radio frequency (RF) circuit comprises an antenna operably coupled to receive an RF signal, a power harvesting circuit operable to generate a first power source from the RF signal, a tank circuit coupled to the antenna and a tuning circuit. The tank circuit includes a selectively variable impedance and the tuning circuit is adapted to dynamically vary the selectively variable impedance of the tank circuit based on resonance of the RF circuit and frequency of the RF signal to substantially align the resonance of the RF circuit with the frequency of the RF signal. When the resonance of the RF circuit is substantially aligned with the frequency of the RF signal, the power harvesting circuit generates a second power source from the RF signal such that the second power source is greater than the first power source.

8 Claims, 4 Drawing Sheets

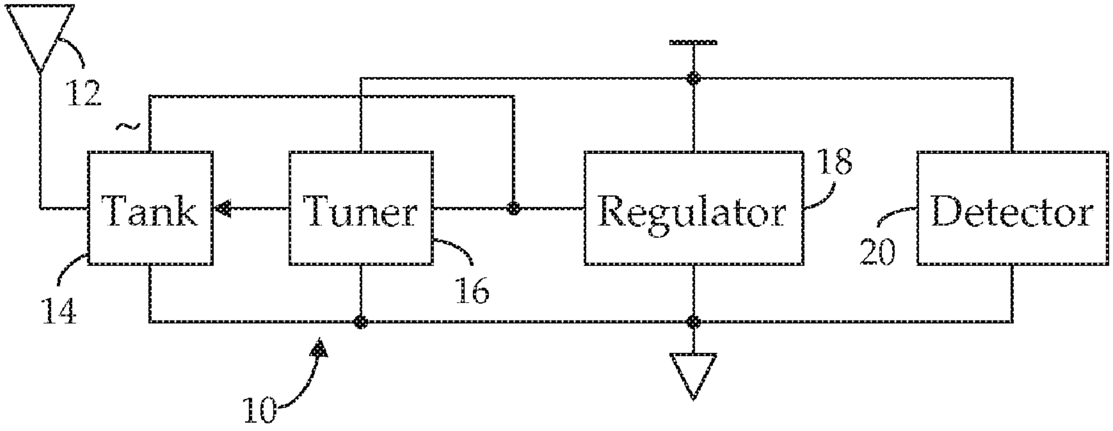

(56) References Cited

U.S. PATENT DOCUMENTS

2007/0008069	A1*	1/2007	Lastinger	G06K 7/0008 340/10.1
2009/0201169	A1*	8/2009	d'Hont	G01S 13/003 340/8.1
2010/0220766	A1*	9/2010	Burgard	G01K 1/024 374/183
2011/0291810	A1*	12/2011	Rokhsaz	H03J 3/20 340/10.1
2011/0300808	A1*	12/2011	Rokhsaz	H03J 3/20 455/67.11
2012/0217311	A1*	8/2012	Rokhsaz	G06K 19/0726 235/492
2014/0227981	A1*	8/2014	Pecen	H01Q 1/52 455/77

* cited by examiner

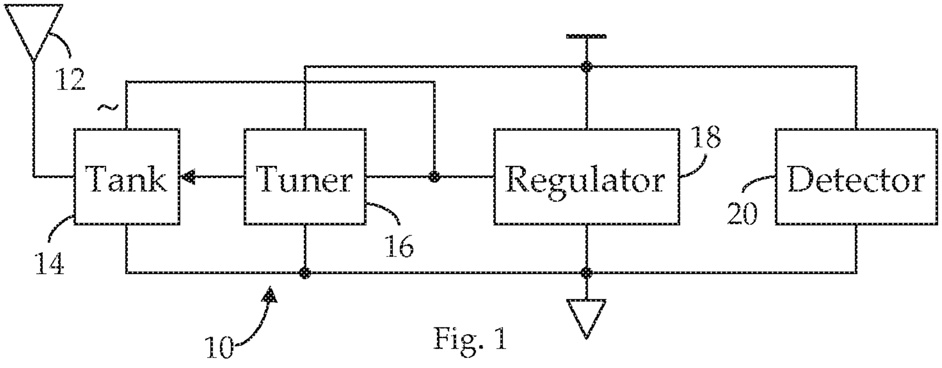
12
~
Tank
Tuner
Regulator
18
Detector
20
14
16
10

Fig. 1

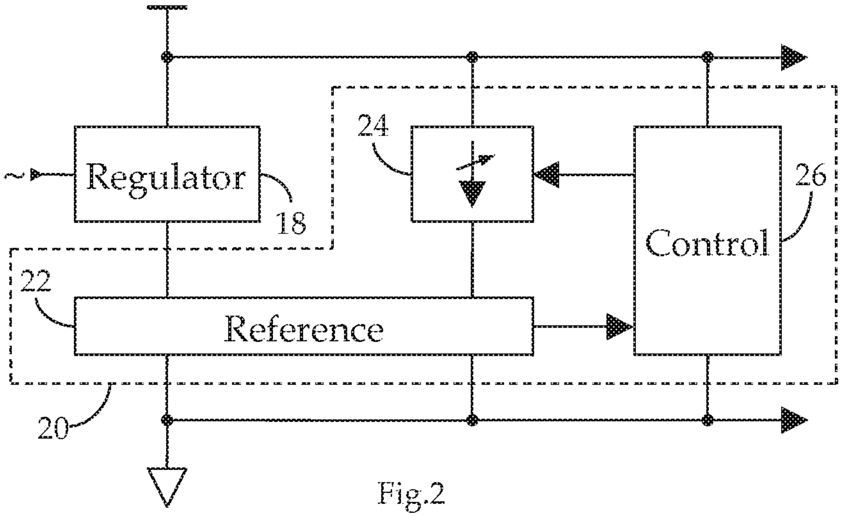
~
Regulator
18
24
Control
26
22
Reference
20

Fig.2

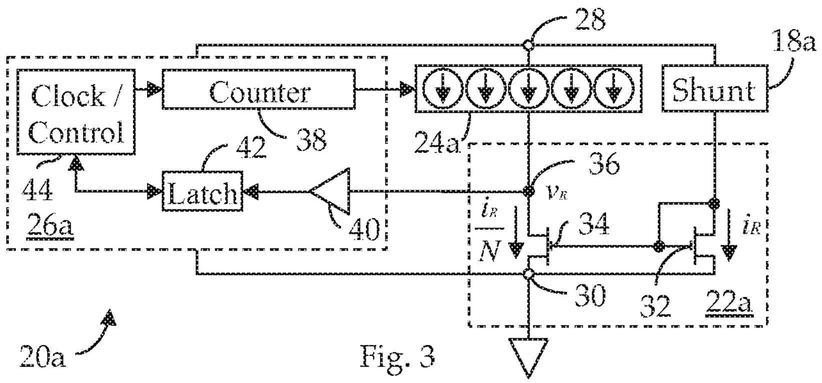
28
18a
Clock /
Control
Counter
Shunt
24a
38
42
36
v_R
44
Latch
26a
40
$\frac{i_R}{N}$
34
i_R
30
32
22a
20a

Fig. 3

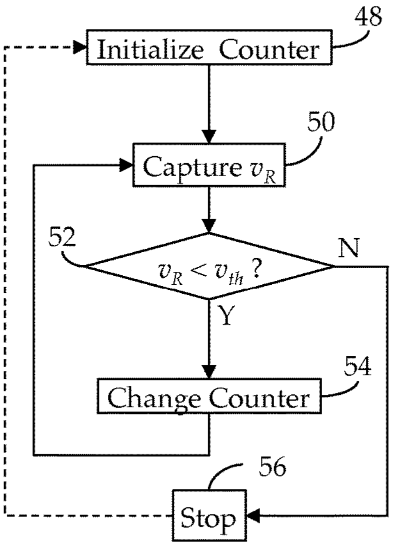
Initialize Counter
48
Capture v_R
50
52
$v_R < v_{th}$?
N
Y
Change Counter
54
56
Stop

Fig. 4

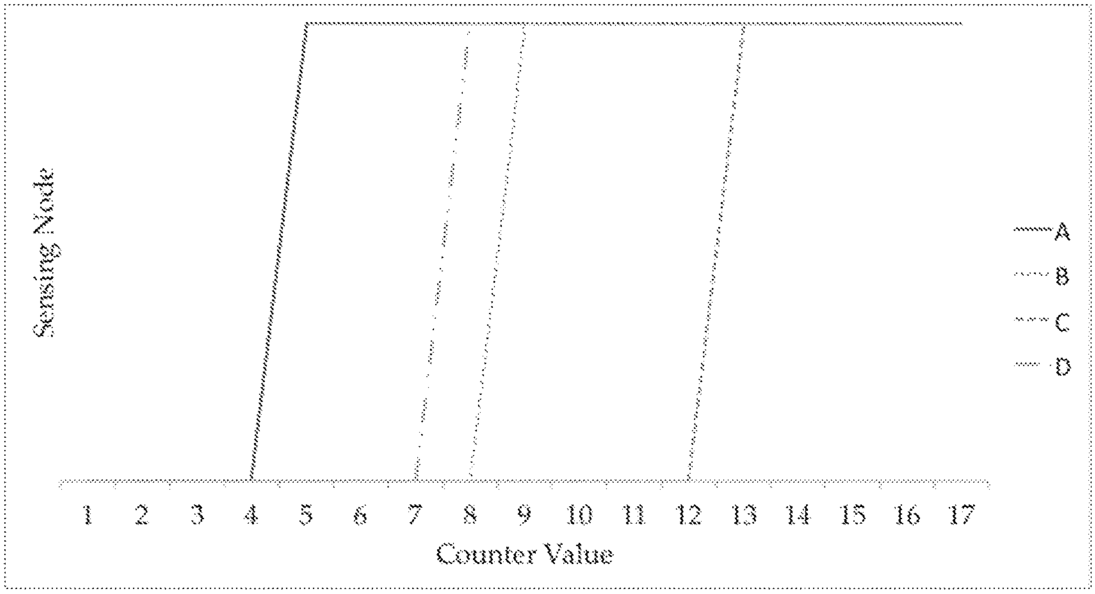
Sensing Node
A
B
C
D
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
Counter Value

Fig. 5

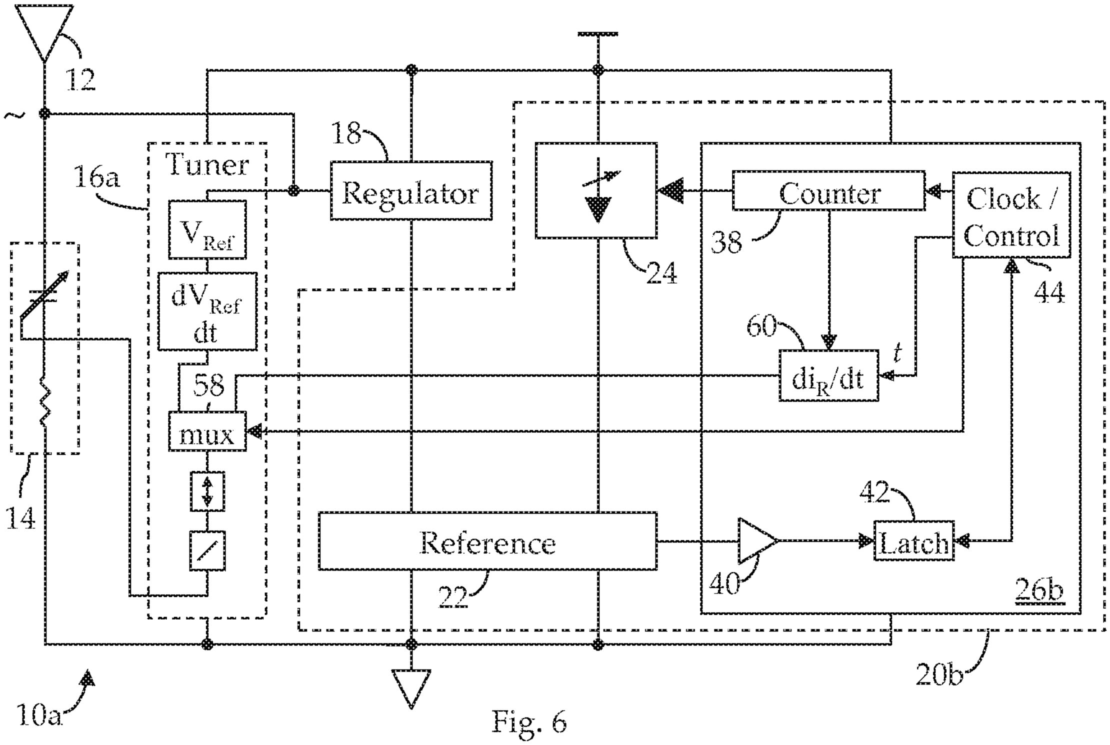
12
~
16a
Tuner
V_{Ref}
$\frac{dV_{Ref}}{dt}$
58
mux
18
Regulator
24
Counter
Clock / Control
38
44
60
di_R/dt
t
14
42
Latch
Reference
22
40
26b
20b
10a

Fig. 6

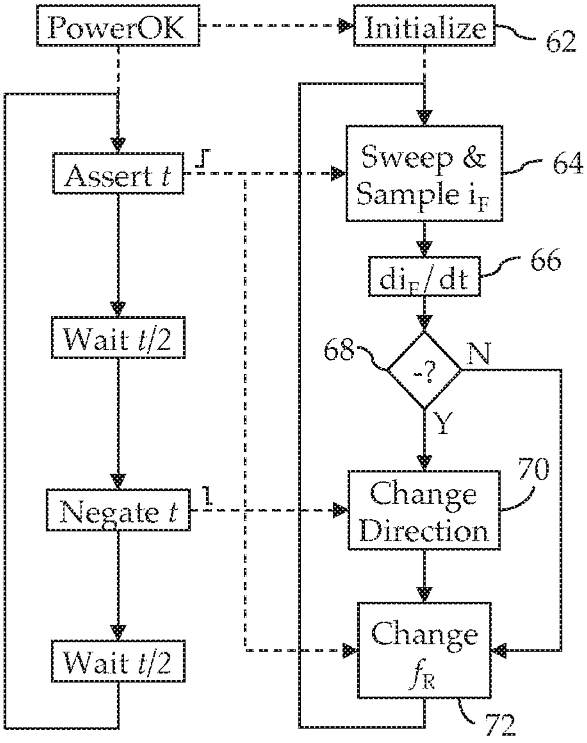
PowerOK
Initialize
62
Assert t
Sweep & Sample i_F
64
di_F/dt
66
Wait t/2
68
-?
N
Y
Negate t
Change Direction
70
Wait t/2
Change f_R
72

Fig. 7

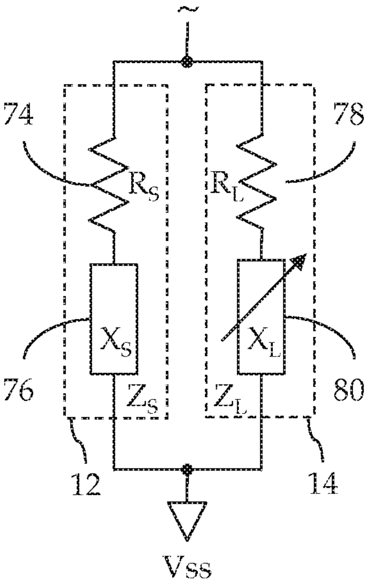
~
74
78
R_S
R_L
X_S
X_L
76
80
Z_S
Z_L
12
14
Vss

Fig. 8

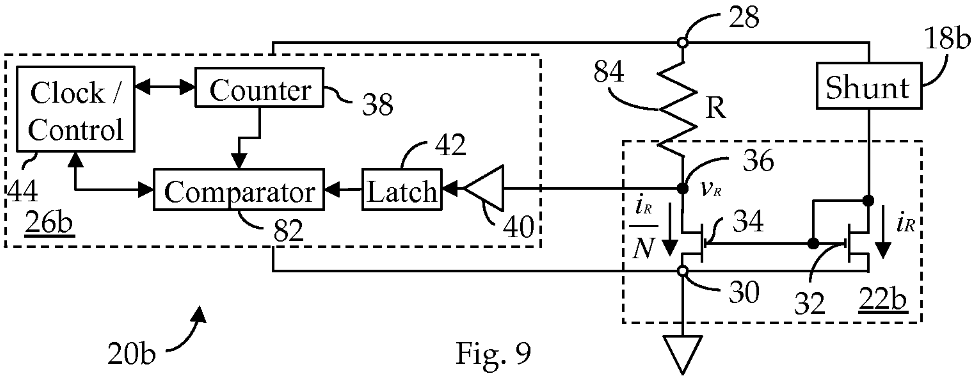
28
18b
84
R
Shunt
Clock / Control
Counter
38
42
36
v_R
Comparator
Latch
44
26b
82
40
$\frac{i_R}{N}$
34
i_R
30
32
22b
20b

Fig. 9

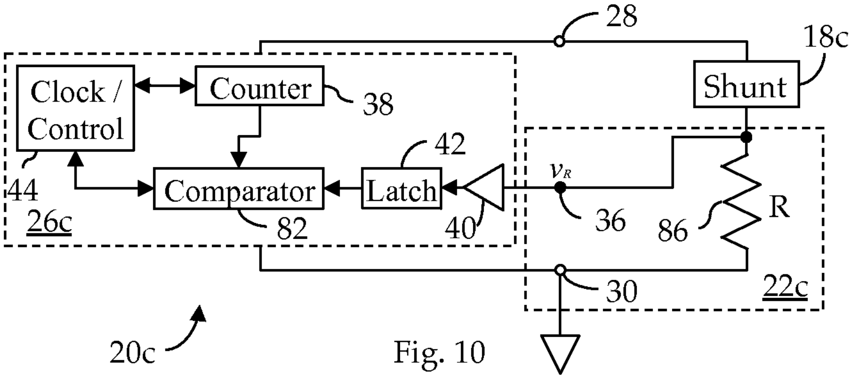
28
18c
Shunt
Clock / Control
Counter
38
42
Comparator
Latch
v_R
R
44
26c
82
40
36
86
30
22c
20c

Fig. 10

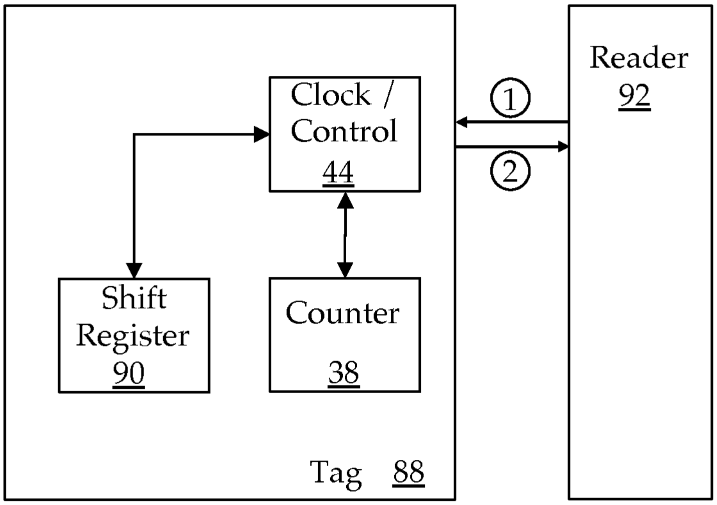
Reader
92
Clock / Control
44
①
②
Shift Register
90
Counter
38
Tag 88

Fig. 11

METHOD AND APPARATUS FOR DETECTING RF FIELD STRENGTH

CROSS-REFERENCE TO RELATED APPLICATIONS

The present U.S. Utility Patent Application claims priority pursuant to 35 U.S.C. §120 as a Continuation of U.S. Utility application Ser. No. 14/644,471, entitled "Method and Apparatus for Detecting RF Field Strength", filed 11 Mar. 2015, issuing as U.S. Pat. No. 9,704,085 on 11 Jul. 2017, which is a Continuation of U.S. Utility application Ser. No. 13/209,425, entitled "Method and Apparatus for Detecting RF Field Strength", filed 14 Aug. 2011, now U.S. Pat. No. 9,048,819, issued 2 Jun. 2015, which is a Continuation-In-Part of application Ser. No. 12/462,331, filed 1 Aug. 2009, now U.S. Pat. No. 8,081,043, issued 20 Dec. 2011 ("Related Application"), which is in turn a Division of application Ser. No. 11/601,085, filed 18 Nov. 2006, now U.S. Pat. No. 7,586,385, issued 8 Sep. 2009 ("Related Patent") (collectively, "Related References"). Application Ser. No. 13/209,425 also claims priority pursuant to 35 U.S.C. §119(e) to U.S. Provisional Application No. 61/428,170, filed 29 Dec. 2010, and U.S. Provisional Application No. 61/485,732, filed 13 May 2011. The subject matter of the Related References, each in its entirety, is expressly incorporated herein by reference.

This application is related to application Ser. No. 13/209,420, filed on 14 Aug. 2011, now U.S. Pat. No. 8,749,319, issued 10 Jun. 2014 ("Related Co-application").

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates generally to detecting RF field strength, and, in particular, to detecting RF field strength in a passive RFID system.

2. Description of the Related Art

In general, in the descriptions that follow, we will italicize the first occurrence of each special term of art that should be familiar to those skilled in the art of radio frequency ("RF") communication systems. In addition, when we first introduce a term that we believe to be new or that we will use in a context that we believe to be new, we will bold the term and provide the definition that we intend to apply to that term. In addition, throughout this description, we will sometimes use the terms assert and negate when referring to the rendering of a signal, signal flag, status bit, or similar apparatus into its logically true or logically false state, respectively, and the term toggle to indicate the logical inversion of a signal from one logical state to the other. Alternatively, we may refer to the mutually exclusive Boolean states as logic_0 and logic_1. Of course, as is well known, consistent system operation can be obtained by reversing the logic sense of all such signals, such that signals described herein as logically true become logically false and vice versa. Furthermore, it is of no relevance in such systems which specific voltage levels are selected to represent each of the logic states.

In accordance with our prior invention previously disclosed in the Related References, the amplitude modulated ("AM") signal broadcast by the reader in an RFID system will be electromagnetically coupled to a conventional antenna, and a portion of the current induced in a tank circuit is extracted by a regulator to provide operating power for all other circuits. Once sufficient stable power is available, the regulator will produce, e.g., a power-on-reset signal to initiate system operation. Thereafter, the method disclosed in the Related References, and the associated apparatus, dynamically varies the capacitance of a variable capacitor component of the tank circuit so as to dynamically shift the f_R of the tank circuit to better match the f_C of the received RF signal, thus obtaining maximum power transfer in the system.

In general, the invention disclosed in the Related References focused primarily on quantizing the voltage developed by the tank circuit as the primary means of matching the f_R of the tank circuit to the transmission frequency, f_C, of the received signal. However, this voltage quantization is, at best, indirectly related to received signal field strength. We submit that what is needed now is an effective and efficient method and apparatus for quantizing the received field strength as a function of induced current. It is further desirable to develop this field quantization in a form and manner that is suitable for selectively varying the input impedance of the receiver circuit to maximize received power, especially during normal system operation. Additionally, in light of the power sensitive nature of RFID systems, it is desirable to vary the input impedance with a minimum power loss.

BRIEF SUMMARY OF THE INVENTION

In accordance with the preferred embodiment of our invention, we provide a sensing system for use in an RFID system. In general, the sensing system comprises an RFID tag and an RFID reader. In one embodiment, the RFID tag comprises a tank circuit having a selectively variable impedance; and a tuning circuit adapted to dynamically vary the impedance of the tank circuit, and to develop a first quantized value representative of the impedance of said tank circuit. In one alternate embodiment, the RFID tag comprises a detector circuit adapted to develop a second quantized value as a function of a field strength of a received RF signal. In yet another embodiment, the RFID tag comprises both the tank and tuning circuit, and the detector circuit. In these embodiments, RFID reader is adapted selectively to retrieve one or both of the first and second values, and, preferably, to use the retrieved values to sense changes to an environment to which the RFID tag is exposed

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

My invention may be more fully understood by a description of certain preferred embodiments in conjunction with the attached drawings in which:

FIG. **1** illustrates, in block diagram form, an RF receiver circuit having a field strength detector constructed in accordance with an embodiment of our invention;

FIG. **2** illustrates, in block diagram form, a field strength detector circuit constructed in accordance with an embodiment of our invention;

FIG. **3** illustrates, in block schematic form, a more detailed embodiment of the field strength detector circuit shown in FIG. **2**;

FIG. **4** illustrates, in flow diagram form, the sequencing of operations in the field strength detector circuit shown in FIG. **3**;

FIG. **5** illustrates, in graph form, the response of the field strength detector circuit shown in FIG. **3** to various conditions;

FIG. **6** illustrates, in block schematic form, an RF receiver circuit constructed in accordance with another embodiment of our invention;

FIG. **7** illustrates, in flow diagram form, the sequencing of the operations in the RF receiver circuit shown in FIG. **6**;

FIG. **8** illustrates, in block schematic form, an alternative representation of the impedance represented by the antenna and the tank circuit of the exemplary RFID receiver circuit.

FIG. **9** illustrates, in block schematic form, an alternative exemplary embodiment of the field strength detector circuit shown in FIG. **3**.

FIG. **10** illustrates, in block schematic form, an alternative exemplary embodiment of the field strength detector circuit shown in FIG. **3**.

FIG. **11** illustrates, in block schematic form, an exemplary RFID sub-system containing tag and reader.

In the drawings, similar elements will be similarly numbered whenever possible. However, this practice is simply for convenience of reference and to avoid unnecessary proliferation of numbers, and is not intended to imply or suggest that our invention requires identity in either function or structure in the several embodiments.

DETAILED DESCRIPTION OF THE INVENTION

Shown in FIG. **1** is an RF receiver circuit **10** suitable for use in an RFID application. As we have described in our Related References, an RF signal electromagnetically coupled to an antenna **12** is received via a tank circuit **14**, the response frequency, f_R, of which is dynamically varied by a tuner **16** to better match the transmission frequency, f_C, of the received RF signal, thus obtaining a maximum power transfer. In particular, as further noted in the Related Applications, the RMS voltage induced across the tank circuit **14** by the received RF signal is quantized by tuner **16** and the developed quantization employed to control the impedance of the tank circuit **14**. As also described in the Related References, the unregulated, AC current induced in the tank circuit **14** by the received RF signal is conditioned by a regulator **18** to provide regulated DC operating power to the receiver circuit **10**. In accordance with our present invention, we now provide a field strength detector **20**, also known as a power detector, adapted to develop a field-strength value as a function of the field strength of the received RF signal. As we have indicated in FIG. **1**, our field strength detector **20** is adapted to cooperate with the regulator **18** in the development of the field-strength value. As we shall disclose below, if desired, our field strength detector **20** can be adapted to cooperate with the tuner **16** in controlling the operating characteristics of the tank circuit **14**.

Shown by way of example in FIG. **2** is one possible embodiment of our field strength or power detector **20**. In this embodiment, we have chosen to employ a shunt-type regulator **18** so that, during normal operation, we can use the shunted 'excess' current as a reference against which we develop the field-strength value. In this regard, we use a reference **22** first to develop a shunt current reference value proportional to the shunted current, and then to develop a mirrored current reference value as a function of both the shunted current and a field strength reference current provided by a digitally-controlled current source **24**. Preferably, once the tuner **16** has completed its initial operating sequence, whereby the f_R of the tank circuit **14** has been substantially matched to the f_C of the received signal, we then enable a digital control **26** to initiate operation of the current source **24** at a predetermined, digitally-established minimum field strength reference current. After a predetermined period of time, control **26** captures the mirrored current reference value provided by the current reference **22**, compares the captured signal against a predetermined threshold value, and, if the comparison indicates that the field strength reference current is insufficient, increases, in accordance with a predetermined sequence of digital-controlled increments, the field strength reference current; upon the comparison indicating that the field strength reference current is sufficient, control **26** will, at least temporarily, cease operation.

In accordance with our invention, the digital field-strength value developed by control **26** to control the field strength current source **24** is a function of the current induced in the tank circuit **14** by the received RF signal. Once developed, this digital field-strength value can be employed in various ways. For example, it can be selectively transmitted by the RFID device (using conventional means) back to the reader (not shown) for reference purposes. Such a transaction can be either on-demand or periodic depending on system requirements. Imagine for a moment an application wherein a plurality of RFID tag devices are distributed, perhaps randomly, throughout a restricted, 3-dimensional space, e.g., a loaded pallet. Imagine also that the reader is programmed to query, at an initial field strength, all tags "in bulk" and to command all tags that have developed a field-strength value greater than a respective field-strength value to remain silent'. By performing a sequence of such operations, each at an increasing field strength, the reader will, ultimately, be able to isolate and distinguish those tags most deeply embedded within the space; once these 'core' tags have been read, a reverse sequence can be performed to isolate and distinguish all tags within respective, concentric 'shells' comprising the space of interest. Although, in all likelihood, these shells will not be regular in either shape or relative volume, the analogy should still be apt.

In FIG. **3**, we have illustrated one possible embodiment of our field strength detector **20***a*. In general, we have chosen to use a shunt circuit **18***a* to develop a substantially constant operating voltage level across supply node **28** and ground node **30**. Shunt regulators of this type are well known in the art, and typically use zener diodes, avalanche breakdown diodes, diode-connected MOS devices, and the like.

As can be seen, we have chosen to implement current reference **22** in the form of a current mirror circuit **22***a*, connected in series with shunt circuit **18***a* between nodes **28** and **30**. As is typical, current mirror circuit **22***a* comprises a diode-connected reference transistor **32** and a mirror transistor **34**. If desired, a more sophisticated circuit such as a Widlar current source may be used rather than this basic two-transistor configuration. For convenience of reference, we have designated the current shunted by shunt circuit **18***a* via reference transistor **32** as i_R; similarly, we have designated the current flowing through mirror transistor **34** as i_R/N, wherein, as is known, N is the ratio of the widths of reference transistor **32** and mirror transistor **34**.

We have chosen to implement the field strength current source **24** as a set of n individual current sources **24***a*, each connected in parallel between the supply node **28** and the mirror transistor **34**. In general, field strength current source **24***a* is adapted to source current at a level corresponding to an n-bit digital control value developed by a counter **38**. In the illustrated embodiment wherein n=5, field strength current source **24***a* is potentially capable of sourcing thirty-two distinct reference current levels. We propose that the initial, minimum reference current level be selected so as to be less than the current carrying capacity of the mirror transistor **34** when the shunt circuit **18***a* first begins to shunt excess induced current through reference transistor **32**; that the maximum reference current level be selected so as to be greater than the current carrying capacity of the mirror transistor **34** when the shunt circuit **18***a* is shunting a maximum anticipated amount of excess induced current; and that the intermediate reference current levels be distributed relatively evenly between the minimum and maximum levels. Of course, alternate schemes may be practicable, and, perhaps, desirable depending on system requirements.

Within control **26***a*, a conventional analog-to-digital converter ("ADC") **40**, having its input connected to a sensing node **36**, provides a digital output indicative of the field strength reference voltage, v_R, developed on sensing node **36**. In one embodiment, ADC **40** may comprise a comparator circuit adapted to switch from a logic_0 state to a logic_1 when sufficient current is sourced by field strength current source **24***a* to raise the voltage on sensing node **36** above a predetermined reference voltage threshold, v_{th}. Alternatively, ADC **40** may be implemented as a multi-bit ADC capable of providing higher precision regarding the specific voltage developed on sensing node **36**, depending on the requirements of the system. Sufficient current may be characterized as that current sourced by the field strength current source **24***a* or sunk by mirror transistor **34** such that the voltage on sensing node **36** is altered substantially above or below a predetermined reference voltage threshold, v_{th}. In the exemplary case of a simple CMOS inverter, v_{th} is, in its simplest form, one-half of the supply voltage (VDD/2). Those skilled in the art will appreciate that v_{th} may by appropriately modified by altering the widths and lengths of the devices of which the inverter is comprised. In the exemplary case a multi-bit ADC, v_{th} may be established by design depending on the system requirements and furthermore, may be programmable by the system.

In the illustrated embodiment, a latch **42** captures the output state of ADC **40** in response to control signals provided by a clock/control circuit **44**. If the captured state is logic_0, the clock/control circuit **44** will change counter **38** to change the reference current being sourced by field strength current source **24***a*; otherwise clock/control circuit **44** will, at least temporarily, cease operation. However, notwithstanding, the digital field-strength value developed by counter **38** is available for any appropriate use, as discussed above.

By way of example, we have illustrated in FIG. **4** one possible general operational flow of our field strength detector **20***a*. Upon activation, counter **38** is set to its initial digital field-strength value (step **48**), thereby enabling field strength current source **24***a* to initiate reference current sourcing at the selected level. After an appropriate settling time, the field strength reference voltage, v_R, developed on sensing node **36** and digitized by ADC **40** is captured in latch **42** (step **50**). If the captured field strength reference voltage, v_R, is less than (or equal to) the predetermined reference threshold voltage, v_{th}, clock/control **44** will change counter **38** (step **54**). This process will repeat, changing the reference current sourced by field strength current source **24***a* until the captured field strength reference voltage, v_R, is greater than the predetermined reference threshold voltage, v_{th}, (at step **52**), at which time the process will stop (step **56**). As illustrated, this sweep process can be selectively reactivated as required, beginning each time at either the initial field-strength value or some other selected value within the possible range of values as desired.

The graph illustrated in FIG. **5** depicts several plots of the voltage developed on sensing node **36** as the field strength detector circuit **20***a* sweeps the value of counter **38** according to the flow illustrated in FIG. **4**. As an example, note that the curve labeled "A" in FIG. **5** begins at a logic_0 value when the value of counter **38** is at a minimum value such as "1" as an exemplary value. Subsequent loops though the sweep loop gradually increase the field strength reference voltage on sensing node **36** until counter **38** reaches a value of "4" as an example. At this point, the "A" plot in FIG. **5** switches from a logic_0 value to a logic_1 value, indicating that the field strength reference voltage, v_R, on sensing node **36** has exceeded the predetermined reference threshold voltage, v_{th}. Other curves labeled "B" through "D" depict incremental increases of reference currents, i_R, flowing through reference device **32**, resulting in correspondingly higher mirrored currents flowing through mirror device **34**. This incrementally higher mirror current requires field strength current source **24** to source a higher current level which in turn corresponds to higher values in counter **38**. Thus, it is clear that our invention is adapted to effectively and efficiently develop a digital representation of the current flowing through sensing node **36** that is suitable for any appropriate use.

One such use, as discussed earlier, of our field strength detector **20** is to cooperate with tuner **16** in controlling the operating characteristics of the tank circuit **14**. FIG. **6** illustrates one possible embodiment where receiver circuit **10***a* uses a field strength detector **20***b* specially adapted to share with tuner **16***a* the control of the tank circuit **14**. In our Related References we have disclosed methods, and related apparatus, for dynamically tuning, via tuner **16***a*, the tank circuit **14** so as to dynamically shift the f_R of the tank circuit **14** to better match the f_C of the received RF signal at antenna **12**. By way of example, we have shown in FIG. **6** how the embodiment shown in FIG. 3 of our Related Patent may be easily modified by adding to tuner **16***a* a multiplexer **58** to facilitate shared access to the tuner control apparatus. Shown in FIG. **7** is the operational flow (similar to that illustrated in FIG. 4 in our Related Patent) of our new field strength detector **20***b* upon assuming control of tank circuit **14**.

In context of this particular use, once tuner **16***a* has completed its initial operating sequences as fully described in our Related Patent, and our field strength detector **20***b* has performed an initial sweep (as described above and illustrated in FIG. **4**) and saved in a differentiator **60** a base-line field-strength value developed in counter **38**, clock/control **44** commands multiplexer **58** to transfer control of the tank circuit **16***a* to field strength detector **20***b* (all comprising step **62** in FIG. **7**). Upon completing a second current sweep, differentiator **60** will save the then-current field-strength value developed in the counter **38** (step **64**). Thereafter, differentiator **60** will determine the polarity of the change of the previously saved field-strength value with respect to the then-current field-strength value developed in counter **38** (step **66**). If the polarity is negative (step **68**), indicating that the current field-strength value is lower than the previously-saved field-strength value, differentiator **60** will assert a change direction signal; otherwise, differentiator **60** will negate the change direction signal (step **70**). In response, the shared components in tuner **16***a* downstream of the multiplexer **58** will change the tuning characteristics of tank circuit **14** (step **72**) (as fully described in our Related References). Now, looping back (to step **64**), the resulting change of field strength, as quantized is the digital field-strength value developed in counter **38** during the next sweep (step **64**), will be detected and, if higher, will result in a further shift in the f_R of the tank circuit **14** in the selected direction or, if lower, will result in a change of direction (step **70**). Accordingly, over a number of such 'seek' cycles, our invention will selectively allow the receiver **10***a* to maximize received field strength even if, as a result of unusual factors, the f_R of the tank circuit **14** may not be precisely matched to the f_C of the received RF signal, i.e., the reactance of the antenna is closely matched with the reactance of the tank circuit, thus achieving maximum power transfer. In an alternative embodiment, it would be unnecessary for tuner **16***a* to perform an initial operating sequence as fully described in our Related Patent. Rather, field strength detector **20***b* may be used exclusively to perform both the initial tuning of the receiver circuit **10***a* as well as the subsequent field strength detection. Note that the source impedance of antenna **12** and load impedance of tank circuit **14** may be represented alternatively in schematic form as in FIG. **8**, wherein antenna **12** is represented as equivalent source resistance R_S **74** and equivalent source reactance X_S **76**, and tank circuit **14** is represented as equivalent load resistance R_L **78** and equivalent, variable load reactance X_L **80**.

In FIG. **9**, we have illustrated an alternate embodiment of our field strength detector illustrated in FIG. **3**. Here, as before, shunt circuit **18***b* is used to develop a substantially constant operating voltage level across supply node **28** and ground node **30**. Also, as before, the current reference **22** is implemented as a current mirror circuit **22***b* connected in series with shunt circuit **18***b* between nodes **28** and **30**. However, in this embodiment, the field strength current source comprises a resistive component **84** adapted to function as a static resistive pull-up device. Many possible implementations exist besides a basic resistor, such as a long channel length transistor, and those skilled in the art will appreciate the various implementations that are available to accomplish analogous functionality. The field strength voltage reference v_R developed on sensing node **36** will be drawn to a state near the supply voltage when the mirrored current flowing though transistor **34** is relatively small, e.g. close to zero amps, indicating a weak field strength. As the field strength increases, the current flowing through mirror transistor **34** will increase, and the field strength voltage reference v_R developed on sensing node **36** will drop proportionally to the mirrored current flowing through mirror transistor **34** as i_R/N. ADC **40**, having its input connected to sensing node **36**, provides a digital output indicative of the field strength reference voltage, v_R, developed on sensing node **36**, as described previously.

In this alternate embodiment, latch **42** captures the output state of ADC **40** in response to control signals provided by a clock/control circuit **44**. As disclosed earlier, the ADC **40** may comprise a comparator circuit. In this instance, ADC **40** is adapted to switch from a logic_1 state to a logic_0 when sufficient current is sunk by mirror transistor **34** to lower the voltage on sensing node **36** below a predetermined reference voltage threshold, v_{th}. Alternatively, ADC **40** may be implemented as a multi-bit ADC capable of providing higher precision regarding the specific voltage developed on sensing node **36**, depending on the requirements of the system.

Comparator **82** subsequently compares the captured output state held in latch **42** with a value held in counter **38** that is selectively controlled by clock/control circuit **44**. In response to the output generated by comparator **82**, clock/control circuit **44** may selectively change the value held in counter **38** to be one of a higher value or a lower value, depending on the algorithm employed. Depending upon the implementation of counter **38** and comparator **82**, clock/control circuit **44** may also selectively reset the value of counter **38** or comparator **82** or both. The digital field-strength value developed by counter **38** is available for any appropriate use, as discussed above.

In FIG. **10**, we have illustrated another alternate embodiment of our field strength detector illustrated in FIG. **3**. Here, as before, shunt circuit **18***c* is used to develop a substantially constant operating voltage level across supply node **28** and ground node **30**. In this embodiment, the current reference **22** is implemented as a resistive component **86** that functions as a static pull-down device. Many possible implementations exist besides a basic resistor, such as a long channel length transistor and those skilled in the art will appreciate the various implementations that are available to accomplish analogous functionality. The field strength voltage reference v_R developed on sensing node **36** will be drawn to a state near the ground node when the current flowing though shunt circuit **18***c* is relatively small, e.g., close to zero amps, indicating a weak field strength. As the field strength increase, the current flowing through shunt circuit **18***c* will increase, and the field strength voltage reference v_R developed on sensing node **36** will rise proportionally to the current flowing through shunt circuit **18***c*. ADC **40**, having its input connected to a sensing node **36**, provides a digital output indicative of the field strength reference voltage, v_R, developed on sensing node **36**, as described previously.

In this alternate embodiment, latch **42** captures the output state of ADC **40** in response to control signals provided by a clock/control circuit **44**. As disclosed earlier, the ADC **40** may comprise a comparator circuit. In this instance, ADC **40** is adapted to switch from a logic_0 state to a logic_1 when sufficient current is sourced by shunt circuit **18***c* to raise the voltage on sensing node **36** above a predetermined reference voltage threshold, v_{th}. Alternatively, ADC **40** may be implemented as a multi-bit ADC capable of providing higher precision regarding the specific voltage developed on sensing node **36**, depending on the requirements of the system.

Comparator **82** subsequently compares the captured output state held in latch **42** with a value held in counter **38** that is selectively controlled by clock/control circuit **44**. In response to the output generated by comparator **82**, clock/control circuit **44** may selectively change the value held in counter **38** to be one of a higher value or a lower value, depending on the algorithm employed. Depending upon the implementation of counter **38** and comparator **82**, clock/control circuit **44** may also selectively reset the value of counter **38** or comparator **82** or both. The digital field-strength value developed by counter **38** is available for any appropriate use, as discussed above.

In another embodiment, our invention may be adapted to sense the environment to which a tag is exposed, as well as sensing changes to that same environment. As disclosed in our Related References, the auto-tuning capability of tuner **16** acting in conjunction with tank circuit **14** detects antenna impedance changes. These impedance changes may be a function of environmental factors such as proximity to interfering substances, e.g., metals or liquids, as well as a function of a reader or receiver antenna orientation. Likewise, as disclosed herein, our field strength (i.e., received power) detector **20** may be used to detect changes in received power (i.e., field strength) as a function of, for example, power emitted by the reader, distance between tag and reader, physical characteristics of materials or elements in the immediate vicinity of the tag and reader, or the like.

Sensing the environment or, at least, changes to the environment is accomplished using one or both of these capabilities.

As an example, the tag **88** of FIG. **11**, contains both a source tag antenna **12** (not shown, but see, e.g., FIG. **6**) and a corresponding load chip tank circuit **14** (not shown, but see, e.g., FIG. **6**). Each contains both resistive and reactive elements as discussed previously (see, e.g., FIG. **8**). A tag **88** containing such a tank circuit **14** mounted on a metallic surface will exhibit antenna impedance that is dramatically different than the same tag **88** in free space or mounted on a container of liquid. Table 1 displays exemplary values for impedance variations in both antenna source resistance **74** as well as antenna source reactance **76** as a function of frequency as well as environmental effects at an exemplary frequency:

TABLE 1

Antenna Impedance Variations			
In Free Air	860 MHz	910 MHz	960 MHz
R_S	1.9	2.5	3.7
X_S	124	136	149
@910 MHz	Free Air	On Water	On Metal
R_S	2.5	26	1.9
X_S	136	136	27

The tuner circuit **16** of our invention as disclosed in the Related References automatically adjusts the load impendence by adjusting load reactance **80** (see, e.g., FIG. **8**) to match source antenna impedance represented by source resistance **74** (see, e.g., FIG. **8**) and source reactance **76** (see, e.g., FIG. **8**). As previously disclosed, matching of the chip load impedance and antenna source impedance can be performed automatically in order to achieve maximum power transfer between the antenna and the chip. My invention as disclosed in the Related References contained a digital shift register **90** for selectively changing the value of the load reactive component, in the present case a variable capacitor, until power transfer is maximized. This digital value of the matched impendence may be used either internally by the tag **88**, or read and used by the reader **92**, to discern relative environmental information to which the tag **88** is exposed. For example, tag **88** may contain a calibrated look-up-table within the clock/control circuit **44** which may be accessed to determine the relevant environmental information. Likewise, a RFID reader **92** may issue commands (see transaction **1** in FIG. **11**) to retrieve (see transaction **2** in FIG. **11**) the values contained in digital shift register **90** via conventional means, and use that retrieved information to evaluate the environment to which tag **88** is exposed. The evaluation could be as simple as referencing fixed data in memory that has already been stored and calibrated, or as complex as a software application running on the reader or its connected systems for performing interpretive evaluations.

Likewise, consider a tag **88** containing our field strength (i.e., received power) detector **20** (not shown, but, e.g., see FIG. **6**) wherein the method of operation of the system containing the tag **88** calls for our field strength detector **20** to selectively perform its sweep function and developing the quantized digital representation of the current via the method discussed earlier. As illustrated in FIG. **11**, counter **38** will contain the digital representation developed by our field strength detector **20** of the RF signal induced current, and may be used either internally by the tag **88**, or read and used by the reader **92**, to discern relative environmental information to which the tag **88** is exposed. For example, reader **92** may issue a command to the tag **88** (see transaction **1** in FIG. **11**) to activate tuner **16** and/or detector **20** and, subsequent to the respective operations of tuner **16** and/or detector **20**, receive (see transaction **2** in FIG. **11**) the digital representations of either the matched impedance or the maximum current developed during those operations. Once again, this digital value of the field strength stored in the counter **38** may be used either internally by the tag **88**, or read and used by the reader **92**, to discern relative environmental information to which the tag **88** is exposed. For example, tag **88** may contain a calibrated look-up-table within the clock and control block **44** which may be accessed to determine the relevant environmental information. Likewise, a RFID reader may issue commands to retrieve the values contained in digital shift register **90**, and use that retrieved information to evaluate the environment to which tag **88** is exposed. The evaluation could be as simple as referencing fixed data in memory that has already been stored and calibrated, or as complex as a software application running on the reader or its connected systems for performing interpretive evaluations. Thus, the combining of the technologies enables a user to sense the environment to which a tag **88** is exposed as well as sense changes to that same environment.

Thus, it is apparent that we have provided an effective and efficient method and apparatus for quantizing the received RF field strength as a function of induced current. We have developed this field quantization in a form and manner that is suitable for selectively varying the impedance of the tank circuit to maximize received power, especially during normal system operation. Those skilled in the art will recognize that modifications and variations can be made without departing from the spirit of our invention. Therefore, we intend that our invention encompass all such variations and modifications as fall within the scope of the appended claims.

What is claimed is:

1. A radio frequency (RF) circuit comprises:

an antenna operably coupled to receive an RF signal;

a power harvesting circuit operable to generate a first power source from the RF signal;

a tank circuit coupled to the antenna, wherein the tank circuit includes a selectively variable impedance; and

a tuning circuit operable to dynamically vary the selectively variable impedance of the tank circuit based on resonance of the RF circuit and frequency of the RF signal to substantially align the resonance of the RF circuit with the frequency of the RF signal, wherein, when the resonance of the RF circuit is substantially aligned with the frequency of the RF signal, the power harvesting circuit generates a second power source from the RF signal, wherein the second power source is greater than the first power source.

2. The RF circuit of claim **1** further comprises:

a processing module operable to generate a digital value representative of dynamic variance of the selectively variable impedance of the tank circuit;

memory operable coupled to the processing module; and

a transmitter section operable to transmit an outbound RF signal to a reader, wherein the outbound RF signal includes the digital value.

3. The RF circuit of claim **2** further comprises:

the antenna operable to receive a second RF signal, wherein the second RF signal has an adjusted transmit power with respect to the RF signal based on an interpretation of the digital value such that the second power source is at least one of a desired voltage level and a desired current level.

4. The RF circuit of claim **1**, wherein the tank circuit comprises:

an inductor; and

a variable capacitor coupled to the inductor, wherein the tuning circuit is operable to adjust capacitance of the variable capacitor such that a resonant frequency of the tank circuit substantially matches a frequency of the RF signal.

5. The RF circuit of claim **1**, wherein the power harvesting circuit comprises:

a regulator coupled to the tank circuit, wherein a current is induced in the tank circuit from the received RF signal, wherein the regulator is operable to extract a portion of the induced current to provide direct current (DC) operating power to the RF circuit.

6. The RF circuit of claim **1** further comprises:

a sensing circuit that includes:

the tank circuit having the selectively variable impedance;

the tuning circuit operable to dynamically vary the impedance of the tank circuit, and to develop a first quantized value representative of the impedance of the tank circuit; and

a detector circuit operable to develop a second quantized value as a function of a field strength of a received RF signal; and

memory, wherein the memory stores the first and second quantized values; and

a control circuit operable selectively to retrieve the first and second quantized values from the memory, wherein the retrieved first and second quantized values are used to sense changes to an environment to which the RF circuit is exposed.

7. The RF circuit of claim **6**, wherein the environment includes one or more of:

free space;

air;

liquids; and

metals.

8. The RF circuit of claim **1**, wherein the selectively variable impedance is set to an initial value.

* * * * *

US009971162B2

(12) **United States Patent**
Banerjee et al.

(10) **Patent No.: US 9,971,162 B2**
(45) **Date of Patent: May 15, 2018**

(54) **APPARATUSES AND METHODS FOR MAKING AN OBJECT APPEAR TRANSPARENT**

(71) Applicant: **Toyota Motor Engineering & Manufacturing North America, Inc.**, Erlanger, KY (US)

(72) Inventors: **Debasish Banerjee**, Ann Arbor, MI (US); **Hideo Iizuka**, Nissin (JP)

(73) Assignee: **Toyota Motor Engineering & Manufacturing North America, Inc.**, Plano, TX (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 74 days.

(21) Appl. No.: **15/185,988**

(22) Filed: **Jun. 17, 2016**

(65) **Prior Publication Data**

US 2017/0227781 A1 Aug. 10, 2017

Related U.S. Application Data

(60) Provisional application No. 62/291,920, filed on Feb. 5, 2016.

(51) **Int. Cl.**
G02B 27/14 (2006.01)
G02B 27/10 (2006.01)
(Continued)

(52) **U.S. Cl.**
CPC ***G02B 27/14*** (2013.01); ***B60R 13/025*** (2013.01); ***G02B 5/3083*** (2013.01);
(Continued)

(58) **Field of Classification Search**
CPC G02B 5/30; G02B 5/3025; G02B 5/3083; G02B 21/14; G02B 27/024;
(Continued)

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,684,255 A * 8/1987 Ford G01J 3/453
356/455
5,941,596 A 8/1999 See
(Continued)

OTHER PUBLICATIONS

Hongsheng Chen et al., "Broadband polygonal invisibility cloak for visible light", Scientific Reports, Feb. 9, 2012, pp. 1-14.
(Continued)

Primary Examiner — Loha Ben
(74) *Attorney, Agent, or Firm* — Dinsmore & Shohl LLP

(57) **ABSTRACT**

A cloaking device includes cloaking region boundary planes oriented non-planar to each other, each of the cloaking region boundary planes having an outward facing mirror surface and an inward facing opaque surface. The cloaking device includes a cloaking region bounded at least partially by the inward facing opaque surfaces of the cloaking region boundary planes. Half mirrors are spaced apart from and generally parallel to the outward facing mirror surfaces such that a half mirror is spaced apart from and generally parallel to each outward facing mirror surface. Light from an object on an object-side of the cloaking device is directed around an article within the cloaking region and forms an image on an image-side of the cloaking device such the article appears transparent to an observer looking towards the object.

20 Claims, 9 Drawing Sheets

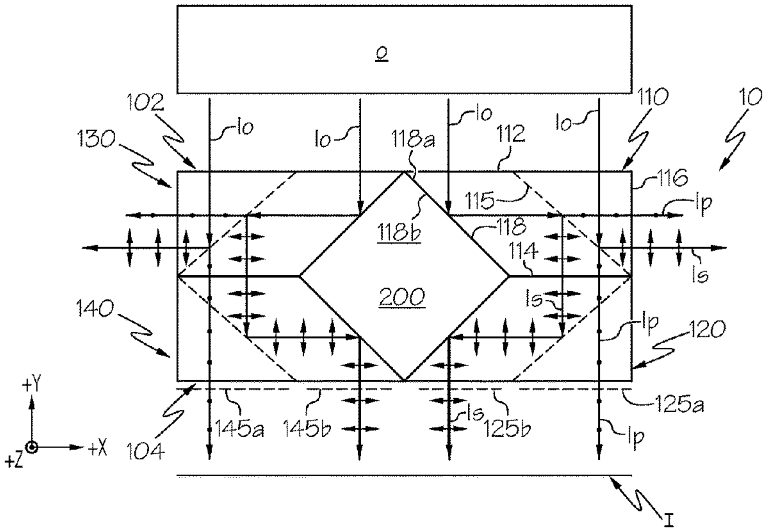

(51) **Int. Cl.**

G02B 27/28	(2006.01)
G02B 5/30	(2006.01)
B60R 13/02	(2006.01)
B60R 1/08	(2006.01)

(52) **U.S. Cl.**

CPC ***G02B 27/1086*** (2013.01); ***G02B 27/283*** (2013.01); *B60R 1/082* (2013.01); *B60R 2013/0287* (2013.01); *G02B 5/30* (2013.01); *G02B 5/3025* (2013.01); *G02B 27/144* (2013.01)

(58) **Field of Classification Search**

CPC .. G02B 27/1086; G02B 27/14; G02B 27/141; G02B 27/143; G02B 27/144; B60R 1/08; B60R 1/082; B60R 13/02; B60R 13/025; B60R 2013/0287; G03B 21/2066; G03B 21/2073

USPC 359/299–304, 493, 618, 629, 634; 396/331, 386, 544; 353/12, 14, 20, 28, 353/66

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,982,343	A	*	11/1999	Iba	G02B 27/0172 345/8
6,130,784	A	*	10/2000	Takahashi	H04N 5/7491 348/E5.145
2015/0248013	A1		9/2015	Chen et al.	

OTHER PUBLICATIONS

John C. Howell et al., "Simple, broadband, optical spatial cloaking of very large objects", Department of Physics and Astronomy, University of Rochester, www.physics.optics, arXiv:1306.0863 </abs/1306.0863> [physics.optics], Jun. 2013.

Yihao Yang et al., "Towards omnidirectional, large scale, full polarization, and broadband practical invisibility cloaks: challenges and progress", EPJ Appl. Metamat. 2014, 1, 7.

"Physics demonstrations: cloaking device?", Apr. 25, 2013; URL: https://skullsinthestars.com/2013/04/25/physics-demonstrations-cloaking-device.

* cited by examiner

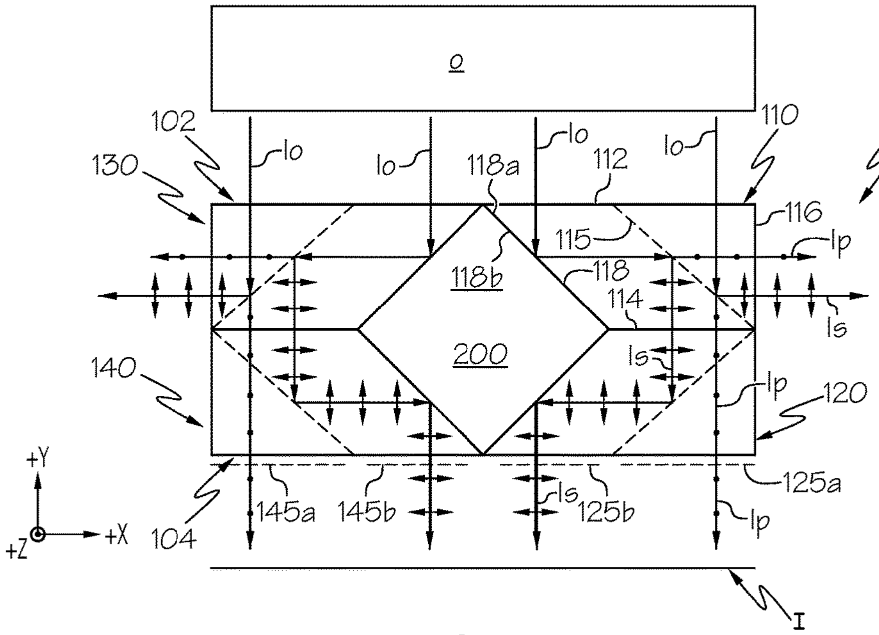
o
102
Io
Io
118a
Io
112
Io
110
10
130
116
115
Ip
118b
118
114
Is
200
Is
140
Ip
120
125a
+Y
145a
145b
Is
125b
Ip
+Z
+X
104
I

FIG. 1

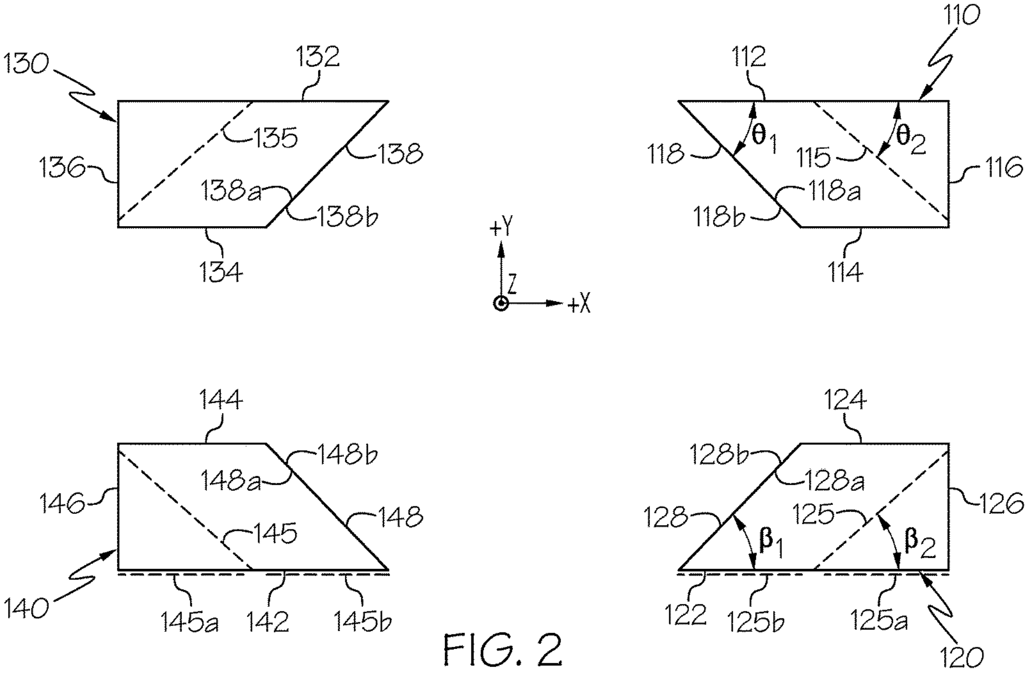
θ1
θ2
+Y
Z
+X
β1
β2

FIG. 2

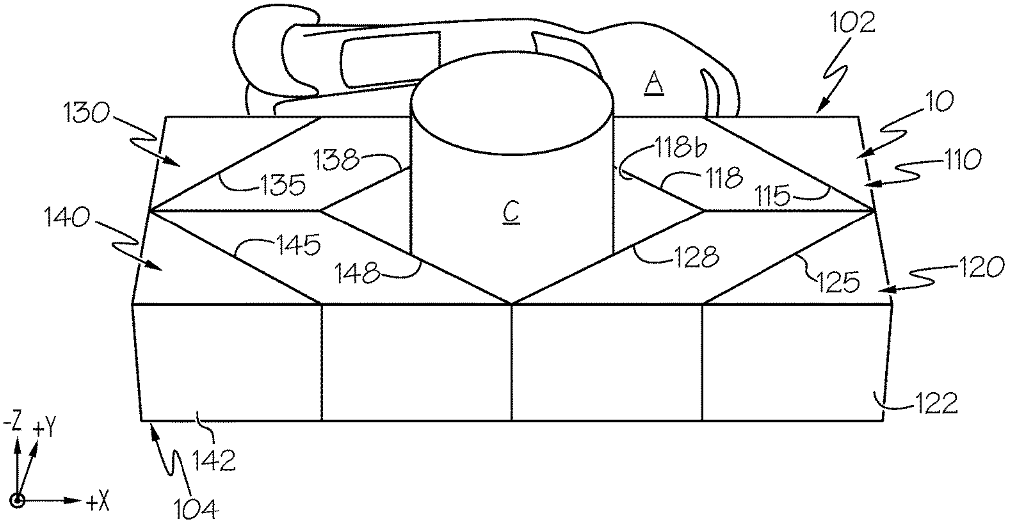
A
C
-Z
+Y
+X

FIG. 3

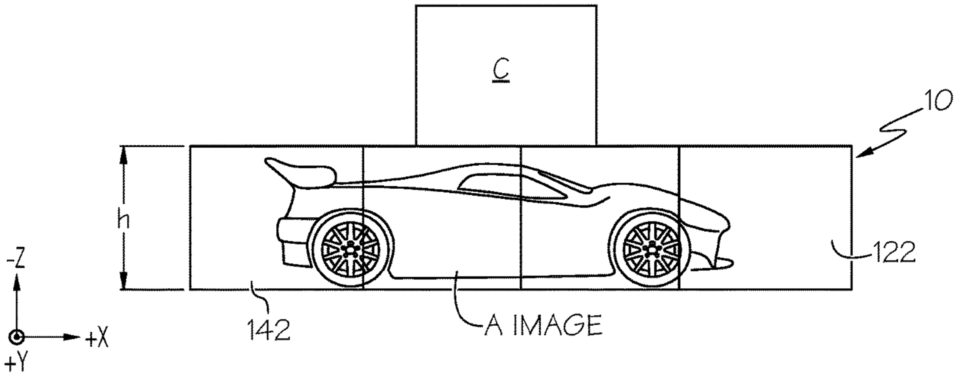
C
10
h
-Z
122
142
A IMAGE
+X
+Y

FIG. 4

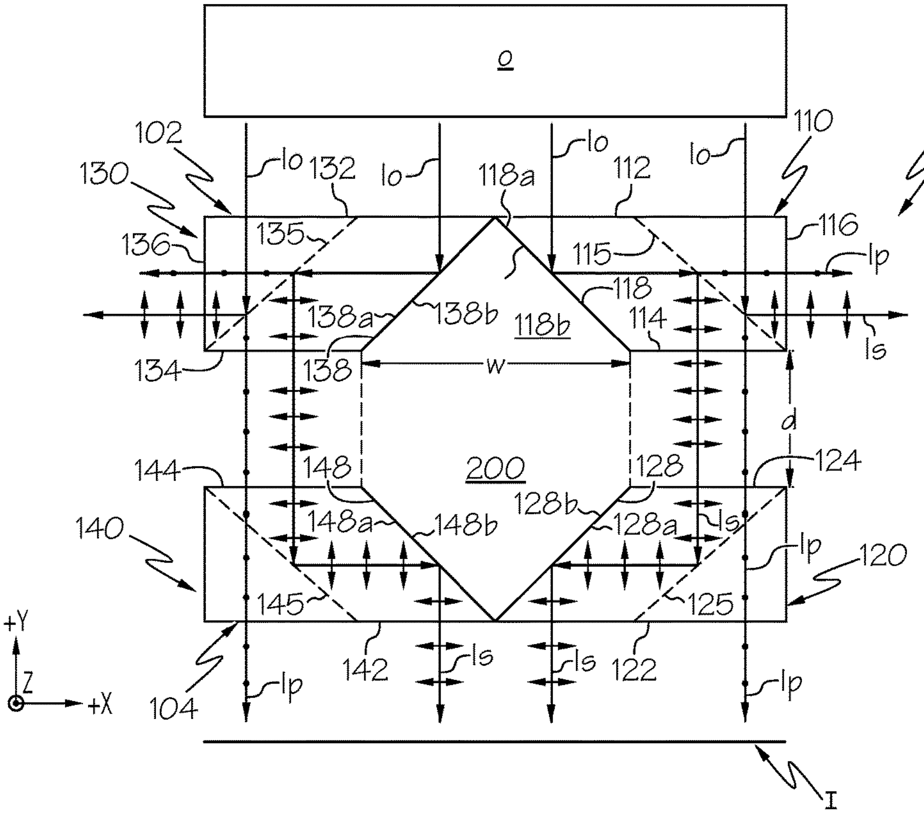
o
Io
Io
Io
Io
Ip
Is
w
d
Is
Ip
+Y
Z
+X
Is
Is
Ip
Ip
I

FIG. 5

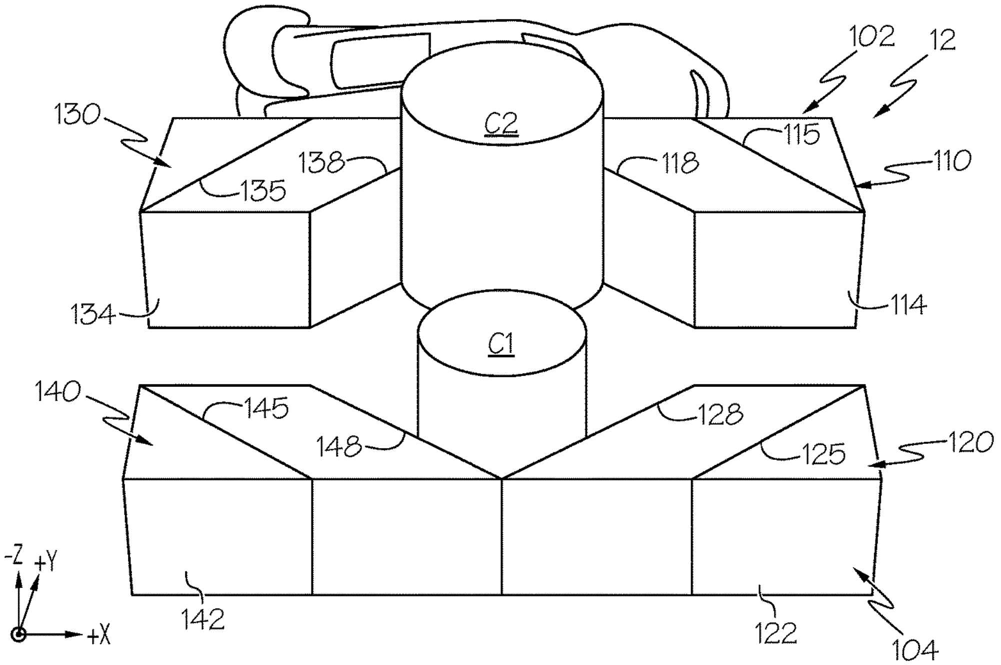
C2
C1
-Z
+Y
+X

FIG. 6

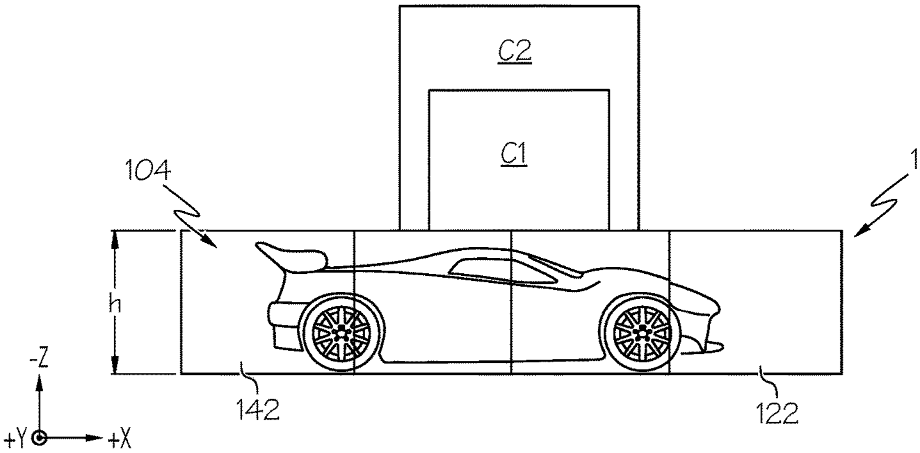
C2
C1
104
12
h
-Z
+Y
+X
142
122

FIG. 7

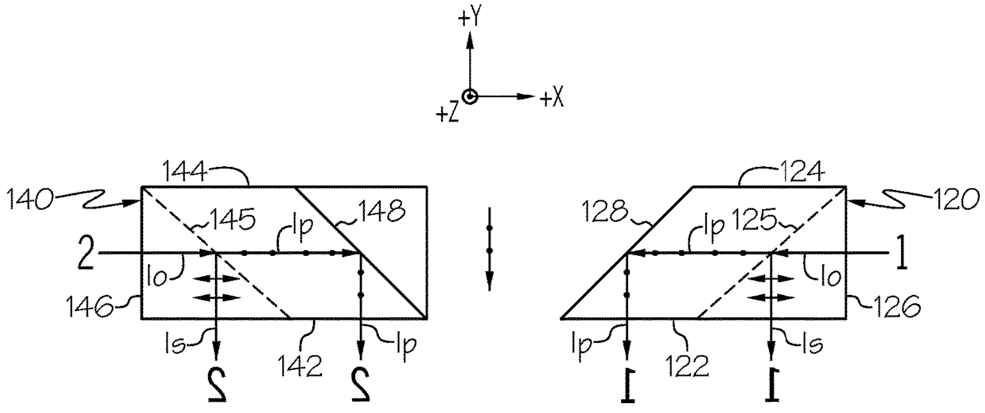
+Y
+Z
+X
140
144
145
Ip
148
2
Io
146
Is
142
Ip
2
2
128
124
Ip
125
120
1
Io
126
Ip
122
Is
1
1

FIG. 8

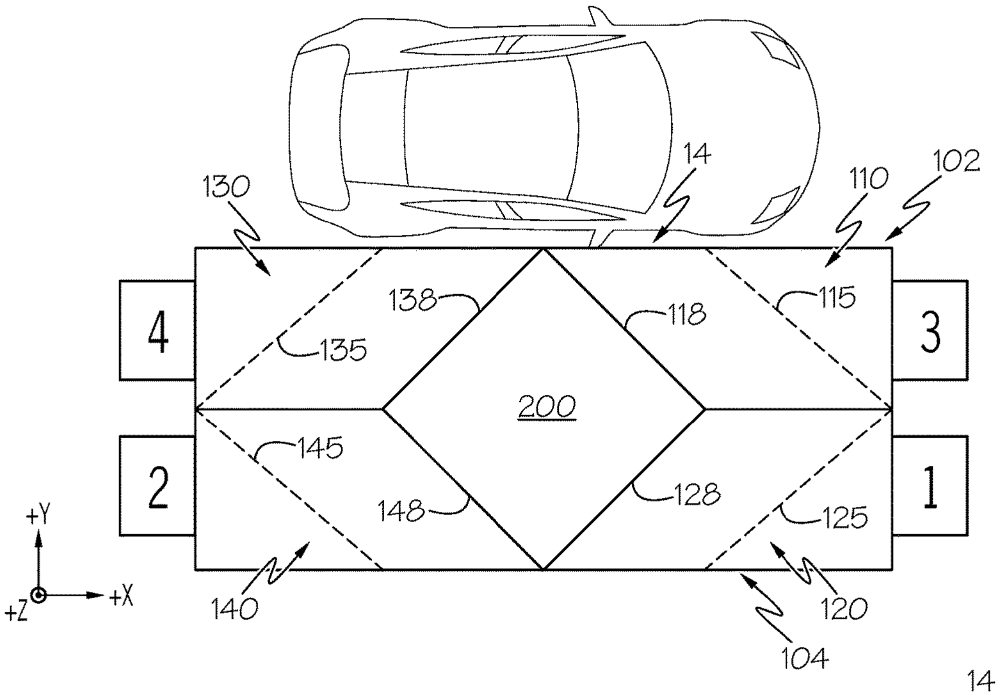
130
14
110
102
4
138
135
118
115
3
200
2
145
148
128
125
1
+Y
+Z
+X
140
120
104

FIG. 9

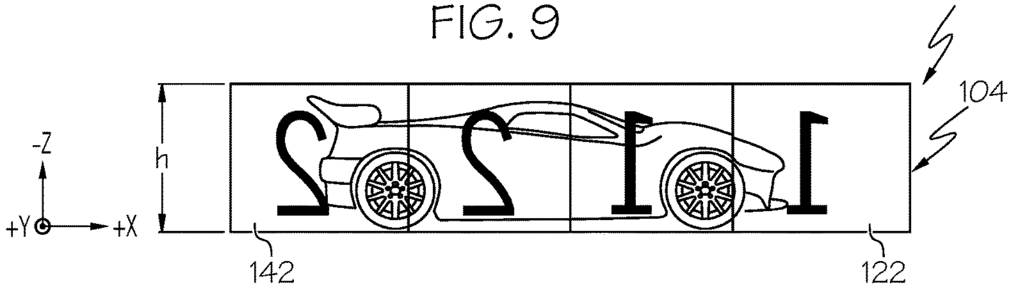
14
104
-Z
+Y
+X
h
2
2
1
1
142
122

FIG. 10

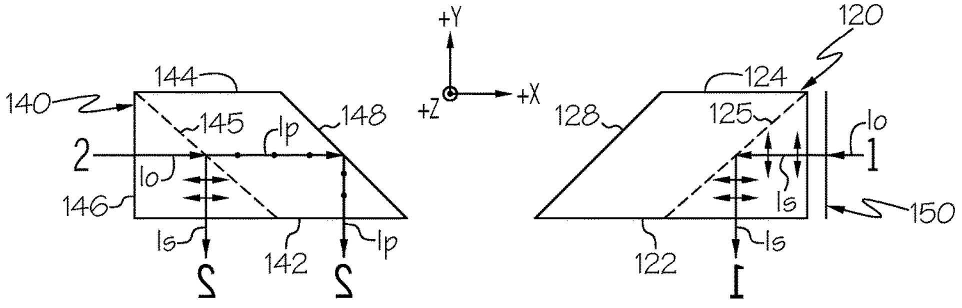
+Y
+Z
+X
120
140
144
124
145
148
128
125
lp
lo
2
1
lo
146
ls
150
ls
lp
142
122
ls
2
2
1

FIG. 11

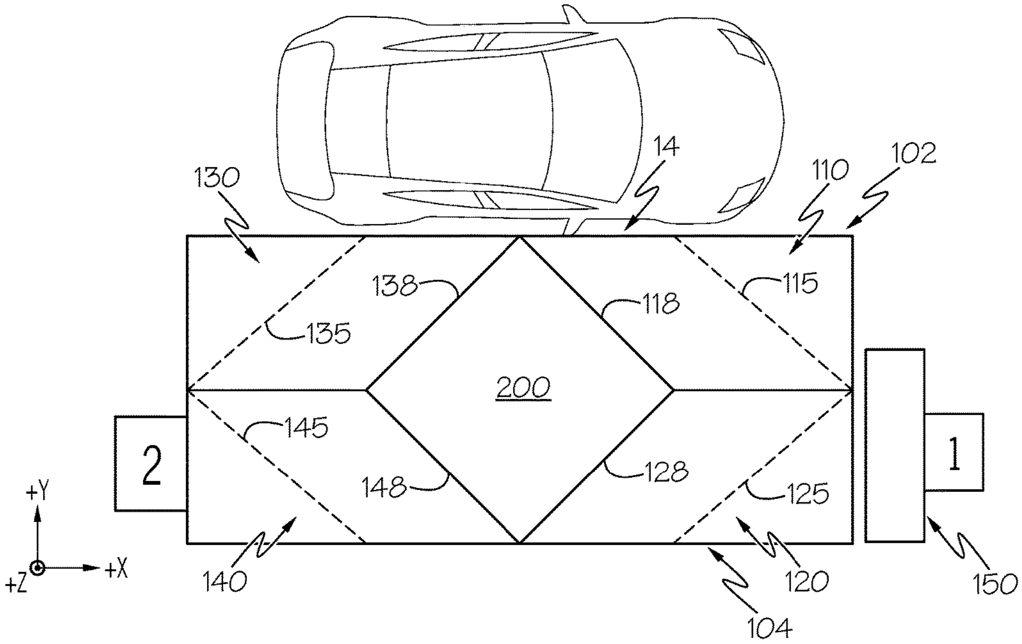
14
130
110
102
138
118
115
135
200
145
2
148
128
125
1
+Y
+Z
+X
140
120
150
104

FIG. 12

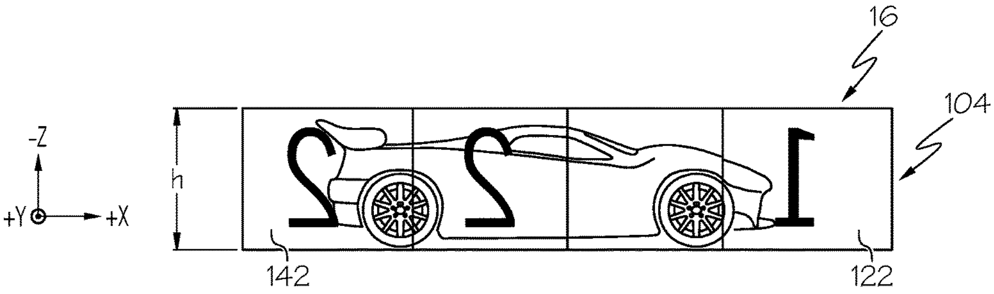
16
104
-Z
h
+Y
+X
2
2
1
142
122

FIG. 13

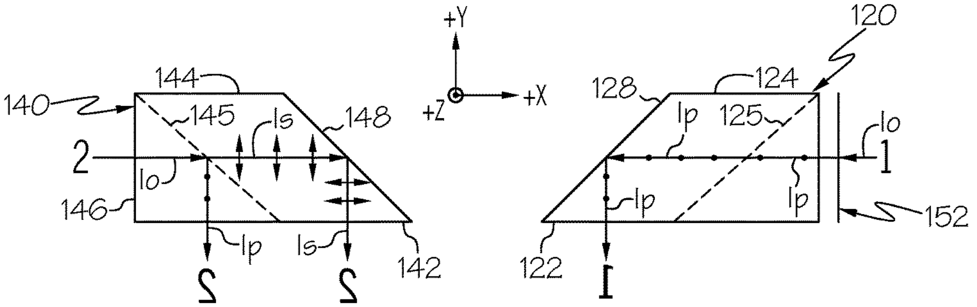
+Y
+X
+Z
140
144
145
Is
148
2
Io
146
Ip
Is
2
2
142
120
128
124
Ip
125
Io
1
Ip
Ip
152
122
1

FIG. 14

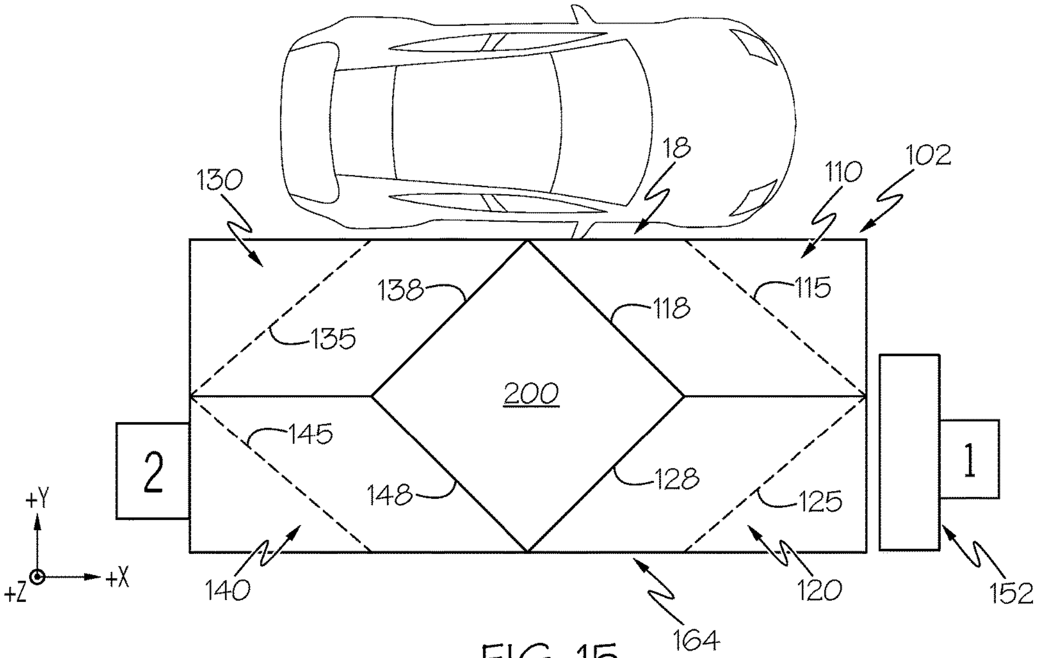
18
130
110
102
138
118
115
135
200
2
1
145
148
128
125
+Y
+Z
+X
140
164
120
152

FIG. 15

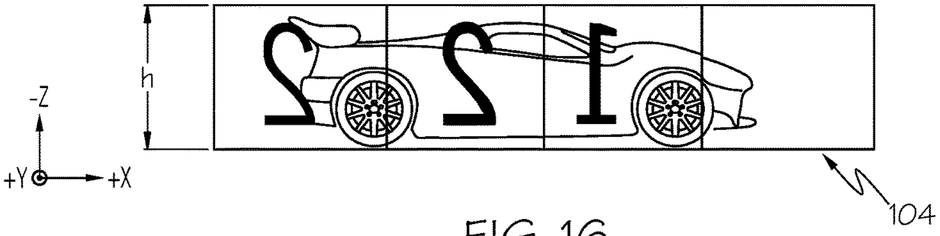
h
-Z
+Y
+X
2
2
1
104

FIG. 16

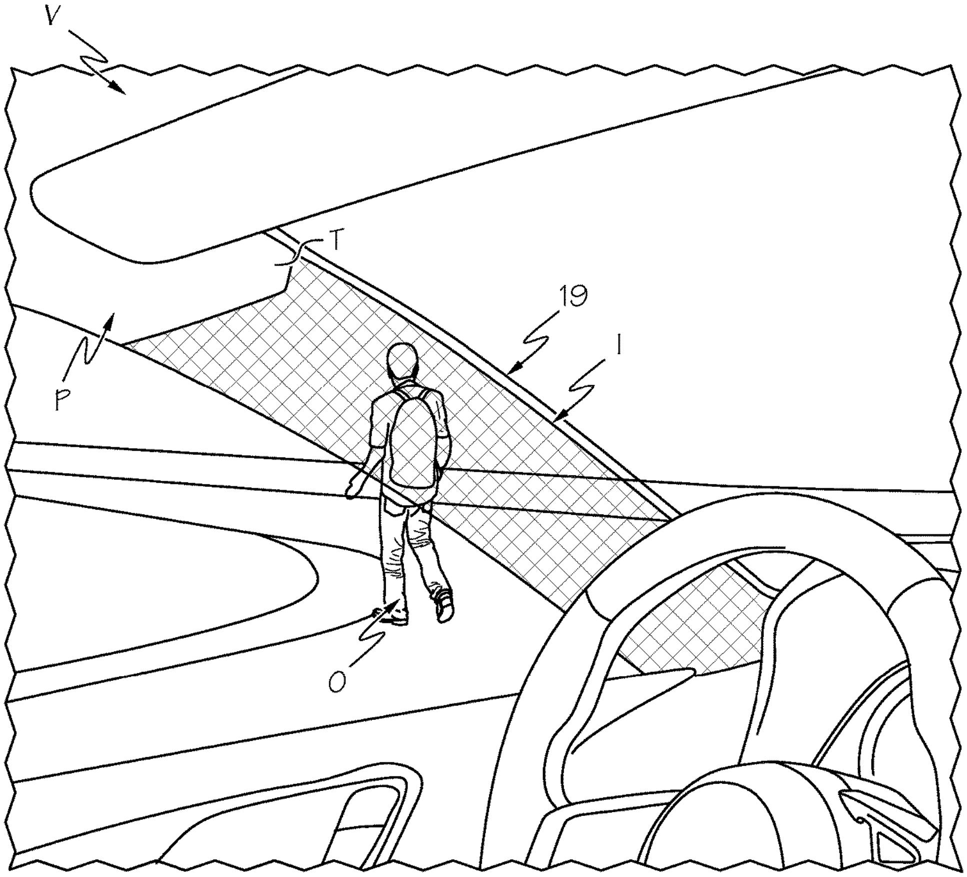

FIG. 17

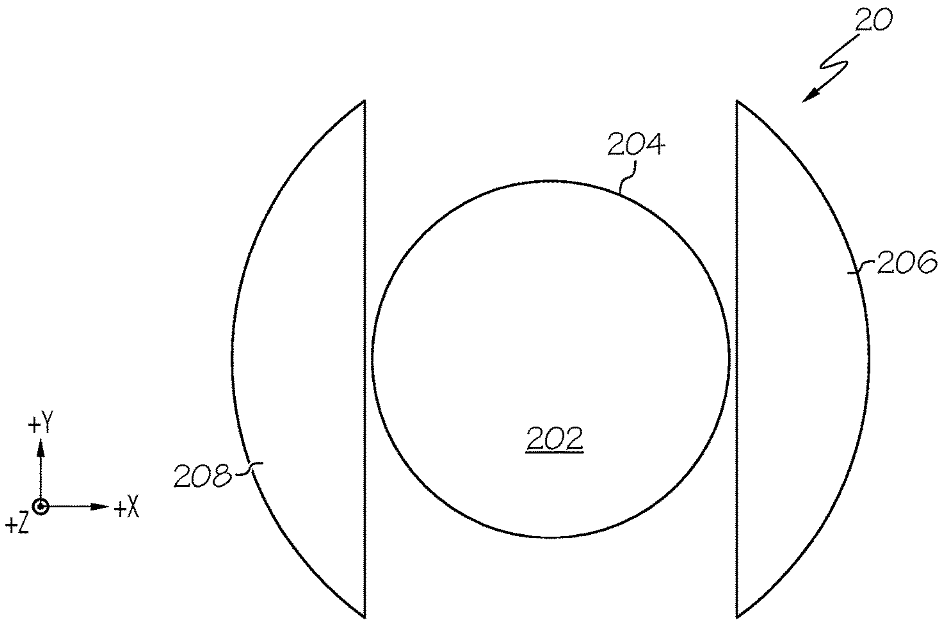
20
204
206
202
208
+Y
+X
+Z

FIG. 18

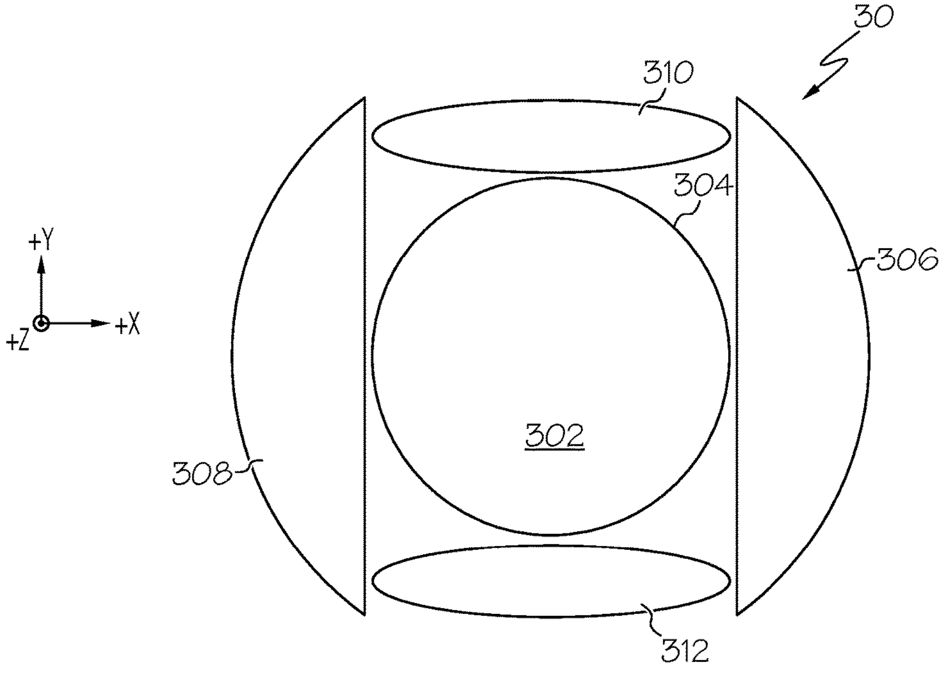
30
310
304
306
302
308
312
+Y
+X
+Z

FIG. 19

APPARATUSES AND METHODS FOR MAKING AN OBJECT APPEAR TRANSPARENT

RELATED APPLICATION

This application claims the benefit of U.S. Provisional Patent Application Ser. No. 62/291,920 filed Feb. 5, 2016, the entirety of which is incorporated herein by reference.

TECHNICAL FIELD

The present specification generally relates to apparatuses and methods for making an object appear transparent and, more specifically, to cloaking devices for pillars of vehicles and methods for making pillars of vehicles appear transparent.

BACKGROUND

Studies on cloaking devices that appear to make a pillar of a vehicle transparent have been published. Such studies disclose the use of metamaterials or the use of video cameras in combination with a display screen to allow an occupant of a vehicle to ostensibly "see" though the vehicle pillar, thereby reducing blind spots in the vehicle. However, metamaterials and video technology use complicated material designs and equipment.

Accordingly, a need exists for alternative devices that appear to make a pillar of a vehicle transparent.

SUMMARY

In one embodiment, a cloaking device includes cloaking region (CR) boundary planes oriented non-planar to each other and each of the CR boundary planes has an outward facing mirror surface and an inward facing opaque surface. The cloaking device includes a CR at least partially bounded by the inward facing opaque surfaces of the CR boundary planes. The cloaking device includes half mirrors spaced apart from and generally parallel to the outward facing mirror surfaces such that a half mirror is spaced apart from and generally parallel to each outward facing mirror surface. An article positioned within the CR is not visible outside of the CR and light reflected from an object on one side of the cloaking device is redirected around the CR and the article positioned within the CR, and forms an image of the object on another side of the cloaking device.

According to another embodiment, a cloaking device for cloaking an article of a vehicle includes a cloaking assembly with an object-side, an image-side and at least two CR boundary planes positioned at least partially between the object-side and the image-side. The at least two CR boundary planes are positioned non-planar to each other and each of the at least two CR boundary planes have an outward facing mirror surface and an inward facing opaque surface. A CR is bounded by the inward facing opaque surfaces of the at least two CR boundary planes and a vehicle article positioned within the CR is not visible from outside the CR. The cloaking assembly has at least two half mirrors positioned at least partially between the object-side and the image-side. Each of the at least two half mirrors is spaced apart from and generally parallel to one of the outward facing mirror surfaces such that a half mirror is spaced apart from and generally parallel to each outward facing mirror surface. Light from an object located on the object-side of the cloaking assembly, and obscured by the vehicle article

positioned within the CR is redirected around the CR, and forms an image of the object on the image-side of the cloaking assembly such that the vehicle article appears transparent to an observer looking in a direction towards the object.

In embodiments, the cloaking assembly has four CR boundary planes and four half mirrors. The four CR boundary planes are oriented generally orthogonal to each other, the CR is bounded by the four CR boundary planes, and each of the half mirrors are spaced apart and generally parallel to one of the outward facing mirror surfaces such that a half mirror is spaced apart and generally parallel to each outward facing mirror surface. Each of the half mirrors may be selected from a p-polarization half mirror or an s-polarization half mirror. Light reflected from the object located on the object-side of the cloaking assembly passes through the cloaking assembly by reflecting off of a first outward facing mirror surface, a first half mirror oriented generally parallel to the first outward facing mirror surface, a second half mirror oriented generally orthogonal to the first half mirror and a second outward facing mirror surface oriented generally parallel to the second half mirror. Light that passes through the cloaking assembly forms an image of the object on the image-side of the cloaking assembly.

In embodiments, the cloaking device with the cloaking assembly has an end surface oriented non-planar to a half mirror. Light from an item located adjacent to the end surface passes through the cloaking assembly and forms an image of the item on the image side of the cloaking assembly.

These and additional features provided by the embodiments described herein will be more fully understood in view of the following detailed description in conjunction with the drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

The embodiments set forth in the drawings are illustrative and exemplary in nature and not intended to limit the subject matter defined by the claims. The following detailed description of the illustrative embodiments can be understood when read in conjunction with the following drawings, where like structure is indicated with like reference numerals and in which:

FIG. **1** schematically depicts a top view of a cloaking device according to one or more embodiments described and illustrated herein;

FIG. **2** schematically depicts top isolated views of four cloaking device components of the cloaking device of FIG. **1** according to one or more embodiments described and illustrated herein;

FIG. **3** schematically depicts a top perspective view of the cloaking device of FIG. **1** with a first object on one side of the cloaking device and a second object within a cloaking region (CR) of the cloaking device;

FIG. **4** schematically depicts a side view of the cloaking device of FIG. **3** with the first object on one side of the cloaking device and the second object within the CR of the cloaking device;

FIG. **5** schematically depicts a top view of a cloaking device according to one or more embodiments described and illustrated herein;

FIG. **6** schematically depicts a top perspective view of the cloaking device of FIG. **5** with a first object on one side of the cloaking device and a second object and a third object within a CR of the cloaking device;

FIG. **7** schematically depicts a side view of the cloaking device of FIG. **6** with the first object on one side of the cloaking device and the second object and the third object within the CR of the cloaking device;

FIG. **8** schematically depicts a top isolated view of two cloaking device components according to one or more embodiments described and illustrated herein;

FIG. **9** schematically depicts a top view of a cloaking device with four cloaking device components according to one or more embodiments described and illustrated herein;

FIG. **10** schematically depicts a side view of the cloaking device shown in FIG. **9**;

FIG. **11** schematically depicts a top isolated view of two cloaking device components according to one or more embodiments described and illustrated herein;

FIG. **12** schematically depicts a top view of a cloaking device with four cloaking device components according to one or more embodiments described and illustrated herein;

FIG. **13** schematically depicts a side view of the cloaking device of FIG. **12**;

FIG. **14** schematically depicts a top isolated view of two cloaking device components according to one or more embodiments described and illustrated herein;

FIG. **15** schematically depicts a top view of four cloaking device components according to one or more embodiments described and illustrated herein;

FIG. **16** schematically depicts a side view of the cloaking device of FIG. **15**;

FIG. **17** schematically depicts a side view of a cloaking device cloaking an A-pillar of a vehicle according to one or more embodiments described and illustrated herein;

FIG. **18** schematically depicts a top view of a cloaking device according to one or more embodiments described and illustrated herein; and

FIG. **19** schematically depicts a top view of a cloaking device with four cloaking device components according to one or more embodiments described and illustrated herein.

DETAILED DESCRIPTION

A cloaking device is provided. The cloaking device may have an object-side, an image-side and a cloaking region (CR) boundary plane having an outward facing mirror surface and an inward facing opaque surface. A CR is at least partially bounded by the inward facing opaque surface of the CR boundary plane and a half mirror is spaced apart from and generally parallel to the outward facing mirror surface. The half mirror may be a p-polarization half mirror or an s-polarization half mirror. In embodiments, an article, e.g. an A-pillar of a vehicle, a B-pillar of a vehicle or a C-pillar of a vehicle, is positioned within the CR and light from an object located on the object-side of the cloaking device and obscured by the CR is redirected around the CR and the article and forms the image of the object on the image-side of the cloaking device such that the article appears transparent. Various embodiments of cloaking devices and methods for using the same will be described in further detail herein with specific reference to the appended drawings.

FIG. **1** generally depicts one embodiment of a cloaking device. The cloaking device includes a cloaking region (CR) that is at least partially bounded by at least two CR boundary planes that are non-planar to each other. Each of the two CR boundary planes has an outward facing mirror surface and an inward facing opaque surface. Spaced apart from and oriented generally parallel to each of the CR boundary planes is a half mirror (e.g., a polarizing half mirror) that allows p-polarized light to pass through and s-polarized light to be reflected or, alternatively, allows s-polarized light to pass through and p-polarized light to be reflected. Referring to the coordinate system in the figure, portions of a target object are located directly behind the CR in the +Y direction (obscured portions of the object which are otherwise not visible through the CR) and portions of the target object are located outside or beyond the CR in the +Y direction (visible portions of the object). Light rays (light) from the obscured portions of the target object travel in the −Y direction and are incident on a first outward facing mirror surface. The first outward facing mirror surface reflects the incident light from the obscured portions of the target object in the +X direction to a first half mirror that is spaced apart from and oriented generally parallel to the first outward facing mirror surface. One mode of polarized light (e.g., p-polarized light) from the light reflected from the first outward facing mirror surface passes through the first half mirror (continues traveling in the +X direction), while another mode of polarized light (e.g. s-polarized light) is reflected by the first half mirror in the −Y direction. The polarized light reflected in the −Y direction by the first half mirror is reflected into the −X direction by a second half mirror. The polarized light reflected into the −X direction by the second half mirror is reflected in the −Y direction by a second outward facing mirror surface that is spaced apart from and oriented generally parallel to the second half mirror. An observer looking at the cloaking device in the +Y direction will see the reflection of the polarized light by the second outward facing mirror surface, i.e. polarized light originating from light from obscured portions of the target object will be visible to a human eye located on an opposite side of the CR from the obscured portions of the target object. Also, light reflected from an article positioned within the CR (i.e., the cloaked article) is not transmitted through the inward facing opaque surfaces and is not seen by the observer looking at the cloaking device in the +Y direction. That is, the cloaking device effectively redirects light from the obscured portions of the target object around the CR such that the light from the target object appears to pass through the cloaked article giving the visual impression that the cloaked article is not present. Light from the visible portions of the target object travels in the −Y direction and is incident on the first half mirror. One mode of the incident light (e.g., s-polarized light) is reflected in the +X direction by the first half mirror and another mode of the incident light (e.g. p-polarized light) is transmitted through the first half mirror and continues traveling in the −Y direction. The transmitted polarized light traveling in the −Y direction reaches the second half mirror and is transmitted through the second half mirror. An observer looking at the cloaking device in the +Y direction will see the transmitted polarized light traveling in the −Y direction, i.e. polarized light originating from light from visible portions of the target object will be visible to a human eye located on an opposite side of the cloaking device from the visible portions of the target object. Accordingly, an individual will see the entire target object, both obscured portions and visible portions, that is located on the opposite side of the CR (and thus on the opposite side of a cloaked article) giving the visual impression that the cloaked article is transparent.

Referring now to FIGS. **1-2**, embodiments of a cloaking device include a cloaking assembly **10** with four CR boundary planes **118**, **128**, **138**, **148**. The CR boundary planes **118**, **128**, **138**, **148** located adjacent to each other are non-planar to each other. In embodiments, opposing CR boundary planes **118**, **148** and **128**, **138** may be oriented parallel to each other and adjacent CR boundary planes **118**, **128**, **138**,

148 may be oriented orthogonal to adjacent CR boundary planes, e.g. CR boundary plane **118** is oriented orthogonal to adjacent CR boundary planes **138**, **128**, CR boundary plane **128** is oriented orthogonal to adjacent CR boundary planes **118**, **148**, etc. Each of the four CR boundary planes has an outward facing mirror surface **118***a*, **128***a*, **138***a*, **148***a*, respectively, and an inward facing opaque surface **118***b*, **128***b*, **138***b*, **148***b*, respectively. The outward facing mirror surfaces **118***a*, **128***a*, **138***a*, **148***a*, can be made from omni-directional photonic crystals or mirrors such that approximately 100% (+/−10%) of light incident on the outward facing mirror surfaces is reflected from the outward facing mirror surfaces. Accordingly, the term "mirror surface" used herein refers to a surface that reflects approximately 100% (+/−10%) of all modes of light (e.g. s-polarized light and p-polarized light) incident on the mirror surface. In embodiments, the CR boundary planes **118**, **128**, **138**, **148** form a CR **200** bound at least partly by the inward facing opaque surfaces **118***b*, **128***b*, **138***b*, **148***b*. The four CR boundary planes **118**, **128**, **138** **148**, have a height 'h' (FIG. **4**) in the Z direction of the coordinate axes in the figures and light reflected or transmitted within the CR **200** does not pass through the inward facing opaque surfaces **118***b*, **128***b*, **138***b*, **148***b*. Accordingly, an article (e.g., a cloaked article) located within the CR **200** is not visible to an observer viewing the cloaking assembly **10** in the +Y direction.

Still referring to FIGS. **1** and **2**, spaced apart from and oriented generally parallel (within +/−2°) to the four CR boundary planes **118**, **128**, **138** **148**, are half mirrors **115**, **125**, **135**, **145**, respectively. The half-mirrors **115**, **125**, **135**, **145** reflect a specific mode of visible light Specifically, each of the half mirrors **115**, **125**, **135**, **145** can be an s-polarizer half mirror or a p-polarizer half mirror. The half mirrors **115**, **125**, **135**, **145** can be in the form of a diffraction grating or thin film polarizer that reflects the s-mode of visible light and allows the p-mode of visible light to pass through (a p-polarization diffraction grating or thin film), or in the alternative, reflects the p-mode of visible light and allows the s-mode of the visible light to pass through (an s-polarization diffraction grating or thin film). It should be appreciated that the half mirrors **115**, **125** are both s-polarizer half mirrors or p-polarizer half mirrors and the half mirrors **135**, **145** are both s-polarizer half mirrors or p-polarizer half mirrors, i.e. the half mirrors **115**, **125** may be s-polarizer mirrors and the half mirrors **135**, **145** may be p-polarizer half mirrors; the half mirrors **115**, **125** may be p-polarizer mirrors and the half mirrors **135**, **145** may be s-polarizer half mirrors; or all of the half mirrors **115**, **125**, **135**, **145** may be s-polarizer half mirrors or p-polarizer half mirrors.

The four CR boundary planes **118**, **128**. **138** **148**, and the corresponding half mirrors **115**, **125**, **135**, **145**, may be part of four cloaking assembly components **110**, **120**, **130**, **140**, respectively. Each of cloaking assembly components **110**, **120**, **130**, **140** has an outward facing surface **112**, **122**, **132**, **142**, respectively, an inward facing surface **114**, **124**, **134**, **144**, respectively, an outer end surface **116**, **126**, **136**, **146**, respectively, and an inner side surface defined by the inward facing opaque surface **118***b*, **128***b*, **138***b*, **148***b*, respectively. The outward facing surfaces **112**, **132** form an object-side **102** of the cloaking device and the outward facing surfaces **122**, **142** form an image-side **104** of the cloaking device. One or more optical filters **125***a*, **125***b* may be positioned on or adjacent to the outward facing surface **122** and one or more optical filters **145***a*, **145***b* may be positioned on or adjacent to the outward facing surface **142**. The optical filter **125***a* is positioned to the right (+X direction) of the half-mirror **125**, the optical filter **125***b* is positioned to the left (−X direction) of the half-mirror **125**, the optical filter **145***a* is positioned to the left (−X direction) of the half-mirror **145**, and the optical filter **145***b* is positioned to the right (+X direction) of the half-mirror **145**. The one or more optical filters **125***a*, **125***b*, **145***a*, **145***b* may be quarter-wave plates that convert linearly polarized light into circularly polarized light, or in the alternative, the one or more optical filters **125***a*, **125***b*, **145***a*, **145***b* may be a quarter-wave plate and a linear polarizer that convert linearly polarized light of one mode, e.g., s-mode or p-mode, into circularly polarized light, and then convert the circularly polarized light into a different linearly polarized mode, e.g., p-mode or s-mode, respectively.

The CR boundary planes **118**, **138** are oriented at an angle θ_1 relative to the outward facing surfaces **112**, **132**, respectively, and the CR boundary planes **128**, **148** are oriented at an angle β_1 relative to the outward facing surfaces **122**, **142**, respectively. In embodiments, the angle θ_1 is equal to the angle β_1. In other embodiments, the angle θ_1 is not equal to the angle β_1. The half mirrors **115**, **135** are oriented at an angle θ_2 relative to the outward facing surfaces **112**, **132**, respectively, and the half mirrors **125**, **145** are oriented at an angle β_2 relative to the outward facing surfaces **122**, **142**, respectively. In embodiments, the angle θ_2 is equal to the angle β_2. In other embodiments, the angle θ_2 is not equal to the angle β_2. In embodiments, the angle θ_1 is equal to the angle θ_2 and the angle β_1 is equal to the angle β_2. For example, θ_1 and θ_2 can be generally equal to 45° (+/−1°), and β_1 and β_2 can be generally equal to 45° (+/−10).

The four cloaking assembly components **110**, **120**, **130**, **140** are transparent to incident light (both polarized and unpolarized) except for the half mirrors **115**, **125**, **135**, **145**, the outward facing mirror surfaces **118***a*, **128***a*, **138***a*, **148***a*, the inward facing opaque surfaces **118***b*, **128***b*, **138***b*, **148***b*, and optionally the outer end surfaces **116**, **126**, **136**, **146**. The four cloaking assembly components **110**, **120**, **130**, **140** can be made from any suitable transparent material, for example a transparent glass or a transparent plastic. In the alternative, the four cloaking assembly components **110**, **120**, **130**, **140** can be a structure, e.g. a frame, that holds the four CR boundary planes **118**, **128**, **138**, **148**, and the four half mirrors **115**, **125**, **135**, **145** in a desired orientation relative to each other with air or another gas present between the four CR boundary planes **118**, **128**, **138**, **148**, and the four half mirrors **115**, **125**, **135**, **145**. The frame may be a plurality of transparent glass or transparent plastic panes that hold the four CR boundary planes **118**, **128**, **138**, **148**, and the four half mirrors **115**, **125**, **135**, **145** in a desired orientation relative to each other with air or another gas present between the four CR boundary planes **118**, **128**, **138**, **148**, and the four half mirrors **115**, **125**, **135**, **145**. In embodiments, transparent glass or transparent plastic panes form the outward facing surfaces **112**, **122**, **132**, **142**, inward facing surfaces **114**, **124**, **134**, **144**, and outer end surfaces **116**, **126**, **136**, **146**, and hold the four CR boundary planes **118**, **128**, **138**, **148**, and the four half mirrors **115**, **125**, **135**, **145** in a desired orientation relative to each other with air or another gas present between the four CR boundary planes **118**, **128**, **138**, **148**, and the four half mirrors **115**, **125**, **135**, **145**.

The outer end surfaces **116**, **126**, **136**, **146** may be transparent, have an absorber layer disposed thereon or be a half-mirror. In embodiments, the outer end surfaces **116**, **126**, **136**, **146** are transparent. In other embodiments, the outer end surfaces **116**, **126**, **136**, **146** have an absorber layer disposed thereon that absorbs incident light from within the four cloaking assembly components **110**, **120**, **130**, **140**, and/or from outside the four cloaking assembly components

110, **120**, **130**, **140**. The absorber layer may be made from any suitable material that absorbs approximately 90% (+/−10%) of incident light, illustratively including dark colored surfaces made of glass, polymers, metals, ceramics, composites, etc. In other embodiments, the outer end surfaces **116**, **126**, **136**, **146** are half-mirrors such that one mode of incident light is reflected (e.g., s-polarized light or p-polarized light) and another mode of incident light is transmitted through the half mirror (e.g., p-polarized light or s-polarized light, respectively). In other embodiments, one or more of the outer end surfaces **116**, **126**, **136**, **146** is transparent and one or more of the outer end surfaces **116**, **126**, **136**, **146** has an absorber layer disposed thereon. In other embodiments, one or more outer end surfaces **116**, **126**, **136**, **146** is transparent and one or more of the outer end surfaces **116**, **126**, **136**, **146** is a half mirror. In other embodiments, one or more of the outer end surfaces **116**, **126**, **136**, **146** has an absorber layer disposed thereon and one or more of the outer end surfaces **116**, **126**, **136**, **146** is a half mirror. In other embodiments, one or more of the outer end surfaces **116**, **126**, **136**, **146** is transparent, one or more of the outer end surfaces **116**, **126**, **136**, **146** has an absorber layer disposed thereon, and one or more of the outer end surfaces **116**, **126**, **136**, **146** is a half mirror.

While FIGS. **1**-**2** illustrate four cloaking assembly components **110**, **120**, **130**, **140**, it should be appreciated that the four CR boundary planes **118**, **128**, **138**, **148**, and the four half mirrors **115**, **125**, **135**, **145** can be contained within any number of cloaking assembly components. For example, in embodiments, the cloaking assembly components **110** and **120** may be a single cloaking assembly component with no inward facing surface **114**, **124**, between the outward facing surfaces **112**, **122**, and the cloaking assembly components **130** and **140** may be a single cloaking assembly component with no inward facing surface **134**, **144**, between the outward facing surfaces **132**, **142**. In other embodiments, the cloaking assembly components **110** and **130** may be a single cloaking assembly component and the two cloaking assembly components **120** and **140** can be a single cloaking assembly component. In other embodiments, three of the four cloaking assembly components **110**, **120**, **130**, **140** may be a single cloaking assembly component assembled with a remaining one of the cloaking assembly components **110**, **120**, **130**, **140**. For example, the three cloaking assembly components **110**, **120**, **130** may be a single cloaking assembly component assembled with the remaining cloaking assembly component **140**. In other embodiments, the four cloaking assembly components **110**, **120**, **130**, **140** may be a single cloaking assembly component.

Still referring to FIGS. **1**-**2**, light 'lo' is reflected from an object 'O' (e.g., a target object) located on the object-side **102** of the cloaking device, travels in the −Y direction, and is incident on and transmitted through the outward facing surfaces **112**, **132**. The light lo reflected from the obscured portions of the object 'O' located directly behind the CR in the +Y direction is reflected by the outward facing mirror surfaces **118***a*, **138***a* in the +X and −X directions, respectively. The light lo reflected in the +X and −X directions is incident on the half mirrors **115**, **135** respectively. As illustrated in FIG. **1**, the half mirrors **115**, **135** are p-polarization half mirrors that reflect the s-mode (ls) of the light lo and allow the p-mode (lp) of the light lo to pass through. The s-polarized light ls reflected from the half mirrors **115**, **135** in the −Y direction is incident on the half mirrors **125**, **145**, respectively. The half mirrors **125**, **145** are also p-polarization mirrors that reflect the s-polarized light ls traveling in the −Y direction into the −X and +X directions, respectively. The s-polarized light ls reflected in the −X and +X directions is incident on the outward facing mirror surfaces **128***a*, **148***a*, respectively, and the outward facing mirror surfaces **128***a*, **148***a* reflect the incident s-polarized light ls into the −Y direction. The s-polarized light reflected into the −Y direction by the outward facing mirror surfaces **128***a*, **148***a*, forms an image of the obscured portion of the target object located directly above the CR **200** in the +Y direction that is visible to an observer viewing the image-side **104** of the cloaking assembly **10**. In embodiments, the optical filter **125***b*, the optical filter **145***b*, or both optical filters **125***b*, **145***b* may be positioned on or adjacent to the outward facing surfaces **122**, **142**, respectively, as described above, and may convert the s-polarized light reflected into the −Y direction by the outward facing mirror surfaces **128***a*, **148***a*, into circularly polarized light or p-polarized light. For example, the optical filter **125***b* the optical filter **145***b*, or both optical filters **125***b*, **145***b* may be quarter wave plates that convert the s-polarized light reflected into the −Y direction by the outward facing mirror surfaces **128***a*, **148***a* into circularly polarized light (not shown) before forming the image of the obscured portion of the target object 'O' located directly above the CR **200** in the +Y direction. In the alternative, the optical filter **125***b*, the optical filter **145***b*, or both optical filters **125***b*, **145***b* may be a quarter-wave plate and a linear polarizer that converts the s-polarized light reflected into the −Y direction by the outward facing mirror surfaces **128***a*, **148***a* into circularly polarized light (not shown) and then converts the circularly polarized light into p-polarized light before forming the image of the obscured portion of the target object 'O' located directly above the CR **200** in the +Y direction.

Light lo reflected from the visible portions of the target object 'O' not located directly behind the CR in the +Y direction is incident on and transmitted through the outward facing surfaces **112**, **132**. The transmitted light lo is incident on the p-polarization half mirrors **115**, **135**. The half mirrors **115**, **135** reflect the s-mode is of the incident light lo in the +X and −X directions, respectively, and transmit the p-mode lp of the incident light lo in the −Y direction. The p-polarized light lp transmitted through the half mirrors **115**, **135** will be incident on and be transmitted through the p-polarization half mirrors **125**, **145**, respectively, in the −Y direction. The p-polarized light lp transmitted in the −Y direction by the half mirrors **125**, **145** forms an image of the visible portions of the target object O not located directly behind the CR **200** in the +Y direction that is visible to an observer viewing the cloaking assembly in the +Y direction. In embodiments, the optical filter **125***a*, the optical filter **145***a*, or both optical filters **125***a*, **145***a* may be positioned on or adjacent to the outward facing surfaces **122**, **142**, respectively, as described above, in order to convert the p-polarized light into circularly polarized light of s-polarized light. For example, the optical filter **125***a*, the optical filter **145***a* or both optical filters **125***a*, **145***a* may be quarter wave plates that convert the p-polarized light transmitted into the −Y direction by the half mirrors **125**, **145** into circularly polarized light (not shown) before forming the image of the visible portions of the target object 'O' not located behind above the CR **200** in the +Y direction. In the alternative, the optical filter **125***a*, the optical filter **145***a*, or both optical filters **125***a*, **145***a* may be a quarter-wave plate and a linear polarizer that convert the p-polarized light reflected into the −Y direction by the outward facing mirror surfaces **128***a*, **148***a* into circularly polarized light (not shown) and then convert the circularly polarized light into s-polarized light before forming the image of the visible portions of the target object 'O' not located directly above the CR **200** in the +Y direction. Accordingly, an image 'I' of the entire target object 'O' (both the obscured portion and the visible portions) is visible to the observer viewing the image-side **104** of the cloaking assembly **10**. While half mirrors **115**, **125**, **135**, **145** are described herein as p-polarization half mirrors, it should be appreciated that half mirrors **115**, **125**, **135**, **145** can be s-polarization half mirrors and the reflected and transmitted portions of the reflected light lo discussed above will be reversed. However, the human eye cannot distinguish between s-polarized light, p-polarized light or circularly polarized light and the image 'I' of the entire object 'O' will be visible to the human eye when an observer is viewing the image-side **104** of the cloaking assembly **10**. Also, it should be appreciated that the one or more of optical filters **125***a*, **125***b*, **145***a*, **145***b* may be positioned on or adjacent to outward facing surfaces **122**, **142** as described above such that only p-polarized light forms the image 'I', only s-polarized light forms the image 'I', only circularly polarized light forms the image 'I', only p-polarized light and circularly polarized light form the image 'I' or only s-polarized light and circularly polarized light form the image 'I'. Such manipulation of the polarized light reflected into the −Y direction by the outward facing mirror surfaces **128***a*, **128***a* and the polarized light transmitted in the −Y direction by the half mirrors **125**, **145** may assist an observer wearing polarized sunglasses in viewing the image-side **104** of the cloaking assembly **10**. Particularly, the one or more optical filters **125***a*, **125***b*, **145***a*, **145***b* ensure that polarized sunglasses that block p-polarized light or s-polarized light do not prevent an observer wearing such polarized sunglasses from seeing the entire target object 'O' when viewing the image-side **104** of the cloaking assembly **10**.

Referring now to FIGS. **1**-**4**, a top perspective view and a side view of a cloaking device according to embodiments (as discussed with respect to FIGS. **1**-**2**) are shown in FIGS. **3**-**4**, respectively. Specifically, FIG. **3** is a top perspective view of an article in the form of a column 'C' within the CR **200** of the cloaking assembly **10** and an automobile 'A' located behind the column C on the object-side **102** of the cloaking assembly **10** in the +Y direction. The column C has a height dimension in the Z direction (increasing height in the −Z direction) greater than the height h of the cloaking device. FIG. **4** is a side view from the +Y direction of the cloaking assembly **10** shown in FIG. **3** and shows the portion of the column C that is within the CR **200** is not visible and the automobile A located behind the column C in the +Y direction is visible to an observer viewing the cloaking assembly **10** in the +Y direction. Accordingly, the column C positioned within the CR **200** is not visible to an observer viewing the image-side **104** of the cloaking assembly **10** and an image of the entire automobile A (both the obscured portion and the visible portions) is visible to the observer viewing the image-side **104**. Although column C in FIGS. **3**-**4** is separate from the CR boundary planes **118**, **128**, **138**, **148**, i.e., column C is a separate object from the cloaking device **10**, it should be appreciated that column C may be structurally part of the cloaking device **10** and have an outer surface that provides or is equivalent to the CR boundary planes **118**, **128**, **138**, **148** with outward facing mirror surfaces **118***a*, **128***a*, **138***a*, **148***a*, respectively.

Referring now to FIGS. **1**-**5**, FIG. **5** shows embodiments of a cloaking assembly **12** with an enlarged CR **200**. Particularly, the cloaking assembly components **110**, **120**, **130**, **140** are arranged such that the cloaking assembly components **110**, **130** are spaced apart in the Y direction from the cloaking assembly components **120**, **140**, and a distance 'd' is between the inward facing surfaces **114**, **124**, and between the inward facing surfaces **134**, **144**. The area of the CR **200** in the X-Y plane is increased by the width 'w' times the distance d (w×d). The volume of the CR **200** is increased by the width w times the distance d times the height h (FIG. **4**) of the CR **200** (w×d×h). While FIG. **5** illustrates four cloaking assembly components **110**, **120**, **130**, **140**, it should be appreciated that the four CR boundary planes **118**, **128**, **138**, **148**, and the four half mirrors **115**, **125**, **135**, **145** can be contained within any number of cloaking assembly components. For example, in embodiments, the cloaking assembly components **110** and **130** may be a single cloaking assembly component and the two cloaking assembly components **120** and **140** can be a single cloaking assembly component.

As shown in the FIG. **5**, light lo is reflected from the target object O, travels in the −Y direction and is incident on and transmitted through the outward facing surfaces **112**, **132**. Light lo reflected from the obscured portions of the object located directly behind the CR in the +Y direction is reflected by the outward facing mirror surfaces **118***a*, **138***a* in the +X and −X direction, respectively. The reflected light lo in the +X and −X directions is incident on the half mirrors **115**, **135** respectively. The half mirrors **115**, **135** are p-polarization half mirrors and reflect the s-mode is of the light lo and transmit the p-mode lp of the light lo. The s-polarized light ls reflected by the half mirrors **115**, **135** in the −Y direction is incident on the half mirrors **125**, **145**, respectively. The half mirrors **125**, **145** are also p-polarization mirrors that reflect the s-polarized light ls traveling in the −Y direction into the −X and +X directions, respectively. The s-polarized light ls traveling in the −X and +X directions is incident on the outward facing mirror surfaces **128***a*, **148***a*, respectively, and the outward facing mirror surfaces **128***a*, **148***a* reflect the incident s-polarized light ls in the −Y direction. The s-polarized light reflected in the −Y direction by the outward facing mirror surfaces **128***a*, **148***a*, forms an image of the obscured portion of the object located behind the CR **200** in the +Y direction that is visible to an observer viewing the image-side **104** of the cloaking assembly **12**. Light lo reflected from the visible portions of the target object O not located behind the CR **200** in the +Y direction is incident on and transmitted through the outward facing surfaces **112**, **132**. The transmitted light lo is incident on the p-polarization half mirrors **115**, **135**, and the half mirrors **115**, **135** reflect the s-mode is of the incident light lo in the +X and −X directions, respectively, and transmit the p-mode lp of the incident light lo in the −Y direction. The p-polarized light lp transmitted through the half mirrors **115**, **135** in the −Y direction is incident on and transmitted through the p-polarization half mirrors **125**, **145**, respectively, in the −Y direction. The p-polarized light lp transmitted in the −Y direction through the half mirrors **125**, **145**, forms an image of the visible portions of the target object O not located behind the CR **200** in the +Y direction that is visible to an observer viewing the image side **104** of the cloaking assembly **12**. Accordingly, an image I of the entire target object O (both the obscured portion and the visible portions) is visible to an observer viewing the image-side **104** of the cloaking assembly **12**. It should be appreciated that half mirrors **115**, **125**, **135**, **145**, can be s-polarization half mirrors and the reflected and transmitted portions of the reflected light lo discussed above will be reversed. Also, it should be appreciated that the one or more of optical filters **125***a*, **125***b*, **145***a*, **145***b* may be included. However, the human eye cannot distinguish between s-polarized light, p-polarized light or circularly polarized light and the image I of the entire object O will be visible to an observer viewing the image-side **104** of the cloaking assembly **12**.

Referring now to FIGS. **1**-**7**, a top perspective view and a side view of a cloaking device according to embodiments (as discussed with respect to FIG. **5**) are shown in FIGS. **6**-**7**, respectively. Specifically, FIG. **6** is a top perspective view of a first article in the form of a column 'C1' and a second article in the form of a column 'C2' within the CR **200** of the cloaking assembly **12** and the automobile 'A' located behind the columns C1, C2 on the object-side **102** of the cloaking assembly **12** in the +Y direction. The column C1 has a first height dimension in the Z direction and the column C2 has a second height dimension that is greater (in the −Z direction) than the first height dimension of column C**1**. The first height dimension of column C**1** and the second height dimension of column C**2** are both greater than the height h of the cloaking assembly **12**. FIG. **7** is a side view from the +Y direction of the cloaking assembly **12** shown in FIG. **6** and shows the image-side **104**. Portions of column C1 and column C2 within the CR **200** are not visible and automobile A located behind column C1 and column C2 in the +Y direction is visible to an observer viewing the image-side **104** of the cloaking assembly **12**. Although columns C1, C2 in FIGS. **6**-**7** are separate from the CR boundary planes **118**, **128**, **138**, **148**, i.e., columns C1, C2 are separate objects from the cloaking device **12**, it should be appreciated that columns C**1**, C**2** may be structurally part of the cloaking device **12** and have an outer surface that provides or is equivalent to the CR boundary planes **118**, **128**, **138**, **148** with outward facing mirror surfaces **118***a*, **128***a*, **138***a*, **148***a*, respectively.

Referring to FIG. **8**, embodiments of a cloaking device that can display additional images in conjunction with the images depicted in FIGS. **1**, **4**, **5** and **7** are shown. Particularly, FIG. **8** shows an item in the form of the number '1' (item 1) and an item in the form of the number '2' (item 2) located adjacent to the outer end surfaces **126**, **146**, respectively, in the +X, −X directions, respectively. Light lo reflected from item 1 and item 2 is incident on and transmitted through the outer end surfaces **126**, **146** in the −X, +X directions, respectively. The light lo transmitted through the outer end surfaces **126**, **146** is incident on the half mirrors **125**, **145**, respectively. The half mirrors **125**, **145** (p-polarization half mirrors) transmit the p-mode lp of the light lo in the −X, +X directions, respectively, and reflect the s-mode is of the light lo into the −Y direction. The transmitted p-polarized light lp in the −X, +X directions is incident on the outward facing mirror surfaces **128***a*, **148***a*, respectively. The outward facing mirror surfaces **128***a*, **148***a*, reflect the incident p-polarized light lp into the −Y direction. The s-polarized light ls reflected by the half mirror **125** in the −Y direction forms a reversed '1' image on the right hand side of the outward facing surface **122** and the s-polarized light ls reflected by the half mirror **145** in the −Y direction forms a reversed '2' image on the left hand side of the outer facing surface **142**. The p-polarized light lp reflected by the outward facing mirror surface **128***a* into the −Y direction forms a reversed '1' image on the left hand side of the outward facing surface **122** and the p-polarized light lp reflected by the outward facing mirror surface **148***a* into the −Y direction forms a reversed '2' image on the right hand side of the outer facing surface **142**. It should be appreciated that the display of additional images as depicted in FIG. **8** in conjunction with images as depicted in FIGS. **1**, **4**, **5** and **7** can be used as part of a heads-up display, a warning system (signal) that an object is located on the object side of the cloaking device, etc.

Referring to FIGS. **1**-**10**, embodiments of a cloaking device with a cloaking assembly **14** having items in the form of '1', '2', '3', '4' adjacent to the outer end surfaces **126**, **146**, **116**, **136**, respectively are shown in FIG. **9**. FIG. **10** is a side view of the cloaking assembly **14** illustrated in FIG. **9** with reversed images of item 1 and item 2, in conjunction with the image of the automobile A, visible on the image-side **104** of the cloaking assembly **14**. It should be appreciated that items 1, 2, 3, 4 depicted in FIGS. **9**-**10** are shown for illustrative purposes only, i.e. the items 1, 2 in FIG. **9** could be oriented in the Y-Z plane in order to provide reversed images of items 1, 2 on the image side **104** (X-Z plane) in FIG. **10**. It should also be appreciated that reversed images of item 3 and item 4 can be formed or provided on the object-side **102** of the cloaking assembly **14** in a similar manner as the reversed images of item 1 and item 2 are formed on the image-side **104**.

Referring to FIGS. **1**-**8** and **11**, embodiments of a cloaking device that can selectively display images in conjunction with the images depicted in FIGS. **1**, **4**, **5** and **7** are shown. Particularly, FIG. **11** shows an item in the form of the number '1' and an item in the form of the number '2' located adjacent to the outer end surfaces **126**, **146**, respectively, in the +X, −X directions, respectively. Light lo from item 1 is incident on an s-polarizer **150**. The s-polarizer **150** allows the s-mode is of the incident light lo to pass through in the −X direction. The s-polarized light ls is incident on and reflected by the p-polarization half mirror **125** into the Y direction and no light is transmitted through or beyond the p-polarization half mirror **125** in the −X direction. Accordingly, a reversed image of item 1 is formed only on the right hand side of the outward facing surface **122**. The light lo from item 2 is transmitted through the outer end surface **146** in the +X direction and is incident on the p-polarization half mirror **145**. The p-polarization half mirror **145** transmits the p-mode lp of the incident light lo in the +X direction and reflects the s-mode is of the incident light lo in the −Y direction. The p-polarized light lp in the +X direction is incident on the outward facing mirror surface **148***a*. The outer facing mirror surface **148***a* reflects the incident p-polarized light lp in the −Y direction. The s-polarized light ls reflected by the half mirror **145** in the −Y direction forms a reversed item 2 image on the left hand side of the outer facing surface **142**. The p-polarized light lp reflected by the outer facing mirror surface **148***a* into the −Y direction provides a reversed item 2 image on the right hand side of the outer facing surface **142**.

Referring to FIGS. **1**-**8** and **11**-**13**, embodiments of a cloaking device with a cloaking assembly **16** having items in the form of '1' and '2' adjacent to the outer end surfaces **126**, **146**, respectively, and the s-polarizer **150** adjacent the outer end surface **126**, are shown in FIG. **12**. FIG. **13** is a side view of the cloaking assembly **16** illustrated in FIG. **12** with the reversed image of item 1 visible on the right hand side of the outer facing surface **122** of the cloaking assembly component **120** and the reversed image of item 2 visible on the left hand side and right hand side of the outward facing surface **142** of the cloaking assembly components **140** shown in conjunction with the image of the automobile A. It should be appreciated that items 1, 2 depicted in FIG. **12** are shown for illustrative purposes only, i.e. the items 1, 2 in FIG. **12** could be oriented in the Y-Z plane in order to provide the reversed images of items 1, 2 on the image-side (**104**) (X-Z plane) shown in FIG. **13**.

Referring to FIGS. **1**-**8** and **14**, another embodiment of a cloaking device that can selectively display additional images in conjunction with the images depicted in FIGS. **1**,

4, **5** and **7** is shown. Particularly, FIG. **14** shows an item in the form of the number '1' and an item in the form of the number '2' located adjacent to the outer end surfaces **126**, **146**, respectively, in the +X, −X directions, respectively. Light lo reflected from the item 1 is incident on a p-polarizer **152** which allows p-polarized light lp to pass through in the −X direction. The p-polarized light lp is incident on the p-polarization half mirror **125**. The incident p-polarized light lp in the −X direction is transmitted through the p-polarization half mirror **125** in the −X direction and there is no s-polarized light is to be reflected by the p-polarization half mirror **125** in the −Y direction. The p-polarized light lp transmitted through the p-polarization half mirror **125** is incident on the outward facing mirror surface **128***a* and the outward facing mirror **128***a* reflects the p-polarized light lp into the −Y direction. Accordingly, a reversed image of item 1 is formed only on the left hand side of the outward facing surface **122**. The light lo reflected from item 2 is transmitted through the outer end surface **146** in the +X direction and is incident on the p-polarization half mirror **145**. The p-polarization half mirror **145** transmits the p-mode lp of the incident light lo in the +X directions and reflects the s-mode ls of the incident light lo into the −Y direction. The p-polarized light lp in the +X is incident on the outer facing mirror surface **148***a*. The outer facing mirror surface **148***a* reflects the incident p-polarized light lp into the −Y direction. The s-polarized light ls reflected by the half mirror **145** in the −Y direction forms a reversed 2 on the left hand side of the outer facing surface **142** and the p-polarized light lp reflected by the outer facing mirror surface **148***a* in the −Y direction forms a reversed 2 image on the right hand side of the outer facing surface **142**.

Referring to FIGS. **1**-**8** and **14**-**16**, embodiments of a cloaking device with a cloaking assembly **18** having items 1 and items 2 adjacent to the outer end surfaces **126**, **146**, respectively, and the p-polarizer **152** adjacent to the outer end surface **126**, are shown in FIG. **15**. FIG. **16** is a side view of the cloaking assembly **18** illustrated in FIG. **15** with the single reversed image of item 1 visible on the left hand side of the outer facing surface **122** of the cloaking assembly component **120** and the reversed image of item 2 visible on the left hand side and right hand side of the outward facing surface **142** of the cloaking assembly components **140** in conjunction with the image of the automobile A. It should be appreciated that items 1, 2 depicted in FIG. **15** are shown for illustrative purposes only, i.e. the items 1, 2 in FIG. **15** could be oriented in the Y-Z plane in order to provide the reverse images of items 1, 2 on the image-side **104** (X-Z plane) shown in FIG. **16**.

Referring to FIGS. **1**, **2**, **5**, **8**-**17**, embodiments of an A-pillar of a vehicle being cloaked by a cloaking device are shown. Particularly, FIG. **17** shows a cloaking device **19** cloaking a portion of an A-pillar P of a vehicle V. A portion of the A-pillar P is positioned within a CR **200** (not shown) of the cloaking device **19** and a portion of the A-pillar P extends beyond the cloaking device and is covered with trim T. Illustrated outside of the vehicle V is a target object O in the form of pedestrian. A portion of the pedestrian O is visible through a side window of the vehicle V and a portion of the pedestrian is visible "through" the A-pillar P cloaked by the cloaking device **19**. The cloaking device **19** redirects light reflected from the pedestrian O around the A-pillar P positioned within the CR **200** of the cloaking device **19** and forms an image I of the pedestrian O on an image-side **104** of the cloaking device **19** that is visible to an occupant of the vehicle V looking towards the pedestrian O. Accordingly, light from the pedestrian O appears to pass through the A-pillar P and a blind spot typically created by the A-pillar P is not present. In embodiments, the A-pillar P itself serves as the CR **200**, i.e. the A-pillar P has an outer surface with one or more outer facing mirror surfaces that assist in redirecting light from the pedestrian) around the A-pillar P. It should be appreciated that cloaking of the A-pillar P with the cloaking device **19** and removing the blind spot produced by the A-pillar P is performed without the use of metamaterials, video images, cameras, sophisticated electronics, etc.

Referring to FIGS. **18**-**19**, alternative embodiments of cloaking devices are shown. Specifically, FIG. **18** shows a cloaking device **20** with a circular CR **202** having an outer mirror surface **204** with half mirrors **206**, **208** around the outer mirror surface **204** having a circular shape. In embodiments, the CR **202** is provided by an object to be cloaked itself, e.g., an A-pillar, that has an outer mirror surface **204**. The cloaking device shown in FIG. **18** can increase the field of view for an observer viewing the cloaking device **20** in the +Y direction. FIG. **19** shows a cloaking device **30** with a circular CR **302** having an outer mirror surface **304** with half mirrors **306**, **308** and optical components **310**, **312** such as lenses. In embodiments, the CR **302** is provided by an object to be cloaked itself, e.g., an A-pillar, that has an outer mirror surface **304**. The cloaking device shown in FIG. **19** allows for an external view from an object located behind the CR **302** (+Y direction) to be manipulated to increase a field of view or to normalize a distorted image.

The cloaking devices described herein may be used to cloak vehicle articles such as a vehicle A-pillar, and remove a "blind spot" caused by the vehicle article. It will be understood that the term "blind spot" as used herein refers to an area around the vehicle that cannot be directly observed by a driver while operating the vehicle. The terms "object," "article," and "item" may interchangeably refer to a visual object or image (2D or 3D) that reflects light or transmits light and the team "light from" may refer to "light reflected from" or "light transmitted from." The terms "generally," "approximately," and "about" may be utilized herein to represent the inherent degree of uncertainty that may be attributed to any quantitative comparison, value, measurement, or other representation. These terms are also utilized herein to represent the degree by which a quantitative representation may vary from a stated reference without resulting in a change in the basic function of the subject matter at issue.

While particular embodiments have been illustrated and described herein, it should be understood that various other changes and modifications may be made without departing from the spirit and scope of the claimed subject matter. Moreover, although various aspects of the claimed subject matter have been described herein, such aspects need not be utilized in combination. It is therefore intended that the appended claims cover all such changes and modifications that are within the scope of the claimed subject matter.

What is claimed is:

1. A cloaking device comprising:

an object-side and an image-side;

a cloaking region (CR) boundary plane having an outward facing mirror surface and an inward facing opaque surface;

a CR at least partially bounded by the inward facing opaque surface of the CR boundary plane;

a half mirror spaced apart from and generally parallel to the outward facing mirror surface;

wherein light from an object located on the object-side of the cloaking device and obscured by the CR is redirected around the CR to form an image of the object on the image-side of the cloaking device such that the light from the object appears to pass through the CR.

2. The cloaking device of claim **1**, wherein the half mirror is selected from a p-polarization half mirror and an s-polarization half mirror.

3. The cloaking device of claim **1**, further comprising an article positioned within the CR, wherein the light from the object located on the object-side of the cloaking device and obscured by the CR is redirected around the CR and the article to form the image of the object on the image-side of the cloaking device such that the article appears transparent.

4. The cloaking device of claim **3**, wherein the article is a vehicle A-pillar.

5. The cloaking device of claim **1**, further comprising:

at least two CR boundary planes positioned non-planar to each other, each of the at least two CR boundary planes having an outward facing mirror surface and an inward facing opaque surface, the CR at least partially bounded by the inward facing opaque surfaces of the at least two CR boundary planes;

at least one quarter-wave plate positioned on or adjacent to the image side; and

at least two half mirrors spaced apart from and generally parallel to the outward facing mirror surfaces, respectively; such that a half mirror of the at least two half mirrors is spaced apart from and generally parallel to each of the outward facing mirror surfaces;

wherein the light from the object located on the object-side of the cloaking device and obscured by the CR is redirected around the CR by the at least two outward facing mirror surfaces and the at least two half mirrors, and the image of the object is formed on the image-side of the cloaking device such that the light from the object located on the object-side of the cloaking device appears to pass through the CR.

6. The cloaking device of claim **5**, wherein the at least two CR boundary planes are positioned orthogonal to each other.

7. The cloaking device of claim **6**, wherein the light from the object on the object-side of the cloaking device is reflected by a first outward facing mirror surface, a first half mirror positioned parallel to the first outward facing mirror surface, a second half mirror positioned orthogonal to the first half mirror, a second outward facing mirror surface positioned parallel to the second half mirror, and transmitted through the at least one quarter-wave plate before forming the image of the object on the image-side of the cloaking device.

8. The cloaking device of claim **7**, further comprising a vehicle article positioned within the CR, wherein the light from the object located on the object-side of the cloaking device and obscured by the CR is redirected around the CR and the vehicle article to form the image of the object on the image-side of the cloaking device and the vehicle article appears transparent.

9. The cloaking device of claim **5**, further comprising an end side, wherein light from an item positioned adjacent to the end side forms an image of the item on the image-side of the cloaking device.

10. A cloaking device for cloaking an article of a vehicle comprising:

a cloaking assembly comprising:

an object-side and an image-side;

at least two cloaking region (CR) boundary planes located at least partially between the object-side and the image-side, the at least two CR boundary planes positioned non-planar to each other and each of the at least two CR boundary planes having an outward facing mirror surface and an inward facing opaque surface;

a CR bounded by the inward facing opaque surfaces of the at least two CR boundary planes;

at least two half mirrors located at least partially between the object-side and the image-side, each of the at least two half mirrors spaced apart from and generally parallel to one of the outward facing mirror surfaces;

at least one quarter-wave plate positioned on or adjacent to the image side;

and

a vehicle article positioned within the CR;

wherein light from an object located on the object-side of the cloaking assembly and obscured by the CR is redirected around the CR and forms an image of the object on the image-side of the cloaking assembly such that the vehicle article appears to be transparent.

11. The cloaking device of claim **10**, wherein the vehicle article is a vehicle A-pillar.

12. The cloaking device of claim **10**, wherein the light from the object is reflected by a first outward facing mirror surface, a first half mirror oriented generally parallel to the first outward facing mirror surface, a second half mirror oriented generally orthogonal to the first half mirror and a second outward facing mirror surface oriented generally parallel to the second half mirror and transmitted through the at least one quarter-wave plate before forming the image on the image-side of the cloaking assembly.

13. The cloaking device of claim **12**, further comprising an end surface, wherein light from an item located adjacent to the end surface passes through the cloaking assembly and forms an image of the item on the image-side of the cloaking assembly.

14. The cloaking device of claim **13**, wherein a p-mode or an s-mode of the light from the item is reflected from the second half mirror before forming the image of the item on the image-side of the cloaking assembly.

15. The cloaking device of claim **13**, wherein a p-mode or an s-mode of the light from the item is transmitted through the second half mirror and reflected by the second outward facing mirror surface before forming the image of the item on the image-side of the cloaking assembly.

16. A cloaking device comprising:

four cloaking region (CR) boundary planes positioned non-planar to each other, each of the four CR boundary planes having an outward facing mirror surface and an inward facing opaque surface;

a CR bounded by the inward facing opaque surfaces of the four CR boundary planes;

an article located at least partially within the CR;

four half mirrors spaced apart from and generally parallel to the outward facing mirror surfaces, respectively, such that a half mirror of the four half mirrors is spaced apart from and generally parallel to each of the outward facing mirror surfaces;

wherein light from an object on an object-side of the cloaking device is redirected around the CR and appears as an image on an image-side of the cloaking device and is visible to an observer viewing the image-side of the cloaking device, and wherein the article within the CR is not visible to the observer viewing the image-side of the cloaking device such that the article within the CR appears transparent to the observer viewing the image-side of the cloaking device.

17. The cloaking device of claim **16**, wherein the light from the object is reflected by a first of the outward facing mirror surfaces, a first of the outward facing mirror surfaces, a first of the four half mirrors oriented generally parallel to the first outward facing mirror surface, a second half mirror oriented generally orthogonal to the first half mirror and a second outward facing mirror surface oriented generally parallel to the second half mirror before forming the image on the image-side of the cloaking device.

18. The cloaking device of claim **17**, wherein the article is a vehicle A-pillar.

19. The cloaking device of claim **18**, further comprising an end side, wherein light from an item located adjacent to the end side is reflected by at least one of the second half mirror and the second outward facing mirror surface and forms an image of the item on the image-side of the cloaking device.

20. The cloaking device of claim **19**, wherein the image of the item is selectively formed on the image-side of the cloaking device.

* * * * *

US010389138B2

(12) **United States Patent**
Dinwiddie et al.

(10) **Patent No.: US 10,389,138 B2**
(45) **Date of Patent: *Aug. 20, 2019**

(54) **POWER RECEIVER FOR EXTRACTING POWER FROM ELECTRIC FIELD ENERGY IN THE EARTH**

(71) Applicant: **Earth Energies, Inc.**, Johns Creek, GA (US)

(72) Inventors: **John Dinwiddie**, Cary, NC (US); **Terry L. Wright**, Suwanee, GA (US); **David R. Ames**, Johns Creek, GA (US)

(73) Assignee: **EARTH ENERGIES, INC.**, Alpharetta, GA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 306 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **15/397,281**

(22) Filed: **Jan. 3, 2017**

(65) **Prior Publication Data**

US 2017/0117714 A1 Apr. 27, 2017

Related U.S. Application Data

(63) Continuation-in-part of application No. 14/509,772, filed on Oct. 8, 2014, now Pat. No. 9,564,268.

(60) Provisional application No. 61/889,894, filed on Oct. 11, 2013.

(51) **Int. Cl.**
H05F 7/00 (2006.01)
H02J 4/00 (2006.01)

(52) **U.S. Cl.**
CPC ***H02J 4/00*** (2013.01); *H05F 7/00* (2013.01); *Y10T 307/549* (2015.04)

(58) **Field of Classification Search**
CPC H05F 7/00
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

8,004,250 B2 * 8/2011 Grandics H01Q 1/248
290/1 R
9,564,268 B2 * 2/2017 Dinwiddie H01T 9/00

FOREIGN PATENT DOCUMENTS

DE 102008029441 A1 * 12/2009 .... H05F 7/00

* cited by examiner

Primary Examiner — Fritz M Fleming
(74) *Attorney, Agent, or Firm* — Coats & Bennett, PLLC

(57) **ABSTRACT**

A resonant transformer connected between a ground terminal and elevated terminal draws current from the earth's electric field through a primary winding of the transformer. An impulse generator applies a high voltage impulse to the primary winding of the resonant transformer to cause current to flow from the ground terminal through the primary winding. The flow of current through the primary winding of the resonant transformer induces a current in the secondary winding, which may be converted and filtered to a usable form, e.g. 60 Hz AC or DC.

18 Claims, 8 Drawing Sheets

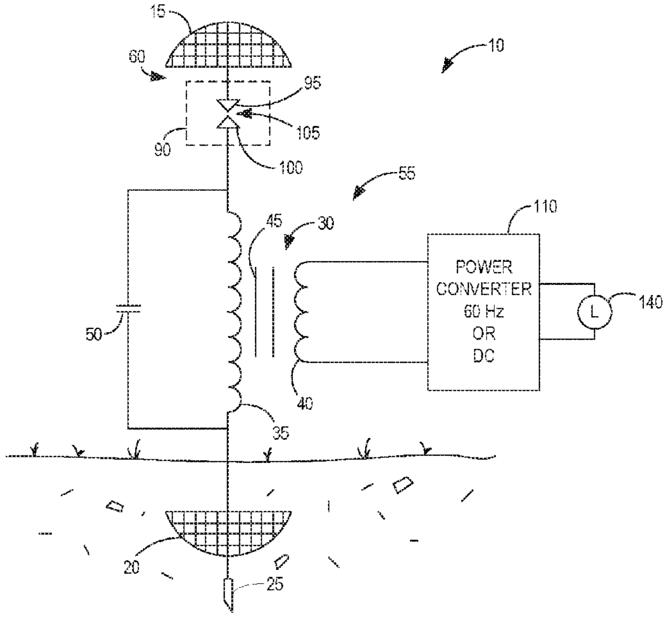

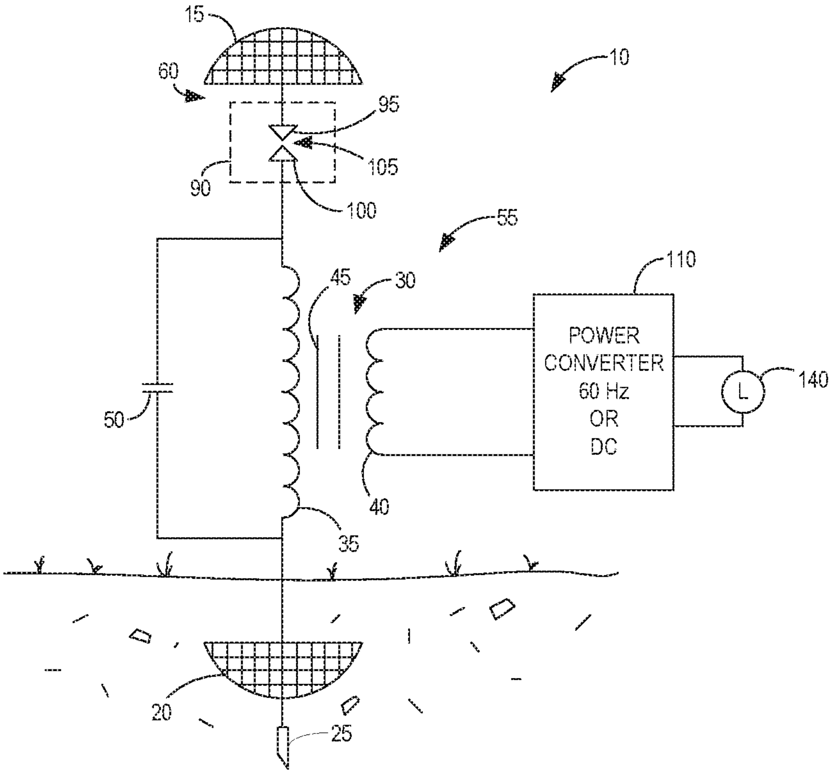
POWER
CONVERTER
60 Hz
OR
DC
L

FIG 1

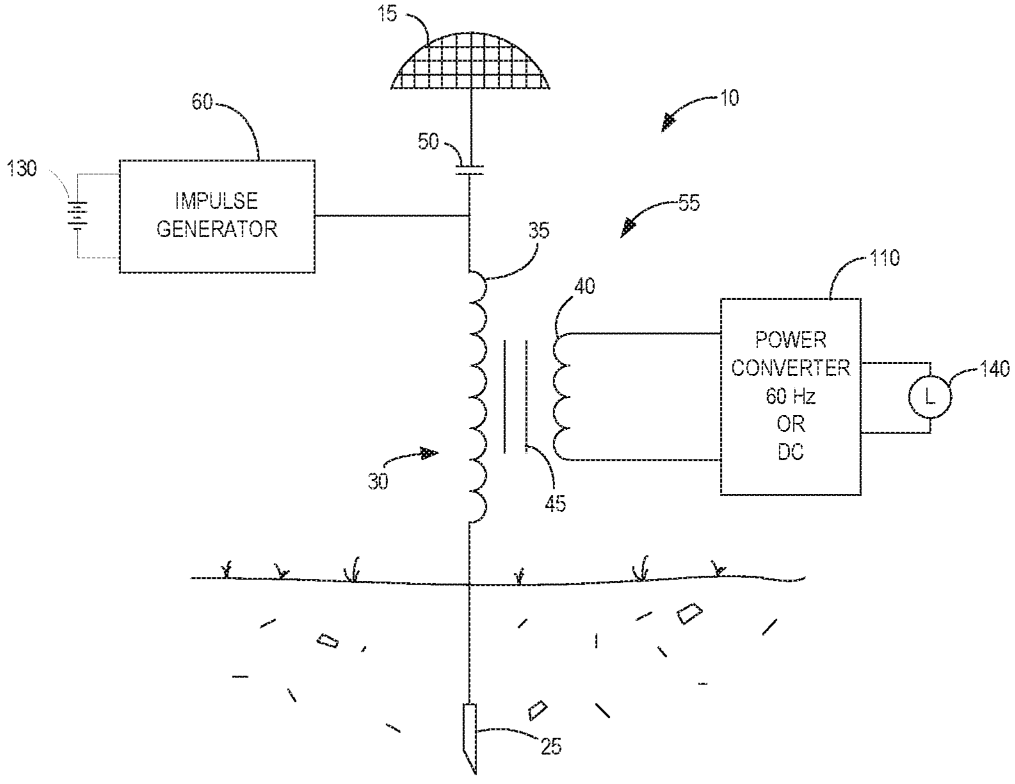
IMPULSE GENERATOR
POWER CONVERTER 60 Hz OR DC
L

FIG 2

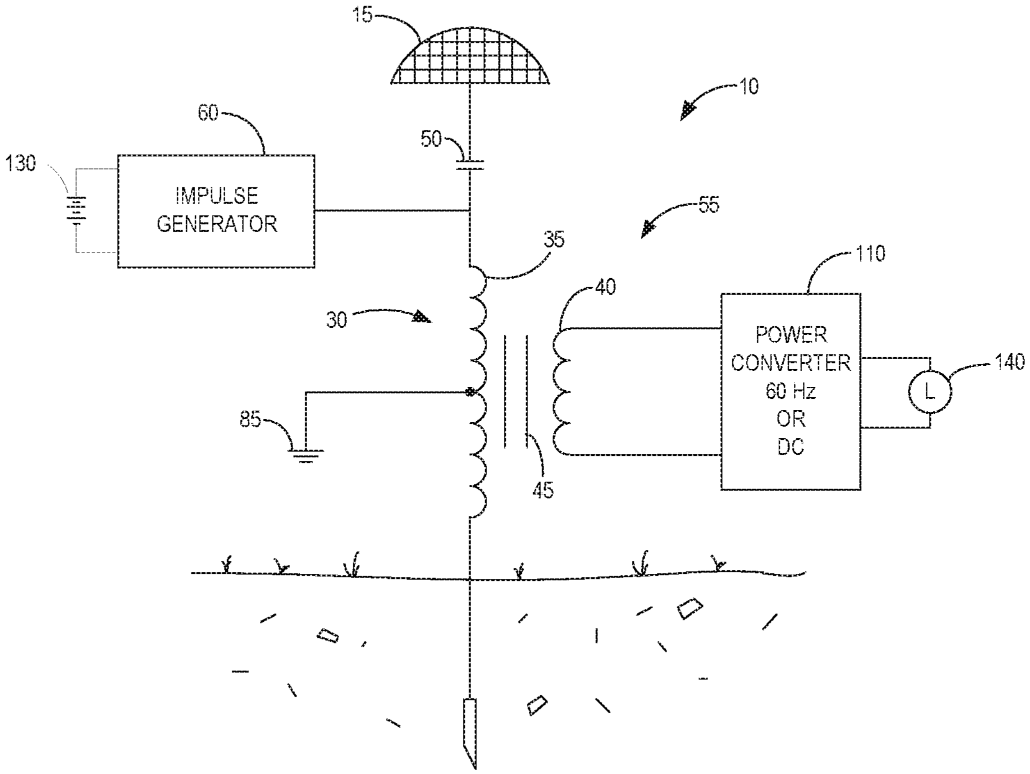
IMPULSE GENERATOR
POWER CONVERTER 60 Hz OR DC
L

FIG 3

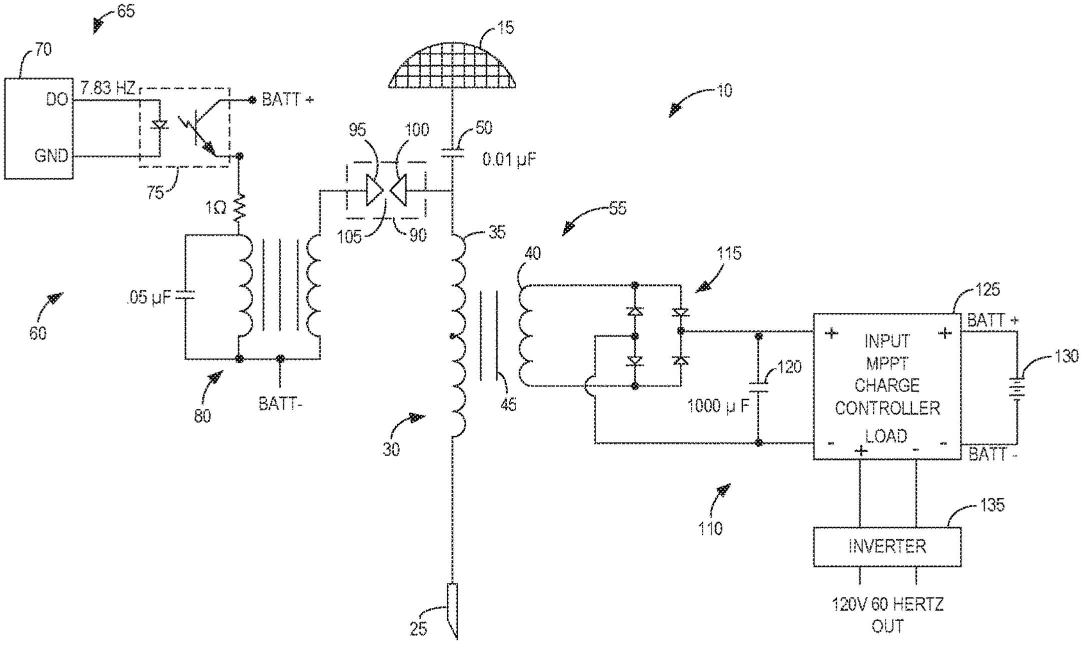
65
70
DO
GND
7.83 HZ
BATT +
75
1Ω
.05 µF
60
80
BATT-
15
95
100
50
0.01 µF
105
90
35
30
25
45
10
55
40
115
120
1000 µ F
125
INPUT
MPPT
CHARGE
CONTROLLER
LOAD
BATT +
BATT -
130
110
135
INVERTER
120V 60 HERTZ
OUT

FIG 4

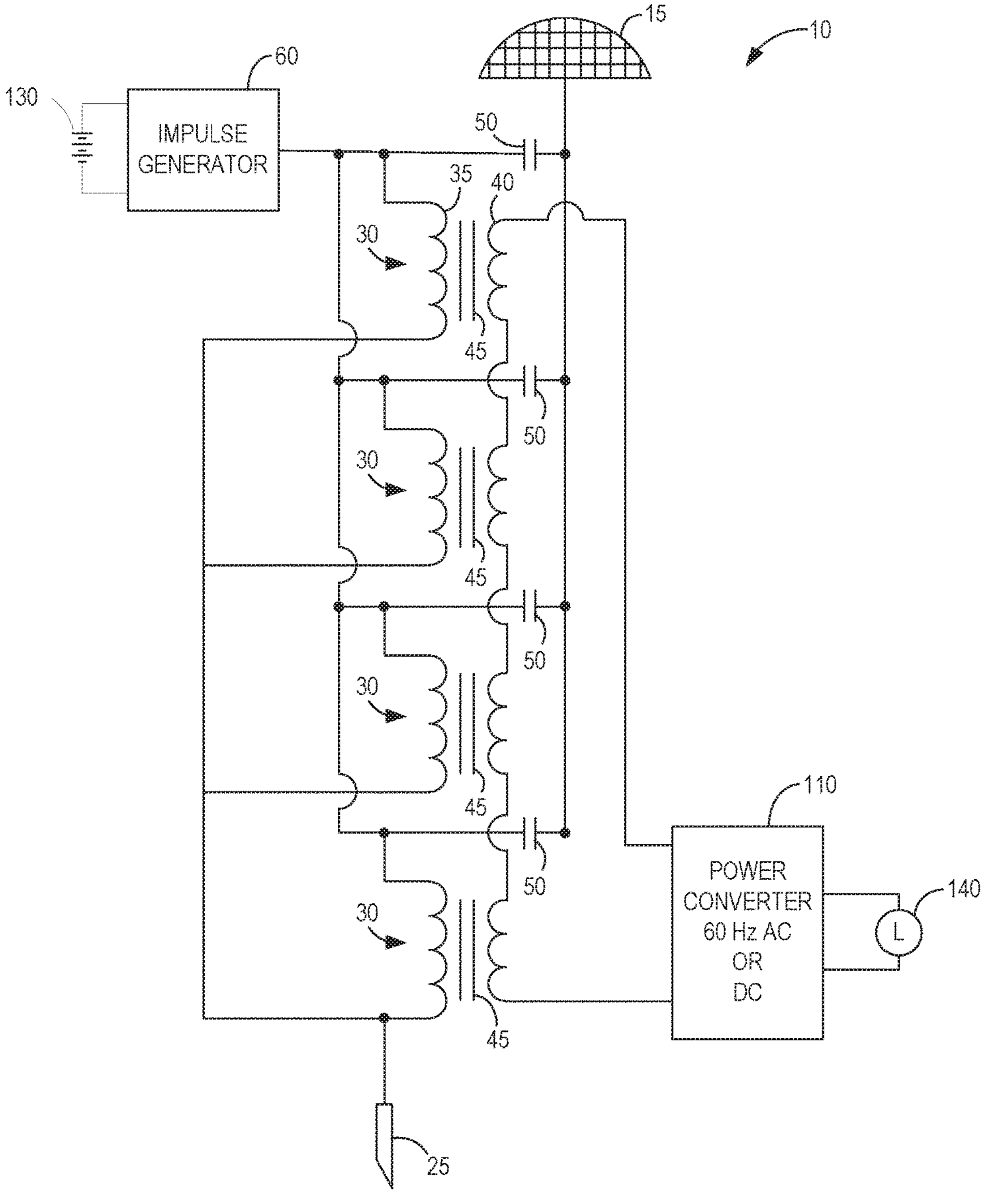
IMPULSE GENERATOR
POWER CONVERTER 60 Hz AC OR DC
L

FIG 5

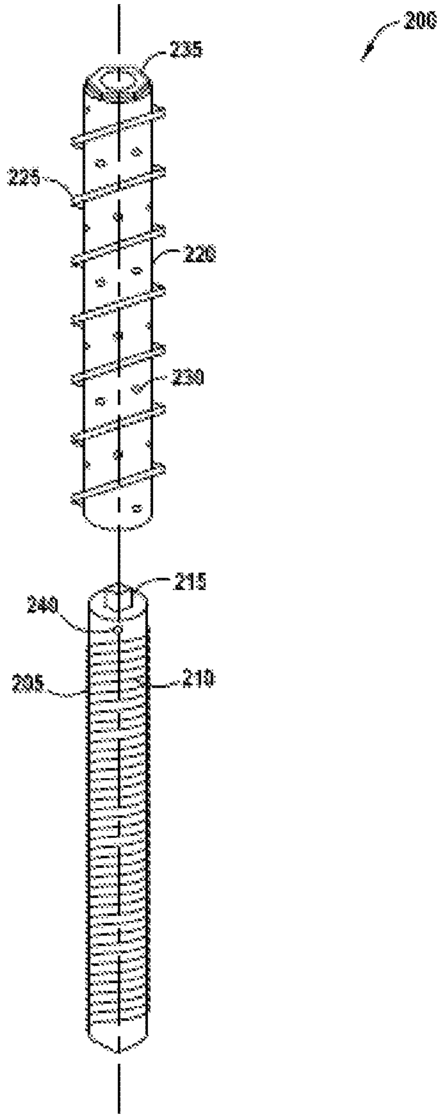
200
235
225
220
230
215
240
205
210

FIG 6A

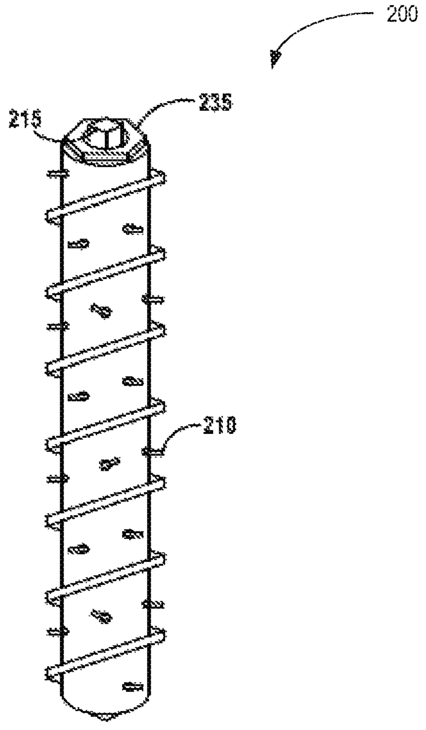
200
215
235
210

FIG 6B

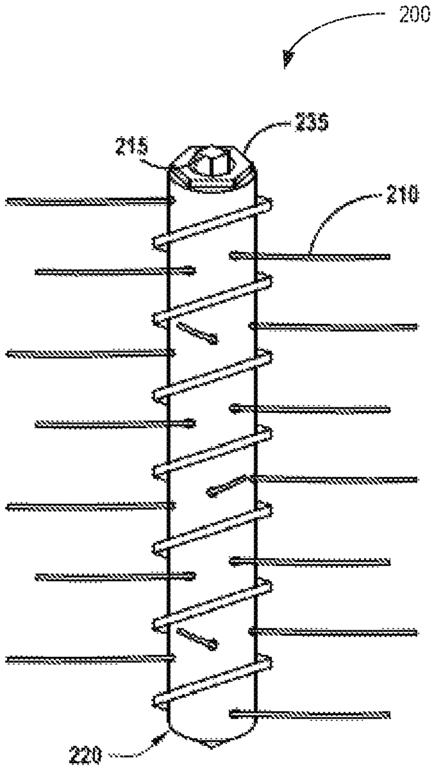
200
215
235
210
220

FIG 6C

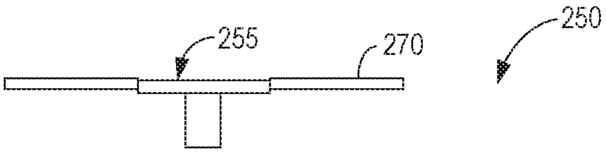

FIG 7A

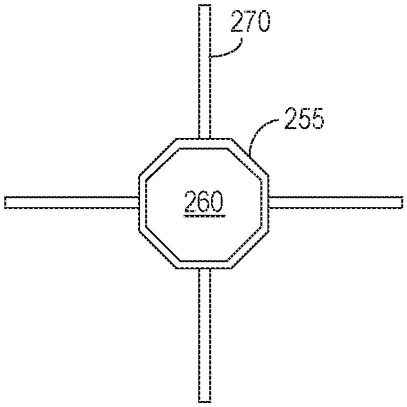

FIG 7B

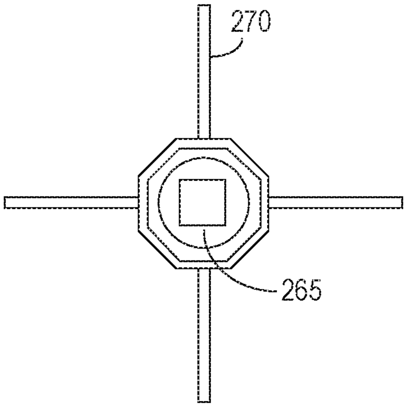

FIG 7C

POWER RECEIVER FOR EXTRACTING POWER FROM ELECTRIC FIELD ENERGY IN THE EARTH

This application is a continuation-in-part of prior U.S. application Ser. No. 14/509772 filed 8 Oct. 2014 which claims the benefit of U.S. Provisional Application No. 61/889,894 filed 11 Oct. 2013, the disclosures of all of which are incorporated by reference herein in their entirety.

TECHNICAL FIELD

The present invention relates generally to renewable energy, and more particularly to methods and apparatus for extracting energy from subsurface electrical fields beneath the earth's surface.

BACKGROUND

The earth and the ionosphere cavity may be viewed as a global electric circuit. Electrical currents are constantly flowing within the earth and its atmosphere. Within the earth, the majority of the earth's energy is carried by extremely low frequency (ELF) and ultralow frequency (ULF) waves in the 0-200 Hz frequency range. The earth's rotating magnetic field and positive lightning are two energy sources that sustain the ELF/ULF waves within the earth and the atmosphere.

A great deal of research has been devoted to studying the electric field present in the earth's ionosphere cavity. Joseph M. Crawley, the "Fair Weather Atmosphere as a Power Source", Proceedings ESA Annual Meeting on Electrostatics 2011; O. Jefimenko, "Operation of Electric Motors from Atmospheric Electric Field," American Journal of Physics, Vol. 39, Pgs. 776-779, 1971; M. L. Breuer, "Usability of Tapping Atmospheric Charge as a Power Source," Renewable Energy, Vol. 28, Pgs. 1121-1127, 2003. Numerous attempts have been made in the past to extract electrical energy from the earth's atmosphere. For example, U.S. Pat. No. 1,540,998 to Plauson describes a system for converting atmospheric electrical energy into usable power. These past attempts have been successful in producing only small amounts of power from the electrical field in the earth's ionosphere cavity. The modest success of these experiments compared to results from other renewable energy sources, such as solar and wind, has tempered further research and prevented widespread use of the electric field in the ionosphere cavity as an energy source.

SUMMARY

The present invention relates to a power receiver for extracting power from electric fields beneath the earth's surface. In embodiments of the present disclosure, a resonant transformer connected to a ground terminal draws current from the earth's electric field through the primary winding of the transformer. Current flow through the resonant transformer is induced by applying a high voltage impulse to the primary winding. The flow of current through the primary winding of the resonant transformer induces a current in the secondary winding, which may be converted and filtered to a usable form, e.g. 60 Hz AC or DC.

In some embodiments of the power receiver, the resonant frequency of the resonant transformers is below 200 Hz.

In some embodiments of the power receiver, the resonant transformer comprises a ferro-resonant transformer.

In some embodiments, the power receiver further comprises an elevated terminal.

In some embodiments of the power receiver, the primary winding of the resonant transformer is connected between the ground terminal and elevated terminal.

In some embodiments of the power receiver, the elevated terminal comprises an upper capacitive plate coupled to the earth's ionosphere cavity.

In some embodiments of the power receiver, the impulse generator comprises the upper capacitive plate and a spark gap connected between the upper capacitive plate and the primary winding of the resonant transformer. The spark gap comprises a pair of electrodes separated by a gap and configured to generate a spark when a voltage difference between the electrodes reaches a predetermined level.

In some embodiments of the power receiver, the impulse generator comprises a pulse generator for generating low voltage pulses, a step-up transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses, and a spark gap connected between the step-up transformer and the primary winding of the resonant transformer to generate a spark responsive to the high voltage impulses from the step-up transformer.

In some embodiments of the power receiver, the impulse generator comprises a pulse generator for generating low voltage pulses, and a step-up transformer connected to the primary winding of the resonant transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses.

In some embodiments of the power receiver, the impulse generator comprises a solid state spark generator.

In some embodiments of the power receiver, the resonant transformer includes a capacitor connected in parallel with the primary winding.

In some embodiments of the power receiver, the resonant transformer includes a capacitor connected in series with the primary winding between the impulse generator and the elevated terminal.

In some embodiments, the power receiver comprises multiple resonant transformers having primary windings connected in parallel between the ground terminal and the elevated terminal.

In some embodiments of the power receiver, the resonant transformers have different resonant frequencies.

In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are all below 200 Hz.

In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

Another embodiment of the power receiver comprises a resonant circuit connected to a ground terminal disposed below the surface of the earth, an impulse generator for generating and applying a high voltage electrical impulse to the resonant circuit to induce current flow from the ground terminal through the resonant circuit, and a power conversion circuit connected to the resonant circuit to convert electrical current flowing through the resonant circuit to a desired form. The resonant circuit has a resonant frequency below 200 Hertz.

In some embodiments of the power receiver, the resonant circuit comprises a resonant transformer having a primary winding, a secondary winding, and resonant capacitor connected in series with the primary winding.

In some embodiments of the power receiver, the resonant circuit comprises multiple resonant transformers having primary windings connected in parallel to the ground terminal.

In some embodiments of the power receiver, the resonant transformers have different resonant frequencies.

In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are all below 200 Hz.

In some embodiments of the power receiver, the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

Other embodiments of the disclosure comprise a ground terminal for a power receiver. In one embodiment, the ground terminal comprises a ground shaft configured for insertion beneath the surface of the earth, a hollow cylinder surrounding the ground shaft and having a plurality of openings, and a plurality of ground wires connected at one end to the ground shaft. The ground wires are wound around the ground shaft and have free ends protruding through respective openings in the hollow shaft so that rotation of the ground shaft relative to the hollow cylinder causes the ground wires to extend radially into the earth.

Other embodiments of the disclosure comprise methods of extracting power from the earth. In one embodiment, the method comprises applying a high voltage impulse to resonant circuit coupled to a ground terminal disposed beneath the surface of the earth to initiate resonance in the resonant circuit and induce the flow of current from the ground terminal to the resonant circuit, and converting the current flowing from the ground terminal into the resonant circuit into a useful form.

In some embodiments of the method, the resonant circuit comprises a resonant transformer including a primary winding coupled to the ground terminal and a second winding coupled to a power converter, and applying a high voltage impulse to resonant circuit comprises applying a high voltage impulse to the primary winding of the resonant transformer.

In some embodiments of the method, applying a high voltage impulse to the primary winding of the resonant transformer comprises applying an impulse in the range to 10,000 to 40,000 volts to primary winding of the transformer.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** illustrates a first exemplary embodiment of a power receiver.

FIG. **2** illustrates a second exemplary embodiment of a power receiver.

FIG. **3** illustrates a third exemplary embodiment of a power receiver.

FIG. **4** illustrates a fourth exemplary embodiment of a power receiver.

FIG. **5** illustrates a fifth exemplary embodiment of a power receiver.

FIG. **6**A is an exploded perspective view of an exemplary ground antenna array for the power receiver.

FIG. **6**B is a perspective view of an assembled ground antenna array before being deployed.

FIG. **6**C is a perspective view of an assembled ground antenna array after being deployed.

FIG. **7**A is a side view of an insertion tool for installing the ground antenna array.

FIG. **7**B is a top view of the insertion tool for installing the ground antenna array.

FIG. **7**C is a bottom view of the insertion tool for installing the ground antenna array.

DETAILED DESCRIPTION

Referring now to the drawings, a power receiver for extracting energy from the earth's electric field are illustrated and indicated generally by the numeral **10**. Various embodiments of the power receiver **10** are described and similar reference numbers are used throughout the description to indicate similar components.

The power receiver **10** converts energy in the ELF/ULF waves to useful form, e.g. 60 Hz AC or DC. The power receiver **10** is essentially a resonance circuit that resonates at the natural resonance frequencies in the earth's electric field. These resonance frequencies, known as Schumann resonance frequencies, occur at 7.83 Hz, 14.3 Hz, 20.8 Hz, 27.3 Hz, and 33.8 Hz. A high voltage impulse initiates resonance within the power receiver **10**. In the resonant mode, the impedance of the power receiver **10** is reduced to near zero thus inducing ground currents to flow into the power receiver **10** where the ground currents are converted to useful form.

FIG. **1** illustrates a first embodiment of the power receiver **10**. The power receiver **10** comprises a resonant transformer **30** connected between an elevated terminal **15** and ground terminal **25**. In this embodiment, the elevated terminal **15** is capacitively coupled to electric fields within the earth's ionosphere cavity and functions as an upper capacitive plate. A lower capacitive plate **20** is connected to the ground terminal **25** beneath the surface of the earth.

The resonant transformer **30** comprises a primary winding **35**, secondary winding **40**, ferromagnetic core **45**, and capacitor **50**. One end of the primary winding **35** is connected to the lower capacitive plate **20** and ground terminal **25**. The opposite end of the primary winding **35** is connected via a spark gap **90** to the elevated terminal **15**. The capacitor **50** is connected in parallel with the primary winding **35** of the resonant transformer **30** to form an LC circuit **55** with a resonance frequency range of between about 0.1 and 200 Hz. In a preferred embodiment, the resonant transformer has a Q of about 10 or greater and resonance frequency in the range of about 0.1-200 Hertz. For example, the resonant transformer **30** may have a resonance frequency of about 7.83 Hz, the fundamental Schumann resonance frequency. The secondary winding **40** of the resonant transformer **30** is connected to a power converter **110** as will be hereinafter described in greater detail. The power converter **110** converts the energy extracted from the earth's electric field by the power receiver **10** into a usable form for driving a load **140**.

The elevated terminal/upper capacitive plate **15** comprises an insulated, dish-shaped plate with a large radius of curvature. The capacitance and resistance of the elevated terminal is chosen for receiving broadband electric field frequencies in the 0-200 Hz range. The upper capacitive plate **15** is sized to maximize to the extent practical coupling with the electric field in the earth's ionosphere cavity.

The lower capacitive plate **20** is also a dish-shaped plate with a large radius of curvature. One function of the lower capacitive plate **20** is to collect charge from the earth's ground currents and provide an instantaneous source of current as hereinafter described. The capacitance and resistance of the lower capacitive plate **20** is selected to promote the flow of current from the ground with minimal losses.

The spark gap **90** connected between the elevated terminal **15** and resonant transformer **30** comprises a pair of electrodes **95**, **100** separated by an evacuated air gap **105**. Electrode **95** is connected to the upper capacitive plate **15**. Electrode **100** is connected to the resonant transformer **30**. The spark gap **105** prevents electrical discharge from the upper capacitive plate **15** to the earth's ionosphere cavity. The spark gap **90** in combination with the elevated terminal **15** function as an impulse generator that applies a high voltage impulse of about 10,000-40,000 volts to the primary winding **35** to initiate resonance in the transformer **30**.

In operation, the capacitive coupling of the upper capacitive plate **15** induces a high voltage operating current in the upper capacitive plate **15**. The upper capacitive plate is connected to a first electrode **95** to the spark gap **90**. When the voltage difference between the electrodes **95** and **100** reaches a threshold, a spark forms across the electrodes **95**, **100** and a high voltage impulse is applied to the primary winding **35** of the resonant transformer **30**. This high voltage impulse initiates resonance within the transformer **30**.

In resonant mode, the impedance of the resonance transformer is reduced to nearly zero allowing current to flow from the capacitive plate **20** and ground terminal **25** through the primary winding **35** of the transformer **30**, which in turn induces current in the secondary winding **40**. Power converter **110** converts the current flowing through the secondary winding **40** into a usable form for driving a load **140**. The transformer **30** will continue to resonate for a short period of time. By providing high voltage impulses to the primary winding **35** of the resonant transformer **30** at periodic intervals, it is possible to maintain a continuous flow of current from the earth into the resonant transformer **30**, thus producing a continuous supply of power.

FIG. **2** discloses a second embodiment of the primary receiver **10**. This embodiment includes a resonant transformer **30** connected between an elevated terminal **15** and ground terminal **25**. The resonant transformer **30** comprises a primary winding **35**, secondary winding **40**, ferromagnetic core **45** and a high voltage capacitor **50**. One end of the primary winding **35** is connected to the ground terminal **25**. The opposite end of the primary winding **35** is connected to the elevated terminal **15**. The capacitor **50** has a capacitance of about 0.01 micro-farads. In contrast to the previous embodiment, capacitor **50** is connected in series with the primary winding **35** and elevated terminal **15** and forms a LC circuit **55** with a Q of about 10 or greater and a resonance frequency in the range of about 0.1 to 200 Hz. In a preferred embodiment, the resonance frequency of the transformer **30** is 7.83 Hz, the fundamental Schumann resonance frequency. An impulse generator **60** is connected between the primary winding **35** of the resonant transformer **30** and the series capacitor **50** and applies a high voltage impulse to the primary winding **35** of the resonant transformer **30**. A battery **130** or other external power source supplies power to the impulse generator **60**. As previously described, the high voltage impulse applied by the impulse generator **60** initiates resonance within the resonant transformer **30** inducing current flow from the ground terminal **25** into the primary winding **35** of the resonant transformer **30**. The flow of current from the ground terminal **25** through the primary winding **35** induces current in the secondary winding **40**. Power converter **110** converts the electrical energy in the current flowing through the primary winding **40** into a usable form.

In contrast to the first embodiment, it is not required to capacitively couple the elevated terminal **15** in the second embodiment to the earth's ionosphere cavity. Rather, the elevated terminal **15** in this embodiment provides lightning protection and dissipates some of the energy flowing into the

power receiver **10** to the earth's ionosphere cavity. Also, in contrast to the first embodiment, the capacitor **50** is connected in series between the primary winding **35** of the transformer **30** and the elevated terminal **15**. Those skilled in the art will appreciate that the capacitor **50** could also be connected in parallel rather than series with the primary winding **35** as shown in FIG. **1**. Another difference is that the impulse generator **60** has an external power source. The amount of energy generated by the power receiver **10**, however, is far greater than the energy needed to generate high voltage impulses. The first embodiment does not require an external power source to generate high voltage impulses.

FIG. **3** illustrates a third embodiment of the power receiver **10**. This embodiment is essentially the same as the embodiment shown in FIG. **2**. The main difference is that a center tap of the primary winding **35** in the resonant transformer **30** is connected to an electrical ground **85**. It should be appreciated that the electrical ground **85** may be different than the earth ground. When the center tap of the resonant transformer **30** is grounded at a distance away from the ground terminal **25** (e.g. 50 ft to 100 ft), the power receiver **10** becomes a transmitter via the ground loop formed.

FIG. **4** illustrates the power receiver **10** of FIG. **2** in greater detail. The power receiver includes a resonant transformer **30** connected between a ground terminal **25** and an elevated terminal **15**. The ground terminal **25** may comprise a ⅝-inch×8-foot copper ground rod, such as the ERICO 615880UPC. The elevated terminal **15** may comprise a 90% copper mesh formed into a hemisphere with a radius of about 9 inches. The elevated terminal **15** may be elevated at a height of approximately 6 feet above the ground.

The resonant transformer **30** includes a primary winding **35**, secondary winding **40**, ferromagnetic core **45** and series capacitor **50** configured as previously described. The resonant transformer **30** may have a Q of about 10 and a resonance frequency in the range of about 0.1 to 200 Hz. The resonant transformer **30** may be made using an Allanson transformer (part #1530BP120R) connected in series with a 0.01 micro-farad capacitor, such as the Condensor Products high voltage capacitor (part #TC 103-17-125). The resonant transformer **30** is used in a step-down configuration. The center tap of the resonant transformer **30** may optionally be connected to a ground.

An impulse generator **60** is connected between the primary winding **35** of the resonant transformer **30** and the series capacitor **50** and applies a high voltage impulse in the range of about 10,000 to 40,000 volts to the primary winding of the transformer **30**. A battery **130** or other external power source supplies power to the impulse generator **60**. The power converter **110** connects to the secondary winding **40** of the resonant transformer **30** for converting current in the secondary winding of the transformer to a useful form.

The impulse generator **60** comprises a pulse generator **65** for generating low voltage pulses, a step-up transformer **80** for converting the low voltage pulses from the pulse generator **65** to high voltage pulses, and a spark gap **90** for generating sparks responsive to the high voltage pulses from the step-up transformer **80**.

The pulse generator **65** comprises a square wave generator **70**, such as a Sinometer VC2002 function signal generator, and solid state relay **75**. The square wave generator **70** generates a digital pulse stream. In one embodiment, the digital pulse stream generates a square waveform with a frequency of about 7.83 Hz. The frequency of the digital pulse stream is selected to match the resonance frequency of the transformer **30**, though such is not necessarily required.

The pulse stream output from the square wave generator **70** is applied to the solid state relay **75**. The solid state relay **75** is connected between a battery or other power source and a first winding of the step-up transformer **80**. The battery may comprise a 12 V, 7.0 A/H sealed lead acid battery, such as the ELB 1270A by Lithonia Lighting. The solid state relay **75** functions as a switch that is activated responsive to the waveform from the square wave generator **70** to provide a continuous stream of low voltage pulses from the battery to the first winding of the step-up transformer **80**. A 1 ohm resistor is connected between the solid state relay **75** and step-up transformer **80**.

The step-up transformer **80** may comprise a Transco 15 kV, 30 mA neon sign transformer (part #S15612). The step-up transformer **80** converts the low voltage pulses from the pulse generator **65** to high voltage pulses that are applied to the spark gap **90**. The step-up transformer has a 0.5 micro-farad capacitor connected in parallel with the primary winding of the step-up transformer **80**. The step-up transformer produces pulses at the output of about 30,000 to 40,000 volts.

The spark gap **90** comprises a pair of electrodes **95**, **100** separated by an air gap **105**. A suitable spark gap electrode pair is the Information Unlimited SPARK05 ¼-inch×1-inch tungsten electrodes. As previously described, when the voltage potential between the electrodes **95**, **100** reaches a threshold, a spark forms between the electrodes **95**, **100** and supplies a nearly instantaneous, high voltage impulse to the primary winding **35** of the resonant transformer **30**. This high voltage impulse initiates resonance in the resonant transformer **30** inducing current flow from the ground terminal **25** through the primary winding **35** of the resonant transformer **30**.

The power converter **110** comprises a bridge rectifier **115**, filter capacitor **120**, charge controller **125**, and inverter **135**. A suitable rectifier is the Micro Commercial Components 10 amp, 1000 volt bridge rectifier (Part #GBJL 1010). The bridge rectifier **115** converts the AC current flowing through the secondary winding **40** of the resonant transformer to a DC current. A filter capacitor **120** removes unwanted frequencies from the DC current. A suitable capacitor **120** is Cornell Dubilier 1000uF 450VDC capacitor (part #383LX102M450N082). The filter capacitor **120** has a capacitance of about 1000 micro-farads. The DC current is input to the charge controller **125**. The charge controller **125** may, for example, comprise a maximum power point tracking (MPPT) charge controller, such as a Tracer 4215 BN MPPT Solar Charge Controller, which is commonly used in solar power generating systems. The charge controller **125** applies a small amount of energy to a battery **30** to charge the battery **130**. As previously noted, the battery **130** serves as a power source for the impulse generator **60**. The remaining current is supplied to an inverter **135**, which converts the DC current to an AC current with a desired voltage and frequency, e.g., 120 volts/60 Hz AC. A suitable inverter **135** is the 1500 W Pure Sine power inverter (AIMS) (part #PWRI1500125).The power converter **110** as shown in FIG. **4** may be utilized in the embodiment shown in FIGS. **1**, **2** and **3**.

FIG. **5** illustrates a power receiver **10** according to another embodiment. The power receiver **10** comprises a plurality of resonant transformers **30** connected between a ground terminal **25** and elevated terminal **15**. Each of the resonant transformers **30** comprises a primary winding **35**, secondary winding **40**, ferromagnetic core **45** and series capacitor **50**. The primary windings **35** of the resonant transformers **30** are connected in parallel. The secondary windings **40** are connected in series with the power converter **110**. An impulse generator **60** applies a high voltage impulse to the primary windings **35** of the resonant transformers **30**. A battery **130** or other external power source supplies power to the impulse generator **60**. The power converter **110** converts the current in the power converter circuit to a usable form for driving a load **140**.

In one embodiment, each of the resonant transformers **30** shown in FIG. **5** is configured to have a different resonant frequency. In one embodiment, the resonant transformers **30** are configured to resonate at frequencies of 7.83 Hz, 14.8 Hz, 20.3 Hz and 26.8 Hz respectively. Additional resonant transformers **30** could be added to operate at other resonance frequencies.

FIGS. **6A-6C** illustrate a high quality ground antenna array **200** which may be used as a ground terminal **25**. The ground antenna array **200** comprises a generally cylindrical ground shaft **205** disposed with a hollow cylinder **220** and a plurality of reinforced, heavy gauge ground wires **210** attached at one end to the ground shaft **205**. The ground shaft **205** and ground wires **210** should be highly conductive and have low resistance to supply current from the ground to the power receiver **10**. In one embodiment, the ground wires **210** may be copper or other highly-conductive metal. The end of the ground shaft may be pointed to facilitate insertion into the earth. A connection port on the ground shaft **220** is provided to electrically connect the ground antenna array **220** to the resonant transformer **30**.

The hollow cylinder **220** has external threads **25** to facilitate insertion into the ground. A rotator nut **235** is fixedly secured to the top end of the hollow shaft **220**. A square shaft **215** protrudes from the top end of the ground shaft **205** into the opening in the rotator nut **235**. FIG. **6B**. A tool **250**, shown in FIG. **7**, engages with the rotator nut **235** and square shaft **215** during insertion of the ground antennas array **200** into the ground as will be hereinafter described.

The insertion tool **250** is shown in FIG. **7**. The insertion tool **250** includes a tool body **255** having a first socket **260** on one side to fit the rotator nut **235** on the hollow cylinder **220** and a second socket **265** on the other side to fit the square shaft **215** on the ground shaft **205**. Arms **270** extend from the outer periphery of the tool body **255** for manually or mechanically turning the insertion tool **250**.

Before the antenna array **200** is deployed, the ground wires **210** are wound around the ground shaft **205** with the free ends protruding slightly from respective openings **230** in the hollow cylinder **220** to a distance not to exceed one half (½) the depth of the external threads **225** on the hollow cylinder **220**. FIG. **6B** illustrates the ground antenna array **200** before deployment. FIG. **6C** illustrates the ground antenna array in a deployed configuration.

Installation of the ground antenna array **200** is performed in two stages. In the first stage, a hole slightly smaller in diameter than the threads **235** of the hollow cylinder **220** is drilled into the Earth to a depth matching the length of the hollow cylinder **220** or slightly longer. The hole is filled with water and the water is allowed to soak into the soil. After the ground is softened, the hollow cylinder **220** is rotated using the insertion tool **250** to insert the ground antenna array **200** into the ground. The first socket **260** of the insertion tool **250** is engaged with the rotator nut **230** and the insertion tool **250** is turned by hand or a mechanized rotating shaft fitted and attached to the tool arms **270** to thread the ground assembly into the hole. During the initial insertion of the ground antenna array **200**, the ground shaft **205** is fixed to the hollow shaft **220** and rotates with the hollow shaft. The hollow cylinder **220** is rotated until it reaches the full depth of the hole.

Once the ground antenna array **200** has been fully inserted into the earth, the insertion tool **250** is flipped over and the second socket **265** of the insertion tool **250** is engaged with the square shaft **215**. The insertion tool **250** is turned by hand or a mechanized rotating shaft fitted and attached to the tool arms **270** to rotate the ground shaft **205**. During the second phase, the ground shaft **205** rotates freely inside the hollow cylinder **220**. Rotation of the ground shaft **205** causes the reinforced ground wires **210** to extend radially into the earth. The ground shaft **220** is rotated until the ground wires are fully extended. The ends of the ground wires may be sharpened to aid in the extension of the ground wires during the second phase.

After the ground antenna array **200** is deployed, a connection cable **280** is attached to a connection port **240** on the ground shaft **220** to electrically connect the ground antenna array **220** to the resonant transformer **30** in the power receiver **10**.

What is claimed is:

1. A power receiver for extracting electrical energy from the earth's electric field, said power receiver comprising:

a resonant transformer connected to a ground terminal disposed below the surface of the earth;

an impulse generator for generating and applying a high voltage electrical impulse to a primary winding of the resonant transformer to induce current flow from the ground terminal through the primary winding of the transformer; and

a power conversion circuit connected to a secondary winding of the resonant transformer to convert electrical current flowing through the secondary winding to a desired form.

2. The power receiver of claim **1** wherein a resonant frequency of the resonant transformer is below 200 Hz.

3. The power receiver of claim **1** wherein the resonant transformer comprises a ferro-resonant transformer.

4. The power receiver of claim **1** further comprising an elevated terminal, and wherein the primary winding of the resonant transformer is connected between the ground terminal and elevated terminal.

5. The power receiver of claim **4** wherein said resonant transformer includes a capacitor connected in series with the primary winding between the impulse generator and the elevated terminal.

6. The power receiver of claim **4** comprising multiple resonant transformers having primary windings connected in parallel between the ground terminal and the elevated terminal.

7. The power receiver of claim **6** wherein the resonant transformers have different resonant frequencies.

8. The power receiver of claim **7** wherein the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

9. The power receiver of claim **1** wherein the impulse generator comprises:

a pulse generator for generating low voltage pulses;

a step-up transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses;

a spark gap connected between the step-up transformer and the primary winding of the resonant transformer to generate a spark responsive to the high voltage impulses from the step-up transformer.

10. The power receiver of claim **1** wherein the impulse generator comprises a solid state spark generator.

11. The power receiver of claim **1** wherein said resonant transformer includes a capacitor connected in parallel with the primary winding.

12. A power receiver for extracting electrical energy from the earth's electric field, said power receiver comprising:

a resonant circuit connected to a ground terminal disposed below the surface of the earth, said resonant circuit having a resonant frequency below 200 Hertz

an impulse generator for generating and applying a high voltage electrical impulse to the resonant circuit to induce current flow from the ground terminal through the resonant circuit; and

a power conversion circuit connected to the resonant circuit to convert electrical current flowing through the resonant circuit to a desired form.

13. The power receiver of claim **12** wherein the impulse generator comprises:

a pulse generator for generating low voltage pulses;

a step-up transformer for converting the low voltage pulses provided by the pulse generator to high voltage impulses;

a spark gap connected between the step-up transformer and resonant circuit to generate a spark responsive to the high voltage impulses from the step-up transformer.

14. The power receiver of claim **12** wherein the resonant circuit comprises a resonant transformer having a primary winding, a secondary winding, and resonant capacitor connected in parallel with the primary winding.

15. The power receiver of claim **12** wherein the resonant circuit comprises a resonant transformer having a primary winding, a secondary winding, and resonant capacitor connected in series with the primary winding.

16. The power receiver of claim **12** wherein the resonant circuit comprises multiple resonant transformers having primary windings connected in parallel.

17. The power receiver of claim **16** wherein the resonant transformers have different resonant frequencies.

18. The power receiver of claim **17** wherein the resonant frequencies of the resonant transformers are matched to respective Schumann resonances.

* * * * *

US 20030193319A1

(19) **United States**

(12) **Patent Application Publication**
Wood et al.

(10) **Pub. No.: US 2003/0193319 A1**

(43) **Pub. Date: Oct. 16, 2003**

(54) **ION POWERED PLATFORM**

(76) Inventors: **James Rick Wood**, Grapevine, TX (US); **Mark K. Browder**, Frisco, TX (US)

Correspondence Address:
WILLIAMS, MORGAN & AMERSON, P.C.
10333 RICHMOND, SUITE 1100
HOUSTON, TX 77042 (US)

(21) Appl. No.: **10/121,226**

(22) Filed: **Apr. 12, 2002**

Publication Classification

(51) **Int. Cl.^{7}** **H02P 9/00**
(52) **U.S. Cl.** .... **322/2 A**

(57) **ABSTRACT**

An apparatus and method for powering a platform from an ambient electrostatic gradient are disclosed. The apparatus includes an ion powered platform comprises a platform body, an electrical load carried by the platform body; and an electrically conductive probe. The electrically conductive probe is electrically connected to the electrical load at a first point thereon and adapted to generate electrical energy from an ambient electrostatic field gradient. The electrically conductive probe extends from the electrical load such that a second point thereon is vertically displaced from the first point and on the opposite connection of the load. The method includes comprises positioning a first point on an electrically conductive probe at least a predetermined vertical distance from a second point on the electrically conductive probe at which the electrically conductive probe is electrically connected to the electrical load; initiating relative movement between an ambient atmosphere and the electrically conductive probe; and supplying energy generated by an electrical potential created by initiating the relative movement to the electrical load.

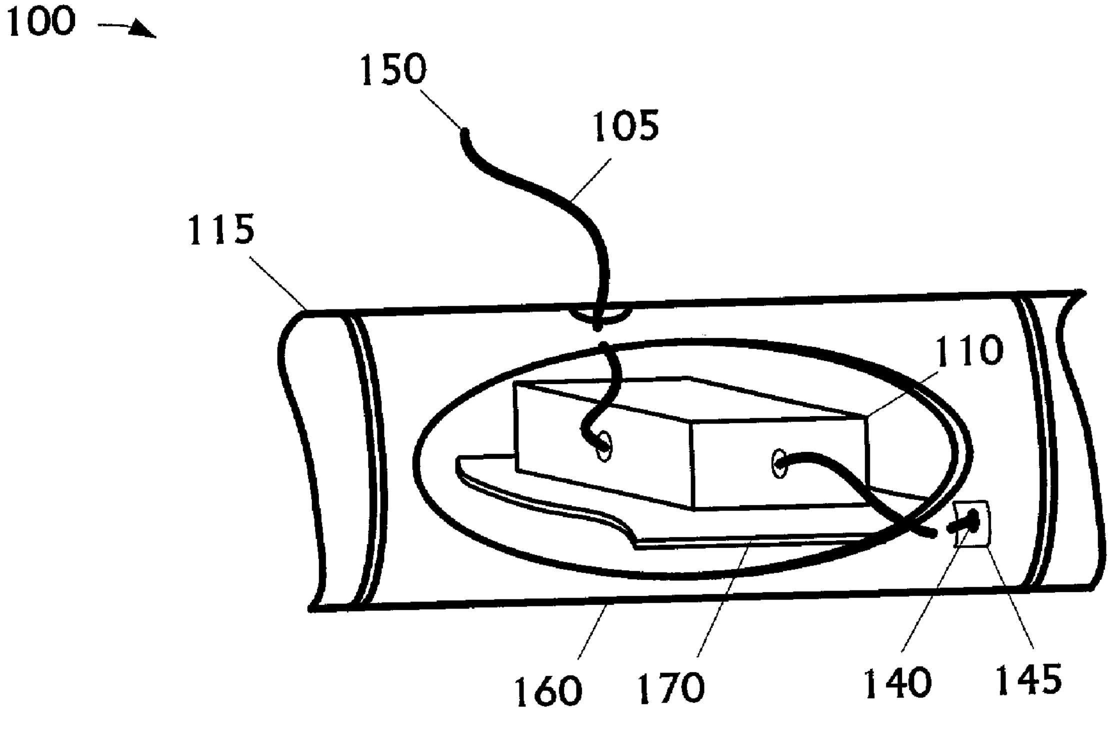

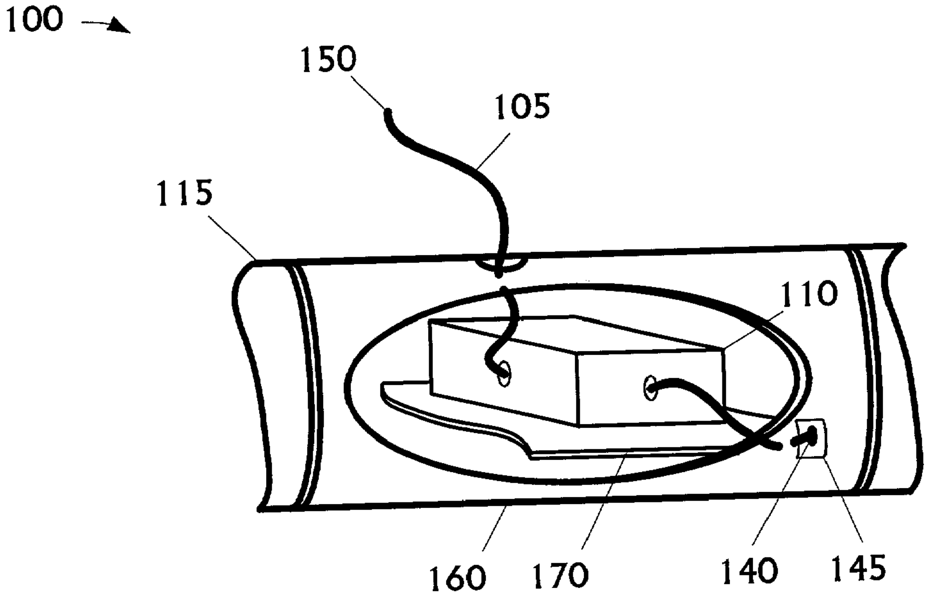

FIG. 1

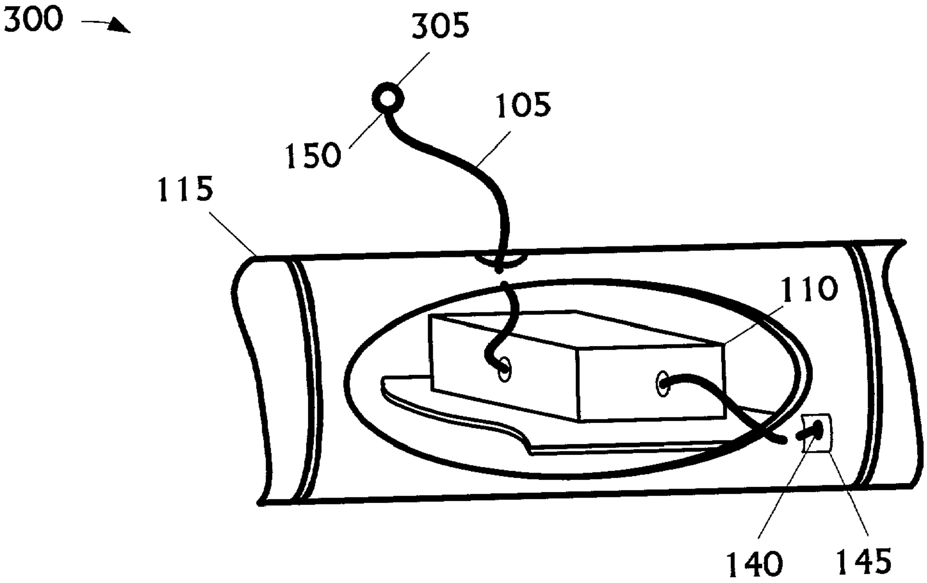

FIG. 3

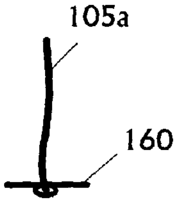

FIG. 2A

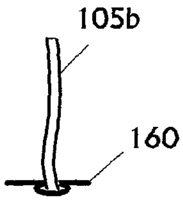

FIG. 2B

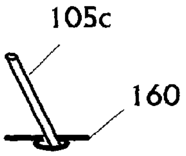

FIG. 2C

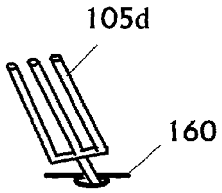

FIG. 2D

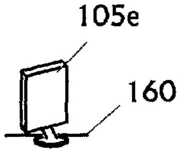

FIG. 2E

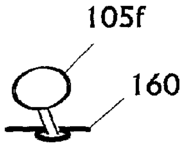

FIG. 2F

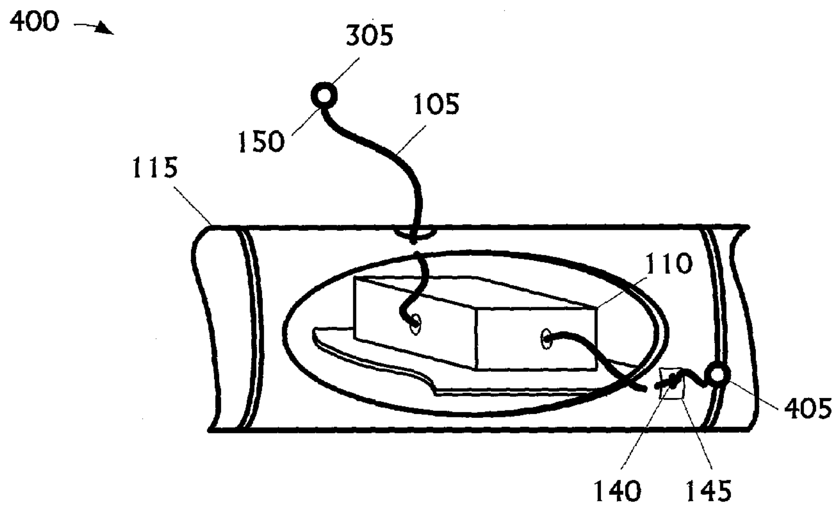

FIG. 4

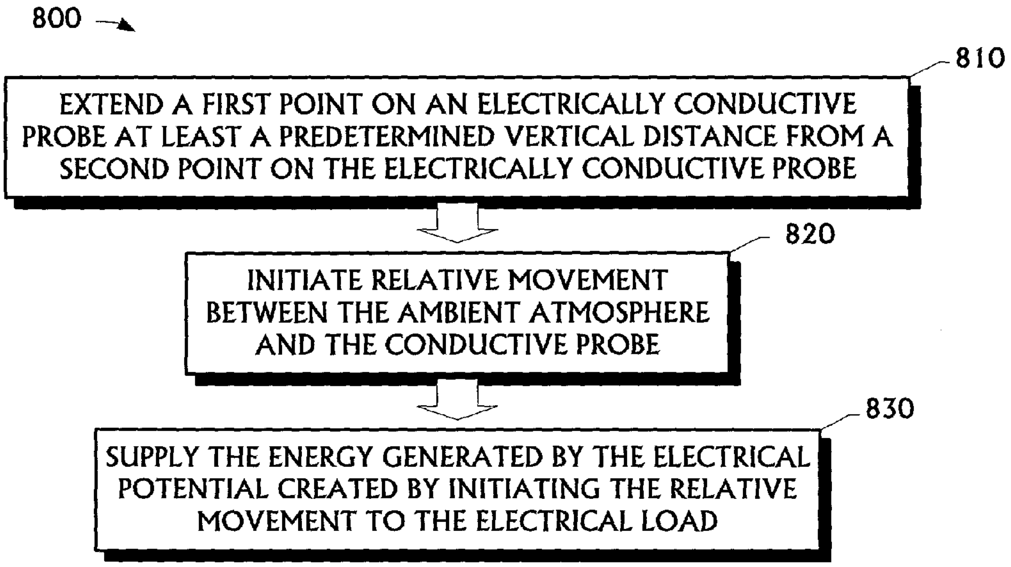

EXTEND A FIRST POINT ON AN ELECTRICALLY CONDUCTIVE PROBE AT LEAST A PREDETERMINED VERTICAL DISTANCE FROM A SECOND POINT ON THE ELECTRICALLY CONDUCTIVE PROBE

INITIATE RELATIVE MOVEMENT BETWEEN THE AMBIENT ATMOSPHERE AND THE CONDUCTIVE PROBE

SUPPLY THE ENERGY GENERATED BY THE ELECTRICAL POTENTIAL CREATED BY INITIATING THE RELATIVE MOVEMENT TO THE ELECTRICAL LOAD

FIG. 8

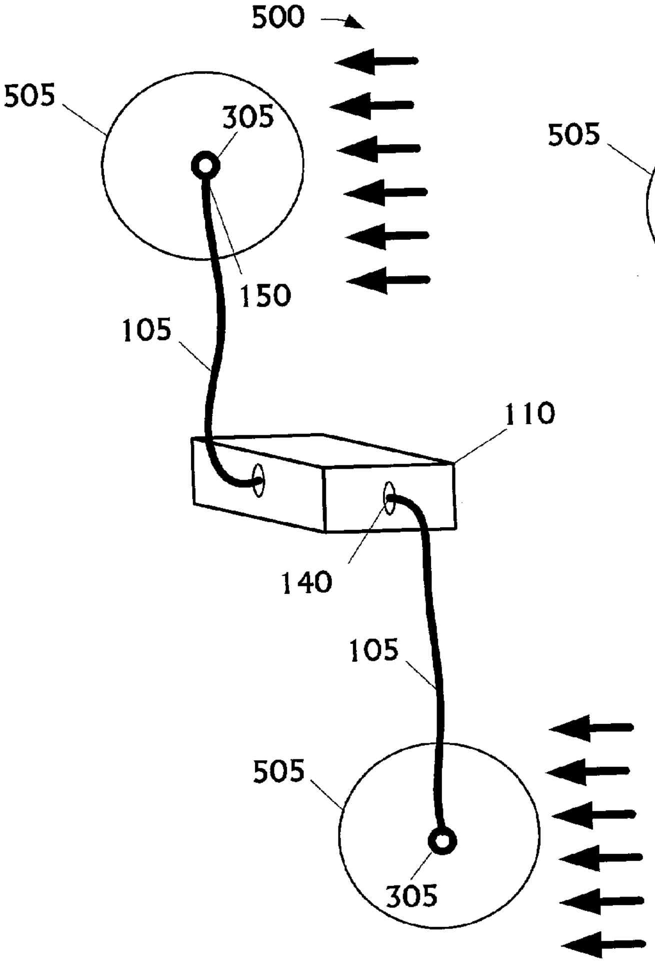

FIG. 5A

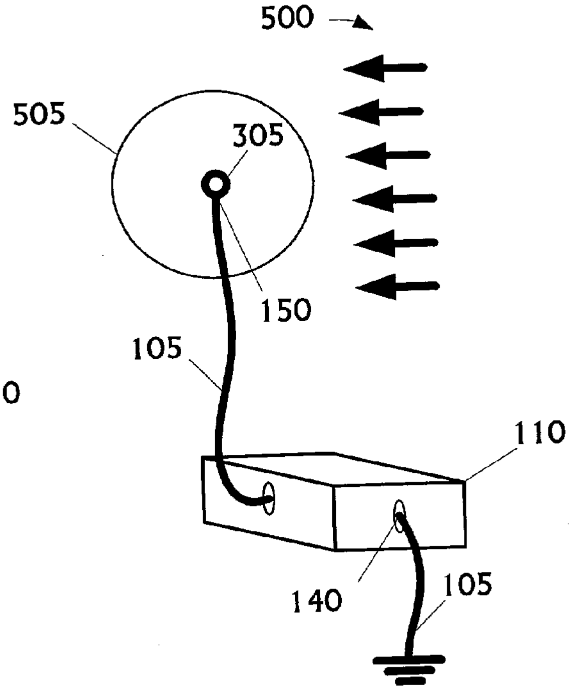

FIG. 5B

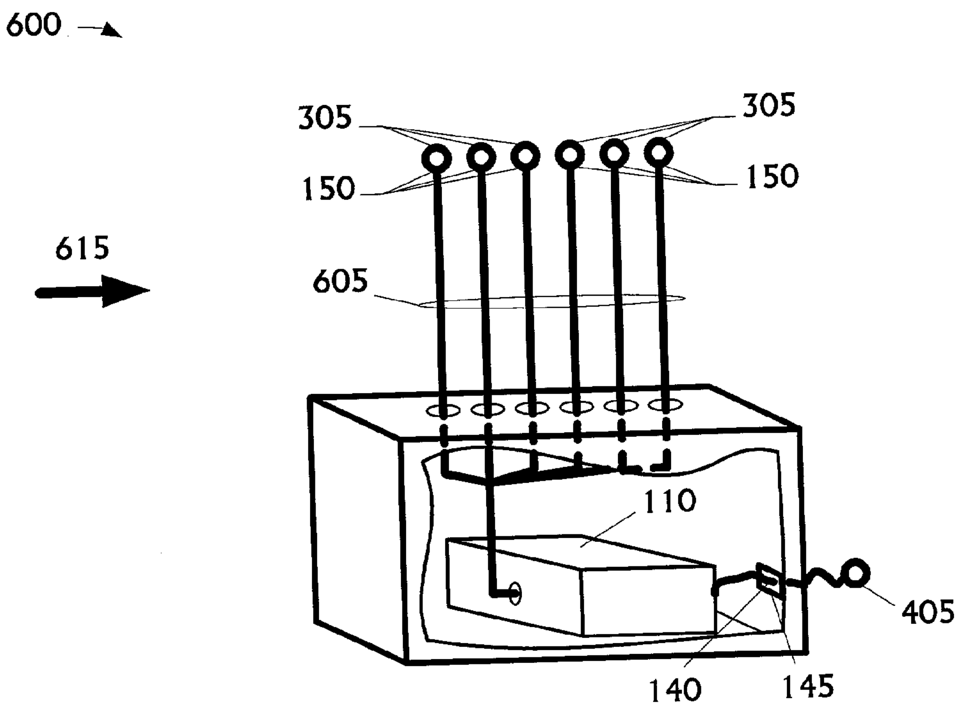

FIG. 6A

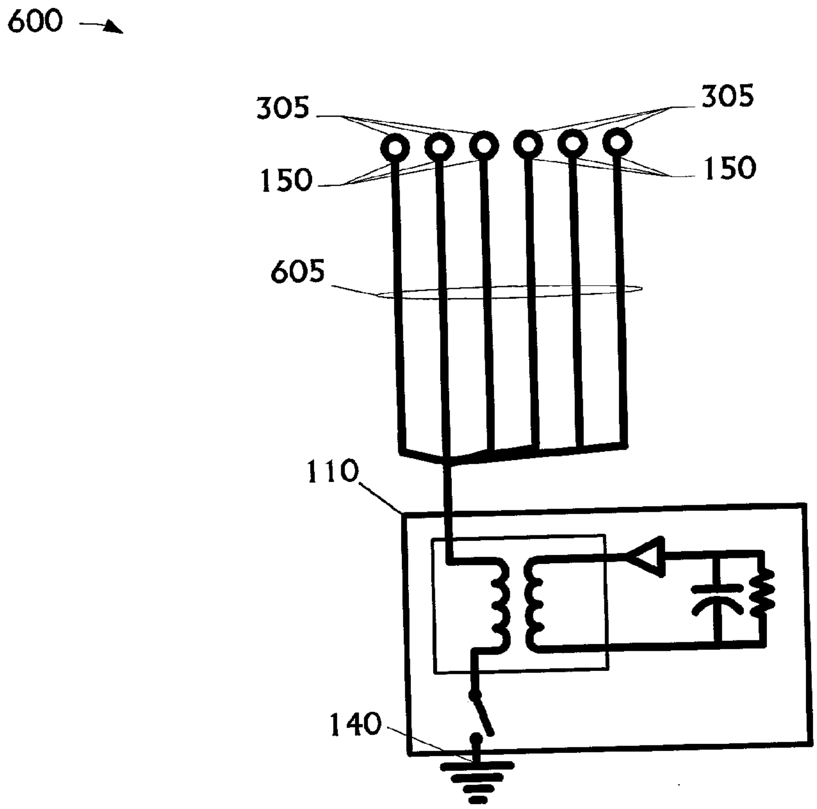

FIG. 6B

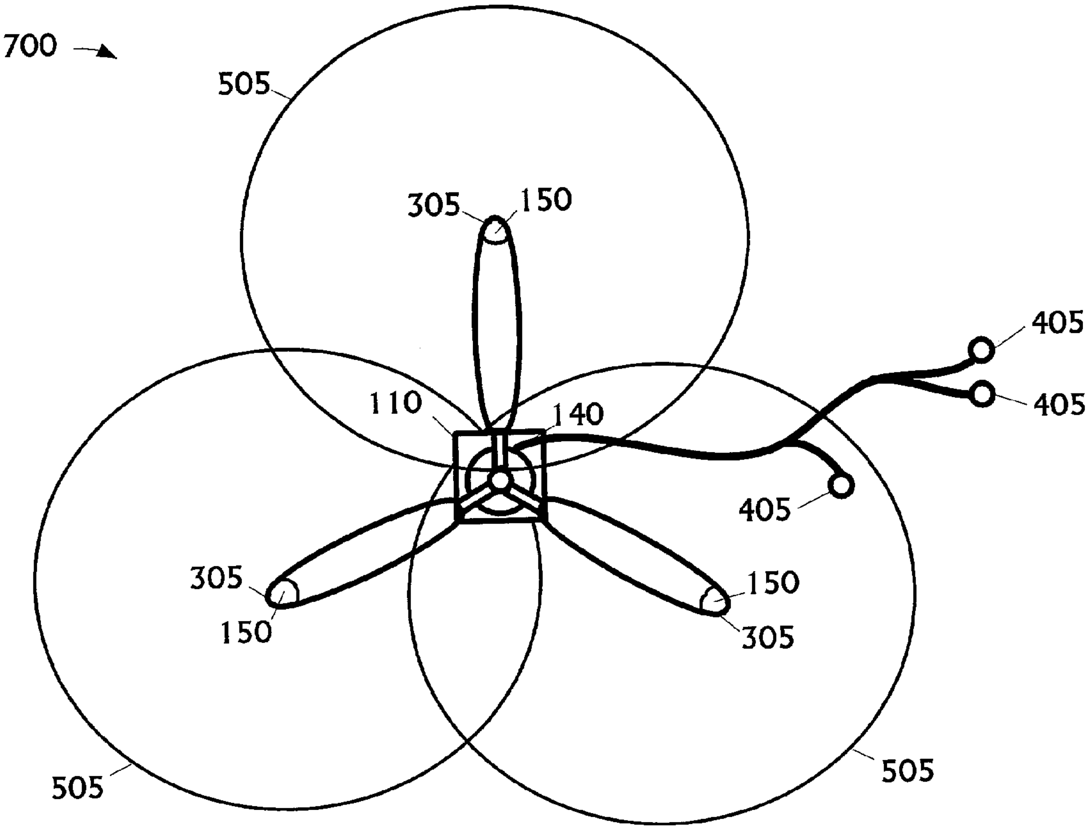
700
505
305
150
405
405
110
140
405
305
150
150
305
505
505

FIG. 7A

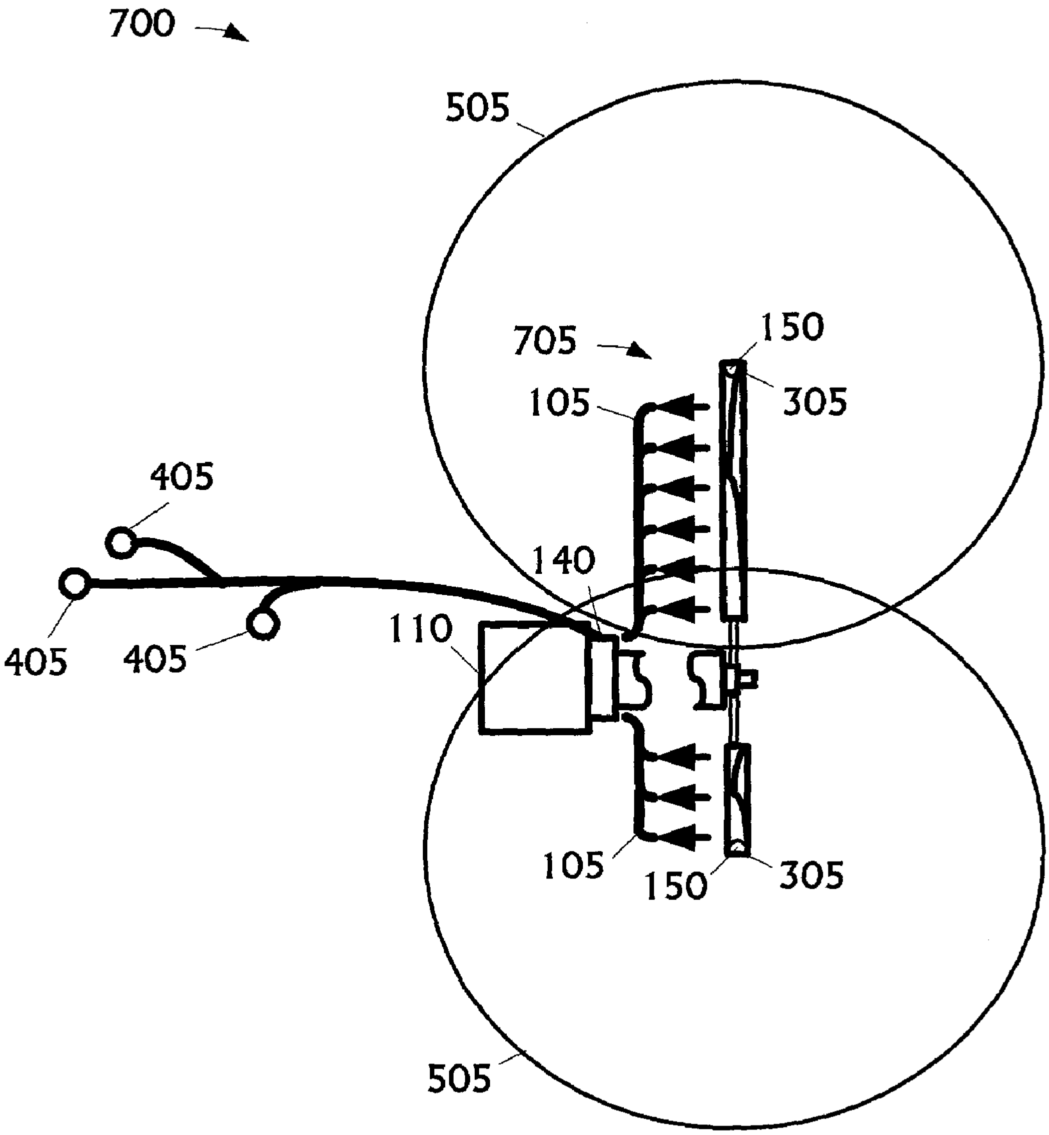
700
505
705
150
105
305
405
140
110
405
405
105
150
305
505

FIG. 7B

ION POWERED PLATFORM

BACKGROUND OF THE INVENTION

[0001] 1. Field of the Invention

[0002] The present invention pertains to an electronic power supply, and, more particularly, to a method and apparatus for powering an electronic device from an ambient (i.e., naturally occurring, as opposed to man-made) electrostatic field gradient.

[0003] 2. Description of the Related Art

[0004] Technological advances enable ever smaller, more capable tools and devices. Electronic components continue to shrink in size while increasing the power of their applications. One example of this phenomenon is the ubiquitous microprocessor that forms the heart of the personal computer. However, the fabrication and design techniques underpinning this phenomenon are applicable to other active and passive electronic components. At the same time, electronic components are finding their way into a broader range of applications. These factors contribute to a discernible trend toward smaller, more capable, and more powerful tools and devices in many areas.

[0005] One factor nevertheless continues to retard this trend—power. All these electronic components require electrical power. Furthermore, their power consumption typically rises as the size decreases and the capability increases. Power is typically supplied either by a battery or through an electrical cord connected to a powered electrical grid or a generator. Batteries continue to be relatively large and heavy despite recent, significant advances in battery technology. Connection to a power grid and/or generator brings numerous limitations regarding portability and convenience arising from the umbilical/electrical cord.

[0006] The present invention is directed to resolving, or at least reducing, one or all of the problems mentioned above.

SUMMARY OF THE INVENTION

[0007] The invention includes an apparatus and method for powering a platform from an ambient electrostatic gradient.

[0008] Thus, in a first aspect of the invention, an ion powered platform comprises a platform body, an electrical load carried by the platform body; and an electrically conductive probe. The electrically conductive probe is electrically connected to the electrical load at a first point thereon and adapted to generate electrical energy from an ambient electrostatic field gradient. The electrically conductive probe extends from the electrical load such that a second point thereon is vertically displaced from the first point and on the opposite connection of the load.

[0009] In a second aspect, the invention includes a method for powering an electrical load on an ion powered platform with ions. The method comprises positioning a first point on an electrically conductive probe at least a predetermined vertical distance from a second point on the electrically conductive probe at which the electrically conductive probe is electrically connected to the electrical load; initiating relative movement between an ambient atmosphere and the electrically conductive probe; and supplying energy generated by an electrical potential created by initiating the relative movement to the electrical load.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] The invention may be understood by reference to the following description taken in conjunction with the accompanying drawings, in which like reference numerals identify like elements, and in which:

[0011] **FIG. 1** conceptually depicts, in a partially sectioned perspective view, one particular embodiment of an ion powered platform constructed in accordance with the present invention;

[0012] **FIG. 2A** to **FIG. 2F** depicts alternative embodiments of the electrically conductive probe of the ion powered platform of **FIG. 1** implemented in electrically conductive solids;

[0013] **FIG. 3** conceptually depicts an alternative embodiment in which an ionization source is located at the distal end of an electrically conductive probe;

[0014] **FIG. 4** conceptually depicts another alternative embodiment employing a secondary ionization source to enhance ionization efficiency;

[0015] **FIG. 5**A and **FIG. 5B** illustrate the operation of embodiments employing a primary ionization source and a primary and a secondary ionization source, respectively;

[0016] **FIG. 6A** and **FIG. 6B** conceptually depict yet another alternative embodiment in which the ion powered platform is stationary;

[0017] **FIG. 7**A and **FIG. 7B** conceptually depict yet another alternative embodiment in which the electrically conductive probe is implemented in an ionized fluid medium; and

[0018] **FIG. 8** illustrates a method practiced in accordance with one aspect of the present invention.

[0019] While the invention is susceptible to various modifications and alternative forms, the drawings illustrate specific embodiments herein described in detail by way of example. It should be understood, however, that the description herein of specific embodiments is not intended to limit the invention to the particular forms disclosed, but on the contrary, the intention is to cover all modifications, equivalents, and alternatives falling within the spirit and scope of the invention as defined by the appended claims.

DETAILED DESCRIPTION OF THE INVENTION

[0020] Illustrative embodiments of the invention are described below. In the interest of clarity, not all features of an actual implementation are described in this specification. It will of course be appreciated that in the development of any such actual embodiment, numerous implementation-specific decisions must be made to achieve the developers' specific goals, such as compliance with system-related and business-related constraints, which will vary from one implementation to another. Moreover, it will be appreciated that such a development effort, even if complex and time-consuming, would be a routine undertaking for those of ordinary skill in the art having the benefit of this disclosure.

[0021] **FIG. 1** conceptually illustrates, in a partially sectioned perspective view, one particular embodiment of an ion powered platform **100** constructed in accordance with the present invention. For instance, Earth's atmosphere includes an ambient electrostatic field gradient of approximately 100 to 300 volts per meter over most of Earth's surface, although this varies with, for instance, altitude and weather conditions. The ion powered platform **100** extends an electrically conductive probe **105** at least a predetermined vertical distance to utilize this electrostatic field gradient to power an electrical load **110**. The movement of the electrically conductive probe **105** through the electrostatic field gradient causes individual molecules in the atmosphere to ionize, thereby generating an electrical potential across the electrically conductive probe **105**. This electrical potential is then harnessed to power the electrical load **110**. Note that the movement of the electrically conductive probe **105** through the electrostatic field gradient is a relative movement. In other words, the atmosphere streams past the conductive probe **105**. This can result from a motive force imparted to or by the ion powered platform **100** causing it to move. Alternatively, this can result from a wind blowing across a stationary, or slow moving, electrically conductive probe **105**.

[0022] More particularly, the ion powered platform **100** comprises a platform body **115**, the electrical load **110** carried by the platform body **115**, and the electrically conductive probe **105**. The electrically conductive probe **105** is electrically connected to the electrical load **110** and is electrically grounded at a first point **140**. In the illustrated embodiment, the first point **140** is grounded by electrically connecting it to an electrically conductive patch **145** on the fuselage **170** of the airframe **160**. Note that, if the fuselage **170** is fabricated from an electrically conductive material, the first point **140** can be grounded at any point thereon. The electrically conductive probe **105** extends from the electrical load **110** such that a second point **150** thereon is vertically displaced from the first point **140**. The electrically conductive probe **105** is adapted to generate electrical energy from an ambient electrostatic field gradient, as is discussed further below

[0023] The invention admits wide variation in the implementation of the platform body **115**. In the illustrated embodiment, the platform **100** is a flying vehicle, and so the platform body **115** includes an airframe **160** to which the electrical load **110** is mounted and a fuselage **170**. Thus, the platform body **115** of the illustrated embodiment houses the electrical load **110**, but this is not necessary to the practice of the invention. The electrical load **110** may be left exposed to the ambient environment in some alternative embodiments. Indeed, the ion powered platform **100** may be implemented in alternative embodiments as, for example, a reconnaissance drone or a ground-based robotic vehicle. The ion powered platform **100** need not even be a vehicle, but might alternatively be, e.g., a container made of wood, metal, plastic, or some other material.

[0024] The identity of the electrical load **110** will be implementation specific. In the illustrated embodiment, the electrical load **110** is a fuse capacitor. The fuse capacitor charges with electrical energy received from the electrically conductive probe **105** at the first point **140**. When the fuse capacitor charges sufficiently, the flying submunition is armed and ready to detonate. However, the invention admits variation in the implementation of the electrical load **110** as wide as that for the ion powered platform body **115**. The electrical load **110** may be, for example, in alternative embodiments, a motor, or some high-power, short-burst communications circuitry. The motor may also be a high voltage or a low voltage motor, depending on the implementation.

[0025] The invention also permits wide variation in implementing the electrically conductive probe **105**. The shape and/or geometry of the electrically conductive probe **105** is not material to the practice of the invention. Alternative embodiments are illustrated in **FIG. 2A** to **FIG. 2E**. The electrically conductive probe **105** of **FIG. 1** may be, for instance:

[0026] an electrically conductive fiber **105***a*, as shown in **FIG. 2A**;

[0027] an electrically conductive ribbon **105***b*, as shown in **FIG.** 2B, for instance, woven or otherwise fabricated from electrically conductive fibers (e.g., the electrically conductive fiber **105***a*) or a flexible material coated by metal or a metallized polymer;

[0028] an electrically conductive rod **105***c*, as shown in **FIG.** 2C, such as a metallized glass rod;

[0029] an array **105***d* of electrically conductive rods (e.g., the electrically conductive rod **105***c*) ribbons (e.g., the electrically conductive ribbon **105***b*), or electrically conductive fibers (e.g, the electrically conductive fiber **105***a*) as shown in **FIG. 2D**;

[0030] an electrically conductive sheet **105***e*, as shown in **FIG.** 2E, fabricated from, e.g., a metal or electrically conductive composite material; or

[0031] an electrically conductive ring **105***f*, as shown in **FIG.** 2F, fabricated from, e.g., a metal or an electrically conductive composite material.

[0032] Thus, the electrically conductive probe **105** may by flexible (e.g., as are the electrically conductive fiber **105***a* and electrically conductive ribbon **105***b*) or rigid (e.g., like the electrically conductive rod **105***c* and electrically conductive sheet **105***e*). Note also that the vertical displacement between the two most vertically distant points on the electrically conductive probe **105** contributes more significantly to its operation than its actual shape and/or geometry.

[0033] The vertical displacement between the two most vertically distant points on the electrically conductive probe **105** will be implementation specific. The vertical displacement will primarily be a factor of the power requirements of the electrical load and the electrostatic field gradient, applied in light of commonly accepted engineering principles. For instance, as was previously mentioned, Earth's electrostatic field gradient varies from approximately 100 to approximately 300 volts per meter in fair weather and at sea level. However, these values will vary with altitude and weather—

bad weather, for example, can greatly increase the magnitude. Note that the invention may find extraterritorial application, e.g., on Venus, Mars, or Europa, whose ambient electrostatic field gradients may vary from that of Earth's. Similarly, the power requirements of the electrical load **110** may vary over time as the ion powered platform **100** enters differing operational modes. Similarly, some applications might be more tolerant of power supply faults than are other, thereby providing greater latitude in design.

[**0034**] Another important factor is the volume of ionized air. Greater volumes tend to produce a greater electrical potential. For this reason, implementations of the electrically conductive probe **105** that maximize this volume of ionized air may be generally preferred for most embodiments. The array **105***d* in **FIG. 2D** might therefore be preferred over the electrically conductive rod **105***c* in **FIG. 2C**. The electrically conductive ribbon **105***b* in **FIG. 2B** might also generally be preferred over the electrically conductive fiber **105***a* in **FIG. 2A**. However, each implementation may be different, and this principle is not universally applicable to all embodiments.

[**0035**] Still referring to **FIG. 1**, in one particular implementation, the electrically conductive element **105** is an electrically conductive fiber, or "streamer," extending from the ion powered platform **100**. The first and second points **140**, **150** are separated by 100 meters and generate a potential difference of 10,000 volts. In the illustrated embodiment, as was mentioned, this potential is used to charge a fuse capacitor in a flying submunition. However, in alternative embodiments, a high voltage capacitor could be used to, for instance, provide current to a switching circuit to create lower voltage higher current pulses adequate to run a small motor. By utilizing this larger current capability, the ability to deliver energy for useful work is made possible.

[**0036**] Design constraints may make it difficult or impractical to provide the requisite vertical displacement in some implementations. One such embodiment **300** is shown in **FIG. 3**. In these embodiments, the ionizing effect of the electrically conductive probe **105**, shown in **FIG. 1**, can be enhanced by including an ionization source **305** at the point **150** of the electrically conductive probe **105** most distal from the electrical load **110**. For embodiments where the point **140** does not contact the electrical ground **145**, the ionization source **305** may be located anywhere on the electrically conductive probe **105**, but the point **150** most distal the electrical load **110** may be used to produce enhanced results. Where the electrically conductive probe **105** is an array (e.g., the array **105***d* in **FIG. 2D**), each element of the array can include an ionization source **305** on the end thereof. The ionization source **305** contains a low level, radioactive source of alpha or beta particles encapsulated in a suitable material. Suitable alpha emitters include, but are not limited to, isotopes of Polonium and Americium (Am). Suitable beta emitters include, but are not limited to, Tritium. Suitable encapsulating materials include, but are not limited to, glass compositions such as silicate, borosilicate, aluminasilicate, and borate. In some embodiments, the point **150** can be doped with an alpha emitter or a beta emitter during manufacture.

[**0037**] Suitable alpha and beta emitters are commercially available. For instance, suitable alpha emitters are sold under the mark RadSPHERES™ by:

MO-SCI Corporation

4000 Enterprise Drive, P.O. Box 2

Rolla, Mo. 65402-0002

573-364-2338 (ph)

573-364-9589(fax)

[**0038**] Additional information is available on the World Wide Web at <http://www.thomasregister.com/olc/mo-sci-corp/radsph.htm>. The RadSPHERES™ alpha emitters are glass microspheres of encapsulated low-level radioactive materials. Alternatively, the low-level radioactive material may be encapsulated in a bar-shaped material. However, other commercially available alpha emitters and beta emitters may also be employed.

[**0039**] Some embodiments successfully employ further ionization efficiency than that provided by the ionization source **305** in **FIG. 3**. One such embodiment **400** is conceptually illustrated in **FIG. 4**. In **FIG. 4, a** second ionization source **405** is on or near the electrical load **110** on the end thereof opposite the first point **140** where the electrically conductive probe **105** is electrically connected to the electrical load **110**. The second ionization source **405** may, like the first ionization source **305**, be any suitable alpha or beta particle emitter. In the illustrated embodiments, the first and second ionization sources **305**, **405** are of the same construction and materials, but this is not necessary to the practice of the invention.

[**0040**] Turning now to **FIG.** 5A and **FIG.** 5B, in operation, the relative movement between the electrically conductive probe **105** and the ambient atmosphere, generally indicated by the arrows **500**, ionizes the air streaming past the electrically conductive probe **105**. Hence, the potential created between the first point **140** and the second point **150** on the electrically conductive probes **105**. Each ionization source **305** creates an ionization region **505**, which is a volume of air ionized by the ionization source **305** without having to stream past the electrically conductive probe **105**. The increase in the volume of ionized air increases the efficiency of the coupling to the electrostatic field gradient in the vicinity of the ionization source **305**, which includes at least a portion of the electrically conductive probe **105**. Thus, as the ionization sources **305** increase the volume of ionized air, they increase the efficiency of the coupling to the electrostatic field gradient.

[**0041**] Note that the embodiments described above contemplate that the ion powered platforms **100, 300, 400,** move relative to Earth's atmosphere. However, this is not necessary to the practice of the invention. The same effect can be achieved by placing a stationary ion powered platform in a location with air movement. Consider, for instance, the ion powered platform **600** shown in **FIG. 6A**. The ion powered platform **600** is a sensor, but is stationary. The ion powered platform **600** includes an electrically conductive probe **605** that is an array. The elements of the array **605** are rigid, and each has an ionization source **305** electrically connected to the end thereof as was described above. The ion powered platform **600** also includes an ionization source **405** on the other side of the electrical load **110**, which is a capacitor in the illustrated embodiment. **FIG. 6B** conceptually illustrates the electrical circuit of the ion powered platform **600**, first shown in **FIG. 6**A, which includes a transformer **620**. The ion powered platform **600** is then placed in an area expected to consistently experience high winds, represented by the arrow **615**. If the winds are high and consistent enough, the fuse capacitor (i.e., the electrical load **110**) will charge and the sensor (i.e., the ion powered platform **600**) may then be activated.

[**0042**] Yet another alternative embodiment **700** is illustrated in **FIG. 7**A and **FIG. 7B**. The embodiment **700** is part of a propeller-driven propulsion system for a flying vehicle not otherwise shown. The propeller **705** includes three blades **710**. The number of blades **710** is immaterial to the practice of the invention. Each blade **710** is doped at the distal end **715** thereof to produce a primary ionization source **305**. Each primary ionization source **305** generates an ionization region **505** as described above relative to the embodiment **500** shown in **FIG. 5**. The embodiment **700** also includes a secondary ionization source **405**. The electrical load **110** in this embodiment is a motor for driving the propeller **705**. The motor may be, for example, a surface acoustic wave linear motor, an ultrasonic motor, or an electrostatic motor, depending on the implementation.

[**0043**] The embodiment **700** is launched from another flying vehicle, and so is moving relative to the atmosphere. As is best shown in **FIG. 7**B, the atmosphere streams past the primary ionization sources **305** on the distal ends **715** of the blades **710** and through the ionization regions **505**, as represented by the arrows **720**. The ionized atmosphere completes an electrical circuit between the first point **140** (i.e., the electrical ground) on the ring **725** and the second points **150** on the distal ends **715**. Thus, the electrically conductive probe **105** in this embodiment is implemented in an ionized fluid medium as opposed to an electrically conductive solid.

[**0044**] Thus, the invention includes, in another aspect, a method **800**, illustrated in **FIG. 8**, for powering an electrical load (e.g., the electrical load **110** in **FIG. 1**) on a ion powered platform (e.g., the ion powered platforms **100**, **300**, **400**, **600**, in **FIG. 1**, **FIG. 3**, **FIG. 4**, and **FIG. 6**A, respectively) with ions. The method **800** begins by extending a first point on an electrically conductive probe at least a predetermined vertical distance from a second point on the electrically conductive probe at which the electrically conductive probe is electrically connected to the electrical load (at **810**). This extension may occur during fabrication, between fabrication and deployment, or after deployment. Next, the method **800** initiates relative movement between the ambient atmosphere and the electrically conductive probe (at **820**). Note that the movement is relative. As is apparent from the discussion above, the relative to some fixed point, the ion powered platform may move through the atmosphere, the atmosphere may move over the ion powered platform, or both the ion powered platform and atmosphere may move. Finally, the method **800** supplies the energy generated by the electrical potential created by initiating the relative movement to the electrical load (at **830**).

[**0045**] This concludes the detailed description. The particular embodiments disclosed above are illustrative only, as the invention may be modified and practiced in different but equivalent manners apparent to those skilled in the art having the benefit of the teachings herein. Furthermore, no limitations are intended to the details of construction or design herein shown, other than as described in the claims below. It is therefore evident that the particular embodiments disclosed above may be altered or modified and all such variations are considered within the scope and spirit of the invention. Accordingly, the protection sought herein is as set forth in the claims below.

What is claimed:

1. An ion powered platform, comprising:

a platform body;

an electrical load carried by the platform body; and

an electrically conductive probe electrically connected to the electrical load at a first point thereon and adapted to generate electrical energy from an ambient electrostatic field gradient, the electrically conductive probe extending from the electrical load such that a second point thereon is vertically displaced from the first point and on the opposite connection of the load.

2. The ion powered platform of claim 1, further comprising an ionization source electrically coupled to the second end of the probe.

3. The ion powered platform of claim 2, wherein the ionization source comprises one of an alpha-particle emitter and a beta-particle emitter.

4. The ion powered platform of claim 3, wherein the alpha-particle emitter includes one of Polonium and Americium.

5. The ion powered platform of claim 4, wherein the Polonium or Americium is encapsulated in one of silicate, borosilicate, aluminasilicate, and borate.

6. The ion powered platform of claim 3, wherein the beta-particle emitter includes Tritium.

7. The ion powered platform of claim 2, further comprising a second ionization source affixed to the electrical load at a third point on the electrical load distal from the first point.

8. The ion powered platform of claim 7, wherein the second ionization source comprises one of an alpha-particle emitter and a beta-particle emitter.

9. The ion powered platform of claim 8, wherein the alpha-particle emitter includes one of Polonium and Americium.

10. The ion powered platform of claim 9, wherein the Polonium or Americium is encapsulated in one of silicate, borosilicate, aluminasilicate, and borate.

11. The ion powered platform of claim 8, wherein the beta-particle emitter includes Tritium.

12. The ion powered platform of claim 1, wherein the electrically conductive probe comprises one of a fiber, a ribbon, a rod, a ring, and a sheet.

13. The ion powered platform of claim 12, wherein the fiber comprises an electrically conductive carbon filament.

14. The ion powered platform of claim 13, wherein at least one of the ribbon, the rod, the ring, and the sheet comprises a plurality of electrically conductive carbon filaments, a metallized glass, a metal coating, or a metallized polymer.

15. The ion powered platform of claim 1, wherein the electrically conductive probe comprises

a first ionization source;

a second ionization source; and

an ionized fluid medium between the first and second ionization sources.

16. The ion powered platform of claim 15, wherein at least one of first and second ionization sources comprises one of an alpha-particle emitter and a beta-particle emitter.

17. The ion powered platform of claim 16, wherein the alpha-particle emitter includes one of Polonium and Americium.

18. The ion powered platform of claim 17, wherein the Polonium or Americium is encapsulated in one of silicate, borosilicate, aluminasilicate, and borate.

19. The ion powered platform of claim 16, wherein the beta-particle emitter includes Tritium.

20. An ion powered platform, comprising:

a platform body;

an electrical load carried by the platform body; and

means for generating electrical energy from an ambient electrostatic field gradient.

21. The ion powered platform of claim 20, further comprising a first means for ionizing a region of air, the ionizing means being electrically coupled to the generating means.

22. The ion powered platform of claim 21, wherein the ionizing means comprises one of means for emitting alpha-particles and means for emitting beta-particles.

23. The ion powered platform of claim 21, further comprising second means for ionizing a region of air electrically coupled to the generating means.

24. The ion powered platform of claim 20, wherein the generating means comprises one of a fiber, a ribbon, a rod, a ring, and a sheet.

25. The ion powered platform of claim 20, wherein the generating means comprises:

first means for ionizing a region of air;

second means for ionizing a region of air; and

an ionized fluid medium between the first and second ionizing means.

26. A method for powering an electrical load on an ion powered platform with ions, comprising:

positioning a first point on an electrically conductive probe at least a predetermined vertical distance from a second point on the electrically conductive probe at which the electrically conductive probe is electrically connected to the electrical load;

initiating relative movement between an ambient atmosphere and the electrically conductive probe; and

supplying energy generated by an electrical potential created by initiating the relative movement to the electrical load.

27. The method of claim 26, further comprising providing an ionization source as the first point.

28. The method of claim 27, further comprising providing a second ionization source on the side of the electrical load opposite the second point.

29. The method of claim 26, further comprising providing an ionization source on the side of the electrical load opposite the second point.

30. The method of claim 26, wherein extending the first point on the electrically conductive probe at least the predetermined vertical distance from the second point includes extending on deployment of the ion powered platform.

31. The method of claim 26, wherein extending the first point on the electrically conductive probe at least the predetermined vertical distance from the second point includes extending on manufacture of the ion powered platform.

32. The method of claim 26, wherein initiating the relative movement includes placing the ion powered platform in motion.

33. The method of claim 26, wherein initiating the relative movement includes deploying the ion powered platform in expectation of a desirable wind pattern.

* * * * *

US 20030209636A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2003/0209636 A1**

St.Clair (43) **Pub. Date: Nov. 13, 2003**

(54) **BOBBIN ELECTROMAGNETIC FIELD PROPULSION VEHICLE**

(76) Inventor: **John Quincy St.Clair**, San Juan, PR (US)

Correspondence Address:
John St.Clair
Hyperspace Research Institute
52 Kings Court, 4A
San Juan, PR 00911 (US)

(21) Appl. No.: **10/142,583**

(22) Filed: **May 9, 2002**

Publication Classification

(51) **Int. Cl.^{7}** **B64G 1/32**

(52) **U.S. Cl.** **244/166**

(57) **ABSTRACT**

This invention relates to a spacecraft which generates its own magnetic moment and magnetic field gradient in order to produce lift on the hull. The magnetic moment is generated by a large area solenoid located in the hull. A toroidal core wrapped with electrical bobbins at intervals along said core produces a traveling magnetic wave along its surface. This magnetic wave creates a spacetime curvature, similar to a tilted plate, which causes the formation of a magnetic field gradient. Power is not critical because the system uses a magnetic vortex wormhole generator to lower the speed of light in order to efficiently create highly relativistic fields due to Lorentz transformation.

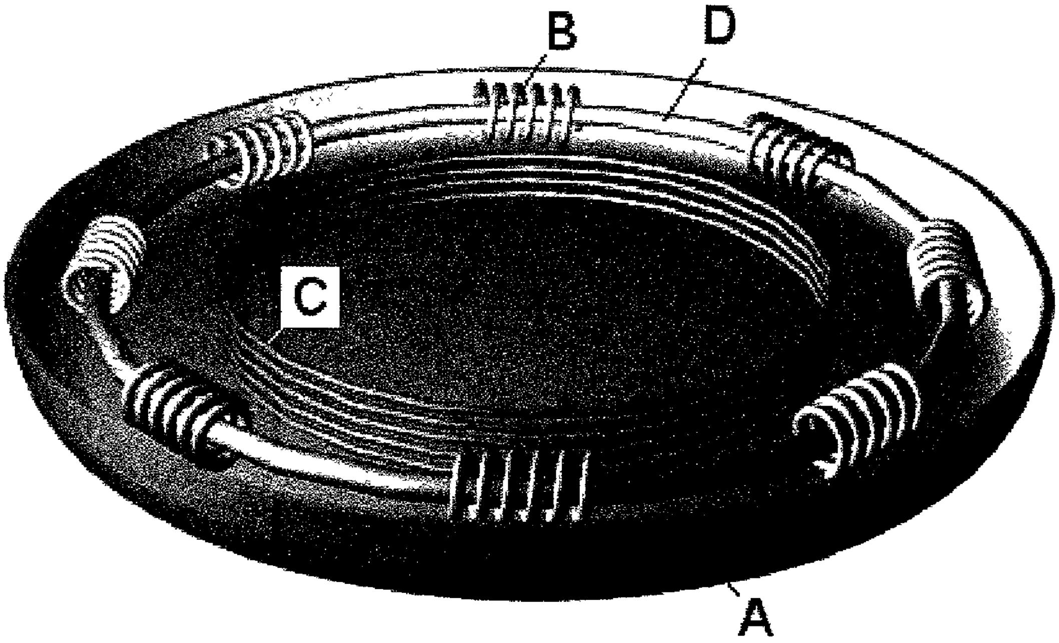

Figure 1

$$\text{Force} = \mu \cdot \frac{\partial B_z}{\partial z} = \text{amp} \cdot \text{m}^2 \frac{\text{kg}}{\text{sec} \cdot \text{coul} \cdot \text{m}} = \frac{\text{coul}}{\text{sec}} \cdot \text{m}^2 \cdot \frac{\text{kg}}{\text{sec} \cdot \text{coul} \cdot \text{m}} = \frac{\text{kg} \cdot \text{m}}{\text{sec}^2} = \text{newton}$$

Figure 2

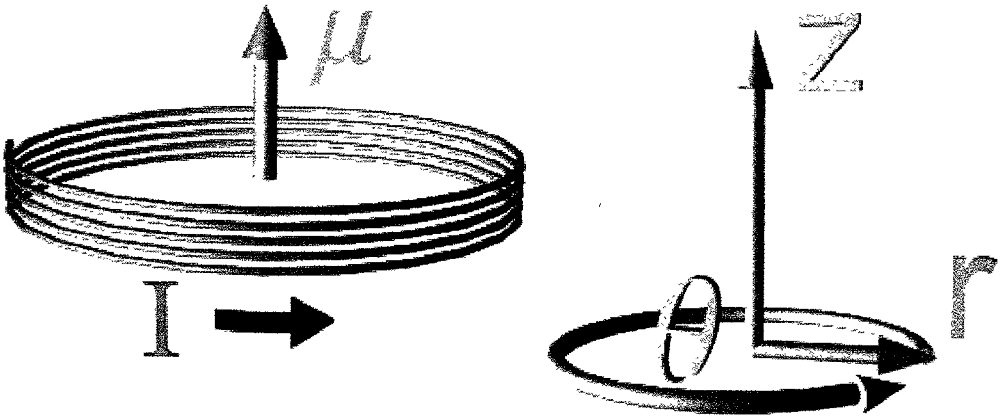

Figure 3

$$B_Z = \mu n I$$

Figure 4

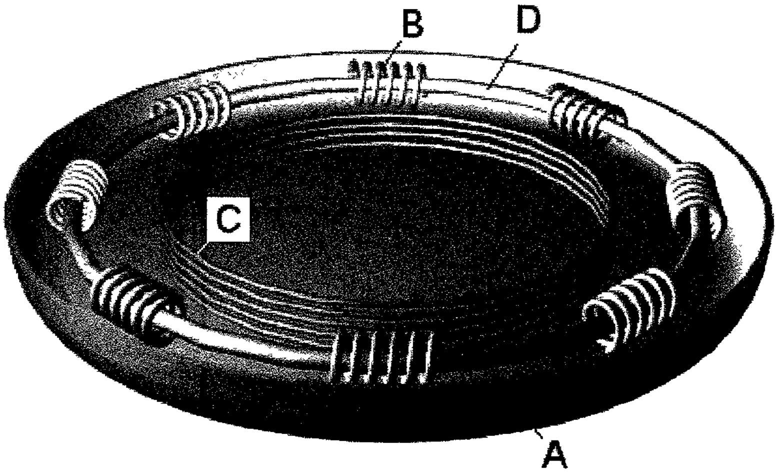

Figure 5

$$v = \sqrt{\frac{\omega}{\sigma\mu}}$$

Figure 6

$$g_{\alpha\phi} = \begin{array}{c|cccc|} & t & r & \theta & z \\ t & -1 & 0 & \mathrm{Sin}[\theta - t] & 0 \\ r & 00 & 1 & 0 & 0 \\ \theta & \mathrm{Sin}[\theta - t] & 0 & r^2 & 0 \\ z & 0 & 0 & 0 & 1 \end{array}$$

Figure 7

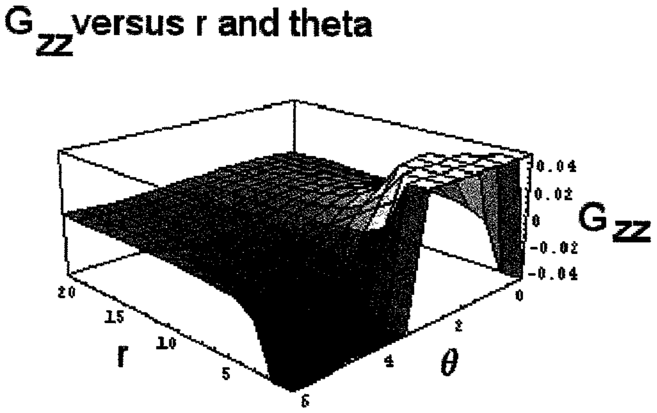
G_{zz} versus r and theta

Figure 8

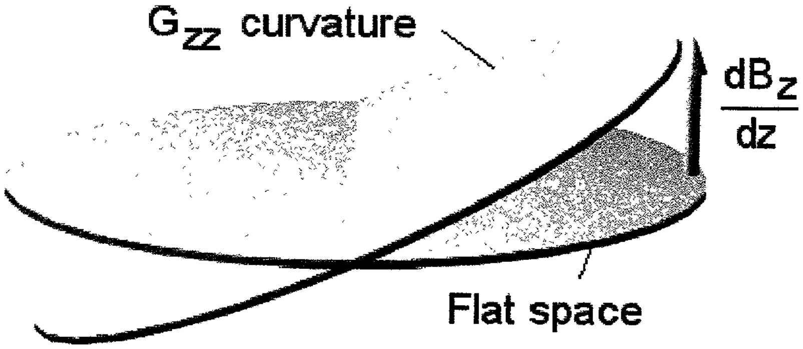
G_{zz} curvature
$\frac{dB_z}{dz}$
Flat space

BOBBIN ELECTROMAGNETIC FIELD PROPULSION VEHICLE

BRIEF SUMMARY OF THE INVENTION

[0001] This invention, which is the subject of my present application, is comprised of a toroidal core around which are wound a plurality of electromagnetic bobbins. The bobbins are pulsed electrically to create an amplified magnetic wave that travels around the core. At the same time, a circular electrical conductor carrying direct current creates a magnetic field around its area which results in the formation of a magnetic moment. This magnetic moment, together with the spacetime curvature distortion created by the traveling magnetic wave, produces a lift force on the vehicle.

REFERENCE PAPERS

[0002] Levitron, Hones, U.S. Pat. No. 5,404,062.

[0003] *Geometry of Electromagnetic Systems,* Paul Hammond, page 179.

BACKGROUND OF THE INVENTION

[0004] The idea for this invention comes from experiments I have done using thin transformer laminations wound in intervals with bobbins of wire connected to a frequency generator. Pulsing the bobbins electrically creates a slow traveling magnetic wave along the surface of the lamination. The velocity of the wave, as shown by Hammond in the reference paper, is the square root of the frequency divided by the conductivity and permeability of the material. From Einstein's General Theory of Relativity, this type of wave around the circumference creates a spacetime curvature distortion in the vertical direction that looks like a tilted plate. The magnetic field traveling around the circumference has to follow this curvature. This creates a magnetic gradient in the z-direction which together with the magnetic moment developed by a solenoid produces lift on the vehicle.

SUMMARY OF THE INVENTION

[0005] It is the object of this invention to create a magnetic field gradient and magnetic moment in order to produce a lift force on the hull of a vehicle. The magnetic field gradient is produced by a traveling magnetic wave which produces a tilting-plate spacetime curvature around the hull. The magnetic moment is created by a simple circular wire carrying direct current around its area.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0006] Not Applicable.

A BRIEF DESCRIPTION OF THE DRAWINGS

[0007] **FIG. 1**. The lift force on the vehicle is equal to a magnetic moment times the magnetic field gradient.

[0008] **FIG. 2**. Magnetic moment μ created by direct current I flowing counterclockwise around the wire coil. Cylindrical coordinates are shown to the right.

[0009] **FIG. 3**. The wire coil solenoid creates a magnetic field in the z-direction equal to the permeability of space times the number of turns per length of the solenoid times the current in the windings.

[0010] **FIG. 4**. Perspective view of bobbin spacecraft.

[0011] **FIG. 5**. Magnetic wave velocity along core.

[0012] **FIG. 6**. The g metric tensor in cylindrical coordinates with the sinusoidal wave function in the $\{t,\theta\}$ slots.

[0013] **FIG. 7**. The spacetime curvature G_{zz} in the vertical direction created by the magnetic wave traveling around the core.

[0014] **FIG. 8**. Tilted plate spacetime curvature showing magnetic gradient.

DETAILED DESCRIPTION OF THE INVENTION

[0015] 1. Referring to the equation in **FIG. 1**, the lift force on the bobbin spacecraft is equal to its magnetic moment times its magnetic field gradient. The magnetic moment has units of electrical current, measured in amps, times the area enclosed by the current. Thus the units of the magnetic moment are amp–m^2. The magnetic moment can be created by a large, circular coil of wire carrying direct current I as shown in **FIG. 2**. The spacetime cylindrical coordinates $\{t, r, \theta, z\}$ are shown to the right of the drawing where t is time, r the radius, the horizontal angle θ, and the vertical height z.

[0016] 2. The wire coil solenoid also produces a magnetic field in the vertical z-direction equal to the permeability of space μ times the number of turns per unit length of the solenoid n, times the current I in the winding. The equation is shown in **FIG. 3** which can be developed from Ampere's law that the magnetic field around a loop is equal to the current passing through the loop.

[0017] 3. Referring to **FIG. 4**, the bobbin spacecraft consists of a large, horizontal, highpermeability toroidal core (D) wrapped at intervals with electrical bobbins (B) which can be pulsed electrically to create a traveling magnetic wave around the core. Interior to this core is a large direct current solenoid (C) which produces said magnetic field in the z-direction. These devices are enclosed in a circular hull (A) containing the coils on the outer rim and a cabin area in the center.

[0018] 4. Referring to **FIG. 5**, the velocity of the magnetic wave on the surface of the core is equal to the square root of the wave frequency ω divided by the conductivity a times the permeability μ of the core material. When the first bobbin is pulsed, a wave starts to propagate along the surface of the core. As the wave passes the second bobbin in sequence, another electrical pulse is generated to amplify the wave. After many cycles, the wave traveling around the core becomes larger and larger in amplitude.

[0019] 5. A traveling wave has a wave function equal to a sinusoidal function with an argument of the angle θ around the periphery less the time t, or $\text{Sin}[\theta-\omega t]$.

[0020] 6. In gravitational physics, there is a g metric tensor which is a measure of length in spacetime coordinates. When mass or electromagnetic fields are involved in a certain region of space, a curvature of space is created. The curvature of space can then be calculated directly from this metric tensor. The result is Einstein's G curvature tensor which shows the spacetime distortion. The g metric tensor is a 4 by 4 matrix having rows and columns equal to time t, radius r, angle theta, and vertical height z. The diagonal from top left to bottom right has a signature equal to $\{-1, 1, r^2, 1\}$.

[0021] 7. Because the magnetic wave traveling around the core is varying with time t in the theta direction, the wave function has to go into the $\{t,\theta\}$ and $\{\theta,t\}$ slots of the metric tensor, as shown in **FIG. 6**.

[0022] 8. From this metric tensor, Einstein's G curvature tensor is calculated using a general relativity software program. The spacetime curvature in the vertical z-direction is contained in the G_{zz} component. A 3D plot of this curvature with respect to radius and theta is shown in **FIG. 7**. The axis on the right side is the angle around the periphery, and the axis on the left is the radius which goes from 0 to 20 meters. The vertical axis of the graph is the spacetime curvature in the vertical z-direction. The center of the vertical axis is zero. At a small radius, there is a sinusoidal curvature which is positive from 0 to π and which is negative from π to 2π. What this looks like is a tilted plate as depicted in **FIG. 8**.

[0023] 9. In flat spacetime with no electromagnetic fields or mass, the curvature would be the horizontal plate as seen **FIG. 8**. Due to the traveling magnetic wave, the spacetime curvature looks like the tilted plate. The flat-space magnetic field was pointing up in the vertical direction, having no divergence and no gradient. In curved spacetime, however, the magnetic field becomes tilted just like the plate and a magnetic gradient dB_z/dz is created. This spacetime curvature gradient of the magnetic field times the magnetic moment of the second coil produces a lift force on the hull which is anchored to these coils.

What I claim as my invention is:

1. A spacecraft that generates a magnetic moment and a magnetic field gradient in the vertical direction in order to create a lift force on the hull.

2. A toroidal core wrapped with electrical bobbins at intervals around the core whose purpose is to create and amplify a magnetic wave that travels along the surface of the core.

3. A circular, direct-current carrying solenoid, located radially inside the toroidal core, to create the magnetic moment.

4. Said magnetic surface wave creates a spacetime curvature, similar to a tilted plate, which produces a magnetic field gradient in the vertical direction.

5. A circular hull, with interior cabin, electrical power supply to drive the coils, and a computer to calculate and sequence the activation of the electrical bobbins.

6. Magnetic sensors, in the form of small coils, located on the toroidal core which can detect the velocity and position of the traveling magnetic wave as feedback to the computer control system.

7. A magnetic vortex generator, either rotating magnet or dual coil, to produce a wormhole through which low linear mass and low speed of light hyperspace energy can enter the hull in order to create highly relativistic electromagnetic fields.

8. A variable current generator to modulate the direct current in the coil so as to produce a variable magnetic moment that can control lift, hover and descent.

9. An electrical power supply using mechanical flywheel stored energy together with a large area capacitor, resupplied with energy by solar cells located on the outer surface of the hull.

* * * * *

US 20040057255A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2004/0057255 A1**

Patrick et al. (43) **Pub. Date: Mar. 25, 2004**

(54) POWER GENERATION WITHIN A MOTIONLESS ELECTROMAGNETIC GENERATOR

(76) Inventors: **Stephen L. Patrick**, Huntsville, AL (US); **Thomas E. Bearden**, Huntsville, AL (US); **James C. Hayes**, Huntsville, AL (US); **Kenneth D. Moore**, Huntsville, AL (US); **James L. Kenny**, Huntsville, AL (US)

Correspondence Address:
Norman Friedland
Suite 400
11300 U. S. Highway One
North Palm Beach, FL 33408 (US)

(21) Appl. No.: **10/081,847**

(22) Filed: **Feb. 26, 2002**

Related U.S. Application Data

(63) Continuation of application No. 09/656,313, filed on Sep. 6, 2000, now Pat. No. 6,362,718.

Publication Classification

(51) **Int. Cl.^{7}** **H02J 1/00**

(52) **U.S. Cl.** **363/15**

(57) ABSTRACT

An electromagnetic generator without moving parts includes a permanent magnet and a magnetic core including first and second magnetic paths. A first input coil and a first output coil extend around portions of the first magnetic path, while a second input coil and a second output coil extend around portions of the second magnetic path. The input coils are alternatively pulsed to provide induced current pulses in the output coils. Driving electrical current through each of the input coils reduces a level of flux from the permanent magnet within the magnet path around which the input coil extends. In an alternative embodiment of an electromagnetic generator, the magnetic core includes annular spaced-apart plates, with posts and permanent magnets extending in an alternating fashion between the plates. An output coil extends around each of these posts. Input coils extending around portions of the plates are pulsed to cause the induction of current within the output coils.

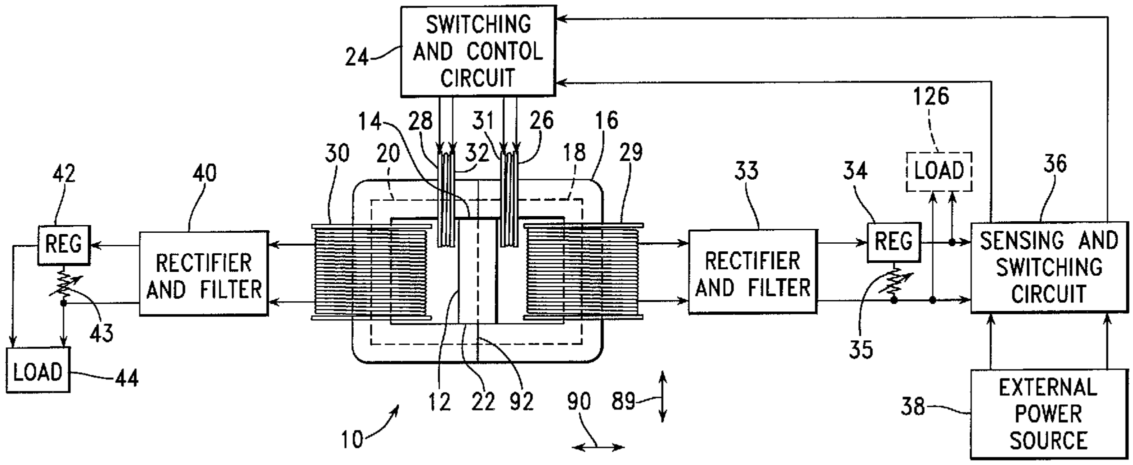
SWITCHING AND CONTOL CIRCUIT
24
14
28
31
26
16
30
20
32
18
29
42
40
33
34
126
LOAD
36
REG
RECTIFIER AND FILTER
RECTIFIER AND FILTER
REG
SENSING AND SWITCHING CIRCUIT
43
LOAD
44
35
12
22
92
90
89
10
38
EXTERNAL POWER SOURCE

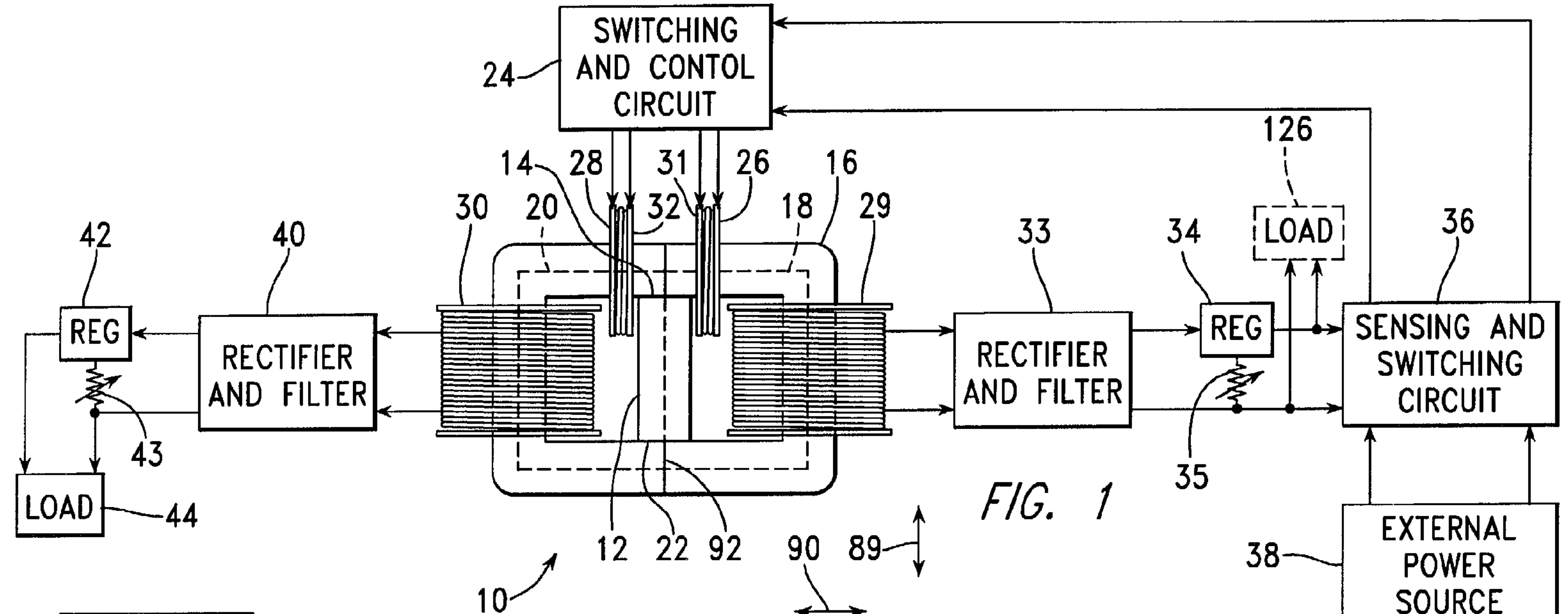
SWITCHING AND CONTOL CIRCUIT
24
14
28
31
26
16
30
20
32
18
29
42
40
33
34
126
LOAD
36
REG
RECTIFIER AND FILTER
RECTIFIER AND FILTER
REG
SENSING AND SWITCHING CIRCUIT
43
35
LOAD
44
12
22
92
90
89
38
EXTERNAL POWER SOURCE
10

FIG. 1

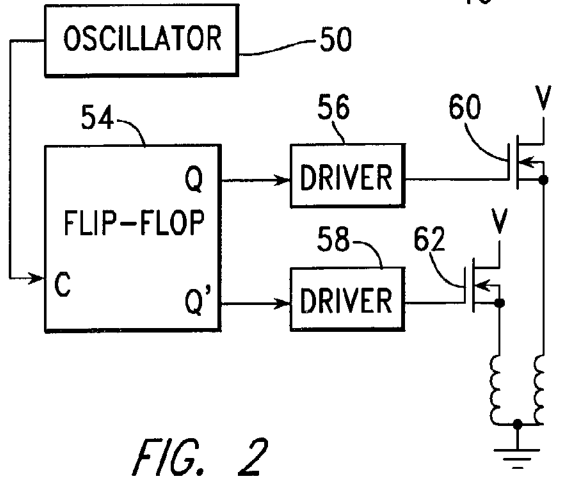
OSCILLATOR
50
54
56
60
V
Q
FLIP-FLOP
C
Q'
DRIVER
58
DRIVER
62
V

FIG. 2

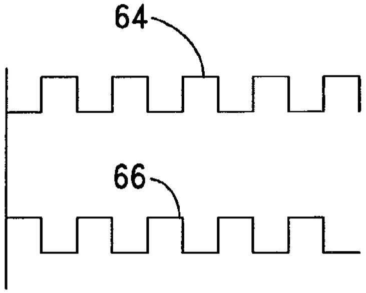

FIG. 3

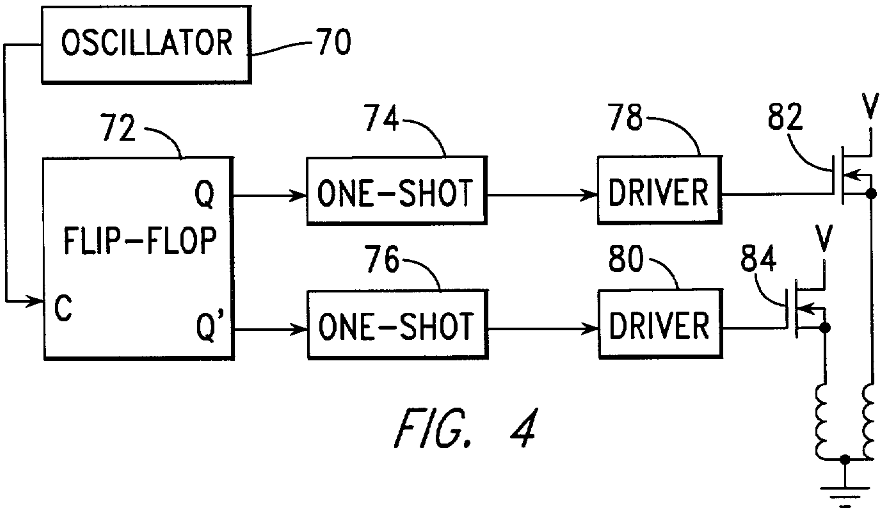
OSCILLATOR
70
72
FLIP-FLOP
Q
C
Q'
74
ONE-SHOT
76
ONE-SHOT
78
DRIVER
80
DRIVER
82
V
84
V

FIG. 4

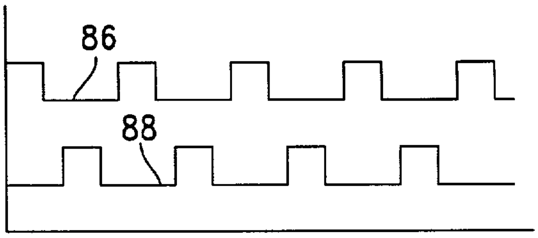

FIG. 5

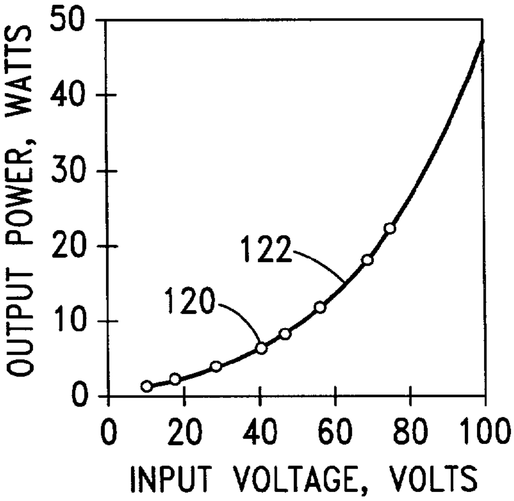

FIG. 7

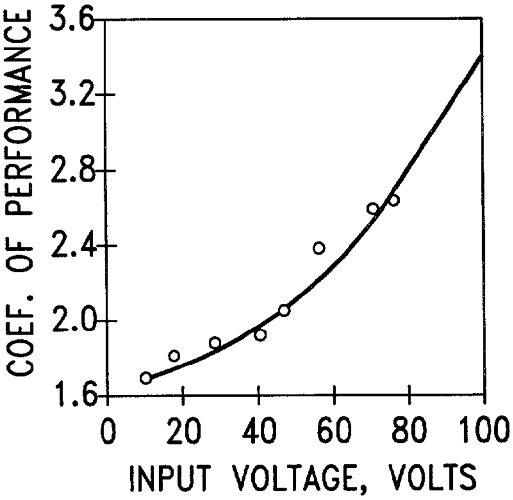

FIG. 8

FIG. 6A

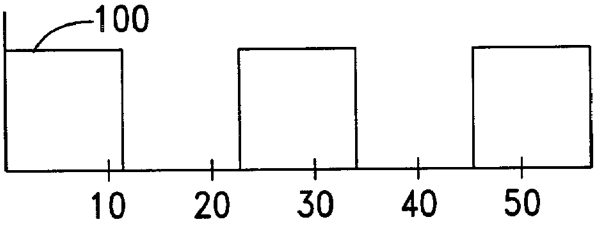

FIG. 6B

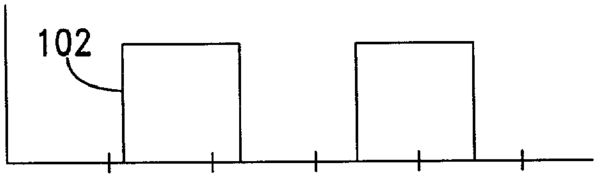

FIG. 6C

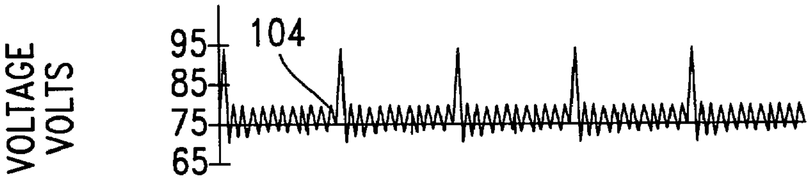

FIG. 6D

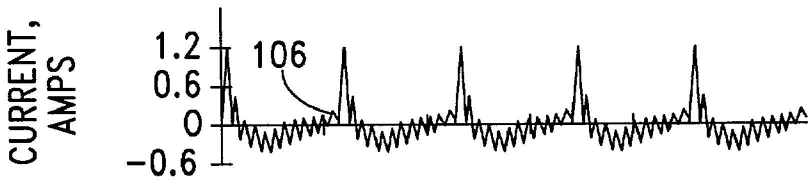

FIG. 6E

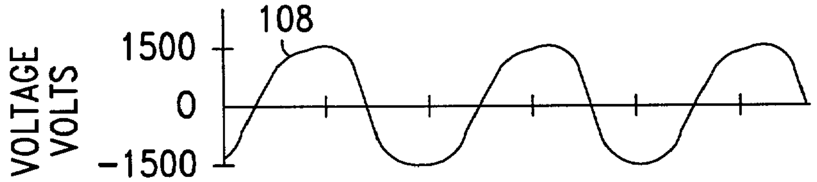

FIG. 6F

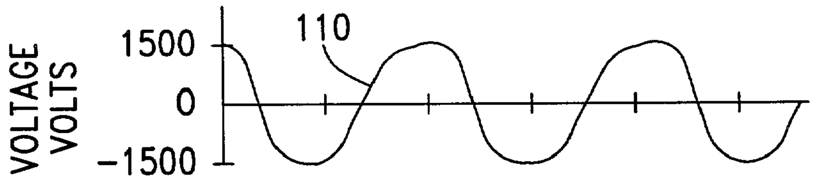

FIG. 6G

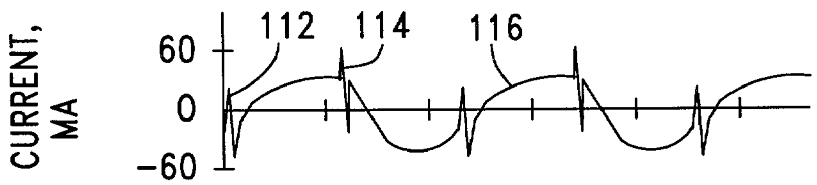

FIG. 6H

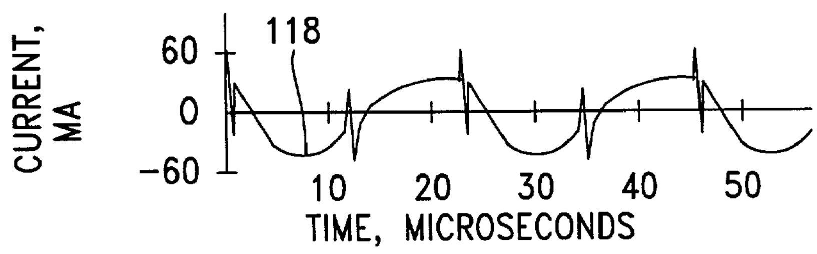

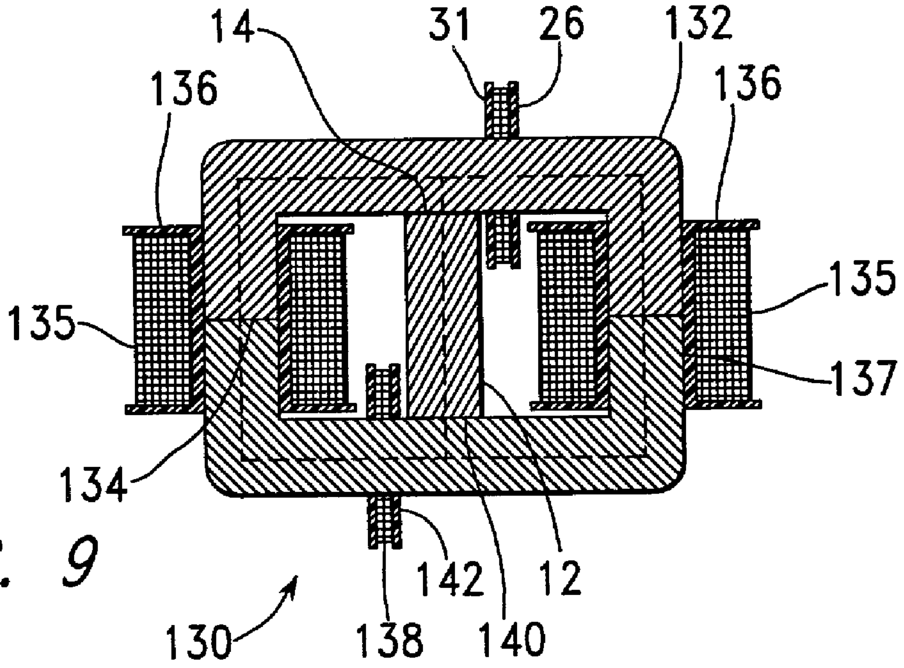

FIG. 9

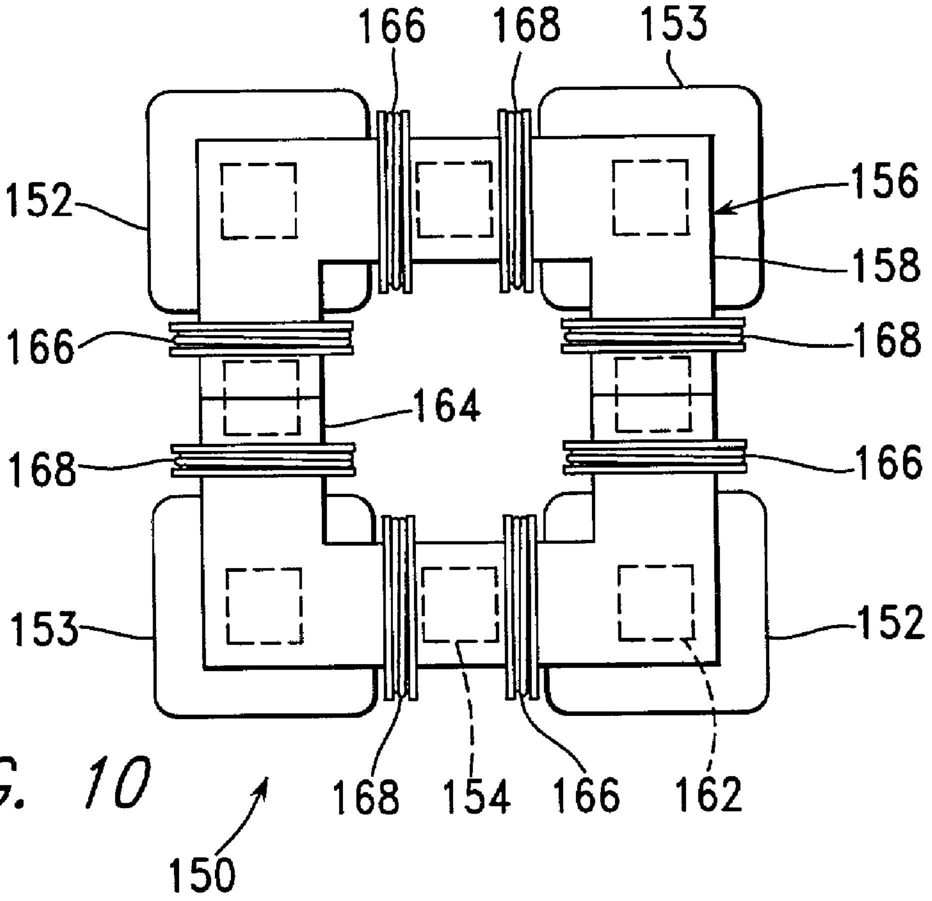

FIG. 10

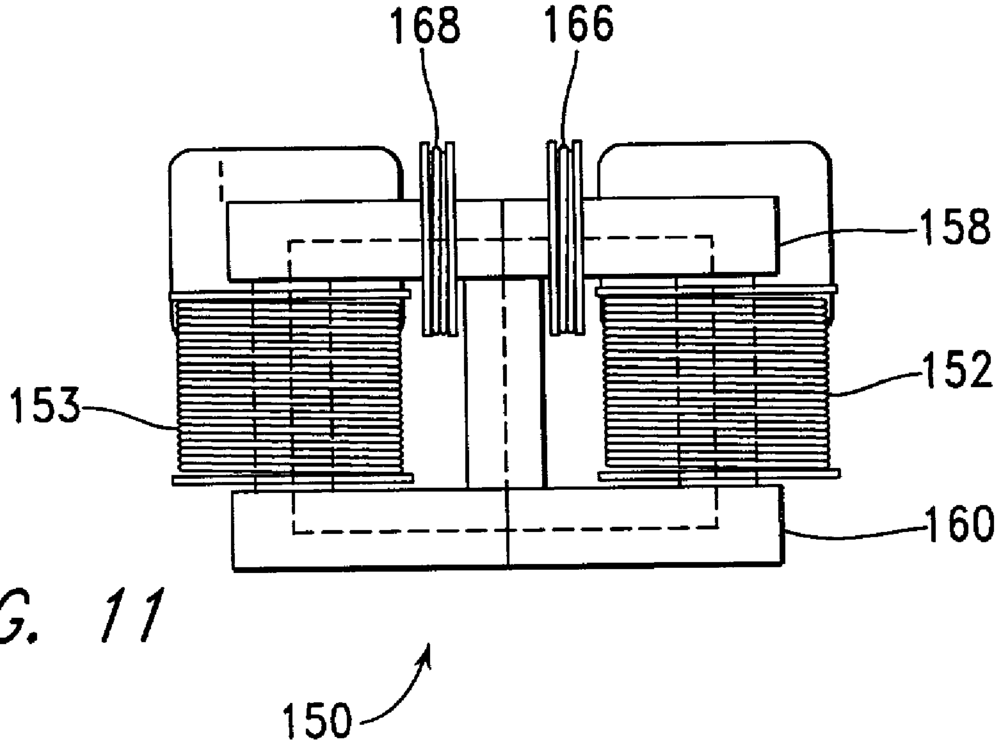
168
166
158
152
153
160
150

FIG. 11

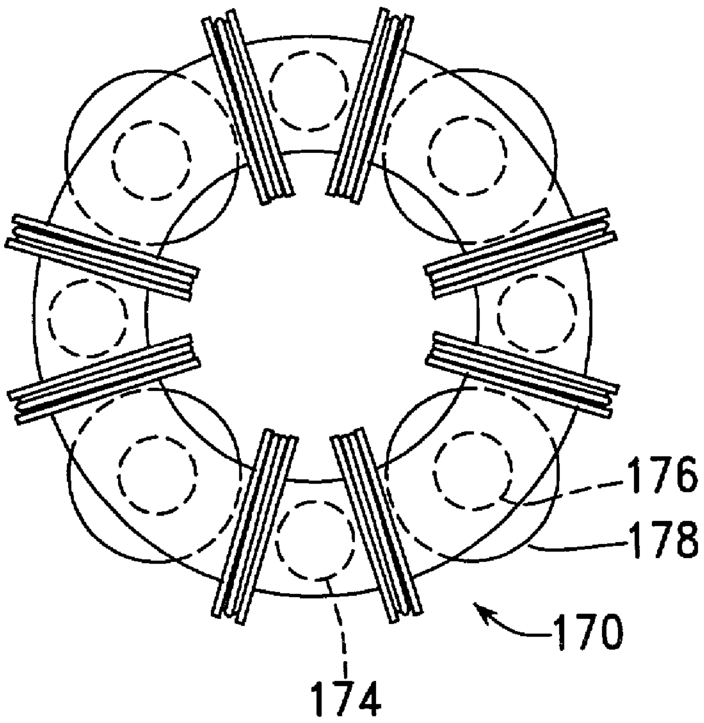
176
178
170
174

FIG. 12

POWER GENERATION WITHIN A MOTIONLESS ELECTROMAGNETIC GENERATOR

[0001] This application is a continuation of a copending U.S. application Ser. No. 09/656,313, filed 09/06/00, titled "Motionless Electromagnetic Generator," for which the issue fee has been paid.

BACKGROUND INFORMATION

[0002] 1. Field of Invention

[0003] This invention relates to a magnetic generator used to produce electrical power without moving parts, and, more particularly, to such a device having a capability, when operating, of producing electrical power without an external application of input power through input coils.

[0004] 2. Description of the Related Art

[0005] The patent literature describes a number of magnetic generators, each of which includes a permanent magnet, two magnetic paths external to the permanent magnet, each of which extends between the opposite poles of the permanent magnet, switching means for causing magnetic flux to flow alternately along each of the two magnetic paths, and one or more output coils in which current is induced to flow by means of changes in the magnetic field within the device. These devices operate in accordance with an extension of Faraday's Law, indicating that an electrical current is induced within a conductor within a changing magnetic field, even if the source of the magnetic field is stationary.

[0006] A method for switching magnetic flux to flow predominantly along either of two magnetic paths between opposite poles of a permanent magnet is described as a "flux transfer" principle by R. J. Radus in *Engineer's Digest,* Jul. 23, 1963. This principle is used to exert a powerful magnetic force at one end of both the north and south poles and a very low force at the other end, without being used in the construction of a magnetic generator. This effect can be caused mechanically, by keeper movement, or electrically, by driving electrical current through one or more control windings extending around elongated versions of the pole pieces **14**. Several devices using this effect are described in U.S. Pat. Nos. 3,165,723, 3,228,013, and 3,316,514, which are incorporated herein by reference.

[0007] Another step toward the development of a magnetic generator is described in U.S. Pat. No. 3,368,141, which is incorporated herein by reference, as a device including a permanent magnet in combination with a transformer having first and second windings about a core, with two paths for magnetic flux leading from each pole of the permanent magnet to either end of the core, so that, when an alternating current induces magnetic flux direction changes in the core, the magnetic flux from the permanent magnet is automatically directed through the path which corresponds with the direction taken by the magnetic flux through the core due to the current. In this way, the magnetic flux is intensified. This device can be used to improve the power factor of a typically inductively loaded alternating current circuit.

[0008] Other patents describe magnetic generators in which electrical current from one or more output coils is described as being made available to drive a load, in the more conventional manner of a generator. For example, U.S. Pat. No. 4,006,401, which is incorporated herein by reference, describes an electromagnetic generator including permanent magnet and a core member, in which the magnetic flux flowing from the magnet in the core member is rapidly alternated by switching to generate an alternating current in a winding on the core member. The device includes a permanent magnet and two separate magnetic flux circuit paths between the north and south poles of the magnet. Each of the circuit paths includes two switching means for alternately opening and closing the circuit paths, generating an alternating current in a winding on the core member. Each of the switching means includes a switching magnetic circuit intersecting the circuit path, with the switching magnetic circuit having a coil through which current is driven to induce magnetic flux to saturate the circuit path extending to the permanent magnet. Power to drive these coils is derived directly from the output of a continuously applied alternating current source. What is needed is an electromagnetic generator not requiring the application of such a current source.

[0009] U.S. Pat. No. 4,077,001, which is incorporated herein by reference, describes a magnetic generator, or dc/dc converter, comprising a permanent magnet having spaced-apart poles and a permanent magnetic field extending between the poles of the magnet. A variable-reluctance core is disposed in the field in fixed relation to the magnet and the reluctance of the core is varied to cause the pattern of lines of force of the magnetic field to shift. An output conductor is disposed in the field in fixed relation to the magnet and is positioned to be cut by the shifting lines of permanent magnetic force so that a voltage is induced in the conductor. The magnetic flux is switched between alternate paths by means of switching coils extending around portions of the core, with the flow of current being alternated between these switching coils by means of a pair of transistors driven by the outputs of a flip-flop. The input to the flip flop is driven by an adjustable frequency oscillator. Power for this drive circuit is supplied through an additional, separate power source. What is needed is a magnetic generator not requiring the application of such a power source.

[0010] U.S. Pat. No. 4,904,926, which is incorporated herein by reference, describes another magnetic generator using the motion of a magnetic field. The device includes an electrical winding defining a magnetically conductive zone having bases at each end, the winding including elements for the removing of an induced current therefrom. The generator further includes two pole magnets, each having a first and a second pole, each first pole in magnetic communication with one base of the magnetically conductive zone. The generator further includes a third pole magnet, the third pole magnet oriented intermediately of the first poles of the two pole electromagnets, the third pole magnet having a magnetic axis substantially transverse to an axis of the magnetically conductive zone, the third magnet having a pole nearest to the conductive zone and in magnetic attractive relationship to the first poles of the two pole electromagnets, in which the first poles thereof are like poles. Also included in the generator are elements, in the form of windings, for cyclically reversing the magnetic polarities of the electromagnets. These reversing means, through a cyclical change in the magnetic polarities of the electromagnets, cause the magnetic flux lines associated with the magnetic attractive relationship between the first poles of the electromagnets and the nearest pole of the third magnet to correspondingly reverse, causing a wiping effect across the magnetically conductive zone, as lines of magnetic flux swing between respective first poles of the two electromagnets, thereby inducing electron movement within the output windings and thus generating a flow of current within the output windings.

[0011] U.S. Pat. No. 5,221,892, which is incorporated herein by reference, describes a magnetic generator in the form of a direct current flux compression transformer including a magnetic envelope having poles defining a magnetic axis and characterized by a pattern of magnetic flux lines in polar symmetry about the axis. The magnetic flux lines are spatially displaced relative to the magnetic envelope using control elements which are mechanically stationary relative to the core. Further provided are inductive elements which are also mechanically stationary relative to the magnetic envelope. Spatial displacement of the flux relative to the inductive elements causes a flow of electrical current. Further provided are magnetic flux valves which provide for the varying of the magnetic reluctance to create a time domain pattern of respectively enhanced and decreased magnetic reluctance across the magnetic valves, and, thereby, across the inductive elements.

[0012] Other patents describe devices using superconductive elements to cause movement of the magnetic flux. These devices operate in accordance with the Meissner effect, which describes the expulsion of magnetic flux from the interior of a superconducting structure as the structure undergoes the transition to a superconducting phase. For example, U.S. Pat. No. 5,011,821, which is incorporated herein by reference, describes an electric power generating device including a bundle of conductors which are placed in a magnetic field generated by north and south pole pieces of a permanent magnet. The magnetic field is shifted back and forth through the bundle of conductors by a pair of thin films of superconductive material. One of the thin films is placed in the superconducting state while the other thin film is in a non-superconducting state. As the states are cyclically reversed between the two films, the magnetic field is deflected back and forth through the bundle of conductors.

[0013] U.S. Pat. No. 5,327,015, which is incorporated herein by reference, describes an apparatus for producing an electrical impulse comprising a tube made of superconducting material, a source of magnetic flux mounted about one end of the tube, a means, such as a coil, for intercepting the flux mounted along the tube, and a means for changing the temperature of the superconductor mounted about the tube. As the tube is progressively made superconducting, the magnetic field is trapped within the tube, creating an electrical impulse in the means for intercepting. A reversal of the superconducting state produces a second pulse.

[0014] None of the patented devices described above use a portion of the electrical power generated within the device to power the reversing means used to change the path of magnetic flux. Thus, like conventional rotary generators, these devices require a steady input of power, which may be in the form of electrical power driving the reversing means of one of these magnetic generators or the torque driving the rotor of a conventional rotary generator. Yet, the essential function of the magnetic portion of an electrical generator is simply to switch magnetic fields in accordance with precise timing. In most conventional applications of magnetic generators, the voltage is switched across coils, creating magnetic fields in the coils which are used to override the fields of permanent magnets, so that a substantial amount of power must be furnished to the generator to power the switching means, reducing the efficiency of the generator.

[0015] Recent advances in magnetic material, which have particularly been described by Robert C. O'Handley in *Modern Magnetic Materials, Principles and Applications,* John Wiley & Sons, New York, pp. 456-468, provide nanocrystalline magnetic alloys, which are particularly well suited for the rapid switching of magnetic flux. These alloys are primarily composed of crystalline grains, or crystallites, each of which has at least one dimension of a few nanometers. Nanocrystalline materials may be made by heat-treating amorphous alloys which form precursors for the nanocrystalline materials, to which insoluble elements, such as copper, are added to promote massive nucleation, and to which stable, refractory alloying materials, such as niobium or tantalum carbide are added to inhibit grain growth. Most of the volume of nanocrystalline alloys is composed of randomly distributed crystallites having dimensions of about 2-40 nm. These crystallites are nucleated and grown from an amorphous phase, with insoluble elements being rejected during the process of crystallite growth. In magnetic terms, each crystallite is a single-domain particle. The remaining volume of nanocrystalline alloys is made up of an amorphous phase in the form of grain boundaries having a thickness of about 1 nm.

[0016] Magnetic materials having particularly useful properties are formed from an amorphous Co—Nb—B (cobalt-niobium-boron) alloy having near-zero magnetostriction and relatively strong magnetization, as well as good mechanical strength and corrosion resistance. A process of annealing this material can be varied to change the size of crystallites formed in the material, with a resulting strong effect on DC coercivity. The precipitation of nanocrystallites also enhances AC performance of the otherwise amorphous alloys.

[0017] Other magnetic materials are formed using iron-rich amorphous and nanocrystalline alloys, which generally show larger magnetization that the alloys based on cobalt. Such materials are, for example, Fe—B—Si—Nb—Cu (iron-boron-silicon-niobium-copper) alloys. While the permeability of iron-rich amorphous alloys is limited by their relatively large levels of magnetostriction, the formation of a nanocrystalline material from such an amorphous alloy dramatically reduces this level of magnetostriction, favoring easy magnetization.

[0018] Advances have also been made in the development of materials for permanent magnets, particularly in the development of materials including rare earth elements. Such materials include samarium cobalt, $SmCo_5$, which is used to form a permanent magnet material having the highest resistance to demagnetization of any known material. Other magnetic materials are made, for example, using combinations of iron, neodymium, and boron.

SUMMARY OF THE INVENTION

[0019] It is a first objective of the present invention to provide a magnetic generator which a need for an external power source during operation of the generator is eliminated.

[**0020**] It is a second objective of the present invention to provide a magnetic generator in which a magnetic flux path is changed without a need to overpower a magnetic field to change its direction.

[**0021**] It is a third objective of the present invention to provide a magnetic generator in which the generation of electricity is accomplished without moving parts.

[**0022**] In the apparatus of the present invention, the path of the magnetic flux from a permanent magnet is switched in a manner not requiring the overpowering of the magnetic fields. Furthermore, a process of self-initiated iterative switching is used to switch the magnetic flux from the permanent magnet between alternate magnetic paths within the apparatus, with the power to operate the iterative switching being provided through a control circuit consisting of components known to use low levels of power. With self-switching, a need for an external power source during operation of the generator is eliminated, with a separate power source, such as a battery, being used only for a very short time during start-up of the generator.

[**0023**] According to a first aspect of the present invention, an electromagnetic generator is provided, including a permanent magnet, a magnetic core, first and second input coils, first and second output coils, and a switching circuit. The permanent magnet has magnetic poles at opposite ends. The magnetic core includes a first magnetic path, around which the first input and output coils extend, and a second magnetic path, around which the second input and output coils extend, between opposite ends of the permanent magnet. The switching circuit drives electrical current alternately through the first and second input coils. The electrical current driven through the first input oil causes the first input coil to produce a magnetic field opposing a concentration of magnetic flux from the permanent magnet within the first magnetic path. The electrical current driven through the second input coil causes the second input coil to produce a magnetic field opposing a concentration of magnetic flux from the permanent magnet within the second magnetic path.

[**0024**] According to another aspect of the present invention, an electromagnetic generator is provided, including a magnetic core, a plurality of permanent magnets, first and second pluralities of input coils, a plurality of output coils, and a switching circuit. The magnetic core includes a pair of spaced-apart plates, each of which has a central aperture, and first and second pluralities of posts extending between the spaced-apart plates. The permanent magnets each extend between the pair of spaced apart plates. Each permanent magnet has magnetic poles at opposite ends, with the magnetic fields of all the permanent magnets being aligned to extend in a common direction. Each input coil extends around a portion of a plate within the spaced-apart plates, between a post and a permanent magnet. An output coil extends around each post. The switching circuit drives electrical current alternately through the first and second pluralities of input coils. Electrical current driven through each input coil in the first plurality of input coils causes an increase in magnetic flux within each post within the first plurality of posts from permanent magnets on each side of the post and a decrease in magnetic flux within each post within the second plurality of posts from permanent magnets on each side of the post. Electrical current driven through each input coil in the second plurality of input coils causes a decrease in magnetic flux within each post within the first plurality of posts from permanent magnets on each side of the post and an increase in magnetic flux within each post within the second plurality of posts from permanent magnets on each side of the post.

BRIEF DESCRIPTION OF THE DRAWINGS

[**0025**] **FIG. 1** is a partly schematic front elevation of a magnetic generator and associated electrical circuits built in accordance with a first version of the first embodiment of the present invention;

[**0026**] **FIG. 2** is a schematic view of a first version of a switching and control circuit within the associated electrical circuits of **FIG. 1**;

[**0027**] **FIG. 3** is a graphical view of drive signals produced within the circuit of **FIG. 2**;

[**0028**] **FIG. 4** is a schematic view of a second version of a switching and control circuit within the associated electrical circuits of **FIG. 1**;

[**0029**] **FIG. 5** is a graphical view of drive signals produced within the circuit of **FIG. 3**;

[**0030**] **FIG. 6A** is a graphical view of a first drive signal within the apparatus of **FIG. 1**;

[**0031**] **FIG. 6B** is a graphical view of a second drive signal within the apparatus of **FIG. 1**;

[**0032**] **FIG. 6C** is a graphical view of an input voltage signal within the apparatus of **FIG. 1**;

[**0033**] **FIG. 6D** is a graphical view of an input current signal within the apparatus of **FIG. 1**;

[**0034**] **FIG. 6E** is a graphical view of a first output voltage signal within the apparatus of **FIG. 1**;

[**0035**] **FIG. 6F** is a graphical view of a second output voltage signal within the apparatus of **FIG. 1**;

[**0036**] **FIG. 6G** is a graphical view of a first output current signal within the apparatus of **FIG. 1**;

[**0037**] **FIG. 6H** is a graphical view of a second output current signal within the apparatus of **FIG. 1**;

[**0038**] **FIG. 7** is a graphical view of output power measured within the apparatus of **FIG. 1**, as a function of input voltage;

[**0039**] **FIG. 8** is a graphical view of a coefficient of performance, calculated from measurements within the apparatus of **FIG. 1**, as a function of input voltage;

[**0040**] **FIG. 9** is a cross-sectional elevation of a second version of the first embodiment of the present invention;

[**0041**] **FIG. 10** is a top view of a magnetic generator built in accordance with a first version of a second embodiment of the present invention;

[**0042**] **FIG. 11** is a front elevation of the magnetic generator of **FIG. 10**; and

[**0043**] **FIG. 12** is a top view of a magnetic generator built in accordance with a second version of the second embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0044] **FIG. 1** is a partly schematic front elevation of an electromagnetic generator **10**, built in accordance with a first embodiment of the present invention to include a permanent magnet **12** to supply input lines of magnetic flux moving from the north pole **14** of the magnet **12** outward into magnetic flux path core material **16**. The flux path core material **16** is configured to form a right magnetic path **18** and a left magnetic path **20**, both of which extend externally between the north pole **14** and the south pole **22** of the magnet **12**. The electromagnetic generator **10** is driven by means of a switching and control circuit **24**, which alternately drives electrical current through a right input coil **26** and a left input coil **28**. These input coils **26**, **28** each extend around a portion of the core material **16**, with the right input coil **26** surrounding a portion of the right magnetic path **18** and with the left input coil **28** surrounding a portion of the left magnetic path **20**. A right output coil **29** also surrounds a portion of the right magnetic path **18**, while a left output coil **30** surrounds a portion of the left magnetic path **20**.

[0045] In accordance with a preferred version of the present invention, the switching and control circuit **24** and the input coils **26**, **28** are arranged so that, when the right input coil **26** is energized, a north magnetic pole is present at its left end **31**, the end closest to the north pole **14** of the permanent magnet **12**, and so that, when the left input coil **28** is energized, a north magnetic pole is present at its right end **32**, which is also the end closest to the north pole **14** of the permanent magnet **12**. Thus, when the right input coil **26** is magnetized, magnetic flux from the permanent magnet **12** is repelled from extending through the right input coil **26**. Similarly, when the left input coil **28** is magnetized, magnetic flux from the permanent magnet **12** is repelled from extending through the left input coil **28**.

[0046] Thus, it is seen that driving electrical current through the right input coil **26** opposes a concentration of flux from the permanent magnet **12** within the right magnetic path **18**, causing at least some of this flux to be transferred to the left magnetic path **20**. On the other hand, driving electrical current through the left input coil **28** opposes a concentration of flux from the permanent magnet **12** within the left magnetic path **20**, causing at least some of this flux to be transferred to the right magnetic path **18**.

[0047] While in the example of **FIG. 1**, the input coils **26**, **28** are placed on either side of the north pole of the permanent magnet **12**, being arranged along a portion of the core **16** extending from the north pole of the permanent magnet **12**, it is understood that the input coils **26**, **28** could as easily be alternately placed on either side of the south pole of the permanent magnet **12**, being arranged along a portion of the core **16** extending from the south pole of the permanent magnet **12**, with the input coils **26**, **28** being wired to form, when energized, magnetic fields having south poles directed toward the south pole of the permanent magnet **12**. In general, the input coils **26**, **28** are arranged along the magnetic core on either side of an end of the permanent magnet forming a first pole, such as a north pole, with the input coils being arranged to produce magnetic fields of the polarity of the first pole directed toward the first pole of the permanent magnet.

[0048] Further in accordance with a preferred version of the present invention, the input coils **26**, **28** are never driven with so much current that the core material **16** becomes saturated. Driving the core material **16** to saturation means that subsequent increases in input current can occur without effecting corresponding changes in magnetic flux, and therefore that input power can be wasted. In this way, the apparatus of the present invention is provided with an advantage in terms of the efficient use of input power over the apparatus of U.S. Pat. No. 4,000,401, in which a portion both ends of each magnetic path is driven to saturation to block flux flow. In the electromagnetic generator **10**, the switching of current flow within the input coils **26**, **28** does not need to be sufficient to stop the flow of flux in one of the magnetic paths **18**, **20** while promoting the flow of magnetic flux in the other magnetic path. The electromagnetic generator **10** works by changing the flux pattern; it does not need to be completely switched from one side to another.

[0049] Experiments have determined that this configuration is superior, in terms of the efficiency of using power within the input coils **26**, **28** to generate electrical power within the output coils **29**, **30**, to the alternative of arranging input coils and the circuits driving them so that flux from the permanent magnet is driven through the input coils as they are energized. This arrangement of the present invention provides a significant advantage over the prior-art methods shown, for example, in U.S. Pat. No. 4,077,001, in which the magnetic flux is driven through the energized coils.

[0050] The configuration of the present invention also has an advantage over the prior-art configurations of U.S. Pat. Nos. 3,368,141 and 4,077,001 in that the magnetic flux is switched between two alternate magnetic paths **18**, **20** with only a single input coil **26**, **28** surrounding each of the alternate magnetic paths. The configurations of U.S. Pat. Nos. 3,368,141 and 4,077,001 each require two input coils on each of the magnetic paths. This advantage of the present invention is significant both in the simplification of hardware and in increasing the efficiency of power conversion.

[0051] The right output coil **29** is electrically connected to a rectifier and filter **33**, having an output driven through a regulator **34**, which provides an output voltage adjustable through the use of a potentiometer **35**. The output of the linear regulator **34** is in turn provided as an input to a sensing and switching circuit **36**. Under start up conditions, the sensing and switching circuit **36** connects the switching and control circuit **24** to an external power source **38**, which is, for example, a starting battery. After the electromagnetic generator **10** is properly started, the sensing and switching circuit **36** senses that the voltage available from regulator **34** has reached a predetermined level, so that the power input to the switching and control circuit **24** is switched from the external power source **38** to the output of regulator **34**. After this switching occurs, the electromagnetic generator **10** continues to operate without an application of external power.

[0052] The left output coil **30** is electrically connected to a rectifier and filter **40**, the output of which is connected to a regulator **42**, the output voltage of which is adjusted by means of a potentiometer **43**. The output of the regulator **42** is in turn connected to an external load **44**.

[0053] **FIG. 2** is a schematic view of a first version of the switching and control circuit **24**. An oscillator **50** drives the clock input of a flip-flop **54**, with the Q and Q' outputs of the flip-flop **54** being connected through driver circuits **56**, **58** to power FETS **60**, **62** so that the input coils **26**, **28** are alternately driven. In accordance with a preferred version of the present invention, the voltage V applied to the coils **26**, **28** through the FETS **60**, **62** is derived from the output of the sensing and switching circuit **36**.

[**0054**] **FIG. 3** is a graphical view of the signals driving the gates of FETS **60**, **62** of **FIG. 2**, with the voltage of the signal driving the gate of FET **60** being represented by line **64**, and with the voltage of the signal driving FET **62** being represented by line **66**. Both of the coils **26**, **28** are driven with positive voltages.

[**0055**] **FIG. 4** is a schematic view of a second version of the switching and control circuit **24**. In this version, an oscillator **70** drives the clock input of a flip-flop **72**, with the Q and Q' outputs of the flip-flop **72** being connected to serve as triggers for one-shots **74**, **76**. The outputs of the one-shots **74**, **76** are in turn connected through driver circuits **78**, **80** to drive FETS **82**, **84**, so that the input coils **26**, **28** are alternately driven with pulses shorter in duration than the Q and Q' outputs of the flip flop **72**.

[**0056**] **FIG. 5** is a graphical view of the signals driving the gates of FETS **82**, **84** of **FIG. 4**, with the voltage of the signal driving the gate of FET **82** being represented by line **86**, and with the voltage of the signal driving the gate of FET **84** being represented by line **88**.

[**0057**] Referring again to **FIG. 1**, power is generated in the right output coil **29** only when the level of magnetic flux is changing in the right magnetic path **18**, and in the left output coil **30** only when the level of magnetic flux is changing in the left magnetic path **20**. It is therefore desirable to determine, for a specific magnetic generator configuration, the width of a pulse providing the most rapid practical change in magnetic flux, and then to provide this pulse width either by varying the frequency of the oscillator **50** of the apparatus of **FIG. 2**, so that this pulse width is provided with the signals shown in **FIG. 3**, or by varying the time constant of the one-shots **74**, **76** of **FIG. 4**, so that this pulse width is provided by the signals of **FIG. 5** at a lower oscillator frequency. In this way, the input coils are not left on longer than necessary. When either of the input coils is left on for a period of time longer than that necessary to produce the change in flux direction, power is being wasted through heating within the input coil without additional generation of power in the corresponding output coil.

[**0058**] A number of experiments have been conducted to determine the adequacy of an electromagnetic generator built as the generator **10** in **FIG. 1** to produce power both to drive the switching and control logic, providing power to the input coils **26**, **28**, and to drive an external load **44**. In the configuration used in this experiment, the input coils **26**, **28** had 40 turns of 18-gauge copper wire, and the output coils **29**, **30** had 450 turns of 18-gauge copper wire. The permanent magnet **12** had a height of 40 mm (1.575 in. between its north and south poles, in the direction of arrow **89**, a width of 25.4 mm (1.00 in.), in the direction of arrow **90**, and in the other direction, a depth of 38.1 mm (1.50 in.). The core **16** had a height, in the direction of arrow **89**, of 90 mm (3.542 in.), a width, in the direction of arrow **90**, of 135 mm (5.315 in.) and a depth of 70 mm (2.756 in.). The core **16** had a central hole with a height, in the direction of arrow **89**, of 40 mm (1.575 mm) to accommodate the magnet **12**, and a width, in the direction of arrow **90**, of 85 mm (3.346 in.). The core **16** was fabricated of two "C"-shaped halves, joined at lines **92**, to accommodate the winding of output coils **29**, **30** and input coils **26**, **28** over the core material.

[**0059**] The core material was a laminated iron-based magnetic alloy sold by Honeywell as METGLAS Magnetic Alloy 2605SA1. The magnet material was a combination of iron, neodymium, and boron.

[**0060**] The input coils **26**, **28** were driven at an oscillator frequency of 87.5 KHz, which was determined to produce optimum efficiency using a switching control circuit configured as shown in **FIG. 2**. This frequency has a period of 11.45 microseconds. The flip flop **54** is arranged, for example, to be set and reset on rising edges of the clock signal input from the oscillator, so that each pulse driving one of the FETS **60**, **62** has a duration of 11.45 microseconds, and so that sequential pulses are also separated to each FET are also separated by 11.45 microseconds.

[**0061**] FIGS. 6A-6H are graphical views of signals which simultaneously occurred within the apparatus of **FIGS. 1 and 2** during operation with an applied input voltage of 75 volts. **FIG. 6A** shows a first drive signal **100** driving FET **60**, which conducts to drive the right input coil **26**. **FIG. 6B** is shows a second drive signal **102** driving FET **62**, which conducts to drive the left input coil **28**.

[**0062**] **FIGS. 6C and 6D** show voltage and current signals associated with current driving both the FETS **60**, **62** from a battery source. **FIG. 6C** shows the level **104** of voltage V. While the nominal voltage of the battery was 75 volts, a decaying transient signal **106** is superimposed on this voltage each time one of the FETS **60**, **62** is switched on to conduct. The specific pattern of this transient signal depends on the internal resistance of the battery, as well as on a number of characteristics of the magnetic generator **10**. Similarly, **FIG. 6D** shows the current **106** flowing into both FETS **60**, **62** from the battery source. Since the signals **104**, **106** show the effects of current flowing into both FETS **60**, **62** the transient spikes are 11.45 microseconds apart.

[**0063**] FIGS. 6E-6H show voltage and current levels measured at the output coils **29**, **30**. **FIG. 6E** shows a voltage output signal **108** of the right output coil **29**, while **FIG. 6F** shows a voltage output signal **110** of the left output coil **30**. For example, the output current signal **116** of the right output coil **29** includes a first transient spike **112** caused when the a current pulse in the left input coil **28** is turned on to direct magnetic flux through the right magnetic path **18**, and a second transient spike **114** caused when the left input coil **28** is turned off with the right input coil **26** being turned on. **FIG. 6G** shows a current output signal **116** of the right output coil **29**, while **FIG. 6H** shows a current output signal **118** of the left output coil **30**.

[**0064**] **FIG. 7** is a graphical view of output power measured using the electromagnetic generator **10** and eight levels of input voltage, varying from 10 v to 75 v. The oscillator frequency was retained at 87.5 KHz. The measurement points are represented by indicia **120**, while the curve **122** is generated by polynomial regression analysis using a least squares fit.

[**0065**] **FIG. 8** is a graphical view of a coefficient of performance, defined as the ratio of the output power to the input power, for each of the measurement points shown in **FIG. 7**. At each measurement point, the output power was substantially higher than the input power. Real power measurements were computed at each data point using measured voltage and current levels, with the results being averaged over the period of the signal. These measurements agree with RMS power measured using a Textronic THS730 digital oscilloscope.

[**0066**] While the electromagnetic generator **10** was capable of operation at much higher voltages and currents without saturation, the input voltage was limited to 75 volts because of voltage limitations of the switching circuits being used. Those skilled in the relevant art will understand that components for switching circuits capable of handling higher voltages in this application are readily available. The experimentally-measured data was extrapolated to describe operation at an input voltage of 100 volts, with the input current being 140 ma, the input power being 14 watts, and with a resulting output power being 48 watts for each of the two output coils **29**, **30**, at an average output current of 12 ma and an average output voltage of 4000 volts. This means that for each of the output coils **29**, **30**, the coefficient of performance would be 3.44.

[**0067**] While an output voltage of 4000 volts may be needed for some applications, the output voltage can also be varied through a simple change in the configuration of the electromagnetic generator **10**. The output voltage is readily reduced by reducing the number of turns in the output windings. If this number of turns is decreased from 450 to 12, the output voltage is dropped to 106.7, with a resulting increase in output current to 0.5 amps for each output coil **29**, **30**. In this way, the output current and voltage of the electromagnetic generator can be varied by varying the number of turns of the output coils **29**, **30**, without making a substantial change in the output power, which is instead determined by the input current, which determines the amount of magnetic flux shuttled during the switching process.

[**0068**] The coefficients of performance, all of which were significantly greater than 1, plotted in **FIG. 8** indicate that the output power levels measured in each of the output coils **29**, **30** were substantially greater than the corresponding input power levels driving both of the input coils **26**, **28**. Therefore, it is apparent that the electromagnetic generator **10** can be built in a self-actuating form, as discussed above in reference to **FIG. 1**. In the example of **FIG. 1**, except for a brief application of power from the external power source **38**, to start the process of power generation, the power required to drive the input coils **26**, **28** is derived entirely from power developed within the right output coil **29**. If the power generated in a single output coil **29**, **30** is more than sufficient to drive the input coils **26**, **28**, an additional load **126** may be added to be driven with power generated in the output coil **29** used to generate power to drive the input coils **26**, **28**. On the other hand, each of the output coils **29**, **30** may be used to drive a portion of the input coil power requirements, for example with one of the output coils **26**, **28** providing the voltage V for the FET **60** (shown in **FIG. 2**), while the other output coil provides this voltage for the FET **62**.

[**0069**] Regarding thermodynamic considerations, it is noted that, when the electromagnetic generator **10** is operating, it is an open system not in thermodynamic equilibrium. The system receives static energy from the magnetic flux of the permanent magnet. Because the electromagnetic generator **10** is self-switched without an additional energy input, the thermodynamic operation of the system is an open dissipative system, receiving, collecting, and dissipating energy from its environment; in this case, from the magnetic flux stored within the permanent magnet. Continued operation of the electromagnetic generator **10** causes demagnetization of the permanent magnet. The use of a magnetic material including rare earth elements, such as a samarium cobalt material or a material including iron, neodymium, and boron is preferable within the present invention, since such a magnetic material has a relatively long life in this application.

[**0070**] Thus, an electromagnetic generator operating in accordance with the present invention should be considered not as a perpetual motion machine, but rather as a system in which flux radiated from a permanent magnet is converted into electricity, which is used both to power the apparatus and to power an external load. This is analogous to a system including a nuclear reactor, in which a number of fuel rods radiate energy which is used to keep the chain reaction going and to heat water for the generation of electricity to drive external loads.

[**0071**] **FIG. 9** is a cross-sectional elevation of an electromagnetic generator **130** built in accordance with a second version of the first embodiment of the present invention. This electromagnetic generator **130** is generally similar in construction and operation to the electromagnetic generator **10** built in accordance with the first version of this embodiment, except that the magnetic core **132** of the electromagnetic generator **10** is built in two halves joined along lines **134**, allowing each of the output coils **135** to be wound on a plastic bobbin **136** before the bobbin **136** is placed over the legs **137** of the core **132**. **FIG. 9** also shows an alternate placement of an input coil **138**. In the example of **FIG. 1**, both input coils **26**, **28** were placed on the upper portion of the magnetic core **16**, with these coils **26**, **28** being configured to establish magnetic fields having north magnetic poles at the inner ends **31**, **32** of the coils **26**, **28**, with these north magnetic poles thus being closest to the end **14** of the permanent magnet **12** having its north magnetic pole. In the example of **FIG. 9, a** first input coil **26** is as described above in reference to **FIG. 1**, but the second input coil **138** is placed adjacent the south pole **140** of the permanent magnet **12**. This input coil **138** is configured to establish a south magnetic pole at its inner end **142**, so that, when input coil **138** is turned on, flux from the permanent magnet **12** is directed away from the left magnetic path **20** into the right magnetic path **18**.

[**0072**] **FIGS. 10 and 11** show an electromagnetic generator **150** built in accordance with a first version of a second embodiment of the present invention, with **FIG. 10** being a top view thereof, and with **FIG. 11** being a front elevation thereof. This electromagnetic generator **150** includes an output coil **152**, **153** at each corner, and a permanent magnet **154** extending along each side between output coils. The magnetic core **156** includes an upper plate **158**, a lower plate **160**, and a square post **162** extending within each output coil **152**, **153**. Both the upper plate **158** and the lower plate **160** include central apertures **164**.

[**0073**] Each of the permanent magnets **154** is oriented with a like pole, such as a north pole, against the upper plate

158. Eight input coils **166**, **168** are placed in positions around the upper plate **158** between an output coil **152**, **153** and a permanent magnet **154**. Each input coil **166**, **168** is arranged to form a magnetic pole at its end nearest to the adjacent permanent magnet **154** of a like polarity to the magnetic poles of the magnets **154** adjacent the upper plate **158**. Thus, the input coils **166** are switched on to divert magnetic flux of the permanent magnets **154** from the adjacent output coils **152**, with this flux being diverted into magnetic paths through the output coils **153**. Then, the input coils **168** are switched on to divert magnetic flux of the permanent magnets **154** from the adjacent output coils **153**, with this flux being diverted into magnetic paths through the output coils **152**. Thus, the input coils form a first group of input coils **166** and a second group of input coils **168**, with these first and second groups of input coils being alternately energized in the manner described above in reference to **FIG. 1** for the single input coils **26**, **28**. The output coils produce current in a first train of pulses occurring simultaneously within coils **152** and in a second train of pulses occurring simultaneously within coils **153**.

[**0074**] Thus, driving current through input coils **166** causes an increase in flux from the permanent magnets **154** within the posts **162** extending through output coils **153** and a decrease in flux from the permanent magnets **154** within the posts **162** extending through output coils **152**. On the other hand, driving current through input coils **168** causes a decrease in flux from the permanent magnets **154** within the posts **162** extending through output coils **153** and an increase in flux from the permanent magnets **154** within the posts **162** extending through output coils **152**.

[**0075**] While the example of **FIGS. 10 and 11** shows all of the input coils **166**, **168** deployed along the upper plate **158**, it is understood that certain of these input coils **166**, **168** could alternately be deployed around the lower plate **160**, in the manner generally shown in **FIG. 9**, with one input coil **166**, **168** being within each magnetic circuit between a permanent magnet **154** and an adjacent post **162** extending within an output coil **152**, **153**, and with each input coil **166**, **168** being arranged to produce a magnetic field having a magnetic pole like the closest pole of the adjacent permanent magnet **154**.

[**0076**] **FIG. 12** is a top view of a second version **170** of the second embodiment of the present invention, which is similar to the first version thereof, which has been discussed in reference to **FIGS. 10 and 11**, except that an upper plate **172** and a similar lower plate (not shown) are annular in shape, while the permanent magnets **174** and posts **176** extending through the output coils **178** are cylindrical. The input coils **180** are oriented and switched as described above in reference to **FIGS. 9 and 10**.

[**0077**] While the example of **FIG. 12** shows four permanent magnets, four output coils and eight input coils it is understood that the principles described above can be applied to electromagnetic generators having different numbers of elements. For example, such a device can be built to have two permanent magnets, two output coils, and four input coils, or to have six permanent magnets, six output coils, and twelve input coils.

[**0078**] In accordance with the present invention, material used for magnetic cores is preferably a nanocrystalline alloy, and alternately an amorphous alloy. The material is preferably in a laminated form. For example, the core material is a cobalt-niobium-boron alloy or an iron based magnetic alloy.

[**0079**] Also in accordance with the present invention, the permanent magnet material preferably includes a rare earth element. For example, the permanent magnet material is a samarium cobalt material or a combination of iron, neodymium, and boron.

[**0080**] While the invention has been described in its preferred versions and embodiments with some degree of particularity, it is understood that this description has been given only by way of example and that numerous changes in the details of construction, fabrication, and use, including the combination and arrangement of parts, may be made without departing from the spirit and scope of the invention.

What is claimed is:

1. A method for generating electrical power, wherein

said method comprises driving electrical current alternately through a first input coil extending around a first portion of a first magnetic path and a second input coil extending around a first portion of a second magnetic path, inducing a flow of electrical current through a first output coil extending around a second portion of said first magnetic path due to changes in magnetic flux within said first magnetic path, and inducing a flow of electrical current through a second output coil extending around a second portion of said first magnetic path due to changes in magnetic flux within said second magnetic path,

said first magnetic path includes a first U-shaped magnetic structure extending in a first direction between opposite poles at opposite ends of a permanent magnet,

said second magnetic path includes a second U-shaped magnetic structure extending in a second direction between said opposite poles at said opposite ends of said permanent magnet,

electrical current driven through said first input coil produces a magnetic field opposing a concentration of magnetic flux from said permanent magnet within said first magnetic path, and

electrical current driven through said second input coil produces a magnetic field opposing a concentration of magnetic flux from said permanent magnet within said second magnetic path.

2. The method of claim 1, wherein

said permanent magnet has a pole of a first type at a first end,

said first input coil is displaced along said first magnetic path adjacent said first end of said permanent magnet,

said second input coil is displaced along said second magnetic path adjacent said first end of said permanent magnet,

driving said electrical current through said first input coil causes a magnetic field to be generated having a pole of said first type at an end of said first input coil adjacent said permanent magnet, and

driving said electrical current through said second input coil causes a magnetic field to be generated having a pole of said first type at an end of said second input coil adjacent said permanent magnet.

3. The method of claim 1, wherein

said permanent magnet has a pole of a first type at a first end and of a second type at a second end, opposite said first end,

said first input coil is displaced along said first magnetic path adjacent said first end of said permanent magnet,

said second input coil is displaced along said second magnetic path adjacent said second end of said permanent magnet,

driving said electrical current through said first input coil causes a magnetic field to be generated having a pole of said first type at an end of said first input coil adjacent said permanent magnet, and

driving said electrical current through said second input coil causes a magnetic field to be generated having a pole of said second type at an end of said second input coil adjacent said permanent magnet.

4. The method of claim 1, additionally comprising:

driving a switching and control circuit by an external power source during a starting process, wherein said switching and control circuit drives said electrical current alternately through said first and second input coils

rectifying a first portion of said flow of electrical current through said first and second output coils to form a first rectified output current; and

driving said switching and control circuit by said first portion of said flow of electrical current following said starting process.

5. The method of claim 4, additionally comprising rectifying a second portion of said flow of electrical current through said first and second output coils to form a second rectified output current flowing through an external load.

6. The method of claim 5, wherein said first portion of said flow of electrical current flows through said first output coil, and said second portion of said flow of electrical current flows through said second output coil.

7. The method of claim 1, wherein said first and second input coils are alternately driven for time periods of approximately 11.5 milliseconds.

8. The method of claim 1, wherein each said U-shaped magnetic structure is composed of a nanocrystalline magnetic alloy.

9. The method of claim 8, wherein said nanocrystalline magnetic alloy is a cobalt-niobium-boron alloy.

10. The method of claim 8, wherein said nanocrystalline magnetic alloy is an iron-based alloy.

11. The method of claim 1, wherein said changes in magnetic flux within said first and second magnetic paths occur without driving said first and second paths to magnetic saturation.

12. A method for generating electrical power, wherein

said method comprises driving electrical current alternately through a first and a second plurality of input coils, and inducing a flow of current within first and second pluralities of output coils by changes in magnetic flux within a magnetic core extending through said input coils and said output coils,

said magnetic core includes an upper plate section extending around an upper aperture, a lower plate section, spaced apart from said upper plate section, extending around a lower aperture, and a plurality of posts extending in a first pattern around said upper and lower apertures and between said upper and lower plates,

a plurality of permanent magnets extend in a second pattern around said upper and lower apertures and between said upper plate section and said lower plate section,

each post within said plurality of posts extends between an adjacent pair of permanent magnets within said plurality of permanent magnets,

each permanent magnet within said plurality of permanent magnets extends between an adjacent pair of posts within said plurality of posts,

all permanent magnets within said plurality of permanent magnets have a pole of a first type at an end adjacent said upper plate and a pole of a second type at an end adjacent said lower plate,

each input coil in said first plurality of input coils extends around a plate section within said magnetic core between a permanent magnet and a post extending through an output coil in said first plurality of output coils adjacent said permanent magnet and spaced apart from said permanent magnet in a first direction along said plate section, being oriented to oppose a concentration of flux extending from said permanent magnet through said input coil when electrical current is driven through said input coil, and

each input coil in said second plurality of input coils extends around a plate section within said magnetic core between a permanent magnet an a post extending through an output coil in said second plurality of output coils adjacent said permanent magnet and spaced apart from said permanent magnet opposite said first direction along said plate section, being oriented to oppose a concentration of flux extending from said permanent magnet through said input coil when electrical current is driven through said input coil.

13. The method of claim 12, wherein each input coil in said first and second pluralities of input coils extends around said upper plate section.

14. The method of claim 12, wherein

each input coil in said first plurality of input coils extends around said upper plate section, and

each input coil in said second plurality of input coils extends around said lower plate section.

* * * * *

US 20060073976A1

(19) **United States**

(12) **Patent Application Publication**

Pohlman

(10) **Pub. No.: US 2006/0073976 A1**

(43) **Pub. Date: Apr. 6, 2006**

(54) **METHOD OF GRAVITY DISTORTION AND TIME DISPLACEMENT**

(76) Inventor: **Marlin B. Pohlman**, Tulsa, OK (US)

Correspondence Address:
STITES & HARBISON PLLC
1199 NORTH FAIRFAX STREET
SUITE 900
ALEXANDRIA, VA 22314 (US)

(21) Appl. No.: **10/954,767**

(22) Filed: **Oct. 1, 2004**

Publication Classification

(51) **Int. Cl.**

F16C	***39/06***	(2006.01)
G01V	***7/00***	(2006.01)
G09B	***23/06***	(2006.01)
F41B	***6/00***	(2006.01)
H01S	***3/09***	(2006.01)

(52) **U.S. Cl.** .... **505/166**; 505/180; 434/300; 73/382 G; 505/164

(57) **ABSTRACT**

A method for employing sinusoidal oscillations of electrical bombardment on the surface of one Kerr type singularity in close proximity to a second Kerr type singularity in such a method to take advantage of the Lense-Thirring effect, to simulate the effect of two point masses on nearly radial orbits in a 2+1 dimensional anti-de Sitter space resulting in creation of circular timelike geodesics conforming to the van Stockum under the Van Den Broeck modification of the Alcubierre geometry (Van Den Broeck 1999) permitting topology change from one spacelike boundary to the other in accordance with Geroch's theorem (Geroch 1967) which results in a method for the formation of Godel-type geodesically complete spacetime envelopes complete with closed timelike curves.

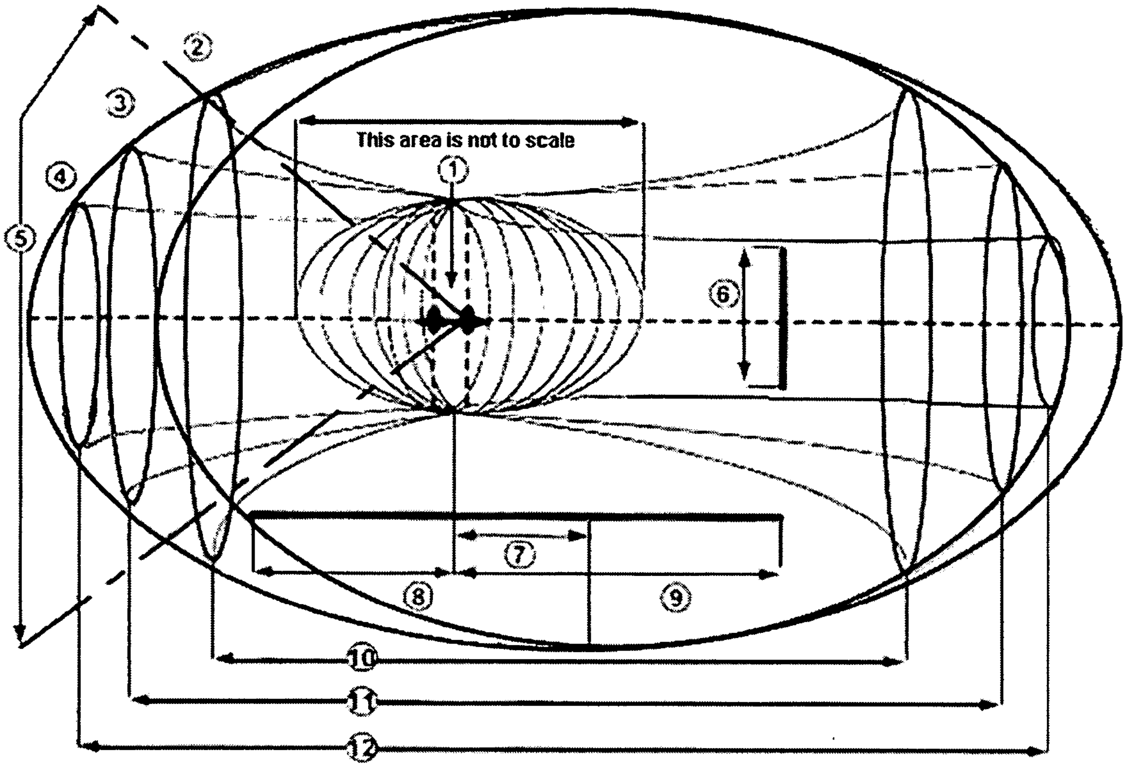

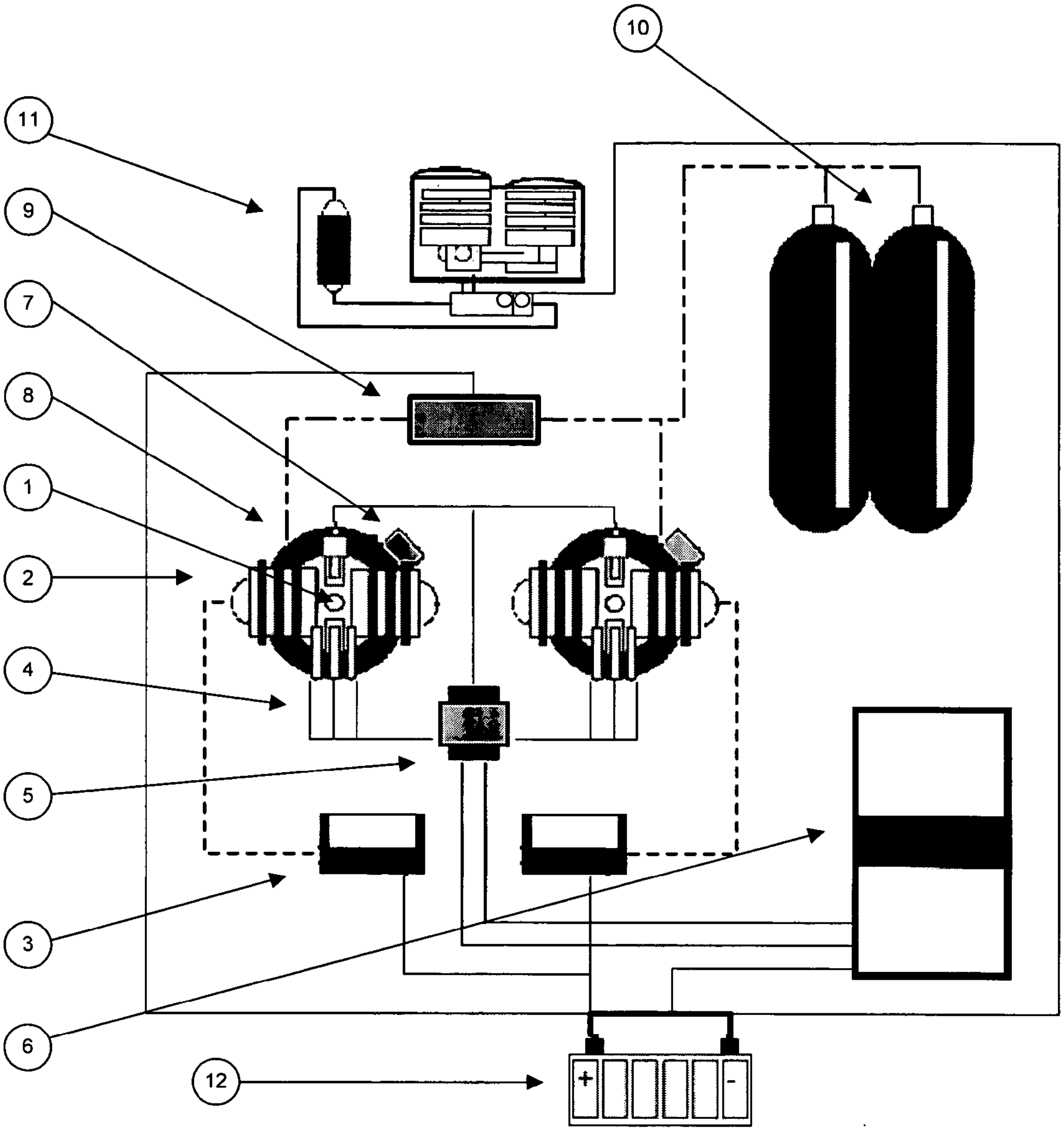
10
11
9
7
8
1
2
4
5
3
6
12
+
-

Figure 1

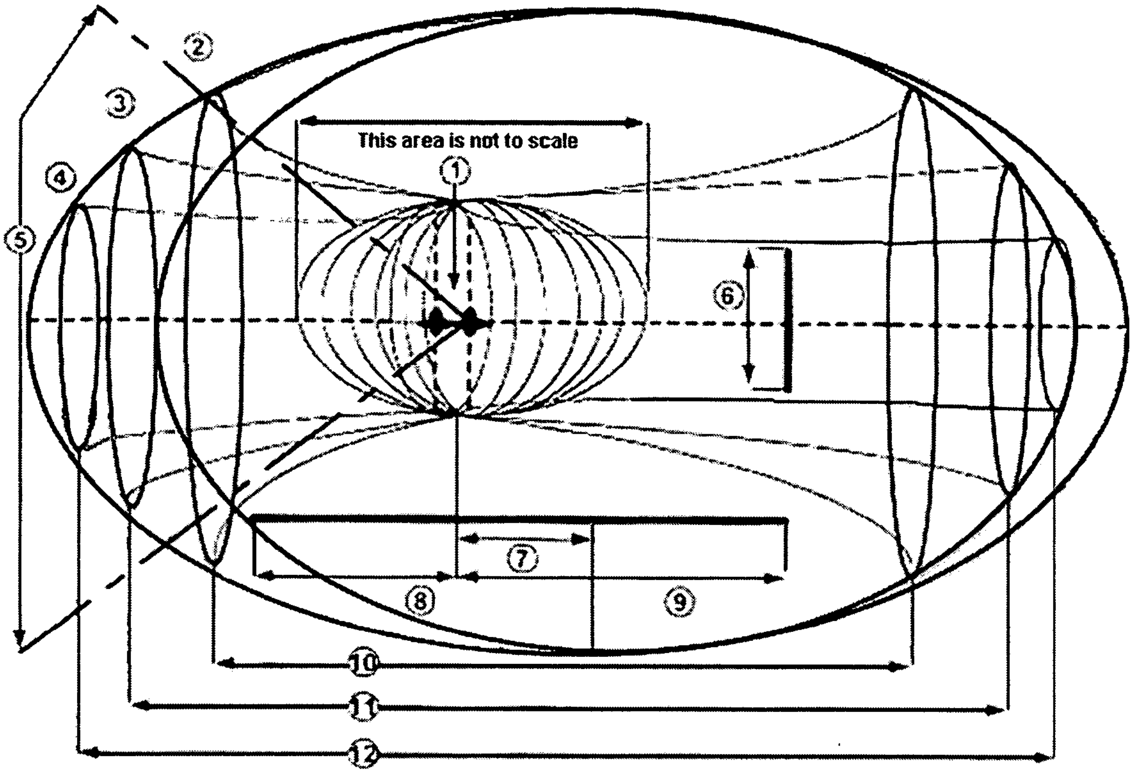
This area is not to scale
①
②
③
④
⑤
⑥
⑦
⑧
⑨
⑩
⑪
⑫

Figure 2

METHOD OF GRAVITY DISTORTION AND TIME DISPLACEMENT

FIELD OF THE INVENTION

[0001] The present invention relates to the use of technical time displacement devices, which operate by the modification of gravitational fields. These drive systems do not depend on the emission of matter to create thrust to take advantage of time dilation, but rather create a change in the curvature of space-time, in accordance with general relativity. This allows travel across topologies by warping space-time, to produce a topology change from one spacelike boundary to the other in accordance with Geroch's theorem (Geroch 1967)

THEORETICAL BACKGROUND OF THE INVENTION

[0002] The concept of gravity should be examined in the light of quantum gravity and in turn as a component of quantum physics itself. The fundamental minimal quantum of energy in quantum physics is Planck's constant; h. Thus in accordance with the energy equivalence formula $E=mc^2$, the fundamental minimum quantity of mass (m_q) can therefore be derived, from known constants by; $m_q=h/c^2$ (1). Taking this minimal mass, it is possible to show that the formation of all matter, the forces of nature and indeed space-time itself derive from this single quintessential quantity.

[0003] Thus if the number of quintessences in a system is; $n_q=m/m_q$: then the total Energy of the system is more logically given by, the energy of a single quintessence (h); directly multiplied by the number of quintessences (n_q) in that system, thus

$$E=hn_q=mc^2 \qquad (1\text{ a}).$$

[0004] Furthermore, this minimal mass, termed quintessence, can form the basis of the existence of a quantum gravitational field in the form of a space-time lattice, from which quantum gravity may be derived from first principles. Furthermore, the conglomeration of these quintessences also accounts for the formation of the elementary particles and the forces acting between them, as in superstring theory. This concept explains the formation of matter and the forces of nature on a quantum mechanical basis and directly explains the existence of wave particle duality. Thus as $n_q=m/m_q$; the frequency of light and matter (f) is determined, directly, from the number of constituent quintessences. This leads automatically to the fundamental equation, derived from (1), $f=n_q=E/h$, where n_q is the number of quintessences, which leads directly to the frequency of both light and matter. This in turn leads directly to a Universal wave equation for matter and light $\lambda=c/\beta n_q=hc/\beta E$ (2), where β is the relative directional velocity, v/c. As the momentum, $p=\beta\times E/c$, then this equation also gives the standard de-Broglie wave equation, $\lambda=h/p$ in agreement with current theory and experiments[1].

[0005] Using the Universal wave equation, the standard equation for special relativity, $m'=m_o/(1-\beta^2)^{1/2}$, derives from first principles. Also from these observations, a modified Dirac wave equation may be derived, $E\psi=(-j\beta\cdot\nabla+\beta m)\psi$ (2a), the results of which have been recently verified by a paper in which the orbitals of electrons were experimentally directly visualised[2]. Moreover, a fundamental equation for general relativity can be formulated, where G is the gravitational constant and r_q is the given radius of quintessence; $G=9(r_q)^2c^4/\lambda\beta E$ (3), such that the Universal wave equation is in direct agreement with general relativity[3]. Thus special and general relativity and quantum mechanics can be unified.

[0006] From here it is possible to proceed in a number of ways; the geometric structure of the electron and the forces of Nature may be derived from first principles and in turn the structure of the quarks, including the top and bottom, otherwise known as truth and beauty can be seen. Moreover, the presence of a space-time lattice results in an understanding of quantum EPR effects. By allowing a theoretical flow of energy through the space-time lattice it can be shown that:

[0007] Energy is not bound by space-time

[0008] Thus logically accounting for phenomena such as entanglement and quantum tunnelling. Quintessence can also be used to explain, logically, the inner physics of a black hole, the missing mass of the Galaxy, the continuing expansion of the Universe, Guth's inflationary theory and the Big Bang. Hence, it is now possible to understand the Universe, including space-time, matter and the forces of nature from the radius, mass and vibration of a single quantity, quintessence.

[0009] With this understanding of space-time, matter and the forces of Nature, and in particular gravity, it is possible to demonstrate that the modification of gravitational fields, and in turn the warping of space-time, can be technically readily achieved.

[0010] Using standard equations for special relativity, $m'=m_0/(1-\beta^2)^{1/2}$, it can be demonstrated that by differentially increasing the velocity of electrons, by applying a differential current, their mass can be increased in a specific way. In turn by increasing the mass of electrons, by general relativity, the number of gravitons emitted from these electrons can be modulated. By multiplying this effect using an ultracentrifugational device the differential graviton emission can be manifestly amplified. This in turn, in accordance with general relativity, will cause a change in the curvature of space-time.

[0011] This effective warping of space-time does not, of necessity, imply superluminal velocities, but does allow the creation of warp drive systems, which do not depend on the creation of thrust by the ejection of material as used in current space technologies.

Part 1—Fundamental Laws of Physics

[0012] Quintessential Mass

[0013] The quantum physical, minimum component of energy is Planck's constant; h. To define the minimal component of mass, using the standard energy equivalence formula; $E=mc^2$, such a minimal mass (m_q) would be required to have the value equivalent to; $m_q=h/c^2$ (1). The total mass of a system (m) would then be; $m=n_qm_q$, where (n_q) is the number of these minimal units. Thence, the total energy of a system can be derived from the minimal energy; h, multiplied by the number of these energy units (n_q). Thus as, $E=mc^2$, then also $E=m_qn_qc^2$ and substituting $m_q=h/c^2$, the energy equivalence formula has the more logical formulation; $E=hn_q$(1a). Thus the energy of a system is equivalent to the minimal energy unit; h, multiplied by the number of those minimal energy units (n_q).

[0014] Taking this minimal mass/energy, it is possible to show that all matter, the forces of nature and space time can be constructed from this single quintessential quantity. Moreover, using this quantity the laws of physics can be derived from first principles. Thus, a priori, all components of the physical universe, including space-time, can be constructed from this minimal mass component, termed quintessence.

[0015] Wave Particle Duality

[0016] If the presence of quintessence accounts for the structure of matter and if matter itself forms from the number of quintessences, then the frequency of matter and thus wave particle duality directly arises from first principles. Specifically the wavelength of matter derives from the vibration of quintessence from which it is constituted. Thus the frequency (f) and in turn the wavelength of light and matter is directly equivalent to the number of quintessences contained within it. We find that the actual frequency of light can be directly derived from first principles from the effective mass of the photon (m,) and thus by the number of quintessences (n_q) it contains.

[0017] Thus for light conventionally:

$$f=E/h$$

[0018] and if $E=mc^2$, and $h=m_qc^2$, then

$$f=m_\gamma c^2/m_q c^2$$

and

$$f=m_\gamma/m_q=n_q$$

Thus

$$f=n_q \tag{4}$$

[0019] Thus the formula for the frequency of light E=hf is now readily explained by the observation that the frequency is determined quite directly from the number of quintessences n_q within the photon.

[0020] The wavelength is thus also given by:

$$\lambda=c/f=m_{qc}/m_\gamma=h/p$$

[0021] We can now show that the frequency of matter also has the same derivation from quintessence, as has the frequency of light. The frequency of matter is again equivalent to the number of quintessences it contains. Thus the wave particle duality of matter itself can be explained by its composition from quintessence. The amount of quintessences contained within a electron sphere will depend on the number of quintessences constituting the electron and those passing through it as a result of its relative velocity β^2 (where $\beta=v/c$); effectively its relativistic momentum (p). The frequency will then be related to the total number of quintessences. Thus for matter,

$$f=\beta^2 n_q \tag{4a}$$

[0022] Thus it is possible to derive the conventional de Broglie wave equation for matter from first principles. Thus, as $\lambda=v/f$, we have:

$$\lambda=v/\beta^2 n_q \tag{5}$$

[0023] thus as $n_q=E/h$

$$\lambda=hc/\beta E \tag{2}$$

[0024] and as conventionally $\beta E/c=p$, then for matter:

$$\lambda=h/p$$

[0025] Provided that in the de Broglie equation, the momentum of the object is calculated using the relativistic mass, thus accounting for the total number of quintessences n_q in an object, this gives an accurate value for the wavelength of matter[1].

[0026] Thus the wavelength of matter follows directly from its constituents, quintessence. As matter is made of quintessence, similarly to light, its frequency depends on the number of quintessences n_q within it, traveling relative to the speed of light. Moreover, $\lambda=hc/\beta E$, underpins a fundamental relationship between wavelength and energy. Furthermore, this is mathematically the same as the term $\lambda=hv/\beta^2 E$, giving a relativistic expression for the wavelength of matter, from which the relativistic equations may be directly derived

[0027] Wave Equations

[0028] The derivation of wave particle duality from first principles also now allows the derivation of a modified wave equation for matter.

[0029] To derive his wave equation Shrodinger commenced with the de Broglie equation using momentum (p). For lower energies the momentum of an electron is conventionally derived from the kinetic energy of the electron and the mass of the electron m_0. Thus conventionally:

$$E_k = \frac{1}{2}mv^{1/2} \quad \text{and} \quad p = m_0 v$$

Thus

$$E_k = p^2/2m_0$$

then

$$p = \sqrt{(E_k \cdot 2m_0)}$$

[0030] and conventionally, the de Broglie equation can also be written as:

$$\lambda=h/p=h/\sqrt{(E_k \cdot 2m_0)}$$

[0031] In turn the Shrodinger wave equation directly derives from the square of the above classical non relativistic term for kinetic energy:

$$\lambda^2 = h^2/E_k \cdot 2m_0$$

$$\text{thus } E_k = \frac{h^2}{2m} \cdot \frac{1}{\lambda^2}$$

$$\text{As } E = E_k + V$$

$$\text{then } E\psi = -\frac{h^2}{2m} \cdot \frac{d^2\psi}{dx^2} + V\psi = jh \cdot \frac{d\psi}{dt}$$

[0032] However, the Shrodinger equation, may be refined by taking into account relativity. Thus the true values for the energy are given by the relativistic momentum (p).

[0033] A fundamental relativistic wave equation for ψ, and its logical derivation may now be developed through the concept of quintessence as a fundamental constituent of matter.

[0034] The amount of quintessences in the electron is determined by the number of quintessences forming the electron at rest, plus the amount of quintessences passing through it due to its relativistic velocity, which will determine the relativistic momentum (p) of a particle.

[0035] The frequency of matter can now be readily calculated from first principles to give a more accurate result. Thus as matter is made of quintessence, similarly to light, its frequency is equal to the number of quintessences n_q within it. The wavelength will depend on its velocity travelling relative to the speed of light and thus multiplied by the relative velocity compared to c(β=v/c);

[0036] Hence for matter as previously shown:

$$\beta=v/\beta^2 n_q=hc/\beta E \tag{2}$$

[0037] And conventionally

$$E=(p^2c^2+m_0^2c^4)^{1/2}$$

[0038] Using these equations, we can now, also, reformulate the Shrodinger wave equation, which has the advantage that relativity can be treated in a quantum mechanical way. Thus if the wave energy of matter is defined as:

$$E_\lambda = \frac{\beta\sqrt{p^2c^2+m_0^2c^4}}{c^2}$$

thus

$$E_\lambda = \sqrt{(\beta^2p^2/c^2)+\beta^2m_0^2}$$

[0039] which in complex space generalises to

$$E_\psi=(-j\beta\cdot\nabla+\beta m)_\psi \tag{2a}$$

[0040] As the term

$$\alpha = \frac{e^2}{hc}4\pi\varepsilon_0;$$

also represents the ground state ratio of the velocity of the electron to c. Thus α=β=v/c=1/137.

[0041] Thus, also

$$E_\psi=(-j\alpha\cdot\nabla+\beta m)_\psi$$

[0042] This is thus the standard relativistic equation that Dirac was able to construct from the Shrodinger wave equation. This relativistic equation can be derived from the modified wave equation. This takes into account the relative mass energy which the quintessential wave equation contains.

[0043] Where importantly the term β·m is the mass m, multiplied by the ratio of the relative velocity to light β=v/c, and the term a is also essentially the relative velocity of the electron.

[0044] The Dirac equation was an empirical formula which worked mathematically, nevertheless even Dirac admitted it was not logically understood. The importance of these equations is that they show that the existence of quintessence allows the waveparticle duality of matter to be explained and mathematically derived from first principles, Thus the frequency of matter or even light is simply determined by the number of quintessences it contains.

[0045] Indeed, a recent publication in Nature has suggested that the direct visualisation of the orbitals of electrons shows that these are in very close agreement with theory. However, there is a significant departure from theory, in the interstitial molecular regions, suggesting that the higher velocities of the electrons obey the modified Dirac equation. Thus these orbitals were in keeping with the modified Dirac equation, which itself may be derived from the wave equation above, λ=hc/βE

[0046] The Shrodinger wave equation will approximate to the correct values until v approaches c. Indeed the Shrodinger equation will give similar answers as that derived from equation, under most experimental conditions.

[0047] However, equation 2 and its derivative may have advantages over standard Shrodinger theory with relativistic speeds. Furthermore, equation 2, conceptually shows that the wave particle duality of matter derives from the principle that the frequency of matter is directly equal to the number of quintessences it contains. Importantly it also mathematically allows relativity and quantum mechanics to be united.

[0048] With v≅c, the modified Dirac equation will yield more accurate results, particularly compared with the Schrodinger equation. We also find that the equation λ=hc/βE is equivalent to the de Broglie wave equation, λ=h/p, provided we use the relativistic mass in the de Broglie equation. Given this, these equations yield accurate experimental results.

[0049] Thus we find that the modified formulation of de Broglie wave equation λ=hc/βE leads directly to a modified Dirac relativistic wave equation and is supported by recent experiments which measure the wavelength of matter and demonstrate the electron orbitals experimentally from these wave equations for matter.

[0050] Wave Particle Duality and Relativity

[0051] From here it is possible to proceed in several ways using the relativistic wave equation. It is apparent that the reintroduction of the term for relative velocity into the wave equations will enable the reintroduction of special relativity into quantum mechanics. In particular we should now be able to derive the term

$$\sqrt{\left(1-\frac{v^2}{c^2}\right)}$$

as a special case of quantum mechanics.

[0052] Thus if:

$$\lambda = hc/\beta E$$

As $E=\sqrt{p^2c^2+m_0^2c^4}$, squaring

$$\lambda^2 = \frac{h^2c^2}{\beta^2\cdot(p^2c^2+m_0^2c^4)}$$

-continued

Conventionally $p^2c^2 = \frac{E^2v^2}{c^2}$

then

$$\lambda^2 = \frac{h^2c^2}{\beta^2 \cdot ((E^2v^2/c^2) + m_0^2c^4)}$$

Thus as $\beta^2 = \frac{v^2}{c^2}$ and $m_0^2c^4 = E_0^2$, then:

$$\beta^4E^2 + \beta^2E_0^2 = \frac{h^2c^2}{\lambda^2}$$

hence $\beta^4E^2 = h^2c^2 \cdot \frac{1}{\lambda^2} - \beta^2m_0^2c^4$ thus

$$\beta^2 = \frac{h^2c^2}{\beta^2E^2} \cdot \frac{1}{\lambda^2} - \frac{\beta^2m_0^2c^4}{\beta^2E^2}$$

As $E^2 = m^2c^4$

$$\beta^2 = \frac{h^2c^2}{\beta^2m^2c^4} \cdot \frac{1}{\lambda^2} - \frac{\beta^2m_0^2c^4}{\beta^2E^2}$$

Substituting $h = m_qc^2$

$$\beta^2 = \frac{m_q^2c^6}{\beta^2m^2c^4} \cdot \frac{1}{\lambda^2} - \frac{\beta^2m_0^2c^4}{\beta^2E^2}$$

As $m_q/'m = 1/n_q$ (eq. 2)

$$\beta^2 = \frac{c^2}{\beta^2n_q^2} \cdot \frac{1}{\lambda^2} - \frac{\beta^2m_0^2c^4}{\beta^2E^4}$$

Thus if $f = \beta^2n_q$; (eq. 7a)

$$\beta^2 = \frac{v^2}{f^2} \cdot \frac{1}{\lambda^2} - \frac{\beta^2m_0^2c^4f^2}{\beta^2E^2}$$

As $1/\lambda^2 = f^2/v^2$

$$\beta^2 = \frac{v^2}{f^2} \cdot \frac{f^2}{v^2} - \frac{\beta^2m_0^2c^4f^2}{\beta^2E^2}$$

Thus

$$\beta^2 = 1 - \frac{\beta^2m_0^2c^4}{\beta^2E^2}$$

As $E^2 = m^2c^4$

$$\beta^2 = 1 - \frac{m_0^2}{'m^2}$$

Hence

$$m_0/'m = (1 - \beta^2)^{1/2}$$

Thus

$$'m = m_0 \Big/ \left[1 - \frac{v^2}{c^2}\right]^{1/2}$$

[0053] Thus this derivation now allows relativity as a universal case of the quintessential wave nature of matter.

[0054] The original premises on which special relativity was based were: that the speed of light is a constant and that all observers are equal. As the speed of light has dimensions of length and time but not apparently of mass, the relativistic change in mass is not accounted for. Using quintessence logically and directly accounts for the relativistic mass changes.

[0055] Moreover, relativity can be derived from the de Broglie equation, and visa versa, directly, thus linking relativity and quantum mechanics by taking into account the existence of quintessence mass.

[0056] Hence, it is now possible to derive the relativistic equations for mass and in turn for space and time from the quintessential wave equation, thus deriving special relativity as a universal case of quantum mechanics and thus uniting special relativity and quantum mechanics. This now allows a further understanding of the nature of space-time.

[0057] The Space-Time Lattice

[0058] The understanding of the true nature of space-time and how it is formulated in three dimensions of real space is crucial. To simply assume that space-time exists, and thence not to question the nature of that existence, denies a deeper understanding of the universe.

[0059] In order to understand the nature of space-time itself, at the quantum level a further look at the nature light and the photon is necessary. Since Einstein's description of light as a particle (the photon) and the description of the photoelectric effect, the standard picture of light as simply a wave can, no longer be applied. If light was to exist as a photon, it could not exist in one dimension, as ordinary waves do, it would need to be three dimensional, with the addition of time. Let us suppose, in this case, that a photon is a three dimensional helical ringlet of light, travelling in the x vector, and spinning around the x-axis. Conventionally this ringlet has a radius; $r=\lambda/2\pi$. The ringlet itself would be vibrating in the y and z vectors. The vectors x, y and z would represent the photon, the substance of which, would be travelling in the x direction and oscillating in the y and z vectors, which would represent oscillatory energy. This in turn would allow it to act as a wave, and create oscillatory electromagnetic fields.

[0060] It is important to re-examine space-time itself in this light, this would have one directional vector with two vector dimensions of energy, one of capacitance and one of electrical permeability, thus accounting for the well known constants of free space; the permittivity of free space (ϵ_0) and the permeability of free space (μ_0) respectively. The vector dimension of direction x, would be the direction of travel and those "quintessences" travelling in an outwardly direction would account for none other than the expansion of the universe. Three of these quintessences would naturally constitute three dimensional visible space-time. These constituents of space-time would interact with the generations of the other vector dimensions reciprocally. Thus one quintessence would sweep out one vector of permeability and one vector of permittivity, through which the other two quintessences could travel, and vica versa, creating a three dimensional space-time lattice.

[0061] The permittivity of free space, (ϵ_0) which is equivalent to capacitance, would as with capacitance plates, be determined by the effective separation between quintessences. The permeability of free space (μ_0) is in fact a force, measured as $4\pi\times10^{-7}$ N/A^2, would result from the force produced by the vibration of quintessence and would be dependent on the density of quintessence. Hence these two parameters would be reciprocal and thus the product of these two would therefore be a constant, which is recognised as none other than the speed of light.

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}$$

[0062] This space time lattice would in effect be created by quintessences travelling in all directions with a speed of c within the lattice. The quintessences of the space time lattice would in effect produce a non-static ether. A non-static ether is fully compatible with special and general relativity. Indeed such an ether explains how space time can be curved as in general relativity. Furthermore, the existence of a non-static ether, was espoused by Einstein in his University of Leyden lecture on general relativity of May 5, 1920. In Einstein's own words;

[0063] "According to the general theory of relativity space without ether is unthinkable."

[0064] Recent evidence from a number of sources now strongly support the presence of this non-static ether, in the form of quintessence. An editorial from a major journal states "combined with other observations such as those of distant Supernova, the QMAP results corroborate the prevailing theory of inflation with the twist that the Universe is only one third matter (both ordinary and dark) and two thirds quintessence, a form of energy possibly inherent in empty space".

[0065] If we take into account the existence of quintessence and as such a three dimensional space-time lattice, matter which is intrinsically made of constituents of charge would interact with this lattice to produce the effects of mass. Mass would be perceived as a result of matter (whose constituent particles appear to contain charge) interacting with this lattice directly due to the inhibition of motion by the lattice's electrical permeability and permittivity vectors, which would form the existence of complex space. These quintessences would in the direction in the y and z vectors produce small vibrations of the order of the Planck length (10^{-35} m), whilst passing through the vectors of permeability and permittivity, thus producing the effects of mass.

[0066] The vibration would endow quintessence itself a (non rest) mass m_q equivalent, to the minimal mass of:

$$m_q = h/c^2 = 7\times373\times10^{-51}\ \text{kg}\cdot\text{sec} \tag{1}$$

[0067] The presence and magnitude of Planck's constant (h) and especially the speed of light (c) is thus explained. Indeed, the speed of light

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}$$

is not in itself a fundamental quantity.

[0068] As the energy equivalence formula is $E=mc^2$, the minimal mass of a single quintessence, would thus be the minimal mass, h/c^2, hence again:

$$m_q = h(\mu_0\epsilon_0) = h/c^2 = 7.373\times10^{-51}\ \text{kg}\cdot\text{sec} \tag{1b}$$

or

$$m_q c^2 = h \tag{1c}$$

[0069] It is postulated by general relativity that the shape of space time itself can be altered, indeed the presence of the space time lattice now allows this to be altered by altering the density of quintessence. It is further clear that if quintessences underly the structure of the space-time lattice, they may also underly the structure of matter itself.

[0070] With regards a single quintessence, this passing through an energy vector of the space-time lattice would appear as a vibrating string. In a similar way to string theory, the conglomeration of these quintessences would produce the constituents of ordinary matter. Thus the general equation for the number of quintessences (n_q) in an object of mass (m) would be

$$m/m_q = n_q$$

[0071] The mass of the electron (m_e) for example, would be directly determined by the number of quintessences in the electron, multiplied by the mass of quintessence.

[0072] Quintessence and Complex Space

[0073] Quintessence is postulated to constitute the fundamental nature of space-time. Three quintessences each travelling in their respective x vectors at 90° to each other would create three dimensional real space-time. These quintessences would in the direction in their respective y and z vectors produce small vibrations of the order of the Planck length (10^{-35} m), this would create the vector dimensions of permeability and permittivity. The result would give space-time 9 dimensions of space as in superstring theory. However, unlike superstring theory the six hidden dimensions would not be "curled up so as to be so small as to be invisible" these six dimensions would be present in complex space. Thus, only three of these dimensions would represent ordinary three dimensional particulate space time i.e. three dimensional objects. The other six dimensions produced by the vibrations of quintessence would form complex space.

[0074] The mathematics of complex space, using imaginary $\sqrt{-1}$ or (j) numbers, is assumed in the standard formulation of the Shrodinger wave equation. Thus the presence of complex space is an integral part of quantum mechanics.

$$-\frac{h^2}{2m}\cdot\frac{d^2\psi}{dx^2} + V\psi = jh\cdot\frac{d\psi}{dt}$$

[0075] The mathematics of complex space is also an essential and integral part of the principles and application of modern electronic and control engineering. Indeed it has been well recognised for some time that each direction vector in electronic engineering can, be associated with complex vectors.

[0076] As this complex space consists of the vectors of permittivity and permeability it would only be "felt" by charged particles as in the electron. Nevertheless, as all particles are fundamentally composed of charged particles the effects of complex space would be felt by endowing these particles with mass and in turn kinetic energy.

[0077] In conventional complex space, a 2 dimensional Cartesian Argand diagram is mathematically used. However, in order to formulate the equations for particles a three dimensional Argand diagram is essential. This will have three dimensional vectors, one real vector and two imaginary vectors. Three of these diagrams will be required to fully describe the nature of particles, each with a real vector in the x, y and z vectors, respectively. Nevertheless, in the instance below the real vector is the x vector and the two imaginary vectors are given by $({}_jy,{}_jz)$

[0078] The Three Dimensional Argand Diagram

[0079] The beauty of a three dimensional Argand diagram is that the complex conjugate (i.e. the mirror image which confers mathematical reality on the coordinates) is formed by the value of the minus coordinate in the other complex vector dimension. Thus the complex conjugate of $(C_x^{1/2}+{}_jc_y^{1/2}+{}_{-j}c_z^{1/2})$ is $(c_x^{1/2}+{}_jc_y^{1/2}+{}_jc_z^{1/2})$. These two sums when multiplied thus give a real number solution.

[0080] Furthermore it is clear that nine dimensions of space time are necessary in the general relativistic equations. By including complex space we thereby create the nine dimensional spacial metrictensor and the metric energy tensor of matter necessary for computations for general relativity From here we can begin to understand the true structure of matter.

[0081] Energy and the Space-Time Lattice

[0082] The presence of numerous experimental data for quantum tunnelling. and indeed the recent observations by Nicholas Gisin, on the entanglement of distant photons now returns us to EPR experiments.

[0083] Using the quintessential modification of the de Broglie wave equation, gives us an insight into these teleportation and EPR effects.

[0084] As

$$\lambda=hc/\beta E \tag{2}$$

and

$$E=hn_q \tag{1a}$$

then

$$\lambda=c/\beta n_q \tag{2b}$$

[0085] Importantly, as indicated by equation (2b), energy having no quintessence; would have a wavelength of infinity. Specifically pure energy containing no quintessences, would have a lambda of infinity. According to quantum mechanics an infinite wavelength would result in the probability of that energy being anywhere. As energy itself has no electrical charge it would not be impeded by the permittivity and permeability of the three dimensional space-time lattice. Moreover, energy would not be detectable in three dimensional space-time, unless it interacted with matter, as in the EPR experiments. Indeed, energy is not observed when not bound to any form of mass or particle. Thus equation 9d, takes us to our original assertion

[0086] Energy is Not Bound by the Space-Time Lattice

[0087] Thus, as the EPR experiments suggest the existence of energy separate from matter and thus separate from the three dimensional space-time lattice, it is interesting to find that experiment suggests the existence of free energy in a continuum separate from space time to produce the effects of quantum teleportation

[0088] This is not, however, teleportation across an additional dimension, this is a term to describe in partially familiar terms the dissociation of energy from the three dimensional space-time lattice. As time is inextricably linked to each dimension of space, the effects of energy would be inextricably linked to the events, such as the creation of virtual particles, we see interacting within space-time.

[0089] It is unlikely that observers have any direct day to day experience to explain quantum events. Nevertheless, quantum physics may have given us a window into the hitherto hidden workings of the Universe. Thereby, the mystery of the uniformity of the Universe, across distances which the speed of light could not apparently traverse, is readily explained by the fact that the free energy contained in the Universe is not bound by the space-time lattice.

[0090] In the case of light, due to the exceedingly small masses involved, there would be relatively easy exchange of matter with free energy within a photon. This would make the photon the ideal experimental tool to look for energy which is not bound by matter and in turn energy which is not bound in space-time. Indeed, very recently Furusawa et at. have reported to have observed the transference of energy as photons from A to B, without those photons traversing space-time. This finding which has been supported using other experimental techniques, is very important as it suggests the existence of such a quantum continuum.

[0091] We have already seen strong experimental data using photons, atomic spins and other data for quantum teleportation which have recently been published which support these findings. According to the above equations the teleportation would vary in a predictable fashion, as with photons, in line with the wavelength of the light used, relative to the size of vibration of quintessence. As regards matter, the results do confirm that the effect of quantum tunnelling is indeed dependant on the wavelength of matter and the size of that matter.

Part II—Particle Physics

[0092] Electron Structure

[0093] Understanding the electron is fundamental to the understanding of the elementary particles. The hidden nature of the electron may recently have been revealed through observations by Horst Stormer, Daniel Tsui and Robert Laughlin for which a Nobel prize has recently been awarded. They describe a quasi electron particle of charge ⅓e. This has been described on a quantum basis as a vortex of energy, bound as a quasi particle in one dimension x, but not bound in the other two dimensions y and z, allowing dispersion in space-time as a vortex. What is more intriguing are the experimental conditions in which this occurs. First of all a two dimensional electron gas is created and held between two capacitance plates. A magnetic force is then applied in the remaining dimension, virtually creating a one dimensional passage through which only a quasi electron appears to be able to pass.

[0094] Given the presence of charge of ⅓e, then three of these quasi electrons could form an entire electron in three dimensional visible space time. Nevertheless, each would have energy and hence a wave function which would be present in the other vectors. This electron could thus follow the probability functions as described by the Shrodinger wave equation for ψ (otherwise termed as "essence" by Shrodinger)

[0095] If the mass of the electron (m_e) is constituted from quintessence, using the formula:

$$m_e/m_q = n_q$$

[0096] Then an electron would be constituted from:

$$\frac{9.11 \times 10^{-3}\ \text{kg}}{7.373 \times 10^{-51}\ \text{kg} \cdot \text{sec}} = 1.235 \times 10^{20}\ \text{quintessences/sec.}$$

[0097] Thus taking into account the mass-energy content of quintessence (m_q) it is independently possible to derive the magnitude of the charge of an electron (e) using the following equation.

$$e = \sqrt{\frac{m_q \varepsilon_0}{\left(\frac{4}{3}\right)\pi hc}} = 1.61 \times 10^{-19} C$$

[0098] This is in close agreement with the experimentally observed charge on the electron of 1.602×10^{-19} C.

[0099] Interestingly substituting $m_q = h/3c^2$ in the above equation we have:

$$e = \sqrt{\frac{\varepsilon_0}{3\left(\frac{4}{3}\pi c^3\right)}} \tag{6}$$

[0100] This can also be written as

$$e = \sqrt{\frac{\varepsilon_0}{3\left(\frac{4}{3}\right)\pi c^3}} \tag{6a}$$

[0101] Equation (6) has a number of very special implications, if re-examined, firstly three of these quasi electron spheres appear to be required to constitute the charge of the electron. More intriguingly, it indicates that the charge is related to the volume of a sphere with an apparent radius of c. Thirdly it indicates that the square of the charge of an electron (e) is proportional to the permittivity of free space (so). The charge given from equation (3) is in close agreement with the measured charge of the electron. Furthermore a more exact value for the charge of the electron (to seven decimal places) can be deduced by taking into account the gravitational field of the Earth (see Gravity and the Charge of the Electron). Furthermore the charge of the electron (e) can now be derived from first principles. Thus, equation (3) corroborates the evidence that the electron is indeed composed of three quasi electrons in keeping with recent experimental findings.

[0102] The significance of the electron, composed of three spheres each with a radius of c, is not immediately clear, but can be understood if the frequency of rotation of the electron is taken into account. Thus if the diameter of the electron was approximately 10^{-19} m, then its spin would need to be

$$\frac{1}{c} \times 10^{-15}$$

m approx. eq. 10^6 cycles/sec. Thus given a very high rotation rate an electron could have an effective radius of 1/c and still occupy subatomic sizes. Indeed these observations might be used to estimate the rate of rotation of the quasi electron and its size (see Appendix 1).

[0103] With regards a single quintessence, this passing through an energy vector of the space-time lattice would appear as a vibrating string. In a similar way to string theory, the conglomeration of these quintessences would produce the constituents of ordinary matter. The electron, for example, would be constituted from approximately 1.235×10^{20} quintessences.

[0104] The dimensions of the equation for the electron can be readily resolved by considering each of the three vector dimensions. The exact dimensions of the equation need to be considered in the light of the nature of space-time itself. These dimensional equations help explain the nature of matter. Indeed the equation for the electron may be necessary for the full understanding of gravity

[0105] Complex Space and Electron Structure

[0106] The presence of complex space also now further explains the conformation of the electron, and its formulation at the quantum level, and the presence of particles, anti-particles and their spin up and spin down characteristics.

[0107] Indeed the short form equation for the charge of the electron (−e) can now be rewritten as a metric tensor with three dimensions in real space and six in complex space.

[0108] Thus if three of the x, y and z vectors are in real space and six vectors in complex space, where c is the speed of light in the real space vector, $_{j}c$ is the speed of light in the complex vector and $_{-j}c$ is the complex conjugate of $_{j}c$, thus the electron can be mathematically represented by the equation:

$$-e = \varepsilon_{qe} \Big/ \left(\frac{4}{3}\pi\right)^{1/2} \cdot \begin{matrix} (c_x)^{1/2} \cdot (jc_y)^{1/2} \cdot (-jc_z)^{1/2} \\ + \quad + \quad + \\ (-jc_x)^{1/2} \cdot (c_y)^{1/2} \cdot (jc_z)^{1/2} \\ + \quad + \quad + \\ (jc_x)^{1/2} \cdot (jc_y)^{1/2} \cdot (c_z)^{1/2} \end{matrix}$$

[0109] Which now elegantly gives the real number solution

$$e = \frac{\varepsilon_{qe}}{3\sqrt{\left(\frac{4}{3}\right)\pi c^3}}$$

[0110] Where e=ϵ_{qe} is given as the permittivity of free space for a single quasi electron Equation 4 represents a "complex" tensor

[0111] Whilst the two dimensional Argand diagram has four quadrants, the three dimensional Argand diagram has eight cubic sectors. Two of these cubic sectors are diametric opposites and can represent "real" particulate objects. These have the primary coordinates x, y, −z; as in the electron described above, and the −x, −y, z, with the real vector x now having a minus sign. These two "real" cubic sectors, therefore, mathematically represent particles and their anti-particles.

[0112] The mathematical presence of the two primary diagonal mirror images (x, y, −z and −x, −y, z) now allow the introduction of the concept of antiparticles. This extension of the maths into a three dimensional Argand diagram thus results in the automatic formulation of the maths of antiparticles. Thus the charge of the positron ($^{+}$e) is formulated by the shortened form equation, where the real vectors now each have the minus sign, and therefore exist in the −x, −y, z sector of the three dimensional Argand diagram.

$$e = \varepsilon_{qe} \Big/ \left(\frac{4}{3}\pi\right)^{1/2} \cdot \begin{matrix} -c_x^{1/2} \cdot_{-j} c_y^{1/2} \cdot_{j} c_z^{1/2} \\ {}_{j}c_x^{1/2} \cdot -c^{1/2y} \cdot_{-j} c_z^{1/2} \\ {}_{-j}c_x^{1/2} \cdot jc_y^{1/2} \cdot -c_z^{1/2} \end{matrix}$$

[0113] The three dimensional Argand diagram also accounts for chirality and indeed the up and down spin of the electron. There are two other "real" primary coordinates in the Argand diagram, these are themselves the partial mirror images of the above coordinates (i.e. x, −y, z and −x, y, −z). In particular the y axis is of the opposite sign, thus in particles the y axis is in the downward direction, to form down spin particles and in anti-particles in the up direction, to form the antiparticle The up spin electron is given by eq. 8 and hence the down spin electron (−e|↓≦) is given by the equation

$$-el\downarrow\rangle = \varepsilon_{qe} \Big/ \left(\frac{4}{3}\pi\right)^{1/2} \cdot \begin{matrix} c_x^{1/2} \cdot_{-j} c_y^{1/2} \cdot_{j} c_z^{1/2} \\ {}_{j}c_x^{1/2} \cdot c^{1/2y} \cdot_{-j} c_z^{1/2} \\ {}_{-j}c_x^{1/2} \cdot jc_y^{1/2} \cdot -c_z^{1/2} \end{matrix} \tag{9}$$

[0114] Thus the three dimensional Argand diagram accounts directly for the presence of antiparticles and the spin up and spin down particles seen in nature. It also accounts for the necessity of the electron to form a square root spherical object, as complex space depends on √−1, otherwise known as j.

[0115] Electron Pairing and Superconductivity

[0116] As the quintessences making up the electron are in a square root conformation, each of these quasi electrons would have a tendancy to pair to form an entire sphere.

[0117] The square root sphere structure of electrons with up and down spins can now superimpose to produce a complete sphere of varying extents. This produces electron pairing as seen at the atomic and molecular levels. It also accounts for the Pauli exclusion principle. This pairing thus accounts for the reactivity of the valence electrons and the electron probability densities, which in turn accounts for the existence of chemistry.

[0118] Furthermore, it is possible to account directly for superconductivity from first principles. For if both the complex and real vectors of the electron combine completely, the product of an up and down spin electron form a perfect superimposed sphere with radius c, with a charge of 2.59×10^{-38} C, denoted by the formula:

$$e^2 = \frac{\varepsilon o}{3\left(\frac{4}{3}\pi c^3\right)} = 2.59\times10^{-38}C \tag{6b}$$

[0119] As with standard superconducting theory, superconductivity can be explained by the formation of "Cooper" electron pairs, where the electrons are forced to pair by the presence of positive crystal charge in particular formation, at supercooled temperatures. In addition the electron pair now forms a stable entity whose angular momentum cancels.

[0120] It additionally becomes clear that the charge of two separate electrons (2e) is 3.2×10^{-19} C, but the charge of the combined electrons (e^2) is 2.59×10^{-38} C. This electron pair thus appears to have 19 orders of magnitude less charge than the electron and in turn 19 orders of magnitude less resistance. It is this effective reduction in charge and in turn resistance, which may account for superconductivity. When observed directly any electrical interaction with the Cooper electron pair will, however, result in the release of the full charge of both electrons, so that the full electrical charge put in will be equal to that coming out of the apparatus.

[0121] The Fine Structure Constant

[0122] Intruiginty from our knowledge of the electron we can further define the term .alpha., the fine structure constant; from the structure of the electron. Thus as the standard term

$$\alpha = \frac{e^2}{hc\times 47\pi\varepsilon_0};$$

substituting the term

$$e^2 = \frac{\varepsilon_0}{3(4/3)\pi c^3} \tag{eq. 6}$$

and

$$h = m_q c^2 \tag{eq. 1}$$

we find:

$$\frac{2\pi}{\alpha} = m_q[3(4/3\pi c^3)]^2$$

*or

$$\frac{2\pi}{\alpha} = \frac{m_q e^4}{\varepsilon_0^2}$$

[0123] For brevity we may represent the quasi electron structure as $(4/3)\pi c^3=\theta$; to signify its threefold symmetry, thus

$$\frac{2\pi}{\alpha} = m_q(3\theta)^2 \tag{10}$$

[0124] Indicating that the fine structure constant of the electron (a) is indeed related to its dimensional structure. Again taking into account the effects of gravity the fine structure constant can be derived from first principles to nine decimal places (see Gravity and the Charge of the Electron).

[0125] Fundamental Forces and Particle Structure

[0126] In order to understand the fundamental forces and the nature of fundamental particles, an overview is required. Thus, there are three major forces; strong, electro-weak and gravity, each mediated by three force particles the gluon, photon and graviton respectively. These in turn, influence three types of particle, the quark, lepton, and by general relativity space-time itself. Each of these are composed of particles with multiples of charge of ⅓, which are themselves in three generations, and are present in three dimensions of real space. It is important that a comprehensive view of nature explains this threefold symmetry.

[0127] Using the Standard Model of particles, it is well accepted there exist quark particle charges of −⅓, −⅔ and +⅓ and +⅔ in quarks and anti-quarks. Given that each particle is made up of three quarks the presence of these fractional charges support the association of the fractional charges in this way to form three dimensional charged particles. In stable particles each of the three quarks would have a vector in one dimension, giving the three quarks together an existence in three dimensional visible space time. The particles that bind the quarks (gluons) are themselves required, in stable particles, to have three different color charges, one color in each dimension, for the particle to exist in three dimensional space-time. Furthermore, there are three generations of quarks (and indeed leptons).

[0128] The Standard Model (or a modification of this) and in particular the observation of quarks and indeed quasi electrons with fractional charge of ⅓ and ⅔ in both cases, indicates that particles are constituted from the equivalent of three of these quasi particles to form an electron and quarks to form baryons. In the normal three dimensions the energy would be carried by the particle, However, because each particle is constituted of three quasi particles and in each quasi particle or quark one visible dimension would be the direction vector, in the other two hidden dimensions of each vector the waves would carry energy. Thus each particle would be associated with vibration, which would account for wave particle duality and Heisenberg's uncertainty principle in three dimensional visible space-time.

[0129] These observations lead us directly to the previous postulate that the structure of the electron is composed of none other than three (root) spheres, and that this equation for the electron allows the determination of the charge of the electron from first principles, thus:

$$e = \frac{\varepsilon 0^2}{[3(4/3\pi c^3)]} \tag{6}$$

[0130] In addition the mass of the proton (m_p) can be directly calculated from the ratio of the mass (m_e) of the electron, given by the equation:

$$\frac{m_e}{m_p} = 5.45 \times 10^{-4} = 3\frac{\pi}{\sqrt{c}} \tag{11}$$

[0131] Strictly we should write,

$$\frac{m_e}{m_p + m_e} = 3\pi\sqrt{c};$$

which is much more elegant.

[0132] Which now gives

$$\frac{m_e}{m_p} = \frac{1}{\sqrt{c}/3\pi - 1} = 5.4462 \times 10^{-3}$$

[0133] This is in very close agreement with the experimentally derived ratio of the proton to electron masses which is also; 5.4462×10^{-3}

[0134] Thus the correlation factor between theory and experiment has a maximum error <0.00001.

[0135] If we combine equation 3:

$$e = \sqrt{\frac{\varepsilon_0}{3(4/3)\pi c^3}}$$

and equation 13:

$$\frac{m_e}{m_p} = 3\frac{\pi}{\sqrt{c}}$$

the positive charge of the proton (e_p) is given by:

$$e_p = \sqrt{\frac{\varepsilon_0}{3(4/3)\pi c^3}} \times m_e \times 3\frac{\pi/\sqrt{c}}{m_p} = e \tag{12}$$

[0136] The stable nuclear proton conformation can thus be represented by the short form equation:

$$p=3^{+}\sqrt{3(4/3)\pi c^3}\times 3(\pi/\sqrt{c}) \tag{13}$$

[0137] This forms a stable 3×3 conformation as with the stable electron structure.

[0138] Importantly the term ($\pi/\sqrt{c}$) is the 90° solution to the Shrodinger wave equation for an electron confined in a space with radius c!.

[0139] Thus the standard equation for an electron confined in a one dimensional box is given by:

$$E\psi(x) = -\frac{h^2}{2m}\frac{d^2\psi(x)}{dx^2}$$

[0140] If the one dimensional box has a length 2L the quantum amplitude (A) can only be non zero between x=0 and X=2L and the standard solution for the amplitude is none other than:

$$A=(1/L^{1/2})$$

[0141] Thus in one dimension the standard solution to the Shrodinger wave equation is:

$$\psi(x)=(\pi/L^{1/2})\mathrm{Sin}\ x/L$$

[0142] Thus not only is the electron charge derived from the equation for three spheres each with a radius of c (eq. 3); but the proton mass and charge can also be derived from the standard solution to the Shrodinger wave equation for a an electron confined in a space of radius c!.

[0143] The term ($\pi/c^{1/2}$) itself would thus most logically represent the gluon which is present in the proton. These gluons would bind the quasi electrons together to form the fundamental particles

[0144] The masses of all the known particles, including the up and down quarks, the W boson, the muon, charm, strange, the tauon, truth and beauty can thus also be derived from first principles in this fashion, and have the quasi electron as their basic constituent particle (see Appendix 1).

[0145] Thus the structure of the muon (μ) can also be derived from the ratio of the mass of the electron (m_e) and the mass of the muon (m_μ):

$$m_e/m_\mu=4.7\times10^{-3}=c^{1/3}$$

Thus

$$\mu=\epsilon_0{}^{1/2}\times m_e/m_\mu\times3^{-}(4/3=\pi c^3)^{1/2}\times(\pi/c^{1/3})$$

[0146] Where the charge of the muon is in this equation equivalent to that of the electron e. In this case ($\pi/c^{1/3}$) can be considered to represent a specific high energy photon. Thus the structure of the muon, written in short form is:

$$\mu=3^{-}(4/3\pi c^3)^{1/2}\times(\pi/c^{1/3}). \tag{14}$$

[0147] Moreover the structure of the tauon can be calculated from the ratio of the mass of the electron and that of the Tauon (1.79 Mev);

[0148] Thus

0.511 Mev/1.79 Gev=2.85×10^{-4}

$$m_e/m_r=(\pi/c)^{1/3}\times(\pi/c)^{1/9}=2.85\times10^{-4}$$

[0149] As the charge of the tauon is equivalent to the charge of the electron, hence the structure of the Tauon is given by the above equation

$$e_\tau=\epsilon_{qe}\times(m_e/m_\tau)\times3^{-}(4/3\pi c^3)^{1/2}\times(\pi/c)^{1/3}\times(\pi/c)^{1/9}=e$$

[0150] This equation accurately predicts the charge −1; and mass of the Tauon (.~1.78 Gev). Thus the structure of the Tauon can in short form be given by the equation

$$\tau=-3(4/3\pi c^3)^{1/2}\times(\pi/c)^{1/3}\times(\pi/c)^{1/9} \tag{15}$$

[0151] Furthermore a more exact value for the mass of the muon and tauon can be deduced by taking into account the gravitational field of the Earth in a similar way to identifying the exact charge of the electron. In addition it may be necessary to take into account a possible mass value of the neutrino to arrive at a precisely accurate mass value of the muon and tauon. Nevertheless, the mathematical proof of these short form equations lies in the fact that they can very closely identify the charge and the masses of these particles, from first principles, as in equations.

[0152] Overall the mathematical geometrical structure of all the particles can be derived from the quasi electron, which is in turn derived from quintessence. Thus, the short form particle structures can now be derived from first principles. This includes the quasi electron (qe) and electron (e), from which the quarks (u,d) and in turn the stable proton (p) and stable neutron (n) and alpha particle (a) respectively are derived. The general structure of the force carrying bosons the photon. (g) and the gluon (γ) and the intermediate vector boson (W) can be given. It will also intriguingly be possible to derive, according to their generation, the structure of the strange (s) charm (c), beauty (b, or bottom) and truth (t or top) quarks directly from the structure of the muon (μ) and Tauon (τ) respectively.

[0153] Using the term $\Theta=(4/3\pi c^3)$, where, −/+ represents the charge of the quasi electron, we find:

[0154] 1st Generation:

$$(q_e)={}^{-}\Theta^{1/2} \tag{6c}$$

$$e=3^{-1/2} \tag{6}$$

$$d={}^{-}\Theta^{1/2}\cdot3(\pi/c^{1/2}) \tag{16}$$

$$u=2^{+}\Theta^{1/2}\cdot3(\pi/c^{1/2}) \tag{17}$$

$$s={}^{-}\Theta^{1/2}\cdot3(\pi/c^{1/2}) \tag{18}$$

[0155] 2nd Generation

$$\mu=3^{-}\Theta^{1/2}\cdot(\pi/c^{1/3}) \tag{14}$$

$$c=2^{+}\Theta^{1/2}\cdot(\pi/c^{1/3})\cdot(\pi/c^{1/4}) \tag{19}$$

$$b={}^{-}\Theta^{1/2}\cdot(\pi/c^{1/9})\cdot(\pi/c)^{1/4} \tag{20}$$

[0156] 3rd Generation

$$\tau=3^{-}\Theta^{1/2}\cdot(\pi/c)^{1/3}\cdot(\pi/c)^{1/9} \tag{15}$$

$$t=2^{+}\Theta^{1/2}\cdot(\pi/c)^{1/3}\cdot(\pi/c)^{1/9}\cdot(\pi/c)^{1/4} \tag{21}$$

[0157] Particle Gluons (g):

$$g_1=(\pi/c^{1/2}) \tag{22}$$

$$g_2=(\pi/c)^{1/2} \tag{22a}$$

$$g_3=(\pi/c^{1/4}) \tag{22b}$$

$$g_4=(\pi/c)^{1/4} \tag{22c}$$

[0158] Particle Photons (.gamma.):

$$\gamma_1=(\pi/c^{1/3}) \tag{23}$$

$$\gamma_2=(\pi/c)^{1/3} \tag{23a}$$

$$\gamma_3=(\pi/c^{1/9}) \tag{23b}$$

$$\gamma_4=(\pi/c)^{1/9} \tag{23c}$$

[0159] Intermediate Vector Boson ($W^{+/-}$):

$$W^{+}=3^{+}\Theta^{1/2}\cdot 2(\pi/c^{1/6})^{6} \tag{24}$$

$$W^{-}=3^{-}\Theta^{1/2}\cdot 2(\pi/c^{1/6})^{6} \tag{25}$$

[0160] Stable* Proton:

$$p=3^{+}\Theta^{1/2}\cdot 3(\pi/c^{1/2}) \tag{13}$$

[0161] Stable* Neutron:

Stable * Neutron:

$$n = \begin{array}{l} {}^{+}\Theta\downarrow^{1/2}\cdot 2(3\pi/c^{1/2})\cdot{}^{-}\Theta\uparrow^{1/2} \\ {}^{+}\Theta\downarrow^{1/2}\cdot 2(3\pi/c^{1/2})\cdot{}^{-}\Theta\uparrow^{1/2} \\ {}^{+}\Theta\downarrow^{1/2}\cdot 2(3\pi/c^{1/2})\cdot{}^{-}\Theta\uparrow^{1/2} \end{array}$$

* Stable nucleonic neutron and proton conformations differ slightly from the Standard Model, this is due to the sharing of quasi electron and quasi positron particles within the nucleus, which allows stabalisation of these particles by the formation of stable 3×3 structures. The Standard conformations which describe non-nucleonic neutrons and protons are additionally given in Appendix 1.

[0162] Alpha particle (α):

$$\text{Alpha particle }(\alpha) := \alpha = \begin{array}{l} [3^{+}\Theta]\downarrow^{1/2}\cdot 2(3\pi/c^{1/2})\cdot[3^{-}\Theta]\uparrow^{1/2} \\ [3^{+}\Theta]\uparrow^{1/2}\cdot 2(3\pi/c^{1/2})\cdot[3^{-}\Theta]\downarrow^{1/2} \\ [3^{+}\Theta]\downarrow^{1/2}\cdot 2(3\pi/c^{1/2})\cdot[3^{-}\Theta]\uparrow^{1/2} \end{array} \tag{27}$$

[0163] The mathematical proof for these structures and their decay mechanisms is lengthy and is thus fully contained in Appendix 1. All the particle structures are accurately mathematically defined by the masses of these particles.

[0164] The structure of these particles all contain the quasi electron and thus the metric tensor structure necessary in the formulation of the gravitational equations is sustained. The respective forces created by the gluon and the photon are important as they tell us the behaviour of matter and also lead to the likely structure of the graviton

[0165] Particle Spin and Size

[0166] The significance of the electron, composed of three spheres each with a radius of 1/c, is not immediately clear, but can be understood if the frequency of rotation of the electron is also taken into account. Knowing the structure of the electron has led us to deduce its charge and thus may lead us estimate its size and spin. Thus these observations might be used to calculate the radius and rate of rotation of the electron.

[0167] Let us suppose, that nature is truly beautiful, and that the radius of the fundamental quasi electron is indeed 1/c, and in turn the radius was balanced by the velocity of rotation 2π/c. This can be directly confirmed mathematically by taking into account the known spin of the electron, h/4π. Thus the actual spin of the electron may be calculated form the known energy of the spin.

[0168] The radius of the electron is not up till now known, but the radius of a quark has been estimated, and this is the radius derived from deep inelastic collisions of the proton. These estimates reveal a radius of approx. $r_p=1.18\times10^{-15}$ m This value may be used to assist in confirming the spin of the proton in revolutions per sec. (revs) and in turn the spin and size of the electron. Firstly we may proceed to estimate the spin of the proton. Thus as h=E·t (Joules×sec) and h=E·t=F·d·t (Joules×sec), then the spin;

$$h/4\pi=F\cdot d\cdot t \tag{28}$$

[0169] As F=ma, where $\alpha=(\text{revs}\cdot 2\pi)^2 r_p$ and m=the mass of the proton, then

$$h/4\pi=m(revs\cdot 2\pi)^2 r_p\cdot d\cdot t$$

[0170] The actual distance (d) traveled in a circle of half integer spin in 1 second is: $\text{revs}\cdot\pi r_p$ thus:

$$h/4=m(revs\cdot 2\pi)^3 r_p^{\,2}/2$$

[0171] Hence:

$$revs=[h/m(2\pi)^4 r_p^{\,2}]^{1/3}$$

[0172] Taking the effective mass the proton as 1.6726×10^{-27} kg, then the rate of spin of the proton in revolutions/sec is:

$$revs=5.65\times10^{6}\text{ cycles/sec}$$

[0173] From the frequency of the specific rotation of the proton, given the half integer spin associated with the proton, we can thus mathematically confirm the relationship between the radius of a particle and its spin:

$$r_p\times revs/2=1/c \tag{29}$$

[0174] Furthermore, the fundamental radius of 1/c seen in geometric structure the quasi electron, is also reflected in the rotation rate and radius for the proton, thus as above 1/c÷½ revs=1.85×10^{-15} m. Moreover, this means the actual half integer velocity of rotation is none other than 2π/c in metres/sec. So that the particle is in harmonic balance.

[0175] Using the fundamental formula h/4π=F·d·t, it is possible to obtain accurate estimates of the radius and spin rates of the electron, or indeed any particle, using the same principle of harmonic balance. Using the formula:

$$revs=[h/m(2\pi)^4 r_e^{\,2}]^{1/3}$$

[0176] It appears there are two unknowns, the radius if the electron and its revolution rate, however, in accordance with the equation, r_p=2/c.revs, which gives the revolution rate of the proton, the same principle may also be used for the electron, by substituting r_e=2/c.revs, such that:

$$revs=hc^2/4m_e(2\pi)^4 \tag{30}$$

[0177] Taking the mass of the electron 9.109382×10^{-31} kg, the rate of revolution of the electron is:

$$revs=1.048\times10^{10}\text{ cycles/sec}$$

[0178] Which gives a predicted radius of the electron as

$$r_e=6.336\times10^{-19}\text{ cycles/sec}$$

[0179] So the half integer rotation velocity (revs.πr) is 2π/c !, for the electron in keeping with the harmonic balance of the electron.

[0180] The same principle may be used to obtain an accurate estimate of the spin and radius of the muon, or any other particle. Using the above formula

$$revs=hc^2/4m_e(2\pi)^4 \tag{31}$$

[0181] Then as the mass of the muon is 1.8823×10^{-28} kg then the revs of the muon $f_\mu=5.070\times10^{-7}$ cycles/sec and the radius r_μ is thus 1.316×10^{-16} m.

[0182] It is now possible to begin to explain how the muon and the other subatomic particles are formed. If a quasi electron is complexed with another structure the total geometric structure needs to maintain harmonic balance. So the frequency of rotation would need to match geometric structure with which the quasi electron was complexed

[0183] Intriguingly we find asymptotic convergence for the formulas for frequency and mass occurs, when the geometric structure complexed with the quasi electron has the structure represented by ($\pi/c^{1/3}$) [giving the frequency divided by two, because the single integer spin of the force carrying particles compares to a half integer spin for the muon]. So that

$$(f_r3\pi/2)^{1/3}=f_\mu$$

[0184] When the ratio of the masses of the electron (m_e) and muon (m_μ) are related, such that:

$$m_e(3\pi/c^{1/3})=m_\mu$$

[0185] Indeed we find that (allowing for the neutrino) this ratio is very close to the actual ratio of the mass of the electron to the mass of the muon, determined experimentally.

[0186] Furthermore, we have seen that these geometric structures, representing harmonics of the speed of light, which either match the frequency or the amplitude of vibration of the quasi electron, mathematically define the masses of the particles and the fundamental forces of Nature.

Part III—Quantum Gravity

[0187] Quantum General Relativity

[0188] Given the overall energy "complex" energy tensor structure of the electron and the metric tensor, assumed in general relativity, the quantum nature of gravity itself can now be explored. The spherical complex tensor for the electron and the positron give the mathematical quantum structure and energy tensor for all the other particles. Together with the time dimension these nine space dimensions account for the 10 parameters present in the metric tensor necessary to formulate the equations for gravity using Riemann geometry and thus forms the basis of quantum gravity. Intriguingly the metric tensor at each point in space time is required to consist of a collection of ten numbers, Consequently, ten dimensional space-time hypotheses, such as this or superstring theory, do automatically yield general relativity.

[0189] Furthermore, the mathematical representation of the graviton and the gravitational constant may be directly estimated from the knowledge of the mass and radius of quintessence. Thence, the force of the vibrations of quintessence lead directly to quantum gravity.

[0190] The radius of quintessence should be approximately in keeping with the Planck length estimate (r), which is conventionally derived from the standard dimensional equation:

$$r_q^2\cong Gh/c^3 \tag{32}$$

[0191] Given the nine spacial parameters present in the metric tensor, used in general relativity we find that the actual formula for r_q^2 is mathematically in agreement with theory when:

$$9r_q^2=Gh/c^3 \tag{33}$$

[0192] This again supports the 9 dimensional view of space and the size of the vibrations of quintessence can thus be estimated.

$$r_q=1.35\times10^{-35}\text{ m} \tag{33a}$$

[0193] This value is in agreement with the Planck length. Indeed if the above equation is correct then we find that we can derive the standard equation for the general relativistic increase in radius, r', (eq. 34) directly from first principles and arrive at a more fundamental equation for quantum gravity. As

$$r'=G\cdot M/3c^3 \tag{34}$$

[0194] By substituting eq. 33) into equation 34, a fundamental relationship between r' and M is obtained.

$$r'/3r_q^2=GMc^3/Ghc^3=Mc/h$$

[0195] And substituting the quintessential equation, $h=m^qc^2$ (eq. 1) then:

$$r'/3r_q^2=M/m_qc=n_q/c \tag{35}$$

[0196] Hence the ratio of the change in radius to that of the radius of quintessence squared, is proportional, by a factor of c, to the ratio of the mass M of an object to that of the mass of quintessence, effectively the number of quintessences. Thus the change in radius, r' due to gravitation, is related to none other than the ratio of the mass and radius of an object to the mass and the square of the radius of quintessence. Thus again the gravitational change in radius is directly related to the number of quintessences.

[0197] Naturally, this would be exactly what would be logically expected if quintessence, like the equation for the charge of the electron (eq. 6) forms from a root sphere. Thus the change in spacial radius of a normal sphere is dependant on the square of the quintessential radius.

[0198] This increase in apparent radius represents none other than the (gravitational) binding energy for quintessence.

[0199] The meaning of the above dimensional equation (33) might itself be further understood by substituting the mass of quintessence (where $m_q=h/c^2$) into the equation. Thus in nine dimensions the gravitational constant (G) may be more logically given as,

$$9(\pi r_q^2/m_q)=G\pi/c \tag{36}$$

[0200] Where πr_q^2 is the cross sectional area of quintessence and m_q is the effective mass of quintessence, and thus

($\pi r_q^2/m_q$) represents the effective mass per unit area which quintessence exerts. This equation reduces to:

$$9r_q^2/m_q=G/c \tag{37}$$

[0201] From this we may derive the standard general relativistic relationship for the apparent change in radius (r') around a mass (M), from an understanding of the mass m_q and number (n_q) of quintessences. As $m_q=M/n_q$, then:

$$3r_q^2=G\cdot M/3c\cdot n_q \tag{38}$$

[0202] Then if

$$n_q=r'c/3r_q^2 \tag{39}$$

[0203] thus directly substituting for n_q in eq. 38:

$$r'=G\cdot M/3c^2 \tag{34}$$

[0204] The importance of this is that the gravitational change in radius now logically derives from equation 36, which describes the gravitational force as resulting directly from the mass of quintessence exerted/per unit area of quintessence.

$$9(\pi r_q^2/m_q)=G\cdot\pi/c \tag{36}$$

[0205] Thus equation 34 is the conventional equation for the general relativistic increase in radius (r') in a gravitational field, which is here derived from the underlying nature of quintessence. Thus the gravitational constant is derived from the mass and radius of vibration squared of quintessence from first principles.

[0206] Indeed it is apparent that a more fundamental equation for gravitation now exists, for equation (39) is mathematically accurate and numerically agrees with eq. 34:

$$r'/3r_q^2=n_q/c \tag{39}$$

[0207] These equations may be readily mathematically verified. If in accordance with standard general relativity, the apparent increase in radius r' is:

$$r'=GM/3c^2 \tag{34}$$

[0208] Then given that the mass of the Earth is 5.9745×10^{24} kg;

$$r'=1.478\times10^{-3}\text{ m}$$

[0209] Accordingly if $r'=3r_q^2\cdot n_q/c$; (eq. 39). Given the number of quintessences n_q constituting the Earth is M_E/m_q, then

$$n_q=5.9745\times10^{24}/7.3725\times10^{-5}=8.104\times10^{74}$$

[0210] As $r_q^2=1.823\times10^{-70}$ (eq. 33a) then:

$$r'=1.478\times10^{-3}\text{ m}$$

[0211] Thus equation 39 gives the same answer as the standard equation and may be understood on a logical basis. Indeed the meaning of c in the equation may be understood as it has been previously shown as being the basis for the radius of matter (eq. 6). Hence the general relativistic change in radius, r', is none other than the effective binding energy for quintessence.

[0212] Quantum Gravity and Wave Particle Duality

[0213] Quantum gravity can now be readily linked with quantum mechanics, indeed any observations which are self consistent must be able to do so easily.

[0214] The frequency of light has been previously derived

$$f=E/h=n_q$$

[0215] Thus the formula for the frequency of light (E=hf) has previously been explained theoretically by the simple observation that the frequency is determined quite directly from the number of quintessences (n_q) within the photon. The same principle has also been shown to apply to matter.

[0216] Let us now follow these equations for matter by calculating the wavelength of a photon from the Gravitational constant as an example; and also as a test of these observations and to demonstrate that the gravitational equations can also apply to the quantum world.

[0217] If

$$n_q=r'c/3r_q^2 \tag{39a}$$

[0218] where r' is the general relativistic increase in radius, and r_q is the radius of quintessence (eq. 33). Where $f=E/h=n_q$, substituting for n_q, then the frequency of the photon f_γ(where β=1) is given by:

$$f_\gamma=r'c/3r_q^2$$

[0219] Using the standard equation, $r'=GM/3c^2$ (eq. 34); we may substitute for r', thus we have:

$$f_\gamma = GM/9r_q^2c\ 15\text{ Thus}$$

$$f_\gamma = \frac{G}{9r_q^2c^3}\cdot m_\gamma c^2 \text{ and as } E = m_\gamma c^2;$$

$$f_\gamma = \frac{GE}{9r_q^2c^3} \tag{40}$$

Indeed as $9r_q^2 = Gh/c^3$, then $f_\gamma = E/h = n_q$

[0220] It is possible to also demonstrate that the same relationship holds for the wave equation for matter. If we take the relativistic wave energy of matter, which has been previously derived,

$$f=\beta^2n_q$$

[0221] This includes the term for the number of quintessences flowing through the electron, in the complex vectors of space-time, to give the relativistic electron momentum (p) and a term for the rest mass, thus substituting into (40)

$$f_\gamma = \frac{GE}{9r_q^2c^3} \tag{40}$$

[0222] As $f=\beta^2n_q$ for matter then the equation expands to:

$$f_m = \frac{G}{9r_q^2c^3}\cdot\beta^2E \text{ As } \lambda = v/f\text{, then}$$

$$\lambda = \frac{9r_q^2c^3v}{G\beta^2E} \tag{41}$$

[0223] Then the equation again reduces to:

$$G = \frac{9r_q^2c^4}{\lambda\beta E} \tag{3}$$

[0224] Equations 3, 40 and 41 are important as they show that the quantum wavelength of any particle of rest mass m can be derived from the gravitational constant G. Thus linking quantum mechanics to quantum gravity.

[0225] It is therefore important to confirm the numerical accuracy of the above equation (40). We can do this by comparing the result to the standard computation of the de Broglie equation, in a range where de Broglie itself is likely to be most accurate; which according to these observations is in the low energy range (see section on Wave Particle Duality).

[0226] If we take an electron with an energy of 0.1 KeV the wavelength is conventionally given (where the kinetic energy of the electron E_k is given by the product of the charge of the electron (C) and the potential applied eV=0.1 Kev), by the standard equation:

$$\lambda = h/p = h/(E_k \cdot 2m_0)^{1/2} \text{ thus}$$

$$\lambda = 6.63\times10^{-34}/[1.602\times10^{-19}\times1\times10^2\times18.22\times10^{-31}]^{1/2}$$

$$\text{hence}$$

$$\lambda \cong 1.23\times10^{-10}m \text{ Using}$$

$$\lambda = \frac{9r_q^2c^4}{G\beta E} \text{ Where } E = \gamma m_0c^2 \tag{3a}$$

[0227] At 0.1 Kev, electron velocity is 6×10^6 r/sec, thus $\beta=2\times10^{-2}$ and $\gamma=1/(1-v^2/c^2)^{1/2}=1.0002$. Thus:

$$\gamma = \frac{9\times1.82\times10^{-70}\times80.78\times10^{32}}{6.76\times10^{-11}\times\beta\times1.0002\times9.11\times10^{-31}\times8.998\times10^{16}}$$

$$\lambda = 1.21\times10^{-10}m$$

[0228] Divergence between the de Broglie equation and the above equation (2) occurs at intermediate and high energies where it is generally accepted that the standard de Broglie equation may be less accurate. The values for eq. 2 and de Broglie are compared to recent experiments, which demonstrate a relativistic curvilinear plot for wavelengths of matter in keeping with eq. 40.

[0229] The de Broglie equation in the non-relativistic format yields a simple log/linear scale, which is not in keeping with relativity; whereas eq. 3 is dependent on relativity and mathematically accounts for both relativity in calculating the wavelength. Indeed recent experiment on quantum tunnelling through a wire mesh strongly suggests that the relationship between energy and wavelength is relativistically curvilinear(ref 1). Furthermore equation 3a suggests a fundamental relationship between energy (E), relative velocity (v/c=.β), gravity (G) and the quantum wavelength (λ)

$$\lambda = \frac{9r_q^2c^4}{G\beta E} \text{ Indeed as } {}^9r_q^2 = Gh/c^3\text{, then} \tag{3a}$$

$$\lambda = hc/\beta E \tag{2}$$

[0230] Equation 2 is the very same as the Universal wave equation derived form first principles for the wavelength of tight and matter, which allowed a relativistic solution to the equations for wave particle duality (see Wave Particle Duality). This now indicates that these quintessential equations are compatible with relativity, quantum mechanics and quantum gravity.

[0231] Graviton Structure

[0232] From these observations, if the value for the gravitational constant is substituted into the equation (35) we may now estimate the probable geometric structure of the gravitation, which is the force particle mediating gravity by acting on quintessence. Thus the Gravitational constant has been previously derived from the vibration of quintessence by the equation:

$$G\cdot(\pi/c)=9(\pi r_q^2/m_q) \tag{36}$$

[0233] This is in accurate agreement with the value for $G(6.67\times10^{-11}\text{ N m}^2\text{ kg}^{-2})$. This suggests that the most probable mathematical representation of the graviton (φ), the third force carrying particle is

$$\phi=(\pi/c) \tag{42}$$

[0234] Thus the gravitational constant (G) can be given by the mass and radius of quintessence and the structure of the graviton

$$G=9\pi r_q^2/\phi m_q \tag{43}$$

[0235] This shows the gravitational force to be related to the fundamental radius of quintessence space time, and the graviton.

[0236] Quantised General Relativity

[0237] The classical general relativistic formula, as given by Einstein is:

$$R_{\mu\nu}-1/2g_{\mu\nu}R=-\kappa T_{\mu\nu}$$

[0238] Where R is effectively the curvature of space-time, $R_{\mu\nu}$ denotes the contracted Riemann tensor of curvature and $T_{\mu\nu}$ is the "energy tensor" of matter.)

[0239] If we substitute the energy tensor matrix of the electron (eq. 9)×time, for the energy tensor of matter $T_{\mu\nu}$; and the metric tensor of the space-time lattice×time for the contracted Riemann tensor we can arrive at the same solutions for general relativity.

[0240] Furthermore, in his published paper on General Relativity, Einstein. defined the constant κ as:

$$\kappa=8\pi G/c^2$$

[0241] Therefore Einstein's equation should be written as

$$R_{\mu\nu} - 1/2g_{\mu\nu}R = \frac{8\pi G}{c^2}\cdot T_{\mu\nu} \tag{43}$$

[0242] Einstein himself was apparently not happy about the right hand component of the equation. However, we find that this part of the equation can now be explained and quantised by substituting the gravitational constant, $G=9\pi r_q^2/\phi m_q$, (eq. 42a),

[0243] Giving:

$$R_{\mu\nu}-1/2g_{\mu\nu}R=-\frac{8\pi^2 9r_q^2}{\varphi m_q c^2}\cdot T_{\mu\nu} \tag{44}$$

[0244] By substituting $m_q \cdot c^2=h$, and further substituting h=h/2π, we arrive at a quantised solution to Einstein's equations. Where A_q is the surface area of quintessence ($A_q=4\pi r_q^2$); ϕ is the graviton [ϕ=(π/c)] and h is Plancks constant. thus:

$$R_{\mu\nu}-1/2g_{\mu\nu}R=-\frac{9A_q}{\varphi\hbar}\cdot T_{\mu\nu} \tag{45}$$

[0245] The gravitational equation can now be further understood on a logical basis. The term $A_q=(4\pi r_q^2)$, where represents standard term for the surface area of a sphere of quintessence for the 9 space dimensions of the space time lattice, h is the energy content of quintessence×time and ϕ is the graviton, thus the right hand term now represents a true "metric energy tensor" of matter.

[0246] This leads directly to the standard solution to the field equations, for the general relativistic increase in radius r' of an object, where A is the surface area of a sphere of a given mass M, such that

$$r=\sqrt{(A/4\pi)}-r=GM/3c^2 \tag{34}$$

[0247] Furthermore, although equation 45, gives the same solutions as Einstein's equation, which is essentially correct, the difference is that the equation is now dependant upon Planck's constant (h), and moreover the radius of quintessence, which now defines a quantised solution to the equations.

[0248] Graviton Force Characteristics

[0249] Similar to the photon, the previously derived equation (42) for the graviton [ϕ=(π/c)] appears to also mathematically represent a helical ringlet of quintessence, but with a spin of 2. For the photon, taking the direction of motion as the x vector and its axis of spin also as the x vector, would account for the electromagnetic force and its attraction and repulsion characteristics. In the case of the gluon component ($\pi/c^{1/2}$), if the direction vector is x, then the axis of spin would be in the y vector, the same as quasi electrons, accounting for the particle binding characteristics of the gluon force. In the case of the graviton, if the direction of motion was in the x vector, the graviton spin axis would be in the z vector thus, as will be demonstrated, accounting for the gravitational force.

[0250] The spin axis of the graviton can also be derived using the known characteristics of the electron. If an electron is travelling in the x direction, then its spin axis is determined by the by the sign of the ${}_j$y vector (up or down). This view is in agreement with conventional theory, which indicates that the electron spin is similar to a rotating planet orbiting the sun, (the electron even appears to have orbital precession). As the electron passes through the space-time lattice, this spin would generate the formation of gravitons. This would occur as a result of the ejection of the excess quintessence passing through the electron. As the electron spins, the ejection of these gravitons would occur at a tangent to the electron's direction of motion. The ejection of the gravitons would occur, similarly to the ejection of energy of a pulsar or quasar, through the equivalent of the north and south poles of the electron. Thus, propelling the graviton in the direction of the electrons y vector. The ejection of the graviton would re-orientate and impart a specific angular momentum to the gravitons which would thus end up spinning on its own z axis. If for instance the graviton is released from an up spin electron the graviton will be rotating clockwise and its leading edge will displace quintessence downwards. In turn this will provide an upwards force.

[0251] This picture accounts for Fleming's left hand rule, is logical and provides an explanation for the magnetic force around a wire. According, to the left hand rule if the direction of the current is in the x vector, the magnetic field is in the z vector, and the force is upwards, in the y vector, in accordance with the above model. Therefore, this particular spin axis and the structure of the graviton results in its force characteristics. As the graviton is very small compared to the electron and both have different rather rapid spin axis it is difficult for these to bind and interact. Nevertheless, because the graviton has a spin of 2, and as it spin axis is perpendicular to its direction of motion, in the z vector, it readily displaces space-time quintessence to produce gravity. Thus because the graviton is able to displace space-time, it is capable of escaping a black hole. How else could the effects of gravity be felt beyond a black hole?

[0252] Quantum Gravity and Electromagnetism

[0253] With the above electron model of graviton production the nature of magnetism can be understood from first principles, Furthermore, the presence of a space-time lattice links relativity, and the forces of gravity with the electromagnetic and other forces of Nature. Indeed, evidence for these links may first date back to the 1820's, when Andre Ampere first defined the Amp. The force of attraction between two parallel wires 1 metre apart each carrying 1 Amp in a vacuum was defined as none other than the permeability of free space (2×10^{-7} N per metre of conductor). Thus conventionally the magnetic field strength around a long straight wire is given as:

$$B=\mu_0 I/2\pi r$$

[0254] Where I is the current and μ_0 is the permeability of free space ($4\pi\times10^{-7}$ N A^{-2})

[0255] The attraction between two wires both carrying negative charge is, however, counterintuitive as negative charges should repel. A conventional explanation overcomes this by invoking the presence of a magnetic field which is created by the current by the production of virtual photons. Thus we appear to have an explanation for the effects of magnetism which involves virtual photons, however, these photons are not observed. More accurately, according to conventional special relativity the magnetic field is none other than the electric field viewed relativistically.

[0256] A more satisfactory explanation, therefore, lies in the interaction between the electrons and the space time lattice. The moving electrons in the two wires interact with the lattice to produce gravitons; which are in phase when both streams of electrons are traveling in the same direction. The gravitonic waves interact constructively to disperse the space time lattice between the wires and induce an attractive force between the two wires, which produces in effect the permeability of free space. Thus this force results from the vibration of quintessence itself.

[0257] Conversely in two wires with current going in opposite directions the graviton waves are in anti-phase and would interact destructively between the wires. The gravitonic waves traveling radially outward from the wires would, however, disperse the lattice outside the two wires and produce apparent repulsion between the wires, which is exactly what is observed. These effects of electricity suggest that gravitons act as waves and that phase is important.

[0258] This effect is also seen with the north and south poles of ferromagnets. Nevertheless, with matter other than iron, cobalt or nickel, the graviton emission cannot be phased as the atoms are unable to align and magnets do not appear to exist with other materials.

[0259] In ordinary magnetic system the release of gravitons from the north pole would be exactly balanced by those released from the south pole of the magnet and hence there would be no net force on the magnet until an external magnet or electrical current were applied.

[0260] Overall the magnitude of the forces in electrical systems where electrical conduction occurs are well defined by the permeability and permittivity of free space μ_0, and ϵ_0. Where v is the constant velocity of the charge and ϵ is the electric field produced by the charge.

$$B=[\mu_0\epsilon_0]v\cdot\epsilon$$

[0261] These observations suggest that the forces of electricity which produce magnetism are indeed related to the permittivity and permeability of free space and that these quantities are exerted by an apparent vacuum. Thus the effects of magnetism could be explained by none other than the phased effects of gravitational waves on the space time lattice.

[0262] Electromagnetism is of further interest to quantum gravity, particularly if we combine the standard equations, $B=\mu_0 I/2\pi r$ and $B=[\mu_0\epsilon_0]v\cdot\epsilon$, substituting for B we have:

$$2\pi r=I/\epsilon_0 v\epsilon \qquad (46)$$

[0263] Thus $2\pi r$ is proportional to the inverse of ϵ_0. Thus as space time is dispersed by gravitons the permittivity field will increase in the same way capacitance increases with separation of plates. Because of the inverse relationship between ϵ_0 and $2\pi r$, as ϵ_0 increases the circumference of a circle and the apparent ratio of π is to r, will appear to diminish in accordance with general relativity. This not an actual diminution in the circumference of a circle but the effective reduction of the resistance to motion in a circular path in this field.

[0264] Incidentally, the above observations, also lead us directly to Schrödinger's formula for the average equilibrium distance (r) between an electron with charge (e) in orbit around a proton, which is conventionally given by:

$$r=n^2 4\pi h^2\epsilon_0/me^2$$

[0265] Where ϵ_0 is again the permittivity of free space, m is the mass of the electron and n is an orbital integer, h is Planck's constant and e is the charge of the electron. Furthermore if $e=[\epsilon_0/3(4/3\pi c^3)]^{1/2}$ (eq. 3); then the equation at n=1, for the electron orbital radius elegantly simplifies to:

$$r=4h^2c^3/m$$

[0266] Hence the orbital radius of the electron is related to spin of the electron (h) and its mass (m).

[0267] Quantum Gravity and the Charge of the Electron

[0268] The equation for the charge of the electron (eq. 1) contains the term so (permittivity of free space) which according to these observations should vary in a gravitational field.

$$e=[\epsilon_0/3(4/3\pi c^3)]^{1/2} \qquad (6a)$$

[0269] If we combine the standard equations, $B=\mu_0 I/2\pi r$ and $B=[\mu_0\epsilon_0]v\cdot\epsilon$, substituting for B we have:

$$2\pi r=I/\epsilon_0 v\epsilon \qquad (46)$$

[0270] Thus $2\pi r$ is proportional to the inverse of ϵ_0. Thus as space time is dispersed by gravitons the permittivity field will increase in the same way capacitance increases with separation of plates. Because of the inverse relationship between ϵ_0 and $2\pi r$, as ϵ_0 increases the circumference of a circle and the apparent ratio of π to r, will appear to diminish in accordance with general relativity.

[0271] Thus ϵ_0 rises when space-time is dispersed by the gravitons that produce the gravitational field, This occurs in a similar way to the process by which capacitance increases with separation of plates in a capacitor.

[0272] Nevertheless, as c is a constant and as $c=[\mu_0\epsilon_0]^{-1/2}$, then if ϵ_0 rises then μ_0 falls. This is entirely consistent as μ_0, which represents the force that quintessence exerts, would be reduced if the quintessence space time lattice is dispersed.

[0273] Furthermore, as $\mu_0=4\pi\times10^{-7}$ N A^{-2}; then as μ_0 falls, then the apparent ration π to r, also falls in a gravitational field. This is largely the same as stating, as does general relativity, that the apparent radius r', rises in a gravitational field. So this view is consistent with general relativity.

[0274] Nevertheless, to derive an exact value for the charge of the electron we must account for gravity in the above equation. We will take the specific example of the Earth's gravitational field in order to obtain the exact value for the electron. If in accordance with standard general relativity, the apparent increase in radius r' is:

$$r'=GM/3c^2 \qquad (34)$$

[0275] Then given that the mass of the Earth is 5.9745×10^{24} kg; then

$$r'=1.47864\times10^{-3}\text{ m}$$

thus

$$2\pi r'=9.29057\times10^{-3}$$

[0276] Which is the incremental factor by which ϵ_0 must increase in Earth's gravitational field. So to correct ϵ_0 to account for gravity, ϵ_0 must be divided by the incremental factor, $2\pi r'$. Similarly as effectively π decreases in a gravitational field, to correct π to account for gravity it must be multiplied by this incremental factor. So the equation for an electron in a zero gravitational field is:

$$e=[\epsilon_0/3(4/3\pi c^3)]^{1/2}\div(1+2\pi r')=1.6022\times10^{-19}\ \text{C} \qquad (6b)$$

[0277] This now gives the charge of the electron as measured in a zero gravitational field as 1.6022×10^{-19} C, which is the same as that measured on Earth. Notably these observations appear to suggest that the charge of the electron is the same irrespective of the gravitational field.

[0278] Virtually unlimited degrees of accuracy for the charge of the electron and for the fine structure constant (α), may be achieved by taking into account 2nd and nth order gravitometric effects. Thus if we take into account the effect of gravity upon the radius of the Earth it is also important to take into account an effect upon the instruments with which we measure quantities, this would be a second order gravitometric effect. Thus taking into account 2nd order effects (r''), we have a very small, but nevertheless relevant change, such that: $r''=r'(1+2r')$. Thus $2r''=9.3180486\times10^{-3}$, and thus:

$$e=[\epsilon_0/3(4/3\pi c^3)]^{1/2}\div(1+2\pi r'')=1.6021765\times10^{-19}\ \text{C} \qquad (6c)$$

[0279] This agrees exactly to the nearest 7 decimal places with the maximum accuracy of the experimental value for the charge of the electron. Furthermore by taking into account the nth order gravitometric effect, it is theoretically possible to predict accuracy for the charge of the electron to 3n decimal places. This mathematically accuracy confirms the structure of the electron from first principles and indeed the theoretical effects of gravity on the permittivity of free space (ϵ_0).

[0280] This returns us directly to the fine structure constant for the electron which is conventionally given by: $\alpha=e^2/hc\cdot4\pi\epsilon_0$. If $\alpha=e^2=\epsilon_0/3(4/3\pi c^3)$, accordingly the quintessential equation for α is structurally given by: $2\pi/\alpha=m^q[3\Theta]^2$ (where $\Theta=4/3\ \pi c^3$; see The Structure of the Electron and Matter), we must now take into account the effects of gravity, as above, thus:

$$2\pi\alpha/m_q[3\Theta]^2\div(1\_\_2\pi r'')^2=0.007297353$$

[0281] Where the gravitational term for the increase in radius r'' allows the mathematical derivation of $\alpha=0.007297353$, and the above equation is in agreement with the conventional experimental value for $\alpha=0.007297353$ to the nearest 9 decimal places.

[0282] Hence the term $(1\_\_2\pi r'')^2$ is in accordance with these observations for the effect of gravity on electromagnetic forces. To a maximum accuracy governed by current knowledge of the mass of the Earth and the Gravitational constant and thus the term for the gravitational increase in radius r'. These observations can also be used to accurately predict the magnetic moment of the electron

[0283] Thus the presence of the fine structure constant can now be further understood, by deriving the constant from first principles; specifically from the actual dimensional conformation for the charge of the electron: $e=[\epsilon^0/3(4/3\pi c^3)]^{1/2}$ (eq. 6).

[0284] Overall the fine structure constant α (allowing for the term r' which is the general relativistic increase in the radius of the Earth due to gravitation) is given by none other than the formula for the mass of quintessence and from the structure of the electron, which can now be derived from first principles to seven decimal places or more.

[0285] Quantum Gravity and the Electron Magnetic Moment

[0286] The theoretical origin and nature of magnetism remains obscure in current electromagnetic theory. An explanation suggests these magnetic effects are produced by photons, although no photons have ever been observed. To get round this difficulty it is postulated by physics that magnetism results from "virtual" photons. However, Maxwell's equation for electromagnetism states that the photon has no net magnetic effect.

$$\delta B_x/\delta B_y/\delta y+\delta B_z/\delta z=0$$

[0287] Thus magnetism could not, by the above standard equation, be derived from a photon real or virtual.

[0288] In addition observational data suggests that black holes have powerful magnetic fields and as in theory photons are unable to escape from black holes (except for small quantities in the form of Hawking radiation), it would be difficult to explain these magnetic fields on the basis of photon emission.

[0289] Einstein postulated that magnetism was merely due to special relativity .sup.(ref 17). The postulate for the nature of magnetism in these current observations, states that the magnetic force results from relativity due to none other than the phased emission of gravitons (why postulate two invisible forces, magnetism and gravity, when one, the graviton, will do). This view as previously discussed (Quantum Gravity and Electromagnetism) is entirely compatible with standard relativity$^{(\text{ref }19)}$. Thus with the graviton origin of magnetism, the equation for the magnetic moment of the electron should have an expression in terms of quintessence and in turn the gravitational force and in particular the graviton.

[0290] The standard term for the magnetic moment of the Bohr Magneton (SIB) is:

$$\mu B=eh/4\pi m_e$$

[0291] In standard quantum mechanics the Bohr Magneton, μB, however, needs to be corrected to agree with experiment. The "correction factor" is termed "ϵ"; where $\epsilon=(\alpha/2\pi)-0.328\alpha^2/\pi^2=0.001159641$. Thus theory reveals μ_e, the magnetic moment of the electron where:

$$\mu_e=(eh/4\pi m_e)[1+(\alpha/2\pi)-0.328\alpha^2/\pi^2]$$

[0292] The conventional derivation of the term ϵ above, is given from the fine structure constant, $(\alpha/2\pi)$ which is theoretically consistent. However, a rather arbitrary mathematical correction term; $0.328\alpha^2/\pi^2$ needs to be used in this standard equation. This appears ad hoc and needless to say, more accurate measurements show, the electron magnetic moment to the Bohr magneton ratio, $1+\epsilon=1.001159652$, which suggests the correction factor is indeed incorrect. Nevertheless, this correction factor is essential for "renormalisation" and thus for quantum mechanics to work.

[0293] Quantum gravity readily explains the discrepancy between the theoretical Bohr Magneton (μB) and the actual measured magnetic moment of the electron (μ_e). In accordance with the above chapter (Quantum Gravity and the Charge of the Electron)

[0294] Thus the significant mathematical discrepancies can be removed by accounting for the effects of quantum gravity.

[0295] Thus taking the charge of the electron (e), using the equation for the Bohr magneton and the effects of quantum gravity such that gravitational change in radius is r". The magnetic moment of the electron is given by:

$$\mu_e=(eh/4\pi m_e)(1+[\alpha/2\pi\div(1+r'')])$$

[0296] This gives an electron magnetic moment to Bohr magneton ratio of 1.00115968. Thus the mathematical term for the magnetic moment of the electron is given, avoiding the arbitrary and dubious term $0.328\alpha^2/\pi^2$ used in the standard equation, simply by accounting for quantum gravity.

[0297] It is now possible to unite the equations for gravity and magnetism by substituting the fundamental key equations of quantum gravity. Thus if: $h=3m_qc^2$ (eq. 1b) and $m_e=m_qn_q$ (eq. 2). Then we can express the magnetic moment of any particle with the charge of the electron, including the proton, in terms of the number of quintessences (n_q) in that particle.

$$\mu B=ec^2/(4/3\pi n_q); \tag{47}$$

[0298] Given that the postulated structure of the graviton is: ($\phi=\pi/c$) (eq. 42), then substituting we have

$$\mu B=3ec/4\phi n_q \tag{48}$$

[0299] Showing that the equations for the magnetic moment are compatible with the gravitational equations given earlier. Principally, the quintessential equations now allow the determination of the magnetic moment of any charged object from the equation for the graviton and directly from the number of quintessences it contains. In conventional physics the magnetic moment of the electron requires a correction factor, $(1+(\alpha/2\pi c)-0.328\alpha^2/\pi^2)$, to derive the correct experimental value. These observations herein, indicate that the correction factor is more logically (1+r"), where r" is the general relativistic increase in radius around a gravitational body. This suggests that magnetism is not only affected by gravity, but can, as shown as above, be derived using the quantum gravitational equations.

[0300] Quantum Gravity and Special Relativity

[0301] Ordinary matter passing through the lattice would produce gravitons which would interact with space-time as described by general relativity. The quantity of gravitons would be determined by the apparent mass and in turn these would apparently curve space time. The geometry of this "curvature" is elegantly described by general relativity using Riemann geometry, specifically using metric tensors. Intriguingly the metric tensor is not a single number, but at each point in space time it is required to consist of a collection of ten numbers, Consequently, ten dimensional space-time hypotheses, such as this or superstring theory, may automatically yield general relativity

[0302] General relativity is indeed very elegant, nevertheless there was a logical step yet to answer. That is, how do gravitons shape space time? This can now be readily answered by considering the interaction of a three dimensional space time lattice with gravitons themselves to produce the effects of gravity. The effects of gravity are as such to compel a body in motion towards the gravitational object and to a much smaller extent visa versa. This effect can only be produced If gravitons repel quintessence (the constituents of the 3D lattice). Indeed, it has been stated that in order to explain cosmic inflation and the "flatness" of the Universe that quintessence must shun (or be shunned by) matter.

[0303] In descriptive terms a body close to a large mass will have a tendency to move toward it because the three dimensional lattice would be less dense as it approached the surface of the large mass. Overall there would be less resistance to motion in the direction of the large mass, and the motion in this direction would be facilitated by the vibration of quintessence.

[0304] In general relativity the principle governing motion is the geodesic of least distance, this can be re-expressed using similar equations using least action. Furthermore, the concept of motion due to the vibrations of quintessence is more logically and experimentally compelling.

[0305] These observations can now be used to link general and special relativity. Thus as we approach the speed of light, the mass of an object travelling through the space-time lattice would approach infinity, directly because the number of quintessences passing through a body would increase with increasing velocity, hence the equation:

$$m'=m_0/(1-v^2/c^2)^{1/2} \tag{50}$$

or

$$m'=m_0/(1-v^{2\,[\mu}{}_0\epsilon_0])^{1/2} \tag{50a}$$

[0306] In turn this would generate increasing gravitons and accordingly this would explain the observed effects of special relativity. Time itself is due to passage through the space-time lattice, and where the space-time lattice is dispersed by gravitons, time and length are reduced with increasing velocity and hence increasing space-time lattice dispersion, similar to the way in which gravity alters space-time

[0307] As a result:

$$t'=t(1-v^2/c^2)^{1/2},\ l'=l(1-v^2/c^2)^{1/2}$$

[0308] Thus resulting in the effects of special relativity.

[0309] Quintessence and Black Holes

[0310] To address the relationship of the space-time lattice to gravity directly, it is important to discuss the concept of quintessence with regard to general relativistic equations. The standard general relativistic equation for the apparent increase in radius (r) due to the curvature of space time around a gravitational object, which has also been previously derived from first principles (eq. 36), is:

$$r'=GM/3c^2 \tag{34}$$

[0311] This can also thus be written as:

$$3r'=GM[\mu_0\epsilon_0] \tag{51}$$

[0312] This standard equation, is in keeping with the above observations. Specifically, as the mass increases, ϵ_0 increases, in turn the radius will appear to increase (relative to π).

[0313] The above observations now allow us to examine the effects with regard to the interior of black holes themselves. The event horizon would represent a critical density for quintessence, in which light could not escape. The Schwarzschild radius would now be given by:

$$R_s=2GM[\mu_0\epsilon_0]$$

[0314] The event horizon will occur at the point at which there is less resistance to circular motion than motion in a straight or partially curved line. Given that π is proportional to $1/\epsilon_0$. The event horizon should occur when the permittivity has increased by a factor of π.

[0315] Effectively because the permittivity of free space rises, π decreases. This is entirely in keeping with general relativity which predicts the effective change in the ratio of the radius to the circumference as given by the conventional equation, where r', is again the apparent change in radius.

$$r=GM/3c^2$$

[0316] Hence π will effectively decrease as we approach the event horizon of a black hole, When π decreases to 1, the circular circumference is equal to the diameter and moreover, inside this limit it is shorter for light to travel in a circle. Thus light cannot escape the event horizon.

[0317] This can give us great insights into the workings of space-time, for flat Euclidean space the standard equation is:

$$e^{i\pi}=-1$$

[0318] In accordance with general relativity, the ratio of the radius to the circumference changes in a gravitational field, and effectively π=1, at the event horizon, thus the boundary condition for the shape of space-time at the event horizon now has the direct equation:

$$e^{f}$$

[0319] Within a black hole as the permittivity of space increases by a factor of 2π an object within it will complete two rotations rather than travel in a straight line. In effect exceeding the speed of light by 2π. Hence, the condition for space-time is represented by the equation:

$$e^{i/2}$$

[0320] Thus an increase in the permittivity of free space by a minimum factor of π, to produce a black hole is estimated to result from an increase in mass by a factor of approx. 10^6 (the ratio of the mass of the earth and that of a putative black hole).

[0321] Continuing with the subject of a black hole, according to the model inside the black hole, the gravitons produced by the matter present would be in equilibrium with the density of the space-time lattice. Increasing the rate of rotation of the matter in the black hole for instance would thus increase the production of gravitons and its effective mass and increase the radius of the event horizon. A density gradient of the space-time lattice would continue to exist within the black hole. Progressively closer to the center of a black hole matter itself would be increasingly compressed and the spherical structure of the quasi electron would be predicted to collapse. This collapse would result in the formation of an exotic form of matter in the form of pure quintessence in a black hole.

[0322] This pure quintessence would produce the singularity at the centre of the black hole. The larger the black hole in terms of mass the more pure quintessence would exist at its core.

[0323] Quintessence and the Big Bang

[0324] Quintessence theory not only predicts the occurrence of the Big Bang, but allows a prediction for the value of the entire mass of the Universe, from first principles.

[0325] In accordance with quintessence theory the big bang resulted from the explosion of an immense black hole singularity, which was constituted from pure quintessence.

[0326] On the basis of quintessence, there will be a critical mass for Big Bang event; thus if entire space-time between quintessence is compacted so that no further quintessence can be accommodated, the addition of further quintessence would destabilize the immense black hole, resulting in the Big Bang.

[0327] It is possible to predict this critical mass, using the radius of quintessence as a benchmark. Given the nine spatial parameters present in the metric tensor, used in general relativity we find that the actual formula for the radius of quintessence; r_q^2, is mathematically in agreement with general relativistic theory when:

$$9r_q^2=Gh/C^3 \tag{33}$$

[0328] This again supports the 9 dimensional view of space (so crucial in superstring theory). Moreover, the size of the vibrations of quintessence can thus be calculated as:

$$r_q=1.35\times10^{-35}\text{ m} \tag{33a}$$

[0329] The volume of each quintessence is thus:

$$4/3\pi r_q^3=1.0306\times10^{-104}\text{ m}^3 \tag{33b}$$

[0330] So to be accommodated within unit volume of space time, with no intervening apparent space time, (given that each of 9 overlapping quintessences are required) would require approx.

$$9\times10^{104}\text{ quintessences}$$

[0331] As the mass of quintessence is $m^q=h/c^2=7.373\times10^{-51}$ kg sec (eq. 1). Then the mass of the Universe, to two decimal places, is:

$$1.18\times10^{53}\text{ kg} \tag{33c}$$

[0332] This is in close agreement with a recent estimate of the mass of the Universe from COBE and other satellite data, which estimates the mass to be 100 trillion trillion trillion trillion tonnes (10^{53} kg)

[0333] Moreover, the early formation of the galaxies can be readily explained, it is likely that in such a big bang some very small black holes might have prevailed and that these formed the seeds of the galaxies we see today.

[0334] The event horizon, calculated from the Schwartzschild radius, of such an immense black hole is about 10^{26} m, which would have allowed Guth's inflationary component to the early expansion of the Universe.

[0335] In addition, inflation may result directly from the observation that once electrons have formed from the primordial soup of quintessence, they emit gravitons which in turn repel space time, which might also result in another cosmic inflationary cycle.

[0336] Most importantly quintessence theory explains the Big Bang from first principles and is capable of accurately predicting the mass of the Universe.

[0337] The Nature of Energy

[0338] These observations allow a fundamental understanding of energy. The quantum physical, minimum component of energy is Planck's constant; h. To define the minimal component of mass, using the standard energy equivalence formula; $E=mc^2$, such a minimal mass (m_q)

would be required to have the value equivalent to; $m_q=h/c^2$ (1). The total mass of a system (m) would then be; $m=m_q n_q$, where (n_q) is the number of these minimal units. Thence, the total energy of a system can be derived from the minimal energy; h, multiplied by the number of these energy units (n_q). Thus as, $E=mc^2$, then also $E=m_q n_q c^2$ and substituting $m_q=h/c^2$, the energy equivalence formula has the more logical formulation;

$$E=hn_q \tag{1a}$$

[0339] Thus the energy of a system is equivalent to the minimal energy unit; h, multiplied by the number of those minimal energy units (n_q)

[0340] This leads directly to a deeper understanding of wave particle duality and the wave nature of matter.

[0341] This is encapsulated by the quintessential energy formulae

[0342] As conventionally $\beta \cdot E/c=p$, then

$$\lambda=h/p=hc/\beta E \tag{2}$$

and

$$E=hn_q \tag{1a}$$

then

$$\lambda=c/\beta n_q \tag{2b}$$

[0343] Importantly, as indicated by equation (2b), energy having no quintessence; would have a wavelength of infinity. Specifically pure energy containing no quintessences, would have a lambda of infinity. According to quantum mechanics an infinite wavelength would result in the probability of that energy being anywhere. As energy itself has no electrical charge it would not be impeded by the permittivity and permeability of the three dimensional space-time lattice. Moreover, energy would not be detectable in three dimensional space-time, unless it interacted with matter, as in the EPR experiments. Indeed, energy is not observed when not bound to any form of mass or particle.

[0344] Thus equation 2b, takes us to our original assertion regarding the existence of pure energy.

[0345] Energy is not Bound by the Space-Time Lattice

[0346] Thus, as the EPR experiments suggest the existence of energy separate from matter and thus separate from the three dimensional space-time lattice, it is interesting to find that experiment suggests the existence of free energy in a continuum separate from space time and matter to produce the effects of quantum teleportation.

[0347] This is not, however, teleportation across an additional dimension, this is a term to describe in partially familiar terms the dissociation of energy from the three dimensional space-time lattice. As time is inextricably linked to each dimension of space, the effects of energy would be inextricably linked to the events, such as the creation of virtual particles, we see interacting within space-time. It is unlikely that observers have any direct day to day experience to explain quantum events. Nevertheless, quintessence theory may have given us a window into the hitherto hidden workings of the Universe. Thereby, the mystery of the uniformity of the Universe, across distances which the speed of light could not apparently traverse, is readily explained by the fact that the free energy contained in the Universe is not bound by the space-time lattice.

[0348] In the case of light, due to the exceedingly small masses involved, there would be relatively easy exchange of matter with free energy within a photon. This would make the photon the ideal experimental tool to look for energy which is not bound by matter and in turn energy which is not bound in space-time. Indeed, very recently Furusawa et al. have reported to have observed the transference of energy as photons from A to B, without those photons traversing space-time. This finding which has been supported using other experimental techniques, is very important as it suggests the existence of such free energy.

[0349] Overall, quintessence theory gives an a priori explanation for the concept of mass, the elementary particles, the forces of nature and quantum effects. It can equally be used logically to explain the inner physics of a black hole, the missing mass in the Galaxy, the expansion of the Universe, Guth's inflationary theory and predicts the Big Bang, from first principles.

[0350] Part IV: Applied Theory with the Intent to Create Closed Timelike Curves

[0351] Electron Bombardment of the Photosphere to Induce Gravitational Shift

[0352] Leveraging the above relationship between mass and quintessence we derive that if an electromagnetic radiation with velocity v strikes the event horizon singularity of rest inertial mass mi, and U is the electromagnetic energy absorbed by the singularity, then, according to Maxwell's prediction, a momentum q=U/v is transferred to it. Mass shift $d \cdot m_b$, dependent on the external electromagnetic energy, equals the inertial mass shift dependent on the increment of energy in the particle. Since in this case the inertial mass shift does not depend on velocity V, i.e., it is related only to the momentum q absorbed, it can be obtained by making p=0 in variation $\Delta H=H'-H=c[q^2+(m_i c)^2]^{1/2}-(m_i c)^2$ from the particles inertial Hamiltonian. Consequently, the expression of d_{m_p}, is written as:

$$d_{m_p}=\Delta H/c^2=m_i\sqrt{1+U/m_i c^2\left[\varepsilon_r u_{r/2}\left\{\sqrt{1+d/w\varepsilon^2+1}\right\}^2=1\right.}$$

[0353] Comparing now the expression of m_i and m_g we have $m_g=m_i-2dm_b$. By replacing m_b in this equation, given by equation above, we obtain the expression of the correlation between gravitational mass and inertial mass, i.e.,

$$m_g=m_t-2\sqrt{1+U/m_t c^2\left[\sqrt{\varepsilon_r u_{r/2}\{[d/we]^2+1\}^2-1}\right.}\cdot m_t$$

[0354] We see that only in the absence of electromagnetic radiation on the event horizon (U=0) is the gravitational mass equivalent to the inertial mass. Note that the electromagnetic characteristics, e, m and s do not refer to the singularity itself, but to the outside medium around the singularity (photosphere) in which the incident radiation is propagating.

[0355] Stable CTC Solution From Modified M-Theory

[0356] Our innovation is a method of creating an event where a dual membrane or dual boundary condition exists. We do this using generalization from a Misner space which has been modified if one analytically continues the maximal extended Misner Metric so that $ds^2=-du^2+dw^2+(dx^2)^2+(dx^3)^2$ to the Euclidean section so that u=iota zeta we obtain a Misner instanton on the section where w and zeta are both real. The Euclidean time, t, and the closed spacelike coordinate are both periodic, the later having a period of $2\Pi t^2$. Going back to the Lorentzian sector we find that the period of the closed coordinate becomes linearly dependent on the physical time. Using automorphic fields in the Hadamard function one can obtain a quantized condition for time. This gives us a figure on the order of the Plank time. This confines such a stable wormhole condition to the general area of the Plank scale in its modified form which is that area defined by the Membrane itself. As such, these Plank scale wormholes are the true source of the true virtual aspects of the vacuum and quintessence. The effects of the dual singularity system can be viewed as an overlap zone of two distinct space-times which have boundary conditions on both sides. As an object accelerates towards C it is this same boundary or horizon that object encounters when the Time defined horizon solution for the universe is imposed. At this point drawing upon Van Den Broeck's alterations for this space-time geometry to create a single closed Friedman-Robertson-Walker spacetime the space-time geometry can be represented by this equation $ds^2=dct^2-B^2[(dx-\beta fdct)^2]+dy^2+dz^2$ B can be any function that is large near the displacement device. We then consider this transformation as extended to four dimensional space-time with arbitrarily time dependent acceleration. We also present the device frame energy density T00 from a four dimensional calculation and note that the 4d classical calculation is everywhere finite.

[0357] Consider an Alcubierre interval given according to a remote frame's cylindrical coordinates by:

$$ds^2=(1-\beta^2f^2)dct^2+2bfdctdz-dz^2-dr^2-r^2df^2$$

where f is a function that is 1 at the location of the device and zero far from it.

[0358] Starting out with the first transformation z'=z−òct beta dct. Where b is first expressed here as a function of time ct. With some algebra for simplification this results in

$$ds^2=[1-\beta^2(1-f)^2]dct^2-2b(1-f)dctdz'-dz^2-dr^2-r^2df^2$$

Let g=1−f and this becomes

$$ds^2=[1-\beta^2g^2]dct^2-2bgdctdz'-dz'^2-dr^2-r^2df^2$$

[0359] Notice that this returned the original intervals form with a reversal on the sign of b and a reversal of the boundary conditions for g. Now we notice that at r=0, this interval becomes the interval for special relativity transformed to cylindrical coordinates. Thus, we have found a transformation to a frame based local to the device. One can also verify that in these coordinates the relevant affine connections vanish at r=0. a further proposed modification to this field we will reintroduce a time dilation term into the devices frame's interval. Only we will use different boundary conditions for it. We will keep A=1 both at the location of the device, and far from it, but allow it to become large in the warped region. This is achieved by the simplest means possible in the proposed field generation method. Since rotating kerr singularity under bombardment would produce an equal negative energy region. With the effect focused outward away from the device slightly by the relative alignment of the kerr singularity the actual inward going portion of such fields would overlap and cancel their effect out in the region of the device. This yields a space normal time region, which is again restored at the fringes of the outward going field. From the standpoint of a geometric picture of space-time around the device and extending outwards one has actually created a dual event horizon situation as far as time goes. The inner one is the shell of the canceled out field inside of which normal time flow is restored. The outer one is formed at the boundary where normal time resumes and within which we have a negative energy field.

SUMMARY OF THE INVENTION

[0360] The present invention is A method for the generation of a pseudo 2+1 dimensional anti-de Sitter space (DeDeo & Gott 2002) using two Kerr type positively charged rotating dilation singularities where one singularity is maintained as a axis of rotation or "reference" singularity, and the other "target" singularity is subjected to a differential electron flow so as to simultaneously pass above the photosphere of said singularity in its direction of rotation—prograde orbit—and contrary to its direction of rotation—retrograde orbit—to release a directed flow of gravitons in a sinusoidal oscillation simulating a rotational effect of the "target" singularity around the axis of rotation provided by the "reference" singularity, resulting in the creation of timelike curves in a compact time-oriented manifold permitting topology change from one spacelike boundary to the other in accordance with Geroch's theorem (Geroch 1967) which results in a method for the formation of G odel-type geodesically complete spacetime envelopes complete with closed timelike curves.

BRIEF DESCRIPTION OF THE DRAWINGS

[0361] **FIG. 1** is a schematic representation of the mechanism employed to house the components necessary to generate a 2+1 dimensional anti-de Sitter space, resulting in the creation of timelike curves in a compact time-oriented manifold

[0362] **FIG. 2** is a schematic representation of the G odel-type geodesically complete spacetime envelope created by the mechanism complete with closed timelike curves

DETAILED DESCRIPTION OF THE INVENTION

[0363] Principles of Gravity Distortion Time Displacement Systems

[0364] The theoretical understanding of quantum gravity allows the design of time displacement systems from first principles. It is unlikely that gravitons or Kerr singularities can be controlled in a precise way using current technology. Nevertheless, an understanding of three dimensional space-time and matter, does allow the design of elementary displacement systems. That is, systems whose displacement rely on direct warping space-time as opposed to the ejection of material to provide thrust resulting in time dilatational effects.

[0365] The background for these systems are already partially understood and quintessence theory allows their further development. For this invention, the formation of black holes in the laboratory represents a crucial step in understanding the mechanisms that underlay gravitational physics and in turn the warping of space-time. The existence of black holes permits the localized application of the Axial torsion Spin-Rotation Coupling Effect (Zhang & Beesham 2002) resulting in a Rotating Frame with Relativistic Factor (Zhang 2003) which can be used in the creation of a Alcubierre space time bubble under the Van Den Broeck modification of the Alcubierre geometry (Van Den Broeck 1999) resulting in a method for the generation of a pseudo 2+1 dimensional anti-de Sitter space (DeDeo & Gott 2002) using two Kerr type positively charged rotating dilation singularities where one singularity is maintained as a axis of rotation or "reference" singularity, and the other "target" singularity is subjected to a differential electron flow so as to simultaneously pass above the photosphere of said singularity in its direction of rotation and contrary to its direction of rotation to release a directed flow of gravitons in a sinusoidal oscillation simulating a rotational effect of the "target" singularity around the axis of rotation provided by the "reference" singularity. In this instance the space-time lattice would be repelled by gravitons in such a way as to disperse space-time quintessence in a circular fashion around each of the singularities, producing multiple event horizons around the simulated axis of rotation provided by the reference singularity.

[0366] The release of gravitons from the target singularity is controlled by differentially governing the electron flow across the photosphere with the use of powerful electric currents. In turn the differential direction of flow across the photosphere of gravitons would determine the direction of motion through the space time lattice and the resulting time displacement from one spacelike boundary to the other in accordance with Geroch's theorem (Geroch 1967) when implementing this approach it is important to remember that the black hole is not excited by the smashing of the clump as it "hits" the horizon. The hole is rather excited when the metric perturbation associated with the clump is "felt" by the background metric. The excitation event therefore constitutes a smooth process whereby in-fall of a clump from .rmb and through r+ serves as a source in the Teukolsky (1973) equation for small perturbations to the Kerr geometry (with appropriate boundary conditions at r+ and r∞). This is an important distinction with a great deal of relevance to the practical engineering of the displacement unit since we need to gauge the "driving" of QNR modes in terms of an e_ective coupling from clump in-fall.

[0367] What results is a method for resonant driving of the quasi-normal ringing (QNR) wave modes of the Kerr geometry of the target singularity. The micro black hole hyper-accreting at rates ÿ M. 1 M□sec. 1 from a neutrino cooled disk is pushed through to oscillate near resonance of its (l,m=2, 2) quadrupole QNR frequency due to the in-fall of compact mass over-densities from the cusp in e_ective potential on a dynamical time scale. This mode is induced via induced magneto-rotationally induced fluid dynamics in the ultra-relativistic region of the flow bounded from below by the marginally bound orbit radius: rmb If the QNR modes are fed resonantly for a few seconds of hyper-accretion, the enhanced amplitude of the oscillations yields a very high rate of energy deposition into gravitational waves. Indeed, the integrated energy deposition is large enough to "evaporate" the equivalent of a factor of a few times the total rest mass-energy of a single clump into gravitational waves, which in turn interact with the reference singularity

[0368] Application of the method described in the previous section results in translation outside of the cosmological horizon, where F(R)<0. The topology of the resulting geodesic, for large constant R, is mathematically equivalent to a Euclidean cylinder of the condition R×Sn where T is the coordinate along the cylinder. I± are located outside the future/past cosmological horizons, where R is timelike and T is spacelike. In the case of a rotating Kerr black hole, there are two circular photon orbits that can exist in the equatorial plane and be exploited to create a tipler sinusoid. One is a prograde orbit moving in the same direction as the black hole's rotation, while the other is a retrograde orbit moving against the black hole's rotation. Their radii are respectively given by

$$r1=2M(1+\mathrm{Cos}(2/3\ \arccos(-|\alpha|/M)))$$

$$r2=2M(1+\mathrm{Cos}(2/3\ \arccos(|\alpha|/M)))$$

where a is the angular momentum per unit mass of the black hole. The orbits fall in the range M greater or equal to r1 greater or equal to 3M greater or equal to r2 greater or equal to 4M The fact that a prograde photon or in or case cooper pair with Bose characteristics orbits the black hole at a smaller radius than a retrograde one can be attributed to the well-known Lense-Thirring effect, i.e., the dragging of inertial frames due to the black hole's rotation which we control and influence via electromagnetic induced load coupling with the open magnetic field lines threading the BH horizon. This dragging would cause charged cooper pairs to revolve around the black hole relative to a static observer at infinity. Thus, to such an observer, a prograde cooper pair would have to orbit at a smaller radius to compensate for the 'extra' angular momentum acquired, while a retrograde one would have to orbit at a larger radius to compensate for the 'lost' angular momentum. Indeed, in the limit of zero rotation, these two orbits coincide at r D 3M, giving the single circular orbit of the Schwarzschild black hole. Now, recall that orbits around the Schwarzschild black hole are necessarily confined to a plane passing through its center, because of the spherical symmetry of the space-time. However, the Kerr black hole space-time has only an axial symmetry (in addition to being stationary), and this raises the possibility of non-planar orbits. One could, for example, contemplate the existence of spherical Boseon orbits—orbits with constant coordinate radii that are not necessarily confined to the equatorial plane—around the Kerr black hole. Such orbits would be a nontrivial generalization of the two circular photon orbits that lie in the equatorial plane. At first it may seem a little surprising that such spherical orbits could even exist, but there is an interesting reason as to why they are possible. Note that an object in a spherical orbit would, in addition to moving around the black hole in the azimuthal direction, be undergoing some periodic motion in the latitudinal direction. This is only possible if there is a conserved quantity associated with motion in this direction, just as angular momentum is necessarily conserved by its rotational motion in the azimuthal direction. (This result can be seen, for example, using action-angle variable) Now, because the Kerr space-time has only axial symmetry, geodesics in it should have only two constants of motion, namely energy and angular momentum. However, Carter discovered the remarkable fact that geodesics in the Kerr space-time possess a third constant of motion. It turns out that Carter's new constant governs the motion of geodesics in the latitudinal direction, although it is not related to any obvious space-time symmetry. Thus, spherical timelike orbits, which assume eigenlike properties around the Kerr black hole, are possible and when coupled with charged Bosons may be exploited to manipulate the relative event horizons of the "target" singularity. The location and the temperature of the modified event horizon depend on the time, charge and angle of incidence of the cooper pairs. The Fermionic spectrum of Dirac particles displays a spin-rotation coupling effect due to the interaction between the particles with spin-1=2 and the black holes with rotation. The effects arise from the interaction between the spin of Dirac particles and the rotation of the evaporating black holes. The feature of this spin-rotation coupling effect is its dependence on different helicity states of coupled particles with spin-½ and its irrelevance to the mass of particles.

[0369] In order to design a mechanism for time displacement exploiting the Carter asymettry cited above we utilize two positively charged top spin rotating kerr type black holes aligned on demand to create a simulated rotational effect around a central axis provided by the "reference" singularity. In the case of the Kerr blackhole (singularity), this is accomplished by an inverse Blandford-Znajek (BZ) process utilizing the magnetic flux of open field lines connecting the horizon and an induced remote load. (Ding-Xiong Wang, Kan Xiao & Wei-Hua Lei, 2001) permitting the microscopic blackhole (singularity) to be rotated along its horizontal axis in at relativistic centrifugational speeds. A differential current is then applied in such a way as to pass through the entire photosphere in the desired direction. As a result the gravity field can be manipulated by three factors that affect it in distinct ways. Adding electric charge to the singularities increases the diameter of the inner event horizons. Adding mass to the singularities increases the area of gravitational influence around the singularities. Rotating and positioning the polar axis of the singularities affects and alters the resulting gravity sinusoid

[0370] In order to create a sinusoid capable of inducing a topology change from one spacelike boundary to the other in accordance with Geroch's theorem The electric charge in the upper half of the photosphere would be maximised. The electrons will have a vector in the left to right direction as the singularity spins clockwise. If a maximised current is applied to the singularity in the same direction this will result in a increase in the velocity of the electrons relative to the centre of gravity of the singularity, due to the flow of current. In turn, according special relativity and to the space-time lattice model, this wilt result in an increase in the relativistic mass if the electrons and in turn by general relativity an increase in the release of gravitons.

[0371] Conversely in the lower half of the singularity the electrons will have a vector of motion in the right to left direction due to the spin of the singularity. This will be relativistically slowed by the differential current applied in the same direction as the current above, and hence in the opposite direction to the direction of rotation. The charge can be separately applied and adjusted to ensure that the electrons are relativistically stationary relative to the centre of gravity. In turn this will minimize the relativistic mass and result in a decrease in the release of gravitons for the lower half of the singularity.

[0372] The overall result will be a greater release of gravitons in one direction and a lesser release of gravitons in the converse direction. The effect will be enhanced by the use of a multi-phasic current simultaneously applied. This will result in the release of multi-phasic gravitons which will disperse space time in the singularity with increase in density in the converse this effect can also be produced and supplemented with the use radio frequency pulses, The radio frequency pulses must be designed to produce a change in the spin of the particle to enhance the release of gravitons in the desired direction

[0373] With the use of large currents the drift velocity of the electrons across the photosphere could be greatly increased. Within this region the electrical resistance is virtually eliminated. Thereby allowing large currents to be induced with minimum total power output.

[0374] The result is the production of cooper paired electrons of high speed and hence high relativistic mass in the desired half of the singularity, whilst producing low speed and thus low mass paired electrons at the converse of the singularity, in accordance with special relativity. The imbalance in the rotating singularity will be continuously present creating a dynamic warping of space-time. In effect, the differential current flow, will produce differential graviton production and in turn, by general relativity, the warping of space-time

[0375] As the cooper pairs cross the photosphere, accretion of the Bose particles results. Near-hole accretion across the target singularity is then motivated by magneto-rotationally induced, ultra-relativistic disk dynamics in the region of the flow bounded from below by the marginally bound geodesic radius rmb. As the particles impelled have high spin values, a largely coherent magnetic field in this region has the dynamical implication of compact mass segregation at the displacement nodes of the non-axisymmetric, MRI modes. This results in prolific gravitational wave emission coincident with the gamma-ray stage. The gravitational wave emissions are then manipulated to influence the reference singularity to produce a variable gravitational sinusoid which is then used to mathematically approximate the gravametric distortion, inducing a topology change from one spacelike boundary to the other in accordance with Geroch's theorem.

BIBLIOGRAPHY

[0376] Mesoscopic behaviour of the neutral Fermi gas ^{3}He confined in quantum wires. S. L. Phillipson, A. M. Genault, S. N Fisher, G. R. Pickett, P. J. Y Thibault. Nature, 395: 578-580.

[0377] Direct Observation of d-orbital Holes and Cu—Cu Bonding in Cu_2. J. M. Zuo et al. Nature 401: 49-52.

[0378] The Foundation of the General Theory of Relativity, A Einstein. Translated from "Die Grundlage der Allgemeinen Relativittstheorie". Annalen der Physik, 49, 1916. Published in, Einstein, The Principle of Relativity, Dover publications, 1952.

[0379] Sidelights on Relativity. Ether and the Theory of Relativity Albert Einstein. Dover Publications 1983.

[0380] Inflation is dead long live inflation. George Musser. Scientific American, July 1998, p 9.

[0381] The flip side of the Universe. George Musser. Scientific American, September 1998, p 13.

[0382] The Fifth Element, Marcus Chown. New Scientist, Apr. 3, 1999, pp 29-32

[0383] The Quantum Universe. Tony Hey and Patrick Walters. Cambridge University Press 1987. (Solving the Shrodinger equation pp. 165-167).

[0384] Complex Numbers, Mathematics for Engineers. W. Bolton; Longman Scientific & Technical; Chap. Introducing Complex Numbers, p 1-11.

[0385] Unconditional Quantum Teleportation. A. Furusawa, I. Sorensen, S. L. Braunstein, C. A. Fuchs, H. J Kimble, E. S. Polzik. Science, Oct. 23, 1998. Vol. 282: 706-709.

[0386] Complete quantum teleportation using nuclear magnetic resonance. M. A. Neilsen, E. Knill, R. Laflamme. Nature, Nov. 5, 1998. Vol. 396: 52-55.

[0387] Experimental quantum teleportation. D. Bouwmeester et al. Nature, 390: 575-579 (1997).

[0388] Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein-Podolski-Rosen channels. D. Boschi, S. Branca, F. De Martini, L. Hardy, S. Popescu. Phys. Rev. Lett. 80: 1121-1125 (1998).

[0389] W. Tittel, J. Brendel, H. Zhuniden, N. Gisin, Phys. Rev. Lett. 81: 3563-3566. (1998) into the Vortex. Charles Seife. New Scientist, Oct. 24, 1998. p 7.

[0390] Direct observation of a fractional charge. R de Picciotto, M Reznikov, M. Heibium, V Umansky, G. Bunin, D. Mahalu. Nature, 389: 162-164.

[0391] Tables of Physical and Chemical Constants. Kay & Laby. Longman Scientific & Technical.

[0392] Physics in Focus. Michael Brimicombe. Section, Wave Mechanics; p 521. Nelson Special Relativity. A. P. French. Nelson. Chapter 8, Special Relativity and Electricity, pp 256.

[0393] Universe In Balance. Jeff Peterson. New Scientist, 168: p 27

[0394] The Accelerating Universe. Mario Livio. John Wiley and Sons, Inc. (2000)

[0395] Escape from Earth, Charles Seife, New Scientist, Feb. 6, 1999. p 6.

[0396] The Discovery of the Top Quark. Tony Liss and Paul Tipton. Scientific American; September 1997, pp. 26-41.

[0397] The Mystery of the Nucleon Spin. Klaus Rith and Andreas Schaifer. Scientific American: July 1999, pp 42-47.

[0398] And you're glue. Frank Wilczek. Nature; 400, pp 21-23.

[0399] The warp drive: hyper-fast travel within general relativity Miguel Alcubierre —Classical and Quantum Gravity, vol. 11, L73-L77, (1994)

[0400] Photon propagation in a stationary warp drive space-time: Claes R. Cramer (1995)

[0401] Some thoughts on the Implications of Faster-Than-Light Interstellar Space Travel: I. A. Crawford, Quarterly Journal of the Royal Astronomical Society, vol. 36, 205-218, (1995)

[0402] Physical and Cosmological Implications of a Possible Class of Particles Able to Travel Faster than Light: Luis Gonzalez-Mestres: Contribution to the 28th International Conference on High Energy Physics, Warsaw (Poland), (1996)

[0403] Warp drive and causality: Allen E. Everett, Physical Review D, vol. 53, 7365-7368, (1996)

[0404] A Superluminal Subway: The Krasnikov Tube, Allen E. Everett & Thomas A. Roman, Physical Review D, vol. 56, 2100-2108, (1997)

[0405] Quantum effects in the Alcubierre warp drive spacetime, William A. Hiscock, Classical and Quantum Gravity, vol. 14, L183-L188 (1997)

[0406] The unphysical nature of "Warp Drive", Michael J. Pfenning & L. H. Ford; Classical and Quantum Gravity, vol. 14, 1743-1751, (1997)

[0407] On the Possibility of a Propulsion Drive Creation Through a Local Manipulation of Spacetime Geometry; Vesselin Petkov; Presented at the 34th

[0408] AIAA/ASME/SAE/ASEE Joint Propulsion Conference, (1998)

[0409] 'Operational' energy conditions, Adam D. Helfer; Classical and Quantum Gravity, vol. 15, 1169-1183, (1998)

[0410] No warp drive; D. H. Coule; Classical and Quantum Gravity, vol. 15, 2523-2527, (1998)

[0411] Quantum Inequality Restrictions on Negative Energy Densities in Curved Spacetimes, Michael John Pfenning; Doctoral Dissertation, (1998)

[0412] Hyperfast Interstellar Travel in General Relativity; S. V. Krasnikov; Physical Review D, vol. 57, 4760, (1998)

[0413] Superluminal travel requires negative energies; Ken D. Olum; Physical Review Letters, vol. 81, 3567-3570, (1998)

[0414] Hyper-fast travel without negative energy, Eric Baird; (1999)

[0415] Warp drives, wavefronts and superluminality, Eric Baird; (1999)

[0416] A traversable wormhole, S. Krasnikov; (1999)

[0417] Speed Limits in General Relativity; Robert J. Low; Classical and Quantum Gravity, vol. 16, 543-549, (1999)

[0418] Null geodesics in the Alcubierre warp drive spacetime: the view from the bridge; Chad Clark, William A. Hiscock & Shane L. Larson; Classical and Quantum Gravity, vol. 16, 3965-3972, (1999)

[0419] A 'warp drive' with more reasonable total energy; Chris Van Den Broeck; Classical and Quantum Gravity, vol. 16, 3973-3979, (1999)

[0420] On the warp drive space-time, Pedro F. Gonzalez-Diaz; Physical Review D, vol. 62, 44005-44012, (2000)

[0421] On the (im)possibility of warp bubbles; Chris Van Den Broeck; Summary of talk delivered at STAIF-2000, (2000)

[0422] Reduced Total Energy Requirements for a Modified Alcubierre Warp Drive Spacetime; F. Loup, D. Waite & E. Halerewicz Jr. (2001)

[0423] Warp Drive With Zero Expansion, Jose Natario; Classical and Quantum Gravity, vol. 19, 1157-1166, (2002)

[0424] A Causally Connected Superluminal Warp Drive Spacetime; F. Loup, R. Held, D. Waite, E. Halerewicz, Jr., M. Stabno, M. Kuntzman & R. Sims, (2002)

[0425] Weak Energy Condition Violation and Superluminal Travel; Francisco Lobo & Paulo Crawford (2002)

[0426] On the Problems of Hazardous Matter and Radiation at Faster than Light Speeds in the Warp Drive Space-Time; C. B. Hart, R. Held, P. K. Hoiland, S. Jenks, F. Loup, D. Martins, J. Nyman, J. P. Pertierra, P. A. Santos, M. A. Shore, R. Sims, M. Stabno & T. O. M. Teage, (2002)

[0427] 2+1 gravity, chaos and time machines; Bengtsson I and Brannlund J, J. Math. Phys. 42 (2001) 3565

[0428] Time Travel in Einstein's Universe: the Physical Possibilities of Travel through Time, Gott J R; Houghton Mifflin Company, Boston, 2001

[0429] Gravitational memory of natural wormholes, Anchordoqui L A Trobo M L Vucetich H Zyserman F Mod Phys Lett A 15(6): 429-438, 2000

[0430] A traversable singularity-free wormhole. The gravitational memory phenomenon Gott time machines, BTZ black hole formation, and Choptuik scaling, Birmingham D, Sen S; Phys Rev Left 84 (2000) 1074-1077

[It is shown that the Gott time machine provides a mechanism for black hole formation]

[0431] Negative energy, wormholes and warp drive, Ford L H, Roman, T A; Sci Am: 46-53, 2000 Jan (reply: 2000 May p. 12)

[0432] Warp drive space-time; Gonzalez-Diaz P F; Phys Rev D62: 044005 (2000)

[0433] Misner-brane cosmology, Gonzalez-Diaz P F; Phys Lett B486: 158-164 (2000)

[0434] Kinks, energy conditions and closed timelike curves; Gonzalez-Diaz P F; Int J Mod Phys D 9(5): 531-541, 2000. (gr-qc/0007065)

[0435] Will we travel back (or forward) in time? Gott J R; Time, Apr. 10, 2000 pp 68-70

[0436] Charged universes of the Godel type with closed timelike curves; Klepac P and Horsky J, Class Quant Grav 17(13): 2547-2555, 2000

[0437] The Ori-Soen time machine, Olum K D; Phys Rev D61 (2000) 124022

[0438] How much energy do closed timelike curves in 2+1 spacetimes need?, Tiglio M H, Phys Rev D61 (2000) 08 1503

[0439] Energy conditions and their cosmological consequences, Visser M and Barcelo C gr-qc/0001099

[0440] Cylindrically symmetric spinning Brans-Dicke spacetimes with closed timelike curves Anchordoqui L A, Perez Bergliaffa S E, Trobo M L, Birman G S; Mod Phys Lett A14 (1999): 1105-1112

[0441] From spinning to non-spinning cosmic string spacetime; De Lorenci V A De Paola R DM Svaiter N F, Class Quant Grav 16(10): 3047-3055, 1999

[0442] Bounds on negative energy densities in static space-times, Fewster C J, Teo E; Phys Rev D 5910(10): 4016, 1999

[0443] Decohering a charged scalar field in a time-machine wormhole background, Flagga M SN; Int J Theor Phys 38(8): 2121-2161, 1999

[0444] Rotating black holes in unified theories; Gibbons G W. Progress of Theoretical Physics Supplement. (136): 18-28, 1999

[0445] Supersymmetric rotating black holes and causality violation; Gibbons G W and Herdeiro C AR, Class Quant Grav 16 (11) 3619-3652, 1999

[0446] Perdurance of multiply connected de Sitter space; Gonzalez-Diaz P F, Phys Rev D 59, 1999, 123513

[0447] Nonorientable spacetime tunneling; Gonzalez-Diaz P F, Garay L J, Phys Rev D 59 (1999) 064026

[0448] Can the universe create itself?; Gott J R, and Li X L, Phys Rev D 58, 1998, 023501. Also: astro-ph/9712344

[0449] Supersymmetric rotating black holes and causality violation; Holst S, Matschull H-J Class Quant Grav 16, 1999, 3095

[0450] Godel-type universes in string-inspired charged gravity; Kanti P, Vayonakis C E, Phys Rev D 60, 1999, 103519

[0451] Quantum field theory and time machines; Krasnikov S., Phys Rev D 59: 4010, 1999 Jan 14 Dec. 1998

[0452] Time machines constructed from anti-de Sitter space; Li L X, Phys Rev D 59 (1999) 084016

[0453] Wormholes in string theory, Vollick D N., Class Quant Grav 16(5):1599-1604, 1999

[0454] Time machines and the breakdown of unitarity; Antonsen F and Bormann K, Int J Theor Phys 37(9): 2383-2393, 1998

[0455] An exterior for the Godel spacetime; Bonnor W B, Santos N O, MacCallum M A H, Class. Quant. Grav. 15 (1998) 357-366

[0456] All the worlds a time machine, Chown M, New Scientist 157(2124): 38-41, 1998 Mar. 7.

[short exposition of M. Hadleys idea that an elementary particle is a local ST warp that it permits closed timelike curves. Hence general relativity could reproduce (explain) the effects of QM]

[0457] Quantum time machine; Gonzalez-Diaz P F Phys Rev D 58 (12): 4011, 1998

[0458] A design for a quantum time machine, Gray A quant-ph/9804008

{misleading title. On QM measurement problem; quantum "seeing into future."}

[0459] Time machines with non-compactly generated cauchy horizons and "handy singularities"; Krasnikov S V, gr-qc/9711040 (To appear in the proceedings of the 8th Marcel Grossman Conference)

[0460] A singularity-free WEC-respecting time machine, Krasnikov S V, Class Quant Grav 15 (1998) 997-1003

[0461] Self-consistent vacuum for Misner space and the chronology protection conjecture, Li L X and Gott J R, Phys Rev Lett 80, 2980, 1998

{Misner space stable to quantum flutuations; against CPC}

[0462] Decoherence caused by topology in a time-machine spacetime; Mensky M B and Novikov I D, Int. J. Mod. Phys. D5 (1996) 1-27

[0463] Godel metric as a squashed anti-de Sitter geometry, Rooman M, Spindel Ph, Class Quant Grav 15 (1998) 3241-3249

[0464] Testing causality violation on spacetimes with closed timelike curves; Rosenberg S, Phys Rev D 57(6):3365-3377, 1998

{politzer TM used to investigate the scattering experiment; interesting results on agreement with the no CTC case if there is enough TM's}

[0465] A new non-Hausdorff spacetime model for resolution of the time travel paradoxes; Sharlow M F Annals of Physics 263(2): 179-197, 1998

[0466] Particle creation near the chronology horizon; Sushkov S V, Phys Rev D 58, 1998, 044006

[0467] Divergences in the effective action of acausal spacetimes, Cassidy M J, Class Quant Grav 14 p 523, 1997 {shows that a self-consistent quantum vacuum for Misner space exist, allowing time travel}

[0468] The 1-loop instability of Misner space; Cassidy M J, Class Quant Grav 15 (8),

[0469] Superluminal subway—the Krasnikov tube; Everett A E and Roman T A, Phys Rev D 56 (4): 2100-2108, 1997

[0470] Time machines and the principle of self-consistency as a consequence of the principle of stationary action (ii)—the Cauchy problem for a self-interacting relativistic particle, Carlini A and Novikov I D, Int J Mod Phys D 5(5): 445-479, 1996 Oct.

[0471] Warp drive and causality, Everett A E, Phys Rev D 53 (12) pp 7365-7368, 1996 Ringholes and closed timelike curves, Gonzalez-Diaz P F Phys Rev D 54(10): 6122-6131, 1996

[0472] Time machine as four-dimensional wormhole; Guts A K gr-qc/9612064

[0473] Gott time machines in the Anti-de Sitter space; Holst S, Gen Rel Grav 28 (1996) 387-403

[0474] The quantum stability of the time machine, Krasnikov S V, Phys Rev D54 (1996) 7322-7327

[0475] Must time machine be unstable against vacuum fluctuations?, Li L X, Class Quant Grav 13 (1996) 2563-2568

[0476] Three-dimensional billiards with time machine, Mensky M B and Novikov I D, Int J Mod Phys D 5(2): 179-192, 1996

[0477] Decoherence caused by topology in a time-machine spacetime, Mensky M B and Novikov I D, Int J Mod Phys D 5(1): 1-27, 1996.

[0478] Improved time-machine model; Soen Y and Ori A, Phys Rev D 54(8): 4858-4861,

[0479] Unitarity restoration in the presence of closed timelike curves; Anderson A, Phys Rev D51 (1995) 5707-5715

[0480] Problems in time-machine construction due to wormhole evolution; Antonsen F and Bormann K, Int J Theor Phys 34 1995, p 2061

[0481] Time machines: the principle of self-consistency as a consequence of the principle of minimal action; Carlini A, Frolov V P, Mensky M B, Novikov I D, and Soleng H H, Int J Mod Phys D4 (1995) 557-580; Erratum-ibid. D5 (1996) 99

[0482] Nonlinearity in quantum theory and closed timelike curves; Cassidy M J, Phys Rev D52 (1995) 5676-5680

[0483] Unitarity of Quantum Theory and Closed Time-Like Curves; Fewster C J, Wells C G, Phys Rev D52 (1995) 5773-5782

[0484] Time machines without closed causal geodesics; Low R J, Class Quant Grav 12(5): L 37-L 38, 1995 May.

[0485] A remark on kinks and time machines; Yurtsever U, Gen Rel Grav 27(6): 691-693, 1995

[0486] The warp drive: hyper-fast travel within general relativity; Alcubierre M, Class Quantum Grav 11, L73 (1994)

[0487] Energy-momentum restrictions on the creation of Gott time machines; Carroll S M, Farhi E, Guth A H and Olum K D, Phys Rev D 50 pp 6190-6206, 1994

[0488] Kinks and time machines; Chamblin A, Gibbons G W and Steif A R, Phys Rev D 50(4) pp R2353-R2355, 1994

[0489] The breakdown of quantum mechanics in the presence of time machine, Goldwirth D S, Perry M J and Piran T, Gen Rel Grav 25 pp 7-13 1994

[0490] Quantum propagator for a nonrelativistic particle in the vicinity of a time machine; Goldwirth D S, Perry M J, Piran T and Thorne K S, Phys Rev D 49(8): 3951-3957, 1994

[0491] Time machine, self-consistency and the foundations of quantum mechanics; Gonella F, Found Phys Lett 7(2): 161-166, 1994

[0492] Time walk through the quantum cosmic string; Gonzalez-Diaz P F, gr-qc/9412034

[0493] Unitarity and causality in generalized quantum mechanics for nonchronal spacetimes; Hartle J B, Phys Rev D 49 (12) pp 6543-6555, 1994

[0494] (2+1)-dimensional spacetimes containing closed timelike curves; Headrick M P, Gott J R; Phys Rev D 50, 1994, 7244-7259

[0495] Time travel, Jackiw R.; Ukrainskii Fizicheskii Zhurnal. 39(1): 6-14, 1994.

[0496] On the classical stability of a time machine; Krasnikov S V, Class Quant Grav 11 (11): 2755-2759, 1994 Nov.

[0497] Isometries between Gott's two-string spacetime and Grant's generalization of Misner space, Laurence D, Phys Rev D 50 pp 4957-4965

[0498] Negative Energy Densities in Extended Sources Generating Closed Timelike Curves in General Relativity with and without Torsion; Soleng H H, Phys Rev D49 (1994) 1124-1125

[0499] Multiparticle solutions in 2+1 gravity and time machines; Steif A R, Int J Mod Phys D 3(1): 277-280, 1994

[0500] VanVleck determinants—traversable wormhole spacetimes; Visser M

[0501] Algebraic approach to quantum field theory on non-globally-hyperbolic spacetimes; Yurtsever U, Class Quant Grav 11 (1994) 999

[0502] Cosmic strings and chronology protection; Grant J DE, Phys Rev D 47, 2388 (1993)

[0503] Complex geometry, quantum tunneling, and time machines; Li L X, Xu J M and Liu L, Phys Rev D 48(10): 4735-4737, 1993

[0504] Quantum field theory in spaces with closed timelike curves; Boulware D G, Phys Rev D46 (1992) 4421-4441

[0505] Global structure of Gott's two-string spacetime; Cutler C, Phys Rev D 45 p. 487-494, 1992 {Cauchy horizons for Gott's machine (two-string space-time)}

[0506] "Cosmic flashing" in four dimensions; Ford L H and Roman T A, Phys Rev D 46, 1328-1339 (1992)

[0507] Unitarity of interacting fields in curved spacetime; Friedman J L Papastamatiou, Phys Rev D 46 4442, 1992

[0508] Feilure of unitarity for interacting fields on spacetimes with closed timelike curves; Friedman J L Papastamatiou, and Simon J Z, Phys Rev D 46 4456-4469, 1992

[0509] Particle creation for time travel through a wormhole; Kim S W, Phys Rev D 46 2428-2434, 1992

[0510] Time machine and self-consistent evolution in problems with self-interaction; Novikov I D, Phys Rev D 45 p. 1989, 1992

[0511] Structure of the singularity inside a realistic rotating black hole, Phys Rev Lett} 68, 2117-2120; Ori A (1992)

[0512] Simple quantum systems in spacetimes with closed timelike curves; Politzer H D, Phys Rev D 46 4470-4476, 1992

[0513] Quantum mechanics near closed timelike lines; Deutsch D, Phys Rev D 44 (10) p 3197-3217 1991

[0514] Billiard balls in wormhole spacetime with closed timelike curves: Classical theory; Echeverria F, Klinkhammer G and Thorne K S, Phys Rev D 44 (4) p. 1077-1099, 1991

[0515] Closed timelike curves produced by pairs of moving cosmic strings: Exact solutions; Gott J R, Phys Rev Left 66, 1126, (1991)

[0516] Inner structure of a charged black hole: An exact mass-inflation solution; Ori A, Rev Left 67: 789, 1991

[0517] On the outcome of Kerr-like collapse; Charlton N and Clarke C J S, Classical and Quantum Gravity 7 No. 5 (May 1990)

[0518] Cauchy problem in spacetimes with closed timelike curves; Friedman J, Morris M S, Novikov I D, Echeverria F, Klinkhammer G, Thorne K S and Yurtsever U, Phys Rev D 42 (6) 1915-1930, 1990 {QM fixes the TM classical billiard paradox (Polchinski's prdx. "CTC's may be not as nasty as people have assumed". The principle of self-consistency in time travel reviewed.}

[0519] Traversable wormholes from surgically modified Schwarzschild spacetimes, Visser M, Nucl Phys B 328 pp 203-212

[0520] Traversable wormholes: some simple examples; Visser M; Phys Rev D 39: 3182-3184 1989

[0521] Minimal acceleration requirements for time travel in Godel spacetime; Malament D, Journal of Mathematical Physics 26, 1985, p 774-777

[0522] Timelike nongeodesic trajectories which violate causality: A rigorous derivation; de Felice F, Nuovo Cimento 65 B pp 224-232, 1981

[0523] Rotating cylinders and the possibility of global causality violation; Tipler F J, Phys Rev D 9 pp 2203-2206, 1974

[0524] Global structure of the Kerr family of gravitational fields, Carter B, Phys Rev 174, p 1559, 1968

[0525] Complete analytic extension of the symmetry axis of Kerr's solution of Einstein's equations, Carter B, Phys Rev 141, p 1242, 1966

[0526] Acta Physica Polonica, Staruszkiewicz A 24 1963, p 734

[0527] An example of a new type of cosmological solutions of Einstein's field equations of gravitation, Goedel K, Rev Mod Phys 21 p. 447, 1949

[0528] Gravitational field of a distribution of particles rotating about an axis of symmetry; Van Stockum W J, Proc R Soc Edin 57 pp 135-154, 1937

[0529] Appendix A: The Synthetic creation of micro singularities. Throughout this patent application the creation of micro singularities has been maintained. While at first glance this may seem to be a deficiency in the patent application it is in fact an engineering problem that is in the process of being solved. In the International Journal of Theoretical Physics 41 (11): 2073-2090, November 2002 L. J. Garay published "Black Holes in Bose-Einstein Condensates". In this article Garay showed that there exist both dynamically stable and unstable dilute-gas Bose-Einstein condensates that, in the hydrodynamic limit, exhibit a behavior completely analogous to that of gravitational black holes. This discovery coupled with creation of color glass condensate (CGC) an extreme form of nuclear matter in which a nucleus travels at near-light (relativistic) speed which flattens like a pancake in its direction of motion and spawns a large number of gluons hold the key to high energy singularity formation.

I claim:

1. A method for the generation of a pseudo 2+1 dimensional anti-de Sitter space comprising the steps of:

creating two Kerr type positively charged rotating dilation singularities, including the steps of

maintaining one of the singularities as a axis of rotation reference singularity,

maintaining the other of the singularities as a target singularity, and

subjecting the target singularity to a differential electron flow so as to simultaneously pass the differential electron flow above a photosphere of said target singularity in a direction of rotation thereof and contrary to the direction of rotation thereof, in order to release a directed flow of gravitons in a sinusoidal oscillation simulating a rotational effect of the target singularity around the axis of rotation provided by the reference singularity.

2. A method of generating a force around a body, comprising the steps of:

employing sinusoidal oscillations of electrical bombardment on the surface of one Kerr type reference singularity in close proximity to a second Kerr type target singularity to take advantage of the Lense-Thirring effect, wherein the electrical currents employed in the bombardment are passed simultaneously across the photosphere of said reference singularity in its direction of rotation and contrary to its direction of rotation to release a directed flow of gravitons in a sinusoidal oscillation simulating a rotational effect of the target singularity around the axis of rotation provided by the reference singularity;

creating timelike curves in a compact time-oriented manifold of Godel-type geodesically complete spacetime envelope under the Van Den Broeck modification of the Alcubierre geometry, resulting in the creation of timelike curves in a compact time-oriented manifold permitting topology change from one spacelike boundary to the other in accordance with Geroch's theorem.

* * * * *

US 20070191075A1

(19) **United States**

(12) **Patent Application Publication**

Greene et al.

(10) **Pub. No.: US 2007/0191075 A1**

(43) **Pub. Date: Aug. 16, 2007**

(54) **IMPLEMENTATION OF AN RF POWER TRANSMITTER AND NETWORK**

(75) Inventors: **Charles E. Greene**, Cabot, PA (US); **Daniel W. Harrist**, Carnegie, PA (US); **John G. Shearer**, Ligonier, PA (US); **Michele Migliuolo**, Pittsburgh, PA (US); **Gregory W. Puschnigg**, Ligonier, PA (US)

Correspondence Address:
Ansel M. Schwartz
Attorney at Law
Suite 304
201 N. Craig Street
Pittsburgh, PA 15213 (US)

(73) Assignee: **Powercast, LLC**

(21) Appl. No.: **11/705,303**

(22) Filed: **Feb. 12, 2007**

Related U.S. Application Data

(60) Provisional application No. 60/773,036, filed on Feb. 13, 2006.

Publication Classification

(51) **Int. Cl.**
H04B 1/10 (2006.01)
H04B 1/38 (2006.01)

(52) **U.S. Cl.** **455/572**; 455/299

(57) **ABSTRACT**

Disclosed is a power transmission system for wirelessly powering a power harvesting device. The system comprises at least one RF power transmitter. The system includes an AC power grid, or a DC grid to which the transmitter is electrically connected. Also disclosed is an adjustable RF power transmitter for powering wirelessly an RF power harvesting- device. Also disclosed is a power transmission system for wirelessly powering an RF power harvesting device. The system can include a computer with an antenna or a lighting fixture or a light or a battery charging unit or a battery. Also disclosed is an apparatus for wirelessly powering a power harvesting device. Also disclosed is a method for wirelessly powering a power harvesting device.

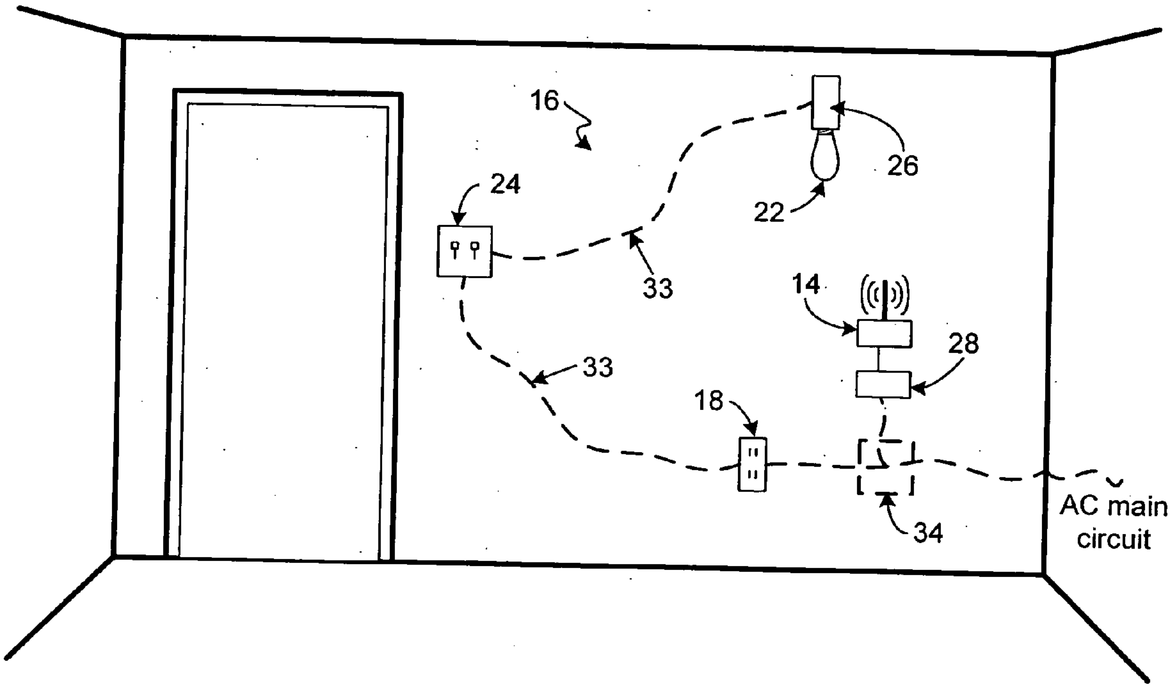

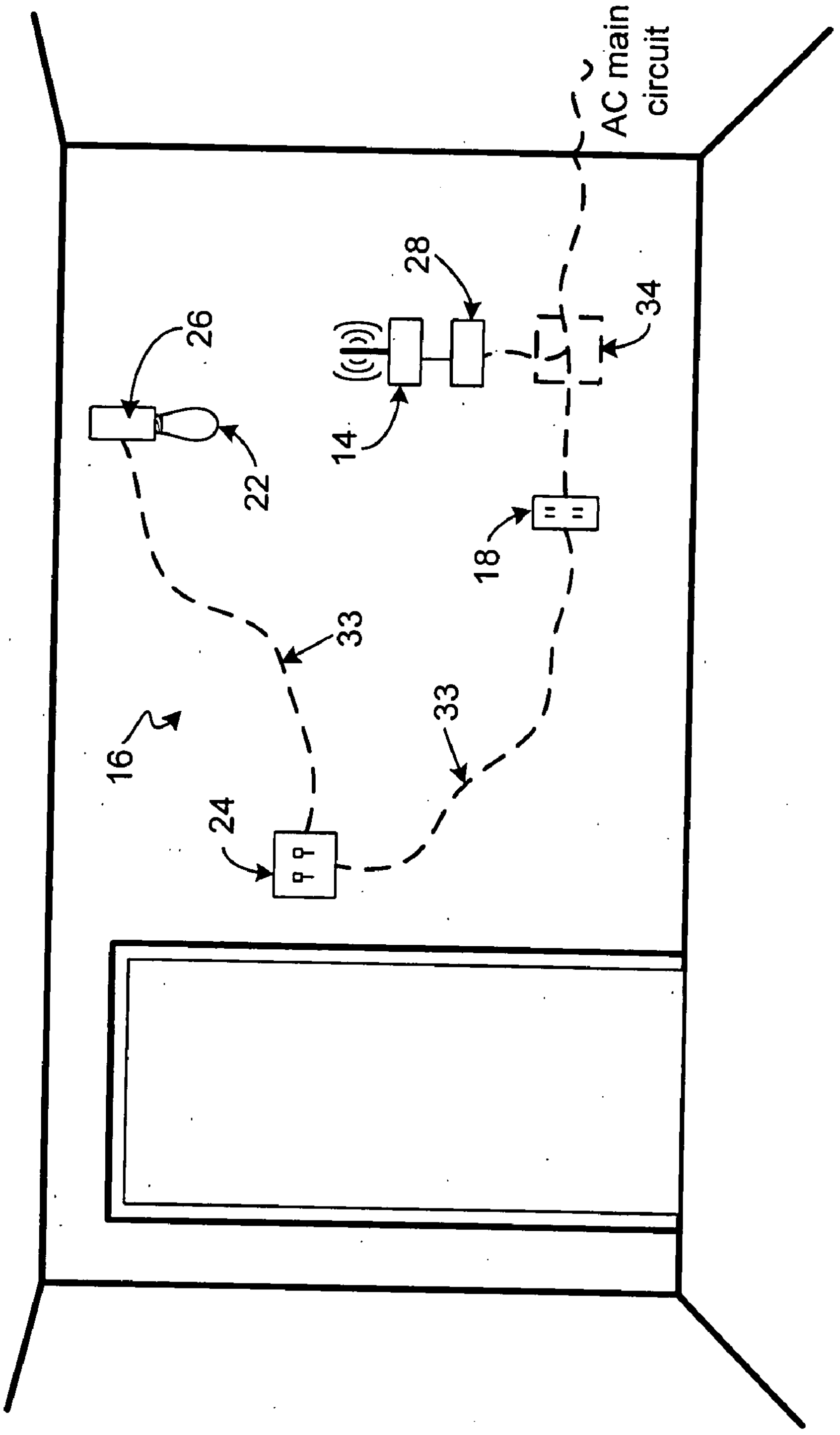
AC main circuit
28
34
26
14
22
18
33
33
16
24

Fig. 1

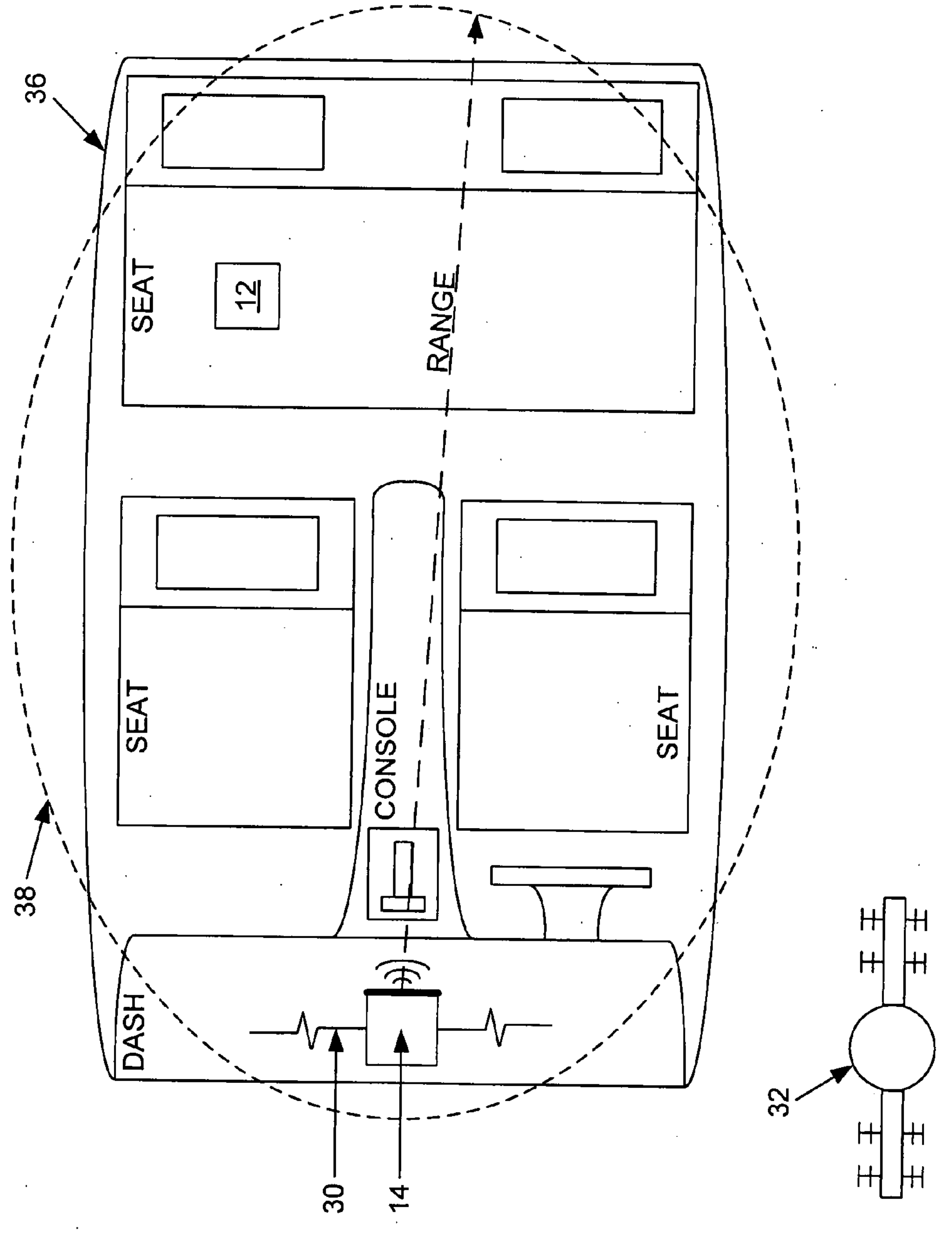
36
SEAT
12
RANGE
SEAT
CONSOLE
SEAT
38
DASH
30
14
32

Fig. 2

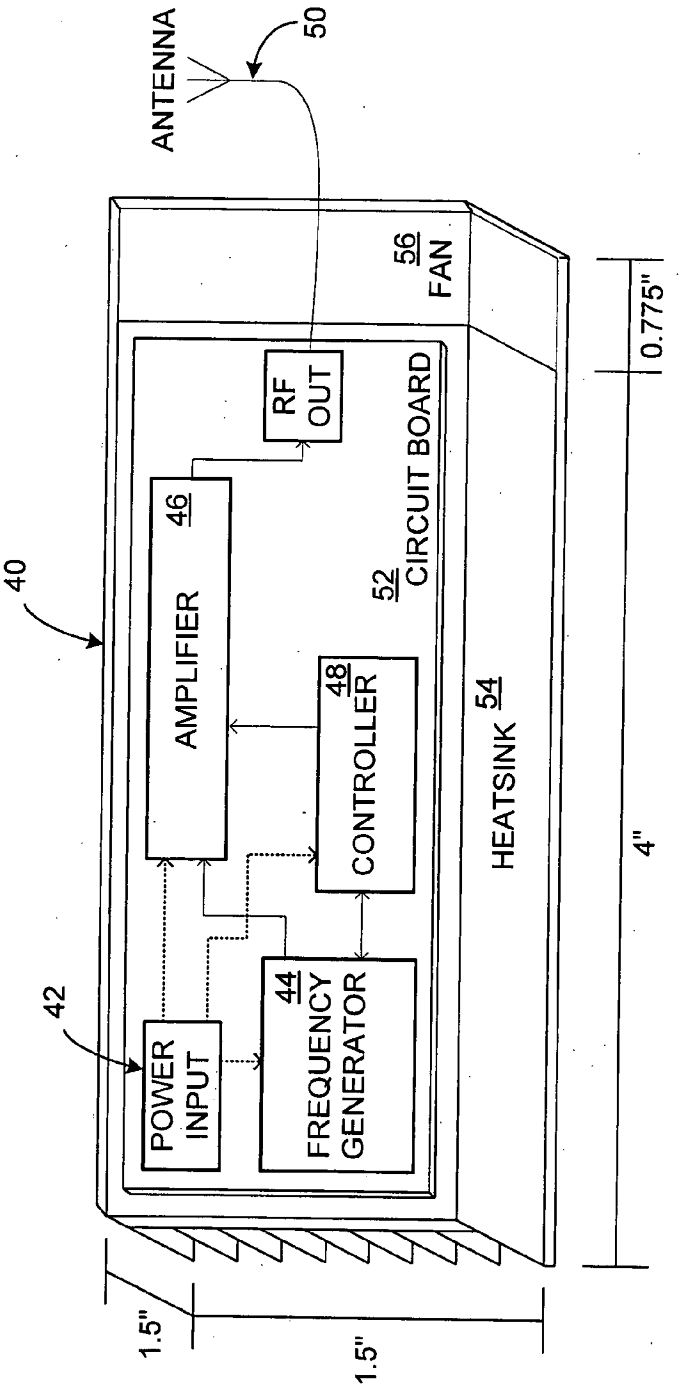
ANTENNA
50
40
FAN
56
RF OUT
0.775"
AMPLIFIER
46
52
CIRCUIT BOARD
48
CONTROLLER
HEATSINK
54
4"
POWER INPUT
42
44
FREQUENCY GENERATOR
1.5"
1.5"

Fig. 3

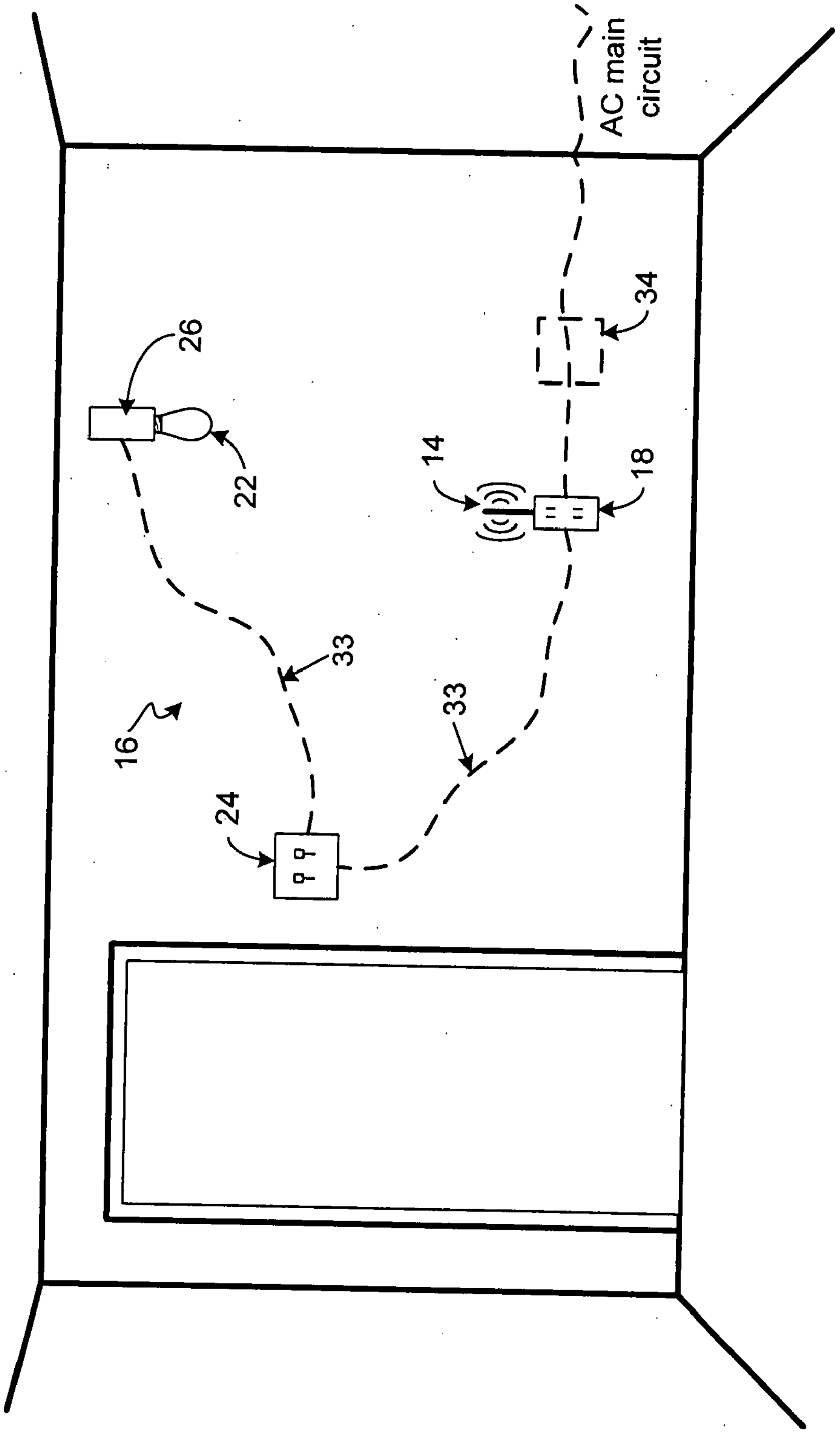
AC main circuit
34
26
22
14
18
33
33
16
24

Fig. 4

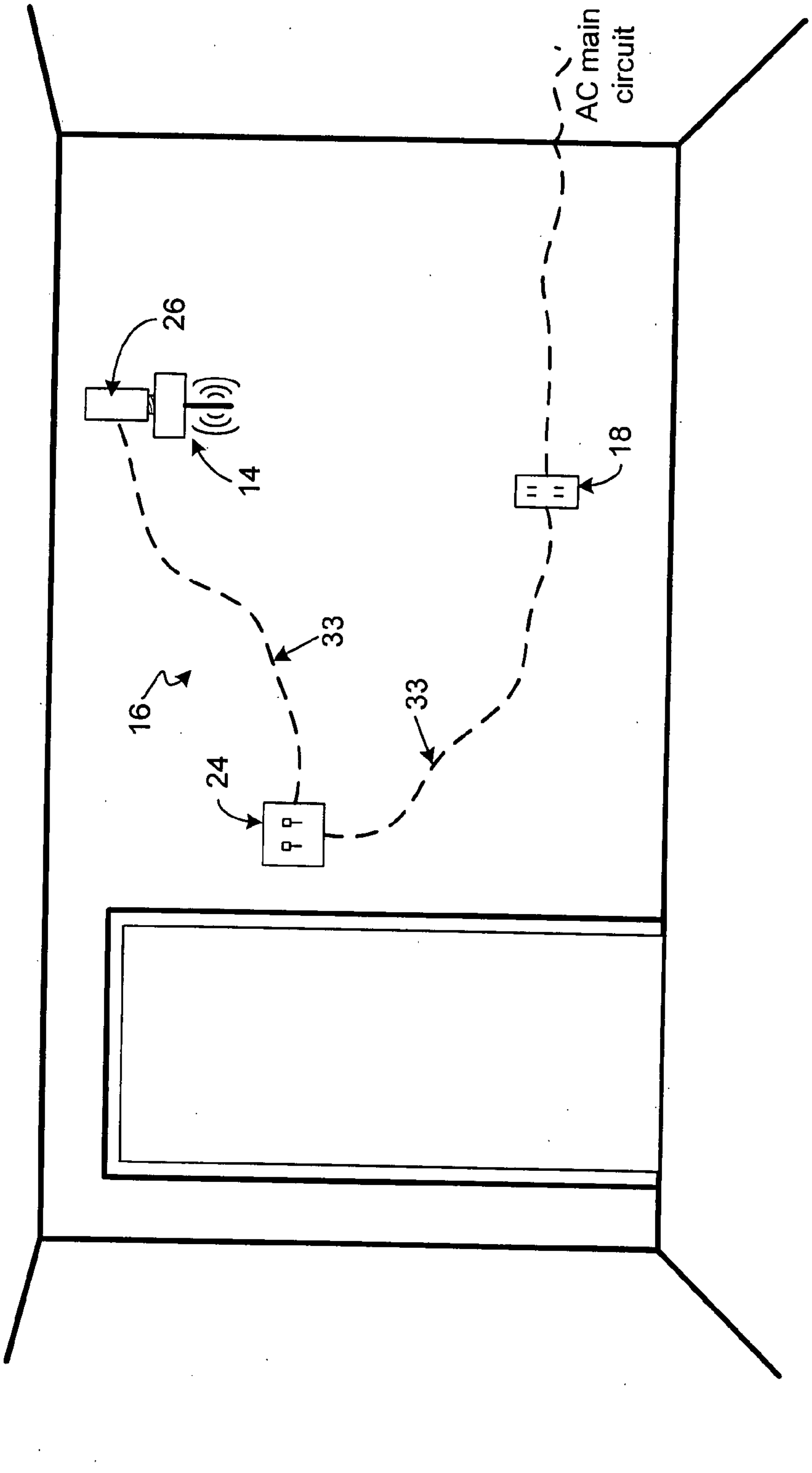
AC main circuit
26
14
18
33
33
16
24

Fig. 5

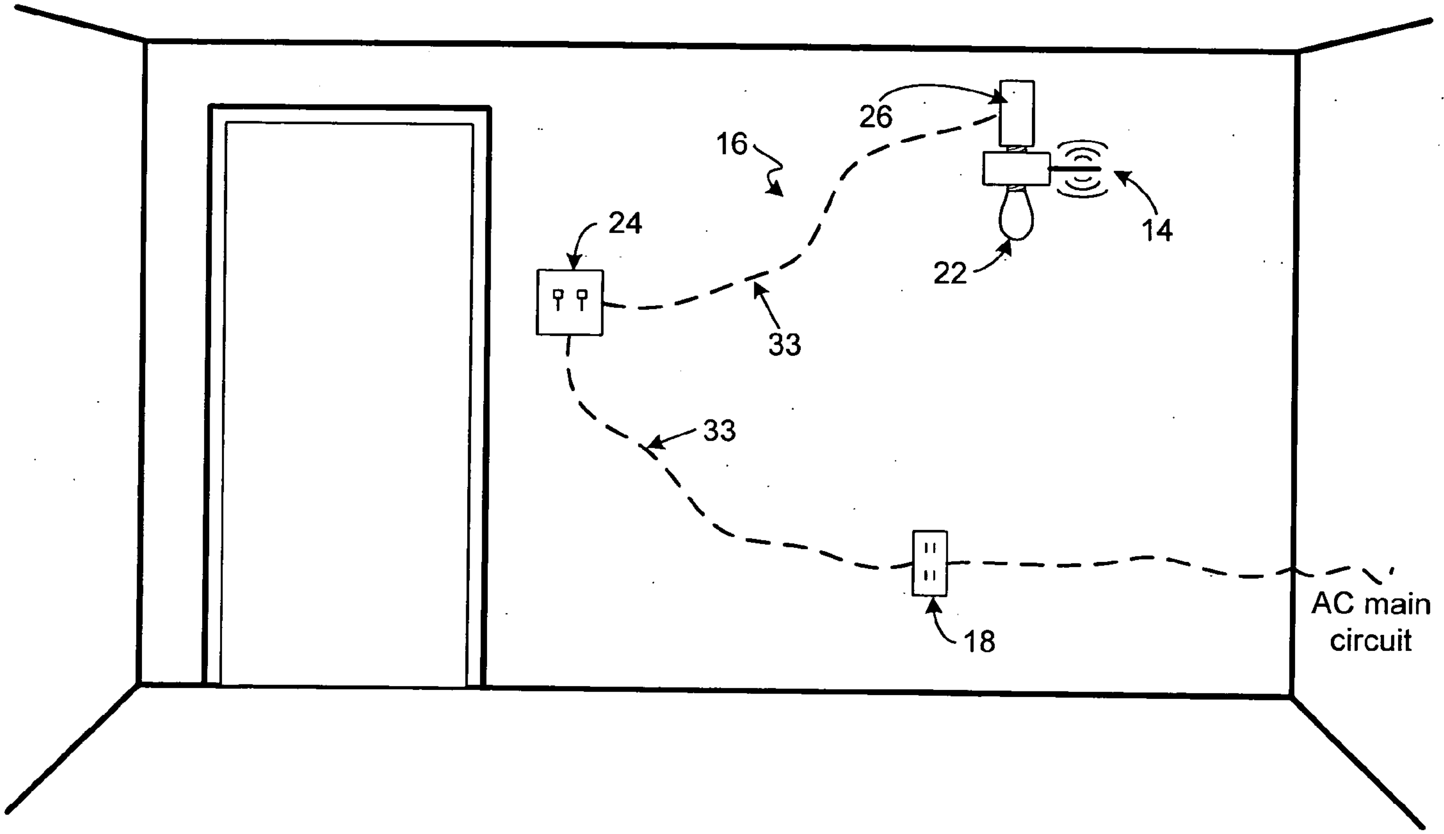
26
16
24
14
22
33
33
18
AC main circuit

Fig. 6

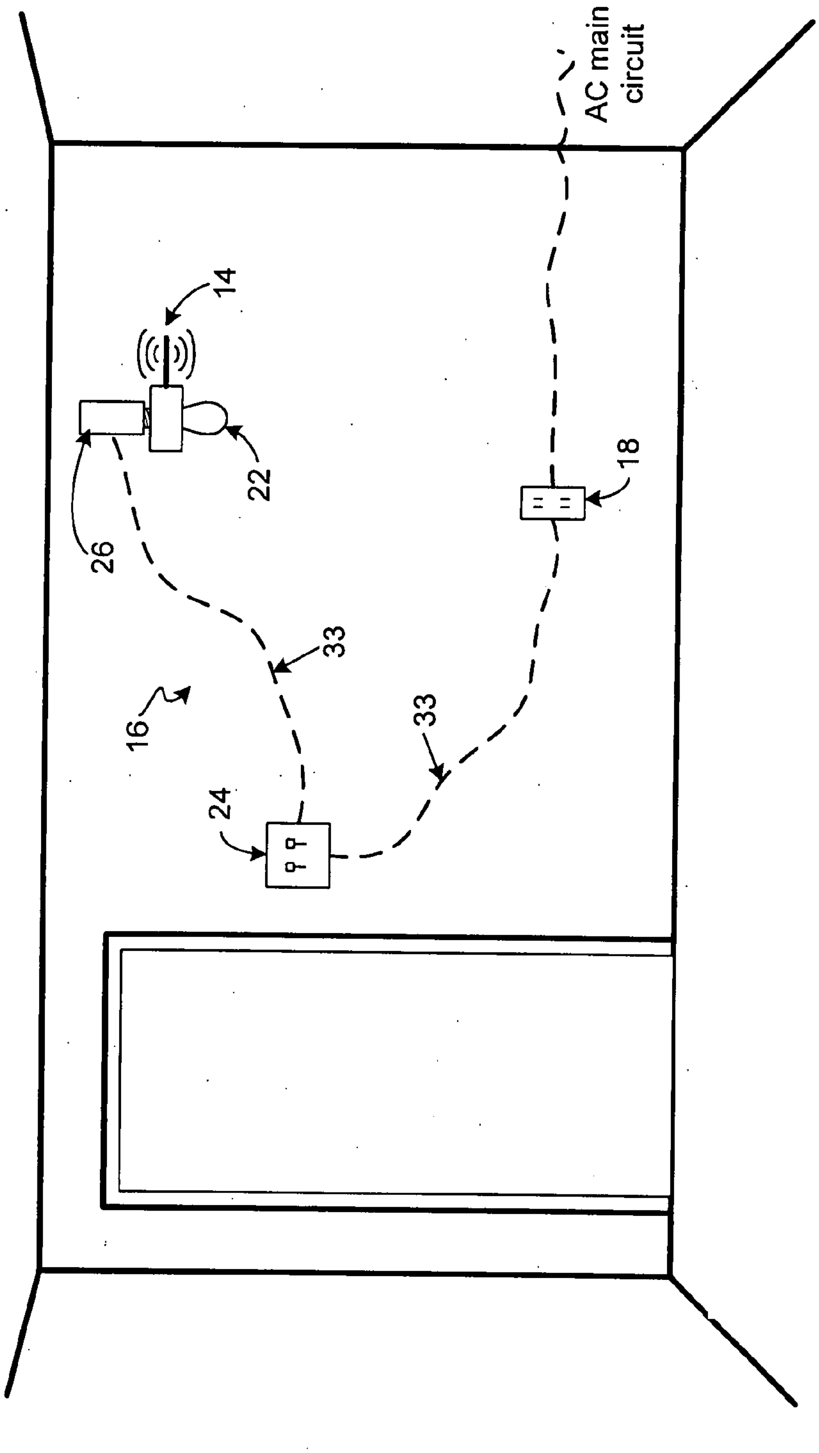
AC main circuit
14
22
26
18
33
33
16
24

Fig. 7

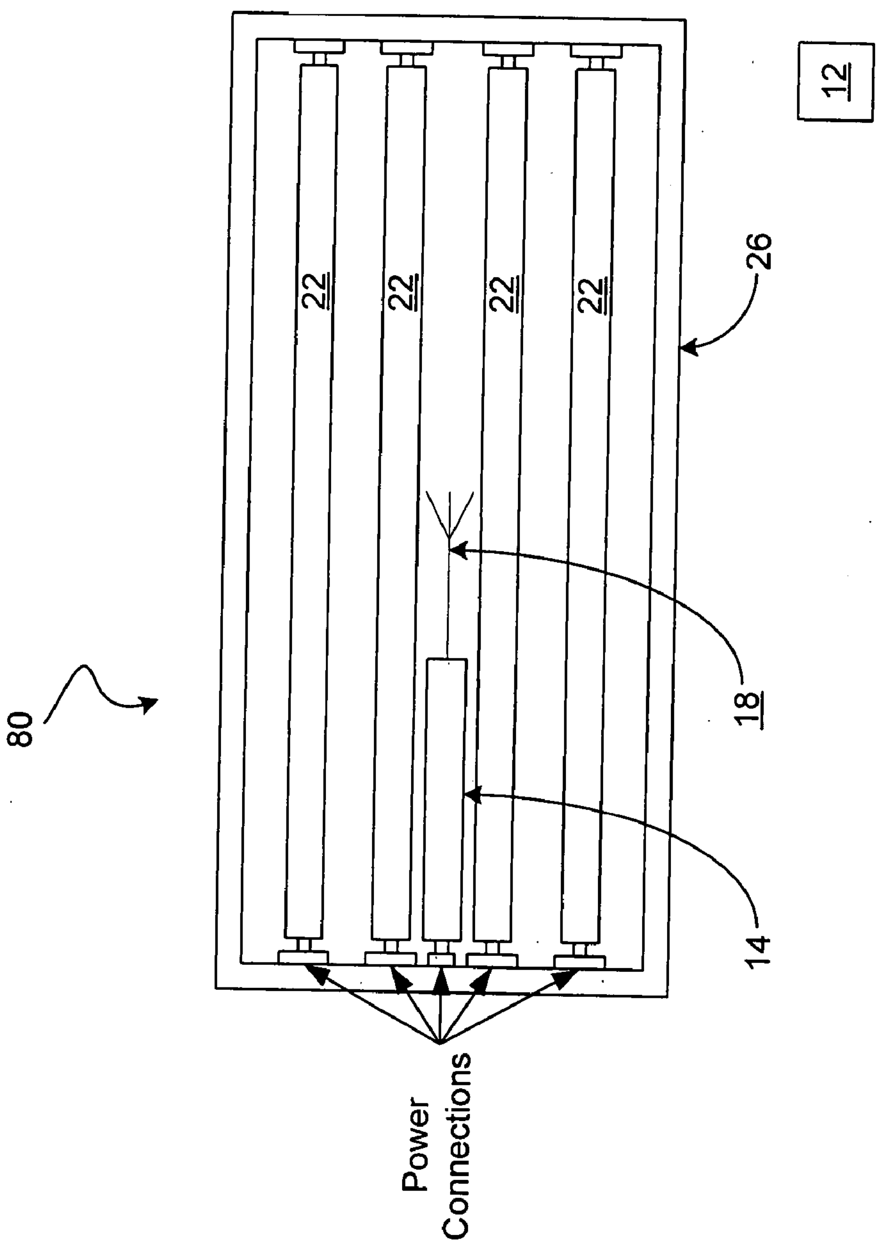
80
12
22
22
22
22
26
18
14
Power Connections

Fig. 8

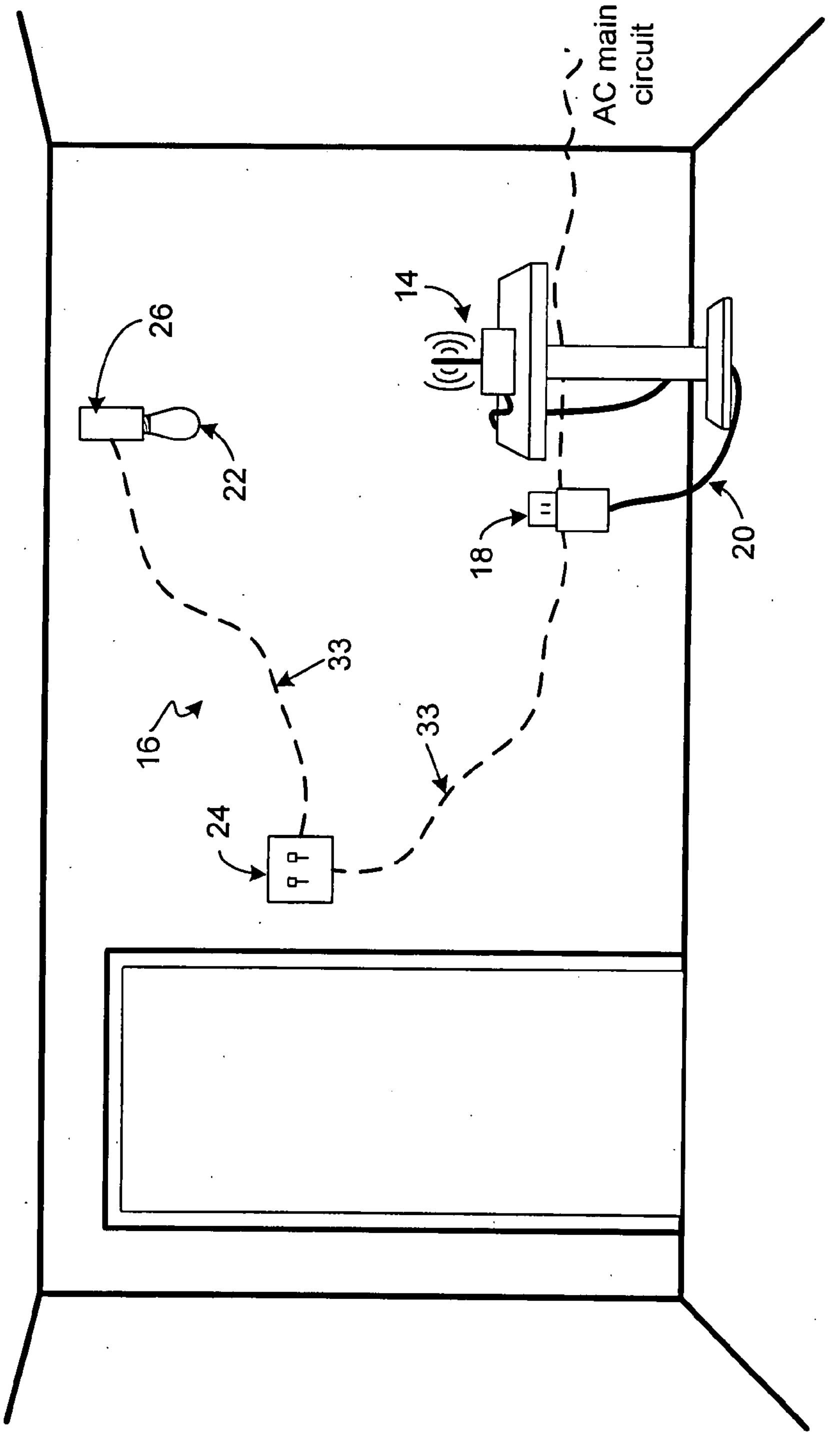
AC main circuit
14
26
22
18
20
33
33
16
24

Fig. 9

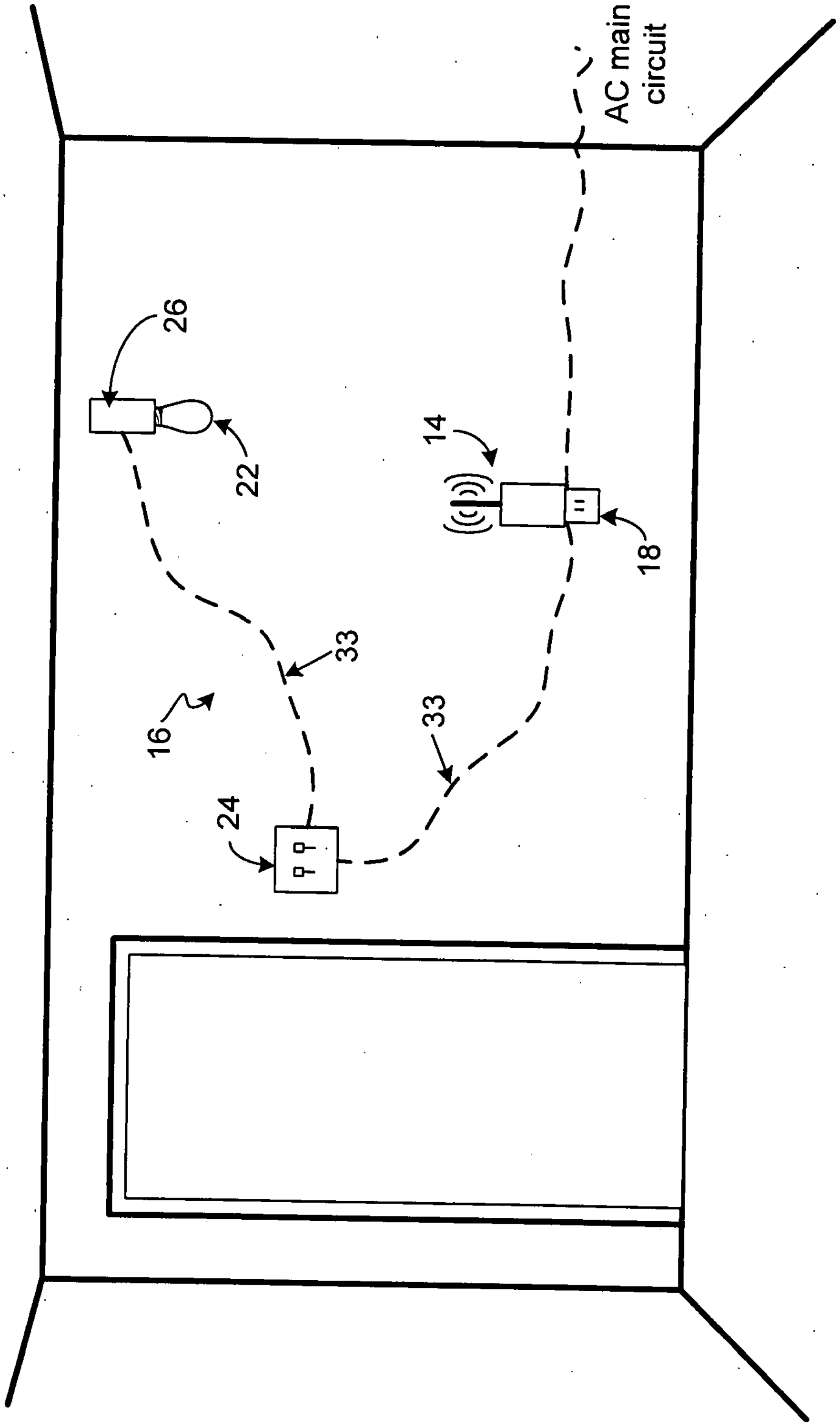
AC main circuit
26
22
14
18
33
33
16
24

Fig. 10

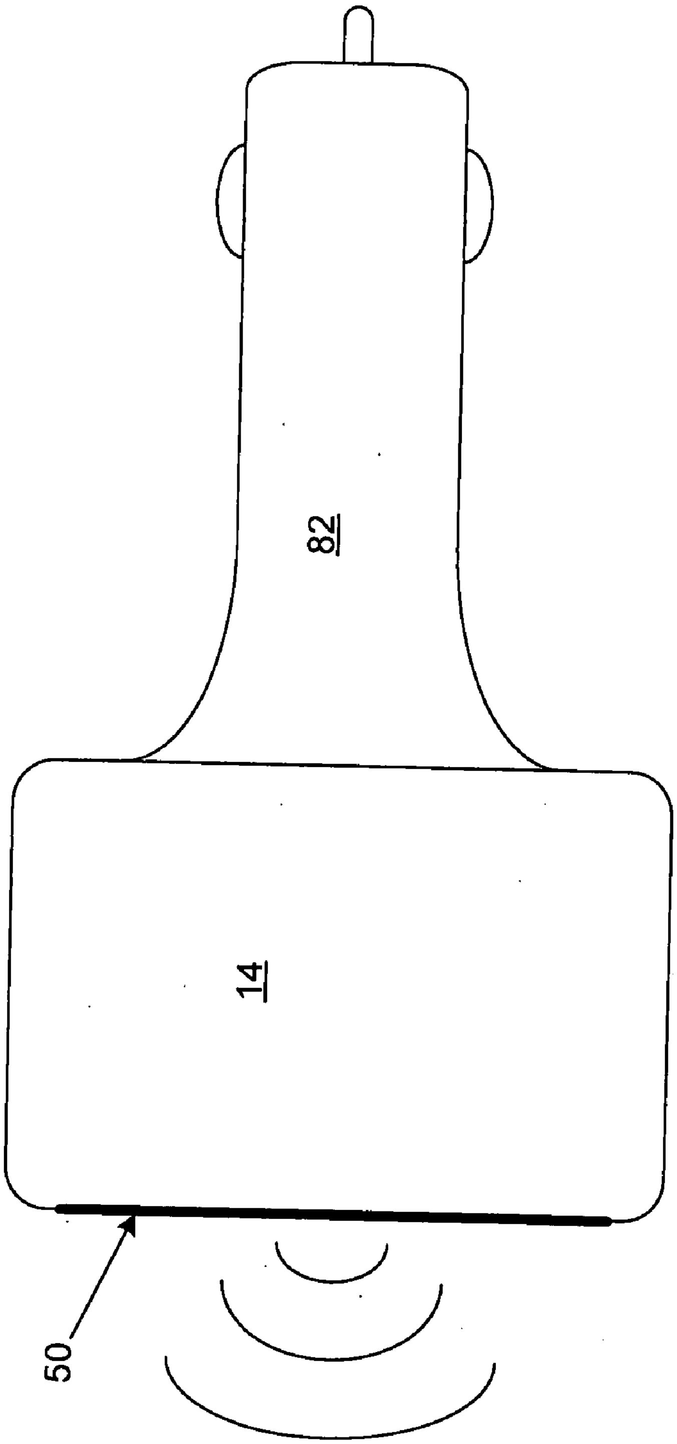

Fig. 11

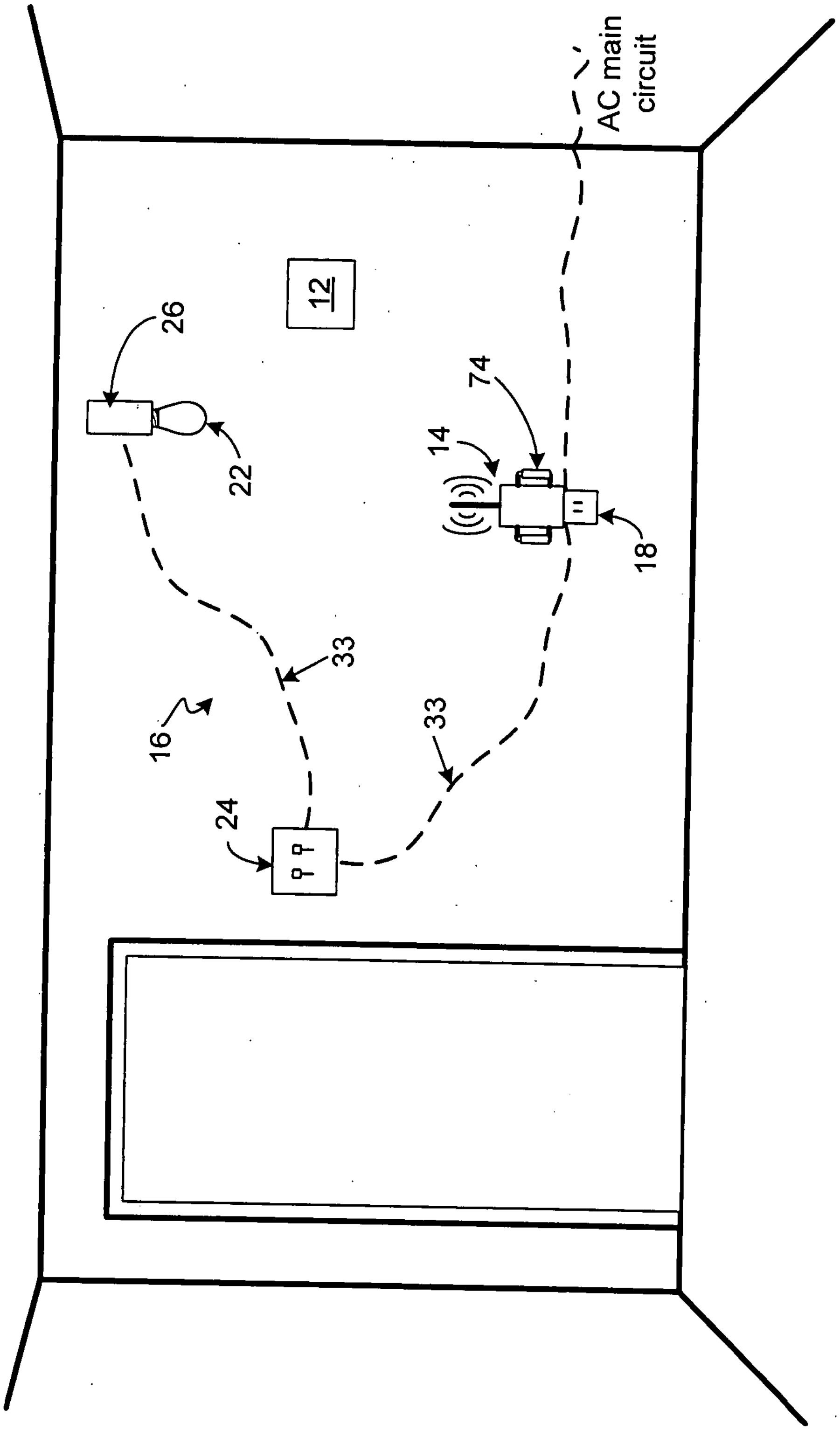
AC main circuit
12
26
22
74
14
18
33
33
16
24

Fig. 12

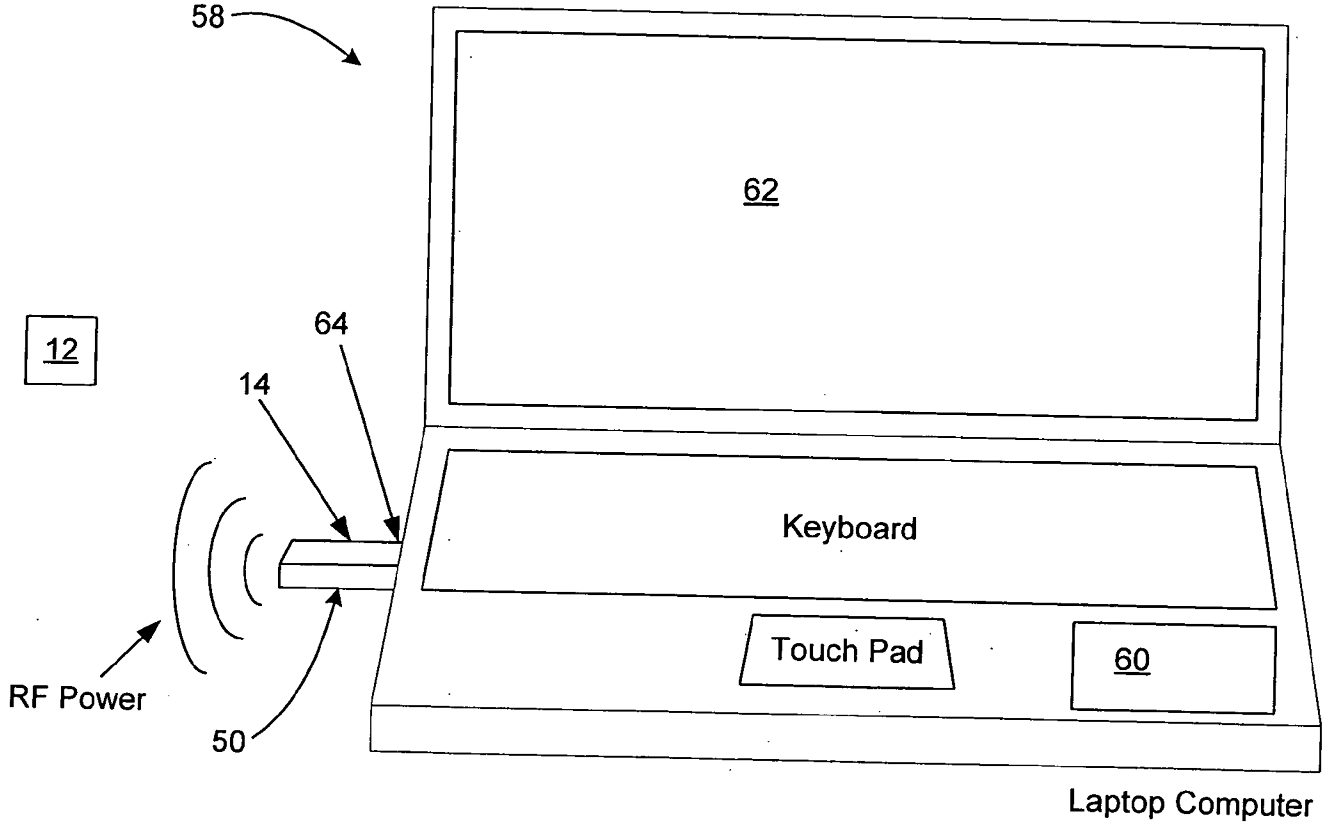
RF Power
Keyboard
Touch Pad
Laptop Computer

Fig. 13

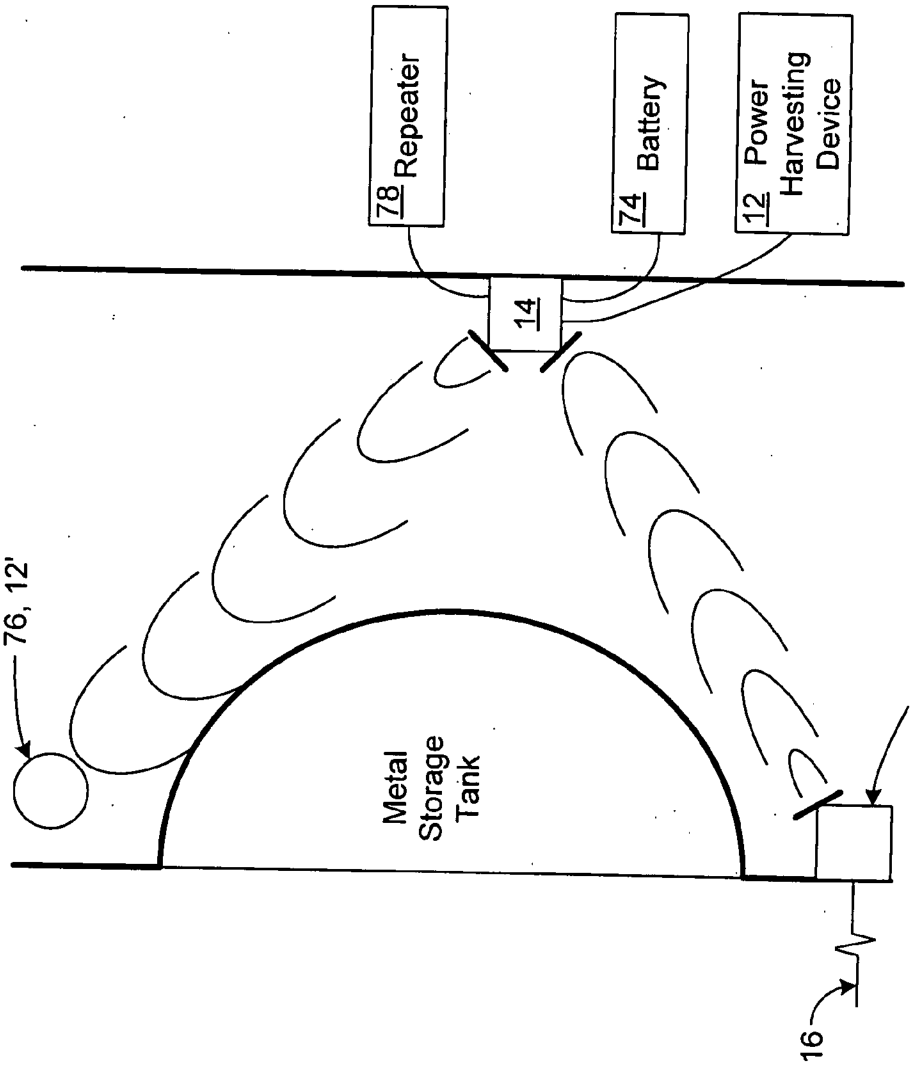
78 Repeater
74 Battery
12 Power Harvesting Device
14
76, 12'
Metal Storage Tank
14'
16

Fig. 14

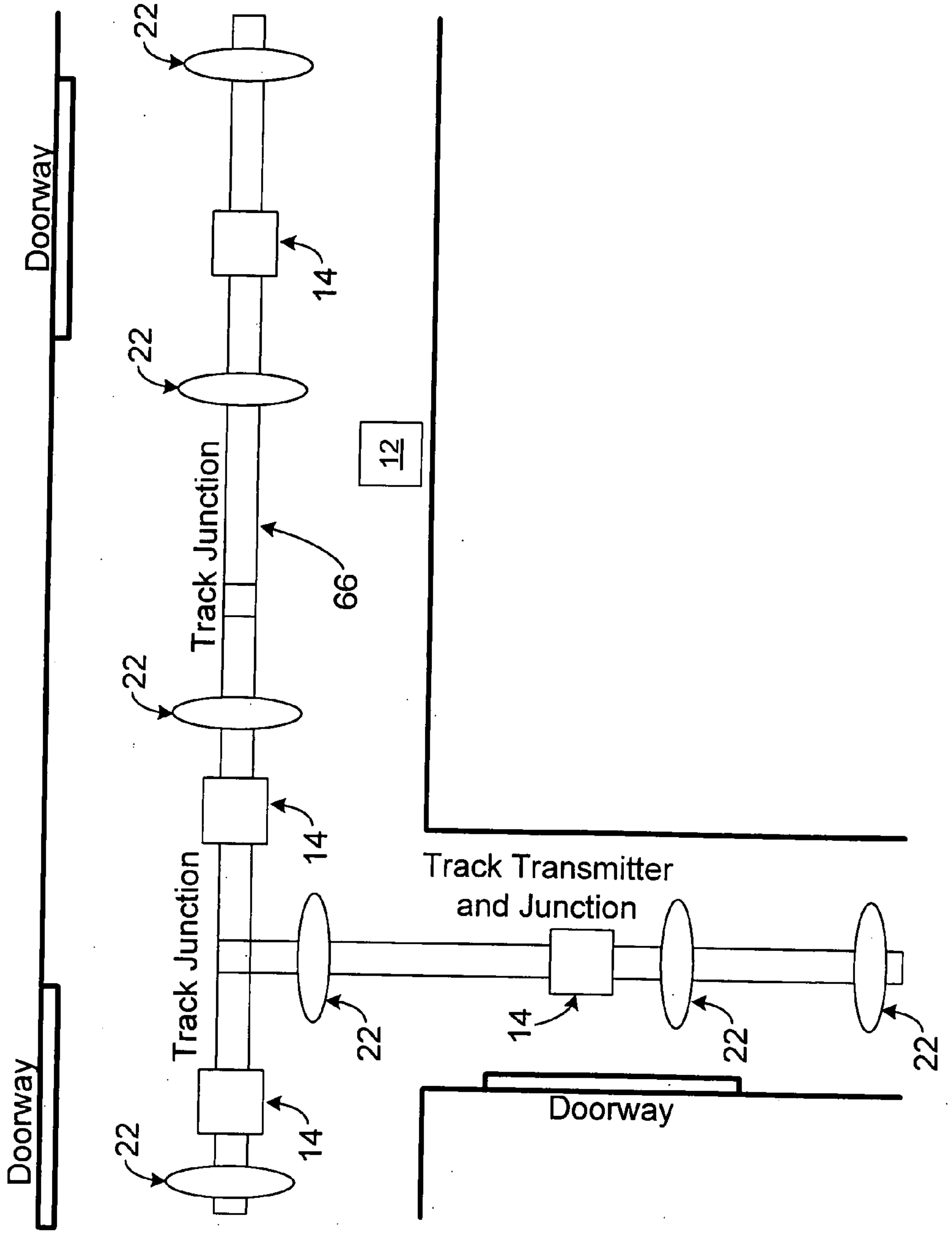
Doorway
Track Junction
Track Junction
Track Transmitter and Junction
Doorway
Doorway

Fig. 15

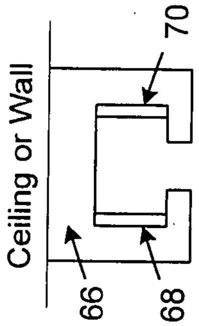
Ceiling or Wall
66
68
70

Fig. 16a

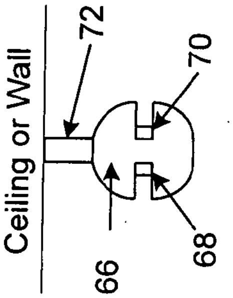
Ceiling or Wall
72
66
68
70

Fig. 16b

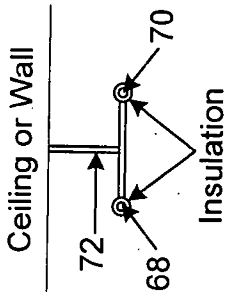
Ceiling or Wall
72
68
70
Insulation

Fig. 16c

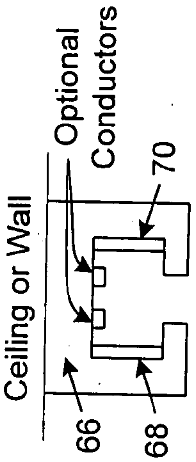
Ceiling or Wall
Optional Conductors
66
68
70

Fig. 16d

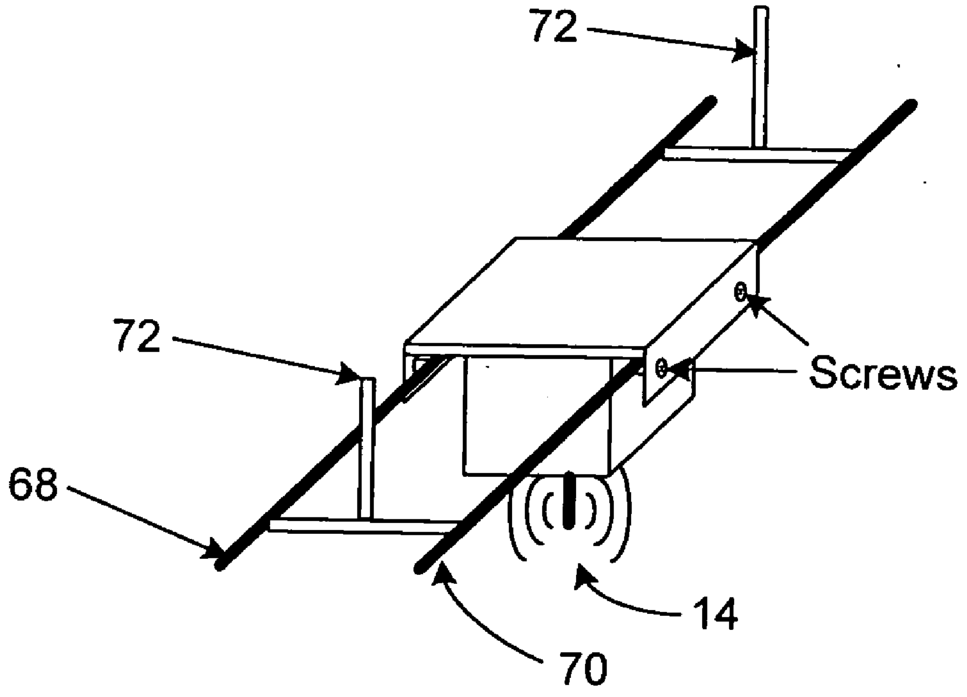
72
72
Screws
68
14
70

Fig. 17

IMPLEMENTATION OF AN RF POWER TRANSMITTER AND NETWORK

BACKGROUND OF THE INVENTION

[0001] 1. Field of the Invention

[0002] The present invention is related to transmitting wirelessly power to a power harvesting device. More specifically, the present invention is related to implementing a power transmitter for the wireless transmission of power to a power harvesting device.

[0003] 2. Description of Related Art

[0004] As processor capabilities have expanded and power requirements have decreased there has been an ongoing explosion of devices that operate completely independent of wires or power cords. These "untethered" devices range from cell phones and wireless keyboards to building sensors and active Radio Frequency Identification (RFID) tags.

[0005] Engineers and designers of these untethered devices continue to have to deal with the limitations of portable power sources, primarily using batteries as the key design parameter. While the performance of processors and portable devices has been doubling every 18-24 months driven by Moore's law, battery technology in terms of capacity has only been growing at measly 6% per year. Even with power conscious designs and the latest in battery technology, many devices do not meet the lifetime cost and maintenance requirements for applications that require a large number of untethered devices, such as logistics and building automation. Today's devices that need two-way communication require scheduled maintenance every three to 18 months to replace or recharge the device's power source (typically a battery). One-way devices that simply broadcast their status without receiving any signals, such as automated utility meter readers, have a better battery life typically requiring replacement within 10 years. For both device types, scheduled power-source maintenance is costly and can be disruptive to the entire system that the device is intended to monitor and/or control. Unscheduled maintenance trips are even more costly and disruptive. On a macro level, the relatively high cost associated with the internal battery also reduces the practical, or economically viable, number of devices that can be deployed.

[0006] The ideal solution to the power problem for untethered devices is a device or system that can collect and harness sufficient energy from the environment. This energy can be harnessed from many different sources, such as sunlight, vibration, heat, or electro-magnetic radiation. The harnessed energy would then either directly power an untethered device or augment a power supply. However, this ideal solution may not always be practical to implement due to low energy in the environment, and site restrictions may limit the ability to use a dedicated energy supply. The proposed invention takes these factors into account and provides a solution for both the ideal situation and also for more restrictive circumstances.

BRIEF SUMMARY OF THE INVENTION

[0007] The present invention pertains to a power transmission system for wirelessly powering a power harvesting device. The system comprises at least one RF power transmitter. The system comprises an AC power grid to which the transmitter is electrically connected.

[0008] The power grid can have an outlet. The transmitter can have a cord which plugs into the outlet. The transmitter can plug directly into the outlet. The power grid can have a light. The transmitter can include an AC to DC converter that can convert the AC power obtained from the grid to a usable DC voltage or current.

[0009] The present invention pertains to a power transmission system for wirelessly powering a power harvesting device. The system comprises at least one RF power transmitter. The system comprises a DC power grid to which the transmitter is electrically connected.

[0010] The present invention pertains to an adjustable RF power transmitter for powering wirelessly an RF power harvesting device. The transmitter comprises a housing having outer dimensions no greater than 3"×3" by 8 inches. The transmitter comprises a power input. The transmitter comprises a frequency generator in communication with the power input. The transmitter comprises an amplifier in communication with the frequency generator. The transmitter comprises a controller connected with the frequency generator. The transmitter comprises an antenna connected to the amplifier.

[0011] The present invention pertains to a power transmission system for wirelessly powering an RF power harvesting device. The system comprises a computer with an antenna. The system comprises an RF transmitter in communication with the antenna. The system comprises a power supply in electrical communication with the RF transmitter and the computer.

[0012] The present invention pertains to an apparatus for wirelessly powering a power harvesting device. The apparatus comprises at least one RF power transmitter. The apparatus comprises a lighting fixture in which the transmitter is disposed and from which the transmitter receives power.

[0013] The lighting fixture can be a fluorescent lighting fixture. The lighting fixture can be an incandescent lighting fixture. The apparatus can include a light source in electrical communication with the lighting fixture.

[0014] The present invention pertains to a power transmission system for wirelessly powering a power harvesting device. The system comprises at least one RF power transmitter. The system comprises a track supplying power to which the transmitter is electrically connected.

[0015] The present invention pertains to a power transmission system for wirelessly powering a power harvesting device. The system comprises at least one RF power transmitter. The system comprises a battery charging unit to which the transmitter is electrically connected.

[0016] The present invention pertains to a power transmission system for wirelessly powering a power harvesting device. The system comprises at least one RF power transmitter. The system comprises at least one rechargeable battery to which the transmitter is electrically connected.

[0017] The present invention pertains to a method for wirelessly powering a power harvesting device. The method comprises the steps of electrically connecting at least one RF power transmitter to an AC power grid. There is the step of transmitting power with the RF power transmitter.

[0018] The present invention pertains to a method for wirelessly powering a power harvesting device. The method comprises the steps of electrically connecting at least one RF power transmitter to a DC power grid. There is the step of transmitting power with the RF power transmitter.

[0019] The present invention pertains to a method for wirelessly powering a power harvesting device. The method comprises the steps of electrically connecting a power supply with an RF transmitter and a computer. There is the step of transmitting power with the RF power transmitter.

[0020] The present invention pertains to a method for wirelessly powering a power harvesting device. The method comprises the steps of electrically connecting at least one RF power transmitter with a lighting fixture in which the transmitter is disposed and from which the transmitter receives power. There is the step of transmitting power with the RF power transmitter.

[0021] The present invention pertains to a method for wirelessly powering a power harvesting device. The method comprises the steps of electrically connecting at least one RF power transmitter to a battery charging unit. There is the step of transmitting power with the RF power transmitter.

[0022] The present invention pertains to a power transmission system for wirelessly powering a power harvesting device. The system comprises at least one RF power transmitter. The system comprises means for providing power to which the transmitter is electrically connected.

[0023] The present invention pertains to an apparatus for wirelessly powering a power harvesting device from a DC power outlet of a vehicle, as shown in FIG. **11**. The apparatus comprises an RF power transmitter. The apparatus comprises a power plug to which the transmitter is attached and electrically connected that plugs into the DC power outlet.

[0024] The present invention pertains to an apparatus for wirelessly powering a power harvesting device from an AC power grid having an AC power outlet. The apparatus comprises an RF power transmitter. The apparatus comprises a power plug to which the transmitter is electrically connected that electrically connects with the AC power outlet.

[0025] The present invention pertains to an apparatus for wirelessly powering a power harvesting device from a DC power outlet of a DC grid. The apparatus comprises an RF power transmitter. The apparatus comprises a power plug to which the transmitter is electrically connected that electrically connects with the DC power outlet.

[0026] The present invention pertains to an apparatus for wirelessly powering a power harvesting device from a computer having an antenna and a power supply. The apparatus comprises an RF power transmitter. The apparatus comprises a power plug to which the transmitter is electrically connected that electrically connects with the computer.

[0027] The present invention pertains to an apparatus for wirelessly powering a power harvesting device from a light fixture. The apparatus comprises an RF power transmitter. The apparatus comprises an electrical interface to which the transmitter is electrically connected that electrically connects with the light fixture.

[0028] The present invention pertains to an apparatus for wirelessly powering a power harvesting device from a track having at least one light. The apparatus comprises an RF power transmitter. The apparatus comprises an electrical interface to which the transmitter is electrically connected that electrically connects with the track.

[0029] The present invention pertains to an apparatus for wirelessly powering a power harvesting device from a battery charging unit. The apparatus comprises an RF power transmitter. The apparatus comprises an electrical interface to which the transmitter is electrically connected that electrically connects with the battery charging unit.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[0030] FIG. **1** is an illustration of an RF power transmitter integrated into an RF power network via direct hardwiring to an AC power grid;

[0031] FIG. **2** is an illustration of an RF power transmitter integrated into a vehicle;

[0032] FIG. **3** is an illustration of an RF power transmitter;

[0033] FIG. **4** is an illustration of an RF power transmitter integrated into an RF power network via replacing an AC outlet on an AC power grid;

[0034] FIG. **5** is an illustration of an RF power transmitter integrated into an RF power network via replacing a light source;

[0035] FIG. **6** is an illustration of an RF power transmitter integrated into an RF power network via use in combination with a light source;

[0036] FIG. **7** is an illustration of an RF power transmitter integrated into an RF power network via integration of the RF power transmitter and a light source;

[0037] FIG. **8** is an illustration of a lighting fixture containing an RF power transmitter;

[0038] FIG. **9** is an illustration of an RF power transmitter integrated into an RF power network via connection to an AC outlet through a cord;

[0039] FIG. **10** is an illustration of an RF power transmitter integrated into an RF power network via direct connection with an AC outlet;

[0040] FIG. **11** is an illustration of an RF power transmitter that plugs into a DC power outlet in a vehicle;

[0041] FIG. **12** is an illustration of an RF power transmitter and battery recharger integrated into an RF power network via direct connection with an AC outlet;

[0042] FIG. **13** is an illustration of an RF power transmitter plugged into a USB port of a laptop computer;

[0043] FIG. **14** is an illustration of an RF powered RF power transmitter used to relay power;

[0044] FIG. **15** is an illustration of RF power transmitters integrated into an RF power network connected to a track that provides AC or DC power;

[0045] FIGS. **16**_a-d_ are illustrations of various tracks useable with the network illustrated in FIG. **15**; and

[0046] FIG. **17** is an illustration of an RF power transmitter useable with a two-cable track.

DETAILED DESCRIPTION OF THE INVENTION

[0047] A complete understanding of the invention will be obtained from the following description when taken in connection with the accompanying drawing figures wherein like reference characters identify like parts throughout.

[0048] For purposes of the description hereinafter, the terms "upper", "lower", "right", "left", "vertical", "horizontal", "top", "bottom", and derivatives thereof shall relate to the invention as it is oriented in the drawing figures. However, it is to be understood that the invention may assume various alternative variations and step sequences, except where expressly specified to the contrary. It is also to be understood that the specific devices and processes illustrated in the attached drawings, and described in the following specification, are simply exemplary embodiments of the invention. Hence, specific dimensions and other physical characteristics related to the embodiments disclosed herein are not to be considered as limiting.

[0049] The present invention pertains to a power transmission system **10** for wirelessly powering a power harvesting device **12**, as shown in FIG. **1**. The system **10** comprises at least one RF power transmitter **14**. The system **10** comprises an AC power grid **16** to which the transmitter **14** is electrically connected.

[0050] The power grid can have an outlet **18**. The transmitter **14** can have a cord **20** which plugs into the outlet **18**, as shown in FIG. **9**. The transmitter **14** can plug directly into the outlet **18**, as shown in FIG. **10**. The power grid can have a light **22**, as shown in FIG. **1**. The transmitter **14** can include an AC to DC converter **28** that can convert the AC power obtained from the grid to a usable DC voltage or current. Preferably, the power grid has a light switch **24** to turn on the light **22**. The grid preferably includes in wall wiring **33**. The transmitter **14** can be integrated with the outlet **18**, as shown in FIG. **4**. The grid can have a lighting fixture **26** and the transmitter **14** contacts the fixture **26**, as shown in FIG. **5**. The grid can have a lighting fixture **26** and the light **22** and the transmitter **14** can contact the fixture, as shown in FIGS. **6** and **7**.

[0051] The grid can include a utility pole **32** to which the transmitter **14** is in contact, as shown in FIG. **2**. The grid can include a junction box **34** to which transmitter **14** is in contact, as shown in FIG. **1**.

[0052] The present invention pertains to a power transmission system **10** for wirelessly powering a power harvesting device **12**, as shown in FIG. **2**. The system **10** comprises at least one RF power transmitter **14**. The system **10** comprises a DC power grid **30** to which the transmitter **14** is electrically connected.

[0053] Preferably, the transmitter **14** is disposed in a vehicle **36**. The transmitter **14** can provide a coverage area **38** over the cabin of the vehicle **36**. The transmitter **14** can be in contact with the dash, trunk, cabin, ceiling, or engine compartment of the vehicle **36**.

[0054] The present invention pertains to an adjustable RF power transmitter **14** for powering wirelessly an RF power harvesting device **12**, as shown in FIG. **3**. The transmitter **14** comprises a housing **40** having outer dimensions no greater than 3"×3" by 8 inches. The transmitter **14** comprises a power input **42**. The transmitter **14** comprises a frequency generator **44** in communication with the power input **42**. The transmitter **14** comprises an amplifier **46** in communication with the frequency generator **44**. The transmitter **14** comprises a controller **48** connected with the frequency generator **44**. The transmitter **14** comprises an antenna **50** connected to the amplifier **46**.

[0055] The transmitter **14** can include a circuit board **52** on which the power input **42**, the frequency generator **44** and the amplifier **46** are disposed. The transmitter **14** can include a heat sink **54** in contact with the circuit board **52**. The transmitter **14** can include a fan **56** disposed adjacent the circuit board **52**.

[0056] The present invention pertains to a power transmission system **10** for wirelessly powering an RF power harvesting device **12**, as shown in FIG. **13**. The system **10** comprises a computer **58** with an antenna **50**. The system **10** comprises an RF transmitter **14** in communication with the antenna **50**. The system **10** comprises a power supply **60** in electrical communication with the RF transmitter **14** and the computer **58**.

[0057] Preferably, the RF transmitter **14** is connected to the computer **58**. The computer **58** can have a power port **64**, and the transmitter **14** is plugged into the power port **64**. The power port **64** can be a USB port **64**. The antenna **50** can be integrated with the transmitter **14**. Alternatively, the system **10** can include a display **62**, and the antenna **50** is in contact with the display **62**.

[0058] The present invention pertains to an apparatus **80** for wirelessly powering a power harvesting device **12**, as shown in FIG. **8**. The apparatus **80** comprises at least one RF power transmitter **14**. The apparatus **80** comprises a lighting fixture **26** in which the transmitter **14** is disposed and from which the transmitter **14** receives power.

[0059] The lighting fixture **26** can be a fluorescent lighting fixture **26**. The lighting fixture **26** can be an incandescent lighting fixture **26**. The lighting fixture **26** can be an LED lighting fixture **26**. The apparatus **80** can include a light source in electrical communication with the lighting fixture **26**.

[0060] The present invention pertains to a power transmission system **10** for wirelessly powering a power harvesting device **12**, as shown in FIG. **15**. The system **10** comprises at least one RF power transmitter **14**. The system **10** comprises a track **66** supplying power to which the transmitter **14** is electrically connected.

[0061] Preferably, there are at least two power transmitters **14**. The system **10** can include at least two lights **22** electrically connected to the track **66**. The track **66** can include a first conductor **68** and a second conductor **70**, as shown in FIGS. **16**_a_-**16**_d_. The track **66** may include a support **72** attached to a wall or ceiling, as shown in FIG. **17**.

[0062] The present invention pertains to a power transmission system **10** for wirelessly powering a power harvesting device **12**, as shown in FIG. **12**. The system **10** comprises at least one RF power transmitter **14**. The system **10** comprises a battery **74** charging unit to which the transmitter **14** is electrically connected.

[**0063**] The present invention pertains to a power transmission system **10** for wirelessly powering a power harvesting device **12**, as shown in FIG. **14**. The system **10** comprises at least one RF power transmitter **14**. The system **10** comprises at least one rechargeable battery **74** to which the transmitter **14** is electrically connected.

[**0064**] The system **10** can include a second RF power transmitter **14'** which transmits power to the power harvesting device **12** electronically connected to the battery **74**. A valve sensor **76** can include the valve sensor **76'** powered by a power harvesting device **12'**. The system **10** can include an RF power repeater **78**.

[**0065**] The present invention pertains to a method for wirelessly powering a power harvesting device **12**. The method comprises the steps of electrically connecting at least one RF power transmitter **14** to an AC power grid **16**. There is the step of transmitting power with the RF power transmitter **14**.

[**0066**] The present invention pertains to a method for wirelessly powering a power harvesting device **12**. The method comprises the steps of electrically connecting at least one RF power transmitter **14** to a DC power grid **30**. There is the step of transmitting power with the RF power transmitter **14**.

[**0067**] The present invention pertains to a method for wirelessly powering a power harvesting device **12**. The method comprises the steps of electrically connecting a power supply **60** with an RF transmitter **14** and a computer **58**. There is the step of transmitting power with the RF power transmitter **14**.

[**0068**] The present invention pertains to a method for wirelessly powering a power harvesting device **12**. The method comprises the steps of electrically connecting at least one RF power transmitter **14** with a lighting fixture **26** in which the transmitter **14** is disposed and from which the transmitter **14** receives power. There is the step of transmitting power with the RF power transmitter **14**.

[**0069**] The present invention pertains to a method for wirelessly powering a power harvesting device **12**. The method comprises the steps of electrically connecting at least one RF power transmitter **14** to a battery **74** charging unit. There is the step of transmitting power with the RF power transmitter **14**.

[**0070**] The present invention pertains to a power transmission system **10** for wirelessly powering a power harvesting device **12**. The system **10** comprises at least one RF power transmitter **14**. The system **10** comprises means for providing power to which the transmitter **14** is electrically connected.

[**0071**] The means can be an AC power grid **16**, a DC power grid **30**, a battery **74** or any other power sources identified herein.

[**0072**] The present invention pertains to an apparatus **80** for wirelessly powering a power harvesting device **12** from a DC power outlet **18** of a vehicle **36**, as shown in FIG. **11**. The apparatus **80** comprises an RF power transmitter **14**. The apparatus **80** comprises a power plug **82** to which the transmitter **14** is attached and electrically connected that plugs into the DC power outlet **18**.

[**0073**] The present invention pertains to an apparatus **80** for wirelessly powering a power harvesting device **12** from an AC power grid **16** having an AC power outlet **18**, as shown in FIG. **1**. The apparatus **80** comprises an RF power transmitter **14**. The apparatus **80** comprises a power plug to which the transmitter **14** is electrically connected that electrically connects with the AC power outlet **18**.

[**0074**] The present invention pertains to an apparatus **80** for wirelessly powering a power harvesting device **12** from a DC power outlet **18** of a DC grid **30**, as shown in FIG. **2**. The apparatus **80** comprises an RF power transmitter **14**. The apparatus **80** comprises a power plug to which the transmitter **14** is electrically connected that electrically connects with the DC power outlet **18**.

[**0075**] The present invention pertains to an apparatus **80** for wirelessly powering a power harvesting device **12** from a computer **58** having an antenna **50** and a power supply, as shown in FIG. **13**. The apparatus **80** comprises an RF power transmitter **14**. The apparatus **80** comprises a power plug to which the transmitter **14** is electrically connected that electrically connects with the computer **58**.

[**0076**] The present invention pertains to an apparatus **80** for wirelessly powering a power harvesting device **12** from a light fixture **26**, as shown in FIG. **5**. The apparatus **80** comprises an RF power transmitter **14**. The apparatus **80** comprises an electrical interface to which the transmitter **14** is electrically connected that electrically connects with the light fixture **26**.

[**0077**] The present invention pertains to an apparatus **80** for wirelessly powering a power harvesting device **12** from a track **66** having at least one light **22**, as shown in FIG. **15**. The apparatus **80** comprises an RF power transmitter **14**. The apparatus **80** comprises an electrical interface to which the transmitter **14** is electrically connected that electrically connects with the track **66**.

[**0078**] The present invention pertains to an apparatus **80** for wirelessly powering a power harvesting device **12** from a battery **74** charging unit, as shown in FIG. **14**. The apparatus **80** comprises an RF power transmitter **14**. The apparatus **80** comprises an electrical interface to which the transmitter **14** is electrically connected that electrically connects with the battery **74** charging unit.

[**0079**] The design of a Radio-Frequency (RF) Power Transmitter and an RF Power Network have been described in detail in U.S. patent application Ser. No. 11/356,892, “Pulse Transmission Method”, U.S. continuation-in-part patent application Ser. No. 11/651,818, “Pulse Transmission Method”, U.S. provisional application Ser. No. 11/438,508, “Power Transmission Network”, and U.S. provisional continuation-in-part application 60/833,864, “Power Transmission Network and Method”, all incorporated by reference herein. The referenced patents give great detail on how an RF Power Transmitter and an RF Power Network can be constructed for various transmitter and antenna **50** combinations. It, however, also becomes advantageous, and the focus of the invention, to describe how the RF Power Transmitters **14** and RF Power Network derive the power used for operating the components necessary to both such as, but not limited to, transmitter(s) **14**, controller(s) **48**, and/or antenna(s) **50**. It should be noted that an RF Power Network is made up of more than one RF power transmitter **14** where the coverage areas **38** may or may not overlap and the RF power transmitter **14** includes one or more antenna **50** for transmitting the generated RF power which may be pulsed or continuous. It should also be noted that RF power transmitters **14** and RF power networks may be used to directly power one or more RF harvesting devices or charge, recharge, or trickle charge a power storage component. An RF power receiver such as, but not limited to, the RF power receivers described in U.S. provisional application **11/584, 983**, "Method and Apparatus for High Efficiency Rectification for Various Loads", may be used with the presented invention although any RF harvesting device may be used. It should be noted that a device containing RF harvesting circuitry may be referred to herein as an RF harvesting device, an RF power harvesting device **12**, or an RF power receiver. Additionally, the apparatus **80**, with or without the receiving antenna **50**, for converting the RF energy to a usable form such as, but not limited to, direct current (DC), may be referred to herein as the RF harvesting circuitry, RF power harvester, or RF power receiver.

[0080] It should be noted that the RF power transmitters **14** and RF power network in the invention should not be confused with RF power transmitters **14** using inductive coupling, which requires the device to be relatively close to the power transmission source. The RFID Handbook by the author Klaus Finkenzeller defines the inductive coupling region as distance between the transmitter **14** and receiver of less than 0.16 times lambda where lambda is the wavelength of the RF wave. The proposed invention can obtain power in the near field (sometimes referred to as inductive) region as well as the far-field region. The far-field region is distances greater than 0.16 times lambda.

[0081] One method for obtaining power for an RF power transmitter **14** and/or RF power network would be to hardwire the RF power transmitter **14** or RF power network to an Alternating Current (AC) power grid or source used to supply power to lights, outlets, and other devices with voltages from 100 to 240 volts. This may be the ideal choice for new construction projects where the wiring can be easily accessed and the RF power transmitters **14** can be installed along with the wiring, lighting fixtures **26**, switches, and outlets. The RF power transmitter **14** and/or RF power network may then contain an AC to DC converter **28** that can convert the AC power obtained from the AC power main or source to a usable DC voltage (or current) such as but not limited to 3.3 to 48 volts. An example of this transmitter **14** and network implementation is shown in FIG. **1**. It should be noted that the RF power transmitter(s) **14** and/or RF power network may be hardwired to a DC network or source, if available, and, if needed, the RF power transmitter(s) **14** and/or RF power network may use a DC to DC converter to obtain the correct operational voltage. An example of a DC network or source includes, but is not limited to, the wiring within an automobile, car, truck, van, recreational vehicle, bus, public transportation, commercial truck, commercial equipment, construction equipment, industrial equipment, farm equipment, airplane, boat, ship, submarine, computer **58**, or any other manned or unmanned apparatus **80** containing a DC network or source. As an example, an RF power transmitter **14** may be hardwired to the 12 volt DC network or source within an automobile. At least one RF power transmitter **14** may be installed in the engine compartment, dash, ceiling, cabin, or trunk of the automobile to provide RF power to RF power harvesting devices **12** inside or outside (if within the coverage area **38**) of the automobile. The RF power harvesting devices **12** may include, but are not limited to, cellular phones, cellular phone accessories, car phones, voice communicating devices, PDAs, music players, laptops, toys, car sensors, or other devices that may require power.

[0082] As an example, an RF power transmitter **14** may be mounted on a utility pole **32** and hardwired to the AC power grid **16** in order to provide RF power to RF energy harvesting devices within the coverage area **38**. Multiple RF power transmitters **14** may be implemented in order to provide an RF power network.

[0083] As a specific example, at least one RF power transmitter **14** may be implemented by direct hardwiring to the DC network in the dashboard of an automobile in order to charge a cellular phone containing RF power harvesting circuitry while the cellular phone is inside the automobile or outside the automobile but still within the coverage area **38** of the RF power transmitter **14**. The coverage area **38** for the automobile in this example would be designed to give coverage over the cabin of the automobile as shown in FIG. **2**. For most automobiles, the coverage area **38** would have a range of 6 to 8 feet from the RF power transmitter **14**. The coverage area **38** is defined by a minimum electric and/or magnetic field strength produced by the RF power transmitter **14**. The range of the RF power transmitter **14** or coverage area **38** is defined as the distance from the RF power transmitter **14** to the outer limit of the coverage area **38** for a specific angle with respect to the RF power transmitter **14**. The coverage area **38** may take on different shapes and sizes and is dependent on numerous factors including, but not limited to, the RF power transmitter **14** power level, the gain and radiation pattern of the RF power transmitting antenna **50**, the environment, and the power needs of the RF power harvesting devices **12** in the coverage area **38**. For an automobile, it may be necessary to provide 1 milli-watt (mW) of power to an RF power harvesting device **12** within the cabin having a maximum range of six feet. If omnidirectional antennas **50** are used by the RF power transmitter **14** and the RF harvesting circuit at 915 mega-Hertz (MHz), the RF power transmitter **14** would need to supply 2 Watts (W) of power to the RF power transmitting antenna **50** in order to supply the needed power to the RF power harvesting device **12** at a range of six feet. An RF power transmitter **14** has been designed and constructed to meet these requirements. The adjustable RF power transmitter **14** is capable of transmitting 0.25 W to 20 W of power as a continuous-wave (CW) or as a pulsed-wave (PW). The transmitter **14** has outside dimensions of 1.5×1.5×4.775 inches as shown in FIG. **3**. The omnidirectional antennas **50** at 915 MHz may be implemented with half-wave dipoles that have a length of 6 inches and a diameter of 0.1 inches.

[0084] It should be noted that the RF power transmitter **14** shown in FIG. **3** may be used with any of the embodiments herein if found to be advantageous. The RF power transmitter **14** may contain a power input **42** for accepting AC or DC power, a frequency generator **44** for generating the appropriate frequency(ies), an amplifier **46** and/or preamplifier for adjusting the output power (gain or attenuation), a controller **48** for controlling the amplifier **46** and frequency generator **44**, a heat sink **54** for dissipating heat from or cooling the RF power transmitter **14**, a fan **56** for providing air flow through or across the heat sink **54** and/or printed circuit board **52** for cooling, a printed circuit board **52** (PCB) for component mounting, and an RF output connection for supplying the RF power to the RF power transmitting antenna **50**. The RF power transmitting antenna **50** may also be integrated onto the PCB.

[0085] It should be noted that one or more antennas **50** may be used with the invention and the antennas **50** may have omnidirectional or directional radiation patterns and may be designed to have linear, circular, elliptical, dual, or any other type of polarization that may be advantageous to the RF power system **10**.

[0086] For the case of a DC network or source with a computer **58**, an RF power transmitter **14** may be hardwired to the **12**-volt power supply **60** in order to supply RF power to devices located in or around the computer **58**. The computer **58** may have one or more antennas **50** located internally or externally in communication with the RF power transmitter **14**. As an example, the RF power transmitter **14** may be located in the computer **58** case while the antennas **50** are mounted in or on the monitor or display. The monitor may have two antennas **50**, one on each side of the screen in order to give a better coverage area **38** or network. The antennas **50** may be connected to the computer **58** case using one or two coaxial cables or the RF power may be supplied through a conductor in the monitor cable. As a specific example, the RF power transmitter **14** may have dimensions of 5.75 by 6.69 by 1.63 inches allowing the RF power transmitter **14** to fix into a bay within the computer **58** tower typically used for CD-ROM and DVD drives. The RF power transmitter **14** may accept a plug **82** from the computer **58** power supply **60**. The RF power transmitting antenna **50** may be external to the computer **58** tower or may be formed on the front of the RF power transmitter **14**. Additionally, the RF power transmitter **14** may be in communication internally with the computer **58** or part of the computer **58** for control of the RF power transmitter **14** or for controlling communication with RF power harvesting devices **12** receiving RF power from the RF power transmitter **14**. The RF power transmitter **14** may also be formed as a card designed to plug into standard computer **58** or laptop slots such as, but not limited to, PCI bus slots, AGP slots, PCI express slots, ISA slots, PCMCIA slots, or any other computer **58** or laptop slot. In certain applications, the RF power transmitter **14** may also be formed on the motherboard of the computer **58** with the RF power antenna **50** being internal or external to the computer **58** tower.

[0087] Additionally, the RF power transmitters **14** may be recessed or flush mounted like an AC or DC outlet **18** or switch, or may replace, or be used in conjunction, with an AC or DC outlet **18** whether existing or specialized to include an RF power transmitter **14** and RF power antenna **50**. An example of this can be seen in FIG. **4** where the AC outlet **18** from FIG. **1** has been replaced with the RF power transmitter **14**. The RF power transmitter **14** is mounted flush with the wall. In this example, the box retains the function of providing AC power to devices that plug into it, but also transmits RF power through an RF power antenna **50** that is mounted behind the wall. Furthermore, the RF power transmitter **14** may fit completely into a standard junction box **34** with the RF power antenna **50** being internal or the antenna **50** may be connected outside the junction box **34** by a connector exiting through the junction box **34** or junction box **34** cover. The RF power transmitter **14** may have dimensions of 3.8 by 3.8 by 2.1 inches in order to fit in a junction box **34**, and the RF power transmitting antenna **50** may have a length of 6 inches and a diameter of 0.1 inches for a 915 MHz RF power transmitter **14**. In certain instances the cover of the junction box **34** could be or contain the RF power antenna **50**. It may also be possible to embed the RF power transmitter **14** and/or RF power antenna **50** on or in, partially or completely, the material of the structure that the RF power transmitter **14** and/or RF power antenna **50** are mounted to, in, or behind, depending on the attenuating properties of the material. The RF power transmitter **14** and/or RF power antenna **50** may also be located behind the material to eliminate the need for an opening in the material for the RF power transmitter **14** and/or RF power antenna **50** to protrude through. As an example, the RF power transmitter **14** and RF power antenna **50** may be implemented by direct connection to the AC power main or source and be completely located behind a low attenuation wall at the frequency of the RF power transmitter **14**.

[0088] In cases of an existing structure or an implementation requiring a simpler installation, it may be advantageous to develop additional methods for deriving the power for the RF Power Transmitter **14** and/or RF Power Network. As an example, in an existing building it may be necessary to access wiring inside the walls or ceiling in order to hardwire the RF power transmitter(s) **14** to the AC power main which may require that an opening be made and repaired. The destruction and construction required for the opening in the wall may require building permits and trained or experienced personal such as an electrician, carpenter, or other contractor. This process may not be an attractive solution for certain implementations. Therefore, it becomes necessary to develop additional methods other than direct hardwiring of the RF power transmitter(s) **14**. One such method is to design a transmitter **14** that can be used in conjunction with an existing lighting fixture **26**, lamp, or other power receptacle for a light source. A light source may include, but is not limited to, a light bulb, an incandescent light, a fluorescent bulb, a fluorescent lamp, a halogen bulb, a light-emitting diode (LED), an organic light-emitting diode (OLED), a full spectrum bulb, or any other light producing device. As an example, a transmitter **14** could be constructed in a fashion that would enable it to screw or plug into an existing lighting fixture **26**, lamp, or other power receptacle for a light **22** source using a standard or custom base, such as but not limited to, Candelabra/E12, Intermediate/E17, Medium/E26, Mogul/E39, Bayonet, (T8) Medium Bi-Pin, (T12) Medium Bi-Pin, (T5) Miniature Bi-Pin, or any other type of connector used to connect the light **22** source to the AC or DC power main. The resulting transmitter **14** would replace the light **22** source to provide coverage of RF power rather than light **22** where the RF energy could be used to deliver power to devices containing RF power harvesting circuitry. An example of this invention can be seen in FIG. **5**. In this example, the RF power transmitter **14**, like the light **22** source it is replacing, protrudes from the lighting fixture **26**, but in some cases the, the RF power transmitter **14** may be recessed into the fixture. The RF power transmitter **14** may have dimensions of 1.6 by 1.6 by 4 inches in order to fit in a lighting fixture **26**, and the

RF power transmitting antenna **50** may have a length of 6 inches and a diameter of 0.1 inches for a 915 MHz RF power transmitter **14**.

[**0089**] When implementing an RF power transmitter **14** solely or within an RF power network, it may not always be practical to completely remove the light **22** source from the lighting fixture **26**, lamp, or other power receptacle for a light **22** source. In this case, the transmitter **14** could, as previously described, screw or plug into the existing lighting fixture **26**, lamp, or other power receptacle for a light **22** source and also include the ability to accept a light **22** source with a standard or custom base, such as but not limited to, Candelabra/E12, Intermediate/E17, Medium/E26, Mogul/E39, Bayonet, (T8) Medium Bi-Pin, (T12) Medium Bi-Pin, (T5) Miniature Bi-Pin, or any other type of connector used to connect the light **22** source to the AC or DC power main. The ability to accept a light **22** source allows the lighting fixture **26**, lamp, or other power receptacle for a light **22** source to contain an RF power transmitter **14** and a light **22** source meaning the lighting fixture **26**, lamp, or other power receptacle for a light **22** source can perform its primary function of supplying light **22** while also performing a secondary function of providing RF power to devices containing RF power harvesting circuitry. An example of this method can be seen in FIG. **6**. It should be noted that the type of base in the lighting fixture **26**, lamp, or other power receptacle for a light **22** source may be different than the type of base in the RF power transmitter **14** that accepts the light **22** source. More specifically, the lighting fixture **26**, lamp, or other power receptacle for a light **22** source may have a Mogul/E39 base while the RF power transmitter **14** accepts a light **22** source with a Medium/E26 base. Additionally, the lighting fixture **26**, lamp, or other power receptacle for a light **22** source may be recessed into the ceiling, wall, or mounting surface. The RF power transmitter **14** may have dimensions of 4 by 4 by 1 inch in order to fit into a recessed lighting fixture **26** that accepts a light bulb, and the RF power transmitting antenna **50** may have a length of 6 inches and a diameter of 0.1 inches for a 915 MHz RF power transmitter **14**.

[**0090**] To simplify the previous example, the light **22** source could be integrated into the RF power transmitter **14** to allow the light **22** source and RF power transmitter **14** to work in conjunction with one another as shown in FIG. **7**. As an example, the antenna **50** of the RF power transmitter **14** could be formed from part of metal contained within the existing light **22** source or the antenna **50** could be integrated into or onto the light **22** source as a custom solution. The antenna **50** may be formed by depositing metal or any other conductive material on the glass of the light **22** source to form a resonant antenna **50** structure. The conductive material may have transparent or semi-transparent properties to allow light **22** to pass through the antenna **50** structure. A transparent antenna **50** may be formed using a material such as, but not limited to, Indium Tin Oxide. The antenna **50** may also be formed inside the light **22** source if found to be advantageous. In cases where a directional light **22** source is used to focus the light **22** to a specific area using a light **22** reflective surface, the antenna **50** could use the reflective surface, if metallic, to also reflect or focus the RF energy transmitted from the antenna **50**. It may become necessary to use a long-life light **22** source such as an LED to reduce the amount of maintenance on each lighting fixture **26**, lamp, or other power receptacle for a light **22** source. When the light **22** source or RF power transmitter **14** cease normal operation, the RF power transmitter **14** and light **22** source combination can be easily replaced by unscrewing or unplugging the RF power transmitter **14** with integrated light **22** source. The used device may be repaired or simply discarded depending on the application. It should be noted that the RF power transmitter **14** and light **22** source combination may be recessed into the lighting fixture **26**. The RF power transmitter **14** with integrated light **22** source may have dimensions of 4 by 4 by 6 inches in order to fit into a recessed lighting fixture **26**, and the RF power transmitting antenna **50** may have a length of 6 inches and a diameter of 0.1 inches for a 915 MHz RF power transmitter **14**.

[**0091**] In the two previous implementations, the RF power transmitter **14** accepted or had a built-in light **22** source. In certain applications, the RF power transmitter **14** may be integrated into an existing or specialized lighting fixture **26**. As an example, typical lighting in an office building is provided by lighting fixtures **26** containing four four-foot fluorescent lights. This type of lighting fixture **26** may be retrofitted with at least one RF power transmitter **14** or the lighting fixture **26** may be redesigned to contain at least one RF power transmitter **14**. An example of this can be seen in FIG. **8**. As an additional example, an RF power transmitter **14** may be used in conjunction with an existing or specialized light bulb and/or fixture within a street light for the purpose of providing power to RF power harvesting devices **12** within the coverage area **38** defined by one or more street lights. Additionally, an RF power transmitter **14** may be implemented with landscape, exterior, emergency, specialty, automobile, or any other type of lighting fixture **26** or light **22** producing source. For the case of the automobile, an RF power transmitter **14** may be implemented with or within the interior lights or headlights to provide RF power to devices within the resulting coverage area **38**.

[**0092**] Another way of implementing an RF power transmitter **14** and/or RF power network is to connect the transmitter(s) **14** to existing outlets **18**, receptacles, ports, or connectors within a building, automobile, device or structure by a plug **82** and cord **20** that can be used to provide AC or DC directly from the outlet **18**. In most cases for an AC power grid **16** or source, the DC power may be obtained from an AC to DC converter **28** located at the outlet **18**, receptacle, port **64**, connector or somewhere between the outlet **18**, receptacle, port **64**, or connector and RF power transmitter **14**. As an example, an RF power transmitter **14** and/or RF power network may be designed to give coverage over a desk, a room, an entire home, an entire building floor, the entire building, or an automobile. The coverage area **38** is defined by a minimum electric and/or magnetic field strength produced by the RF power transmitter **14**. As in the case of the desk or room, a single RF power transmitter **14** may be sufficient to give coverage over the required area. Therefore, the RF power transmitter **14** may be designed to plug into an existing outlet **18** near the desk or somewhere within the room. The ability to have a cord **20** gives flexibility to the RF power network design by allowing the use of an existing AC or DC outlet **18** with the ability to place the RF power transmitter **14** away from the AC or DC outlet **18**. As an example, it may be necessary to provide RF power coverage over a bedroom in order to recharge a medical implant within a patient while they are located within their bedroom. The RF power transmitter **14** can then be located on the nightstand or can be attached to the headboard in order to supply RF power to the medical implant to recharge the implant's battery **74** or power storage component. The RF power transmitter **14** can obtain its operational power by plugging into one of the AC outlets **18** in the room where the AC outlet **18** may be located several feet from the RF power transmitter **14** location as shown in FIG. **9**. Several RF power transmitters **14** have been designed at 915 MHz that plug into an existing AC outlet **18** using a power cord **20**. The first RF power transmitter **14** had dimensions of 2.6 by 4.25 by 1.26 inches and an output power of 0.5 W while the second had dimensions of 4.4 by 6.4 by 2 inches and an output power of 5 W. The RF power transmitting antenna **50** for the first transmitter **14** was a monopole with a length of 3 inches while the second RF power transmitting antenna **50** was a dipole and had a length of 6 inches and a diameter of 0.1 inches.

[0093] For the case of an automobile, the RF power transmitter **14** may plug into the 12V DC power outlet **18** or cigarette lighter outlet **18** through a cord **20** and the RF power transmitter **14** may then be placed on the dash or the center console in order to provide RF power to devices containing RF power harvesting circuitry in the coverage area **38** provided by the RF power transmitter **14** and RF power antenna **50**.

[0094] For the case of a computer **58**, the RF power transmitter **14** may plug into existing or specialized computer **58** ports **64**, such as, but not limited to, USB, serial, parallel, FireWire, or any other power carrying port **64**, through a cord **20** in order to supply power to an RF power transmitter **14**.

[0095] It should be noted that an RF power transmitter **14** may plug into other devices, directly or with a cord **20**, such as, but not limited to, a console gaming system, computer **58**, laptop computer **58**, or any other device have an outlet **18**, receptacle, port **64**, or connector that may be used to obtain power for an RF power transmitter **14**.

[0096] It should also be noted that any of the RF power transmitters **14** described herein may have the ability to supply power to other devices by having an outlet. **18**, receptacle, port **64**, or connector that may be the same or a different type than the one supplying the RF power transmitter **14**. As an example, an RF power transmitter **14** being powered from a USB or Ethernet port **64** may have a USB or Ethernet port **64** to allow other devices to use the same USB or Ethernet port **64** as the RF power transmitter **14**.

[0097] In certain applications, it may not be necessary to run a cord **20** from the outlet **18**, receptacle, port **64**, or connector to the RF power transmitter **14**. The AC or DC outlet **18**, receptacle, port **64**, or connector may be positioned in a location that provides the required RF energy coverage when the RF power transmitter **14** is located at the location of the outlet **18**, receptacle, port **64**, or connector. In these cases, the RF power transmitter **14** can simply plug into the outlet **18**, receptacle, port **64**, or connector without the need for an extension cord **20**. The RF power transmitter **14** may be supported in whole or part by the friction created from the AC or DC prongs inserted into the outlet **18**, receptacle, port **64**, or connector. Additionally, the RF power transmitter **14** may pass the AC or DC power to at least one AC or DC outlet **18**, receptacle, port **64**, or connector located on the RF power transmitter **14** in order to enable other devices to plug into the AC or DC main or source through the RF power transmitter **14**. The RF power transmitter **14** may have one or more antenna **50** that are used to radiate and/or direct the RF power away from the outlet **18**, receptacle, port **64**, or connector to an RF power receiving device containing RF power harvesting circuitry which can harvest the available RF power to power a device or charge or recharge a charge storage component such as a battery **74**, capacitor, or other charge storage component. FIG. **10** shows an example of an RF power transmitter **14** that plugs directly into an AC outlet **18**. An RF power transmitter **14** plugging directly into an AC outlet **18** may have dimensions of 2.6 by 4.25 by 1.26 inches, and the RF power transmitting antenna **50** may have a length of 6 inches and a diameter of 0.1 inches for a 915 MHz RF power transmitter **14**.

[0098] For the case of an automobile, the RF power transmitter **14** may plug directly into the 12V DC power outlet **18** or cigarette lighter outlet **18** without the need for a cord **20** in order to provide RF power to devices containing RF power harvesting circuitry in the coverage area **38** provided by the RF power transmitter **14** and RF power antenna **50**. An example of an RF power transmitter **14** that plugs directly into the DC power outlet **18** of an automobile can be seen in FIG. **11**. An RF power transmitter **14** plugging directly into a DC outlet **18** may have dimensions of 2 by 2 by 1 inch, and the RF power transmitting antenna **50** may be internal or external to the RF power transmitter **14** and have a length of 6 inches and a diameter of 0.1 inches for a 915 MHz RF power transmitter **14**.

[0099] For the case of a computer **58**, the RF power transmitter **14** may plug directly into existing or specialized computer **58** ports **64**, such as, but not limited to, USB, serial, parallel, FireWire, or any other power carrying port **64**, in order to supply power to an RF power transmitter **14**.

[0100] It may be beneficial in certain applications to include a battery **74** charger or power storage component charger with the RF power transmitter **14**. This solution is of particular interest when the RF power harvesting device **12** may require more power than the RF power transmitter **14** or RF power network can provide or if the RF power harvesting device **12** needs to obtain a fast charge such as when the battery **74** voltage level has fallen below the minimum threshold for operation of the device. The battery **74** or other charge storage component that is normally charged or recharged from the RF power transmitted by the RF power transmitters **14** and/or RF power network could be removed from the device and placed in the battery **74** or charge storage component recharger built-in to the RF power transmitter **14** for a faster charging where the recharger is powered directly by the AC or DC power main. A battery **74** or charge storage component charger could be included in any of the implementations described in this document. An example of including a battery **74** charger with an RF power transmitter **14** is shown in FIG. **12** for the case of the RF power transmitter **14** that directly plugs into an AC outlet **18**. An RF power transmitter **14** with a battery **74** charger plugged directly into an AC outlet **18** may have dimensions of 2.6 by 4.25 by 1.26 inches, and the RF power transmitting antenna **50** may have a length of 6 inches and a diameter of 0.1 inches for a 915 MHz RF power transmitter **14**.

[0101] For the case of an automobile, the RF power transmitter **14** may plug directly into the 12V DC outlet **18**

or may have a cord **20** and the RF power transmitter **14** may then be placed on the dash or the center console in order to provide RF power to devices containing RF power harvesting circuitry in the coverage area **38** provided by the RF power transmitter **14** and RF power antenna **50**. Additionally, the RF power transmitter **14** may contain a battery **74** charger or charge storage component charger in order to obtain a faster charge cycle. The charger may be designed to accept standard battery **74** sizes such as AA, AAA, C, and/or D cell batteries or may be designed to accept a product specific battery **74** that may or may not be attached to the device at the time of charging. As an example, a cellular phone may contain RF power harvesting circuitry for capturing RF power when within the coverage area **38** provided by the RF power transmitter **14**. The RF power transmitter **14** may also contain a cradle with charging connections that would allow the cellular phone to be directly charged by a hardwired connection in order to obtain a faster charge.

[**0102**] In certain applications, the RF power transmitter **14** may obtain operational power from a battery **74** or charge storage component in order to transmit RF power. The battery **74** or charge storage component may include, but is not limited to, rechargeable batteries, capacitors, fuel cells, generators, other charge storage components, or other charge generating components. In some cases, the RF power transmitter **14** may draw its power from a battery **74** or charge storage device that is supplying power to other devices simultaneously. For example, a laptop computer **58** uses a battery **74** for operational power. An RF power transmitter **14** could be attached with or without a cord **20** to the laptop computer **58**, by means disclosed herein such as through the USB port **64** of the laptop computer **58**, and would use the same battery **74** for operational power that the laptop computer **58** is using. An example of the RF power transmitter **14** directly connected to a computer **58** is shown in FIG. **13**. The RF power transmitter **14** could then supply power to computer **58** peripherals or other devices within its coverage area **38**, such as, but not limited to, keyboards, mice, game controllers, cellular phones, cellular phone accessories, PDAs or other peripherals or devices that are designed with RF power harvesting circuitry. An RF power transmitter **14** plugging directly into the USB or other port **64** of a computer **58** may have dimensions of 3 by 0.75 by 0.75 inches, and the RF power transmitting antenna **50** may be integrated with the RF power transmitter **14**.

[**0103**] In certain applications, the battery **74** or charge storage component used to run the first RF power transmitter **14** may be receiving power from a second RF power transmitter **14** for the purpose of charging the battery **74** or charge storage element in the first RF power transmitter **14**. Power for the second RF power transmitter **14** may be obtained from an AC or DC power network or by other means described herein. As an example, it may be necessary to supply power to an RF power-harvesting device, which is located in a position that does not allow direct line-of-sight or a low attenuation transmission path. As a specific example, an RF power transmitter **14** may be required to supply power to a valve sensor **76** in an industrial application. However, the AC power grid **16** used for obtaining operational power for the RF power transmitter **14** may be located on one side of a large metal storage tank while the valve sensor **76** requiring power may be located on the opposite side. In order to obtain sufficient power at the valve sensor **76**, an additional RF powered RF power transmitter **14** may be required in order to direct, relay, or bounce the power around the metal storage tank as shown in FIG. **14**. The RF powered RF power transmitter **14** may be an RF power transmitter **14** also containing an RF power harvesting device **12** or may be implemented with a passive RF repeater **78**. The passive RF repeater **78** receives the power with one antenna **50** and passes the RF power to a second antenna **50**, which retransmits the power in a different direction.

[**0104**] When deploying multiple RF power transmitters **14** in an RF power network, it becomes necessary to develop a method that allows the installer or user to easily and quickly install or add RF power transmitters **14** to the RF power network. One solution is to build an RF power transmitter **14** that can be installed into the track **66** of existing track **66** lighting. The RF power transmitters **14** can then be easily retrofitted into existing structures or places containing track **66** lighting. The RF power transmitter **14** can simply snap or screw into the track **66** to obtain AC or DC power depending on the type of track **66** lighting. The power track **66** could then contain both lights **22** and RF power transmitters **14** although the track **66** may contain only RF power transmitters **14**. It should be noted that the track **66** may contain RF power transmitters **14** that accept or have built-in light **22** sources as previously described herein.

[**0105**] It is also possible to develop a specialized type of track **66** that allows the tracks **66** to be concatenated with track **66** junctions to incorporate a large number of RF power transmitters **14** in order to cover a large area. The tracks **66** may be, but are not limited to, six feet in length. The track **66** junctions may contain, but are not limited to, a connector (whether a plug **82**, snap, or clip) or a slip-in fitting in order to concatenate tracks **66** to obtain longer lengths. It is also possible to design a track **66** junction that could be used to connect two or more tracks **66** together by connectors or slip-in fittings in order to change the direction of the track(s) **66** or to connect multiple tracks **66**. Additionally, it is possible to design an RF power transmitter **14** that can be used as an RF power transmitter **14** and a track **66** junction. An example of a track **66** system for implementation of an RF power network can be seen in FIG. **15**. It should be noted that the track **66** may contain light **22** sources.

[**0106**] Each track **66** may contain at least two conducting portions in order to provide an input and return for the AC or DC power for the RF power transmitters **14**. As an example for an AC main implementation, the input line is the AC hot wire, typically the black wire, and the return line is the AC neutral wire, typically the white wire. The track **66** and supporting structure, if metal, may be connected to the AC ground for safety purposes. There may also be a ground wire and a wire used for communicating between the RF power transmitters **14** in order for an RF power transmitter **14** to obtain information about other RF power transmitters **14** operations such as, but not limited to, pulse timing, polarization, frequency, power level, transmission algorithm, antenna **50** gain, or other pertinent information. The communication between the RF power transmitters **14** may be done by, but not limited to, a microcontroller integrated in the RF power transmitter **14** with each having a unique identification or a master/slave configuration. It should be noted that for large implementations of RF power transmitters **14**, it may become necessary to split the communicating portion of the network into multiple smaller networks which could be accomplished with, but not limited to, a special track **66** junction that only passes the AC or DC power and isolates the communication conductors.

[**0107**] The tracks **66** used for the invention can take many different forms. The proposed invention can be implemented with any type including, but not limited to, snap-in tracks **66**, screw-in tracks **66**, sliding tracks **66**, concatenatable tracks **66**, AC tracks **66**, DC tracks **66**, or any other track **66** that can supply current to at least one RF power transmitter **14**. It should be noted that the track **66** may take on various shapes including, but not limited to, those shown in FIG. **16**.

[**0108**] One track **66** that is particularly advantageous is the coated cable shown in FIG. **16***c*. One cable is the input path while the other cable acts as the return path. The cable could be used to supply either AC or DC power to the RF power transmitter **14** although DC would have numerous advantages due to safety issues, fire concerns, and building regulations. The RF power transmitters **14** could be set on top of the cable, which would supply power and support the RF power transmitters **14**. The RF power transmitters **14** could have at least one screw for each cable in order to secure the RF power transmitter **14** to the cable and to pierce the non-conducting protective coating on the outside of the conducting cable. An example of an RF power transmitter **14** connected to the cable track **66** system **10** can be seen in FIG. **17**. It should be noted that additional cables could be used for, but not limited to, ground, communication, or some other signal if found to be advantageous. As with the network shown in FIG. **15**, the cable track **66** system **10** could be used to provide coverage over a hallway or hallways. The main advantage of the cable track **66** system **10** is its easy installation. The cable may be a large spool of cable allowing long runs of track **66** without the need for track **66** junctions. It is also possible to implement curved tracks **66** using the cable track **66** system **10**, which would allow the tracks **66** to turn corners or be installed in a circular fashion in a large room. Track **66** junctions may be used to connect multiple tracks **66** together as previously described. The supports **72** for the tracks **66** may simply snap or clamp to the cable to provide the proper spacing and support to the cable and RF power transmitters **14** and RF power antennas **50**. The RF power cable track **66** system **10** may be implemented behind a material for aesthetic purposes such as, but not limited to, a wall, ceiling, or drop ceiling.

[**0109**] An RF power transmitter **14** has been designed and constructed to meet the requirements of the track **66** system **10**. The adjustable RF power transmitter **14** is capable of transmitting 0.25 W to 20 W of power as a continuous-wave (CW) or as a pulsed-wave (PW). The transmitter **14** has outside dimensions of 1.5×1.5×4.775 inches. The RF power transmitting antenna **50** at 915 MHz may be implemented with a half-wave dipole that has a length of 6 inches and a diameter of 0.1 inches.

[**0110**] It should be noted that the RF power transmitters **14** described herein may contain communication circuitry and a communication antenna **50** in order to obtain operational information such as, but not limited to, timing, transmitted power, transmission algorithm, frequency, antenna **50** characteristics, or any other information from other RF power transmitters **14**. Additionally, the RF power transmitters **14** may contain a power sensor and antenna **50** for measuring the amount of power transmitted by other RF power transmitters **14** in order to obtain information such as, but not limited to, timing, transmitted power, transmission algorithm, frequency, antenna **50** characteristics, or any other information from other RF power transmitters **14**.

[**0111**] It should be noted that the RF power transmitters **14** described herein may be implemented as a single RF power transmitter **14** or as part of an RF power network where the coverage area **38** of each RF power transmitter **14** may or may not overlap.

[**0112**] Numerous examples have been given herein that describe the physical size of the RF power transmitter **14** and RF power transmitting antenna **50** which may be dependent on one or more of several factors including, but not limited to, transmitted RF power, RF power transmitting antenna **50** gain, the frequency(ies) of the RF power transmitter **14**, the required RF power coverage area **38**, heat sink **54** size, amount of air movement by the fan **56** or by the environment, ambient temperature, and the type of operational power available for the RF power transmitter **14**. These factors may be adjusted or modified to obtain the desired physical size needed in order to implement the RF power transmitter **14** in a practical application such as, but not limited to, using an RF power transmitter **14** plugged directly into a computer **58** to supply RF power to an RF power harvesting device **12** that has been installed into a cellular phone, or using an RF power network to provide an RF power coverage area **38** that covers an office.

[**0113**] The coverage areas **38** and range of the RF transmitters **14** described herein may be dependent on one or more of numerous factors including, but not limited to, transmitted RF power, RF power transmitting antenna **50** gain, the frequency(ies) of the RF power transmitter **14**, the type and amount of operational power available for the RF power transmitter **14**, and the maximum amount of RF power needed to operate the RF power harvesting device(s) **12**. These factors may be adjusted or modified to obtain the desired coverage area **38** needed in order to implement the RF power transmitting and RF power harvesting system **10**.

[**0114**] It should be noted that the operational power for an RF power transmitter **14** and/or RF power network described in the invention herein may be derived from numerous AC or DC sources including, but not limited to, an AC power network, AC power grid **16**, AC power main, DC power network, DC power grid **30**, DC power main, telephone lines or jacks, Ethernet cable or jacks, cable network, or any other AC or DC source. The wiring for these sources may include, but is not limited to, building wire (**10-2, 10-3, 12-2, 12-3, 14-2, 14-3**), telephone wire, CAT-3, CAT-5, CAT-6, coaxial cable, or any other wiring or cable. The way to connect to these wires or to a device may include, but is not limited to, a 2-prong plug **82**, 3-prong plug **82**, DC power plug **82**, vehicle cigarette lighter or power receptacle, RJ-45 connector, RJ-11 connector, F-type connector, screw-on plug **82** or connector, SMA connector, BNC connector, N-type connector, other coaxial connectors, USB connector, mini-USB connector, Firewire connector, product specific connectors, specialized connectors or any other type of connector, plug **82**, or receptacle.

[**0115**] It will be understood by those skilled in the art that while the foregoing description sets forth in detail preferred embodiments of the present invention, modifications, additions, and changes might be made thereto without departing from the spirit and scope of the invention.

The invention claimed is:

1. A power transmission system for wirelessly powering a power harvesting device comprising:

at least one RF power transmitter; and

an AC power grid to which the transmitter is electrically connected.

2. A system as described in claim 1 wherein the power grid has an outlet.

3. A system as described in claim 2 wherein the transmitter has a cord which plugs into the outlet.

4. A system as described in claim 2 wherein the transmitter plugs directly into the outlet.

5. A system as described in claim 2 wherein the power grid has a light.

6. A system as described in claim 5 wherein the power grid has a light switch to turn on the light.

7. A system as described in claim 1 wherein the grid includes in wall wiring.

8. A system as described in claim 7 wherein the transmitter is integrated with the outlet.

9. A system as described in claim 2 wherein the grid has a lighting fixture and the transmitter contacts the fixture.

10. A system as described in claim 5 wherein the grid has a lighting fixture and one or the other of the light and the transmitter contact the fixture.

11. A system as described in claim 1 wherein the grid includes a utility pole to which the transmitter is in contact.

12. A system as described in claim 1 wherein the grid includes a junction box to which transmitter is in contact.

13. A system as described in claim 1 wherein the transmitter includes an AC to DC converter that can convert the AC power obtained from the grid to a usable DC voltage or current.

14. A power transmission system for wirelessly powering a power harvesting device comprising:

at least one RF power transmitter; and

a DC power grid to which the transmitter is electrically connected.

15. A system as described in claim 14 wherein the transmitter is disposed in a vehicle.

16. A system as described in claim 15 wherein the transmitter provides a coverage area over the cabin of the vehicle.

17. A system as described in claim 16 wherein the transmitter is in contact with the dash, trunk, cabin, ceiling, or engine compartment of the vehicle.

18. An adjustable RF power transmitter for powering wirelessly an RF power harvesting device comprising:

a housing having outer dimensions no greater than 3"×3" by 8 inches;

a power input;

a frequency generator in communication with the power input;

an amplifier in communication with the frequency generator;

a controller connected with the frequency generator; and

an antenna connected to the amplifier.

19. A transmitter as described in claim 18 including a circuit board on which the power input, the frequency generator and the amplifier are disposed.

20. A transmitter as described in claim 19 including a heat sink in contact with the circuit board.

21. A transmitter as described in claim 20 including a fan disposed adjacent the circuit board.

22. A power transmission system for wirelessly powering an RF power harvesting device comprising:

a computer;

an antenna;

an RF transmitter in communication with the antenna and the computer; and

a power supply in electrical communication with the RF transmitter and the computer.

23. A system as described in claim 22 wherein the RF transmitter is disposed in the computer.

24. A system as described in claim 22 wherein the computer has a power port and the transmitter is plugged into the power port.

25. A system as described in claim 24 wherein the power port is a USB port.

26. A system as described in claim 22 wherein the antenna is integrated with the transmitter.

27. A system as described in claim 22 including a display, and the antenna is in contact with the display.

28. An apparatus for wirelessly powering a power harvesting device comprising:

at least one RF power transmitter; and

a lighting fixture to which the transmitter is connected and from which the transmitter receives power.

29. An apparatus as described in claim 28 wherein the lighting fixture is a fluorescent lighting fixture.

30. An apparatus as described in claim 28 wherein the lighting fixture is an incandescent lighting fixture.

31. An apparatus as described in claim 28 wherein the lighting fixture is an LED lighting fixture.

32. An apparatus as described in claim 28 including a light source in electrical communication with the lighting fixture.

33. A power transmission system for wirelessly powering a power harvesting device comprising:

at least one RF power transmitter; and

a track supplying power to which the transmitter is electrically connected.

34. A system as described in claim 33 wherein there are at least two power transmitters.

35. A system as described in claim 34 including at least two lights electrically connected to the track.

36. A system as described in claim 35 wherein the track includes a first conductor and a second conductor.

37. A system as described in claim 36 wherein the track includes a support attached to a wall or ceiling.

38. A power transmission system for wirelessly powering a power harvesting device comprising:

at least one RF power transmitter; and

a battery charging unit to which the transmitter is electrically connected.

39. A power transmission system for wirelessly powering a power harvesting device comprising:

at least one RF power transmitter; and

at least one rechargeable battery to which the transmitter is electrically connected.

40. A system as described in claim 39 including a second RF power transmitter which transmits power to the power harvesting device electronically connected to the battery.

41. A system as described in claim 40 including a valve sensor powered by a power harvesting device.

42. A method for wirelessly powering a power harvesting device comprising the steps of:

electrically connecting at least one RF power transmitter to an AC power grid; and

transmitting power with the RF power transmitter.

43. A method for wirelessly powering a power harvesting device comprising the steps of:

electrically connecting at least one RF power transmitter to a DC power grid; and

transmitting power with the RF power transmitter.

44. A method for wirelessly powering a power harvesting device comprising the steps of:

electrically connecting a power supply with an RF transmitter and a computer; and

transmitting power with the RF power transmitter.

45. A method for wirelessly powering a power harvesting device comprising the steps of:

electrically connecting at least one RF power transmitter with a lighting fixture to which the transmitter is in contact with and from which the transmitter receives power; and

transmitting power with the RF power transmitter.

46. A method for wirelessly powering a power harvesting device comprising the steps of:

electrically connecting at least one RF power transmitter to a battery charging unit; and

transmitting power with the RF power transmitter.

47. A power transmission system for wirelessly powering a power harvesting device comprising:

at least one RF power transmitter; and

means for providing power to which the transmitter is electrically connected.

48. An apparatus for wirelessly powering a power harvesting device from a DC power outlet of a vehicle comprising:

an RF power transmitter; and

a power plug to which the transmitter is attached and electrically connected that plugs into the DC power outlet.

49. A power transmission system for wirelessly powering an RF power harvesting device comprising:

an antenna;

an RF transmitter in communication with the antenna; and

a connector in communication with the RF transmitter and configured to be placed in communication with a device.

50. A system as described in claim 49, wherein the device is a computer.

51. A system as described in claim 49, wherein the connector is a USB connector.

52. An apparatus for wirelessly powering a power harvesting device from an AC power grid having an AC power outlet comprising:

an RF power transmitter; and

a power plug to which the transmitter is electrically connected that electrically connects with the AC power outlet.

53. An apparatus for wirelessly powering a power harvesting device from a DC power outlet of a DC grid comprising:

an RF power transmitter; and

a power plug to which the transmitter is electrically connected that electrically connects with the DC power outlet.

54. An apparatus for wirelessly powering a power harvesting device from a computer having an antenna and a power supply comprising:

an RF power transmitter; and

a power plug to which the transmitter is electrically connected that electrically connects with the computer.

55. An apparatus for wirelessly powering a power harvesting device from a light fixture comprising:

an RF power transmitter; and

an electrical interface to which the transmitter is electrically connected that electrically connects with the light fixture.

56. An apparatus for wirelessly powering a power harvesting device from a track having at least one light comprising:

an RF power transmitter; and

an electrical interface to which the transmitter is electrically connected that electrically connects with the track.

57. An apparatus for wirelessly powering a power harvesting device from a battery charging unit comprising:

an RF power transmitter; and

an electrical interface to which the transmitter is electrically connected that electrically connects with the battery charging unit.

* * * * *

US 20070249300A1

(19) **United States**
(12) **Patent Application Publication** (10) **Pub. No.: US 2007/0249300 A1**
Sorrells et al. (43) **Pub. Date: Oct. 25, 2007**

(54) **SYSTEMS AND METHODS OF RF TOWER TRANSMISSION, MODULATION, AND AMPLIFICATION, INCLUDING EMBODIMENTS FOR COMPENSATING FOR WAVEFORM DISTORTION**

(76) Inventors: **David F. Sorrells**, Middleburg, FL (US); **Gregory S. Rawlins**, Heathrow, FL (US); **Michael W. Rawlins**, Lake Mary, FL (US)

Correspondence Address:
STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
1100 NEW YORK AVENUE, N.W.
WASHINGTON, DC 20005

(21) Appl. No.: **11/509,031**

(22) Filed: **Aug. 24, 2006**

Related U.S. Application Data

(60) Provisional application No. 60/794,121, filed on Apr. 24, 2006, provisional application No. 60/797,653, filed on May 5, 2006, provisional application No. 60/798,705, filed on May 9, 2006.

Publication Classification

(51) **Int. Cl.**
H04B 1/04 (2006.01)
H01Q 11/12 (2006.01)
(52) **U.S. Cl.** **455/114.2**; 455/127.3

(57) **ABSTRACT**

Methods and systems for vector combining power amplification are disclosed herein. In one embodiment, a plurality of signals are individually amplified, then summed to form a desired time-varying complex envelope signal. Phase and/or frequency characteristics of one or more of the signals are controlled to provide the desired phase, frequency, and/or amplitude characteristics of the desired time-varying complex envelope signal. In another embodiment, a time-varying complex envelope signal is decomposed into a plurality of constant envelope constituent signals. The constituent signals are amplified equally or substantially equally, and then summed to construct an amplified version of the original time-varying envelope signal. Embodiments also perform frequency up-conversion.

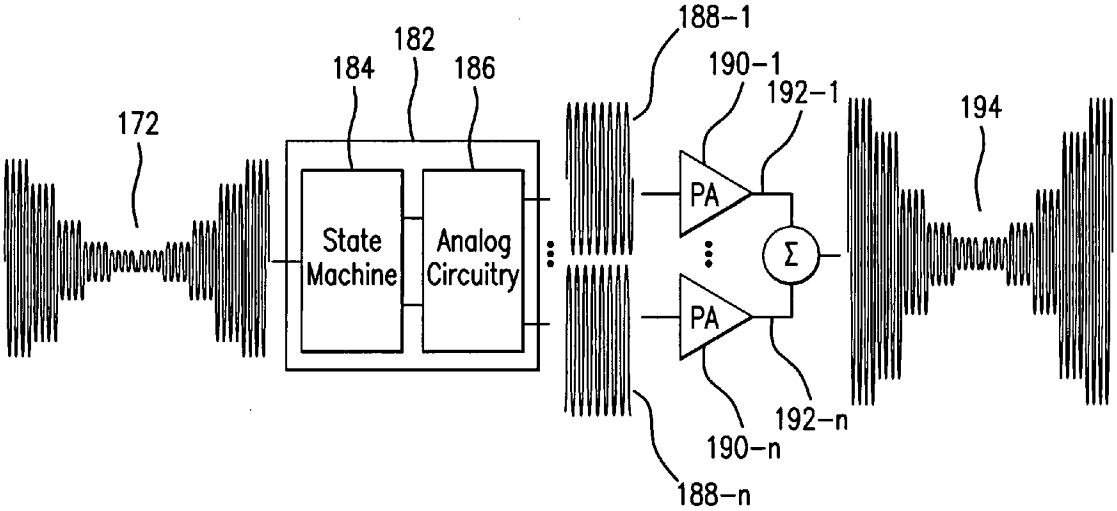

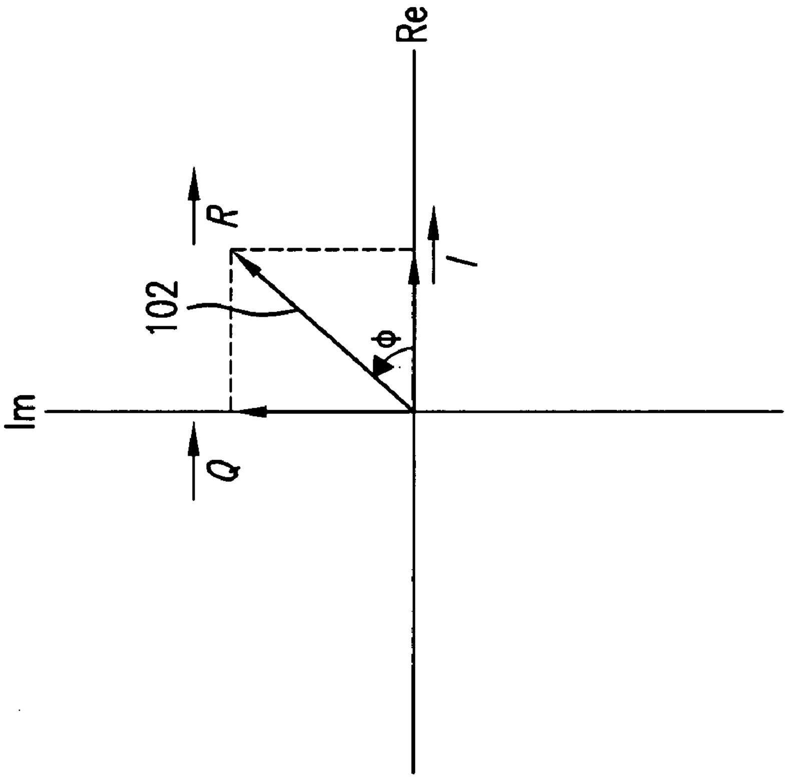
Re
Im
R
I
Q
102
φ

FIG. 1

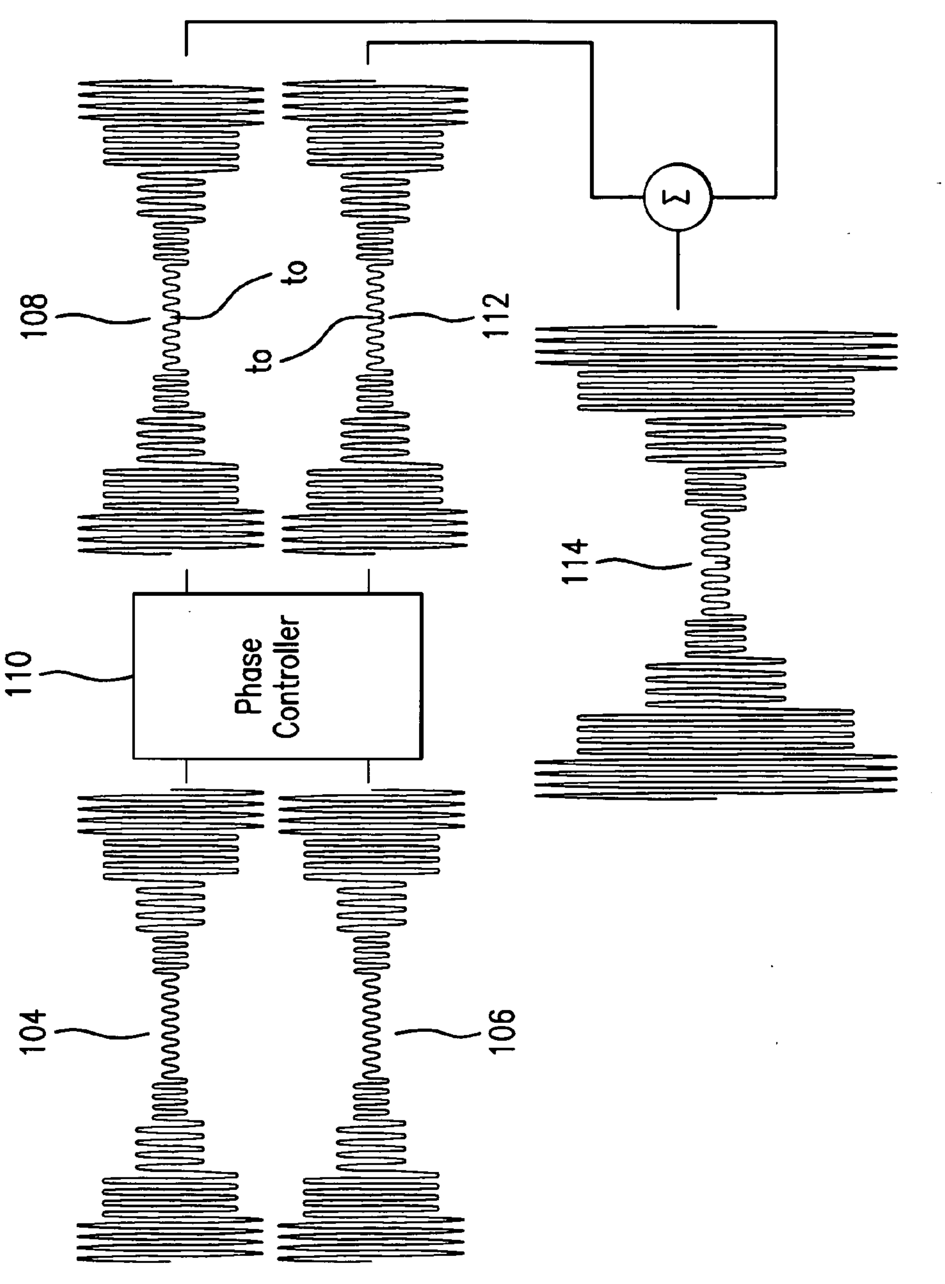
108
to
to
112
110
Phase Controller
Σ
114
104
106

FIG. 1A

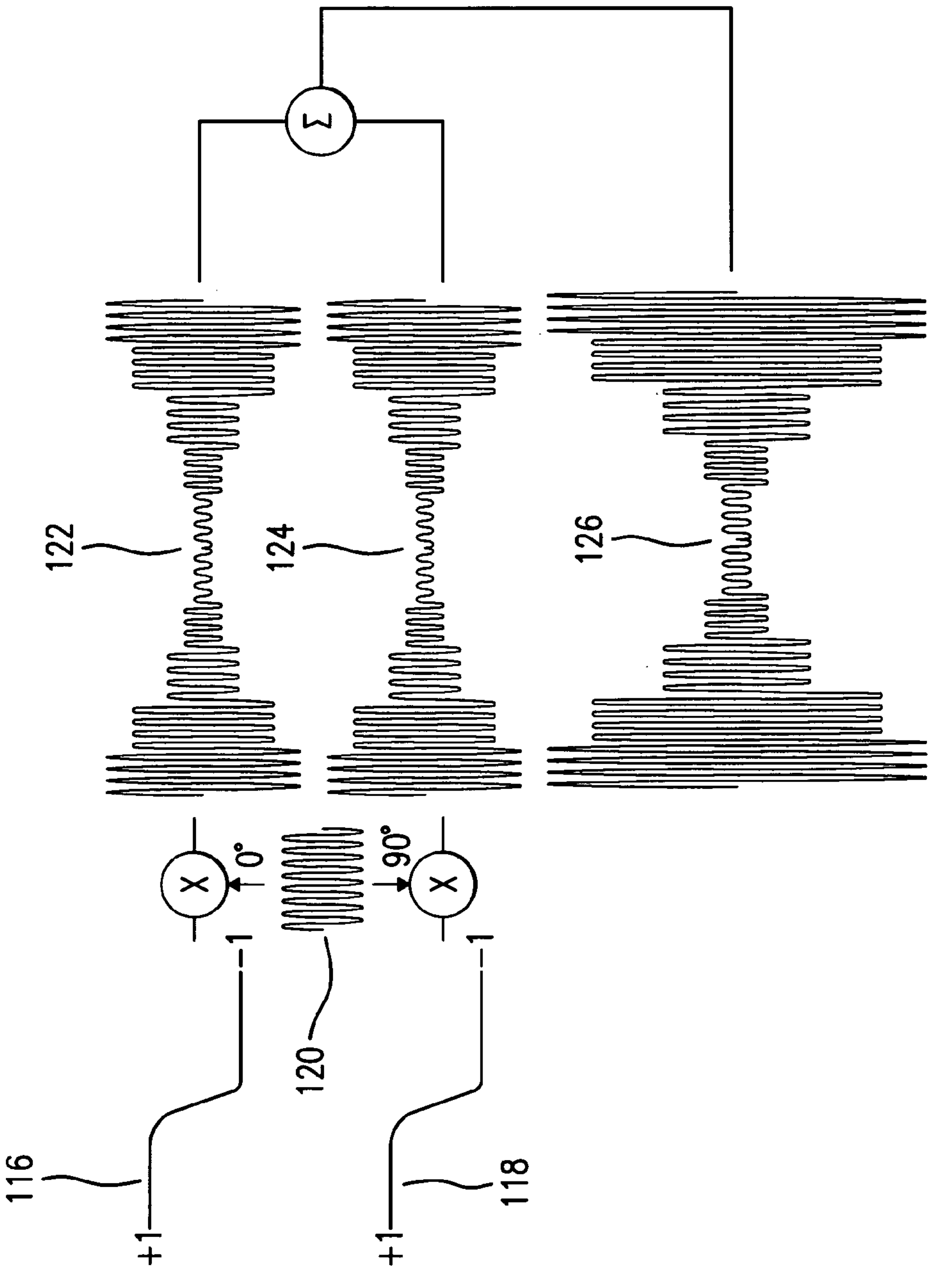
116
+1
-1
118
+1
-1
120
X
0°
X
90°
122
124
126
Σ

FIG.1B

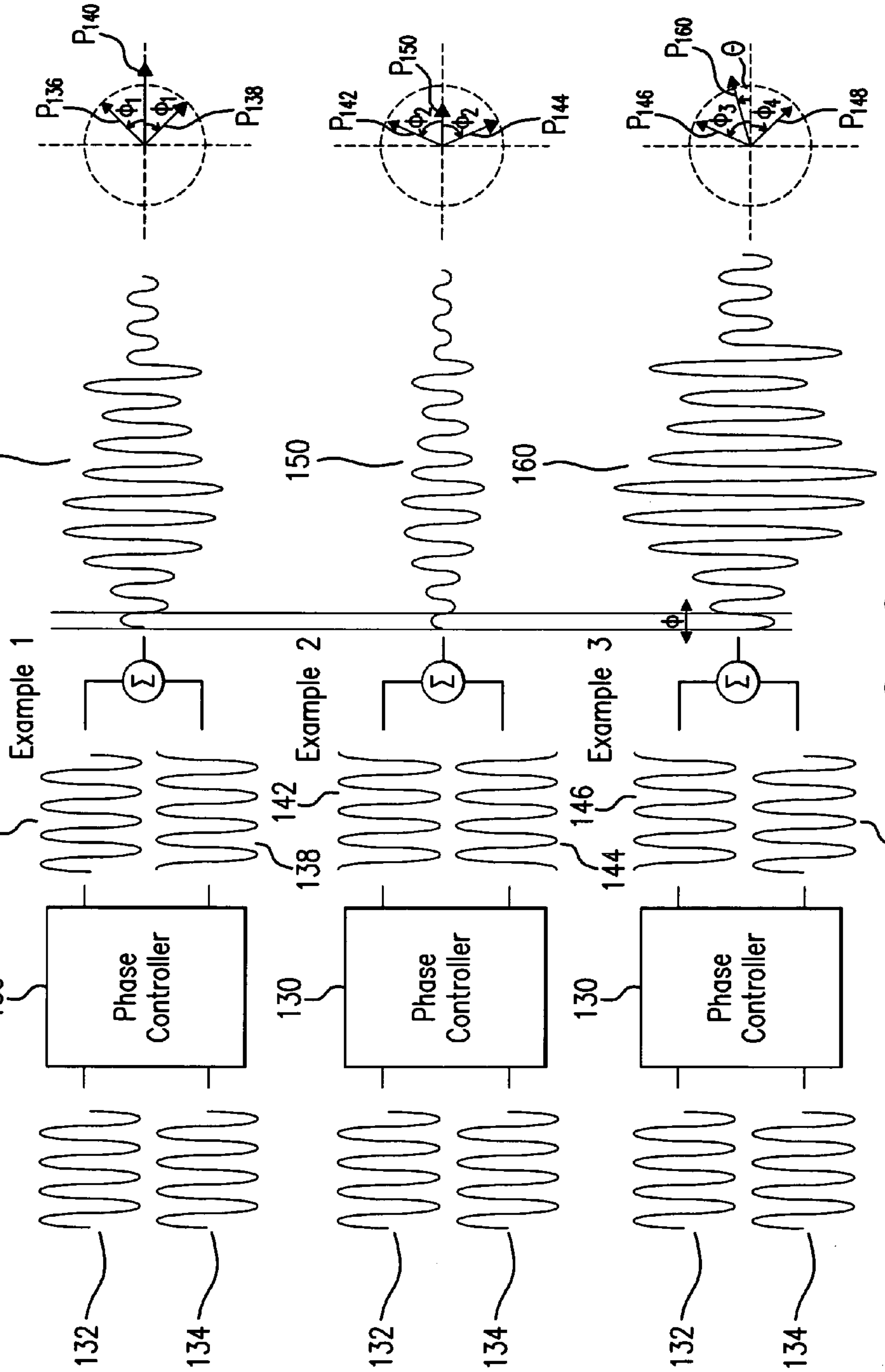
132
134
130
Phase Controller
136
138
142
144
146
148
Example 1
Example 2
Example 3
Σ
Φ
140
150
160
P_{136}
P_{138}
P_{140}
ϕ_1
P_{142}
P_{144}
P_{150}
ϕ_2
P_{146}
P_{148}
P_{160}
ϕ_3
ϕ_4
θ

FIG. 1C

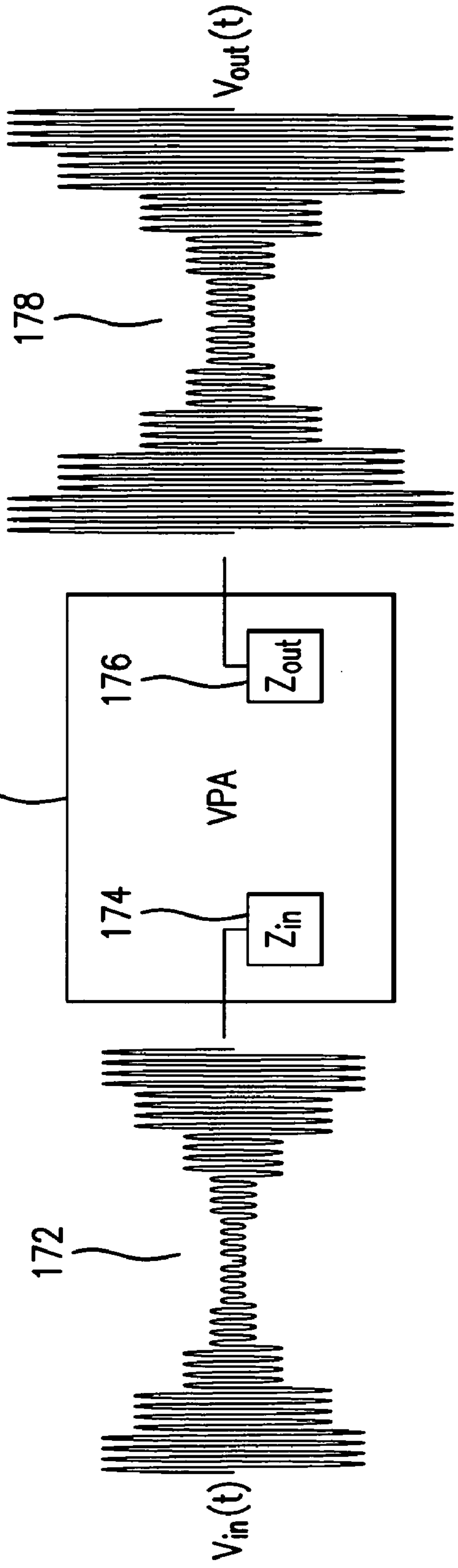
$V_{out}(t)$
178
176
Z_{out}
170
VPA
174
Z_{in}
172
$V_{in}(t)$

FIG.1D

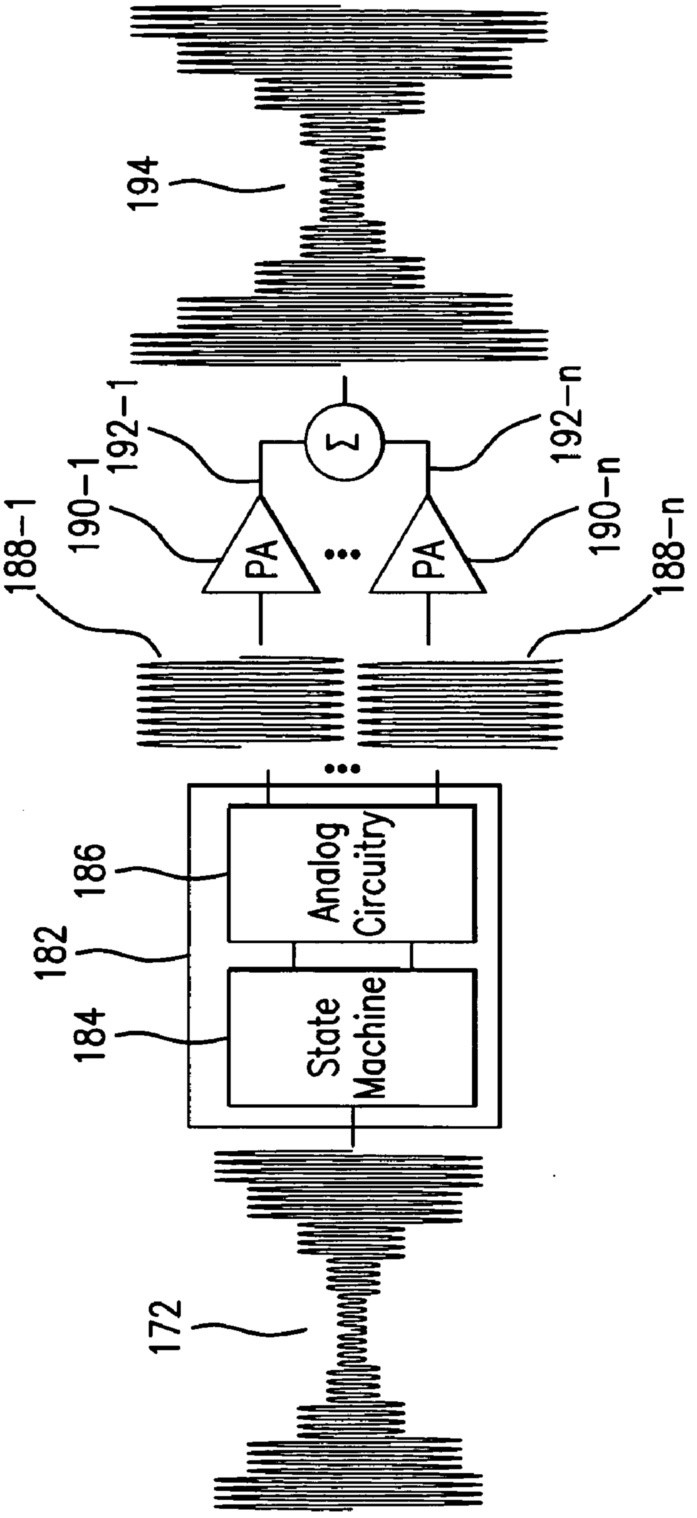
194
192-1
190-1
188-1
Σ
PA
PA
192-n
190-n
188-n
186
182
184
Analog Circuitry
State Machine
172

FIG. 1E

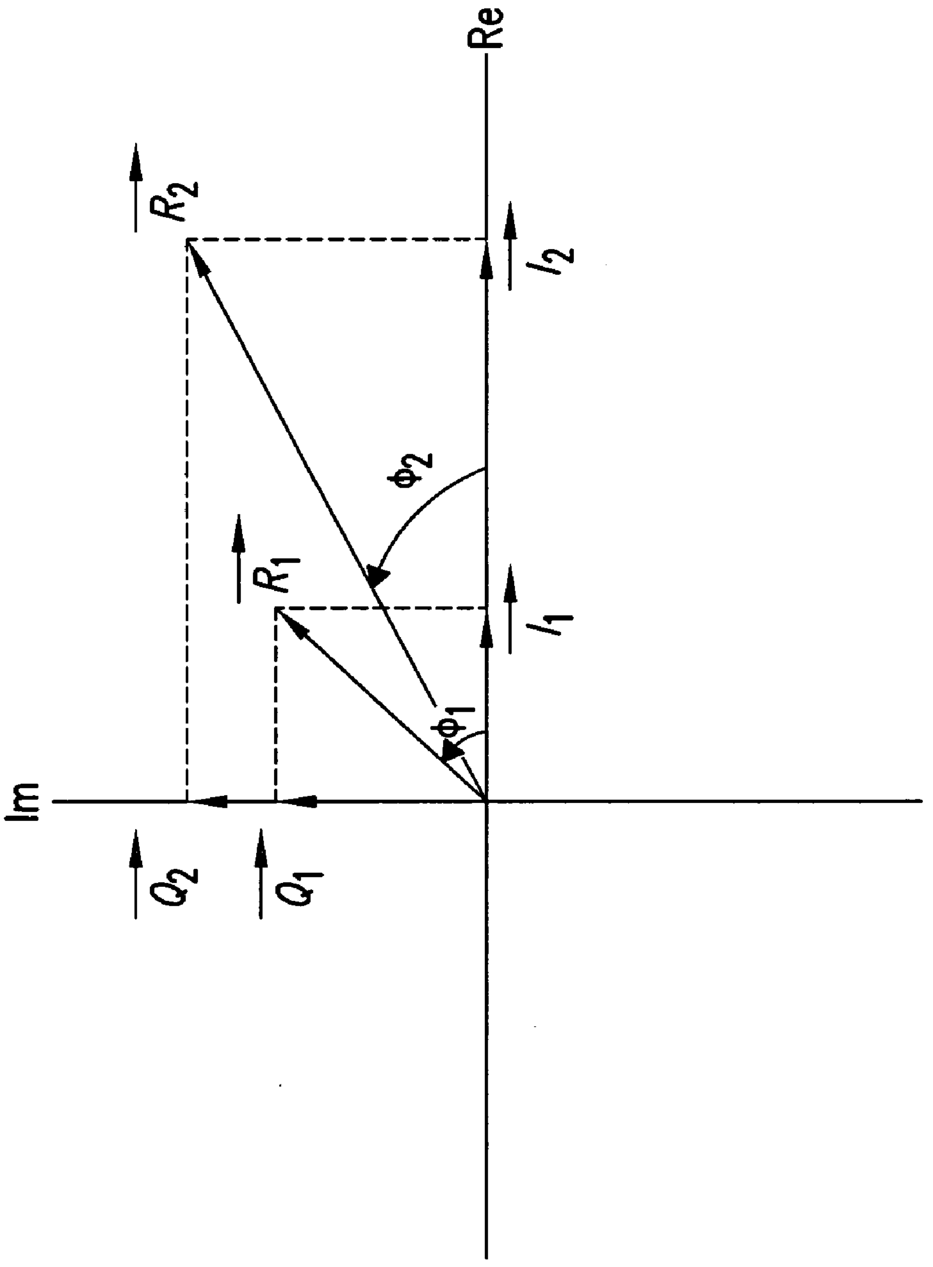
Re
Im
$\vec{R}_2$
$\vec{R}_1$
$\vec{I}_2$
$\vec{I}_1$
$\vec{Q}_2$
$\vec{Q}_1$
ϕ_2
ϕ_1

FIG.2

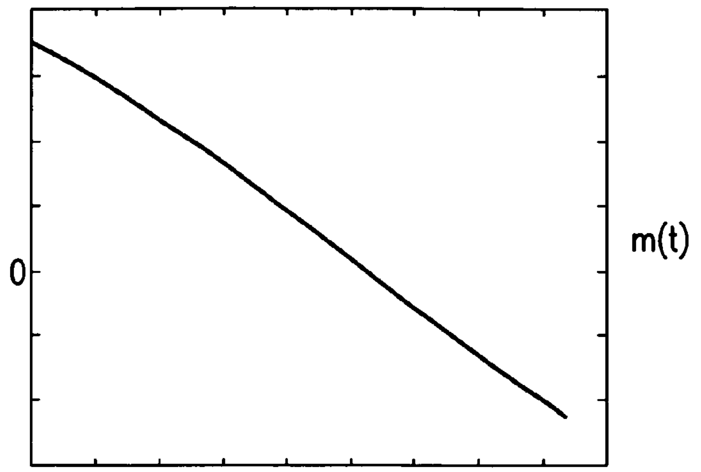

FIG.3A

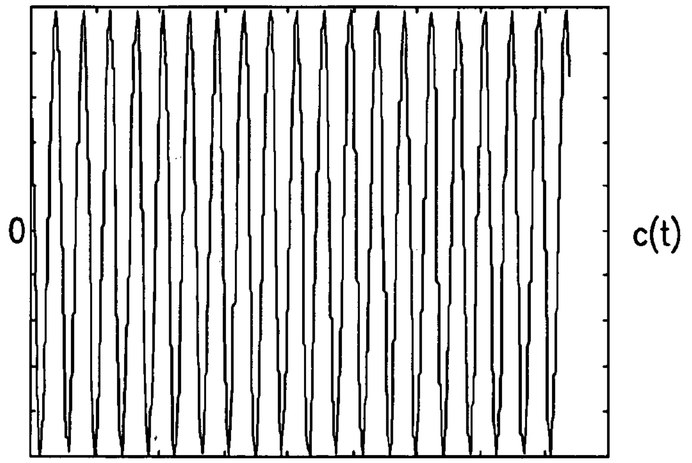

FIG.3B

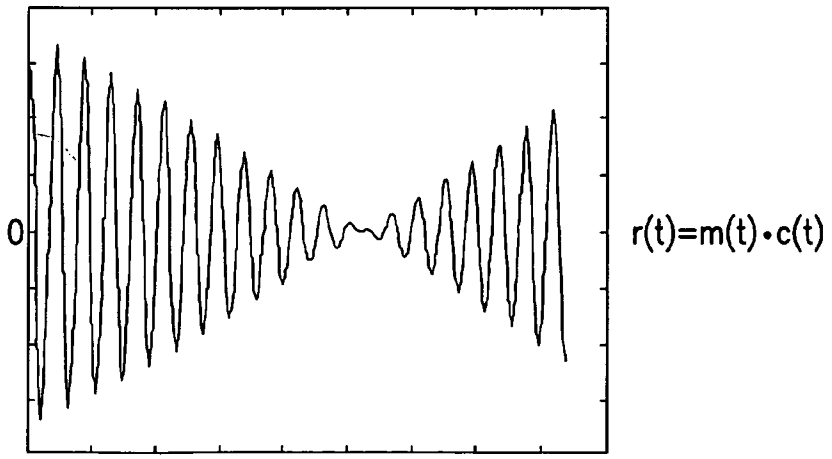

FIG.3C

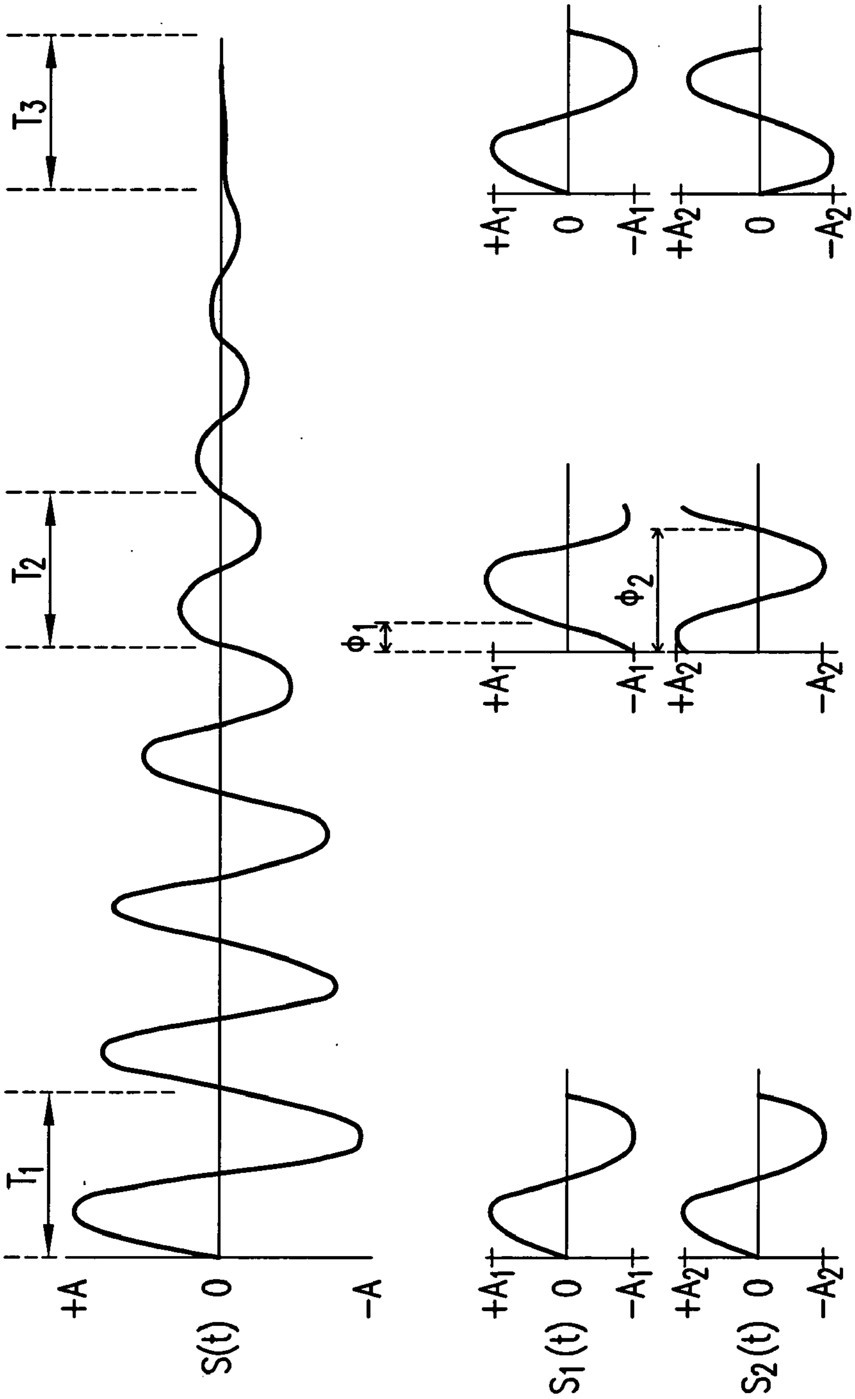
+A
S(t) 0
-A
T1
T2
T3
S1(t)
+A1
0
-A1
S2(t)
+A2
0
-A2
ϕ1
ϕ2

FIG.3D

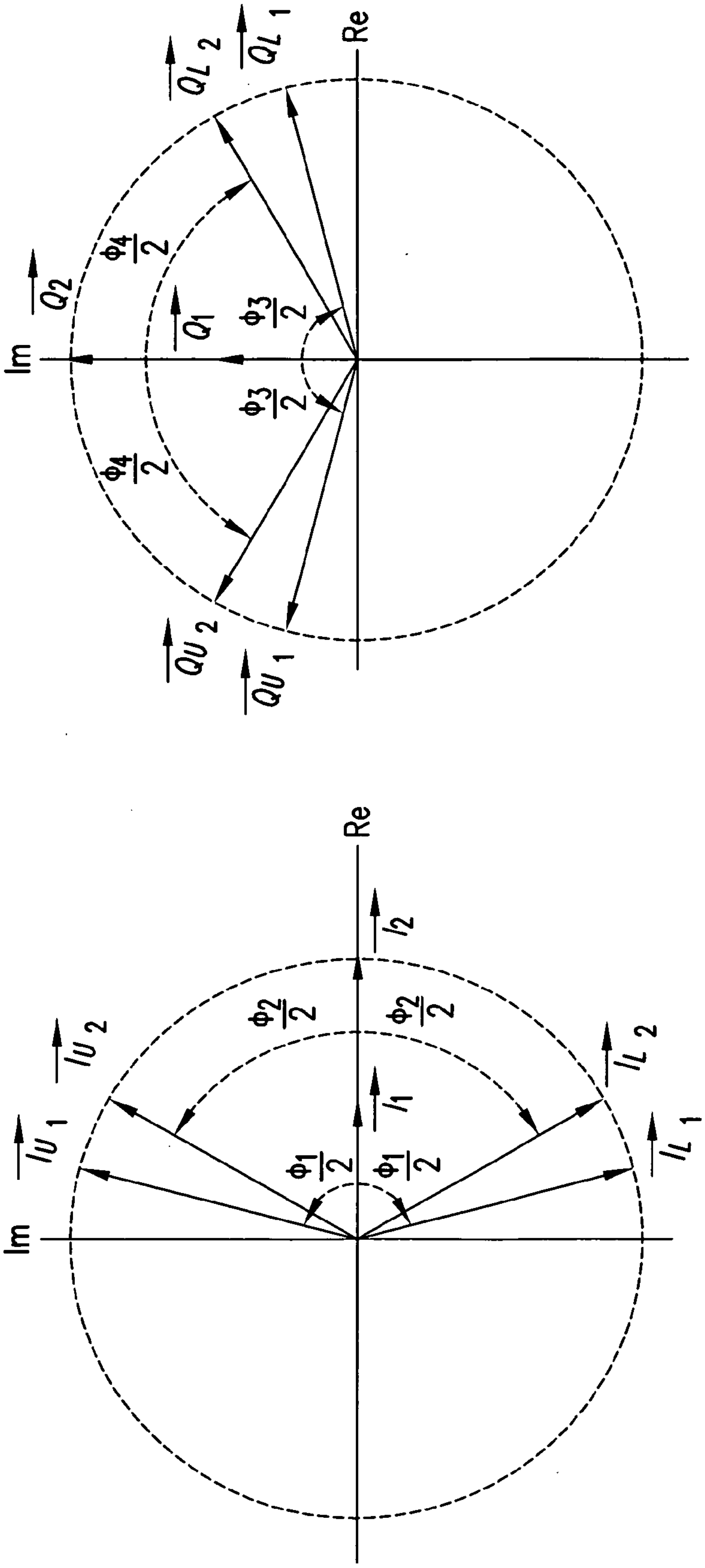
Im
Re
$\overrightarrow{Q_2}$
$\overrightarrow{Q_1}$
$\overrightarrow{QL_2}$
$\overrightarrow{QL_1}$
$\overrightarrow{QU_2}$
$\overrightarrow{QU_1}$
$\frac{\phi_3}{2}$
$\frac{\phi_3}{2}$
$\frac{\phi_4}{2}$
$\frac{\phi_4}{2}$
Im
Re
$\overrightarrow{IU_1}$
$\overrightarrow{IU_2}$
$\overrightarrow{I_1}$
$\overrightarrow{I_2}$
$\overrightarrow{IL_1}$
$\overrightarrow{IL_2}$
$\frac{\phi_1}{2}$
$\frac{\phi_1}{2}$
$\frac{\phi_2}{2}$
$\frac{\phi_2}{2}$

FIG.4

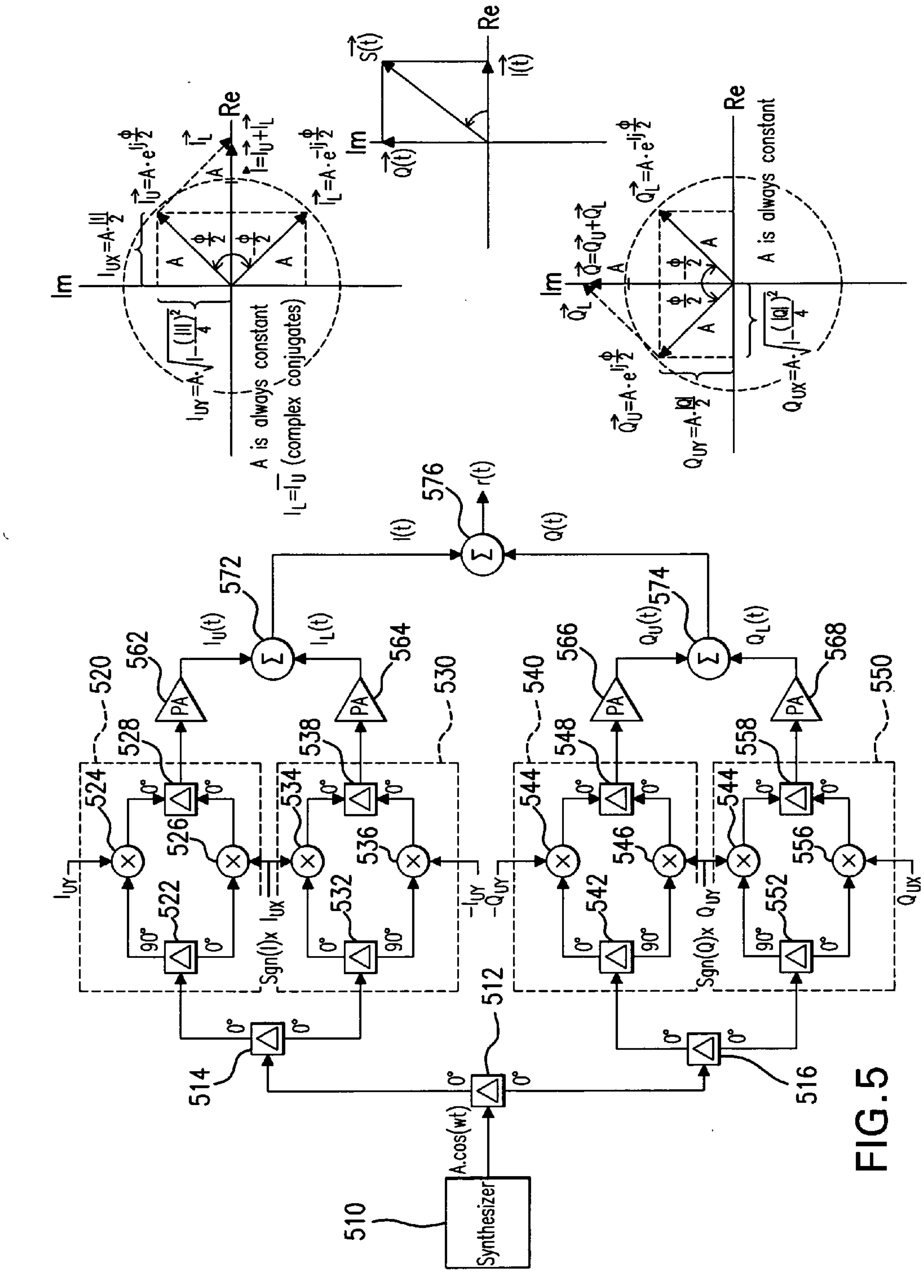
510
Synthesizer
A.cos(wt)
512
514
516
0°
90°
520
522
524
526
528
530
532
534
536
538
540
542
544
546
548
550
552
556
558
562
564
566
568
572
574
576
PA
Σ
I_{UY}
I_{UX}
$-I_{UY}$
$-Q_{UY}$
Q_{UY}
Q_{UX}
Sgn(I)x
Sgn(Q)x
$I_U(t)$
$I_L(t)$
$Q_U(t)$
$Q_L(t)$
I(t)
Q(t)
r(t)
Im
Re
$I_{UX}=A\cdot\frac{|I|}{2}$
$\vec{I}_U=A\cdot e^{j\frac{\phi}{2}}$
$\vec{I}_L$
$\vec{I}=\vec{I}_U+\vec{I}_L$
$\vec{I}_L=A\cdot e^{-j\frac{\phi}{2}}$
$I_{UY}=A\cdot\sqrt{1-\frac{(|I|)^2}{4}}$
A is always constant
$I_L=\overline{I_U}$ (complex conjugates)
$\frac{\phi}{2}$
$-\frac{\phi}{2}$
A
$\vec{S}(t)$
$\vec{I}(t)$
$\vec{Q}(t)$
$\vec{Q}_L$
$\vec{Q}=\vec{Q}_U+\vec{Q}_L$
$\vec{Q}_L=A\cdot e^{-j\frac{\phi}{2}}$
$\vec{Q}_U=A\cdot e^{j\frac{\phi}{2}}$
$Q_{UY}=A\cdot\frac{|Q|}{2}$
$Q_{UX}=A\cdot\sqrt{1-\frac{(|Q|)^2}{4}}$
A is always constant

FIG.5

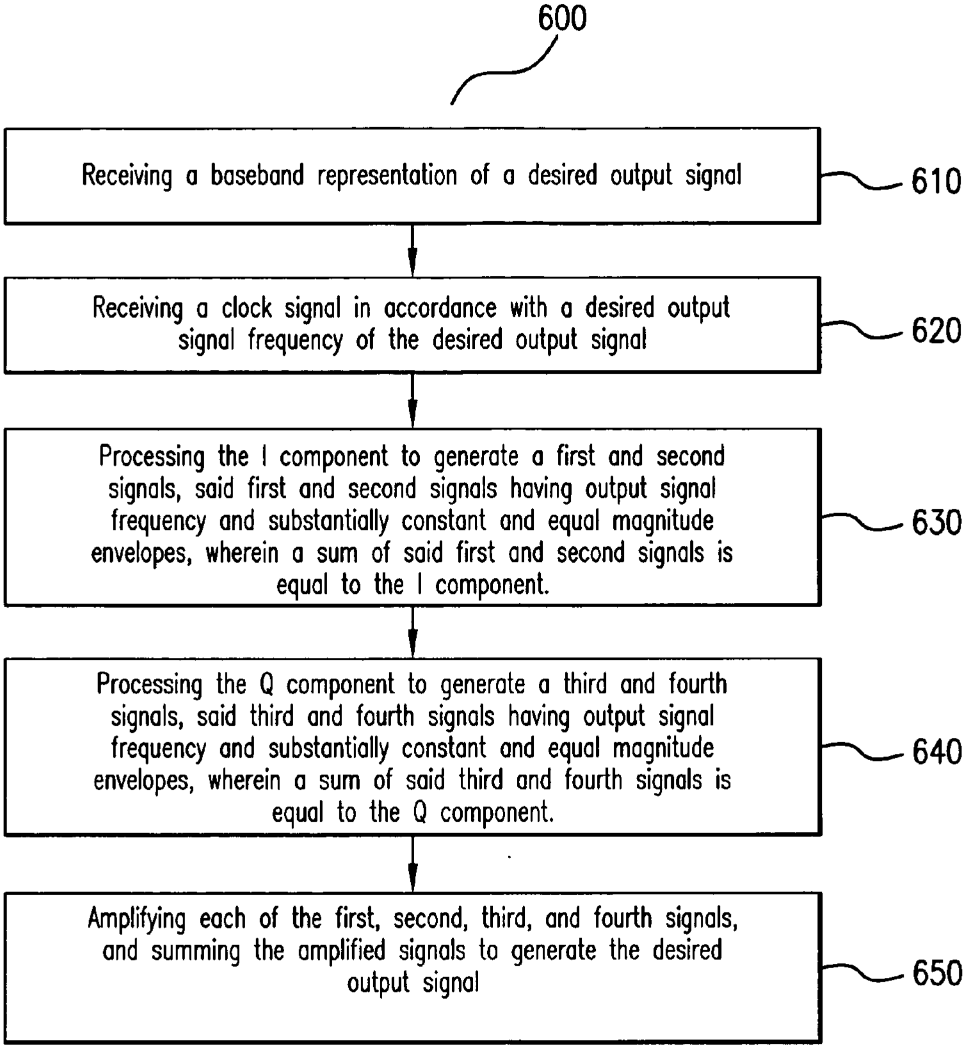

Receiving a baseband representation of a desired output signal

Receiving a clock signal in accordance with a desired output signal frequency of the desired output signal

Processing the I component to generate a first and second signals, said first and second signals having output signal frequency and substantially constant and equal magnitude envelopes, wherein a sum of said first and second signals is equal to the I component.

Processing the Q component to generate a third and fourth signals, said third and fourth signals having output signal frequency and substantially constant and equal magnitude envelopes, wherein a sum of said third and fourth signals is equal to the Q component.

Amplifying each of the first, second, third, and fourth signals, and summing the amplified signals to generate the desired output signal

FIG.6

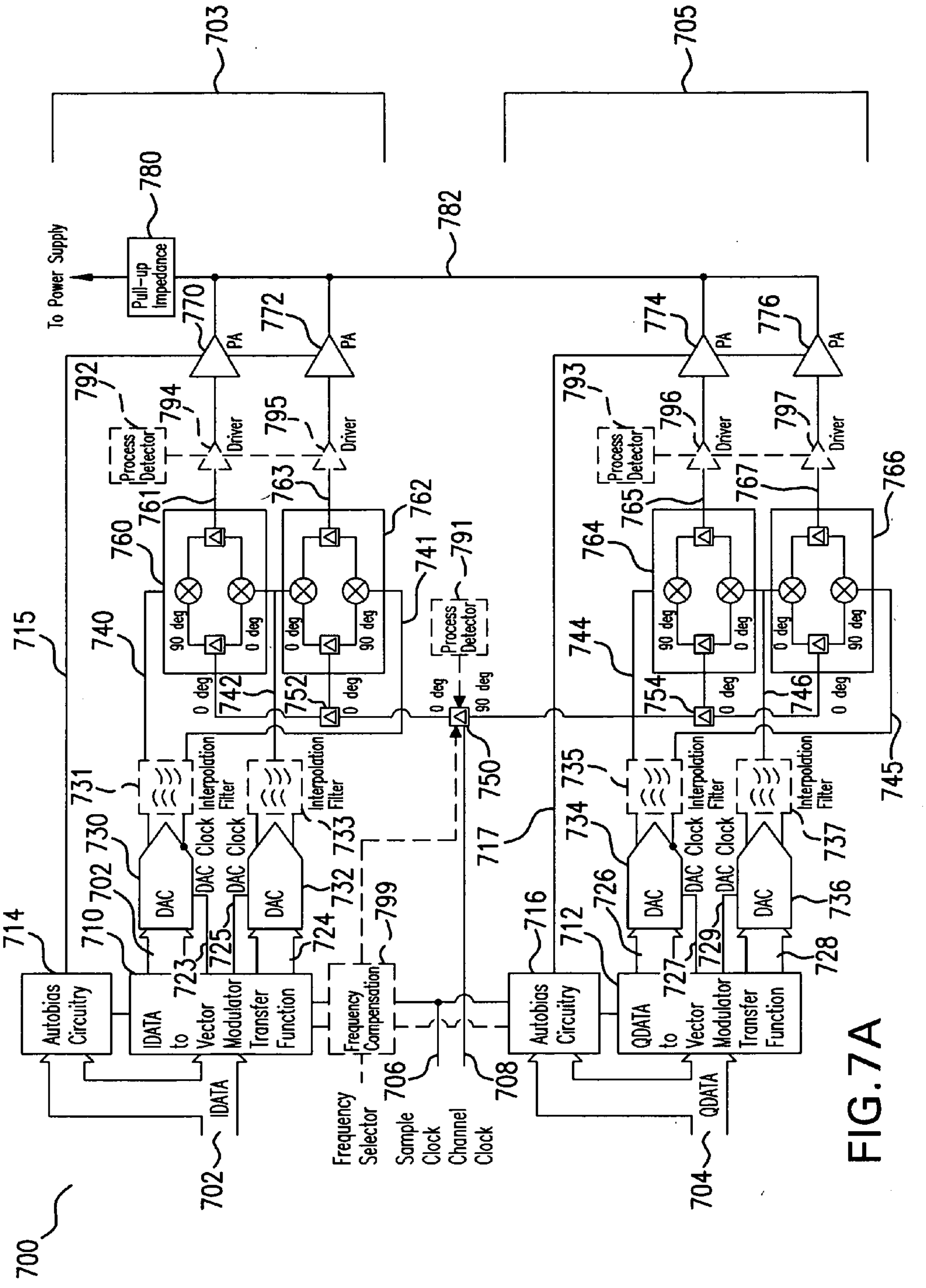
IDATA
QDATA
Frequency Selector
Sample Clock
Channel Clock
Autobias Circuitry
IDATA to Vector Modulator Transfer Function
QDATA to Vector Modulator Transfer Function
Frequency Compensation
DAC
DAC Clock
Interpolation Filter
0 deg
90 deg
Process Detector
Driver
PA
Pull-up Impedance
To Power Supply

FIG. 7A

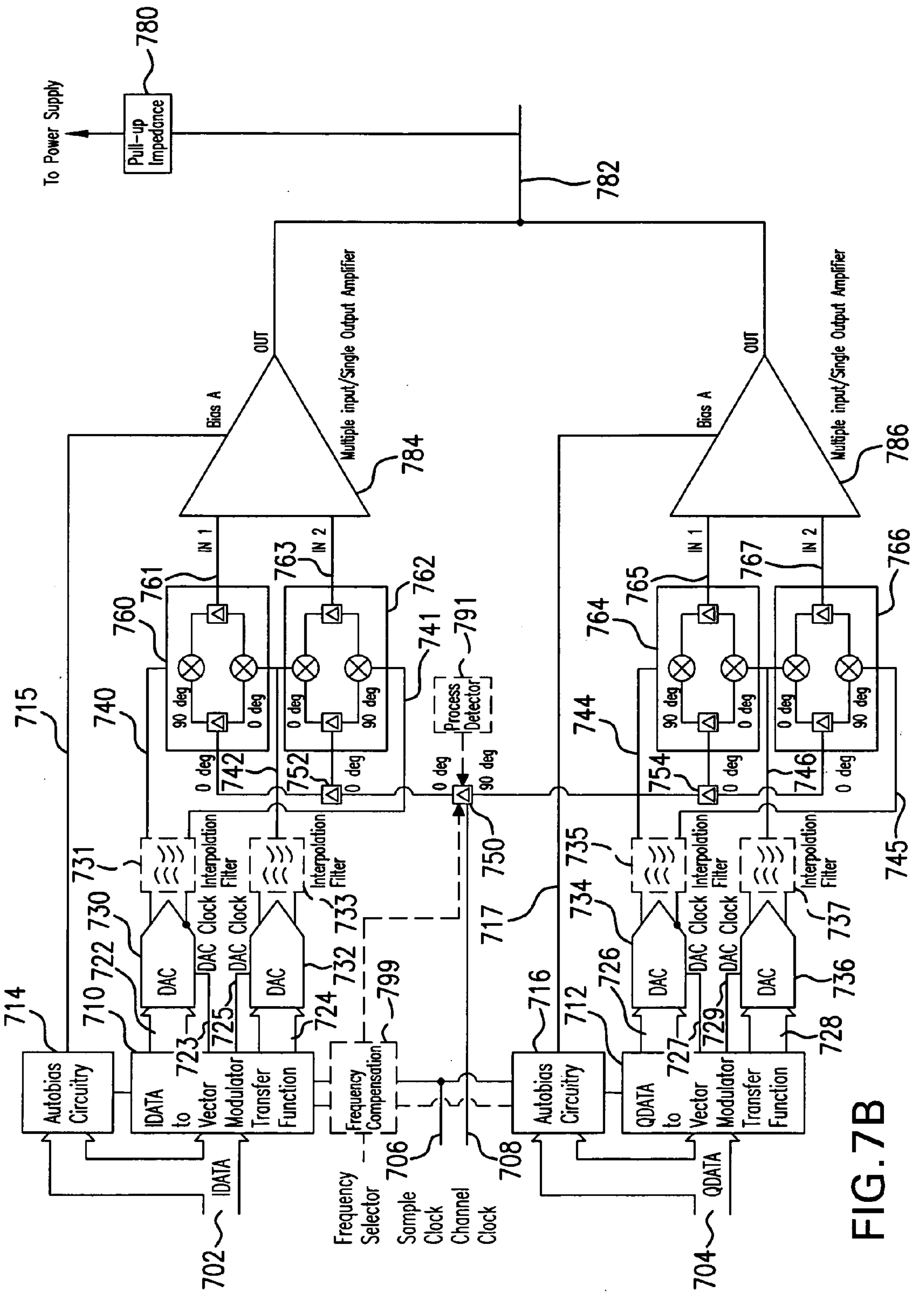
To Power Supply
Pull-up Impedance
Multiple input/Single Output Amplifier
OUT
Bias A
IN 1
IN 2
Autobias Circuitry
IDATA to Vector Modulator Transfer Function
QDATA to Vector Modulator Transfer Function
IDATA
QDATA
DAC
DAC Clock
Interpolation Filter
0 deg
90 deg
Process Detector
Frequency Compensation
Frequency Selector
Sample Clock
Channel Clock

FIG. 7B

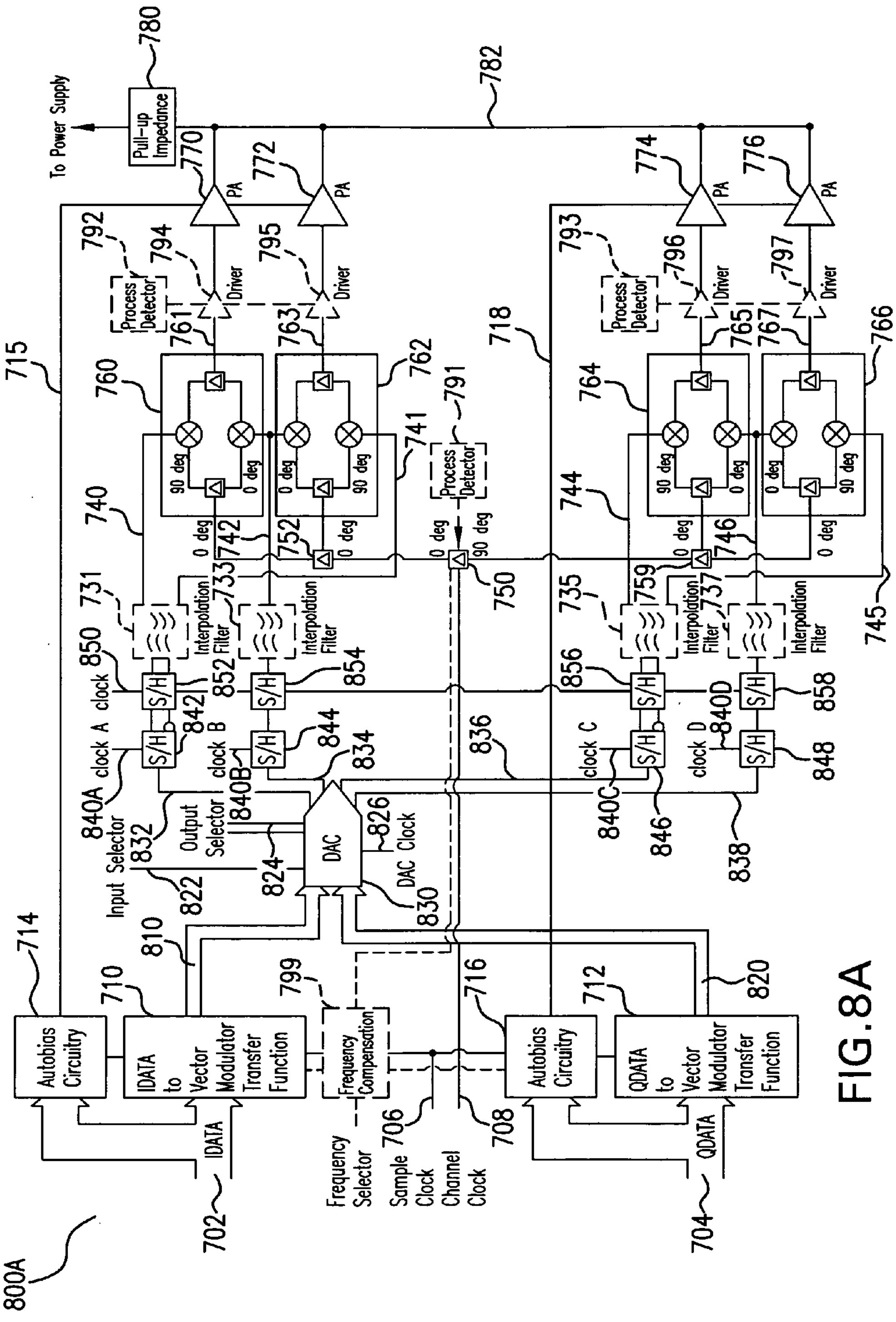
IDATA
QDATA
Frequency Selector
Sample Clock
Channel Clock
Autobias Circuitry
IDATA to Vector Modulator Transfer Function
Autobias Circuitry
QDATA to Vector Modulator Transfer Function
Frequency Compensation
Input Selector
Output Selector
DAC
DAC Clock
clock A
clock B
clock C
clock D
clock
S/H
Interpolation Filter
0 deg
90 deg
Process Detector
Process Detector
Driver
PA
Pull-up Impedance
To Power Supply

FIG.8A

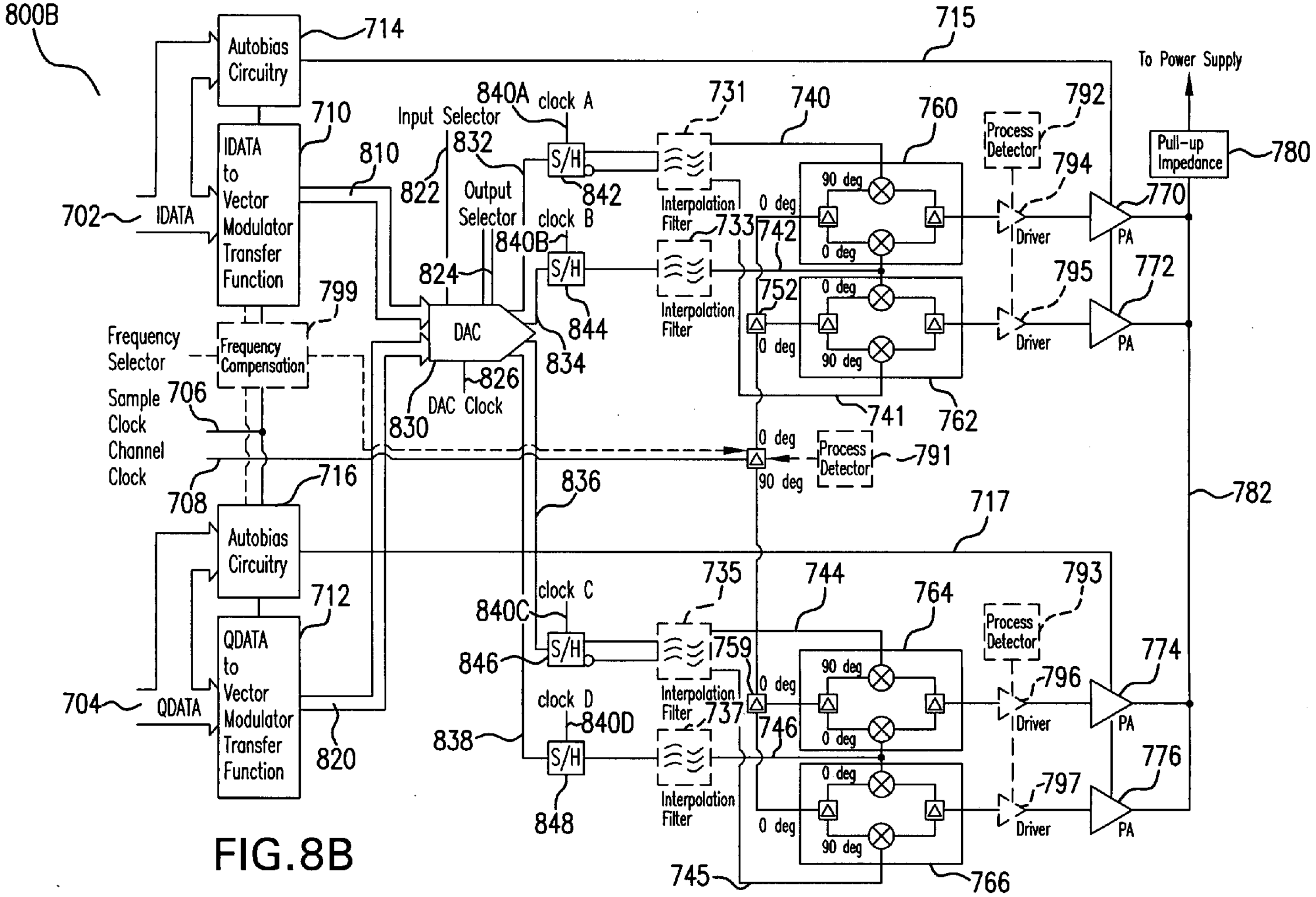
IDATA
QDATA
Autobias Circuitry
IDATA to Vector Modulator Transfer Function
Autobias Circuitry
QDATA to Vector Modulator Transfer Function
Frequency Selector
Frequency Compensation
Sample Clock
Channel Clock
Input Selector
Output Selector
DAC
DAC Clock
clock A
S/H
clock B
clock C
clock D
Interpolation Filter
0 deg
90 deg
Process Detector
Driver
PA
Pull-up Impedance
To Power Supply

FIG.8B

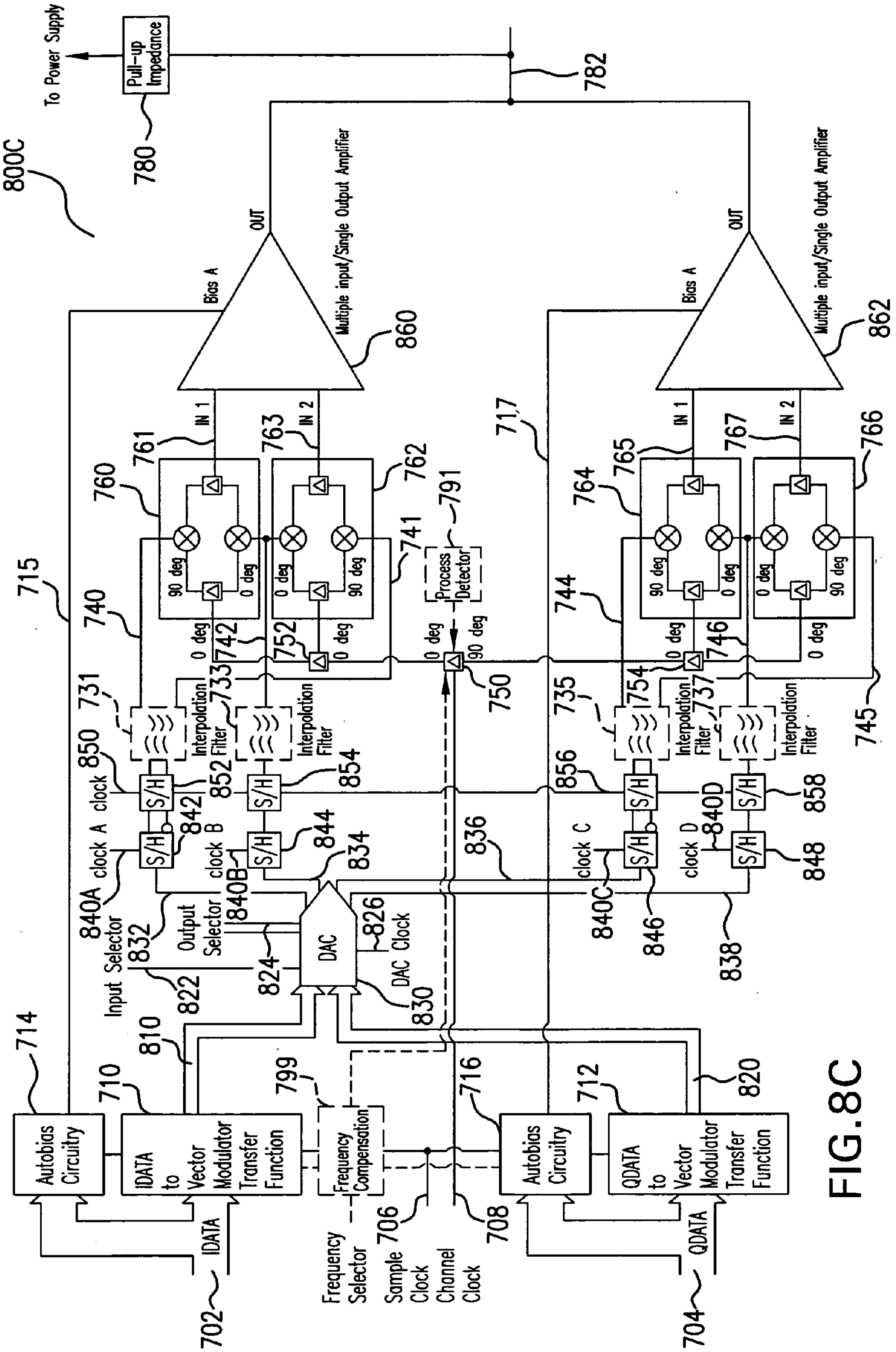
To Power Supply
Pull-up Impedance
OUT
Multiple input/Single Output Amplifier
Bias A
IN 1
IN 2
90 deg
0 deg
Process Detector
Interpolation Filter
S/H
clock A
clock B
clock C
clock D
clock
Output Selector
Input Selector
DAC
DAC Clock
Autobias Circuitry
IDATA to Vector Modulator Transfer Function
QDATA to Vector Modulator Transfer Function
Frequency Compensation
IDATA
QDATA
Frequency Selector
Sample Clock
Channel Clock

FIG.8C

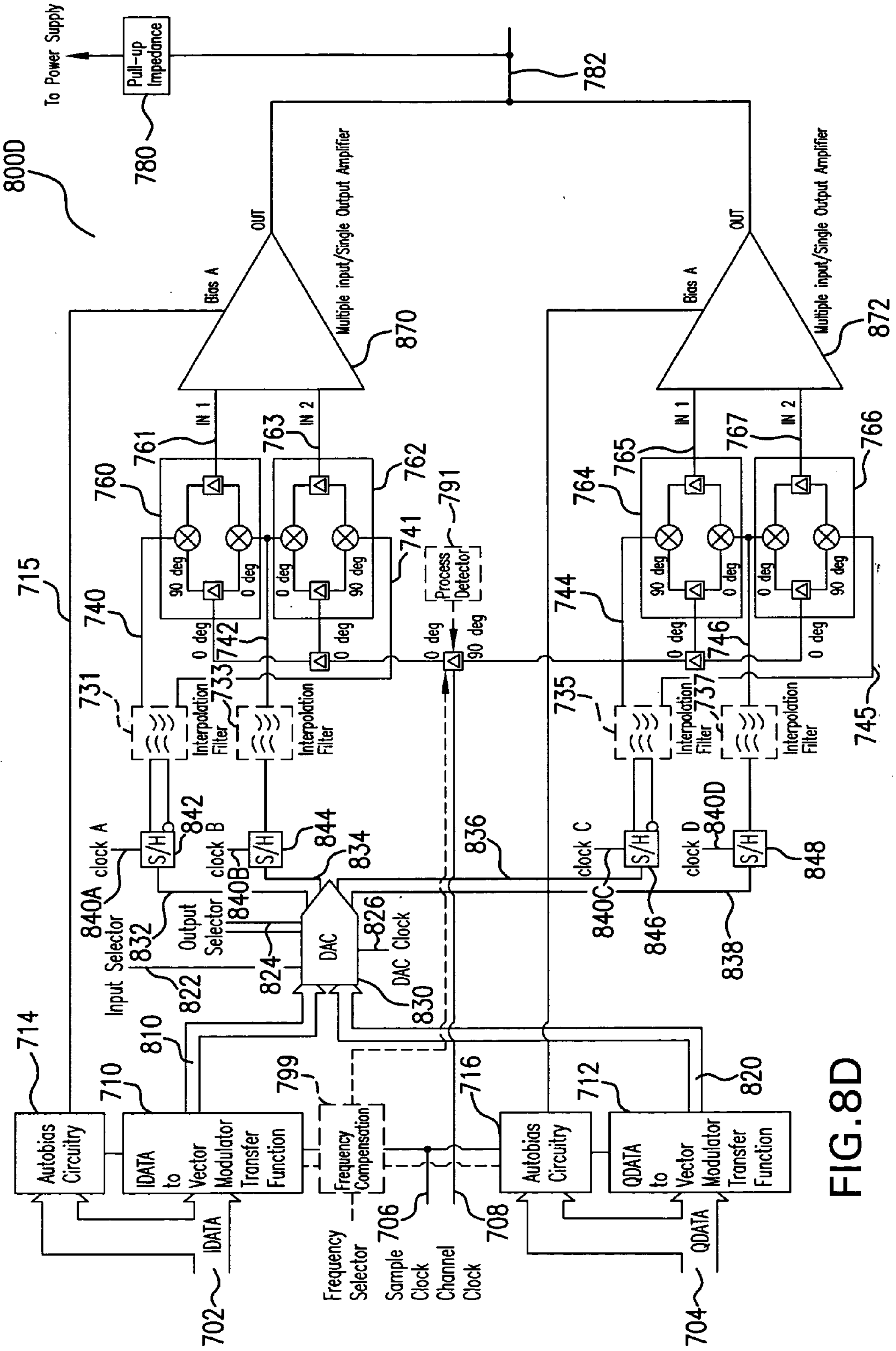
To Power Supply
Pull-up Impedance
OUT
Multiple input/Single Output Amplifier
Bias A
IN 1
IN 2
90 deg
0 deg
Process Detector
Interpolation Filter
clock A
clock B
clock C
clock D
S/H
Input Selector
Output Selector
DAC
DAC Clock
Autobias Circuitry
IDATA to Vector Modulator Transfer Function
QDATA to Vector Modulator Transfer Function
Frequency Compensation
Frequency Selector
Sample Clock
Channel Clock
IDATA
QDATA

FIG.8D

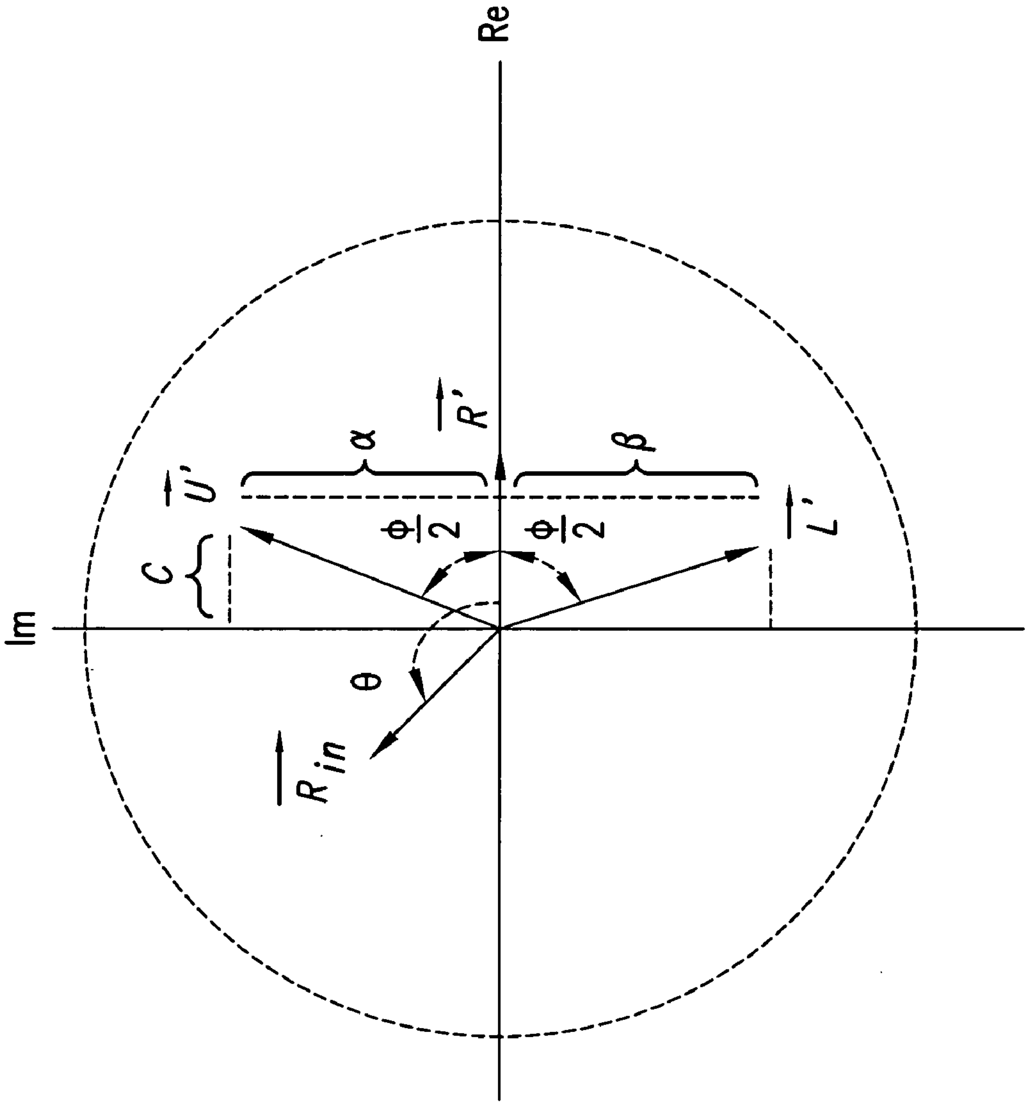
Re
Im
$\overrightarrow{R'}$
$\overrightarrow{U'}$
$\overrightarrow{L'}$
α
β
$\frac{\phi}{2}$
$\frac{\phi}{2}$
C
θ
$\overrightarrow{R_{in}}$

FIG.9A

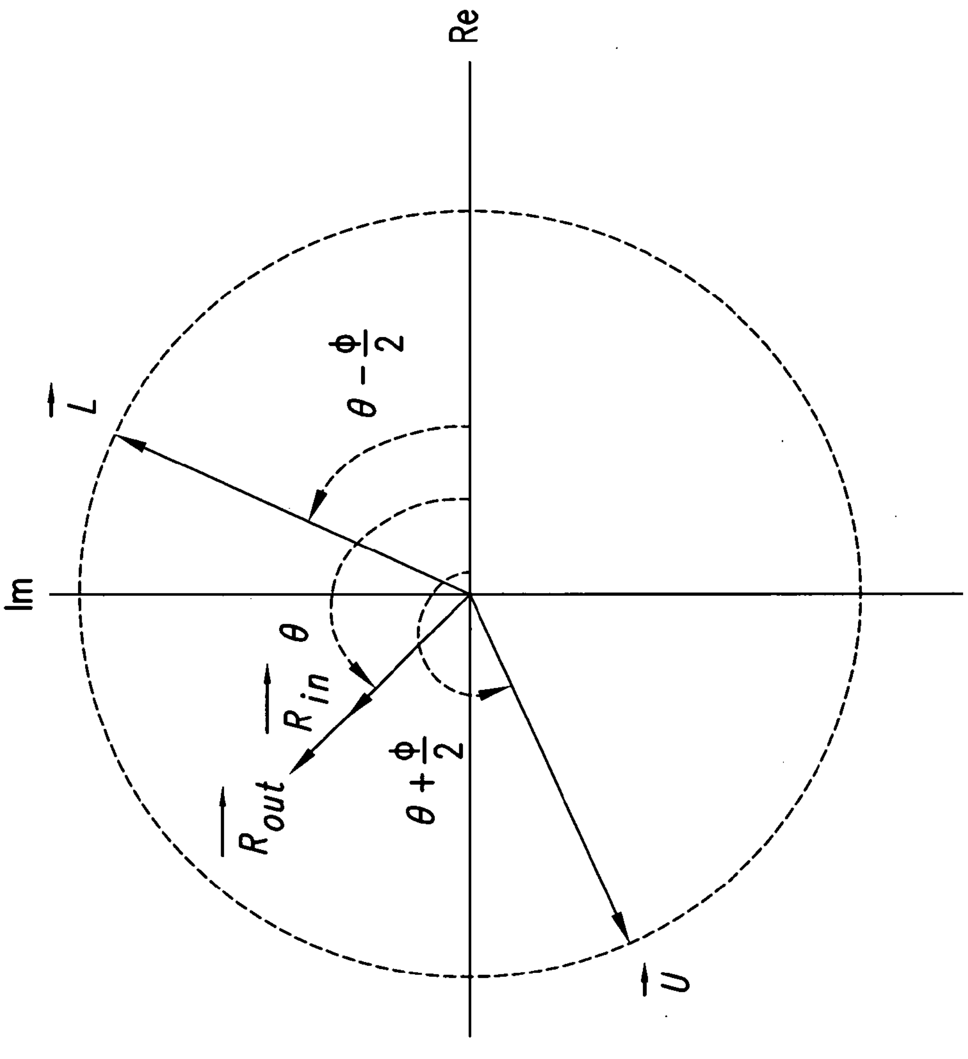
Re
Im
$\vec{L}$
$\theta - \frac{\Phi}{2}$
θ
$\vec{R_{in}}$
$\vec{R_{out}}$
$\theta + \frac{\Phi}{2}$
$\vec{U}$

FIG.9B

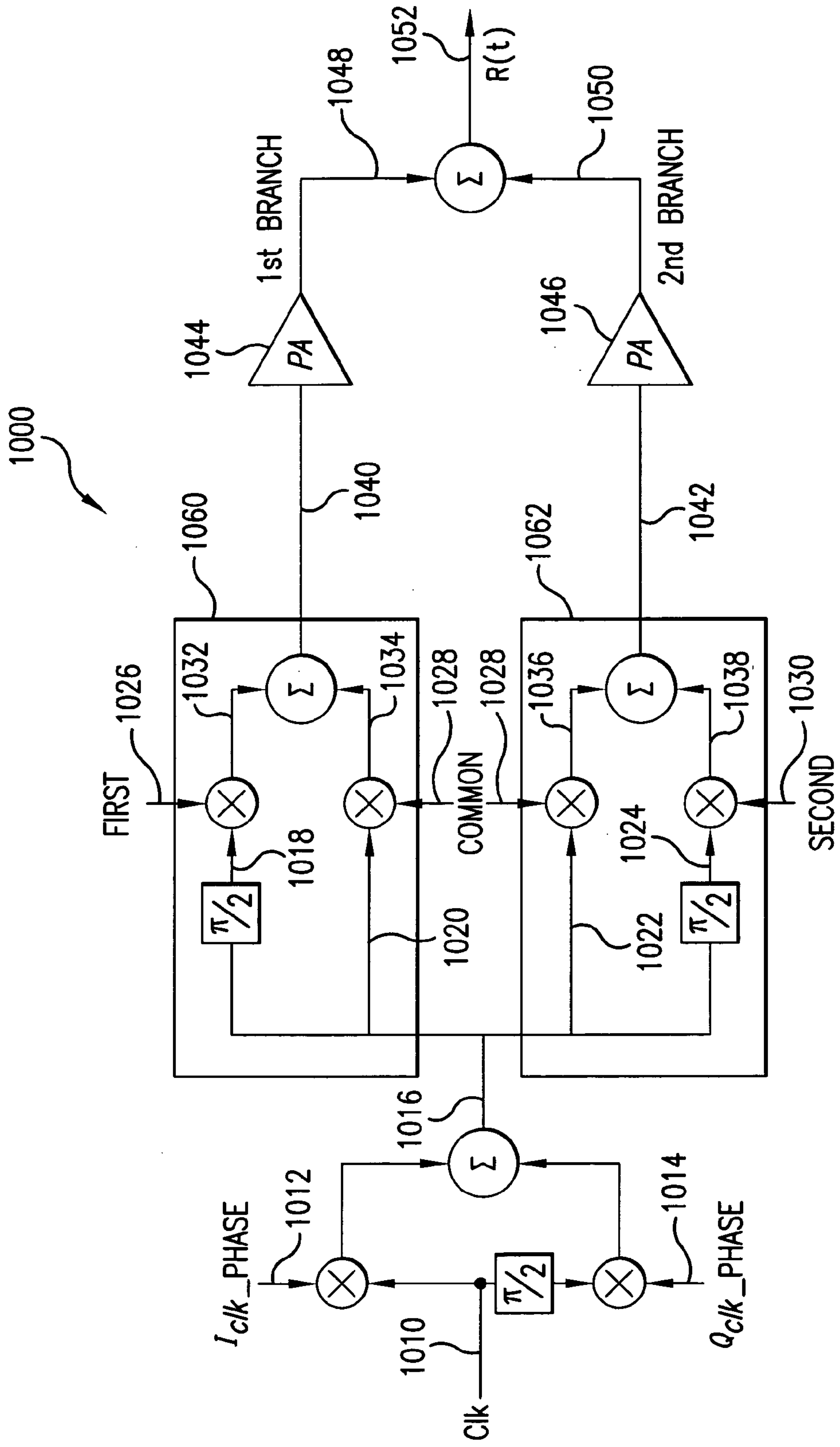
1000
I_{clk}_PHASE
1012
Clk
1010
$\pi/2$
Q_{clk}_PHASE
1014
Σ
1016
FIRST
1026
1018
$\pi/2$
1020
1032
Σ
1034
1060
COMMON
1028
1028
1036
Σ
1038
1024
1022
$\pi/2$
SECOND
1030
1062
1040
PA
1044
1042
PA
1046
1st BRANCH
2nd BRANCH
1048
1050
Σ
R(t)
1052

FIG.10

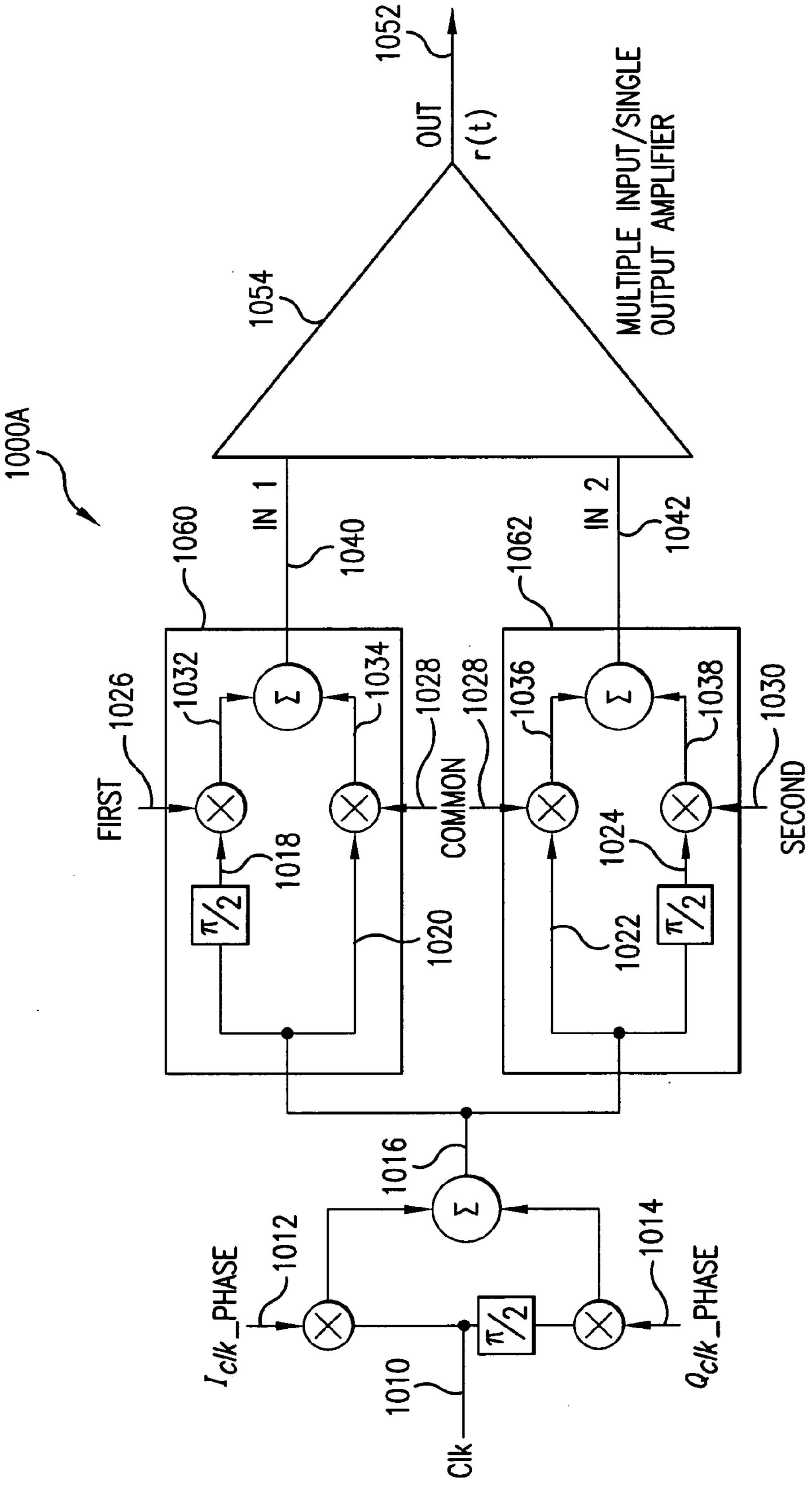
1000A
1052
OUT
r(t)
MULTIPLE INPUT/SINGLE OUTPUT AMPLIFIER
1054
IN 1
1040
IN 2
1042
1060
1062
FIRST
1026
1032
Σ
1034
1028
COMMON
1028
1036
Σ
1038
1030
SECOND
1018
π/2
1020
1024
1022
π/2
1016
Σ
1012
1014
I_{clk}_PHASE
Q_{clk}_PHASE
π/2
1010
Clk

FIG.10A

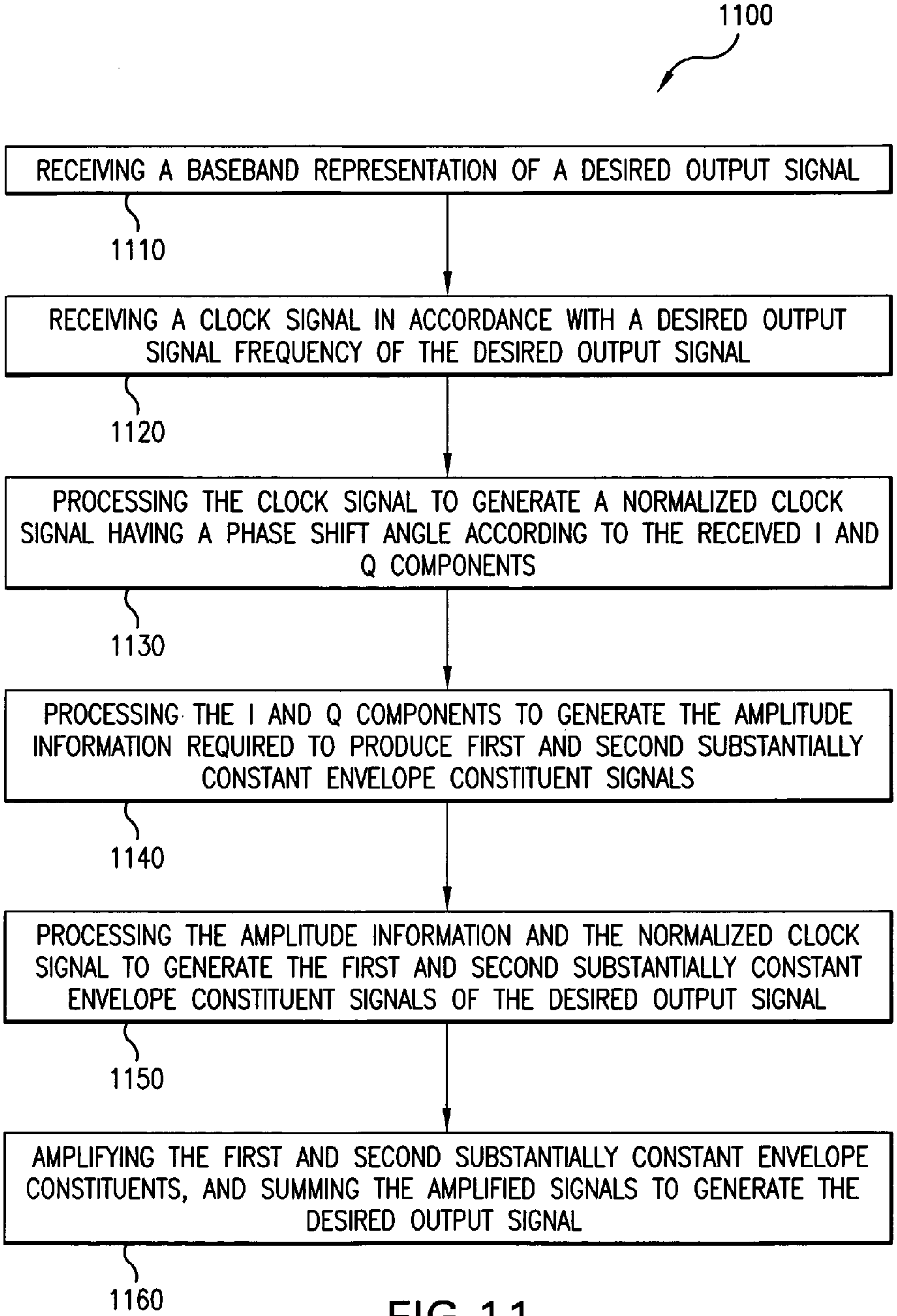
1100
RECEIVING A BASEBAND REPRESENTATION OF A DESIRED OUTPUT SIGNAL
1110
RECEIVING A CLOCK SIGNAL IN ACCORDANCE WITH A DESIRED OUTPUT SIGNAL FREQUENCY OF THE DESIRED OUTPUT SIGNAL
1120
PROCESSING THE CLOCK SIGNAL TO GENERATE A NORMALIZED CLOCK SIGNAL HAVING A PHASE SHIFT ANGLE ACCORDING TO THE RECEIVED I AND Q COMPONENTS
1130
PROCESSING THE I AND Q COMPONENTS TO GENERATE THE AMPLITUDE INFORMATION REQUIRED TO PRODUCE FIRST AND SECOND SUBSTANTIALLY CONSTANT ENVELOPE CONSTITUENT SIGNALS
1140
PROCESSING THE AMPLITUDE INFORMATION AND THE NORMALIZED CLOCK SIGNAL TO GENERATE THE FIRST AND SECOND SUBSTANTIALLY CONSTANT ENVELOPE CONSTITUENT SIGNALS OF THE DESIRED OUTPUT SIGNAL
1150
AMPLIFYING THE FIRST AND SECOND SUBSTANTIALLY CONSTANT ENVELOPE CONSTITUENTS, AND SUMMING THE AMPLIFIED SIGNALS TO GENERATE THE DESIRED OUTPUT SIGNAL
1160

FIG. 11

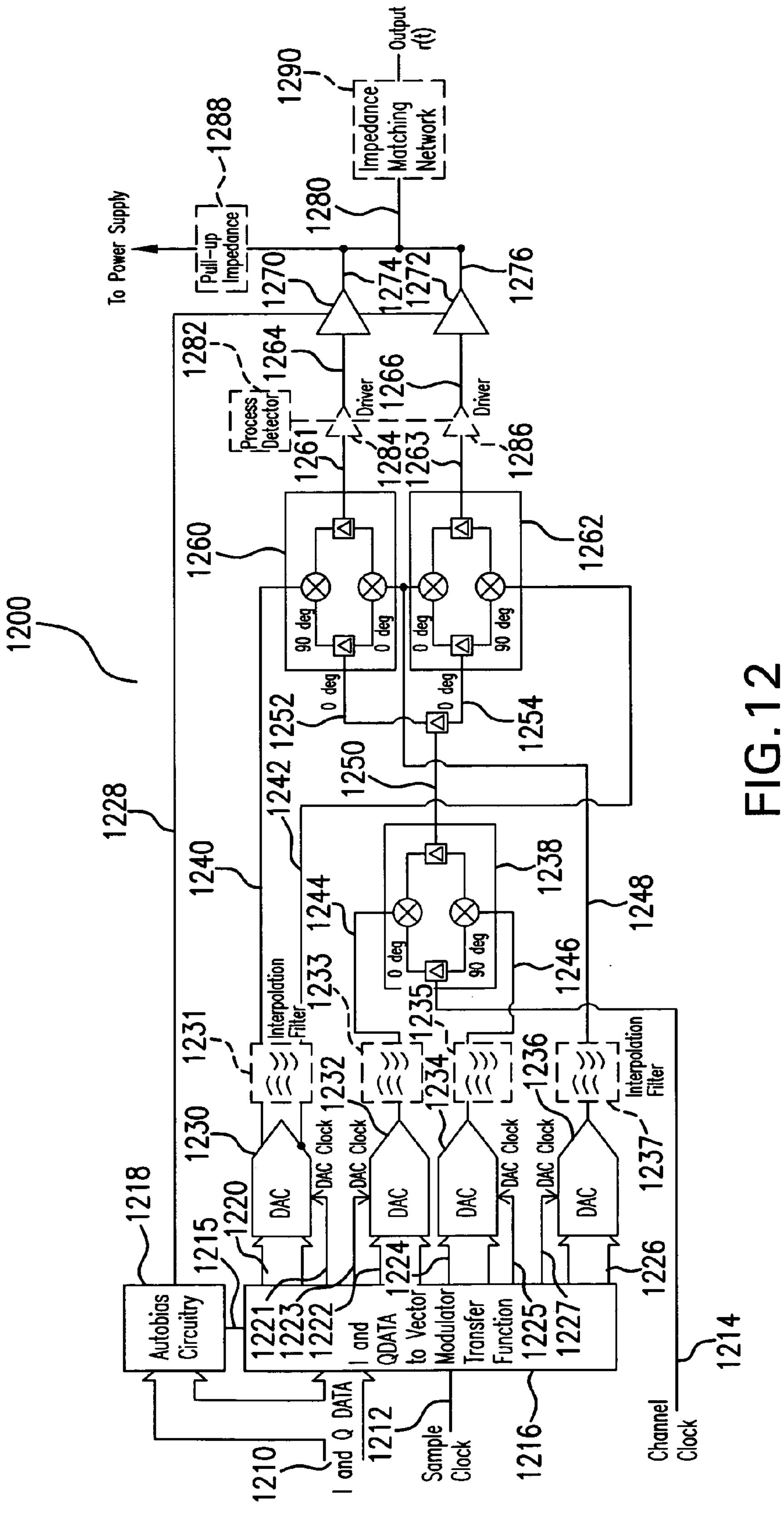
I and Q DATA
Sample Clock
Channel Clock
Autobias Circuitry
I and QDATA to Vector Modulator Transfer Function
DAC
DAC Clock
Interpolation Filter
0 deg
90 deg
Driver
Process Detector
Pull-up Impedance
To Power Supply
Impedance Matching Network
Output r(t)

FIG. 12

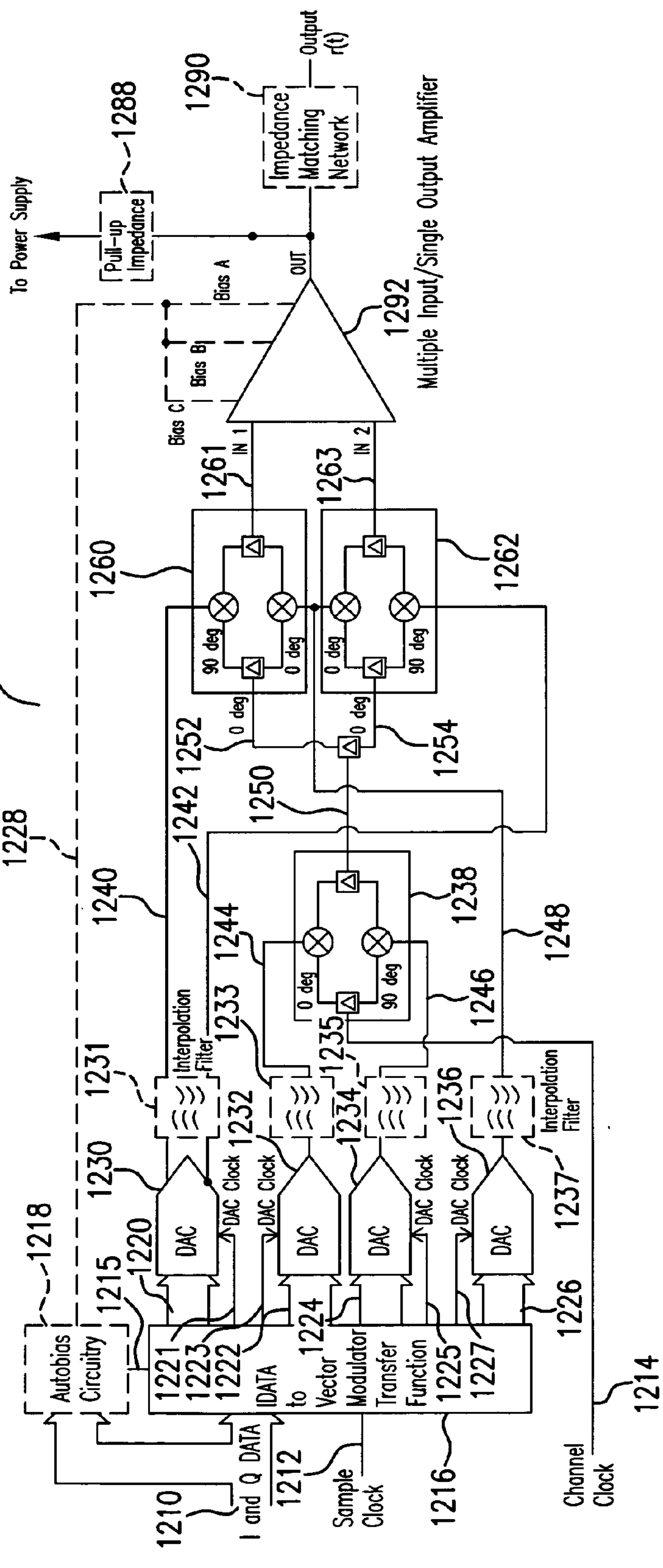
I and Q DATA
Sample Clock
Channel Clock
Autobias Circuitry
IDATA to Vector Modulator Transfer Function
DAC
DAC Clock
Interpolation Filter
0 deg
90 deg
IN 1
IN 2
Bias A
Bias B
Bias C
OUT
Multiple Input/Single Output Amplifier
Pull-up Impedance
To Power Supply
Impedance Matching Network
Output r(t)

FIG.12A

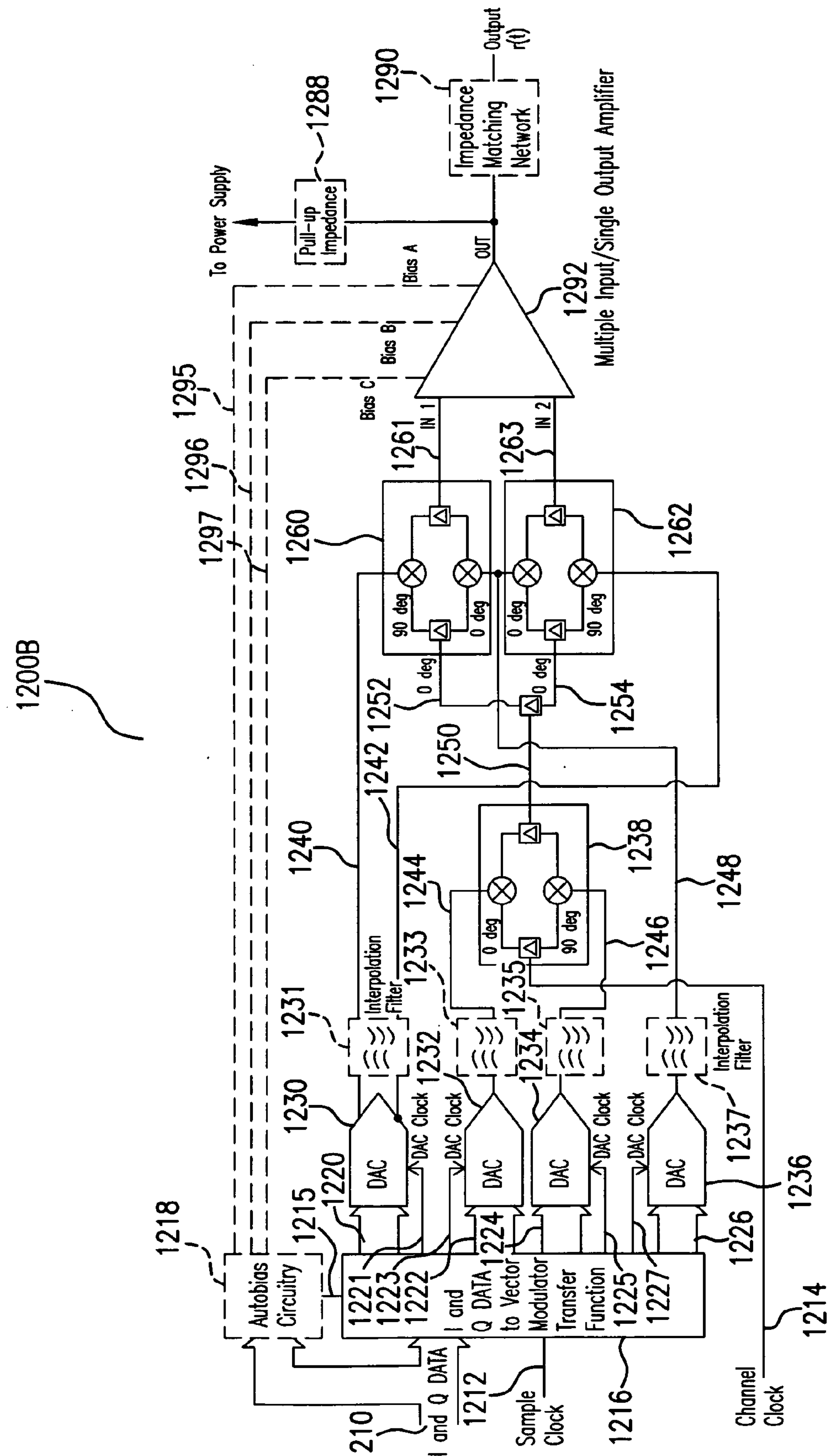
I and Q DATA
Sample Clock
Channel Clock
Autobias Circuitry
I and Q DATA to Vector Modulator Transfer Function
DAC
DAC Clock
Interpolation Filter
0 deg
90 deg
IN 1
IN 2
Bias A
Bias B
Bias C
OUT
Multiple Input/Single Output Amplifier
To Power Supply
Pull-up Impedance
Impedance Matching Network
Output r(t)

FIG. 12B

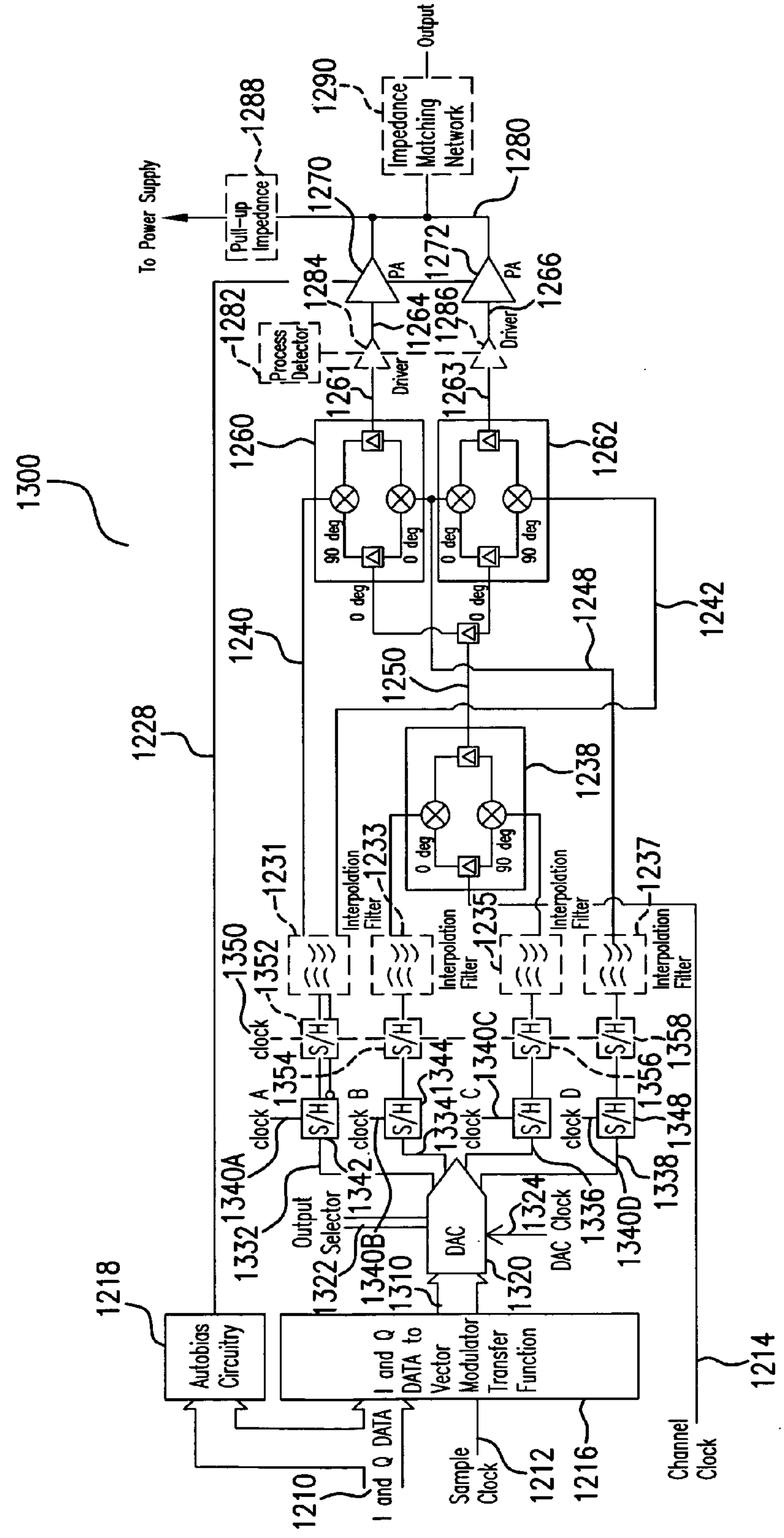
I and Q DATA
Sample Clock
Channel Clock
Autobias Circuitry
I and Q DATA to Vector Modulator Transfer Function
DAC
DAC Clock
Output Selector
clock A
S/H
clock B
clock C
clock D
clock
Interpolation Filter
0 deg
90 deg
Driver
PA
Process Detector
Pull-up Impedance
To Power Supply
Impedance Matching Network
Output

FIG. 13

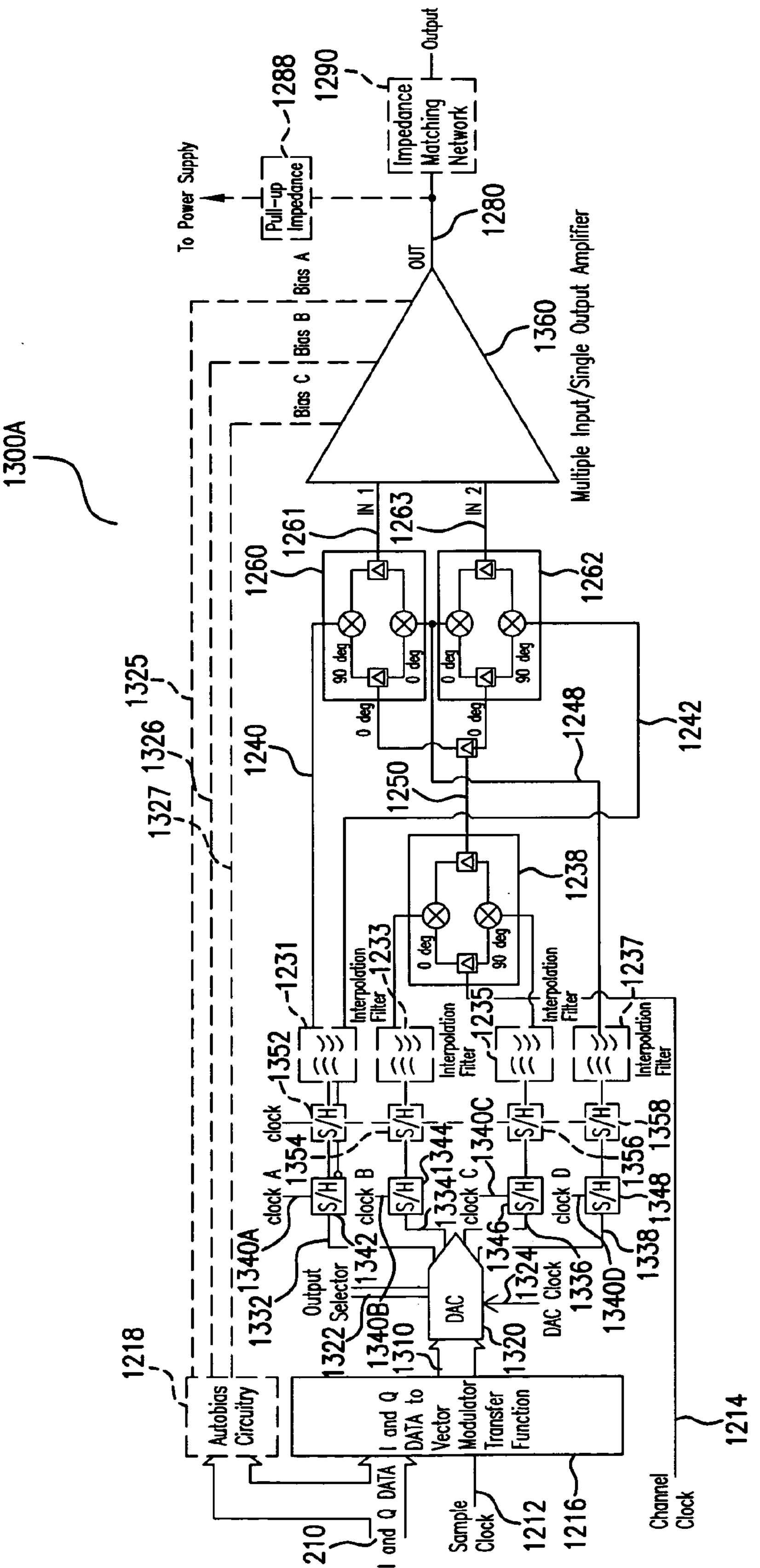
I and Q DATA
Sample Clock
Channel Clock
Autobias Circuitry
I and Q DATA to Vector Modulator Transfer Function
DAC
DAC Clock
Output Selector
clock A
clock B
clock C
clock D
clock
S/H
Interpolation Filter
0 deg
90 deg
IN 1
IN 2
OUT
Multiple Input/Single Output Amplifier
Bias A
Bias B
Bias C
To Power Supply
Pull-up Impedance
Impedance Matching Network
Output

FIG.13A

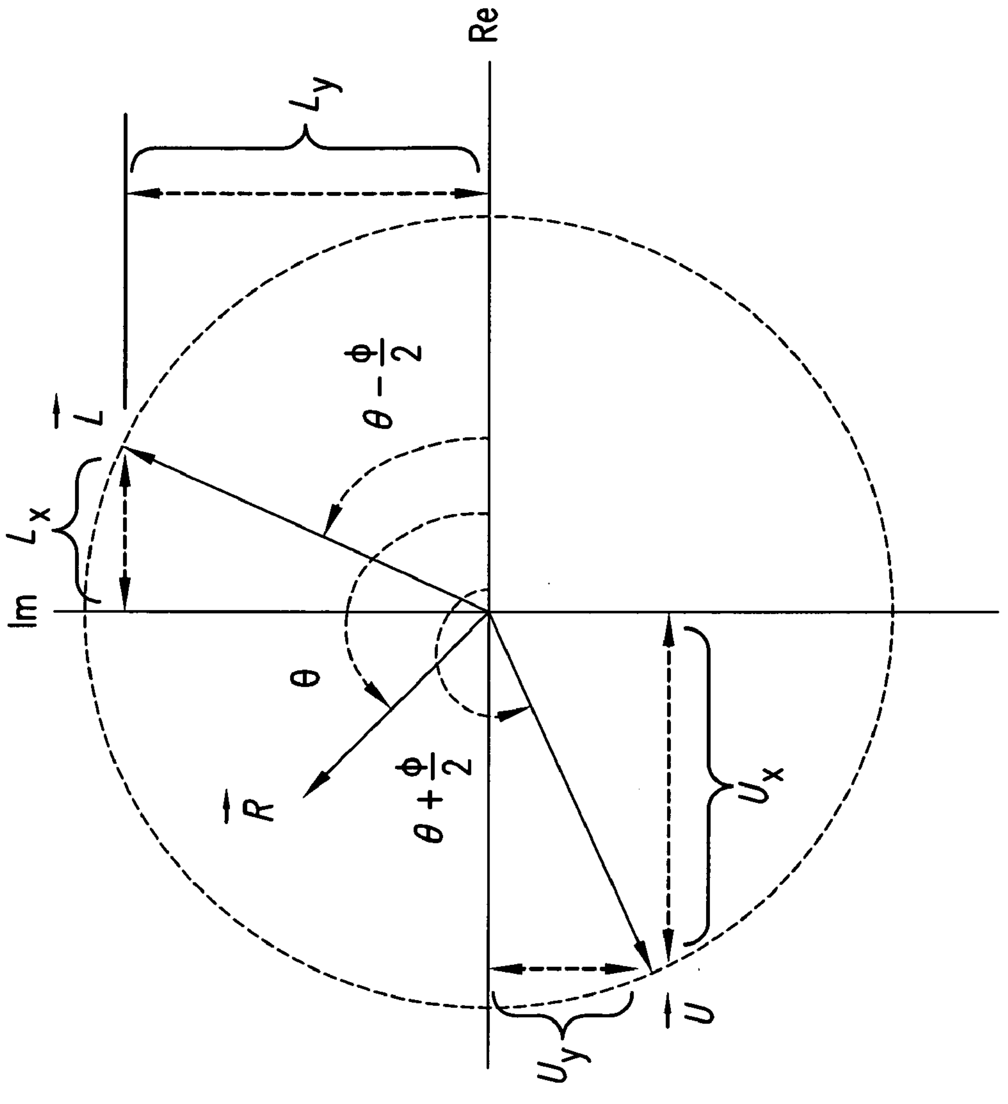
Re
L_y
$\vec{L}$
$\theta - \frac{\phi}{2}$
L_x
Im
θ
$\vec{R}$
$\theta + \frac{\phi}{2}$
U_x
$\vec{U}$
U_y

FIG.14

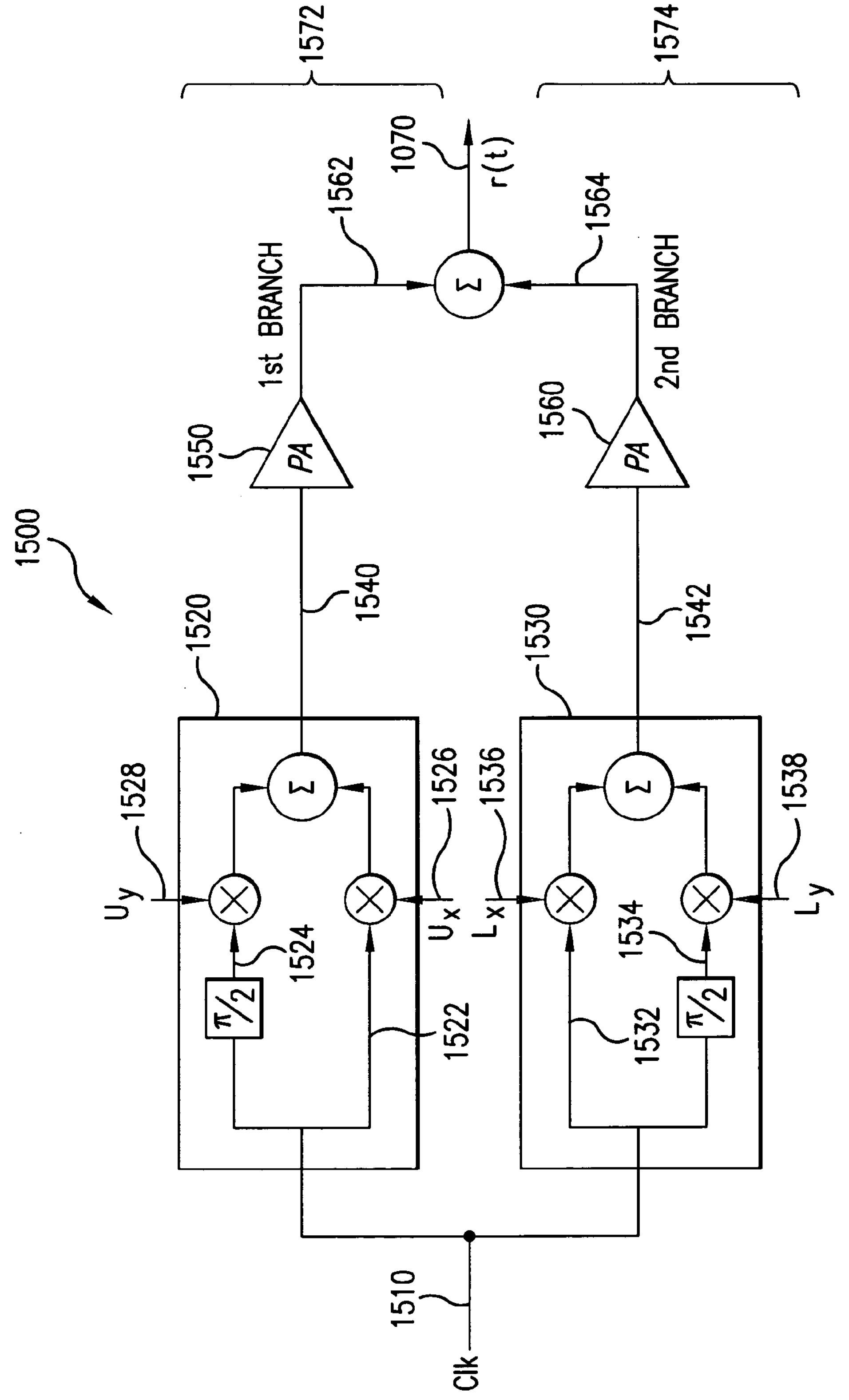
1500
1572
1574
1070
r(t)
1562
1564
Σ
1st BRANCH
2nd BRANCH
1550
PA
1560
PA
1540
1542
1520
1530
Σ
Σ
1528
U_y
1526
U_x
1536
L_x
1538
L_y
1524
1534
π/2
π/2
1522
1532
1510
Clk

FIG. 15

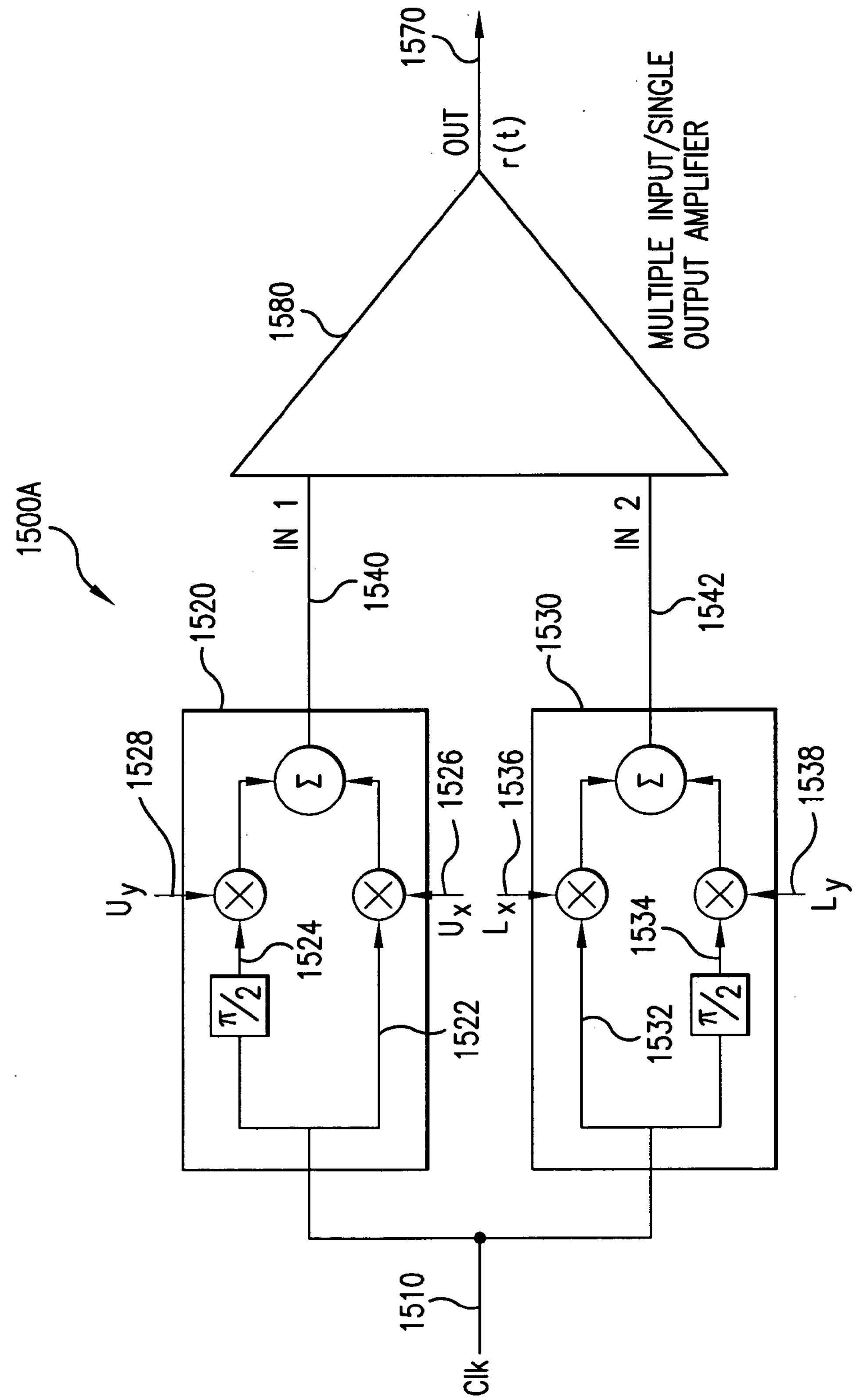
1500A
Clk
1510
1520
1528
Uy
1524
π/2
1522
Σ
Ux
1526
1540
IN 1
1530
Lx
1536
1532
1534
π/2
Σ
Ly
1538
1542
IN 2
1580
MULTIPLE INPUT/SINGLE OUTPUT AMPLIFIER
OUT
r(t)
1570

FIG.15A

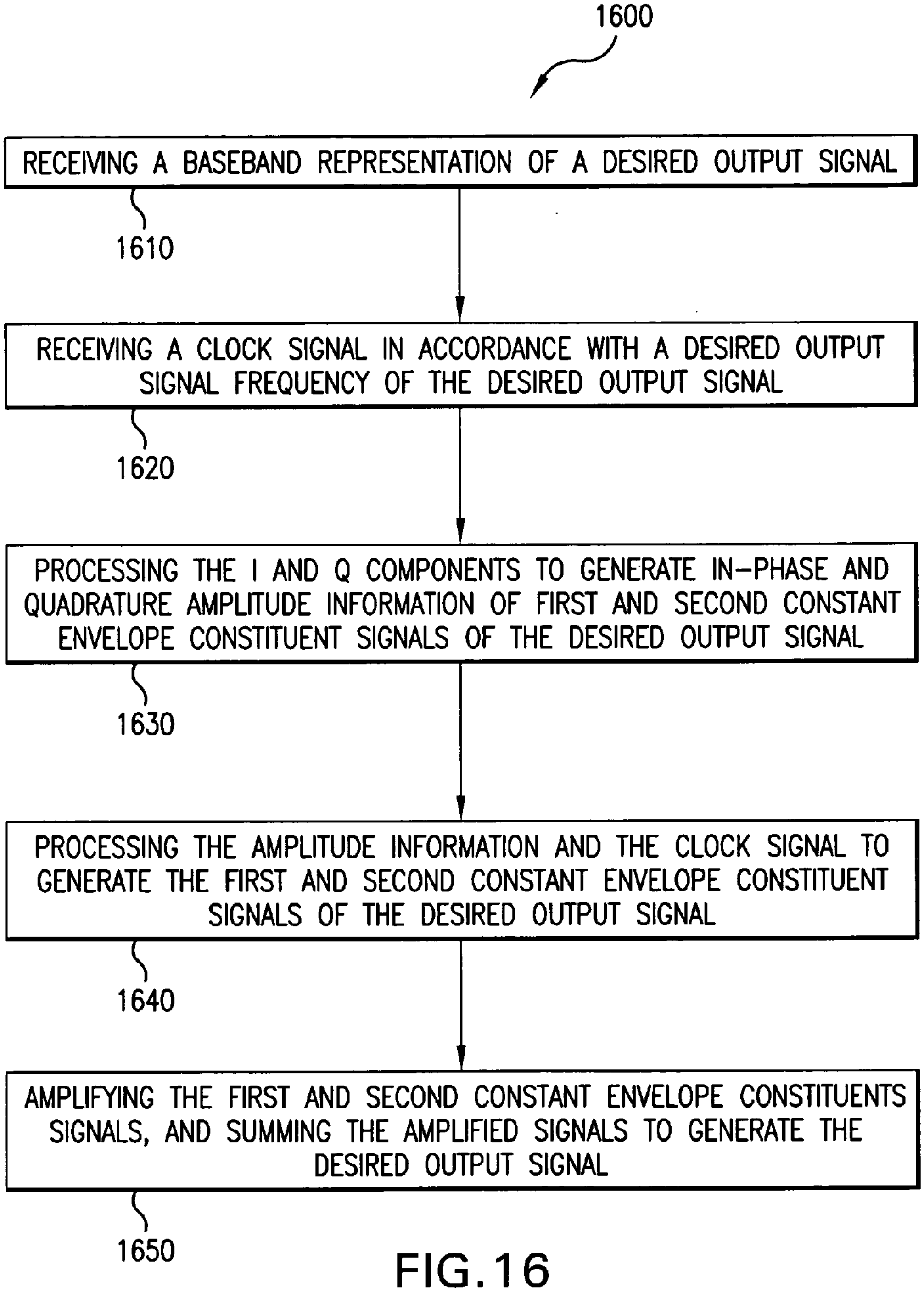
1600
RECEIVING A BASEBAND REPRESENTATION OF A DESIRED OUTPUT SIGNAL
1610
RECEIVING A CLOCK SIGNAL IN ACCORDANCE WITH A DESIRED OUTPUT SIGNAL FREQUENCY OF THE DESIRED OUTPUT SIGNAL
1620
PROCESSING THE I AND Q COMPONENTS TO GENERATE IN-PHASE AND QUADRATURE AMPLITUDE INFORMATION OF FIRST AND SECOND CONSTANT ENVELOPE CONSTITUENT SIGNALS OF THE DESIRED OUTPUT SIGNAL
1630
PROCESSING THE AMPLITUDE INFORMATION AND THE CLOCK SIGNAL TO GENERATE THE FIRST AND SECOND CONSTANT ENVELOPE CONSTITUENT SIGNALS OF THE DESIRED OUTPUT SIGNAL
1640
AMPLIFYING THE FIRST AND SECOND CONSTANT ENVELOPE CONSTITUENTS SIGNALS, AND SUMMING THE AMPLIFIED SIGNALS TO GENERATE THE DESIRED OUTPUT SIGNAL
1650

FIG. 16

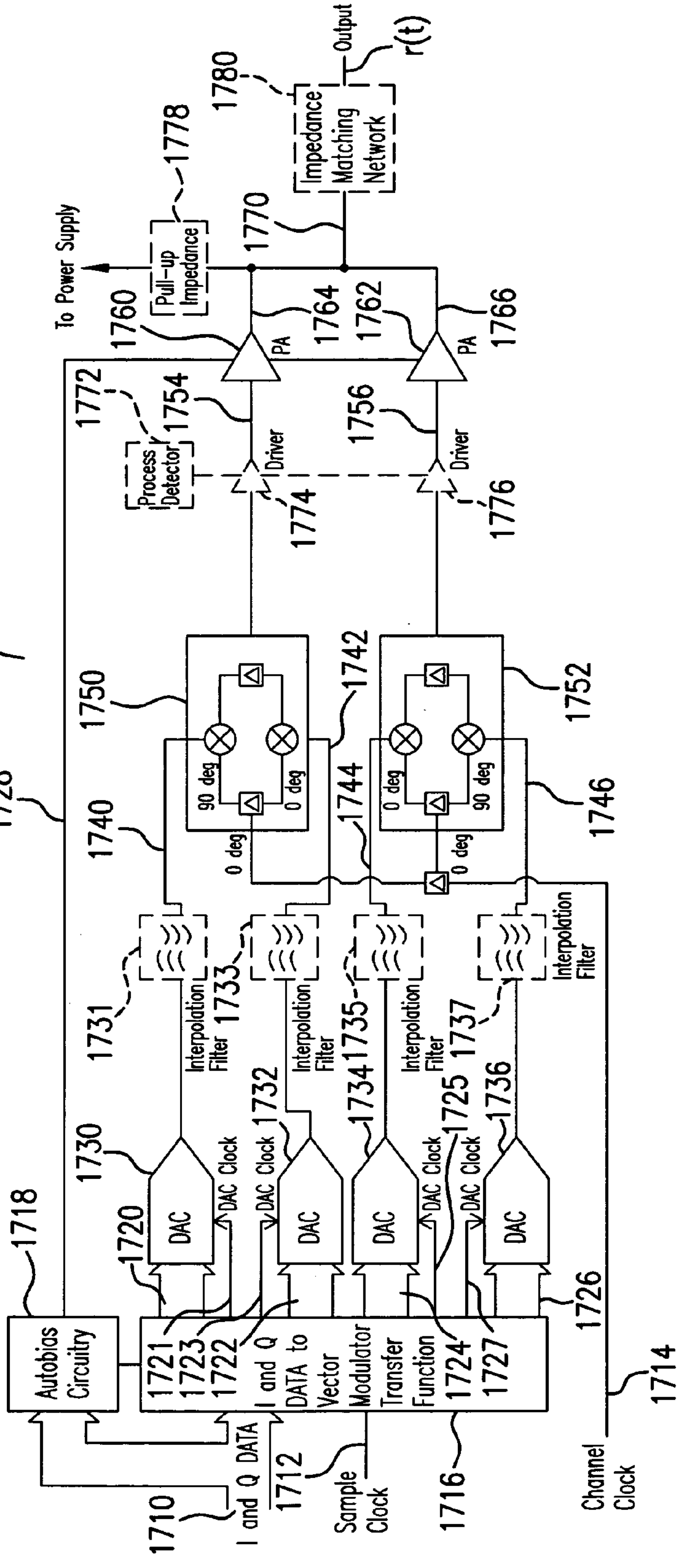
1700
1710
I and Q DATA
1712
Sample Clock
1716
Channel Clock
1714
Autobias Circuitry
1718
I and Q DATA to Vector Modulator Transfer Function
1720
1721
1723
1722
1724
1727
1725
1726
DAC
DAC Clock
1730
1732
1734
1736
Interpolation Filter
1731
1733
1735
1737
1728
1740
1742
1744
1746
1750
1752
90 deg
0 deg
Driver
1754
1756
1772
Process Detector
1774
1776
PA
1760
1762
1764
1766
To Power Supply
Pull-up Impedance
1778
1770
Impedance Matching Network
1780
Output
r(t)

FIG.17

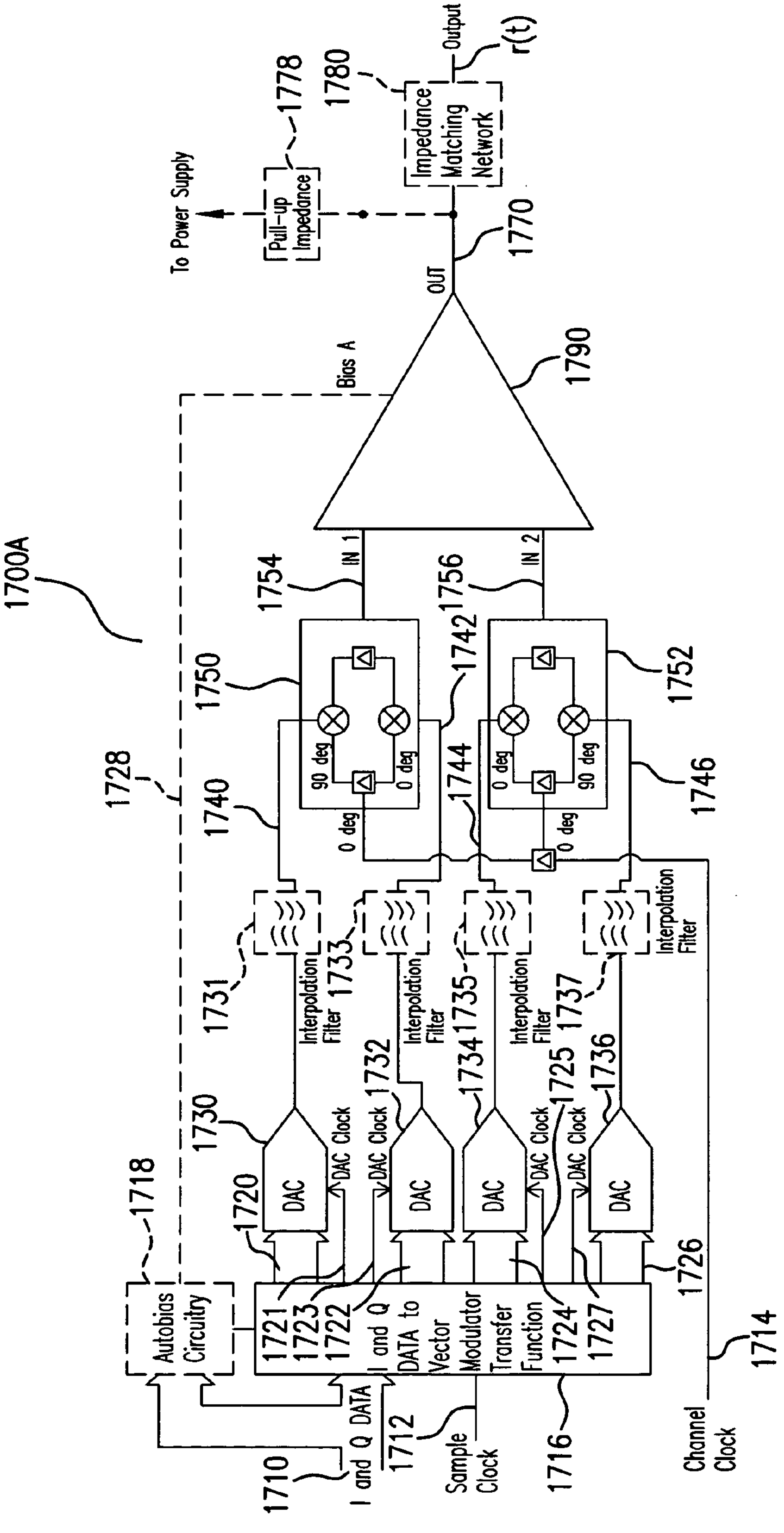
1700A
1710
I and Q DATA
1712
Sample Clock
1716
Channel Clock
1714
1718
Autobias Circuitry
I and Q DATA to Vector Modulator Transfer Function
1721
1723
1722
1724
1727
1720
1726
1725
DAC
DAC Clock
1730
1732
1734
1736
Interpolation Filter
1731
1733
1735
1737
1728
1740
1742
1744
1746
1750
1752
90 deg
0 deg
1754
1756
IN 1
IN 2
Bias A
1790
OUT
1770
Pull-up Impedance
1778
To Power Supply
Impedance Matching Network
1780
Output
r(t)

FIG.17A

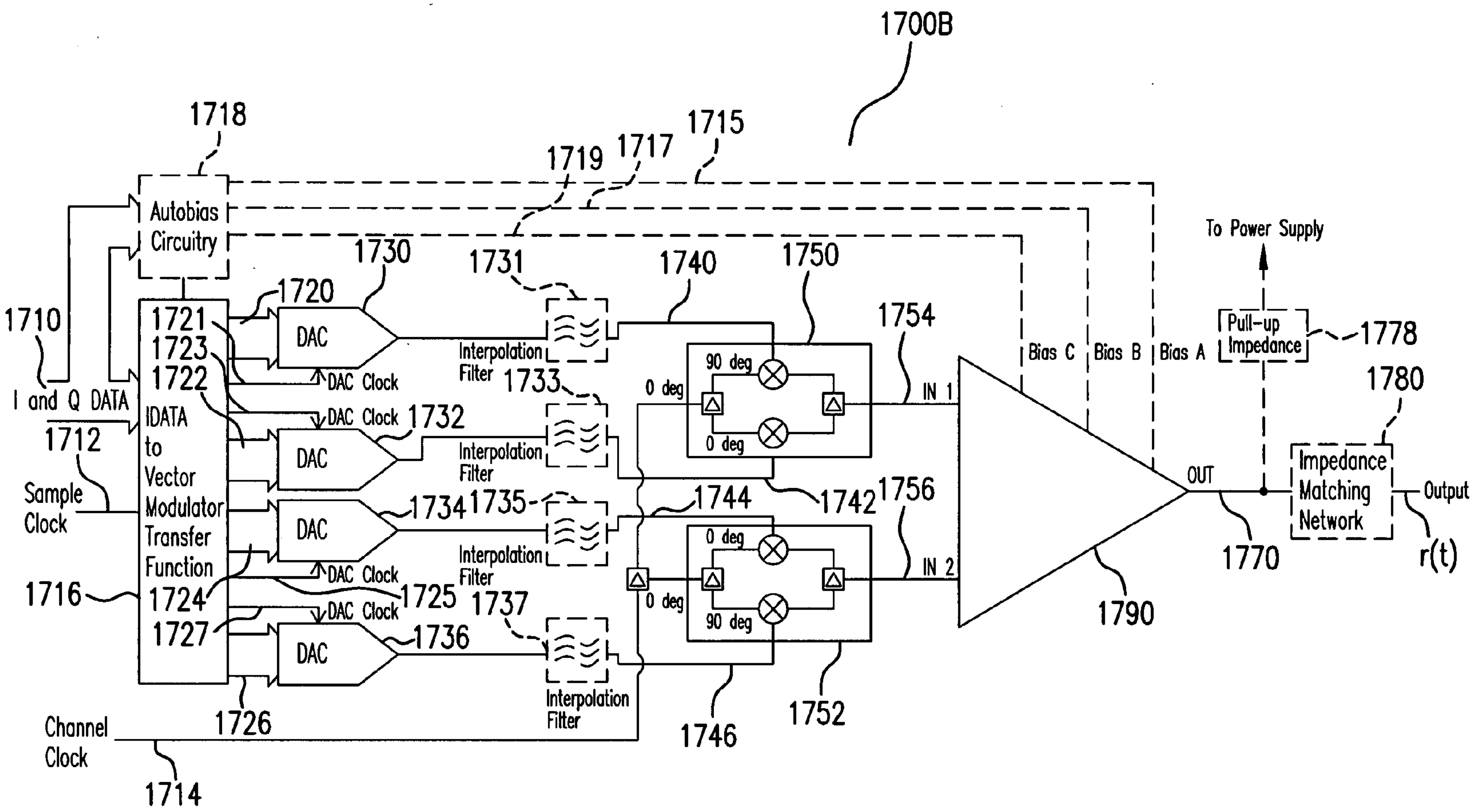
1700B
1718
Autobias
Circuitry
1719
1717
1715
1710
I and Q DATA
1712
Sample
Clock
1716
IDATA
to
Vector
Modulator
Transfer
Function
1721
1723
1722
1724
1727
1726
Channel
Clock
1714
1720
DAC
DAC Clock
1730
DAC
DAC Clock
1732
DAC
DAC Clock
1734
1725
DAC
DAC Clock
1736
1731
Interpolation
Filter
1733
Interpolation
Filter
1735
Interpolation
Filter
1737
Interpolation
Filter
1740
1750
90 deg
0 deg
0 deg
1742
1744
0 deg
0 deg
90 deg
1746
1752
1754
IN 1
1756
IN 2
Bias C
Bias B
Bias A
1790
OUT
1770
To Power Supply
Pull-up
Impedance
1778
1780
Impedance
Matching
Network
Output
r(t)

FIG. 17B

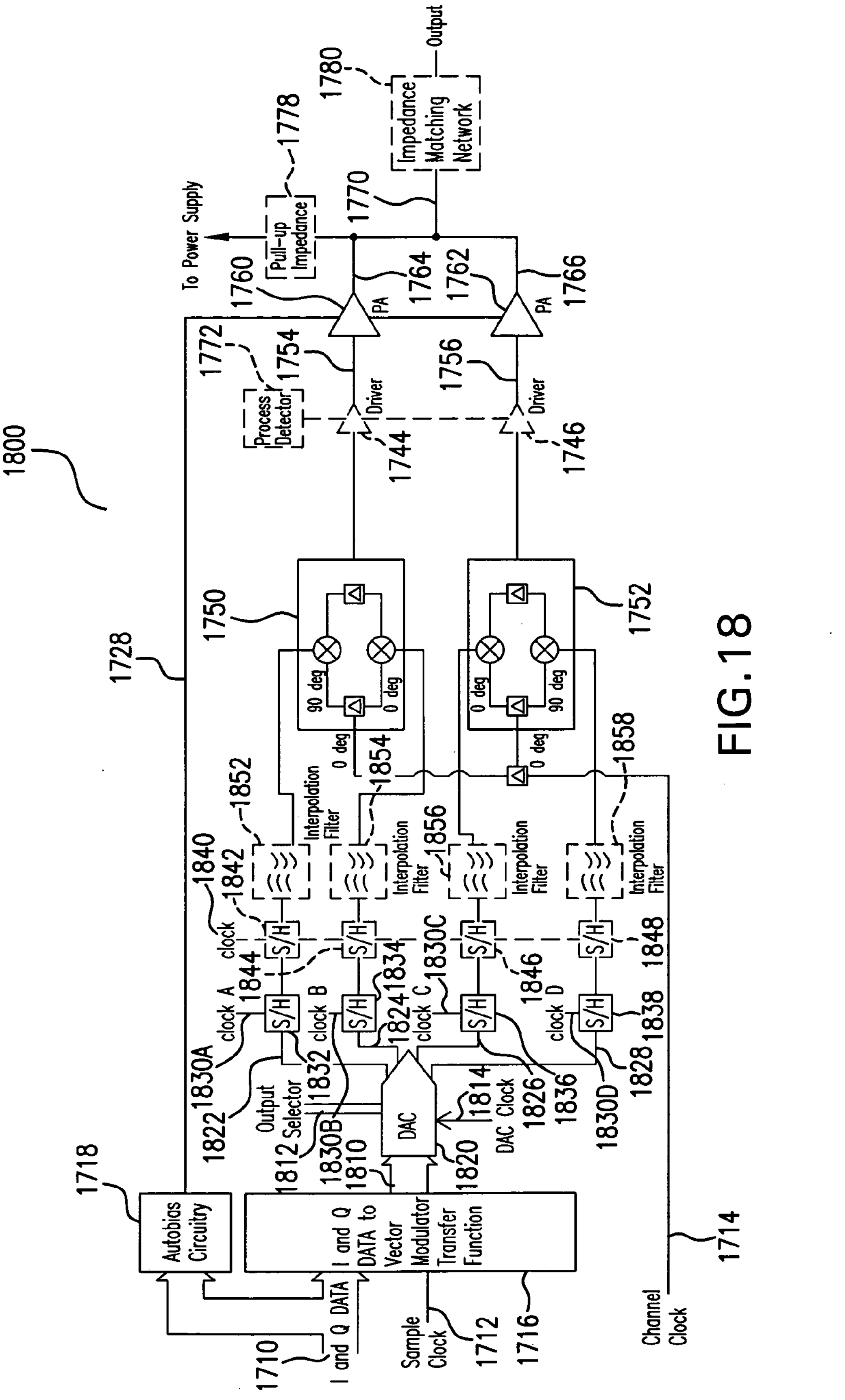
1800
1710
I and Q DATA
Sample Clock
1712
1716
Channel Clock
1714
1718
Autobias Circuitry
I and Q DATA to Vector Modulator Transfer Function
1728
1810
1812
1820
DAC
DAC Clock
1814
Output Selector
1822
1830A
1830B
1830C
1830D
1824
1826
1828
1832
1834
1836
1838
clock A
clock B
clock C
clock D
S/H
clock
1840
1842
1844
1846
1848
Interpolation Filter
1852
1856
1854
1858
0 deg
90 deg
1750
1752
1744
1746
Driver
1754
1756
Process Detector
1772
PA
1760
1762
1764
1766
To Power Supply
Pull-up Impedance
1778
1770
Impedance Matching Network
1780
Output

FIG. 18

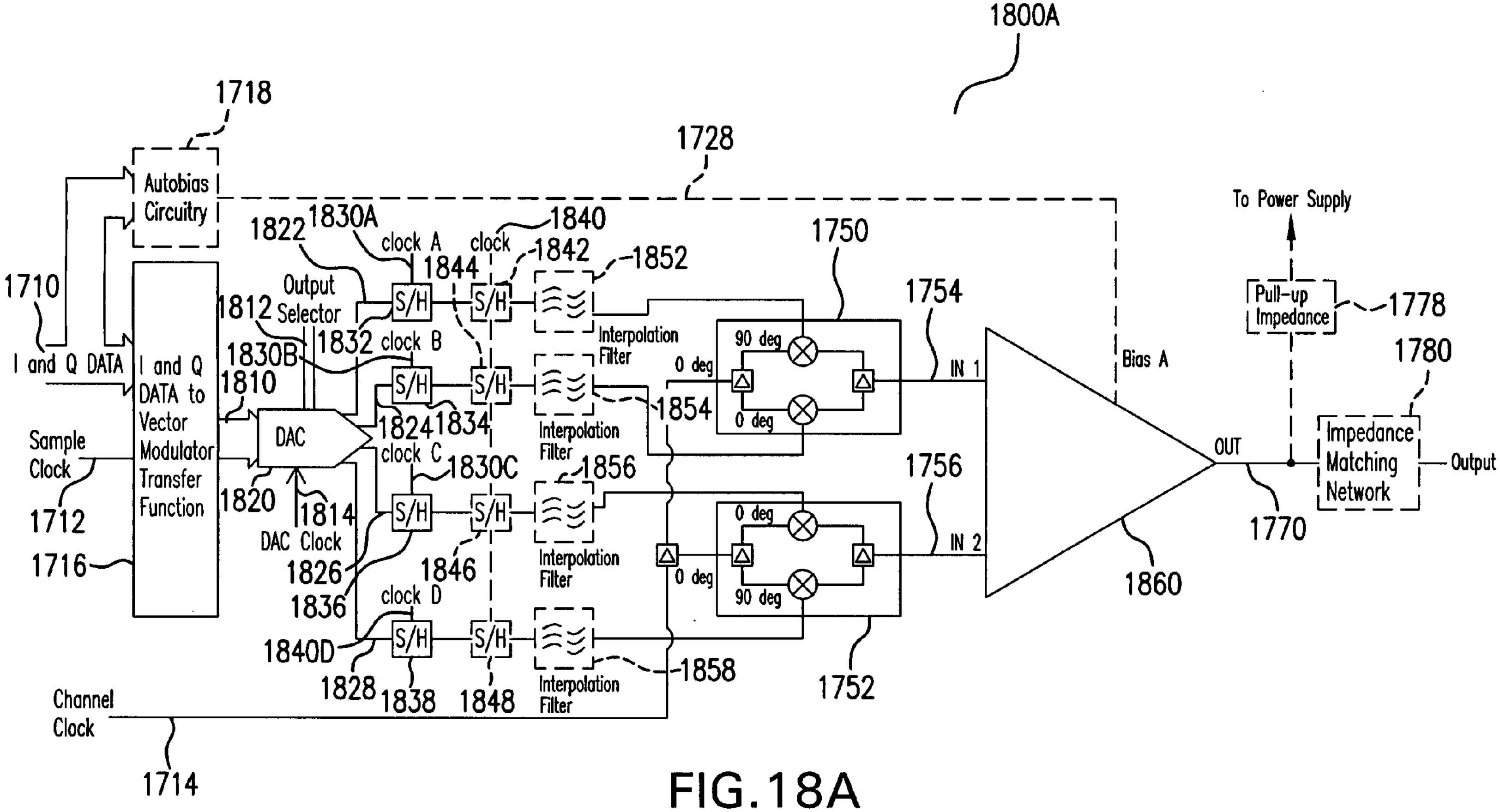
1800A
1718
Autobias Circuitry
1728
1710
I and Q DATA
I and Q DATA to Vector Modulator Transfer Function
Sample Clock
1712
1716
1812
1822
1830A
Output Selector
1832
1830B
1810
DAC
1820
1814
DAC Clock
clock A
clock B
clock C
clock D
clock
1840
1842
1844
1824
1834
1830C
1826
1836
1846
1840D
1828
1838
1848
S/H
1852
1854
1856
1858
Interpolation Filter
0 deg
90 deg
1750
1752
1754
1756
IN 1
IN 2
Bias A
1860
OUT
1770
To Power Supply
Pull-up Impedance
1778
1780
Impedance Matching Network
Output
Channel Clock
1714

FIG.18A

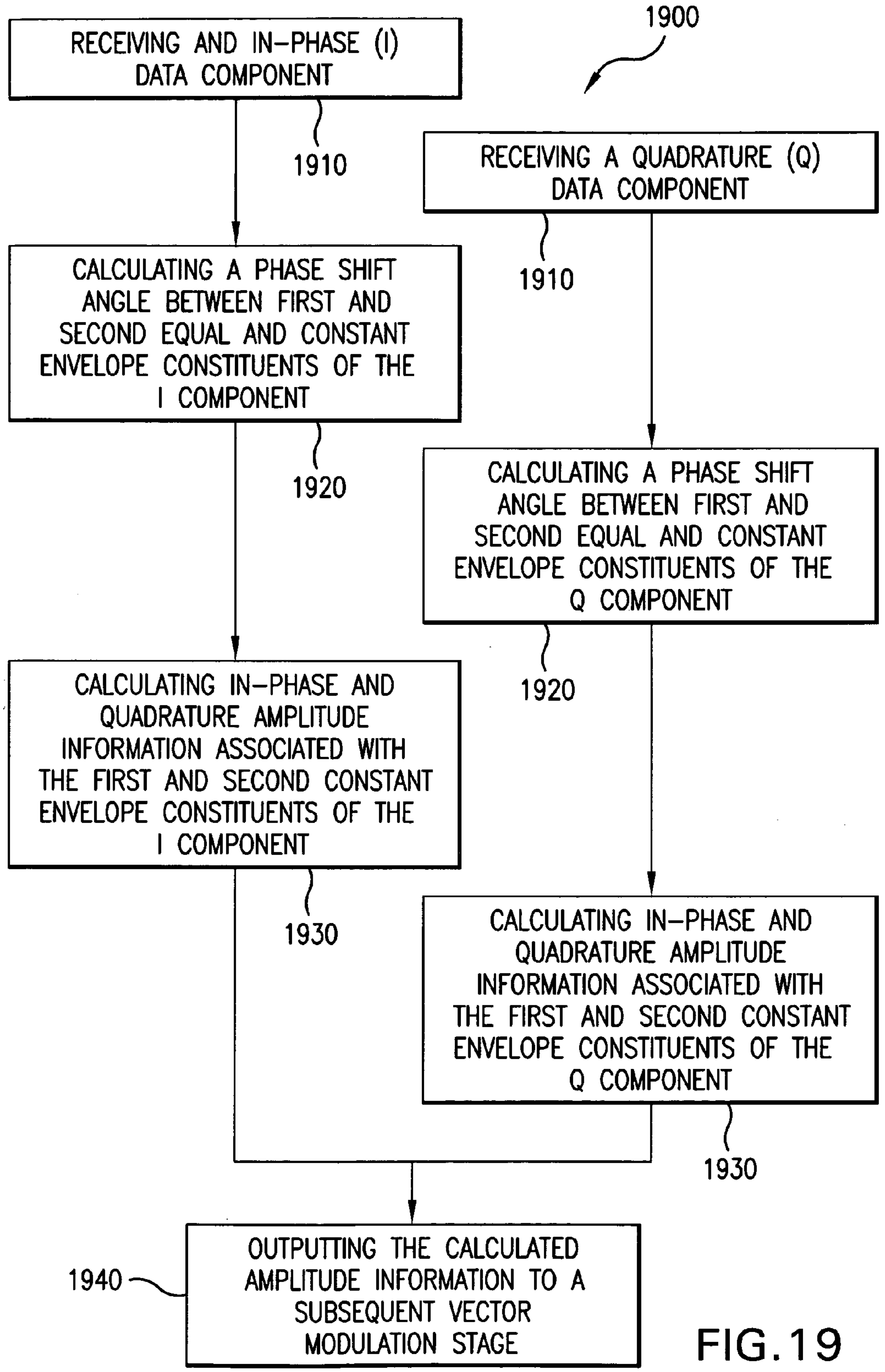
1900
RECEIVING AND IN-PHASE (I) DATA COMPONENT
1910
RECEIVING A QUADRATURE (Q) DATA COMPONENT
1910
CALCULATING A PHASE SHIFT ANGLE BETWEEN FIRST AND SECOND EQUAL AND CONSTANT ENVELOPE CONSTITUENTS OF THE I COMPONENT
1920
CALCULATING A PHASE SHIFT ANGLE BETWEEN FIRST AND SECOND EQUAL AND CONSTANT ENVELOPE CONSTITUENTS OF THE Q COMPONENT
1920
CALCULATING IN-PHASE AND QUADRATURE AMPLITUDE INFORMATION ASSOCIATED WITH THE FIRST AND SECOND CONSTANT ENVELOPE CONSTITUENTS OF THE I COMPONENT
1930
CALCULATING IN-PHASE AND QUADRATURE AMPLITUDE INFORMATION ASSOCIATED WITH THE FIRST AND SECOND CONSTANT ENVELOPE CONSTITUENTS OF THE Q COMPONENT
1930
1940
OUTPUTTING THE CALCULATED AMPLITUDE INFORMATION TO A SUBSEQUENT VECTOR MODULATION STAGE

FIG.19

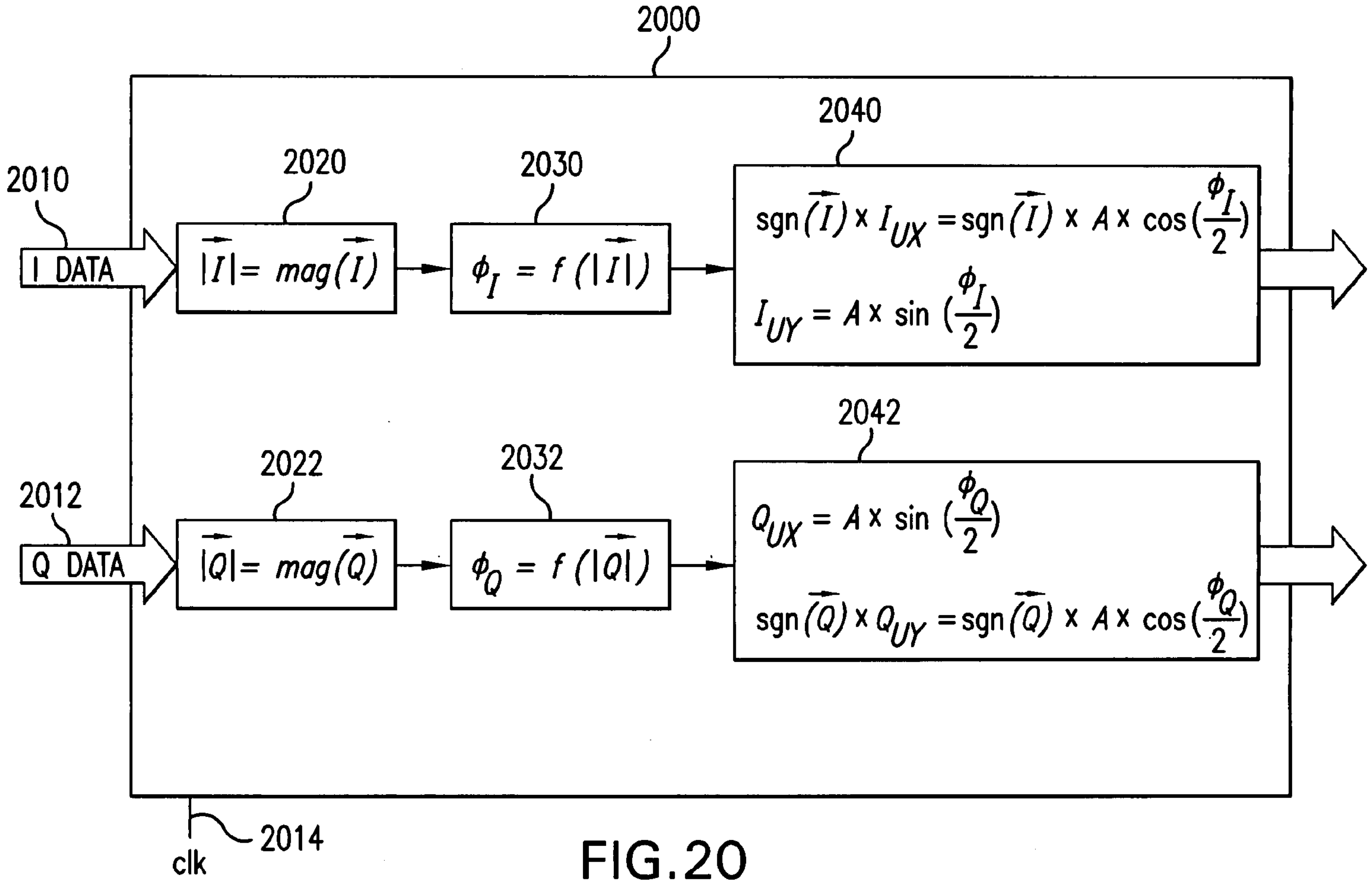
2000
2010
I DATA
2012
Q DATA
2014
clk
2020
$|\vec{I}| = mag(\vec{I})$
2022
$|\vec{Q}| = mag(\vec{Q})$
2030
$\phi_I = f(|\vec{I}|)$
2032
$\phi_Q = f(|\vec{Q}|)$
2040
$sgn(\vec{I}) \times I_{UX} = sgn(\vec{I}) \times A \times \cos(\frac{\phi_I}{2})$
$I_{UY} = A \times \sin(\frac{\phi_I}{2})$
2042
$Q_{UX} = A \times \sin(\frac{\phi_Q}{2})$
$sgn(\vec{Q}) \times Q_{UY} = sgn(\vec{Q}) \times A \times \cos(\frac{\phi_Q}{2})$

FIG.20

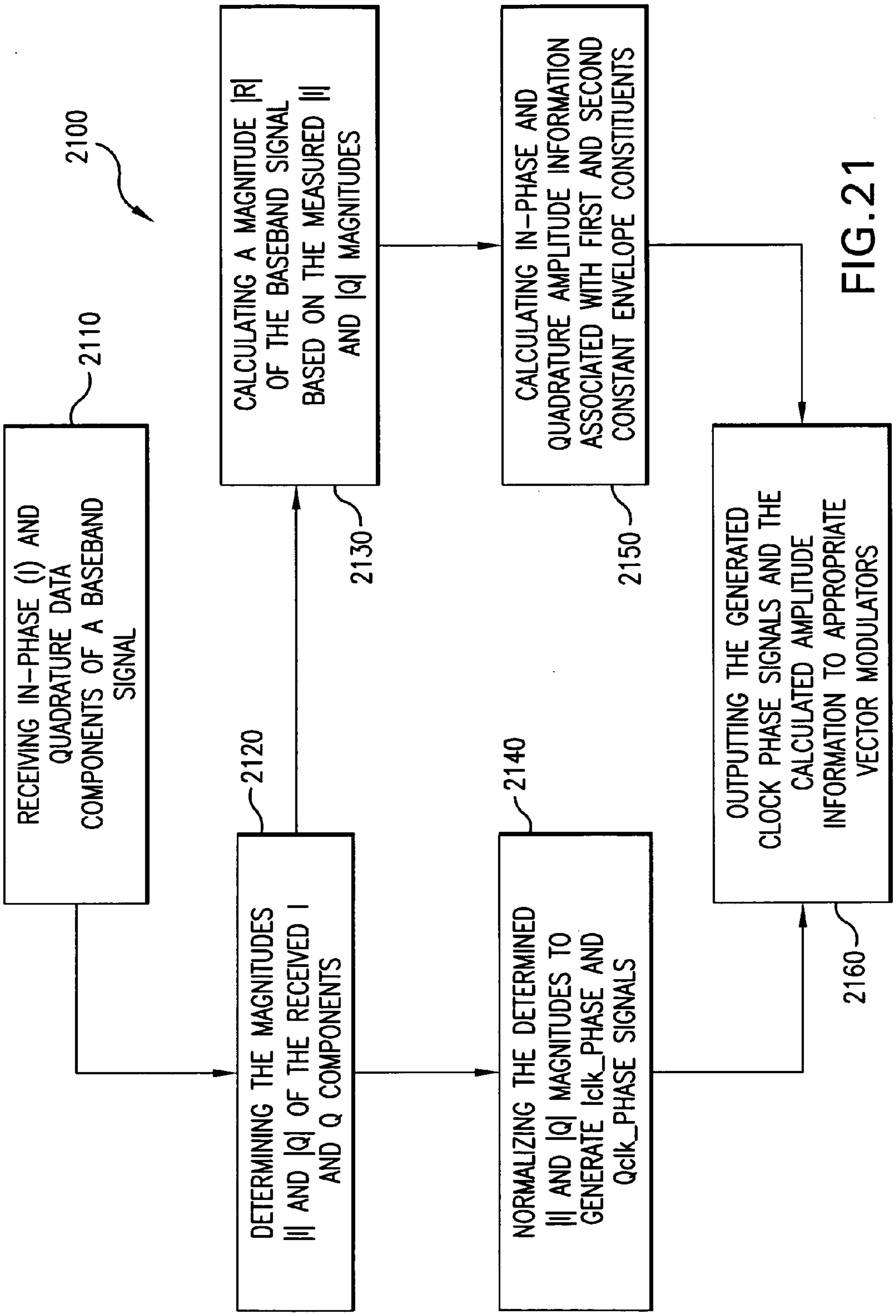
2100
2110
RECEIVING IN-PHASE (I) AND QUADRATURE DATA COMPONENTS OF A BASEBAND SIGNAL
2120
DETERMINING THE MAGNITUDES |I| AND |Q| OF THE RECEIVED I AND Q COMPONENTS
2130
CALCULATING A MAGNITUDE |R| OF THE BASEBAND SIGNAL BASED ON THE MEASURED |I| AND |Q| MAGNITUDES
2140
NORMALIZING THE DETERMINED |I| AND |Q| MAGNITUDES TO GENERATE Iclk_PHASE AND Qclk_PHASE SIGNALS
2150
CALCULATING IN-PHASE AND QUADRATURE AMPLITUDE INFORMATION ASSOCIATED WITH FIRST AND SECOND CONSTANT ENVELOPE CONSTITUENTS
2160
OUTPUTTING THE GENERATED CLOCK PHASE SIGNALS AND THE CALCULATED AMPLITUDE INFORMATION TO APPROPRIATE VECTOR MODULATORS

FIG.21

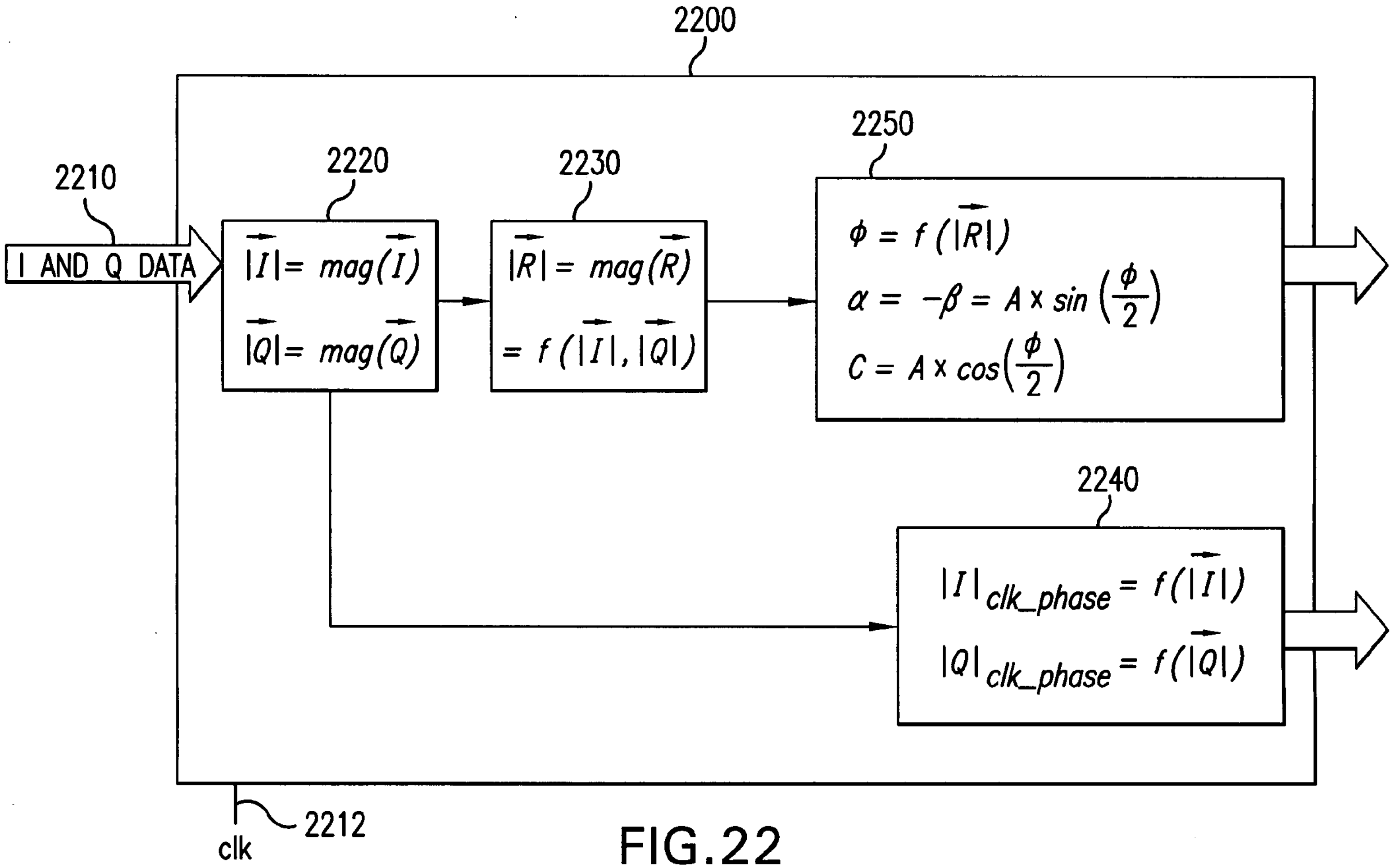
2200
2210
I AND Q DATA
2220
$|\vec{I}| = mag(\vec{I})$
$|\vec{Q}| = mag(\vec{Q})$
2230
$|\vec{R}| = mag(\vec{R})$
$= f(|\vec{I}|, |\vec{Q}|)$
2250
$\phi = f(|\vec{R}|)$
$\alpha = -\beta = A \times \sin\left(\frac{\phi}{2}\right)$
$C = A \times \cos\left(\frac{\phi}{2}\right)$
2240
$|I|_{clk\_phase} = f(|\vec{I}|)$
$|Q|_{clk\_phase} = f(|\vec{Q}|)$
clk
2212

FIG.22

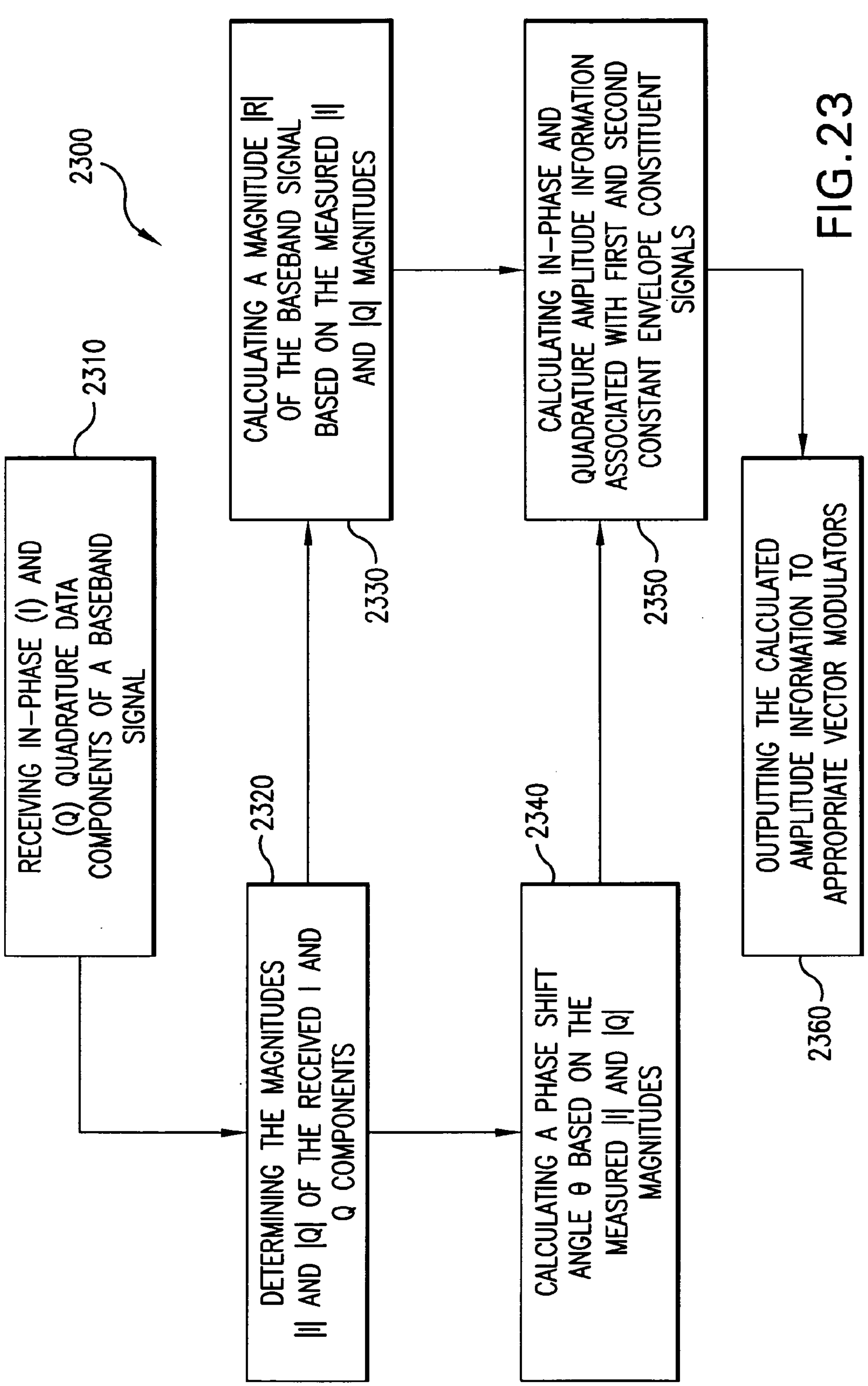
2300
RECEIVING IN-PHASE (I) AND (Q) QUADRATURE DATA COMPONENTS OF A BASEBAND SIGNAL
2310
DETERMINING THE MAGNITUDES |I| AND |Q| OF THE RECEIVED I AND Q COMPONENTS
2320
CALCULATING A MAGNITUDE |R| OF THE BASEBAND SIGNAL BASED ON THE MEASURED |I| AND |Q| MAGNITUDES
2330
CALCULATING A PHASE SHIFT ANGLE θ BASED ON THE MEASURED |I| AND |Q| MAGNITUDES
2340
CALCULATING IN-PHASE AND QUADRATURE AMPLITUDE INFORMATION ASSOCIATED WITH FIRST AND SECOND CONSTANT ENVELOPE CONSTITUENT SIGNALS
2350
OUTPUTTING THE CALCULATED AMPLITUDE INFORMATION TO APPROPRIATE VECTOR MODULATORS
2360

FIG.23

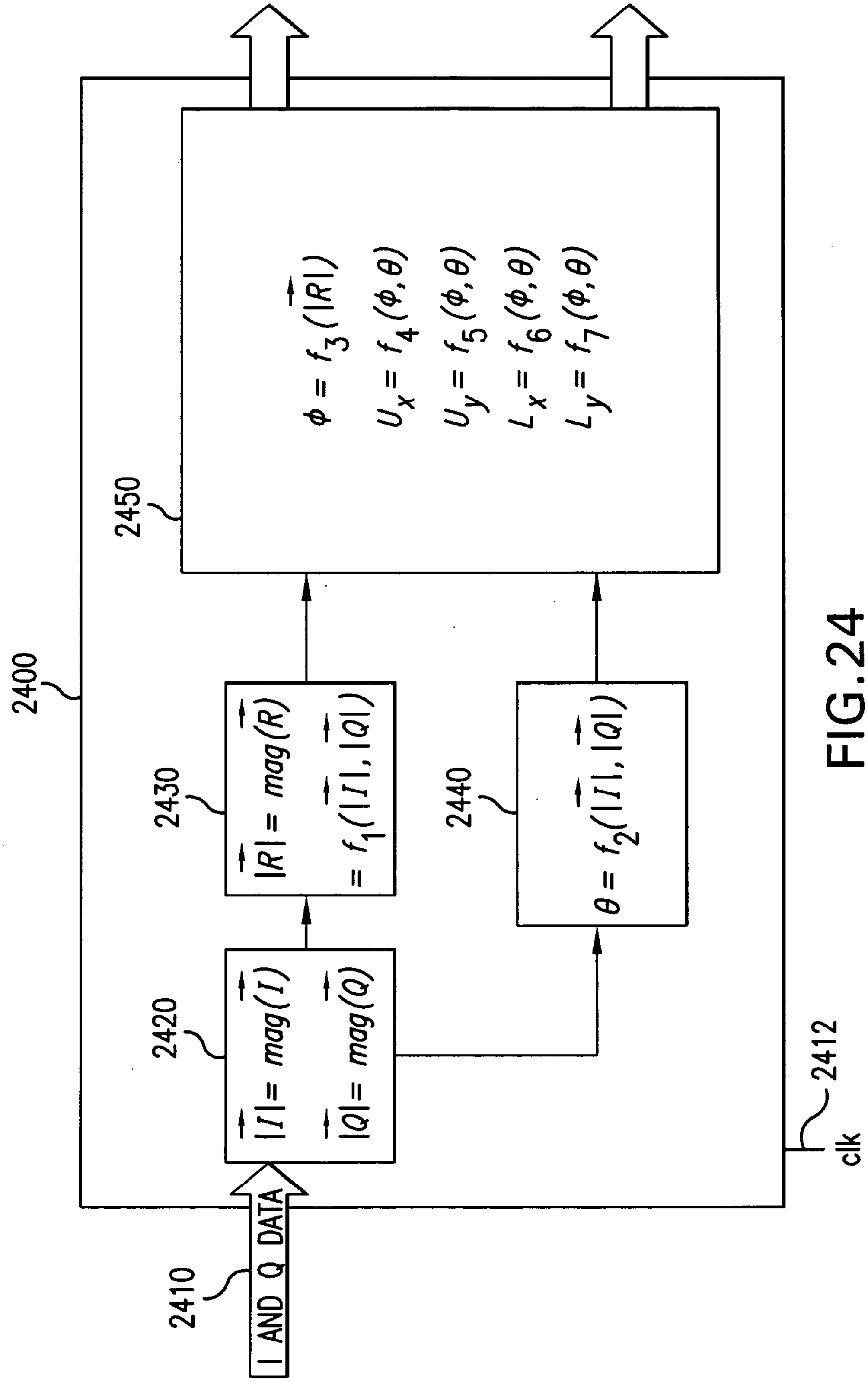
2400
2410
I AND Q DATA
2412
clk
2420
$|\vec{I}| = mag(\vec{I})$
$|\vec{Q}| = mag(\vec{Q})$
2430
$|\vec{R}| = mag(\vec{R})$
$= f_1(|\vec{I}|, |\vec{Q}|)$
2440
$\theta = f_2(|\vec{I}|, |\vec{Q}|)$
2450
$\phi = f_3(|\vec{R}|)$
$U_x = f_4(\phi, \theta)$
$U_y = f_5(\phi, \theta)$
$L_x = f_6(\phi, \theta)$
$L_y = f_7(\phi, \theta)$

FIG. 24

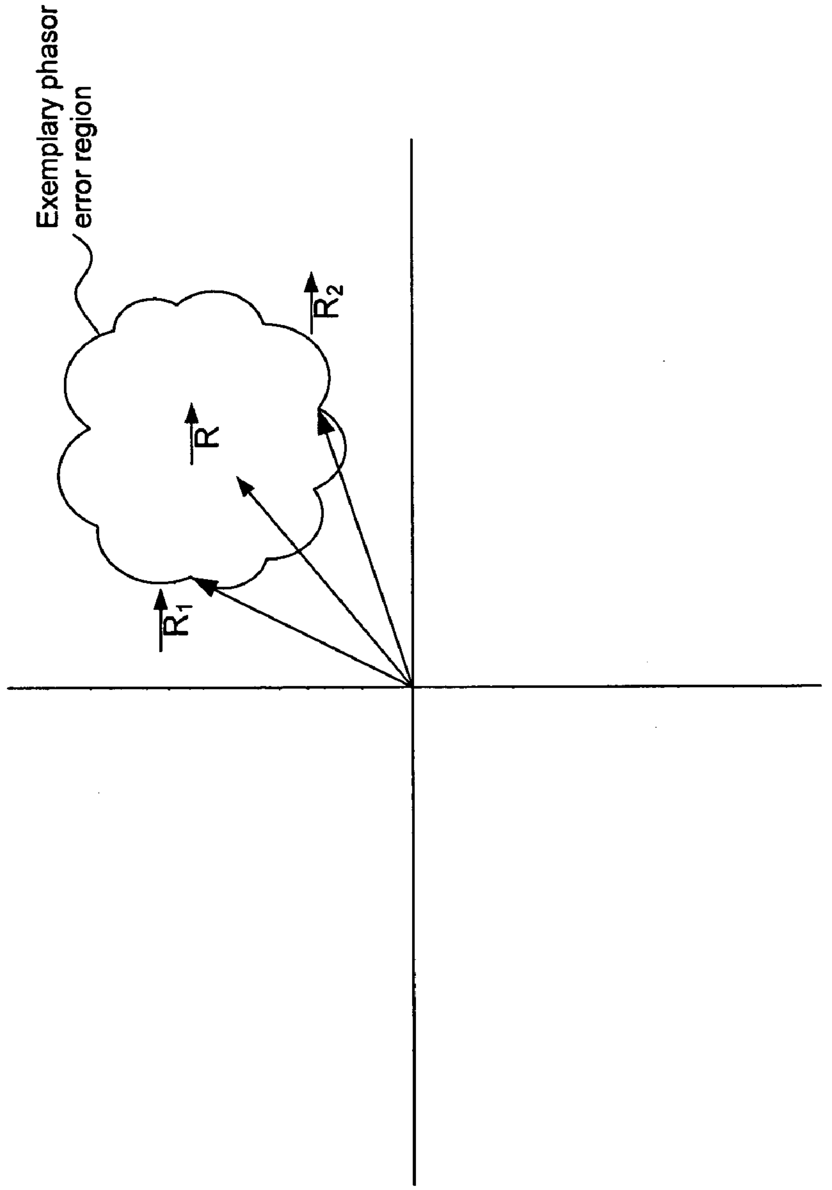
Exemplary phasor error region
$\vec{R}_2$
$\vec{R}$
$\vec{R}_1$

FIG. 25

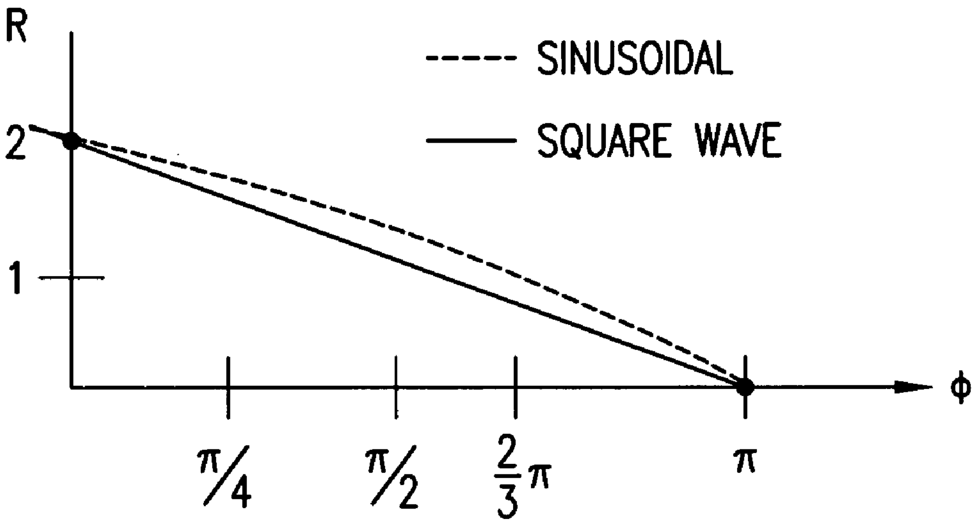

FIG.26

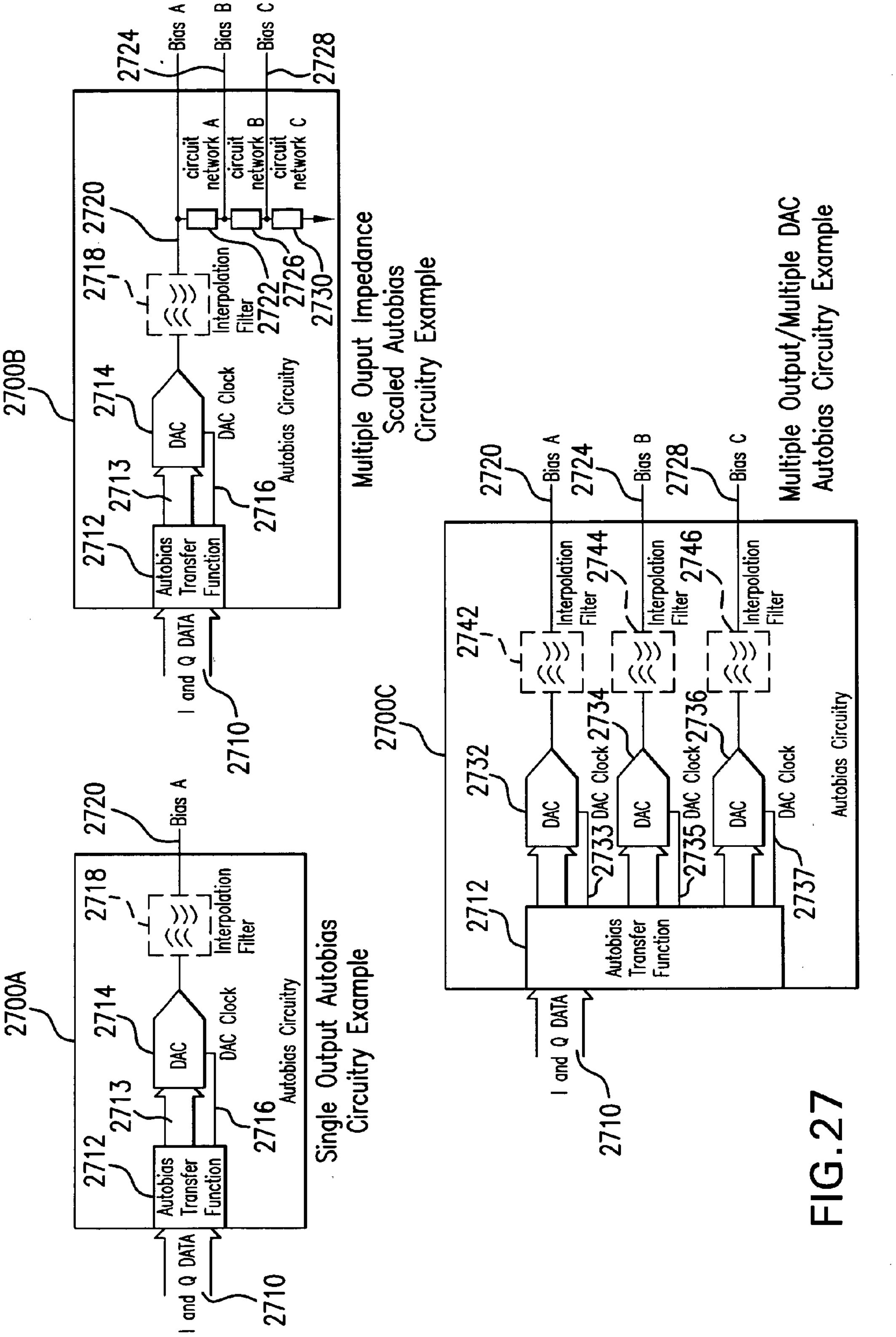
2700A
2712
2713
2714
2718
2720
Bias A
Autobias Transfer Function
DAC
DAC Clock
Interpolation Filter
Autobias Circuitry
2716
I and Q DATA
2710
Single Output Autobias Circuitry Example
2700B
2712
2713
2714
2718
2720
2724
Bias A
Bias B
Bias C
2728
circuit network A
circuit network B
circuit network C
Autobias Transfer Function
DAC
DAC Clock
Interpolation Filter
2722
2726
2730
Autobias Circuitry
2716
I and Q DATA
2710
Multiple Ouput Impedance Scaled Autobias Circuitry Example
2700C
2712
2732
2734
2736
2742
2744
2746
2720
Bias A
2724
Bias B
2728
Bias C
Autobias Transfer Function
DAC
DAC Clock
2733
2735
2737
Interpolation Filter
Autobias Circuitry
I and Q DATA
2710
Multiple Output/Multiple DAC Autobias Circuitry Example

FIG.27

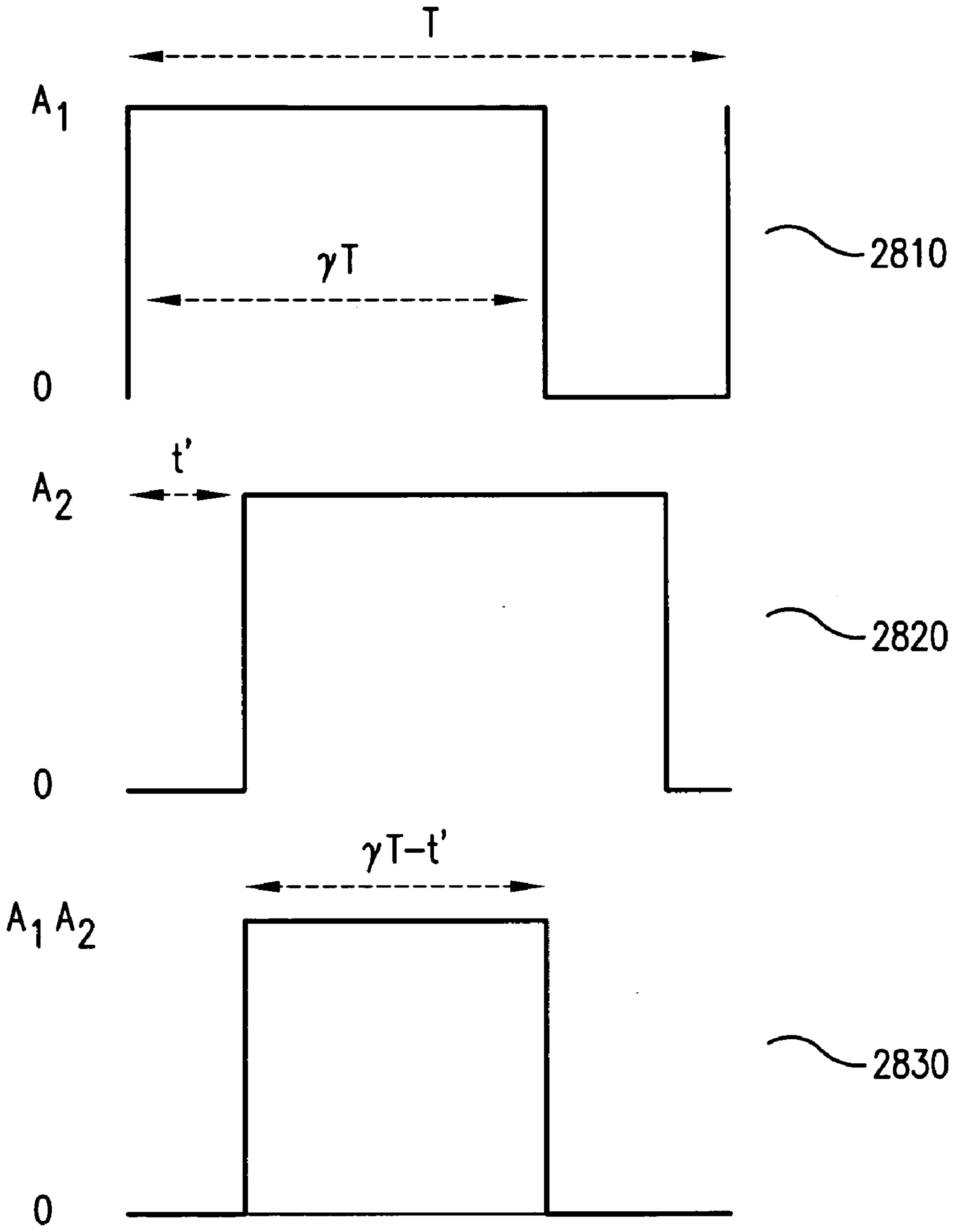
T
A_1
γT
0
2810
t'
A_2
2820
0
$\gamma T-t'$
$A_1 A_2$
2830
0

FIG. 28

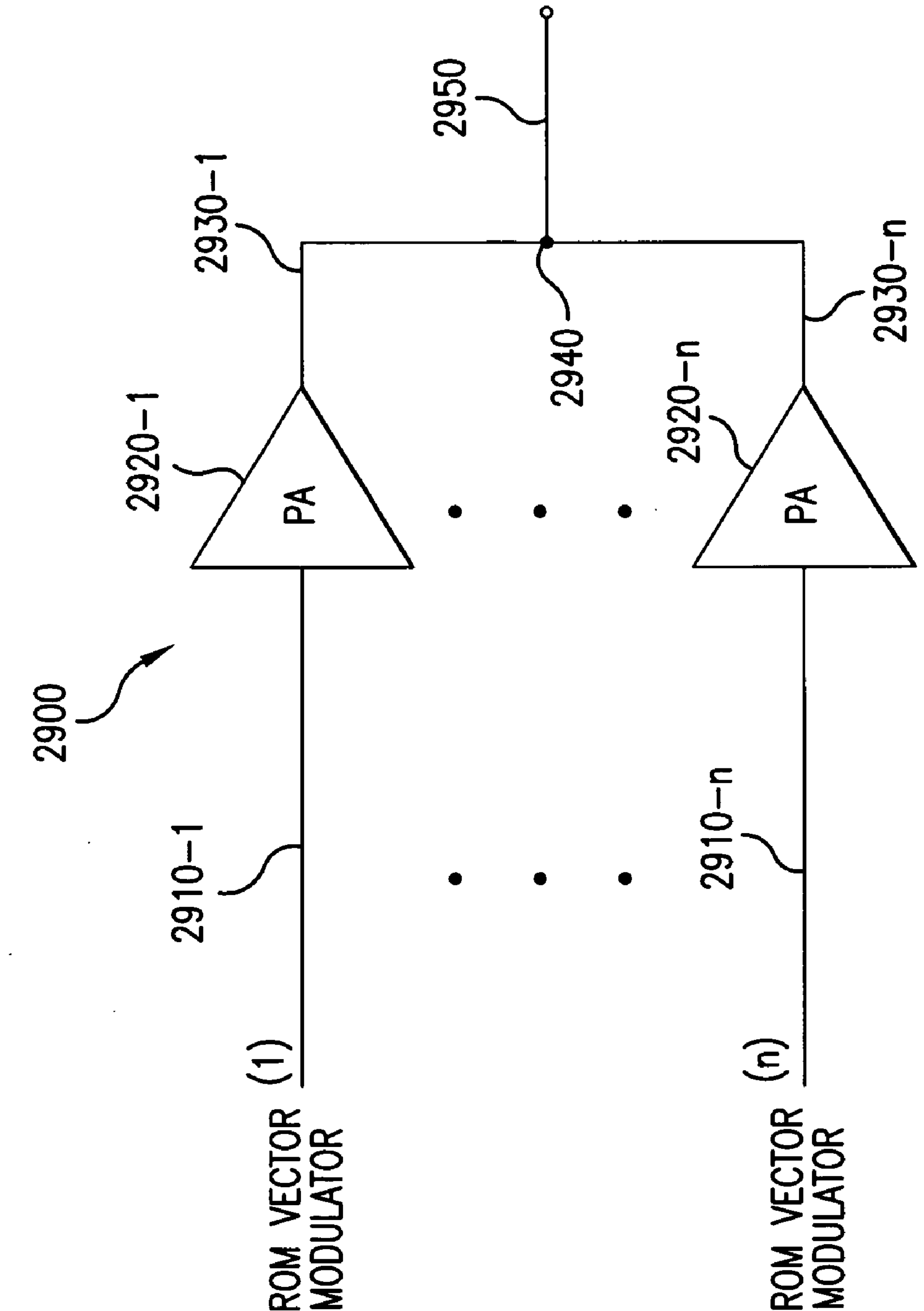
2900
FROM VECTOR MODULATOR
(1)
2910-1
PA
2920-1
2930-1
2940
2950
FROM VECTOR MODULATOR
(n)
2910-n
PA
2920-n
2930-n

FIG.29

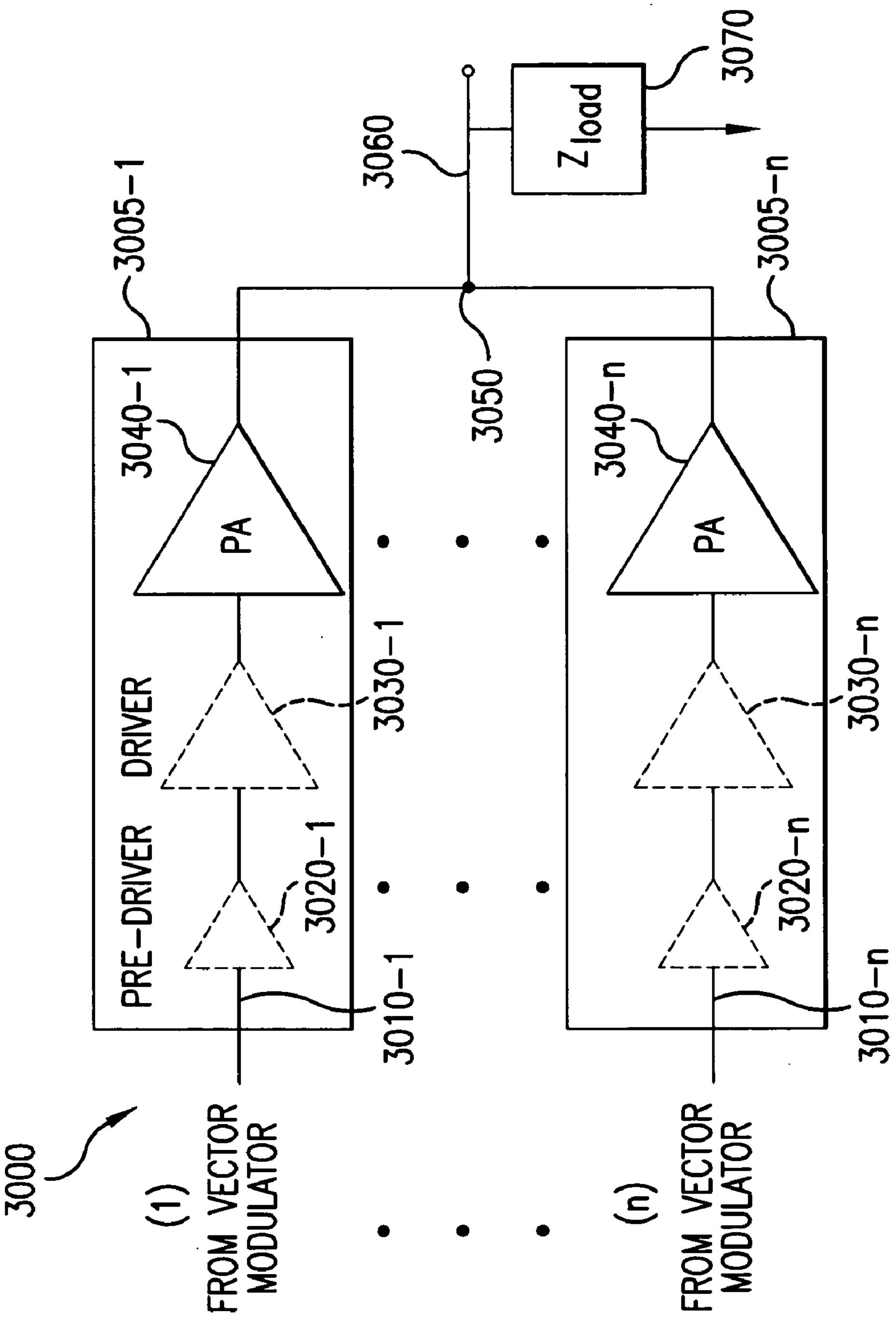
3000
(1)
FROM VECTOR
MODULATOR
(n)
FROM VECTOR
MODULATOR
3005-1
3005-n
PRE-DRIVER
DRIVER
PA
PA
3010-1
3010-n
3020-1
3020-n
3030-1
3030-n
3040-1
3040-n
3050
3060
Z_{load}
3070

FIG.30

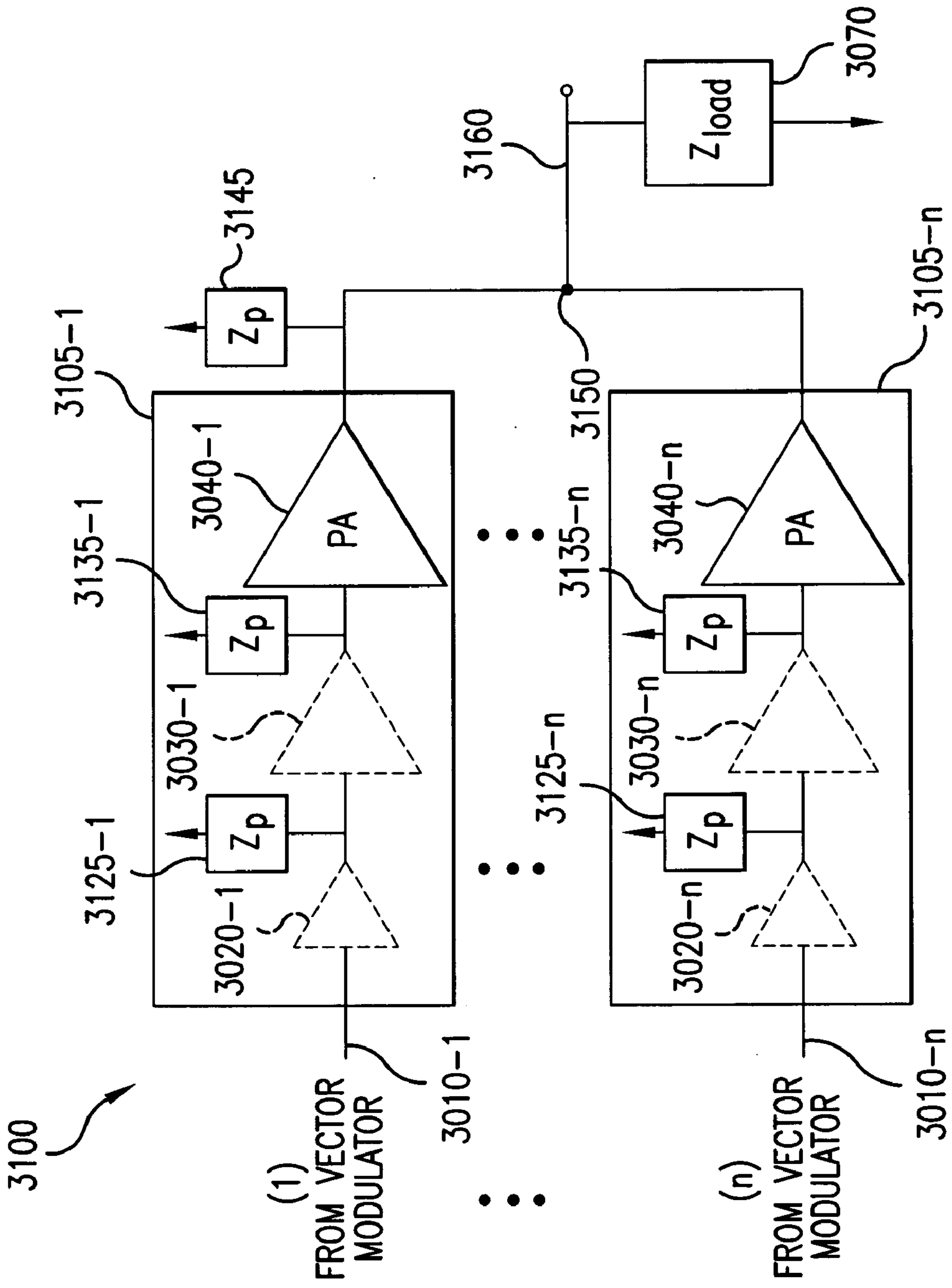
3100
(1)
FROM VECTOR
MODULATOR
3010–1
3105–1
3020–1
3125–1
Zp
3030–1
3135–1
Zp
3040–1
PA
Zp
3145
3150
3160
Zload
3070
(n)
FROM VECTOR
MODULATOR
3010–n
3105–n
3020–n
3125–n
Zp
3030–n
3135–n
Zp
3040–n
PA

FIG.31

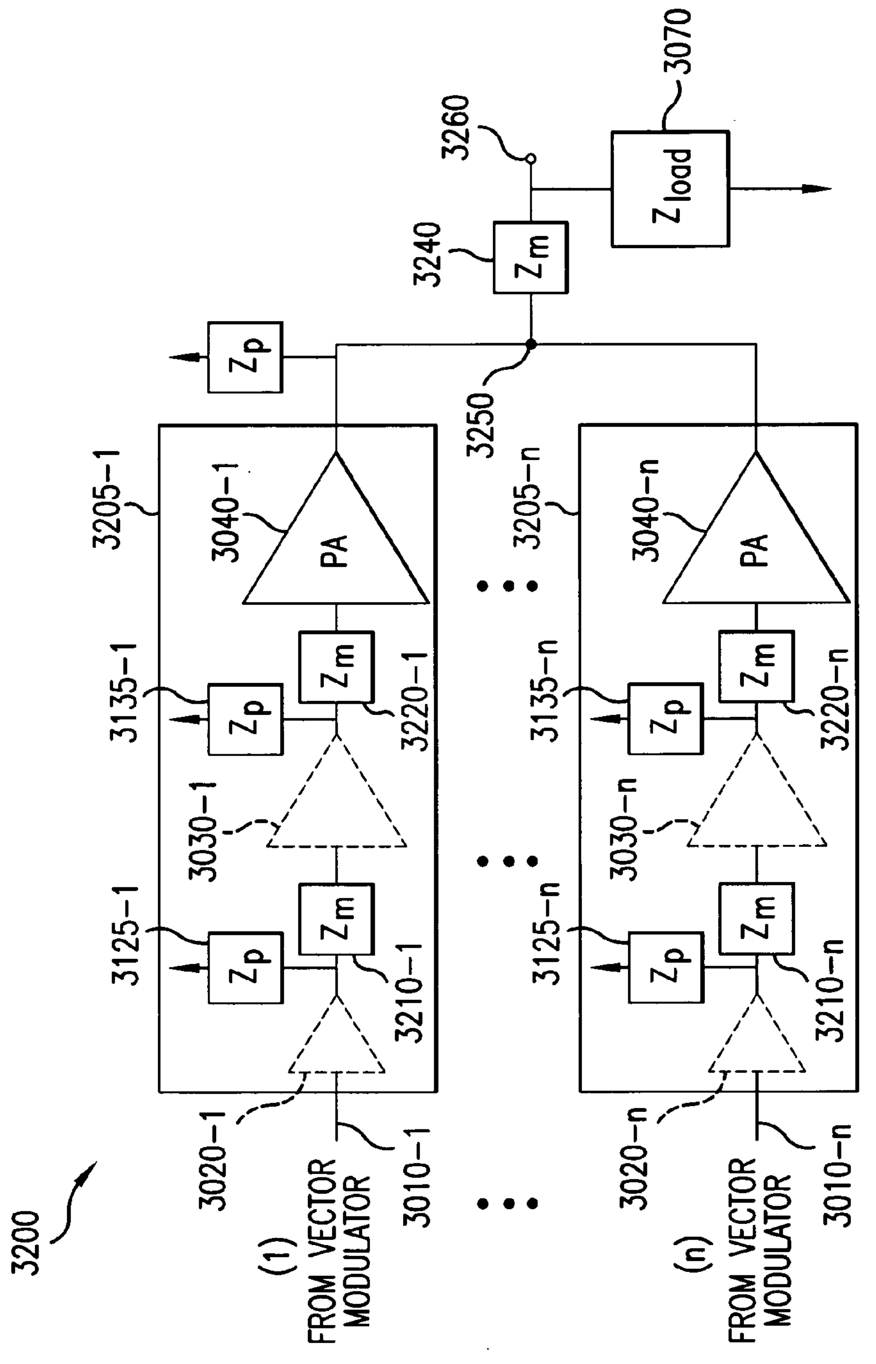
3200
(1)
FROM VECTOR
MODULATOR
3010-1
3020-1
3210-1
Zp
Zm
3125-1
3030-1
3135-1
Zp
Zm
3220-1
3040-1
PA
3205-1
Zp
3250
3240
Zm
3260
Zload
3070
(n)
FROM VECTOR
MODULATOR
3010-n
3020-n
3210-n
Zp
Zm
3125-n
3030-n
3135-n
Zp
Zm
3220-n
3040-n
PA
3205-n

FIG.32

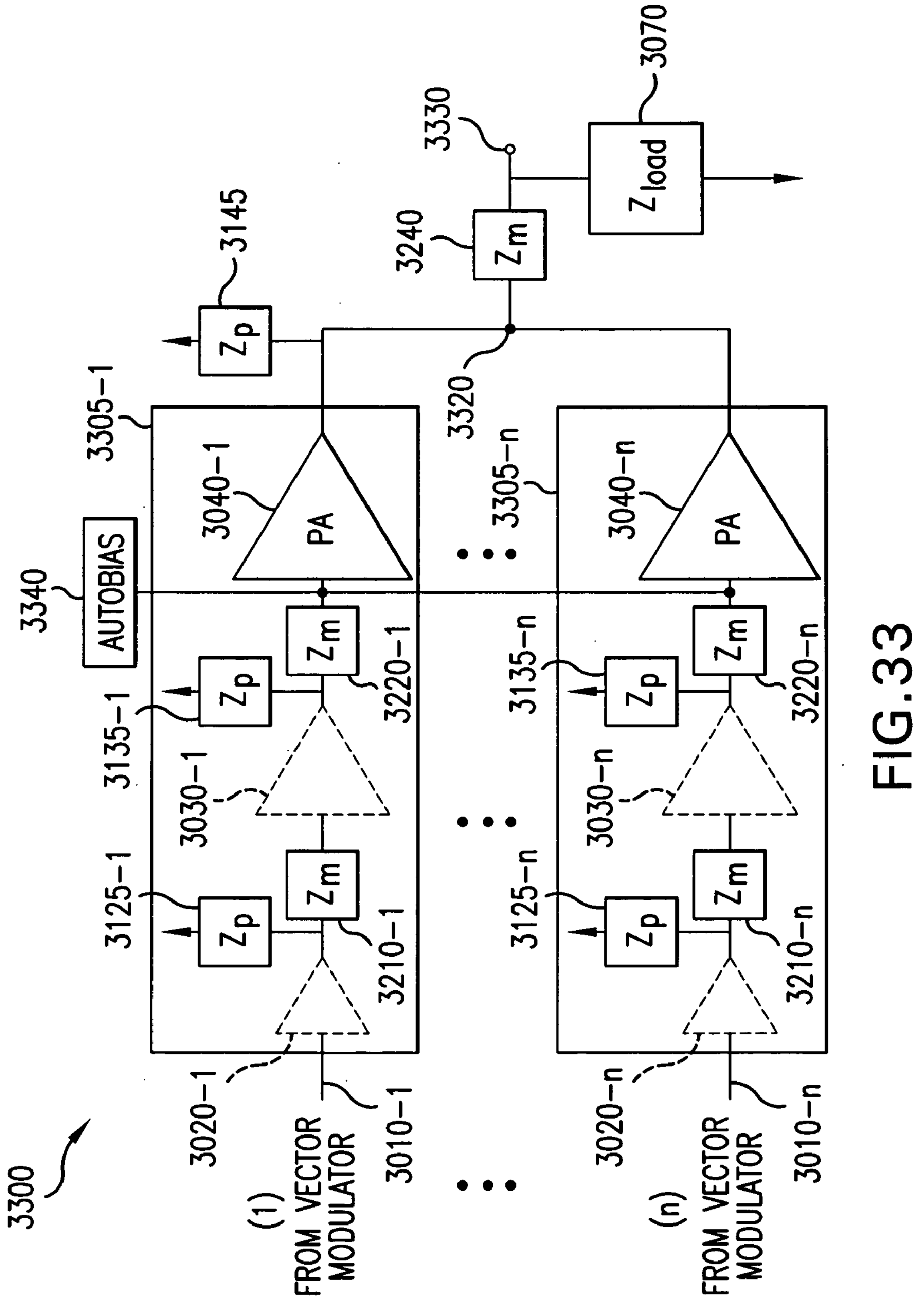
3300
(1)
FROM VECTOR
MODULATOR
3010-1
3020-1
3125-1
Zp
Zm
3210-1
3030-1
3135-1
Zp
Zm
3220-1
3340
AUTOBIAS
3040-1
PA
3305-1
Zp
3145
3320
Zm
3240
3330
Zload
3070
(n)
FROM VECTOR
MODULATOR
3010-n
3020-n
3125-n
Zp
Zm
3210-n
3030-n
3135-n
Zp
Zm
3220-n
3040-n
PA
3305-n

FIG.33

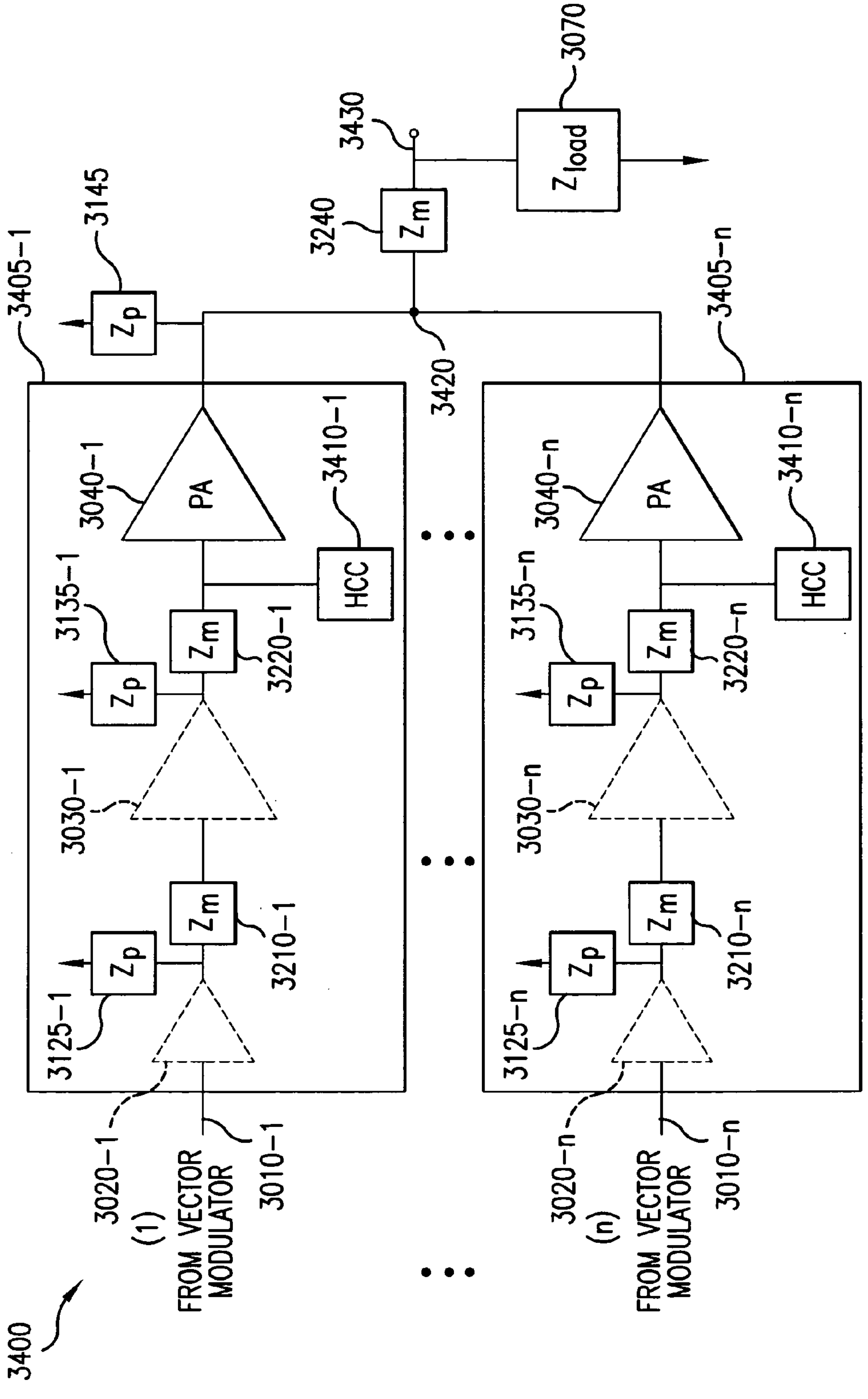
3400
FROM VECTOR MODULATOR
(1)
3020-1
3010-1
3125-1
Zp
Zm
3210-1
3030-1
Zp
3135-1
Zm
3220-1
HCC
3410-1
PA
3040-1
3405-1
Zp
3145
3420
Zm
3240
3430
Zload
3070
FROM VECTOR MODULATOR
(n)
3020-n
3010-n
3125-n
Zp
Zm
3210-n
3030-n
Zp
3135-n
Zm
3220-n
HCC
3410-n
PA
3040-n
3405-n

FIG.34

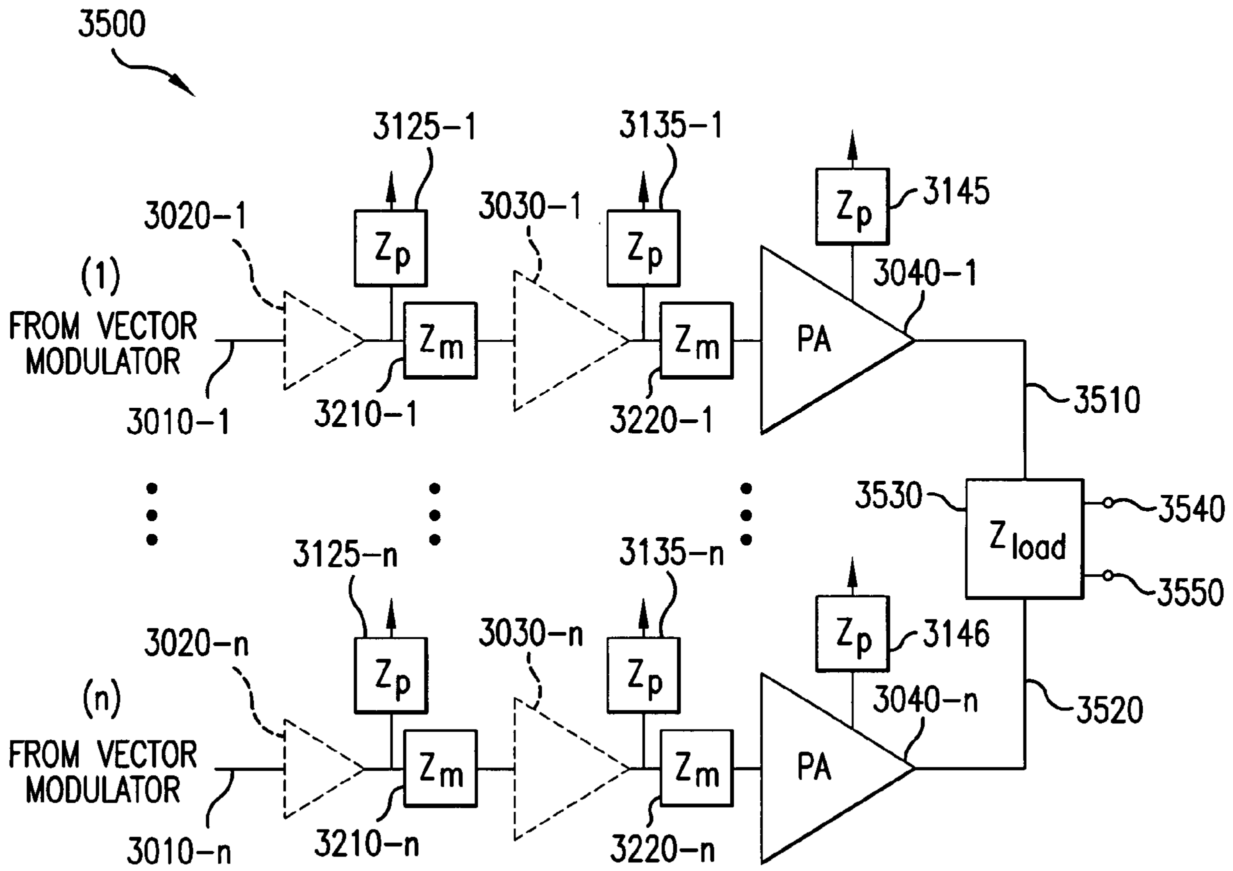
3500
(1)
FROM VECTOR MODULATOR
3010-1
3020-1
3125-1
Z_p
Z_m
3210-1
3030-1
3135-1
Z_p
Z_m
3220-1
PA
Z_p
3145
3040-1
3510
3530
Z_{load}
3540
3550
(n)
FROM VECTOR MODULATOR
3010-n
3020-n
3125-n
Z_p
Z_m
3210-n
3030-n
3135-n
Z_p
Z_m
3220-n
PA
Z_p
3146
3040-n
3520

FIG.35

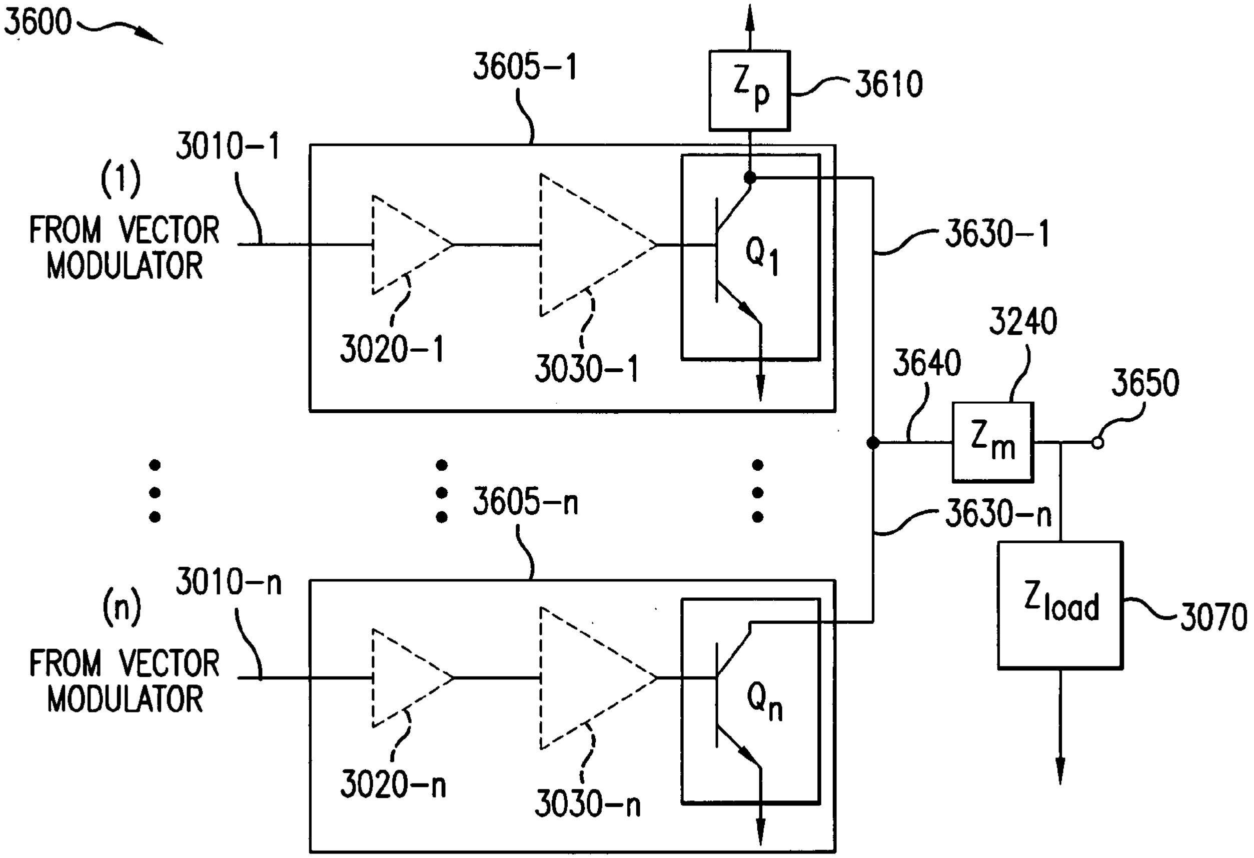
3600
3605-1
3610
Z_p
3010-1
(1)
FROM VECTOR MODULATOR
Q_1
3630-1
3020-1
3030-1
3240
3640
3650
Z_m
3605-n
3630-n
3010-n
(n)
FROM VECTOR MODULATOR
Q_n
Z_{load}
3070
3020-n
3030-n

FIG.36

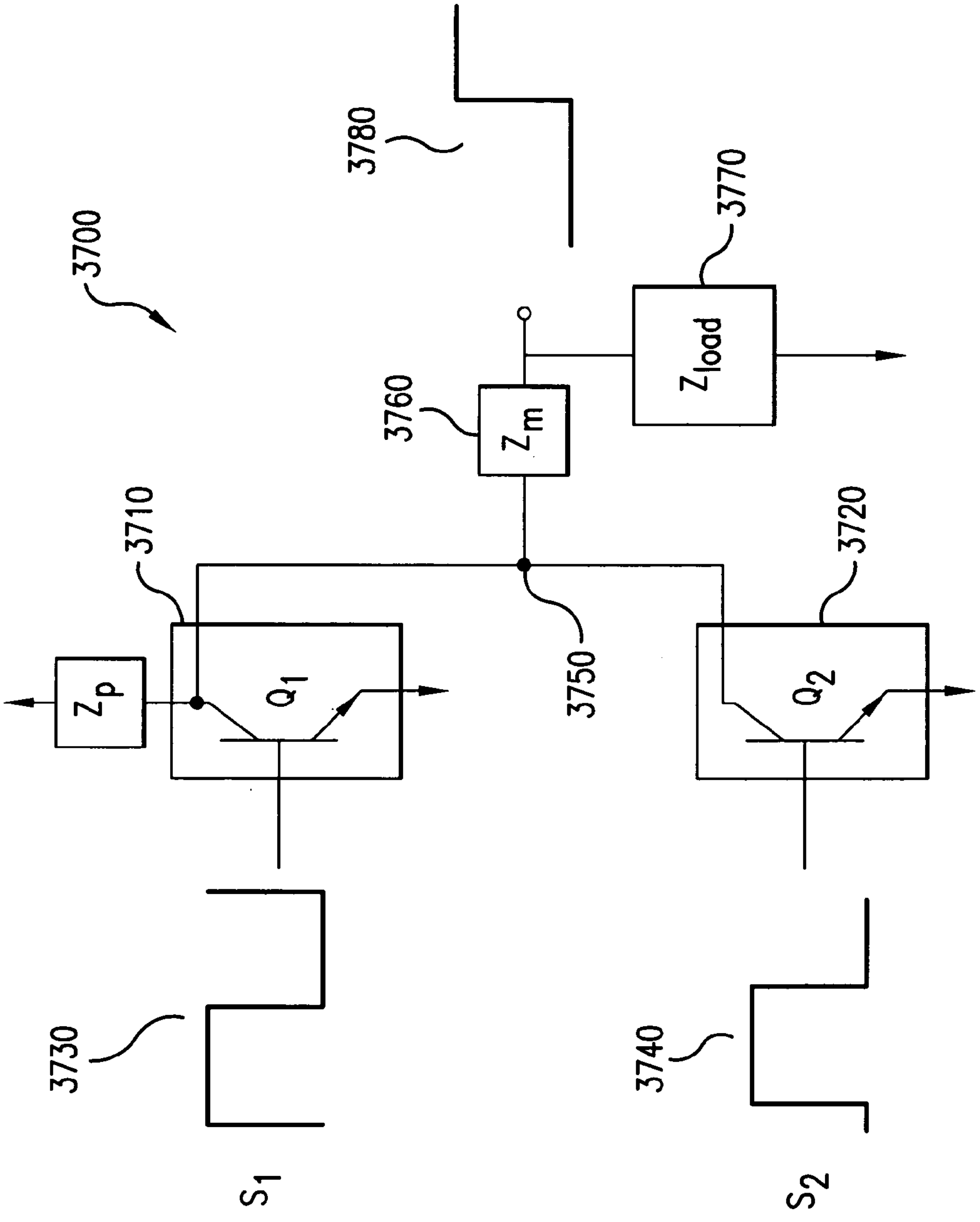
3700
3780
3770
Z_{load}
3760
Z_m
3710
Z_p
Q_1
3750
3720
Q_2
3730
3740
S_1
S_2

FIG.37

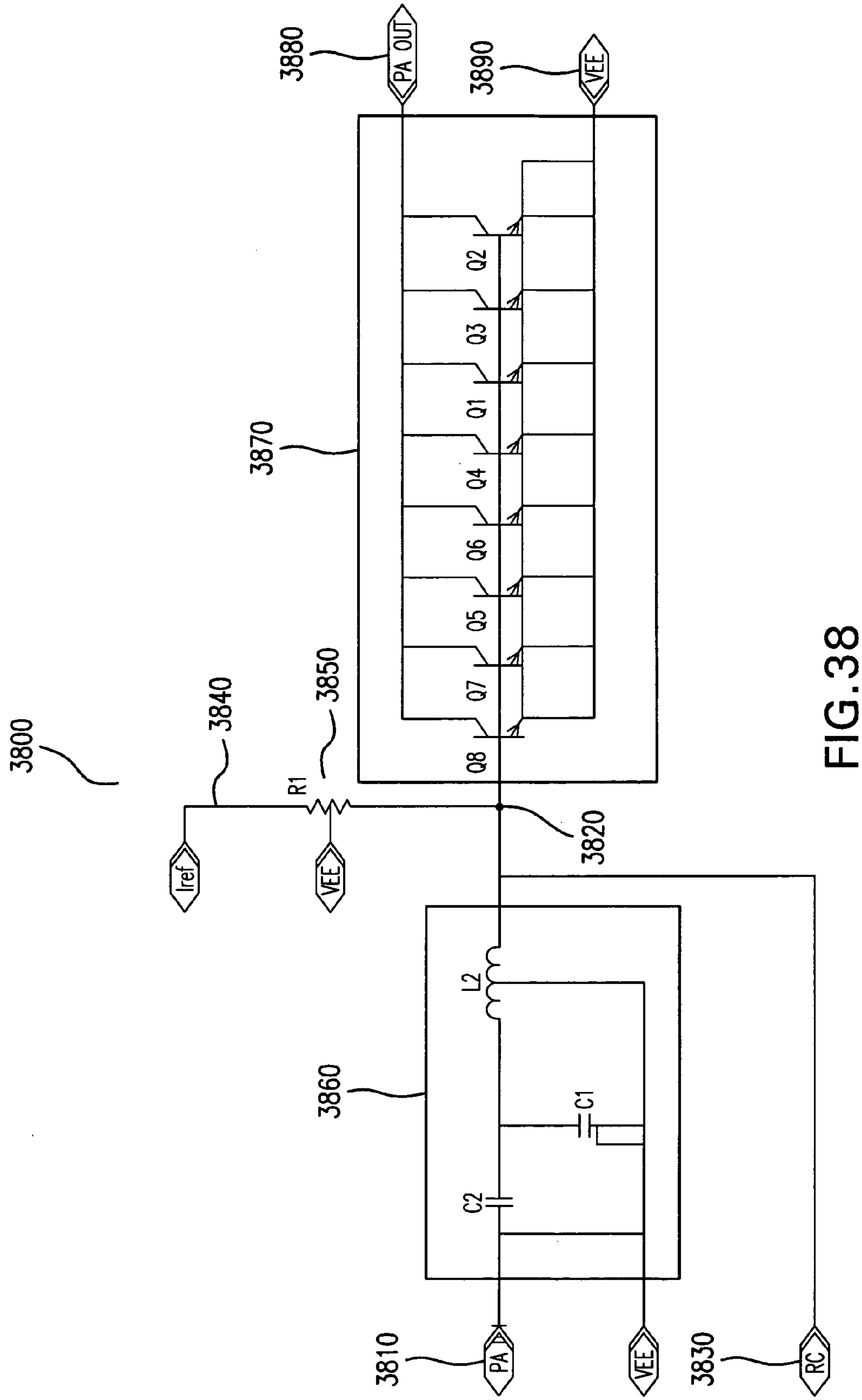
3880
PA OUT
3890
VEE
3870
Q2
Q3
Q1
Q4
Q6
Q5
Q7
Q8
3800
3840
3850
R1
Iref
VEE
3820
3860
L2
C1
C2
3810
PA
VEE
3830
RC

FIG.38

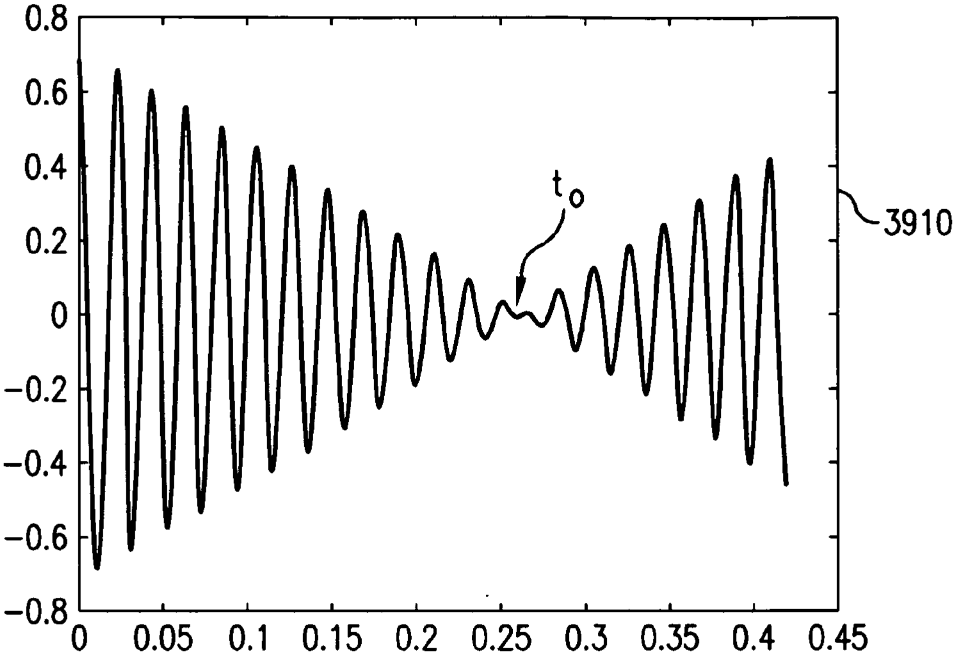
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
t_0
3910

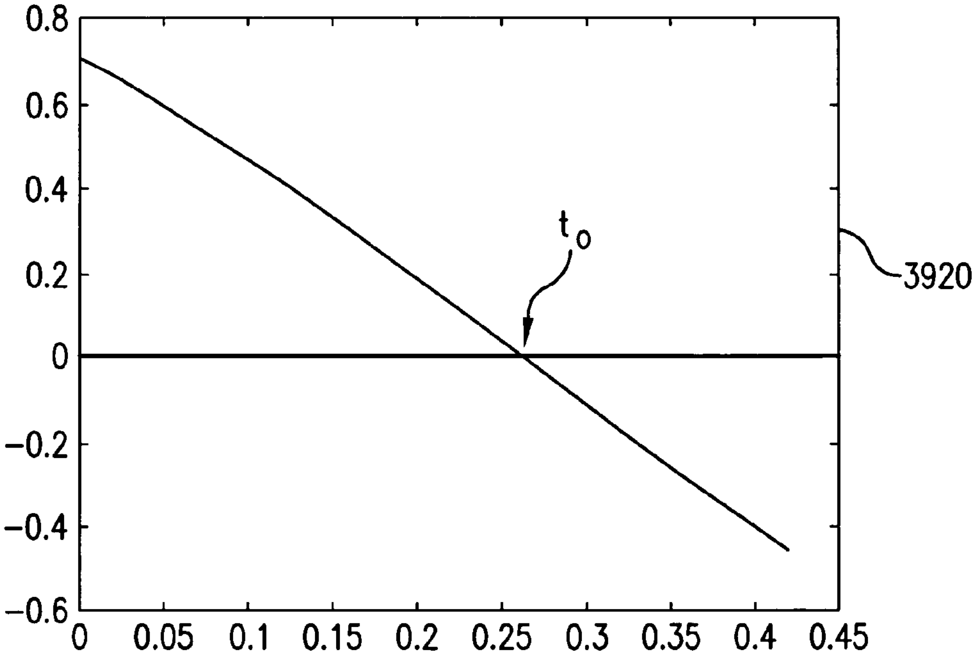
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
t_0
3920

FIG.39

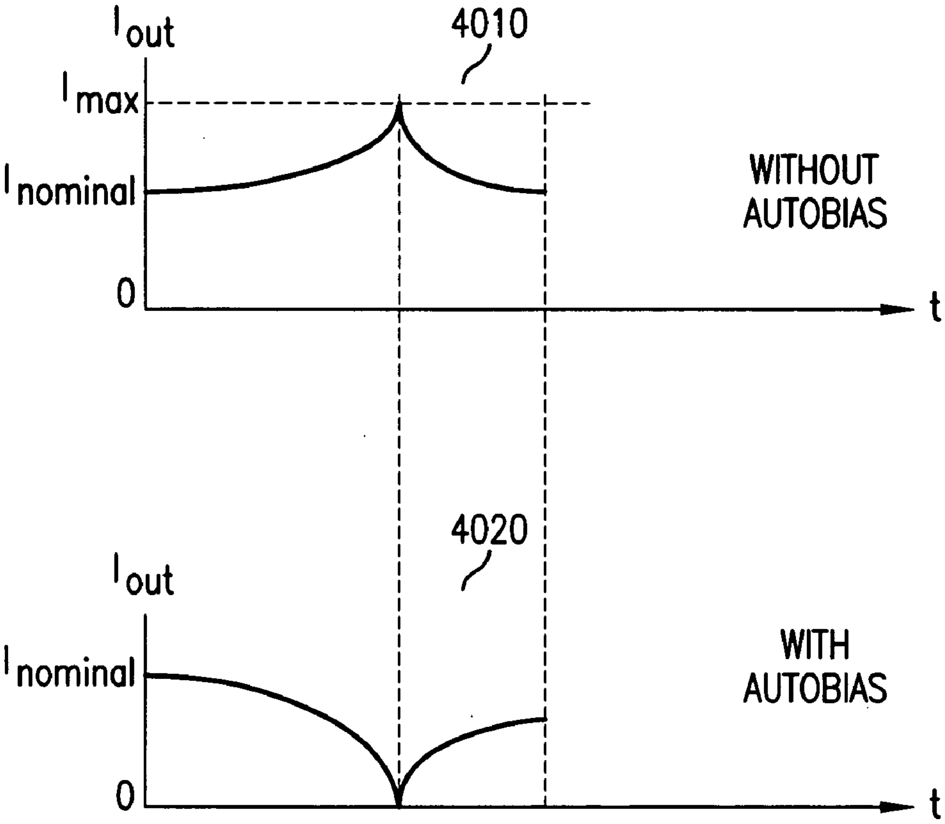
I out
4010
I max
I nominal
0
t
WITHOUT AUTOBIAS
4020
I out
I nominal
0
t
WITH AUTOBIAS

FIG.40

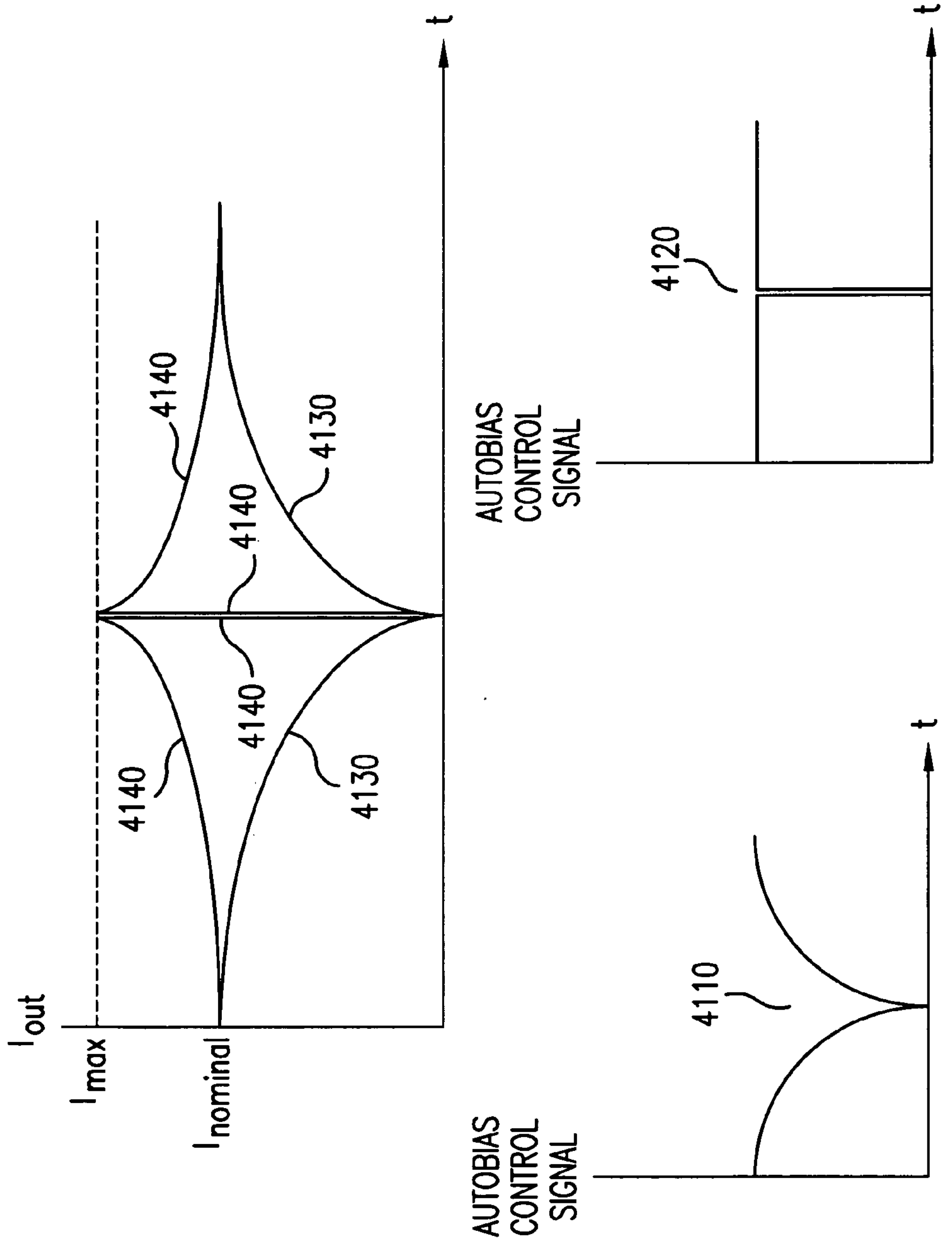
I_{out}
I_{max}
$I_{nominal}$
4140
4140
4140
4140
4130
4130
t
AUTOBIAS CONTROL SIGNAL
4110
t
AUTOBIAS CONTROL SIGNAL
4120
t

FIG.41

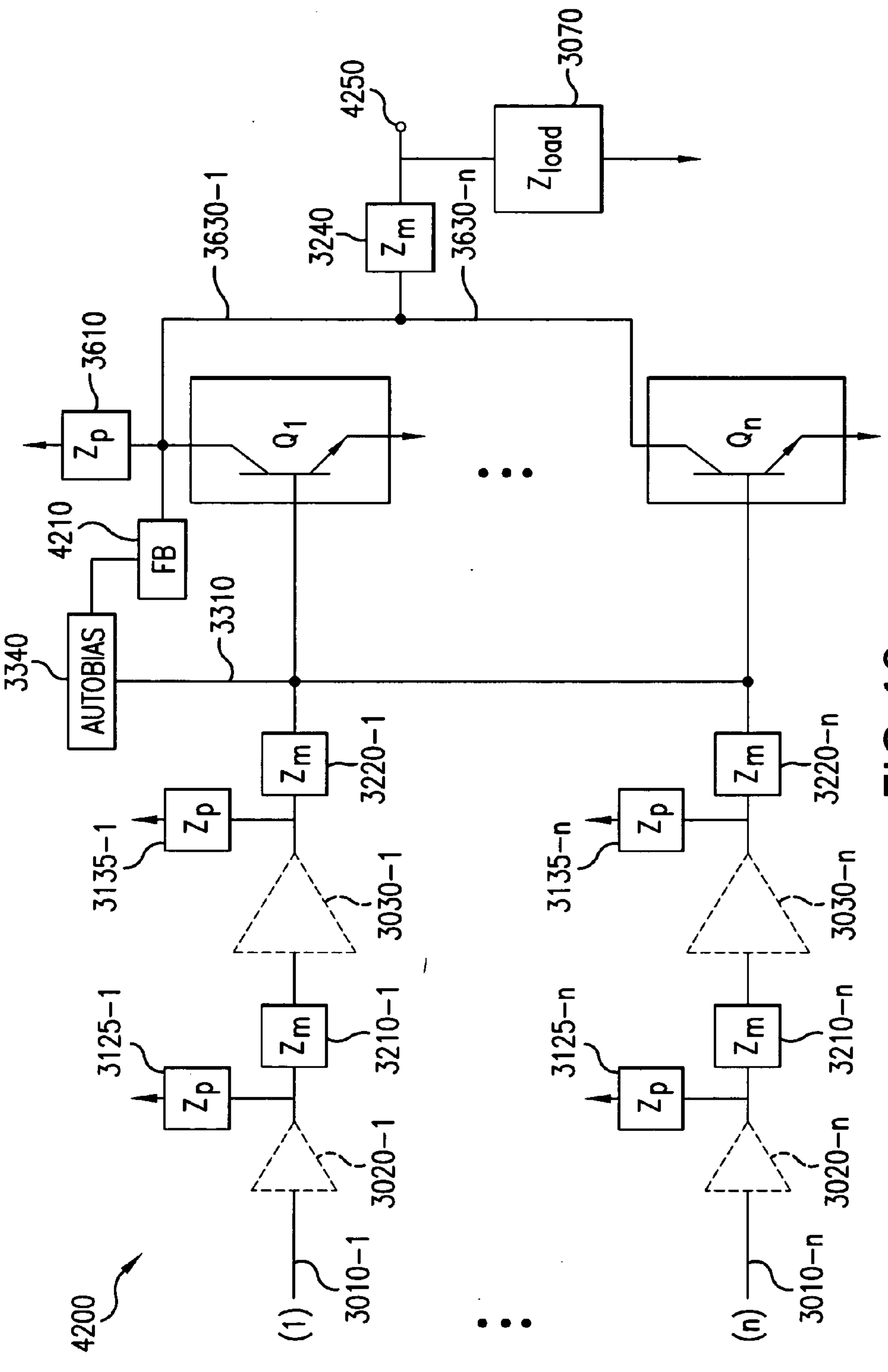
4200
(1)
3010-1
3020-1
Zp
3125-1
Zm
3210-1
3030-1
Zp
3135-1
Zm
3220-1
(n)
3010-n
3020-n
Zp
3125-n
Zm
3210-n
3030-n
Zp
3135-n
Zm
3220-n
3340
AUTOBIAS
3310
FB
4210
Zp
3610
Q1
Qn
3630-1
3240
Zm
3630-n
4250
Zload
3070

FIG.42

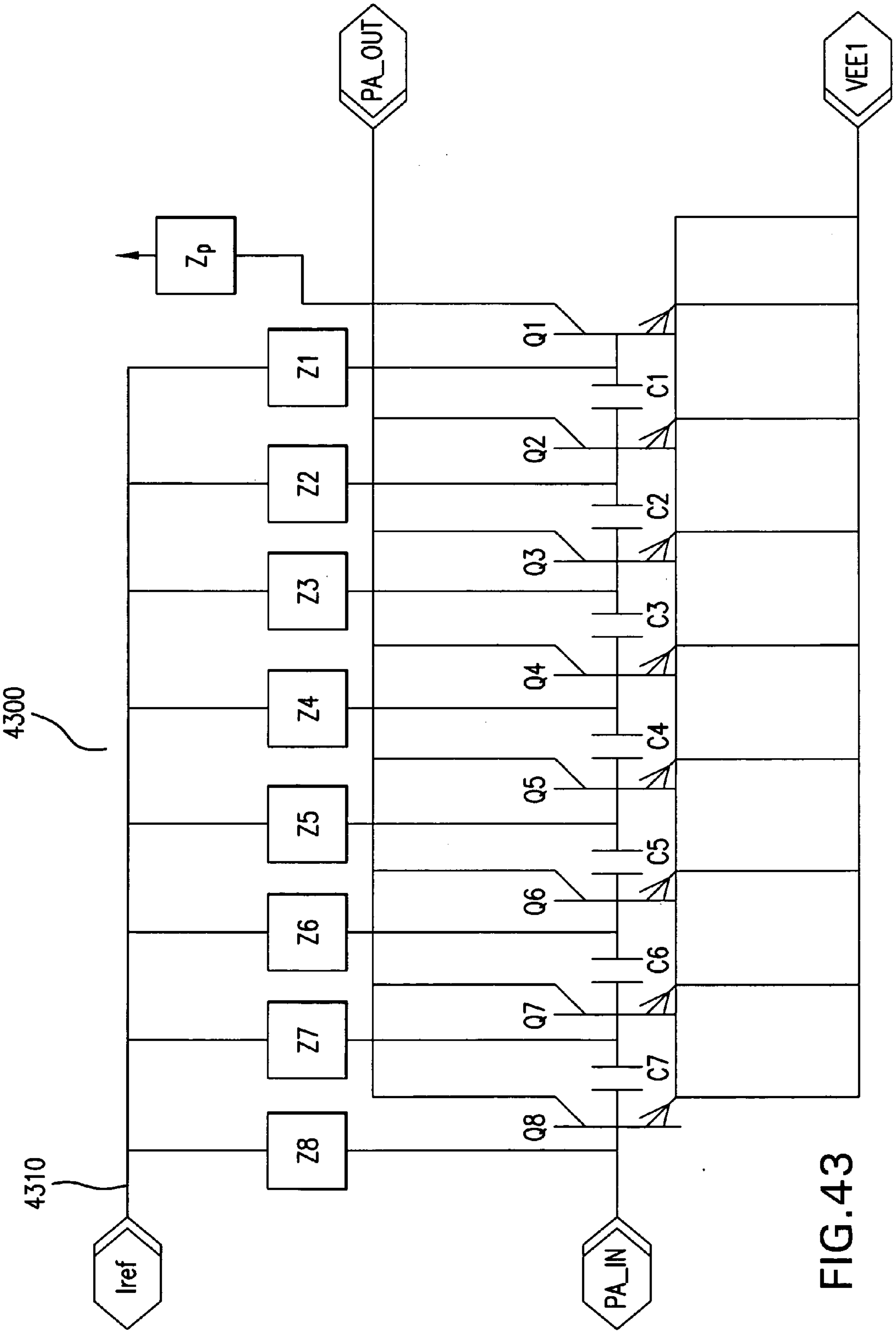
PA_OUT
VEE1
Zp
Z1
Z2
Z3
Z4
Z5
Z6
Z7
Z8
Q1
Q2
Q3
Q4
Q5
Q6
Q7
Q8
C1
C2
C3
C4
C5
C6
C7
4300
4310
Iref
PA_IN

FIG.43

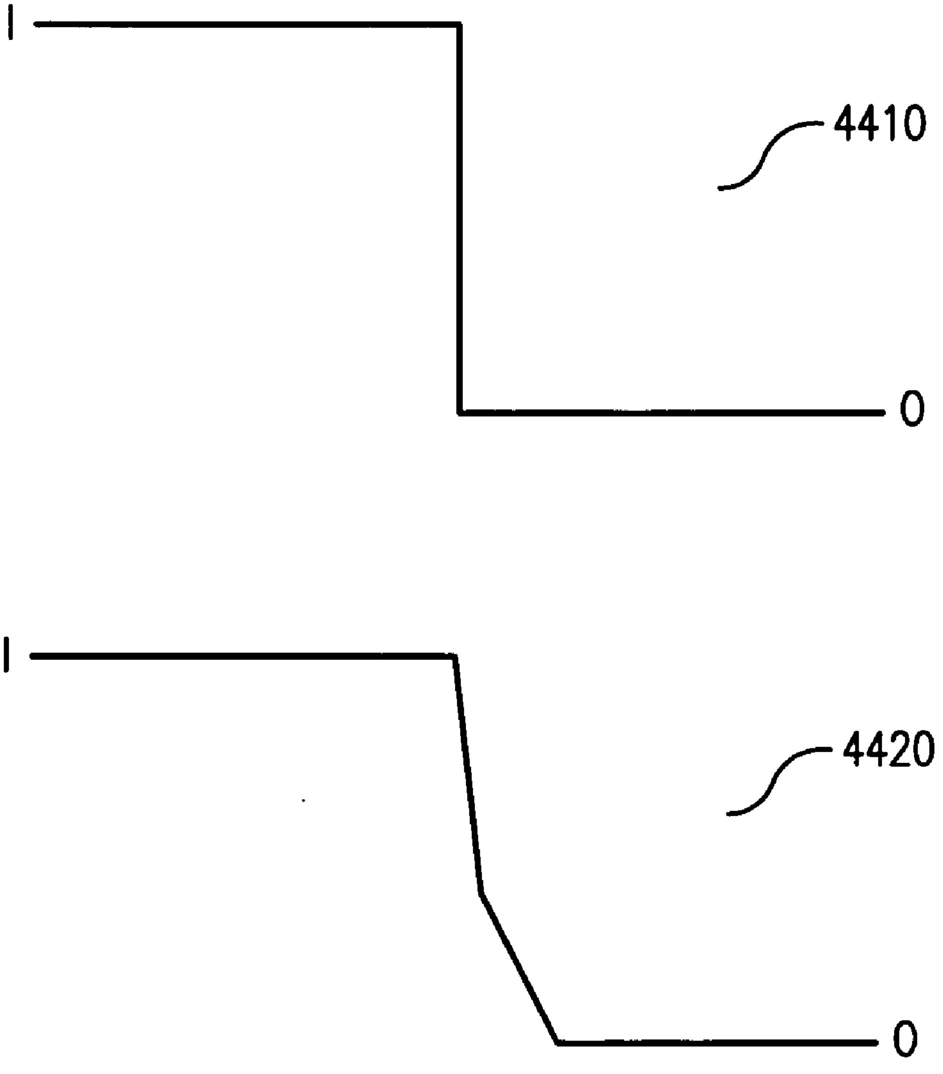
I
4410
0
I
4420
0

FIG.44

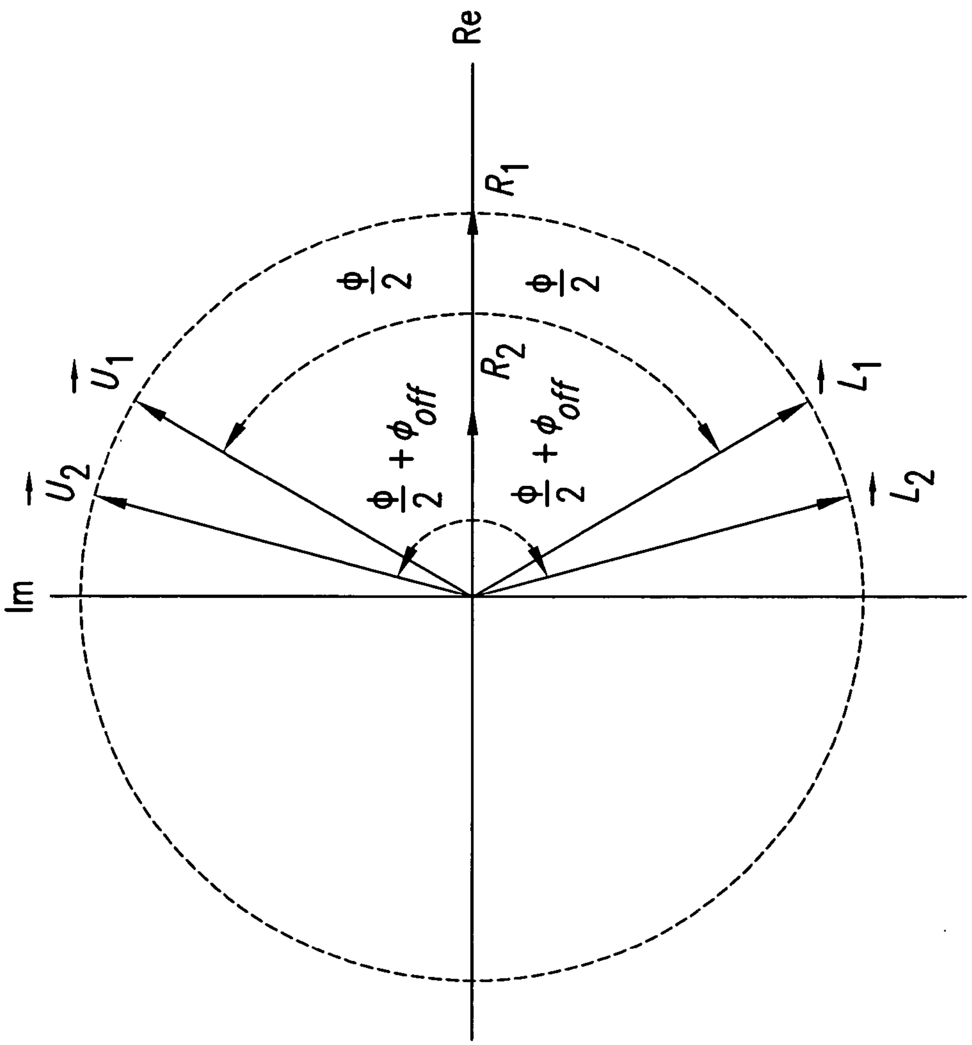
Re
Im
$\vec{R_1}$
$\vec{R_2}$
$\vec{U_1}$
$\vec{U_2}$
$\vec{L_1}$
$\vec{L_2}$
$\frac{\phi}{2}$
$\frac{\phi}{2}$
$\frac{\phi}{2}+\phi_{off}$
$\frac{\phi}{2}+\phi_{off}$

FIG.45

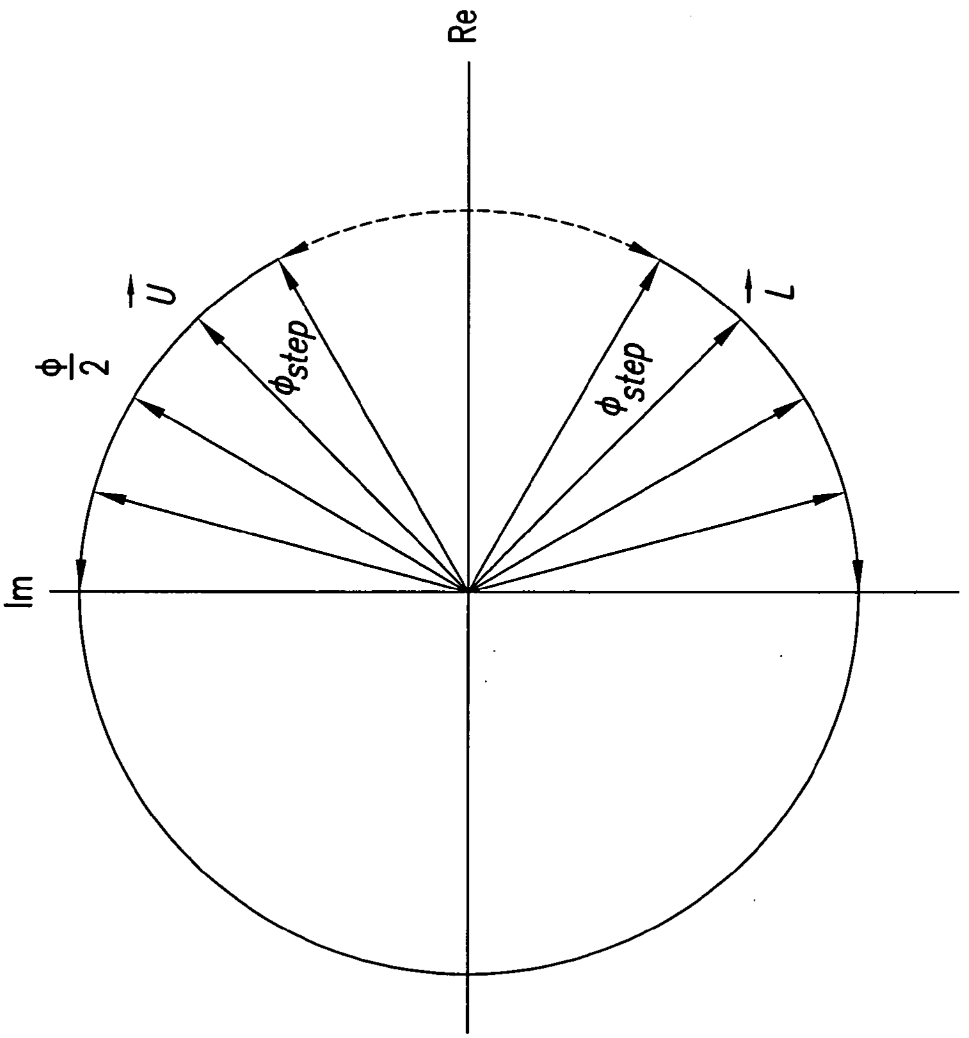
Re
Im
$\frac{\phi}{2}$
$\vec{U}$
ϕ_{step}
ϕ_{step}
$\vec{L}$

FIG.46

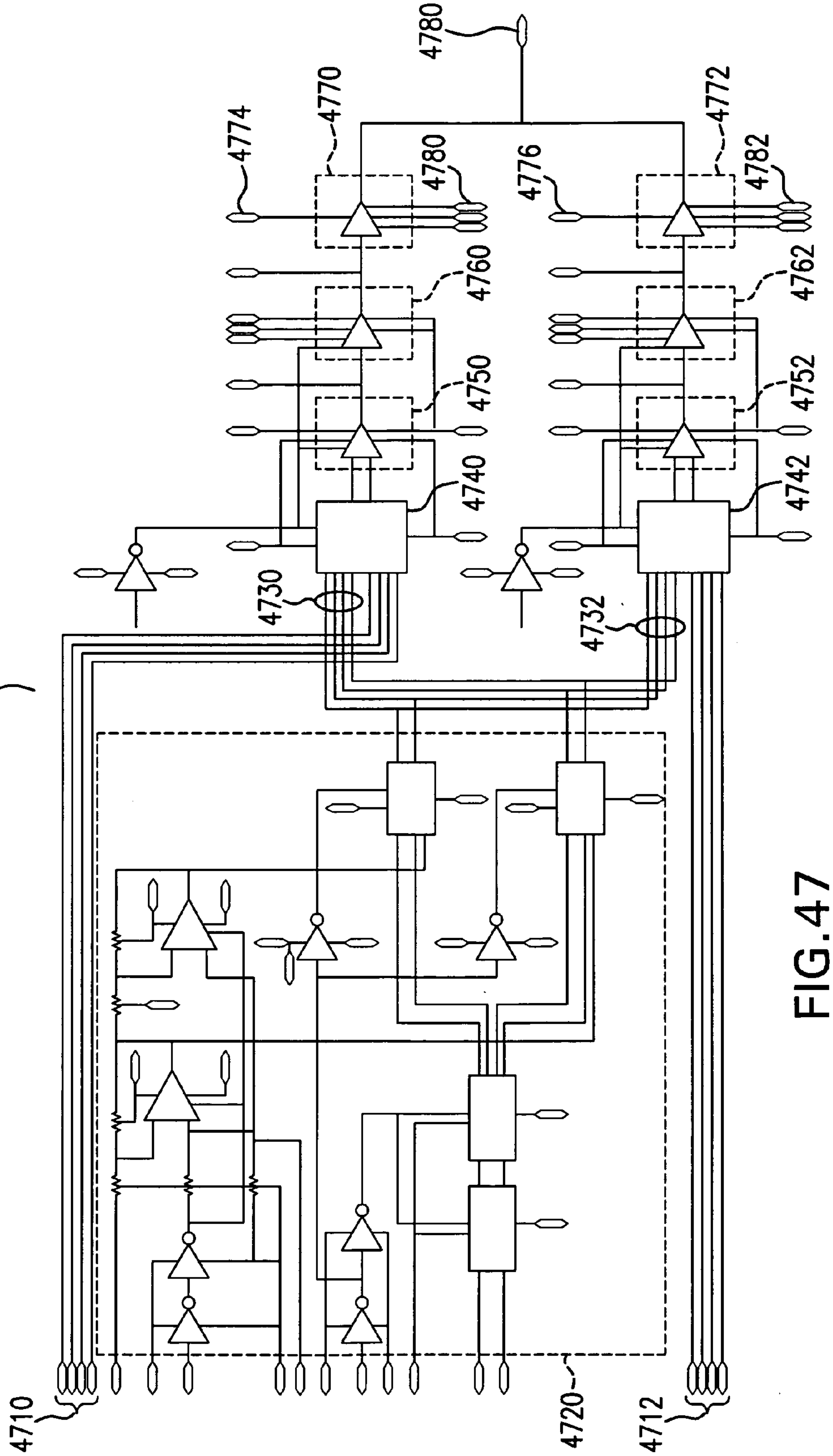

FIG.47

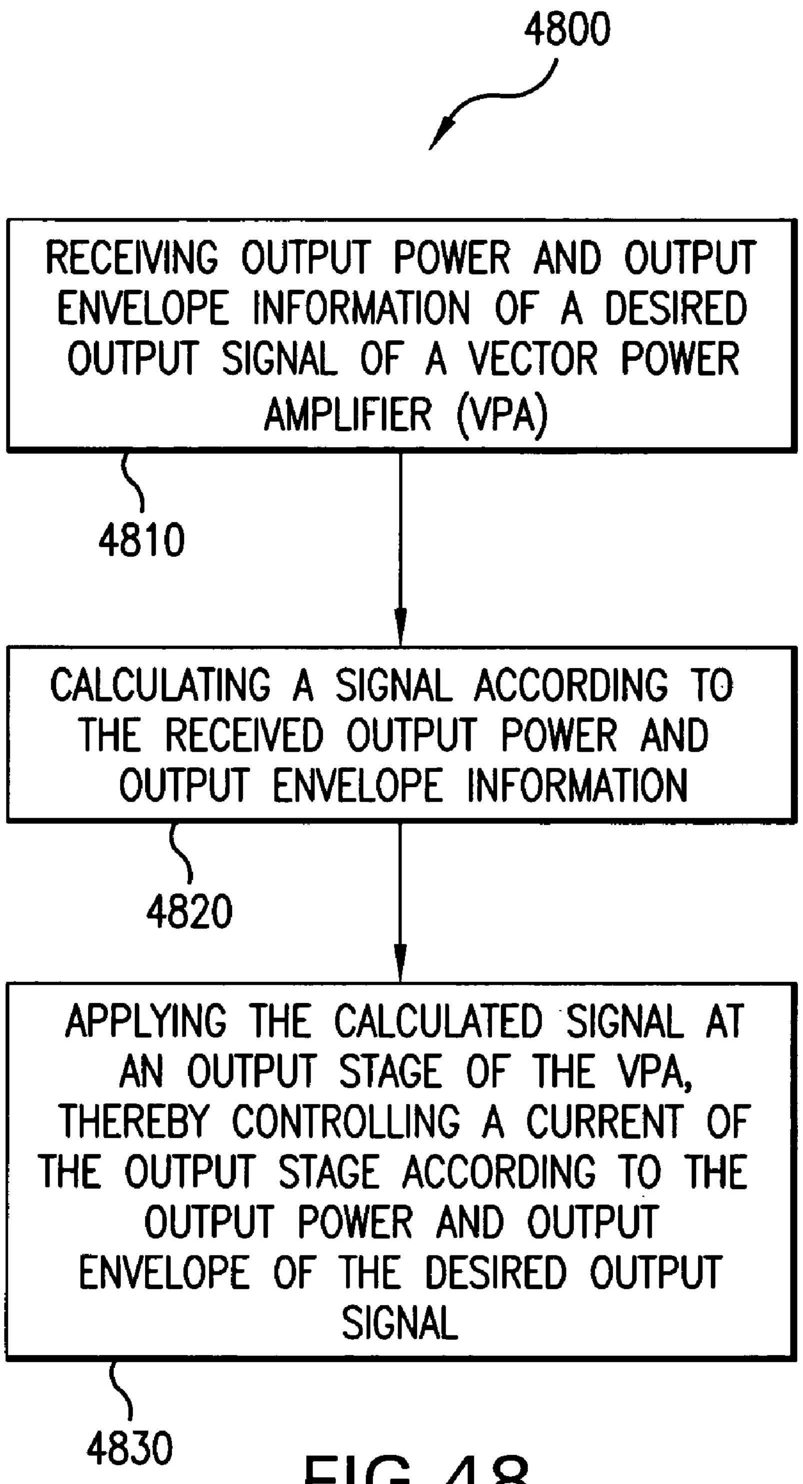
4800
RECEIVING OUTPUT POWER AND OUTPUT ENVELOPE INFORMATION OF A DESIRED OUTPUT SIGNAL OF A VECTOR POWER AMPLIFIER (VPA)
4810
CALCULATING A SIGNAL ACCORDING TO THE RECEIVED OUTPUT POWER AND OUTPUT ENVELOPE INFORMATION
4820
APPLYING THE CALCULATED SIGNAL AT AN OUTPUT STAGE OF THE VPA, THEREBY CONTROLLING A CURRENT OF THE OUTPUT STAGE ACCORDING TO THE OUTPUT POWER AND OUTPUT ENVELOPE OF THE DESIRED OUTPUT SIGNAL
4830

FIG.48

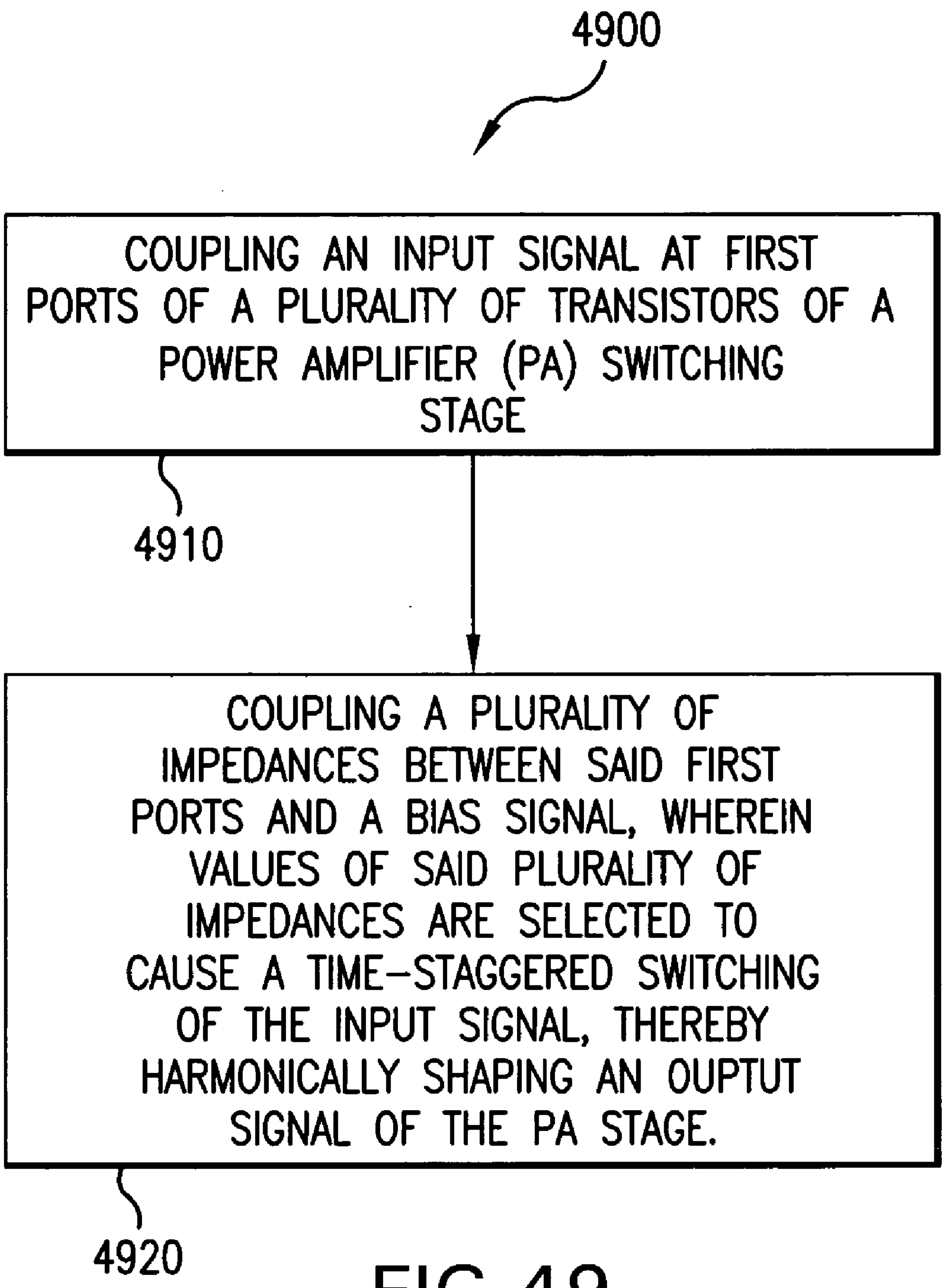
COUPLING AN INPUT SIGNAL AT FIRST PORTS OF A PLURALITY OF TRANSISTORS OF A POWER AMPLIFIER (PA) SWITCHING STAGE
COUPLING A PLURALITY OF IMPEDANCES BETWEEN SAID FIRST PORTS AND A BIAS SIGNAL, WHEREIN VALUES OF SAID PLURALITY OF IMPEDANCES ARE SELECTED TO CAUSE A TIME-STAGGERED SWITCHING OF THE INPUT SIGNAL, THEREBY HARMONICALLY SHAPING AN OUPTUT SIGNAL OF THE PA STAGE.

FIG.49

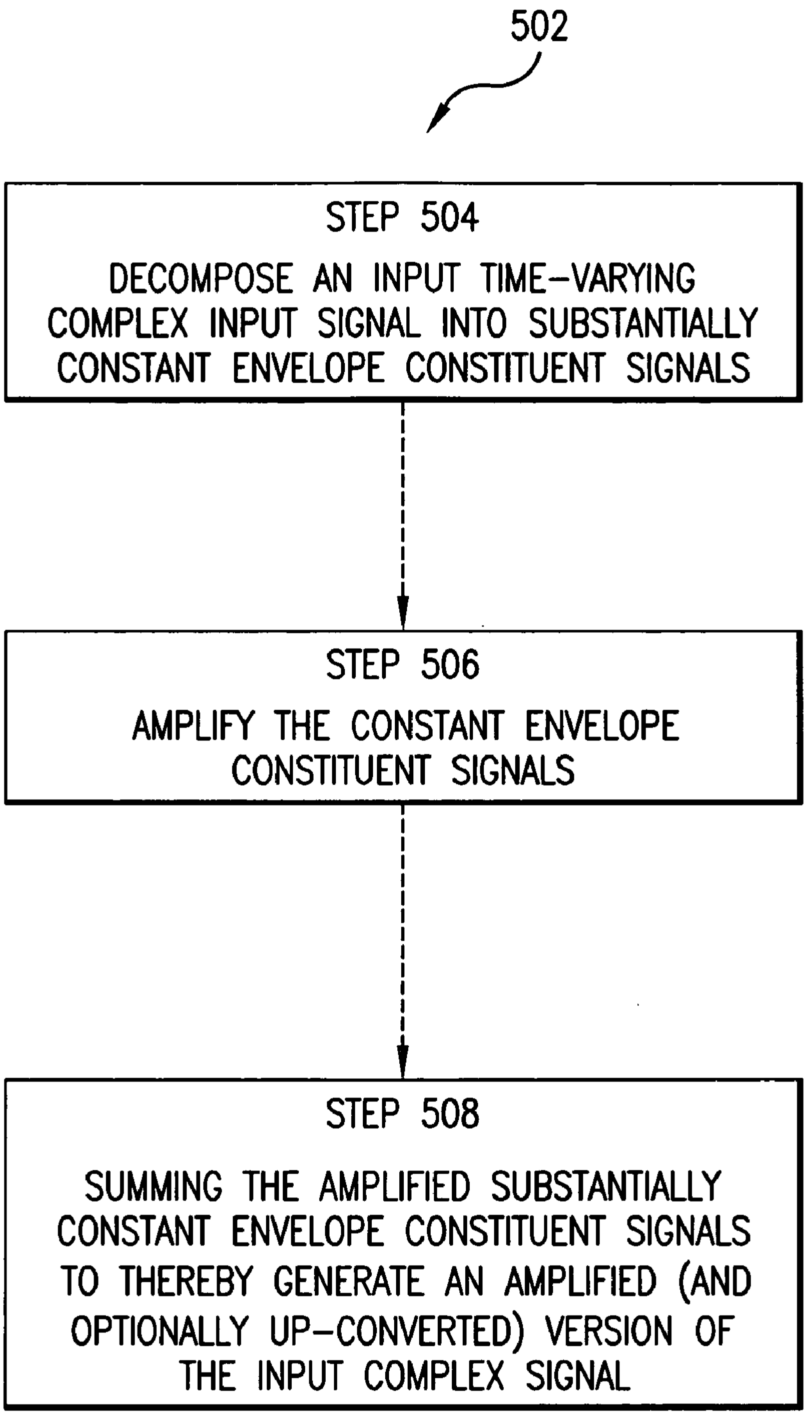
502
STEP 504
DECOMPOSE AN INPUT TIME-VARYING COMPLEX INPUT SIGNAL INTO SUBSTANTIALLY CONSTANT ENVELOPE CONSTITUENT SIGNALS
STEP 506
AMPLIFY THE CONSTANT ENVELOPE CONSTITUENT SIGNALS
STEP 508
SUMMING THE AMPLIFIED SUBSTANTIALLY CONSTANT ENVELOPE CONSTITUENT SIGNALS TO THEREBY GENERATE AN AMPLIFIED (AND OPTIONALLY UP-CONVERTED) VERSION OF THE INPUT COMPLEX SIGNAL

FIG.50

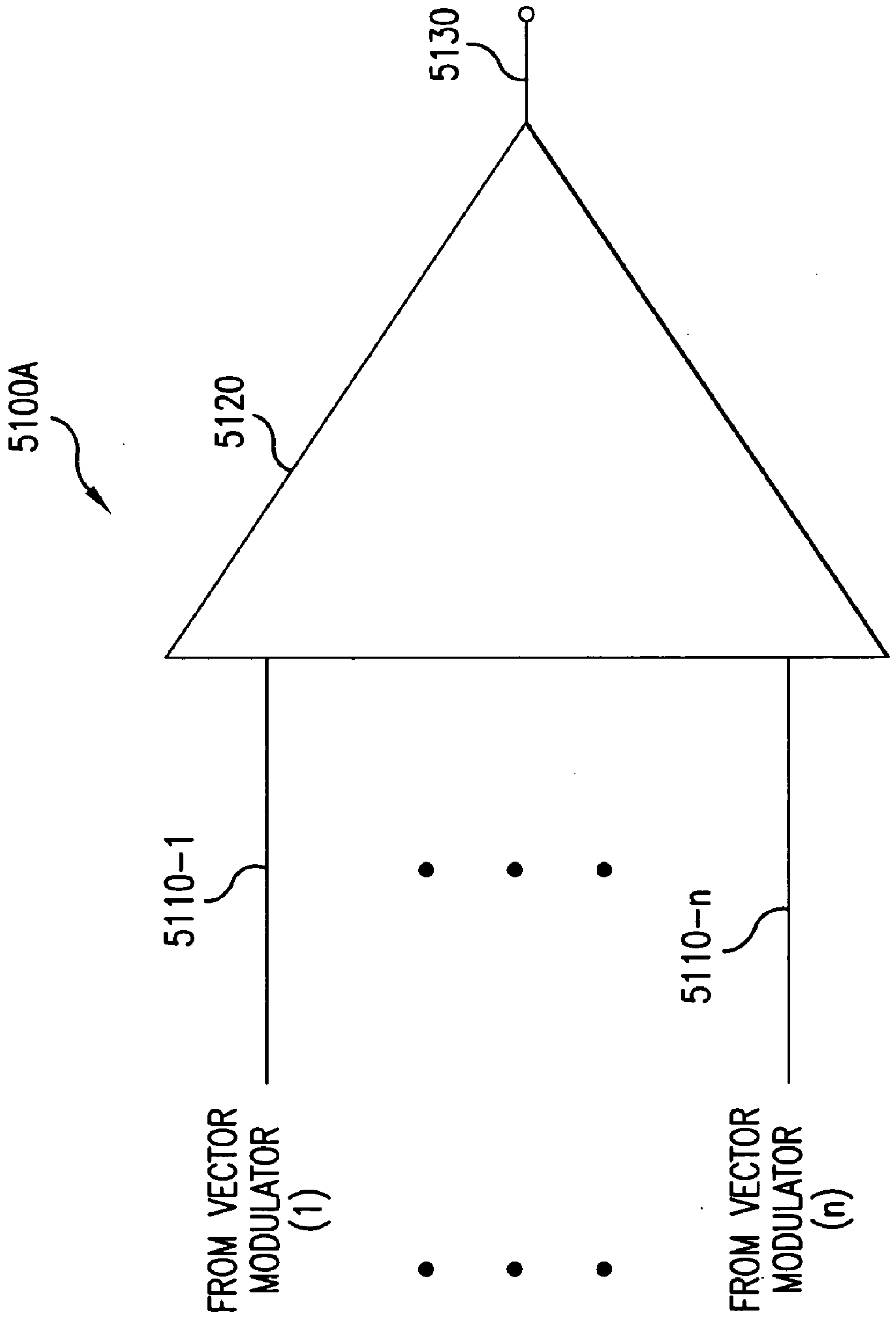
5100A
5120
5130
5110-1
5110-n
FROM VECTOR MODULATOR (1)
FROM VECTOR MODULATOR (n)

FIG.51A

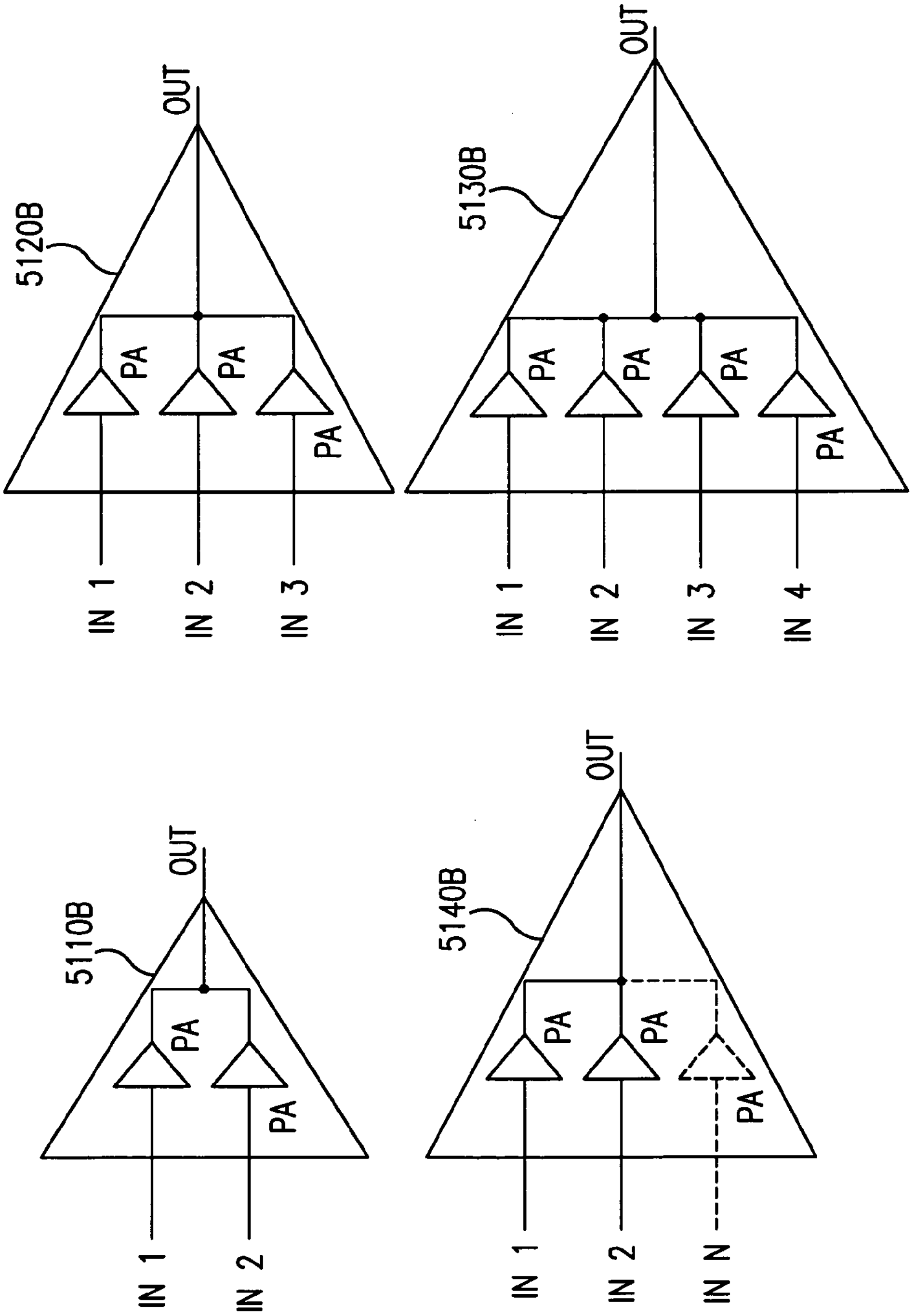
5120B
OUT
PA
PA
PA
IN 1
IN 2
IN 3
5130B
OUT
PA
PA
PA
PA
IN 1
IN 2
IN 3
IN 4
5110B
OUT
PA
PA
IN 1
IN 2
5140B
OUT
PA
PA
PA
IN 1
IN 2
IN N

FIG.51B

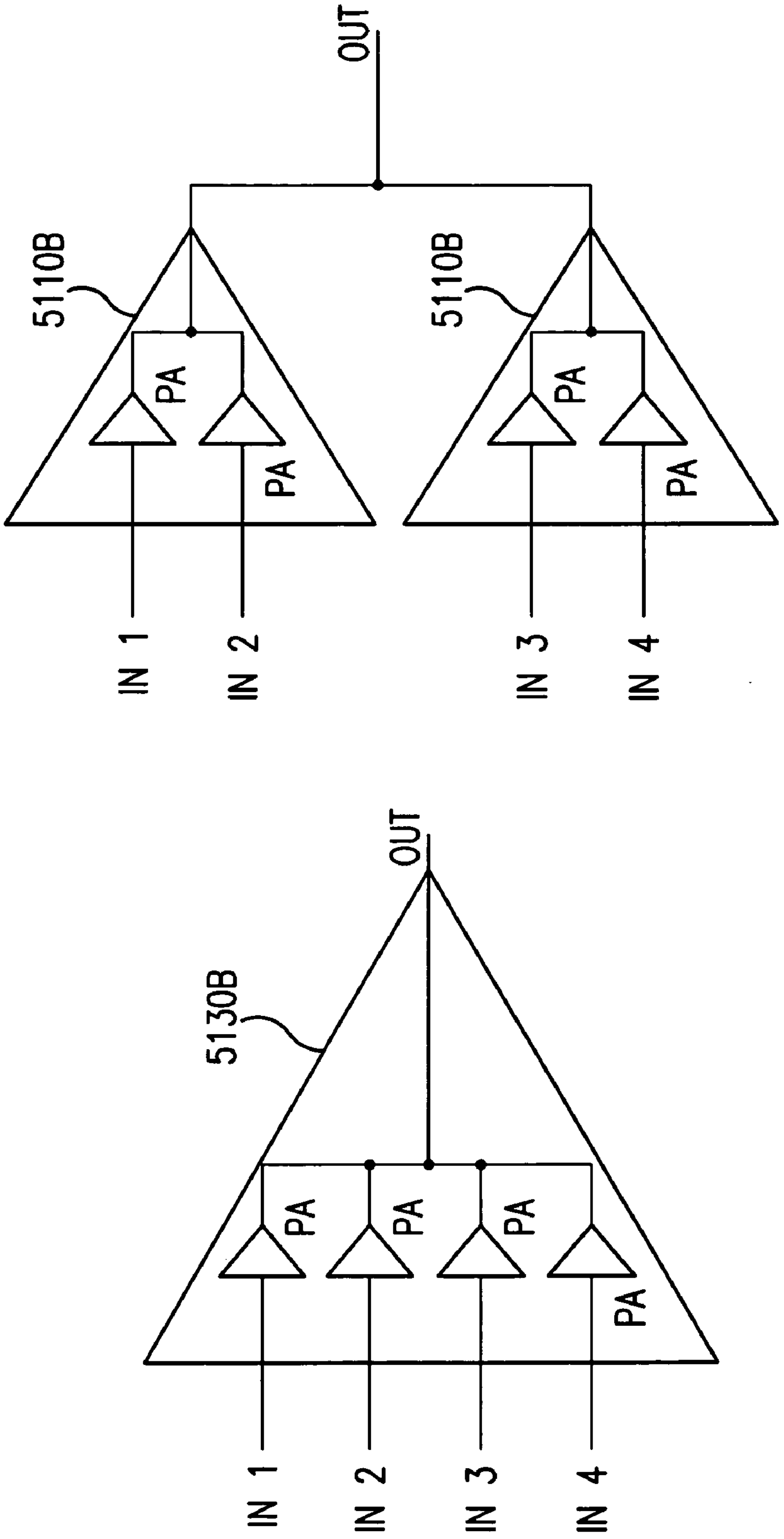
OUT
5110B
PA
PA
IN 1
IN 2
5110B
PA
PA
IN 3
IN 4
OUT
5130B
PA
PA
PA
PA
IN 1
IN 2
IN 3
IN 4

FIG.51C

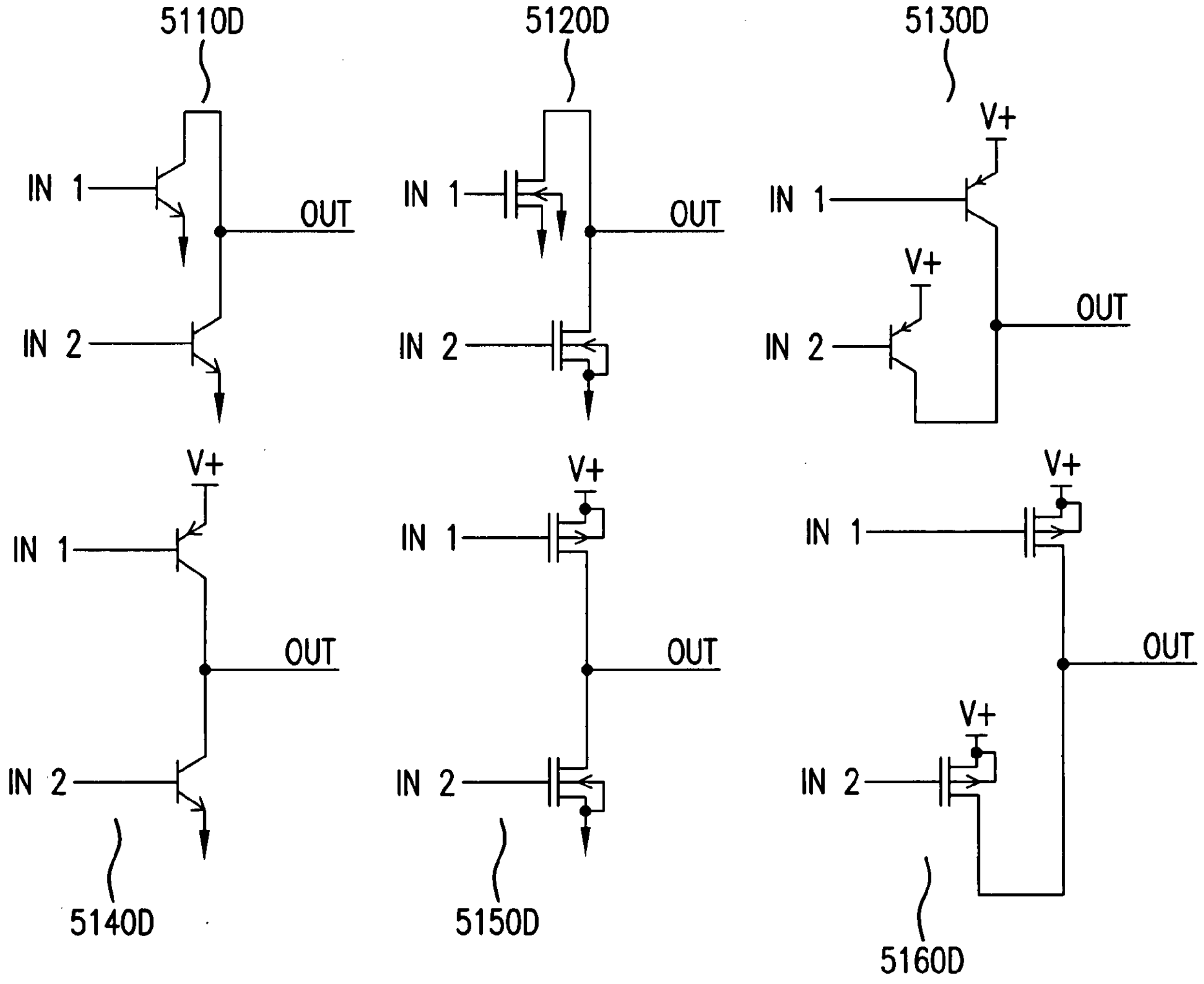
5110D
IN 1
IN 2
OUT
5120D
IN 1
IN 2
OUT
5130D
V+
IN 1
V+
IN 2
OUT
V+
IN 1
IN 2
OUT
5140D
V+
IN 1
IN 2
OUT
5150D
V+
IN 1
OUT
V+
IN 2
5160D

FIG.51D

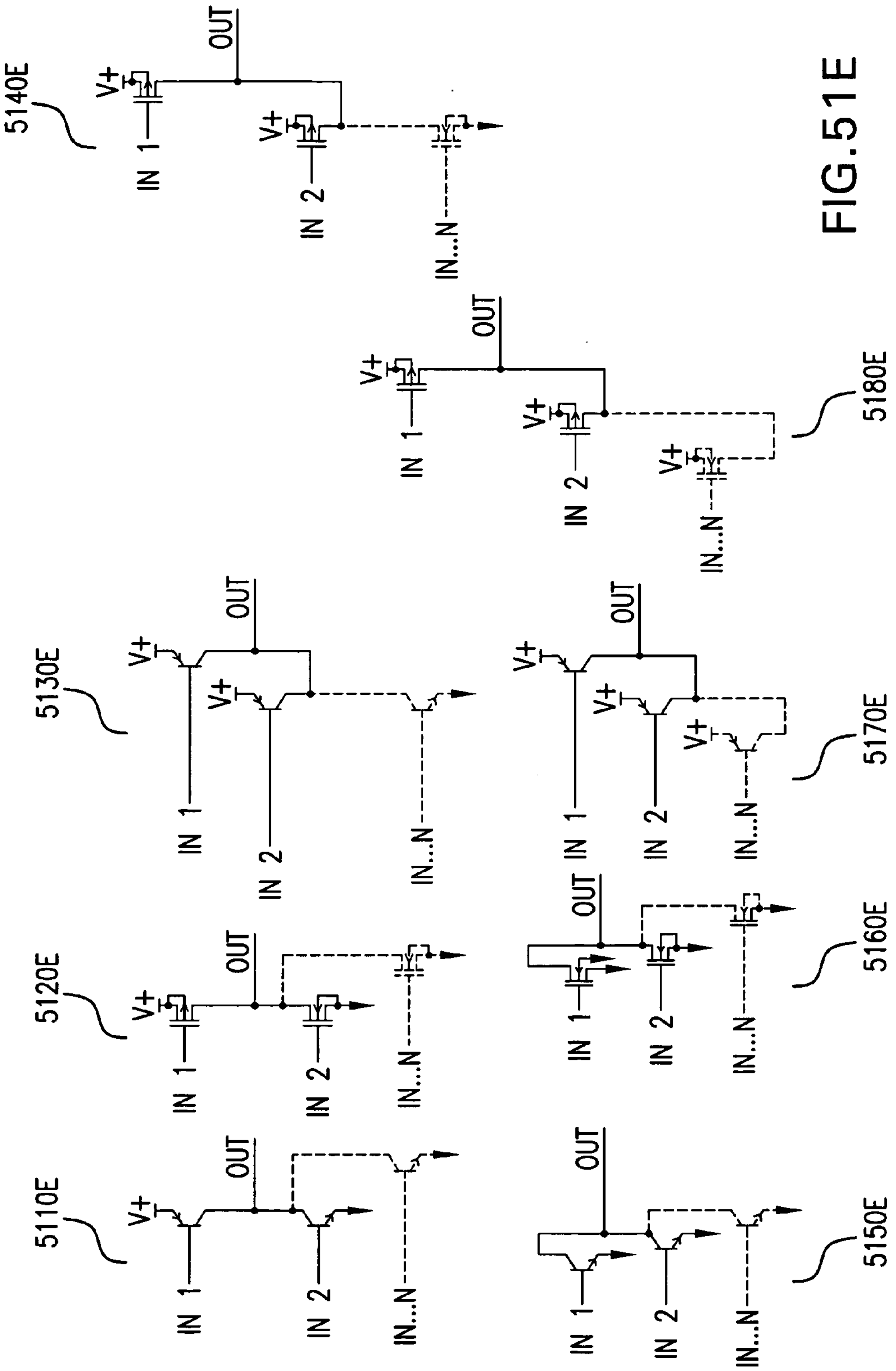
V+
OUT
IN 1
IN 2
IN...N

FIG.51E

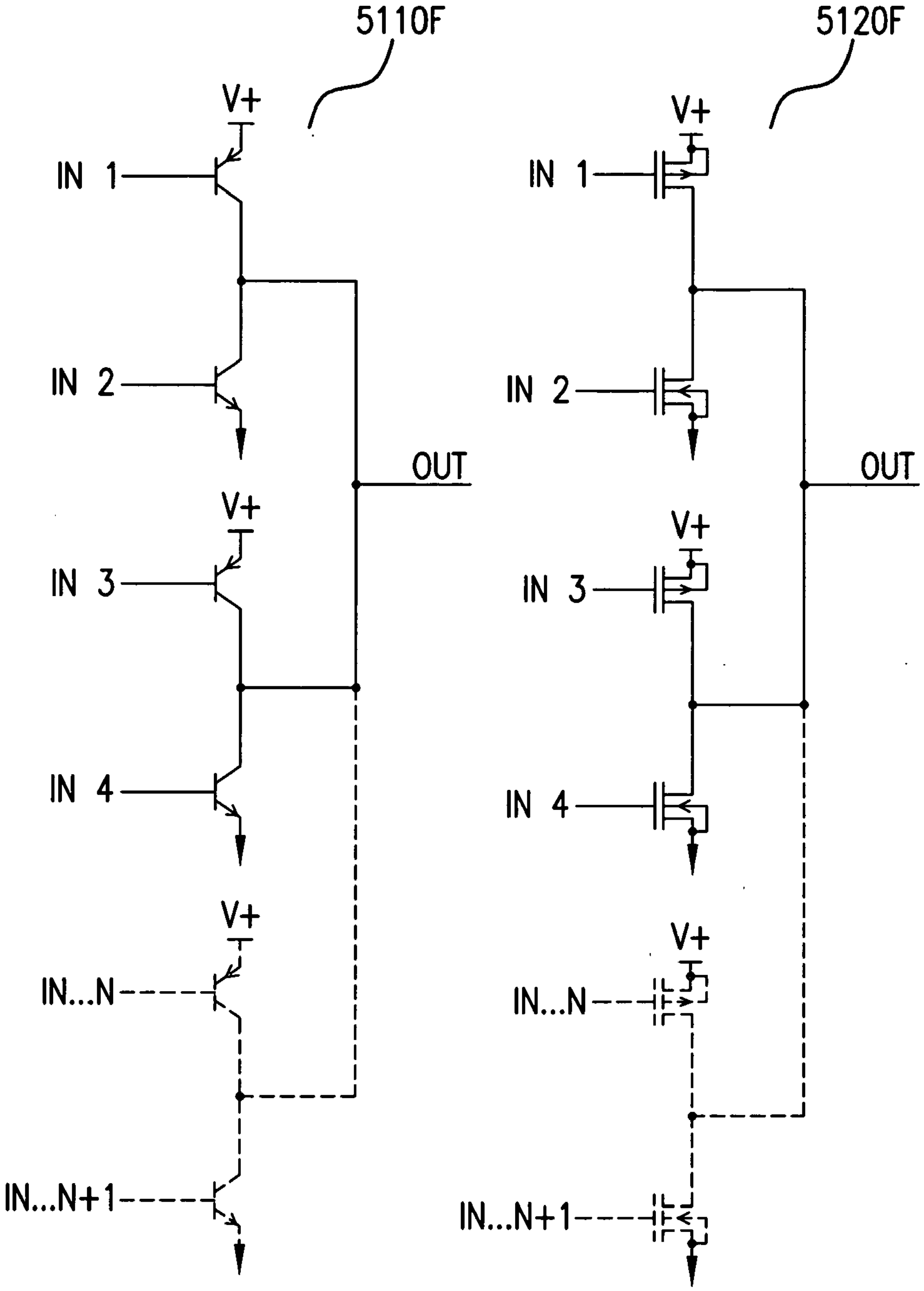
5110F
V+
IN 1
IN 2
OUT
V+
IN 3
IN 4
V+
IN...N
IN...N+1
5120F
V+
IN 1
IN 2
OUT
V+
IN 3
IN 4
V+
IN...N
IN...N+1

FIG.51F

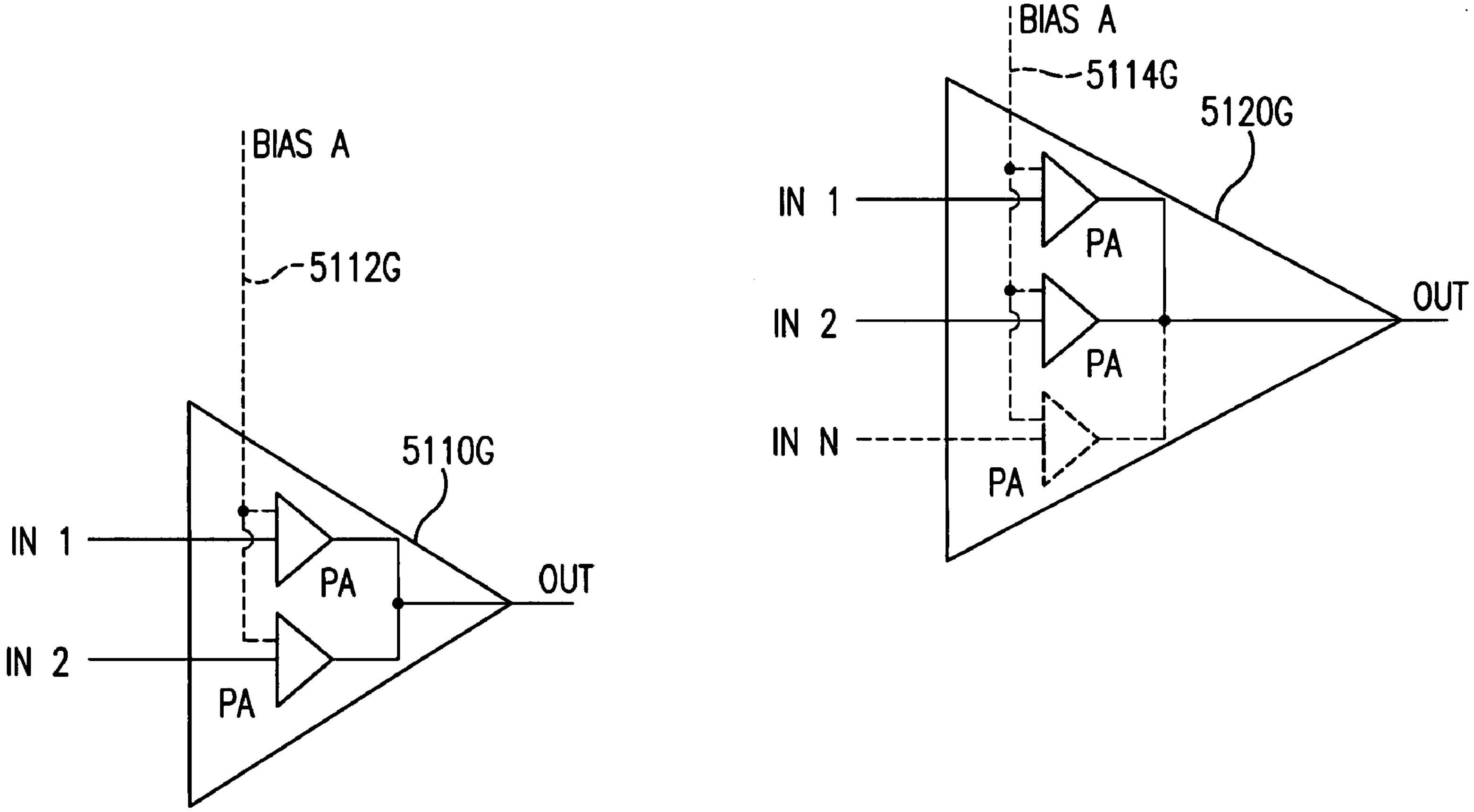
BIAS A
5112G
5110G
IN 1
IN 2
PA
PA
OUT
BIAS A
5114G
5120G
IN 1
IN 2
IN N
PA
PA
PA
OUT

FIG.51G

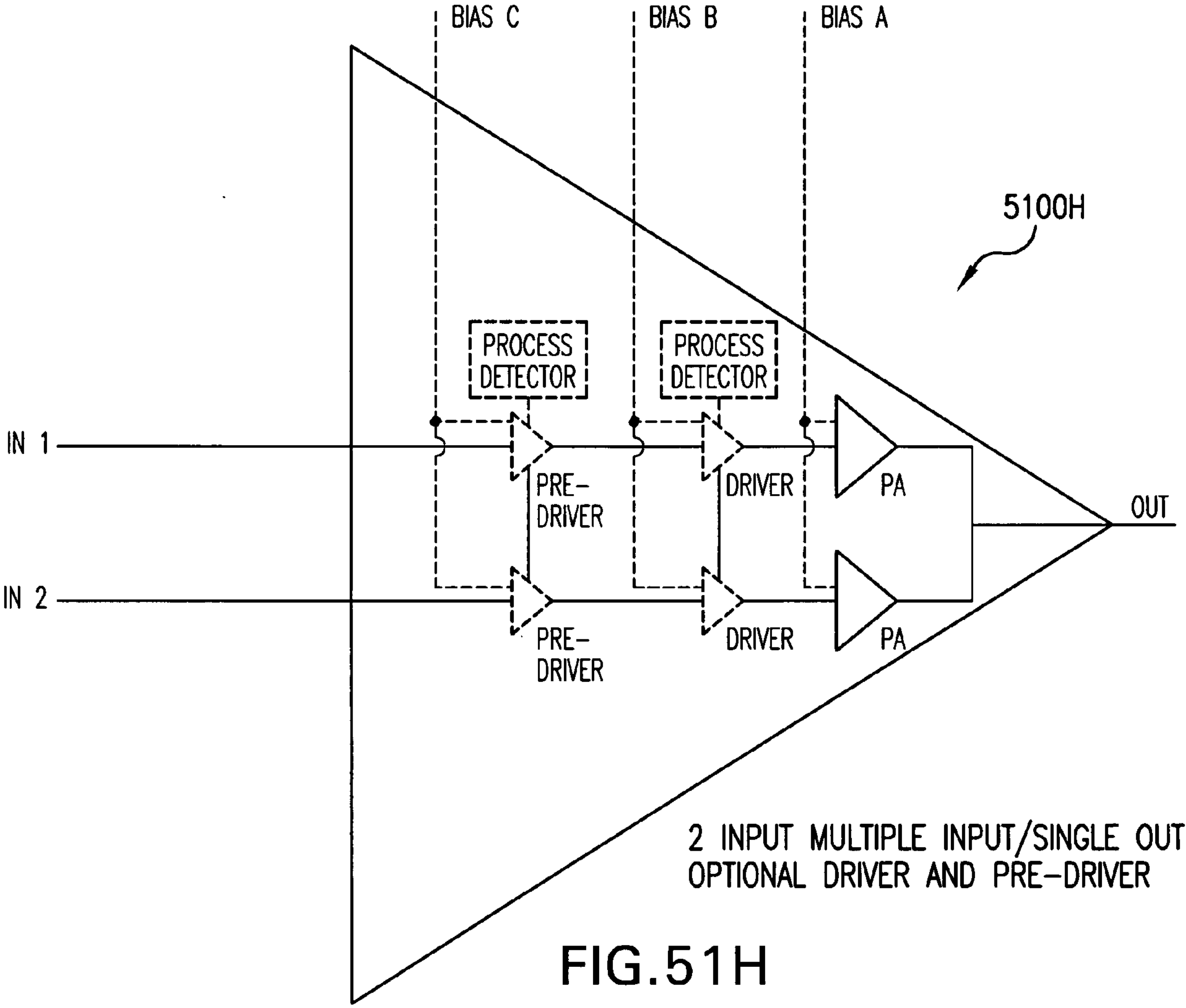
BIAS C
BIAS B
BIAS A
5100H
PROCESS DETECTOR
PROCESS DETECTOR
IN 1
IN 2
PRE-DRIVER
DRIVER
PA
PRE-DRIVER
DRIVER
PA
OUT

2 INPUT MULTIPLE INPUT/SINGLE OUTPUT AMPLIFIER WITH OPTIONAL DRIVER AND PRE-DRIVER

FIG.51H

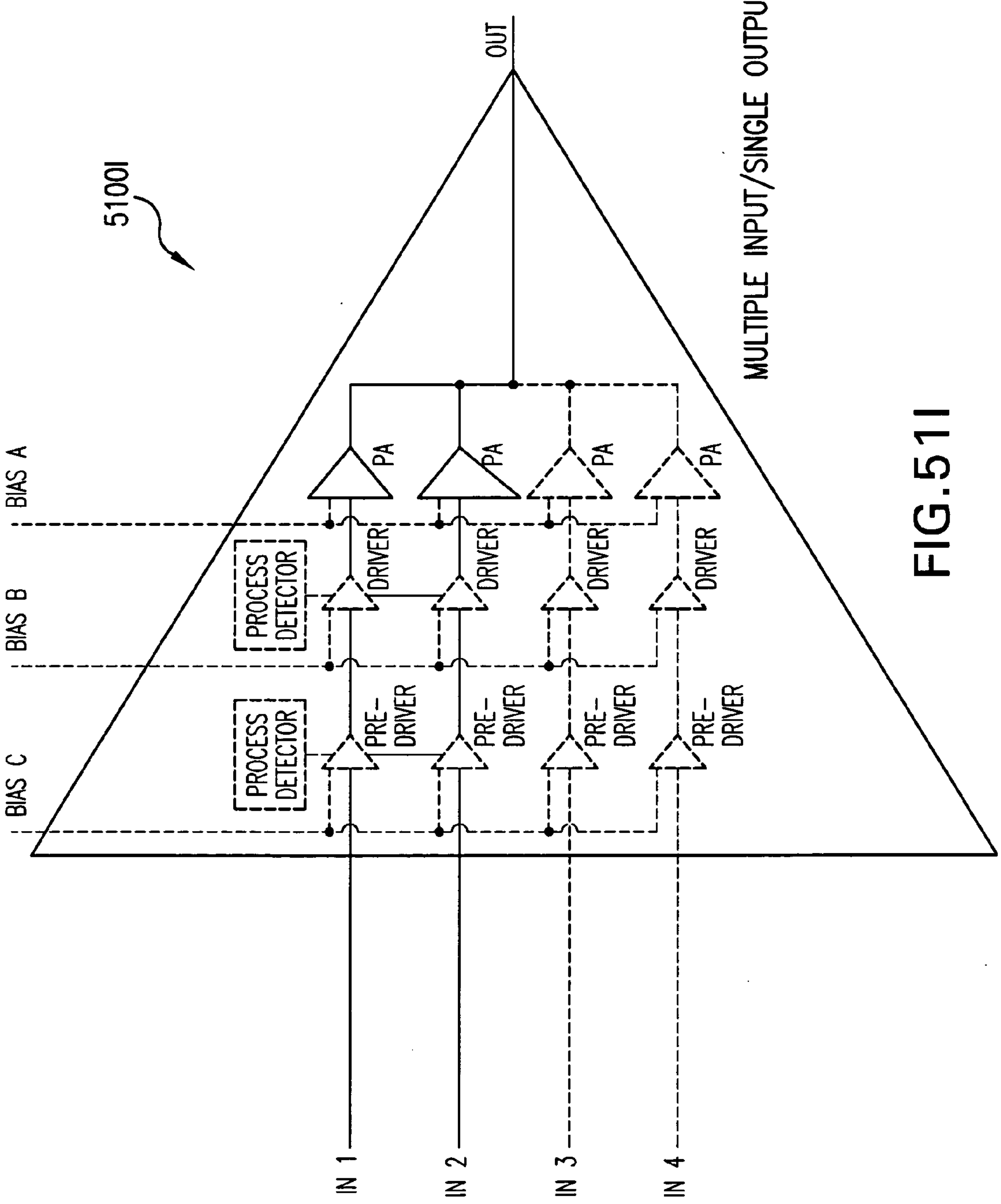
5100I
BIAS A
BIAS B
BIAS C
PROCESS DETECTOR
PROCESS DETECTOR
PA
PA
PA
PA
DRIVER
DRIVER
DRIVER
DRIVER
PRE-DRIVER
PRE-DRIVER
PRE-DRIVER
PRE-DRIVER
OUT
IN 1
IN 2
IN 3
IN 4

MULTIPLE INPUT/SINGLE OUTPUT AMPLIFIER

FIG.51I

FIG. 52

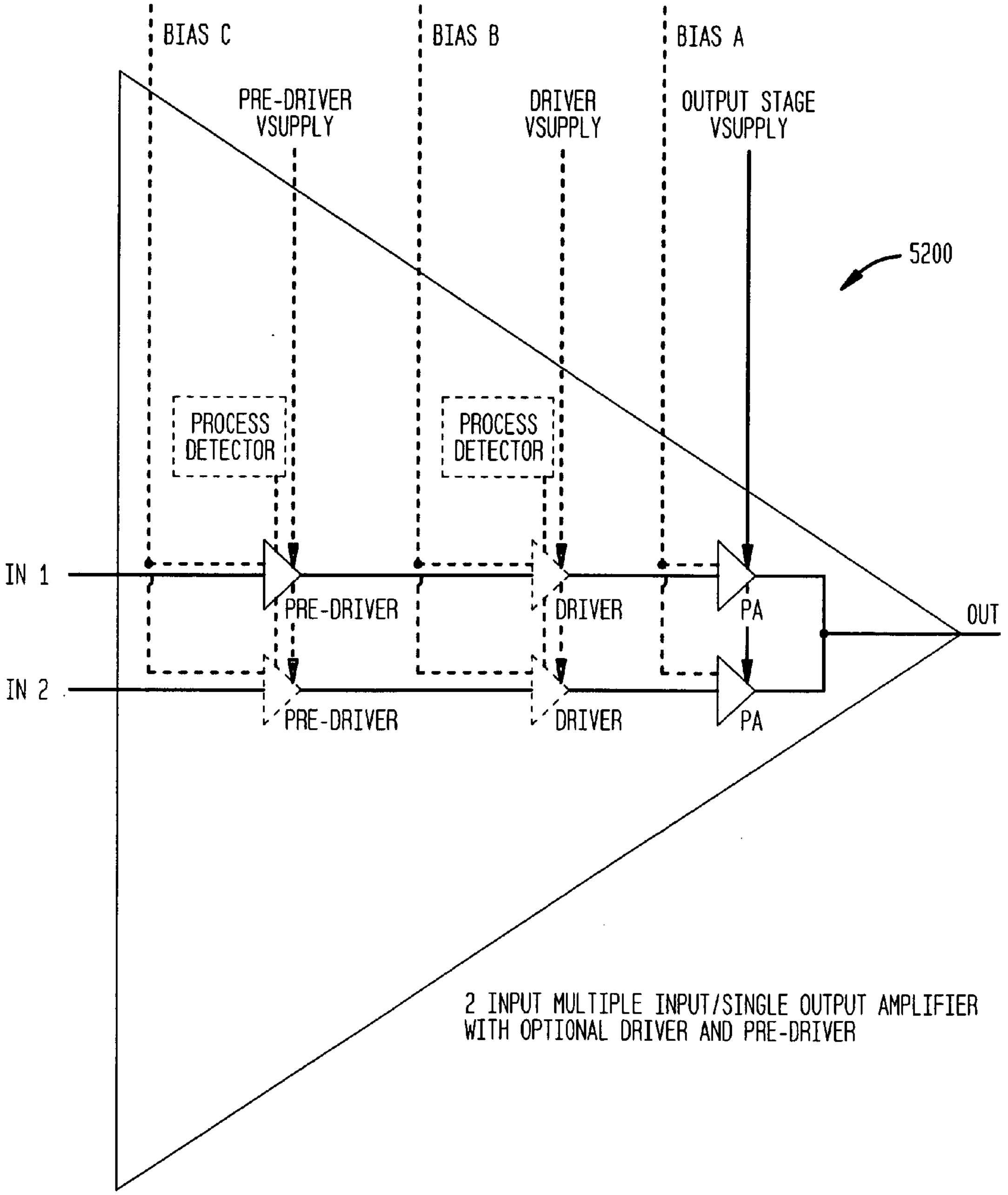
BIAS C
BIAS B
BIAS A
PRE-DRIVER VSUPPLY
DRIVER VSUPPLY
OUTPUT STAGE VSUPPLY
5200
PROCESS DETECTOR
PROCESS DETECTOR
IN 1
IN 2
PRE-DRIVER
DRIVER
PA
PRE-DRIVER
DRIVER
PA
OUT
2 INPUT MULTIPLE INPUT/SINGLE OUTPUT AMPLIFIER WITH OPTIONAL DRIVER AND PRE-DRIVER

FIG. 53

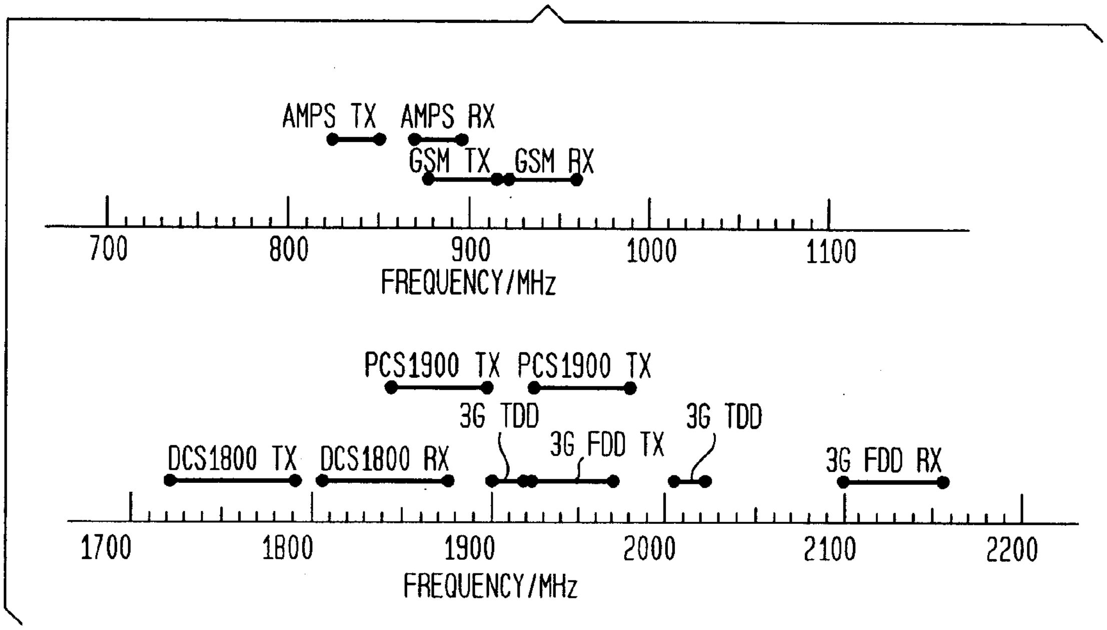
AMPS TX
AMPS RX
GSM TX
GSM RX
700
800
900
1000
1100
FREQUENCY/MHz
PCS1900 TX
PCS1900 TX
3G TDD
3G FDD TX
3G TDD
DCS1800 TX
DCS1800 RX
3G FDD RX
1700
1800
1900
2000
2100
2200
FREQUENCY/MHz

FIG. 54A

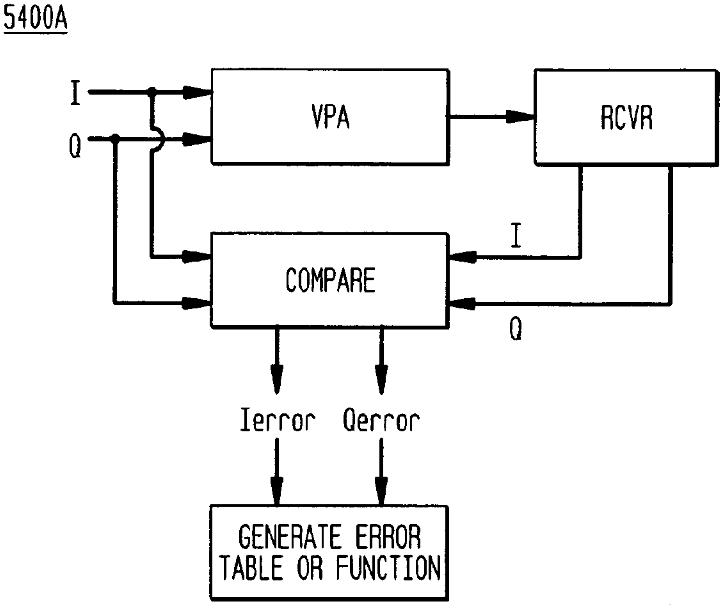
5400A
I
Q
VPA
RCVR
I
COMPARE
Q
Ierror
Qerror
GENERATE ERROR TABLE OR FUNCTION

FIG. 54B

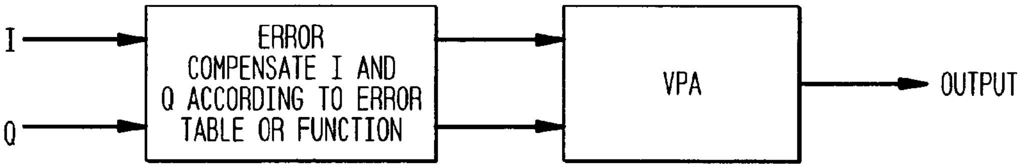
I
Q
ERROR COMPENSATE I AND Q ACCORDING TO ERROR TABLE OR FUNCTION
VPA
OUTPUT

FIG. 55

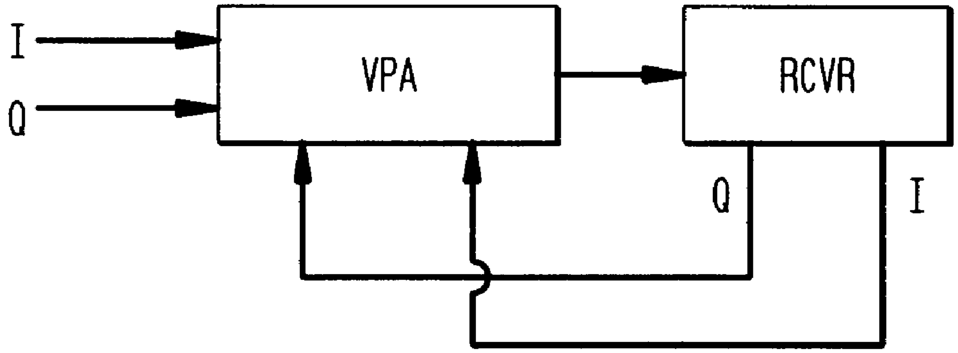
I
Q
VPA
RCVR
Q
I

FIG. 56

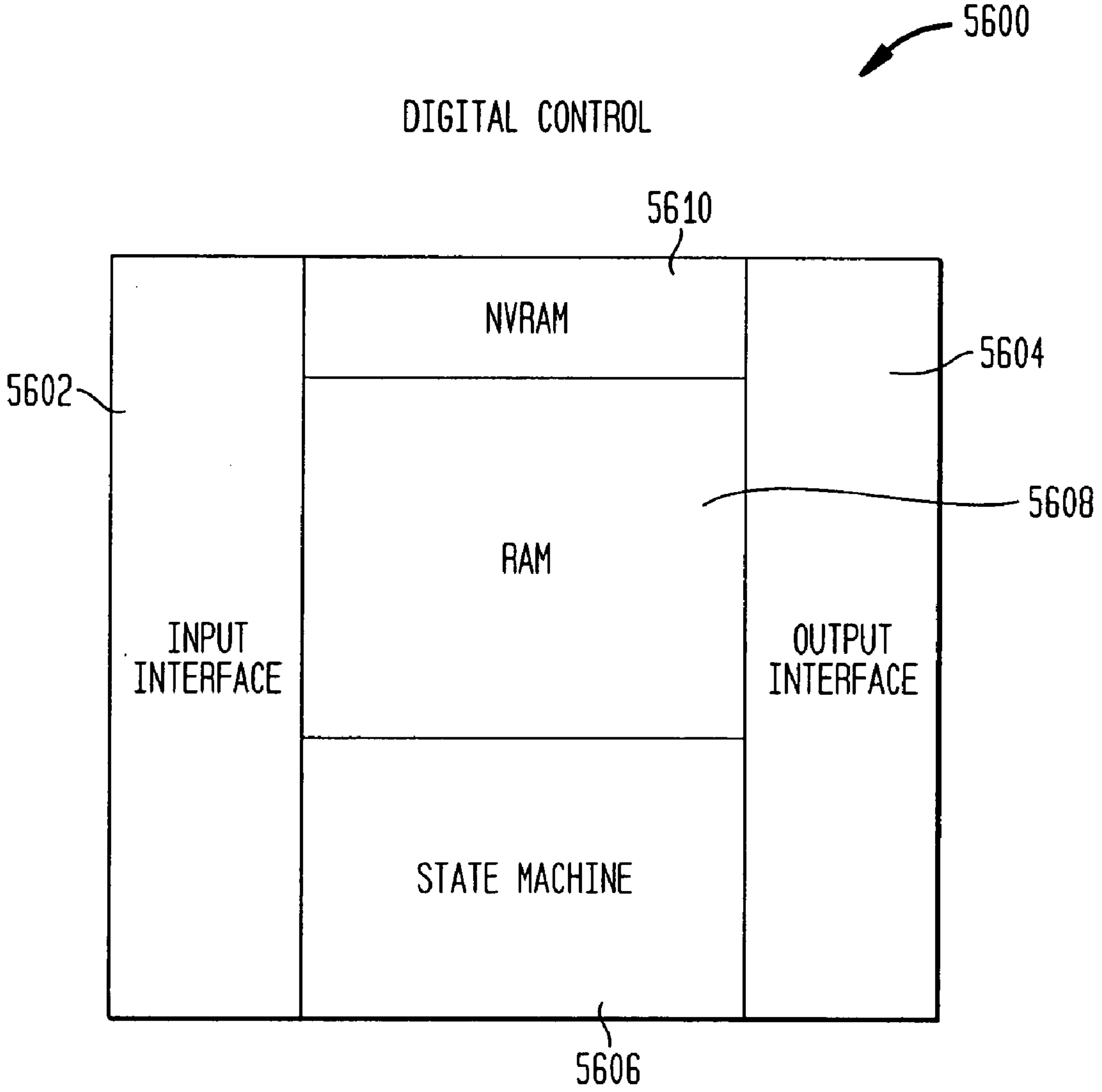
5600
DIGITAL CONTROL
5610
NVRAM
5602
5604
5608
RAM
INPUT INTERFACE
OUTPUT INTERFACE
STATE MACHINE
5606

FIG. 57

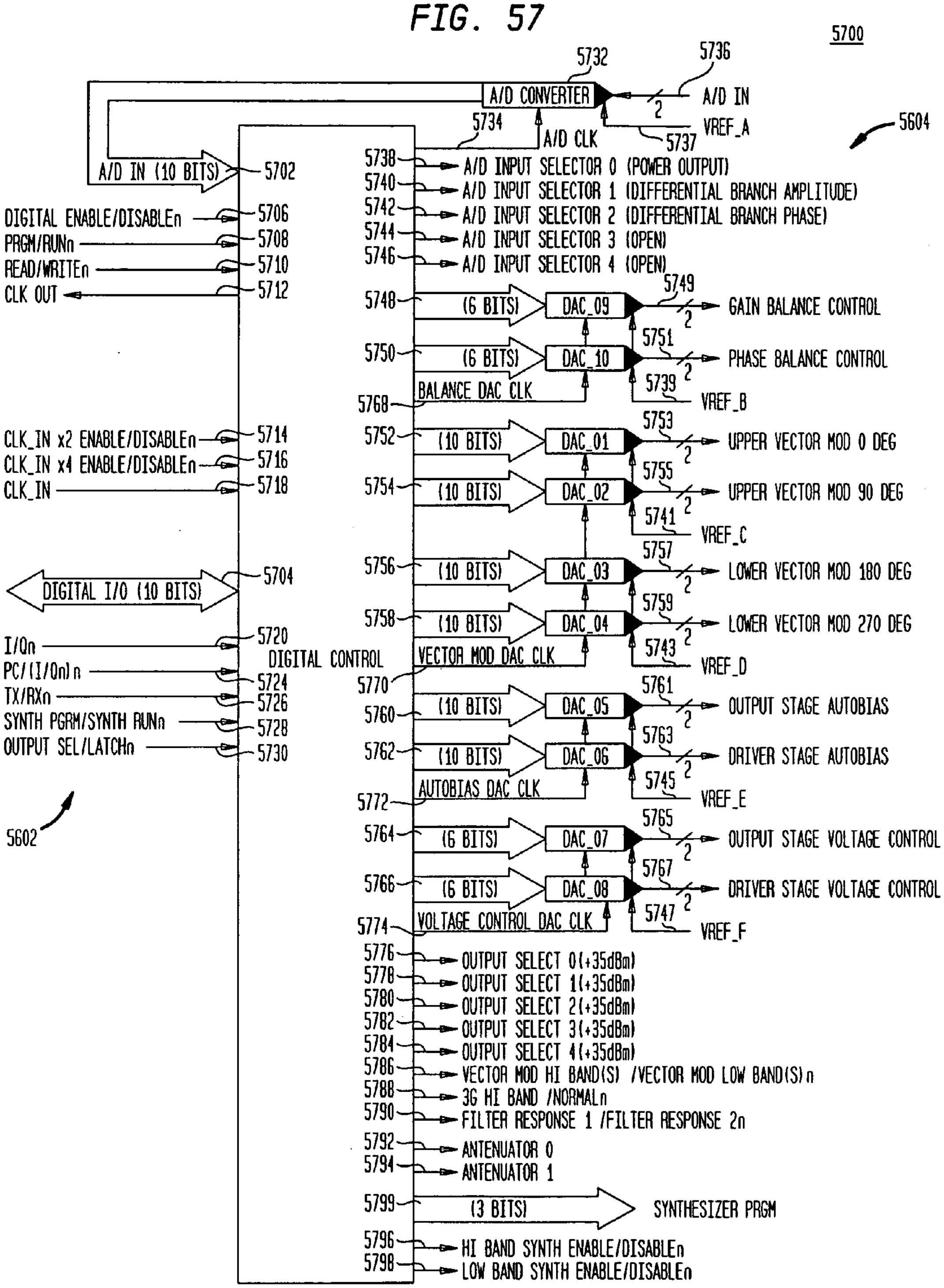
5700
5732
5736
A/D CONVERTER
2
A/D IN
VREF_A
5737
5734
A/D CLK
5604
A/D IN (10 BITS)
5702
DIGITAL ENABLE/DISABLEn
5706
PRGM/RUNn
5708
READ/WRITEn
5710
CLK OUT
5712
5738
A/D INPUT SELECTOR 0 (POWER OUTPUT)
5740
A/D INPUT SELECTOR 1 (DIFFERENTIAL BRANCH AMPLITUDE)
5742
A/D INPUT SELECTOR 2 (DIFFERENTIAL BRANCH PHASE)
5744
A/D INPUT SELECTOR 3 (OPEN)
5746
A/D INPUT SELECTOR 4 (OPEN)
5749
5748
(6 BITS)
DAC_09
2
GAIN BALANCE CONTROL
5751
5750
(6 BITS)
DAC_10
2
PHASE BALANCE CONTROL
BALANCE DAC CLK
5739
VREF_B
5768
5753
CLK_IN x2 ENABLE/DISABLEn
5714
CLK_IN x4 ENABLE/DISABLEn
5716
CLK_IN
5718
5752
(10 BITS)
DAC_01
2
UPPER VECTOR MOD 0 DEG
5755
5754
(10 BITS)
DAC_02
2
UPPER VECTOR MOD 90 DEG
5741
VREF_C
5757
DIGITAL I/O (10 BITS)
5704
5756
(10 BITS)
DAC_03
2
LOWER VECTOR MOD 180 DEG
5759
5758
(10 BITS)
DAC_04
2
LOWER VECTOR MOD 270 DEG
5743
VREF_D
I/Qn
5720
DIGITAL CONTROL
VECTOR MOD DAC CLK
PC/(I/Qn)n
5724
5770
TX/RXn
5726
SYNTH PGRM/SYNTH RUNn
5728
OUTPUT SEL/LATCHn
5730
5761
5760
(10 BITS)
DAC_05
2
OUTPUT STAGE AUTOBIAS
5763
5762
(10 BITS)
DAC_06
2
DRIVER STAGE AUTOBIAS
AUTOBIAS DAC CLK
5745
VREF_E
5772
5602
5765
5764
(6 BITS)
DAC_07
2
OUTPUT STAGE VOLTAGE CONTROL
5767
5766
(6 BITS)
DAC_08
2
DRIVER STAGE VOLTAGE CONTROL
5774
VOLTAGE CONTROL DAC CLK
5747
VREF_F
5776
OUTPUT SELECT 0(+35dBm)
5778
OUTPUT SELECT 1(+35dBm)
5780
OUTPUT SELECT 2(+35dBm)
5782
OUTPUT SELECT 3(+35dBm)
5784
OUTPUT SELECT 4(+35dBm)
5786
VECTOR MOD HI BAND(S) /VECTOR MOD LOW BAND(S)n
5788
3G HI BAND /NORMALn
5790
FILTER RESPONSE 1 /FILTER RESPONSE 2n
5792
ANTENUATOR 0
5794
ANTENUATOR 1
5799
(3 BITS)
SYNTHESIZER PRGM
5796
HI BAND SYNTH ENABLE/DISABLEn
5798
LOW BAND SYNTH ENABLE/DISABLEn

FIG. 58

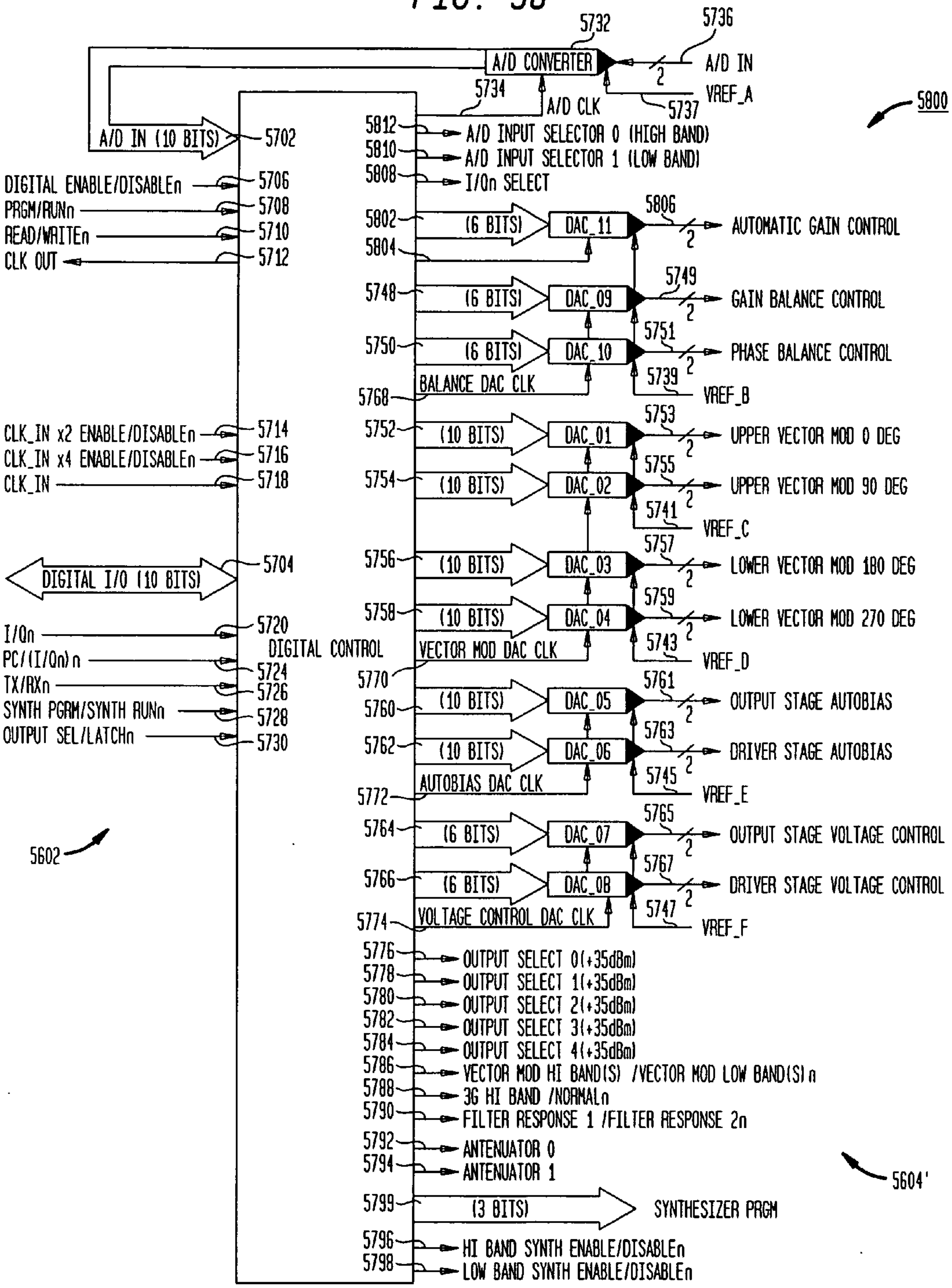
5732
5736
A/D CONVERTER
A/D IN
2
VREF_A
5737
5734
A/D CLK
5800
A/D IN (10 BITS)
5702
5812
A/D INPUT SELECTOR 0 (HIGH BAND)
5810
A/D INPUT SELECTOR 1 (LOW BAND)
5808
I/Qn SELECT
DIGITAL ENABLE/DISABLEn
5706
PRGM/RUNn
5708
READ/WRITEn
5710
CLK OUT
5712
5802
(6 BITS)
DAC_11
5806
2
AUTOMATIC GAIN CONTROL
5804
5748
(6 BITS)
DAC_09
5749
2
GAIN BALANCE CONTROL
5750
(6 BITS)
DAC_10
5751
2
PHASE BALANCE CONTROL
BALANCE DAC CLK
5739
VREF_B
5768
CLK_IN x2 ENABLE/DISABLEn
5714
CLK_IN x4 ENABLE/DISABLEn
5716
CLK_IN
5718
5752
(10 BITS)
DAC_01
5753
2
UPPER VECTOR MOD 0 DEG
5754
(10 BITS)
DAC_02
5755
2
UPPER VECTOR MOD 90 DEG
5741
VREF_C
DIGITAL I/O (10 BITS)
5704
5756
(10 BITS)
DAC_03
5757
2
LOWER VECTOR MOD 180 DEG
5758
(10 BITS)
DAC_04
5759
2
LOWER VECTOR MOD 270 DEG
I/Qn
5720
DIGITAL CONTROL
VECTOR MOD DAC CLK
5743
VREF_D
PC/(I/Qn)n
5724
TX/RXn
5726
SYNTH PGRM/SYNTH RUNn
5728
OUTPUT SEL/LATCHn
5730
5770
5760
(10 BITS)
DAC_05
5761
2
OUTPUT STAGE AUTOBIAS
5762
(10 BITS)
DAC_06
5763
2
DRIVER STAGE AUTOBIAS
AUTOBIAS DAC CLK
5745
VREF_E
5772
5602
5764
(6 BITS)
DAC_07
5765
2
OUTPUT STAGE VOLTAGE CONTROL
5766
(6 BITS)
DAC_08
5767
2
DRIVER STAGE VOLTAGE CONTROL
VOLTAGE CONTROL DAC CLK
5747
VREF_F
5774
5776
OUTPUT SELECT 0(+35dBm)
5778
OUTPUT SELECT 1(+35dBm)
5780
OUTPUT SELECT 2(+35dBm)
5782
OUTPUT SELECT 3(+35dBm)
5784
OUTPUT SELECT 4(+35dBm)
5786
VECTOR MOD HI BAND(S) /VECTOR MOD LOW BAND(S)n
5788
3G HI BAND /NORMALn
5790
FILTER RESPONSE 1 /FILTER RESPONSE 2n
5792
ANTENUATOR 0
5794
ANTENUATOR 1
5604'
5799
(3 BITS)
SYNTHESIZER PRGM
5796
HI BAND SYNTH ENABLE/DISABLEn
5798
LOW BAND SYNTH ENABLE/DISABLEn

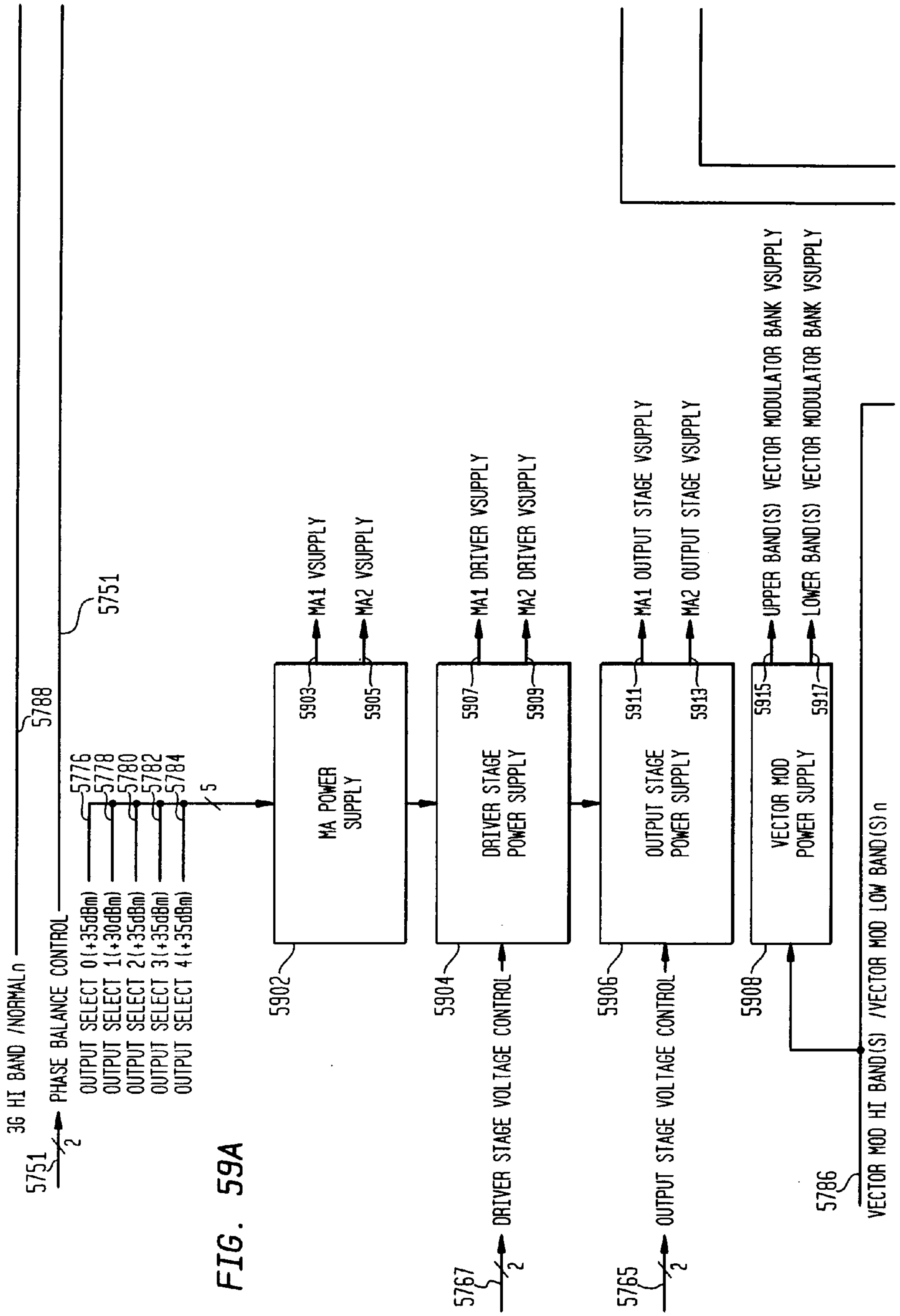
5751
2
3G HI BAND /NORMALn
5788
PHASE BALANCE CONTROL
5751
OUTPUT SELECT 0(+35dBm)
OUTPUT SELECT 1(+30dBm)
OUTPUT SELECT 2(+35dBm)
OUTPUT SELECT 3(+35dBm)
OUTPUT SELECT 4(+35dBm)
5776
5778
5780
5782
5784
5
MA POWER SUPPLY
5902
5903
5905
MA1 VSUPPLY
MA2 VSUPPLY
5767
2
DRIVER STAGE VOLTAGE CONTROL
DRIVER STAGE POWER SUPPLY
5904
5907
5909
MA1 DRIVER VSUPPLY
MA2 DRIVER VSUPPLY
5765
2
OUTPUT STAGE VOLTAGE CONTROL
OUTPUT STAGE POWER SUPPLY
5906
5911
5913
MA1 OUTPUT STAGE VSUPPLY
MA2 OUTPUT STAGE VSUPPLY
5786
VECTOR MOD HI BAND(S) /VECTOR MOD LOW BAND(S)n
VECTOR MOD POWER SUPPLY
5908
5915
5917
UPPER BAND(S) VECTOR MODULATOR BANK VSUPPLY
LOWER BAND(S) VECTOR MODULATOR BANK VSUPPLY

FIG. 59A

FIG. 59B

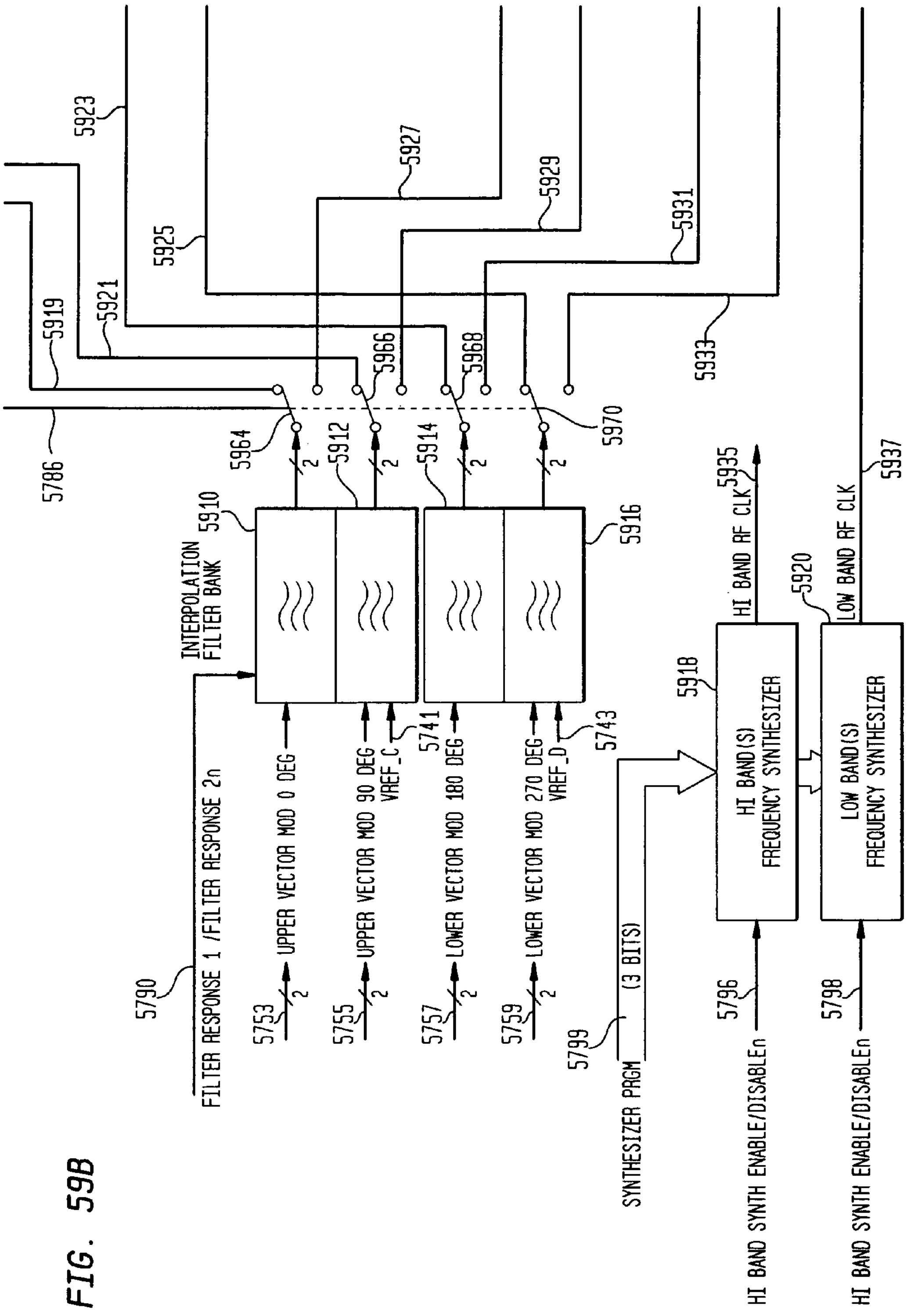
INTERPOLATION FILTER BANK
FILTER RESPONSE 1 /FILTER RESPONSE 2n
UPPER VECTOR MOD 0 DEG
UPPER VECTOR MOD 90 DEG
VREF_C
LOWER VECTOR MOD 180 DEG
LOWER VECTOR MOD 270 DEG
VREF_D
SYNTHESIZER PRGM
(3 BITS)
HI BAND(S) FREQUENCY SYNTHESIZER
HI BAND RF CLK
LOW BAND(S) FREQUENCY SYNTHESIZER
LOW BAND RF CLK
HI BAND SYNTH ENABLE/DISABLEn
HI BAND SYNTH ENABLE/DISABLEn

FIG. 59C

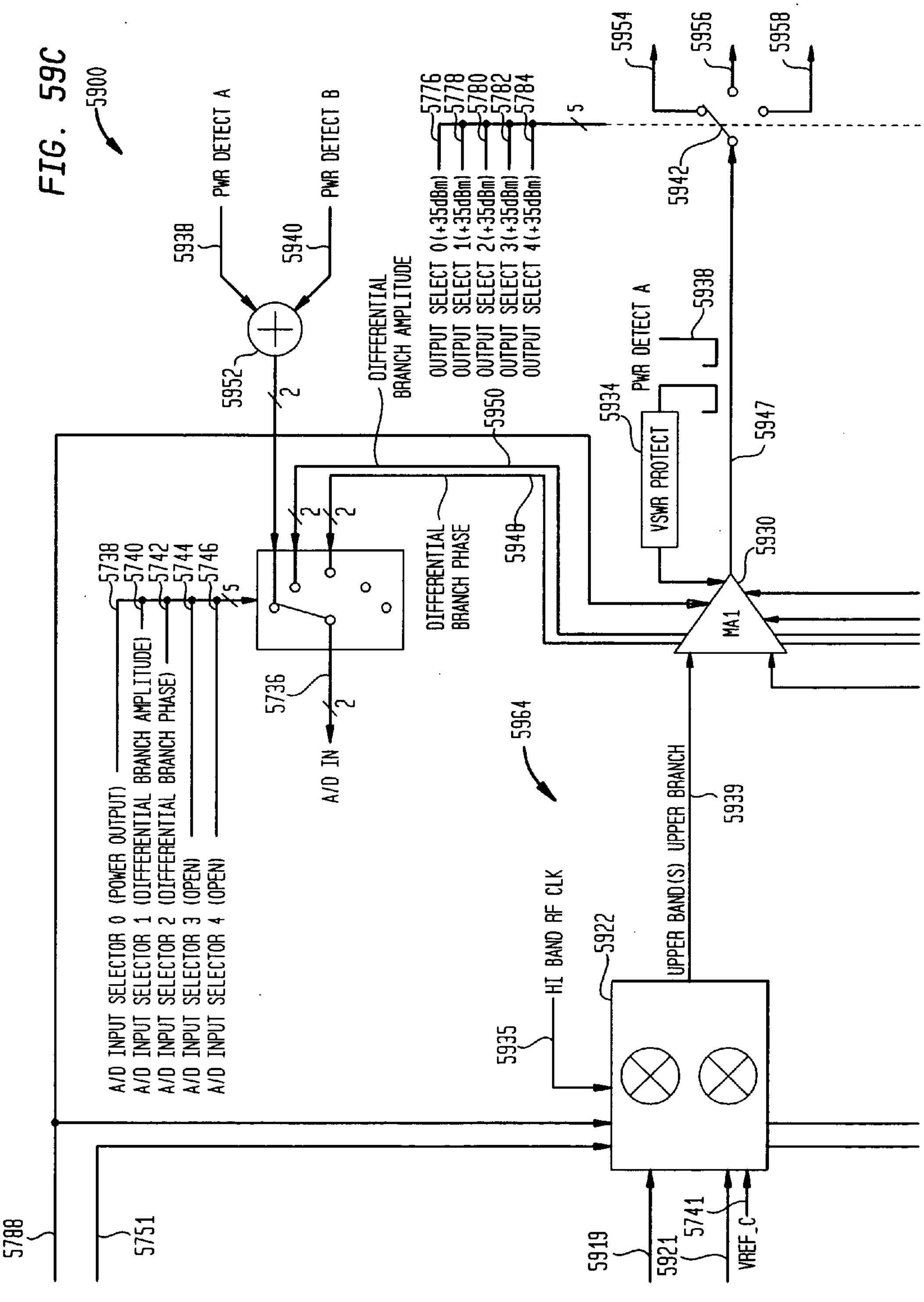
PWR DETECT A
PWR DETECT B
DIFFERENTIAL BRANCH AMPLITUDE
OUTPUT SELECT 0(+35dBm)
OUTPUT SELECT 1(+35dBm)
OUTPUT SELECT 2(+35dBm)
OUTPUT SELECT 3(+35dBm)
OUTPUT SELECT 4(+35dBm)
PWR DETECT A
VSWR PROTECT
MA1
DIFFERENTIAL BRANCH PHASE
A/D INPUT SELECTOR 0 (POWER OUTPUT)
A/D INPUT SELECTOR 1 (DIFFERENTIAL BRANCH AMPLITUDE)
A/D INPUT SELECTOR 2 (DIFFERENTIAL BRANCH PHASE)
A/D INPUT SELECTOR 3 (OPEN)
A/D INPUT SELECTOR 4 (OPEN)
A/D IN
UPPER BAND(S) UPPER BRANCH
HI BAND RF CLK
VREF_C

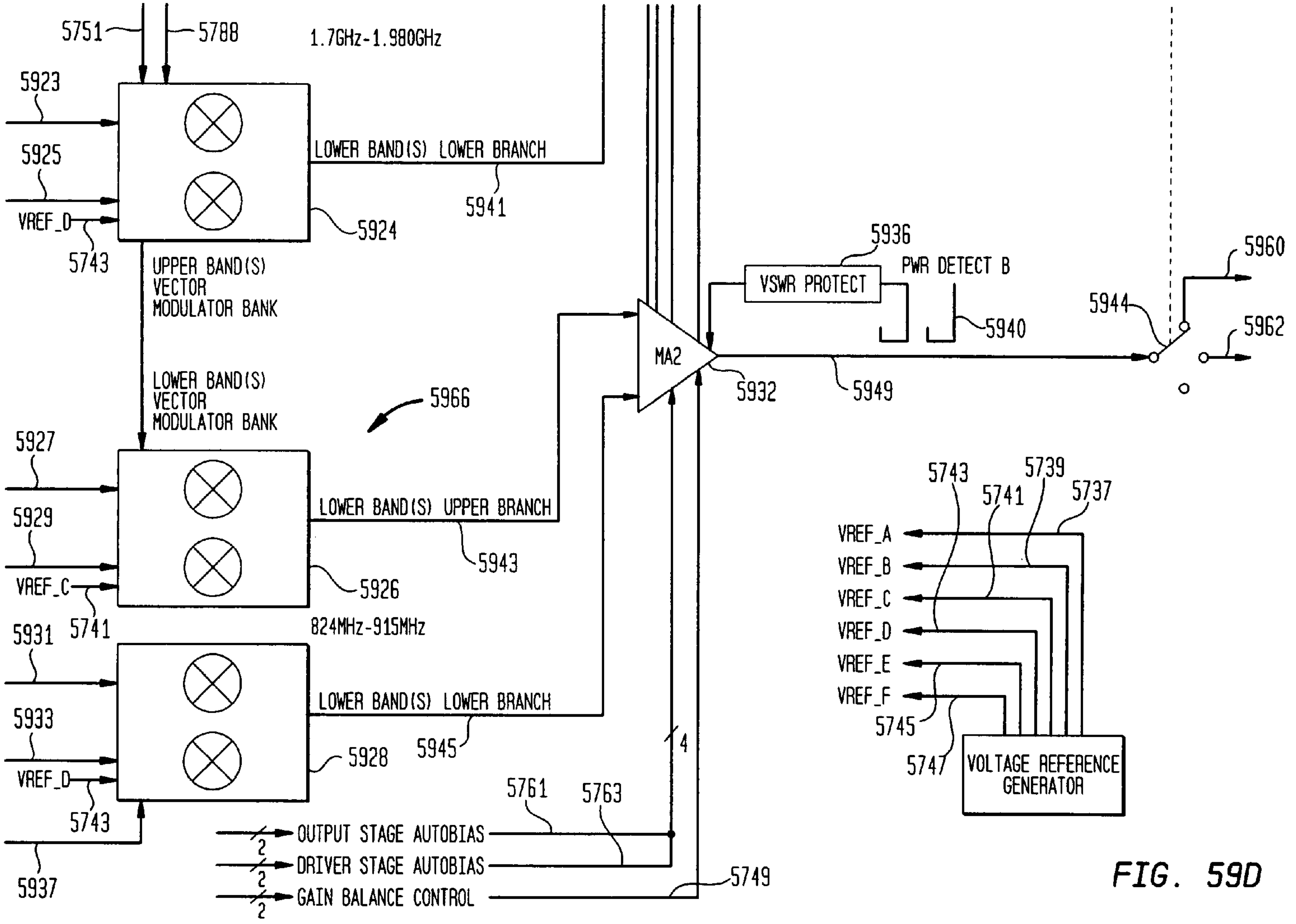
1.7GHz-1.980GHz
VREF_D
LOWER BAND(S) LOWER BRANCH
UPPER BAND(S) VECTOR MODULATOR BANK
LOWER BAND(S) VECTOR MODULATOR BANK
VREF_C
LOWER BAND(S) UPPER BRANCH
824MHz-915MHz
VREF_D
LOWER BAND(S) LOWER BRANCH
OUTPUT STAGE AUTOBIAS
DRIVER STAGE AUTOBIAS
GAIN BALANCE CONTROL
MA2
VSWR PROTECT
PWR DETECT B
VREF_A
VREF_B
VREF_C
VREF_D
VREF_E
VREF_F
VOLTAGE REFERENCE GENERATOR

FIG. 59D

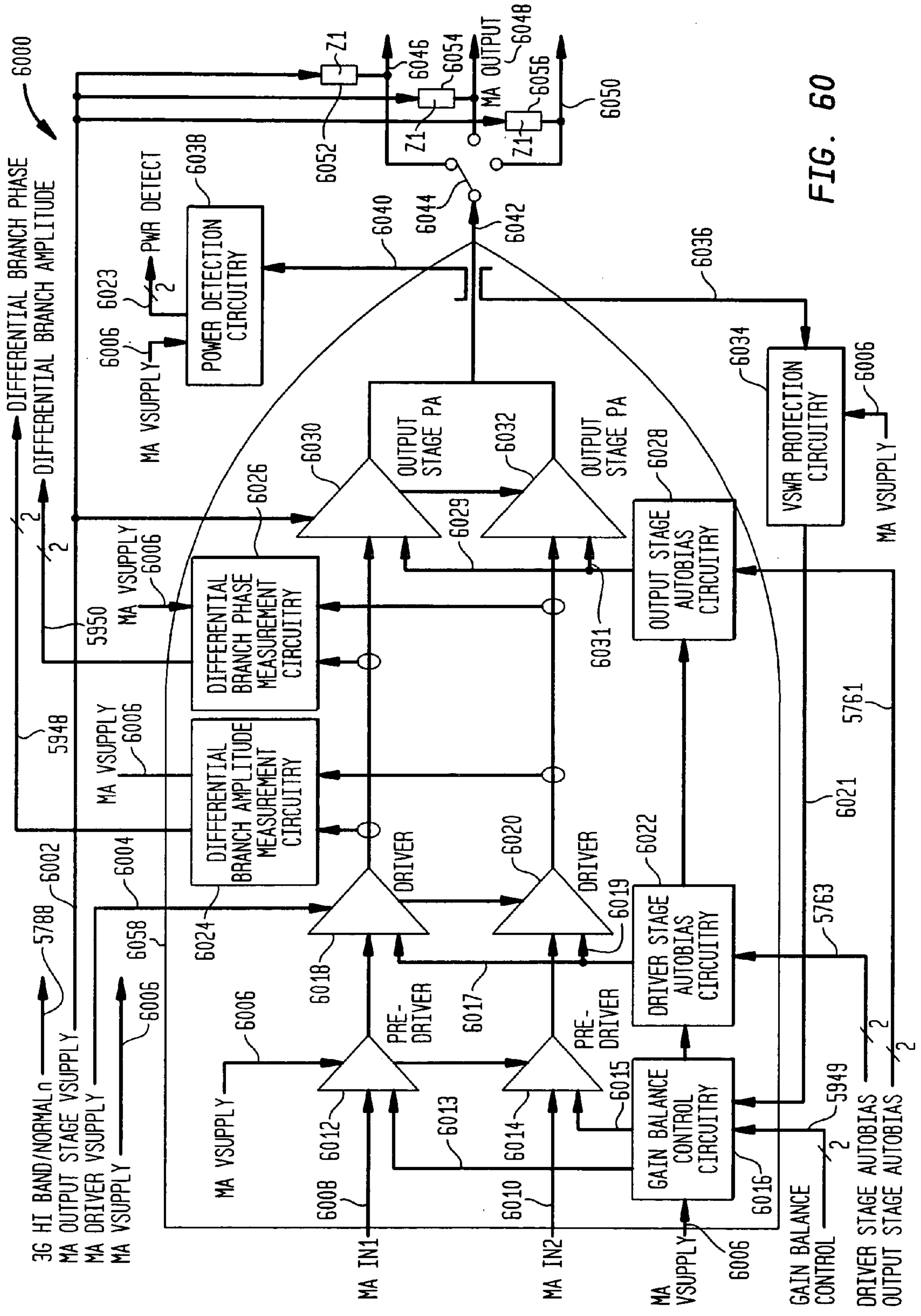
6000
DIFFERENTIAL BRANCH PHASE
DIFFERENTIAL BRANCH AMPLITUDE
3G HI BAND/NORMALn
MA OUTPUT STAGE VSUPPLY
MA DRIVER VSUPPLY
MA VSUPPLY
5788
6002
6006
6004
6058
6024
MA IN1
MA IN2
6008
6010
MA VSUPPLY
6006
6012
6013
6014
PRE-DRIVER
PRE-DRIVER
6015
GAIN BALANCE CONTROL CIRCUITRY
6016
MA VSUPPLY
6006
5949
GAIN BALANCE CONTROL
DRIVER STAGE AUTOBIAS
OUTPUT STAGE AUTOBIAS
2
6018
DRIVER
DRIVER
6017
6019
6020
DRIVER STAGE AUTOBIAS CIRCUITRY
6022
5763
DIFFERENTIAL BRANCH AMPLITUDE MEASUREMENT CIRCUITRY
MA VSUPPLY
6006
5948
DIFFERENTIAL BRANCH PHASE MEASUREMENT CIRCUITRY
MA VSUPPLY
6006
5950
6026
6030
6032
OUTPUT STAGE PA
OUTPUT STAGE PA
6029
6031
6028
OUTPUT STAGE AUTOBIAS CIRCUITRY
6021
5761
VSWR PROTECTION CIRCUITRY
6034
6006
MA VSUPPLY
6036
POWER DETECTION CIRCUITRY
6038
PWR DETECT
6023
MA VSUPPLY
6006
6040
6042
6044
6052
Z1
Z1
Z1
6046
6054
MA OUTPUT
6048
6056
6050

FIG. 60

FIG. 61A

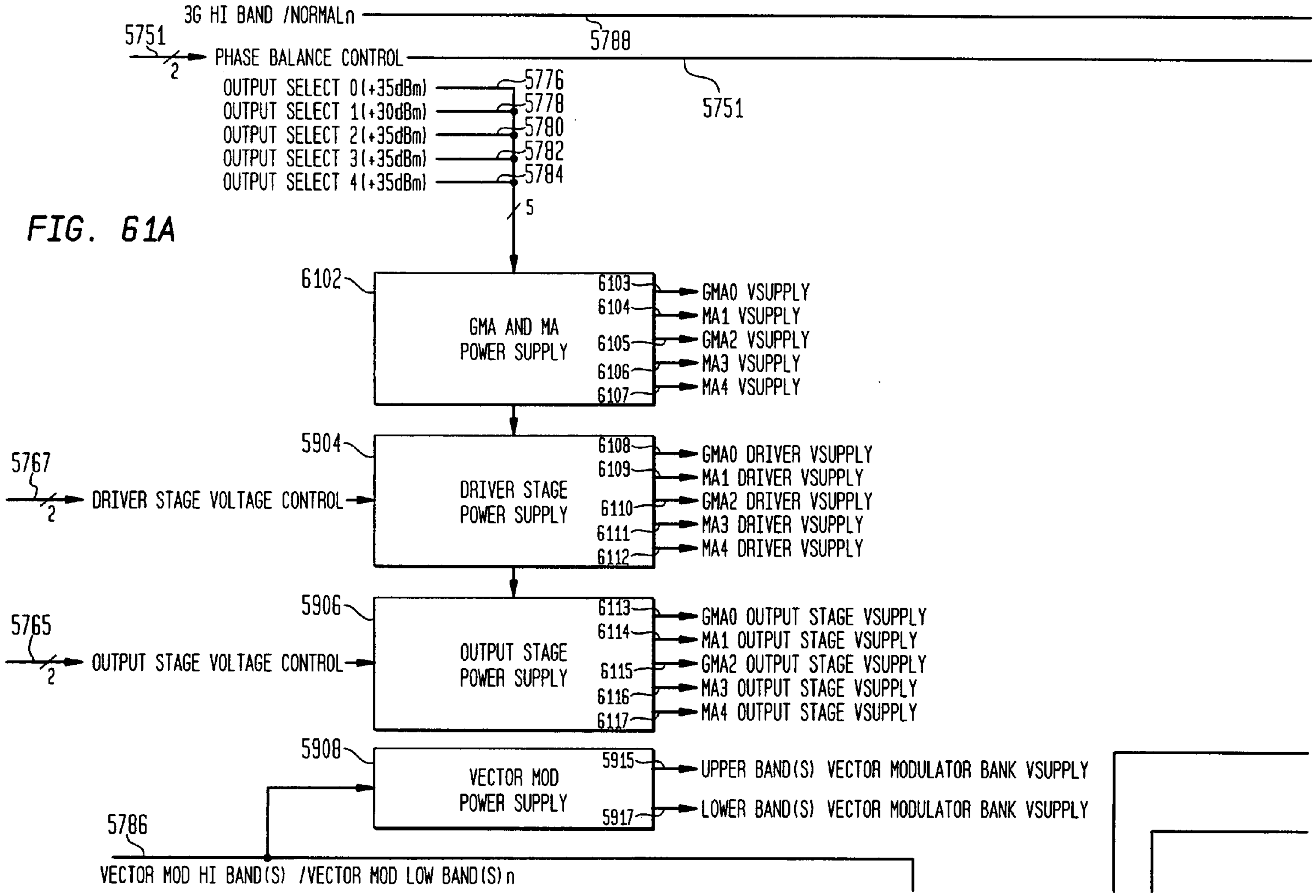
3G HI BAND /NORMALn
5788
5751
2
PHASE BALANCE CONTROL
5751
OUTPUT SELECT 0(+35dBm)
OUTPUT SELECT 1(+30dBm)
OUTPUT SELECT 2(+35dBm)
OUTPUT SELECT 3(+35dBm)
OUTPUT SELECT 4(+35dBm)
5776
5778
5780
5782
5784
5
6102
GMA AND MA
POWER SUPPLY
6103
6104
6105
6106
6107
GMA0 VSUPPLY
MA1 VSUPPLY
GMA2 VSUPPLY
MA3 VSUPPLY
MA4 VSUPPLY
5904
5767
2
DRIVER STAGE VOLTAGE CONTROL
DRIVER STAGE
POWER SUPPLY
6108
6109
6110
6111
6112
GMA0 DRIVER VSUPPLY
MA1 DRIVER VSUPPLY
GMA2 DRIVER VSUPPLY
MA3 DRIVER VSUPPLY
MA4 DRIVER VSUPPLY
5906
5765
2
OUTPUT STAGE VOLTAGE CONTROL
OUTPUT STAGE
POWER SUPPLY
6113
6114
6115
6116
6117
GMA0 OUTPUT STAGE VSUPPLY
MA1 OUTPUT STAGE VSUPPLY
GMA2 OUTPUT STAGE VSUPPLY
MA3 OUTPUT STAGE VSUPPLY
MA4 OUTPUT STAGE VSUPPLY
5908
VECTOR MOD
POWER SUPPLY
5915
5917
UPPER BAND(S) VECTOR MODULATOR BANK VSUPPLY
LOWER BAND(S) VECTOR MODULATOR BANK VSUPPLY
5786
VECTOR MOD HI BAND(S) /VECTOR MOD LOW BAND(S)n

FIG. 61B

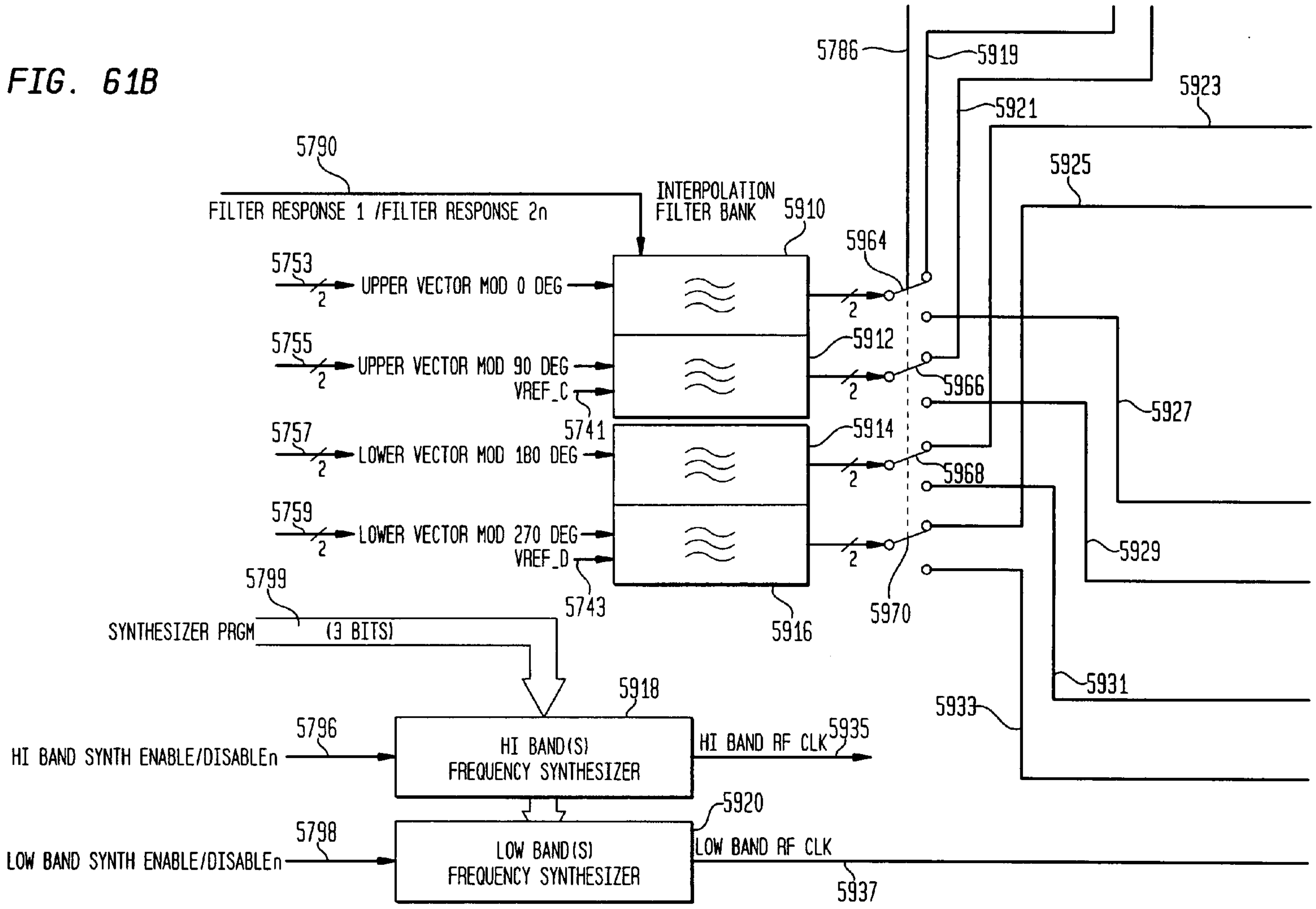
FILTER RESPONSE 1 /FILTER RESPONSE 2n
INTERPOLATION
FILTER BANK
UPPER VECTOR MOD 0 DEG
UPPER VECTOR MOD 90 DEG
VREF_C
LOWER VECTOR MOD 180 DEG
LOWER VECTOR MOD 270 DEG
VREF_D
SYNTHESIZER PRGM
(3 BITS)
HI BAND SYNTH ENABLE/DISABLEn
HI BAND(S)
FREQUENCY SYNTHESIZER
HI BAND RF CLK
LOW BAND SYNTH ENABLE/DISABLEn
LOW BAND(S)
FREQUENCY SYNTHESIZER
LOW BAND RF CLK

FIG. 61C

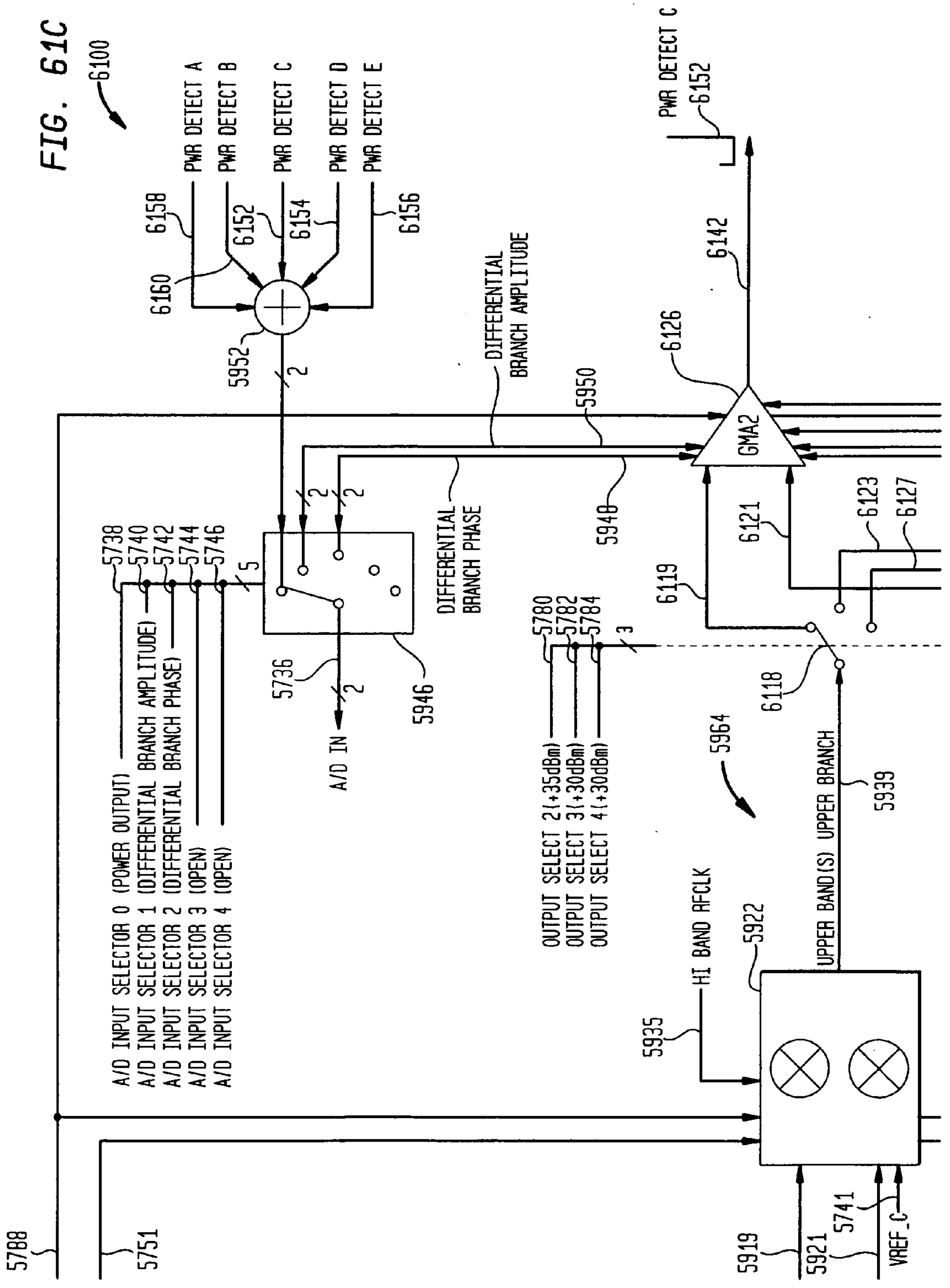
PWR DETECT A
PWR DETECT B
PWR DETECT C
PWR DETECT D
PWR DETECT E
DIFFERENTIAL BRANCH AMPLITUDE
DIFFERENTIAL BRANCH PHASE
GMA2
PWR DETECT C
A/D INPUT SELECTOR 0 (POWER OUTPUT)
A/D INPUT SELECTOR 1 (DIFFERENTIAL BRANCH AMPLITUDE)
A/D INPUT SELECTOR 2 (DIFFERENTIAL BRANCH PHASE)
A/D INPUT SELECTOR 3 (OPEN)
A/D INPUT SELECTOR 4 (OPEN)
A/D IN
OUTPUT SELECT 2(+35dBm)
OUTPUT SELECT 3(+30dBm)
OUTPUT SELECT 4(+30dBm)
UPPER BAND(S) UPPER BRANCH
HI BAND RFCLK
VREF_C

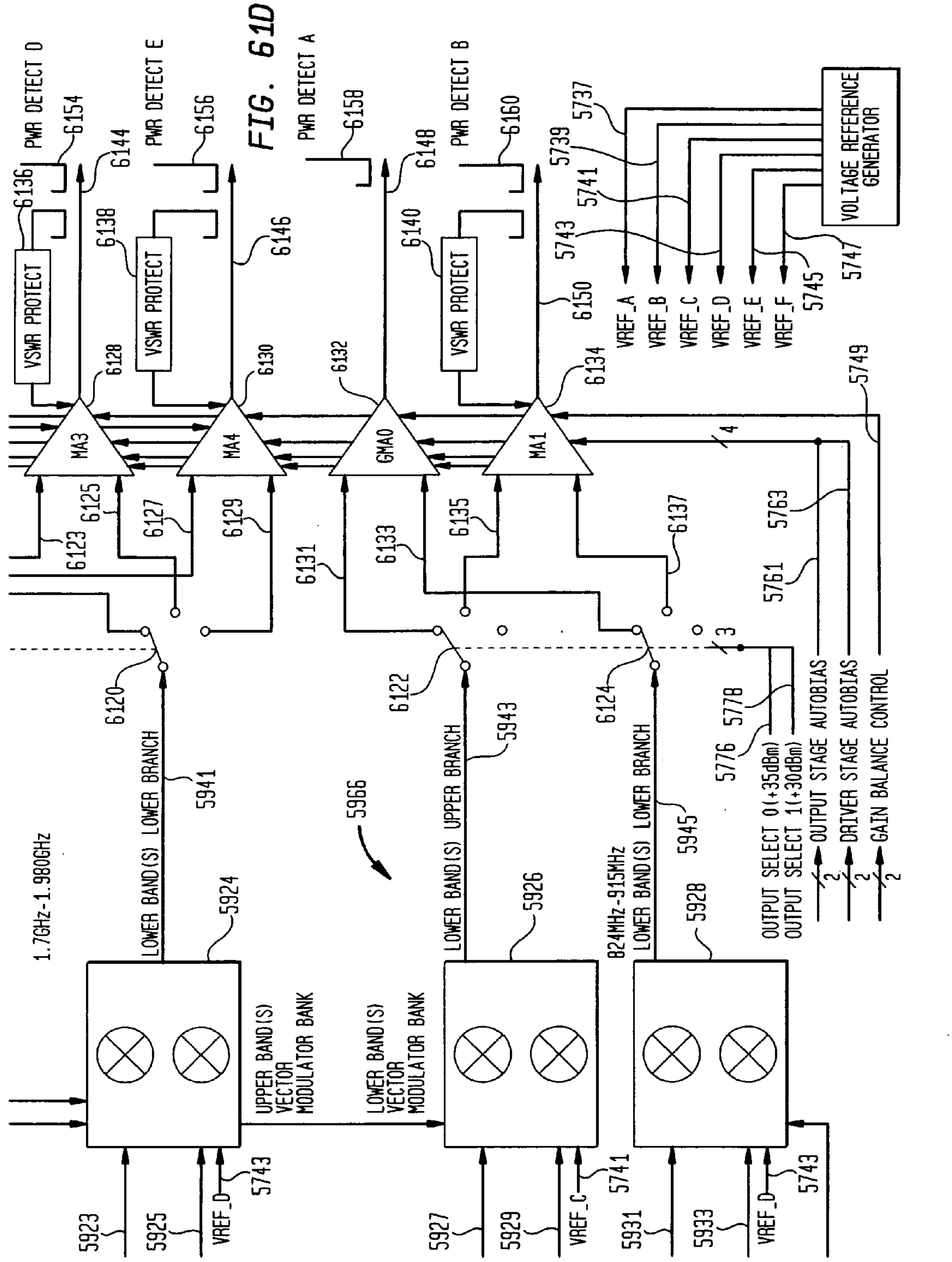
PWR DETECT D
PWR DETECT E
PWR DETECT A
PWR DETECT B
VSWR PROTECT
VSWR PROTECT
VSWR PROTECT
MA3
MA4
GMA0
MA1
VOLTAGE REFERENCE GENERATOR
VREF_A
VREF_B
VREF_C
VREF_D
VREF_E
VREF_F
OUTPUT SELECT 0(+35dBm)
OUTPUT SELECT 1(+30dBm)
OUTPUT STAGE AUTOBIAS
DRIVER STAGE AUTOBIAS
GAIN BALANCE CONTROL
1.7GHz-1.980GHz
LOWER BAND(S) LOWER BRANCH
LOWER BAND(S) UPPER BRANCH
824MHz-915MHz
LOWER BAND(S) LOWER BRANCH
UPPER BAND(S) VECTOR MODULATOR BANK
LOWER BAND(S) VECTOR MODULATOR BANK
VREF_D
VREF_C
VREF_D

FIG. 61D

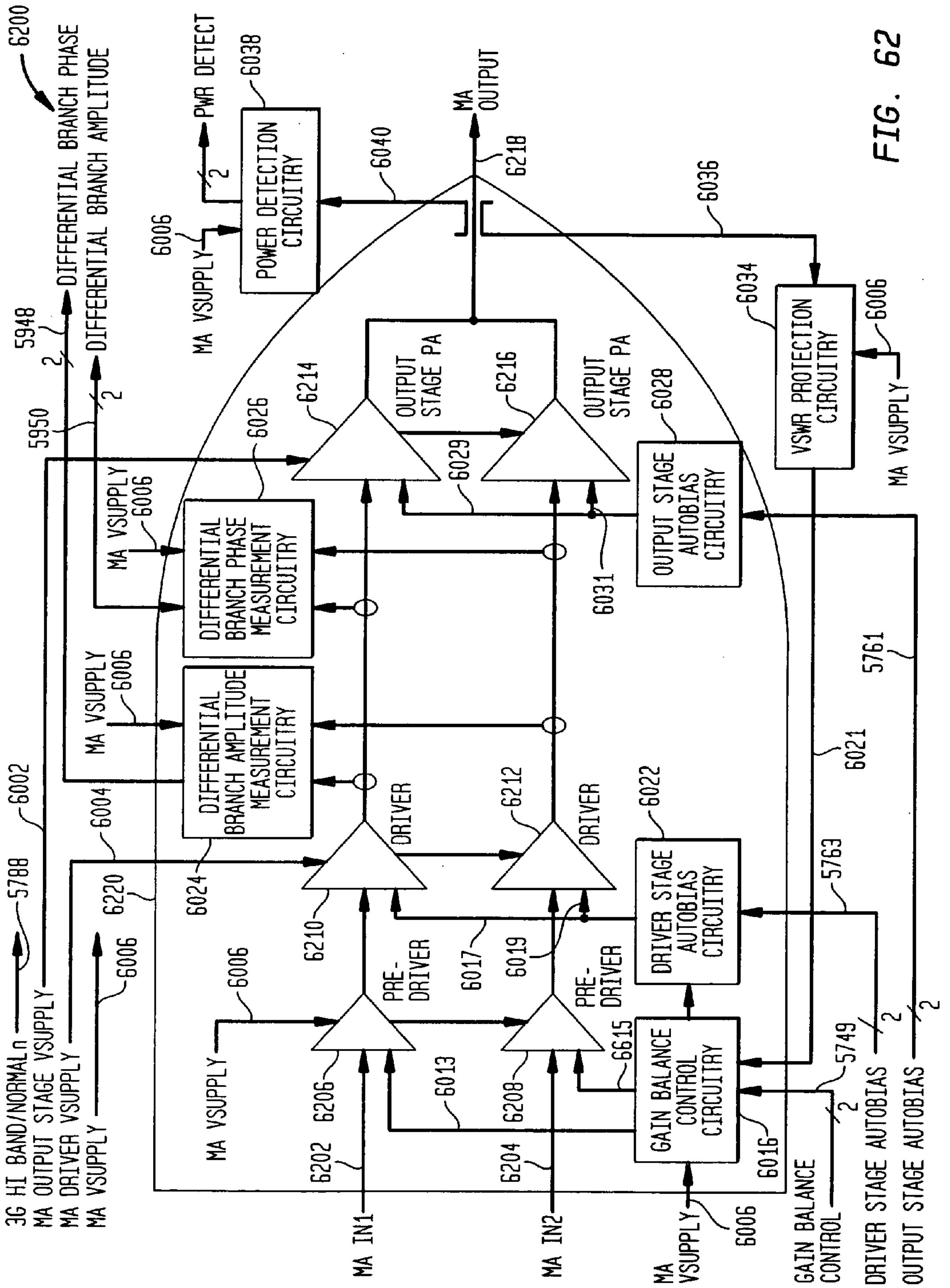
3G HI BAND/NORMALn
MA OUTPUT STAGE VSUPPLY
MA DRIVER VSUPPLY
MA VSUPPLY
DIFFERENTIAL BRANCH PHASE
DIFFERENTIAL BRANCH AMPLITUDE
PWR DETECT
MA VSUPPLY
POWER DETECTION CIRCUITRY
MA OUTPUT
VSWR PROTECTION CIRCUITRY
MA VSUPPLY
OUTPUT STAGE PA
OUTPUT STAGE PA
OUTPUT STAGE AUTOBIAS CIRCUITRY
MA VSUPPLY
DIFFERENTIAL BRANCH PHASE MEASUREMENT CIRCUITRY
MA VSUPPLY
DIFFERENTIAL BRANCH AMPLITUDE MEASUREMENT CIRCUITRY
DRIVER
DRIVER
DRIVER STAGE AUTOBIAS CIRCUITRY
MA VSUPPLY
PRE-DRIVER
PRE-DRIVER
GAIN BALANCE CONTROL CIRCUITRY
MA IN1
MA IN2
MA VSUPPLY
GAIN BALANCE CONTROL
DRIVER STAGE AUTOBIAS
OUTPUT STAGE AUTOBIAS

FIG. 62

FIG. 63A

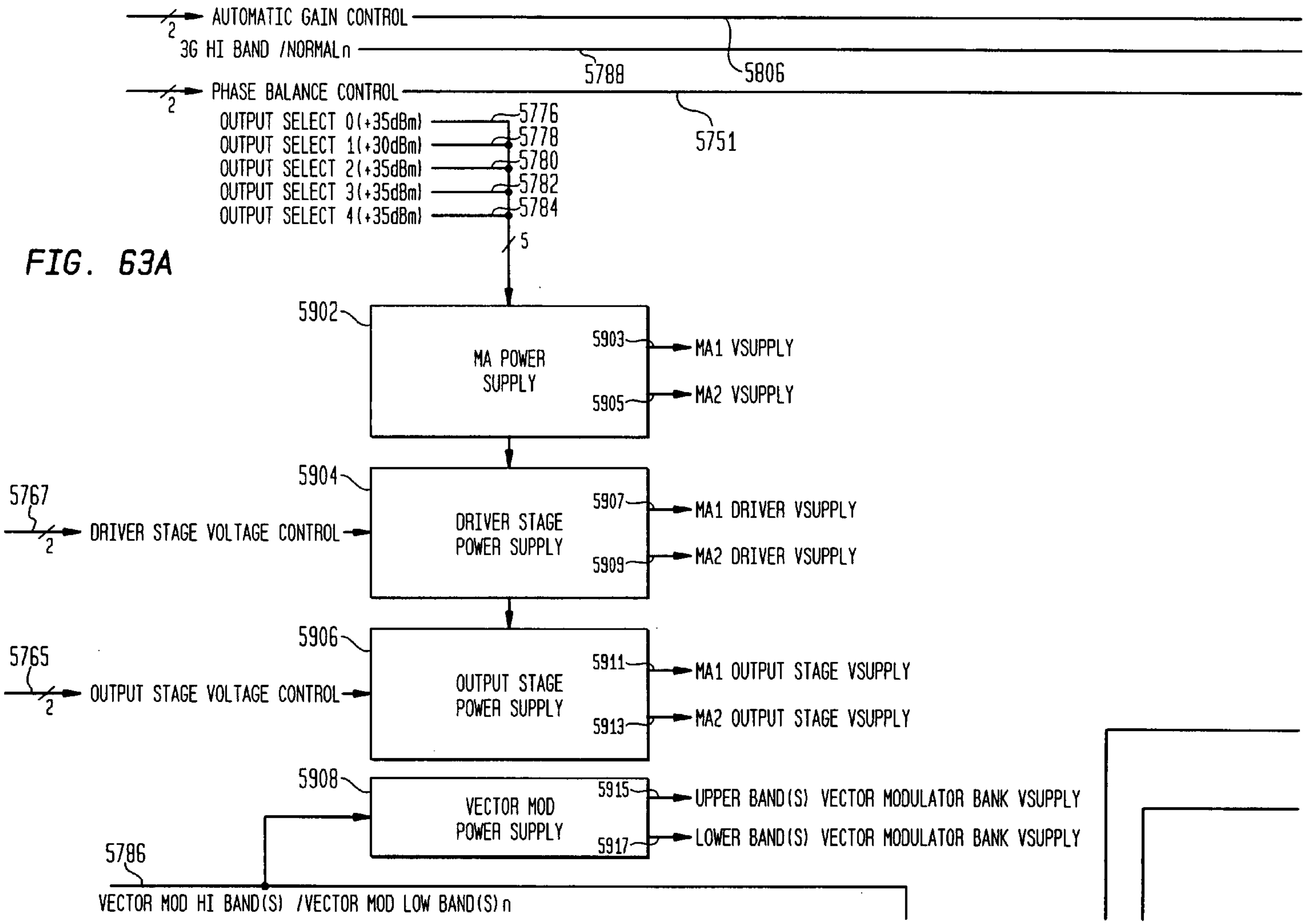
AUTOMATIC GAIN CONTROL
2
3G HI BAND /NORMALn
5788
5806
PHASE BALANCE CONTROL
2
5751
OUTPUT SELECT 0(+35dBm)
OUTPUT SELECT 1(+30dBm)
OUTPUT SELECT 2(+35dBm)
OUTPUT SELECT 3(+35dBm)
OUTPUT SELECT 4(+35dBm)
5776
5778
5780
5782
5784
5
5902
MA POWER SUPPLY
5903
5905
MA1 VSUPPLY
MA2 VSUPPLY
5904
5767
2
DRIVER STAGE VOLTAGE CONTROL
DRIVER STAGE POWER SUPPLY
5907
5909
MA1 DRIVER VSUPPLY
MA2 DRIVER VSUPPLY
5906
5765
2
OUTPUT STAGE VOLTAGE CONTROL
OUTPUT STAGE POWER SUPPLY
5911
5913
MA1 OUTPUT STAGE VSUPPLY
MA2 OUTPUT STAGE VSUPPLY
5908
VECTOR MOD POWER SUPPLY
5915
5917
UPPER BAND(S) VECTOR MODULATOR BANK VSUPPLY
LOWER BAND(S) VECTOR MODULATOR BANK VSUPPLY
5786
VECTOR MOD HI BAND(S) /VECTOR MOD LOW BAND(S)n

FIG. 63B

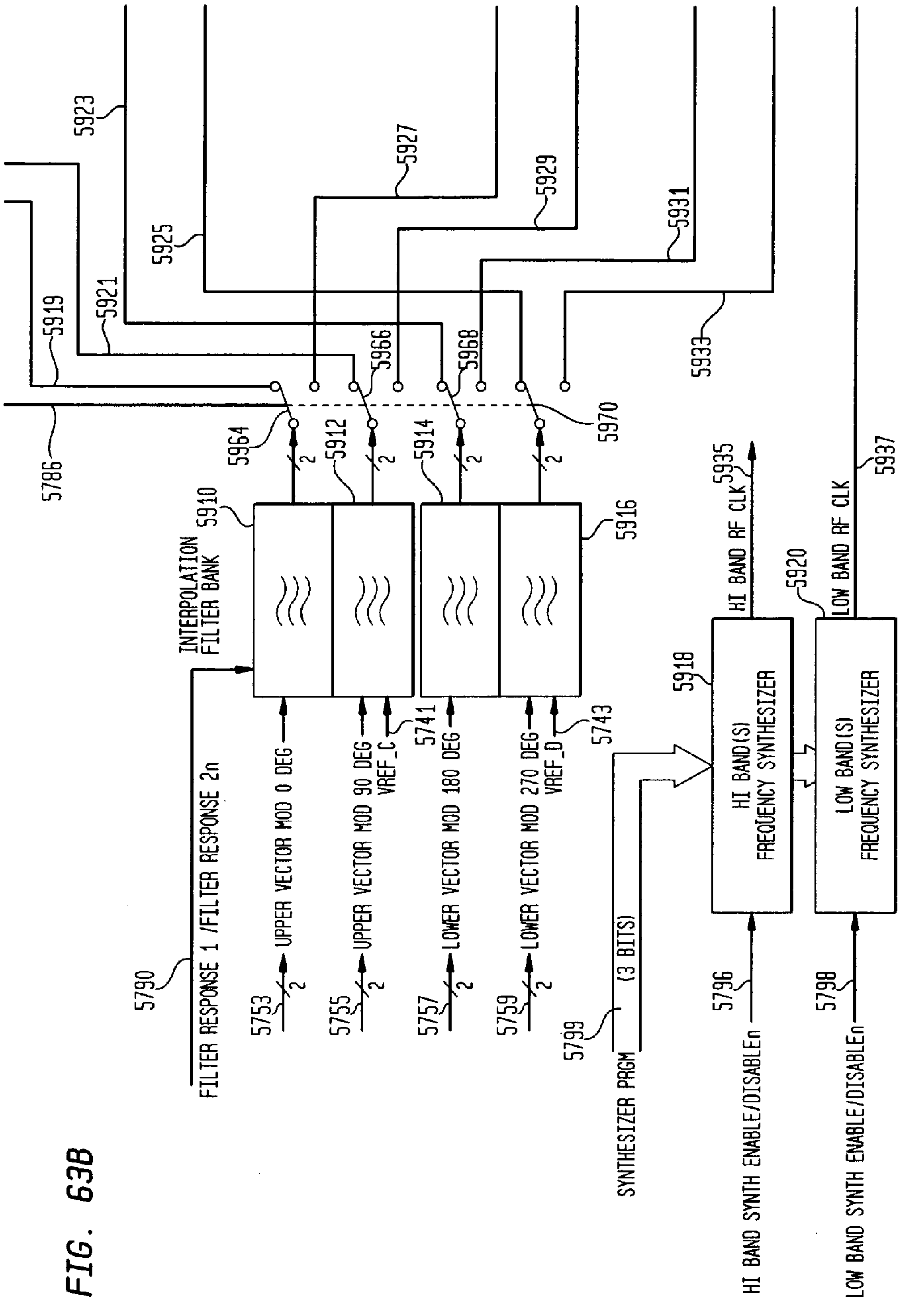
INTERPOLATION FILTER BANK
HI BAND RF CLK
LOW BAND RF CLK
HI BAND(S) FREQUENCY SYNTHESIZER
LOW BAND(S) FREQUENCY SYNTHESIZER
FILTER RESPONSE 1 /FILTER RESPONSE 2n
UPPER VECTOR MOD 0 DEG
UPPER VECTOR MOD 90 DEG
VREF_C
LOWER VECTOR MOD 180 DEG
LOWER VECTOR MOD 270 DEG
VREF_D
SYNTHESIZER PRGM
(3 BITS)
HI BAND SYNTH ENABLE/DISABLEn
LOW BAND SYNTH ENABLE/DISABLEn

FIG. 63C

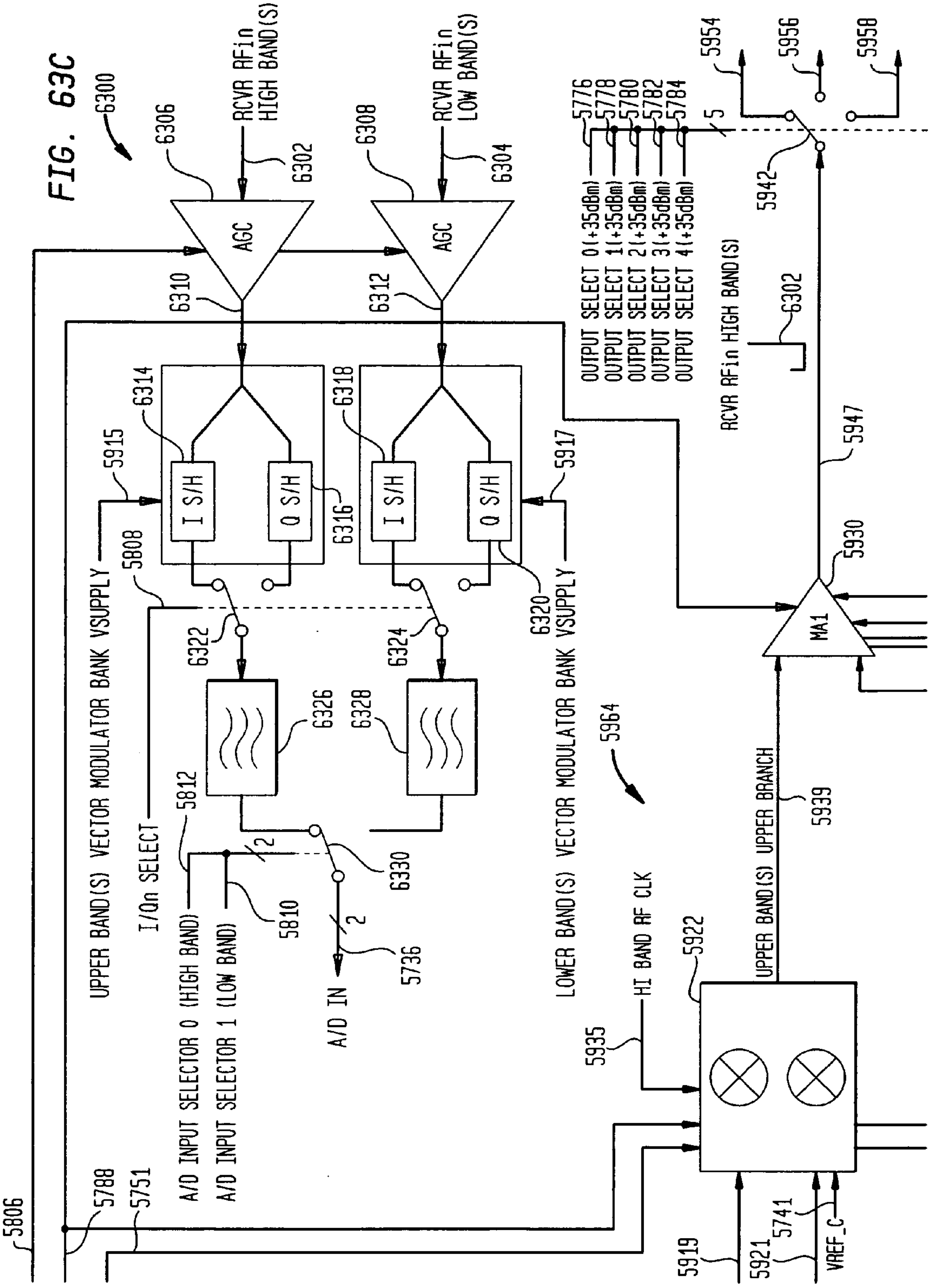
AGC
RCVR RFin
HIGH BAND(S)
AGC
RCVR RFin
LOW BAND(S)
I S/H
Q S/H
I S/H
Q S/H
UPPER BAND(S) VECTOR MODULATOR BANK VSUPPLY
LOWER BAND(S) VECTOR MODULATOR BANK VSUPPLY
I/Qn SELECT
A/D INPUT SELECTOR 0 (HIGH BAND)
A/D INPUT SELECTOR 1 (LOW BAND)
A/D IN
OUTPUT SELECT 0(+35dBm)
OUTPUT SELECT 1(+35dBm)
OUTPUT SELECT 2(+35dBm)
OUTPUT SELECT 3(+35dBm)
OUTPUT SELECT 4(+35dBm)
RCVR RFin HIGH BAND(S)
MA1
UPPER BAND(S) UPPER BRANCH
HI BAND RF CLK
VREF_C

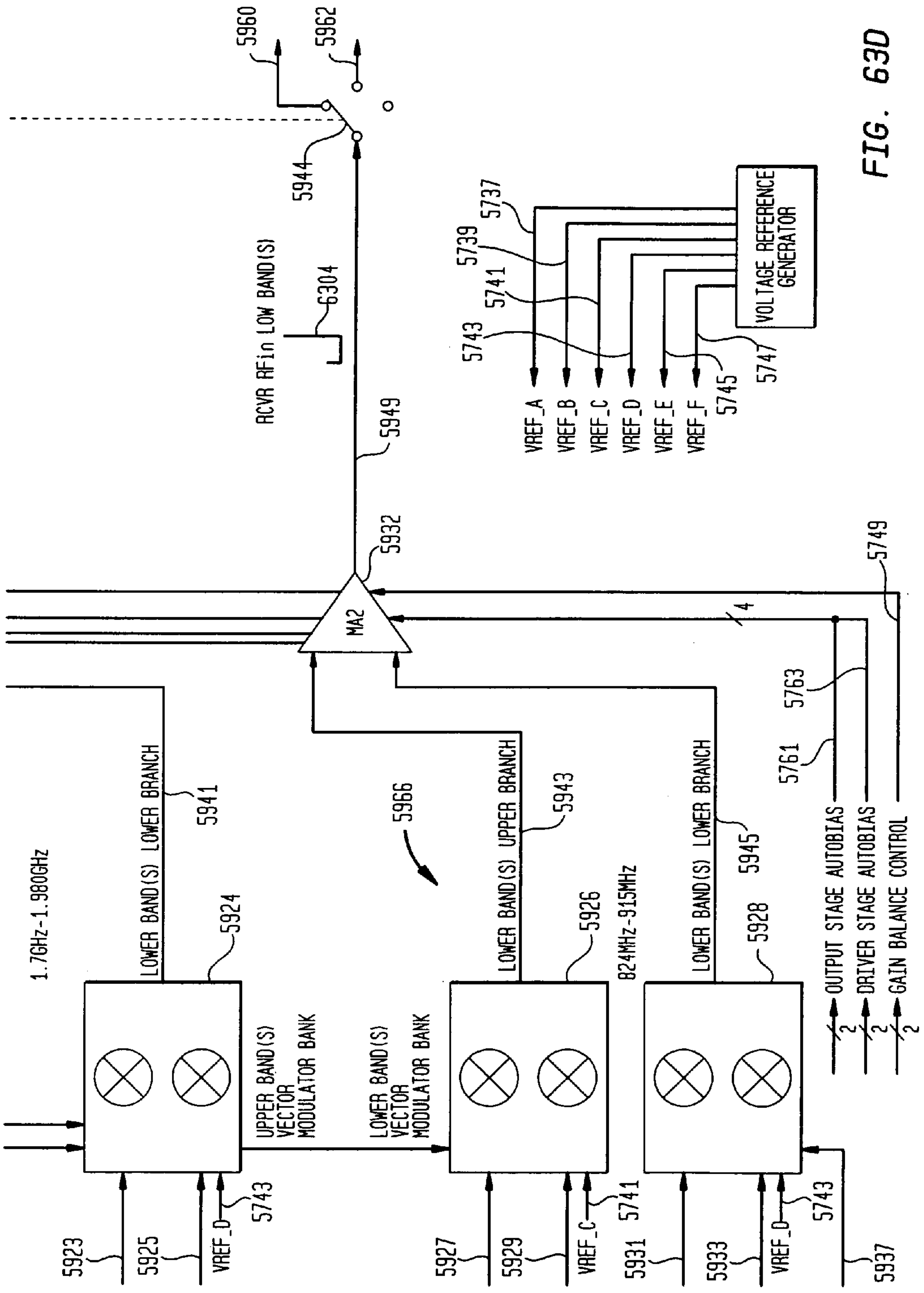
RCVR RFin LOW BAND(S)
MA2
VOLTAGE REFERENCE GENERATOR
VREF_A
VREF_B
VREF_C
VREF_D
VREF_E
VREF_F
OUTPUT STAGE AUTOBIAS
DRIVER STAGE AUTOBIAS
GAIN BALANCE CONTROL
1.7GHz-1.980GHz
LOWER BAND(S) LOWER BRANCH
LOWER BAND(S) UPPER BRANCH
824MHz-915MHz
LOWER BAND(S) LOWER BRANCH
UPPER BAND(S) VECTOR MODULATOR BANK
LOWER BAND(S) VECTOR MODULATOR BANK
VREF_D
VREF_C
VREF_D

FIG. 63D

FIG. 64

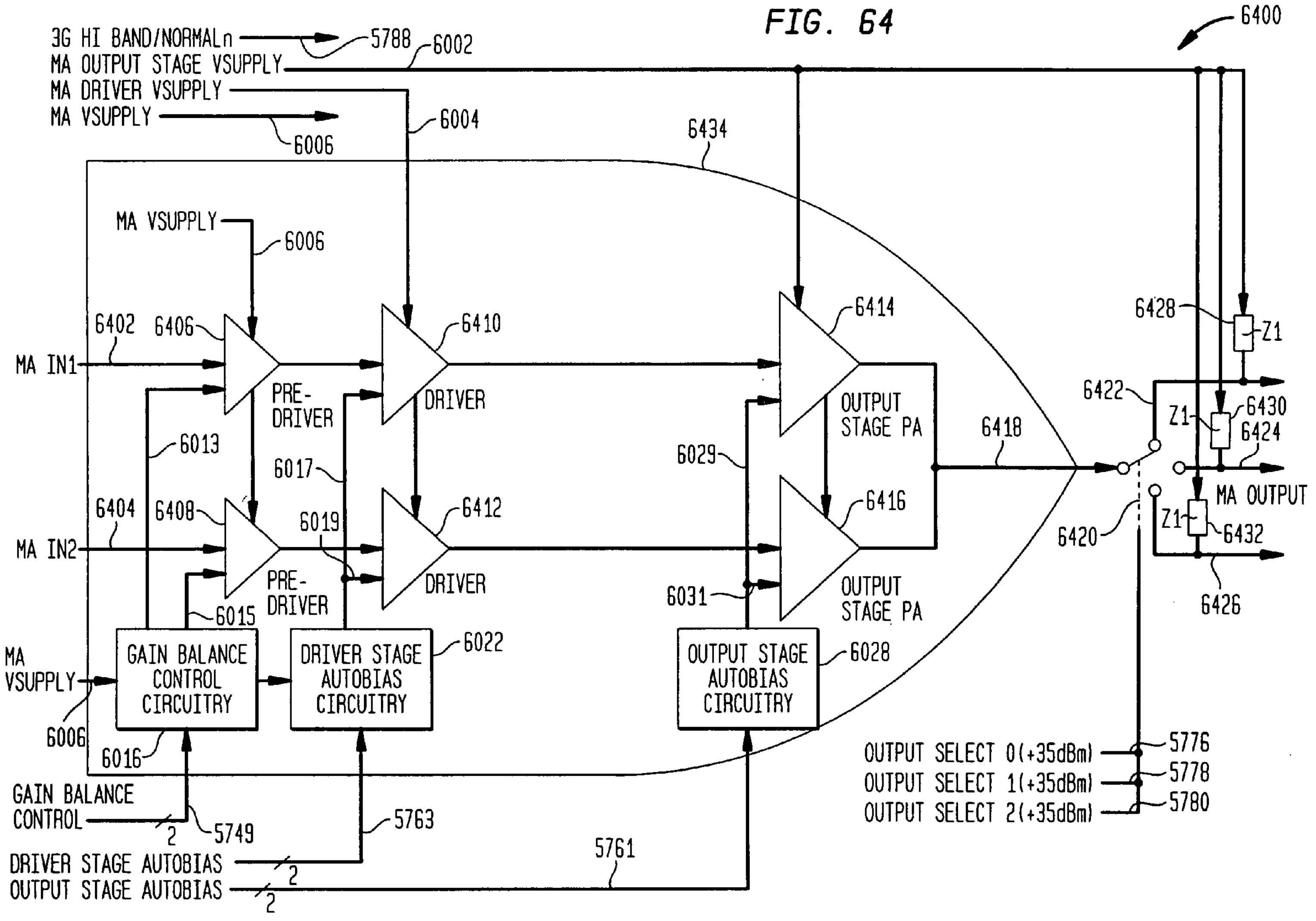
3G HI BAND/NORMALn
MA OUTPUT STAGE VSUPPLY
MA DRIVER VSUPPLY
MA VSUPPLY
MA VSUPPLY
MA IN1
PRE-DRIVER
DRIVER
OUTPUT STAGE PA
MA IN2
PRE-DRIVER
DRIVER
OUTPUT STAGE PA
GAIN BALANCE CONTROL CIRCUITRY
DRIVER STAGE AUTOBIAS CIRCUITRY
OUTPUT STAGE AUTOBIAS CIRCUITRY
MA VSUPPLY
GAIN BALANCE CONTROL
DRIVER STAGE AUTOBIAS
OUTPUT STAGE AUTOBIAS
Z1
Z1
MA OUTPUT
Z1
OUTPUT SELECT 0(+35dBm)
OUTPUT SELECT 1(+35dBm)
OUTPUT SELECT 2(+35dBm)

FIG. 65

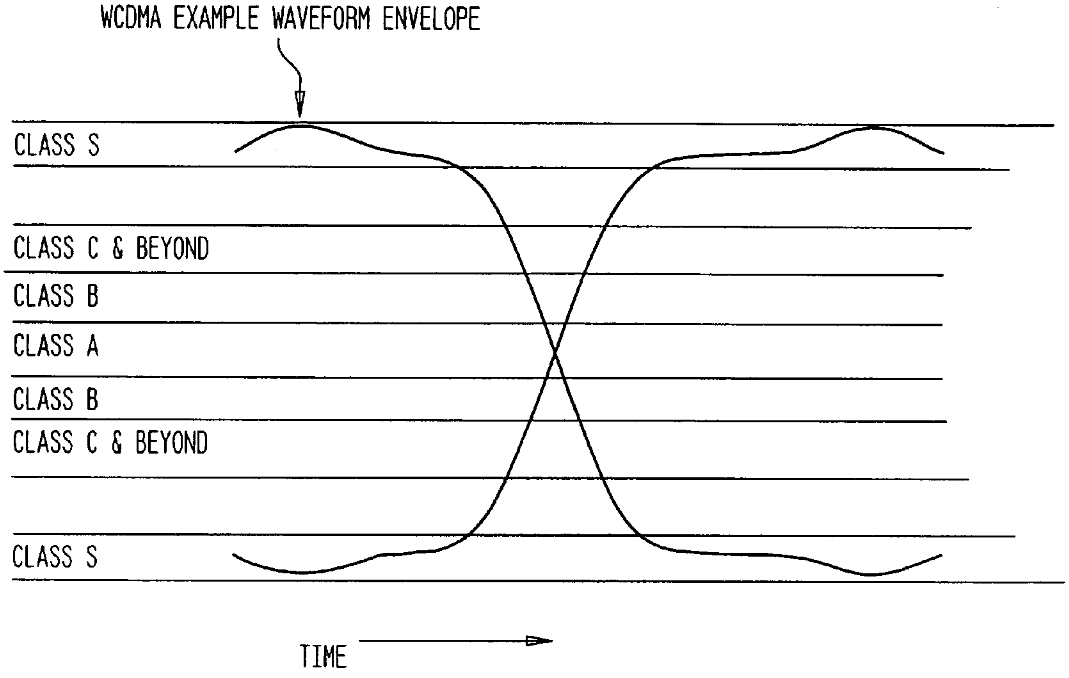
WCDMA EXAMPLE WAVEFORM ENVELOPE
CLASS S
CLASS C & BEYOND
CLASS B
CLASS A
CLASS B
CLASS C & BEYOND
CLASS S
TIME

FIG. 66

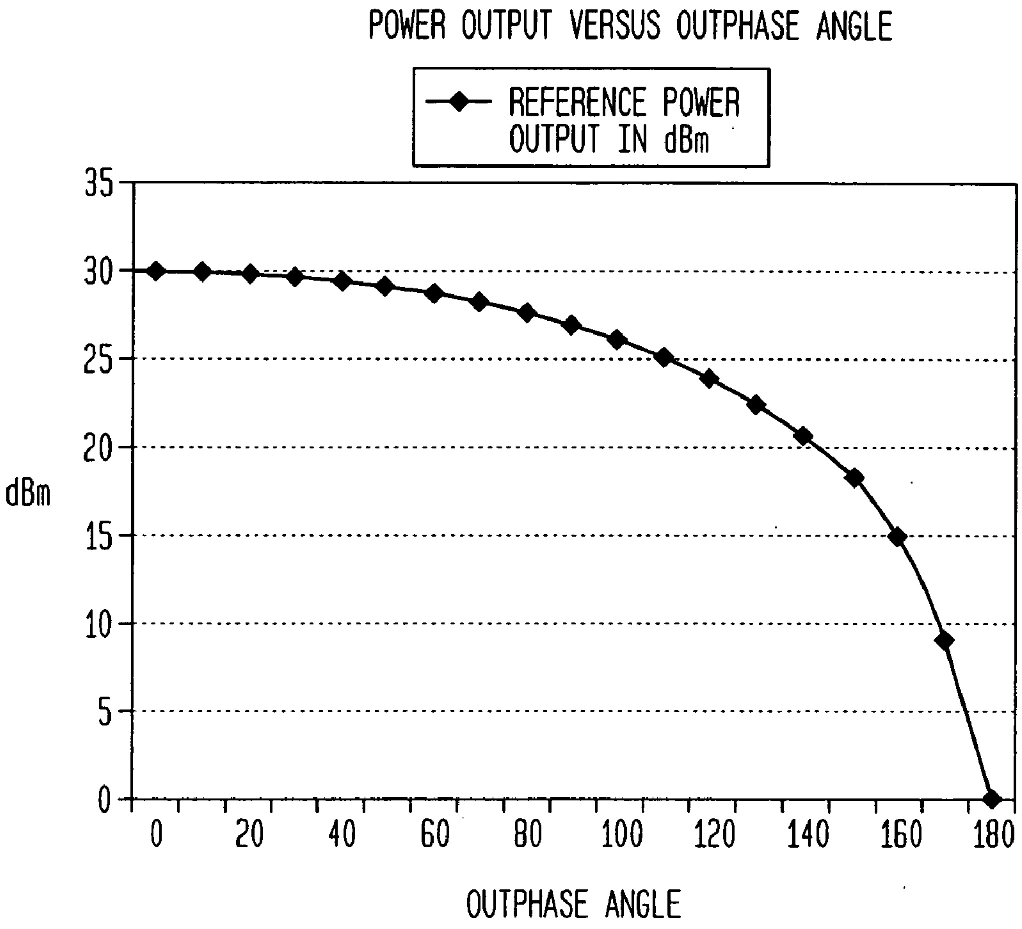
POWER OUTPUT VERSUS OUTPHASE ANGLE
REFERENCE POWER OUTPUT IN dBm
35
30
25
20
15
10
5
0
dBm
0
20
40
60
80
100
120
140
160
180
OUTPHASE ANGLE

FIG. 67

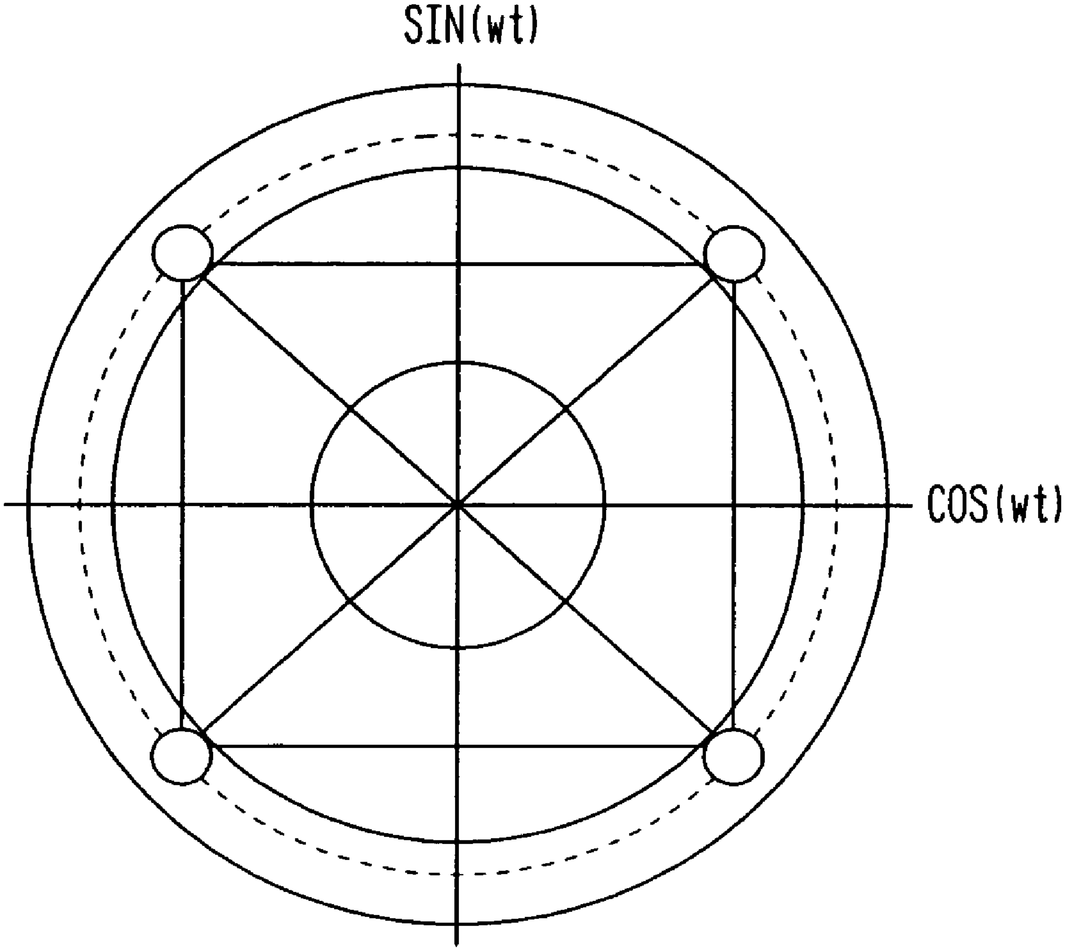
SIN(wt)
COS(wt)

FIG. 68

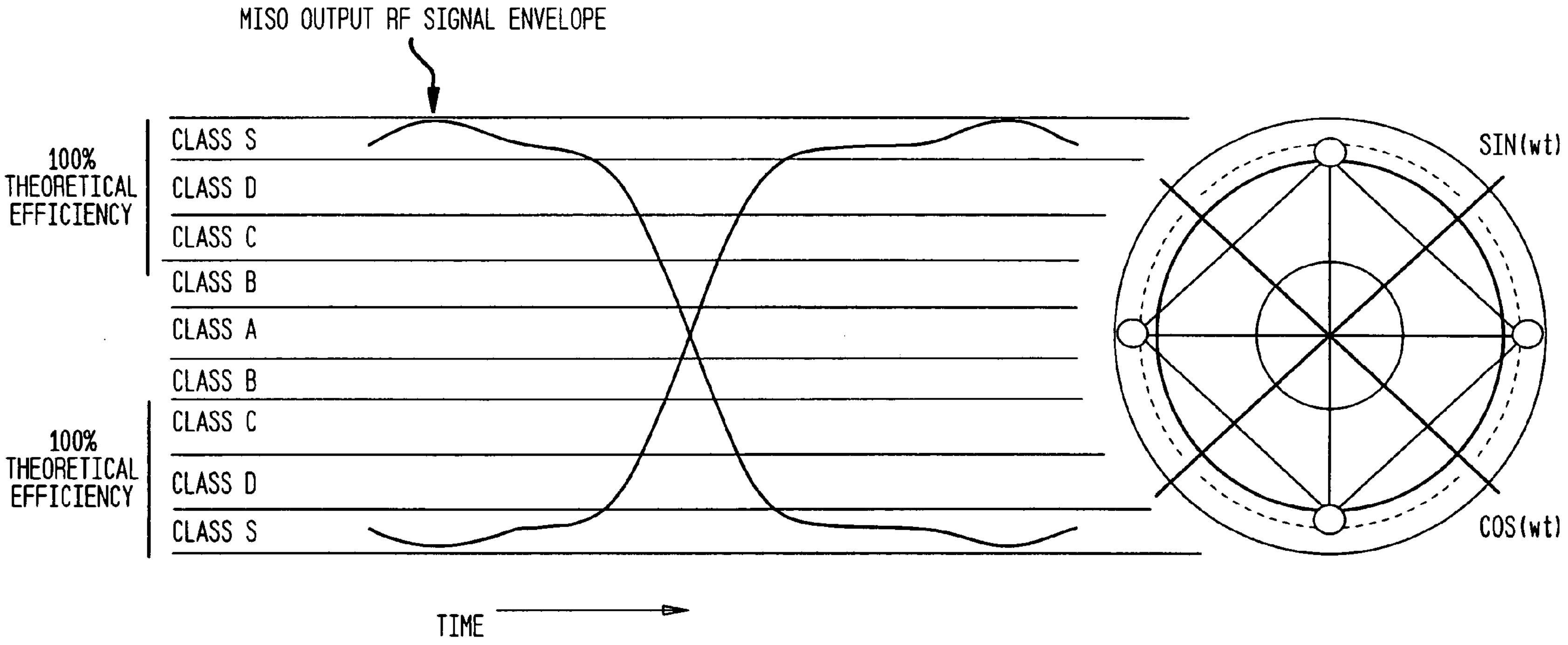
MISO OUTPUT RF SIGNAL ENVELOPE
100% THEORETICAL EFFICIENCY
CLASS S
CLASS D
CLASS C
CLASS B
CLASS A
CLASS B
100% THEORETICAL EFFICIENCY
CLASS C
CLASS D
CLASS S
TIME
SIN(wt)
COS(wt)

FIG. 69

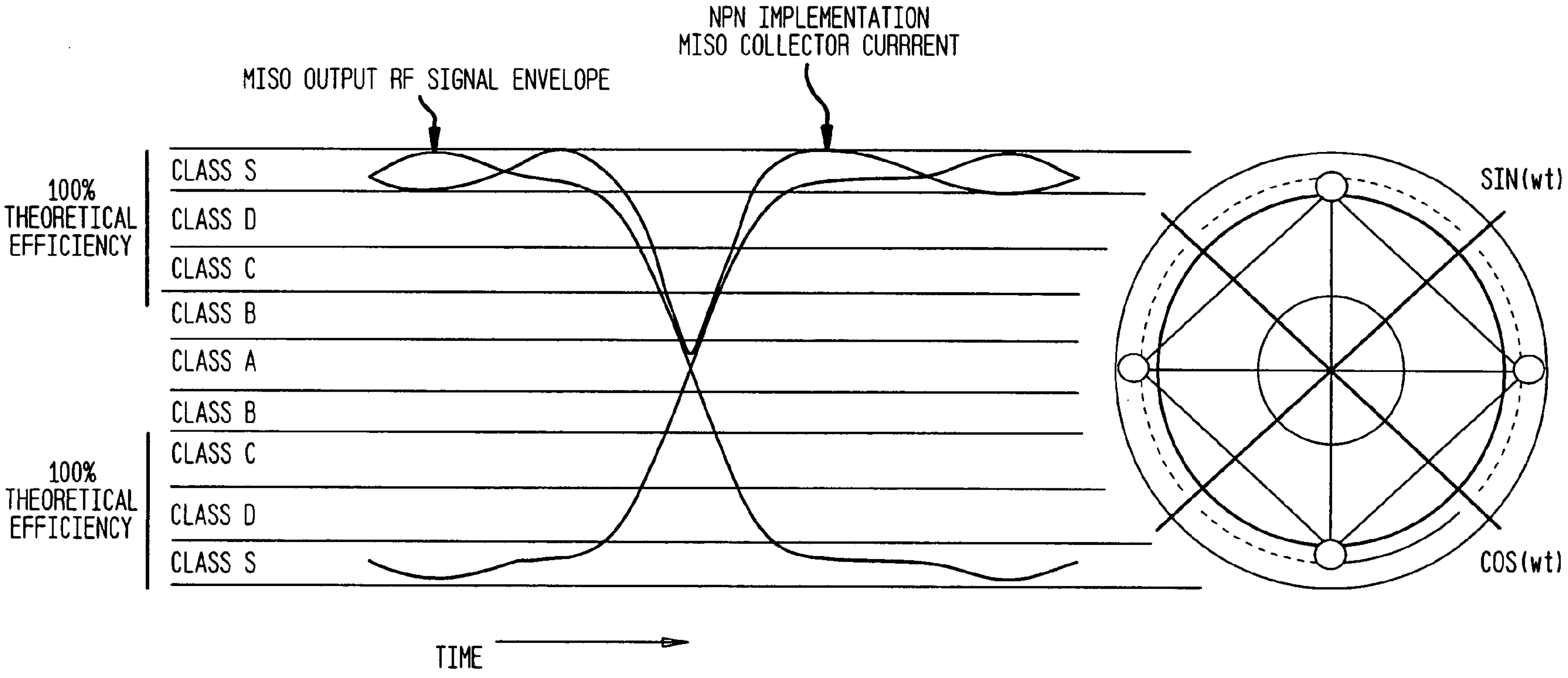
NPN IMPLEMENTATION
MISO COLLECTOR CURRRENT
MISO OUTPUT RF SIGNAL ENVELOPE
100%
THEORETICAL
EFFICIENCY
CLASS S
CLASS D
CLASS C
CLASS B
CLASS A
CLASS B
CLASS C
CLASS D
CLASS S
100%
THEORETICAL
EFFICIENCY
SIN(wt)
COS(wt)
TIME

FIG. 70

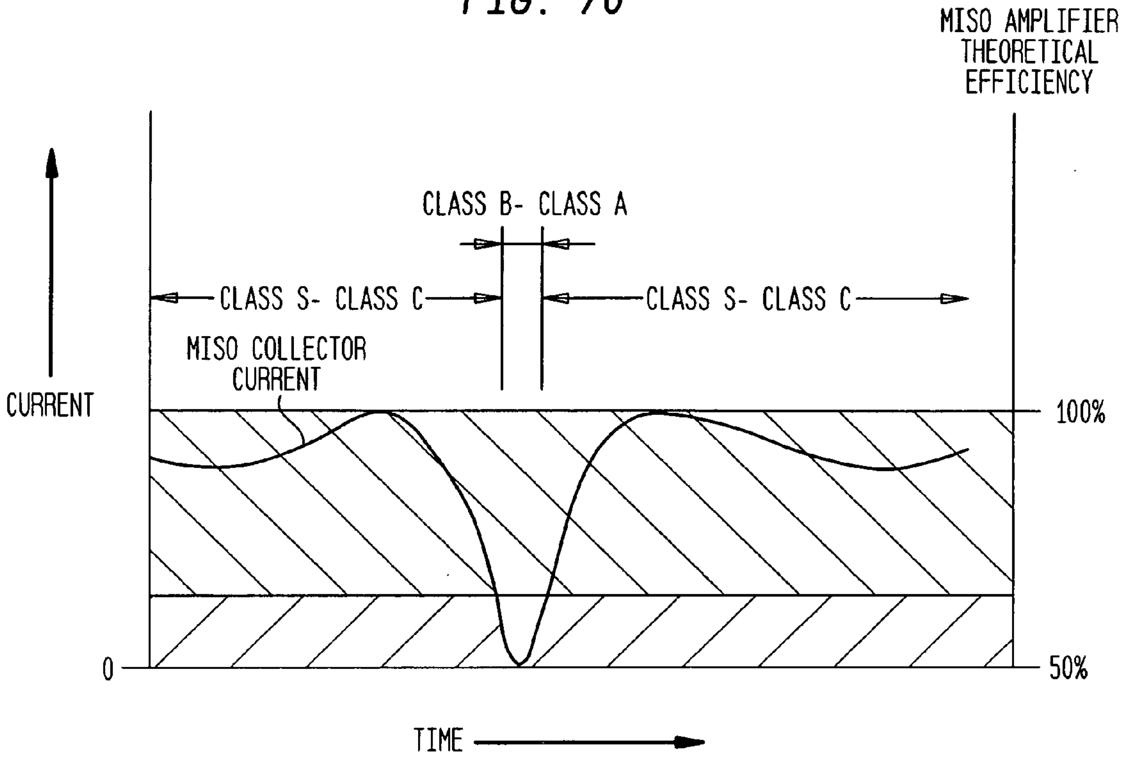
MISO AMPLIFIER THEORETICAL EFFICIENCY
CLASS B- CLASS A
CLASS S- CLASS C
CLASS S- CLASS C
MISO COLLECTOR CURRENT
CURRENT
100%
0
50%
TIME

FIG. 71

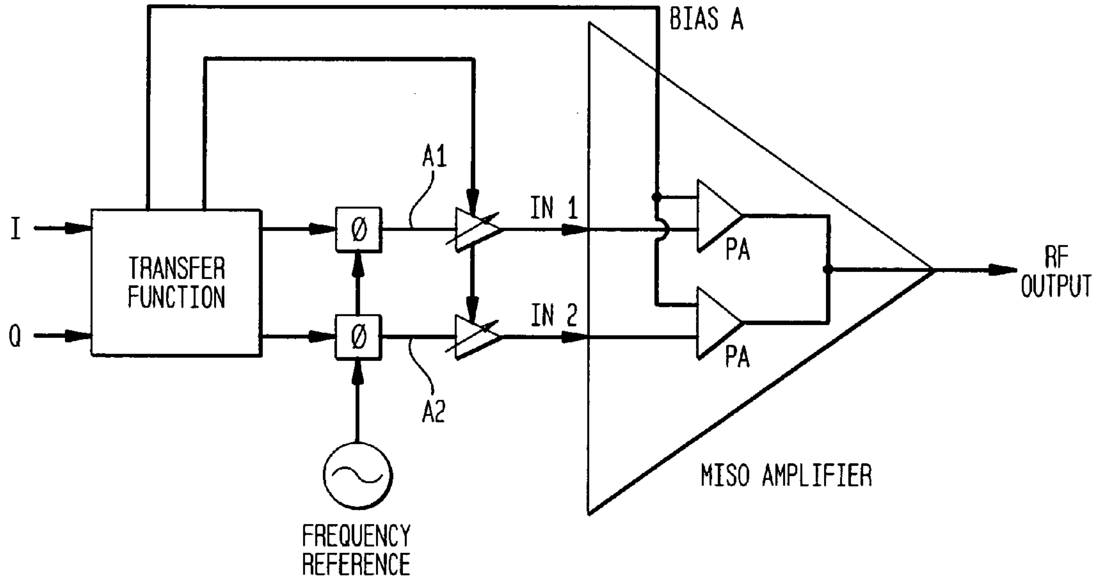
BIAS A
A1
IN 1
I
TRANSFER FUNCTION
Q
Ø
Ø
IN 2
PA
PA
RF OUTPUT
A2
MISO AMPLIFIER
FREQUENCY REFERENCE

FIG. 72

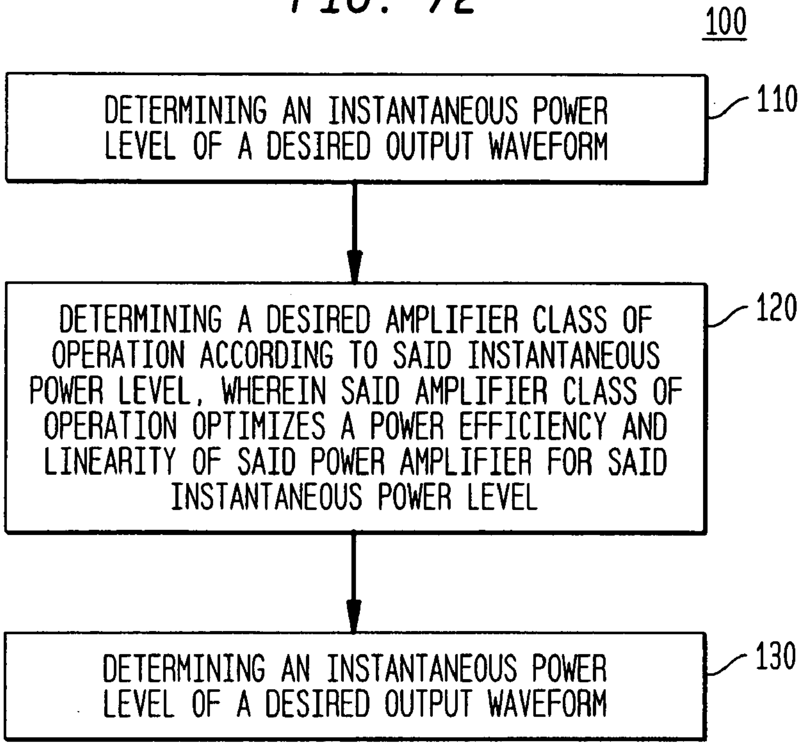
100
DETERMINING AN INSTANTANEOUS POWER LEVEL OF A DESIRED OUTPUT WAVEFORM
110
DETERMINING A DESIRED AMPLIFIER CLASS OF OPERATION ACCORDING TO SAID INSTANTANEOUS POWER LEVEL, WHEREIN SAID AMPLIFIER CLASS OF OPERATION OPTIMIZES A POWER EFFICIENCY AND LINEARITY OF SAID POWER AMPLIFIER FOR SAID INSTANTANEOUS POWER LEVEL
120
DETERMINING AN INSTANTANEOUS POWER LEVEL OF A DESIRED OUTPUT WAVEFORM
130

FIG. 73

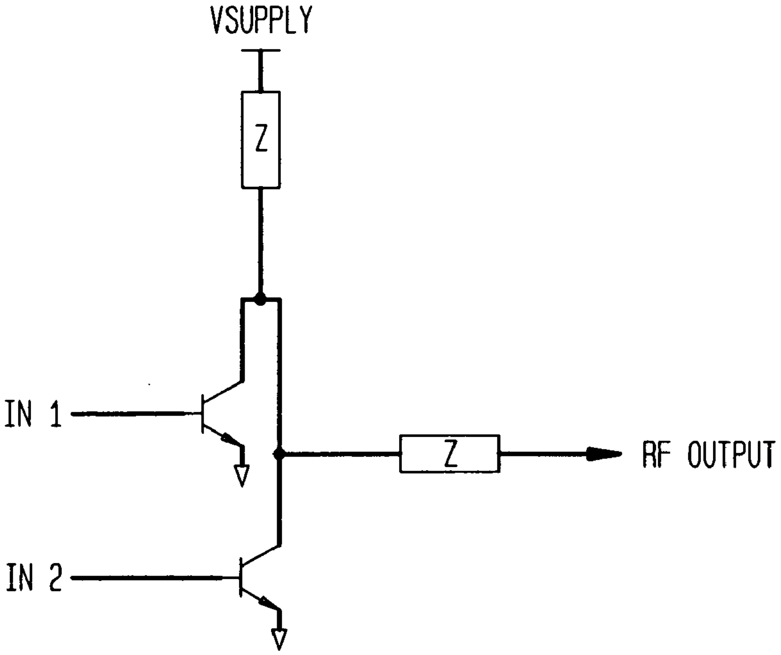
VSUPPLY
Z
IN 1
Z
RF OUTPUT
IN 2

FIG. 74

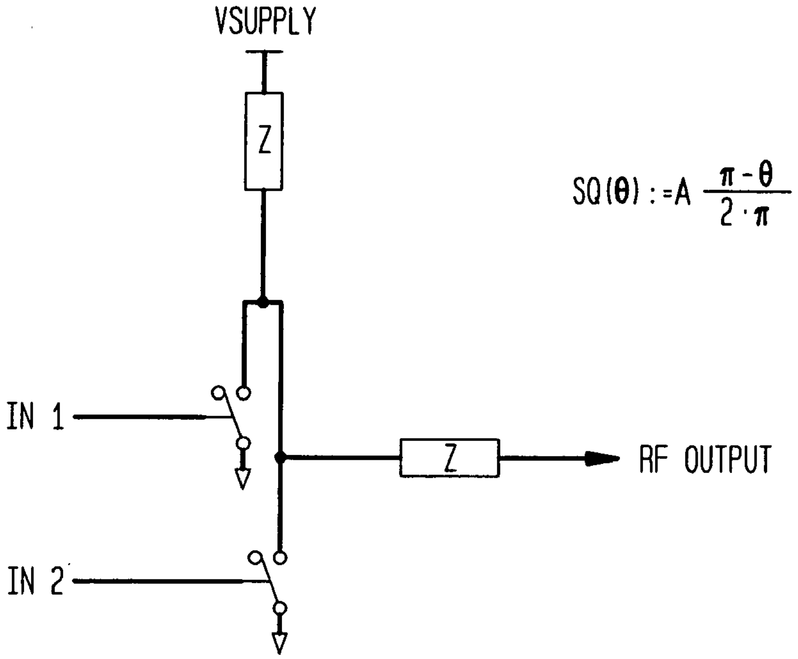
VSUPPLY
Z
IN 1
IN 2
Z
RF OUTPUT

$$SQ(\theta) := A\,\frac{\pi-\theta}{2\cdot\pi}$$

FIG. 75

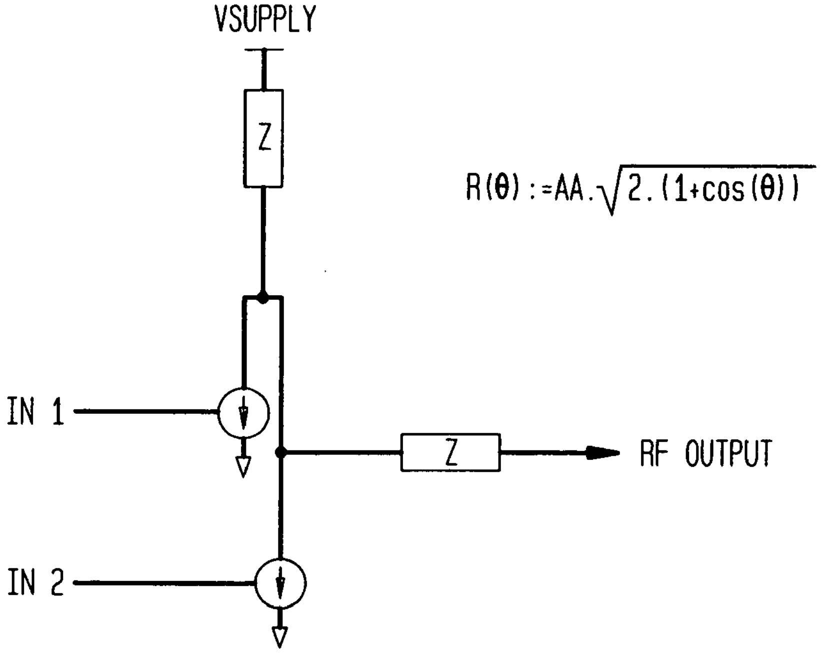
VSUPPLY
Z
IN 1
IN 2
Z
RF OUTPUT

$$R(\theta) := AA.\sqrt{2.(1+\cos(\theta))}$$

FIG. 76

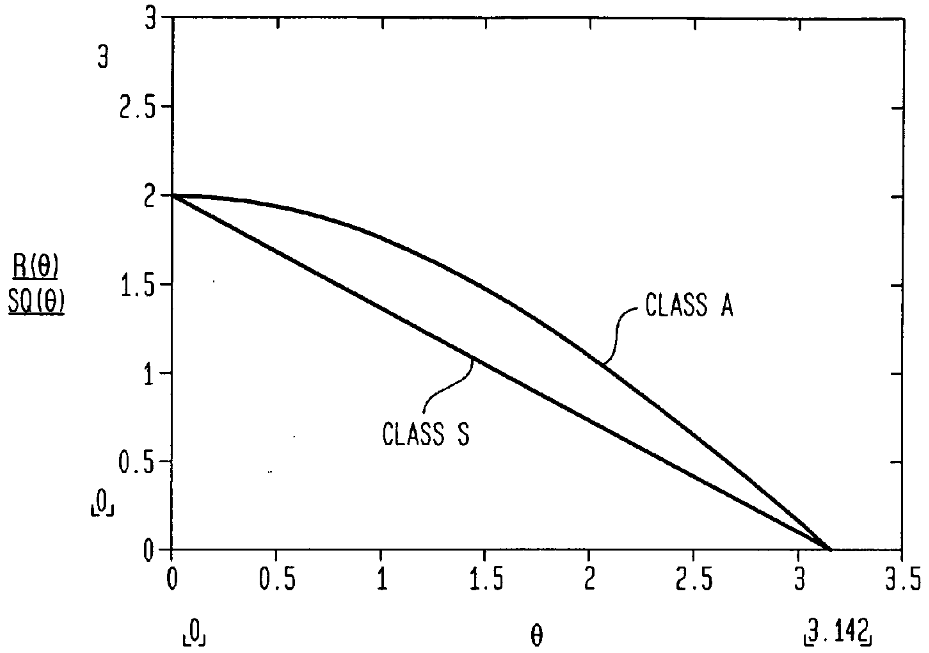
3
3
2.5
2
1.5
1
0.5
0
0
R(θ)
SQ(θ)
CLASS A
CLASS S
0
0.5
1
1.5
2
2.5
3
3.5
0
θ
3.142

FIG. 77

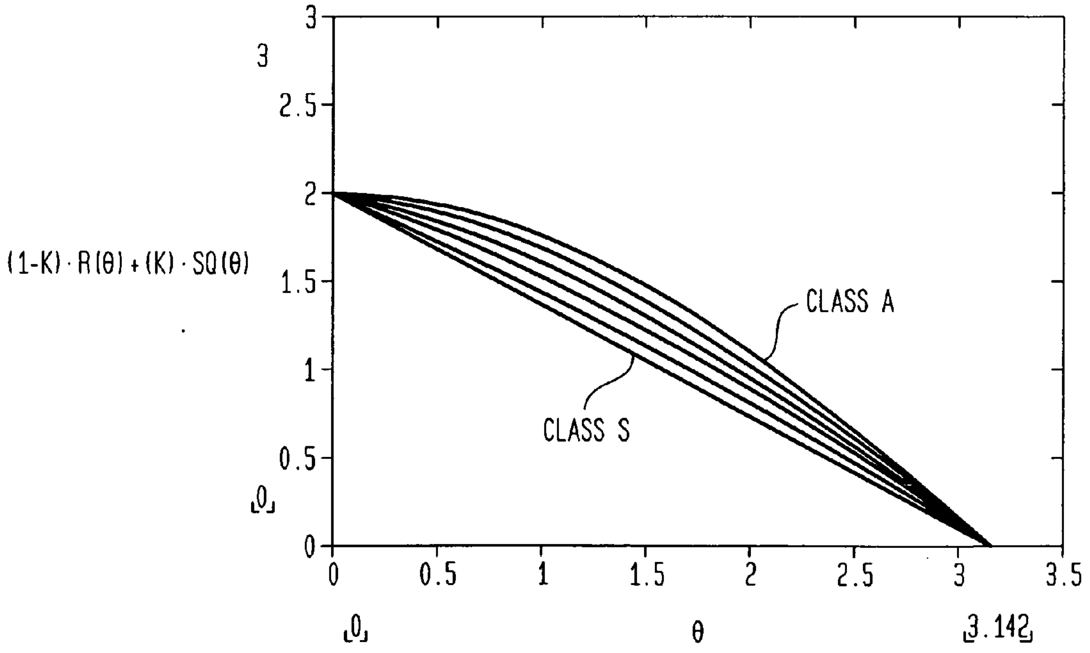
3
3
2.5
2
1.5
1
0.5
0
0
(1-K)·R(θ)+(K)·SQ(θ)
CLASS A
CLASS S
0
0.5
1
1.5
2
2.5
3
3.5
0
θ
3.142

FIG. 78

$$R\cdot\sin(\omega\cdot t+\delta)=A1\ \sin(\omega\cdot t+\phi e1(t))\ +\ A2\ \sin(\omega\cdot t+\theta(t))\ +\phi e2(t))$$

$$R\cdot\sin(\omega\cdot t+\delta)=$$

$$A1\cdot\sin(\omega\cdot t)\cdot\cos(\phi e1(t))\ +\ A1\cdot\sin(\phi e1(t))\cdot\cos(\omega\cdot t)\ +\ A2\cdot\sin(\omega\cdot t+\theta(t))\cdot\cos(\phi e2(t))\ +\ A2\cdot\sin(\phi e2(t))\cdot\cos(\omega\cdot t+\theta(t))\quad \text{Angle sum relationship (6.1.13)}:\ \sin(\beta)=\sin(X)*\cos(\beta)+\cos(X)*\sin(\beta)$$

$$A1\cdot\sin(\omega\cdot t)\cdot\cos(\phi e1(t))\ +\ A1\cdot\sin(\phi e1(t))\cdot\cos(\omega\cdot t)\ +\ A2\cdot\sin(\omega\cdot t+\theta(t))\cdot\cos(\phi e2(t))\ +\ A2\cdot\sin(\phi e2(t))\cdot\cos(\omega\cdot t+\theta(t))\quad \text{Angle sum relationship (6.1.13)}:\ \cos(\beta)=\cos(X)*\cos(\beta)-\sin(X)*\sin(\beta)$$

$$A1\cdot\sin(\omega\cdot t)\cdot\cos(\phi e1(t))\ +\ A1\cdot\sin(\phi e1(t))\cdot\cos(\omega\cdot t)\ +\ A2\cdot\sin(\omega\cdot t)\cdot\cos(\theta(t))\cdot\cos(\phi e2(t))\ +\ A2\cdot\sin(\theta(t))\cdot\cos(\omega\cdot t)\cdot\cos(\phi e2(t))\ +\ A2\cdot\cos(\omega\cdot t)\cdot\cos(\theta(t))\cdot\sin(\phi e2(t))\ -\ A2\cdot\sin(\omega\cdot t)\cdot\sin(\theta(t))\cdot\sin(\phi e2(t))$$

$$\sin(\omega\cdot t)\cdot(A1\cdot\cos(\phi e1(t))\ +\ A2\cdot\cos(\theta(t))\cdot\cos(\phi e2(t))\ -\ A2\cdot\sin(\theta(t))\cdot\sin(\phi e2(t))\ +\ \cos(\omega\cdot t)\cdot(A1\cdot\sin(\phi e1(t))\ +\ A2\cdot\sin(\theta(t))\cdot\cos(\phi e2(t))\ +\ A2\cdot\cos(\theta(t))\cdot\sin(\phi e2(t))$$

$$R(t)=\sqrt{(A1\cdot\cos(\phi e1(t))\ +\ A2\cdot\cos(\theta(t))\cdot\cos(\phi e2(t))\ -\ A2\cdot\sin(\theta(t))\cdot\sin(\phi e2(t))^{2}\ +\ (A1\cdot\sin(\phi e1(t))\ +\ A2\cdot\sin(\theta(t))\cdot\cos(\phi e2(t))\ +\ A2\cdot\cos(\theta(t))\cdot\sin(\phi e2(t))^{2}}$$

$$R(t)=(2\cdot A1\cdot\cos(\phi e1(t))\cdot A2\cdot\cos(\theta(t))\cdot\cos(\phi e2(t))\ -\ 2\cdot A1\cdot\cos(\phi e1(t))\cdot A2\cdot\sin(\theta(t))\cdot\sin(\phi e2(t))\ +\ A2^{2}\ +\ A1^{2}\ +\ 2\cdot A1\cdot\sin(\phi e1(t))\cdot A2\cdot\sin(\theta(t))\cdot\cos(\phi e2(t))\ +\ 2\cdot A1\cdot\sin(\phi e1(t))\cdot A2\cdot\cos(\theta(t))\cdot\sin(\phi e2(t)))$$

$$\delta(t)=\operatorname{atan}\left[\frac{(A1\cdot\sin(\phi e1(t))\ +\ A2\cdot\sin(\theta(t))\cdot\cos(\phi e2(t))\ +\ A2\cdot\cos(\theta(t))\cdot\sin(\phi e2(t))}{A1\cdot\cos(\phi e1(t))\ +\ A2\cdot\cos(\theta(t))\cdot\cos(\phi e2(t))\ -\ A2\cdot\sin(\theta(t))\cdot\sin(\phi e2(t))}\right]$$

SYSTEMS AND METHODS OF RF TOWER TRANSMISSION, MODULATION, AND AMPLIFICATION, INCLUDING EMBODIMENTS FOR COMPENSATING FOR WAVEFORM DISTORTION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims the benefit of U.S. Provisional Patent Application No. 60/794,121 filed on Apr. 24, 2006, U.S. Provisional Patent Application No. 60/797,653 filed on May 5, 2006, and U.S. Provisional Patent Application No. 60/798,750 filed on May 9, 2006, all of which are incorporated herein by reference in their entireties.

BACKGROUND OF THE INVENTION

[0002] 1. Field of the Invention

[0003] The present invention relates generally to RF power transmission, modulation, and amplification. More particularly, the invention relates to methods and systems for vector combining power amplification.

[0004] 2. Background Art

[0005] In power amplifiers, a complex tradeoff typically exists between linearity and power efficiency.

[0006] Linearity is determined by a power amplifier's operating range on a characteristic curve that relates its input to output variables—the more linear the operating range the more linear the power amplifier is said to be. Linearity is a desired characteristic of a power amplifier. In one aspect, for example, it is desired that a power amplifier uniformly amplifies signals of varying amplitude, and/or phase and/or frequency. Accordingly, linearity is an important determiner of the output signal quality of a power amplifier.

[0007] Power efficiency can be calculated using the relationship of the total power delivered to a load divided by the total power supplied to the amplifier. For an ideal amplifier, power efficiency is 100%. Typically, power amplifiers are divided into classes which determine the amplifier's maximum theoretical power efficiency. Power efficiency is clearly a desired characteristic of a power amplifier—particularly, in wireless communication systems where power consumption is significantly dominated by the power amplifier.

[0008] Unfortunately, the traditional tradeoff between linearity and efficiency in power amplifiers is such that the more linear a power amplifier is the less power efficient it is. For example, the most linear amplifier is biased for class A operation, which is the least efficient class of amplifiers. On the other hand, higher class amplifiers such as class B,C,D, E, etc, are more power efficient, but are considerably nonlinear which can result in spectrally distorted output signals.

[0009] The tradeoff described above is further accentuated by typical wireless communication signals. Wireless communication signals, such as OFDM, CDMA, and W-CDMA for example, are generally characterized by their peak-to-average power ratios. The larger the signal's peak to average ratio the more non-linear distortion will be produced when non-linear amplifiers are employed.

[0010] Outphasing amplification techniques have been proposed for RF amplifier designs. In several aspects, however, existing outphasing techniques are deficient in satisfying complex signal amplification requirements, particularly as defined by wireless communication standards, for example.

[0011] In one aspect, existing outphasing techniques employ an isolating and/or a combining element when combining constant envelope constituents of a desired output signal. For example, it is commonly the case that a power combiner is used to combine the constituent signals. This combining approach, however, typically results in a degradation of output signal power due to insertion loss and limited bandwidth, and, correspondingly, a decrease in power efficiency.

[0012] In another aspect, the typically large size of combining elements precludes having them in monolithic amplifier designs.

[0013] What is needed therefore are power amplification methods and systems that solve the deficiencies of existing power amplifying techniques while maximizing power efficiency and minimizing non-linear distortion. Further, power amplification methods and systems that can be implemented without the limitations of traditional power combining circuitry and techniques are needed.

BRIEF SUMMARY OF THE INVENTION

[0014] Embodiments for vector combining power amplification are disclosed herein.

[0015] In one embodiment, a plurality of substantially constant envelope signals are individually amplified, then combined to form a desired time-varying complex envelope signal. Phase and/or frequency characteristics of one or more of the signals are controlled to provide the desired phase, frequency, and/or amplitude characteristics of the desired time-varying complex envelope signal.

[0016] In another embodiment, a time-varying complex envelope signal is decomposed into a plurality of substantially constant envelope constituent signals. The constituent signals are amplified, and then re-combined to construct an amplified version of the original time-varying envelope signal.

[0017] Embodiments of the invention can be practiced with modulated carrier signals and with baseband information and clock signals. Embodiments of the invention also achieve frequency up-conversion. Accordingly, embodiments of the invention represent integrated solutions for frequency up-conversion, amplification, and modulation.

[0018] Embodiments of the invention can be implemented with analog and/or digital controls. The invention can be implemented with analog components or with a combination of analog components and digital components. In the latter embodiment, digital signal processing can be implemented in an existing baseband processor for added cost savings.

[0019] Additional features and advantages of the invention will be set forth in the description that follows. Yet further features and advantages will be apparent to a person skilled in the art based on the description set forth herein or may be learned by practice of the invention. The advantages of the invention will be realized and attained by the structure and methods particularly pointed out in the written description and claims hereof as well as the appended drawings.

[0020] It is to be understood that both the foregoing summary and the following detailed description are exemplary and explanatory and are intended to provide further explanation of embodiments of the invention as claimed.

BRIEF DESCRIPTION OF THE FIGURES

[0021] Embodiments of the present invention will be described with reference to the accompanying drawings, wherein generally like reference numbers indicate identical or functionally similar elements. Also, generally, the left-most digit(s) of the reference numbers identify the drawings in which the associated elements are first introduced.

[0022] FIG. **1**A is an example that illustrates the generation of an exemplary time-varying complex envelope signal.

[0023] FIG. **1**B is another example that illustrates the generation of an exemplary time-varying complex envelope signal.

[0024] FIG. **1**C is an example that illustrates the generation of an exemplary time-varying complex envelope signal from the sum of two or more constant envelope signals.

[0025] FIG. **1**D illustrates the power amplification of an example time-varying complex envelope signal according to an embodiment of the present invention.

[0026] FIG. **1**E is a block diagram that illustrates a vector power amplification embodiment of the present invention.

[0027] FIG. **1** illustrates a phasor representation of a signal.

[0028] FIG. **2** illustrates a phasor representation of a time-varying complex envelope signal.

[0029] FIGS. **3**A-**3**C illustrate an example modulation to generate a time-varying complex envelope signal.

[0030] FIG. **3**D is an example that illustrates constant envelope decomposition of a time-varying envelope signal.

[0031] FIG. **4** is a phasor diagram that illustrates a Cartesian 4-Branch Vector Power Amplification (VPA) method of an embodiment of the present invention.

[0032] FIG. **5** is a block diagram that illustrates an exemplary embodiment of the Cartesian 4-Branch VPA method.

[0033] FIG. **6** is a process flowchart embodiment for power amplification according to the Cartesian 4-Branch VPA method.

[0034] FIG. **7**A is a block diagram that illustrates an exemplary embodiment of a vector power amplifier for implementing the Cartesian 4-Branch VPA method.

[0035] FIG. **7**B is a block diagram that illustrates another exemplary embodiment of a vector power amplifier for implementing the Cartesian 4-Branch VPA method.

[0036] FIG. **8**A is a block diagram that illustrates another exemplary embodiment of a vector power amplifier according to the Cartesian 4-Branch VPA method.

[0037] FIG. **8**B is a block diagram that illustrates another exemplary embodiment of a vector power amplifier according to the Cartesian 4-Branch VPA method.

[0038] FIG. **8**C is a block diagram that illustrates another exemplary embodiment of a vector power amplifier according to the Cartesian 4-Branch VPA method.

[0039] FIG. **8**D is a block diagram that illustrates another exemplary embodiment of a vector power amplifier according to the Cartesian 4-Branch VPA method.

[0040] FIGS. **9**A-**9**B are phasor diagrams that illustrate a Cartesian-Polar-Cartesian-Polar (CPCP) 2-Branch Vector Power Amplification (VPA) method of an embodiment of the present invention.

[0041] FIG. **10** is a block diagram that illustrates an exemplary embodiment of the CPCP 2-Branch VPA method.

[0042] FIG. **10**A is a block diagram that illustrates another exemplary embodiment of the CPCP 2-Branch VPA method.

[0043] FIG. **11** is a process flowchart embodiment for power amplification according to the CPCP 2-Branch VPA method.

[0044] FIG. **12** is a block diagram that illustrates an exemplary embodiment of a vector power amplifier for implementing the CPCP 2-Branch VPA method.

[0045] FIG. **12**A is a block diagram that illustrates another exemplary embodiment of a vector power amplifier for implementing the CPCP 2-Branch VPA method.

[0046] FIG. **12**B is a block diagram that illustrates another exemplary embodiment of a vector power amplifier for implementing the CPCP 2-Branch VPA method.

[0047] FIG. **13** is a block diagram that illustrates another exemplary embodiment of a vector power amplifier for implementing the CPCP 2-Branch VPA method.

[0048] FIG. **13**A is a block diagram that illustrates another exemplary embodiment of a vector power amplifier for implementing the CPCP 2-Branch VPA method.

[0049] FIG. **14** is a phasor diagram that illustrates a Direct Cartesian 2-Branch Vector Power Amplification (VPA) method of an embodiment of the present invention.

[0050] FIG. **15** is a block diagram that illustrates an exemplary embodiment of the Direct Cartesian 2-Branch VPA method.

[0051] FIG. **15**A is a block diagram that illustrates another exemplary embodiment of the Direct Cartesian 2-Branch VPA method.

[0052] FIG. **16** is a process flowchart embodiment for power amplification according to the Direct Cartesian 2-Branch VPA method.

[0053] FIG. **17** is a block diagram that illustrates an exemplary embodiment of a vector power amplifier for implementing the Direct Cartesian 2-Branch VPA method.

[0054] FIG. **17**A is a block diagram that illustrates another exemplary embodiment of a vector power amplifier for implementing the Direct Cartesian 2-Branch VPA method.

[0055] FIG. **17**B is a block diagram that illustrates another exemplary embodiment of a vector power amplifier for implementing the Direct Cartesian 2-Branch VPA method.

[0056] FIG. **18** is a block diagram that illustrates another exemplary embodiment of a vector power amplifier for implementing the Direct Cartesian 2-Branch VPA method.

[0057] FIG. **18**A is a block diagram that illustrates another exemplary embodiment of a vector power amplifier for implementing the Direct Cartesian 2-Branch VPA method.

[0058] FIG. **19** is a process flowchart that illustrates an I and Q transfer function embodiment according to the Cartesian 4-Branch VPA method.

[0059] FIG. **20** is a block diagram that illustrates an exemplary embodiment of an I and Q transfer function according to the Cartesian 4-Branch VPA method.

[0060] FIG. **21** is a process flowchart that illustrates an I and Q transfer function embodiment according to the CPCP 2-Branch VPA method.

[0061] FIG. **22** is a block diagram that illustrates an exemplary embodiment of an I and Q transfer function according to the CPCP 2-Branch VPA method.

[0062] FIG. **23** is a process flowchart that illustrates an I and Q transfer function embodiment according to the Direct Cartesian 2-Branch VPA method.

[0063] FIG. **24** is a block diagram that illustrates an exemplary embodiment of an I and Q transfer function according to the Direct Cartesian 2-Branch VPA method.

[0064] FIG. **25** is a phasor diagram that illustrates the effect of waveform distortion on a representation of a signal phasor.

[0065] FIG. **26** illustrates magnitude to phase transform functions according to an embodiment of the present invention.

[0066] FIG. **27** illustrates exemplary embodiments of biasing circuitry according to embodiments of the present invention.

[0067] FIG. **28** illustrates a method of combining constant envelope signals according to an embodiment the present invention.

[0068] FIG. **29** illustrates a vector power amplifier output stage embodiment according to the present invention.

[0069] FIG. **30** is a block diagram of a power amplifier (PA) output stage embodiment.

[0070] FIG. **31** is a block diagram of another power amplifier (PA) output stage embodiment.

[0071] FIG. **32** is a block diagram of another power amplifier (PA) output stage embodiment.

[0072] FIG. **33** is a block diagram of another power amplifier (PA) output stage embodiment according to the present invention.

[0073] FIG. **34** is a block diagram of another power amplifier (PA) output stage embodiment according to the present invention.

[0074] FIG. **35** is a block diagram of another power amplifier (PA) output stage embodiment according to the present invention.

[0075] FIG. **36** is a block diagram of another power amplifier (PA) output stage embodiment according to the present invention.

[0076] FIG. **37** illustrates an example output signal according to an embodiment of the present invention.

[0077] FIG. **38** illustrates an exemplary PA embodiment.

[0078] FIG. **39** illustrates an example time-varying complex envelope PA output signal and a corresponding envelop signal.

[0079] FIG. **40** illustrates example timing diagrams of a PA output stage current.

[0080] FIG. **41** illustrates exemplary output stage current control functions.

[0081] FIG. **42** is a block diagram of another power amplifier (PA) output stage embodiment.

[0082] FIG. **43** illustrates an exemplary PA stage embodiment.

[0083] FIG. **44** illustrates an exemplary waved-shaped PA output signal.

[0084] FIG. **45** illustrates a power control method.

[0085] FIG. **46** illustrates another power control method.

[0086] FIG. **47** illustrates an exemplary vector power amplifier embodiment.

[0087] FIG. **48** is a process flowchart for implementing output stage current shaping according to an embodiment of the present invention.

[0088] FIG. **49** is a process flowchart for implementing harmonic control according to an embodiment of the present invention.

[0089] FIG. **50** is a process flowchart for power amplification according to an embodiment of the present invention.

[0090] FIGS. **51**A-I illustrate exemplary multiple-input single-output (MISO) output stage embodiments.

[0091] FIG. **52** illustrates an exemplary MISO amplifier embodiment.

[0092] FIG. **53** illustrates frequency band allocation on lower and upper spectrum bands for various communication standards.

[0093] FIGS. **54**A-B illustrate feedforward techniques for compensating for errors.

[0094] FIG. **55** illustrates a receiver-based feedback error correction technique.

[0095] FIG. **56** illustrates a digital control module embodiment.

[0096] FIG. **57** illustrates another digital control module embodiment.

[0097] FIG. **58** illustrates another digital control module embodiment.

[0098] FIG. **59** illustrates a VPA analog core embodiment.

[0099] FIG. **60** illustrates an output stage embodiment according to the VPA analog core embodiment of FIG. **60**.

[0100] FIG. **61** illustrates another VPA analog core embodiment.

[0101] FIG. **62** illustrates an output stage embodiment according to the VPA analog core embodiment of FIG. **61**.

[0102] FIG. **63** illustrates another VPA analog core embodiment.

[0103] FIG. **64** illustrates an output stage embodiment according to the VPA analog core embodiment of FIG. **63**.

[0104] FIG. **65** illustrates real-time amplifier class control using an exemplary waveform, according to an embodiment of the present invention.

[0105] FIG. **66** is an example plot of output power versus outphasing angle.

[0106] FIG. **67** illustrates exemplary power control mechanisms using an exemplary QPSK waveform, according to an embodiment of the present invention.

[0107] FIG. **68** illustrates real-time amplifier class control using an exemplary waveform, according to an embodiment of the present invention.

[0108] FIG. **69** illustrates real-time amplifier class control using an exemplary waveform, according to an embodiment of the present invention.

[0109] FIG. **70** illustrates an exemplary plot of VPA output stage theoretical efficiency versus VPA output stage current, according to an embodiment of the present invention.

[0110] FIG. **71** illustrates an exemplary VPA according to an embodiment of the present invention.

[0111] FIG. **72** is a process flowchart that illustrates a method for real-time amplifier class control in a power amplifier, according to an embodiment of the present invention.

[0112] FIG. **73** illustrates an example VPA output stage.

[0113] FIG. **74** illustrates an equivalent circuit for amplifier class S operation of the VPA output stage of FIG. **73**.

[0114] FIG. **75** illustrates an equivalent circuit for amplifier class A operation of the VPA output stage of FIG. **73**.

[0115] FIG. **76** is a plot that illustrates exemplary magnitude to phase shift transform functions for amplifier class A and class S operation of the VPA output stage of FIG. **73**.

[0116] FIG. **77** is a plot that illustrates a spectrum of magnitude to phase shift transform functions corresponding to a range of amplifier classes of operation of the VPA output stage of FIG. **73**.

[0117] FIG. **78** illustrates a mathematical derivation of the magnitude to phase shift transform in the presence of branch phase and amplitude errors.

[0118] The present invention will be described with reference to the accompanying drawings. The drawing in which an element first appears is typically indicated by the leftmost digit(s) in the corresponding reference number.

DETAILED DESCRIPTION OF THE INVENTION

Table of Contents

[0119] 1. Introduction
[0120] 1.1. Example Generation of Time-Varying Complex Envelope Input Signals
[0121] 1.2. Example Generation of Time-Varying Complex Envelope Signals from Constant Envelope Signals
[0122] 1.3. Vector Power Amplification Overview
[0123] 2. General Mathematical Overview
[0124] 2.1. Phasor Signal Representation
[0125] 2.2. Time-Varying Complex Envelope Signals
[0126] 2.3. Constant Envelope Decomposition of Time-Varying Envelope Signals
[0127] 3. Vector Power Amplification (VPA) Methods and Systems
[0128] 3.1. Cartesian 4-Branch Vector Power Amplifier
[0129] 3.2. Cartesian-Polar-Cartesian-Polar (CPCP) 2-Branch Vector Power Amplifier
[0130] 3.3. Direct Cartesian 2-Branch Vector Power Amplifier
[0131] 3.4. I and Q Data to Vector Modulator Transfer Functions
[0132] 3.4.1. Cartesian 4-Branch VPA Transfer Function
[0133] 3.4.2. CPCP 2-Branch VPA Transfer Function
[0134] 3.4.3. Direct Cartesian 2-Branch VPA Transfer Function
[0135] 3.4.4. Magnitude to Phase Shift Transform
[0136] 3.4.4.1. Magnitude to Phase Shift Transform for Sinusoidal Signals
[0137] 3.4.4.2. Magnitude to Phase Shift Transform for Square Wave Signals
[0138] 3.4.5. Waveform Distortion Compensation
[0139] 3.5. Output Stage
[0140] 3.5.1. Output Stage Embodiments
[0141] 3.5.2. Output Stage Current Shaping
[0142] 3.5.3. Output Stage Protection
[0143] 3.6. Harmonic Control
[0144] 3.7. Power Control
[0145] 3.8. Exemplary Vector Power Amplifier Embodiment
[0146] 4. Additional Exemplary Embodiments and Implementations
[0147] 4.1. Overview
[0148] 4.1.1. Control of Output Power and Power Efficiency
[0149] 4.1.2. Error Compensation and/or Correction
[0150] 4.1.3. Multi-Band Multi-Mode Operation
[0151] 4.2. Digital Control Module
[0152] 4.3. VPA Analog Core
[0153] 4.3.1. VPA Analog Core Implementation A
[0154] 4.3.2. VPA Analog Core Implementation B
[0155] 4.3.3. VPA Analog Core Implementation C
[0156] 5. Real-Time Amplifier Class Control of VPA Output Stage
[0157] 6. Summary
[0158] 7. Conclusions

1. Introduction

[0159] Methods, apparatuses and systems for vector combining power amplification are disclosed herein.

[0160] Vector combining power amplification is an approach for optimizing linearity and power efficiency simultaneously. Generally speaking, and referring to flowchart **502** in FIG. **50**, in step **504** a time-varying complex envelope input signal, with varying amplitude and phase, is decomposed into constant envelope constituent signals. In step **506**, the constant envelope constituent signals are amplified, and then in step **508** summed to generate an amplified version of the input complex envelope signal. Since substantially constant envelope signals may be amplified with minimal concern for non-linear distortion, the result of summing the constant envelope signals suffers minimal non-linear distortion while providing optimum efficiency.

[0161] Accordingly, vector, combining power amplification allows for non-linear power amplifiers to be used to efficiently amplify complex signals whilst maintaining minimal non-linear distortion levels.

[0162] For purposes of convenience, and not limitation, methods and systems of the present invention are sometimes referred to herein as vector power amplification (VPA) methods and systems.

[0163] A high-level description of VPA methods and systems according to embodiments of the present invention is now provided. For the purpose of clarity, certain terms are first defined below. The definitions described in this section are provided for convenience purposes only, and are not limiting. The meaning of these terms will be apparent to persons skilled in the art(s) based on the entirety of the teachings provided herein. These terms may be discussed throughout the specification with additional detail.

[0164] The term signal envelope, when used herein, refers to an amplitude boundary within which a signal is contained as it fluctuates in the time domain. Quadrature-modulated signals can be described by $r(t)=i(t)\cdot\cos(\omega c\cdot t)+q(t)\cdot\sin(\omega\cdot t)$ where i(t) and q(t) represent in-phase and quadrature signals with the signal envelope e(t), being equal to $e(t)=\sqrt{i(t)^2+q(t)^2}$ and the phase angle associated with r(t) is related to arctan (q(t)/i(t)).

[0165] The term constant envelope signal, when used herein, refers to in-phase and quadrature signals where $e(t)=\sqrt{i(t)^2+q(t)^2}$, with e(t) having a relatively or substantially constant value.

[0166] The term time-varying envelope signal, when used herein, refers to a signal having a time-varying signal envelope. A time-varying envelope signal can be described in terms of in-phase and quadrature signals as $e(t)=\sqrt{i(t)^2+q(t)^2}$, with e(t) having a time-varying value.

[0167] The term phase shifting, when used herein, refers to delaying or advancing the phase component of a time-varying or constant envelope signal relative to a reference phase.

1.1) Example Generation of Complex Envelope Time-Varying Input Signals

[0168] FIGS. **1**A and **1**B are examples that illustrate the generation of time-varying envelope and phase complex input signals. In FIG. **1**A, time-varying envelope carrier signals **104** and **106** are input into phase controller **110**. Phase controller **110** manipulates the phase components of signals **104** and **106**. In other words, phase controller **110** may phase shift signals **104** and **106**. Resulting signals **108** and **112**, accordingly, may be phased shifted relative to signals **104** and **106**. In the example of FIG. **1**A, phase controller **110** causes a phase reversal (180 degree phase shift) in signals **104** and **106** at time instant to, as can be seen from signals **108** and **112**. Signals **108** and **112** represent time-varying complex carrier signals. Signals **108** and **112** have both time-varying envelopes and phase components. When summed, signals **108** and **112** result in signal **114**. Signal **114** also represents a time-varying complex signal. Signal **114** may be an example input signal into VPA embodiments of the present invention (for example, an example input into step **504** of FIG. **50**).

[0169] Time-varying complex signals may also be generated as illustrated in FIG. **1**B. In FIG. **1**B, signals **116** and **118** represent baseband signals. For example, signals **116** and **118** may be in-phase (I) and quadrature (Q) baseband components of a signal. In the example of FIG. **1**B, signals **116** and **118** undergo a zero crossing as they transition from +1 to −1. Signals **116** and **118** are multiplied by signal **120** or signal **120** phase shifted by 90 degrees. Signal **116** is multiplied by a 0 degree shifted version of signal **120**. Signal **118** is multiplied by a 90 degree shifted version of signal **120**. Resulting signals **122** and **124** represent time-varying complex carrier signals. Note that signals **122** and **124** have envelopes that vary according to the time-varying amplitudes of signals **116** and **118**. Further, signals **122** and **124** both undergo phase reversals at the zero crossings of signals **116** and **118**. Signals **122** and **124** are summed to result in signal **126**. Signal **126** represents a time-varying complex signal. Signal **126** may represent an example input signal into VPA embodiments of the present invention. Additionally, signals **116** and **118** may represent example input signals into VPA embodiments of the present invention.

1.2) Example Generation of Time-Varying Complex Envelope Signals from Constant Envelope Signals

[0170] The description in this section generally relates to the operation of step **508** in FIG. **50**. FIG. **1**C illustrates three examples for the generation of time-varying complex signals from the sum of two or more substantially constant envelope signals. A person skilled in the art will appreciate, however, based on the teachings provided herein that the concepts illustrated in the examples of FIG. **1**C can be similarly extended to the case of more than two constant envelope signals.

[0171] In example **1** of FIG. **1**C, constant envelope signals **132** and **134** are input into phase controller **130**. Phase controller **130** manipulates phase components of signals **132** and **134** to generate signals **136** and **138**, respectively. Signals **136** and **138** represent substantially constant envelope signals, and are summed to generate signal **140**. The phasor representation in FIG. **1**C, associated with example 1 illustrates signals **136** and **138** as phasors P_{136} and P_{138}, respectively. Signal **140** is illustrated as phasor P_{140}. In example 1, P_{136} and P_{138} are symmetrically phase shifted by an angle ϕ_1 relative to a reference signal assumed to be aligned with the real axis of the phasor representation. Correspondingly, time domain signals **136** and **138** are phase shifted in equal amounts but opposite directions relative to the reference signal. Accordingly, P_{140}, which is the sum of P_{136} and P_{138}, is in-phase with the reference signal.

[0172] In example 2 of FIG. **1**C, substantially constant envelope signals **132** and **134** are input into phase controller **130**. Phase controller **130** manipulates phase components of signals **132** and **134** to generate signals **142** and **144**, respectively. Signals **142** and **144** are substantially constant envelope signals, and are summed to generate signal **150**. The phasor representation associated with example 2 illustrates signals **142** and **144** as phasors P_{142} and P_{144}, respectively. Signal **150** is illustrated as phasor P_{150}. In example 2, P_{142} and P_{144} are symmetrically phase shifted relative to a reference signal. Accordingly, similar to P_{140}, P_{150} is also in-phase with the reference signal. P_{142} and P_{144}, however, are phase shifted by an angle whereby $\phi_2 \neq \phi_1$ relative to the reference signal. P_{150}, as a result, has a different magnitude than P_{140} of example 1. In the time domain representation, it is noted that signals **140** and **150** are in-phase but have different amplitudes relative to each other.

[0173] In example 3 of FIG. **1**C, substantially constant envelope signals **132** and **134** are input into phase controller **130**. Phase controller **130** manipulates phase components of signals **132** and **134** to generate signals **146** and **148**, respectively. Signals **146** and **148** are substantially constant envelope signals, and are summed to generate signal **160**. The phasor representation associated with example 3 illustrates signals **146** and **148** as phasors P_{146} and P_{148}, respectively. Signal **160** is illustrated as phasor P_{160}. In example 3, P_{146} is phased shifted by an angle ϕ_3 relative to the reference signal. P_{148} is phase shifted by an angle ϕ_4 relative to the reference signal. ϕ_3 and ϕ_4 may or may not be equal. Accordingly, P_{160}, which is the sum of P_{146} and P_{148}, is no longer in-phase with the reference signal. P_{160} is phased shifted by an angle Θ relative to the reference signal. Similarly, P_{160} is phase shifted by Θ relative to P_{140} and P_{150} of examples 1 and 2. P_{160} may also vary in amplitude relative to P_{140} as illustrated in example 3.

[0174] In summary, the examples of FIG. **1**C demonstrate that a time-varying amplitude signal can be obtained by the sum of two or more substantially constant envelope signals (Example 1). Further, the time-varying signal can have amplitude changes but no phase changes imparted thereon by equally shifting in opposite directions the two or more substantially constant envelope signals (Example 2). Equally shifting in the same direction the two or more constant envelope constituents of the signal, phase changes but no amplitude changes can be imparted on the time-varying signal. Any time-varying amplitude and phase signal can be generated using two or more substantially constant envelope signals (Example 3).

[0175] It is noted that signals in the examples of FIG. **1**C are shown as sinusoidal waveforms for purpose of illustration only. A person skilled in the art will appreciate based on the teachings herein that other types of waveforms may also have been used. It should also be noted that the examples of FIG. **1**C are provided herein for the purpose of illustration only, and may or may not correspond to a particular embodiment of the present invention.

1.3) Vector Power Amplification Overview

[0176] A high-level overview of vector power amplification is now provided. FIG. **1**D illustrates the power amplification of an exemplary time-varying complex input signal **172**. Signals **114** and **126** as illustrated in FIGS. **1**A and **1**B may be examples of signal **172**. Further, signal **172** may be generated by or comprised of two or more constituent signals such as **104** and **106** (FIG. **1**A), **108** and **112** (FIG. **1**A), **116** and **118** (FIG. **1**B), and **122** and **124** (FIG. **1**B).

[0177] In the example of FIG. **1**D, VPA **170** represents a VPA system embodiment according to the present invention. VPA **170** amplifies signal **172** to generate amplified output signal **178**. Output signal **178** is amplified efficiently with minimal distortion.

[0178] In the example of FIG. **1**D, signals **172** and **178** represent voltage signals $V_{in}(t)$ and $V_{olt}(t)$, respectively. At any time instant, in the example of FIG. **1**D, $V_{in}(t)$ and $V_{olt}(t)$ are related such that $V_{olt}(t)=Kev_{in}(tat')$, where K is a scale factor and t' represents a time delay that may be present in the VPA system. For power implication,

$$\frac{V_{out}^2(t)}{Z_{out}} > \frac{V_{in}^2(t)}{Z_{in}},$$

where output signal **178** is a power amplified version of input signal **172**.

[0179] Linear (or substantially linear) power amplification of time-varying complex signals, as illustrated in FIG. **1**D, is achieved according to embodiments of the present as shown in FIG. **1**E.

[0180] FIG. **1**E is an example block diagram that conceptually illustrates a vector power amplification embodiment according to embodiments of the present invention. In FIG. **1**E, input signal **172** represents a time-varying complex signal. For example, input signal **172** may be generated as illustrated in FIGS. **1**A and **1**B. In embodiments, signal **172** may be a digital or an analog signal. Further, signal **172** may be a baseband or a carrier-based signal.

[0181] Referring to FIG. **1**E, according to embodiments of the present invention, input signal **172** or equivalents thereof are input into VPA **182**. In the embodiment of FIG. **1**E, VPA **182** includes a state machine **184** and analog circuitry **186**. State machine **184** may include digital and/or analog components. Analog circuitry **186** includes analog components. VPA **182** processes input signal **172** to generate two or more signals **188**-{**1**, . . . , n}, as illustrated in FIG. **1**E. As described with respect to signals **136**, **138**, **142**, **144**, and **146**, **148**, in FIG. **1**C, signals **188**-{**1**, . . . , n} may or may not be phase shifted relative to each other over different periods of time. Further, VPA **182** generates signals **188**-{**1**, . . . , n} such that a sum of signals **188**-{**1**, . . . , n} results in signal **194** which, in certain embodiments, can be an amplified version of signal **172**.

[0182] Still referring to FIG. **1**E, signals **188**-{**1**, . . . , n} are substantially constant envelope signals. Accordingly, the description in the prior paragraph corresponds to step **504** in FIG. **50**.

[0183] In the example of FIG. **1**E, generally corresponding to step **506** in FIG. **50**, constant envelope signals **188**-{**1**, . . . , n} are each independently amplified by a corresponding power amplifier (PA) **190**-{**1**, . . . , n} to generate amplified signals **192**-{**1**, . . . , n}. In embodiments, PAs **190**-{**1**, . . . , n} amplify substantially equally respective constant envelope signals **188**-{**1**, . . . , n}. Amplified signals **192**-{**1**, . . . , n} are substantially constant envelope signals, and in step **508** are summed to generate output signal **194**. Note that output signal **194** can be a linearly (or substantially linearly) amplified version of input signal **172**. Output signal **194** may also be a frequency-upconverted version of input signal **172**, as described herein.

2. General Mathematical Overview

2.1) Phasor Signal Representation

[0184] FIG. **1** illustrates a phasor representation $\vec{R}$ **102** of a signal r(t). A phasor representation of a signal is explicitly representative of the magnitude of the signal's envelope and of the signal's phase shift relative to a reference signal. In this document, for purposes of convenience, and not limitation, the reference signal is defined as being aligned with the real (Re) axis of the orthogonal space of the phasor representation. The invention is not, however, limited to this embodiment. The frequency information of the signal is implicit in the representation, and is given by the frequency of the reference signal. For example, referring to FIG. **1**, and assuming that the real axis corresponds to a $\cos(\omega t)$ reference signal, phasor $\vec{R}$ would translate to the function $r(t)=R(t)\cos(\omega t+\phi(t))$, where R is the magnitude of $\vec{R}$.

[0185] Still referring to FIG. **1**, it is noted that phasor $\vec{R}$ can be decomposed into a real part phasor $\vec{I}$ and an imaginary part phasor $\vec{Q}$. $\vec{I}$ and $\vec{Q}$ are said to be the in-phase and quadrature phasor components of $\vec{R}$ with respect to the reference signal. It is further noted that the signals that correspond to $\vec{I}$ and $\vec{Q}$ are related to r(t) as $I(t)=R(t)\cdot\cos(\phi(t))$ and $Q(t)=R(t)\cdot\sin(\phi(t))$, respectively. In the time domain, signal r(t) can also be written in terms of its in-phase and quadrature components as follows:

$$r(t)=I(t)\cdot\cos(\omega t)+Q(t)\cdot\sin(\omega t)=R(t)\cdot\cos(\phi(t))\cdot\cos(\omega t)+R(t)\cdot\sin(\phi(t))\cdot\sin(\omega t) \tag{1}$$

[0186] Note that, in the example of FIG. **1**, R(t) is illustrated at a particular instant of time.

2.2) Time-Varying Complex Envelope Signals

[0187] FIG. **2** illustrates a phasor representation of a signal r(t) at two different instants of time t**1** and t**2**. It is noted that the magnitude of the phasor, which represents the magnitude of the signal's envelope, as well as its relative phase shift both vary from time t**1** to time t**2**. In FIG. **2**, this is illustrated by the varying magnitude of phasors $\vec{R_1}$ and $\vec{R_2}$ and their corresponding phase shift angles ϕ_1 and ϕ_2. Signal r(t), accordingly, is a time-varying complex envelope signal.

[0188] It is further noted, from FIG. **2**, that the real and imaginary phasor components of signal r(t) are also time-varying in amplitude. Accordingly, their corresponding time domain signals also have time-varying envelopes.

[0189] FIGS. 3A-3C illustrate an example modulation to generate a time-varying complex envelope signal. FIG. 3A illustrates a view of a signal m(t). FIG. 3B illustrates a view of a portion of a carrier signal c(t). FIG. 3C illustrates a signal r(t) that results from the multiplication of signals m(t) and c(t).

[0190] In the example of FIG. 3A, signal m(t) is a time-varying magnitude signal. m(t) further undergoes a zero crossing. Carrier signal c(t), in the example of FIG. 3B, oscillates at some carrier frequency, typically higher than that of signal m(t).

[0191] From FIG. 3C, it can be noted that the resulting signal r(t) has a time-varying envelope. Further, it is noted, from FIG. 3C, that r(t) undergoes a reversal in phase at the moment when the modulating signal m(t) crosses zero. Having both non-constant envelope and phase, r(t) is said to be a time-varying complex envelope signal.

2.3) Constant Envelope Descomposition of Time-Varying Envelope Signals

[0192] Any phasor of time-varying magnitude and phase can be obtained by the sum of two or more constant magnitude phasors having appropriately specified phase shifts relative to a reference phasor.

[0193] FIG. 3D illustrates a view of an example time-varying envelope and phase signal S(t). For ease of illustration, signal S(t) is assumed to be a sinusoidal signal having a maximum envelope magnitude A. FIG. 3D further shows an example of how signal S(t) can be obtained, at any instant of time, by the sum of two constant envelope signals $S_1(t)$ and $S_2(t)$. Generally, $S_1(t)=A_1 \sin(\omega t+\phi_1(t))$ and $S_1(t)=A_2 \sin(\omega+\phi_2(t))$.

[0194] For the purpose of illustration, three views are provided in FIG. 3D that illustrate how by appropriately phasing signals $S_1(t)$ and $S_2(t)$ relative to S(t), signals $S_1(t)$ and $S_2(t)$ can be summed so that $S(t)=K(S_1(t)+S_2(t))$ where K is a constant. In other words, signal S(t) can be decomposed, at any time instant, into two or more signals. From FIG. 3D, over period T_1, $S_1(t)$ and $S_2(t)$ are both in-phase relative to signal S(t), and thus sum to the maximum envelope magnitude A of signal S(t). Over period T_3, however, signals $S_1(t)$ and $S_2(t)$ are 180 degree out-of-phase relative to each other, and thus sum to a minimum envelope magnitude of signal S(t).

[0195] The example of FIG. 3D illustrates the case of sinusoidal signals. A person skilled in the art, however, will understand that any time-varying envelope, which modulates a carrier signal that can be represented by a Fourier series or Fourier transform, can be similarly decomposed into two or more substantially constant envelope signals. Thus, by controlling the phase of a plurality of substantially constant envelope signals, any time-varying complex envelope signal can be generated.

3. Vector Power Amplification Methods and Systems

[0196] Vector power amplification methods and systems according to embodiments of the present invention rely on the ability to decompose any time-varying envelope signal into two or more substantially constant envelope constituent signals or to receive or generate such constituent signals, amplify the constituent signals, and then sum the amplified signals to generate an amplified version of the time-varying complex envelope signal.

[0197] In sections 3.1-3.3, vector power amplification (VPA) embodiments of the present invention are provided, including 4-branch and 2-branch embodiments. In the description, each VPA embodiment is first presented conceptually using a mathematical derivation of underlying concepts of the embodiment. An embodiment of a method of operation of the VPA embodiment is then presented, followed by various system level embodiments of the VPA embodiment.

[0198] Section 3.4 presents various embodiments of control modules according to embodiments of the present invention. Control modules according to embodiments of the present invention may be used to enable certain VPA embodiments of the present invention. In some embodiments, the control modules are intermediary between an input stage of the VPA embodiment and a subsequent vector modulation stage of the VPA embodiment.

[0199] Section 3.5 describes VPA output stage embodiments according to embodiments of the present invention. Output stage embodiments are directed to generating the output signal of a VPA embodiment.

[0200] Section 3.6 is directed to harmonic control according to embodiments of the present invention. Harmonic control may be implemented in certain embodiments of the present invention to manipulate the real and imaginary power in the harmonics of the VPA embodiment, thus increasing the power present in the fundamental frequency at the output.

[0201] Section 3.7 is directed to power control according to embodiments of the present invention. Power control may be implemented in certain embodiments of the present invention in order to satisfy power level requirements of applications where VPA embodiments of the present invention may be employed.

3.1) Cartesian 4-Branch Vector Power Amplifier

[0202] According to one embodiment of the invention, herein called the Cartesian 4-Branch VPA embodiment for ease of illustration and not limitation, a time-varying complex envelope signal is decomposed into 4 substantially constant envelope constituent signals. The constituent signals are equally or substantially equally amplified individually, and then summed to construct an amplified version of the original time-varying complex envelope signal.

[0203] It is noted that 4 branches are employed in this embodiment for purposes of illustration, and not limitation. The scope of the invention covers use of other numbers of branches, and implementation of such variations will be apparent to persons skilled in the art based on the teachings contained herein.

[0204] In one embodiment, a time-varying complex envelope signal is first decomposed into its in-phase and quadrature vector components. In phasor representation, the in-phase and quadrature vector components correspond to the signal's real part and imaginary part phasors, respectively.

[0205] As described above, magnitudes of the in-phase and quadrature vector components of a signal vary proportionally to the signal's magnitude, and are thus not constant envelope when the signal is a time-varying envelope signal. Accordingly, the 4-Branch VPA embodiment further decomposes each of the in-phase and quadrature vector components of the signal into four substantially constant envelope components, two for the in-phase and two for the quadrature signal components. This concept is illustrated in FIG. **4** using a phasor signal representation.

[0206] In the example of FIG. **4**, phasors $\vec{I_1}$ and $\vec{I_2}$ correspond to the real part phasors of an exemplary time-varying complex envelope signal at two instants of time t**1** and t**2**, respectively. It is noted that phasors $\vec{I_1}$ and $\vec{I_2}$ have different magnitudes.

[0207] Still referring to FIG. **4**, at instant t**1**, phasor $\vec{I_1}$ can be obtained by the sum of upper and lower phasors $\vec{I_{U_1}}$ and $\vec{I_{L_1}}$. Similarly, at instant t**2**, phasor $\vec{I_2}$ can be obtained by the sum of upper and lower phasors $\vec{I_{U_2}}$ and $\vec{I_{L_2}}$. Note that phasors $\vec{I_{U_1}}$ and $\vec{I_{U_2}}$ have equal or substantially equal magnitude. Similarly, phasors $\vec{I_{L_1}}$ and $\vec{I_{L_2}}$ have substantially equal magnitude. Accordingly, the real part phasor of the time-varying envelope signal can be obtained at any time instant by the sum of at least two substantially constant envelope components.

[0208] The phase shifts of phasors $\vec{I_{U_1}}$ and $\vec{I_{L_1}}$ relative to $\vec{I_1}$, as well as the phase shifts of phasors $\vec{I_{U_2}}$ and $\vec{I_{L_2}}$ relative to $\vec{I_2}$ are set according to the desired magnitude of phasors $\vec{I_1}$ and $\vec{I_2}$, respectively. In one case, when the upper and lower phasors are selected to have equal magnitude, the upper and lower phasors are symmetrically shifted in phase relative to the phasor. This is illustrated in the example of FIG. **4**, and corresponds to $\vec{I_{U_1}}$, $\vec{I_{L_1}}$, $\vec{I_{U_2}}$, and $\vec{I_{L_2}}$ all having equal magnitude. In a second case, the phase shift of the upper and lower phasors are substantially symmetrically shifted in phase relative to the phasor. Based on the description herein, anyone skilled in the art will understand that the magnitude and phase shift of the upper and lower phasors do not have to be exactly equal in value

[0209] As an example, it can be further verified that, for the case illustrated in FIG. **4**, the relative phase shifts, illustrated as

$$\frac{\phi_1}{2} \text{ and } \frac{\phi_2}{2}$$

in FIG. **4**, are related to the magnitudes of normalized phasors $\vec{I_1}$ and $\vec{I_2}$ as follows:

$$\frac{\phi_1}{2} = \cot^{-1}\left(\frac{I_1}{2\sqrt{1-\frac{I_1^2}{4}}}\right); \text{ and} \tag{2}$$

$$\frac{\phi_2}{2} = \cot^{-1}\left(\frac{I_2}{2\sqrt{1-\frac{I_2^2}{4}}}\right), \tag{3}$$

[0210] wherein I_1 and I_2 represent the normalized magnitudes of phasors $\vec{I_1}$ and $\vec{I_2}$, respectively, and wherein the domains of I_1 and I_2 are restricted appropriately according to the domain over which equation (2) and (3) are valid. It is noted that equations (2) and (3) are one representation for relating the relative phase shifts to the normalized magnitudes. Other, solutions, equivalent representations, and/or simplified representations of equations (2) and (3) may also be employed. Look up tables relating relative phase shifts to normalized magnitudes may also be used.

[0211] The concept describe above can be similarly applied to the imaginary phasor or the quadrature component part of a signal r(t) as illustrated in FIG. **4**. Accordingly, at any time instant t, imaginary phasor part $\vec{Q}$ of signal r(t) can be obtained by summing upper and lower phasor components $\vec{Q_U}$ and $\vec{Q_L}$ of substantially equal and constant magnitude. In this example, $\vec{Q_U}$ and $\vec{Q_L}$ are symmetrically shifted in phase relative to $\vec{Q}$ by an angle set according to the magnitude of $\vec{Q}$ at time t. The relationship of $\vec{Q_U}$ and $\vec{Q_L}$ to the desired phasor $\vec{Q}$ are related as defined in equations 2 and 3 by substituting Q_1 and Q_2 for I_1 and I_2 respectively.

[0212] It follows from the above discussion that, in phasor representation, any phasor $\vec{R}$ of variable magnitude and phase can be constructed by the sum of four substantially constant magnitude phasor components:

$$\vec{R}=\vec{I_U}+\vec{I_L}+\vec{Q_U}+\vec{Q_L};$$

$$\vec{I_U}+\vec{I_L}=\vec{I};$$

$$\vec{Q_U}+\vec{Q_L}=\vec{Q};$$

$$I_U=I_L=\text{constant};$$

$$Q_U=Q_L=\text{constant}; \tag{4}$$

where I_U, I_L, Q_U, and Q_L represent the magnitudes of phasors $\vec{I_U}$, $\vec{I_L}$, $\vec{Q_U}$, and $\vec{Q_L}$, respectively.

[0213] Correspondingly, in the time domain, a time-varying complex envelope sinusoidal signal $r(t)=R(t)\cos(\omega t+\phi)$ is constructed by the sum of four constant envelope signals as follows:

$$r(t) = I_U(t) + I_L(t) + Q_U(t) + Q_L(t); \tag{5}$$

$$I_U(t) = sgn(\vec{I}) \times I_U\cos\left(\frac{\phi_I}{2}\right)\times\cos(\omega t) + I_U \times \sin\left(\frac{\phi_I}{2}\right)\times\sin(\omega t);$$

$$I_L(t) = sgn(\vec{I}) \times I_L\cos\left(\frac{\phi_I}{2}\right)\times\cos(\omega t) - I_L \times \sin\left(\frac{\phi_I}{2}\right)\times\sin(\omega t);$$

$$Q_U(t) = sgn(\vec{Q}) \times Q_U\cos\left(\frac{\phi_Q}{2}\right)\times\sin(\omega t) + Q_U \times \sin\left(\frac{\phi_Q}{2}\right)\times\cos(\omega t);$$

$$Q_L(t) = sgn(\vec{Q}) \times Q_L\cos\left(\frac{\phi_Q}{2}\right)\times\sin(\omega t) - Q_L \times \sin\left(\frac{\phi_Q}{2}\right)\times\cos(\omega t).$$

where $sgn(\vec{I})=\pm 1$ depending on whether $\vec{I}$ is in-phase or 180° degrees out-of-phase with the positive real axis. Similarly, $sgn(\vec{Q})=\pm 1$ depending on whether $\vec{Q}$ is in-phase or 180° degrees out-of-phase with the imaginary axis.

$$\frac{\phi_I}{2}$$

corresponds to the phase shift of $\vec{I_U}$ and $\vec{I_L}$ relative to the real axis. Similarly,

$$\frac{\phi_Q}{2}$$

corresponds to the phase shift of $\vec{Q}_U$ and $\vec{Q}_L$ relative to the imaginary axis.

$$\frac{\phi_I}{2} \text{ and } \frac{\phi_Q}{2}$$

can be calculated using the equations given in (2) and (3).

[0214] Equations (5) can be further simplified as:

$$r(t) = I_U(t) + I_L(t) + Q_U(t) + Q_L(t); \tag{6}$$

$$I_U(t) = sgn(\vec{I}) \times I_{UX} \times \cos(\omega t) + I_{UY} \times \sin(\omega t);$$

$$I_L(t) = sgn(\vec{I}) \times I_{UX} \times \cos(\omega t) - I_{UY} \times \sin(\omega t);$$

$$Q_U(t) = -Q_{UX} \times \cos(\omega t) + sgn(\vec{Q}) \times Q_{UY} \times \sin(\omega t);$$

$$Q_L(t) = Q_{UY} \times \cos(\omega t) - sgn(\vec{Q}) \times Q_{UY} \times \sin(\omega t).$$

$$I_{UX} = I_U \times \cos\left(\frac{\phi_I}{2}\right) = I_L \times \cos\left(\frac{\phi_I}{2}\right), \text{ where}$$

$$I_{UY} = I_U \times \sin\left(\frac{\phi_I}{2}\right) = I_L \times \sin\left(\frac{\phi_I}{2}\right),$$

$$Q_{UX} = Q_U \times \sin\left(\frac{\phi_Q}{2}\right) = Q_L \times \sin\left(\frac{\phi_Q}{2}\right), \text{ and}$$

$$Q_{UY} = Q_U \times \cos\left(\frac{\phi_Q}{2}\right) = Q_L \times \cos\left(\frac{\phi_Q}{2}\right).$$

[0215] It can be understood by a person skilled in the art that, whereas the time domain representations in equations (5) and (6) have been provided for the case of a sinusoidal waveform, equivalent representations can be developed for non-sinusoidal waveforms using appropriate basis functions. Further, as understood by a person skilled in the art based on the teachings herein, the above-describe two-dimensional decomposition into substantially constant envelope signals can be extended appropriately into a multi-dimensional decomposition.

[0216] FIG. **5** is an example block diagram of the Cartesian 4-Branch VPA embodiment. An output signal r(t) **578** of desired power level and frequency characteristics is generated from baseband in-phase and quadrature components according to the Cartesian 4-Branch VPA embodiment.

[0217] In the example of FIG. **5**, a frequency generator such as a synthesizer **510** generates a reference signal A*cos(ωt) **511** having the same frequency as that of output signal r(t) **578**. It can be understood by a person skilled in the art that the choice of the reference signal is made according to the desired output signal. For example, if the desired frequency of the desired output signal is 2.4 GHz, then the frequency of the reference signal is set to be 2.4 GHz. In this manner, embodiments of the invention achieve frequency up-conversion.

[0218] Referring to FIG. **5**, one or more phase splitters are used to generate signals **521**, **531**, **541**, and **551** based on the reference signal **511**. In the example of FIG. **5**, this is done using phase splitters **512**, **514**, and **516** and by applying 0° phase shifts at each of the phase splitters. A person skilled in the art will appreciate, however, that various techniques may be used for generating signals **521**, **531**, **541**, and **551** of the reference signal **511**. For example, a 1:4 phase splitter may be used to generate the four replicas **521**, **531**, **541**, and **551** in a single step or in the example embodiment of FIG. **5**, signal **511** can be directly coupled to signals **521**, **531**, **541**, **551** Depending on the embodiment, a variety of phase shifts may also be applied to result in the desired signals **521**, **531**, **541**, and **551**.

[0219] Still referring to FIG. **5**, the signals **521**, **531**, **541**, and **551** are each provided to a corresponding vector modulator **520**, **530**, **540**, and **550**, respectively. Vector modulators **520**, **530**, **540**, and **550**, in conjunction with their appropriate input signals, generate four constant envelope constituents of signal r(t) according to the equations provided in (6). In the example embodiment of FIG. **5**, vector modulators **520** and **530** generate the I_U(t) and I_L(t) components, respectively, of signal r(t). Similarly, vector modulators **540** and **550** generate the Q_U(t) and Q_L(t) components, respectively, of signal r(t).

[0220] The actual implementation of each of vector modulators **520**, **530**, **540**, and **550** may vary. It will be understood by a person skilled in the art, for example, that various techniques exist for generating the constant envelope constituents according to the equations in (6).

[0221] In the example embodiment of FIG. **5**, each of vector modulators **520**, **530**, **540**, **550** includes an input phase splitter **522**, **532**, **542**, **552** for phasing the signals **522**, **531**, **541**, **551**. Accordingly, input phase splitters **522**, **532**, **542**, **552** are used to generate an in-phase and a quadrature components or their respective input signals.

[0222] In each vector modulator **520**, **530**, **540**, **550**, the in-phase and quadrature components are multiplied with amplitude information. In FIG. **5**, for example, multiplier **524** multiplies the quadrature component of signal **521** with the quadrature amplitude information I_{UY} of I_U(t). In parallel, multiplier **526** multiplies the in-phase replica signal with the in-phase amplitude information sgn(I)×I_{UX} of I_U(t).

[0223] To generate the I_U(t) constant envelope constituent signals **525** and **527** are summed using phase splitter **528** or alternate summing techniques. The resulting signal **529** corresponds to the IU(t) component of signal r(t).

[0224] In similar fashion as described above, vector modulators **530**, **540**, and **550**, respectively, generate the I_L(t), Q_U(t), and Q_L(t) components of signal r(t). I_L(t), Q_U(t), and Q_L(t), respectively, correspond to signals **539**, **549**, and **559** in FIG. **5**.

[0225] Further, as described above, signals **529**, **539**, **549**, and **559** are characterized by having substantially equal and constant magnitude envelopes. Accordingly, when signals **529**, **539**, **549**, and **559** are input into corresponding power amplifiers (PA) **562**, **564**, **566**, and **568**, corresponding amplified signals **563**, **565**, **567**, and **569** are substantially constant envelope signals.

[0226] Power amplifiers **562**, **564**, **566**, and **568** amplify each of the signals **529**, **539**, **549**, **559**, respectively. In an embodiment, substantially equal power amplification is applied to each of the signals **529**, **539**, **549**, and **559**. In an embodiment, the power amplification level of PAs **562**, **564**, **566**, and **568** is set according to the desired power level of output signal r(t).

[0227] Still referring to FIG. **5**, amplified signals **563** and **565** are summed using summer **572** to generate an amplified version **573** of the in-phase component $\vec{I}$ (t) of signal r(t).

Similarly, amplified signals **567** and **569** are summed using summer **574** to generate an amplified version **575** of the quadrature component $\vec{Q}$ (t) of signal r(t).

[0228] Signals **573** and **575** are summed using summer **576**, as shown in FIG. **5**, with the resulting signal corresponding to desired output signal r(t).

[0229] It must be noted that, in the example of FIG. **5**, summers **572**, **574**, and **576** are being used for the purpose of illustration only. Various techniques may be used to sum amplified signals **563**, **565**, **567**, and **569**. For example, amplified signals **563**, **565**, **567**, and **569** may be summed all in one step to result in signal **578**. In fact, according to various VPA embodiments of the present invention, it suffices that the summing is done after amplification. Certain VPA embodiments of the present invention, as will be further described below, use minimally lossy summing techniques such as direct coupling via wire. Alternatively, certain VPA embodiments use conventional power combining techniques. In other embodiments, as will be further described below, power amplifiers **562**, **564**, **566**, and **568** can be implemented as a multiple-input single-output power amplifier.

[0230] Operation of the Cartesian 4-Branch VPA embodiment shall now be further described with reference to the process flowchart of FIG. **6**. The process begins at step **610**, which includes receiving the baseband representation of the desired output signal. In an embodiment, this involves receiving in-phase (I) and quadrature (Q) components of the desired output signal. In another embodiment, this involves receiving magnitude and phase of the desired output signal. In an embodiment of the Cartesian 4-Branch VPA embodiment, the I and Q are baseband components. In another embodiment, the I and Q are RF components and are down-converted to baseband.

[0231] Step **620** includes receiving a clock signal set according to a desired output signal frequency of the desired output signal. In the example of FIG. **5**, step **620** is achieved by receiving reference signal **511**.

[0232] Step **630** includes processing the I component to generate first and second signals having the output signal frequency. The first and second signals have substantially constant and equal magnitude envelopes and a sum equal to the I component. The first and second signals correspond to the $I_U(t)$ and $I_L(t)$ constant envelope constituents described above. In the example of FIG. **5**, step **630** is achieved by vector modulators **520** and **530**, in conjunction with their appropriate input signals.

[0233] Step **640** includes processing the Q component to generate third and fourth signals having the output signal frequency. The third and fourth signals have substantially constant and equal magnitude envelopes and a sum equal to the Q component. The third and fourth signals correspond to the $Q_U(t)$ and $Q_L(t)$ constant envelope constituents described above. In the example of FIG. **5**, step **630** is achieved by vector modulators **540** and **550**, in conjunction with their appropriate input signals.

[0234] Step **650** includes individually amplifying each of the first, second, third, and fourth signals, and summing the amplified signals to generate the desired output signal. In an embodiment, the amplification of the first, second, third, and fourth signals is substantially equal and according to a desired power level of the desired output signal. In the example of FIG. **5**, step **650** is achieved by power amplifiers **562**, **564**, **566**, and **568** amplifying respective signals **529**, **539**, **549**, and **559**, and by summers **572**, **574**, and **576** summing amplified signals **563**, **565**, **567**, and **569** to generate output signal **578**.

[0235] FIG. **7**A is a block diagram that illustrates an exemplary embodiment of a vector power amplifier **700** implementing the process flowchart **600** of FIG. **6**. In the example of FIG. **7**A, optional components are illustrated with dashed lines. In other embodiments, additional components may be optional.

[0236] Vector power amplifier **700** includes an in-phase (I) branch **703** and a quadrature (Q) branch **705**. Each of the I and Q branches further comprises a first branch and a second branch.

[0237] In-phase (I) information signal **702** is received by an I Data Transfer Function module **710**. In an embodiment, I information signal **702** includes a digital baseband signal. In an embodiment, I Data Transfer Function module **710** samples I information signal **702** according to a sample clock **706**. In another embodiment, I information signal **702** includes an analog baseband signal, which is converted to digital using an analog-to-digital converter (ADC) (not shown in FIG. **7**A) before being input into I Data Transfer Function module **710**. In another embodiment, I information signal **702** includes an analog baseband signal which input in analog form into I Data Transfer Function module **710**, which also includes analog circuitry. In another embodiment, I information signal **702** includes a RF signal which is down-converted to baseband before being input into I Data Transfer Function module **710** using any of the above described embodiments.

[0238] I Data Transfer Function module **710** processes I information signal **702**, and determines in-phase and quadrature amplitude information of at least two constant envelope constituent signals of I information signal **702**. As described above with reference to FIG. **5**, the in-phase and quadrature vector modulator input amplitude information corresponds to $\mathrm{sgn}(I)\times I_{UX}$ and I_{UY}, respectively. The operation of I Data Transfer Function module **710** is further described below in section 3.4.

[0239] I Data Transfer Function module **710** outputs information signals **722** and **724** used to control the in-phase and quadrature amplitude components of vector modulators **760** and **762**. In an embodiment, signals **722** and **724** are digital signals. Accordingly, each of signals **722** and **724** is fed into a corresponding digital-to-analog converter (DAC) **730** and **732**, respectively. The resolution and sample rate of DACs **730** and **732** is selected to achieve the desired I component of the output signal **782**. DACs **730** and **732** are controlled by DAC clock signals **723** and **725**, respectively. DAC clock signals **723** and **725** may be derived from a same clock signal or may be independent.

[0240] In another embodiment, signals **722** and **724** are analog signals, and DACs **730** and **732** are not required.

[0241] In the exemplary embodiment of FIG. **7**A, DACs **730** and **732** convert digital information signals **722** and **724** into corresponding analog signals, and input these analog signals into optional interpolation filters **731** and **733**, respectively. Interpolation filters **731** and **733**, which also serve as anti-aliasing filters, shape the DACs outputs to produce the desired output waveform. Interpolation filters **731** and **733** generate signals **740** and **742**, respectively. Signal **741** represents the inverse of signal **740**. Signals **740**-**742** are input into vector modulators **760** and **762**.

[0242] Vector modulators **760** and **762** multiply signals **740-742** with appropriately phased clock signals to generate constant envelope constituents of I information signal **702**. The clock signals are derived from a channel clock signal **708** having a rate according to a desired output signal frequency. A plurality of phase splitters, such as **750** and **752**, for example, and phasors associated with the vector modulator multipliers may be used to generate the appropriately phased clock signals.

[0243] In the embodiment of FIG. **7**A, for example, vector modulator **760** modulates a 90° shifted channel clock signal with quadrature amplitude information signal **740**. In parallel, vector modulator **760** modulates an in-phase channel clock signal with in-phase amplitude information signal **742**. Vector modulator **760** combines the two modulated signals to generate a first modulated constant envelope constituent **761** of I information signal **702**. Similarly, vector modulator **762** generates a second modulated constant envelope constituent **763** of I information signal **702**, using signals **741** and **742**. Signals **761** and **763** correspond, respectively, to the $I_U(t)$ and $I_L(t)$ constant envelope components described with reference to FIG. **5**.

[0244] In parallel and in similar fashion, the Q branch of vector power amplifier **700** generates at least two constant envelope constituent signals of quadrature (Q) information signal **704**.

[0245] In the embodiment of FIG. **7**A, for example, vector modulator **764** generates a first constant envelope constituent **765** of Q information signal **704**, using signals **744** and **746**. Similarly, vector modulator **766** generates a second constant envelope constituent **767** of Q information signal **704**, using signals **745** and **746**.

[0246] As described above with respect to FIG. **5**, constituent signals **761**, **763**, **765**, and **767** have substantially equal and constant magnitude envelopes. In the exemplary embodiment of FIG. **7**A, signals **761**, **763**, **765**, and **767** are, respectively, input into corresponding power amplifiers (PAs) **770**, **772**, **774**, and **776**. PAs **770**, **772**, **774**, and **776** can be linear or non-linear power amplifiers. In an embodiment, PAs **770**, **772**, **774**, and **776** include switching power amplifiers.

[0247] Circuitry **714** and **716** (herein referred to as "autobias circuitry" for ease of reference, and not limitation) and in this embodiment, control the bias of PAs **770**, **772**, **774**, and **776** according to I and Q information signals **702** and **704**. In the embodiment of FIG. **7**A, autobias circuitry **714** and **716** provide, respectively, bias signals **715** and **717** to PAs **770**, **772** and PAs **774**, **776**. Autobias circuitry **714** and **716** are further described below in section 3.5. Embodiments of PAs **770**, **772**, **774**, and **776** are also discussed below in section 3.5.

[0248] In an embodiment, PAs **770**, **772**, **774**, and **776** apply substantially equal power amplification to respective substantially constant envelope signals **761**, **763**, **765**, and **767**. In other embodiments, PA drivers are additionally employed to provide additional power amplification. In the embodiment of FIG. **7**A, PA drivers **794**, **795**, **796**, and **797** are optionally added between respective vector modulators **760**, **762**, **764** **766** and respective PAs **770**, **772**, **774**, and **776**, in each branch of vector power amplifier **700**.

[0249] The outputs of PAs **770**, **772**, **774**, and **776** are coupled together to generate output signal **782** of vector power amplifier **700**. In an embodiment, the outputs of PAs **770**, **772**, **774**, and **776** are directly coupled together using a wire. Direct coupling in this manner means that there is minimal or no resistive, inductive, or capacitive isolation between the outputs of PAs **770**, **772**, **774**, and **776**. In other words, outputs of PAs **770**, **772**, **774**, and **776**, are coupled together without intervening components. Alternatively, in an embodiment, the outputs of PAs **770**, **772**, **774**, and **776** are coupled together indirectly through inductances and/or capacitances that result in low or minimal impedance connections, and/or connections that result in minimal isolation and minimal power loss. Alternatively, outputs of PAs **770**, **772**, **774**, and **776** are coupled using well known combining techniques, such as Wilkinson, hybrid, transformers, or known active combiners. In an embodiment, the PAs **770**, **772**, **774**, and **776** provide integrated amplification and power combining in a single operation. In an embodiment, one or more of the power amplifiers and/or drivers described herein are implemented using multiple input, single output power amplification techniques, examples of which are shown in FIGS. **7**B, and **51**A-H.

[0250] Output signal **782** includes the I and Q characteristics of I and Q information signals **702** and **704**. Further, output signal **782** is of the same frequency as that of its constituents, and thus is of the desired up-converted output frequency. In embodiments of vector power amplifier **700**, a pull-up impedance **780** is coupled between the output of vector amplifier **700** and a power supply. Output stage embodiments according to power amplification methods and systems of the present invention will be further described below in section 3.5.

[0251] In other embodiments of vector power amplifier **700**, process detectors are employed to compensate for any process variations in circuitry of the amplifier. In the embodiment of FIG. **7**A for example, process detectors **791-793** are optionally added to monitor variations in PA drivers **794-797** and phase splitter **750**. In further embodiments, frequency compensation circuitry **799** may be employed to compensate for frequency variations.

[0252] FIG. **7**B is a block diagram that illustrates another exemplary embodiment of vector power amplifier **700**. Optional components are illustrated with dashed lines, although other embodiments may have more or less optional components.

[0253] The embodiment illustrates a multiple-input single-output (MISO) implementation of the amplifier of FIG. **7**A. In the embodiment of FIG. **7**B, constant envelope signals **761**, **763**, **765** and **767**, output from vector modulators **760**, **762**, **764**, and **766**, are input into MISO PAs **784** and **786**. MISO PAs **784** and **786** are two-input single-output power amplifiers. In an embodiment, MISO PAs **784** and **786** include elements **770**, **772**, **774**, **776**, **794-797** as shown in the embodiment of FIG. **7**A or functional equivalence thereof. In another embodiment, MISO PAs **784** and **786** may include other elements, such as optional pre-drivers and optional process detection circuitry. Further, MISO PAs **784** and **786** are not limited to being two-input PAs as shown in FIG. **7**B. In other embodiments as will be described further below with reference to FIGS. **51**A-H, PAs **784** and **786** can have any number of inputs and outputs.

[0254] FIG. **8**A is a block diagram that illustrates another exemplary embodiment **800**A of a vector power amplifier according to the Cartesian 4-Branch VPA method shown in FIG. **6**. Optional components are illustrated with dashed lines, although other embodiments may have more or less optional components.

[0255] In the embodiment of FIG. **8**A, a DAC **830** of sufficient resolution and sample rate replaces DACs **730**, **732**, **734**, and **736** of the embodiment of FIG. **7**A. DAC **830**'s sample rate is controlled by a DAC clock signal **826**.

[0256] DAC **830** receives in-phase and quadrature information signals **810** and **820** from I Data Transfer Function module **710** and Q Data Transfer Function module **712**, respectively, as described above. In an embodiment, a input selector **822** selects the order of signals **810** and **820** being input into DAC **830**.

[0257] DAC **830** may output a single analog signal at a time. In an embodiment, a sample and hold architecture may be used to ensure proper signal timing to the four branches of the amplifier, as shown in FIG. **8**A.

[0258] DAC **830** sequentially outputs analog signals **832**, **834**, **836**, **838** to a first set of sample-and-hold circuits **842**, **844**, **846**, and **848**. In an embodiment, DAC **830** is clocked at a sufficient rate to emulate the operation of DACs **730**, **732**, **734**, and **736** of the embodiment of FIG. **7**A. An output selector **824** determines which of output signals **832**, **834**, **836**, and **838** should be selected for output.

[0259] DAC **830**'s DAC clock signal **826**, output selector signal **824**, input selector **822**, and sample-and-hold clocks **840**A-D, and **850** are controlled by a control module that can be independent or integrated into transfer function modules **710** and/or **712**.

[0260] In an embodiment, sample-and-hold circuits (S/H) **842**, **844**, **846**, and **848** sample and hold the received analog values from DAC **830** according to a clock signals **840**A-D. Sample-and-hold circuits **852**, **854**, **856**, and **858** sample and hold the analog values from sample and hold circuits **842**, **844**, **846**, and **848** respectively. In turn, sample-and-hold circuits **852**, **854**, **856**, and **858** hold the received analog values, and simultaneously release the values to vector modulators **760**, **762**, **764**, and **766** according to a common clock signal **850**. In another embodiment, sample-and-hold circuits **852**, **854**, **856**, and **858** release the values to optional interpolation filters **731**, **733**, **735**, and **737** which are also anti-aliasing filters. In an embodiment, a common clock signal **850** is used in order to ensure that the outputs of S/H **852**, **854**, **856**, and **858** are time-aligned.

[0261] Other aspects of vector power amplifier **800**A substantially correspond to those described above with respect to vector power amplifier **700**.

[0262] FIG. **8**B is a block diagram that illustrates another exemplary embodiment **800**B of a vector power amplifier according to the Cartesian 4-Branch VPA method shown in FIG. **6**. Optional components are illustrated with dashed lines, although other embodiments may have more or less optional components.

[0263] Embodiment **800**B illustrates another single DAC implementation of the vector power amplifier. However, in contrast to the embodiment of FIG. **8**A, the sample and hold architecture includes a single set of sample-and-hold (S/H) circuits. As shown in FIG. **8**B, S/H **842**, **844**, **846**, and **848** receive analog values from DAC **830**, illustrated as signals **832**, **834**, **836**, and **838**. Each of S/H circuits **842**, **844**, **846** and **848** release its received value according to a different clock **840**A-D as shown. The time difference between analog samples used for to generate signals **740**, **741**, **742**, **744**, **745**, and **746** can be compensated for in transfer functions **710** and **712**. According to the embodiment of FIG. **8**B, one level of S/H circuitry can be eliminated relative to the embodiment of FIG. **8**A, thereby reducing the size and the complexity of the amplifier.

[0264] Other aspects of vector power amplifier **800**B substantially correspond to those described above with respect to vector power amplifiers **700** and **800**A.

[0265] FIG. **8**C is a block diagram that illustrates another exemplary embodiment **800**C of vector power amplifier **700**. Optional components are illustrated with dashed lines, although other embodiments may have more or less optional components. The embodiment of FIG. **8**C illustrates a multiple-input single-output (MISO) implementation of the amplifier of FIG. **8**A. In the embodiment of FIG. **8**C, constant envelope signals **761**, **763**, **765** and **767**, output from vector modulators **760**, **762**, **764**, and **766**, are input into MISO PAs **860** and **862**. MISO PAs **860** and **862** are two-input single-output power amplifiers. In an embodiment, MISO PAs **860** and **862** include elements **770**, **772**, **774**, **776**, **794-797** as shown in the embodiment of FIG. **7**A or functional equivalence thereof. In another embodiment, MISO PAs **860** and **862** may include other elements, such as optional pre-drivers and optional process detection circuitry. In another embodiment, MISO PAs **860** and **862** may include other elements, such as pre-drivers, not shown in the embodiment of FIG. **7**A. Further, MISO PAs **860** and **862** are not limited to being two-input PAs as shown in FIG. **8**C. In other embodiments as will be described further below with reference to FIGS. **51**A-H, PAs **860** and **862** can have any number of inputs and outputs.

[0266] Other aspects of vector power amplifier **800**C substantially correspond to those described above with respect to vector power amplifiers **700** and **800**A.

[0267] FIG. **8**D is a block diagram that illustrates another exemplary embodiment **800**D of vector power amplifier **700**. Optional components are illustrated with dashed lines, although other embodiments may have more or less optional components. The embodiment of FIG. **8**D illustrates a multiple-input single-output (MISO) implementation of the amplifier of FIG. **8**B. In the embodiment of FIG. **8**D, constant envelope signals **761**, **763**, **765** and **767**, output from vector modulators **760**, **762**, **764**, and **766**, are input into MISO PAs **870** and **872**. MISO PAs **870** and **872** are two-input single-output power amplifiers. In an embodiment, MISO PAs **870** and **872** include elements **770**, **772**, **774**, **776**, **794-797** as shown in the embodiment of FIG. **7**A or functional equivalence thereof. In another embodiment, MISO PAs **870** and **872** may include other elements, such as optional pre-drivers and optional process detection circuitry. In another embodiment, MISO PAs **870** and **872** may include other elements, such as pre-drivers, not shown in the embodiment of FIG. **7**A. Further, MISO PAs **870** and **872** are not limited to being two-input PAs as shown in FIG. **8**D. In other embodiments as will be described further below with reference to FIGS. **51**A-H, PAs **870** and **872** can have any number of inputs and outputs.

[0268] Other aspects of vector power amplifier **800**D substantially correspond to those described above with respect to vector power amplifiers **700** and **800**B.

3.2) Cartesian-Polar Cartesian-Polar 2-Branch Vector Power Amplifier

[0269] A Cartesian-Polar-Cartesian-Polar (CPCP) 2-Branch VPA embodiment shall now be described (The name of this embodiment is provided for ease of reference, and is not limiting).

[0270] According to the Cartesian-Polar-Cartesian-Polar (CPCP) 2-Branch VPA method, a time-varying complex envelope signal is decomposed into 2 substantially constant envelope constituent signals. The constituent signals are individually amplified, and then summed to construct an amplified version of the original time-varying complex envelope signal. In addition, the phase angle of the time-varying complex envelope signal is determined and the resulting summation of the constituent signals are phase shifted by the appropriate angle.

[0271] In one embodiment of the CPCP 2-Branch VPA method, a magnitude and a phase angle of a time-varying complex envelope signal are calculated from in-phase and quadrature components of a signal. Given the magnitude information, two substantially constant envelope constituents are calculated from a normalized version of the desired time-varying envelope signal, wherein the normalization includes implementation specific manipulation of phase and/or amplitude. The two substantially constant envelope constituents are then phase shifted by an appropriate angle related to the phase shift of the desired time-varying envelope signal. The substantially constant envelope constituents are then individually amplified substantially equally, and summed to generate an amplified version of the original desired time-varying envelope signal.

[0272] FIGS. **9**A and **9**B conceptually illustrate the CPCP 2-Branch VPA embodiment using a phasor signal representation. In FIG. **9**A, phasor $\vec{R}_{in}$ represents a time-varying complex envelope input signal r(t). At any instant of time, $\vec{R}_{in}$ reflects a magnitude and a phase shift angle of signal r(t). In the example shown in FIG. **9**A, $\vec{R}_{in}$ is characterized by a magnitude R and a phase shift angle θ. As described above, the phase shift angle is measured relative to a reference signal.

[0273] Referring to FIG. **9**A, $\vec{R}'$ represents the relative amplitude component of $\vec{R}_{in}$ generated by $\vec{U}'$ and $\vec{L}'$.

[0274] Still referring to FIG. **9**A, it is noted that, at any time instant, $\vec{R}'$ can be obtained by the sum of an upper phasor $\vec{U}'$ and a lower phasor $\vec{L}'$. Further, $\vec{U}'$ and $\vec{L}'$ can be maintained to have substantially constant magnitude. The phasors, $\vec{U}'$ and $\vec{L}'$, accordingly, represent two substantially constant envelope signals. r'(t) can thus be obtained, at any time instant, by the sum of two substantially constant envelope signals that correspond to phasors $\vec{U}'$ and $\vec{L}'$.

[0275] The phase shifts of phasors $\vec{U}'$ and $\vec{L}'$ relative to $\vec{R}'$ are set according to the desired magnitude R of $\vec{R}'$. In the simplest case, when upper and lower phasors $\vec{U}'$ and $\vec{L}'$ are selected to have equal magnitude, upper and lower phasors $\vec{U}'$ and $\vec{L}'$ are substantially symmetrically shifted in phase relative to $\vec{R}'$. This is illustrated in the example of FIG. **9**A. It is noted that terms and phrases indicating or suggesting orientation, such as but not limited to "upper and lower" are used herein for ease of reference and are not functionally or structurally limiting.

[0276] It can be verified that, for the case illustrated in FIG. **9**A, the phase shift of $\vec{U}'$ and $\vec{L}'$ relative to $\vec{R}'$, illustrated as angle

$$\frac{\phi}{2}$$

in FIG. **9**A, is related to the magnitude of $\vec{R}'$ as follows:

$$\frac{\phi}{2}=\cot^{-1}\left(\frac{R}{2\sqrt{1-\frac{R^2}{4}}}\right) \tag{7}$$

where R represents a normalized magnitude of phasor $\vec{R}'$.

[0277] Equation (7) can further be reduced to

$$\frac{\phi}{2}=\cos^{-1}\left(\frac{R}{2}\right) \tag{7.10}$$

where R represents a normalized magnitude of phasor $\vec{R}'$.

[0278] Alternatively, any substantially equivalent mathematical equations or other substantially equivalent mathematical techniques such as look up tables can be used.

[0279] It follows from the above discussion that, in phasor representation, any phasor $\vec{R}'$ of variable magnitude and phase can be constructed by the sum of two constant magnitude phasor components:

$$\vec{R}'=\vec{U}'+\vec{L}'|\vec{U}|=|\vec{L}|=A=\text{constant} \tag{8}$$

[0280] Correspondingly, in the time domain, a time-varying envelope sinusoidal signal r'(t)=R(t)×cos(ωt) is constructed by the sum of two constant envelope signals as follows:

$$r'(t)=U'(t)+L'(t); \tag{9}$$
$$U'(t)=A\times\cos\left(\omega t+\frac{\phi}{2}\right);$$
$$L'(t)=A\times\cos\left(\omega t-\frac{\phi}{2}\right);$$

where A is a constant and

$$\frac{\phi}{2}$$

is as shown in equation (7).

[0281] From FIG. **9**A, it can be further verified that equations (9) can be rewritten as:

$$r'(t)=U'(t)+L'(t);$$
$$U'(t)=C\cos(\omega t)+\alpha\sin(\omega t);$$
$$L'(t)=C\cos(\omega t)-\beta\sin(\omega t); \tag{10}$$

where C denotes the real part component of phasors $\vec{U}'$ and $\vec{L}'$ and is equal to

$$A \times \cos\left(\frac{\phi}{2}\right).$$

Note that C is a common component of $\overrightarrow{U'}$ and $\overrightarrow{L'}$. α and β denote the imaginary part components of phasors $\overrightarrow{U'}$ and $\overrightarrow{L'}$, respectively.

[0282]

$$\alpha = \beta = A \times \sin\left(\frac{\phi}{2}\right).$$

Accordingly, from equations (12),

[0283]

$$r'(t) = 2C \times \cos(\omega t) = 2A \times \cos\left(\frac{\phi}{2}\right) \times \cos(\omega t).$$

As understood by a person skilled in the art based on the teachings herein, other equivalent and/or simplified representations of the above representations of the quantities A, B, and C may also be used, including look up tables, for example.

[0284] Note that $\overrightarrow{R_{in}}$ is shifted by θ degrees relative to $\overrightarrow{R'}$. Accordingly, using equations (8), it can be deduced that:

$$\overrightarrow{R_{in}} = \overrightarrow{R'}e^{j\theta} = (\overrightarrow{U'} + \overrightarrow{L'})e^{j\theta} = \overrightarrow{U'}e^{j\theta + \overrightarrow{L'}ej\theta} \tag{11}$$

[0285] Equations (11) imply that a representation of $\overrightarrow{R_{in}}$ can be obtained by summing phasors $\overrightarrow{U'}$ and $\overrightarrow{L'}$, described above, shifted by θ degrees. Further, an amplified output version, $\overrightarrow{R_{out}}$, of $\overrightarrow{R_{in}}$ can be obtained by separately amplifying substantially equally each of the θ degrees shifted versions of phasors $\overrightarrow{U'}$ and $\overrightarrow{\ '}$, and summing them. FIG. **9**B illustrates this concept. In FIG. **9**B, phasors $\overrightarrow{U}$ and $\overrightarrow{L}$ represent θ degrees shifted and amplified versions of phasors $\overrightarrow{U'}$ and $\overrightarrow{L'}$. Note that, since $\overrightarrow{U'}$ and $\overrightarrow{L'}$ are constant magnitude phasors, $\overrightarrow{U}$ and $\overrightarrow{L}$ are also constant magnitude phasors. Phasors $\overrightarrow{U}$ and $\overrightarrow{L}$ sum, as shown FIG. **9**B, to phasor $\overrightarrow{R_{out}}$, which is a power amplified version of input signal $\overrightarrow{R_{in}}$.

[0286] Equivalently, in the time domain, it can be shown that:

$$r_{out}(t) = U(t) + L(t);$$

$$U(t) = K[C\cos(\omega t+\theta) + \alpha\sin(\omega t+\theta)];$$

$$L(t) = K\cdot[C\cos(\omega t+\theta) - \beta\sin(\omega t+\theta)]. \tag{12}$$

where $r_{out}(t)$ corresponds to the time domain signal represented by phasor $\overrightarrow{R_{out}}$, U(t) and L(t) correspond to the time domain signals represents by phasors $\overrightarrow{U}$ and $\overrightarrow{L}$, and K is the power amplification factor.

[0287] A person skilled in the art will appreciate that, whereas the time domain representations in equations (9) and (10) have been provided for the case of a sinusoidal waveform, equivalent representations can be developed for non-sinusoidal waveforms using appropriate basis functions.

[0288] FIG. **10** is a block diagram that conceptually illustrates an exemplary embodiment **1000** of the CPCP 2-Branch VPA embodiment. An output signal r(t) of desired power level and frequency characteristics is generated from in-phase and quadrature components according to the CPCP 2-Branch VPA embodiment.

[0289] In the example of FIG. **10**, a clock signal **1010** represents a reference signal for generating output signal r(t). Clock signal **1010** is of the same frequency as that of desired output signal r(t).

[0290] Referring to FIG. **10**, an Iclk_phase signal **1012** and a Qclk_phase signal **1014** represent amplitude analog values that are multiplied by the in-phase and quadrature components of Clk signal **1010** and are calculated from the baseband I and Q signals.

[0291] Still referring to FIG. **10**, clock signal **1010** is multiplied with Iclk_phase signal **1012**. In parallel, a 90° degrees shifted version of clock signal **1010** is multiplied with Qclk_phase signal **1014**. The two multiplied signals are combined to generate Rclk signal **1016**. Rclk signal **1016** is of the same frequency as clock signal **1010**. Further, Rclk signal **1016** is characterized by a phase shift angle according to the ratio of Q(t) and I(t). The magnitude of Rclk signal **1016** is such that $R^2clk = I^2clk\_phase + Q^2clk\_phase$. Accordingly, Rclk signal **1016** represents a substantially constant envelope signal having the phase characteristics of the desired output signal r(t).

[0292] Still referring to FIG. **10**, Rclk signal **1016** is input, in parallel, into two vector modulators **1060** and **1062**. Vector modulators **1060** and **1062** generate the U(t) and L(t) substantially constant envelope constituents, respectively, of the desired output signal r(t) as described in (12). In vector modulator **1060**, an in-phase Rclk signal **1020**, multiplied with Common signal **1028**, is combined with a 90° degree shifted version **1018** of Rclk signal, multiplied with first signal **1026**. In parallel, in vector modulator **1062**, an in-phase Rclk signal **1022**, multiplied with Common signal **1028**, is combined with a 90° degrees shifted version **1024** of Rclk signal, multiplied with second signal **1030**. Common signal **1028**, first signal **1026**, and second signal **1030** correspond, respectively, to the real part C and the imaginary parts α and β described in equation (12).

[0293] Output signals **1040** and **1042** of respective vector modulators **1060** and **1062** correspond, respectively, to the U(t) and L(t) constant envelope constituents of input signal r(t).

[0294] As described above, signals **1040** and **1042** are characterized by having substantially equal and constant magnitude envelopes. Accordingly, when signals **1040** and **1042** are input into corresponding power amplifiers (PA) **1044** and **1046**, corresponding amplified signals **1048** and **1050** are substantially constant envelope signals.

[0295] Power amplifiers **1044** and **1046** apply substantially equal power amplification to signals **1040** and **1042**, respectively. In an embodiment, the power amplification level of PAs **1044** and **1046** is set according to the desired power level of output signal r(t). Further, amplified signals **1048** and **1050** are in-phase relative to each other. Accordingly, when summed together, as shown in FIG. **10**, resulting signal **1052** corresponds to the desired output signal r(t).

[0296] FIG. **10**A is another exemplary embodiment **1000**A of the CPCP 2-Branch VPA embodiment. Embodiment **1000**A represents a Multiple Input Single Output (MISO) implementation of embodiment **1000** of FIG. **10**.

[0297] In embodiment **1000**A, constant envelope signals **1040** and **1042**, output from vector modulators **1060** and **1062**, are input into MISO PA **1054**. MISO PA **1054** is a two-input single-output power amplifier. In an embodiment, MISO PA **1054** may include various elements, such as pre-drivers, drivers, power amplifiers, and process detectors (not shown in FIG. **10**A), for example. Further, MISO PA **1054** is not limited to being a two-input PA as shown in FIG. **10**A. In other embodiments, as will be described further below with reference to FIGS. **51**A-H, PA **1054** can have any number of inputs.

[0298] Operation of the CPCP 2-Branch VPA embodiment is depicted in the process flowchart **1100** of FIG. **11**.

[0299] The process begins at step **1110**, which includes receiving a baseband representation of the desired output signal. In an embodiment, this involves receiving in-phase (I) and quadrature (Q) components of the desired output signal. In another embodiment, this involves receiving magnitude and phase of the desired output signal.

[0300] Step **1120** includes receiving a clock signal set according to a desired output signal frequency of the desired output signal. In the example of FIG. **10**, step **1120** is achieved by receiving clock signal **1010**.

[0301] Step **1130** includes processing the clock signal to generate a normalized clock signal having a phase shift angle according to the received I and Q components. In an embodiment, the normalized clock signal is a constant envelope signal having a phase shift angle according to a ratio of the I and Q components. The phase shift angle of the normalized clock is relative to the original clock signal. In the example of FIG. **10**, step **1130** is achieved by multiplying clock signal **101O**'s in-phase and quadrature components with Iclk_phase **1012** and Qclk_phase **1014** signals, and then summing the multiplied signal to generate Rclk signal **1016**.

[0302] Step **1140** includes the processing of the I and Q components to generate the amplitude information required to produce first and second substantially constant envelope constituent signals.

[0303] Step **1150** includes processing the amplitude information of step **1140** and the normalized clock signal Rclk to generate the first and second constant envelope constituents of the desired output signal. In an embodiment, step **1150** involves phase shifting the first and second constant envelope constituents of the desired output signal by the phase shift angle of the normalized clock signal. In the example of FIG. **10**, step **1150** is achieved by vector modulators **1060** and **1062** modulating Rclk signal **1016** with first signal **1026**, second signal **1030**, and common signal **1028** to generate signals **1040** and **1042**.

[0304] Step **1160** includes individually amplifying the first and second constant envelope constituents, and summing the amplified signals to generate the desired output signal. In an embodiment, the amplification of the first and second constant envelope constituents is substantially equal and according to a desired power level of the desired output signal. In the example of FIG. **10**, step **1160** is achieved by PAs **1044** and **1046** amplifying signals **1040** and **1042** to generate amplified signals **1048** and **1050**.

[0305] FIG. **12** is a block diagram that illustrates an exemplary embodiment of a vector power amplifier **1200** implementing the process flowchart **1100**. Optional components are illustrated with dashed lines, although in other embodiments more or less components may be optional.

[0306] Referring to FIG. **12**, in-phase (I) and quadrature (Q) information signal **1210** is received by an I and Q Data Transfer Function module **1216**. In an embodiment, I and Q Data Transfer Function **1216** samples signal **1210** according to a sample clock **1212**. I and Q information signal **1210** includes baseband I and Q information of a desired output signal r(t).

[0307] In an embodiment, I and Q Data Transfer Function module **1216** processes information signal **1210** to generate information signals **1220**, **1222**, **1224**, and **1226**. The operation of I and Q Data Transfer Function module **1216** is further described below in section 3.4.

[0308] Referring to FIG. **12**, information signal **1220** includes quadrature amplitude information of first and second constant envelope constituents of a baseband version of desired output signal r(t). With reference to FIG. **9**A, for example, information signal **1220** includes the α and β quadrature components. Referring again to FIG. **12**, information signal **1226** includes in-phase amplitude information of the first and second constant envelope constituents of the baseband version of signal r(t). With reference to FIG. **9**A, for example, information signal **1226** includes the common C in-phase component.

[0309] Still referring to FIG. **12**, information signals **1222** and **1224** include normalized in-phase Iclk_phase and quadrature Qclk_phase signals, respectively. Iclk_phase and Qclk_phase are normalized versions of the I and Q information signals included in signal **1210**. In an embodiment, Iclk_phase and Qclk_phase are normalized such that that (I^2clk_phase+Q^2clk_phase=constant). It is noted that the phase of signal **1250** corresponds to the phase of the desired output signal and is created from Iclk_phase and Qclk_phase. Referring to FIG. **9**B, Iclk_phase and Qclk_phase are related to I and Q as follows:

$$\theta = \tan^{-1}\left(\frac{Q}{I}\right) = \tan^{-1}\left(\frac{Q_{\text{clk\_phase}}}{I_{\text{clk\_phase}}}\right)$$

(12.1)

where θ represents the phase of the desired output signal, represented b

[0310] phasor $\overrightarrow{R_{out}}$ in FIG. **9**B. The sign information of the baseband I and Q information must be taken into account to calculate θ for all four quadrants.

[0311] In the exemplary embodiment of FIG. **12**, information signals **1220**, **1222**, **1224**, and **1226** are digital signals. Accordingly, each of signals **1220**, **1222**, **1224**, and **1226** is fed into a corresponding digital-to-analog converter (DAC) **1230**, **1232**, **1234**, and **1236**. The resolution and sample rate of DACs **1230**, **1232**, **1234**, and **1236** is selected according to specific signaling schemes. DACs **1230**, **1232**, **1234**, and **1236** are controlled by DAC clock signals **1221**, **1223**, **1225**, and **1227**, respectively. DAC clock signals **1221**, **1223**, **1225**, and **1227** may be derived from a same clock signal or may be independent.

[0312] In other embodiments, information signals **1220**, **1222**, **1224**, and **1226** are generated in analog format and no DACs are required.

[0313] Referring to FIG. **12**, DACs **1230**, **1232**, **1234**, and **1236** convert digital information signals **1220**, **1222**, **1224**, and **1226** into corresponding analog signals, and input these analog signal into optional interpolation filters **1231**, **1233**, **1235**, and **1237**, respectively. Interpolation filters **1231**, **1233**, **1235**, and **1237**, which also serve as anti-aliasing filters, shape the DACs output signals to produce the desired output waveform. Interpolation filters **1231**, **1233**, **1235**, and **1237** generate signals **1240**, **1244**, **1246**, and **1248**, respectively. Signal **1242** represents the inverse of signal **1240**.

[0314] Still referring to FIG. **12**, signals **1244** and **1246**, which include Iclk_phase and Qclk_phase information, are input into a vector modulator **1238**. Vector modulator **1238** multiplies signal **1244** with a channel clock signal **1214**. Channel clock signal **1214** is selected according to a desired output signal frequency. In parallel, vector modulator **1238** multiplies signal **1246** with a 90° shifted version of channel clock signal **1214**. In other words, vector modulator **1238** generates an in-phase component having amplitude of Iclk_phase and a quadrature component having amplitude of Qclk_phase.

[0315] Vector modulator **1238** combines the two modulated signals to generate Rclk signal **1250**. Rclk signal **1250** is a substantially constant envelope signal having the desired output frequency and a phase shift angle according to the I and Q data included in signal **1210**.

[0316] Still referring to FIG. **12**, signals **1240**, **1242**, and **1248** include the U, L, and Common C amplitude components, respectively, of the complex envelope of signal r(t). Signals **1240**, **1242**, and **1248** along with Rclk signal **1250** are input into vector modulators **1260** and **1262**.

[0317] Vector modulator **1260** combines signal **1240**, multiplied with a 90° shifted version of Rclk signal **1250**, and signal **1248**, multiplied with a 0° shifted version of Rclk signal **1250**, to generate output signal **1264**. In parallel, vector modulator **1262** combines signal **1242**, multiplied with a 90° shifted version of Rclk signal **1250**, and signal **1248**, modulated with a 0° shifted version of Rclk signal **1250**, to generate output signal **1266**.

[0318] Output signals **1264** and **1266** represent substantially constant envelope signals. Further, phase shifts of output signals **1264** and **1266** relative to Rclk signal **1250** are determined by the angle relationships associated with the ratios α/C and β/C, respectively. In an embodiment, $\alpha=-\beta$ and therefore output signals **1264** and **1266** are symmetrically phased relative to Rclk signal **1250**. With reference to FIG. **9**B, for example, output signals **1264** and **1266** correspond, respectively, to the $\vec{U}$ and $\vec{L}$ constant magnitude phasors.

[0319] A sum of output signals **1264** and **1266** results in a channel-clock-modulated signal having the I and Q characteristics of baseband signal r(t). To achieve a desired power level at the output of vector power amplifier **1200**, however, signals **1264** and **1266** are amplified to generate an amplified output signal. In the embodiment of FIG. **12**, signals **1264** and **1266** are, respectively, input into power amplifiers (PAs) **1270** and **1272** and amplified. In an embodiment, PAs **1270** and **1272** include switching power amplifiers. Autobias circuitry **1218** controls the bias of PAs **1270** and **1272** as further described below in section 3.5.2. In the embodiment of FIG. **12**, for example, autobias circuitry **1218** provides a bias voltage **1228** to PAs **1270** and **1272**.

[0320] In an embodiment, PAs **1270** and **1272** apply substantially equal power amplification to respective constant envelope signals **1264**-**1266**. In an embodiment, the power amplification is set according to the desired output power level. In other embodiments of vector power amplifier **1200**, PA drivers and/or pre-drivers are additionally employed to provide additional power amplification capability to the amplifier. In the embodiment of FIG. **12**, for example, PA drivers **1284** and **1286** are optionally added, respectively, between vector modulators **1260** and **1262** and subsequent PAs **1270** and **1272**.

[0321] Respective output signals **1274** and **1276** of PAs **1270** and **1272** are substantially constant envelope signals. Further, when output signals **1274** and **1276** are summed, the resulting signal has minimal non-linear distortion. In the embodiment of FIG. **12**, output signals **1274** and **1276** are coupled together to generate output signal **1280** of vector power amplifier **1200**. In an embodiment, no isolation is used in coupling the outputs of PAs **1270** and **1272**. Accordingly, minimal power loss is incurred by the coupling. In an embodiment, the outputs of PAs **1270** and **1272** are directly coupled together using a wire. Direct coupling in this manner means that there is minimal or no resistive, inductive, or capacitive isolation between the outputs of PAs **1270** and **1272**. In other words, outputs of PAs **1270** and **1272** are coupled together without intervening components. Alternatively, in an embodiment, the outputs of PAs **1270** and **1272** are coupled together indirectly through inductances and/or capacitances that result in low or minimal impedance connections, and/or connections that result in minimal isolation and minimal power loss. Alternatively, outputs of PAs **1270** and **1272** are coupled using well known combining techniques, such as Wilkinson, hybrid combiners, transformers, or known active combiners. In an embodiment, the PAs **1270** and **1272** provide integrated amplification and power combining in a single operation. In an embodiment, one or more of the power amplifiers and/or drivers described herein are implemented using multiple input, single output power amplification techniques, examples of which are shown in FIGS. **12**A, **12**B, and **51**A-H.

[0322] Output signal **1280** represents a signal having the I and Q characteristics of baseband signal r(t) and the desired output power level and frequency. In embodiments of vector power amplifier **1200**, a pull-up impedance **1288** is coupled between the output of vector power amplifier **1200** and a power supply. In other embodiments, an impedance matching network **1290** is coupled at the output of vector power amplifier **1200**. Output stage embodiments according to power amplification methods and systems of the present invention will be further described below in section 3.5.

[0323] In other embodiments of vector power amplifier **1200**, process detectors are employed to compensate for any process variations in circuitry of the amplifier. In the exemplary embodiment of FIG. **12**, for example, process detector **1282** is optionally added to monitor variations in PA drivers **1284** and **1286**.

[0324] FIG. **12**A is a block diagram that illustrates another exemplary embodiment of a vector power amplifier **1200**A implementing the process flowchart **1100**. Optional components are illustrated with dashed lines, although in other embodiments more or less components may be optional.

[0325] Embodiment **1200**A illustrates a multiple-input single-output (MISO) implementation of embodiment **1200**. In embodiment **1200**A, constant envelope signals **1261** and **1263**, output from vector modulators **1260** and **1262**, are input into MISO PA **1292**. MISO PA **1292** is a two-input single-output power amplifier. In an embodiment, MISO PA **1292** includes elements **1270**, **1272**, **1282**, **1284**, and **1286** as shown in the embodiment of FIG. **12**. In another embodiment, MISO PA **1292** may include other elements, such as pre-drivers, not shown in the embodiment of FIG. **12**. Further, MISO PA **1292** is not limited to being a two-input PA as shown in FIG. **12**A. In other embodiments as will be described further below with reference to FIGS. **51**A-H, PA **1292** can have any number of inputs and outputs.

[0326] Still referring to FIG. **12**A, embodiment **1200**A illustrates one implementation for delivering autobias signals to MISO PA **1292**. In the embodiment of FIG. **12**A, Autobias signal **1228** generated by Autobias circuitry **1218**, has one or more signals derived from it to bias different stages of MISO PA **1292**. As shown in the example of FIG. **12**A, three bias control signals Bias A, Bias B, and Bias C are derived from Autobias signal **1228**, and then input at different stages of MISO PA **1292**. For example, Bias C may be the bias signal to the pre-driver stage of MISO PA **1292**. Similarly, Bias B and Bias A may be the bias signals to the driver and PA stages of MISO PA **1292**.

[0327] In another implementation, shown in embodiment **1200**B of FIG. **12** B, Autobias circuitry **1218** generates separate Autobias signals **1295**, **1296**, and **1295**, corresponding to Bias A, Bias B, and Bias C, respectively. Signals **1295**, **1296**, and **1297** may or may not be generated separately within Autobias circuitry **1218**, but are output separately as shown. Further, signals **1295**, **1296**, and **1297** may or may not be related as determined by the biasing of the different stages of MISO PA **1294**.

[0328] Other aspects of vector power amplifiers **1200**A and **1200**B substantially correspond to those described above with respect to vector power amplifier **1200**.

[0329] FIG. **13** is a block diagram that illustrates another exemplary embodiment **1300** of a vector power amplifier according to the CPCP 2-Branch VPA embodiment. Optional components are illustrated with dashed lines, although in other embodiments more or less components may be optional.

[0330] In the exemplary embodiment of FIG. **13**, a DAC of sufficient resolution and sample rate **1320** replaces DACs **1230**, **1232**, **1234** and **1236** of the embodiment of FIG. **12**. DAC **1320** is controlled by a DAC clock **1324**.

[0331] DAC **1320** receives information signal **1310** from I and Q Data Transfer Function module **1216**. Information signal **1310** includes identical information content to signals **1220**, **1222**, **1224** and **1226** in the embodiment of FIG. **12**.

[0332] DAC **1320** may output a single analog signal at a time. Accordingly, a sample-and-hold architecture may be used as shown in FIG. **13**.

[0333] DAC **1320** sequentially outputs analog signals **1332**, **1334**, **1336**, **1336** to a first set of sample-and-hold circuits **1342**, **1344**, **1346**, and **1348**. In an embodiment, DAC **1230** is clocked at a sufficient rate to replace DACs **1230**, **1232**, **1234**, and **1236** of the embodiment of FIG. **12**. An output selector **1322** determines which of output signals **1332**, **1334**, **1336**, and **1338** should be selected for output.

[0334] DAC **1320**'s DAC clock signal **1324**, output selector signal **1322**, and sample-and-hold clocks **1340**A-D and **1350** are controlled by a control module that can be independent or integrated into transfer function module **1216**.

[0335] In an embodiment, sample-and-hold circuits (S/H) **1342**, **1344**, **1346**, and **1348** hold the received analog values and, according to a clock signal **1340**A-D, release the values to a second set of sample-and-hold circuits **1352**, **1354**, **1356**, and **1358**. For example, S/H **1342** release its value to S/H **1352** according to a received clock signal **1340**A. In turn, sample-and-hold circuits **1352**, **1354**, **1356**, and **1358** hold the received analog values, and simultaneously release the values to interpolation filters **1231**, **1233**, **1235**, and **1237** according to a common clock signal **1350**. A common clock signal **1350** is used in order to ensure that the outputs of S/H **1352**, **1354**, **1356**, and **1358** are time-aligned.

[0336] In another embodiment, a single layer of S/H circuitry that includes S/H **1342**, **1344**, **1346**, and **1348** can be employed. Accordingly, S/H circuits **1342**, **1344**, **1346**, and **1348** receive analog values from DAC **1320**, and each releases its received value according to a clock independent of the others. For example, S/H **1342** is controlled by clock **1340**A, which may not be synchronized with clock **1340**B that controls S/H **1344**. To ensure that outputs of S/H circuits **1342**, **1344**, **1346**, and **1348** are time-aligned, delays between clocks **1340**A-D are pre-compensated for in prior stages of the amplifier. For example, DAC **1320** outputs signal **1332**, **1334**, **1336**, and **1338** with appropriately selected delays to S/H circuits **1342**, **1344**, **1346**, and **1348** in order to compensate for the time differences between clocks **1340**A-D.

[0337] Other aspects of vector power amplifier **1300** are substantially equivalent to those described above with respect to vector power amplifier **1200**.

[0338] FIG. **13**A is a block diagram that illustrates another exemplary embodiment **1300**A of a vector power amplifier according to the CPCP 2-Branch VPA embodiment. Optional components are illustrated with dashed lines, although in other embodiments more or less components may be optional. Embodiment **1300**A is a MISO implementation of embodiment **1300** of FIG. **13**.

[0339] In the embodiment of FIG. **13**A, constant envelope signals **1261** and **1263** output from vector modulators **1260** and **1262** are input into MISO PA **1360**. MISO PA **1360** is a two-input single-output power amplifier. In an embodiment, MISO PA **1360** includes elements **1270**, **1272**, **1282**, **1284**, and **1286** as shown in the embodiment of FIG. **13**. In another embodiment, MISO PA **1360** may include other elements, such as pre-drivers, not shown in the embodiment of FIG. **13**, or functional equivalents thereof. Further, MISO PA **1360** is not limited to being a two-input PA as shown in FIG. **13**A. In other embodiments as will be described further below with reference to FIGS. **51**A-H, PA **1360** can have any number of inputs.

[0340] The embodiment of FIG. **13**A further illustrates two different sample and hold architectures with a single or two levels of S/H circuitry as shown. The two implementations have been described above with respect to FIG. **13**.

[0341] Embodiment **1300**A also illustrates optional bias control circuitry **1218** and associated bias control signal

1325, **1326**, and **1327**. Signals **1325**, **1326**, and **1327** may be used to bias different stages of MISO PA **1360** in certain embodiments.

[0342] Other aspects of vector power amplifier **1300**A are equivalent to those described above with respect to vector power amplifiers **1200** and **1300**.

3.3) Direct Cartesian 2-Branch Vector Power Amplifier

[0343] A Direct Cartesian 2-Branch VPA embodiment shall now be described. This name is used herein for reference purposes, and is not functionally or structurally limiting.

[0344] According to the Direct Cartesian 2-Branch VPA embodiment, a time-varying envelope signal is decomposed into two constant envelope constituent signals. The constituent signals are individually amplified equally or substantially equally, and then summed to construct an amplified version of the original time-varying envelope signal.

[0345] In one embodiment of the Direct Cartesian 2-Branch VPA embodiment, a magnitude and a phase angle of a time-varying envelope signal are calculated from in-phase and quadrature components of an input signal. Using the magnitude and phase information, in-phase and quadrature amplitude components are calculated for two constant envelope constituents of the time-varying envelope signal. The two constant envelope constituents are then generated, amplified equally or substantially equally, and summed to generate an amplified version of the original time-varying envelope signal R_{in}.

[0346] The concept of the Direct Cartesian 2-Branch VPA will now be described with reference to FIGS. **9**A and **14**.

[0347] As described and verified above with respect to FIG. **9**A, the phasor $\vec{R}'$ can be obtained by the sum of an upper phasor $\vec{U}'$ and a lower phasor $\vec{L}'$ appropriately phased to produce $\vec{R}'$. $\vec{R}'$ is calculated to be proportional to the magnitude R_{in}. Further, $\vec{U}'$ and $\vec{L}'$ can be maintained to have substantially constant magnitude. In the time domain, $\vec{U}'$ and $\vec{L}'$ represent two substantially constant envelope signals. The time domain equivalent r'(t) of $\vec{R}'$ can thus be obtained, at any time instant, by the sum of two substantially constant envelope signals.

[0348] For the case illustrated in FIG. **9**A, the phase shift of $\vec{U}'$ and $\vec{L}'$ relative to $\vec{R}'$, illustrated as angle

$$\frac{\phi}{2}$$

in FIG. **9**A, is related to the magnitude of $\vec{R}'$ as follows:

$$\frac{\phi}{2} = \cot^{-1}\left(\frac{R}{2\sqrt{1-\frac{R^2}{4}}}\right) \tag{13}$$

where R represents the normalized magnitude of phasor $\vec{R}'$.

[0349] In the time domain, it was shown that a time-varying envelope signal, r'(t)=R(t)cos(ωt) for example, can be constructed by the sum of two constant envelope signals as follows:

$$r'(t)=U'(t)+L'(t);$$

$$U'(t)=C\times\cos(\omega t)+\alpha\times\sin(\omega t);$$

$$L'(t)=C\times\cos(\omega t)-\beta\times\sin(\omega t). \tag{14}$$

where C denotes the in-phase amplitude component of phasors $\vec{U}'$ and $\vec{L}'$ and is equal or substantially equal to

$$A\times\cos\left(\frac{\phi}{2}\right)$$

(A being a constant). α and β denote the quadrature amplitude components of phasors $\vec{U}'$ and $\vec{L}'$, respectively.

[0350]

$$\alpha=\beta=A\times\sin\left(\frac{\phi}{2}\right).$$

Note that equations (14) can be modified for non-sinusoidal signals by changing the basis function from sinusoidal to the desired function.

[0351] FIG. **14** illustrates phasor $\vec{R}$ and its two constant magnitude constituent phasors $\vec{U}$ and $\vec{L}$. $\vec{R}$ is shifted by θ degrees relative to $\vec{R}'$ in FIG. **9**A. Accordingly, it can be verified that:

$$R=\vec{R}'\times e^{j\theta}=(\vec{U}'+\vec{L}')\times e^{j\theta}=\vec{U}+\vec{L};$$

$$\vec{U}=\vec{U}'\times e^{j\theta};$$

$$\vec{L}=\vec{L}'\times e^{j\theta}. \tag{15}$$

[0352] From equations (15), it can be further shown that:

$$\vec{U}=\vec{U}'\times e^{j\theta}=(C+j\alpha)\times e^{j\theta};\Rightarrow\vec{U}=(C+j\alpha)(\cos\theta+j\sin\theta)=(C\cos\theta-\alpha\sin\theta)+j(C\sin\theta+\alpha\cos\theta). \tag{16}$$

[0353] Similarly, it can be shown that:

$$\vec{L}=\vec{L}'\times e^{j\theta}=(C+j\beta)\times e^{j\theta};\Rightarrow\vec{L}=(C+j\beta)(\cos\theta+j\sin\theta)=(C\cos\theta-\beta\sin\theta)+j(C\sin\theta+\beta\cos\theta). \tag{17}$$

[0354] Equations (16) and (17) can be re-written as:

$$\vec{U}=(C\cos\theta-\alpha\sin\theta)+j(C\sin\theta+\alpha\cos\theta)=U_x+jU_y;$$

$$\vec{L}=(C\cos\theta-\beta\sin\theta)+j(C\sin\theta+\beta\cos\theta)=L_x+jL_y. \tag{18}$$

[0355] Equivalently, in the time domain:

$$U(t)=U_x\phi_1(t)+U_y\phi_2(t);$$

$$L(t)=L_x\phi_1(t)+L_y\phi_2(t); \tag{19}$$

[0356] where $\phi_1(t)$ and $\phi_2(t)$ represent an appropriately selected orthogonal basis functions.

[0357] From equations (18) and (19), it is noted that it is sufficient to calculate the values of α, β, C and $\sin(\Theta)$ and $\cos(\Theta)$) in order to determine the two constant envelope constituents of a time-varying envelope signal r(t). Further, α, β and C can be entirely determined from magnitude and phase information, equivalently I and Q components, of signal r(t).

[0358] FIG. **15** is a block diagram that conceptually illustrates an exemplary embodiment **1500** of the Direct Cartesian 2-Branch VPA embodiment. An output signal r(t) of desired power level and frequency characteristics is generated from in-phase and quadrature components according to the Direct Cartesian 2-Branch VPA embodiment.

[0359] In the example of FIG. **15**, a clock signal **1510** represents a reference signal for generating output signal r(t). Clock signal **1510** is of the same frequency as that of desired output signal r(t).

[0360] Referring to FIG. **15**, exemplary embodiment **1500** includes a first branch **1572** and a second branch **1574**. The first branch **1572** includes a vector modulator **1520** and a power amplifier (PA) **1550**. Similarly, the second branch **1574** includes a vector modulator **1530** and a power amplifier (PA) **1560**.

[0361] Still referring to FIG. **15**, clock signal **1510** is input, in parallel, into vector modulators **1520** and **1530**. In vector modulator **1520**, an in-phase version **1522** of clock signal **1510**, multiplied with U_x signal **1526**, is summed with a 90° degrees shifted version **1524** of clock signal **1510**, multiplied with U_y signal **1528**. In parallel, in vector modulator **1530**, an in-phase version **1532** of clock signal **1510**, multiplied with Lx signal **1536**, is summed with a 90° degrees shifted version **1534** of clock signal **1510**, multiplied with Ly signal **1538**. U_x signal **1526** and U_y signal **1528** correspond, respectively, to the in-phase and quadrature amplitude components of the U(t) constant envelope constituent of signal r(t) provided in equation (19). Similarly, L_x signal **1536**, and L_y signal **1538** correspond, respectively, to the in-phase and quadrature amplitude components of the L(t) constant envelope constituent of signal r(t) provided in equation (19).

[0362] Accordingly, respective output signals **1540** and **1542** of vector modulators **1520** and **1530** correspond, respectively, to the U(t) and L(t) constant envelope constituents of signal r(t) as described above in equations (19). As described above, signals **1540** and **1542** are characterized by having equal and constant or substantially equal and constant magnitude envelopes.

[0363] Referring to FIG. **15**, to generate the desired power level of output signal r(t), signals **1540** and **1542** are input into corresponding power amplifiers **1550** and **1560**.

[0364] In an embodiment, power amplifiers **1550** and **1560** apply equal or substantially equal power amplification to signals **1540** and **1542**, respectively. In an embodiment, the power amplification level of PAs **1550** and **1560** is set according to the desired power level of output signal r(t).

[0365] Amplified output signals **1562** and **1564** are substantially constant envelope signals. Accordingly, when summed together, as shown in FIG. **15**, resulting signal **1570** corresponds to the desired output signal r(t).

[0366] FIG. **15**A is another exemplary embodiment **1500**A of the Direct Cartesian 2-Branch VPA embodiment. Embodiment **1500**A represents a Multiple Input Signal Output (MISO) implementation of embodiment **1500** of FIG. **15**.

[0367] In embodiment **1500**A, constant envelope signals **1540** and **1542**, output from vector modulators **1520** and **1530**, are input into MISO PA **1580**. MISO PA **1580** is a two-input single-output power amplifier. In an embodiment, MISO PA **1580** may include various elements, such as pre-drivers, drivers, power amplifiers, and process detectors (not shown in FIG. **15**A), for example. Further, MISO PA **1580** is not limited to being a two-input PA as shown in FIG. **15**A. In other embodiments, as will be described further below with reference to FIGS. **51**A-H, PA **1580** can have any number of inputs.

[0368] Operation of the Direct Cartesian 2-Branch VPA embodiment is depicted in the process flowchart **1600** of FIG. **16**. The process begins at step **1610**, which includes receiving a baseband representation of a desired output signal. In an embodiment, the baseband representation includes I and Q components. In another embodiment, the I and Q components are RF components that are down-converted to baseband.

[0369] Step **1620** includes receiving a clock signal set according to a desired output signal frequency of the desired output signal. In the example of FIG. **15**, step **1620** is achieved by receiving clock signal **1510**.

[0370] Step **1630** includes processing the I and Q components to generate in-phase and quadrature amplitude information of first and second constant envelope constituent signals of the desired output signal. In the example of FIG. **15**, the in-phase and quadrature amplitude information is illustrated by U_x, U_y, L_x, and L_y.

[0371] Step **1640** includes processing the amplitude information and the clock signal to generate the first and second constant envelope constituent signals of the desired output signal. In an embodiment, the first and second constant envelope constituent signals are modulated according to the desired output signal frequency. In the example of FIG. **15**, step **1640** is achieved by vector modulators **1520** and **1530**, clock signal **1510**, and amplitude information signals **1526**, **1528**, **1536**, and **1538** to generate signals **1540** and **1542**.

[0372] Step **1650** includes amplifying the first and second constant envelope constituents, and summing the amplified signals to generate the desired output signal. In an embodiment, the amplification of the first and second constant envelope constituents is according to a desired power level of the desired output signal. In the example of FIG. **15**, step **1650** is achieved by PAs **1550** and **1560** amplifying respective signals **1540** and **1542** and, subsequently, by the summing of amplified signals **1562** and **1564** to generate output signal **1574**.

[0373] FIG. **17** is a block diagram that illustrates an exemplary embodiment of a vector power amplifier **1700** implementing the process flowchart **1600**. Optional components are illustrated with dashed lines, although other embodiments may have more or less optional components.

[0374] Referring to FIG. **17**, in-phase (I) and quadrature (Q) information signal **1710** is received by an I and Q Data Transfer Function module **1716**. In an embodiment, I and Q Data Transfer Function module **1716** samples signal **1710** according to a sample clock **1212**. I and Q information signal **1710** includes baseband I and Q information.

[0375] In an embodiment, I and Q Data Transfer Function module **1716** processes information signal **1710** to generate information signals **1720**, **1722**, **1724**, and **1726**. The operation of I and Q Data Transfer Function module **1716** is further described below in section 3.4.

[0376] Referring to FIG. **17**, information signal **1720** includes vector modulator **1750** quadrature amplitude information that is processed through DAC **1730** to generate signal **1740**. Information signal **1722** includes vector modulator **1750** in-phase amplitude information that is processed through DAC **1732** to generate signal **1742**. Signals **1740** and **1742** are calculated to generate a substantially constant envelope signal **1754**. With reference to FIG. **14**, for example, information signals **1720** and **1722** include the upper quadrature and in-phase components U_y and U_x, respectively.

[0377] Still referring to FIG. **17**, information signal **1726** includes vector modulator **1752** quadrature amplitude information that is processed through DAC **1736** to generate signal **1746**. Information signal **1724** includes vector modulator **1752** in-phase amplitude information that is processed through DAC **1734** to generate signal **1744**. Signals **1744** and **1746** are calculated to generate a substantially constant envelope signal **1756**. With reference to FIG. **14**, for example, information signals **1724** and **1726** include the lower in-phase and quadrature components L_x and L_y, respectively.

[0378] In the exemplary embodiment of FIG. **17**, information signals **1720**, **1722**, **1724** and **1726** are digital signals. Accordingly, each of signals **1720**, **1722**, **1724** and **1726** is fed into a corresponding digital-to-analog converter (DAC) **1730**, **1732**, **1734**, and **1736**. The resolution and sample rates of DACs **1730**, **1732**, **1734**, and **1736** are selected according to the specific desired signaling schemes. DACs **1730**, **1732**, **1734**, and **1736** are controlled by DAC clock signals **1721**, **1723**, **1725**, and **1727**, respectively. DAC clock signals **1721**, **1723**, **1725**, and **1727** may be derived from a same clock or may be independent of each other.

[0379] In other embodiments, information signals **1720**, **1722**, **1724** and **1726** are generated in analog format and no DACs are required.

[0380] Referring to FIG. **17**, DACs **1730**, **1732**, **1734**, and **1736** convert digital information signals **1720**, **1722**, **1724**, and **1726** into corresponding analog signals, and input these analog signals into optional interpolation filters **1731**, **1733**, **1735**, and **1737**, respectively. Interpolation filters **1731**, **1733**, **1735**, and **1737**, which also serve as anti-aliasing filters, shape the DACs output signals to produce the desired output waveform. Interpolation filters **1731**, **1733**, **1735**, and **1737** generate signals **1740**, **1742**, **1744**, and **1746**, respectively.

[0381] Still referring to FIG. **17**, signals **1740**, **1742**, **1744**, and **1746** are input into vector modulators **1750** and **1752**. Vector modulators **1750** and **1752** generate first and second constant envelope constituents. In the embodiment of FIG. **17**, channel clock **1714** is set according to a desired output signal frequency to thereby establish the frequency of the output signal **1770**.

[0382] Referring to FIG. **17**, vector modulator **1750** combines signal **1740**, multiplied with a 90° shifted version of channel clock signal **1714**, and signal **1742**, multiplied with a 0° shifted version of channel clock signal **1714**, to generate output signal **1754**. In parallel, vector modulator **1752** combines signal **1746**, multiplied with a 90° shifted version of channel clock signal **1714**, and signal **1744**, multiplied with a 0° shifted version of channel clock signal **1714**, to generate output signal **1756**.

[0383] Output signals **1754** and **1756** represent constant envelope signals. A sum of output signals **1754** and **1756** results in a carrier signal having the I and Q characteristics of the original baseband signal. In embodiments, to generate a desired power level at the output of vector power amplifier **1700**, signals **1754** and **1756** are amplified and then summed. In the embodiment of FIG. **17**, for example, signals **1754** and **1756** are, respectively, input into corresponding power amplifiers (PAs) **1760** and **1762**. In an embodiment, PAs **1760** and **1762** include switching power amplifiers. Autobias circuitry **1718** controls the bias of PAs **1760** and **1762**. In the embodiment of FIG. **17**, for example, autobias circuitry **1718** provides a bias voltage **1728** to PAs **1760** and **1762**.

[0384] In an embodiment, PAs **1760** and **1762** apply equal or substantially equal power amplification to respective constant envelope signals **1754** and **1756**. In an embodiment, the power amplification is set according to the desired output power level. In other embodiments of vector power amplifier **1700**, PA drivers are additionally employed to provide additional power amplification capability to the amplifier. In the embodiment of FIG. **17**, for example, PA drivers **1774** and **1776** are optionally added, respectively, between vector modulators **1750** and **1752** and subsequent PAs **1760** and **1762**.

[0385] Respective output signals **1764** and **1766** of PAs **1760** and **1762** are substantially constant envelope signals. In the embodiment of FIG. **17**, output signals **1764** and **1766** are coupled together to generate output signal **1770** of vector power amplifier **1700**. In embodiments, it is noted that the outputs of PAs **1760** and **1762** are directly coupled. Direct coupling in this manner means that there is minimal or no resistive, inductive, or capacitive isolation between the outputs of PAs **1760** and **1762**. In other words, outputs of PAs **1760** and **1762** are coupled together without intervening components. Alternatively, in an embodiment, the outputs of PAs **1760** and **1762** are coupled together indirectly through inductances and/or capacitances that result in low or minimal impedance connections, and/or connections that result in minimal isolation and minimal power loss. Alternatively, outputs of PAs **1760** and **1762** are coupled using well known combining techniques, such as Wilkinson, hybrid couplers, transformers, or known active combiners. In an embodiment, the PAs **1760** and **1762** provide integrated amplification and power combining in a single operation. In an embodiment, one or more of the power amplifiers and/or drivers described herein are implemented using multiple input, single output (MISO) power amplification techniques, examples of which are shown in FIGS. **17**A, **17**B, and **51**A-H.

[0386] Output signal **1770** represents a signal having the desired I and Q characteristics of the baseband signal and the desired output power level and frequency. In embodiments of vector power amplifier **1700**, a pull-up impedance **1778** is coupled between the output of vector power amplifier **1700** and a power supply. In other embodiments, an impedance matching network **1780** is coupled at the output of vector power amplifier **1700**. Output stage embodiments according to power amplification methods and systems of the present invention will be further described below in section 3.5.

[0387] In other embodiments of vector power amplifier **1700**, process detectors are employed to compensate for any process and/or temperature variations in circuitry of the amplifier. In the exemplary embodiment of FIG. **17**, for example, process detector **1772** is optionally added to monitor variations in PA drivers **1774** and **1776**.

[0388] FIG. **17**A is a block diagram that illustrates another exemplary embodiment **1700**A of a vector power amplifier implementing process flowchart **1600**. Optional components are illustrated with dashed lines, although other embodiments may have more or less optional components. Embodiment **1700**A illustrates a multiple-input single-output (MISO) implementation of the amplifier of FIG. **17**. In the embodiment of FIG. **17**A, constant envelope signals **1754** and **1756**, output from vector modulators **1750** and **1760**, are input into MISO PA **1790**. MISO PA **1790** is a two-input single-output power amplifier. In an embodiment, MISO PA **1790** include elements **1760**, **1762**, **1772**, **1774**, and **1776** as shown in the embodiment of FIG. **17**, or functional equivalents thereof. In another embodiment, MISO PA **1790** may include other elements, such as pre-drivers, not shown in the embodiment of FIG. **17**. Further, MISO PA **1790** is not limited to being a two-input PA as shown in FIG. **17**A. In other embodiments, as will be described further below with reference to FIGS. **51**A-H, PA **1790** can have any number of inputs.

[0389] In another embodiment of embodiment **1700**, shown as embodiment **1700**B of FIG. **17**B, optional Autobias circuitry **1218** generates separate bias control signals **1715**, **1717**, and **1719**, corresponding to Bias A, Bias B, and Bias C, respectively. Signals **1715**, **1717**, and **1719** may or may not be generated separately within Autobias circuitry **1718**, but are output separately as shown. Further, signals **1715**, **1717**, and **1719** may or may not be related as determined by the biasing required for the different stages of MISO PA **1790**.

[0390] FIG. **18** is a block diagram that illustrates another exemplary embodiment **1800** of a vector power amplifier according to the Direct Cartesian 2-Branch VPA embodiment of FIG. **16**. Optional components are illustrated with dashed lines, although other embodiments may have more or less optional components.

[0391] In the exemplary embodiment of FIG. **18**, a DAC **1820** of sufficient resolution and sample rate replaces DACs **1730**, **1732**, **1734**, and **1736** of the embodiment of FIG. **17**. DAC **1820** is controlled by a DAC clock **1814**.

[0392] DAC **1820** receives information signal **1810** from I and Q Data Transfer Function module **1716**. Information signal **1810** includes identical information content to signals **1720**, **1722**, **1724**, and **1726** in the embodiment of FIG. **17**.

[0393] DAC **1820** may output a single analog signal at a time. Accordingly, a sample-and-hold architecture may be used as shown in FIG. **18**.

[0394] In the embodiment of FIG. **18**, DAC **1820** sequentially outputs analog signals **1822**, **1824**, **1826**, and **1828** to sample-and-hold circuits **1832**, **1834**, **1836**, and **1838**, respectively. In an embodiment, DAC **1820** is of sufficient resolution and sample rate to replace DACs **1720**, **1722**, **1724**, and **1726** of the embodiment of FIG. **17**. An output selector **1812** determines which of output signals **1822**, **1824**, **1826**, and **1828** are selected for output.

[0395] DAC **1820**'s DAC clock signal **1814**, output selector signal **1812**, and sample-and-hold clocks **1830**A-D, and **1840** are controlled by a control module that can be independent or integrated into transfer function module **1716**.

[0396] In an embodiment, sample-and-hold circuits **1832**, **1834**, **1836**, and **1838** sample and hold their respective values and, according to a clock signal **1830**A-D, release the values to a second set of sample-and-hold circuits **1842**, **1844**, **1846**, and **1848**. For example, S/H **1832** release's its value to S/H **1842** according to a received clock signal **1830**A. In turn, sample-and-hold circuits **1842**, **1844**, **1846**, and **1848** hold the received analog values, and simultaneously release the values to interpolation filters **1852**, **1854**, **1856**, and **1858** according to a common clock signal **1840**.

[0397] In another embodiment, a single set of S/H circuitry that includes S/H **1832**, **1834**, **1836**, and **1838** can be employed. Accordingly, S/H circuits **1832**, **1834**, **1836**, and **1838** receive analog values from DAC **1820**, and each samples and holds its received value according to independent clocks **1830**A-D. For example, S/H **1832** is controlled by clock **1830**A, which may not be synchronized with clock **1830**B that controls S/H **1834**. For example, DAC **1820** outputs signals **1822**, **1824**, **1826**, and **1828** with appropriately selected analog values calculated by transfer function module **1716** to S/H circuits **1832**, **1834**, **1836**, and **1838** in order to compensate for the time differences between clocks **1830**A-D.

[0398] Other aspects of vector power amplifier **1800** correspond substantially to those described above with respect to vector power amplifier **1700**.

[0399] FIG. **18**A is a block diagram that illustrates another exemplary embodiment **1800**A of a vector power amplifier according to the Direct Cartesian 2-Branch VPA embodiment. Optional components are illustrated with dashed lines, although in other embodiments more or less components may be optional. Embodiment **1800**A is a Multiple Input Single Output (MISO) implementation of embodiment **1800** of FIG. **18**.

[0400] In the embodiment of FIG. **18**A, constant envelope signals **1754** and **1756**, output from vector modulators **1750** and **1752**, are input into MISO PA **1860**. MISO PA **1860** is a two-input single-output power amplifier. In an embodiment, MISO PA **1860** includes elements **1744**, **1746**, **1760**, **1762**, and **1772** as shown in the embodiment of FIG. **18**, or functional equivalents thereof. In another embodiment, MISO PA **1860** may include other elements, such as predrivers, not shown in the embodiment of FIG. **17**. Further, MISO PA **1860** is not limited to being a two-input PA as shown in FIG. **18**A. In other embodiments as will be described further below with reference to FIGS. **51**A-H, PA **1860** can have any number of inputs.

[0401] The embodiment of FIG. **18**A further illustrates two different sample and hold architectures with a single or two levels of S/H circuitry as shown. The two implementations have been described above with respect to FIG. **18**.

[0402] Other aspects of vector power amplifier **1800**A are substantially equivalent to those described above with respect to vector power amplifiers **1700** and **1800**.

3.4) I and Q Data to Vector Modulator Transfer Functions

[0403] In some of the above described embodiments, I and Q data transfer functions are provided to transform received I and Q data into amplitude information inputs for subsequent stages of vector modulation and amplification. For example, in the embodiment of FIG. **17**, I and Q Data Transfer Function module **1716** processes I and Q information signal **1710** to generate in-phase and quadrature amplitude information signals **1720**, **1722**, **1724**, and **1726** of first and second constant envelope constituents **1754** and **1756** of signal r(t). Subsequently, vector modulators **1750** and **1752**

utilize the generated amplitude information signals **1720**, **1722**, **1724**, and **1726** to create the first and second constant envelope constituent signals **1754** and **1756**. Other examples include modules **710**, **712**, and **1216** in FIGS. **7**, **8**, **12**, and **13**. These modules implement transfer functions to transform I and/or Q data into amplitude information inputs for subsequent stages of vector modulation and amplification.

[0404] According to the present invention, I and Q Data Transfer Function modules may be implemented using digital circuitry, analog circuitry, software, firmware or any combination thereof.

[0405] Several factors affect the actual implementation of a transfer function according to the present invention, and vary from embodiment to embodiment. In one aspect, the selected VPA embodiment governs the amplitude information output of the transfer function and associated module. It is apparent, for example, that I and Q Data Transfer Function module **1216** of the CPCP 2-Branch VPA embodiment **1200** differs in output than I and Q Data Transfer Function module **1716** of the Direct Cartesian 2-Branch VPA embodiment **1700**.

[0406] In another aspect, the complexity of the transfer function varies according to the desired modulation scheme (s) that need to be supported by the VPA implementation. For example, the sample clock, the DAC sample rate, and the DAC resolution are selected in accordance with the appropriate transfer function to construct the desired output waveform(s).

[0407] According to the present invention, transfer function embodiments may be designed to support one or more VPA embodiments with the ability to switch between the supported embodiments as desired. Further, transfer function embodiments and associated modules can be designed to accommodate a plurality of modulation schemes. A person skilled in the art will appreciate, for example, that embodiments of the present invention may be designed to support a plurality of modulation schemes (individually or in combination) including, but not limited to, BPSK, QPSK, OQPSK, DPSK, CDMA, WCDMA, W-CDMA, GSM, EDGE, MPSK, MQAM, MSK, CPSK, PM, FM, OFDM, and multi-tone signals. In an embodiment, the modulation scheme(s) may be configurable and/or programmable via the transfer function module.

3.4.1) Cartesian 4-Branch VPA Transfer Function

[0408] FIG. **19** is a process flowchart **1900** that illustrates an example I and Q transfer function embodiment according to the Cartesian 4-Branch VPA embodiment. The process begins at step **1910**, which includes receiving an in-phase data component and a quadrature data component. In the Cartesian 4-Branch VPA embodiment of FIG. **7**A, for example, this is illustrated by I Data Transfer Function module **710** receiving I information signal **702**, and Q Data Transfer Function module **712** receiving Q information signal **704**. It is noted that, in the embodiment of FIG. **7**A, I and Q Data Transfer Function modules **710** and **712** are illustrated as separate components. In implementation, however, I and Q Data Transfer Function modules **710** and **712** may be separate or combined into a single module.

[0409] Step **1920** includes calculating a phase shift angle between first and second substantially equal and constant envelope constituents of the I component. In parallel, step **1920** also includes calculating a phase shift angle between first and second substantially equal and constant envelope constituents of the Q component. As described above, the first and second constant envelope constituents of the I components are appropriately phased relative to the I component. Similarly, the first and second constant envelope constituents of the Q components are appropriately phased relative to the Q component. In the embodiment of FIG. **7**A, for example, step **1920** is performed by I and Q Data Transfer Function modules **710** and **712**.

[0410] Step **1930** includes calculating in-phase and quadrature amplitude information associated with the first and second constant envelope constituents of the I component. In parallel, step **1930** includes calculating in-phase and quadrature amplitude information associated with the first and second constant envelope constituents of the Q component. In the embodiment of FIG. **7**A, for example, step **1930** is performed by and I and Q Data Transfer Function modules **710** and **712**.

[0411] Step **1940** includes outputting the calculated amplitude information to a subsequent vector modulation stage. In the embodiment of FIG. **7**A, for example, I and Q Transfer Function modules **710** and **712** output amplitude information signals **722**, **724**, **726**, and **728** to vector modulators **760**, **762**, **764**, and **766** through DACs **730**, **732**, **734**, and **736**.

[0412] FIG. **20** is a block diagram that illustrates an exemplary embodiment **2000** of a transfer function module, such as transfer function modules **710** and **712** of FIG. **7**A, implementing the process flowchart **1900**. In the example of FIG. **20**, transfer function module **2000** receives I and Q data signals **2010** and **2012**. In an embodiment, I and Q data signals **2010** and **2012** represent I and Q data components of a baseband signal, such as signals **702** and **704** in FIG. **7**A.

[0413] Referring to FIG. **20**, in an embodiment, transfer function module **2000** samples I and Q data signals **2010** and **2012** according to a sampling clock **2014**. Sampled I and Q data signals are received by components **2020** and **2022**, respectively, of transfer function module **2000**. Components **2020** and **2022** measure, respectively, the magnitudes of the sampled I and Q data signals. In an embodiment, components **2020** and **2022** are magnitude detectors.

[0414] Components **2020** and **2022** output the measured I and Q magnitude information to components **2030** and **2032**, respectively, of transfer function module **2000**. In an embodiment, the measured I and Q magnitude information is in the form of digital signals. Based on the I magnitude information, component **2030** calculates a phase shift angle ϕ_I between first and second equal and constant or substantially equal and constant envelope constituents of the sampled I signal. Similarly, based on the Q magnitude information, component **2032** calculates phase shift angle ϕ_Q between a first and second equal and constant or substantially equal and constant envelope constituents of the sampled Q signal. This operation shall now be further described.

[0415] In the embodiment of FIG. **20**, ϕ_I and ϕ_Q are illustrated as functions $f(|\vec{I}|)$ and $f(|\vec{Q}|)$ of the I and Q magnitude signals. In embodiments, functions $f(|\vec{I}|)$ and $f(|\vec{Q}|)$ are set according to the relative magnitudes of the baseband I and Q signals respectively. $f(|\vec{I}|)$ and $f(|\vec{Q}|)$ according to embodiments of the present invention will be further described below in section 3.4.4.

[0416] Referring to FIG. **20**, components **2030** and **2032** output the calculated phase shift information to components **2040** and **2042**, respectively. Based on phase shift angle ϕ_I, component **2040** calculates in-phase and quadrature amplitude information of the first and second constant envelope constituents of the sampled I signal. Similarly, based on phase shift angle ϕ_Q, component **2042** calculates in-phase and quadrature amplitude information of the first and second constant envelope constituents of the sampled Q signal. Due to symmetry, in embodiments of the invention, calculation is required for 4 values only. In the example of FIG. **20**, the values are illustrated as $\mathrm{sgn}(I)\times I_{UX}$, I_{UY}, Q_{UX}, and $\mathrm{sgn}(Q)\times Q_{UY}$, as provided in FIG. **5**.

[0417] Components **2040** and **2042** output the calculated amplitude information to subsequent stages of the vector power amplifier. In embodiments, each of the four calculated values is output separately to a digital-to-analog converter. As shown in the embodiment of FIG. **7**A for example, signals **722**, **724**, **726**, and **728** are output separately to DACs **730**, **732**, **734**, and **736**, respectively. In other embodiments, signals **722**, **724**, **726**, and **728** are output into a single DAC as shown in FIGS. **800**A and **800**B.

3.4.2) CPCP 2-Branch VPA Transfer Function

[0418] FIG. **21** is a process flowchart **2100** that illustrates an example I and Q transfer function embodiment according to the CPCP 2-Branch VPA embodiment. The process begins at step **2110**, which includes receiving in-phase (I) and quadrature (Q) data components of a baseband signal. In the CPCP 2-Branch VPA embodiment of FIG. **12**, for example, this is illustrated by I and Q Data Transfer Function module **1216** receiving I and Q information signal **1210**.

[0419] Step **2120** includes determining the magnitudes $|I|$ and $|Q|$ of the received I and Q data components.

[0420] Step **2130** includes calculating a magnitude $|R|$ of the baseband signal based on the measured $|I|$ and $|Q|$ magnitudes. In an embodiment, $|R|$ is such that $|R|^2=|I|^2+|Q|^2$. In the embodiment of FIG. **12**, for example, steps **2120** and **2130** are performed by I and Q Data Transfer Function module **1216** based on received information signal **1210**.

[0421] Step **2140** includes normalizing the measured $|I|$ and $|Q|$ magnitudes. In an embodiment, $|I|$ and $|Q|$ are normalized to generate an Iclk_phase and Q_{clk}_phase signals (as shown in FIG. **10**) such that $|I_{clk_phase}|^2+|Q_{clk_phase}|^2=$constant. In the embodiment of FIG. **12**, for example, step **2140** is performed by I and Q Data Transfer Function module **1216** based on received information signal **1210**.

[0422] Step **2150** includes calculating in-phase and quadrature amplitude information associated with first and second constant envelope constituents. In the embodiment of FIG. **12**, for example, step **2150** is performed by I and Q Data Transfer Function module **1216** based on the envelope magnitude $|R|$.

[0423] Step **2160** includes outputting the generated Iclk_phase and Qclk_phase (from step **2140**) and the calculated amplitude information (from step **2150**) to appropriate vector modulators. In the embodiment of FIG. **12**, for example, I and Q Data Transfer Function module **1216** output information signals **1220**, **1222**, **1224**, and **1226** to vector modulators **1238**, **1260**, and **1262** through DACs **1230**, **1232**, **1234**, and **1236**.

[0424] FIG. **22** is a block diagram that illustrates an exemplary embodiment **2200** of a transfer function module (such as module **1216** of FIG. **12**) implementing the process flowchart **2100**. In the example of FIG. **22**, transfer function module **2200** receives I and Q data signal **2210**. In an embodiment, I and Q data signal **2210** includes I and Q components of a baseband signal, such as signal **1210** in the embodiment of FIG. **12**, for example.

[0425] In an embodiment, transfer function module **2200** samples I and Q data signal **2210** according to a sampling clock **2212**. Sampled I and Q data signals are received by component **2220** of transfer function module **2200**. Component **2220** measures the magnitudes $|\vec{I}|$ and $|\vec{Q}|$ of the sampled I and Q data signals.

[0426] Based on the measured $|\vec{I}|$ and $|\vec{Q}|$ magnitudes, component **2230** calculates the magnitude $|R|$ of the baseband signal. In an embodiment, $|\vec{R}|$ is such that $|\vec{R}|^2=|\vec{I}|^2+|\vec{Q}|^2$.

[0427] In parallel, component **2240** normalizes the measured $|\vec{I}|$ and $|\vec{Q}|$ magnitudes. In an embodiment, $|\vec{I}|$ and $|\vec{Q}|$ are normalized to generate Iclk_phase and Qclk_phase signals such that $|\text{Iclk_phase}|^2+|\text{Qclk_phase}|^2=$constant, where |Iclk_phase| and |Qclk_phase| represent normalized magnitudes of $|\vec{I}|$ and $|\vec{Q}|$. Typically, given that the constant has a value A, the measured $|\vec{I}|$ and $|\vec{I}|$ magnitudes are both divided by the quantity

$$\frac{A}{\sqrt{|\vec{I}|^2+|\vec{Q}|^2}}$$

[0428] Component **2250** receives the calculated $|\vec{R}|$ magnitude from component **2230**, and based on it calculates a phase shift angle ϕ between first and second constant envelope constituents. Using the calculated phase shift angle ϕ, component **2050** then calculates in-phase and quadrature amplitude information associated with the first and second constant envelope constituents.

[0429] In the embodiment of FIG. **22**, the phase shift angle ϕ is illustrated as a function $f(|\vec{R}|)$ of the calculated magnitude $|\vec{R}|$.

[0430] Referring to FIG. **22**, components **2240** and **2250** output the normalized |Iclk_phase| and |Qclk_phase| magnitude information and the calculated amplitude information to DAC's for input into the appropriate vector modulators. In embodiments, the output values are separately output to digital-to-analog converters. As shown in the embodiment of FIG. **12**, for example, signals **1220**, **1222**, **1224**, and **1226** are output separately to DACs **1230**, **1232**, **1234**, and **1236**, respectively. In other embodiments, signals **1220**, **1222**, **1224**, and **1226** are output into a single DAC as shown in FIGS. **13** and **13**A.

3.4.3) Direct Cartesian 2-Branch Transfer Function

[0431] FIG. **23** is a process flowchart **2300** that illustrates an example I and Q transfer function embodiment according to the Direct Cartesian 2-Branch VPA embodiment. The process begins at step **2310**, which includes receiving in-phase (I) and quadrature (Q) data components of a baseband signal. In the Direct Cartesian 2-Branch VPA embodiment of

FIG. **17**, for example, this is illustrated by I and Q Data Transfer Function module **1716** receiving I and Q information signal **1710**.

[0432] Step **2320** includes determining the magnitudes |I| and |Q| of the received I and Q data components.

[0433] Step **2330** includes calculating a magnitude |R| of the baseband signal based on the measured |I| and |Q| magnitudes. In an embodiment, |R| is such that $|R|^2=|I|^2+|Q|^2$. In the embodiment of FIG. **17**, for example, steps **2320** and **2330** are performed by I and Q Data Transfer Function module **1716** based on received information signal **1710**.

[0434] Step **2340** includes calculating a phase shift angle θ of the baseband signal based on the measured |I| and |Q| magnitudes. In an embodiment, θ is such that

$$\theta=\tan^{-1}\left(\frac{|Q|}{|I|}\right),$$

and wherein the sign of I and Q determine the quadrant of θ. In the embodiment of FIG. **17**, for example, step **2340** is performed by I and Q Data Transfer Function module **1216** based on I and Q data components received in information signal **1210**.

[0435] Step **2350** includes calculating in-phase and quadrature amplitude information associated with a first and second constant envelope constituents of the baseband signal. In the embodiment of FIG. **17**, for example, step **2350** is performed by I and Q Data Transfer Function module **1716** based on previously calculated magnitude |R| and phase shift angle θ.

[0436] Step **2360** includes outputting the calculated amplitude information to DAC's for input into the appropriate vector modulators. In the embodiment of FIG. **17**, for example, I and Q Data Transfer Function module **1716** output information signals **1720**, **1722**, **1724**, and **1726** to vector modulators **1750** and **1752** through DACs **1730**, **1732**, **1734**, and **1736**. In other embodiments, signals **1720**, **1722**, **1724**, and **1726** are output into a single DAC as shown in FIGS. **18** and **18**A.

[0437] FIG. **24** is a block diagram that illustrates an exemplary embodiment **2400** of a transfer function module implementing the process flowchart **2300**. In the example of FIG. **24**, transfer function module **2400** (such as transfer function module **1716**) receives I and Q data signal **2410**, such as signal **1710** in FIG. **17**. In an embodiment, I and Q data signal **2410** includes I and Q data components of a baseband signal.

[0438] In an embodiment, transfer function module **2400** samples I and Q data signal **2410** according to a sampling clock **2412**. Sampled I and Q data signals are received by component **2420** of transfer function module **2200**. Component **2420** measures the magnitudes $|\vec{I}|$ and $|\vec{Q}|$ of the sampled I and Q data signals.

[0439] Based on the measured $|\vec{I}|$ and $|\vec{Q}|$ magnitudes, component **2430** calculates the magnitude $|\vec{R}|$. In an embodiment, $|\vec{R}|$ is such that $|\vec{R}|^2=|\vec{I}|^2+|\vec{Q}|^2$.

[0440] In parallel, component **2240** calculates the phase shift angle θ of the baseband signal. In an embodiment, θ is such that

$$\theta=\tan^{-1}\left(\frac{|\vec{Q}|}{|\vec{I}|}\right),$$

where the sign of I and Q determine the quadrant of θ.

[0441] Component **2450** receives the calculated $|\vec{R}|$ magnitude from component **2430**, and based on it calculates a phase shift angle ϕ between first and second constant envelope constituent signals. In the embodiment of FIG. **24**, the phase shift angle ϕ is illustrated as a function $f_3|\vec{R}|)$ of the calculated magnitude $|\vec{R}|$. This is further described in section 3.4.4.

[0442] In parallel, component **2450** receives the calculated phase shift angle θ from component **2440**. As functions of ϕ and θ, component **2450** then calculates in-phase and quadrature amplitude information for the vector modulator inputs that generate the first and second constant envelope constituents. In an embodiment, the in-phase and quadrature amplitude information supplied to the vector modulators are according to the equations provided in (18).

[0443] Component **2450** outputs the calculated amplitude information to subsequent stages of the vector power amplifier. In embodiments, the output values are separately output to digital-to-analog converters. As shown in the embodiment of FIG. **17**, for example, signals **1720**, **1722**, **1724**, and **1726** are output separately to DACs **1730**, **1732**, **1734**, and **1736**, respectively. In other embodiments, signals **1720**, **1722**, **1724**, and **1726** are output into a single DAC as shown in FIGS. **18** and **18**A.

3.4.4) Magnitude to Phase Shift Transform

[0444] Embodiments of f(|I|), f(|Q|) of FIG. **20** and f(|R|) of FIGS. **22** and **24** shall now be further described.

[0445] According to the present invention, any periodic waveform that can be represented by a Fourier series and a Fourier transform can be decomposed into two or more constant envelope signals.

[0446] Below are provided two examples for sinusoidal and square waveforms.

3.4.4.1) Magnitude to Phase Shift Transform for Sinusoidal Signals

[0447] Consider a time-varying complex envelope sinusoidal signal r(t). In the time domain, it can be represented as:

$$r(t)=R(t)\sin(\omega t+\delta(t)) \tag{20}$$

where R(t) represents the signal's envelope magnitude at time t, δ(t) represents the signal's phase shift angle at time t, and ω represents the signal's frequency in radians per second.

[0448] It can be verified that, at any time instant t, signal r(t) can be obtained by the sum of two appropriately phased equal and constant or substantially equal and constant envelope signals. In other words, it can be shown that:

$$R(t)\sin(\omega t+\delta(t))=A\ \sin(\omega t)+A\ \sin(\omega t+\phi(t)) \tag{21}$$

for an appropriately chosen phase shift angle ϕ(t) between the two constant envelope signals. The phase shift angle ϕ(t) will be derived as a function of R(t) in the description below. This is equivalent to the magnitude to phase shift transform for sinusoidal signals.

[0449] Using a sine trigonometric identity, equation (21) can be re-written as:

$$R(t)\sin(\omega t+\delta(t))=A\ \sin(\omega t)+A\ \sin(\omega t))\cos\ \phi t)+\mathrm{A}\ \sin(\phi(t))\cos\ \omega t;\ \Rightarrow R(t)\sin(\omega t+\delta(t))=A\ \sin(\phi(t))\cos\ \omega t+A(1+\cos\ \phi(t))\sin\ \omega t. \tag{22}$$

[0450] Note, from equation (22), that signal r(t) is written as a sum of an in-phase component and a quadrature component. Accordingly, the envelope magnitude R(t) can be written as:

$$R(t)=\sqrt{(A\sin(\phi(t)))^2+(A(1+\cos(\phi(t))))^2};\ \Rightarrow R(t)=\sqrt{2A(A+\cos(\phi(t)))}. \tag{23}$$

[0451] Equation (23) relates the envelope magnitude R(t) of signal r(t) to the phase shift angle φ(t) between two constant envelope constituents of signal r(t). The constant envelope constituents have equal or substantially equal envelope magnitude A, which is typically normalized to 1.

[0452] Inversely, from equation (23), the phase shift angle φ(t) can be written as a function of R(t) as follows:

$$\phi(t)=\arccos\left(\frac{R(t)^2}{2A^2}-1\right). \tag{24}$$

[0453] Equation (24) represents the magnitude to phase shift transform for the case of sinusoidal signals, and is illustrated in FIG. **26**.

3.4.4.2) Magnitude to Phase Shift Transform for Square Wave Signals:

[0454] FIG. **28** illustrates a combination of two constant envelope square wave signals according to embodiments of the present invention. In FIG. **28**, signals **2810** and **2820** are constant envelope signals having a period T, a duty cycle γT (0<γ<1), and envelope magnitudes A1 and A2, respectively.

[0455] Signal **2830** results from combining signals **2810** and **2820**. According to embodiments of the present invention, signal **2830** will have a magnitude equal or substantially equal to a product of signals **2810** and **2820**. In other words, signal **2830** will have a magnitude of zero whenever either of signals **2810** or **2820** has a magnitude of zero, and a non-zero magnitude when both signals **2810** and **2820** have non-zero magnitudes.

[0456] Further, signal **2830** represents a pulse-width-modulated signal. In other words, the envelope magnitude of signal **2830** is determined according to the pulse width of signal **2830** over one period of the signal. More specifically, the envelope magnitude of signal **2830** is equal or substantially to the area under the curve of signal **2830**.

[0457] Referring to FIG. **28**, signals **2810** and **2820** are shown time-shifted relative to each other by a time shift t'. Equivalently, signals **2810** and **2820** are phase-shifted relative to each other by a phase shift

angles $\phi=\left(\frac{t'}{T}\right)\times 2\pi$ radians.

[0458] Still referring to FIG. **28**, note that the envelope magnitude R of signal **2830**, in FIG. **28**, is given by:

$$R=A_1\times A_2\times(\gamma T-t') \tag{25}$$

[0459] Accordingly, it can be deduced that φ is related to R according to:

$$\phi=\left[\gamma-\frac{R}{T(A_1A_2)}\right]\times(2\pi). \tag{26}$$

[0460] Note, from equation (26), that R is at a maximum of γA1A2 when φ=0. In other words, the envelope magnitude is at a maximum when the two constant envelope signals are in-phase with each other.

[0461] In typical implementations, signals **2810** and **2820** are normalized and have equal or substantially equal envelope magnitude of 1. Further, signals **2810** and **2820** typically have a duty cycle of 0.5. Accordingly, equation (26) reduces to:

$$\phi=\left[0.5-\frac{R}{T}\right]\times(2\pi). \tag{27}$$

[0462] Equation (27) illustrates the magnitude to phase shift transform for the case of normalized and equal or substantially equal envelope magnitude square wave signals. Equation (27) is illustrated in FIG. **26**.

3.4.5) Waveform Distortion Compensation

[0463] In certain embodiments, magnitude to phase shift transforms may not be implemented exactly as theoretically or practically desired. In fact, several factors may exist that require adjustment or tuning of the derived magnitude to phase shift transform for optimal (or at least improved) operation. In practice, phase and amplitude errors may exist in the vector modulation circuitry, gain and phase imbalances can occur in the vector power amplifier branches, and distortion may exist in the MISO amplifier itself including but not limited to errors introduced by directly combining at a single circuit node transistor outputs within the MISO amplifier described herein. Each of these factors either singularly or in combination will contribute to output waveform distortions that result in deviations from the desired output signal r(t). When output waveform distortion exceeds system design requirements, waveform distortion compensation may be required.

[0464] FIG. **25** illustrates the effect of waveform distortion on a signal using phasor signal representation. In FIG. **25**, $\vec{R}$ represents a phasor representation of a desired signal r(t). In the example of FIG. **25**, waveform distortion can cause the actual output phasor to vary from r(t) anywhere within the phasor error region. An exemplary phasor error region is illustrated in FIG. **25**, and is equal or substantially equal to the maximum error vector magnitude. Phasors $\vec{R}_1$ and $\vec{R}_2$ represent examples of potential output phasors that deviate from the desired r(t).

[0465] According to embodiments of the present invention, waveform distortions can be measured, calculated, or estimated during the manufacture of the system and/or in real time or non-real time operation. FIG. **54**A and FIG. **55** are examples of methods that can be used for phasor error measurement and correction. These waveform distortions can be compensated for or reduced at various points in the system. For example, a phase error between the branch amplifiers can be adjusted by applying an analog voltage offset to the vector modulation circuitry, within the transfer function, and/or using real time or non-real time feedback techniques as shown in the example system illustrated in FIGS. **58**, **59** and **60**. Similarly, branch amplification imbalances can be adjusted by applying an analog voltage offset to the vector modulation circuitry, within the transfer function, and/or using real time or non-real time feedback techniques as shown in FIGS. **58**, **59** and **60**. In the system illustrated in FIGS. **58**, **59** and **60**, for example, waveform distortion adjustment is performed, as illustrated in FIG. **60**, using Differential Branch Amplitude Measurement Circuitry **6024** and Differential Branch Phase Measurement Circuitry **6026**, which provide a Differential Branch Amplitude signal **5950** and a Differential Branch Phase signal **5948**, respectively. These signals are input into an A/D Converter **5732** by input signal selector **5946**, with the values generated by AID converter **5732** being input into Digital Control Module **5602**. Digital Control Module **5602** uses the values generated by A/D converter **5732** to calculate adjusted or offset values to provide control voltages for phase adjustments to Vector modulation circuitry **5922**, **5924**, **5926**, and **5928** and control voltages for amplitude adjustments to Gain Balance control circuitry **6016**. In FIG. **58**, these control voltages are illustrated using Gain Balance Control signal **5749** and Phase Balance Control signal **5751**. The feedback approach described above also compensates for process variations, temperature variations, IC package variations, and circuit board variations by ensuring the system amplitude and phase errors remain with a specified tolerance. Additional example feedback and feedforward error measurement and compensation techniques are further described in section 4.1.2.

[0466] In other embodiments, the measured, calculated, or estimated waveform distortions are compensated for at the transfer function stage of the power amplifier. In this approach, the transfer function is designed to factor in and correct the measured, calculated, and/or estimated waveform distortions. FIG. **78** illustrates a mathematical derivation of the magnitude to phase shift transform in the presence of amplitude and phase errors in branches of the VPA. Equation (28) in FIG. **78** takes into account both phase and amplitude errors in an exemplary embodiment. Note that $R^*\sin(\acute{\omega}^*t+\delta)$ in FIG. **78** can be representative of either $\vec{R_1}$ or $\vec{R_2}$ in FIG. **25**, for example. Equation (28) assumes that amplitudes A**1** and A**2** of the VPA branches can be different and that each branch can contain a respective phase error ϕe**1**(t) and ϕe**2**(t). For reference purposes, in a theoretically perfect system, A**1**=A**2** and ϕe**1**(t)=ϕe**2**(t)=0. δ(t) is adjusted by quadrant based on the sign value of the input vectors I(t) and Q(t). As such, with no amplitude or phase errors, the phasor corresponding to $R^*\sin(\acute{\omega}^*t+\delta)$ is aligned with the desired phasor $\vec{R}$ in FIG. **25**.

[0467] In some embodiments, in practice, amplitude and phase components of the phasor corresponding to $R^*\sin(\acute{\omega}^*t+\delta)$ are compared to the desired phasor $\vec{R}$ to generate system amplitude and phase error deviations. These amplitude and phase error deviations from the desired phasor $\vec{R}$, as shown in FIG. **25**, can be accounted for in the system transfer function. In an embodiment, A**1** and A**2** can be substantially equalized and ϕe**1**(t) and ϕe**2**(t) can be minimized by properly adjusting the control inputs to the vector modulation circuitry. In an embodiment, as illustrated in FIG. **57**, this is performed by the digital control module, which provides, using digital-to-analog converters DAC_**01**, DAC_**02**, DAC_**03**, and DAC_**04**, control inputs to the vector modulation circuitry.

[0468] Accordingly, given the fact that equations such as equation (28) can be used to calculate the resultant phasor at any instant in time based on the values of A**1** and A**2** and ϕe**1**(t) and ϕe**2**(t), transfer function modification(s) can be made to compensate for the system errors, and such transfer function modification(s) will be apparent to persons skilled in the relevant art(s) based on the teachings contained herein. Exemplary methods for generating error tables and/or mathematical functions to compensate for system errors are described in Section 4.1.2. It will be apparent to persons skilled in the relevant art(s) that these waveform distortion correction and compensation techniques can be implemented in either the digital or the analog domains, and implementation of such techniques will be apparent to persons skilled in the relevant art(s) based on the teachings contained herein.

3.5) Output Stage

[0469] An aspect of embodiments of the present invention lies in summing constituent signals at the output stage of a vector power amplifier (VPA). This is shown, for example, in FIG. **7** where the outputs of PAs **770**, **772**, **774**, and **776** are summed. This is similarly shown in FIGS. **8**, **12**, **13**, **17**, and **18**, for example. Various embodiments for combining the outputs of VPAs are described herein. While the following is described in the context of VPAs, it should be understood that the following teachings generally apply to coupling or summing the outputs of any active devices in any application.

[0470] FIG. **29** illustrates a vector power amplifier output stage embodiment **2900** according to an embodiment of the present invention. Output stage **2900** includes a plurality of vector modulator signals **2910**-{**1**, . . . , n} being input into a plurality of corresponding power amplifiers (PAs) **2920**-{**1**, . . . , n}. As described above, signals **2910**-{**1**, . . . , n} represent constituent signals of a desired output signal of the vector power amplifier.

[0471] In the example of FIG. **29**, PAs **2910**-{**1**, . . . , n} equally amplify or substantially equally amplify input signals **2910**-{**1**, . . . , n} to generate amplified output signals **2930**-{ **1**, . . . , n}. Amplified output signals **2930**-{**1**, . . . , n} are coupled together directly at summing node **2940**. According to this example embodiment of the present invention, summing node **2940** includes no coupling or isolating element, such as a power combiner, for example. In the embodiment of FIG. **29**, summing node **2940** is a zero-impedance (or near-zero impedance) conducting wire. Accordingly, unlike in conventional systems that employ combining elements, the combining of output signals according to this embodiment of the present invention incurs minimal power loss.

[0472] In another aspect, output stage embodiments of the present invention can be implemented using multiple-input single-output (MISO) power amplifiers.

[0473] In another aspect, output stage embodiments of the present invention can be controlled to increase the power efficiency of the amplifier by controlling the output stage current according to the desired output power level.

[0474] In what follows, various output stage embodiments according to VPA embodiments of the present invention are provided in section 3.5.1. In section 3.5.2, embodiments of output stage current shaping functions, for increasing the power efficiency of certain VPA embodiments of the present invention, are presented. Section 3.5.3 describes embodiments of output stage protection techniques that may be utilized for certain output stage embodiments of the present invention.

3.5.1) Output Stage Embodiments

[0475] FIG. **30** is a block diagram that illustrates a power amplifier (PA) output stage embodiment **3000** according to an embodiment of the present invention. Output stage embodiment **3000** includes a plurality of PA branches **3005**-{**1**, . . . , n}. Signals **3010**-{**1**, . . . , n} incoming from respective vector modulators represent inputs for output stage **3000**. According to this embodiment of the present invention, signals **3010**-{**1**, . . . , n} represent equal and constant or substantially equal and constant envelope constituent signals of a desired output signal of the power amplifier.

[0476] PA branches **3005**-{**1**, . . . , n} apply equal or substantially equal power amplification to respective signals **3010**-{**1**, . . . , n}. In an embodiment, the power amplification level through PA branches **3005**-{**1**, . . . , n} is set according to a power level requirement of the desired output signal.

[0477] In the embodiment of FIG. **30**, PA branches **3005**-{**1**, . . . , n} each includes a power amplifier **3040**-{**1**, . . . , n}. In other embodiments, drivers **3030**-{**1**, . . . , n} and pre-drivers **3020**-{**1**, . . . , n}, as illustrated in FIG. **30**, may also be added in a PA branch prior to the power amplifier element. In embodiments, drivers and pre-drivers are employed whenever a required output power level may not be achieved in a single amplifying stage.

[0478] To generate the desired output signal, outputs of PA branches **3005**-{**1**, . . . , n} are coupled directly at summing node **3050**. Summing node **3050** provides little or no isolation between the coupled outputs. Further, summing node **3050** represents a relatively lossless summing node. Accordingly, minimal power loss is incurred in summing the outputs of PAs **3040**-{**1**, . . . , n}.

[0479] Output signal **3060** represents the desired output signal of output stage **3000**. In the embodiment of FIG. **30**, output signal **3060** is measured across a load impedance **3070**.

[0480] FIG. **31** is a block diagram that illustrates another power amplifier (PA) output stage embodiment **3100** according to the present invention. Similar to the embodiment of FIG. **30**, output stage **3100** includes a plurality of PA branches **3105**-{**1**, . . . , n}. Each of PA branches **3105**-{**1**, . . . , n} may include multiple power amplification stages represented by a pre-driver **3020**-{**1**, . . . , n}, driver **3030**-{**1**, . . . , n}, and power amplifier **3040**-{**1**, . . . , n}. Output stage embodiment **3100** further includes pull-up impedances coupled at the output of each power amplification stage to provide biasing of that stage. For example, pull-up impedances **3125**-{**1**, . . . , n} and **3135**-{**1** , . . . , n}, respectively, couple the pre-driver and driver stage outputs to power supply or independent bias power supplies. Similarly, pull-up impedance **3145** couples the PA stage outputs to the power supply or an independent bias power supply. According to this embodiment of the present invention, pull-up impedances represent optional components that may affect the efficiency but not necessarily the operation of the output stage embodiment.

[0481] FIG. **32** is a block diagram that illustrates another power amplifier (PA) output stage embodiment **3200** according to the present invention. Similar to the embodiment of FIG. **30**, output stage **3200** includes a plurality of PA branches **3205**-{**1**, . . . , n}. Each of PA branches **3205**-{**1**, . . . , n} may include multiple power amplification stages represented by a pre-driver **3020**-{**1**, . . . , n}, driver **3030**-{**1**, . . . , n}, and power amplifier **3040**-{**1**, . . . , n}. Output stage embodiment **3200** also includes pull-up impedances coupled at the output of each power amplification stage to achieve a proper biasing of that stage. Further, output stage embodiment **3200** includes matching impedances coupled at the outputs of each power amplification stage to maximize power transfer from that stage. For example, matching impedances **3210**-{**1**, . . . , n} and **3220**-{**1**, . . . , n}, are respectively coupled to the pre-driver and driver stage outputs. Similarly, matching impedance **3240** is coupled at the PA stage output. Note that matching impedance **3240** is coupled to the PA output stage subsequent to summing node **3250**.

[0482] In the above-described embodiments of FIGS. **30**-**32**, the PA stage outputs are combined by direct coupling at a summing node. For example, in the embodiment of FIG. **30**, outputs of PA branches **3005**-{**1**, . . . , n} are coupled together at summing node **3050**. Summing node **3050** is a near zero-impedance conducting wire that provides minimal isolation between the coupled outputs. Similar output stage coupling is shown in FIGS. **31** and **32**. It is noted that in certain embodiments of the present invention, output coupling, as shown in the embodiments of FIGS. **30**-**32** or embodiments subsequently described below, may utilize certain output stage protection measures. These protection measures may be implemented at different stages of the PA branch. Further, the type of protection measures needed may be PA implementation-specific. A further discussion of output stage protection according to an embodiment of the present invention is provided in section 3.5.3.

[0483] FIG. **33** is a block diagram that illustrates another power amplifier (PA) output stage embodiment **3300** according to the present invention. Similar to the embodiment of FIG. **30**, output stage **3300** includes a plurality of PA branches **3305**-{**1**, . . . , n}. Each of PA branches **3305**-{**1**, . . . , n} may include multiple power amplification stages represented by a pre-driver **3020**-{**1**, . . . , n}, driver **3030**-{**1**, . . . , n}, and power amplifier **3040**-{**1**, . . . , n}. Output stage embodiment **3300** may also include pull-up impedances **3125**-{**1**, . . . , n}, **3135**-{**1**, . . . , n}, and **3145** coupled at the output of each power amplification stage to achieve a proper biasing of that stage. Additionally, output stage embodiment **3300** may include matching impedances **3210**-{**1**, . . . , n}, **3220**-{**1**, . . . , n}, and **3240** coupled at the output of each power amplification stage to maximize power transfer from that stage. Further, output stage embodiment **3300** receives an autobias signal **3310**, from an Autobias module **3340**, coupled at the PA stage input of each PA branch **3305**-{**1**, . . . , n}. Autobias module **3340** controls the bias of PAs **3040**-{**1**, . . . , n}. In an embodiment, autobias signal **3340** controls the amount of current flow through the PA stage according to a desired output power level and signal envelope of the output waveform. A further description of the operation of autobias signal and the autobias module is provided below in section 3.5.2.

[0484] FIG. **34** is a block diagram that illustrates another power amplifier (PA) output stage embodiment **3400** according to the present invention. Similar to the embodiment of FIG. **30**, output stage **3400** includes a plurality of PA branches **3405**-{**1**, . . . , n}. Each of PA branches **3405**-{**1**, . . . , n} may include multiple power amplification stages represented by a pre-driver **3020**-{**1**, . . . , n}, driver **3030**-{**1**, . . . , n}, and power amplifier **3040**-{**1**, . . . , n}. Output stage embodiment **3400** may also include pull-impedances **3125**-{**1**, . . . , n}, **3135**-{**1**, . . . , n}, and **3145** coupled at the output of each power amplification stage to achieve desired biasing of that stage. Additionally, output stage embodiment **3400** may include matching impedances **3210**-{**1**, . . . , n}, **3220**-{**1**, . . . , n}, and **3240** coupled at the output of each power amplification stage to maximize power transfer from that stage. Further, output stage embodiment **3400** includes a plurality of harmonic control circuit networks **3410**-{**1**, . . . , n} coupled at the PA stage input of each PA branch {**1**, . . . , n}. Harmonic control circuit networks **3410**-{**1**, . . . , n} may include a plurality of resistance, capacitance, and/or inductive elements and/or active devices coupled in series or in parallel. According to an embodiment of the present invention, harmonic control circuit networks **3410**-{**1**, . . . , n} provide harmonic control functions for controlling the output frequency spectrum of the power amplifier. In an embodiment, harmonic control circuit networks **3410**-{**1**, . . . , n} are selected such that energy transfer to the fundamental harmonic in the summed output spectrum is increased while the harmonic content of the output waveform is decreased. A further description of harmonic control according to embodiments of the present invention is provided below in section 3.6.

[0485] FIG. **35** is a block diagram that illustrates another power amplifier (PA) output stage embodiment **3500** according to the present invention. Output stage embodiment **3500** represents a differential output equivalent of output stage embodiment **3200** of FIG. **32**. In embodiment **3500**, PA stage outputs **3510**-{**1**, . . . , n} are combined successively to result in two aggregate signals. The two aggregate signals are then combined across a loading impedance, thereby having the output of the power amplifier represent the difference between the two aggregate signals. Referring to FIG. **35**, aggregate signals **3510** and **3520** are coupled across loading impedance **3530**. The output of the power amplifier is measured across the loading impedance **3530** as the voltage difference between nodes **3540** and **3550**. According to embodiment **3500**, the maximum output of the power amplifier is obtained when the two aggregate signals are 180 degrees out-of-phase relative to each other. Inversely, the minimum output power results when the two aggregate signals are in-phase relative to each other.

[0486] FIG. **36** is a block diagram that illustrates another output stage embodiment **3600** according to the present invention. Similar to the embodiment of FIG. **30**, output stage **3600** includes a plurality of PA branches **3605**-{**1**, . . . , n}. Each of PA branches {**1**, . . . , n} may include multiple power amplification stages represented by a pre-driver **3020**-{**1**, . . . , n}, a driver **3030**-{**1**, . . . , n}, and a power amplifier (PA) **3620**-{**1**, . . . , n}.

[0487] According to embodiment **3600**, PA's **3620**-{**1**, . . . , n} include switching power amplifiers. In the example of FIG. **36**, power amplifiers **3620**-{**1**, . . . , n} include npn bipolar junction transistor (BJT) elements Q**1**, . . . , Qn. BJT elements Q**1**, . . . , Qn have common collector nodes. Referring to FIG. **36**, collector terminals of BJT elements Q**1**, . . . , Qn are coupled together to provide summing node **3640**. Emitter terminals of BJT elements Q**1**, . . . , Qn are coupled to a ground node, while base terminals of BJT elements Q**1**, . . . , Qn provide input terminals into the PA stage.

[0488] FIG. **37** is an example (related to FIG. **36**) that illustrates an output signal of the PA stage of embodiment **3600** in response to square wave input signals. For ease of illustration, a two-branch PA stage is considered. In the example of FIG. **37**, square wave signals **3730** and **3740** are input, respectively, into BJT elements **3710** and **3720**. Note than when either of BJT elements **3710** or **3720** turns on, summing node **3750** is shorted to ground. Accordingly, when either of input signals **3730** or **3740** is high, output signal **3780** will be zero. Further, output signal **3780** will be high only when both input signals **3730** and **3740** are zero. According to this arrangement, PA stage **3700** performs pulse-width modulation, whereby the magnitude of the output signal is a function of the phase shift angle between the input signals.

[0489] Embodiments are not limited to npn BJT implementations as described herein. A person skilled in the art will appreciate, for example, that embodiments of the present invention may be implemented using pnp BJTs, CMOS, NMOS, PMOS, or other type of transistors. Further, embodiments can be implemented using GaAs and/or SiGe transistors with the desired transistor switching speed being a factor to consider.

[0490] Referring back to FIG. **36**, it is noted that while PAs **3620**-{**1**, . . . , n) are each illustrated using a single BJT notation, each PA **3620**-{**1**, . . . , n} may include a plurality of series-coupled transistors. In embodiments, the number of transistors included within each PA is set according to a required maximum output power level of the power amplifier. In other embodiments, the number of transistors in the PA is such that the numbers of transistors in the pre-driver, driver, and PA stages conform to a geometric progression.

[0491] FIG. **38** illustrates an exemplary PA embodiment **3800** according to an embodiment of the present invention. PA embodiment **3800** includes a BJT element **3870**, a LC network **3860**, and a bias impedance **3850**. BJT element **3870** includes a plurality of BJT transistors Q**1**, . . . , Q**8** coupled in series. As illustrated in FIG. **38**, BJT transistors Q**1**, . . . , Q**8** are coupled together at their base, collector, and emitter terminals. Collector terminal **3880** of BJT element **3870** provides an output terminal for PA **3800**. Emitter terminal **3890** of BJT element **3870** may be coupled to substrate or to an emitter terminal of a preceding amplifier stage. For example, emitter terminal **3890** is coupled to an emitter terminal of a preceding driver stage.

[0492] Referring to FIG. **38**, LC network **3860** is coupled between PA input terminal **3810** and input terminal **3820** of BJT element **3870**. LC network **3860** includes a plurality of capacitive and inductive elements. Optionally, a Harmonic Control Circuit network **3830** is also coupled at input terminal **3820** of BJT element **3870**. As described above, the HCC network **3830** provides a harmonic control function for controlling the output frequency spectrum of the power amplifier.

[0493] Still referring to FIG. **38**, bias impedance **3850** couples Iref signal **3840** to input terminal **3820** of BJT element **3870**. Iref signal **3840** represents an autobias signal that controls the bias of BJT element **3870** according to a desired output power level and signal envelope characteristics.

[0494] It is noted that, in the embodiment of FIG. **38**, BJT element **3870** is illustrated to include 8 transistors. It can be appreciated by a person skilled in the art, however, that BJT element **3870** may include any number of transistors as required to achieve the desired output power level of the power amplifier.

[0495] In another aspect, output stage embodiments can be implemented using multiple-input single-output (MISO) power amplifiers. FIG. **51**A is a block diagram that illustrates an exemplary MISO output stage embodiment **5100**A. Output stage embodiment **5100**A includes a plurality of vector modulator signals **5110**-{**1**, . . . , n} that are input into MISO power amplifier (PA) **5120**. As described above, signals **5110**-{**1**, . . . , n} represent constant envelope constituents of output signal **5130** of the power amplifier. MISO PA **5120** is a multiple input single output power amplifier. MISO PA **5120** receives and amplifies signals **5110**-{**1**, . . . , n} providing a distributed multi signal amplification process to generate output signal **5130**.

[0496] It is noted that MISO implementations, similar to the one shown in FIG. **51**A, can be similarly extended to any of the output stage embodiments described above. More specifically, any of the output stage embodiments of FIGS. **29**-**37** can be implemented using a MISO approach. Additional MISO embodiments will now be provided with reference to FIGS. **51**B-I. It is noted that any of the embodiments described above can be implemented using any of the MISO embodiments that will now be provided.

[0497] Referring to FIG. **51**A, MISO PA **5120** can have any number of inputs as required by the substantially constant envelope decomposition of the complex envelope input signal. For example, in a two-dimensional decomposition, a two-input power amplifier can be used. According to embodiments of the present invention, building blocks for creating MISO PAs for any number of inputs are provided. FIG. **51**B illustrates several MISO building blocks according to an embodiment of the present invention. MISO PA **5110**B represents a two-input single-output PA block. In an embodiment, MISO PA **5110**B includes two PA branches. The PA branches of MISO PA **5110**B may be equivalent to any PA branches described above with reference to FIGS. **29**-**37**, for example. MISO PA **5120**B represents a three-input single-output PA block. In an embodiment, MISO PA **5120**B includes three PA branches. The PA branches of MISO PA **5120**B may equivalent to any PA branches described above with reference to FIGS. **29**-**37**, for example.

[0498] Still referring to FIG. **51**B, MISO PAs **5110**B and **5120**B represent basic building blocks for any multiple-input single-output power amplifier according to embodiments of the present invention. For example, MISO PA **5130**B is a four-input single-output PA, which can be created by coupling together the outputs of two two-input single-output PA blocks, such as MISO PA **5110**B, for example. This is illustrated in FIG. **51**C. Similarly, it can be verified that MISO PA **5140**B, an n-input single-output PA, can be created from the basic building blocks **5110**B and **5120**B.

[0499] FIG. **51**D illustrates various embodiments of the two-input single output PA building block according to embodiments of the present invention.

[0500] Embodiment **5110**D represents an npn implementation of the two-input single output PA building block. Embodiment **5110**D includes two npn transistors coupled together using a common collector node, which provides the output of the PA. A pull-up impedance (not shown) can be coupled between the common collector node and a supply node (not shown).

[0501] Embodiment **5130**D represents a pnp equivalent of embodiment **5110**. Embodiment **5130**D includes two pnp transistors coupled at a common collector node, which provides the output of the PA. A pull-down impedance (not shown) can be coupled between the common collector node and a ground node (not shown).

[0502] Embodiment **5140**D represents a complementary npn/pnp implementation of the two-input single output PA building block. Embodiment **5140**D includes an npn transistor and a pnp transistor coupled at a common collector node, which provides the output of the PA.

[0503] Still referring to FIG. **51**D, embodiment **5120**D represents a NMOS implementation of the two-input single output PA building block. Embodiment **5120**D includes two NMOS transistors coupled at a common drain node, which provides the output of the PA.

[0504] Embodiment **5160**D represents an PMOS equivalent of embodiment **5120**D. Embodiment **5120**D includes two PMOS transistors coupled at a common drain node, which provides the output of the PA.

[0505] Embodiment **5150**D represents a complementary MOS implementation of the two-input single-output PA building block. Embodiment **5150**D includes a PMOS transistor and an NMOS transistor coupled at common drain node, which provides the output of the PA.

[0506] Two-input single-output embodiments of FIG. **51**D can be further extended to create multiple-input single-output PA embodiments. FIG. **51**E illustrates various embodiments of multiple-input single-output PAs according to embodiments of the present invention.

[0507] Embodiment **5150**E represents an npn implementation of a multiple-input single-output PA. Embodiment **5150**E includes a plurality of npn transistors coupled together using a common collector node, which provides the output of the PA. A pull-up impedance (not shown) can be coupled between the common collector node and a supply voltage (not shown). Note that an n-input single-output PA according to embodiment **5150**E can be obtained by coupling additional npn transistors to the two-input single-output PA building block embodiment **5110**D.

[0508] Embodiment **5170**E represents a pnp equivalent of embodiment **5150**E. Embodiment **5170**E includes a plurality of pnp transistors coupled together using a common collector node, which provides the output of the PA. A pull-down impedance (not shown) may be coupled between the common collector node and a ground node (not shown). Note than an n-input single-output PA according to embodiment **5170**E can be obtained by coupling additional pnp transistors to the two-input single-output PA building block embodiment **5130**D.

[0509] Embodiments **5110**E and **5130**E represent complementary npn/pnp implementations of a multiple-input single-output PA. Embodiments **5110**E and **5130**E may include a plurality of npn and/or pnp transistors coupled together using a common collector node, which provides the output of the PA. Note that an n-input single-output PA according to embodiment **5110**E can be obtained by coupling additional npn and/or pnp transistors to the two-input single-output PA building block embodiment **5140**D. Similarly, an n-input single-output PA according to embodiment **5130**E can be obtained by coupling additional npn and/or pnp transistors to the two-input single-output PA building block embodiment **5130**D.

[0510] Embodiment **5180**E represents an PMOS implementation of a multiple-input single-output PA. Embodiment **5180**E includes a plurality of PMOS transistors coupled together using a common drain node, which provides the output of the PA. Note that an n-input single-output PA according to embodiment **5180**E can be obtained by coupling additional NMOS transistors to the two-input single-output PA building block embodiment **5160**D.

[0511] Embodiment **5160**E represents a NMOS implementation of multiple-input single-output PA. Embodiment **5160**E includes a plurality of NMOS transistors coupled together using a common drain node, which provides the output of the PA. Note that an n-input single-output PA according to embodiment **5160**E can be obtained by coupling additional PMOS transistors to the two-input single-output PA building block embodiment **5120**D.

[0512] Embodiments **5120**E and **5140**E complementary MOS implementations of a multiple-input single-output PA. Embodiments **5120**E and **5140**E include a plurality of npn and pnp transistors coupled together using a common drain node, which provides the output of the PA. Note that a n-input single-output PA according to embodiment **5120**E can be obtained by coupling additional NMOS and/or PMOS transistors to the two-input single-output PA building block **5150**D. Similarly, an n-input single-output PA according to embodiment **5140**E can be obtained by coupling additional NMOS and/or PMOS transistors to the two-input single-output PA building block **5160**D.

[0513] FIG. **51**F illustrates further multiple-input single-output PA embodiments according to embodiments of the present invention. Embodiment **5110**F represents a complementary npn/pnp implementation of a multiple-input single-output PA. Embodiment **5110**F can be obtained by iteratively coupling together embodiments of PA building block **5140**D. Similarly, embodiment **5120**F represents an equivalent NMOS/PMOS complementary implementation of a multiple-input single-output PA. Embodiment **5120**F can be obtained by iteratively coupling together embodiments of PA building block **5150**D.

[0514] It must be noted that the multiple-input single-output embodiments described above may each correspond to a single or multiple branches of a PA. For example, referring to FIG. **29**, any of the multiple-input single-output embodiments may be used to replace a single or multiple PAs **2920**-{**1**, . . . , n}. In other words, each of PAs **2920**-{**1**, . . . , n} may be implemented using any of the multiple-input single-output PA embodiments described above or with a single-input single-output PA as shown in FIG. **29**.

[0515] It is further noted that the transistors shown in the embodiments of FIGS. **51**D, **51**E, and **51**F may each be implemented using a series of transistors as shown in the exemplary embodiment of FIG. **38**, for example.

[0516] FIG. **51**G illustrates further embodiments of the multiple-input single-output PA building blocks. Embodiment **5110**G illustrates an embodiment of the two-input single-output PA building block. Embodiment **5110**G includes two PA branches that can each be implemented according to single-input single-output or multiple-input single-output PA embodiments as described above. Further, embodiment **5110**G illustrates an optional bias control signal **5112**G that is coupled to the two branches of the PA embodiment. Bias control signal **5112**G is optionally employed in embodiment **5110**G based on the specific implementation of the PA branches. In certain implementations, bias control will be required for proper operation of the PA. In other implementations, bias control is not required for proper operation of the PA, but may provide improved PA power efficiency, output circuit protection, or power on current protection.

[0517] Still referring to FIG. **51**G, embodiment **5120**G illustrates an embodiment of the three-input single-output PA building block. Embodiment **5120**G includes three PA branches that can each be implemented according to single-input single-output or multiple-input single-output PA embodiments as described above. Further, embodiment **5120**G illustrates an optional bias control signal **5114**G that is coupled to the branches of the PA embodiment. Bias control signal **5114**G is optionally employed in embodiment **5120**G based on the specific implementation of the PA branches. In certain implementations, bias control will be required for proper operation of the PA. In other implementations, bias control is not required for proper operation of the PA, but may provide improved PA power efficiency.

[0518] FIG. **51**H illustrates a further exemplary embodiment **5100**H of the two-input single-output. PA building block. Embodiment **5100**H includes two PA branches that can each be implemented according to single-input single-output or multiple-input single-output PA embodiments as described above. Embodiment **5100**H further includes optional elements, illustrated using dashed lines in FIG. **51**H, that can be additionally employed in embodiments of embodiment **5100**H. In an embodiment, PA building block **5100**H may include a driver stage and/or pre-driver stage in each of the PA branches as shown in FIG. **51**H. Process detectors may also be optionally employed to detect process and temperature variations in the driver and/or pre-driver stages of the PA. Further, optional bias control may be provided to each of the pre-driver, driver, and/or PA stages of each branch of the PA embodiment. Bias control may be provided to one or more the stages based on the specific implementation of that stage. Further, bias control may be required for certain implementations, while it can be optionally employed in others.

[0519] FIG. **51**I illustrates a further exemplary embodiment **5100**I of a multiple-input single-output PA. Embodiment **5100**I includes at least two PA branches that can each be implemented according to single-input single-output or multiple-input single-output PA embodiments as described above. Embodiment **5100**I further includes optional elements that can be additionally employed in embodiments of embodiment **5100**I. In an embodiment, the PA may include driver and/or pre-driver stages in each of the PA branches as shown in FIG. **51**I. Process detectors may also be optionally employed to detect process and temperature variations in the driver and/or pre-driver stages of the PA. Further, optional bias control may be provided to each of the pre-driver, driver, and/or PA stages of each branch of the PA embodiment. Bias control may be provided to one or more the stages based on the specific implementation of that stage.

Further, bias control may be required for certain implementations, while it can be optionally employed in others.

3.5.2) Output Stage Current Control —Autobias Module

[0520] Embodiments of the output stage and optional pre-driver and driver stage bias and current control techniques according to embodiments of the present invention are described below. In certain embodiments, output stage current control functions are employed to increase the output stage efficiency of a vector power amplifier (VPA) embodiment In other embodiments, output stage current control is used to provide output stage protection from excessive voltages and currents which is further describe in section 3.5.3. In embodiments, output stage current control functions are performed using the Autobias module described above with reference to FIG. **33**. A description of the operation of the Autobias module in performing these current control functions is also presented below according to an embodiment of the present invention.

[0521] According to embodiments of the present invention, power efficiency of the output stage of a VPA can be increased by controlling the output stage current of the VPA as a function of the output power and the envelope of the output waveform.

[0522] FIG. **37**, illustrates a partial schematic of a Multiple Input Single Output amplifier comprised of two NPN transistors with input signals S1 and S2. When S1 and S2 are designed to be substantially similar waveforms and substantially constant envelope signals, any time varying complex-envelope output signal can be created at circuit node **3750** by changing the phase relationship of S1 and S2.

[0523] FIG. **39** illustrates an example time varying complex-envelope output signal **3910** and its corresponding envelope signal **3920**. Note than signal **3910** undergoes a reversal of phase at an instant of time t_0. Correspondingly, envelope signal **3920** undergoes a zero crossing at time to. Output signal **3910** exemplifies output signals according to typical wireless signaling schemes such as W-CDMA, QPSK, and OFDM, for example.

[0524] FIG. **40** illustrates example diagram FIG. **37**'s output stage current in response to output signal **3910**. I_{out} signal **4010** represents output stage current without autobias control, and I_{out} signal **4020** represents output stage current with autobias control. Without autobias control, as the phase shift between S1 and S2 changes from 0 to 180 degrees, the output current I_{out} increases. With autobias control, the output current I_{out} decreases and can be minimized when at or near t_0 of FIG. **39**.

[0525] Note that I_{out} signal **4020** varies as a function of envelope signal **3920**. Accordingly, I_{out} signal **4020** is at the maximum when a maximum output power is required, but decreases as the required output power goes down. Particularly, I_{out} signal **4020** approaches zero as the associated output power goes to zero. Accordingly, a person skilled in the art will appreciate that output stage current control, according to embodiments of the present invention, results in significant power savings and increases the power efficiency of the power amplifier.

[0526] According to embodiments of the present invention, output stage current control may be implemented according to a variety of functions. In an embodiment, the output stage current can be shaped to correspond to the desired output power of the amplifier. In such an embodiment, the output stage current is a function that is derived from the envelope of the desired output signal, and the power efficiency will increase.

[0527] FIG. **41** illustrates exemplary autobias output stage current control functions **4110** and **4120** according to embodiments of the present invention. Function **4110** may represent a function of output power and signal envelope as described above. On the other hand, function **4120** may represent a simple shaping function that goes to a minimum value for a pre-determined amount of time when the output power is below a threshold value. Accordingly, functions **4110** and **4120** represent two cases of autobias output stage current control functions with autobias control signal **4110** resulting in I_{out} response **4130** and autobias control signal **4120** resulting in I_{out} response **4140**. The invention, however, is not limited to those two exemplary embodiments. According to embodiments of the present invention, output stage autobias current control functions may be designed and implemented to accommodate the efficiency and current consumption requirements of a particular vector power amplifier design.

[0528] In implementation, several approaches exist for performing output stage current control. In some embodiments, output stage current shaping is performed using the Autobias module. The Autobias module is illustrated as autobias circuitry **714** and **716** in the embodiments of FIGS. **7** and **8**. Similarly, the Autobias module is illustrated as autobias circuitry **1218** in the embodiments of FIGS. **12** and **13**, and as autobias circuitry **1718** in the embodiments of FIGS. **17** and **18**.

[0529] Output stage current control using Autobias is depicted in process flowchart **4800** of the embodiment of FIG. **48**. The process begins in step **4810**, which includes receiving output power and output signal envelope information of a desired output signal of a vector power amplifier (VPA). In some embodiments, implementing output stage current control using Autobias requires a priori knowledge of the desired output power of the amplifier. Output power information may be in the form of envelope and phase information. For example, in the embodiments of FIGS. **7**, **8**, **12**, **13**, **17**, and **18**, output power information is included in I and Q data components received by the VPA embodiment. In other embodiments, output power information may be received or calculated using other means.

[0530] Step **4820** includes calculating a signal according to the output power and output envelope signal information. In embodiments, an Autobias signal is calculated as a function of some measure of the desired output power. For example, the Autobias signal may be calculated as a function of the envelope magnitude of the desired output signal. Referring to the embodiments of FIGS. **7**, **8**, **12**, **13**, **17**, and **18**, for example, it is noted that the Autobias signal (signals **715** and **717** in FIGS. **7** and **8**, signal **1228** in FIGS. **12** and **13**, and signals **1728** in FIGS. **17** and **18**) is calculated according to received I and Q data components of a desired output signal. In certain embodiments, such as the ones described in FIGS. **7**, **8**, **12**, **13**, **17**, and **18**, the Autobias signal is calculated by an Autobias module being provided output power information. In other embodiments, the Autobias signal may be calculated by the I and Q Data Transfer Function module(s) of the VPA. In such embodiments, an Autobias module may not be required in implementation. In embodiments, the I and Q Data Transfer Function module calculates a signal, outputs the signal to a DAC which output signal represents the Autobias signal.

[0531] Step **4830** includes applying the calculated signal at an output stage of the VPA, thereby controlling a current of the output stage according to the output power of the desired output signal. In embodiments, step **4830** includes coupling the Autobias signal at the PA stage input of the VPA. This is illustrated, for example, in the embodiments of FIGS. **33** and **42** where Autobias signal **3310** is coupled at the PA stage input of the VPA embodiment. In these embodiments, Autobias signal **3310** controls the bias of the PA stage transistors according to the output power of the desired output signal of the VPA embodiment. For example, Autobias signal **3310** may cause the PA stage transistors to operate in cutoff state when the desired output power is minimal or near zero, thereby drawing little or no output stage current. Similarly, when a maximum output power is desired, Autobias signal **3310** may bias the PA stage transistors to operate in class C,D,E, etc. switching mode. Autobias signal **3310** may also cause the PA stage transistors or FETs to operate in forward or reverse biased states according to the desired output power and signal envelope characteristics.

[0532] In other embodiments, step **4830** includes coupling the Autobias signal using pull-up impedances at the PA stage input and optionally the inputs of the driver and pre-driver stages of the VPA. FIGS. **38** and **43** illustrate such embodiments. For example, in the embodiment of FIG. **38**, bias impedance **3850** couples Autobias Iref signal **3840** to input terminal **3820** of BJT element **3870**. BJT element **3870** represents the PA stage of one PA branch of an exemplary VPA embodiment. Similarly, in the embodiment of FIG. **43**, Autobias signal **4310** is coupled to transistors **Q1**, . . . , **Q8** through corresponding bias impedances **Z1**, . . . , **Z8**. Transistors **Q1**, . . . , **Q8** represent the PA stage of one branch of an exemplary VPA embodiment.

[0533] Embodiments for implementing the Autobias circuitry described above will now be provided. FIG. **27** illustrates three embodiments **2700**A, **2700**B, and **2700**C for implementing the Autobias circuitry. These embodiments are provided for illustrative purposes, and are not limiting. Other embodiments will be apparent to persons skilled in the art(s) based on the teachings contained herein.

[0534] In embodiment **2700**A, Autobias circuitry **2700**A includes an Autobias Transfer Function module **2712**, a DAC **2714**, and an optional interpolation filter **2718**. Autobias circuitry **2700**A receives an I and Q Data signal **2710**. Autobias Transfer Function module **2712** processes the received I and Q Data signal **2710** to generate an appropriate bias signal **2713**. Autobias Transfer Function module **2712** outputs bias signal **2713** to DAC **2714**. DAC **2714** is controlled by a DAC clock **2716** which may be generated in Autobias transfer module **2712**. DAC **2714** converts bias signal **2713** into an analog signal, and outputs the analog signal to interpolation filter **2718**. Interpolation filter **2718**, which also serves as an anti-aliasing filter, shapes the DAC's output to generate Autobias signal **2720**, illustrated as Bias A in embodiment **5112**G. Autobias signal **2720** may be used to bias the PA stage and/or the driver stage, and/or the pre-driver stage of the amplifier. In an embodiment, Autobias signal **2720** may have several other Autobias signals derived therefrom to bias different stages within the PA stage. This can be done using additional circuitry not included in embodiment **2700**A.

[0535] In contrast, embodiment **2700**B illustrates an Autobias circuitry embodiment in which multiple Autobias signals are derived within the Autobias circuitry. As shown in embodiment **2700**B, circuit networks **2722**, **2726**, and **2730**, illustrated as circuit networks A, B, and C in embodiment **2700**B, are used to derive Autobias signals **2724** and **2728** from Autobias signal **2720**. Autobias signals **2720**, **2724**, and **2728** are used to bias different amplification stages.

[0536] Embodiment **2700**C illustrates another Autobias circuitry embodiment in which multiple Autobias signals are generated independently within the Autobias Transfer Function module **2712**. In embodiment **2700**C, Autobias Transfer Function module **2712** generates multiple bias signals according to the received I and Q Data signal **2710**. The bias signals may or may not be related. Autobias Transfer Function module **2712** outputs the generated bias signals to subsequent DACs **2732**, **2734**, and **2736**. DACs **2732**, **2734**, and **2736** are controlled by DAC clock signals **2733**, **2735**, and **2737**, respectively. DACs **2732**, **2734**, and **2736** convert the received bias signals into analog signals, and output the analog signals to optional interpolation filters **2742**, **2744**, and **2746**. Interpolation filters **2742**, **2744**, and **2746**, which also serve as anti-aliasing filters, shape the DACs outputs to generate Autobias signals **2720**, **2724**, and **2728**. Similar to embodiment **2700**B, Autobias signals **2720**, **2724**, and **2728** are used to bias different amplification stages such as the pre-driver, driver, and PA.

[0537] As noted above, Autobias circuitry embodiments according to the present invention are not limited to the ones described in embodiments **2700**A, **2700**B, and **2700**C. A person skilled in the art will appreciate, for example, that Autobias circuitry can be extended to generate any number of bias control signals as required to control the bias of various stages of amplification, and not just three as shown in embodiments **5200**B and **5200**C, for example.

3.5.3) Output Stage Protection

[0538] As described above, output stage embodiments according to embodiments of the present invention are highly power efficient as a result of being able to directly couple outputs at the PA stage using no combining or isolating elements. Certain output stage embodiments in certain circumstances and/or applications, however, may require additional special output stage protection measures in order to withstand such direct coupling approach. This may be the case for example for output stage embodiments such as **5110**D, **5120**D, **5130**D, **5160**D, **5150**E, **5160**E, **5170**E, and **5180**E illustrated in FIGS. **51**D and **51**E. Note that, generally, complementary output stage embodiments, such as embodiments **5140**D, **5150**D, **5110**E, **5120**E, **5130**E, and **5140**E of FIGS. **51**D and **51**E, do not require (but may optionally use) the same output stage protection measures as will be described herein in this section. Output stage protection measures and embodiments to support such measures are now provided.

[0539] In one aspect, transistors of distinct branches of a PA stage should generally not simultaneously be in opposite states of operation for extended periods of time. Following a restart or power on with no inputs being supplied to the final PA stages, transients within the PA branches may cause this mode to occur resulting in the PA stage transistors potentially damaging one another or circuit elements connected to the output. Accordingly, embodiments of the present invention further constrain the Autobias module to limit the output current in the PA stage.

[0540] In another aspect, it may be desired to ensure that the Autobias module limits the output voltages below the breakdown voltage specification of the PA stage transistors. Accordingly, in embodiments of the present invention, such as the one illustrated in FIG. **42** for example, a feedback element **4210** is coupled between the common collector node of the PA stage and the Autobias module. Feedback element **4210** monitors the collector to base voltage of the PA stage transistors, and may constrain the Autobias signal as necessary to protect the transistors and/or circuit elements.

[0541] A person skilled in the art will appreciate that other output stage protection techniques may also be implemented. Furthermore, output stage protection techniques may be implementation specific. For example, depending on the type of PA stage transistors (npn, pnp, NMOS, PMOS, npn/pnp, NMOS/PMOS), different protection functions may be required.

3.6) Harmonic Control

[0542] According to embodiments of the present invention, an underlying principle for each branch PA is to maximize the transfer of power to a fundamental harmonic of the output spectrum. Typically, each branch PA may be multi-stage giving rise to a harmonically rich output spectrum. In one aspect, transfer of real power is maximized for the fundamental harmonic. In another aspect, for non-fundamental harmonics, real power transfer is minimized while imaginary power transfer may be tolerated. Harmonic control, according to embodiments of the present invention, may be performed in a variety of ways.

[0543] In one embodiment, real power transfer onto the fundamental harmonic is maximized by means of waveshaping of the PA stage input signals. In practice, several factors play a role in determining the optimal wave shape that results in a maximum real power transfer onto the fundamental harmonic. Embodiment **3400** of the present invention, described above, represents one embodiment that employs waveshaping of PA stage input signals. In embodiment **3400**, a plurality of harmonic control circuitry (HCC) networks **3410**-{**1**, . . . , n} are coupled at the PA stage input of each PA branch {**1**, . . . , n}. HCC networks **3410**-{**1**, . . . , n} have the effect of waveshaping the PA stage inputs, and are typically selected so as to maximize real power transfer to the fundamental harmonic of the summed output spectrum. According to embodiments of the present invention, waveshaping can be used to generate variations of harmonically diverse waveforms. In other embodiments, as can be apparent to a person skilled in the art, waveshaping can be performed at the pre-driver and/or the driver stage.

[0544] In another embodiment, harmonic control is achieved by means of waveshaping of the PA stage output. FIG. **43** illustrates an exemplary PA stage embodiment **4300** of the present invention. In embodiment **4300**, Autobias signal **4310** is coupled to transistors Q**1**, . . . , Q**8** through corresponding bias impedances Z**1**, . . . , Z**8**. Notice that when impedances Z**1**, . . . , Z**8** have different values, transistors Q**1**, . . . , Q**8** have different bias points and can be turned on at different times. This approach of biasing transistors Q**1**, . . . , Q**8** is referred to as staggered bias. Note that using staggered bias, the PA output waveform can be shaped in a variety of ways depending on the values assigned to bias impedances Z**1**, . . . , Z**8**.

[0545] Harmonic control using staggered bias is depicted in process flowchart **4900** of the embodiment of FIG. **49**. The process begins in step **4910**, which includes coupling an input signal at first ports of a plurality of transistors of a power amplifier (PA) switching stage. In the example embodiment of FIG. **43**, for example, step **4910** corresponds to coupling PA_IN signal **4310** at base terminals of the plurality of transistors Q**1**, . . . , Q**8**.

[0546] Step **4920** includes coupling a plurality of impedances between the first ports of the plurality of transistors and a bias signal. In the example embodiment of FIG. **43**, for example, step **4920** is achieved by coupling impedances Z**1**, . . . , Z**8** between base terminals of respective transistors Q**1**, . . . , Q**8** and Iref signal. In an embodiment, values of the plurality of impedances are selected to cause a time-staggered switching of the input signal, thereby harmonically shaping an output signal of the PA stage. In embodiments, a multi-stage staggered output may be generated by selecting multiple distinct values of the plurality of impedances. In other embodiments, switching is achieved by selecting the plurality of impedances to have equal or substantially equal value.

[0547] FIG. **44** illustrates an exemplary wave-shaped PA output using a two-stage staggered bias approach. In a two-stage staggered bias approach, a first set of the PA transistors is first turned on before a second set is turned on. In other words, the bias impedances take two different values. Waveform **4410** represents an input waveform into the PA stage. Waveform **4420** represents the wave-shaped PA output according to a two-stage staggered bias. Notice that output waveform **4420** slopes twice as it transitions from 1 to 0, which corresponds to the first and second sets of transistors turning on successively.

[0548] According to embodiments of the present invention, a variety of multi-stage staggered bias approaches may be designed. Bias impedance values may be fixed or variable. Furthermore, bias impedance values may be equal or substantially equal, distinct, or set according to a variety of permutations. For example, referring to the example of FIG. **43**, one exemplary permutation might set Z**1**=Z**2**=Z**3**=Z**4** and Z**5**=Z**6**=Z**7**=Z**8** resulting in a two-stage staggered bias.

3.7) Power Control

[0549] Vector power amplification embodiments of the present invention intrinsically provide a mechanism for performing output power control.

[0550] FIG. **45** illustrates one approach for performing power control according to an embodiment of the present invention. In FIG. **45**, phasors $\vec{U}_1$ and $\vec{L}_1$ represent upper and lower constituents of a first phasor $\vec{R}_1$. $\vec{U}_1$ and $\vec{L}_1$**0** are constant magnitude and are symmetrically shifted in phase relative to $\vec{R}_1$ by a phase shift angle

$$\frac{\phi}{2}.$$

Phasors $\vec{U}_2$ and $\vec{L}_2$ represent upper and lower constituents of a second phasor $\vec{R}_2$. $\vec{U}_2$ and $\vec{L}_2$ are constant magnitude and are symmetrically shifted in phase relative to $\vec{R}_2$ by a phase shift angle

$$\frac{\phi}{2}+\phi_{off}.$$

[0551] It is noted, from FIG. **45**, that $\vec{R}_1$ and $\vec{R}_2$ are in-phase relative to each other but only differ in magnitude. Furthermore, $\vec{U}_2$ and $\vec{L}_2$ are equally or substantially equally phased shifted relative to $\vec{U}_1$ and $\vec{L}_1$, respectively. Accordingly, it can be inferred that, according to the present invention, a signal's magnitude can be manipulated without varying its phase shift angle by equally or substantially equally shifting symmetrically its constituent signals.

[0552] According to the above observation, output power control can be performed by imposing constraints on the phase shift angle of the constituent signals of a desired output signal. Referring to FIG. **45**, for example, by constraining the range of values that phase shift angle

$$\frac{\phi}{2}$$

can take, magnitude constraints can be imposed on phasor $\vec{R}_1$.

[0553] According to embodiments of the present invention, a maximum output power level can be achieved by imposing a minimum phase shift angle condition. For example, referring to FIG. **45**, by setting a condition such that

$$\frac{\phi}{2}\geq\phi_{ff},$$

the magnitude of phasor $\vec{R}_1$ is constrained not to exceed a certain maximum level. Similarly, a maximum phase shift angle condition imposes a minimum magnitude level requirement.

[0554] In another aspect of power control, output power resolution is defined in terms of a minimum power increment or decrement step size. According to an embodiment of the present invention, output power resolution may be implemented by defining a minimum phase shift angle step size. Accordingly, phase shift angle values are set according to a discrete value range having a pre-determined step size. FIG. **46** illustrates an exemplary phase shift angle spectrum, whereby phase shift angle

$$\frac{\phi}{2}$$

is set according to a pre-determined value range having a minimum step ϕ_{step}.

[0555] A person skilled in the art will appreciate that a variety of power control schemes may be implemented in a fashion similar to the techniques described above. In other words, various power control algorithms can be designed, according to the present invention, by setting corresponding constraints on phase shift angle values. It is also apparent, based on the description above of data transfer functions, that power control schemes can be naturally incorporated into a transfer function implementation.

3.8) Exemplary Vector Power Amplifier Embodiment

[0556] FIG. **47** illustrates an exemplary embodiment **4700** of a vector power amplifier according to the present invention. Embodiment **4700** is implemented according to the Direct Cartesian 2-Branch VPA method.

[0557] Referring to FIG. **47**, signals **4710** and **4712** represent incoming signals from a transfer function stage. The transfer function stage is not shown in FIG. **47**. Block **4720** represents a quadrature generator which may be optionally implemented according to an embodiment of the present invention. Quadrature generator **4720** generates clock signals **4730** and **4732** to be used by vector modulators **4740** and **4742**, respectively. Similarly, signals **4710** and **4712** are input into vector modulators **4740** and **4742**. As described above, vector modulators **4740** and **4742** generate constant envelope constituents that are, subsequently, processed by a PA stage. In embodiment **4700**, the PA stage is multi-stage, whereby each PA branch includes a pre-driver stage **4750-4752**, a driver stage **4760-4762**, and a power amplifier stage **4770-4772**.

[0558] Further illustrated in FIG. **47** are Autobias signals **4774** and **4776**, and terminals **4780** and **4782** for coupling harmonic control circuitry and networks. Terminal node **4780** represents the output terminal of the vector power amplifier, and is obtained by direct coupling of the two PA branches' outputs.

4. Additional Exemplary Embodiments and Implementations

4.1) Overview

[0559] Exemplary VPA implementations according to embodiments of the present invention will be provided in this section. Advantages of these VPA implementations will be appreciated by persons skilled in the art based on the teachings herein. We briefly describe below some of these advantages before presenting in more detail the exemplary VPA implementations.

4.1.1) Control of Output Power and Power Efficiency

[0560] The exemplary VPA implementations enable several layers of functionality for performing power control and/or for controlling power efficiency using circuitry within the VPA. FIG. **52** illustrates this functionality at a high level using a MISO VPA embodiment **5200**. MISO VPA embodiment **5200** is a 2 input single output VPA with optional driver and pre-driver stages in each branch of the VPA. As in previously described embodiments, the input bias voltage or current to each amplification stage (e.g., pre-driver stage, driver stage, etc.) of the VPA is controlled using a bias signal (also referred to as Autobias in other embodiments). In embodiment **5200**, separate bias signals Bias C, Bias B, and Bias A are coupled to the pre-driver, driver, and PA stages, respectively, of the VPA. Additionally, VPA embodiment **5200** includes power supply signals (Pre-Driver VSUPPLY, Driver VSUPPLY, and Output Stage VSUPPLY) that are used to power respective stages of the VPA. In embodiments, these power supply signals are generated using voltage controlled power supplies and can be further used to bias their respective amplifications stages, thereby providing additional functionality for controlling the overall power efficiency of the VPA and for performing power control, as well as other functions of the VPA. For example, when controlled independently, the power supply signals and bias signals can be used to operate different amplification stages of the VPA at different power supply voltages and bias points, enabling a wide output power dynamic range for the VPA. In embodiments the voltage controlled power supplies can be implemented as continuously variable supplies such as voltage controlled switching supplies which provide variable voltage supplies to the appropriate amplification stage. In other embodiments the voltage controlled power supply can be implemented by using switches to provide different power supply voltages. For example, a VPA output stage and/or optional driver stages and/or optional pre-driver stages power supply could be switched between 3.3V, 1.8V, and 0V depending on the desired operating parameters.

4.1.2) Error Compensation and/or Correction

[0561] The exemplary VPA implementations provide different approaches for monitoring and/or compensating for errors in the VPA. These errors may be due, among other factors, to process and/or temperature variations in the VPA, phase and amplitude errors in the vector modulation circuitry, gain and phase imbalances in branches of the VPA, and distortion in the MISO amplifier (see, for example, Section 3.4.5 above). In previously described VPA embodiments, part of this functionality was embodied in the process detector circuitry (e.g., process detector **792** in FIG. **7**A, process detector **1282** in FIG. **12**, process detector **1772** in FIG. **17**). These approaches can be classified as feedforward, feedback, and hybrid feedforward/feedback techniques, and can be implemented in a variety of ways as will be further discussed in the following sections that describe the exemplary VPA implementations. A conceptual description of these error monitoring and compensation approaches will be now provided.

[0562] FIGS. **54**A and **54**B are block diagrams that illustrate at a high level feedforward techniques for compensating for errors in a VPA. Feedforward techniques rely on a priori knowledge of expected errors in the VPA in order to pre-compensate for these errors within the VPA. Thus, feedforward techniques include an error measurement phase (typically performed in a test and characterization process) and a pre-compensation phase using the error measurements.

[0563] FIG. **54**A illustrates a process **5400**A for generating an error table or function that describes expected errors in I data and Q data at the output of the VPA (error measurement phase). Such errors are typically due to imperfections in the VPA. Process **5400**A is typically performed in a testing lab prior to finalizing the VPA design, and includes measuring at the output of a receiver I and Q values that correspond to a range of I and Q values at the input of the VPA. Typically, the input I and Q values are selected to generate a representative range of the 360° degrees polar space (for example, the I and Q values may be selected at a uniform spacing of 30° degrees). Subsequently, error differences between the input I and Q values and the output I and Q values are calculated. For example, after measuring I and Q at the output of the receiver for a particular set of I and Q input values, a compare circuitry calculates as I_{error} and Q_{error} or the differences in I data and Q data between the input I and Q values and the receiver output I and Q values. I_{error} and Q_{error} represent the expected errors in I and Q at the output of the VPA for the particular set of I and Q input values.

[0564] In an embodiment, the receiver is integrated with the VPA, or is provided by an external calibration and/or testing device. Alternatively, the receiver is the receiver module in the device employing the VPA (e.g., the receiver in a cellular phone). In this alternative embodiment, the VPA error table and/or feedback information can be generated by this receiver module in the device.

[0565] The calculated I_{error} and Q_{error} values are used to generate an error table or function representative of expected I and Q errors for various I and Q input values. In embodiments, the calculated I_{error} and Q_{error} values are further interpolated to generate error values for an augmented range of I and Q input values, based on which the error table or function is generated.

[0566] FIG. **54**B illustrates feedforward error pre-compensation (pre-compensation phase) according to an embodiment of the present invention. As illustrated, I and Q input values are corrected for any expected I_{error} and Q_{error} values as determined by an error table or function, prior to amplification by the VPA. I and Q error pre-compensation may be performed at different stages and/or at different temperatures and/or at different operating parameters within the VPA. In the embodiment of FIG. **54**B, error correction occurs prior to the amplification stage of the VPA. For example, I and Q error correction may be performed by the transfer function module of the VPA, such as transfer function modules **1216** and **1726** of FIGS. **12** and **17**, for example. Several methods exist for implementing I and Q error correction in the transfer function module of the VPA including using look up tables and/or digital logic to implement an error function. Typically, feedforward techniques require data storage such as RAM or NVRAM, for example, to store data generated in the measurement phase.

[0567] In contrast to feedforward techniques, feedback techniques do not pre-compensate for errors but perform real-time measurements inside or at the output of the VPA to detect any errors or deviations due to process or temperature variations, for example. FIG. **55** is a block diagram that conceptually illustrates an exemplary Cartesian feedback error correction technique according to embodiments of the present invention. As will be further described below, FIG. **55** illustrates a receiver-based feedback technique, in which the output of the VPA is received by a receiver, before being fed back to the VPA. Other feedback techniques according to embodiments of the present invention will be further described below. Feedback techniques may require additional circuitry to perform these real-time measurements, which may be made at different stages within the VPA, but require minimal or no data storage. Several implementations exist for feedback error correction as will be further described in the description of the exemplary VPA implementations below.

[0568] Hybrid feedforward/feedback techniques include both feedforward and feedback error pre-compensation and/or correction components. For example, a hybrid feedforward/feedback technique may pre-compensate for errors but may also use low rate periodical feedback mechanisms to supplement feedforward pre-compensation.

4.1.3) Multi-Band Multi-Mode VPA Operation

[0569] The exemplary VPA implementations provide several VPA architectures for concurrently supporting multiple frequency bands (e.g., quad band) and/or multiple technology modes (e.g., tri mode) for data transmission. Advantages of these VPA architectures will be appreciated by a person skilled in the art based on the teachings to be provided herein. In embodiments, the VPA architectures allow for using a single PA branch for supporting both TDD (Time Division Duplex) and FDD (Frequency Division Duplex) based standards. In other embodiments, the VPA architectures allow for the elimination of costly and power inefficient components at the output stage (e.g., isolators), typically required for FDD based standards. For the purpose of illustration and not limitation, frequency band allocation on lower and upper spectrum bands for various communication standards is provided in FIG. **53**. Note that the DCS **1800** (Digital Cellular System **1800**) and the PCS **1900** (Personal Communications Service **1900**) bands can support different GSM-based implementations, also known as GSM-**1800** and GSM-**1900**. The 3G TDD bands are allocated for third generation time division duplex standards such as UMTS TDD (Universal Mobile Telephone System) and TD-SCDMA (Time Division-Synchronous Code Division Multiple Access), for example. The 3G FDD bands are allocated for third generation frequency division duplex standards such as WCDMA (Wideband CDMA), for example.

[0570] As will be appreciated by persons skilled in the art based on the teachings herein, advantages enabled by the exemplary VPA implementations exist in various aspects in addition to those described above. In the following, a more detailed description of the exemplary VPA implementations will be provided. This includes a description of different implementations of the digital control circuitry of the VPA followed by a description of different implementations of the analog core of the VPA. Embodiments of the present invention are not limited to the specific implementations described herein. As will be understood by persons skilled in the art based on the teachings herein, several other VPA implementations may be obtained by combining features provided in the exemplary VPA implementations. Accordingly, the exemplary VPA implementations described below do not represent an exhaustive listing of VPA implementations according to embodiments of the present invention, and other implementations based on teachings contained herein are also within the scope of the present invention. For example, certain digital control circuitry could be integrated or combined with a baseband processor. In addition, certain analog control circuitry such as quadrature generators and vector modulators can be implemented using digital control circuitry. In an embodiment, the VPA system can be implemented in its entirety using digital circuitry and can be integrated completely with a baseband processor.

4.2) Digital Control Module

[0571] The digital control module of the VPA includes digital circuitry that is used, among other functions, for signal generation, performance monitoring, and VPA operation control. In Section 3, the signal generation functions of the digital control module (i.e., generating constant envelope signals) were described in detail with reference to the transfer function module (state machine) of the digital control module, in embodiments **700**, **1200**, and **1700**, for example. The performance monitoring functions of the digital control module include functions for monitoring and correcting for errors in the operation of the VPA and/or functions for controlling the bias of different stages of the VPA. The VPA operation control functions of the digital control module include a variety of control functions related to the operation of the VPA (e.g., powering up or programming VPA modules). In certain embodiments, these control functions may be optional. In other embodiments, these control functions are accessible through the digital control module to external processors connected to the VPA. In other embodiments, these functions are integrated with baseband processors or other digital circuitry. Other functions are also performed by the digital control module in addition to those described above. Digital control module functions and implementations will now be provided in further detail.

[0572] FIG. **56** is a high level illustration of a digital control module embodiment **5600** according to an embodiment of the present invention. Digital control module embodiment **5600** includes an input interface **5602**, an output interface **5604**, a state machine **5606**, a RAM (Random Access Memory) **5608**, and a NVRAM (Non-Volatile RAM) **5610**. In embodiments, Ram **5608**, and/or NVRAM **5610** may be optional.

[0573] Input interface **5602** provides a plurality of buses and/or ports for inputting signals into digital control module **5600**. These buses and/or ports include, for example, buses and/or ports for inputting I and Q data signals, control signals provided by an external processor, and/or clock signals. In an embodiment, input interface **5602** includes an I/O bus. In another embodiment, input interface **5602** includes a data bus for receiving feedback signals from the analog core of the VPA. In another embodiment, input interface **5602** includes ports for reading values out of digital control module **5600**. In an embodiment, values are read out of digital control module **5600** by an external processor (e.g., a baseband processor) connected to digital control module **5600**.

[0574] Output interface **5604** provides a plurality of output buses and/or ports for outputting signals from digital control module **5600**. These output buses and/or ports include, for example, buses and/or ports for outputting amplitude information signals (used to generate constant envelope signals), bias control signals (Autobias signals), voltage control signals (power supply signals), and output select signals.

[0575] State machine **5606** performs various functions related to the signal generation and/or performance monitoring functions of digital control module **5600**. In an embodiment, state machine **5606** includes a transfer function module, as described in Section 3, for performing signal generation functions. In another embodiment, state machine **5606** includes modules for generating, among other types of signals, bias control signals, power control signals, gain control signals, and phase control signals. In another embodiment, state machine **5606** includes modules for performing error pre-compensation in a feedforward error correction system.

[0576] RAM **5608** and/or NVRAM **5610** are optional components of digital control module **5600**. In embodiments, RAM **5608** and NVRAM **5610** reside externally of digital control module **5600** and may be accessible to digital control module **5600** through data buses connected to digital control module **5600** via input interface **5602**, for example. RAM **5608** and/or NVRAM **5610** may or may not be needed depending on the specific VPA implementation. For example, a VPA implementation employing feedforward techniques for error pre-compensation may require RAM **5608** or NVRAM **5610** to store error tables or functions. On the other hand, a feedback technique for error correction may solely rely on digital logic modules in the state machine and may not require RAM **5608** or NVRAM **5610** storage. Similarly, the amount of RAM **5608** and NVRAM **5610** storage may depend on the specific VPA implementation. Typically, when used, NVRAM **5610** is used for storing data that is not generated in real time and/or that must be retained when power is turned off. This includes, for example, error tables and/or error values such as scalar values and angular values generated in the testing and characterization phase of the VPA system and/or look up tables used by transfer functions modules.

[0577] FIG. **57** illustrates an exemplary digital control module implementation **5700** according to an embodiment of the present invention. Digital control module implementation **5700** illustrates in particular an exemplary input interface **5602** and an exemplary output interface **5604** of an exemplary VPA digital control module **5700**. As will be further described below, signals of the input and output interfaces **5602** and **5604** of VPA digital control module **5700** correlate directly with signals from the analog core of the VPA and/or signals to/from one or more external processors/controllers connected to the VPA. In the example embodiments described in the sections above, the analog core of the VPA was represented by analog circuitry **186** together with PA stage **190**-{**1**, . . . , n} in FIG. **1**E, for example. It is noted that bit widths of data buses and/or signals of the input and output interfaces in FIG. **57** are provided for the purpose of illustration only and are not limiting.

[0578] The input interface **5602** of exemplary digital control module **5700** includes an A/D IN bus **5702**, a digital I/O bus **5704**, and a plurality of control signals **5706**-**5730**. In other digital control module implementations, the input interface **5602** may include more or less data buses, programming buses, and/or control signals.

[0579] A/D IN bus **5702** carries feedback information from the analog core of the VPA to the digital control module **5700**. Feedback information can be used, among other functions, to monitor the output power of the VPA and/or for amplitude and/or phase variations in branches of the VPA. As illustrated in FIG. **57**, an A/D converter **5732** converts from analog to digital feedback information received from the analog core of the VPA (using A/D IN signal **5736**) before sending it on A/D IN bus **5702** to the digital control module **5700**. In an embodiment, the digital control module **5700** controls a clock signal A/D CLK **5734** of the A/D converter **5732**. In another embodiment, the digital control module **5700** controls an input selector to the AID converter **5732** to select between multiple feedback signals at the input of the A/D converter **5732**. In an embodiment, this is performed using A/D Input Selector signals **5738**-**5746**.

[0580] Digital I/O bus **5704** carries data and control signals into and out of the digital control module **5700** from and to one or more processors or controllers that may be connected to the VPA. In an embodiment, some of control signals **5706**-**5730** are used to inform the digital control module **5700** of the type of information to expect on (or that is present on) digital I/O bus **5704**. For example, PC/(I/Q)n signal **5724** indicates to the digital control module **5700** whether power control information or I/Q data is being sent over digital I/O bus **5704**. Similarly, I/Qn signal **5720** indicates to the digital control module **5700** whether I or Q data is being sent over digital I/O bus **5704**.

[0581] Other control signals of the input interface **5602** of the VPA digital control module **5700** include Digital Enable/Disablen **5706**, PRGM/RUNn **5708**, READ/WRITEn **5710**, CLK OUT **5712**, CLK_IN×2 Enable/Disablen **5714**, CLK_IN×4 Enable/Disablen **5716**, CLK_IN **5718**, TX/RXn **5726**, SYNTH PRGM/SYNTH RUNn **5728**, and OUTPUT SEL/LATCHn **5730**.

[0582] Digital Enable/Disablen signal **5706** controls the power-up, reset, and shut down of the VPA. Signals to power-up, reset, or shut down the VPA typically come from a processor connected to the VPA. For example, when used in a cellular phone, a baseband processor or controller of the cellular phone may shut down the VPA in receive mode and enable it in transmit mode.

[0583] PRGM/RUNn signal **5708** indicates to the digital control module **5700** whether it is in programming or in run mode. In programming mode, the digital control module **5700** can be programmed to enable the desired operation of the VPA. For example, memory (RAM **5608**, NVRAM **5610**) bits of the digital control module **5700** can be programmed to indicate the standard to be used (e.g., WCDMA, EDGE, GSM, etc.) for communication. Programming of digital control module **5700** is done using digital I/O bus **5704**.

[0584] In an embodiment, the VPA is programmed and/or re-programmed (partially or completely) after it is installed in (or integrated with) the final product or device employing the VPA. For example, when used in a cellular phone, the VPA can be programmed after the cellular phone is manufactured to provide the cellular phone with new, additional, modified or different features, such as features related to (1) supported waveforms, (2) power control, (3) enhanced efficiency, and/or (4) power-up and power-down profiles. The VPA can also be programmed to remove waveforms or other features as desired by the network provider.

[0585] Programming of the VPA may be payment based. For example, the VPA may be programmed to include features and enhancements selected and purchased by the end-user.

[0586] In an embodiment, the VPA is programmed after the device is manufactured using any well known method or technique, including but not limited to: (1) programming the VPA using the programming interface of the device employing the VPA; (2) programming the VPA by storing programming data on a memory card readable by the device (a SIM card, for example, in the case of a cellular phone); and/or (3) programming the VPA by transferring programming data to the VPA wirelessly by the network provider or other source.

[0587] READ/WRITEn signal **5710** indicates to the digital control module **5700** whether data is to be read from or written to the digital control module storage (RAM **5608** or NVRAM **5610**) via digital I/O bus **5704**. When data is being read out of the digital control module **5700**, CLK OUT signal **5712** indicates timing information for reading from digital I/O bus **5704**.

[0588] CLK_IN signal **5718** provides a reference clock signal to the digital control module **5700**. Typically, the reference clock signal is selected according to the communication standards supported by the VPA. For example, in a dual-mode WCDMA/GSM system, it is desirable that the reference clock signal be a multiple of the WCDMA chip rate (3.84 MHz) and the GSM channel raster (200 KHz), with 19.2 MHz being a popular rate as the least common multiple of both. Further, CLK_IN signal **5718** can be made a multiple of the reference clock signal. In an embodiment, CLK_IN×2 Enable/Disablen **5714**, CLK_IN×4 Enable/Disablen **5716** can be used to indicate to the VPA digital control module **5700** that a multiple of the reference clock is being provided.

[0589] TX/RXn signal **5726** indicates to the digital control module **5700** when the system (e.g., cellular phone) employing the VPA is going into transmit or receive mode. In an embodiment, the digital control module **5700** is notified a short amount of time prior to the system going into transmit mode in order for it to power up the VPA. In another embodiment, the digital control module **5700** is notified when the system is going into receive mode in order for it to enter a sleep mode or to shutdown the VPA.

[0590] SYNTH PRGM/SYNTH RUNn signal **5728** is used to program the synthesizer that provides the reference frequency to the VPA (such as synthesizers **5918** and **5920** shown in FIG. **59**). When SYNTH PRGM **5728** is high, the VPA digital control module **5700** can expect to receive data for programming the synthesizer on digital I/O bus **5704**. Typically, programming of the synthesizer is needed when selecting the VPA transmission frequency. When SYNTH RUN **5728** goes high, the synthesizer is instructed to run. The synthesizer may be integrated with the VPA system or provided as an external component or subsystem.

[0591] OUTPUT SEL/LATCHn signal **5730** is used to select the VPA output to be used for transmission. This may or may not be needed depending on the number of outputs of the VPA. When OUTPUT SEL **5730** goes high, the digital control module **5700** expects to receive data for selecting the output on digital I/O bus **5704**. When LATCH **5730** goes high, the digital control module **5700** ensures that the VPA output used for transmission is held (cannot be changed) for the duration of the current transmit sequence.

[0592] The output interface **5604** of exemplary digital control module **5700** includes a plurality of data buses (**5748**, **5750**, **5752**, **5754**, **5756**, **5758**, **5760**, **5762**, **5764**, and **5766**), a programming bus **5799**, and a plurality of control signals (**5768**, **5770**, **5772**, **5744**, **5776**, **5778**, **5780**, **5782**, **5784**, **5786**, **5788**, **5790**, **5792**, **5794**, **5796**, and **5798**). In other embodiments of digital control module **5700**, the output interface **5604** may have more or less data buses, programming buses, and/or control signals.

[0593] Data buses **5752**, **5754**, **5756**, and **5758** carry digital information from the digital control module **5700** that is used to generate the substantially constant envelope signals in the analog core of the VPA. Note that exemplary digital control module **5700** may be used in a 4-Branch VPA embodiment (see Section 3.1) or a 2-Branch VPA embodiment (see Section 3.3). For example, digital information carried by data buses **5752**, **5754**, **5756**, and **5758** correspond to signals **722**, **724**, **726**, and **728** in the embodiment of FIG. **7**A or signals **1720**, **1722**, **1724**, and **1726** in the embodiment of FIG. **17**, and may be generated by the digital control module **5700** according to equations (5) (for a 4-Branch VPA embodiment) and (18) (for a 2-Branch VPA embodiment). Digital information carried by data buses **5752**, **5754**, **5756**, and **5758** is converted from digital to analog using respective Digital-to-Analog Converters (DACs **01-04**) to generate analog signals **5753**, **5755**, **5757**, and **5759**, respectively. Analog signals **5753**, **5755**, **5757**, and **5759** are input into vector modulators in the analog core of the VPA as will be further described below with reference to the VPA analog core implementations. In an embodiment, DACs **01-04** are controlled and synchronized by a Vector MOD DAC CLK signal **5770** provided by the digital control module. Further, DACs **01-04** are provided the same central reference voltage VREF_D signal **5743**.

[0594] Data buses **5760** and **5762** carry digital information from the digital control module **5700** that is used to generate bias voltage signals for the PA amplification stage and the driver amplification stage of the VPA (see FIG. **52** for illustration of different amplification stages of the VPA). In another embodiments additional control functions such as pre-driver Stage Bias Control is used. Digital information carried by data bus **5760** is converted from digital to analog using DAC_**05** to generate output stage bias signal **5761**. Similarly, digital information carried by data bus **5762** is converted from digital to analog using DAC_**06** to generate driver stage bias signal **5763**. Output stage bias signal **5761** and driver stage bias signal **5763** correspond, for example, to bias signals A and B illustrated in embodiment **5100**H. In an embodiment, DACs **05** and **06** are controlled and synchronized using an Autobias DAC CLK signal **5772**, and are provided the same central reference voltage VREF_E signal **5745**.

[0595] Data buses **5764** and **5766** carry digital information from the digital control module **5700** that is used to generate voltage control signals for the output stage and the driver stage of the VPA. Digital information carried by data bus **5764** is converted from digital to analog using DAC_**07** to generate output stage voltage control signal **5765**. Similarly, digital information carried by data bus **5766** is converted from digital to analog using DAC_**08** to generate driver stage voltage control signal **5767**. Output stage voltage control signal **5765** and driver stage voltage control **5767** are used to generate supply voltages for the output stage and the driver stage, providing a further method for controlling the voltage of the output stage and driver stage of the VPA. In an embodiment, DACs **07** and **08** are controlled and synchronized using a Voltage Control DAC CLK signal **5774**, and are provided the same central reference voltage VREF_F signal **5747**.

[0596] Data buses **5748** and **5750** carry digital information from the digital control module **5700** that is used to generate gain and phase balance control signals. In an embodiment, the gain and phase balance control signals are generated in response to feedback gain and phase information received from the analog core of the VPA on A/D IN bus **5702**. Digital information carried by data bus **5748** is converted from digital to analog using DAC_**09** to generate analog gain balance control signal **5749**. Similarly, digital information carried by data bus **5750** is converted from digital to analog using DAC_**10** to generate analog phase balance control **5751**. Gain and phase balance control signals **5749** and **5751** provide one mechanism for regulating gain and phase in the analog core of the VPA. In an embodiment, DACs **09** and **10** are controlled and synchronized using a Balance DAC CLK signal **5768**, and are provided the same central reference voltage VREF_B **5739**.

[0597] Programming bus **5799** carries digital instructions from the digital control module **5700** that are used to program frequency synthesizer or synthesizers in the analog core of the VPA. In an embodiment, digital instructions carried by programming bus **5799** are generated according to data received on digital I/O bus **5704**, when SYNTH PRGM signal **5728** is high. Digital instructions for programming the frequency synthesizers include instructions for setting the appropriate synthesizer (HI Band or Low Band) to generate a frequency according to the selected communication standard. In an embodiment, programming bus **5799** is a 3-wire programming bus.

[0598] In addition to the data and programming buses described above, the output interface **5604** includes a plurality of control signals.

[0599] In conjunction with programming bus **5799**, used for programming the frequency synthesizers of the analog VPA core, HI Band Enable/Disablen and Low Band Enable/Disablen control signals **5796** and **5798** are generated to control which of a high band frequency synthesizer and a low band frequency synthesizer of the analog VPA core is enabled/disabled.

[0600] Control signals **5738**, **5740**, **5742**, **5744**, and **5746** control an input selector for multiplexing feedback signals from the analog core of the VPA onto A/D IN input signal **5736** of A/D converter **5732**. In an embodiment, control signals **5738**, **5740**, **5744**, and **5746** control the multiplexing of a power output feedback signal, a differential branch amplitude feedback signal, and a differential branch phase feedback signal on A/D IN signal **5736**. Other feedback signals may be available in other embodiments. In an embodiment, the feedback signals are multiplexed according to a pre-determined multiplexing cycle. In another embodiment, certain feedback signals are periodically carried by A/D IN signal **5736**, while others are requested on-demand by the digital control module.

[0601] Output select control signals **5776**, **5778**, **5780**, **5782**, and **5784** are generated by the digital control module **5700** in order to select a VPA output, when the particular VPA implementation supports a plurality of outputs for different frequency bands and/or technology modes. In an embodiment, output select control signals **5776**, **5778**, **5780**, **5782**, and **5782** are generated according to digital control module input signal **5730**. In the example implementation of FIG. **57**, the digital control module **5700** provides five output select control signals for selecting one of five different VPA outputs. In an embodiment, output select control signals **5776**, **5778**, **5780**, **5782**, and **5784** control circuitry within the analog core of the VPA in order to power up circuitry corresponding to the selected VPA output and to power off circuitry corresponding to the remaining unselected VPA outputs. In embodiments, at any time, output select control signals **5776**, **5778**, **5780**, **5782**, and **5784** ensure that circuitry corresponding to a single VPA output are powered up, when the VPA is in transmit mode. A different digital control module embodiment may have more or less output select control signals depending on the particular number of VPA outputs supported by the particular analog core implementation.

[0602] Vector MOD HI Band(s)/Vector MOD Low Band (s)n control signal **5786** is generated by the digital control module **5700** to indicate whether a high band frequency modulation set or a low band frequency modulation set of vector modulators is to be used in the analog core of the VPA. In an embodiment, the high band and the low band vector modulators have different characteristics, allowing each set to be more suitable for a range of modulation frequencies. Control signal **5786** is generated according to the selected output of the VPA. In an embodiment, control signal **5786** controls circuitry within the analog core of the VPA in order to ensure that the selected set of vector modulators is powered up and that the other set(s) of vector modulators are powered off. In another embodiment, control signal **5786** controls circuitry within the analog core of the VPA in order to couple a set of interpolation filters to the selected set of vector modulators.

[0603] 3G HI Band/Normaln control signal **5788** is an optional control signal which may be used, if necessary, to enable the VPA to support the wide range High frequency band. In an embodiment, control signal **5788** may force more current through the output stage circuitry of the analog core and/or modify the output impedance characteristics of the VPA.

[0604] Filter Response **1**/Filter Response **2**n control signal **5790** is an optional control signal which may be used to dynamically change the response of interpolation filters in the analog core of the VPA. This may be needed as the interpolation filters have different optimal responses for different communication standards. For example, the optimal filter response has a 3 dB corner frequency around 5 MHz for WCDMA or EDGE, while this frequency is around 400 KHz for GSM. Accordingly, control signal **5790** allows for optimizing the interpolation filters according to the used communication standard.

[0605] Attenuator control signals **5792** and **5794** are optional control signals which may be used, if necessary, to provide additional output power control features and functions. For example, attenuator control signals **5792** and **5794** could be configured to enable/disable RF attenuators on the output of the VPA. These attenuators may be required based on the specific VPA implementation, which could be fabricated using Silicon, GaAs, or CMOS processes.

[0606] FIG. **58** illustrates another exemplary digital control module **5800** according to an embodiment of the present invention. Exemplary digital control module **5800** is similar in many respects to digital control module **5700**. In particular, both embodiments **5700**, **5800** have the same input interface **5602**, and substantial portions of the output interface (the output interface in FIG. **58** is labeled with reference number **5604'**). The differences between exemplary embodiments **5700** and **5800** relate to the type of feedback information being provided to the digital control module. Specifically, the two embodiments **5700** and **5800** are designed to operate with distinctly different feedback mechanisms for error correction. These mechanisms will be further described below in Section 4.3 with reference to the exemplary analog core implementations.

[0607] Exemplary implementation **5800** includes different input select control signals **5808**, **5810**, and **5812** compared to exemplary implementation **5700**. Input select control signals **5810** and **5812** control whether feedback information is to be received from the high band or the low band analog circuitry of the VPA, depending on which band is in use. Input select control signal I/Qn **5808** controls the multiplexing of I and Q feedback data from the analog core of the VPA. In an embodiment, control signal **5812** allows sequential switching between I data and Q data on A/D IN signal **5736**.

[0608] In further distinction to exemplary embodiment **5700**, exemplary embodiment **5800** include an additional data bus **5802**, which carries digital information from the digital control module **5800** used to generate an automatic gain control signal **5806**. Automatic gain control signal **5806**

is used to control the gain of an amplifier circuit used in the feedback mechanism in the analog core of the VPA. Further description of this component of the feedback mechanism will be provided below. In an embodiment, digital information carried by data bus **5802** is converted from digital to analog by DAC_**11** to generate analog signal **5806**. DAC_**11** is controlled by a clock signal **5804** provided by the digital control module, and is provided VREF_B signal **5739** as a central reference voltage.

[0609] It is noted that exemplary digital control modules **5700** and **5800** illustrate some of the typical input and output digital control module signals that may be used in a digital control module implementation. More or less input and output signals may also be used, as will be appreciated by a person skilled in the art based on the teachings herein, depending on the system in which the VPA is being used and/or the specific VPA analog core to be used with the digital control module. In an embodiment, exemplary digital control module implementations **5700** and **5800** may be used in conjunction with a VPA analog core using feedback only, feedforward only, or both feedback and feedforward error correction. When used in a feedforward only approach, feedback elements and/or signals (e.g., A/D IN **5702**, control signals **5738**, **5740**, **5742**, **5744**, **5746**, gain and phase balance control signals **5749** and **5751**) may be disabled or eliminated. Accordingly, variations of exemplary digital control module implementations **5700** and **5800** are within the scope of embodiments of the present invention.

4.3) VPA Analog Core

[0610] In this section, various exemplary implementations of the VPA analog core will be provided. As will be described below, the various exemplary implementations share a large number of components, circuits, and/or signals, with the main differences relating to the output stage architecture, the adopted error correction feedback mechanism, and/or the actual semiconductor material used in chip fabrication. As will be understood by a person skilled in the art based on the teachings herein, other VPA analog core implementations are also conceivable by interchanging, adding, and/or removing features among the various exemplary implementations described below. Accordingly, embodiments of the present invention are not to be limited to the exemplary implementations described herein.

4.3.1) VPA Analog Core Implementation A

[0611] FIG. **59** illustrates a VPA analog core implementation **5900** according to an embodiment of the present invention. In an embodiment, the input signals of analog core **5900** connect directly or indirectly (through DACs) to output signals from the output interface **5604** of digital control module **5600**. Similarly, feedback signals from analog core **5900** connect directly or indirectly (through DACs) to the input interface of the digital control module **5600**. For illustrative purposes, the analog core **5900** is shown in FIG. **59** as being connected to digital control module **5700**, as indicated by the same numeral signals on both FIG. **57** and FIG. **59**.

[0612] Analog core implementation **5900** is a 2-Branch VPA embodiment. This implementation **5900**, however, can be readily modified to a 4-Branch or a CPCP VPA embodiment, as will be apparent to persons skilled in the art based on the teachings herein.

[0613] At a high level, analog core **5900** includes an input stage for receiving data signals from the digital control module **5700**, a vector modulation stage for generating substantially constant envelope signals, and an amplification output stage for amplifying and outputting the desired VPA output signal. Additionally, analog core **5900** includes power supply circuitry for controlling and delivering power to the different stages of the analog core, optional output stage protection circuitry, and optional circuitry for generating and providing feedback information to the digital control module of the VPA.

[0614] The input stage of VPA analog core **5900** includes an optional interpolation filter bank (**5910**, **5912**, **5914**, and **5916**) and a plurality of switches **5964**, **5966**, **5968**, and **5970**. Interpolation filters **5910**, **5912**, **5914**, and **5916**, which may also serve as anti-aliasing filters, shape the analog outputs **5753**, **5755**, **5757**, and **5759** of DACs **01**-**04** to generate the desired output waveform. In an embodiment, the response of interpolation filters **5910**, **5912**, **5914**, and **5916** is dynamically changed using control signal **5790** from the digital control module **5700**. Digital control module signal **5790** may, for example, control switches within interpolation filters **5910**, **5912**, **5914**, and **5916** to cause a change in active circuitry (enable/disable RC circuitry) within filters **5910**, **5912**, **5914**, and **5916**. This may be needed as interpolation filters **5910**, **5912**, **5914**, and **5916** have different optimal responses for different communication standards. It should be noted that interpolation filters **5910**, **5912**, **5914**,and **5916** can be implemented using digital circuitry such as FIR filters or programmable FIR filters. When implemented digitally, these filters can be included within the VPA system or integrated with a baseband processor.

[0615] Subsequently, the outputs of interpolation filters **5910**, **5912**, **5914**, and **5916** are switched using switches **5964**, **5966**, **5968**, and **5970** to connect to either an upper band path **5964** or a lower band path **5966** of the VPA analog core **5900**. This determination between the upper and lower band paths is usually made by the digital control module **5700** based on the selected frequency range for transmission by the VPA. For example, the lower band path **5966** is used for GSM-**900**, while the upper band path **5964** is used for WCDMA. In an embodiment, switches **5964**, **5966**, **5968**, and **5970** are controlled by Vector MOD HI Band(s)/Vector MOD Low Band(s)n signal **5786**, provided by the digital control module **5700**. Signal **5786** controls the coupling of each of switches **5964**, **5966**, **5968**, and **5970** to respective first or second inputs, thereby controlling the coupling of the outputs of interpolation filters **5910**, **5912**, **5914**, and **5916** to the either the upper path **5964** or lower path **5966** of the VPA analog core **5900**.

[0616] The vector modulation stage of VPA analog core **5900** includes a plurality of vector modulators **5922**, **5924**, **5926**, and **5928**, divided between the upper band path **5964** and the lower band path **5966** of the analog core **5900**. Based on the selected band of operation, either the upper band path vector modulators (**5922**, **5924**) or the lower band path vector modulators (**5926**, **5928**) are active.

[0617] In an embodiment, the operation of vector modulators **5922**, **5924** or **5926**, **5928** is similar to the operation of vector modulators **1750** and **1752** in the embodiment of FIG. **17**, for example. Vector modulators **5922** and **5924** (or **5926** and **5928**) receive input signals **5919**, **5921**, **5923**, and **5925** (**5927**, **5929**, **5931**, and **5933**) from optional interpolation filters **5910**, **5912**, **5914**, and **5916**, respectively. Input signals **5919**, **5921**, **5923**, and **5925** (or **5927**, **5929**, **5931**, and **5933**) include amplitude information that is used to generate the constant envelope signals by the vector modulators. Further, vector modulators **5922** and **5924** (or **5926** and **5928**) receive a HI Band RF_CLK signal **5935** (LOW BAND RF_CLK signal **5937**) from a HI Band(s) Frequency Synthesizer **5918** (Low Band(s) Frequency Synthesizer **5920**). HI Band(s) Frequency Synthesizer **5918** (Low Band (s) Frequency Synthesizer **5920**) are optionally located externally or in the VPA analog core. In an embodiment, HI Band(s) Frequency Synthesizer **5918** (Low Band(s) Frequency Synthesizer **5920**) generates RF frequencies in the upper band range of 1.7-1.98 GHz (lower band range of 824-915 MHz). In another embodiment, HI Band(s) Frequency Synthesizer **5918** and Low Band(s) Frequency Synthesizer **5920** are controlled by digital control module signals **5796** and **5798**, respectively. Signals **5796** and **5798**, for example, power up the appropriate frequency synthesizer according to the selected transmission frequency band, and instruct the selected synthesizer to generate a RF frequency clock according to the selected transmission frequency.

[0618] Vector modulators **5922** and **5924** (or **5926** and **5928**) modulate input signals **5919**, **5921**, **5923**, and **5925** (**5927**, **5929**, **5931**, and **5933**) with HI BAND RF_CLK signal **5935** (LOW BAND RF_CLK signal **5937**). In an embodiment, vector modulators **5922** and **5924** (or **5926** and **5928**) modulate the input signals with appropriately derived and/or phase shifted versions of HI BAND RF_CLK signal **5935** (LOW BAND RF_CLK signal **5937**), and combine the generated modulated signals to generate substantially constant envelope signals **5939** and **5941** (**5943** and **5945**).

[0619] In another embodiment, vector modulators **5922** and **5924** (or **5926** and **5928**) further receive a phase balance control signal **5751** from the VPA digital control module. Phase balance control signal **5751** controls vector modulators **5922** and **5924** (or **5926** and **5928**) to cause a change in phase in constant envelope signals **5939** and **5941** (or **5943** and **5945**), in response to phase feedback information from the analog core. The amplitude and phase feedback mechanism is further discussed below. Optionally, upper band path vector modulators **5922** and **5924** also receive a 3G HI Band/Normaln signal **5788** from the digital control module. Signal **5788** can be used, if necessary, to further support driving the vector modulators at the highest frequencies of the upper band.

[0620] The output stage of VPA analog core **5900** includes a plurality of MISO amplifiers **5930** and **5932**, divided between the upper band path **5964** and the lower band path **5966** of the analog core **5900**. Based on the selected band of operation, either the upper band path MISO amplifier **5930** or the lower band path MISO amplifier **5932** is active.

[0621] In an embodiment, MISO amplifier **5930** (or **5932**) receives substantially constant envelope signals **5939** and **5941** (or **5943** and **5945**) from vector modulators **5922** and **5924** (or **5926** and **5928**). MISO amplifier **5930** (or **5932**) individually amplifies signals **5939** and **5941** (or **5943** and **5945**) to generate amplified signals, and combines the amplified signals to generate output signal **5947** (or **5949**). In an embodiment, MISO amplifier **5930** (or **5932**) combines the amplified signals via direct coupling, as described herein. Other modes of combining the amplified signals according to embodiments of the present invention have been described above in Section 3.

[0622] The output stage of VPA analog core **5900** is capable of supporting multi-band multi-mode VPA operation. As shown in FIG. **59**, the output stage includes two MISO amplifiers **5930** and **5932** for upper band and lower band operation, respectively. In addition, the output of each of the upper band **5964** and the lower band **5966** is further switched between one or more output paths according to the selected transmission mode (e.g., GSM, WCDMA, etc.). Typically, separate output paths are needed for different transmission modes since FDD-based modes (e.g., WCDMA) require the presence of duplexers at the output, while TDD-based modes (e.g., GSM, EDGE) have T/R switched outputs.

[0623] In analog core **5900**, the output **5947** of MISO amplifier **5930** can be coupled to one of three output paths **5954**, **5956**, and **5958**, with each output path **5954**, **5956**, **5958** being the one that is coupled to an antenna (not shown) or connector (not shown) for a particular mode of transmission. Similarly, the output **5949** of MISO amplifier **5932** can be coupled to one of two output paths **5960** and **5962**. In an embodiment, output select signals **5776**, **5778**, **5780**, **5782**, and **5784**, provided by the digital control module, control switches **5942** and **5944** to couple the output of the active MISO amplifier to the appropriate output path, based on the selected transmission mode. It is noted that more or less output paths **5954**, **5956**, **5958**, **5960**, and **5962** may be used.

[0624] Accordingly, with only two MISO amplifiers **5930** and **5932**, analog core **5900** supports multiple different transmission modes. In an embodiment, analog core **5900** allows for using a single MISO amplifier to support GSM, EDGE, WCDMA, and CDMA **2000**. It is clear therefore that one of the advantages of this exemplary VPA analog core according to implementation **5900** is in the reduction in the number of PAs per supported output paths This directly corresponds to a reduction in required chip area for the VPA analog core **5900**.

[0625] In an embodiment, the output stage of analog core **5900** receives optional output stage autobias signal **5761**, driver stage autobias signal **5763**, and gain balance control signal **5749** from the digital control module. Output stage autobias signal **5761** and driver stage autobias signal **5763** may or may not be needed according to the particular type of transistors used in the actual MISO implementation. In an embodiment, output stage autobias signal **5761** and driver stage autobias signal **5763** control the bias of MISO amplification stages to cause a change in the power output and/or the power efficiency of the VPA. Similarly, gain balance control signal **5749** may cause a change in the gain levels of different MISO amplification stages, in response to power output feedback information received by the digital control module from the analog core. Further discussion of these optional output stage input signals will be provided below.

[0626] In an embodiment, the output stage of analog core **5900** provides optional feedback signals to the digital control module **5700** of the VPA. Typically, these feedback signals are used by the digital control module **5700** to correct for amplitude and phase variations in branches of the VPA and/or for controlling the output power of the VPA. In the specific implementation of analog core **5900**, a differential feedback approach is employed to monitor for amplitude and phase variations, using a differential branch amplitude signal **5950** and a differential branch phase signal **5948** provided by the output stage. Further, output power monitoring is provided using signals PWR Detect A **5938** and

PWR Detect B **5940**, which measure the output power of MISO amplifiers **5930** and **5932**, respectively. Since only one of MISO amplifiers **5930** and **5932** can be active at any time, in an embodiment, PWR Detect A **5938** and PWR Detect **5940** are summed together using summer **5942**, to generate a signal that corresponds to the output power of the VPA.

[0627] In an embodiment, the feedback signals from the output stage are multiplexed using an input selector **5946** controlled by the digital control module **5700**. In another embodiment, the digital control module **5700** uses A/D Input Selector signals **5738**, **5740**, **5742**, **5744**, and **5746** to control input selector **5946** and select the feedback signal to be received. It is noted that monitoring of feedback signals may not need to occur in real-time rate and may only need to be performed periodically at a low rate. For example, for the purpose of branch amplitude and phase error correction, the rate at which feedback monitoring is performed depends on several factors such as the degree of feedforward correction being performed in the digital control module, process variations due to temperature, or operation changes such as changing battery or supply voltages.

[0628] Above, the tradeoffs between feedforward and feedback error compensation and/or correction techniques have been described. Accordingly, parameters governing the rates at which feedback monitoring is performed are design choices typically selected by the actual designer of the VPA. As a result, analog core implementation **5900** can be programmed to operate as a pure feedback implementation by disabling any feedforward correction in the digital control module, a pure feedforward implementation by disabling the monitoring of feedback signals, or as a hybrid feedforward/feedback implementation with variable feedforward/feedback utilization.

[0629] In an embodiment, the output stage of analog core **5900** includes optional output stage protection circuitry. In FIG. **59**, this is illustrated using VSWR (Voltage-Standing-Wave-Ratio) Protect circuitry **5934** and **5936** coupled respectively to MISO amplifiers **5930** and **5932**. VSWR protection circuitry **5934**, **5936** may or may not be needed depending on the actual MISO amplifier implementation. In an embodiment, VSWR Protect circuitry **5934** and **5936** protect the output stage PAs (see PAs **6030** and **6032** in FIG. **60**, for example) from going into thermal shutdown or device breakdown, when the output voltage level could cause the output stage breakdown voltage to be exceeded. In conventional systems, this is achieved by using an RF isolator at the output of the PAs, which is both expensive and lossy (typically causes around 1.5 dB in power loss). Accordingly, VSWR Protect circuitry **5934**, **5936** eliminate the need for isolators at the output stage, further reducing the cost, size, and power loss of the VPA. In an embodiment, VSWR Protect circuitry **5934**, **5936** enable an isolator-free output stage capable of supporting WCDMA. VSWR protection circuitry **5934** and **5936** also enable the VPA to operate into any VSWR level without damaging the VPA. VSWR protection circuitry can be designed to deliver the maximum output power of a particular implementation of a VPA into any VSWR level.

[0630] As described above, analog core **5900** includes power supply circuitry for controlling and delivering power to the different stages of the analog core **5900**. In one aspect, the power supply circuitry provides means for powering up active portions of the VPA analog core **5900**. In another aspect, the power supply circuitry provides means for controlling the power efficiency and/or the output power of the VPA.

[0631] In analog core implementation **5900**, the power supply circuitry includes MA Power Supply **5902**, Driver Stage Power Supply **5904**, Output Stage Power Supply **5906**, and Vector Mods Power Supply **5908**. In an embodiment, the power supply circuitry is controlled by output select signals **5776**, **5778**, **5780**, **5782**, and **5784**, provided by the digital control module **5700**.

[0632] MA Power Supply **5902** includes circuitry for controlling the powering up of active portions of the VPA analog core **5900**. In analog core **5900**, MA Power Supply **5902** has two outputs MA**1** VSUPPLY **5903** and MA**2** VSUPPLY **5905**. At any time, only one of MA**1** VSUPPLY **5903** or MA**2** VSUPPLY **5905** is active, ensuring that only the upper band **5964** or the lower band **5966** portion of the VPA analog core **5900** is powered up. In an embodiment, the active output of MA Power Supply **5902** is coupled to all active circuitry of the VPA analog core **5900**, with the exception of circuitry having unique power supply signals as described below. MA Power Supply **5902** receives output select signals from the digital control module, which enable one or the other of output signals MA**1** VSUPPLY **5903** or MA**2** VSUPPLY **5905**, based on the selected output of the VPA.

[0633] Driver Stage Power Supply **5904** includes circuitry for providing power to the driver stage circuitry of the MISO amplifiers **5930**, **5932**. Similar to MA Power Supply **5902**, Driver Stage Power Supply **5904** has two outputs MA**1** Driver VSUPPLY **5907** and MA**2** Driver VSUPPLY **5909**, with only one of the two outputs being active at any time. Driver Stage Power Supply **5904** is also controlled by output select signals **5776**, **5778**, **5780**, **5782**, and **5784** according to the selected output of the VPA. In addition, Driver Stage Power Supply **5904** receives a Driver Stage Voltage Control signal **5767** from the digital control module **5700**. In an embodiment, the outputs MA**1** Driver VSUPPLY **5907** and MA**2** Driver VSUPPLY **5909** are generated according to the received Driver Stage Voltage Control signal **5767**. In another embodiment, Driver Stage Voltage Control signal **5767** causes Driver Stage Power Supply **5904** to increase or decrease MA**1** Driver VSUPPLY **5907** or MA**2** Driver VSUPPLY **5909** to control the driver stage power amplification level. In another embodiment, Driver Stage Voltage Control signal **5767** is used by the digital control module **5700** to affect a change, using Driver Stage Power Supply **5904**, in the power supply voltage of the driver stage of the active MISO amplifier **5930** or **5932**, thereby controlling the power efficiency of the VPA.

[0634] Output Stage Power Supply **5906** includes circuitry for providing power to the PA stage circuitry of the MISO amplifiers **5930**, **5932**. Similar to MA Power Supply **5902**, Output Stage Power Supply **5906** has two outputs MA**1** Output Stage VSUPPLY **5911** and MA**2** Output Stage VSUPPLY **5913**, with only one of the two outputs being active at any time. Output Stage Power Supply **5906** is also controlled by output select signals **5776**, **5778**, **5780**, **5782**, and **5784** according to the selected output of the VPA. In addition, Output Stage Power Supply **5906** receives an Output Stage Voltage Control signal **5765** from the digital control module **5700**. In an embodiment, the outputs MA**1** Output Stage VSUPPLY **5911** and MA**2** Output Stage VSUPPLY **5913** are generated according to the received

Output Stage Voltage Control signal **5765**. In another embodiment, Output Stage Voltage Control signal **5765** causes Output Stage Power Supply **5906** to increase or decrease MA**1** Output Stage VSUPPLY **5911** or MA**2** Output Stage VSUPPLY **5913** to control the PA stage power amplification level. In another embodiment, Output Stage Voltage Control signal **5765** is used by the digital control module **5700** to affect a change, using Output Stage Power Supply **5906**, in the power supply voltage of the PA stage of the active MISO amplifier **5930** or **5932**, thereby controlling the power efficiency of the VPA.

[0635] Vector Mods Power Supply **5908** includes circuitry for providing power to the vector modulators **5922**, **5924**, **5926**, and **5928** of the analog core **5900**. In analog core **5900**, Vector Mods Power Supply **5908** has two outputs **5915** and **5917** for powering up the upper band vector modulators **5922** and **5924** and the lower band vector modulators **5926** and **5928**, respectively. At any time, only one of outputs **5915** or **5917** is active, ensuring that only the upper band or the lower vector modulators of the analog core **5900** are powered up. Vector Mods Power Supply **5908** receives a vector mod select signal **5786** from the digital control module **5700**, which controls which of its two outputs **5915** and **5917** is active, according to the selected transmission frequency requirements.

[0636] In addition to the above described power supply circuitry, analog core **5900** may optionally include voltage reference generator circuitry. The voltage reference generator circuitry may reside externally or within the VPA analog core **5900**. The voltage reference generator circuitry generates reference voltages for different circuits within the VPA. In an embodiment, as illustrated in FIG. **57**, the voltage reference generator circuitry provides reference voltages to DACs **01-10**, coupled to data outputs of the digital control module. In another embodiment, as illustrated in FIG. **59**, the voltage reference generator circuitry provides reference voltages to the interpolation filters and/or the vector modulators in the VPA analog core. In an embodiment, circuits of the same branch of the VPA are provided with the same reference voltage. For example, note that DACs **01** and **02**, interpolation filters **5910** and **5912**, and vector modulators **5922** and **5924**, which represent a VPA branch or data path, all share the same reference voltage VREF_C **5741**. For different implementations and system performance requirements, the voltage reference signals can be provided as a single reference voltage or multiple reference voltages.

[0637] FIG. **60** illustrates an output stage embodiment **6000** according to VPA analog core implementation **5900**. Output stage embodiment **6000** includes a MISO amplifier stage **6058**, an optional output switching stage (embodied by switch **6044**), and optional output stage protection and power detection circuitry.

[0638] In an embodiment, MISO amplifier stage **6058** corresponds to MISO amplifier **5930** in analog core **5900**. Accordingly, MA VSUPPLY signal **6006**, MA Driver VSUPPLY signal **6004**, and MA Output Stage VSUPPLY signal **6002** correspond respectively to signals **5903**, **5907**, and **5911** in FIG. **59**. Similarly, MA IN**1** and MA IN**2** input signals **6008** and **6010** and MA Output signals **6046**, **6048**, and **6050** correspond respectively to MISO input signals **5939** and **5941** and output signals **5954**, **5956**, and **5958** in FIG. **59**. PWR Detect signal **6023** corresponds to PWR Detect A signal **5938** in FIG. **59**. (Generally, implementation of MISO amplifier **5932** could also be based on MISO amplifier stage **6058** in FIG. **60**.)

[0639] MISO amplifier stage **6058** in embodiment **6000** includes a pre-driver amplification stage, embodied by Pre-Drivers **6012** and **6014**, a driver amplification stage, embodied by Drivers **6018** and **6020**, and a PA amplification stage, embodied by output stage PAs **6030** and **6032**. In an embodiment, substantially constant envelope input signals MA IN**1** **6008** and MA IN**2** **6010** are amplified at each stage of MISO amplifier **6058**, before being summed at the outputs of the PA stage.

[0640] In an embodiment, MISO amplifier stage **6058** is powered by power supply signals provided by voltage controlled power supply circuits. As described with reference to FIG. **59**, the power supply signals are generated by power supply circuitry of the VPA analog core **5900**. In an embodiment, the power supply signals are used to control the power supply voltages of the different amplification stages of MISO amplifier stage **6058**, thereby affecting the power efficiency of the VPA under various operating conditions. In another embodiment, the power supply signals are used to control the gain of each of the different amplification stages of MISO amplifier stage **6058**, thereby enabling a power control mechanism. Further, the power supply signals can be controlled independently of each other, allowing for independent control of power and/or efficiency for each of the different amplification stages of MISO amplifier stage **6058**. This independent control allows, for example, for shutting off one or more amplification stages of MISO amplifier **6058** according to the desired output power of the VPA. In FIG. **60**, the power supply signals are illustrated using signals **6002**, **6004**, and **6006**.

[0641] In an embodiment, MISO amplifier stage **6058** includes bias control circuitry. The bias control circuitry may be optional according to the particular MISO amplifier implementation. In an embodiment, the bias control circuitry provides a mechanism for controlling efficiency and/or power at each amplification stage of MISO amplifier **6058**. This mechanism is independent of the mechanism described above with reference to the power supply signals. Further, this mechanism provides for independently and individually controlling each amplification stage. In FIG. **60**, the bias control circuitry is illustrated using Gain Balance Control Circuitry **6016**, Driver Stage Autobias Circuitry **6022**, and Output Stage Autobias Circuitry **6028**.

[0642] In an embodiment, Gain Balance Control Circuitry **6016** is coupled to the inputs of the pre-driver amplification stage as illustrated in FIG. **60**. Gain Balance Control Circuitry **6016** receives a Gain Balance Control signal **5749** from the digital control module **5700** (through a DAC), and outputs input bias control signals **6013** and **6015**. Driver Stage Autobias Circuitry **6022** is coupled to the inputs of the driver amplification stage as illustrated in FIG. **60**. Driver Stage Autobias Circuitry **6022** receives Driver Stage Autobias signal **5763** from the digital control module **5700** (through a DAC), and outputs input bias control signals **6017** and **6019**. Similarly, Output Stage Autobias Circuitry **6028** is coupled to the inputs of the PA amplification stage as illustrated in FIG. **60**. Output Stage Autobias Circuitry **6028** receives Output Stage Autobias signal **5761** from the digital control module **5700** (through a DAC), and outputs input bias control signals **6029** and **6031**.

[0643] In an embodiment, the digital control module **5700** independently controls the bias of the pre-driver stage, the driver stage, and the PA stage of MISO amplifier **6058** using Gain Balance Control signal **5749**, Driver Stage Autobias signal **5763**, and Output Stage Autobias signal **5761**, respectively. In another embodiment, the digital control module **5700** may affect a change in the bias of the pre-driver stage, the driver stage, and/or the PA stage of MISO amplifier **6058** only using Gain Balance Control signal **5749**. As illustrated in FIG. **60**, Gain Balance Control Circuitry **6016** is coupled to Driver Stage Autobias Circuitry **6022** and Output Stage Autobias Circuitry **6028**. In an embodiment, a change in the overall gain of the VPA is affected by digital control module **5700** first by controlling the bias at the pre-driver stage. If further gain change is needed, bias control is performed at the driver stage, and subsequently at the PA stage.

[0644] In an embodiment, MISO amplifier stage **6058** includes circuits for enabling an error correction and/or compensation feedback mechanism. In output stage embodiment **6000**, a differential feedback mechanism is adopted, whereby Differential Branch Amplitude Measurement Circuitry **6024** and Differential Branch Phase Measurement Circuitry **6026** respectively measure differences in amplitude and phase between branches of MISO amplifier **6058**. In an embodiment, Differential Branch Amplitude Measurement Circuitry **6024** and Differential Branch Phase Measurement Circuitry **6026** are coupled at the inputs of the PA stage (PAs **6030** and **6032**) of MISO amplifier **6058**. In other embodiments, circuitry **6024** and **6026** may be coupled at the inputs of prior stages of MISO amplifier **6058**. In an embodiment, Differential Branch Amplitude Measurement Circuitry **6024** and Differential Branch Phase Measurement Circuitry **6026** respectively output Differential Branch Amplitude signal **5950** and Differential Branch Phase signal **5948**, which are fed back to digital control module **5700** (through A/D converters). Since digital control module **5700** knows at any particular time the correct differences in amplitude and/or phase between the branches of MISO amplifier **6058**, it may determine any errors in amplitude and/or phase based on Differential Branch Amplitude signal **5950** and Differential Branch Phase signal **5948**.

[0645] Output stage embodiment **6000** includes optional output stage protection circuitry. The output stage protection circuitry may or may not be needed according to the particular MISO amplifier implementation. In FIG. **60**, the output stage protection circuitry is illustrated using VSWR Protection Circuitry **6034**. In an embodiment, VSWR Protection Circuitry **6034** monitors the output of the PA stage, and controls the gain of MISO amplifier **6058** to protect PAs **6030** and **6032**. In embodiment **6000**, VSWR Protection Circuitry **6034** receives a signal **6036**, which is coupled either directly or indirectly to the output of the PA stage. In an embodiment, VSWR Protection Circuitry **6034** ensures that the voltage level at the output of the PA stage remains below a certain level, to prevent PAs **6030** and **6032** from going into thermal shutdown or experiencing device breakdown. In an embodiment, VSWR Protection Circuitry **6034** ensures that a breakdown voltage of PAs **6030** and **6032** is not exceeded. Accordingly, whenever the voltage level at the output of PAs **6030** and **6032** is above a pre-determined threshold, VSWR Protection Circuitry **6034** may cause a decrease in the gain of the MISO amplification stages. In an embodiment, VSWR Protection Circuitry **6034** is coupled to Balance Gain Control Circuitry **6016**, which in turn is coupled to both Driver Stage Autobias Circuitry **6022** and Output Stage Autobias Circuitry **6028**. In an embodiment, VSWR Protection Circuitry **6034** responds to a pre-determined voltage level at the output stage PAs by decreasing gain first at the pre-driver stage, then at the driver stage, and finally at the PA stage. As described above, VSWR Protection Circuitry **6034** may or may not be needed according to the particular MISO amplifier implementation. For example, a GaAs (Gallium Arsenide) MISO amplifier implementation would not require VSWR Protection Circuitry, as typical breakdown voltages of GaAs transistors are too large to be exceeded in many RF scenarios.

[0646] Output stage embodiment **6000** includes optional power detection circuitry. In an embodiment, the power detection circuitry serves as a means for providing power level feedback to the digital control module. In FIG. **60**, the power detection circuitry is illustrated using Power Detection Circuitry **6038**. In an embodiment, Power Detection Circuitry **6038** is coupled to the output of the PA stage of MISO amplifier **6058**. Power Detection Circuitry **6038** may be coupled directly or indirectly to the output of the PA stage as illustrated by signal **6040** in FIG. **60**. In an embodiment, Power Detection Circuitry **6038** outputs a PWR Detect signal **6023**. PWR Detect signal **6023** may be equivalent to PWR Detect A signal **5938** or PWR Detect B signal **5940** shown in FIG. **59**, which are fed back (through A/D converters) into the digital control module of the VPA. The digital control module uses PWR Detect signal **6023** to regulate the output power of the VPA as desired.

[0647] The optional output switching stage of output stage embodiment **6000** is embodied by a switch **6044** in FIG. **60**. In an embodiment, switch **6044** is coupled to one of three outputs **6046**, **6048**, or **6050** of the VPA. As described earlier, the switch is controlled by a set of output select signals **5776**, **5778**, and **5780**, provided by the digital control module. Switch **6044** is coupled to the proper output according to the select transmission mode and/or desired output frequency requirements (e.g., GSM, WCDMA, etc.).

[0648] Accordingly, pull-up impedance coupling at the output of the VPA can be done in various ways. In an embodiment, as shown in FIG. **60**, pull-up impedances **6052**, **6054**, and **6056** are respectively coupled between outputs **6046**, **6048**, and **6050** and MA Output Stage VSUPPLY **6002**. In another embodiment, a single pull-up impedance is used and is coupled between the output **6042** of the PA stage and MA Output Stage VSUPPLY **6002**. The advantage of the first approach lies in that, by placing the pull-impedance after the switch **6044**, the impedance characteristics of switch **6044** can be taken into account when selecting values for impedances **6052**, **6054**, and/or **6056**, allowing the VPA designer to exploit a further aspect to increase the efficiency of the VPA. On the other hand, the second approach requires a smaller number of pull-up impedances.

[0649] According to the particular MISO amplifier implementation, output stage embodiment **6000** may include more or less circuitry than to what is illustrated in FIG. **60**.

[0650] According to embodiments of the present invention, output stage embodiment **6000** including MISO amplifier stage **6058**, the optional output switching stage (switch **6044**), and the optional output protection and power detection circuitry may be fabricated using a SiGe (Silicon-Germanium) material. In another embodiment, MISO amplifier stage **6058** is fabricated using SiGe, and the output switching stage is fabricated using GaAs. In another embodiment, the PA stage (PAs **6030** and **6032**) and the output switching stage are fabricated using GaAs, while other circuitry of MISO amplifier stage **6058** and optional circuitry of the output stage are fabricated using SiGe. In another embodiment, the PA stage, the driver stage, and the output switching stage are fabricated using GaAs, while other circuitry of MISO amplifier stage **6058** and optional circuitry of the output stage are fabricated using SiGe. In another embodiment, the PA stage, the driver stage, the pre-driver stage, and the output switching stage are fabricated using GaAs. In another embodiment, the VPA system may be implemented using CMOS for all circuitry except for the output stage (**6030** or **6032**) which could be implemented in SiGe or GaAs material. In another embodiment, the VPA system may be implemented in its entirety in CMOS. Other variations and/or combinations of fabrication material(s) used for circuitry of the output stage are also possible, as can be understood by a person skilled in the art, and are therefore also within the scope of embodiments of the present invention.

[0651] Accordingly, as different semiconductor materials have different costs and performance, embodiments of the present invention provide a variety of VPA designs encompassing a wide range of cost and performance options.

4.3.2) VPA Analog Core Implementation B

[0652] FIG. **61** illustrates an alternative VPA analog core implementation **6100** according to an embodiment of the present invention. For illustrative purposes, the VPA analog core **6100** is shown in FIG. **61** as being connected to digital control module **5700**, although alternatively other digital control modules could be used. The physical connection between analog core **6100** and digital control module implementation **5700** is illustrated in FIG. **61**, as indicated by the same numeral signals on both FIG. **57** and FIG. **61**.

[0653] Analog core implementation **6100** is corresponds to a 2-Branch VPA embodiment. This implementation, however, can be readily modified to a 4-Branch or a CPCP VPA embodiment, as will be apparent to persons skilled in the art based on the teachings herein.

[0654] Analog core implementation **6100** has the same input stage and vector modulation stage as analog core implementation **5900**, described above. Accordingly, similar to analog core implementation **5900**, analog core **6100** includes an upper band path **5964** and a lower band path **5966** for upper band and lower band operation of the VPA, respectively.

[0655] One of the differences between analog core **5900** and analog core **6100** lies in the output stage of the VPA. In contrast to the output stage of analog core **5900**, which includes two MISO amplifiers **5930** and **5932**, the output stage of analog core **6100** includes five MISO amplifiers **6126**, **6128**, **6130**, **6132**, and **6134**, divided between the upper band path **5964** and the lower band path **5966** of the analog core. In an embodiment, the output stage includes a combination of SiGe and GaAs MISO amplifiers. In an embodiment, the upper band path **5964** includes three MISO amplifiers **6126**, **6128**, and **6130**, and the lower band path **5966** includes two MISO amplifiers **6132** and **6134**. Based on the selected band of operation, a single MISO amplifier, either in the upper band path **5964** or the lower band path **5966**, is active. In an embodiment, each of MISO amplifiers **6126**, **6128**, **6130**, **6132**, and **6134** can be dedicated to a single transmission mode (e.g., WCDMA, GSM, EDGE, etc.) of the VPA. This is in contrast to analog core **5900**, where each of MISO amplifiers **5930** and **5932** supports more than one transmission modes. Advantages and disadvantages of each architecture will be further discussed below.

[0656] As a result of having more than one MISO amplifiers per path, a switching stage is needed to couple the vector modulation stage to the MISO amplifiers in analog core **6100**. In FIG. **61**, this is illustrated using switches **6118**, **6120**, **6122**, and **6124**. In an embodiment, according to the selected transmission mode, switches **6118** and **6120** couple the outputs **5939** and **5941** of vector modulators **5922** and **5924** to one of MISO amplifiers **6126**, **6128**, and **6130**. Similarly, switches **6122** and **6124** couples the outputs **5943** and **5945** to one of MISO amplifiers **6132** and **6134**, according to the selected transmission mode and/or frequency requirements.

[0657] In an embodiment, MISO amplifier **6126** (or **6128**, **6130**, **6132**, **6134**) receives constant envelope signals **6119** and **6121** (or **6123** and **6125**, **6127** and **6129**, **6131** and **6133**, **6135** and **61137**). MISO amplifier **6126** (or **6128**, **6130**, **6132**, **6134**) individually amplifies signals **6119** and **6121** (or **6123** and **6125**, **6127** and **6129**, **6131** and **6133**, **6135** and **6137**) to generate amplified signals, and combines the amplified signals to generate output signal **6141** (**6144**, **6146**, **6148**, **6150**). In an embodiment, MISO amplifier **6126** (or **6128**, **6130**, **6132**, **6134**) combines the amplified signals via direct coupling, as described herein. Other modes of combining the amplified signals according to embodiments of the present invention have been described above in Section 3.

[0658] The output stage of VPA analog core **6100** is capable of supporting multi-band multi-mode VPA operation. Further, since the output stage of analog core **6100** can dedicate one MISO amplifier for each supported transmission mode, the output switching stage (embodied in analog core **5900** by switches **5942** and **5944**) can be eliminated. This results in a more efficient output stage (no power loss due switching stage), but at the expense of a larger chip area. This summarizes the main tradeoff between the architecture of analog core **5900** and that of analog core **6100**.

[0659] In an embodiment, the output stage of analog core **6100** receives optional bias control signals from digital control module **5700**. These are output stage autobias signal **5761**, driver stage autobias signal **5763**, and gain balance control signal **5749**, which have been described above with reference to analog core **5900**.

[0660] In an embodiment, the output stage of analog core **6100** provides optional feedback signals to digital control module **5700** of the VPA. These feedback signals include Differential Branch Amplitude signal **5950** and Differential Branch Phase signal **5948**, described above with reference to analog core **5900**, to enable a differential feedback approach to monitor for amplitude and phase variations in branches of the VPA. Also, similar to analog core **5900**, output power monitoring is provided using PWR Detect signals **6152**, **6154**, **6156**, **6158**, and **6160**, each of which measuring one of outputs **6142**, **6144**, **6146**, **6148**, and **6150** of the VPA. Since only one of the VPA outputs can be active at any time, PWR Detect signals **6152**, **6154**, **6156**, **6158**, and **6160** are summed together, in an embodiment, using summer **5952**, to generate a signal that corresponds to the current output power of the VPA.

[0661] Similar to analog core **5900**, the feedback signals from the output stage are multiplexed using an input selector **5946** controlled by the digital control module. Other aspects of the multiplexing of the feedback signals are described above with reference to analog core **5900**.

[0662] Similar to analog core **5900**, analog core **6100** can be designed to operate as a pure feedback implementation by disabling any feedforward correction in the digital control module, a pure feedforward implementation by disabling the monitoring of feedback signals, or as a hybrid feedforward/feedback implementation with variable feedforward/feedback utilization.

[0663] In an embodiment, the output stage of analog core **6100** includes optional output stage protection circuitry. In FIG. **61**, this is illustrated using VSWR (Voltage-Standing-Wave-Ratio) Protect circuitry **6136**, **6138**, and **6140** coupled respectively to MISO amplifiers **6128**, **6130**, and **6134**. VSWR protection circuitry may or may not be needed depending on the actual MISO amplifier implementation. For example, note that MISO amplifiers **6126** and **6132**, which are GaAs amplifiers, require no VSWR protection circuitry for many applications. Functions and advantages of VSWR Protection circuitry according to embodiments of the present invention are described above with reference to analog core **5900**.

[0664] Analog core **6100** includes power supply circuitry for controlling and delivering power to the different stages of the analog core. In one aspect, the power supply circuitry provides means for powering up active portions of the VPA analog core. In another aspect, the power supply circuitry provides means for controlling the power efficiency and/or the output power of the VPA.

[0665] The power supply circuitry of analog core **6100** is substantially similar to the power supply circuitry of analog core **5900**, with the difference being that analog core **6100** includes five MISO amplifiers as opposed to two in analog core **5900**. In FIG. **61**, the power supply circuitry is embodied in GMA and MA Power Supply circuitry **6102**, Driver Stage Power Supply circuitry **5904**, Output Stage Power Supply circuitry **5908**, and Vector Mods Power Supply circuitry **5908**. Each of circuitry **6102**, **5904**, and **5906** has five output power supply signals, with a single one of these five output signals being active at any time, according to the active MISO amplifier of the VPA. Function and operation of the power supply circuitry of analog core **6100** are substantially similar to those of the power supply circuitry of analog core **5900**, described above.

[0666] FIG. **62** illustrates an output stage embodiment **6200** according to VPA analog core implementation **6100**. Output stage embodiment **6200** includes a MISO amplifier stage **6220** and optional output stage protection and power detection circuitry.

[0667] MISO amplifiers **6126**, **6128**, **6130**, **6132** and/or **6134** shown in FIG. **61** can be implemented using an amplifier such as MISO amplifier stage **6220**.

[0668] Output stage embodiment **6200** is substantially similar to output stage embodiment **6000** illustrated in FIG. **60**, with the main difference being in the elimination of the output switching stage (embodied by switch **6044** in FIG. **60**) in embodiment **6200**.

[0669] Similar to embodiment **6000**, MISO amplifier stage **6220** in embodiment **6200** includes a pre-driver amplification stage, embodied by Pre-Drivers **6206** and **6208**, a driver amplification stage, embodied by Drivers **6210** and **6212**, and a PA amplification stage, embodied by output stage PAs **6214** and **6216**. In an embodiment, substantially constant envelope input signals MA IN1 **6202** and MA IN **6204** are amplified at each stage of MISO amplifier **6220**, before being summed at the outputs of the PA stage. Input signals MA IN1 **6202** and MA IN **6204** correspond to signals **6123** and **6125** in FIG. **61**, for example.

[0670] In an embodiment, MISO amplifier stage **6220** of output stage embodiment **6200** is powered by power supply signals provided by voltage controlled power supply circuits. In another embodiment, MISO amplifier stage **6220** includes optional bias control circuitry controllable by the digital control module. In another embodiment, MISO amplifier stage **6220** includes circuits for enabling an error correction and/or compensation feedback mechanism. In another embodiment, output stage embodiment **6000** includes optional output stage protection circuitry and power detection circuitry. These aspects (power supply, bias control, error correction, output protection, and power detection) of output stage embodiment **6200** are substantially similar to what have been described above with respect to output stage embodiment **6000**.

[0671] According to embodiments of the present invention, output stage embodiment **6200** may be fabricated using a SiGe (Silicon-Germanium) material including MISO amplifier stage **6220** and the optional output protection and power detection circuitry. In another embodiment, MISO amplifier stage **6220** is fabricated using SiGe in its entirety. In another embodiment, the PA stage (PAs **6214** and **6216**) of MISO amplifier stage **6220** is fabricated using GaAs, while other circuitry of MISO amplifier stage **6220** and optional circuitry of the output stage are fabricated using SiGe. In another embodiment, the PA stage and the driver stage (Drivers **6210** and **6212**) of MISO amplifier stage **6220** are fabricated using GaAs, while other circuitry of MISO amplifier stage **6220** and optional circuitry of the output stage are fabricated using SiGe. In another embodiment, the PA stage, the driver stage, and the pre-driver stage (Pre-Drivers **6206** and **6208**) are fabricated using GaAs. In another embodiment, the VPA system may be implemented using CMOS for all circuitry except for the output stage (**6030** or **6032**) which could be implemented in SiGe or GaAs material. In another embodiment, the VPA system may be implemented in its entirety in CMOS. Other variations and/or combinations of fabrication material(s) used for circuitry of the output stage are also possible, as can be understood by a person skilled in the art, and are therefore also within the scope of embodiments of the present invention. Further, output stages within the same the VPA may be fabricated using different material, as illustrated in FIG. **61** for example, where MISO amplifiers **6128**, **6130**, and **6134** are SiGe amplifiers and MISO amplifiers **6126** and **6132** are GaAs amplifiers (one or more stages of their output stage are GaAs).

4.3.3) VPA Analog Core Implementation C

[0672] FIG. **63** illustrates another VPA analog core implementation **6300** according to an embodiment of the present invention. For illustrative purposes, example analog core **6300** is shown in FIG. **63** as being connected to digital control module **5800**, although other digital control modules could alternatively be used. The physical connection between analog core **6300** and digital control module **5800** is indicated by the same numeral signals on both FIG. **58** and FIG. **63**.

[0673] Analog core implementation **6300** corresponds to a 2-Branch VPA embodiment. This implementation, however, can be readily modified to a 4-Branch or a CPCP VPA embodiment, as will be apparent to a person skilled in the art based on the teachings herein.

[0674] Analog core implementation **6300** includes similar input stage, vector modulation stage, and amplification output stage as analog core **5900** of FIG. **59**. Function, operation and control of these stages is described above with reference to FIG. **59**.

[0675] Similar to analog core **5900**, analog core **6300** includes a feedback error correction and/or compensation mechanism. In contrast to analog core **5900**, however, analog core **6300** employs a receiver-based feedback mechanism, as opposed to a differential feedback mechanism in analog core **5900**. A receiver-based feedback mechanism is one that is based on having a receiver that receives the active output of the VPA, generates I data and Q data from the received output, and feeds back the generated I and Q data to the digital control module. By estimating the delay between the input and the output of the VPA, the feedback I and Q signals can be properly aligned with their corresponding input I and Q signals. In another embodiment, the receiver feedback includes the complex output signal (magnitude and phase polar information) instead of Cartesian I and Q data signals.

[0676] In an embodiment, this is done by coupling a receiver (not shown) at the active output of the VPA (**5947** or **5949**). In FIG. **63**, signals **6302** and **6304** respectively represent upper band and lower band RF inputs into the receiver. Only one of signals **6302** and **6304** can be active at any time, depending on whether the upper band path **5964** or the lower band path **5966** of analog core **6300** is being used. Similarly, the receiver-based feedback mechanism includes an upper band path and a lower band path. In an embodiment, each of the upper band and lower band feedback paths include an Automatic Gain Controller (AGC) (**6306** and **6308**), I/Q sample-and-hold (S/H) circuitry (**6314**, **6316** and **6318**, **6320**), switching circuitry (**6322** and **6324**), and optional interpolation filters (**6326** and **6328**). In an embodiment, a switch **6330**, controlled by the digital control module by means of input select signals **5810** and **5812**, couples either the upper band or the lower band feedback paths to the digital control module. Further, based on the coupled feedback path, digital control module I/Qn Select signal **5808** controls switching circuitry **6322** or **6324** to alternate the coupling of I data and Q data to the digital control module. Other implementations are also possible as can be understood by a person skilled in the art based on the teachings herein.

[0677] In an embodiment, the AGC circuitry is used to allow the receiver to feedback useful I and Q information over a wide dynamic range of VPA output power. For example, output signals **5954**, **5956**, **5958**, **5960**, and **5962** can vary from +35 dBm to −60 dBm in certain cell phone applications. For I and Q data to contain accurate feedback information, the I and Q output of the receiver needs to be scaled to utilize the majority of the input voltage range of the A/Din signal **5736**, independently of the output signal power. Digital Control module **5800** is designed to control the VPA to the required output power, which allows digital control module **5800** to determine an appropriate receiver gain to achieve the proper A/D input voltage which is digitized through A/D **5732**.

[0678] A VPA analog core with a receiver-based feedback mechanism can be implemented as a pure feedback, feedforward, or hybrid feedback/feedforward system. As described above, a pure feedback implementation requires a minimal amount of or no memory (RAM **5608**, NVRAM **5610**) in the digital control module. This may represent one advantage of an analog core implementation according to analog core **6300**, in addition to the elimination of differential feedback measurement circuitry from the analog core. Nonetheless, analog core **6300** can be programmed to operate as a pure feedback implementation by disabling any feedforward correction in digital control module **5800**, a pure feedforward implementation by disabling the monitoring of feedback signals, or as a hybrid feedforward/feedback implementation with variable feedforward/feedback utilization.

[0679] In an embodiment, the output stage of analog core **6300** includes optional output stage protection circuitry. This is not shown in FIG. **63**, but has been described above with respect to analog core implementations **5900** and **6100**. Other aspects of analog core **6300** (bias control, power supply, etc.) are substantially similar to analog core **5900**, and are described above with reference to FIG. **59**.

[0680] FIG. **64** illustrates an output stage embodiment **6400** according to VPA analog core implementation **6300**. Output stage embodiment **6400** includes a MISO amplifier stage **6434** and an output switching stage. In an embodiment, MISO amplifier stage **6434** corresponds to MISO amplifier **5930** and/or **5932**, shown in FIG. **63** (that is, either or both of MISO amplifiers **5930**, **5932** can be implemented using an amplifier such as MISO amplifier stage **6434**).

[0681] Output stage embodiment **6400** is substantially similar to output stage embodiment **6000** illustrated in FIG. **60**, with the main difference being in the elimination of the differential branch measurement circuitry (**6024** and **6026**) due to the use a receiver-based feedback mechanism.

[0682] Similar to embodiment **6000**, MISO amplifier stage **6434** in embodiment **6400** includes a pre-driver amplification stage, embodied by Pre-Drivers **6406** and **6408**, a driver amplification stage, embodied by Drivers **6410** and **6412**, and a PA amplification stage, embodied by output stage PAs **6414** and **6416**. In an embodiment, constant envelope input signals MA IN1 **6402** and MA IN **6404** are amplified at each stage of MISO amplifier stage **6434**, before being summed at the outputs of the PA stage of MISO amplifier stage **6434**.

[0683] In an embodiment, MISO amplifier stage **6434** of output stage embodiment **6400** is powered by power supply signals provided by voltage controlled power supply circuits. In another embodiment, MISO amplifier stage **6434** includes optional bias control circuitry controllable by the digital control module. In another embodiment, output stage embodiment **6400** includes optional output stage protection circuitry (not shown in FIG. **64**). These aspects (power supply, bias control, and output protection) of output stage embodiment **6400** are substantially similar to what have been described above with respect to output stage embodiment **6000**.

[0684] According to embodiments of the present invention, output stage embodiment **6400** may be fabricated using a SiGe (Silicon-Germanium) material including the MISO amplifier stage **6434**, the output switching stage **6420**, and the optional output protection circuitry. In another embodiment, MISO amplifier stage **6434** is fabricated using SiGe, and the output switching stage **6420** is fabricated using GaAs. In another embodiment, the PA stage (PAs **6414** and **6416**) of MISO amplifier stage **6434** and the output switching stage **6420** are fabricated using GaAs, while other circuitry of MISO amplifier stage **6434** and optional circuitry of the output stage are fabricated using SiGe. In another embodiment, the PA stage, the driver stage (Drivers **6410** and **6412**), and the output switching stage **6420** are fabricated using GaAs, while other circuitry of MISO amplifier stage **6434** and optional circuitry of the output stage are fabricated using SiGe. In another embodiment, the PA stage, the driver stage, the pre-driver stage (Pre-drivers **6406** and **6408**), and the output switching stage **6420** are fabricated using GaAs. In another embodiment, the VPA system may be implemented using CMOS for all circuitry except for the output stage (**6030** or **6032**) which could be implemented in SiGe or GaAs material. In another embodiment, the VPA system may be implemented in its entirety in CMOS. Other variations and/or combinations of fabrication material(s) used for circuitry of the output stage are also possible, as can be understood by a person skilled in the art, and are therefore also within the scope of embodiments of the present invention. Further, output stages within the same the VPA may be fabricated using different material, as illustrated in FIG. **61** for example, where MISO amplifiers **6128**, **6130**, and **6134** are SiGe amplifiers and MISO amplifiers **6126** and **6132** are GaAs amplifiers (one or more stages of their output stage are GaAs).

5. Real-Time Amplifier Class Control of VPA Output Stage

[0685] According to embodiments of the present invention, a VPA output stage can be controlled to vary its amplifier class of operation according to changes in its output waveform trajectory. This concept is illustrated in FIG. **65** with reference to an exemplary WCDMA waveform. The graph in FIG. **65** illustrates a timing diagram of a WCDMA output waveform envelope versus the class of operation of the VPA output stage. Note that the output waveform envelope is directly proportional to the output power of the VPA output stage.

[0686] It is noted that the VPA output stage amplifier class traverses from a class S amplifier to a class A amplifier as the output waveform envelope decreases from its maximum value towards zero. At the zero crossing, the VPA output stage operates as a class A amplifier, before switching to higher class amplifier operation as the output waveform envelope increases.

[0687] One important problem overcome by this real-time ability to control the VPA output stage amplifier class of operation is the phase accuracy control problem. With regard to the example shown in FIG. **65**, the phase accuracy control problem lies in the fact that in order to produce high quality waveforms, at any given power level, a 40 dB of output power dynamic range is desirable. However, the phase accuracy required to produce a 40 dB output power dynamic range (around 1.14 degrees or 1.5 picoseconds) is well beyond the tolerance of practical circuits in high volume applications. As will be appreciated, the specific power ranges cited in this paragraph, and elsewhere herein, are provided solely for illustrative purposes, and are not limiting.

[0688] Embodiments according to the present invention solve the phase accuracy control problem by transiting multiple classes of operation based on waveform trajectory so as to maintain the best balance of efficiency versus practical control accuracy for all waveforms. In embodiments, the output power dynamic range of the VPA output stage exceeds 90 dB.

[0689] In an embodiment, at higher instantaneous signal power levels, the amplifier class in operation (class S) is highly efficient and phase accuracy is easily achieved using phase control. At lower instantaneous signal power levels, however, phase control may not be sufficient to achieve the required waveform linearity. This is illustrated in FIG. **66**, which shows a plot of the VPA output power (in dBm) versus the outphasing angle between branches of the VPA. It can be seen that at high power levels, a change in outphasing angle results in a smaller output power change than at lower power levels. Accordingly, phase control provides higher resolution power control at higher power levels than at lower power levels.

[0690] Accordingly, to support high resolution power control at lower power levels, other mechanisms of control are needed in addition to phase control. FIG. **67** illustrates exemplary power control mechanisms according to embodiments of the present invention using an exemplary QPSK waveform. The QPSK constellation is imposed on a unit circle in the complex domain defined by cos(wt) and sin(wt). The constellation space is partitioned between three concentric and non-intersecting regions: an outermost "phase control only" region, a central "phase control, bias control, and amplitude control" region, and an innermost "bias control and amplitude control" region. According to embodiments of the present invention, the outermost, central, and innermost regions define the type of power control to be applied according to the power level of the output waveform. For example, referring to FIG. **67**, at lower power levels (points falling in the innermost region), bias control and amplitude control are used to provide the required waveform linearity. On the other hand, at higher power levels (points falling in the outermost region), phase control (by controlling the outphasing angle) only is sufficient.

[0691] As can be understood by persons skilled in the art, the control regions illustrated in FIG. **67** are provided for purposes of illustration only and are not limiting. Other control regions can be defined according to embodiments of the present invention. Typically, but not exclusively, the boundaries of the control regions are based on the Complementary Cumulative Density Function (CCDF) of the desired output waveform and the sideband performance criteria. Accordingly, the control regions' boundaries change according to the desired output waveform of the VPA.

[0692] In embodiments, the power control mechanisms defined by the different control regions enable the transition of the VPA output stage between different class amplifiers. This is shown in FIG. **68**, which illustrates, side by side, the output stage amplifier class operation versus the output waveform envelope and the control regions imposed on a unit circle. FIG. **69** further shows the output stage current in response to the output waveform envelope. It is noted that the output stage current closely follows the output waveform envelope. In particular, it is noted that the output stage current goes completely to zero when the output waveform envelope undergoes a zero crossing.

[0693] FIG. **70** illustrates the VPA output stage theoretical efficiency versus the output stage current. Note that the output stage current waveform of FIG. **70** corresponds to the one shown in FIG. **69**. In an embodiment, the VPA output stage operates at 100% theoretical efficiency for 98% (or greater) of the time. It is also noted from FIG. **70** the transition of the output stage between different amplifier classes of operation according to changes in the output stage current.

[0694] FIG. **71** illustrates an exemplary VPA according to an embodiment of the present invention. For illustrative purposes, and not purposes of limitation, the exemplary embodiment of FIG. **71** will be used herein to further describe the various control mechanisms that can be used to cause the transitioning of the VPA output stage (illustrated as a MISO amplifier in FIG. **71**) between different amplifier classes of operation.

[0695] The VPA embodiment of FIG. **71** includes a transfer function module, a pair of vector modulators controlled by a frequency reference synthesizer, and a MISO amplifier output stage. The transfer function module receives I and Q data and generates amplitude information that is used by the vector modulators to generate substantially constant envelope signals. The substantially constant envelope signals are amplified and summed in a single operation using the MISO amplifier output stage.

[0696] According to embodiments of the present invention, the MISO amplifier output stage can be caused to transition in real time between different amplifier classes of operation according to changes in output waveform trajectory. In an embodiment, this is achieved by controlling the phases of the constant envelope signals generated by the vector modulators. In another embodiment, amplitudes of the MISO amplifier input signals are controlled using the transfer function. In another embodiment, the MISO amplifier inputs are biased (biasing of the MISO inputs can be done at any amplification stage within the MISO amplifier) using the transfer function to control the MISO amplifier class of operation. In other embodiments, combinations of these control mechanisms (phase, input bias and/or input amplitude) are used to enable the MISO amplifier stage to transition between different amplifier classes of operation.

[0697] FIG. **72** is a process flowchart **100** that illustrates a method for real-time amplifier class control in a power amplifier, according to changes in output waveform trajectory, according to an embodiment of the invention. Process flowchart **100** begins in step **110**, which includes determining an instantaneous power level of a desired output waveform. In an embodiment, the instantaneous power level is determined as a function of the desired output waveform envelope.

[0698] Based on the determined instantaneous power level, step **120** of process flowchart **100** includes determining a desired amplifier class of operation, wherein said amplifier class of operation optimizes the power efficiency and linearity of the power amplifier. In an embodiment, determining the amplifier class of operation depends on the specific type of desired output waveform (e.g., CDMA, GSM, EDGE).

[0699] Step **130** includes controlling the power amplifier to operate according to the determined amplifier class of operation. In an embodiment, the power amplifier is controlled using phase control, bias control, and/or amplitude control methods, as described herein.

[0700] According to process flowchart **100**, the power amplifier is controlled such that it transitions between different amplifier classes of operation according to the instantaneous power level of the desired output waveform. In other embodiments, the power amplifier is controlled such that it transitions between different amplifier classes of operation according the average output power of the desired output waveform. In further embodiments, the power amplifier is controlled such that it transitions between different amplifier classes of operation according to both the instantaneous power level and the average output power of the desired output waveform.

[0701] According to embodiments of the present invention, the power amplifier can be controlled to transition from a class A amplifier to a class S amplifier, while passing through intermediary amplifier classes (AB, B, C, and D).

[0702] Embodiments of the invention control transitioning of the power amplifier(s) to different amplifier classes as follows:

[0703] To achieve a class A amplifier, the drive level and bias of the power amplifier are controlled so that the output current conduction angle is equal to 360 degrees. The conduction angle is defined as the angular portion of a drive cycle in which output current is flowing through the amplifier.

[0704] To achieve a class AB amplifier, the drive level and bias of the power amplifier are controlled so that the output current conduction angle is greater than 180 degrees and less than 360 degrees.

[0705] To achieve a class B amplifier, the drive level and bias of the power amplifier are controlled so that the output current conduction angle is approximately equal to 180 degrees.

[0706] To achieve a class C amplifier, the drive level and bias of the power amplifier are controlled so that the output current conduction angle is less than 180 degrees.

[0707] To achieve a class D amplifier, the drive level and bias of the power amplifier are controlled so that the amplifier is operated in switch mode (on/off).

[0708] To achieve a class S amplifier, the amplifier is controlled to generate a Pulse Width Modulated (PWM) output signal.

[0709] In an embodiment, the above described real-time amplifier class control of the VPA output stage is accompanied by a dynamic change in the transfer function being implemented in the digital control module of the VPA. This is further described below with respect to FIGS. **73-77**.

[0710] FIG. **73** illustrates an example VPA output stage according to an npn implementation with two branches. Each branch of the VPA output stage receives a respective substantially constant envelope signal. The substantially constant envelope signals are illustrated as IN**1** and IN**2** in FIG. **73**. Transistors of the VPA output stage are coupled together by their emitter nodes to form an output node of the VPA.

[0711] When the VPA output stage operates as a class S amplifier, it effectuates Pulse Width Modulation (PWM) on the received substantially constant envelope signals IN**1** and IN**2**. A theoretical equivalent circuit of the VPA output stage in this amplifier class of operation is illustrated in FIG. **74**. Note that transistors of the VPA output stage are equivalent to switching amplifiers in this class of operation. The output of the VPA as a function of the outphasing angle θ between the substantially constant envelope signals IN**1** and IN**2** (assuming that IN**1** and IN**2** have substantially equal amplitude of value A) is given by

$$SQ(\theta) = A\frac{\pi - \theta}{2\pi}.$$

A plot of this function, described previously as the magnitude to phase shift transform, is illustrated in FIG. 76.

[0712] On the other hand, when the VPA output stage operates as a class A amplifier, it emulates a perfect summing node. A theoretical, equivalent circuit of the VPA output stage in this amplifier class of operation is illustrated in FIG. **75**. Note that transistors of the VPA output stage are equivalent to current sources in this class of operation. The output of the VPA as function of the outphasing angle θ between the substantially constant envelope signals IN**1** and IN**2** (assuming that IN**1** and IN**2** have substantially equal amplitude of value A) is given by $R(\theta)=AA\sqrt{2(1+\cos(\theta))}$. A plot of this function, described previously as the magnitude to phase shift transform, is illustrated in FIG. **76**.

[0713] According to an embodiment of the present invention, amplifier classes of operation A and S represent two extremes of the amplifier operating range of the VPA output stage. However, as described above, the VPA output stage may transition a plurality of other amplifier classes of operation including, for example, classes AB, B, C, and D. Accordingly, the transfer function implemented by the digital control module of the VPA varies within a spectrum of magnitude to phase shift transform functions, with the transform functions illustrated in FIG. **76** representing the boundaries of this spectrum. This is shown in FIG. **77**, which illustrates a spectrum of magnitude to phase shift transform functions corresponding to a range of amplifier classes of operation of the VPA output stage. FIG. **77** illustrates 6 functions corresponding to the six amplifier classes of operation A, AB, B, C, D, and S. In general, however, an infinite number of functions can be generated using the functions corresponding to the two extreme classes of operation A and S. In an embodiment, this is performed using a weighted sum of the two functions and is given by $(1-K)\times R(\theta)+K\times SQ(\theta)$, with $0\leqq K\leqq 1$.

6. Summary

[0714] Mathematical basis for a new concept related to processing signals to provide power amplification and up-conversion is provided herein. These new concepts permit arbitrary waveforms to be constructed from sums of waveforms which are substantially constant envelope in nature. Desired output signals and waveforms may be constructed from substantially constant envelope constituent signals which can be created from the knowledge of the complex envelope of the desired output signal. Constituent signals are summed using new, unique, and novel techniques not available commercially, not taught or found in literature or related art. Furthermore, the blend of various techniques and circuits provided in the disclosure provide unique aspects of the invention which permits superior linearity, power added efficiency, monolithic implementation and low cost when compared to current offerings. In addition, embodiments of the invention are inherently less sensitive to process and temperature variations. Certain embodiments include the use of multiple input single output amplifiers described herein.

[0715] Embodiments of the invention can be implemented by a blend of hardware, software and firmware. Both digital and analog techniques can be used with or without microprocessors and DSP's.

[0716] Embodiments of the invention can be implemented for communications systems and electronics in general. In addition, and without limitation, mechanics, electro mechanics, electro optics, and fluid mechanics can make use of the same principles for efficiently amplifying and transducing signals.

7. Conclusion

[0717] The present invention has been described above with the aid of functional building blocks illustrating the performance of specified functions and relationships thereof. The boundaries of these functional building blocks have been arbitrarily defined herein for the convenience of the description. Alternate boundaries can be defined so long as the specified functions and relationships thereof are appropriately performed. Any such alternate boundaries are thus within the scope and spirit of the claimed invention. One skilled in the art will recognize that these functional building blocks can be implemented by discrete components, application specific integrated circuits, processors executing appropriate software and the like and combinations thereof.

[0718] While various embodiments of the present invention have been described above, it should be understood that they have been presented by way of example only, and not limitation. Thus, the breadth and scope of the present invention should not be limited by any of the above-described exemplary embodiments, but should be defined only in accordance with the following claims and their equivalents.

1. A method of power transmission, modulation, and amplification including output waveform distortion compensation, comprising:

receiving a frequency reference signal;

receiving information;

generating a plurality of control signals from said received information;

generating a plurality of substantially constant envelope signals using said frequency reference signal and said control signals;

amplifying and combining said substantially constant envelope signals to create an output signal; and

reducing waveform distortion in said output signal by reducing phase and amplitude errors in said output signal relative to a desired output signal.

2. The method according to claim **1**, wherein said reducing waveform distortion step comprises:

generating phase and amplitude error values in said output signal relative to said desired output signal.

3. The method according to claim **2**, wherein said step of generating phase and amplitude error values comprises one or more of:

measuring said phase and amplitude error values in said output signal relative to said desired output signal;

calculating said phase and amplitude error values in said output signal relative to said desired output signal; and

estimating said phase and amplitude error values in said output signal relative to said desired output signal.

4. The method according to claim **1**, wherein said reducing waveform distortion step comprises reducing phase and amplitude errors in said output signal due to one or more of:

(a) circuitry process variations;
(b) circuitry temperature variations;
(c) IC package variations;
(d) circuit board variations;
(e) said step of generating a plurality of control signals;
(f) said step of generating a plurality of substantially constant envelope signals;
(g) amplifying said substantially constant envelope signals; and
(h) combining said amplified substantially constant envelope signals at a single power amplifier node.

5. The method according to claim **4**, wherein phase and amplitude errors due to one or more of (a)-(h) are estimated at system manufacture time.

6. The method according to claim **4**, wherein phase and amplitude errors due to one or more of (a)-(h) are at least one of measured, calculated, and estimated during real-time system operation.

7. The method according to claim **4**, wherein said phase and amplitude errors due to (e) are caused by vector modulation circuitry that performs said step of generating a plurality of substantially constant envelope signals.

8. The method according to claim **4**, wherein said phase and amplitude errors due to (g) result in gain and phase imbalances between power amplifier branches.

9. The method according to claim **1**, wherein said reducing waveform distortion step is performed when output waveform distortion at least satisfies system design requirements.

10. The method according to claim **1**, wherein said reducing waveform distortion step is performed to ensure that phase and amplitude errors in said output signal remain within a specified system error tolerance.

11. The method according to claim **1**, wherein said reducing waveform distortion step is performed by pre-compensating for said phase and amplitude errors in said output signal relative to said desired output signal.

12. The method according to claim **3**, wherein said reducing waveform distortion step comprises feedforward error compensation that includes pre-compensating for said estimated phase and amplitude error values in said output signal.

13. The method according to claim **3**, wherein said reducing waveform distortion step comprises feedback error correction that includes correcting for said measured/calculated phase and amplitude error values in said output signal.

14. The method according to claim **3**, wherein said reducing waveform distortion step comprises hybrid feedforward/feedback error compensation that includes at least one of pre-compensating for estimated phase and amplitude error values in said output signal and correcting for measured/calculated phase and amplitude error values in said output signal.

15. The method according to claim **1**, wherein said reducing waveform distortion step comprises adjusting transfer functions that are used to generate said plurality of control signals.

16. The method according to claim **15**, wherein said transfer functions are implemented in a digital control module.

17. The method according to claim **1**, wherein said reducing waveform distortion step comprises adjusting control inputs used to generate said plurality of substantially constant envelope signals.

18. The method according to claim **17**, wherein said control inputs comprise control voltages of vector modulation circuitry that generate said substantially constant envelope signals.

19. The method according to claim **1**, wherein said reducing waveform distortion step comprises applying error pre-compensation to said received information.

20. The method according to claim **19**, further comprising:

generating expected phase and amplitude error values in said output signal; and
adjusting said received information to correct for said generated expected phase and amplitude error values.

21. The method according to claim **20**, further comprising:

generating an error table or function, said error table or function defining expected phase and amplitude errors for a representative range of said received information.

22. The method according to claim **19**, wherein said step of generating expected phase and amplitude error values is performed during a system test and characterization process.

23. An apparatus for power transmission, modulation, and amplification, comprising:

input circuitry that receives information, and that generates a plurality of control signals from said received information;
vector modulation circuitry, coupled to said input circuitry, that receives said control signals and a frequency reference signal, and that generates a plurality of substantially constant envelope signals using said frequency reference signal and said control signals;
a power amplifier that receives said substantially constant envelope signals and amplifies and combines said substantially constant envelope signals to create an output signal; and
means for reducing waveform distortion in said output signal by reducing phase and amplitude errors in said output signal relative to a desired output signal.

24. The apparatus of claim **23**, wherein said means for reducing waveform distortion comprises:

means for generating phase and amplitude error values in said output signal relative to said desired output signal.

25. The apparatus of claim **24**, wherein said means for generating phase and amplitude error values comprises one or more of:

means for measuring said phase and amplitude error values in said output signal relative to said desired output signal;
means for calculating said phase and amplitude error values in said output signal relative to said desired output signal; and
means for estimating said phase and amplitude error values in said output signal relative to said desired output signal.

26. The apparatus of claim **23**, wherein said means for reducing waveform distortion comprises means for reducing phase and amplitude errors in said output signal due to one or more of:

(a) circuitry process variations;
(b) circuitry temperature variations;
(c) IC package variations;
(d) circuit board variations;
(e) said input circuitry;
(f) said vector modulation circuitry;
(g) said power amplifier;
(h) amplifying said substantially constant envelope signals using said power amplifier; and
(i) combining outputs of said power amplifier at a single circuit node of said power amplifier.

27. The apparatus of claim **26**, further comprising:
means for estimating phase and amplitude errors due to one or more of (a)-(i), at system manufacture time.

28. The apparatus of claim **26**, further comprising:
means for at least one of measuring, calculating, and estimating phase and amplitude errors due to one or more of (a)-(i), during real-time system operation.

29. The apparatus of claim **26**, further comprising:
means for reducing gain and phase imbalances between branches of said power amplifier.

30. The apparatus of claim **23**, further comprising:
means for monitoring said phase and amplitude errors in said output signal relative to said desired output signal; and
means for triggering said means for reducing waveform distortion when said phase and amplitude errors at least satisfy system design requirements.

31. The apparatus of claim **30**, wherein said means for monitoring said phase and amplitude errors comprises differential branch phase and differential branch amplitude measurement circuitry.

32. The apparatus of claim **23**, wherein said means for reducing waveform distortion comprises means for pre-compensating for said phase and amplitude errors in said output signal relative to said desired output signal.

33. The apparatus of claim **25**, further comprising:
a feedforward error compensation module configured to pre-compensate for said estimated phase and amplitude error values in said output signal.

34. The apparatus of claim **25**, further comprising:
a feedback error compensation module configured to correct for said measured/calculated phase and amplitude error values in said output signal.

35. The apparatus of claim **25**, further comprising:
a hybrid feedforward/feedback error compensation module configured to compensate for said estimated phase and amplitude error values and to correct for said measured/calculated phase and amplitude error values.

36. The apparatus of claim **23**, wherein said means for reducing waveform distortion comprises means for adjusting transfer functions of a digital control module of said input circuitry.

37. The apparatus of claim **23**, wherein said means for reducing waveform distortion comprises means for adjusting said control signals according to said phase and amplitude errors.

38. The apparatus of claim **23**, wherein said means for reducing waveform distortion comprises means for adjusting control inputs of said vector modulation circuitry.

39. The apparatus of claim **23**, wherein said means for reducing waveform distortion comprises means for applying offset voltages to said vector modulation circuitry.

40. The apparatus of claim **23**, wherein said means for reducing waveform distortion comprises means for pre-compensating said received information for said phase and amplitude errors.

41. The apparatus of claim **40**, wherein said means for pre-compensating said received information comprises:
means for generating expected phase and amplitude error values in said output signal; and
means for adjusting said received information to correct for said expected phase and amplitude error values.

42. The apparatus of claim **41**, wherein said means for pre-compensating said received information further comprises:
an error table or function, wherein said error table or function defines expected phase and amplitude errors for a representative range of said received information.

* * * * *

US 20070279014A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2007/0279014 A1**

Wadle et al. (43) **Pub. Date: Dec. 6, 2007**

(54) **POWER FROM A NON-ANIMAL ORGANISM**

(76) Inventors: **Gordon W. Wadle**, Thomson, IL (US); **Kris J. Lagadinos**, Brookline, MA (US)

Correspondence Address:
MINTZ, LEVIN, COHN, FERRIS, GLOVSKY AND POPEO, P.C.
ONE FINANCIAL CENTER
BOSTON, MA 02111

(21) Appl. No.: **11/622,713**

(22) Filed: **Jan. 12, 2007**

Related U.S. Application Data

(63) Continuation-in-part of application No. 11/302,709, filed on Dec. 13, 2005.

(60) Provisional application No. 60/758,847, filed on Jan. 13, 2006, provisional application No. 60/761,822, filed on Jan. 24, 2006, provisional application No. 60/784,537, filed on Mar. 21, 2006, provisional application No. 60/803,059, filed on May 24, 2006, provisional application No. 60/805,676, filed on Jun. 23, 2006.

Publication Classification

(51) **Int. Cl.**
H02N 11/00 (2006.01)

(52) **U.S. Cl.** .... **322/2 R**

(57) **ABSTRACT**

A method for drawing electricity from a non-animal organism, the method including coupling a first electrical conductor to the non-animal organism, coupling a second electrical conductor to a ground rod, embedding the ground rod into soil at a predetermined depth as a function of a desired current level, whereby the current available from the non-animal organism is increased by increasing the depth that the ground rod is embedded into the soil, coupling an electrical load between the first electrical conductor and the second electrical conductor, the electrical load being configured to draw electricity from the non-animal organism via the first electrical conductor, and operating the electrical load using electricity drawn from the non-animal organism.

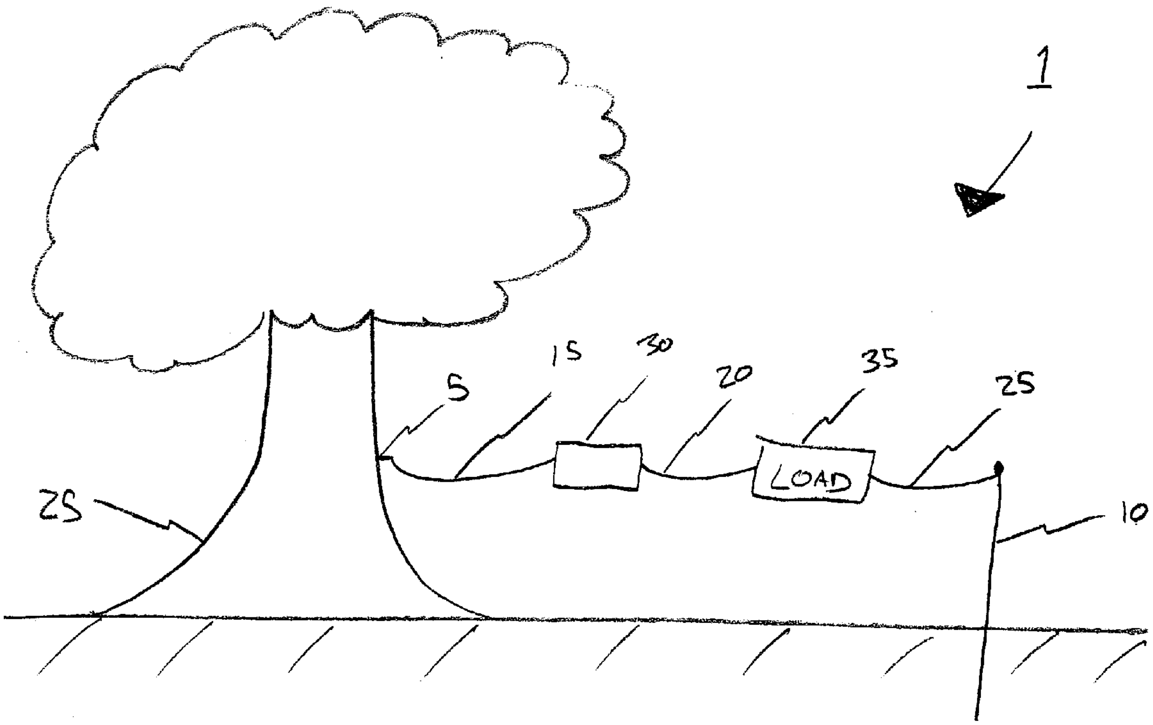

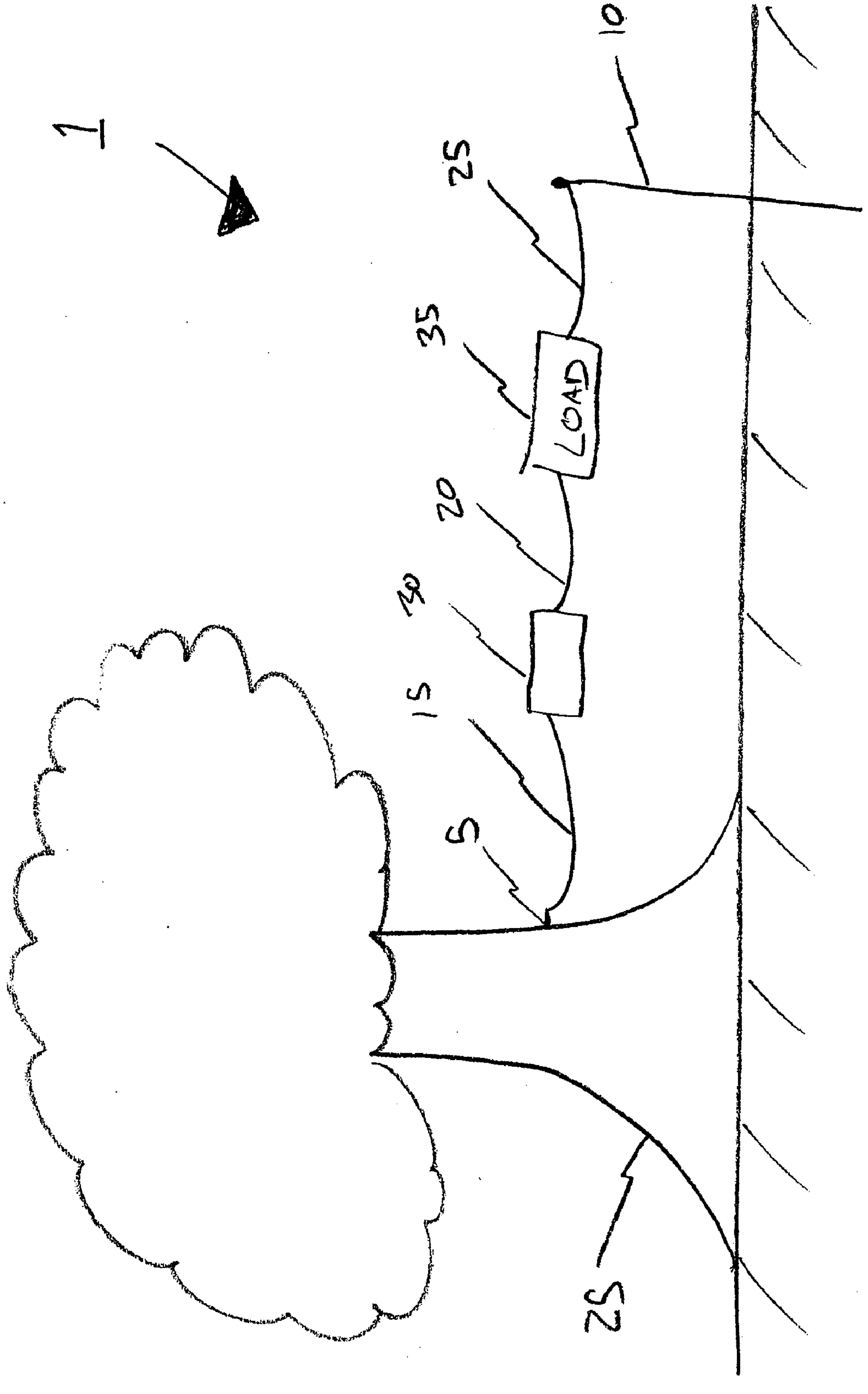
1
10
25
35
LOAD
20
30
15
5
25

FIG. 1

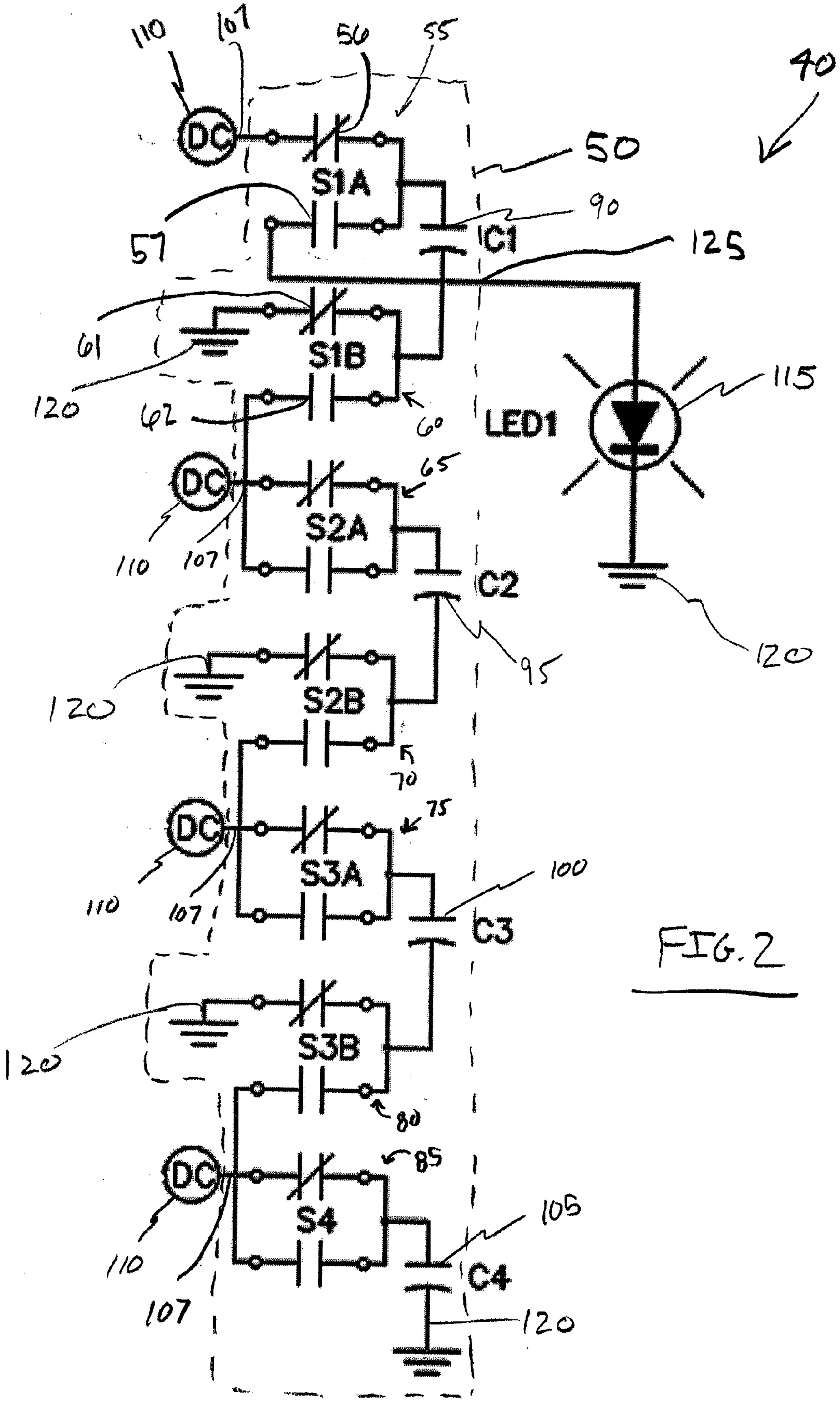
DC
S1A
C1
S1B
LED1
DC
S2A
C2
S2B
DC
S3A
C3
S3B
DC
S4
C4

FIG. 2

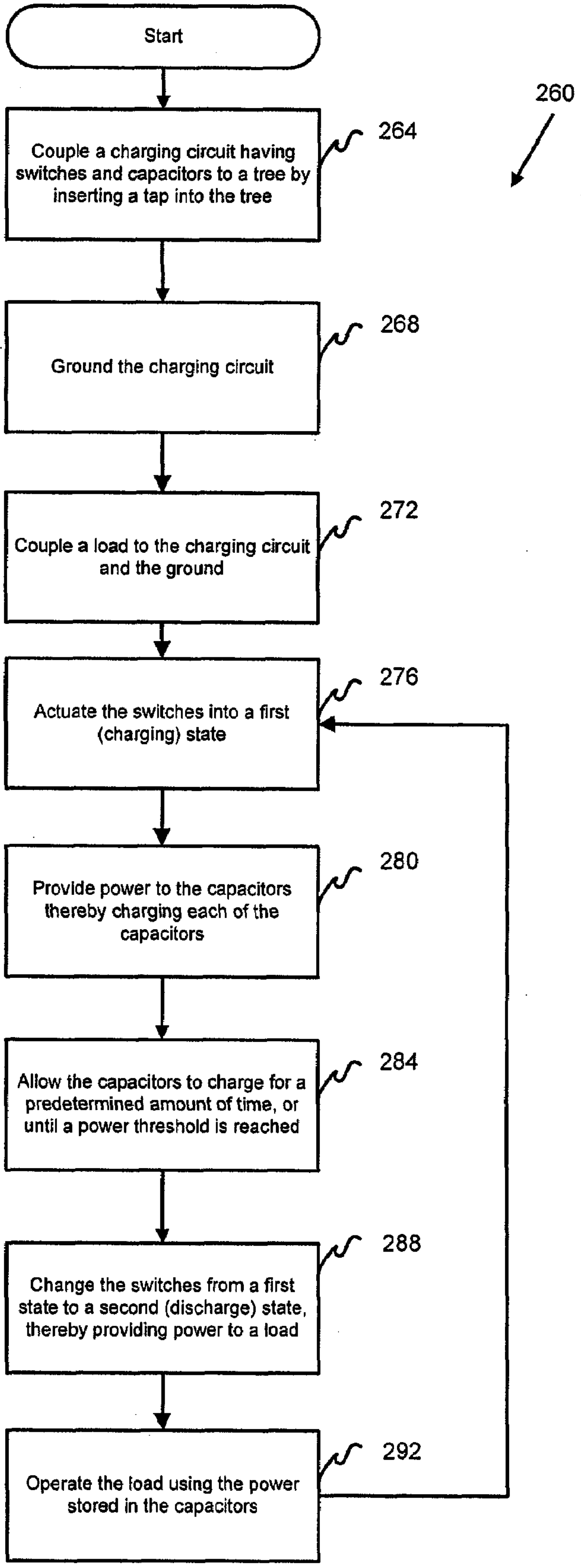
Start
Couple a charging circuit having switches and capacitors to a tree by inserting a tap into the tree
264
Ground the charging circuit
268
Couple a load to the charging circuit and the ground
272
Actuate the switches into a first (charging) state
276
Provide power to the capacitors thereby charging each of the capacitors
280
Allow the capacitors to charge for a predetermined amount of time, or until a power threshold is reached
284
Change the switches from a first state to a second (discharge) state, thereby providing power to a load
288
Operate the load using the power stored in the capacitors
292
260

FIG. 3

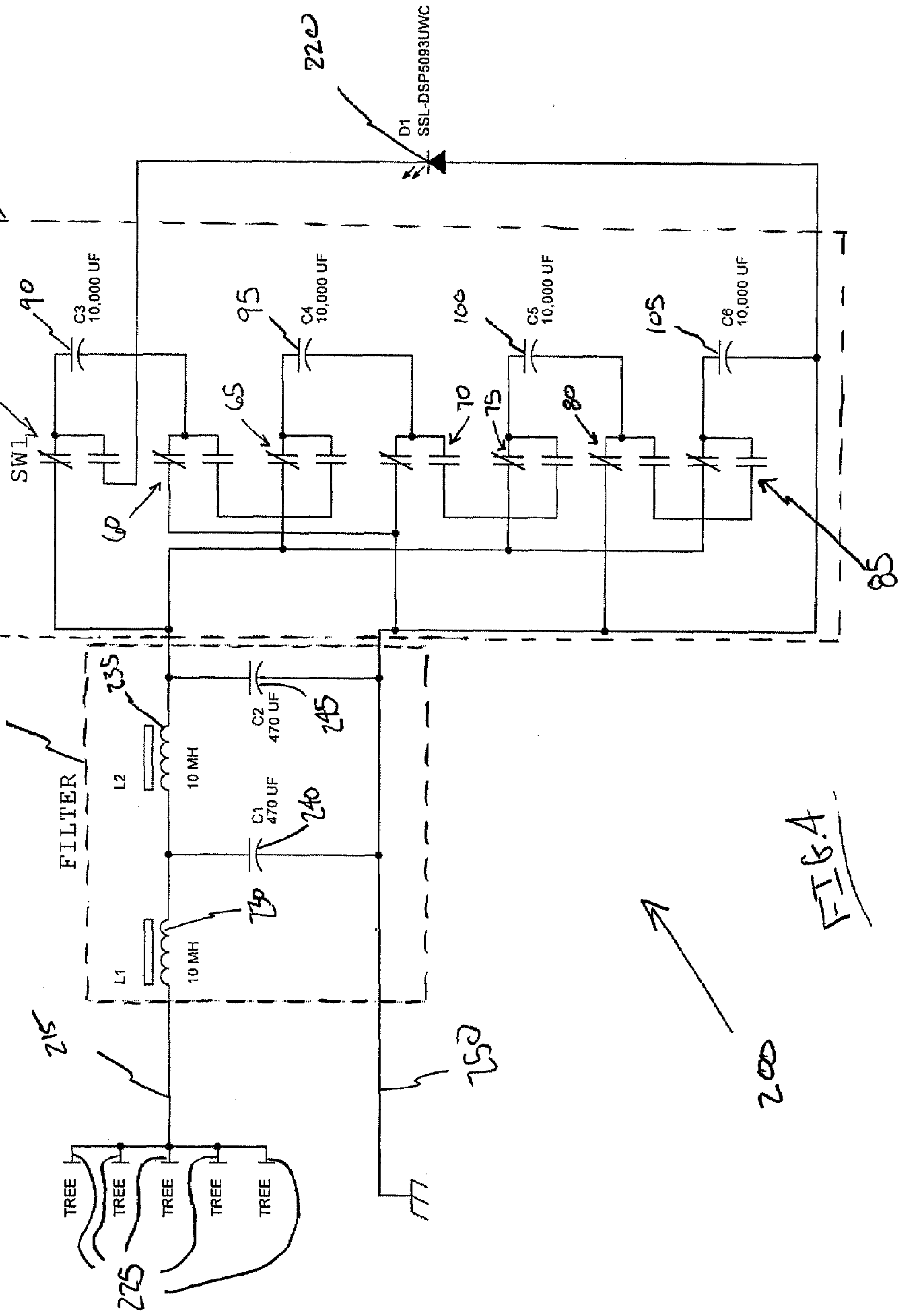
50
55
SW1
220
D1
SSL-DSP5093UWC
90
C3
10,000 UF
95
C4
10,000 UF
100
C5
10,000 UF
105
C6
10,000 UF
60
65
70
75
80
85
205
FILTER
235
L2
10 MH
C2
470 UF
245
C1
470 UF
240
230
L1
10 MH
215
250
TREE
TREE
TREE
TREE
TREE
225
200

FIG. 4

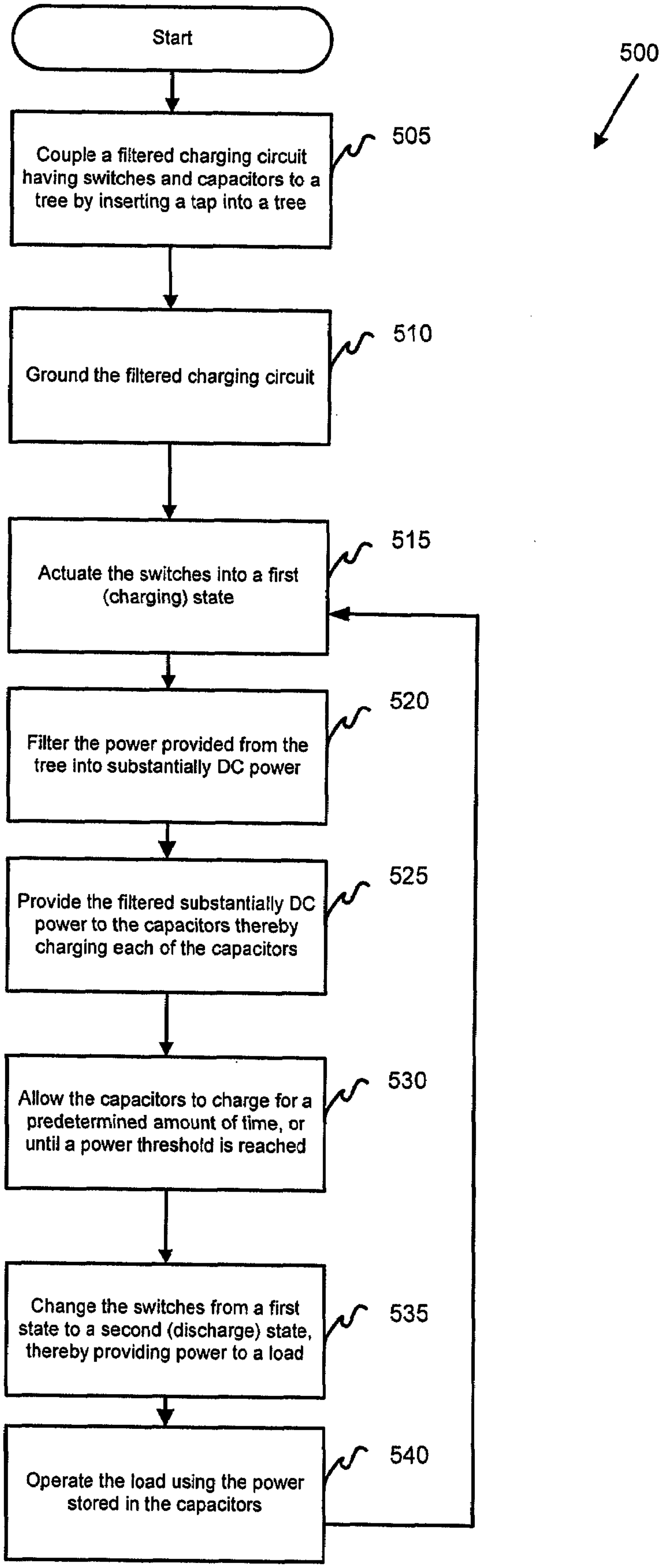
500
Start
Couple a filtered charging circuit having switches and capacitors to a tree by inserting a tap into a tree
505
Ground the filtered charging circuit
510
Actuate the switches into a first (charging) state
515
Filter the power provided from the tree into substantially DC power
520
Provide the filtered substantially DC power to the capacitors thereby charging each of the capacitors
525
Allow the capacitors to charge for a predetermined amount of time, or until a power threshold is reached
530
Change the switches from a first state to a second (discharge) state, thereby providing power to a load
535
Operate the load using the power stored in the capacitors
540

FIG. 5

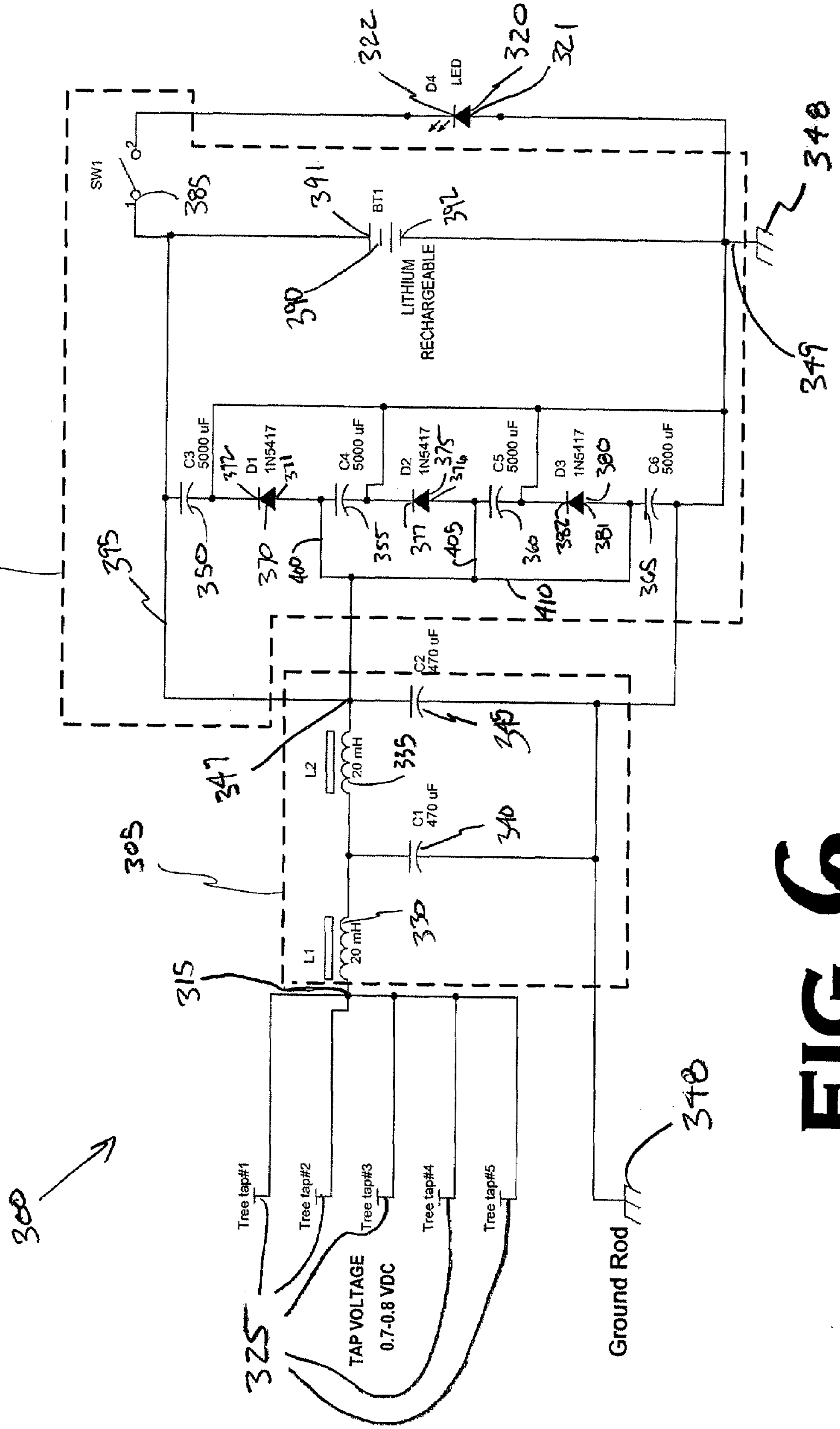
300
305
310
315
320
321
322
325
330
335
340
341
345
348
348
349
350
355
360
365
370
371
372
375
376
377
380
381
382
385
390
391
392
395
400
405
410
D4
LED
SW1
1
2
BT1
LITHIUM
RECHARGEABLE
C3
5000 uF
D1
1N5417
C4
5000 uF
D2
1N5417
C5
5000 uF
D3
1N5417
C6
5000 uF
C2
470 uF
L2
20 mH
C1
470 uF
L1
20 mH
Tree tap#1
Tree tap#2
Tree tap#3
Tree tap#4
Tree tap#5
TAP VOLTAGE
0.7-0.8 VDC
Ground Rod

FIG. 6

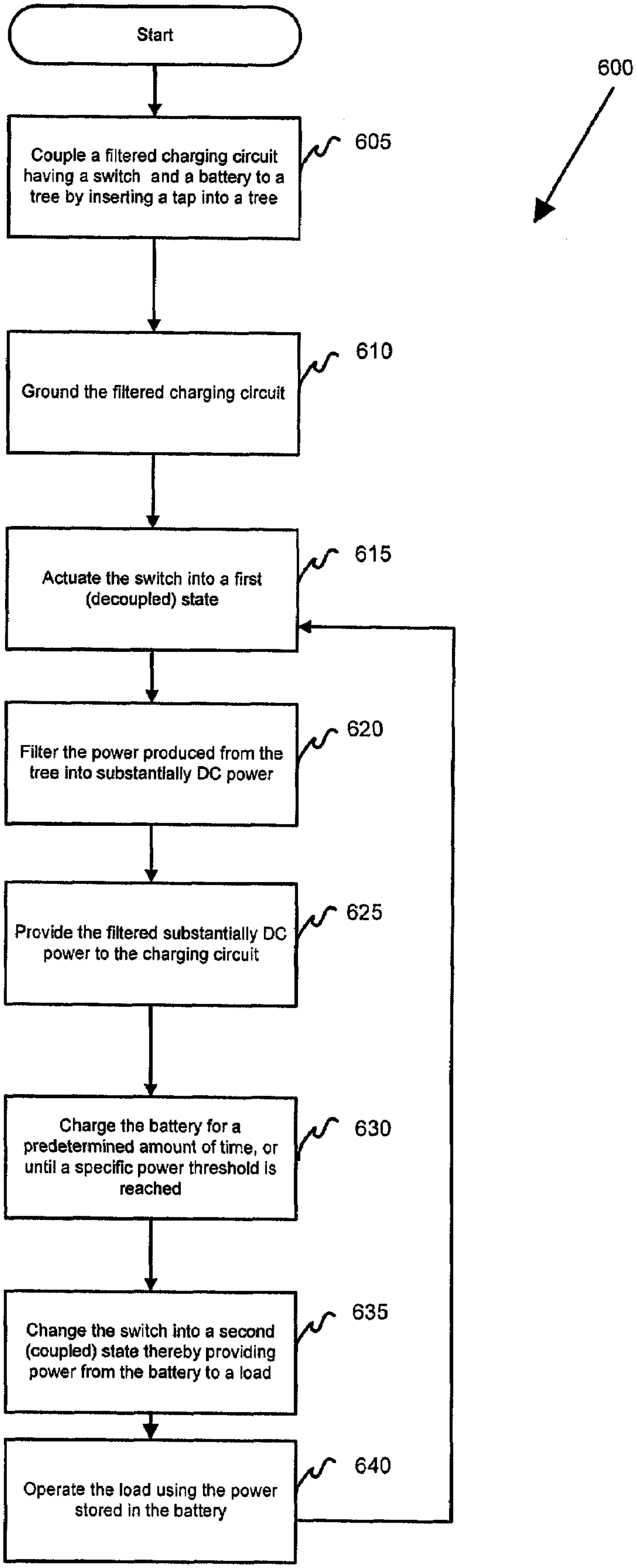
Start
600
Couple a filtered charging circuit having a switch and a battery to a tree by inserting a tap into a tree
605
Ground the filtered charging circuit
610
Actuate the switch into a first (decoupled) state
615
Filter the power produced from the tree into substantially DC power
620
Provide the filtered substantially DC power to the charging circuit
625
Charge the battery for a predetermined amount of time, or until a specific power threshold is reached
630
Change the switch into a second (coupled) state thereby providing power from the battery to a load
635
Operate the load using the power stored in the battery
640

FIG. 7

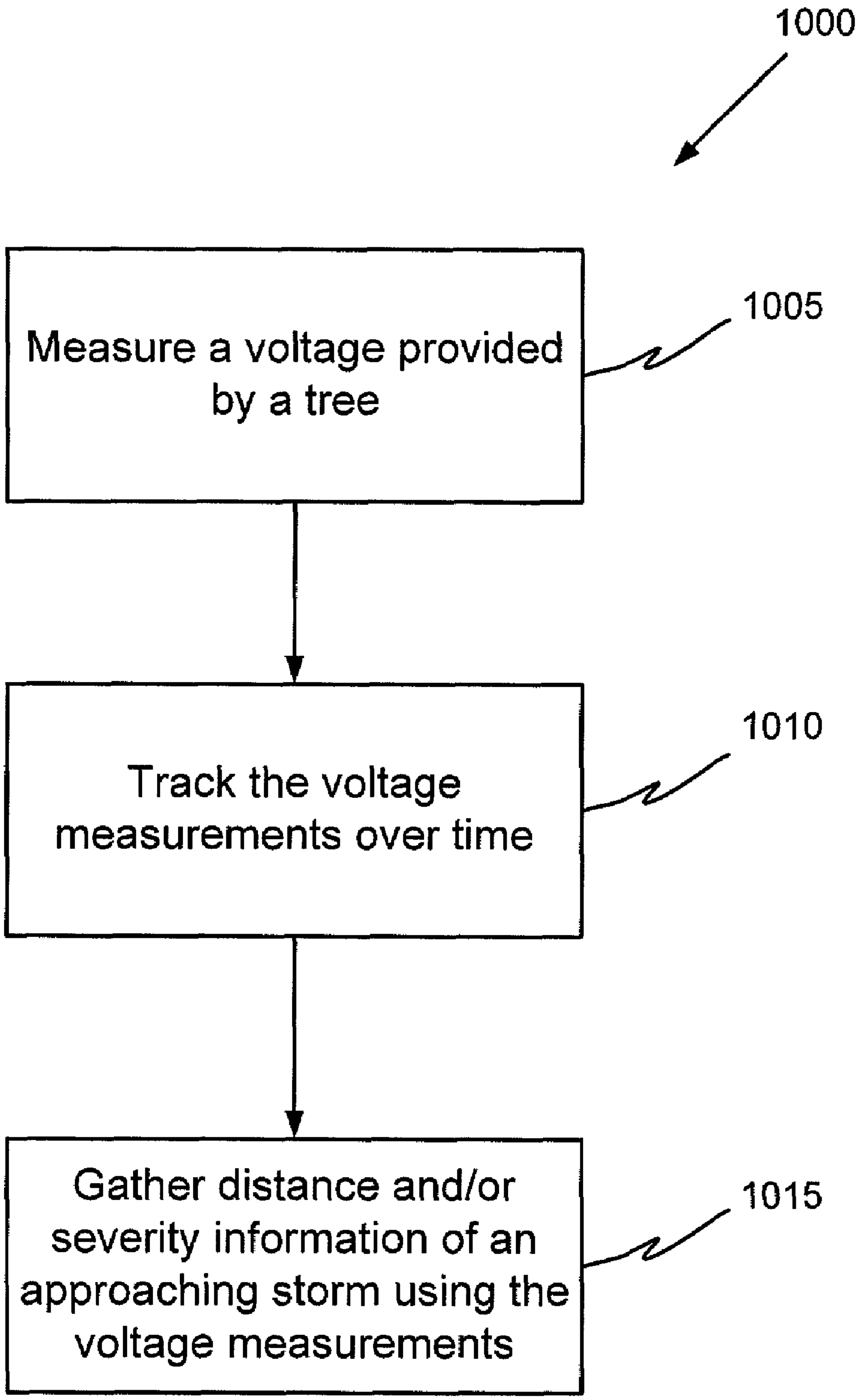
1000
Measure a voltage provided by a tree
1005
Track the voltage measurements over time
1010
Gather distance and/or severity information of an approaching storm using the voltage measurements
1015

FIG. 8

POWER FROM A NON-ANIMAL ORGANISM

RELATED APPLICATIONS

[0001] This application is a continuation-in-part of U.S. application Ser. No. 11/302,709, filed Dec. 13, 2005, which is incorporated by reference herein in its entirety. This application claims priority to U.S. Provisional Application No. 60/758,847, filed Jan. 13, 2006; U.S. Provisional Application No. 60/761,822, filed Jan. 24, 2006; U.S. Provisional Application No. 60/784,537, filed Mar. 21, 2006; U.S. Provisional Application No. 60/803,059, filed May 24, 2006; and U.S. Provisional Application No. 60/805,676, filed Jun. 23, 2006, all of which are incorporated by reference herein in their entirety.

BACKGROUND

[0002] Since the late-nineteenth century the use of, and uses for, electricity has increased tremendously, becoming a fundamental part of everyday life for most people. One only has to look at remote parts of the world to see how drastically different life is without electricity. Most electric devices in use today typically draw between a few milliwatts to several megawatts of power, depending on the application. Higher costs for the fuels needed to generate electricity, and a higher electrical demand in general, however, have led to increased electricity costs, thereby increasing the attractiveness of alternative power sources.

[0003] One typical use of electricity is a light emitting diode (LED). LEDs have seen increasing popularity in recent times due to a lower per unit cost and a greater number of available colors. LEDs are more energy efficient (i.e., less power is consumed) and generally have a much longer life expectancy than conventional filament-based light bulbs. In general, LEDs draw approximately 20 mA at 2V (i.e., 40 mW) when illuminated, which is far less than conventional light bulbs.

[0004] Distribution of electricity from a generation plant to the end-user is not a trivial problem. Thousands of miles of wires and cables creating a transmission network are involved in delivering power to consumers. The transmission network adds costs such as material costs and the cost of lost energy due to the resistance of the transmission wires. For the average consumer of electricity, the transmission costs generally equal the cost of the electricity itself. Furthermore, portions of the world have no electricity because it is simply too far from the nearest transmission line or the terrain itself prohibits installation of transmission lines.

SUMMARY

[0005] A method for drawing electricity from a non-animal organism, the method including coupling a first electrical conductor to the non-animal organism, coupling a second electrical conductor to a ground rod, embedding the ground rod into soil at a predetermined depth as a function of a desired current level, whereby the current available from the non-animal organism is increased by increasing the depth that the ground rod is embedded into the soil, coupling an electrical load between the first electrical conductor and the second electrical conductor, the electrical load being configured to draw electricity from the non-animal organism via the first electrical conductor, and operating the electrical load using electricity drawn from the non-animal organism.

[0006] In general, in another aspect, the invention provides a system including a non-animal organism, a first electrical conductor electrically coupled to the non-animal organism, a plurality of ground rods embedded into soil wherein a quantity of the plurality of ground rods is a function of a desired current level from the non-animal organism, whereby the current available from the non-animal organism is increased by increasing the quantity of the plurality of ground rods, a second electrical conductor coupled to the plurality of ground rods, and an electrical load coupled between the first electrical conductor and the second electrical conductor to draw electricity from the non-animal organism, the electrical load using electricity drawn from the non-animal organism.

[0007] In general, in another aspect, the invention provides a method of predicting weather using electricity from a non-animal organism, the method including coupling a first electrical conductor to the non-animal organism, coupling a second electrical conductor to a ground, coupling a voltmeter between the first electrical conductor and the second electrical conductor, measuring a voltage potential between the first and second electrical connectors, providing a weather prediction as a function of the measured voltage potential.

[0008] Implementations of the invention may further include the following features. The method of predicting weather including determining a baseline voltage reading for the non-animal organism under a first weather condition, determining a plurality of voltage readings over time, comparing each of the plurality of voltage readings to the baseline voltage reading to determine differences between the baseline voltage reading and each of the plurality of voltage readings, and calculating information indicative of a second, future weather condition as a function of the differences.

[0009] In general, in another aspect, the invention provides a system for use with live vegetative matter growing in soil, the system including a non-animal organism, a first electrical conductor electrically coupled to the non-animal organism, and a second electrical conductor coupled to the first electrical conductor and coupled to the live vegetative matter, the second electrical conductor providing electricity from the non-animal organism to the live vegetative matter, wherein the growth of the live vegetative matter is stimulated by the electricity provided by the non-animal organism.

[0010] In general, in another aspect, the invention provides a system including a non-animal organism, a first electrical conductor electrically coupled to the non-animal organism, a second electrical conductor coupled to a ground, and an electrical load coupled between the first electrical conductor and the second electrical conductor to draw electricity from the non-animal organism, the electrical load using electricity drawn from the non-animal organism, wherein the load is one of a battery, a battery charging circuit, a sensor, a radio frequency identification chip, a transmitter, a receiver, a global positioning service (GPS) device, a light emitting device, and a fire ignition system.

[0011] Implementations of the invention may include one or more of the following features. The load is the sensor and the sensor is one of an oxygen sensor, an air-speed sensor, a humidity sensor, a barometric pressure sensor, a camera, a photoelectric sensor, an altitude sensor, a smoke detector, a microphone, and a vibration sensor. The load is the GPS device and the GPS device is one of a GPS receiver, a GPS transmitter, a GPS guidance system, and a GPS navigation system. The load is the light emitting device and the light emitting device is one of a light emitting diode configured to emit visible light, and an infrared light emitting diode configured to emit an infrared signal.

[0012] Various aspects of the invention may provide one or more of the following capabilities. A non-animal organism, such as a member of the plant and/or fungi kingdom, may supply electricity to a load. Electricity may be available in remote areas without an electricity transmission network. Alternative "clean" electricity can be produced. An LED may be powered from a non-animal organism. Infra-red LEDs used in military operations may be powered. A traffic light may be powered from a non-animal organism. A security light may be powered from a non-animal organism. Dependence on fossil fuels to generate electricity may be reduced. Lighting may be provided at campgrounds and/or ski areas using power provided from non-animal organisms. Power derived from non-animal organisms may be used to recharge batteries in hybrid vehicles. Advance storm warning can be obtained by measuring the voltage provided by the non-animal organism.

[0013] These and other capabilities of the invention, along with the invention itself, will be more fully understood after a review of the following figures, detailed description, and claims.

BRIEF DESCRIPTION OF THE FIGURES

[0014] FIG. **1** is a diagram of an apparatus for drawing power from a tree.

[0015] FIG. **2** is a diagram of a charging circuit used to provide power derived from a tree to a load.

[0016] FIG. **3** is a flowchart of a process for deriving power from a tree using the charging circuit shown in FIG. **2**.

[0017] FIG. **4** is a circuit diagram of a filtered charging circuit used in providing electricity from a tree to a load, including the charging circuit of FIG. **2** and a filter.

[0018] FIG. **5** is a flowchart of a process of deriving power from a tree using the electrical circuit shown in FIG. **4**.

[0019] FIG. **6** is a circuit diagram of a filtered charging circuit used in providing electricity from a tree to a load and including a battery.

[0020] FIG. **7** is a flowchart of a process of deriving power from a tree using the electrical circuit shown in FIG. **6**.

[0021] FIG. **8** is a flowchart of a process of determining storm distance and/or severity using voltage measurements taken from a tree.

DETAILED DESCRIPTION

[0022] Embodiments of the invention provide techniques for drawing electricity from non-animal organisms such as members of the plant and/or fungi kingdom, and providing the electricity to a load. Non-animal, non-mammal organisms such as spermatophytes, pteridophytes, succulents, Marattiales ferns, Ophioglossales ferns, Leptosporangiate ferns, Mycophycota fungi, Zygomycota fungi, Basidiomycota fungi, and Ascomycota fungi may be used. Specifically, electricity can be drawn from vegetative matter such as a living tree. The amount of available electricity has been found to depend on the location and type of non-animal organism, and to be approximately 0.5-2 volts DC, plus some AC current. For example, an apparatus for using this energy includes a conductor inserted into a tree and connected to a positive terminal of a load. A negative conductor of the load is connected to a grounded conductor, thereby completing a circuit. Other circuitry, such as charging circuits and/or voltage step-up circuits, may also be used. Other embodiments are within the scope of the invention.

[0023] Referring to FIG. **1**, an apparatus **1** for deriving electricity from a tree **25** includes a tap **5**, a conductor **10**, wires **15**, **20**, and **25**, a circuit **30**, and a load **35**. The tap **5** is configured to attach to, and to conduct current flow from, the tree **25**. For example, the tap **5** may be configured to be inserted into the tree **25**, although other configurations are possible (e.g., a non-invasive transformer core that surrounds the circumference of the tree **25**). The wire **15** is electrically coupled to the tap **5** and the circuit **30**. The wire **20** is electrically coupled to the circuit **30** and the load **35**. The wire **25** is electrically coupled to the load **35** and the conductor **10**. The conductor **10** is electrically conductive and is configured to be inserted approximately two feet into the ground while protruding above the ground, although the conductor **10** may be configured to be inserted to other depths. By increasing the depth that the conductor **10** is inserted into the ground and/or using multiple conductors **10**, the load **35** can draw more current from the tree **25**. The conductor **10** is preferably a tinned copper rod. Other materials and/or configurations of the conductor **10** are possible. For example the conductor **10** may be aluminum and/or connected to a "ground" connection of a typical household electrical system. The circuit **30** is electrically conductive and is configured to filter the power provided by the tree, to step-up (or step-down) the voltage supplied by the tree **25**, and/or to store the power provided by the tree **25**. The circuit **30** may perform functions other than those listed above. Also, embodiments of the apparatus **1** without the circuit **30** are possible (e.g., connecting a load directly between the tree **25** and the conductor **10**).

[0024] The load **35** can be any of a number of different devices used for a variety of purposes. For example, the load **35** can include a lithium battery that is charged by the tree **25**, a sensor (e.g., capable of sensing temperature, air speed, humidity, barometric pressure, video, audio, light, vibration, altitude, oxygen levels), a remote sensor (e.g., over a LAN, WAN, the Internet, WiFi), a radio frequency identification (RFID) chip, a transmitter, and/or a receiver. The load **35** can be a device for use with a global positioning system (GPS) such as a GPS receiver, a GPS transmitter, a GPS guidance system, and/or a GPS navigation system. The load **35** can include a fire and/or smoke detection system, a system configured to charge a battery powered device (e.g., a mobile phone, a laptop computer, a portable GPS system, a flashlight, a radio), a lighting system (e.g., for recreational use, for military use)(including, e.g., one or more light emitting diodes (LEDs) such as infrared LEDs), a fire ignition system (e.g., for campground use), a weather detection and/or monitoring system, an emergency alert/assistance beacon, a solar lighting backup system, and/or a wireless transmission system for use with a computer. The load **35** can include a plant (e.g., as described below in Experiment 4).

[0025] Various embodiments of the tap **5** are possible. Preferably, the tap **5** is a stainless steel rod, e.g., a nail, having an outside diameter of about 0.125 inches, but other materials and sizes are possible. For example, brass plated or aluminum rods having an outside diameter of about 0.06 inches may be used. The tap **5** is electrically conductive material and is preferably of a material (e.g., stainless steel) that has a relatively high corrosion resistance, thus inhibiting increased resistance caused by corrosion. For extended use, the tap **5** is preferably not copper (at least on its exterior) as this can negatively affect (e.g., kill) many types of trees. The tap **5** is preferably configured to be inserted between about 0.375 inches and about 0.75 inches into the tree **25**, although other depths are possible. In trees with thick bark, the tap **5** may be inserted further into the tree **25**. For example, if a tree has bark 1 inch thick, the tap **5** may be inserted about 1.5 inches into the tree **25**. The tap **5** is preferably inserted into the tree **25** between about one and about six feet above ground level, although other heights may be used. While the apparatus **1** includes the one tap **5**, multiple taps may be used. Using multiple taps in a single tree has been found to increase the amount of current available from the tree. The taps may all be the same, or one or more taps may be different (e.g., a different material, configured for different insertion depth, etc.) than another tap.

[0026] Referring also to FIG. **2**, an exemplary embodiment **40** of the apparatus **1** including an LED load **115**, and an exemplary circuit **30** that is a charging circuit **50**, which includes switches **55**, **60**, **65**, **70**, **75**, **80**, and **85**, and capacitors **90**, **95**, **100**, and **105**. The switches **55**, **60**, **65**, **70**, **75**, **80**, and **85** are single-pole double-throw (SPDT) switches. The switch **55** includes selective connections **56** and **57**. The switch **57** is connected on one side to the switch **56** and the capacitor **90** and on its other side to an output **125** configured to be connected to the load **115**. The switch **60** also includes selective connections **61** and **62**. When the switches **55** and **60** are in a first state, the connections **56** and **61** are closed and the connections **57** and **62** are open, thereby coupling the capacitor **90** between a power source **110** (here, a tree) and a ground **120**. When the switches **55** and **60** are in a second state, the connections **56** and **61** are open, and the connections **57** and **62** are closed, thereby coupling the capacitor **90** between the load LED **115**, and the switch **65**. Each of the switches **55**, **65**, **75**, and **85** are coupled to the tree **110** via the tap **107**. The switches **65**, **70**, **75**, **80**, and **85** operate as described with respect to the switches **55** and **60**.

[0027] The capacitors **90**, **95**, **100**, and **105** are coupled to the switches **55**, **60**, **65**, **70**, **75**, **80**, and **85** such that when the switches **55**, **60**, **65**, **70**, **75**, **80**, and **85** are in a first state, the circuit **50** is in a charging state and each of the capacitors **90**, **95**, **100**, and **105** are coupled between the power source **110** and the ground **120**. When the switches **55**, **60**, **65**, **70**, **75**, **80**, and **85** are in the first state the capacitors **90**, **95**, **100**, and **105** accumulate an electrical charge. The capacitors **90**, **95**, **100**, and **105** are further coupled to the switches **55**, **60**, **65**, **70**, **75**, **80**, and **85** such that when the switches **55**, **60**, **65**, **70**, **75**, **80**, and **85** are in a second state, the circuit **50** is in a discharging state and the capacitors **90**, **95**, **100**, and **105** are coupled in series between the ground **120** and a load **115** thus providing power to the load **115**. The voltage provided to the load **115** is substantially equal to the sum of the voltages across each of the capacitors **90**, **95**, **100** and **105**. The capacitors **90**, **95**, **100**, and **105** are preferably about 10,000 μF, but other capacitances are possible. While an LED is shown as the load **115**, other loads may be used.

[0028] While the charging circuit **50** is shown coupled to a single tree (i.e., the tree **110**), other configurations are possible. For example, each of the switches **55**, **65**, **75**, and **85** may be connected to separate trees. The switches **55**, **65**, **75**, and **85** could each be connected to multiple trees (or other non-animal organisms). One or more of the switches **55**, **65**, **75**, and **85** could each be connected to a single tree with multiple taps **107**. One of the switches **55**, **65**, **75**, and **85** could be connected to a single tree with a single tap, with the remainder of the switches **55**, **65**, **75**, and **85** being connected to multiple trees, each with multiple taps. One of the switches **55**, **65**, **75**, and **85** could be connected to a single tree with multiple taps, with the remainder of the switches **55**, **65**, **75**, and **85** being coupled to a single tree with multiple taps. Each of the switches **55**, **65**, **75**, and **85** may be coupled to a single tree or multiple trees using more than one of the tap **107**.

[0029] In operation, referring to FIG. **3**, with further reference to FIG. **2**, a process **260** for providing power derived from a tree to a load using the apparatus **40** includes the stages shown. The process **260**, however, is exemplary only and not limiting. The process **260** may be altered, e.g., by having stages added, removed, or rearranged.

[0030] At stage **264**, the charging circuit **50** is coupled to the living non-animal organism power source **110**, such as a tree, a plant, etc. Preferably, the tap **107** is inserted into the power source **110**. The tap **107** is inserted approximately 0.375 inches to 0.75 inches into the tree. Alternatively, a non-invasive tap may be used, e.g., a transformer core can be placed around a circumference of the tree.

[0031] At stage **268**, the charging circuit **50** is grounded. Preferably, the charging circuit **50** is coupled to a ground rod, or other suitable electrical ground, such as a ground connection in a typical residential power system. More current may be drawn from the living non-animal organism by the load **115** by increasing the depth that the ground rod is inserted into the ground and/or using multiple ground rods.

[0032] At stage **272**, the load **115** is coupled between the charging circuit **50** and the ground **120**. The load **115** is coupled on one side to the output **125** of the charging circuit **50** and on its other side to the ground **120**.

[0033] At stage **276**, the switches **55**, **60**, **65**, **70**, **75**, **80**, and **85** are actuated into the first (charging) state. The connections **56** and **61** of the switches **55** and **60** are closed, the connections **57** and **62** of the switches **55** and **60** are opened, and likewise for the switches **65**, **70**, **75**, **80**, and **85**. This couples the capacitors **90**, **95**, **100**, and **105** to the taps **107**.

[0034] At stage **280**, the power is provided from the tree **110** to the capacitors **90**, **95**, **100**, and **105**. The capacitors **90**, **95**, **100**, and **105** store energy received from the taps **107**.

[0035] At stage **284**, the capacitors **90**, **95**, **100**, and **105** are allowed to charge. The amount of time the capacitors **90**, **95**, **100**, and **105** are charged may vary to suit a specific application. For example, to provide sufficient power to illuminate the LED, each of the capacitors **90**, **95**, **100**, and **105** is charged to 0.5 Vdc. The amount of time for the capacitors **90**, **95**, **100**, and **105** to reach 0.5 Vdc may vary depending on the amount of power supplied by a particular power source.

[0036] At stage **288**, the switches **55**, **60**, **65**, **70**, **75**, **80**, and **85** are changed from the first state to the second state to discharge the power accumulated in the capacitors **90**, **95**, **100**, and **105**, thereby providing power to the load **115**.

[0037] The power from the capacitors **90**, **95**, **100**, and **105** is used to operate the load **115**, here causing the LED to emit light. The process **260** returns to stage **276** where the switches **55**, **60**, **65**, **70**, **75**, **80**, and **85** are changed from the second state to the first state, thereby providing power from the taps **107** to the capacitors **90**, **95**, **100**, and **105**.

[0038] Referring to FIGS. **2** and **4**, a filtered charging circuit **200** includes a filter circuit **205** and the charging circuit **50**, which are coupled to a power input **215**, a load **220** (in FIGS. **2** and **4** an LED), and a ground connector **250**. The filter circuit **205** is coupled between the power input **215** and the charging circuit **50**, and is configured to provide substantially DC power to the charging circuit **50**. The power input **215** is coupled to multiple taps **225** configured to be inserted into one or more trees. As described above with reference to FIG. **2**, the charging circuit can provide the load **220** with a stepped-up, substantially DC voltage.

[0039] The filter circuit **205** includes inductors **230** and **235**, and capacitors **240** and **245**. The inductors **230** and **235** are coupled in series between the power input **215** and the charging circuit **50** to inhibit high-frequency power produced by the tree from reaching the charging circuit **50**. The capacitor **240** is coupled between the junction of the inductors **230** and **235** and the ground **250**. The capacitor **245** is coupled between the junction of the inductor **235** and the charging circuit **50** and the ground **250**. For example, the inductors **230** and **235**, and the capacitors **240** and **245** are arranged in a 2-stage pie filter configuration. The capacitors short-out (e.g., ground) high-frequency power produced by the tree, further inhibiting non-DC power from being conducted to the charging circuit **50**. The inductors **230** and **235** are preferably about 10 mH, although other inductances are possible. The capacitors **240** and **245** are preferably about 470 μF, although other capacitances are possible. The charging circuit **50** is configured to receive substantially DC power from the filter circuit **205**, and to output intermittent DC power to the load **220** similar to the description provided above with respect to FIG. **2**.

[0040] In operation, referring to FIG. **5**, with further reference to FIG. **4**, a process **500** for providing power derived from a tree to the load **220** using the filtered charging circuit **200** includes the stages shown. The process **500**, however, is exemplary only and not limiting. The process **500** may be altered, e.g., by having stages added, removed, or rearranged.

[0041] At stage **505**, the filtered charging circuit **200** is coupled to the power input **215** such as a tree, a fungus, or other suitable non-animal organism, here by inserting the taps **225** into a single tree. Each of the taps **225** is inserted approximately 0.375 inches to approximately 0.75 inches into the tree. If any of the taps **225** are non-invasive, then that (those) taps(s) **225** (e.g., a transformer core) is (are) mounted accordingly. (e.g., placed around the circumference of a tree).

[0042] At stage **510**, the filtered charging circuit **200** is coupled to ground. The filtered charging circuit **200** is connected to the ground connector **250**, such as a rod, or other suitable electrical ground connector (e.g., a ground connection in a typical residential power system). More current may be drawn from the living non-animal organism by the load **220** by increasing the depth that the ground rod is inserted into the ground and/or using multiple ground rods.

[0043] At stage **515**, the switches **55**, **60**, **65**, **70**, **75**, **80**, and **85** are actuated into a first (charging) state coupling the capacitors **90**, **95**, **100**, and **105** to the filter circuit **205**. Power flows from the filter circuit **205** to the capacitors **90**, **95**, **100**, and **105**.

[0044] At stage **520**, the power derived from the tree is filtered to substantially remove alternating current (AC). At stage **520** the filter circuit **205** filters the power derived from the taps **225** into substantially DC power. The combination of the inductors **230** and **235** and the capacitors **240** and the **245** substantially filters out non-DC frequencies produced by the tree. The inductors **230** and **235** choke the high-frequencies produced by the tree. The capacitors **240** and **245** inhibit low frequency power and conduct high-frequency power to the ground connector **250**. The filter circuit **205** provides the filtered substantially DC power to the charging circuit **50**.

[0045] At stage **525** the filtered substantially DC power from the filter circuit **205** is provided to the capacitors **90**, **95**, **100**, and **105**. The switches **55**, **60**, **65**, **70**, **75**, **80**, and **85** are put in the first state to couple the circuit **205** to the capacitors **90**, **95**, **100**, and **105** to provide power to, and charge, the capacitors **90**, **95**, **100**, and **105**. At stage **530**, the capacitors **90**, **95**, **100**, and **105** are allowed to charge. The amount of time the capacitors **90**, **95**, **100**, and **105** are charged varies, and may be tailored to suit a specific application. For example, to provide sufficient power to illuminate the load **220**, each of the capacitors is charged to 0.5 Vdc. The amount of time required to reach 0.5 Vdc may vary depending on the amount of power supplied by a particular power source.

[0046] At stage **535**, the switches **55**, **60**, **65**, **70**, **75**, **80**, and **85** are changed from the first state to the second state to discharge the power accumulated in the capacitors **90**, **95**, **100**, and **105**, thereby providing power to the load **220**.

[0047] After stage **535**, the switches **55**, **60**, **65**, **70**, **75**, **80**, and **85** are actuated from the second state to the first state, thereby providing filtered substantially DC power from the filter circuit **205** to the charging circuit **50**. The stages **515**, **520**, **525**, and **530** may be repeated.

[0048] At stage **540**, the power from the capacitors **90**, **95**, **100**, and **105** is used to operate the load **220**, here causing the LED to emit light. The process **500** returns to stage **515** where the switches **55**, **60**, **65**, **70**, **75**, and **85** are changed from the second state to the first state, thereby providing power from the taps **225** to the capacitors **90**, **95**, **100**, and **105**

[0049] Referring to FIG. **6**, a filtered charging circuit **300** includes a filter circuit **305** and a charging circuit **310**, which are coupled to a power input **315** and a load **320** (in FIG. **6**, an LED). The filter **305** is coupled between the power input **315** and the charging circuit **310**, and is configured to provide substantially DC power to the charging circuit **310**. The power input **315** is coupled to multiple taps **325** that are each configured to be inserted into a tree. The load **320** is preferably a SSL-DSP5093UWC LED (manufactured by Lumex Incorporated, of Palatine, Ill.), although other LEDs, and other types of loads, may be used.

[0050] The filter circuit **305** includes inductors **330** and **335**, capacitors **340** and **345**, and an output node **347**. The inductors **330** and **335** are coupled in series between the power input **315** and the output node **347** and are of inductances to serve as chokes of any high-frequencies received at the power input **315**. The capacitor **340** is coupled between the junction of the inductors **330** and **335** and the ground **348**. The capacitor **345** is coupled between the output node **347** and the ground **348**. For example, the inductors **330** and **335**, and the capacitors **340** and **345** are arranged in a 2-stage pie filter configuration. The inductors **330** and **335** are preferably about 10 mH, although other inductances are possible. The capacitors **340** and **345** work in conjunction with the inductors **330** and **335** shorting-out high frequency signals that may have passed through the inductors **330** and **335**, respectively. The capacitors **340** and **345** are preferably about 470° F., although other capacitances are possible.

[0051] The charging circuit **310** includes capacitors **350**, **355**, **360**, and **365**, diodes **370**, **375**, and **380**, a switch **385**, a battery **390**, and a ground connection **349** connected to the ground **348**. Coupled between the output node **347** and the ground connection **349** are the capacitors **350**, **355**, **360**, and **365**, and the diodes **370**, **375**, and **380**, in an alternating series of capacitors and diodes. Anodes **371**, **376**, and **381** of the diodes **370**, **375**, and **380**, respectively, are coupled to the output node **347**. Cathodes **372**, **377**, and **382** of the diodes **370**, **375**, and **380**, respectively, are coupled to the ground connection **349**. The capacitor **350** is coupled between the cathode **372** of the diode **370** and the output node **347**. The capacitor **365** is coupled between the anode **381** of the diode **380** and the ground connection **349**. The capacitors **350**, **355**, **360**, and **365**, and the diodes **370**, **375**, and **380** act as a voltage multiplier circuit to allow filtered substantially DC power to charge the capacitors **350**, **355**, **360**, and **365** (e.g., by summing the voltages across the capacitors **350**, **355**, **360**, and **365**). Using the capacitors **350**, **355**, **360**, and **365**, and the diodes **370**, **375**, and **380**, a higher voltage (e.g., 2-2.5 V) is produced to charge the battery **390**. The capacitors **350**, **355**, **360**, and **365** are 5,000 μF, although other capacitances are possible, such as 10,000 μF. The diodes **370**, **375**, and **380** are preferably 1N5417 diodes, but other diodes are possible.

[0052] The battery **390** is coupled between the output node **347** and the ground **348** such that it may receive power from the output node **347**. The battery **390** is preferably a lithium-ion battery, but other batteries may be used. A positive terminal **391** of the battery **390** is coupled to the output node **347** and the switch **385**. A negative terminal **392** of the battery **390** is coupled to the ground **348**. Other configurations are possible (e.g., coupling the negative terminal **392** to the output node **347**, and coupling the positive terminal **391** to the ground **348**).

[0053] The switch **385** is coupled between a terminal **322** of the load **320** and output node **347** to control a power flow to the load **320**. When the switch **385** is in an open state (as shown), power is inhibited (and preferably prevented) from flowing to the load **320**. When the switch **385** is in a closed state, power may flow to the load **320**. A terminal **321** of the load **320** is coupled to the ground **348**.

[0054] In operation, referring to FIG. **7**, with further reference to FIG. **6**, a process **600** for providing power derived from a tree to the load **320** using the filtered charging circuit **300** includes the stages shown. The process **600**, however, is exemplary only and not limiting. The process **600** may be altered, e.g., by having stages added, removed, or rearranged.

[0055] At stage **605**, the filtered charging circuit **300** is coupled to the power input **315** such as a tree, a fungus, or other suitable non-animal organism, here by inserting the taps **325** into a single tree. Each of the taps **325** is inserted approximately 0.375 inches to approximately 0.75 inches into the tree. If any of the taps **325** are non-invasive, then that (those) taps(s) **325** (e.g., a transformer core) is (are) mounted accordingly. (e.g., placed around the circumference of a tree).

[0056] At stage **610**, the filtered charging circuit **300** is coupled to ground. The filtered charging circuit **300** is connected to the ground connector **349**, such as a rod, or other suitable electrical ground connector (e.g., a ground connection in a typical residential power system). More current may be drawn from the living non-animal organism by the load **320** by increasing the depth that the ground rod is inserted into the ground.

[0057] At stage **615**, the switch **385** is actuated into the first state (i.e., open) where the load **320** is disconnected from the filtered charging circuit **300** and current is inhibited/prevented from reaching/operating the LED **320**.

[0058] At stage **620**, the power derived from the tree is filtered to substantially remove alternating current (AC). At stage **620** the filter circuit **305** filters the power derived from the taps **325** into substantially DC power. The combination of the inductors **330** and **335** and the capacitors **340** and the **345** substantially filters out non-zero frequencies produced by the tree. The inductors **330** and **335** choke the high-frequencies produced by the tree. The capacitors **340** and **345** inhibit low frequency power and conduct high-frequency power to the ground connector **349**. The filter circuit **305** provides the filtered substantially DC power to the charging circuit **310**.

[0059] At stage **625**, the filtered substantially DC power is provided to the charging circuit **310** via the output node **347**. Power provided from the output node **347** is conducted through the capacitors **350**, **355**, **360**, and **365**, and the diodes **370**, **375**, and **380**. The configuration of the diodes **370**, **375**, and **380** allows substantially only filtered DC power to charge the capacitors **350**, **355**, **360**, and **365**.

[0060] At stage **630**, the battery **390** is charged using power from the output node **347** and the capacitors **350**, **355**, **360**, and **365**. The amount of time the battery **390** is charged varies, and may be tailored to suit a specific application. The battery **390** may be charged for a specific predetermined amount of time, or may be charged until a certain power threshold is reached.

[0061] At stage **635** the switch **385** is actuated into the second state (e.g., closed) coupling the load **320** across the terminals **391** and **392** of the battery **392**, thereby providing power from the battery **390** to the LED **320**. Power may also be provided to the load **320** from the output node **347** and/or the capacitors **350**, **355**, **360**, and **365**. The stages **615**, **620**, **625**, **630**, and **635** may be repeated.

[0062] At stage **640**, the power from the capacitors **350**, **355**, **360**, and **365**, and the battery **390** is used to operate the load **320**, here causing the LED to emit light. The process **600** returns to stage **615** where the switch **385** is changed from the second state to the first state, thereby decoupling the load **320** from the positive terminal **391** of the battery **390**, the output node **347**, and the capacitor **350**. The switch **385** thus alternates between the first state and the second state to provide intermittent power to the LED **320**. Alternatively, the switch **385** can remain in the second state to provide substantially constant power to the LED **320**. Other modes of operation are also possible.

[0063] In operation, referring to FIG. **8**, with further reference to FIG. **1**, a process **1000** for determining storm distance and/or severity by measuring the voltage provided by the tree **25** includes the stages shown. The process **1000**, however, is exemplary only and not limiting. The process **1000** may be altered, e.g., by having stages added, removed, or rearranged.

[0064] At stage **1005**, the voltage provided by the tree **25** is measured using the apparatus **1**. Voltage values are recorded, e.g., at regular time intervals such as every 30 seconds, although other intervals are possible. Preferably, the apparatus **1** is not used to provide power to a load (e.g., the load **35**) during stage **1005**, although the apparatus **1** can provide power to a load simultaneously with the voltage measurements. The voltage can be measured, for example, by a computer and/or manually.

[0065] At stage **1010**, the voltage measurements are tracked. For example, a computer system can collect the voltage readings at regular intervals and store the values in a data table with each entry in the table representing a discrete voltage measurement at a known time. Alternatively, a person taking manual measurements can record the measurements manually.

[0066] At stage **1015**, the voltage measurements are compared to a baseline voltage for the tree **25** (e.g., a voltage value collected on a clear day). If the voltage measurements decrease relative to the baseline voltage of the tree **25**, then a conclusion can be reached and an indication can be provided that a storm (e.g., a lightning storm) is approaching. The amount of the voltage drop and/or the speed of the voltage drop when compared to the baseline voltage can be used to determine the severity and/or the distance of an approaching storm. For example, a 0.5V drop in twenty minutes (with the baseline voltage as a reference point) can result in a determination that a more severe storm is approaching than a 0.2V drop in an hour (with the baseline voltage as a reference point). The voltage readings collected and tracked at stages **1005** and **1010** can be used at stage **1015** to determine information about an approaching storm alone (e.g., distance and/or severity), or can be combined with other weather tools, such as Radar and/or satellite imagery, used in predicting weather conditions.

Experiment 1

[0067] Referring to Appendix A, exemplary results of voltage yield tests from different trees using different tap configurations, different ground rod quantities, and different numbers of taps are shown. The tests were performed using the configuration shown in FIG. **1**, and described in the corresponding written description, where the load was a voltmeter. The circuit **30**, however, as shown in FIG. **1**, was omitted in the tests. The tests were performed selecting different geographic locations of the trees, different types of trees, different tap materials, different tap depths, different tap diameters, different tap heights (i.e., height from ground level), different tree altitudes, varying numbers of taps, and varying soil conditions. As shown in Appendix A, factors such as the species and/or the variety of a particular plant, e.g., tree, affects the available voltage and/or current. For example, an oak tree located 40 feet above sea level and a maple tree located 200 feet above sea level provided differing amounts of voltage and/or current. Trees produced a substantially constant DC voltage (and some AC voltage), while other plants produced a less-constant DC voltage than trees. Furthermore, two trees, providing about 0.75V and 0.8V (DC), respectively, were coupled in series. Approximately 0.8V was measured from the second of the two tree coupled in series.

Experiment 2

[0068] The charging circuit **50** (of FIG. **2**) was used to successfully power an LED. The charging circuit **50**, using four 10,000 μF (35 Vdc) capacitors, successfully illuminated an SSL-DSP5093UWC LED (manufactured by Lumex Incorporated, of Palatine, Ill.) for approximately one second. The charging circuit **50** was placed in the charging state for approximately 1.75 hours, thereby charging the capacitors **90**, **95**, **100**, and **105**. At the end of the charging period, there was approximately a 0.5 Vdc potential in each of the capacitors **90**, **95**, **100**, and **105**, storing approximately 0.0125 Joules of energy in each of the capacitors **90**, **95**, **100**, and **105**. To light the LED, the switches **55**, **60**, **65**, **70**, **75**, **80**, and **85** were actuated, changing the switches **55**, **60**, **65**, **70**, **75**, **80**, and **85** from the first (charging) state, to the second (discharge) state, thereby providing 2 Vdc to the LED (4×0.5 Vdc) and illuminating the LED. After approximately one second of the LED being illuminated, the voltage across the LED dropped to 1.5 Vdc and the LED no longer illuminated (the lower operating threshold of the SSL-DSP5093UWC LED is approximately 1.5V). The capacitors **90**, **95**, **100**, and **105** were allowed to recharge for approximately one hour to again reach a 0.5 Vdc potential across each of the capacitors **90**, **95**, **100**, and **105**.

Experiment 3

[0069] The apparatus was used to collect weather related information (exemplary data is shown in Appendix B). Voltage readings were collected as a lightning storm approached from the West of a test site including a tree. As the storm approached the test site, a voltage provided by the tree decreased relative to prior levels. The closer the storm was relative to the test site, the larger the voltage drop. For example, when the storm was several miles away, the voltage provided by the tree dropped about 0.2V compared to a voltage measured from the tree on a clear day. As the storm had substantially reached the test site, the voltage provided by the tree had dropped approximately 0.5V compared to the voltage measured from the tree on a clear day. The approaching storm was an intense lightning storm, including positive lightning. Data consistent with the above description was recorded during other lightning storms. Observations indicate that stronger electrical activity (e.g., lightning) produced by a storm caused a quicker and larger voltage drop. Thus, by measuring the voltage provided by the tree **25**, it was possible to gather information regarding the severity of an approaching storm. After a storm had passed over the test site, the voltage provided by the tree would return to "normal" levels within about thirty-five to forty minutes.

Experiment 4

[0070] A modified version of the apparatus **1** shown in FIG. **1** was used to stimulate/enhance the growth of plants including tomato and broccoli plants. Providing electricity produced by a tree to a plant was found to increase growth of the plant, to increase the plant's resistance to pests, and to increase the plant's resistance to frost. A tree was coupled to a plant using the tap **5** and the wire **15**, with the plant being the load **35**. The plant's root system replaced the conductor **10**.

Broccoli Plant

[0071] One of several broccoli plants in a group near each other was coupled to an apple tree as described above during an entire growing season. Prior to coupling the apple tree to the broccoli plant, the apple tree produced about 1.1 Vdc and the broccoli plant produced an average of about 0.3 Vdc. As the growing season progressed, the "energized" broccoli plant showed increased growth and increased resistance to pests relative to the other neighboring broccoli plants. For example, the energized broccoli plant grew taller than the other neighboring broccoli plants, and produced a larger center head and more side heads than the other neighboring broccoli plants. An additional experiment was performed by energizing the smallest and weakest broccoli plant of the group of broccoli plants. Within about two to three days of being energized, the newly-energized broccoli plant was about the same size and height as the neighboring non-energized broccoli plants.

[0072] The energized broccoli plant was not bothered by pests, whereas the non-energized broccoli plants were attacked by pests. As determined by several visual inspections during the growing season, the energized broccoli plant was substantially untouched by pests, whereas the non-energized broccoli plants' leaves were eaten by pests. As a further experiment, a worm was placed on the energized broccoli plant and then onto one of the other broccoli plants. After being placed on the "energized" broccoli plant, the worm did not eat the broccoli plant and fell off. When the same worm was placed on the non-energized broccoli plant, the worm began eating the broccoli plant soon thereafter. An additional experiment was performed by energizing a pest-inhabited broccoli plant. Within about one hour of being energized, the pests inhabiting the broccoli plant vacated the plant.

Tomato Plant

[0073] One of several Cherokee Purple tomato plants in a group near each other was coupled to an elm tree. Prior to coupling the elm tree to the tomato plant, the elm tree produced about 1.2 Vdc. The energized/connected tomato plant included four shoots, each of which were coupled to the elm tree. Visual inspections of the tomato plant revealed that the energized tomato plant grew approximately thirty-three percent higher than the non-energized plants. The energized tomato plant also produced more tomatoes than the non-energized tomato plants. Furthermore, the energized tomato plant survived the first two frosts of the winter season, whereas the non-energized tomato plants died after the first frost.

[0074] Other embodiments are within the scope and spirit of the invention, including the appended claims. Features implementing functions may be physically located at various positions, including being distributed such that portions of functions are implemented at different physical locations. Loads other than LEDs may be used, such as a transmitter, receiver, microchip, incandescent light source, infrared light source, a laser, a DC/DC voltage converter, a DC/AC inverter, etc. Power may be drawn from non-animal organisms other than trees. For example, broccoli plants, tomato plants, soybean plants, and mushrooms may be used. Also, potted plants, and potted trees may be used. The tap may be inserted into a branch of the tree. The load can draw more current from the tree using multiple ground rods.

[0075] While the tap has been disclosed as a nail, other configurations are possible such as a staple. Non-invasive embodiments of the tap are possible, e.g., a donut-shaped transformer core placed around the circumference of a tree. The surface area of a tap may be increased by, for example, being threaded (e.g., being a screw) and/or placing outwardly disposed barbs on the tap. A tap may have a flange disposed around the circumference of the tap to help a user insert the tap correctly into a tree (e.g., to the correct depth). A tap may include a handle to help in insertion into the tree and/or removal from the tree.

[0076] While the terms "connected," "connector," "coupled," and "connection" have been used to indicate a direct connection, other configurations are possible. For example, referring to FIG. **6**, when the diode **380** is "coupled" to the capacitor **360**, this may include indirect connection through another component (e.g., a resistor coupled between the cathode **382** of the diode **380** and the capacitor **360**).

[0077] The word "or" is to be construed as including the conjunctive and disjunctive definition.

[0078] Further, while the description refers to the invention, the description may include more than one invention.

APPENDIX A

				POWER SOURCE DATA COLLECTION			Height			
Test No.	Time Intervals	Voltage DC	Tree Type	Nail Type	Nail Penetration Depth	Nail Diameter	from Ground	No. of Nails	Soil Type	Altitude
1	7:00 PM	0.9 VDC	PINE	STAINLESS	¾"	⅛"	3 FT	2	LOAM	
2	7:25	0.9 VDC	PINE	STAINLESS	¾"	⅛"	4 FT	2	LOAM	
3	7:40	0.9 VDC	PINE	STAINLESS	¾"	⅛"	5 FT	2	LOAM	
1	1:00 PM	1.0 VDC	PINE	STAINLESS	¾"	⅛"	5 FT	2	CLAY-SAND	
1	10 MIN	−1.2	EIM	¾"	⅜		18"	1	SAND	40
2		−1.6	BLUE SPRUCE	¾"	⅜		18"	1	SAND	40
3		−1.0	MAPLE	¾"	⅜		18"	1	SAND	40
4		−1.1	MAPLE	¾"	⅜		18"	1	SAND	40
5		−1.2	EIM	¾"	⅜		18"	1	SAND	40
6		−1.1	WALNUT	¾"	⅜		18"	1	SAND	40
7		−0.8	LILAC BUSH	¾"	⅜		18"	1	SAND	40
8		−1.1	ELM	¾"	⅜		18"	1	SAND	40

APPENDIX A-continued

				POWER SOURCE DATA COLLECTION						
Test No.	Time Intervals	Voltage DC	Tree Type	Nail Type	Nail Penetration Depth	Nail Diameter	Height from Ground	No. of Nails	Soil Type	Altitude
9		−1.6	BLUE SPRUCE	¾"	⅜		18"	1	SAND	40
10		−1.1	MAPLE	¾"	⅜		18"	1	SAND	40
11		−1.1	MAPLE	¾"	⅜		18"	1	SAND	40
12		−1.4	BIRCH	¾"	⅜		18"	1	SAND	40
13		−1.4	BIRCH	¾"	⅜		36"	1	SAND	40
14		−1.5	BIRCH	¾"	⅜		2"	1	SAND	40
15		−1.2	OAK	¾"	⅜		18"	4	SAND	40
16		−1.2	ELM	¾"	⅜		18"	1	SAND	40
17		−1.5	APPLE	¾"	⅜		18"	1	SAND	40
18		−1.5	APPLE	¾"	⅜		36"	1	SAND	40
19		−1.3	OAK	¾"	⅜		18"	1	SAND	40
20		−1.2	MAPLE	¾"	⅜		18"	1	SAND	40
21		−0.8	? BUSH	¾"	⅜		12"	1	SAND	40
22		−1.1	ELDER	¾"	⅜		18"	1	SAND	40
23		−1.6	SPRUCE	¾"	⅜		18"	1	SAND	40
24		−1.2	OAK	¾"	⅜		18"	1	SAND	40
25		−1.1	GREEN	¾"	⅜		18"	1	SAND	40
26		−1.1	SPRUCE	¾"	⅜		36"	1	SAND	40
27		−1.1		¾"	⅜		48"	1	SAND	40
28		−1.1		¾"	⅜		8"	1	SAND	40
29		−1.1		¾"	⅜		2"	1	SAND	40
30		−1.1		¾"	⅜		4"	1	SAND	40
31		−1.0	BIRCH	¾"	⅜		18"	1	SAND	40
32		−1.0	BIRCH	¾"	⅜		12"	1	SAND	40
33		−1.0	BIRCH	¾"	⅜		5"	1	SAND	40
34		−1.1	MAPLE	¾"	⅜		18"	1	SAND	40
35		−1.4	OAK	¾"	⅜		18"	1	SAND	40
36		−0.9	?	¾"	⅜		12"	1	SAND	40
37		1.1	ELM	¾"	⅜		18"	1	SAND	40
38		1.2	ELM	¾"	¼ to ⅜		18"	1	SAND	APPOX 60
39		1.1	OAK	¾"	¼ to ⅜		18"	1	SAND	APPOX 60
40		1.1	OAK	¾"	¼ to ⅜		18"	1	SAND	APPOX 60
41		1.2	ELM	¾"	¼ to ⅜		18"	1	SAND	APPOX 60
42		1.0	BIRCH	¾"	¼ to ⅜		18"	1	SAND	APPOX 60
43		1.2	MAPLE	¾"	¼ to ⅜		18"	1	SAND	APPOX 60
44		1.4	BLUE SPRUCE	¾"	¼ to ⅜		18"	1	SAND	APPOX 60
45		1.1	ELM	¾"	¼ to ⅜		18"	1	SAND	APPOX 60
45		1.3	MAPLE	¾"	¼ to ⅜		18"	1	SAND	APPOX 60
47		1.1	MAPLE	¾"	¼ to ⅜		18"	1	SAND	APPOX 60
48		−1.2	APPLE	¾"	¼ to ⅜		18"	5	SAND	40
49		−1.2	APPLE	¾"	¼ to ⅜		30"	4	SAND	40
50		−1.3	WILLOW	¾"	¼ to ⅜		18"	1	SAND	40
51		−1.3	WILLOW	¾"	¼ to ⅜		24"	1	SAND	40
52		−1.3	WILLOW	¾"	¼ to ⅜		36"	1	SAND	40
53		−1.0	MAPLE	¾"	¼ to ⅜		18"	1	SAND	40
54		−1.1	MAPLE	¾"	¼ to ⅜		0"	1	SAND	40
55		−1.2	ELM	¾"	¼ to ⅜		18"	1	SAND	40
56		1.3	OAK	¾"	¼ to ⅜		18"	1	SAND CLAY	120
57		1.1	ELM	¾"	¼ to ⅜		18"	1	SAND CLAY	120
58		1.4	SASAFRAS	¾"	¼ to ⅜		18"	1	SAND CLAY	120
59		1.0	OAK	¾"	¼ to ⅜		18"	1	SAND CLAY	120
60		1.0	OAK	¾"	¼ to ⅜		38"	1	SAND CLAY	120
61		1.2	OAK	¾"	¼ to ⅜		0"	1	SAND CLAY	120
62		1.3	SPRUCE	¾"	¼ to ⅜		18"	1	SAND CLAY	120
63		1.4	SPRUCE	¾"	¼ to ⅜		30"	1	SAND CLAY	120
64		1.2	MAPLE	¾"	¼ to ⅜		18"	1	SAND CLAY	120
65		1.1	CEDAR	¾"	¼ to ⅜		18"	1	SAND	40
66		1.4	CHERRY	¾"	¼ to ⅜		18"	1	SAND	40
67		1.4	CHERRY	¾"	¼ to ⅜		12"	1	SAND	40
68		1.5	CHERRY	¾"	¼ to ⅜		0"	1	SAND	40
69		1.4	CHERRY	¾"	¼ to ⅜		36"	1	SAND	40
70		1.1	CEDAR	¾"	¼ to ⅜		18"	1	SAND	40
71		1.2	MAPLE	¾"	¼ to ⅜		18"	1	SAND	40
72		1.2	MAPLE	¾"	¼ to ⅜		36"	1	SAND	40
73		1.3	MAPLE	¾"	¼ to ⅜		0"	1	SAND	40
112		0.9	CEDAR	¾"	¼ to ⅜		18"	1	SAND	40'
113		0.9	CEDAR	¾"	¼ to ⅜		36"	1	SAND	40'
114		1.0	CEDAR	¾"	¼ to ⅜		0"	1	SAND	40'
115		1.0	CEDAR	¾"	¼ to ⅜		0"	1	SAND	40'
116		1.3	OAK	¾"	¼ to ⅜		18"	1	SAND	200'

APPENDIX A-continued

				POWER SOURCE DATA COLLECTION						
Test No.	Time Intervals	Voltage DC	Tree Type	Nail Type	Nail Penetration Depth	Nail Diameter	Height from Ground	No. of Nails	Soil Type	Altitude
117		1.3	OAK	¾"	¼ to ⅜		24"	1	SAND	200'
118		1.3	OAK	¾"	¼ to ⅜		48"	1	SAND	200'
119		1.3	OAK	¾"	¼ to ⅜		0"	1	SAND	200'
120		1.1	MAPLE	¾"	¼ to ⅜		18"	1	SAND	200'
121		1.1	MAPLE	¾"	¼ to ⅜		24"	1	SAND	200'
122		1.2	MAPLE	¾"	¼ to ⅜		0"	1	SAND	200'
123		1.4	SPRUCE	¾"	¼ to ⅜		18"	1	SAND	200'
124		1.5	SPRUCE	¾"	¼ to ⅜		0"	1	SAND	200'
125		1.2	OAK	¾"	¼ to ⅜		18"	1	SAND	200'
126		1.2	OAK	¾"	¼ to ⅜		24"	1	SAND	200'
127		1.3	OAK	¾"	¼ to ⅜		0"	1	SAND	200'
128		1.0	MAPLE	¾"	¼ to ⅜		18"	1	SAND	200'
129		1.0	MAPLE	¾"	¼ to ⅜		24"	1	SAND	200'
130		1.0	MAPLE	¾"	¼ to ⅜		36"	1	SAND	200'
131		1.1	MAPLE	¾"	¼ to ⅜		0"	1	SAND	200'
132		−1.2	MAPLE	¾"	¼ to ⅜		24"	1	SAND	200'
133		−1.2	MAPLE	¾"	¼ to ⅜		24"	4	SAND	200'
134		−1.2	ELM	¾"	¼ to ⅜		18"	1	SAND	200'
135		−1.2	ELM	¾"	¼ to ⅜		30"	1	SAND	200'
136		−1.3	ELM	¾"	¼ to ⅜		44"	8	SAND	200'
137		−1.2	ELM	¾"	¼ to ⅜		60"	1	SAND	200'
138		−1.4	SPRUCE	¾"	¼ to ⅜		8"	1	SAND	200'
139		1.2	ELM	¾"	¼ to ⅜		20"	1	SAND	140'
140		1.2	ELM	¾"	¼ to ⅜		28"	1	SAND	140'
141		1.2	ELM	¾"	¼ to ⅜		36"	1	SAND	140'
142		1.4	ELM	¾"	¼ to ⅜		0"	1	SAND	140'
143		1.6	SPRUCE	¾"	¼ to ⅜		18"	1	SAND	140'
144		1.6	SPRUCE	¾"	¼ to ⅜		30"	1	SAND	140'
145		1.4	SPRUCE	¾"	¼ to ⅜		0"	1	SAND	140'
146		1.1	MAPLE	¾"	¼ to ⅜		18"	1	SAND	140'
147		1.1	MAPLE	¾"	¼ to ⅜		24"	1	SAND	140'
148		1.1	MAPLE	¾"	¼ to ⅜		46"	1	SAND	140'
149		1.3	MAPLE	¾"	¼ to ⅜		0"	1	SAND	140'
150		1.1	OAK	¾	¼–⅜		18"		SAND	140'
151		1.1	OAK	¾	¼–⅜		28"		SAND CLAY	140'
152		1.1	OAK	¾	¼–⅜		38"		SAND CLAY	140'
153		1.2	OAK	¾	¼–⅜		49"		SAND CLAY	140'
154		1.2	OAK	¾	¼–⅜		0"		SAND CLAY	140'
155		0.9	RED OAK	¾	¼–⅜		18"	1	SAND CLAY	140'
156		0.9	RED OAK	¾	¼–⅜		30"	1	SAND CLAY	140'
157		0.8	RED OAK	¾	¼–⅜		56"	1	SAND CLAY	140'
158		1.1	RED OAK	¾	¼–⅜		0"	1	SAND CLAY	140'
159		1.2	SUGAR MAPLE	¾	¼–⅜		18"	1	SAND CLAY	140'
160		1.2	SUGAR MAPLE	¾	¼–⅜		25"	1	SAND CLAY	140'
161		1.3	SUGAR MAPLE	¾	¼–⅜		0"	1	SAND CLAY	140'
162		1.4	SUGAR MAPLE	¾	¼–⅜		18"	1	SAND CLAY	140'
163		1.2	BLACK CHERRY	¾	¼–⅜		17"	1	SAND CLAY	140'
164		1.2	BLACK CHERRY	¾	¼–⅜		25"	1	SAND CLAY	140'
165		1.3	BLACK CHERRY	¾	¼–⅜		0"	1	SAND CLAY	140'
166		1.4	BLACK CHERRY	¾	¼–⅜		20"	12	SAND CLAY	140'
167		1.4	PEAR	¾	¼–⅜		0"	1	SAND CLAY	140'
168		1.1	PEAR	¾	¼–⅜		18"	1	SAND CLAY	140'
169		1.1	WILLOW	¾	¼–⅜		27"	1	SAND CLAY	140'
170		1.3	WILLOW	¾	¼–⅜		0"	1	SAND CLAY	140'
171		1.6	WILLOW	¾	¼–⅜		18"	1	SAND CLAY	140'
172		1.1	SPRUCE	¾	¼–⅜		20"	1	SAND CLAY	140'
173		1.1	BEECH	¾	¼–⅜		30"	1	SAND	40'
174		1.1	BEECH	¾	¼–⅜		40"	1	SAND	40'
175		1.1	BEECH	¾	¼–⅜		50"	1	SAND	40'
176		1.0	BEECH	3 inch	¼–⅜		20"	1	SAND	40'
177		1.0	BEECH	5 inch	¼–⅜		20"	1	SAND	40'
178		1.2	BEECH	staple	¼–⅜		20"	1	SAND	40'
179		1.0	ELM	¾	¼–⅜		18"	1	SAND	40'
180		1.0	ELM	3 inch	¼–⅜		36"	1	SAND	40'
181		0.9	ELM	5 inch	¼–⅜		36"	1	SAND	40'
182		1.2	ELM	staple	¼–⅜		36"	1	SAND	40'
183		1.1	BIRCH	¾	¼–⅜		18"	1	SAND	40'
184		1.3	ELM	¾	¼–⅜		18"	1	SAND	140'
185		1.3	ELM	¾	¼–⅜		36"	1	SAND	140'
186		1.4	ELM	¾	¼–⅜		0"	1	SAND	140'

APPENDIX A-continued

				POWER SOURCE DATA COLLECTION						
Test No.	Time Intervals	Voltage DC	Tree Type	Nail Type	Nail Penetration Depth	Nail Diameter	Height from Ground	No. of Nails	Soil Type	Altitude
187		1.4	SPRUCE	¾"	¼ to ⅜		18"	1	SAND	140
188		1.4	SPRUCE	¾"	¼ to ⅜		34"	1	SAND	140
189		1.5	SPRUCE	¾"	¼ to ⅜		0	1	SAND	140
190		1.3	OAK	¾"	¼ to ⅜		18"	1	SAND	140
191		1.3	OAK	¾"	¼ to ⅜		36"	1	SAND	140
192		1.3	OAK	¾"	¼ to ⅜		48"	1	SAND	140
193		1.4	OAK	¾"	¼ to ⅜		0	1	SAND	140
194		1.3	APPLE?	¾"	¼ to ⅜		18"	1	SAND	140
195		1.3	APPLE	¾"	¼ to ⅜		30"	1	SAND	140
196		1.1	PINE	¾"	¼ to ⅜		18"	1	SAND	140
197		1.1	PINE	¾"	¼ to ⅜		36"	1	SAND	140
198		1.0	MAPLE	¾"	¼ to ⅜		18"	1	SAND	40
199		1.0	MAPLE	¾"	¼ to ⅜		36"	1	SAND	40
200		1.2	BLACK CHERRY	¾"	¼ to ⅜		12"	1	SAND	40
201		1.2	BLACK CHERRY	¾"	¼ to ⅜		20"	1	SAND	40
202		1.2	BLACK CHERRY	¾"	¼ to ⅜		48"	1	SAND	40
203		1.3	BLACK CHERRY	¾"	¼ to ⅜		0	1	SAND	40
204		1.1	LILAC	¾"	¼ to ⅜		14"	1	SAND	40
205		1.1	LILAC	¾"	¼ to ⅜		22"	1	SAND	40
206		1.1	LILAC	¾"	¼ to ⅜		40"	1	SAND	40
207		1.1	ELM	¾"	¼ to ⅜		18"	1	SAND	40
208		1.1	ELM	¾"	¼ to ⅜		36"	1	SAND	40
209		1.1	ELM	¾"	¼ to ⅜		50"	1	SAND	40
210		1.3	SPRUCE	¾"	¼ to ⅜		18"	1	SAND	40
211		1.3	SPRUCE	¾"	¼ to ⅜		30"	1	SAND	40
212		1.3	SPRUCE	¾"	¼ to ⅜		50"	1	SAND	40
213		1.3	SPRUCE	¾"	¼ to ⅜		74"	1	SAND	40
214		−1.2	ELM	¾"	¼ to ⅜		20"	8	SAND	40
215		−1.2	ELM	¾"	¼ to ⅜		36"	1	SAND	40
216		−1.3	ELM	¾"	¼ to ⅜		0	1	SAND	40
217		−1.1	MAPLE	¾"	¼ to ⅜		18"	1	SAND	40
218		−1.3	APPLE	¾"	¼ to ⅜		14"	1	SAND	40
219		−1.3	APPLE	¾"	¼ to ⅜		25"	1	SAND	40
220		−1.3	APPLE	¾"	¼ to ⅜		50"	1	SAND	40
221		−1.4	SPRUCE	¾"	¼ to ⅜		14"	1	SAND	40
222		−1.4	SPRUCE	¾"	¼ to ⅜		22"	1	SAND	40
223		−1.4	SPRUCE	¾"	¼ to ⅜		36"	1	SAND	40
224		−1.1	MAPLE	¾"	¼ to ⅜		18"	1	SAND	40
225		−1.1	MAPLE	¾"	¼ to ⅜		36"	1	SAND	40
226		−1.0	ELM	¾"	¼ to ⅜		20"	1	SAND	40
227		−1.0	ELM	¾"	¼ to ⅜		40"	1	SAND	40
228		−1.0	ELM	¾"	¼ to ⅜		50"	1	SAND	40
229		−1.2	BEECH	¾"	¼ to ⅜		18"	1	SAND	40
230		−1.2	BEECH	¾"	¼ to ⅜		24"	1	SAND	40
231		−1.2	BEECH	¾"	¼ to ⅜		38"	1	SAND	40
232		−1.3	OAK	¾"	¼ to ⅜		16"	1	SAND	40
233		−1.3	OAK	¾"	¼ to ⅜		28"	1	SAND	40
234		−1.3	OAK	¾"	¼ to ⅜		38"	1	SAND	40
235		−1.4	OAK	¾"	¼ to ⅜		0	1	SAND	40
236		−1.2	BIRCH	¾"	¼ to ⅜		18"	1	SAND	40
237		−1.3	BIRCH	¾"	¼ to ⅜		30"	1	SAND	40
238		−1.3	BIRCH	¾"	¼ to ⅜		44"	1	SAND	40
239		−1.2	BIRCH	¾"	¼ to ⅜		0	1	SAND	40
240		−1.1	POPLAR	¾"	¼ to ⅜		18"	1	SAND	40
241		−1.1	POPLAR	¾"	¼ to ⅜		24"	1	SAND	40
242		−1.2	POPLAR	¾"	¼ to ⅜		36"	1	SAND	40
243		−1.2	POPLAR	¾"	¼ to ⅜		48"	1	SAND	40
244		−1.1	ELM	¾"	¼ to ⅜		18"	1	SAND	40
245		−1.1	ELM	¾"	¼ to ⅜		28"	1	SAND	40
246		−1.2	BLACKBERRY	¾"	¼ to ⅜		?10"?	1	SAND	40
247		−1.2	BLACKBERRY	¾"	¼ to ⅜		16"	1	SAND	40
248		−0.9	WILLOW	¾"	¼ to ⅜		12"	1	SAND	40
249		−1.0	WILLOW	¾"	¼ to ⅜		20"	1	SAND	40
250		−1.1	WILLOW	¾"	¼ to ⅜		0	1	SAND	40
251		−0.8	BROCOLLI	¾"	¼ to ⅜		8"	1	SAND	40
252		−0.7	BROCOLLI	¾"	¼ to ⅜		LEAF	1	SAND	40
253		−1.1	ELM	¾"	¼ to ⅜		14"	1	SAND	40
254		−1.1	ELM	¾"	¼ t0 ⅜		20"	1	SAND	40
255		−1.0	ELM	¾"	¼ to ⅜		18"	1	SAND	40
256		−1.0	ELM	¾"	¼ to ⅜		36"	1	SAND	40

APPENDIX A-continued

				POWER SOURCE DATA COLLECTION			Height			
Test No.	Time Intervals	Voltage DC	Tree Type	Nail Type	Nail Penetration Depth	Nail Diameter	from Ground	No. of Nails	Soil Type	Altitude
257		−1.1	WALNUT	¾"	¼ to ⅜		18"	1	SAND	40
258		−0.3	WALNUT	¾"	¼ to ⅜		18"	1	SAND	40
259		−0.4	PINE	¾"	¼ to ⅜		16"	1	SAND	40
260		−0.9	PINE	¾"	¼ to ⅜		0	1	SAND	40
261		−1.2	PINE	¾"	¼ to ⅜		20"	1	SAND	40
262		−1.2	PINE	¾"	¼ to ⅜		40"	1	SAND	40
263		−1.3	PINE	¾"	¼ to ⅜		0	1	SAND	40
264		−1.1	LILAC	¾"	¼ to ⅜		12"	1	SAND	40
265		−1.1	LILAC	¾"	¼ to ⅜		18"	1	SAND	40
266		−1.0	MAPLE	¾"	¼ to ⅜		2"	1	SAND	40
267		−1.0	MAPLE	¾"	¼ to ⅜		0	1	SAND	40
268		−1.1	PINE	¾"	¼ to ⅜		18"	1	SAND	40
269		−1.0	PINE	¾"	¼ to ⅜		32"	1	SAND	40
270		−1.3	LEMON	¾"	¼ to ⅜		18"	1	SAND	40
271		−0.9	TOMATO	¾"	¼ to ⅜		6"	1	SAND	40
272		−0.8	CAULIFLOWER	¾"	¼ to ⅜		2"	1	SAND	40
273		0.0	GRASS	¾"	¼ to ⅜		0	Alligator clip	SAND	40
274		−1.1	PINE	¾"	¼ to ⅜		16"	1	SAND	40
275		−1.1	MAPLE	¾"	¼ to ⅜		15"	1	SAND	40
276		−1.1	MAPLE	¾"	¼ to ⅜		28"	1	SAND	40
277		−1.0	MAPLE	¾"	¼ to ⅜		36"	1	SAND	40
278		−1.0	ELM	¾"	¼ to ⅜		25"	1	SAND	40
279		−1.1	ELM	¾"	¼ to ⅜		35"	1	SAND	40
280		−0.9	MAPLE	¾"	¼ to ⅜		18"	1	SAND	40
281		−1.0	MAPLE	¾"	¼ to ⅜		36"	1	SAND	40
282		−1.0	CEDAR	¾"	¼ to ⅜		18"	1	SAND	40
283		−1.1	CEDAR	¾"	¼ to ⅜		30"	1	SAND	40
284		−1.0	BASSWOOD	¾"	¼ to ⅜		20"	1	SAND	40
285		−1.0	BASSWOOD	¾"	¼ to ⅜		36"	1	SAND	40
286		−1.0	BASSWOOD	¾"	¼ to ⅜		48"	1	SAND	40
287		−1.0	BASSWOOD	¾"	¼ to ⅜		65"	1	SAND	40
290		0.0	TELE POLE	¾"	¼ to ⅜		24"	1	SAND	40
291		−0.9	LILAC	¾"	¼ to ⅜		16"	1	SAND	40
293		−1.4	SPRUCE	¾"	¼ to ⅜		18"	1	SAND	40
294		−1.4	SPRUCE	¾"	¼ to ⅜		28"	1	SAND	40
295		−1.3	SPRUCE	¾"	¼ to ⅜		40"	1	SAND	40
296		−1.1	ELM	¾"	¼ to ⅜		16"	1	SAND	40
297		−1.2	APPLE	¾"	¼ to ⅜		16"	1	SAND	40
298		−1.2	APPLE	¾"	¼ to ⅜		24"	1	SAND	40
299		−1.1	MAPLE	¾"	¼ to ⅜		18"	1	SAND	40
300		−1.1	MAPLE	¾"	¼ to ⅜		30"	1	SAND	40
301		−1.2	MAPLE	¾"	¼ to ⅜		0	1	SAND	40
302		−1.2	APPLE	¾"	¼ to ⅜		16"	1	SAND	40
303		−1.2	APPLE BROCCOLI	¾"	¼ to ⅜			1	SAND	40
304		−1.2	APPLE BROCCOLI	¾"	¼ to ⅜			1	SAND	40
305		−1.2	APPLE BROCCOLI	¾"	¼ to ⅜			1	SAND	40

No. of taps	Material	No. of ground rods	Ground material	DC Voltage	AC Voltage	Current (mA)
1	Roofing Nail	1	Copper	1.02	1.20	15
1	Roofing Nail	2	Copper	1.02	1.20	21
1	Roofing Nail	3	Copper	1.02	1.20	28
1	Roofing Nail	6	Copper	1.00	1.20	45
2	Roofing Nail	1	Copper	1.02	1.20	20
2	Roofing Nail	2	Copper	1.00	1.20	27
2	Roofing Nail	3	Copper	1.00	1.20	35
2	Roofing Nail	6	Copper	1.01	1.20	57

Conductor 1	Media	Conductor 2	Media	DC Voltage	AC Voltage	Current (mADC)
Copper	Tree	Copper	Earth	0.50	0.60	10
Copper	Tree	Copper	Tree	0.01	0.00	0.00
Roofing Nail	Tree	Copper	Earth	0.72	0.80	30
Roofing Nail	Tree	Copper	Tree	0.85	0.00	20
Roofing Nail	Tree	Roofing Nail	Tree	0.02	0.00	0.00
Roofing Nail	Tree	Roofing Nail	Earth	0.46	0.50	1.0

Conductor 1	Media	Conductor 2	Media	DC Voltage	AC Voltage	Current (μADC)	Elevation
Roofing Nail	Potted Tree	Copper	Earth	0.60	0.20	22	Ground level
Roofing Nail	Potted Tree	Copper	Earth	0.60	0.20	21	1" thick pine board
Roofing Nail	Potted Tree	Copper	Earth	0.59	0.59	21	16" wooden box
Roofing Nail	Potted Tree	Copper	Earth	0.00	0.00	0.00	Held waist high

[0079] The potted tree was a Norfolk Island Pine approximately three feet tall, which was potted in a plastic pot about 40 mils. thick.

APPENDIX B

Test 1:

Time	Storm distance	DC Voltage from tree
Baseline Voltage	—	1.2 V
11:00 AM	About 100 miles	1.1 V
12:00 PM	About 50-60 miles	1.0 V
		0.5 V
		1.0 V
		0.5 V
2:00 PM	Dissipated	1.2 V

[0080] The 12:00 PM measurements reflect fluctuations when lightning strikes occurred.

Test 2:

Time	Storm distance	DC Voltage from tree
Baseline Voltage	—	1.2 V
3:00 PM	50-60 miles	1.1 V
3:15 PM	40-50 miles	1.0 V
		0.3 V
		1.0 V
		0.3 V
3:30 PM	Dissipated	1.2 V

[0081] The 3:15 PM measurements reflect fluctuations when lightning strikes occurred.

Test 3:

Time	Storm distance	DC Voltage from tree
Baseline Voltage	—	1.1 V
7:45 PM	50-60 miles	1.1 V
7:55 PM	40-50 miles	0.72 V
		0.85 V
		0.72 V
9:16 PM	Dissipated	1.1 V

[0082] The 7:55 PM measurements reflect fluctuations when lightning strikes occurred.

What is claimed is:

1. A method for drawing electricity from a non-animal organism, the method comprising:

coupling a first electrical conductor to the non-animal organism;

coupling a second electrical conductor to a ground rod;

embedding the ground rod into soil at a predetermined depth as a function of a desired current level, whereby the current available from the non-animal organism is increased by increasing the depth that the ground rod is embedded into the soil;

coupling an electrical load between the first electrical conductor and the second electrical conductor, the electrical load being configured to draw electricity from the non-animal organism via the first electrical conductor; and

operating the electrical load using electricity drawn from the non-animal organism.

2. A system comprising:

a non-animal organism;

a first electrical conductor electrically coupled to the non-animal organism;

a plurality of ground rods embedded into soil wherein a quantity of the plurality of ground rods is a function of a desired current level from the non-animal organism, whereby the current available from the non-animal organism is increased by increasing the quantity of the plurality of ground rods;

a second electrical conductor coupled to the plurality of ground rods; and

an electrical load coupled between the first electrical conductor and the second electrical conductor to draw electricity from the non-animal organism, the electrical load using electricity drawn from the non-animal organism.

3. A method of predicting weather using electricity from a non-animal organism, the method comprising:

coupling a first electrical conductor to the non-animal organism;

coupling a second electrical conductor to a ground;

coupling a voltmeter between the first electrical conductor and the second electrical conductor;

measuring a voltage potential between the first and second electrical connectors;

providing a weather prediction as a function of the measured voltage potential.

4. The method of predicting weather of claim **3** further comprising:

determining a baseline voltage reading for the non-animal organism under a first weather condition;

determining a plurality of voltage readings over time;

comparing each of the plurality of voltage readings to the baseline voltage reading to determine differences between the baseline voltage reading and each of the plurality of voltage readings; and

calculating information indicative of a second, future weather condition as a function of the differences.

5. A system for use with live vegetative matter growing in soil, the system comprising:

a non-animal organism;

a first electrical conductor electrically coupled to the non-animal organism; and

a second electrical conductor coupled to the first electrical conductor and coupled to the live vegetative matter, the second electrical conductor providing electricity from the non-animal organism to the live vegetative matter;

wherein the growth of the live vegetative matter is stimulated by the electricity provided by the non-animal organism.

6. A system comprising:

a non-animal organism;

a first electrical conductor electrically coupled to the non-animal organism;

a second electrical conductor coupled to a ground; and

an electrical load coupled between the first electrical conductor and the second electrical conductor to draw electricity from the non-animal organism, the electrical load using electricity drawn from the non-animal organism;

wherein the load is selected from the group consisting of:

a battery;
a battery charging circuit;
a sensor;
a radio frequency identification chip;
a transmitter;
a receiver;
a global positioning service (GPS) device;
a light emitting device; and
a fire ignition system.

7. The system of claim **6** wherein the load is the sensor and the sensor is selected from the group consisting of:

an oxygen sensor;
an air-speed sensor;
a humidity sensor;
a barometric pressure sensor;
a camera;
a photoelectric sensor;
a altitude sensor;
a smoke detector;
a microphone; and
a vibration sensor.

8. The system of claim **6** wherein the load is the GPS device and the GPS device is selected from the group consisting of:

a GPS receiver;
a GPS transmitter;
a GPS guidance system; and
a GPS navigation system.

9. The system of claim **6** wherein the load is the light emitting device and the light emitting device is selected from the group consisting of:

a light emitting diode configured to emit visible light; and

an infrared light emitting diode configured to emit an infrared signal.

* * * * *

US 20090026579A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2009/0026579 A1**

Silver et al. (43) **Pub. Date: Jan. 29, 2009**

(54) **EM RECTIFYING ANTENNA SUITABLE FOR USE IN CONJUNCTION WITH A NATURAL BREAKDOWN DEVICE**

(76) Inventors: **Guy Silver**, Sunnyvale, CA (US); **Juinerong Wu**, Sunnyvale, CA (US)

Correspondence Address:
MACPHERSON KWOK CHEN & HEID LLP
2033 GATEWAY PLACE, SUITE 400
SAN JOSE, CA 95110 (US)

(21) Appl. No.: **11/577,172**

(22) PCT Filed: **Oct. 7, 2005**

(86) PCT No.: **PCT/US05/36398**

§ 371 (c)(1),
(2), (4) Date: **Apr. 12, 2007**

(30) **Foreign Application Priority Data**

Oct. 12, 2004 (US) 10/963,357

Publication Classification

(51) **Int. Cl.**

H01L 29/861	(2006.01)
H01L 21/20	(2006.01)
H01Q 21/08	(2006.01)
H01P 11/00	(2006.01)

(52) **U.S. Cl.** **257/603**; 438/380; 343/824; 29/600; 136/244; 257/E21.09; 257/E29.335

(57) **ABSTRACT**

A rectenna capable of power conversion from electromagnetic (EM) waves of high frequencies is provided. In one embodiment, a rectenna element generates currents from two sources—based upon the power of the incident EM wave and from an n-type semiconductor, or another electron source attached to a maximum voltage point of an antenna element. The combined current from both sources increases the power output of the antenna, thereby increasing the detection sensitivity of the antenna of a low power signal. Full wave rectification is achieved using a novel diode connected to a gap in the antenna element of a rectenna element. The diode is conductive at forward bias voltage or reverse bias voltage, and rectifies the antenna signal generated by the desired EM wave received by antenna raise from The rectenna element of the present invention may be used as a building block to create large rectenna arrays.

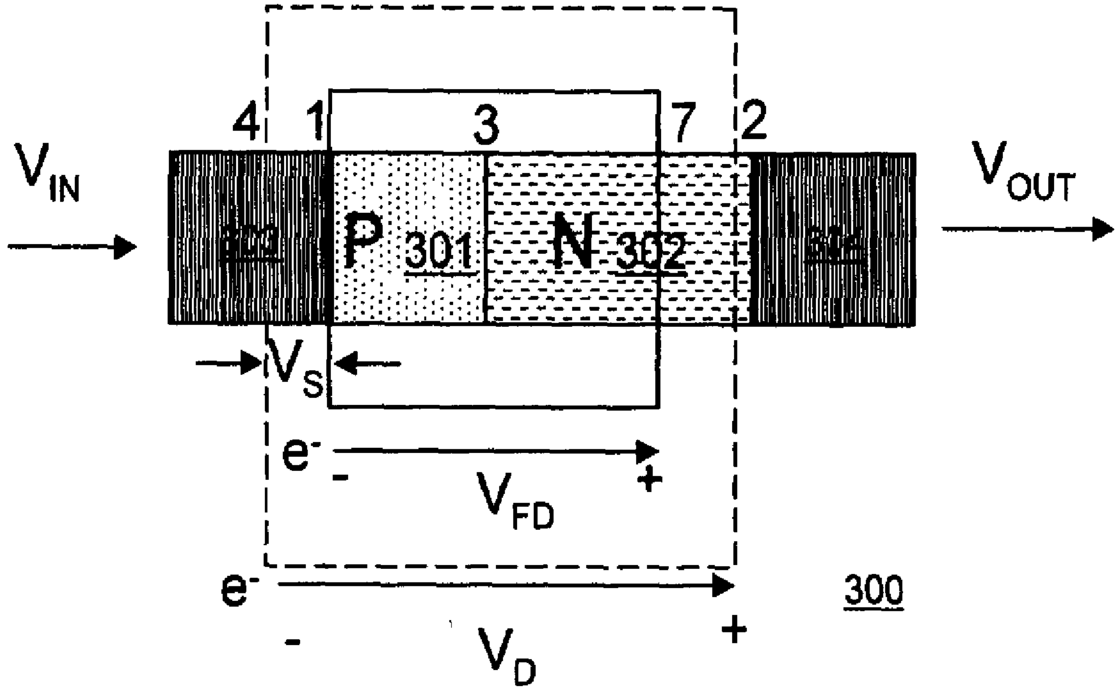

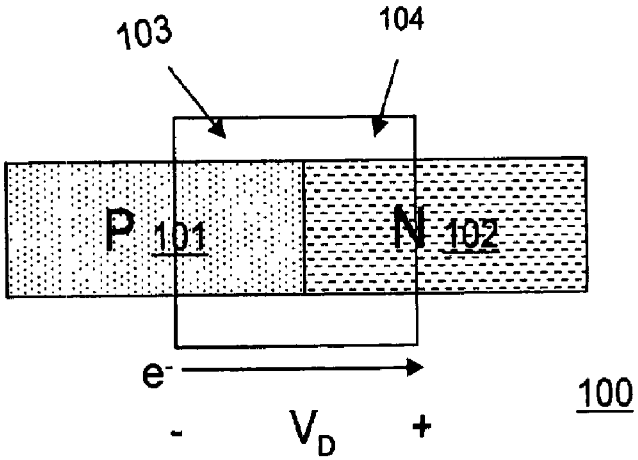
103
104
P 101
N 102
e-
-
V_D
+
100

Figure 1 (prior art)

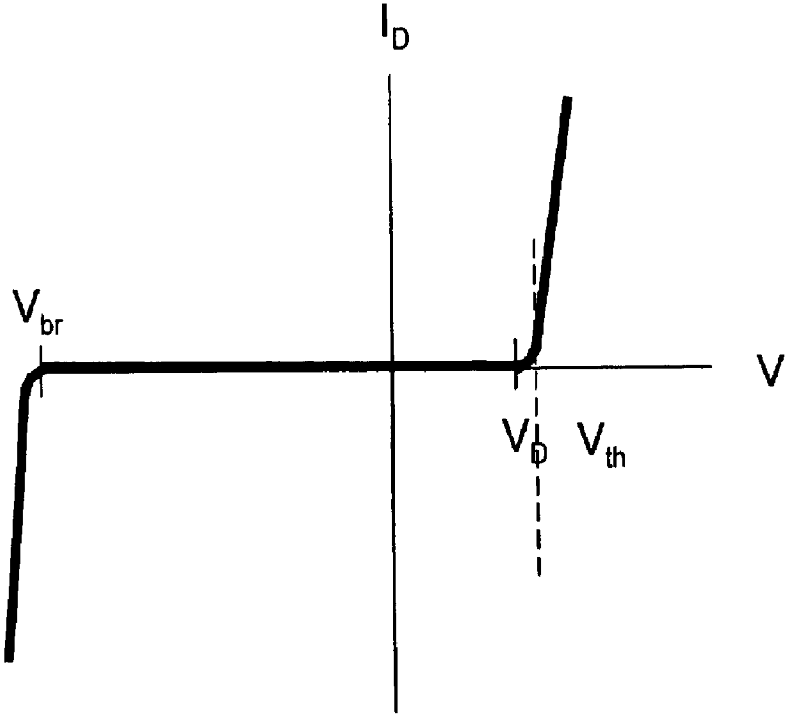

Figure 2 (prior art)

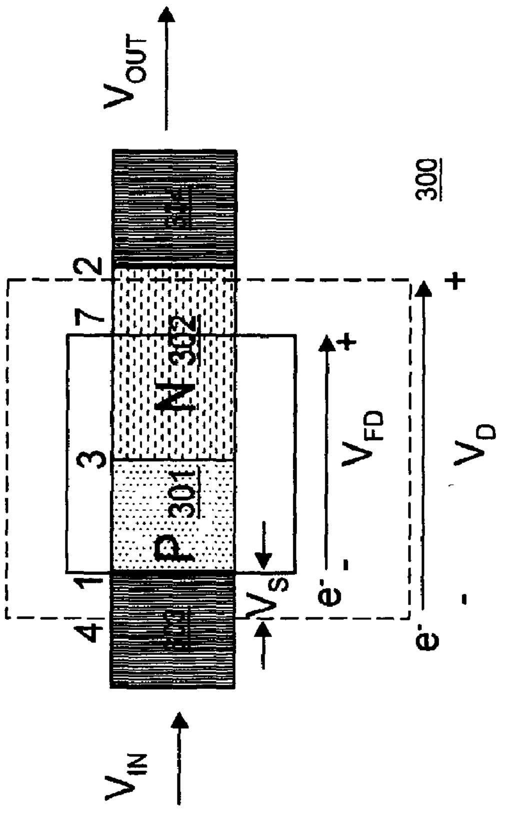
V_{IN}
V_{OUT}
4
1
3
7
2
P
301
N
302
V_S
e^-
-
+
V_{FD}
V_D
300

Figure 3

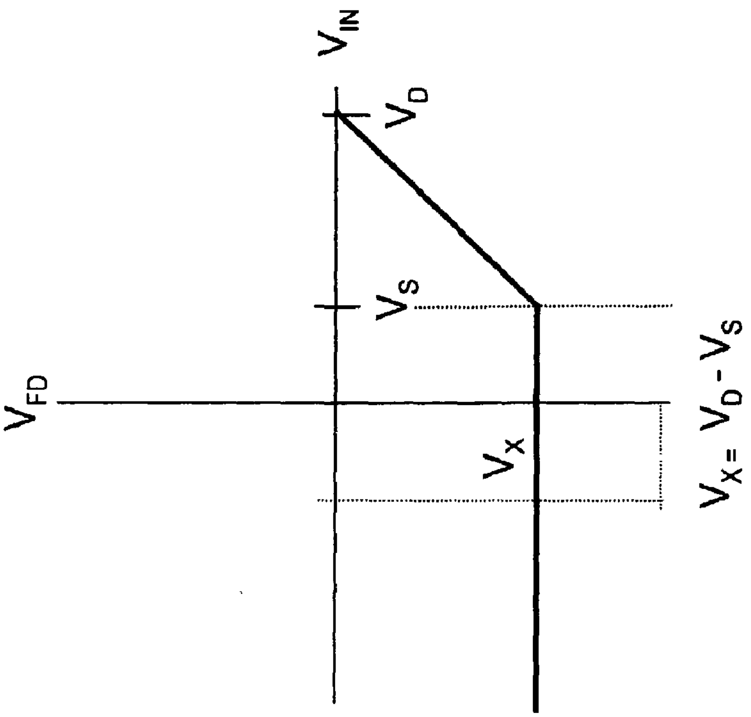

Figure 4

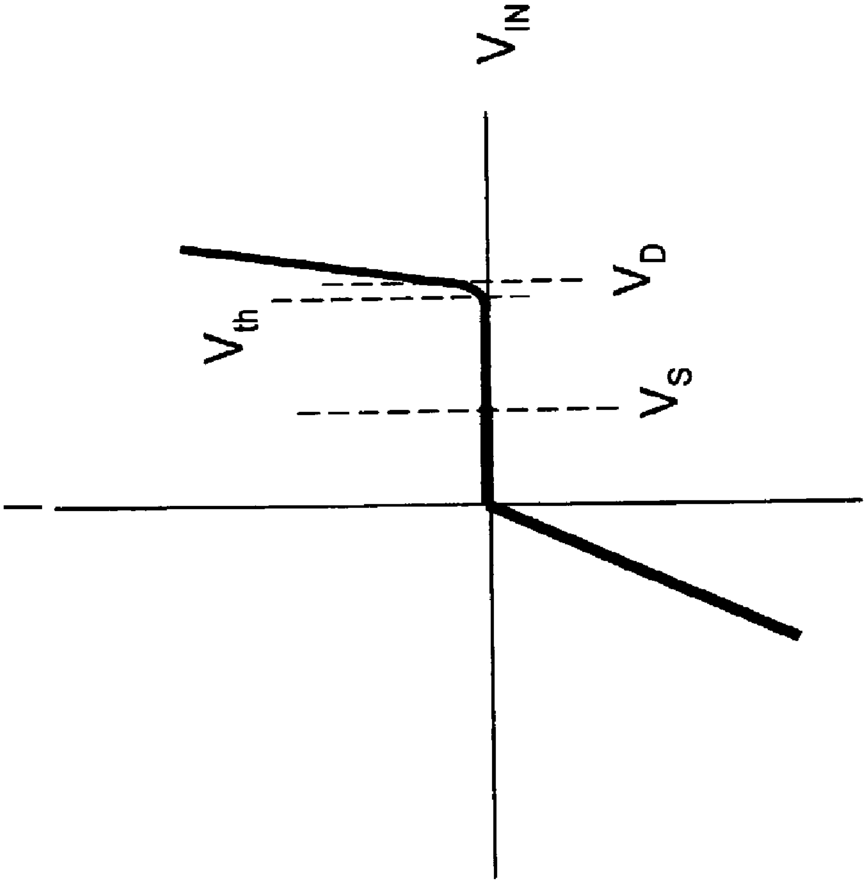

Figure 5

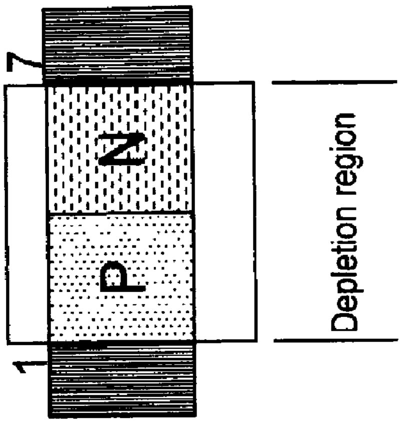
7
1
P
N
Depletion region
3

Figure 6 (b)

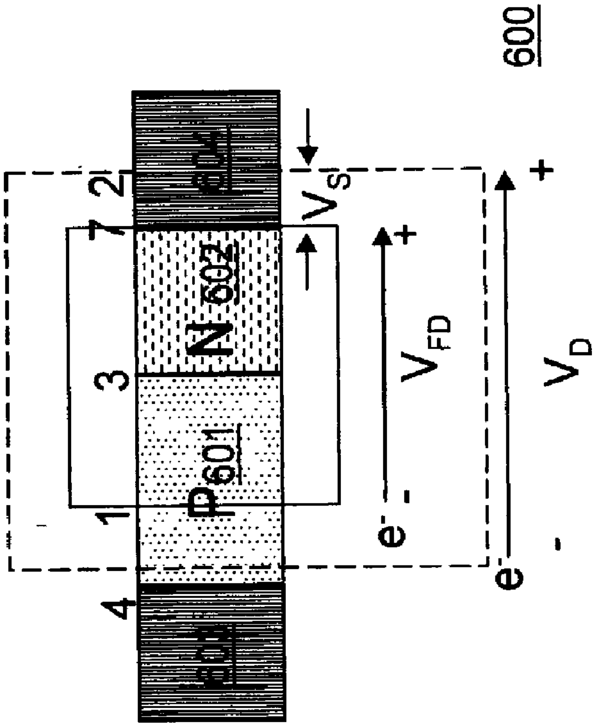
600
4
1
3
7
2
603
P
601
N
602
604
V_S
e⁻
-
+
V_FD
e⁻
-
+
V_D

Figure 6 (a)

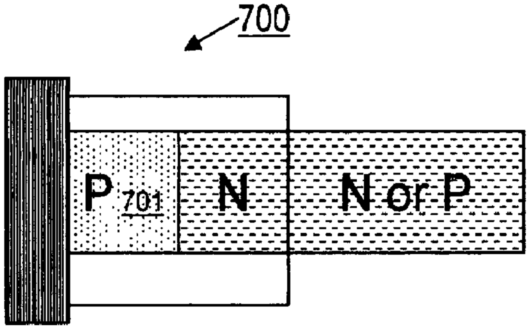
700
P 701
N
N or P

Figure 7(a)

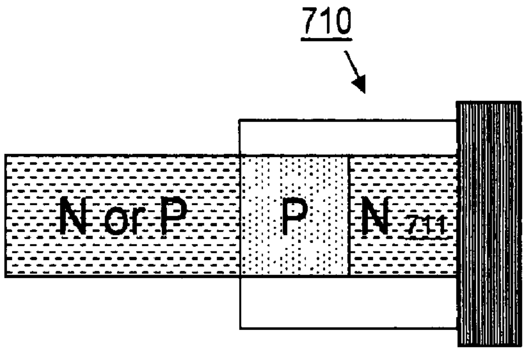
710
N or P
P
N 711

Figure 7(b)

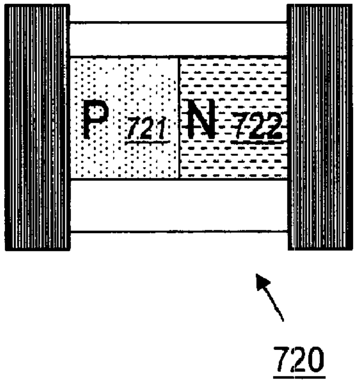
P 721
N 722
720

Figure 7(c)

☐ Depletion band

▨ Conductor

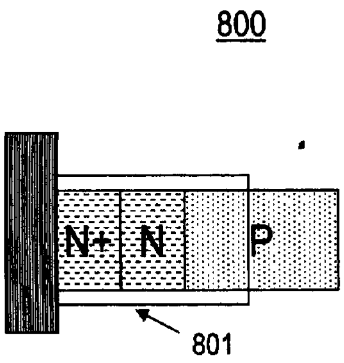

Figure 8(a)

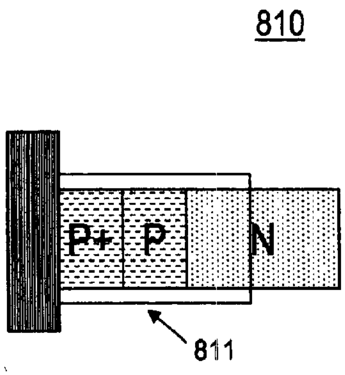

Figure 8(b)

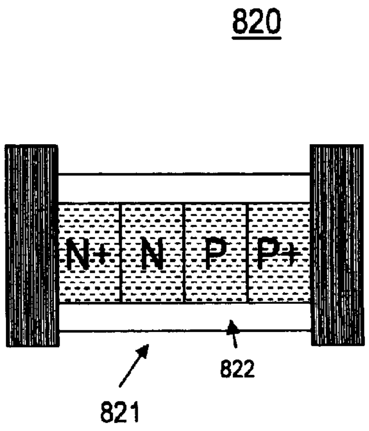
820
N+
N
P
P+
821
822

Figure 8(c)

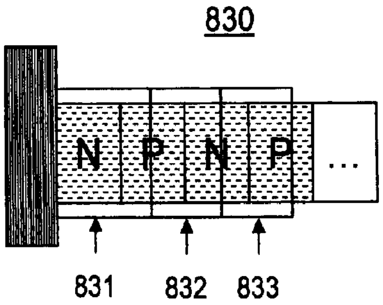
830
N
P
N
P
...
831
832
833

Figure 8(d)

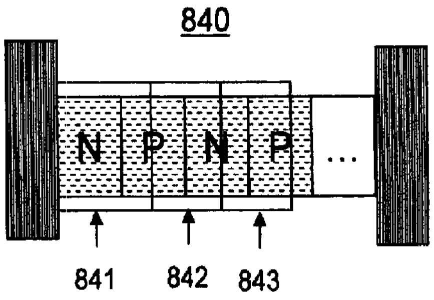
840
N
P
N
P
...
841
842
843

Figure 8(e)

☐ Depletion band

▨ Conductor

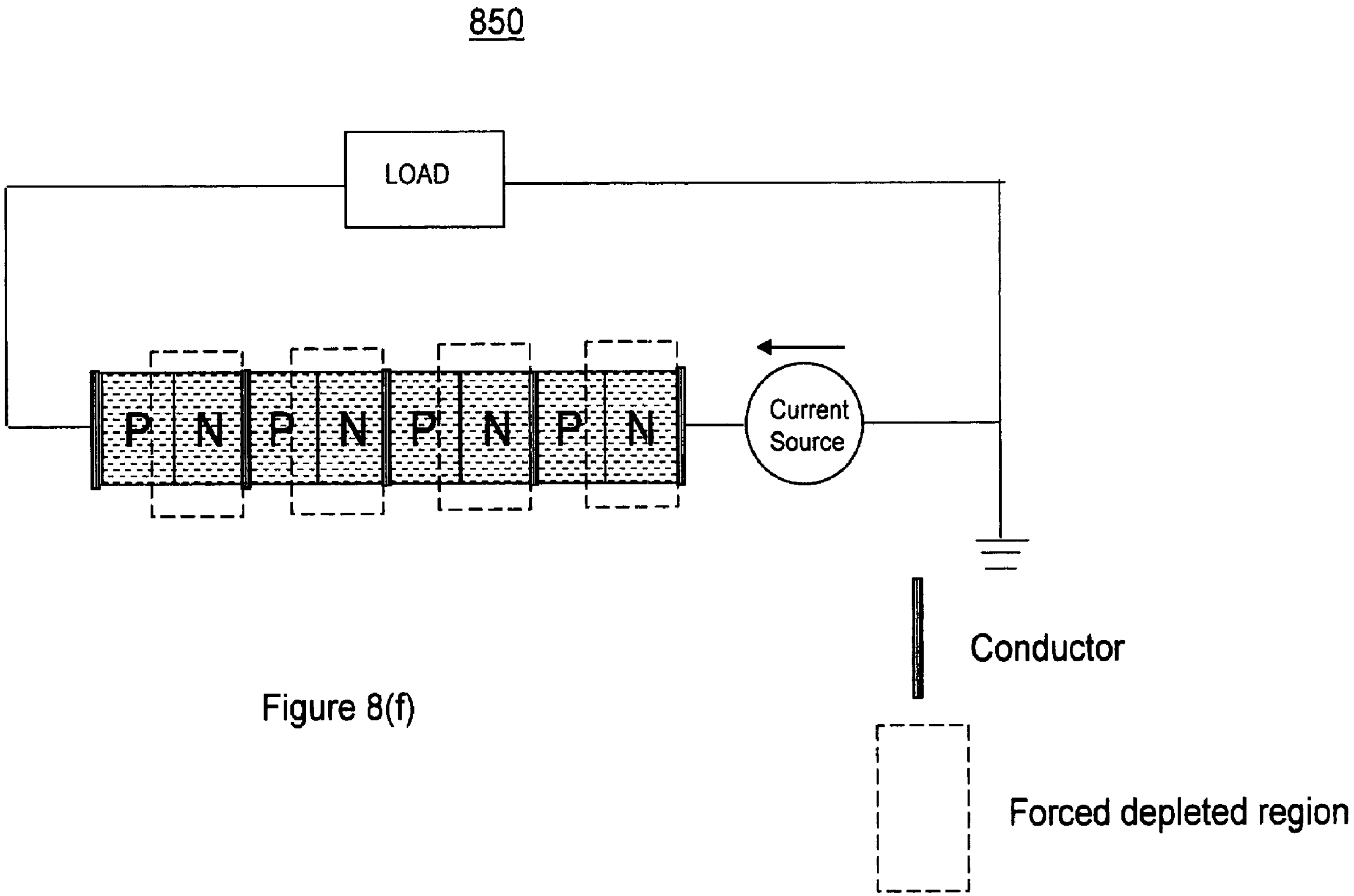
850
LOAD
P
N
P
N
P
N
P
N
Current Source
Conductor
Forced depleted region

Figure 8(f)

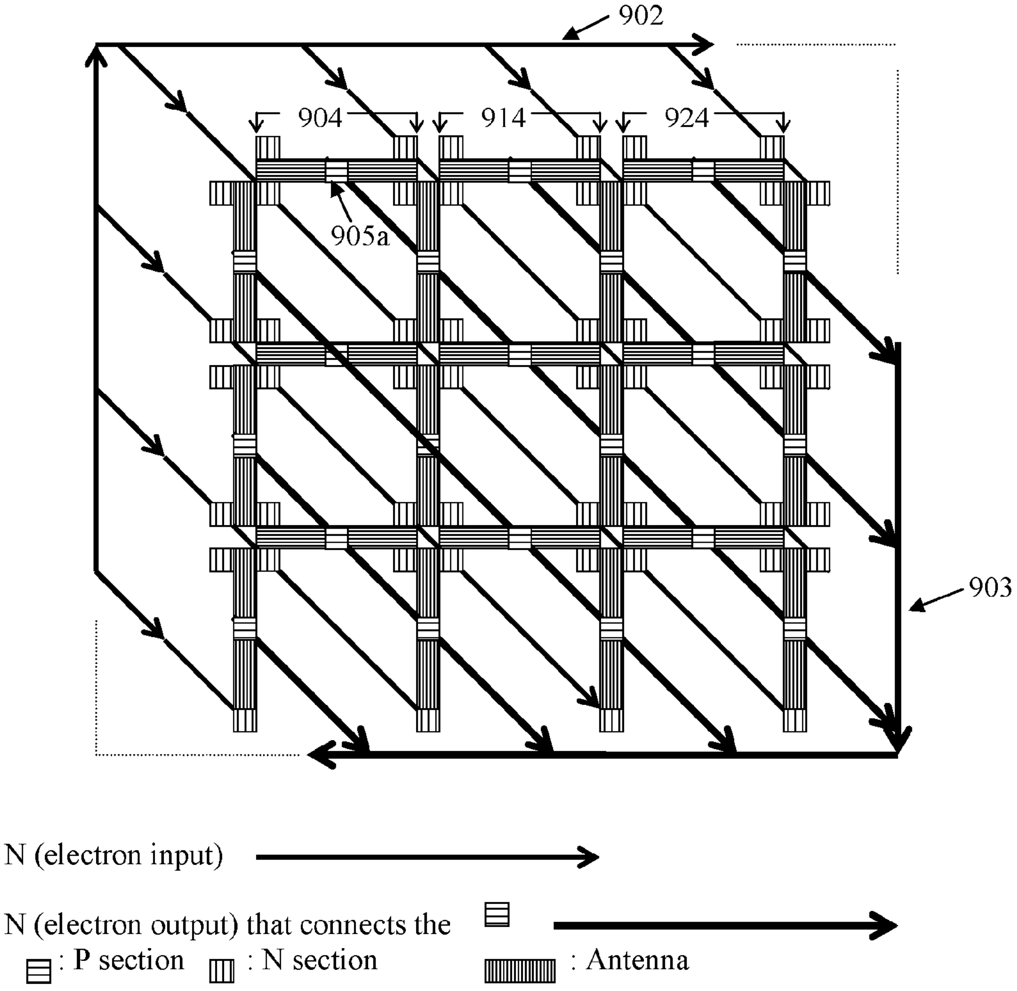
902
904
914
924
905a
903
N (electron input)
N (electron output) that connects the ▤
▤ : P section
▥ : N section
▥ : Antenna

Side view:

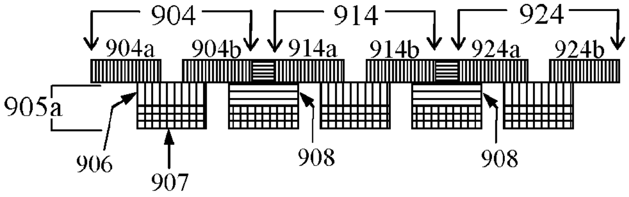
904
914
924
904a
904b
914a
914b
924a
924b
905a
906
907
908
908

anatenna: ▥

antenna going inside the page: ▤

P type block of XBD: ▤

N type channel to supply input electron to antenna : ▥

N type block of XBD and the input/output channels: ▦

Figure 9

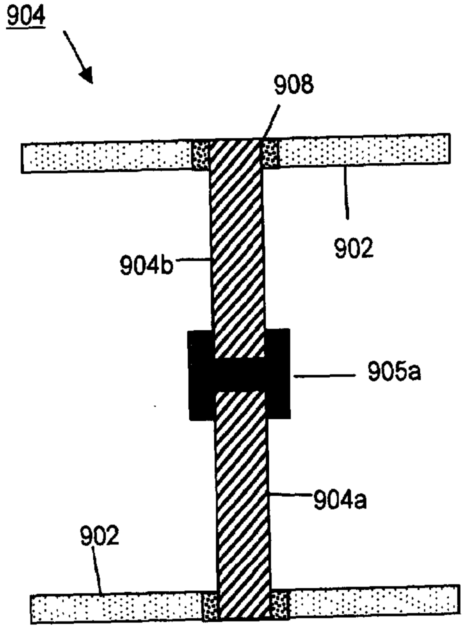
904
908
902
904b
905a
904a
902

Figure 10(a)

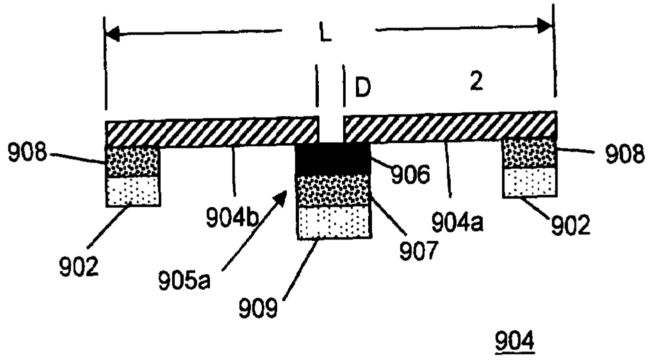
L
D
2
908
906
908
904b
904a
902
907
902
905a
909
904

Figure 10(b)

▩ N+ -type
░ N -type
■ P+ -type
▨ conductor

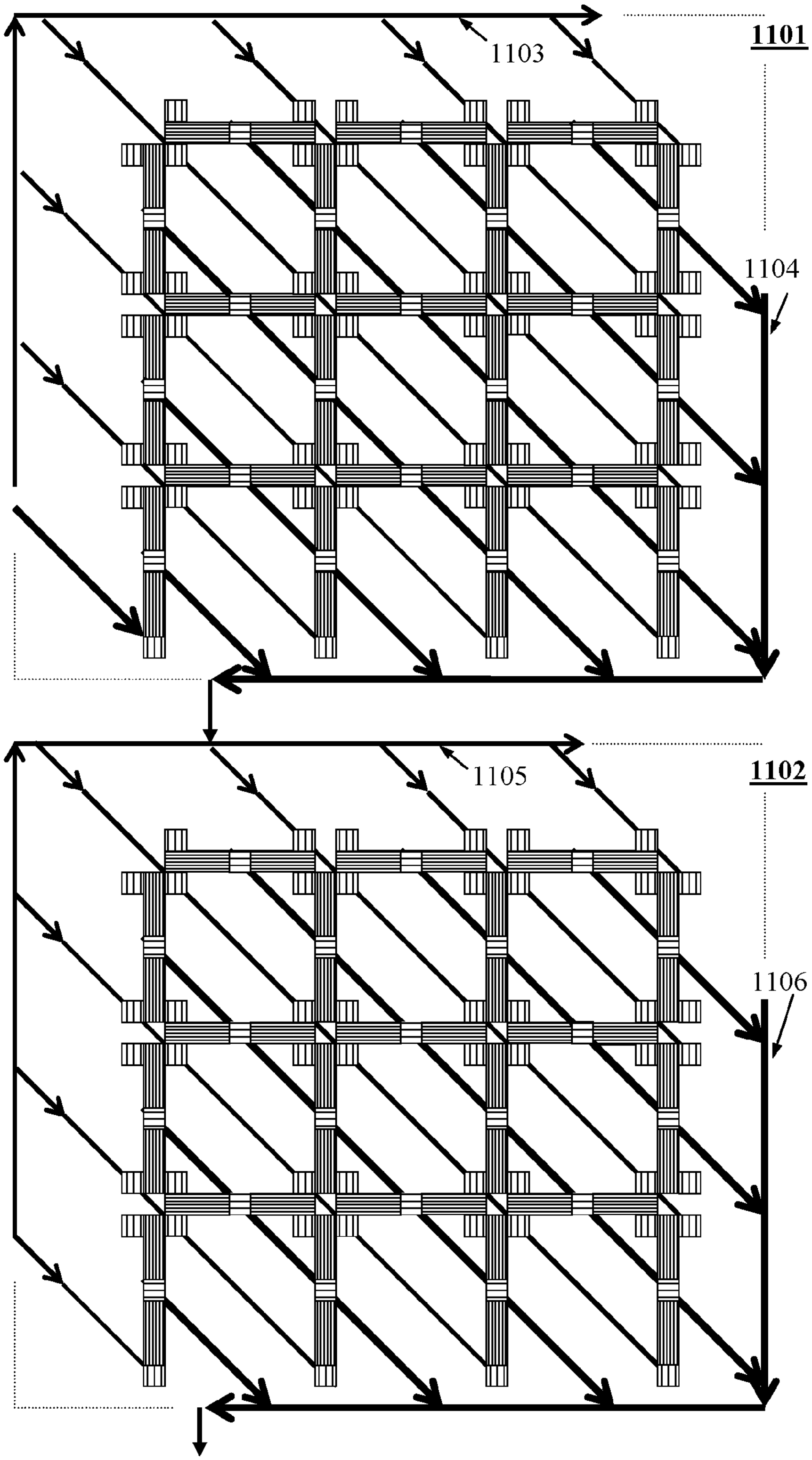
1101
1103
1104
1102
1105
1106

Figure 11

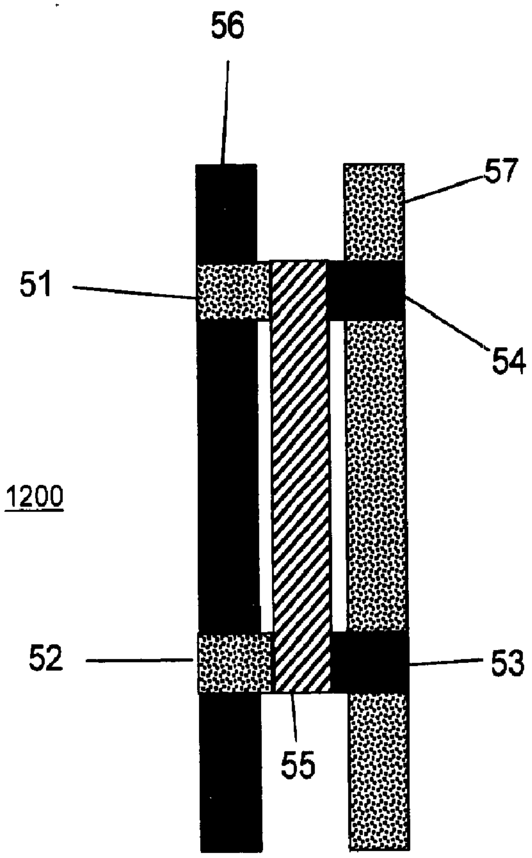
56
57
51
54
1200
52
53
55

Top View

Figure 12(a)

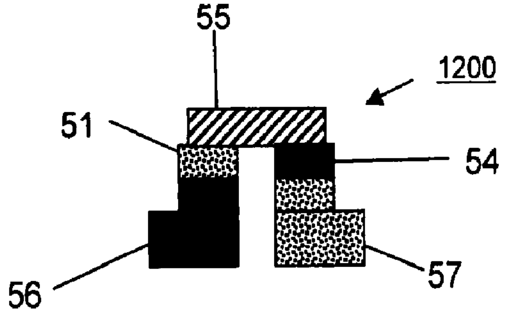
55
1200
51
54
56
57

Side View

Figure 12(b)

▨ N-type

■ P-type

▧ Conductor

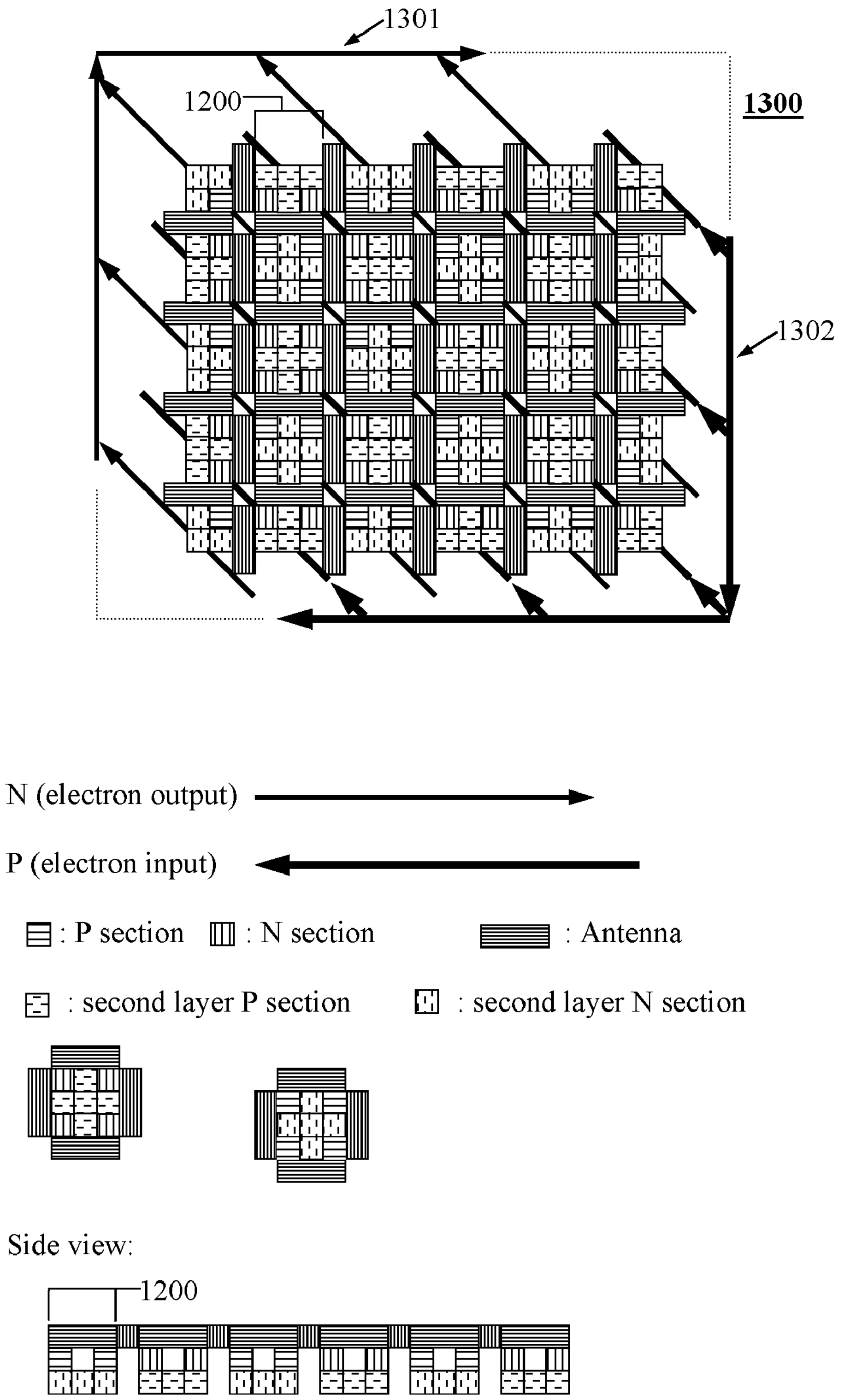
1301
1200
1300
1302
N (electron output) →
P (electron input) ←
▤ : P section
▥ : N section
▤ : Antenna
: second layer P section
: second layer N section
Side view:
1200

Figure 13

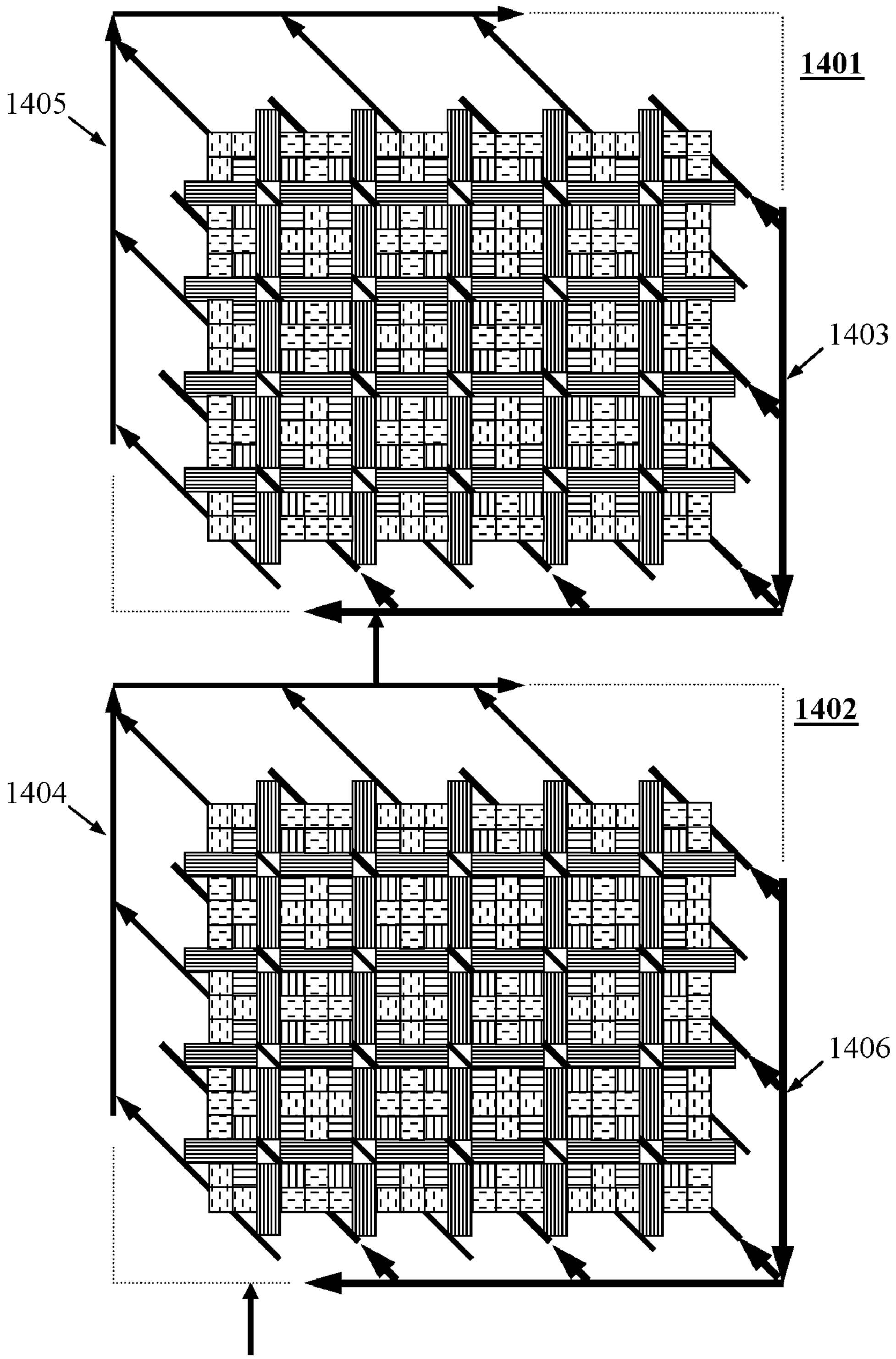
1401
1405
1403
1402
1404
1406

Figure 14

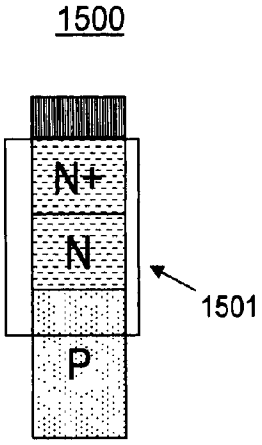
1500
N+
N
P
1501

Figure 15(a)

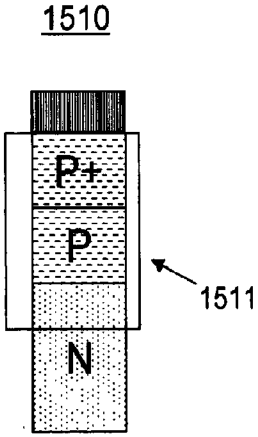
1510
P+
P
N
1511

Figure 15(b)

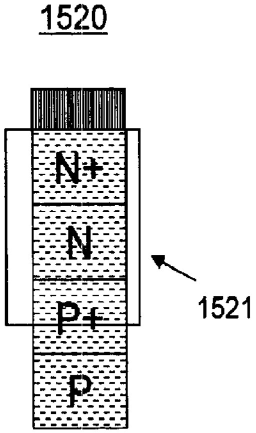
1520
N+
N
P+
P
1521

Figure 15(c)

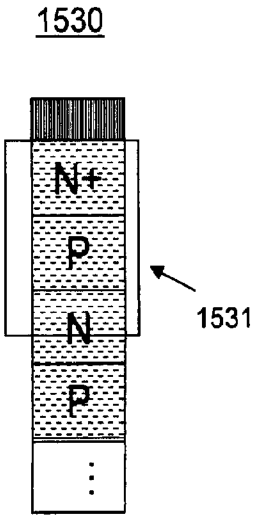
1530
N+
P
N
P
⋮
1531

Figure 15(d)

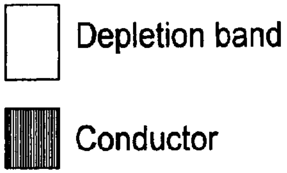
Depletion band
Conductor

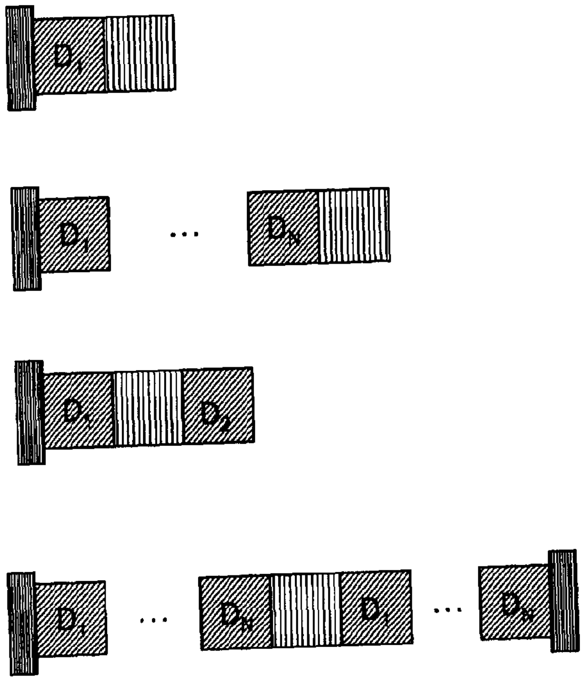
D_1
D_1
...
D_N
D_1
D_2
D_1
...
D_N
D_1
...
D_N

Figure 16(a)

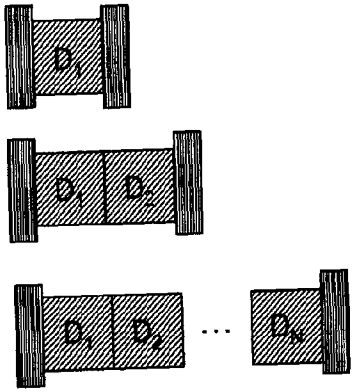
D_1
D_1
D_2
D_1
D_2
...
D_N

Figure 16(b)

▨ D_N P or N Completely Depleted

▮ Conductor

▥ P or N not complete Depleted

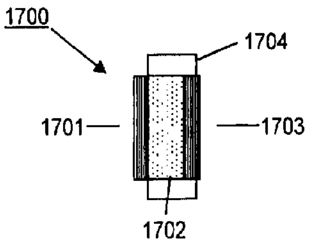
1700
1704
1701
1703
1702

Figure 17(a)

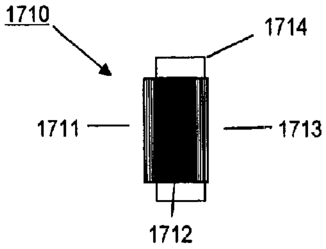
1710
1714
1711
1713
1712

Figure 17(b)

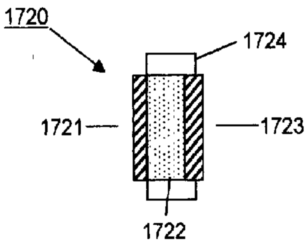
1720
1724
1721
1723
1722

Figure 17(c)

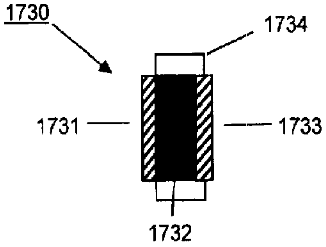
1730
1734
1731
1733
1732

Figure 17(d)

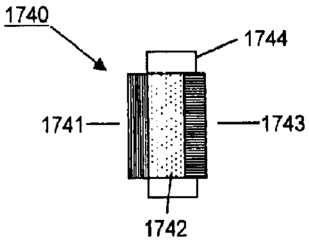
1740
1744
1741
1743
1742

Figure 17(e)

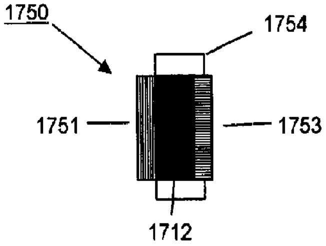
1750
1754
1751
1753
1712

Figure 17(f)

□ Depletion band

▒ N-type

■ P-type

▨ conductor for Ohmic contact

▥ Conductor type 1 for Schottky barrier

▤ Conductor type 2 for Schottky barrier

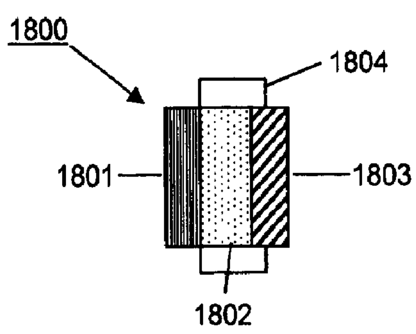
1800
1804
1801
1803
1802

Figure 18(a)

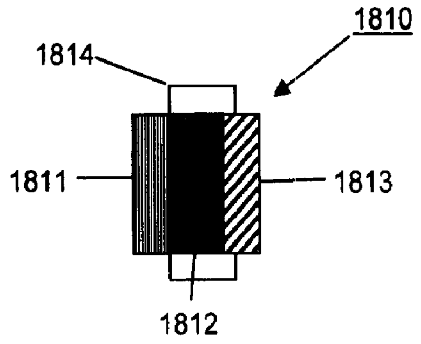
1810
1814
1811
1813
1812

Figure 18(b)

☐ Depletion band

▒ N-type

■ P-type

▨ Conductor for Ohmic contact

▥ Conductor for Schottky barrier

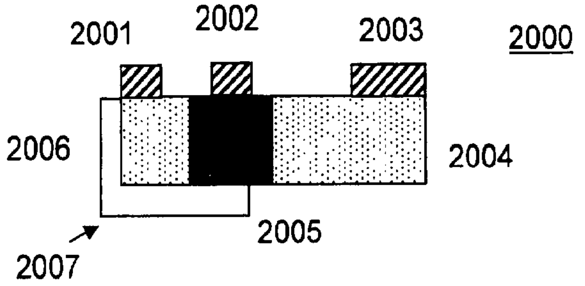
2001
2002
2003
2000
2006
2004
2005
2007

Figure 19(a)

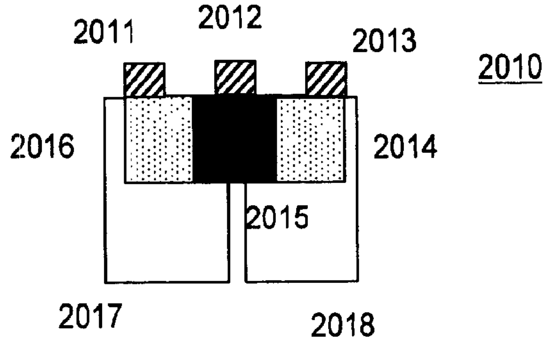
2011
2012
2013
2010
2016
2014
2015
2017
2018

Figure 19(b)

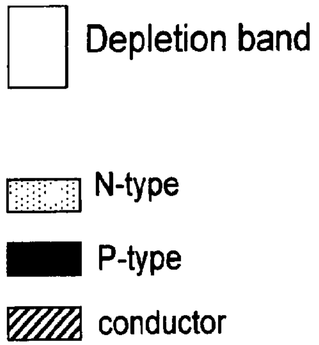
Depletion band
N-type
P-type
conductor

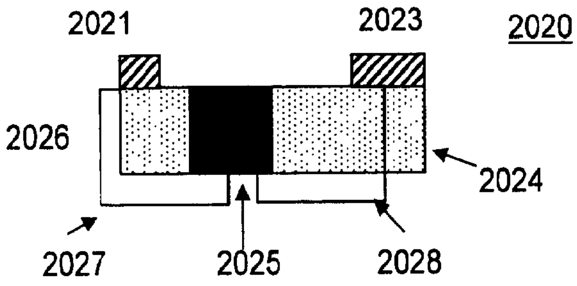
2021
2023
2020
2026
2024
2027
2025
2028

Figure 19(c)

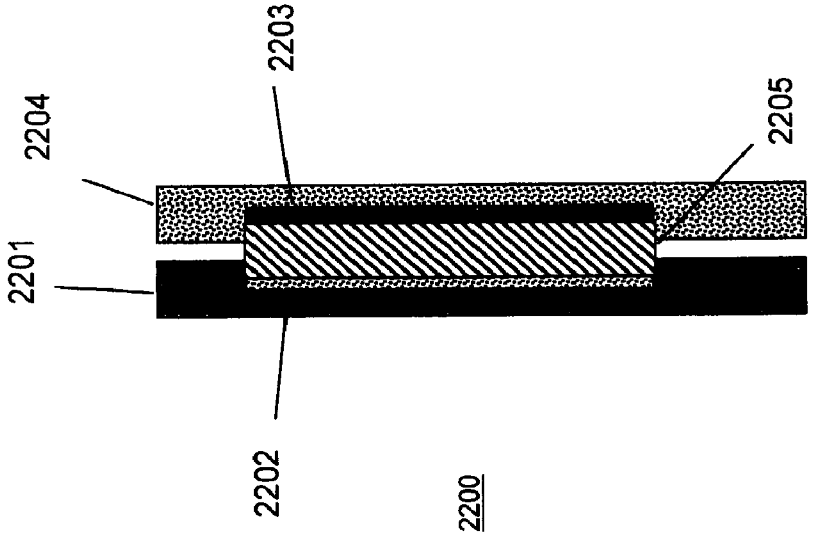
2201
2204
2203
2205
2202
2200

Top View

Figure 20(a)

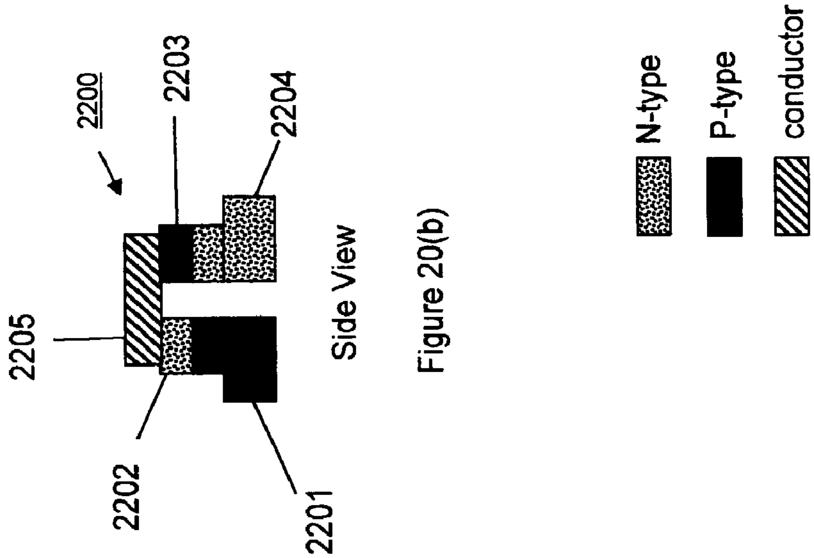
2205
2200
2203
2204
2202
2201
N-type
P-type
conductor

Side View

Figure 20(b)

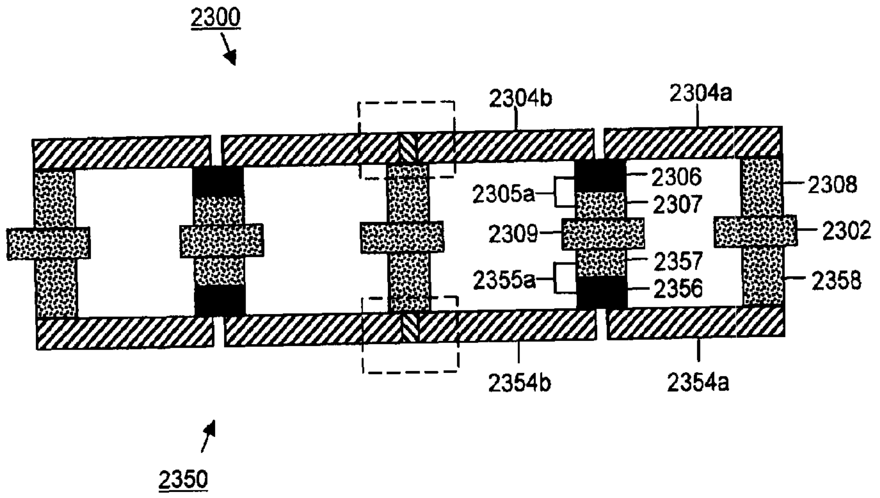
2300
2304b
2304a
2306
2305a
2307
2308
2309
2302
2357
2355a
2356
2358
2354b
2354a
2350

Figure 21

▩ N-type

■ P-type

▨ Antenna section

▧ Antenna section going into paper

⬚ Mark the end of antenna section going into paper

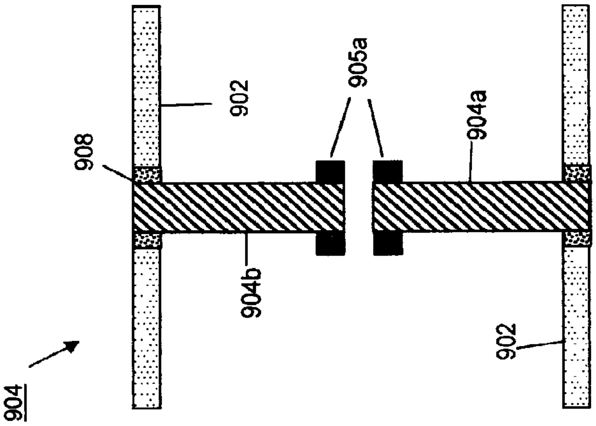
904
908
902
905a
904a
904b
902

Figure 22(a)

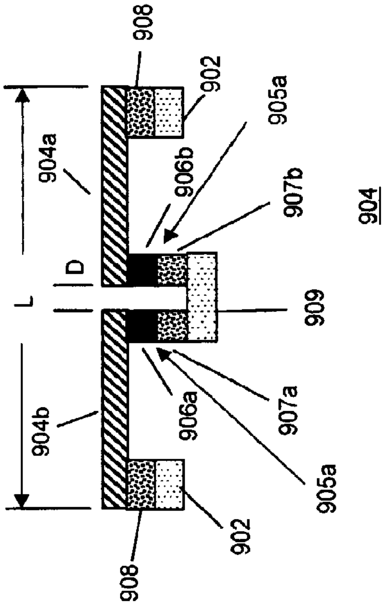
908
902
905a
904a
906b
907b
904
D
L
909
904b
906a
907a
905a
902
908

Figure 22(b)

▩ N+ -type
░ N -type
■ P+ -type
▨ conductor

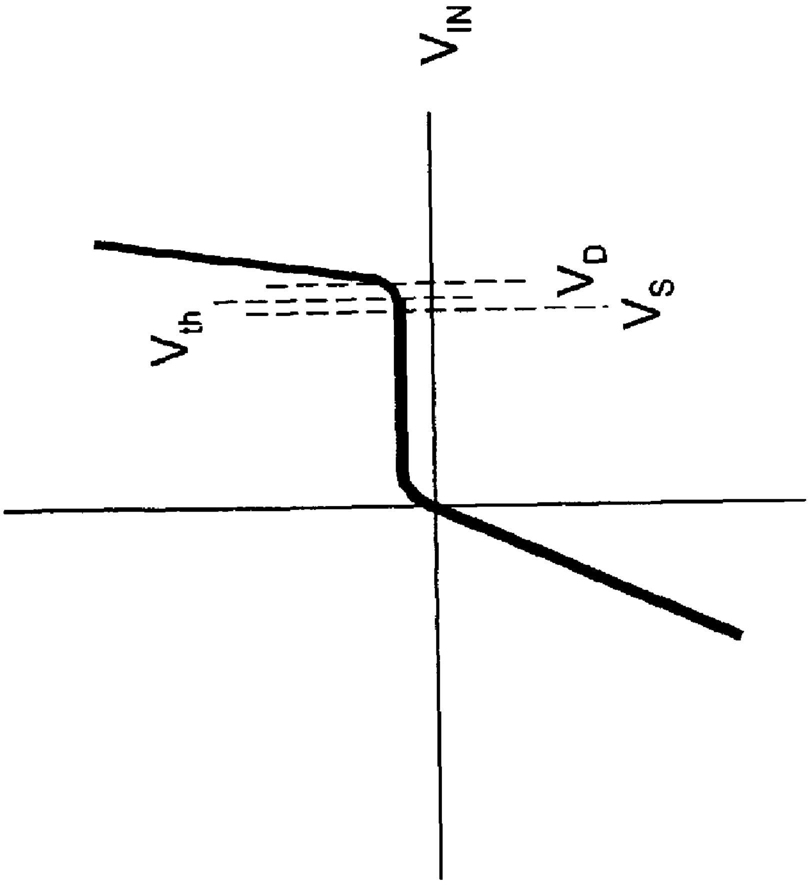

Figure 23

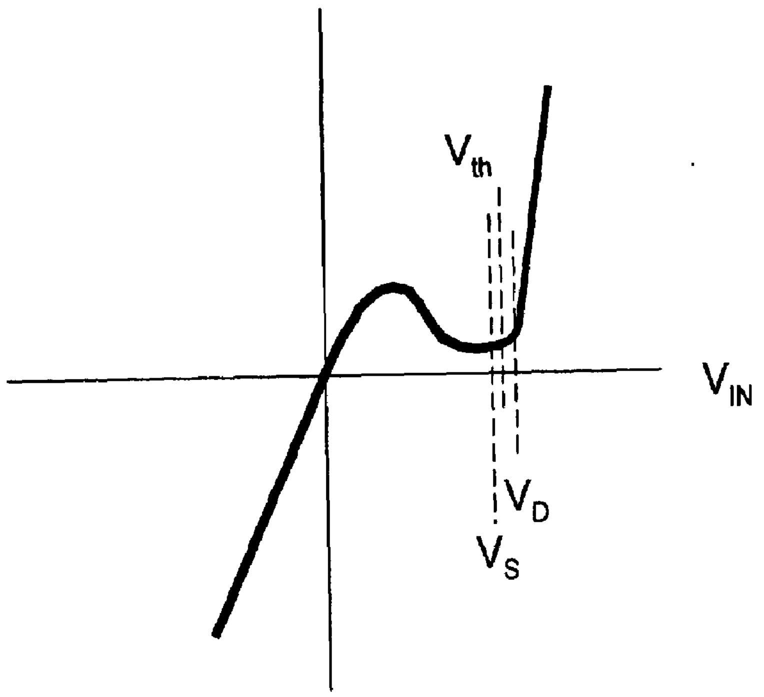

Figure 24a

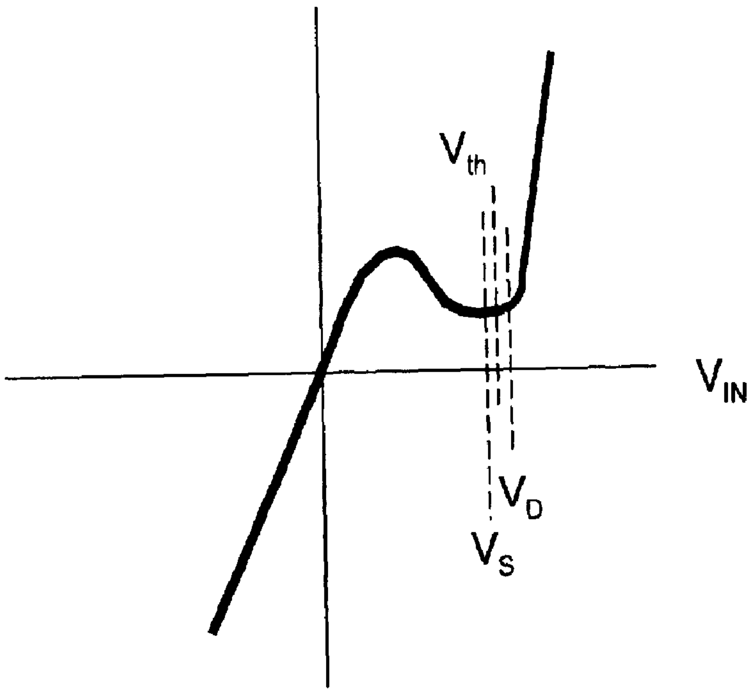

Figure 24b

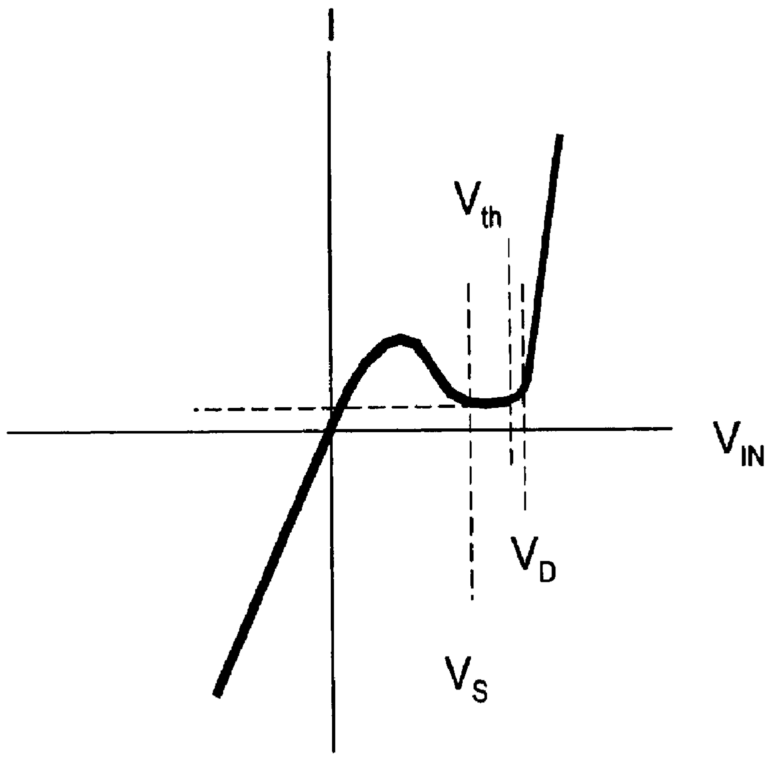

Figure 25(a)

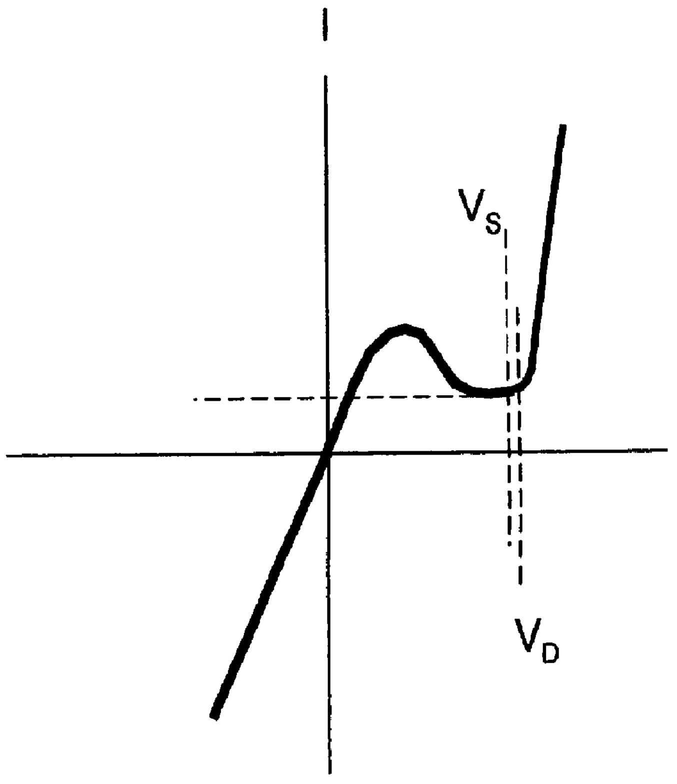

Figure 25(b)

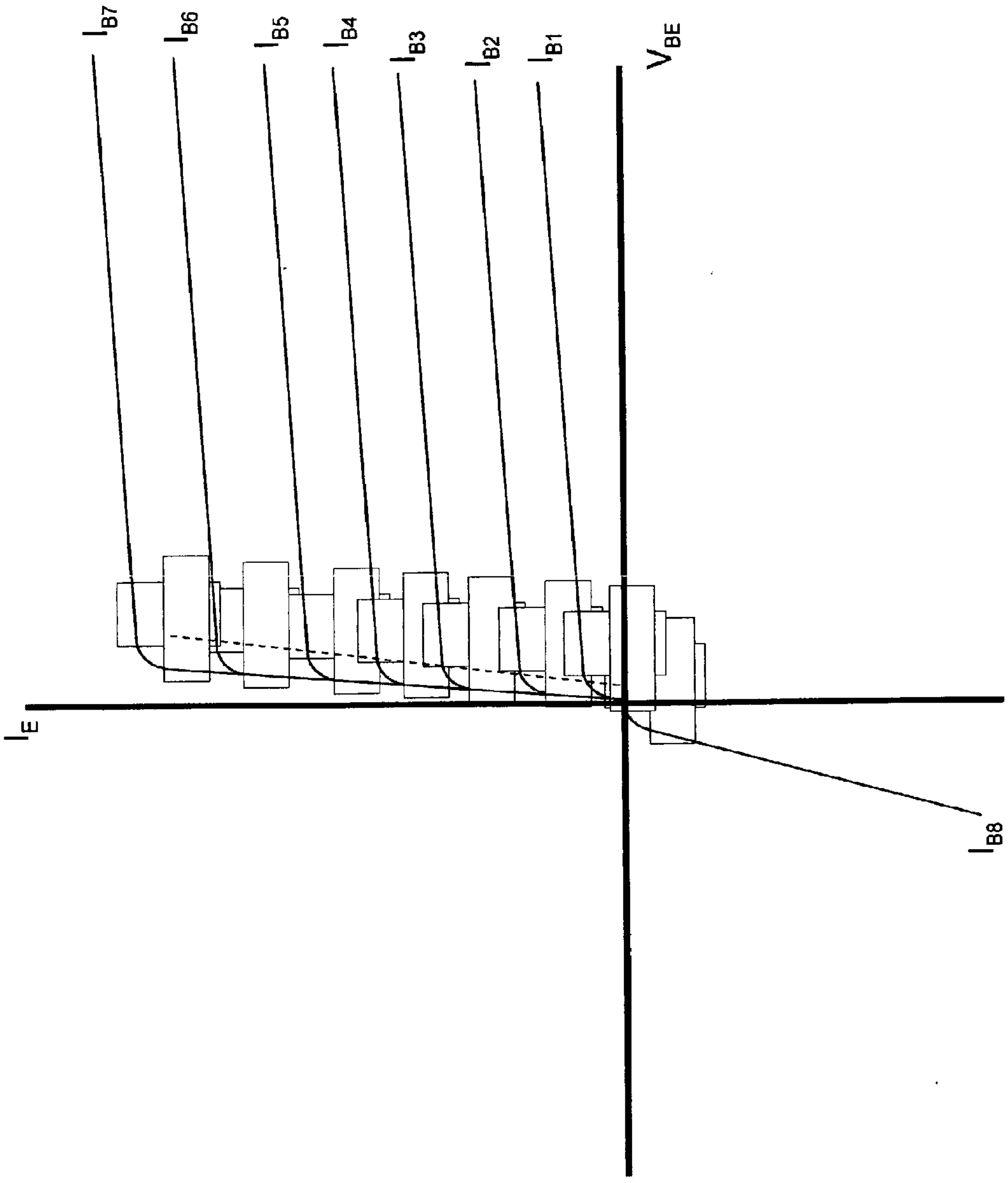

Figure 26

<table>
<tr><th>NBD device structure</th><th>Depleted Region Fermi Level Adjusted</th><th>Characteristics during increasing</th><th>Characteristics during decreasing biasing</th><th>Near zero reverse current with reverse potential</th><th>Maintains same forward threshold voltage</th></tr>
<tr><td>Sch | N non-F | Ohm</td><td>No</td><td>Depletion region smaller</td><td>Depletion region bigger</td><td>Yes</td><td>Yes</td></tr>
<tr><td>Sch | N non-F | Sch</td><td>No</td><td>- Depletion regions: Sch/N smaller, N/Sch bigger</td><td>- Depletion regions: Sch/N biggger, N/Sch smaller</td><td>No</td><td>Yes</td></tr>
<tr><td>Sch | N non-F | P | Ohm</td><td>No</td><td>- Depletion regions: Sch/N smaller, N/P bigger</td><td>- Depletion regions: Sch/N bigger, N/P smaller</td><td>No</td><td>Yes</td></tr>
<tr><td>Sch | N non-F | P | Sch</td><td>No</td><td>- Depletion regions: Sch/N smaller, N/P same, P/Sch smaller</td><td>- Depletion regions: Sch/N bigger, N/P same, P/Sch bigger</td><td>No</td><td>Yes</td></tr>
<tr><td>Ohm | N non-F | P | Ohm</td><td>No</td><td>Depletion region smaller</td><td>Depletion region bigger</td><td>Yes</td><td>Yes</td></tr>
<tr><td>Ohm | N non-F | P | Sch</td><td>No</td><td>- Depletion regions: Ohm/N same, N/P smaller, P/Sch smaller</td><td>- Depletion regions: Ohm/N same, N/P bigger, P/Sch bigger</td><td>No</td><td>Yes</td></tr>
<tr><td>Sch | N non-F | P | N</td><td>No</td><td>- Depletion regions: Sch/N smaller, N/P same, P/N smaller</td><td>- Depletion regions: Sch/N bigger, N/P same, P/N bigger</td><td>No</td><td>Yes</td></tr>
<tr><td>Ohm | N non-F | P | N</td><td>No</td><td>- Depletion regions: Ohm/N same, N/P smaller, P/N smaller</td><td>- Depletion regions: Ohm/N same, N/P bigger, P/N bigger</td><td>No</td><td>Yes</td></tr>
<tr><td>Sch | N Forced | Ohm</td><td>Yes</td><td>Depletion region same</td><td>Depletion region same</td><td>Yes</td><td>Yes</td></tr>
</table>

Figure 27.1

Sch | N Forced | Sch	Yes	- Depletion regions: Sch/N same, N/Sch same - N fermi -> Efn	- Depletion regions: Sch/N same, N/Sch same - N fermi -> Efi	No	Yes
Sch | N Forced | P | Ohm	Yes	- Depletion regions: Sch/N same, N/P same - N fermi ->Efn	- Depletion regions: Sch/N same, N/P same - Charge storage	No	Yes
Sch | N Forced | P | Sch	Yes	- Depletion regions: Sch/N same, N/P same, P/Sch smaller - N fermi -> Efn	- Depletion regions: Sch/N same, N/P same, P/Sch extra big - N fermi -> same	No	No after 1 fwd inc/dec cycle Sch/N, N/P lower
Ohm | N Forced | P | Ohm	Yes	Depletion region same - N fermi ->Efn	Depletion region same, - Charge storage	Yes	Yes
Ohm | N Forced | P | Sch	Yes	- Depletion regions: Ohm/N same, N/P same, P/Sch smaller - N fermi -> Efn	- Depletion regions: Ohm/N same, N/P same, P/Sch extra big - N fermi -> same	No	No after 1 fwd inc/dec cycle N/P Vth smaller
Sch | N Forced | P | N	Yes	- Depletion regions: Sch/N same, N/P same, P/N smaller - N fermi -> Efn	- Depletion regions: Sch/N same, N/P same, P/N extra big - N fermi -> same	No	No after 1 fwd inc/dec cycle Sch/N, N/P lower
Ohm | N Forced | P | N	Yes	- Depletion regions: Ohm/N same, N/P same, P/N smaller - N fermi -> Efn	- Depletion regions: Ohm/N same, N/P same, P/N extra big - N fermi -> same	No	No after 1 fwd inc/dec cycle N/P Vth smaller
Sch | P non-F | Ohm	No	Depletion region smaller	Depletion region bigger	Yes	Yes

Figure 27.2

Sch | P non-F | Sch	No	- Depletion regions: Sch/P smaller, P/Sch bigger	- Depletion regions: Sch/P bigger, P/Sch smaller	No	Yes
Sch | P non-F | N | Ohm	No	- Depletion regions: Sch/P smaller, P/N bigger	- Depletion regions: Sch/P bigger, P/N smaller	No	Yes
Sch | P non-F | N | Sch	No	- Depletion regions: Sch/P smaller, P/N same, N/Sch smaller	- Depletion regions: Sch/P bigger, P/N same, N/Sch bigger	No	Yes
Ohm | P non-F | N | Ohm	No	Depletion region smaller	Depletion region bigger	Yes	No
Ohm | P non-F | N | Sch	No	- Depletion regions: Ohm/P same, P/N smaller, N/Sch smaller	- Depletion regions: Ohm/P same, P/N bigger, N/Sch bigger	No	Yes
Sch | P non-F | N | P	No	- Depletion regions: Sch/P smaller, P/N same, N/P smaller	- Depletion regions: Sch/P bigger, P/N same, N/P bigger	No	Yes
Ohm | P non-F | N | P	No	- Depletion regions: Ohm/P same, P/N smaller, N/P smaller	- Depletion regions: Ohm/P same, P/N bigger, N/P bigger	No	Yes
Sch | P Forced | Ohm	Yes	Depletion region same	Depletion region same	Yes	Yes
Sch | P Forced | Sch	Yes	- Depletion regions: Sch/P same, P/Sch same - P fermi -> Efp	- Depletion regions: Sch/P same, P/Sch same - P fermi -> Efi	No	Yes
Sch | P Forced | N | Ohm	Yes	- Depletion regions: Sch/P same, P/N same - P fermi ->Efp	- Depletion regions: Sch/P same, P/N same - Charge storage	No	Yes

Figure 27.3

Sch | P Forced | N | Sch	Yes	- Depletion regions: Sch/P same, P/N same, N/Sch smaller - P fermi -> Efp	- Depletion regions: Sch/P same, P/N same, N/Sch extra big - P fermi -> same	No	No after 1 fwd inc/dec cycle Sch/P, P/N lower
Ohm | P Forced | N | Ohm	Yes	Depletion region same, P fermi ->Efp	Depletion region same, Charge storage	Yes	Yes
Ohm | P Forced | N | Sch	Yes	- Depletion regions: Ohm/P same, P/N same, N/Sch smaller - P fermi -> Efp	- Depletion regions: Ohm/P same, P/N same, N/Sch extra big - P fermi -> same	No	No after 1 fwd inc/dec cycle P/N Vth smaller
Sch | P Forced | N | P	Yes	- Depletion regions: Sch/P same, P/N same, N/P smaller - P fermi -> Efp	- Depletion regions: Sch/P same, P/N same, N/P extra big - P fermi -> same	No	No after 1 fwd inc/dec cycle Sch/P, P/N lower
Ohm | P Forced | N | P	Yes	- Depletion regions: Ohm/P same, P/N same, N/P smaller - P fermi -> Efp	- Depletion regions: Ohm/P same, P/N same, N/P extra big - P fermi -> same	No	No after 1 fwd inc/dec cycle P/N threshold smaller, N/P larger

Figure 27.4

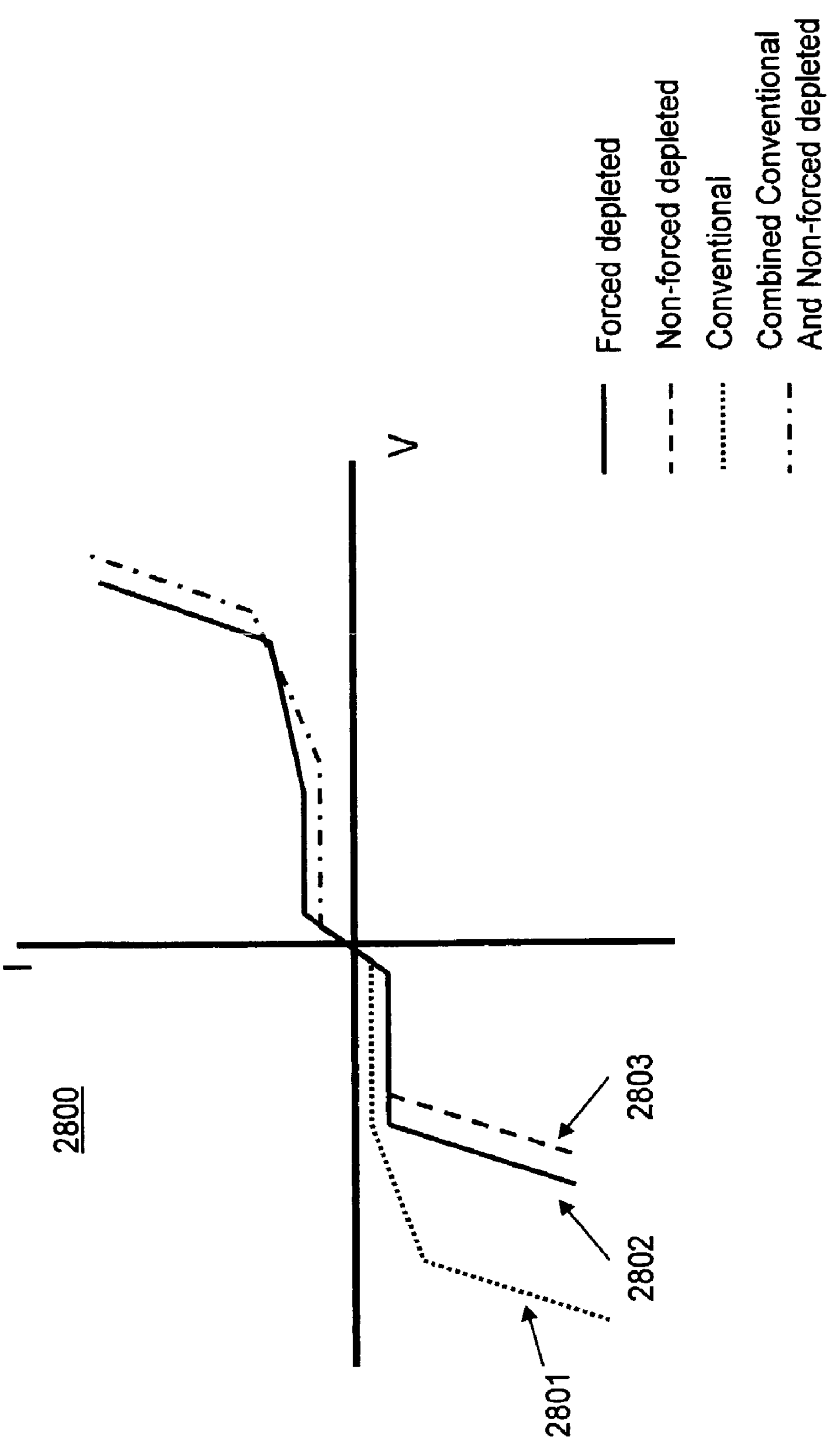
2800
I
V
Forced depleted
Non-forced depleted
Conventional
Combined Conventional And Non-forced depleted
2801
2802
2803

Figure 28

EM RECTIFYING ANTENNA SUITABLE FOR USE IN CONJUNCTION WITH A NATURAL BREAKDOWN DEVICE

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] The present application is a continuation-in-part application of, and claims priority to, co-pending U.S. patent application ("Co-pending Patent Application"), Ser. No. 10/963,357, entitled "EM Rectifying Antenna Suitable for use in Conjunction with a Natural Breakdown Device," filed on Oct. 12, 2004, bearing Attorney Docket No. M-15617 US. The Co-pending patent application is hereby incorporated by reference in its entirety.

BACKGROUND OF THE INVENTION

[0002] 1. Field of the Invention

[0003] The present invention relates to an energy conversion device using rectifying antennae ("rectannae"). In particular, the present invention relates to an energy conversion device which converts electromagnetic wave energy to electrical energy using a semiconductor device that rectifies a small high frequency signal.

[0004] 2. Discussion of the Related Art

[0005] FIG. **1** shows the current versus voltage characteristics of a conventional pn junction diode. FIG. **2** is a schematic representation of conventional abrupt pn junction diode **100**. As shown in FIG. **2**, conventional pn junction diode **100** includes p-region **101** and n-region **102**. P-region **101** may be doped, for example, using a p-type dopant (i.e., electron acceptor, such as boron) and n-region **102** may be doped using an n-type dopant (i.e., an electron donor, such as phosphorus). Near the abrupt junction between p-region **101** and n-region **102**, equilibrium due to the difference in electrochemical potentials of the two regions and the diffusion of charge carriers (e.g., electrons and "holes") between the two regions deplete the charge carriers to form "depletion regions" **103** and **104** in p-region **101** and n-region **102**, respectively. Under a so-called "abrupt junction approximation", the widths x_p of depletion region **103** and x_n for depletion region **104**, with an externally imposed voltage V across the pn junction, are given, respectively by:

$$x_n = \sqrt{\frac{2\varepsilon_s N_A(\phi_i - V)}{qN_D(N_A + N_D)}}$$

$$x_p = \sqrt{\frac{2\varepsilon_s N_D(\phi_i - V)}{qN_A(N_A + N_D)}}$$

where ϵ_s is the electrical permittivity of silicon, q is the charge of an electron, ϕ_i is the "built-in" potential of the pn junction, N_A and N_B are the doping concentrations of p-region **101** and n-region **102**, respectively.

[0006] As shown in FIG. **1**, the horizontal axis shows the voltage V across the pn junction, and the vertical axis shows the diode current I_D across the pn junction. As shown in FIG. **1**, when voltage V across the pn junction is greater than zero volts and greater than voltage V_{th} (the "threshold voltage"), the pn junction is strongly "forward biased" and the diode current I_D grows exponentially with the voltage V. When the voltage V across the pn junction is less than 0 volts, but not less than the voltage V_{br} (the "breakdown voltage"), the pn junction is "reverse biased" and the diode current I_D is very small. Under reversed bias, as the voltage grows in magnitude, the carriers generated increases in energy, leading to phenomena such as tunneling and impact ionization at voltage V_{br}. At voltage V_{br}, the diode current I_D becomes very large and the diode has "broken down." At breakdown, the magnitude of the average electrical field (in volts per centimeter) across the pn junction is given by the empirical expression:

$$|E_{br}| = \frac{4.0 \times 10^5}{1 - \frac{1}{3}\log\frac{N_D}{10^{16}}}$$

where N_D is the lesser of N_A and N_B.

SUMMARY

[0007] The present invention provides a rectifying antenna ("rectenna") capable of converting power from high frequency electromagnetic (EM) waves in free space to electrical energy. According to one embodiment of the present invention, an antenna generates currents from two sources—from a potential difference generated by an incident EM wave in the antenna, and from an n-type semiconductor or another electron source attached to a maximum voltage point of an antenna. The combined current from both sources increases the power output of the antenna, thereby increasing the detection sensitivity of the antenna of a low power signal.

[0008] According to one embodiment of the present invention, full wave rectification is achieved using a novel diode connected to an antenna. In that embodiment, the diode is conductive at a reverse bias voltage, stores charge at forward bias, and rectifies the antenna signal generated by the targeted electromagnetic (EM) radiation within a selected spectrum received by antenna. The diode and the antenna form a rectenna. In one embodiment of the present invention, half wave rectification is achieved using the diode.

[0009] The rectenna of the present invention may be used as a building block to create large rectenna arrays. Thus, microwave, sub-millimeter wave, THz wave, IR wave and visible wave applications may be found in a rectenna of this invention. A rectenna of the present invention is scalable, so that future improvements are possible with improvements of semiconductor technology, material science, device physics and antenna design. There is also no upper limit to the frequencies that can be received using such a rectenna. The present invention is applicable to EM-to-DC power generation and supply, heat shield, humidity, food production, dermatology (e.g., examination of burned skin structure without biopsy) and tomography (and other medical imaging sensing). The present invention allows T-rays replacement of Roentgen rays (X-rays) in imaging of objects, such as those used to inspect pallets in certain airport safety procedures or in drug detection procedures. The present invention is also applicable to chemical analysis using flames and gases (e.g., toxicity analysis, Fourier spectrum), quality control procedures (e.g., detection of holes in plastic parts, such as those used in automobile applications), and radar applications (e.g., measurement of reflection)

[0010] According to another aspect of the present invention, a "natural breakdown device" ("NBD") allows current to flow even with a zero bias voltage. NBD is a natural breakdown device of the diode type, according to one embodiment of natural breakdown invention. An NBD may be used as a rectifier on the rectenna of the present invention to rectify small amplitude high frequency electromagnetic signals.

[0011] The present invention is better understood upon consideration of the detailed description below and the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] FIG. **1** shows the current (I) versus voltage (V) characteristics of a conventional pn junction diode.

[0013] FIG. **2** is a schematic representation of conventional pn junction diode **100**.

[0014] FIG. **3** is a schematic representation of a natural breakdown device (NBD) **300**, according to one embodiment of the present invention.

[0015] FIG. **4** shows the relationship between the built-in voltage V_{FD} and bias voltage V_{IN} across NBD **300** during operation.

[0016] FIG. **5** shows the current-voltage (IV) characteristics of NBD **300** when the value of voltage V_s is not close to the threshold voltage V_{th} of NBD **300**.

[0017] FIG. **6**(*a*) is a schematic representation of a natural breakdown device (NBD) **600**, having an N-type region that is fully depleted at zero bias voltage, according to one embodiment of the present invention.

[0018] FIG. **6**(*b*) shows an NBD **600** that has both p-type and n-type regions fully depleted at zero bias voltage, according to one embodiment of the present invention.

[0019] FIGS. **7**(*a*) to **7**(*c*) show NBD **700**, NBD **710** and NBD **720**, each including a depleted region (e.g., depletion regions **701**, **711**, **721** or **722**), at zero applied voltage bias.

[0020] FIGS. **8**(*a*) to **8**(*f*) show NBD **800**, NBD **810** and NBD **820** each including a forced depletion region (e.g., depletion region **801**, **811**, **821** and **822**) having two or more doping concentrations at zero applied voltage bias. In particular, FIG. **8**(*f*) shows multiple NBD **600**s connected in series at zero bias.

[0021] FIG. **9** shows rectenna array **900** including numerous antennae (e.g., antenna **904**, **914** and **924**) constructed in accordance with one embodiment of the present invention.

[0022] FIGS. **10**(*a*) and **10**(*b*) show top and side views of one embodiment of the invention, illustrating gap dipole antenna **904** in rectenna array **900** of FIG. **9**, in accordance with one embodiment of the present invention.

[0023] FIG. **11** shows gap rectenna arrays **1101** and **1102** connected in series for DC power collection, showing electron output channel **1104** of rectenna array **1101** being connected to electron input channel **1105** of rectenna array **1102**.

[0024] FIGS. **12**(*a*) and **12**(*b*) show, respectively, the top and side views of gapless dipole antenna **1200**, in accordance with one embodiment of the present invention.

[0025] FIG. **13** shows rectenna array **1300** including numerous gapless antennae.

[0026] FIG. **14** shows rectenna arrays **1401** and **1402** connected in series for DC power collection, showing electron output channel **1404** of rectenna array **1402** being connected to electron input channel **1403** of rectenna array **1401**.

[0027] FIG. **15**(*a*) to **15**(*d*) shows schematic representations of various NBDs at zero applied voltage bias, according to one embodiment of the present invention.

[0028] FIG. **16**(*a*) to **16**(*b*) shows, schematic representations of various configurations of NBDs at zero applied bias voltages, according to one embodiment of the present invention.

[0029] FIGS. **17**(*a*) to **17**(*f*) show schematic representations of various NBDs at zero applied voltages, according to one embodiment of the present invention.

[0030] FIG. **18**(*a*) is a schematic representation of NBD **1800** at zero applied bias voltage, according to one embodiment of the present invention; NBD **1800** represents a natural breakdown Schottky diode under a forced depletion condition.

[0031] FIG. **18**(*b*) is a schematic representation of NBD **1810** at zero applied bias voltage, according to one embodiment of the present invention; NBD **1810** represents a natural breakdown Schottky diode under a forced depletion condition.

[0032] FIG. **19**(*a*) is a schematic representation of NBD **2000** at zero applied bias voltage, according to one embodiment of the present invention; NBD **2000** represents an NPN bipolar transistor with one region under a forced depletion condition.

[0033] FIG. **19**(*b*) is a schematic representation of NBD **2010** at zero applied voltage bias, according to one embodiment of the present invention; NBD **2010** represents a natural breakdown NPN bipolar transistor having two regions under a forced depletion conditions.

[0034] FIG. **19**(*c*) is a schematic representation of NBD **2020** at zero applied voltage bias, according to one embodiment of the present invention; NBD **2020** represents a natural breakdown NPN device with one region fully depleted.

[0035] FIGS. **20**(*a*) and **20**(*b*) show, respectively, the top and side views of gapless dipole rectenna element **2200**, in accordance with one embodiment of the present invention.

[0036] FIG. **21** shows, gap rectenna arrays **2300** and **2350** connected in parallel, according to one embodiment of present invention.

[0037] FIGS. **22**(*a*) and **22**(*b*) show rectifier structure **905***a* provided for signal rectification, according to one embodiment of the present invention.

[0038] FIG. **23** shows the current-voltage (IV) characteristics of NBD **300** when contact resistance is very low.

[0039] FIG. **24** shows the current-voltage (IV) characteristics of NBD **300** when voltage V_s is close to threshold voltage V_{th}.

[0040] FIG. **25**(*a*) shows the current-voltage (IV) characteristics of NBD **300** utilizing degenerate material when the value of V_s is not close to threshold voltage V_{th}.

[0041] FIG. **25**(*b*) shows the current-voltage (IV) characteristics of NBD **300** utilizing degenerate material when the voltage V_s is close to threshold voltage V_{th}.

[0042] FIG. **26** shows the current-voltage (IV) characteristics of NBD **2000**.

[0043] FIGS. **27.1** to **27.2** are tables of various NBD structures with their corresponding characteristics.

[0044] FIG. **28** shows the current-voltage (IV) characteristics of NBD **2020**.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0045] While the following description discloses mainly dipole rectangular antenna ("rectenna") elements in square array configurations (i.e., two-dimensional antenna arrays each having the same number of antenna elements along each dimension), the present invention is applicable to rectennae of other configurations, such as a two-dimensional configuration having different number of antenna elements along the two dimensions. The present invention provides an array of dipole rectennae elements configured to capture and harvest electromagnetic energy for power generation.

[0046] One embodiment of the present invention is shown in FIG. **9**. FIG. **9** shows rectenna array **900** including numerous antennae (e.g., antennae **904**, **914** and **924**). FIGS. **10**(*a*) and **10**(*b*) are top and side of rectenna array **900** of FIG. **9**, showing dipole antenna **904**, which includes antenna sections **904***a* and **904***b* contacting rectifier structure **905***a*. Rectifier structure **905***a* rectifies the signal received at antenna **904***a* and **904***b*. Rectifier structure **905***a* may be provided, according to this embodiment of the present invention, by a "natural breakdown device" (NBD), which is further discussed below in this detailed description and referred to as a "Natural Breakdown Device" (NBD) or), a natural breakdown PN junction diode.

[0047] As shown in FIGS. **10**(*a*) and **10**(*b*), rectifier structure **905***a* includes an NBD which includes p-region **906** and n-region **907**. P-region **906** has a width that is fully depleted at a zero bias voltage condition. In this description, when a depletion region has a width that is smaller than the width of the depletion at zero applied voltage bias at an abrupt junction for the same doping concentrations, that depletion region is said to be "forced". The forced depleted p-region **906** of rectifier structure **905***a* connects directly to the antenna. The n-region **907** of rectifier structure **905***a* connects directly to an output channel **909**. Rectifier structure **905***a* may also be provided by two NBDs, according to one embodiment of the present invention, as shown in FIGS. **22**(*a*) and **22**(*b*). Using two NBDs, each connected to a different one of the antenna sections **904***a* and **904***b*, shorting of the antenna sections by the NBD may be avoided.

[0048] Some rectifier qualities of rectifier structure, such as **905***a*, which allow dipole antenna **904** and other antennae presented in the present invention operate efficiently include: (1) being conducting near zero reverse bias voltage or near zero forward bias voltage. (2) having a current that is significantly greater than a leakage current of a conventional pn-junction diode in reverse bias, (3) rectifying input signal by producing a current in a predetermined direction, and (4) when the rectifier structure is connected with an voltage source, the input voltage is raised.

[0049] As shown in FIG. **10**(*b*), each dipole antenna **904** has a length L and is made up from two antenna sections (e.g., sections **904***a* and **904***b*) separated in the center of the antenna by a gap of length D. Length L is ideally one-half the wavelength of the electromagnetic (EM) wave that antenna is designed to receive, D is typically significantly smaller than the smallest wavelength of the electromagnetic (EM) wave the antenna is expected to receive. P-region **906** of rectifier structure **905***a* (NBD **300**) is connected, on one side, next to the gap of antenna **904**, to both sections **904***a* and **904***b* and, on the other side, to n-region **907**. Heavily doped n+-regions **908** connect to both ends of antenna **904**. A rectenna element is thus formed by antenna **904**, rectifier structure **905***a*, and n+-regions **908**. N+-regions **908** of multiple rectenna elements are connected to constitute input channel **902** for electrons. Similarly, heavily doped n+-regions **909** connect the n-regions of multiple rectifier structures (e.g., n-region **907** of rectifier structure **905***a*) in parallel to form output channel **903** for electrons. Antenna sections of each antenna are isolated from the antenna sections of other antennae, so that each antenna may attain whatever voltage level due to the EM wave incident to that antenna. N+ regions **908** and electron input channel **902** have the same conductivity type. In another embodiment, n+-regions **908** and **902** may be replaced by heavily doped p+-regions.

[0050] When an EM wave of wavelength **2**L is received by the dipole antenna, a voltage difference is created between the ends of the antenna sections. Relative to the anode portion of the rectifier structure at the center of the dipole antenna, the voltages at the opposite ends of the dipole antennae are at opposite polarities. The resistances between n+-regions **908** to antenna **904** and between antenna **904** and the NBD are very small. The voltage difference causes electrons to flow in both antenna sections from n+-regions **908** of FIG. **10** to rectifier structure **905***a*. This is the first source of electron current in rectifier structure **905***a*. The second source of electron current in rectifier **905***a* is the electron current induced in the antenna by the magnetic field of the EM wave. This electron current has a magnitude that depends on the power of the incident EM wave. The direction of each source of electron currents at any given time is determined by the phase of the EM wave. In this configuration, the two sources provide electron currents to the dipole antenna. Maximum currents in both antenna sections occur at the center of the dipole antenna. The gap length D of antenna **904** is designed to be much smaller than ¼ the wavelength of the EM wave that antenna **904** is designed to receive. Ideally, each of antenna sections **904***a* and **904***b* is less than ¼ the wavelength of the EM wave being received. The magnitude of the current from n+-regions **908** is determined by the magnitude of the voltage difference and the magnitude of resistance between the junction of the n+-region and the contacting antenna section of the dipole antenna.

[0051] In this configuration, where the rectifier structure at the center of the dipole antenna is an NBD similar to that discussed for NBD **300** of FIG. **3**[1], the electron currents in the dipole antenna are shunted by the NBD **300** to the output channel (e.g., output channel **903**). Rectifier structure may be provided by other similar devices. As explained below, the NBD has a raised voltage at the output channel, thus the electron currents of multiple antennae are summed at electron output channel **903**. For example, on a dipole antenna as shown in rectenna array **900** of FIG. **9**, an electron current from electron input channel **902** flows to the n-regions (e.g., n+ regions **908** of FIG. **10**) surrounding the antenna sections, through the antenna sections to the NBD **300**s, and from there the electron current flows to electron output channel **903**. Thus, rectenna array **900** may be seen as a battery providing a voltage (i.e., the voltage across the NBD) between electron input channel **902** and electron output channel **903**. DC power is collected along the N-type channels diagonally and each center NBD of gap rectenna. Extra electrons are supplied by the connections of n-type channels diagonally at the ends of the antenna. The alternative embodiment shown in FIGS. **22**(*a*) and **22**(*b*) perform similarly.

---

[1] In this description, when a device requires a diode similar to NBD **300** discussed in FIG. **3**, the required diode is referred as "NBD **300**". Similarly, when a device requires a diode similar to NBD **600** of FIG. **6**, the required diode is referred as "NBD **600**".

[0052] FIG. **11** shows rectenna arrays **1101** and **1102** connected in series by connecting electron output channel **1104** of rectenna array **1101** to electron input channel **1105** of rectenna array **1102**. Thus connected, rectenna arrays **1101** and **1102** may be seen as a battery having a voltage that is twice the NBD voltage between electron input channel **1103** and electron output channel **1106**. In this manner, multiple rectenna arrays may be connected to provide a device having a voltage that is N times the NBD voltage, where N may be any positive integer greater than or equal to 1. Alternatively, rectenna arrays **1101** and **1102** may be connected together in parallel if they have the same output voltage, with electron input channels **1103** and **1105** connected together and electron output channels **1104** and **1106** connected together, to create a device having twice the current of an individual rectenna array.

[0053] As explained below, an NBD may be used to detect and rectify very high frequency signals. Accordingly, by detecting and rectifying very high frequency EM waves efficiently, the present invention provides a device for detecting and converting into DC power using frequencies of the EM spectrum that are hitherto impractical or impossible to use for a power conversion application, such as frequencies higher than the visible spectrum. The present invention finds applications in power conversion applications using a wide range of EM wave frequencies, such as infra-red (IR) and visible light (e.g., solar energy). The present invention may be utilized as a fast response photo-sensor in optical sensing and optical networks. The antenna in each rectenna element (or rectenna) can be any size and any shape of any type. A rectenna array can be configured out of many different rectenna elements. Also, each rectenna array can be different from the rectenna arrays of FIGS. **9** and **11**. The rectenna array of the present invention collects the electrons current flowing from all the input channels of the rectenna elements to all the output channels of the rectenna elements. The fully depleted region of an NBD is directly connected to the antenna.

[0054] The presence of a gap in each dipole antenna limits the highest frequency under which a rectenna (or a rectenna array) of the present invention may operate. According to one embodiment, a gapless dipole antenna is shown in top and side views in FIGS. **12**(*a*) and **12**(*b*), respectively. NBD **600** may be used as an input NBD for gapless dipole antenna, which is further discussed below in this detailed description. As shown in FIGS. **12**(*a*) and **12**(*b*), antenna **55** of dipole antenna **1200** is connected to two NBDs (formed by p-regions **53** and **54** and n-region **57**), forming an electron output terminal on one side, and to two NBDs (formed by p-region **56** and n-regions **51** and **52**), forming an electron input terminal on the other side. In other words, a rectenna element **1200** is formed by antenna **1200**, two NBDs for input at an electron input terminal and two NBDs for output at an electron output terminal. A voltage difference between the electron input terminal and the electron output terminal of a dipole rectenna element **1200** is provided by the sum of the voltages across the two NBDs. Thus, dipole antenna **1200** has a resolution of two NBDs, which is twice the resolution of dipole antenna **904** of FIGS. **9** and **10**. When an EM wave impinges on the antenna, a voltage difference is created between the ends of this antenna. The voltages at the opposite ends of the dipole antenna are of opposite polarities. The end of the antenna having the positive voltage will have electron flow induced from the input NBDs to the output NBDs because the resistance between NBD **300**s and the antenna **55** is very small. The input NBDs are provided for electron input. The output NBDs provide a rectifier structure. The input NBD directs the electrons from the electron input channels to the antenna. The output NBD serves as a rectifier structure for antenna signals. The output NBD has a forced depleted p-region that connects directly to antenna. NBD **300** may be used as an output NBD because the p-region that connects to the antenna section is a force depleted region. The output NBD is not required to have a forced depleted p-region, according to one embodiment of the present invention. A non-forced NBD[2] may be used. Unlike a forced depleted p-region, a non-forced depletion region does not rectify the electron current at forward bias. The n-region of an input NBD can be either a non-forced depleted region or a forced depleted region. The forced depleted input NBDs will build up electrons at forward bias and deliver better rectenna efficiency than the non-forced depleted input NBDs. N-region of NBD **600** is forced depleted and may be used as an input NBD for gapless rectenna design. N-region of non-forced NBD **600** is non-forced depleted and may be used as an input NBD for a gapless rectenna design. The fully depleted regions of input NBDs or output NBDs are directly connected to the antennae.

[0055] The phase of the EM wave determines which end of the antenna has an electron current. In this configuration, there is only one current source for the antenna (rectenna or rectenna element). The source of electron current is the p+-regions at the end of the antenna next to the input NBDs (e.g., p+-region **56** of FIGS. **12**(*a*) and **12**(*b*)). The magnitude of the current from the p+-regions is determined by the voltage difference and the resistance between the junction of the p+-region and the contacting antenna section of the gapless antenna. The DC voltage at both ends of antenna element **55** is equal to the NBD voltage, relative to either the input terminal or the output terminal or dipole rectenna element **1200**. The NBD can raise output voltage for signals at reverse bias. If the voltage induced by the EM wave on antenna **55** is within this input voltage range, then the voltage difference between antenna **55** and output n-channel **57** is raised. Therefore, a small signal of the EM waves may be rectified for suitable applications which were not achievable by other rectenna. DC power is collected by connecting the n-type channels under the output NBDs (p-regions **53** and **54** in FIG. **12**(*a*)). Extra electrons are supplied by the connections of P-type channels under the input NBDs. When two gapless antennae are lined up end to end with a separation distance D then this antenna pair will operate like a single gap antenna as described above. This is because the input NBDs at the far ends of the gapless antenna pair act like the gap antenna n+-type input channel and the two output NBDs in the center of the gapless antenna pair act as the center output NBD of the gap antenna.

[0056] FIG. **13** shows rectenna array **1300** including numerous gapless dipole antennae (e.g., antenna **1200** of FIG. **12**) with their respective electron input terminals connected together to form electron input channel **1302** and their respective electron output terminals connected to form output channel **1301**. Numerous such rectenna arrays may be connected in series or in parallel, as shown in FIG. **14**. As shown in FIG. **14**, electron output channel **1404** of rectenna array **1402** is connected serially with electron input channel **1403** of rectenna array **1401** through a metal trace (not shown) to provide current at twice the output voltage of the individual rectenna array. The metal trace prevents the electron input channel (p-type) and electron output channel (n-type) from forming a pn-junction, as a pn junction causes an undesirable voltage drop. Rectenna array **1401** and rectenna array **1402** can be connected in parallel by having electron input channels **1403** and **1406** connected together and electron output channels **1405** and **1404** connected together. Connecting two of the same rectenna arrays in parallel provides twice the current of an individual rectenna array.

---

[2] As discussed below, the term non-forced NBD refers to an NBD having a width in one region that is exactly equal to the depletion width of the corresponding conductivity type in an abrupt P-N junction of the same dopant concentrations.

[0057] Input device such as the NBDs provided for electron input may be provided by similar devices. Gapless rectenna element **1200** provides full wave rectification. In one embodiment, rectenna element may include an input NBD and an output NBD at two ends of the antenna for half-wave rectification. Alternatively, half-wave rectification may be achieved using either one input NBD (e.g., NBD **51**) and one output NBD (e.g., **53**), or one input NBD (e.g., NBD **52**) and output NBD (e.g., NBD **57**). In other words, a rectenna element is formed by antenna **1200**, either one NBD (e.g., NBD **51**) for electron input at an electron input terminal and one NBD (e.g., NBD **53**) for electron output at an electron output terminal, or one NBD (e.g., NBD **52**) for electron input at an electron input terminal and one NBD (e.g., NBD **57**) for electron output at an electron output terminal.

[0058] Gapless rectenna element **1200** collects current between the electron input channels and the output NBDs. Another gapless antenna **2200** utilizes two sources of current, similar to gap antenna **904**, according to another embodiment of the present invention. Dipole antenna **2200** is shown in top and side views in FIGS. **20**(*a*) and **20**(*b*), respectively. As shown in FIGS. **20**(*a*) and **20**(*b*), antenna **2205** of dipole antenna **2200** is connected to an NBD formed by p-regions **2203** and n-region **2204**, forming an electron output terminal on one side, and to an NBD formed by p-region **2201** and n-regions **2202**, forming an electron input terminal on the other side. The NBD formed by p-regions **2203** and n-region **2204** is a rectifier structure, in either of p-region **2203** and n-region **2204** can be a forced depleted region or non-forced depleted region. The NBD formed by p-region **2201** and n-regions **2202** drives electrons to the antenna **2205**, and may be implemented by an NBD, such as NBD **600** discussed below. A rectenna element includes antenna **2205**, an electron input terminal and an electron output terminal. DC power is collected by connecting n-type channels under the output NBD (e.g., n-regions **2204** in FIG. **20**(*a*)). Extra electrons are supplied by connecting p-type channels under the input NBD. The input terminals of rectenna elements are connected together and the output terminals of rectenna elements are connected together to form the rectenna array.

[0059] FIG. **21** shows gap rectenna arrays **2300** and **2350** connected in parallel in one package such that input channels and output channels are shared by both rectenna arrays. The sharing of input and output channels is shown in FIG. **21** by input channel **2302** and output channel **2309**. Rectifiers **2305***a* and **2355***a* output electrons to output **2309** and n-regions **2308** and **2358** receive electrons from input channel **2302**. Using this type of packaging provides a higher density design providing a way to increase the received EM wave bandwidth and power output. Non-gap rectenna arrays can also be arranged in parallel similarly. Full-wave rectification results from using a forced NBD in such an rectenna design. However, using a non-forced NBD achieves half-wave rectification. To perform a full-wave rectification using two half-wave rectifications, antenna **2304***a* and antenna **2304***b* each connects to a separate **2305***a* which connect to output channel **2309**, and antenna **2354***a* and antenna **2354***b* each connects to a separate **2355***a* which connect to output channel **2309**.

[0060] According to one embodiment of present invention, full-wave rectification can be achieved using an NBD for a gap dipole antenna, while half-wave rectification can be achieved using a non-forced NBD. According to one embodiment of present invention, full-wave rectification can be achieved by using either a forced or a non-forced NBD (e.g., NBD **300**) for rectifiers, but using a forced or non-forced NBD (e.g., NBD **600**) for electron input with gapless dipole antenna. Full-wave rectification may also be achieved by using two half-wave rectifications, according to one embodiment of the present invention.

[0061] Rectifier structures of the EM rectenna of the present invention may include different numbers and types of NBDs, according to one embodiment of the present invention. These NBDs of a rectifier can be connected in series, in parallel or both. Therefore, the rectifier structures of the EM rectenna can be in different configurations depending on the requirements of the desired result or performance. When the input voltage is varied within a predetermined negative bias voltage range of a NBD, the output voltage of a NBD may be higher than the input voltage by a substantially constant value. This means that the output voltage of an antenna can be raised by a DC voltage when the rectifier has serially connected NBDs. Similarly, the output current of an antenna can be regulated when rectifier has NBDs connects in parallel. By varying the output voltage or current, a variation of the EM rectennae can be provided to meet different needs and to achieve different efficiencies.

[0062] Rectifier structures of the EM rectenna of the present invention may use other devices for rectification instead of NBDs according to one embodiment of the present invention. Other rectifying devices include Schottky diodes, back diodes and tunnel diodes. Suitable rectifier devices would be any device that conduct appreciatable current in one bias direction. Any device which have rectifier qualities as described above are within the scope of this invention.

[0063] The present invention uses semiconductor material (i.e. n-type and p-type material) for electron input and output channels instead of traditional higher conductive metallic material. Semiconductor material was chosen to reduce interference between electron channels and antennae, especially in highly packed antennae structures as discussed above.

[0064] Since each antenna gathers electromagnetic waves at a frequency consistent with its size and shape, and the diode (i.e., the rectifier structure) rectifies that energy into a D.C. current. As in both gap and gapless rectenna embodiments of the present invention, each rectenna element can be any size and any shape, and can be provided by any antenna type. A rectenna array can be made out of many different rectenna elements. The present invention is not limited to a dipole rectenna design, nor to the full-wave rectifying rectenna described above. According to another aspect of the present invention, an NBD capable of rectifying a small-amplitude, high frequency signal is provided, which is suitable for use in conjunction with the energy conversion device described above. This NBD is a “forced depletion” natural breakdown device, as explained below, according to one embodiment of the present invention.

[0065] The present invention referred a p-type or n-type region as fully depleted when the whole region is depleted of its respective majority carriers. This region may include different materials in any forms, shapes, dimensions, conductivities or concentrations.

[0066] The present invention referred a p-type or n-type region as fully depleted when the whole region is depleted of its respective majority carriers. This region may include different materials in any forms, shapes, dimensions, conductivities or concentrations. The doping level and other parameters may be varied independently to achieve the desired results. Although the examples and drawing shown herein for NBDs shown regions of homogeneous dopant concentrations, such regions are only provided for illustration purpose only. The present invention is equally applicable in devices where the dopant concentrations are non-homogeneous.

[0067] According to one embodiment of the present invention, a NBD type diode includes a semiconductor (say, p-type region) that has a width w_p that is less than or equal to the depletion width x_p of a conventional abrupt pn-junction without an externally imposed voltage. That is:

$$w_p \leq w_p = \sqrt{\frac{2\varepsilon_s N_D \phi_i}{q N_A (N_A + N_D)}}$$

[0068] where ϵ_s the electrical permittivity of silicon, q is the charge of an electron, ϕ_i is the "built in" potential of the pn junction, N_A and N_D are the doping concentrations of p-region **101** and n region **102**, respectively. As mentioned above, when $w_p = x_p$ ($w_n = x_n$), the region is referred to as a non-forced depleted region, and when $w_p \leqq x_p$ ($w_n \leqq x_n$), the region is referred to as a forced depleted region

[0069] FIG. **3** is a schematic representation of NBD **300**, according to one embodiment of the present invention. As shown in FIG. **3**, NED **300** includes p-region **301** and n-region **302**, with p-region **301** having a width w_p, which is less than or equal to the depletion width x_p of a corresponding depletion width in the p-region side of a conventional pn junction. (A semiconductor region having a depletion width that is less than the depletion width for the corresponding semicondcutor type in a conventional pn junction is referred herein as having a "forced depletion" width). In contrast, the width of n-region **302** may be smaller than, greater than or equal to the conventional depletion region x_n for the n region of a conventional pn junction diode of comparable doping concentrations. One embodiment of the present invention is a forced depletion condition on p-region **301**, and n-region **302** having a width greater than x_n. Also shown are contact regions **303** and **304** which are contacts to allow NED **300** to be connected to an electronic circuit. The doping concentrations in p region **301** and n-region **302** are sufficiently high such that contacts **303** and **304** are ohmic contacts. Contact region **303** and **304** may be formed, for example, by depositing a conventional interconnect conductor, such as aluminum or copper, using conventional chemical vapor deposition techniques, or other means known to those skilled in the art. P region **301** and n region **302** may be formed in a conventional silicon substrate using ion implantation, or other means known to those skilled in the art.

[0070] Width w_p of an NBD **300** may be calculated based upon the doping concentration. The predetermined width w_p for NBD **300** may be calculated using the following steps:

[0071] (1) First choose doping concentrations for a p-region and an n-region of a conventional PN junction diode such that, under the zero applied bias voltage, the p-region has a depletion width X_p between point **4** and point **3** and the n-region has a depletion width X_n between point **3** and point **2**, as shown in FIG. **3**. These dimensions create a built-in voltage V_D volts in the conventional pn junction diode. In an NBD, the built-in voltage V_{FD} is less than V_D. Voltage V_{th} is the threshold voltage for the conventional pn junction diode. The Fermi level of the NBD at zero applied bias voltage is shifted towards the Fermi level of the intrinsic semiconductor and away from the equilibrium Fermi level of a conventional pn junction diode.

[0072] (2) Select a voltage V_s, which is the maximum forward bias voltage that can be applied on NBD **300** before the region that is fully depleted at zero applied bias voltage ceases to be fully depleted. A number of desired features (e.g., constant tunneling current at low forward bias voltage, charge storage, voltage range of tunneling) may be achieved by the NBD between zero applied bias voltage and voltage V_s, depending on the application of the NBD. When the applied bias voltage is increased from zero towards voltage V_s, the Fermi level of the NBD is decreased towards the Fermi level of the conventional pn junction diode. However, if the applied bias voltage is decreased in this range, the Fermi level maintains a constant level, so that a charge is accumulated in the n-type region as the applied voltage decreases from V_s to zero.

[0073] As the value of voltage V_s is between zero and the conventional pn junction diode built-in voltage V_D, the voltage across the pn-junction is given by $V_x = V_D - V_S$. In one embodiment of the present invention, the value of voltage V_S is selected to be close to the built-in voltage $V_D (V_S \sim V_D)$. The width of the fully depleted depletion region required to achieve this condition is very thin. In another embodiment of the present invention, the value of V_s is not selected to be close to V_D. According to one embodiment of the present invention, the value of V_s can be tuned or selected to accommodate an active voltage range for the NBD operation. The depletion width of NBD **300** remains unchanged as long as the applied bias voltage is less than V_S. As explained above, when an applied bias voltage on an NBD is between zero and V_s, a charge is accumulated. An NBD with a smaller V_S stores less charge than an NBD which has a larger V_S.

[0074] (3) Calculate the depletion width w_p of p-region **301** such that, when voltage V_s is imposed on **303** toward p-region **301**, the whole p-region **301** remains depleted. Assuming an abrupt junction approximation, the width w_p may be calculated using the following equation:

$$x_p = \left[\frac{\frac{kT}{q}\ln\left[\frac{N_D N_A}{n_i^2}\right] - V_S}{\frac{e}{2\varepsilon_S}\left(\frac{N_A^2 + N_A N_D}{N_D}\right)}\right]^{\frac{1}{2}}$$

[0075] There are other ways to calculate w_p as known by those skilled in the art. The built-in voltage V_{FD} of the NBD **300** equals to—($V_D - V_S$) volts. If $w_p = x_p$, V_s is zero volts.

[0076] Note that the width w_p is calculated above using an abrupt junction approximation. Other suitable methods may also be used. The width w_p may be calculated using different junction approximations, depending on the application. As explained above, the condition $w_p < x_p$ is referred to as a "forced depletion condition" and, under such a condition, p-region **301** is referred to as a "forced depletion region", according to one embodiment of the present invention. When p-region **301** is in a forced depletion condition, the value of V_S is not equal to zero. The condition $w_p = x_p$, is referred to as a "non-forced depletion condition" and, under such a condition, p-region **301** will be referred to as a "non-forced depletion region", according to another embodiment of the present invention.

[0077] When p-region **301** is in non-forced depletion condition, the value of V_S is equal to zero. Once w_p is determined, an NBD **300** with the p-region **301** fully depleted between contact region **303** and n-region **302** may be created with different width of n-region **302**. W_n is the depletion region width of n-region **302** on NBD **300**. The width of n-region **302** may range from w_n to larger than x_n. The difference in the width of n-region **302** creates variations for the NBD **300**. NBDs can be created having a width for n-region **302** greater than x_n. When the external voltages applied to contacts **303** and **304** is zero (zero applied voltage bias or without bias), p-region **301** of NBD **300** is fully depleted. One embodiment of the invention provides a forced depletion condition on p-region **301** with the w_P is less than x_p and the width of n-region **302** being a value between w_n to larger than x_n. One embodiment of the invention is a non-forced depletion condition in which p-region **301** with a w_p equal to x_p and the width of n-region **302** may range from w_n to larger than x_n.

[0078] When the applied bias voltage is less than zero volts, even by a small amount, the accumulated charge is relaxed. Thus, the NBD is conducting under a reversed bias condition.

[0079] Alternatively, the forced or non-forced depletion region dimensions can be obtained by starting with the dimensions of a conventional pn-junction, and adjusting the dopant concentrations of the p-region and n-region such that the width of one of the region ("first region") is exactly equal to its depletion region (non-forced depletion condition), if the regions form a conventional pn-junction, or further increasing the concentration of the other region, such that the width of the first region is now less than its expected depletion width, calculated as if the regions form a conventional pn-junction.

[0080] In another embodiment of the present invention, shown in FIG. **6**, n-region **602** may also be put under a forced depletion condition without a forced depletion or non-force depletion in p-region **601**, NBD **600** variations. An NBD can have more than one fully depleted region. The differences in the width of p-region **601** create variations for NBD **600**. One embodiment of the invention provides a non-forced depletion n-region **602** (i.e., a w_n that is equal to x_n) and a width of p-region **301** between w_p to larger than x_p. The differences in the width of p-region **601** create variations for this embodiment. Another embodiment of the invention provides a forced depletion condition on n-region **602** with the w_n is smaller than X_n and the width of p-region **301** may range from w_p to larger than x_p, according to this embodiment. The differences in the width of p-region **601** create variations for this embodiment. These variations of NBD are natural breakdown diodes. An NBD can have more than one fully depleted region. Each of these regions can be either a forced depletion region or a non-forced depletion region.

[0081] An NBD has one of the p-region or n-region fully depleted under zero biasing. NBD **300** has p-region **301** in a forced depletion condition and an n-region **302** with its width larger than x_n. The operations of NBD **300** are explained with respect to an external applied bias voltage under following conditions:

[0082] a) When a zero forward bias voltage is applied on NBD (i.e., at zero bias), the input voltage is V_{IN}, (i.e., $V_{IN}=0$).

[0083] b) When a forward bias voltage is applied on NBD **300** between 0 and V_s (i.e., $0<V_{IN}<=V_S$).

[0084] c) When a forward bias voltage applied on NBD **300** is between V_s and V_{th} (i.e., $V_s<V_{IN}<V_{th}$).

[0085] d) When a forward bias voltage applied on NBD **300** is larger than V_{th} (i.e., $V_{IN}>V_{th}$).

[0086] e) When a reverse bias voltage V_{IN} is applied on NBD **300**.

[0087] Accordingly, (a) When $V_{IN}=0$, an electric field (referred to as the "secondary electric field") with a voltage difference of V_s is created along the edge of the depletion region in n-region **302**. No current flows in the device. The voltage difference between the depletion region electric field and the built-in potential is V_s.

[0088] (b) When V_s is selected to be not close to V_D, $0<V_{IN}<=V_s$ and V_{IN} is applied at contact **303** on NBD **300**, the depletion width in p-region **301** remains the entire width of p-region **301** because V_{IN} is smaller than the voltage potential across the secondary electric field in n-region **302**. As V_{IN} increases, the Fermi level between the depletion region and the equilibrium Fermi level between the p-region **301** and n-region **302** decreases. As the applied bias voltage V_{IN} increases, the secondary electric field strength decreases in n-region **302**. Once V_{IN} reaches V_s, the secondary electric field in n-region **302** becomes zero. NBD **300** responds to a signal change faster than a conventional pn junction, so that NBD **300** is suitable for rectifying high speed signals. This fast response time characteristic is present in all NBDs. A device that utilizes NBD concepts will have a faster response time than the same device not utilizing NBD concepts.

[0089] When V_s is selected to be close to V_D, for an applied bis voltage V_{IN} less than V_s (i.e., $0<V_{IN}<=V_s$), a constant forward current flows in NBD **300** due to tunneling through the sufficiently thin depletion region. The depletion width in p-region **301** remains the entire width of p-region **301**.

[0090] (c) When the applied bias voltage V_{IN} is between V_s and the threshold voltage (i.e., $V_S<V_{IN}<Vth$), the depletion widths in both p-region **301** and n-region **302** reduce. The voltage drop across the depletion regions reduces also. In this regime, a small forward leakage current proportional to V_{IN} flows in NBD **300**. As V_{IN} becomes very close to V_{th}, the depletion width in NBD **300** becomes significantly small to allow a significant current to flow.

[0091] (d) When the applied voltage V_{IN} exceeds the threshold voltage (i.e., $V_{IN}>=V_{th}$), NBD **300** conducts current.

[0092] (e) When a reverse bias voltage V_{IN} is applied to NBD **300**, the depletion width remains the same. As the reverse bias voltage increases the reverse current increases substantially linearly based upon the resistance between contact **303** and p-region **301** and the resistance between contact **304** and n-region **302**.

[0093] (f) When reversed biased, NBD **300** operates as a majority carrier device (electrons being injected in to the n-region) as apposed to a minority carrier device when forward biased. The switching times of majority carrier device are typically faster than switching times of minority carrier device.

[0094] FIG. **5** is a plot of the current versus voltage (IV) characteristics of NBD **300**, when V_s is selected not to be close to V_D. NBD **300** can conduct appreciable reverse current when input voltage is negative and conduct forward current when input voltage is larger than V_{th}. NBD **300** conducts a negligible leakage current when the bias voltage is between 0 and V_{th}. When a forward bias voltage between 0 and V_s is applied to an NBD, the depletion region width does not change as long as the Fermi level of the device is greater than the depletion region Fermi potential. This behavior is illustrated by FIG. **4** which shows the relationship between NBD **300** depletion region voltage V_{FD} and the input voltage V_{IN}, where the behavior of the depletion width is also shown. FIG. **24** is a plot of the current versus voltage (IV) characteristics of NBD **300**, when V_S is selected to be very close to V_D, where a constant forward current due to tunneling is shown for an applied bias voltage between 0 volts and V_s.

[0095] In one embodiment the width of NBD **300** p-region for creating a constant current is calculated by the following:

[0096] 1) Using the IV curve of a conventional diode utilizing the same materials and doping concentrations for NBD device, find forward bias voltage that matches desired current on curve. The bias voltage needs to be less than V_{th}. This bias voltage will be the V_s value.

[0097] 2) Determine width of depletion band of p-region in conventional diode when conventional diode is biased at V_s determined in step 1). This can be done through calculation or simulation of device as known by those skilled in the art. This width is W_P.

[0098] 3) Create an NBD with a p-region of width W_p as determined in step 2) using the same material and doping concentration of conventional diode used in step 1).

[0099] An NBD device behaves asymmetrically for an increasing applied bias voltage between 0 volts and voltage Vs and a decreasing applied bias voltage over the same voltage range. As the applied bias voltage increases, the Fermi level of the device decreases, so that the secondary electric field decreases. However, when the applied bias voltage decreases over the same range, a third electric field appears at the secondary electric field edge in the n-region **302** neutral region, which continues to increase without causing a change in the Fermi level. If the bias voltage is again increased, this third electric field decreases. Once the applied bias voltage becomes negative, the charge in the third electric field is released, thereby causing a reverse current surge. The maximum voltage across the third electric field is V_s. Therefore, applying a negative bias voltage after applying a positive bias voltage on NBD **300** would result in a reverse current.

[0100] Another embodiment of the present invention has a forced depletion condition on p-region **301** and an n-region **302** with its width less than X_n where the depletion band width does not completely cover n-region **302**. The behavior of this device is the same as NBD **300** described above. For the situation when a forced depletion condition is created, the depletion width of n-region **302** covers the whole n-region **302** between contact **304** and complete depleted p-region **301**, ("NBD **310**"): The behavior of NBD **310** under condition (a) of $V_{IN}=0$ is the same as described above. The behavior of NBD **300** under condition (b) when $0<V_{IN}<=V_S$, status is the same as described above until the n-region **302** becomes fully depleted. Once n-region **302** is fully depleted, NBD **300** reaches condition (c) prior to $V_{IN}=V_s$. The behavior of NBD **300** under condition (c) is the same as described above.

[0101] When a NBD **300** is connected in series to a current source (i.e. a signal from an antenna or sensor), a voltage appears across NBD **300**. Since NBD **300** conducts current at reverse biased voltage, a negative voltage potential appears when a current source imposes a reverse bias current.

[0102] Thus, NBD **300** can provide a small positive current when a positive bias voltage V_{IN} less than V_s is applied (i.e., $0<V_{IN}<=V_s$) is applied at contact region **303**. NBD **300** can also rectify negative input signals (i.e., $V_{IN}<0$). NBD **300**, with the capability of storing charge between $0<V_{IN}<=V_s$, and conducting a reverse current at reverse bias, can rectify very high frequency signals, within the regime from V_s to $-V_s$ and negative voltage signals.

[0103] Another embodiment of the present invention provides a forced depletion p-region **301** and a forced depletion n-region **302** (i.e., having a width less than x_n). The behavior of NBD is the same as NBD **300** described above.

[0104] Another embodiment according to the present invention has a non-forced depletion width (i.e., $w_p=x_p$ or $w_n=X_n$). In this configuration, the resulting NBD is conductive at reverse bias. A non-forced NBD has the same reverse bias characteristics of an NBD **300**. However, when a forward bias voltage is applied on a non-forced NBD, its current and voltage characteristics are the same as those of a conventional pn junction diode.

[0105] To summarize, an NBD of the present invention allows a conductive current flow when a reverse bias voltage applied.

[0106] If an NBD and a current source are connected in series, the output voltage appears across the NBD. By selecting a value of V_s at forward bias, NBD may be provided desirable features such as charge storage and constant current. If the applied bias voltage exceeds the threshold voltage, the NBD provides a conductive current. NBD **300** stores charge by a decreasing applied bias voltage between 0 volts and V_s; the stored charge is released when applied bias voltage becomes negative. Thus, an NBD can full wave rectify even a small sinusoidal signal, thus is suitable for applications such as solar energy generation. The application of the NBD invention to conventional PN junction diodes created two new active bias voltage ranges, namely the reverse bias voltage and forward bias between zero and Vs volts. These two new active regions enable NBD modified PN junction diodes to have three active regions to be utilized for applications.

[0107] An NBD does not have a breakdown voltage. As one of the regions of a pn junction is fully depleted, no leakage current is observed even under a reverse bias condition. The built-in electric field provides high conductivity at reverse bias by zeroing the distance for electrons and holes to travel across the pn-junction to the opposite region. A reverse current flows when the applied bias voltage V_{IN} is negative (i.e., $0>V_{IN}$). Constant current or charge storage at forward bias can be provided by a non-zero V_s value, depending on the application.

[0108] FIG. **23** shows the IV curve of NBD **300** when contact resistance is very low, according to one embodiment of present invention. In this embodiment, the IV curve is very close to the ideal diode IV curve. Because the value of V_s is not equal to zero, the NBD **300** will have charge stored when forward bias decreasing toward zero volts. This charge will be released when the NBD **300** is reverse-biased and creates a reverse current surge. The NBD **300** may be considered as an ideal diode with reverse current surge capability. In FIG. **23**, when $V_S=0$, the IV curve is the same as the IV curve of non-forced NBD **300** with very low contact resistance, according to one embodiment of present invention. The non-forced NBD **300** may be considered as an ideal diode with a breakdown voltage at the forward threshold voltage.

[0109] According to another embodiment of the present invention, as discussed above, FIG. **6**(*a*) shows NBD **600** with n-type region **602** having a width w_n that is less than x_n—the depletion width of a conventional pn junction diode:

$$w_n \le x_n = \sqrt{\frac{2\varepsilon_s N_A \phi_i}{qN_D(N_A + N_D)}}$$

[0110] A similar determination provides width W_n for NBD **600**. The operation of a forced depletion condition with n-region **602** with its depletion band width $w_n<X_n$ and p-region **601** larger or equal to x_p behaves as follows:

[0111] At (a) zero applied bias voltage, no current flows in the device. As $W_n<x_n$, an excess number of holes builds up along the edge of the depletion region in p-region **601**, thereby creating a secondary electric field. The voltage potential across this secondary electric field is V_s.

[0112] At (b), when voltage V_s is selected not to be close to V_D and the applied bias voltage is between 0 volts and V_s (i.e., $0<V_{IN}<=V_S$), n-region **602** remains fully depleted. As the applied bias voltage V_{IN} increases, the secondary electric field decreases. Once V_{IN} equals V_s, the secondary electric field in p-region **601** becomes zero. Thus, NBD **600** responds to a signal change faster than a conventional pn junction because the secondary electric field width is smaller than the depletion width of a conventional pn junction. Thus, NBD **600** is suitable for rectifying high speed signals.

[0113] When voltage V_s is selected to be close to V_D and for an applied bias voltage between 0 volts and V_s (i.e., $0<V_{IN}<=V_s$), a constant forward current due to tunneling through the sufficiently thin depletion region. N-region **602** remains fully depleted the entire width of n-region **602**.

[0114] (c) When the applied bias voltage is between V_s and the threshold voltage (i.e., $V_S<V_{IN}<V_{th}$), the depletion widths in both p-region **601** and n-region **602** reduce. The current in the device is a small forward leakage current that is proportional to V_{IN}. The output voltage is V_{IN}-V_D. When V_{IN} is very close to V_{th}, the depletion width on NBD **600** is significantly small for conductive current to flow.

[0115] (d) When the applied bias voltage exceeds the threshold voltage (i.e., $V_{IN}>=V_{th}$), NBD **600** conducts current.

[0116] (e) When a reverse bias voltage V_{IN} is applied to NBD **600**, the depletion width remains the same. As the reverse bias voltage increases the reverse current also increases substantially linearly.

[0117] NBD **600** has the same current versus voltage characteristics as NBD **300**. NBD **600** also stores charge under a forward-biased voltage between 0 and V_s in the same way as NBD **300**. Also, like the NBD **300**, when NBD **600** is connected in series to a current source a voltage appears across NBD **300**.

[0118] Another embodiment according to the present invention provides $w_n=x_n$ (i.e., the non-forced depletion condition). Non-forced NBD **600** has the same behavior as a non-forced NBD **300**.

[0119] According to one embodiment of the present invention, FIG. **6**(*b*) shows an NBD having two regions (p-region and n-region) which are both under forced depletion condition. According to other embodiments of the present invention, FIGS. **7**(*a*), **7**(*b*) and (*c*) show different NBD configurations, represented by NBD **700**, NBD **710** and NBD **720**, each including a region (e.g., **701**, **711**, **721** or **722**) under the forced depletion condition. Also, when region **702** is p-type and region **712** is n-type, then NBD **700** and NBD **710** can be seen as NPN and PNP transistors respectively. FIGS. **8**(*a*), **8**(*b*), **8**(*c*), **8**(*d*) and **8**(*e*) show different NBD configurations **800**, **810**, **820**, **830** and **840** each including a region (e.g., **801**, **811**, **821**, **822**, **831**, **832**, **833**, **841**, **842**, and **843**) having two or more doping concentrations under forced depletion conditions. FIGS. **15**(*a*) to **15**(*d*), **16**(*a*) and **16**(*b*) show some of the embodiments of the NBD, according to the present invention.

[0120] FIGS. **17**(*a*) to **17**(*f*) show some NBDs at a zero applied bias voltage, each having a forced depletion p-region (or n-region) formed adjacent to a Schottky barrier or an ohmic contact. The Schottky barrier or ohmic contact imposes a fully depletion p-region or a fully depletion n-region. Each of these fully depleted region may be considered non-forced or forced depleted regions.

[0121] FIG. **18**(*a*) shows NBD **1800** at a zero applied bias voltage, and conductors **1801** and **1803** which are contacts provided for connecting NBD **1800** to an electronic circuit. In NBD **1800**, n-region **1802** is fully depleted under a forced depletion condition, having a width less than depletion width **1804**. The doping concentration in n-region **1802** is sufficiently high such that the junction between conductor **1803** and n-region **1802** is an ohmic contact and conductor **1801** forms a Schottky barrier to n-region **1802**. NBD **1800** under reverse bias voltage performs substantially in the manner described above for NBD **600**, having the characteristics such as fast switching, conducting at a reverse bias, and a reverse voltage when connected with a current source in series. NBD **1800** under forward bias performs substantially the same as a conventional n-type Schottky diode under forward biased utilizing comparable materials. The forced depletion width described above for the pn junction configuration may be applied to determine the forced depletion width of a fully depleted p-region **1812** (or n-region **1802**).

[0122] FIG. **18**(*b*) shows NBD **1810** at a zero bias, and conductors **1811** and **1813**, which are contacts providing for connecting NBD **1810** to an electronic circuit. P-region **1812** is fully depleted under a forced depletion condition, so that the width of p-region **1812** is less than depletion widths **1814**. The doping concentration in p-region **1812** is sufficiently high, such that the junction between conductor **1813** and p-region **1812** is an ohmic contact, and conductor **1811** forms a Schottky barrier to p-region **1812**. NBD **1810** when reverse biased performs substantially in the manner describe above for NBD **300**, having characteristics such as fast switching, conducting at reverse bias, and a reverse voltage when connected with a current source in series. NBD **1810** under forward bias performs substantially the same as a conventional p-type Schottky diode under forward biased utilizing comparable materials.

[0123] To determine a forced depletion width for NBD **1800**, (1) a depletion width x_n, built-in voltage V_D and threshold voltage V_{th} of a conventional Schottky diode are determined using an n-region doping concentration at a zero bias, (2) a forward-bias work voltage V_s that is between zero and V_{th} is selected that can be used with Schottky diode **1800**, and (3) the depletion width w_n of n-region **1802** is calculated such that, when a forward bias voltage V_s is applied on NBD **1800**, n-region **1802** remains depleted with a built-in voltage V_{FD} that is equal to $-V_x$ (V_x is given by $V_x=V_D-V_s$). Regions **1802** and **1812** include, respectively, multiple p-type and n-type sections of different doping concentrations. According to another embodiment of present invention, an NBD may also be formed using three semiconductor regions, one or more of which is fully depleted. For example, all three regions may be fully depleted. A transistor having at least one fully depleted semiconductor region is an NBD within the scope of the present invention, which may used, for example, in conjunction with the rectenna described above.

[0124] According to another embodiment of the present invention, FIG. **19**(*c*) shows NPN bipolar device **2020** with no external voltages being applied. Device **2020** has n-type region **2026** fully depleted, p-type region **2025** sufficiently large to have a neutral region between depletion regions **2027** and **2028** and n-type region **2024** sufficiently large to have a neutral region between depletion region **2028** and contact **2023**. The p-type region **2025** (neutral region) is sufficiently large to prevent overlap of minority diffusion from depletion regions **2027** and **2028**. Device **2020** is considered a forced depleted device when n-type region **2026** is forced depleted and is considered non-forced depleted device when n-type region **2026** is non-forced depleted. Device **2020** is considered a conventional device when n-type region **2026** has a neutral region between contact **2021** and depletion region **2027**.

[0125] FIG. **28** shows IV curves **2801**, **2802** and **2803** for a conventional device, a non-forced depleted device and a forced depleted device respectively. IV curves **2801**, **2802** and **2803** are exaggerated and provided only for illustration purposes only. The bias voltage is with respect to fully depleted n-type region **2026** therefore forward biasing device **2020** would be to have contact **2023** at a positive voltage relative to contact **2021**. When device **2020** is a conventional device, a forward bias decreases the depletion region **2027** width and depletion region **2028** width increases. From IV curve **2801**, as the forward bias increases across conventional device **2020** current remains small and constant (i.e. leakage current) until depletion region **2027** conducts and depletion region **2028** width increases sufficiently so that its minority diffusion reaches depletion region **2027**. As the forward bias is further increased across device **2020** (conventional), the current increases until depletion region **2028** reaches depletion region **2027** where current increases exponentially. TV curve **2801** is based on device **2020** being conventional and having n-type region **2024** and n-type region **2026** each having the same neutral region width. Therefore, when device **2020** is conventional, it has similar characteristics reverse biased as forward biased in that depletion region **2028** width decreases and depletion region **2027** width increases.

[0126] Comparing a non-forced depleted device IV curve **2803** to conventional device TV curve **2801**, it can be seen that during forward biasing both devices behave similarly except during reverse biasing the non-forced depleted device has exponential increase in current at a smaller voltage magnitude then the conventional device. This behavior can be understood by the fact that a non-forced depleted region behaves the same as a conventional region when forward biased but conducts immediately when reverse biased, as explained above. Therefore when non-forced depleted device is reversed biased exponential increase of reverse current will occur as soon as depletion region **2028** start having exponential increase in current instead of when depletion regions **2027** and **2028** contact each other.

[0127] Comparing forced depleted device IV curve **2802** to conventional device IV curve **2801**, it can be seen that there are differences in both forward biasing and reverse biasing. As explained above, a forced depleted region behaves differently during increasing forward biasing verses decreasing forward biasing. During increasing forward biasing between zero and V_s depletion region **2027** width remains the same until, depletion region **2028** width increases and the n-type region **2026** Fermi level moves toward its non-forced Fermi level. During decreasing forward biasing between zero and VS depletion region **2027** width remains the same and depletion region **2028** width also remains the same. This is due to the potential voltage created across p-type region **2025** neutral region maintaining the same potential voltage between p-type region **2025** and n-type region **2024**. After an increase or decrease forward bias cycle between zero and V_s, the threshold voltage for the n-type region **2026** or p-type region **2025** junction will be reduced by V_s volts. Also after this forward bias cycle depletion region **2028** will be increased (i.e. the threshold voltage increased) by V_s volts. It can be seen that a voltage of V_s has been transferred from a threshold voltage in forced depleted region (i.e. n-type region **2026**) to a threshold voltage in an adjacent semiconductor region (i.e. p-type region **2025**).

[0128] IV curve **2802** shows the IV characteristics for forced depleted device **2020** after an increasing/decreasing forward biasing between zero and V_s. The reduced threshold voltage in forced depleted n-type region **2026** means during forward biasing current will flow across depletion region **2027** at a lower voltage than either the non-forced depleted device and conventional device. Also, during forward biasing, exponential current increase occurs at a lower voltage due to the increase width of depletion region **2028**. At reverse bias the increased width of depletion region **2028** requires a larger negative bias before exponential current will occur.

[0129] A fully depleted region within a NBD of the present invention will obey the following rules:

[0130] 1) If the fully depleted region is forced depleted, then at thermal equilibrium the Fermi level in the forced depleted region will adjust toward the intrinsic semiconductor Fermi level by the amount necessary to create a net zero current through the device.

[0131] 2) A voltage difference in neutral regions adjacent to fully depleted regions is created before the fully depleted region's Fermi level is adjusted toward the intrinsic semiconductor Fermi level.

[0132] 3) A fully depleted region's Fermi level will adjust toward its original level before voltage differences are created in neutral regions adjacent to the fully depleted region.

[0133] Based on the above rules characteristics for forced depleted regions within a NBD are as follows:

[0134] At thermal equilibrium prior to any bias voltage:

- **[0135]** The forced depleted region's Fermi level is forced toward the semiconductor intrinsic level. The voltage associated to this energy shift is V_s.
- **[0136]** All depletion regions within fully depleted regions will have reduced widths compared to non-fully depleted and conventional region of same material as fully depleted region.

[0137] During increasing forward bias between zero and V_s:

- **[0138]** The forced depleted region's Fermi level moves toward its original Fermi level.
- **[0139]** Depletion regions associated with the forced depleted region maintain there same widths.

[0140] During decreasing forward bias between zero and V_s:

[0141] If there are no adjacent neutral regions to the forced depleted region, then the forced depleted region's Fermi level moves toward intrinsic level.

[0142] If there are no depletion regions which can increase in width within adjacent neutral regions to the forced depleted region then charge will build up in adjacent neutral regions. The forced depleted region's Fermi level will stay the same. Depletion regions associated with the forced depleted region maintain there same widths.

[0143] Any depletion region which can increase in width within adjacent neutral regions to the forced depleted region will increase in width. Built-in potentials associated with the forced depleted region will be reduced making their threshold voltages lower. The forced depleted region's Fermi level stays the same. Depletion regions associated with the forced depleted region maintain their same widths.

[0144] During reverse bias:

[0145] If the device has only one depletion region, then the device will conduct at near zero bias (no threshold voltage), any charge storage due to decreasing forward biasing is immediately released, the forced depleted region's Fermi level moves toward its semiconductor intrinsic level until reaching the thermal equilibrium level. The depletion region in forced depleted region creates a negative potential voltage through device.

[0146] If the device has more than one depletion region, then the device will have a reverse threshold voltage (i.e. a breakdown voltage). Any charge stored stays until the device conducts a reverse current. The forced depleted region adds a negative potential voltage when the device is conducting reverse current.

[0147] Based on the above rules characteristics for non-forced depleted regions within a NBD are as follows:

[0148] Thermal equilibrium

[0149] The same as when utilizing a non-fully depleted region.

[0150] Forward bias increasing

[0151] The same as when utilizing a non-fully depleted region.

[0152] Forward bias decreasing

[0153] The same as when utilizing a non-fully depleted region.

[0154] Reverse bias

[0155] If the device has only one depletion region, then the device conducts at near zero bias (no threshold voltage). The depletion region in non-forced depleted region creates a negative voltage through device.

[0156] If the device has more than one depletion region, then the device has a reverse threshold voltage (breakdown voltage). The non-forced depleted region adds a negative voltage when the device is conducting a reverse current.

[0157] FIGS. **27.1** to **27.4** show a table of NBD structures with their expected characteristics. The first column on the left side of the table is the structure of the NBD as if the structure where laid out in linear manner. The terms used to indicate the structure are the following:

[0158] "Sch"—Contact-Semiconctor Schottky barrier

[0159] "Ohm"—Contact-Semiconductor Ohmic barrier,

[0160] "N non-F"—Non-forced depleted n-type region

[0161] "P non-F"—Non-forced depleted p-type region

[0162] "N Forced"—Forced depleted n-type region

[0163] "P Forced"—Forced depleted p-type region

[0164] "N"—n-type region with a neutral region

[0165] "P"—p-type region with a neutral region

[0166] The second column of the table indicates whether or not the fully depleted region has its Fermi level adjusted toward the semiconductor intrinsic level. A forced depleted region has its Fermi level adjusted (i.e. "Yes") while a non-forced depleted region does not (i.e. "No"). The third and forth columns define the characteristics that the structure has during increasing and decreasing forward biasing respectively. Terms used in these columns are the following:

[0167] "->Efn"—Indicates that the Fermi level of fully depleted region moves toward its original Fermi level prior to being fully depleted.

[0168] "->Efn"—Indicates that the Fermi level of fully depleted region moves towards the semiconductor intrinsic level.

[0169] The fifth column indicates whether or not the NBD structure conducts a reverse current at near zero reverse bias with a reverse potential across structure. The sixth column indicates whether or not the structure maintains the same bias voltage to conduct current (i.e. threshold voltage) after biasing the structure. A "No" means that after an increasing forward bias followed by a decreasing forward bias (a forward bias cycle) that one or more junctions have a different potential barrier then prior to the forward cycle.

[0170] According to another embodiment of the present invention, FIGS. **19**(*a*) and **19**(*b*) show NPN bipolar transistors **2000** and **2010** with no external voltages being applied. As shown in FIG. **19**(*a*), NPN bipolar transistor **2000** has a fully depleted n-type region **2006** which may function either as an emitter or as a collector for NPN bipolar transistor **2000**. FIG. **19**(*b*) shows NPN **2010** having fully depleted n-type regions **2014** and **2016** as an emitter and a collector, respectively. Different transistor characteristics may be achieved using one or more different semiconductor regions that are under a forced or non-forced depletion condition. For example, depleted regions **2006**, **2014** and **2016** in NPN transistors **2000** and **2010** may be put under forced or non-forced depletion conditions. Base P-type regions **2005** and **2015** may also be made fully depleted. PNP bipolar transistor structures may also be formed according to principles described for NPN transistors **2000** and **2010** above.

[0171] The behaviors of NPN and PNP transistors change when one or more of their semiconductor regions are completed depleted. Using NPN transistor **2000** configured in a common-emitter configuration (i.e. with emitter terminal **2003** grounded) as an example, when emitter region **2006** is fully depleted, a current flows from emitter region **2006** to base region **2005** (i.e. electrons flow from base region **2005** to emitter region **2006**) when emitter region **2006** is reversed biased or zero biased relative to base region **2005**. This current affects the cut-off mode operation of NPN transistor **2000**. Depending on whether or not emitter region **2006** of NPN transistor **2000** is in a forced depleted condition, determines base-to-emitter voltage V_{BE} is in a range that maintains emitter region **2006** fully depleted. FIG. **26** shows an example of the IV characteristics of NPN transistor **2000** when the emitter is forced depleted. The dashed line shows the IV characteristics of a conventional NPN transistor.

[0172] According to another embodiment of the present invention, four or more regions in different combinations of fully depleted or un-depleted (i.e., conventional) semiconductor regions may also be used to form NBDs. Further, whether or not a center semiconductor region may be fully depleted between two fully depleted adjacent semiconductor regions results in different NBDs. Therefore, all variations in an NBD with respect to the semiconductor material types, the numbers of fully depleted regions (whether or not in a forced-depletion condition), the numbers of un-depleted regions, the alignment or orientation of the semiconductor regions, the different doping concentrations are within the scope of the present invention. The forced and non-forced depletion conditions of the present invention can be applied to any device having one or more semiconductor regions that is not fully depleted to modify the behavior for such a device. This may also result in new devices.

[0173] It is known in the art that there are no contact materials that create a true p-type ohmic contact. Instead a p-type Ohmic contact may be emulated using a p-type Schottky contact with a sufficiently thin depletion region. The thin depletion region allows tunneling, as it is created using a highly doped p-type material. Using a highly doped p-type material may be undesirable or may not provide a low enough resistance. Another embodiment of the present invention creates a p-type Ohmic contact using an NBD structure which includes a contact, a forced depleted n-type region adjacent to the contact and a p-type region adjacent n-type region opposite the contact. As explained above, the built-in potentials between the forced depleted n-type region and the contact and the p-type region are reduced after an increasing and decreasing forward bias cycle if the p-type region has another depletion region within its neutral region. The forced depleted n-type region is sufficiently thin to allow tunneling between the contact and the p-type region. This p-type ohmic NBD structure may reduce the resistance between a p-type region and a contact in some circumstances. In another embodiment of the present invention a contact, a forced depleted p-type region adjacent to contact and a n-type region adjacent to n-type region opposite contact forms a n-type ohmic NBD structure.

[0174] A fully depleted region may be created using a p-type or an n-type material adjacent to an intrinsic material. In another embodiment of the present invention, an NBD uses a forced depleted region created by a p-type or an n-type semiconductor region adjacent to at least one intrinsic semiconductor region. In another embodiment of the present invention a NBD uses a non-forced depleted region created by a p-type or an n-type semiconductor region adjacent to at least one intrinsic semiconductor region.

[0175] In an NBD, when a p-type region is fully depleted at zero applied bias voltage, no drift current flows in it. Therefore, a thermally generated current carrying species of a non-fully depleted region, normally associated in a conventional pn junction diode with a leakage current does not occur in a fully depleted region. This is because an electric field forces any such electrons to move from the fully depleted p-region to the non-fully depleted region and any such holes to move from the non-fully depleted region to the fully depleted p-region, so that the NBD is conductive at zero bias. In some NBDs, such as a Schottky diode, the contacts to the devices can act as a p-type or n-type region.

[0176] FIG. **8**(*f*) shows multiple NBD **600**s connected in series in order to generate a voltage from a current source, thus supplying power to a load, according to application of the present invention. Therefore, the output voltage of multiple NBDs connected in series to generate a desired output voltage that is multiple times of a single NBD output voltage. NBDs may be connected in parallel to prevent an input current that is greater than the saturation current of any of the serially connected NBDs. By varying the number of NBDs and the manner of connection, variations of desired output voltages or currents may be achieved.

[0177] According to one of the embodiment of present invention, NBD **300** has characteristics as following: (1) low reverse current threshold voltage, (2) rectification of small signals (**3**) rectification of high frequency signals (**4**) conductivity at reverse-bias with appreciable amount of current and (5) reverse output voltage when connected in series with a current source. In addition, in a reverse bias mode, the conducting current of NBD **300** is significantly greater than the leakage current of a conventional diode prior to breakdown. Further, as the built-in voltage raises the output voltage level, NBD **300** indirectly raises the output power. New types of circuits for network switching, digital computing, signaling and waveform shaping (such as clipping and clamping) using NBDs are thus possible. Special diodes using P-type and N-type materials, including Step-Recovery (SRD), PIN and Zener diode types, may be created by modifying the depletion widths which may be determined, for example, by the steps described above. By having multiple p-type and n-type regions of different doping concentrations, other diode characteristics such as the saturation current, leakage current and input/output resistance can be created accordingly.

[0178] According to one embodiment of the present invention, different characteristics of an NBD may be achieved using degenerate materials. Devices using degenerate materials typically exhibit tunneling and negative resistance characteristics within voltage ranges that are less then V_{th}. An NBD with different values of V_s can shift tunneling and negative resistance voltage ranges. FIG. **25**(*a*) shows the current-voltage (IV) characteristics of NBD **300** fabricated using degenerate materials for one value of voltage V_s which is not close to V_{th}. NBD **300** in this configuration shows a larger voltage range in which tunneling occurs and a smaller voltage range in which negative resistance voltage occurs, relative to a conventional pn junction diode utilizing the same degenerate materials. FIG. **25**(*b*) shows the current-voltage (IV) characteristics of NBD **300** utilizing degenerate materials for another value of V_s which is selected to be closer to the threshold voltage V_{th}. NBDs in this configuration have larger minimum forward currents and reduced negative resistance values than a conventional pn junction diode utilizing the same degenerate materials.

[0179] The above examples show that tunneling devices can be provided by NBDs, in accordance with the present invention. Such tunneling devices have desirable characteristics in voltage range, current range and resistance. In one embodiment of present invention, tunneling diodes with different peak voltage, peak currents, valley voltages, or valley currents are also possible. High frequency applications using NBD type tunnel diodes are possible. By varying and mixing different types of materials in NBD devices, such as using a combination of nondegenerate and degenerate materials, new devices or new features may be achieved for existing devices.

[0180] Furthermore, the device in accordance with the present teaching are essentially low voltage devices or devices which can operate with a high signal-to-noise ratio, which makes them well suited to uses as detectors in optical communication systems. Drawings and figures for the EM rectenna of the present invention are provided for explanations and do not represent the relationship of connecting parts in scale, size or position. For example, top views of antenna design shown NBDs and electron input/output channels exposing, actual implementations may include other considerations. The figures used for describing the rectenna embodiments in this invention show semiconductor regions disproportionately large and placed in relationship to antennae for explanatory purposes only. Adjustments on the rectenna system may be required to obtain higher packing density or efficiency. According to the present invention, a highly doped n-type material can act as a contact under some situations. Also, a highly doped p-type material can act as a contact under some situations.

[0181] A fully depleted region in a device at zero applied bias voltage achieves: (1) When a region is fully depleted at zero applied bias voltage, electrons at the contact/semiconductor junction move from fully depleted p-region to the n-region (positive particle (holes) at the contact/semiconductor junction move from fully depleted n-region to the p-region) as a result of the electric field created by the depletion region. In this case, the distance between the fully depleted p-region and external electrons at the contact/semiconductor junction (the distance between the fully depleted n-region and external positive particles (holes) at the contact/semiconductor junction) is zero. As a result, a reverse bias conducting current is achieved. (2) If a fully depleted p-region has external electrons at the contact/semiconductor junction (a fully depleted n-region has external positive particles (holes) at the contact/semiconductor junction), the reverse bias conductivity occurs without a threshold voltage. (3) The electron at the contact semiconductor junction move against the direction of the electric field created by the depletion region (the external positive particles (holes) at the contact semiconductor junction move in the same direction of the electrical field). For example, an NBD conducts at reverse bias without having a breakdown voltage, as in the case of a conventional diode. (4) An NBD partitions conventional PN junction depletion regions into two electric fields.

[0182] The present invention is applicable also to rectenna elements other than dipole rectenna elements. Also, the NBDs of the present invention can be used with practically any antenna types and size to achieve an EM wave to DC power conversion. For an antenna type in which points of maximum voltages or currents may be determined, and for which a gap may be positioned without affecting the operating EM wave frequency response, a gap with an associated NBD may be placed at one or more of such maximum current points. Otherwise, for such an antenna type, a non-gap or gap-less rectenna element may be used by finding maximum voltage points and, at each maximum voltage point along the antenna, place a pair of NBDs to form an input terminal and an output terminal, such as shown in FIGS. **20**(*a*) and **20**(*b*).

[0183] Using the parallel and series connection discussed above for connecting dipole rectennae elements into rectenna arrays, a device can be designed to output any voltage with any antenna type. Further, the rectenna elements on each device may be different, so that a single device may be made to capture various frequencies within a large spectrum of EM waves. Also, different portions of a single device may include rectenna elements of different antenna types and these different portions may be used to perform different functions. The present invention is not limited to a dipole rectenna design, nor to the full wave rectifying rectenna described above. In accordance with the embodiments of the invention, each rectenna element can be of any size, shape or type. In addition, a rectenna array can include more than one type of rectenna elements. For example, in an RFID application, a portion of a rectenna array may be sensitive to one frequency and is used to capture the EM wave to power the RFID circuit, while another portion of the rectenna array may be used for RF transmission and reception. Such an arrangement reduces the required number of parts and creates a more compact design. The single plane construction is not limited to contiguous dipoles but is also applicable to the more usual case of separate dipoles. Although the above configurations mentioned only planar arrays, the present invention is applicable also to non-planar arrays. The invention can also be carried out using discrete parts. Because an NBD of the present invention can rectify very high frequency signals, an EM rectenna of the present invention can be used to detect and rectify very high frequency EM waves efficiently. Thus, detection and power conversion from new frequencies in the EM spectrum can now be achieved where previously were impractical or impossible. This includes generating DC power from the IR and visible light spectra (e.g., solar energy) and fast response photo-sensors for optical sensing and optical networks. The present invention ability to absorb IR frequencies at high efficiencies may be used in such applications as heat shields and cooling devices.

[0184] One of the main obstacles preventing the proliferation of solar energy conversion systems is efficiency. The main issue with efficiency is mostly on the rectifiers and the ways that the rectifiers are connected to the antenna. The embodiments presented in the EM rectenna of the present invention improve the efficiency of converting electromagnetic (EM) waves to DC electricity over the prior art by providing the improvements in: (1) full wave rectification for a broader input signal frequency range (e.g., frequencies at or above infra-red), (2) solving the voltage drop problems in rectifiers[3], (3) providing an output voltage level for higher efficiency by utilizing special rectifiers that output a voltage in response to a small signal, (4) full wave rectification with fewer rectifiers, and (5) providing two sources of input currents, one from the received EM wave and the other one from the electron input channels.

[3] A rectifier (typically a diode) in the prior art requires a bias voltage drop across it before it conducts current. A voltage drop creates power loss, especially for low power signals. The voltage drop issue is important in solar engines, since the voltage drop reduces the voltage supplied to the load.

[0185] While only dipole antennae are used to illustrate the rectennae of the present invention, the methods discussed above may be used to incorporate other antenna types and sizes into the rectenna of the present invention.

[0186] The detailed description above is provided to illustrate the specific embodiments above and is not intended to be limiting. Numerous modifications and variations within the scope of the present invention are possible. The present invention is set forth in the following claims.

We claim:

1. A semiconductor device, comprising a semiconductor region that is fully depleted at zero bias voltage across the semiconductor device.

2. A semiconductor device as in claim **1**, further comprising a contact adjacent the first semiconductor region.

3. A semiconductor device as in claim **1**, further comprising a second semiconductor region forming a pn junction with the first semiconductor region.

4. A semiconductor device as in claim **3**, wherein the semiconductor device conducts a current substantially linearly, when a reverse bias voltage is imposed across the pn-junction.

5. A semiconductor device as in claim **1**, wherein the first semiconductor region is a forced depletion region.

6. A semiconductor device as in claim **1**, wherein the first semiconductor region is a non-forced depletion region.

7. A semiconductor device as in claim **5**, wherein the Fermi level of the semiconductor region is closer to the intrinsic Fermi level of the semiconductor region than the Fermi level of the semiconductor device as zero bias voltage.

8. A semiconductor device as in claim **5**, further comprising a voltage difference equal to voltage difference associated to energy difference between the fermi level of the semiconductor region and the fermi level of the semiconductor device.

9. A method for creating a natural breakdown device, comprising:

providing a semiconductor device with a junction formed by a first semiconductor region and a region of a predetermined material, wherein the first semiconductor region has a width greater than its depletion width;

adjusting a parameter of the region of the predetermined material such that the width of the first semiconductor is equal to or less than the expected depletion width of the first semiconductor region.

10. A method as in claim **9**, wherein the predetermined material comprises a second semiconductor region and the parameter adjusted comprises a dopant concentration.

11. A method as in claim **9**, wherein the predetermined material comprises a conductor material forming a contact and the parameter adjusted comprises a work function of the conductor material.

12. A method as in claim **9**, wherein the parameter adjusted is a dimension of the region of the predetermined material relative the width the dimension of the semiconductor region.

13. A rectenna, comprising:

a conductive portion including a first point and a second point separated by a predetermined distance, such that when the conductive portion is irradiated by an electromagnetic wave, the first point and the second point constitute points of maximum voltage; and

a semiconductor diode adjacent said first point comprising a p-region in contact with the conductive portion and an n-region in contact with the p-region.

14. A rectenna as in claim **13**, further comprising a second semiconductor diode adjacent the second point comprising a p-region in contact with the conductive portion and an n-region in contact with the p-region.

15. A rectenna as in claim **13**, wherein the rectenna is one of a plurality of antennae forming an rectenna array by connecting the p-regions of the second semiconductor diodes of the rectennae to form an input channel and connecting the n-regions of the first semiconductor diodes of the rectennae to form an output channel.

16. A rectenna as in claim **15**, wherein the rectennae in the rectenna array perform two or more functions.

17. A rectenna as in claim **13**, wherein the p-region and the n-region have respectively a first doping concentration and a second doping concentration and wherein the p-region is fully depleted at a zero applied voltage across the p-region and the n-region.

18. A rectenna, comprising:

a first conductive portion;

a second conductive portion separated from the first conductive portion by a gap that is substantially smaller than the combined lengths of the first conductive portion and the second conductive portion; and

a semiconductor device having a p-region and an n-region, the p-region being in contact with the first conductive portion at the gap.

19. A rectenna as in claim **18**, wherein the p-region of the semiconductor device contacts the second conductive portion at the gap.

20. A rectenna as in claim **18**, further comprising a second semiconductor device having a p-region and an n-region, the p-region being in contact with the second portion at the gap wherein the first and second semiconductor device is isolated from each other.

21. A rectenna as in claim **18**, further comprising an n-region in contact with an end portion of the first conductive portion away from the gap.

22. A rectenna as in claim **18**, wherein the p-region and the n-region of the semiconductor device have respectively a first doping concentration and a second doping concentration and wherein the p-region is fully depleted at zero bias voltage across the p-region and the n-region.

23. A rectenna as in claim **18**, further comprising an n-region in contact with one end of the second conductive portion away from the gap.

24. A rectenna as in claim **18**, wherein the rectenna is one of a plurality of rectennae in a rectenna array.

25. A rectenna as in claim **24**, wherein the rectenna array is formed by connecting the n-regions of the semiconductor devices of the rectennae to form an output channel to form an output channel and by connecting the n-regions of those ends of the first conductive portions and the second conductive portions away from the gap to form an input channel.

26. A rectenna as in claim **25**, wherein the rectennae in the rectenna array perform two or more functions.

27. A method to modify the Fermi energy level within a semiconductor region comprising,

providing a semiconductor device with a first semiconductor region of a first conductivity type and at first Fermi level, the first semiconductor region being forced depleted at zero biasing across the semiconductor device;

adjusting the first semiconductor region to a second Fermi level at thermal equilibrium, the second Fermi level being closer to the first semiconductor region intrinsic semiconductor Fermi level than the first Fermi level; providing a sub-threshold bias voltage across the semiconductor device;

28. A method as in claim **27**, wherein a voltage difference in a neutral region adjacent the first semiconductor region is created before the Fermi level of the first semiconductor region moves toward the second Fermi level

29. A method as in claim **27**, wherein the Fermi level of first semiconductor region is adjusted toward the first Fermi level before a voltage difference is created in a neutral region adjacent to the first semiconductor region.

30. A natural breakdown diode, comprising:

a first semiconductor region of a first conductivity type having a first doping concentration;

a second semiconductor region of a second conductivity type opposite first conductivity type having a second doping concentration; and

an ohmic contact adjacent the first semiconductor region; wherein the first semiconductor region is fully depleted at zero bias voltage across the first semiconductor region and the second semiconductor region.

31. A natural breakdown diode as in claim **30**, further comprises a third semiconductor region of first semiconductor type between the first semiconductor region and the ohmic contact, and wherein the first semiconductor region and the third semiconductor region are fully depleted at a zero bias voltage across the first semiconductor region, second semiconductor region and third semiconductor region.

32. A natural breakdown diode as in claim **31**, wherein the natural breakdown diode is conducting when a negative bias voltage is imposed across the first semiconductor region and second semiconductor region.

33. A natural breakdown diode as in claim **30**, wherein the natural breakdown diode provides a substantially constant output when a bias voltage between zero volts and a predetermined voltage V_s is imposed across the first semiconductor region and second semiconductor region.

34. A natural breakdown diode as in claim **30**, further comprising a third semiconductor region of second semiconductor type between the second semiconductor region and first semiconductor region.

35. A natural breakdown diode as in claim **30**, wherein the n-region is fully depleted at zero bias voltage across the p-region and the n-region.

36. A method for providing a natural breakdown diode, comprising:

(1) determining the doping concentrations of a p-region and an n-region of a conventional pn junction diode;

(2) selecting a forward-bias voltage V_s less than the conventional pn junction diode threshold voltage V_{th} and which is less than a predetermined maximum bias voltage for the diode; and

(3) calculating the depletion width w_p of the conventional pn junction diode when a bias voltage of V_s is imposed across the p-region and the n-region of the conventional pn junction diode;

(4) providing w_p as the width of a p-region the natural breakdown diode.

37. A method as in claim **36**, further comprising calculating the electric field across a pn junction in the natural breakdown diode and adjusting the doping concentrations of the p-region and an n-region of the natural breakdown diode such that natural breakdown diode is conducting at zero bias voltage.

38. A method for constructing a rectenna, comprising:

providing a conductive portion including a first point and a second point separated by a predetermined distance, such that when the conductive portion is irradiated by an electromagnetic wave, the first point and the second point constitute points of maximum voltage; and

providing a semiconductor diode adjacent said first point comprising a p-region in contact with the conductive portion and an n-region in contact with the p-region.

39. A method as in claim **38**, further comprising providing a second semiconductor diode adjacent the first point comprising a first n-region in contact with the conductive portion and a p-region in contact with the n-region.

40. A method as in claim **38**, further comprising providing a second semiconductor diode adjacent the second point comprising a p-region in contact with the conductive portion and an n-region in contact with the p-region.

41. A method as in claim **38**, further comprising:

providing a plurality of rectennae arranged in a rectenna array;

connecting the p-regions of the second semiconductor diodes of the rectennae to form an input channel; and

connecting the n-regions of the first semiconductor diodes of the rectennae to form an output channel.

42. A method as in claim **41**, further comprising providing two or more functions to be performed by the rectennae in the rectenna array.

43. A method as in claim **42**, further comprising providing the p-region and the n-region to have respectively a first doping concentration and a second doping concentration and providing that the p-region is fully depleted at a zero bias voltage across the p-region and the n-region.

44. A method as in claim **41**, further comprising providing a plurality of the rectenna arrays by connecting the input channels of the rectenna arrays together; and connecting the output channels of rectenna arrays together.

45. A method as in claim **36**, further comprising:

providing an increasing bias voltage across the natural breakdown diode;

applying a decreasing bias voltage across the natural breakdown diode, such that a charge builds up in a neutral region of the n-region.

46. A method as in claim **27**, wherein a depletion region associated with the first semiconductor region has the same width during the application of the decreasing bias voltage.

47. A method for providing a rectenna, comprising:

providing a first conductive portion;

providing a second conductive portion separated from the first conductive portion by a gap that is substantially smaller than the combined lengths of the first conductive portion and the second conductive portion; and

providing a semiconductor device having a p-region and an n-region, the p-region being in contact with both the first conductive portion and the second conductive portion at the gap.

48. A method as in claim **47**, wherein the p-region of the semiconductor device contacts the second conductive portion at the gap.

49. A method as in claim **47**, further comprising a second semiconductor device having a p-region and an n-region, the p-region being in contact with the second portion at the gap.

50. A method as in claim **49**, wherein the first and second semiconductor device is isolated from each other.

51. A method as in claim **50**, further comprising providing an n-region in contact with an end portion of the fist conductive portion away from the gap.

52. A method as in claim **51**, further comprising providing the p-region and the n-region of the semiconductor device to have respectively a first doping concentration and a second doping concentration and providing that the p-region is fully depleted at zero bias voltage across the p-region and the n-region.

53. A method as in claim **47**, further comprising an n-region in contact with one end of the first conductive portion away from the gap.

54. A method as in claim **47**, further comprising providing the rectenna to be one of a plurality of rectennae in a rectenna array.

55. A method as in claim **51**, wherein providing the rectenna to be one of a plurality of rectennae in a rectenna array comprises:

providing the plurality of rectennae;

connecting the n-regions of the semiconductor devices of the rectennae to form an output channel to form an output channel; and

connecting the n-regions of those ends of the first conductive portions and the second conductive portions away from the gap to form an input channel.

56. A method in claim **55**, wherein the rectenna is one of a plurality of rectennae arranged in a plurality of rectenna arrays, the method further comprising:

providing a plurality of the rectennae arrays;

connecting the input channels of the rectenna arrays together; and

connecting the output channels of rectenna arrays together.

57. A method as in claim **56**, wherein the rectenna array is one of a plurality of rectenna arrays, the method further comprising:

providing the plurality of rectenna arrays;

connecting the p-regions of the second semiconductor diodes of the rectennae to form an input channel; and

connecting the n-regions of the first semiconductor diodes of the rectennae to form an output channel.

58. A method as in claim **57**, wherein the rectennae in the rectenna array perform two or more functions.

59. A method as in claim **57**, wherein the first conductive portion and the second conductive portion are provided a predetermined combined length which is substantially half of an odd multiple of a wavelength of the electromagnetic wave.

60. A method for providing a natural breakdown diode, comprising:

providing a first semiconductor region of a first conductivity type having a first doping concentration and providing a second semiconductor region of a second conductivity type opposite first conductivity having a second doping concentration;

providing a width for the first semiconductor region that is less than or equal to a first semiconductor region depletion width of a conventional diode having a first semiconductor region of the first doping concentration and a second doping concentration; and

providing an ohmic contact adjacent the first semiconductor region.

61. A method as in claim **60**, further comprises providing a third semiconductor region of same conductivity type as first conductivity type between the first semiconductor region and the ohmic contact having a third doping concentration, wherein the first semiconductor region and the third semiconductor region are fully depleted at zero bias voltage across first semiconductor region, the second semiconductor region and the third semiconductor region.

62. A method as in claim **60**, further comprising providing the diode to operate at break down mode when a zero bias voltage is imposed across the first semiconductor region and the second semiconductor region.

63. A method as in claim **60**, further comprising providing the diode a constant voltage output when a bias voltage between zero volts and a predetermined voltage V_s is imposed across the p-region and the n-region.

64. A method as in claim **60**, further comprising providing a third semiconductor region of same conductivity type as second conductivity type between the second semiconductor region and the first semiconductor region.

65. A method as in claim **60**, further providing the second semiconductor region to be fully depleted at zero bias voltage across the first semiconductor region and the second semiconductor region.

66. A method for collecting electromagnetic radiation power, comprising:

providing a plurality of antennae elements forming an array having an input channel and an output channel, wherein each antenna element comprising:

a conductive portion including a first point and a second point separated by a predetermined distance, such that when the conductive portion is irradiated by an electromagnetic wave, the first point and the second point constitute points of maximum voltage; and

a semiconductor diode adjacent said first point comprising a p-region in contact with the conductive portion and an n-region in contact with the p-region; and

providing a power storage device for receiving a current from the antenna array, the power storage device being connected across the input channel and the output channel.

67. A method as in claim **66**, wherein the antenna element further comprises a second semiconductor diode adjacent the first point comprising a first n-region in contact with the conductive portion and a p-region in contact with the n-region.

68. A method as in claim **66**, wherein the p-regions of the second semiconductor diodes of the antenna array to form the input channel and connecting the n-regions of the first semiconductor diodes of the antenna array to form the output channel.

69. A method as in claim **66**, wherein the antenna element further comprises a second semiconductor diode adjacent the second point comprising a p-region in contact with the conductive portion and an n-region in contact with the p-region.

70. An antenna element, comprising:

an input terminal;

an output terminal;

a first conductor and a second conductor separated by a gap each conductor having a first end and a second end, respectively, wherein the second ends of each conductor is located at the gap, the first end of each conductor is connected to the input terminal, and the total lengths of the first conductor, the second conductor and the gap equal to a predetermined value related to the wavelength an electromagnetic radiation to be received by the antenna element; and

a first rectifier coupled between the second end of the first conductor and the output terminal.

71. An antenna element as in claim **70**, wherein a second rectifier is coupled between the second end of the second conductor and the output terminal.

72. An antenna element as in claim **70** wherein, when an electromagnetic radiation is incident to the antenna element, a first electrical current is induced in the antenna element due to a voltage difference along each conductor, and a second electrical current is induced in the antenna element by a magnetic field of the electromagnetic radiation.

73. An antenna element as in claim **70**, wherein the rectifier is conducting at a zero bias voltage

74. An antenna array comprising a plurality of antenna elements interconnected by semiconductor devices.

75. A rectifier, comprising:

a conductor having a first end and a second end; and

a semiconductor device connected to the first end of the conductor, wherein the semiconductor device including a semiconductor region that is fully depleted at zero bias voltage.

76. A rectifier as in claim **75**, further comprising a second semiconductor device, connected to the conductor that includes a semiconductor region that is fully depleted at zero bias voltage.

77. A rectifier as in claim **75**, further including a second conductor separated from the first conductor by a gap.

78. In a semiconductor device, a method for transferring a potential from a first built-in potential to a second build-in potential, comprising:

providing a first forced fully-depleted semiconductor region of a first conductivity type;

providing a second semiconductor region of a second conductivity type different from the first conductivity type, the second conductor region being adjacent the first semiconductor region and having a first built-in potential;

providing s second built-in potential adjacent the second semiconductor region opposite first built-in potential;

providing an increasing forward bias relative to first semiconductor region; and

providing a decreasing forward bias relative to first semiconductor region, such that first built-in potential is decreased and second built-in potential is increased.

79. A method as in claim **78**, wherein the threshold voltage of the semiconductor device is less than the threshold voltage of a conventional pn junction.

80. A method for modifying a built-in potential of a semiconductor device, comprising:

creating in a semiconductor substrate a pn-junction of a predetermined built-in potential; and

providing in the semiconductor substrate at a predetermined distance from the pn-junction a semiconductor region that is fully depleted when the semiconductor device is at zero bias voltage.

81. A method as in claim **80**, wherein the fully depleted semiconductor region is provided by an ohmic contact to a suitably doped semiconductor region.

82. A method as in claim **80**, wherein the fully depleted semiconductor region is provided by a Schottky contact to a suitably doped semiconductor region.

83. A method for signal power storage in a semiconductor, comprising:

providing a semiconductor device including a semiconductor region that is fully depleted at zero bias voltage; and

applying a sub-threshold bias voltage across the semiconductor region.

84. An Ohmic contact within a semiconductor device comprising:

a first semiconductor region of a first conductivity type;

a second semiconductor region of a second conductivity type opposite first conductivity type;

a contact adjacent to the first semiconductor region, the second semiconductor region adjacent to the first semiconductor region opposite a contact, the first semiconductor region being sufficiently thin to create an ohmic coupling between the second semiconductor region and the contact.

85. A method for providing an Ohmic contact within a semiconductor device, comprising:

providing a first semiconductor region of a first conductivity type; and

providing a second semiconductor region of a second conductivity type opposite the first conductivity type adjacent to the first semiconductor region; and

providing a contact adjacent to first semiconductor region opposite second semiconductor region wherein the first semiconductor region is sufficiently thin to create an ohmic coupling between the second semiconductor region and the contact.

86. A method to fully rectify an input signal with a single semiconductor device, comprising,

providing a first semiconductor region of a first conductivity type;

providing a second semiconductor region of a second conductivity type opposite first conductivity type adjacent to the first semiconductor region;

providing an ohmic contact adjacent to first semiconductor region opposite the second semiconductor region;

providing the first semiconductor region is forced depleted at zero bias voltage across the contact, the first semiconductor region and the second semiconductor region; and

providing the input signal.

* * * * *

US 20090102296A1

(19) **United States**

(12) **Patent Application Publication**

Greene et al.

(10) **Pub. No.: US 2009/0102296 A1**

(43) **Pub. Date: Apr. 23, 2009**

(54) **POWERING CELL PHONES AND SIMILAR DEVICES USING RF ENERGY HARVESTING**

(75) Inventors: **Charles E. Greene**, Cabot, PA (US); **Daniel W. Harrist**, Carnegie, PA (US); **Michael Thomas McElhinny**, Port Vue, PA (US)

Correspondence Address:
Ansel M. Schwartz
Attorney at Law
Suite 304, 201 N. Craig Street
Pittsburgh, PA 15213 (US)

(73) Assignee: **Powercast Corporation**

(21) Appl. No.: **12/005,696**

(22) Filed: **Dec. 28, 2007**

Related U.S. Application Data

(60) Provisional application No. 60/878,816, filed on Jan. 5, 2007.

Publication Classification

(51) **Int. Cl.**
H02J 17/00 (2006.01)

(52) **U.S. Cl.** **307/149**

(57) **ABSTRACT**

A device for receiving wireless power includes a point of reception, wherein the point of reception is positionable in at least a first position and a second position. A method for receiving wireless power. The method includes the steps of positioning a point of reception in contact with a housing to a first position. There is the step of receiving wireless power at the point of reception and providing it to a power harvester in the housing. There is the step of converting the wireless power to usable DC with the power harvester. There is the step of providing the usable DC to core components in the housing. There is the step of using the DC by the core components. There is the step of repositioning the point of reception to a second position. There is the step of receiving wireless power at the point of reception at the second position and providing it to the power harvester. There is the step of converting the wireless power received by the point of reception in the second position to usable DC with the power harvester. There is the step of providing the usable DC to the core components in the housing. There is the step of using the DC by the core components.

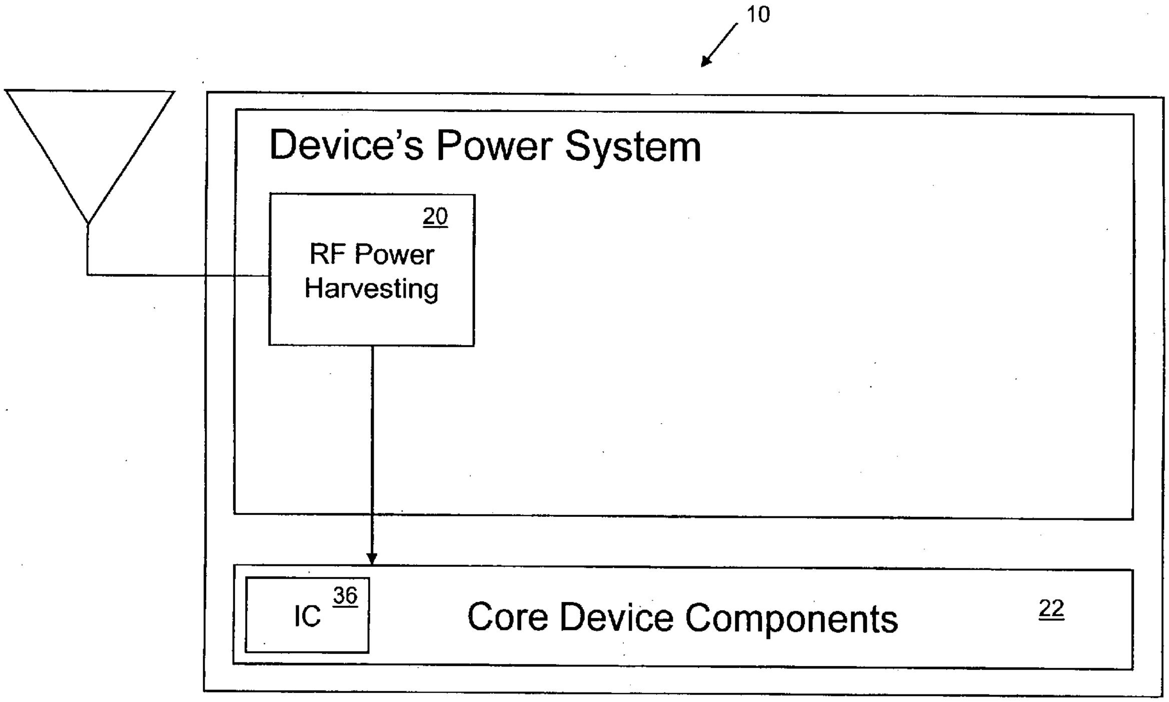

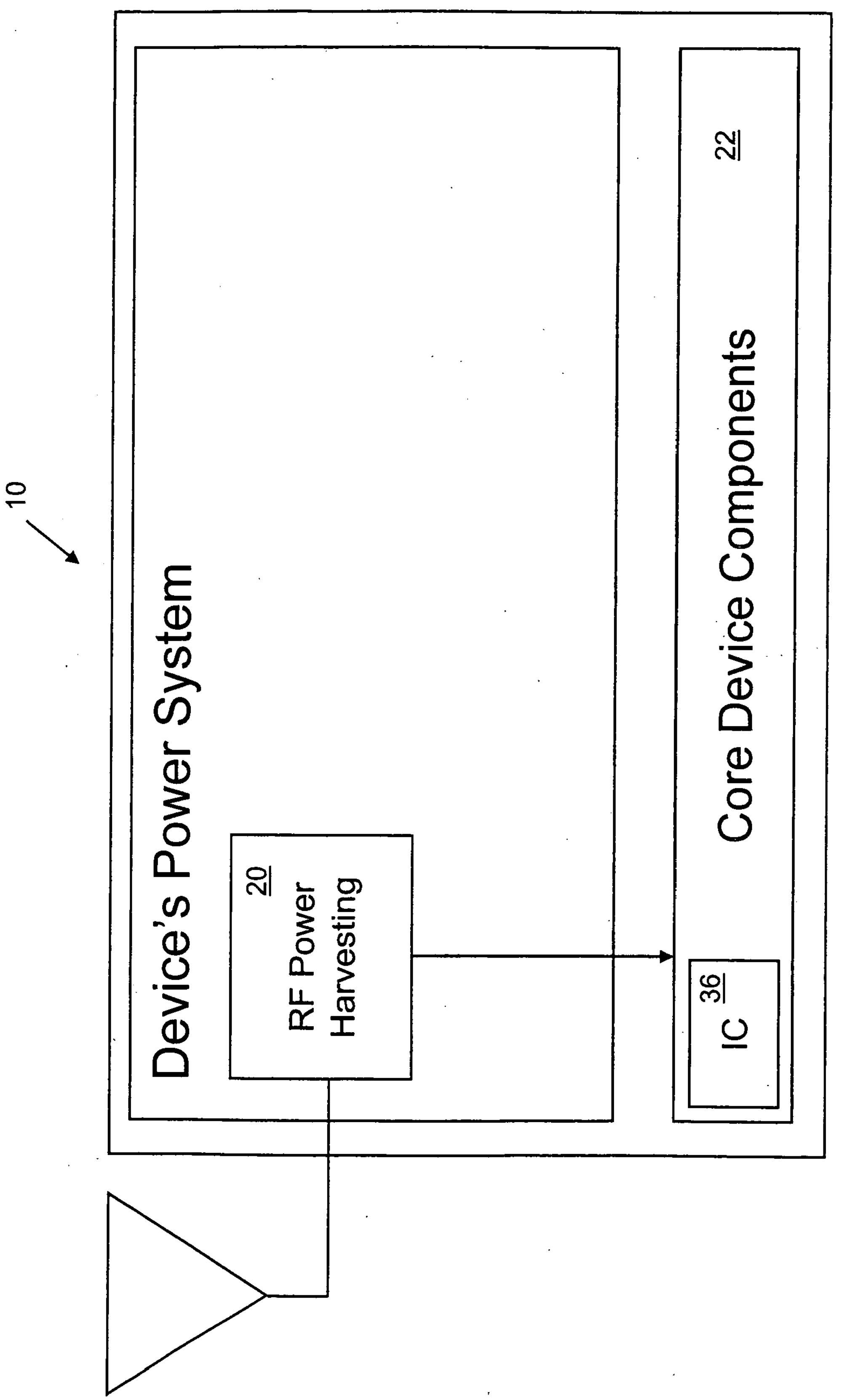
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
IC
36

Fig. 1

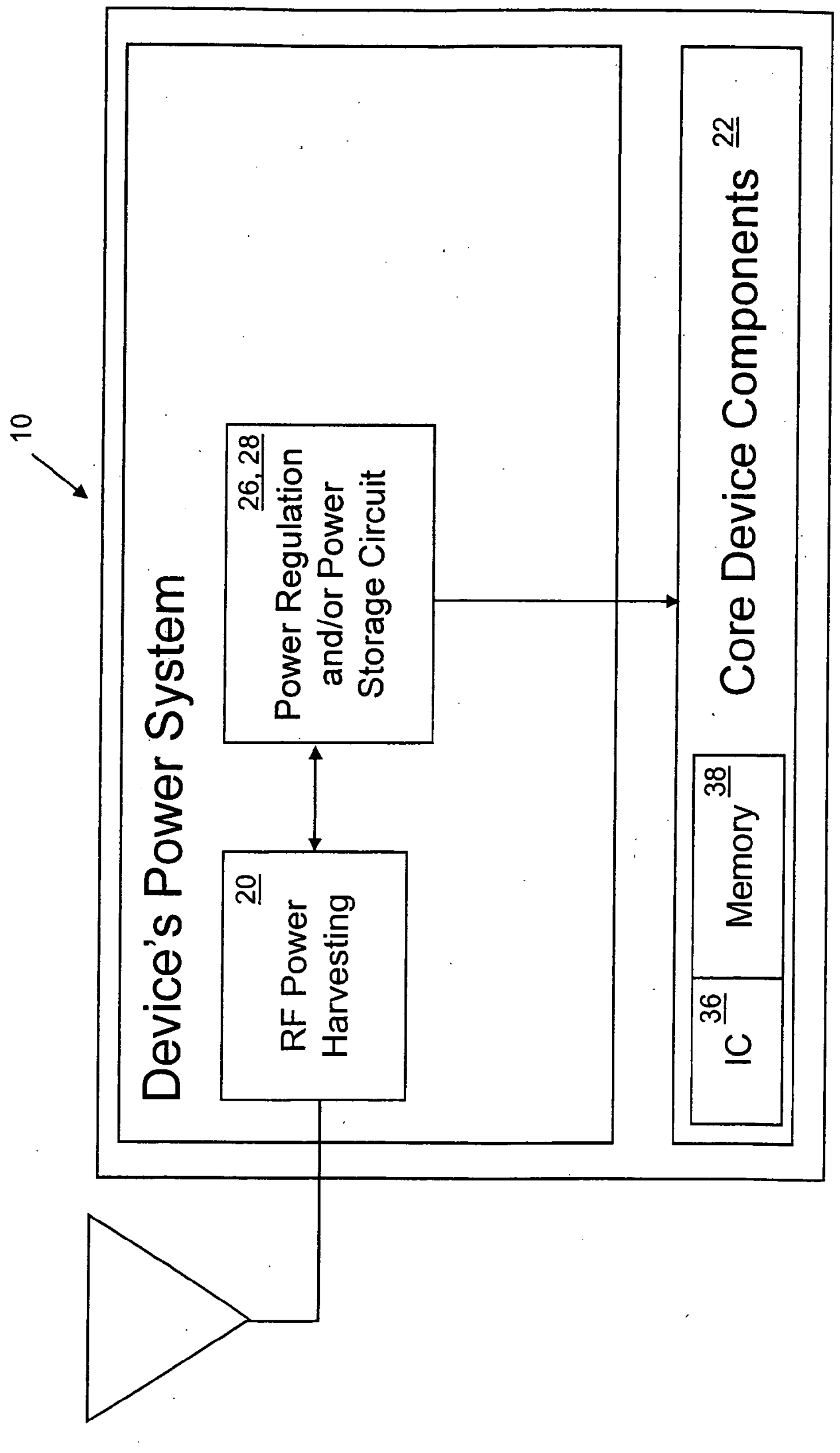
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Core Device Components
22
IC
36
Memory
38

Fig. 2

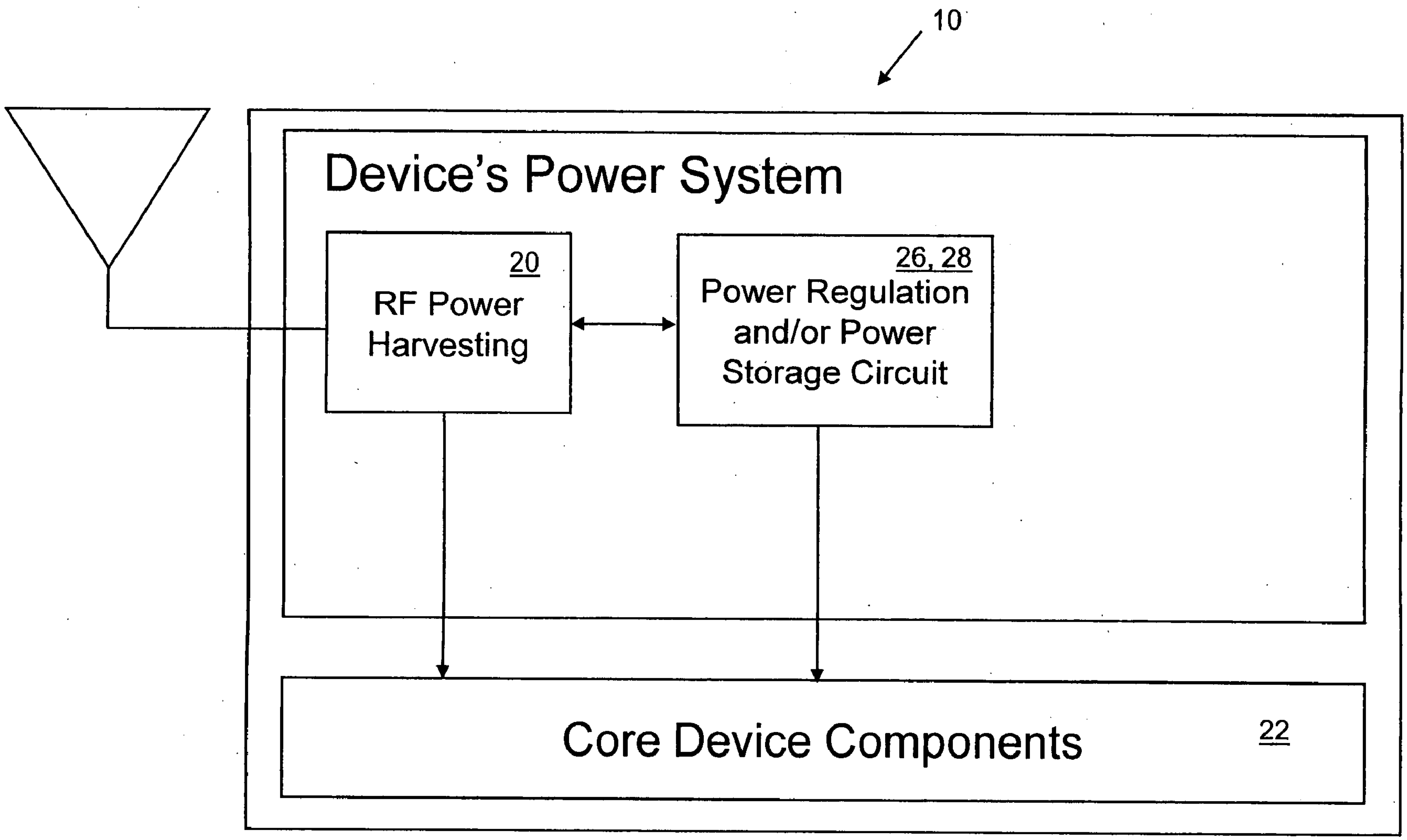
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Core Device Components
22

Fig. 3

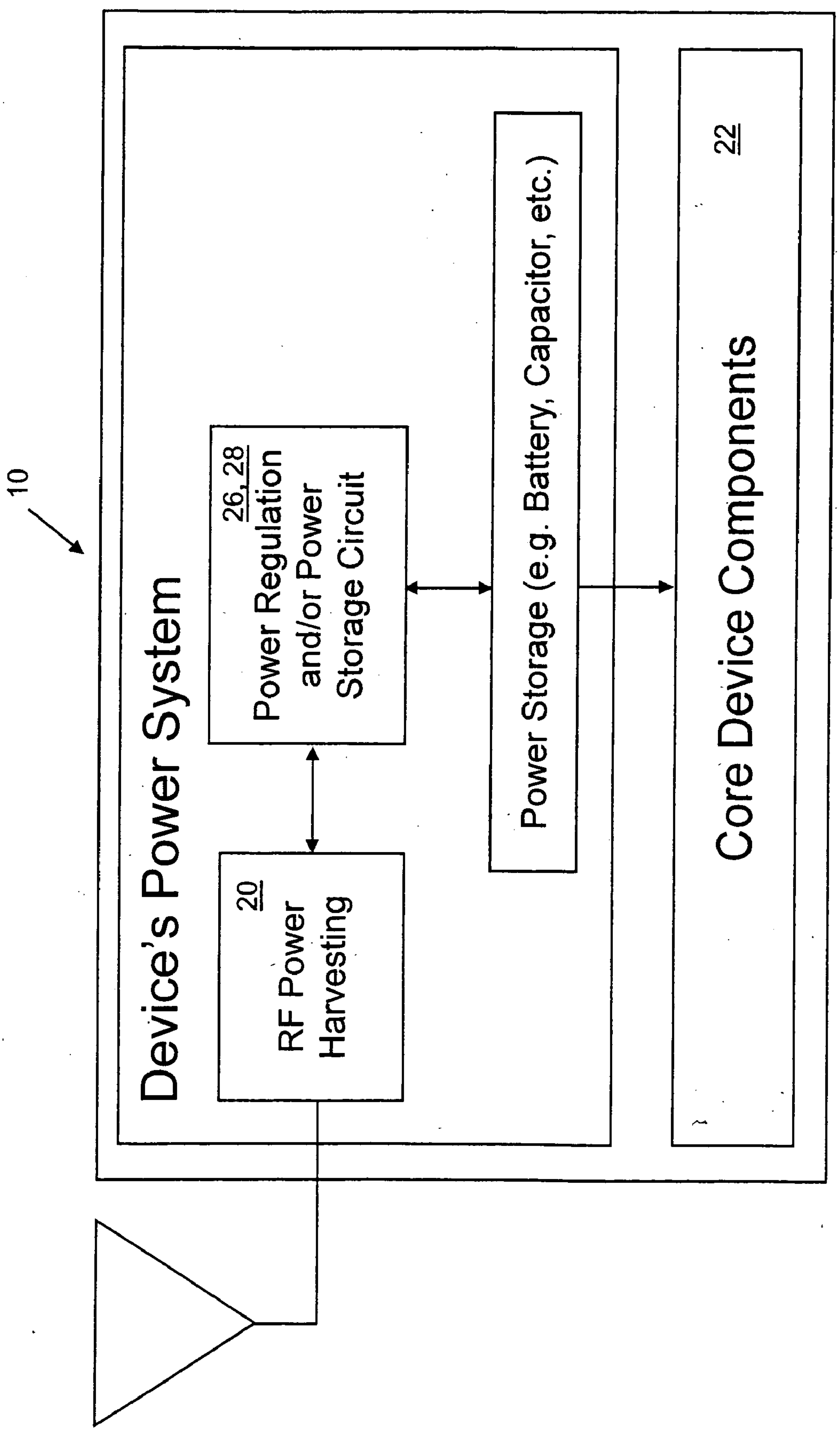
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 4

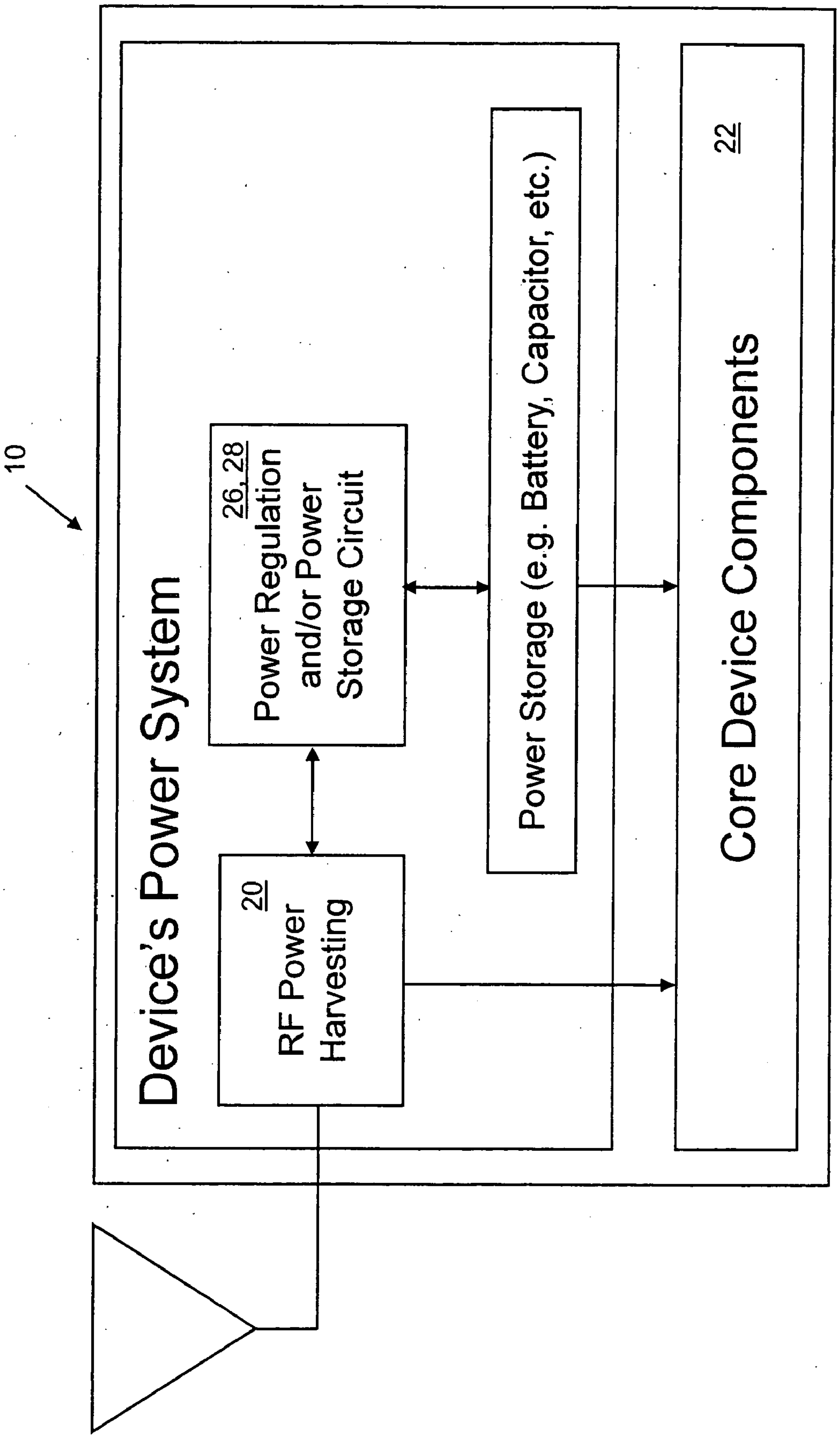
10
Device's Power System
26, 28
Power Regulation and/or Power Storage Circuit
20
RF Power Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 5

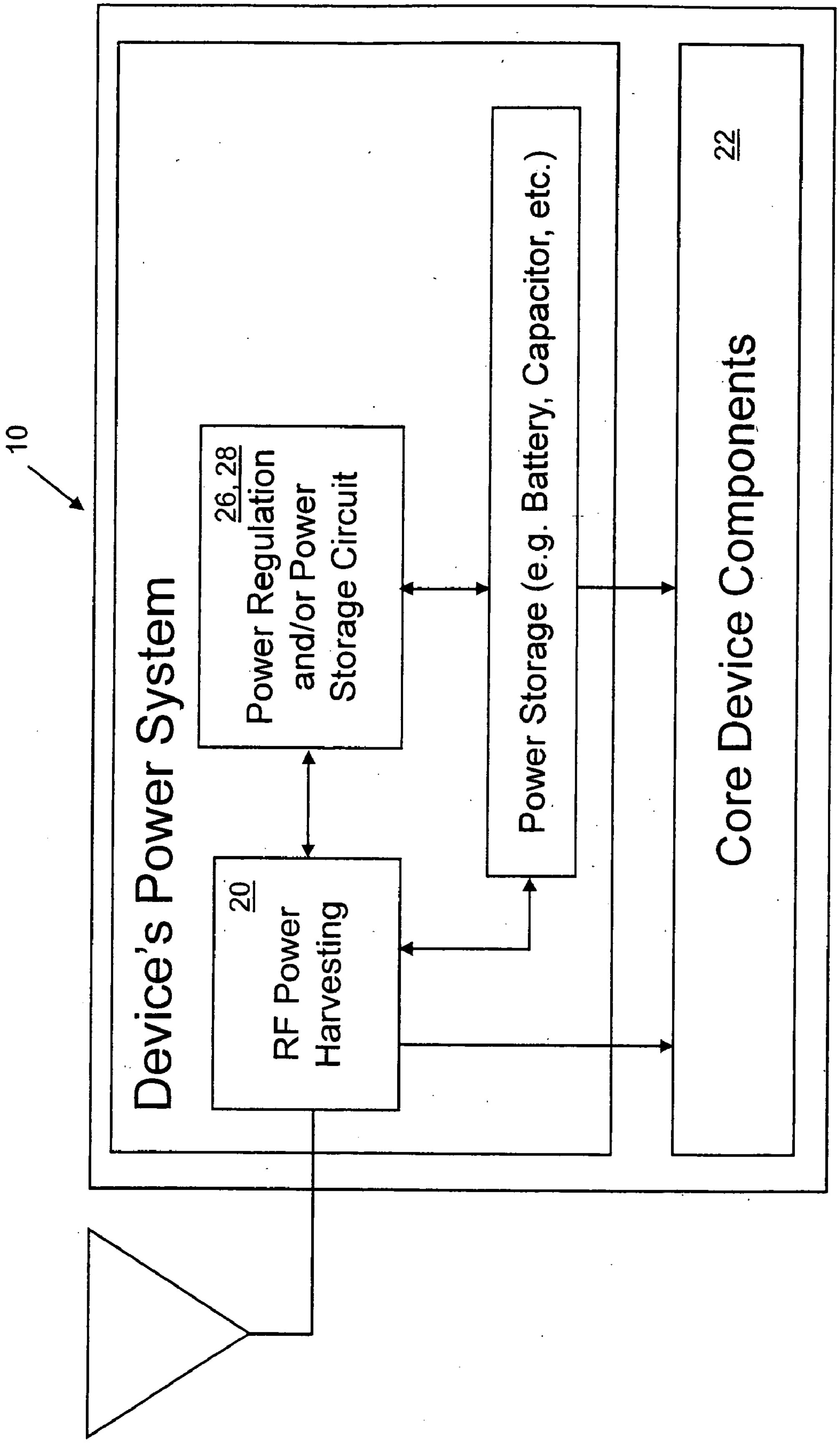
10
Device's Power System
26, 28
Power Regulation and/or Power Storage Circuit
20
RF Power Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 6

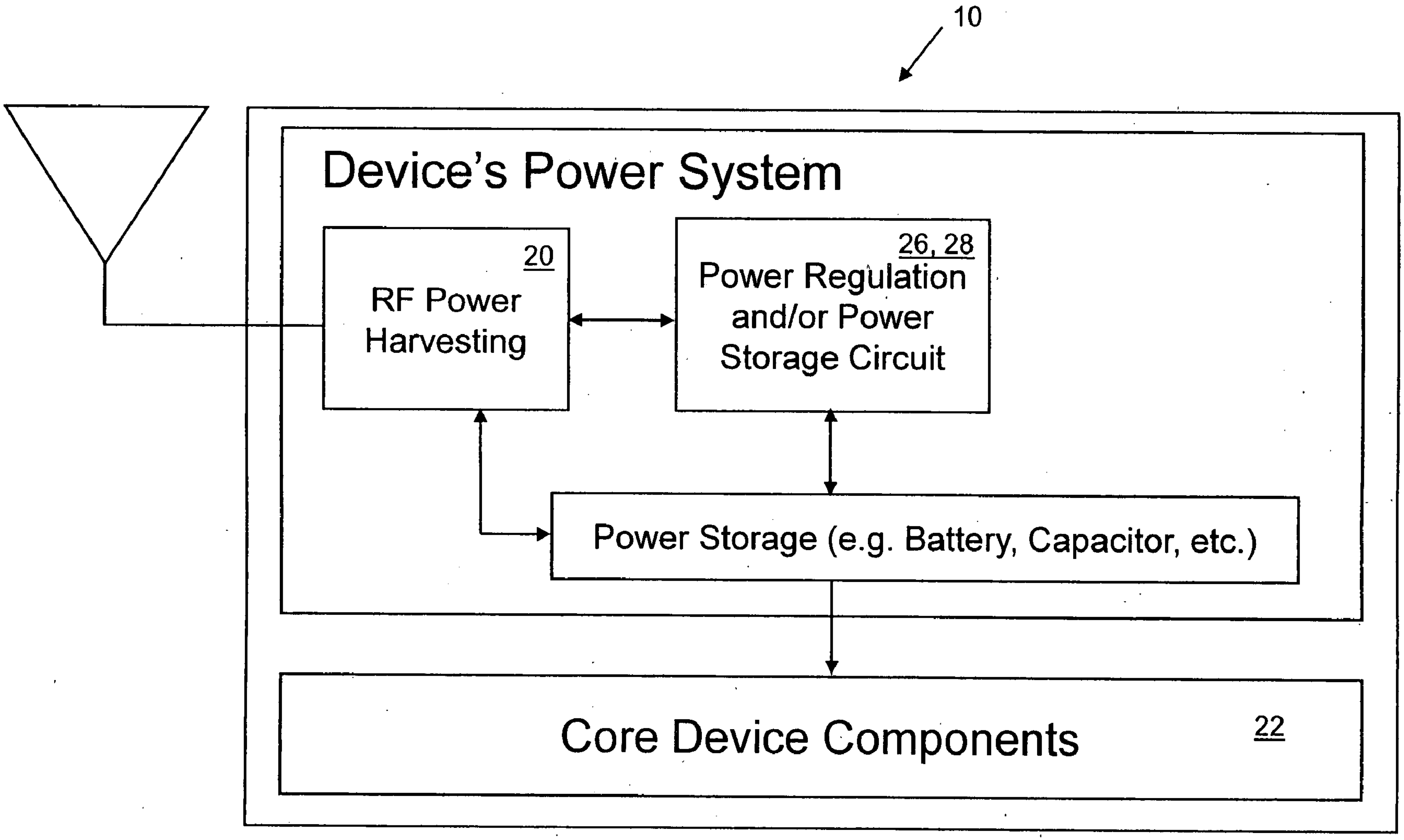
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 7

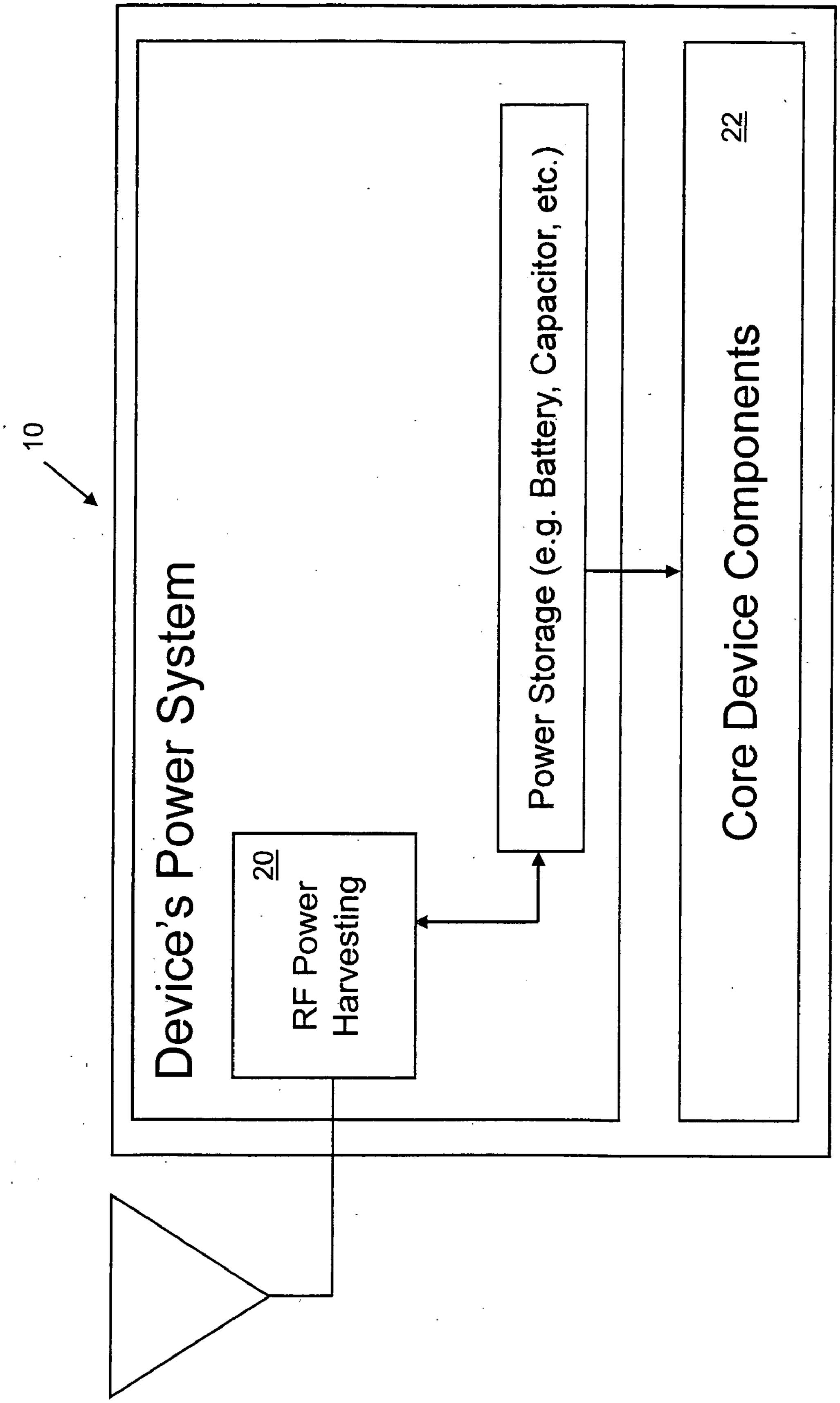
10
Device's Power System
20
RF Power Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 8

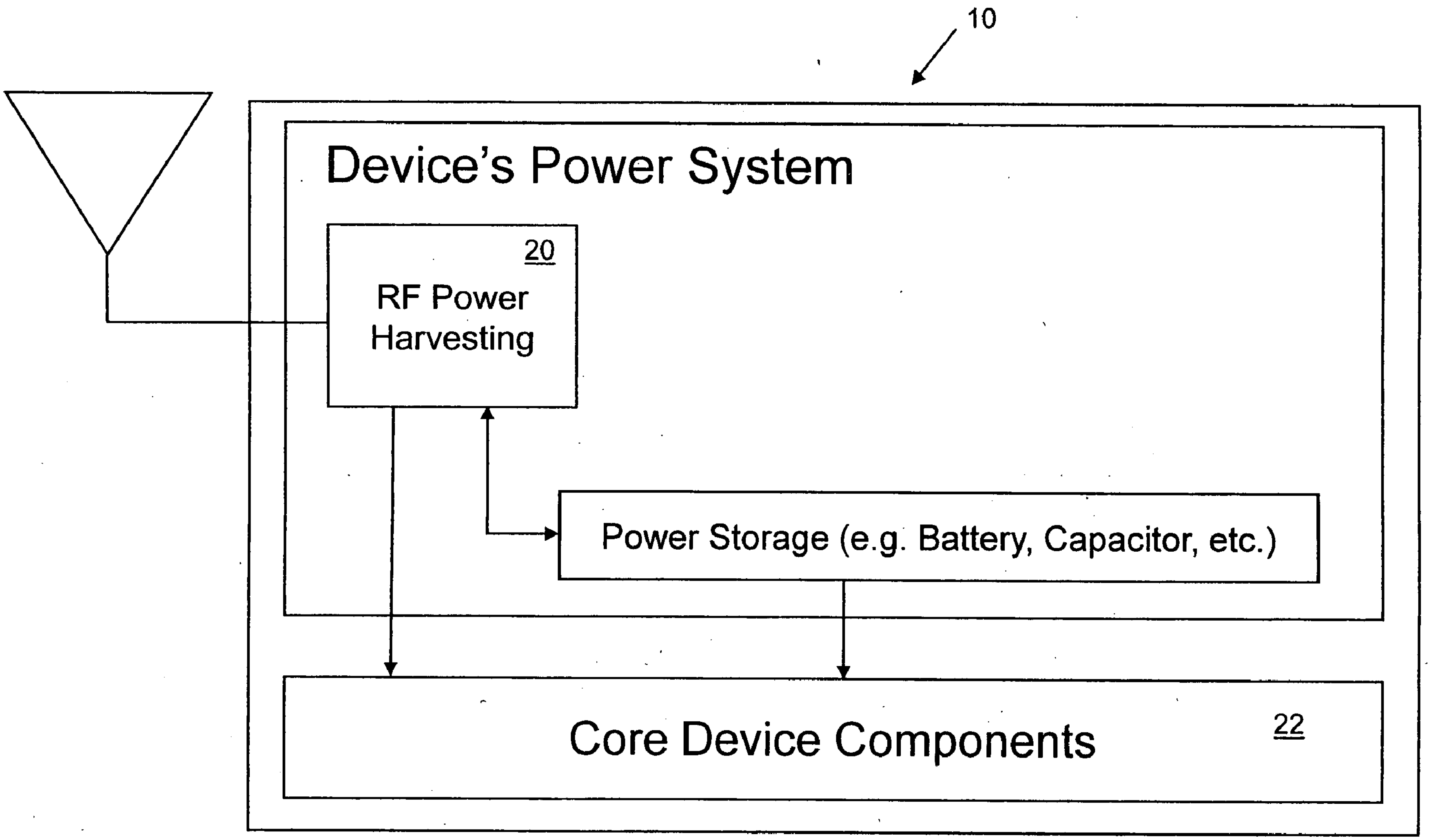
10
Device's Power System
20
RF Power
Harvesting
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 9

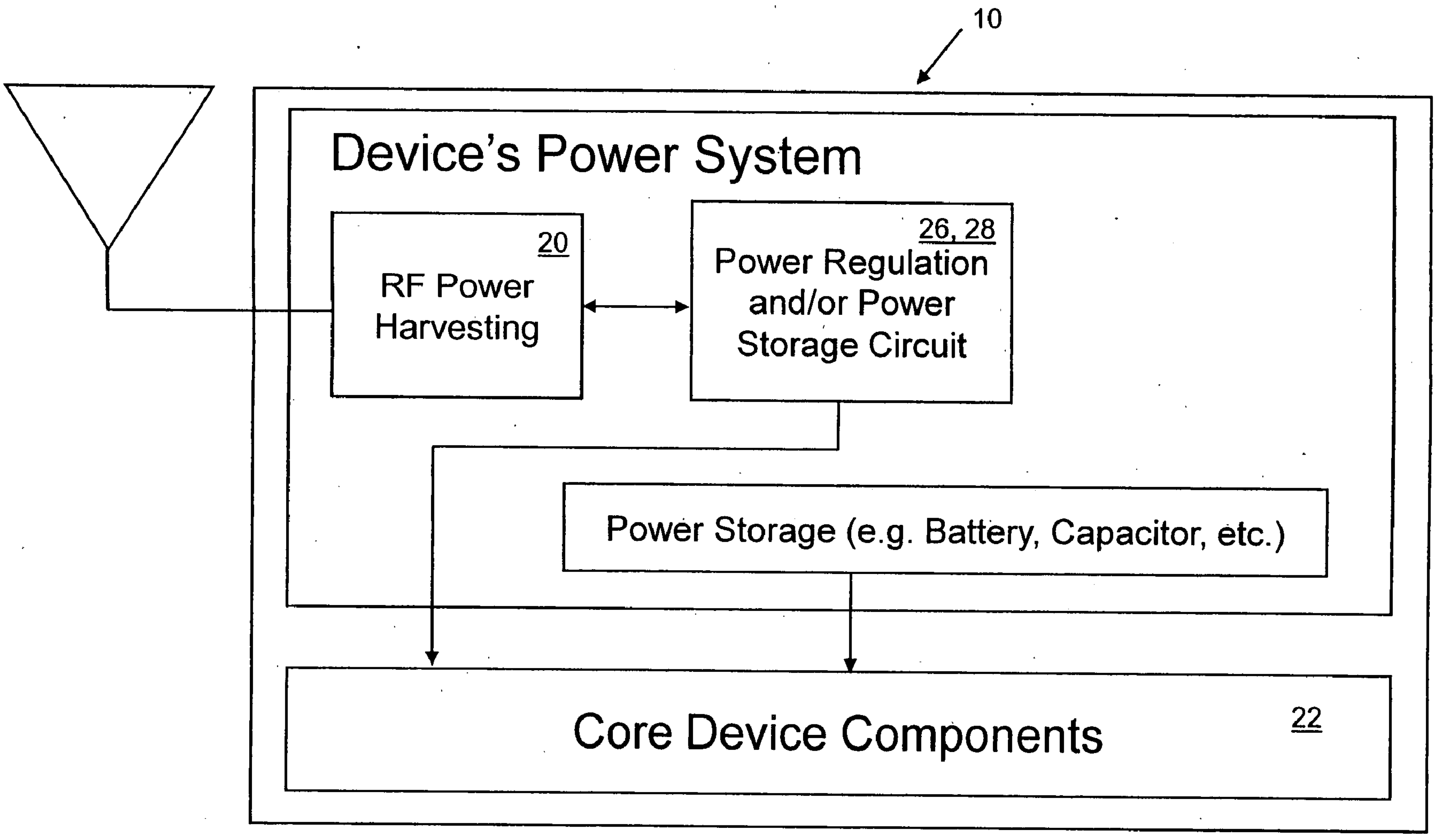
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 10

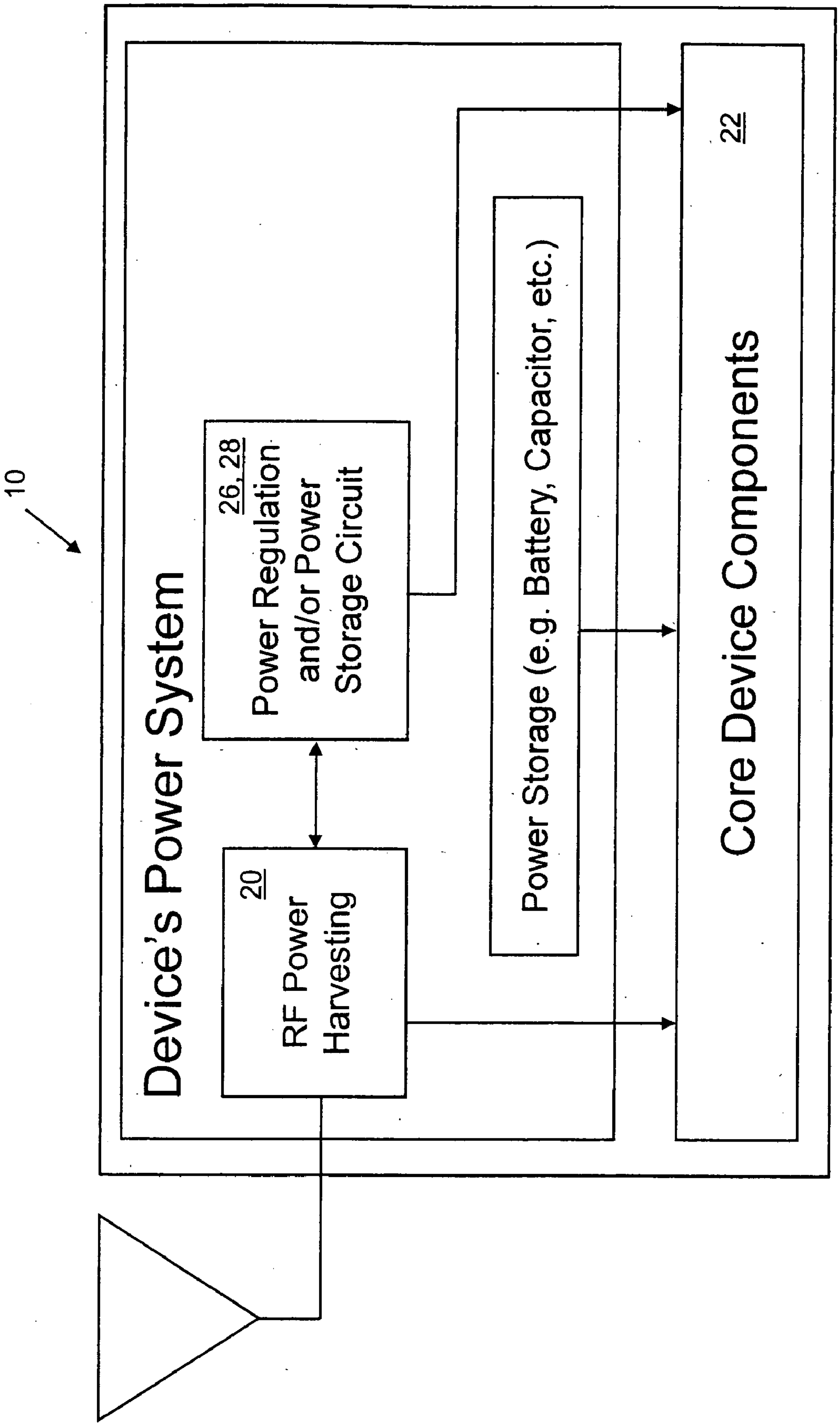
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 11

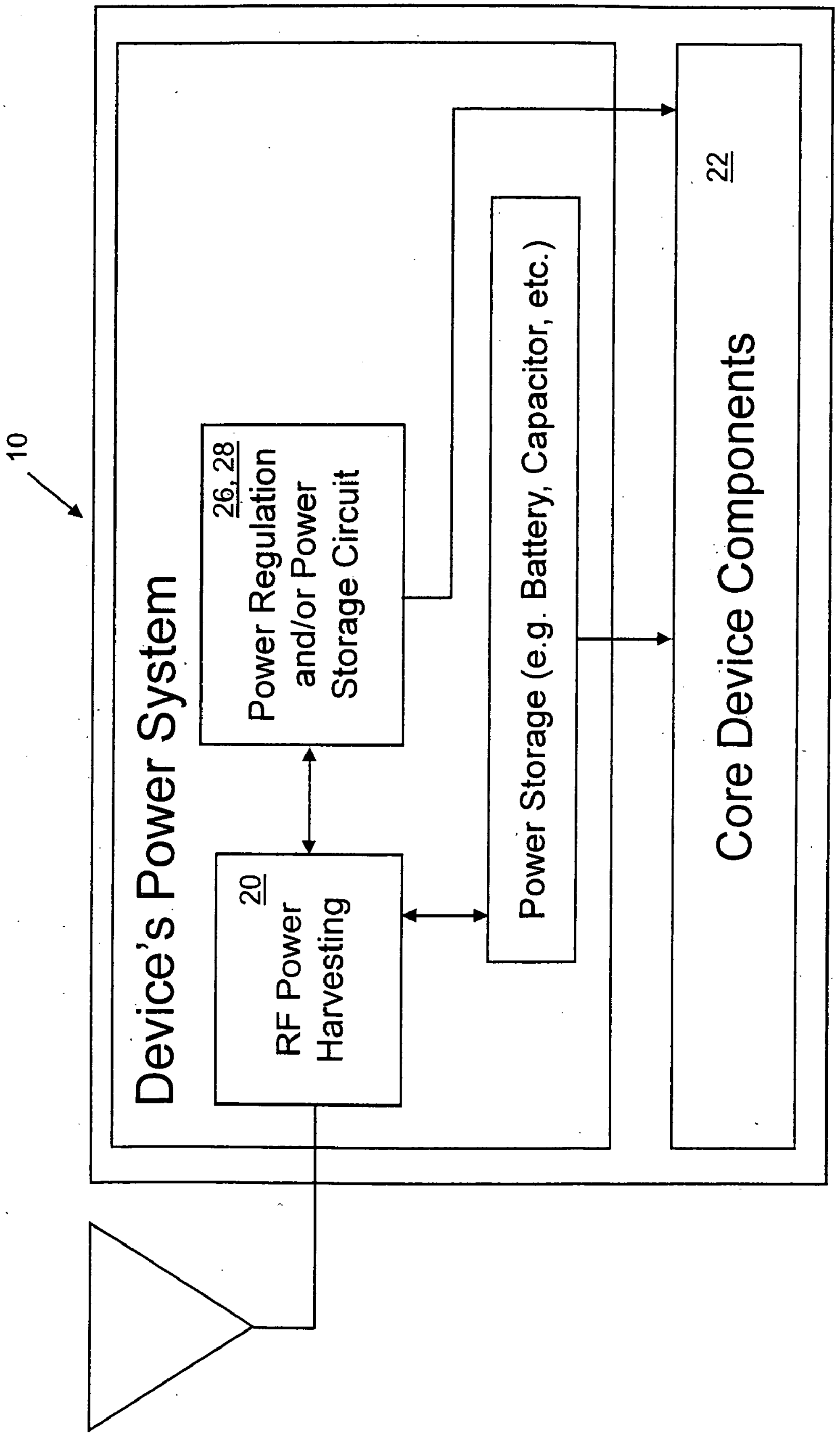
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 12

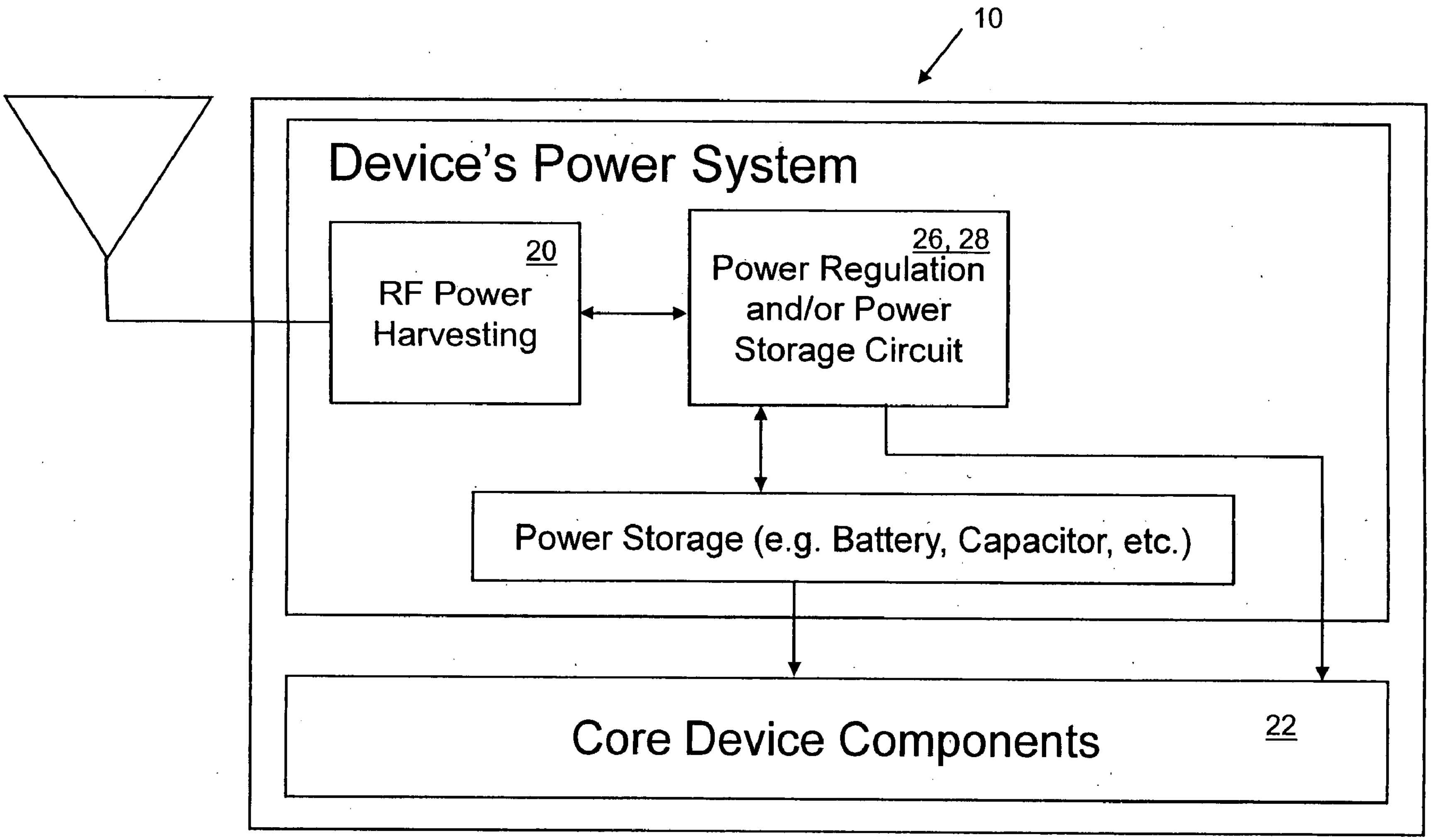
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 13

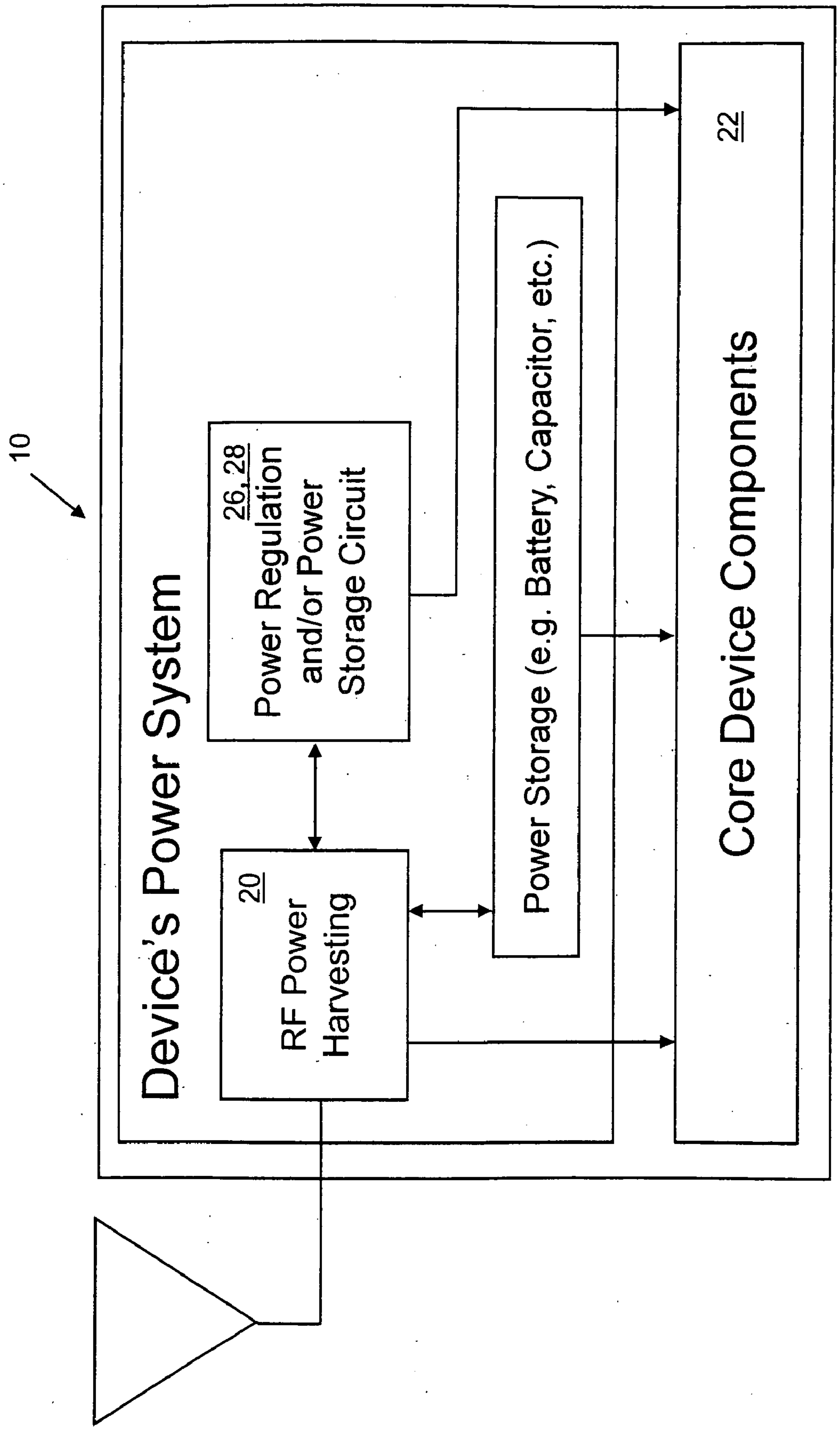
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 14

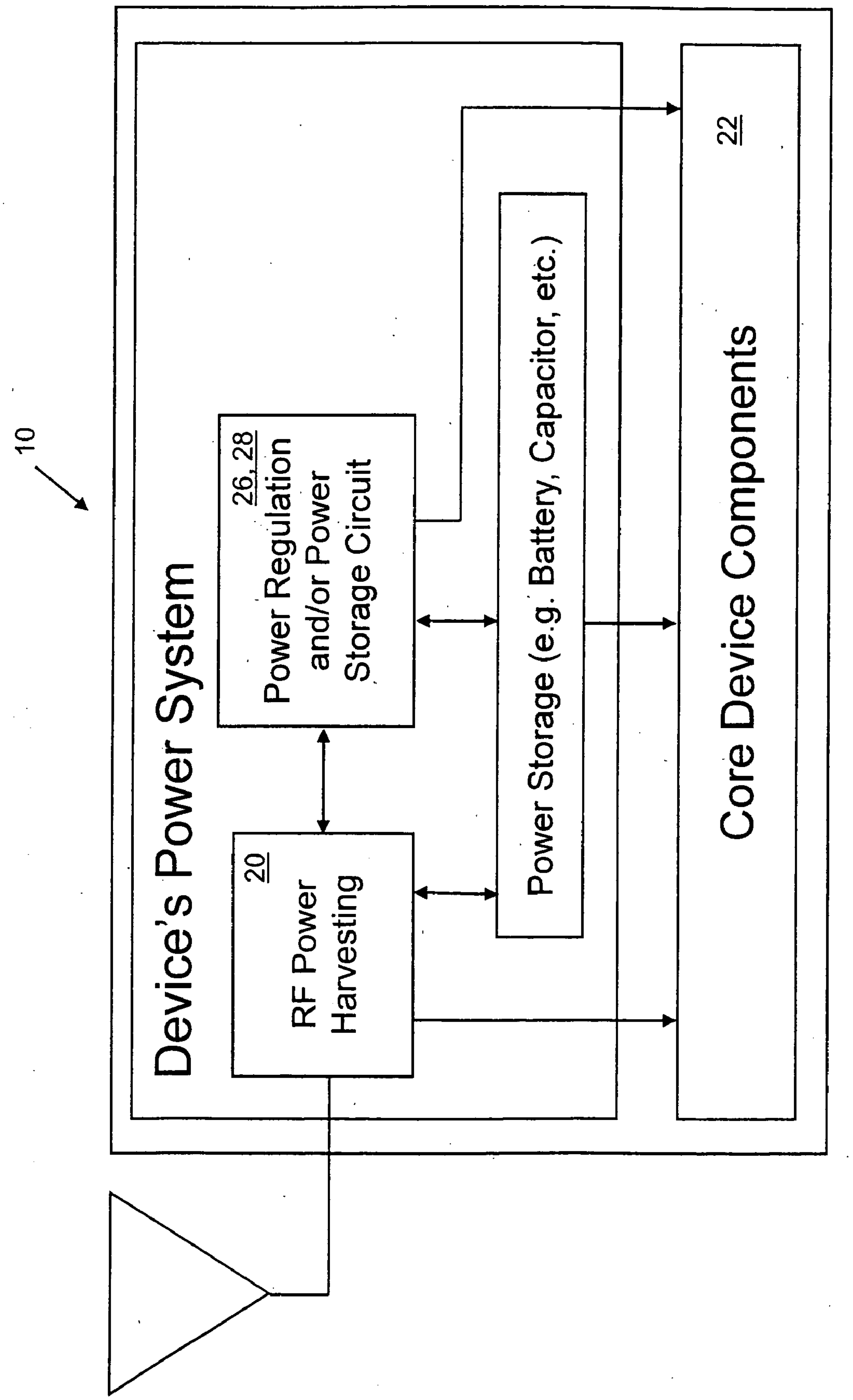
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 15

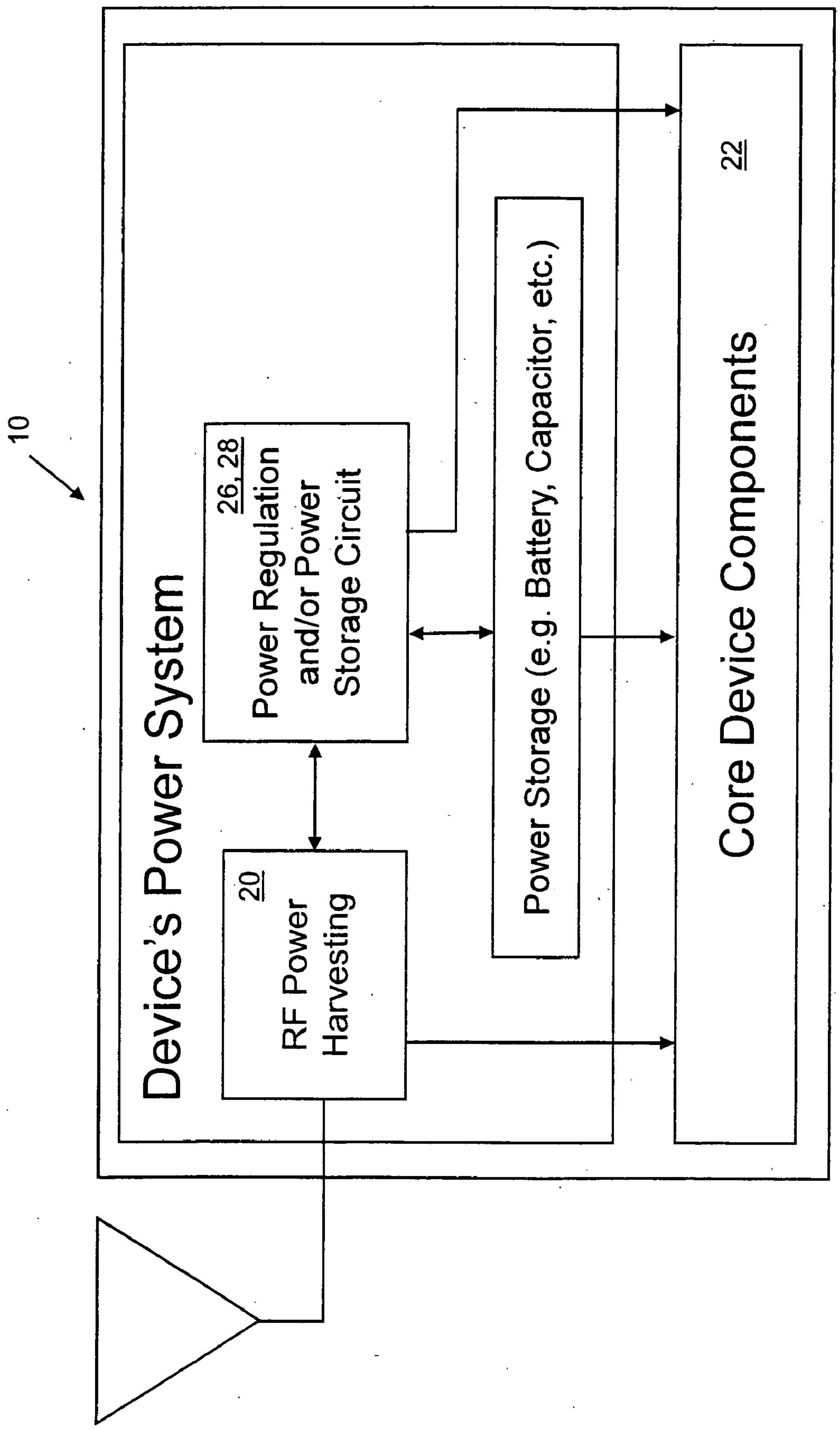
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 16

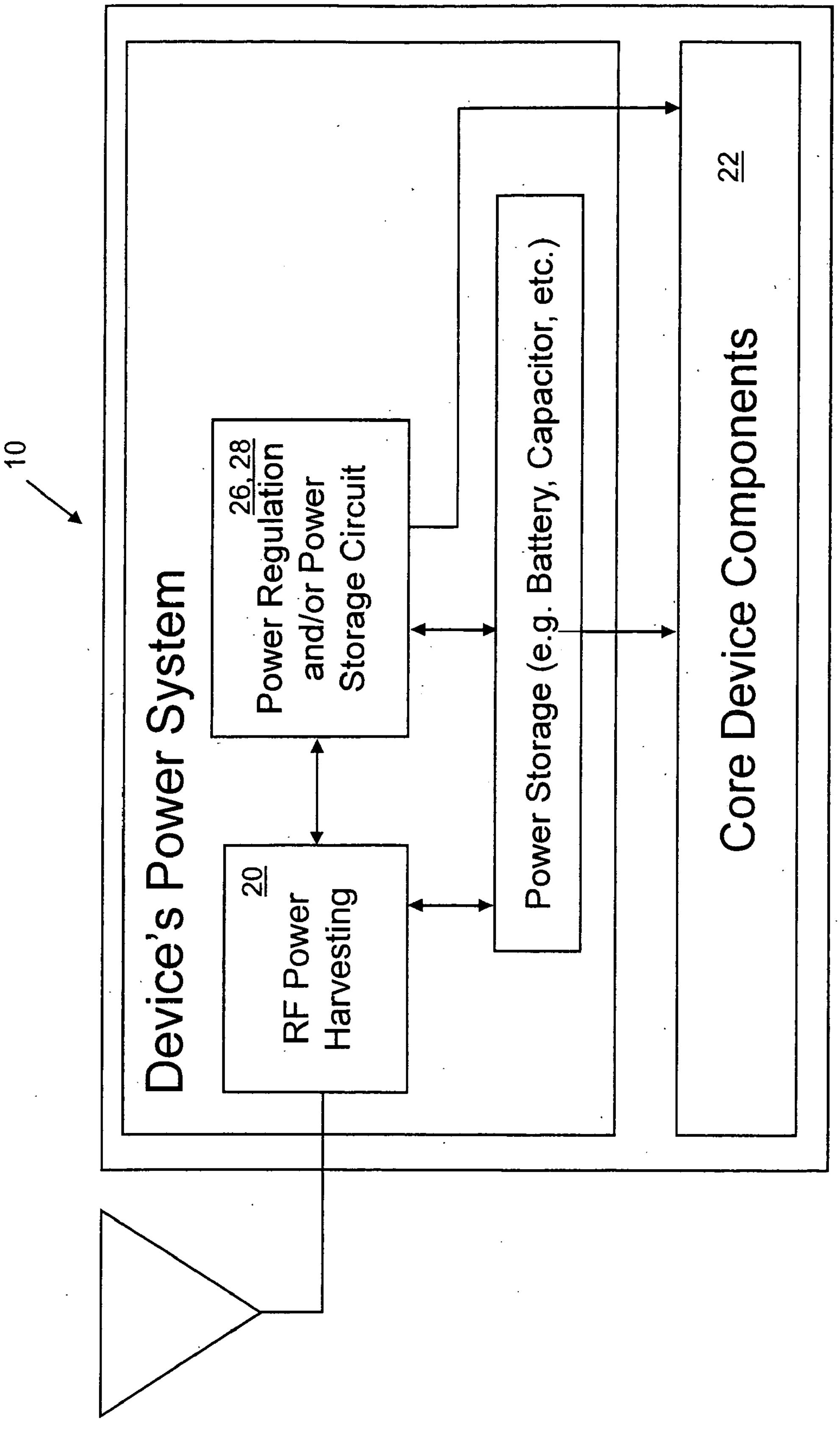
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 17

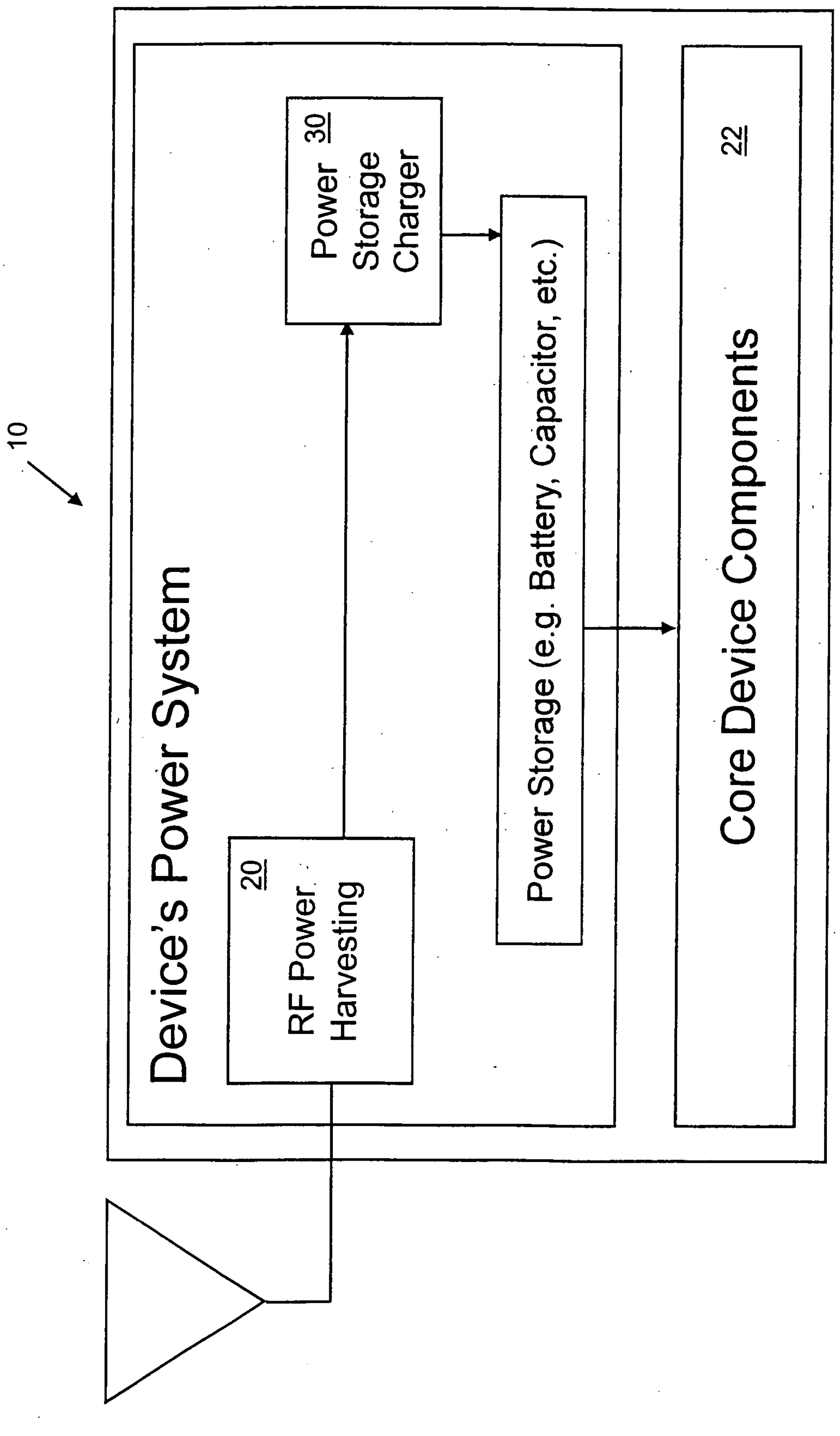
10
Device's Power System
RF Power Harvesting 20
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 18

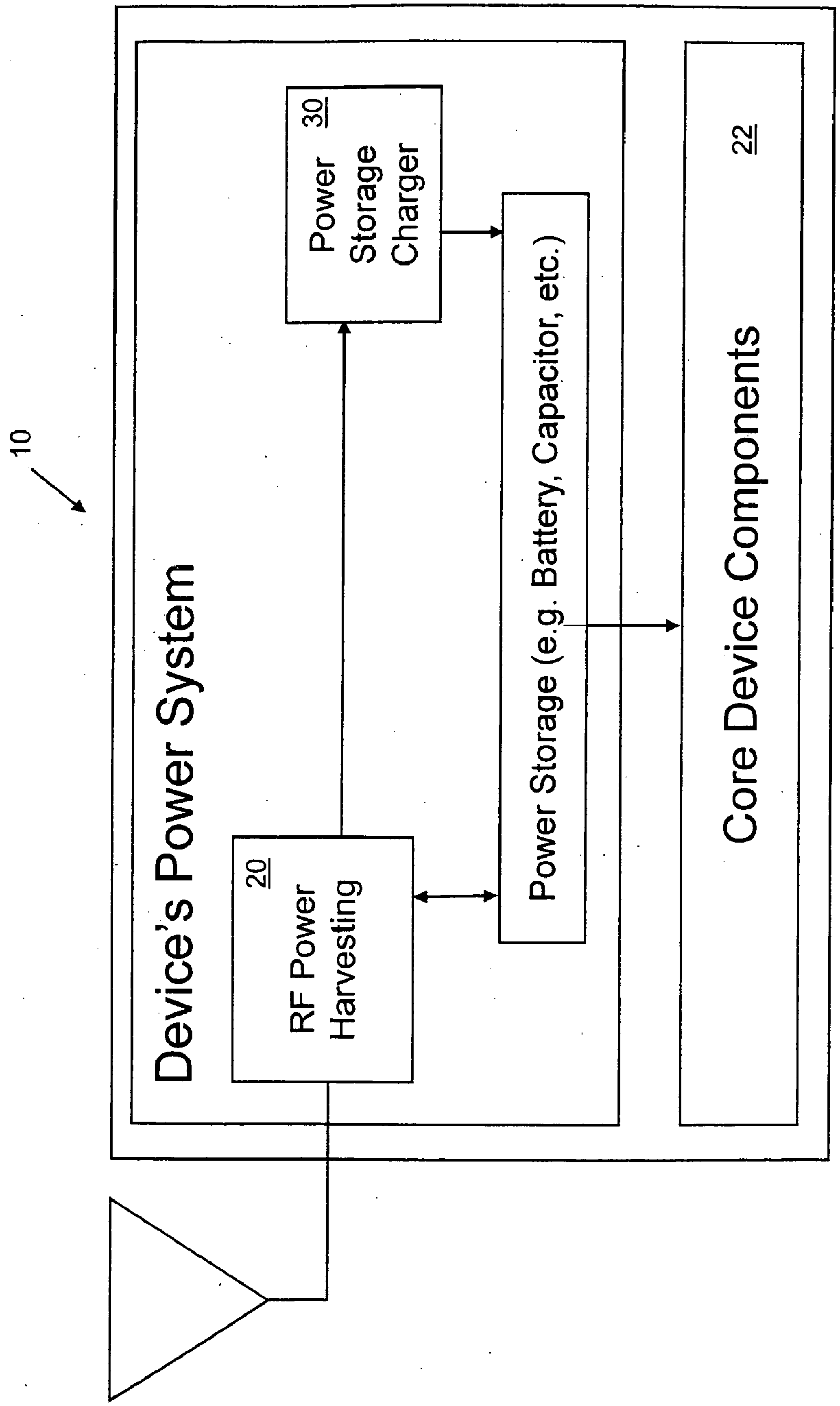
10
Device's Power System
20
RF Power Harvesting
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 19

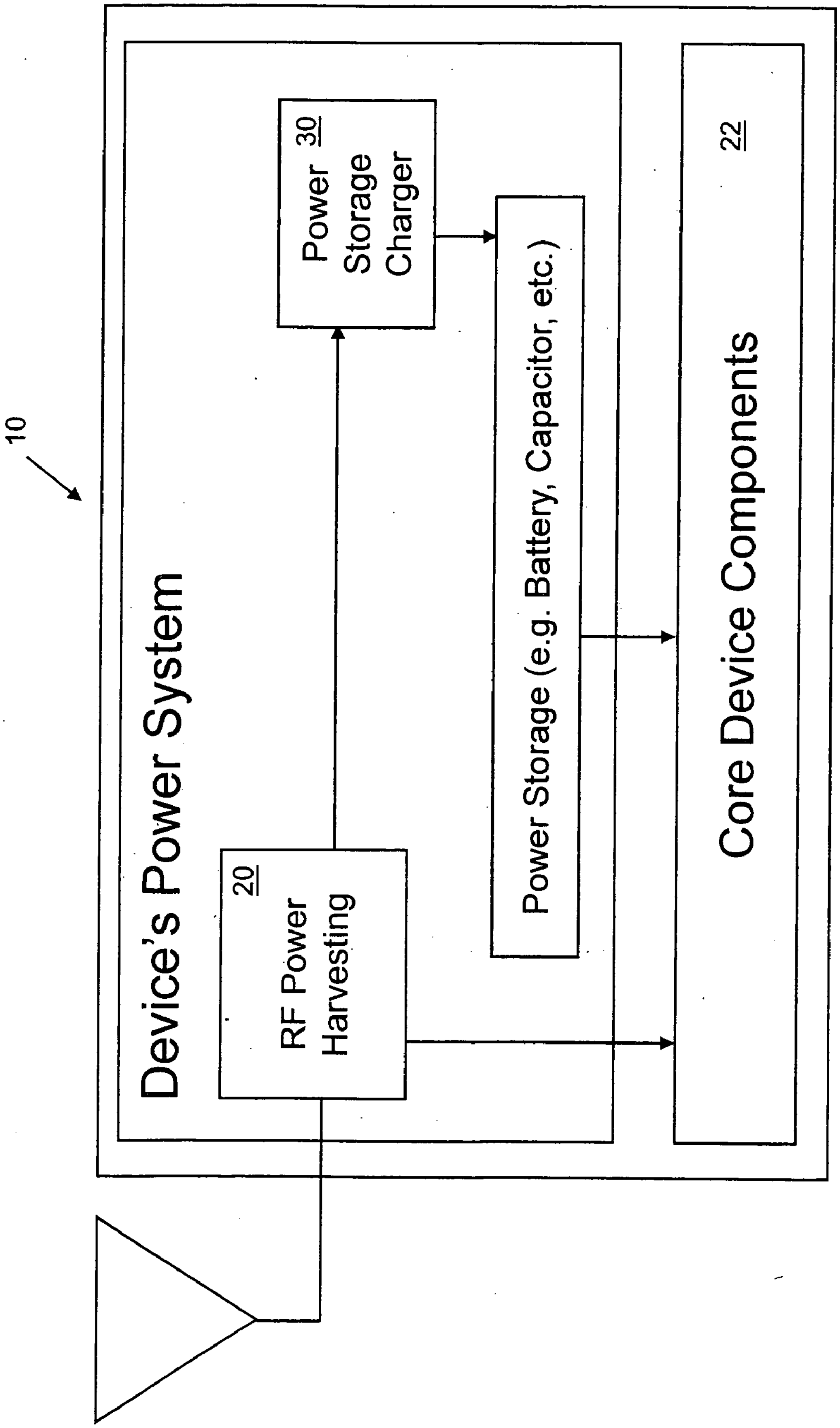
10
Device's Power System
RF Power Harvesting 20
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 20

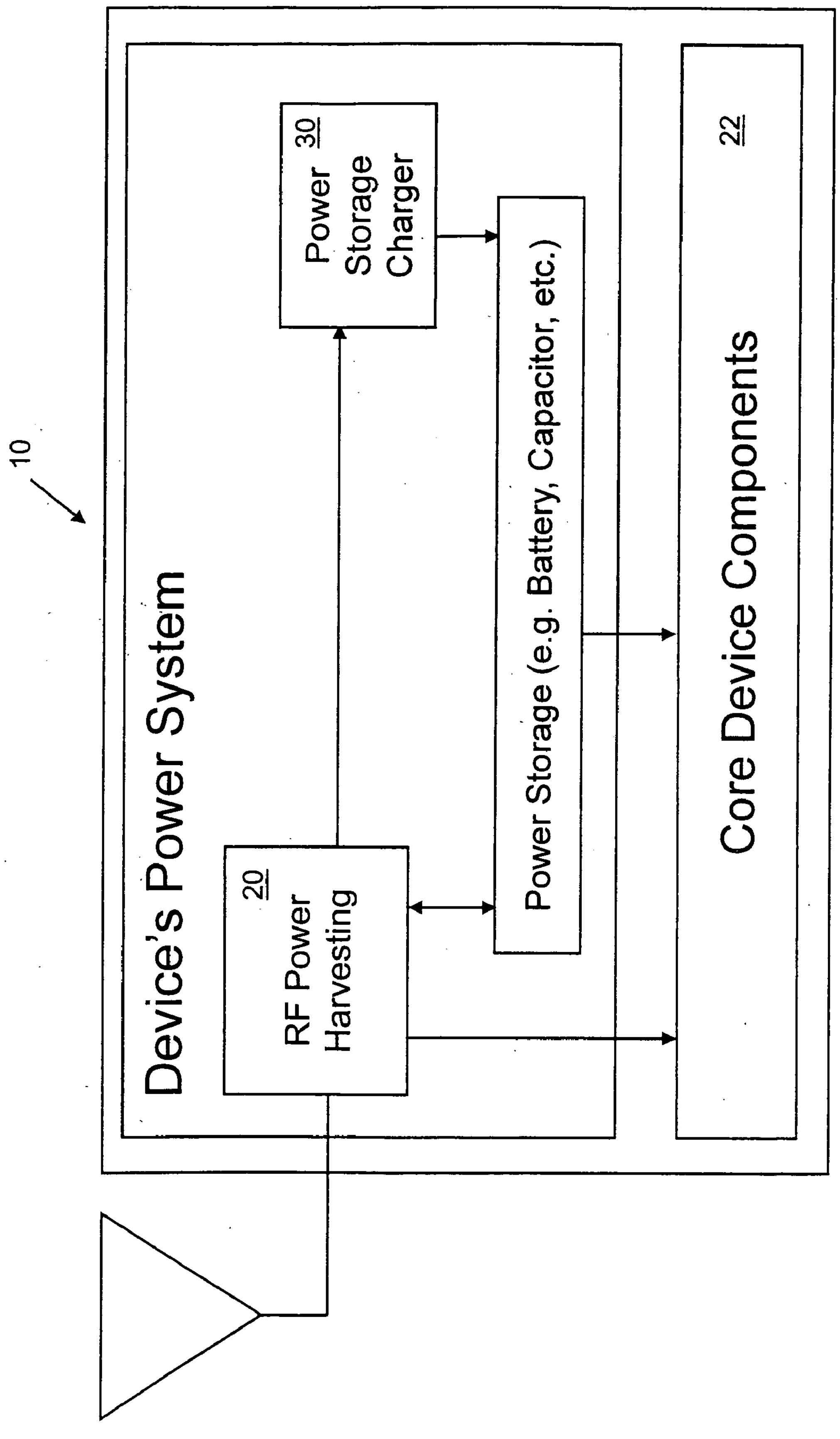
10
Device's Power System
20
RF Power Harvesting
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 21

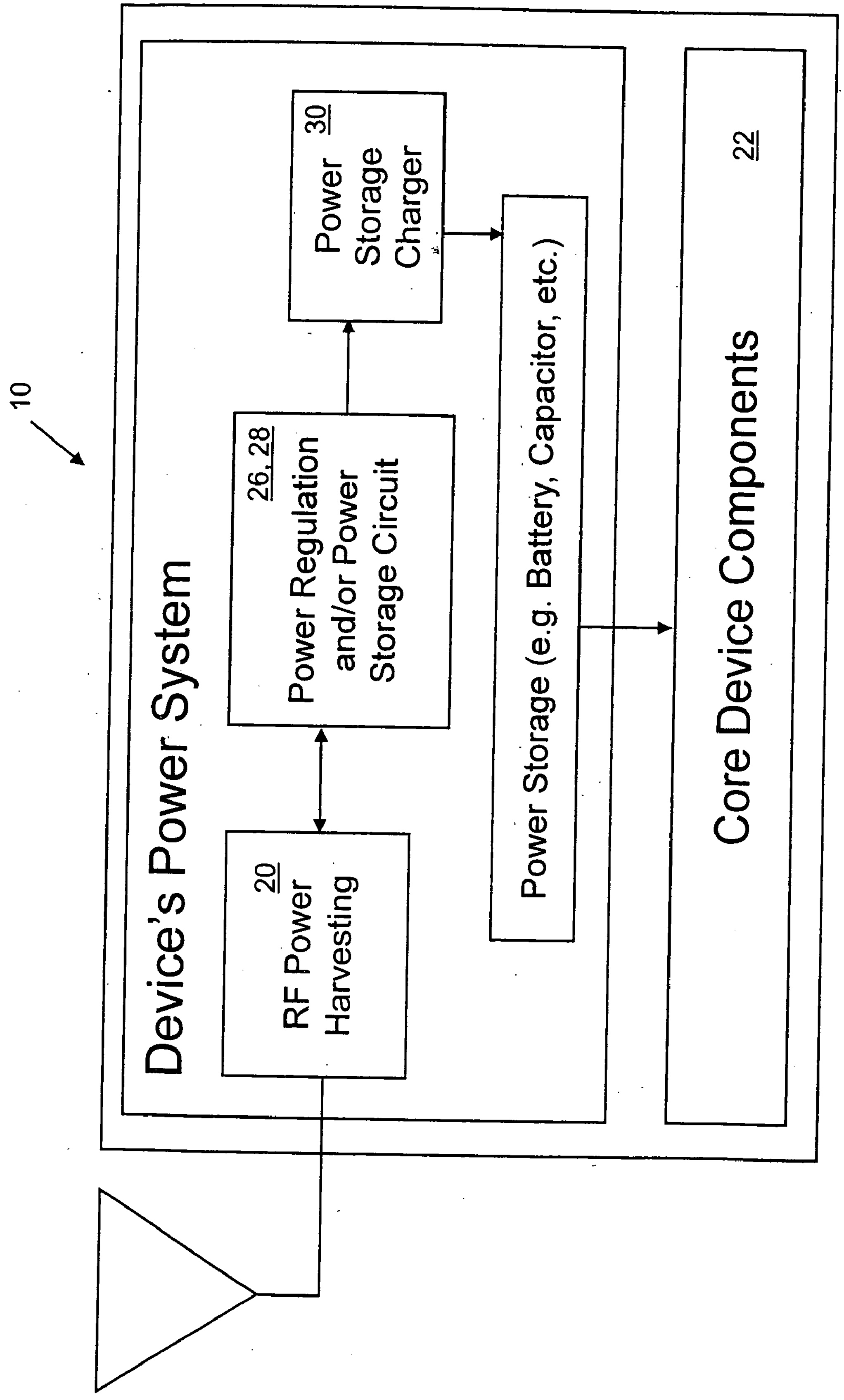
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 22

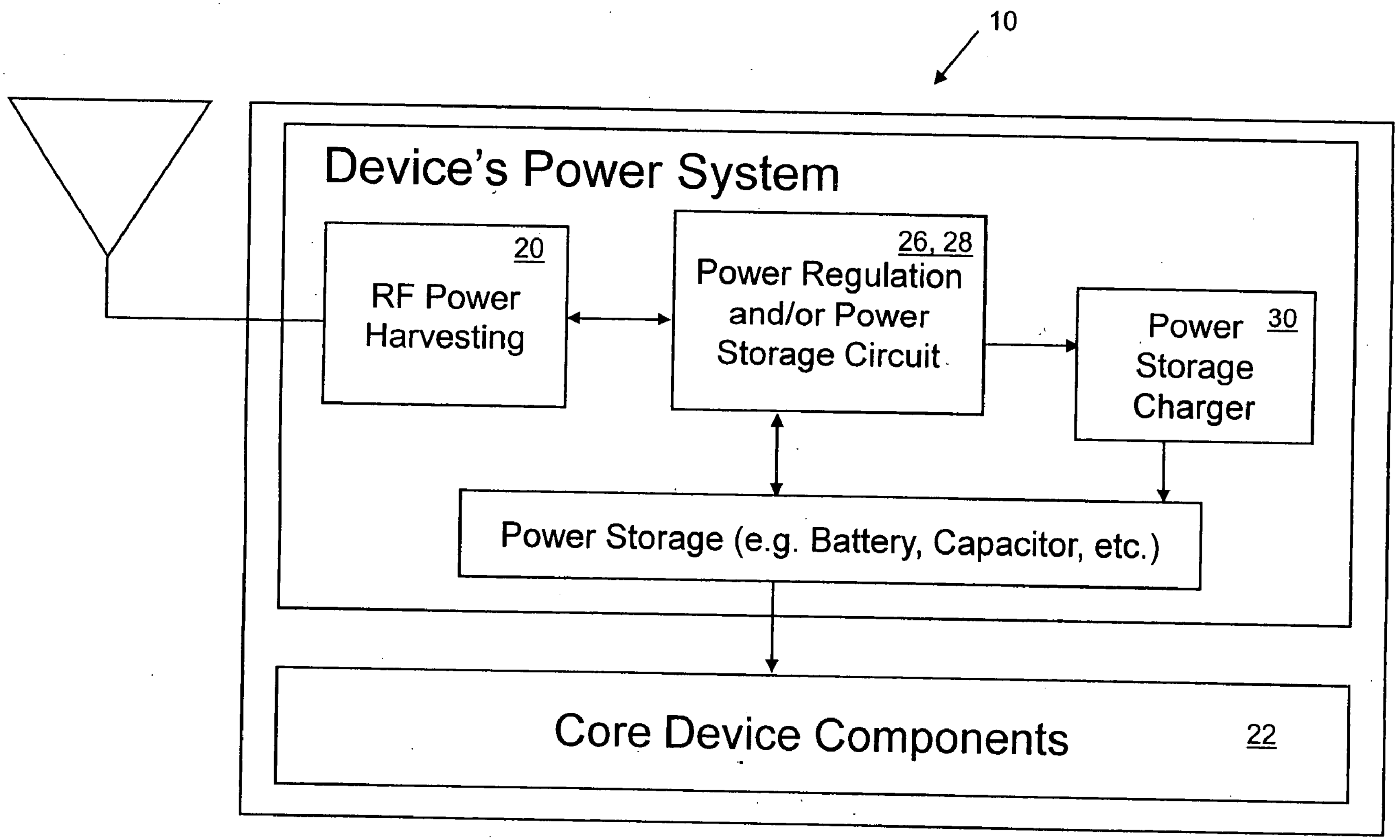
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 23

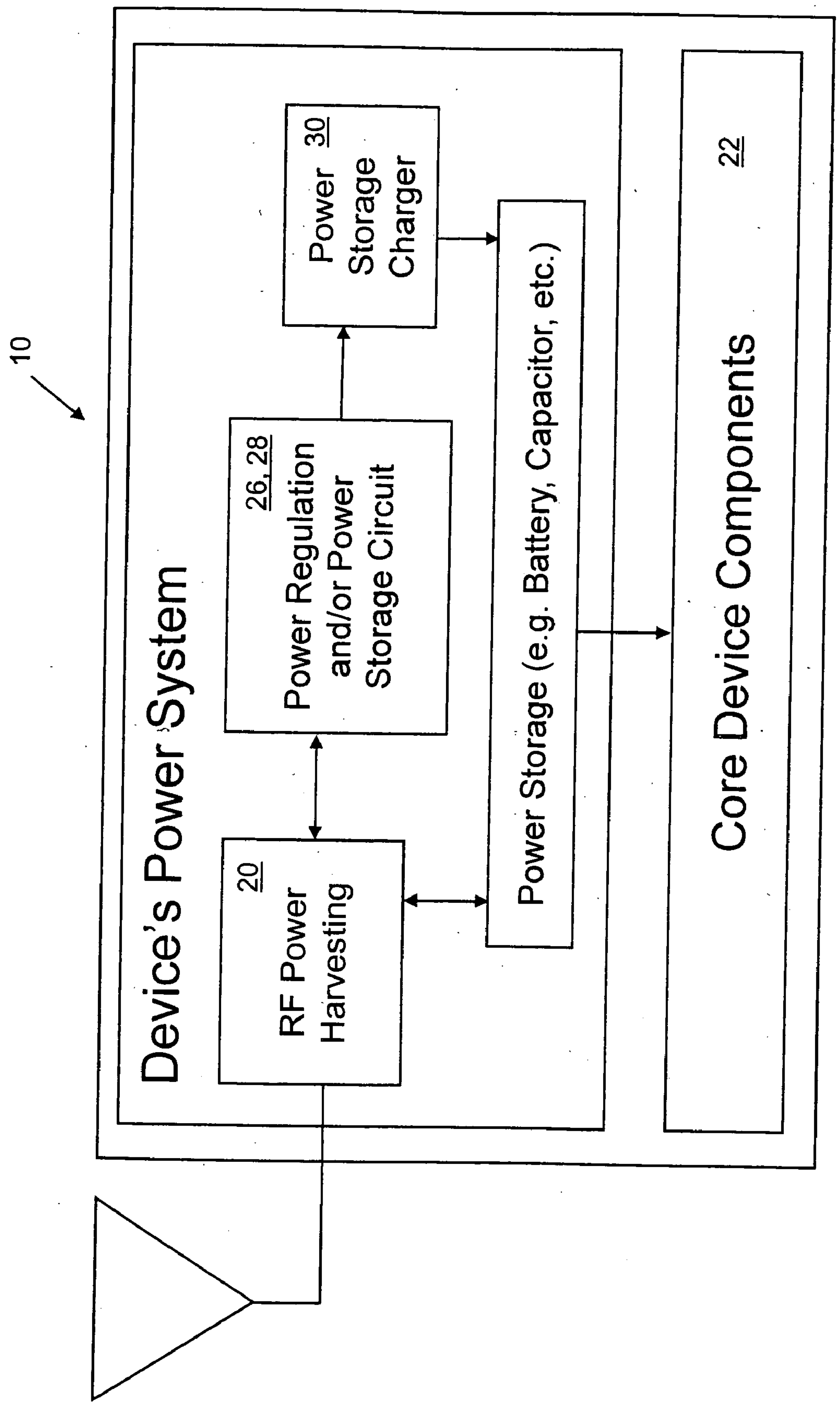
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 24

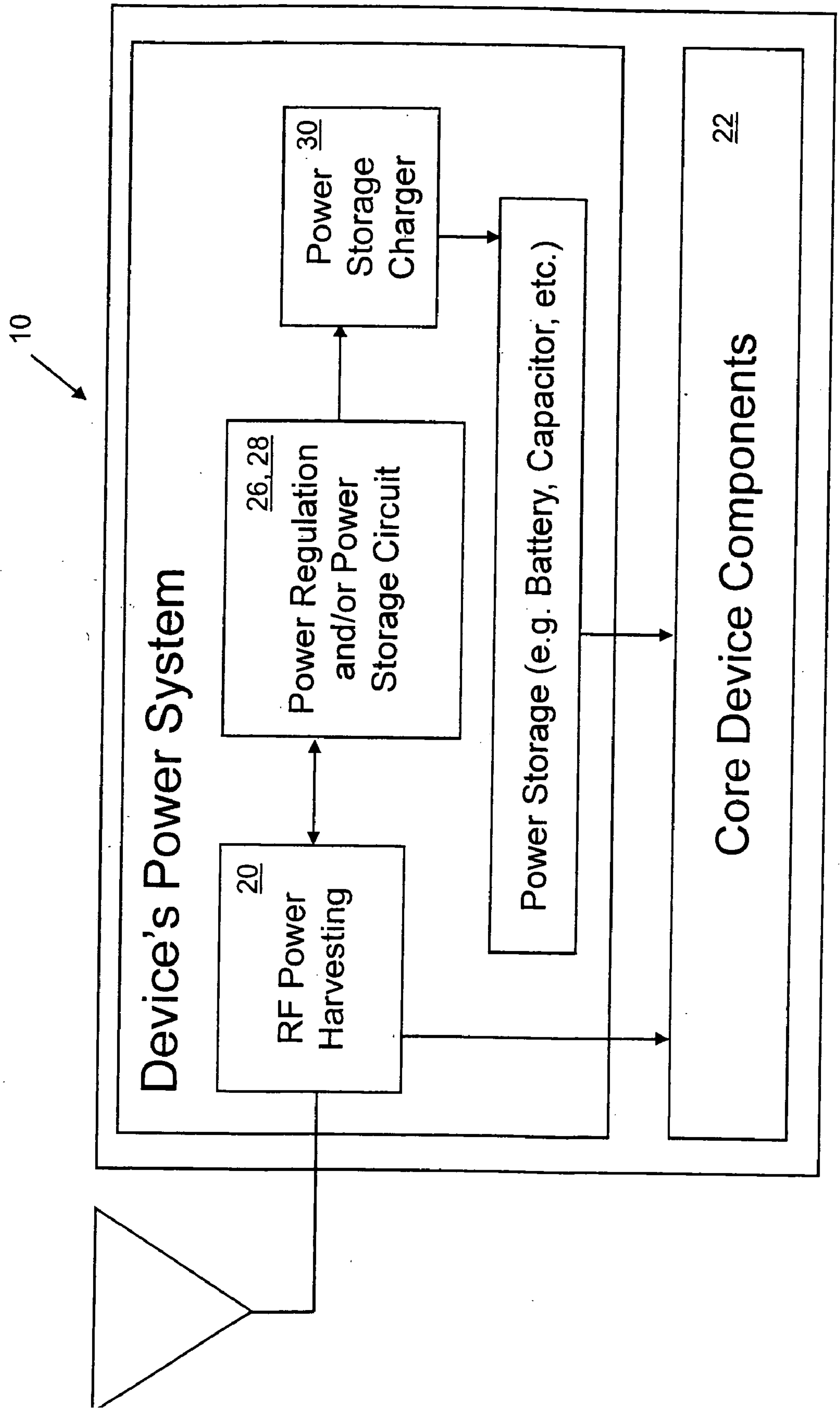
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 25

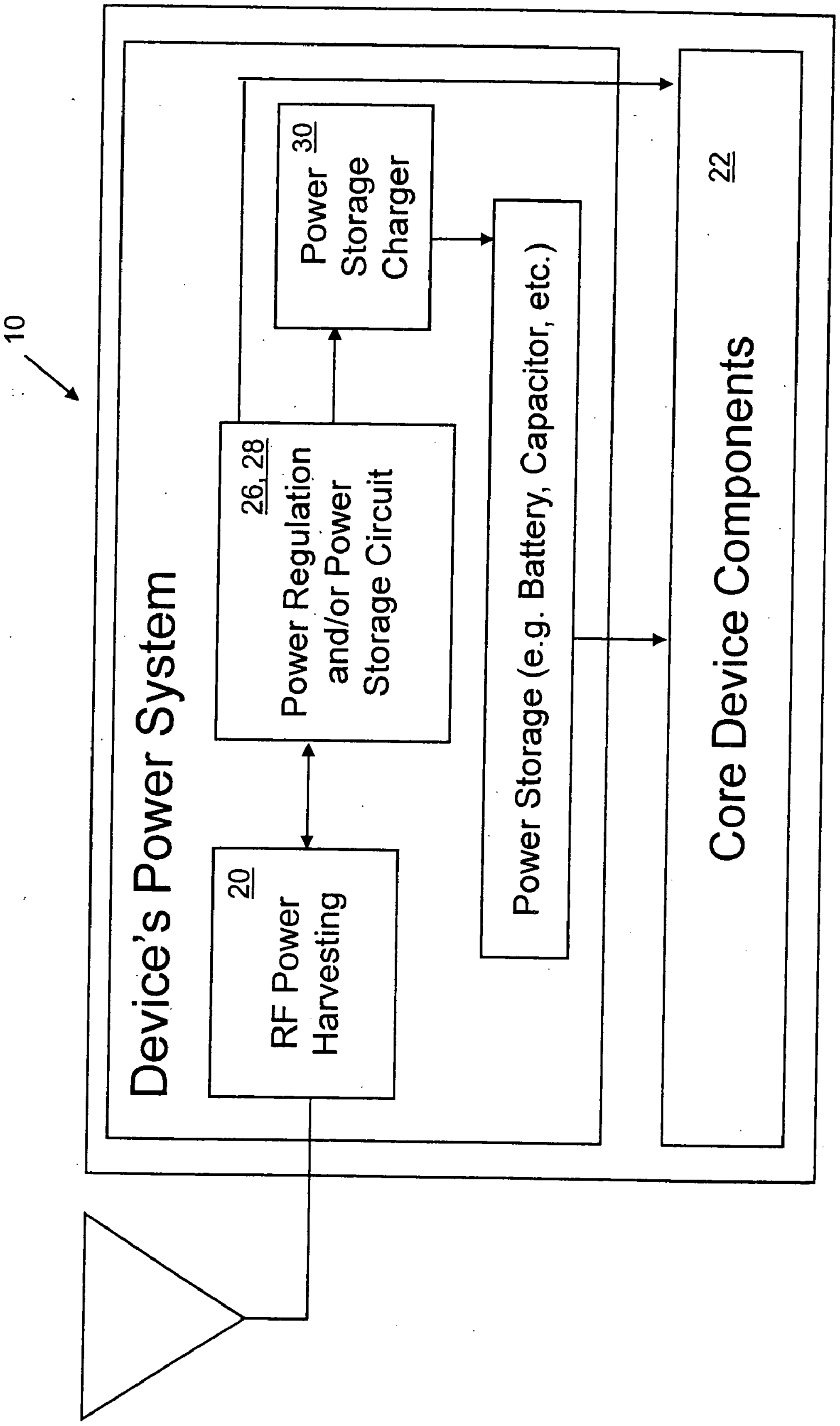
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 26

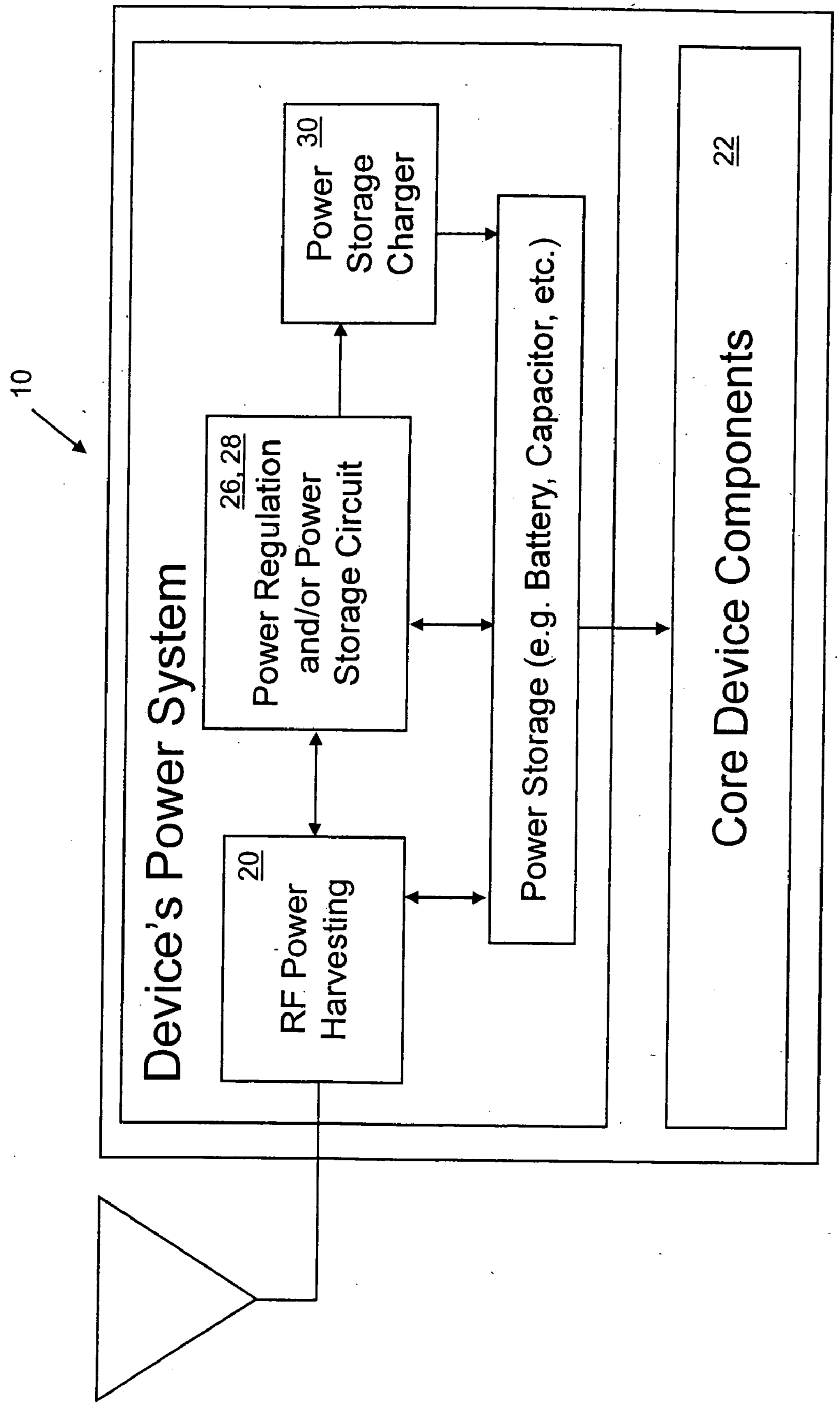
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 27

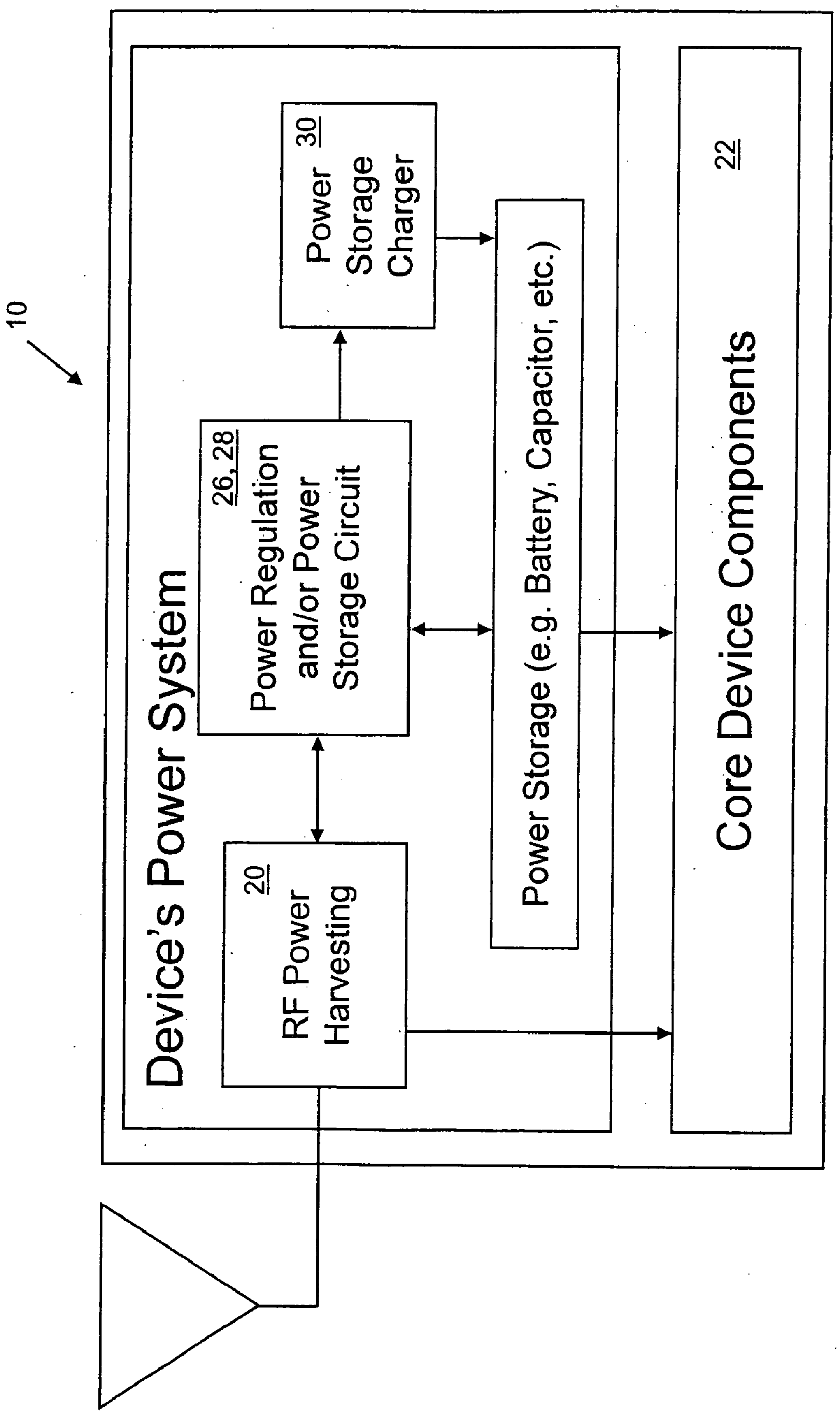
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 28

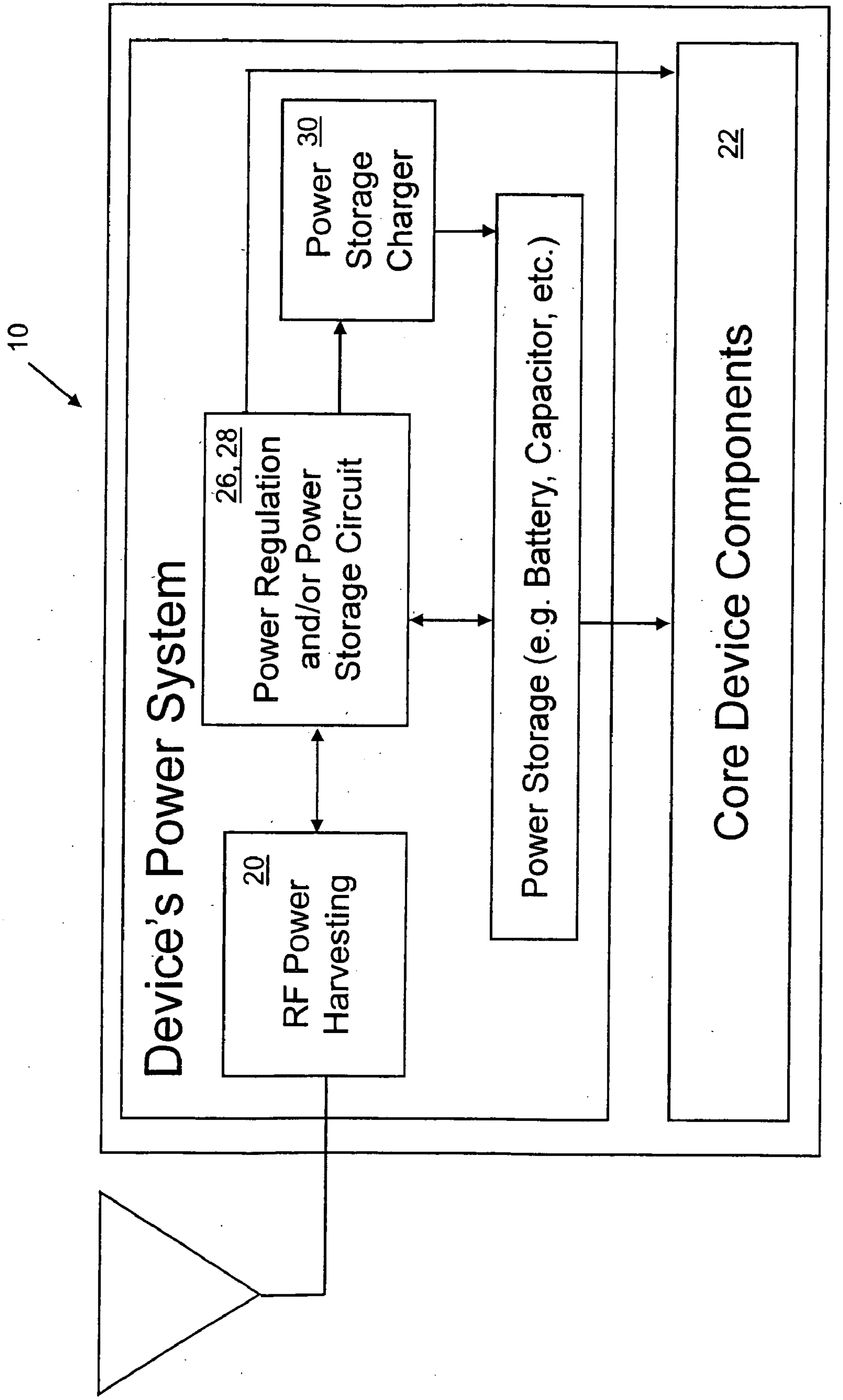
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 29

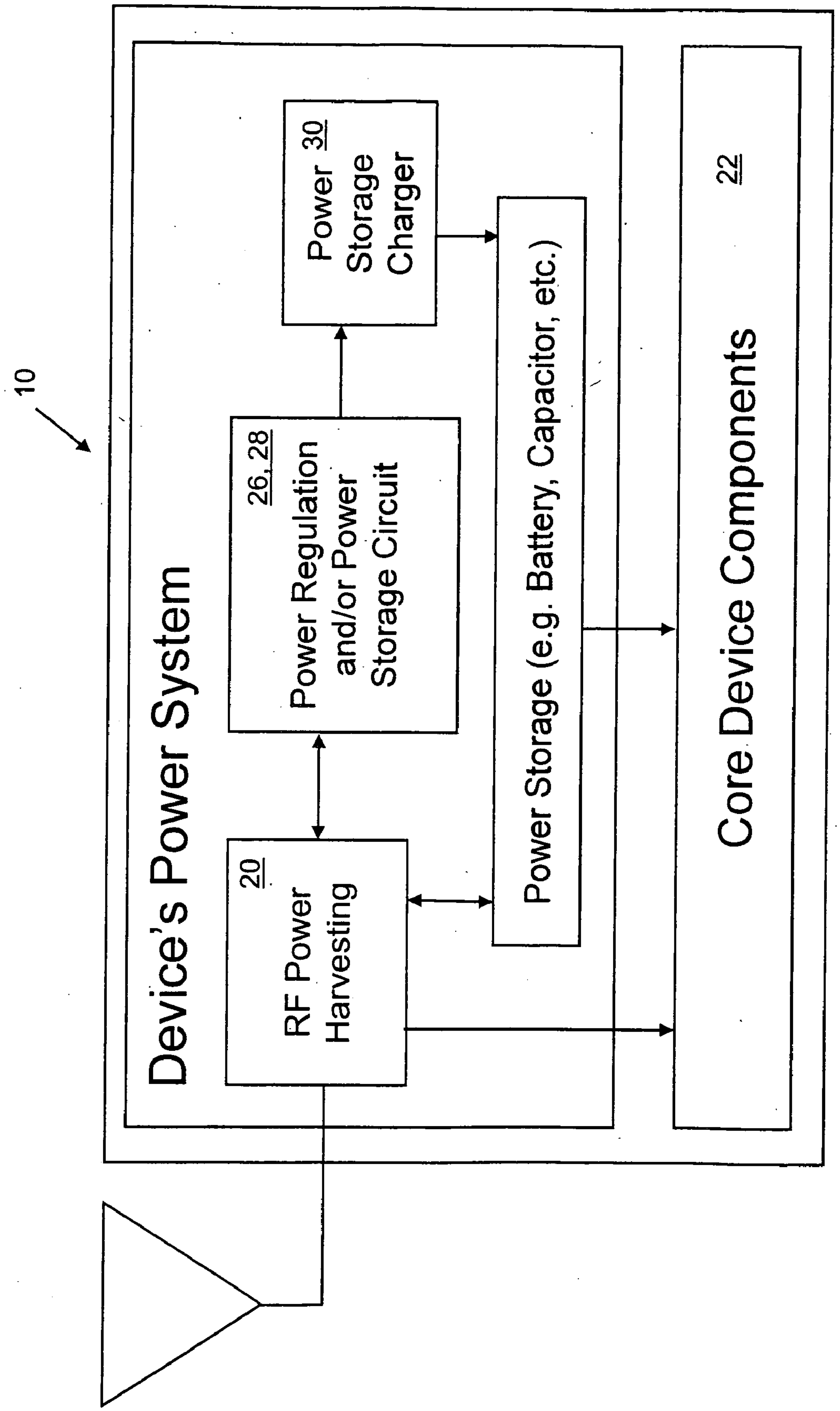
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 30

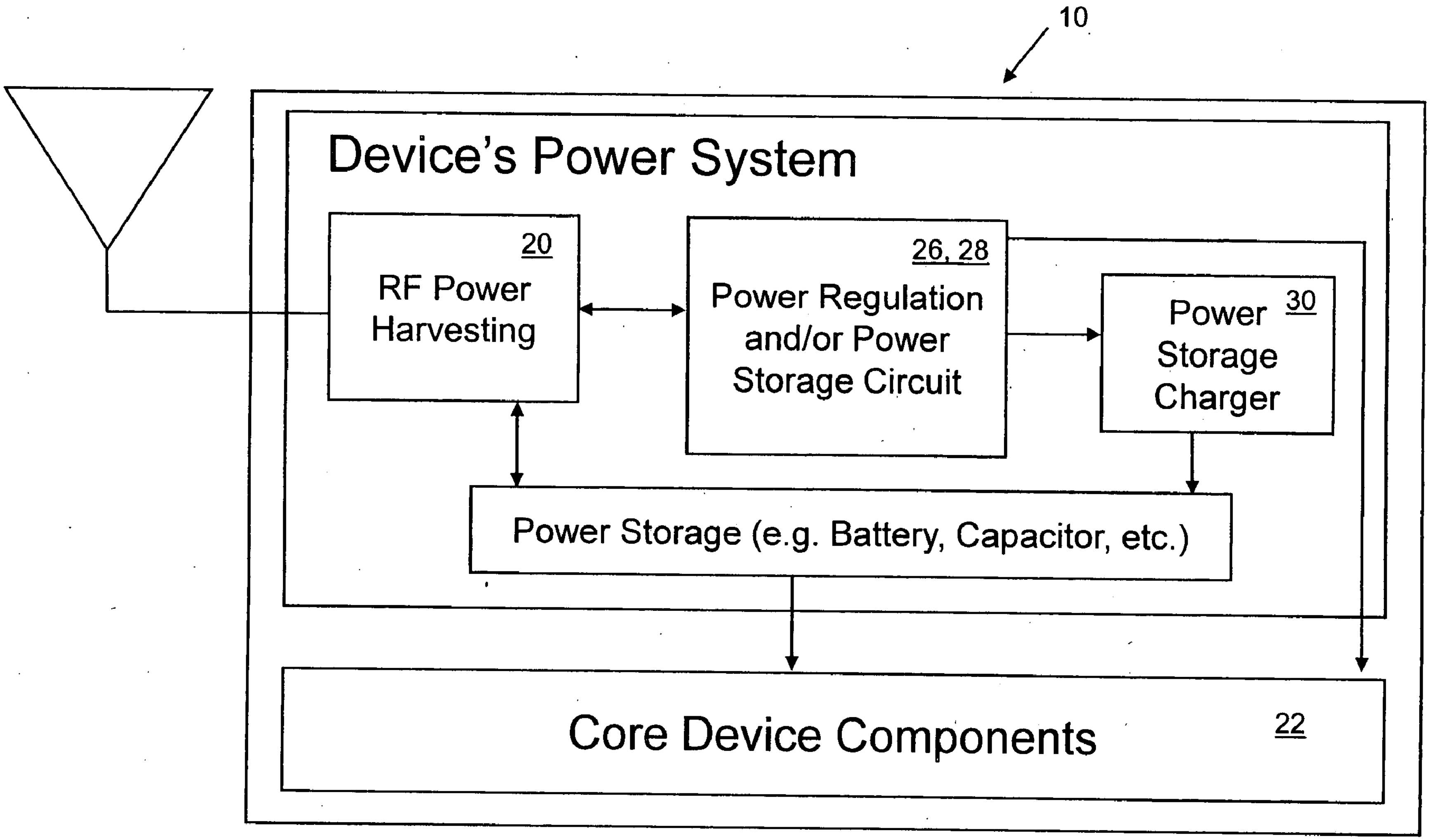
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 31

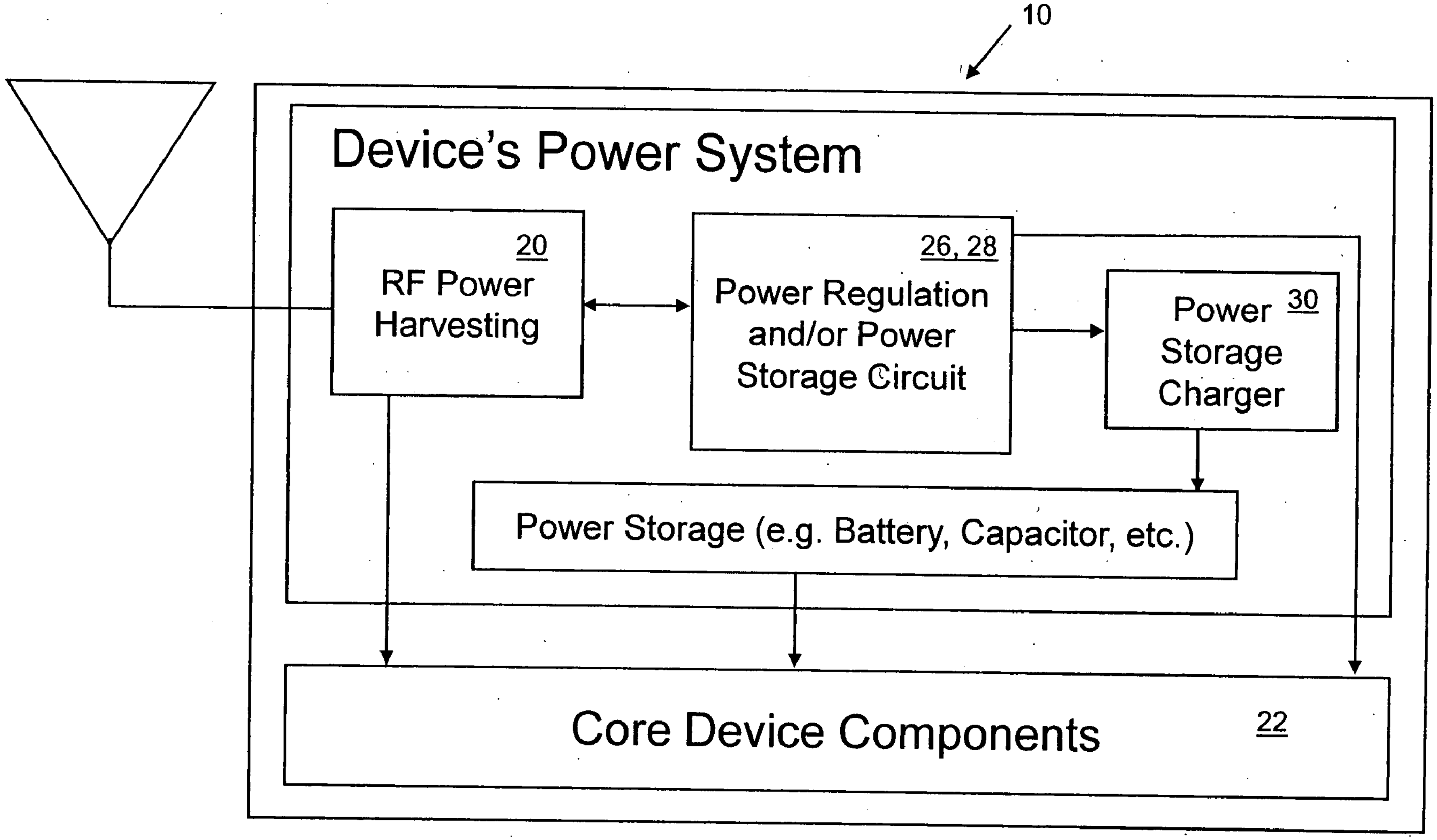
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 32

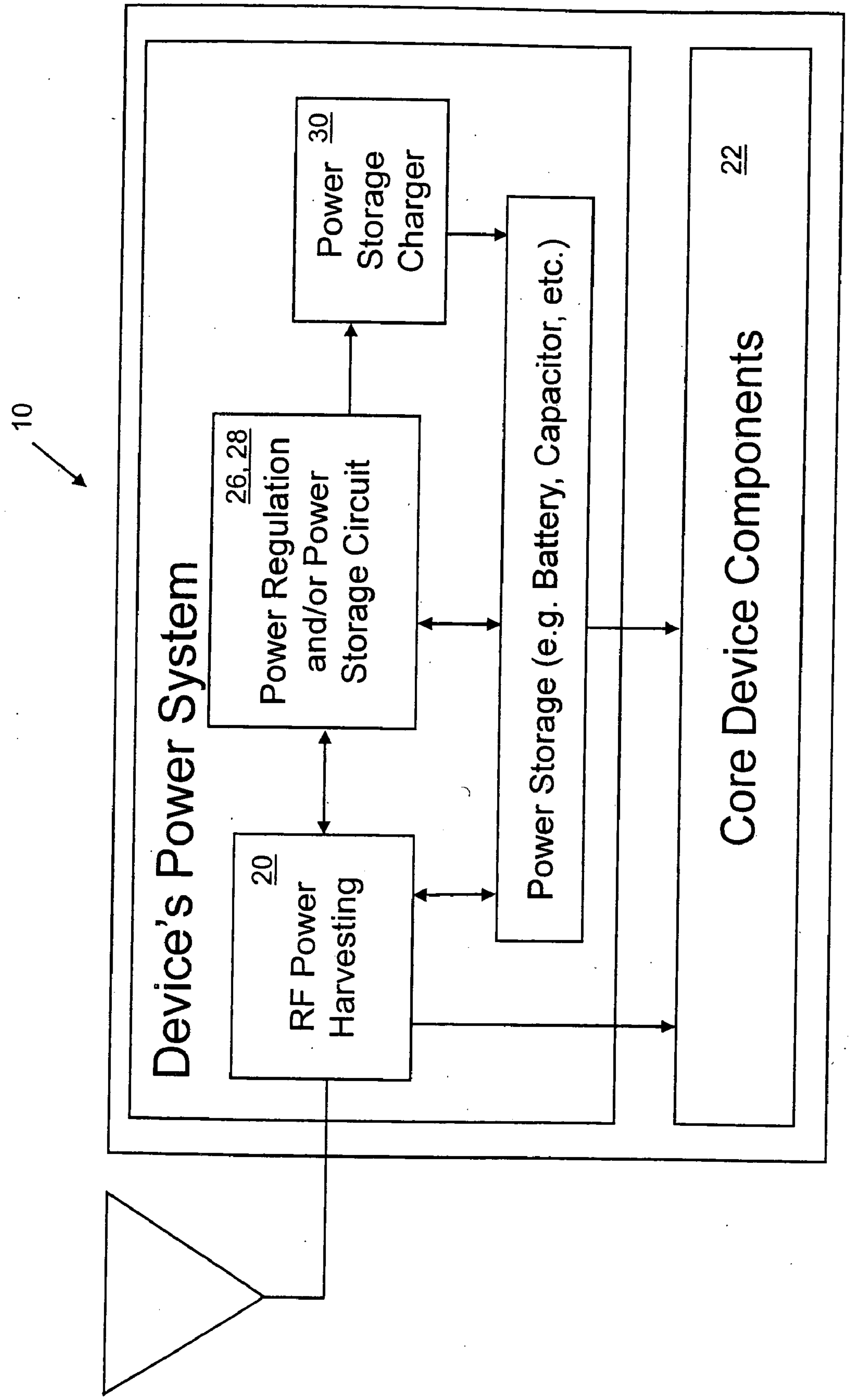
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 33

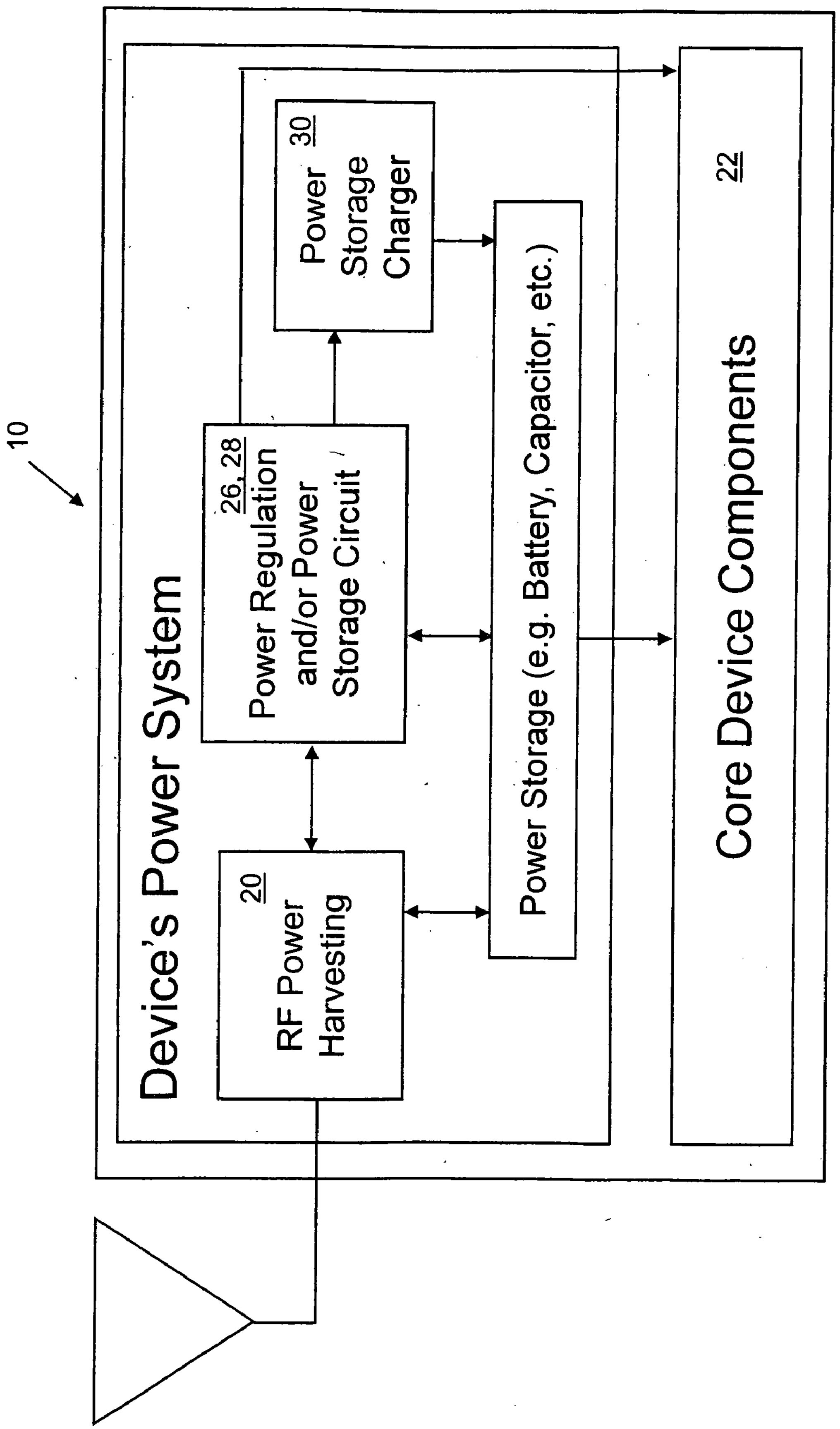
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 34

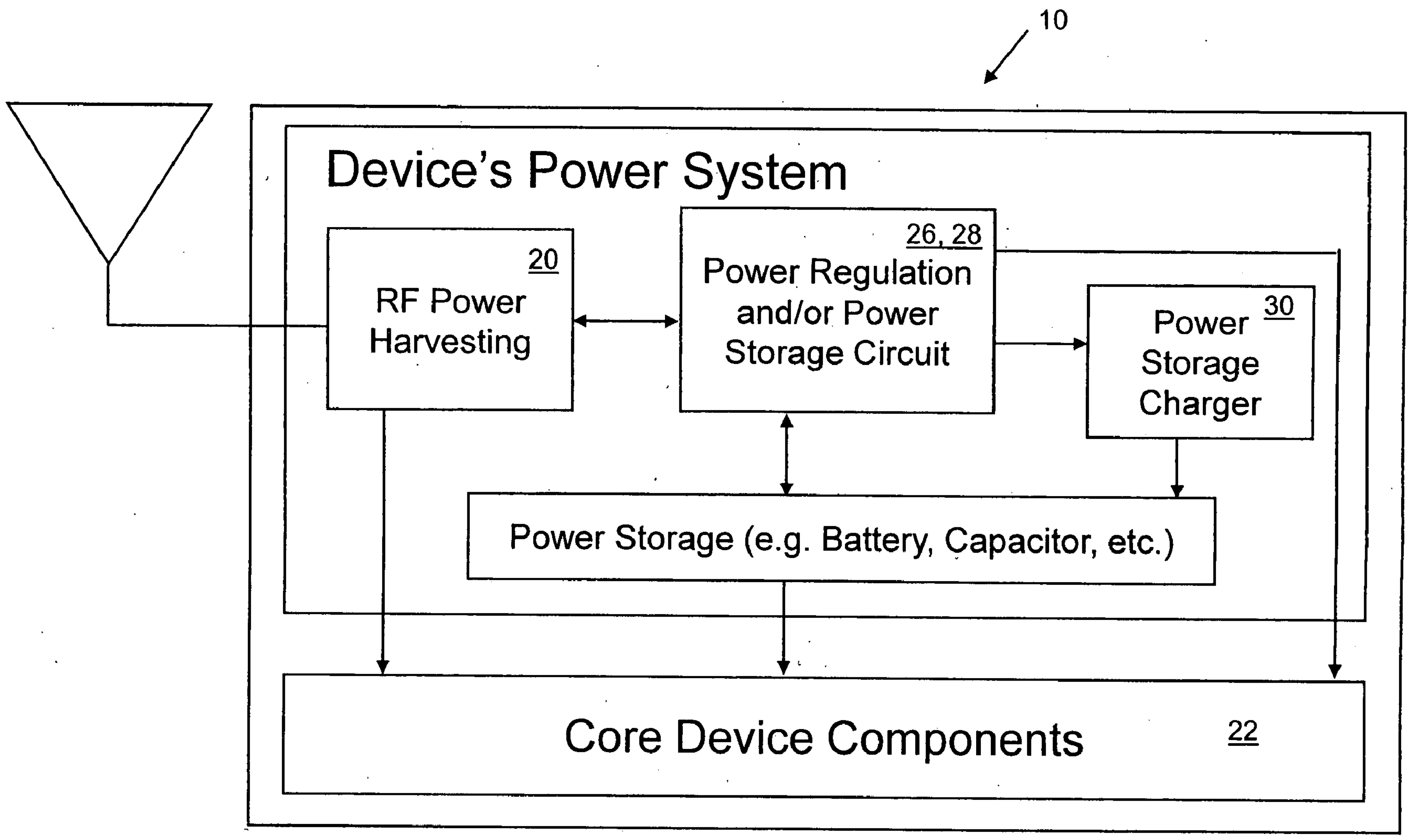
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 35

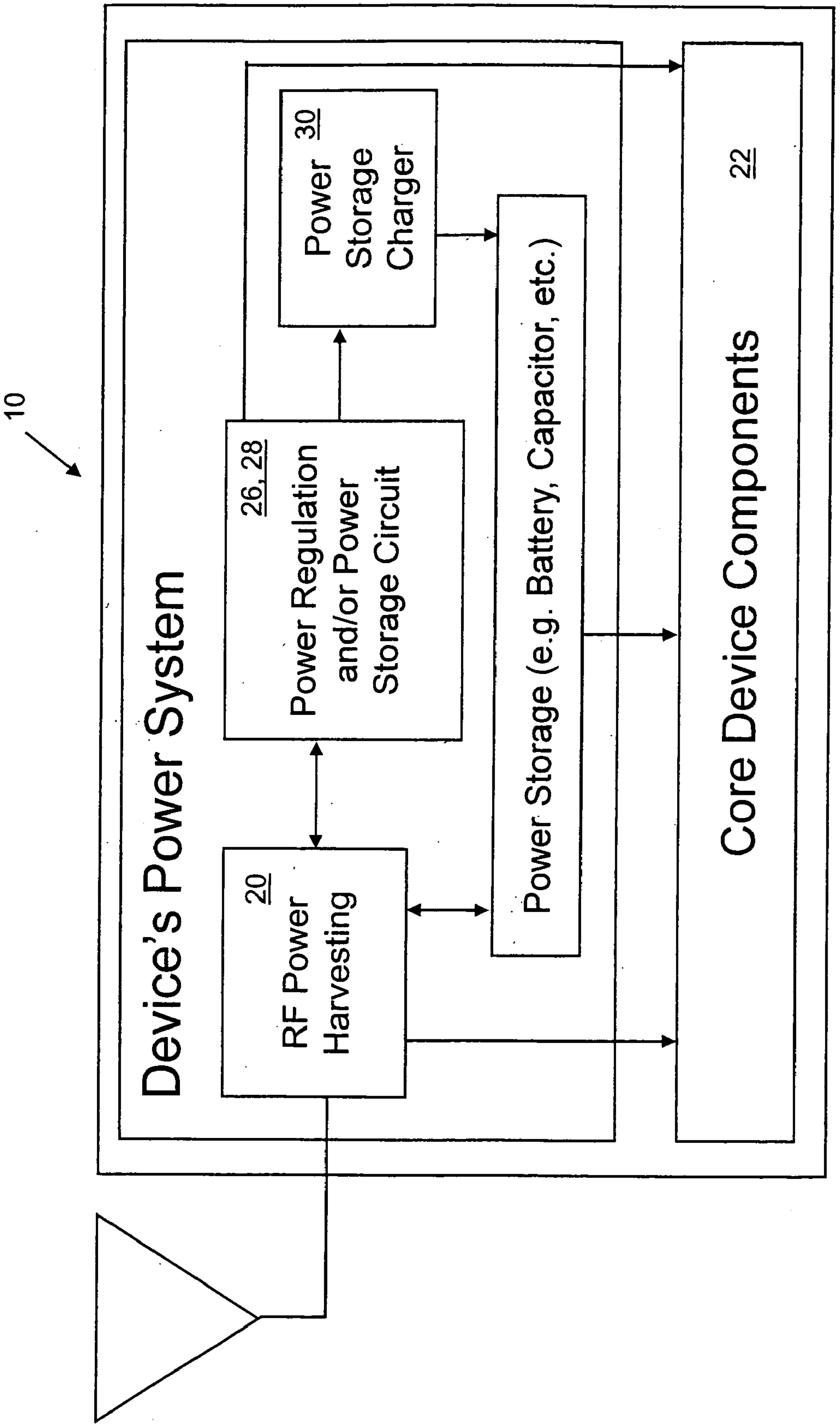
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 36

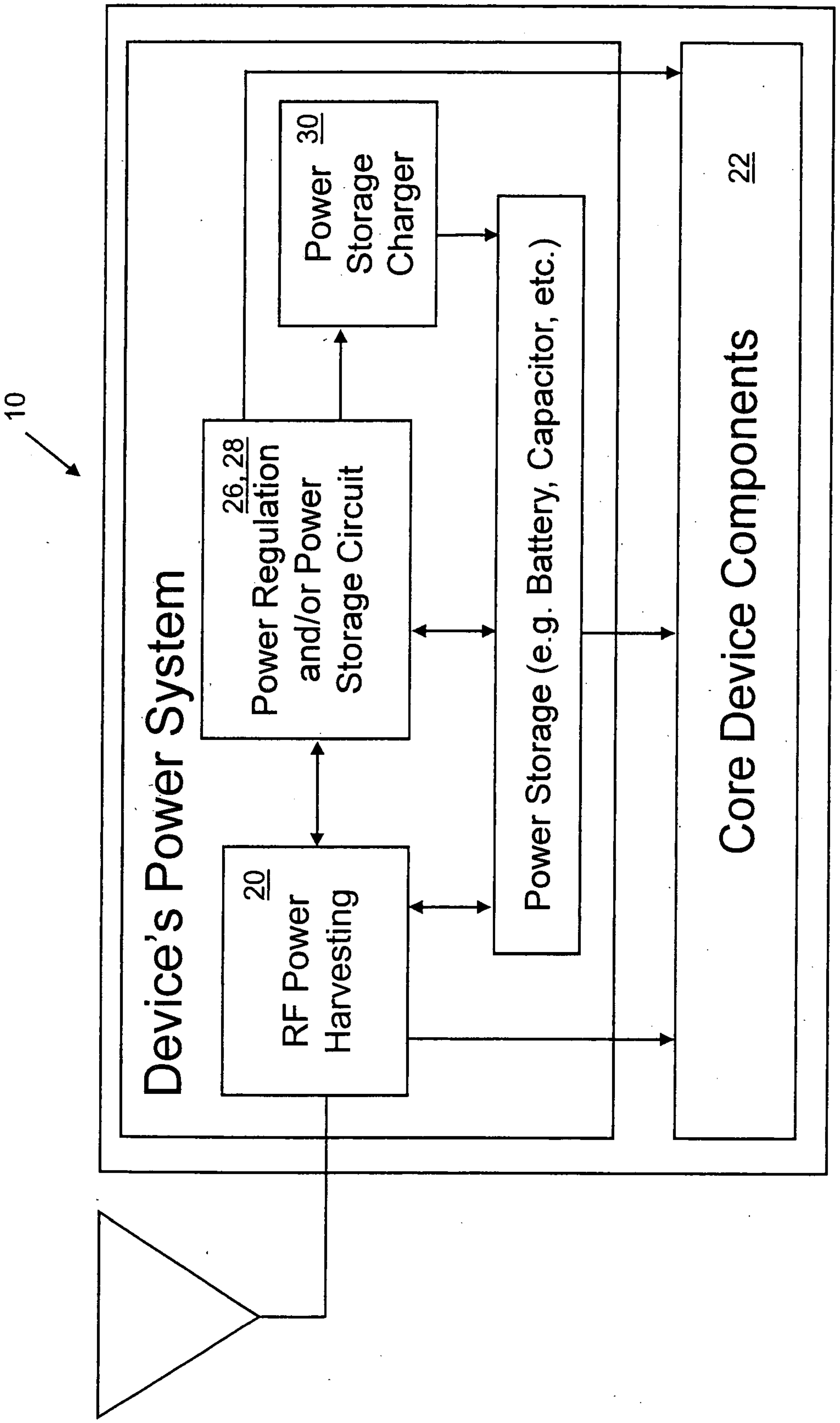
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 37

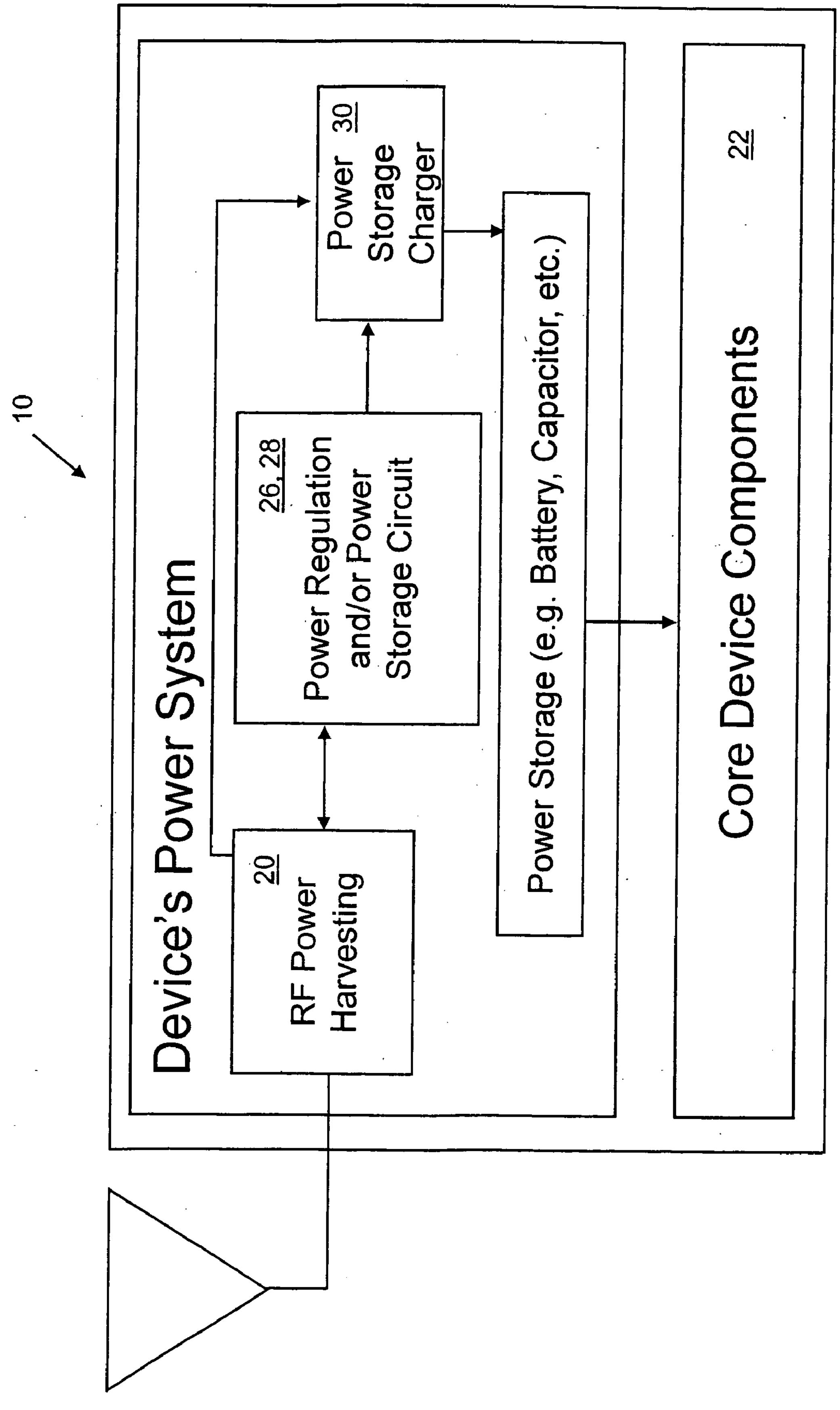
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 38

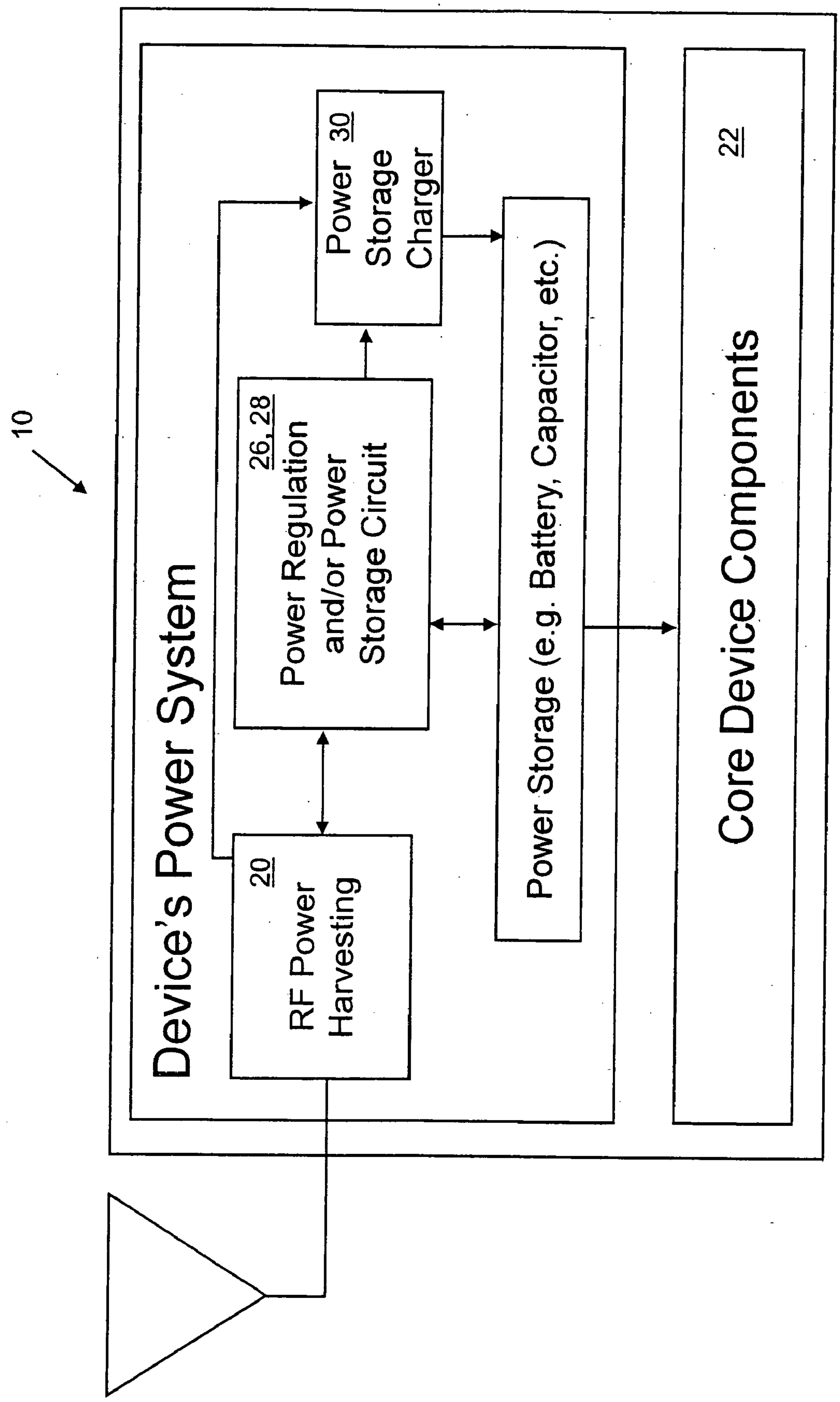
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 39

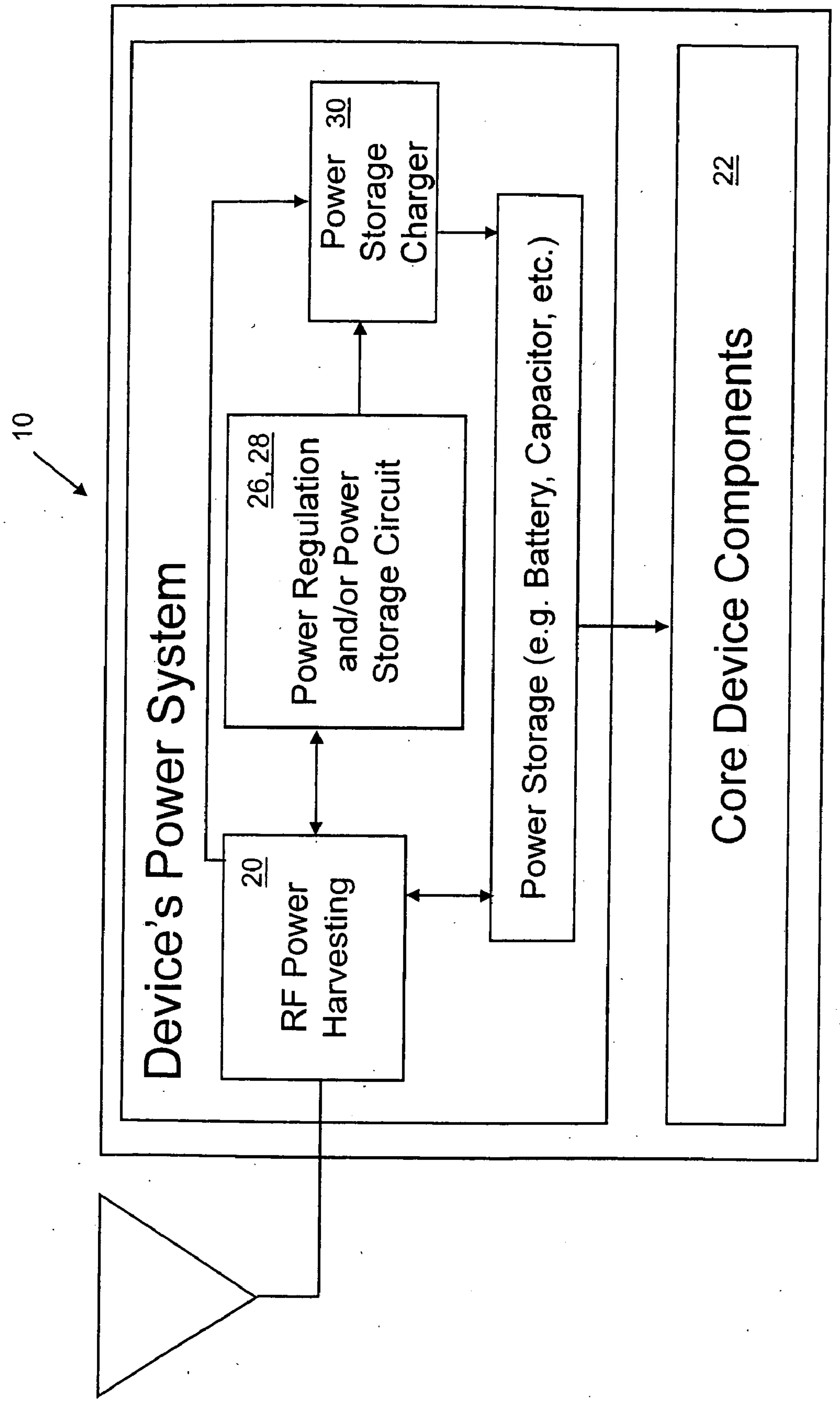
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 40

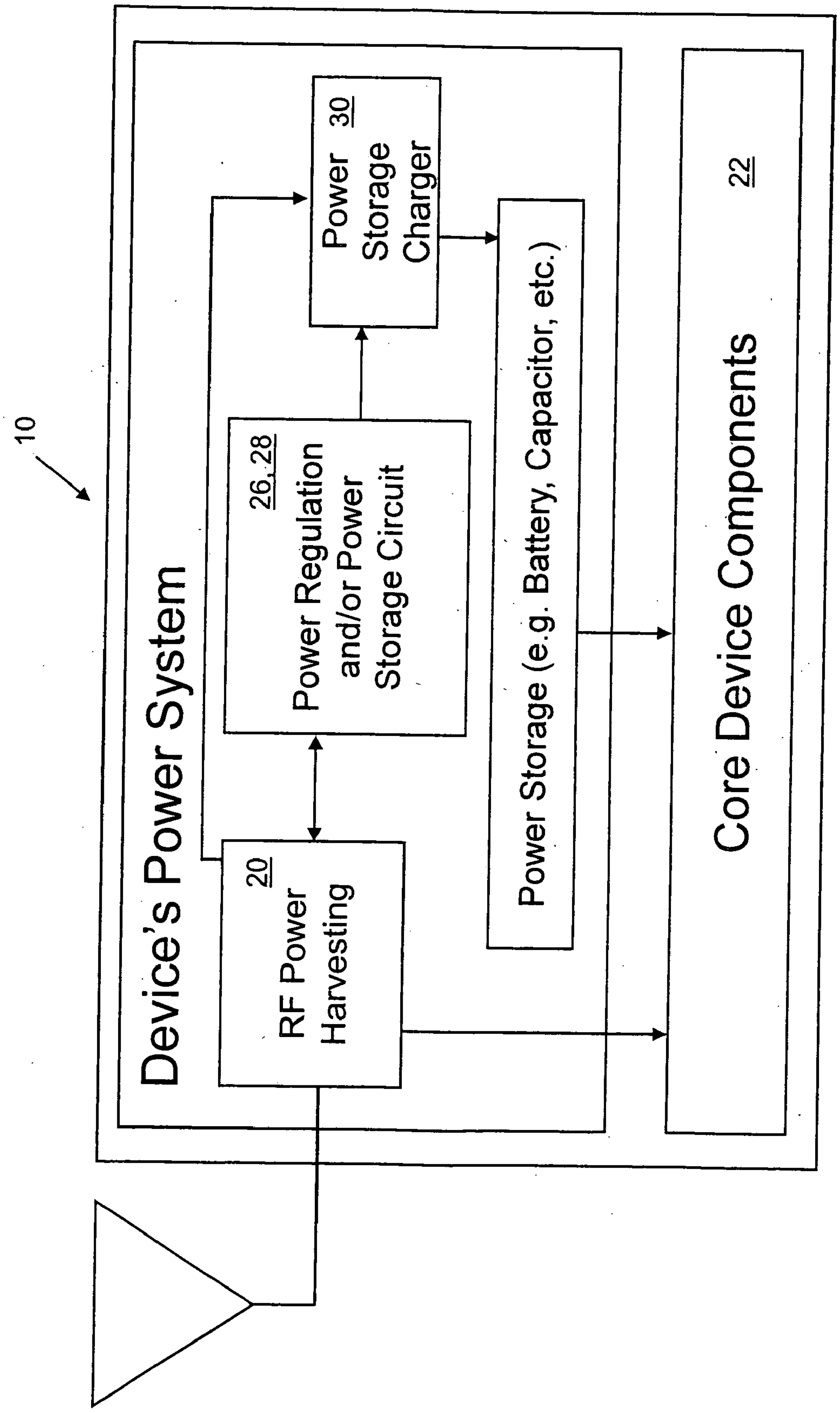
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 41

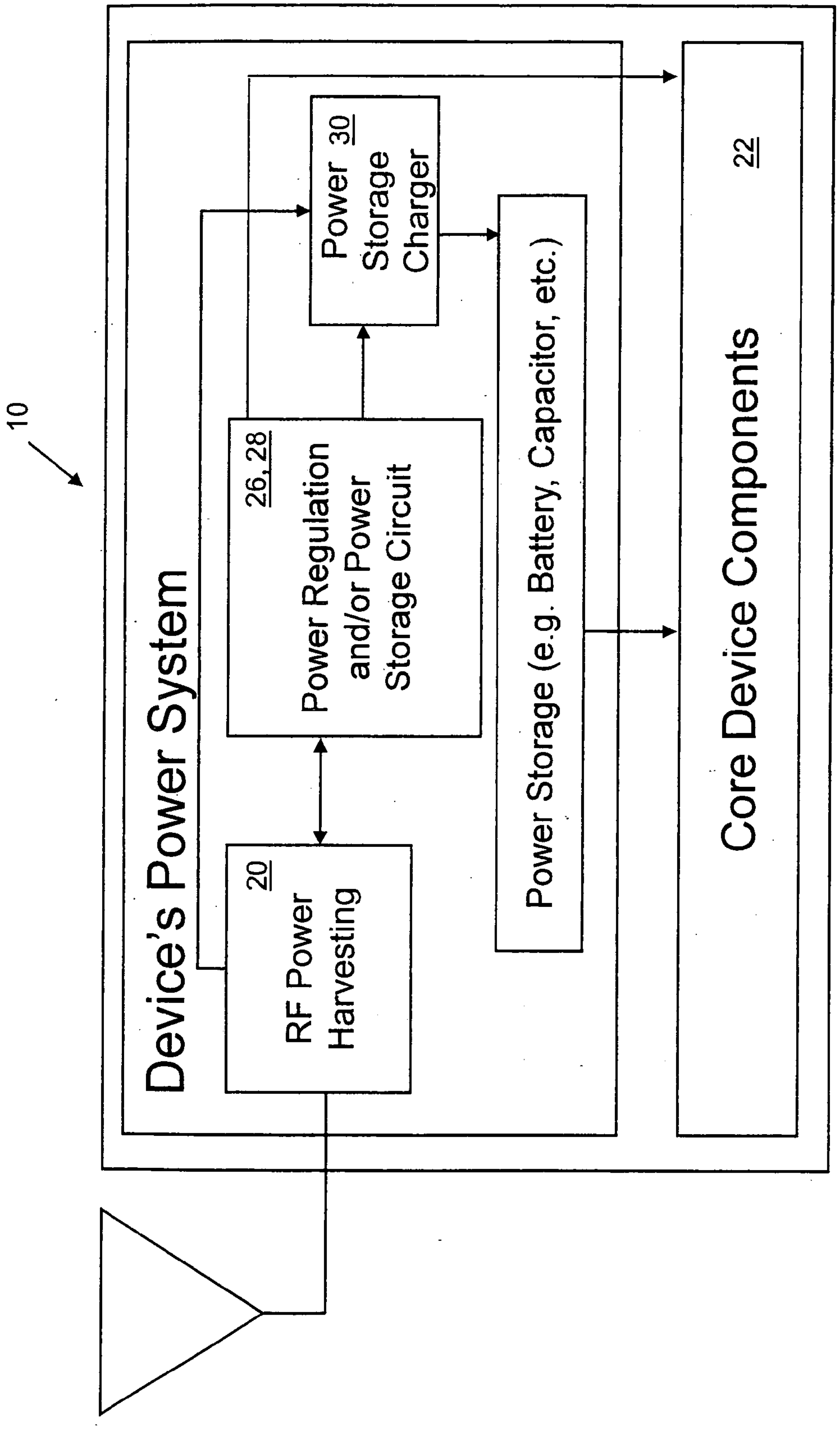
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 42

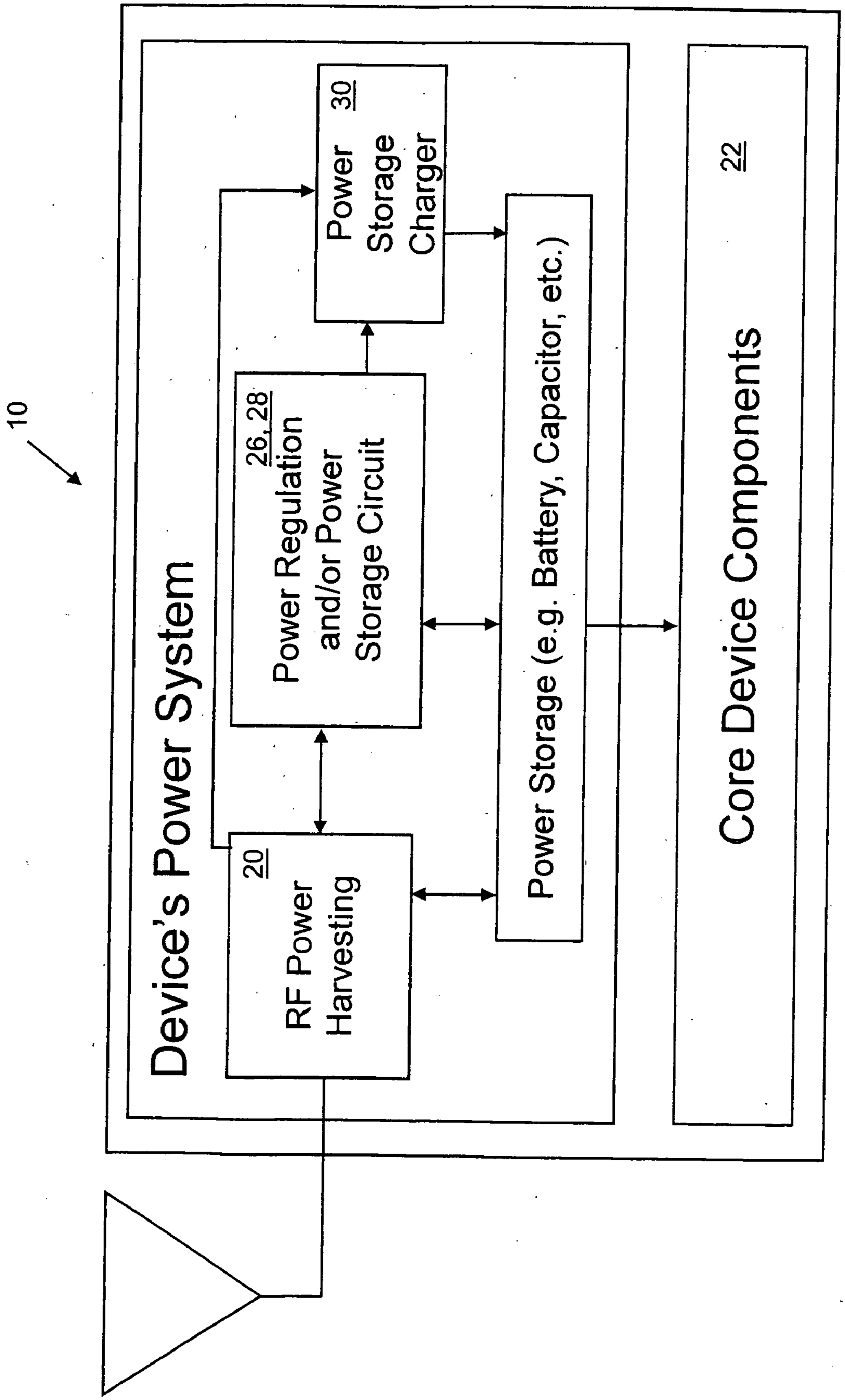
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 43

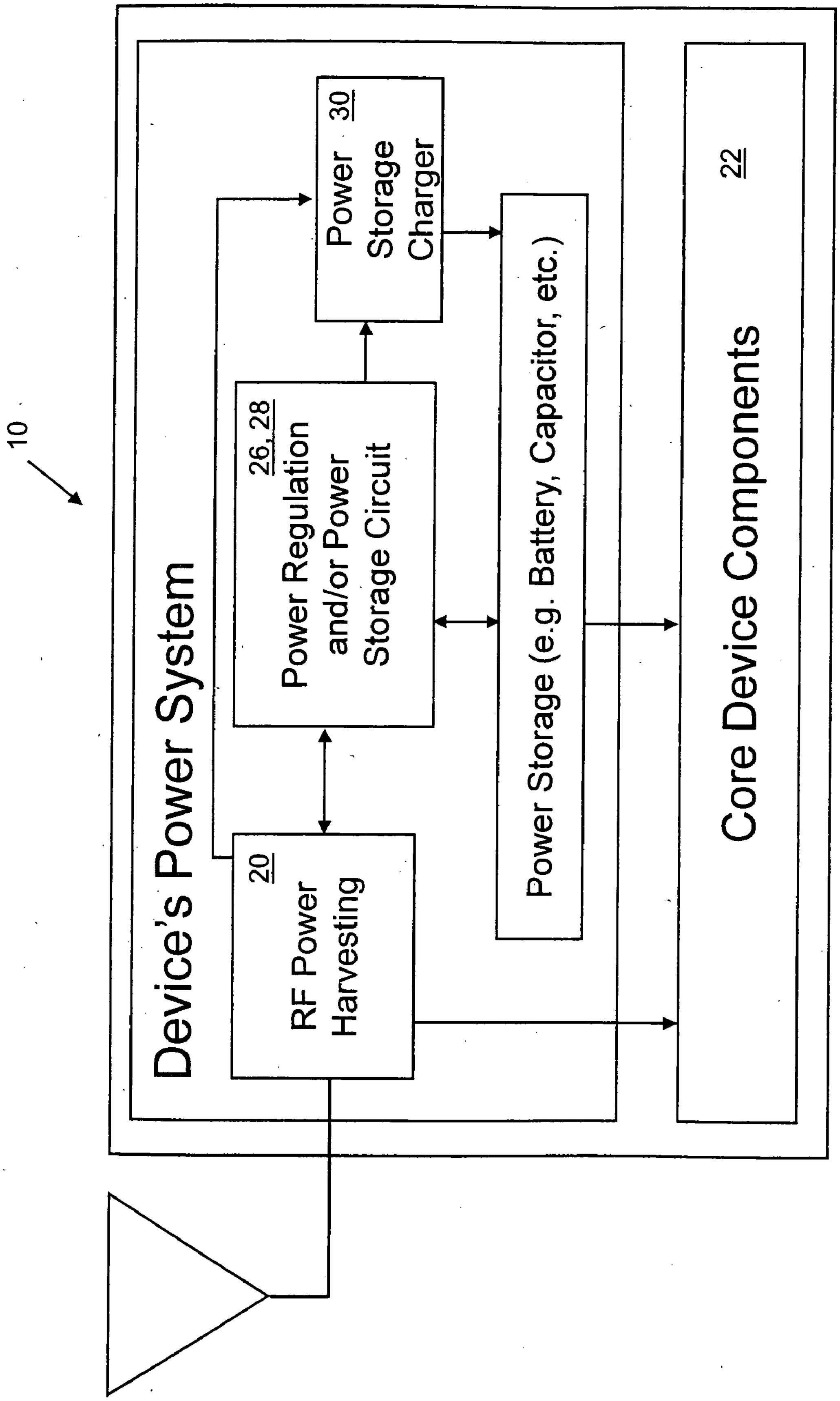
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 44

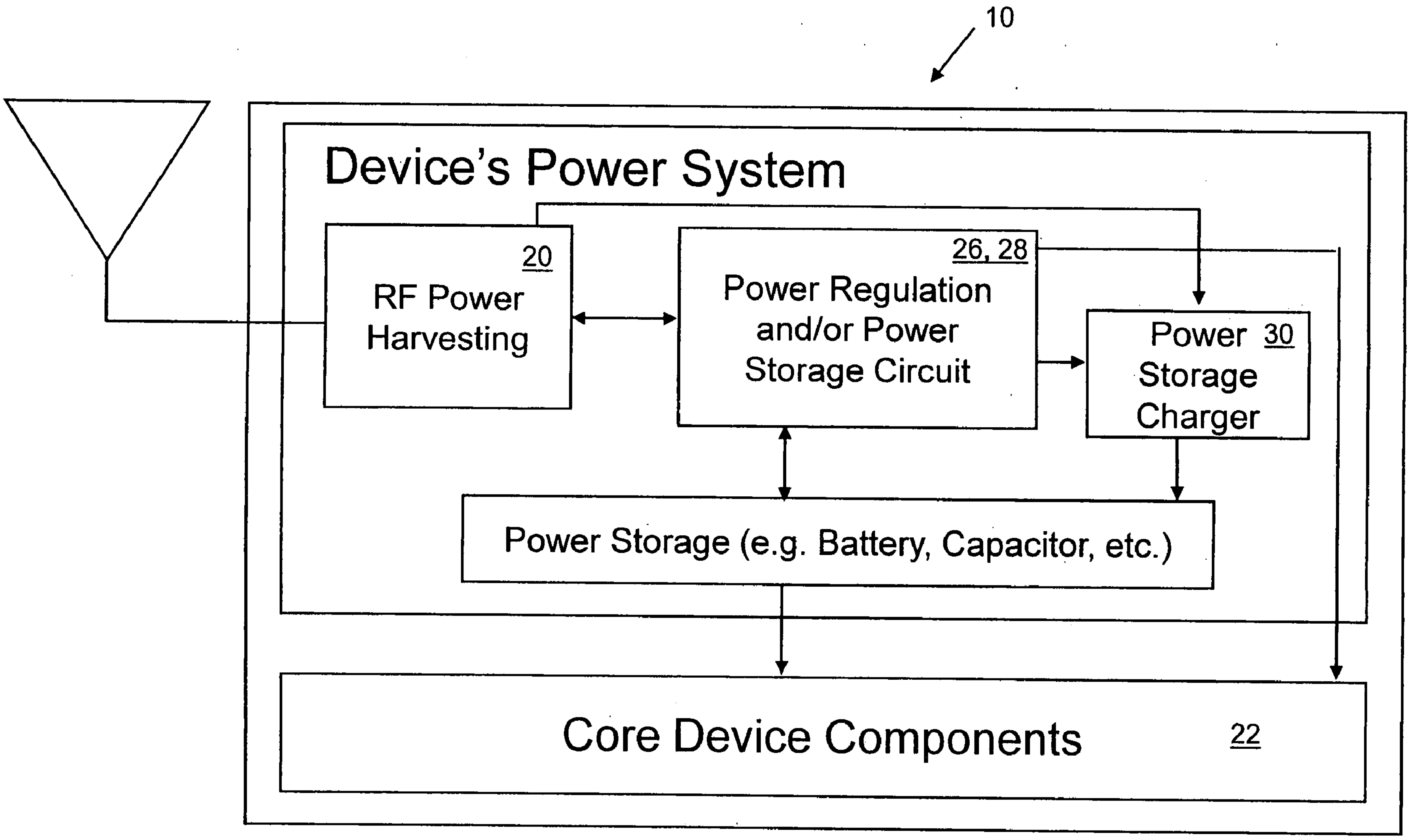
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 45

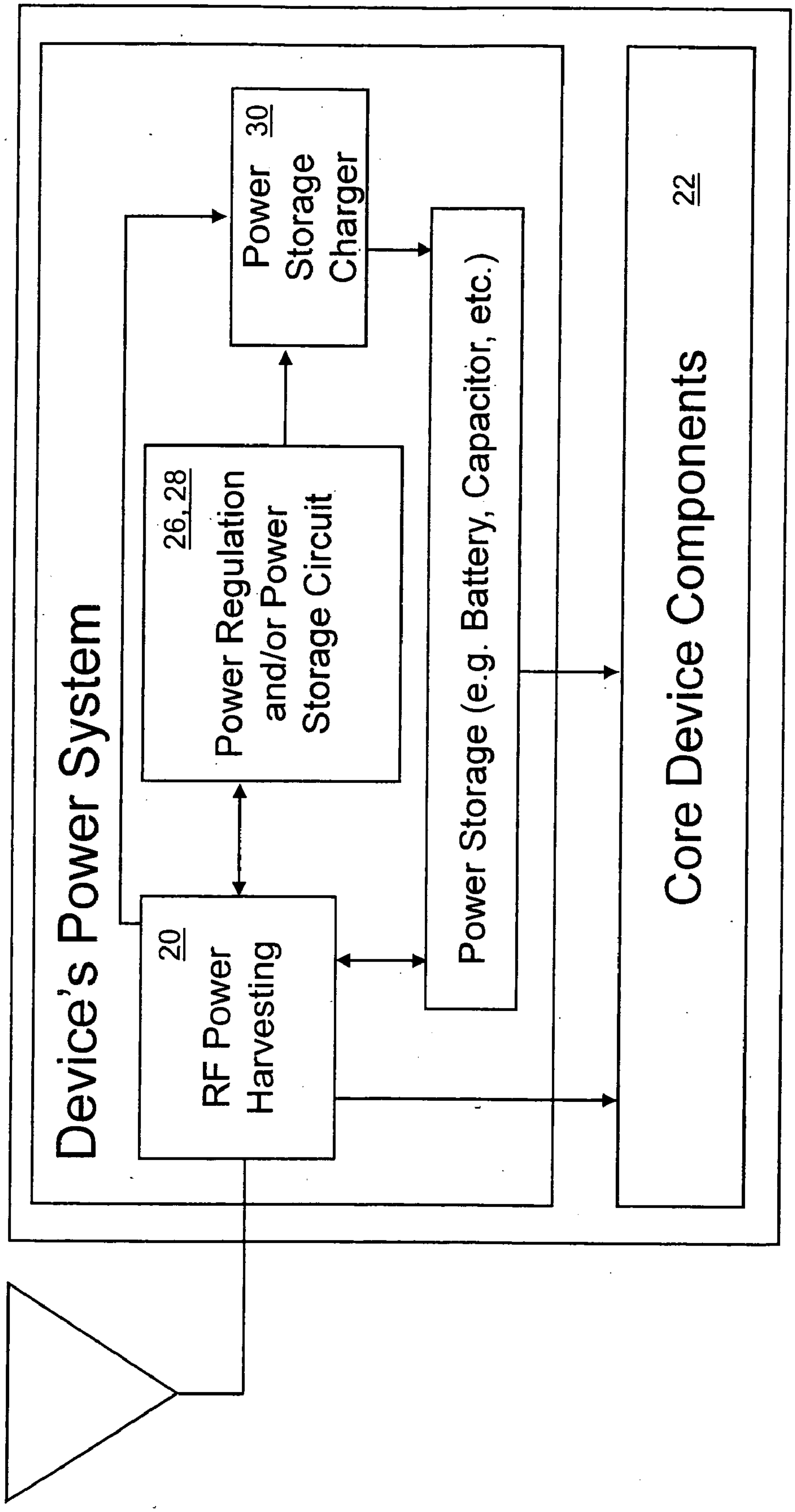
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage
Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 46

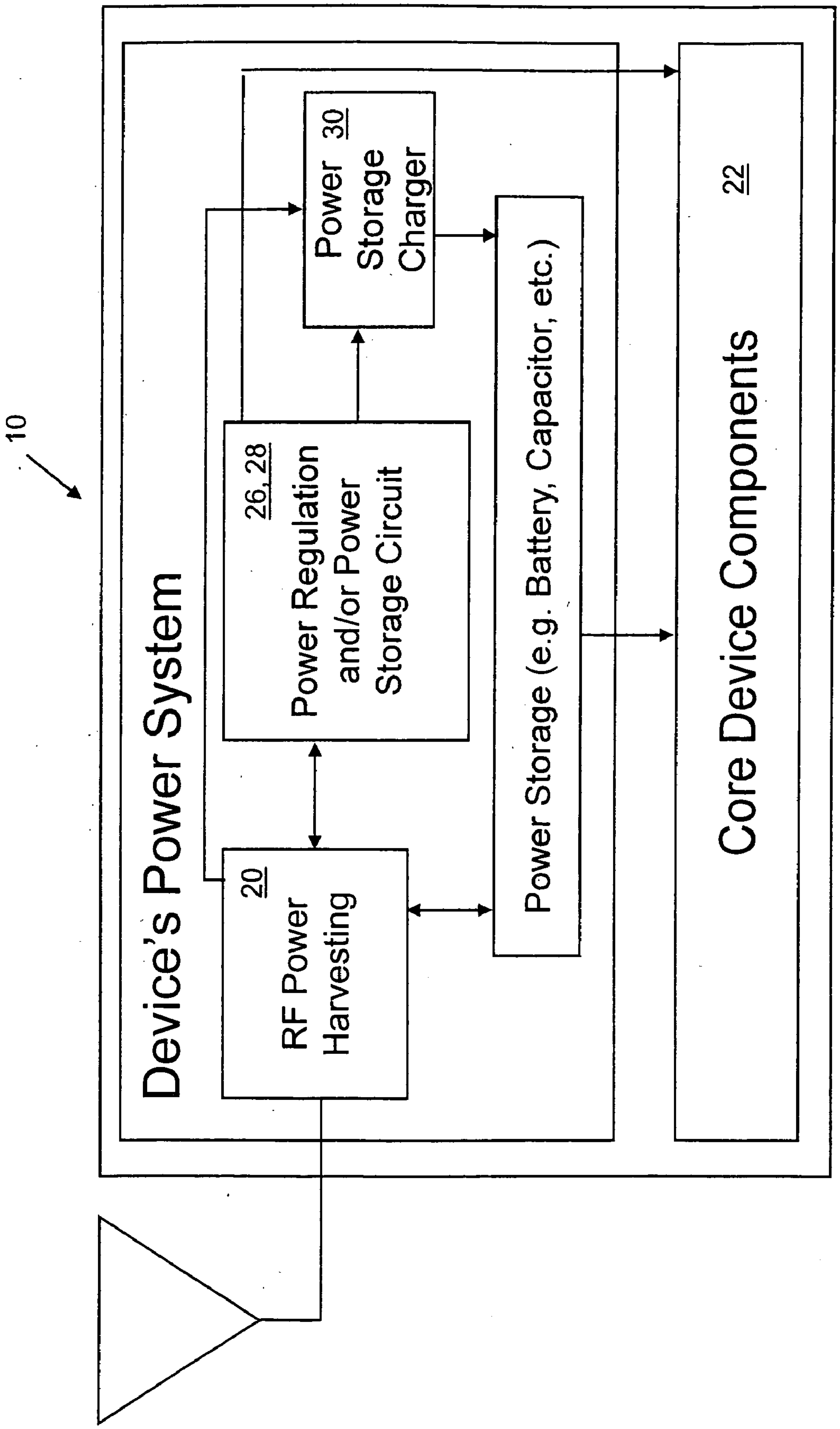
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage
Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 47

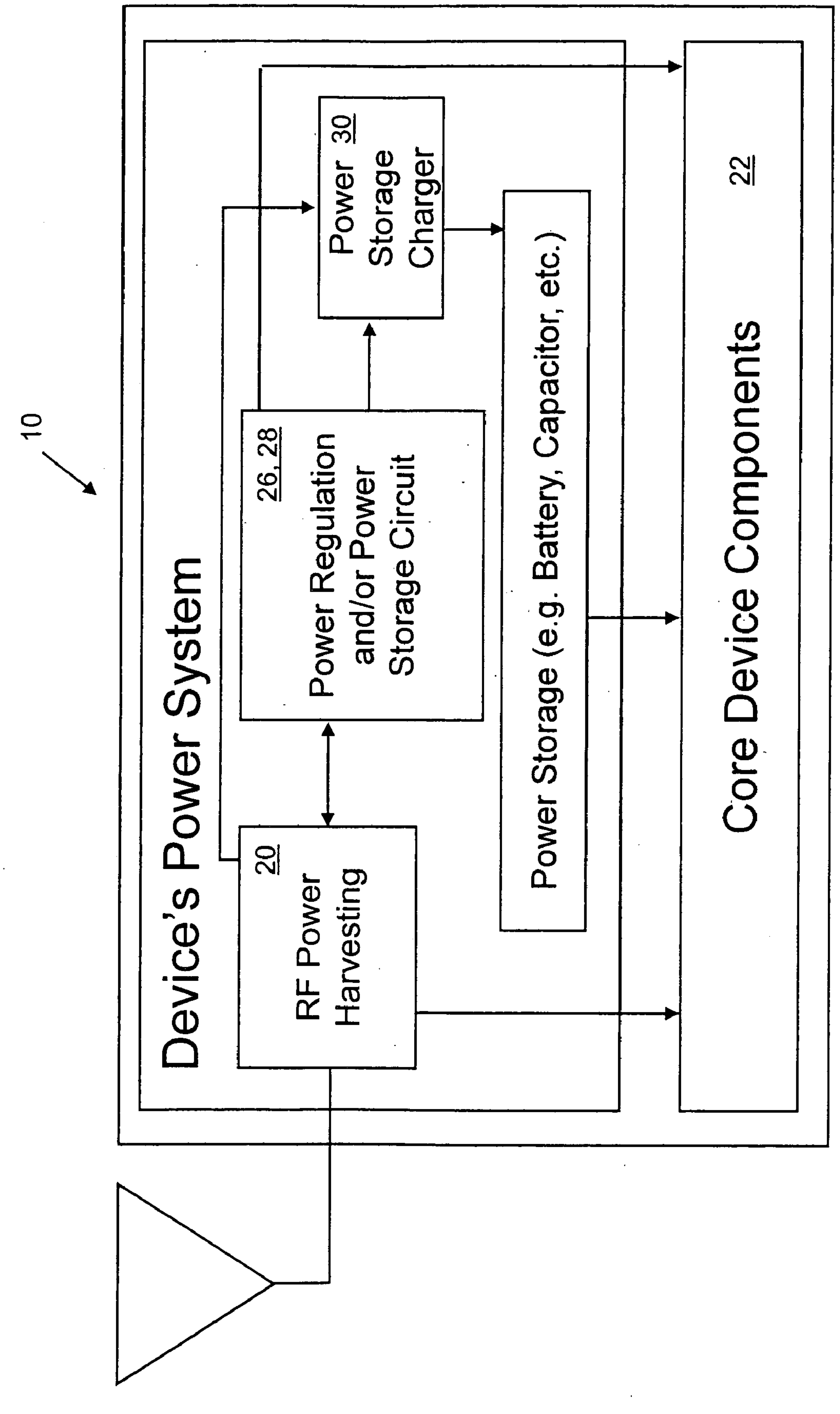
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage
Charger
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 48

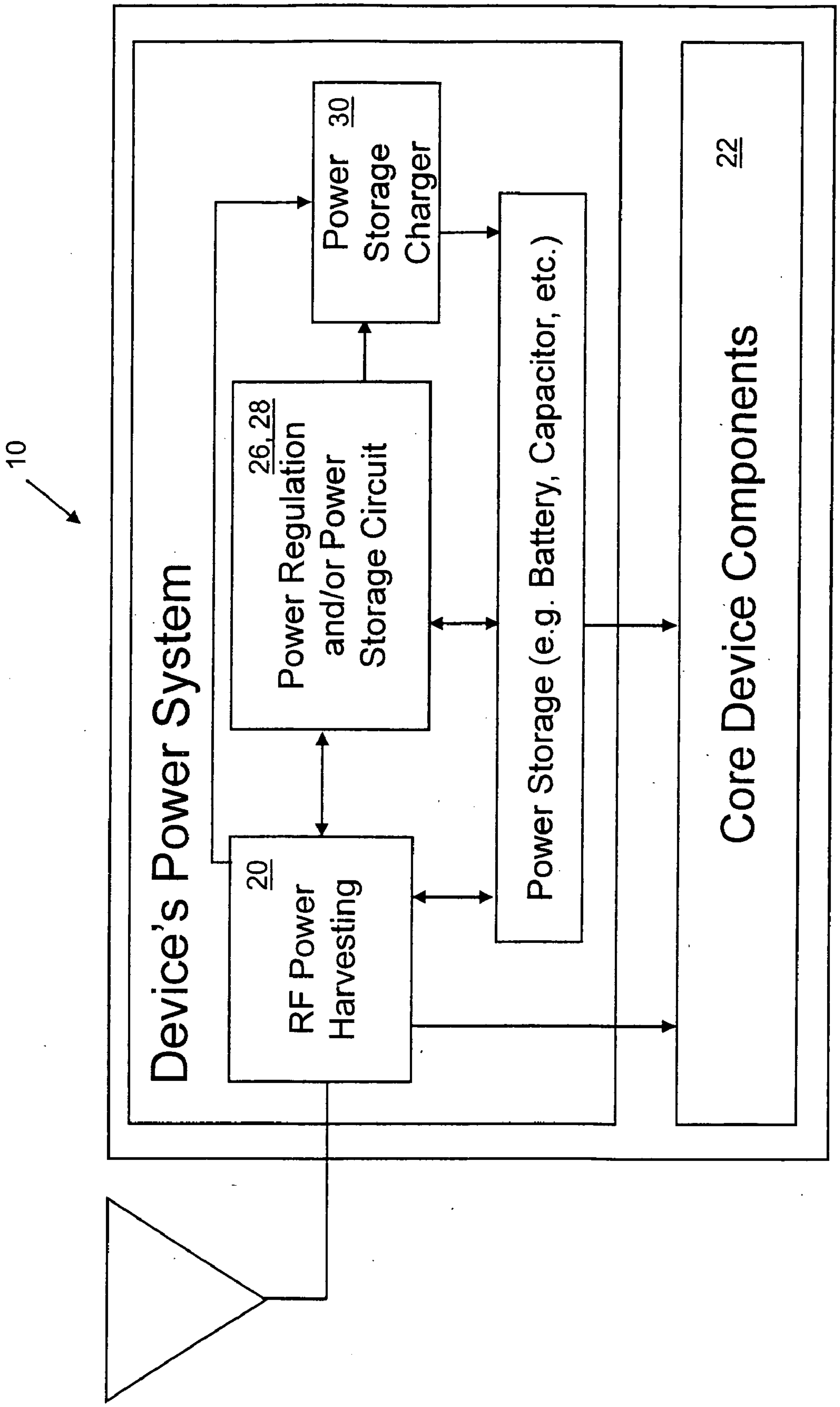
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
30
Power Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 49

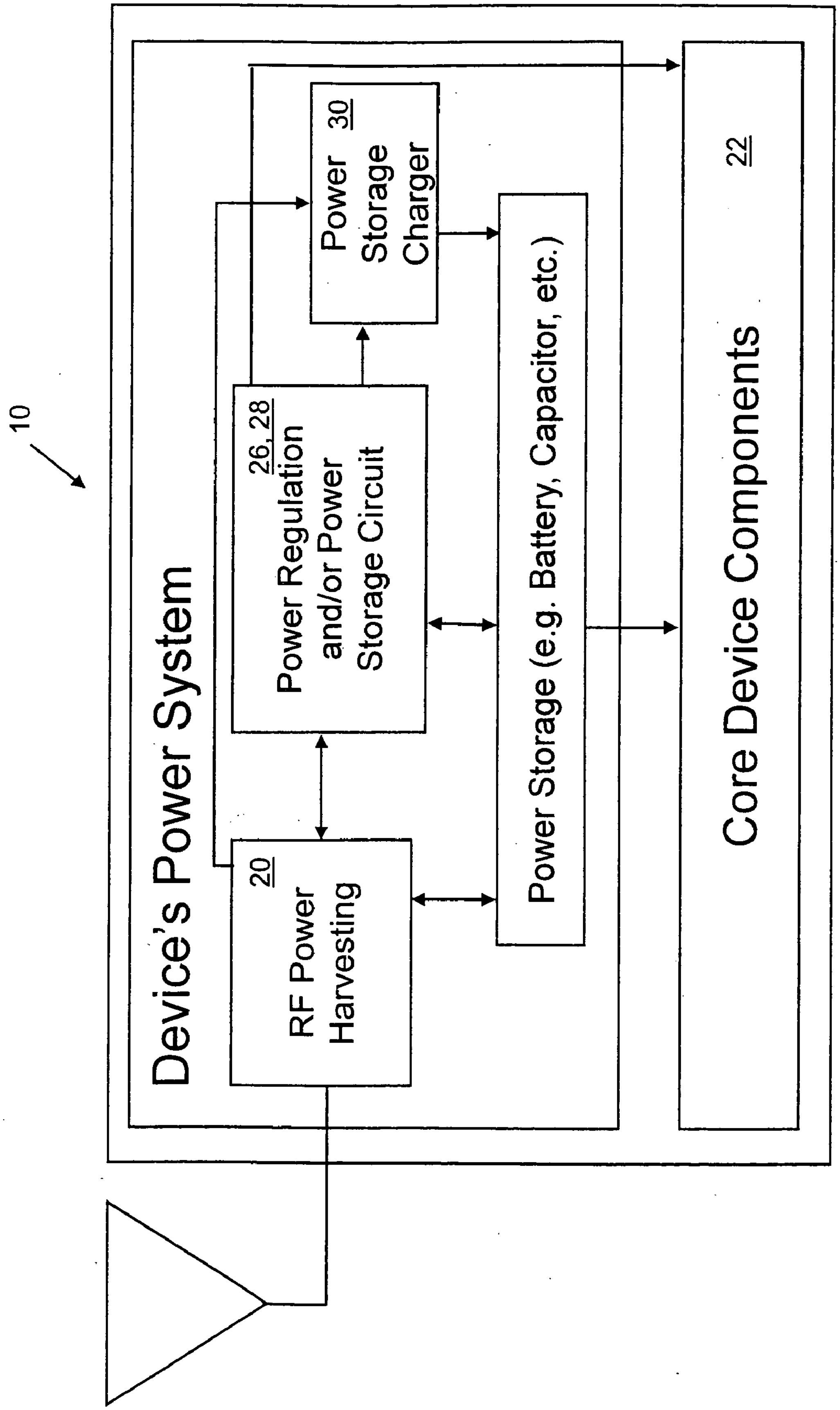
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 50

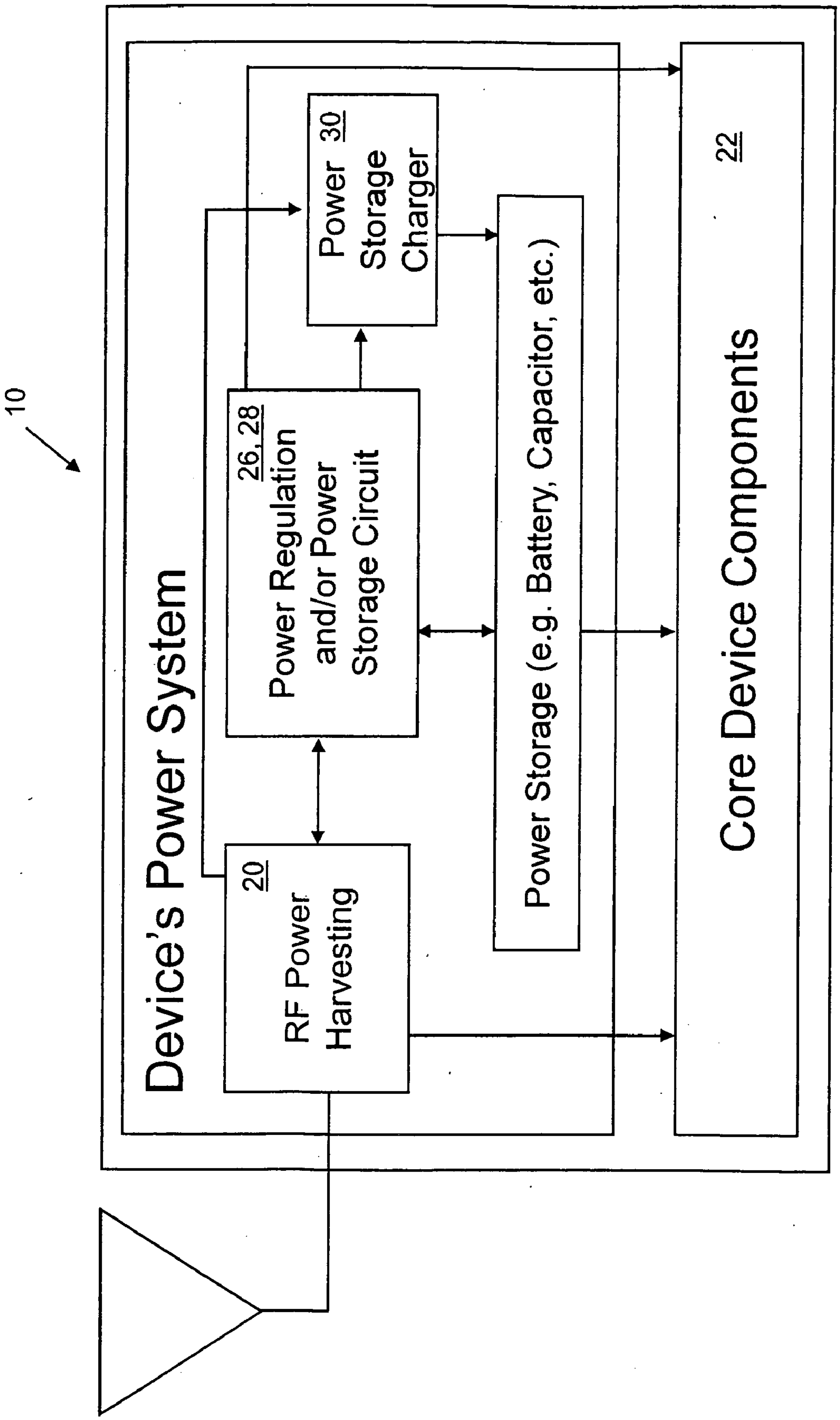
10
Device's Power System
20
RF Power Harvesting
26, 28
Power Regulation and/or Power Storage Circuit
Power 30
Storage Charger
Power Storage (e.g. Battery, Capacitor, etc.)
22
Core Device Components

Fig. 51

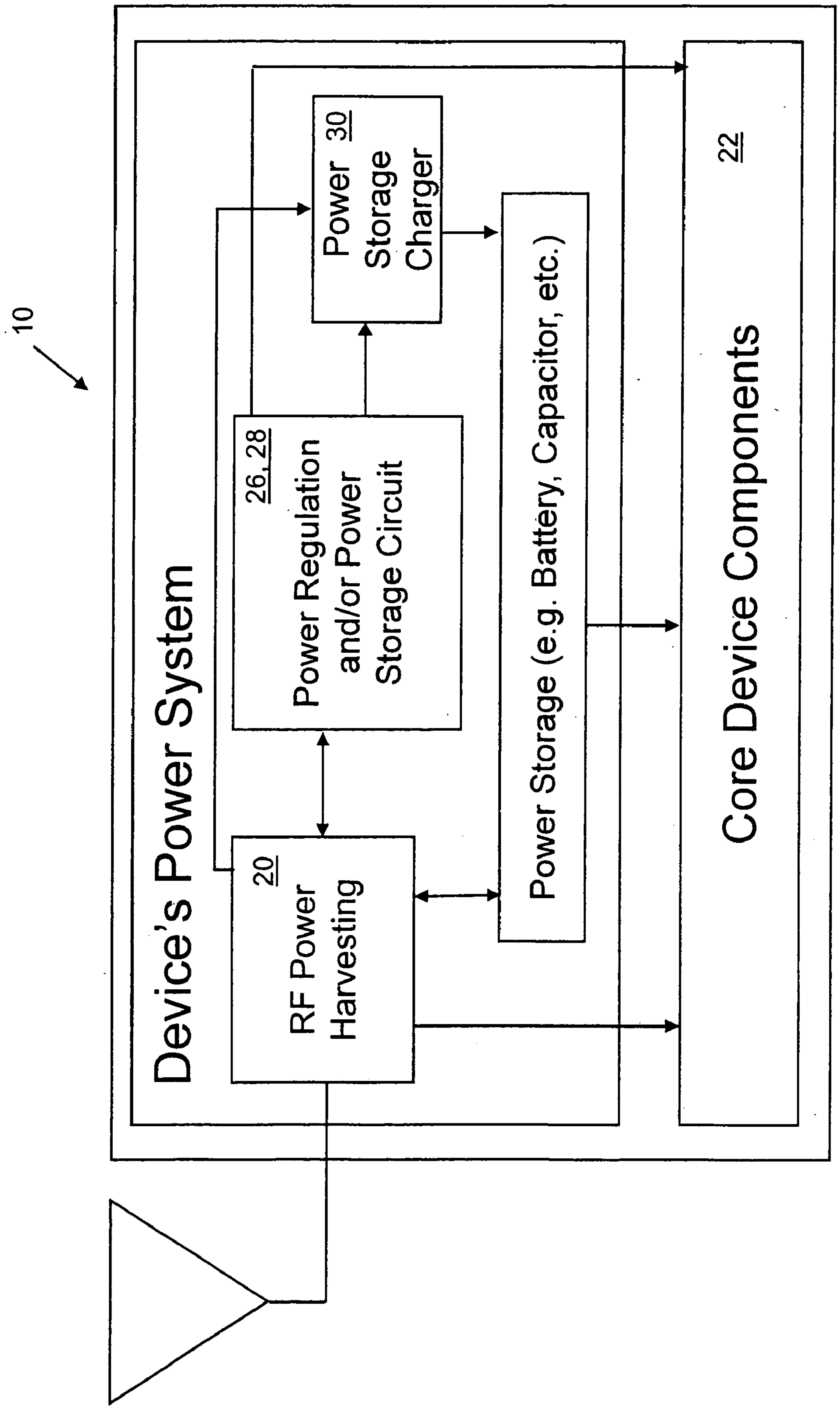
10
Device's Power System
RF Power Harvesting 20
Power Regulation and/or Power Storage Circuit 26, 28
Power Storage Charger 30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components 22

Fig. 52

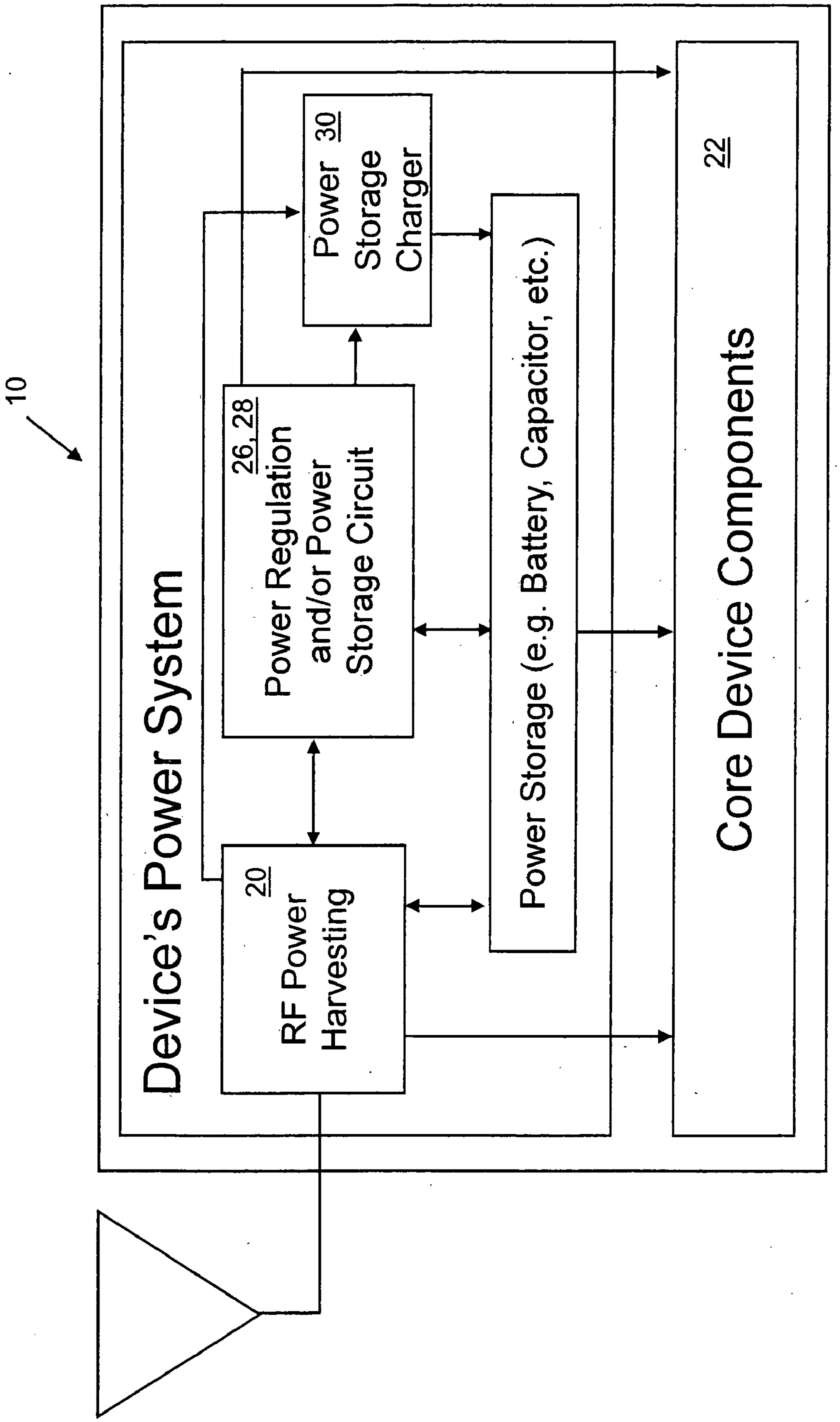
10
Device's Power System
RF Power Harvesting
20
Power Regulation and/or Power Storage Circuit
26, 28
Power Storage Charger
30
Power Storage (e.g. Battery, Capacitor, etc.)
Core Device Components
22

Fig. 53

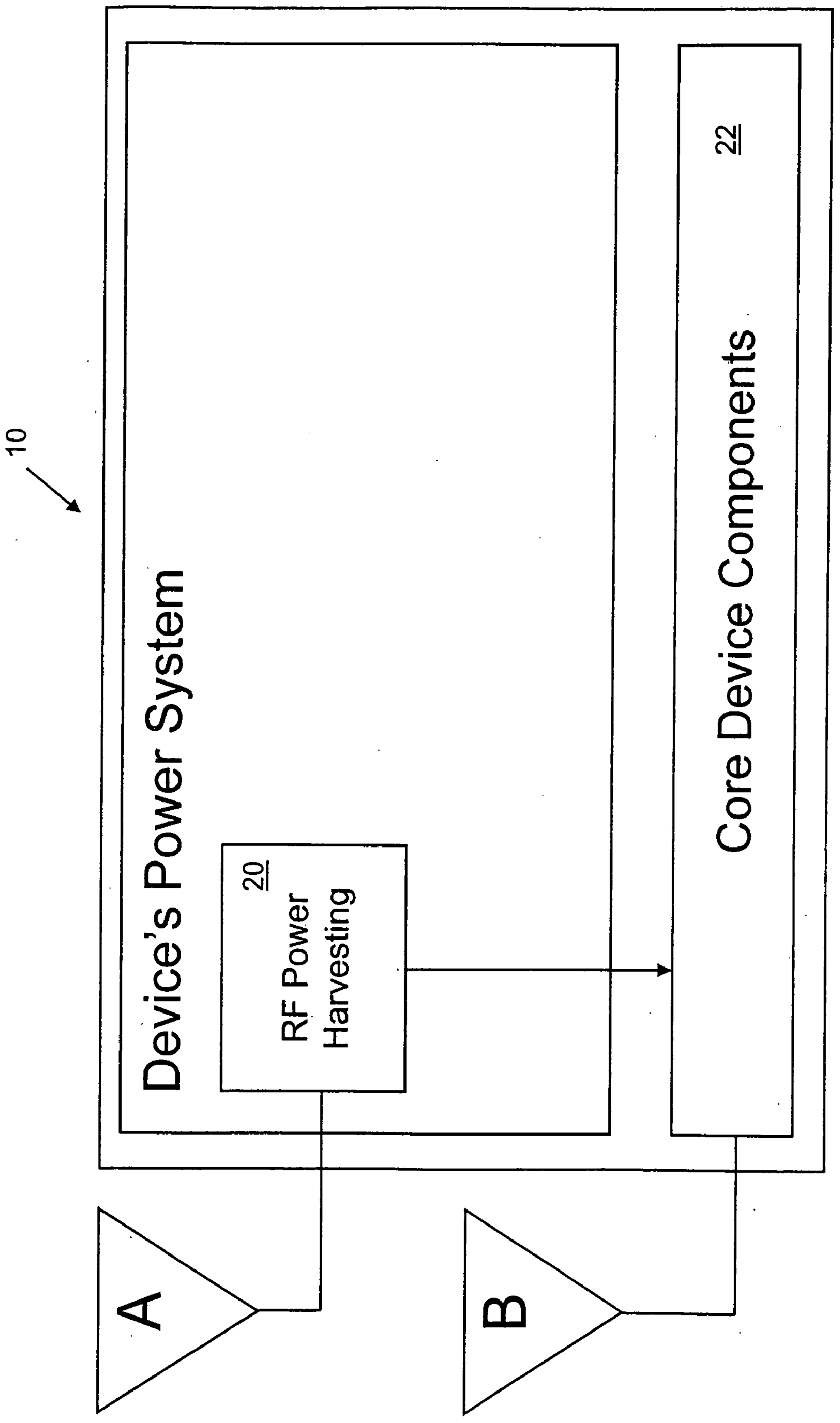
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
A
B

Fig. 54

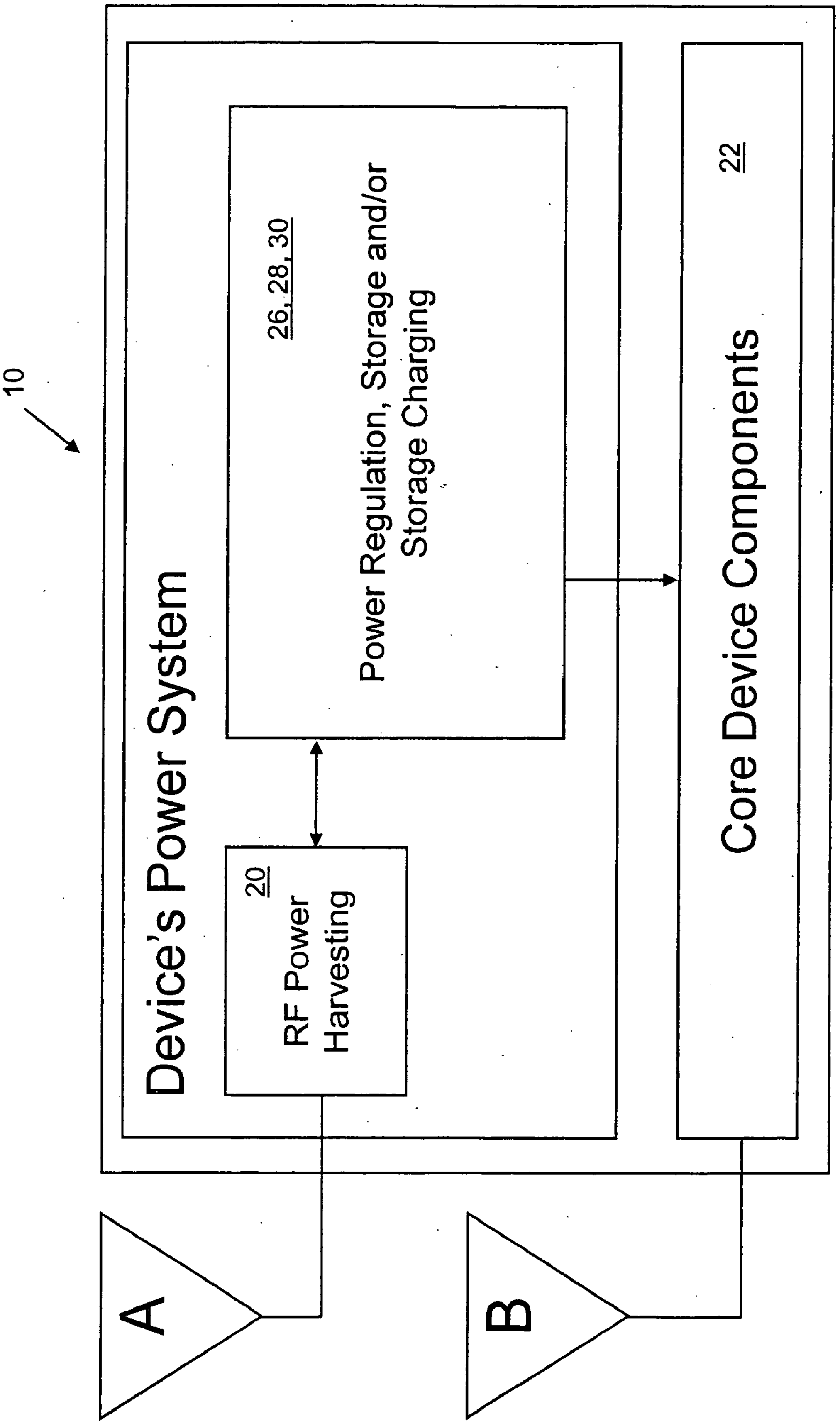
10
Device's Power System
RF Power Harvesting
20
Power Regulation, Storage and/or Storage Charging
26, 28, 30
Core Device Components
22
A
B

Fig. 55

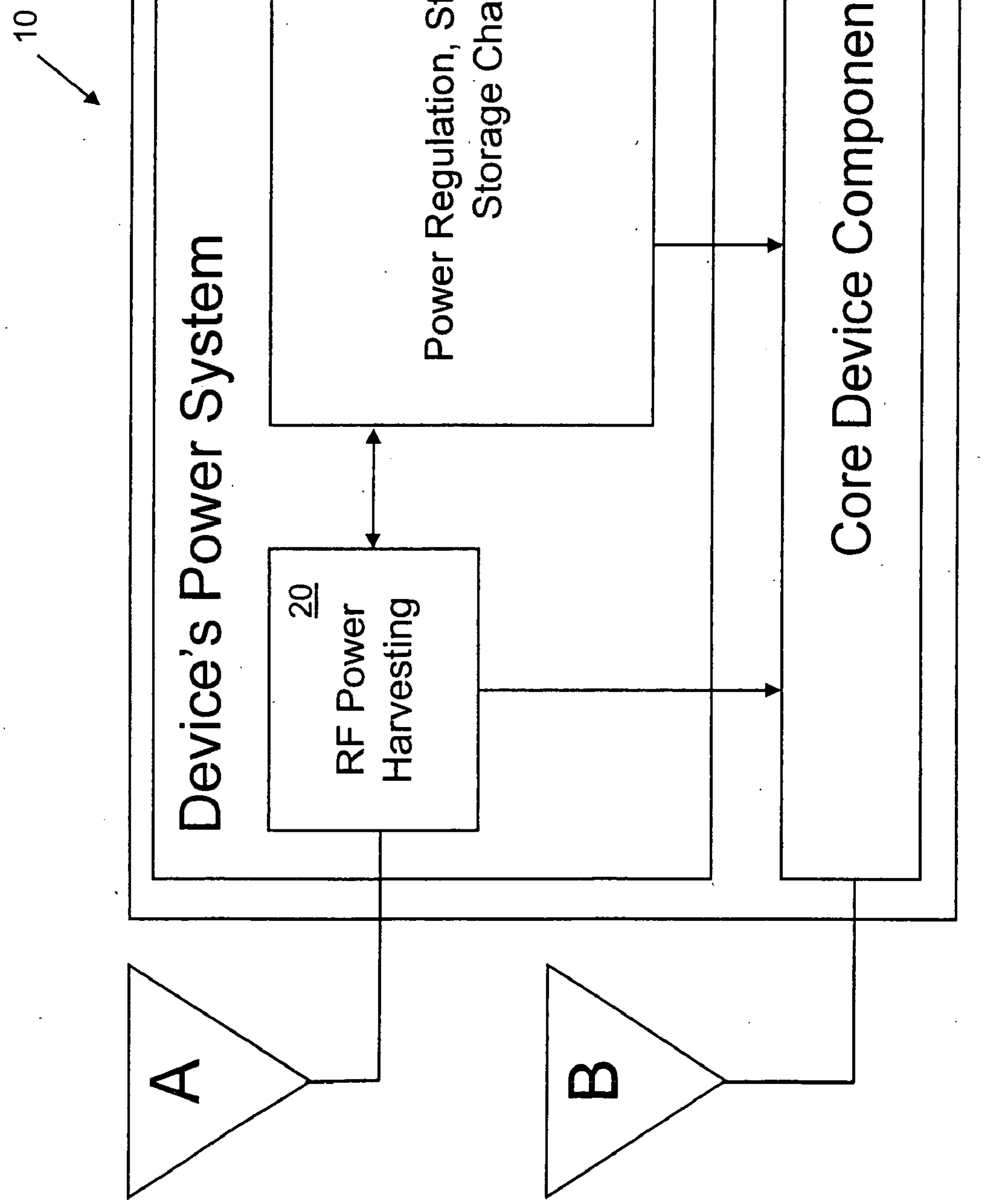
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
22
Core Device Components
A
B

Fig. 56

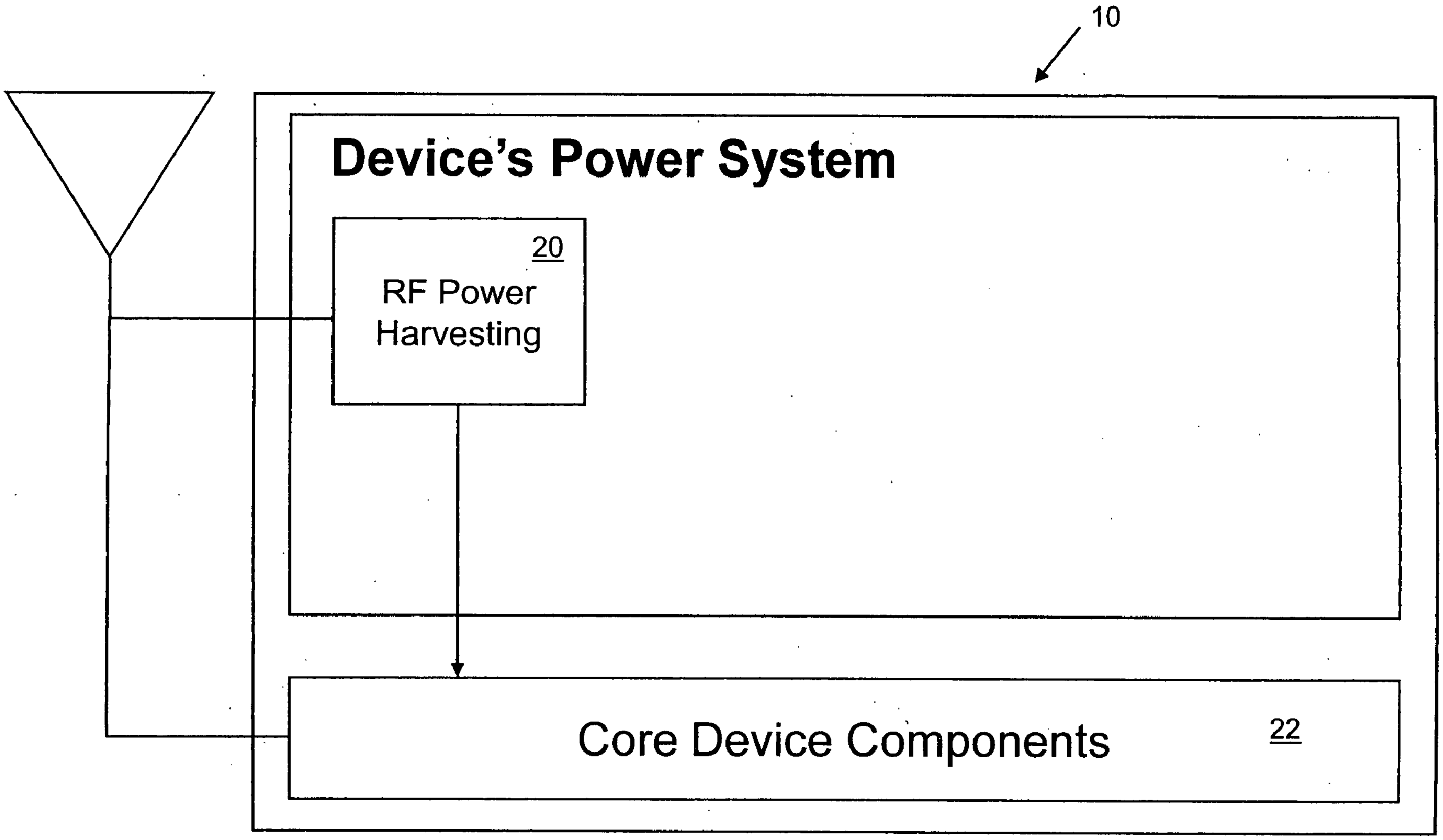
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22

Fig. 57

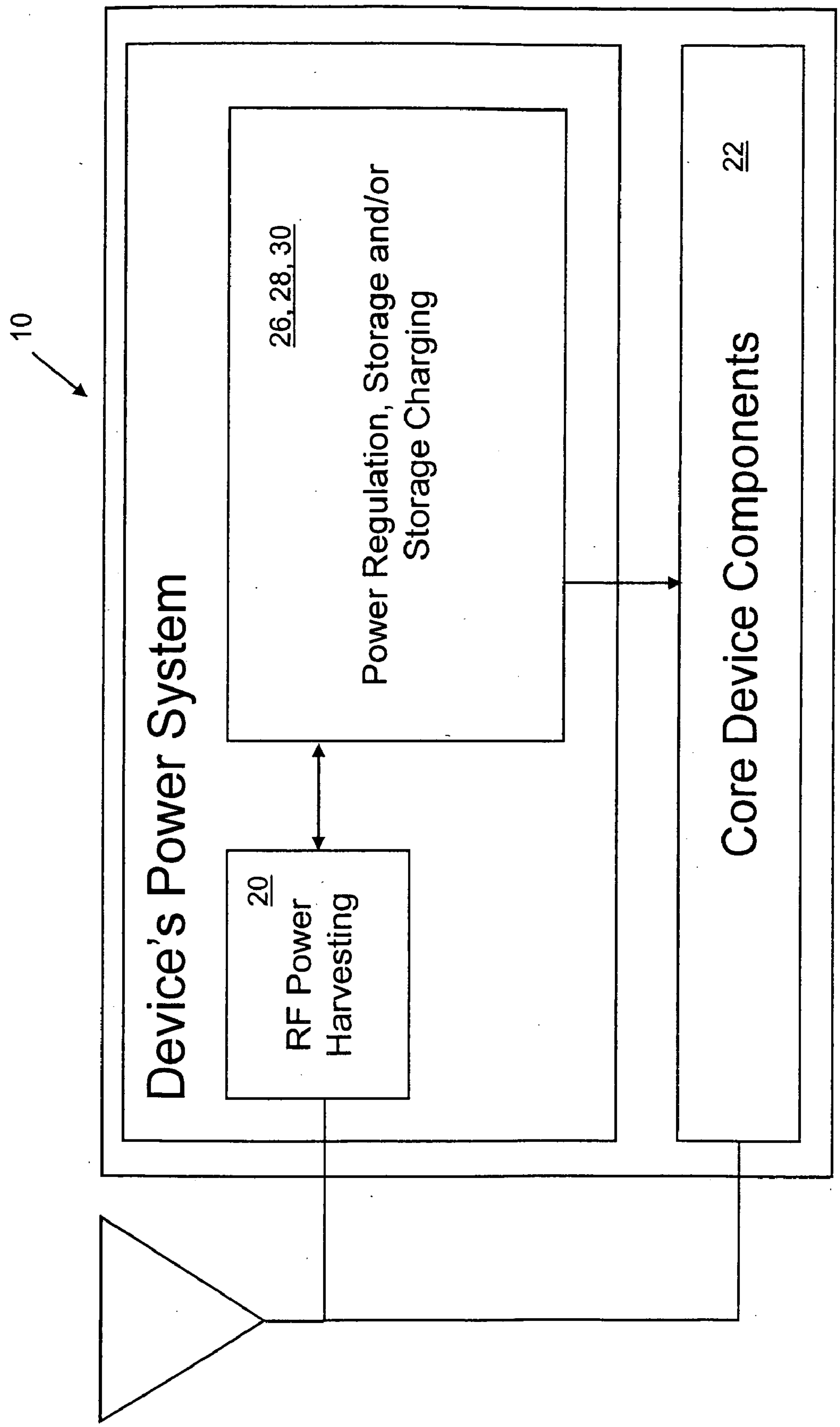
10
Device's Power System
RF Power Harvesting
20
Power Regulation, Storage and/or Storage Charging
26, 28, 30
Core Device Components
22

Fig. 58

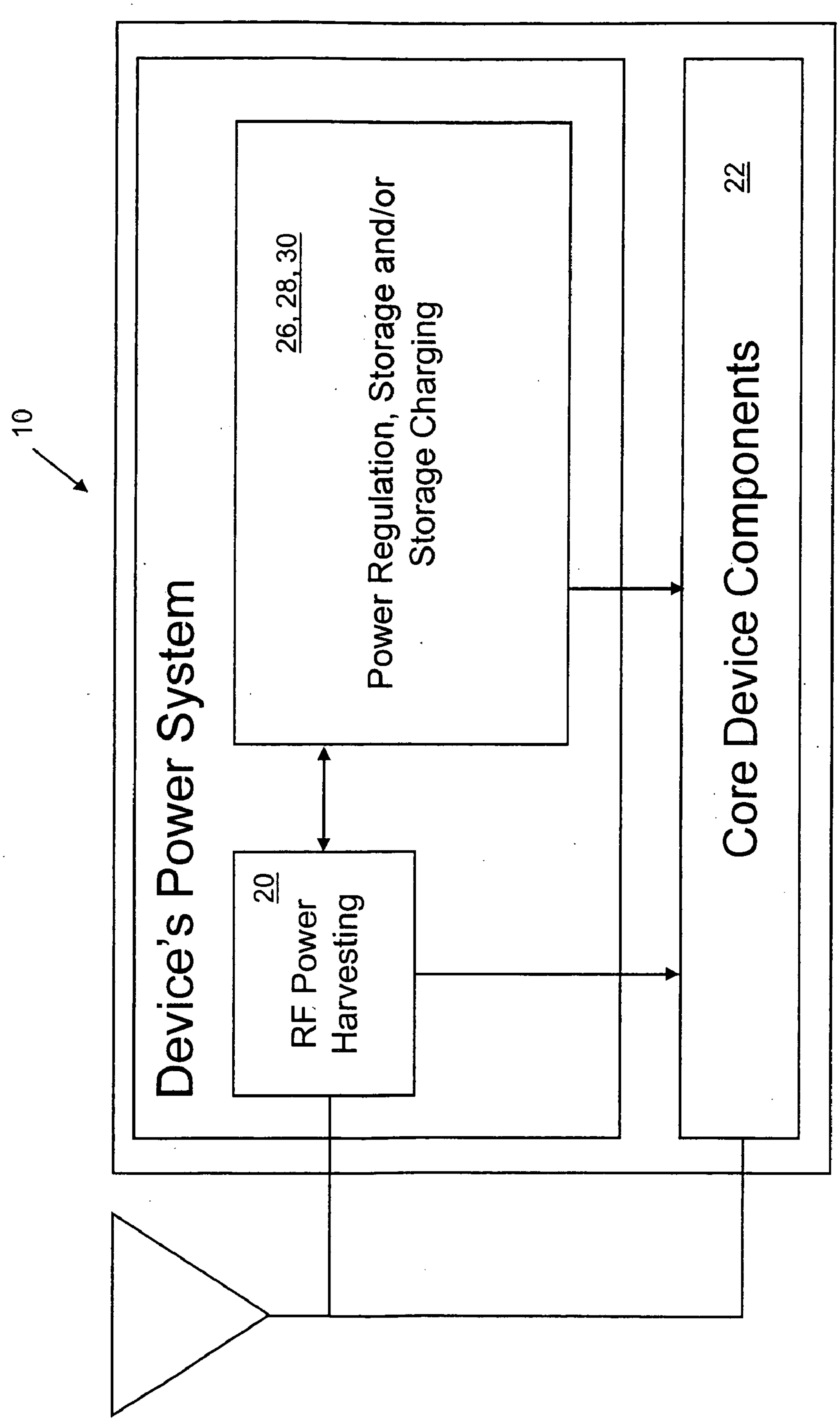
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22

Fig. 59

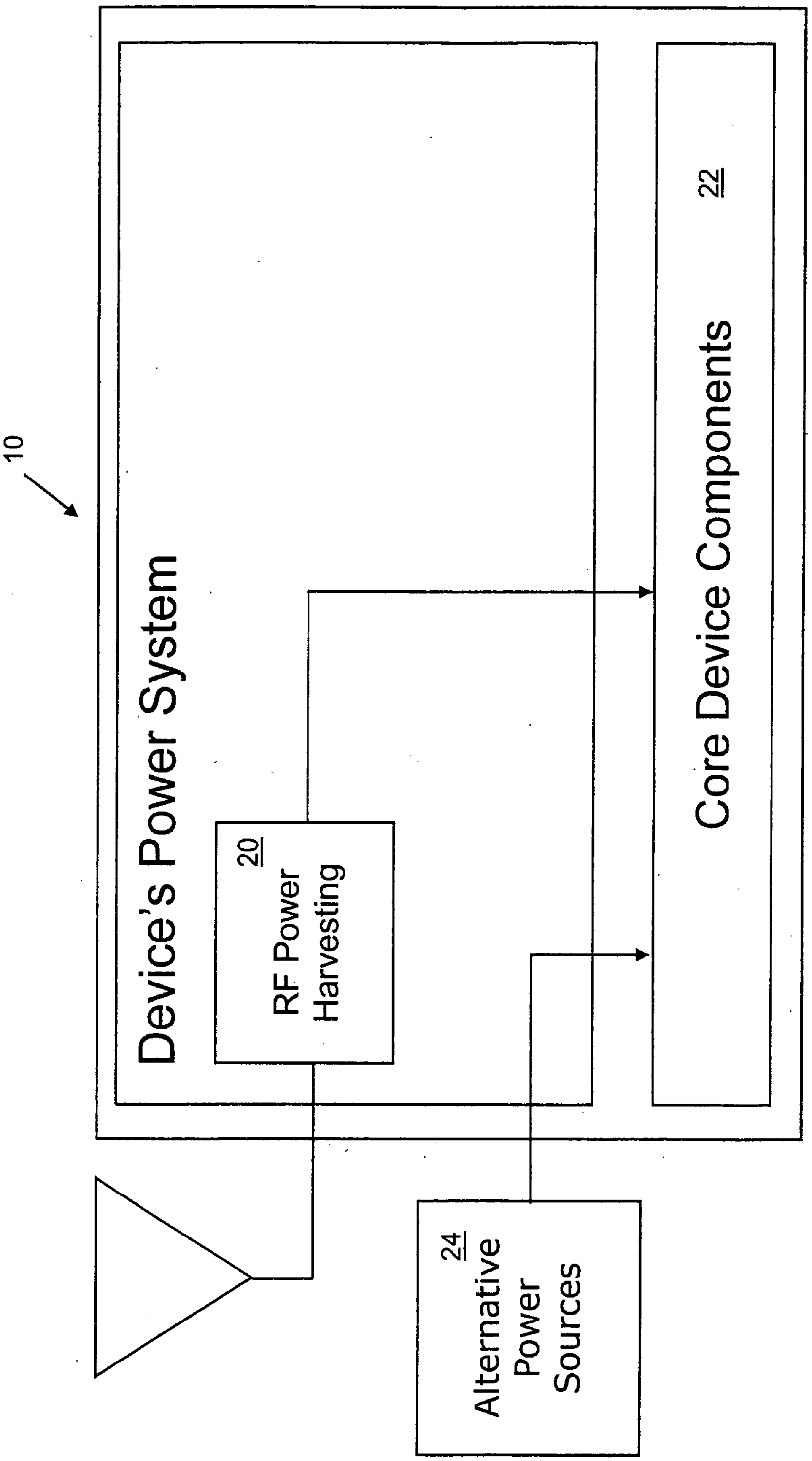
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 60

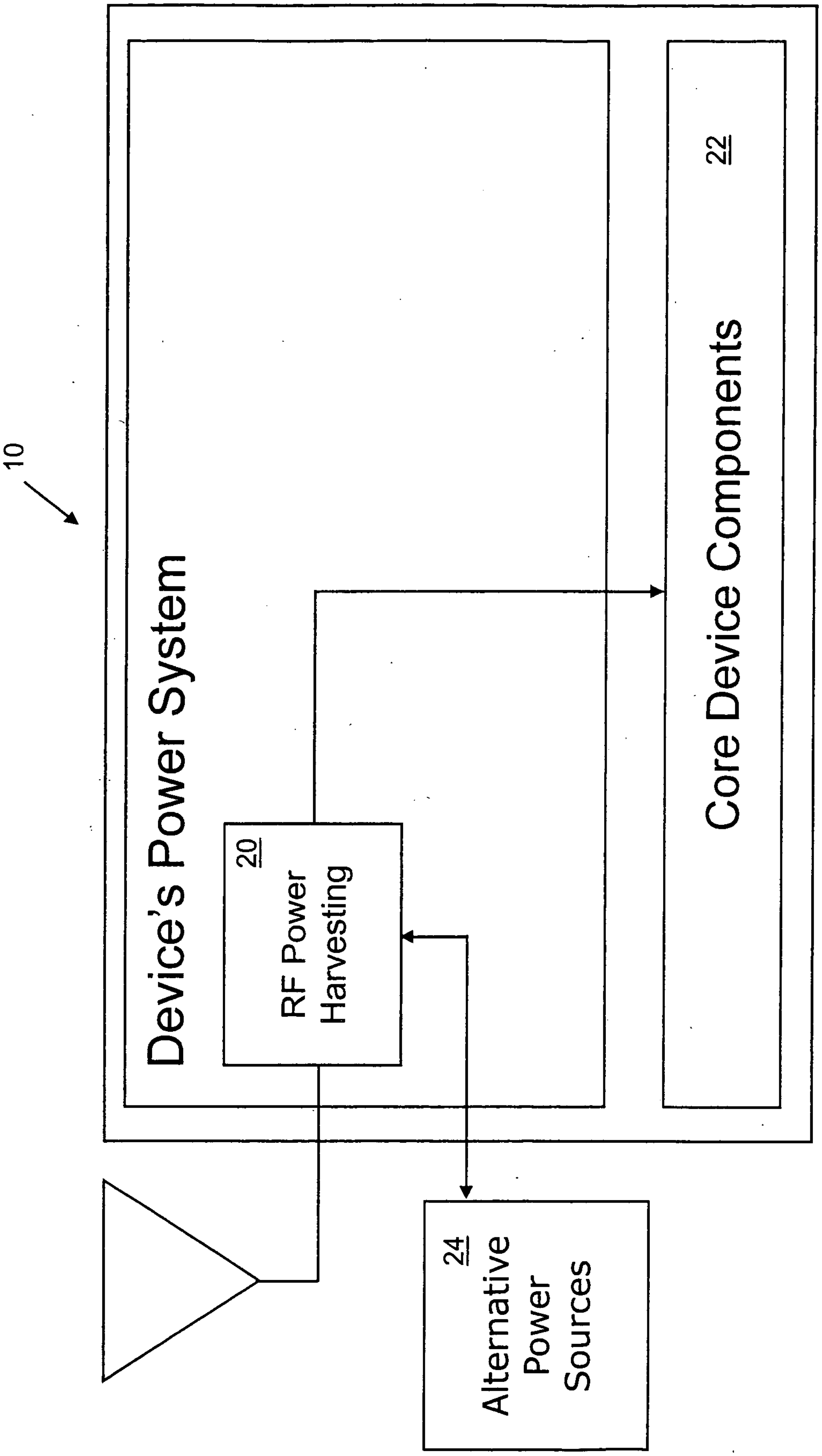
10
Device's Power System
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 61

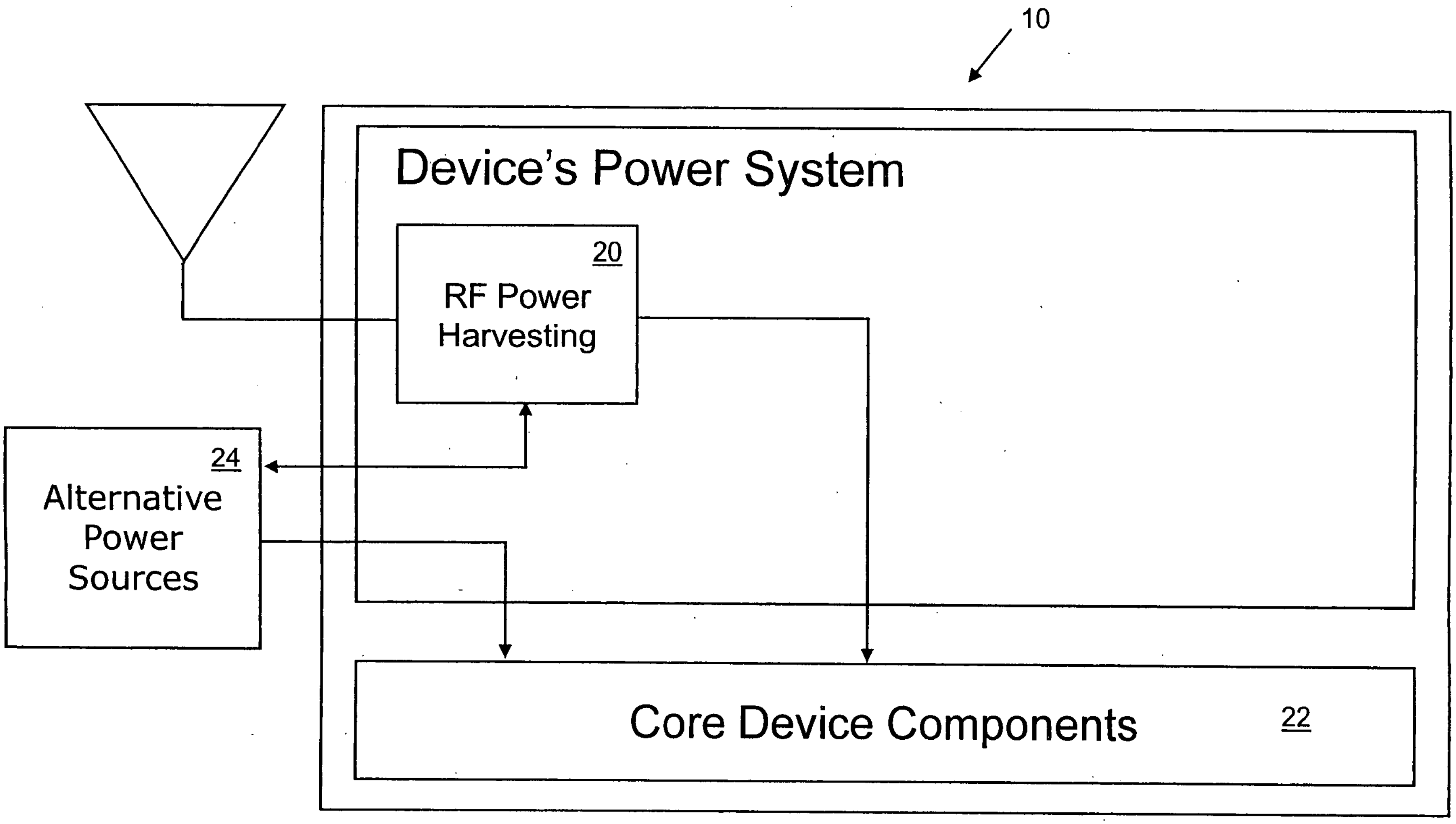
10
Device's Power System
20
RF Power Harvesting
24
Alternative Power Sources
Core Device Components
22

Fig. 62

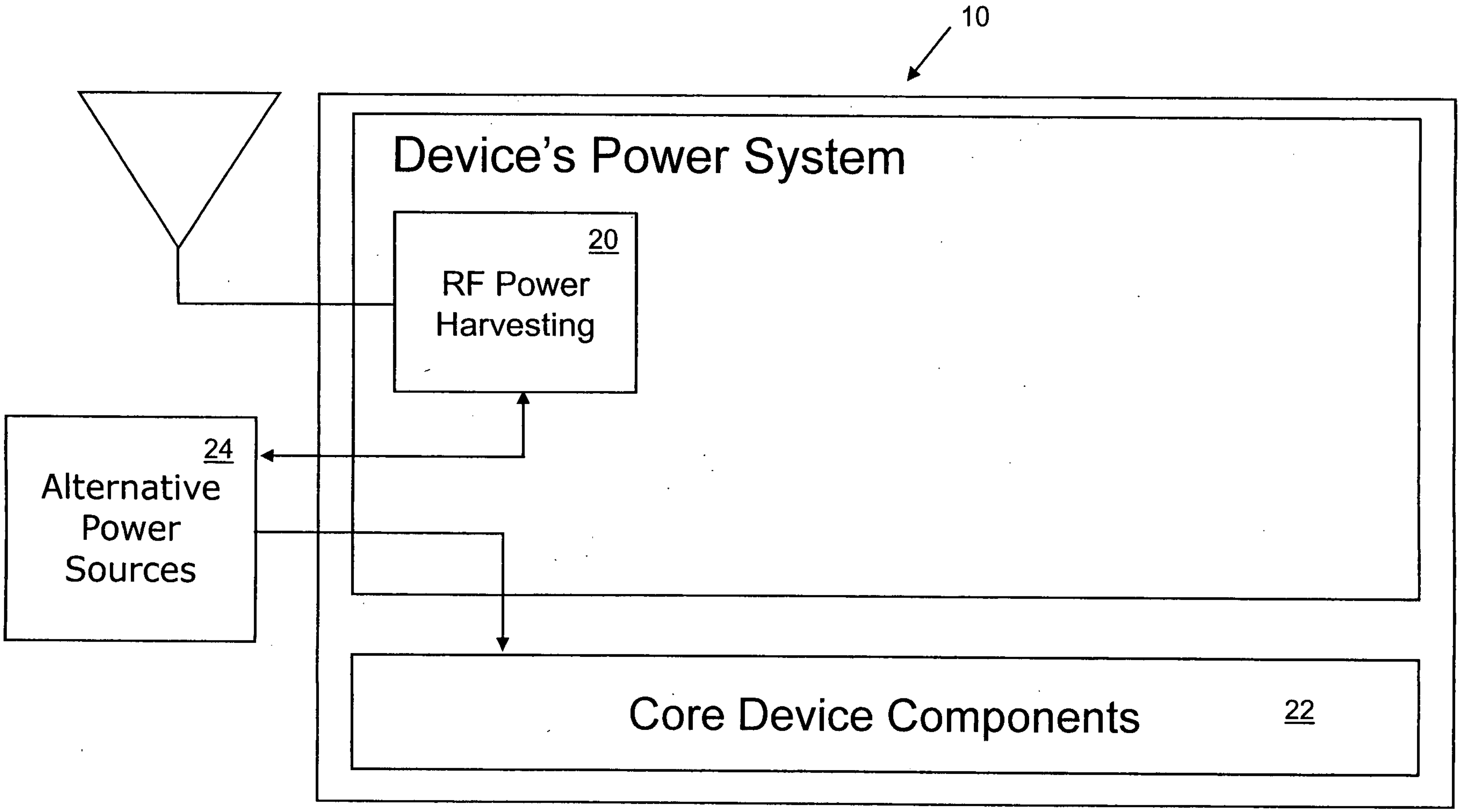
10
Device's Power System
20
RF Power Harvesting
24
Alternative Power Sources
Core Device Components
22

Fig. 63

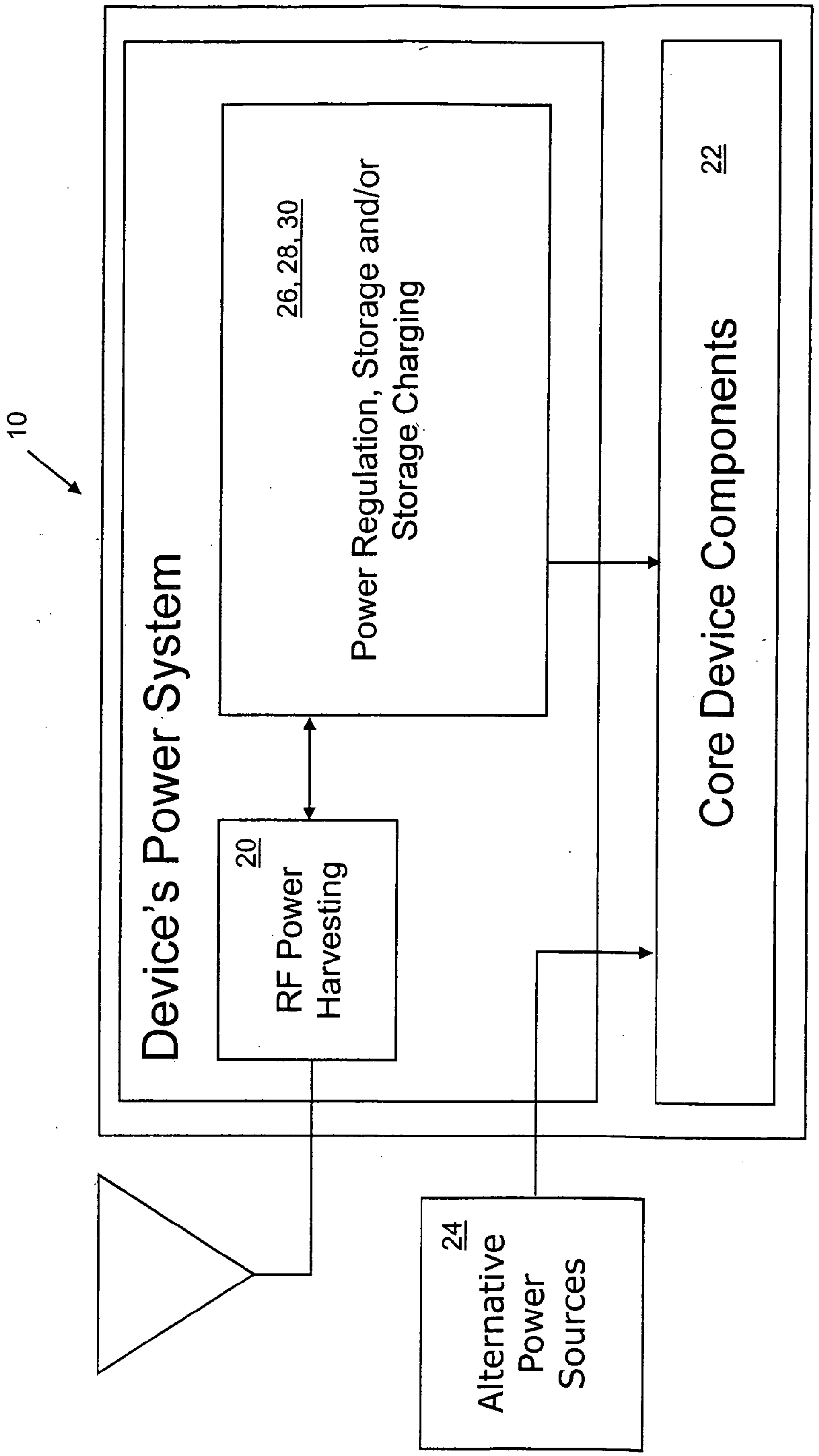
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
22
Core Device Components
24
Alternative Power Sources

Fig. 64

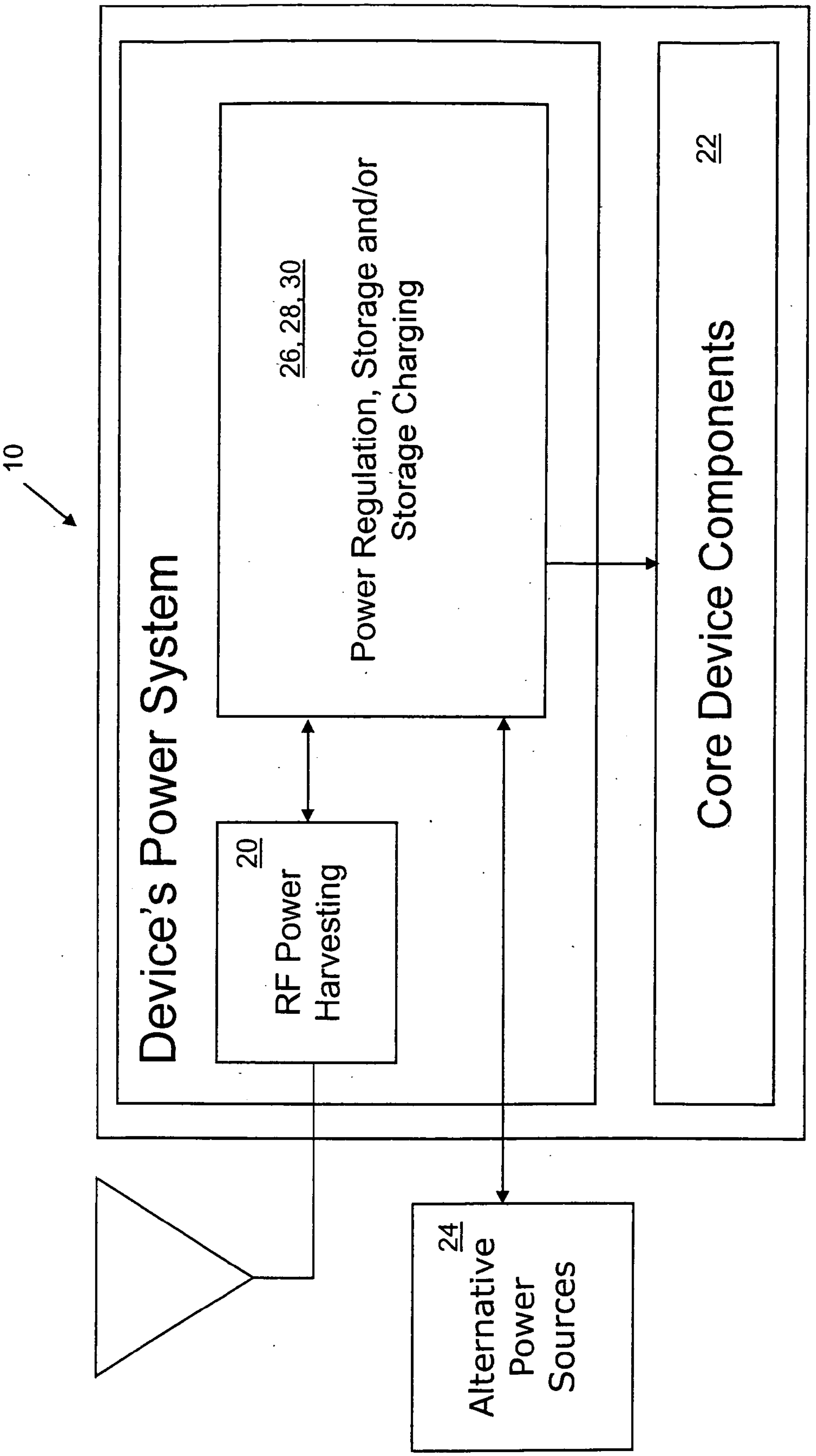
10
Device's Power System
RF Power Harvesting
20
Power Regulation, Storage and/or Storage Charging
26, 28, 30
Core Device Components
22
Alternative Power Sources
24

Fig. 65

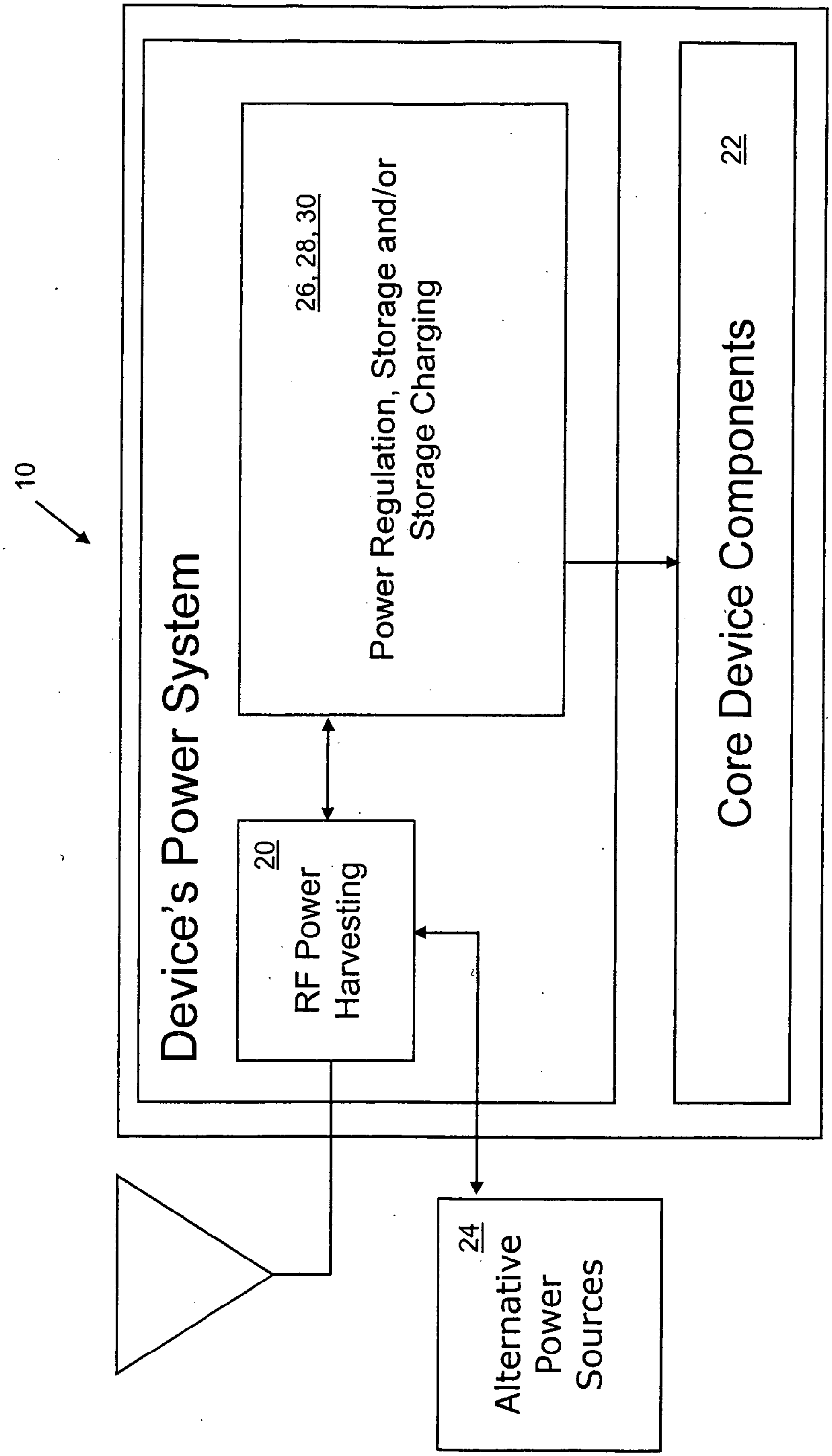
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 66

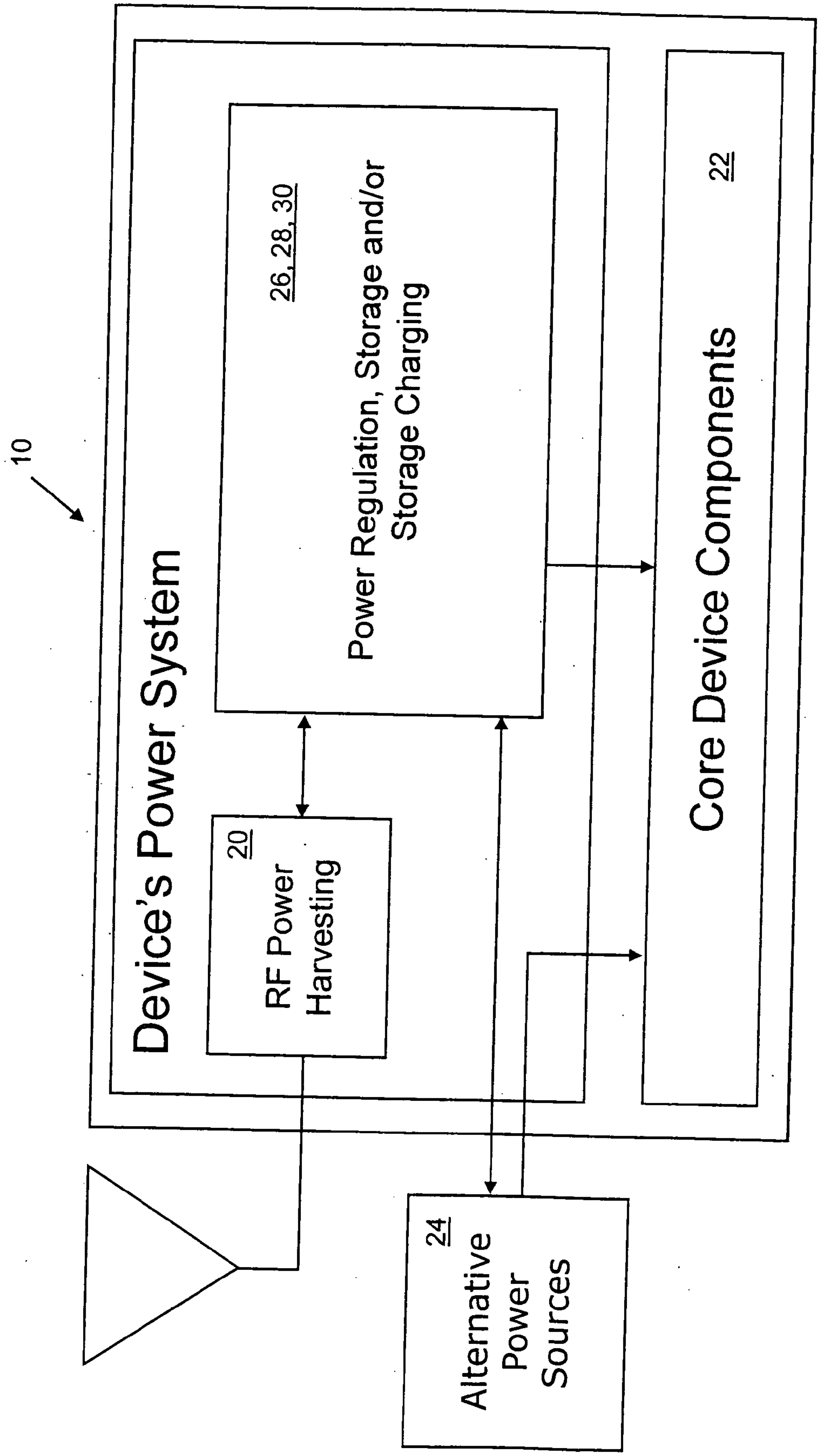
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 67

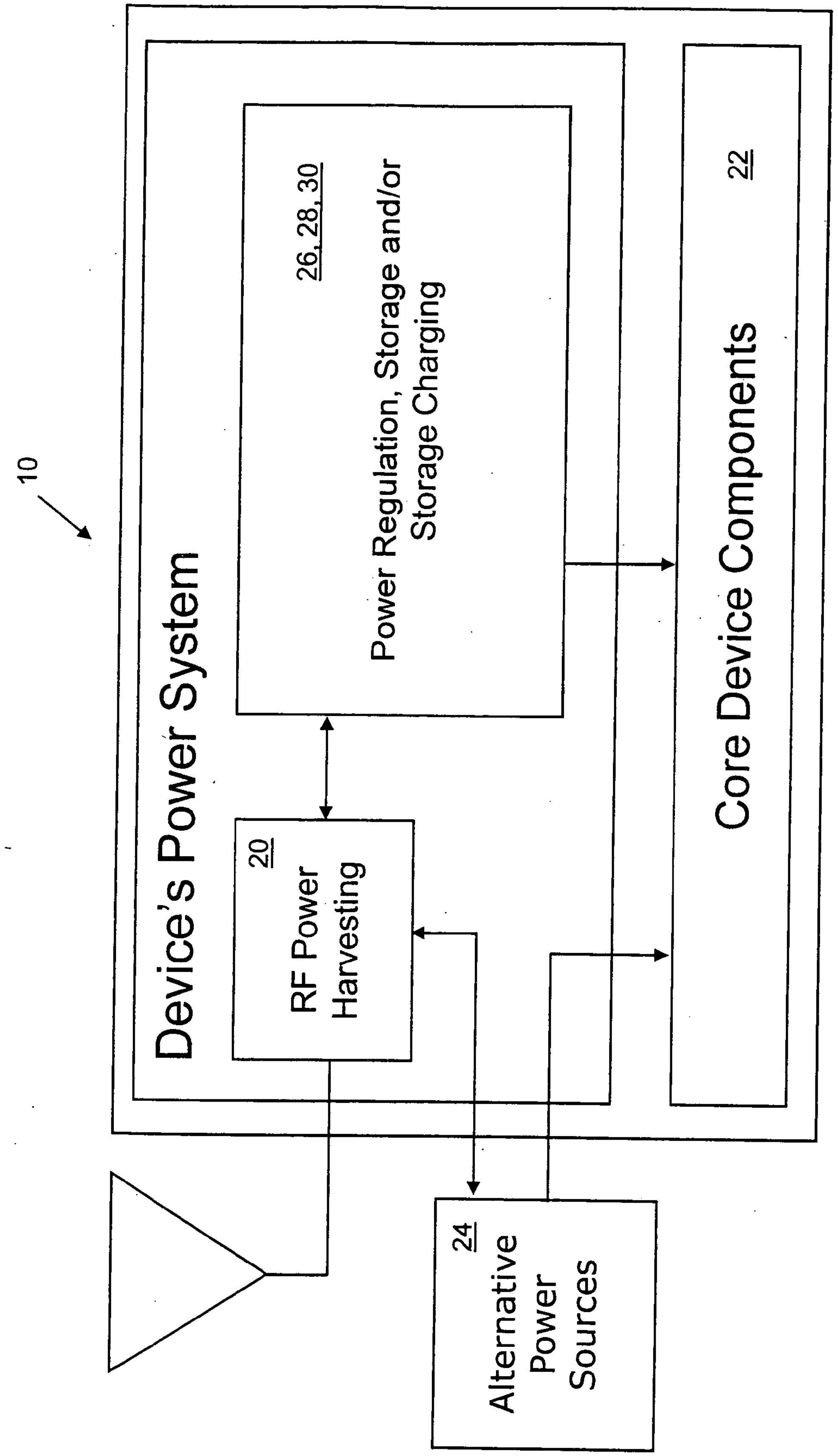
10
Device's Power System
RF Power Harvesting
20
Power Regulation, Storage and/or Storage Charging
26, 28, 30
Core Device Components
22
Alternative Power Sources
24

Fig. 68

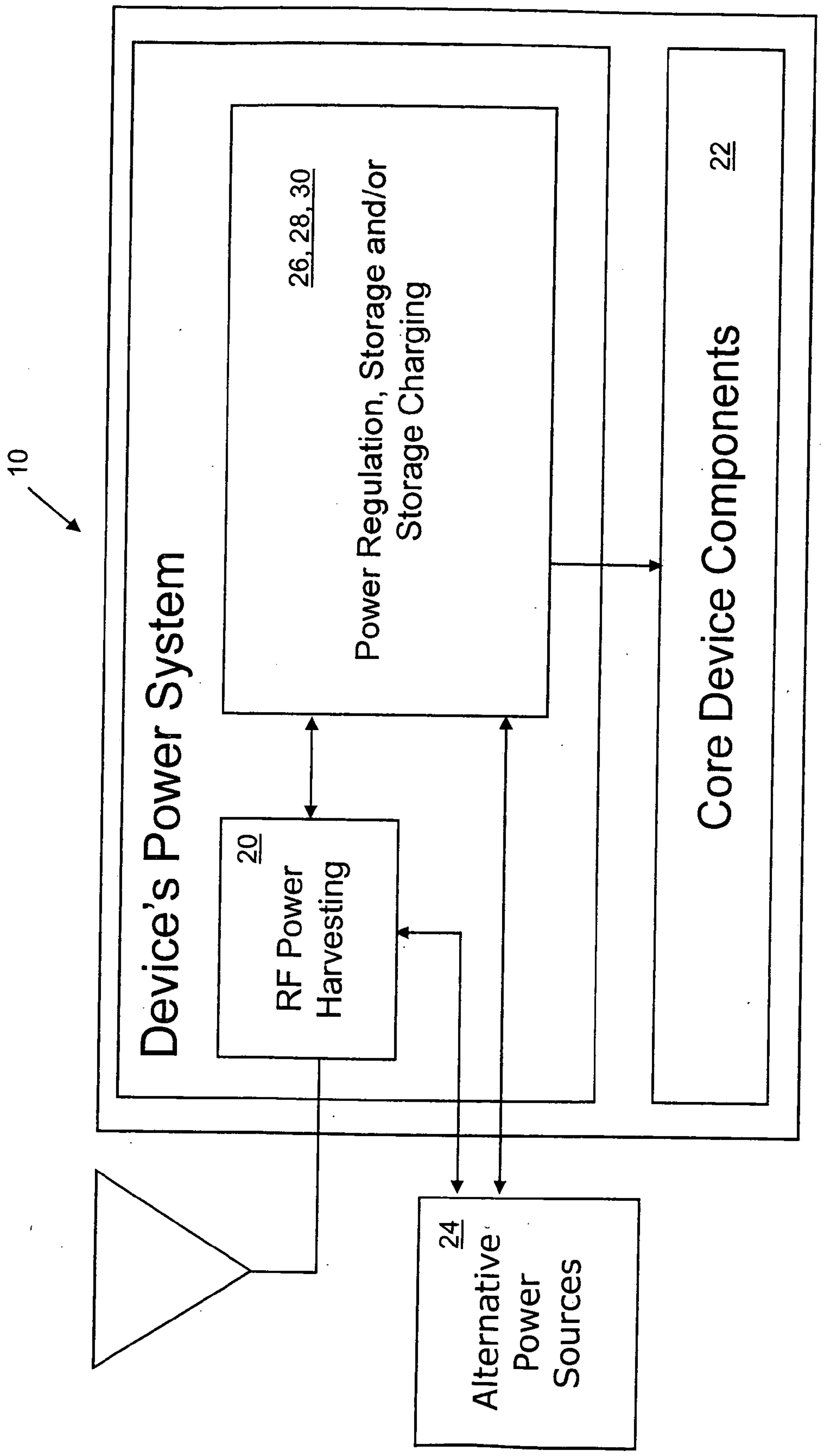
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 69

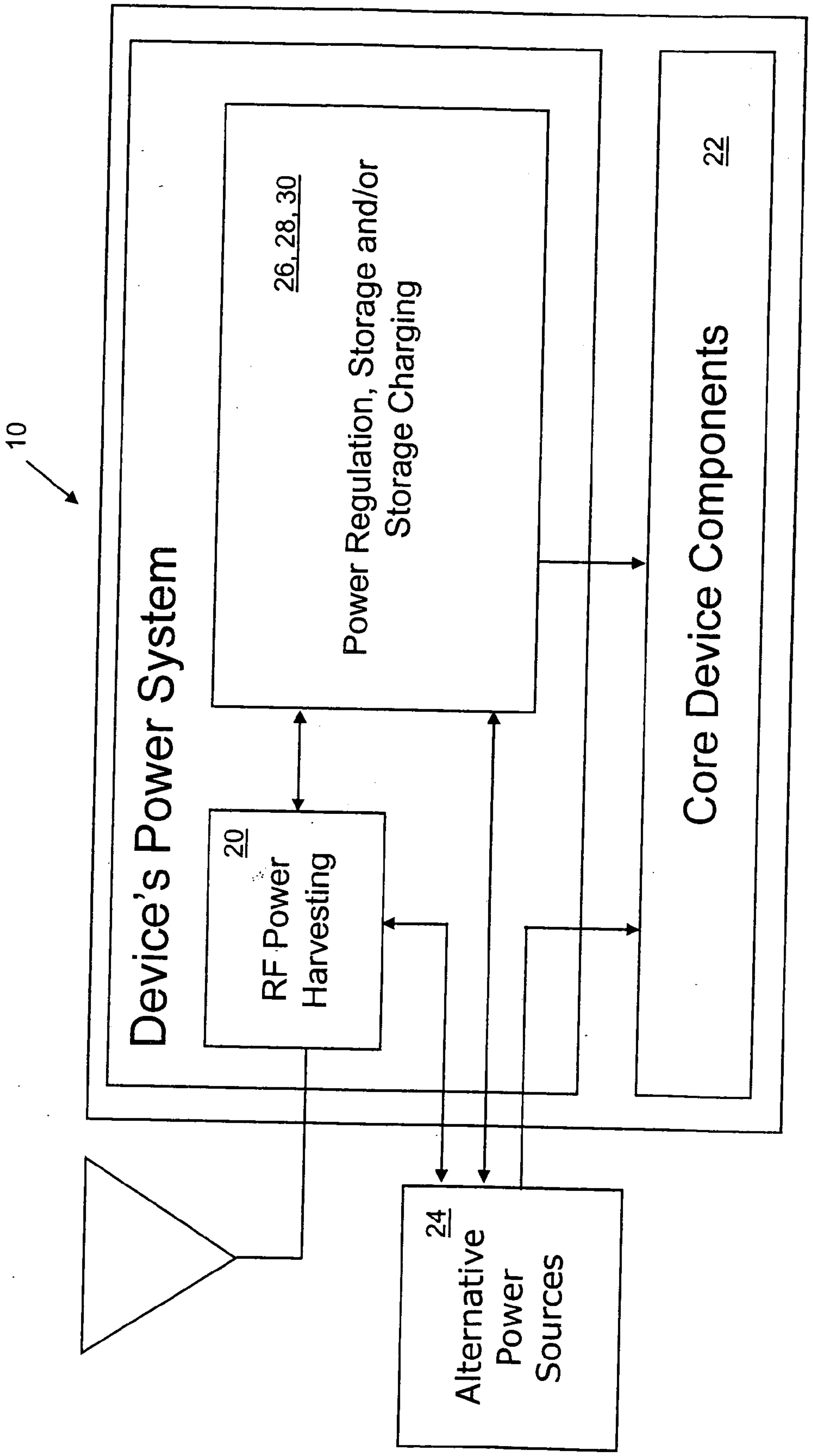
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 70

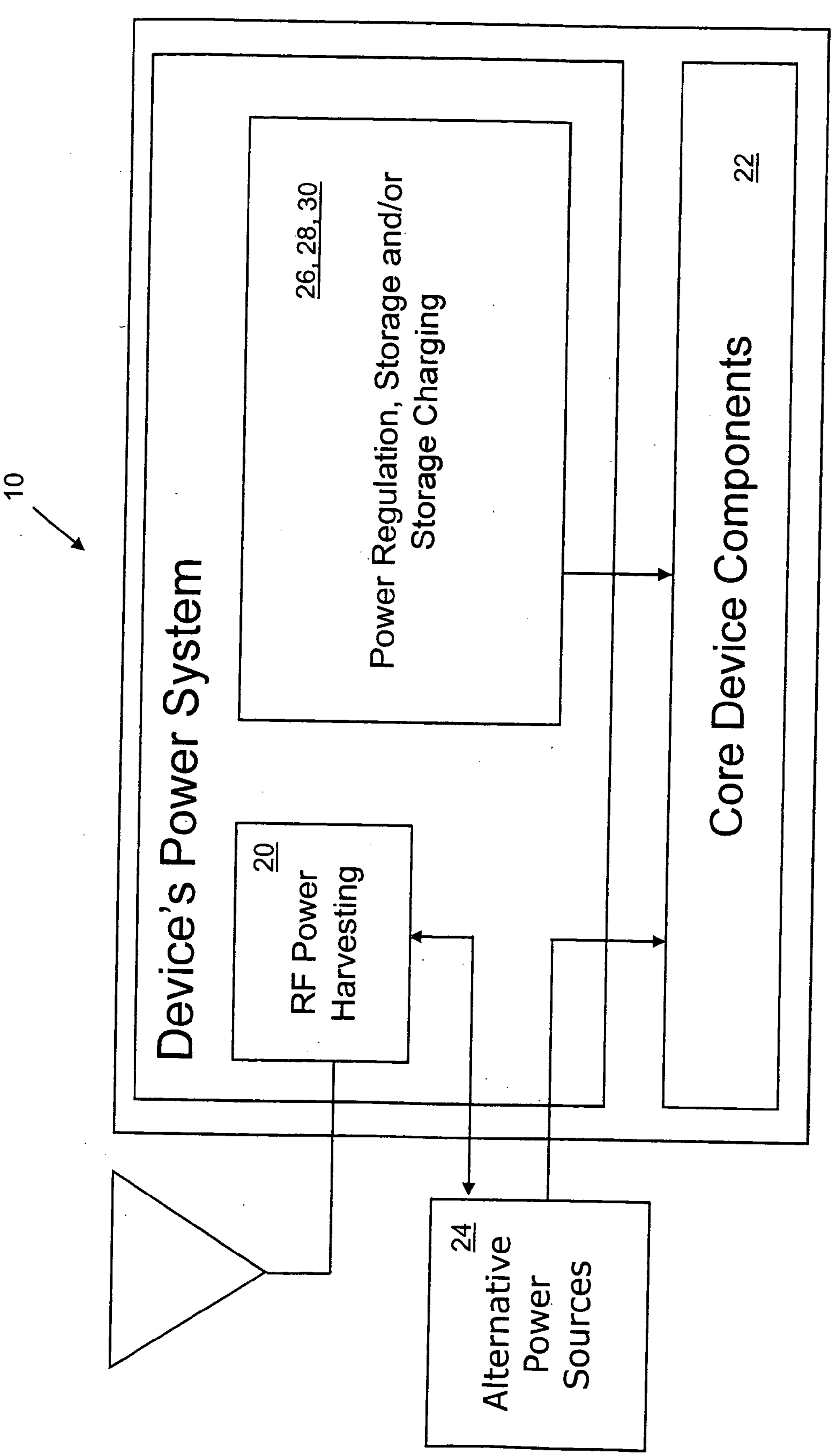
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 71

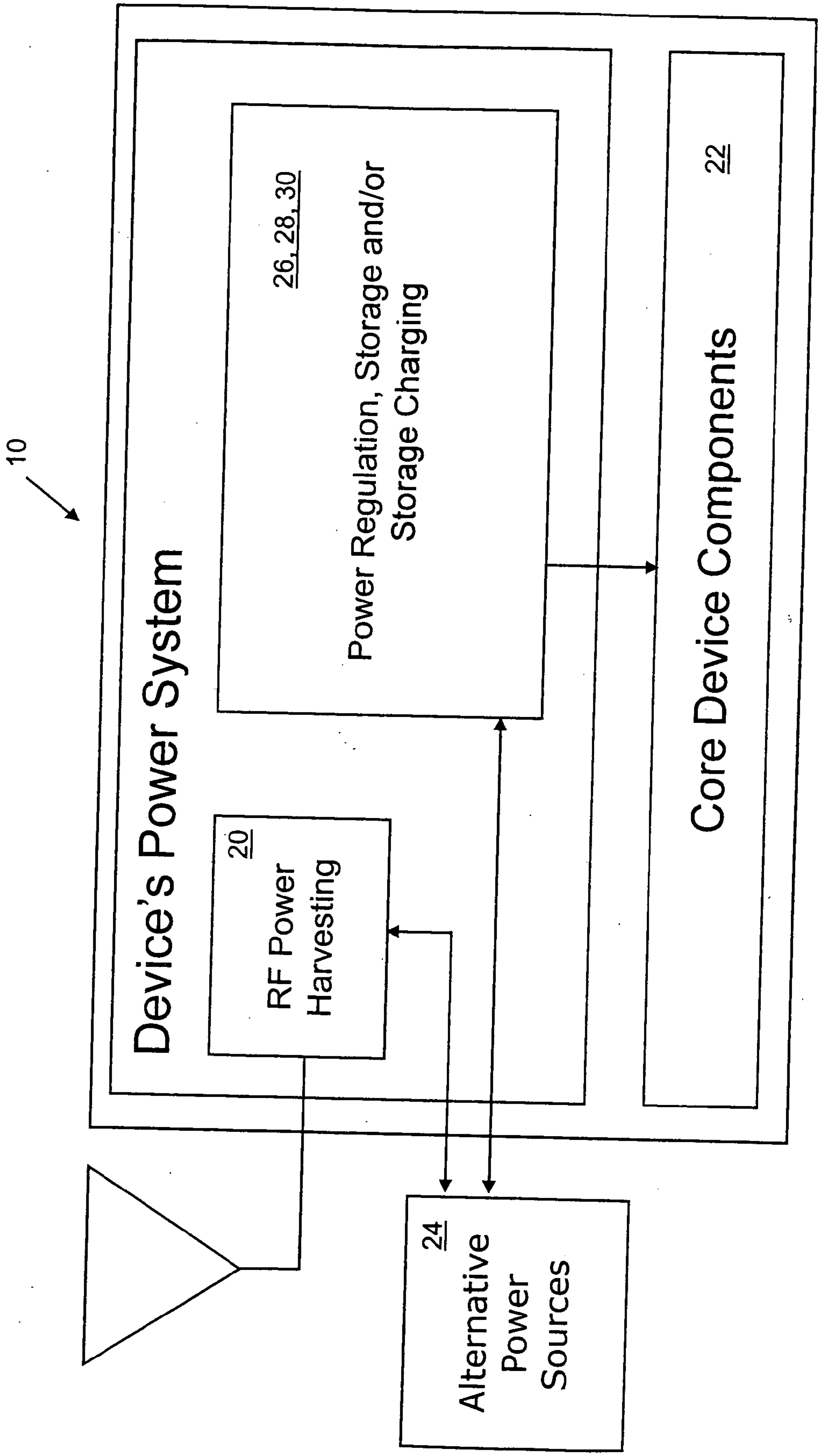
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 72

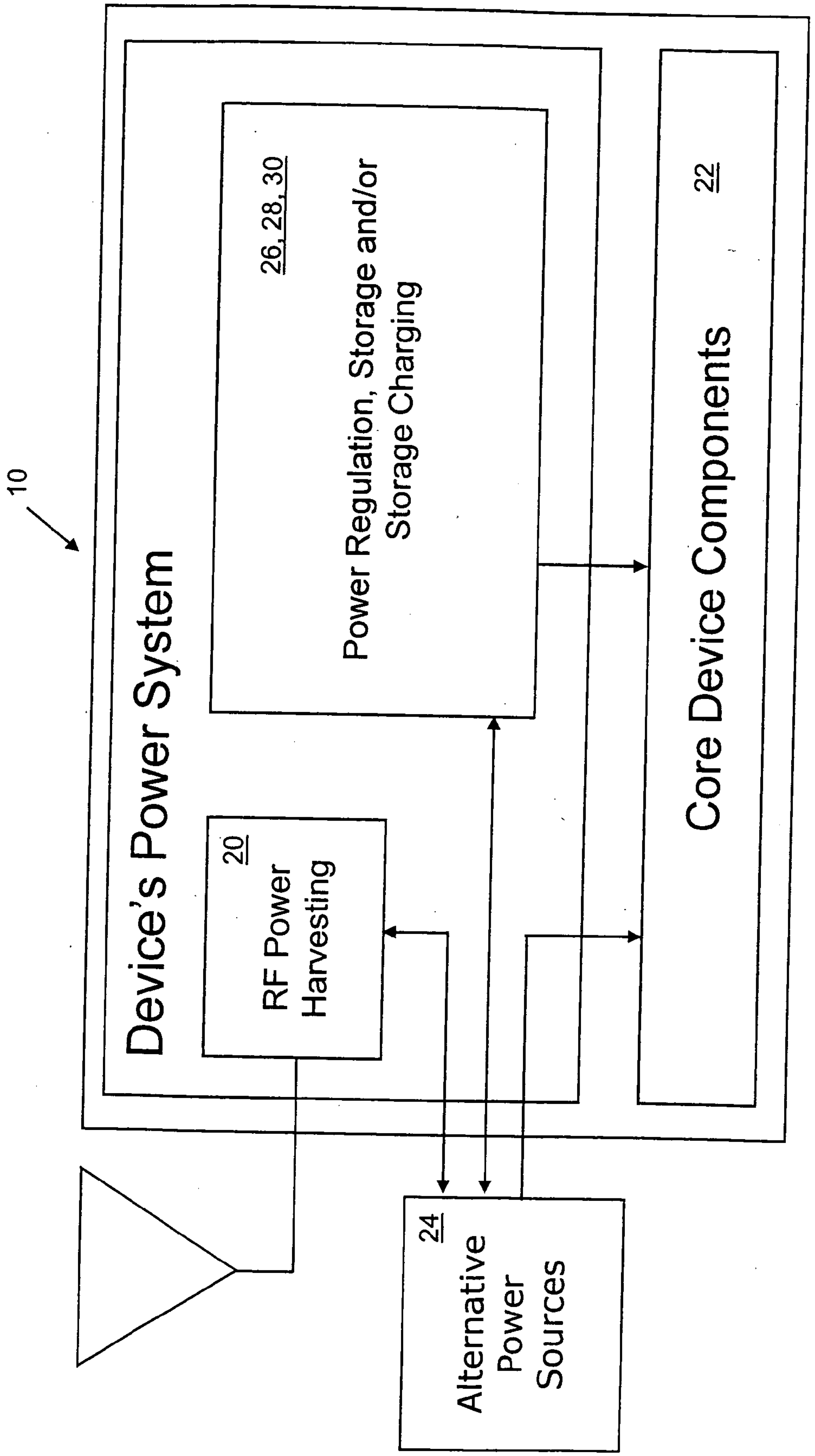
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 73

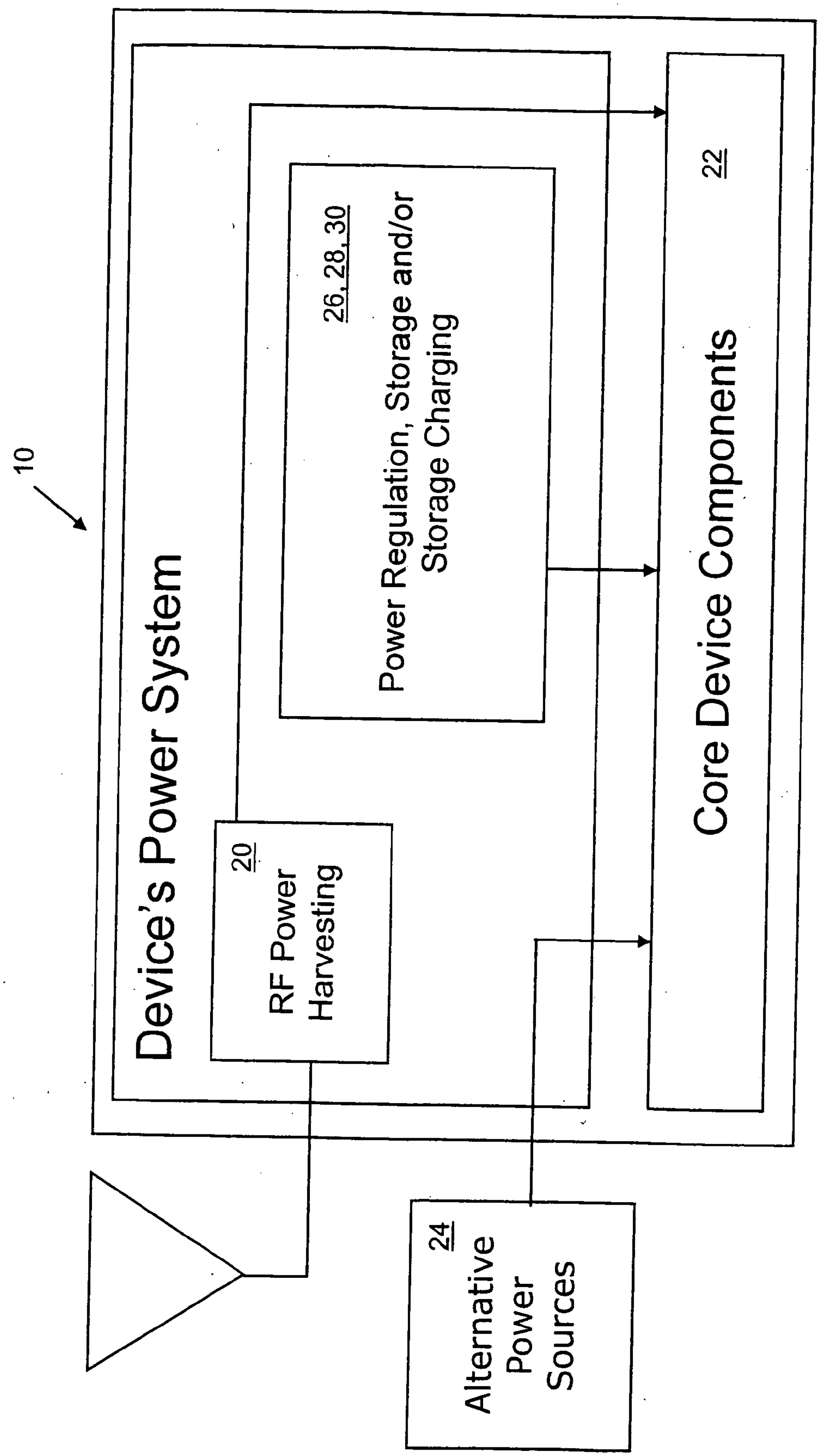
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 74

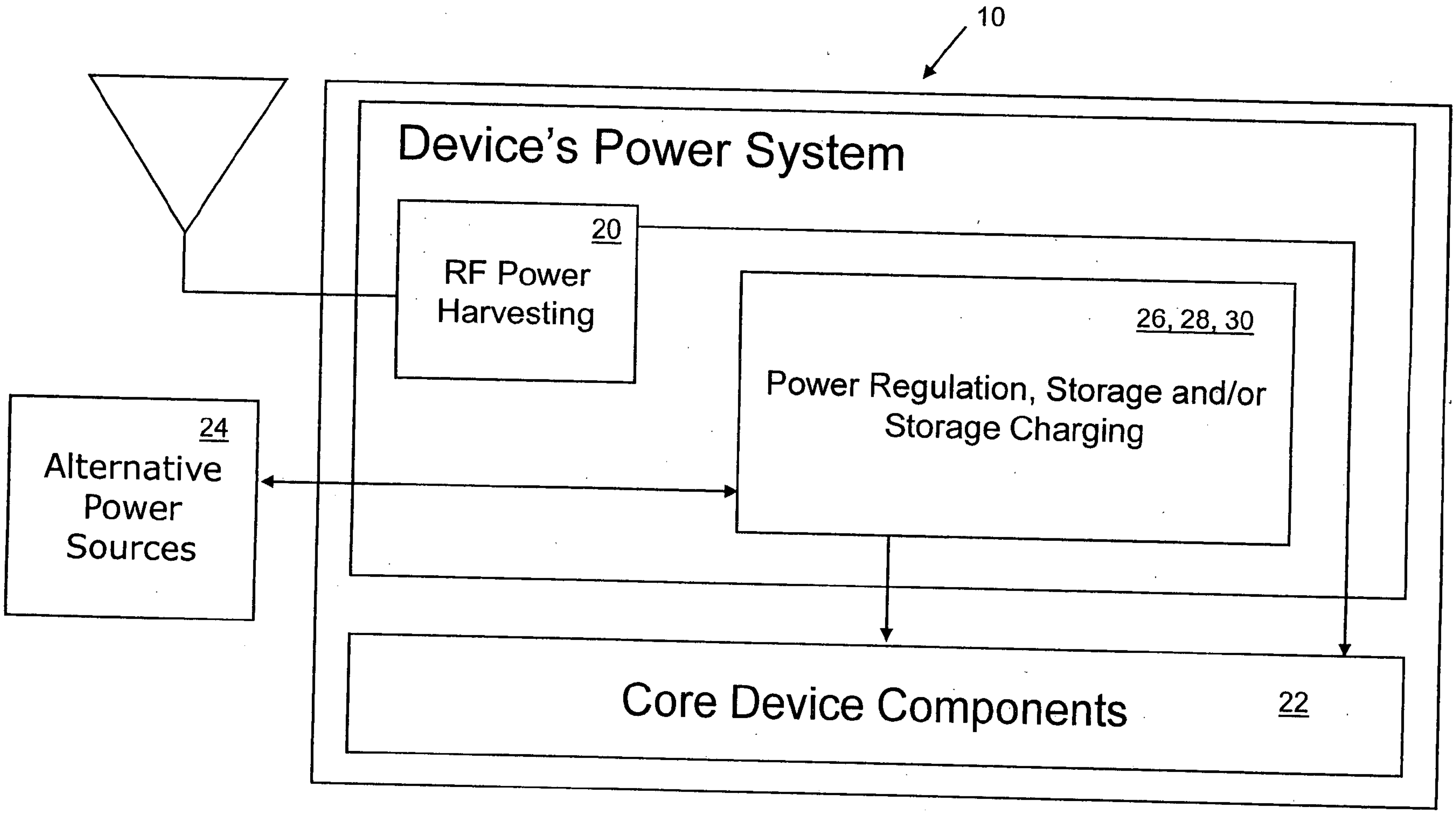
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 75

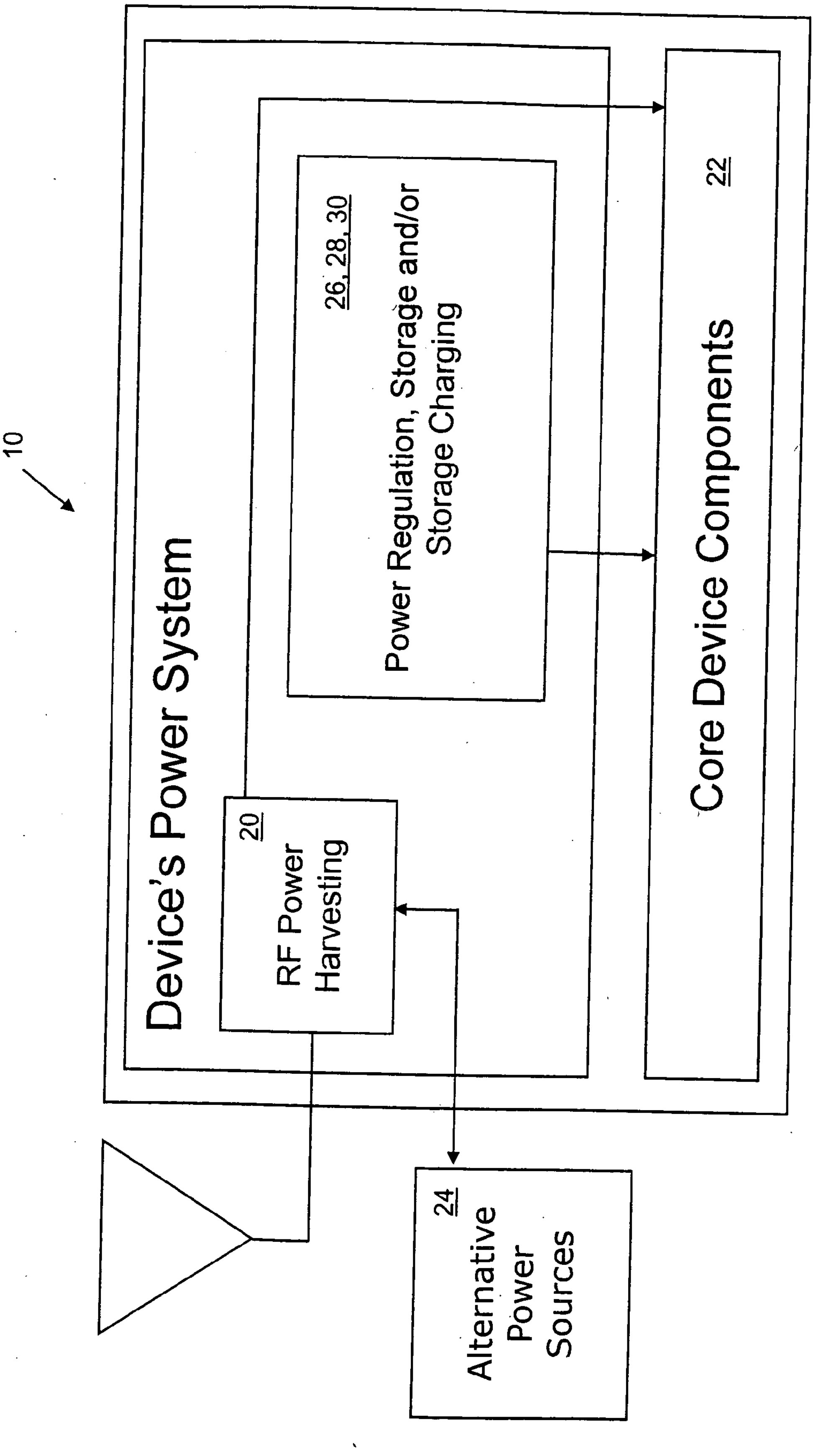
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 76

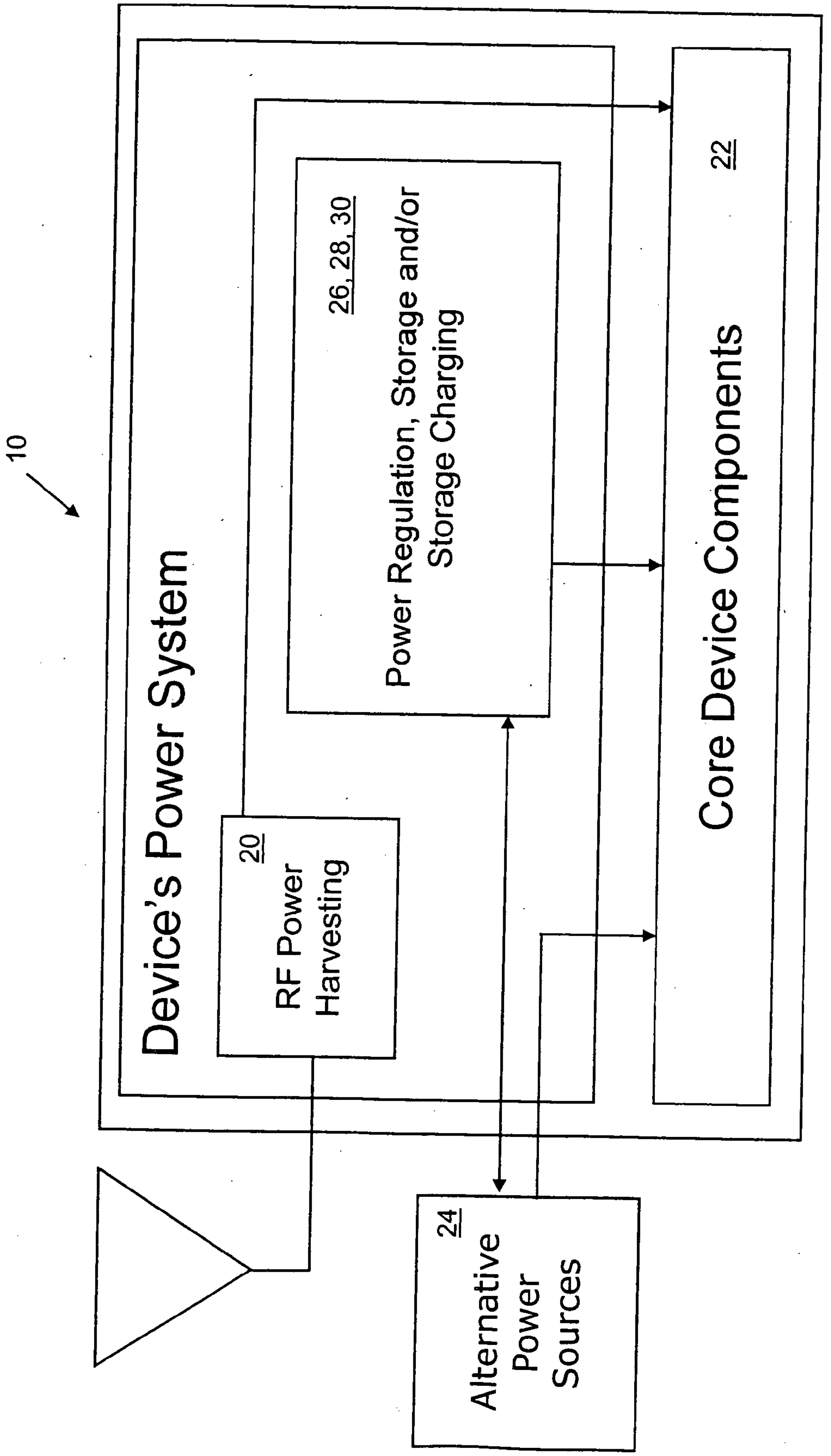
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
22
Core Device Components
24
Alternative Power Sources

Fig. 77

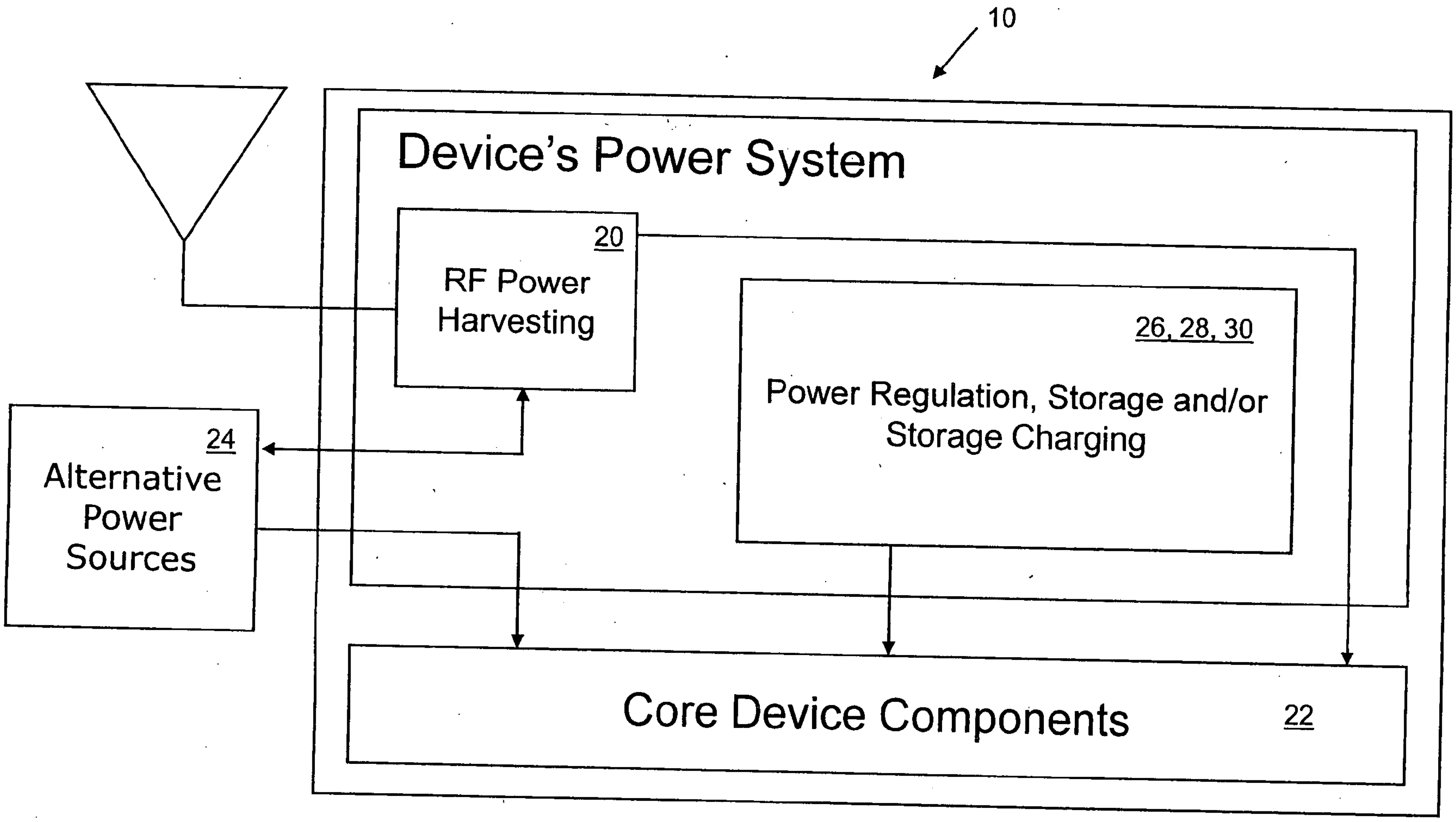
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 78

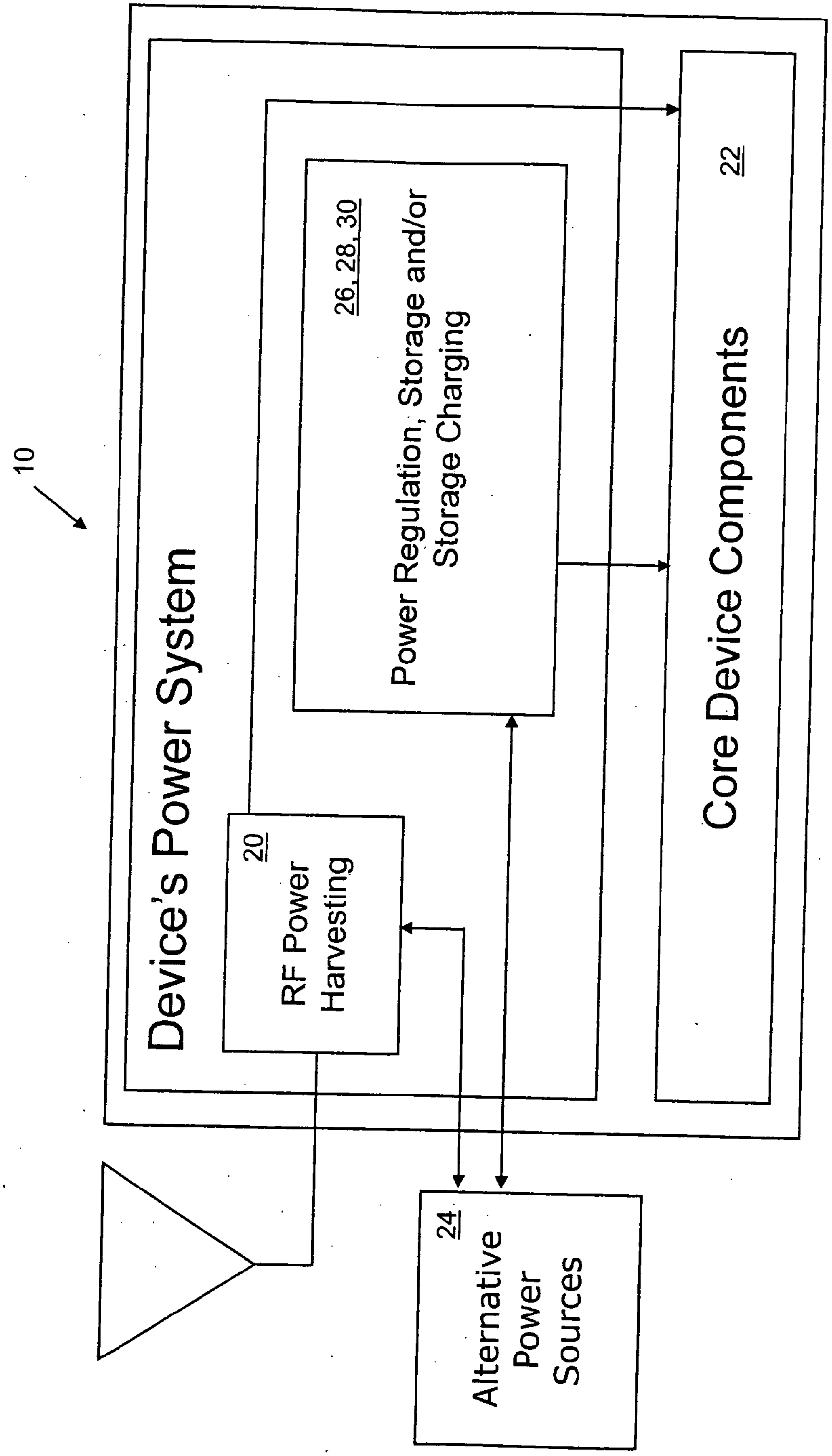
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 79

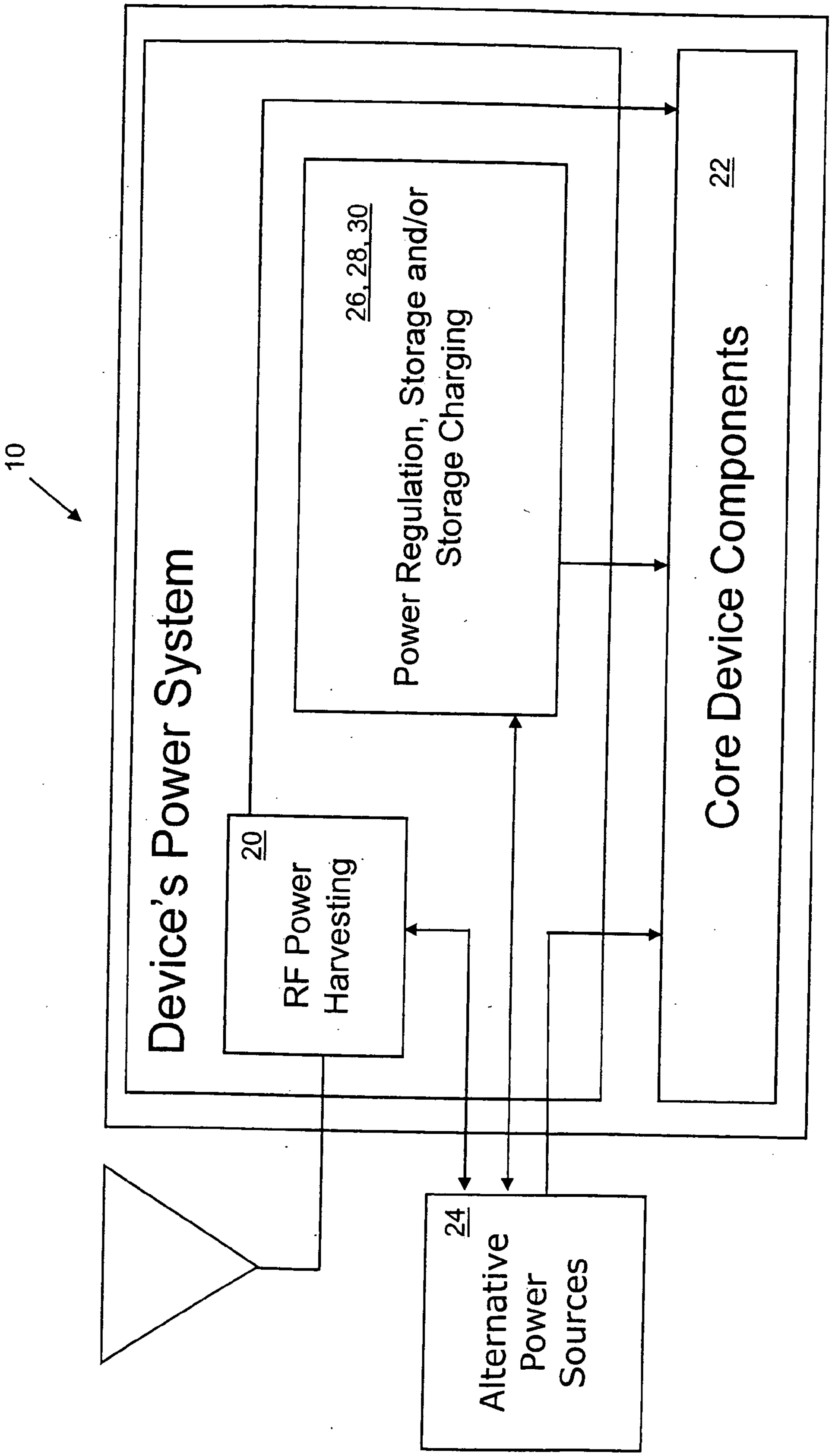
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 80

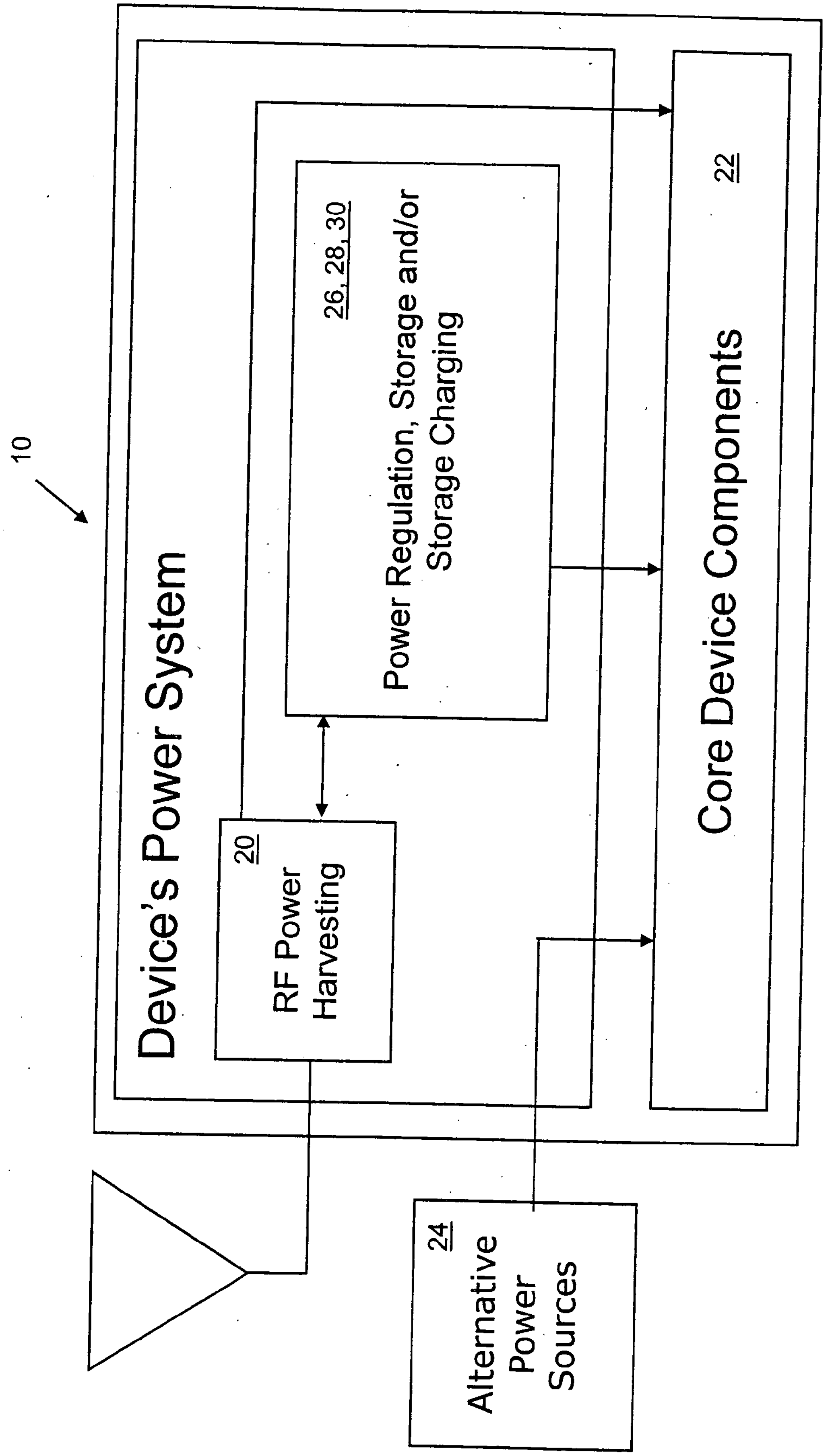
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 81

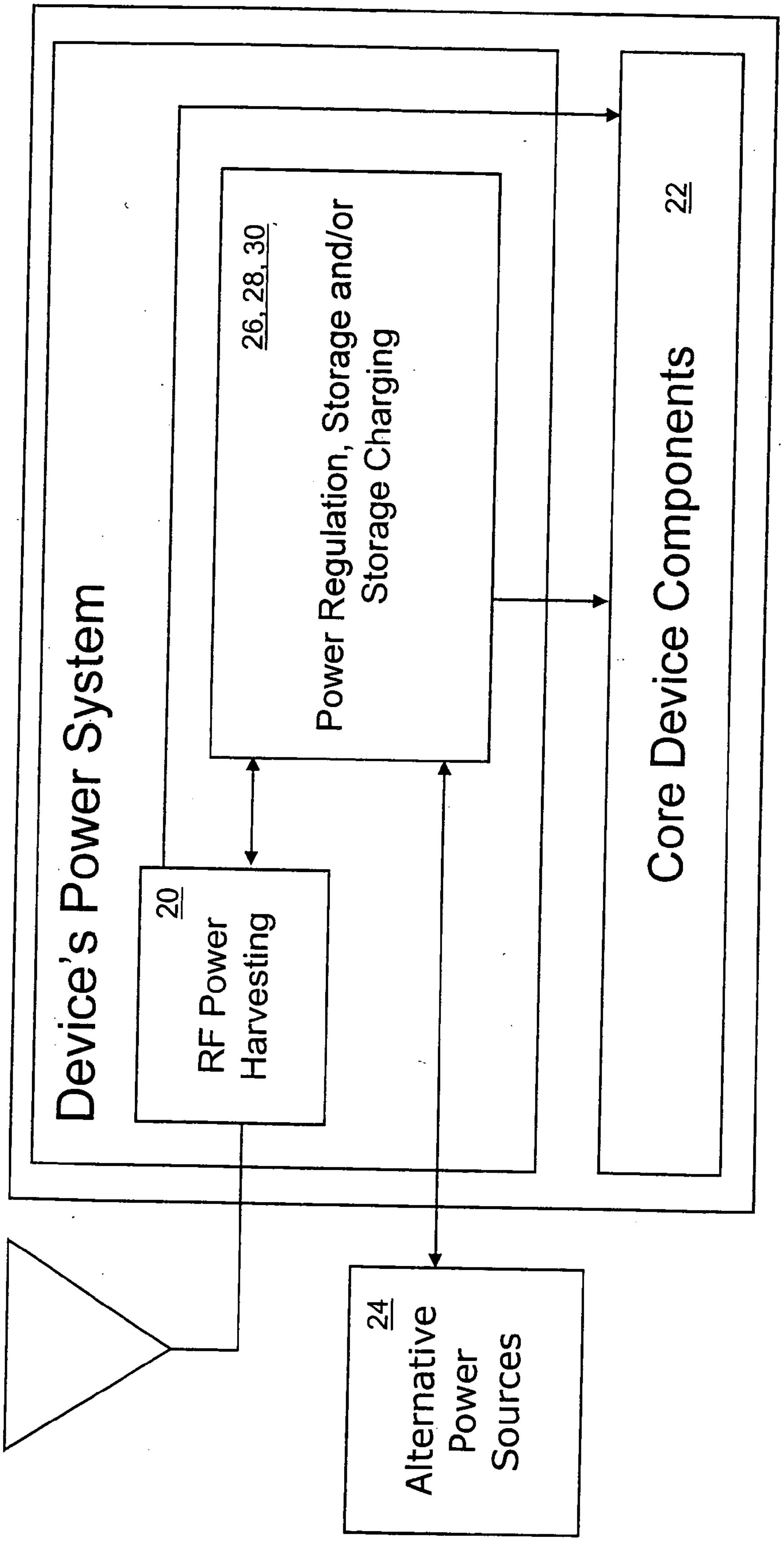
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 82

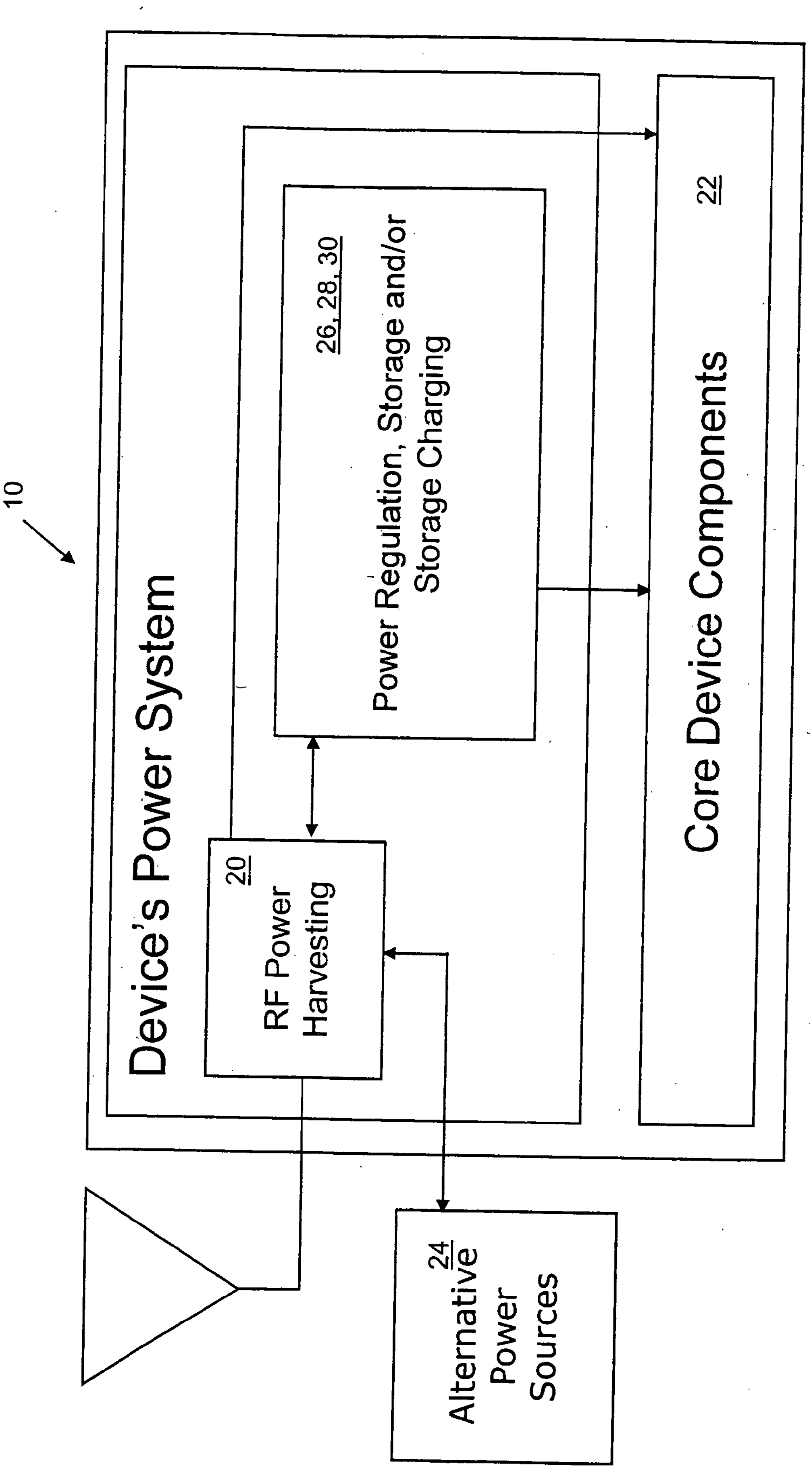
10
Device's Power System
26, 28, 30
Power Regulation, Storage and/or Storage Charging
20
RF Power Harvesting
Core Device Components
22
24
Alternative Power Sources

Fig. 83

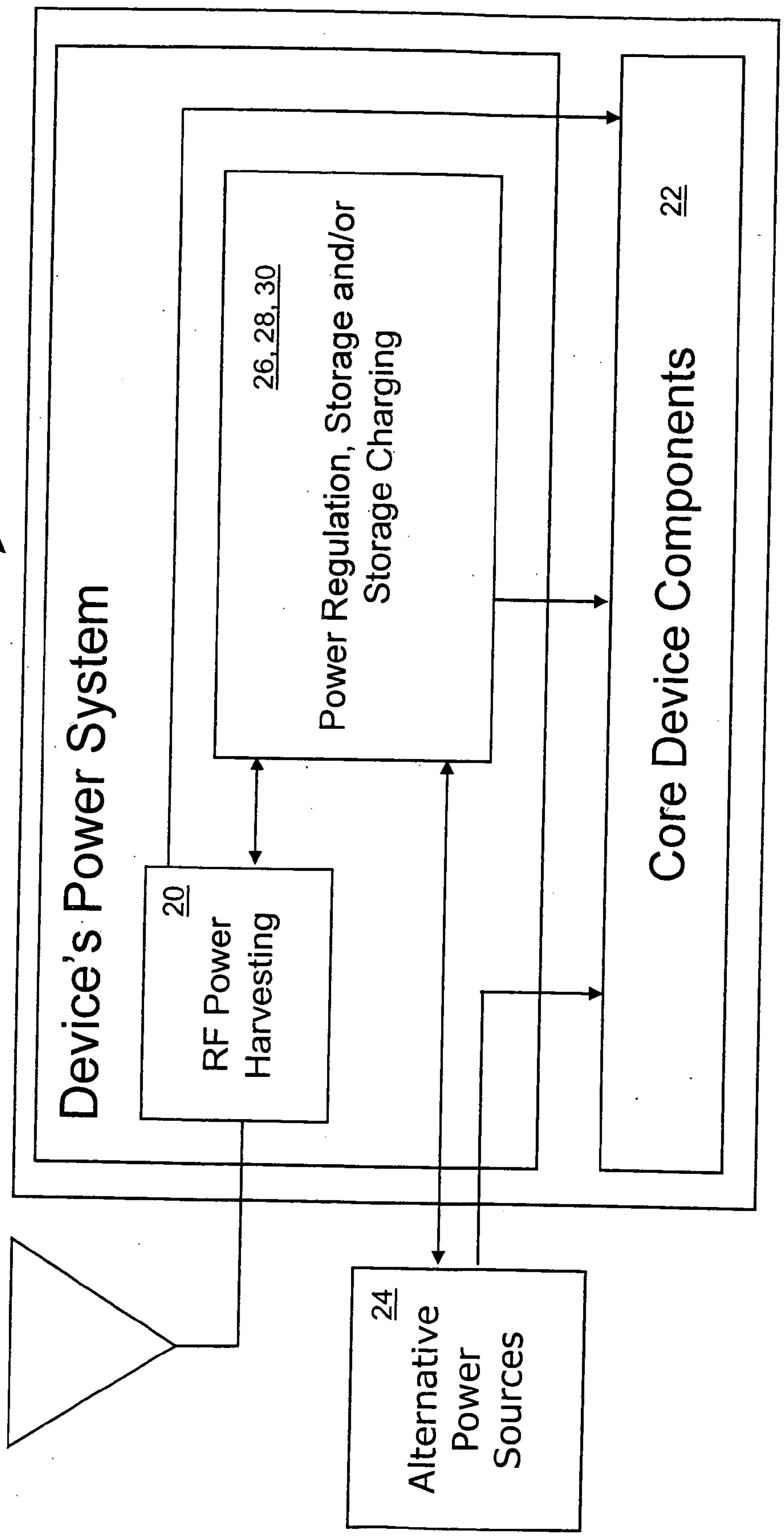
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 84

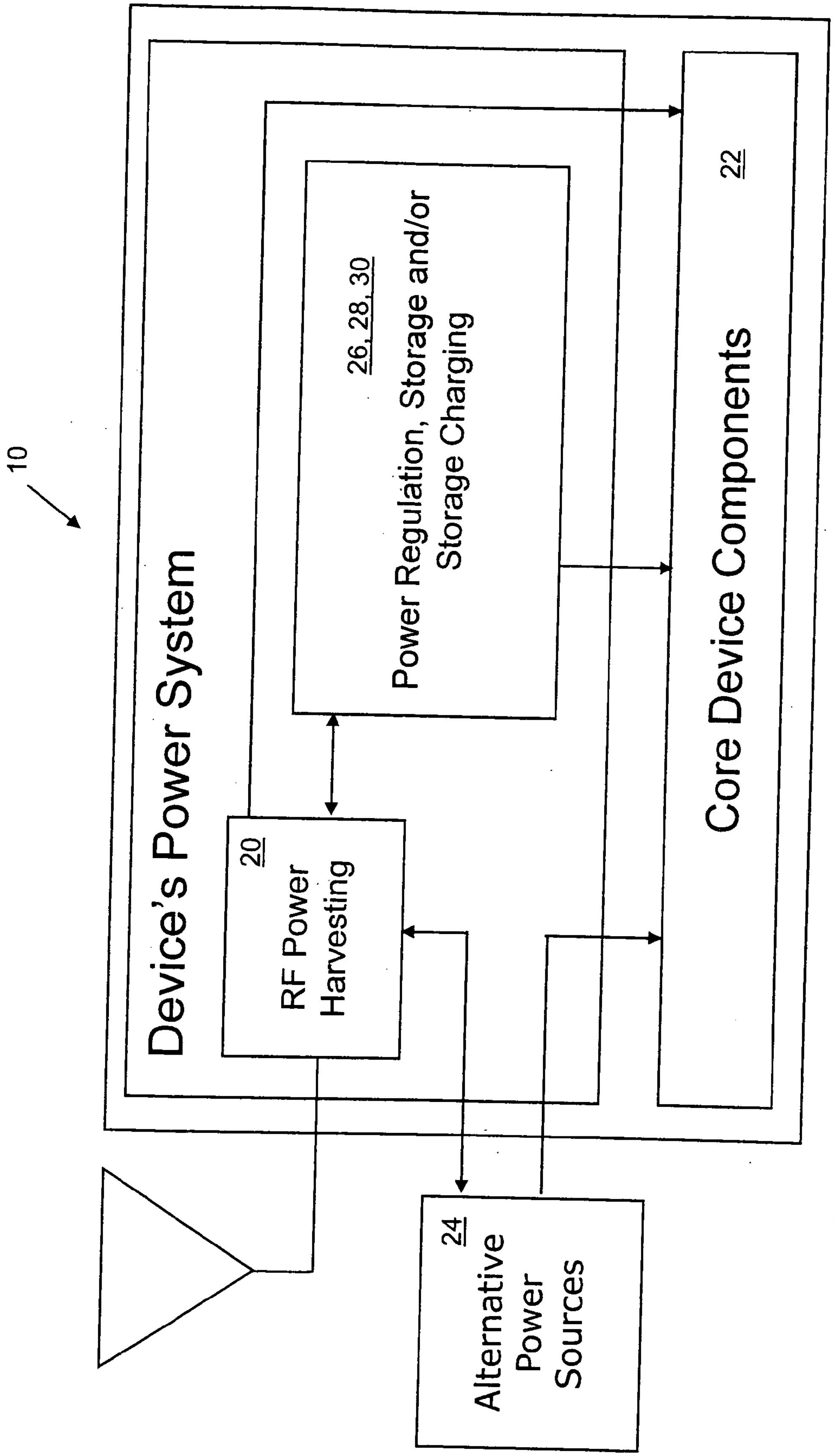
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 85

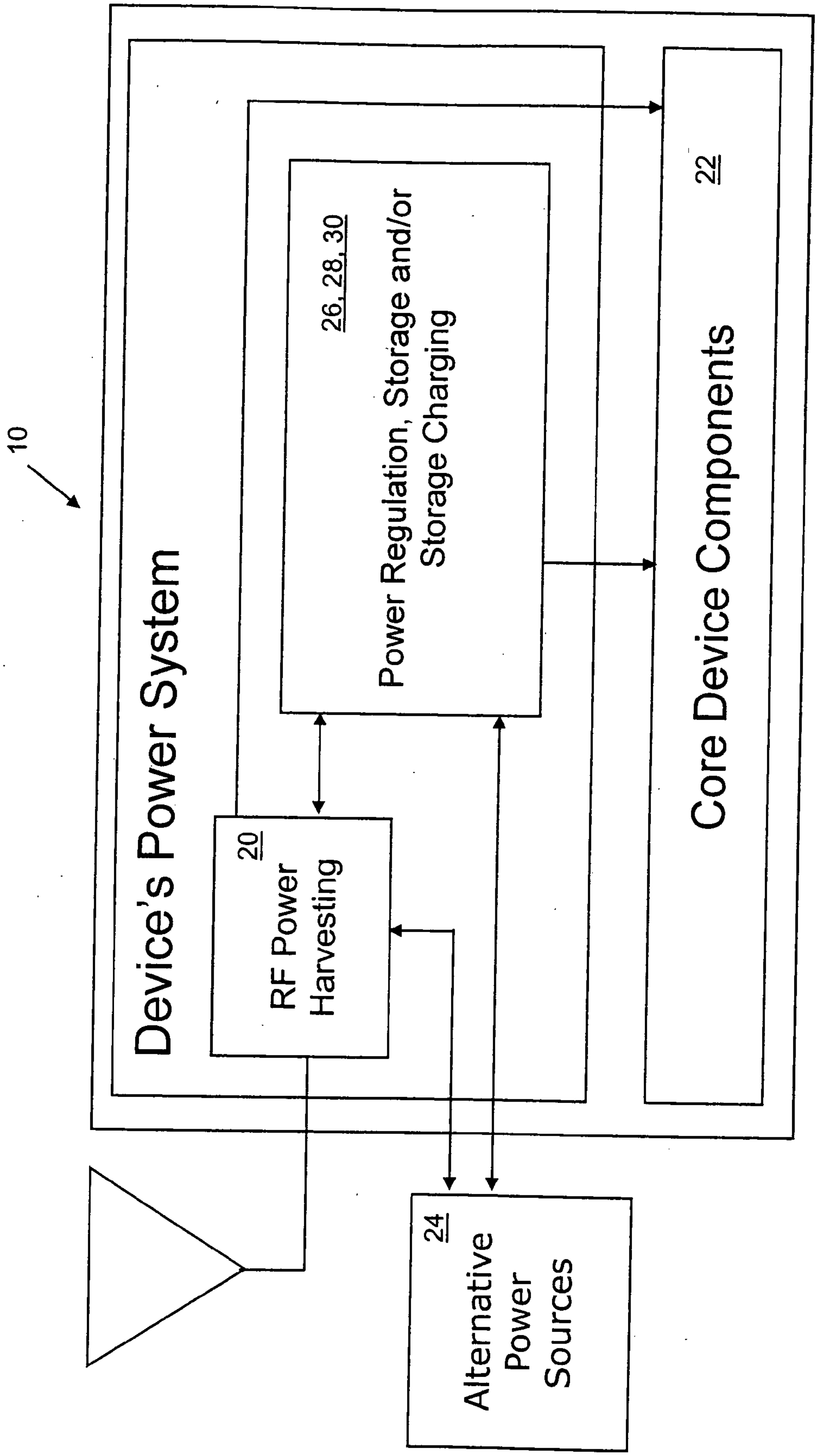
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
Core Device Components
22
24
Alternative Power Sources

Fig. 86

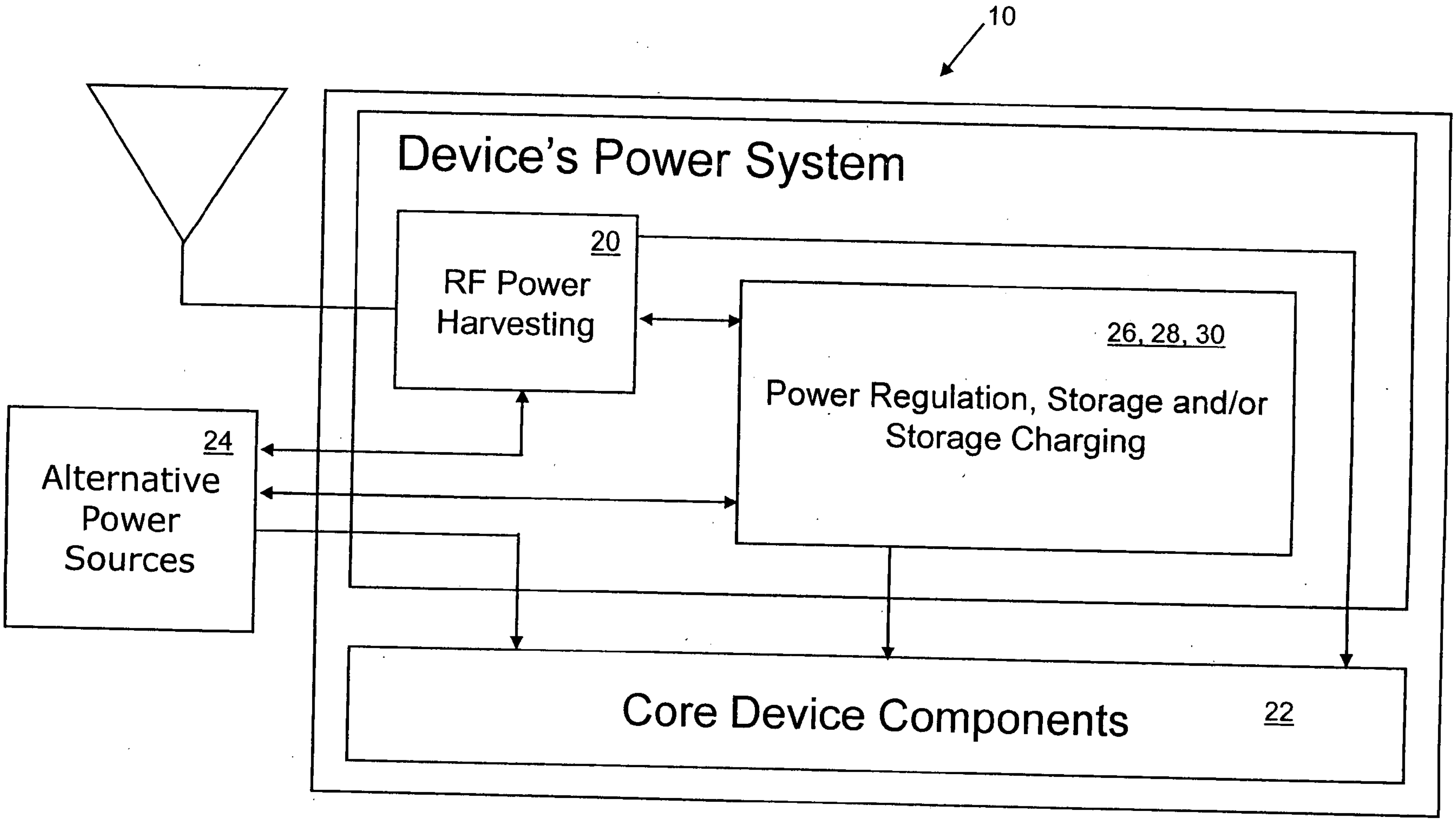
10
Device's Power System
20
RF Power Harvesting
26, 28, 30
Power Regulation, Storage and/or Storage Charging
24
Alternative Power Sources
Core Device Components
22

Fig. 87

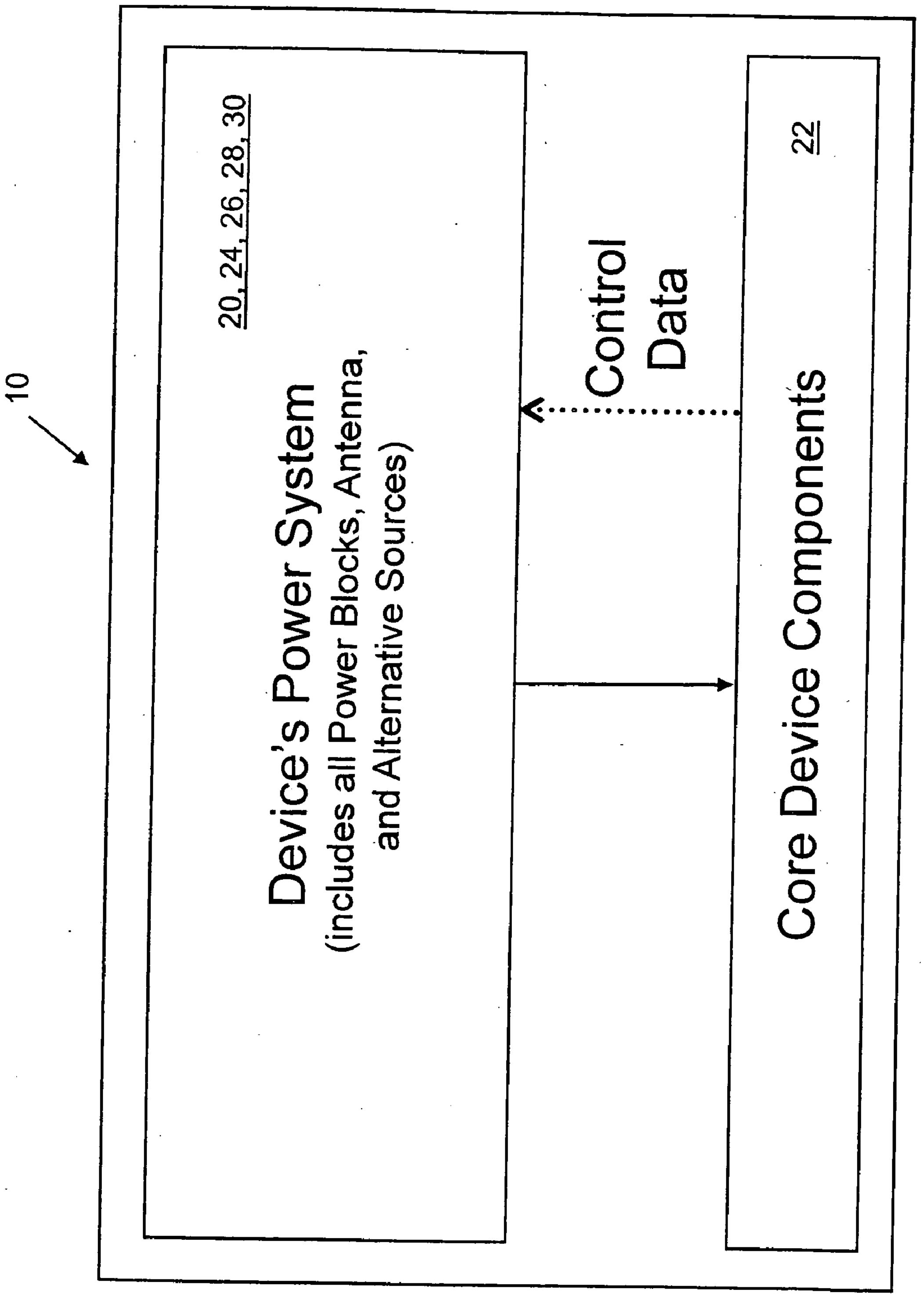
10
Device's Power System
(includes all Power Blocks, Antenna, and Alternative Sources)
20, 24, 26, 28, 30
Control Data
Core Device Components
22

Fig. 88

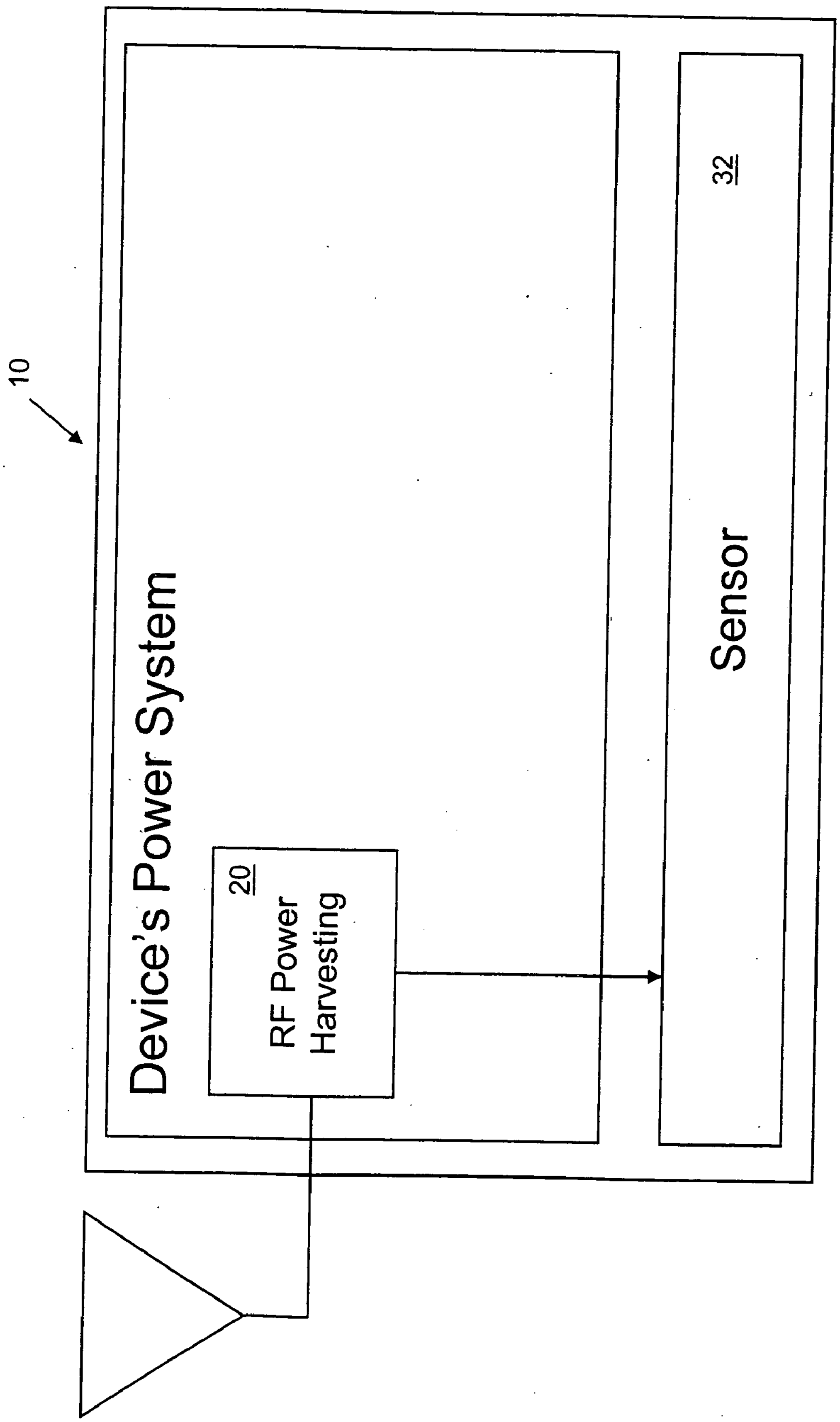
10
Device's Power System
20
RF Power Harvesting
Sensor
32

Fig. 89

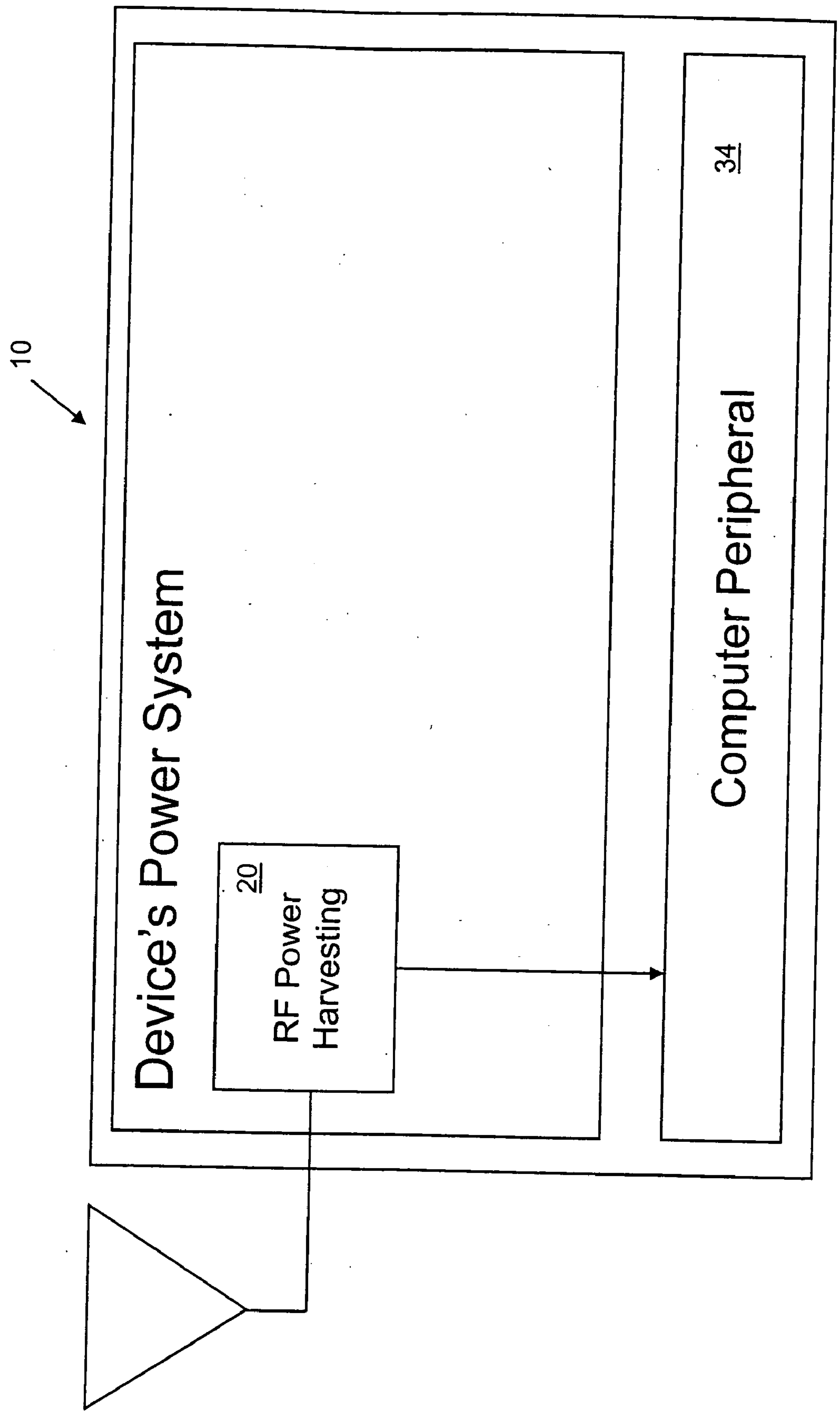
10
Device's Power System
20
RF Power Harvesting
Computer Peripheral
34

Fig. 90

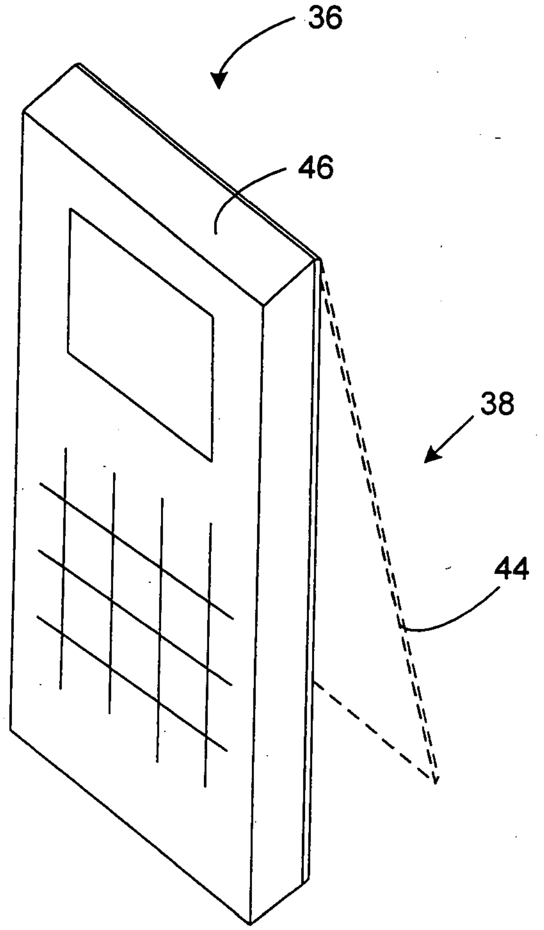

Fig. 91(a)

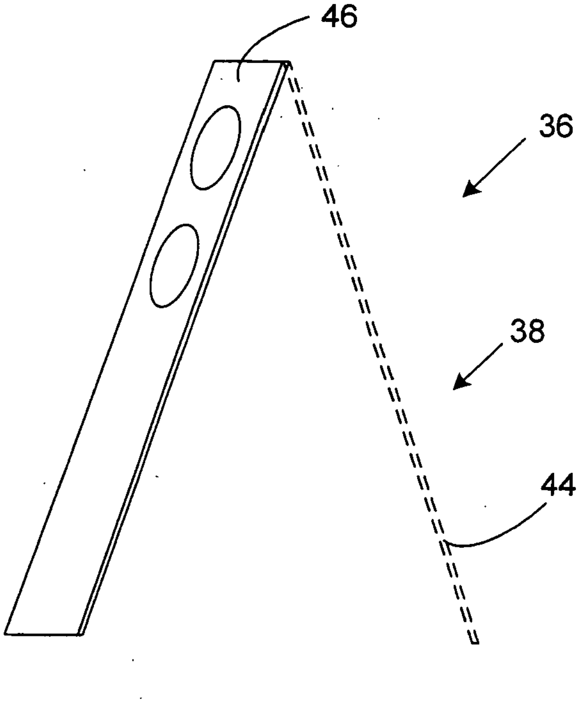

Fig. 91(b)

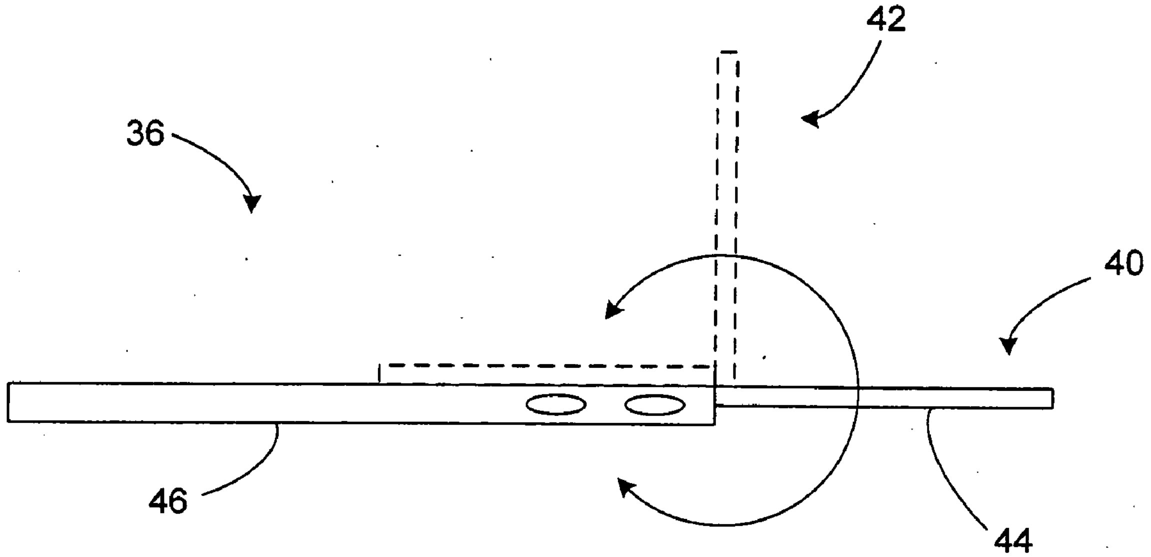
42
36
40
46
44

Fig. 92

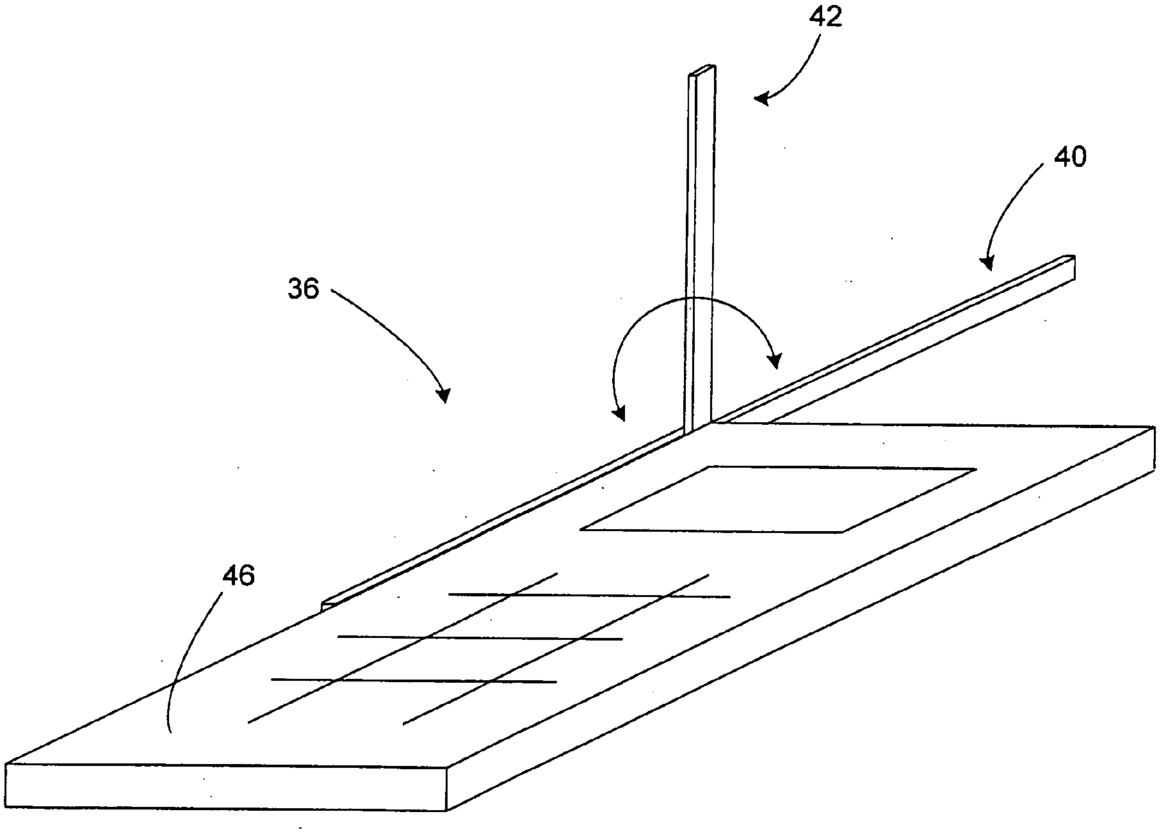

Fig. 93

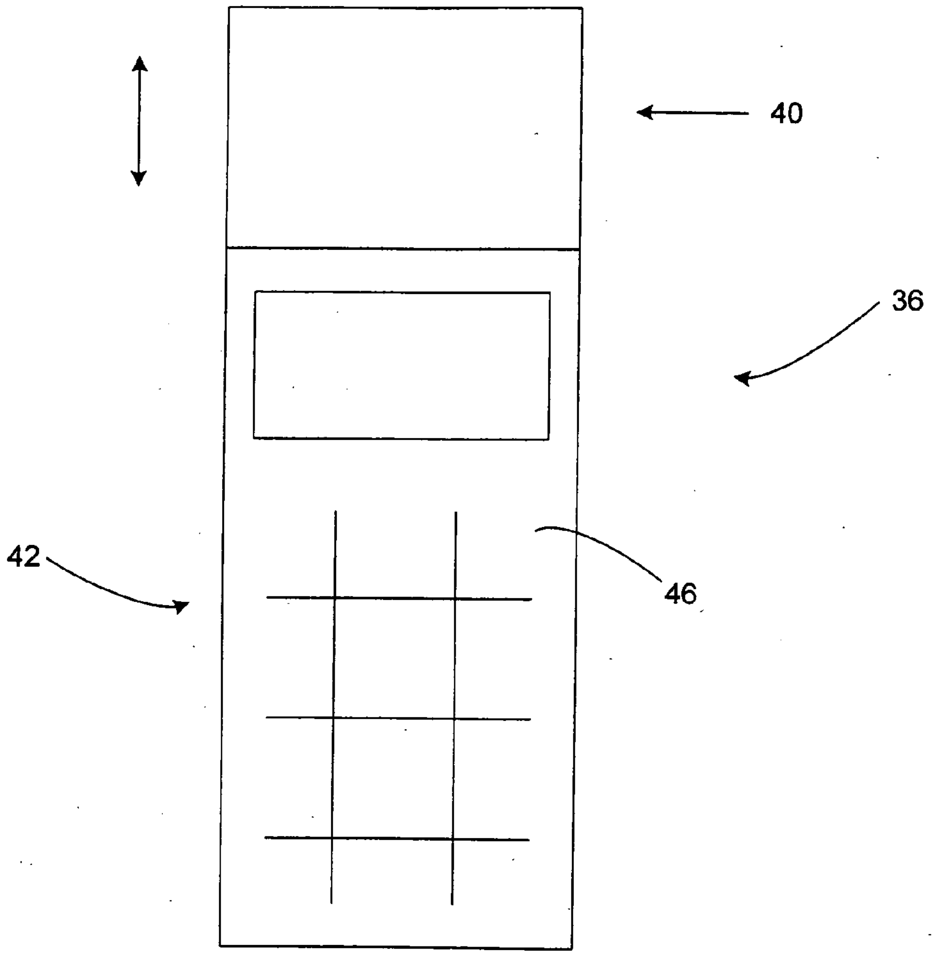

Fig. 94(a)

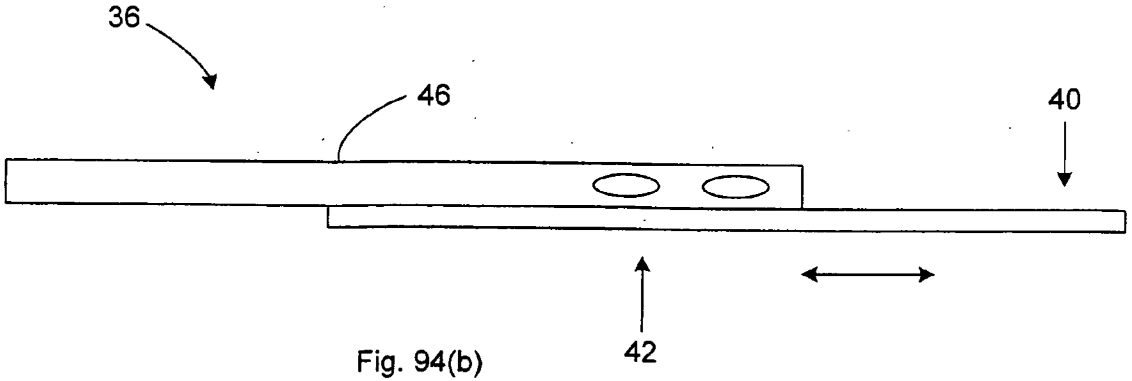

Fig. 94(b)

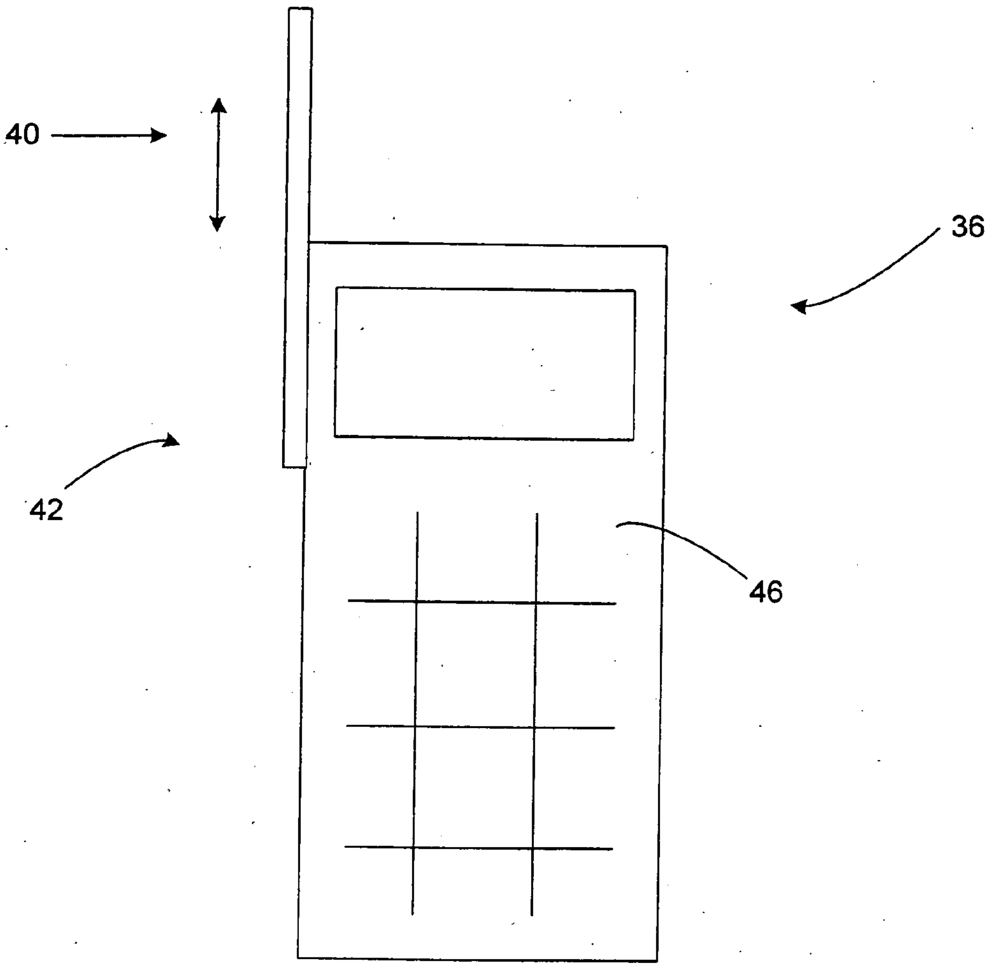

Fig. 95(a)

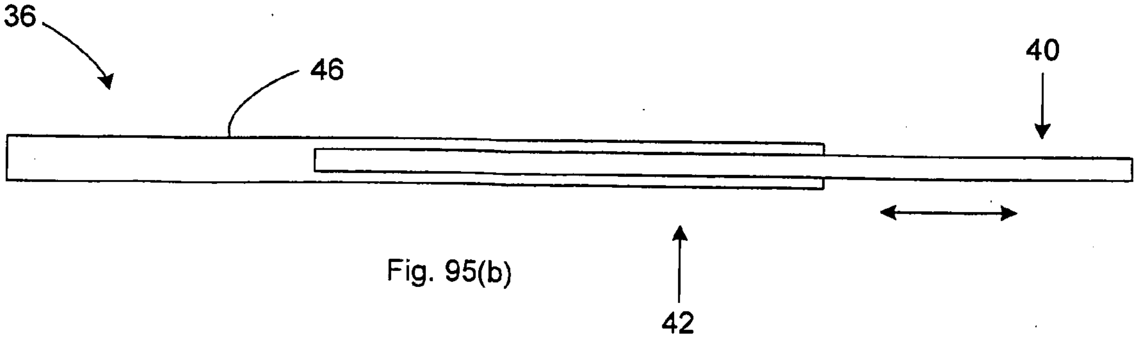

Fig. 95(b)

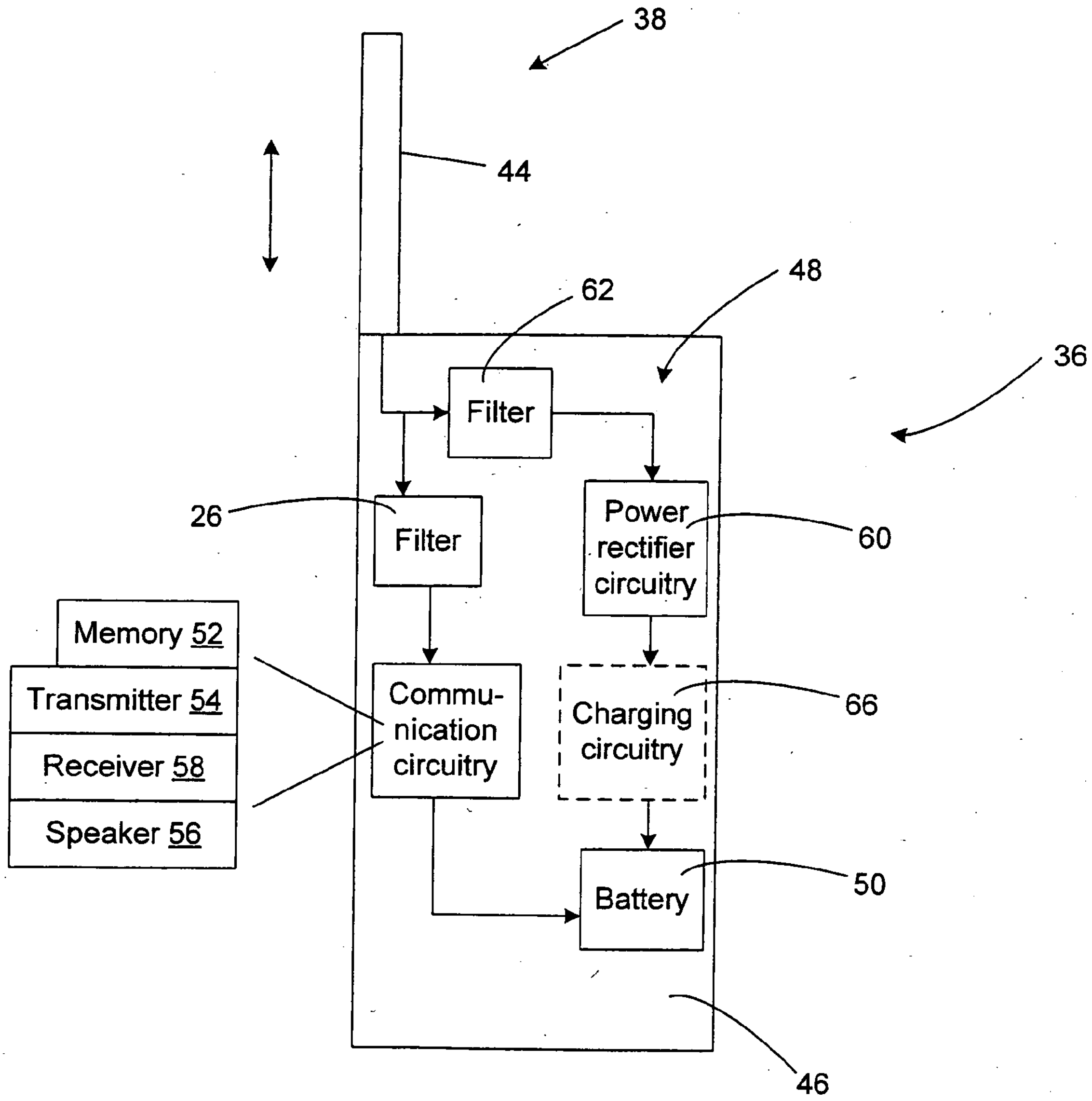
38
44
62
48
36
Filter
Filter
26
Power rectifier circuitry
60
Memory 52
Transmitter 54
Receiver 58
Speaker 56
Commu-nication circuitry
Charging circuitry
66
Battery
50
46

Fig. 96

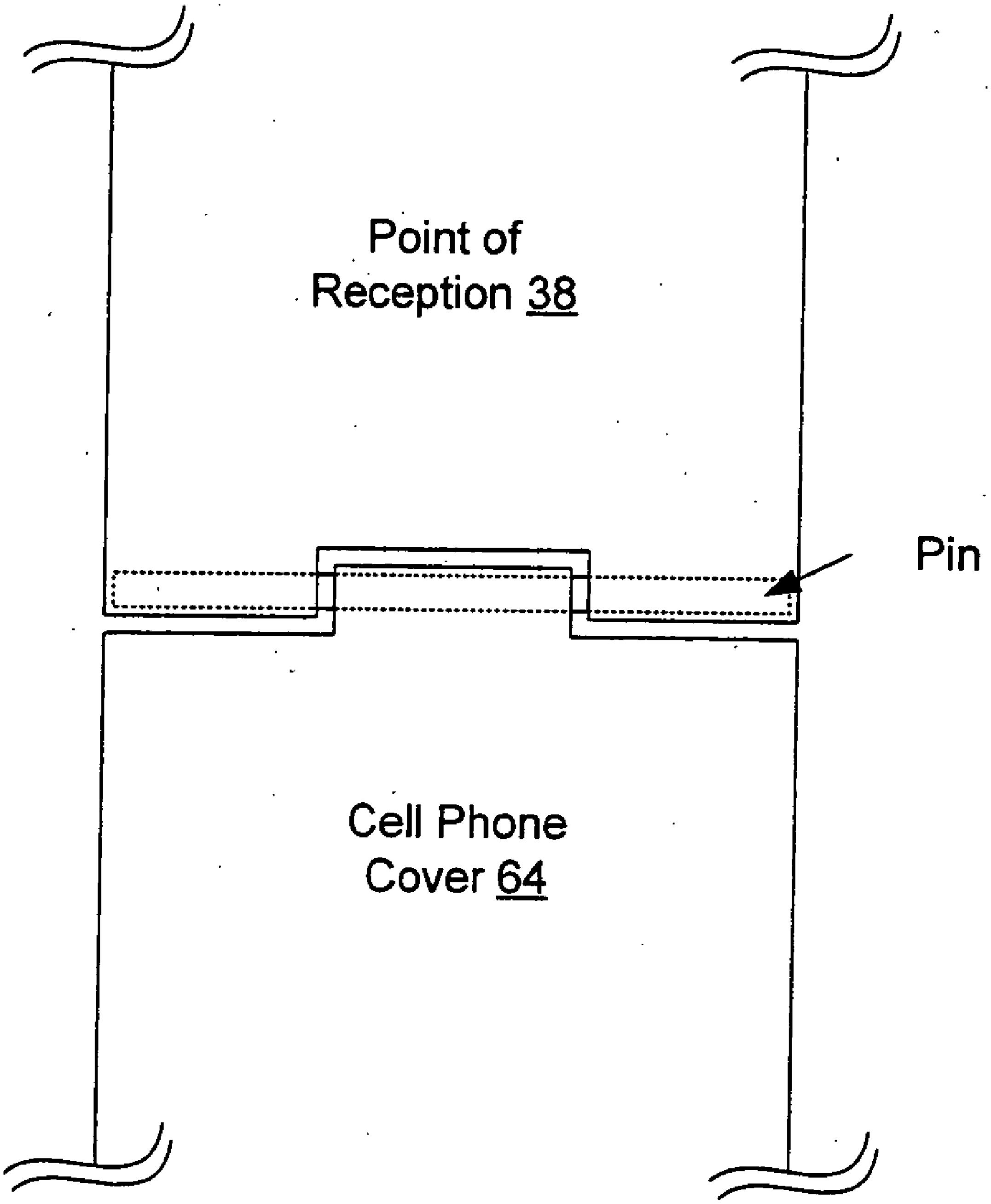
Point of
Reception 38
Pin
Cell Phone
Cover 64

Fig. 97

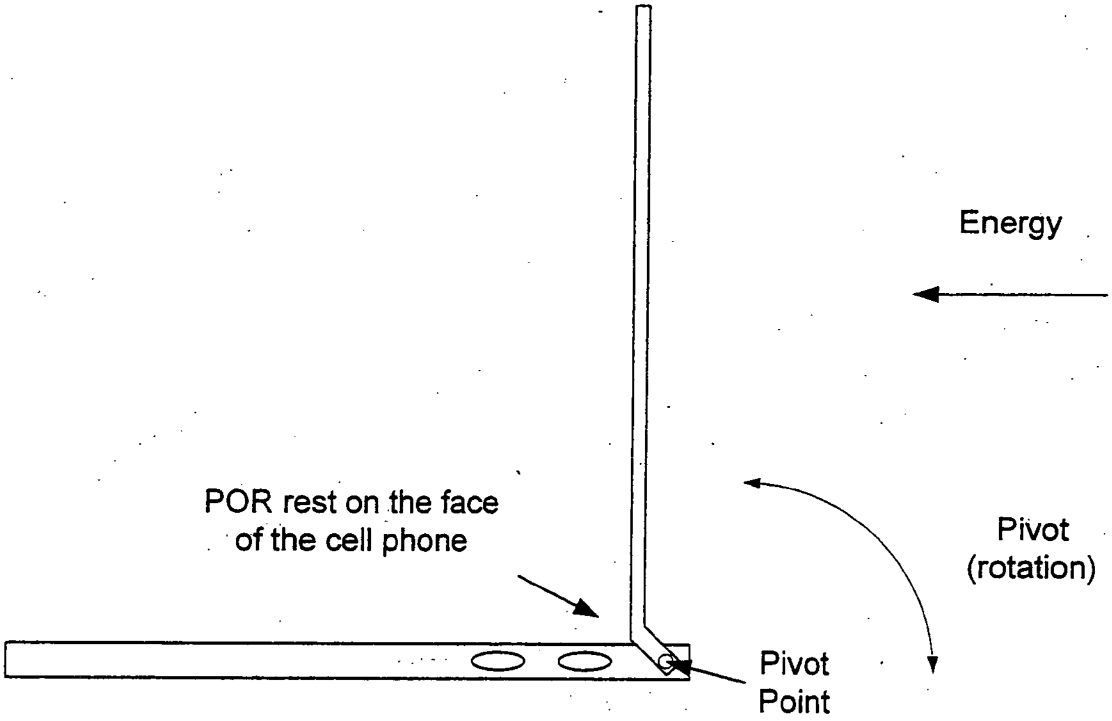
Energy
POR rest on the face of the cell phone
Pivot (rotation)
Pivot Point

Figure 98 (a) - Cell phone with point of reception (shown as a patch antenna) located in charging position

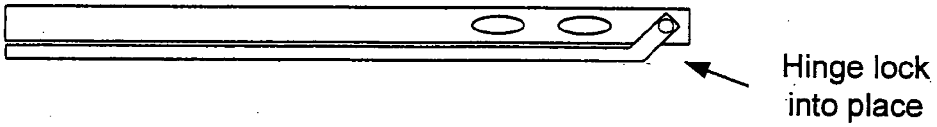
Hinge lock into place

Figure 98 (b) - Cell phone with point of reception (shown as a patch antenna) located in normal use position

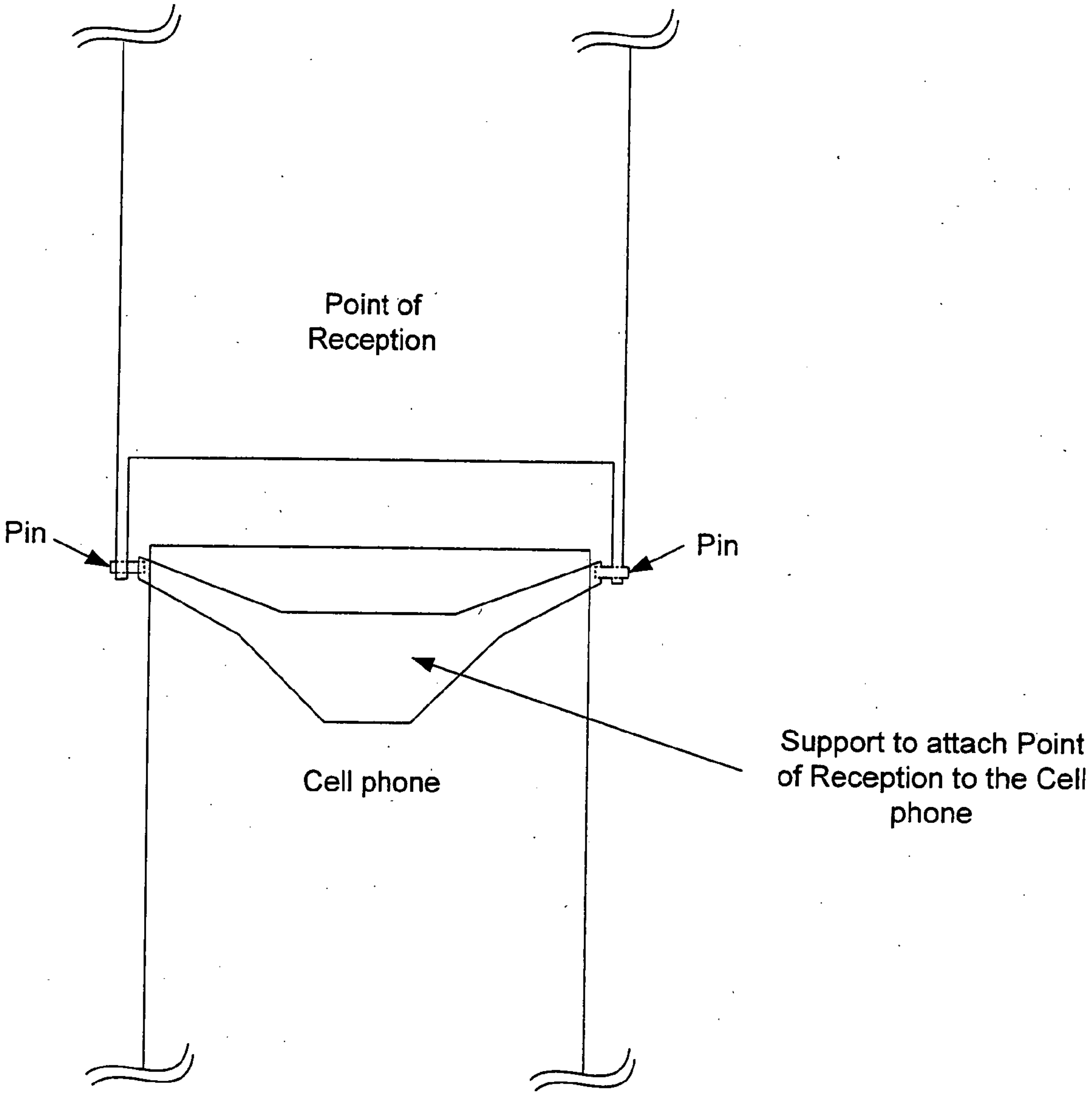
Point of
Reception
Pin
Pin
Cell phone
Support to attach Point of Reception to the Cell phone

Fig. 99

POWERING CELL PHONES AND SIMILAR DEVICES USING RF ENERGY HARVESTING

FIELD OF THE INVENTION

[0001] The present invention is related to the wireless powering of devices. More specifically, the present invention is related to the wireless powering of devices, namely cell phones and the like, with a power harvester.

BACKGROUND OF THE INVENTION

[0002] As processor capabilities have expanded and power requirements have decreased there has been an ongoing explosion of devices that operate completely independent of wires or power cords. These "untethered" devices range from cell phones, and wireless keyboards to building sensors and active RFID tags.

[0003] Engineers and designers of these untethered devices continue to have to deal with the limitations of portable power sources, primarily batteries as the key design parameter. While performance of processors and portable devices have been doubling every 18-24 months driven by Moore's law, battery technology in terms of capacity has only been growing at measly 6% per year. Even with power conscious designs and the latest in battery technology, many devices do not provide the lifetime cost and maintenance requirements for applications that require a large number of untethered devices such as logistics, and building automation. Today's devices that need two-way communication require scheduled maintenance every three to 18 months to replace or recharge the device's power source (typically a battery). One-way devices simply broadcasting their status (one-way) such as automated utility meter readers have a better battery life, typically requiring replacement within 10 years. For both device types, scheduled power-source maintenance is costly and disruptive to the entire system that a device is intended to monitor and/or control. Unscheduled maintenance trips are even more costly and disruptive. On a macro level, the relatively high cost associated with the internal battery also reduces the practical, or economically viable, number of devices that can be deployed.

[0004] The ideal solution to the power problem for untethered devices is a device or system that can collect and harness sufficient energy from the external environment. The harnessed energy would then either directly power an untethered device or augment a battery or other storage component. Directly powering an untethered device enables the device to be constructed without the need for a battery. Augmenting a storage component could be along two lines: 1) increasing the overall life of the device or 2) by providing more power to the device to increase the functionality of the device. The other parameters for an ideal solution is that the harnessing device could be used in a wide range of environments including harsh and sealed environments (e.g. nuclear reactors), would be inexpensive to produce, would be safe for humans, and would have a minimal effect on the basic size, weight and other physical characteristics of the untethered device.

BRIEF SUMMARY OF THE INVENTION

[0005] The present invention pertains to a device for receiving wireless power. The device comprises a point of reception, wherein the point of reception is positionable in at least a first position and a second position.

[0006] The present invention pertains to a method for receiving wireless power. The method comprises the steps of positioning a point of reception in contact with a housing to a first position. There is the step of receiving wireless power at the point of reception and providing it to a power harvester in the housing. There is the step of converting the wireless power to usable DC with the power harvester. There is the step of providing the usable DC to the core components in the housing. There is the step of using the DC by the core components. There is the step of repositioning the point of reception to a second position. There is the step of receiving wireless power at the point of reception at the second position and providing it to the power harvester. There is the step of converting the wireless power received by the point of reception in the second position to usable DC with the power harvester. There is the step of providing the usable DC to the core component in the housing. There is the step of using the DC by the core components.

[0007] The present invention pertains to an apparatus for an application. The apparatus comprises a core device preferably having an integrated circuit for the application. The apparatus comprises a power harvester connected to the core device to power the core device.

[0008] The present invention pertains to an apparatus for an application. The apparatus comprises a core device having an integrated circuit for the application. The apparatus comprises means for receiving energy wirelessly and providing power from the energy to the core device to power the integrated circuit of the core device. The receiving means is connected to the core device.

[0009] The present invention pertains to a method for an application. The method comprises the steps of converting RF energy into usable energy. There is the step of preferably powering an integrated circuit of the core device with the usable energy.

[0010] This invention pertains to a technique that uses radio frequency (RF) energy as a source of energy to directly power or augment a power storage component in an untethered device. The present invention meets the requirements described in the previous "Background of the Invention" section.

[0011] Traditional RF receiving devices have focused on maximizing selectivity of the frequency to isolate and to be coherent without interference from other sources. In contrast, while this methodology operates at a specific frequency or range of frequencies, the device accepts any interference to supplement the output power of the device. Also, the research related to power harvesting that uses RF energy as a source has primarily focused on devices in close proximity of the source. In most cases, prior research assumed a dedicated or directed source of RF to power the device.

[0012] It is an object of this invention to provide a method and apparatus to

[0013] 1. remotely energize an untethered device without using direct wiring

[0014] 2. power or augment the life of the power storage component so it matches the life of the device and, ultimately, powers the off-grid device with or without the use of batteries

[0015] 3. allow untethered devices to be virtually maintenance free

[0016] 4. provide augmentation for other energy harvesting technologies (solar, piezoelectric, etc.)

[0017] 5. provide backup power to tethered devices

[0018] It is a further object of this invention to directly power or augment a power storage component in an untethered device in conjunction with other power harvesting technologies and storage elements.

[0019] With this method and apparatus a device's power storage components do not require replacement, thus enabling the device to be permanently placed off-grid, where it may be physically impractical, costly, or dangerous (due to a harsh environment) to provide maintenance.

[0020] For devices on-grid (tethered) or with reliable power sources, RF power harvesting can be used as a backup in case the primary power source is lost.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING

[0021] In the accompanying drawings, the preferred embodiment of the invention and preferred methods of practicing the invention are illustrated in which:

[0022] FIG. **1** is a block diagram of the RF Power Harvesting block used to directly supply power to the Core Device Components.

[0023] FIG. **2** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0024] FIG. **3** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to the Core Device Components.

[0025] FIG. **4** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0026] FIG. **5** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to the Core Device Components.

[0027] FIG. **6** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the Power Storage block and used to supply power to the Core Device Components.

[0028] FIG. **7** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the Power Storage block.

[0029] FIG. **8** is a block diagram of the RF Power Harvesting block in communication with the Power Storage block.

[0030] FIG. **9** is a block diagram of the RF Power Harvesting block in communication with the Power Storage block and used to supply power to the Core Device Components.

[0031] FIG. **10** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0032] FIG. **11** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to the Core Device Components.

[0033] FIG. **12** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block.

[0034] FIG. **13** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0035] FIG. **14** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to the Core Device Components.

[0036] FIG. **15** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to the Core Device Components.

[0037] FIG. **16** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to the Core Device Components.

[0038] FIG. **17** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block.

[0039] FIG. **18** is a block diagram of the RF Power Harvesting block supplying power to the Power Storage Charger.

[0040] FIG. **19** is a block diagram of the RF Power Harvesting block supplying power to the Power Storage Charger and the RF Power Harvesting block in communication with the Power Storage block.

[0041] FIG. **20** is a block diagram of the RF Power Harvesting block supplying power to the Power Storage Charger and the Core Device Components.

[0042] FIG. **21** is a block diagram of the RF Power Harvesting block supplying power to the Power Storage Charger and the Core Device Components and the RF Power Harvesting block in communication with the Power Storage block.

[0043] FIG. **22** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0044] FIG. **23** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0045] FIG. **24** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the Power Storage block.

[0046] FIG. **25** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to the Core Device Components.

[0047] FIG. **26** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0048] FIG. **27** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the Power Storage block.

[0049] FIG. **28** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the RF Power Harvesting block supplying power to the Core Device Components.

[0050] FIG. **29** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit.

[0051] FIG. **30** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to the Core Device Components.

[0052] FIG. **31** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the Power Storage block.

[0053] FIG. **32** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power the Core Device Components.

[0054] FIG. **33** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the Power Storage block and used to supply power to the Core Device Components.

[0055] FIG. **34** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block.

[0056] FIG. **35** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to the Core Device Components.

[0057] FIG. **36** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the Power Storage block and used to supply power to the Core Device Components.

[0058] FIG. **37** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to the Core Device Components.

[0059] FIG. **38** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and supplies power to the Power Storage Charger.

[0060] FIG. **39** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and supplies power to the Power Storage Charger.

[0061] FIG. **40** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the Power Storage block and supplies power to the Power Storage Charger.

[0062] FIG. **41** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to the Core Device Components and supplies power to the Power Storage Charger.

[0063] FIG. **42** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and supplies power to the Power Storage Charger.

[0064] FIG. **43** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the Power Storage block and supplies power to the Power Storage Charger.

[0065] FIG. **44** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the RF Power Harvesting block supplying power to the Core Device Components and supplies power to the Power Storage Charger.

[0066] FIG. **45** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and supplies power to the Power Storage Charger.

[0067] FIG. **46** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to the Core Device Components and supplies power to the Power Storage Charger.

[0068] FIG. **47** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the Power Storage block and supplies power to the Power Storage Charger.

[0069] FIG. **48** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power the Core Device Components and supplies power to the Power Storage Charger.

[0070] FIG. **49** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the Power Storage block and used to supply power to the Core Device Components and supplies power to the Power Storage Charger.

[0071] FIG. **50** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and supplies power to the Power Storage Charger.

[0072] FIG. **51** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and used to supply power to the Core Device Components and supplies power to the Power Storage Charger.

[0073] FIG. **52** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and the Power Storage block and used to supply power to the Core Device Components and supplies power to the Power Storage Charger.

[0074] FIG. **53** is a block diagram of the RF Power Harvesting block in communication with a Power Regulation and/or Power Storage Circuit and Power Storage block and used to supply power to the Core Device Components and supplies power to the Power Storage Charger.

[0075] FIG. **54** is a block diagram of the RF Power Harvesting block using Antenna A to directly supply power to the Core Device Components.

[0076] FIG. **55** is a block diagram of the RF Power Harvesting block using Antenna A to supply power to the Power Regulation, Storage and/or Storage Charging block.

[0077] FIG. **56** is a block diagram of the RF Power Harvesting block using Antenna A to supply power to the Power Regulation, Storage and/or Storage Charging block and used to supply power to the Core Device Components.

[0078] FIG. **57** is a block diagram of the RF Power Harvesting block used to directly supply power to the Core Device Components.

[0079] FIG. **58** is a block diagram of the RF Power Harvesting block used to supply power to the Power Regulation, Storage and/or Storage Charging block.

[0080] FIG. **59** is a block diagram of the RF Power Harvesting block used to supply power to the Power Regulation, Storage and/or Storage Charging block and used to supply power to the Core Device Components.

[0081] FIG. **60** is a block diagram of the RF Power Harvesting block used to directly supply power to the Core Device Components.

[0082] FIG. **61** is a block diagram of the RF Power Harvesting block used to directly supply power to the Core Device Components and in communication with the Alternative Power Sources block.

[0083] FIG. **62** is a block diagram of the RF Power Harvesting block in communication with the Alternative Power Sources block and used to directly supply power to the Core Device Components.

[0084] FIG. **63** is a block diagram of the RF Power Harvesting block in communication with the Alternative Power Sources block.

[0085] FIG. **64** is a block diagram of the RF Power Harvesting block in communication with the Power Regulation, Storage and/or Storage Charging block.

[0086] FIG. **65** is a block diagram of the RF Power Harvesting block in communication with the Power Regulation, Storage and/or Storage Charging block.

[0087] FIG. **66** is a block diagram of the RF Power Harvesting block in communication with the Power Regulation, Storage and/or Storage Charging block and the Alternative Power Sources block.

[0088] FIG. **67** is a block diagram of the RF Power Harvesting block in communication with the Power Regulation, Storage and/or Storage Charging block.

[0089] FIG. **68** is a block diagram of the RF Power Harvesting block in communication with the Power Regulation, Storage and/or Storage Charging block and the Alternative Power Sources block.

[0090] FIG. **69** is a block diagram of the RF Power Harvesting block in communication with the Alternative Power Sources block and Power Regulation, Storage and/or Storage Charging block.

[0091] FIG. **70** is a block diagram of the RF Power Harvesting block in communication with the Alternative Power Sources block and Power Regulation, Storage and/or Storage Charging block.

[0092] FIG. **71** is a block diagram of the Alternative Power Sources block.

[0093] FIG. **72** is a block diagram of the RF Power Harvesting block in communication with the Alternative Power Sources block.

[0094] FIG. **73** is a block diagram of the RF Power Harvesting block in communication with the Alternative Power Sources block.

[0095] FIG. **74** is a block diagram of the RF Power Harvesting block used to supply power to the Core Device Components.

[0096] FIG. **75** is a block diagram of the RF Power Harvesting block used to supply power to the Core Device Components.

[0097] FIG. **76** is a block diagram of the RF Power Harvesting block used to supply power to the Core Device Components and in communication with the Alternative Power Sources block.

[0098] FIG. **77** is a block diagram of the RF Power Harvesting block used to supply power to the Core Device Components.

[0099] FIG. **78** is a block diagram of the RF Power Harvesting block used to supply power to the Core Device Components and in communication with the Alternative Power Sources block.

[0100] FIG. **79** is a block diagram of the RF Power Harvesting block used to supply power to the Core Device Components and in communication with the Alternative Power Sources block.

[0101] FIG. **80** is a block diagram of the RF Power Harvesting block used to supply power to the Core Device Components and in communication with the Alternative Power Sources.

[0102] FIG. **81** is a block diagram of the RF Power Harvesting block used to supply power to the Core Device Components and in communication with the Power Regulation, Storage and/or Storage Charging block.

[0103] FIG. **82** is a block diagram of the RF Power Harvesting block used to supply power to the Core Device Components and in communication with the Power Regulation, Storage and/or Storage Charging block.

[0104] FIG. **83** is a block diagram of the RF Power Harvesting block used to supply power to the Core Device Components and in communication with the Power Regulation, Storage and/or Storage Charging block and the Alternative Power Sources block.

[0105] FIG. **84** is a block diagram of the RF Power Harvesting block used to supply power to the Core Device Components and in communication with the Power Regulation, Storage and/or Storage Charging block.

[0106] FIG. **85** is a block diagram of the RF Power Harvesting block used to supply power to the Core Device Components and in communication with the Power Regulation, Storage and/or Storage Charging block and the Alternative Power Sources block.

[0107] FIG. **86** is a block diagram of the RF Power Harvesting block used to supply power to the Core Device Components and in communication with the Alternative Power Sources block and Power Regulation, Storage and/or Storage Charging block.

[0108] FIG. **87** is a block diagram of the RF Power Harvesting block used to supply power to the Core Device Components and in communication with the Alternative Power Sources block and Power Regulation, Storage and/or Storage Charging block.

[0109] FIG. **88** is a block diagram of the entire power system for the device.

[0110] FIG. **89** is a block diagram of a power harvesting block used to supply power to a core device having a sensor.

[0111] FIG. **90** is a block diagram of a power harvesting block used to supply power to a core device having a computer peripheral.

[0112] FIG. **91**(*a*) is a perspective view illustration of a first embodiment of a cell phone according to the present invention.

[0113] FIG. **91**(*b*) is a side view illustration of the first embodiment of the cell phone.

[0114] FIG. **92** is a side view illustration of a second embodiment of a cell phone according to the present invention.

[0115] FIG. **93** is a perspective view illustration of a third embodiment of a cell phone according to the present invention.

[0116] FIG. **94**(*a*) is a front view illustration of a fourth embodiment of a cell phone according to the present invention.

[0117] FIG. **94**(*b*) is a side view illustration of the fourth embodiment of the cell phone.

[0118] FIG. **95**(*a*) is a front view illustration of a fifth embodiment of a cell phone according to the present invention.

[0119] FIG. **95**(*b*) is a side view illustration of the fifth embodiment of the cell phone.

[0120] FIG. **96** is an illustration of a sixth embodiment of a cell phone according to the present invention.

[0121] FIG. **97** is an illustration of a first embodiment hinge of a seventh embodiment of a cell phone according to the present invention.

[0122] FIGS. **98**(*a*) and (*b*) are side view illustrations of a second embodiment hinge of a seventh embodiment of a cell phone according to the present invention.

[0123] FIG. **99** is an illustration of a third embodiment hinge of a seventh embodiment of a cell phone according to the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0124] A complete understanding of the invention will be obtained from the following description when taken in connection with the accompanying drawing figures wherein like reference characters identify like parts throughout.

[0125] There is shown an apparatus **10** for an application. The apparatus **10** comprises a core device **22** preferably having an integrated circuit for the application. The apparatus **10** comprises a power harvester **20** connected to the core device **22** to power the core device **22**.

[0126] The apparatus **10** preferably includes an alternative power source **24** connected to the core device **22** to power the core device **22** in conjunction with the power harvester **20**. Preferably, the apparatus **10** includes a power regulator **26** and/or power storage circuit **28** connected to the power harvester **20**. The apparatus **10** preferably includes a power storage charger **30** connected to the power harvester **20**. Preferably, the apparatus **10** includes a power storage connected to the power harvester **20**.

[0127] Preferably, the core device **22** includes a memory connected to the integrated circuit and to the power harvester **20** to power memory.

[0128] The core device **22** can include a sensor **32**, as shown in FIG. **89**. The sensor **32** can include a proximity sensor, an intrusion sensor, an environmental sensor, a chemical sensor, a biological sensor, a sensor in contact with an automobile, an occupancy sensor, a motion sensor, a position sensor, a metal detector, or a sensor **32** in contact with an aircraft. The sensor **32** can include an alarm connected with the power harvester **20** to power the alarm, a display connected with the power harvester **20** to power the display, a sensor **32** disposed in a building, an industrial automation sensor, a sensor **32** in contact with an elevator, a temperature sensor, a fire sensor, an accelerometer, or a level sensor.

[0129] The sensor **32** can include a gas level sensor, a fluid level sensor, a light level sensor, a flow sensor, or a gas flow sensor, a fluid flow sensor, a light flow sensor, or a plasma flow sensor.

[0130] The sensor **32** can include a pressure sensor, a gas pressure sensor or a fluid pressure sensor, a fluid pressure sensor, a light sensor, an infrared light sensor, an ultraviolet light sensor, an x-ray sensor, a cosmic ray sensor, a visible light sensor, or a gamma ray sensor, a stress sensor, a strain sensor, a depth sensor, or an electrical characteristic sensor.

[0131] The sensor **32** includes a voltage sensor, a current sensor, a viscosity sensor, an acoustical sensor, a sound sensor, a listening sensor, a thickness sensor, a density sensor, a surface quality sensor, a volume sensor, a physical sensor, a mass sensor, a weight sensor, a conductivity sensor, a distance sensor, an orientation sensor, or a vibration sensor.

[0132] The sensor **32** can include a radioactivity sensor, a field strength sensor, an electric field sensor or a magnetic field sensor, a smoke detector, a carbon monoxide detector, a radon detector, an air quality sensor, a humidity sensor, a glass breakage sensor, or a break beam detector. The sensor can include a thermal energy sensor, an electromagnetic sensor, a mechanical sensor, an optical sensor, a radiation sensor, a sensor in contact with a vehicle, or a water craft.

[0133] The present invention pertains to an apparatus **10** for an application. The apparatus **10** comprises a core device **22** having an integrated circuit for the application. The apparatus **10** comprises means for receiving energy wirelessly and providing power from the energy to the core device **22** to power the integrated circuit of the core device **22**. The receiving means is connected to the core device **22**. Preferably, the core device **22** includes means for sensing.

[0134] Alternatively, the core device **22** can include a computer peripheral **34**, as shown in FIG. **90**. The computer peripheral **34** can include a handheld game, a gaming system, a game controller, a controller, a keyboard, a mouse, a computer terminal, computer storage, or computer equipment.

[0135] The present invention pertains to a method for an application. The method comprises the steps of converting rf energy into usable energy. There is the step of preferably powering an integrated circuit of the core device **22** with the usable energy.

[0136] Preferably, there is the step of regulating the usable energy provided to the core device **22**. There is preferably the step of storing the usable energy. Preferably, there is the step of providing power to the core device **22** from an alternative power source **24** in conjunction with the usable energy.

[0137] The present invention can be implemented in numerous ways. Most of these ways are depicted in FIGS. **1-88**. These figures contain multiple blocks that are configured in multiple ways. In the figures, an arrow represents the flow of power unless otherwise stated. Single-headed (one-way) arrows represent that the power is flowing from one block to another. The single-headed arrow may represent multiple wires that provide power from one block to multiple parts in the other block. Two-headed (two-way) arrows represent a single wire that can have power flow in either direction or multiple wires each having power flow in a single direction. As an example, a two-headed arrow between the RF Power Harvesting block and the Power Regulation and/or Power Storage circuit **28** block can represent a single wire that allows harvested power to flow into a storage device such as a capacitor. The same block diagram can also represent two wires between the two blocks with the first wire allowing harvested power to flow into a voltage regulator **26**. The second wire can allow the regulated voltage to feedback to the RF Power Harvesting block to provide power to internal components such as transistors to increase the performance of the RF Power Harvesting block. Each block is described in detail below. Each block represents the functionality described below associated with it. For instance, the RF power harvesting block describes a power harvester **20**, and the power regulation block describes a power regulator **26**.

[0138] RF Power Harvesting Block

[0139] The RF Power Harvesting Block is used to convert the energy captured by the antenna into usable power such as DC voltage. This block may include antenna matching, rectifying circuitry, voltage transforming circuitry, and/or other performance optimizing circuitry. The rectifying circuitry may include a diode(s), a transistor(s), or some other rectifying device or combination. Examples of the rectifying circuitry include but are not limited to half-wave, full-wave, and voltage doubling circuits. The RF power harvesting block is connected to an antenna that may or may not be used as the communications antenna for the core device **22** components. The output of the RF Power Harvesting Block is a DC voltage or current. The RF Power Harvesting Block may accept feedback (or input) from other circuitry or blocks, which may be used to control the harvesting circuitry to improve the performance or vary the output. This feedback may include but is not limited to a DC voltage or a clock from the Core device **22** Components. U.S. Pat. No. 6,615,074 (FIGS. 8, 9, 12a, 12b, 13, 14), incorporated by reference, herein, shows numerous examples of RF power harvesting circuits that can be used to implement the block and function described.

[0140] Power Regulation and/or Power Storage Circuit **28** Block

[0141] It may be necessary to regulate the converted power (hold the power at a constant level) for specific devices. The devices that would need this block require a fairly constant voltage or current. Deviations from the required values may cause the device to not perform within its specifications. The regulation can be implemented in many different ways. The block can be as simple as using a Zener diode, or as complicated as using an integrated circuit such as a linear voltage regulator **26** or switching regulator **26** to hold the voltage at a constant level. Certain devices have a more tolerable power requirement. For these devices, the regulation stage may be excluded. This block may also include, with or without the regulation, a storage device such as a capacitor, a battery, or some other device able to store charge. The output from the Power Regulation and/or Power Storage circuit **28** Block may be used as feedback to other blocks within the Device's Power System or to the Alternative power source **24** if they require a regulated supply voltage or stored power. U.S. Pat. No. 6,894,467 (FIGS. 1, 3), Linear Voltage Regulator, incorporated by reference, herein, is an example of a practical application of implementing the regulation described in the block. U.S. Pat. No. 6,297,618 (FIGS. 1-4), Power storage device and method of measuring voltage of storage battery, incorporated by reference, herein, is an example of a practical application of implementing the storage described in the block.

[0142] Power Storage Charger **30** Block

[0143] The Power Storage Charger **30** Block may be needed if the storage component requires a special charging mechanism such as pulse charging or trickle charging. This block controls how the captured and converted power is supplied to the storage device. U.S. Pat. No. 6,836,095 (FIGS. 1-3), Battery Charging Method and Apparatus, incorporated by reference herein, is an example of a practical application of implementing the special charging mechanism described in the block.

[0144] Power Storage Block

[0145] If a device has intermittent power requirements, it may be necessary to store the captured power for use at a later time. The power can be stored in the Power Storage block, which could include a battery, a capacitor, and/or another type of power storage component. Storage components include but are not limited to batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other storage elements. The output from the Power Storage Block may be used as feedback to other blocks within the Device's Power System or to the Alternative power source **24** if they require a dedicated and predictable supply voltage. U.S. Pat. No. 6,297,618 (FIGS. 1-4), Power Storage Device and Method of Measuring Voltage of Storage Battery, incorporated by reference herein, is an example of a practical application of implementing the storage described in the block. U.S. Pat. No. 6,835,501, Alkaline Rechargeable Battery, incorporated by reference herein, is also an example of a practical application of implementing the storage described in the block.

[0146] Core Device **22** Components Block

[0147] The Core Device **22** Components Block is the device that is receiving power from the system. This block may be but is not limited to the devices listed in the subsequent pages of this document. It may be advantageous for the Core Device **22** Components to communicate with any of the blocks that are supplying power to it. This communication can include but is not limited to a feedback control signal such as a clock or an ON/OFF command. As an example, the device may want to turn off the Alternative Power Sources **24** block if it is receiving sufficient power from the RF Power Harvesting block.

[0148] Alternative Power Sources **24** Block

[0149] RF energy harvesting also has the ability to be augmented by other types of power harvesting, storage components, or dedicated sources (e.g. power line). The Alternative Power Sources **24** Block shows how this type of system could be implemented. The augmenting power harvesting technologies include but are not limited to solar, light (visible and non-visible), piezoelectric, vibration, acoustic, thermal, microgenerators, wind, and other environmental elements. This block can work independently or have communication with other blocks. U.S. Pat. No. 6,784,358, Solar Cell Structure Utilizing an Amorphous Silicon Discrete By-Pass Diode, incorporated by reference herein, is an example of a practical application of implementing an alternative power source **24** described by the block. U.S. Pat. No. 6,858,970, Multi-Frequency Piezoelectric Energy Harvester, incorporated by reference herein, is also an example of a practical application of implementing an alternative power source **24** described by the block.

[0150] Power Regulation, Storage and/or Storage Charging Block

[0151] The Power Regulation, Storage and/or Storage Charging block contains all the combinations of the Power Regulation and/or Power Storage circuit **28**, Power Storage Charger **30**, and Power Storage block. This block is used in the later figures to reduce the number of figures needed to show how the blocks can interconnect.

[0152] The disclosed invention is the application for retrieving radio frequency (RF) energy by an antenna, converting that energy into direct current (DC) power, regulating that energy using an optimized circuit, storing that energy in an optimized component, and/or supplying the power for specific devices. FIGS. **1-53** show how the system could be implemented.

[0153] Retrieval of RF Energy

[0154] The RF energy is retrieved from the environment by the use of an antenna. The antenna can be shared or standalone with respect to an antenna used for the device's wireless communication. FIGS. **54-56** show a device that has an antenna A for use by the RF harvesting apparatus **10** and an antenna B used for wireless one-way or two-way communication. FIGS. **57-59** show a device where the antenna is shared by the both the device's communication module and RF harvesting apparatus **10**. In terms of form factor, the antenna used by the apparatus **10** can be a separate component or integrated directly into the form factor of the device. The antenna is able to capture two types of available RF energy. The first type of energy exists as ambient RF energy. This type of RF surrounds us in our day-to-day lives and is usually generated to carry one way or two-way combinations of voice, video and data communications. The sources that the antenna can harvest from include medium-frequency AM radio broadcast, very-high-frequency (VHF) FM radio broadcast and television broadcast, ultra-high-frequency (UHF) broadcast, cellular base stations, wireless data access points, super-high-frequency (SHF) frequencies, and the industrial, scientific, and medical (ISM) bands. These sources cover transmission frequencies from 300 kHz to 30 GHz.

[0155] The second type of energy available is directed RF energy. This type of RF energy is directed from a transmitter specifically designed to deliver RF energy for harvesting by the antenna. The transmitter can be configured as a standalone device or integrated into an existing device.

[0156] Conversion of the Energy into DC

[0157] The RF energy captured by the antenna must be converted into a useful form of energy for the specific device. This conversion is shown in block form in all FIGS. (**1-88**) as the RF Power Harvesting Block. The most common form of useable energy is DC energy. To perform this conversion, the block includes circuitry to rectify the captured alternating current (AC) energy to create DC energy. The rectification in this block can be done with a diode(s), a transistor(s), or some other rectifying device or combination.

[0158] Regulation of the Energy

[0159] It may be necessary to regulate the converted power (hold the power at a constant level) for specific devices. FIGS. **2-7**, **10-17**, and **22-53** show how this regulation can be added to the circuit using a Power Regulation and/or Power Storage block. The devices that would need this block require a fairly constant voltage or current. Deviations from the required values may cause the device to not perform within its specifications. The regulation can be implemented in many different ways. The block can be as simple as using a Zener diode, or as complicated as using an integrated circuit such as a linear voltage regulator **26** or a switching regulator **26** to hold the voltage at a constant level. Certain devices have a more tolerable power requirement. For these devices, the regulation stage may be excluded.

[0160] Storage of the Energy

[0161] If a device has intermittent power requirements, such as the devices exampled by FIGS. **2-53**, it may be necessary to store the captured power for use at a later time. The power can be stored in the Power Storage block or the Power Regulation and/or Power Storage block. Storage devices can include but are not limited to a battery, a capacitor, or another type of power storage component. In certain applications, it may be necessary to include additional circuitry that controls how the power is transferred to the storage device. The Power Storage Charger **30** block is shown in FIGS. **18-53**. This may be needed if the storage component requires a special charging mechanism such as pulse charging or trickle charging. Storage components include but are not limited to batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other storage elements. There are devices that will not require storage. These devices can run directly off of the converted power. These devices also may or may not require regulation of the captured power.

[0162] Supplying the Power

[0163] The captured DC power, which may or may not be regulated and/or stored, is supplied to the device, which is represented by the Core Device **22** Components block in the figures. This may be a single connection or it may supply multiple parts of the device with power.

[0164] RF energy harvesting also has the ability to be augmented by other types of power harvesting or storage components. Other power harvesting technologies include but are not limited to solar, light (visible and non-visible), piezoelectric, vibration, acoustic, thermal, microgenerators, wind, and other environmental elements. Storage components include but are not limited to batteries (rechargeable and non-rechargeable), capacitors, inductors, fuel cells, and other storage elements. FIGS. **60-88** show how the alternative power can be connected to an RF energy harvesting system. These figure show how the RF energy harvesting components and the alternative power source **24** (or sources) can work independently or have communication with each other. The antenna configurations shown in FIGS. **54-59** are still applicable with the addition of an alternative power source(s) **24**. These antenna configurations can be applied to FIGS. **60-88**.

[0165] RF energy harvesting also has the ability to provide a backup to devices on-grid (tethered) or with reliable power sources, which can be used in case the primary power source is lost. As an example, it may be mandated by regulations that a sensor has auxiliary power in case the primary supply is lost. It could be possible to use a rechargeable battery that obtains its charge from the primary supply when in operation. However, if the primary supply is lost for a time greater than the life of the rechargeable battery, the specification of uninterrupted power is not met. RF energy could be used to supply power to the described device while the primary supply is not available. The primary supply could include but is not limited to an on-grid connection, a generator, a battery, or other reliable power supply.

[0166] RF energy harvesting with or without alternative source augmentation is applicable to provide electric power directly or indirectly to a range of electronic components contained in any specific electrical or electronic device and includes but is not limited to:

[0167] Passive electronic components, active electronic components

[0168] Resistors, fixed resistors, variable resistors, thermistors, thyristor, thermocouple

[0169] Capacitors, Electrolytic Capacitors, Tantalum Capacitors, Ceramic Capacitors, Multilayer Ceramic Capacitors, Polystyrene Film Capacitors, Electric Double Layer Capacitors (Super Capacitors), Polyester Film Capacitors, Polypropylene Capacitors, Mica Capacitors, Metallized Polyester Film Capacitors, Variable Capacitors

[0170] Diodes, Voltage regulation diodes, light-emitting diodes, organic light-emitting diodes, Variable capacitance diodes, Rectification diodes, Switching diodes, Regulation Diodes, Diode bridges, Schottky barrier diodes, tunnel diodes, PIN diodes, Zener diodes, Avalanche diodes, TVSs

[0171] Integrated circuits, microcontroller unit (MCU), microprocessor unit (MPU), logic circuits, memory, printed circuits, circuit boards, printed wiring boards

[0172] Transistors, MOSFETs, FETs, BJTs, JFETs, IGBTs, Relays, Antennas, semiconductors, conductors, inductors, relays, diacs, triacs, SCRs, MOVs

[0173] Fuses, circuit breakers

[0174] Batteries, Non-rechargeable batteries, rechargeable batteries, coin cell batteries, button cell batteries, alkaline batteries, lithium batteries, lithium ion batteries, lithium polymer batteries, NIMH batteries, NICAD batteries, Lead acid batteries, Zinc air batteries, Manganese Lithium batteries, Niobium Titanium Lithium batteries, Vanadium Pentoxide Lithium batteries, Carbon Zinc batteries, Zinc Chloride batteries, Lithium Thionyl Chloride batteries, Manganese Dioxide batteries, Lithium Poly-Carbonmonofluoride batteries, Lithium Manganese Dioxide batteries, Lithium Chloride batteries, Lead Acid Calcium batteries, Lead Acid Tin batteries, Oxy Nickel batteries, Silver Oxide batteries, Magnesium batteries

[0175] Inductors, Coils, High Frequency Coils, Toroidal Coils, Transformers, switches, chokes

[0176] Motors, DC motors, stepper motor, AC motors, Fans

[0177] Crystals, Oscillators, Clocks, Timers

[0178] Displays, LCDs, LED displays

[0179] RF energy harvesting with or without alternative source augmentation is applicable across a range of markets and specific devices and includes but is not limited to:

[0180] Consumer Electronics

[0181] Electronic equipment, wired devices, battery powered devices, wireless communication devices, cell phones, telephones, phones, cordless phones, portable phones, Bluetooth devices, Bluetooth headsets, hands-free headsets, headsets, headphones, Wireless headsets, radios, AM/FM radios, shortwave radios, weather radios, Two-way radios, portable radio, lights, lanterns, portable lights, flashlights, nightlights, spotlights, search lights, calculators, graphing calculators, desk calculators, clocks, alarm clocks, wall clocks, desk clocks, travel clocks, watches, wristwatches, pocket watches, stop watches, timers, voice recorders, Dictaphones, laser pointers, power tools, cordless power tools, electronic razors, electric razors, handheld games, gaming systems, game controllers, wireless game controllers, remote controls, battery chargers, computers, portable computers, keyless entry, toys, toy guns, toy laser guns, games, microphones, musical instruments, musical effects processors, musical instrument tuners, metronomes, electronic chord charts, door openers, garage door openers, PDA, Cameras, Video recorders, Multimeter, electronic test equipment, hand-held electronics, portable electronics, wireless pens, sound generators, noise generators, language translators, electric toothbrushes, portable televisions, pagers, transceivers, toy vehicles, remote control vehicles, toy planes, remote control planes, pet containment systems, invisible fence pet sensors, memory backup, base station battery backups, appliance battery backups, uninterrupted power supplies, GPS devices, memory retention power supplies, metal detectors, stud finders, metal stud finders, stun guns, tazers, wearable devices, baby monitors, intercoms, doorbells, wireless doorbells, electronic office supplies, electronic staplers, radar jammers, radar detectors, digital scales, microfilm cassettes, video head testers, compasses, noise canceling headphones, air samplers, depth finders, barometers, weather measurement instruments, data transfer devices, automatic distress signaling unit, Wireless audio speakers, Satellite radios, Police scanners, Car navigation systems (GPS devices), Decorative lights, Christmas lights, garden lights, lawn lights, ornamental lights, porch lights

[0182] Multi-media players: MP3, DVD, analog music players, CD players, tape players, digital music players, digital video players, minidisc

[0183] Computer: keyboards, mice, peripherals, computer equipment, electronic computers, computer storage, computer terminals

[0184] Building/Industrial Automation

[0185] Sensors: Position, elevator, temperature, fire, accelerometers, level, gas level, fluid level, light level, flow, gas flow, fluid flow, light flow, plasma flow, pressure, gas pressure, fluid pressure, motion, light, infrared light, ultraviolet light, X-rays, cosmic rays, visible light, gamma rays, chemical, stress, strain, depth, electrical characteristics, voltage, current, viscosity, acoustical, sound, listening, thickness, density, surface quality, volume, physical, mass, weight, force, conductivity, distance, orientation, vibration, radioactivity, field strength, electric field strength, magnetic field strength, occupancy, smoke detector, carbon monoxide detector, radon detector, air quality, humidity, glass breakage, break beam detector

[0186] Controls: Position, elevator, temperature, fire, accelerometers, level, gas level, fluid level, light level, flow, gas flow, fluid flow, light flow, plasma flow, pressure, gas pressure, fluid pressure, motion, light, infrared light, ultraviolet light, X-rays, cosmic rays, visible light, gamma rays, chemical, stress, strain, depth, electrical characteristics, voltage, current, viscosity, acoustical, sound, listening, thickness, density, surface quality, volume, physical, mass, weight, force, conductivity, distance, orientation, vibration, radioactivity, field strength, electric field strength, magnetic field strength, occupancy, smoke detector, carbon monoxide detector, radon detector, air quality, humidity

[0187] Devices: Thermostats, light switches, door locks, smart-card door locks, lighting, emergency lighting, motion lighting, safety lighting, highway lighting, construction lighting, sign lighting, roadway sign lighting, construction sign lighting, automatic flushing units, automatic soap dispenser, automatic paper towel dispenser, automatic faucets, automatic door sensors, identification reader, fingerprint reader, credit card readers, card readers, valve actuators, gauges, analog gauges, digital gauges, fire extinguishers, wireless switches, remotely operated inspection equipment, gas/oil pipeline monitoring systems, robotic pipeline inspection gauges, “auto-reclosers” for electric power lines, sonar buoys, telemetry systems, electronic record tracking systems, robbery tracking devices, interrogators, programmers, emergency exit alarms, alarms, flood alarms, gas alarms, electronic entry systems, security keypads, silo transducers, data recorders, signal tracers, anti-static strap testers, radiosonde weather balloons, utilities load controllers, profilometers, noise cancellation equipment, infrared beacons

[0188] Military/Government

[0189] Tracking tags: Weapons, vehicle, soldier, gear/assets, staff, general population, security access badges

[0190] Sensors: Proximity, intrusion, environmental, chemical/biological

[0191] Equipment: Battery charger, surveillance, card readers, identification reader, fingerprint reader, retinal scanners, satellites, rockets, space vehicles, search and rescue transponders (SARTs), emergency position-indicating rescue beacons (EPIRBs), emergency locator transmitters (ELTs), military radios, electronic toll collection systems, postal tracking systems, communications, thermal imaging, night vision, training targets, field medical equipment, house arrest monitors, laser tags, electronic parking meters, multiple integrated laser engagement system, munitions and mines, ship sensors

[0192] Utility

[0193] Gas consumption meters, water consumption meters, and electric consumption meters

[0194] Logistics & Supply Chain Management

[0195] Radio-frequency identification devices (RFID), RFID readers

[0196] Tracking: Asset tags, cargo container location beacons, transponders, transceivers

[0197] Devices: Smart price tags, smart shelving, handheld barcode scanner, barcode scanners, credit card readers, card readers, retail signage, hotel door locks

[0198] Homeland Security

[0199] Sensors: Occupancy, proximity, environmental, chemical/biological, motion, position

[0200] Metal detector wand

[0201] Medical

[0202] Implantable: cochlear implants, neural stimulators, pace makers, medication administration, defibrillator

[0203] Body function monitors: pressure, temperature, respiration, blood oxygenation, insulin, hearing aid, pulse, EKG, heart, Holter,

[0204] Tracking tags: Patient, baby identification, assets, supplies, staff, medication, instruments

[0205] Devices: Home healthcare equipment, ambulatory infusion pumps, blood analyzers, biofeedback systems, bone growth stimulators, thermometers, digital thermometers, stimulators, galvanic stimulators, muscle stimulators, pediatric scales

[0206] Agriculture—livestock tracking and asset tracking.

[0207] Tracking: livestock, asset, wildlife tracking devices

[0208] Equipment: cattle prods

[0209] Automotive

[0210] Automotive antennas, Automotive Audio Systems, Automotive Lighting, Automotive Video Systems, Computers, Processors, Controls, Switches, Electric Motors, Actuators, Ignition Systems, Starter Systems, Injection Systems, Powertrain Electronics, Radar Detectors, Proximity Detectors, Safety Systems, Security Systems, Sensors, Regulators, Distributors, Vehicle Control, Wiper Systems, Washer Systems, Radio, Video Systems, Entertainment Systems, Navigation Systems, GPS systems, Power Mirror Systems, Emission control systems

[0211] Appliances

[0212] Monitoring systems and control systems for major and small appliances including washing machines, dryers, refrigerators, freezers, coolers, air conditioners, humidifiers, dehumidifiers, air purifiers, air filters, fans, furnaces, water heaters, boilers, space heaters, sowing machines, ice makers, microwave ovens, convection ovens, ovens, toaster ovens, ranges, range hoods, cooktops, stoves, stovetops, crock pots, hot plates, dishwashers, garbage disposals, can openers, vacuum cleaners, blenders, mixers, food processors, irons, coffee makers, toasters, grills, hair dryers, electric tooth brushes, electric razors, electric drills, electric screwdrivers, chainsaws, lawnmowers, push mowers, riding mowers, trimmers, brush cutters, pruners, edgers, vending machines

[0213] Ventilation, Heating, Air-Conditioning, and Commercial Refrigeration Equipment

[0214] Monitoring systems, control systems

[0215] Engine, Turbine, and Power Transmission Equipment

[0216] Monitoring systems, control systems

[0217] Other General Purpose Machinery Manufacturing

[0218] Monitoring systems, control systems

[0219] Telecommunications

[0220] Monitoring systems, control systems

[0221] Portable

[0222] Aircraft

[0223] Monitoring systems, control systems, actuator systems, sensors

[0224] It should be noted that devices within a specific category may be applicable across multiple areas even if they are not specifically listed. (e.g. temperature sensors apply to Industrial and Building Automation).

[0225] To retrofit or redesign the devices listed, it is possible to implement the described systems in numerous ways. It may be advantageous to leave the device design as is including the existing power supply. As an example, a device may use non-rechargeable batteries to operate. The device will most likely have a protection circuit to prevent damage if the batteries are installed incorrectly. The protection mechanism is commonly a diode inline with the positive terminal of the battery. In this case, the RF Power Harvesting Source with or without an Alternative power source **24** could be inserted, with an antenna, into the device. The power generated by the RF Power Harvesting Source (and alternative power source **24**, if applicable) could be connected to the device after the protection mechanism described to avoid potential charging of a non-rechargeable battery.

[0226] Another way to configure the system is to replace the non-rechargeable batteries with rechargeable batteries. In this instance, the output from the RF Power Harvesting Source (and alternative power source **24**) could be connected to either side of the protections device. If the connection is before the protection mechanism, the system will recharge the battery and supply power to the device. If the connection is after the protection mechanism, the system will supply power to the device and battery will supply any extra power needed that could not be supplied by the system. It should be noted that the protection device in this case is unneeded for proper operation. Its only function would be to protect the batteries from being installed incorrectly. An antenna could be contained inside or placed on the outside of the device.

[0227] Another configuration of the system is to remove the existing batteries and install the RF Power Harvesting Source (and alternative power source **24**) in the enclosure provided for the batteries. An antenna could be contained inside or placed on the outside of the device.

[0228] Yet another method of configuring the system would be to reduce the number of batteries and replace them with the RF Power Harvesting Source (and alternative power source **24**). In this case, the output from the system would be connected to the batteries in series or parallel depending on the original battery configuration. An antenna could be contained inside or placed on the outside of the device.

[0229] An additional option, would be to completely redesign the product and integrate the require circuitry and storage components into the device. This method is probably the most advantageous because it can fully take advantage of the benefits offered by the RF Power Harvesting Source (and alternative power source **24**). An antenna could be contained inside or placed on the outside of the device.

[0230] If the RF Power Harvesting Source (and alternative power source **24**) is used as a backup to the primary power supply, a switch could be implemented into the system in order to switch the RF Power Harvesting Source (and alternative power source **24**) in when the primary source is lost. In this case, an antenna could be contained inside or placed on the outside of the device.

[0231] To show the flexibility of RF energy harvesting, several products were retrofitted to include RF energy harvesting circuitry. These products include a wireless keyboard, a wall clock, and a desk calculator.

[0232] The wireless keyboard is an example of recharging and augmenting a battery to supply power to a device. This system is shown in FIG. **13**. The output from the regulation circuitry recharges the battery and supplies power to the keyboard. The battery is also used to supply power to the keyboard. The keyboard also includes a separate antenna for receiving power and for data communications. The antenna configuration can be seen in FIG. **55**.

[0233] The wall clock is an example of a direct powering system. The wall clock was retrofitted to include energy harvesting circuitry and the internal AA battery was removed. This system is shown in FIG. **2**. The wall clock did not need regulation but did require a capacitor for storage to supply the pulse of power to move the second hand.

[0234] The calculator is an example of using RF energy harvesting with another energy harvesting technology. The calculator had an internal 1.5V coin cell battery and a small solar panel. The internal battery was removed, however, the solar panel was left intact. This system is shown in FIG. **60**. In this system, the calculator can receive power from both the solar panel and the RF energy harvesting circuitry to eliminate the need for a battery.

[0235] As an additional example, an RF energy harvesting circuit similar to the ones shown in U.S. Pat. No. 6,615,074 (FIGS. 8, 9, 12a, 12b, 13, 14), incorporated by reference, herein, was connected in series with a 0.5V solar cell. Individually, the solar cell was able to provide 0.480V to a 10 kilo-ohm resistor, which was being used to simulate the Core Device **22** Components. This corresponds to 23 microwatts. The RF power harvesting circuit by itself was able to provide 2.093V across the 10 kilo-ohm resistor when being supplied by 1 milliwatt of RF power. This corresponds to 438 microwatts. The two circuit outputs were then combined in series by connecting the output from the RF energy harvesting circuit to the ground of the solar cell. The output of the solar cell was then connected to the resistor. The other end of the resistor was connected to the ground of the RF energy harvesting circuit. The voltage across the resistor with the circuits connected, as shown in FIG. **63**, was 2.445V. This corresponds to 598 microwatts. As can be seen, the combination of the two technologies produces a result higher than the addition of the individual powers. From this, it can be determined that the two technologies can cooperate in a way that produces favorable results. In the example given, the solar cell produces current to supply the load and helps to bias the RF rectifying diodes, which allows the RF energy harvesting circuit to operate a higher efficiency. The solar cell also changes the impedance seen by the RF energy harvesting circuit, which produces a beneficial result. To be more specific, when examining the power output of the individual circuit (solar and RF power harvesting), the sum of the power captured by the individual circuits was 23 uW+438 uW=461 uW. However, when the two circuits are combined and are allowed to work in conjunction with one another, the output power becomes 598 uW. This results shows that combining the two power-harvesting technologies produces a 30 percent increase in the output power for this example. This same technique can be applied to multiple energy harvesting technologies to produce even greater output power. The equations for this example are shown below.

[0236] Individual Circuits

$$P_I=P_1+P_2+\dots+P_N$$

[0237] Combined Circuits

$$P_C>P_I=P_1+P_2+\dots+P_N$$

where P_I is the sum of the individual output powers

[0238] P_C is the output of the combined circuit

[0239] P_1 is the output power from the first power harvesting technology

[0240] P_2 is the output power from the second power harvesting technology

[0241] P_N is the output power from the Nth power harvesting technology

[0242] N is the number of power harvesting technologies or circuits

[0243] The present invention pertains to a device **36** for receiving wireless power. The device **36** comprises a point of reception, wherein the point of reception is positionable in at least a first position **40** and a second position **42**.

[0244] The present invention pertains to a method for receiving wireless power. The method comprises the steps of positioning a point of reception in contact with a housing **46** to a first position **40**. There is the step of receiving wireless power at the point of reception and providing it to a power harvester **20** in the housing **46**. There is the step of converting the wireless power to usable DC with the power harvester **20**. There is the step of providing the usable DC to core components **48** in the housing **46**. There is the step of using the DC by the core components **48**. There is the step of repositioning the point of reception to a second position **42**. There is the step of receiving wireless power at the point of reception at the second position **42** and providing it to the power harvester **20**. There is the step of converting the wireless power received by the point of reception in the second position **42** to usable DC with the power harvester **20**. There is the step of providing the usable DC to the core component in the housing **46**. There is the step of using the DC by the core components **48**.

[0245] Another example of a product that was retrofitted to include RF energy harvesting circuitry was a cell phone. The cell phone is an example of recharging and augmenting a battery to supply power to a device. This system is shown in FIGS. **91-96**.

[0246] The cell phone is one example from a family of similar products, including, personal digital assistants, MP3 players, etc. Any of these devices may be configured to receive wireless power with or without communications data. The device includes a point of reception which receives the wireless power. For example, the point of reception may be an antenna. The point of reception is connected to the power harvester.

[0247] The point of reception is positionable in at least two positions: a first position and a second position. The first and second positions are designed such that in the first position, the cell phone is in normal operation and in the second position, the cell phone is capable of efficiently being charged/recharged.

[0248] The first and second positions may also be designed such that reception at each position may vary depending on the location of the device. Preferably, either the first position or the second position of the point of reception will provide better reception of the wireless power for a given location of the device (e.g., for charging and use, or for optimal charging).

[0249] Some or all of the positions may be applicable to a given embodiment of a cell phone or other device. In other words, various permutations of positions for the point of reception may be desirable and designed into the device. Additionally, the positions may be "infinite" in that the point of reception may be positioned anywhere desired as allowed by the particular design. Any mechanism for attaching the point of reception to the device is contemplated as is dictated by the particular application. For example, the mechanism may be a hinge (single pin, dual pin), a ball and socket joint, etc.

[0250] The device may include a stop mechanism configured to assist in positioning the point of reception in a desired position. The stop mechanism may be integral with the housing, the point of reception, or both. As an example, the point of reception may be an antenna that is contained in an antenna housing, for example, a plastic housing. The antenna housing may have a ridge that fits into one or more notches formed on the housing or device as the antenna housing moves with respect to the notched part of the housing or the device.

[0251] The point of reception may be designed into the device or connectable to the communications port of the device.

[0252] The device may include a communications antenna. The point of reception and the communications antenna may be co-located in an area of the device.

[0253] The device may have a single antenna configured to act as both a point of reception for wireless power and a communications antenna. A filter separates the received wireless power and the received communications data. A rectifier (i.e., the power harvester) converts the wireless power to a form usable by the device, such as DC.

[0254] The device may be configured such that the device automatically determines when it needs to be charged. At such time, the device sends a message to a wireless power transmitter indicating that the device needs to be charged by having the wireless power transmitter send wireless power to the device. The message may be sent using any means capable of indicating that the device needs to be charged, such as RF or infrared. The wireless power transmitter receives the message and begins to send wireless power to the device.

[0255] The wireless power transmitter may or may not stop sending power depending on the application. When the device is fully charged, the device may send a message to the wireless power transmitter to indicate that it no longer needs power. The wireless power transmitter may stop sending wireless power or continue to send a lower power level to supply operation current or to keep the battery or batteries charged while they are being drained by active, sleep, or leakage currents. Alternatively, the wireless power transmitter may send wireless power for a predetermined amount of time. If multiple devices are present, the wireless power transmitter may continue to send power even if one device has fully charged.

[0256] The wireless power transmitter may require periodic indications from the device that the device is still present in order to continue sending power. This would help to avoid sending power if the device is moved mid-way through charging, that is, if no device is present to receive the wireless power.

[0257] The device may indicate power requirements or battery size to set wireless power transmitter output power level. If multiple devices are present, the highest power level may be chosen.

[0258] If multiple wireless power transmitters are involved, the wireless power transmitters may communicate with each other to coordinate power transfer.

[0259] The device may send charging status information to the wireless power transmitter or other data device, such as a computer.

[0260] The device preferably includes a housing having a front, a back, a side, and an end. The point of reception is connected to the housing.

[0261] The point of reception may be pivotally connected to the housing, for example, at the end or the side of the housing. For example, referring to FIG. **91**, a cell phone may have an antenna that pivots from a first position substantially juxtaposed to the housing of the cell phone to a second position at an angle to the housing. In the second position, the antenna may be used to support the cell phone in an upright position, as shown.

[0262] Expanding on the example shown in FIG. **91** and referring to FIG. **92**, the antenna may pivot to a third position substantially extending parallel from the back of the housing. The antenna may further pivot to a fourth position substantially at a right angle to the front of the housing. The antenna may further rotate to a fifth position substantially juxtaposed to the front of the housing (for example, to protect a screen and other elements of the device).

[0263] For another example, referring to FIG. **93**, the antenna may pivot from a first position substantially juxtaposed to the side of the housing to a second position substantially at a right angle to the side of the housing. The antenna may further rotate to a third position substantially extending parallel to the side of the housing.

[0264] The point of reception may be slideably connected to the housing, for example, at the back, the side, or the front of the housing. For example, referring to FIG. **94**, the cell phone may have an antenna that slides from a first position substantially juxtaposed to the back of the housing to a second position substantially extending from the back of the housing. Similarly, the antenna may be slideably juxtaposed to the front of the housing (not shown).

[0265] For another example, referring to FIG. **95**, the antenna may slide from a first position substantially juxtaposed to the side of the housing to a second position substantially extending parallel from the side of the housing.

[0266] The point of reception may be retractably connected to the housing, for example, at the end of the housing. For example, referring to FIG. **96**, the point of reception may be co-located or integral with a communications antenna, where the antenna(s) is retracted into the housing in the first position and pulls out of the housing into the second position.

[0267] Filters are used to separate the incoming power and communications signals and to route the separated signals to the appropriate circuitry. A first filter may be designed to pass the frequency(ies) of the power signal while having a high impedance for the frequency(ies) of the communications signal. A second filter may be designed to pass the frequency (ies) of the communications signal while having a high impedance for the frequency(ies) of the power signal. The output of the first filter may be supplied to the power rectification circuitry that converts the power to a usable form, such as DC. The output of the power rectification circuitry may or may not be connected to charging circuitry. The charging circuitry monitors and/or regulates the voltage and/or current supplied to the battery to ensure proper charging.

[0268] The point of reception may be rotatably connected to the housing, for example, at the end or the side of the housing. For example, a cell phone may have an antenna that rotates from a first position substantially extending parallel to the back of the housing of the cell phone to a second position substantially extending parallel to the back of the housing, but where a face of the antenna is in a different position than the face when in the first position.

[0269] It should be noted that any of the previous embodiments of the cell phone may include an indicator to inform the user of the charging status. The indicator may also inform the user of the amount of wireless power being received. The indicator could then not only be used to position the device to achieve the desired charging rate, but also to position the antenna to achieve the desired charging rate. Examples of indicators include LEDs, LCDs, or other indicating components.

[0270] It should be noted that any of the previous embodiments of the cell phone may have the point of reception achieved by a user (for example, manually sliding the point of reception with respect to the housing) or automatically (for example, via spring loading).

[0271] A cell phone charger/recharger was designed to retrofit the SLVR cell phone from Motorola. The device was constructed as shown in FIGS. **91**(*a*) and (*b*). The back cover (housing) of the cell phone was removed and replaced with a specially designed cover (housing) that included a hinge at the top just below the lens of the camera portion of the phone.

[0272] The point of reception was designed to angle away from the cell phone using a pin hinge (FIG. **97**) in order to maximize the power transfer for the application. When angled away from the cell phone, the point of reception (implemented in this example as a patch antenna designed on Rogers 4003 material) acted as an antenna and as a support for the phone. The patch antenna was probe fed with the rectification circuitry being located near the middle of the antenna behind the ground plane. The patch antenna was designed to receive the maximum amount of energy when vertically polarized and the back of the cell phone point toward the source.

[0273] The rectification circuitry used was disclosed in U.S. patent application Ser. No. 11/584,983 filed Oct. 23, 2006, incorporated herein by reference. The output of the rectifier was connected to a charging circuit used to ensure that the battery contained within the cell phone was not over charged in terms of voltage or current.

[0274] The charging circuit was also connected to an indicator to show the user that the phone was charging. The indicator could also be used to show the charging status such as fully charged. The retrofitted cell phone used an LED as the indicator to show whether or not the cell phone was charging.

[0275] The output of the charging circuit was connected to the battery of the cell phone using a flexible printed circuit board (flex PCB), although a ribbon cable or other similar mechanism may be used. The flex PCB was thin enough to run under the back cover of the cell phone from the battery to a small notch where the flex PCB exited the cell phone and was connected to the charging circuit on the back of the antenna.

[0276] The antenna, rectifier, and charging circuit were encased in a plastic enclosure. The enclosure was connected to the hinge that also connected to the specially designed back cover. The hinge was designed to be resistive in order to allow the user to increase the angle between the phone and the antenna to a desired position without the need for a stopping mechanism, such as grooves.

[0277] A cell phone charger/recharger was also designed as described in the previous example, but using the design shown in FIG. **92**. The point of reception (implemented using a patch antenna designed on Rogers 4003 material) was designed to be positioned by the user on the back of the cell phone during normal cell phone use and located perpendicular to the front (face) of the cell phone for recharging as shown in FIG. **98**.

[0278] The patch antenna was probe fed with the rectification circuitry being located near the middle of the antenna behind the ground plane. The patch antenna was designed to receive the maximum amount of energy when vertically polarized and positioned perpendicular to the face of the phone with the top of the phone pointed toward the source.

[0279] The rectification circuitry used was disclosed in U.S. patent application Ser. No. 11/584,983 filed Oct. 23, 2006, incorporated by reference herein. The output of the rectifier was connected to a charging circuit used to ensure that the battery contained within the cell phone was not over charged in terms of voltage or current.

[0280] The charging circuit was also connected to an indicator to show the user that the phone was charging. The indicator could also be used to show the charging status such as fully charged. The retrofitted cell phone used an LED as the indicator to show whether or not the cell phone was charging.

[0281] The output of the charging circuit was connected to the battery of the cell phone using a flexible printed circuit board (flex PCB) although a ribbon cable or similar mechanism could be used. The flex PCB was thin enough to run under the back cover of the cell phone from the battery to a small notch where the flex PCB exited the cell phone and was connected the charging circuit on the back of the antenna.

[0282] The point of reception was designed to swing from the back of the cell phone to a position perpendicular to the face of the phone using two pin hinges located along the sides of the phone (shown in FIG. **99**) in order to maximize the power transfer for the application. If found to be advantageous, the electrical connection from the cell phone battery to the output of the charging circuit could be made through the hinges of the device. As an example, each pin hinge could be made with a metal pin where the right pin was connected to the positive connection of the battery and charging circuit and the left pin was connected to the negative connection of the battery and charging circuit.

[0283] It should be noted that in both of the previous examples, the wireless charger/recharger was designed to retrofit an existing cell phone. It is also possible to design the device into the cell phone.

[0284] As can be seen by the previous examples, RF energy harvesting can be used alone or in conjunction with alternative power sources **24** to power a wide range of devices. The addition of RF energy harvesting technology to the device allows for increased battery life, increased functionality, or the removal of the primary battery.

[0285] For purposes herein the following definition is applicable. A portable electronic device is defined to be less than about 25 pounds and preferably less than about 5 pounds in weight. It can be carried by one person either with or without some type of strap and preferably with only one arm or hand of the person. It has a device or circuitry that is powered by electricity.

[0286] Besides the various applications listed above, the RF energy harvesting can be used with any device requiring an antenna, although an antenna is necessarily needed in all embodiments, and includes radios and walkie talkies, besides cell phones, PDAs and MP3 players, to mention but a few of the many possible electronic devices.

[0287] Although the invention has been described in detail in the foregoing embodiments for the purpose of illustration, it is to be understood that such detail is solely for that purpose and that variations can be made therein by those skilled in the art without departing from the spirit and scope of the invention except as it may be described by the following claims.

1. A device for receiving wireless power, comprising:
- a point of reception, wherein the point of reception is positionable in at least a first position and a second position.

2. The device as described in claim **1**, wherein when the device is in a first location, when the point of reception is in the first position, the point of reception receives better reception than when in the second position.

3. The device as described in claim **2**, wherein when the device is in a second location, when the point of reception is in the second position, the point of reception receives better reception than when in the first position.

4. The device as described in claim **1**, wherein the point of reception is an antenna.

5. The device as described in claim **1**, further including a communications antenna.

6. The device as described in claim **5**, wherein the point of reception is co-located with the communications antenna.

7. The device as described in claim **1**, wherein the point of reception is connected to a communications port of the device.

8. The device as described in claim **1**, wherein the device includes a housing.

9. The device as described in claim **8**, wherein the point of reception is connected to the housing.

10. The device as described in claim **9**, wherein the point of reception is pivotally connected to the housing.

11. The device as described in claim **10**, wherein the point of reception is connected to an end of the housing.

12. The device as described in claim **10**, wherein the point of reception is connected to a side of the housing.

13. The device as described in claim **9**, wherein the point of reception is slideably connected to the housing.

14. The device as described in claim **13**, wherein the point of reception is connected to a back of the housing.

15. The device as described in claim **13**, wherein the point of reception is connected to a side of the housing.

16. The device as described in claim **9**, wherein the point of reception is retractably connected to the housing.

17. The device as described in claim **16**, wherein the point of reception is connected to an end of the housing.

18. The device as described in claim **9**, further including a stop mechanism configured to position the point of reception.

19. The device as described in claim **9**, wherein the point of reception is rotatably connected to the housing.

20. The device as described in claim **1** including a power harvester for converting wireless energy into usable DC, and core components in electrical communication with the power harvester to receive the DC to power the core components.

21. The device as described in claim **20**, further including an alternative power source connected to the core components to power the core components in conjunction with the power harvester.

22. The device as described in claim **20**, further including a power regulator and/or power storage circuit connected to the power harvester.

23. The device as described in claim **20**, further including a power storage charger connected to the power harvester.

24. The device as described in claim **20**, further including power storage connected to the power harvester.

25. The device as described in claim **20**, wherein the core components include a memory connected to an integrated circuit and to the power harvester to power memory.

26. The device as described in claim **20** wherein the core components includes a transmitter.

27. The device as described in claim **26** wherein the core components include a speaker.

28. The device as described in claim **27** wherein the core components include a receiver for receiving spoken words and converting the words into signals to be transmitted by the transmitter.

29. A device for receiving wireless power and communications data, comprising:
- an antenna configured to receive wireless power and communications data;
- a filter to separate the wireless power and the communications data; and
- a rectifier to convert the wireless power into a usable form.

30. A method for receiving wireless power comprising the steps of:
- positioning a point of reception in contact with a housing to a first position;
- receiving wireless power at the point of reception and providing it to a power harvester in the housing;
- converting the wireless power to usable DC with the power harvester;
- providing the usable DC to core components in the housing;
- using the DC by the core components;
- repositioning the point of reception to a second position;
- receiving wireless power at the point of reception at the second position and providing it to the power harvester;
- converting the wireless power received by the point of reception in the second position to usable DC with the power harvester;
- providing the usable DC to the core components in the housing; and
- using the DC by the core components.

* * * * *

US 20090174361A1

(19) **United States**

(12) **Patent Application Publication**

Duron et al.

(10) **Pub. No.: US 2009/0174361 A1**

(43) **Pub. Date: Jul. 9, 2009**

(54) **ENERGY HARVESTING IN RFID SYSTEMS**

(75) Inventors: **Mark Duron**, East Patchogue, NY (US); **Russell Calvarese**, Stony Brook, NY (US); **Robert Sandler**, Melville, NY (US); **Thomas Wulff**, North Patchogue, NY (US)

Correspondence Address:
MOTOROLA, INC.
1303 EAST ALGONQUIN ROAD, IL01/3RD
SCHAUMBURG, IL 60196

(73) Assignee: **Symbol Technologies, Inc.**, Holtsville, NY (US)

(21) Appl. No.: **11/971,760**

(22) Filed: **Jan. 9, 2008**

Publication Classification

(51) **Int. Cl.**
H02J 7/00 (2006.01)
H02M 7/217 (2006.01)

(52) **U.S. Cl.** **320/101**; 320/128; 320/114; 363/127

(57) **ABSTRACT**

Methods, systems, and apparatuses for providing power to devices that are part of RFID systems are described. Energy is harvested at portable/mobile devices, stored and conditioned to provide on-going power as needed for various circuits/components. The energy may be generated in a variety of ways, including using a vibratory energy harvesting device, a magnetic energy harvesting device, and an optical energy harvesting devices.

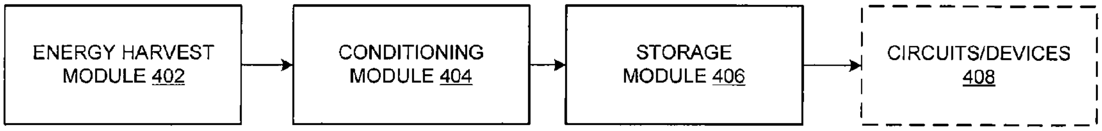

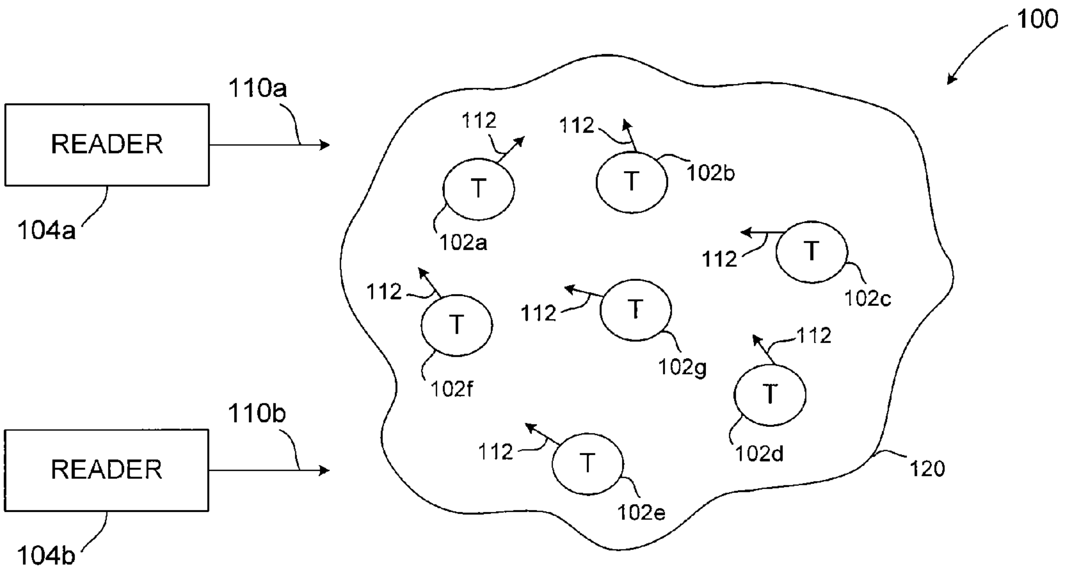
READER
READER
T
T
T
T
T
T
T

FIG. 1

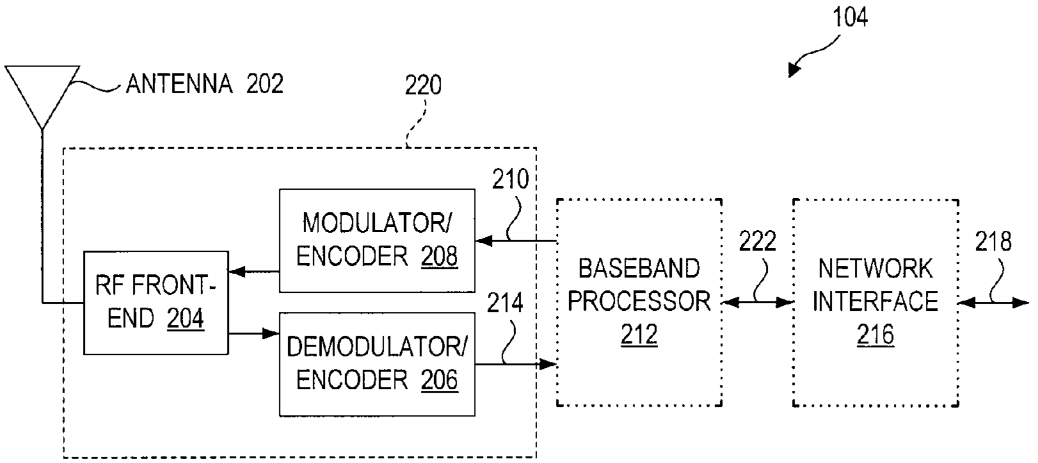
104
ANTENNA 202
220
210
MODULATOR/
ENCODER 208
RF FRONT-
END 204
214
DEMODULATOR/
ENCODER 206
BASEBAND
PROCESSOR
212
222
NETWORK
INTERFACE
216
218

FIG. 2

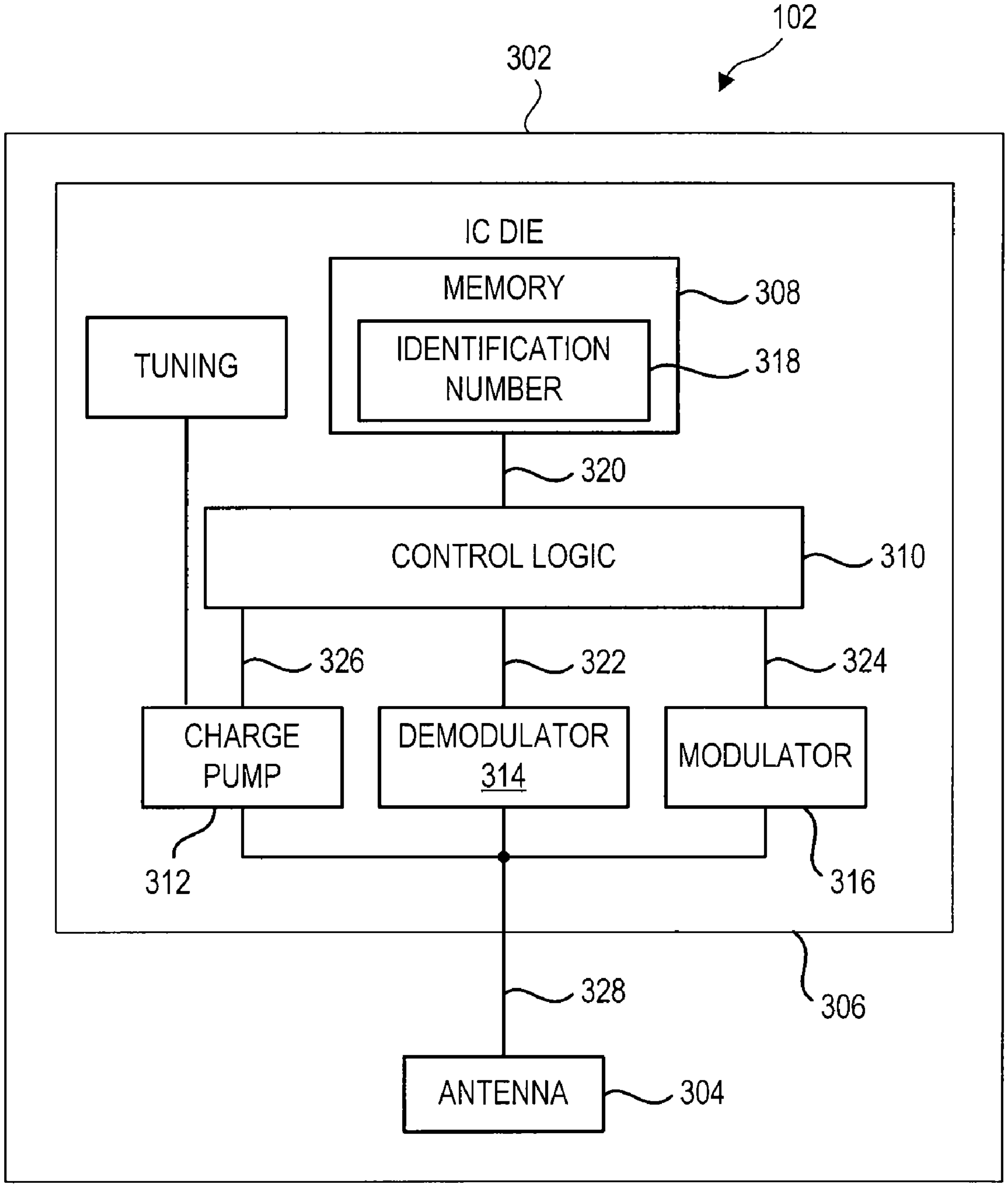
IC DIE
MEMORY
IDENTIFICATION NUMBER
TUNING
CONTROL LOGIC
CHARGE PUMP
DEMODULATOR 314
MODULATOR
ANTENNA

FIG. 3

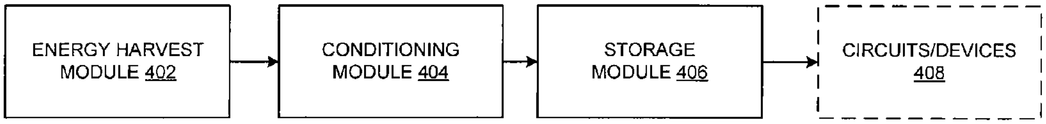
ENERGY HARVEST MODULE 402
CONDITIONING MODULE 404
STORAGE MODULE 406
CIRCUITS/DEVICES 408

FIG. 4

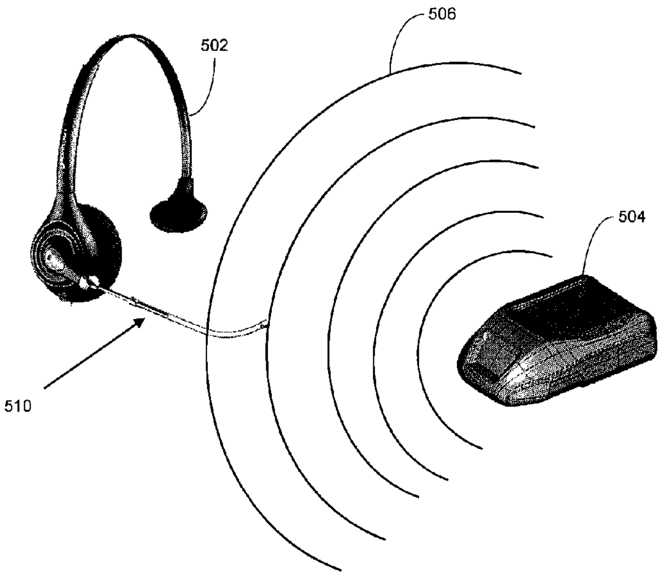

FIG. 5

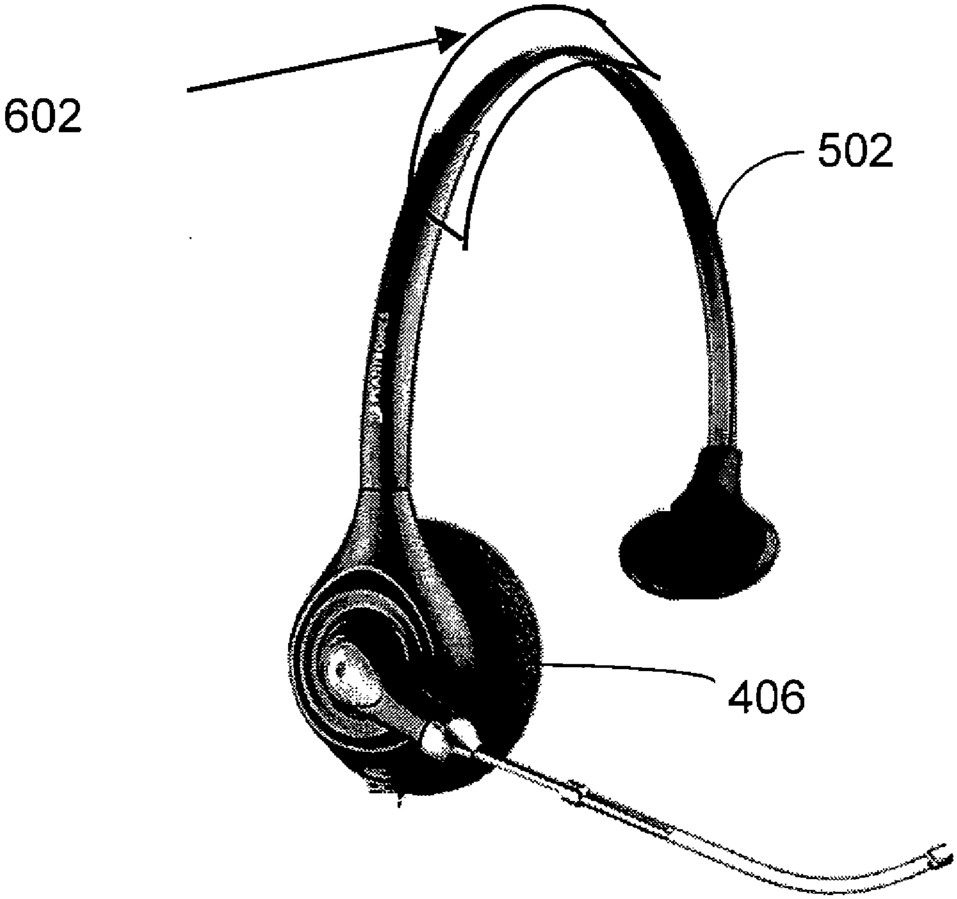

FIG. 6

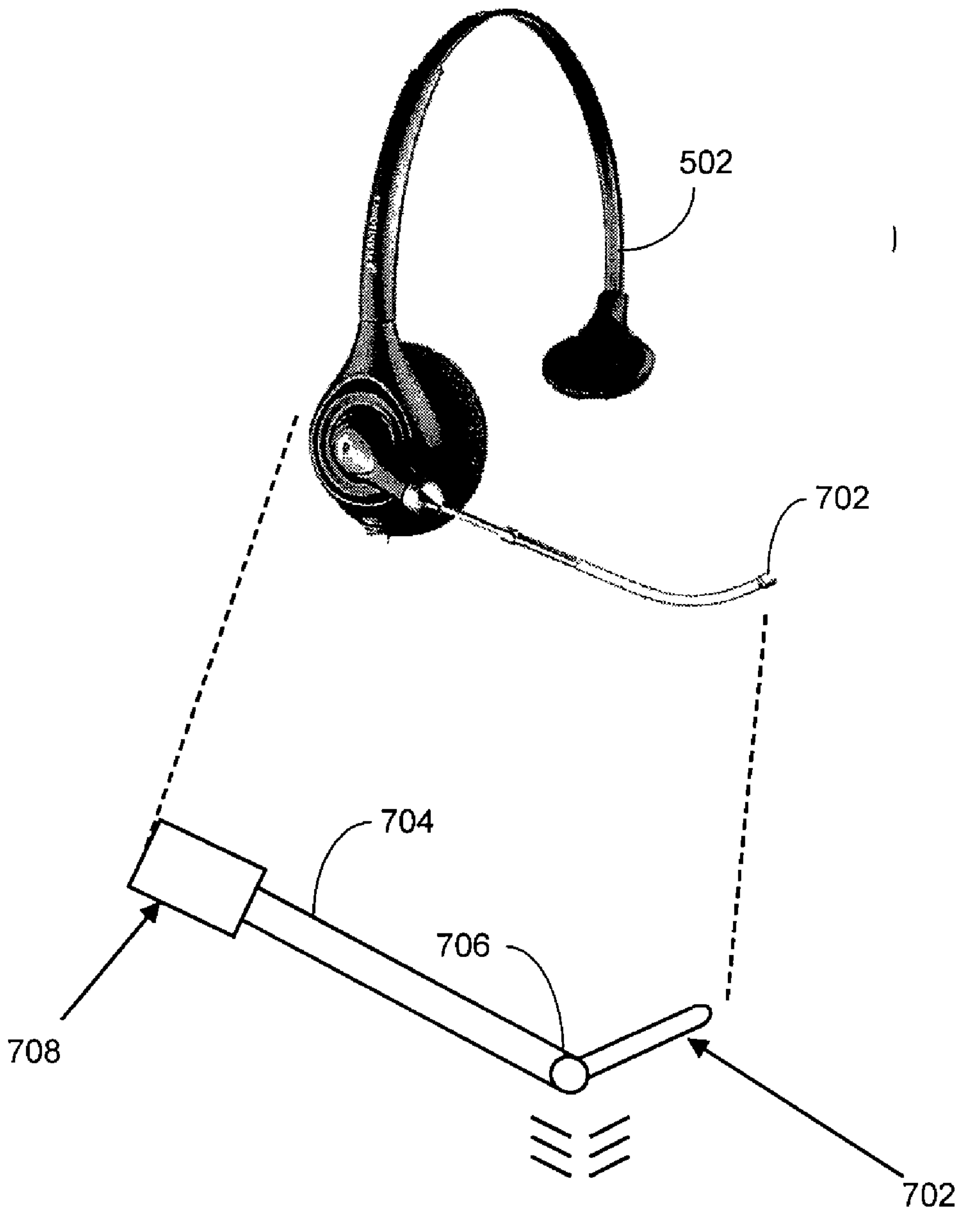
502
702
704
706
708
702

FIG. 7

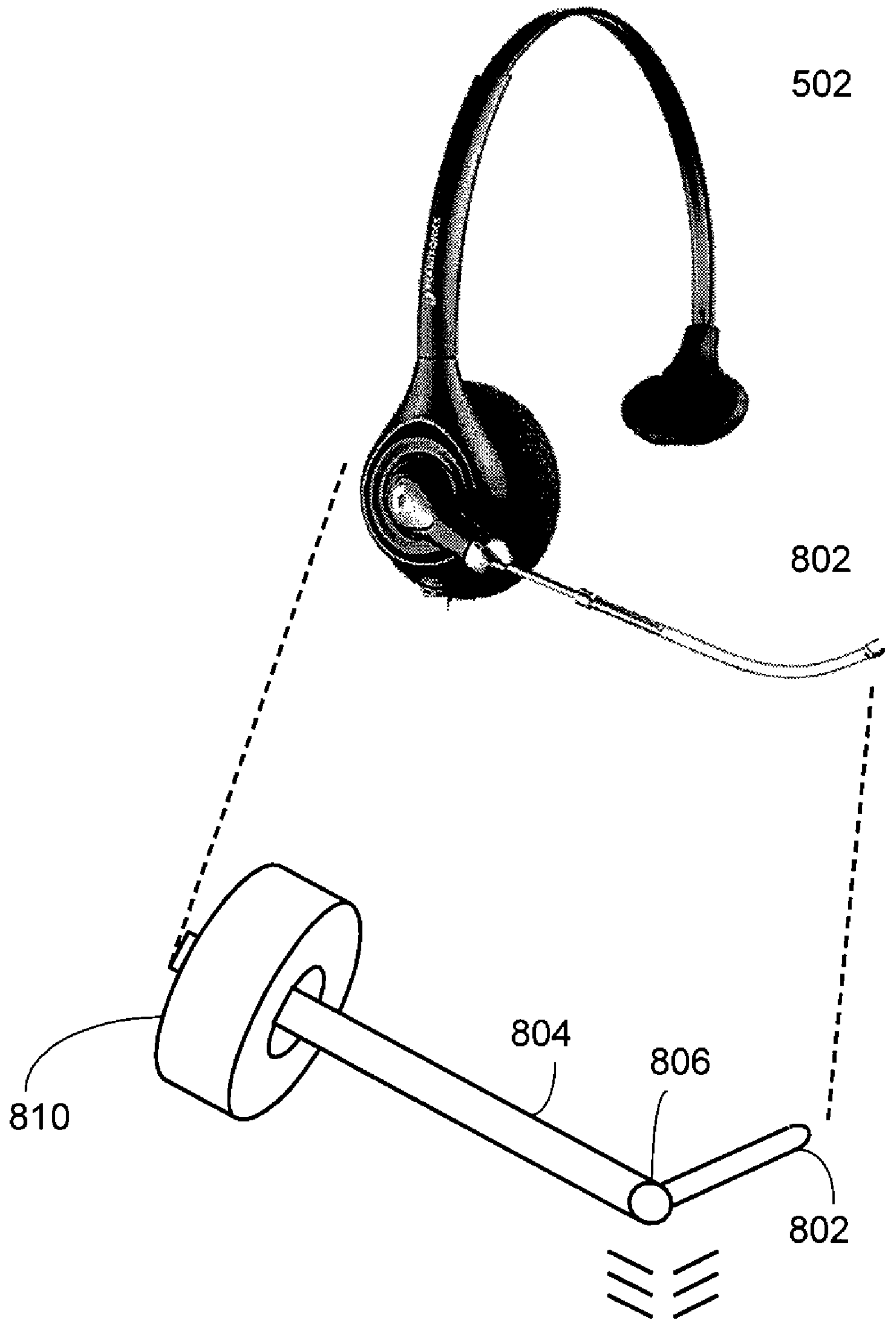
502
802
804
806
810
802

FIG. 8

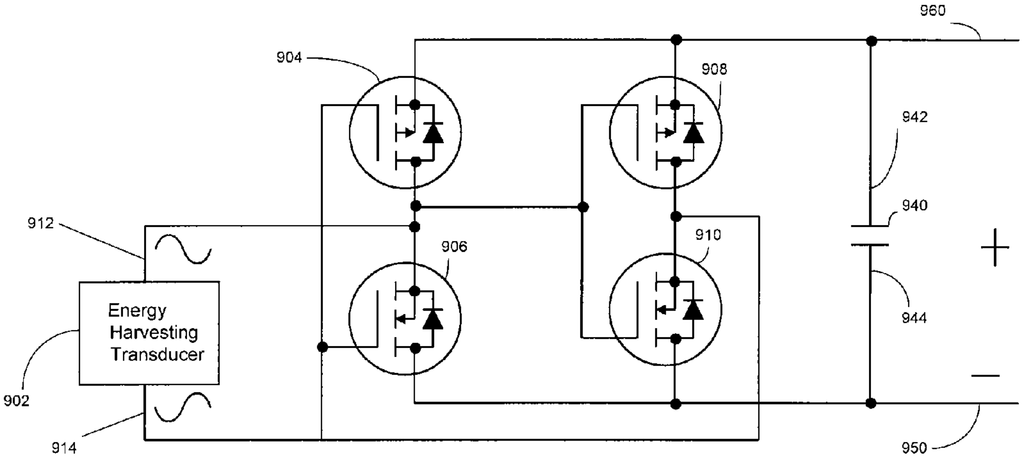
+
Energy Harvesting Transducer
—

FIG. 9

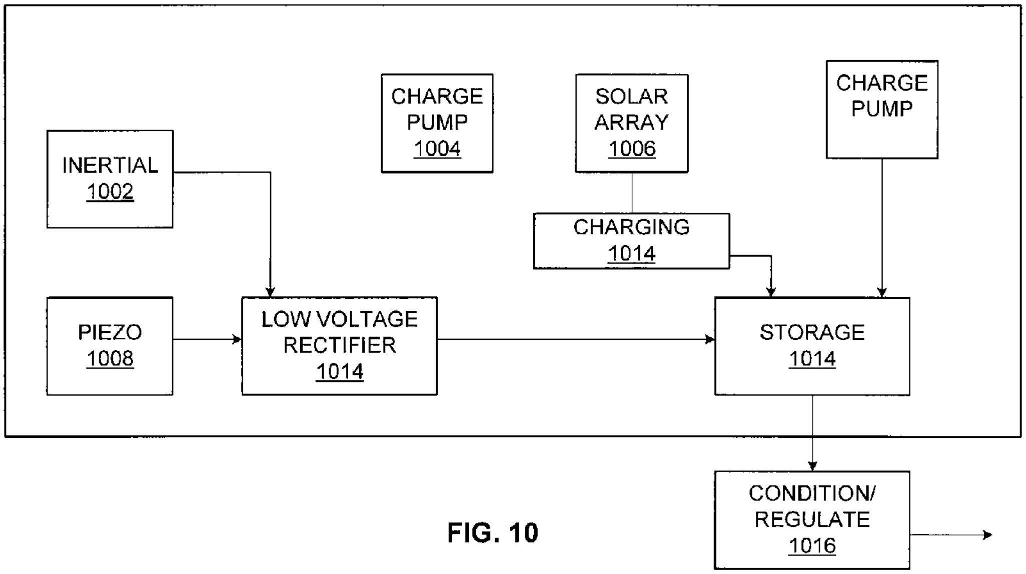
CHARGE PUMP 1004
SOLAR ARRAY 1006
CHARGE PUMP
INERTIAL 1002
CHARGING 1014
PIEZO 1008
LOW VOLTAGE RECTIFIER 1014
STORAGE 1014
CONDITION/ REGULATE 1016

FIG. 10

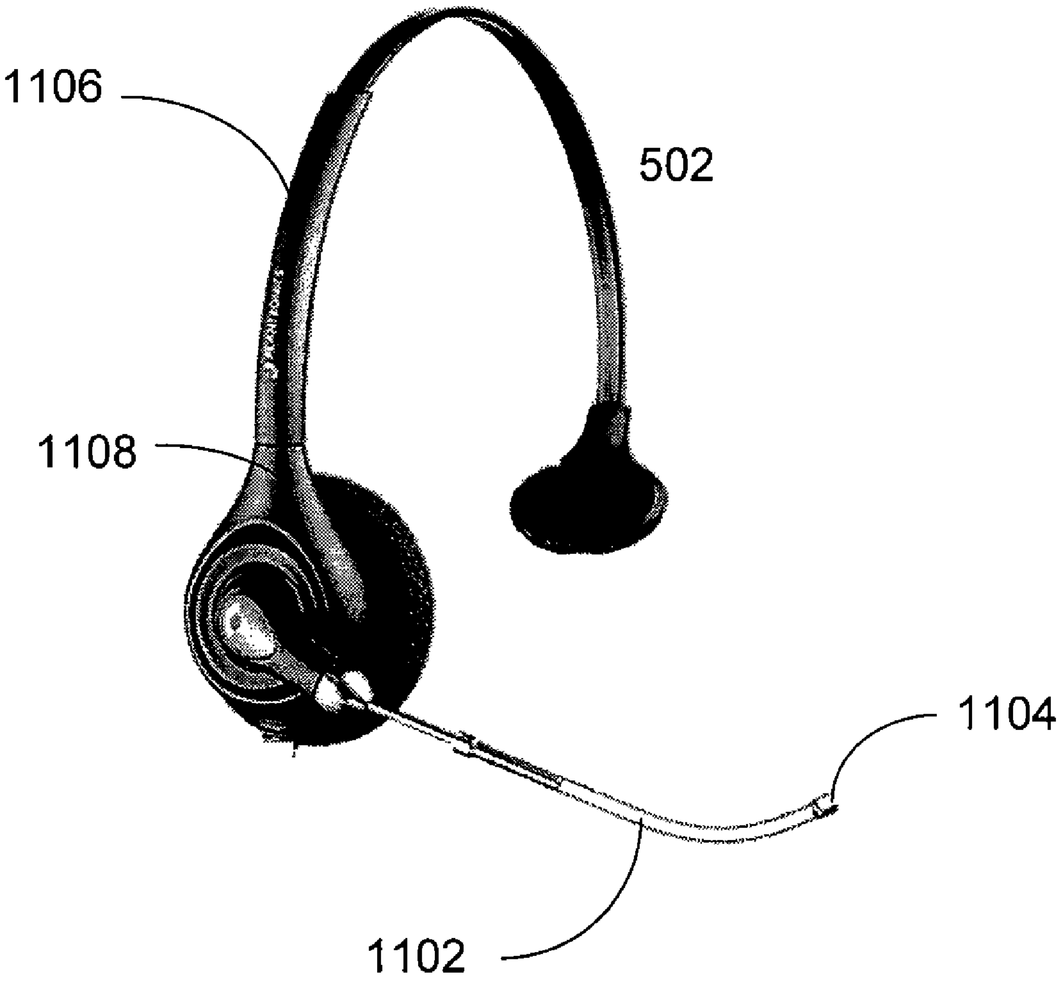
1106
502
1108
1104
1102

FIG. 11

ENERGY HARVESTING IN RFID SYSTEMS

BACKGROUND OF THE INVENTION

[0001] 1. Field of the Invention

[0002] The invention relates to radio frequency identification (RFID) systems, and in particular, to harvesting energy in an RFID systems environment.

[0003] 2. Background Art

[0004] Radio frequency identification (RFID) tags are electronic devices that may be affixed to items whose presence is to be detected and/or monitored. The presence of an RFID tag, and therefore the presence of the item to which the tag is affixed, may be checked and monitored wirelessly by devices known as "readers." Readers typically have one or more antennas transmitting radio frequency signals to which tags respond. Because the reader "interrogates" RFID tags, and receives signals back from the tags in response to the interrogation, the reader is sometimes termed a "reader interrogator" or simply "interrogator."

[0005] With the maturation of RFID technology, efficient communication between tags and interrogators has become a key enabler in supply chain management, especially in manufacturing, shipping, and retail industries, as well as in building security installations, healthcare facilities, libraries, airports, warehouses etc.

[0006] With the increased use of peripheral devices in RFID systems, it is becoming increasingly important to devise new ways to provide power to such devices. For example, a mobile computer with an integrated RFID interrogator may be used by an operator wearing a headset to communicate with the computer or a central communications point. A reader interrogator may be remote from a conventional power source, etc.

[0007] Thus, what is needed are improved ways of providing power to various devices included in RFID systems.

BRIEF SUMMARY OF THE INVENTION

[0008] The invention provides various techniques for harvesting power in the context of RFID systems. It also provides techniques for accumulating power harvested by various techniques and utilizing that power to operate various devices.

[0009] In exemplary embodiments, power is harvested by a device worn by an operator. One example of such a device is a headset. Power is harvested using various techniques/structures including: a solar array, an inertial harvesting device, a piezo energy harvester, and a charge pump.

[0010] There is provided a low voltage rectifier circuit for use in harvesting power from a variety of sources.

[0011] Power storage devices are provided for cumulating power derived from one or multiple sources.

[0012] In various embodiments, energy may be generated by a variety of energy harvesting devices, including a vibratory energy harvesting device, electromagnetic energy harvesting device, a magnetic energy harvesting device, a photo-optical energy harvesting device, a heat energy harvesting device, or a mechanical energy harvesting device.

[0013] In an embodiment, a vibratory energy harvesting device generates the energy from vibration of the mobile structure. For example, the vibratory energy harvesting device comprises a piezoelectric transducer. In an example implementation, a moment arm mounts the piezoelectric transducer. A capacitor is coupled to the energy storage device. Vibration of the arm causes the arm to deflect the piezoelectric transducer. The piezoelectric transducer generates a current due to the deflection. The current charges the capacitor.

[0014] In another example aspect, a magnetic energy harvesting device generates the energy from magnetism related to the movement of the mobile structure.

[0015] In a further example aspect, the magnetic energy harvesting device includes a magnet and a coil. The energy is generated by movement of the magnet with respect to the coil.

[0016] In another example aspect, an photo-optical energy harvesting device includes an photo-optical-to-electrical transducer that converts light received at the mobile structure into energy.

[0017] In another example aspect, a heat energy harvesting device converts heat generated by operation of the mobile structure into energy.

[0018] In another example aspect, a mechanical energy harvesting device converts friction due to operation of the mobile structure into energy. In another example aspect, the mechanical energy harvesting device uses a rotational wheel mechanism that rotates due to movement of the mobile structure to generate energy.

[0019] These and other objects, advantages and features will become readily apparent in view of the following detailed description of the invention. Note that the Summary and Abstract sections may set forth one or more, but not all exemplary embodiments of the present invention as contemplated by the inventor(s).

BRIEF DESCRIPTION OF THE DRAWINGS/FIGURES

[0020] The accompanying drawings, which are incorporated herein and form a part of the specification, illustrate the invention and, together with the description, further serve to explain the principles of the invention and to enable a person skilled in the pertinent art to make and use the invention.

[0021] In the drawings, like reference numbers indicate identical or functionally similar elements.

[0022] Additionally, references numbers which are the same, but vary by virtue of an appended letter of the alphabet (for example, **412**, **412**R, **412**P, **412**S) or an appended letter and number (for example, **412**, **412**S**1**, **412**S**2**) indicate elements which may be substantially the same or similar, but represent variations or modifications of the basic element. In some cases, the reference number without the appended letter or without the appended letter and number (for example, **412**) may indicate a generic form of the element, while reference numbers with an appended letter or an appended letter and number (for example, **412**S, **412**S**1**, **412**S**2**, **412**P) may indicate a more particular or modified form of the element.

[0023] Generally, the leftmost digit(s) of a reference number identifies the drawing in which the reference number first appears. For example, an element labeled **412** typically indicates that the element first appeared in FIG. **4**.

[0024] FIG. **1** shows an environment where RFID readers (interrogators) communicate with an exemplary population of RFID tags.

[0025] FIG. **2** is a block diagram of receiver and transmitter portions of an RFID reader.

[0026] FIG. **3** is a block diagram of an exemplary radio frequency identification (RFID) tag.

[0027] FIG. **4** is a general block diagram illustrating a system for harvesting energy and utilizing it to power circuits/devices.

[0028] FIG. **5** is a schematic diagram of an embodiment of the invention illustrating harvesting energy from a signal transmitted by an RFID interrogator.

[0029] FIG. **6** is a schematic diagram of an embodiment of the invention illustrating harvesting energy from a photo-optical array.

[0030] FIG. **7** is a schematic diagram of an embodiment of the invention illustrating harvesting mechanical energy using a piezo-electric device.

[0031] FIG. **8** is a schematic diagram of an embodiment of the invention illustrating harvesting mechanical energy using a magnetic induction of an inertial device.

[0032] FIG. **9** is a schematic diagram of a low voltage rectifier circuit suitable for use with various energy harvesting transducer/devices.

[0033] FIG. **10** is a general block diagram illustrating how power can be harvested from two or more devices and combined.

[0034] FIG. **11** is a schematic diagram illustrating how energy harvesting components can be placed in a worn headset.

[0035] The present invention will now be described with reference to the accompanying drawings. In the drawings, like reference numbers indicate identical or functionally similar elements. Additionally, the left-most digit(s) of a reference number identifies the drawing in which the reference number first appears.

DETAILED DESCRIPTION OF THE INVENTION

Introduction

[0036] Power arrangements for RFID system components are described herein.

[0037] Supplying power to RFID system devices located in real world installations can be difficult, particularly for mobile devices. Providing cabled power to a user of an RFID system "in the field" may be problematic. The power cable may be damaged by wear associated with the movement and actuation of the structure A battery may be mounted to the mobile structure to avoid the need for a power cable. However, RFID devices may require a large amount of power to carry out their functions, and thus require regular recharging/changing of batteries.

[0038] The invention provides various techniques for harvesting power in the context of RFID systems. It also provides techniques for cumulating power harvested by various techniques and utilizing that power to operate various devices.

[0039] In exemplary embodiments, power is harvested by a headset worn by an operator. Power is harvested using various techniques/structures including: RF signal transmitted by an RFID interrogator, solar array, inertial harvesting device, and piezo energy harvester.

[0040] There is provided a low voltage rectifier circuit for use in harvesting power from a variety of sources.

[0041] Power storage devices are provided for cumulating power derived from one or multiple sources.

[0042] In aspects, the energy may be generated by a variety of energy harvesting devices, including a vibratory energy harvesting device, a magnetic energy harvesting device, an optical energy harvesting device, a heat energy harvesting device, or a mechanical energy harvesting device.

[0043] In an example aspect, a vibratory energy harvesting device generates the energy from vibration of the mobile structure during operation of the mobile structure.

[0044] In a further example aspect, the vibratory energy harvesting device comprises a piezoelectric transducer. In an example implementation, a moment arm mounts the piezoelectric transducer. A capacitor is coupled to the energy storage device. Vibration of the arm causes the arm to deflect the piezoelectric transducer. The piezoelectric transducer generates a current due to the deflection. The current charges the capacitor.

[0045] In another example aspect, a magnetic energy harvesting device generates the energy from magnetism related to the movement of the mobile structure.

[0046] In a further example aspect, the magnetic energy harvesting device includes a magnet and a coil. The energy is generated by movement of the magnet with respect to the coil.

[0047] In another example aspect, an optical energy harvesting device includes an optical-to-electrical transducer that converts light received at the mobile structure into energy.

[0048] In another example aspect, a heat energy harvesting device converts heat generated by operation of the mobile structure into energy.

[0049] In another example aspect, a mechanical energy harvesting device converts friction due to operation of the mobile structure into energy. In another example aspect, the mechanical energy harvesting device uses a rotational wheel mechanism that rotates due to movement of the mobile structure to generate energy.

[0050] Embodiments of the present invention overcome problems with powering readers present in conventional systems. For example, according to embodiments, energy is generated on the mobile structure on which the reader is disposed, such as in the form of electrical energy. The energy is stored on the mobile structure, and used to power the reader. In this manner, the need for replacement of batteries, battery charging cycle times, power cables, and/or extremely large batteries is reduced or eliminated.

[0051] The present specification discloses one or more embodiments that incorporate the features of the invention. The disclosed embodiment(s) merely exemplify the invention. The scope of the invention is not limited to the disclosed embodiment(s). The invention is defined by the claims appended hereto.

[0052] References in the specification to "one embodiment," "an embodiment," "an example embodiment," etc., indicate that the embodiment described may include a particular feature, structure, or characteristic, but every embodiment may not necessarily include the particular feature, structure, or characteristic. Moreover, such phrases are not necessarily referring to the same embodiment. Further, when a particular feature, structure, or characteristic is described in connection with an embodiment, it is submitted that it is within the knowledge of one skilled in the art to effect such feature, structure, or characteristic in connection with other embodiments whether or not explicitly described.

[0053] Furthermore, it should be understood that spatial descriptions (e.g., "above," "below," "up," "down," "top," "bottom," "vertical," "horizontal," etc.) used herein are for purposes of illustration only, and that practical implementations of the structures described herein can be spatially arranged in any orientation or manner.

[0054] In the following description, numerous specific details are set forth in order to provide a thorough understanding of the invention. However, it will be apparent to those skilled in the art that the invention, including structures, systems, and methods, may be practiced without these specific details. The description and representation herein are the common means used by those experienced or skilled in the art to most effectively convey the substance of their work to others skilled in the art. In other instances, well-known methods, procedures, components, and circuitry have not been described in detail to avoid unnecessarily obscuring aspects of the invention.

[0055] References in the specification to "one embodiment," "an embodiment," "an example embodiment," etc., indicate that the embodiment described may include a particular feature, structure, or characteristic, but every embodiment may not necessarily include the particular feature, structure, or characteristic. Moreover, such phrases are not necessarily referring to the same embodiment. Further, when a particular feature, structure, or characteristic is described in connection with an embodiment, it is submitted that it is within the knowledge of one skilled in the art to effect such feature, structure, or characteristic in connection with other embodiments whether or not explicitly described.

Exemplary Operating Environment

[0056] Before describing embodiments of the invention in detail, it is helpful to describe an example RFID communications environment in which the inventions may be implemented. FIG. **1** illustrates an environment **100** where RFID tag readers **104** (readers **104***a* and **104***b* shown in FIG. **1**) communicate with an exemplary population **120** of RFID tags **102**. As shown in FIG. **1**, the population **120** of tags includes seven tags **102***a*-**102***g*. A population **120** may include any number of tags **102**.

[0057] Environment **100** includes any number of one or more readers **104**. For example, environment **100** includes a first reader **104***a* and a second reader **104***b*. Readers **104***a* and/or **104***b* may be requested by an external application to address the population of tags **120**. Alternatively, reader **104***a* and/or reader **104***b* may have internal logic that initiates communication, or may have a trigger mechanism that an operator of a reader **104** uses to initiate communication. Readers **104***a* and **104***b* may also communicate with each other in a reader network (see FIG. **2**).

[0058] As shown in FIG. **1**, reader **104***a* "reads" tags **120** by transmitting an interrogation signal **110***a* to the population of tags **120**. Interrogation signals may have signal carrier frequencies or may comprise a plurality of signals transmitted in a frequency hopping arrangement. Readers **104***a* and **104***b* typically operate in one or more of the frequency bands allotted for this type of RF communication. For example, the Federal Communication Commission (FCC) defined frequency bands of 902-928 MHz and 2400-2483.5 MHz for certain RFID applications.

[0059] Tag population **120** may include tags **102** of various types, such as, for example, various classes of tags as enumerated above. Thus, in response to interrogation signals, the various tags **102** may transmit one or more response signals **112** to an interrogating reader **104**. Some of the tags, for example, may respond by alternatively reflecting and absorbing portions of signal **104** according to a time-based pattern or frequency. This technique for alternatively absorbing and reflecting signal **104** is referred to herein as backscatter modulation. Typically, such backscatter modulation may include one or more alpha-numeric characters that uniquely identify a particular tag. Readers **104***a* and **104***b* receive and obtain data from response signals **112**, such as an identification number of the responding tag **102**. In the embodiments described herein, a reader may be capable of communicating with tags **102** according to various suitable communication protocols, including Class 0, Class 1, EPC Gen 2, other binary traversal protocols and slotted aloha protocols, and any other protocols mentioned elsewhere herein, and future communication protocols. Additionally, tag population **120** may include one or more tags having the packed object format described herein and/or one or more tags not using the packed object format (e.g., standard ISO tags).

[0060] FIG. **2** shows a block diagram of an example RFID reader **104**. Reader **104** includes one or more antennas **202**, a receiver and transmitter portion **220** (also referred to as transceiver **220**), a baseband processor **212**, and a network interface **216**. These components of reader **104** may include software, hardware, and/or firmware, or any combination thereof, for performing their functions.

[0061] Baseband processor **212** and network interface **216** are optionally present in reader **104**. Baseband processor **212** may be present in reader **104**, or may be located remote from reader **104**. For example, in an embodiment, network interface **216** may be present in reader **104**, to communicate between transceiver portion **220** and a remote server that includes baseband processor **212**. When baseband processor **212** is present in reader **104**, network interface **216** may be optionally present to communicate between baseband processor **212** and a remote server. In another embodiment, network interface **216** is not present in reader **104**.

[0062] In an embodiment, reader **104** includes network interface **216** to interface reader **104** with a communications network **218**. As shown in FIG. **2**, baseband processor **212** and network interface **216** communicate with each other via a communication link **222**. Network interface **216** is used to provide an interrogation request **210** to transceiver portion **220** (optionally through baseband processor **212**), which may be received from a remote server coupled to communications network **218**. Baseband processor **212** optionally processes the data of interrogation request **210** prior to being sent to transceiver portion **220**. Transceiver **220** transmits the interrogation request via antenna **202**.

[0063] Reader **104** has at least one antenna **202** for communicating with tags **102** and/or other readers **104**. Antenna (s) **202** may be any type of reader antenna known to persons skilled in the relevant art(s), including for example and without limitation, a vertical, dipole, loop, Yagi-Uda, slot, and patch antenna type.

[0064] Transceiver **220** receives a tag response via antenna **202**. Transceiver **220** outputs a decoded data signal **214** generated from the tag response. Network interface **216** is used to transmit decoded data signal **214** received from transceiver portion **220** (optionally through baseband processor **212**) to a remote server coupled to communications network **218**. Baseband processor **212** optionally processes the data of decoded data signal **214** prior to being sent over communications network **218**.

[0065] In embodiments, network interface **216** enables a wired and/or wireless connection with communications network **218**. For example, network interface **216** may enable a wireless local area network (WLAN) link (including a IEEE 802.11 WLAN standard link), a BLUETOOTH link, and/or other types of wireless communication links. Communications network **218** may be a local area network (LAN), a wide area network (WAN) (e.g., the Internet), and/or a personal area network (PAN).

[0066] In embodiments, a variety of mechanisms may be used to initiate an interrogation request by reader **104**. For example, an interrogation request may be initiated by a remote computer system/server that communicates with reader **104** over communications network **218**. Alternatively, reader **104** may include a finger-trigger mechanism, a keyboard, a graphical user interface (GUI), and/or a voice activated mechanism with which a user of reader **104** may interact to initiate an interrogation by reader **104**.

[0067] In the example of FIG. **2**, transceiver portion **220** includes a RF front-end **204**, a demodulator/decoder **206**, and a modulator/encoder **208**. These components of transceiver **220** may include software, hardware, and/or firmware, or any combination thereof, for performing their functions. Example description of these components is provided as follows.

[0068] Modulator/encoder **208** receives interrogation request **210**, and is coupled to an input of RF front-end **204**. Modulator/encoder **208** encodes interrogation request **210** into a signal format, such as, for example, one of pulse-interval encoding (PIE), FM0, or Miller encoding formats, modulates the encoded signal, and outputs the modulated encoded interrogation signal to RF front-end **204**.

[0069] RF front-end **204** may include one or more antenna matching elements, amplifiers, filters, an echo-cancellation unit, a down-converter, and/or an up-converter. RF front-end **204** receives a modulated encoded interrogation signal from modulator/encoder **208**, up-converts (if necessary) the interrogation signal, and transmits the interrogation signal to antenna **202** to be radiated. Furthermore, RF front-end **204** receives a tag response signal through antenna **202** and down-converts (if necessary) the response signal to a frequency range amenable to further signal processing.

[0070] Demodulator/decoder **206** is coupled to an output of RF front-end **204**, receiving a modulated tag response signal from RF front-end **204**. In an EPC Gen 2 protocol environment, for example, the received modulated tag response signal may have been modulated according to amplitude shift keying (ASK) or phase shift keying (PSK) modulation techniques. Demodulator/decoder **206** demodulates the tag response signal. For example, the tag response signal may include backscattered data formatted according to FM0 or Miller encoding formats in an EPC Gen 2 embodiment. Demodulator/decoder **206** outputs decoded data signal **214**.

[0071] The configuration of transceiver **220** shown in FIG. **2** is provided for purposes of illustration, and is not intended to be limiting. Transceiver **220** may be configured in numerous ways to modulate, transmit, receive, and demodulate RFID communication signals, as would be known to persons skilled in the relevant art(s).

[0072] The invention described herein is applicable to any type of RFID tag, with suitable additional features, as described in further detail below in conjunction with FIG. **4** and beyond. FIG. **3** is a schematic block diagram of an example radio frequency identification (RFID) tag **102** as already known to those practiced in the art. Tag **102** includes a substrate **302**, an antenna **304**, and an integrated circuit (IC) **306**. Antenna **304** is formed on a surface of substrate **302**. Antenna **304** may include any number of one, two, or more separate antennas of any suitable antenna type, including for example dipole, loop, slot, and patch. IC **306** includes one or more integrated circuit chips/dies, and can include other electronic circuitry. IC **306** is attached to substrate **302**, and is coupled to antenna **304**. IC **306** may be attached to substrate **302** in a recessed and/or non-recessed location.

[0073] IC **306** controls operation of tag **102**, and transmits signals to, and receives signals from RFID readers using antenna **304**. In the example of FIG. **3**, IC **306** includes a memory **308**, a control logic **310**, a charge pump **312**, a demodulator **314**, and a modulator **316**. Inputs of charge pump **312**, and demodulator **314**, and an output of modulator **316** are coupled to antenna **304** by antenna signal **328**.

[0074] Demodulator **314** demodulates a radio frequency communication signal (e.g., interrogation signal **110**) on antenna signal **328** received from a reader by antenna **304**. Control logic **310** receives demodulated data of the radio frequency communication signal from demodulator **314** on an input signal **322**. Control logic **310** controls the operation of RFID tag **102**, based on internal logic, the information received from demodulator **314**, and the contents of memory **308**. For example, control logic **310** accesses memory **308** via a bus **320** to determine whether tag **102** is to transmit a logical "1" or a logical "0" (of identification number **318**) in response to a reader interrogation. Control logic **310** outputs data to be transmitted to a reader (e.g., response signal **112**) onto an output signal **324**. Control logic **310** may include software, firmware, and/or hardware, or any combination thereof. For example, control logic **310** may include digital circuitry, such as logic gates, and may be configured as a state machine in an embodiment.

[0075] Modulator **316** is coupled to antenna **304** by antenna signal **328**, and receives output signal **324** from control logic **310**. Modulator **316** modulates data of output signal **324** (e.g., one or more bits of identification number **318**) onto a radio frequency signal (e.g., a carrier signal transmitted by reader **104**) received via antenna **304**. The modulated radio frequency signal is response signal **112** (see FIG. **1**), which is received by reader **104**. In one example embodiment, modulator **316** includes a switch, such as a single pole, single throw (SPST) switch. The switch is configured in such a manner as to change the return loss of antenna **304**. The return loss may be changed in any of a variety of ways. For example, the RF voltage at antenna **304** when the switch is in an "on" state may be set lower than the RF voltage at antenna **304** when the switch is in an "off" state by a predetermined percentage (e.g., 30 percent). This may be accomplished by any of a variety of methods known to persons skilled in the relevant art(s).

[0076] Charge pump **312** (or other type of power generation module) is coupled to antenna **304** by antenna signal **328**. Charge pump **312** receives a radio frequency communication signal (e.g., a carrier signal transmitted by reader **104**) from antenna **304**, and generates a direct current (DC) voltage level that is output on tag power signal **326**. Tag power signal **326** powers circuits of IC die **306**, including control logic **320**.

[0077] Charge pump **312** rectifies a portion of the power of the radio frequency communication signal of antenna signal **328** to create a voltage power. Charge pump **312** increases the voltage level of the rectified power to a level sufficient to power circuits of IC die **306**. Charge pump **312** may also include a regulator to stabilize the voltage of tag power signal **326**. Charge pump **312** may be configured in any suitable way known to persons skilled in the relevant art(s). For description of an example charge pump applicable to tag **102**, refer to U.S. Pat. No. 6,734,797, titled "Identification tag Utilizing Charge Pumps for Voltage Supply Generation and Data

Recovery," which is incorporated by reference herein in its entirety. Alternative circuits for generating power in a tag, as would be known to persons skilled in the relevant art(s), may be present. Further description of charge pump **312** is provided below.

[0078] It will be recognized by persons skilled in the relevant art(s) that tag **102** may include any number of modulators, demodulators, charge pumps, and antennas. Tag **102** may additionally include further elements, including an impedance matching network and/or other circuitry. Furthermore, although tag **102** is shown in FIG. **3** as a passive tag, tag **102** may alternatively be an active tag (e.g., powered by a battery, not shown).

[0079] Memory **308** is typically a non-volatile memory, but can alternatively be a volatile memory, such as a DRAM. Memory **308** stores data, including an identification number **318**. In a Gen-2 tag, tag memory **308** may be logically separated into four memory banks.

[0080] As further described below, according to embodiments of the present invention, energy is generated on the mobile structure on which the reader is disposed. The generated energy is stored on the mobile structure, and used to power the reader. In this manner, the need for power cables and/or extremely large batteries with regard to mobile structures is reduced or eliminated.

[0081] Embodiments of the present invention are described in further detail below. Such embodiments may be implemented in environment **100** shown in FIG. **1**, with reader **104** shown in FIG. **2**, and/or in alternative environments and RFID devices.

Energy Harvesting—General Concept

[0082] Energy harvesting systems are described herein for providing energy to various RFID system devices. Energy harvesting device generate energy in various ways, some examples of which are described in specific embodiments herein. The energy harvesting devices may be associated with a mobile/portable structure in various ways. For example, an energy harvesting device may be disposed directly on mobile structure, integrated into a mobile/portable structure or otherwise attached to it.

[0083] Various embodiments are described that feature a headset worn by an operator. Although the inventions described herein are particularly useful for powering headsets, they are not limited to headsets. Harvested energy can be used to power other devices, such as mobile/portable computers, communication devices, etc.

[0084] Embodiments for the energy harvesting systems can be implemented wherever convertible energy may be available. Systems can be implemented in a commercial or industrial environment, such as in a warehouse, a factory, a business, or store, and in a military or other non-commercial environment. For example, an RFID interrogator may be attached to a stationary structure or to a mobile structure. The energy generating systems enable deployment of devices that form RFID systems or are attached to RFID systems without the need for power cables and with potentially less space required for batteries.

[0085] FIG. **4** is a general block diagram illustrating a system for harvesting energy and utilizing it to power circuits/devices. Main system components include an energy harvesting module **402**, an energy conditioning circuit **404**, and an energy storage device **406**. Energy accumulated and stored in storage device **406** can be used to power various circuits/devices **408**, as needed. The energy storage device **406** is typically a rechargeable battery, but may alternatively be another type of battery or storage device otherwise known or future developed that can receive and store energy. Example materials/battery types for a rechargeable battery include Lithium (e.g., Li-ion, Li-polymer), Nickel-Cadmium (NiCD), Nickel-Metal Hydride (NiMH), Zinc-air, or other material. Further examples for energy storage device **406** include fuel cells, nano-enabled energy storage materials, capacitors, inertial energy storage devices, or other energy storage devices. In an embodiment, energy is transferred to energy storage device from energy harvesting device in the form of an electric current over a suitable wire, cable, or bundle of wires and/or cables. Energy storage device is disposed on mobile structure including being carried by, mounted on, or directly attached to mobile structure.

[0086] Energy harvesting module **402** may include one or more energy harvesting devices including but not limited to: solar array, piezo device, inertial magnetic device, RF energy harvester, etc. A plurality of energy harvesting devices can be coupled so as to cumulate their respective energy outputs (see FIG. **10**).

[0087] In embodiments, different types of energy harvesting devices can be used. For example, FIG. **7** shows a piezo device **708** as the energy harvesting device.

RF Energy Harvesting

[0088] FIG. **5** is a schematic diagram of an embodiment of the invention illustrating harvesting energy from a signal transmitted by an RFID interrogator.

[0089] A headset is worn by an operator. The operator may be carrying an RFID interrogator **504** that regularly transmits an interrogation signal **506**. The operator may be carrying and operating a mobile computer with which the headset communicates. Headset **502** has an integrated antenna **510**, in which a current is induced by signal **506**. Current induced in antenna **510** may be conditioned by a rectifier circuit and accumulated in a storage device. One suitable rectifier circuit is shown in FIG. **9**. Other types of rectifier circuits can be used as well. One suitable storage device is a capacitor or bank of capacitors. Such capacitors may be conventional capacitors, polymer capacitors or other suitable type of capacitor. Another suitable energy storage device is a battery. There are other suitable storage devices as well.

[0090] As an alternative, a conventional RFID tab can be integrated into headset **502**. The charge pump (add ref numeral) of the RFID tag (FIG. **3**) can serve as the energy harvesting device.

Solar Array Energy Harvesting

[0091] FIG. **6** is a schematic diagram of an embodiment of the invention illustrating harvesting energy from a solar array, or more generally, an optical energy harvesting device. Headset **502** is fitted with an array of solar cells **602**, preferably on an upper portion of the headset. Solar cells generate an electric current that is appropriately conditioned and stored in an energy storage device.

[0092] Optical energy harvesting converts light received by optical energy into a current, to generate energy to charge energy storage device **406**. Optical energy harvesting device **1202** can incorporate a variety of optical energy harvesting materials/devices to generate energy. For example, optical energy harvesting device **1202** may include optical-to-electrical transducers such as solar cells (or photovoltaic cells) to convert light to energy, photodiodes, optoelectronic transducers, other light sensitive elements that convert light to a current, and any other optical energy harvesting materials/devices known to persons skilled in the relevant art(s). The optical-to-electrical transducers may be positioned on a mobile structure, such as headset **502**, such that solar and/or other light energy is received in sufficient quantities to generate a charging current.

Piezo Electic Energy Harvesting

[0093] FIG. **7** is a schematic diagram of an embodiment of the invention illustrating harvesting mechanical energy using a piezo electric device. Piezoelectric materials exhibit a "piezoelectric effect," such that when they are subjected to a compressive or tensile stress, an electric field is generated across the material, creating a voltage gradient and subsequent current flow. Example suitable piezoelectric materials include polyvinylidene fluoride (PVDF) and lead zirconate titanate (PZT).

[0094] This embodiment is constructed and arranged to capture some of the vibratory energy of motion, such as for example, when an operator wearing headset **502** moves about. A headset microphone **702** is constructed in a manner that it acts like a moment arm. It is connected to a boom element **704** either directly or through a connecting element **706**. Motion of the operator results in a vibration of microphone **702** and boom element **704**. Boom element **704** drives a piezo-electric device **708** which generates an electric current as it mechanically shifts. The electric current is appropriately conditioned and stored.

[0095] Piezo device **708** could also be attached to a piece of movable machinery, such as, for example, a conveying device, such as a fork lift. Piezo device **708** can include a variety of vibration sensing/converting mechanisms, including piezoelectric transducers, magnets, and other mechanisms. Piezoelectric device **708** is attached to a moment arm formed by boom element **704**. Piezoelectric device **708** is also anchored to headset **502**. When headset **502** moves, the moment arm formed by boom **704** vibrates. This vibration deflects piezoelectric device **708**. Piezoelectric device **708** converts this mechanical deflection to a voltage gradient that charges a storage device, such as a capacitor, that ultimately provides a trickle charge to energy storage device **406**.

[0096] Piezoelectric device **708** may include a piezoelectric transducer and optionally, a rectifier, filter, DC-DC converter, etc. Some piezoelectric transducers output an AC (alternating current) current due to the piezoelectric effect caused by vibration. If such transducers are used, piezo device **708** should include a rectifier to convert the AC output current of the piezoelectric transducer into a DC (direct current) current signal. Such a rectifier may include one or more diodes arranged in a rectifier configuration. A filter can be included to smooth a rectified DC current signal. Such a filter may include a capacitor.

[0097] For further description regarding piezoelectric transducers and circuit **900**, refer to Katz, Andrew, "Residential Piezoelectric Energy Sources," Delta Smart House, Jul. 21, 2004 (all pages) (http://delta.pratt.duke.edu/downloads/piezoelectrics_andrew.doc), which is incorporated by reference herein in its entirety.

Inertial Energy Harvesting

[0098] FIG. **8** is a schematic diagram of an embodiment of the invention illustrating harvesting mechanical energy using an inertial device. This embodiment is constructed and arranged to capture some of the mechanical energy of motion when an operator wearing headset **502** moves about. A headset microphone **802** is constructed in a manner that it acts like a moment arm. It is connected to a boom element **804** either directly or through a connecting element **806**. Boom element **804** is made of a magnetic material. Motion of the operator results in a vibration of microphone **802** and boom element **804**. Boom element **804** is constructed and arranged so that its motion induces a current in an induction coil **810** as it mechanically shifts. The electric current induced in induction coil **810** is appropriately conditioned and stored.

[0099] The inertial harvesting device can be configured in a variety of ways to use magnetism to generate energy. For example, one or more magnets can be positioned near a coil of wire. When vibrations (such as due to the vibratory motion of boom element **804**) cause relative motion between the magnetic material and induction coil, a current is generated in the wire of the coil that provides a trickle charge to a storage device (not shown in FIG. **8**). The storage device may be an intermediate storage element, such as one or more capacitors, for holding charge prior to transfer to energy storage device **406**, or may be energy storage device **406** itself.

Low Voltage Rectifier Circuit

[0100] FIG. **9** is a schematic diagram of a low voltage rectifier circuit suitable for use with various energy harvesting transducer/devices. It is assumed in this embodiment that energy is being harvested by an energy harvesting transducer **902** that provides an AC signal that must be rectified so that power can be stored in a battery, capacitor, or the like. Rectification is carried out by four insulated gate field effect transistors (IGFETs) **904**, **906**, **908** and **910**. IGFETs are used because of their low-voltage properties. They can operate at the very low voltages generated by energy harvesting transducers already described, such as the piezo-electric transducer and inertial transducer. IGFETs **904**, **906**, **908** and **910** are connected in a rectifier circuit arrangement. Transducer **902** has outputs **912** and **914** characterized by an AC signal being formed between them. Gates of the IGFETS **904** and **906** are connected to node **914** of energy harvesting transducer **902**. The drain of IGFET **904** and the source of IGFET **906** are connected to a node **912** of the energy harvesting transducer, and to the gates of IGFETs **908** and **910**. Node **914** of the energy harvesting transducer is also connected to the drain of IGFET **908** and to the source of IGFET **910**. A capacitor **940** has a first node **942** connected to the source of IGFET **908** and a second node **944** connected to the drain of IGFET **910**, harvested power being made available across the first and second nodes of the capacitor. Capacitor **940** smoothes voltage swings across the outputs **950** and the positive output **960** of the rectifier circuit.

Multiple Energy Harvesting Devices

[0101] More than one energy harvesting device can be connected in a manner that allows power to be accumulated from them. FIG. **10** is a general block diagram of an embodiment illustrating how power can be harvested from two or more devices and combined. In this embodiment, power is harvested using an inertial device **1002**, an RFID tag charge pump **1004**, a solar array **1006** and a piezo device **1008**. Power from each of these harvesting devices is conditioned as necessary before being stored in an energy storage device

1014. Power requiring low voltage rectification, such as from the inertial device **1002** and the piezo device **1008** is coupled to a low voltage rectifier **1010**. An exemplary embodiment of a low voltage rectifier is shown in FIG. **9**. Solar array **1006** is coupled to a charging circuit **1012**. Storage device **1014** can be any suitable energy storage device. Examples include a rechargeable battery, a capacitor, a bank of capacitors, etc. Power stored in storage device **1014** can then be utilized as needed by coupling the storage device **1014** to an energy conditioning circuit **1016** such as a regulator. For example, conditioning circuit **1016** could be a buck boost regulator. Power stored in storage device **1014** can, for example, be used to power a headset, a computer to which a headset is tethered, a portable/mobile communication device to which a headset is tethered, an RFID interrogator, etc.

Position of Components

[0102] FIG. **11** is a schematic diagram illustrating how energy harvesting components can be placed in a worn headset **502**. One or more RF harvesting antennas can be placed in/integrated within a boom element **1102** supporting a microphone **1104**. As an alternative, RF harvesting antennas can be placed in/integrated within a head clasp support structure **1106** of headset **502**. Power storage devices can be placed in a body portion **1108** or the head clasp support structure **1106** of headset **502**. Power conditioning circuits can also be located with the head clasp support structure **1106** or the body portion **1108** of headset **502**.

ALTERNATIVE EMBODIMENTS

[0103] Although the embodiments feature a headset, the principles of the invention apply to other mobile/portable devices as well. Examples of such mobile/portable structures include conveying devices, such as a forklift, warehouse box crushers, conveyor belts, cars, trucks, etc.

CONCLUSION

[0104] While various embodiments of the present invention have been described above, it should be understood that they have been presented by way of example only, and not limitation. It will be apparent to persons skilled in the relevant art that various changes in form and detail can be made therein without departing from the spirit and scope of the invention. Thus, the breadth and scope of the present invention should not be limited by any of the above-described exemplary embodiments, but should be defined only in accordance with the following claims and their equivalents.

What is claimed is:

1. A method for providing power to a mobile device in an RFID system, comprising:

harvesting energy at the mobile device;

charging an energy storage device disposed on the mobile structure with the generated energy; and

powering one or more circuits of the mobile device with energy stored in the energy storage device.

2. The method of claim **1**, wherein the harvesting comprises:

(a) generating the energy from vibration of the mobile structure during operation of the mobile structure.

3. The method of claim **2**, wherein step (a) comprises:

generating the energy with a piezoelectric transducer mounted to the mobile structure.

4. The method of claim **1**, wherein said generating comprises:

generating energy produced by relative movement of a magnet and a coil that are mounted to the mobile structure.

5. The method of claim **1**, wherein said generating comprises:

converting light received at the mobile device into energy.

6. The method of claim **1**, wherein said converting light comprises:

converting light received at the mobile device into energy using an optical-to-electrical transducer.

7. The method of claim **1**, wherein said generating comprises:

obtaining power from a charge pump of an RFID tag associated with the mobile device.

8. The method of claim **1**, wherein said generating comprises:

inducing a current in an antenna receiving a signal from an RFID interrogator; and

conditioning the current induced in the antenna for storage in the energy storage device.

9. The method of claim **1**, further comprising:

conditioning power from the energy storage device.

10. The method of claim **9**, wherein the conditioning comprises conditioning power with a buck boost regulator.

11. An arrangement for providing power to a mobile device in an RFID system, comprising:

an energy harvesting device associated with the mobile device;

an energy storage device disposed on the mobile device constructed and arranged to store power harvested by the energy harvesting device; and

a powering conditioning circuit for obtaining power from the energy storage device and providing power to one or more circuits of the mobile device.

12. An arrangement according to claim **11** further comprising a low voltage rectifier circuit, including:

first and second P-channel FETs; and

first and second N-channel FETs, the P and N-channel FETs being arranged in an H-bridge configuration.

13. An arrangement according to claim **11** wherein the energy harvesting device is a charge pump.

14. An arrangement according to claim **11** wherein the energy harvesting device is a piezo-electric device.

15. An arrangement according to claim **11** wherein the energy harvesting device is a magnetic device.

16. An arrangement according to claim **11** wherein the energy harvesting device is a photo-optic device.

17. An arrangement according to claim **11** wherein the energy storing device is a capacitor.

18. An arrangement according to claim **11** wherein the energy storing device is a battery.

19. A system for providing power to a portable device that is part of a radio frequency identification (RFID) system, comprising:

an energy storage device disposed on the portable device structure and coupled to the reader; and

an energy harvesting device disposed on the mobile structure that generates energy;

wherein the energy storage device stores the generated energy.

20. An arrangement according to claim **19** further comprising a low voltage rectifier circuit, including:

first and second P-channel FETs; and

first and second N-channel FETs, the P and N-channel FETs being arranged in an H-bridge configuration.

21. The system of claim **19**, wherein the mobile structure is a headset.

22. The system of claim **19**, wherein said energy harvesting device comprises:

a vibratory energy harvesting device that generates the energy from vibration of the mobile structure during operation of the mobile structure.

23. The system of claim **19**, wherein said vibratory energy harvesting device comprises:

a piezoelectric transducer.

24. The system of claim **19**, further comprising:

a moment arm that mounts the piezoelectric transducer; and

a capacitor coupled to the energy storage device;

wherein vibration of the arm causes the arm to deflect the piezoelectric transducer;

wherein the piezoelectric transducer generates a current due to the deflection, and

wherein the current charges the capacitor.

25. The system of claim **19**, wherein said energy harvesting device comprises:

a magnetic induction energy harvesting device.

26. The system of claim **25**, wherein said magnetic energy harvesting device comprises:

a magnet; and

a coil;

wherein the energy is generated by movement of the magnet through the coil field.

27. The system of claim **19**, wherein said energy harvesting device comprises:

an photo-optical-to-electrical transducer that converts light received at the mobile structure into energy.

28. A low voltage rectifier circuit, comprising:

first and second P-channel FETs; and

first and second N-channel FETs, the P and N-channel FETs being arranged in an H-bridge configuration.

* * * * *

US 20100007218A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2010/0007218 A1**

Ogram (43) **Pub. Date: Jan. 14, 2010**

(54) **ATMOSPHERIC ELECTRICAL GENERATOR**

(76) Inventor: **Mark Ellery Ogram**, Tucson, AZ (US)

Correspondence Address:
Mark E. Ogram (30343)
Ste. 110-186, 6990 E. 22nd Street
Tucson, AZ 85710-5192 (US)

(21) Appl. No.: **12/218,297**

(22) Filed: **Jul. 14, 2008**

Publication Classification

(51) **Int. Cl.**
H02G 11/00 (2006.01)

(52) **U.S. Cl.** **307/145**

(57) **ABSTRACT**

A mechanism to tap an electrical source which uses an aircraft (preferably a lighter than air balloon) tethered by a conductive line. The conductive line is extended/withdrawn by a winch motor to adjust the altitude of the aircraft. The conductive line is isolated from the ground and an electrical conductor is connected to the conductive line and to an electrical load. In this manner, static electricity generated in the atmosphere is gathered for use.

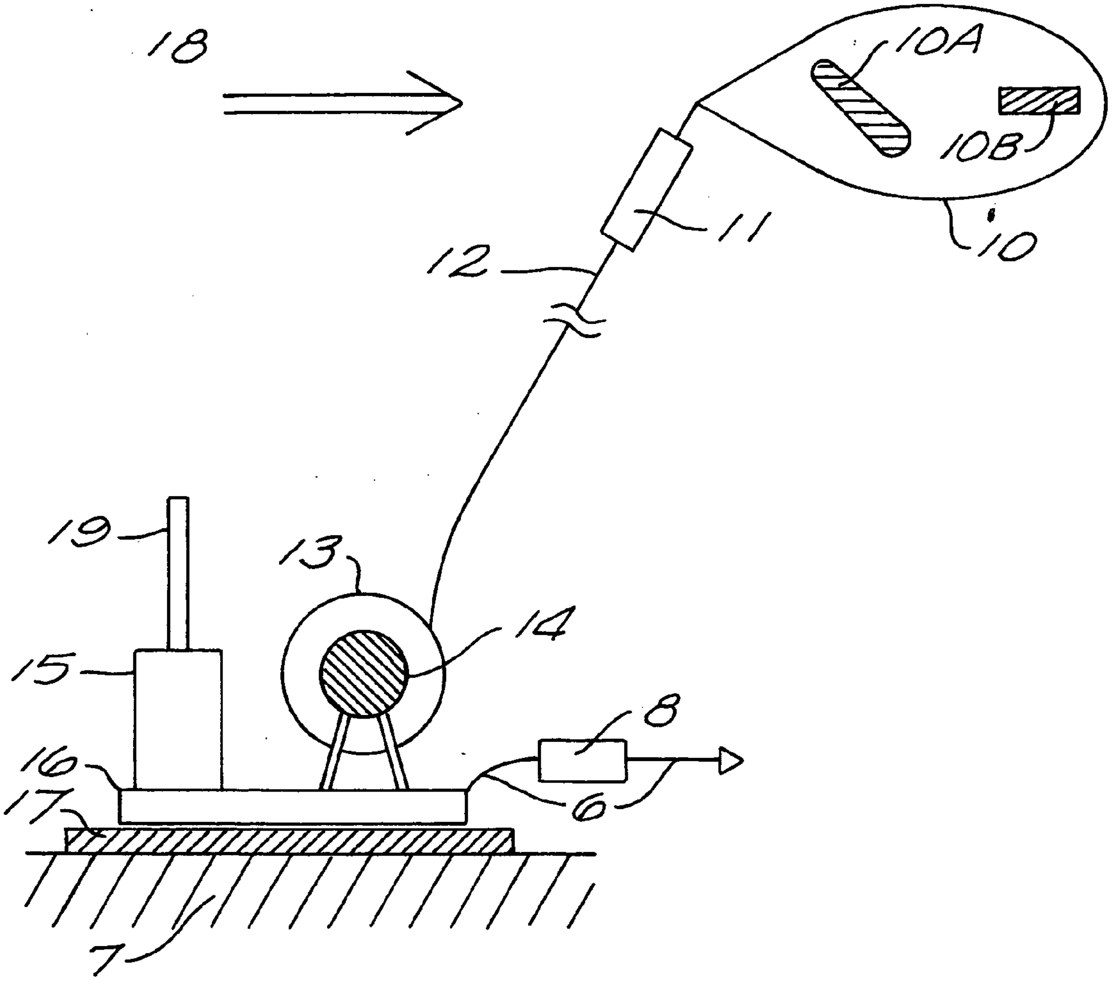

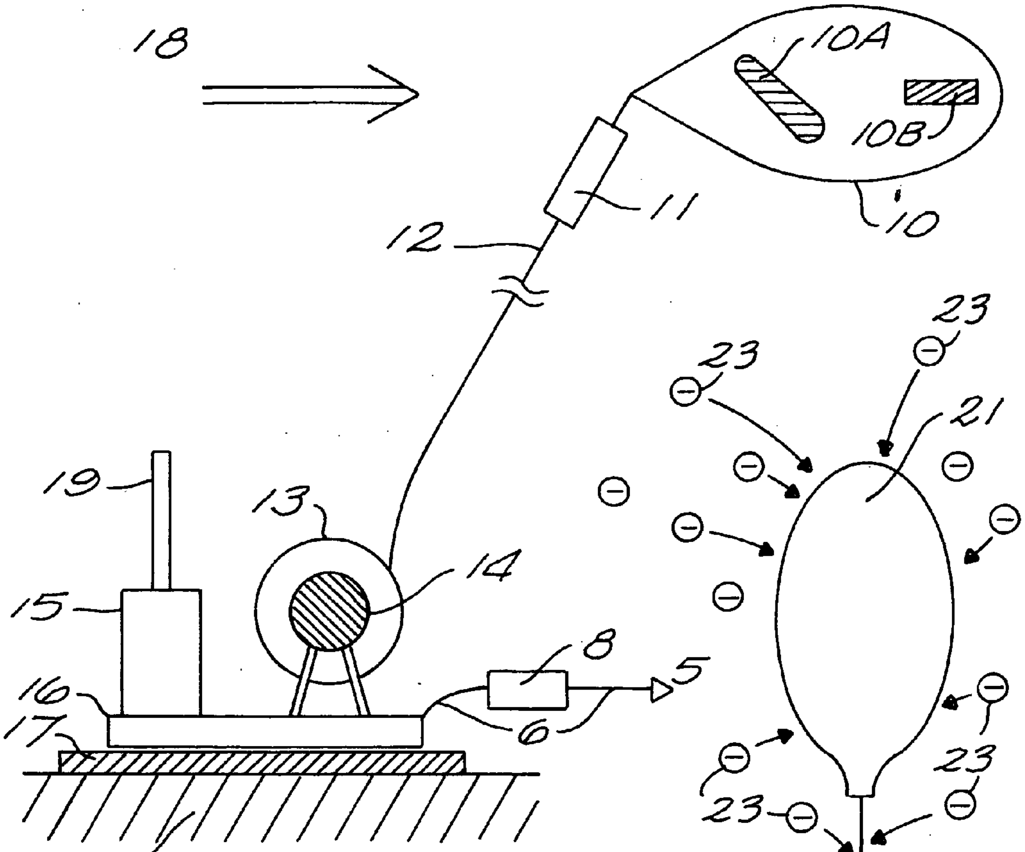
18
10A
10B
10
11
12
19
13
14
15
8
5
16
6
17
7

FIG. 1

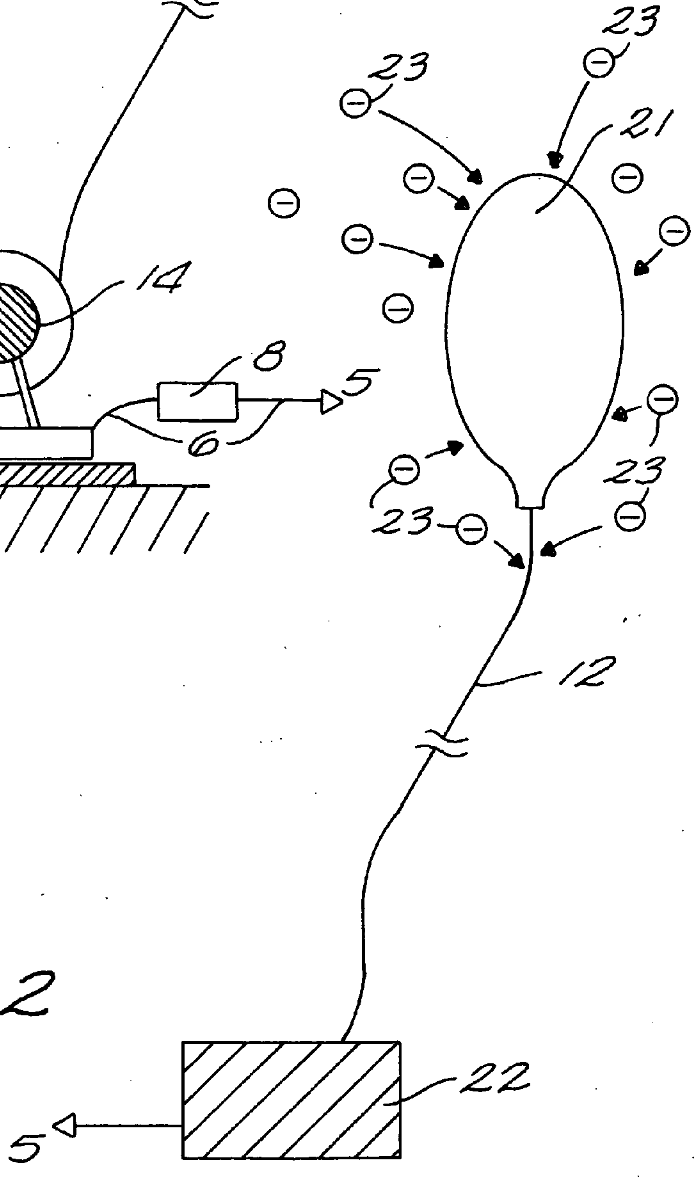
23
23
21
23
23
23
12
22
5

FIG. 2

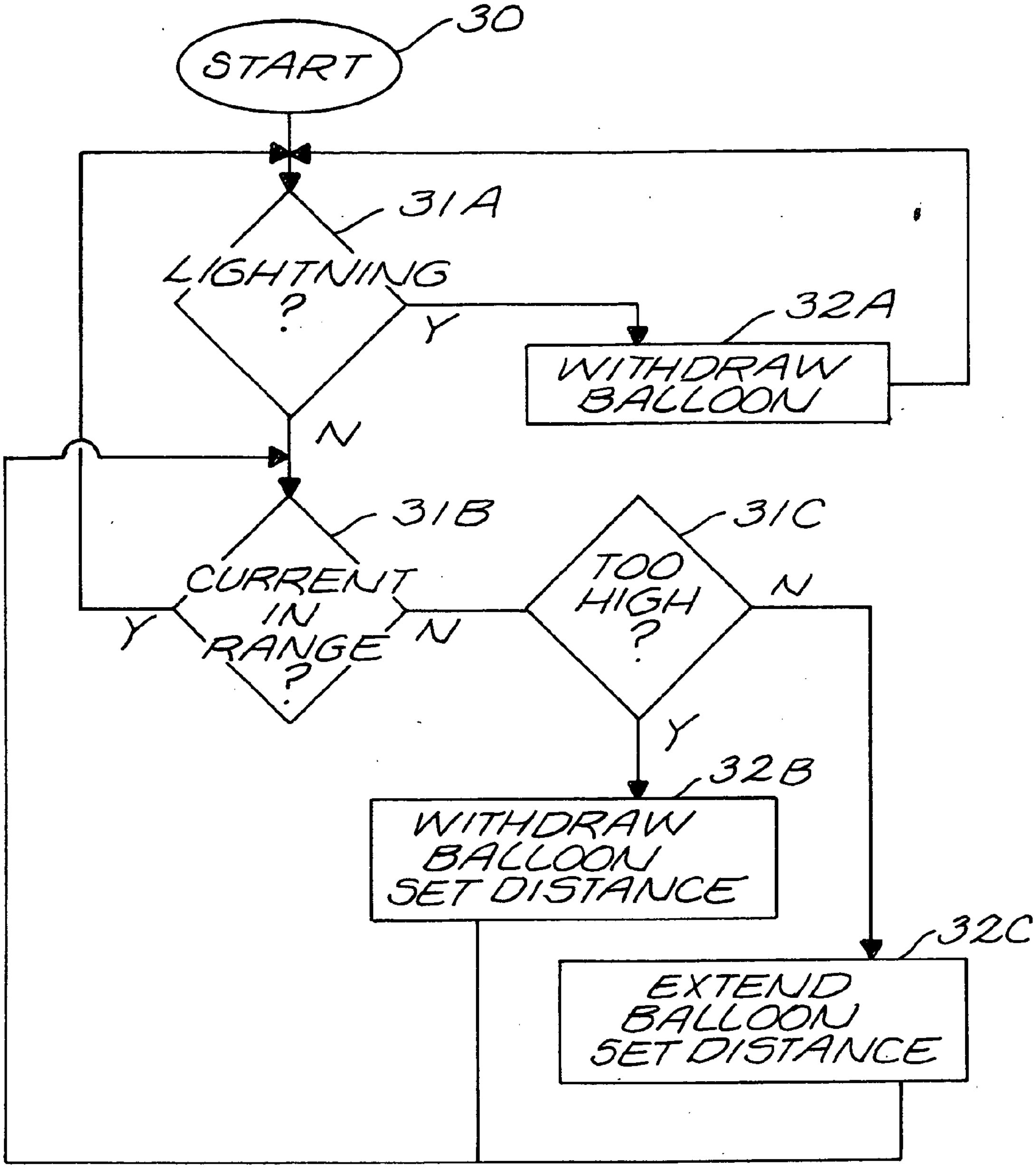
30
START
31A
LIGHTNING ?
Y
N
32A
WITHDRAW BALLOON
31B
CURRENT IN RANGE ?
Y
N
31C
TOO HIGH ?
N
Y
32B
WITHDRAW BALLOON SET DISTANCE
32C
EXTEND BALLOON SET DISTANCE

FIG. 3

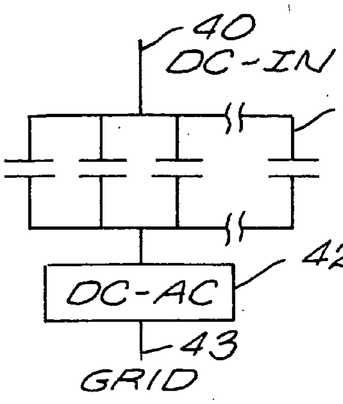
40
DC-IN
41
42
DC-AC
43
GRID

FIG. 4A

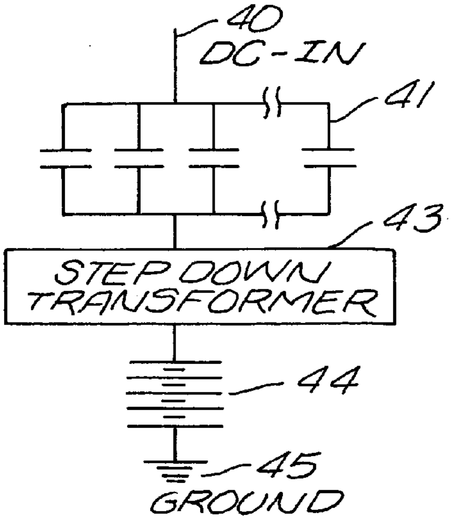
40
DC-IN
41
43
STEP DOWN TRANSFORMER
44
45
GROUND

FIG. 4B

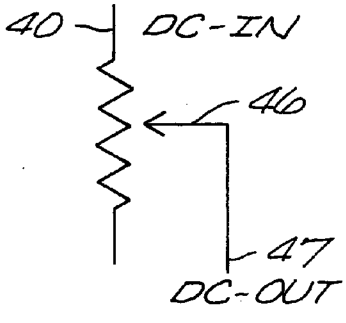
40
DC-IN
46
47
DC-OUT

FIG. 4C

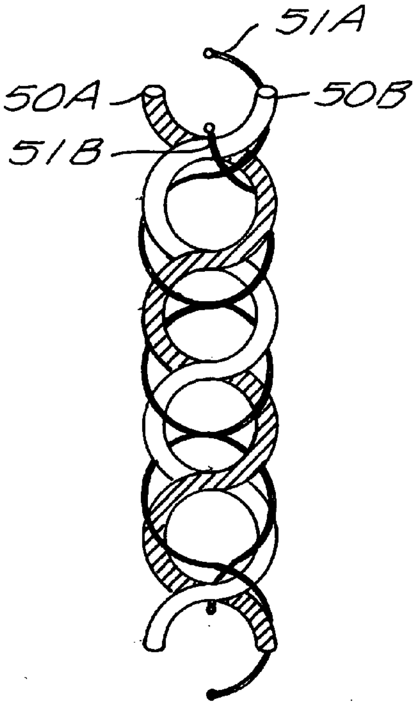
51A
50A
50B
51B

FIG. 5

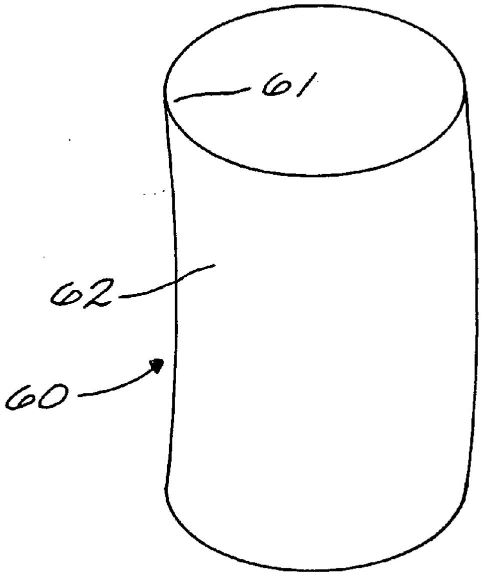

FIG. 6A

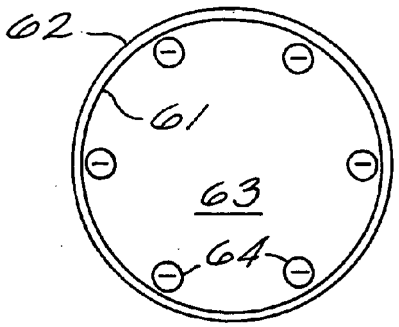

FIG. 6B

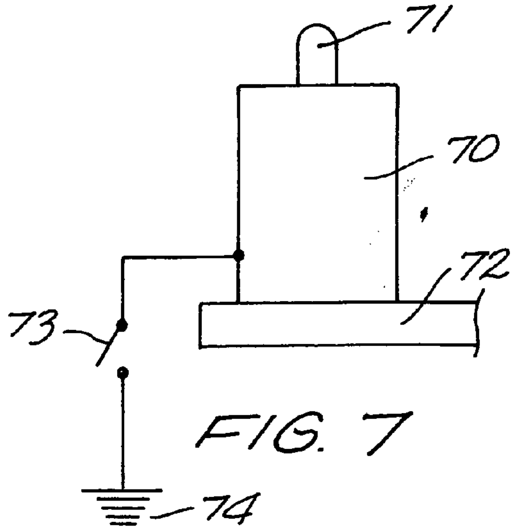
71
70
72
73
74

FIG. 7

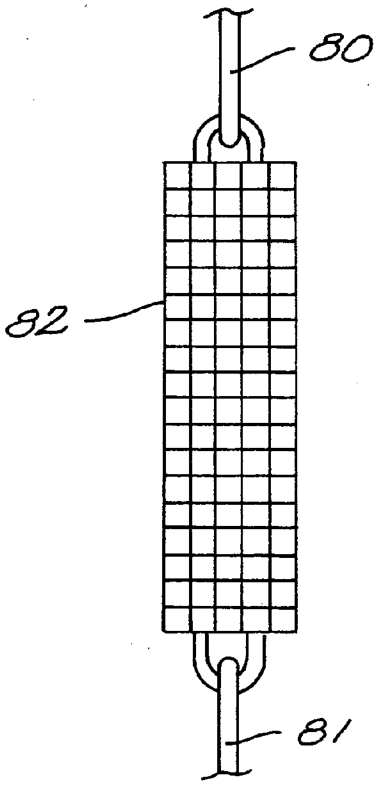

FIG. 8A

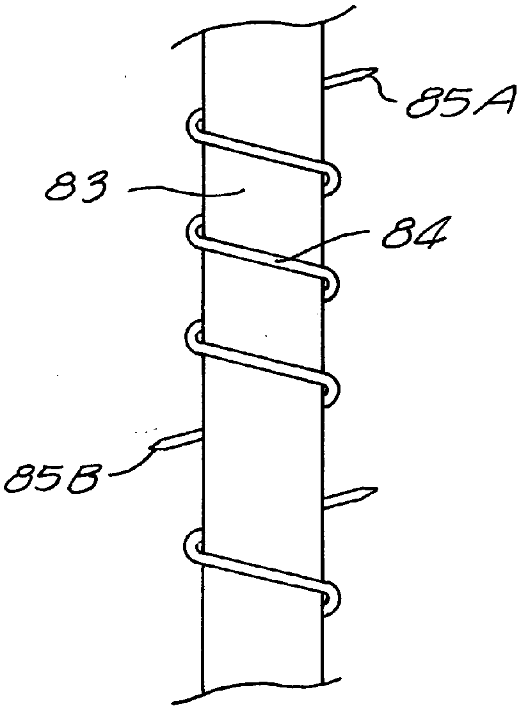
85A
83
84
85B

FIG. 8B

ATMOSPHERIC ELECTRICAL GENERATOR

BACKGROUND OF THE INVENTION

[0001] This invention relates to the production of electrical energy and more particularly to the production of electrical energy from the atmosphere.

[0002] Everyone is familiar with Benjamin Franklin's kite experiment of 1752. Using a kite whose string had become wet, negative charges from the passing clouds flowed into the string, down to the suspended key, and then into a Leyden jar via a thin metal wire. Franklin was protected by a dry silk string; but, when Franklin's knuckle came too close to the key, he received a strong shock. Fortunately, Benjamin Franklin was not killed, others who tried this same experiment were not so lucky.

[0003] Since then, the formation of lightning has remained something of a mystery. Lightning bolts are triggered when a negatively charged cloud base induces a positive charge from the ground, thereby forming a "pathway" for the discharge of the collected electrical energy.

[0004] Lightning travels up to 60,000 miles per hour with a flash that is brighter than ten million 100-watt lightbulbs. This wattage is as much power as is produced by all of the electricity plants in the United States and with a voltage of up to 300 million volts.

[0005] It is this very fact, the power within lightning is immense, that has prevented any successful collection of the electrical energy from lightning. The electricity in lightning is far too extreme for current technology to harness.

[0006] While lightning has attracted a energy starved industrial world, no one has developed any technique to harness this naturally occurring electrical source.

[0007] It is clear there is a continuing need for an electrical source other than carbon-based fuels and that the naturally occurring electricity in the atmosphere is being ignored.

SUMMARY OF THE INVENTION

[0008] The invention is a mechanism which taps into the naturally occurring static electricity in the atmosphere. Whereas heretofore, the attempt to garner electricity from the atmosphere has focused exclusively on capturing lightning, the present invention syphons off the static electricity which is generated from any agitated air and voids lightning.

[0009] Lightning is only the final discharge of the static electricity, whether that lightning is intra-cloud lightning, cloud-to-ground lightning, or inter-cloud lightning. Other types of final discharges are known as heat lightning, summer lightning, sheet lightning, ribbon lightning, silent lightning, ball lightning, bead lightning, elves, jets, and sprites. Well before these discharges are observed, as the atmosphere becomes agitated by wind or thermal, static electricity is being generated.

[0010] The present invention recognizes that this static electricity is being formed and creates a mechanism to capture it.

[0011] The mechanism of this invention utilizes an aircraft such as a lighter than air balloon. While the preferred embodiment uses a foil balloon, a variety of other aircraft are obvious to those of ordinary skill in the art, including, but not limited to: gliders, rubber balloons (such as weather balloons), biaxially-oriented polyethylene terephthalate polyester film balloons, and latex balloons.

[0012] Within this discussion, the balloon is referenced, but, the invention is not intended to be limited solely to balloons.

[0013] The balloon is sent aloft and is tethered by a conductive line. In this context, the conductive line may be any obvious to those of ordinary skill in the art. For the preferred embodiment, the conductive line is a generically referred to as a "poly-rope" and is commercially available through a variety of sources. A suitable conductive line is described in U.S. Pat. No. 5,203,542, entitled "Apparatus for Improved Electric Fence Wire Construction for use with Intensive Grazing" issued Apr. 20, 1993, to Coley, et al. and incorporated hereinto by reference.

[0014] The conductive line is played out of a winch to control the altitude of the balloon. The motor controlling the winch is able to reverse direction to both extend and withdraw the conductive line which is wrapped around a spool on the winch. The winch/spool combination are part of a base unit.

[0015] In some embodiments of the invention, the spool is constructed of rubber so as to insulate the conductive line from the winch assembly. In this embodiment, only the conductive line is charged by the atmospheric static electricity while the winch remains neutral.

[0016] In yet another embodiment, the winch/spool are part of a base unit which is itself isolated from the ground by an insulator. In this embodiment, the entire base unit is charged by the atmospheric static electricity.

[0017] A conductor, such as an insulated wire, is electrically connected to the conductive line. In one embodiment, where the conductive line is electrically isolated from the spool and winch motor, the conductor is connected to the conductive line. In the embodiment where the conductive line is electrically connected to the base unit, then the conductor is connected anywhere on a metallic base unit.

[0018] The other end of the conductor is connected to a load. The load in this case can be any of a variety of electrical loads well known to those of ordinary skill in art, including, but not limited to a motor, a battery system, or the electrical grid for the system.

[0019] In the preferred embodiment, a sensor array is used to monitor the activities both at the base unit (such as electrical flow within the conductor) and in the surrounding locale.

[0020] A sensor monitoring the electrical flow (i.e. voltage and/or current) within the conductor is used to monitor the electrical activity within the conductor.

[0021] In the preferred embodiment, a lightning sensor monitors for lightning activity within the locale. As noted earlier, the electrical characteristic of lightning is so extreme that ideally this discharge is avoided as it might damage the mechanism of this invention.

[0022] The sensor array is utilized by a controller, such as microprocessor, programmed to operate the mechanism as outlined herein.

[0023] The controller operates the winch motor to extend or withdraw the conductive line and by extension the altitude of the balloon. The controller is programmed to operate the winch by monitoring the electrical characteristics of the conductor and adjusting the balloon's altitude to maintain these characteristics within the conductor within a preset range.

[0024] This preset range is established either in the base programming of the controller or is established by an operator of the system.

[0025] As example, by controlling the amount of current being withdrawn from the atmosphere, the mechanism oper-

ates within a safe range and also provides a relatively stable current flow from which a variety of activities can take place (such as DC-AC conversion).

[0026] The controller also utilizes the lightning sensor to protect the mechanism from a lightning strike. Should lightning be detected within a pre-determined range (as established by the software or defined by an operator), then the balloon is pulled down to minimize the risk of damage from a lightning strike.

[0027] The invention, together with various embodiments thereof will be more fully explained by the following description of the accompanying drawings.

DRAWINGS IN BRIEF

[0028] FIG. **1** diagrams the preferred embodiment of the invention.

[0029] FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

[0030] FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

[0031] FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

[0032] FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

[0033] FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

[0034] FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

[0035] FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

DRAWINGS IN DETAIL

[0036] FIG. **1** diagrams the preferred embodiment of the invention.

[0037] Balloon **10** is an aircraft which, in this illustration, is a lighter than air balloon. Wings **10**A, extending from the body of balloon **10**, provide additional lift in air flow **18**. Tail **10**B helps to stabilize balloon **10**.

[0038] Balloon **10** is tethered to the ground via conductive line **12**. As noted earlier, a variety of configurations and materials are available to serve as conductive line **12**. In this illustration, a poly-wire is used. Poly-wire is commercially available through a variety of vendors, including, but not limited to: Jeffers Livestock and Sareba Systems, Inc. of Ellendale, Minn.

[0039] In this embodiment, located proximate to balloon **10**, is an electrical collection enhancement lead **11** which assists in the collection of the static electrical charge in the atmosphere. Electrical collection enhancement lead **11** is configured to attract the static charge and conduct the charge into the conductive line **12**.

[0040] The electricity flows down the conductive line into spool **13**, where the conductive line **12** is collected and either withdrawn or dispensed through operation of winch motor **14**.

[0041] Winch motor **14** and spool **13** are mounted onto base unit **16** which is electrically isolated from ground **7** using insulator **17**. Note, in this embodiment of the invention, when electricity is being collected from the atmosphere, the entire base unit **16** becomes charged. In another embodiment of the invention, spool **13** is constructed of rubber, thereby preventing base unit **16** from becoming charged, thereby restricting the charging from the atmosphere to only conductive line **12**.

[0042] In this embodiment, conductor **6** is connected to base unit **16** (since the entire base unit **16** is charged and the base unit is metallic) to communicate the electrical current to load **5**. Conductor **6** is ideally an insulated wire.

[0043] The electrical current through conductor **6** is measured using sensor **8**.

[0044] In the alternative embodiment discussed above, where only the conductive line **12** is charged, then conductor **6** is connected to conductive line **12**.

[0045] Controller **15**, located in this embodiment on base unit **16**, operates winch motor **14** in response to signals from sensor **8** (measuring the current being discharged to load **5**) to maintain the current flow within a pre-defined range. As the current flow diminishes, then the conductive line **12** extended from spool **13** to increase the altitude of balloon **10** to that more static charge from the atmosphere is gathered; as the current flow falls exceeds a preset level, conductive line **12** is withdrawn onto spool **13** to decrease the static charge being collected from the atmosphere.

[0046] The range of current flow through conductor **6** is ideally set by the program, although some embodiments of the invention permit an operator to establish this range of operation.

[0047] In an alternative embodiment, the sensor monitoring conductor **6** monitors the voltage therein.

[0048] In the preferred embodiment of the invention, controller **15** is also equipped with a lightning sensor **19**. In this embodiment, when lightning is sensed within a preset range, then substantially all of conductive line **12** is wound onto spool **13** to pull balloon **10** near the ground and protect the entire mechanism from being damaged from a lightning discharge.

[0049] In the preferred embodiment, the "safe" distance form lightning is set in the programming of controller **15** and is ideally two miles; other embodiments permit the operator to "safe" distance.

[0050] There are a variety of lightning sensors well known to those of ordinary skill in the art, including, but not limited to those described in: U.S. Pat. No. 7,016,785, entitled "Lightning Detection" issued to Makela, et al. on Mach 21, 2006; U.S. Pat. No. 6,829,911, entitled "Lightning Detection and Prediction Alarm Device" issued to Jones, et al. on Dec. 7, 2004; U.S. Pat. No. 7,200,418, entitled "Detection of Lightning" issued to Karikuranta, et al. on Apr. 3, 2007; and U.S. Pat. No. 6,961,662, entitled "Systems and Methods for Spectral Corrected Lightning Detection" issued to Murphy on Nov. 1, 2005; all of which are incorporated hereinto by reference.

[0051] In another embodiment of the invention, controller **15** is not located on base unit **16**, rather it is remote and communicates its control signals to winch motor **14** using radio waves.

[0052] FIG. **2** illustrates the collection of the negative charged particles in the atmosphere.

[0053] Static charges **23** are generated in the atmosphere by agitated air. These static charges are often collected at the bottom of clouds, but exist in other environments as well.

[0054] Balloon **21** is extended into this strata of static charges **23** which are then attracted to conductive line **12** to flow to base unit **22** and then onto load **5**.

[0055] By increasing or decreasing the altitude of balloon **21** (defined by the length of the extended conductive line **12**),

conductive line **12** is selectively exposed to varying densities and levels of the static charge strata, and by extension, the current flow or voltage is increased or decreased.

[0056] FIG. **3** is a flow-chart of the operation of the controller for the preferred embodiment of the invention.

[0057] Once the program starts **30**, the lightning sensor is checked to determine if lightning has occurred within the unsafe range **31**A, if it has, then the balloon is lowered **32**A, and the program continues monitoring the status of lightning until no lightning is detected.

[0058] When the lightning status is acceptable, then the current within the conductor is checked to see if the current is within the prescribed range **31**B. If the current is acceptable (within range) the program returns to check the lightning status **31**A; otherwise a determination is made to see if the current is above the prescribed range **31**C.

[0059] If the current is above the prescribed range, then the altitude of the balloon is withdrawn a set amount **32**B (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

[0060] If the current is below the prescribed range, then the altitude of the balloon is extended a set amount **32**C (ideally twenty-five feet) and the program loops back to see if the current is within range **31**B.

[0061] In this manner of feed-back and minor adjustments in the altitude of the balloon, the current is maintained within a prescribed range which can be handled by the downstream electrical system.

[0062] As noted earlier, some embodiments of the invention monitor the voltage instead of the current.

[0063] FIGS. **4**A, **4**B, and **4**C are electrical schematics for handling the static charge from the atmosphere.

[0064] By maintaining the voltage being collected in a prescribed range, an electrical conversion system is easily designed. While FIGS. **4**A, **4**B, and **4**C illustrate some electrical configurations, those of ordinary skill in the art readily recognize a variety of other configurations which will serve the same function.

[0065] Referencing FIG. **4**A, Direct Current In (DC IN) **40** is buffered by a gang of capacitors **41** before being communicated to a DC/AC converter **42**. The DC/AC converter converts the direct current into a an alternating current suitable for placement over an existing electrical grid **43** such as normally found from a power-plant.

[0066] Those of ordinary skill in the art readily recognize a variety of DC/AC converters, including, but not limited to: U.S. Pat. No. 7,394,671, entitled "Controller IC, DC-AC Conversion Apparatus, and parallel running system of DC-AC Conversion Apparatuses" issued to Fukumoto, et al. on Jul. 1, 2008; and, U.S. Pat. No. 7,330,366, entitled "DC-AC Converter" issued to Lee, et al. on Feb. 12, 2008; all of which are incorporated hereinto by reference.

[0067] FIG. **4**B illustrates an electrical arrangement suitable for use in charging a battery. DC IN **40** is buffered by capacitor bank **41** before entering into a step down transformer **43**. Step down transformer **43** reduces the voltage so that the voltage can safely be introduced into battery **44** which is connected to ground **45** at the battery's other pole.

[0068] Those of ordinary skill in the art readily recognize a variety of batteries which will work in this capacity, including, but not limited to those described in: U.S. Pat. No. 7,378,181, entitled "Electric Storage Battery Construction and Manufacture" issued to Skinlo on May 27, 2008; U.S. Pat. No. 7,388,350, entitled "Battery with Electronic Compartment" issued to Wright on Jun. 17, 2008; U.S. Pat. No. 7,397,220, entitled "Connection Member and Battery Pack" issued to Uchida, et al. on Jul. 8, 2008; and, U.S. Pat. No. 7,375,492, entitled "Inductively Charged Battery Pack" issued to Calhoon, et al. on May 20, 2008; all of which are incorporated hereinto by reference.

[0069] In FIG. **4**C, DC IN **40** is fed into an adjustable rheostat **46** which is controlled by the controller so that the DC OUT **47** falls within a specified range.

[0070] FIG. **5** illustrates a conductive line used in the preferred embodiment of the invention.

[0071] This type of conductive line is commonly called poly-wire and consists of multiple interwoven strands of plastic **50**A and **50**B woven into a cord or rope arrangement having intertwined therein exposed metal wires **51**A and **51**B. While this illustration shows two plastic strands and two metal wires, any number of possible combinations is possible.

[0072] The exposed metal wires **51**A and **51**B attract the atmospheric static charge and transmit the charge down to the base unit (not shown).

[0073] FIGS. **6**A and **6**B illustrate an alternative conductive line creating an ionized pathway for the flow of the static charges from the atmosphere.

[0074] This conductive line utilizes a tube **60** having an outer layer **62** of PET Film (Biaxially-oriented polyethylene terephtalate polyester film) which provides exceptionally high tensile strength and is chemically and dimensionally stable. The tube has an ideal diameter of between two and three inches.

[0075] An interior metal coating **61** provides an initial conduit for the flow of static charge. The static charge through the metal forces the tube to expand due to the repulsion experienced by like charges. Further, the flow of electricity causes the interior of the tube **60** to become ionized to provide an additional pathway for the atmospheric static charges to the base unit (not shown).

[0076] Because outer layer **62** provides a gas barrier, the resulting ionization is not dissipated by air currents, thereby providing a highly stable pathway.

[0077] FIG. **7** illustrates the controller of an alternative embodiment and the associated safety devices.

[0078] In this embodiment, controller box **70**, resting on insulating pad **72**, is in communication with the sensors as described above. Using the input from these sensors, when there is flow of electricity through the base unit, warning flashing light **71** is illuminated. To electrically neutralize the mechanism, switch **73** is activated to pass any existing current into the ground **74**.

[0079] FIGS. **8**A and **8**B illustrate two embodiments of enhanced electrical collection leads.

[0080] Referencing FIG. **8**A, enhanced electrical collection lead **82** is a wire mesh which is in electrical communication with conductive line **81** and balloon **80**. Because of the significant amount of metal exposed by enhanced electrical collection lead **82**, more static electricity from the atmosphere is drawn to the collection lead **82**, and then down conductive line **81** to the base unit (not shown).

[0081] Conductive lead **82** is positioned proximate to balloon **80**.

[0082] In FIG. **4**B, poly-wire **83** has enhanced electrical collection leads **84** wrapped therearound. Collection leads **84** have pointed ends **85**A and **85**B which have a propensity to attract more electricity than rounded ends do.

[0083] It is clear from the foregoing that the present invention captures an entirely new source of electrical energy.

What is claimed is:

1. A mechanism to tap an electrical source comprising:

a) a lighter than air balloon;

b) a base unit having a spool of conductive line on a winch motor, one end of said conductive material secured to said lighter than air balloon, said winch motor capable of selectively extending or withdrawing said conductive line from said spool, one end of said conductive line connected to said lighter than air balloon;

c) an insulator electrically isolating said conductive line from ground; and,

d) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load.

2. The mechanism to tap an electrical source according to claim **1**, further including a controller, said controller selectively operating said winch motor based upon a voltage level in said conductor.

3. The mechanism to tap an electrical source according to claim **2**,

a) further including an electrical flow sensor monitoring electrical flow through said conductor and generating an electrical flow indicia indicative of said electrical flow in said conductor, said electrical flow indicia communicated to said controller; and,

b) wherein said controller operates said winch motor such that said electrical flow indicia remains within a selected operating range.

4. The mechanism to tap an electrical source according to claim **3**, wherein said selected operating range is established by an operator.

5. The mechanism to tap an electrical source according to claim **4**, further including a warning light activated when said electrical flow indicia is non-zero.

6. The mechanism to tap an electrical source according to claim **3**,

a) further including a lightning sensor generating a lightning presence indicia indicative of lightning within a prescribed range, said presence indicia being communicated to said controller; and,

b) wherein said controller, in response to said lightning presence indicia, operates said winch motor to withdraw substantially all of said conductive line onto said spool.

7. The mechanism to tap an electrical source according to claim **6**, wherein said prescribed range is established by an operator.

8. The mechanism to tap an electrical source according to claim **3**, further including an electrical collection enhancement lead in electrical contact with a first end said conductive material, said electrical collection enhancement lead configured to attract static electricity.

9. The mechanism to tap an electrical source according to claim **8**, wherein said electrical collection enhancement lead is positioned proximate to said lighter than air balloon.

10. The mechanism to tap an electrical source according to claim **9**, wherein said electrical collection enhancement lead includes at least two pointed electrical conductors.

11. The mechanism to tap an electrical source according to claim **1**, wherein said insulator electrically isolates said conductive line from said winch motor.

12. The mechanism to tap an electrical source according to claim **1**, wherein said insulator electrically isolates said base unit from ground.

13. A mechanism comprising:

a) an airborne aircraft tethered by a conductive line secured to a winch capable of extending or withdrawing said conductive line from a spool, thereby adjusting an altitude of the airborne aircraft;

b) an insulator electrically isolating said conductive line from ground; and,

c) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load.

14. The mechanism according to claim **13**, further including a controller, said controller selectively operating said winch motor based upon a voltage level in said conductor.

15. The mechanism to tap an electrical source according to claim **14**,

a) further including a lightning sensor communicating with said controller; and,

b) wherein said controller, in response to selected signals from said lightning sensor withdraws substantially all of said conductive line.

16. A mechanism to tap an electrical source comprising:

a) a lighter than air balloon;

b) a base unit having a spool of conductive line on a winch motor, one end of said conductive material secured to said balloon, said winch motor capable of adjusting an altitude of said lighter than air balloon by selectively extending or withdrawing said conductive line from said spool;

c) an insulator electrically isolating said conductive line from ground; and,

d) a conductor having a first end electrically connected to said conductive line and a second end electrically connected to a load;

e) a sensor array having,

   1) an electrical flow sensor monitoring electrical flow through said conductor, and,

   2) a lightning sensor monitoring existence of proximate lightning; and,

f) a controller responsive to signals from said sensor array to selectively operate said winch motor.

17. The mechanism to tap an electrical source according to claim **16**,

a) further including a warning light; and,

b) wherein said controller activates said warning light when said electrical flow in said conductor is non-zero.

18. The mechanism to tap an electrical source according to claim **17**, further including an electrical collection enhancement lead in electrical contact with said conductive material, said electrical collection enhancement lead configured to attract static electricity and positioned proximate to said lighter than air balloon.

* * * * *

US 20100044496A1

(19) **United States**

(12) **Patent Application Publication**

Gosling

(10) **Pub. No.: US 2010/0044496 A1**

(43) **Pub. Date: Feb. 25, 2010**

(54) **ANNULAR AIRBORNE VEHICLE**

(75) Inventor: **Harry George Dennis Gosling**, Bristol (GB)

Correspondence Address:
BARNES & THORNBURG LLP
P.O. BOX 2786
CHICAGO, IL 60690-2786 (US)

(73) Assignee: **GO SCIENCE LIMITED**, Bristol (GB)

(21) Appl. No.: **12/594,719**

(22) PCT Filed: **Apr. 16, 2008**

(86) PCT No.: **PCT/GB2008/001329**

§ 371 (c)(1),
(2), (4) Date: **Oct. 5, 2009**

(30) **Foreign Application Priority Data**

Apr. 18, 2007 (GB) 0707512.0

Publication Classification

(51) **Int. Cl.**
B64C 15/12 (2006.01)
B64C 9/00 (2006.01)

(52) **U.S. Cl.** **244/12.5**; 244/87

(57) **ABSTRACT**

An airborne vehicle having a wing-body which defines a wing-body axis and appears substantially annular when viewed along the wing-body axis, the interior of the annulus defining a duct which is open at both ends. A propulsion system is provided comprising one or more pairs of propulsion devices, each pair comprising a first propulsion device mounted to the wing-body and positioned on a first side of a plane including the wing-body axis, and a second propulsion device mounted to the wing-body and positioned on a second side of the plane including the wing-body axis. A direction of thrust of the first propulsion device can be adjusted independently of the direction of thrust of the second propulsion device and/or a magnitude of thrust of the first propulsion device can be adjusted independently of the magnitude of thrust of the second propulsion device. In certain embodiments the wing-body appears swept forward when viewed from a first viewing angle, and swept backward when viewed from a second viewing position at right angles to the first viewing angle.

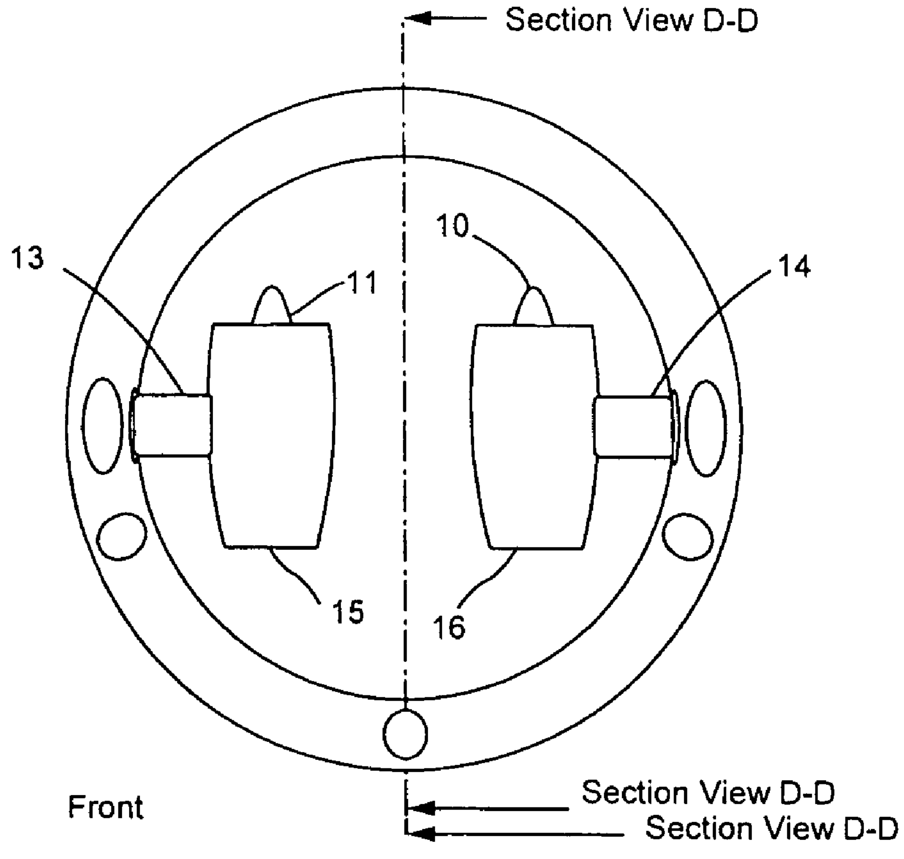

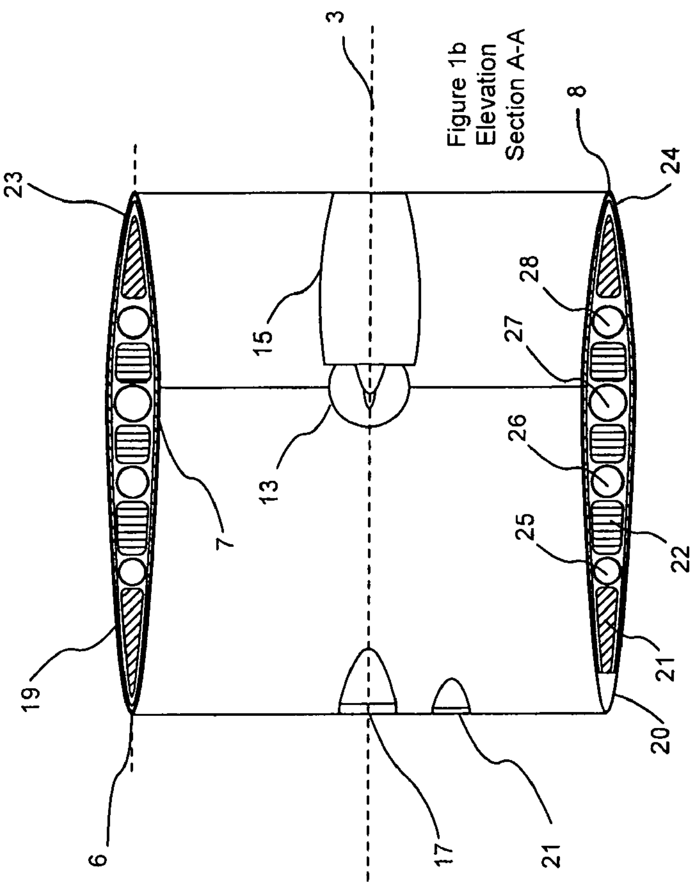
3
23
8
24
15
28
27
13
26
7
25
22
19
21
20
6
17
21

Figure 1b
Elevation
Section A-A

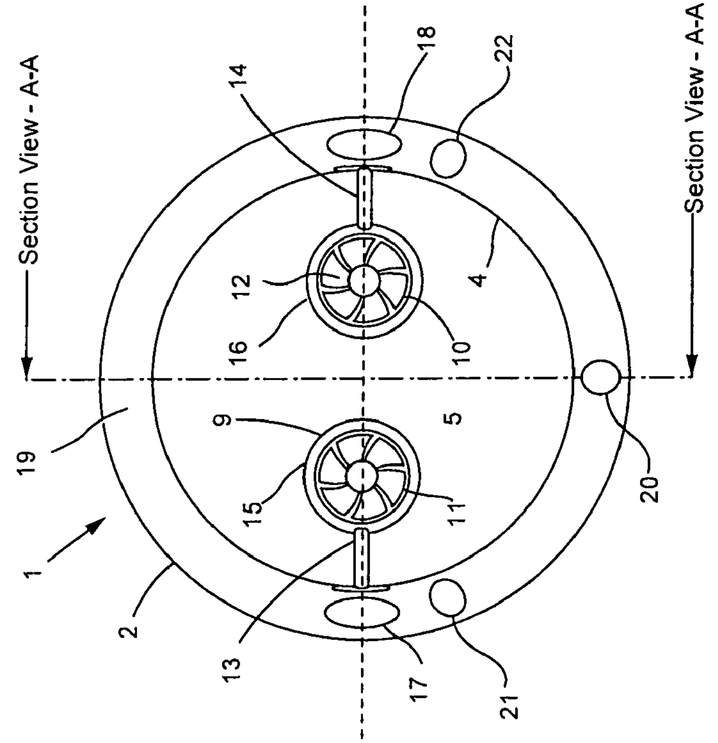
Section View - A-A
Section View - A-A
1
19
2
13
17
21
15
9
11
5
20
16
12
10
4
14
18
22

Figure 1a
Front

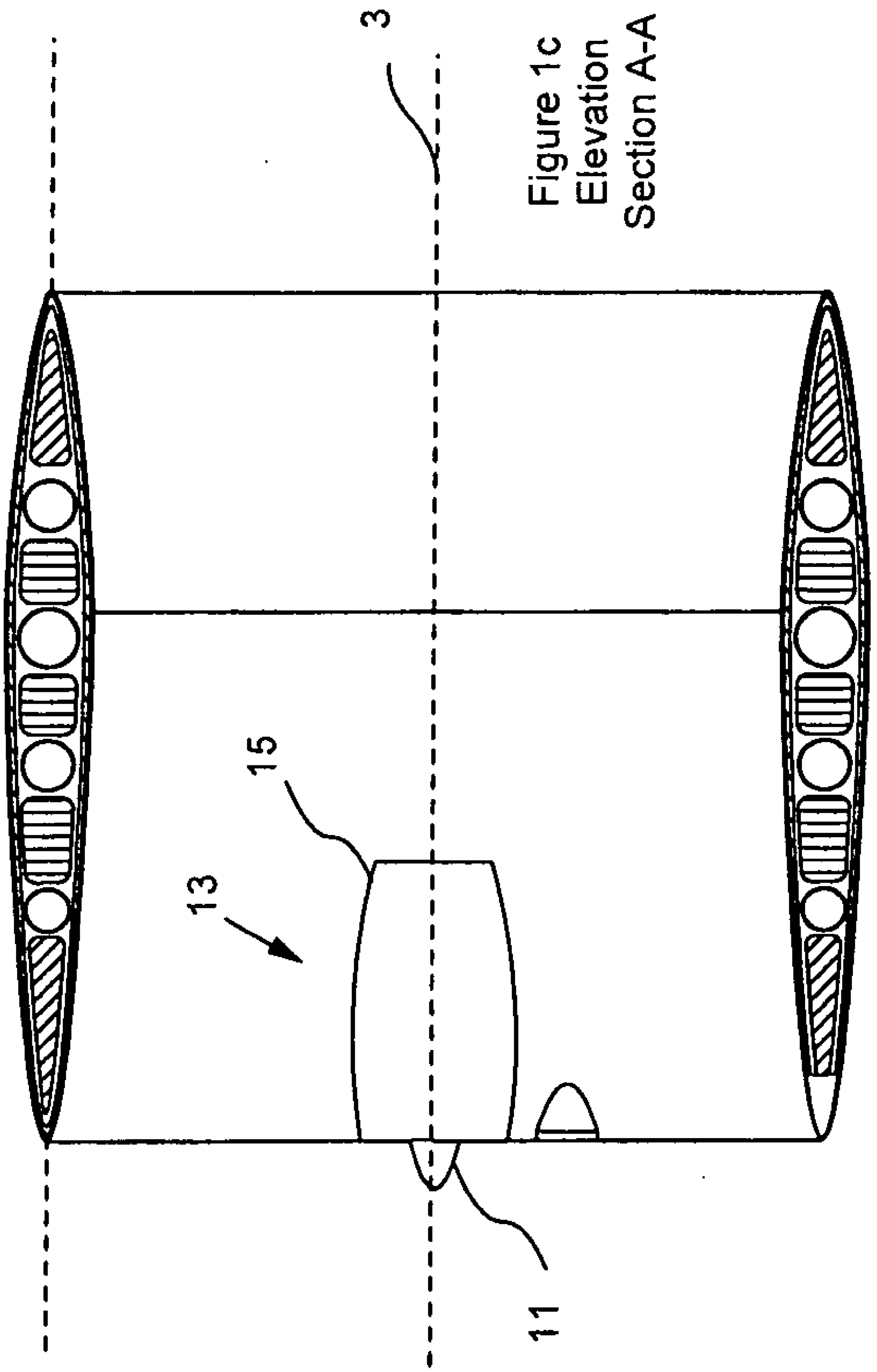

Figure 1c
Elevation
Section A-A

Figure 2a
Elevation
Section B-B

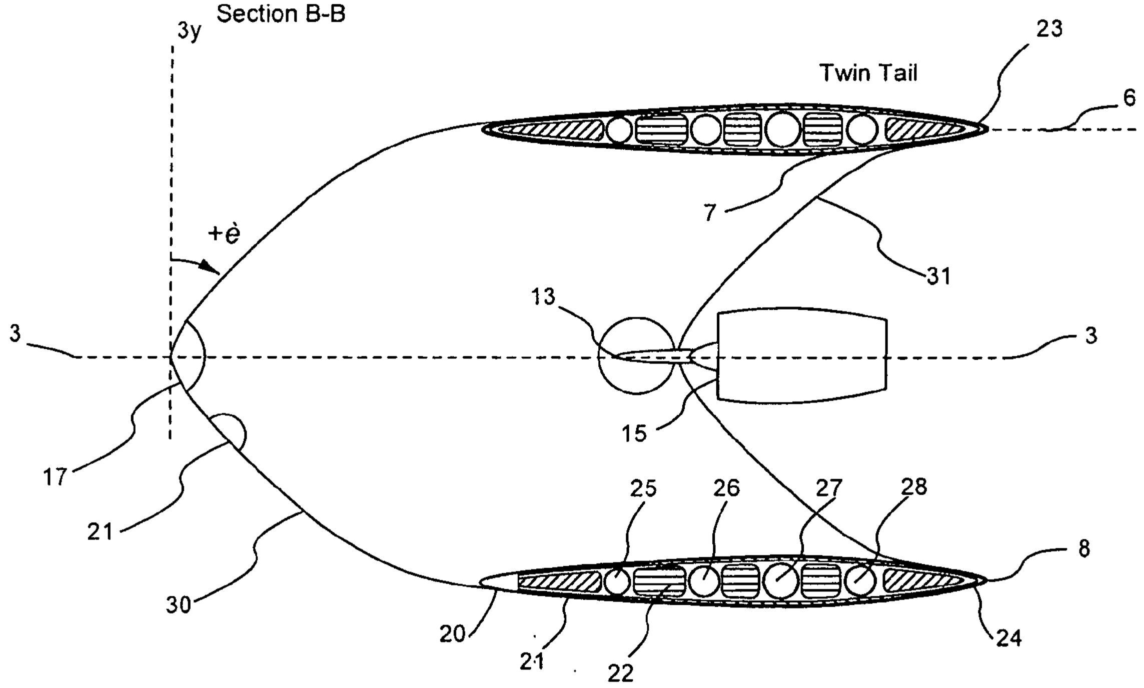
3y
Twin Tail
23
6
7
31
+è
13
3
3
17
15
21
25
26
27
28
8
30
20
21
22
24

Figure 2b
Plan

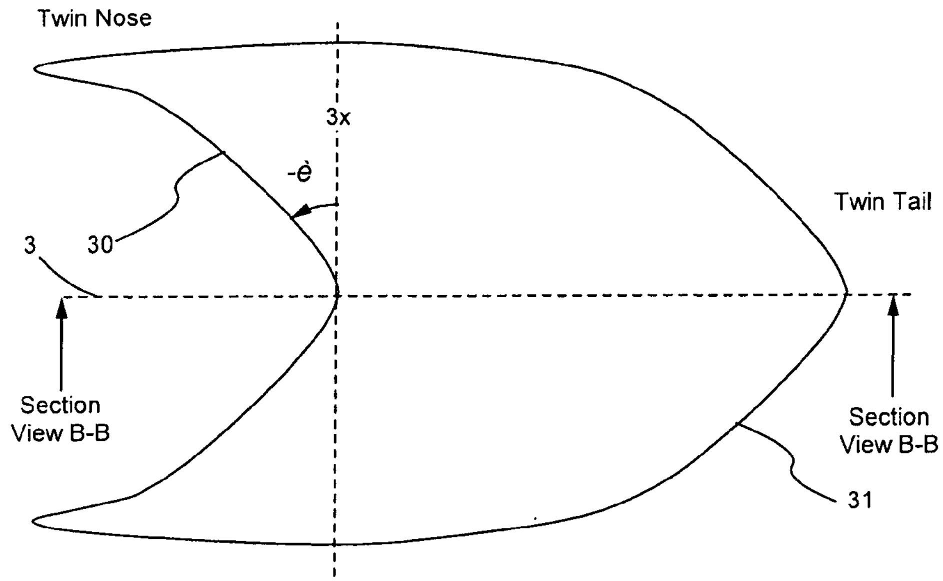
Twin Nose
3x
-è
Twin Tail
30
3
30
Section
View B-B
Section
View B-B
31

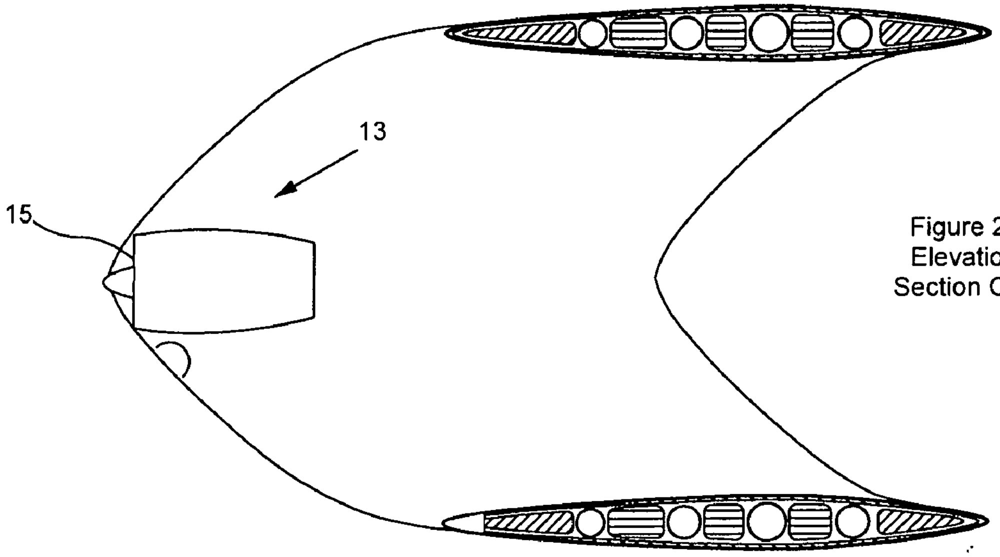

Figure 2c
Elevation
Section C-C

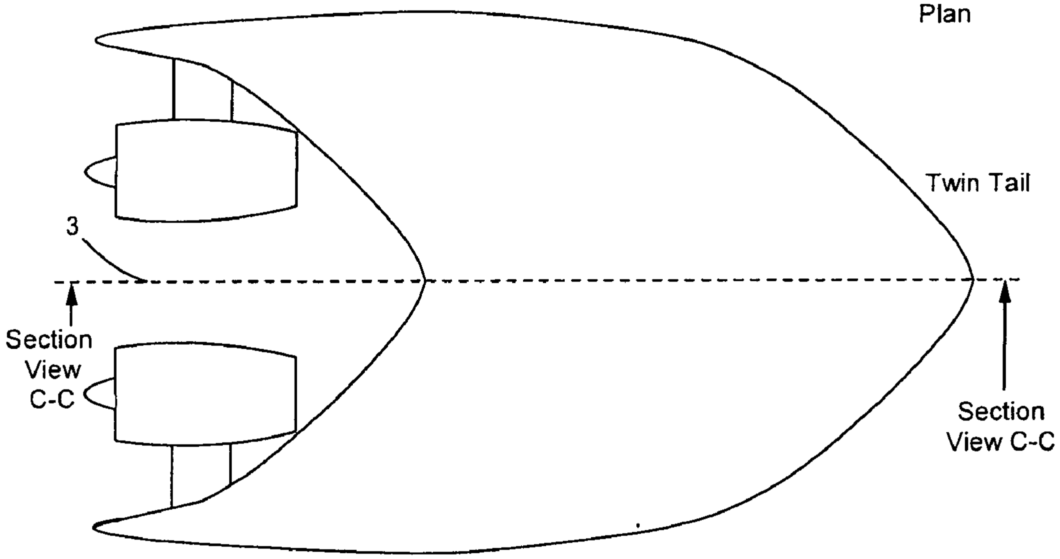
Twin Nose
Twin Tail
3
Section View C-C
Section View C-C

Figure 2d
Plan

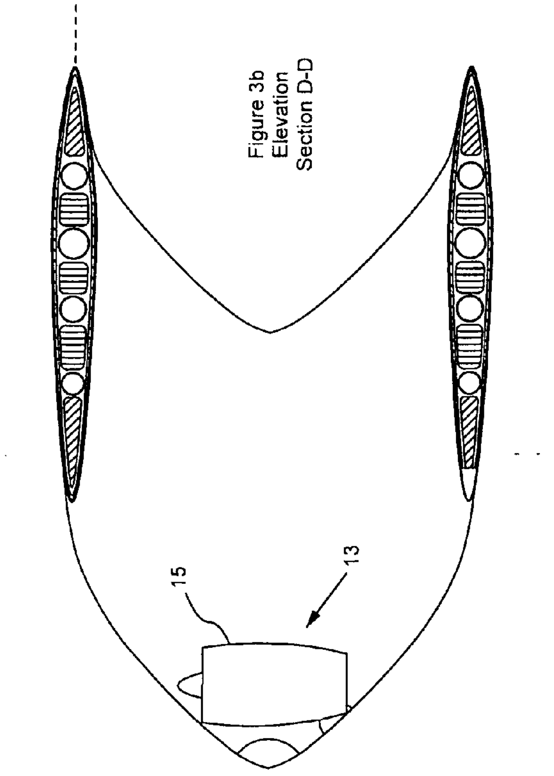

Figure 3b
Elevation
Section D-D

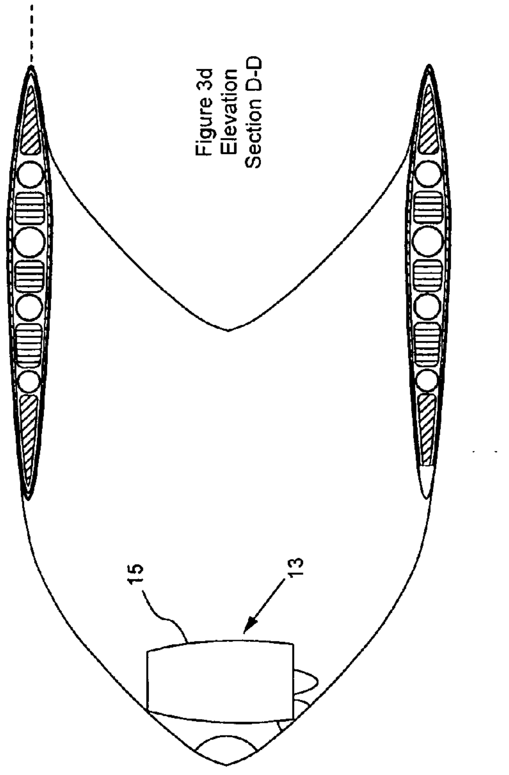

Figure 3d
Elevation
Section D-D

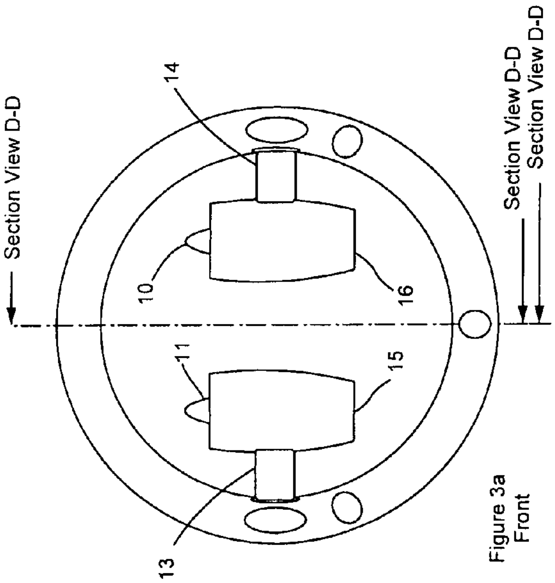
Section View D-D
14
10
16
11
15
13
Section View D-D

Figure 3a
Front

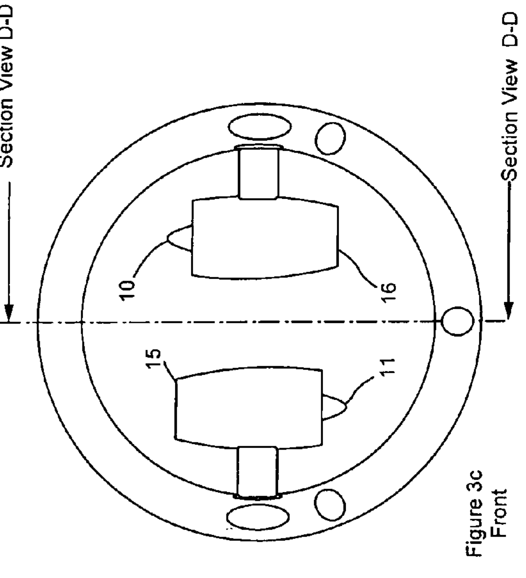
Section View D-D
10
16
15
11
Section View D-D

Figure 3c
Front

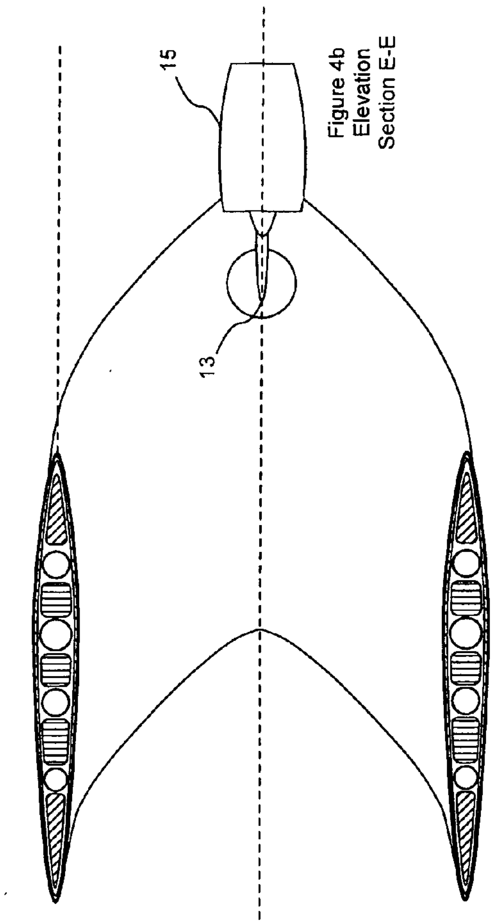

Figure 4b
Elevation
Section E-E

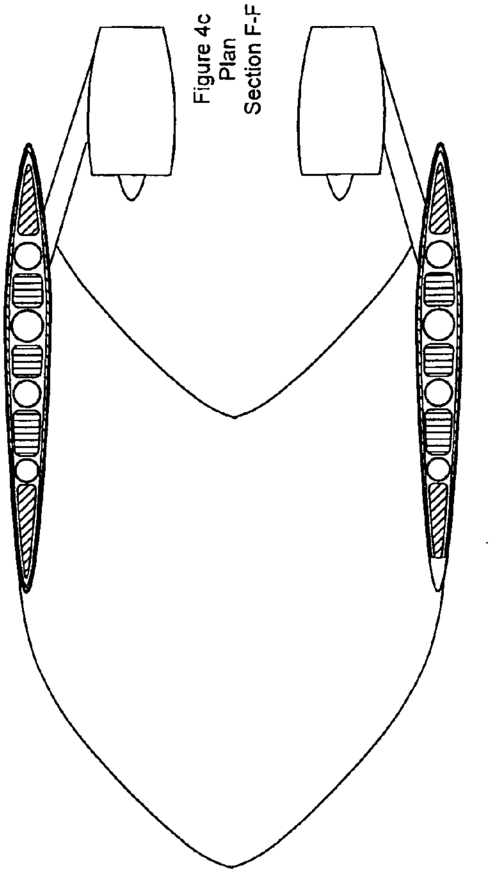

Figure 4c
Plan
Section F-F

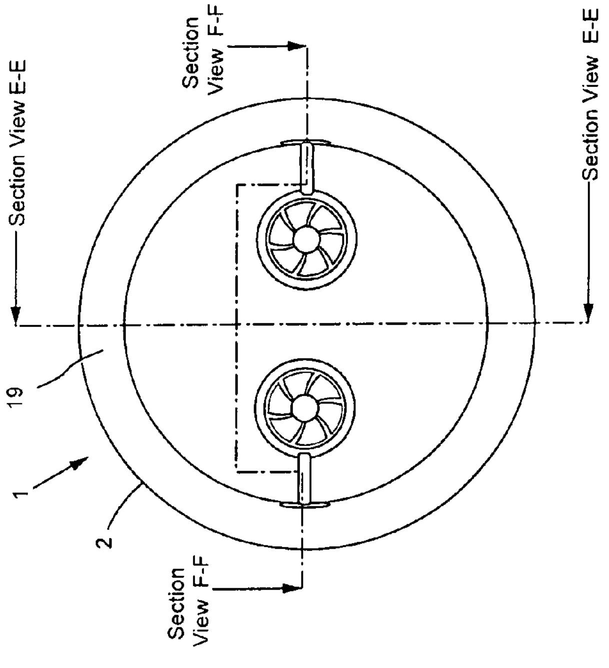
Section View E-E
Section View F-F
Section View E-E
Section View F-F
1
19
2

Figure 4a
Front

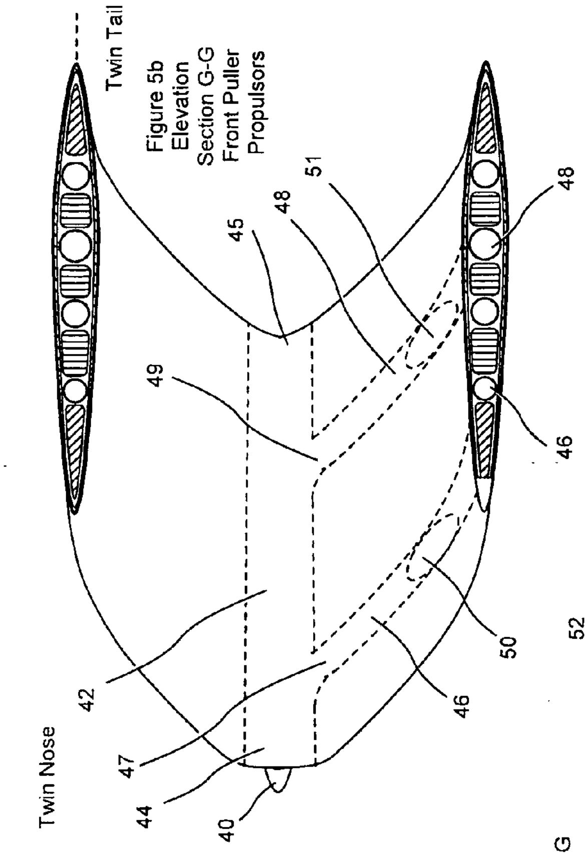
Twin Tail
Twin Nose
40
44
47
42
46
50
49
45
48
51
48
46

Figure 5b
Elevation
Section G-G
Front Puller
Propulsors

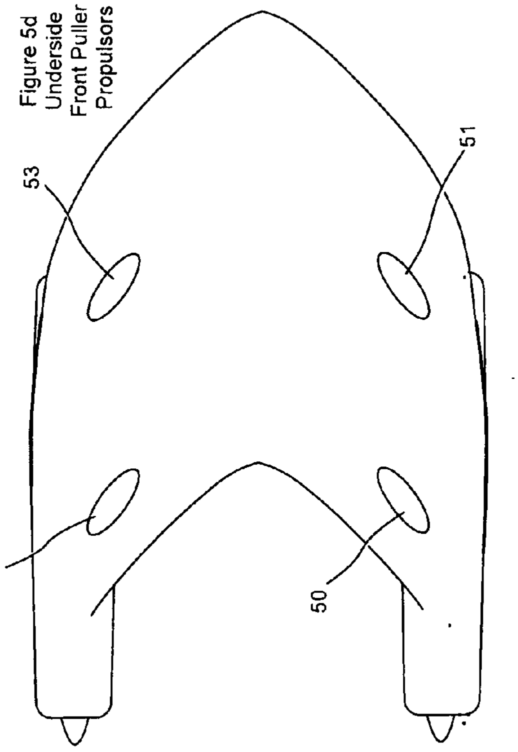

Figure 5d
Underside
Front Puller
Propulsors

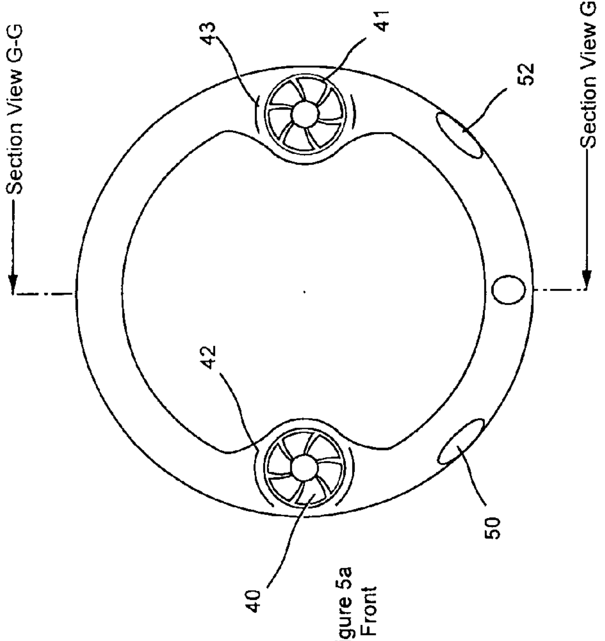
Section View G-G
Section View G-G
43
41
52
42
40
50

Figure 5a
Front

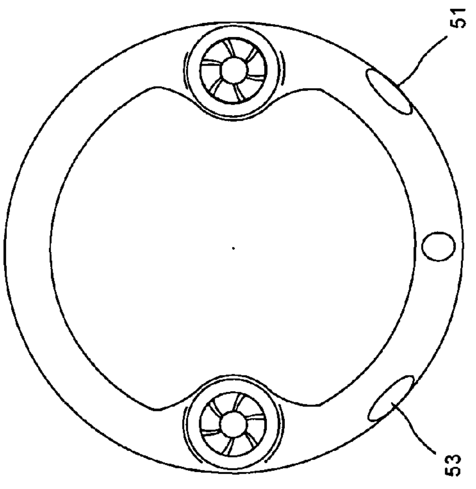

Figure 5c
Rear

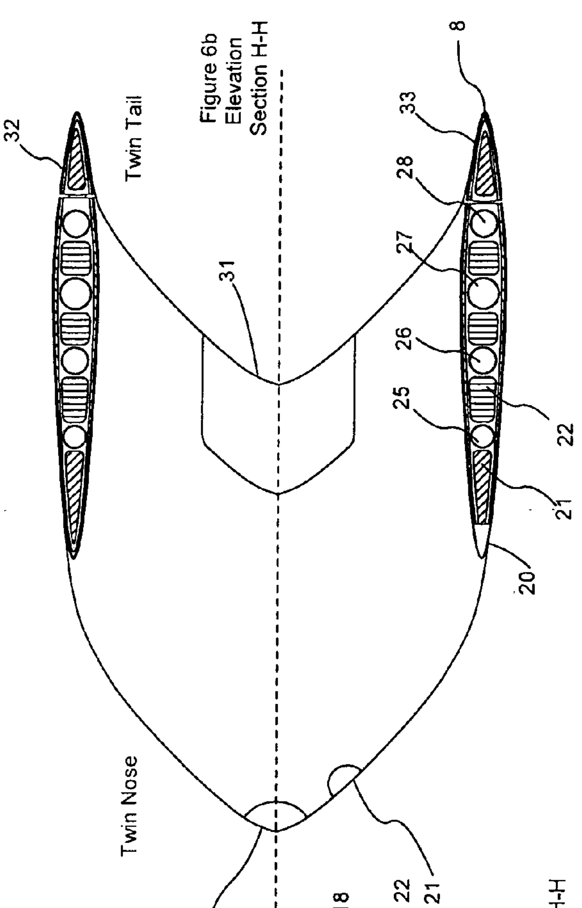
Twin Tail
32
8
33
28
27
31
26
25
22
21
20
Twin Nose
14
17
21

Figure 6b
Elevation
Section H-H

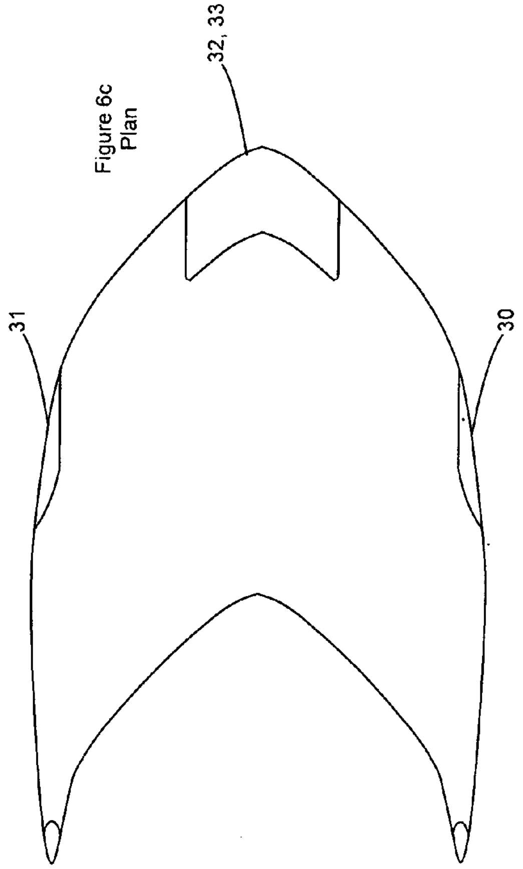

Figure 6c
Plan

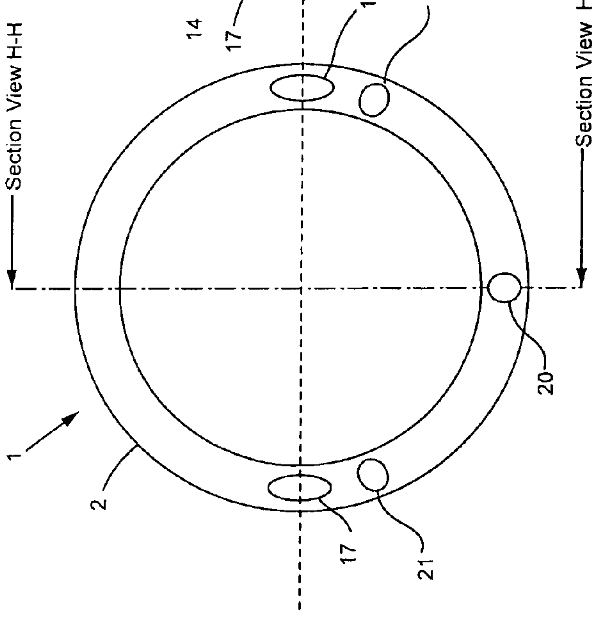
Section View H-H
Section View H-H
18
22
1
2
17
21
20

Figure 6a
Front

Front View
Figure 7a

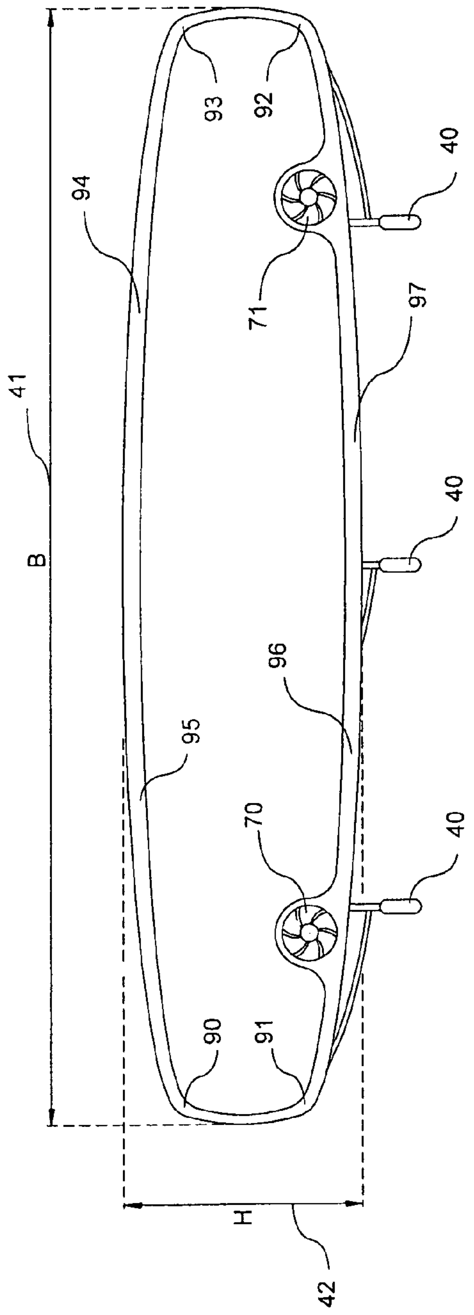
41
B
94
93
92
71
40
97
40
96
95
70
40
90
91
H
42

Elevation View
Figure 7b

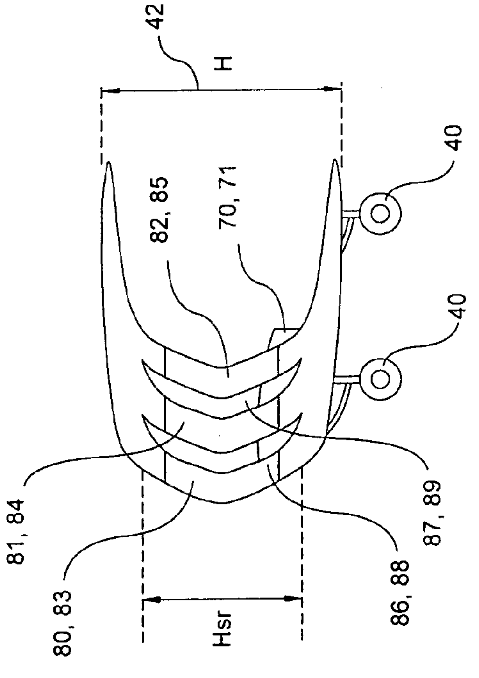
42
H
40
82, 85
70, 71
40
81, 84
87, 89
80, 83
86, 88
Hsr

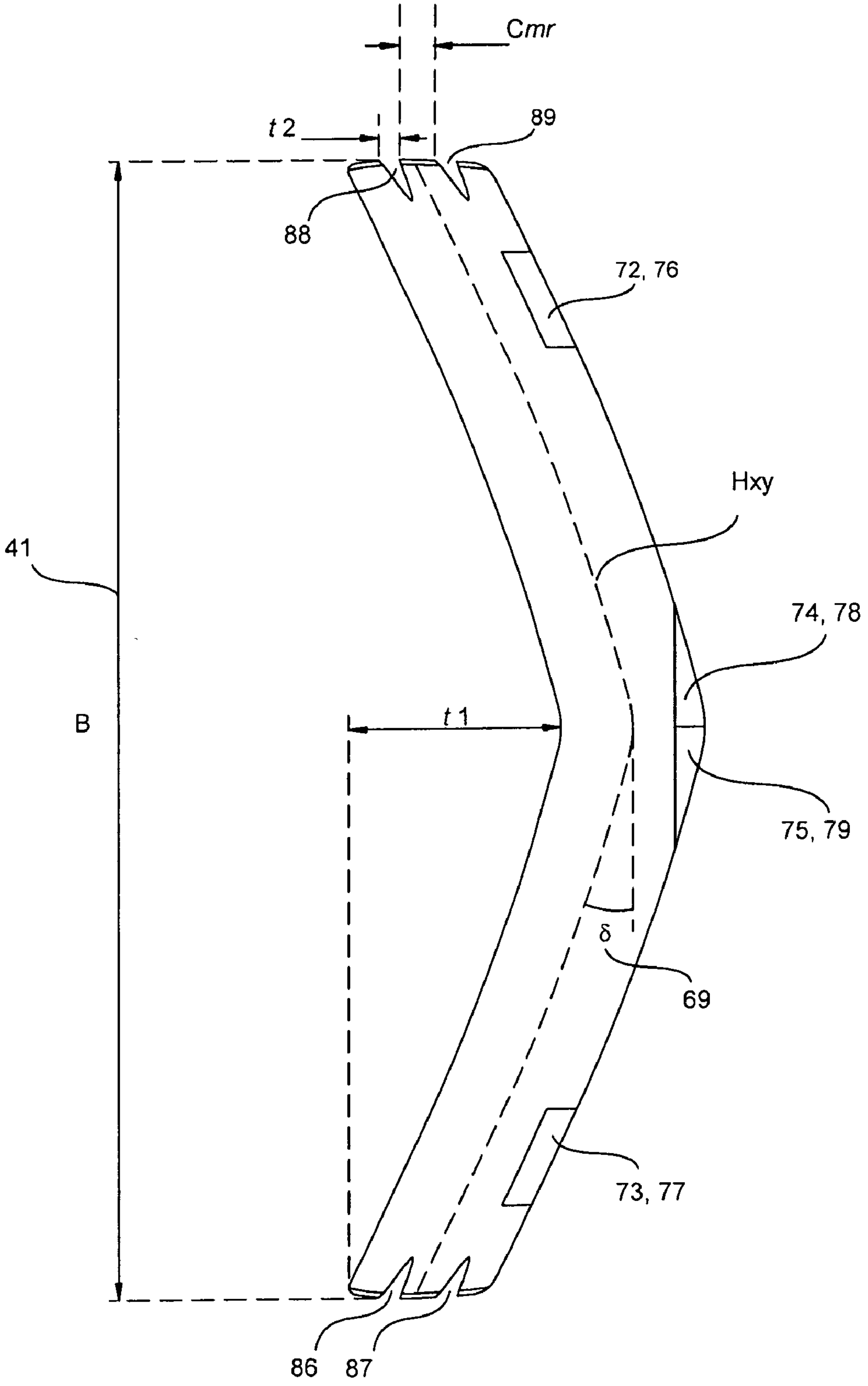
Cmr
t 2
89
88
72, 76
Hxy
41
74, 78
B
t 1
75, 79
δ
69
73, 77
86
87

Plan View
Figure 7c

ANNULAR AIRBORNE VEHICLE

[0001] The present invention relates to an annular airborne vehicle—that is, a vehicle having an outer wing-body which defines a wing-body axis and appears substantially annular when viewed along the wing-body axis, the interior of the annulus defining a duct which is open at both ends.

[0002] It is well understood that aircraft should be designed for minimal drag and mass characteristics, while maximising lift and aeroelastic characteristics so that fundamental aircraft design parameters may be adapted to achieve specific goals including better lift to drag ratios, better fuel efficiency, longer endurance and higher payload capacities, for example.

[0003] These fundamental design goals remain true whether the aircraft might be a miniature unmanned air vehicle (UAV), or a glider, or a passenger aircraft, or indeed a large transport aircraft.

[0004] It is an objective in many UAV applications to provide means for short take off and landing (STOL) or vertical take off and landing (VTOL), so that devices may be deployed and recovered without difficulty when runways may not be available. It is also desirable for UAVs to include means to hover, where surveillance, localisation, or communications activities may be required with little or no aircraft motion relative to the ground. Under such scenarios it is also desirable that the UAV should be capable of transition into an efficient forward flight phase, and vice versa, so that vehicle endurance would not necessarily be compromised by flight inefficiencies during VTOL, STOL, or hover phases. Under some circumstances it is advantageous to deploy such UAVs not only from ground but also from aircraft which may already be in flight, or land vehicles which may themselves be in motion, and therefore such UAVs would require special capabilities in order to withstand the demanding conditions imposed by such deployment envelopes.

[0005] Large UAVs are used in high altitude, long endurance scenarios where reconnaissance, or surveillance or imaging or communications functions may be required in order to achieve mission goals. Under such scenarios the endurance characteristics of the UAV become more important than take off or landing aspects, where lift to drag ratios must be high, mass must be low, and strength to weight ratios must also be high. It follows from aerodynamic principles that induced drag may be minimised by increasing the aspect ratio (AR) of the wing, where:

$$AR=B^2/S$$

(B=span; S=projected planform area)

[0006] which produces long slender wings where maximal efficiency must be balanced with aeroelastic constraints where wing structures must also survive other stresses induced during severe weather or take off or landing. Therefore such UAVs require improved vehicle forms in order to deliver better performance in terms of endurance, or range, or fuel efficiency, or payload capacity.

[0007] Therefore it is an object of this invention to disclose a common annular air vehicle form that may be utilised efficiently across a variety of UAV applications including those based upon miniature, agile vehicles, and large high altitude long endurance vehicles, and gliders.

[0008] It is also another object of this invention to disclose a common annular air vehicle form that may be utilised efficiently across a variety of manned aircraft applications including those based upon glider, light transit, heavy transit and passenger aircraft.

[0009] A first aspect of the present invention provides an airborne vehicle having a wing-body which defines a wing-body axis and appears substantially annular when viewed along the wing-body axis, the interior of the annulus defining a duct which is open at both ends; and a propulsion system comprising one or more pairs of propulsion devices, each pair comprising a first propulsion device mounted to the wing-body and positioned on a first side of a plane including the wing-body axis, and a second propulsion device which is mounted to the wing-body, positioned on a second side of the plane including the wing-body axis, and operable independently of the first propulsion device.

[0010] Preferably a direction of thrust of the first propulsion device can be adjusted independently of the direction of thrust of the second propulsion device, for instance by rotating the propulsion device. Alternatively each propulsion device comprises a thrust generator and a plurality of ducts arranged to receive propulsion gas from the thrust generator, and the direction of thrust of each propulsion device can be adjusted independently of the direction of thrust of the other propulsion device by adjusting the flow of propulsion gas in the ducts. In this case typically each duct is contained within the wing-body and at least some of the ducts have an inlet and an outlet in the wing-body.

[0011] In certain embodiments a thrust vector of each propulsion device can be adjusted between a first configuration in which the thrust vectors are co-directed and a second configuration in which the thrust vectors are contra-directed.

[0012] In one embodiment of the first aspect of the invention a magnitude of thrust of the first propulsion device can be adjusted independently of the magnitude of thrust of the second propulsion device, but not the direction of thrust.

[0013] Thus in general terms the propulsion devices are operable independently in the sense that either the direction of thrust of the first propulsion device can be adjusted independently of the direction of thrust of the second propulsion device; or a magnitude of thrust of the first propulsion device can be adjusted independently of the magnitude of thrust of the second propulsion device; or both.

[0014] Preferably a controller device is provided which is configured to independently operate the propulsion devices by issuing respective control signals to the propulsion devices.

[0015] A second aspect of the invention provides an airborne vehicle having a wing-body which defines a wing-body axis and appears substantially annular when viewed along the wing-body axis, the interior of the annulus defining a duct which is open at both ends, wherein the wing-body appears swept forward when viewed from a first viewing angle, and swept backward when viewed from a second viewing position at right angles to the first viewing angle.

[0016] In certain embodiments of the invention, the double-swept configuration of the second aspect of the invention provides several advantages:

- **[0017]** it enables the center of gravity of the annular vehicle to be more easily separated from its center of pressure, and therefore provides for better static and dynamic margins in pitch stability;
- **[0018]** aeroelastic forces which would otherwise subject conventional planar forwardly swept wings to excessive

tensile or flutter loads are effectively braced and dampened by the resilient annular structure of the wing-body;

[0019] the element of forward sweep makes the vehicle more tolerant of relatively high angles of attack, with lower susceptibility to stall, which is important when executing landings, take-off or other manoeuvres;

[0020] the element of forward sweep improves lift over drag ratios under some circumstances.

[0021] Preferably the wing-body has a leading edge with two or more noses, and a trailing edge with two or more tails which may be rotationally offset in relation to the noses (for instance by 90 degrees).

[0022] Typically at least part of the leading and/or trailing edge of the wing-body follows a substantially helical curve.

[0023] A third aspect of the invention provides an airborne vehicle having a wing-body which defines a wing-body axis and appears substantially annular when viewed along the wing-body axis, the interior of the annulus defining a duct which is open at both ends, wherein the wing-body carries at least one rudder on its left side and at least one rudder its right side.

[0024] In some embodiments the wing body carries two or more rudders on its left side and two or more rudders on its right side, and the wing-body is formed with a slot between each adjacent pair of rudders.

[0025] Preferably the wing-body has a projected planform area S, and a maximum outer diameter B normal to the wing-body axis, and wherein the ratio B^2/S is greater than 0.5. The relatively large diameter wing-body enables an array of two or more sensors to be well spaced apart on the wing-body, providing a large sensor baseline. In this way the effective acuity of the sensor array increases in proportion to the length of the sensor baseline. Also, the relatively high ratio B^2/S gives a high ratio of lift over drag, enabling the vehicle to be operated efficiently as a glider.

[0026] The duct may be fully closed along all or part of its length, or partially open with a slot running along its length. The duct may also include slots or ports to assist or modify its aerodynamic performance under certain performance conditions.

[0027] Various embodiments of the invention will now be described by way of example with reference to the accompanying drawings, in which:

[0028] FIG. **1***a* is a front view of a first propelled vehicle;

[0029] FIG. **1***b* is a cross-section of the right-hand side of the vehicle taken along the wing-body axis and along a line A-A in FIG. **1**;

[0030] FIG. **1***c* is a cross-section of the right-hand side of a second propelled vehicle, where its propellers are located within the forward half of the duct;

[0031] FIG. **2***a* is a cross-section of the right-hand side of a third propelled vehicle taken along the wing-body axis and along a line B-B in FIG. **2***b*, where its propellers are located within the rearward half of the duct;

[0032] FIG. **2***b* is a plan view of the third propelled vehicle;

[0033] FIG. **2***c* is a cross-section of the right-hand side of a fourth propelled vehicle taken along the wing-body axis and along a line C-C in FIG. **2***d*, where its propellers are located within the forward half of the duct;

[0034] FIG. **2***d* is a plan view of the fourth propelled vehicle;

[0035] FIG. **3***a* is a front view of the fourth propelled vehicle with its propellers in an up thrust configuration;

[0036] FIG. **3***b* is a cross-section of the right-hand side of the fourth propelled vehicle taken along the wing-body axis and along a line D-D in FIG. **3***a;*

[0037] FIG. **3***c* is a front view of the fourth propelled vehicle with its propellers in a contra-directed spin thrust configuration;

[0038] FIG. **3***d* is a cross-section of the right-hand side of the fourth propelled vehicle taken along the wing-body axis and along a line D-D in FIG. **3***c;*

[0039] FIG. **4***a* is a front view of a fifth propelled vehicle with its propellers in a forward thrust configuration and located behind the rearward half of the duct;

[0040] FIG. **4***b* is a cross-section of the right-hand side of the fifth propelled vehicle taken along the wing-body axis and along a line E-E in FIG. **4***a;*

[0041] FIG. **4***c* is a cross-section plan view of the fifth propelled vehicle taken along a line F-F in FIG. **4***a;*

[0042] FIG. **5***a* is a front view of a sixth propelled vehicle with its propellers mounted conformally within the annular wing body;

[0043] FIG. **5***b* is a cross-section of the right-hand side of the sixth propelled vehicle taken along a line G-G in FIG. **5***a;*

[0044] FIG. **5***c* is a rear view of the vehicle of FIG. **5***a;*

[0045] FIG. **5***d* is an underside view of the vehicle of FIG. **5***a;*

[0046] FIG. **6***a* is a front view of a glider;

[0047] FIG. **6***b* is a cross-section of the right-hand side of the glider taken along the wing-body axis and along a line H-H in FIG. **6***a;*

[0048] FIG. **6***c* is a plan view of the glider;

[0049] FIG. **7***a* is a front view of a seventh propelled vehicle;

[0050] FIG. **7***b* is a side view of the vehicle of FIG. **7***a;* and

[0051] FIG. **7***c* is a plan view of the vehicle of FIG. **7***a*.

[0052] Referring to FIGS. **1***a* and **1***b*, an airborne vehicle **1** has an outer wing-body **2** which is evolved from a laminar flow aerofoil profile (shown in FIG. **1***b*) as a body of revolution around a wing-body axis **3**. Thus the outer wing-body **2** appears annular when viewed along the wing-body axis as shown in FIG. **1***a*. An inner wall **4** of the annulus defines a duct **5** which is open fore and aft so that air flows through the duct as the vehicle moves through the air, generating aerodynamic lift by differential fluid flow across upper and lower aerofoil surfaces, which in this example arises whenever the axis of the annulus maintains an angle of incidence to its trajectory.

[0053] As shown in FIG. **1***b*, the aerofoil profile tapers outwardly gradually from a narrow nose end **6** to a widest point **7**, then tapers inwardly more rapidly to a tail end **8**. In this particular embodiment the widest point **7** is positioned approximately two-thirds of the distance between the nose and tail ends. The particular aerofoil section as described has symmetrical upper and lower surfaces which in this example provides for low minimum drag under cruising conditions, and may be modified in variants of this and other vehicles so as to modify the coefficients of lift, drag and pitch moment to suit the particular mission envelope and associated fluid flow regime as determined by the relevant range of Reynolds numbers.

[0054] A pair of propulsors **9**, **10** are mounted symmetrically on opposite sides of a vertical plane containing the wing-body axis **3**. In this particular embodiment the propulsors are shown as ducted electric fan motors comprising propellers (or impellers) **11**, **12** which are mounted on L-shaped support shafts **13**, **14** which in turn are mounted to the wing-body in line with the widest point **7** as shown in FIG. **1***b*. The propellers are mounted within shrouds **15**, **16** in such a way that their efficiency is increased within a proportion of the motor operating envelope. Each I,-shaped shaft is pivotally mounted to the wing-body so that it can rotate independently of the other shaft by 360 degrees relative to the wing-body about an axis parallel to the pitch axis of the vehicle, thus providing thrust-vectored propulsion. Both the shroud and L-shaped shaft have an aerofoil section using a ratio between chord length and height similar to that described for the outer wing-body, where the optimal relationship between height and chord is once again determined by the relative fluid flow regime as described by the relevant Reynolds number for these elements. Thus for example the propulsors **9**, **10** can be rotated between the co-directed configuration shown in FIGS. **1***a* and **1***b*, in which they provide a thrust force to propel the vehicle forward and along the wing-body axis, to a contra-directed configuration (not shown) in which they cause the vehicle to roll continuously around the wing-body axis.

[0055] The vehicle uses four independently operable motors (not shown) within its propulsion system: two brushless DC electric motors to drive the propellers, and two DC electric brushless motors to drive the L-shaped support shafts upon which the propeller motors are mounted, where a mechanical worm drive gear reduction mechanism is used to transfer drive and loads between the motor and the L-shaped shafts. Alternative motor types such as stepper motors may be used for the latter scheme, so long as operating loads are consistent with the rating of the motors. Similarly alternative motor types may be adopted for the propulsors, including internal combustion, gas turbine or solid propellant motors.

[0056] The thrust-vectored propulsors provide the means for motion along the wing-body axis, either forward or in reverse, and spin or roll around the wing-body axis, and pitch or yaw about the vehicle's centre of gravity (CofG), which in this embodiment is located below the wing-body axis, above the lower annulus section, within a vertical plane that contains the wing-body axis, and approximately at 15 to 23% of the chord length when measured from the nose. As described earlier it is clear that the two propulsors may be contra-directed in order to induce vehicle roll. The two propulsors may also be co-directed. For instance when both are directed down so that their thrust vectors lies above the CofG, then the vehicle will pitch nose down. Similarly when the two propulsors are directed up so that their thrust vector lies below the CofG, then the vehicle will pitch nose up. It is also clear that varying degrees of propulsor pitch in relation to the vehicle and each other may be used to achieve vehicle pitch, roll and yaw. Yaw may also be induced by differential thrust application when differential propeller revolution rates are adopted. Thus it can be seen that the vehicle is able to dive, turn, roll and climb under its own autonomous control, since the vehicle also includes a controller device that includes the means to measure linear accelerations in three orthogonal axes, and angular accelerations in three orthogonal axes, and processing methods to calculate the control demands that would be delivered to the propulsion system.

[0057] Thus the direction of thrust of the propulsor **9** can be adjusted by the controller device independently of the direction of thrust of the propulsor **10** by issuing control signals to the electric motors which drive the L-shaped support shafts. Also, the magnitude of thrust of the propulsor **9** can be adjusted by the controller device independently of the magnitude of thrust of the propulsor **10** by issuing control signals to the electric motors which drive the the propellers.

[0058] The vehicle has a high degree of manoeuvrability, since its thrust vectored propulsion may be arranged for high turn rates under dynamic control. It is also clear that the vehicle has a high degree of stability. In the first instance when motion is along the axis of the wing-body then relatively high speeds may be achieved with contra-rotating propellers that cancel induced torque, while contra-directed propulsors provide for further roll stability. In the second instance when spin motion around the wing-body axis is induced, then angular momentum is increased and once again the stability of the vehicle is increased, where this may be measured as a reduction in vehicle attitude or position errors when subject to external forces.

[0059] A significant advantage offered by this propulsion system is the provision of effective attitude control at slow speeds, when alternative control surfaces would be much less effective. It follows therefore that large thrust vector demands may be used effectively at slow speeds, and during launch or recovery, or take off and landing, while low thrust vector demands would satisfy the majority of performance envelope demands when traveling at speed. Another significant advantage of this arrangement is the relative ease with which the vehicle may transition from launch to forward transit, or to hover, or VTOL/STOL, and vice versa. A further significant advantage of this arrangement is the low drag characteristic of the annular wing-body, which preferentially allows for high speed transit in normal forward flight.

[0060] The nose of the vehicle carries a pair of video cameras **17**,**18** for collision avoidance and imaging applications. The relatively large diameter of the wing-body enables the cameras to be well spaced apart, thus providing a long stereoscopic baseline that provides for accurate range estimation by measurement of parallax between objects located within both camera fields of view. An ultrasonic transmitter **20** and two receivers **21**, **22** are provided for acoustic imaging and sensing. Again, the wide baseline is an advantage since localization accuracy will be improved by differential time of flight processing between the separated receivers. The outer wing-body **2** contains an interior space which can be seen in FIG. **1***a*. This outer wing-body is preferentially manufactured from a stiff composite material using glass or carbon fibre filaments laminated alternately between layers of epoxy resin. Alternatively a cheaper, less resilient wing-body may be moulded from a suitable hard polymer such as polyurethane or high density polyethylene. Under certain conditions where mass reduction is paramount then the outer wing-body skin may be formed instead from a tough polymer film such as mylar. It is also possible to manufacture the outer wing-body from aluminium, should the wing-body be pressurised. The interior space may be pressurized, and houses a pair of battery packs **21**, **22**, a pair of tail sensors **23**, **24**, and four toroidal vessels **25**-**28** spaced apart along the wing-body axis. The vessels may contain the vehicle electronics, some propulsion sub-system elements and other items, and are joined by axial or tri-lateral struts (not shown) which would share the construction methods adopted for the toroidal vessels. In this particular embodiment the toroidal vessels are preferentially manufactured from stiff composites using either glass or carbon fibre filaments wound helically around the toroid and alternately laminated between layers of epoxy resin. Alternatively the toroidal vessels may be manufactured from a suitable grade of metal such as aluminium, stainless or galvanized steel, or titanium. The toroidal and annular structures disclosed herein are designed to provide for superior aeroelastic resilience around the wing-body so that mass may effectively and safely be reduced without risk of major stress loading or concentration at any section of the annulus.

[0061] The aspect ratio (AR) of the annular wing-body is defined as follows:

$$AR=2B^2/S$$

[0062] where B is the span of the wing-body (defined by the maximum outer diameter of the wing-body) and where S is the projected planform area of the wing-body.

[0063] In the vehicle of FIG. **1***a,* the AR is approximately 2, although this number may be modified in other embodiments where the application may demand other ratios. It is evident that the vehicle form may be adjusted by simple variation of its toroidal diameter to reflect narrow vehicles where aspect ratio is low, or to reflect broad vehicles where aspect ratio is high. In either case specific advantages may be gained under certain circumstances, since relatively high coefficients of lift may be achieved using a toroidal form with low aspect ratio, while optimal glide slope ratios, or equivalent ratios of lift over drag may be achieved using a toroidal form with high aspect ratio.

[0064] In the embodiment of FIGS. **1***a* and **1***b,* the propulsion devices are mounted towards the tail of the vehicle—that is towards the right-hand side of FIG. **1***b.* In the second embodiment shown in FIG. **1***c* the propulsion devices are mounted towards the nose of the vehicle—that is towards the left-hand side of FIG. **1***c.* The front view of the vehicle of FIG. **1***c* is identical to that of the vehicle of FIG. **1***b.*

[0065] A swept-wing vehicle is shown in FIGS. **2***a* and **2***b.* The front view of the vehicle of FIGS. **2***a* and **2***b* is identical to that of the vehicle of FIG. **1***b.*

[0066] The leading edge **30** of the wing-body of the vehicle describes a form of helical curve around one quadrant of the circumference of the wing body, where such a helix subtends an angle +θ with a vertical line $\mathbf{3}_y$ perpendicular to the wing-body axis **3**, when viewed in elevation as shown in FIG. **2***a.* This helical curve is used identically around the third quadrant of the circumference of the wing-body leading edge, while its mirror image is used to form the second and third quadrants of the leading edge. The four helical elements are joined, so that the leading edge forms a closed double chevron sweep around the forward circumference of the wing-body. Within this particular embodiment, the trailing edge forms an identical closed double chevron sweep which is translated along the axis **3** by the chord length of the aerofoil. In other embodiments the relationship between the leading and trailing double chevron sweeps may be modified, for example to vary the relationship between orthogonal lift and rudder stabilization surfaces.

[0067] Similarly the trailing edge **31** of the wing-body of the vehicle subtends a positive angle (not labeled) with the line $\mathbf{3}_y$. Thus, because the mid-chord line of the wing-body (that is, a line at the mid-point between the leading and trailing edges) subtends a positive angle with the line $\mathbf{3}_y$ on both the upper and lower sides of the wing-body, the wing-body appears swept backward when viewed in elevation. In contrast, the leading edge **30** of the wing-body of the vehicle subtends an angle −θ with a horizontal line $\mathbf{3}_x$ perpendicular to the wing-body axis **3**, when viewed in plan as shown in FIG. **2***b.* Similarly the trailing edge **31** of the wing-body of the vehicle subtends a negative angle (not labeled) with the line $\mathbf{3}_x$. Thus, because the mid-chord line of the wing-body (that is, a line at the mid-point between the leading and trailing edges) subtends a negative angle with the line $\mathbf{3}_x$ on both the left and right-hand sides of the wing body, the wing-body appears swept forward when viewed in plan.

[0068] The sweep angle δ is described in FIG. **7***c* (and also labelled with numeral **69**) under an alternative embodiment and may vary between 0 to 60°, and 0 to −60°, where δ is defined as the angle subtended by the helix H_{xy} (i.e, the mid-chord line of the wing-body) and any plane which lies normal to the axis of the annular wing-body, and where H_{xy} is transcribed by the mid-chord line around the annular wing-body, and where δ is constrained by the axial displacement t**1** between the two chords whose tangential planes are vertical (i.e, left and right sides) in relation to the chords whose tangential planes are horizontal (i.e top and bottom sections). In this example and other embodiments the mean value of δ may be determined within each of four quadrants bounded by two orthogonal planes that coincide with the wing body axis **3**, at which boundaries δ becomes zero. The axial displacement t**1** is greater than or equal to 0.1 times the chord of the annulus. These relationships are described again in FIG. **7***c.*

[0069] In the embodiment of FIGS. **2***a* and **2***b,* the propulsion devices are mounted towards the tail of the vehicle—that is towards the right-hand side of FIGS. **2***a* and **2***b.* In the embodiment shown in FIGS. **2***c* and **2***d,* the propulsion devices are mounted towards the nose of the vehicle—that is towards the left-hand side of FIGS. **2***c* and **2***d.* The front view of the vehicle of FIGS. **2***c* and **2***d* is identical to that of the vehicle of FIG. **1***b.*

[0070] FIGS. **3***a* and **3***b* illustrate the vehicle of FIGS. **2***c* and **2***d* with its propellers in an up thrust configuration, and FIGS. **3***c* and **3***d* illustrate the vehicle of FIGS. **2***c* and **2***d* with its propellers in a contra-directed spin thrust configuration.

[0071] FIGS. **4***a*-**4***c* show a variant of the vehicle of FIGS. **2***a* and **2***b.* Whereas the wing-body in the vehicle of FIGS. **2***a* and **2***b* is swept forward when viewed in plan, the wing-body in FIGS. **4***a*-**4***c* is swept backward when viewed in plan. In this configuration the static and dynamic pitch stability margins are reduced by comparison with the wing-body scheme described in FIG. **2**, however the propulsors have greater degrees of freedom in thrust vector control with lower interference to lifting surface fluid flows. As a consequence this scheme offers further improvements in vehicle agility.

[0072] FIG. **5***a* is a front view of a sixth propelled vehicle, employing a conformal propulsion system. A starboard propeller (or impeller) **40** is mounted in the inlet port **44** of a starboard main duct **42** which runs along the starboard side of the hull to a main outlet port **45**, and a port propeller **41** (or impeller) is mounted in the inlet port of a port main duct **43** which runs along the port side of the hull to a main outlet port (not shown).

[0073] The ducting network of the propulsion system is symmetrical, so only the starboard elements will be described in detail with reference to FIG. **5***b.* The starboard main duct **42** is coupled to a forward duct **46** at an inlet **47**, and to a rear duct **48** at an inlet **49**. The forward duct **46** has port and starboard underside outlet ports **52**, **50** in the underside of the hull, and the rear duct **48** has port and starboard underside outlet ports **53**, **51** in the underside of the hull.

[0074] The port main duct **43** is also coupled to the ducts **46**,**48** via inlets (not shown) on the port side of the vehicle, similar to the inlets **47**,**49**.

[0075] Note that all of the ducts described above are contained within the annular wing-body.

[0076] Valves (such as butterfly valves or sliding plate valves) are provided in the inlets **47**, **49** (and the equivalent inlets on the port side of the vehicle), in the underside outlet ports **50-53**, in the main outlet port **45** (and the equivalent main outlet port on the port side of the vehicle), and in the forward and rear ducts **46**, **48** between the port and starboard underside outlet ports. The valves may be opened and closed digitally by pulse width modulation control.

[0077] The valves can be operated independently to adjust the flow of propulsion gas in the ducts, and thus adjust the magnitude and direction of the thrust generated by the port and starboard propulsion systems. For example forward thrust can be achieved by closing the valves in the inlets **47**, **49** (and the equivalent inlets on the port side of the vehicle), and opening the valves in the port and starboard main outlet ports. VTOL, STOL, or hover phases can be achieved by closing (fully or partially) the valves in the port and starboard main outlet ports, and opening all other valves.

[0078] The underside outlet ports **50-53** are arranged to enclose a vertical axis that locates the centre of gravity (CofG) of the vehicle within a horizontal frame of reference. Therefore during VTOL, STOL or hover phases the propellers, motors, ducts, and valves may be controlled by a suitable device so that thrust may be adjusted between the four ports **50-53** in such a way that vehicle attitude and resultant vehicle accelerations and velocities may be precisely controlled.

[0079] This controller device is described above for earlier variants of this annular wing-body. Once again the embodiment is particularly described with ducted electric fan motors, although these may be replaced by suitable alternative motors including internal combustion, gas turbine, or solid propellant motors.

[0080] A glider vehicle is shown in FIGS. **6***a-c*. The wing-body of the vehicle has an annular construction as shown in FIG. **6***a*, and adopts a forwardly swept shape similar to the wing-body of FIGS. **2***a* and **2***b*. The wing-body uses similar construction, and houses various sensors, battery packs, and toroidal vessels in common with the vehicles shown in FIGS. **1-4**.

[0081] Because the glider has no propulsion devices, it has a fully conformal outer shape with no superstructure either inside the duct or projecting from the exterior of the vehicle, other than its two elevators **32**, **33** and two rudders **30**, **31** which are conformally and orthogonally mounted around the rear of the wing-body. The elevators and rudders are secured to the wing-body by linkage mechanisms that provide for rotation within arcs of ±30°. The elevators and rudders are controlled by worm drive mechanisms as disclosed for propulsor controls in FIGS. **1** to **5**.

[0082] FIGS. **7***a*-**7***c* show a propelled vehicle according to a further embodiment of the invention. As described for vehicles disclosed in FIGS. **2** to **6**, this particular embodiment of the annular wing-body also includes wing sweep, and in particular a symmetrical double chevron sweep as explained for the vehicles described in FIG. **2**, only in this embodiment the sweep angle δ **69** is approximately 21 degrees.

[0083] The vehicle has a twin propulsion system **70**, **71** similar to that shown in FIGS. **2***a*/**2***b* although in this particular embodiment the motors are mounted on the lower half of the annulus. Also, in contrast to the propulsion system of FIGS. **2***a*/**2***b* (in which the direction and magnitude of thrust of each propulsion device can be adjusted independently of the direction and magnitude of thrust of the other propulsion device) vector control actuation elements are not included in the embodiment of FIGS. **7***a*-**7***c* so in this case only the magnitude of thrust of each propulsion device can be adjusted independently of the magnitude of thrust of the other propulsion device, and not the direction of thrust (other than by switching between forward and reverse modes). Note also that in a further embodiment (not shown) the propulsion system of FIGS. **2***a*/**2***b* may be fitted to the vehicle of FIGS. **7***a*-**7***c*.

[0084] The position of the motors around the annulus may be adjusted symmetrically about the vertical plane that contains the wing-body axis, where absolute thrust vectors would need to be aligned to best fit the range of flight regimes intended. Elevators **74**, **75** and ailerons **72**, **73** are included on the upper cambered wing-body section, while elevators **78**, **79** and ailerons **76**, **77** are included on the lower cambered wing-body section, and rudders **80** to **82** are included on the left side of the wing-body, while rudders **83** to **85** are included on the right side of the wing-body. The form of this annular vehicle is particularly suited to large, high endurance platforms which seek to minimize mass and maximize operational efficiency and where agility concerns would not be so prevalent. Once again the annular wing-body form seeks to maximize a broad range of performance criteria including resilience to aeroelastic loading which become more dominant when aspect ratios become large.

[0085] In contrast with the circular annulus shown in FIG. **1***a*, the annulus is wider than it is tall, as shown in the front view of FIG. **7***a*. In this embodiment the relationship between span B and chord is 8 to 1, while span B to wing-body height H is 5 to 1. The aspect ratio of this embodiment is derived from:

$$AR=2B^2/S$$

[0086] Where AR=16

[0087] This particular aspect ratio is suited to long endurance operations where efficient soaring will be required.

[0088] The wing-body is arranged as four cambered sections **94** to **97** joined by four curved elements **90** to **93** that provide both for structural bracing and conformal housing of twin motors, fuel, batteries, undercarriage, control systems, stowage bays and payload items. The construction of the wing-body follows the methods and materials described for the earlier wing-bodies of FIGS. **1** to **6**, where composite technologies would be used extensively within its structures. Similarly the motors may be based upon ducted electric fan, or internal combustion, or gas turbine technologies in either of push or pull configuration, augmented by hybrid energy sources including photovoltaic and fuel cell technologies.

[0089] The wing-body vehicle has three undercarriage elements **40** for providing stability during take-off and landing operations. Two elevator and two aileron elements **72** to **79** are included on each of the upper and lower sections of the annular wing-body, where the linkage mechanism that secures these elements is similar to that disclosed for the vehicle form described in FIG. **6***b*, while its control mechanism uses a worm drive as disclosed for vehicle forms disclosed in FIGS. **1** to **6**. A plurality of rudder elements **80** to **85** are arranged at either vertical side of the annulus, using a linkage mechanism similar again to that described in FIG. **6***b*, and a control mechanism based again upon a worm drive as disclosed for vehicle forms described in FIGS. **1** to **6**, together with flared slots **86** to **89** between each adjacent pair of rudders which provide for longitudinal separation τ_2 between each rudder element where the ratio S_r between the separation distance t**2** and the mid-chord length of the rudder element C_{mr} may vary between 0.5 and 5, such that:

$$S_r = t2/C_{mr}$$

[0090] and where the vertical height H_{sr} subtended by the slots may vary between 0.3 and 0.8 times the height H of the annulus. Each rudder element incorporates an appropriate aerofoil to ensure smooth aerodynamic flow along their discrete surfaces, where the precise aerofoil shape would be similar to that adopted for the other sections of the annular wing body, with adjustments as might be necessary for the variation in Reynolds number associated with rudder chord reduction.

[0091] In this particular embodiment three rudder elements **80** to **82** and **83** to **85** are shown on each of left and right sides of the annular wing-body, and two flared slots **86**, **87** and **88**, **89** are also shown between the rudder elements on each of left and right sides. Each rudder houses a secure linkage that bonds with upper and lower wing-body sections to ensure structural integrity, and to allow for rotation within a prescribed arc about a vertical axis, where the arc would normally be constrained to less than ±30°.

[0092] In this embodiment of the annular wing-body the combination of annulus, slots and multiple rudders serve to minimise induced drag, parasitic drag, mass and cross-flow effects for the whole of the annual wing-body vehicle while providing for stability and control in yaw. The double chevron sweep configuration also improves the soaring capability of the annular wing-body, where lift over drag ratios are increased by the effective forward sweep of upper and lower cambered lifting sections of the annular wing-body.

[0093] The high rotational symmetry of many of the wing-body shapes (as viewed along the wing-body axis) described herein gives other advantages where the vehicle is to be operated in a continuous or transient roll mode. However, the invention also covers alternative embodiments of the invention (not shown) including:

- **[0094]** embodiments in which the duct is divided into two or more separate ducts by suitable partitions
- **[0095]** embodiments in which the outer wing-body itself defines two or more separate ducts
- **[0096]** embodiments in which the outer wing-body is evolved from an aerofoil as a body of revolution around the wing-body axis by an angle less than 360 degrees. In this case, instead of being closed, the duct will be partially open with a slot running along its length. By making the angle greater than 180 degrees, and preferably close to 360 degrees, the wing-body will remain substantially annular so as to provide aerodynamic lift at any angle of roll.

[0097] The annular air vehicles described above achieve significant improvements in overall performance when measured across a broad range of whole vehicle criteria including lift to drag ratios, agility, stability, endurance, launch and recovery, take off or landing, power to weight ratios, fuel efficiency and resilience to aeroelastic loading.

1. An airborne vehicle having a wing-body which defines a wing-body axis and appears substantially annular when viewed along the wing-body axis, the interior of the annulus defining a duct which is open at both ends; and a propulsion system comprising one or more pairs of propulsion devices, each pair comprising a first propulsion device mounted to the wing-body and positioned on a first side of a plane including the wing-body axis, and a second propulsion device which is mounted to the wing-body, positioned on a second side of the plane including the wing-body axis, and operable independently of the first propulsion device, wherein a direction of thrust of the first propulsion device can be adjusted independently of the direction of thrust of the second propulsion device.

2. The vehicle of claim **1** wherein the direction of thrust of the first propulsion device can be adjusted independently of the direction of thrust of the second propulsion device by rotating the propulsion device.

3. The vehicle of claim **1** wherein each propulsion device comprises a thrust generator and a plurality of ducts arranged to receive propulsion gas from thrust generator, and wherein the direction of thrust of the first propulsion device can be adjusted independently of the direction of thrust of the second propulsion device by adjusting the flow of propulsion gas in the ducts.

4. The vehicle of claim **3** wherein each duct is contained within the wing-body.

5. The vehicle of claim **1** wherein a thrust vector of each propulsion device can be adjusted between a first configuration in which the thrust vectors are co-directed and a second configuration in which the thrust vectors are contra-directed.

6. The vehicle of claim **1** wherein at least part of the wing-body is swept with respect to the wing-body axis.

7. The vehicle of claim **6** wherein the wing-body appears swept forward when viewed from a first viewing angle, and swept backward when viewed from a second viewing position at right angles to the first viewing angle.

8. The vehicle of claim **7** wherein the outer wing-body appears swept forward when viewed in elevation, and swept backward when viewed in plan.

9. The vehicle of claim **1** wherein the first propulsion device is positioned on a first side of a vertical plane including the wing-body axis, and the second propulsion device is positioned on a second side of the vertical plane including the wing-body axis.

10. The vehicle of claim **1** further comprising a controller device which is configured to independently operate the propulsion devices by issuing respective control signals to the propulsion devices.

11. An airborne vehicle having a wing-body which defines a wing-body axis and appears substantially annular when viewed along the wing-body axis, the interior of the annulus defining a duct which is open at both ends, wherein the wing-body appears swept forward when viewed from a first viewing angle, and swept backward when viewed from a second viewing angle at right angles to the first viewing angle.

12. The vehicle of claim **11** wherein the wing-body appears swept forward when viewed in plan, and swept backward when viewed in elevation.

13. The vehicle of claim **11** wherein the wing-body has a leading edge with two or more noses, and a trailing edge with two or more tails.

14. The vehicle of claim **13** wherein the noses are rotationally offset in relation to the tails.

15. The vehicle of claim **11** wherein at least part of the leading and/or trailing edge of the wing-body follows a substantially helical curve.

16. An airborne vehicle having a wing-body which defines a wing-body axis and appears substantially annular when viewed along the wing-body axis, the interior of the annulus defining a duct which is open at both ends, wherein the wing-body carries at least one rudder on its left side and at least one rudder its right side.

17. The vehicle of claim **16** wherein the wing body carries two or more rudders on its left side and two or more rudders on its right side.

18. The vehicle of claim **17** wherein the wing-body is formed with a slot between each adjacent pair of rudders.

19. The vehicle of claim **1** wherein the wing-body has a projected planform area S, and a maximum outer diameter B normal to the wing-body axis, and wherein the ratio B2/S is greater than 0.5.

20. The vehicle of claim **1** further comprising three or more undercarriage elements for providing stability during take-off and landing operations.

21. The vehicle of claim **1** further comprising an energy source housed at least partially inside the outer wing-body.

* * * * *

US 20100172217A1

(19) **United States**

(12) **Patent Application Publication**

von Bezold

(10) **Pub. No.: US 2010/0172217 A1**

(43) **Pub. Date: Jul. 8, 2010**

(54) **SYSTEM OF METHODS AND APPARATI FOR HANDLING AND MODIFYING INFORMATION ENERGY AND MATTER BY FORMATIVE RESONANCE**

(76) Inventor: **Ernst von Bezold**, Richmond Hill (CA)

Correspondence Address:
Ernst von Bezold
RR2, 1020 Rutherford Rad
Maple, Ontario L6A 1S2

(21) Appl. No.: **12/350,960**

(22) Filed: **Jan. 8, 2009**

Publication Classification

(51) **Int. Cl.**
G01V 1/04 (2006.01)

(52) **U.S. Cl.** **367/190**

(57) **ABSTRACT**

This invention relates to an improvement in systems, methods and apparati for handling and processing information, energy and matter by making use of potentials of new technologies, science and industrial art, principally in cymatics, quantum field theory and psychoenergetic science. It comprises a system of methods and apparati for handling information, energy and material using multimodal cymatic (standing wave patterning, whether stationary or moving) generation, and manipulation, principally of material and energy in vorticoid including concatenated and grouped vorticoid forms, to concentrate, cleanse, separate, transform, produce etc., materials and energy, including without limitation the production of clean water and manufactures materials, the production of net usable energy by effects such as Schauberger and vacuum potential-related, the collection, processing and communication of information by nonlocal quantum coherence (entrainment) of dynamic systems, initially cymatic, viz., vorticoid, the similar transport of lower-dimensionality material patterns, i.e., materials, also to facilitate transport of living organisms, and the tuning of higher-dimensional energy forms, i.e., formative fields, for applications such as healing; a means of fine-tune controlling such systems directly by means of psychoenergetic operator interaction and indirectly by means of conditioned space technologies and vibratory enhancing technologies such as "scalar" treated materials and systems; and by other means described as "gauge-symmetry"-level raising, here, e.g., Bearden; and their combination; and a means of facilitating learning to generate Tiller-effect spacial conditioning (higher gauge symmetry) enhancement using motivated feedback in utilising these systems with spacial conditioning technology assist.

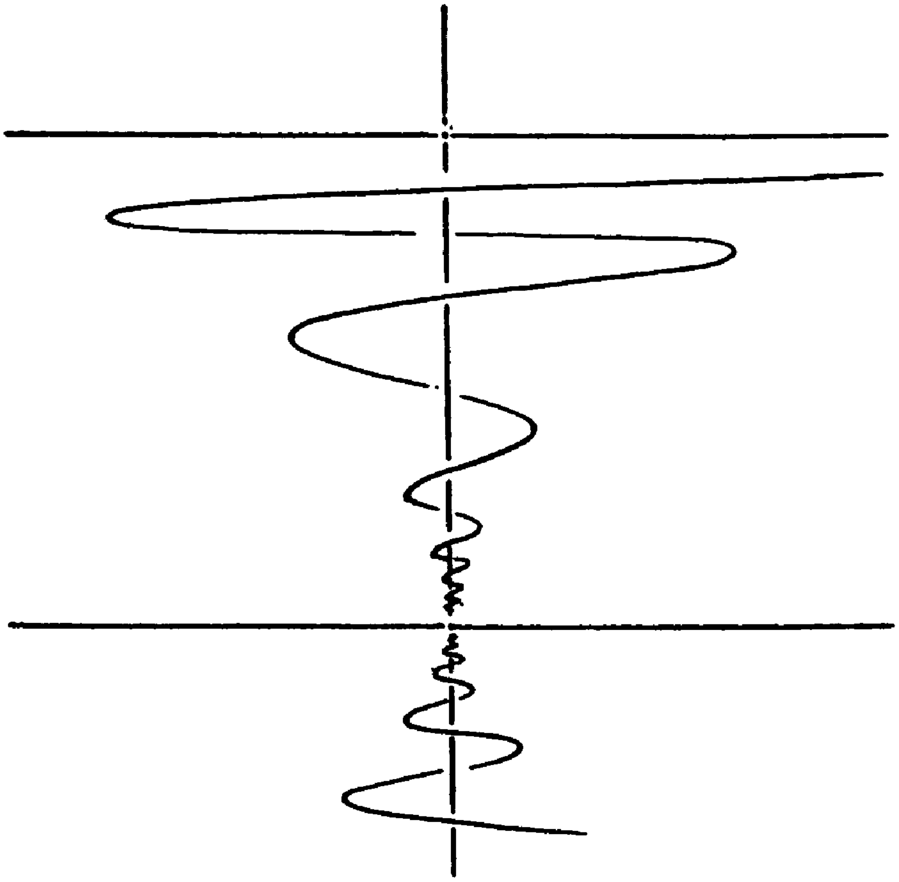

Figure 1

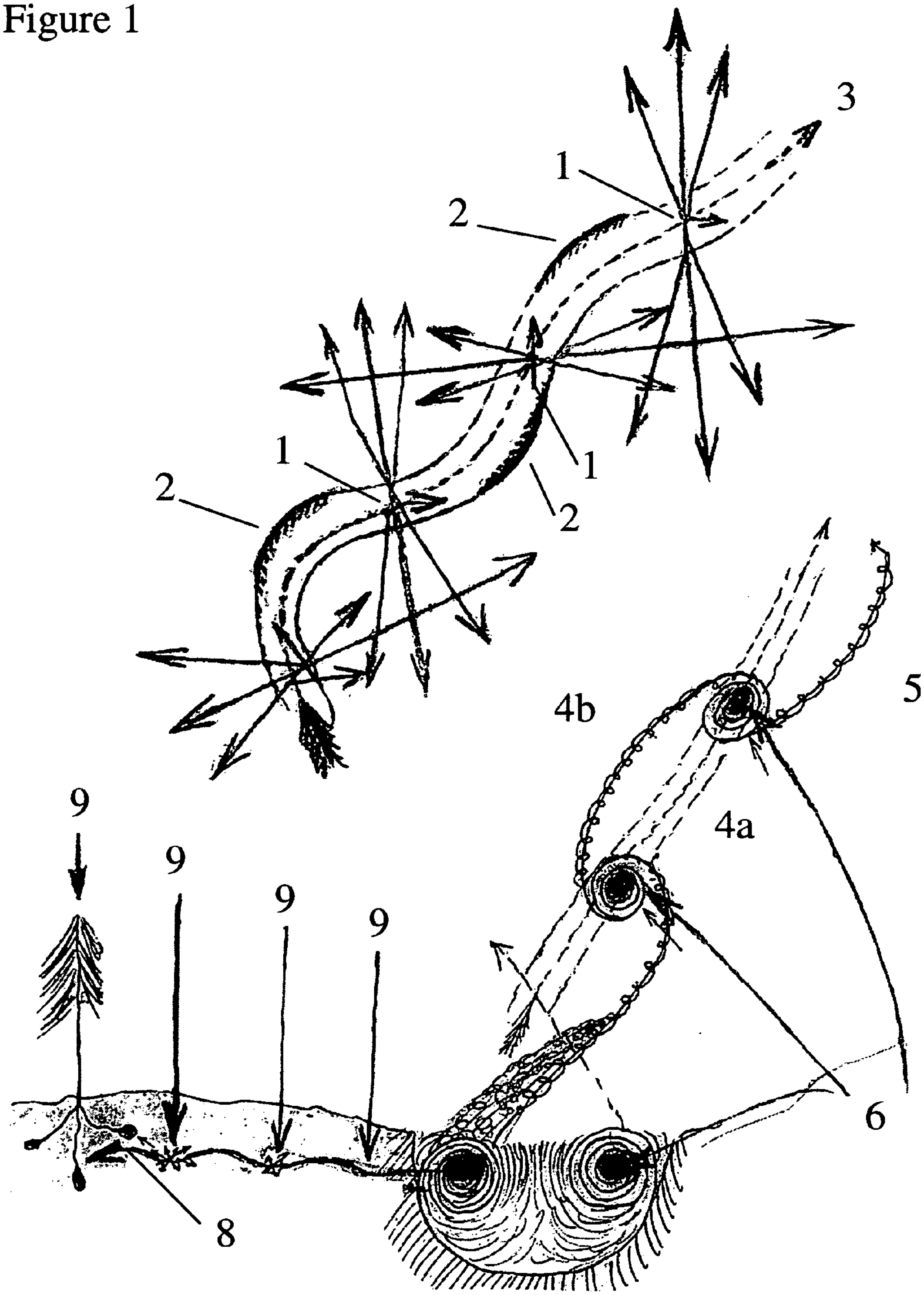
3
1
2
1
2
1
2
4b
5
4a
9
9
9
9
6
8
7

Figure 2

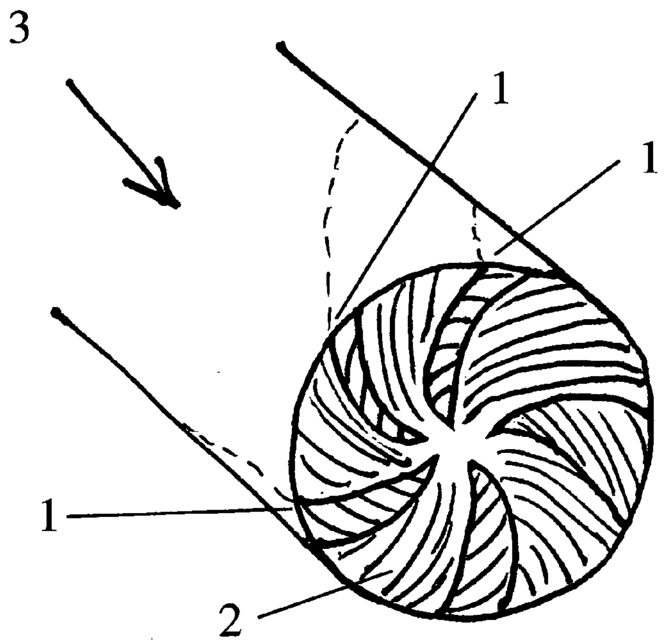

Figure 3

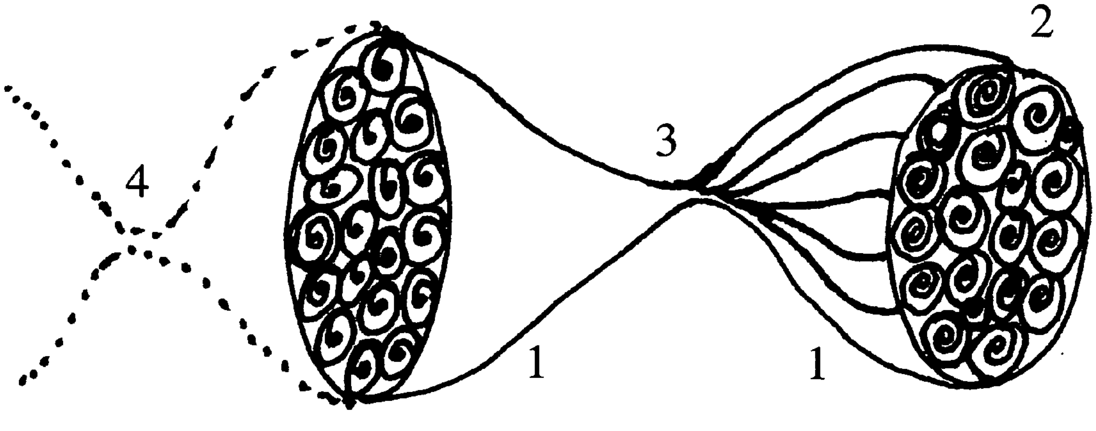

Figure 4a

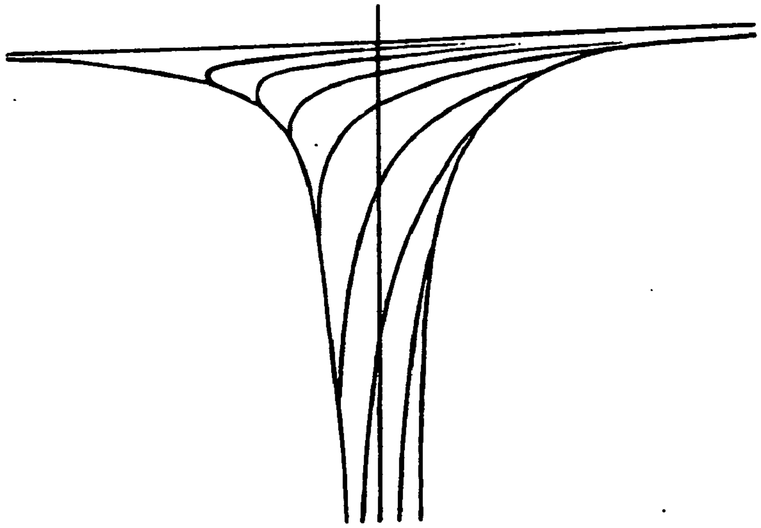

Figure 4b

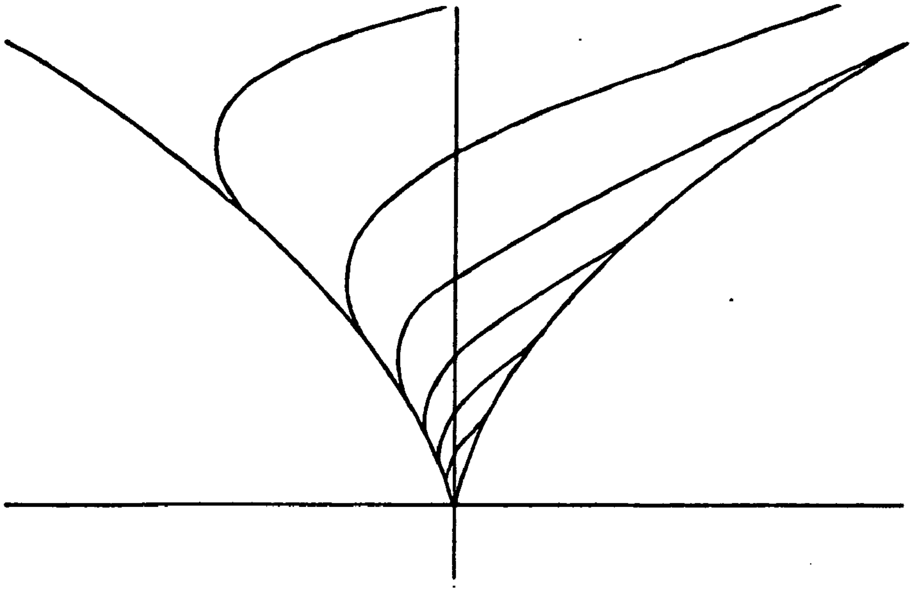

Figure 4c

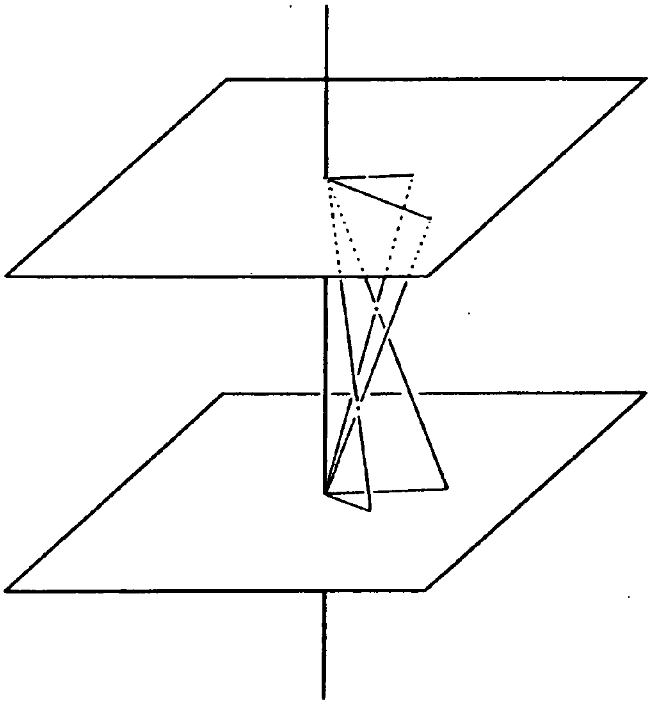

Figure 4d

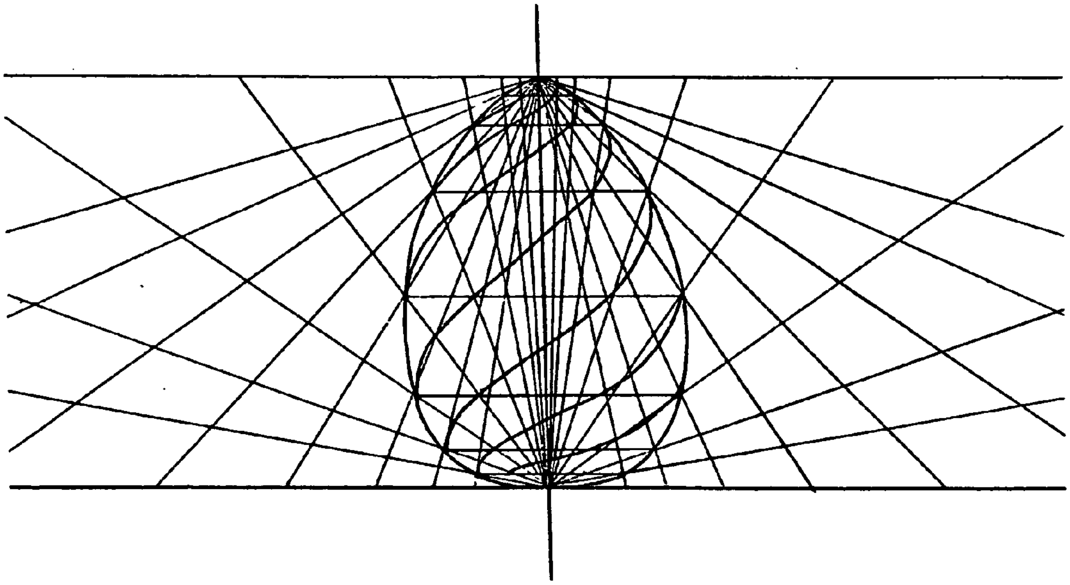

Figure 4e

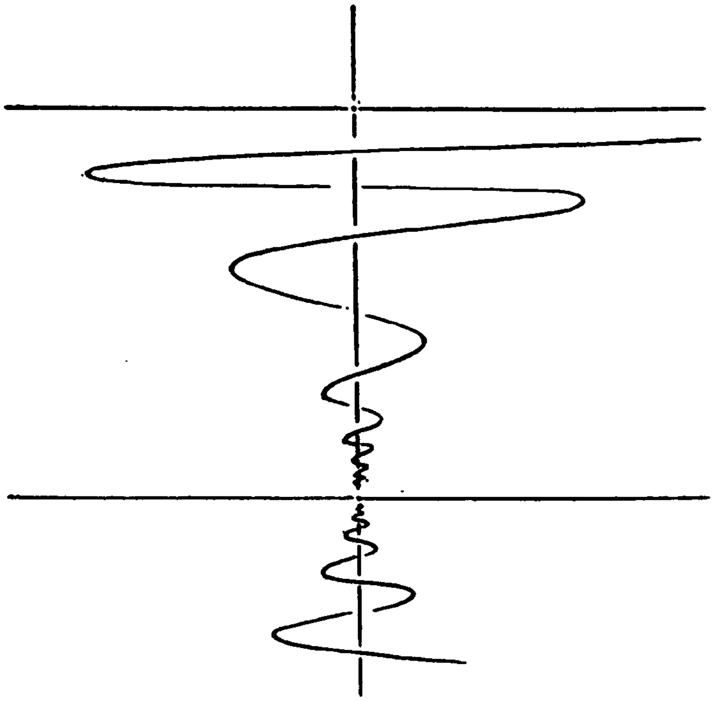

Figure 4f

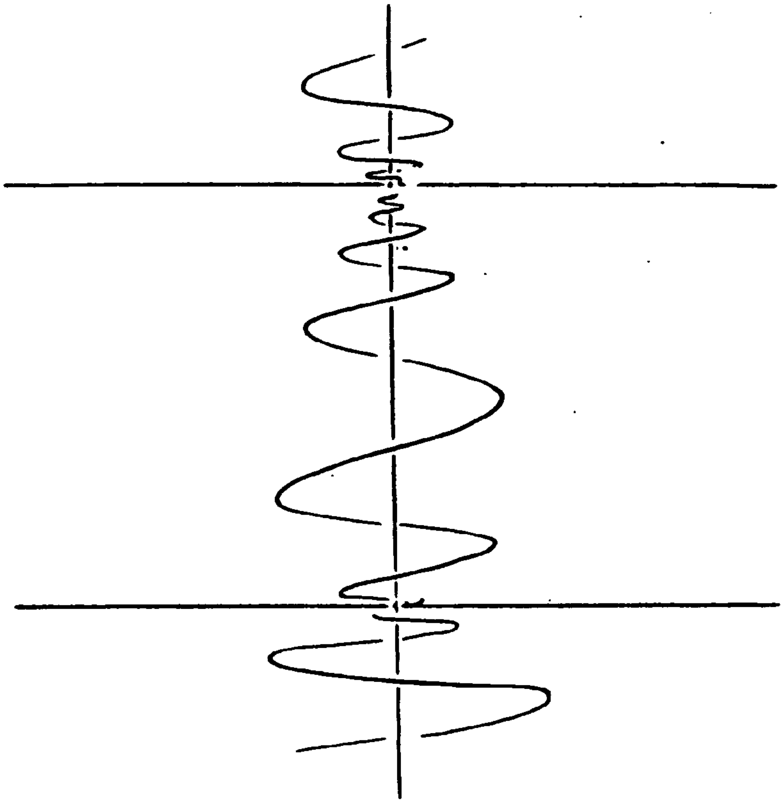

SYSTEM OF METHODS AND APPARATI FOR HANDLING AND MODIFYING INFORMATION ENERGY AND MATTER BY FORMATIVE RESONANCE

BACKGROUND OF THE INVENTION

[0001] This invention relates to an improvement in methods of handling and processing information, energy and matter. It was developed initially to make use of potentials of new technologies, science and industrial art to problems of non-chemical purification, with little or no chemicals added beyond water, of pre-combustion inputs and post-combustion outputs in thermoelectric power generation. The invention extends their application to new, nonobvious and useful systems of methods and apparati for the production, transformation and distribution of matter, energy and information.

PRIOR ART

[0002] With respect to the system of methods and apparati which is the subject of this application, and embodiments therefor, the work of the following, their predecessors and experimentally, clinically or epistemologically near neighbours, form the principal precedent contributions drawn on by this writer: Viktor Schauberger (Austria), Theodor Schwenk (Germany), Hans Jenny (Switzerland), Lawrence Edwards (England), Shafica Karagulla (Canada, United States), Marcel Vogel (United States), Fritz-Albert Popp (Germany), Thomas E. Bearden (United States), Ernst von Bezold (Canada) and William A. Tiller (Canada, United States). The following discussion outlines those contributions applicable to this work as understood by this writer and builds on them to show their relationship to novel, nonobvious and useful aspects of the invention here disclosed.

[0003] Viktor Schauberger was an Austrian forester and technologist whose works are now also widely available in English (1). He observed that when a current in a flowing stream winds centripetally and then centrifugally with an opposite sense of spin, at the turning point of the change of direction of rotation, oxygenation occurs and from this place also there is an energetic flow transverse to the direction of the current, into the stream banks, stimulating the growth of vegetation in those areas. Working with this and similar observations Schauberger was able to invent and frequently to patent 'naturalesque' technologies for diverse applications including trans mountain lumber transport and for transportation; for water purification, energy production, which he identified as environmentally healthier and therefore sustainable in comparison with still-prevalent environmentally destructive mechanistic technologies.

[0004] The centripetal-centrifugal spiral effects he reported (see figure) and made use of ('Schauberger effect') have been applied here in guiding the development of the invention, in light also of observation in the human being's energy field by numerous independent, highly sensitive and skilled observers, of major chakras (see for example S. Karagulla), whirling energetic flexible cones extending well beyond the physical body, situated at and through major glandular centres and aligned in pairs tip-to-tip pointing in opposite directions from the front and back of the body (e.g., B. Brennan and others) set along a central energy channel (called Sushumna in the orient) which runs vertically coincident with the human spine.

[0005] Theodor Schwenk's work on water as a spiraling and projective-geometrically characterisable cosmic sense organ demonstrated, by wheat germination trials using shaken water, the existence and rhythms of cosmic, i.e., extra-terrestrial, atmospherically significant Zeitgeber environmental rhythms, and therefore the importance of taking into account not only the use of vortices but also the significance of macroenvironmental rhythms in cleaning and enlivening water organisms to function harmoniously with the syntropic potential of the biospheric macroenvironment.

[0006] Hans Jenny was a physician living in Dornach, Switzerland who developed the field of Cymatics, creating and filming steady and moving three-dimensional standing waves generally produced with fluid agitated to a standing wave by sound through a loudspeaker base, based on Ernst Chladni's acoustically produced two-dimensional standing wave figures and on an appreciation of the contribution of vibration to the creation of form. Hans Jenny patented a device for making speech visible [ca. 1968; reference].

[0007] In a Chladni figure such as that formed by sprinkled fine sand vibrating on the back of a bowed horizontally held violin, the distribution of nodes and antinodes in the particular mode of vibration of the substrate medium, i.e., the violin back, at the frequencies at which it is being excited, may be displayed by the redistribution of the scattered sand in a pattern corresponding to the position of nodal and or to antinodal points and boundaries, evidently dependent primarily on the size and density of the sand particles. This acoustically mediated resonance has been researched by Jenny, Armin Husemann, Sherry Edwards and others as one of a number of formative principles in the human body. A formative faculty (classically, the facultas formatrix) was cited by Johannes Kepler to explain snow crystal form variation (the body of formative forces of the Earth); was researched in embryology by Erich Blechschmidt; and considered more broadly, has been studied as a phenomenon of internal communication demonstrable optically and functionally e.g., electrically, in all living organisms, by Fritz-Albert Popp and coworkers.

[0008] These examples of energetic formative processes in whole systems, i.e., in bodies of connected parts, demonstrate a formative paradigm which is applied and extended with multimodal energy flows or activity sources, e.g., electro-magneto-gravitic-(/levitatory)-]optic-acoustic and scalar, as well as psychoenergetic, for application in the invention.

[0009] Lawrence Edwards, an English mathematics teacher and biology researcher, showed that tree buds breathe, or fluctuate rhythmically, in winter around an ideal middle form whose projective geometric path curve is related to the golden mean. His vortex spiral calculation methods and examples are cited here as prior art for the production of vortices and their energetic as well as material products which will enliven nature functionally, in the sense Schauberger described and demonstrated, rather than degrading nature.

[0010] Dr. Shafica Karagulla left the Montreal Neurological Institute to do research on higher sense perception, eventually founding with Viola Neal the Higher Sense Perception Foundation in Los Angeles. Karagulla's clinical reporting of the correlation between chakra structure and activity on the one hand, and physiological health status such as patient endocrinology on the other (1965), provide along with the work of V. V. Hunt (UCLA) and other investigators such as R. Steiner a reliable modern base of published observation of some typical human psychoenergetic anatomy and its corresponding function.. This enables those less skilled in the art of such observation to make better use of transitory or fragmentary experimental data and to apply with greater confidence, along with those not making such observations, the findings of others as a source of design modeling information.

[0011] Marcel Vogel, a solid state chemist was instrumental in developing the first magnetic memory disk drive coating for IBM. Vogel published work (1973) demonstrating that psychoenergetic telekinetic phenomena become reproducible with sufficient experimenter self-training of the mind which is the instrument of manifesting such psychoenergetic phenomena. Taken together with the contribution of Tiller, Vogel's work exemplifies a shift to scientific knowledge and ordinary technical application of human capacities which have become a part of the state of the art in healing and are applied in the invention for process control enhancement of which water purification, energy production and telecommunication are three examples.

[0012] Fritz-Albert Popp founded the International Institute of Biophysics (Neuss, Germany, a collaborative research undertaking of dozens of scientists globally. His paradigm shifting discoveries have shown that the activity of a quantitatively small amount of energy, as light, directs and maintains the organisation of the body's biochemical activity, and that essentially all measured bioenergetic functions of organisms, e.g., skin electrical resistivity as measured in humans at hundreds of points on the hands, demonstrate an internal coherence predictive and therefore indicative of health. Normative self-validating measurement of coherence (cf. L. Edwards) in the functioning of cymatic structures is therefore incorporated as a monitoring and feedback control criterion of a preferred embodiment of the invention. Popp patent:

[0013] Thomas E. Bearden, a career research scientist (US Army, retired), wrote the book on Energy from the Vacuum (2002). His paper on Giant Negentropy from the Common Dipole provides a physical science foundation for the Schauberger effect which in its higher order gauge symmetry application appears consistent with W. A. Tiller's description of a second, more psychoenergetically malleable level of the physical world access to which depends on an engineerable coupling function. Bearden's presentation of a corrected and extended physics, taken together with technology which demonstrates it (Patrick et al., U.S. Pat. No. 6,362,718, MEG) provides with Schauberger's work the basis in the present invention justifying a the strategic objective of a water cleaning design for power plants which can be upgraded and modified functionally over time to increase net energy production process efficiencies so as to enable combustibles to be reduced and eventually phased out. Bearden's scientific-historical apparatus also provides a point of comparison with Tiller for quantitative modeling of conditioned space effects incorporated as a functional energy control option within the invention.

[0014] Ernst von Bezold published on the BDnow! Agricultural email list (2001), a description which then was archived by colorado.edu, of a class of technologies, i.e., using shapes combined with rhythmitized materials, based in part on L. Edwards' work on shapes applied to the functional renewal, by applying a thin coating of rhythmitized copper ointment (a therapeutic preparation manufactured by, e.g., Wala or Weleda in Europe, and Merkhur Laboratories in Canada, whose detailed experimental scientific basis is referenced by R. Hauschka), of cow horns used in bio-dynamic agriculture, e.g., in making manure or powdered quartz preparations. The principle of cohering shape and material, intensified by rhythmical treatment of the material (treatment such as may also be found in N. Tesla's purple plates, evidently in the Austrian Johannes Grander's water treatment technology, or in more recent "scalar"-resonant liquid treatment vessels and jewelry) has a parallel in the technology of the great pyramid at Giza and an extension in the biogeometry of A. Kareem as and as presented by R. Gilbert, et al. As described earlier in relation to Edwards' work, this principle is applied here with respect to optimised vortex shaping (cf. Edwards) and material, in this case water, formed and controlled cymatically, with psychoenergetic system control enhancement, and eventually, generation.

[0015] William A. Tiller, former department chair and professor emeritus of material science and engineering at Stanford University, wrote Psychoenergetic Science as a simplification and compilation of three previous related volumes documenting 30 years of paradigm-shifting experimental science on the demonstrable, potential, and instrumentally engineerable enhancement of, the integral functional contribution of consciousness to the matter-energy equilibrium of the object-oriented physical science paradigm presently being transformed by the more interactive understanding of reality shown in quantum physics being applied in such diverse fields such as cryptography, quantum computing and psychoneuroimmunologic therapies. Tiller's empirical work on space conditioning for process modulation, together with his work and that of Vogel on experimenter training for replicability (cf. Charles Tart, "state-specific science") applies and extends the implications of the Indian preceptor Patanjali (ca, 8^{th} century) in a unique and distinctly instrumentally modern, scientifically verified and technologically accessible form. The invention applies both the trained-operator effect and the instrument-mediated spacial conditioning effects to control and enhance multimodal cymatic processing and to and anticipate its transformative progression in part as an operator-initiated and controlled activity with corresponding strategic extension of options of scope and scale of application.

REFERENCE NOTES

[0016] 1. 1. Drawings, and descriptions in English, are available from and, respectively for, the work of V. Schauberger e.g., translated and edited by Callum Coates, Ecotechnology v. 1-4. Gateway, Dublin, Ireland, 1997-2000.

[0017] 2. Hans Jenny: Cymatics. Basel, Switzerland.

[0018] 3. Shafica Karagulla: Breakthrough to Creativity. DeVorss, Santa Monica Calif., ca, 1965

[0019] 4. Marcel Vogel: Reproducible long distance communication with plants detailed in Proceedings of the Second International Psychotronics Conference, Prague, 1973.

[0020] 5. Interview for Canadian Broadcasting Corporation 'Ideas' series "Physics and Beyond", 1973.

[0021] 6. Fritz-Albert Popp: Director International Institute of Biophysics, Neuss, Germany. For bibliography see www.lifescientists.de

[0022] 7. Edwards, Lawrence: The Vortex of Life nature's patterns in space and time. Floris, Edinburgh, 1993. ISBN 0-86315-148-5

[0023] 8. Bearden, T. E.: "Giant Negentropy from the Common Dipole," cited from *Journal of New Energy,* 5(1), Summer 2000, p. 11-23. Offprint by the Planetary Association for Clean Energy, Inc., Ottawa Canada; ISBN 0-919969-11-9 [2001].

[0024] 9. Energy From the Vacuum concepts and principles. Xxviii, 952 p. Cheniere Press, Santa Barbara Calif., 2002.

[0025] 10. Patrick; Stephen L. Bearden; Thomas E., Hayes; James C., Moore; Kenneth D., Kenny; James L. Motionless electromagnetic generator. U.S. Pat. No. 6,362,718. Mar. 26, 2002

[0026] 11. Tiller, W. A.: Psychoenergetic science: a second Copernican-scale revolution. Pavior Publishing, Walnut Creek, Calif. c. 2007. ISBN 978-1-4243-3863-4 www.tiller.org

BRIEF SUMMARY OF THE INVENTION

[0027] Essentially, and without limitation of this description to its application in an illustrative preferred embodiment, a cymatic (3- or higher-dimensional Chladni figure-generating) resonant vibrational process is used to induce a rotating standing wave figure in a fluid, such as water, air, plasma, or a slurry including solids, producing a centripetal vortex or vortex of vortices which passing inward through the focus of the tip unwinds from its tip in the same or by induction in an opposite rotation, or meets another incoming vortex tip or tips, at the juncture comprising the tip(s) focus producing differentiable active and potential energetic outputs, and from the focus a likewise differentiable outflow of materials in directions primarily other than that of the axis of tip focus, such as transverse, of least resistance as shaped and sorted within transporting flows of standing waves e.g., whose node-antinode frequencies, amplitudes and spacial configurations are appropriately resonant with selected materials and material-energetic configurations. The forming and modulation of the form may be by external or internal, direct or interferometrically induced, multimodal electro-magneto-gravitic-opto-acoustic and scalar potential drivers controlled by feedback of functional parameters of the manipulated material flow and feedback of the driving energy inputs and their energetic transforms, and having regard to macroenvironmental conditional-probability fluctuations, e.g., in turbulent flow, and, in early embodiments, being fine tuned—later, consequent on operator learning based on feedback with the system as an instrument of learning, increasingly being generated—by operator interaction directly (Vogel), or indirectly using conditioned space effect technology (Tiller) amplification-modulation or other higher gauge broken symmetry operations (Bearden), to bias threshold condition probabilities favourably in respect of intended states of the system.

BRIEF DESCRIPTION OF THE DRAWINGS

[0028] Annotated List of FIGS. **1**. to **4**(*a*)-(*f*)

[0029] FIG. **1**. Background of the invention and cross-reference for FIG. **3**. which illustrates a feature of a preferable embodiment of the invention:

[0030] An example from river current flow, "The Inner Picture of the Dynamics of Current Flow", showing inward spiraling with reversal of spiral direction at the point, i.e., in the region, of start of outward spiraling—and indicating transverse-vectored radiant propagation of energy, as well as in this context reactive enhancement, i.e., oxygenation—is from Schauberger, 1941, as reproduced in Coates, Vol. 1, 1997.

[0031] FIGS. **2** and **3**. Illustrations of features of preferable embodiments of the invention:

[0032] FIG. **2**. A sketch of a sample cross-section functionally spiraling pathway, i.e. here as spirally-fluted piping, whether formed structurally, or through the formed and formative injection of materials such as water into and as a part of the pathway, or in part or in whole energetic-dynamically, or by some combination or permutation of these.

[0033] FIG. **3**. A sketch of a sample complex vortex of vortices, modeled on the inventor's observation as well as on descriptions, of naturally-occurring and developed or "built" complex vortex structures which may be characterised as dynamically stable n-dimensional "standing wave" structures. The example is of a collinear pair of complex vortices shown in oblique cross-section; to the left in the figure is appended a dotted-line indicator of a contiguous, expansion-contraction vortex figure such as may be found in FIG. **1** with a second point of maximal vortex centripetal concentration, as may also be seen in FIG. **4**(*f*). This vortex meeting, or transition, point is a focal region of energy concentration and release which as a focal point of modulation (depending also on vortex spin directions) can be used to amplify the modulating signal's effect in the centrifugal or expansive phase or, respectively depending on vortex spin directions, to modulate energy released along another vector from the focal region; for a naturalesque example; see FIG. **1**.

[0034] FIG. **4** note: Such sets of vortices can be concatenated in various geometries—linearly, in rings, nested, etc.—as information energy and materials handling and transformation structures, e.g., in the processing of water of whatever phase, and gases or other materials, of energy, and of information, associated with combustion, of coal, etc; with other manufacture of energy and of material substance in preferred composition, form, state, and the production of purposed information-rich structures or forms and structures or forms for processing information; with the sensing or other collection, processing, communication of information and associated energetic and material entities, e.g., as information carriers, viz., modulating and as modulators, locally and nonlocally, for example using functionally entrained (i.e., resonantly tuned) nonlocally paired single or doubled (for energetic balance) sets of more or less triple vortices, e.g., three-dimensionally depending on intended function, which are monitored and modulated at the origin (see **4**(*e*) below), e.g., using one or more members of the class of modulators listed above, especially operator- or user-specifiable modulators, to enable transfer of informational, energetic, material pattern momentum between the nonlocally paired vortex sets, in order to facilitate convenient transfer of process-related patterning for the effective replication of data, of energy states, or transfer under competent operation conditions of low-dimensional-number patterns that act as localizing carriers for higher-dimensionally-resonant homeostatic structural fields and their associated states, i.e., principally in living organisms. The same principle can be applied reflexively to enhance holotropic pattern stability of organisms with respect to an operator-programmable template field; e.g., as two examples among many, for conservation and improvement of biotic function and its patterned energetic basis using an algorithmically-controlled amplifier to enhance food stability and therefore conservation or to support healing based on modulation of a resonantly shaped and directed energy flow using a reflexively applied template which is functionally operator-enhanced.

[0035] As cited above, a mathematical apparatus for calculating projective-geometrically visualisable vortex and related path curves, including but not limited to a sample computer program, and sample illustrations of the families of curves employed here, which it is proposed are a functional correspondent class of the shapes observed and then designed and used by V. Schauberger, may be found in Lawrence Edwards' work, inter alia: The Vortex of Life, 1993. See also articles in Mathematical-Physical Correspondence; and, The Field of Form, 1982. FIG. **4** illustrates examples of some applicable projective-geometric curves and their calculation from Edwards (1993); of classes of curves to be applied and based on the mathematical descriptions of their classes, to be used in examples, indicated descriptively in the foregoing (i.e., in the accompanying) text, of preferred embodiments of the invention.

DETAILED DESCRIPTION OF THE INVENTION

[0036] The present invention now will be described more fully with reference to the general inventive concept, and to illustrate embodiments of the invention. As this invention may be embodied in many different forms, its scope should not be construed as limited to the exemplary embodiments set forth; rather, the examples are provided to indicate a scope of potential preferred embodiments to those skilled in the art.

[0037] The present invention of a system of methods to handle information, energy and materials and to modify their properties relates to the modification by manipulation iteratively (including but not limited to, in particular, recursively; the term "iteratively" is so used throughout), simultaneously and prospectively or retrospectively, of the potential and manifest energetic, temporal and spacial coordinate, motional and orientational, coherence and informational or entropic status, dimensionality including but not limited to metric (dimensional curvature), combination or dissociation, of materials in amounts and states ranging from particle to compound and bulk mass, including but not limited to plasma, gas, solid and liquid states singly, intermixed or combined, making use of one or a coordinated combination (including but not limited to, as the case may be, ordered combination, i.e., permutation; "combination" is so used as an inclusive term, throughout) of more than one structure and system of vibration and resonance of material or energetic shape, including but not limited to environmental Zeitgebers or related factors or their artificial analogues, in a feedback or resonance-controlled manner permitting dynamically active, i.e., reflexively, iteratively and synchronically varying differentiation, integration, transformation including but not limited to metamorphosis or conversion from one state to another, of informational, energetic and associated material states, systems, forms and entities.

[0038] The invention consists of a unimodal or multimodal integrated system of methods as outlined here to handle information, energy and material and modify their properties, including but not limited to local and nonlocal interaction, patterning and coherence.

[0039] An example of a class of embodiments is as it relates to communication of information, energy and material, particularly between the holographic sub-quantum and quantum-level entrainment and cross-modulation of asymptotically and hyperbolically n-dimensionally pattern-matched, e.g., including but not limited to in particular functionally synchronous nonlocal, systems of information, energy and material.

[0040] Another example of a class of embodiments is as it relates to holotropically "self"-similar precursorial re-equilibration of organisms and entities, including but not limited to in particular their energetic (viz. their persisting or durable sub- and quantum-resonant) template or formative pattern in a range of applications from medicine to physical device and structures maintenance, to re-establish a proximal functional and structural wholeness at a higher level of holistic functional coherence, in particular also therefore to re-tune them optimally with and to a resonant pattern of macro-environmentally current syntropy with its accessible dimensional and nondimensional substates. "Vis medicatrix naturae per facultatem formatrix": i.e., re-tune organism to their archetypal (basic, healthy) formative patterns in harmony with the macrobiosphere, "The healing life of nature [accessed] through the formative faculty".

[0041] An example of a class of embodiments preferable for their obviously immediately beneficial and pertinent application relates principally to the energies and material substance associated with solids, with liquid and gaseous fluids and solids associated with them; to plasma, and associated quantum-energetic (viz., respectively, subquantum-energetic, i.e., including but not limited to, in particular, ["quantum"-]nonlocally entrained energetic-potential such as so-called zero-point) states.

[0042] A specific application in this class of embodiment is its use for the pre-, real-time ongoing, and post-treatment of fluids, in particular, water and gases including but not limited to related particulates, associated with combustion, in the processing, particularly the cleaning, of pre-, combusting and post-combustion gases, in particular, respectively, gaseous/ified combustants, products and proximal energy, materials and energetic patterns associated with combustion of e.g., hydrocarbons such as coal; by the use singly and in combinations; continuously, and discontinuously, through design-resonant, or forced pulsation inducing turbulence, pulsation, cavitation; of physical-structurally, or energetically, e.g., acoustically; magnetically; etc.; functionally spiraling, e.g., spirally-fluted piping, pathways which may be used to induce or facilitate; optionally with 2, 3 or more dimensional Chladni-figure, i.e., cymatic, dynamically shifting resonance/feedback-controlled electro-magneto-gravitic-levitatory-opto-acoustic-temporal-spacial standing-wave patterns, including but not limited to scalar Whittaker fine structure signaling particularly for thermal behaviour control of materials, and application of Grander and Tiller field-effect systems; and also, optionally with a class of directly operator-driven phenomena, e.g., Keely-effect, Vogel-method or "Hutchison-effect" induction;

[0043] to condense or rarefy and drive spacially, i.e., to segregate, mix, react, shape, etc., said fluids, viz. gases and their constituent and included substances including but not limited to precipitates, elements and compounds to be formed or dissociated, fractionated out, otherwise modified, and to assist in the production of energy; especially, in particular, where certain forms are approached or striven for or referenced, such as the mean as well as ideal value cycloid space curve families described by Lawrence Edwards in the life of nature, materials in these forms being evidently energizing resonant transducers of water quality renewal and related effects; and forms such as the bio-geometrically-analysed designs—e.g., concavities in structures, such as, e.g. a gradual half-degree indentation of each of the outer upper surfaces of the Great Pyramid at Giza used to induce "horizontal negative green", and other, cone-pendulum-response states as presented and described by A. Kareem, R. Gilbert and others—which may be favourable or conducive to the improvement of informational content, energy and material, e.g., i.e., in this case of quality of water and gases and their forms and behaviours, especially combustion-associated gases and water in whatever phase, and of their environment, and of related cleansing, reactive or structural-energetic-chemical including but not limited to biochemical stabilizing activity in this setting as well as in other settings such as biochemical cleansing, stabilisation, renewal of living organisms, their environment and precedents including but not limited to inputs and succedents including but not limited to outputs.

[0044] Modulation preferentially may be varied in feedback, resonantly simultaneous with, or preceding (e.g., within entrainment parameter values range) synchrony with, environmental Zeitgebers and their measures, e.g., using rhythmitized shapes (rhythmitised shaping functions in classes such as that of the Great pyramid, cow horns and other natural structures as resonators, described as mathematical function classes by L. Edwards), with respect to the patterns shown by astronomical ephemerides, and the mathematical descriptions of, in particular, planes and other shapes or figures formed among massive bodies during their transits, by astronomically related variations in the values of the tensor Tau sub jk and associated described functions including but not limited to those identified by G. Unger of Dornach, Switzerland, and by associated indicators; to optimise negentropic as well as entropic functions severally and in combination.

[0045] A specific class of instances of this kind of embodiment would utilise rotating multimodal fields and resonant "radiative" and "field" effects, whether by gross energetic induction as described or by resonant rhythmitised shaping or by indirect Tiller-type or directly operator-programmed tuning or combinations to induce essentially, preferably by means without moving parts,

[0046] the enlivening of energy and of material, i.e., of water;

[0047] pre-cleaning and conditioning of gases; including but not limited to, for both water and combustion-related gases, electrical or other energetic charging or programming, e.g., magnetic induction of appropriate coherent spin and precession of particles and other structures, e.g., hydrogen proton precession;

[0048] cymatic vortical motion including but not limited to, based on the work of Viktor Schauberger, Theodor Schwenk and others, centripetal as well as, as may be useful or conducive in application, centrifugal, and alternating centripetal/centrifugal, with "braiding" of flows to facilitate blending—respectively separation of succedent materials and states—of spiraling gases and water; leading to improved mixing, modification and separation of materials, and improved control of energy production, improved cleansing and production of associated materials in usable higher quality lower cost form than might otherwise be the case;

[0049] modification of gasification, precipitation, and endo/exothermic processes by Whittaker fine structure signaling, by Grander spacial and magnetic induction, et. seq., especially for real-time process- and flow-control stabilisation;

[0050] modification of the foregoing and of other processes, states and conditions through operator-controlled intent and concentration of energies, indirectly by programmable-unit Tiller "field effect" action, viz. gauge-symmetry level modulation, as well as directly, including but not limited to Keely-technology and Vogel-methodology effects.

[0051] An example of a design which preferably embodies elements of the foregoing in an application to facilitate production of clean combustion by-products of coal and of other combustibles or pre-combustible feedstocks and their combustion processing materials and systems, the materials being principally water and air, with an optimal net production of environmentally life-enhancing (i.e., also therefore, minimally life-degrading) quality and quantity of energy, material and information especially bioinformation, is as follows:

[0052] Iteratively, and by resonance including but not limited to pre-emptive or pattern-anticipatory entrainment-leveraging e.g., phase-lock, signaling,

[0053] 1. Form water and gas flows architecturally, i.e., physical-structurally using the patented and unpatented built as well as the designed examples from among: R. Steiner; especially, V. Schauberger with numerous illustrations of examples in patents and other literature to be cited; T. Schwenk; J. Grander; and for informational, energetic and materials modulating effects using shaped rhythmitised materials, E. von Bezold (2001); and, by means of a "complex" vortex, i.e., a set of complex vortices each comprised of multiple vortices arranged in groups as dynamically stable standing-wave "cells", the focal point of each of which may centre on and be contiguous with the focal point of a corresponding collinear vortex whose mouth is pointed in the opposite direction to the first, or with the focal points of a number of vortices dynamically balanced energetically and in space, all with one another.

[0054] 2. Also shape the streams, e.g., for a first approximation as exemplified in the foregoing paragraph, vibrationally by tuned coherent energetic induction and modulation magnetically/diamagnetically, opto-acoustically-thermally using dynamic Chladni-figure standing-wave vortices, i.e., after the cymatics of H. Jenny et. al.—also approached by the flow forms of Wilkes, articulated by A. Husemann and others, also alluded to mathematically by G. Adams, by P. Forchheimer, later by T. Schwenk (1954) who cites I Prigogine in the context of sensitive control-response; discussed mathematically by R. Abram and concurrently and subsequently especially in the work of L. Edwards—in the classes of shapes expressed mathematically by L. Edwards (1993) who provides calculation methods, computer software and diagrammatic illustrations: oscillating, varying, within and around syntropic, life optima "antenna" shapes, e.g., mean as well as ideal shapes, tuned over time to work with principal macro-environmental (Zeitgeber) rhythms; and transformed (e.g., by rotation and by translation in flow, e.g., induced acoustically or by optically-tunnelled electromagnetic radiation, e.g., amplitude-, frequency- and waveform-specific light, such as squeezed light or other coherent as well as incoherent light) metrically and physico-chemically, pulsated, and otherwise modulated to induce enhanced laminar flow, controlled turbulence, and cavitation-reformative collapse cycles, including but not limited to cycles that are rhythmically- i.e. resonantly-enhanced; and further to modulate the sensitive, cf. T. Schwenk, non-equilibrium, cf. I. Prigogine, self-organising activity.

[0055] 3. Concurrently, modify the processes locally and globally, in part and in whole, using as appropriate the means described in 1. and 2., and other available signalling energizing systems as modulators for the above control systems and as net additional inputs and, as appropriate, use energetic outputs to be fed back into the system as controlling modulators; including but not limited to, as examples, electrostatic dispersion; complementary electrical charging; electrolysis; magnetic, endo/exothermal, other emf or energetic treatment including but not limited to macroenvironmental signals, unamplified and amplified; to, e.g., increase, maintain or reduce solubility; flow formation, as described above, to enliven fluids; centripetal cyclones, compression vortex mixing and transformation of inputs, e.g., vortical compression wave precipitation threshold modification; space state, time state and information coherence system behaviour state programming, e.g., via proximity field conditioning, e.g., Tesla purple plate technology or other "scalar treated" programmed resonators, characterisable for example in quartz by a Cartesian matrix spaced throughout the material of silicon atoms each containing a proton with a doubled Lissajous figure; Tiller-effect units; Grander units; Whittaker structure modulation to improve e.g., output cleansing, output precipitation conditions as well as to optimise thermal production efficiency exothermically internally and endothermically at specified system functional boundaries to eliminate net environmental negative impact and, as with the whole system, optimise positive impact; further to the description in 2., use of sound and acoustically-tuned EMF, e.g., as heterodyned fields, and other means to induce mass hyperacceleration and velocity effects, and other methods of catalysis, e.g., to improve reaction chemistry, hydroxyl radical and hydrogen chemistry optimization, etc.; further applications integrated synergically with above processes, of acoustic, especially pulsating, and EMR (electromagnetic radiation) and other radiative effects other than magnetic; e.g., for improved control of combination-dissociation processes by sound, acoustico-magnetic, infrared, UV, etc., within transient standing-wave cells whose opto- and magneto-acoustics are, for example, like those demonstrated optically by a cyclone, a tornado or an optical fibre, i.e., distinctly self-contained or transmissive in a parametrically-specific manner.

[0056] Specific terms employed for illustration should be read in a generic and descriptive sense only and not for purposes of limitation. The scope of the invention is further set forth in the following claims.

1. A system of methods and apparati for handling information, energy and material using multimodal cymatic (standing wave patterning, whether stationary or moving) generation, and manipulation, principally of material and energy in vorticoid including concatenated and grouped vorticoid forms, to concentrate, cleanse, separate, transform, produce etc., materials and energy, including without limitation the production of clean water and manufactures materials, the production of net usable energy by effects such as Schauberger and vacuum potential-related, the collection, processing and communication of information by nonlocal quantum coherence (entrainment) of dynamic systems, initially cymatic, viz., vorticoid, the similar transport of lower-dimensionality material patterns, i.e., materials, also to facilitate transport of living organisms, and the tuning of higher-dimensional energy forms, i.e., formative fields, for applications such as healing.

2. A means of fine-tune controlling such systems directly by means of psychoenergetic operator interaction and indirectly by means of conditioned space technologies and vibratory enhancing technologies such as "scalar" treated materials and systems; and by other means described as "gauge-symmetry"-level raising, here, e.g., Bearden; and their combination.

3. Means of facilitating learning to generate Tiller-effect spacial conditioning (higher gauge symmetry) enhancement using motivated feedback in utilising these systems with spacial conditioning technology assist.

4. Further, based on the foregoing, all related dependent claims which may properly be construed from the foregoing.

* * * * *

US 20100270967A1

(19) **United States**

(12) **Patent Application Publication**

CHO et al.

(10) **Pub. No.: US 2010/0270967 A1**

(43) **Pub. Date: Oct. 28, 2010**

---

(54) **APPARATUS FOR HARVESTING ENERGY FROM MICROWAVE**

(75) Inventors: **In-Kui CHO**, Daejeon (KR); **Jung-Ick MOON**, Daejeon (KR); **Je-Hoon YUN**, Daejeon (KR); **Jong-Hwa KWON**, Daejeon (KR)

Correspondence Address:
RABIN & Berdo, PC
1101 14TH STREET, NW, SUITE 500
WASHINGTON, DC 20005 (US)

(73) Assignee: **ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE**, Daejeon (KR)

(21) Appl. No.: **12/696,442**

(22) Filed: **Jan. 29, 2010**

(30) **Foreign Application Priority Data**

Apr. 28, 2009 (KR) 10-2009-0037190

Publication Classification

(51) **Int. Cl.**
H01M 10/46 (2006.01)
H02J 7/00 (2006.01)

(52) **U.S. Cl.** **320/101**; 320/107

(57) **ABSTRACT**

Provided is an apparatus for harvesting energy from a microwave. An apparatus for harvesting energy from a microwave in an electricity storage system of a building, includes: a plurality of rectennas arranged with a predetermined length according to characteristic of microwave to be absorbed, and configured to collect microwave in the atmosphere and convert the collected microwave into electric energy; a current converter configured to convert the electric energy outputted from the rectennas into a storable current; and a charging tank configured to store the current outputted from the current converter, wherein the arranged rectennas are attached to an outer wall of the building, and the rectennas attached to the top of the building are printed in a transparent conductor on the surface of a solar cell.

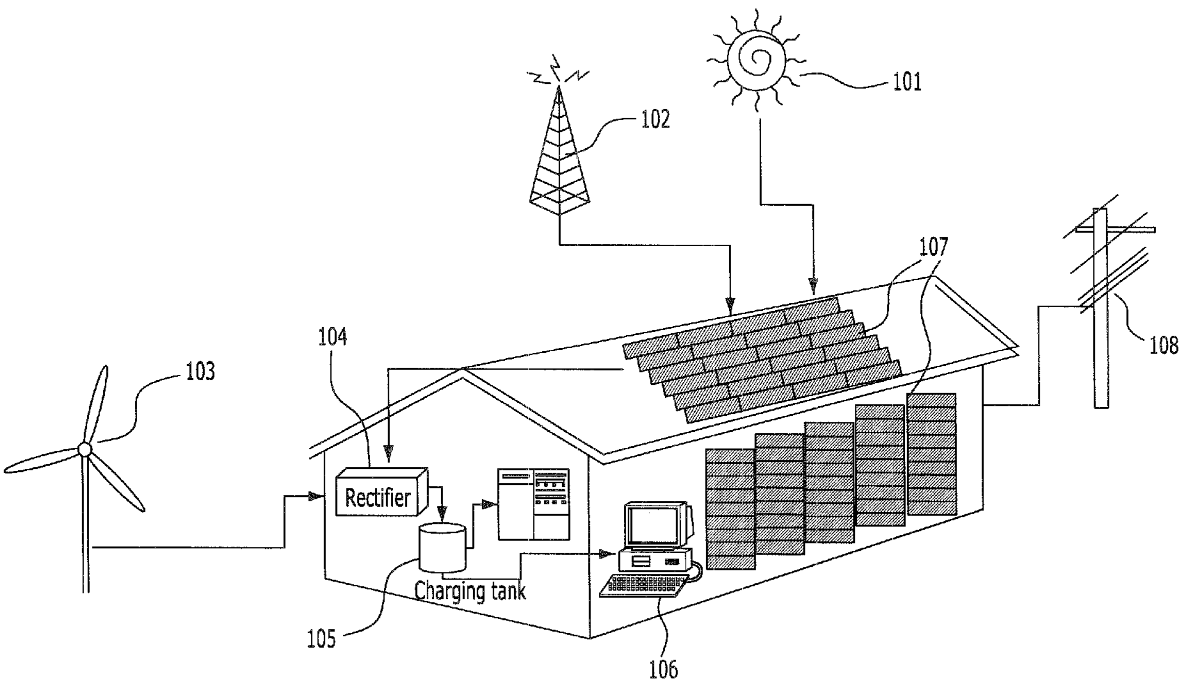

FIG. 1

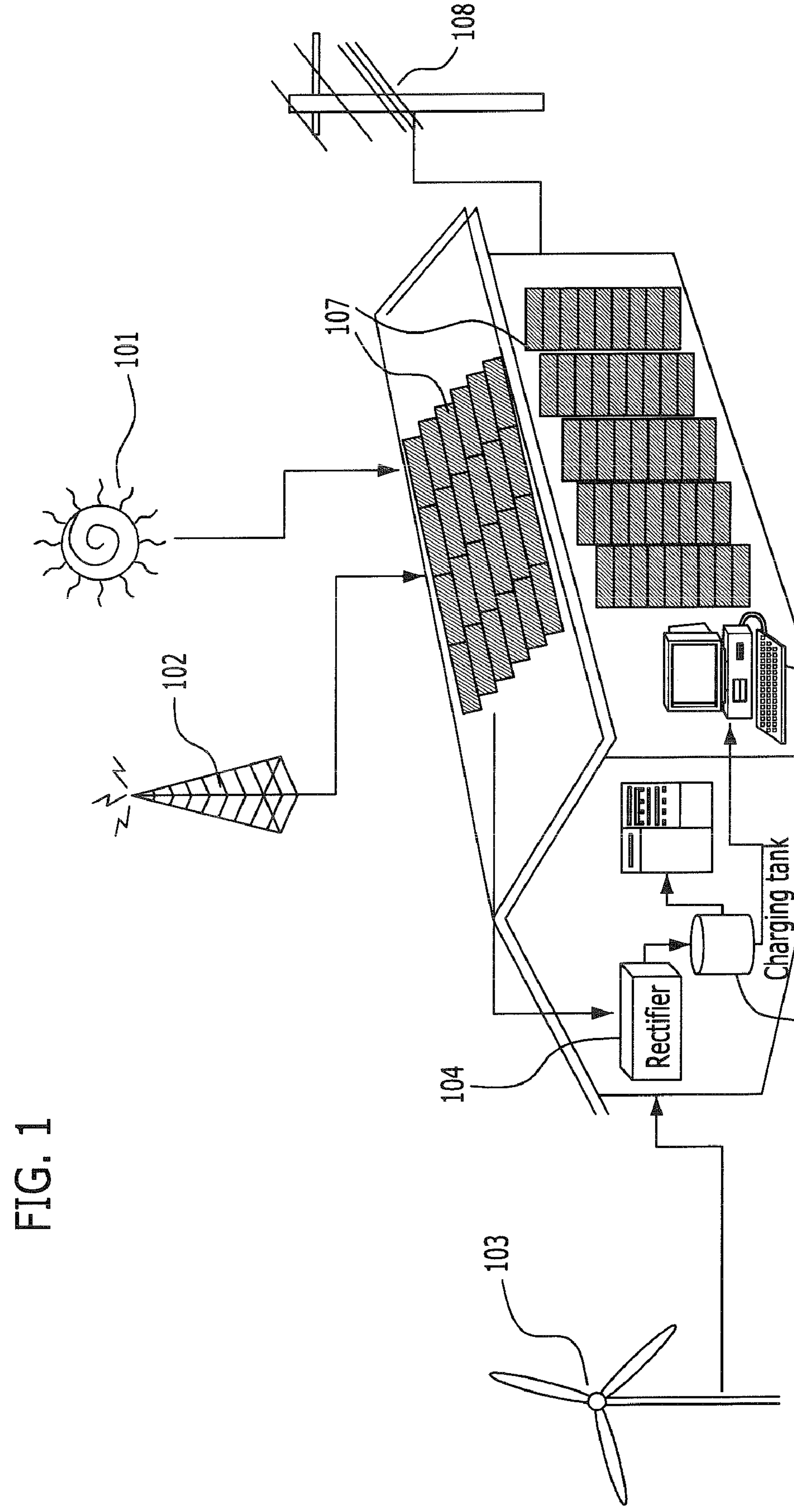
101
102
103
104
105
106
107
107
108
Rectifier
Charging tank

FIG. 2

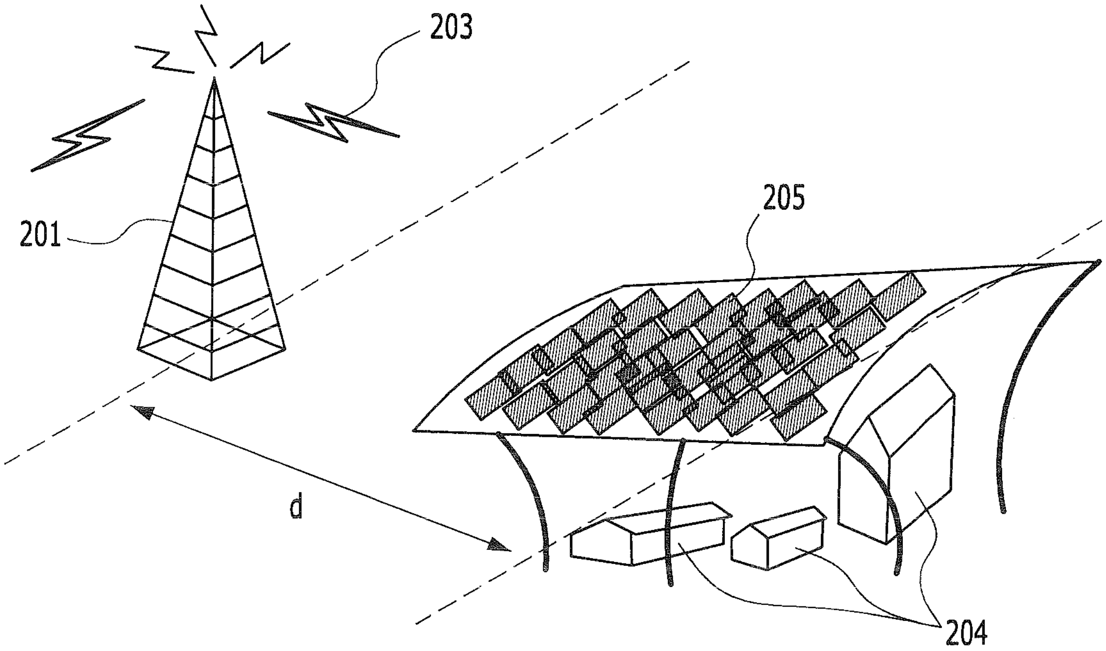
203
205
201
d
204

FIG. 3

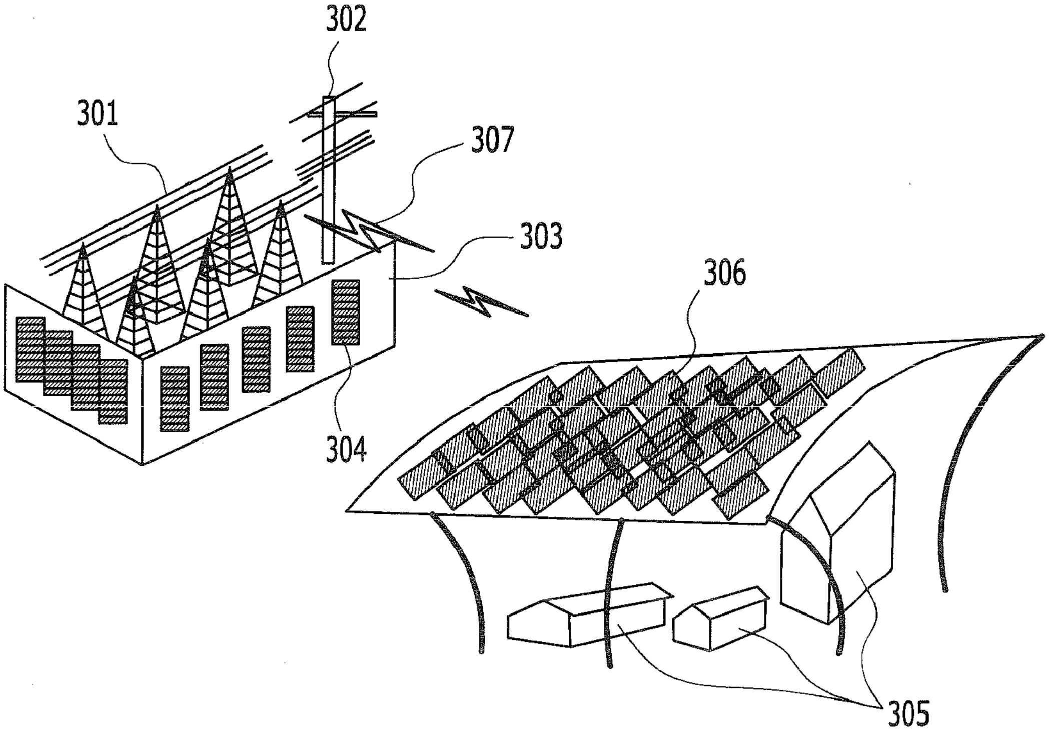
302
301
307
303
306
304
305

FIG. 4

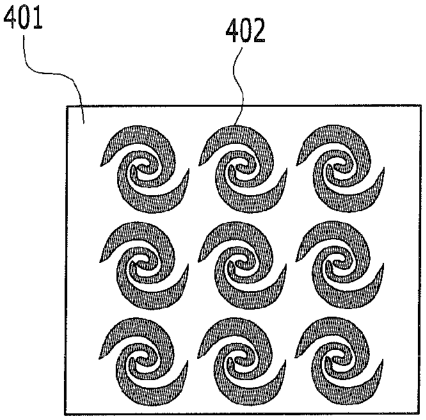

FIG. 5

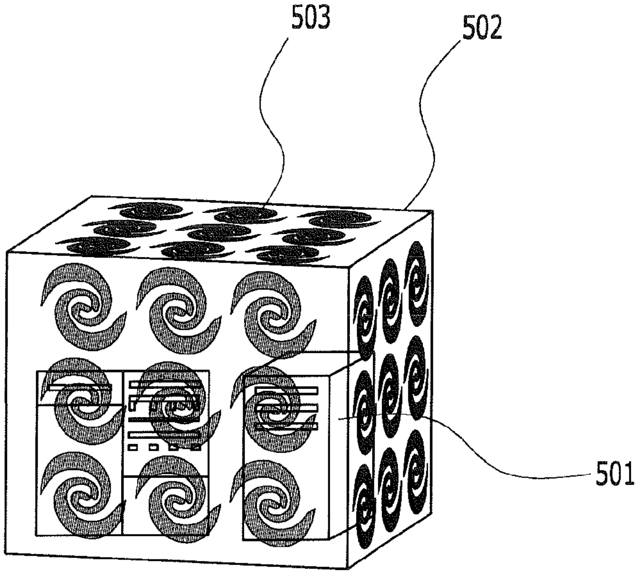

FIG. 6

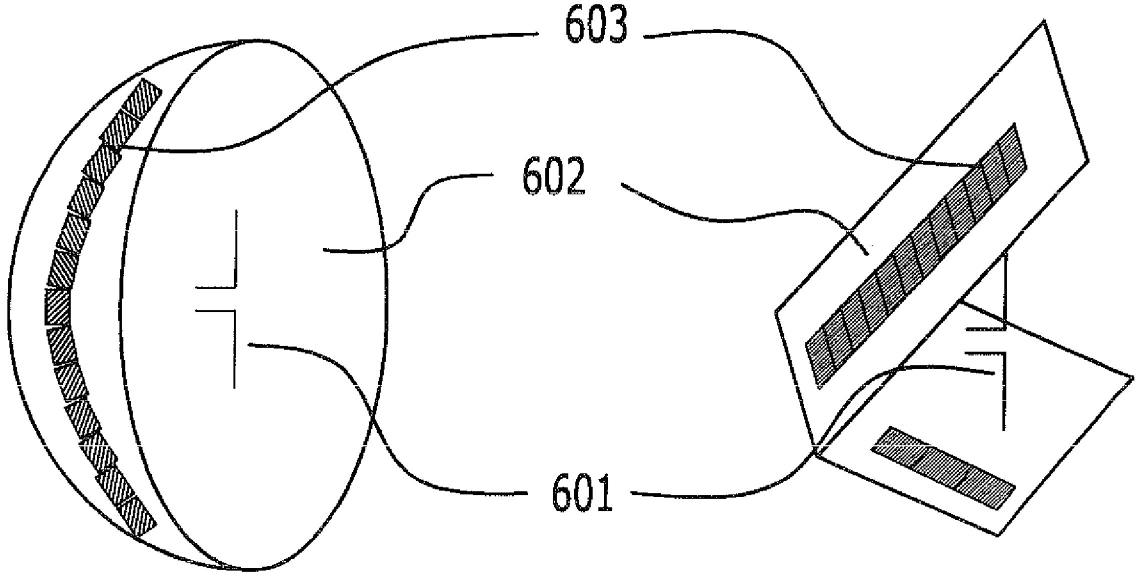

APPARATUS FOR HARVESTING ENERGY FROM MICROWAVE

CROSS-REFERENCE(S) TO RELATED APPLICATIONS

[0001] The present invention claims priority of Korean Patent Application No(s). 10-2009-0037190, filed on Apr. 28, 2009, which is incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION

[0002] 1. Field of the Invention

[0003] The present invention relates to an energy harvesting apparatus; and, more particularly, to an apparatus for harvesting energy from a microwave.

[0004] 2. Description of Related Art

[0005] Generally, a microwave is energy generated from a flow of electricity and magnetism and is also called a radio wave. That is, an electric field and a magnetic field are simultaneously generated around an area where vibration occurs while electricity is flowing, and a wave generated while the electric field and the magnetic field vary periodically is called a microwave. Such a microwave exists everywhere.

[0006] A microwave may be classified into electric fields generated by a voltage, and magnetic fields generated by a current. The microwave is generated from any product using electricity, for example, power transmission/distribution lines, electric appliances, and so on.

[0007] An electric field is generated in a perpendicular direction from a source in a straight-line shape and is easily removed or weakened by trees, buildings, human being's skins, and so on. A magnetic field is formed in a circular shape around a source and is not easily removed or weakened by certain objects or materials.

[0008] A microwave is classified into a gamma ray, an X ray, an ultraviolet ray, a visible ray (light), an infrared ray, and a radio wave (ultrahigh frequency radio wave, high frequency radio wave, low frequency radio wave, and very low frequency radio wave) in descending order of frequency. Accordingly, the radio wave refers to a microwave with a frequency of 3,000 GHz or less, and it is indispensable to daily life because it is used in various types. However, the very low frequency radio wave and the low frequency radio wave generate an electric field and a magnetic field. Hence, if a human body is exposed to the very low frequency radio wave or the low frequency radio wave for a long time, a body temperature is changed and a biorhythm is unbalanced, which is very likely to cause diseases. For example, search results were reported that the very low frequency radio wave and the low frequency radio wave might cause the reduction in the number of sperms for men, and might cause menstrual irregularity and birth defects for women. Furthermore, since a brain tumor might be caused in serious cases, the World Health Organization (WHO) conducted investigations, and harmfulness of the microwave has been revealed.

[0009] Examples of microwaves around our surroundings may include effective isotropically radiated power (hereinafter, referred to as "EIRP") radiated from a near base station, a magnetic field generated around a high pressure power station or a transformer substation, and strong electromagnetic interference (hereinafter, referred to as "EMI") generated in communication/electricity systems with high power consumption. Therefore, there is a need for methods for ensuring safety from the microwave environmental problems. Moreover, methods for reproducing harmful microwave into electric energy have attracted attention.

[0010] Meanwhile, due to concerns about exhaustion of finite resources such as petroleum energy, there is a need for methods for ensuring stable and regular energy. The microwave always exists around us. Therefore, wireless power transmission and energy reproduction technologies using microwave and radio wave have attracted attention as new research fields due to the exhaustion of fossil energy.

[0011] A representative new energy reproduction technology is to use a solar heat. The technology using a solar heat is applied in many places. However, an energy reproduction technology using a solar heat has a large number of variable parameters. Sufficient energy can be obtained during the daytime, but energy is consumed at night. Also, the energy efficiency is changed according to climatic condition. Specifically, efficiency of energy is high when a solar heat is strong, but it is low when it rains or snows. An amount of energy produced is also always varied. Moreover, since the sun always moves from the east to the west, optimal energy collection is difficult.

[0012] Like the solar heat, a wind power also has a large number of variable parameters. A blowing wind rotates blades of a wind power generator and an electric generator is operated by the rotation of the blades, resulting in production of energy. However, when no wind blows, energy cannot be produced. Also, the wind power generator must be installed in a windy area. Like this, location requirements for the wind power generator are hard to meet.

[0013] Therefore, there is a need for stable energy supply, instead of variable energy production of a solar heat or wind power.

SUMMARY OF THE INVENTION

[0014] An embodiment of the present invention is directed to providing an apparatus for ensuring a stable energy source.

[0015] Another embodiment of the present invention is directed to providing an energy harvesting apparatus capable of solving energy harvesting limitation caused by existence of variable parameters and so on.

[0016] Another embodiment of the present invention is directed to providing an energy harvesting apparatus capable of reducing influence of microwave on a human body.

[0017] In accordance with an aspect of the present invention, there is provided an apparatus for harvesting energy from a microwave in an electricity storage system of a building, which includes: a plurality of rectennas arranged with a predetermined length according to characteristic of microwave to be absorbed, and configured to collect microwave in the atmosphere and convert the collected microwave into electric energy; a current converter configured to convert the electric energy outputted from the rectennas into a storable current; and a charging tank configured to store the current outputted from the current converter, wherein the arranged rectennas are attached to an outer wall of the building, and the rectennas attached to the top of the building are printed in a transparent conductor on the surface of a solar cell.

[0018] In accordance with another aspect of the present invention, there is provided a system for harvesting energy from a microwave in an electricity storage system, which includes: a plurality of rectennas having a predetermined length and configured to collect a magnetic field among microwaves radiated from a high voltage cable and convert the collected magnetic field into electric energy; a current converter configured to convert the electric energy outputted from the rectennas into a storable current; and a charging tank configured to store the converted current, wherein the rectennas are spaced apart from the high voltage cable by a predetermined distance and arranged in a form of a fence so that a magnetic field does not affect a human body.

[0019] In accordance with another aspect of the present invention, there is provided an apparatus for harvesting energy from a microwave radiated from a system with an antenna, which includes: a plurality of rectennas configured to absorb a microwave radiated in an unwanted direction among microwaves radiated from the antenna and convert the absorbed microwave into electric energy; a current converter configured to convert the electric energy outputted from the rectennas into a storable current, and a charging tank configured to store the current outputted from the current converter, wherein the rectennas are spaced apart from the antenna by a predetermined distance in the unwanted direction and arranged in a certain form to shield the microwave.

[0020] Other objects and advantages of the present invention can be understood by the following description, and become apparent with reference to the embodiments of the present invention. Also, it is obvious to those skilled in the art to which the present invention pertains that the objects and advantages of the present invention can be realized by the means as claimed and combinations thereof.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] FIG. **1** is a conceptual diagram explaining a method for harvesting energy from a microwave by using an existing solar cell system in accordance with an embodiment of the present invention.

[0022] FIG. **2** is a conceptual diagram explaining an energy harvesting method using a microwave radiated from a base station in accordance with an embodiment of the present invention.

[0023] FIG. **3** is a conceptual diagram explaining an energy harvesting method using a microwave generated around a high pressure power station and a transformer substation in accordance with an embodiment of the present invention.

[0024] FIG. **4** illustrates an array structure of a rectenna for harvesting energy in accordance with an embodiment of the present invention.

[0025] FIG. **5** is a conceptual diagram explaining an energy harvesting apparatus using an EMI microwave generated from high-power wireless facilities and communication equipment in accordance with an embodiment of the present invention.

[0026] FIG. **6** is a conceptual diagram explaining an energy harvesting apparatus using microwave energy radiated from side lobes among microwaves radiated from a general antenna in accordance with an embodiment of the present invention.

DESCRIPTION OF SPECIFIC EMBODIMENTS

[0027] The advantages, features and aspects of the invention will become apparent from the following description of the embodiments with reference to the accompanying drawings, which is set forth hereinafter. Exemplary embodiments set forth herein may be modified in many different forms, and these embodiments are provided so that this disclosure will fully convey the concept of the invention to those skilled in the art. In the drawings, antennas, devices, or buildings are exaggerated or illustrated in different forms for clarity.

[0028] Prior to description of embodiments with reference to the accompanying drawings, an overview of the invention will be presented.

[0029] Wireless microwaves always exist around us. Examples of the wireless microwaves include microwave generated from a power station, EMI generated from a wireless device, and RF wave, microwave and millimeter wave always existing in the atmosphere.

[0030] The present invention relates to an apparatus for harvesting energy from these microwaves. Unlike a solar heat, microwaves always exist. Meanwhile, damage is often caused by EMI, or microwave environment is frequently misunderstood. Therefore, the present invention is directed to providing methods for ensuring stable and regular energy around us, under serious concerns about microwave environment problems and exhaustion of petroleum energy.

[0031] Methods for harvesting energy in accordance with embodiments of the present invention will be described below.

[0032] A first embodiment of the present invention relates to a method that uses an existing power collector where a rectenna is further installed in a solar cell which has been widely applied. The rectenna is a power collecting antenna made of a transparent conductor.

[0033] A second embodiment of the present invention relates to a method that generates new electricity by using EIRP radiated from a near base station and ensures safety from microwave outputted from the base station.

[0034] A third embodiment of the present invention relates to a method that ensures safety around a high pressure power station and a transformer substation, while reproducing electric energy from a magnetic field, which is a main cause of microwave generated at a high pressure power station.

[0035] A fourth embodiment of the present invention relates to a microwave energy reproducing method employing a method for shielding strong EMI generated in a communication and electric system with high power consumption.

[0036] A fifth embodiment of the present invention relates to an electric energy reproducing method that uses energy corresponding to side lobes among radiation patterns in order to control and shield microwave radiated from a communication repeater, a broadcast repeater and a base station to other areas.

First Embodiment

[0037] FIG. **1** is a conceptual diagram explaining a method for harvesting energy from a microwave by using an existing solar cell system in accordance with a first embodiment of the present invention.

[0038] Generally, most private homes or offices such as buildings receive AC power from an external transformer **108**, and most electronic appliances use electric energy by employing an AC-DC converter. However, since a large amount of energy is lost during the conversion process, many studies have been conducted to reduce the number of AC-DC conversions by providing a DC power supply system.

[0039] Therefore, in the new energy reproduction technology, energy is stored and used by employing a DC storage battery for a DC power supply.

[0040] A representative new energy reproduction technology is to use the sun **101**. Electricity is generated by using light and heat radiated from the sun **101**. As the method using the sun **101**, there is a solar heat technology and a photovoltaic power generation technology. In the following description, the sun **101** means both the solar light and the solar heat. In the case where the sun **101** means the solar light and the case where the sun **101** means the solar heat, the same reference numeral is used.

[0041] The solar heat technology uses a heat collector to convert solar radiant energy into heat energy, and the photovoltaic power generation technology uses a solar cell to convert a solar radiant energy into an electric energy. The solar heat technology is a photothermal field using a wave property of the sun's rays and is applied to air conditioning of buildings and water heating through absorption, storage and heat conversion of the solar heat **101**. The photovoltaic power generation technology directly converts the solar light **101** into electric energy and generates power by using a solar cell which generates electricity due to the photoelectric effect when receiving the sunlight.

[0042] A photovoltaic power generation system includes a solar cell module **107**, a rectifier **104** used as a power converter, and a charging tank **105** configured to store a rectified DC current.

[0043] The solar cell module **107** is installed in the roof or exterior of the building, where an amount of sunshine is sufficient, in order to receive a large amount of the solar light **101**. Energy generated from the solar cell module **107** by the use of the solar light **101** rectifies an AC current into a DC current having one direction through the rectifier **104**. The DC current outputted from the rectifier **104** is stored in the charging tank **105** and used in a computer **106** or other household electric appliances.

[0044] In addition to the solar heat **101**, a wind power generator **103** is used as another new and renewable energy technology. The wind power generator **103** generates induced electricity by converting the kinetic energy of the wind into torque and supplies the induced electricity to power systems or consumers.

[0045] The wind power generator **103** includes a momentum conversion system configured to absorbing and converting energy of the wind power, a power transmission system, a power conversion system, and a control system. The respective components that store the electric energy by using the wind power are mutually coupled to function as a whole system. The induced electricity generated during the execution of the function is converted into a DC current trough the rectifier **104** and stored in the charging tank **105**.

[0046] The current stored in the charging tank **105** by the solar heat **101** or the wind power generator **103** may be used in household electric appliances. However, the solar heat **101** or the wind power generator **103** has a changeability, that is, the quantity of energy generated is changed according to circumstances.

[0047] Microwave has a very low energy, compared with the wind power or the solar heat **101**, but it radiates the same energy at all times. Also, since microwave has less variable factors, the system can be easily configured. Furthermore, compared with the solar cell using semiconductor, the antenna configuration is inexpensive and simple.

[0048] The rectenna, which is an antenna configured to collect power from microwave, is a compound word of "antenna" and "rectifier." The antenna is configured to receive microwave, and the rectifier is configured to convert the received microwave into electric energy. In order to receive maximum power, the rectenna has a matching circuit and input/output impedance different from those of a general antenna. Like the general antenna, the rectenna may be fabricated in various configurations. Among them, a patch configuration is advantageous to a packaging and easy to fabricate in an array form because of its two-dimensional structure.

[0049] In addition, the rectenna may be made of a transparent conductor or a general conductor. Power generated through the rectenna may be used in the rectifier **104** and the charging tank **105**, which is the existing renewable energy system, without modifications. Hence, additional installation and fabrication costs can be remarkably reduced.

[0050] Furthermore, the rectenna may be fabricated using an inexpensive painting scheme. The rectenna configured by applying the painting scheme to a transparent conductor may be fabricated in a planar type and used by attaching on the solar cell module **107**. If the rectenna is implemented with the transparent conductor, energy can be ensured twofold by overlapping the rectenna on the solar cell. In addition, the rectenna may be attached to the outer wall of the building, the top of the building, or the window where energy can be easily collected. The rectenna may convert the absorbed microwave into electric energy. If a desired microwave is absorbed through the rectenna, the rectenna resonates at a specific frequency. Accordingly, if using a frequency function, the length of the rectenna may be determined according to the frequency of the microwave.

[0051] Moreover, as illustrated in FIG. **4**, the rectenna may be arranged in an array form in order to improve energy received through the power collecting antenna.

[0052] FIG. **4** illustrates an array structure of a rectenna for harvesting energy in accordance with an embodiment of the present invention.

[0053] In the case of one energy collecting antenna **402**, an amount of energy collected is small. Hence, a plurality of antennas **402** are arranged in an array form. Compared with the case of using one antenna **402**, an energy collecting antenna unit **401** configured in an array form can increase an amount of energy received. When designing the energy collecting antenna unit **401** in the array form, the antennas **402** must be arranged considering antenna area and inter-antenna interference.

[0054] The energy collecting antenna **401** may be fabricated not in an antenna structure but in a resonance structure or an electromagnetic induction structure. In particular, in the case of using the rectenna to collect low frequency microwave, the resonance structure and the electromagnetic induction structure may be advantageous.

[0055] As described above, the use of the power collecting antenna can solve the problems of microwave generated inside the house or existing around the house. That is, various kinds of microwaves generated in a living space can be removed, and the electric energy can be reproduced by using the microwave.

Second Embodiment

[0056] FIG. **2** is a conceptual diagram explaining a method for harvesting energy from a microwave radiated from a base station in accordance with a second embodiment of the present invention.

[0057] For convenience of life, a large number of mobile communication base stations **201** are installed around us. However, the intensity of microwave will be very strong if the mobile communication base stations **201** are installed at locations very near from the house or village **204**. The above-described example is not limited to the mobile communication base station, and may include a broadcast repeater and so on. In this case, a shield **205** may be installed in order for stable communication and protection of the house or village **204** from the strong microwave **203**.

[0058] Like the embodiment of FIG. **4**, the shield **205** may be configured in an array form, or may be made of a transparent conductor and installed over the village, considering an amount of sunshine. The shield **205** is not limited to the embodiment of FIG. **5**, and may be modified in various forms. For example, the shield **205** may be installed in a fence form.

[0059] Furthermore, the installation of the shield **205** makes it possible to receive the strong microwave **203** radiated from the mobile communication base station **201** and then reproduce energy.

[0060] Generally, the energy radiated from the mobile communication base station **201** is EIRP of more than 10,000 W. The radiant energy decreases with the inverse square of distance (d). As one example, in the case of using the existing rectenna technology and charging technology, the energy of 480 mW is produced at the distance of 1 m in the antenna area of 6×6 cm^2, the energy of 4.8 mW is produced at the distance of 10 m, and the energy of 48 μW is produced at the distance of 100 m. Also, the energy of 13.3 mW is produced at a distance of 100 m in the antenna area of 1 m^2, and the energy of 1.3 W is generated in the antenna area of 10 m^2.

Third Embodiment

[0061] FIG. **3** is a conceptual diagram explaining a method for harvesting energy from a microwave generated around a high pressure power station and a transformer substation in accordance with a firth embodiment of the present invention.

[0062] The high pressure power station or transformer substation **301** or the high-voltage power transmission line **302** is considered as obnoxious facilities, and persons are reluctant to live around there. This is because persons are concerned about damage from microwave exposure because low frequency microwave is generated from those places. A low frequency generates a strong magnetic field. The control of such a magnetic field may be impossible even by the installation of a high fence.

[0063] Therefore, the method of FIG. **3** is to keep a magnetic field, which is a main cause among microwaves generated at the high pressure power station, reproduce electric energy, ensure the safety around the transformer substation, and harvest energy by using the magnetic field.

[0064] As an antenna for detection of a magnetic field, a low-impedance loop antenna is generally used. Like the embodiment of the present invention, when a frequency is very low just like microwave radiated from the high pressure power station and the transformer substation, magnetic field components can be collected by using a resonant loop antenna or an inductive coil.

[0065] In this embodiment, the resonant loop antenna, the magnetic resonator, and the electromagnetic induction coil for removing the magnetic field **307** may be fabricated in a form indicated by reference numeral **304**, and they are arranged around the fence **303** so that microwave is prevented from leaking out of the high pressure power station **301** and the high-voltage power transmission line **302**. Also, the magnetic field **307** and the microwave collected at the fence **303** configured with the above-described components are prevented from leaking out and converted into electric energy. That is, by providing a reproduction unit (not shown) for storing energy by converting energy of the magnetic field and so on collected at the fence **303** into an electric form, energy stored in the electric condenser (not shown) may be used.

[0066] Furthermore, by installing the shield **306** as well as the fence **303** at the village **305** near the high pressure power station **301** or the high-voltage power transmission line **302**, like the embodiment of FIG. **2**, energy can be reproduced by collecting microwave that is not prevented by the fence **303**. Moreover, the installation of the shield **306** can protect the house or village **305** from microwave.

Fourth Embodiment

[0067] FIG. **5** is a conceptual diagram explaining an energy harvesting apparatus using EMI microwave generated from high-power wireless facilities and communication equipment in accordance with a fourth embodiment of the present invention.

[0068] The high-power wireless facilities or communication/electric system **501** consumes high power, which may cause strong EMI. Therefore, a method for protecting from EMI and reproducing energy by collecting EMI noises will be described below.

[0069] An energy collecting antenna **503** is installed in the outside of the high-power wireless facilities or communication/electric system **501** in order to remove EMI noises generated when driving the high-power wireless facilities or communication/electric system **501**. As described above, the energy collecting antenna **503** may be implemented in an array antenna type, and a printing scheme may be applied on a transparent substrate. In the case of applying the printing scheme on the transparent substrate as described above, operation states of the high-power wireless facilities or communication/electric system **501** can be easily confirmed. Furthermore, microwave radiated from the high-power wireless facilities or communication/electric system **501** can be removed. Moreover, after collecting microwave, the collected microwave may be converted into electric energy and then stored, or may be immediately supplied to other equipment as an energy source.

[0070] In the antenna array arrangement method, as illustrated in FIG. **5**, antennas are arranged on all sides in order to completely shield EMI generated in the system. In this manner, leaking noises can be reduced, and energy can be reproduced by using EMI leaking from the system.

Fifth Embodiment

[0071] FIG. **6** is a conceptual diagram explaining an energy harvesting apparatus using microwave energy radiated from side lobes among microwaves radiated from a general antenna in accordance with a fifth embodiment of the present invention.

[0072] The fifth embodiment of the present invention is directed to providing a method for reproducing electric energy by using energy corresponding to side lobes among radiation patterns in order to control and shield microwave radiated from the communication/broadcast repeater and the base station to other areas.

[0073] In this embodiment, an energy collecting antenna **603** is installed in the rear side of the communication/broadcast antenna **601** or the region belonging to the side lobes. The installation of the energy collecting antenna **603** makes it possible to obtain energy by using microwave unnecessary to communication. The receiving energy can be increased by implementing the energy collecting antenna **603** in an array antenna type. A reproduction unit (not shown) may be provided for converting the collected energy into an electric form and then storing energy, and the energy stored in an electric condenser (not shown) can be used.

[0074] Moreover, the radiation direction of the communication/broadcast antenna **601** can be stabilized by adjusting the position and angle of the energy collecting antenna **603**. That is, the energy collecting antenna **603** is installed at the position corresponding to the side lobes among the radiation patterns of the communication/broadcast antenna **601**, considering the angle of the energy collecting antenna **603**. By installing the energy collecting antenna **603** in the above-described manner, microwave radiated to other areas is blocked. Thus, the interference problem can be solved, and the energy can be harvested.

[0075] The energy harvesting apparatus in accordance with the embodiments of the present invention can ensure a stable energy source, solve energy harvesting limitation caused by existence of variable parameters and so on, and reduce influence of microwave on a human body.

[0076] While the present invention has been described with respect to the specific embodiments, it will be apparent to those skilled in the art that various changes and modifications may be made without departing from the spirit and scope of the invention as defined in the following claims.

What is claimed is:

1. An apparatus for harvesting energy from a microwave in an electricity storage system of a building, the apparatus comprising:

a plurality of rectennas arranged with a predetermined length according to characteristic of microwave to be absorbed, and configured to collect microwave in the atmosphere and convert the collected microwave into electric energy;

a current converter configured to convert the electric energy outputted from the rectennas into a storable current; and

a charging tank configured to store the current outputted from the current converter,

wherein the arranged rectennas are attached to an outer wall of the building, and the rectennas attached to the top of the building are printed in a transparent conductor on the surface of a solar cell.

2. The apparatus of claim **1**, further comprising a wind power generator configured to generate induced electricity by using wind power and provide the induced electricity to the charging tank.

3. The apparatus of claim **1**, further comprising a plurality of rectennas configured to collect microwave generated from a device generating microwave inside the building, convert the collected microwave into an electric current, and provide the electric current to the current converter.

4. The apparatus of claim **3**, wherein the rectennas are printed on a transparent conductor in array type, and four sides of the device generating the microwave are shielded by the rectennas, thereby preventing leakage of the microwave.

5. The apparatus of claim **1**, wherein, in case where a mobile communication repeater or a broadcast repeater is installed in the outside or top of the building, the apparatus further comprises a plurality of array rectennas configured to prevent microwaves generated at the mobile communication repeater or the broadcast repeater from flowing into the building, and to convert the microwave generated at the mobile communication repeater or the broadcast repeater into electric energy.

6. A system for harvesting energy from a microwave in an electricity storage system, the system comprising:

a plurality of rectennas having a predetermined length and configured to collect a magnetic field among microwaves radiated from a high voltage cable and convert the collected magnetic field into electric energy;

a current converter configured to convert the electric energy outputted from the rectennas into a storable current; and

a charging tank configured to store the converted current,

wherein the rectennas are spaced apart from the high voltage cable by a predetermined distance and arranged in a form of a fence so that a magnetic field does not affect a human body.

7. The system of claim **6**, wherein the fence is made of a transparent conductor, and the array rectennas are printed on the fence.

8. The system of claim **6**, wherein the fence is configured in a resonance structure or an electromagnetic induction structure.

9. The system of claim **6**, wherein, in case where a mobile communication repeater or a broadcast repeater is installed in the outside or top of the building near the electricity storage system, the system further comprises a plurality of array rectennas configured to collect a microwave generated from the mobile communication repeater or the broadcast repeater and convert the collected microwave into electric energy.

10. The system of claim **9**, wherein the array rectennas are spaced apart by a predetermined distance and arranged in a form of a fence so that the microwave does not affect a human body.

11. An apparatus for harvesting energy from a microwave radiated from a system with an antenna, the apparatus comprising:

a plurality of rectennas configured to absorb a microwave radiated in an unwanted direction among microwaves radiated from the antenna and convert the absorbed microwave into electric energy;

a current converter configured to convert the electric energy outputted from the rectennas into a storable current, and

a charging tank configured to store the current outputted from the current converter,

wherein the rectennas are spaced apart from the antenna by a predetermined distance in the unwanted direction and arranged in a certain form to shield the microwave.

12. The apparatus of claim **11**, wherein, when the antenna is a parabolic antenna, the rectennas are arranged on a rear plate of the parabolic antenna to absorb a microwave radiated in a rear direction among microwaves radiated from the parabolic antenna and convert the absorbed microwave into electric energy.

13. The apparatus of claim **12**, wherein the rectennas are implemented on a transparent conductor by using a printing scheme.

14. The apparatus of claim **11**, wherein, when the antenna is a dipole antenna, the rectennas are arranged to absorb a microwave radiated in an unwanted direction among microwaves radiated from the antenna and convert the absorbed microwave into electric energy.

15. The apparatus of claim **14**, wherein the rectennas are implemented on a transparent conductor by using a printing scheme.

* * * * *

US 20110018516A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2011/0018516 A1**

Notman et al. (43) **Pub. Date: Jan. 27, 2011**

(54) DC-DC CONVERTERS

(76) Inventors: **Andrew Notman**, Edinburgh (GB); **Mark McCloy-Stevens**, Edinburgh (GB)

Correspondence Address:
DICKSTEIN SHAPIRO LLP
1825 EYE STREET NW
Washington, DC 20006-5403 (US)

(21) Appl. No.: **12/841,688**

(22) Filed: **Jul. 22, 2010**

Related U.S. Application Data

(60) Provisional application No. 61/228,002, filed on Jul. 23, 2009, provisional application No. 61/261,123, filed on Nov. 13, 2009.

(30) **Foreign Application Priority Data**

Jul. 22, 2009 (GB) 0912745.7
Oct. 27, 2009 (GB) 0918788.1

Publication Classification

(51) **Int. Cl.**
G05F 1/00 (2006.01)

(52) **U.S. Cl.** **323/284**

(57) ABSTRACT

Methods and apparatus for control of DC-DC converters. The DC-DC converter is operable so that the low side supply switch may be inhibited from turning on in a cycle following the high side supply switch turning off. Turn on of the low side switch is inhibited if the time between turn off of the high side switch and the inductor (L) current reaching zero is less than a predetermined duration. Inhibiting the low side switch from turning on can prevent the inductor current from going negative, which would reduce the efficiency of the converter. When turn on of the low side switch is inhibited the inductor current flows through a parallel path, such as a parasitic body diode associated with the low side switch, which allows current flow in one direction only.

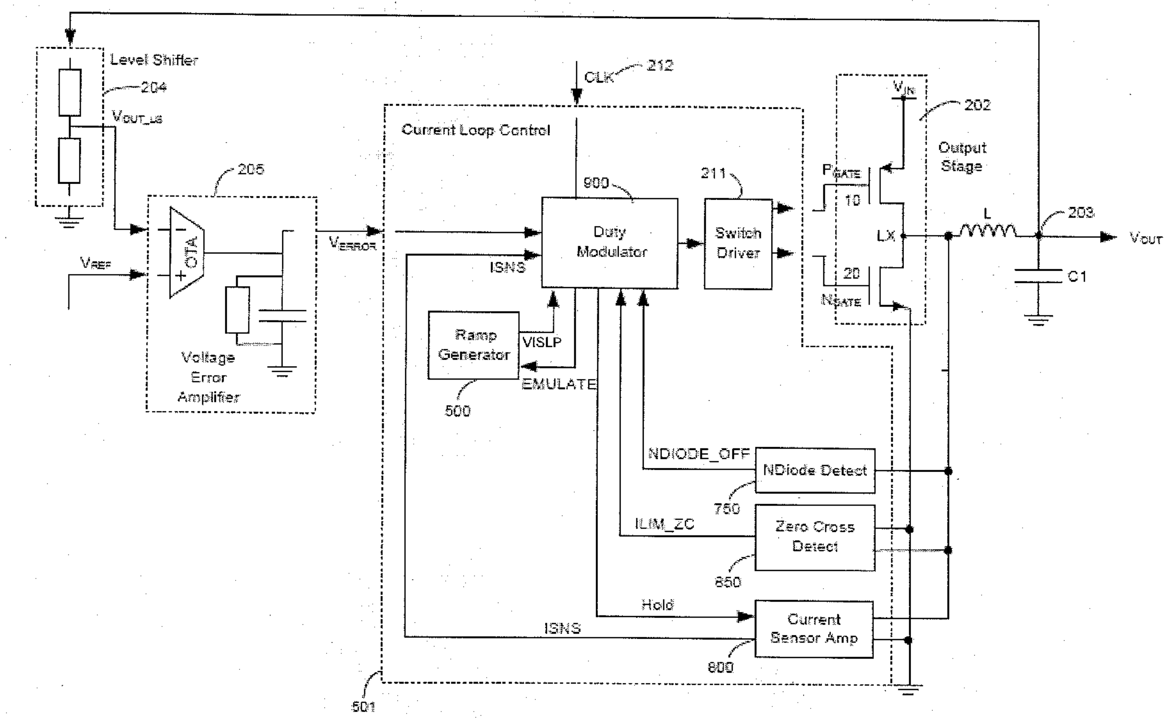
Level Shifter
204
V_{OUT_LS}
205
OTA
V_{REF}
Voltage Error Amplifier
V_{ERROR}
CLK
212
Current Loop Control
900
Duty Modulator
211
Switch Driver
P_{GATE}
10
20
N_{GATE}
V_{IN}
202
Output Stage
LX
L
203
V_{OUT}
C1
ISNS
Ramp Generator
VISLP
EMULATE
500
NDIODE_OFF
NDiode Detect
750
ILIM_ZC
Zero Cross Detect
850
Hold
Current Sensor Amp
800
ISNS
501

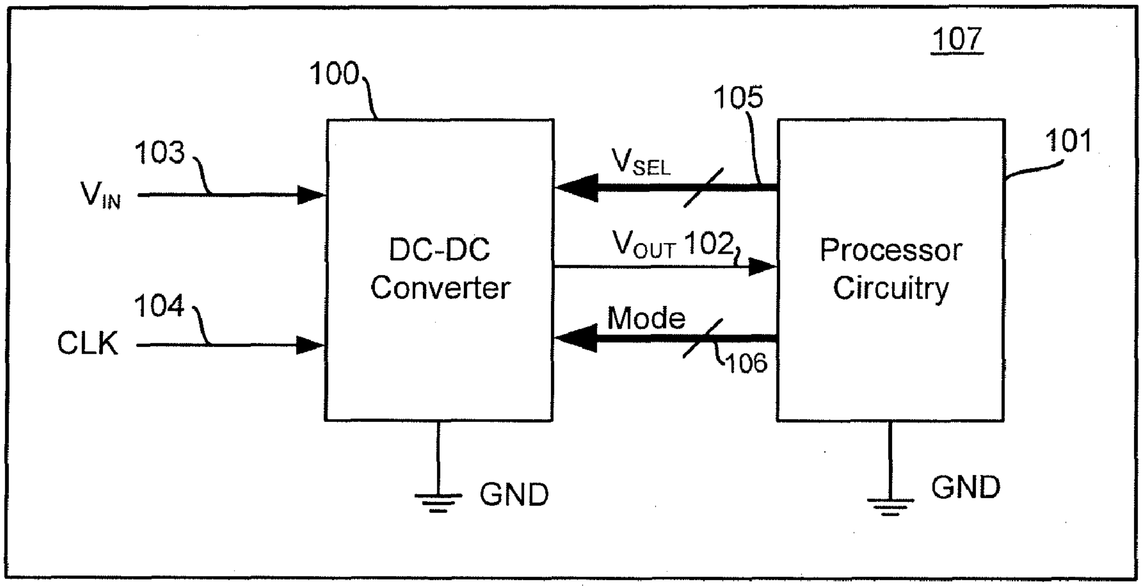
107
100
105
101
103
V_{IN}
V_{SEL}
DC-DC Converter
V_{OUT} 102
Processor Circuitry
104
CLK
Mode
106
GND
GND

Figure 1

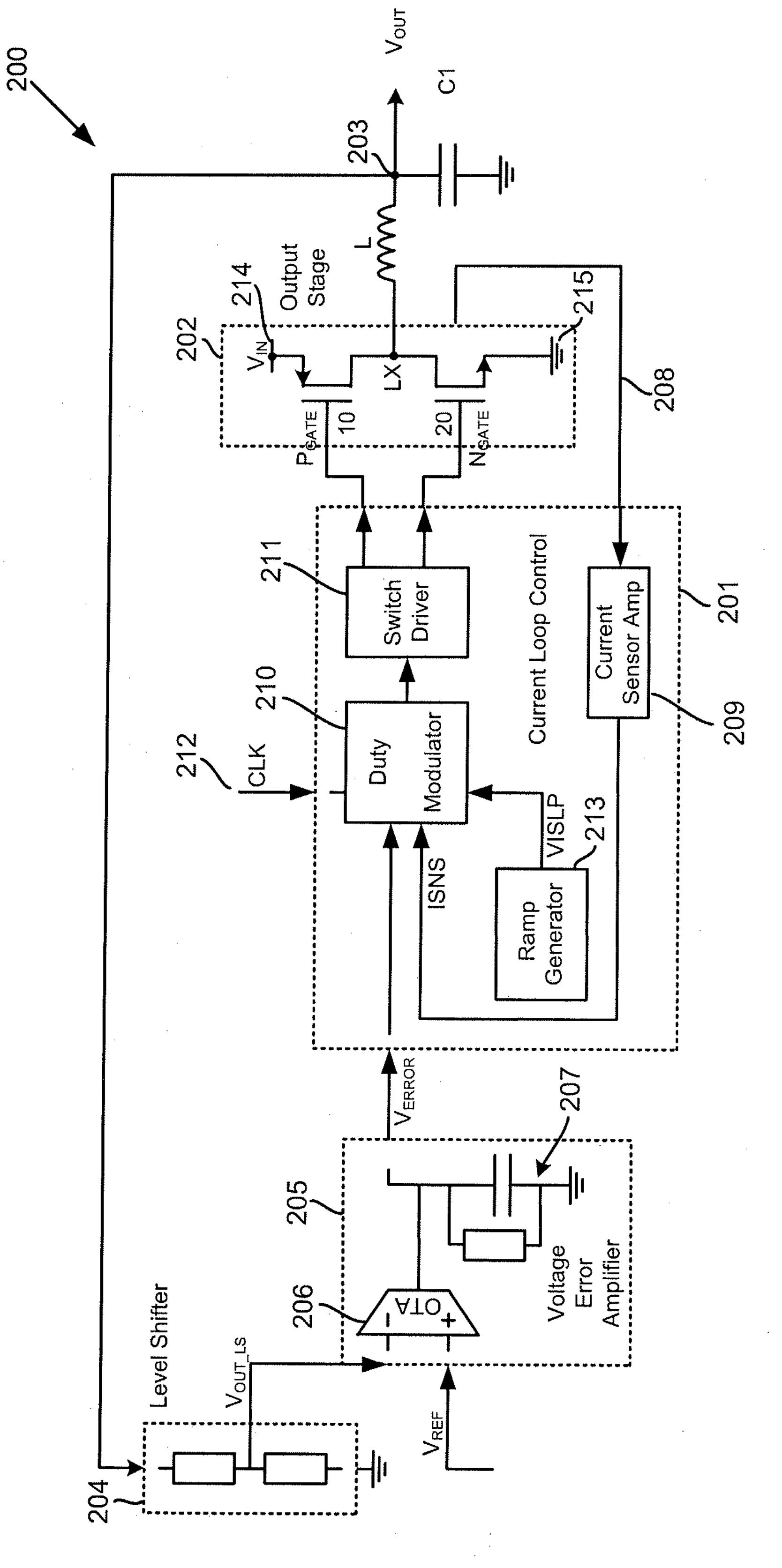
200
V_{OUT}
C1
203
L
214
Output Stage
202
V_{IN}
LX
215
10
20
P_{GATE}
N_{GATE}
208
211
Switch Driver
201
Current Loop Control
Current Sensor Amp
209
210
Duty Modulator
212
CLK
VISLP
213
ISNS
Ramp Generator
V_{ERROR}
207
205
206
OTA
-
+
Voltage Error Amplifier
Level Shifter
V_{OUT_LS}
V_{REF}
204

Figure 2

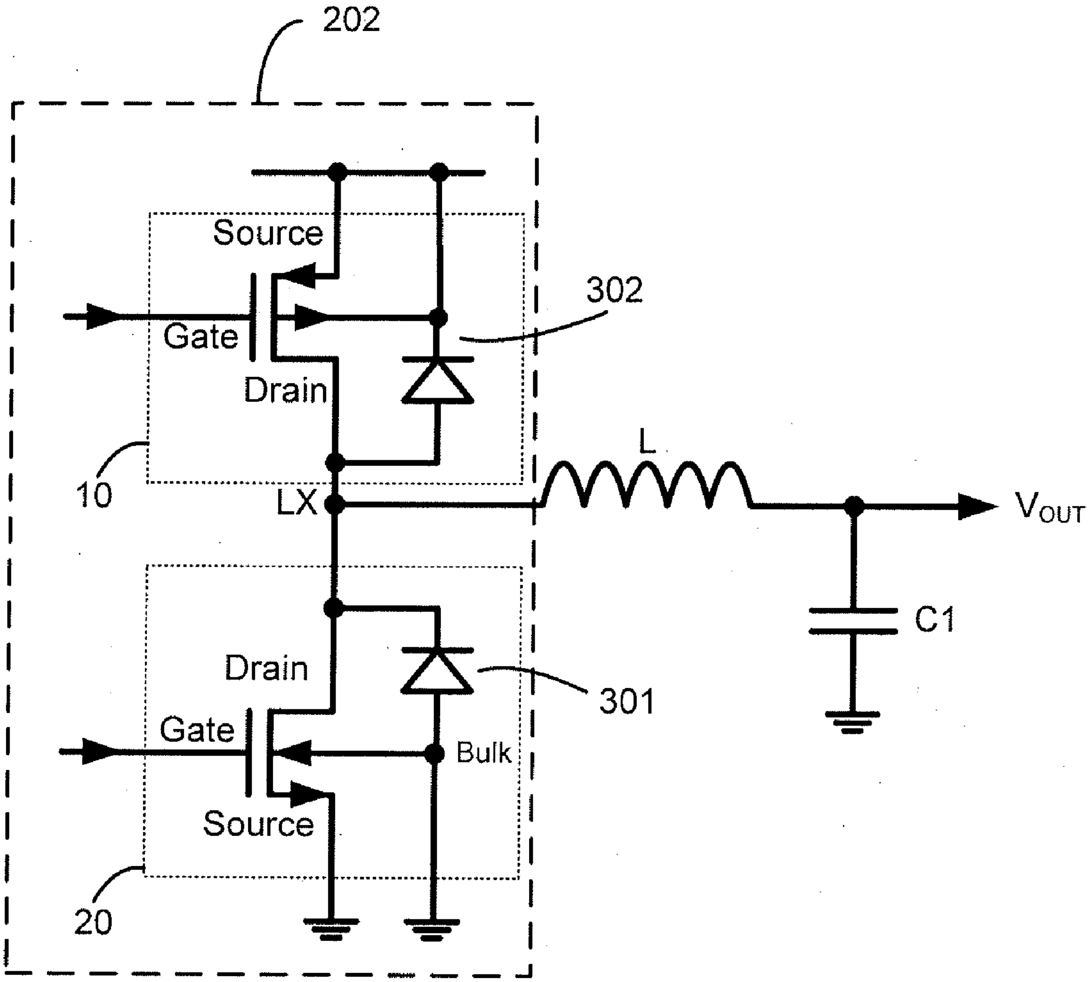
202
Source
Gate
Drain
302
10
LX
L
V_{OUT}
C1
Drain
Gate
301
Bulk
Source
20

Figure 3

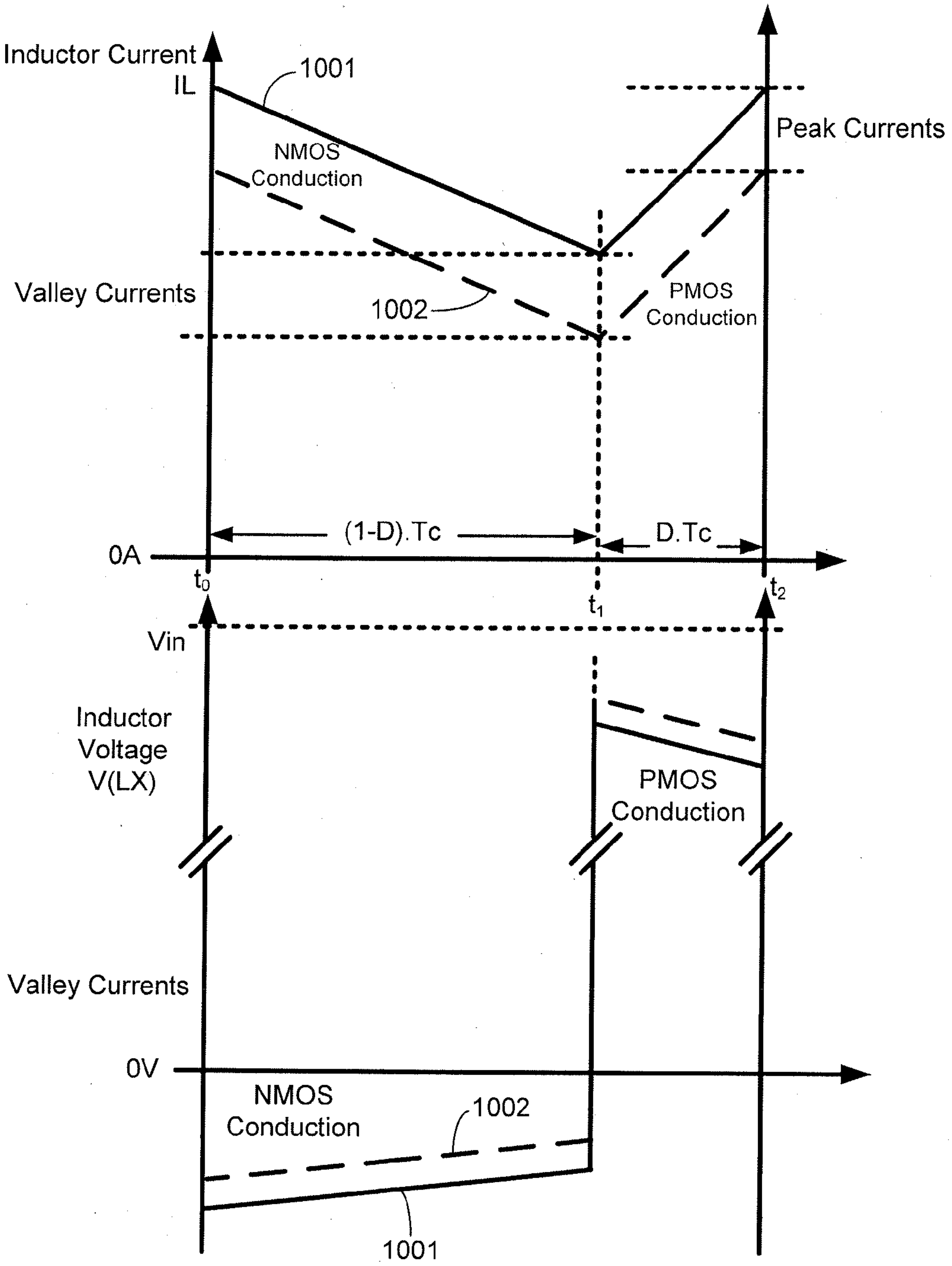
Inductor Current
IL
1001
NMOS
Conduction
Peak Currents
Valley Currents
1002
PMOS
Conduction
(1-D).Tc
D.Tc
0A
t_0
t_1
t_2
Vin
Inductor
Voltage
V(LX)
PMOS
Conduction
Valley Currents
0V
NMOS
Conduction
1002
1001

Figure 4

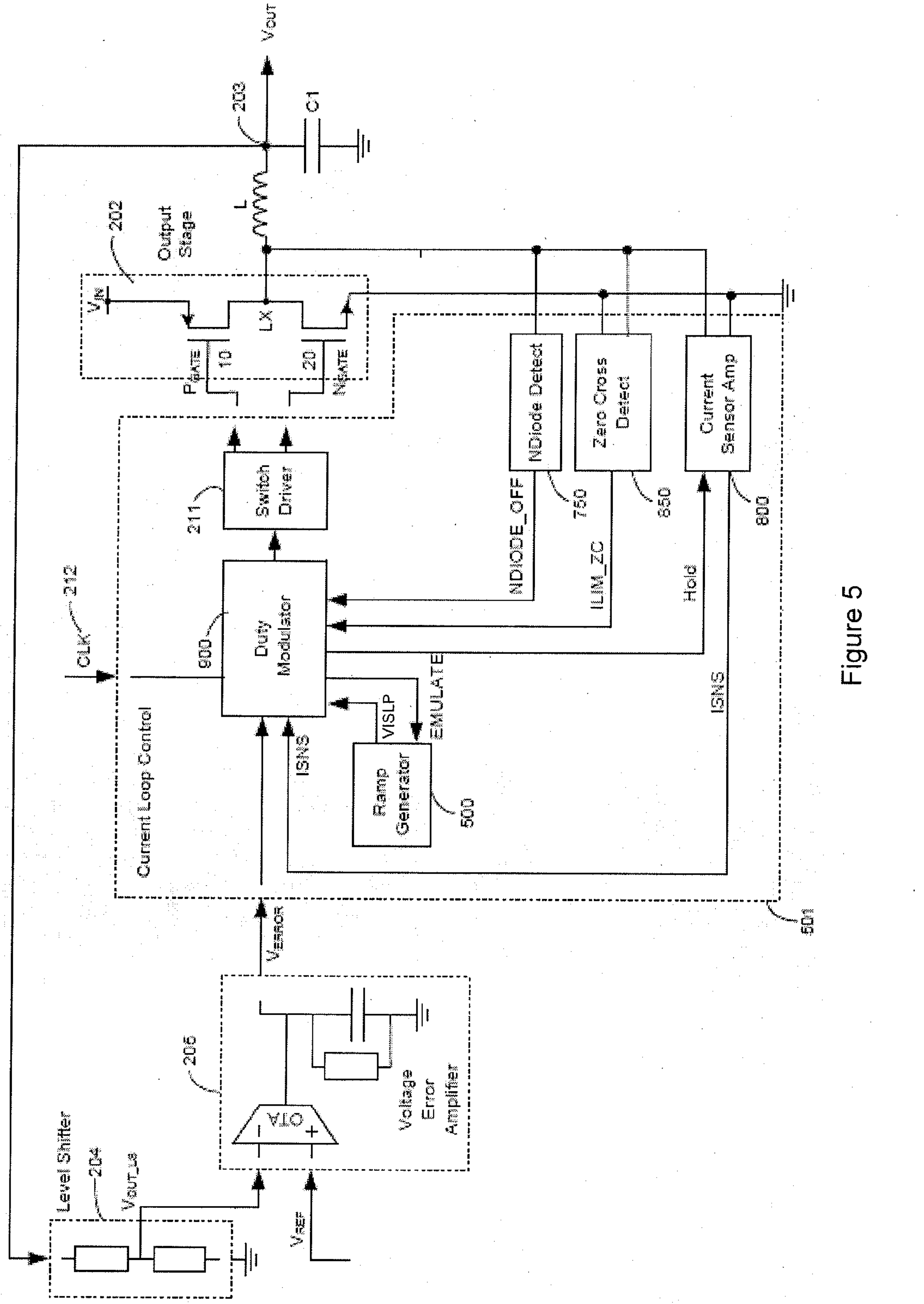

V_{OUT}
203
C1
202
Output Stage
L
V_{IN}
LX
P_{GATE}
10
20
N_{GATE}
NDiode Detect
Zero Cross Detect
Current Sensor Amp
750
850
800
NDIODE_OFF
ILIM_ZC
Hold
ISNS
Switch Driver
211
CLK
212
900
Duty Modulator
VISLP
EMULATE
ISNS
Ramp Generator
500
Current Loop Control
501
V_{ERROR}
206
OTA
-
+
Voltage Error Amplifier
Level Shifter
204
V_{OUT_LS}
V_{REF}

Figure 5

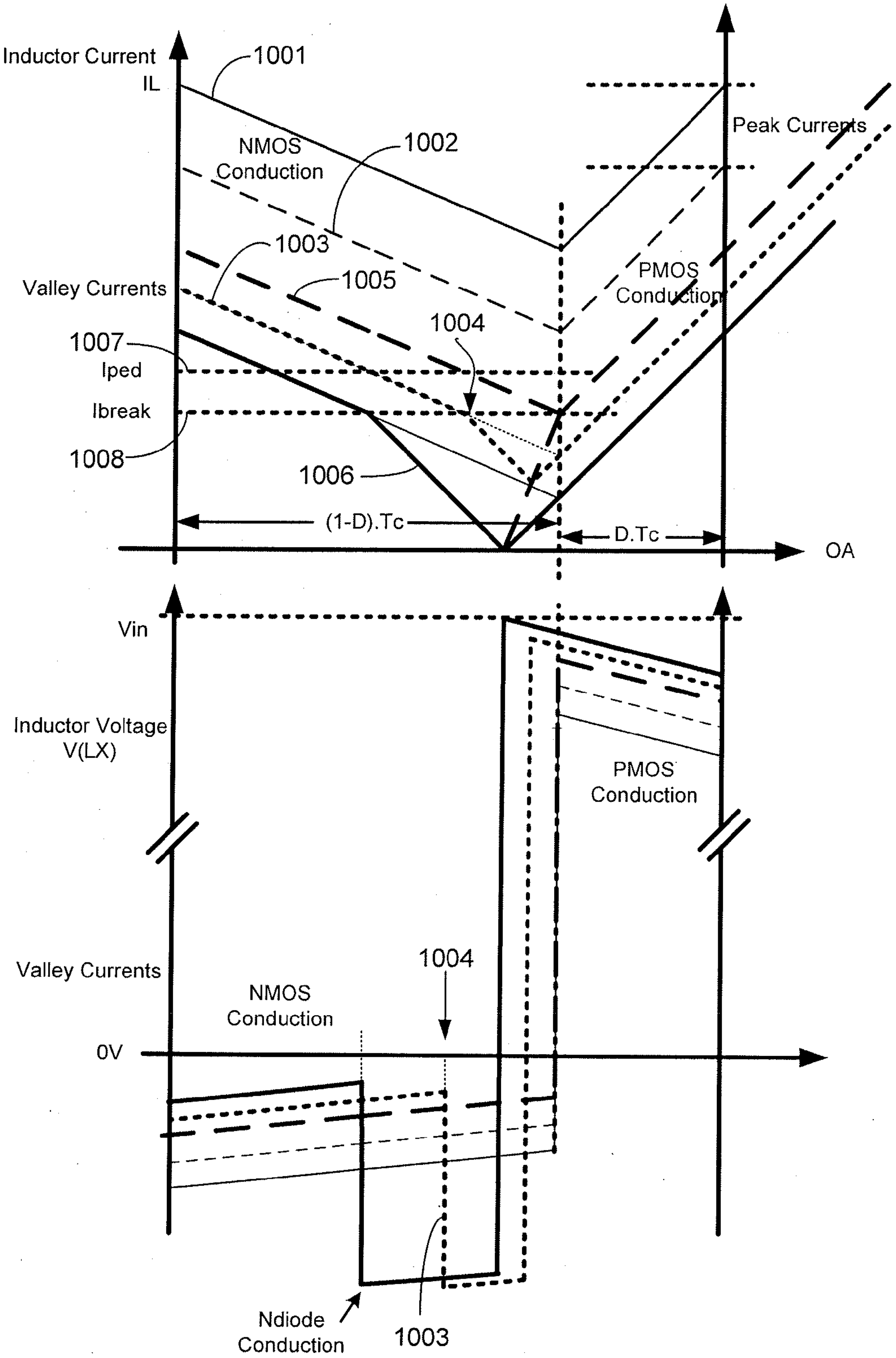
Inductor Current
IL
1001
NMOS
Conduction
1002
1003
Valley Currents
1005
1004
1007
Iped
Ibreak
1008
1006
(1-D).Tc
D.Tc
OA
Peak Currents
PMOS
Conduction
Vin
Inductor Voltage
V(LX)
PMOS
Conduction
Valley Currents
1004
NMOS
Conduction
0V
Ndiode
Conduction
1003

Figure 6

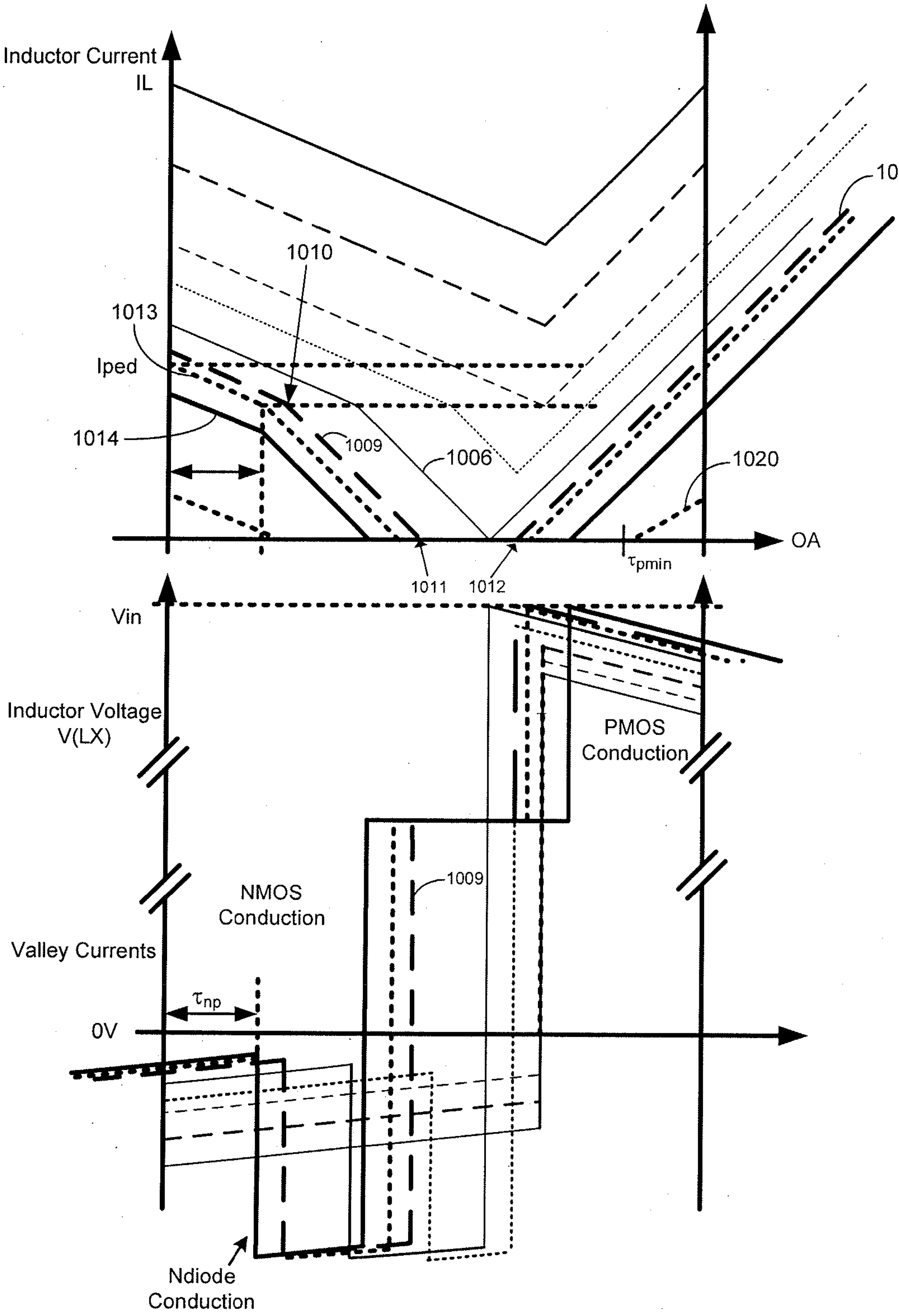
Inductor Current
IL
1010
1013
Iped
1014
1009
1006
1009
1020
OA
τ_{pmin}
1011
1012
Vin
Inductor Voltage
V(LX)
PMOS
Conduction
NMOS
Conduction
1009
Valley Currents
τ_{np}
0V
Ndiode
Conduction

Figure 7

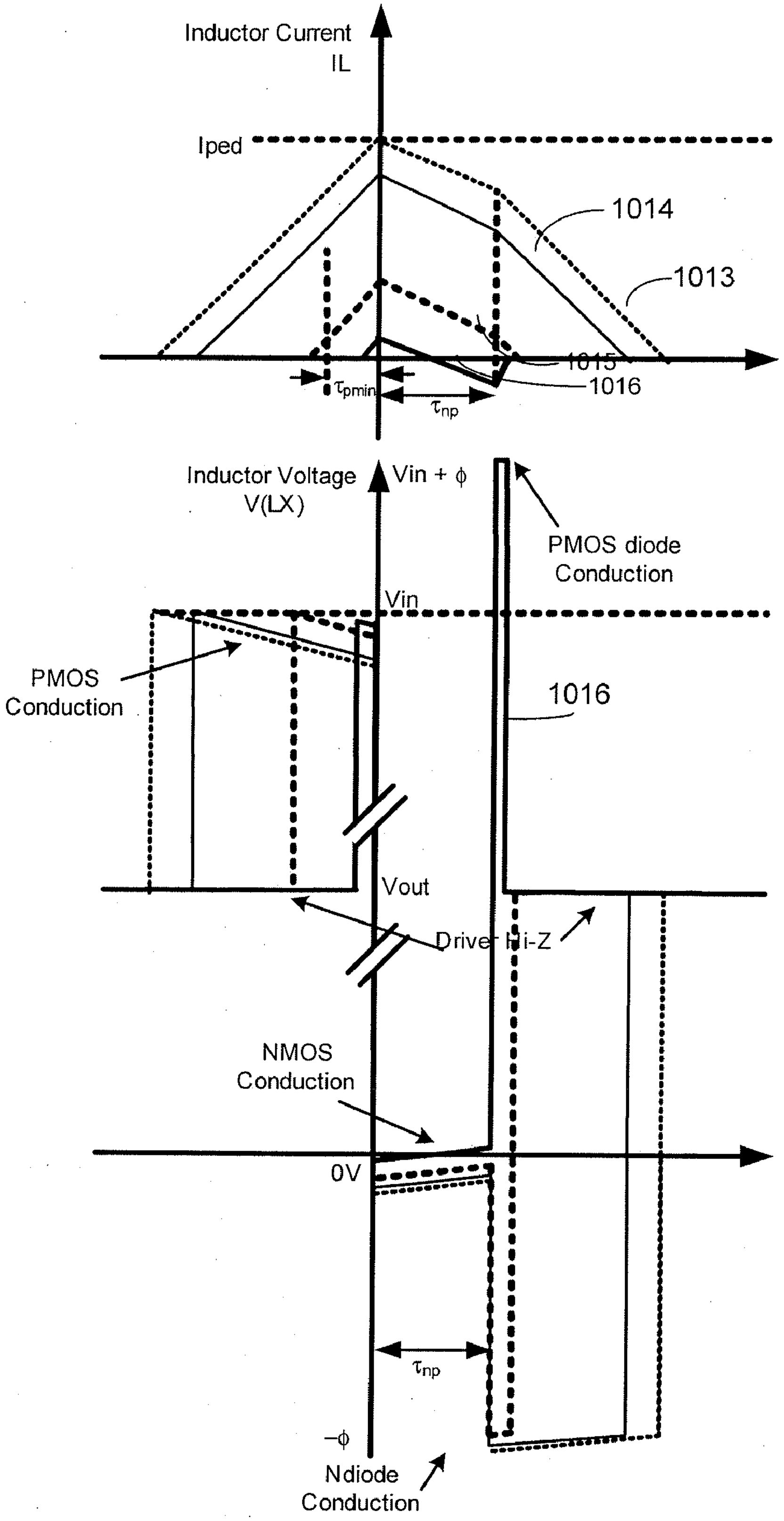
Inductor Current
IL
Iped
1014
1013
1015
1016
τ_{pmin}
τ_{np}
Inductor Voltage
V(LX)
Vin + ϕ
PMOS diode
Conduction
Vin
PMOS
Conduction
1016
Vout
Driver Hi-Z
NMOS
Conduction
0V
τ_{np}
−ϕ
Ndiode
Conduction

Figure 8

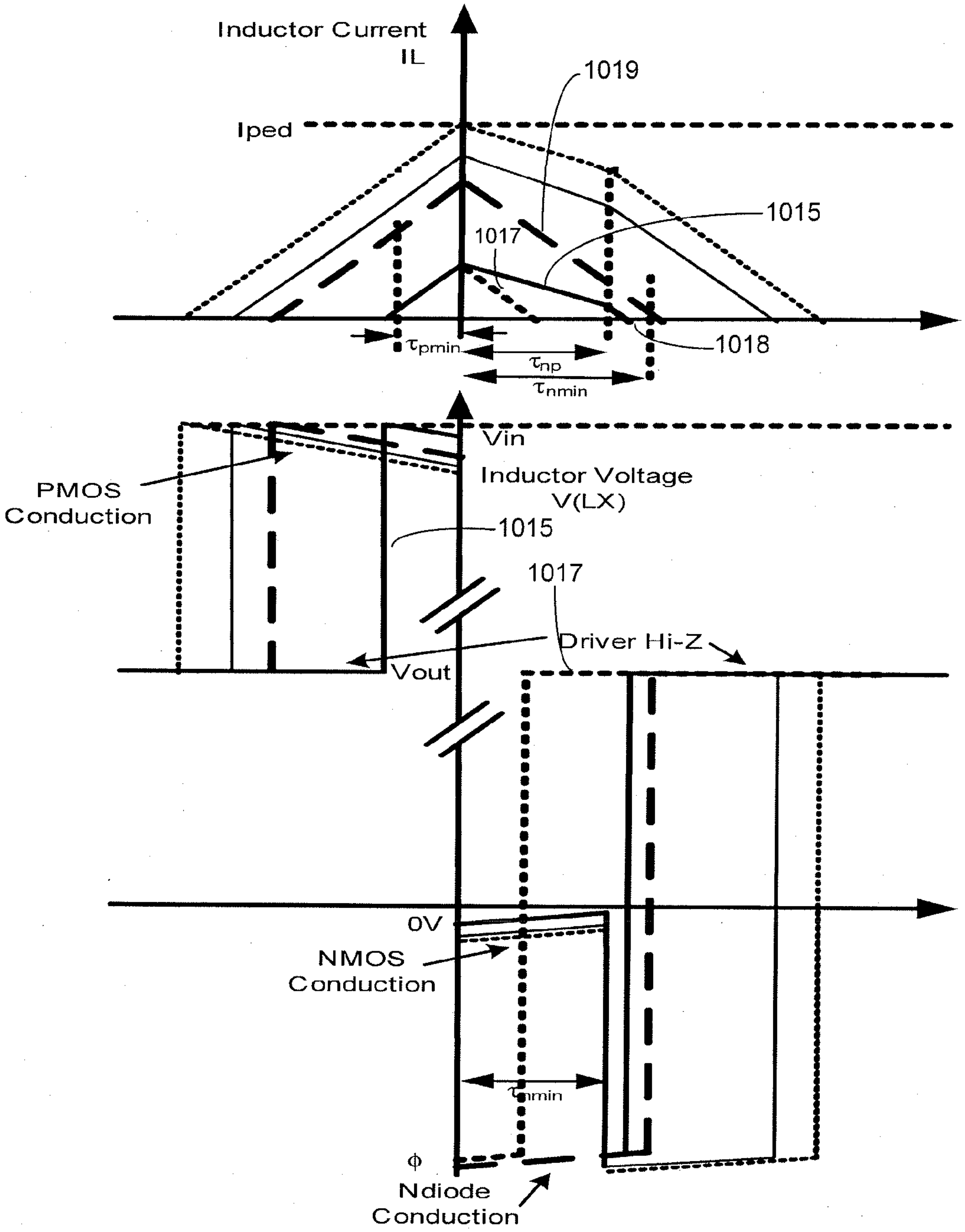
Inductor Current
IL
1019
Iped
1015
1017
1018
τ_{pmin}
τ_{np}
τ_{nmin}
Vin
Inductor Voltage
V(LX)
PMOS
Conduction
1015
1017
Driver Hi-Z
Vout
0V
NMOS
Conduction
τ_{nmin}
ϕ
Ndiode
Conduction

Figure 9

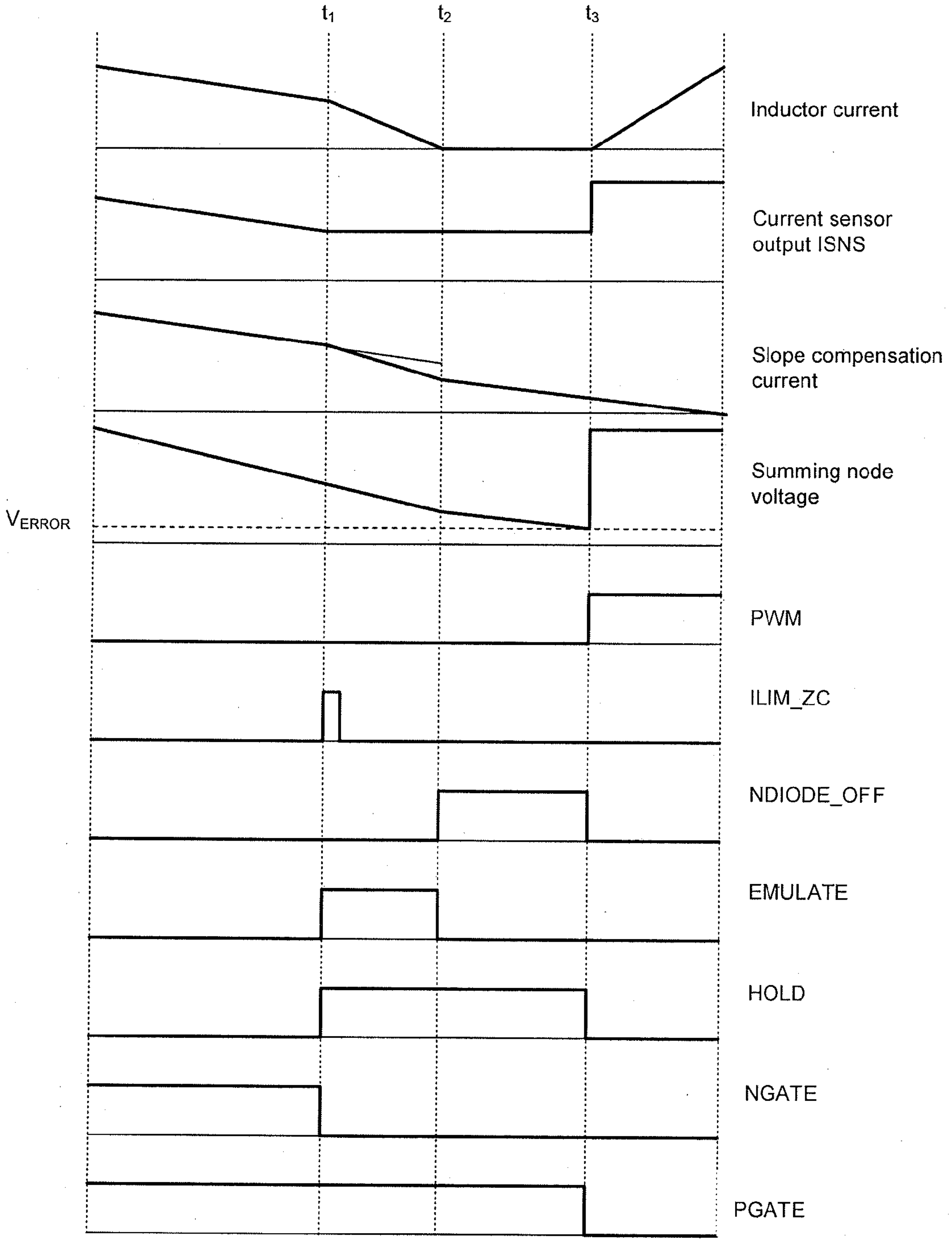

t_1
t_2
t_3
Inductor current
Current sensor output ISNS
Slope compensation current
Summing node voltage
V_{ERROR}
PWM
ILIM_ZC
NDIODE_OFF
EMULATE
HOLD
NGATE
PGATE

Figure 10a

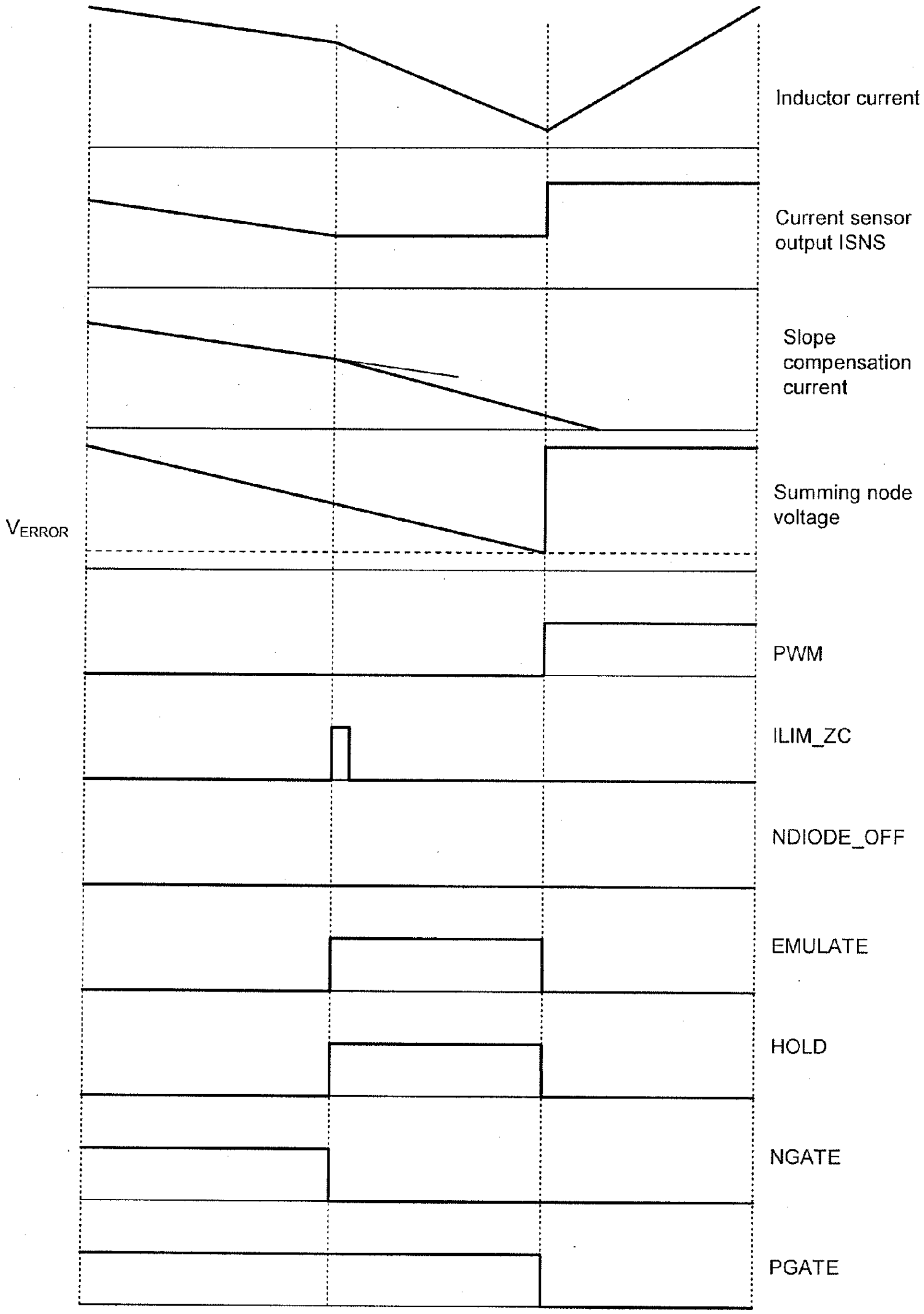
Inductor current
Current sensor output ISNS
Slope compensation current
V_{ERROR}
Summing node voltage
PWM
ILIM_ZC
NDIODE_OFF
EMULATE
HOLD
NGATE
PGATE

Figure 10b

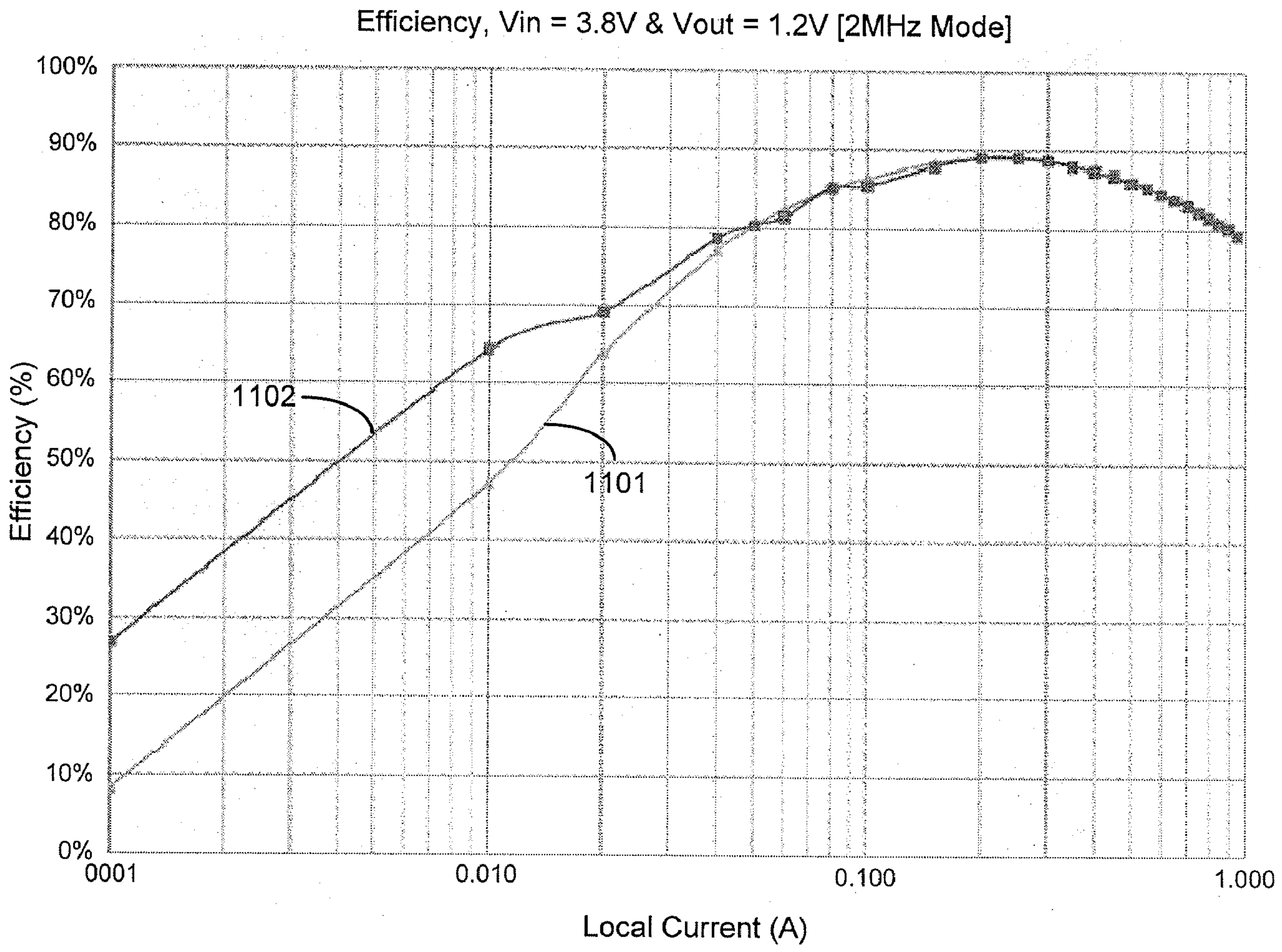
Efficiency, Vin = 3.8V & Vout = 1.2V [2MHz Mode]
Efficiency (%)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0001
0.010
0.100
1.000
Local Current (A)
1102
1101

Figure 11

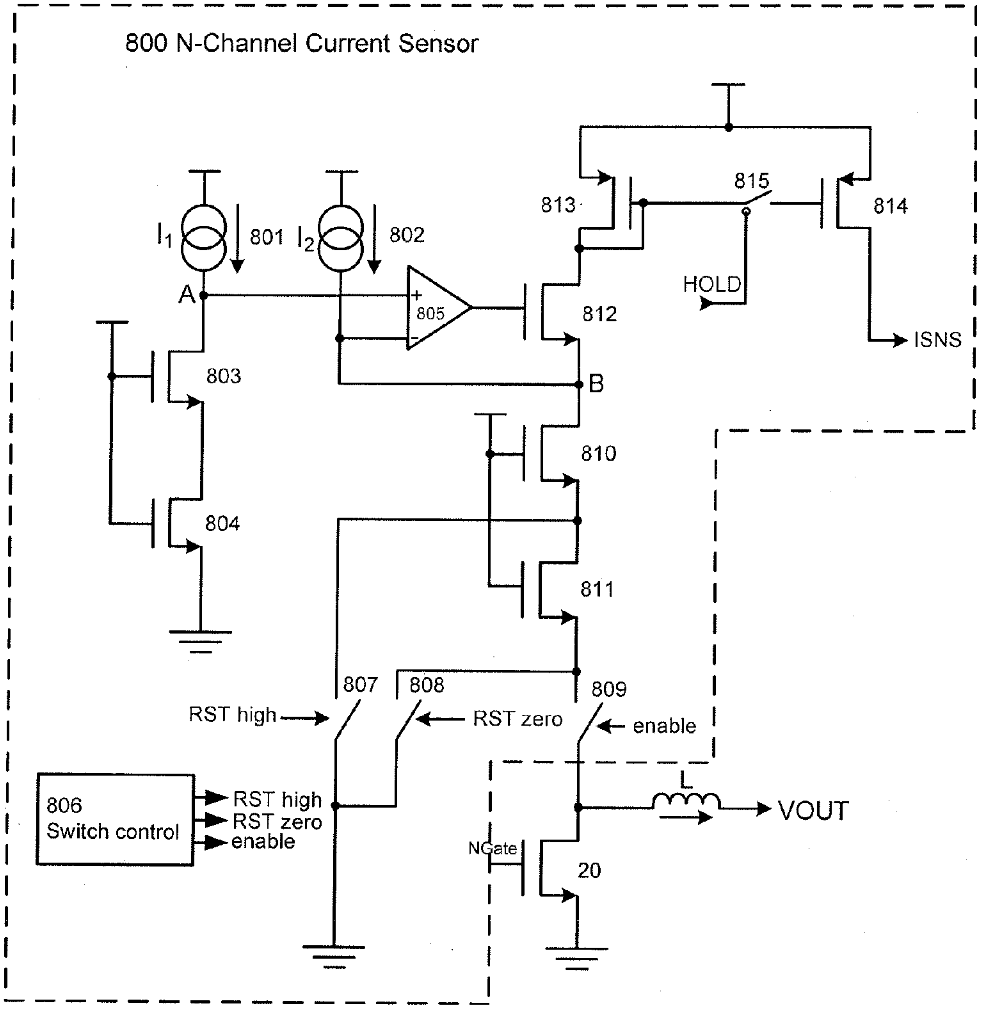
800 N-Channel Current Sensor
I1
801
I2
802
A
805
+
−
803
804
813
815
814
812
HOLD
ISNS
B
810
811
807
808
809
RST high
RST zero
enable
806
Switch control
RST high
RST zero
enable
L
VOUT
NGate
20

Figure 12

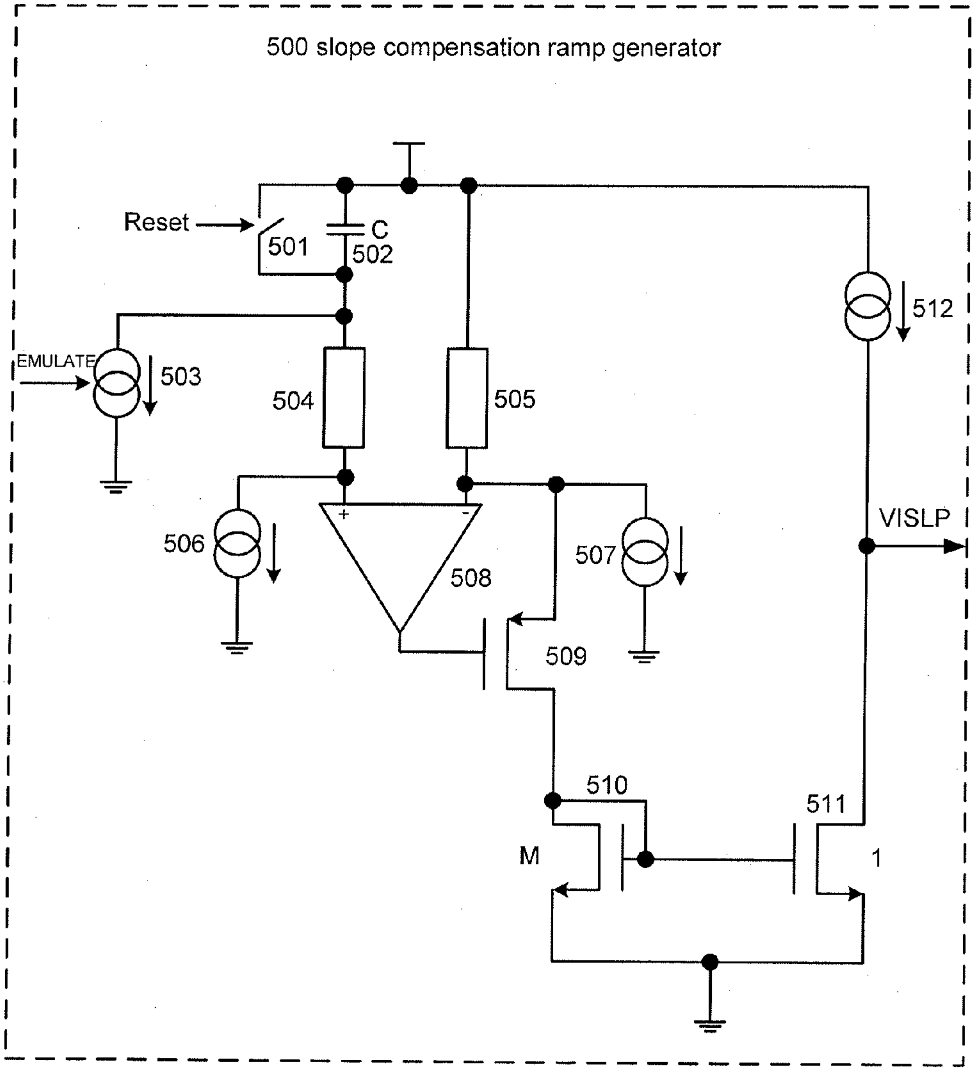
500 slope compensation ramp generator
Reset
C
EMULATE
VISLP
M

Figure 13

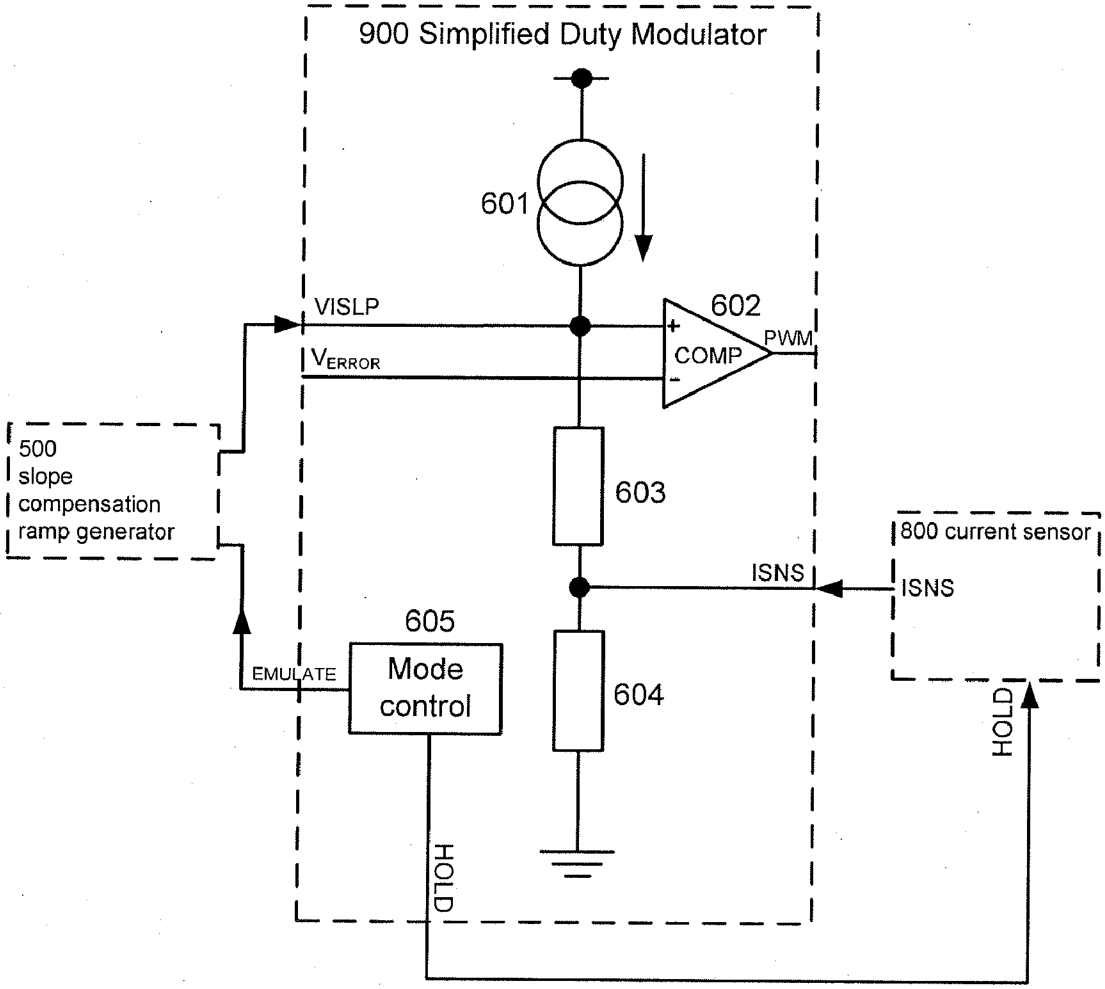
900 Simplified Duty Modulator
601
602
COMP
+
-
PWM
VISLP
V_{ERROR}
603
604
605
Mode control
ISNS
EMULATE
HOLD
500 slope compensation ramp generator
800 current sensor
ISNS
HOLD

Figure 14

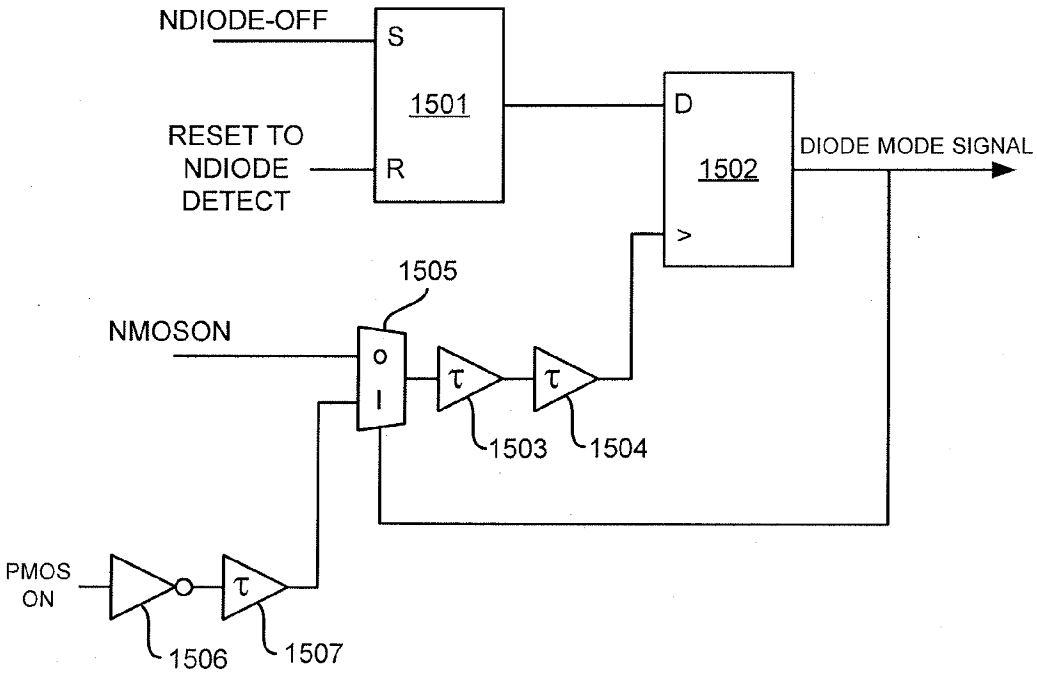
NDIODE-OFF
S
1501
R
RESET TO NDIODE DETECT
D
1502
>
DIODE MODE SIGNAL
1505
NMOSON
0
1
τ
τ
1503
1504
PMOS ON
τ
1506
1507

Figure 15

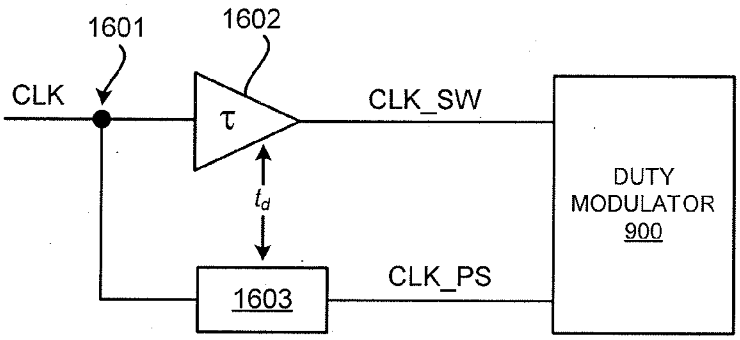
1601
1602
CLK
τ
CLK_SW
t_d
1603
CLK_PS
DUTY MODULATOR 900

Figure 16

DC-DC CONVERTERS

[0001] This application claims the benefit of U.S. Provisional Application No. 61/228,002, filed Jul. 23, 2009, and U.S. Provisional Application No. 61/261,123, filed Nov. 13, 2009.

BACKGROUND OF THE INVENTION

[0002] 1. Field of the Invention

[0003] This invention relates to voltage converters, especially to DC-DC or switched voltage regulators and to methods and apparatus for control of such regulators.

[0004] 2. Description of the Related Art

[0005] In electronic equipment in general, and particularly for fast moving segments such as portable consumer devices such as mobile phones, MP3 players etc, there is a relentless push to use the latest processor technology to increase the device capability and feature set while reducing power and cost. As the next generation of processors start to become available, they need to operate from lower voltages than previous generation processors so as to allow use of the reduction in process feature size, i.e. W/L, that enables a greater level of integration such that it is beneficial in terms of reduced die size, reduced die cost and reduced power consumption. Typically such processors are powered via a DC-DC voltage converter which provides a regulated voltage to the processor.

[0006] These trends result in two design challenges for DC-DC converters to service these applications: one induced by the choice of the value of the low voltage; the other by battery technology lagging behind the change to lower processor supply voltages.

> **[0007]** (i) The reduction in processor supply voltage requires a much tighter control—in absolute terms—of the DC-DC converter output voltage under all prevailing processor load and battery conditions. If the control over the processor supply voltage is not sufficient, problems with under or overvoltage may occur: both equally undesirable.
>
> **[0008]** (ii) Since the battery terminal voltage has not dropped appreciably, and the duty cycle of a DC-DC converter is given by the ratio of V_{OUT}/V_{IN}, duty cycles must therefore reduce. This, coupled with the desire for small external components pushes the DC-DC converter to high operating frequencies, resulting in extremely short switch on i.e. conduction, times. The increase in switching speed afforded by the reduction in transistor feature size is not normally available for the power switches since the interface components must be rated to battery voltage.

[0009] Since small conduction periods, i.e. on times, of the power switches are difficult to control, it is becoming increasingly difficult to control the lower processor output voltages with adequate accuracy using traditional Peak Current Mode control methods. Valley Current Mode (VCM) is an alternative method of control of a DC-DC converter which has been proposed. This method of DC-DC loop control controls the input transistor off, i.e. non-conduction, time, rather than the on time. For the low duty cycle required, the non-conduction time is longer than the conduction time, so is therefore easier to control. Also VCM is known to offer an inherently higher bandwidth and an improved transient response.

SUMMARY OF THE INVENTION

[0010] It is therefore an aim of the present invention to provide improved control of DC-DC converters.

[0011] Thus according to the present invention there is provided a DC-DC converter comprising: an inductor operably connected between a first node and an output node; a high side switch operably connected between a high side supply input node and the first node; a low side switch operably connected between a low side supply input node and the first node; switch control circuitry for controlling said switches, wherein the switch control circuitry comprises first monitoring circuitry for determining whether a first time period between the high side switch turning off and the inductor current reaching zero is less than a threshold time period, wherein if the first time period in one duty cycle is less than the threshold time period the switch control circuitry inhibits the low side switch from turning on in a subsequent cycle.

[0012] The DC-DC converter of this aspect of the invention thus has an inductor connected to a first node connected between high and low side supply switches. The high and low side supply switches are any suitable switches, such as power switches, for example PMOS and NMOS switches respectively, for connecting, in use, to high and low side supplies, e.g. a voltage input V_{IN} and ground.

[0013] As the skilled person will appreciate in operation the first node, i.e. the input node for the inductor, is typically switched between the high side supply and the low side supply in a duty cycle determined by the input and output voltages. The inductor current rises during the period that the high side switch is turned on, the low side switch being turned off during this period, and falls during the period where the low side switch is turned on, the high side switch being turned off during this period.

[0014] In current mode control a signal representing the inductor current is compared with a threshold derived from the voltage output of the converter to control the duty cycle. As the skilled person will appreciate one or more additional signals, for instance a slope compensation ramp signal, may be added to one or both of the current or voltage signals prior to comparison.

[0015] The converter may be controlled in a peak current mode, where the current is measured during the period when the high side switch is on and the comparison determines when the low side switch turns on, i.e. the switching occurs at the peak current. Alternatively the converter may be controlled in a valley current mode, where the current is measured during the period when the low side switch is on (and hence is decreasing) and the comparison determines when the high side switch turns on, i.e. the switching occurs at the valley current.

[0016] In either mode of control, when the low side switch is turned on, the current in the inductor drops steadily until the low side switch turns off—which could be due to the end of the cycle in Peak Mode control or the appropriate Valley current being reached in valley mode control. In some operating conditions however, at relatively low load current demand, the inductor current would drop to zero and then go negative, i.e. the direction of current flow would reverse, before the high side switch was turned on. This reversal of current in the inductor means that current would be flowing away from the load (i.e. away from an output capacitor) and to the low side supply. This is inefficient as charge previously supplied during via the high side supply is effectively wasted.

[0017] In some converters therefore there may be circuitry to prevent the low side switch from passing a significant negative current. This circuitry may involve monitoring the current flow through the low side switch and/or inductor in order to turn the switch off before a significant reverse current can flow. Typically however there will be propagation delays associated with such monitoring circuitry as will be described in detail later and hence there will be a minimum conduction period for the low side switch. In other words when the low side switch is turned on it will be on for at least this minimum conduction period before it is turned off.

[0018] At low current demands the inductor current may be low enough such that the low side switch being turned even on for the minimum conduction period would result in a reverse current flow.

[0019] According to this aspect of the present invention therefore the first monitoring circuitry determines whether a first time period between the high side switch turning off and the inductor current reaching zero is less than a threshold time period. This threshold time period is set to be greater than the minimum conduction period mentioned above. If the inductor current does decay to zero in a period less than this time threshold period the low side switch is inhibited from turning on in a subsequent cycle.

[0020] Operation of the high side switch is not affected and thus the high side switch may be turned on as appropriate in accordance with the mode of control. However the low side switch would remain off when the high side switch is subsequently turned off.

[0021] As will be described in detail later, if one of the high or low side switches is turned off, and the other switch is not turned on, at a point when there is current flow in the inductor then current may continue to flow through a parallel path. For example the current may flow through an NMOS body diode associated with a NMOS power switch. As will be explained later in more detail an NMOS power switch will usually have a parasitic diode associated with it. When the NMOS power switch is turned on the parasitic diode will have virtually no effect. However if the NMOS switch is turned off and the high side switch is also turned off, any current flowing in the inductor may flow through the NMOS body diode. As is usual with a diode the current flow is substantially in one direction only and hence this current will not reverse. Thus by allowing the current to flow through the NMOS body diode the inductor current can decay to zero but it will not go negative, i.e. the direction of current flow will not reverse. This mode of operation therefore prevents the inductor current from going substantially negative.

[0022] Instead of relying on an inherent body diode associated with the low side switch, in some embodiments a separate diode (or other unidirectional current device) may be specifically provided in parallel with the low side switch to provide a parallel path and control current flow when the low side switch is off. The separate diode should be arranged to allow forward current to flow if the low side switch is turned off (before the high side switch is turned on) but to prevent current reversal in the inductor. Any suitable diode, such as a low drop diode or Schottky diode could be used.

[0023] As mentioned, if, in one cycle, the first time period between the high side switch turning off and the inductor current reaching zero is less than a threshold time period, then the low side switch is inhibited from turning on in a subsequent cycle. The low side switch may be prevented from turning on in the very next cycle but, if the converter has switched to a pulse skip mode of operation the low side switch would not be turned on anyway. As will be described in more detail later in some instances the converter may operate in a mode, termed a pulse skip mode, where the high side switch is not turned on during a particular cycle. In pulse skip mode the low side switch is also kept turned off. There may therefore be one or more cycles in which in the high side switch, and hence the low side switch, are not turned on.

[0024] Preferably, if the time period between the high side switch turning off and the inductor current reaching zero is less than the time threshold period but the converter then enters pulse skip mode (thus skipping the turn on of the high side switch and hence the low side switch in one or more cycles), the inhibition of the low side switch is maintained until the converter exits pulse skip mode. In other words, even if there have been one or more cycles in which the high side switch has not turned on, the next time that the high side switch is turned on the low side switch is subsequently inhibited from turning on. The subsequent cycle (in which turn on of the low side switch is inhibited) is therefore preferably the next one in which the low side switch would otherwise turn on following the high side switch turning off.

[0025] It will be appreciated that the term cycle refers to a switching cycle wherein, in non pulse skip mode of operation and where turn on of the switches is not inhibited, each of the high side and low side switches will be on for a period of time. In embodiments of the present invention the cycle is defined by a clock signal. An appropriate edge of a clock pulse, i.e. the leading or trailing edge is used to start a cycle and the same edge of a subsequent pulse is used to define the start of the next cycle (and hence the end of the present cycle). Such DC-DC converters allow constant frequency operation as fixed by the clock signal.

[0026] DC-DC converters allowing constant frequency operation offer several advantages as compared to other modes of operation, such as constant on-time converters where the frequency of operation is variable. Operating at a constant frequency means that the frequency of operation is known in advance and the designer of the device that uses the DC-DC converter can take this into account. Thus interference due to the switching frequency of the DC-DC converter can be avoided and line filters or notch filters and the like can be designed and used in the device with a knowledge of the switching frequencies to reduce interference. Further in a device with several constant frequency DC-DC converters the clock signals for the converters may be phased to reduce the total combined input current ripple, i.e. to arrange the high side on time of one converter to occur during, or at least overlap with, the high side off time of another converter. This can reduce the stress on the supply input capacitor and the peak current requirements of the common supply.

[0027] In a preferred embodiment of the present invention therefore the converter is one where the switching cycle is defined by clock edges of a clock signal. The skilled person will of course appreciate that the converter may be designed to operate at more than one different switching frequency and an appropriate frequency may be selected in use by selecting an appropriate clock signal. For example a converter may be operable at a first fixed frequency based on a first clock signal or a second fixed frequency based on a second clock signal. This allows the converter to balance efficiency against transient performance. As the skilled person will be aware higher switching frequencies mean more switching losses but generally better transient performance. At lower current loads, transient performance may be less of an issue and the frequency may be switched to a lower frequency clock signal to improve efficiency. In any case however the possible frequencies of operation are known in advance and fixed by the first and second clock signals and are not continually variable according to small changes in load.

[0028] The skilled person will also appreciate that it is common to apply small amounts of dither to a clock signal to vary the switching frequency slightly. This is a known change however which is achieved (generally) though prearranged variation of the clock signal. Again the device designed will be aware of the degree of dither that will be applied and the switching frequency will be maintained within a small set range defined by the amount of dither. As used in this specification the term constant frequency applied to switching frequencies shall be taken to mean that the switching cycle is controlled based on clock edges of a clock signal, including clock signals that are dithered.

[0029] In a peak current mode converter the turn on of the high side switch is controlled based on the relevant clock edge. Thus at the start of the cycle the high side switch is turned on unless the converter is in pulse skip mode in which case both switches remain off for the entire cycle. Once turned on, the high side switch is subsequently turned off based on a comparison of the current signal with the voltage error (plus any slope compensation). At this point the low side switch is turned on, unless inhibited from doing so, and in any case the time until the inductor current reaches zero is compared to the time threshold period.

[0030] In a valley current mode converter the clock edge controls turn on of the low side switch. Assuming that the high side switch was turned on in the previous cycle (i.e. the converter was not in pulse skip mode) the start of the cycle sees the low side switch turn on, unless the low side switch is inhibited from doing so. In either case the time between the high side switch turning off (i.e. end of the previous cycle) and the inductor current reaching zero is compared to the time threshold period. Turn on of the high side switch is based on a comparison of the current signal with the voltage error (plus any slope compensation).

[0031] As briefly mentioned above and as will be explained in detail later, when a low side switch, such as a NMOS switch, is turned off, when current is still flowing in the inductor (as there would be when the high side switch has just been turned off), current will still flow through a parallel path. For instance the current may flow through a parasitic body diode associated with the NMOS switch. The inductor current will still therefore flow and will decrease (faster than it would have done had the switch been on) until the inductor current reaches zero. The parasitic diode will not conduct in the reverse direction however and thus there will be no reverse current and hence no efficiency loss. In some embodiments a separate diode may be provided to provide the parallel path and to provide forward conduction only when the low side switch is off and the high side switch has not turned on.

[0032] If a DC-DC converter according to this aspect of the invention is operating normally and the current load drops (for constant output voltage) the inductor current will reach zero earlier and earlier in each cycle. Were the demand to continue to fall at some point the inductor current will reach zero in a time period less than the time threshold period. The duration of time threshold period is chosen in view of an expected maximum cycle-to-cycle change for the particular converter such that the inductor current in the cycle in which the time threshold period is reached will not have undergone any significant reversal. As mentioned above there is a practical minimum conduction period for the low side switch. If the ideal on-time for the low side switch drops below the minimum conduction period it is not possible to turn the low side switch off in time and current reversal may occur. The time threshold period is chosen to be sufficiently larger than the minimum conduction period such that the converter is unlikely to have a low side switch on-time in one cycle which is greater than the time threshold period but which, in the very next cycle, has an ideal low side switch on-time which is shorter than the minimum conduction period.

[0033] When the inductor current reaches zero (from the high side switch turning off) in a time period shorter than the threshold time period this is indicative that, were the current drop to continue, current reversal would be likely. The low side switch is therefore inhibited from turning on in a subsequent cycle to avoid the potential for current reversal. This embodiment of the present invention therefore effectively pre-emptively prevents a situation where current reversal could occur.

[0034] The cycle in which the low side switch is inhibited from turning on may be the very next cycle but, as mentioned above, in some converters there may be a pulse skip mode where the high side switch is not turned on in every cycle, depending on the loading. Clearly if the high side switch is not turned on the low side switch need not be turned on either. However current reversal is still a possibility were the low side switch to turn on immediately after the next high side switch conduction period. Therefore the switch control circuitry is arranged to inhibit the low side switch from turning on following the next time that the high side switch has been on, i.e. in the next non pulse skip cycle.

[0035] The converter operates in this mode, inhibiting turn on of the low side switch in a subsequent cycle for as long as the inductor current reaches zero before the time threshold. Only when the operating conditions change, and the inductor current reaches zero after the time threshold period does normal operation resume, with the low side switch turning on when the high side switch turns off.

[0036] As mentioned it can be beneficial for efficiency reasons to prevent the inductor current from going negative, i.e. flowing away from the load during the duty cycle. The switch control circuitry may therefore be operable, in a cycle where turn on of the low side switch is not inhibited, to turn the low side switch off before the high side switch is turned on. This means that there may be a period in the duty cycle where both of the high and low side switches are off. This may be referred to as a discontinuous switching mode as there is a discontinuity in the operation of the switches. The switch control circuitry may comprise second monitoring circuitry for determining the current flow in the inductor from the electrical properties of the low side switch when the low side switch is on and monitoring current flow in the inductor against a current threshold and wherein the switch control circuitry is arranged to turn the low side power switch off when the current threshold is reached.

[0037] The current flowing through the low side switch, and hence through the inductor, can be determined by the electrical properties of the low side switch when it is turned on. For a transistor switch, such as an NMOS switch, the source-drain voltage of the transistor is proportional to the current flow through the channel. By using the electrical properties of the low side switch to derive the inductor current no additional sense resistor is required. As will be appreciated the use of sense resistors purely to allow sensing of the inductor current would introduce associated resistive losses and hence inefficiency into the operation of the converter.

[0038] By setting the low side switch to turn off when the inductor current reaches a threshold current level at least some of the period of negative, or reverse, inductor current flow may be prevented. The threshold may be set at zero current but propagation delays and the like mean that there would be a delay between the inductor current reaching the threshold and the low side switch actually being turned off. During this time the inductor current may have decreased further, i.e. a reverse current flow may have started. The threshold is therefore preferably set to a positive level such that the low side switch is turned off before the current flow in the inductor reverses direction. Preferably the threshold level is set as low as possible allowing for voltage offsets and propagation delays to ensure that the inductor current does not reverse direction. However a small current reversal may be tolerated in some applications. The threshold level may be fixed or may be configurable and may be changed in operation depending on the operating conditions of the converter, such as the target output voltage.

[0039] The first monitoring circuitry, i.e. the circuitry for monitoring whether the inductor current reaches zero before the time threshold period, may comprise a voltage threshold monitoring circuit for comparing the voltage of the first node with a threshold voltage when the low side switch is turned off. When low side switch is turned off and an inductor current is flowing, as mentioned above there will be conduction through a parallel path such as a unidirectional switch associated with the low side switch. The first node will therefore be at a diode voltage below the low side supply, e.g. ground at this point. Once the inductor current extinguishes the node will fly high. Thus detecting the voltage on this node crossing a voltage threshold can be used to determine when the inductor current has reached zero. The voltage threshold may be set to zero volts.

[0040] Various arrangements for the first monitoring circuitry exist. The first monitoring circuitry may for instance comprise a first timer circuit configured, in response to the high side switch turning off, to generate a first output signal for a period equal to the time threshold period and thereafter to generate a second output signal. For example the timer circuit may have a logic high or logic low output. If the output is logic high, for example, the time threshold period has not expired. Thus the monitoring circuit may comprise a simple logic arrangement. If the inductor current reaches zero, e.g. the output of voltage threshold monitoring circuit goes high, when the output of the timer circuit is also high, an inhibit low switch signal is generated. This could easily be implemented by an AND gate arrangement although many other options exist as will be appreciated by the skilled person. In an alternative arrangement the first monitoring circuitry may comprise a first timer circuit configured to determine whether the inductor current has reached zero at a time equal to the time threshold period after the high side switch turns off, i.e. in order words the inductor current is sampled at a time equal to the first time period after the high side switch turns off.

[0041] The first timer circuit may be arranged to receive a control signal synchronous with a high side switch control signal (e.g. a copy of the high side switch control signal or feedback from the switch driver) and delay said control signal by at least one delay equal to the threshold time period using a delay element. The period between receipt of the control signal and receipt of the delayed control signal can be used to determine the first time threshold period. Other timer arrangements are possible however. A ramp signal could be generated which reaches a threshold at the end of the time threshold time period. A counter could count the cycles of a sufficiently fast clock signal.

[0042] The first timer circuit may alternatively be arranged to receive a low side control signal synchronous with a low side control signal (e.g. a copy of the low side switch control signal or feedback from the switch driver) and delay said low side control signal by at least one delay such that the time period between the high side switch turning off and the delayed low side control signal is substantially equal to the time threshold period. It will be appreciated that the low side switch will turn on (if not inhibited from doing so) a short time after the high side switch turns off. Hence the delay applied to a low side control signal may be slightly shorter than the threshold time period as there will be some inherent delay in low side switch turning on after the low side switch turns off.

[0043] It will be clear however that the time period between the time at which the low side switch turns on (or would otherwise turn on if not inhibited from doing so) and the inductor current reaching zero may be monitored as a way of monitoring the time between the high side switch turning off and the inductor current reaching zero. There is a generally fixed delay between the time at which the high side switch turns off and the time at which the low side switch turns on (or would otherwise turn on if not inhibited from doing so).

[0044] The time threshold period may be fixed or may be configurable. The period may be configurable, i.e. different time threshold periods may be chosen, to cope with different operating conditions.

[0045] The low side switch may be an NMOS transistor and the high side switch may be a PMOS transistor.

[0046] The DC-DC converter, in some embodiments, may be operable in valley mode control. The converter may comprise circuitry for keeping a current feedback loop active when the low side switch is off. The circuitry for keeping a current feedback loop active when the low side switch is off comprises emulation circuitry for emulating the change in inductor current when the low side switch is off.

[0047] As will be appreciated valley mode control uses the current in the inductor during the period when the low side switch is off to control switching. However this aspect of the present invention inhibits turn on of the low side switch in certain circumstances. In valley mode control, when the low side switch is turned off, current mode control of the converter may be difficult to achieve. This embodiment of the present invention therefore comprises emulation circuitry for emulating the change in inductor current during a period when both the high side and low side switches are turned off. The emulated change in inductor current can be used by the switch control circuitry to control when to turn on the high side switch.

[0048] The switch control circuitry may comprise current sense circuitry for determining the current flow in the inductor from the electrical properties of the low side switch when the low side switch is turned on and wherein the output of said current sense circuit comprises the first signal when the low side switch is turned on. This provides a lossless current mode control. However when the low side switch is off the electrical properties of the switch can no longer be used to derive the inductor current. The emulation circuit however emulates the change of inductor current when the low side switch is turned off and hence the emulated change in current can be used when the low side switch is off. This embodiment of the present invention therefore allows lossless current sensing but maintains control of the current loop when the low side switch is off.

[0049] In some embodiments the DC-DC converter may be operable in peak mode control.

[0050] The switch control circuitry may be operable in a pulse skip mode to prevent turning the high side switch on in a cycle. As mentioned above, at low operating current it may be advantageous to skip pulses, i.e. not turn the high side switch on in certain cycles.

[0051] Where the converter is operable in valley mode control the switch control circuitry may prevent turning the high side switch on in a cycle if a second time period between turning the high side switch on and turning the high side switch off would be less than a second threshold time period. This embodiment of the present invention implements a pulse skip mode based on a minimum high side conduction time.

[0052] The switch control circuitry may comprise a second timer circuit for producing an inhibit signal which inhibits turn on of the high side switch, for a time period substantially equal to the second time threshold period, before a clock edge which would cause the high side switch to turn off. The second timer circuit may comprise an input for receiving an input clock signal, a delay for producing a delayed clock signal and circuitry for generating the inhibit signal based on the input clock signal.

[0053] In one embodiment of DC-DC converter the cycle frequency of the converter is constant. However the embodiments of the present invention are applicable to variable frequency converters and/or converters that apply dither to the clock edges.

[0054] The DC-DC converter according to these embodiments of the invention therefore provide a converter that can be controlled to low load current demand and which offers high efficiency. Consider a buck converter, with an inductor switched between a supply VIN and Ground by a PMOS and an NMOS transistor respectively. Under high or medium current demand, the inductor current, composed of an average component and a ripple, remains above zero through every cycle. This is termed Continuous Conduction Mode (CCM). As the load current demand decreases, it is advantageous for efficiency reasons to alter the control so that the current in the inductor goes to zero for some of each cycle to avoid ripple being large enough to cause a reverse in the current in the inductor. This is termed Discontinuous Current Mode (DCM). At even lower load currents it is advantageous to “pulse-skip” so the PMOS switch does not recharge the inductor in some cycles, but this can lead to undesired behaviour.

[0055] These various modes present different control problems and dynamics, and it is important to be able to transition between modes seamlessly without transients appearing during change over. This aspect of the present invention provides a control mechanism that achieves these efficiency benefits and maintains control over the current loop of the converter.

[0056] In another aspect of the present invention there is provided a method of controlling a DC-DC converter comprising an inductor operably connected between a first node and an output node; a high side switch operably connected between a high side supply input node and the first node; and a low side switch operably connected between a low side supply input node and the first node; the method comprising: monitoring whether a first time period between the high side switch turning off and the inductor current reaching zero is less than a threshold time period, and where the first time period in one duty cycle is less than the threshold time period the switch control circuitry inhibiting the low side switch from turning on in a subsequent cycle.

[0057] This method of the present invention offers all of the same advantages and can be used in all of the same embodiments as described with reference to the first aspect of the invention.

[0058] In another aspect of the invention there is provided a DC-DC converter comprising: a PMOS switch operably connected between a voltage input node and the first node; an NMOS switch operably connected between a ground input node and the first node; control circuitry operable to control said switches, wherein the switch control circuitry detects whether a first time period between the PMOS switch turning off and the voltage at the first node reaching zero is less than a time threshold, wherein if the first time period is less than the threshold time period the switch control circuitry prevents the NMOS switch from turning on in the next cycle.

[0059] In a further aspect of the invention there is provided a DC-DC converter comprising: an inductor operably connected between a first node and an output node; a high side switch operably connected between a high side supply input node and the first node; a low side switch operably connected between a low side supply input node and the first node; switch control circuitry arranged to monitor the time between the high side switch turning off and the voltage at the first node reaching zero and, if said monitored time is less than a first time threshold, preventing the low side switch from subsequently turning on the next time the high side switch is turned on.

[0060] The invention, in another aspect, provides a DC-DC converter comprising: an inductor operably connected between a first node and an output node; a high side switch operably connected between a high side supply input node and the first node; a low side switch operably connected between a low side supply input node and the first node; and switch control circuitry operable in a first mode where the low side switch is not turned on following the high side switch being turned off.

[0061] A further aspect of the invention provides a DC-DC converter operable in a mode wherein a low side switch is not turned on following activation and deactivation of a high side supply switch.

[0062] In general the problem of NMOS current reversal in light loading is mitigated by detecting if the output voltage crosses zero before a time threshold that is longer than the minimum NMOS on time set by the propagation delay from the zero-current comparator, and if so, disabling the NMOS in the next cycle. The advantage is that valley mode control is maintained down to very light loads and the efficiency is only sacrificed when direct NMOS control is no longer possible. The diode mode of operation could also occur in non-pulse-skipping mode, say in DCM with a very short PMOS on-time and a high V_{OUT} relative to V_{IN}-V_{OUT} giving a fast NMOS current slew rate, giving power dissipation savings.

[0063] Conveniently the output voltage of the NMOS switch is used to determine the zero crossing time. In general however this aspect of the invention relates to means of preventing current reversal and especially to circuitry for determining a forward current duration for the NMOS and for preventing the NMOS from turning on in the subsequent cycle if the forward current duration is less than a first duration. Such means may include a sense resistor for sensing the inductor current and/or a current transformer.

[0064] The present invention provides a practical Valley Current Mode converter that can be used, with current mode control, for a variety of applications. Embodiments of the present invention realise an integrated circuit (IC) arrangement and in particular a power management integrated circuit (PMIC) arrangement comprising a Valley Current Mode converter. An embodiment of the Valley Mode converter comprises a fixed frequency valley mode converter, i.e. a converter having a switching frequency which is based on a predetermined clock signal. The Valley Current Mode converter may be operable in Continuous Current Mode (CCM), Discontinuous Current Mode (DCM) and/or a transitional phase between DCM and CCM. The valley mode converter may comprise a valley mode converter operable in a pulse skip mode. The valley current mode converter may be controlled based on the same control loop in all of these modes of operation, i.e. the loop that provides control in CCM also provides control in DCM.

[0065] Embodiments of the invention provide current mode converters, especially valley current mode converters, with lossless current sensing. Lossless sensing techniques avoid the use of separate sense resistors with the associated losses. Preferably the converter is fully integrated in that the high and low side switches are integrated within an IC/PMIC in addition to the converter and the internal (preferably lossless) current sense circuits. A fully integrated lossless valley current mode converter represents a novel embodiment of switched converter. Preferably the lossless and/or fully integrated converter is operable at a fixed switching frequency. A converter comprising lossless sensing and enabling CCM and DCM operation is thus provided by the present invention and a novel IC/PMIC may comprise such a converter.

[0066] The ability to operate at a fixed frequency is a particular advantage of the embodiments of the present invention and in general one aspect of the invention relates to a fixed frequency converter which is operable in non-CCM modes of operation, whether full DCM and/or a transitional phase between CCM and full DCM and/or a pulse skip mode of operation.

[0067] The various aspects of the present invention are described principally with reference to buck converters but other embodiments of the invention provide boost converters or buck-boost converters.

[0068] Although described above in relation to DC-DC converters the embodiments of the invention can be applied generally to switching regulators. All aspects of the invention may be implemented in a power management apparatus such as a power management integrated circuit. The embodiments of the present invention are applicable to use in electronic devices and in particular to portable devices.

BRIEF DESCRIPTION OF THE DRAWINGS

[0069] The invention will now be described by way of example only with reference to the following drawings, of which:

[0070] FIG. **1** shows a typical arrangement of a DC-DC converter arranged to power processor circuitry;

[0071] FIG. **2** illustrates a conventional DC-DC converter;

[0072] FIG. **3** illustrates the parasitic diodes associated with the switches of the output stage;

[0073] FIG. **4** shows inductor current and voltage waveforms at a certain level of current demand;

[0074] FIG. **5** shows a DC-DC converter according to an embodiment of the invention;

[0075] FIG. **6** shows inductor current and voltage waveforms in a DC-DC converter according to an embodiment of the invention at a level of current demand lower than that shown in FIG. **4**;

[0076] FIG. **7** shows inductor current and voltage waveforms in a DC-DC converter according to an embodiment of the invention at a level of current demand lower than that shown in FIG. **6**;

[0077] FIG. **8** shows inductor current and voltage waveforms between cycles;

[0078] FIG. **9** shows current and voltage waveforms illustrating a mode of operation at low current demand;

[0079] FIG. **10***a* illustrates various control waveforms of an embodiment of a DC-DC Converter operating in DCM;

[0080] FIG. **10***b* illustrates various control waveforms of an embodiment of a DC-DC Converter operating in a transitional mode between CCM and DCM;

[0081] FIG. **11** illustrates the relative efficiency of operating a DC-DC in modes according to embodiments of the invention as compared to Forced CCM;

[0082] FIG. **12** illustrates one embodiment of a current sense circuit;

[0083] FIG. **13** illustrates one embodiment of a ramp generator capable of emulating a change in inductor current;

[0084] FIG. **14** illustrates a simplified embodiment of a duty modulator;

[0085] FIG. **15** illustrates one embodiment of a circuit arrangement for implementing diode mode of operation; and

[0086] FIG. **16** illustrates one embodiment of a circuit arrangement for implementing pulse skip operation.

DETAILED DESCRIPTION

[0087] FIG. **1** shows a typical application where processor circuitry **101**, which may, for example, be a processor of a portable electronic device **107**, is supplied with a voltage V_{OUT} **102** by a DC-DC converter **100**. The DC-DC converter **100** receives an input voltage V_{IN} **103** and an external clock signal CLK **104** and outputs the required voltage output V_{OUT} **102**. It is usual for the supply voltage of a processor to be reduced when the processor is idling in order to save power, and then to ramp up to a more normal operating voltage where the processor may achieve full operating speed. The processor circuitry **101** therefore provides voltage select signals V_{SEL} **105** to the DC-DC converter **100** to select an appropriate voltage output V_{OUT}. The voltage select signals may be digital signals for controlling a programmable element of the DC-DC converter, such as a level shifter, as will be described later. The DC-DC converter **100** may also be operable in various modes, as will be described later, and the processor circuitry may select a particular mode of operation by appropriate mode control signals MODE **106**. It will be appreciated that DC-DC converters may be used to provide power to device sub-systems other than processors and the embodiments described herein are generally applicable to many DC-DC converters or switched voltage regulators used for many applications.

[0088] A conventional current-mode buck (i.e. step down) DC-DC converter **200** is shown in simplified form in FIG. **2**. The converter **200** comprises two nested feedback loops, an inner Current Control loop and an outer Voltage Control loop.

[0089] The Current Control loop block **201** takes an input signal V_{ERROR} and a current sense signal ISNS fed back from the output stage and generates pulse-width modulated drive signals for the output stage **202**. The voltage on the output stage output node LX is switched between ground and supply, V_{IN}, at a controlled duty cycle, resulting in a triangular current waveform in the inductor L. The inductor L and output capacitor C1 act as a filter to ensure an average voltage V_{OUT} at an output node **203**.

[0090] In operation the inductor current is sensed, and compared with V_{ERROR}. So this feedback loop generates an output sensed current varying according to the input signal V_{ERROR}. In many conventional DC-DC converters the sensed current is a peak current, although it is known to use an average current in some converters. In embodiments of the present invention to be described the minimum or "valley" current is used to control the duty cycle of the converter.

[0091] Variation of the delivered output current, smoothed by the output filter L, C**1**, modulates the output voltage at V_{OUT}. This voltage V_{OUT} is fed back, translated down to an appropriate voltage V_{OUT_LS} by a Level Shifter, or Voltage Shifter, block **204**, to the input of a Voltage Error Amplifier block **205**. The Voltage Error Amplifier block **205** compares this processed version of V_{OUT} with a supplied reference voltage V_{REF} and provides the error signal V_{ERROR} which drives the above described inner feedback loop to close the outer feedback loop and thus stabilise V_{OUT} at the desired voltage.

[0092] The Level Shifter **204** is illustrated as a resistive potential divider, but could be other voltage scaling or voltage-shifting circuitry, or even be a direct connection if the required V_{OUT} is equal to V_{REF}. The level shifter applies a conversion to the level of V_{OUT} such that, when V_{OUT} is equal to the desired or target output voltage, the level shifted signal V_{OUT_LS} has a known relationship to the reference voltage V_{REF} (e.g. the level shifted signal V_{OUT_LS} may be equal to V_{REF} when V_{OUT} is exactly the desired output voltage). The Level Shifter may be programmable, mechanically or digitally, to provide different voltage scaling or shifting circuitry, so as to allow the converter to be configurable to output different values of V_{OUT}. For instance, it may be programmable by a digital multi-bit signal, such as the VSEL signal generated by a processor as shown in FIG. **1**.

[0093] The Voltage Error Amplifier Block **205** is illustrated as including an Operational Transconductance Amplifier (OTA) **206** driving an RC network **207**, but could include some other amplifier. The RC network **207**, or other passive impedances, provide closed-loop stabilisation.

[0094] The Current Loop Control block **201** receives a signal **208** from the output stage **202** which passes through a Current Sensor Amp block **209** to pre-condition it, for instance to scale or strobe the signal, representative of the inductor current, to generate a convenient current sense signal ISNS. A duty modulator **210** compares the ISNS signal to the input V_{ERROR} to derive drive signals of the appropriate duty cycle to drive the output stage devices **10**, **20** on and off via a Switch Driver buffer stage **211**. The Duty Modulator may require a clock signal **212** and a Ramp Generator **213** to generate the necessary sequence of pulses as would be understood by one skilled in the art. Ramp Generator **213** may generate a slope compensation ramp signal which may be added, either in whole or in part, to the ISNS signal and/or the V_{ERROR} signal so as to prevent sub-harmonic oscillations as would be well understood by one skilled in the art.

[0095] The Output Stage **202** in general will have a high-side driver device such as a PMOS transistor **10** to switch the output to the high-side supply rail **214** (V_{IN}) and a low-side driver device such as an NMOS transistor **20** to switch the output to the low-side supply rail **215** (Ground). The Output Stage **202** is also required to supply information, i.e. an indication, of inductor current signal **208**, to feed back to the Current Control block **201**.

[0096] The inductor current could be sensed using a series resistor in series with the inductor or the respective transistor. However the use of such series resistors introduces an extra source of resistive power loss and thus reduces the efficiency of the converter. Efficiency is, especially for battery powered devices, an important consideration. It is preferable therefore to use "lossless" sensing techniques, for example to sense the drain-source voltage across the PMOS due to its on resistance. This gives a voltage proportional to the PMOS current.

[0097] In operation of a conventional peak mode DC-DC converter the PMOS switch is turned on at an edge of the clock signal **212**. In the lossless current sensing approach, i.e. in embodiments not having a sense resistor, the drain-source voltage of the PMOS is monitored to derive a current signal proportional to the current through the PMOS **10** and hence the inductor L. This current signal, appropriately scaled, is compared to the threshold V_{ERROR} signal. When the current reaches the threshold the PMOS is turned off by switch driver **211** and the NMOS is turned on, i.e. the switching occurs at the peak of the inductor current. Practically a slope compensation ramp signal is applied to either or both the current signal or V_{ERROR} signal prior to the comparison to prevent sub-harmonic oscillation. The effect is therefore that the current signal is compared with a threshold that ramps across each cycle.

[0098] The bandwidth of the current control loop is high, generating the appropriate pulse width on a cycle-by-cycle basis, whereas the bandwidth of the outer, voltage control, loop is relatively low, so the current loop may be analysed assuming a substantially static input voltage V_{ERROR}.

[0099] The terminal LX of the inductor L is switched to V_{IN} for a fraction D of each clock cycle and to ground for the remaining fraction 1-D of each cycle. The average voltage at node LX is thus D^*V_{IN}. The output capacitor C**1** is large enough for the voltage V_{OUT} to be substantially constant within each cycle, so the average voltage at V_{OUT} is also D^*V_{IN}. So for small V_{OUT} relative to V_{IN}, the duty cycle $D=V_{OUT}/V_{IN}$ may be small.

[0100] Also there is a push towards faster clocking frequencies to allow the use of lower value, physically smaller, inductors, which would further reduce the PMOS on-time, and the time available to sense its current.

[0101] As mentioned previously control of very short switching times is difficult. Some embodiments of the present invention therefore use Valley Current Mode (VCM) control. In VCM the NMOS switch **20** is turned on at a clock edge and the current in the inductor during the NMOS conduction period is monitored. When the inductor current drops to the V_{ERROR} threshold (as modified by the slope compensation ramp) the NMOS is turned off and the PMOS is turned on, i.e. the switching is controlled by the lowest inductor current or the valley current. For short duty cycles the NMOS switch may be turned on for significantly longer than the PMOS switch and so Valley Current Mode control can ease some aspects of control of a DC-DC converter at low voltages.

[0102] FIG. **4** illustrates the inductor current and the voltage at the output stage output terminal, LX, during one cycle of a set of repetitive cycles. The solid curves **1001** show waveforms for a higher average current, the dashed curves **1002** show waveforms for a lower current (assuming constant V_{IN} and V_{OUT}). The upper part of FIG. **4** shows the current curves at node LX and the lower part of the figure illustrates the voltages at node LX. During the first part of the illustrated cycle, i.e. between t_0 and t_1, the NMOS is conducting and LX is near ground while the other end of the inductor is kept at V_{OUT} by the output capacitor. The current therefore decreases at a slope $dI_L/dt=V_{OUT}/L$. During the second part of the cycle, i.e. between t_1 and t_2, the PMOS is conducting and the voltage of node LX is near supply, V_{IN}, while the other end of the inductor is kept at V_{OUT} by the output capacitor, so the current increases at a slope $dI_L/dt=(V_{IN}-V_{OUT})/L$. If the load current demand decreases the current waveform maintains substantially the same slopes, but moves downwards to reduce the average current supplied to the capacitor and eventually to the load, as illustrated by dashed curve **1002** representing a lower average current. As shown in the lower part of the FIG. **4**, the voltage at LX does not quite reach ground or V_{IN} while passing this current, due to the I.R drop of the conducting NMOS or PMOS. For a converter with good efficiency, these I.R drops are small compared to V_{OUT}, so do not greatly affect the current waveforms, but are exaggerated in FIG. **4** for the purposes of explanation.

[0103] As mentioned, if the load current demand decreases, for a constant V_{OUT} and V_{IN} the current waveforms shown in FIG. **4** will fall. The required valley current may fall to zero or may go negative as the current demand decreases. The inductor current going negative is inefficient as this means that the output capacitor is charging back into the converter (and to ground through the NMOS switch). In essence charge supplied to the output capacitor **C1** is therefore wasted.

[0104] One aspect of the present invention therefore provides a mode of operating a DC-DC converter in Valley Current Mode that prevents the inductor current from reversing, i.e. the current flow going negative and charging back to the DC-DC converter. This aspect of the present invention relates to control mechanisms for DC-DC converters that maintain a high efficiency.

[0105] In this mode of operation the low side supply switch, i.e. the NMOS, is turned off before the inductor current goes negative. The low side switch is therefore turned off before the high-side PMOS switch is turned on. In other words there is a divorce in the points in the inductor current cycle between the turn-off of the NMOS power switch, and the turn-on of the PMOS power switch, so there is a phase where neither conducts, and possibly even a phase of no inductor current.

[0106] Operation in a mode where the inductor current is maintained at zero for part of the duty cycle is known as Discontinuous Conduction Mode (DCM) or sometimes as Discontinuous Current Mode. DCM is a known mode of operation in Peak Mode Control DC-DC converters. This aspect of the present invention implements DCM in a Valley Current Mode control system. Operation of a DC-DC converter where there is no divorce in switching between the PMOS and NMOS is referred to a Continuous Conduction (or Current) Mode (CCM).

[0107] FIG. **5** shows a DC-DC converter according to an embodiment of the invention. The DC-DC converter comprises an output stage **202** coupled to an inductor L and output capacitor C1 as described above with relation to FIG. **2**. The DC-DC converter also has a level shifter **204** and voltage error block **205** as described above in relation to FIG. **2**. However the current loop block **501** of this embodiment comprises further elements Zero Cross Detect circuitry **850** and NDiodeDetect circuitry **750** connected to the drain of the NMOS **20**. Also the Ramp Generator **500** is modified to include an optional extra slope, and the Current Sensor Amp **800** has a Hold function. Duty modulator **900** is also modified to provide the additional mode of control. The operation of these blocks will be described below. Embodiments of current sense circuit **800**, ramp modulator **500** and duty modulator **900** are shown in FIGS. **12** to **14** respectively.

[0108] The operation of the circuit including inventive features will be described with respect to the operational waveforms shown in FIGS. **6** to **9**, starting with operation at high load currents, and progressing through different regimes of operation at progressively lower currents. It is assumed that V_{IN} and V_{OUT} are constant during this sequence.

[0109] As stated above, there is now a divorce in the points in the inductor current cycle between the turn off of the NMOS power switch, and the turn on of the PMOS power switch. Ideally, the NMOS should be turned off exactly when the current through it has decayed to zero. This could be detected by monitoring when the drain-source voltage across the NMOS has fallen to zero, for instance by using a comparator. However, inevitably there will be propagation delays in comparator responding and in driving through logic and pre-driver circuitry to actually control the NMOS to turn off. Also the comparator may have a random input offset voltage.

[0110] In this DC-DC converter therefore a signal derived from the source-drain voltage of the NMOS is compared, by Zero Cross Detect circuitry, to a threshold level corresponding to some pedestal current level Iped. The pedestal current level Iped is chosen such that, with worst case delays and offset, the NMOS switch **20** will be effectively switched off before the current through it changes polarity. When the pedestal current limit Iped is reached the Zero Cross Detect Circuitry **850** generates a signal, ILIM_ZC instructing the Duty Modulator to turn the low-side switch **20** off. The zero cross detection circuit **850** may therefore comprise a comparator arrangement as will be understood by one skilled in the art. The comparator arrangement may be a zero crossing comparator as described in which case the pedestal level Iped is achieved by choice of an appropriate offset. It would alternatively be possible to use a comparator with zero offset, or a different offset, and compare against an appropriate threshold level.

[0111] This Iped value may be fixed by design, or may be designed to automatically alter in synchronism with say the V_{SEL} signal (setting the desired output voltage) or a measured output voltage. The Iped value may be programmable to account for a specified inductor value. In general Iped may be programmable to account for variables that may alter current slew rates and thus the current headroom needed for a fixed delay time. The skilled person will be aware of how to generate a reference level equal to a desired Iped for use in the comparison circuit.

[0112] This deliberate offset in the comparator threshold introduces a zone of operation on the boundary between CCM and DCM where inductor current never quite decays to zero. Embodiments described herein allow current mode control to be maintained in such a zone of operation, and into full DCM which offers advantages in terms of seamless control, reduced transients and the ability to enter pulse skip mode of operation.

[0113] Transition between CCM and DCM

[0114] To illustrate operation in this region, FIG. **6** shows a selection of inductor currents and voltages at node LX for the same V_{IN} and V_{OUT}, as the average output current is reduced further. To account for offset voltages in the zero cross detector **850** and propagation delays before the output NMOS effectively switches, and ensure that the NMOS turns off before the inductor current reaches zero, i.e. before the NMOS drain voltage reaches zero, the current threshold level is set to a positive level, Iped, illustrated as threshold **1007** in FIG. **6**. Typically this would correspond to a drain voltage of the order of a few tens of mV. As mentioned above, because of the propagation delays, the power NMOS will usually shut off a short time later and thus at a current Ibreak (line **1008**) somewhat less than Iped. To a good approximation, the delays will not vary much with current, and the slope of the current waveform is a constant V_{OUT}/L, so the difference between Iped and Ibreak will be constant and hence break will also be constant (for inductor currents that are above Iped when the NMOS turns on—operation in the regime when the inductor current in below Iped when the NMOS turns on will be described later).

[0115] At the instant (I=Ibreak) when the NMOS effectively switches off, there will still be inductor current flowing towards the load. Since the NMOS is off, this current is forced to flow through the drain-bulk diode instead. The skilled person will appreciate that the NMOS transistor will have a parasitic drain bulk diode. FIG. **3** shows the output stage **202** of DC-DC converter with the parasitic diode **301** of the NMOS switch **20** shown as a separate parallel diode for the purposes of explanation. The PMOS switch **10** likewise has a similar parasitic diode **302**.

[0116] When the NMOS device **20** is turned on, its associated parasitic diode **301** is shunted to ground by this NMOS **20** and, so long as the current through the switch **20** is not sufficient to induce an I.R drop larger than a diode's on-voltage (say 0.5V) across the parasitic diode **301**, the diode **301** will not turn on. The diode **301** is however active if the NMOS device **20** turns off while there is still a current flowing from ground to the drain node: this will always be true if the current in the output filtering inductor L is flowing to the output.

[0117] When the inductor current reaches zero, the diode current will extinguish and the LX node will move from its negative voltage state, and will, if sufficient time is permitted, reach a steady-state output voltage substantially equal to that on the output voltage V_{OUT}. (In practice there will be some damped oscillations, not illustrated, at this transient due to stray capacitances and inductances on this node).

[0118] It should be noted that if there is a current in the opposite direction when the NMOS **20** is turned off, the node LX will fly high until it can find a sink for its current, possibly a similar body diode **302** inherent to a PMOS **10** providing the high-side drive of this output stage **202**.

[0119] Whilst an NMOS switch will have an inherent body diode associated with it, a separate diode (or other unidirectional current device) could be specifically provided in parallel with the low side switch **20** to control current flow when the low side switch is off. This separate diode could be arranged to allow forward current to flow if the low side switch is turned off (before the PMOS is turned on) but to prevent current reversal in the inductor. Thus the diode **301** shown in FIG. **3** could alternatively be a separate external diode. Any suitable diode, preferably such as a low drop diode or Schottky diode could be used. If required an external diode could be provided for the high side switch as well. The rest of the description will refer to a body diode or parasitic diode of an NMOS switch for clarity but it will be understood that other external diodes could be used instead.

[0120] Referring back to FIG. **6** conduction through the low side parasitic diode **301** causes a change in inductor slope from V_{OUT}/L to $(V_{OUT}+\Phi)/L$ where Φ is the diode voltage. (This diode voltage will vary somewhat as the current decays, but may be of the order of 0.5V at these low currents). As described above, the break point of the slope will be at a current Ibreak largely independent of load current demand in this regime, i.e. mode, of operation.

[0121] Curves **1003** show the inductor current and voltage at node LX for operation in this mode. It can be seen that these waveforms correspond to a lower average current demand than those of **1001** and **1002** discussed above with relation to FIG. **4**. During a first part of the cycle the NMOS **20** is switched on at a clock edge as described previously. During this time the node LX is near ground and the inductor current decreases with a slope of V_{OUT}/L. At a point during the cycle the inductor current drops to Iped which starts the process of turning the NMOS switch **20** off. Propagation delays and offset errors however mean that the switch **20** actually only turns off at point **1004** later when the inductor current has reached Ibreak. At this point the NMOS **20** is off but there is still current flowing in the inductor and so current now flows through the body diode **301** of the NMOS **20**. The voltage at node LX jumps to a diode voltage below ground and current flows in the inductor with the increased slope of $(V_{OUT}+\Phi)/L$.

[0122] It will be appreciated therefore that a switch such as the NMOS **20** and PMOS **10** has various different states of operation. In an on-state the channel of the NMOS **20** (or PMOS **10**) is conductive and current can flow in either direction through the switch **20** (or **10**). In an off-state the channel is not conductive. However, under certain conditions current may flow through a body diode associated with the switch. Such a body diode passes current in one direction only i.e. the body diode is a uni-directional automatic switch. Once the current flow through the body diode is extinguished there is no current flow through the switch element at all. As used in this specification the terms "on" and "off", when referring to switches, refer to whether the controlled region of the switch is conducting or not. Thus, turning the NMOS **20** off refers to turning its channel region off so that no current flows through the channel region.

[0123] Having turned the NMOS **20** off, the PMOS **10** still needs to be turned on at some point. The required switching point for the PMOS turn-on can be derived graphically from the plot of FIG. **6**. For a given peak current, the slope of the PMOS conduction region is $(V_{IN}-V_{OUT})/L$, that of the initial NMOS conduction region is V_{OUT}/L, until I=Ibreak, and then $(V_{OUT}+\Phi)/L$. The required switching point is then defined graphically. As can be seen from FIG. **6**, this switching point moves to the left, denoting a higher duty cycle—physically the PMOS has to switch on longer to compensate for the higher loss of current during the phase when both NMOS and PMOS are off.

[0124] As mentioned above the use of a deliberate offset, Iped, such that the NMOS switch turns off before the inductor current reaches zero introduces a zone of operation on the boundary between CCM and DCM where the inductor current never quite decays to zero before the PMOS is turned on. The boundary of this zone is illustrated by curves **1005** and **1006**. Curve **1005** represents the case where the inductor current reaches the Iped threshold but the propagation delays mean that the NMOS is turned off just as the PMOS is turned on. Curve **1006** represents the boundary where the NMOS is actually turned off and conduction through the parasitic unidirectional switch, i.e. body diode, **301** of the NMOS occurs but the inductor current only just reaches zero as the PMOS is turned on.

[0125] This mode of operation therefore represents a transitional mode between traditional CCM and DCM. This mode may therefore be referred to as a "discontinuous switching mode" as the step of switching the NMOS off is not performed as part of the same process of switching the PMOS on. The skilled person will of course appreciate that a DC-DC converter may be designed such that the NMOS always turn off before the PMOS turns on (or that the PMOS turns off before the NMOS turns on) to avoid a conduction path straight from V_{IN} to ground. However such staggered switching is performed as part of the same switching process. In this mode of operation the NMOS switch is turned off as part of a separate process to the PMOS turning on as will be described.

[0126] For a continuous transition between CCM and DCM, it is desirable that the PMOS turn-on still be defined by the Duty Modulator **900**, i.e. by a comparison of the sensed current signal ISNS with the voltage loop error signal V_{ERROR} and a ramp signal waveform, possibly with slope compensation included in the ramp signal. However, as evident from the voltage waveform, the inductor current can no longer be estimated from the NMOS drain-source voltage. The inductor current is now passing through the body diode, and the voltage jumps down to a diode voltage drop below ground.

[0127] On a normal, un-isolated n-well CMOS process, the anode of diode **301** shown in FIG. **3** will be the substrate into which all the devices are manufactured. It is thus impossible to isolate just this part of the substrate, for example to place a series resistor between the anode and ground to sense the diode current. Thus, any current passing through the diode **301** cannot be detected. An isolated process, either with extra process steps to build a junction-isolated pocket of p silicon under the NMOS, or a silicon-on-insulator technology, would allow a sense resistor between the anode of the diode and ground, probably also connected to the NMOS source to sense the total current. Were a separate external diode present a sense resistor could be connected to the external diode. However, as discussed above, a series resistance will cause undesirable ohmic power losses.

[0128] Rather than sense the current by monitoring the NMOS itself, a separate series sense resistor could be used. If attached to LX, this will require differential sensing that would have to recover from the large common-mode signal when the PMOS **10** is on. Again the I.R drop required to produce an adequate signal would also impair the converter efficiency. A sense resistor in series with the source terminal would be more convenient but would fail to gather at least some of the current from the drain diode, much of which would flow to substrate and guard rings rather than laterally to the source. If the converter included an external diode in parallel with the low side switch to allow unidirectional current flow after the low side switch is turned off a sense resistor could be used in a path common to both the low side switch and external diode but as mentioned above use of a sense resistor would still add an I.R drop.

[0129] The embodiment shown in FIG. **5** therefore employs a lossless approach to detecting the current flow in the inductor but still allows for current loop control when the NMOS is off and hence the current flow can not be directly detected. As shown in FIG. **5** the current sense amplifier monitors the source drain voltage of the low side supply switch **20** to derive a signal indicative of the current flow through the NMOS **20** when the NMOS **20** is on. An aspect of the circuit shown in FIG. **5** is that when the zero cross detect circuitry **850** detects that the current flow through the NMOS reaches the threshold level Iped, and the output signal ILIM_ZC becomes true, the current sensor amp, **800**, is placed into a hold mode (by the derived HOLD signal shown), whereby the ISNS output current prevailing at the time that the hold signal became true is held on the ISNS output. In other words a value for the current sense signal, indicative of the inductor current, at the time just before the NMOS turns off is held as the ISNS signal after the NMOS has turned off i.e. the inductor current is effectively sampled and stored just before the low-side switch **20** turned off.

[0130] The true inductor current is then emulated by adding a component to the held ISNS value with a slope of $-(V_{OUT}+\Phi)/L$, before this signal is used for comparison in the Duty Modulator **900**. Practically, it is more convenient to add this slope to the ramp signal slope to get the same effect and thus, in one embodiment, ramp generator **500** comprises circuitry for adding an additional slope. This slope represents the emulated change in inductor current when the NMOS **20** is off and the change in inductor current is due to conduction through the parasitic diode **301** of the NMOS **20**. As mentioned above this ramp signal may be added to the slope compensation signal (which itself may be applied to either the ISNS signal or V_{ERROR} signal) but alternatively a ramp could be applied to the ISNS current signal to produce an emulated current signal. The emulated ramp could alternatively be applied to the V_{ERROR} signal. The emulated change in inductor current could be emulated by two or more ramp signals which are applied to various different signals so as to produce a combined effect which represents the change in inductor current when the NMOS is off.

[0131] The embodiment shown in FIG. **5** comprises circuitry for emulating the change in inductor current during a period when the low side switch, i.e. NMOS **20**, has turned off but there may still be conduction due to the fact that the NMOS switch **20** is designed to turn off before the current reaches zero. As described above the change in inductor current may be emulated accurately taking into account the exact voltage drop across the inductor during the period when the NMOS is turned off, however a reasonable approximation to the change in the inductor current is sufficient for many converters. By emulating the change in inductor current is meant providing, by any means, a representation or replication of the change of the change in inductor current, whether exact or estimated or approximated, i.e. any means of imitating or mimicking the expected change in inductor current.

[0132] Various solutions for emulating the change in inductor current could be envisaged and could be applied in alternative embodiments. For example, in an alternative arrangement when signal ILIM_ZC goes true, indicating that the NMOS **20** is about to turn off, the Duty Modulator **900** may switch to use a different current signal which, for example, emulates not only the change in inductor current but also the absolute value of inductor current, based on a knowledge of Ibreak. However, holding a value of the current sense signal ISNS is convenient as it preserves the actual value of the signal and means that the same signals are used by the Duty Modulator **900** during the period when the NMOS **20** is on, i.e. conducting, and the period when the NMOS **20** is off, i.e. non-conducting. Also, as will be described later, at lower currents again the value of Ibreak may change depending on the starting current when the NMOS is turned on. Similarly adding a ramp signal to the slope compensation signal simply involves changing the slope of a ramp which is already applied.

[0133] As will be clear from the discussion above the change in inductor current during the period when the NMOS is off, but the NMOS body diode is passing current, has a slope equal to $-(V_{OUT}+\Phi)/L$. For simplicity, however the emulated change in inductor current may be approximated by $-V_{OUT}/L$ without greatly disturbing the loop accuracy or dynamics, any error being corrected for by the voltage feedback loop. And in cases where V_{OUT} is programmed, for example by a digital input signal V_{SEL}, this extra ramp rate may be digitally programmed by the V_{SEL} signals, rather than by monitoring the output voltage directly. This avoids loading the output, making the converter more efficient, and removing the need for an electrically and physically large on-chip resistor. Any slope which approximates the actual change in inductor current may be used for the ramp but the more accurate the ramp slope the better the overall performance.

[0134] If the load current demand decreases further, eventually the real valley current reaches zero, i.e. the current through the NMOS body diode has time to decay to zero before the PMOS is due to turn on. The converter is now operating in a full DCM mode.

[0135] DCM Operation

[0136] Operation in full DCM is illustrated in FIG. **7**. Waveform **1009** illustrates a current level where the NMOS turns off at point **1010** and the inductor current reduces to zero at point **1011** before the PMOS turns on a point **1012**.

[0137] As described above, once node LX is no longer driven by either the NMOS or the PMOS, and the diode conduction stops, this node presents a high impedance to the inductor L. Ideally the node LX would just step up to a voltage equal to V_{OUT}. In practice the effect of parasitic inductances and capacitances on this high impedance node will result in voltage oscillations that may take some time to dampen down, but this does not affect the operation of the circuit.

[0138] The extinguishing of the inductor current is detected by the NDIODE DETECT block, **750**, of FIG. **5**. This circuitry monitors the voltage seen at the LX node. When the diode current extinguishes, the LX voltage starts to rise from a diode voltage below ground. When a threshold, such as zero, is met, the NDIODE_OFF signal of FIG. **5** goes true, causing the extra slope to no longer be added to the slope compensation signal. NDiode Detect circuit **750** therefore comprises a relatively simple comparator circuit as would be well understood by one skilled in the art.

[0139] At the point at which the NDIODE_OFF signal goes true the inductor current is zero and the change in the inductor current is also zero. Therefore the slope of the additional ramp signal is no longer applied. In the embodiment shown in FIG. **5** the ramp generator slope therefore changes back to the slope due to the slope compensation ramp only. The current sense signal ISNS is still maintained at the constant value held since the NMOS was switched off. Although the inductor current is now zero the change in inductor current has been added to the slope compensation signal to effectively null out the contribution from the ISNS signal and hence the ISNS signal must be maintained. Again however there are many ways in which the emulation of change in inductor current could be achieved. In other arrangements, where the additional ramp signal is applied separately the value of the additional ramp signal should ideally exactly compensate for the value of the held ISNS at the point at which the NDIODE_OFF signal goes true. It would therefore be possible to discontinue both of these signals at the same time and simply drop to a value of zero on the ISNS signal. In practice however there may be a small offset between the value of the additional ramp signal and the held ISNS signal at the point at which NDIODE_OFF goes true. A small offset will not greatly affect the current loop control and so the values of both signals may be maintained to avoid a sudden jump to zero. This will avoid possible hunting or discontinuity of operation of the feedback loop around this loop operating point, and so assist the smoothness of the transition between operational regimes.

[0140] Whilst the NMOS body diode is conducting the rate of change of the inductor current is emulated and when the NMOS body diode stops conducting, indicating that the inductor current is zero, the emulated change is inductor current is also reduced to zero. It will be appreciated that the circuitry of this embodiment of the invention therefore effectively emulates the inductor current during the whole period that the NMOS switch is off but that for part of this period the emulated current remains substantially at zero. Therefore, both zero and non-zero current may be emulated. As mentioned, due to inaccuracies in the signals produced the effective emulated current may not be exactly zero. An effective emulated current of near zero is sufficient to allow the current control loop to function correctly. In general the emulation circuitry effectively emulates a changing inductor current in a first period, when the NMOS is off and current is flowing through the body diode, and a substantially static inductor current at or near zero in a second period when the inductor current has stopped and before the PMOS turns on.

[0141] The ramp including slope compensation still continues however, and eventually causes the Duty Modulator **900** to switch the PMOS **10** back on at the time required. This operation is described in more detail later with reference to FIG. **10***a*.

[0142] Also, as the current demand falls, the PMOS will be switched on later and later. (Note that in full DCM mode, such as illustrated in FIG. **7**, a drop in current demand leads to a reduction in duty cycle whereas in the transition phase between DCM and CCM, as illustrated in FIG. **6**, a drop in current demand actually leads to an increase in duty cycle). Thus, operating in DCM a drop in current demand may lead to a shorter duty cycle and a shorter period during which the PMOS is turned on.

[0143] However there is a minimum feasible pulse width for control of the PMOS switch and thus a minimum period in which the PMOS switch may be operational. This means that there is a minimum charge that may be passed by the PMOS each cycle. If current demand falls even further, then this minimum charge passed by the PMOS each cycle may be more than required by the load, so the output voltage would continue ramping up.

[0144] Pulse Skip Mode

[0145] In one embodiment a minimum conduction period T_{pmin} is defined for the PMOS, to give the PMOS time to turn on before reaching the clock edge where it must turn off. If the desired duty cycle reduces such that PMOS on time would be below this minimum conduction period the PMOS is prevented from turning on until the output voltage V_{OUT} has drooped enough to require recharging.

[0146] Thus, in this embodiment, a Pulse-Skip mode is enabled, not with an error voltage threshold or with an average current, but rather with this predetermined minimum power PMOS switch conduction period limit T_{pmin}.

[0147] A pulse skip mode of operation is thus advantageous in that it allows operation at low current demands. Pulse skip mode is also advantageous in that any switching losses associated with switching the power switches **10** and **20** are avoided whilst skipping pulses. This can avoid significant losses associated with operation of the converter herein illustrated and thus improve the efficiency of the converter at low current demands.

[0148] If the PMOS conduction period would be less than T_{pmin} in any particular DCM clock cycle, then the power PMOS switch conduction period is omitted in that particular cycle. While PMOS pulses are suppressed, the output voltage will sag, forcing the error voltage, V_{ERROR}, to rise eventually. The current, previously sampled by the current sensor block **800** at the last NMOS conduction period, is still held, the slope compensation is still active, and the loop will eventually recover normally by turning on the power PMOS switch **10** for a time greater than T_{pmin}. The PMOS switch will be turned on at a time defined as normal by the Duty Modulator based on the previously sampled current, the current slope compensation ramp, and the error voltage V_{ERROR}. This regime of operation is very much under voltage-mode control but this does not matter since it is guaranteed that the output current waveform is discontinuous, so the normal double-pole response from an inductor and capacitor filter so characteristic of voltage mode CCM, drops to a single pole response.

[0149] FIG. **7** shows a current waveform **1020** that would be suppressed by the pulse skipping.

[0150] It will be clear that in order to implement this embodiment one needs to determine the duration that the PMOS will be on for, prior to turning the PMOS on or not.

[0151] This duration, in VCM, corresponds to time between the current signal reaching the threshold level and the next clock edge turning the PMOS off, i.e. the start of the next cycle. The duration may be determined by monitoring the time between the last clock edge, i.e. start of the current cycle and the threshold being reached and determining the remaining cycle time based on a knowledge of the cycle frequency.

[0152] In a preferred embodiment however the system clock is delayed by an accurate delay. Thus the external clock signal **212** is delayed by a fixed delay, greater than or equal to the minimum PMOS conduction period, before being used to control the switching of the PMOS and NMOS. A signal derived from the undelayed clock signal can then be used to control timings. For example if the system clock is delayed by a period P and the minimum PMOS conduction period is T_{pmin} then, if the switch control signal is generated before a time T_1, where $T_1=P-T_{pmin}$, after the undelayed clock edge, the PMOS is switched on whereas as the switch control signal is generated after T_1 the pulse is skipped.

[0153] The apparatus may therefore comprise a delay means for delaying the clock edge used for switching by a known delay and a timer circuit which operates based on an undelayed (or less delayed) clock edge. The timer circuit may assert a no_switch signal a certain time after the undelayed clock edge until reset by the delayed clock edge. In this way the no_switch signal is asserted for the period of T_{pmin} before the delayed clock edge. Turn on of the PMOS is inhibited when the no_switch signal is asserted.

[0154] FIG. **16** illustrates one embodiment of a circuit arrangement that may be used to implement a pulse skip mode of operation. This circuit may form part of the Duty Modulator **900** shown in FIG. **5** or it may be a separate circuit for supplying signals to the Duty Modulator. A master clock signal CLK is received at an input **1601**. This signal is input to delay element **1602** to produce a delayed version of the clock signal. One delay element **1602** is shown in FIG. **16** but more than one delay element may be used. The delay element **1602** delays the master clock signal by a period t_d. The delayed signal CLK_SW is used as the master switching clock in the state machine of the Duty Modulator. A pulse generator (not shown) may be located before or after delay element **1602** to ensure that the delayed clock signal has a clock pulse with a desired pulse width, for example 20 ns or so. The master CLK signal is also received by a pulse generator **1603** which generates a pulse skip signal CLK_PS. Pulse skip signal CLK_PS goes high in response to the rising edge of the master clock pulse and remains high for a period at least equal to t_d. Preferably pulse skip signal CLK_PS has a pulse width greater than t_d so that it overlaps with the CLK_SW signal to avoid glitches. For example the CLK_PS signal may remain high for a period equal to t_d plus the duration of the pulse of the delayed clock signal.

[0155] The CLK_PS signal is received by another part of the Duty Modulator circuit which is arranged so that the PMOS switch can not be turned on whilst the CLK_PS signal is high. Thus, for a period of time before the start of the next cycle (as defined by the CLK_SW) signal the PMOS switch is inhibited from being turned on. The duration of this period, T_{pmin}, is effectively set by t_d and any propagation delays inherent in the logic. Ignoring propagation delays the period T_{pmin} is equal to the delay duration t_d. The delay duration may be set according to the frequency of operation of the converter.

[0156] It will be appreciated that other clock signals may be derived from the master clock, for instance a reset clock signal for resetting various logic blocks and/or one or more clock signals used to clock various logic block, flip-flops etc. However the only the signals directly relevant for pulse skip are shown for clarity.

[0157] As the current demand falls even further, the current through the NMOS will decay below the threshold Iped earlier and earlier. FIG. **8** shows waveforms for lower current levels, but with the time axis shifted half a period to focus on the region of interest. Also the relative slopes of the charging and discharging of the inductor have been changed, with a slower PMOS slope, representing a smaller V_{IN}-V_{OUT} relative to V_{OUT}, to illustrate a problem of operation.

[0158] As discussed previously there is a propagation delay T_{np} between when the sensed current passes the Iped threshold and when the NMOS actually turns off. So in the case of a current that is exactly Iped when the PMOS turns off and NMOS turns on, as illustrated by waveform **1013**, the NMOS will be kept on for a time T_{np} turning off at Ibreak. Also the sensed current is preset high before the PMOS/NMOS transition and the comparator of the Zero Crossing Detect circuit is reset, so even if the current at the PMOS/NMOS transition is less then Iped, as shown by waveform **1014**, the NMOS will still be kept on for a time T_{np}. So the sensed current when the NMOS turns off will decrease below the previous value of Ibreak, rather than stay constant. Thus the current decays initially at a slope V_{OUT}/L until time T_{np} and then, as the output node jumps to a diode voltage drop below ground, decays at $(V_{OUT}+\Phi)/L$ until decayed to zero.

[0159] The PMOS then turns on at the appropriate time before the next cycle starts to give a steady state peak current and hence the required average output current to the load. As above, the turn-on is still under control of the Duty Modulator **900**, with the sensed current sampled and held corresponding to Ibreak (with Ibreak now decreasing) and with the emulated slope added during the diode conduction period. If the current demand were to decrease further then the PMOS would turn on later, eventually getting into the pulse-skip mode as discussed above.

[0160] However, as shown in the solid curve **1016** in FIG. **8**, at very low average output currents, when the power NMOS switch **20**, is forced on for its minimum T_{npp}, the condition will eventually be reached when the inductor current is zero or negative by the end of the NMOS conduction period, i.e. the NMOS will start to sink current from the load. When the NMOS eventually turns off, the inductor will still require current in this direction, so node LX will fly high until being caught by the PMOS body diode **302**, which will then pass this current. The current will decay, at a rate $(V_{IN}+\Phi-V_{OUT})/L$, to zero, whereupon the driver output will become high impedance and the output will fall to V_{OUT}, albeit probably with some voltage oscillation in a real circuit. This voltage spike is undesirable per se, but more importantly this activity will waste power, first in unnecessary dissipation in the NMOS and the NMOS taking power away from the load, then in power loss in the PMOS body diode. In detail, on an n-well CMOS process, the PMOS "diode" will actually act as a vertical pnp transistor, with the drain as the emitter and V_{IN} as the base, and the substrate as collector, so that much of the diode current will flow to substrate through a voltage drop of VIN+Φ, which is much larger than just the diode drop Φ. Also there is the possibility that the NMOS channel current can cross zero early enough for the subsequent flow of charge from the load to actually be equal or greater than the charge supplied to the load earlier in the cycle, i.e. to actually fail to charge the load at all.

[0161] The outer loop would eventually compensate for these effects, but the reverse current through the NMOS would disadvantageously cause extra power losses.

[0162] Furthermore the loop might not be able to operate in a pulse-skipping mode in these conditions, so the power saving possible by pulse-skipping would not be realised. Under some conditions of V_{IN} and V_{OUT} as the current demand falls the condition described above, where the minimum NMOS conduction time would result in a negative current flow in the inductor, may occur before the duty cycle has reduced to less than the defined minimum PMOS conduction time T_{pmin}, i.e. before pulse skip mode has been entered. This power wastage may keep the current demand required sufficiently high that the duty cycle is always above the PMOS minimum conduction time. Thus pulse skipping mode would not be entered.

[0163] Diode Mode

[0164] To avoid this problem, in accordance with the present invention, the DC-DC converter is operable in a mode, which may be termed a "diode mode", where the high side switch may be turned on but, when it is turned off, the low side switch is not turned on and current flows through a parallel path. In one embodiment of the invention, after the PMOS switches off, the current is carried by the NMOS body diode (which will automatically shut off when it becomes reverse biased) rather than by the NMOS channel (which can conduct in both directions). In other words, rather than operate such that the NMOS is always turned on after the PMOS has been turned on, in diode mode the PMOS may be turned off and the NMOS maintained in an off state such that any current is conducted via the body diode only. As mentioned above an NMOS has an inherent body diode but, in another embodiment of the present invention, a separate external diode, or other unidirectional switch element, could be provided in parallel with the low side switch instead. In this diode mode the PMOS may be turned on when the (emulated) current signal crosses the V_{ERROR} threshold (as modified with a slope compensation ramp) in a first cycle and then again in a subsequent without the NMOS being turned on in between. This diode mode of operation may occur for several successive cycles.

[0165] In principle, the NMOS could be turned off as soon as the current crosses zero. But, due to the same propagation delay and offset issues described above, the NMOS would only actually turn off after a delay similar to T_{np}—which would be too late.

[0166] One embodiment of the invention therefore detects whether the current has reversed, i.e. crossed zero, during one cycle, and, if so, forces the NMOS off in the next cycle. In other words a negative current is detected in one cycle and such detection prevents a negative current situation from reoccurring in the next cycle. Following such detection, the NMOS may be inhibited from turning on at the next clock edge. However by then there has already been one cycle with undesired behaviour. Further, such an arrangement requires a detection of current flow in the NMOS. Thus the NMOS has to be on in one cycle so that the diode mode can be instigated for the subsequent cycle. In an extreme case this could lead to the undesired behaviour every other cycle.

[0167] In a preferred embodiment a minimum NMOS conduction time threshold, T_{nmin}, that is somewhat longer than t_{np}, is defined. If the current reaches zero before this extended time, then the control circuitry forces the NMOS not to turn on the next cycle. This is illustrated in FIG. **9**. Solid curve **1015** illustrates a waveform at a low current demand. The initial part of the curve shows the PMOS on phase where the inductor current increases. At the next clock edge the PMOS turns off and the NMOS turns on. The current in the NMOS is already less than Iped but the propagation delays mean that the NMOS does not turn off until the end of period T_{np}. At this point the inductor current is still positive and so diode conduction occurs until the inductor current reaches zero at point **1018**. The point at which the diode current reaches zero is less than the defined minimum conduction period, indicating that the current demand is close to the limit where the minimum NMOS on time, T_{np}, may lead to a negative current. Hence the converter enters diode mode and inhibits the NMOS from turning on in a subsequent cycle. In this subsequent cycle the PMOS may still turn on at the same time and hence the first part of curve **1015** would be repeated. However the NMOS would be prevented from turning on at the next clock edge.

Thus in this cycle, the drain-bulk diode of the power NMOS switch sources the current set up in the inductor by the PMOS. As described previously this slew rate will be higher, $(V_{OUT}+\Phi)/L$ rather than V_{OUT}/L, so the current will decay even earlier, as shown by curve **1017** which represents the current and voltage waveforms in diode mode.

[0168] The system remains in this mode until the current demand increases or the PMOS current recovers enough to drive the zero-cross point past T_{nmin} (illustrated by dashed curve **1019**) after which the previous mode of operation will resume. As mentioned, when operating in diode mode, the slope of current decrease is greater than in non-diode mode. It can be seen therefore that there is an inherent degree of hysteresis in entering and exiting diode mode in that the peak inductor current required to enter diode mode is lower than the peak inductor current to exit diode mode. It will, of course, be appreciated that once operating in diode mode, as the current decays more quickly than in non-diode mode operation, the converter may need to slightly increase the charge supplied during the PMOS conduction time to compensate.

[0169] By setting a conduction time threshold which is longer than the minimum propagation delay T_{np}, this embodiment of the present invention pre-empts the undesired behaviour. If the NMOS is on for a duration in one cycle which is less than a time threshold (i.e. the defined duration T_{nmin})—but which is greater than the minimum propagation time T_{np}—there will be no reverse current but the next cycle the NMOS will not turn on. The use of diode mode guarantees that pulse skip mode can be entered when appropriate.

[0170] It should be noted that in diode mode the turn on of the NMOS is inhibited in a subsequent cycle. The inhibition could be applied in the very next cycle. However, if in the next cycle the converter enters pulse skip mode there will be no current flow in the inductor at all in the next cycle. In such a case the diode mode preferably inhibits the NMOS from turning on after the next time that the PMOS has been on. In other words, if the inductor current reaches zero in a time period shorter than T_{nmin} in a time period, the NMOS is inhibited from turning on at the clock edge following the next time that the PMOS has been switched on. It is preferable to maintain diode mode until the PMOS has been on to prevent diode mode being missed. Alternatively, diode mode may only apply to the very next cycle but if the converter has been operating in a pulse skip mode it could be arranged such that the first cycle where a pulse in not skipped (following a skipped pulse) is operated in diode mode automatically. In other words it may be assumed a converter exiting pulse skip mode will need to be operated in diode mode.

[0171] FIG. **15** illustrates an embodiment of a circuit arrangement, which may for example be implemented within Duty Modulator **900**, for implementing diode mode. The circuit arrangement samples the NDIODE_OFF signal every cycle at a period effectively equal to T_{nmin} to determine whether or not the inductor current has reached zero. The NDIODE_OFF signal is input to a latch **1501** which latches when the NDIODE_OFF signal goes true, i.e. the output from the comparator indicates that the voltage at Node LX reaches zero. The latch **1501** rejects oscillations on node LX after the conduction of the NMOS body diode has ended. Although shown as part of the diode mode circuit the latch **1501** could be implemented within the NDIODE Detect circuitry. Latch **1501** is reset by a reset signal sent to the NDIODE detector comparator. The output of latch **1501** is provided to D type flip flop **1502** and sampled at an appropriate clock edge. The clock edge is derived from the NMOS ON signal, which is derived from feedback from the switch driver, passed through delays **1503** and **1504**. Two delays are shown in FIG. **15** for explanation purpose but a single delay or more than two delays may be used to provide the correct delay. Delay **1503** provides a delay which effectively provides a delay matched to the zero cross comparator and NMOS switch off time, i.e. the minimum conduction time of the NMOS. Delay **1504** provides a delay to allow for delays in the NDIODE detect comparator and logic in generating the NDIODE_OFF signal.

[0172] When the delayed clock edge is received by D type flip flop **1502** the state of the latched NDIODE_OFF signal is sampled. If the output is false, this means that inductor current has not yet reached zero and diode mode is not initiated. If the output is true this means that the inductor current reached zero in a time less than T_{nmin} and Diode Mode is enabled. The output of the flip flop **1502** therefore comprises a diode mode signal which is used by the Duty Modulator to determine whether or not to inhibit turn on off the NMOS at the start of the next cycle.

[0173] It will be appreciated that in Diode Mode the NMOS is not turned on and so the feedback from the NMOS switch driver can not be used. In this instance an appropriate signal is created. The PMOS ON signal is inverted by inverter **1506** and delayed by delay **1507** by a delay corresponding to the delay between switching the PMOS off and switching the NMOS on. The result is a signal that goes high at the time that the NMOS would have turned on were the circuit not in Diode Mode. This signal is selected by multiplexer **1505** responsive to the Diode Mode Signal output.

[0174] It should be noted that diode mode has been described in relation to Valley Current Mode control of a DC-DC converter. However the principle of diode mode is equally applicable to a Peak Current Mode controller. In Peak Mode Control, the PMOS is turned on at a clock edge and is turned off when the current reaches the threshold. At this point the PMOS is turned off and the NMOS is turned on. To prevent power wastage, as described above, it may be preferred to switch the NMOS switch off before the current goes negative. However, as described above there will again be propagation delays and offsets associated with any comparison circuitry and hence there will be a minimum effective NMOS on time. Thus the minimum period T_{nmin} may be defined as described above and if, in any cycle, the NMOS current reaches zero in a duration less than T_{nmin} the NMOS may be inhibited from turning on in a subsequent period.

[0175] The operation of the embodiments of the present invention is now described including more detail of the Current Loop block and waveforms therein rather than focussing on the output waveforms.

[0176] FIG. **10***a* shows the control signals in detail when the DC-DC converter is operating in full DCM mode, as described above with relation to FIG. **7**.

[0177] The first trace shows the inductor current, similar to the current curve shown in FIG. **7**, with a first phase, to time t**1**, of decay with a slope of V_{OUT}/L when the NMOS is ON, a steeper slope of $(V_{OUT}+\Phi)/L$ as it passes through the NMOS body diode until the current decays to zero at t**2**, a third phase of zero current, until t**3**, when the PMOS is turned on.

[0178] The second trace shows the sensed current, processed by the block **800**. Initially this follows the droop of the inductor current, until it drops to Ibreak, where it is held constant. Once the PMOS turns on, the sensed current output is set to a higher value to ensure that it is well above the Iped threshold during any transient when it is enabled as the PMOS turns off and the NMOS turns on.

[0179] The third trace shows the current due to slope compensation plus additional ramp. This starts of with a slope equivalent to $(V_{IN}-V_{OUT})/L$ but increases in slope by nominally $(V_{OUT}+\Phi)/L$ between t**1** and t**2** to emulate the increase in slope of the actual inductor current. This extra slope is removed after the current goes to zero at t**2**.

[0180] The fourth trace shows the voltage at an internal summing node, where the ISNS and the slope compensation current are summed. The PMOS turns on when this crosses V_{ERROR}. Note the voltage loop will slowly settle to an appropriate V_{ERROR} value to generate the correct voltage after any change in load current demand.

[0181] The remaining traces are the control signals. PWM is an internal node pulsed high after the summing node voltage crosses V_{ERROR}, and reset by the edge of external clock CLK. ILIM_ZC gives a triggering pulse when the NMOS current crosses the pedestal threshold Iped, as determined by Zero cross detect circuit **850**. This sets the EMULATE signal which is the signal to add the additional ramp to the generated slope compensation to emulate the change in inductor current. EMULATE is reset by the NDIODE_OFF signal generated by the NDiode detect block **750** which senses the output voltage passing from negative to positive when the body diode stops conducting. The NDIODE_OFF signal is itself reset when the PMOS is turned on. HOLD is a combination of EMULATE and NDIODE_OFF and is used to cause the current sensor block **800** to hold the sensed current value. NGATE and PGATE are the drive signals for the NMOS and PMOS gates.

[0182] FIG. **10***b* shows operation in the transition phase between CCM and DCM operation, in conditions similar to those of the dotted curve in FIG. **6**, where the actual valley current is above zero but below Iped and Ibreak. The same signal traces are shown as in FIG. **10***a*.

[0183] The control scheme has to incorporate the current control loop in this region since the inductor current has not decayed to zero and the overall loop would be intrinsically second-order in nature without some current information to compensate it.

[0184] Unfortunately, the current sensor, **800**, cannot supply any current information if the power NMOS switch **20** is off. Instead, the current information is emulated by sampling the current prevailing in the sensor during the time when the switch is turned off, and is supplemented by the additional emulated current ramp while the drain-bulk diode **301** is conducting.

[0185] Operation is similar to the true DCM above except that the actual inductor current never decays to zero, so NDIODE_OFF is never triggered, and so the extra slope compensation ramp is still applied when the summing node voltage passes V_{ERROR}. Also the net effect of this emulation is that the overall inductor current ramp plus slope compensation ramp—to first order—is identical to that prevailing in CCM where the average inductor current is high enough so that the zero cross detection circuit never trips. The transitions between full CCM, CCM under NMOS diode conduction, and DCM are therefore smooth.

[0186] The embodiments of the present invention therefore provide a DC-DC converter operable in valley current mode and operable in a number of different modes. The use of the different modes allows the DC-DC converter to maintain accurate current mode control in CCM, and through a transitional phase to full DCM. In the transitional phase, although the NMOS is turned off the current loop is still active and current loop control is provided by emulating the change in inductor current. This embodiment of the present invention therefore provides seamless transfer between CCM and DCM and vice versa. The control mechanism uses a lossless current sensing approach and thus losses associated with a series sense resistor are avoided. The DC-DC converter may operate a fixed frequency in CCM, the transitional phase and DCM—although variable frequency operation may be implemented if desired. A pulse skip mode is also available which allows for significant efficiency gains at low current demands by preventing switching when unnecessary and thus reducing the relatively switching losses. A diode mode of operation may be used at low current demands to guarantee that pulse skip mode will be entered. These modes of operation are automatic in operation (if enabled). The onset of the transitional phase is set by the value of Iped and the converter will automatically operate in CCM, transitional phase or DCM depending on the load current. If the minimum conduction time periods are enabled the converter will also automatically enter pulse skip and/or diode mode as appropriate and the transition between modes will be automatic and continuous, i.e. the control method is the same between all modes and there is no change from one set of control signals to another. The control loop is maintained in each of CCM, the transitional phase, DCM and pulse skip, even though no current information may be available.

[0187] The skilled person will appreciate that CCM mode control offers the most stable control method but the other modes of operation offer efficiency savings. FIG. **11** illustrates the relative efficiency of a DC-DC converter operating purely in CCM, a forced CCM mode (FCCM), with a converter that automatically operates switches to DCM, via the transitional phase, and then to pulse skip mode at increasing low current demands. The curves show efficiency against current on a logarithmic scale. The curves shown in FIG. **11** apply to identical converters where only the mode of operation is changed and assume a constant V_{IN} and a constant V_{OUT} throughout the operating range.

[0188] The lower curve **1101** shows the efficiency of the converter operating in FCCM. At relatively high current loads, around 200-300 mA, the converter is very efficient. However as the load current falls the efficiency drops markedly, due partly to the fact that the NMOS will pass negative current for a part of each cycle.

[0189] The upper curve **1102** shows operation of a converter with automatic mode switching enabled. At currents of 200 mA or above the converter is operating in CCM mode and hence the efficiency is effectively the same as for a converter operating in FCCM. Below 200 mA the transitional phase through to DCM proper is starting and the curves start to deviate with the converter operating in DCM being more efficient. Below about 20 mA in this example pulse skipping starts and there is a jump in efficiency as the switching losses are avoided. At lower currents still it can be seen that the converter operating in pulse skip mode is significantly more efficient than a corresponding converter operating in FCCM.

[0190] These significant efficiency saving can have a large impact on battery life for portable electronic device applications. However in some applications the efficiency gains may not be as much of a concern and it may be preferred to operate in FCCM to provide the best control over voltage ripple and the like. Conveniently therefore a mode control signal input to the duty controller, for example mode control signal **106** shown in FIG. **1**, may determine whether the converter operates in FCCM or whether DCM and pulse skip modes of operation are enabled.

[0191] Various aspect and embodiments of the circuitry for implementing the modes described above will now be described.

[0192] FIG. **12** illustrates one embodiment of a current sensor that has a Hold capability.

[0193] The N-channel current sensor **800** interface to the power switch **20** is shown.

[0194] A reference voltage V_A is generated on the non-inverting input node of amplifier **805**. V_A is proportional to the current **801** and the series combination of the resistances of NMOS devices **803** and **804**. This bias voltage V_A is replicated at the inverting input node of amplifier **805** by the action of feedback, ensuring that node B, the drain of device **810**, is held at V_A.

[0195] In normal operation, switch **809** is closed and switches **807** and **808** are open, allowing a current path through devices **20**, **809**, **811**, **810**, **812**, and **813**. Since node B is held at a constant voltage the current through **809**, **811**, **810** is given by $(V_A-V(LX))/R_{SX}$ where R_{SX} is R**809**+R**811**+R**810**, i.e. the sum of the resistances of these three devices. Thus as the inductor current varies, the current through the NMOS also varies (the current through R**809** is small compared to that through NMOS **20** and inductor L), so the voltage on LX varies due to the varying I.R drop across NMOS **20**. This variation modulates the voltage across Rsx. Since Rsx is also a MOS channel resistance which will to first order track proportionately that of NMOS **20**, the net result is a modulation in current through Rsx through **812** to the output that is independent of MOS channel resistance and is a pure ratioed copy of the current through NMOS **20**, with ratio dependent mainly on the respective aspect ratios of the transistors involved.

[0196] Normally the current in the inductor is towards the load, as shown. This pulls the NMOS drain node LX below ground and increases the current through **809**. However in some modes or during transients, the current may be of opposite polarity. Ignoring current source **802**, the maximum reverse current detectable would be that causing LX to rise to a voltage V_A, i.e. V_A/R**20** since **812** can only source, and not sink, current. Including current source **802** reduces this maximum reverse current.

[0197] When the current sensor is in track mode, i.e. when the NMOS is on and the actual current through the NMOS is being monitored, switch **815** is closed, and the current seen at node ISNS is created by the current mirror comprising devices **813**, **814**, and **815**. When in Hold mode, as determined by the HOLD control signal (which is inverted for control of switch **815**), the gate of mirror device **814** is open-circuited, and the gate stays at constant voltage, and hence its output current remains at the same value until reconnected. Known techniques, such as half-size devices clocked in anti-phase for example, are used to cancel the charge injection onto this gate node that would be caused by this switch turning off. Extra capacitance between gate and supply may be added, though proved to be unnecessary in this design as a gradual droop over time has little effect on the regulator output.

[0198] To obtain an accurate scaling value for ISNS versus absolute externally applied current, the current mirror **813**, **814**, and **815** may be made a digitally programmable current mirror composed of multiple elements to give a programmable current transfer ratio by standard techniques. In manufacture, a predetermined external current may be applied and the voltage developed across an internal on-chip resistor may be monitored and the current transfer ratio modified until a predetermined target voltage is developed across the on-chip resistor. This then fixes the regulator's "sense resistance" (strictly a transresistance) R_{SENSE} to a known value relative to other on-chip resistances, enabling a tight tolerance of the loop transfer function. There will still be some variation due to other on-chip resistances/capacitances over temperature, but manufacturing variations in the short channel NMOS transistor resistance and offsets and current ratio errors in the Sensor Amp can be compensated.

[0199] The currents of devices **801** and **802** are inversely proportional to on-chip resistances similar to those used in the duty modulator block. Devices **803**, **804**, **810**, **811**, and switches **807**, **808**, and **809** are all scaled analogues of the resistance of the master power switch **20** in that they track each other over process, voltage supply, and temperature. That is the devices and switches are formed in the same area of the silicon so that any process variations affect all devices equally and the devices experience the same temperature and supply variations.

[0200] At the start of a period of duty cycle control, the current output of this current sensor should be reset to a value higher than that seen under normal conditions in order to prevent false triggering by the duty modulator block **900** of FIG. **5**. This reset path is achieved by closing switch **807** while keeping switches **808** and **809** open. This decouples the sensor from the power device **20** and, by virtue of the series combination of devices **810** and **807** being less than the series combination of devices **810**, **811**, and **809**, pushes the output current much higher than usual.

[0201] In order to save power, during some phases in DCM and pulse-skip modes, the path comprising devices **808**, **810**, **811**, **812**, and **813** is engaged which sets the ISNS output to the pedestal value seen during normal operation, rather than the higher reset value. This is achieved because the resistances of devices **808** and **809** are equal.

[0202] During CCM, switch **815** is closed. When the DC-DC converter is placed into the transitional mode between CCM and DCM, once zero current has been detected—actually always above zero, i.e. at Iped, to allow for offsets and propagation delays in the drive train—switch **815** is opened and the current produced on the ISNS output is held constant by the gate capacitance of device **814** at a level similar to that flowing just before the switch was opened.

[0203] FIG. **13** illustrates one embodiment of a ramp generator **500** which is capable of producing a slope compensation signal which may include a component emulating the change in inductor current.

[0204] It is apparent that the output current of the block is given by:

$$Iout(t) \equiv I_{512} + \frac{I_{507}}{M} - \frac{R_{504}}{R_{505} * M} * I_{506} - \frac{I_{506} + I_{503}}{M * C_{502} * R_{505}} * t$$

[0205] If the currents I_{503}, I_{506}, I_{507}, and I_{512} are ratiometric with an external current reference, the added voltage contribution seen at the output node due to the ramp generator block is a fixed voltage offset that is dependant on the manufacturing tolerance of the on-chip resistors, plus a time dependent ramp voltage independent of manufacturing tolerance in the on-chip resistors. The voltage offset may be reduced to zero to avoid dependence on on-chip resistance and to set the ramp to zero at the start of each cycle by judicious scaling of R_{504}, R_{505}, and M. The current source **503** may be controlled to change the slope of ramp to include the emulated change in inductor current when the NMOS is off and the body diode is conducting.

[0206] FIG. **14** shows a greatly simplified depiction of the duty modulator block with relevant connections as used in FIG. **5**

[0207] It consists of a bandgap voltage generator and internal resistor derived current source, **601**, a voltage comparator, **602**, two resistors, **603** and **604**, and a mode control block, **605**.

[0208] Mode control block comprises logic for generating a HOLD signal when signal ILIM_ZC goes true and keeping the hold signal high until reset—when the summing node voltage reaches V_{ERROR} as described above with reference to FIGS. **10***a* and **10***b*. It also comprises logic for generating an EMULATE signal when ILIM_ZC goes true until NDIODE_OFF goes true.

[0209] Mode control block may also receive an external mode control signal for switching between FCCM mode or automatic DCM and pulse skipping. When operating in FCCM the signal ILIM_ZC may simply be kept continually false and the converter will only switch the NMOS off when the signal on the summing node reaches V_{ERROR}. It may be preferable however to effectively deactivate NDiode detect block **750** and zero cross detect block **850** when operating in FCCM.

[0210] Whilst the above embodiments have been described in relation to a DC-DC buck converter the embodiments of the invention can be applied generally to switching regulators. The switching regulator may be part of a power management apparatus such as a power management integrated circuit (i.e. a PMIC). The embodiments of the invention may be useful for power management of any sub-systems of any form of electrical device, whether having a single power source or multiple power sources and whether portable or not. The embodiments of the present invention are particularly applicable to portable devices however such as: mobile computing devices such as laptops, netbooks, PDAs and the like; mobile communication devices such as radio telephones, cellular telephone, mobile email devices and the like; personal media players such as MP3 or other audio players, personal radios, video players; portable video game consoles and devices; personal navigation devices such as satellite navigators and GPS receivers, whether in-vehicle or hand-held or any other portable or battery operated device.

[0211] It should be noted that the above-mentioned embodiments illustrate rather than limit the invention, and that those skilled in the art will be able to design many alternative embodiments without departing from the scope of the appended claims. The word "comprising" does not exclude the presence of elements or steps other than those listed in a claim, "a" or "an" does not exclude a plurality, and a single processor or other unit may fulfil the functions of several units recited in the claims. Any reference signs in the claims shall not be construed so as to limit their scope.

1. A DC-DC converter comprising:

an inductor operably connected between a first node and an output node;

a high side switch operably connected between a high side supply input node and the first node;

a low side switch operably connected between a low side supply input node and the first node;

switch control circuitry for controlling said switches, wherein the switch control circuitry comprises first monitoring circuitry for determining whether a first time period between the high side switch turning off and the inductor current reaching zero is less than a threshold time period,

wherein if the first time period in one duty cycle is less than the threshold time period the switch control circuitry inhibits the low side switch from turning on in a subsequent cycle.

2. A DC-DC converter as claimed in claim **1** wherein if the first time period in one duty cycle is less than the threshold time period the switch control circuitry inhibits the low side switch from turning on following the next time that the high side switch has been on.

3. A DC-DC converter as claimed in claim **1** wherein the switch control circuitry is operable, in a cycle where turn on of the low side switch is not inhibited, to turn the low side switch off before the high side switch is turned on.

4. A DC-DC converter as claimed in claim **3** wherein the switch control circuitry comprises second monitoring circuitry for determining the current flow in the inductor from the electrical properties of the low side switch when the low side switch is on and monitoring current flow in the inductor against a current threshold and wherein the switch control circuitry is arranged to turn the low side switch off when the current threshold is reached.

5. A DC-DC converter as claimed in claim **4** wherein said current threshold is set such that the low side switch is turned off before the current flow through in the inductor reverses direction.

6. A DC-DC converter as claimed in claim **4** wherein the current threshold level is varied in response to a variation of a target output voltage.

7. A DC-DC converter as claimed in claim **1** wherein the first monitoring circuitry comprises a voltage threshold monitoring circuit for comparing the voltage of the first node with a threshold voltage when the low side switch is turned off.

8. A DC-DC converter as claimed in claim **7** wherein the threshold voltage is zero volts.

9. A DC-DC converter as claimed in claim **1** wherein the first monitoring circuitry comprises a first timer circuit configured to determine whether the inductor current has reached zero at a time equal to the time threshold period after the high side switch turns off.

10. A DC-DC converter as claimed in claim **9** wherein the first timer circuit is arranged to receive a control signal synchronous with a high side switch control signal and wherein the first timer circuit comprises at least one delay element for delaying said control signal by a delay equal to the threshold time period.

11. A DC-DC converter as claimed in claim **1** wherein the time threshold period is configurable.

12. A DC-DC converter as claimed in claim **1** wherein the low side switch is an NMOS transistor.

13. A DC-DC converter as claimed in claim **1** wherein the DC-DC converter is operable in valley mode control.

14. A DC-DC converter as claimed in claim **13** wherein the converter comprises circuitry for keeping a current feedback loop active when the low side switch is off.

15. A DC-DC as claimed in claim **14** wherein the circuitry for keeping a current feedback loop active when the low side switch is off comprises emulation circuitry for emulating the change in inductor current when the low side switch is off

16. A DC-DC converter as claimed in claim **13** wherein the switch control circuitry comprises current sense circuitry for determining the current flow in the inductor from the electrical properties of the low side switch when the low side switch is on.

17. A DC-DC converter as claimed in claim **1** wherein the DC-DC converter is operable in peak mode control.

18. A DC-DC converter as claimed in claim **1** wherein the switch control circuitry is operable in a pulse skip mode to prevent turning the high side switch on in a cycle.

19. A DC-DC converter as claimed in claim **18**, operable in valley mode control, wherein the switch control circuitry prevents turning the high side switch on in a cycle if a second time period between turning the high side switch on and turning the high side switch off would be less than a second threshold time period.

20. A DC-DC converter as claimed in claim **19** wherein the switch control circuitry comprises a second timer circuit for producing an inhibit signal which inhibits turn on of the high side switch, for a time period substantially equal to the second time threshold period, before a clock edge which would cause the high side switch to turn off.

21. A DC-DC converter as claimed in claim **20** wherein the second timer circuit comprises an input for receiving an input clock signal, a delay for producing a delayed clock signal and circuitry for generating the inhibit signal based on the input clock signal.

22. A DC-DC converter as claimed in claim **1** wherein the cycle frequency of the converter is constant.

23. A DC-DC converter as claimed in claim **1** comprising a diode in parallel with the low side switch to allow forward current to flow if the low side switch is turned off before the high side switch is turned on but to prevent current reversal in the inductor.

24. A method of controlling a DC-DC converter comprising an inductor operably connected between a first node and an output node; a high side switch operably connected between a high side supply input node and the first node; and a low side switch operably connected between a low side supply input node and the first node; the method comprising:

monitoring whether a first time period between the high side switch turning off and the inductor current reaching zero is less than a threshold time period, and

where the first time period in one duty cycle is less than the threshold time period the switch control circuitry inhibiting the low side switch from turning on in a subsequent cycle.

25. (canceled)

26. A DC-DC converter comprising:

an inductor operably connected between a first node and an output node;

a high side switch operably connected between a high side supply input node and the first node;

a low side switch operably connected between a low side supply input node and the first node;

switch control circuitry arranged to monitor the time between the high side switch turning off and the voltage at the first node reaching zero and, if said monitored time is less than a first time threshold, preventing the low side switch from subsequently turning on the next time the high side switch is turned on.

27. A DC-DC converter comprising:

an inductor operably connected between a first node and an output node;

a high side switch operably connected between a high side supply input node and the first node;

a low side switch operably connected between a low side supply input node and the first node; and

switch control circuitry operable in a first mode where the low side switch is not turned on following the high side switch being turned off.

28. A DC-DC converter operable in a mode wherein a low side switch is not turned on following activation and deactivation of a high side supply switch.

29. A power management integrated circuit comprising a DC-DC converter as claimed in claim **1**.

30. An electronic device comprising a DC-DC converter as claimed in claim **1**.

31. An electronic device as claimed in claim **30** wherein the device is one of: a portable computing device; a laptop computer; a personal data assistant; a personal media player; an mp3 player; a portable television; a mobile communications device; a mobile telephone; a navigation aid; a GPS device; a game console.

32. (canceled)

* * * * *

US 20130229064A1

(19) **United States**
(12) **Patent Application Publication**
MOON et al.
(10) **Pub. No.: US 2013/0229064 A1**
(43) **Pub. Date: Sep. 5, 2013**

(54) **APPARATUS FOR HARVESTING LEAKAGE ENERGY**

(71) Applicant: **ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE**, (US)

(72) Inventors: **Jung Ick MOON**, Daejeon (KR); **In Kui CHO**, Daejeon (KR); **Seong-Min KIM**, Daejeon (KR); **Je Hoon YUN**, Daejeon (KR); **Woo Jin BYUN**, Daejeon (KR)

(73) Assignee: **ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE**, Daejeon (KR)

(21) Appl. No.: **13/773,873**

(22) Filed: **Feb. 22, 2013**

(30) **Foreign Application Priority Data**

Feb. 27, 2012 (KR) 10-2012-0019957

Publication Classification

(51) **Int. Cl.**
H04B 5/00 (2006.01)

(52) **U.S. Cl.**
CPC *H04B 5/0037* (2013.01)
USPC **307/104**

(57) **ABSTRACT**

Disclosed is an apparatus for harvesting leakage energy. The apparatus for harvesting leakage energy includes: an energy harvesting unit configured to harvest energy leaked while energy radiated from a transmitting unit of an energy transmitting apparatus is transmitted to a receiving unit of an energy receiving apparatus; and a harvesting circuit module configured to supply energy harvested by the energy harvesting unit to a load. In accordance with the embodiment of the present invention, energy efficiency of overall system can be increased by harvesting the leakage energy without affecting performance of a wireless power transmission system.

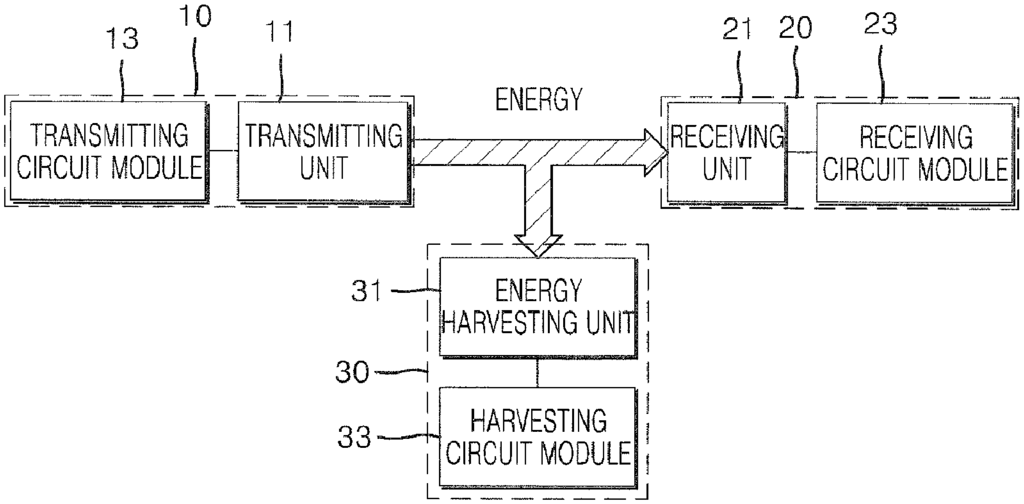
13
10
11
TRANSMITTING CIRCUIT MODULE
TRANSMITTING UNIT
ENERGY
21
20
23
RECEIVING UNIT
RECEIVING CIRCUIT MODULE
31
ENERGY HARVESTING UNIT
30
33
HARVESTING CIRCUIT MODULE

FIG. 1

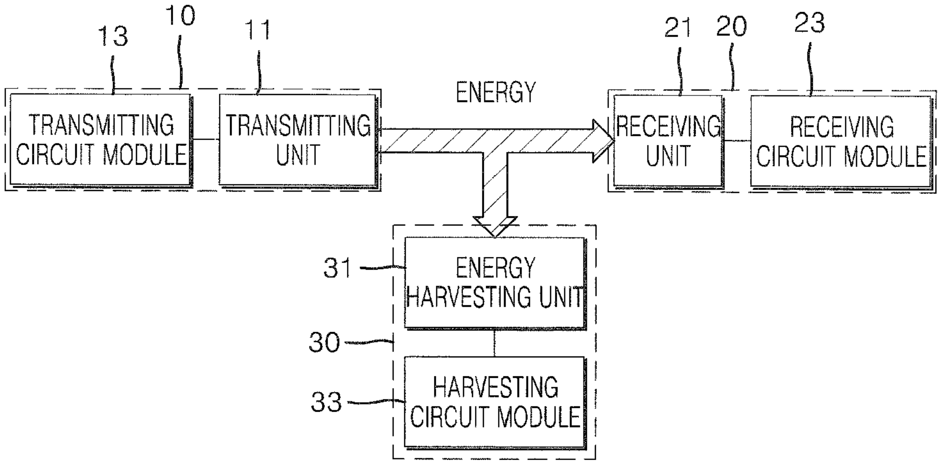
13
10
11
TRANSMITTING CIRCUIT MODULE
TRANSMITTING UNIT
ENERGY
21
20
23
RECEIVING UNIT
RECEIVING CIRCUIT MODULE
31
ENERGY HARVESTING UNIT
30
33
HARVESTING CIRCUIT MODULE

FIG. 2

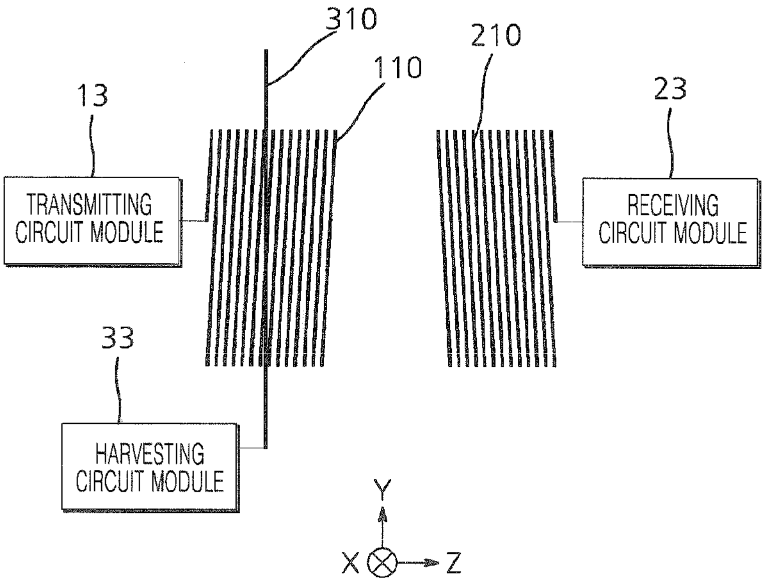
310
110
210
13
23
TRANSMITTING CIRCUIT MODULE
RECEIVING CIRCUIT MODULE
33
HARVESTING CIRCUIT MODULE
Y
X
Z

FIG. 3

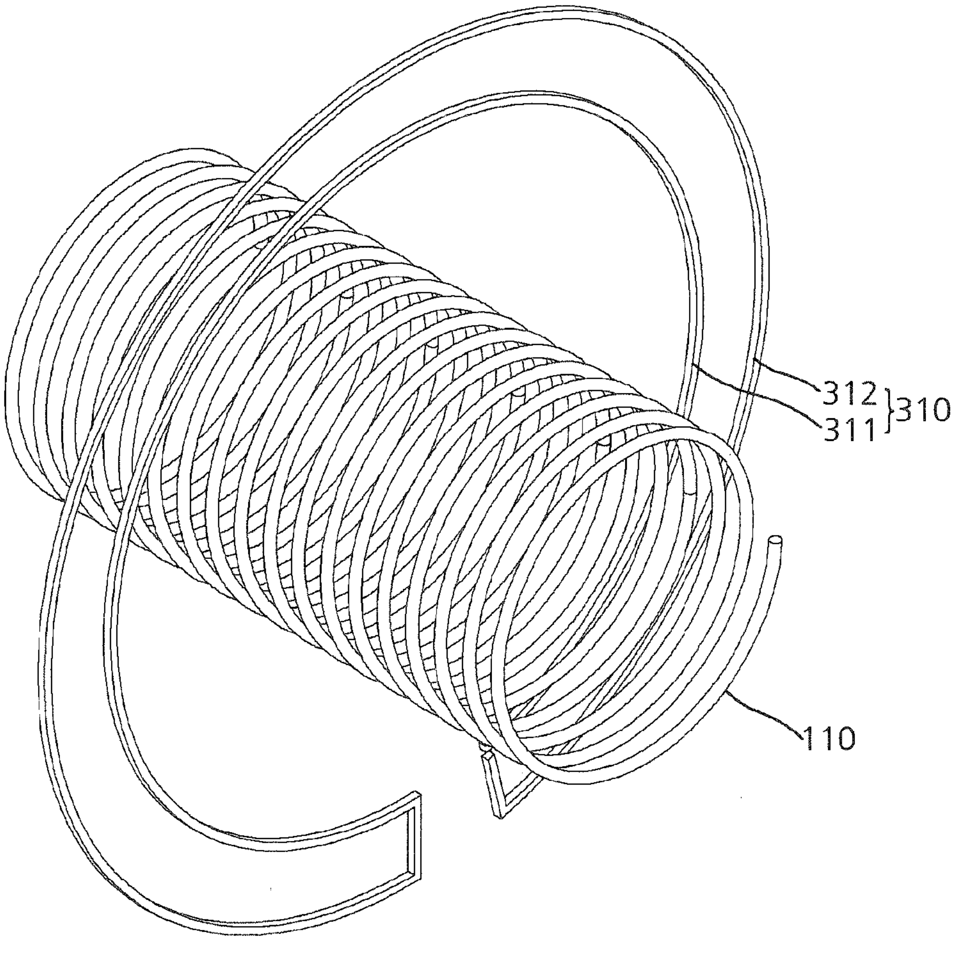
312
311
310
110

FIG. 4

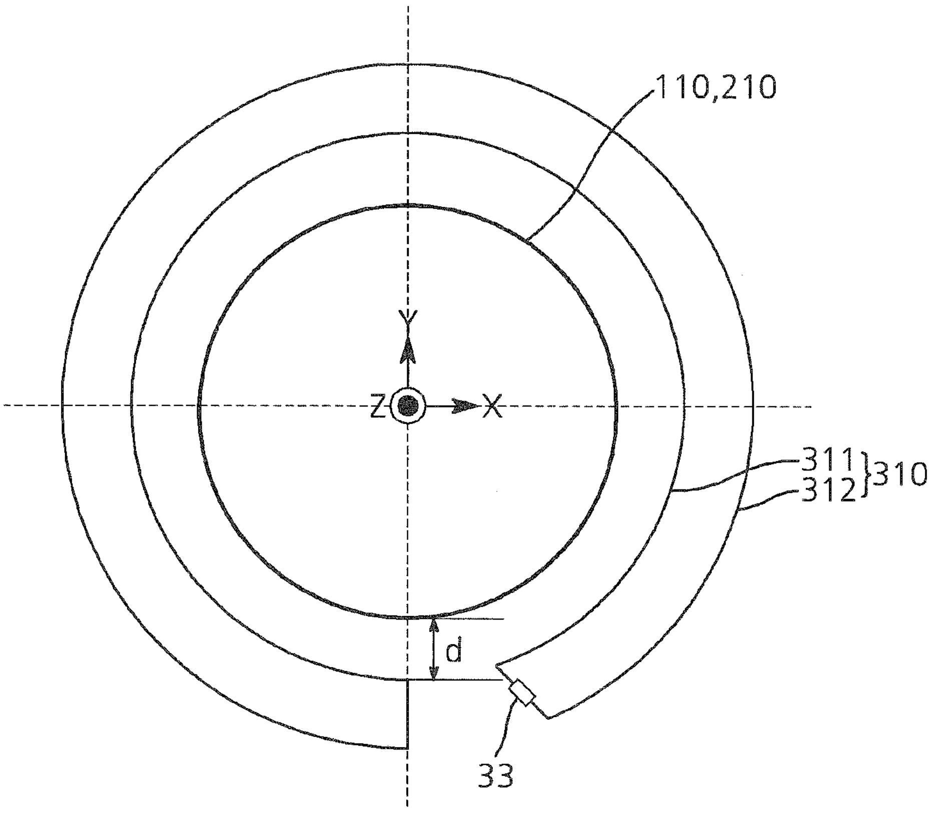
110,210
Y
Z
X
311
312
310
d
33

FIG. 5

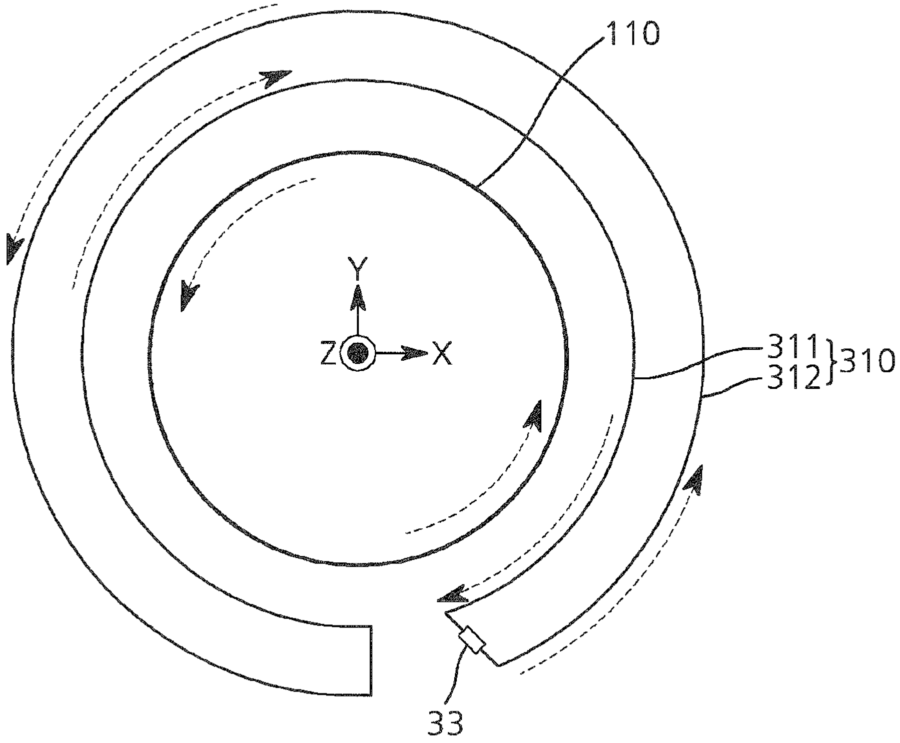
110
Y
Z
X
311
312
310
33

FIG. 6

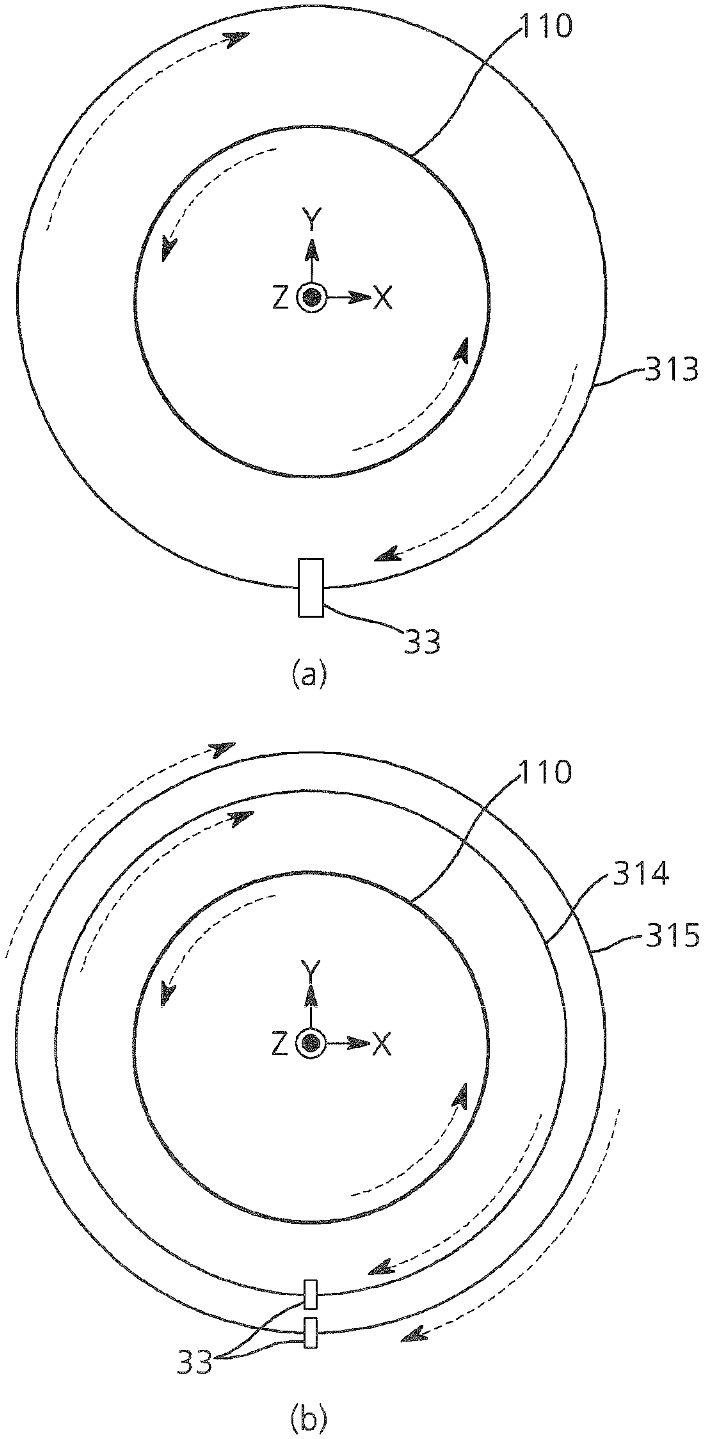
110
Y
Z
X
313
33
(a)
110
314
315
Y
Z
X
33
(b)

FIG. 7

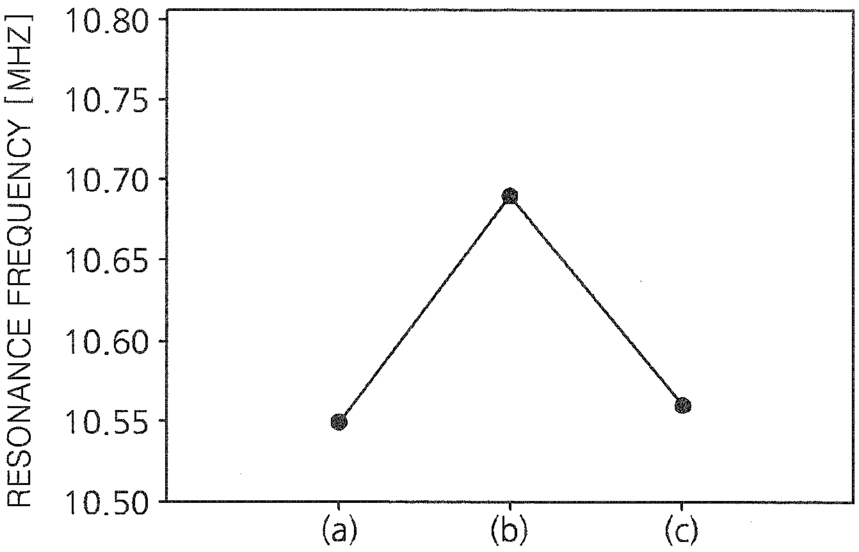
RESONANCE FREQUENCY [MHZ]
10.80
10.75
10.70
10.65
10.60
10.55
10.50
(a)
(b)
(c)

FIG. 8

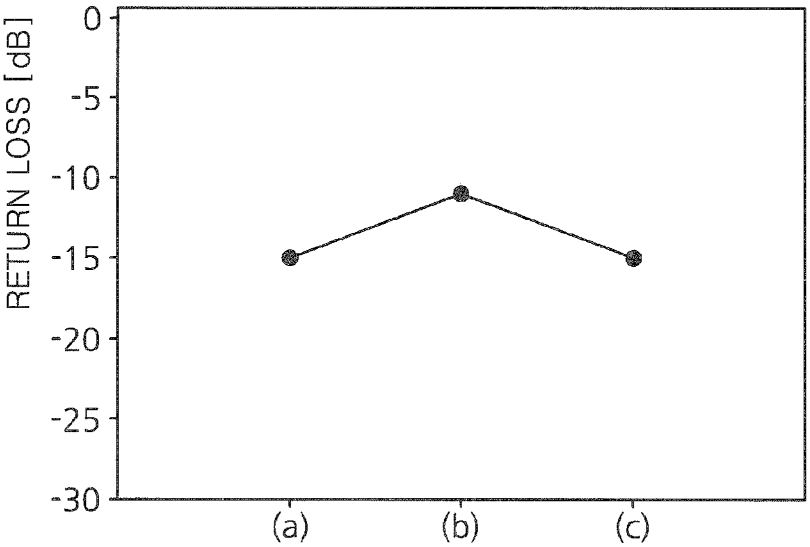
RETURN LOSS [dB]
0
-5
-10
-15
-20
-25
-30
(a)
(b)
(c)

FIG. 9

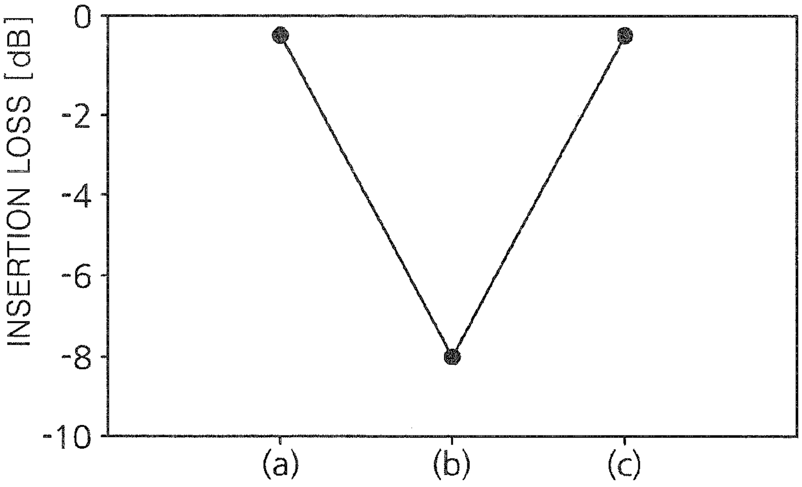

FIG. 10

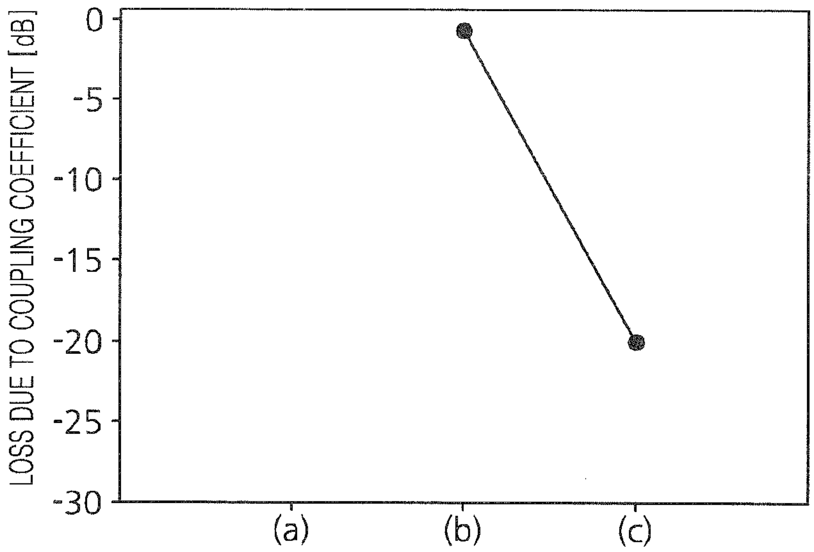

APPARATUS FOR HARVESTING LEAKAGE ENERGY

CROSS-REFERENCES TO RELATED APPLICATIONS

[0001] The present application claims priority under 35 U.S.C 119(a) to Korean Application No. 10-2012-0019957, filed on Feb. 27, 2012, in the Korean Intellectual Property Office, which is incorporated herein by reference in its entirety set forth in full.

BACKGROUND

[0002] Exemplary embodiments of the present invention relate to an apparatus for harvesting leakage energy, and more particularly, to an apparatus for harvesting leakage energy in a wireless power transmission system using a magnetic resonance or electromagnetic induction phenomenon.

[0003] Generally, devices driven using electricity as an energy source are supplied with power through a power line. However, there are problems in that a use of the power line hinders mobility of the devices or has a limitation in a position.

[0004] Recently, with a rapid use of a mobile device in addition to smart phones, a demand for wireless charging technologies has been increased. Therefore, research into a wireless power transmission system capable of transferring power wirelessly has been actively conducted.

[0005] Technologies of transmitting power wirelessly convert electric energy into an electromagnetic wave or a light wave that can be transmitted wirelessly and transmit power using electric field, magnetic field, or electromagnetic field and may be largely classified into an electromagnetic induction type, a magnetic resonance type, a microwave type, and the like.

[0006] The electromagnetic induction type is a type of generating induction current at two coils and transmitting energy, the magnetic resonance type is a type of using an evanescent wave coupling phenomenon of moving an electromagnetic wave from one medium to another medium through near field when two media are resonated at the same frequency, and the microwave type is a type of transmitting energy by radiating a microwave signal to the air through an antenna.

[0007] A physical quantity leaked while the physical quantity radiated from a resonator or a radiator is transmitted to a receiver by passing through a space is essentially present in the wireless power transmission system.

[0008] The energy leaked to the space acts as loss in the wireless power transmission system, thereby degrading overall transmission efficiency and energy efficiency.

[0009] As the related art, KR Patent Laid-Open No. 2011-0117732 (Oct. 27, 2011: Wireless Energy Transmission)

[0010] The above-mentioned technical configuration is a background art for helping understanding of the present invention and does not mean related arts well known in a technical field to which the present invention pertains.

SUMMARY

[0011] An embodiment of the present invention is directed to an apparatus for harvesting leakage energy capable of increasing energy efficiency by harvesting leakage energy without affecting a wireless power transmission system.

[0012] An embodiment of the present invention relates to an apparatus for harvesting leakage energy, including: an energy harvesting unit configured to harvest energy leaked while energy radiated from a transmitting unit of an energy transmitting apparatus is transmitted to a receiving unit of an energy receiving apparatus; and a harvesting circuit module configured to supply energy harvested by the energy harvesting unit to a load.

[0013] The energy harvesting unit may include: an inner loop coil configured to be disposed at an outer side of the transmitting unit or the receiving unit and having an opened side; and an outer loop coil configured to be disposed at an outer side of the inner loop coil and having an opened side, wherein the opened sides of the inner loop coil and the outer loop coil are connected to each other to form a single closed loop.

[0014] The energy harvesting unit may be disposed at the transmitting unit.

[0015] The energy harvesting unit may be disposed at a predetermined distance from the transmitting unit or the receiving unit.

[0016] The transmitting unit and the receiving unit may each be a transmitting coil and a receiving coil.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] The above and other aspects, features and other advantages will be more clearly understood from the following detailed description taken in conjunction with the accompanying drawings, in which:

[0018] FIG. **1** is a block diagram of a wireless power transmission system including an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention;

[0019] FIG. **2** is a side view for describing an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention;

[0020] FIG. **3** is a perspective view for describing an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention;

[0021] FIG. **4** is a front view for describing an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention;

[0022] FIG. **5** is a diagram for describing an energy harvesting principle of an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention;

[0023] FIGS. **6**A and **6**B are diagrams for describing an effect of an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention;

[0024] FIG. **7** is a graph for describing a change in a resonance frequency of a wireless power transmission system when using an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention;

[0025] FIG. **8** is a graph for describing a change in a return loss of a wireless power transmission system when using an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention;

[0026] FIG. **9** is a graph for describing a change in an insertion loss of a wireless power transmission system when using an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention;

[0027] FIG. **10** is a graph for describing a change in loss according to a coupling coefficient between an energy transmitting unit and an energy harvesting unit in a wireless power transmission system when using an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention.

DESCRIPTION OF SPECIFIC EMBODIMENTS

[0028] Hereinafter, an apparatus for harvesting leakage energy in accordance with embodiments of the present invention will be described with reference to the accompanying drawings. During the process, a thickness of lines, a size of components, or the like, illustrated in the drawings may be exaggeratedly illustrated for clearness and convenience of explanation. Further, the following terminologies are defined in consideration of the functions in the present invention and may be construed in different ways by intention or practice of users and operators. Therefore, the definitions of terms used in the present description should be construed based on the contents throughout the specification.

[0029] FIG. **1** is a block diagram of a wireless power transmission system including an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention and FIG. **2** is a side view for describing an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention.

[0030] In addition, FIG. **3** is a perspective view for describing an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention and FIG. **4** is a front view for describing an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention.

[0031] As illustrated in FIG. **1**, a wireless power transmission system in accordance with an embodiment of the present invention includes an energy transmitting apparatus **10**, an energy receiving apparatus **20**, and an apparatus for harvesting leakage energy **30**.

[0032] The energy transmitting unit **10** includes a transmitting circuit module **13** generates an RF signal to be transmitted and amplifies and outputs the generated RF signal and a transmitting unit **11** that receives the RF signal from the transmitting circuit module **13** to radiate energy to the energy receiving apparatus **20**.

[0033] The energy receiving apparatus **20** includes a receiving unit **21** that receives energy radiated from the transmitting unit **11** and transmits the received energy to a receiving circuit module **23** and the receiving circuit module **23** that receives energy from the receiver **21** and supplies the received energy to a load.

[0034] In this configuration, the transmitting unit **11** and the receiving unit **21** may be a transmitting coil **110** and a receiving coil **210** that may each transmit or receive energy by electromagnetic induction and a resonator (not illustrated) that may transmit or receive energy using a magnetic resonance type.

[0035] Meanwhile, as illustrated in FIG. **1**, the apparatus for harvesting leakage energy **30** in accordance with the embodiment of the present invention harvest energy leaked during the transmission of the energy radiated from the energy transmitting apparatus **10** to the energy receiving apparatus **20** and includes an energy harvesting unit **31** and a harvesting circuit module **33**.

[0036] The energy harvesting unit **31** may be disposed at a predetermined distance from an outer side of the transmitting unit **11** to harvest energy leaked from the transmitting unit **11**.

[0037] In this case, the energy harvesting unit **31** may be disposed at a predetermined distance from an outer side of the receiving unit **31** rather than the transmitting unit **11** to harvest energy leaked from the receiving unit **21** and may be disposed at various positions but may be preferably disposed at the transmitting unit **11** side to which energy is radiated in terms of the energy harvesting efficiency.

[0038] As illustrated in FIGS. **2** and **3**, the energy harvesting unit **31** may be a harvesting coil **310** in a loop form that harvests the energy leaked from the transmitting coil **110** or the receiving coil **210** and as illustrated in FIG. **4**, the harvesting coil **310** may be configured to include an inner loop coil **311** and an outer loop coil **312**.

[0039] In this configuration, the harvesting coil **310** including the inner loop coil **311** and the outer loop coil **312** may be configured of a conductor in which current may flow.

[0040] Referring to FIG. **4**, the inner loop coil **311** may be formed to have a diameter larger by a predetermined distance d than the transmitting coil **110** or the receiving coil **210** and the outer loop coil **312** may be formed to have a diameter larger than the inner loop coil **311**.

[0041] That is, the inner loop coil **311** and the outer loop coil **312** are spaced apart from the transmitting coil **110** or the receiving coil **210** by the predetermined distance d and thus, may be disposed in a concentric shape.

[0042] As illustrated in FIG. **4**, the inner loop coil **311** and the outer loop coil **312** each have opened sides and the opened sides of the inner loop coil **311** and the outer loop coil **312** may be connected to each other to form a single closed loop and the detailed description thereof will be described below.

[0043] The harvesting circuit module **33** receives the energy harvested in the energy harvesting unit **31** and supplies the received energy to loads such as elements, devices, and the like.

[0044] Since the amount of energy harvested in the energy harvesting unit **31** is small, the harvesting circuit module **33** preferably supplies energy to elements or devices having low power consumption, such as a light emitting diode (LED) lamp, and the like. By this operation, the energy efficiency of the overall system can be increased.

[0045] FIG. **5** is a diagram for describing an energy harvesting principle of an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention and FIG. **6** is a diagram for describing an effect of an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention.

[0046] Meanwhile, as described above, the inner loop coil **311** and the outer loop coil **312** configuring the harvesting coil **310** may each have the opened sides. As illustrated in FIG. **4**, each opened side of the inner loop coil **311** and the outer loop coil **312** may be connected to each other to form the single closed loop.

[0047] In this case, the energy harvesting quantity may be in proportion to an area of a closed curved surface formed by the inner loop coil **311** and the outer loop coil **312** to harvest energy as desired by appropriately adjusting the diameter of the inner loop coil **311** and the outer loop coil **312**.

[0048] As described above, the opened sides of the inner loop coil **311** and the outer loop coil **312** are connected to each other to form the single closed loop, which can harvest leakage energy without affecting the wireless power transmission system including the energy transmitting apparatus **10** and the energy receiving apparatus **20**.

[0049] FIG. **5** illustrates a direction of current flowing in each coil when the opened sides of the inner loop coil **311** and the outer loop coil **312** are connected to each other to form the single closed loop.

[0050] Referring to FIG. **5**, provided that current flows in the transmitting coil **110** counterclockwise, current flows in the inner loop coil **311** clockwise that is a direction opposite to the transmitting coil **110** according to a Len's law.

[0051] In this case, the inner loop coil **311** and the outer loop coil **312** are connected to each other in the single closed loop and therefore, current flows in the outer loop coil **312** counterclockwise that is a direction opposite to the inner loop coil **311**.

[0052] The current flowing in a direction opposite to the transmitting coil **110** prevents power from being smoothly transmitted from the transmitting coil **110**. However, in the present invention, current flows in the outer loop coil **312** in the same direction as the transmitting coil **110** and therefore, the influence due to the inner loop coil **311** may be offset, which little affect the power transmission of the transmitting coil **110**.

[0053] FIG. **6** illustrates a direction of current flowing in each coil when harvesting energy using a single loop coil **313** and harvesting energy using two loop coils **314** and **315** that are not connected to each other.

[0054] As illustrated in FIG. **6**A, when using the single loop coil **313**, current flows in the single loop coil **313** in a direction opposite to current flowing in the transmitting coil **110** to prevent the power transmission of the transmitting coil **100**.

[0055] In addition, as illustrated in FIG. **6**B, when using two loop coils **314** and **315** that are not connected to each other, the influence of energy generated from the transmitting coil **110** is dominated and therefore, current flows in both of the two loop coils **314** and **315** that are not connected to each other in a direction opposite to the transmitting coil **110** to prevent the power transmission of the transmitting coil **110**.

[0056] As described above, according to the apparatus for harvesting leakage energy **30** of the present invention, the opened sides of the inner loop coil **311** and the outer loop coil **312** are connected to each other to form the single closed loop, which can harvest the leakage energy without affecting the power transmission of the transmission coil **110**. As a result, the energy efficiency can be increased.

[0057] The effect of the present invention will be described in more detail with reference to FIGS. **7** to **10**.

[0058] FIG. **7** is a graph for describing a change in a resonance frequency of a wireless power transmission system when using an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention and FIG. **8** is a graph for describing a change in a return loss of a wireless power transmission system when using an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention.

[0059] FIG. **9** is a graph for describing a change in an insertion loss of a wireless power transmission system when using an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention and FIG. **10** is a graph for describing a change in loss according to a coupling coefficient between an energy transmitting unit and an energy harvesting unit in a wireless power transmission system when using an apparatus for harvesting leakage energy in accordance with an embodiment of the present invention.

[0060] FIGS. **7** to **10** commonly illustrate calculated simulation values in three cases.

[0061] In each figure of FIGS. **7** to **10**, (a) illustrates the case in which there is no apparatus for harvesting leakage energy **30**, (b) illustrates the case in which the single loop coil **313** is used as illustrated in FIG. **6**A, and (c) illustrates the case in which the apparatus for harvesting leakage energy **30** in accordance with the present invention is used.

[0062] Simulation execution conditions are illustrated in the following Table **1**.

TABLE 1

Common Condition	Transmitting Coil (110)	Diameter 60 cm, 5.5 turns
	Receiving Coil (210)	Diameter 60 cm, 5.5 turns
	Distance Between Transmitting Coil (110) and Receiving Coil (210)	90 cm (Center-Center)
(b)	Harvesting Coil (310)	Diameter 129 cm, 1 turn
	Distance Between Transmitting Coil (110) and Harvesting Coil (310)	34.5 cm (Outermost-Outermost, Based On Vertical Axis)
(c)	Harvesting Coil (310)	Inner Diameter 109 cm, Outer Diameter 129 cm 1 turn
	Distance Between Transmitting Coil (110) and Harvesting Coil (310)	24.5 cm (Outermost-Outermost, Based On Vertical Axis)

[0063] Referring to FIG. **7**, in the case of (b) using the single loop coil **313**, the resonance is generated at 10.70 [MHz] and thus, the resonance frequency is larger changed than (a) in which the resonance is generated at 10.55 [MHz]. However, in the case of (c) in accordance with the embodiment of the present invention, it can be appreciated that there is little difference from (a) in terms of the resonance frequency.

[0064] Similarly, referring to FIGS. **8** and **9**, in the case of (c) in accordance with the embodiment of the present invention unlike (b) using the single loop coil, it can be appreciated that there is little difference from (a) in terms of the return loss and the insertion loss.

[0065] In addition, FIG. **10** illustrates the loss due to the coupling coefficient between the transmitting coil **110** and the harvesting coil **310**. As the loss approximates '0', the coupling coefficient between the transmitting coil **110** and the harvesting coil **310** is increased and thus, more energy is transmitted to the harvesting coil **310** than the receiving coil **210** to be originally transmitted.

[0066] However, as illustrated in FIG. **10**, in the (c) of the embodiment of the present invention, the loss due to the coupling coefficient is about −20 [dB], which corresponds to a value of about 1/100. Therefore, it can be appreciated that the power transmission of the transmitting coil **110** is little affected.

[0067] When the power transmittal is maximal, in the case of (b) using the single loop coil **313**, strong coupling is generated between magnetic field and the loop coil and as a result, the magnetic distribution near the transmitting coil **110** is distorted and energy is not originally transmitted to the receiving coil **210**.

[0068] On the other hand, in the case of (c) in accordance with the embodiment of the present invention, the approximately same magnetic distribution as (a) that does not have the apparatus for harvesting leakage energy **30** is maintained as it is and as a result, the apparatus for harvesting leakage energy **30** in accordance with the embodiment of the present invention little affects the wireless transmission system.

[0069] As described above, according to the apparatus for harvesting leakage energy **30** of the present invention, the energy leaked while the energy radiated from the energy transmitting apparatus **10** of the wireless power transmission system is transmitted to the energy receiving apparatus **20** can be harvested through the harvesting coil **310** in a loop form.

[0070] In addition, the present invention forms the single closed loop by connecting the inner loop coil **311** included in the apparatus for harvesting leakage energy **30** with the opened side of the outer loop coil **312**, thereby harvesting the leakage energy without affecting the performance of the wireless power transmission system.

[0071] In addition, in accordance with the embodiments of the present invention, it is possible to increase the energy efficiency of the overall system by driving the elements or the devices having the low power consumption using the harvested energy.

[0072] Meanwhile, in the embodiment of the present invention, the case in which the harvesting coil **310** harvesting energy is disposed at the transmitting coil **110** side is mainly described, but the position at which the harvesting coil **310** is disposed is not limited thereto and therefore, the harvesting coil **310** can be disposed at the receiving coil **210** side.

[0073] In accordance with the embodiments of the present invention, it is possible to harvest energy leaked while the energy radiated from the energy transmission apparatus of the wireless power transmission system is transmitted to the energy receiving apparatus through the harvesting coil in the loop form.

[0074] Further, in accordance with the embodiments of the present invention, it is possible to harvest the leakage energy without affecting the performance of the wireless power transmission system by forming the single closed loop by connecting the inner loop coil included in the apparatus for harvesting leakage energy with the opened side of the outer loop coil.

[0075] In addition, in accordance with the embodiments of the present invention, it is possible to increase the energy efficiency of the overall system by driving the elements or the devices having the low power consumption using the harvested energy.

[0076] Although the embodiments of the present invention have been described in detail, they are only examples. It will be appreciated by those skilled in the art that various modifications and equivalent other embodiments are possible from the present invention. Accordingly, the actual technical protection scope of the present invention must be determined by the spirit of the appended claims.

What is claimed is:

1. An apparatus for harvesting leakage energy, comprising:

an energy harvesting unit configured to harvest energy leaked while energy radiated from a transmitting unit of an energy transmitting apparatus is transmitted to a receiving unit of an energy receiving apparatus; and

a harvesting circuit module configured to supply energy harvested by the energy harvesting unit to a load.

2. The apparatus of claim **1**, wherein the energy harvesting unit includes:

an inner loop coil configured to be disposed at an outer side of the transmitting unit or the receiving unit and to have an opened side; and

an outer loop coil configured to be disposed at an outer side of the inner loop coil and to have an opened side,

the opened sides of the inner loop coil and the outer loop coil being connected to each other to form a single closed loop.

3. The apparatus of claim **1**, wherein the energy harvesting unit is disposed at the transmitting unit.

4. The apparatus of claim **1**, wherein the energy harvesting unit is disposed at a predetermined distance from the transmitting unit or the receiving unit.

5. The apparatus of claim **1**, wherein the transmitting unit and the receiving unit each are a transmitting coil and a receiving coil.

* * * * *

US 20150214841A1

(19) **United States**

(12) **Patent Application Publication**

Ramorini et al.

(10) **Pub. No.: US 2015/0214841 A1**

(43) **Pub. Date: Jul. 30, 2015**

(54) **DC-DC CONVERTER WITH IMPROVED ENERGY MANAGEMENT, METHOD FOR OPERATING THE DC-DC CONVERTER, ENVIRONMENTAL ENERGY HARVESTING SYSTEM USING THE DC-DC CONVERTER, AND APPARATUS USING THE ENERGY HARVESTING SYSTEM**

(71) Applicant: **STMicroelectronics S.r.I**, Agrate Brianza (IT)

(72) Inventors: **Stefano Ramorini**, Arluno (IT); **Alessandro Gasparini**, Cusano Milanino (IT); **Alberto Cattani**, Cislago (IT)

(73) Assignee: **STMicroelectronics S.r.I**, Agrate Brianza (IT)

(21) Appl. No.: **14/603,597**

(22) Filed: **Jan. 23, 2015**

(30) **Foreign Application Priority Data**

Jan. 28, 2014 (IT) TO2014A000059

Publication Classification

(51) **Int. Cl.**
H02M 3/158 (2006.01)

(52) **U.S. Cl.**
CPC *H02M 3/158* (2013.01)

(57) **ABSTRACT**

A converter includes an inductor configured to receive an input signal and output configured to supply an electrical load with an output signal. The converter operates to charge the inductor until a maximum pre-set current value is reached during a first operating condition in which the electrical load is not supplied. Next, the converter actively supplies the electrical load by partially discharging the inductor during a first time interval of a second operating condition. Then, the converter passively supplies the electrical load by the residual charge of the inductor during a second time interval, subsequent to the first time interval, of the second operating condition, by discharging the inductor completely.

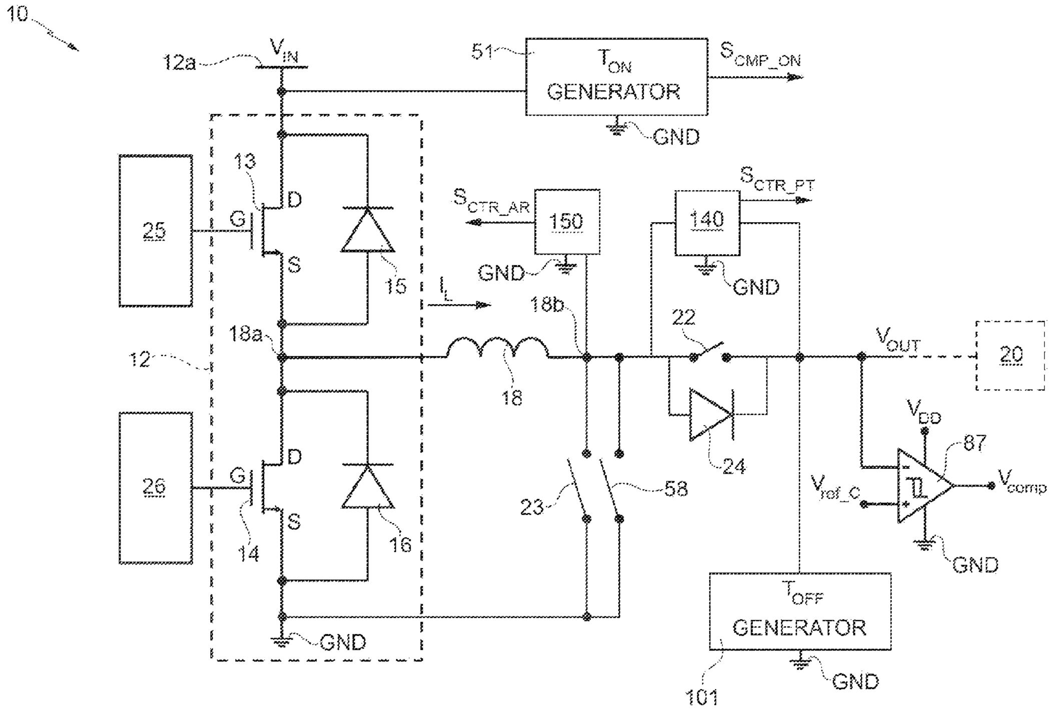
10
12a
V_{IN}
51
T_{ON} GENERATOR
S_{CMP_ON}
GND
13
D
G
S
25
15
S_{CTR_AR}
150
GND
140
S_{CTR_PT}
GND
I_L
18b
18a
22
12
18
V_{OUT}
20
V_{DD}
87
24
26
D
G
S
16
14
23
58
V_{ref_C}
V_{comp}
GND
T_{OFF} GENERATOR
GND
101
GND

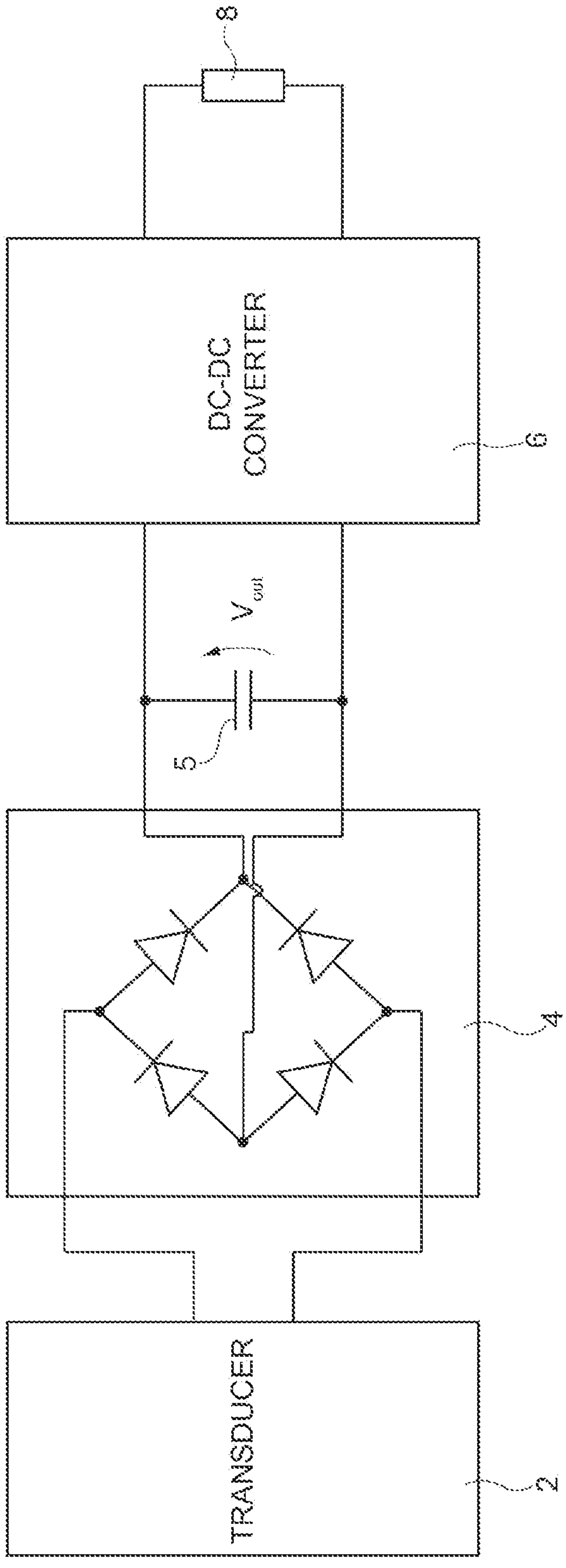
8
DC-DC CONVERTER
6
V_{out}
5
4
TRANSDUCER
2

FIG. 1

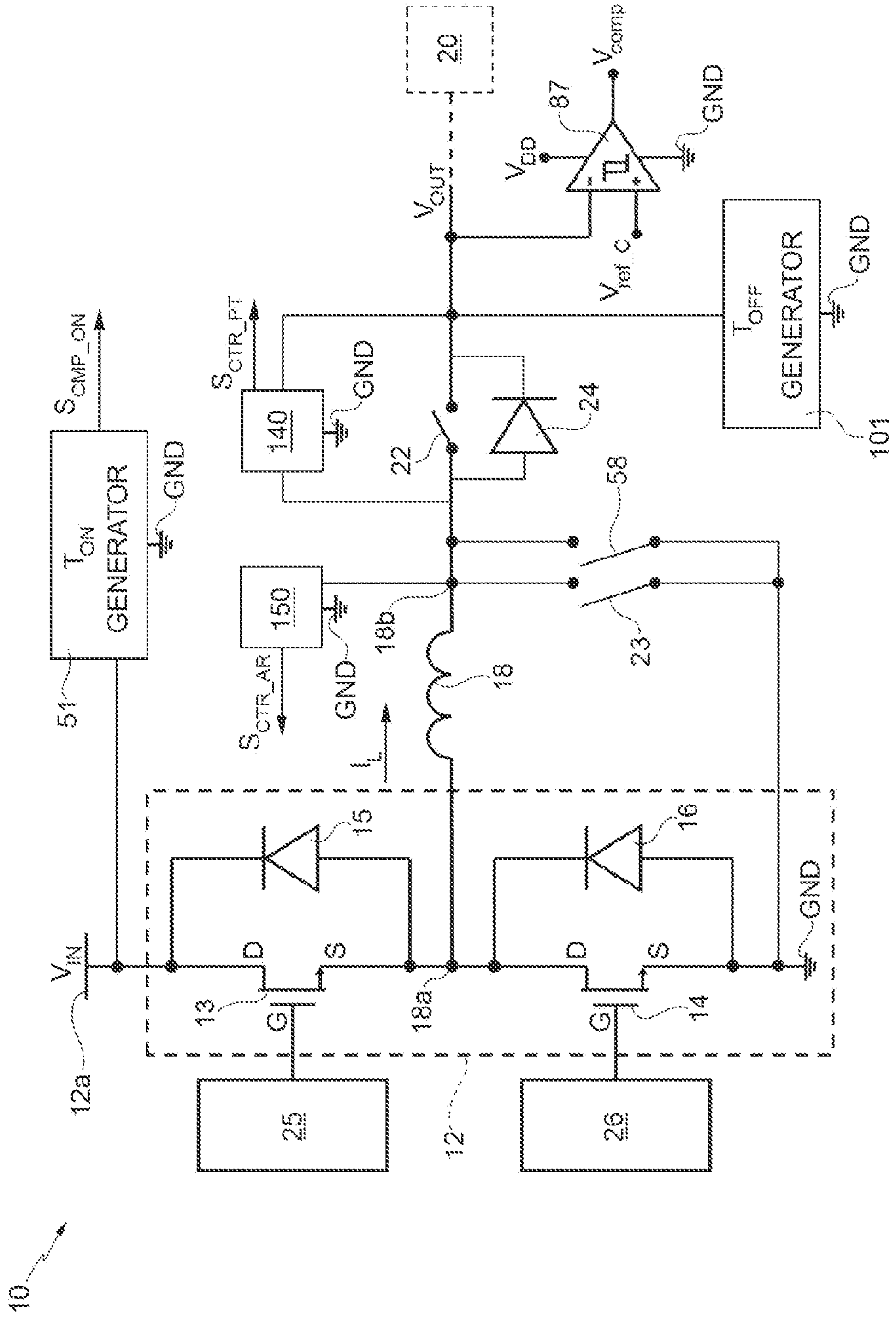
10
12a
V_{IN}
51
T_{ON} GENERATOR
GND
S_{CMP_ON}
25
26
12
13
14
15
16
G
D
S
18a
18
18b
I_L
150
S_{CTR_AR}
140
S_{CTR_PT}
22
24
23
58
V_{OUT}
20
V_{DD}
87
V_{comp}
V_{ref_c}
T_{OFF} GENERATOR
101

FIG. 2

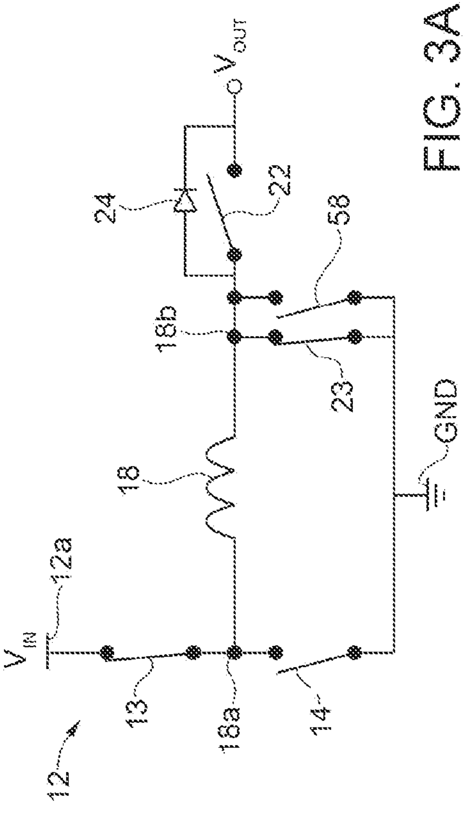
12
13
18a
14
V_IN
12a
18
GND
23
18b
58
24
22
V_OUT

FIG. 3A

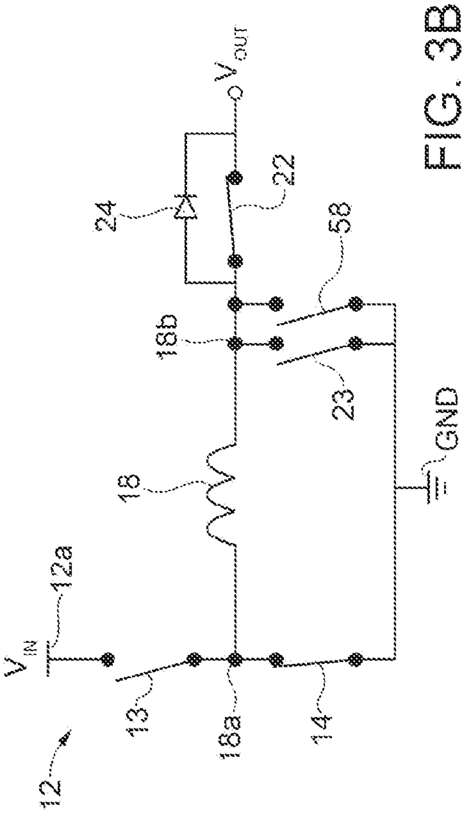
12
13
18a
14
V_IN
12a
18
GND
23
18b
58
24
22
V_OUT

FIG. 3B

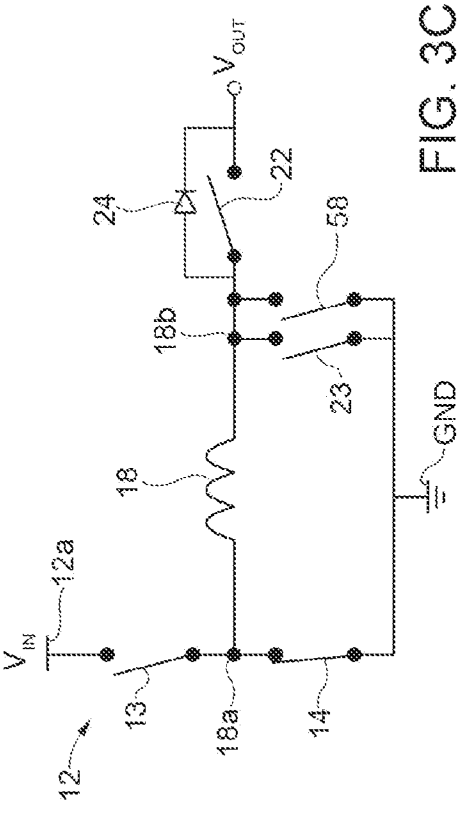
12
13
18a
14
V_IN
12a
18
GND
23
18b
58
24
22
V_OUT

FIG. 3C

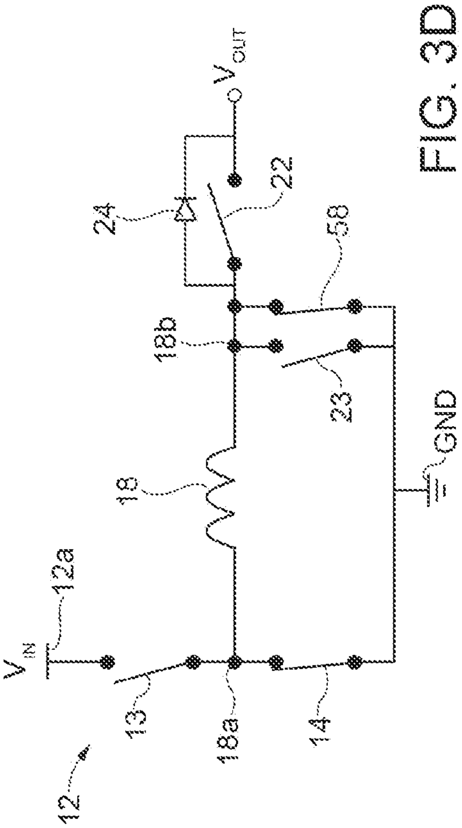
12
13
18a
14
V_IN
12a
18
GND
23
18b
58
24
22
V_OUT

FIG. 3D

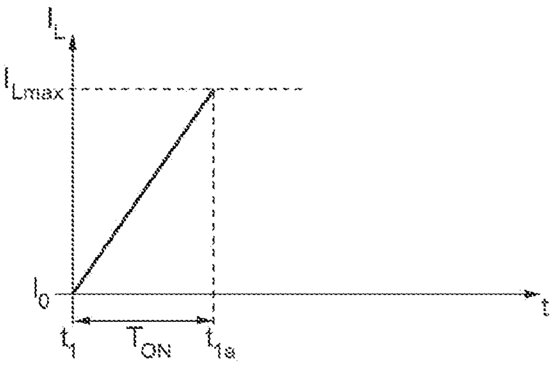

FIG. 4A

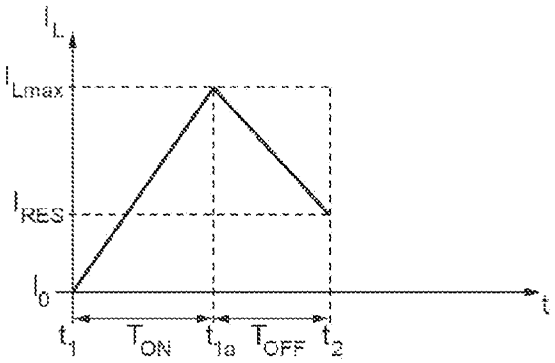

FIG. 4B

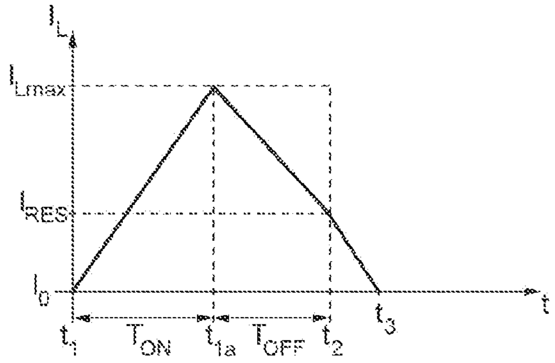

FIG. 4C

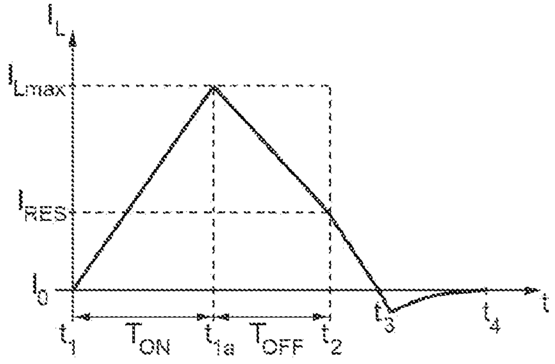

FIG. 4D

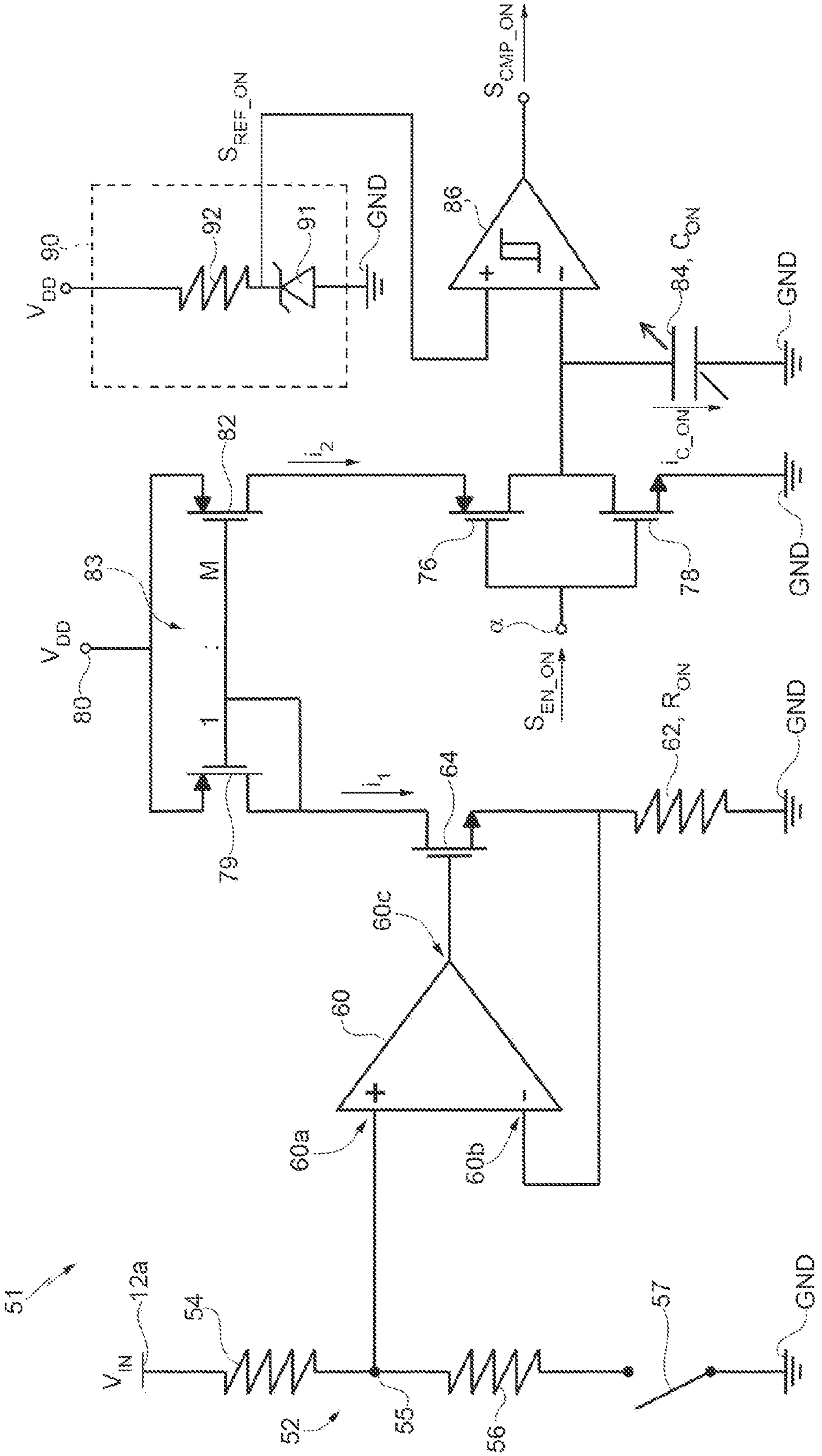
51
52
V_{IN}
12a
54
55
56
57
GND
60
+
-
60a
60b
60c
64
i_1
62, R_{ON}
GND
79
80
V_{DD}
83
1
:
M
82
i_2
76
78
GND
α
S_{EN_ON}
i_{C_ON}
84, C_{ON}
GND
86
+
-
S_{CMP_ON}
S_{REF_ON}
V_{DD}
90
92
91
GND

FIG. 5

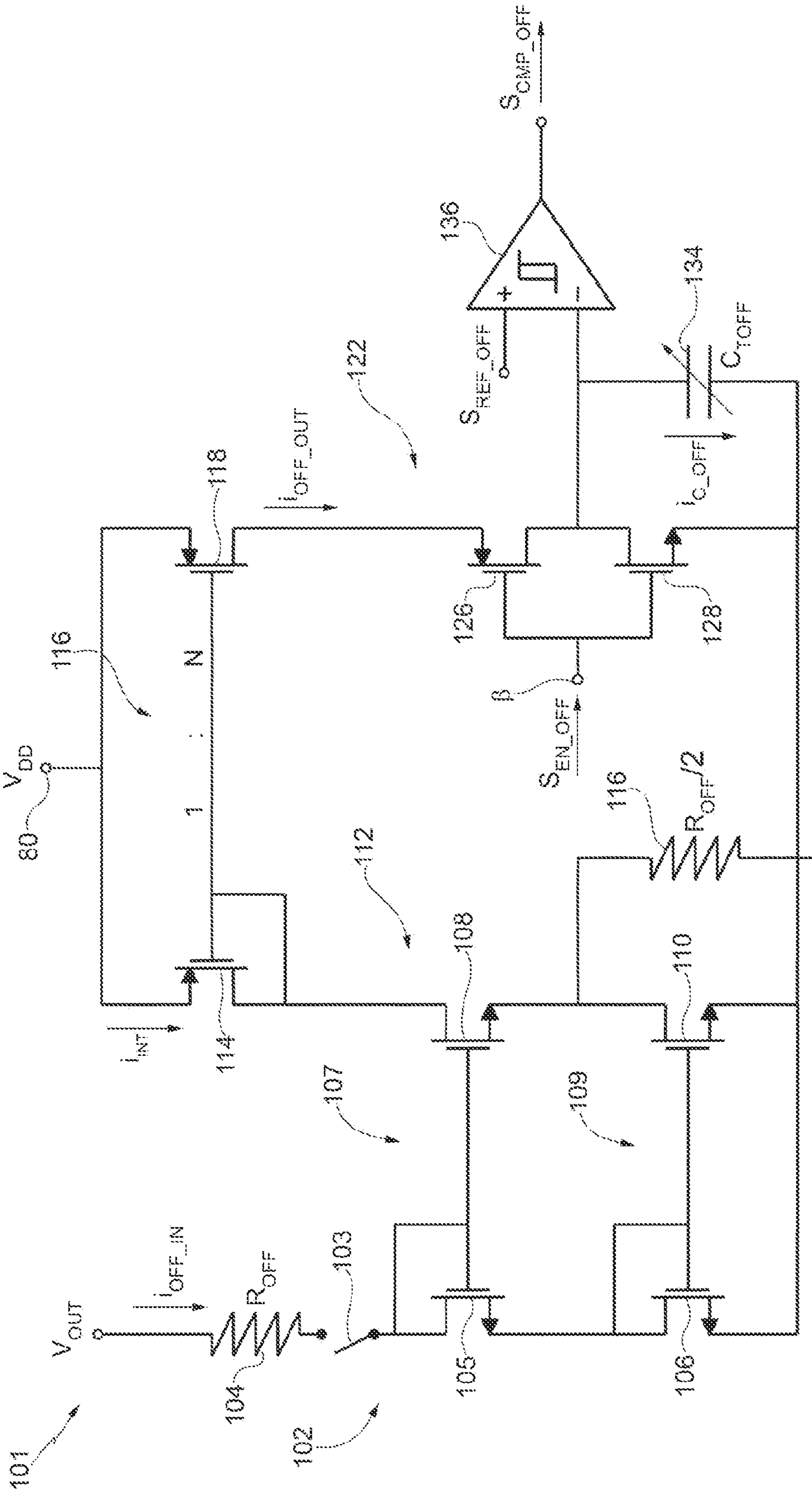
101
V_{OUT}
i_{OFF_IN}
R_{OFF}
104
103
102
105
106
107
109
108
110
112
114
i_{INT}
80
V_{DD}
116
1
:
N
118
i_{OFF_OUT}
122
126
128
S_{EN_OFF}
β
116
$R_{OFF}/2$
GND
S_{REF_OFF}
+
−
136
S_{CMP_OFF}
134
C_{TOFF}
i_{C_OFF}

FIG. 6

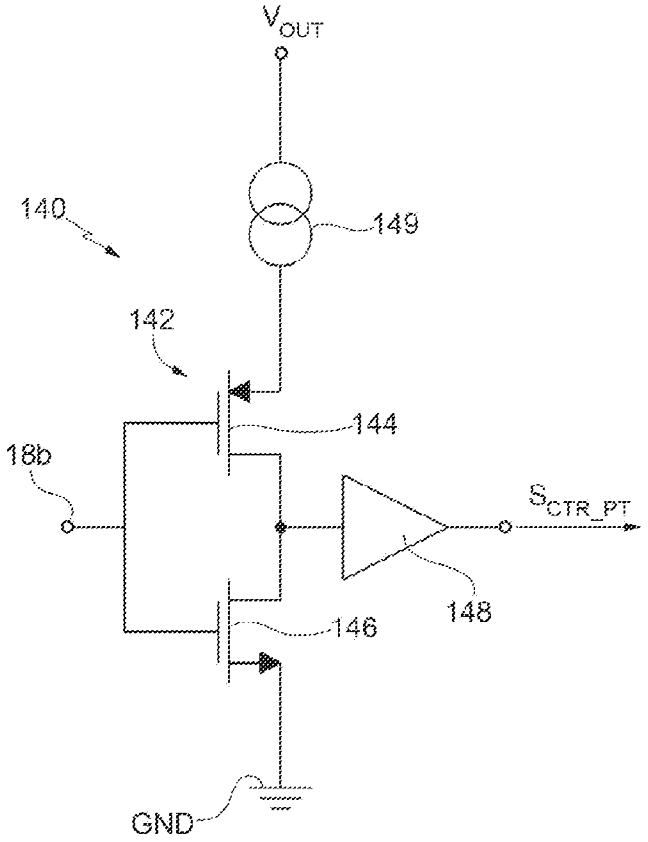
V_{OUT}
S_{CTR_PT}
GND

FIG. 7

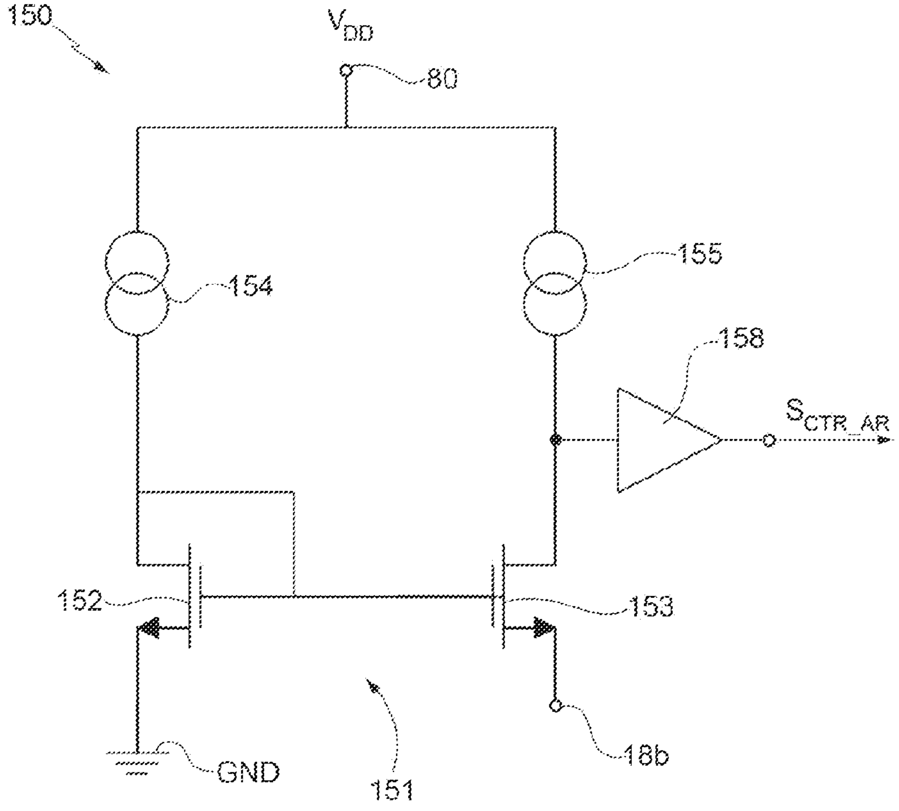
V_{DD}
S_{CTR_AR}
GND

FIG. 8

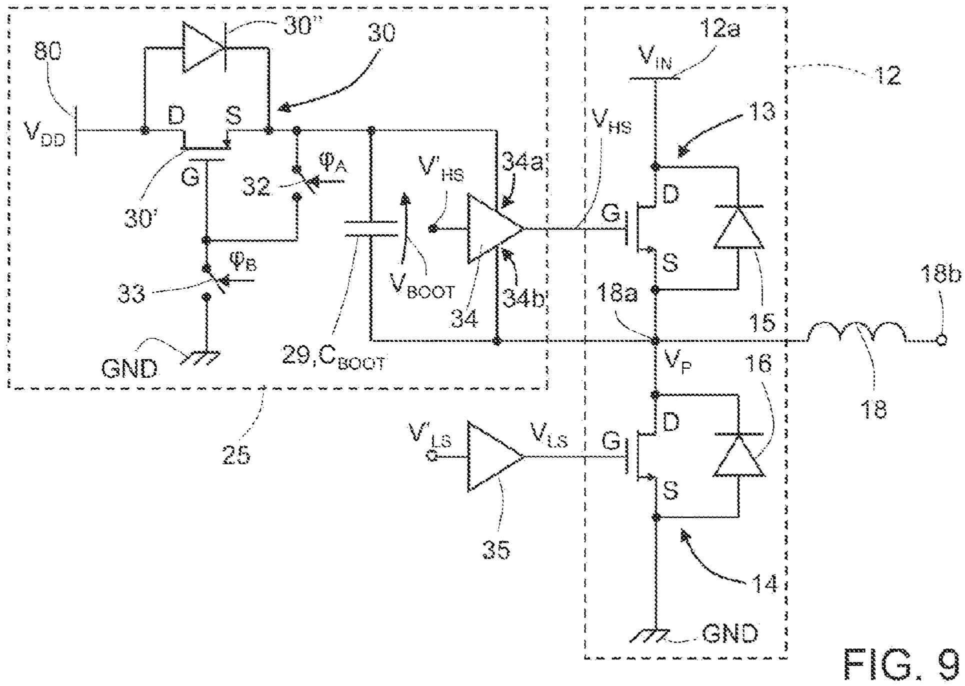
80
V_{DD}
30"
30
D
S
G
30'
φ_A
32
φ_B
33
GND
29, C_{BOOT}
V'_{HS}
V_{BOOT}
34a
34
34b
25
V_{HS}
V_{IN}
12a
12
13
G
D
S
15
18a
V_P
16
18b
18
V'_{LS}
35
V_{LS}
G
D
S
14
GND

FIG. 9

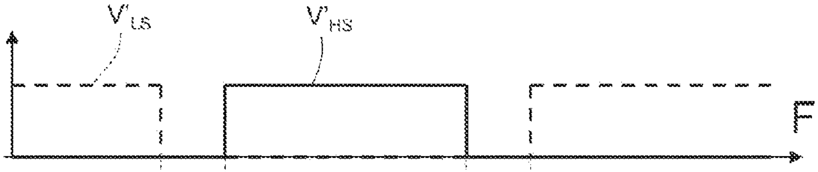

FIG. 10A

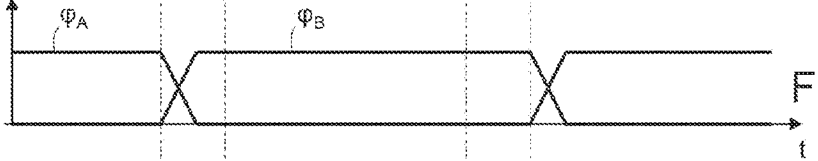

FIG. 10B

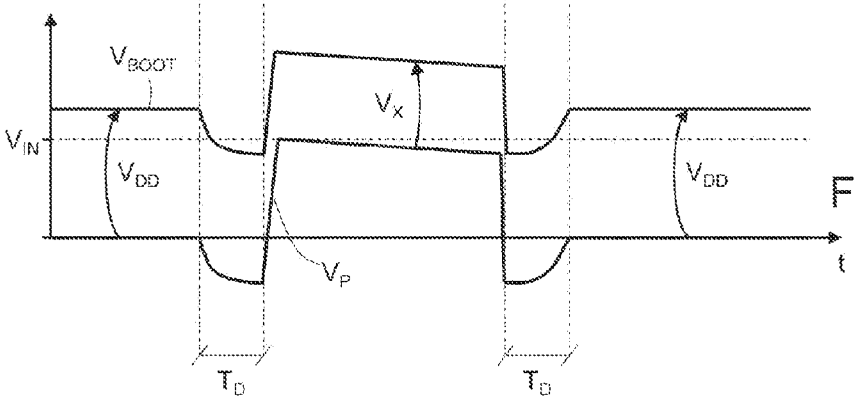

FIG. 10C

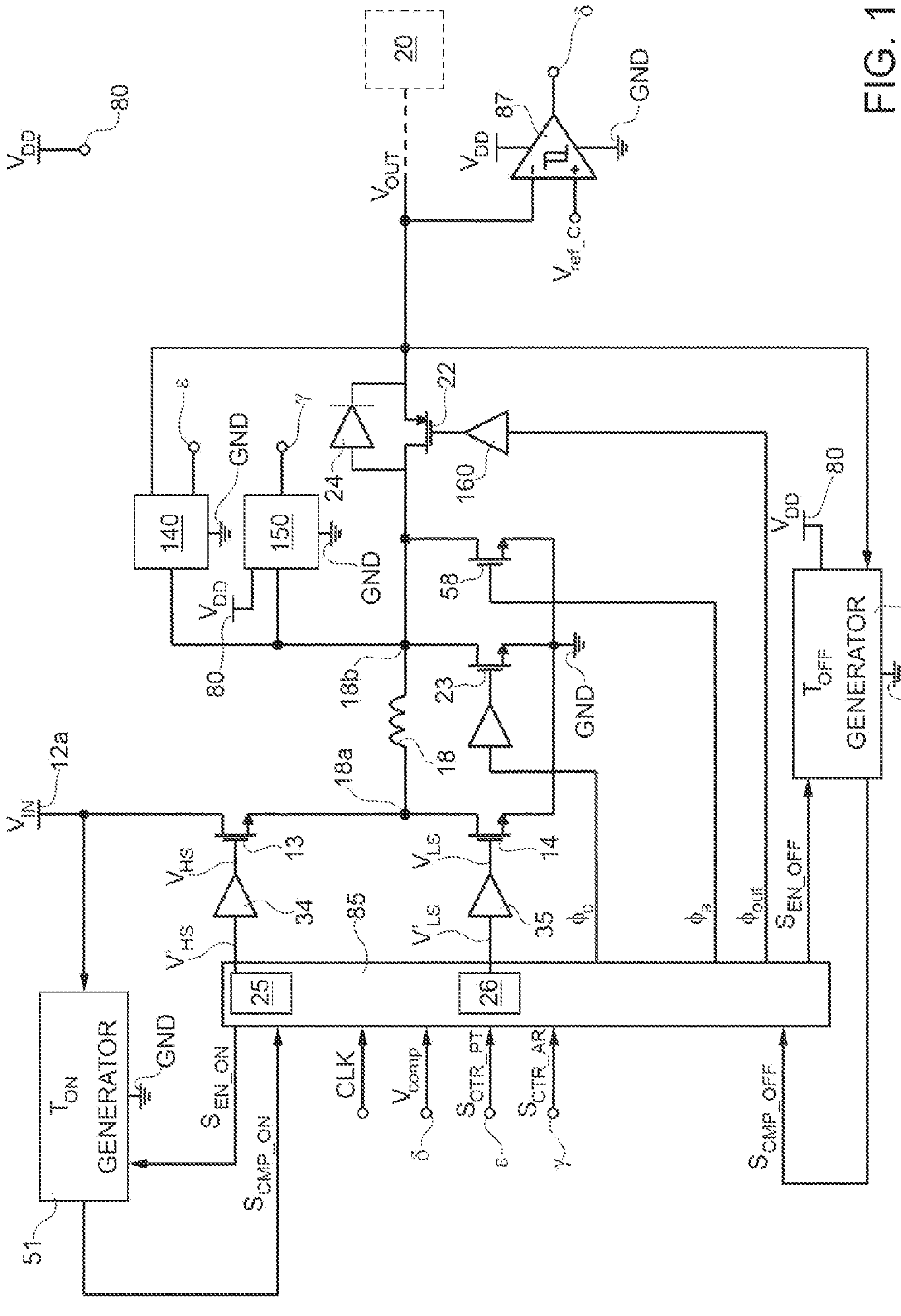
V_{DD}
80
V_{IN}
12a
T_{ON} GENERATOR
GND
51
S_{EN_ON}
S_{CMP_ON}
25
26
85
CLK
V_{comp}
S_{CTR_PT}
S_{CTR_AR}
δ
ε
γ
S_{CMP_OFF}
V'_{HS}
V_{HS}
34
13
V'_{LS}
V_{LS}
35
14
ϕ_b
ϕ_a
ϕ_{out}
S_{EN_OFF}
18a
18
18b
23
GND
80
V_{DD}
140
150
GND
58
24
22
160
V_{DD}
80
T_{OFF} GENERATOR
GND
101
V_{OUT}
20
V_{DD}
87
V_{ref_CO}
GND
δ
ε
γ

FIG. 11

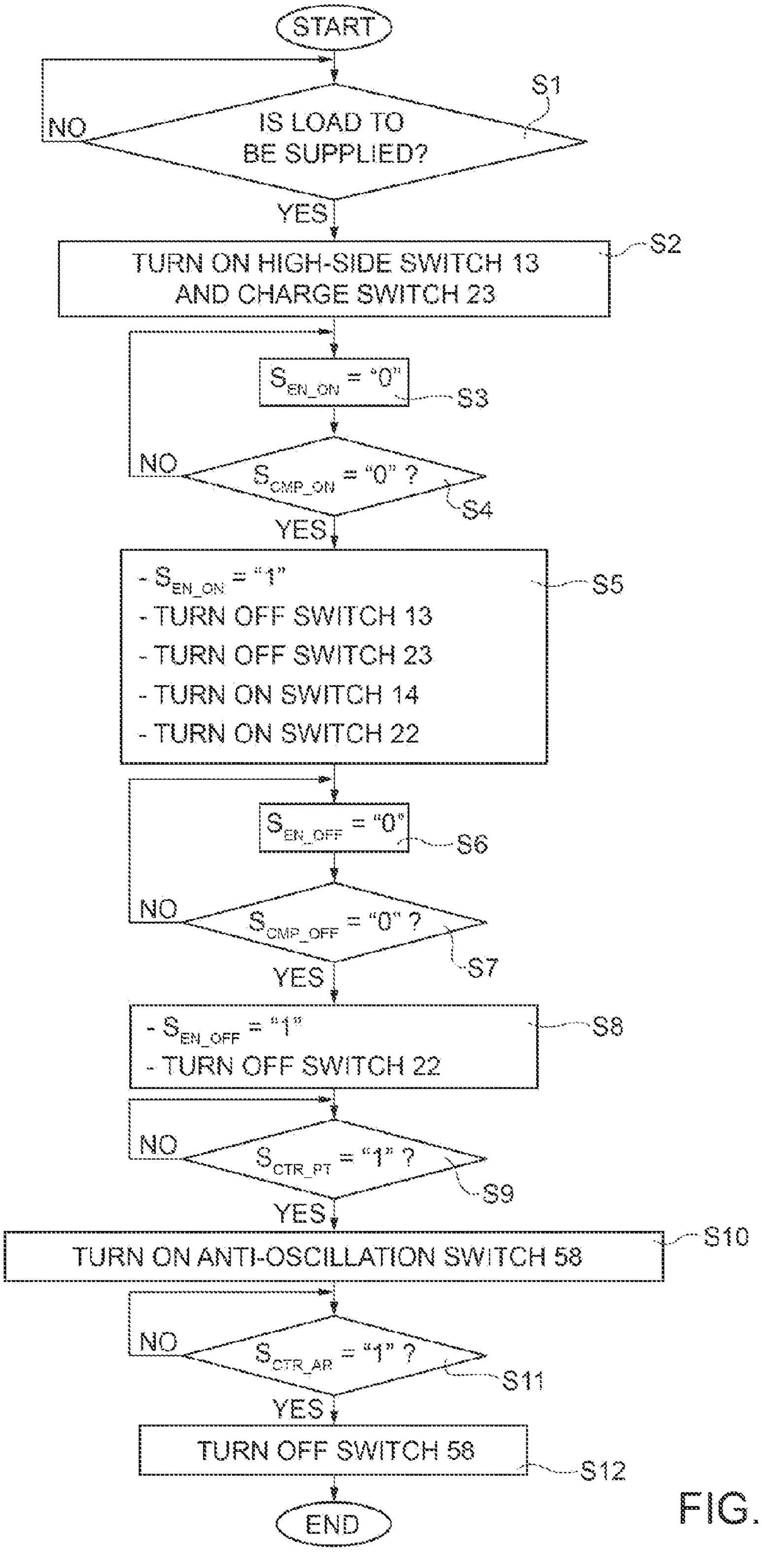
START
IS LOAD TO BE SUPPLIED?
NO
YES
S1
TURN ON HIGH-SIDE SWITCH 13 AND CHARGE SWITCH 23
S2
S_{EN_ON} = "0"
S3
NO
S_{CMP_ON} = "0" ?
S4
YES
- S_{EN_ON} = "1"
- TURN OFF SWITCH 13
- TURN OFF SWITCH 23
- TURN ON SWITCH 14
- TURN ON SWITCH 22
S5
S_{EN_OFF} = "0"
S6
NO
S_{CMP_OFF} = "0" ?
S7
YES
- S_{EN_OFF} = "1"
- TURN OFF SWITCH 22
S8
NO
S_{CTR_PT} = "1" ?
S9
YES
TURN ON ANTI-OSCILLATION SWITCH 58
S10
NO
S_{CTR_AR} = "1" ?
S11
YES
TURN OFF SWITCH 58
S12
END

FIG. 12

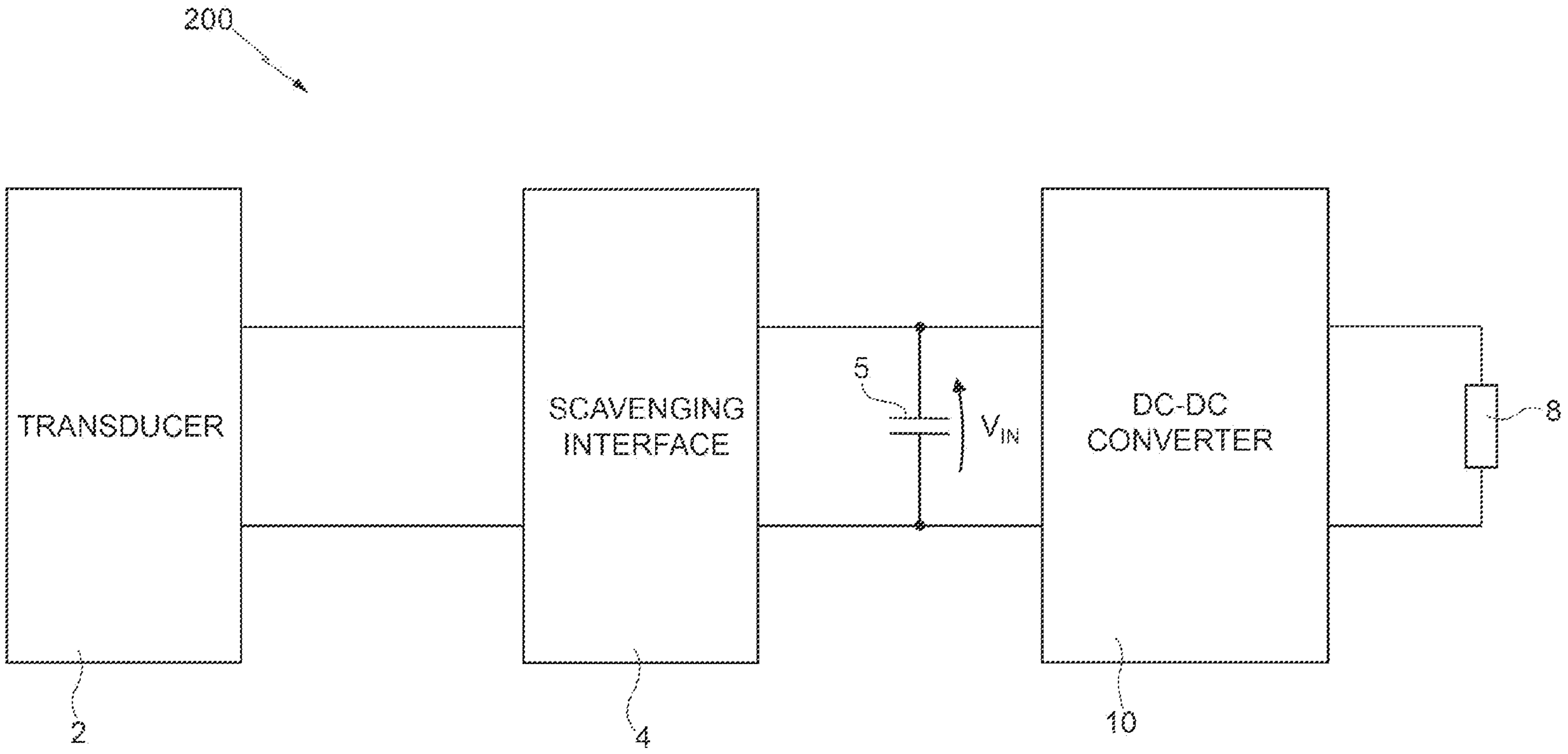
TRANSDUCER
SCAVENGING INTERFACE
V_{IN}
DC-DC CONVERTER

FIG. 13

DC-DC CONVERTER WITH IMPROVED ENERGY MANAGEMENT, METHOD FOR OPERATING THE DC-DC CONVERTER, ENVIRONMENTAL ENERGY HARVESTING SYSTEM USING THE DC-DC CONVERTER, AND APPARATUS USING THE ENERGY HARVESTING SYSTEM

PRIORITY CLAIM

[0001] This application claims priority from Italian Application for Patent No. TO2014A000059 filed Jan. 28, 2014, the disclosure of which is incorporated by reference.

TECHNICAL FIELD

[0002] The disclosure relates to a DC-DC converter with improved energy management and to a method for operating the DC-DC converter. The disclosure further relates to an environmental energy harvesting system comprising the DC-DC converter and to an apparatus comprising said environmental energy harvesting system.

BACKGROUND

[0003] As is known, systems for harvesting energy (also known as "energy scavenging systems") from intermittent environmental energy sources (i.e., ones that supply energy in an irregular way) have aroused and continue to arouse considerable interest in a wide range of fields of technology. Typically, energy harvesting systems are configured to harvest, store and transfer energy generated by mechanical or thermal sources to a generic load of an electrical type.

[0004] Low-frequency vibrations, such as, for example, mechanical vibrations of disturbance in systems with moving parts may be a valid source of energy. The mechanical energy is converted, by one or more appropriate transducers (for example, piezoelectric or electromagnetic devices) into electrical energy, which may be used for supplying an electrical load. In this way, the electrical load does not require batteries or other supply systems that are cumbersome and present a low resistance in regard to mechanical stresses.

[0005] FIG. **1** is a schematic illustration in the form of functional blocks of an energy harvesting system of a known type.

[0006] The energy harvesting system **1** of FIG. **1** comprises: a transducer **2**, for example of an electromagnetic or piezoelectric type, subject in use to environmental mechanical vibrations and configured to convert mechanical energy into electrical energy, typically into AC (alternating current) voltages; a scavenging interface **4**, for example comprising a diode-bridge rectifier circuit (also known as Graetz bridge), configured to receive at input the AC signal generated by the transducer **2** and supplying at output a DC (direct current) signal for charging a capacitor **5** connected to the output of the rectifier circuit **4**; and a DC-DC converter **6**, connected to the capacitor **5** for receiving at input the electrical energy stored by the capacitor **5** and supplying it to an electrical load **8**. Thus, the capacitor **5** has the function of element for storing energy, which is made available, when required, to the electrical load **8** for operation of the latter.

[0007] The transducer **2** is, for example, an electrochemical transducer, or an electromechanical transducer, or an electroacoustic transducer, or an electromagnetic transducer, or a photoelectric transducer, or an electrostatic transducer, or a thermoelectric transducer.

[0008] The main disadvantage of the configuration according to FIG. **1** regards the fact that the maximum voltage supplied at output by the scavenging interface **4** is limited by the input dynamics of the DC-DC converter **8**.

[0009] In applications in which the transducer **2** converts mechanical energy into electrical energy in a discontinuous way and/or the power P_{LOAD} required by the electrical load **8** varies significantly in time, also the voltage V_{OUT} consequently has an evolution variable in time. This causes, for example, a variation of the efficiency factor of the DC-DC converter **6**, which is inversely proportional to the value assumed by V_{OUT}. The maximum value of V_{OUT} is further limited by the range of input voltages accepted by the DC-DC converter.

[0010] European Patent reference No. EP 2518878 (incorporated by reference) describes a DC-DC converter with a high efficiency factor even in conditions of light load and wide dynamic range of the input voltage. According to this reference, it is important to monitor the inductor current to prevent it from assuming negative values (i.e., to prevent energy from being absorbed by the electrical load). For this purpose, according to the reference, a current-detection circuit is provided, which constantly monitors the inductor current and uncouples the electrical load from the inductor when the current monitored assumes the zero value. A solution of this type requires, in the design step, an accurate trade-off between the desired accuracy of the measurement supplied by the current-detection circuit and the consumption of current of said current-detection circuit.

[0011] It is consequently desirable to have available a DC-DC converter that will enable management, in an efficient and, at the same time, accurate way, of supply of the electrical load connected thereto, preventing the DC-DC converter from absorbing energy from the load itself. It is likewise desirable for the management of supply of the load to be carried out automatically and with low current consumption.

SUMMARY

[0012] In an embodiment, a DC-DC converter is provided with improved energy management, a method for operating the DC-DC converter is provided, an energy harvesting system comprising the DC-DC converter is provided and an apparatus comprising the energy harvesting system that will enable the aforementioned advantages to be achieved is provided. In particular, the DC-DC converter enables a high efficiency factor to be maintained even in conditions of light load (e.g., 10 μW-100 μW) and of variable input and output voltages (e.g., 0.25 V-5 V at input, 1 V-5 V at output), minimizing the power consumption of the DC-DC converter.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] For a better understanding of the present invention, preferred embodiments thereof are now described, purely by way of non-limiting example and with reference to the attached drawings, wherein:

[0014] FIG. **1** shows an energy harvesting system of a known type;

[0015] FIG. **2** is a schematic illustration of a DC-DC converter;

[0016] FIGS. **3**A-**3**D show the converter of FIG. **2** in successive operating steps;

[0017] FIGS. **4**A-**4**D show the evolution of the inductor current in the DC-DC converter of FIG. **2**;

[0018] FIGS. **5**-**8** show respective circuits belonging to the converter of FIG. **2** for control of the inductor current in order to implement, respectively, the steps of FIGS. **4A**-**4D**;

[0019] FIG. **9** shows a circuit for driving the converter of FIG. **2** used during the steps of FIGS. **3A**-**3D**;

[0020] FIGS. **10A**-**10C** show driving signals used during operation of the driving circuit of FIG. **9**;

[0021] FIG. **11** shows an embodiment of the DC-DC converter comprising a control logic;

[0022] FIG. **12** shows steps of a method implemented by the control logic of FIG. **11** for operating the DC-DC converter, in order to implement the steps of FIGS. **3A**-**3D** and **4A**-**4D**; and

[0023] FIG. **13** shows an environmental energy harvesting system comprising the DC-DC converter of FIG. **2** or FIG. **11**.

DETAILED DESCRIPTION OF THE DRAWINGS

[0024] FIG. **2** is a schematic illustration of a converter **10**, in particular configured to carry out a DC-DC conversion, of a buck-boost type, comprising an inductor **18** coupled to the input of the converter **10** and at least one output for supplying a respective load **20**. This type of converter is also known as SISO (single-inductor single-output) converter. It is evident that, according to a different embodiment, the converter **10** may be of a SIMO (single-inductor multiple-output) type, having a plurality of outputs configured to supply a respective plurality of electrical loads.

[0025] For simplicity of description, in what follows reference will be made to a converter of a SISO type, without this implying any loss of generality.

[0026] In the description, reference will be made indifferently to switches “open” or “OFF” to indicate that said switches do not conduct electric current and, in the case of transistors, have their control terminal (e.g., the gate terminal) biased in such a way as not to conduct electric current between the source and drain terminals. Further, reference will be made indifferently to switches “closed” or “ON” to indicate that said switches conduct electric current and, in the case of transistors, have their control terminal (e.g., the gate terminal) biased in such a way as to conduct electric current between the source and drain terminals.

[0027] In greater detail, the converter **10**, comprises a main bridge **12**, including a supply terminal **12***a* at a voltage V_{IN} (with V_{IN} ranging, for example, between 0.25 V and 5 V) and a reference terminal GND (connected to ground; however, other reference voltages may be used, for example $-V_{IN}$). The main bridge **12** further includes a high-side switch **13** and a low-side switch **14**, connected together in series between the supply terminal **12***a* and the ground reference terminal GND. In particular, the high-side switch **13** is directly connected to the supply terminal **12***a* for receiving the signal V_{IN} and the low-side switch **14** is directly connected to the ground reference terminal GND.

[0028] According to one embodiment, as shown in FIG. **2**, the high-side switch **13** is a MOSFET, in particular an n-channel double-diffusion MOS (DMOS) transistor, with an ON-resistance (R_{ON}) of approximately 0.2Ω at 100 mA. The low-side switch **14** is of the same type as the transistor **13**. Alternatively, in a way not shown, the high-side switch **13** and the low-side switch **14** may be produced using a different technology, for example they may be p-channel MOSFETs, or NPN or PNP bipolar transistors, IGBTs, or simply diodes.

[0029] Once again with reference to FIG. **2**, a diode **15** and a diode **16** are connected between a respective source terminal S and a respective drain terminal D of the high-side switch **13** and of the low-side switch **14**. The diodes **15** and **16** are connected in configuration known as “antiparallel” (with respect to the normal direction of flow of the current through the high-side switch **13** and the low-side switch **14**). As is known, a characteristic of a MOSFET is that of presenting, in certain operating conditions, the electrical properties of a diode (parasitic diode). Said diode is electrically set (integrated) between the source and drain terminals of the MOSFET. In other words, the high-side switch **13** and the low-side switch **14** may present the electrical behavior of a diode, in which the cathode of the diode corresponds to the drain terminal and the anode to the source terminal of the respective high-side switch **13** and low-side switch **14** (vice versa, in the case of MOSFETs of a p type). In greater detail, the drain terminal D of the high-side switch **13** is connected to the supply terminal **12***a*, the source terminal S of the high-side switch **13** is connected to the drain terminal of the low-side switch **14** and the source terminal S of the low-side switch **14** is connected to the ground reference terminal GND. The high-side switch **13** and low-side switch **14** are driven in conduction by a respective first driving circuit **25** and second driving circuit **26**, which are described more fully hereinafter. The driving circuits **25**, **26** are connected to the control terminal G (or gate terminal) of the transistors that form the high-side switch **13** and the low-side switch **14**, respectively.

[0030] The DC-DC converter **10** further comprises an inductor **18**, having an inductance L for example comprised between approximately 1 μH and approximately 100 μH, in particular 10 μH. The inductor **18** includes a first conduction terminal **18***a* connected between the high-side switch **13** and the low-side switch **14** (in particular, connected to the source terminal S and drain terminal D of the high-side switch **13** and of the low-side switch **14**, respectively) and a second terminal **18***b*, connected to the output bridge of the converter **10**. The second terminal **18***b* is electrically coupled to the electrical load **20** by a coupling switch **22** and may be electrically connected to the reference terminal GND by a charge switch **23**. Further, a diode **24** is connected in parallel to the coupling switch **22**, with the anode coupled to the terminal **18***b* of the inductor **18** and the cathode coupled to the load **20**. According to one embodiment, which in no way limits the scope of the present invention, the coupling switch **22** is a p-channel MOSFET. The diode **24** is the parasitic diode set (integrated) between the source and drain terminals of the MOSFET. In other words, when the coupling switch **22** is OFF, it assumes the electrical behavior of a diode, in which the cathode of the diode corresponds to the source terminal and the anode to the drain terminal of the coupling switch **22**.

[0031] As is described more fully hereinafter, the coupling switch **22** has the function of enabling transfer of the energy stored in the inductor **18** towards the load **20** (coupling switch **22** ON, condition of load supplied and of the inductor current I_L discharged on the output), or else of insulating electrically the load **20** from the inductor **18** (coupling switch **22** OFF, condition of load not supplied in an active way). The presence of the diode **24**, with the cathode facing the load **20**, does not allow transfer of energy from the electrical load **20** to the inductor **18** when the coupling switch **22** is OFF, but instead allows transfer of current in a passive way from the inductor **18** to the electrical load **20** when the coupling switch **22** is OFF and the diode **24** is appropriately biased.

[0032] The electrical load **20** may, for example, require a value of supply voltage V_{OUT} comprised between 1 V and 5 V, or other values still, according to the application, for example sensors of various types, microcontrollers and communication systems.

[0033] In use, the energy required by the electrical load **20** for its operation is, as has been anticipated, supplied by the inductor **18**, which, in turn, is charged and discharged by appropriately controlling the high-side switch **13** and the low-side switch **14** of the main bridge **12** and the switches **23** and **24** of the secondary output bridge. The high-side switch **13** and the low-side switch **14** are driven (opened/closed; turned on/turned off) in such a way as to prevent a direct connection between the terminal **12***a* at the supply voltage V_{IN} and the ground reference terminal GND. Reference may be made for this purpose to FIGS. **9** to **10**C, which are described hereinafter.

[0034] The converter **10** further comprises, in parallel to the charge switch **23**, electrically coupled between the terminal **18***b* of the inductor **18** and the ground reference terminal GND, an anti-oscillation switch **58**, operation of which is described more fully with reference to FIG. **3**D.

[0035] According to one embodiment, both the charge switch **23** and the anti-oscillation switch **58** are transistors, for example n-channel MOSFETs.

[0036] FIGS. **3**A-**3**D show respective operating steps of the converter **10** of FIG. **2** for implementing a method of control of the primary bridge and of the secondary bridge for charging and discharging the inductor **18** during each step of supply of the load **20**.

[0037] For simplicity of representation, all the switches are represented using the graphic symbol of a generic switch and not as MOSFETs. For the same reason, also the respective diodes have been omitted, except for the diode **24**.

[0038] FIGS. **4**A-**4**D show time plots of the variation of the energy (in particular, the variation of the current I_L) in the inductor **18** during each of the operating steps of FIGS. **3**A-**3**D, respectively.

[0039] With reference to FIG. **3**A, in this initial step the inductor **18** is in a discharged state and has to be charged. For this purpose, the high-side switch **13** of the main bridge **12** is closed, whereas the low-side switch **14** is open. Further, in this step, also the charge switch **23** is closed, thus coupling the terminal **18***b* of the inductor **18** to the ground reference terminal GND. The anti-oscillation switch **58** is, instead, open. With joint reference to FIGS. **3**A and **4**A, a current flows between the supply terminal **12***a*, biased at a voltage V_{IN} and the ground reference terminal GND, through the inductor **18**, which is thus energized (charged). Reference may be made to FIG. **4**A, which shows a curve representing the evolution of the current I_L stored in the inductor **18** during the step of FIG. **3**A. The current I_L increases, in a substantially linear way, starting from a minimum value I_0 (in particular the zero value, or a value close to zero) until it reaches a maximum value I_{Lmax}, in a time interval comprised between the instants t_1 and t_{1a}. The maximum value I_{Lmax} is a predefined value, chosen on the basis of the characteristics of the inductor **18** used and on the basis of the electrical load to be supplied, for example 100 mA.

[0040] The step of FIG. **3**A terminates when the energy stored in the inductor **18** reaches the predefined maximum value (current $I_L = I_{Lmax}$).

[0041] The next step is represented in FIG. **3**B. With reference to this figure, the high-side switch **13** and the charge switch **23** are opened, whereas the low-side switch **14** is closed. Opening of the high-side switch **13** is preferably carried out prior to closing of the low-side switch **14**. There is thus prevented direct connection of the supply terminal **12***a* to the ground reference voltage GND, thus avoiding phenomena of cross conduction between the terminal **12***a* and the ground reference terminal GND.

[0042] Then, once again with reference to FIG. **3**B, the coupling switch **22** is closed, thus connecting the load **20** to the terminal **18***b* of the inductor **18**. The terminal **18***a* of the inductor **18** is connected to the ground reference terminal GND via the low-side switch **14** and the energy stored in the inductor **18** flows towards the load **20**, thus supplying it.

[0043] With reference to FIG. **4**B, in this step of supply of the load, the current I_L in the inductor **18** decreases until it reaches a predefined low value I_{RES}, at time t_2. The predefined low value I_{RES} is not the zero value of the current I_L, but a value comprised, for example, between 5% and 20% of the peak value I_{Lmax}.

[0044] Then, FIG. **3**C, the coupling switch **22** is opened and the low-side switch **14** is kept closed. The diode **24** continues to conduct electric current, in so far as, in this situation, it is forward biased. The residual current stored in the inductor **18** consequently continues to flow towards the load **20**, through the diode **24** until it reaches the zero value. This condition is illustrated in FIG. **4**C, which shows the decrease of the current I_L from the value I_{RES} to the value I_0 (zero current value), at time t_3. This step is a step of passive transfer of the energy towards the load **20**.

[0045] During passive transfer of charge from the inductor **18** to the electrical load **20**, the transistor that constitutes the coupling switch **22** is OFF and transfer of charge takes place via the diode integrated in the transistor **22** (in the figures, this is the diode designated by the reference number **24**). Discharge of the inductor current is a function of the output supply voltage V_{OUT} and of a difference of potential across the coupling switch **22**, corresponding to the voltage V_{BE} across the diode **24**.

[0046] The next step is represented in FIG. **3**D. Here, the anti-oscillation switch **58** (which up to now was open) is driven into a closed state, whereas the switch **23** is kept open. Both of the terminals **18***a* and **18***b* of the inductor **18** are thus coupled to the ground reference terminal GND. In this way, any spurious voltage oscillations on the terminals **18***a* and **18***b*, caused by residual energy possibly stored in the inductor **18** and in the parasitic capacitors associated to the terminals **18***a* and **18***b*, are prevented. In fact, any possible undesirable oscillations may cause problems of an electromagnetic nature, in addition to a possible reduction of the global efficiency of the converter **10**.

[0047] With reference to FIG. **4**D, it may be noted that the current I_L in the inductor **18** drops to negative values (i.e., in the inductor **18** a current flows from the terminal **18***b* to the terminal **18***a*) and settles, at time t_4, to the value I_0.

[0048] Then, the cycle of complete charge and discharge of the inductor **18** may restart, if necessary, for supplying the load **20** one again (i.e., the cycle returns to the step of FIGS. **3**A, **4**A).

[0049] The duration of the time interval t_1-t_{1a}, in what follows referred to as interval T_{ON}, is not predefined, but is calculated in an appropriate way so that, as the supply voltage V_{IN} varies, the interval T_{ON} will always enable charging of the inductor **18** to the desired maximum value I_{Lmax}. In other words, a voltage V_{IN} of a high value will charge the inductor

18 in a short time interval T_{ON}; instead, a voltage V_{IN} of a low value will charge the inductor **18** in a longer time interval T_{ON}.

[0050] With reference to FIG. **2**, the function of definition of the appropriate interval T_{ON} is performed by the first adaptive control block **51**, electrically coupled to the supply terminal **12***a* for receiving the voltage V_{IN} on the basis of which the length of the interval T_{ON} is to be determined in an adaptive way.

[0051] FIG. **5** shows, according to one embodiment, a circuit implementation of the first adaptive control block **51**, for calculation of T_{ON}.

[0052] The first adaptive control block **51** comprises a branch **52** including a resistive voltage divider formed by a resistor **54** having a resistance R of 6 MΩ and by a resistor **56**, connected in series to the first resistor **54**, having a resistance equal to R/2, i.e., 3 MΩ. The resistors **54** and **56** are connected together in series. In particular, the resistor **54** has a first terminal coupled to V_{IN} and a second terminal coupled to an intermediate node **55**; the resistor **56** has a first terminal **56***a* coupled to the intermediate node **55** and a second terminal that may be coupled to the ground reference terminal GND. Electrical coupling and uncoupling of the second terminal of the resistor **56** to the ground reference terminal GND is managed by a switch **57** (e.g., a MOSFET), connected in series to the second resistor **56** between the second terminal and the ground reference terminal GND. The switch **57** is in a closed state during the operating steps of the adaptive control block **51** and is opened when the converter **10** is OFF or when the adaptive control block **51** is not functioning in order to prevent any undesirable dissipation of current. The switch **57** may, however, not be present.

[0053] An amplifier **60** includes a non-inverting input **60***a* electrically coupled to the intermediate node **55**, configured to receive, in use, a voltage substantially equal to $V_{IN}/3$ (assuming, for simplicity, that the resistance introduced by the switch **58** is ideally substantially zero). The inverting terminal **60***b* of the amplifier **60** is connected to the ground reference terminal GND through a resistor **62** having a resistance R_{ON}=3 MΩ. The output terminal **60***c* of the amplifier **60** is configured to bias the gate terminal of a transistor, e.g. an n-channel MOSFET, **64**. Further, the source terminal S of the transistor **64** is electrically coupled to the inverting terminal **60***b* and, via the resistor **62**, to the ground reference terminal GND.

[0054] At the instant t_1 (see FIG. **4**A), when it is desired to start charging of the inductor **18**, the switch **57** is closed. In this way, the non-inverting terminal **60***a* is biased at a voltage equal to $V_{IN}/3$, as indicated previously. The output **60***c* of the amplifier **60** assumes a value such as to bias the gate of the transistor **64**, which starts to conduct an electric current $i_1=V_{IN}/(IR_{ON})$. In practice, the stage formed by the amplifier **60**, the transistor **64** and the resistor **62** forms a voltage-to-current converter.

[0055] The drain terminal of the transistor **64** is connected to a supply terminal **80** at a voltage V_{DD} (e.g., comprised between approximately 2 V and approximately 5 V, for example approximately 3.5 V) via a transistor **79**. The voltage V_{DD} is, for example, supplied by a battery, not shown in the figure. The transistor **79** comprises a source terminal connected to the supply terminal **80** and a drain terminal and gate terminal connected together and to the drain terminal of the transistor **64**. A further transistor **82** is connected in current-mirror configuration **83** with the transistor **79**. In particular, the transistor **82** comprises a gate terminal connected to the gate terminal of the transistor **79** and a source terminal connected to the supply terminal **80**. The current mirror **83** has a gain ratio 1:M and has the function of generating a current i_2, through the transistor **82**, of a value $i_2=M\cdot i_1$. According to one embodiment, the value of M is chosen equal to 8 and is obtained by sizing appropriately the transistor **82**, in a per se known manner.

[0056] The first adaptive control block **51**, thus, converts the input voltage V_{IN} into a current signal i_2 that is inversely proportional to the voltage V_{IN}.

[0057] The current signal i_2 is used for charging a capacitor **84** (current signal i_{C_ON}).

[0058] As may be noted from FIG. **5**, the drain terminal of the transistor **82** is connected to the ground reference terminal GND by a pair of transistors **76**, **78**. The transistor **76** is a p-channel MOSFET, whereas the transistor **78** is an n-channel MOSFET and, together, constitute a current-limited inverter.

[0059] The transistors **76**, **78** have their respective gate terminals coupled to the same terminal α for receiving an enable signal S_{EN_ON}. The enable signal S_{EN_ON} is a binary signal configured to assume two values: a "high" value configured to turn on the transistor **78** and, at the same time, to turn off the transistor **76**; and a "low" value configured to turn off the transistor **78** and, at the same time, to turn on the transistor **76**. In a way evident to a person skilled in the branch, the "high" and "low" values are chosen appropriately on the basis of the biasing characteristics of the transistors **76**, **78** used and it is possible to envisage use of an appropriate buffer for adequately biasing the gates of the transistors **76** and **78**.

[0060] In use, during the step of charging of the inductor **18** (FIGS. **3**A, **4**A) it is expedient to monitor the time interval T_{ON} in order for the inductor **18** to be charged up to the required maximum value I_{Lmax}. For this purpose, at time t_1, the enable signal S_{EN_ON} assumes a low value for turning on the transistor **76** and turn off the transistor **78**. The current i_2 thus flows from the supply terminal **80**, through the transistors **82** and **76**, to the capacitor **84**, thus charging it with a current $i_{C\_ON}=i_2$.

[0061] The first adaptive control block **51** further comprises a comparator **86**, configured to receive on its inverting terminal the value of voltage across the capacitor **84** and, on its non-inverting terminal, a reference signal S_{REF_ON} (voltage signal). Said reference value S_{REF_ON} is chosen in such a way that the time of charging of the capacitor **84**, from a zero value to a value S_{REF_ON}, is equal to the desired value for the interval T_{ON}. In this way, the time taken by the capacitor **84** to be charged up to the value S_{REF_ON} is the interval T_{ON}, which is inversely proportional to the input voltage V_{IN}. The interval T_{ON} may thus be determined directly on the basis of the time taken by the capacitor **84** to reach the value S_{REF_ON} starting from a discharged condition of the capacitor **84**.

[0062] According to one embodiment, the reference value S_{REF_ON}, is equal to approximately 1 V. The signal S_{REF_ON} is, as has been said, a voltage reference signal and is generated in a per se known manner, for example by a reference-signal-generator block **90**. The reference-signal-generator block **90** is connected between the supply terminal **80**, for receiving the voltage V_{DD} and the ground reference terminal GND. The supply voltage V_{DD} is applied across a branch including a Zener diode **91** set in series with a resistor **92**. In greater detail, the anode of the Zener diode **91** is connected to ground

GND, the cathode of the Zener diode **91** is connected to a terminal of the resistor **92** and the other terminal of the resistor **92** is connected to the supply terminal **80**. The reference signal S_{REF_ON} is taken at the cathode of the Zener diode **91**. Operation of the reference-signal-generator block **90** is in itself evident. The Zener diode **91** operates as voltage limiter and the resistor **92** has the function of introducing a biasing current in the Zener diode **91**. The value of the resistance **92** is to be calculated on the basis of the specific requirements, i.e., as a function of the effective value of the biasing current of the Zener diode **91**. To obtain the effective value of resistance of the resistor **92** it is sufficient to use the formula $(V_{DD}-V_Z)/i_A$, where V_{DD} is the supply voltage, V_Z is the operating voltage of the Zener diode **91** and i_A is the current of absorption of the block **90** (absorption of the Zener diode **91**). The value of V_Z defines, as is evident, the voltage value desired for reference signal S_{REF_ON} (a Zener diode **91** with an operating voltage equal to the value desired for the reference S_{REF_ON} will be used).

[0063] According to one embodiment, the capacitor **84** is a variable-capacitance capacitor. In this way, at the end of the steps of manufacture of the DC-DC converter **10**, i.e., during a step of testing of the DC-DC converter **10**, it is possible to set a desired value of capacitance of the capacitor **84**, which may have undergone variations on account of process spread. According to one embodiment, the value of capacitance C_{ON} of the capacitor **84** is chosen and set at 900 fF.

[0064] The comparator **86** is, in particular, a hysteretic comparator, configured to generate a hysteresis around the value set by the reference signal S_{REF_ON}, for example a hysteresis interval of approximately 20 mV. This is useful in order to compensate for possible undesirable variations of the signals involved in the comparison step.

[0065] The comparator **86** is configured to generate at output a comparison signal S_{CMP_ON}, configured to assume two values: a "low" value (logic value "0") at time t_{1a} (which identifies the end of the interval T_{ON}); and a "high" value (logic value "1") otherwise.

[0066] In use, when the voltage across the capacitor **84** reaches the value set by the reference signal S_{REF_ON}, the comparison signal S_{CMP_ON} changes state (e.g., passing from a high logic value to a low logic value), thus indicating that the current in the inductor has reached the maximum value I_{Lmax}. The next step is represented in FIGS. **3**B, **4**B. During this step (and during the subsequent steps) the capacitor **84** is discharged completely. For this purpose, the enable signal S_{EN_ON} assumes a "high" value, such as to turn off the transistor **76** and turn on the transistor **78**. The capacitor **84** has, in this condition, both of the terminals coupled to the ground reference terminal GND and discharges. The comparison signal S_{CMP_ON} once again assumes the high logic value.

[0067] The comparison signal S_{CMP_ON} thus generated is used for determining the end of the interval T_{ON} and consequently a falling edge thereof determines turning-off of the high-side switch **13**. In this way, the duration of the interval T_{ON} varies as a function of the value of the input voltage V_{IN} in so far as the time of charging of the capacitor **84** is inversely proportional to the value of the input voltage V_{IN}. Consequently, once the value S_{REF_ON} is appropriately fixed, the peak current I_{Lmax} stored in the inductor **18** assumes a value that is always constant, even as V_{IN} varies. The time interval T_{ON} during which the inductor **18** is being charged assumes, instead, a value variable according to the value of input voltage V_{IN} (which is inversely proportional to V_{IN}) or, rather, according to the value assumed by the current signal i_2, which is proportional to the current signal i_1 generated starting from the voltage value V_{IN}.

[0068] In greater detail, it is found that the value of current $i_{C\_ON}=i_2$ is equal to $(M\cdot V_{IN})/(3R_{ON})$, where R_{ON} is the resistance of the resistor **62** and M is the value of current gain of the current mirror **83** (e.g., M=8). Consequently, we have that the value of the interval T_{ON} is equal to $(3R_{ON}\cdot C_{ON}\cdot S_{REF\_ON})/(M\cdot V_{IN})$, where C_{ON} is the value of capacitance of the capacitor **84**.

[0069] The value I_{Lmax} is given by $(T_{ON}\cdot V_{IN})/L$, where L is the value of inductance of the inductor **18**. It follows that the predefined maximum value I_{Lmax} for the current I_L stored in the inductor **18** is irrespective of the value of the input voltage V_{IN} and is given by

$$I_{Lmax}=(3R_{ON}\cdot C_{ON}\cdot S_{REF\_ON})/(M\cdot L)$$

[0070] To return to FIG. **2**, the converter **10** further comprises a second adaptive control block **101**, configured to generate a discharge time interval T_{OFF} during which the step described with reference to FIGS. **3**B and **4**B is carried out. During the discharge time interval T_{OFF} the charge stored in the inductor **18** flows to the load **20**, thus supplying it. As has been said, this step of discharge (active discharge) occurs between the time t_{1a} and the time t_2 by driving the coupling switch **22** in conduction (transistor ON); a subsequent step of passive discharge occurs by opening the switch **22** (transistor OFF) and causing flow of the residual charge through the diode **24**. The discharge time interval T_{OFF} is an interval during which there is a process of active discharge of the inductor **18** (i.e., using active elements, namely, the transistor that constitutes the switch **22**). During the step of active discharge, the output supply signal V_{OUT} is applied also to the node **18***b* of the inductor, since the coupling switch **22** is ON.

[0071] With reference to FIG. **6**, the second adaptive control block **101** comprises an input branch **102** connected between the load **20** and the ground reference terminal GND. The input branch **102** comprises, connected in series together, a resistor **104**, having a resistance R_{OFF}, for example equal to 9 MΩ, a transistor **105**, of an n-channel type and a transistor **106**, of an re-channel type.

[0072] The resistor **104** has one conduction terminal coupled to the supply voltage for supply of the load **20** and the other conduction terminal coupled to the drain terminal of the transistor **105**. In turn, the transistor **105** has its source terminal coupled to the drain terminal of the transistor **106** and the latter has its source terminal electrically coupled to the ground reference terminal GND.

[0073] According to the embodiment shown in FIG. **6**, a switch **103** is set in series to the resistor **104**, between the resistor **104** and the transistor **105**. During operation of the second adaptive control block **101**, this switch **103** is in a closed state (i.e., it conducts current). When the second adaptive control block **101** is not used, or when the converter **10** is OFF, the switch **103** is open (i.e., it does not conduct). The switch **10** is optional and is configured to guarantee selective electrical insulation of the second adaptive control block **101** in order to prevent undesirable consumption of current when the converter **10** is OFF or when the second adaptive control block **101** is not used.

[0074] In greater detail, the transistor **105** forms, together with a further transistor **108**, a current mirror **107**. Also the transistor **106** forms, together with a further transistor **110**, a respective current mirror **109**. For this purpose, the gate terminal of the transistor **105** is electrically connected to the gate terminal of the transistor **108**; further, the drain terminal of the transistor **105** is electrically connected to the gate terminals of the transistors **105** and **108**. Likewise, the gate terminal of the transistor **106** is electrically connected to the gate terminal of the transistor **110**; further, the drain terminal of the transistor **106** is electrically connected to the gate terminals of the transistors **106** and **110**.

[0075] The transistors **108** and **110** are connected together in series on an intermediate branch **112** of the adaptive control block **101**. The intermediate branch **112** further comprises a p-channel transistor **114** connected between the drain terminal of the transistor **108** and a supply terminal (the latter, in particular, coinciding with the supply terminal **80** shown in FIG. **5** and being supplied by the voltage V_{DD} itself). In greater detail, the source terminal of the transistor **114** is connected to the voltage V_{DD}, the drain terminal of the transistor **114** is connected to the drain terminal of the transistor **108**, the source terminal of the transistor **108** is connected to the drain terminal of the transistor **110** and the source terminal of the transistor **110** is connected to the ground reference terminal GND. Further connected between the drain terminal of the transistor **110** and the ground reference terminal GND (i.e., in parallel to the transistor **110**) is a resistor **116**, which has, for example, a value of resistance equal to $R_{OFF}/2$.

[0076] In use, when an output voltage V_{OUT} is present on the terminal **18***b*, we have that a current i_{OFF_IN} flows on the input branch **102**, through the resistor **104** and the transistors **105**, **106**. The voltage drop on the resistor **104** is thus equal to $V_{OUT}-2V_{GS}$, where V_{GS} is the voltage drop between gate and source terminals of each of the transistors **105** and **106**, which are assumed as being the same as one another. The current t_{OFF_IN} is thus given by

$$i_{OFF\_IN}=(V_{OUT}-2V_{GS})/R_{OFF}$$

[0077] On the basis of operation of the current mirrors **107** and **109** we have that the current i_{INT} that flows in the intermediate branch **112** is equal to $i_{OFF\_IN}+2V_{GS}/R_{OFF}$, where, also in this case, V_{GS} is the voltage drop between the gate and source terminals of the transistor **110**, which is assumed as being the same as the transistors **105** and **106** and thus as having the same voltage V_{GS}. As a result of operation of the current mirrors **107** and **109** and of the presence of the resistor **116** with resistance $R_{OFF}/2$, we thus have: $i_{INT}=V_{OUT}/R_{OFF}$; namely, the current that flows in the intermediate branch **112** is irrespective of the voltages V_{GS} of the transistors belonging to the current mirrors **107** and **109** (which are assumed as being the same as one another).

[0078] The second adaptive control block **101** further comprises an additional current mirror **116**, formed by the transistor **114** (belonging to the intermediate branch **112**) and by a transistor **118**, of a p-channel type, which belongs to an output branch **122**.

[0079] The transistors **114** and **118** have their respective source terminals connected to the supply terminal **80** and respective gate terminals connected together and to the drain terminal of the transistor **114**. The current mirror **116** has a gain ratio 1:N and has the function of generating a current i_{OFF_OUT} of a value $N\cdot i_{INT}$. According to one embodiment of the present invention, the value of N is chosen equal to 10 and is obtained by appropriately sizing the transistor **118** in a per se known manner.

[0080] The output branch **122** further comprises a pair of transistors **126**, **128**. The transistor **126** is a p-channel MOSFET, whereas the transistor **128** is an n-channel MOSFET and, together, provide a current-limited inverter.

[0081] The transistors **126**, **128** have their respective gate terminals coupled to the same terminal β for receiving an enable signal S_{EN_OFF}. The enable signal S_{EN_OFF} is a signal configured to assume two values: a "high" value configured to turn on the transistor **128** and, at the same time, to turn off the transistor **126**; and a "low" value configured to turn off the transistor **128** and, at the same time, to turn on the transistor **126**. In a way evident for the person skilled in the branch, the "high" and "low" values are chosen appropriately on the basis of the biasing characteristics of the transistors **126**, **128** used.

[0082] In use, during the step of active discharge of the inductor **18** (FIGS. 3B, 4B) it is expedient to manage the time interval T_{OFF} so that the inductor **18** will be discharged to the intermediate value I_{RES} envisaged. For this purpose, at time t_{1a}, the enable signal S_{EN_OFF} assumes a low logic value, thus turning on the transistor **126** and turning off the transistor **128**. The current i_{OFF_OUT} thus flows from the supply terminal **80**, through the transistors **118** and **126**, to the capacitor **134**, charging it with a current $i_{C\_OFF}=i_{OFF\_OUT}$.

[0083] The second adaptive control block **101** further comprises a comparator **136**, which is configured to receive on the inverting terminal the value of voltage across the capacitor **134** and, on the non-inverting terminal, a reference signal S_{REF_OFF} (voltage signal) configured to set a maximum voltage value that the voltage across the capacitor **134** may assume. Said reference value S_{REF_OFF} is chosen in such a way that the time of charging of the capacitor **134** from the value that it assumes when it is discharged to the reference value S_{REF_OFF} is equal to the value desired for the interval T_{OFF}. In this way, the time taken by the capacitor **134** to charge up to the reference value S_{REF_OFF} corresponds to the interval T_{OFF}.

[0084] According to one embodiment, the reference value S_{REF_OFF}, is approximately 1 V. The signal S_{REF_OFF} is, as has been said, a voltage reference signal and is generated in a per se known manner, for example by a reference-signal-generator block similar to the reference-signal-generator block **90** of FIG. **5** and thus not described any further.

[0085] The capacitor **134** is, according to one embodiment, a variable-capacitance capacitor. In this way, at the end of the steps of manufacture of the DC-DC converter **10**, i.e., during a step of testing of the DC-DC converter **10**, it is possible to set a desired value of capacitance of the capacitor **134**, which may have undergone variations on account of process spread. According to one embodiment, the value of capacitance C_{OFF} of the capacitor **134** is chosen and set at 900 fF.

[0086] The comparator **136** is, in particular, a hysteretic comparator, configured to generate a hysteresis around the reference value S_{REF_OFF}, for example a hysteresis interval of approximately 10 mV. This is useful in order to compensate for possible undesirable variations of the signals involved in the comparison step.

[0087] The comparator **136** is configured to generate at output a comparison signal S_{CMP_OFF}, configured to assume two values: a "low" value (logic value "0") at time t_2 (i.e., the time that indicates the end of the interval T_{OFF}); and a "high" value (logic value "1") otherwise.

[0088] In use, when the voltage across the capacitor **134** reaches the reference value S_{REF_OFF}, the comparison signal S_{CMP_OFF} changes state (e.g., passing from the high logic value to the low logic value), thus indicating that the current in the inductor has reached the intermediate value I_{RES}. The next step is represented in FIGS. **3**C, **4**C. During this step (and during the subsequent steps) the capacitor **134** is discharged completely. For this purpose, the enable signal S_{EN_OFF} assumes a "high" value, such as to turn off the transistor **126** and turn on the transistor **128**. The capacitor **134** has, in this condition, both of the terminals coupled to the ground reference terminal GND and discharges.

[0089] The coupling switch **22** is kept in the ON state only between t_{1a} and t_2 and is OFF when the comparison signal S_{CMP_OFF} assumes a low value. In this way, the intermediate value I_{RES} of the current that supplies the load **20** through the inductor **18** assumes a constant value even as the supply voltage of the load **20** varies, i.e., as T_{OFF} varies. The time interval T_{OFF} during which the inductor **18** is discharged assumes in fact a value that varies according to the value of output voltage V_{OUT}.

[0090] In greater detail, it is found that the value of current $i_{C\_OFF}=i_{OFF\_OUT}$ is equal to $N \cdot V_{OUT}/(R_{OFF})$, where, it is recalled, N is the value of current gain of the current mirror **116** (for example, if N=10 then $i_{C\_OFF}=10V_{OUT}/R_{OFF}$). Consequently, we have that the value of the interval T_{OFF} is equal to $(R_{OFF} \cdot C_{OFF} \cdot S_{REF\_OFF})/(N \cdot V_{OUT})$, where C_{OFF} is the value of capacitance of the capacitor **134**.

[0091] The variation of current ΔI between the maximum value I_{Lmax} and the intermediate value I_{RES} (i.e., $\Delta I=I_{Lmax}-I_{RES}$) is given by $(T_{OFF} \cdot V_{OUT})/L$, where L is the inductance of the inductor **18**. From this it follows that the value of intermediate current I_{RES} is given by $I_{Lmax}-\Delta I$:

$$I_{RES}=(3R_{ON} \cdot C_{ON} \cdot S_{REF\_ON})/(M \cdot L)-(R_{OFF} \cdot C_{OFF} \cdot S_{REF\_OFF})/(N \cdot L)$$

[0092] To return to FIG. **2**, the converter **10** further comprises a first voltage comparator **140** electrically coupled between the inductor **18** and the coupling switch **22** (i.e., electrically connected to the terminal **18***b* of the inductor **18**).

[0093] The first voltage comparator **140** is shown in greater detail in FIG. **7** and comprises a first input terminal, which is connected between the terminal **18***b* of the inductor **18** and the anode of the diode **24**, a second input terminal, which is connected to the cathode of the diode **24** and receives the output voltage V_{OUT}, a third input terminal, which is coupled to the ground reference terminal GND and an output terminal, which is configured to supply a control signal S_{CTR_PT} that identifies whether the step of passive transfer (i.e., transfer obtained by passive components, such as the diode **24**) of energy to the load **20** has terminated or is in progress (step of FIGS. **3**C, **4**C). The first voltage comparator **140** comprises a current-limited inverter **142** formed by a transistor **144**, in particular a MOSFET of a p-channel type and by a transistor **146**, in particular a MOSFET of an n-channel type. The transistor **144** and the transistor **146** have their respective drain terminals connected together and their respective gate terminals connected to the terminal **18***b* of the inductor **18**. Further, the source terminal of the transistor **146** is coupled to the ground reference GND, whereas the source terminal of the transistor **144** is coupled to the cathode of the diode **24** for receiving the output voltage V_{OUT}. The first voltage comparator **140** further comprises a buffer **148**, having an input coupled to the drain terminals of the transistors **144** and **146** and an output configured to supply the control signal S_{CTR_PT}. The buffer **148** has the function of rectifier, for generating at output a signal that is correlated to the signal that it receives at input but has faster rising and falling edges. The buffer **148** is configured to receive at input a signal varying in the range $0-V_{OUT}$ and generate at output a signal having a low logic value "0" if the input is below a predefined threshold, for example equal to $V_{OUT}/2$, or else a high logic value "1" if the input is above the predefined threshold, or vice versa. Preferably, the buffer **148** has a threshold with hysteresis.

[0094] The presence of the current generator **149**, between the inverter **142** and the second input terminal at voltage V_{OUT}, represents a current limitation, for example obtained using a current mirror with p-channel transistors, for limiting the current through the inverter **142** to a value of 100 nA.

[0095] In use, during the steps of active transfer (FIGS. **3**B, **4**B) and passive transfer (FIGS. **3**C, **4**C), the signal present on the gate terminal of the transistors **144** and **146** is such as to drive the transistor **144** into the OFF state and the transistor **146** into the ON state.

[0096] In use, when the coupling switch **22** is OFF and transfer of energy to the load **20** occurs only via the diode **24**, the inductor current I_L drops to the zero value. Also the voltage V_{BE} across the diode **24** drops until it reaches to the zero value. We thus have a zero voltage drop on the diode **24** and the terminal **18***b* is found at an output voltage V_{OUT}. Supply of the load **20** is thus concluded. However, the inductor current I_L continues to drop (i.e., an inverse current is present in the inductor **18**) as a result of the parasitic capacitances that couple the terminal **18***b* of the inductor to the ground reference terminal GND (there should here be considered, in general, all the parasitic capacitances present and deriving from the manufacturing steps, for example from production of the transistors that provide the switches **23** and **58**). The drop in voltage on the terminal **18***b* causes turning-off of the transistor **146** and turning-on of the transistor **144**. Consequently, the control signal S_{CTR_PT} changes its logic state, going from the low logic value "0" to the high logic value "1". The control signal S_{CTR_PT}="1" identifies the end of the step of FIGS. **3**C and **4**C and determines passage to the next step of FIGS. **3**D and **4**D.

[0097] Passage from the step of passive supply of the load **20** to the step of stabilization of the inductor current I_L is obtained by closing the anti-oscillation switch **58**. The converter **10** remains in the step of FIG. **3**D until the inductor current I_L settles to the zero value (i.e., to the reference voltage value GND). FIG. **8** shows a circuit implementation of a second voltage comparator **150** configured to monitor, during the step of FIG. **3**D, the voltage on the node **18***b*, which is proportional to the inductor current I_L. For this purpose, the second voltage comparator **150** is electrically coupled between the terminal **18***b* of the inductor **18** and the anode of the diode **24**, to the ground reference terminal GND and to the supply voltage V_{DD} (supply terminal **80**). The second voltage comparator **150** comprises a current mirror **151**. The current mirror **151** includes a transistor **152**, of an n-channel type, coupled between the terminal **80** at the supply voltage V_{DD} and the ground reference terminal GND and a transistor **153**, of an re-channel type, coupled between the terminal **80** at the supply voltage V_{DD} and the terminal **18***b* of the inductor **18**. The current generators **154**, **155**, set between the transistor **152** and the supply terminal **80** and, respectively, between the transistor **153** and the supply terminal **80**, represent schematically a current limitation and are, for example, implemented by a pair of current mirrors with p-channel transistors, for limiting the current that flows in the respective branches to a value of some hundreds of nanoamps. The second voltage comparator **150** further comprises a buffer **158**, having the function of rectifier, electrically coupled to the drain terminal of the transistor **153**, for collecting the signal that flows in the branch that includes the generator **155** and the transistor **153** and for generating at output a control signal S_{CTR_AR} that is correlated to the signal that it receives at input but has faster rising and falling edges. The buffer **158** is thus configured to receive at input a signal that varies in the range 0-V_{DD} and generating at output a signal having a low logic value "0" if the input is below a predefined threshold (said threshold is equal to $V_{DD}/2$), or else a high logic value "1" if the input is above the predefined threshold. Preferably, the buffer **158** has a threshold with hysteresis.

[0098] According to one embodiment, the current mirror **151** may have a mirror ratio 10:2 in order to speed up the state transitions of the signal set at input to the buffer **158**. However, a current mirror **151** with a different ratio, for example 1:1, may be used.

[0099] In use, when the anti-oscillation switch **58** is ON, an inverse current $-I_L$ flows in the inductor **18** (from the terminal **18***b* to the terminal **18***a*), generating a negative voltage on the terminal **18***b*, equal to $-I_L \cdot R_{AR}$, where R_{AR} is the ON-resistance of the anti-oscillation switch **58**. The output of the buffer **158** assumes a low logic value "0" as long as a negative voltage remains on the terminal **18***b* and then goes to a high logic value "1" when the voltage on the terminal **18***b* reaches the predefined threshold value.

[0100] FIG. **9** shows the first and second driving circuits **25**, **26** connected to the main bridge **12**, according to one embodiment of the present invention.

[0101] With reference to the high-side switch **13**, the first driving circuit **25** coupled thereto is configured to drive the high-side switch **13** alternatively into an open state, in which the high-side switch **13** does not conduct current and into a closed state, in which the high-side switch **13** conducts current. With reference to a high-side switch **13** of a n-channel MOSFET type, the first driving circuit **25** is configured to appropriately bias the gate terminal G of the high-side switch **13** in such a way that, when it is necessary to close the high-side switch **13**, the voltage between the source terminal S and the gate terminal G will be higher than the threshold voltage of turning-on of the high-side switch **13**. For this purpose, the first driving circuit **25** comprises a bootstrap circuit. In detail, the first driving circuit **25** is coupled to: the supply terminal **80**, at a voltage V_{DD}; a recharge switch **30** (in particular, FIG. **9** shows a MOSFET **30'**, of a p type, with integrated diode **30"** having the drain terminal D connected to the supply terminal **80**); and a capacitor **29**, having a capacitance C_{BOOT} comprised between approximately 200 pF and approximately 700 pF, for example approximately 400 pF, connected between the source terminal S of the recharge switch **30** and the terminal **18***a* of the inductor **18**.

[0102] The recharge switch **30** is configured to couple the capacitor **29** to the supply terminal **80** for charging the capacitor **29** by the voltage V_{DD} and, alternatively, to uncouple the capacitor **29** from the supply terminal **80**. The first driving circuit **25** further comprises additional switches **32** and **33**. The switch **32** is connected between the source terminal S and the gate terminal G of the recharge switch **30**, whereas the switch **33** is connected between the gate terminal G of the recharge switch **30** and the ground reference terminal GND.

[0103] The switches **32** and **33** are, for example, MOSFETs driven in conduction and inhibition by a respective control signal ϕ_A and ϕ_B, applied to the gate terminal of the respective switch **32**, **33**. The control signals ϕ_A and ϕ_{dB} are generated by a logic external to the first driving circuit **25** for implementing proper operation of the circuit **25** during the steps described with reference to FIGS. **3**A-**3**D and **4**A-**4**D.

[0104] The high-side switch **13** and the low-side switch **14** have their gate terminals G connected to a respective driving device **34**, **35**. For example, the first and second driving devices **34**, **35** are each formed by a chain of "m" inverters, where "m" is an even number.

[0105] With reference to the driving device **34**, the latter comprises a first supply input **34***a* and a second supply input **34***b*, which are connected to a floating supply, between V_P and V_{BOOT} (where V_P is the voltage on the terminal **18***a* of the inductor **18**), for generating at output a signal V_{HS} configured to drive (open/close) the high-side transistor **13** on the basis of a signal V_{HS}' that it receives at input. Since the driving device **34** is supplied at a voltage comprised between V_P and V_{BOOT}, in use it is able to generate a driving voltage on the gate terminal of the high-side switch **13** higher than the voltage V_P applied to the source terminal S (in particular higher than the threshold voltage of turning-on of the high-side switch **13**); see, for example, in FIG. **10**C, the voltage V_X.

[0106] With reference to the driving device **35**, the latter receives at input a signal V_{LS}' (which is also generated by the control logic) and generates at output a signal V_{LS} configured to drive (open/close) the low-side transistor **14**. Since the low-side transistor **14** has its source terminal S connected to the ground reference GND, it is not necessary to provide a supply circuit similar to the one described with reference to the driving device **34** that will guarantee voltages on the gate terminal G that vary as a function of the voltage assumed by the source terminal S.

[0107] FIG. **10**A shows, as a function of time t (axis of the abscissae), the timing diagram of the signals V_{HS}' and V_{LS}' applied, respectively, to the driving device **34** and **35**. FIG. **10**B shows, using the same time scale as that of FIG. **10**A, the timing diagram of the control signals ϕ_A and ϕ_B, which are configured to drive in opening and closing the switches **32** and **33** of FIG. **9**. FIG. **10**C shows, using the same scale time as that of FIGS. **10**A and **10**B, the plot of the voltage signal V_{BOOT} and of the voltage signal V.

[0108] When the low-side switch **14** is closed (signal V_{LS}' high), the terminal **18***a* is connected to the ground reference voltage GND and the signal V_P is thus at the ground reference voltage GND (e.g., approximately 0 V).

[0109] During this time interval the capacitor **29** is charged by the voltage V_{DD}. This is made possible by connecting the capacitor **29** to the supply terminal **25***a* by closing the switch **30** (the switch **33** is closed and the switch **32** is open). As has already been said, during the time interval in which the low-side switch **14** is closed, the inductor **18** is connected to ground GND and a discharge current flows through the low-side switch **14**.

[0110] When it is necessary the recharge the inductor **18**, the switch **30** is opened (by opening the switch **33** and closing the switch **32**). Preferably, the low-side switch **14** is opened prior to closing of the high-side switch **13** in order to prevent phenomena of cross conduction. The capacitor **29**, which had previously been charged, maintains the charge stored and supplies the driving device **34**, which, in turn, biases the control terminal of the high-side switch **13**, driving it into conduction. The supply terminal **12***a* is thus connected to the terminal **18***a* of the inductor **18**, enabling supply of the electrical load as explained previously.

[0111] From FIGS. **10**A and **10**B it may be noted that the signal V_{LS}' is at a high value when the control signal ϕ_A is at a high value (switch **32** open) and the signal V_{HS}' is at a low value. During this interval, the capacitor **29** is recharged. When the signal V_{LS}' drops, also the control signal ϕ_A drops (closing the switch **32**). At the same time, the control signal ϕ_B rises (opening the switch **33**). After a certain time interval, also the signal V_{HS}' rises. Between the falling edge of the signal V_{LS}' and the rising edge of the signal V_{HS}' a dead time is in fact envisaged to prevent phenomena of cross conduction.

[0112] With closing of the high-side switch **13** the voltage V_P rises to the value V_{IN}.

[0113] In this step, in which the switch **33** is open and the switch **32** is closed, the capacitor **29** is in a charged condition ($V_{BOOT} \approx V_{DD}$). With turning-on of the high-side switch **13**, the node at V_P rises in voltage and, as a result of the (capacitive) bootstrap, also the voltage on the opposite plate of the capacitor **29** (bootstrap capacitor) rises. The voltage V_X across the capacitor **29** is kept substantially constant (but for minor losses). In this way, during the step of turning-on of the high-side switch **13**, the driving device **34** is supplied and is thus able to turn on the high-side switch **13**.

[0114] The voltage drop V_X across the capacitor **29** when the high-side switch **13** is closed is given by

$$V_X = V_{DD}\left(1 - \frac{C_{GS}}{C_{BOOT} + C_{GS}}\right)$$

[0115] where C_{GS} is the capacitance between the gate terminal and the source terminal of the high-side transistor **13**.

[0116] The embodiment of FIG. **9** enables minimization of the area required for the ensemble formed by the high-side switch **13** and the first driving circuit **25**, maintaining good characteristics of performance as regards the ON-resistance (R_{ON}) and enabling a complete integration of the first driving circuit **25** and of the high-side switch **13** without the need to use external components.

[0117] To return to FIG. **2**, the converter **10** further comprises, according to one embodiment, a hysteretic comparator **87** (of a per se known type, in which the hysteresis is defined between a first threshold V_{th}^- and a second threshold V_{th}^+, with $V_{th}^- < V_{th}^+$). The hysteretic comparator **87** comprises an inverting input configured to receive the output signal V_{OUT} and a non-inverting input configured to receive a reference signal V_{ref_c}.

[0118] The reference signal V_{ref_c} is a bandgap reference, independent of the supply voltage and of the temperature.

[0119] The output of the hysteretic comparator **87** indicates the level of voltage assumed by the output signal V_{OUT}. The signal V_{comp} generated at output by the hysteretic comparator **87** is used during the steps of supply of the electrical load **20** for detecting whether the electrical load **20** needs to be supplied or not. The electrical load **20**, in fact, is supplied (recharged) only when necessary, i.e., when the output voltage signal V_{OUT} drops below the threshold defined by the reference signal V_{ref_c}.

[0120] The hysteretic comparator **87** has two possible output logic levels, in particular the ground reference value GND (or equivalent) and the value of the supply signal (V_{DD}). When the output signal V_{comp} of the hysteretic comparator **87** is equal to V_{DD}, then the output voltage V_{OUT} has dropped below the reference V_{ref_c} and the respective electrical load **20** must be supplied. When the output signal V_{comp} of the hysteretic comparator **87** is equal to the ground reference GND, then the output voltage V_{OUT} is higher than the reference V_{ref_c} and the electrical load **20** does not have be supplied. The characteristic of the comparator **87** is centered around the reference V_{ref_c} and is the classic hysteretic characteristic (indicatively with thresholds $V_{th}^+ > V_{ref\_c}$ and $V_{th}^- < V_{ref\_c}$).

[0121] In order to decide whether the electrical load **20** needs to be supplied, the output signal V_{OUT} is monitored in a continuous way or at regular intervals. If the output signal V_{OUT} drops below the threshold V_{th}^- of the hysteretic comparator **87**, the electrical load **20** is supplied and charged. The charging step is not carried out if the output signal V_{OUT} exceeds the threshold V_{th}^+. Each comparator **87** is configured to have the hysteresis equal to the maximum value of ripple tolerated by the specific application, for example approximately 10-50 mV.

[0122] In order to manage the operating steps of the converter **10**, i.e., the steps described with reference to FIGS. **3**A-**3**D and **4**A-**4**D, for passage from one step to the next and for verifying whether the electrical load **20** needs to be supplied or not, the converter **10** further comprises a control logic **85**, for example an integrated digital controller, as shown schematically in FIG. **11**. The control logic **85** carries out generation of the signals to turn-on/turn-off the high-side switch **13** and the low-side switch **14** and generation of the control signals of the switches **57** (FIG. **5**) and **103** (FIG. **6**), of the anti-oscillation switch **58**, of the switch **23** and of the enable signals S_{EN_ON} and S_{EN_OFF} in order to implement the operating steps described previously.

[0123] In greater detail, the control logic **85** implements the steps of the method for operating the converter **10** according to FIG. **12**, functioning as a state machine.

[0124] With joint reference to FIGS. **11** and **12**, during an initial step (step S**1**), the control logic **85** verifies in the first place whether the electrical load **20** needs to be supplied. In this step, the low-side switch **14** and the charge switch **23** are ON, whereas all the other switches are OFF.

[0125] The above verification step is carried out, as has been said, on the basis of the value assumed by the signal V_{comp} generated by the comparator **87**. If the electrical load **20** does not have to be supplied (output NO from step S**1**), the logic returns into the mode of observation of the signal V_{comp} and waits for the electrical load **20** to require charging. Otherwise, if the electrical load **20** is to be supplied (output YES from step S**1**) control passes to step S**2**.

[0126] In the case where the electrical load **20** does not need to be supplied, the remaining steps of FIG. **12** are not carried out. In this case, it is further envisaged that the control logic **85** will send a signal for selective turning-off of portions of the converter **10** so that a further energy saving will be guaranteed and thus an increase of efficiency of the converter itself. In detail, the control logic electrically uncouples the adaptive control blocks **51** and **101** by closing (turning off) the switches **57** shown in FIGS. **5** and **103** shown in FIG. **6**.

[0127] In this initial resting step, the high-side switch **13**, as likewise the coupling switch **22** and the anti-oscillation switch **58**, are open (i.e., in the case of a MOSFET, OFF). Only the low-side switch **14** and the charge switch **23** are ON, thus connecting both of the terminals of the inductor **18** to the reference terminal GND.

[0128] Passage to the subsequent steps, however, determines closing of the switches **57** and **103** so that the adaptive-control blocks **51** and **101** are supplied and immediately available for operation.

[0129] In step S**2**, the control logic **85** opens the low-side switch **14** (by the signal V_{LS}) and then closes the high-side switch **13** (i.e., it turns on the transistor **13** by the signal V_{HS}). This operation is carried out by generating the signals ϕ_A and ϕ_B of FIG. **9** to close the switch **32** and open the switch **33**. The signals ϕ_A and ϕ_B generated by the control logic **85** are voltage signals that have an amplitude appropriate for controlling the switches **32** and **33** (for example, in the case of switches **32**, **33** of a MOSFET type, the voltage signals generated by the driving devices as a function of the signals ϕ_A and ϕ_B are such as to drive the respective MOSFETs into conduction by generating a gate-to-source voltage higher than the threshold voltage of the respective MOSFET).

[0130] As described previously, following upon generation of the signals Φ_A and Φ_B, the signal V_{HS}' that enables closing of the high-side switch **13** via the driving device **34** is asserted. Likewise, together with the high-side switch **13**, also the charge switch **23** remains closed by the control logic **85**, by the signal Φ_c. The terminal **18***b* is thus coupled to the ground reference terminal GND. A current may thus start to flow through the inductor **18**, from the terminal **12***a* to the terminal GND.

[0131] As soon as the high-side switch **13** conducts, the voltage on the terminal **18***a* starts to rise, locking to the value V_{IN}. At the same time, owing to the capacitive effect, since the difference of potential across the capacitor **29** remains unvaried, the voltage V_{BOOT} rises, enabling the driving device **34** to generate a signal V_{HS} such as to keep the high-side switch **13** in conduction. The inductor **18** may thus be charged.

[0132] As shown in step S**3**, which is temporally simultaneous to step S**2**, or immediately subsequent to step S**2**, monitoring of the time interval T_{ON} starts by activating the first adaptive control block **51**. For this purpose, the enable signal S_{EN_ON} assumes a low logic value "0", such as to drive the transistor **76** of FIG. **5** into conduction, thus charging the capacitor **84**.

[0133] The control logic **85** (step S**4**) immediately starts monitoring of the comparison signal S_{CMP_ON} generated at output by the adaptive control block **51**. During the interval T_{ON}, the comparison signal S_{CMP_ON} has a high logic value "1"; passage of the value of the comparison signal S_{CMP_ON} to a low logic value "0" is interpreted by the control logic **85** as the end of the interval T_{ON}. As long as the comparison signal S_{CMP_ON} has a high logic value "1" (output NO from step S**4**), the control logic **85** remains in step S**4** (i.e., in the condition of FIGS. **3**A, **4**A). Otherwise (output YES from step S**4**), control passes to step S**5**.

[0134] Entry into step S**5** determines transition from the operating condition of FIGS. **3**A, **4**A to the operating condition of FIGS. **3**B, **4**B.

[0135] In step S**5**, the high-side switch **13** is opened (the control logic **85** generates an appropriate signal V_{HS}' such that, via the driving device **34**, the high-side switch **13** is driven into the open state—transistor **13** OFF). Instead, the low-side switch **14** is closed (transistor **14** ON). Preferably, the control logic introduces a delay (dead time) between opening of the high-side switch **13** and closing of the low-side switch **14**.

[0136] The low-side switch **14** is driven into the closed state by generating the signal V_{LS}', which is applied, via the driving device **35**, to the control terminal of the low-side switch **14**. The signals V_{LS}' and V_{HS}' are, for example, CMOS logic signals of amplitude equal to V_{DD}.

[0137] Simultaneously with opening of the high-side switch **13** and, preferably, prior to closing of the low-side switch **14**, the charge switch **23** is opened, thus uncoupling the terminal **18***b* from the ground reference GND.

[0138] The control logic **85** further generates, following upon (or simultaneously with) closing of the low-side switch **14**, a signal Φ_{out} for driving the coupling switch **22** (possibly via a buffer **160** for matching the voltage level of the signal Φ_{out} to the voltage level accepted at input by the gate terminal of the transistor that constitutes the coupling switch **22**). The signal Φ_{out} drives the coupling switch **22** into conduction, thus connecting the inductor **18** to the electrical load **20**. The inductor **18** is thus discharged in the active mode.

[0139] During step S**5** of discharge of the inductor **18**, the time interval T_{OFF} is monitored by the control logic **85** by operating the second adaptive control block **101**. For this purpose (step S**6**), simultaneously with (or immediately after) closing of the switch **22**, the control logic **85** sets the signal S_{EN_OFF} to the low logic value "0", thus turning on the transistor **126** and charging the capacitor **134**.

[0140] The control logic **85** monitors (step S**7**) constantly, or at regular intervals, the value of the comparison signal S_{CMP_OFF} generated at output by the adaptive control block **101** in order to identify the end of the interval T_{OFF}. A high logic value "1" of the comparison signal S_{CMP_OFF} (output NO from step S**7**) maintains the control logic **85** in a wait state. The variation of the comparison signal S_{CMP_OFF} from the high logic value "1" to the low logic value "0" (output YES from step S**7**) determines the end of the interval T_{OFF} and passage from the operating condition of FIGS. **3**B, **4**B to that of FIGS. **3**C, **4**C (passive supply of the load **20**). This passage corresponds to the transition from step S**7** to step S**8**.

[0141] In step S**8**, the control logic **85** sets the enable signal S_{EN_OFF} to the high logic value "1", thus discharging the capacitor **134**. Active supply of the load **20** is concluded and consequently the control logic **85** turns off the coupling switch **22** by the signal Φ_{out}.

[0142] Control then passes to step S**9**, in which the control logic **85** monitors the value of the control signal S_{CTR_PT}: as long as the logic value of the control signal S_{CTR_PT} is low, "0", the control logic **85** remains in step S**9**. Instead, when the logic value of the control signal S_{CTR_PT} passes to the high value "1" (output YES from step S**9**), then the control logic **85** drives the anti-oscillation switch **58** into the closed state (step S**10**) by generating an appropriate signal Φ_a. This step corresponds to the operating condition of FIGS. **3**D, **4**D.

[0143] After step S**10**, the control logic **85** goes immediately to the next step S**11**, where the control signal S_{CTR_AR} generated by the block **150** is acquired and monitored. Transition of the logic state of the control signal S_{CTR_AR}, as already described with reference to FIG. **8**, determines the end of step S**11**. The operating condition of FIGS. **3**D, **4**D terminates with opening of the anti-oscillation switch **58**. The steps of supply of the load **20** have terminated and the control logic **85** returns to step S**1**, activating the charge switch **23** and the low-side switch **14** to maintain the terminals **18***a* and **18***b* of the inductor **18** at the reference terminal GND.

[0144] Steps S**2**-S**12** are, for example, managed on the basis of a timing supplied by a clock signal (CLK) generated outside the circuit converter **10** or generated by a clock circuit of an integrated type. The clock frequency CLK is, for example comprised between approximately 100 kHz and approximately 400 kHz, for example approximately 230 kHz. The clock signal CLK may be generated by an oscillator circuit of a known type. On each rising edge of the clock signal a time slot is generated for the respective step of supply of the load **20**, having a duration equal to the period of oscillation of the clock signal. When the output does not need to be recharged, the clock is set in sleep, i.e., low-consumption, mode and so the time slot is generated. As soon as the output needs to be charged (as determined by step **S1**), the state machine wakes up and restarts with generation of the time slots for supplying the electrical loads that need to be supplied.

[0145] It is evident that use of a clock signal CLK is not necessary and may be omitted. In this case, the control logic starts the operations of steps **S1**-**S12** immediately according to the need and the entire sequence is carried out without a timing supplied by a clock.

[0146] FIG. **13** shows an energy harvesting system **200** comprising the converter **10**. The energy harvesting system **200** is similar to the energy harvesting system **1** of FIG. **1** (elements that are in common are designated by the same reference numbers) and is not described any further herein.

[0147] The transducer **2** may be chosen in the group comprising: electrochemical transducers (configured to convert chemical energy into an electrical signal), electromechanical transducers (configured to convert mechanical energy into an electrical signal), electroacoustic transducers (configured to convert variations of pressure into an electrical signal), electromagnetic transducers (configured to convert a magnetic field into an electrical signal), photoelectric transducers (configured to convert light energy into an electrical signal), electrostatic transducers and thermoelectric transducers.

[0148] The converter **10** is connected to the output of the scavenging interface **4**. The energy stored in the storage capacitor of the scavenging interface **4** (which is of a per se known type) supplies the converter **10**. The input voltage of the converter **10** is thus the voltage produced by the scavenging interface **4**.

[0149] According to one application of the energy harvesting system **200**, generation of electrical energy exploits the mechanical energy produced by an individual when he is walking or running. In this case, the energy harvesting system **200** is set inside the shoes of the individual in question (for example, inside the sole). In systems aimed at fitness, where counting of the steps is of particular interest, it is useful to harvest energy from the vibrations induced by walking/running to be able to supply without the use of a battery acceleration sensors and/or wireless transmitters (for example, RFID, BT, etc.), which are able to communicate with cellphones, music-player devices, or with any other apparatus concerned with providing information on the steps performed.

[0150] A further application of the energy harvesting system **200** envisages harvesting the thermal energy supplied by a source of thermal energy, e.g., through a thermoelectric transducer.

[0151] From an examination of the characteristics of the invention obtained according to the present disclosure the advantages that it affords are evident.

[0152] In particular, the converter **10** according to the present invention enables electrical supply of loads that require low supply voltages, with high efficiency using a single inductor **18**.

[0153] Further, the converter **10** may be completely integrated in an energy harvesting system **200**, which must typically guarantee a high level of performance and resistance in regard to stresses. The high integratability is enabled by the presence of the bootstrap network to turn on the high-side switch, which is of a type set inside (integrated in) the converter **10**.

[0154] Further, the adaptive control circuit **51** enables the converter **10** to be operated in a constant-peak-current mode over a wide range of input voltages, enabling compensation of the time interval T_{ON} for different values of the input voltage V_{IN}. Likewise, the adaptive control circuit **101**, together with the control logic of the coupling switch **22**, enables supply of the electrical load in a precise way, discharging the inductor completely at each step of supply of the load **20**, without absorbing current from the load **20** itself and using, during the final discharging step, passive elements, thus minimizing the consumption of current by the converter **10**.

[0155] Finally, it is clear that modifications and variations may be made to what has been described and illustrated herein, without thereby departing from the scope of the present invention, as defined in the annexed claims.

[0156] In particular, according to a different embodiment (not described in detail), the converter **10** may be of a SIMO (single-inductor multiple-output) type, having a plurality of outputs configured to supply a respective plurality of electrical loads. In this case, a respective plurality of coupling switches **22** is present, each of which is configured to couple the terminal **18*b*** of the inductor **18** with a respective electrical load. For supplying the plurality of electrical loads it is possible to use, for example, a time-multiplexing technique, for example of the type described in the patent application No. EP 2518878.

[0157] Further, the present invention is irrespective of the circuit architecture of the DC-DC converter. It may be applied to any DC-DC converter of a SIMO or SISO type, for example of a buck type, of a boost type, or else of a buck-boost type.

[0158] Only in the case of a DC-DC converter of a buck type does the operating condition of charging (interval T_{ON}) have a duration calculated as a function of the value assumed by the input supply signal V_{IN} minus the output signal V_{OUT}.

[0159] Likewise, only in the case of a DC-DC converter of a boost type does the time interval of active discharge of the inductor (interval T_{OFF}) have a duration $t_{1a}-t_2$ calculated as a function of the value assumed by the output supply signal V_{OUT} minus the input signal V_{IN}.

What is claimed is:

1. A method for operating a converter including an inductor receiving an input supply signal and supplying an electrical load with an output supply signal, the method comprising:

charging the inductor by storing electrical charge until a maximum pre-set current value is reached, during a first operating condition in which said electrical load is not supplied;

actively supplying said electrical load by the electrical charge stored in the inductor during a first time interval of a second operating condition by partially discharging the inductor until an intermediate value of residual charge stored in the inductor is reached; and

passively supplying said electrical load by said residual charge stored in the inductor during a second time interval, subsequent to the first time interval, of the second operating condition.

2. The method according to claim **1**, wherein said first operating condition has a duration calculated as a function of the value assumed by the input supply signal.

3. The method according to claim **1**, wherein the first time interval of the second operating condition has a duration calculated as a function of the value assumed by the output supply signal.

4. The method according to claim **1**, wherein the converter further comprises a charge-transfer switch and a charge-transfer diode coupled together in series between the inductor and the electrical load, the charge-transfer switch being configured to operate in an active operating state, in which the charge-transfer switch is closed and the electrical load is supplied through the charge-transfer switch and in a passive operating state, in which the charge-transfer switch is open and the electrical load is supplied through the charge-transfer diode, wherein actively supplying includes driving the charge-transfer switch into the active operating state and said step of passive supply of said electrical load including driving the charge-transfer switch into the passive operating state.

5. The method according to claim **1**, wherein the converter further comprises a first adaptive control circuit including a first electrical-charge-storage device and a first signal comparator and a second adaptive control circuit including a second electrical-charge-storage device and a second signal comparator, the method further comprising:

charging the first electrical-charge-storage device, during the first operating condition, by a first charging signal correlated to the input supply signal;

generating, by the first signal comparator, a first control signal when the charge stored in the first electrical-charge-storage device reaches a first predefined threshold;

charging the second electrical-charge-storage device, during the second operating condition, by a second charging signal correlated to the output supply signal;

generating, by the second signal comparator, a second control signal when the charge stored in the second electrical-charge-storage device reaches a second predefined threshold, wherein said first predefined threshold indicates a maximum value that the electrical charge stored in said inductor may assume during the first operating condition and the second predefined threshold indicates the intermediate value that the electrical charge stored in said inductor may assume during the second operating condition.

6. The method according to claim **5**, wherein the converter further comprises a charge switch and a discharge switch, coupled in series to one another between a source of electrical energy that supplies the input supply signal and a reference terminal, said inductor having a first conduction terminal and a second conduction terminal, the first conduction terminal being electrically coupled between the charge switch and the discharge switch and the second conduction terminal being electrically coupled to the ground reference terminal via a coupling switch,

the method further comprising:

turning on the charge switch, by turning on the coupling switch and turning off the discharge switch, in order to charge the inductor during the first operating condition;

passing from the first operating condition to the second operating condition upon reception of the first control signal, by turning off the charge switch and the coupling switch, by turning on the discharge switch and by driving the charge-transfer switch into the active operating state; and

driving a transition from the active operating state to the passive operating state of the charge-transfer switch upon reception of the second control signal.

7. A converter for supplying an electrical load by an output supply signal, comprising:

an input stage operatively coupled between a source of electrical energy configured to supply an input supply signal and a reference terminal;

an inductor having a first conduction terminal and a second conduction terminal, the first conduction terminal being electrically coupled to the input stage and the second conduction terminal being selectively electrically coupled to the ground reference terminal; and

a charge-transfer switch and a charge-transfer diode coupled in series to one another between the second conduction terminal of the inductor and an electrical load, the charge-transfer switch being configured to electrically uncouple the second conduction terminal from the electrical load during a first operating condition in which the inductor is charged by the input supply signal and electrically couple the second conduction terminal to the electrical load during a second operating condition in which the electrical load is supplied,

wherein the charge-transfer switch is configured to operate in an active operating state in which the charge-transfer switch is closed and the electrical load is supplied through the charge-transfer switch and in a passive operating state in which the charge-transfer switch is open and the electrical load is supplied through the charge-transfer diode.

8. The converter of claim **7**, further comprising:

a first adaptive control circuit including a first electrical-charge-storage device operatively coupled to the source of electrical energy and a first signal comparator, wherein the first electrical-charge-storage device is configured to be charged, during the first operating condition, by a first charging signal correlated to the input supply signal and the first signal comparator is configured to generate a first control signal when the charge stored in the first electrical-charge-storage device reaches a first predefined threshold; and

a second adaptive control circuit including a second electrical-charge-storage device operatively coupled to the electrical load and a second signal comparator, wherein the second electrical-charge-storage device is configured to be charged, during the second operating condition, by a second charging signal correlated to the output supply signal and the second signal comparator is configured to generate a second control signal when the charge stored in the second electrical-charge-storage device reaches a second predefined threshold,

wherein said first predefined threshold indicates a maximum value that the electrical charge stored in said inductor may assume during the first operating condition and the second predefined threshold indicates an intermediate value, comprised between a zero value and said maximum value, that the electrical charge stored in said inductor may assume during the second operating condition.

9. The converter of claim **7**, further comprising a control logic configured to:

control the input stage for coupling, during the first operating condition, the inductor to the source of electrical energy thus charging the inductor;

pass from the first operating condition to the second operating condition upon reception of the first control signal by driving the input stage for uncoupling the inductor from the source of electrical energy and driving the charge-transfer switch into the active operating state; and

drive a transition from the active operating state to the passive operating state of the charge-transfer switch upon reception of the second control signal.

10. The converter according to claim **7**, wherein said charge-transfer switch is a transistor having an integrated diode, said active operating state corresponding to a flow of electrical charge through the channel of said transistor obtained by turning on the transistor and said passive operating state corresponding to a flow of electrical charge through the integrated diode when the transistor is OFF.

11. The converter according to claim **8**, wherein said first adaptive control circuit further comprises:

a first voltage-to-current converter, electrically coupled to the source of electrical energy for receiving the input supply signal and configured to generate a first intermediate current signal correlated to the input supply signal; and

a first amplifier, configured to receive the first intermediate current signal and generating said first charging signal by amplifying the first intermediate current signal.

12. The converter according to claim **11**, wherein said first adaptive control circuit further comprises a first enabling switch configured to receive the first charging signal and supply the first charging signal to said first electrical-charge-storage device and a second enabling switch configured to couple the first electrical-charge-storage device to a ground reference terminal, and

control logic being configured to:

turn on said first enabling switch and turning off said second enabling switch, during the first operating condition, thus charging the first electrical-charge-storage device; and

turn off said first enabling switch and turning on said second enabling switch, outside the first operating condition, thus discharging the first electrical-charge-storage device.

13. The converter according to claim **8**, wherein said second adaptive control circuit further comprises:

a second voltage-to-current converter configured to receive the output supply signal and configured to generate a second intermediate current signal correlated to the output supply signal; and

a second amplifier configured to receive the second intermediate current signal and generating said second charging signal by amplifying the second intermediate current signal.

14. The converter according to claim **13**, wherein said second adaptive control circuit further comprises a third enabling switch configured to receive the second charging signal and supply the second charging signal to said second electrical-charge-storage device and a fourth enabling switch configured to couple the second electrical-charge-storage device to a ground reference terminal, and

control logic configured to:

turn on said third enabling switch and turning off said fourth enabling switch, during the second operating condition, thus charging the second electrical-charge-storage device; and

turn off said third enabling switch and turning on said fourth enabling switch, outside of the second operating condition, thus discharging the second electrical-charge-storage device.

15. The converter according to claim **7**, further comprising an output comparator configured to acquire said output supply signal and comparing the output supply signal alternatively with a lower supply-threshold value and with an upper supply-threshold value to generate an output comparison signal, and

control logic configured to acquire the output comparison signal and turn off the converter if the output supply signal reaches the upper supply-threshold value and, alternatively, re-activating the converter, by coupling the inductor to the source of electrical energy, if the output supply signal reaches the lower supply-threshold value.

16. The converter according to claim **7**, further comprising:

a monitoring circuit configured to monitor the electrical charge transferred through the charge-transfer switch during the passive operating state of the charge-transfer switch and to generate a monitoring signal indicating a substantially zero transfer of electrical charge to the electrical load;

an anti-oscillation switch connected between the second terminal of the inductor and the ground reference terminal, and

control logic configured to turn on the discharge switch and the anti-oscillation switch upon reception of the monitoring signal.

17. The converter according to claim **7**, wherein the input stage comprises a charge switch and a discharge switch connected together in series, the charge switch being coupled to the source of electrical energy and the discharge switch being coupled to the reference terminal, and

wherein the first conduction terminal of the inductor is electrically coupled between the charge switch and the discharge switch and the second conduction terminal is electrically coupled to the ground reference terminal via a coupling switch.

18. An energy harvesting system configured to supply at least one electrical load (**20**), comprising:

a transducer configured to convert energy coming from an energy source external to said system into an AC/DC electrical signal;

a rectifier circuit configured to receive at input the AC/DC electrical signal generated by the transducer and supplying at output a DC output signal;

a first storage element coupled to the rectifier circuit for receiving at input the DC output signal and configured to store electrical energy; and

a converter according to claim **7** connected between the rectifier circuit and the electrical load and configured to receive at input the DC output signal generated by the rectifier circuit and supply the electrical load.

19. A converter, comprising:

an inductor configured to receive an input supply signal and supply an electrical load with an output supply signal;

a switching circuit coupled to said inductor; and

a control circuit configured to control operation of the switching circuit to:

charge the inductor by storing electrical charge until a maximum pre-set current value is reached during a first operating condition in which said electrical load is not supplied;

actively supply said electrical load by the electrical charge stored in the inductor during a first time interval of a second operating condition that partially discharges the inductor to an intermediate value of residual charge; and

passively supply said electrical load by said residual charge stored in the inductor during a second time interval, subsequent to the first time interval, of the second operating condition.

20. The converter according to claim **19**, wherein the switching circuit comprises:

a charge-transfer switch and a charge-transfer diode coupled between the inductor and the electrical load,

wherein the control circuit closes the charge-transfer switch during active supply of electrical charge to the load, and opens the charge-transfer switch during passive supply of electrical charge to the load, wherein said passive supply is made through the charge-transfer diode.

21. The converter according to claim **19**, further comprising:

a first adaptive control circuit including a first electrical-charge-storage device and a first signal comparator; and

a second adaptive control circuit including a second electrical-charge-storage device and a second signal comparator;

wherein the first electrical-charge-storage device is charged during the first operating condition by a first charging signal correlated to the input supply signal, said the first signal comparator generating a first control signal when the charge stored in the first electrical-charge-storage device reaches a first predefined threshold; and

wherein the second electrical-charge-storage device is charged during the second operating condition by a second charging signal correlated to the output supply signal, said second signal comparator generating a second control signal when the charge stored in the second electrical-charge-storage device reaches a second predefined threshold.

22. The converter according to claim **21**, wherein said first predefined threshold indicates a maximum value that the electrical charge stored in said inductor may assume during the first operating condition and the second predefined threshold indicates the intermediate value that the electrical charge stored in said inductor may assume during the second operating condition.

23. The converter according to claim **21**, wherein the switching circuit comprises:

a charge switch;

a discharge switch;

wherein the charge and discharge switches are coupled in series to one another between a source of electrical energy that supplies the input supply signal and a reference terminal; and

a coupling switch;

wherein said inductor has a first conduction terminal and a second conduction terminal, the first conduction terminal being electrically coupled between the charge switch and the discharge switch and the second conduction terminal being electrically coupled to the ground reference terminal via said coupling switch.

24. The converter according to claim **23**, wherein the control circuit is configured to:

turn on the charge switch, turn on the coupling switch and turn off the discharge switch in order to charge the inductor during the first operating condition;

turn off the charge switch, turn off the coupling switch and turn on the discharge switch during the first time interval of the second operating condition in response to a first control signal; and

turn off the discharge switch during the second time interval of the second operating condition.

* * * * *

US 20150256097A1

(19) **United States**

(12) **Patent Application Publication**
Gudan et al.

(10) **Pub. No.: US 2015/0256097 A1**

(43) **Pub. Date: Sep. 10, 2015**

(54) SYSTEM FOR AMBIENT ENERGY HARVESTING

(71) Applicants: **Ken Gudan**, Sunnyvale, CA (US); **Sergey Chemishkian**, Menlo Park, CA (US); **Matthew S. Reynolds**, Menlo Park, CA (US); **Stewart Thomas**, Menlo Park, CA (US); **Joshua Ensworth**, Menlo Park, CA (US)

(72) Inventors: **Ken Gudan**, Sunnyvale, CA (US); **Sergey Chemishkian**, Menlo Park, CA (US); **Matthew S. Reynolds**, Menlo Park, CA (US); **Stewart Thomas**, Menlo Park, CA (US); **Joshua Ensworth**, Menlo Park, CA (US)

(21) Appl. No.: **14/198,357**

(22) Filed: **Mar. 5, 2014**

Publication Classification

(51) **Int. Cl.**
H02M 7/04 (2006.01)

(52) **U.S. Cl.**
CPC *H02M 7/04* (2013.01)

(57) ABSTRACT

A method and apparatus is disclosed herein for harvesting ambient energy. In one embodiment, an energy harvester comprises: a first RF rectifier to output a first voltage determined by rectified RF energy in response to received RF energy; a first energy reservoir coupled to the first RF rectifier to store energy at the first voltage; a DC/DC converter coupled to the first energy reservoir to convert the first voltage to a second voltage; a second reservoir coupled to the DC/DC converter to store energy at the second voltage, the second voltage being greater than the first voltage; and a third reservoir coupled to the second reservoir to receive energy transferred from the second reservoir periodically.

300

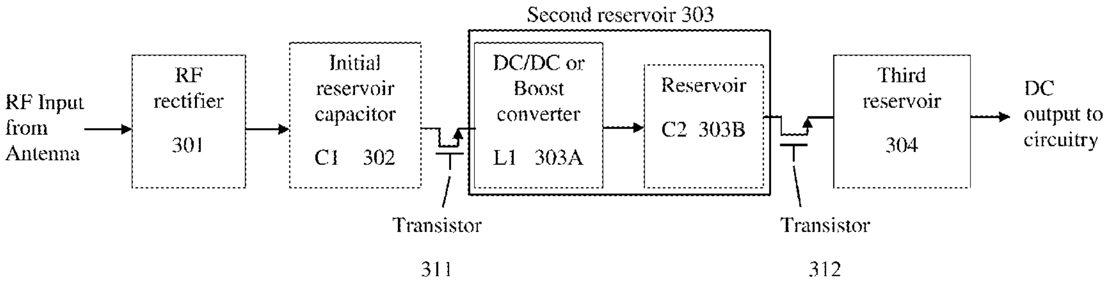
Second reservoir 303
RF Input from Antenna
RF rectifier
301
Initial reservoir capacitor
C1 302
DC/DC or Boost converter
L1 303A
Reservoir
C2 303B
Third reservoir
304
DC output to circuitry
Transistor
311
Transistor
312

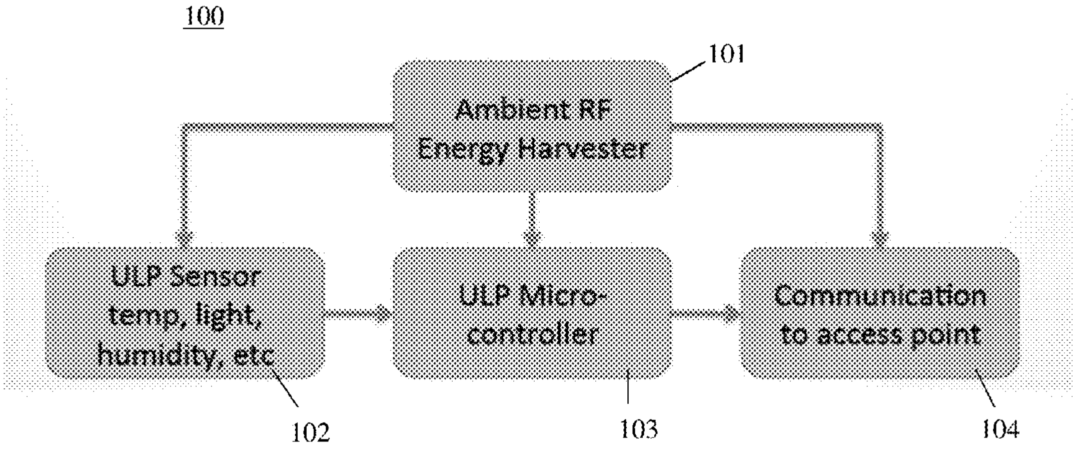
100
101
Ambient RF Energy Harvester
ULP Sensor temp, light, humidity, etc
ULP Micro-controller
Communication to access point
102
103
104

Figure 1

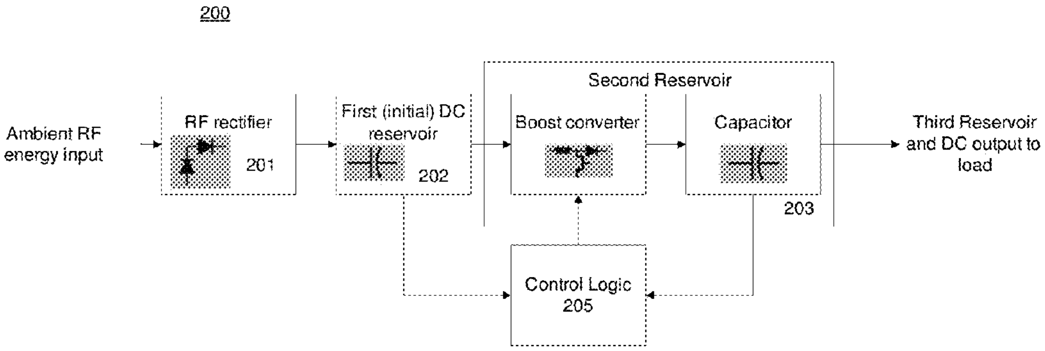
200
Ambient RF energy input
RF rectifier
201
First (initial) DC reservoir
202
Second Reservoir
Boost converter
Capacitor
203
Third Reservoir and DC output to load
Control Logic
205

Figure 2

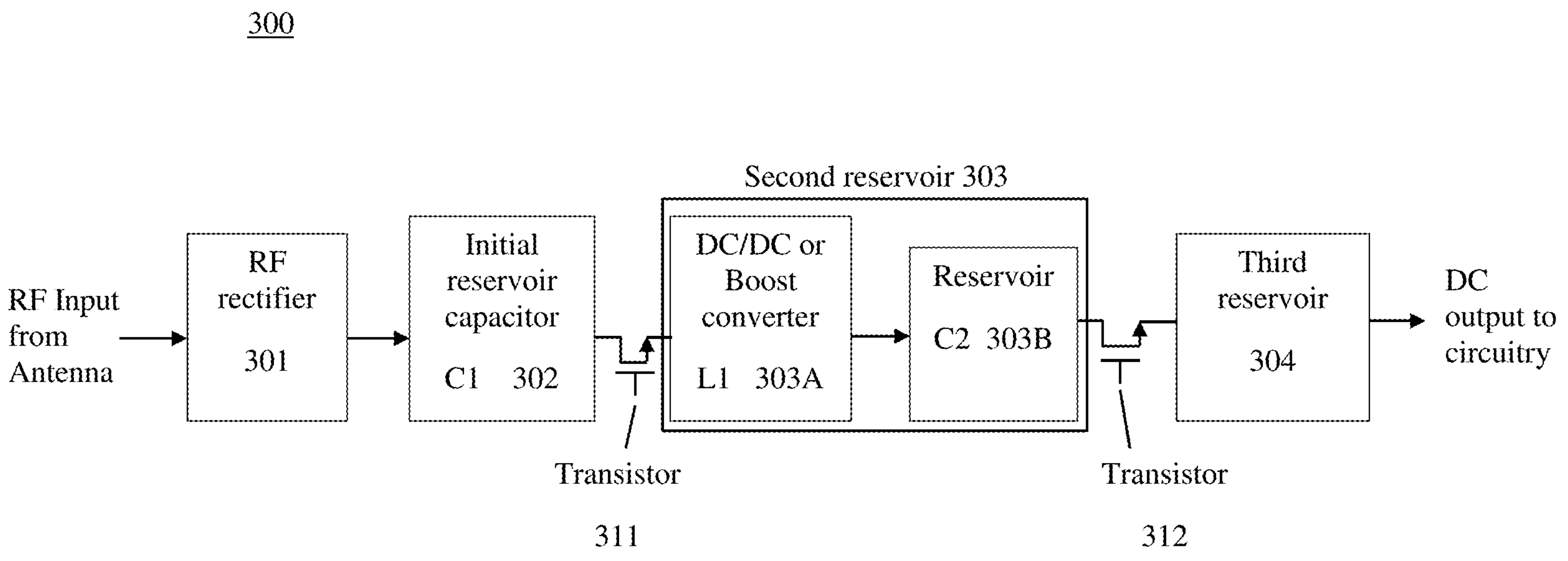
300
RF Input from Antenna
RF rectifier 301
Initial reservoir capacitor C1 302
Second reservoir 303
DC/DC or Boost converter L1 303A
Reservoir C2 303B
Third reservoir 304
DC output to circuitry
Transistor 311
Transistor 312

Figure 3

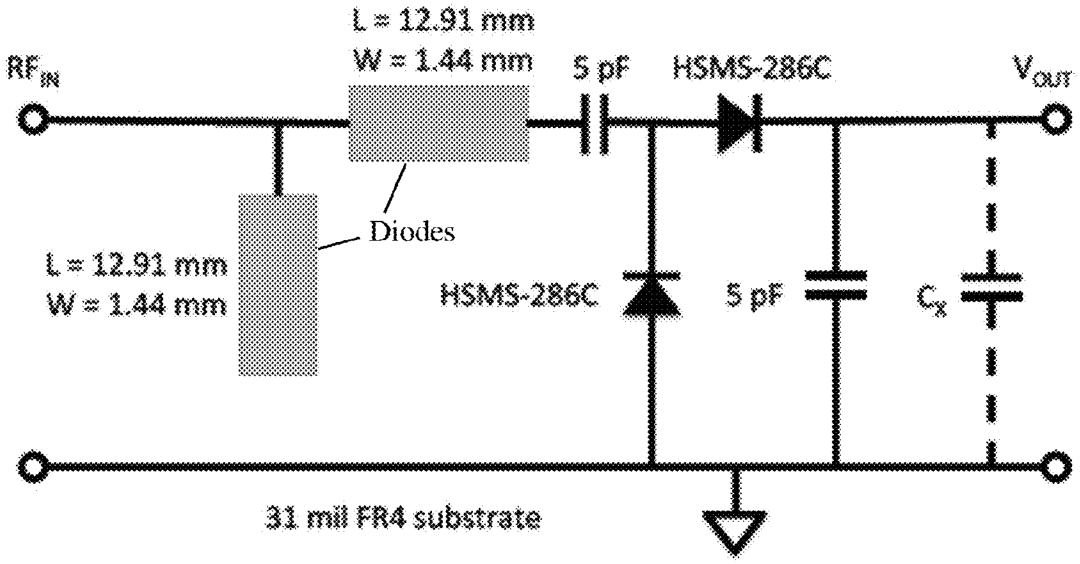
RF_IN
L = 12.91 mm
W = 1.44 mm
5 pF
HSMS-286C
V_OUT
Diodes
L = 12.91 mm
W = 1.44 mm
HSMS-286C
5 pF
C_x
31 mil FR4 substrate

Figure 4

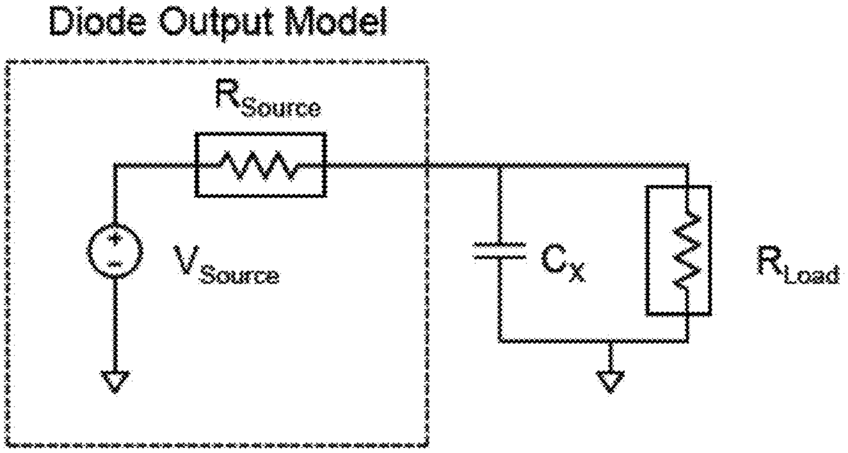
Diode Output Model
R_Source
V_Source
C_X
R_Load

Figure 5

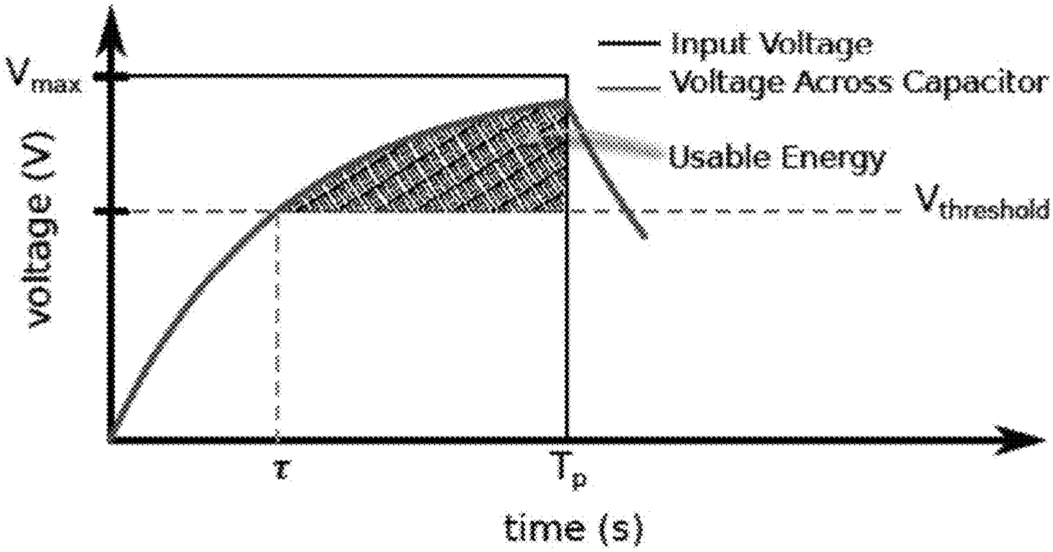

Figure 6A

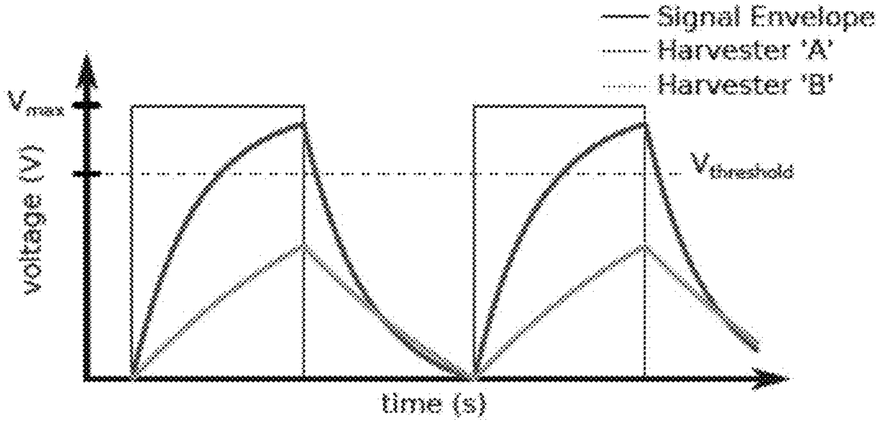

Figure 6B

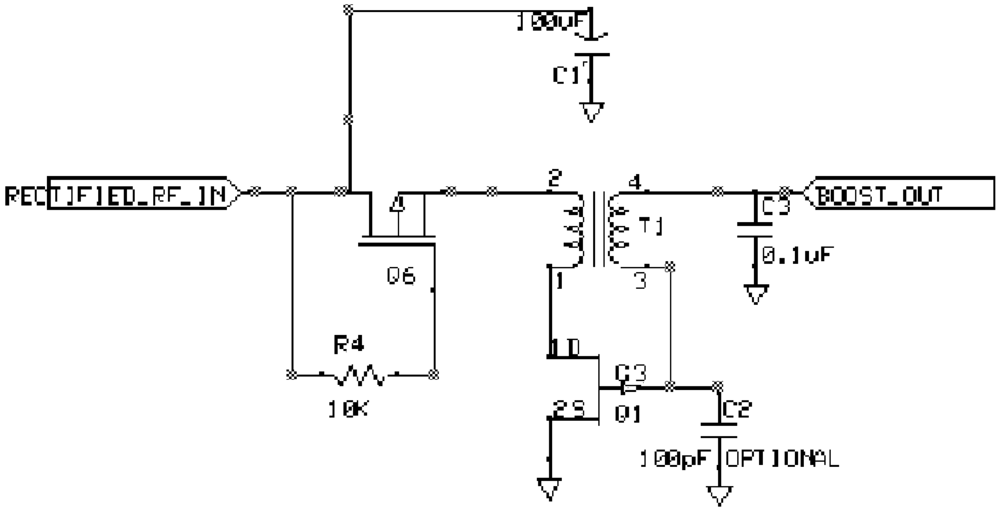
100uF
C1
RECTIFIED_RF_IN
4
Q6
R4
10K
2
4
T1
1
3
C3
BOOST_OUT
0.1uF
1 D
G 3
2 S
Q1
C2
100pF OPTIONAL

Figure 7

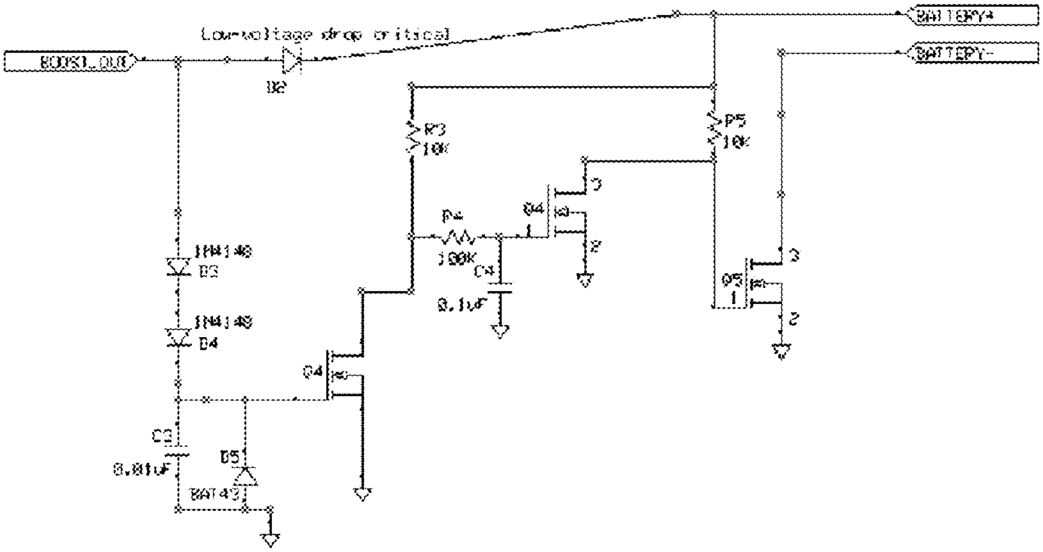
Low-voltage drop critical
BOOST_OUT
D2
BATTERY+
BATTERY-
R3
10K
R5
10K
R4
100K
C4
0.1uF
Q4
3
1
2
1N4148
D3
1N4148
D4
Q4
Q5
3
1
2
C3
0.01uF
D5
BAT43

Figure 8

SYSTEM FOR AMBIENT ENERGY HARVESTING

FIELD OF THE INVENTION

[0001] Embodiments of the present invention relate to the field of wireless power transfer; more particularly, embodiments of the present invention relate to radio-frequency (RF) energy harvesting devices that harvest ambient RF energy and store the received RF energy.

BACKGROUND OF THE INVENTION

[0002] Next generation sensor networks may be powered by energy harvesting techniques to avoid requiring battery maintenance. Energy harvesting is a process by which energy is derived from external sources (e.g., radio frequency energy, solar power, thermal energy, wind energy, salinity gradients, or kinetic energy), captured and stored.

[0003] Energy may be harvested from radio frequency signals propagating wirelessly. With RF harvesting, wireless energy comes from a radio frequency transmitting device that is some distance away from a device that harvests energy from the radio frequency transmission. Properties of an energy harvester include its ability to harvest energy efficiently from available RF signals, its ability to store the harvested energy efficiently with minimal storage loss, and its ability to make the stored energy available to meet the voltage, current, and duty cycle requirements of a desired application.

[0004] One of the more popular forms of RF used today is Wi-Fi (also referred to as IEEE 802.11a/b/g/n etc.) communications. Today, most Wi-Fi communications are in the 2.4 GHz and 5.8 GHz frequency bands and there are many local area networks that are based on Wi-Fi in which access points enable Wi-Fi clients to gain access to networks such as the Internet. Furthermore, the 2.4 GHz and 5.8 GHz bands also support other networking standards, such as Zigbee and Bluetooth, and other proprietary networks, each transmitting energy by communicating in this same frequency band. Additionally there are other frequency bands that support different communication protocols, each of which transmit energy when they are communicating. These include, for example, digital television (DTV) and Global System for Mobile Communications (GSM) signals.

[0005] In a traditional RFID-like setup the RFID readers (interrogators) usually deliver RF power as a continuous wave (CW), i.e. a sinusoidal signal of a particular frequency. In contrast to systems designed for use with continuous wave (CW) signals such as RFID, some RF energy harvesters are designed for use with non-CW signals such as WiFi and GSM. These waveform aware harvesters include ambient energy harvesters that harvest signals of opportunity that are already present in the environment. For these types of harvesters, the system designer must work around the waveforms and power levels that are already present in the environment.

SUMMARY OF THE INVENTION

[0006] A method and apparatus is disclosed herein for harvesting ambient energy. In one embodiment, an energy harvester comprises: a first RF rectifier to output a first voltage determined by rectified RF energy in response to received RF energy; a first energy reservoir coupled to the first RF rectifier to store energy at the first voltage; a DC/DC converter coupled to the first energy reservoir to convert the first voltage to a second voltage; a second reservoir coupled to the DC/DC converter to store energy at the second voltage, the second voltage being greater than the first voltage; and a third reservoir coupled to the second reservoir to receive energy transferred from the second reservoir periodically.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] The present invention will be understood more fully from the detailed description given below and from the accompanying drawings of various embodiments of the invention, which, however, should not be taken to limit the invention to the specific embodiments, but are for explanation and understanding only.

[0008] FIG. **1** is a block diagram of one embodiment of a sensor device.

[0009] FIG. **2** is a block diagram of one embodiment of an RF energy harvester.

[0010] FIG. **3** is another more detailed block diagram of one embodiment of an RF energy harvester.

[0011] FIG. **4** illustrates a diode-based rectifier schematic.

[0012] FIG. **5** illustrates a diode output circuit model.

[0013] FIG. **6**A illustrates the usable energy stored in capacitor C_X during one transmission burse of duration T_p.

[0014] FIG. **6**B illustrates examples of good and bad choices of C_X.

[0015] FIG. **7** illustrates one embodiment of a self-excited oscillator.

[0016] FIG. **8** illustrates one embodiment of a charger schematic.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

[0017] A system for ambient energy harvesting and method for doing the same are described. In one embodiment, the energy harvesting system includes three DC reservoirs to store energy generated from ambient radio frequency (RF) input signals. These reservoirs are coupled together in a cascaded fashion to provide DC energy to a load.

[0018] In one embodiment, the ambient RF energy harvester is part of a sensor/tag device and is used to provide energy to one or more components of the sensor/tag device such as, for example, the sensor, a controller, a communication unit a processing device, etc.

[0019] In the following description, numerous details are set forth to provide a more thorough explanation of the present invention. It will be apparent, however, to one skilled in the art, that the present invention may be practiced without these specific details. In other instances, well-known structures and devices are shown in block diagram form, rather than in detail, in order to avoid obscuring the present invention.

[0020] Some portions of the detailed descriptions which follow are presented in terms of algorithms and symbolic representations of operations on data bits within a computer memory. These algorithmic descriptions and representations are the means used by those skilled in the data processing arts to most effectively convey the substance of their work to others skilled in the art. An algorithm is here, and generally, conceived to be a self-consistent sequence of steps leading to a desired result. The steps are those requiring physical manipulations of physical quantities. Usually, though not necessarily, these quantities take the form of electrical or magnetic signals capable of being stored, transferred, combined, compared, and otherwise manipulated. It has proven convenient at times, principally for reasons of common usage, to refer to these signals as bits, values, elements, symbols, characters, terms, numbers, or the like.

[0021] It should be borne in mind, however, that all of these and similar terms are to be associated with the appropriate physical quantities and are merely convenient labels applied to these quantities. Unless specifically stated otherwise as apparent from the following discussion, it is appreciated that throughout the description, discussions utilizing terms such as "processing" or "computing" or "calculating" or "determining" or "displaying" or the like, refer to the action and processes of a computer system, or similar electronic computing device, that manipulates and transforms data represented as physical (electronic) quantities within the computer system's registers and memories into other data similarly represented as physical quantities within the computer system memories or registers or other such information storage, transmission or display devices.

[0022] The present invention also relates to apparatus for performing the operations herein. This apparatus may be specially constructed for the required purposes, or it may comprise a general purpose computer selectively activated or reconfigured by a computer program stored in the computer. Such a computer program may be stored in a computer readable storage medium, such as, but is not limited to, any type of disk including floppy disks, optical disks, CD-ROMs, and magnetic-optical disks, read-only memories (ROMs), random access memories (RAMs), EPROMs, EEPROMs, magnetic or optical cards, or any type of media suitable for storing electronic instructions, and each coupled to a computer system bus.

[0023] The algorithms and displays presented herein are not inherently related to any particular computer or other apparatus. Various general purpose systems may be used with programs in accordance with the teachings herein, or it may prove convenient to construct more specialized apparatus to perform the required method steps. The required structure for a variety of these systems will appear from the description below. In addition, the present invention is not described with reference to any particular programming language. It will be appreciated that a variety of programming languages may be used to implement the teachings of the invention as described herein.

[0024] A machine-readable medium includes any mechanism for storing or transmitting information in a form readable by a machine (e.g., a computer). For example, a machine-readable medium includes read only memory ("ROM"); random access memory ("RAM"); magnetic disk storage media; optical storage media; flash memory devices; etc.

One Embodiment of a Sensor Device

[0025] FIG. **1** is a block diagram of one embodiment of a sensor device. The sensor device may be part of a tag. Referring to FIG. **1**, sensor device **100** comprises an ambient radio-frequency (RF) energy harvester **101**, an ultra-low-power (ULP) sensor **102** (e.g., temperature, light, humidity, etc.), a micro-controller **103**, and a communication unit **104**. Although not shown, sensor device **100** includes one or more antennas. The antennas may be part of communication unit **104** or may be separate, such as in the case of an antenna used to harvest energy.

[0026] In one embodiment, RF energy harvester **101** harvests ambient RF power to run ultra-low power tags and/or sensors **102**, micro-controller **103**, and/or communication unit **104**, based on received RF input energy received by an antenna (not shown) of sensor device **100**. There are a number of sources of ambient RF power, including, but not limited to WiFi, DTV and GSM signals. Other sources of ambient RF power include, for example, toys, proprietary communication protocols, and even stray energy leakage from devices, such as microwave ovens. The antenna receiving the RF energy input may be part of communication unit **104** or a separate antenna. The antenna(s) may be designed to receive the energy in multiple frequency bands, or there may be separate antenna(s) for these different frequency bands.

[0027] In one embodiment, sensor **102** comprises an ultra-low-power (ULP) sensor and/or tag and micro-controller **103** comprises an ultra-low-power (ULP) micro-controller. Sensor **102** may comprise a temperature sensor, a light sensor, a motion sensor, air flow sensor, image sensor, accelerometer, gyroscope, gas detector, microphone, atmospheric pressure sensor, humidity sensor, mechanical pressure sensor, etc. In one embodiment, sensor **102**, as a ULP sensor, has a power requirement of 1-20 uW and a duty cycle of 1% to 10% (where the duty cycle varies in response to available energy over a period of time or the specific sensing needs, usually seconds to hours).

[0028] In one embodiment, micro-controller **103** performs control operations and/or processing for the sensor device **100**. In one embodiment, controller **103** is responsive to and receives sensor data from sensor **102** and provides it to communication unit **104** for transmission (e.g., wireless transmission) from device **100**. Controller **103** may also log sensor data for a time, and transmit the entire log to communication unit **104** for transmission later. Note that in this case the sensing rate and the communication rate can be different.

[0029] Communication unit **104** provides a communication mechanism for sensor device **100**. In one embodiment communication unit **104** is a wireless communication unit. In one embodiment, communication unit **104** has an RF transmission along with an antenna for receiving and transmitting RF signals, including RF signals transmitted with an access point or other locally positioned device.

[0030] In one embodiment, the ambient RF energy harvester is optimized for overall efficiency and a minimum of start-up power. These optimizations improve the range at which energy can be usefully harvested and improve the number of places where ambient energy is practical, particularly in view of the fact that ambient RF energy levels are small and often bursty (traffic-dependent), rectified voltages are very low (<100 mV) at typical ranges, and typical Complementary Metal Oxide Semiconductor (CMOS) loads such as MCU etc. require 800+mV to operate.

[0031] An energy harvester and method for performing energy harvesting are described. In one embodiment, the energy harvester converts incident radio frequency (RF) energy, or ambient RF energy, to direct current energy. In one embodiment, the energy harvester comprises: a first RF rectifier to output a first voltage determined by rectified RF energy in response to received RF energy; a first energy reservoir coupled to the first RF rectifier to store energy at the first voltage; a DC/DC boost converter coupled to the first energy reservoir to convert the first voltage to a second voltage; a second reservoir coupled to the DC/DC converter to store energy at the second voltage, where the second voltage is greater than the first voltage; and a third reservoir (e.g., a longer-term energy storage device such as a low-leakage capacitor or battery, etc.) coupled to the second reservoir to receive energy transferred from the second reservoir periodically.

[0032] In one embodiment, the RF rectifier comprises one selected from a group consisting of: a Schottky diode, a P-N junction diode, a diode-connected field effect transistor, and a tunnel diode.

[0033] In one embodiment, the first reservoir comprises a capacitor with a capacitance value determined at least in part by the envelope of the received RF energy. In one embodiment, the first reservoir comprises a capacitor having a value based on stored energy above a threshold voltage given ambient RF input power and a duty cycle of the input power. In another embodiment, the first reservoir comprises a first capacitor and the second reservoir comprises a second capacitor, wherein capacitance of the second capacitor is approximately equal to a product of the first capacitor capacitance times a square of a ratio of an expected value of the second voltage divided by an expected value of the first voltage.

[0034] In one embodiment, the second reservoir comprises an inductor used to boost the voltage, wherein energy stored in the inductor is approximately equal to energy stored in the first reservoir.

[0035] In one embodiment, the DC/DC boost converter comprises a self-excited oscillator. In one embodiment, the self-excited oscillator comprises a transistor coupled to a transformer. In one embodiment, the frequency of the self-excited oscillator is controlled at least in part by an envelope of the received RF energy. In one embodiment, the DC/DC converter is a switch mode converter with switching intervals controlled at least in part by an envelope of the received RF energy.

[0036] In one embodiment, the third reservoir is coupled to the second reservoir via a switch, which is closed periodically to transfer energy from the second reservoir to the third reservoir. In one embodiment, the switch is controlled at least in part by the envelope of the received RF energy. In one embodiment, the third reservoir is coupled to a load, and the first voltage is insufficient to operate the load directly, and the third voltage is sufficient to operate the load. In one embodiment, the third reservoir has an energy storage capacity that is greater than an energy storage capacity of the second reservoir.

[0037] FIG. **2** is a block diagram of one embodiment of an RF energy harvester. Referring to FIG. **2**, RF energy harvester **200** comprises RF rectifier **201**, a first direct current (DC) reservoir (storage) **202**, a boost converter and second reservoir **203**, a third reservoir **204**, and control logic **205**.

[0038] RF rectifier **201** receives ambient RF energy input received from an antenna and rectifies the RF energy to produce a DC voltage.

[0039] First DC reservoir **202** stores the voltage generated from the rectified RF signals received by device **100**. In one embodiment, reservoir **202** comprises a capacitor. In one embodiment, the size (value) of the capacitor that is selected for use is based on the stored energy above a voltage threshold for the ambient RF input power.

[0040] Second DC reservoir **203** receives the voltage and increases the voltage. To increase the voltage, second DC reservoir **203** includes a DC/DC converter. In one embodiment, the DC/DC converter comprises a boost converter which includes a self-excited oscillator for operation. The self-excited oscillator includes a junction field effect transistor (JFET) coupled with a transformer. In one embodiment, the amount of the boost conversion is a function of the input energy, and the DC/DC converter is designed to maximize the output voltage as a function of input voltage in this low-energy environment. FIG. **7** illustrates one embodiment of a self-excited oscillator. FIG. **8** illustrates one embodiment of a charger schematic that receives the output of the boost converter, such as, for example, the one shown in FIG. **7**.

[0041] The upconverted voltage from the DC/DC converter is stored in a storage device of second reservoir **203**. In one embodiment, second reservoir **203** uses a capacitor to store the upconverted voltage. In another embodiment, second reservoir **203** uses an inductor to store the upconverted voltage, but this is a much shorter-term storage option.

[0042] When the energy stored in second reservoir **203** reaches a certain level, the energy is transferred to a third reservoir (not shown). Thus, the transfer occurs periodically at a rate determined by the amount of input energy. In one embodiment, the third reservoir comprises a battery, a capacitor, a supercap, etc. The third reservoir provides DC energy to a load.

[0043] In one embodiment, the load comprises a sensor, processor or microcontroller, and/or a communication unit. In one embodiment, the DC voltage from the third reservoir is able to run the load, while the voltage produced by rectifier **201** would not be able to do so.

[0044] Control logic **205** controls the energy harvesting process. In one embodiment, control logic **205** comprises a separate unit that generates control signals to control each of the reservoirs **202-3203**. In another embodiment, the control is distributed through various discrete gating components such that a separate control unit is not necessary but the function is still performed.

[0045] FIG. **3** is another more detailed block diagram of one embodiment of an RF energy harvester. Referring to FIG. **3**, RF energy harvester **300** comprises RF rectifier **301**, a first direct current (DC) reservoir (storage) **302**, second reservoir **303**, and a third reservoir **304**. In one embodiment, the input to RF rectifier **301** is an RF input from an antenna on the sensor device. RF rectifier **301** converts the RF input signals to a voltage. In one embodiment, the voltage may range between 10 mV and 500 mV. The voltage is stored in reservoir **302**. In one embodiment, reservoir **302** comprises a ceramic capacitor 1 uF to 100 uF or more.

[0046] In one embodiment, the stored energy from reservoir **302** is transferred to second reservoir **303** via a transistor **311**, self-gated as the voltage rises in reservoir **302**. In one embodiment, second reservoir **303** comprises a boost converter **303**A. The boost converter **303**A operates to upconvert the voltage stored in reservoir **302** and store that upconverted voltage into reservoir **303**B.

[0047] The energy stored in second reservoir **303** is transferred to third reservoir **304**. In one embodiment, third reservoir **304** comprises a tantalum capacitor, supercap 100 uF to 1 F or a battery. In one embodiment, transistor **312** (e.g., a FET) couples second reservoir **303** to third reservoir **304**, which turns on when second reservoir **303** reaches an optimal voltage. In one embodiment, transistor **312** is controlled at least in part by the envelope of the received RF energy such that when the voltage stored in reservoir **303**B reaches a certain watermark level, transistor **312** closes/turns on and the energy is transferred. In one embodiment, through the selection of high gate-threshold transistors, the voltage can build up fairly high in reservoir **303** (as much as 1-2V) before the transistor **312** closes/turns on. Transistor **312** closes/turns on automatically when the gate threshold of the transistor design is reached in reservoir **303**. As soon as transistor **312** closes/turns on, the energy is transferred from reservoir **303** into reservoir **304**, thus depleting the energy in reservoir **303**, and the gate threshold falls below the watermark level and the transistor **312** closes/opens again. The process then repeats when the voltage in reservoir **303** becomes sufficiently high.

[0048] Note that transistor **311** and **312** may be replaced with other types of switches.

Waveform Aware Harvesting

[0049] In one embodiment, the RF waveform harvester comprises a RF to DC converter that is optimized for use with non-continuous wave (CW) signals. In another embodiment, the RF waveform harvesters are waveform aware and are able to perform ambient energy harvesting based on a waveform model, where the available RF energy is in the form of communication waveforms of a variety of types. Examples of commonly-harvested ambient RF energy sources with non-continuous wave signals include, but are not limited to, digital television (DTV) broadcasts, cellular base station transmissions (e.g., GSM, Code Division Multiple Access (CDMA)), Wi-Fi transmissions, etc.

[0050] For example, in one embodiment, WiFi is the communications waveform that is the ambient form of energy to be harvested. WiFi is bursty and has low amplitude compared to other sources of energy, where a bursty signal has a high peak to average ratio over ms to s time scale. For example, some WiFi signals have WiFi transmission burst times on the order of 1 ms. For a bursty signal, the harvester can only contribute energy to the reservoir capacitor (C_X) when the ambient source is transmitting.

[0051] FIG. **4** illustrates a diode-based harvester schematic. Referring to FIG. **4**, in one embodiment, the harvester is designed to be well matched at input power levels below −20 dBm where the majority of ambient Wi-Fi signals are expected to fall. Thus, harvester is characterized both in terms of its input match from 2.4 GHz to 2.5 GHz, as well as for harvesting efficiency at its maximum power point (MPP). Harvesting efficiency is defined as the ratio of input RF power at the harvester to harvested DC power at the load (η=PDC/PRF).

[0052] The purpose of capacitor C_X in FIG. **4** is to store charge for any subsequent active electronic circuitry. In one embodiment, typical circuit power-up threshold voltages V_{TH} equals 0.8 V, as is typical for CMOS logic in commodity processes. In another embodiment, the power-up threshold voltages V_{TH} equals 50 mV, as is typical for ultra-low startup voltage boost converters using JFETs. Alternatively, specially doped CMOS transistors, floating gate devices, or Silicon on Insulator (SOI) transistors may be capable of operation down to voltages approaching V_{TH}=50-100 mV as well.

[0053] If the usable energy is defined as the portion of the stored energy above some threshold voltage V_{TH}, the following describes how to find a value of C_X given the ambient RF input power and the signal's duty cycle. This is based on the maximum power point curve for the harvester, the typical burst timings for the ambient signal of interest (e.g., WiFi with some traffic model), and a time domain response for the harvester as a function of C_X. Thus, in one embodiment, the storage capacitor selection C_X is optimized on the time scale of network traffic for the signal being harvested.

[0054] FIG. **5** illustrates a diode output circuit model. Because the diode is a non-linear device, the power delivered to the load varies dramatically with load impedance. The Maximum Power Point Tracking (MPPT) circuit functions as a load impedance converter to maximize the power delivered to the load, given a particular nonlinear operating point.

[0055] Referring to FIG. **5**, the diode output model consists of an open-circuit voltage V_{source} along with a nonlinear source resistance R_{source}. The pair, V_{source}, R_{source}, represents a fixed operating point at maximum power-point (MPP). The load resistance R_{load} includes any leakage currents in C_X as well as the quiescent current of the active circuitry attached to the harvester. In other words, the load impedance R_{load} is chosen a priori based on the MPP of the harvester at a given ambient signal level (the maximum power point of the harvester given the input power). Using this model, only the effect of varying C_X (not other system conditions) is considered.

[0056] FIG. **6**A illustrates the usable energy stored in capacitor C_X during one transmission burse of duration T_p. Referring to FIG. **6**A, the usable energy in C_X is represented as the shaded portion of the curve above the voltage threshold and below the voltage across the capacitor.

[0057] FIG. **6**B illustrates examples of good and bad choices of C_X. If C_X is too large, the active circuitry will never be powered because the input energy is insufficient to generate voltage above the threshold. On the other hand, if C_X is minimized, little energy will be stored in the reservoir C_X and the runtime of the active circuitry will not extend beyond the transmission intervals of the RF source. Referring to FIG. **6**, harvesters 'A' and 'B' represent different choices for C_X with the same conditions otherwise. In harvester A, C_X looks appropriate as the output voltage exceeds V_{TH} during each burst, while in harvester B, C_X is too big and the output voltage never exceeds V_{TH} so there is no usable energy.

[0058] An expression is derived for the useful-energy-maximizing reservoir capacitance C_X as a function of ambient RF signal duty cycle and the load impedance RL at a diode harvester's maximum power point. The expression may be used to calculate the usable energy as a function of the known parameters and the unknown capacitance C_X.

[0059] The total energy in C_X may be expressed as follows:

$$E = \frac{1}{2}CV^2$$

where C is the capacitance and V is the terminal voltage of C_X. It is assumed that the energy stored on the capacitor is only usable when the capacitor voltage is above the desired threshold V_{TH}. The usable energy is thus the difference between the stored energy at the end of an RF transmission (at time t=T_p) and the stored energy when C_X is charged only to V_{TH} and is expressed as follows:

$$E_{usable} = \frac{1}{2}C\left[V_{T_p}^2 - V_{th}^2\right]$$

[0060] where V_{Tp} is the terminal voltage on C_X at the end of an RF transmission (e.g., at the end of each burst) of duration T_p and is expressed as:

$$V_{T_p} = V_{src}(1 - e^{-T_p/R_LC})$$

[0061] By substitution, the equation for the usable energy may be re-written in terms of C and the known values of R_L, V_{Th} and T_p as follows:

$$E_{usable} = \frac{1}{2}CV_{T_p}^2\left[e^{-\frac{2T_p}{R_LC}} - 2e^{-\frac{T_p}{R_LC}} + \left(1 - \left(\frac{V_{th}}{V_{T_p}}\right)^2\right)\right]$$

The equation may be solved for E_{usable} as a function of C_X.

[0062] Thus, a new class of RF energy harvesters, referred to herein as waveform aware harvesters, have been described.

These RF energy harvesters are RF to DC converters that are optimized for their performance with non-CW signals.

[0063] Whereas many alterations and modifications of the present invention will no doubt become apparent to a person of ordinary skill in the art after having read the foregoing description, it is to be understood that any particular embodiment shown and described by way of illustration is in no way intended to be considered limiting. Therefore, references to details of various embodiments are not intended to limit the scope of the claims which in themselves recite only those features regarded as essential to the invention.

We claim:

1. An energy harvester to convert incident radio frequency (RF) energy to direct current energy, the energy harvester comprising:

a first RF rectifier to output a first voltage determined by rectified RF energy in response to received RF energy;

a first energy reservoir coupled to the first RF rectifier to store energy at the first voltage;

a DC/DC converter coupled to the first energy reservoir to convert the first voltage to a second voltage;

a second reservoir coupled to the DC/DC converter to store energy at the second voltage, the second voltage being greater than the first voltage; and

a third reservoir coupled to the second reservoir to receive energy transferred from the second reservoir periodically.

2. The energy harvester defined in claim **1** wherein the third reservoir is coupled to the second reservoir via a first switch, the first switch being closed periodically to transfer energy from the second reservoir to the third reservoir.

3. The energy harvester defined in claim **2** wherein the first switch is controlled at least in part by the envelope of the received RF energy.

4. The energy harvester defined in claim **1** wherein the third reservoir is coupled to a load.

5. The energy harvester defined in claim **4** wherein the first voltage is insufficient to operate the load directly, and the third voltage is sufficient to operate the load.

6. The energy harvester defined in claim **1** wherein the DC/DC converter comprises a self-excited oscillator.

7. The energy harvester defined in claim **6** wherein frequency of the self-excited oscillator is controlled at least in part by an envelope of the received RF energy.

8. The energy harvester defined in claim **1** wherein the DC/DC converter comprises a boost converter.

9. The energy harvester defined in claim **1** wherein the first reservoir comprises a capacitor with a capacitance determined at least in part by the envelope of the received RF energy.

10. The energy harvester defined in claim **1** wherein the first reservoir comprises a first capacitor and the second reservoir comprises a second capacitor, wherein capacitance of the second capacitor is approximately equal to a product of the first capacitor capacitance times a square of a ratio of an expected value of the second voltage divided by an expected value of the first voltage.

11. The energy harvester defined in claim **1** wherein the DC/DC converter is a switch mode converter with switching intervals controlled at least in part by an envelope of the received RF energy.

12. The energy harvester defined in claim **1** wherein the RF rectifier comprises one or more diodes selected from a group consisting of: a Schottky diode, a P-N junction diode, a diode-connected field effect transistor, and a tunnel diode.

13. The energy harvester defined in claim **1** wherein the third reservoir has an energy storage capacity that is greater than an energy storage capacity of the second reservoir.

14. The energy harvester defined in claim **1** wherein the second reservoir comprises an inductor, wherein energy stored in the inductor is approximately equal to energy stored in the first reservoir.

15. The energy harvester defined in claim **1** wherein the first reservoir comprises a capacitor having a value based on stored energy above a threshold voltage given ambient RF input power and a duty cycle of the input power.

16. A device comprising:

an antenna;

a load; and

an energy harvester coupled to the antenna and the load, wherein the energy harvester comprises

a first RF rectifier to output a first voltage determined by rectified RF energy in response to received RF energy;

a first energy reservoir coupled to the first RF rectifier to store energy at the first voltage;

a DC/DC converter coupled to the first energy reservoir to convert the first voltage to a second voltage;

a second reservoir coupled to the DC/DC converter to store energy at the second voltage, the second voltage being greater than the first voltage; and

a third reservoir coupled to the second reservoir to receive energy transferred from the second reservoir periodically.

17. The device defined in claim **16** wherein the third reservoir is coupled to the second reservoir via a switch, the switch being closed periodically to transfer energy from the second reservoir to the third reservoir.

18. The device defined in claim **16** wherein the third reservoir is coupled to a load, and wherein the first voltage is insufficient to operate the load directly, and the third voltage is sufficient to operate the load.

19. The device defined in claim **16** wherein the DC/DC converter comprises a boost converter.

* * * * *

US 20150372541A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2015/0372541 A1**

Guo et al. (43) **Pub. Date: Dec. 24, 2015**

(54) **RECTENNA CIRCUIT ELEMENTS, CIRCUITS, AND TECHNIQUES FOR ENHANCED EFFICIENCY WIRELESS POWER TRANSMISSION OR AMBIENT RF ENERGY HARVESTING**

(71) Applicant: **NATIONAL UNIVERSITY OF SINGAPORE**, Singapore (SG)

(72) Inventors: **Yong Xin Guo**, Singapore (SG); **Zheng Zhong**, Singapore (SG); **Hucheng Sun**, Singapore (SG)

(73) Assignee: **National University of Singapore**, Singapore (SG)

(21) Appl. No.: **14/767,125**

(22) PCT Filed: **Feb. 27, 2014**

(86) PCT No.: **PCT/SG2014/000091**

§ 371 (c)(1),
(2) Date: **Aug. 11, 2015**

Related U.S. Application Data

(60) Provisional application No. 61/769,837, filed on Feb. 27, 2013, provisional application No. 61/956,998, filed on Jun. 21, 2013, provisional application No. 61/856,791, filed on Jul. 22, 2013.

Publication Classification

(51) **Int. Cl.**

H02J 17/00	(2006.01)
H02M 1/32	(2006.01)
H02M 7/217	(2006.01)

(52) **U.S. Cl.**

CPC ***H02J 17/00*** (2013.01); ***H02M 7/217*** (2013.01); ***H02M 1/32*** (2013.01)

(57) **ABSTRACT**

A rectenna includes (a) a multi-band multi-channel (MBMC) matching network, and/or (b) an adaptively reconfigurable rectifier or a breakdown-protected rectifier. An MBMC matching network includes a plurality of T-shaped transmission line matching structures coupled in series. An adaptively reconfigurable rectifier circuit includes a low input power rectifying portion, a high high input power rectifying portion, and a set of transistors configured for selectively and automatically transitioning the adaptively reconfigurable rectifier between a low input power operating configuration and a high input RF power operating configuration, in a manner correlated with input RF power level. A breakdown-protected rectifier includes a transistor-protected diode structure having a diode coupled to a transistor in a manner that protects the diode from direct exposure to negative voltages that would ordinarily cause the diode to break down in the absence of the transistor.

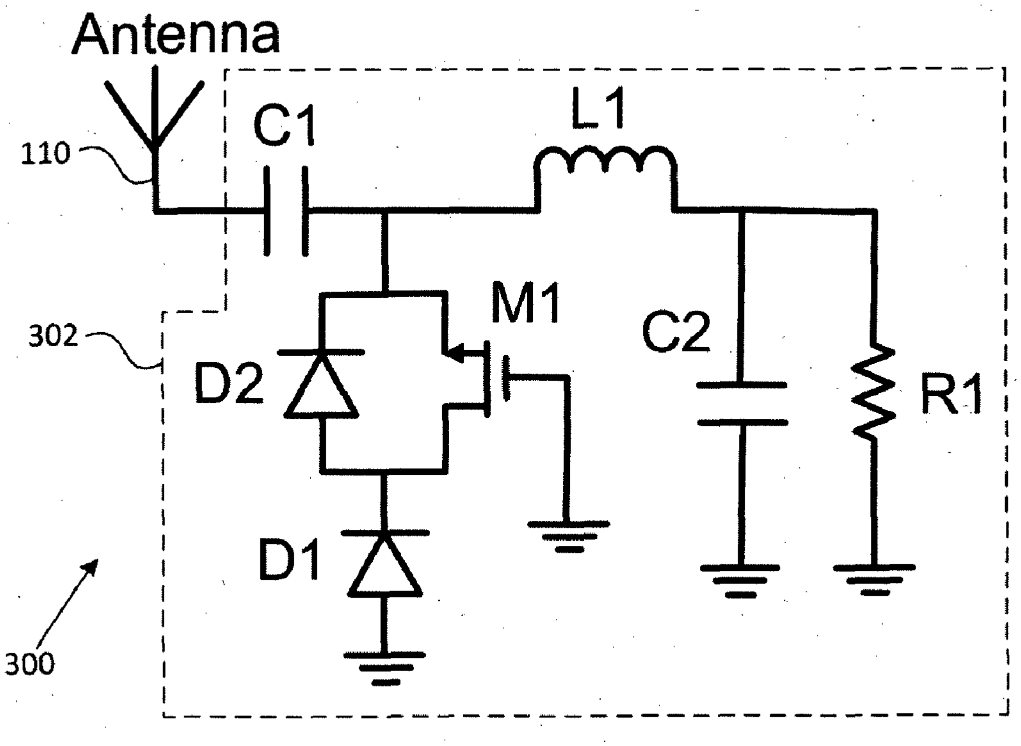

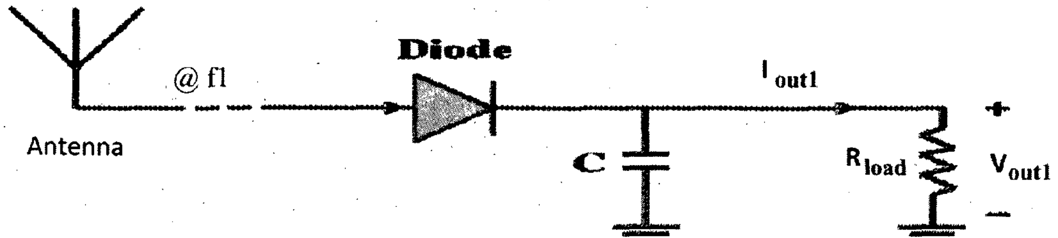
Diode
@ f1
Antenna
I_{out1}
C
R_{load}
+
V_{out1}
–

FIG. 1 (PRIOR ART)

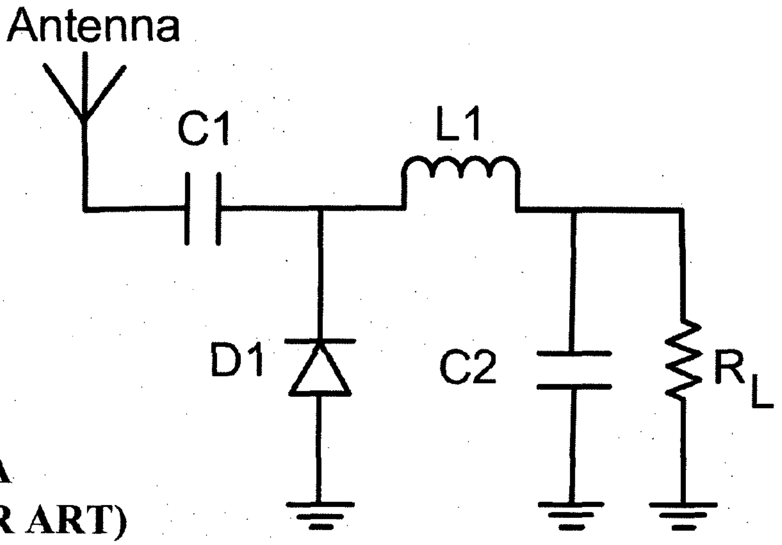
Antenna
C1
L1
D1
C2
R_L

FIG. 2A
(PRIOR ART)

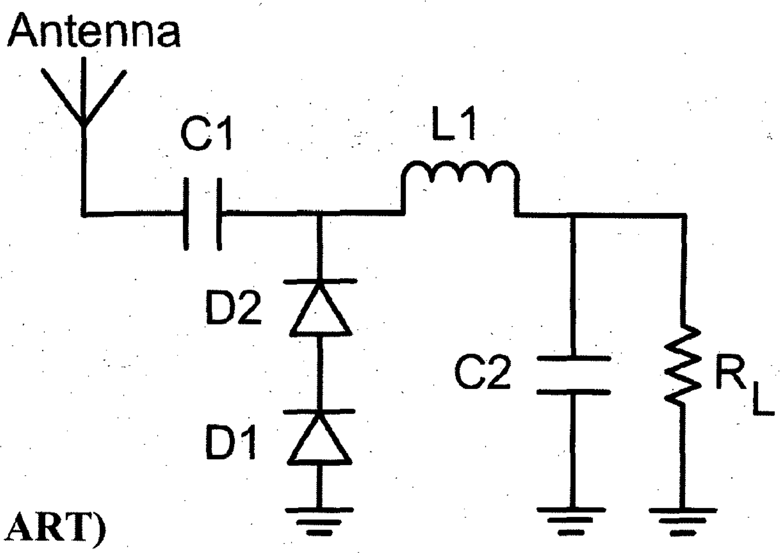
Antenna
C1
L1
D2
D1
C2
R_L

FIG. 2B
(PRIOR ART)

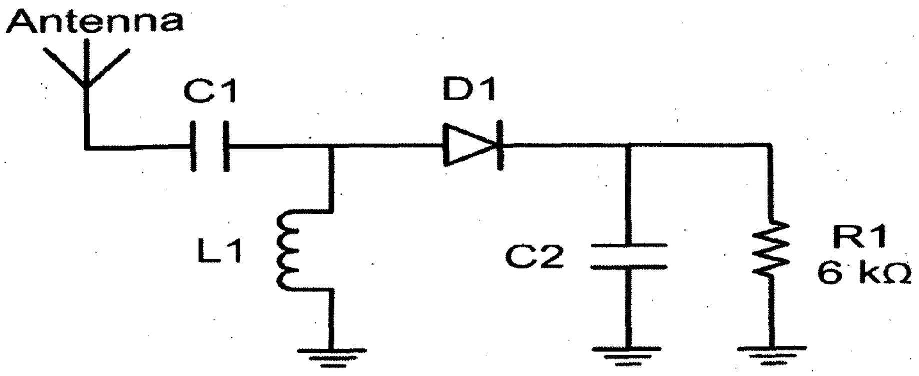
Antenna
C1
D1
L1
C2
R1
6 kΩ

FIG. 2C (PRIOR ART)

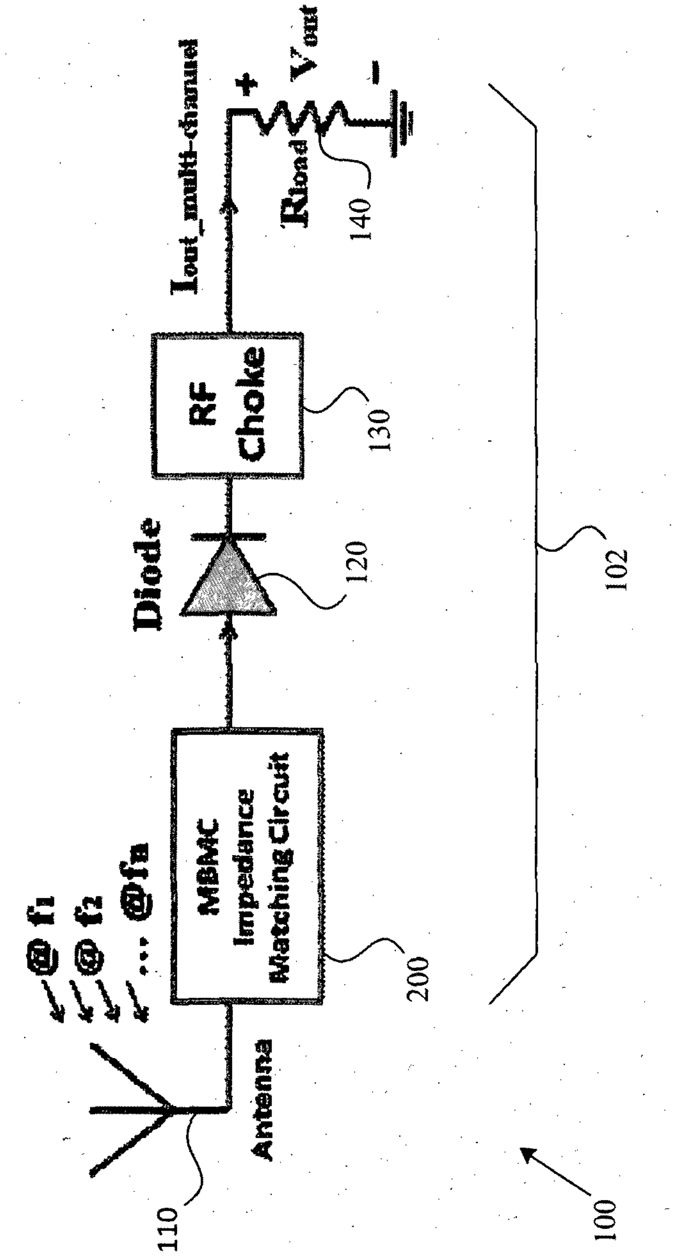
110
Antenna
@ f_1
@ f_2
... @f_n
MBMC Impedance Matching Circuit
200
Diode
120
RF Choke
130
$I_{out\_multi\text{-}channel}$
+
V_{out}
−
R_{load}
140
102
100

FIG. 3

FIG. 4A

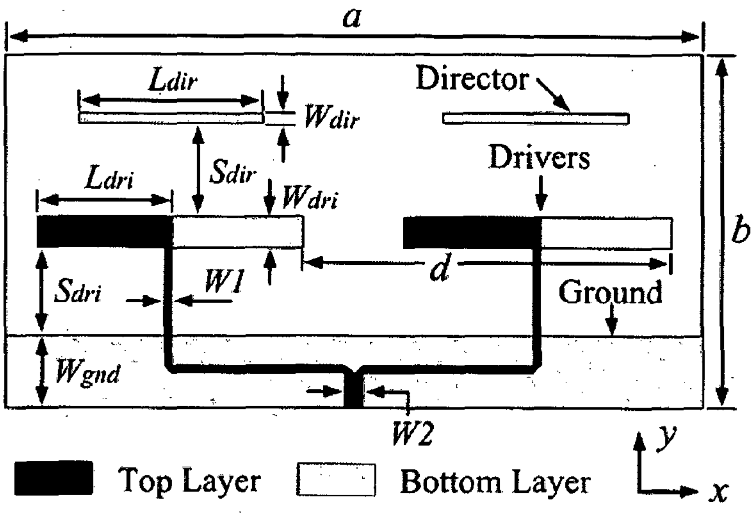
a
Ldir
Wdir
Director
Sdir
Drivers
Ldri
Wdri
b
d
Sdri
W1
Ground
Wgnd
W2
y
x
Top Layer
Bottom Layer

FIG. 4B

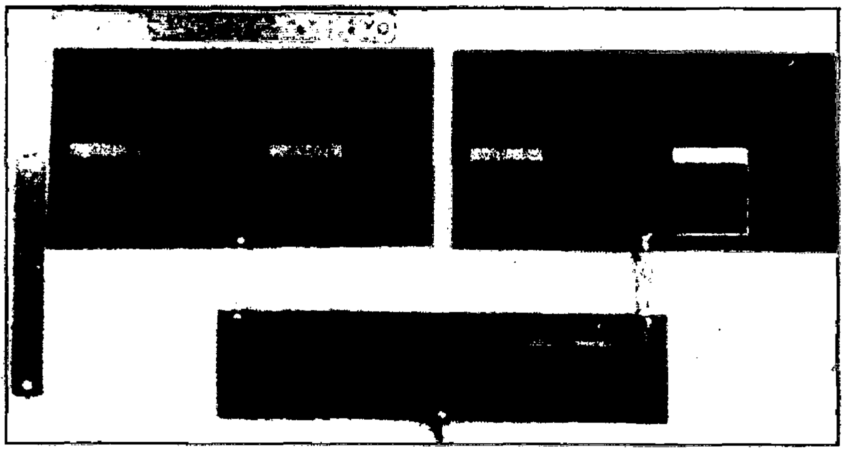

FIG. 4C

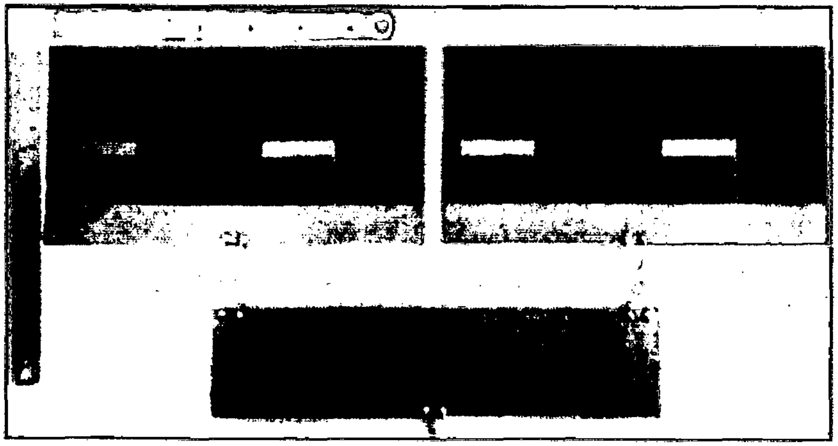

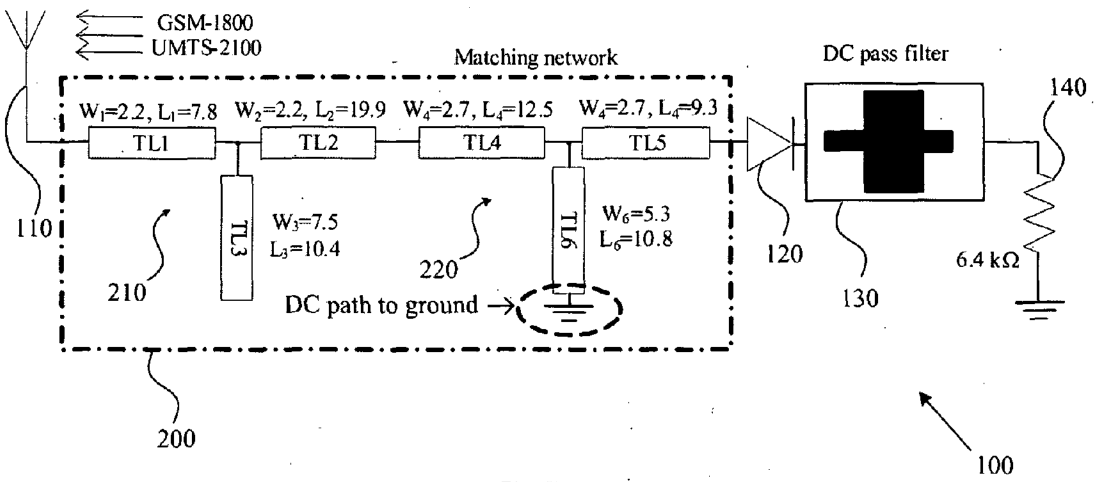
GSM-1800
UMTS-2100
Matching network
DC pass filter
W_1=2.2, L_1=7.8
W_2=2.2, L_2=19.9
W_4=2.7, L_4=12.5
W_4=2.7, L_4=9.3
TL1
TL2
TL4
TL5
TL3
TL6
W_3=7.5
L_3=10.4
W_6=5.3
L_6=10.8
DC path to ground
140
110
210
220
120
130
6.4 kΩ
200
100

FIG. 5A

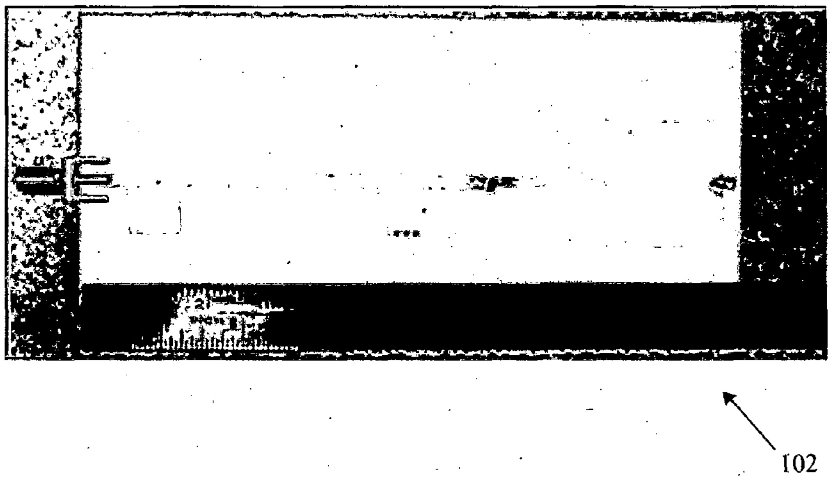
102

FIG. 5B

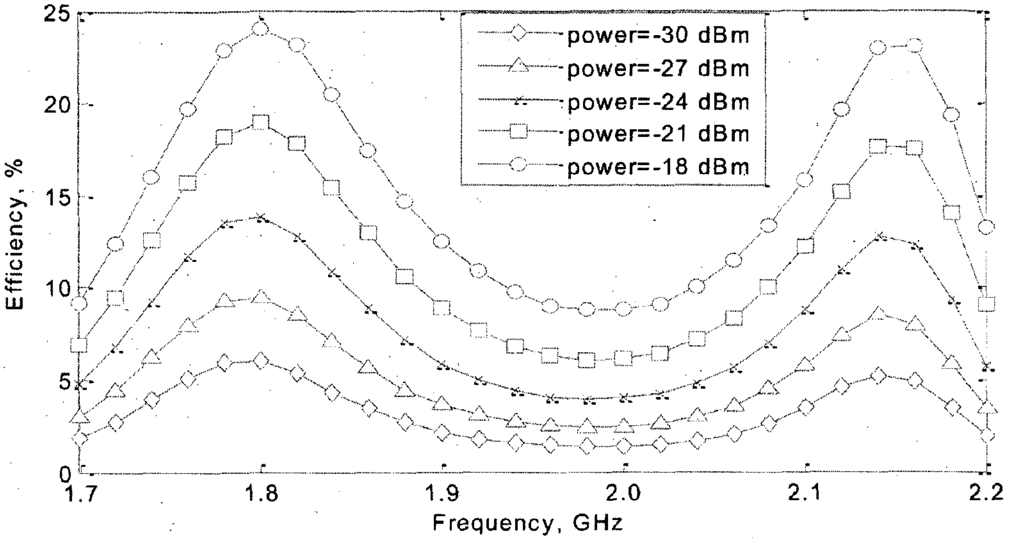
power=-30 dBm
power=-27 dBm
power=-24 dBm
power=-21 dBm
power=-18 dBm
Efficiency, %
0
5
10
15
20
25
Frequency, GHz
1.7
1.8
1.9
2.0
2.1
2.2

FIG. 6

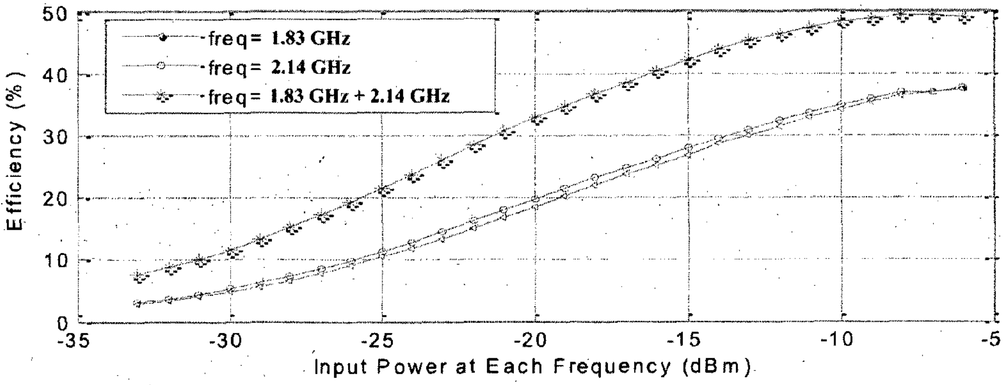
freq= 1.83 GHz
freq= 2.14 GHz
freq= 1.83 GHz + 2.14 GHz
Efficiency (%)
0
10
20
30
40
50
Input Power at Each Frequency (dBm)
-35
-30
-25
-20
-15
-10
-5

FIG. 7

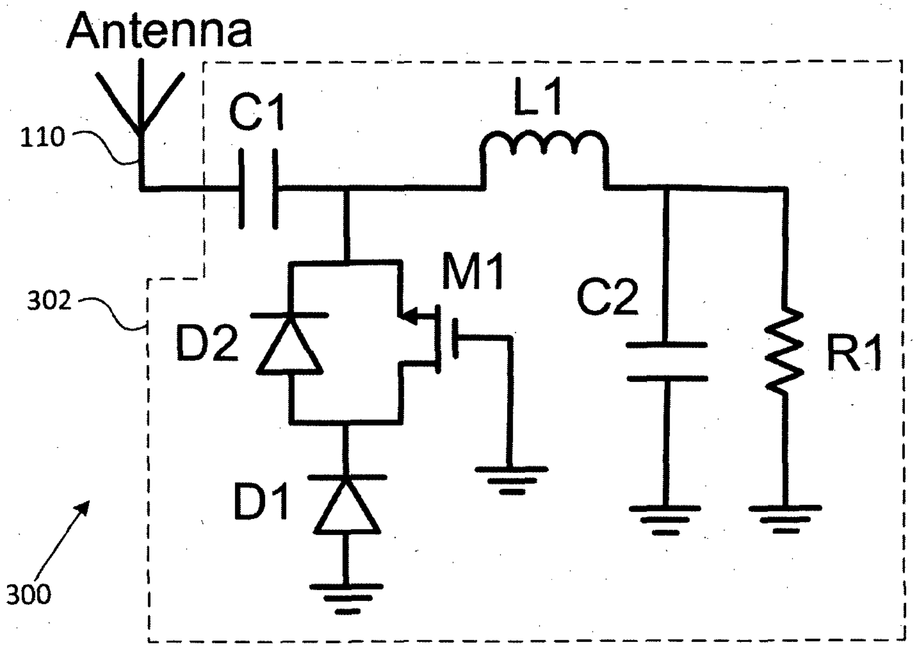
Antenna
110
C1
L1
302
M1
C2
D2
R1
D1
300

FIG. 8A

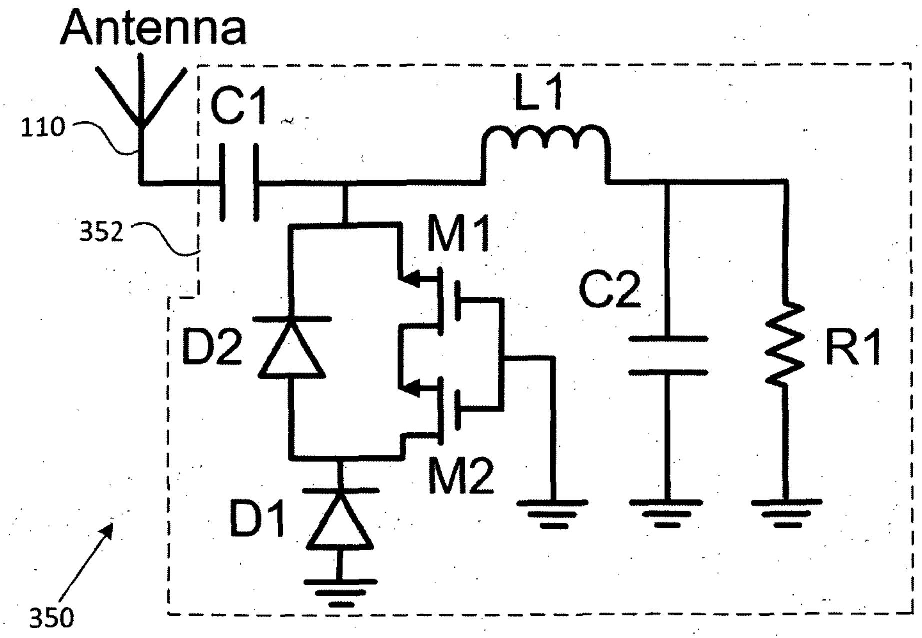
Antenna
110
C1
L1
352
M1
C2
D2
R1
M2
D1
350

FIG. 8B

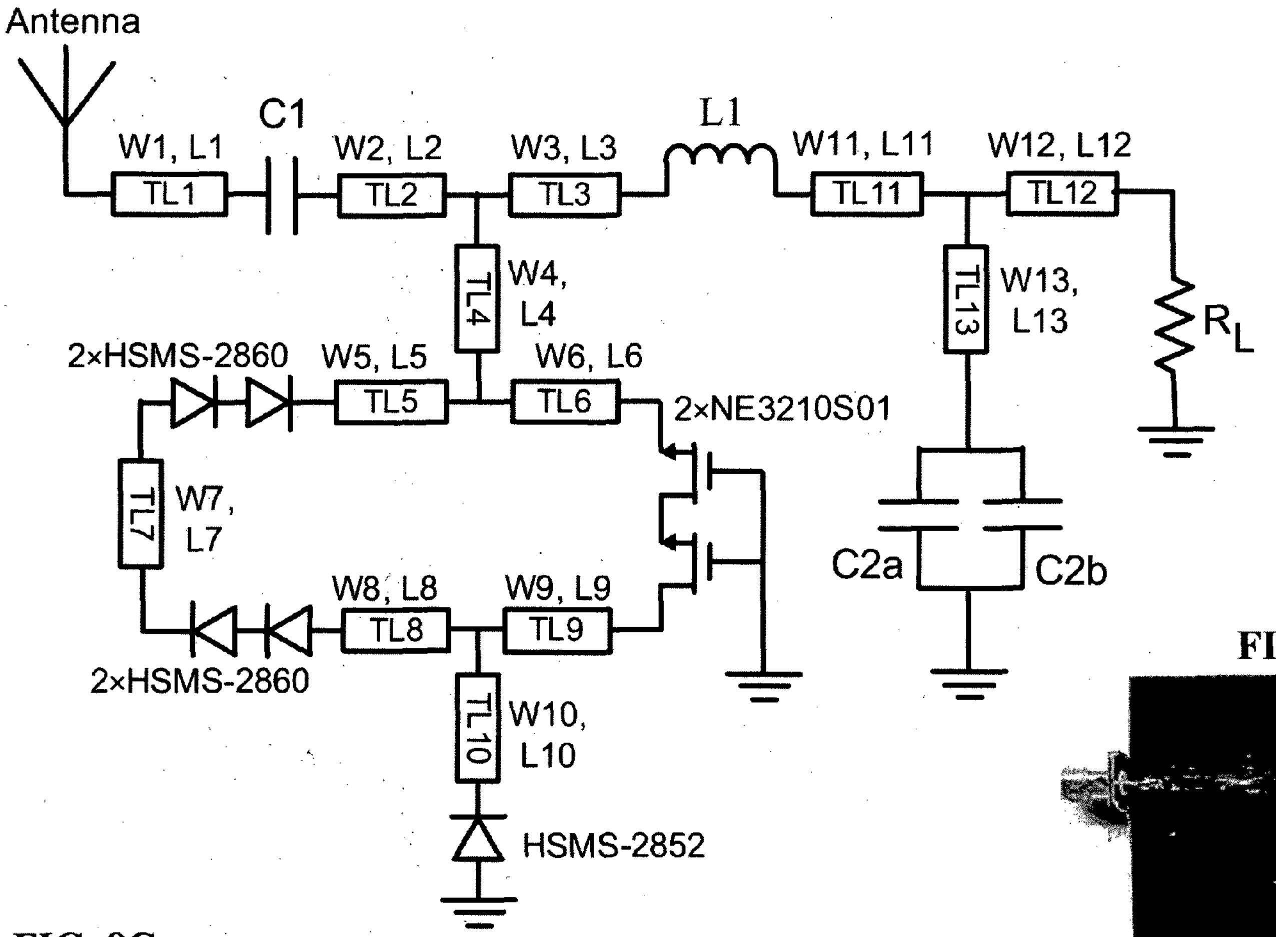
Antenna
C1
L1
350
W1, L1
TL1
W2, L2
TL2
W3, L3
TL3
W11, L11
TL11
W12, L12
TL12
TL4
W4,
L4
TL13
W13,
L13
R_L
2×HSMS-2860
W5, L5
TL5
W6, L6
TL6
2×NE3210S01
TL7
W7,
L7
C2a
C2b
W8, L8
TL8
W9, L9
TL9
2×HSMS-2860
TL10
W10,
L10
HSMS-2852

FIG. 8C

FIG. 8D

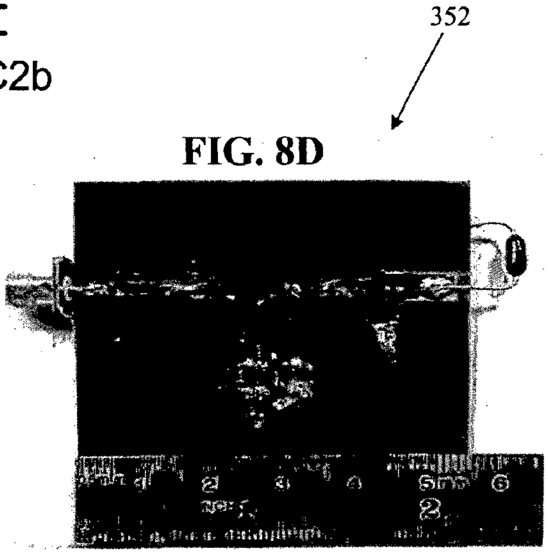
352

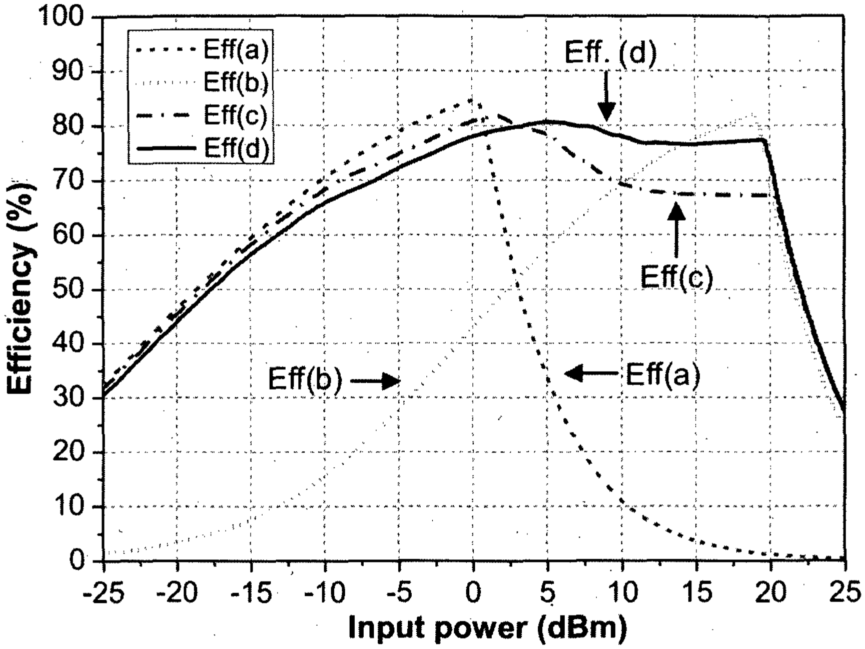
Efficiency (%)
100
90
80
70
60
50
40
30
20
10
0
Input power (dBm)
-25
-20
-15
-10
-5
0
5
10
15
20
25
Eff(a)
Eff(b)
Eff(c)
Eff(d)
Eff. (d)
Eff(c)
Eff(b)
Eff(a)

FIG. 9A

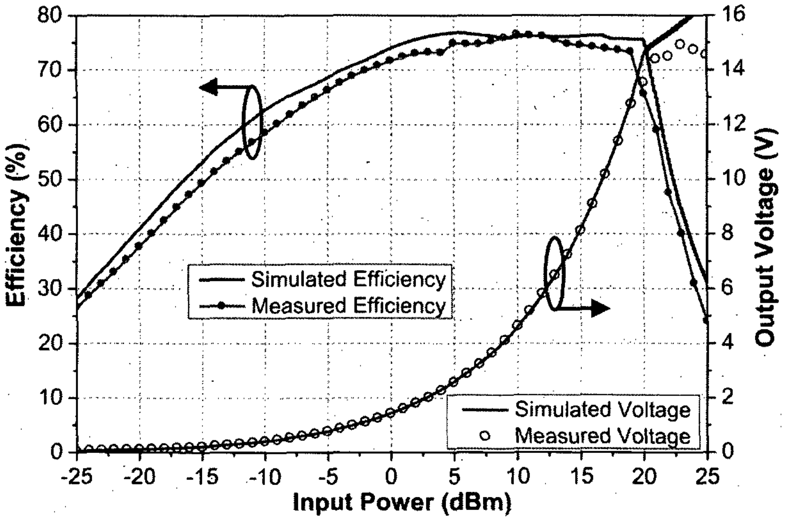
Efficiency (%)
80
70
60
50
40
30
20
10
0
Output Voltage (V)
16
14
12
10
8
6
4
2
0
Input Power (dBm)
-25
-20
-15
-10
-5
0
5
10
15
20
25
Simulated Efficiency
Measured Efficiency
Simulated Voltage
Measured Voltage

FIG. 9B

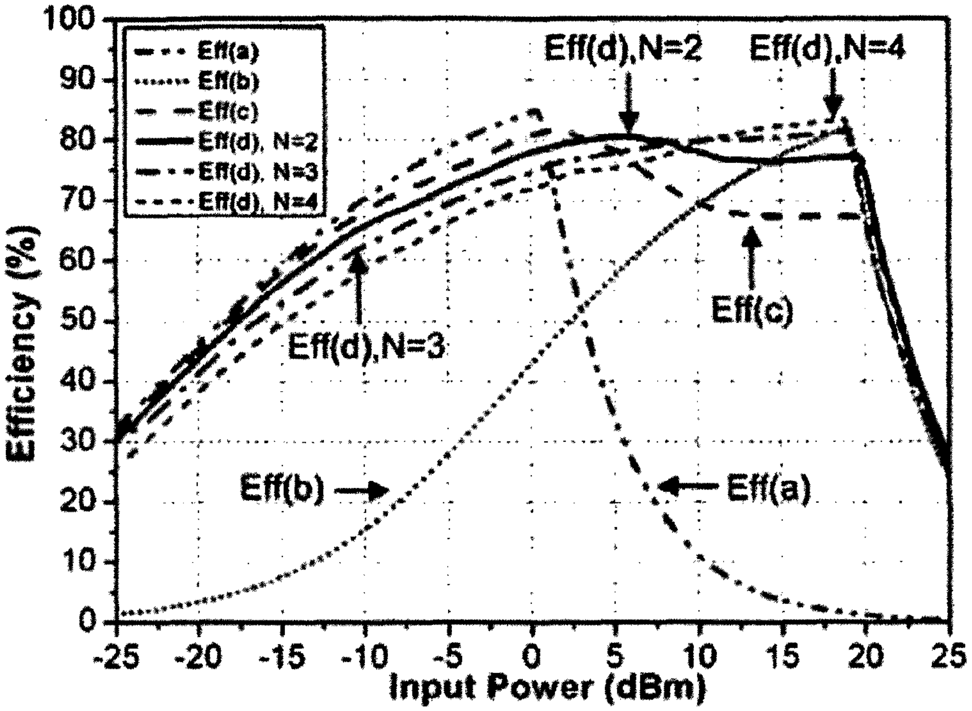
Eff(a)
Eff(b)
Eff(c)
Eff(d), N=2
Eff(d), N=3
Eff(d), N=4
Eff(d),N=2
Eff(d),N=4
Eff(d),N=3
Eff(c)
Eff(b)
Eff(a)
Efficiency (%)
Input Power (dBm)
100
90
80
70
60
50
40
30
20
10
0
-25
-20
-15
-10
-5
0
5
10
15
20
25

FIG. 9C

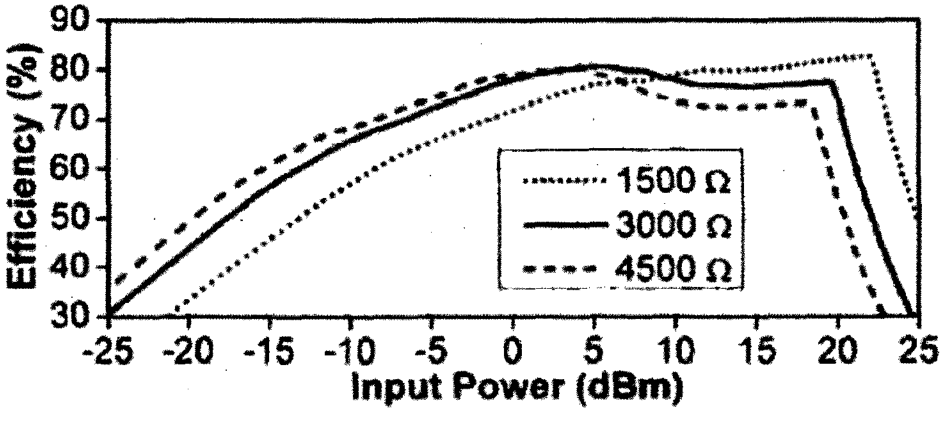
1500 Ω
3000 Ω
4500 Ω
Efficiency (%)
Input Power (dBm)
90
80
70
60
50
40
30
-25
-20
-15
-10
-5
0
5
10
15
20
25

FIG. 9D

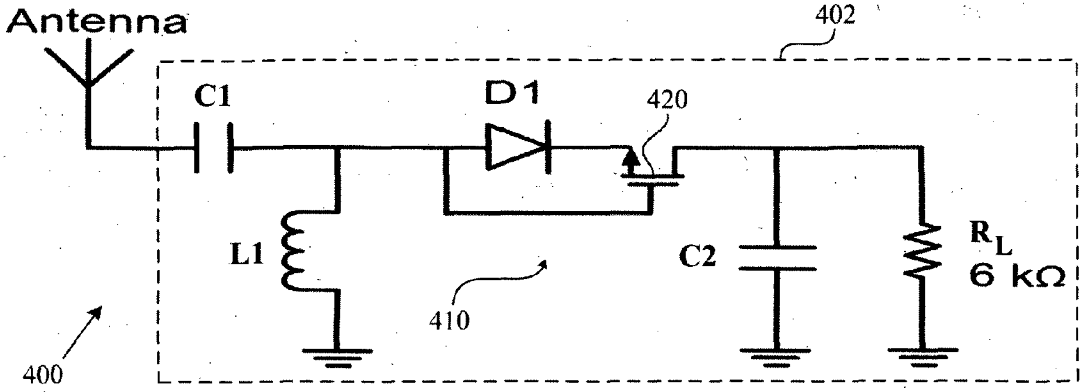
Antenna
402
C1
D1
420
L1
410
C2
R_L
6 kΩ
400

FIG. 10A

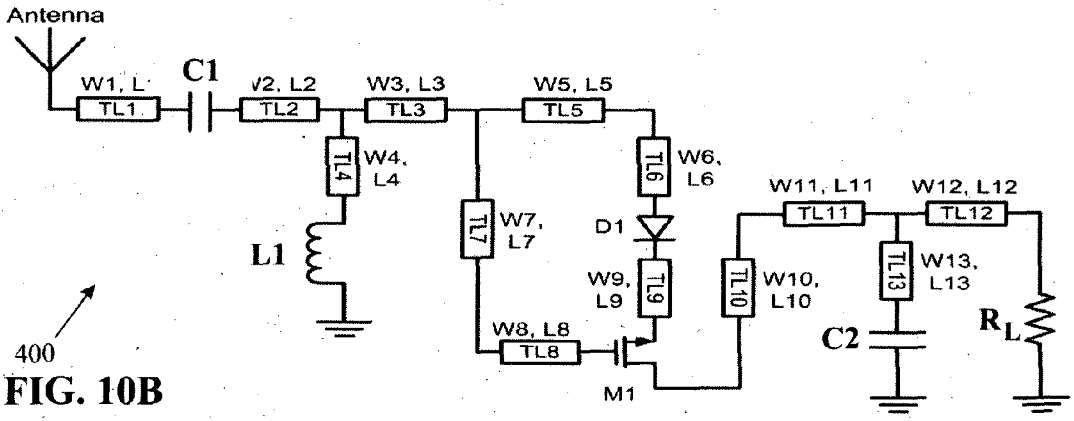
Antenna
W1, L
C1
V2, L2
W3, L3
W5, L5
TL1
TL2
TL3
TL5
TL4
W4, L4
L1
TL7
W7, L7
TL6
W6, L6
D1
TL9
W9, L9
W8, L8
TL8
M1
W11, L11
W12, L12
TL11
TL12
TL10
W10, L10
TL13
W13, L13
C2
R_L
400

FIG. 10B

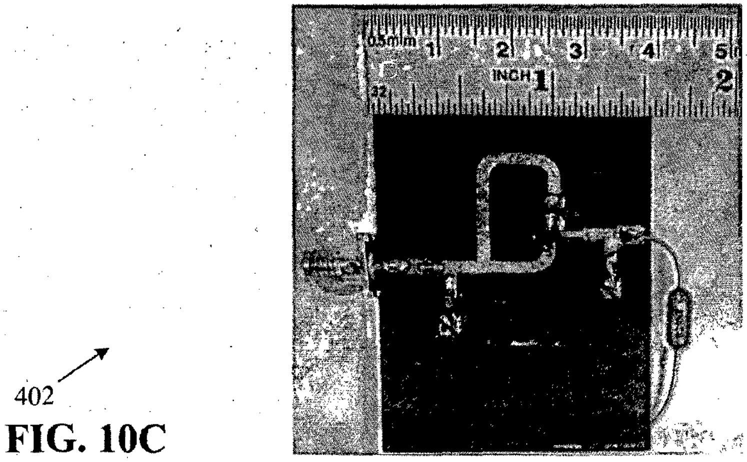
0.5mm
1
2
3
4
5
INCH
1
2
402

FIG. 10C

FIG. 11A

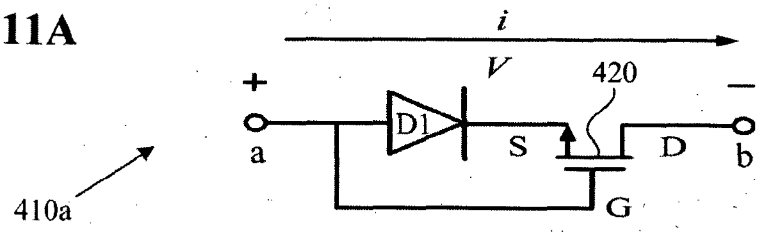
i
V
+
420
—
D1
a
S
D
b
G
410a

FIG. 11B

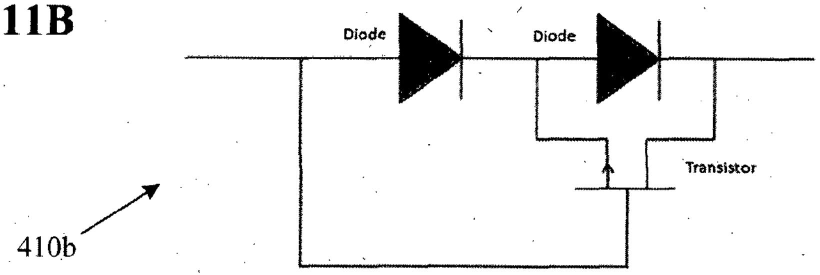
Diode
Diode
Transistor
410b

FIG. 11C

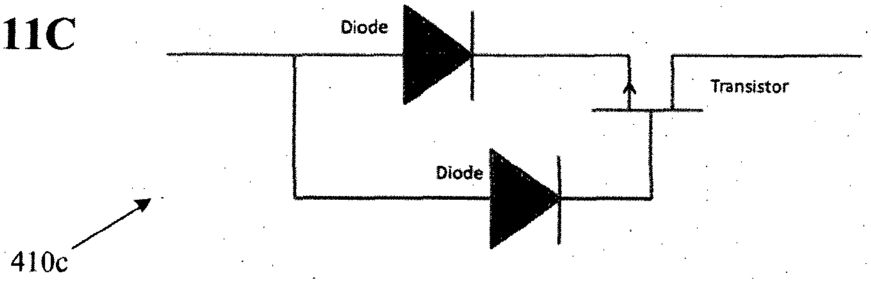
Diode
Transistor
Diode
410c

FIG. 11D

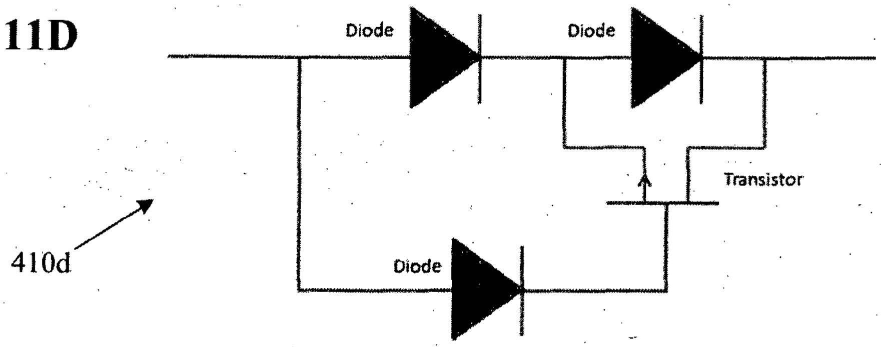
Diode
Diode
Transistor
Diode
410d

FIG. 11E

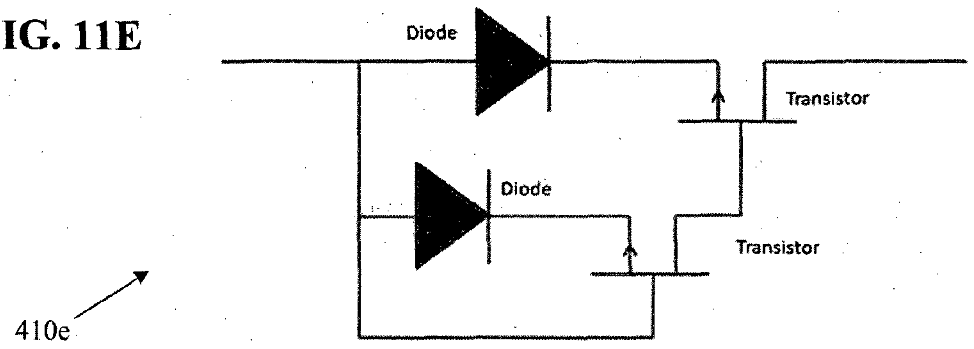
Diode
Transistor
Diode
Transistor
410e

FIG. 11F

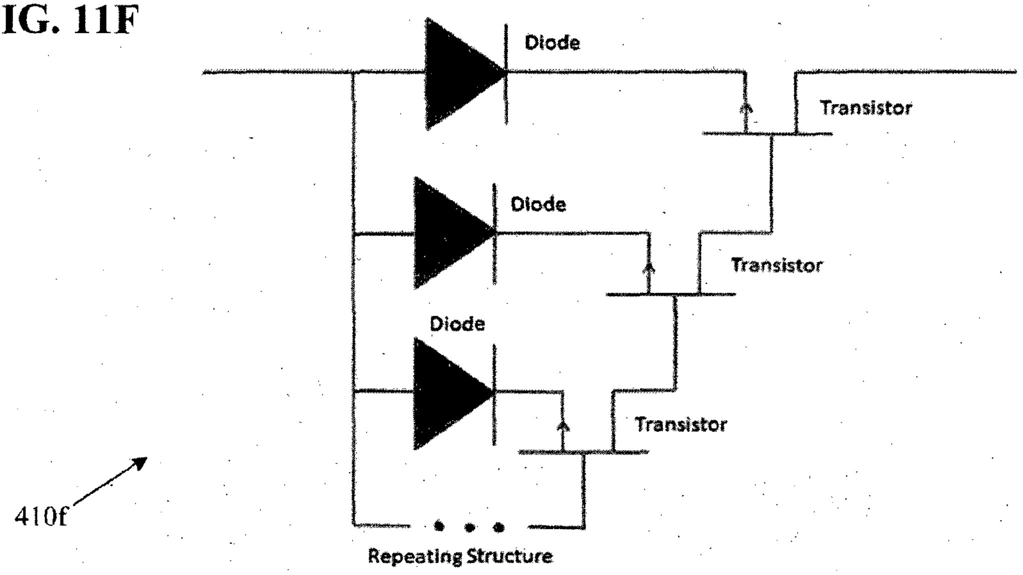
Diode
Transistor
Diode
Transistor
Diode
Transistor
410f
Repeating Structure

FIG. 11G

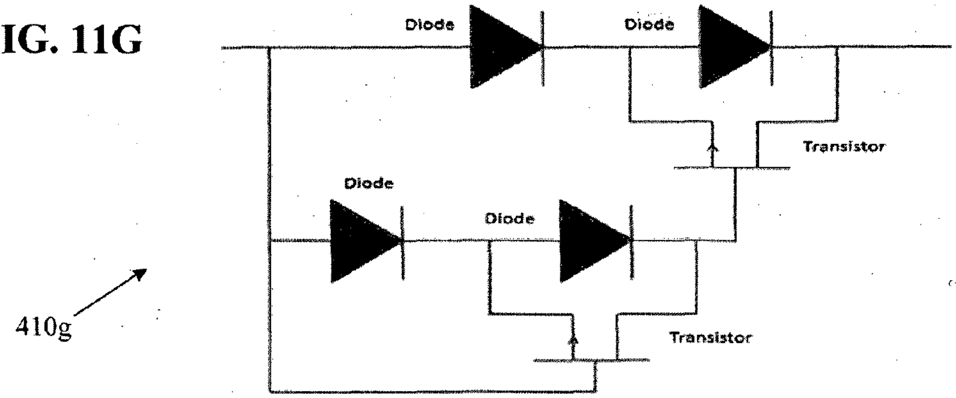
Diode
Diode
Transistor
Diode
Diode
Transistor
410g

FIG. 11H

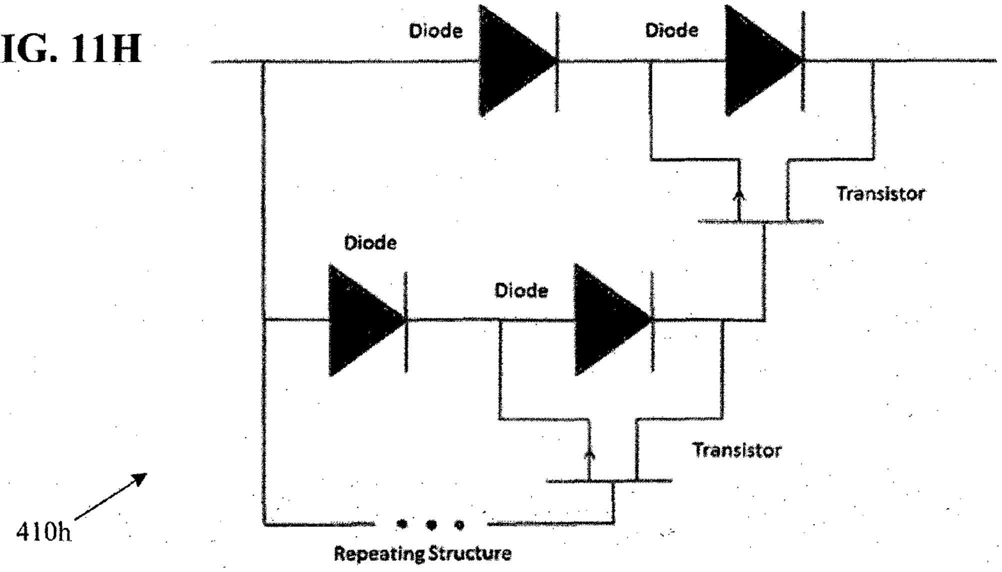
Diode
Diode
Transistor
Diode
Diode
Transistor
Repeating Structure
410h

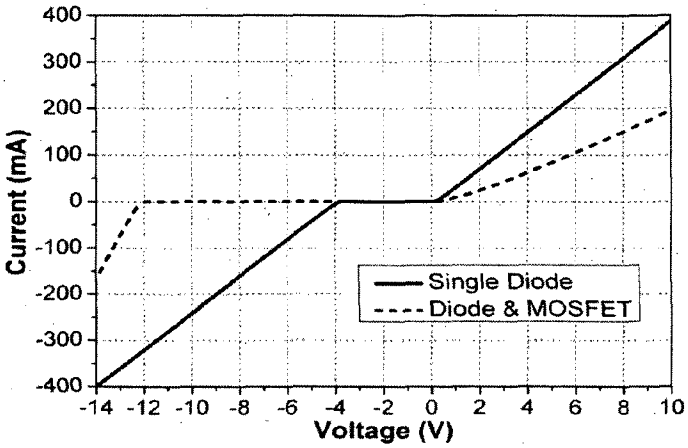
400
300
200
100
0
-100
-200
-300
-400
Current (mA)
-14
-12
-10
-8
-6
-4
-2
0
2
4
6
8
10
Voltage (V)
Single Diode
Diode & MOSFET

FIG. 12

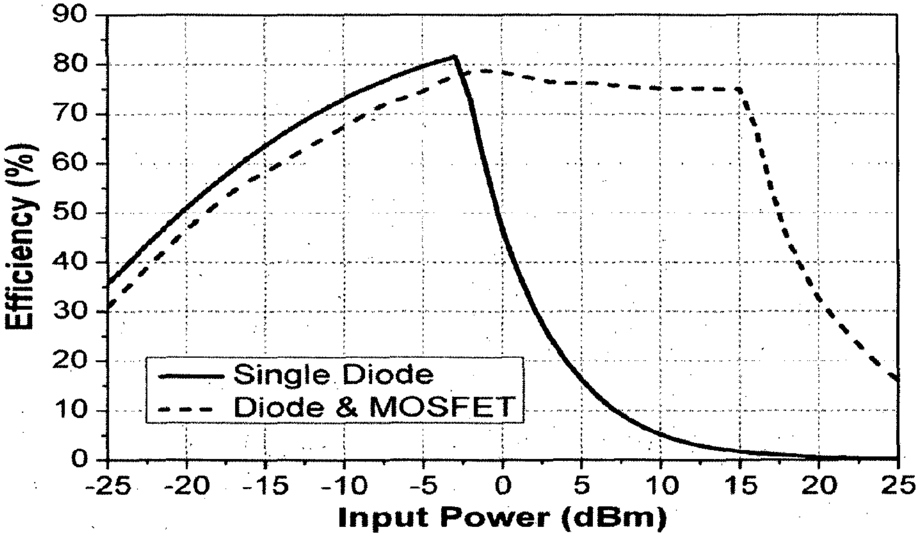
90
80
70
60
50
40
30
20
10
0
Efficiency (%)
-25
-20
-15
-10
-5
0
5
10
15
20
25
Input Power (dBm)
Single Diode
Diode & MOSFET

FIG. 13

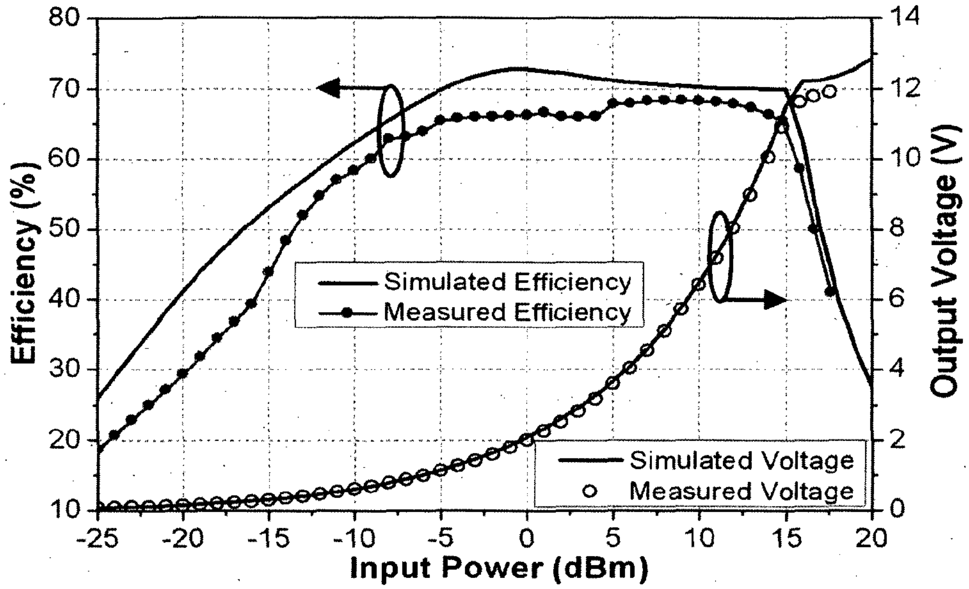
80
70
60
50
40
30
20
10
Efficiency (%)
14
12
10
8
6
4
2
0
Output Voltage (V)
-25
-20
-15
-10
-5
0
5
10
15
20
Input Power (dBm)
Simulated Efficiency
Measured Efficiency
Simulated Voltage
Measured Voltage

FIG. 14

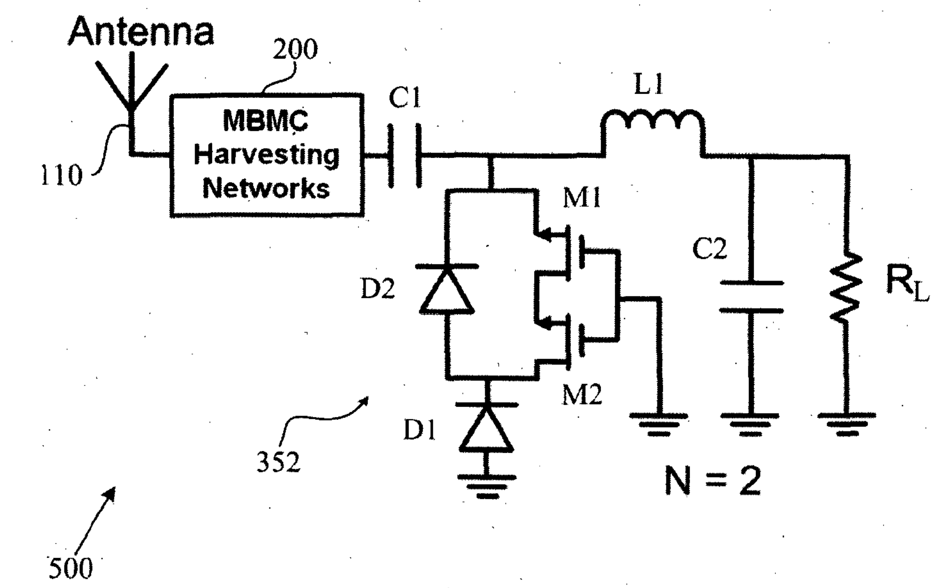
Antenna
200
110
MBMC Harvesting Networks
C1
L1
M1
C2
D2
R_L
M2
D1
352
N = 2
500

FIG. 15A

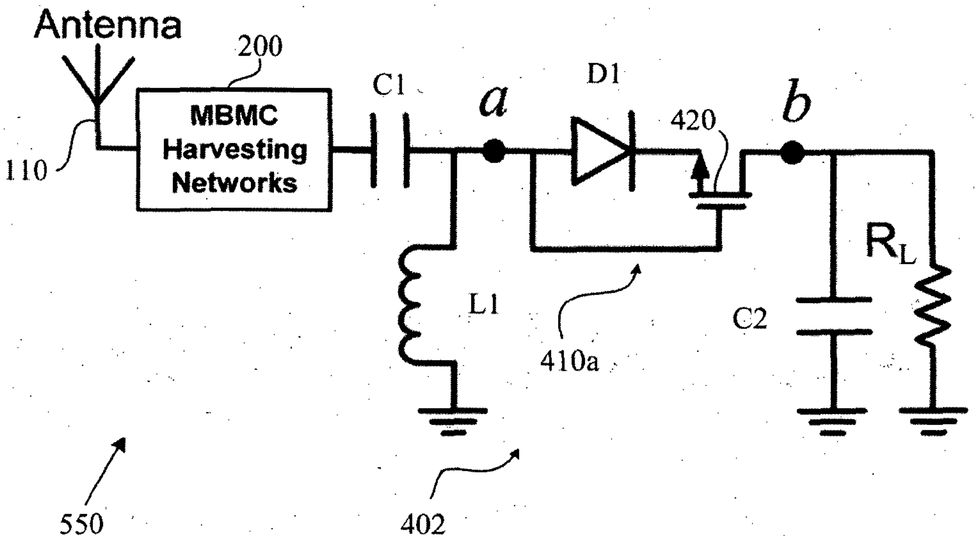
Antenna
200
110
MBMC Harvesting Networks
C1
a
D1
420
b
R_L
L1
410a
C2
550
402

FIG. 15B

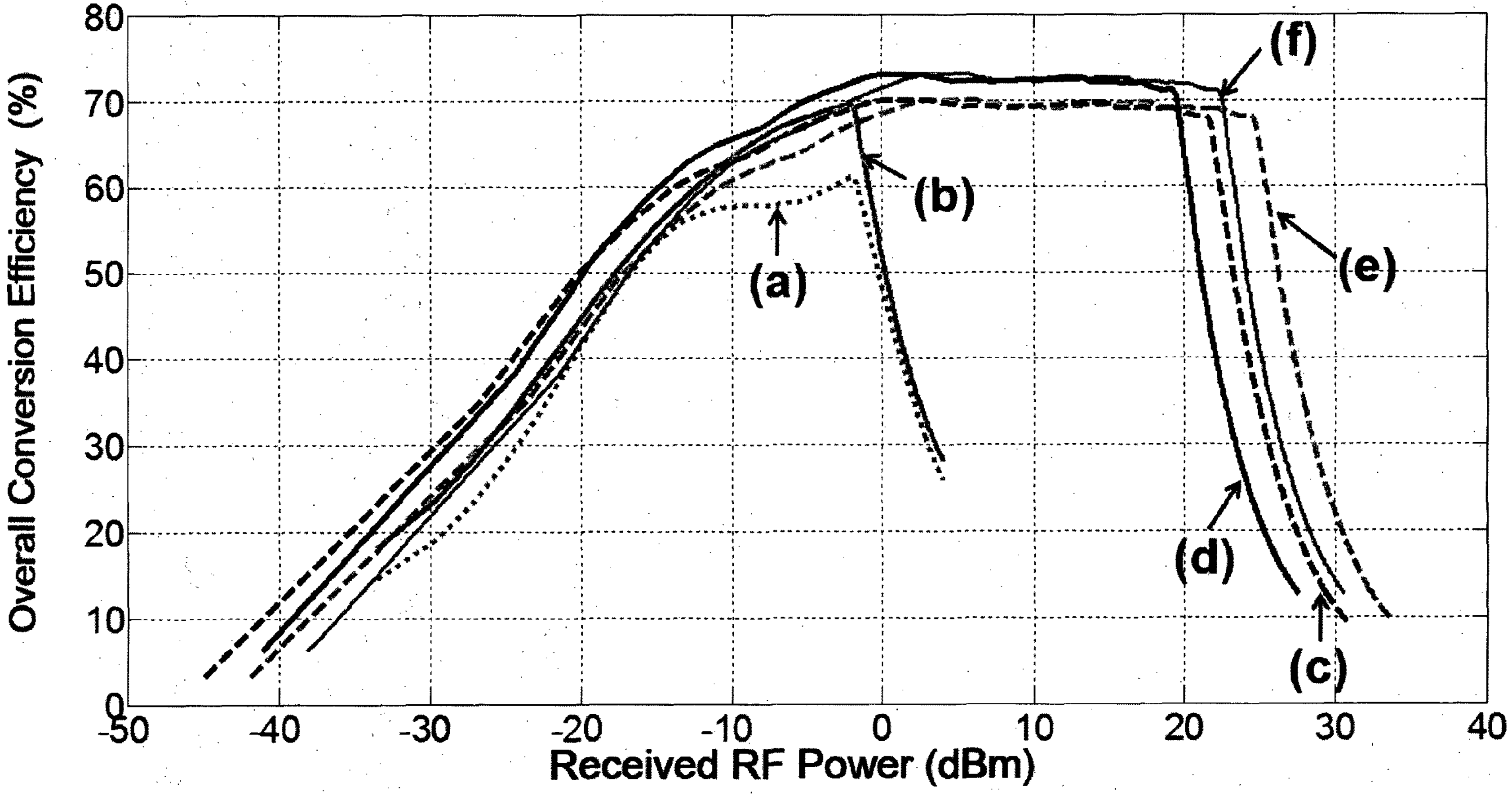
Overall Conversion Efficiency (%)
80
70
60
50
40
30
20
10
0
-50
-40
-30
-20
-10
0
10
20
30
40
Received RF Power (dBm)
(a)
(b)
(c)
(d)
(e)
(f)

FIG. 16

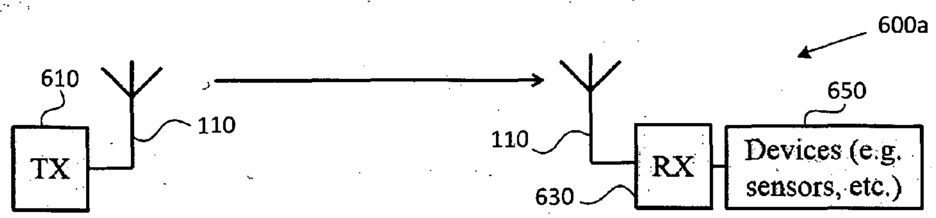
600a
610
TX
110
110
RX
630
650
Devices (e.g. sensors, etc.)

FIG. 17A

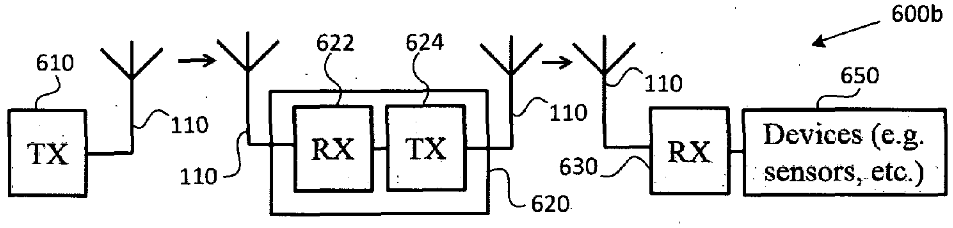
600b
610
TX
110
622
624
RX
TX
110
110
620
110
630
RX
650
Devices (e.g. sensors, etc.)

FIG. 17B

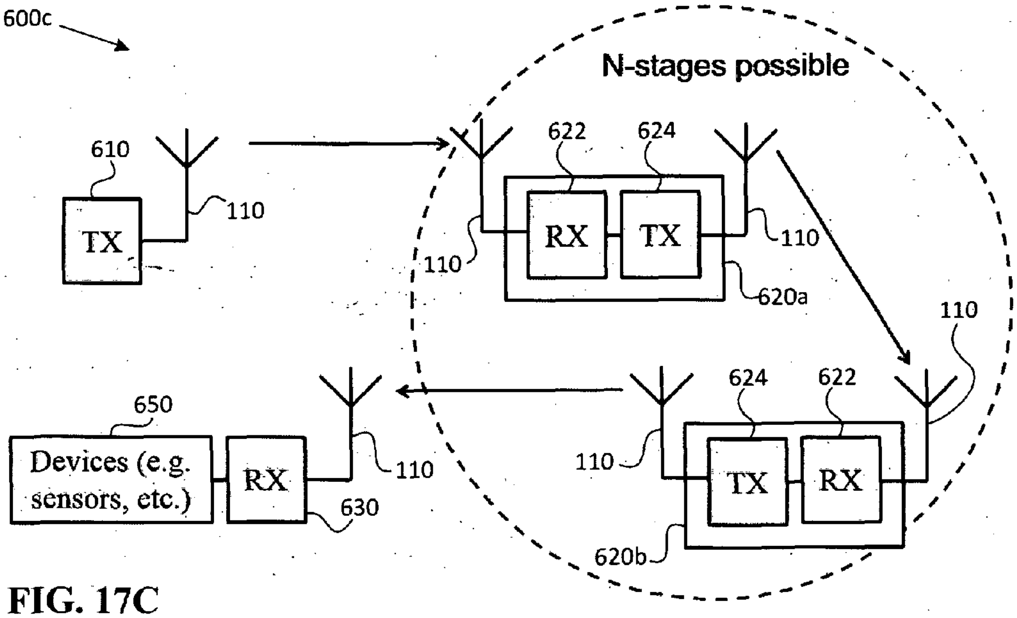
600c
N-stages possible
610
TX
110
622
624
RX
TX
110
110
620a
110
650
Devices (e.g. sensors, etc.)
RX
110
630
624
622
TX
RX
110
620b

FIG. 17C

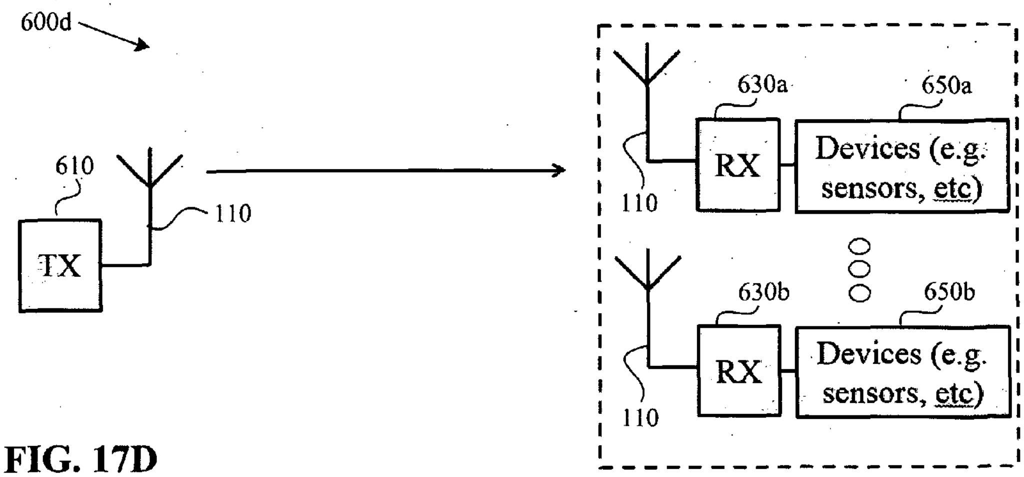
600d
610
TX
110
630a
650a
RX
Devices (e.g. sensors, etc)
110
630b
650b
RX
Devices (e.g. sensors, etc)
110

FIG. 17D

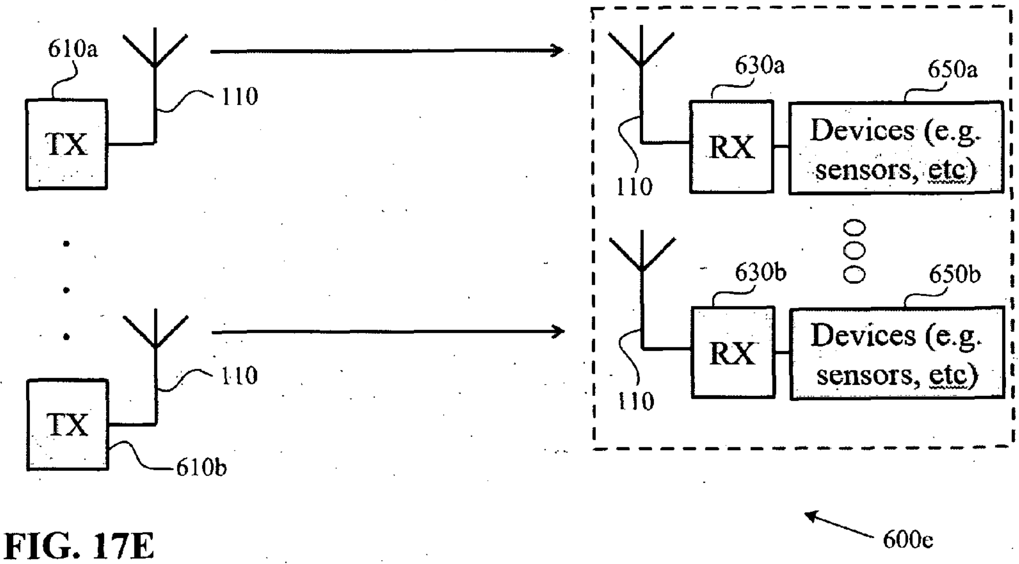
610a
TX
110
110
TX
610b
630a
650a
RX
Devices (e.g. sensors, etc)
110
630b
650b
RX
Devices (e.g. sensors, etc)
110
600e

FIG. 17E

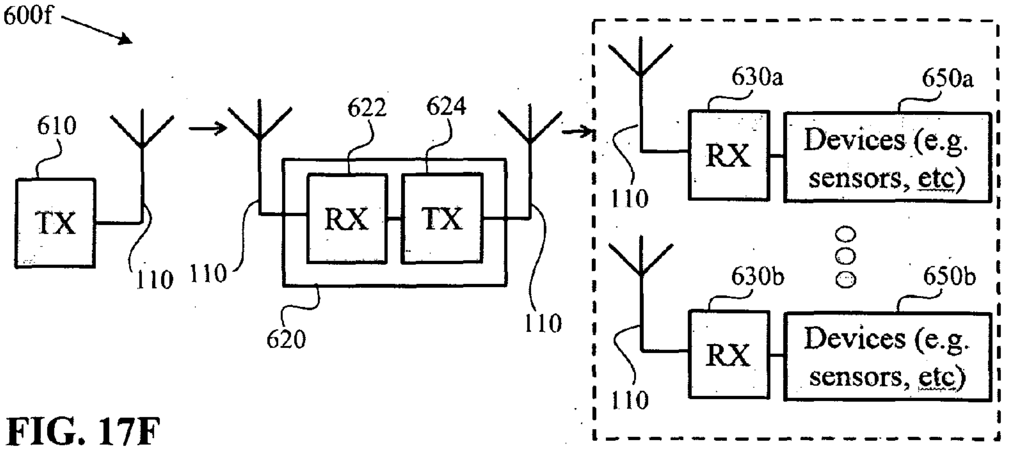
600f
610
TX
110
110
RX
TX
622
624
620
110
630a
RX
650a
Devices (e.g. sensors, etc)
110
630b
RX
650b
Devices (e.g. sensors, etc)
110

FIG. 17F

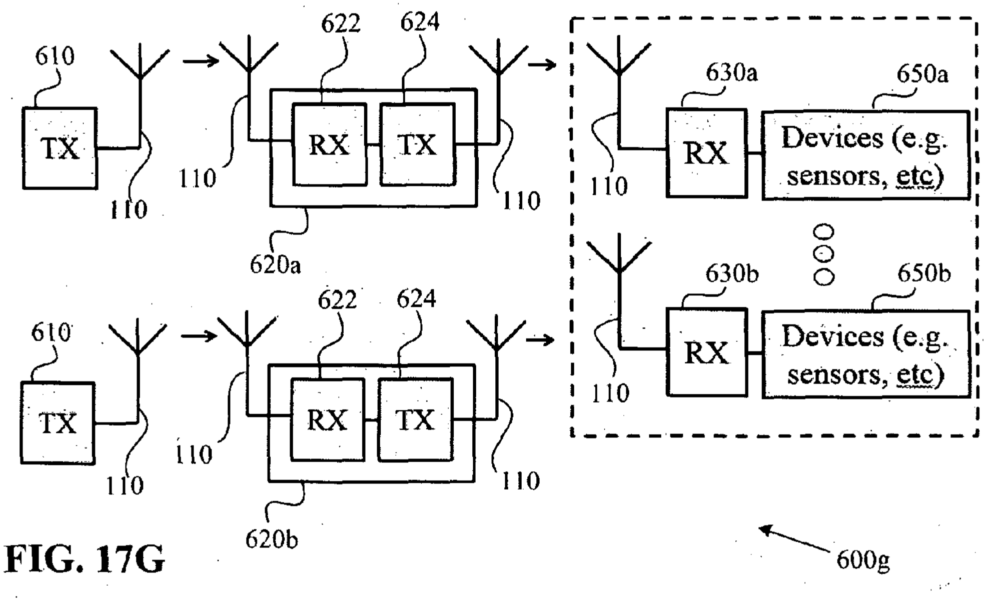
610
TX
110
110
RX
TX
622
624
620a
110
610
TX
110
110
RX
TX
622
624
620b
110
630a
RX
650a
Devices (e.g. sensors, etc)
110
630b
RX
650b
Devices (e.g. sensors, etc)
110
600g

FIG. 17G

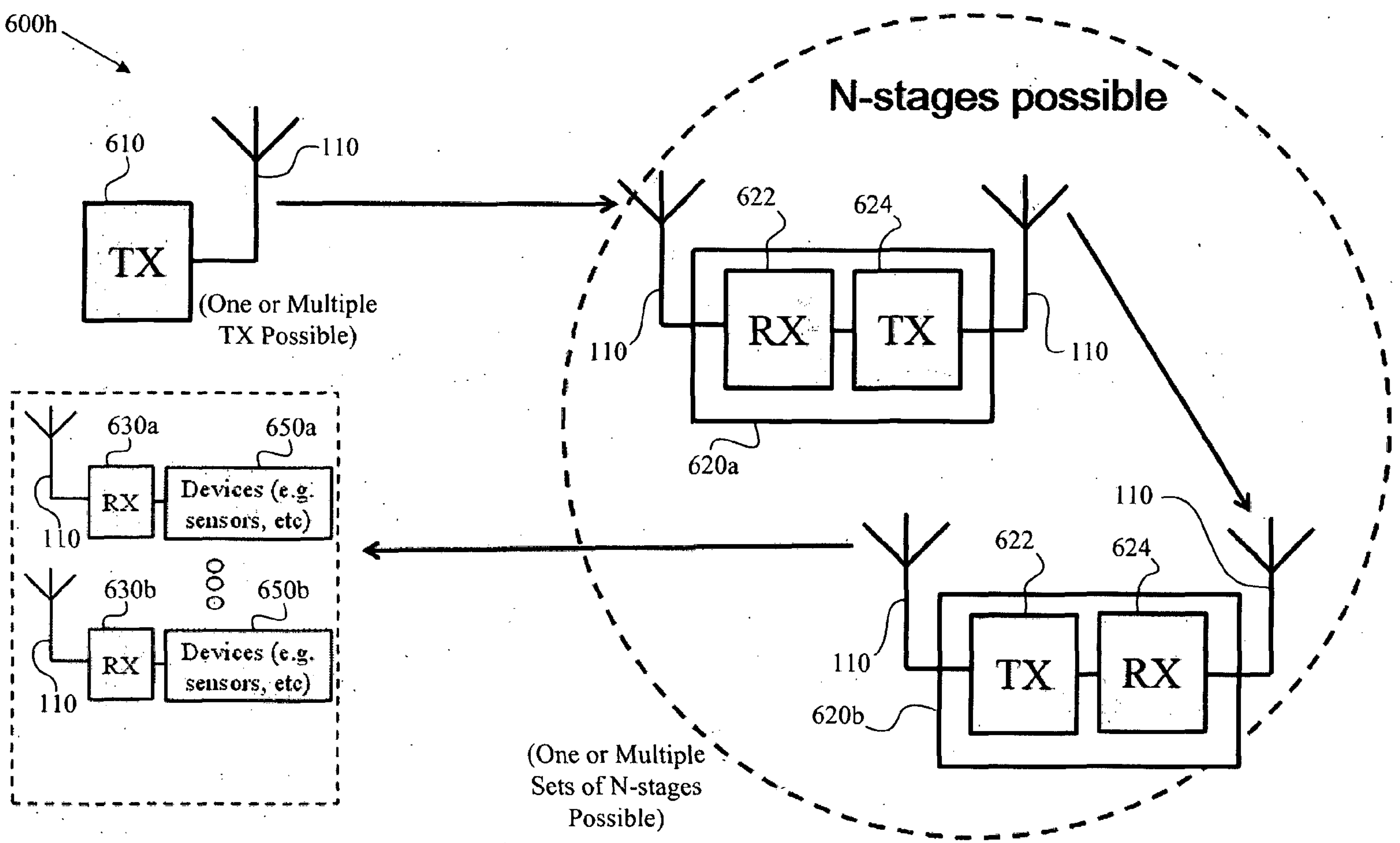
600h
610
TX
110
(One or Multiple
TX Possible)
N-stages possible
622
624
RX
TX
110
110
620a
110
622
624
TX
RX
110
620b
(One or Multiple
Sets of N-stages
Possible)
630a
650a
RX
Devices (e.g.
sensors, etc)
110
630b
650b
RX
Devices (e.g.
sensors, etc)
110

FIG. 17H

RECTENNA CIRCUIT ELEMENTS, CIRCUITS, AND TECHNIQUES FOR ENHANCED EFFICIENCY WIRELESS POWER TRANSMISSION OR AMBIENT RF ENERGY HARVESTING

TECHNICAL FIELD

[0001] Aspects of the present disclosure are directed to rectenna circuit elements, circuits, and techniques for enhancing power conversion efficiency (PCE) in wireless power transmission (WPT) or ambient RF energy harvesting circuits by way of (a) a multi-band multi-channel (MBMC) RF matching network, and/or (b) an extended input power range rectifier, which can include an adaptively reconfigurable rectifier or a breakdown-protected rectifier.

BACKGROUND

[0002] Wireless Power Transfer (WPT) involves the transfer of RF electrical energy from a predetermined source, or particular types or sources, to a receiving antenna that is coupled to a rectifier. The combination of the receiving antenna and the rectifier is known as a rectenna. The rectenna converts received RF energy into DC power, which can be used to power a system or device, such as a sensor or a medical device. Ambient RF energy harvesting involves the collection of RF energy that is inherently present in modern environments as a result of wireless communication between one or more types of wireless systems and devices (e.g., public wireless telecommunication base stations and mobile telephones/smartphones), such that RF energy which would otherwise be wasted can be put to use, such as for battery recharging. As with WPT, ambient RF energy harvesting relies upon a rectenna, which converts harvested RF energy into DC power.

[0003] FIG. **1** is a schematic illustration of a conventional rectenna that is configured for receiving RF energy within a single RF frequency band, and converting RF energy within this single frequency band to DC power. Because this rectenna is limited to a single RF frequency band, RF energies outside of the single frequency band for which the rectenna is designed cannot be converted into DC power, thereby limiting the utility of this conventional rectenna in WPT and/or RF energy harvesting applications. For instance, various types of implantable medical devices, such as implantable pulse generators, have a power source such as a battery that needs to be recharged. For implanted medical devices, instantaneous RF power applied to the human body must be maintained below a predetermined level for safety purposes. Hence, it is desirable to utilize multiple RF frequency bands to apply pulsed RF power in a simultaneous manner for recharging implanted power sources. Additionally, for RF energy harvesting applications, it is desirable to simultaneously harvest RF energy within multiple frequency bands in a simultaneous manner, in order to enhance or maximize the conversion of available ambient RF energy into DC power.

[0004] Multi-frequency rectennas have been developed; however, such rectennas typically utilize multiple antennas and multiple rectifiers, with each antenna and its associated rectifier corresponding to a particular RF frequency band. Consequently, such conventional multi-frequency rectenna designs result in less efficient RF energy conversion, and increased space and cost.

[0005] The RF-to-DC power conversion efficiency of a rectenna critically depends on the nature of the rectifier therein. FIG. **2**A is a schematic illustration of a conventional rectenna based upon a conventional rectifier, which is suitable for low. RF input power conditions. More particularly, for this rectenna, RF-to-DC power conversion efficiency (PCE) increases linearly with input RF power until input RF power exceeds the breakdown voltage of diode D**1**, after which RF-to-DC PCE degrades rapidly.

[0006] FIG. **2**B is a schematic illustration of a conventional rectenna that utilizes a conventional high RF input power rectifier. For the rectenna of FIG. **2**B, a set of diodes D**2** (e.g., a total of four diodes, arranged pairwise in parallel) is coupled in series with diode D**1**, thereby extending the overall breakdown voltage of the combination of diode D**1** and the set of diodes D**2**. Unfortunately, this rectenna exhibits undesirably low RF-to-DC PCE at low input RF power levels.

[0007] FIG. **2**C is a schematic illustration of conventional rectenna that utilizes another type of conventional low RF input power rectifier. Here again, once the input RF power exceeds the breakdown voltage of diode D**1**, the RF-to-DC PCE of this rectenna degrades rapidly.

[0008] A need exists for a rectenna that utilizes a single antenna and a single rectifier, and which exhibits enhanced RF-to-DC PCE across a significantly wider range of input RF operating power conditions compared to prior rectennas.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] FIG. **1** is a schematic illustration of a conventional rectenna

[0010] FIG. **2**A is a schematic illustration of a conventional rectenna that utilizes a conventional low RF input power rectifier.

[0011] FIG. **2**B is a schematic illustration of a conventional rectenna that utilizes a conventional high RF input power rectifier.

[0012] FIG. **2**C is a schematic illustration of a conventional rectenna that utilizes another conventional low RF input power rectifier.

[0013] FIG. **3** is a schematic illustration of a representative multi-band multi-channel (MBMC) rectenna, which includes an MBMC matching network in accordance with an embodiment of the present disclosure.

[0014] FIG. **4**A is a top schematic illustration of a quasi-Yagi antenna 1×2 subarray in accordance with an embodiment of the present disclosure.

[0015] FIG. **4**B is a front or top side image of a 1×4 quasi-Yagi antenna array having a bandwidth that covers each of the GSM-1800 and UTMS-2100 frequency bands, which is formed by connecting two quasi-Yagi 1×2 antenna subarrays of FIG. **4**A together with a T-junction power divider in accordance with an embodiment of the present disclosure.

[0016] FIG. **4**C is a back or bottom side image of the 1×4 quasi-Yagi antenna array of FIG. **4**B.

[0017] FIG. **5**A is a schematic illustration of a representative embodiment of an MBMC matching network within an MBMC rectenna in accordance with an embodiment of the present disclosure.

[0018] FIG. **5**B is an image of an as-fabricated representative implementation of this MBMC matching network in accordance with an embodiment of the present disclosure.

[0019] FIG. **6** plots measured RF-to-DC PCE against frequency at different power levels (−30 dBm to −18 dBM) for a representative implementation of the MBMC rectenna of FIG. **3**.

[0020] FIG. **7** plots measured PCE against single and dual-tone input power for input signal frequencies of 1.83 GHz and 2.14 GHz, which are the approximate center frequencies of the 1.80-1.88 GHz and 2.11-2.17 GHz frequency bands, for the representative implementation of the MBMC rectenna of FIG. **3**.

[0021] FIG. **8**A is a schematic illustration of an adaptively reconfigurable rectenna or adaptive rectenna based upon a first adaptively reconfigurable rectifier or first adaptive rectifier in accordance with an embodiment of the present disclosure.

[0022] FIG. **8**B is a schematic illustration of an adaptively reconfigurable rectenna or adaptive rectenna based upon a second adaptively reconfigurable rectifier or second adaptive rectifier in accordance with an embodiment of the present disclosure.

[0023] FIG. **8**C illustrates a circuit configuration of a representative implementation of an adaptively reconfigurable rectifier shown FIG. **8**B.

[0024] FIG. **8**D is an image of the representative second adaptively reconfigurable rectifier of FIG. **8**C, as-fabricated.

[0025] FIG. **9**A is a graph showing RF-to-DC PCE simulation results for (a) the conventional low RF input power rectifier shown in FIG. **2**A; (b) the conventional high RF input power rectifier shown in FIG. **2**B; (c) the first adaptively reconfigurable rectifier shown in FIG. **8**A; and (d) the second adaptively reconfigurable rectifier shown in FIG. **8**B.

[0026] FIG. **9**B is a graph showing measured versus simulated RF-to-DC PCE, as well as measured versus simulated output voltage, as a function of input RF power for a representative implementation of the second adaptively reconfigurable rectifier of FIGS. **8**B-**8**D.

[0027] FIG. **10**A is a schematic illustration of a breakdown-protected rectenna based upon a breakdown-protected rectifier in accordance with an embodiment of the present disclosure.

[0028] FIG. **10**B is a schematic illustration of a representative implementation of the breakdown-protected rectifier of FIG. **10**A.

[0029] FIG. **10**C is an image of the as-fabricated representative implementation of the breakdown-protected rectifier of FIG. **10**B.

[0030] FIGS. **11**A-**11**H are schematic illustrations of a first through a seventh transistor-protected diode structure, respectively, in accordance with an embodiment of the present disclosure.

[0031] FIG. **12** is a graph illustrating I-V characteristics for an individual diode D1 and the first transistor-protected diode structure **410***a* of FIG. **11**A.

[0032] FIG. **13** is a graph illustrating simulated RF-to-DC PCE versus input RF power level for the rectifier shown in FIG. **2**C as well as the breakdown-protected rectifier of FIG. **10**A.

[0033] FIG. **14** is a graph showing measured versus simulated RF-to-DC PCE, as well as measured versus simulated output voltage, as a function of input RF power for the representative implementation of the breakdown-protected rectifier of FIGS. **10**B and **10**C.

[0034] FIG. **15**A illustrates a rectenna having each of an MBMC matching network and a second adaptively reconfigurable rectifier in accordance with an embodiment of the present disclosure.

[0035] FIG. **15**B illustrates a rectenna having each of an MBMC matching network and a breakdown-protected rectifier in accordance with an embodiment of the present disclosure.

[0036] FIG. **16** is a graph illustrating simulated RF-to-DC PCE versus input RF power for representative implementations of each of (a) the conventional low input RF power rectifier of FIG. **2**A; (b) the MBMC rectifier of FIG. **3**; (c) the second adaptively reconfigurable rectifier shown in FIG. **8**B; (d) the MBMC matching network coupled to the second adaptively reconfigurable rectifier **352**, in a manner shown in FIG. **15**A; (e) the breakdown-protected rectifier of FIG. **10**A; and (f) the MBMC matching network coupled to the breakdown-protected rectifier **402**, in a manner shown in FIG. **15**B.

[0037] FIGS. **17**A-**17**H are block diagrams illustrating representative types of wireless networks in which a set of rectennas, including one or more MBMC rectennas, adaptively reconfigurable rectennas, and/or breakdown-protected rectennas, in accordance with an embodiment of the present disclosure can be deployed.

SUMMARY

[0038] In accordance with an aspect of the present disclosure, a circuit configured for receiving RF energy distributed in multiple RF frequency bands includes a multi-band multichannel (MBMC) matching network. The MBMC matching network includes: a plurality of T-shaped transmission line matching structures, each T-shaped transmission line matching structure coupled in series to another T-shaped transmission line matching structure, wherein each T-shaped transmission line matching structure includes a first portion, a second portion, and a third portion coupled between the first portion and the second portion, and wherein the first portion of a first T-shaped transmission line matching structure within the plurality of T-shaped transmission line structures forms an input of the MBMC matching network, the second portion of a second T-shaped transmission line matching structure within the plurality of T-shaped transmission line matching structures forms an output of the MBMC matching network, and the third portion of at least one T-shaped transmission line matching structure within the plurality of T-shaped transmission line matching structures is couplable to electrical ground.

[0039] Each individual T-shaped transmission line matching structure within the MBMC matching network is configured for providing an impedance match corresponding to RF frequencies within a particular RF frequency band. The circuit can form a portion of a wireless power transmission (WPT) system or a portion of an ambient RF energy harvesting system.

[0040] The circuit can include or be a rectenna, which includes: an antenna coupled, to the input of the MBMC matching network; and a rectifier coupled to the output of the MBMC matching network. The rectifier includes one of: (a) an adaptively reconfigurable rectifier having a low input RF power rectifying portion, a high input RF power rectifying portion, and a set of transistors coupled to one of the high input RF power rectifying portion the low input RF power rectifying portion, the set of transistors configured for automatically transitioning the adaptively reconfigurable rectifier between a low input RF input power operating configuration and a high input RF power operating configuration based upon input RF power level; and (b) a breakdown-protected rectifier having at least transistor-protected diode structure, the transistor-protected diode structure comprising at least one diode coupled to a transistor in a manner that protects the diode from direct exposure to negative voltages that would ordinarily cause the diode to break down in the absence of the transistor.

[0041] In accordance with another aspect of the present disclosure, a circuit configured for rectifying signals (e.g., wirelessly received RF signals) includes one of: (a) an adaptively reconfigurable rectifier having a low input power rectifying portion, a high input power rectifying portion, and a set of transistors coupled to one of the high input power rectifying portion the low input power rectifying portion, the set of transistors configured for selectively and automatically transitioning the adaptively reconfigurable rectifier between a low input power operating configuration and a high input RF power operating configuration in a manner correlated with a power level of a signal input to the circuit; and (b) a breakdown-protected rectifier having at least one transistor-protected diode structure, the at least one transistor-protected diode structure comprising a first diode coupled to a transistor in a manner that protects the first diode from direct exposure to negative voltages that would ordinarily cause the diode to break down in the absence of the transistor.

[0042] In various embodiments, the low input power portion and the high input power portion of the adaptively reconfigurable rectifier are coupled in series, and the set of transistors of the adaptively reconfigurable is be coupled in parallel with the high input power portion. A control terminal of each transistor within the set of transistors is coupled to electrical ground. The set of transistors of the adaptively reconfigurable rectifier can include a plurality of transistors coupled to each other in series. Each transistor within the set of transistors can be a Field Effect Transistor (FET) (e.g., a MOSFET, a MESFET, or another type of FET).

[0043] In various embodiments, the low input power rectifying portion of the adaptively reconfigurable rectifier includes a first set of diodes, and the high input power rectifying portion includes a second set of diodes. At least one of the first set of diodes and the second set of diodes can include a plurality of diodes.

[0044] The breakdown-protected rectifier includes a first transistor-protected diode structure including a first diode and a transistor, wherein the transistor includes a control terminal coupled to the anode of the first diode, a first terminal coupled to the cathode of the first diode, and a second terminal couplable to circuitry external to the transistor-protected diode structure. The anode of the first diode is coupled to an input of the breakdown-protected rectifier and is configured to receive an alternating current (AC) signal, and wherein the second terminal of the transistor is coupled to an output of the breakdown-protected rectifier.

[0045] The transistor-protected diode structure can further include a second diode coupled in parallel with the second terminal and the third terminal of the transistor, such that the anode of the second diode is coupled to the cathode of the first diode. The transistor-protected diode structure can additionally or alternatively includes another diode having its anode coupled to the anode of the first diode, and its cathode coupled to the control terminal of the transistor.

[0046] In some embodiments, the breakdown-protected rectifier includes the first transistor-protected diode structure combined with a second transistor-protected diode structure coupled in a cascaded manner to the first transistor-protected diode structure, such that a negative voltage across the combination of the first and second transistor-protected diode structures is distributed across the transistor of the first transistor-protected diode structure and a transistor of the second transistor-protected diode structure. The second transistor-protected diode structure includes a first diode and a transistor having a control input, a first terminal, and a second terminal, wherein the anode of the first diode of the second transistor-protected diode structure is coupled to the anode of the first diode of the first transistor-protected diode structure, the cathode of the first diode of the second transistor-protected diode structure is coupled to the second terminal of the transistor of the second diode-protected transistor structure, and the third terminal of the transistor of the second diode-protected transistor structure is coupled to the control input of the transistor of the first diode-protected transistor structure.

[0047] In accordance with another aspect of the present disclosure, a wireless network includes: a transmission (TX) stage configured for wirelessly transmitting electromagnetic signals; and a reception (RX) stage configured for receiving electromagnetic signals transmitted by the TX stage. The RX stage includes one of: (a) a multi-band multi-channel (MBMC) matching network having: a plurality of T-shaped transmission line matching structures, each T-shaped transmission line matching structure coupled in series to another T-shaped transmission line matching structure, wherein each T-shaped transmission line matching structure includes a first portion, a second portion, and a third portion coupled between the first portion and the second portion, and wherein the first portion of a first T-shaped transmission line matching structure within the plurality of T-shaped transmission line structures forms an input of the MBMC matching network, the second portion of a second T-shaped transmission line matching structure within the plurality of T-shaped transmission line matching structures forms an output of the MBMC matching network, and the third portion of at least one T-shaped transmission line matching structure within the plurality of T-shaped transmission line matching structures is couplable to electrical ground; (b) an adaptively reconfigurable rectifier having a low input power rectifying portion, a high input power rectifying portion, and a set of transistors coupled to one of the high input power rectifying portion the low input power rectifying portion, the set of transistors configured for selectively and automatically transitioning the adaptively reconfigurable rectifier between a low input power operating configuration and a high input RF power operating configuration in a manner correlated with a power level of a signal input to the RX stage; and (c) a breakdown-protected rectifier having at least one transistor-protected diode structure, the at least one transistor-protected diode structure comprising a first diode coupled to a transistor in a manner that protects the first diode from direct exposure to negative voltages that would, ordinarily cause the diode to break down in the absence of the transistor. In some embodiments, the RX stage includes the MBMC matching network and one of the adaptively reconfigurable rectifier and the breakdown-protected rectifier.

[0048] The wireless network can further include one or more repeater/amplification stages configured to receive electromagnetic signals from the TX stage and transmit electromagnetic signals to the RX stage. One or more of the repeater/amplification: stages include one of: (a) a multi-band multi-channel (MBMC) matching network having: a plurality of T-shaped transmission line matching structures, each T-shaped transmission line matching structure coupled in series to another T-shaped transmission line matching structure, wherein each T-shaped transmission line matching structure includes a first portion, a second portion, and a third portion coupled between the first portion and the second portion, and wherein the first portion of a first T-shaped transmission line matching structure within the plurality of T-shaped transmission line structures forms an input of the MBMC matching network, the second portion of a second T-shaped transmission line matching structure within the plurality of T-shaped transmission line matching structures forms an output of the MBMC matching network, and the third portion of at least one T-shaped transmission line matching structure within the plurality of T-shaped transmission line matching structures is couplable to electrical ground; (b) an adaptively reconfigurable rectifier having a low input power rectifying portion, a high input power rectifying portion, and a set of transistors coupled to one of the high input power rectifying portion the low input power rectifying portion, the set of transistors configured for selectively and automatically transitioning the adaptively reconfigurable rectifier between a low input power operating configuration and a high input RF power operating configuration in a manner correlated with a power level of a signal input to the RX stage; and (c) a breakdown-protected rectifier having at least one transistor-protected diode structure, the at least one transistor-protected diode structure comprising a first diode coupled to a transistor in a manner that protects the first diode from direct exposure to negative voltages that would ordinarily cause the diode to break down in the absence of the transistor. In some embodiments, at least one repeater/amplification stage includes an MBMC matching network and one of an adaptively reconfigurable rectifier and a breakdown-protected rectifier.

DETAILED DESCRIPTION

[0049] In the present disclosure, depiction of a given element or consideration or use of a particular element number in a particular FIG. or a reference thereto in corresponding descriptive material can encompass the same, an equivalent, or an analogous element or element number identified in another FIG. or descriptive material associated therewith. The use of "/" in a FIG. or associated text is understood to mean "and/or" unless otherwise indicated. The recitation of a particular numerical value or value range herein is understood to include or be a recitation of an approximate numerical value or value range, for instance, to within +/−20%, +/−15%, +/−10%, or +/−5%.

[0050] As used herein, the term "set" corresponds to or is defined as a non-empty finite organization of elements that mathematically exhibits a cardinality of at least 1 (i.e., a set as defined herein can correspond to a unit, singlet, or single element set, or a multiple element set), in accordance with known mathematical definitions (for instance, in a manner corresponding to that described in *An Introduction to Mathematical Reasoning: Numbers, Sets, and Functions*, "Chapter 11: Properties of Finite Sets" (e.g., as indicated on p. 140), by Peter J. Eccles, Cambridge University Press (1998)). In general, an element of a set can include or be a system, an apparatus, a device, a structure, an object, a circuit, a circuit element, a process, a physical parameter, or a value depending upon the type of set under consideration.

[0051] Embodiments in accordance with the present disclosure are directed to rectenna circuit elements, circuits, and techniques that facilitate or enable enhanced power conversion efficiency (PCE) in wireless power transmission (WPT) or ambient RF energy harvesting systems, devices, or circuits. More particularly, systems, devices, or circuits in accordance with various embodiments of the present disclosure include rectennas having (a) a multi-band multi-channel (MBMC) RF matching network, and/or (b) an extended input power range rectifier, which can include an adaptively reconfigurable rectifier circuit, or a breakdown-protected rectifier circuit, as further detailed below.

[0052] Representative Rectenna with MBMC Matching Network

[0053] FIG. **3** is a schematic illustration of a representative multi-band multi-channel (MBMC) rectenna **100** in accordance with an embodiment of the present disclosure. The MBMC rectenna **100** includes an antenna or antenna structure **110** configured for receiving RF energy corresponding to or spanning multiple RF frequency bands. The antenna **100** includes an output coupled to an input of an MBMC matching network **200**, such that RF energy received by the antenna **110** is delivered to the MBMC matching network **200**. The MBMC matching network **200** includes an output coupled to an input of a rectifier **120**, which has an output coupled to an input of a DC pass filter or RF choke **130**, and which further has an output coupled to a load resistance **140**. Taken together, the MBMC matching network **200** and the rectifier **120**, possibly in combination with the DC pass filter **130** and the load resistance **140**, can be considered as or defined to be an MBMC rectifier **102**.

[0054] The antenna **110** has a bandwidth spanning multiple RF frequency bands, for instance, multiple predetermined frequency bands suitable for a particular WPT application or an intended or expected RF energy harvesting environment under consideration. For instance, depending upon embodiment details, the antenna **110** can have a bandwidth suitable for receiving RF energy spanning multiple center frequencies across multiple frequency bands that are suitable or allocated for (a) recharging one or more types of implantable/implanted medical devices, such that the MBMC rectenna **100** can form a portion of a system, device, or circuit for WPT based medical device recharging; and/or (b) public/mobile telephony or wireless network communications, such that the MBMC rectenna **100** can form a portion of a system, device, or circuit for harvesting ambient RF energy corresponding to mobile telephony or other types of wireless communication signals that are known or expected to exist in a given modern urban environment.

[0055] To aid understanding, in a representative modern urban environment at the National University of Singapore campus, ambient power density was measured by way of a conventional wideband horn antenna (0.8-18 GHz) and a conventional spectrum analyzer. It was found that the ambient power density was dominated by downlink channel power in three public communication bands, namely, GSM-900, GSM-1800, and UTMS-2100, with corresponding measured values given in Table 1

TABLE 1

Measured Ambient Power Density across Each Channel of Different Public Telecommunication Bands

Band	Downlink Frequency (MHz)	Received Power (dBm)	Antenna Gain (dBi)	Power Density ($\mu W/m^2$)
GSM-900	925-960	−40~−30	2	7.1~71.4
GSM-1800	1805-1880	−30~−20	10	45.2~452.3
UTMS-2100	2110-2170	−30~−20	10	61.6~615.8

[0056] Furthermore, Table 1 indicates that the power densities in the GSM-1800 and UTMS-2100 bands were significantly larger than the power density in the GSM-900 band. Additionally, the total RF power of all ambient communication bands that the horn antenna received was measured by a power meter, which indicated a total ambient RF power of approximately −20 to −17 dBm (or approximately 10-20 μW).

[0057] To further aid understanding, in a representative implementation directed to capturing RF power in the GSM-1800 and UTMS-2100 bands, the antenna **110** has a bandwidth that covers the Global System for Mobile (GSM) 1800 MHz frequency band, as well as the Universal Mobile Telecommunication System (UMTS) 2100 MHz frequency band. FIG. **4**A is a top schematic illustration of a quasi-Yagi antenna 1×2 subarray, which indicates corresponding quasi-Yagi 1×2 antenna subarray design parameters. Two quasi-Yagi 1×2 antenna subarrays can be connected together with a T-junction power divider to form a 1×4 quasi-Yagi antenna array.

[0058] FIG. **4**B is a front or top side image of a 1×4 quasi-Yagi antenna array, and FIG. **4**C is a back or bottom side image of the 1×4 quasi-Yagi antenna array in accordance with an embodiment of the present disclosure, which forms an antenna **110** having a bandwidth that covers each of the GSM-1800 and UMTS-2100 bands. More particularly, using values for the 1×2 quasi-Yagi antenna sub-array parameters indicated in FIG. **4**A, in millimeters, of a=190, b=100, W_1=1.2, W_2=4.4, W_{gnd}=20, L_{dri}=37, W_{dri}=8, L_{dir}=50, W_{dri}=0.5, S_{dir}=26, S_{dri}=24, and d=100, the 1×4 quasi-Yagi antenna array exhibits a −10 dB bandwidth from approximately 1.8 GHz to 2.2 GHz as indicated in FIG. **4**D, which illustrates simulated and measured $|S_{11}|$ and gain of this representative implementation of the antenna **110**. An individual having ordinary skill in the relevant art will readily recognize that the antenna **110** can exhibit other parameter values or other types of designs, based upon an intended rectenna application or environment under consideration, and a plurality of RF frequency bands associated therewith. For instance, in an RF energy harvesting application, the antenna **110** can be a conventional broadband or wideband antenna, such as a wideband horn antenna.

[0059] The MBMC matching network **200** is configured for providing an impedance match between the antenna **110** and the rectifier **120** with respect to multiple RF frequency bands that fall within the antenna's bandwidth, for purpose of enhancing or maximizing the transfer of RF power within multiple RF frequency bands from the antenna **110** to the rectifier **120**. More particularly, the MBMC matching network **200** is configured for providing an impedance match relative to multiple predetermined RF frequency bands (e.g., a plurality of RF frequency bands defined or set aside for a given purpose by a standards or regulatory agency/body) that fall within the bandwidth of the antenna **110**. As a result, the MBMC impedance matching network **200** enhances the amount of RF power across multiple RF frequency bands that can be coupled into the rectifier **120**.

[0060] In various embodiments, the MBMC matching network **200** includes a plurality of T-shaped transmission line matching structures, which are coupled in series between the input and the output of the MBMC matching network **200**. More particularly, in several embodiments, the MBMC matching network **200** includes an individual T-shaped transmission line matching structure corresponding to each individual RF frequency band for which the MBMC matching network **200** is intended to provide an impedance match. Consequently, an MBMC matching network **200** in accordance with an embodiment of the present disclosure can easily and exactly perform complex MBMC impedance transforming/matching in view of a given set of RF frequency bands and impedance matching requirements under consideration. Furthermore, MBMC matching networks **200** in accordance with embodiments of the present disclosure can realize frequency-dependent complex impedance matching at any two or more arbitrary frequencies. Several (e.g., two, three, or more) T-shaped transmission line matching structures can be cascaded (e.g., serially coupled to each other) in order to satisfy or solve MBMC impedance matching requirements. Such combinations of T-shaped transmission line matching structures provide highly flexible MBMC impedance matching solutions with respect to any given complex impedance matching situation at hand. The concept of these combinations is simple and surprisingly effective, which facilitates the applicability and extendibility of embodiments in accordance with the present disclosure to the use of closed-form equations for such cascaded MBMC impedance matching topologies.

[0061] At least one of the T-shaped transmission line matching structures is additionally coupled to electrical ground to provide a DC path. To further aid understanding, an MBMC matching network embodiment that is configured to provide matching for two RF frequency bands (e.g., the GSM-1800 and UTMS-2100 frequency bands), and which correspondingly includes two T-shaped transmission line matching structures, is described in detail below.

[0062] FIG. **5**A is a schematic illustration of a representative embodiment of an MBMC matching network **200** within an MBMC rectenna **100** in accordance with an embodiment of the present disclosure, where the MBMC matching network **200** is configured to provide impedance matching with respect to two RF frequency bands. Thus, this MBMC matching network **200** correspondingly includes a first **210** and a second **220** T-shaped transmission line matching structure. As indicated in FIG. **5**A, each T-shaped transmission line matching structure **210**, **220** includes three transmission lines. More specifically, the first T-shaped transmission line matching structure **210** includes (a) a first transmission line TL**1** that forms a first or left portion, branch, leg, or side of the first T-shaped transmission line matching structure **210**, and which provides an input of the MBMC matching network **200** that is coupled to the antenna **110**; (b) a second transmission line TL**2** coupled in series to the first transmission line TL**1** in a manner that forms a second or right portion, branch, leg, or side of the first T-shaped transmission line matching structure **210**, and which provides an output of the first T-shaped transmission line matching structure **210**; and (c) a third transmission line TL**3** coupled in parallel between the first transmission line TL**1** and the second transmission line TL**2**, which forms a third portion, branch, leg, or side of the first T-shaped transmission line matching structure. In an analogous manner, the second T-shaped transmission line matching structure **220** includes (d) a fourth transmission line TL**4** that forms a first or left portion, branch, leg, or side of the second T-shaped transmission line matching structure **220**, which provides an input of the second T-shaped transmission line matching structure **220** that is coupled to the second transmission line TL**2** of the first T-shaped transmission line matching structure **210**; (e) a fifth transmission line TL**5** coupled in series to the fourth transmission line TL**4** in a manner that forms a second or right portion, branch, leg, or side of the second T-shaped transmission line matching structure **220**, and which provides an output of the MBMC matching network **200**; and (f) a sixth transmission line TL**6** coupled in parallel between the fourth transmission line TL**4** and the fifth transmission line TL**5**, which forms a third portion, branch, leg, or side of the second T-shaped transmission line matching structure. Additionally, a third portion, branch, leg, or side of at least one T-shaped transmission line matching structure is coupled to electrical ground. For instance, the sixth transmission line TL**6**, that is, the third portion of the second T-shaped transmission line matching structure, can be coupled to electrical ground.

[0063] FIG. **5**A further indicates the width and length of each of the first through sixth transmission lines TL**1**-TL**6** for a representative implementation of the MBMC matching network **200** configured to provide impedance matching with respect to the GSM-1800 frequency band and the UTMS-2100 frequency band. FIG. **5**B is an image of an as-fabricated implementation the MBMC rectifier **102**, which includes this representative implementation of the MBMC matching network **200** having first through sixth transmission lines TL**1**-TL**6** fabricated with the dimensions indicated in FIG. **5**A; a conventional Schottky diode (e.g., an Avago HSMS-2852, C_j=0.18 pF, R_s=2512, Avago Technologies, Singapore) coupled to the output of the fifth transmission line TL**5**; and a DC pass filter **130** coupled to the Schottky diode. The DC pass filter **130** is implemented as a conventional stepped-impedance microstrip low pass filter. Each of the aforementioned elements was fabricated on a conventional circuit board, such as an RO4003 32 mil high frequency laminate circuit board (Rogers Corporation, Rogers, Conn. USA). Additionally, for this representative implementation, the resistive load **140** is a 6.4 kΩresistor.

[0064] To verify the frequency-power characteristics or performance of a rectenna **100** implemented as set forth above, single tone RF-to-DC power conversion efficiency (PCE) was tested using a signal generator (e.g., an Agilent E8257D, Agilent Technologies, Palo Alto, Calif. USA). FIG. **6** plots measured RF-to-DC PCE against frequency at different input RF power levels (−30 dBm to −18 dBM) for the representative implementation of the MBMC rectenna **100**. As indicated in FIG. **6**, a higher PCE is achieved at approximately 1.80-1.88 GHz and approximately 2.11-2.17 GHz than for other frequency bands, for all power levels, in accordance with intended objectives for the representative implementation of the MBMC rectenna **100**. Additionally, it will be readily understood by one having ordinary skill in the relevant art that when multiple or several channels in either of the 1.80-1.88 GHz band or the 2.11-2.17 GHZ band are activated, the RF power in such channels can be simultaneously converted to DC power at an enhanced or high PCE.

[0065] To further verify the PCE efficiency enhancement of this representative implementation of the MBMC rectenna **100**, two-tone RF-to-DC PCE performance was also examined using two Agilent E8257D signal generators and a ZFRSC-42 power splitter (Mini-Circuits, Brooklyn, N.Y. USA). FIG. **7** plots measured PCE against single and dual-tone input power for input signal frequencies of 1.83 GHz and 2.14 GHz, which are the approximate center frequencies of the 1.80-1.88 GHz and 2.11-2.17 GHz frequency bands. For purpose of simplicity, the input power for each of these tone frequencies was identical, across a predetermined range of input powers.

[0066] As indicated in FIG. **7**, RF-to-DC PCE is greater for dual tone input compared to single tone input across the tested range of input powers. Thus, the addition of RF input signals at different frequencies facilitates the generation of higher converted DC power and a higher PCE compared to single-tone RF input situations. FIG. **7** clearly indicates that the MBMC rectenna **100** can simultaneously harvest RF power that is dispersed among or across separate frequency bands and different channels within such frequency bands with an enhanced or considerably higher RF-to-DC PCE. Furthermore, when input power is approximately −30 dBm, which is expected to correspond to a relatively low ambient RF power level, the RF-to-DC PCE is only 5.29% and 4.79% for the individual 1.83 GHz and 2.14 GHz tones, respectively. However, the input of the corresponding two tone signal by way of simultaneous input of each of the 1.83 GHz and 2.14 GHz tones results in an RF-to-DC PCE of 11.65%, which surprisingly is more than twice the RF-to-DC PCE for either single tone input situation.

[0067] An additional representative implementation of the MBMC rectenna **100** was tested with respect to its ambient RF energy harvesting performance. In this additional representative implementation of the MBMC rectenna **100**, the antenna **110** was a wideband antenna **100**, rather than a dual band antenna. More specifically, in this additional representative implementation of the MBMC rectenna **100**, the MBMC matching network **200** was connected to the same wideband horn antenna utilized for generating the measurements of Table 1. The MBMC rectenna **100** was also positioned in the same modern urban environment as that used for generating the measurements of Table 1.

[0068] As indicated in Table 2, for this additional representative implementation of the MCMC rectenna **100**, received RF power was approximately −20 dBm, which is approximately 10 times greater than the −30 dBm relatively low ambient RF power level considered above for the dual tone situation. The output voltage of the rectifier **120** was approximately 120 mV. Therefore, the RF-to-DC PCE of this additional representative implementation of the MBMC rectenna **100** was approximately 22.5%, which greatly exceeds the RF-to-DC PCE measured for the above single tone situation, and which significantly exceeds the 11.65% RF-to-DC PCE measured for the above dual tone situation.

TABLE 2

Comparison of RF-to-dc PCE for single-tone, dual-tone, and multi-channel RF energy harvesting

Frequency Band	Received Power (dBm)	RF-to-DC PCE (%)
Single tone (1.83 GHz)	−30	5.29
Dual tone (1.83 & 2.14)	−30	11.65

TABLE 2-continued

Comparison of RF-to-dc PCE for single-tone, dual-tone, and multi-channel RF energy harvesting

Frequency Band	Received Power (dBm)	RF-to-DC PCE (%)
Urban area (multi-band multi-channel)	−20	22.5
Nearby transmission tower	−15	38.8

[0069] Finally, the additional representative implementation of the MBMC rectenna **100** was positioned near a commercial mobile telephony transmission tower, where the ambient RF energy would clearly be higher than in a modern urban area that is away or distant from the mobile transmission tower. When positioned near the mobile transmission tower, the measured RF received power varied between approximately −15 dBm and −20 dBm, in accordance with RF energy variations in multiple channels of the GSM-1800 and UMTS-2100 bands. Additionally, in this high ambient RF energy environment, the maximum DC voltage output by the rectifier **120** was approximately 280 mV. The RF-to-DC PCE for this additional implementation of the MBMC rectenna **100** in the high ambient RF energy environment can be derived according to the maximum input power (−15 dBm, −0.032 mW) to be approximately 38.8%.

[0070] Aspects of Representative Rectenna Rectifier Architectures

[0071] In addition or as an alternative to the foregoing, various embodiments in accordance with the present disclosure can enhance rectenna RF-to-DC PCE by way of a rectifier that can operate over an extended input power range. Depending upon embodiment details, an extended input power range rectifier can be, an adaptively reconfigurable rectifier, or a breakdown-protected rectifier, as described in detail hereafter.

[0072] Adaptively Reconfigurable Rectifier

[0073] FIG. **8**A is a schematic illustration of an adaptively reconfigurable rectenna or adaptive rectenna **300** based upon a first adaptively reconfigurable rectifier or first adaptive rectifier **302** in accordance with an embodiment of the present disclosure. Compared to the conventional high RF input power rectifier circuit shown in FIG. **2**B, the first adaptively reconfigurable rectifier **302** includes a transistor **310** coupled in parallel with the set of diodes D**2**. More particularly, in several embodiments the transistor **310** is a MOSFET having its gate coupled to ground; its drain and source coupled in parallel with the set of diodes D**2** (e.g., its drain can be coupled to the output of D**1**, and its source can be coupled between C**1** and L**1**.

[0074] Because the gate of the MOSFET **310** is maintained at ground potential, the MOSFET's gate-to-source voltage (V_{gs}) is equal to the inverse of the DC voltage output across load resistor R**1**. Consequently, the MOSFET **310** can serve as an adaptive switch that automatically selectively shunts the set of diodes D**2** based upon the DC voltage output across load resistor R**1**. More particularly, in a representative implementation, the MOSFET **310** is an n-channel depletion mode MOSFET, such as an NE3210S01 (NEC Corporation, Tokyo, Japan). When the input RF power is low, V_{gs} is near 0 Volts, and the MOSFET **310** is in the ON state, thereby shunting the set of diodes D**2**. Under such operating conditions, the first adaptively reconfigurable rectifier **302** behaves like the conventional low input power rectifier shown in FIG. **2**A. As input RF power increases, the DC voltage output across load resistor R_L increases, V_{gs} decreases, the MOSFET's channel resistance R_{ds} rises, and resistive loss within the MOSFET **310** increases. This resistive loss results in a slight adverse effect on RF-to-DC PCE as input RF power increases from a low to a moderate level (in contrast to the conventional low RF input power rectifier shown in FIG. **2**A).

[0075] As input RF power increases further, V_{gs} eventually reaches the cutoff voltage of the MOSFET **310**, and the MOSFET **310** is thus transitioned to its OFF state. Therefore, under higher/high input RF power conditions, the first adaptively reconfigurable rectifier **302** behaves like the conventional high input RF power rectifier shown in FIG. **2**B. Thus, the MOSFET **310** automatically reconfigures the first adaptively reconfigurable rectifier **302** to a more efficient rectifier topology (i.e., a high RF input power rectifier topology) under high input RF power conditions. Analogously, the MOSFET **310** automatically reconfigures the first adaptively reconfigurable rectifier **302** to a more efficient rectifier topology (i.e., a low input RF power rectifier topology) under low input RF power conditions. The first adaptively reconfigurable rectenna **300** can thus exhibit a greatly or dramatically increased input RF power operating range compared to either of the conventional rectennas of FIGS. **2**A and **2**B.

[0076] Notwithstanding the foregoing, within or across a given or predetermined input RF power range, V_{gs} is not sufficiently negative to completely turn off the MOSFET **310**, thereby resulting in an amount of resistive loss that can be non-negligible. This resistive loss problem can be solved by another adaptive rectifier topology, as described in detail hereafter.

[0077] FIG. **8**B is a schematic illustration of an adaptively reconfigurable rectenna or adaptive rectenna **350** based upon a second adaptively reconfigurable rectifier or second adaptive rectifier **352** in accordance with another embodiment of the present disclosure. In an embodiment, the second adaptively reconfigurable rectifier **352** includes two transistors **310**, **312** coupled in parallel across the set of diodes D**2**, rather than a single transistor **310** coupled in parallel across the set of diodes D**2**, in order to overcome the aforementioned resistive loss problem. In several embodiments, the two transistors **310**, **312** include a first and a second MOSFET **310**, **312** coupled or connected in series, where the gate of each MOSFET **310**, **312** is coupled to electrical ground. The source of the first MOSFET **310** is coupled between C**1** and L**1**, and the drain of the first MOSFET **310** is coupled to the source of the second MOSFET **312**. The source of the second MOSFET **312** is coupled to diode D**1**.

[0078] FIG. **8**C illustrates a circuit configuration of a representative implementation of the second adaptively reconfigurable rectenna **350** of FIG. **8**B, configured for operation at 100 MHz for purpose of example. In this representative implementation, diode D**1** is an HSMS-2852 diode (Avago Technologies, Singapore); the set of diodes D**2** is a set of four HSMS-2850 diodes, coupled pairwise between transmission line TL**7** in the manner shown; and the first and second MOSFETS **310**, **312** are NE2110S01 MOSFETS. (NEC Corporation, Tokyo, Japan). Additionally, C**1** is 5.1 pF; L**1** is 290 nH; and C**2** is 100 pF. Transmission lines TL**1** through TL**13** have widths (W) and lengths. (L) as follows, in millimeters: W**1**=2.4; L**1**=15; W**2**=W**3**=3.5; L**2**=L**3**=7.5; W**4**=2; L**4**=13; W**5**=W**6**=W**8**=W**9**=0.7; L**5**=L**6**=L**8**=L**9**=3; W**7**=0.7; L**7**=1.8; W**10**=2; L**10**=1; W**11**=3; L**11**=7; W**12**=5; L**12**=8; W**13**=4; and L**13**=5. Additionally, R_L=2.8 kΩ. FIG. **8**D is an image of the corresponding representative second adaptively reconfigurable rectifier **352**, as fabricated on a 31 mil Rogers RT/duroid 5880 substrate having a dielectric constant of 2.2 and a loss tangent of 0.0012. One having ordinary skill in the relevant art will understand that depending upon embodiment and/or implementation details, diode D**1** can be a set of diodes (i.e., one or more diodes); and/or the set of diodes D**2** can be a single diode rather than multiple diodes.

[0079] For comparison purposes, simulations of the RF-to-DC PCE versus input RF power for the conventional low RF input power rectifier shown in FIG. **2**A, the conventional high RF input power rectifier shown in FIG. **2**B, the first adaptively reconfigurable rectifier **302** shown in FIG. **8**A, and the second adaptively reconfigurable rectifier **352** of FIG. **8**B were performed using diode D**1**, the set of diodes D**2**, capacitors C**1** and C**2**, inductance L**1**, and load resistance R_L defined as given above. Each such simulation corresponded to an operating frequency of 100 MHz, for purpose of example.

[0080] FIG. **9**A is a graph showing RF-to-DC PCE simulation results for (a) the conventional low RF input power rectifier shown in FIG. **2**A; (b) the conventional high RF input power rectifier shown in FIG. **2**B; (c) the first adaptively reconfigurable rectifier **302** shown in FIG. **8**A; and (d) the second adaptively reconfigurable rectifier **352** shown in FIG. **8**B. From FIG. **9**A, one can readily note the following:

> **[0081]** (a) With respect to the Conventional low RF input power rectifier shown in FIG. **2**A, it is clear that its RF-to-DC PCE increases with increasing input RF power until reaching a maximum of approximately 85% of input RF power at approximately 0.3 dBm. When the input RF power exceeds this level, the RF-to-DC PCE of the conventional rectifier shown in FIG. **2**A decreases rapidly, since the voltage swing across D**1** exceeds D**1**'s breakdown voltage. Thus, the conventional low RF input power rectifier shown in FIG. **2**A is suitable only for low input RF power situations, and is unsuitable for high input RF power situations.

> **[0082]** (b) With respect to the conventional high RF input power rectifier shown in FIG. **2**B, the RF-to-DC PCE at low RF input powers is less than that for the conventional low RF input power rectifier shown in FIG. **2**A. For instance, when the input RF power is less than about −5 dBm, the RF-to-DC PCE of the conventional high RF input power rectifier remains below approximately 30%. However, the RF-to-DC PCE at high RF input powers is much better than that for the conventional low RF input power rectifier shown in FIG. **2**A, reaching a peak of approximately 82% at about 19 dBm input RF power. Thus, the conventional rectifier shown FIG. **2**B is suitable only for high input RF power situations.

[0083] (c) With respect to the first adaptively reconfigurable rectifier **302** shown in FIG. **8**A, its RF-to-DC PCE generally matches that for the conventional rectifier shown in FIG. **2**A for input RF power less than −15 dBm. As input RF power increases from approximately −15 dBm to approximately 0 dBm, the RF-to-DC PCE of the first adaptively reconfigurable rectifier **302** is slightly reduced compared to that of the conventional low RF input power rectifier because of the aforementioned resistive loss due to the MOSFET **310**. When the input RF power exceeds approximately 10 dBm, V_{gs} reaches the cutoff voltage of the MOSFET **310**, thereby switching the MOSFET **310** to the OFF state. Under such high RF input power conditions, the first adaptively reconfigurable rectifier **302** operates in a manner that is generally similar or analogous to that of the conventional high RF input power rectifier of FIG. **2**B.

> **[0084]** (d) With respect to the second adaptively reconfigurable rectifier **352**, FIG. **9** indicates that the second adaptively reconfigurable rectifier **352** achieves a better OFF state than the first adaptively reconfigurable rectifier **302** across a range of input RF power between approximately 3 dBm and 20 dBm. Thus, the second adaptively reconfigurable rectifier **352** provides significantly enhanced RF-to-DC PCE across a wide input RF power range, for instance, from approximately 3 dBm to 20 dBm, compared to the other rectifiers considered in this simulation.

[0085] In addition to the foregoing, the RF-to-DC PCE of the representative implementation of the second adaptively reconfigurable rectifier **352** shown in FIGS. **8**C-**8**D was measured versus input RF power. More particularly, input RF power and DC voltage output across the resistive load R_L were simultaneously measured, and the RF-to-DC PCE was calculated in accordance with the following equation:

$$\eta=(V_L)^2/R_L\times(1/P_{in})\times100 \quad (1)$$

in a manner readily understood by those having ordinary skill in the relevant art.

[0086] FIG. **9**B is a graph showing measured versus simulated RF-to-DC PCE, as well as measured versus simulated output voltage, as a function of input RF power for the aforementioned representative implementation of the second adaptively reconfigurable rectifier **352**. The measured and simulated results exhibit good agreement with each other. FIG. **9**B indicates that an RF-to-DC PCE exceeding 50% can be achieved across an extended or enhanced input RF power range, from approximately −14 dBm to 21 dBm. Additionally, an RF-to-DC PCE greater than equal to approximately 75% can be reached for input RF powers between approximately 7 dBm and 14 dBm. Also, a highest DC output voltage of about 14.9V can be reached when input RF power approximately equals 23 dBm. The measured results shown in FIG. **9**B clearly indicate that the second adaptively reconfigurable rectifier **352** is suitable for WPT and/or RF energy harvesting applications in which a wide operating input RF power range is desired or required.

[0087] Further to the foregoing, an adaptively reconfigurable rectifier can include more than two transistors (e.g., more than two MOSFETs) coupled in parallel with the set of diodes D**2**. For instance, an adaptively reconfigurable rectifier can include three, four, or more MOSFETs coupled in parallel with the set of diodes D**2**, where such MOSFETs are stacked in series. FIG. **9**C is a graph showing simulated RF-to-DC PCE versus input RF power corresponding to FIG. **9**A, but which also includes simulation results for an adaptively reconfigurable rectifier having three MOSFETs coupled across/in parallel with the set of diodes D**2**, as well as an adaptively reconfigurable rectifier having four MOSFETs coupled across/in parallel with the set of diodes D**2**. More particularly, FIG. **9**C illustrates simulated RF-to-DC PCE versus input RF power for (a) the conventional low RF input power rectifier of FIG. **2**A (indicated as "Eff(a)" in FIG. **9**C); (b) the conventional high RF input power rectifier of FIG. **2**B (indicated as "Eff(b)" in FIG. **9**C); (c) the first adaptively reconfigurable rectifier **302** of FIG. **8**A (indicated as "Eff(c)" in FIG. **9**C); and (d) three versions of the second adaptively reconfigurable rectifier **352** of FIG. **8**B, namely, (i) a version with N=2 MOSFETs **310**, **312**; a version with N=3 MOSFETs; and (iii) a version with N=4 MOSFETs (respectively indicated as "Eff(d) N=2", "Eff(d) N=3", and "Eff(d) N=4" in FIG. **9**C), where such MOSFETs in each version are serially stacked and coupled in parallel with diode D**1**. FIG. **9**C indicates that as the number of serially stacked MOSFETs increases, the RF-to-DC PCE tends to be higher in the high RF input power region, and lower in the low RF input power region.

[0088] Still further to the foregoing, FIG. **9**D is a graph of simulated RF-to-DC PCE versus RF input power for different values of the load resistance R_L for the second adaptively reconfigurable rectifier **352**. FIG. **9**D indicates that as load resistance R_L increases, the RF-to-DC PCE tends to be higher in the low RF input power region, and lower in the high RF input power region.

[0089] Thus, there can be certain tradeoffs in terms of RF-to-DC PCE performance with respect to the number of MOSFETs stacked in series, and/or load resistance R_L. Furthermore, the threshold voltage of the MOSFET(s) should be selected such that under high RF input power conditions, the instantaneous V_{gs} will turn the FET(s) off when the reverse bias across diode D**1** approaches D**1**'s breakdown voltage, as will be readily understood by one having ordinary skill in the relevant art.

[0090] Breakdown-Protected Rectifier

[0091] FIG. **10**A is a schematic illustration of a breakdown-protected rectenna **400** based upon a breakdown-protected rectifier **402** in accordance with an embodiment of the present disclosure. In various embodiments, the breakdown-protected rectifier **402** includes a transistor-protected diode structure **410** that includes at least one diode coupled to a transistor in a manner that protects the diode from direct exposure to negative voltages that would ordinarily cause the diode to break down in the absence of the transistor. For an embodiment in which the transistor-protected diode structure **410** includes a single diode D**1** and a single transistor, a first or control terminal of a transistor is coupled to the anode of diode D**1**; a second transistor terminal is coupled in series to the output or cathode of diode D**1**; and a third transistor terminal is coupled to a load resistance R_L. As a result of such transistor coupling, a negative voltage having a magnitude equal to or greater than the reverse breakdown voltage of diode D**1**, to which diode D**1** would normally be directly exposed in the absence of the transistor, is instead expressed across the combination of the transistor and diode D**1**. The transistor thus protects diode D**1** from direct exposure to a negative voltage having a magnitude that exceeds diode D**1**'s breakdown voltage, until the transistor itself is exposed to a negative voltage between its control terminal and its third terminal that causes junction breakdown within transistor itself.

[0092] FIG. **11**A is a schematic illustration of a first transistor-protected diode structure **410***a* in accordance with an embodiment of the present disclosure. In several embodiments, the first transistor-protected diode structure **410***a* includes a MOSFET **420** having a gate coupled to the input or anode of diode D**1**; a source coupled to the output or cathode of diode D**1**; and a drain couplable to a load resistance R_L. In a representative implementation, diode D**1** can be a Schottky diode, such as an Avago HSMS-2852, having a forward threshold voltage of approximately 150 mV and a reverse breakdown voltage of approximately −3.8 V; and the MOSFET **420** can be an n-channel depletion mode MOSFET, such as a BF998 MOSFET (NXP Semiconductors, Eindhoven, NL), which has a threshold voltage V_{th} of approximately −0.8V. When the combination of diode D**1** and the MOSFET **420** is forward biased ($V_a>V_b$), the MOSFET is in the ON state because V_{gs} is positive, and the combination behaves like a single diode having approximately the same built-in voltage as diode D**1** by itself. One having ordinary skill in the relevant art will understand that there will be a small amount of resistive loss in the MOSFET **420** under such operating conditions.

[0093] When the combination of diode D**1** and the MOSFET **420** is reverse biased ($V_a<V_b$) and V_{ab}, is negative but larger than the breakdown voltage of diode D**1**, V_{gs} is smaller than the MOSFET's threshold voltage. Therefore, MOSFET **420** transitions to an OFF state before diode D**1** breaks down. Consequently, even if the overall bias voltage V_{ab} across the combination of diode D**1** and the MOSFET **420** is beyond the breakdown voltage of diode D**1**, the actual voltage across diode D**1** remains smaller that its breakdown voltage because the major portion of the voltage is sustained between the drain and the source of the MOSFET **420**. When the overall bias voltage V_{ab} is beyond the maximum gate-to-drain voltage V_{gd} that the MOSFET **420** can withstand, the MOSFET **420** will break down.

[0094] FIG. **12** is a graph illustrating the I-V characteristics of each of diode D**1** and the first transistor-protected diode structure **410***a*, in accordance with the aforementioned representative implementations thereof. As clearly indicated in FIG. **12**, the first transistor-protected diode structure **410***a* exhibits a breakdown voltage having a magnitude that is dramatically greater than that of diode D**1** alone. As a result, incorporation of a transistor-protected diode structure **420** into a rectifier can be expected to significantly or dramatically enhance the rectifier's RF-to-DC PCE.

[0095] FIG. **13** is a graph illustrating simulated RF-to-DC PCE versus input RF power level for the rectifier shown in FIG. **2**C as well as the breakdown-protected rectifier **402** of FIG. **10**A at a representative operating frequency of 50 MHz. Each such circuit utilizes C**1**=8 pF, L**1**=750 nH, C**2**=10 pF, and R_L=6.0 kΩ. Diode D**1** and the MOSFET **420** are defined in accordance with the representative implementations thereof indicated above. The conventional rectifier of FIG. **2**C provides an RF-to-DC PCE above 50% for input RF power levels between approximately −20.5 to −0.5 dBm. For higher input RF power levels, the RF-to-DC PCE of the conventional rectifier of FIG. **2**C degrades sharply because the voltage swings across diode D**1** exceed its breakdown voltage. Thus, the conventional rectifier of FIG. **2**C is suitable only for low RF input power conditions.

[0096] As shown in FIG. **13**, the RF-to-DC PCE of the breakdown-protected rectifier **402** of FIG. **10**A is slightly less than that of the conventional rectifier of FIG. **2**C for input RF power between approximately −25 dBm to −2.5 dBm, which can be attributed to the small resistive loss associated with the MOSFET **420**. However, the RF-to-DC PCE of the breakdown-protected rectifier **402** remains above 50% for input RF power levels from approximately −18.5 to 17.5 dBm, clearly indicating that the operating input RF power range of the breakdown-protected rectifier **402** is significantly extended compared to that of the conventional rectifier of FIG. **2**A.

[0097] FIG. **10**B is a schematic illustration of a representative implementation of the breakdown-protected rectifier **402** of FIG. **10**A. In this representative implementations, the widths and lengths of transmission lines TL**1** through TL**13** in millimeters are as follows: W**1**=2.4; L**1**=5.0; W**2**=W**3** W**4**=W**5**=W**6**=W**7**=1.8; L**2**=2.9; L**3**=4.8; L**4**=4.0; L**5**=−9.25; L**6**=1.2; L**7**=25.6; W**8**=3; L**8**=2; W**9**=1; L**9**=1.8; W**10**=1.0; L**10**=4.2; W**11**=W**13**=1.8; W**12**=3; L**11**=3.9; L**12**=4.9; and L**13**=4.0. As above, C**1**=8 pF; L**1**=750 nH; C**2**=100 pF, and RL=6.0 kΩ. This representative implementation of the breakdown-protected rectifier **400** was fabricated on a 31 mil thick Rogers RT/duroid 5880 substrate, which has a dielectric constant of 2.2 and a loss tangent of 0.0012. FIG. **10**C is an image of the as-fabricated representative implementation of the breakdown-protected rectifier **402**.

[0098] FIG. **14** is a graph showing measured versus simulated RF-to-DC PCE, as well as measured versus simulated output voltage, as a function of input RF power for the aforementioned representative implementation of the breakdown-protected rectifier **402** of FIGS. **10**B and **10**C. Measured and simulated results exhibit good agreement with each other. The measured RF-to-DC PCE was greater than 50% across an input RF power range from approximately −13.5 dBm to 16.7 dBm. Additionally, the highest DC output voltage reached approximately 11.9V for input RF power of about 17.6 dBm. Thus, the breakdown-protected rectifier **402** can enhance RF-to-DC PCE in WPT and/or ambient RF energy harvesting applications in which a wide operating RF input power range is desired or required.

[0099] In addition to the first transistor-protected diode structure **410***a*, a variety of other transistor-protected diode structures **410** exist. For instance, FIGS. **11**B-**11**H are schematic illustrations of a second through a seventh transistor-protected diode structure **410***b-h*, which can be incorporated into a rectifier to form a breakdown-protected rectifier **400** in accordance with an embodiment of the present disclosure. As indicated in FIG. **11**B, a second diode D**2** can be coupled such that it is coupled in parallel with the source and drain of the MOSFET **420**. More particularly, the anode of diode D**2** can be coupled to the cathode of diode D**1**, and hence to the source of the MOSFET **420**; and the cathode of diode D**2** is coupled to the drain of the MOSFET **420**, and hence to a load resistor R_L. The presence of diode D**2** can facilitate the reduction of minimization of forward resistive losses within the MOSFET **420** under forward biasing conditions, in a manner readily understood by one having ordinary skill in the relevant art.

[0100] As indicated in FIGS. **11**C and **11**D, as well as FIGS. **11**E-**11**H, circuit elements can be coupled to the gate of the MOSFET **420** to distribute reverse voltages across the gate-to-drain junction of the MOSFET **420** in addition to one or more types of circuit elements under consideration, thereby further increasing the overall magnitude of the breakdown voltage for the transistor-protected diode structure **410** under consideration relative to the breakdown voltage of the first transistor-protected diode structure **410***a* of FIG. **11**A. FIGS. **11**E-**11**H illustrate manners in which particular individual transistor-protected diode structures **410** can be cascaded or iteratively coupled to achieve breakdown voltage enhancement compared to the first transistor-protected diode structure **410***a*.

[0101] Further Representative Rectenna Designs

[0102] Rectennas in accordance with embodiments of the present disclosure can include an MBMC matching network **200**, either alone or in combination with either of an adaptively reconfigurable rectifier **300**, **350** or a transistor-protected rectifier **400**. For instance, FIG. **15**A illustrates a rectenna **500** having each of an MBMC matching network **200** and a second adaptively reconfigurable rectifier **352** in accordance with an embodiment of the present disclosure. FIG. **15**B illustrates a rectenna **550** having each of an MBMC matching network **200** and a breakdown-protected rectifier **402** in accordance with an embodiment of the present disclosure.

[0103] Further Performance Comparisons

[0104] FIG. **16** is a graph illustrating simulated RF-to-DC PCE versus input RF power for representative implementations of each of (a) the conventional low input RF power rectifier of FIG. **2**A; (b) the MBMC rectifier **102** of FIG. **3**; (c) the second adaptively reconfigurable rectifier **352** shown in FIG. **8**B; (d) the MBMC matching network **200** coupled to the second adaptively reconfigurable rectifier **352**, in a manner shown in FIG. **15**A; (e) the breakdown-protected rectifier **402** of FIG. **10**A; and (f) the MBMC matching network **200** coupled to the breakdown-protected rectifier **402**, in a manner shown in FIG. **15**B.

[0105] As indicated in FIG. **16**, each of the rectifiers corresponding to simulation results (b) through (f) exhibit enhanced RF-to-DC PCE compared to the conventional low input RF power rectifier shown in FIG. **2**A. Furthermore, each of the second adaptively reconfigurable rectifier **352**, the breakdown-protected rectifier **402**, and the combination of the MBMC matching network **200** coupled with either of the second adaptively reconfigurable rectifier **352** or the breakdown-protected rectifier **402** exhibit a significantly or dramatically enhanced or extended input RF power operating range compared to the conventional low input RF power rectifier shown in FIG. **2**A.

[0106] The main function of a MBMC network **200** within an MBMC rectifier **102** in accordance with an embodiment of the present disclosure is to collect more ambient RF power in order to increase the input RF power level so as to increase the overall RF-to-DC PCE of a WPT or an ambient RF energy harvesting system, apparatus, device, or circuit. On the other hand, an adaptively reconfigurable rectifier **302**, **352** or a breakdown-protected rectifier **402** in accordance with an embodiment of the present disclosure is mainly used to enhance the breakdown voltage of a rectifier's semiconductor diode component(s), to thereby significantly enhance the input RF power operating range of the rectifier.

[0107] The combination of an MBMC matching network **200** in accordance with an embodiment of the present disclosure with an embodiment of either of an adaptively reconfigurable rectifier **302**, **352** (e.g., as shown in FIG. **15**A) or a breakdown-protected rectifier **402** (e.g., as shown in FIG. **15**B) provides an enhanced or higher RF-to-DC PCE and much higher breakdown power performance when compared with a conventional low RF input power rectifier or a rectifier that includes the MBMC matching network **200**, but which excludes adaptively reconfigurable circuit elements or transistor-protected diodes, respectively, in accordance with embodiments described herein.

[0108] An adaptively reconfigurable rectifier **302**,**352** or a breakdown protected rectifier **402** that is coupled to or which includes an MBMC matching network **200** can collect more RF power than a rectifier configured to receive RF power within a single RF frequency band, and can have a higher RF-to-DC PCE compared with an adaptively reconfigurable rectifier **302**,**352** or a breakdown protected rectifier **402** that is not coupled to or which excludes an MBMC matching network **200**. However, an adaptively reconfigurable rectifier **302**, **352** or a breakdown-protected rectifier **402** that is coupled to or which an MBMC matching network **200** will reach its breakdown power level at a lower overall input RF power level than an adaptively reconfigurable rectifier **302**, **352** or a breakdown-protected rectifier **402** that is not coupled to or which excludes an MBMC matching network **200**, due to greater RF energy collection or harvesting by the MBMC matching network **200** under the same RF environmental conditions.

[0109] Wireless Network Considerations

[0110] A rectenna or rectifier in accordance with an embodiment of the present disclosure (e.g., essentially any rectenna that includes an MBMC matching network **200** and/or one of an adaptively reconfigurable rectifier **350**, **352** or a breakdown-protected rectifier **400**) can be utilized in or form a portion of a wireless network configured for transmitting/receiving (TX/RX) electromagnetic signals. The use of a particular rectenna or rectifier structure or topology in accordance with an embodiment of the present disclosure within a given wireless network, and the implementation thereof, can be based upon a wireless network architecture or application, input RF power operating range, a set of RF frequency bands under, and/or particular RF-to-DC PCE objectives/requirements under consideration.

[0111] FIGS. **17**A-**17**H are block diagrams illustrating representative types of wireless networks **600** in which a set of rectennas (e.g., including one or more MBMC rectennas **100**, adaptively reconfigurable rectennas **300**, **350**, and/or breakdown-protected rectennas **400**) in accordance with one or more embodiments of the present disclosure can be deployed. More particularly, FIG. **17**A illustrates a one-to-one TX/RX wireless network **600***a* in which a single TX stage **610** wirelessly transfers signals to a single RX stage **630**. The RX stage **630** is coupled to a set of devices **650** such as a set of sensors. In a WPT application, the set of devices **650** can receive power output by the RX stage **630**, which converts RF signals received from the TX stage **610** into DC signals in a manner identical, essentially identical, or analogous to that described herein (e.g., by way of an MBMC rectenna **100**, an adaptively reconfigurable rectenna **300**, **350**, or a breakdown-protected rectenna **400**), in a manner readily understood by one having ordinary skill in the relevant art.

[0112] FIG. **17**B illustrates a TX/RX wireless network **600***b* including one "repeater"/"amplification" stage **620** disposed between a TX stage **610** and an RX stage **630**, where; the RX stage **630** is coupled to a set of devices **650** (e.g., a set of sensors) as above. The TX stage **610** can wirelessly transfer RF signals to the repeater/amplification stage **620**, which can further wirelessly transfer RF signals to the RX stage **630**. The repeater/amplification stage **620** includes an adjunct RX stage **622** configured for receiving RF signals from the TX stage **610**; and an adjunct TX stage **624** configured for transmitting RF signals to the RX stage **630**. The presence of the repeater/amplification stage **620** can extend the physical range across which the TX stage **610** and the RX stage **630** are separable/separated, and/or enhance the likelihood of reliable wireless signal reception by the RX stage **630**. One or both of the repeater/amplification stage **620** and the RX stage **630** can receive power by way of WPT and/or ambient RF energy harvesting (e.g., by way of an MBMC rectenna **100**, an adaptively reconfigurable rectenna **300**, **350**, a breakdown-protected rectenna **400**, or a combination of an MBMC rectenna **100** with one of an adaptively reconfigurable rectenna **300**, **350** or a breakdown-protected rectenna **400**), depending upon embodiment details. Furthermore, depending upon embodiment details, the RX stage **630** and the adjunct RX stage **622** can have identical, similar, or different structures or topologies.

[0113] In a manner analogous to that shown in FIG. **17**B, FIG. **17**C illustrates a TX/RX wireless network **600***c* including multiple repeater/amplification **620***a,b* stages disposed between a TX stage **610** and an RX stage **630**. Each repeater/amplification stage **620** includes an adjunct RX stage **622** and an adjunct TX stage **624**, in a manner analogous to that described above. The RX stage **630** and/or one or more of the repeater/amplification stages **620***a,b* can receive power by way of WPT and/or ambient RF energy harvesting (e.g., by way of an MBMC rectenna **100**, an adaptively reconfigurable rectenna **300**, **350**, a breakdown-protected rectenna **400**, or a combination of an MBMC rectenna **100** with one of an adaptively reconfigurable rectenna **300**, **350** or a breakdown-protected rectenna **400**), depending upon embodiment details. Here again, the presence of each repeater/amplification stage **620***a,b* can extend the physical range across which the TX stage **610** and the RX stage **630** are separable/separated, and/or enhance the likelihood of reliable wireless signal reception by the RX stage **630**. Such a wireless network **600***c* can be scaled to include N repeater/amplification stages **620**, in a manner understood by one having ordinary skill in the relevant art. Depending upon embodiment details, the RX stage **630** and one or more adjunct RX stages **622** can have identical, similar, or different structures or topologies.

[0114] FIGS. **17**D-**17**H illustrate further wireless network architectures or configurations **600***d-h* in which one or more RX stages **630** and/or one or more repeater/amplification stages **620** can employ an MBMC rectenna **100**, an adaptively reconfigurable rectenna **300**, **350**, or a breakdown-protected rectenna **400**, in a manner analogous to that described above. In general, one or more rectenna embodiments in accordance with the present disclosure can be utilized in or form portions of an m-to-p TX/RX network, where m indicates the number of TX stages **610** and p indicates the number of RX stages **630** present in the network. Each of m and p can be one or more (e.g., m can equal 1, indicating a single TX stage **610**; and p can equal 1, several, dozens, hundreds, thousands, or more, indicating the presence of one to possibly hundreds, thousands, or more RX stages **630**). Similarly, one or more rectenna embodiments in accordance with the present disclosure can be utilized in or form portions of an m-to-n-to-p TX/repeater-amplification/RX network, where m indicates the number of TX stages **610**; n indicates the number of repeater/amplification stages **620**; and p indicates the number of RX stages **630** in the network. Each of m, n, and p can be one or more, in a manner identical, essentially identical, or analogous to that previously indicated.

[0115] Thus, in view of the foregoing, one or more rectennas in accordance with various embodiments of the present disclosure can be utilized in or form portions of essentially any type of wireless network organized in accordance with essentially any type of wireless network topology, wherein electromagnetic signals are received and rectified. Furthermore, one or more RX stages **630** and/or one or more repeater/amplification stages **620** within such wireless networks can receive power by way of WPT and/or ambient RF energy harvesting, such as by way of an MBMC rectenna **100**, an adaptively reconfigurable rectenna **300**, **350**, or a breakdown-protected rectenna **400** in a manner identical, analogous, or generally analogous to that described above, depending upon embodiment details.

[0116] Aspects of particular embodiments of the present disclosure address at least one aspect, problem, limitation, and/or disadvantage associated with existing rectenna and/or rectifier circuit elements, circuits, or techniques. While features, aspects, and/or advantages associated with certain embodiments have been described in the disclosure, other Embodiments may also exhibit such features, aspects, and/or advantages, and not all embodiments need necessarily exhibit such features, aspects, and/or advantages to fall within the scope of the disclosure. It will be appreciated by a person of ordinary skill in the art that several of the above-disclosed systems, components, processes, or alternatives thereof, may be desirably combined into other different systems, components, processes, and/or applications. In addition, various modifications, alterations, and/or improvements may be made to various embodiments that are disclosed by a person of ordinary skill in the art within the scope and spirit of the present disclosure.

1. A circuit configured for receiving RF energy distributed in multiple RF frequency bands, the circuit including a multi-band multi-channel (MBMC) matching network, the MBMC matching network comprising:

a plurality of T-shaped transmission line matching structures, each T-shaped transmission line matching structure coupled in series to another T-shaped transmission line matching structure,

wherein each T-shaped transmission line matching structure includes a first portion, a second portion, and a third portion coupled between the first portion and the second portion,

wherein the first portion of a first T-shaped transmission line matching structure within the plurality of T-shaped transmission line structures forms an input of the MBMC matching network, the second portion of a second T-shaped transmission line matching structure within the plurality of T-shaped transmission line matching structures forms an output of the MBMC matching network, and the third portion of at least one T-shaped transmission line matching structure within the plurality of T-shaped transmission line matching structures is couplable to electrical ground.

2. The circuit of claim **1**, wherein each individual T-shaped transmission line matching structure within the MBMC matching network is configured for providing an impedance match corresponding to RF frequencies within a particular RF frequency band.

3. The circuit of claim **1**, wherein the circuit includes a rectenna comprising:

an antenna coupled to the input of the MBMC Matching network; and

a rectifier coupled to the output of the MBMC matching network.

4. The circuit of claim **3**, wherein the circuit forms a portion of a wireless power transmission (WPT) system or a portion of an ambient RF energy harvesting system.

5. The circuit of claim **3**, wherein the rectifier comprises one of:

an adaptively reconfigurable rectifier having a low input RF power rectifying portion, a high input RF power rectifying portion, and a set of transistors coupled to one of the high input RF power rectifying portion the low input RF power rectifying portion, the set of transistors configured for automatically transitioning the adaptively reconfigurable rectifier between a low input RF input power operating configuration and a high input RF power operating configuration based upon input RF power level; and

a breakdown-protected rectifier having at least transistor-protected diode structure, the transistor-protected diode structure comprising at least one diode coupled to a transistor in a manner that protects the diode from direct exposure to negative voltages that would ordinarily cause the diode to break down in the absence of the transistor.

6. A circuit configured for rectifying signals, comprising one of:

an adaptively reconfigurable rectifier having a low input power rectifying portion, a high input power rectifying portion, and a set of transistors coupled to one of the high input power rectifying portion the low input power rectifying portion, the set of transistors configured for selectively and automatically transitioning the adaptively reconfigurable rectifier between a low input power operating configuration and a high input RF power operating configuration in a manner correlated with a power level of a signal input to the circuit; and

a breakdown-protected rectifier having at least one transistor-protected diode structure, the at least one transistor-protected diode structure comprising a first diode coupled to a transistor in a manner that protects the first diode from direct exposure to negative voltages that would ordinarily cause the diode to break down in the absence of the transistor.

7. The circuit of claim **6**, wherein the low input power portion and the high input power portion of the adaptively reconfigurable rectifier are coupled in series.

8. The circuit of claim **7**, wherein the set of transistors of the adaptively reconfigurable is coupled in parallel with the high input power portion.

9. The circuit of claim **6**, wherein a control terminal of each transistor within the set of transistors of the adaptively reconfigurable rectifier is coupled to electrical ground.

10. The circuit of claim **6**, wherein the set of transistors of the adaptively reconfigurable rectifier includes a plurality of transistors coupled to each other in series.

11. The circuit of claim **6**, wherein each transistor within the set of transistors is a Field Effect Transistor (FET).

12. The circuit of claim **6**, wherein the low input power rectifying portion of the adaptively reconfigurable rectifier comprises a first set of diodes, and the high input power rectifying portion comprises a second set of diodes.

13. The circuit of claim **12**, wherein at least one of the first set of diodes and the second set of diodes includes a plurality of diodes.

14. The circuit of claim **6**, wherein the breakdown-protected rectifier comprises a first transistor-protected diode structure including a first diode and a transistor, and wherein the transistor includes a control terminal coupled to the anode of the first diode, a first terminal coupled to the cathode of the, first diode, and a second terminal couplable to circuitry external to the transistor-protected diode structure.

15. The circuit of claim **14**, wherein the anode of the first diode is coupled to an input of the breakdown-protected rectifier and is configured to receive an alternating current (AC) signal, and wherein the second terminal of the transistor is coupled to an output of the breakdown-protected rectifier.

16. The circuit of claim **14**, wherein the transistor-protected diode structure further includes a second diode coupled in parallel with the second terminal and the third terminal of the transistor, such that the anode of the second diode is coupled to the cathode of the first diode.

17. The circuit of claim **14**, wherein the transistor-protected diode structure further includes a second diode having its anode coupled to the anode of the first diode, and its cathode coupled to the control terminal of the transistor.

18. The circuit of claim **14**, wherein the breakdown-protected rectifier comprises the first transistor-protected diode structure combined with a second transistor-protected diode structure coupled in a cascaded manner to the first transistor-protected diode structure, such that a negative voltage across the combination of the first and second transistor-protected diode structures is distributed across the transistor of the first transistor-protected diode structure and a transistor of the second transistor-protected diode structure.

19. The circuit of claim **18**, wherein the second transistor-protected diode structure comprises a first diode and a transistor having a control input, a first terminal, and a second terminal, and wherein the anode of the first diode of the second transistor-protected diode structure is coupled to the anode of the first diode of the first transistor-protected diode structure, the cathode of the first diode of the second transistor-protected diode structure is coupled to the second terminal of the transistor of the second diode-protected transistor structure, and the third terminal of the transistor of the second diode-protected transistor structure is coupled to the control input of the transistor of the first diode-protected transistor structure.

20. A wireless network comprising:

a transmission (TX) stage configured for wirelessly transmitting electromagnetic signals; and

a reception (RX) stage configured for receiving electromagnetic signals transmitted by the TX stage, the RX stage comprising one of:

(a) a multi-band multi-channel (MBMC) matching network comprising: a plurality of T-shaped, transmission line matching structures, each T-shaped transmission line matching structure coupled in series to another T-shaped transmission line matching structure,

wherein each T-shaped transmission line matching structure includes a first portion, a second portion, and a third portion coupled between the first portion and the second portion,

wherein the first portion of a first T-shaped transmission line matching structure within the plurality of T-shaped transmission line structures forms an input of the MBMC matching network, the second portion of a second T-shaped transmission line matching structure within the plurality of T-shaped transmission line matching structures forms an output of the MBMC matching network, and the third portion of at least one T-shaped transmission line matching structure within the plurality of T-shaped transmission line matching structures is couplable to electrical ground;

(b) an adaptively reconfigurable rectifier having a low input power rectifying portion, a high input power rectifying portion, and a set of transistors coupled to one of the high input power rectifying portion the low input power rectifying portion, the set of transistors configured for selectively and automatically transitioning the adaptively reconfigurable rectifier between a low input power operating configuration and a high input RF power operating configuration in a manner correlated with a power level of a signal input to the RX stage; and

(c) a breakdown-protected rectifier having at least one transistor-protected diode structure, the at least one transistor-protected diode structure comprising a first diode coupled to a transistor in a manner that protects the first diode from direct exposure to negative voltages that would ordinarily cause the diode to break down in the absence of the transistor.

21. The wireless network of claim **20**, wherein the RX stage includes the MBMC matching network and one of the adaptively reconfigurable rectifier and the breakdown-protected rectifier.

22. The wireless network of claim **20**, further comprising at least one repeater/amplification stage configured to receive electromagnetic signals from the. TX stage and transmit electromagnetic signals to the RX stage, the at least one repeater/amplification stage comprising one of:

(a) a multi-band multi-channel (MBMC) matching network comprising:

a plurality of T-shaped transmission line matching structures, each T-shaped transmission line matching structure coupled in series to another T-shaped transmission line matching structure,

wherein each T-shaped transmission line matching structure includes a first portion, a second portion, and a third portion coupled between the first portion and the second portion,

wherein the first portion of a first T-shaped transmission line matching structure within the plurality of T-shaped transmission line structures forms an input of the MBMC matching network, the second portion of a second T-shaped transmission line matching structure within the plurality of T-shaped transmission line matching structures forms an output of the MBMC matching network, and the third portion of at least one T-shaped transmission line matching structure within the plurality of T-shaped transmission line matching structures is couplable to electrical ground;

(b) an adaptively reconfigurable rectifier having a low input power rectifying portion, a high input power rectifying portion, and a set of transistors coupled to one of the high input power rectifying portion the low input power rectifying portion, the set of transistors configured for selectively and automatically transitioning the adaptively reconfigurable rectifier between a low input power operating configuration and a high input RF power operating configuration in a manner correlated with a power level of a signal input to the RX stage; and

(c) a breakdown-protected rectifier having at least one transistor-protected diode structure, the at least one transistor-protected diode structure comprising a first diode coupled to a transistor in a manner that protects the first diode from direct exposure to negative voltages that would ordinarily cause the diode to break down in the absence of the transistor.

23. The wireless network of claim **22**, wherein the at least one repeater/amplification stage includes an MBMC matching network and one of an adaptively reconfigurable rectifier and a breakdown-protected rectifier.

* * * * *

US 20160020631A1

(19) **United States**

(12) **Patent Application Publication**

Manova-Elssibony

(10) **Pub. No.: US 2016/0020631 A1**

(43) **Pub. Date: Jan. 21, 2016**

(54) **ENERGY HARVESTING SYSTEM**

(71) Applicant: **HUMAVOX LTD.**, Kfar Saba (IL)

(72) Inventor: **Asaf Manova-Elssibony**, Petah-Tikva (IL)

(21) Appl. No.: **14/377,777**

(22) PCT Filed: **Feb. 5, 2013**

(86) PCT No.: **PCT/IL2013/050106**

§ 371 (c)(1),
(2) Date: **Aug. 8, 2014**

Publication Classification

(51) **Int. Cl.**

H02J 7/02	(2006.01)
A61N 1/378	(2006.01)
H02J 5/00	(2006.01)
H03H 7/01	(2006.01)
H02J 17/00	(2006.01)
H03H 7/38	(2006.01)

(52) **U.S. Cl.**

CPC ***H02J 7/025*** (2013.01); ***H02J 17/00*** (2013.01); ***H03H 7/38*** (2013.01); ***H02J 5/00*** (2013.01); ***H03H 7/0138*** (2013.01); ***A61N 1/3785*** (2013.01)

(57) **ABSTRACT**

A system is presented for harvesting electromagnetic energy propagating in surroundings. The system comprises an antenna unit, a harvesting unit, and an input signal adapting circuit. The antenna unit is configured for receiving external electromagnetic radiation from the surroundings and producing a corresponding electric output. The harvesting unit comprises at least one energy harvesting circuit each configured and operable for receiving signals indicative of the output of the antenna unit and generating and storing corresponding electric charge, the harvesting circuit comprising: a rectifying unit comprising a plurality of rectifiers each configured and operable to receive AC electric signals and generate corresponding DC electric power; and a charge collection unit configured and operable to receive the plurality of DC electric powers from said rectifying unit and converting and accumulating them into the electric charge presenting harvested energy. The input signal adapting circuit has an input connected to the antenna unit and an output connected to the rectifying unit, the input signal adapting circuit being configured and operable for adjusting a predetermined electrical property of the antenna unit and rectifying unit to thereby optimize receipt of the electric output of the antenna unit to the harvesting circuit.

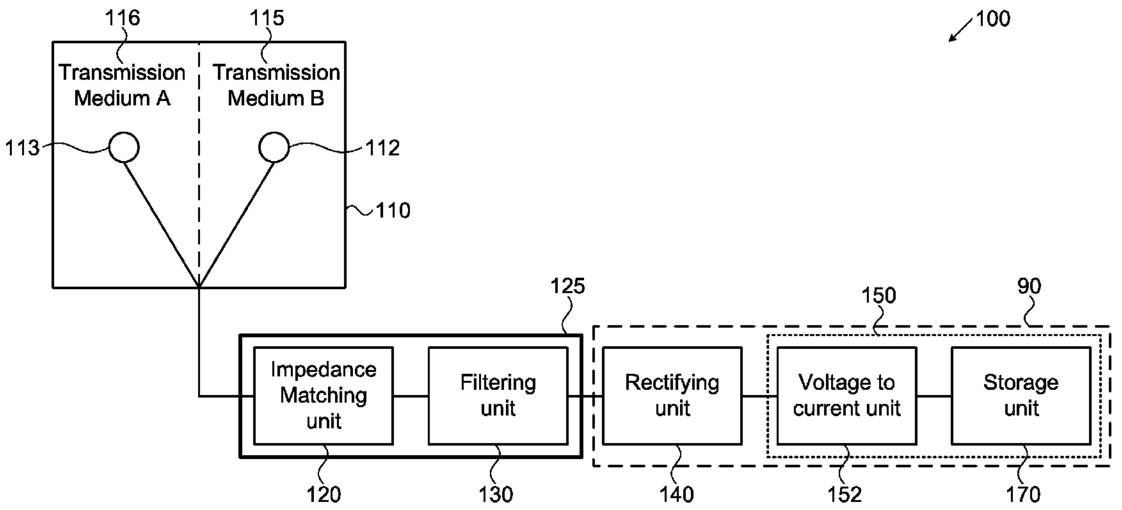

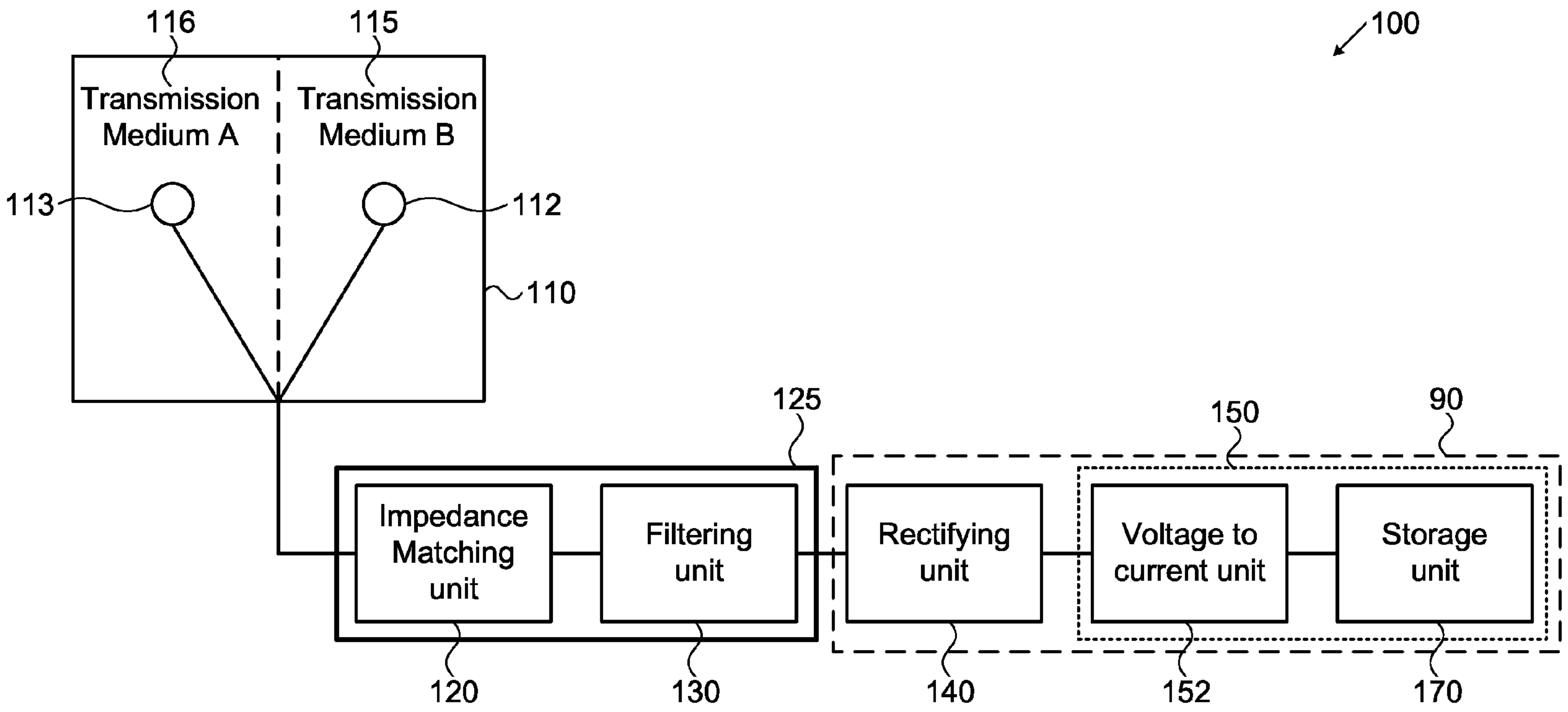
Transmission Medium A
Transmission Medium B
Impedance Matching unit
Filtering unit
Rectifying unit
Voltage to current unit
Storage unit

FIGURE 1A

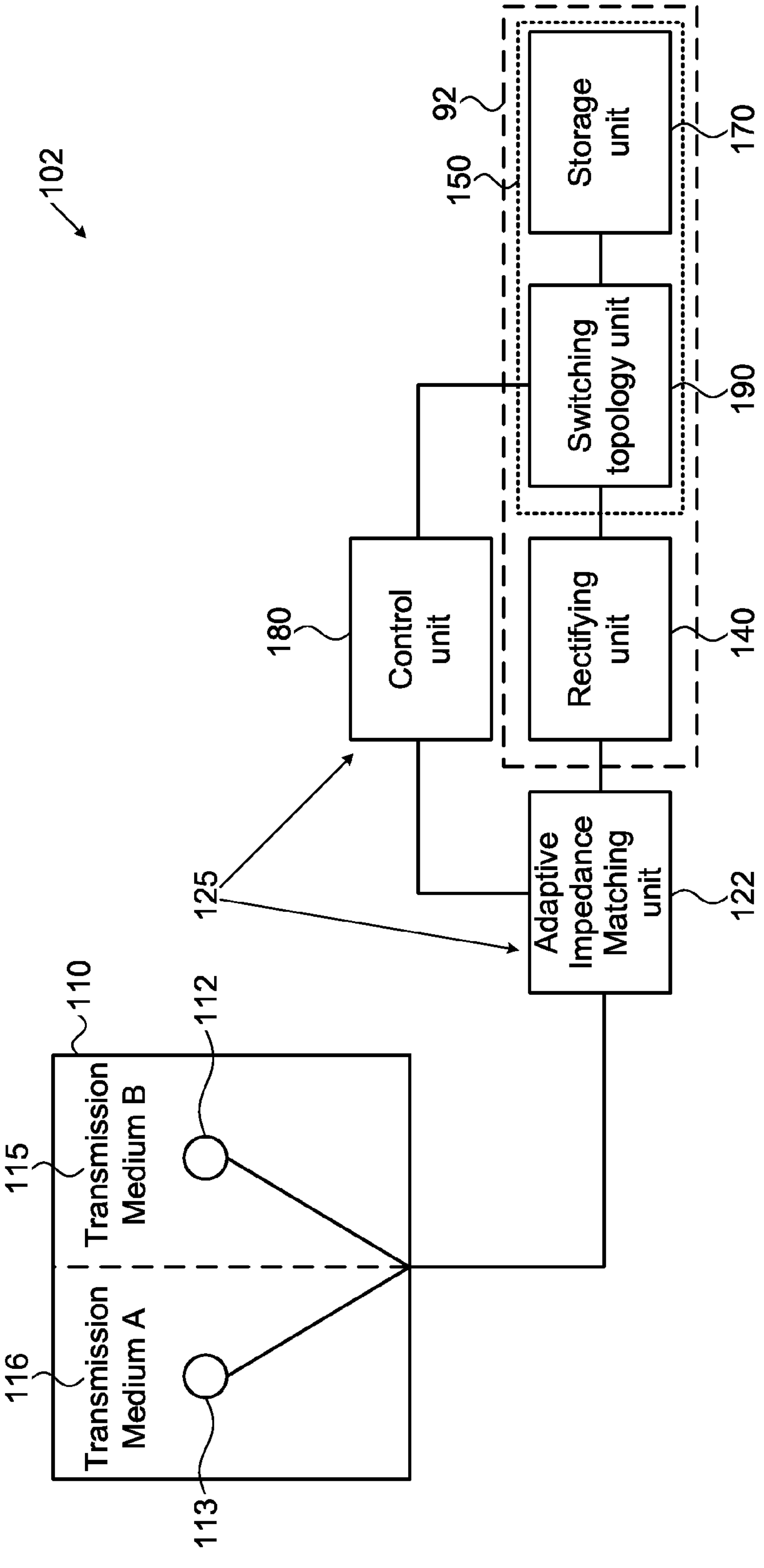
Transmission Medium A
Transmission Medium B
Adaptive Impedance Matching unit
Control unit
Rectifying unit
Switching topology unit
Storage unit

FIGURE 1B

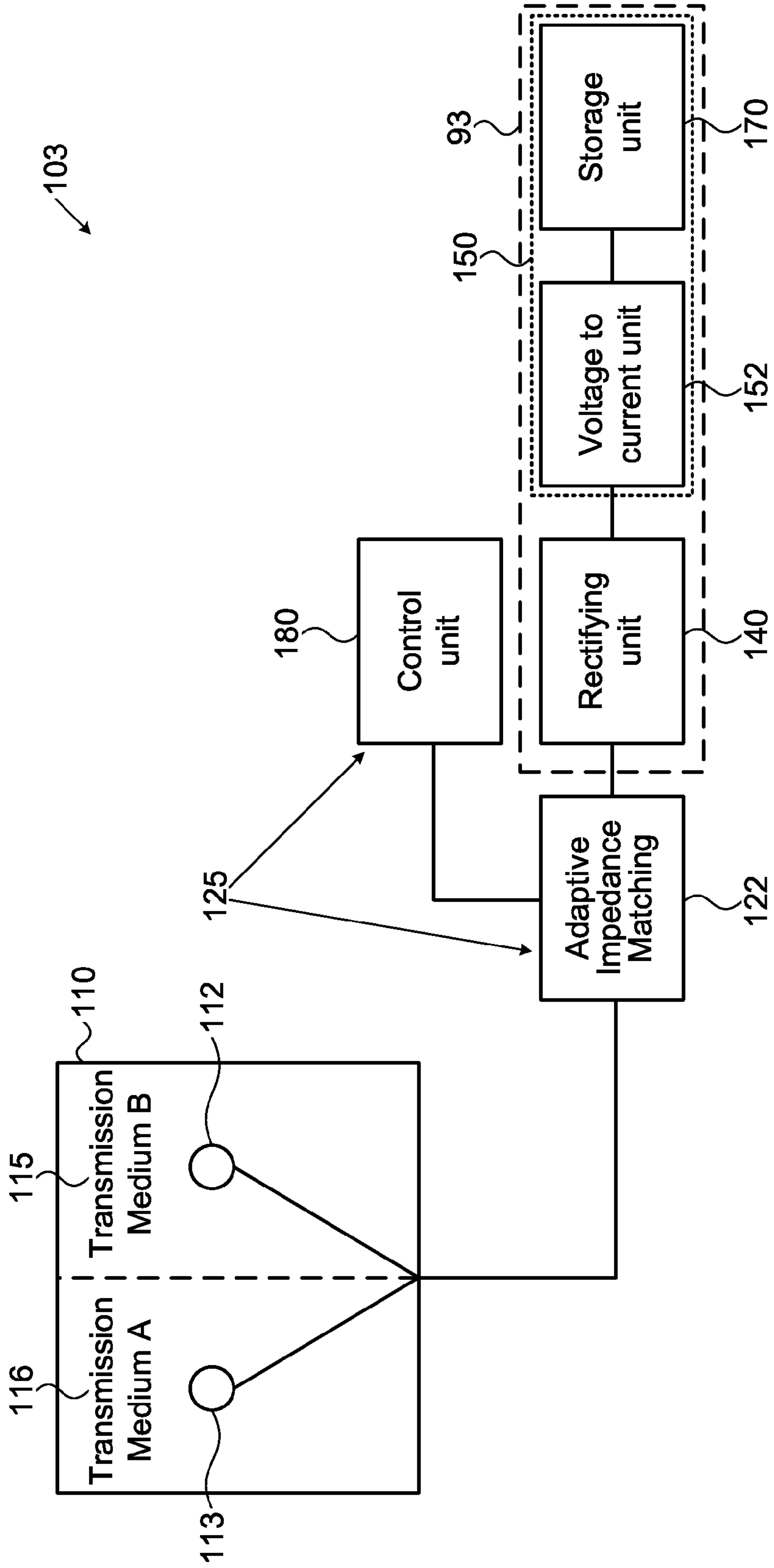
Transmission Medium A
Transmission Medium B
Adaptive Impedance Matching
Control unit
Rectifying unit
Voltage to current unit
Storage unit

FIGURE 1C

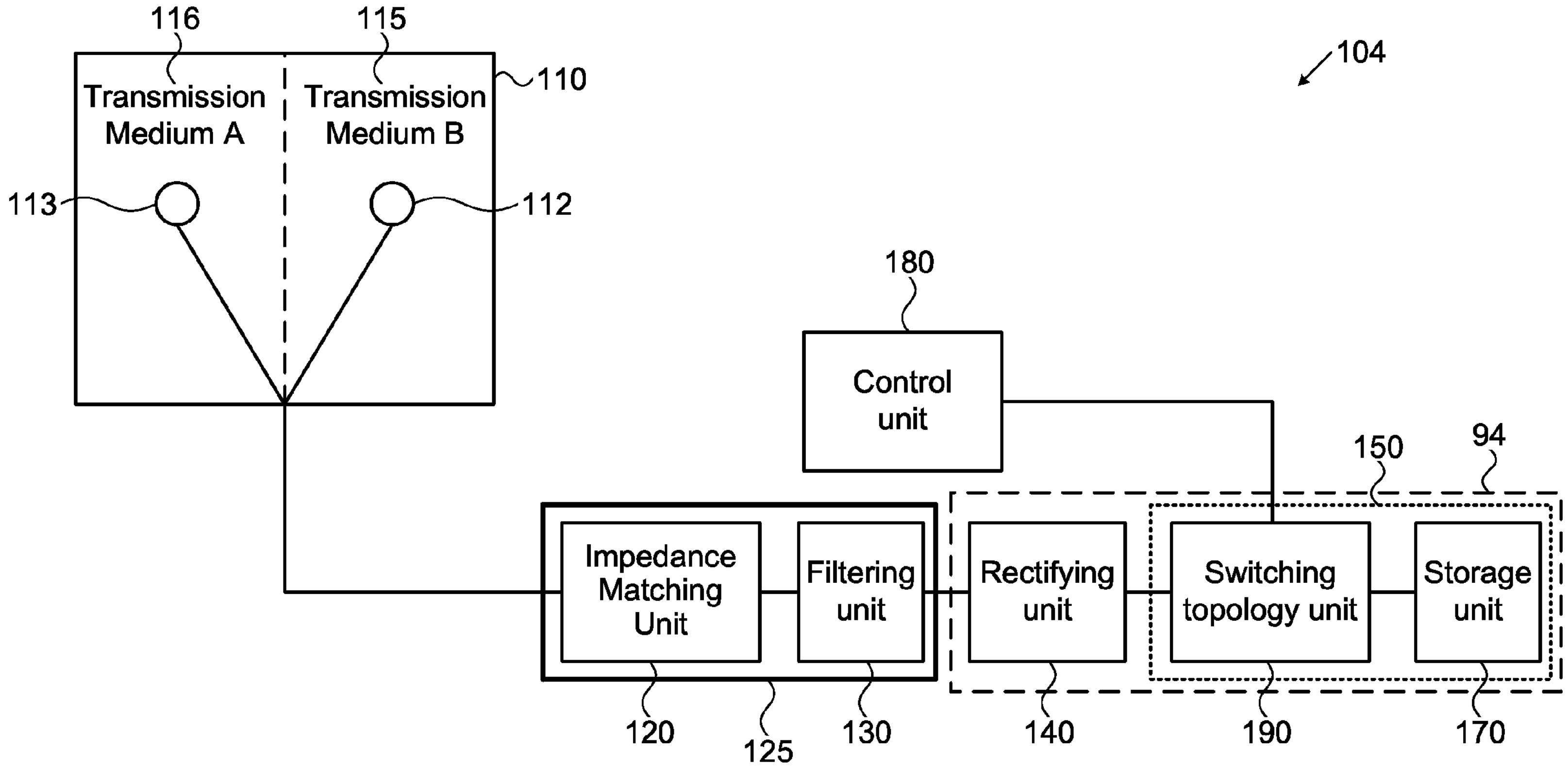
Transmission Medium A
Transmission Medium B
Control unit
Impedance Matching Unit
Filtering unit
Rectifying unit
Switching topology unit
Storage unit

FIGURE 1D

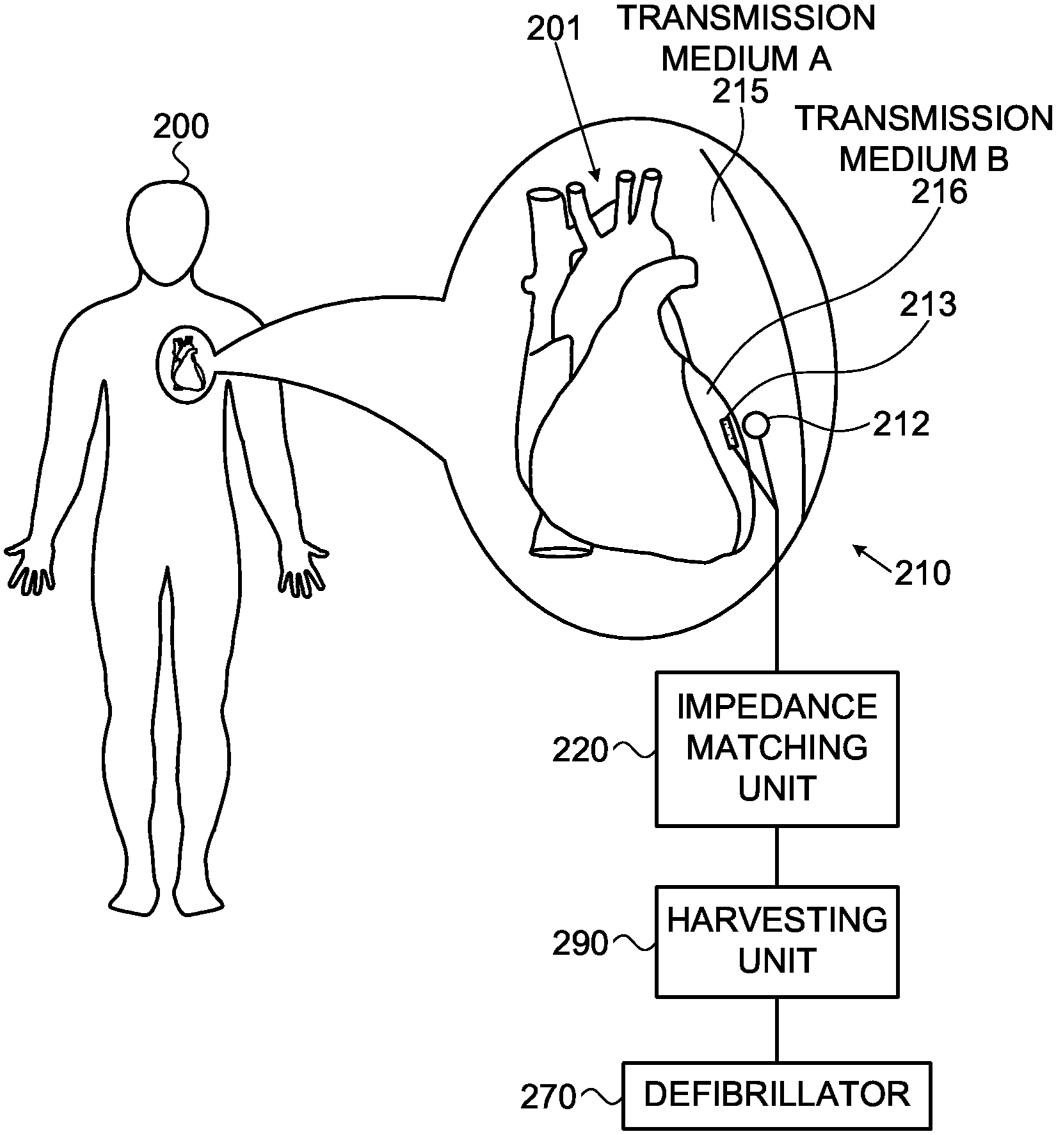
TRANSMISSION MEDIUM A
TRANSMISSION MEDIUM B
IMPEDANCE MATCHING UNIT
HARVESTING UNIT
DEFIBRILLATOR

FIGURE 2

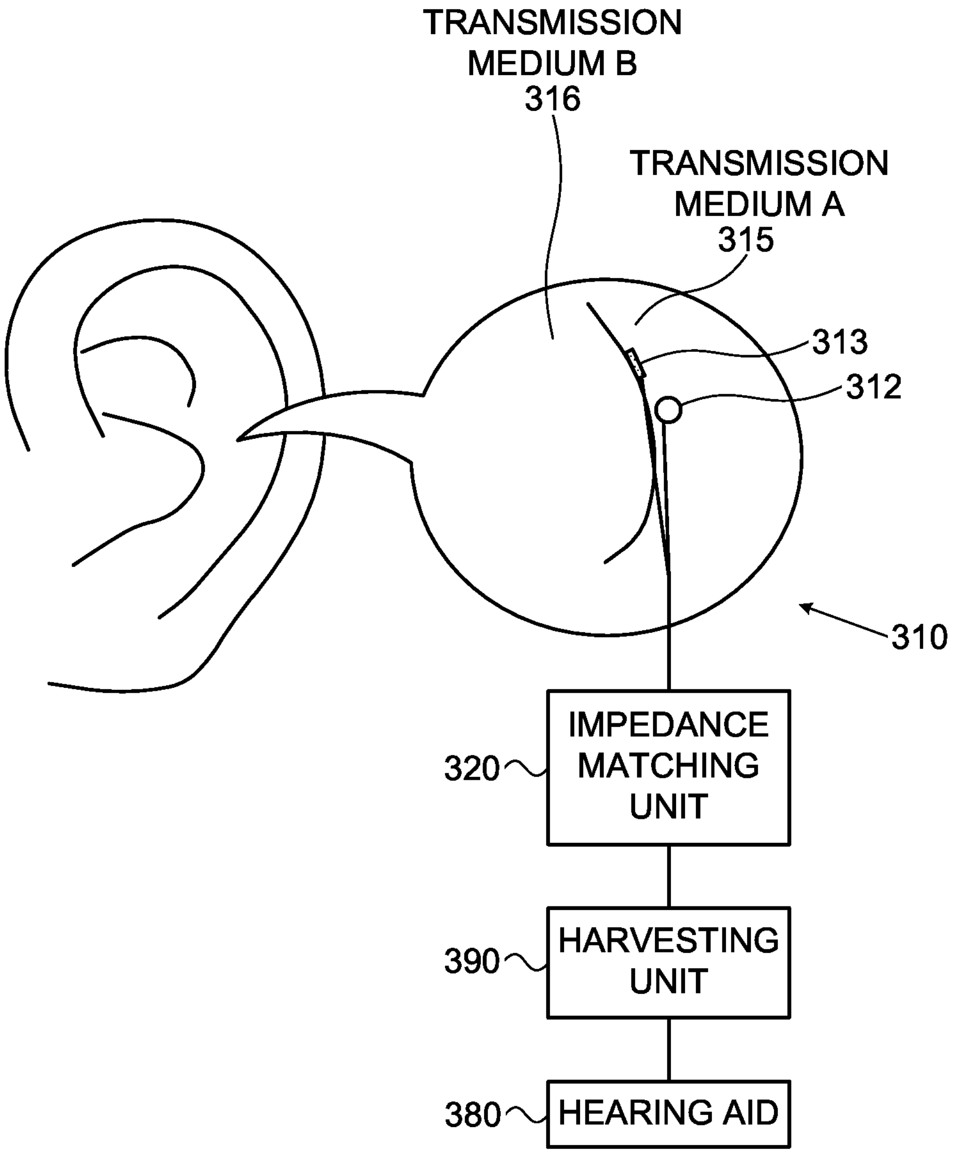
TRANSMISSION MEDIUM B
TRANSMISSION MEDIUM A
IMPEDANCE MATCHING UNIT
HARVESTING UNIT
HEARING AID

FIGURE 3

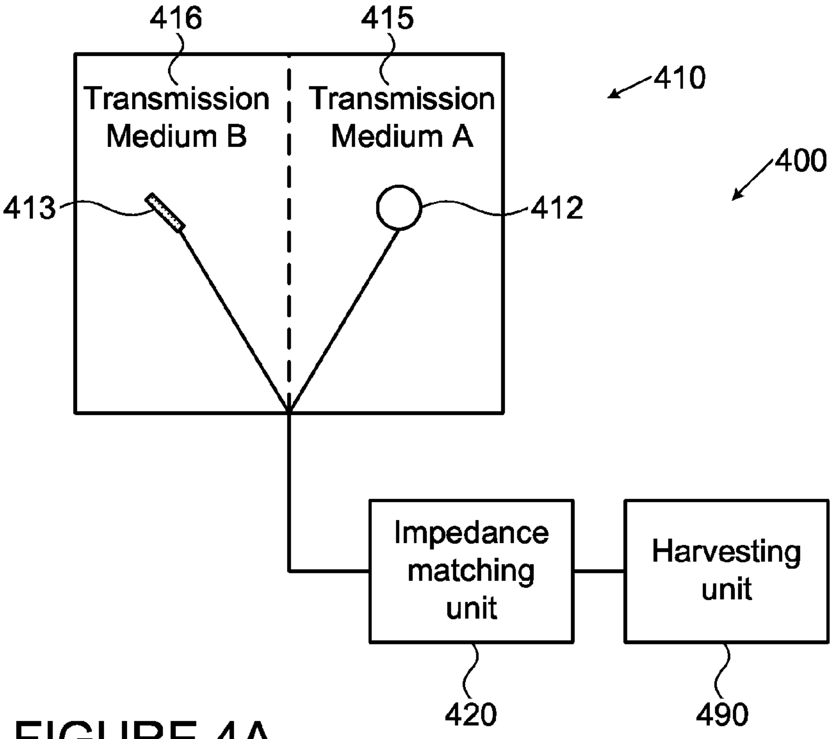
416
415
410
Transmission Medium B
Transmission Medium A
400
413
412
Impedance matching unit
Harvesting unit
420
490

FIGURE 4A

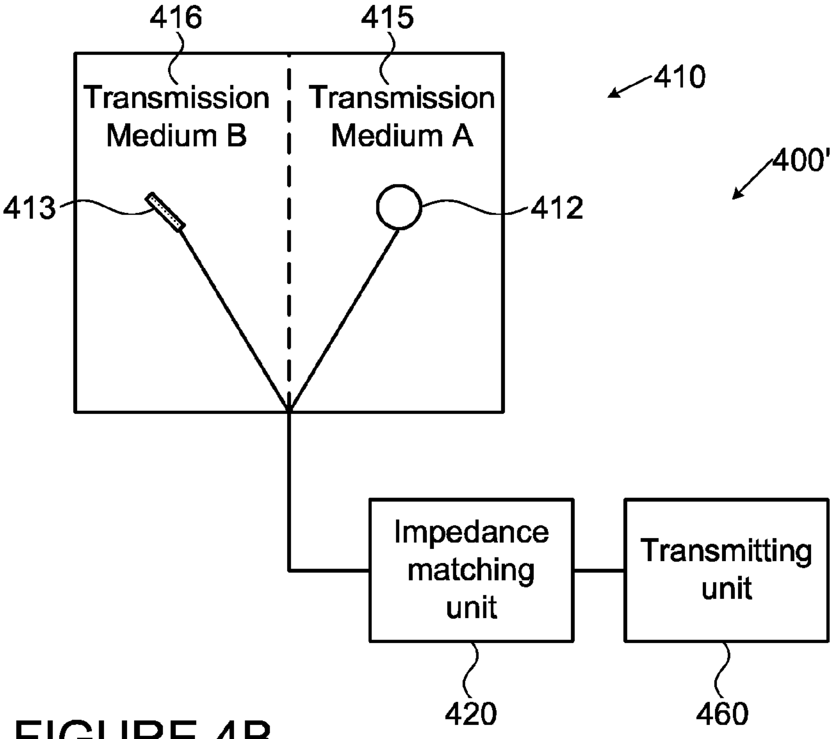
416
415
410
Transmission Medium B
Transmission Medium A
400'
413
412
Impedance matching unit
Transmitting unit
420
460

FIGURE 4B

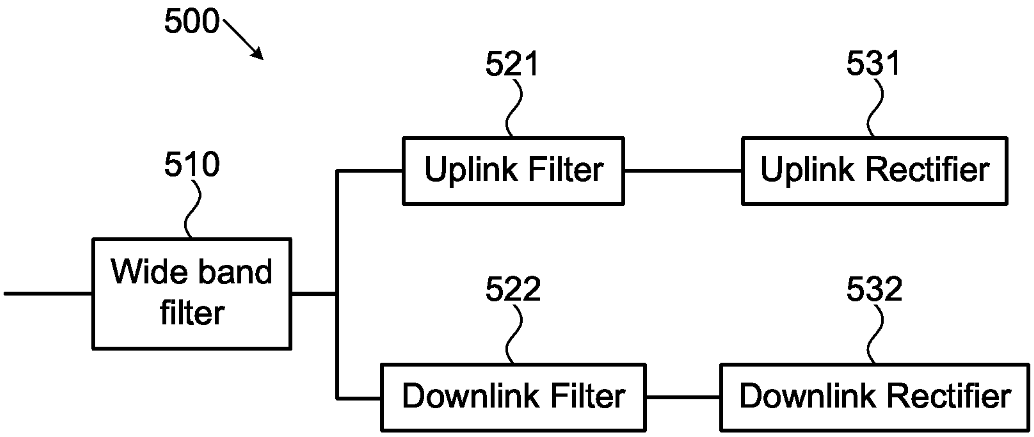
500
521
531
510
Wide band filter
Uplink Filter
Uplink Rectifier
522
532
Downlink Filter
Downlink Rectifier

FIGURE 5

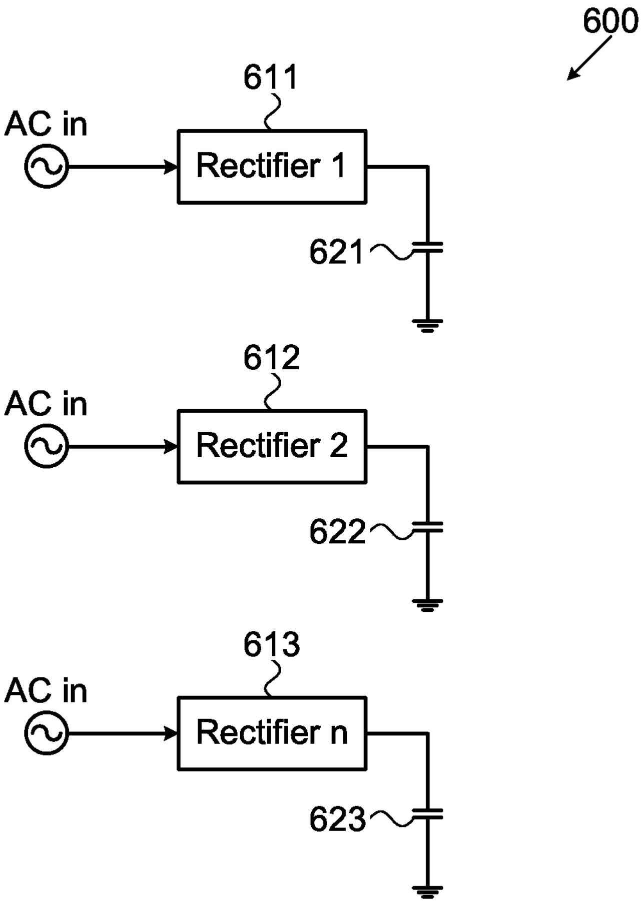
600
611
AC in
Rectifier 1
621
612
AC in
Rectifier 2
622
613
AC in
Rectifier n
623

FIGURE 6

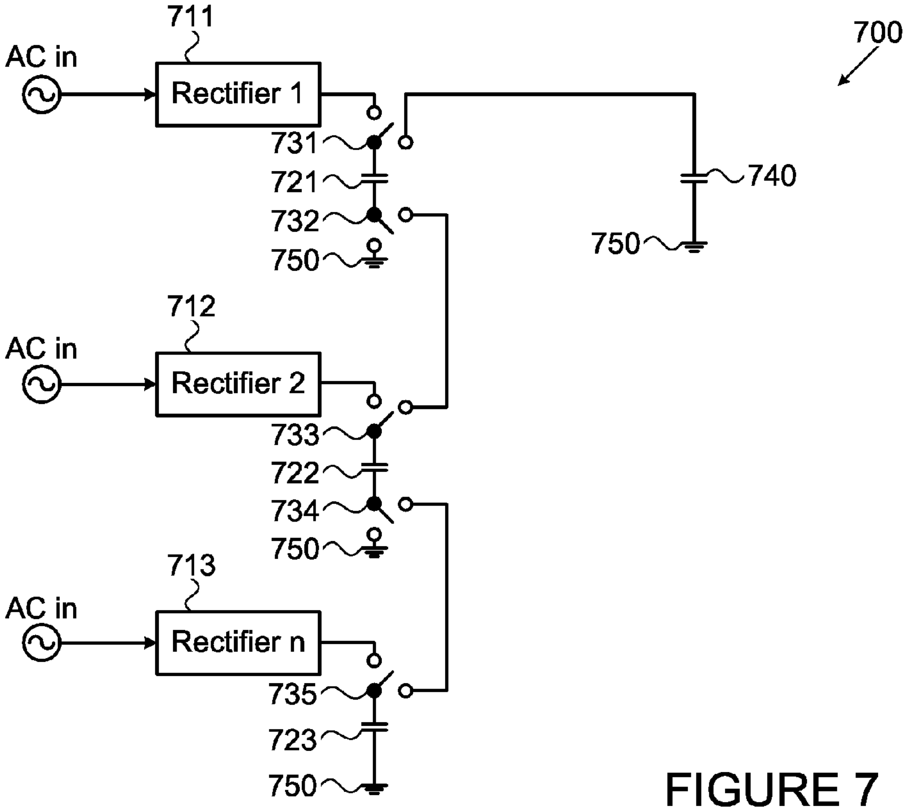
AC in
Rectifier 1
AC in
Rectifier 2
AC in
Rectifier n

FIGURE 7

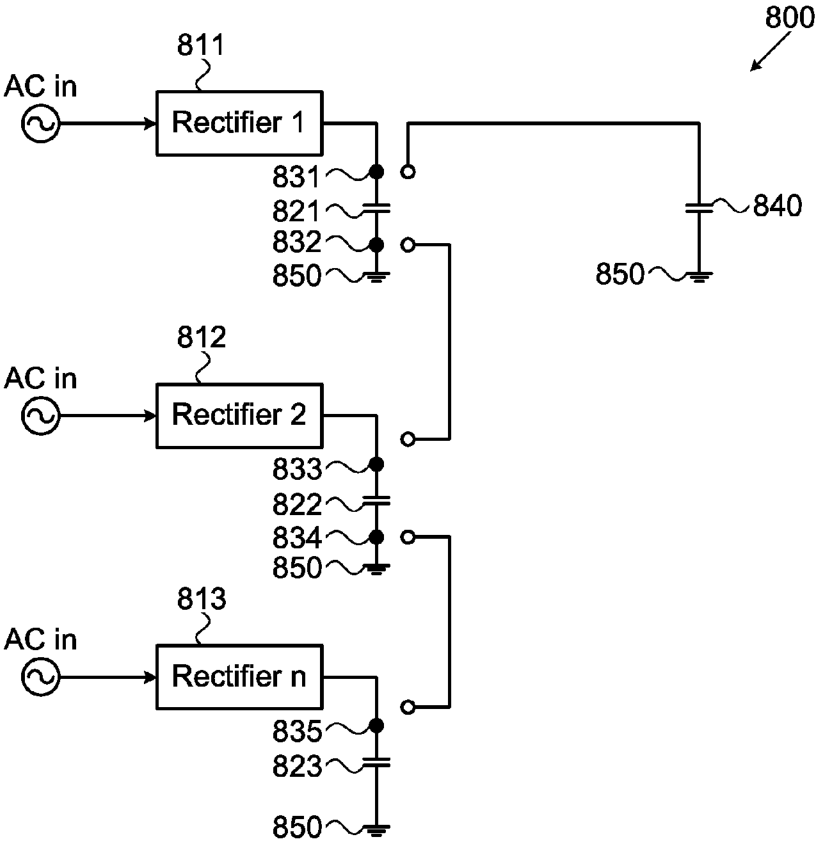
AC in
Rectifier 1
AC in
Rectifier 2
AC in
Rectifier n

FIGURE 8

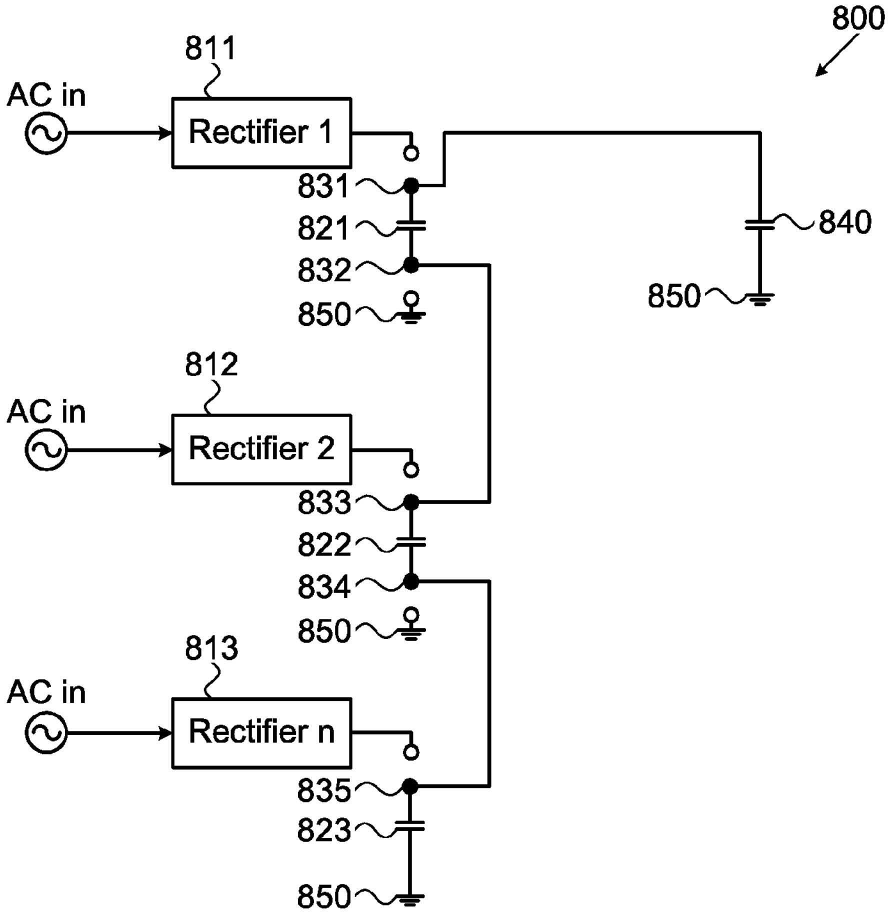
AC in
Rectifier 1
AC in
Rectifier 2
AC in
Rectifier n

FIGURE 9

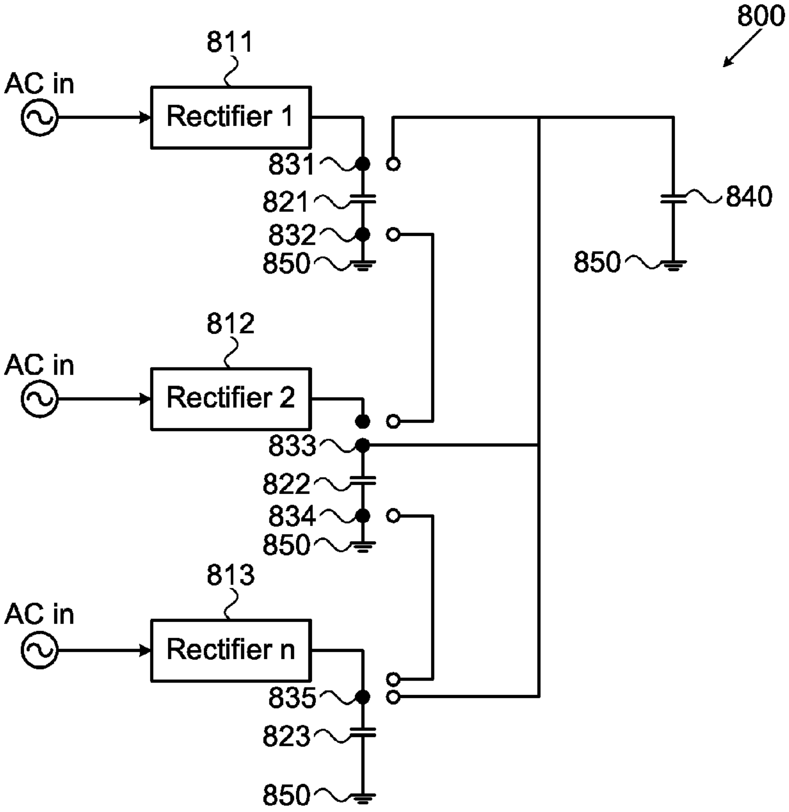
AC in
Rectifier 1
AC in
Rectifier 2
AC in
Rectifier n

FIGURE 10

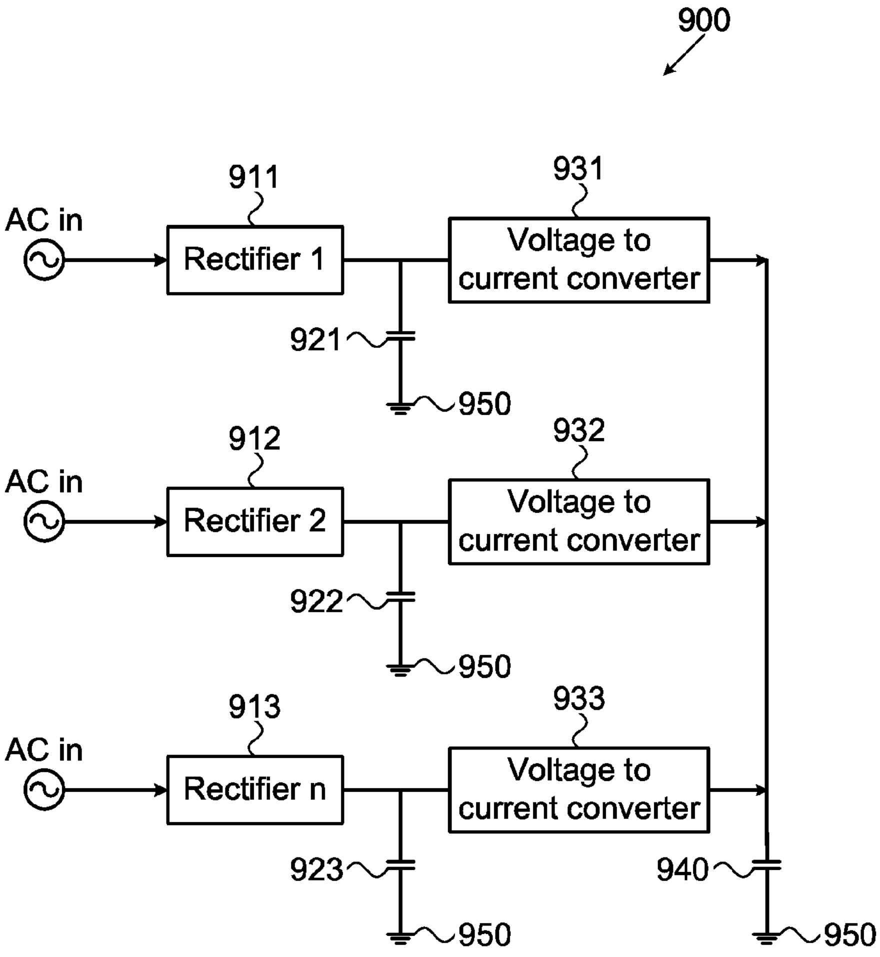
AC in
Rectifier 1
Voltage to current converter
AC in
Rectifier 2
Voltage to current converter
AC in
Rectifier n
Voltage to current converter

FIGURE 11

ENERGY HARVESTING SYSTEM

TECHNOLOGICAL FIELD AND BACKGROUND

[0001] The present invention is in the field of energy harvesting techniques and relates to an energy harvesting system which may be useful in various applications including medical devices (e.g. implantable devices), as well as electric vehicles, and various domestic devices.

[0002] In the last two decades, the usage of electromagnetic communication, such as RF communication, has dramatically increased in all aspects of human life, including mobile/satellite communication, medical treatments, and the like, while different parts of the radio frequency spectrum can be used for different radio transmission technologies and applications (e.g., as TV (television) or cellular applications). Generally, the radio frequency spectrum is shared by civil, government, and military users of all nations, according to International Telecommunications Union (ITU) radio regulations. Thus, people all over the world are surrounded by a variety of electromagnetic signals (radiation), which propagate in the air and have different frequencies and strength.

[0003] There have been several attempts in the art to collect electromagnetic signals which propagate in the air, and to generate electrical power. Some techniques of the kind specified are described for example in the following patents: EP 1,722,284, US 2009/105,782, and JP 10146077.

[0004] In recent years, there has been a rapid technological advance in medical implanted and attached devices. However, the majority of these devices, if not all, require electricity as their energy source. Each such medical device uses batteries as a disposable electricity source. When it comes to implanted devices such as a heart pace maker or a defibrillator, a change of a battery per se requires repetitive surgeries that endanger the patient and expose him to various risks along with the fears before a surgery, inconvenient of the procedure, and recovery period that comes after. When it comes to attached devices such as a hearing aid, the change of batteries that may vary from once a day to once a week according to the device type is merely a nuisance.

[0005] Thus, it is obvious that a mechanism, capable of charging a battery and possibly eliminating a need for battery in such devices, is desired.

GENERAL DESCRIPTION

[0006] There is a need in the art in a novel approach for energy harvesting enabling efficient collection and use of electromagnetic radiation (e.g., RF signals) and storage of the harvested energy in order to be supplied to an electrical load, such as a rechargeable battery or an electric device to enable for example, charging said battery or operating said electric device, in a most efficient manner. A technique for harvesting electromagnetic energy is described in U.S. patent application Ser. No. 13/116,554 of the same inventor incorporated herein in its entirety by reference. Such system functionally allows gathering the already existing RF radiation from the surrounding and converting it into usable energy source.

[0007] The present invention provides a novel energy harvesting system and method, as well as a novel antenna unit for use in an energy harvesting system. The antenna unit of the invention is configured for receiving external electromagnetic radiation from the surroundings and producing a corresponding output electric signal. The antenna unit comprises at least two electrically conductive elements associated with (e.g. positioned in) different transmission media having different electric and/or magnetic properties.

[0008] When harvesting RF energy from sources available in the air, the efficiency of the system is of critical importance due to the variety of frequencies and intensities of received signals. One of important factors determining the efficiency of a harvesting system is a degree of match/adjustment of a predetermined electrical property of the antenna unit and that of the rectifying unit. More specifically, this may be a match between the impedance of a rectifying circuit and the intensity of an input signal. The rectifier impedance is dynamically changing as a function of the intensity of the input signal. More specifically, when the input signal is a low power signal relative to the conductive threshold of the rectifier, the impedance of the rectifier is very high. As the intensity of the input signal increases and approaches the conductive threshold of the rectifier, the impedance of the rectifier decreases, until the intensity of the signal exceeds the conductive threshold of the rectifier and the impedance of the rectifier decreases. In a static system, the rectifier is configured such that it impedance matches a certain level of intensity of an input signal, but in other intensities it is not matched and the efficiency decreases significantly. Thus, there might be a need for an input signal adapting unit/circuit, e.g. an adaptive impedance matching unit, that is configured and operable for dynamically/selectively matching the rectifier impedance according to the changing input signal intensity to the impedance of the antenna unit.

[0009] One way of fulfilling a requirement for a dynamic change of impedance of the rectifier is by providing a dynamic active change of the input impedance. This technique however suffers from energy consumption leading to energy waste.

[0010] It is thus desired in some applications that such a dynamic impedance matching would be implemented by a passive unit. In this connection, it should be understood that the terms “passive unit” and “active unit” used herein refer to units/circuits which are, respectively, autonomously operable based on their predefined working parameters/conditions and operable by a dedicated controller to vary one or more of the working parameters/conditions. The invention provides such a passive filtering unit that operates to match the impedances of the rectifiers of different input intensity signals (received from the antenna unit) and maximize the efficiency of the harvesting system in all input intensity levels.

[0011] When harvesting electromagnetic energy, a capacitor of each rectifier is usually small. However, there is a need for storage of large quantities of charge originated from several rectifiers and from the same rectifier over time. The present invention provides a novel charge collection unit which may be based on either one of the two following novel approaches for gathering the small charges obtained during the harvesting process into a large storage (i.e. capacitor with relatively large capacitance). This is implemented in the invention either by using a switching topology circuit or by using appropriate conversion of voltage to current allowing transferring the small charges obtained to a large capacitor to enable large electric charge storage. To this end, the invention provides a charge collection unit comprising a signal summing unit which is configured and operable to receive a plurality of DC electric powers from multiple rectifiers of a rectifying unit and convert and accumulate them into the electric charge (via a voltage to current conversion or a switching topology circuit), where this charges presents the harvested energy.

[0012] Thus, according to one broad aspect of the present invention, there is provided a system for harvesting electromagnetic energy propagating in the surroundings to which the system is exposed. The system comprises: an antenna unit configured for receiving external electromagnetic radiation from the surroundings and producing a corresponding electric output; a harvesting unit comprising one or more harvesting circuits wherein each harvesting circuit is configured and operable for receiving signals indicative of the output of the antenna unit and generating and storing corresponding electric charge presenting the harvested energy; and an input signal adapting circuit which has an input connected to the antenna unit and an output connected to said at least one harvesting circuit of the harvesting unit and is operable for adjusting a predetermined electrical property of the antenna unit and said at least one harvesting circuit to thereby optimize receipt of the electric output of the antenna unit by said at least one harvesting circuit.

[0013] It should be noted that in this disclosure terms "unit" and "circuit" are used interchangeably because each of them actually refers to one or more electric circuits defining certain functionality for affecting/processing an electric input.

[0014] The harvesting circuit comprises a rectifying unit comprising a plurality of rectifiers each configured and operable to receive AC electric signals and generate corresponding DC electric power; and a charge collection unit configured and operable to receive the plurality of DC electric powers from said rectifying unit and converting and accumulating them into the electric charge presenting harvested energy.

[0015] Generally, the antenna unit may be of any known suitable configuration. Preferably, however, a novel antenna structure of the invention can be used. Such novel antenna structure comprises at least two electrically conductive elements associated with different transmission media differing in at least one of electric and magnetic properties. It should be understood that such a structure is termed herein as "antenna unit" solely in the meaning that it is responsive to electromagnetic radiation to which the structure is exposed (typically RF radiation) and providing a corresponding electric output (response signal).

[0016] In some embodiments, such at least two different transmission media may for example include at least two different animate tissues (in which case the antenna unit may be implanted in a subject); or may include animate tissue and air (in which case the antenna unit is attached to a subject). It should, however, be understood that the antenna unit of the invention is not limited to these examples, as well as is not limited to any specific media.

[0017] The input signal adapting unit is configured and operable to optimize receipt of the antenna output to the rectifying unit of the harvesting circuit by matching between the electric parameter, e.g. impedance, of the antenna unit and the rectifying unit.

[0018] In some embodiments, the input signal adapting unit comprises an impedance matching unit and a filtering unit connected to the impedance matching unit. The impedance matching unit is configured to match between the impedances of the antenna unit and the filtering unit. The filtering unit is configured and operable for sorting the received RF band propagating in the surroundings into a predefined number of subbands with predetermined values of central frequency, band width and strength for each subband. The predefined number of subbands corresponds to a number of the rectifiers in the rectifying unit.

[0019] In some embodiments, the input signal adjusting unit comprises an adaptive impedance matching unit, and a control unit connectable to the adaptive impedance matching unit and configured and operable operation thereof. The adaptive impedance matching unit is configured and operable to controllably adjust impedance of the rectifiers of the rectifying unit to signal intensities of multiple frequency bands received by the antenna unit.

[0020] In some embodiments, the charge collection unit comprises a signal summing unit comprising a switching topology circuit and a controller. The switching topology circuit comprises a predetermined number of capacitors corresponding to a number of the rectifiers in the rectifying unit, all electrically connected to a storage capacitor. The capacitors connected to output of the respective rectifiers have relatively small capacitance values, while the storage capacitor has a relatively large capacitance value. The switching topology circuit is adapted to controllably switch between the plurality of DC signals received from the plurality of rectifiers respectively with the corresponding plurality of the "small" capacitors for power summation and storage of corresponding charge on the "large" capacitor.

[0021] In some embodiments, the charge collection unit comprises a signal summing unit comprising a voltage to current unit. The voltage to current unit comprises a predetermined number of voltage to current conversion circuits having inputs connected to the outputs of the corresponding number of the rectifiers respectively, and having outputs connected to a storage a capacitor. The voltage to current unit is thereby adapted to convert and sum the plurality of DC signals/powers received from the plurality of rectifiers via the plurality of voltage to current conversion circuits.

[0022] The system may further comprise a control unit connected to at least one of the input signal adapting unit (e.g. impedance matching unit) and the charge collection unit. Considering the control unit is connected to at least the charge collection unit comprising a switching topology circuit described above, the control unit operates the switching topology to controllably switch between the plurality of DC signals received from the plurality of rectifiers respectively with the corresponding plurality of "small" capacitors of the relatively small capacitance values for power summation and storage of corresponding charge on the "large" capacitor.

[0023] It should be noted that the same input signal adapting unit may be associated with multiple harvesting circuits. Also, the same charge collection unit may be associated with multiple harvesting circuits by adjusting the number of inputs of the charge collection unit to the total number of rectifiers defined by the multiple harvesting circuits.

[0024] The harvesting unit may comprise a connection port for connecting output of the charge collection unit (defined by charge storage utility in the form of a large capacitor) to at least one electrical load. The at least one electrical load may include at least one of the following: a rechargeable battery, a medical device, an electric vehicle component, an electric device for home use. The medical device may include at least one of the following: a pacemaker, a defibrillator, and a hearing aid. The electric vehicle component may include at least one of the following: electrical accessories, audio systems and amplifiers, TV and DVD systems, GPS systems, air conditioning systems, alarm systems, lights and wiring systems.

[0025] In accordance with another broad aspect of the invention there is provided an antenna unit configured and operable for receiving external electromagnetic radiation from surrounding and producing a corresponding output electric signal. The antenna unit comprises at least two conductive elements positioned in at least two different transmission media being different in electric and/or magnetic properties. The received external electromagnetic radiation creates a potential difference between the at least two conductive elements producing an output electric signal, thereby allowing uptake of the external electromagnetic radiation from the surroundings.

[0026] As mentioned above, in some specific but not limiting examples, the different transmission media may include different animate tissues; or animate tissue and air.

[0027] In accordance with yet another broad aspect of the present invention, there is provided an input signal adapting unit for use with an antenna exposed to electromagnetic radiation of RF band propagating in surroundings. The input signal adapting unit comprises an impedance matching unit and a filtering unit connected to the impedance matching unit. The impedance matching unit is configured and operable to match between impedances of the antenna unit and the filtering unit, and the filtering unit is adapted for sorting the RF band into a predefined number of subbands of a predetermined frequency, bandwidth and strength for each subband.

[0028] The present invention in its further broad aspect is directed to a charge collection unit comprising a signal summing unit adapted to receive a plurality of DC electric signals and sum them in the form of an electric charge on a storage capacitor, the signal summing unit comprising at least one of a switching topology unit and a voltage to current unit.

[0029] The switching topology unit may comprise a predetermined number of capacitors having relatively small capacitance values and being connectable to output of a corresponding number of rectifiers respectively, and connected to said storage capacitor of a relatively large capacitance value. The switching topology unit is adapted to controllably switch between DC sources received from the rectifiers with the predetermined number of the capacitors for power summation.

[0030] The voltage to current unit may comprise a predetermined number of voltage to current conversion circuits having inputs connectable to outputs of a corresponding number of rectifiers respectively, and outputs connected to the storage capacitor. The voltage to current unit is thereby configured and operable to convert and sum a plurality of DC sources received from the plurality of rectifiers via the voltage to current conversion circuits.

BRIEF DESCRIPTION OF THE DRAWINGS

[0031] Examples illustrative of variations of the disclosure are described below with reference to figures attached hereto. In the figures, identical structures, elements or parts that appear in more than one figure are generally labeled with the same numeral in all the figures in which they appear. Dimensions of components and features shown in the figures are generally chosen for convenience and clarity of presentation and are not necessarily shown to scale. The figures presented are in the form of schematic illustrations and, as such, certain elements may be drawn greatly simplified or not-to-scale, for illustrative clarity. The figures are not intended to be production drawings.

[0032] The figures (Figs.) are listed below.

[0033] FIG. **1**A-**1**D are block diagrams schematically illustrating several examples of the configuration of an energy harvesting system of the present invention; wherein in the example of FIG. **1**A the operation mode of major components of the system is a passive mode that do not require additional adaptations for obtaining proper operation of the harvesting system; in the example of FIG. **1**B the operation mode of major components of the system is an active mode that requires adjustment of the operation of specific units of the system to a current condition of the harvesting process; and FIGS. **1**C and **1**D illustrate two examples of the system utilizing semi-active mode of the operation mode of major components of the system.

[0034] FIG. **2** is a schematic illustration of an example of an antenna unit of the present invention, being used with a heart pacemaker device implanted into a human body.

[0035] FIG. **3** is a schematic illustration of another example of the antenna unit of the invention being used with a hearing aid inserted into an individual's ear.

[0036] FIGS. **4**A-**4**B are block diagrams schematically illustrating an example of a two-elements antenna system of the invention configured and operable for receiving energy from the surrounding (FIG. **4**A), or connected to a transmitting unit (FIG. **4**B).

[0037] FIG. **5** is a schematic block diagram illustration of a filtering unit configured for maximizing the efficiency in an RF harvesting system.

[0038] FIG. **6** is a schematic block diagram illustration of a standard circuit for charging capacitors from rectifying elements.

[0039] FIG. **7** is a schematic block diagram illustration of a system for accumulating charge from small capacitors to a large capacitor.

[0040] FIG. **8** is a schematic block diagram illustration of a system for accumulating charge from small capacitors to a large capacitor in a charging switches position.

[0041] FIG. **9** is a schematic block diagram illustration of a system for accumulating charge from small capacitors to a large capacitor in a discharging switches position.

[0042] FIG. **10** is a schematic block diagram illustration of a system for accumulating charge from small capacitors to a large capacitor in a single discharging switch position.

[0043] FIG. **11** is a schematic block diagram illustration of a system for accumulating charge from small capacitors to a large capacitor by converting the stored electric charge to electric current source.

DETAILED DESCRIPTION OF THE EMBODIMENTS

[0044] Although various features of the disclosure may be described in the context of a single embodiment, the features may also be provided separately or in any suitable combination. Conversely, although the disclosure may be described herein in the context of separate embodiments for clarity, the disclosure may also be implemented in a single embodiment. Furthermore, it should be understood that the disclosure can be carried out or practiced in various ways, and that the disclosure can be implemented in embodiments other than the exemplary ones described herein below. The descriptions, examples and materials presented in the description, as well as in the claims, should not be construed as limiting, but rather as illustrative.

[0045] Reference is made to FIGS. **1A-1D** which show, by way of block diagrams, several embodiments of the configuration of an energy harvesting system of the present invention. The system includes such main constructional parts as an antenna unit **110**, and an energy harvesting unit **90** including one or more harvesting circuits each receiving and processing an output of the antenna unit **110**; and an input signal adapting unit **125**. The harvesting circuits includes a rectifying unit **140**, and a charge collection arrangement **150** which either includes a storage capacitor, or is connected to a separate storage unit **170** having such storage capacitor, for accumulating charge presenting harvested energy.

[0046] FIG. **1A** illustrates the energy harvesting system configuration **100**, in which the operation mode of major components is a passive mode, i.e. the mode that doesn't require additional adaptations for obtaining proper operation of the harvesting system. The system **100** includes the antenna unit **110**, the energy harvesting unit **90**, and the input signal adapting unit **125**. The harvesting unit **90** includes the rectifying unit **140**, and the charge collection arrangement **150**. In this example, the input signal adapting unit **125** includes an impedance matching unit **120** and a filtering unit **130** the operation of which will be described below. Also, in this example, the charge collection arrangement **150** includes a voltage to current converter **152**, as will be described more specifically further below.

[0047] Generally, the antenna unit may be of any known suitable kind. Preferably, the antenna unit **110** used in the system **100** utilizes the novel antenna configuration of the present invention. The antenna unit **110** of this novel configuration includes at least two electric conductors (conductive elements) **112** and **113**, which are positioned in at least two different substances constituting transmission media **115** and **116**, having different electrical and/or magnetic properties. Due to the different electrical and/or magnetic properties of substances **115** and **116**, RF radiation in the surrounding of the antenna unit creates different electric potentials on the conductive elements **112** and **113**. The potential difference between the two conductive elements is received by the harvesting unit **90** via the harvesting circuit and is then processed by the input adapting unit **125**. To this end, the signal undergoes impedance matching by unit **120** and filtering by unit **130** that operate together to adjust the impedance of antenna unit **110** to the impedance of the rectifying unit **140** to thereby allow optimal delivery of the received signal.

[0048] Generally, impedance matching unit **120** is used to enable transferring maximal electromagnetic energy/power from the antenna unit to the harvesting unit. In the configuration illustrated in this figure, the impedance of antenna unit **110** includes both the impedances of the transmitting media and the conductive elements that transfer the received energy into the harvesting unit. In order to transfer maximal electromagnetic power from the antenna into the harvesting unit, the impedance of the antenna, $Z_{Antenna}$, needs to be equal (generally as close as possible) to the complex conjugate of the harvesting unit impedance, $Z^*_{Harvesting\ unit}$, namely the following condition is to be satisfied:

$$Z_{Antenna}=Z^*_{Harvesting\ unit}$$

[0049] Impedance matching unit **120** may comprise appropriate electrical circuits which may contain a combination of capacitive elements and inductive elements. Generally, the impedance matching unit **120** may be of any suitable configuration known in the art.

[0050] The received signals are then delivered to the filtering unit **130**. Filtering unit **130** is configured and operable as a spectral filter and includes one or more arrays/sets of filters which sorts/divides the signals received from antenna unit **110** to predefined frequency bands. The spectral filtering process in accordance with the present invention is a non active process, in the meaning that is mainly directed to sort different signal frequencies that are characterized by having different strengths. A specific but not limiting example of a filtering unit is described in details further below with reference to FIG. **5**. The output of each of the last filters in the sets/arrays of filters is connectable to a dedicated rectifier in the plurality of rectifiers comprised in the rectifying unit **140**. The rectifier is configured to convert the respective filtered RF signal into a DC signal according to pre-expected signal strength, thus, eliminating the need for active matching process that usually consumes energy.

[0051] The rectifying unit **140** includes an array of rectifiers associated with a signal summing utility of the charge collection unit. The output of each rectifier is connected to such a signal summing utility that is adapted to receive plurality of DC electric signals, and perform their summation so as to allow storage thereof in a storage capacitor/unit **170**.

[0052] In this specific but not limiting example of FIG. **1A**, the output of each rectifier in the rectifying unit **140** is connected to the respective voltage to current conversion circuit in the voltage to current unit **152**. Voltage to current convertor unit **152** is configured and operable for summing several DC sources into one storage unit **170**. It should be noted that the voltage to current unit **152** is preferably configured to allow for passively accumulating several DC sources into the common storage without the need for active operations and without causing shortcuts and/or other interfering influences between the summed DC sources as normally occurs while connecting directly several DC voltage sources, as will be described in details hereinafter with reference to FIG. **11**. The storage unit **170** is configured to collect the entire harvested energy, and it may be further connected to one or more electrical loads (not shown) such as a rechargeable battery or an electric device.

[0053] FIG. **1B** illustrates another example of the energy harvesting system configuration **102** in which the operation mode of major components is an active mode that requires a control unit for adjusting the operation of specific units of the system to a current condition of the harvesting process. The system **102** is configured generally similar to the above described system **100** in that it comprises an antenna unit **110** which is preferably configured according to the invention, namely includes two conductors **112** and **113** positioned in different transmission media **115** and **116**, and comprises a harvesting unit **92** and an input signal adapting unit **125**. The harvesting unit **92**, similar to the above-described example includes, a rectifying unit **140**, and a charge collection arrangement **150**. In this example, the input signal adapting unit **125** includes an adaptive impedance matching unit **122** and a control utility **180**. The adaptive impedance matching unit **122** is configured to adjust the impedance of antenna unit **110** to the impedance of the rectifying unit **140** to thereby allow optimal delivery of the received signal.

[0054] Generally, in order to harvest energy from electromagnetic waves there is a need to rectify the harvested energy that is received from the antenna unit to thereby receive a DC electrical energy that can be reused as an energy source. The input characteristic impedance of the rectifying unit is not stable and can be varied from some Ohms to Mega Ohms as a function of the input RF level on the rectifier input. In addition, the antenna unit characteristic impedance can be stable or unstable. One technique that allows for adapting stable or unstable antenna to an unstable input impedance front end circuit is an adaptive impedance matching technique.

[0055] Adaptive impedance matching unit **122** is configured and operable by the control unit **180** to constantly adapt the antenna impedance into an unstable characteristic impedance of the harvesting circuit. To this end, the control unit **180** operates to control the parameters of the impedance matching unit to match the impedance of the rectifier to that of the antenna unit. The control unit **180** is operable to vary the impedance of the impedance matching unit **122** in accordance with the input voltage on the harvesting circuit. The control unit **180** is configured and operable to maximize the efficiency of the electrical system by varying the impedance of the impedance matching unit **122** in accordance with electrical characteristics measured in predetermined locations along the circuit. The control unit **180** operates to vary at least one of input and output impedance values of the impedance matching unit to maximize the voltage in predetermined locations. The control unit may be an electronic unit preferably configured of linear electronic elements or a software utility operating on an electronic device. The control unit has at least two input ports for detecting electrical parameters at said predetermined locations, a comparator module for comparing the detected parameters, and at least one output port for providing a control signal to the impedance matching unit to thereby vary its impedance. In the specific example illustrated in FIG. 1B, the control unit **180** is adapted to control the adaptation of the impedance according to the strength of the signal received.

[0056] The received signal is then delivered to a rectifier in the rectifying unit **140** that is adjusted to convert the RF signal into a DC signal. The output of each rectifier of the rectifying unit **140** is connected to the charge collection arrangement **150**. In the example of FIG. 1B, the charge collection arrangement **150** includes a switching topology unit **190** and a control utility and is configured and operable to allow summing several DC sources into one storage capacitor unit **170**. The control utility of the charge collection arrangement **150** and that of the input signal adapting unit **125** may be incorporated in separated controllers or may be parts of the same control unit **180**. The switching topology unit **190** includes several switching circuits, the operation of which is controlled by the respective control utility (e.g. in the control unit **180**). A detailed description of an example of the switching topology unit is described below with reference to FIG. **6-10**. The storage unit **170** collects the entire harvested energy, and as mentioned above may further be connected to a rechargeable battery or to an electrical device.

[0057] FIG. 1C exemplifies a harvesting system **103** in which the operation mode of major components is a semi-active mode that requires a control unit for adjusting the operation of the system to the current stage of the harvesting process. The system **103** is configured generally similar to the above described systems **100** and **102** comprising an antenna unit **110** preferably including two conductors **112** and **113** positioned in different transmission media **115** and **116**, a harvesting unit **93** and an input signal adapting unit **125**. The harvesting unit **93** has at least one harvesting circuit for receiving the RF output of the antenna unit. The harvesting unit **93** each harvesting circuit thereof, similar to the above-described examples, includes a rectifying unit **140**, and a charge collection arrangement **150** including or connected to an energy storage capacitor unit **170**. The system **103** is similar to the above-described unit **102** (FIG. 1B) in that input signal adapting unit **125** includes an adaptive impedance matching unit **122** and a control utility constituted by a control unit **180**. The adaptive impedance matching unit **122** is configured and controlled by the control unit **180** to adjust the impedance of the antenna unit **110** to the impedance of the harvesting unit **93** (rectifying unit) to thereby allow optimal delivery of the received signal, e.g. by controlling the adaptation of the impedance according to the strength of the signal received. The harvesting unit **93** is similar to the above-described unit **90** (FIG. 1A) in that its charge collection arrangement **150** is configured as a voltage to current convertor. Thus, the received antenna signal is delivered to a rectifier in the rectifying unit **140** that is adjusted to convert the RF signal into a DC signal. The output of each rectifier is connected to the voltage to current unit **152** which allows summing several DC sources into one storage unit **170** configured to collect the entire harvested energy. As indicated above, the storage unit may further be connected to a rechargeable battery or to an electrical device (not shown).

[0058] FIG. 1D illustrates yet another example of an energy harvesting system **104** utilizing a semi-active mode of operation of the major components of the system. The system **103** is configured generally similar to the above described systems **100**, **102** and **103** comprising an antenna unit **110** preferably including two conductors **112** and **113** positioned in different transmission media **115** and **116**, a harvesting unit **94** having at least one harvesting circuit for receiving the RF output of the antenna unit, and an input signal adapting unit **125**. The harvesting unit **94** (each harvesting circuit thereof), similar to the above-described examples, includes a rectifying unit **140**, and a charge collection arrangement **150** including or connected to an energy storage capacitor unit **170**. The system **104** is similar to the above-described unit **100** (FIG. 1A) in that its input signal adapting unit **125** includes an impedance matching unit **120** and a filtering unit **130** that are configured to adjust the impedance of antenna unit **110** to the impedance of rectifying unit **140** to thereby allow optimal delivery of the received signal. The harvesting unit **94** is similar to the above-described unit **92** (FIG. 1B) in that its charge collection arrangement **150** includes a switching topology unit **190** and a control utility constituted by a control unit **180** and is configured and operable to allow summing several DC sources into one storage unit **170**.

[0059] The received signal is delivered to a rectifier comprised in rectifying unit **140** that is adjusted to convert the RF signal into a DC signal. In accordance with the variation illustrated in this figure, the output of each rectifier of rectifying unit **140** is functionally connected to the switching topology circuit in switching topology unit **190** that allows summing several DC sources into one storage capacitor unit **170**. The switches position for charging/discharging configuration is determined by control unit **180**.

[0060] It should be noted that the principles of the present invention are not limited to any specific configuration of any one of the above mentioned units and any known suitable configuration can be used in the system of the invention as any men skilled in the art may configure, provided these units have the above described functional features for implementing the technique of the invention. In addition, the harvesting system provided herein may comprise off the shelf antenna, while preferably at least two antenna elements are positioned in different substance (transmission media). As described above, such antenna unit of the present invention provides maximal gain enhancement due to different transmission media surrounding the conductive elements that functionally create an antenna effect that allows uptake of electromagnetic radiation from the surroundings.

[0061] The conductive elements are inserted or attached to at least two different transmission media with different electrical and/or magnetic properties (e.g. permeability, permittivity, and conductivity) in order to increase the antenna maximal gain to functionally allow better performance of the antenna. In some applications of the invention, at least one conductive element of such novel antenna unit is positioned in contact with a user's body. The different transmission media can thus be, but not limited to, skin and air if the antenna is attached to a body or it can be, but not limited to, two different tissue types such as muscle and fat, muscle and intracellular fluid, muscle and fat tissue, dermis and fat tissue, and any other combination thereof, when the antenna is implanted in a body. The novel antenna may also be applicable for uses outside animate bodies, as long as the conductive elements are positioned in at least two different substances that differ from each other in their electrical and/or magnetic properties.

[0062] The general theory of electromagnetic phenomena is based on Maxwell's equations, which constitute a set of four coupled first-order partial-differential equations resulting the space and time changes of electric and magnetic fields to their scalar source densities (divergence) and vector source densities (curl). For stationary media, Maxwell's equations in the differential form are:

$$\nabla\cdot D(r,t)=\rho(r,t)\text{(Gauss's electric law)} \tag{1}$$

$$\nabla\cdot B(r,t)=0\text{(Gauss's magnetic law)} \tag{2}$$

$$\nabla\times E(r,t)=-\partial B(r,t)/\partial t\text{(Faraday's law)} \tag{3}$$

$$\nabla\times H(r,t)=\partial D(r,t)/\partial t+J(r,t)\text{(Ampere's law)} \tag{4}$$

wherein E is the electric field intensity [volt/meter]; H is the magnetic field intensity [Ampere/meter]; D is the electric flux density [Coulomb/square meter]; B is the magnetic flux density [Tesla or Weber/square meter]; ρ is the free electric charge density [Coulomb/cubic meter]; and J is the free electric current density [Ampere/square meter].

[0063] Maxwell's equations involve only macroscopic electromagnetic fields and, explicitly, only macroscopic densities of free charge p(r,t) giving rise to the free current density J(r,t). The effect of the macroscopic charges and current densities bounded to the medium's molecules is indicated by auxiliary magnitudes D and H, which are related to the electric and magnetic fields E and H by the so-called constitutive equations that describe the behavior of the medium. In general, the quantities in these equations are functions of the position (r) and the time (t).

[0064] The power density that the electromagnetic field carries in the free space is defined by Poynting's vector:

$$S=E\times H\text{[Watt/square meter]} \tag{5}$$

which represents the power passing through a unit area perpendicular to the propagation vector r. Inside matter, the power density S_v [Watt/cubic meter] is related to the work (or heat) supplied to the charge distribution:

$$Sv=dS/dv=E\cdot J \tag{6}$$

known as the point form of Joule's law.

[0065] Inside matter, additional physical quantities are introduced to describe the interaction of waves with molecules. The electric permittivity ∈ is connected to the electric field and to the electric flux, but also to a new macroscopic vector P [Coulomb/square meter] called the electric polarization vector, such as:

$$D=\in E+P \tag{7}$$

[0066] For most materials, being linear isotropic media, especially the human body tissues, the macroscopic vector P can be considered as co-linear and perpendicular to the applied electric field. Thus, we get:

$$P=\in\chi_e E \tag{8}$$

wherein χ_e is the electric susceptibility of the matter (i.e. capability of the matter to be polarized) and D can be written as:

$$D=\in_o\in_r E \tag{9}$$

[0067] The magnetic behavior in matter involves a similar treatment with the following relations:

$$H=B/\mu-M \tag{10}$$

$$M=\chi_m H \tag{11}$$

$$B=\mu_o\mu_r H \tag{12}$$

wherein μ is the magnetic permeability and χ_m is the magnetic susceptibility. Since the human body has very minor magnetic effects we neglect here the magnetic nature of the tissues and consider only their electric nature. Very often the relation between the electric field E and the current density J is given at any specific point by Ohm's law:

$$J(r,t)=\sigma E(r,t) \tag{13}$$

wherein σ [mho/meter] is the conductivity of the matter.

[0068] Now the configuration of electromagnetic waves is considered in the boundaries between air ("transmission medium A") and matter ("transmission medium B"). In free space, the existing fields are pure electric field E and magnetic field H and the power density obeys Poynting's law with $\in_r=1$ and $\mu_r=1$. In the human body, the fields are presented by the electric flux density D and magnetic field intensity H (because in case of no magnetic effects B=H), and the local parameters are electric permittivity c (actually a complex number written as $(\in_o(\in'+j\in''))$ and conductivity of the matter σ.

[0069] When a plane wave propagating in free space interacts with the human body, part of the energy of the wave is reflected back to air and part of the energy penetrates into the body. From the penetrating part, some portion continues to propagate inside the body and some portion is absorbed by the tissues and bones. The absorbed power can be summed as the integral of many local interactions:

$$S\text{(absorbed)}=0.5\int Pv\,dv=0.5\int(E\cdot J)dv=0.5\int\sigma|E|^2dv \tag{14}$$

[0070] The factor of 0.5 is indicative of that the dissipation is related to average power and not to the peak power. The absorbed power in the human body is often expressed by the term SAR (Specific Absorption Rate) which is no other than:

$$\text{SAR}=(\sigma/2\rho)|E|^2 \tag{15}$$

wherein σ is the conductivity [mho/meter] in any specified area and p is the mass density [kg/cubic meter] in any specified area. The units of the SAR quantity are therefore [Watt/kg]. SAR is the most commonly used indicator and measure for safety standards in RF exposure.

[0071] The next table exemplifies the tissues dielectric constant and conductivity for several tissues at 900 MHz:

Tissue	ϵ_r	σ (S/m)
Skin	35	0.6
Muscle	58	1.4
Fat	6	0.08
Spinal cord	49	1.1
Blood	62	1.5
Cartilage	35	0.6

[0072] The novel antenna unit described above preferably utilizes miniature antenna elements. For example, the maximal size of the conducting element of the antenna is significantly smaller than the wavelength for which the antenna is designed. Thus, it reduces substantially the form factor of the antenna and therefore, reduces the burden of carrying it. Moreover, the miniature size of the conductive elements allows it to be transplanted in an animate body in general and in a human body in particular.

[0073] The conductive elements may be of any suitable shape, such as but not limited to a ring shape (e.g. conductive pad) or of a line shapes (e.g. conductive wire).

[0074] In some applications of the invention, the antenna unit is used with a medical device. The antenna receives external electromagnetic radiation from the surroundings and produces a corresponding output electric signal for operating the medical device.

[0075] It should be noted that although in the above described examples the novel antenna unit of the present invention is shown as being used in electromagnetic harvesting systems to provide an improved intake of energy from the surroundings, this antenna unit may be utilized in communication systems for transmitting and receiving data, as well as medical applications for communication, for example, for data/signal exchange between an implanted medical device and an external element/system.

[0076] In a specific but not limiting example of the invention, the novel antenna unit may be implanted in and/or attached to a human body being used along with a medical device such as a pacemaker, a cardiac defibrillator, a hearing aid, an orthopedic electrical pulse producing device, and any other electric device that may be implanted in or attached to an animate body.

[0077] In this connection, reference is made to FIG. **2** that schematically illustrates the use of the system of the invention with an implanted medical device, where the antenna unit **210** is connected to an internal organ in the human body **200**. In this example, the antenna unit **210** is implanted in a heart **201** of a human body **200** and a harvesting unit **90** is associated with (integral in or connected to) a defibrillator **270**. Conductive elements **212** and **213** of the antenna unit **210** are imbedded in two different transmission media **215** and **216** respectively, which may, for example, be a fat layer surrounding the heart and a heart muscle. Due to the difference in the electrical properties of these two media **215** and **216**, they differently interact with RF radiation in the surrounding and consequently different electric potentials are obtained in the substances and accordingly on the conductors **212** and **213** thereby creating an antenna effect and transmitting the electromagnetic radiation from the surrounding to the harvesting unit **290**. Although in the illustration in the figure an input signal adapting unit is shown as a separate unit connected to the harvesting unit **290**, it should be understood that such an input signal adapting circuit may be integral with the harvesting unit as shown in the non limiting examples of FIGS. **1**A-**1**D. Also, it should be noted that although in this example, the input signal adapting unit is referred to as an impedance matching unit **220**, the input signal adapting unit, as well as the entire harvesting unit (receiver) **290** may be configured and operable according to any of the above-described examples of FIGS. **1**A-**1**D. Alternatively, as mentioned above, the system may utilize the novel antenna unit of the invention connected to any known suitable energy harvesting device having any known suitable impedance matching circuit. Thus, the antenna unit **210** is connected via the harvesting unit **290** to the defibrillator **270**. It should be clear that all circuits and devices described in this figure are drawn schematically to emphasize the overall configuration.

[0078] Reference is made to FIG. **3** that schematically illustrates how the present invention can be used in another medical application. Here, the system of the invention is associated with a hearing aid **380**. As shown in the figure, an antenna unit **310** is attached to an ear **300** such that a conductive element **312** of the antenna unit is held in air (constituting a transmission medium **315**), while a conductive element **313** is by its one end attached to ear **300** (constituting a different transmission medium **316**). The conductive elements **312** and **313** by their free ends are connected to a harvesting unit **390**; in the present not limiting example, the connection is via a separate input signal adapting unit, which has an impedance matching circuit **320**. The output of harvesting unit **390** (e.g. its storage unit) that is connected to the hearing aid **380**. All circuits and devices are drawn schematically to emphasize the overall configuration. As mentioned above, it should be understood that the harvesting unit together with impedance matching circuit may have any known suitable configuration, but preferably is configured as exemplified above with reference to FIGS. **1**A-**1**D.

[0079] Reference is now made to FIGS. **4**A and **4**B that illustrate two examples of antenna systems of the present invention denoted **400** and **400'** respectively. The systems are configured as a two-element receiving antenna, in which the antenna unit **410** includes conductive elements **412** and **413** embedded in different transmission media **415** and **416** respectively. In these specific non-limiting examples, the conductive element **412** has a closed-loop configuration or is a ring-like element (e.g. termed "conductive pad"), while conductive element **413** is a substantially straight strip-like element (e.g. termed "conductive wire"). It should be understood that the conductive elements may have any shape/geometry, e.g. both may be in the form of conductive pads or conductive wires.

[0080] In the example of FIG. **4**A, conductive elements **412** and **413** are connected via an impedance matching unit/circuit **420** to a harvesting unit **490** (e.g. configured according to the invention as described above, or any other energy collection unit) the output of which may be further connected to an electrical load (not shown). In the example of FIG. **4**B, the conductive elements **412** and **413** are connected through an impedance matching unit/circuit **420** to a transmitting unit **460** configured for transmission of data.

[0081] The impedance matching unit **420** may include an adaptive impedance matching circuit as exemplified with reference to FIGS. **1**B and **1**C, or alternatively may include an impedance matching circuit connected to a filtering unit as exemplified in general above with reference to FIGS. **1**A and **1**D and will now be described more specifically with reference to FIG. **5**.

[0082] The present invention, in its another aspect, provides a novel input signal adapting unit configured for maximizing the efficiency of the energy collection system. To this end, the input signal adapting unit is configured as a dedicated passive impedance matching filtering circuit formed by an impedance matching unit/module and a filtering unit/module. The configuration and operation of the filtering unit is based on the fact that in an RF harvesting system from the environment, the frequency and intensity (strength) of most harvested signals can be predefined, i.e. are either known a priori or are predictable, and therefore by using frequency filtering one can actually differentiate/sort the received signal for different intensities (strengths).

[0083] For example, there are typically two major categories of signals expected in be input signals for an RF harvesting system from the environment, both originated from the mobile phone RF system. The first is the signal from the base station to the cellular phone (downlink) and the second is the signal from the cellular phone to the base station (uplink) The downlink signal is characterized by a relatively low level signal and a different frequency than the uplink signal having a higher intensity. RF radio signals are also expected to be with the same frequency and intensity characteristics of as the downlink signals.

[0084] The filtering unit may include a preliminary wide band pass filter circuit with a wide enough frequency band to include both uplink and downlink frequency bands, and two or more narrow band filter circuits (defining at least one pair of such narrow band filters), one or more of them configured for the uplink frequency band(s) and one or more other of them configured for the downlink frequency band(s). The first wide band filter is responsible for receiving RF signal from the harvesting antenna and allowing passage of signals of both uplink and downlink frequencies. Each narrow band filters is responsible for allowing passage of AC signals in the appropriate frequency band, and accordingly in the appropriate corresponding intensity, to a dedicated rectifier so as to match the impedance of the rectifier according to the expected input intensity. The narrow band filters may be operable in parallel so as to simultaneously allow passage of AC signals in different narrow frequency bands.

[0085] It is understood that in different geographic locations, the expected frequency range and expected intensity may be different, and therefore the selection of the wide band filter and the narrow band filters may vary from country to country and from region to region and/or the filter circuit may include more than one pair (e.g. a large number of pairs) of filters to cover multiple combinations of the wide band filter and the narrow band filters.

[0086] It is understood that change in mobile (cellular) technology and/or mobile (cellular) transmission protocols may define various frequency bands that will be available for harvesting from the environment. However, due to ability of predicting their frequency and intensity, a dedicated narrow band filter and a corresponding rectifier can be designed.

[0087] It should be understood that the cellular uplink and downlink RF signals discussed above are only examples of available RF signals in the air, and other signals are within the scope of the present invention. The above aspect of the invention is exemplified in FIG. **5** which is a schematic block diagram illustration of a filtering unit **500** of the invention for maximizing the efficacy in an RF harvesting system. The filtering unit **500** includes a wide band filter **510** having an input that may receive an alternating current radiofrequency signal from a harvesting system antenna (not shown), and an output which is connected to narrow band filters—two such filters **521** and **522** being shown in the example of FIG. **5**. Narrow band filter **521** is designed to allow passage of signals in the uplink frequency range of the cellular system and is connected in parallel to the narrow band filter **522**, which is designed to allow passage of signals in the downlink frequency range of the cellular system. Wide band filter **510** is configured for passing both of the uplink and the downlink frequency bands.

[0088] Narrow band filter **521** is connected to a dedicated rectifier **531** designed for receiving signals of an expected strength, and narrow band filter **522** is connected to another dedicated rectifier **532** that is similarly designed for optimally obtaining signals of the expected frequency and strength from filter **522**. The impedance of each of the rectifiers **531** and **532** is matched according to the expected intensity parameters (strength) in the specific frequency band, originated from the RF system.

[0089] For realization of the exemplary filtering unit described above, a numeric example is herein provided: A radiation band having a central frequency of 1 GHz and bandwidth (BW) of 800 MHz is received by the antenna. This band contains desired frequencies that need to be harvest. The frequencies in accordance with this specific example are 915 MHz and 950 MHz. The received band is transferred from the antenna to the first wide band filter that allows transmission of a 60 MHz band centered at 930 MHz. The received signals obtained from the first filter are thus between the frequencies of 900 MHz to 960 MHz. The parallel second and third narrow band filters are now receiving signals between 900 MHz to 960 MHz, while the second filter is set to center frequency 915 MHz with BW of 30 MHz and the third filter is set to center frequency of 945 MHz with BW of 30 MHz. The received signals' strength in the band of 915 MHz are predicted to be in the range of (−30) dbm to (−20) dbm, and the received signals strength in the band of 950 MHz are predicted to be in the range of (−5) dbm to 5 dbm.

[0090] The rectifier unit connected to the filtering unit includes a corresponding number of rectifiers, the rectifier connected to the second band pass filter is matched to rectify signal strength in the range between (−30) dbm to (−20) dbm and the rectifier connected to the third band pass filter is matched to rectify signal strength in the range between (−5) dbm to 5 dbm.

[0091] It should be understood that the invention is neither limited to a number of filters in the filtering unit described above, nor to any specific frequency bands. Additional signals propagating in the surroundings of the antenna and suitable to be used by such filtering circuit, include those having the priori known frequency and strength of a commercial radio channel. The expected signal strength received from a commercial radio channel is from (−50) dbm to −40 dbm. In this case, the filter may be set for the channel frequency and the corresponding rectifier may be matched to a strength range of (−50) to (−40) dbm. In case of parallel radio channels, for example, between 88 MHz and 108 MHz, the receiver (e.g.

harvesting unit) can be matched to the strength of the entire band received from the band pass filter.

[0092] The present invention in its yet further aspect provides a novel charge collecting unit suitable to be used in the energy harvesting system. The charge collecting unit of the present invention is configured as a signal summing utility adapted for accumulating charge from relatively small capacitor units to a large capacitor unit.

[0093] In some embodiments of the invention, such signal summing utility is configured to appropriately accumulate charge mainly by changing the topology of various components of the unit. In this case, the summing utility may include such functional modules/circuits as a relatively small storage charging circuit, and a switching circuit, connected to a relatively large storage charging circuit (which may be that of the storage unit in the harvesting system).

[0094] The small storage charging circuit includes a plurality of capacitor units. The small storage charging circuit may for example be adapted for accumulating charge obtained from the output of the rectifying unit (i.e. plurality of rectifiers) and for storing the charge created on multiple capacitor units. Each of the capacitor units in the small storage charging circuit may include a plurality of capacitors being in serial and/or parallel connections.

[0095] The switching circuit may be adapted for determining the mode of the circuit, defining the function thereof as a charging function or as a discharging function. The discharging of the small storage charging circuit may be performed simultaneous or single charging unit at a time, or according to predetermined time pattern for multiple storage charging units. The plurality of switches in the switching circuit may be operated independently, as well as may be connected and switched (operated) according to a predetermined switching pattern, for example a certain switch is activated depending on the operative state of the locally adjacent switches. The plurality of switches in the switching circuit may be controllably operated as exemplified above with reference to FIGS. **1**B and **1**D.

[0096] The charge collection unit of the present invention may be designed to be associated with a plurality of harvesting units. Each harvesting unit may capture RF energy, rectify and store the energy as a charge on a capacitor unit. In some embodiments of the present invention, the plurality of harvesting units may capture plurality of frequency bands, wherein each harvesting unit may be adapted to capture a specific frequency band. Additionally or alternatively, at least two harvesting units may be adapted to capture similar frequency band.

[0097] The large storage charging circuit may be adapted for accumulating and storing the charge pre-stored in each of the plurality of small capacitor units in a large storage charging unit for further usage. The large storage charging circuit may include a plurality of capacitors in serial and/or parallel connections. The output of the large storage charging circuit may for example be followed by one or more of the following: a step up convertor; a voltage stabilizer; a battery charging circuit; a capacitor array, an electrical load. In some embodiments, the output of the large storage charging subunit may be either switched or constantly short to the circuit.

[0098] Some examples of various switching options are provided below with reference to FIGS. **6-10**.

[0099] FIG. **6** is a schematic block diagram illustrating a standard charge collection circuit **600** that might be used in the energy harvesting system of the invention. The collection circuit **600** includes a plurality of charging capacitor units—three such units **621**, **622**, and **623** being shown in this non-limiting example. In this example, the charge collection circuit **600** operates for collecting charge from a rectifying unit including a plurality of rectifier elements—three such elements **611**, **612** and **613** being shown in this non-limiting example. Each charging capacitor unit is connected to a corresponding one of the rectifier elements and connected to ground. Each of the rectifiers **611**, **612** and **613** receives AC signal from a harvesting antenna unit or from any other element positioned between the antenna unit and the rectifier (e.g. input signal adapting unit, e.g. impedance adaptor), rectifies the signal, and the corresponding energy is stored as electric charge in the corresponding one of the capacitor unit **621**, **622**, and **623**. It is understood that each capacitor unit may be configured as a single capacitor unit or multi-capacitor unit where the multiple capacitors may be connected in any serial and/or parallel connections. It is understood also that each rectifier element may receive AC signal in a similar or different frequency band. It should be noted that the circuit **600** is neither limited to a specific number of rectifier-capacitor arrangements nor to a specific electrical connection of the capacitors, and FIG. **6** is just an example of the operational principles of the charge collection unit.

[0100] Reference is now made to FIG. **7** which is a schematic block diagram illustration of a charge collection unit **700** configured and operable according to the present invention. The charge collection unit **700** is operable as a signal summing unit, and includes a plurality of small capacitor units—thee such units **721**, **722** and **723** being shown in this non-limiting example. The capacitor units **721**, **722** and **723** are at one ends connected via switching circuits to rectifiers **711**, **712** and **713** respectively, and at the other ends are connected to ground **750**. Also provided in the charge collection unit **700** is a large capacitor unit **740** which is connected to the capacitor units **721**, **722** and **723** via the switching circuits. The switching circuits include a plurality of switches **731**, **732**, **733**, **734** and **735**, each configured with two working/operative positions **1** and **2**. The distribution of switches positions between positions **1** and **2** determine the functionality of the switching circuit, as will be described with reference to FIGS. **8** and **9**. As shown in the figure, the switches **731**, **733** and **735** selectively connects the charge collection unit **700** to the rectifying unit.

[0101] Thus, each rectifier receives AC signal from a harvesting antenna or from any other intermediate circuit communicating the antenna signal to the rectifier. The energy rectified by each of the multiple rectifiers of the rectifying unit is being stored as electric charge in the corresponding capacitor unit of the charge collection unit **700**. As shown in the figure in a self-explanatory manner, each capacitor unit is on its other end selectively connectable either to ground **750** or to the large capacitor unit **740** through the corresponding two-state switches, and on its other end is selectively connectable either to the respective rectifier or to the large capacitor unit **740** through the other two-state switches.

[0102] The charge collection unit operates in the following manner. The switches **731-735** are controlled by the controller, which can turn them to any given mode (charging/discharging) based on real time monitoring of each of the capacitors **721-723** connected to the outputs of the rectifiers. Based on the information (voltage level) received from each capacitor, the controller can determined as to whether to keep the circuit in the charging mode or whether it is ready to discharge into the storage capacitor. The received information from the capacitors **721**-**723** can also be indicative of whether the harvesting process is ongoing, and whether the reached voltage level is maximal at the time. Based on this information, the controller turns the switches into charging or discharging mode.

[0103] Further, the controller can define a full discharge mode (when all the capacitors are discharged together), semi discharge mode (only some of the capacitors are discharged while the others remain in charging mode), and charging mode (all capacitors are being charged).

[0104] Another option for the controller operation is based on a programmable pattern (i.e. predefined values). When the voltage of the capacitor **721**-**723** reaches a certain voltage level (as being defined), the switches are turned to the discharge mode. When the discharge is complete (can be defined as well), the switches are being turned back into the charging mode. This method allows three modes: full discharging, semi discharging and full charging.

[0105] It is understood that each of capacitor units **721**, **722**, and **723** may be configured as a single capacitor unit or as a multi-capacitor unit in which an array of capacitors are connected in some combination of serial and parallel connections. Similarly, that the large capacitor unit **740** may have a single-capacitor configuration or may be a multi-capacitor circuit utilizing a proper serial and/or parallel connection between the capacitors.

[0106] It should be understood that the configuration exemplified herein is neither limited to a number of rectifiers in the rectifying unit nor to a number of small capacity units in the charge collection unit **700**. It should also be clear that all circuits and devices described in this figure are drawn schematically to emphasize the overall configuration.

[0107] FIGS. **8** and **9** schematically illustrate another example of a charge collection or signal summing unit **800** of the invention. The charge collection unit **800** is configured generally similar to the above described unit **700**, namely includes multiple small capacitor units denoted as **821**, **822** and **823** connected to multiple rectifiers **811**, **812**, **813**, respectively, of a rectifying unit, and connectable to a large capacitor unit **840** in a charging configuration via switching circuit including multiple switches **831**, **832**, **833**, **934** and **835**. The capacitor units **821**, **822** and **823** are at one ends connected to the rectifiers, and at the other ends are connected to ground **850**.

[0108] As mentioned above, the small capacitor units **821**, **822**, and **823** as well as large capacitor unit **840** may include single-capacitor configuration(s) or multi-capacitor configuration(s) connected to one another in serial and/or parallel connections. Also, the invention is not limited to a number of rectifiers in the rectifying unit; the number of small capacity units is selected in accordance with the number of the rectifiers. Further, as also mentioned above, all circuits and devices described in this figure are drawn schematically to emphasize the overall configuration.

[0109] The configuration **800** is different from the above-described example of FIG. **7** in a somewhat different configuration of switching elements **831**, **832**, **833**, **834** and **835**. Each of the switches **831**, **833**, **835** selectively connects the respective capacitor unit to the corresponding rectifier or to the large capacitor unit **840**, and the other switches selectively connect the respective capacitor units to ground **850** or to the large capacitor unit **840**.

[0110] FIG. **8** shows the switching configuration of the summing unit **800** corresponding to a charging configuration. Each capacitor unit is connected directly to its corresponding rectifier and is being charged. It is understood that capacitor units **821**, **822** and **823** may be charged in different charges.

[0111] FIG. **9** shows the signal summing unit **800** of FIG. **8** in a discharging configuration. As shown, the switching elements **831**, **832**, **833**, **834** and **835** are in a switch position topology connecting each of small capacitor units **821**, **822** and **823** to the large capacitor unit **840**. In this discharging configuration all small capacitor units **821**, **822** and **823** are connected in series and their resultant capacitor is connected in parallel to the large capacitor unit **840**, i.e. electrical charge is being transferred from small capacitor units **821**, **822**, and **823** to large capacitor unit **840**. The change of the switching state from the topology described in FIG. **8** to the topology described in FIG. **9** results in charge transfer from small capacitor units **821**, **822** and **823** to large capacitor unit **840**. Since the voltage across the small capacitor units is higher than the voltage across the large capacitor unit **840**, the flow of charge is directed from small capacitor units to large capacitor unit until equilibrium is reached.

[0112] FIG. **10** exemplifies a signal summing unit **800** of FIGS. **8**-**9** modified to have a single discharging configuration. In the scenario illustrated in this figure, switching elements **831**, **832**, and **835** are in a switch position topology connecting small capacitor units **821**, and **823** to rectifiers **811** and **813** respectively on one side, and to ground **850** on the other side, while being disconnected from the large capacitor unit **840**. This switching configuration is a charging configuration of capacitors **821** and **823** from rectifiers **811** and **813** respectively associated with harvesting unit/units. However, switching elements **833** and **834** are positioned in a topology that allows for connecting small capacitor unit **822** to the large capacitor unit **840** in a parallel configuration, i.e., in a discharging position. In other words, in the scenario illustrated in FIG. **10**, one or more of the capacitor units may be discharged while one or more other capacitor units remain charged: small capacitor unit **822** may be discharged to large capacitor unit **840** while the other capacitor units **821** and **823** remain charged.

[0113] In accordance with the example of FIG. **10**, switching elements **833** and **835** are configured with three optional positions, each changing the topology of the circuit and its functionality. These switching elements, when being in the first operational state/position, connect the small capacitor units **822** and **823** to respective rectifiers **812** and **813** thus allowing charging of capacitor units **822** and **823**. When in the second operational position of switching elements **833** and **835** they create serial connection of the small capacitor units **821**, **822** and **823** to thereby allow simultaneous discharge of the small capacitor units to the large capacitor unit **840**. The third operational position of these switching elements allows direct discharge of the small capacitor unit **822** or **823** to the large capacitor unit **840**. It should be understood that the small capacitor unit **821** might be capable of being discharged autonomously, directly to the large capacitor unit **840**.

[0114] As indicated above, this embodiment of the summing unit is also not limited to any specific number of small capacitor units, and this number corresponds to a number of rectifiers in a rectifying unit, and that all circuits and devices described in this figure are drawn schematically to emphasize the overall configuration.

[0115] As described above with reference to FIGS. **1**A and **1**C, in some embodiments of the invention the charge collection unit or signal summing unit allows accumulation of charge in the harvesting system by converting the electric voltage developed on capacitors, connected to the outputs of rectifiers, into electric current source. In such variation, the summing unit includes a small storage charging utility/circuit, a voltage to current converter utility/circuit; and a large storage charging utility/circuit.

[0116] The small storage charging circuit might be mainly adapted for storing the charge obtained via the plurality of rectifiers by using a plurality of capacitor units as described above. The voltage to current converter circuit may be configured for converting, upon reaching a threshold, the voltage, developed in the small storage charging utility, into current, to be delivered to a large storage charging utility. The large storage charging utility may be configured for accumulating and storing the current arriving from voltage to current conversion in the form of charge in a large storage charging unit for further usage.

[0117] In some embodiments of the present invention, the small storage charging circuit may be designed as a plurality of harvesting units. Each harvesting unit may capture RF energy, rectify and store the energy as a charge on a capacitor unit.

[0118] The plurality of harvesting units may capture plurality of frequency bands, wherein each harvesting unit may be adapted to capture a specific frequency band, or at least two harvesting units may be adapted to capture similar frequency bands different from that/those of other harvesting units.

[0119] The plurality of harvesting circuits may have plurality of capacitor values. In some embodiments, each capacitor unit in the plurality of capacitor units in the small storage charging circuit may include a plurality of capacitors in serial and/or parallel connections (capacitors array).

[0120] The discharging of the small storage charging units may be performed by converting the voltage developed on each of the capacitor units in the plurality of capacitor units to an electric current, by using a voltage to current converter units. The multiple voltage to current converter units of the voltage to current converter circuit may be operable independently; and/or may be operable by a controller that controls all voltage to current converter units or parts thereof. The voltage to current converter units may have voltage thresholds, defining the minimal voltage triggering the conversion of voltage to current. The voltage thresholds of the voltage to current converter units may be different; as well as may be changed by the controller.

[0121] Multiple voltage to current converter units may operate substantially simultaneously, as each unit operates independently and do not interfere to the current converted by other units.

[0122] Each voltage to current converter unit may restrict current flow backwards from the large storage charging circuit to the small storage charging circuit, as the voltage to current converter has the ability to force flow of charge in a certain direction.

[0123] The large storage charging circuit may include a plurality of capacitors in serial and/or parallel connections. In some embodiments, the output of the large storage charging circuit may be connected to one or more of the following: a step up converter; a voltage stabilizer; a battery charging circuit; a capacitor array, and an electric load. In some embodiments, the output of the large storage charging circuit may be either switched or constantly short to the circuit.

[0124] FIG. **11** schematically illustrates an example of a charge collection unit or signal summing unit **900** for accumulating charge corresponding to output of a plurality of rectifiers **911**, **912**, **913** of a rectifying utility of a harvesting system. The signal summing unit **900** includes a small capacitor circuit which in the present example is formed by small capacitor units designated as **921**, **922** and **923**; a voltage to current convertor circuit which in this example is formed by a plurality of voltage to current converters **931**, **932** and **933**; and a large capacitor unit/circuit **940**. The small capacitor units **921**, **922** and **923** are at one end connected to the rectifiers **911**, **912** and **913** respectively and to the voltage to current converters **931**, **932** and **933**, and on the other end are connected to ground **950**.

[0125] Each rectifier receives AC signal from a harvesting antenna unit (or from any other intermediate utility positioned between the antenna unit and the rectifiers), and rectifies the signal. Each of the rectifiers **911**, **912** and **913** is also connected to the respective one of the voltage to current converters **931**, **932** and **933**. The charge of each of the rectifiers **911**, **912** and **913** is being stored, up to a predetermined threshold, as an electric charge in the corresponding one of the small capacitor units **921**, **922** and **923**. Each of the voltage to current converters **931**, **932** and **933** has an intrinsic voltage threshold value that switches the voltage to current converter from a disconnection position (inoperative state) to a connection position (operative state). In the disconnection position of one of the voltage to current converters **931**, **932** and **933**, all current from the corresponding rectifier flows to the corresponding small capacitor unit. Thus, the charge accumulation in the corresponding small capacitor unit increases the voltage on the corresponding small capacitor unit. When voltage threshold is reached, the corresponding voltage to current converter operates as a unidirectional connector, forcing the charge developed in the small capacitor unit to flow directly towards the large capacitor unit **940**.

[0126] It is understood that each of small capacitor units **921**, **922**, and **923** as well as the large capacitor unit **940** may be configured as a single capacitor unit or may include an array of capacitors connected in some combination of serial and parallel connections. It should also be clear that any number of rectifiers may be used, and the number of small capacity units and the number of voltage to current converters is selected accordingly, and that all circuits and devices described in this figure are drawn schematically to emphasize the overall configuration.

1. A system for harvesting electromagnetic energy propagating in surroundings, said system comprising:

an antenna unit configured for receiving external electromagnetic radiation from the surroundings and producing a corresponding electric output;

a harvesting unit comprising at least one energy harvesting circuit each configured and operable for receiving signals indicative of the output of the antenna unit and generating and storing corresponding electric charge, the harvesting circuit comprising: a rectifying unit comprising a plurality of rectifiers each configured and operable to receive AC electric signals and generate corresponding DC electric power; and a charge collection unit configured and operable to receive the plurality of DC electric powers from said rectifying unit and converting and accumulating them into the electric charge presenting harvested energy; and

an input signal adapting circuit having an input connected to the antenna unit and an output connected to the rectifying unit, the input signal adapting circuit being configured and operable for adjusting a predetermined electrical property of the antenna unit and rectifying unit to thereby optimize receipt of the electric output of the antenna unit to the harvesting circuit.

2. A system according to claim **1**, wherein said antenna unit comprises at least two conductive elements positioned in at least two different transmission media being different in at least one of electric and magnetic properties.

3. A system according to claim **1**, wherein said at least two different transmission media include at least two different animate tissues.

4. A system according to claim **1**, wherein said at least two different transmission media include animate tissue and air.

5. A system according to claim **1**, wherein said input signal adapting unit is configured to adjust impedance of the antenna unit to impedance of the rectifying unit.

6. A system according to claim **5**, wherein said input signal adapting unit comprises an impedance matching unit and a filtering unit connected to the impedance matching unit, the impedance matching unit being configured to match between the impedances of the antenna unit and the filtering unit, and the filtering unit being configured and operable for sorting the received RF band propagating in the surrounding into a predefined number of subbands with a predetermined central frequency, band width and strength for each subband, the predefined number of subbands corresponding to a number of the rectifiers in the rectifying unit.

7. A system according to claim **5**, wherein said input signal adjusting unit comprises: an adaptive impedance matching unit configured and operable to controllably adjust impedance of the rectifiers of the rectifying unit to signal intensities of multiple frequency bands received by the antenna unit; and a control unit connectable to the adaptive impedance matching unit and configured and operable operation thereof.

8. A system according to claim **1**, wherein said charge collection unit comprises a signal summing unit comprising a switching topology circuit and a controller, the switching topology circuit comprising a predetermined number of capacitors having relatively small capacitance values and being connected to output of the corresponding number of the rectifiers respectively, and a storage capacitor of a relatively large capacitance value, the switching topology circuit being adapted to controllably switch between the plurality of DC signals received from the plurality of rectifiers respectively with the corresponding plurality of said capacitors of the relatively small capacitance values for power summation and storage of corresponding charge on said capacitor of the relatively large capacitance value.

9. A system according to claim **1**, wherein said charge collection unit comprises a signal summing unit comprising a voltage to current unit comprising a predetermined number of voltage to current conversion circuits having inputs connected to outputs of the corresponding number of the rectifiers respectively, and outputs connected to a storage a capacitor, the voltage to current unit being thereby adapted to convert and sum the plurality of DC signals received from the plurality of rectifiers via the plurality of voltage to current conversion circuits.

10. A system according to claim **1**, comprising a control unit connected to at least one of the input signal adapting unit and the charge collection unit and configured and operable to control operation of the harvesting circuit.

11. A system according to claim **10**, wherein said control unit is connected to at least said charge collection unit which comprises a switching topology circuit comprising a predetermined number of capacitors having relatively small capacitance values and being connected to output of the corresponding number of the rectifiers respectively, and a storage capacitor of a relatively large capacitance value, the switching topology circuit being adapted to controllably switch between the plurality of DC signals received from the plurality of rectifiers respectively with the corresponding plurality of said capacitors of the relatively small capacitance values for power summation and storage of corresponding charge on said capacitor of the relatively large capacitance value.

12. A system according to claim **1**, wherein said harvesting circuit comprises a connection port for connecting output of the charge collection unit to at least one electrical load.

13-16. (canceled)

17. An antenna unit configured and operable for receiving external electromagnetic radiation from surroundings and producing a corresponding output electric signal, said antenna unit comprising at least two conductive elements positioned in at least two different transmission media being different in at least one of electric and magnetic properties, such that received external electromagnetic radiation creates a potential difference between the at least two conductive elements producing an output electric signal, thereby allowing uptake of the external electromagnetic radiation from the surroundings.

18. An antenna unit according to claim **17**, wherein said at least two different transmission media include different animate tissues.

19. An antenna unit according to claim **17**, being attached to an animate body, wherein said at least two different transmission media include animate tissue and air.

20. (canceled)

21. An input signal adapting unit for use with an antenna exposed to electromagnetic radiation of Radio Frequency (RF) band propagating in surroundings, the input signal adapting unit comprising an impedance matching unit and a filtering unit connected to the impedance matching unit, the impedance matching unit being configured to match between impedances of the antenna unit and the filtering unit, and the filtering unit being adapted for sorting the RF band into a predefined number of subbands of a predetermined frequency, bandwidth and strength for each subband.

22. (canceled)

23. A charge collection unit for use in an energy harvesting circuit, the charge collection unit comprising a signal summing unit adapted to receive a plurality of DC electric powers and sum them in the form of an electric charge on a storage capacitor, the signal summing unit comprising at least one of a switching topology unit and a voltage to current unit.

24. A charge collection unit according to claim **23**, wherein the summing unit comprises the switching topology unit and a controller, the switching topology unit comprising a predetermined number of capacitors having relatively small capacitance values and being connectable to output of a corresponding number of rectifiers respectively, and connected to said storage capacitor of a relatively large capacitance value, the switching topology unit being adapted to controllably switch between DC sources received from the rectifiers with said predetermined number of the capacitors for power summation.

25. A charge collection unit according to claim **23**, wherein said voltage to current unit comprises a predetermined number of voltage to current conversion circuits having inputs connectable to outputs of a corresponding number of rectifiers respectively, and outputs connected to the storage capacitor, the voltage to current unit being thereby configured and operable to convert and sum a plurality of DC sources received from the plurality of rectifiers via the voltage to current conversion circuits.

26. An energy harvesting unit comprising at least one energy harvesting circuit, wherein each energy harvesting circuit is configured and operable for receiving signals from an antenna unit and generating and storing corresponding electric charge, the harvesting circuit comprising:

an input signal adapting circuit connectable to output of an antenna unit; a rectifying unit connected to output of the input signal adapting circuit, the rectifying unit comprising a plurality of rectifiers each configured and operable to receive AC electric signals and generate corresponding DC electric power; and a charge collection unit configured and operable to receive the plurality of DC electric powers from said rectifying unit and converting and accumulating them into the electric charge presenting harvested energy;

the harvesting circuit having at least one of the following configurations:

(a) the input signal adapting unit comprises an impedance matching unit and a filtering unit connected to the impedance matching unit, the impedance matching unit being configured to match between impedances of the antenna unit and the filtering unit, and the filtering unit being adapted for sorting the RF band into a predefined number of subbands of a predetermined frequency, bandwidth and strength for each subband; and

(b) the charge collection unit comprises a signal summing unit adapted to receive a plurality of DC electric powers and sum them in the form of an electric charge on a storage capacitor, the signal summing unit comprising at least one of a switching topology unit and a voltage to current unit.

* * * * *

US 20160094091A1

(19) **United States**

(12) **Patent Application Publication**

SHIN et al.

(10) **Pub. No.: US 2016/0094091 A1**

(43) **Pub. Date: Mar. 31, 2016**

(54) **WIDEBAND RECTENNA AND RECTIFYING APPARATUS FOR RECTENNA**

(71) Applicants: **Jin Woo SHIN**, Daejeon (KR); **Jun Ho CHOI**, Daejeon (KR); **Joon Ho SO**, Daegu (KR)

(72) Inventors: **Jin Woo SHIN**, Daejeon (KR); **Jun Ho CHOI**, Daejeon (KR); **Joon Ho SO**, Daegu (KR)

(73) Assignee: **AGENCY FOR DEFENSE DEVELOPMENT**, Daejeon (KR)

(21) Appl. No.: **14/645,479**

(22) Filed: **Mar. 12, 2015**

(30) **Foreign Application Priority Data**

Sep. 25, 2014 (KR) 10-2014-0128530

Publication Classification

(51) **Int. Cl.**

H02J 17/00	(2006.01)
H02J 5/00	(2006.01)
H01Q 9/04	(2006.01)

(52) **U.S. Cl.**

CPC ***H02J 17/00*** (2013.01); ***H01Q 9/0407*** (2013.01); ***H02J 5/005*** (2013.01)

(57) **ABSTRACT**

A rectenna according to the present invention includes a circular-polarized patch antenna having dual slots fed by a microstrip and configured to receive and output a radio frequency (RF) signal, and a rectifying circuit configured to convert for output the RF signal, received by the circular-polarized patch antenna, into a direct current (DC) signal and transfer the DC signal from the antenna to a load, wherein the rectifying circuit comprises at least one radial stub.

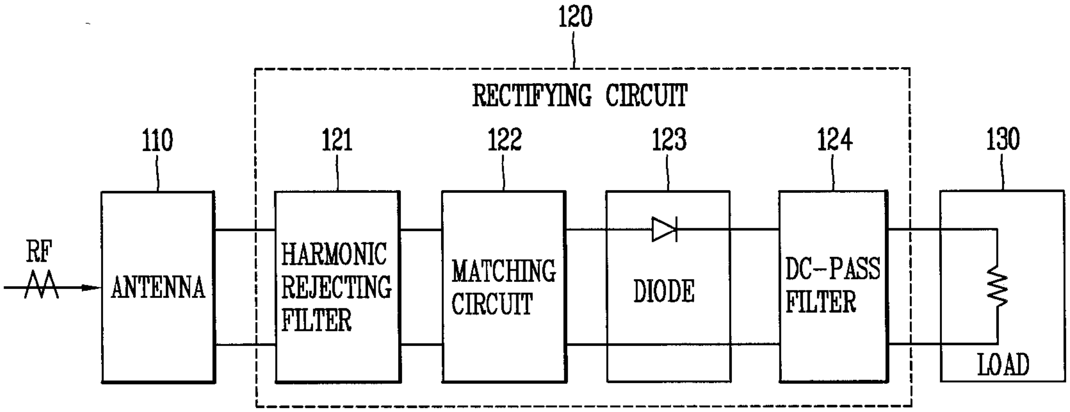
120
RECTIFYING CIRCUIT
110
121
122
123
124
130
RF
ANTENNA
HARMONIC REJECTING FILTER
MATCHING CIRCUIT
DIODE
DC-PASS FILTER
LOAD

FIG. 1

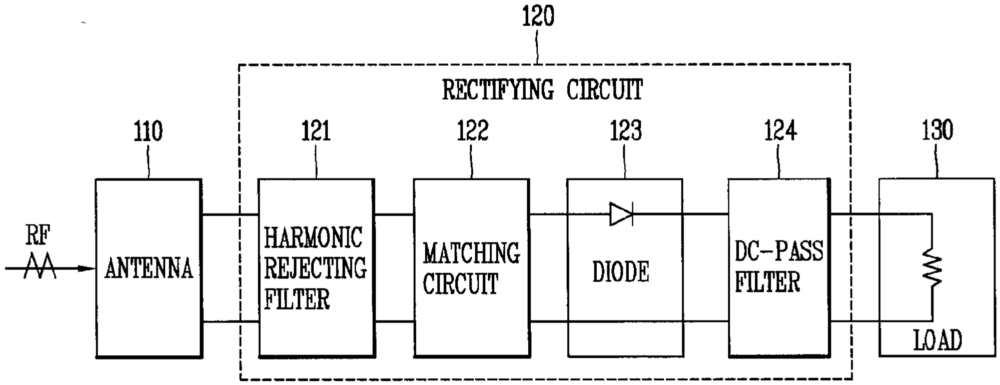
120
RECTIFYING CIRCUIT
110
121
122
123
124
130
RF
ANTENNA
HARMONIC REJECTING FILTER
MATCHING CIRCUIT
DIODE
DC-PASS FILTER
LOAD

FIG. 2A

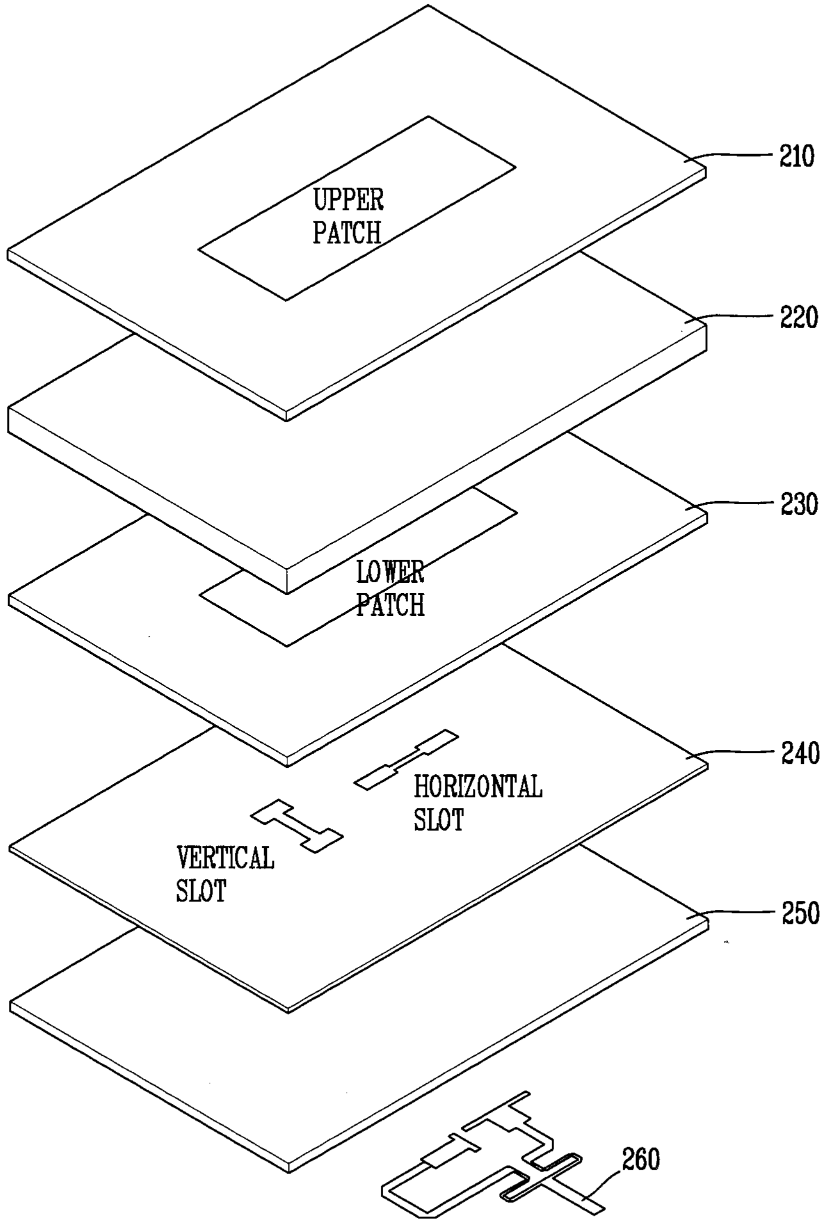
210
UPPER
PATCH
220
230
LOWER
PATCH
240
HORIZONTAL
SLOT
VERTICAL
SLOT
250
260

FIG. 2B(a)

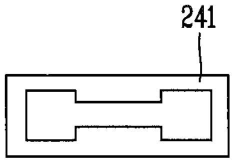

FIG. 2B(b)

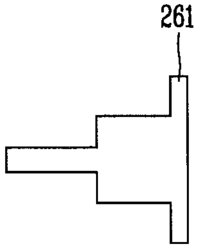

FIG. 2B(c)

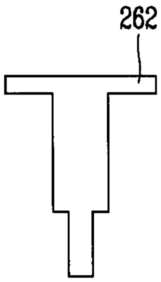

FIG. 2B(d)

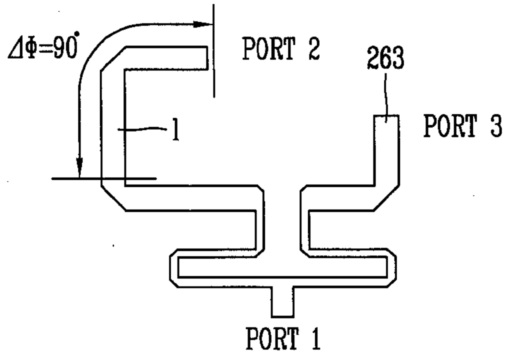
ΔΦ=90°
PORT 2
263
PORT 3
1
PORT 1

FIG. 2c

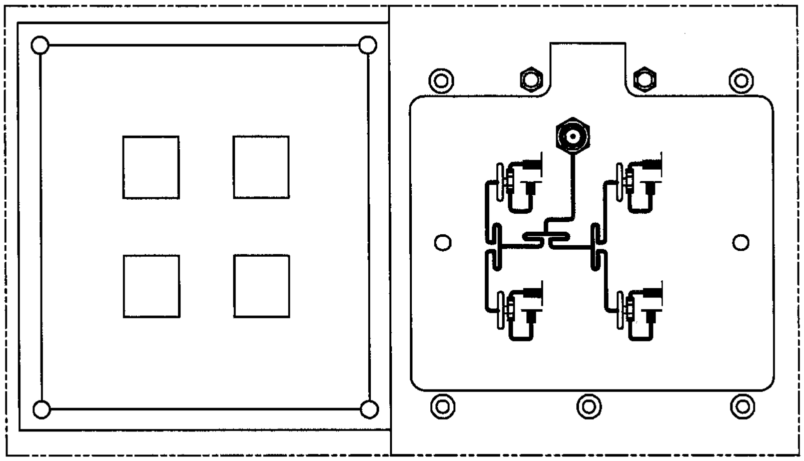

FIG. 3A

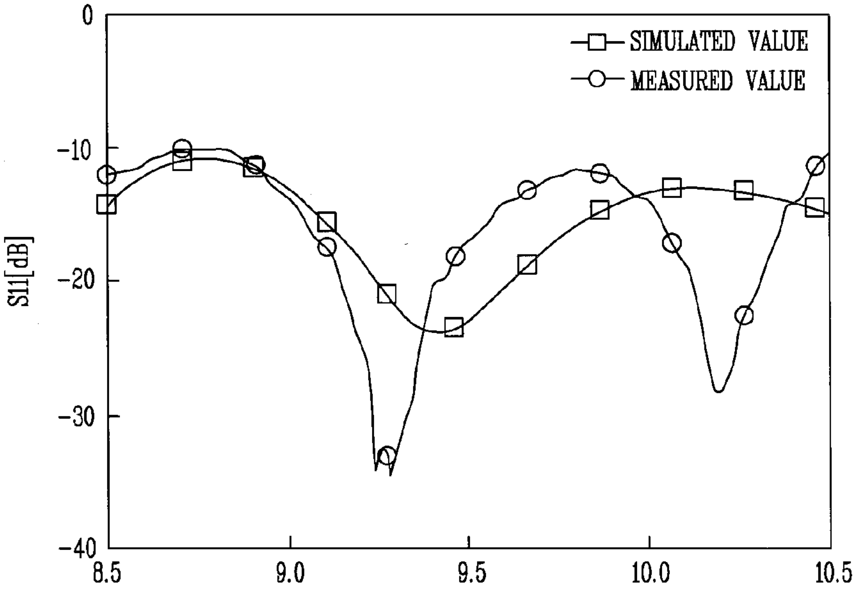
0
-10
-20
-30
-40
S11[dB]
8.5
9.0
9.5
10.0
10.5
SIMULATED VALUE
MEASURED VALUE

FIG. 3B

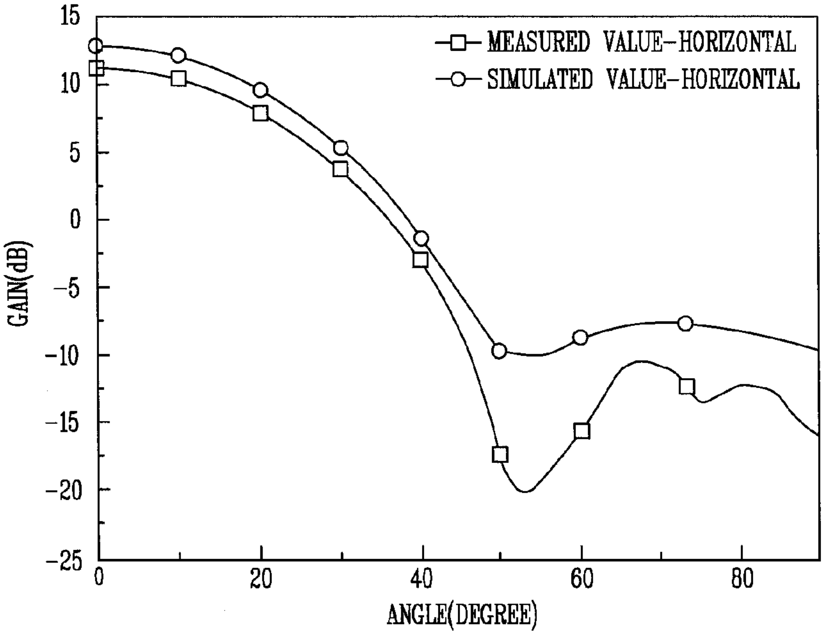
15
10
5
0
-5
-10
-15
-20
-25
GAIN(dB)
0
20
40
60
80
ANGLE(DEGREE)
MEASURED VALUE-HORIZONTAL
SIMULATED VALUE-HORIZONTAL

FIG. 4A

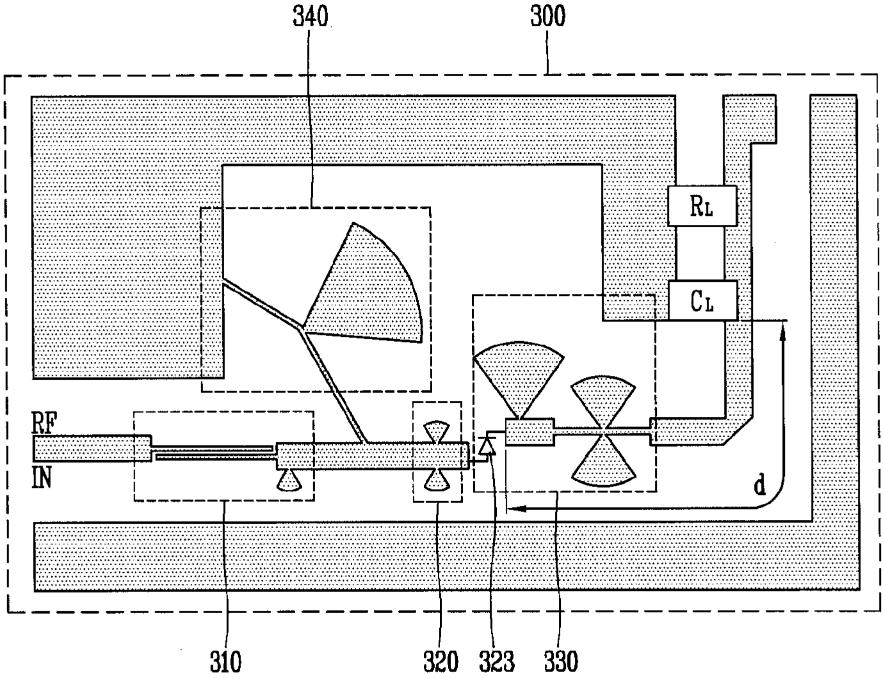
340
300
RL
CL
RF
IN
d
310
320
323
330

FIG. 4B

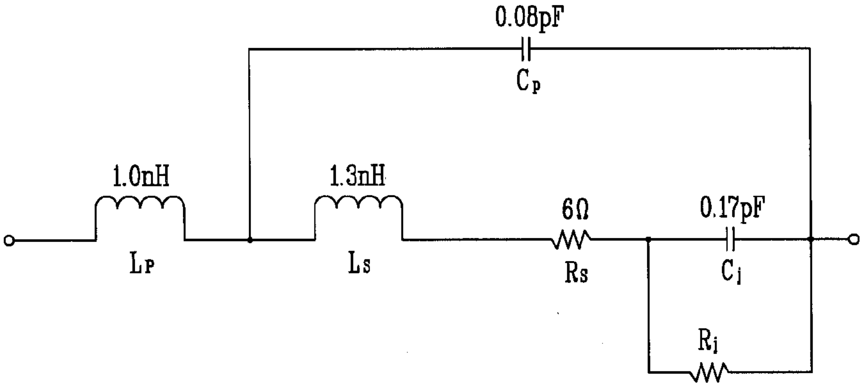
0.08pF
Cp
1.0nH
Lp
1.3nH
Ls
6Ω
Rs
0.17pF
Cj
Rj

FIG. 5A

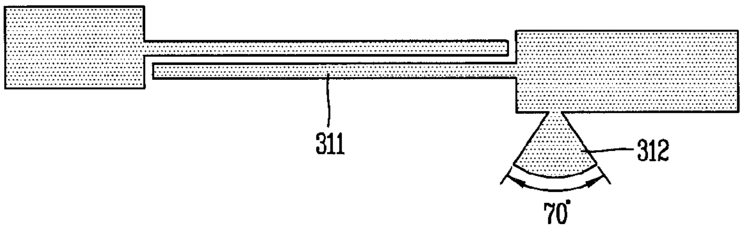
311
312
70°

FIG. 5B

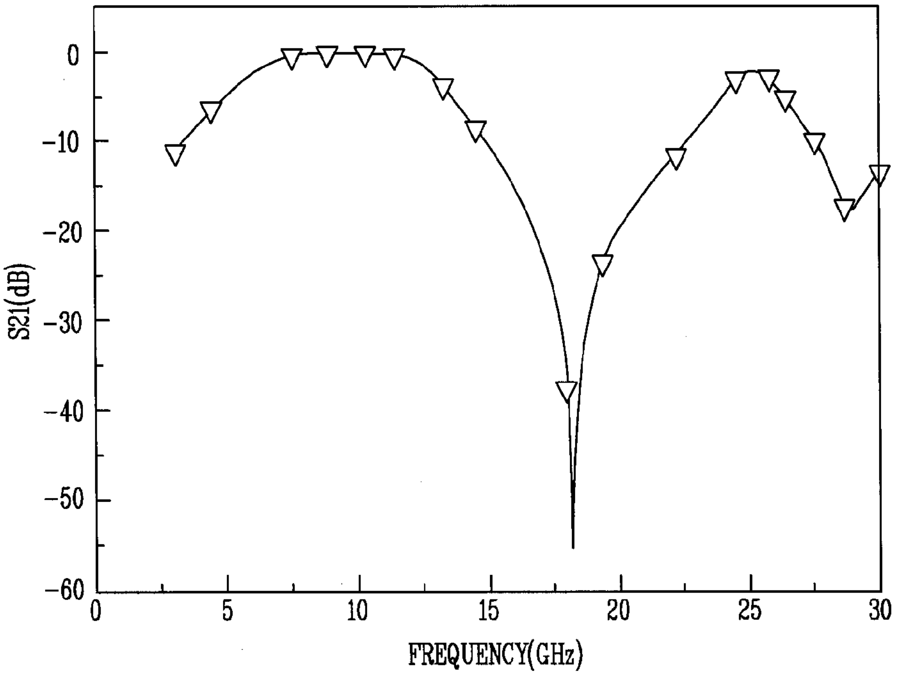
0
-10
-20
-30
-40
-50
-60
S21(dB)
0
5
10
15
20
25
30
FREQUENCY(GHz)

FIG. 6A

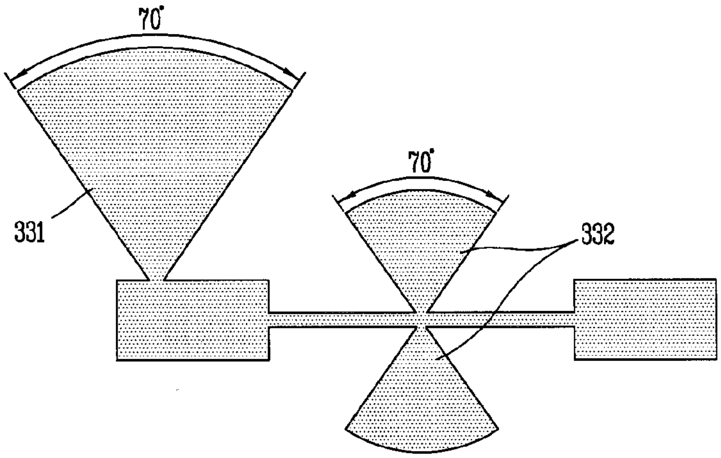
70°
70°
331
332

FIG. 6B

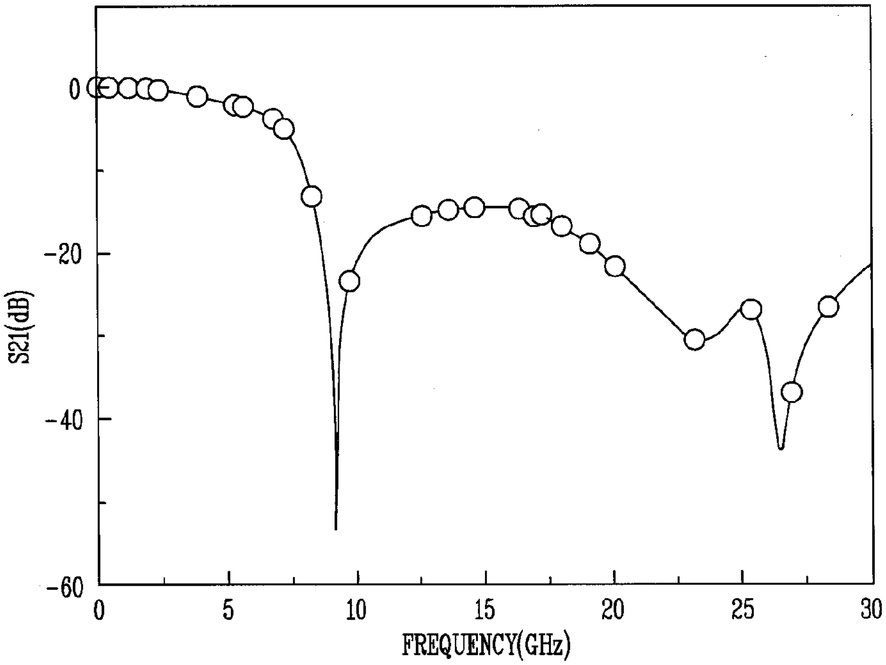

FIG. 7

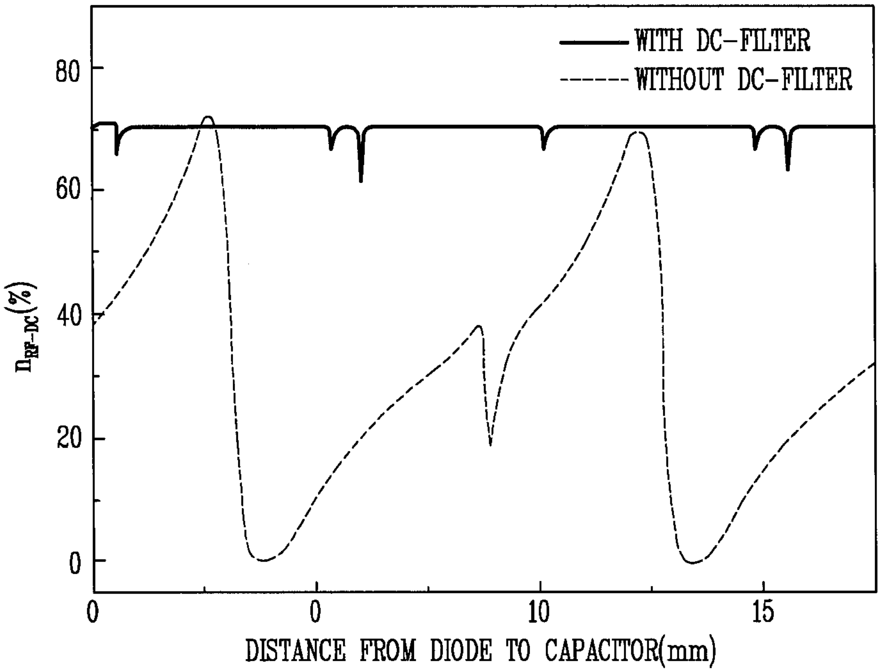
WITH DC-FILTER
WITHOUT DC-FILTER
80
60
40
20
0
n_{RF-DC}(%)
0
0
10
15
DISTANCE FROM DIODE TO CAPACITOR(mm)

FIG. 8A

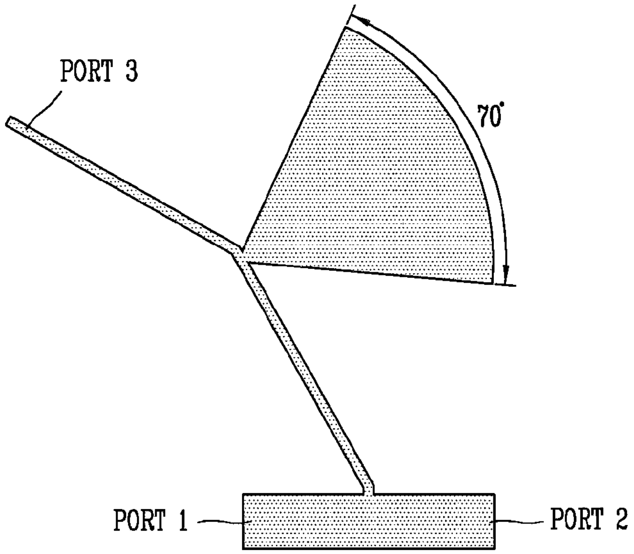
PORT 3
70°
PORT 1
PORT 2

FIG. 8B

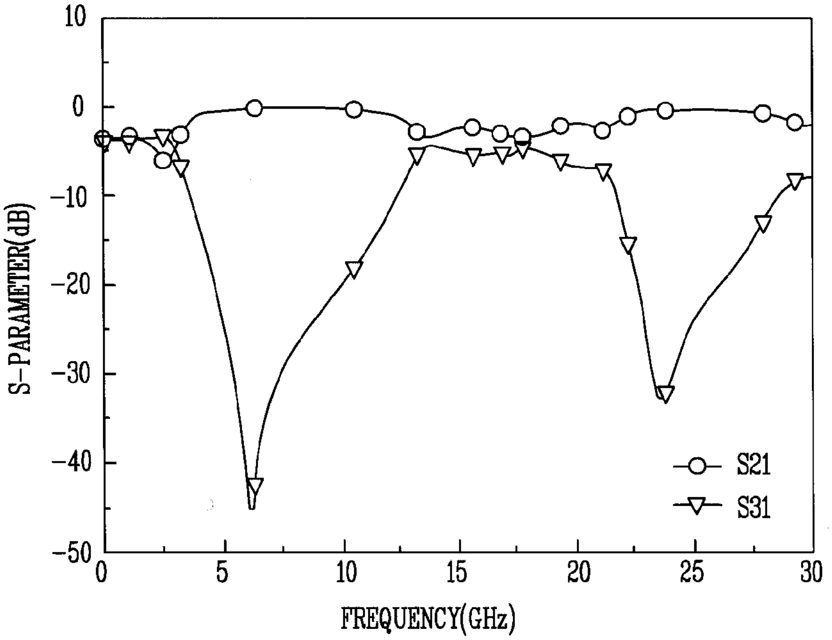
10
0
-10
-20
-30
-40
-50
S-PARAMETER(dB)
0
5
10
15
20
25
30
FREQUENCY(GHz)
S21
S31

FIG. 9A

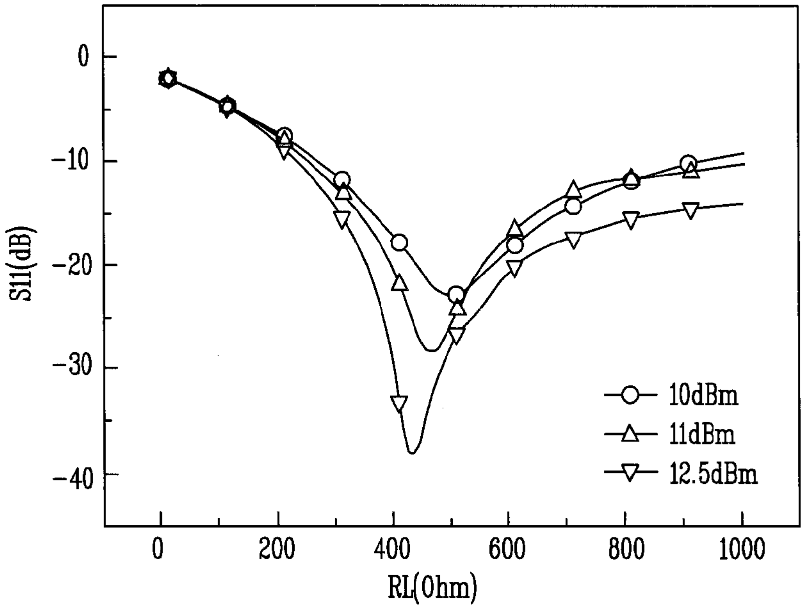
S11(dB)
0
-10
-20
-30
-40
0
200
400
600
800
1000
RL(Ohm)
10dBm
11dBm
12.5dBm

FIG. 9B

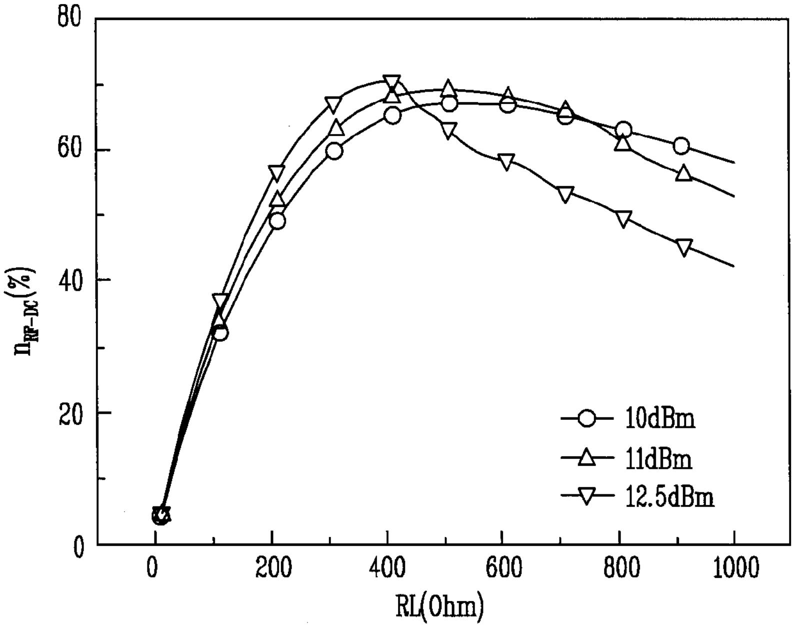
η_{RF-DC}(%)
80
60
40
20
0
0
200
400
600
800
1000
RL(Ohm)
10dBm
11dBm
12.5dBm

FIG. 10A

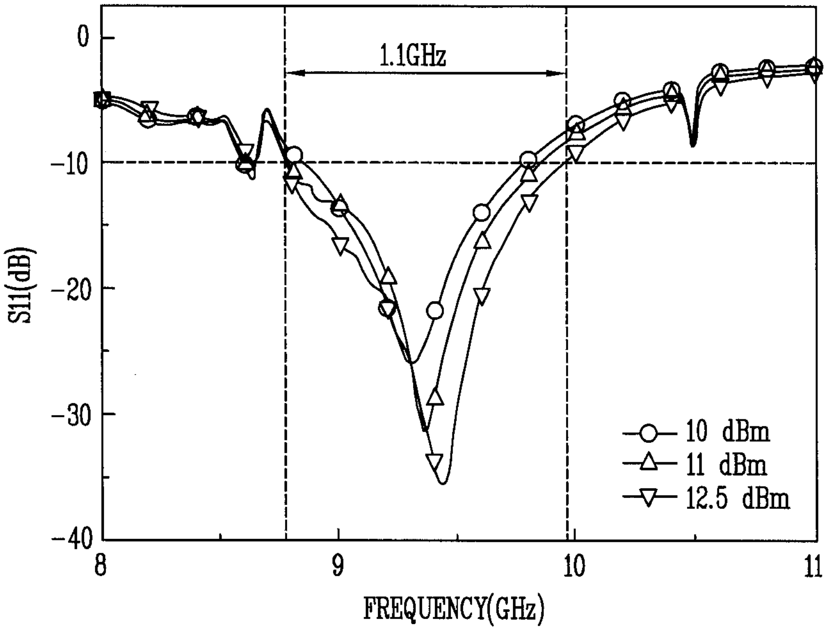

FIG. 10B

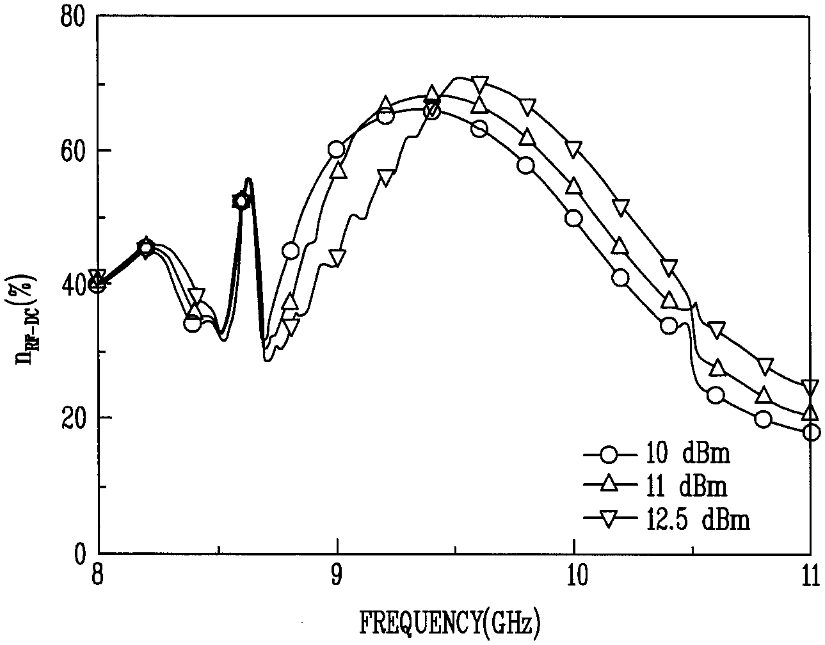

FIG. 11A

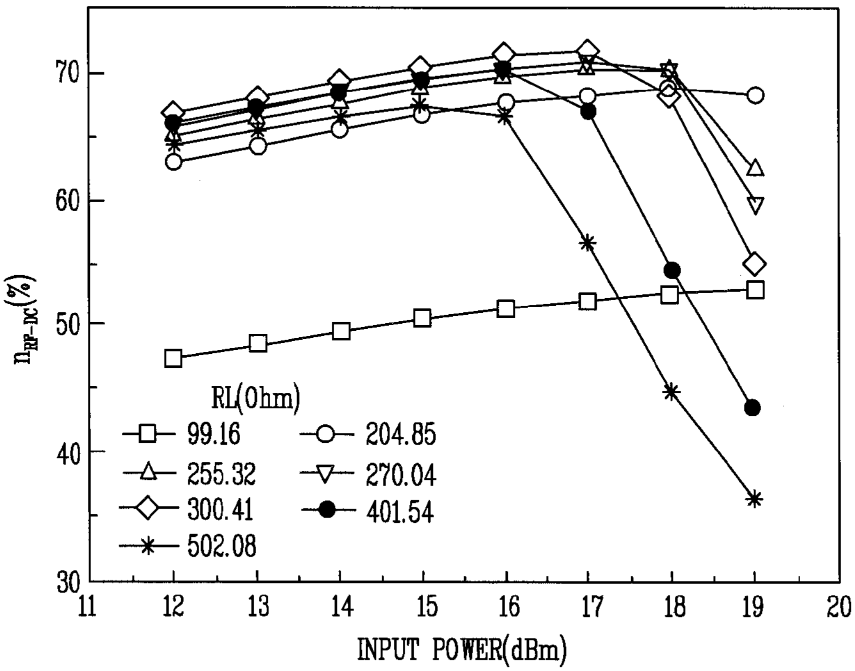

FIG. 11B

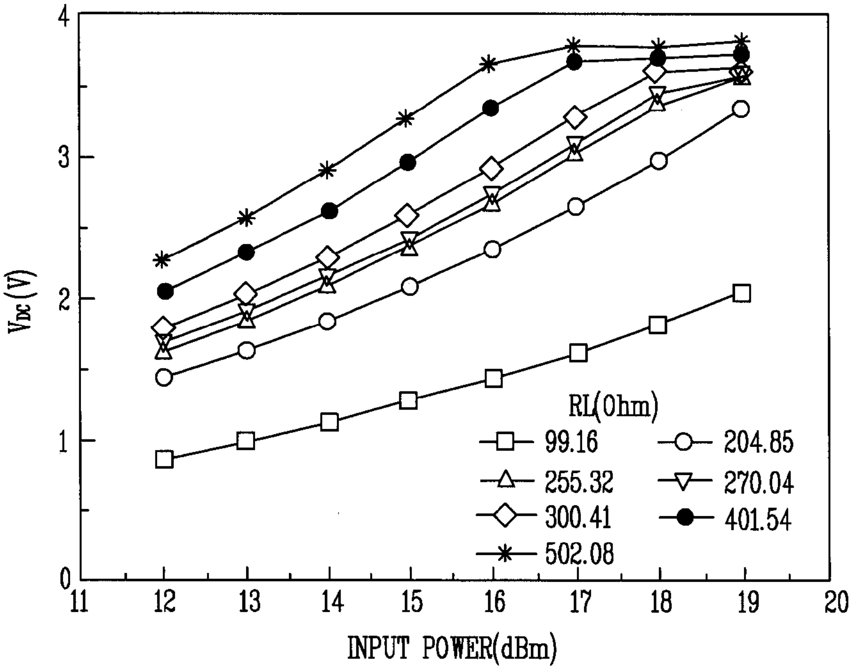

FIG. 12A

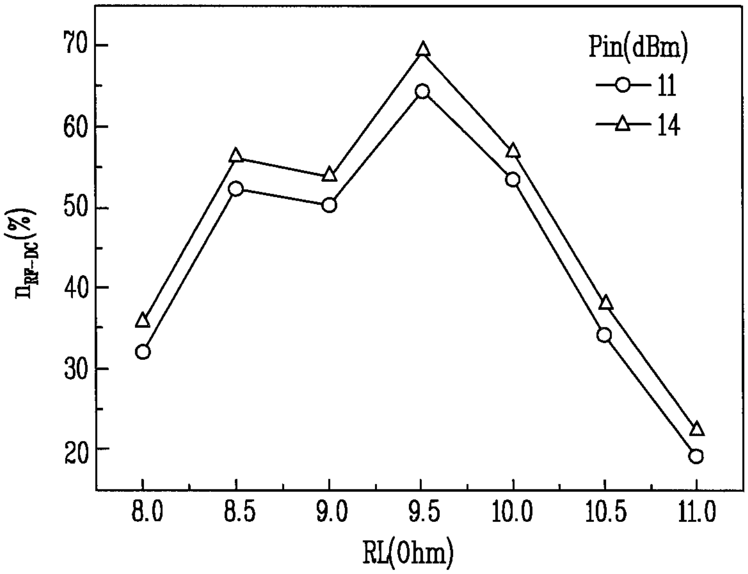

FIG. 12B

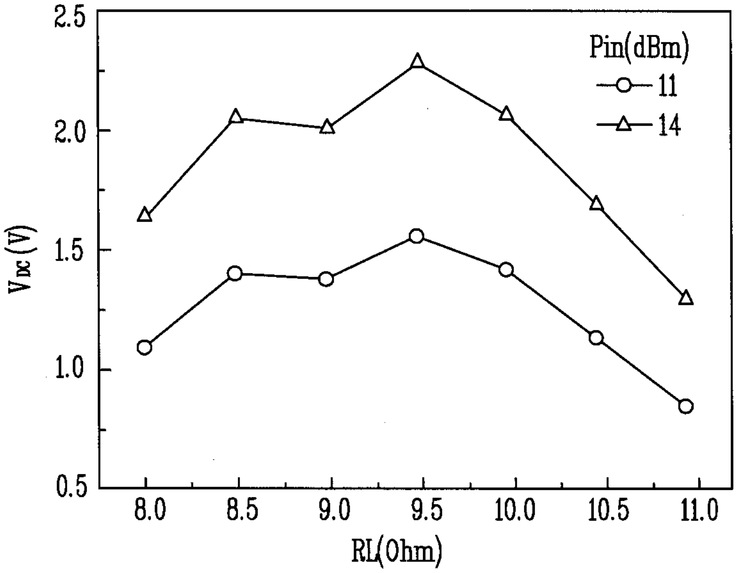

WIDEBAND RECTENNA AND RECTIFYING APPARATUS FOR RECTENNA

CROSS-REFERENCE TO RELATED APPLICATION

[0001] Pursuant to 35 U.S.C. §119(a), this application claims the benefit of earlier filing date and right of priority to Korean Application No. 10-2014-0128530, filed on Sep. 25, 2014, the contents of which is incorporated by reference herein in its entirety.

BACKGROUND OF THE DISCLOSURE

[0002] 1. Field of the Disclosure

[0003] This specification relates to a rectenna, and more particularly, a wideband rectenna.

[0004] 2. Background of the Disclosure

[0005] A rectenna, which converts radio frequency (RF) energy to direct-current (DC) power, is a key component of free-space microwave power transmission (MPT) systems for the application of microwave powered unmanned aerial vehicle (UAVs), space solar power satellite stations (SSPS) and the like. The constraints on the design of MPT systems are the choice of an operating frequency and a system configuration of a transmitter and a receiver.

[0006] Recently, an MPT consisting of an active phased array, a rectenna array system, and a beam controlled system was tested and found to have RF to DC conversion efficiency ($\eta_{RF\text{-}DC}$) of 54% at 1 mW RF input for the unit element and 5.8 GHz. This system, which operates for more than 8 hours, has transmitting and receiving array elements with circular polarization, allowing it to obtain high RF-DC conversion efficiency.

[0007] In order to design a compact rectenna array, a high operating frequency should be chosen to reduce an area of a receiving array aperture. For this reason, there have been some attempts to design rectennas at 35 GHz and 95 GHz. However, there are limitations such as low RF-DC conversion efficiency and high fabricating costs to develop such high power transmitting and receiving systems.

[0008] Therefore, there have been tradeoffs with a compact size and a light weight in terms of choosing an operating frequency for microwave powered aerial vehicle (MPAV) systems, which operate at X-band.

SUMMARY OF THE DISCLOSURE

[0009] Therefore, to obviate the aforementioned drawbacks, an aspect of the detailed description is to provide a wideband rectenna providing high RF-DC conversion efficiency while operating at a wideband, and a rectifying apparatus for rectenna.

[0010] Another aspect of the detailed description is to provide a wideband rectenna for realizing a transmitting and receiving system with a compact size and a light weight, and a rectifying apparatus for rectenna.

[0011] To achieve these and other advantages and in accordance with the purpose of this specification, as embodied and broadly described herein, there is provided a rectenna including a circular-polarized patch antenna having dual slots fed by a microstrip and configured to receive and output a radio frequency (RF) signal, and a rectifying circuit configured to convert for output the RF signal, received by the circular-polarized patch antenna, into a direct current (DC) signal and transfer the DC signal from the antenna to a load, wherein the rectifying circuit comprises at least one radial stub.

[0012] The radial stub may include a first radial stub used for rejecting harmonics of the RF signal, a second radial stub used for impedance matching between the antenna and a rectifying diode, a third radial stub used for the DC power transfer, and a fourth radial stub used for lowering an RF ripple voltage to stably supply the DC power to the load.

[0013] The first radial stub may operate as a band-stop filter.

[0014] The circular-polarized patch antenna may include a first patch substrate with a first patch disposed thereon, a second patch substrate with a second patch disposed thereon, a foam substrate disposed between the first patch substrate and the second patch substrate, and a feed line substrate attached to a lower portion of the second patch.

[0015] The feed line substrate may include H-shaped dual slots, T-shaped feed lines configured to feed the H-shaped dual slots, and a power divider configured to divide power to the feed lines.

[0016] The rectifying circuit may include a harmonic rejecting unit having coupled microstrip lines and a first radial stub and configured to prevent re-radiation of RF harmonics, a matching unit located at an output of the harmonic rejecting unit, having a second radial stub, and configured to perform impedance matching between the antenna and a rectifying diode, a rectifying diode located at an output of the matching unit, a direct current (DC)-pass filter having a third radial stub and configured to attenuate RF output by the rectifying diode to transfer DC power, and a DC return unit having a fourth radial stub and configured to lower a ripple voltage from a signal output by the DC-pass filter so as to supply the DC power stably to the load.

[0017] The harmonic rejecting unit may block noise signals introduced from the antenna.

[0018] The coupled microstrip lines each may operate as a band-pass filter.

[0019] The radial stub may have an angle of circumference of 70°.

[0020] The rectifying circuit may have an operating frequency near 9.5 GHz.

[0021] Also, the present invention provides a rectifying apparatus, including a harmonic rejecting unit having coupled microstrip lines and a first radial stub and configured to prevent re-radiation of radio frequency (RF) harmonics, a matching unit located at an output of the harmonic rejecting unit, having a second radial stub, and configured to perform impedance matching between an antenna and a rectifying diode, a rectifying diode located at an output of the matching unit, a direct current (DC)-pass filter having a third radial stub and configured to attenuate RF output by the rectifying diode to transfer DC power, and a DC return unit having a fourth radial stub and configured to lower a ripple voltage from a signal output by the DC-pass filter so as to supply the DC power stably to a load.

[0022] The first radial stub may be an open-circuited radial stub and operate as a band-stop filter.

[0023] The second radial stub may include two radial stubs symmetric to each other.

[0024] The third radial stub may include a single radial stub, and a pair of radial stubs symmetric to each other.

[0025] The fourth radial stub may include a signal radial stub.

[0026] The harmonic rejecting unit may block noise signals introduced from the antenna.

[0027] The coupled microstrip lines each may operate as a band-pass filter.

[0028] The first radial stub may operate as a band-stop filter.

[0029] The first to fourth radial stubs each may have an angle of circumference of 70°.

[0030] The rectifying apparatus may have an operating frequency near 9.5 GHz.

[0031] A rectenna and a rectifying apparatus for rectenna according to the present invention can provide high RF-DC conversion efficiency while operating at a wideband.

[0032] Also, the use of the rectifying apparatus and the rectenna according to the present invention may allow for implementing a compact system.

[0033] Further scope of applicability of the present application will become more apparent from the detailed description given hereinafter. However, it should be understood that the detailed description and specific examples, while indicating preferred embodiments of the disclosure, are given by way of illustration only, since various changes and modifications within the spirit and scope of the disclosure will become apparent to those skilled in the art from the detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0034] The accompanying drawings, which are included to provide a further understanding of the disclosure and are incorporated in and constitute a part of this specification, illustrate exemplary embodiments and together with the description serve to explain the principles of the disclosure.

[0035] In the drawings:

[0036] FIG. **1** is a configuration view of a general rectenna;

[0037] FIG. **2**A is a disassembled perspective view of a wideband circularly-polarized patch antenna in accordance with one embodiment disclosed herein;

[0038] FIG. **2**B(a) to **2**B(d) are views illustrating feeding circuits of an antenna in accordance with one embodiment disclosed herein;

[0039] FIG. **2**C is a view illustrating a fabricated four-element patch array structure in accordance with one embodiment disclosed herein;

[0040] FIGS. **3**A and **3**B are graphs showing simulation results of a four-element array antenna in accordance with one embodiment disclosed herein;

[0041] FIG. **4**A is a configuration view of a rectifying circuit in accordance with one embodiment disclosed herein;

[0042] FIG. **4**B is a view of a diode equivalent circuit for the circuit of FIG. **4**A;

[0043] FIG. **5**A is a view illustrating a layout of an input harmonic rejecting filter in accordance with one embodiment disclosed herein;

[0044] FIG. **5**B is a graph showing a transmission coefficient characteristic;

[0045] FIG. **6**A is view illustrating a layout of a DC-pass filter in accordance with one embodiment disclosed herein;

[0046] FIG. **6**B is a graph showing a transmission coefficient parameter;

[0047] FIG. **7** is a graph showing conversion efficiency of a DC-pass filter according to a distance from a diode to a capacitor;

[0048] FIG. **8**A is a view illustrating a layout of a DC-return circuit in accordance with one embodiment disclosed herein;

[0049] FIG. **8**B is a graph showing S-parameter;

[0050] FIG. **9**A is a graph showing simulated reflection coefficients of a rectifying circuit disclosed herein according to different input power;

[0051] FIG. **9**B is a graph showing simulated conversion efficiencies of the rectifying circuit disclosed herein according to different input power;

[0052] FIG. **10**A is a view illustrating a simulated value of a reflection coefficient according to different input power;

[0053] FIG. **10**B is a graph showing a simulated RF-DC conversion efficiency;

[0054] FIG. **11**A is a graph showing RF-DC conversion efficiency with respect to actually-measured input power in accordance with one embodiment of the present invention according to different loads;

[0055] FIG. **11**B is a graph showing a converted voltage according to input power;

[0056] FIG. **12**A is a graph showing a measured conversion efficiency in accordance with one embodiment disclosed herein;

[0057] FIG. **12**B is a graph showing a converted voltage.

DETAILED DESCRIPTION OF THE DISCLOSURE

[0058] The present invention may be varied by various ways and have various embodiments, and specific embodiments will be illustrated in the accompanying drawings and described in detail.

[0059] However, this is not to limit the present invention to the specific embodiments, and it should be understood that the present invention includes all the variations, equivalents and replacements belonging to the scope and technical range of the present invention.

[0060] The terms including an ordinal number such as first, second, etc. can be used to describe various elements, but the elements should not be limited by those terms. The terms are used merely for the purpose to distinguish an element from the other element. For example, a first element may be named to a second element, and similarly, a second element may be named to a first element without departing from the scope of right of the invention. The term "and/or" includes any and all combinations of one or more of the associated listed items.

[0061] It will be understood that when an element is referred to as being "connected with" another element, the element can be connected with the other element or intervening elements may also be present. In contrast, when an element is referred to as being "directly connected with" another element, there are no intervening elements present.

[0062] Terms used herein are merely intended to explain specific embodiments, and it should not be understood as limiting the present invention. A singular representation may include a plural representation unless it represents a definitely different meaning from the context. Terms such as "include" or "has" are used herein and should be understood that they are intended to indicate an existence of several characteristics, numbers, steps, operations, components, or combinations thereof, disclosed in the specification, and it is also understood that greater or fewer characteristics, numbers, steps, operations, components, or combinations thereof may likewise be utilized.

[0063] Unless otherwise defined, all terms used herein, including technical or scientific terms, have the same meanings as those generally understood by those with ordinary knowledge in the field of art to which the present invention belongs. Such terms as those defined in a generally used dictionary are to be interpreted to have the meanings equal to the contextual meanings in the relevant field of art, and are not to be interpreted to have ideal or excessively formal meanings unless clearly defined in the present application.

[0064] Hereinafter, description will be given of the preferred embodiments disclosed herein in more detail with reference to the accompanying drawings. In describing the present invention, in order to help understanding of the present invention, the same component on the drawings are designated by the same reference numeral, and repetitive description of the same component will be omitted.

[0065] FIG. **1** is a configuration view of a general rectenna.

[0066] As illustrated in FIG. **1**, a general rectenna may include an antenna **110**, and a rectifying circuit **120**. A load **130** may be connected to the rectenna, more particularly, to an output of the rectifying circuit **120**.

[0067] The rectifying circuit **120** may include a harmonic rejecting filter **121**, a matching circuit **122**, a diode **123** and a direct-current (DC) pass filter **124**.

[0068] Here, the non-linear characteristic of the diode **123** may generate unwanted harmonic signals that propagate backward to the antenna **110** and forward to the load **130**. Therefore, the harmonic rejecting filter **121**, the DC-pass filter **124** and the matching circuit **122** should be carefully designed to obtain an optimized RF-DC conversion efficiency ($\eta_{RF\text{-}DC}$).

[0069] Hereinafter, description will be given of a patch antenna structure according to the present invention.

[0070] FIG. **2**A is a disassembled perspective view of a wideband circularly-polarized patch antenna in accordance with one embodiment disclosed herein.

[0071] FIG. **2**A illustrates an antenna element which has a resonance cavity resulting from first and second patch substrates **210** and **230** having two layered square patches, and foam disposed between the substrates **210** and **230**.

[0072] Also, a substrate **240** having a vertical slot and a horizontal slot may be disposed between the second patch substrate **230** and a feed line substrate **250**. The vertical slot and the horizontal slot provide orthogonal excitation to the cavity.

[0073] Circular polarization is obtained by a Wilkinson power divider and a delay line with a 90-degree phase difference between dual slots.

[0074] That is, the patch antenna can be designed to obtain broadband and circular-polarization characteristics using H-shaped slots and T-shaped feed lines that are stacked in a multi-layer substrate structure.

[0075] As illustrated in FIG. **2**A, the patch antenna according to the one embodiment disclosed herein may be configured in a manner that the feed line substrate **250** for a feed line **260** is 0.38 mm thick, the substrate **230** for a lower patch is 0.64 mm thick, the substrate **220** for foam is 0.1 mm thick, and the substrate **210** for an upper patch is 0.25 mm thick.

[0076] Also, relative dielectric constants ($\in_r$) of the substrate **250** for the feed line, the substrate **230** for the lower patch, the substrate for the foam and the substrate **210** for the upper patch may be 3.5, 2.17, 1.02, and 2.17, respectively.

[0077] Here, the substrates for the dual patches and the foam may be implemented with the low relative dielectric constants so as to obtain maximum radiation efficiency and wide bandwidth, while the feed line substrate may be implemented with the relatively higher relative dielectric constant than those of the substrates for the dual patches and the foam so as to improve maximum feed efficiency.

[0078] The resonance frequency of the cavity is mainly determined by a side length of a square patch. Sizes of the upper patch and the lower patch can be designed to have a small difference in order to obtain the wideband characteristics of the antenna element.

[0079] For example, the lengths of the upper square patch and the lower square patch may be 9.64 mm and 8.51 mm, respectively.

[0080] FIG. **2**B(a) to **2**B(d) are views illustrating feeding circuits of an antenna in accordance with one embodiment disclosed herein.

[0081] In detail, FIG. **2**B(a) illustrates an H-shaped slot **241** used for a wideband circular-polarized patch antenna, FIG. **2**B(b) and FIG. **2**B(c) illustrate T-shaped feed lines **261** and **262**, and FIG. **2**B(d) illustrates a power divider **263**. Optimized geometrical parameters are illustrated in FIG. **2**B(a) to FIG. **2**B(d).

[0082] Here, a T-shaped microstrip open stub may be electrically connected to ports **2** and **3** of the power divider **263** illustrated in FIG. **2**B(d), respectively. The T-shaped feed line may be designed to occupy a smaller space than the general microstrip open stub with a length of λ/4. These stubs feed the H-shaped dual slots **241**.

[0083] The power divider **263** may include a delay line **1** which extends more along one of the port **2** or the port **3** to be longer than the other, in order to create a 90-degree phase difference between the port **2** and the port **3**.

[0084] Each port of the power divider **263** disclosed herein preferably has characteristic impedance of 65Ω instead of 50Ω, in order to provide an enough array spacing by reducing a width of the divider **263** having higher impedance.

[0085] FIG. **2**C is a view illustrating a fabricated four-element patch array structure in accordance with one embodiment disclosed herein.

[0086] As illustrated in FIG. **2**C, an actually-fabricated four-element patch array preferably has an element spacing of 0.8λ. Here, λ is preferably the wavelength at 9.5 GHz.

[0087] The patch array structure of FIG. **2**C may have a Wilkinson power divider to feed the four patches. A patch cavity and a microstrip feed circuit are isolated by a conductor plane having the H-shaped slots, which may result in the merit of easy fabrication of a rectifier circuit on a feeding circuit layer for a large rectenna array.

[0088] FIG. **3** is a graph showing simulation results of a four-element array antenna in accordance with one embodiment disclosed herein.

[0089] In FIG. **3**A, simulated and measured reflection coefficients for the four-element array antenna are indicated with S_{11} with respect to a frequency.

[0090] The measured S_{11} has the lowest value at around 9.5 GHz, and a wideband characteristic in the range from −10 dB over 2 GHz, which is similar to the wideband characteristic of the simulated S_{11}.

[0091] The measured lowest value is −34 dB at 9.3 GHz.

[0092] FIG. **3**B is a graph showing gains of a fabricated four-element array antenna in accordance with one embodiment of the present invention.

[0093] The graph of FIG. **3**B shows simulated and measured gains of the four-element array antenna at 9.5 GHz.

[0094] FIG. **3**B shows horizontal radiation characteristics of the gains. Because the patch array is symmetric, vertical characteristics are similar to the horizontal characteristics.

[0095] Meanwhile, it can be noticed that the measured gain of the four-element array antenna is in a good agreement with the simulated gain. The measured and simulated gains are 11.2 dBi and 12.7 dBi, respectively.

[0096] It can also be understood that a circular-polarization (CP) characteristic of the four-element array antenna is measured in an axial ratio of 1.1 dB at a boresight angle of the antenna.

[0097] FIG. **4**A is a configuration view of a rectifying circuit in accordance with one embodiment disclosed herein, and FIG. **4**B is a view of a diode equivalent circuit for the circuit of FIG. **4**A.

[0098] A rectifying circuit **300** illustrated in FIG. **4**A is an X-band rectifying circuit. A rectifier designed according to one embodiment disclosed herein is compact because a single X-band rectenna is used for a large rectenna array application with a device spacing of 0.8λ. For this application, a feeding network of the antenna and the rectifier are designed using a microstrip line structure.

[0099] The rectifier (for example, a half-wave rectifier) may include a harmonic rejecting unit **310**, a matching unit **320**, a rectifying diode **323**, a DC-patch unit **330** and a DC return unit.

[0100] Here, the diode **323** is one of the most critical components for obtaining high RF-DC conversion efficiency (η_{RF-DC}) in the rectifying circuit **300**. Here, the RF-DC conversion efficiency (η_{RF-DC}) of the rectifying circuit **300** may be defined as the following Formula 1.

$$\eta_{RF-DC} = \frac{P_{DC}}{P_{RF}} = \frac{V_{DC}^2 / R_L}{P_{RF}}$$ [Formula 1]

[0101] Here, P_{DC}, P_{RF}, V_{DC} and R_L denote converted DC power, received RF power, converted DC voltage and load impedance, respectively.

[0102] The rectifying diode **323** of the rectenna as illustrated in FIG. **4**A, as aforementioned, is the most important component in the design of the rectifying circuit **300**. Therefore, the rectenna antenna according to the present invention is preferably set by several conditions to be explained later.

[0103] Conditions for the diode according to the present invention include a high breakdown voltage V_{br} which is related to efficiency of high input power, a low forward voltage V_f which is related to capability of detecting (capturing) small signals, low series resistance R_s which is related to overall efficiency of the rectifier, and low junction capacitance C_j which is related to an operation bandwidth.

[0104] The diode disclosed herein may preferably be a packaged Schottky barrier diode of type HSMS-8101 from Avago Technologies, for example. Since the diode optimally operates over a frequency range from 10 GHz to 14 GHz, it is suitable for use at 9.5 GHz.

[0105] Referring to FIG. **4**B which shows a diode equivalent circuit for the circuit of FIG. **4**A, the series resistance voltage R_s is 6 (R_s=6), the junction capacitance C_j is 0.17 pF (C_j=0.17 pF), the forward voltage V_f is in the range of 0.25 to 0.35 V, and the breakdown voltage V_{br} is greater than −4V (V_{br}<−4V). The junction capacitance C_j is variable according to operating conditions.

[0106] Under these conditions, if it is assumed that a 9.5 GHz operating frequency, 10 mW input power and 100Ω load are provided, upon having 50Ω source impedance, a calculated input impedance of the diode using a circuit simulation is (22-j20) Ω. The impedance difference between the antenna and the diode causes a reduction of power transfer and diode efficiency. Therefore, in order to maximize the RF power transfer from the antenna to the diode, the impedance matching circuit **320** may be designed as illustrated in FIG. **4**A.

[0107] That is, because the input impedance of the diode changes according to the operating frequency, the RF input power and the load resistance, the matching circuit **320**, as illustrated in FIG. **4**A, is preferably designed using two radial stubs for wideband operations.

[0108] Here, the radial stubs for impedance matching between the antenna and the diode **323** preferably have an angle of circumference of 70°.

[0109] In general, non-linearity of the diode generates higher-order harmonics, which causes the reduction of RF-DC conversion efficiency. Thus, the harmonic rejecting filter **310** and the DC-pass filter **330** are necessary to suppress the higher-order harmonics.

[0110] FIG. **5**A is a view illustrating a layout of an input harmonic rejecting filter in accordance with one embodiment disclosed herein, and FIG. **5**B is a graph showing transmission coefficient S_{21} characteristics of the input harmonic rejecting filter.

[0111] The harmonic rejecting filter **310** illustrated in FIG. **5**A may include coupled microstrip lines **311** and a radial stub **312**.

[0112] The harmonic rejecting filter **310** prevents re-radiation of RF harmonics from the diode **323** and blocks unwanted noise signals received from the antenna.

[0113] The transmission coefficient S_{21} may be analyzed using a simulator (for example, an advanced design system (ADS) momentum simulator). The coupled microstrip bandpass filter **311** of the harmonic rejecting filter **310** are centered at 9.5 GHz operating frequency with −3 dB, and has 7.3 GHz bandwidth.

[0114] In other words, the harmonic rejecting filter which includes an open-circuited radial stub **312** operating as a band-stop filter and the band-pass filter **311**, as illustrated in FIG. **5**B, attenuates a first harmonic frequency near 19 GHz and a second harmonic frequency near 28.5 GHz, so as to lower the transmission coefficients S_{21} to −26.6 dB and −16.5 dB, respectively.

[0115] Here, the radial stub **312** for rejecting the harmonics of the RF signals, as illustrated in FIG. **5**A, preferably has an angle of circumference of 70°.

[0116] FIG. **6**A is view illustrating a layout of a DC-pass filter in accordance with one embodiment disclosed herein, and FIG. **6**B is a graph showing a transmission coefficient S_{21} characteristic of the DC-pass filter.

[0117] The DC-pass filter according to the one embodiment disclosed herein attenuates the RF so as to transfer only DC power to the load. The DC-pass filter may be positioned at an output of the diode **323**. The DC-pass filter **330** includes a single radial stub **331** and a pair of radial stubs **332** symmetric to each other, and attenuates the transmission coefficient parameters S_{21} of fundamental, second and third harmonics to be lower than 15 dB, namely, down to −26 dB, −18.8 dB and −25.5 dB, respectively.

[0118] Here, the single radial stub **331** for DC power transfer or the pair of radial stubs **332** for lowering an RF ripple voltage, as illustrated in FIG. **6**A, preferably have an angle of circumference of 70°.

[0119] FIG. **7** is a graph showing conversion efficiency of a DC-pass filter according to a distance from a diode to a capacitor.

[0120] Here, a distance from the diode **323** to a capacitor C_L may be indicated with a reference numeral d, as illustrated in FIG. **4**A.

[0121] In the rectifying circuit **300**, a capacitor which is connected in parallel to the load lowers a ripple voltage and supplies DC power stably to the load.

[0122] As illustrated in FIG. **7**, it can be understood that the DC-pass filter **330** maintains the RF-DC conversion efficiency almost to be a constant with respect to the distance which changes.

[0123] On the other hand, in case of an operation without the DC-pass filter **330** disclosed herein, it can be noticed that the RF-DC conversion efficiency is drastically degraded and an optimum length for the maximum RF-DC conversion efficiency is present.

[0124] Therefore, the DC-pass filter **330** disclosed herein may be a key component for the stable operation of the rectifying circuit **300**.

[0125] FIG. **8**A is a view illustrating a layout of a DC-return unit in accordance with one embodiment disclosed herein, and FIG. **8**B is a graph showing S-parameter.

[0126] The DC return unit **340** is used to prevent re-reflection of harmonics generated by the diode **323** when a reverse current flows into the diode **323**. The DC return unit **340** may reject a countercurrent from the ground (port **3** of FIG. **8**A) back to the rectifying circuit and prevent RF re-reflection to the antenna.

[0127] The DC return unit **340** may function as a band-stop filter to block an X-band transmission between the ground (port **3**) and ports **1** and **2**.

[0128] On the other hand, the RF transmission from the port **1** to the port **2** at the frequency ranging from 3.2 to 13.5 GHz (3 dB band-pass region) is not affected by the ground.

[0129] Here, as illustrated in FIG. **8**A, the DC return unit **340** may further include a radial stub between the ground (port **3**) and the ports **1** and **2**. To lower an RF ripple voltage and supply DC power stably to the load, the radial stub preferably has an angle of circumference of 70°.

[0130] FIG. **9**A is a graph showing simulated reflection coefficients of a rectifying circuit disclosed herein according to different input power, and FIG. **9**B is a graph showing simulated conversion efficiencies of the rectifying circuit disclosed herein according to different input power.

[0131] FIGS. **9**A and **9**B show calculated S_{11} and RF-DC conversion efficiency for the estimation of the rectifying circuit performance under various operation conditions.

[0132] The RF-DC conversion efficiency ($\eta_{RF\text{-}DC}$) was calculated according to Formula 1, as aforementioned, with respect to load resistance of 50Ω to 1000Ω and input power of 10 dBm, 11 dBm and 12.5 dBm.

[0133] Since the input impedance of the diode changes according to the operation conditions, as illustrated in FIGS. **9**A and **9**B, the rectifying circuit has been optimized with respect to the load resistance according to the input power at 9.5 GHz. When the input power is 12.5 dBm and the load resistance is 400Ω, the maximum efficiency and voltage are 71% and 2.25 V, respectively.

[0134] Because the impedances of the diode and each of the components of the rectifying circuit vary according to changes of the operation conditions, it is difficult to design the rectifying circuit for wideband operation. To overcome the difficulty in the present invention, the components of the rectifying circuit were designed using radial stubs to allow for operations over wideband frequencies.

[0135] FIGS. **10**A and **10**B are graphs showing simulation results for S_{11} and RF-DC conversion efficiency according to a frequency and different input power.

[0136] According to the simulation results illustrated in FIG. **10**A, it can be understood that the reflection coefficient S_{11} is lower than −10 dB in a frequency range from 8.8 to 9.9 GHz. Also, it can be checked in FIG. **10**B that the RF-DC conversion efficiency is higher than 50% in a frequency range from 9.1 to 10.2 GHz with respect to input power of 12.5 dBm.

[0137] FIGS. **11**A and **11**B are graphs showing RF-DC conversion efficiency and a converted voltage with respect to actually-measured input power according to different loads.

[0138] It can be understood that the overall tendency of the measured RF-DC conversion efficiency with respect to the input power illustrated in FIG. **11**A is similar to the simulation results. However, an optimal condition for the maximum RF-DC conversion efficiency is slightly different. The simulated maximum RF-DC conversion efficiency is 71% with respect to 12.5 dBm under the load of 40Ω, but the measured maximum RF-DC conversion efficiency is 71.9% with respect to 17 dBm under the load of 300Ω. It is obvious from the measurement results that the diode and the rectifying circuit according to the present invention can be used at higher input power than the simulated value.

[0139] Referring to FIG. **11**B, it can also be noticed that the converted DC voltage increases as the input power increases and the DC voltage at the maximum efficiency is 3.2 V.

[0140] Compared to the conventional conversion circuit which has exhibited 63.2% efficiency at 8.51 GHz or 21% efficiency at 9.3 GHz, it can be noticed that the efficiency of the RF-DC conversion circuit according to the present invention is higher than those of such conventional circuits at the X-band frequency.

[0141] FIGS. **12**A and **12**B are graphs showing measured conversion efficiency and converted voltage in accordance with one embodiment disclosed herein.

[0142] As illustrated in FIG. **12**A, it can be seen that the rectifying circuit disclosed herein has the maximum efficiency of 71.9% and keeps having efficiency more than 50% over 1 GHz frequency bandwidth.

[0143] The rectifying circuit according to the present invention is optimized compactly with dimensions of 35 mm by 21 mm (35×21) at 9.5 GHz, so as to be applied to a large rectenna array system.

[0144] The present invention has been described with reference to the foregoing embodiments, but it can be understood that various modifications and changes can be obtained from the present invention without departing from the scope of the present invention as defined in the appended claims.

What is claimed is:

1. A rectenna comprising:

a circular-polarized patch antenna configured to receive a radio frequency (RF) signal; and

a rectifying circuit configured to convert for output the RF signal, received by the circular-polarized patch antenna, into a direct current (DC) signal and transfer the DC signal from the antenna to a load,

wherein the rectifying circuit comprises at least one radial stub.

2. The rectenna of claim **1**, wherein the radial stub comprises at least one of a first radial stub used for rejecting harmonics of the RF signal, a second radial stub used for impedance matching between the antenna and a rectifying diode, a third radial stub used for the DC power transfer, and a fourth radial stub used for lowering an RF ripple voltage to stably supply the DC power to the load.

3. The rectenna of claim **2**, wherein the first radial stub operates as a band-stop filter.

4. The rectenna of claim **1**, wherein the circular-polarized patch antenna comprises:

a first patch substrate with a first patch disposed thereon;

a second patch substrate with a second patch disposed thereon;

a foam substrate disposed between the first patch substrate and the second patch substrate; and

a feed line substrate attached to a lower portion of the second patch.

5. The rectenna of claim **4**, wherein the feed line substrate comprises:

H-shaped dual slots;

T-shaped feed lines configured to feed the H-shaped dual slots; and

a power divider configured to divide power to the feed lines.

6. The rectenna of claim **1**, wherein the rectifying circuit comprises:

a harmonic rejecting unit configured to prevent re-radiation of RF harmonics, the harmonic rejecting unit having coupled microstrip lines and a first radial stub;

a matching unit located at an output of the harmonic rejecting unit and configured to perform impedance matching between the antenna and a rectifying diode, the matching unit having a second radial stub;

a rectifying diode located at an output of the matching unit;

a direct current (DC)-pass filter having a third radial stub and configured to attenuate RF output by the rectifying diode to transfer DC power; and

a DC return unit having a fourth radial stub and configured to lower a ripple voltage from a signal output by the DC-pass filter so as to supply the DC power stably to the load.

7. The rectenna of claim **6**, wherein the harmonic rejecting unit blocks noise signals introduced from the antenna.

8. The rectenna of claim **6**, wherein the coupled microstrip lines each operates as a band-pass filter.

9. The rectenna of claim **1**, wherein the radial stub has an angle of circumference of 70°.

10. The rectenna of claim **1**, wherein the rectifying circuit has an operating frequency near 9.5 GHz.

11. A rectifying apparatus comprising:

a harmonic rejecting unit having coupled microstrip lines and a first radial stub and configured to prevent re-radiation of radio frequency (RF) harmonics;

a matching unit located at an output of the harmonic rejecting unit and configured to perform impedance matching between an antenna and a rectifying diode, the matching unit having a second radial stub;

a rectifying diode located at an output of the matching unit;

a direct current (DC)-pass filter having a third radial stub and configured to attenuate RF output by the rectifying diode to transfer DC power; and

a DC return unit having a fourth radial stub and configured to lower a ripple voltage from a signal output by the DC-pass filter so as to supply the DC power stably to a load.

12. The apparatus of claim **11**, wherein the coupled microstrip lines each operates as a band-pass filter.

13. The apparatus of claim **11**, wherein the first radial stub is an open-circuited radial stub and operates as a band-stop filter.

14. The apparatus of claim **11**, wherein the second radial stub comprises two radial stubs symmetric to each other.

15. The apparatus of claim **11**, wherein the third radial stub comprises a single radial stub, and a pair of radial stubs symmetric to each other.

16. The apparatus of claim **11**, wherein the fourth radial stub comprises a signal radial stub.

17. The apparatus of claim **11**, wherein the harmonic rejecting unit blocks noise signals introduced from the antenna.

18. The apparatus of claim **11**, wherein the first radial stub operates as a band-stop filter.

19. The apparatus of claim **11**, wherein the first to fourth radial stubs each has an angle of circumference of 70°.

20. The apparatus of claim **11**, wherein the rectifying apparatus has an operating frequency near 9.5 GHz.

* * * * *

US 20170227781A1

(19) **United States**

(12) **Patent Application Publication**

Banerjee et al.

(10) **Pub. No.: US 2017/0227781 A1**

(43) **Pub. Date: Aug. 10, 2017**

(54) **APPARATUSES AND METHODS FOR MAKING AN OBJECT APPEAR TRANSPARENT**

(71) Applicant: **Toyota Motor Engineering & Manufacturing North America, Inc.**, Erlanger, KY (US)

(72) Inventors: **Debasish Banerjee**, Ann Arbor, MI (US); **Hideo Iizuka**, Nissin-shi (JP)

(21) Appl. No.: **15/185,988**

(22) Filed: **Jun. 17, 2016**

Related U.S. Application Data

(60) Provisional application No. 62/291,920, filed on Feb. 5, 2016.

Publication Classification

(51) **Int. Cl.**

G02B 27/14	(2006.01)
B60R 13/02	(2006.01)
G02B 5/30	(2006.01)
G02B 27/28	(2006.01)
G02B 27/10	(2006.01)

(52) **U.S. Cl.**

CPC ***G02B 27/14*** (2013.01); ***G02B 27/283*** (2013.01); ***G02B 27/1086*** (2013.01); ***G02B 5/3083*** (2013.01); ***B60R 13/025*** (2013.01); *B60R 2013/0287* (2013.01)

(57) **ABSTRACT**

A cloaking device includes cloaking region boundary planes oriented non-planar to each other, each of the cloaking region boundary planes having an outward facing mirror surface and an inward facing opaque surface. The cloaking device includes a cloaking region bounded at least partially by the inward facing opaque surfaces of the cloaking region boundary planes. Half mirrors are spaced apart and generally parallel to the outward facing mirror surfaces such that a half mirror is spaced apart and generally parallel to each outward facing mirror surface. Light from an object on an object-side of the cloaking device is directed around an article within the cloaking region and forms an image on an image-side of the cloaking device such the article appears transparent to an observer looking towards the object.

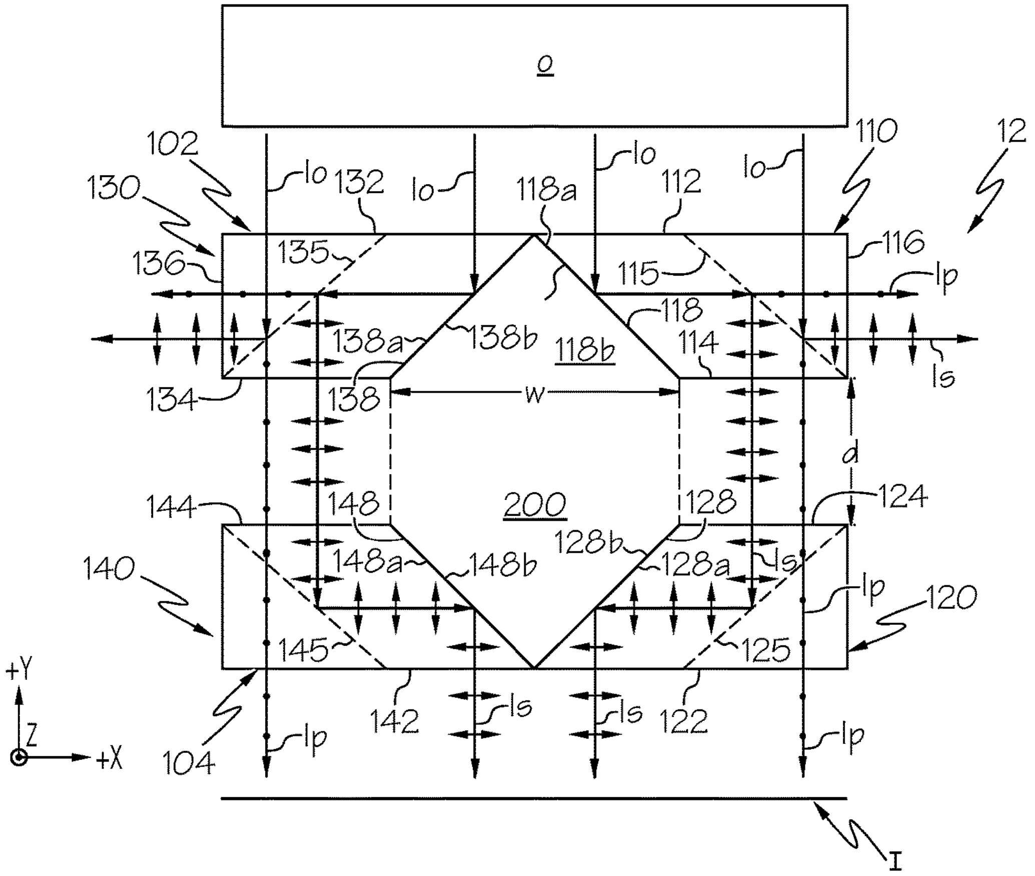

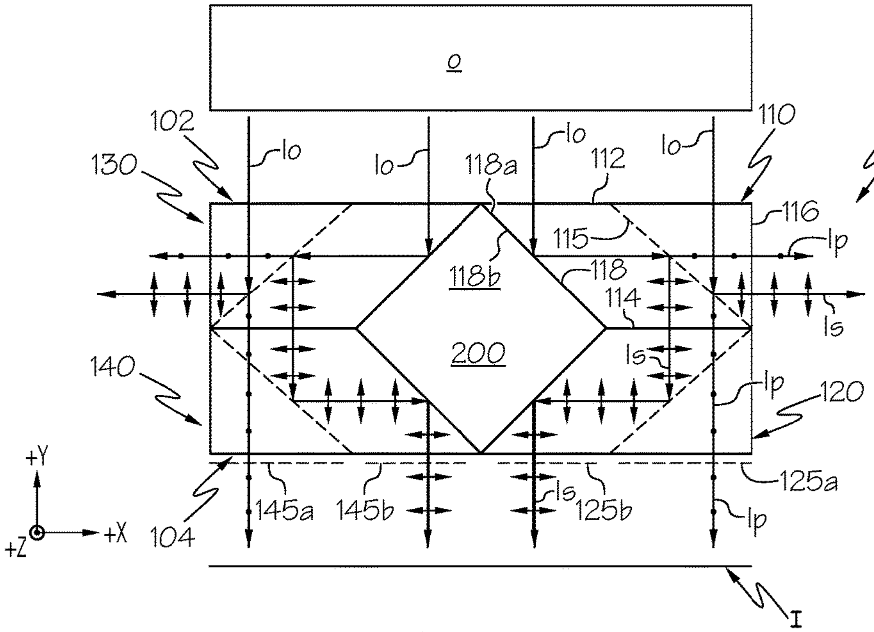
O
102
130
Io
Io
118a
Io
Io
110
10
112
116
115
Ip
118b
118
114
Is
200
Is
140
Ip
120
125a
+Y
145a
145b
Is
125b
+Z
+X
104
Ip
I

FIG. 1

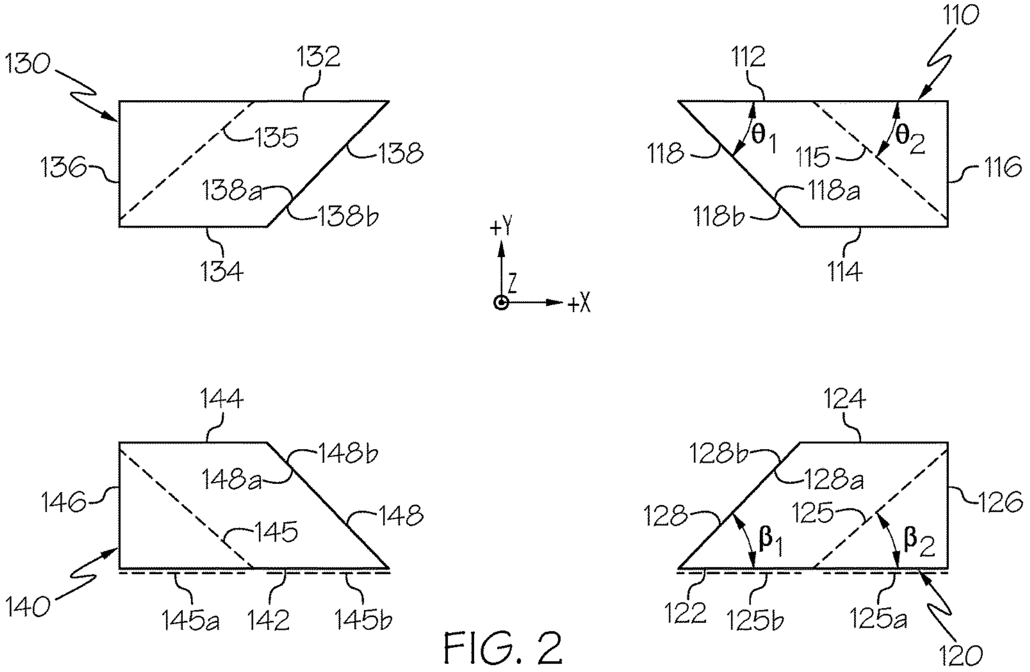
θ1
θ2
+Y
Z
+X
β1
β2

FIG. 2

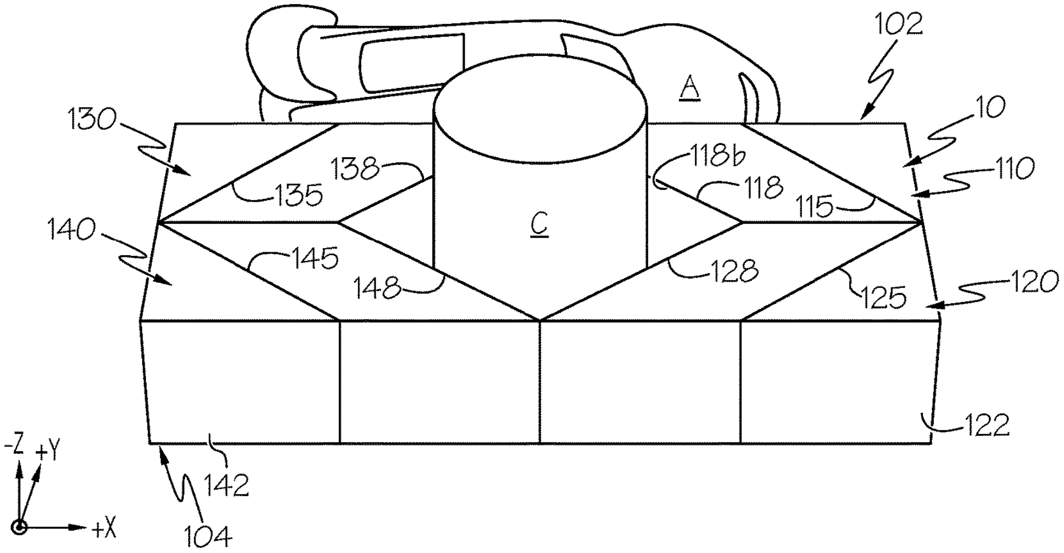
A
C
-Z
+Y
+X

FIG. 3

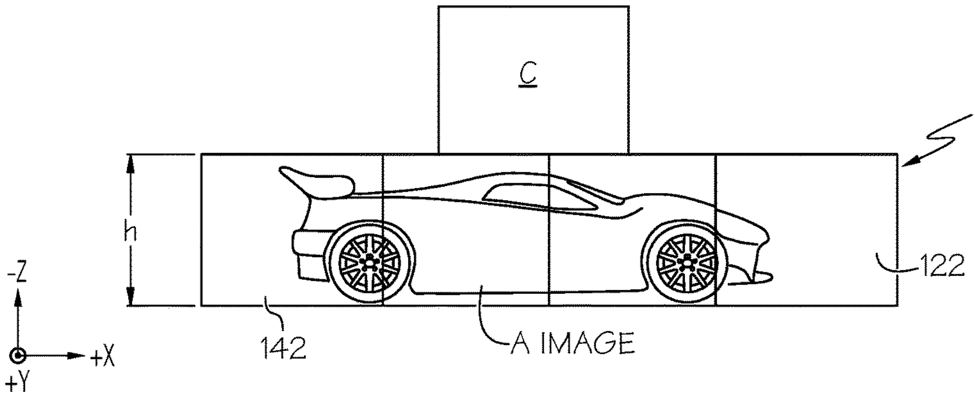
C
10
h
122
-Z
+X
+Y
142
A IMAGE

FIG. 4

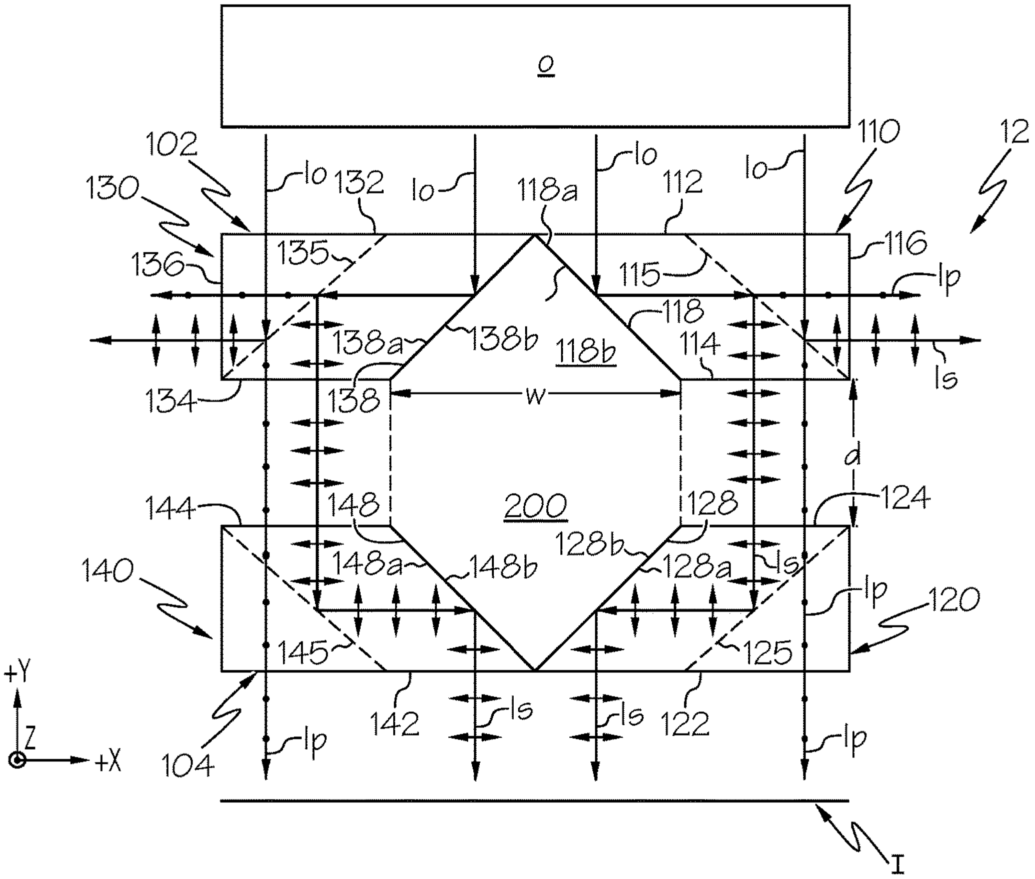
O
Io
Io
Io
Io
Ip
Is
w
d
Is
Ip
+Y
Z
+X
Is
Is
Ip
Ip
I

FIG. 5

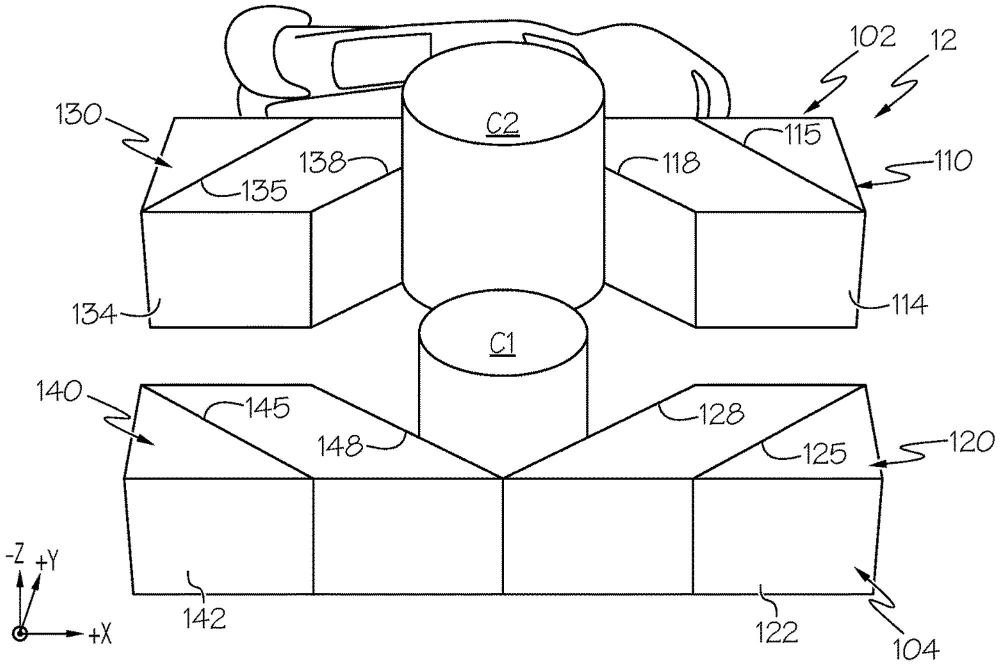
C2
C1
-Z
+Y
+X

FIG. 6

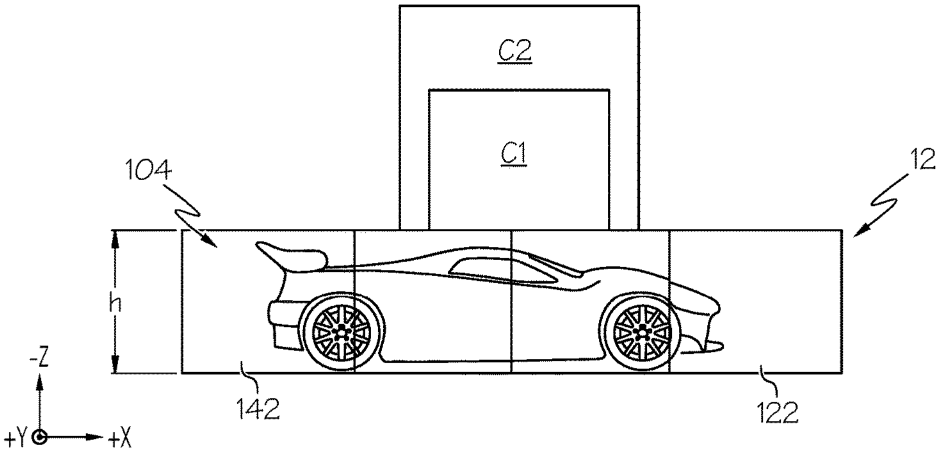
C2
C1
104
12
h
-Z
+Y
+X
142
122

FIG. 7

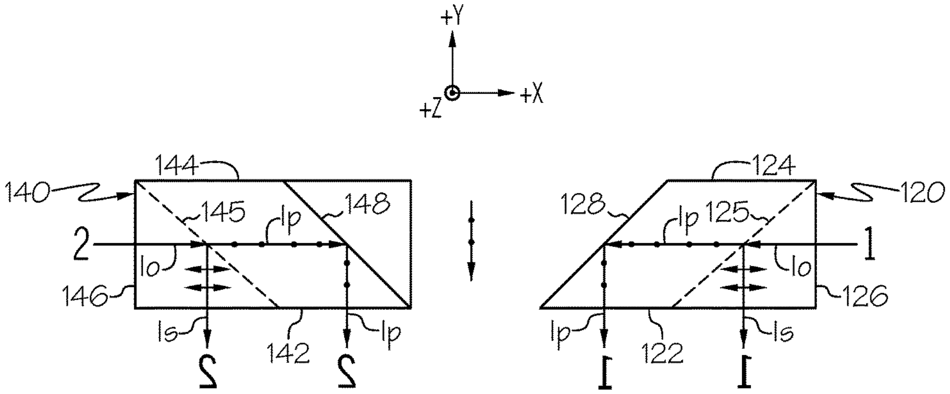
+Y
+X
+Z
140
144
145
Ip
148
2
Io
146
Is
142
2
2
Ip
128
124
120
Ip
125
1
Io
126
Ip
122
Is
1
1

FIG. 8

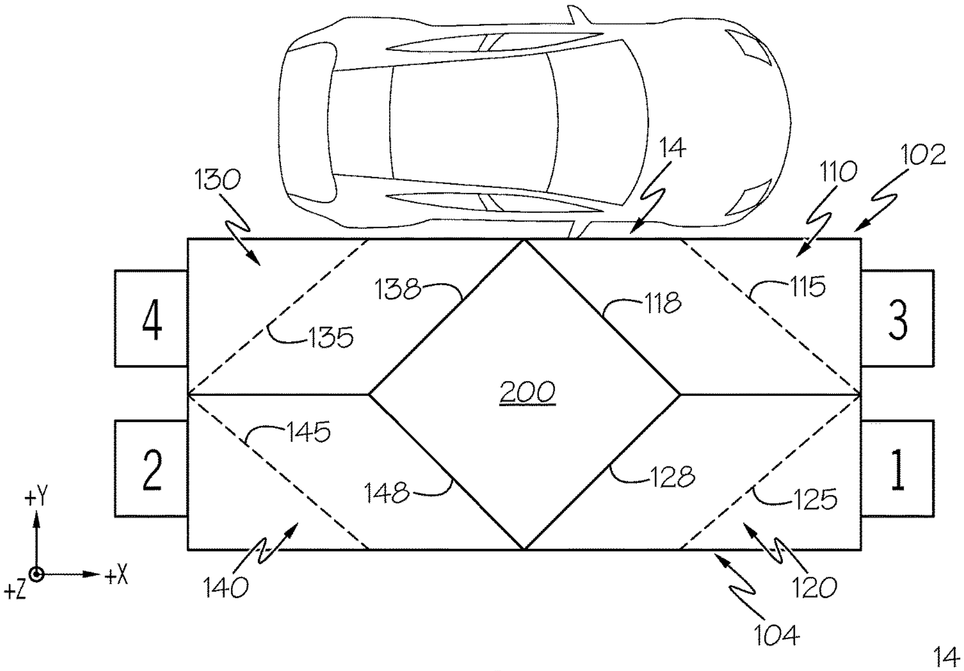
130
14
110
102
138
118
135
115
4
3
200
2
1
145
148
128
125
+Y
+X
+Z
140
120
104

FIG. 9

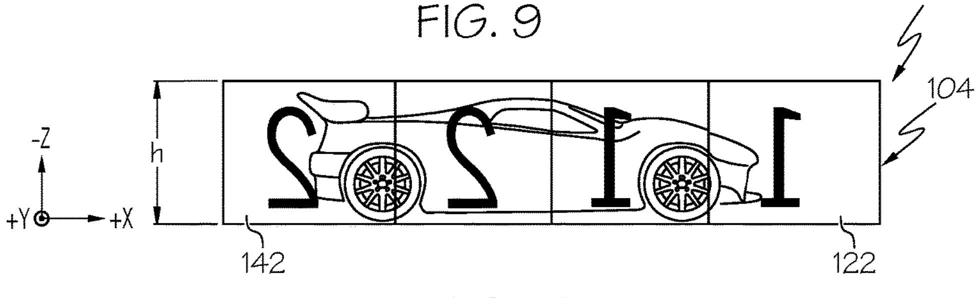
14
104
-Z
h
+Y
+X
2
2
1
1
142
122

FIG. 10

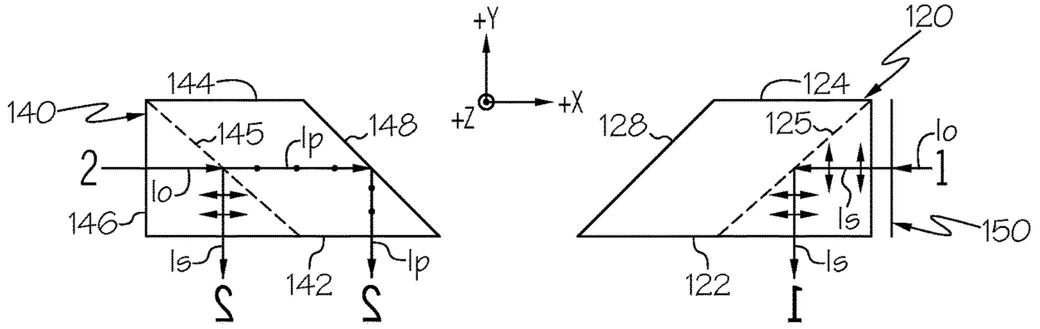
+Y
+Z
+X
140
144
145
148
Ip
2
Io
146
Is
142
Ip
2
2
120
124
128
125
Io
1
Is
150
122
Is
1

FIG. 11

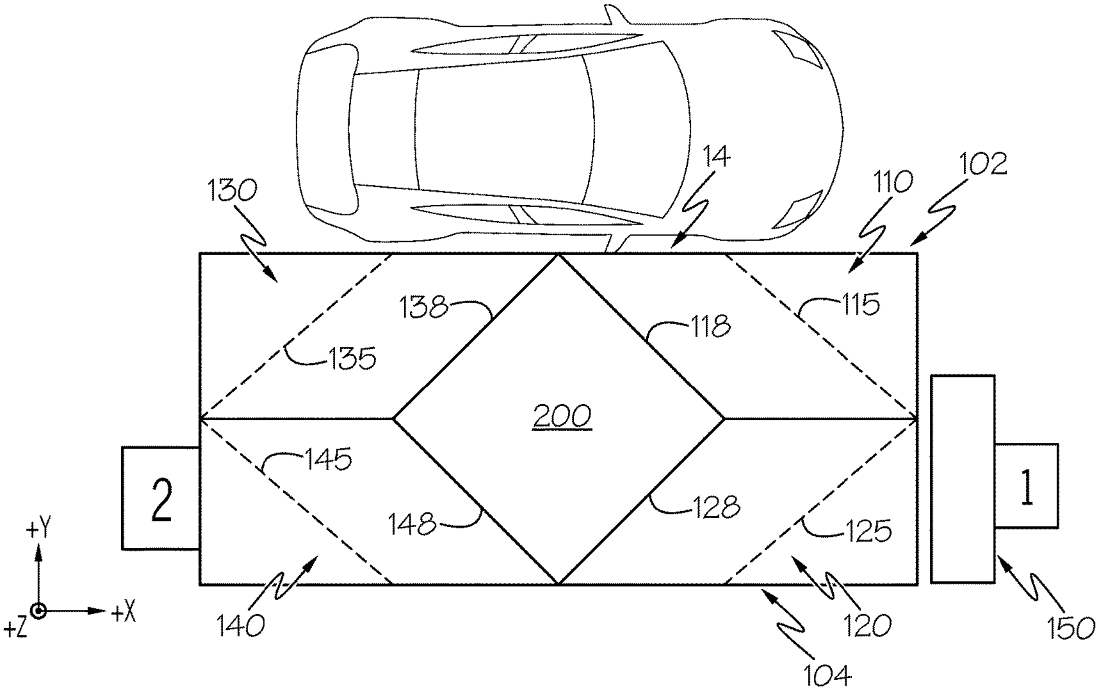
14
130
110
102
138
118
115
135
200
145
2
1
148
128
125
+Y
+Z
+X
140
120
150
104

FIG. 12

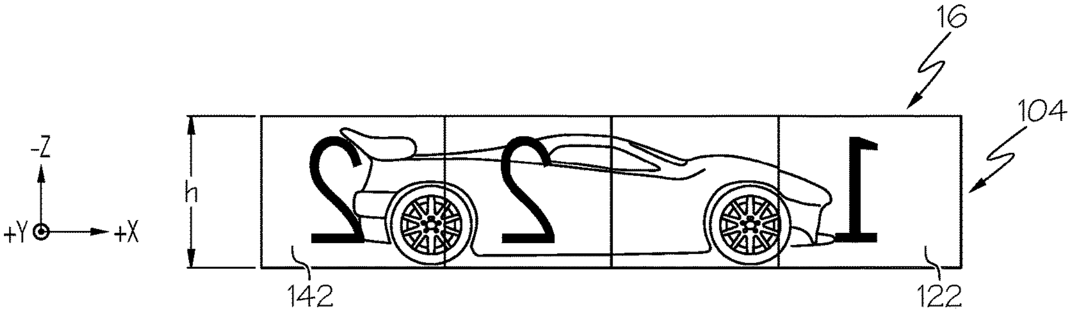
16
104
-Z
h
+Y
+X
2
2
1
142
122

FIG. 13

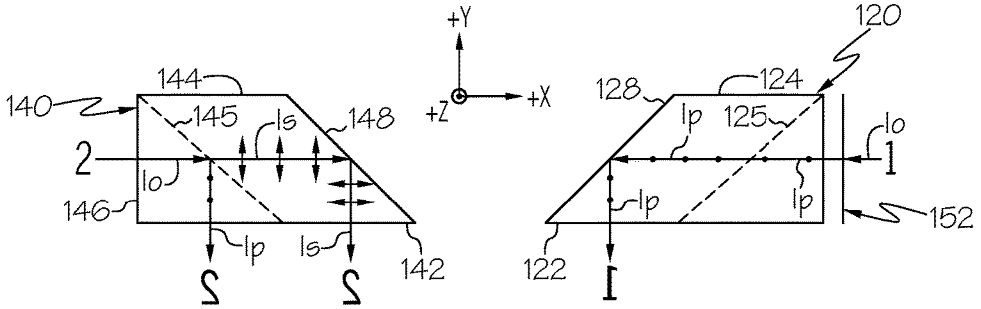
+Y
+Z
+X
120
140
144
145
Is
148
128
124
Ip
125
Io
2
Io
1
146
Ip
Is
142
122
Ip
Ip
152
2
2
1

FIG. 14

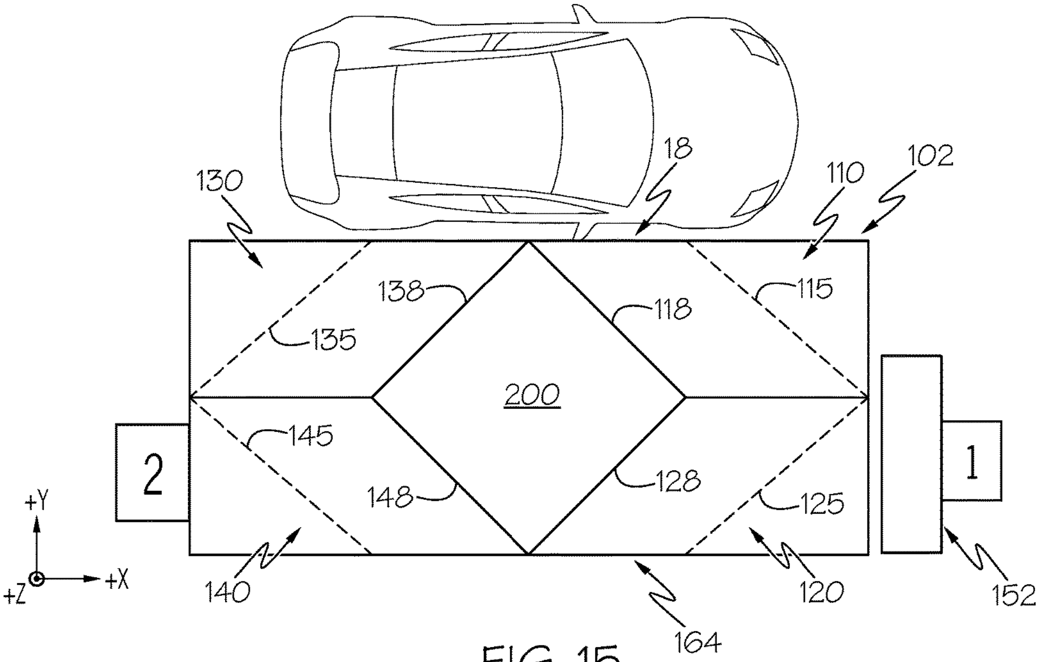
18
130
110
102
138
118
115
135
200
2
1
145
148
128
125
+Y
+Z
+X
140
120
152
164

FIG. 15

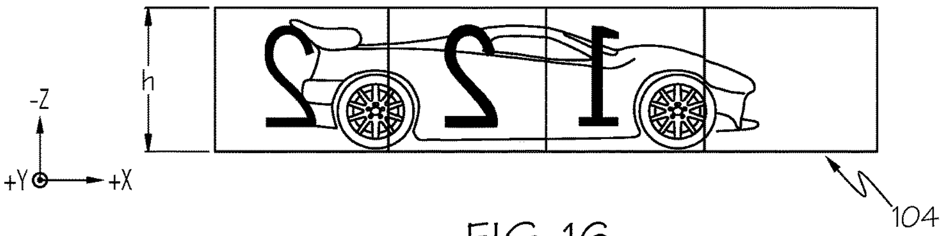
h
2
2
1
-Z
+Y
+X
104

FIG. 16

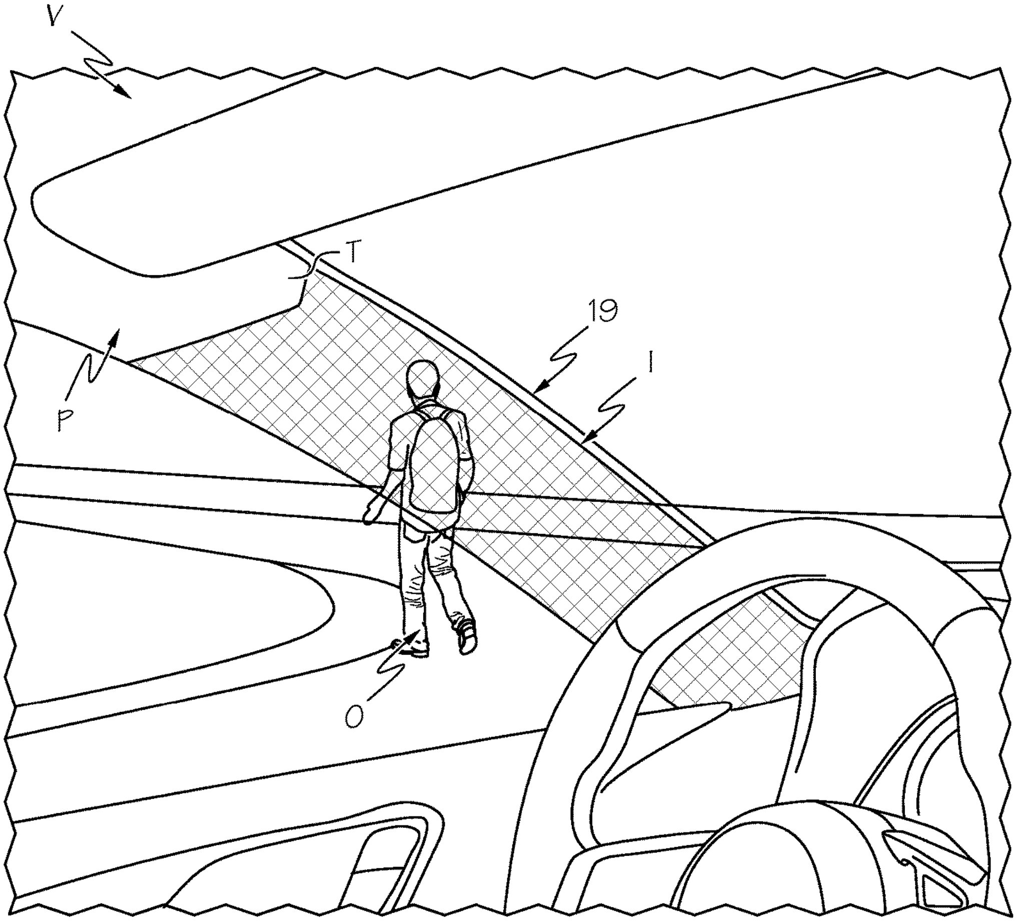

FIG. 17

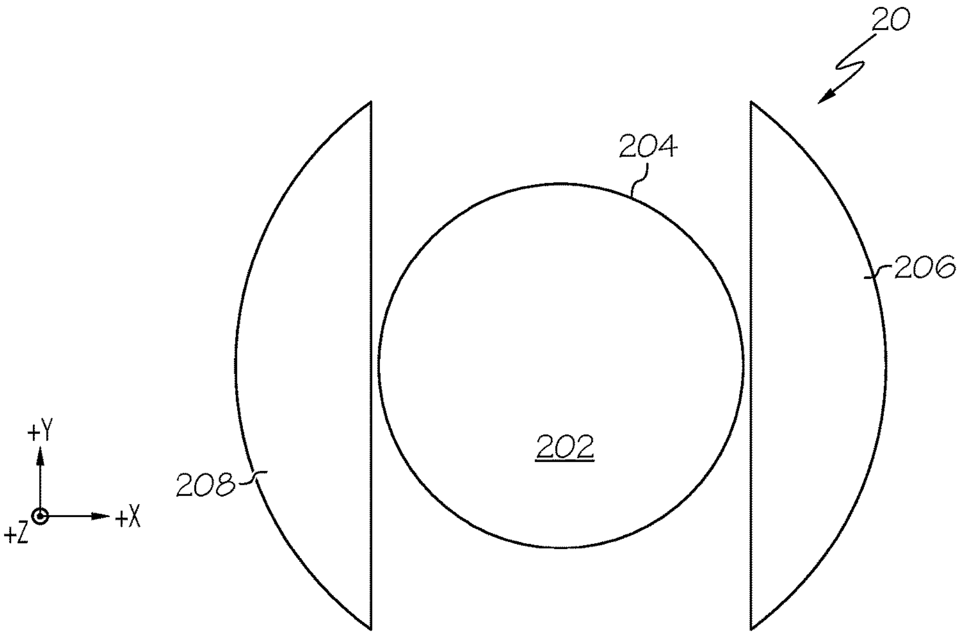
20
204
206
202
208
+Y
+X
+Z

FIG. 18

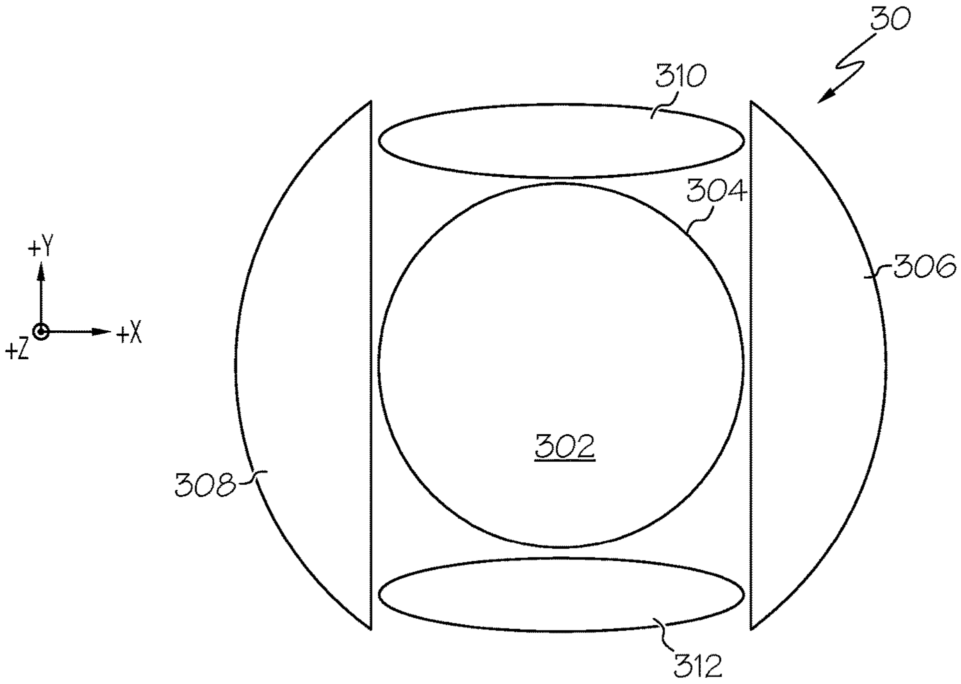
30
310
304
306
302
308
312
+Y
+X
+Z

FIG. 19

APPARATUSES AND METHODS FOR MAKING AN OBJECT APPEAR TRANSPARENT

RELATED APPLICATION

[0001] This application claims the benefit of U.S. Provisional Patent Application Ser. No. 62/291,920 filed Feb. 5, 2016, the entirety of which is incorporated herein by reference.

TECHNICAL FIELD

[0002] The present specification generally relates to apparatuses and methods for making an object appear transparent and, more specifically, to cloaking devices for pillars of vehicles and methods for making pillars of vehicles appear transparent.

BACKGROUND

[0003] Studies on cloaking devices that appear to make a pillar of a vehicle transparent have been published. Such studies disclose the use of metamaterials or the use of video cameras in combination with a display screen to allow an occupant of a vehicle to ostensibly "see" though the vehicle pillar, thereby reducing blind spots in the vehicle. However, metamaterials and video technology use complicated material designs and equipment.

[0004] Accordingly, a need exists for alternative devices that appear to make a pillar of a vehicle transparent.

SUMMARY

[0005] In one embodiment, a cloaking device includes cloaking region (CR) boundary planes oriented non-planar to each other and each of the CR boundary planes has an outward facing mirror surface and an inward facing opaque surface. The cloaking device includes a CR at least partially bounded by the inward facing opaque surfaces of the CR boundary planes. The cloaking device includes half mirrors spaced apart and generally parallel to the outward facing mirror surfaces such that a half mirror is spaced apart and generally parallel to each outward facing mirror surface. An article positioned within the CR is not visible outside of the CR and light reflected from an object on one side of the cloaking device is redirected around the CR and the article positioned within the CR, and forms an image of the object on another side of the cloaking device.

[0006] According to another embodiment, a cloaking device for cloaking an article of a vehicle includes a cloaking assembly with an object-side, an image-side and at least two CR boundary planes positioned at least partially between the object-side and the image-side. The at least two CR boundary planes are positioned non-planar to each other and each of the at least two CR boundary planes have an outward facing mirror surface and an inward facing opaque surface. A CR is bounded by the inward facing opaque surfaces of the at least two CR boundary planes and a vehicle article positioned within the CR is not visible from outside the CR. The cloaking assembly has at least two half mirrors positioned at least partially between the object-side and the image-side. Each of the at least two half mirrors are spaced apart and generally parallel to one of the outward facing mirror surfaces such that a half mirror is spaced apart and generally parallel to each outward facing mirror surface. Light from an object located on the object-side of the cloaking assembly, and obscured by the vehicle article positioned within the CR is redirected around the CR, and forms an image of the object on the image-side of the cloaking assembly such that the vehicle article appears transparent to an observer looking in a direction towards the object.

[0007] In embodiments, the cloaking assembly has four CR boundary planes and four half mirrors. The four CR boundary planes are oriented generally orthogonal to each other, the CR is bounded by the four CR boundary planes, and each of the half mirrors are spaced apart and generally parallel to one of the outward facing mirror surfaces such that a half mirror is spaced apart and generally parallel to each outward facing mirror surface. Each of the half mirrors may be selected from a p-polarization half mirror or an s-polarization half mirror. Light reflected from the object located on the object-side of the cloaking assembly passes through the cloaking assembly by reflecting off of a first outward facing mirror surface, a first half mirror oriented generally parallel to the first outward facing mirror surface, a second half mirror oriented generally orthogonal to the first half mirror and a second outward facing mirror surface oriented generally parallel to the second half mirror. Light that passes through the cloaking assembly forms an image of the object on the image-side of the cloaking assembly.

[0008] In embodiments, the cloaking device with the cloaking assembly has an end surface oriented non-planar to a half mirror. Light from an item located adjacent to the end surface passes through the cloaking assembly and forms an image of the item on the image side of the cloaking assembly.

[0009] These and additional features provided by the embodiments described herein will be more fully understood in view of the following detailed description in conjunction with the drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] The embodiments set forth in the drawings are illustrative and exemplary in nature and not intended to limit the subject matter defined by the claims. The following detailed description of the illustrative embodiments can be understood when read in conjunction with the following drawings, where like structure is indicated with like reference numerals and in which:

[0011] FIG. **1** schematically depicts a top view of a cloaking device according to one or more embodiments described and illustrated herein;

[0012] FIG. **2** schematically depicts top isolated views of four cloaking device components of the cloaking device of FIG. **1** according to one or more embodiments described and illustrated herein;

[0013] FIG. **3** schematically depicts a top perspective view of the cloaking device of FIG. **1** with a first object on one side of the cloaking device and a second object within a cloaking region (CR) of the cloaking device;

[0014] FIG. **4** schematically depicts a side view of the cloaking device of FIG. **3** with the first object on one side of the cloaking device and the second object within the CR of the cloaking device;

[0015] FIG. **5** schematically depicts a top view of a cloaking device according to one or more embodiments described and illustrated herein;

[0016] FIG. **6** schematically depicts a top perspective view of the cloaking device of FIG. **5** with a first object on one side of the cloaking device and a second object and a third object within a CR of the cloaking device;

[0017] FIG. **7** schematically depicts a side view of the cloaking device of FIG. **6** with the first object on one side of the cloaking device and the second object and the third object within the CR of the cloaking device;

[0018] FIG. **8** schematically depicts a top isolated view of two cloaking device components according to one or more embodiments described and illustrated herein;

[0019] FIG. **9** schematically depicts a top view of a cloaking device with four cloaking device components according to one or more embodiments described and illustrated herein;

[0020] FIG. **10** schematically depicts a side view of the cloaking device shown in FIG. **9**;

[0021] FIG. **11** schematically depicts a top isolated view of two cloaking device components according to one or more embodiments described and illustrated herein;

[0022] FIG. **12** schematically depicts a top view of a cloaking device with four cloaking device components according to one or more embodiments described and illustrated herein;

[0023] FIG. **13** schematically depicts a side view of the cloaking device of FIG. **12**;

[0024] FIG. **14** schematically depicts a top isolated view of two cloaking device components according to one or more embodiments described and illustrated herein;

[0025] FIG. **15** schematically depicts a top view of four cloaking device components according to one or more embodiments described and illustrated herein;

[0026] FIG. **16** schematically depicts a side view of the cloaking device of FIG. **15**;

[0027] FIG. **17** schematically depicts a side view of a cloaking device cloaking an A-pillar of a vehicle according to one or more embodiments described and illustrated herein;

[0028] FIG. **18** schematically depicts a top view of a cloaking device according to one or more embodiments described and illustrated herein; and

[0029] FIG. **19** schematically depicts a top view of a cloaking device with four cloaking device components according to one or more embodiments described and illustrated herein.

DETAILED DESCRIPTION

[0030] A cloaking device is provided. The cloaking device may have an object-side, an image-side and a cloaking region (CR) boundary plane having an outward facing mirror surface and an inward facing opaque surface. A CR is at least partially bounded by the inward facing opaque surface of the CR boundary plane and a half mirror is spaced apart and generally parallel to the outward facing mirror surface. The half mirror may be a p-polarization half mirror or an s-polarization half mirror. In embodiments, an article, e.g. an A-pillar of a vehicle, a B-pillar of a vehicle or a C-pillar of a vehicle, is positioned within the CR and light from an object located on the object-side of the cloaking device and obscured by the CR is redirected around the CR and the article and forms the image of the object on the image-side of the cloaking device such that the article appears transparent. Various embodiments of cloaking devices and methods for using the same will be described in further detail herein with specific reference to the appended drawings.

[0031] FIG. **1** generally depicts one embodiment of a cloaking device. The cloaking device includes a cloaking region (CR) that is at least partially bounded by at least two CR boundary planes that are non-planar to each other. Each of the two CR boundary planes has an outward facing mirror surface and an inward facing opaque surface. Spaced apart and oriented generally parallel to each of the CR boundary planes is a half mirror (e.g., a polarizing half mirror) that allows p-polarized light to pass through and s-polarized light to be reflected or, alternatively, allows s-polarized light to pass through and p-polarized light to be reflected. Referring to the coordinate system in the figure, portions of a target object are located directly behind the CR in the +Y direction (obscured portions of the object which are otherwise not visible through the CR) and portions of the target object are located outside or beyond the CR in the +Y direction (visible portions of the object). Light rays (light) from the obscured portions of the target object travel in the −Y direction and are incident on a first outward facing mirror surface. The first outward facing mirror surface reflects the incident light from the obscured portions of the target object in the +X direction to a first half mirror that is spaced apart and oriented generally parallel to the first outward facing mirror surface. One mode of polarized light (e.g., p-polarized light) from the light reflected from the first outward facing mirror surface passes through the first half mirror (continues traveling in the +X direction), while another mode of polarized light (e.g. s-polarized light) is reflected by the first half mirror in the −Y direction. The polarized light reflected in the −Y direction by the first half mirror is reflected into the −X direction by a second half mirror. The polarized light reflected into the −X direction by the second half mirror is reflected in the −Y direction by a second outward facing mirror surface that is spaced apart and oriented generally parallel to the second half mirror. An observer looking at the cloaking device in the +Y direction will see the reflection of the polarized light by the second outward facing mirror surface, i.e. polarized light originating from light from obscured portions of the target object will be visible to a human eye located on an opposite side of the CR from the obscured portions of the target object. Also, light reflected from an article positioned within the CR (i.e., the cloaked article) is not transmitted through the inward facing opaque surfaces and is not seen by the observer looking at the cloaking device in the +Y direction. That is, the cloaking device effectively redirects light from the obscured portions of the target object around the CR such that the light from the target object appears to pass through the cloaked article giving the visual impression that the cloaked article is not present. Light from the visible portions of the target object travel in the −Y direction and are incident on the first half mirror. One mode of the incident light (e.g., s-polarized light) is reflected in the +X direction by the first half mirror and another mode of the incident light (e.g. p-polarized light) is transmitted through the first half mirror and continues traveling in the −Y direction. The transmitted polarized light traveling in the −Y direction reaches the second half mirror and is transmitted through the second half mirror. An observer looking at the cloaking device in the +Y direction will see the transmitted polarized light traveling in the −Y direction, i.e. polarized light originating from light from visible portions of the target object will be visible to a human eye located on an opposite side of the cloaking device from the visible portions of the target object. Accordingly, an individual will see the entire target object, both obscured portions and visible portions, that is located on the opposite side of the CR (and thus on the opposite side of a cloaked article) giving the visual impression that the cloaked article is transparent.

[0032] Referring now to FIGS. **1**-**2**, embodiments of a cloaking device include a cloaking assembly **10** with four CR boundary planes **118**, **128**, **138**, **148**. The CR boundary planes **118**, **128**, **138**, **148** located adjacent to each other are non-planar to each other. In embodiments, opposing CR boundary planes **118**, **148** and **128**, **138** may be oriented parallel to each other and adjacent CR boundary planes **118**, **128**, **138**, **148** may be oriented orthogonal to adjacent CR boundary planes, e.g. CR boundary plane **118** is oriented orthogonal to adjacent CR boundary planes **138**, **128**, CR boundary plane **128** is oriented orthogonal to adjacent CR boundary planes **118**, **148**, etc. Each of the four CR boundary planes have an outward facing mirror surface **118***a*, **128***a*, **138***a*, **148***a*, respectively, and an inward facing opaque surface **118***b*, **128***b*, **138***b*, **148***b*, respectively. The outward facing mirror surfaces **118***a*, **128***a*, **138***a*, **148***a*, can be made from omnidirectional photonic crystals or mirrors such that approximately 100% (+/−10%) of light incident on the outward facing mirror surfaces is reflected from the outward facing mirror surfaces. Accordingly, the term “mirror surface” used herein refers to a surface that reflects approximately 100% (+/−10%) of all modes of light (e.g. s-polarized light and p-polarized light) incident on the mirror surface. In embodiments, the CR boundary planes **118**, **128**, **138**, **148** form a CR **200** bound at least partly by the inward facing opaque surfaces **118***b*, **128***b*, **138***b*, **148***b*. The four CR boundary planes **118**, **128**, **138** **148**, have a height ‘h’ (FIG. **4**) in the Z direction of the coordinate axes in the figures and light reflected or transmitted within the CR **200** does not pass through the inward facing opaque surfaces **118***b*, **128***b*, **138***b*, **148***b*. Accordingly, an article (e.g., a cloaked article) located within the CR **200** is not visible to an observer viewing the cloaking assembly **10** in the +Y direction.

[0033] Still referring to FIGS. **1** and **2**, spaced apart and oriented generally parallel (within +/−2°) to each of the four CR boundary planes **118**, **128**, **138** **148**, are half mirrors **115**, **125**, **135**, **145**, respectively. The half-mirrors **115**, **125**, **135**, **145** reflect a specific mode of visible light. Specifically, each of the half mirrors **115**, **125**, **135**, **145**, can be an s-polarizer half mirror or a p-polarizer half mirror. The half mirrors **115**, **125**, **135**, **145**, can be in the form of a diffraction grating or thin film polarizer that reflects the s-mode of visible light and allows the p-mode of visible light to pass through (a p-polarization diffraction grating or thin film), or in the alternative, reflects the p-mode of visible light and allows the s-mode of the visible light to pass through (an s-polarization diffraction grating or thin film). It should be appreciated that the half mirrors **115**, **125** are both s-polarizer half mirrors or p-polarizer half mirrors and the half mirrors **135**, **145** are both s-polarizer half mirrors or p-polarizer half mirrors, i.e. the half mirrors **115**, **125** may be s-polarizer mirrors and the half mirrors **135**, **145** may be p-polarizer half mirrors; the half mirrors **115**, **125** may be p-polarizer mirrors and the half mirrors **135**, **145** may be s-polarizer half mirrors; or all of the half mirrors **115**, **125**, **135**, **145** may be s-polarizer half mirrors or p-polarizer half mirrors.

[0034] The four CR boundary planes **118**, **128**, **138** **148**, and the corresponding half mirrors **115**, **125**, **135**, **145**, may be part of four cloaking assembly components **110**, **120**, **130**, **140**, respectively. Each cloaking assembly component **110**, **120**, **130**, **140**, has an outward facing surface **112**, **122**, **132**, **142**, respectively, an inward facing surface **114**, **124**, **134**, **144**, respectively, an outer end surface **116**, **126**, **136**, **146**, respectively, and an inner side surface defined by the inward facing opaque surface **118***b*, **128***b*, **138***b*, **148***b*, respectively. The outward facing surfaces **112**, **132** form an object-side **102** of the cloaking device and the outward facing surfaces **122**, **142** form an image-side **104** of the cloaking device. One or more optical filters **125***a*, **125***b* may be positioned on or adjacent to the outward facing surface **122** and one or more optical filters **145***a*, **145***b* may be positioned on or adjacent to the outward facing surface **142**. The optical filter **125***a* is positioned to the right (+X direction) of the half-mirror **125**, the optical filter **125***b* is positioned to the left (−X direction) of the half-mirror **125**, the optical filter **145***a* is positioned to the left (−X direction) of the half-mirror **145**, and the optical filter **145***b* is positioned to the right (+X direction) of the half-mirror **145**. The one or more optical filters **125***a*, **125***b*, **145***a*, **145***b*, may be quarter-wave plates that convert linearly polarized light into circularly polarized light, or in the alternative, the one or more optical filters **125***a*, **125***b*, **145***a*, **145***b*, may be a quarter-wave plate and a linear polarizer that convert linearly polarized light of one mode, e.g., s-mode or p-mode, into circularly polarized light, and then convert the circularly polarized light into a different linearly polarized mode, e.g., p-mode or s-mode, respectively.

[0035] The CR boundary planes **118**, **138**, are oriented at an angle θ_1 relative to the outward facing surfaces **112**, **132**, respectively, and the CR boundary planes **128**, **148** are oriented at an angle β_1 relative to the outward facing surfaces **122**, **142**, respectively. In embodiments, the angle θ_1 is equal to the angle β_1. In other embodiments, the angle θ_1 is not equal to the angle β_1. The half mirrors **115**, **135**, are oriented at an angle θ_2 relative to the outward facing surfaces **112**, **132**, respectively, and the half mirrors **125**, **145** are oriented at an angle θ_2 relative to the outward facing surfaces **122**, **142**, respectively. In embodiments, the angle θ_2 is equal to the angle β_2. In other embodiments, the angle θ_2 is not equal to the angle β_2. In embodiments, the angle θ_1 is equal to the angle θ_2 and the angle β_1 is equal to the angle β_2. For example, θ_1 and θ_2 can be generally equal to 45° (+/−1°), and β_1 and β_2 can be generally equal to 45° (+/−1°).

[0036] The four cloaking assembly components **110**, **120**, **130**, **140**, are transparent to incident light (both polarized and unpolarized) except for the half mirrors **115**, **125**, **135**, **145**, the outward facing mirror surfaces **118***a*, **128***a*, **138***a*, **148***a*, the inward facing opaque surfaces **118***b*, **128***b*, **138***b*, **148***b*, and optionally the outer end surfaces **116**, **126**, **136**, **146**. The four cloaking assembly components **110**, **120**, **130**, **140**, can be made from any suitable transparent material, for example a transparent glass or a transparent plastic. In the alternative, the four cloaking assembly components **110**, **120**, **130**, **140**, can be a structure, e.g. a frame, that holds the four CR boundary planes **118**, **128**, **138**, **148**, and the four half mirrors **115**, **125**, **135**, **145** in a desired orientation relative to each other with air or another gas present between the four CR boundary planes **118**, **128**, **138**, **148**, and the four half mirrors **115**, **125**, **135**, **145**. The frame may be a plurality of transparent glass or transparent plastic panes that hold the four CR boundary planes **118**, **128**, **138**, **148**, and the four half mirrors **115**, **125**, **135**, **145** in a desired orientation relative to each other with air or another gas present between the four CR boundary planes **118**, **128**, **138**, **148**, and the four half mirrors **115**, **125**, **135**, **145**. In embodiments, transparent glass or transparent plastic panes form the outward facing surfaces **112**, **122**, **132**, **142**, inward facing surfaces **114**, **124**, **134**, **144**, and outer end surfaces **116**, **126**, **136**, **146**, and hold the four CR boundary planes **118**, **128**, **138**, **148**, and the four half mirrors **115**, **125**, **135**, **145** in a desired orientation relative to each other with air or another gas present between the four CR boundary planes **118**, **128**, **138**, **148**, and the four half mirrors **115**, **125**, **135**, **145**.

[0037] The outer end surfaces **116**, **126**, **136**, **146**, may be transparent, have an absorber layer disposed thereon or be a half-mirror. In embodiments, the outer end surfaces **116**, **126**, **136**, **146**, are transparent. In other embodiments, the outer end surfaces **116**, **126**, **136**, **146**, have an absorber layer disposed thereon that absorbs incident light from within the four cloaking assembly components **110**, **120**, **130**, **140**, and/or from outside the four cloaking assembly components **110**, **120**, **130**, **140**. The absorber layer may be made from any suitable material that absorbs approximately 90% (+/−10%) of incident light, illustratively including dark colored surfaces made of glass, polymers, metals, ceramics, composites, etc. In other embodiments, the outer end surfaces **116**, **126**, **136**, **146**, are half-mirrors such that one mode of incident light is reflected (e.g., s-polarized light or p-polarized light) and another mode of incident light is transmitted through the half mirror (e.g., p-polarized light or s-polarized light, respectively). In other embodiments, one or more of the outer end surfaces **116**, **126**, **136**, **146**, is transparent and one or more of the outer end surfaces **116**, **126**, **136**, **146**, has an absorber layer disposed thereon. In other embodiments, one or more outer end surfaces **116**, **126**, **136**, **146**, is transparent and one or more of the outer end surfaces **116**, **126**, **136**, **146**, is a half mirror. In other embodiments, one or more of the outer end surfaces **116**, **126**, **136**, **146**, has an absorber layer disposed thereon and one or more of the outer end surfaces **116**, **126**, **136**, **146**, is a half mirror. In other embodiments, one or more of the outer end surfaces **116**, **126**, **136**, **146**, is transparent, one or more of the outer end surfaces **116**, **126**, **136**, **146**, has an absorber layer disposed thereon, and one or more of the outer end surfaces **116**, **126**, **136**, **146**, is a half mirror.

[0038] While FIGS. **1-2** illustrate four cloaking assembly components **110**, **120**, **130**, **140**, it should be appreciated that the four CR boundary planes **118**, **128**, **138**, **148**, and the four half mirrors **115**, **125**, **135**, **145** can be contained within any number of cloaking assembly components. For example, in embodiments, the cloaking assembly components **110** and **120** may be a single cloaking assembly component with no inward facing surface **114**, **124**, between the outward facing surfaces **112**, **122**, and the cloaking assembly components **130** and **140** may be a single cloaking assembly component with no inward facing surface **134**, **144**, between the outward facing surfaces **132**, **142**. In other embodiments, the cloaking assembly components **110** and **130** may be a single cloaking assembly component and the two cloaking assembly components **120** and **140** can be a single cloaking assembly component. In other embodiments, three of the four cloaking assembly components **110**, **120**, **130**, **140** may be a single cloaking assembly component assembled with a remaining one of the cloaking assembly components **110**, **120**, **130**, **140**. For example, the three cloaking assembly components **110**, **120**, **130** may be a single cloaking assembly component assembled with the remaining cloaking assembly component **140**. In other embodiments, the four cloaking assembly components **110**, **120**, **130**, **140** may be a single cloaking assembly component.

[0039] Still referring to FIGS. **1-2**, light 'lo' is reflected from an object 'O' (e.g., a target object) located on the object-side **102** of the cloaking device, travels in the −Y direction, and is incident on and transmitted through the outward facing surfaces **112**, **132**. The light lo reflected from the obscured portions of the object 'O' located directly behind the CR in the +Y direction is reflected by the outward facing mirror surfaces **118***a*, **138***a* in the +X and −X directions, respectively. The light lo reflected in the +X and −X directions is incident on the half mirrors **115**, **135** respectively. As illustrated in FIG. **1**, the half mirrors **115**, **135** are p-polarization half mirrors that reflect the s-mode (ls) of the light lo and allow the p-mode (lp) of the light lo to pass through. The s-polarized light ls reflected from the half mirrors **115**, **135** in the −Y direction is incident on the half mirrors **125**, **145**, respectively. The half mirrors **125**, **145** are also p-polarization mirrors that reflect the s-polarized light ls traveling in the −Y direction into the −X and +X directions, respectively. The s-polarized light ls reflected in in the −X and +X directions is incident on the outward facing mirror surfaces **128***a*, **148***a*, respectively, and the outward facing mirror surfaces **128***a*, **148***a* reflect the incident s-polarized light ls into the −Y direction. The s-polarized light reflected into the −Y direction by the outward facing mirror surfaces **128***a*, **148***a*, forms an image of the obscured portion of the target object located directly above the CR **200** in the +Y direction that is visible to an observer viewing the image-side **104** of the cloaking assembly **10**. In embodiments, the optical filter **125***b*, the optical filter **145***b*, or both optical filters **125***b*, **145***b* may be positioned on or adjacent to the outward facing surfaces **122**, **142**, respectively, as described above, and may convert the s-polarized light reflected into the −Y direction by the outward facing mirror surfaces **128***a*, **148***a*, into circularly polarized light or p-polarized light. For example, the optical filter **125***b* the optical filter **145***b*, or both optical filters **125***b*, **145***b* may be quarter wave plates that convert the s-polarized light reflected into the −Y direction by the outward facing mirror surfaces **128***a*, **148***a* into circularly polarized light (not shown) before forming the image of the obscured portion of the target object 'O' located directly above the CR **200** in the +Y direction. In the alternative, the optical filter **125***b*, the optical filter **145***b*, or both optical filters **125***b*, **145***b* may be a quarter-wave plate and a linear polarizer that converts the s-polarized light reflected into the −Y direction by the outward facing mirror surfaces **128***a*, **148***a* into circularly polarized light (not shown) and then converts the circularly polarized light into p-polarized light before forming the image of the obscured portion of the target object 'O' located directly above the CR **200** in the +Y direction.

[0040] Light lo reflected from the visible portions of the target object 'O' not located directly behind the CR in the +Y direction is incident on and transmitted through the outward facing surfaces **112**, **132**. The transmitted light lo is incident on the p-polarization half mirrors **115**, **135**. The half mirrors **115**, **135** reflect the s-mode is of the incident light lo in the +X and −X directions, respectively, and transmit the p-mode lp of the incident light lo in the −Y direction. The p-polarized light lp transmitted through the half mirrors **115**, **135** will be incident on and be transmitted through the p-polarization half mirrors **125**, **145**, respectively, in the –Y direction. The p-polarized light lp transmitted in the –Y direction by the half mirrors **125**, **145** forms an image of the visible portions of the target object O not located directly behind the CR **200** in the +Y direction that is visible to an observer viewing the cloaking assembly in the +Y direction. In embodiments, the optical filter **125***a*, the optical filter **145***a*, or both optical filters **125***a*, **145***a* may be positioned on or adjacent to the outward facing surfaces **122**, **142**, respectively, as described above, in order to convert the p-polarized light into circularly polarized light of s-polarized light. For example, the optical filter **125***a*, the optical filter **145***a* or both optical filters **125***a*, **145***a* may be quarter wave plates that convert the p-polarized light transmitted into the –Y direction by the half mirrors **125**, **145** into circularly polarized light (not shown) before forming the image of the visible portions of the target object 'O' not located behind above the CR **200** in the +Y direction. In the alternative, the optical filter **125***a*, the optical filter **145***a*, or both optical filters **125***a*, **145***a* may be a quarter-wave plate and a linear polarizer that convert the p-polarized light reflected into the –Y direction by the outward facing mirror surfaces **128***a*, **148***a* into circularly polarized light (not shown) and then convert the circularly polarized light into s-polarized light before forming the image of the visible portions of the target object 'O' not located directly above the CR **200** in the +Y direction. Accordingly, an image 'I' of the entire target object 'O' (both the obscured portion and the visible portions) is visible to the observer viewing the image-side **104** of the cloaking assembly **10**. While half mirrors **115**, **125**, **135**, **145** are described herein as p-polarization half mirrors, it should be appreciated that half mirrors **115**, **125**, **135**, **145** can be s-polarization half mirrors and the reflected and transmitted portions of the reflected light lo discussed above will be reversed. However, the human eye cannot distinguish between s-polarized light, p-polarized light or circularly polarized light and the image 'I' of the entire object 'O' will be visible to the human eye when an observer is viewing the image-side **104** of the cloaking assembly **10**. Also, it should be appreciated that the one or more of optical filters **125***a*, **125***b*, **145***a*, **145***b* may be positioned on or adjacent to outward facing surfaces **122**, **142** as described above such that only p-polarized light forms the image 'I', only s-polarized light forms the image 'I', only circularly polarized light forms the image 'I', only p-polarized light and circularly polarized light form the image 'I' or only s-polarized light and circularly polarized light form the image 'I'. Such manipulation of the polarized light reflected into the –Y direction by the outward facing mirror surfaces **128***a*, **128***a* and the polarized light transmitted in the –Y direction by the half mirrors **125**, **145** may assist an observer wearing polarized sunglasses in viewing the image-side **104** of the cloaking assembly **10**. Particularly, the one or more optical filters **125***a*, **125***b*, **145***a*, **145***b* ensure that polarized sunglasses that block p-polarized light or s-polarized light do not prevent an observer wearing such polarized sunglasses from seeing the entire target object 'O' when viewing the image-side **104** of the cloaking assembly **10**.

[0041] Referring now to FIGS. **1-4**, a top perspective view and a side view of a cloaking device according to embodiments (as discussed with respect to FIGS. **1-2**) are shown in FIGS. **3-4**, respectively. Specifically, FIG. **3** is a top perspective view of an article in the form of a column 'C' within the CR **200** of the cloaking assembly **10** and an automobile 'A' located behind the column C on the object-side **102** of the cloaking assembly **10** in the +Y direction. The column C has a height dimension in the Z direction (increasing height in the –Z direction) greater than the height h of the cloaking device. FIG. **4** is a side view from the +Y direction of the cloaking assembly **10** shown in FIG. **3** and shows the portion of the column C that is within the CR **200** is not visible and the automobile A located behind the column C in the +Y direction is visible to an observer viewing the cloaking assembly **10** in the +Y direction. Accordingly, the column C positioned within the CR **200** is not visible to an observer viewing the image-side **104** of the cloaking assembly **10** and an image of the entire automobile A (both the obscured portion and the visible portions) is visible to the observer viewing the image-side **104**. Although column C in FIGS. **3-4** is separate from the CR boundary planes **118**, **128**, **138**, **148**, i.e., column C is a separate object from the cloaking device **10**, it should be appreciated that column C may be structurally part of the cloaking device **10** and have an outer surface that provides or is equivalent to the CR boundary planes **118**, **128**, **138**, **148** with outward facing mirror surfaces **118***a*, **128***a*, **138***a*, **148***a*, respectively.

[0042] Referring now to FIGS. **1-5**, FIG. **5** shows embodiments of a cloaking assembly **12** with an enlarged CR **200**. Particularly, the cloaking assembly components **110**, **120**, **130**, **140** are arranged such that the cloaking assembly components **110**, **130** are spaced apart in the Y direction from the cloaking assembly components **120**, **140**, and a distance 'd' is between the inward facing surfaces **114**, **124**, and between the inward facing surfaces **134**, **144**. The area of the CR **200** in the X-Y plane is increased by the width 'w' times the distance d (w x d). The volume of the CR **200** is increased by the width w times the distance d times the height h (FIG. **4**) of the CR **200** (w x d x h). While FIG. **5** illustrates four cloaking assembly components **110**, **120**, **130**, **140**, it should be appreciated that the four CR boundary planes **118**, **128**, **138**, **148**, and the four half mirrors **115**, **125**, **135**, **145** can be contained within any number of cloaking assembly components. For example, in embodiments, the cloaking assembly components **110** and **130** may be a single cloaking assembly component and the two cloaking assembly components **120** and **140** can be a single cloaking assembly component.

[0043] As shown in the FIG. **5**, light lo is reflected from the target object O, travels in the –Y direction and is incident on and transmitted through the outward facing surfaces **112**, **132**. Light lo reflected from the obscured portions of the object located directly behind the CR in the +Y direction is reflected by the outward facing mirror surfaces **118***a*, **138***a* in the +X and –X direction, respectively. The reflected light lo in the +X and –X directions is incident on the half mirrors **115**, **135** respectively. The half mirrors **115**, **135** are p-polarization half mirrors and reflect the s-mode is of the light lo and transmit the p-mode lp of the light lo. The s-polarized light ls reflected by the half mirrors **115**, **135** in the –Y direction is incident on the half mirrors **125**, **145**, respectively. The half mirrors **125**, **145** are also p-polarization mirrors that reflect the s-polarized light ls traveling in the –Y direction into the –X and +X directions, respectively. The s-polarized light ls traveling in the –X and +X directions is incident on the outward facing mirror surfaces **128***a*, **148***a*, respectively, and the outward facing mirror surfaces **128***a*, **148***a* reflect the incident s-polarized light ls in the –Y direction. The s-polarized light reflected in the –Y direction by the outward facing mirror surfaces **128***a*, **148***a*, forms an image of the obscured portion of the object located behind the CR **200** in the +Y direction that is visible to an observer viewing the image-side **104** of the cloaking assembly **12**. Light lo reflected from the visible portions of the target object O not located behind the CR **200** in the +Y direction is incident on and transmitted through the outward facing surfaces **112**, **132**. The transmitted light lo is incident on the p-polarization half mirrors **115**, **135**, and the half mirrors **115**, **135** reflect the s-mode is of the incident light lo in the +X and −X directions, respectively, and transmit the p-mode lp of the incident light lo in the −Y direction. The p-polarized light lp transmitted through the half mirrors **115**, **135** in the −Y direction is incident on and transmitted through the p-polarization half mirrors **125**, **145**, respectively, in the −Y direction. The p-polarized light lp transmitted in the −Y direction through the half mirrors **125**, **145**, forms an image of the visible portions of the target object O not located behind the CR **200** in the +Y direction that is visible to an observer viewing the image side **104** of the cloaking assembly **12**. Accordingly, an image I of the entire target object O (both the obscured portion and the visible portions) is visible to an observer viewing the image-side **104** of the cloaking assembly **12**. It should be appreciated that half mirrors **115**, **125**, **135**, **145**, can be s-polarization half mirrors and the reflected and transmitted portions of the reflected light lo discussed above will be reversed. Also, it should be appreciated that the one or more of optical filters **125***a*, **125***b*, **145***a*, **145***b* may be included. However, the human eye cannot distinguish between s-polarized light, p-polarized light or circularly polarized light and the image I of the entire object O will be visible to an observer viewing the image-side **104** of the cloaking assembly **12**.

[0044] Referring now to FIGS. **1-7**, a top perspective view and a side view of a cloaking device according to embodiments (as discussed with respect to FIG. **5**) are shown in FIGS. **6-7**, respectively. Specifically, FIG. **6** is a top perspective view of a first article in the form of a column '**C1**' and a second article in the form of a column '**C2**' within the CR **200** of the cloaking assembly **12** and the automobile 'A' located behind the columns C**1**, C**2** on the object-side **102** of the cloaking assembly **12** in the +Y direction. The column C**1** has a first height dimension in the Z direction and the column C**2** has a second height dimension that is greater (in the −Z direction) than the first height dimension of column C**1**. The first height dimension of column C**1** and the second height dimension of column C**2** are both greater than the height h of the cloaking assembly **12**. FIG. **7** is a side view from the +Y direction of the cloaking assembly **12** shown in FIG. **6** and shows the image-side **104**. Portions of column C**1** and column C**2** within the CR **200** are not visible and automobile A located behind column C**1** and column C**2** in the +Y direction is visible to an observer viewing the image-side **104** of the cloaking assembly **12**. Although columns C**1**, C**2** in FIGS. **6-7** are separate from the CR boundary planes **118**, **128**, **138**, **148**, i.e., columns C**1**, C**2** are separate objects from the cloaking device **12**, it should be appreciated that columns C**1**, C**2** may be structurally part of the cloaking device **12** and have an outer surface that provides or is equivalent to the CR boundary planes **118**, **128**, **138**, **148** with outward facing mirror surfaces **118***a*, **128***a*, **138***a*, **148***a*, respectively.

[0045] Referring to FIG. **8**, embodiments of a cloaking device that can display additional images in conjunction with the images depicted in FIGS. **1**, **4**, **5** and **7** are shown. Particularly, FIG. **8** shows an item in the form of the number '1' (item 1) and an item in the form of the number '2' (item 2) located adjacent to the outer end surfaces **126**, **146**, respectively, in the +X, −X directions, respectively. Light lo reflected from item 1 and item 2 is incident on and transmitted through the outer end surfaces **126**, **146** in the −X, +X directions, respectively. The light lo transmitted through the outer end surfaces **126**, **146** is incident on the half mirrors **125**, **145**, respectively. The half mirrors **125**, **145** (p-polarization half mirrors) transmit the p-mode lp of the light lo in the −X, +X directions, respectively, and reflect the s-mode is of the light lo into the −Y direction. The transmitted p-polarized light lp in the −X, +X directions is incident on the outward facing mirror surfaces **128***a*, **148***a*, respectively. The outward facing mirror surfaces **128***a*, **148***a*, reflect the incident p-polarized light lp into the −Y direction. The s-polarized light ls reflected by the half mirror **125** in the −Y direction forms a reversed '1' image on the right hand side of the outward facing surface **122** and the s-polarized light ls reflected by the half mirror **145** in the −Y direction forms a reversed '2' image on the left hand side of the outer facing surface **142**. The p-polarized light lp reflected by the outward facing mirror surface **128***a* into the −Y direction forms a reversed '1' image on the left hand side of the outward facing surface **122** and the p-polarized light lp reflected by the outward facing mirror surface **148***a* into the −Y direction forms a reversed '2' image on the right hand side of the outer facing surface **142**. It should be appreciated that the display of additional images as depicted in FIG. **8** in conjunction with images as depicted in FIGS. **1**, **4**, **5** and **7** can be used as part of a heads-up display, a warning system (signal) that an object is located on the object side of the cloaking device, etc.

[0046] Referring to FIGS. **1-10**, embodiments of a cloaking device with a cloaking assembly **14** having items in the form of '1', '2', '3', '4' adjacent to the outer end surfaces **126**, **146**, **116**, **136**, respectively are shown in FIG. **9**. FIG. **10** is a side view of the cloaking assembly **14** illustrated in FIG. **9** with reversed images of item 1 and item 2, in conjunction with the image of the automobile A, visible on the image-side **104** of the cloaking assembly **14**. It should be appreciated that items 1, 2, 3, 4 depicted in FIGS. **9-10** are shown for illustrative purposes only, i.e. the items 1, 2 in FIG. **9** could be oriented in the Y-Z plane in order to provide reverse images of items 1, 2 on the image side **104** (X-Z plane) in FIG. **10**. It should also be appreciated that reverse images of item 3 and item 4 can be formed or provided on the object-side **102** of the cloaking assembly **14** in a similar manner as the reversed images of item 1 and item 2 are formed on the image-side **104**.

[0047] Referring to FIGS. **1-8** and **11**, embodiments of a cloaking device that can selectively display images in conjunction with the images depicted in FIGS. **1**, **4**, **5** and **7** are shown. Particularly, FIG. **11** shows an item in the form of the number '1' and an item in the form of the number '2' located adjacent to the outer end surfaces **126**, **146**, respectively, in the +X, −X directions, respectively. Light lo from item 1 is incident on an s-polarizer **150**. The s-polarizer **150** allows the s-mode is of the incident light lo to pass through in the −X direction. The s-polarized light ls is incident on and reflected by the p-polarization half mirror **125** into the Y direction and no light is transmitted through or beyond the p-polarization half mirror **125** in the −X direction. Accordingly, a reversed image of item 1 is formed only on the right hand side of the outward facing surface **122**. The light lo from item 2 is transmitted through the outer end surface **146** in the +X direction and is incident on the p-polarization half mirror **145**. The p-polarization half mirror **145** transmits the p-mode lp of the incident light lo in the +X direction and reflects the s-mode is of the incident light lo in the −Y direction. The p-polarized light lp in the +X direction is incident on the outward facing mirror surface **148***a*. The outer facing mirror surface **148***a* reflects the incident p-polarized light lp in the −Y direction. The s-polarized light ls reflected by the half mirror **145** in the −Y direction forms a reversed item 2 image on the left hand side of the outer facing surface **142**. The p-polarized light lp reflected by the outer facing mirror surface **148***a* into the −Y direction provides a reversed item 2 image on the right hand side of the outer facing surface **142**.

[0048] Referring to FIGS. **1-8** and **11-13**, embodiments of a cloaking device with a cloaking assembly **16** having items in the form of '1' and '2' adjacent to the outer end surfaces **126**, **146**, respectively, and the s-polarizer **150** adjacent the outer end surface **126**, are shown in FIG. **12**. FIG. **13** is a side view of the cloaking assembly **16** illustrated in FIG. **12** with the reversed image of item 1 visible on the right hand side of the outer facing surface **122** of the cloaking assembly component **120** and the reversed image of item 2 visible on the left hand side and right hand side of the outward facing surface **142** of the cloaking assembly components **140** shown in conjunction with the image of the automobile A. It should be appreciated that items 1, 2 depicted in FIG. **12** are shown for illustrative purposes only, i.e. the items 1, 2 in FIG. **12** could be oriented in the Y-Z plane in order to provide the reverse images of items 1, 2 on the image-side (**104**) (X-Z plane) shown in FIG. **13**.

[0049] Referring to FIGS. **1-8** and **14**, another embodiment of a cloaking device that can selectively display additional images in conjunction with the images depicted in FIGS. **1**, **4**, **5** and **7** is shown. Particularly, FIG. **14** shows an item in the form of the number '1' and an item in the form of the number '2' located adjacent to the outer end surfaces **126**, **146**, respectively, in the +X, −X directions, respectively. Light lo reflected from the item 1 is incident on a p-polarizer **152** which allows p-polarized light lp to pass through in the −X direction. The p-polarized light lp is incident on the p-polarization half mirror **125**. The incident p-polarized light lp in the −X direction is transmitted through the p-polarization half mirror **125** in the −X direction and there is no s-polarized light is to be reflected by the p-polarization half mirror **125** in the −Y direction. The p-polarized light lp transmitted through the p-polarization half mirror **125** is incident on the outward facing mirror surface **128***a* and the outward facing mirror **128***a* reflects the p-polarized light lp into the −Y direction. Accordingly, a reversed image of item 1 is formed only on the left hand side of the outward facing surface **122**. The light lo reflected from item 2 is transmitted through the outer end surface **146** in the +X direction and is incident on the p-polarization half mirror **145**. The p-polarization half mirror **145** transmits the p-mode lp of the incident light lo in the +X directions and reflects the s-mode ls of the incident light lo into the −Y direction. The p-polarized light lp in the +X is incident on the outer facing mirror surface **148***a*. The outer facing mirror surface **148***a* reflects the incident p-polarized light lp into the −Y direction. The s-polarized light ls reflected by the half mirror **145** in the −Y direction forms a reversed 2 on the left hand side of the outer facing surface **142** and the p-polarized light lp reflected by the outer facing mirror surface **148***a* in the −Y direction forms a reversed 2 image on the right hand side of the outer facing surface **142**.

[0050] Referring to FIGS. **1-8** and **14-16**, embodiments of a cloaking device with a cloaking assembly **18** having items 1 and items 2 adjacent to the outer end surfaces **126**, **146**, respectively, and the p-polarizer **152** adjacent to the outer end surface **126**, are shown in FIG. **15**. FIG. **16** is a side view of the cloaking assembly **18** illustrated in FIG. **15** with the single reversed image of item 1 visible on the left hand side of the outer facing surface **122** of the cloaking assembly component **120** and the reversed image of item 2 visible on the left hand side and right hand side of the outward facing surface **142** of the cloaking assembly components **140** in conjunction with the image of the automobile A. It should be appreciated that items 1, 2 depicted in FIG. **15** are shown for illustrative purposes only, i.e. the items 1, 2 in FIG. **15** could be oriented in the Y-Z plane in order to provide the reverse images of items 1, 2 on the image-side **104** (X-Z plane) shown in FIG. **16**.

[0051] Referring to FIGS. **1**, **2**, **5**, **8-17**, embodiments of an A-pillar of a vehicle being cloaked by a cloaking device are shown. Particularly, FIG. **17** shows a cloaking device **19** cloaking a portion of an A-pillar P of a vehicle V. A portion of the A-pillar P is positioned within a CR **200** (not shown) of the cloaking device **19** and a portion of the A-pillar P extends beyond the cloaking device and is covered with trim T. Illustrated outside of the vehicle V is a target object O in the form of pedestrian. A portion of the pedestrian O is visible through a side window of the vehicle V and a portion of the pedestrian is visible "through" the A-pillar P cloaked by the cloaking device **19**. The cloaking device **19** redirects light reflected from the pedestrian O around the A-pillar P positioned within the CR **200** of the cloaking device **19** and forms an image I of the pedestrian O on an image-side **104** of the cloaking device **19** that is visible to an occupant of the vehicle V looking towards the pedestrian O. Accordingly, light from the pedestrian O appears to pass through the A-pillar P and a blind spot typically created by the A-pillar P is not present. In embodiments, the A-pillar P itself serves as the CR **200**, i.e. the A-pillar P has an outer surface with one or more outer facing mirror surfaces that assist in redirecting light from the pedestrian) around the A-pillar P. It should be appreciated that cloaking of the A-pillar P with the cloaking device **19** and removing the blind spot produced by the A-pillar P is performed without the use of metamaterials, video images, cameras, sophisticated electronics, etc.

[0052] Referring to FIGS. **18-19**, alternative embodiments of cloaking devices are shown. Specifically, FIG. **18** shows a cloaking device **20** with a circular CR **202** having an outer mirror surface **204** with half mirrors **206**, **208** around the outer mirror surface **204** having a circular shape. In embodiments, the CR **202** is provided by an object to be cloaked itself, e.g., an A-pillar, B-pillar, C-pillar, D-pillar, etc., that has an out outer mirror surface **204**. The cloaking device shown in FIG. **18** can increase the field of view for an observer viewing the cloaking device **20** in the +Y direction. FIG. **19** shows a cloaking device **30** with a circular CR **302** having an outer mirror surface **304** with half mirrors **306**, **308** and optical components **310**, **312** such as lenses. In embodiments, the CR **302** is provided by an object to be cloaked itself, e.g., an A-pillar, B-pillar, C-pillar, D-pillar, etc., that has an out outer mirror surface **304**. The cloaking device shown in FIG. **19** allows for an external view from an object located behind the CR **302** (+Y direction) to be manipulated to increase a field of view or to normalize a distorted image.

[0053] The cloaking devices described herein may be used to cloak vehicle articles such as a vehicle A-pillar, B-pillar, C-pillar, D-pillar, etc., and remove a "blind spot" caused by the vehicle article. It will be understood that the term "blind spot" as used herein refers to an area around the vehicle that cannot be directly observed by a driver while operating the vehicle. The terms "object," "article," and "item" may interchangeably refer to a visual object or image (2D or 3D) that reflects light or transmits light and the team "light from" may refer to "light reflected from" or "light transmitted from." The terms "generally," "approximately," and "about" may be utilized herein to represent the inherent degree of uncertainty that may be attributed to any quantitative comparison, value, measurement, or other representation. These terms are also utilized herein to represent the degree by which a quantitative representation may vary from a stated reference without resulting in a change in the basic function of the subject matter at issue.

[0054] While particular embodiments have been illustrated and described herein, it should be understood that various other changes and modifications may be made without departing from the spirit and scope of the claimed subject matter. Moreover, although various aspects of the claimed subject matter have been described herein, such aspects need not be utilized in combination. It is therefore intended that the appended claims cover all such changes and modifications that are within the scope of the claimed subject matter.

What is claimed is:

1. A cloaking device comprising:
- an object-side and an image-side;
- a cloaking region (CR) boundary plane having an outward facing mirror surface and an inward facing opaque surface;
- a CR at least partially bounded by the inward facing opaque surface of the CR boundary plane;
- a half mirror spaced apart and generally parallel to the outward facing mirror surface;
- wherein light from an object located on the object-side of the cloaking device and obscured by the CR is redirected around the CR to form an image of the object on the image-side of the cloaking device such that the light from the object appears to pass through the CR.

2. The cloaking device of claim **1**, wherein the half mirror is selected from a p-polarization half mirror and an s-polarization half mirror.

3. The cloaking device of claim **1**, further comprising an article positioned within the CR, wherein the light from the object located on the object-side of the cloaking device and obscured by the CR is redirected around the CR and the article to form the image of the object on the image-side of the cloaking device such that the article appears transparent.

4. The cloaking device of claim **3**, wherein the article is selected from the group consisting of a vehicle A-pillar, a vehicle B-pillar and a vehicle C-pillar.

5. The cloaking device of claim **1**, further comprising:
- at least two CR boundary planes positioned non-planar to each other, each of the at least two CR boundary planes having an outward facing mirror surface and an inward facing opaque surface, the CR at least partially bounded by the inward facing opaque surfaces of the at least two CR boundary planes;
- at least one quarter-wave plate positioned on or adjacent to the image side; and
- at least two half mirrors spaced apart and generally parallel to the outward facing mirror surfaces such that a half mirror is spaced apart and generally parallel to each outward facing mirror surface;
- wherein the light from the object located on the object-side of the cloaking device and obscured by the CR is redirected around the CR by the at least two outward facing mirror surfaces and the at least two half mirrors, and the image of the object is formed on the image-side of the cloaking device such that the light from the object located on the object-side of the cloaking device appears to pass through the CR.

6. The cloaking device of claim **5**, wherein the at least two CR boundary planes are positioned orthogonal to each other.

7. The cloaking device of claim **6**, wherein the light from the object on the object-side of the cloaking device is reflected by a first outward facing mirror surface, a first half mirror positioned parallel to the first outward facing mirror surface, a second half mirror positioned orthogonal to the first half mirror, a second outward facing mirror surface positioned parallel to the second half mirror, and transmitted through the at least one quarter-wave plate before forming the image of the object on the image-side of the cloaking device.

8. The cloaking device of claim **7**, further comprising a vehicle article positioned within the CR, wherein the light from the object located on the object-side of the cloaking device and obscured by the CR is redirected around the CR and the vehicle article to form the image of the object on the image-side of the cloaking device and the vehicle article appears transparent.

9. The cloaking device of claim **5**, further comprising an end side, wherein light from an item positioned adjacent to the end side forms an image of the item on the image-side of the cloaking device.

10. A cloaking device for cloaking an article of a vehicle comprising:
- a cloaking assembly comprising:
  - an object-side and an image-side;
  - at least two CR boundary planes located at least partially between the object-side and the image-side, the at least two CR boundary planes positioned non-planar to each other and each of the at least two CR boundary planes having an outward facing mirror surface and an inward facing opaque surface;
  - a CR bounded by the inward facing opaque surfaces of the at least two CR boundary planes;
  - at least two half mirrors located at least partially between the object-side and the image-side, each of the at least two half mirrors spaced apart and generally parallel to one of the outward facing mirror surfaces such that a half mirror is spaced apart and generally parallel to each outward facing mirror surface;
  - at least one quarter-wave plate positioned on or adjacent to the image side;
  - and
  - a vehicle article positioned within the CR;

wherein light from an object located on the object-side of the cloaking assembly and obscured by the CR is redirected around the CR and forms an image of the object on the image-side of the cloaking assembly such that the vehicle article appears to be transparent.

11. The cloaking device of claim **10**, wherein the vehicle article is selected from the group consisting of a vehicle A-pillar, a vehicle B-pillar and a vehicle C-pillar.

12. The cloaking device of claim **10**, wherein the light from the object is reflected by a first outward facing mirror surface, a first half mirror oriented generally parallel to the first outward facing mirror surface, a second half mirror oriented generally orthogonal to the first half mirror and a second outward facing mirror surface oriented generally parallel to the second half mirror and transmitted through the at least one quarter-wave plate before forming the image on the image-side of the cloaking assembly.

13. The cloaking device of claim **12**, further comprising an end surface, wherein light from an item located adjacent to the end surface passes through the cloaking assembly and forms an image of the item on the image-side of the cloaking assembly.

14. The cloaking device of claim **13**, wherein a p-mode or an s-mode of the light from the item is reflected from the second half mirror before forming the image of the item on the image-side of the cloaking assembly.

15. The cloaking device of claim **13**, wherein a p-mode or an s-mode of the light from the item is transmitted through the second half mirror and reflected by the second outward facing mirror surface before forming the image of the item on the image-side of the cloaking assembly.

16. A cloaking device comprising:

four CR boundary planes positioned non-planar to each other, each of the four CR boundary planes having an outward facing mirror surface and an inward facing opaque surface;

a CR bounded by the inward facing opaque surfaces of the four CR boundary planes;

an article located at least partially within the CR;

four half mirrors spaced apart and generally parallel to the outward facing mirror surfaces such that a half mirror is spaced apart and generally parallel to each outward facing mirror surface;

wherein light from an object on an object-side of the cloaking device is redirected around the CR and appears as an image on an image-side of the cloaking device and is visible to an observer viewing the image-side of the cloaking device and the article within the CR is not visible to the observer viewing the image-side of the cloaking device such that the article within the CR appears transparent to the observer viewing the image-side of the cloaking device.

17. The cloaking device of claim **16**, wherein the light from the object is reflected by a first outward facing mirror surface, a first half mirror oriented generally parallel to the first outward facing mirror surface, a second half mirror oriented generally orthogonal to the first half mirror and a second outward facing mirror surface oriented generally parallel to the second half mirror before forming the image on the image-side of the cloaking device.

18. The cloaking device of claim **17**, wherein the article is selected from at least one of a vehicle A-pillar, a vehicle B-pillar and a vehicle C-pillar.

19. The cloaking device of claim **18**, further comprising an end side, wherein light from an item located adjacent to the end side is reflected by at least one of the second half mirror and the second outward facing mirror surface and forms an image of the item on the image-side of the cloaking device.

20. The cloaking device of claim **19**, wherein the image of the item is selectively formed on the image-side of the cloaking device.

* * * * *

Burn things out, mess things up—that's how you learn.

Make: Electronics

Learn by
Discovery

Charles Platt

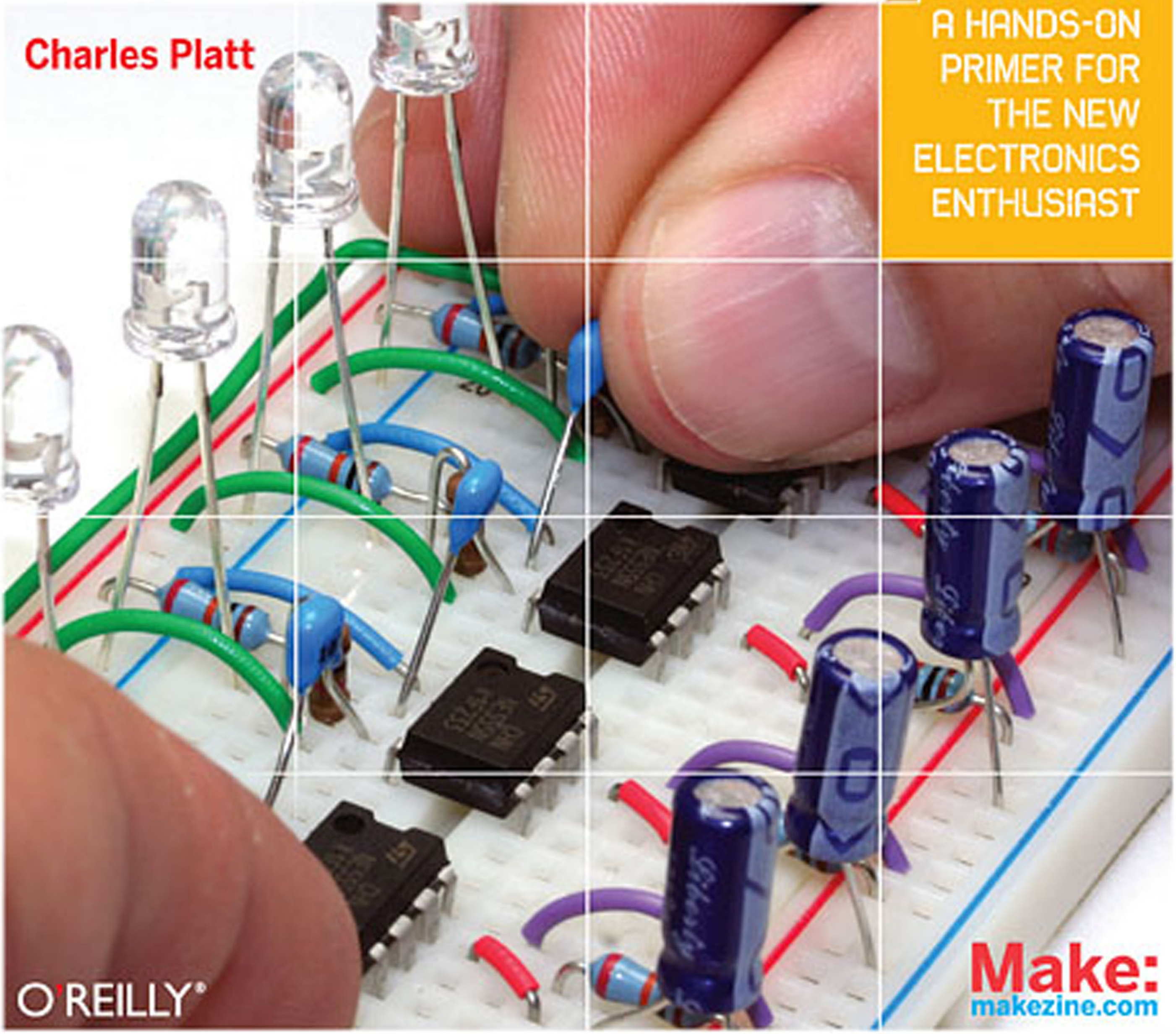

A HANDS-ON
PRIMER FOR
THE NEW
ELECTRONICS
ENTHUSIAST

O'REILLY

Make:
makezine.com

Make: Electronics

Make: Electronics

Learning by Discovery

Charles Platt

with photographs and illustrations by the author

O'REILLY®
Beijing • Cambridge • Farnham • Köln • Sebastopol • Taipei • Tokyo

Make: Electronics

by Charles Platt

Copyright © Helpful Corporation. All rights reserved.
Printed in Canada.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly Media books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*my.safaribooksonline.com*). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

Editors: Dale Dougherty and Brian Jepson

Development Editor: Gareth Branwyn

Production Editor: Rachel Monaghan

Technical Editor: Andrew "Bunnie" Huang

Copyeditor: Nancy Kotary

Proofreader: Nancy Reinhardt

Indexer: Julie Hawks

Cover Designer: Mark Paglietti

Interior Designer: Ron Bilodeau

Illustrator/Photographer: Charles Platt

Cover Photographer: Marc de Vinck

Print History:

December 2009: First Edition.

The O'Reilly logo is a registered trademark of O'Reilly Media, Inc. Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and O'Reilly Media, Inc., was aware of a trademark claim, the designations have been printed in caps or initial caps.

While every precaution has been taken in the preparation of this book, the publisher and author assume no responsibility for errors or omissions, or for damages resulting from the use of the information contained herein.

ISBN: 978-0-596-15374-8
[TI]

For my dearest Erico

Contents

Preface ix

1. Experiencing Electricity 1
Shopping List: Experiments 1 Through 5 1
Experiment 1: Taste the Power! 5
Experiment 2: Let's Abuse a Battery! 9
Experiment 3: Your First Circuit 13
Experiment 4: Varying the Voltage 18
Experiment 5: Let's Make a Battery 32

2. Switching Basics and More 39
Shopping List: Experiments 6 Through 11 39
Experiment 6: Very Simple Switching 43
Experiment 7: Relay-Driven LEDs 55
Experiment 8: A Relay Oscillator 60
Experiment 9: Time and Capacitors 68
Experiment 10: Transistor Switching 73
Experiment 11: A Modular Project 82

3. Getting Somewhat More Serious 95
Shopping List: Experiments 12 Through 15 95
Experiment 12: Joining Two Wires Together 104
Experiment 13: Broil an LED 114
Experiment 14: A Pulsing Glow 117
Experiment 15: Intrusion Alarm Revisited 127

4. Chips, Ahoy! 147
Shopping List: Experiments 16 Through 24 147
Experiment 16: Emitting a Pulse 153
Experiment 17: Set Your Tone 162
Experiment 18: Reaction Timer 170
Experiment 19: Learning Logic 181
Experiment 20: A Powerful Combination 197
Experiment 21: Race to Place 205
Experiment 22: Flipping and Bouncing 211
Experiment 23: Nice Dice 214
Experiment 24: Intrusion Alarm Completed 223

5. What Next? 227
Shopping List: Experiments 25 Through 36 228
Customizing Your Work Area 228
Reference Sources 233
Experiment 25: Magnetism 236
Experiment 26: Tabletop Power Generation 239
Experiment 27: Loudspeaker Destruction 242
Experiment 28: Making a Coil React 246
Experiment 29: Filtering Frequencies 249
Experiment 30: Fuzz 257
Experiment 31: One Radio, No Solder, No Power 262
Experiment 32: A Little Robot Cart 268
Experiment 33: Moving in Steps 284
Experiment 34: Hardware Meets Software 293
Experiment 35: Checking the Real World 306
Experiment 36: The Lock, Revisited 311
In Closing 317

Appendix. Online Retail Sources and Manufacturers 319

Index 325

Preface

How to Have Fun with This Book

Everyone uses electronic devices, but most of us don't really know what goes on inside them.

Of course, you may feel that you don't need to know. If you can drive a car without understanding the workings of an internal combustion engine, presumably you can use an iPod without knowing anything about integrated circuits. However, understanding some basics about electricity and electronics can be worthwhile for three reasons:

- By learning how technology works, you become better able to control your world instead of being controlled by it. When you run into problems, you can solve them instead of feeling frustrated by them.
- Learning about electronics can be fun—so long as you approach the process in the right way. The tools are relatively cheap, you can do all the work on a tabletop, and it doesn't consume a lot of time (unless you want it to).
- Knowledge of electronics can enhance your value as an employee or perhaps even lead to a whole new career.

Learning by Discovery

Most introductory guides begin with definitions and facts, and gradually get to the point where you can follow instructions to build a simple circuit.

This book works the other way around. I want you to start putting components together right away. After you see what happens, you'll figure out what's going on. I believe this process of *learning by discovery* creates a more powerful and lasting experience.

Stay Within the Limits!

Although I believe that everything suggested in this book is safe, I'm assuming that you will stay within the limits that I suggest. Please always follow the instructions and pay attention to the warnings, denoted by the icon you see here. If you go beyond the limits, you will expose yourself to unnecessary risks.

Learning by discovery occurs in serious research, when scientists notice an unusual phenomenon that cannot be explained by current theory, and they start to investigate it in an effort to explain it. This may ultimately lead to a better understanding of the world.

We're going to be doing the same thing, although obviously on a much less ambitious level.

Along the way, you will make some mistakes. This is good. Mistakes are the best of all learning processes. I want you to burn things out and mess things up, because this is how you learn the limits of components and materials. Since we'll be using low voltages, there'll be no chance of electrocution, and so long as you limit the flow of current in the ways I'll describe, there will be no risk of burning your fingers or starting fires.

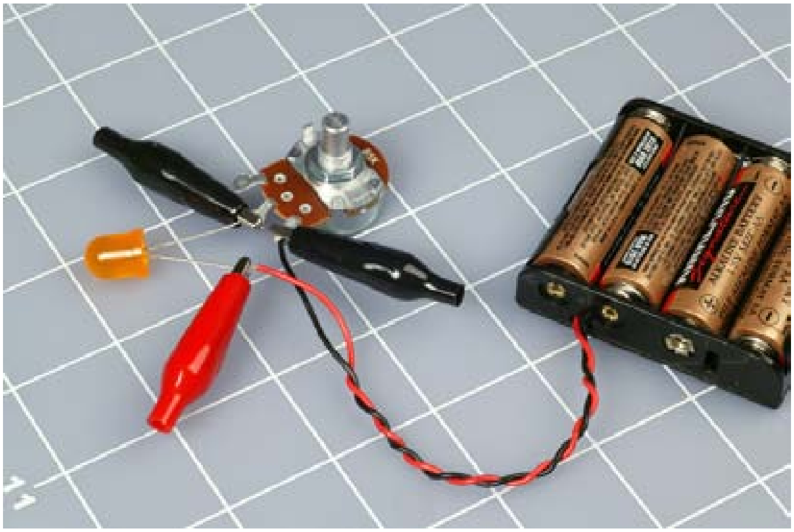

Figure P-1. *Learning by discovery allows you to start building simple circuits right away, using a handful of cheap components, a few batteries, and some alligator clips.*

How Hard Will It Be?

I assume that you're beginning with no prior knowledge of electronics. So, the first few experiments will be ultra-simple, and you won't even use solder or prototyping boards to build a circuit. You'll be holding wires together with alligator clips.

Very quickly, though, you'll be experimenting with transistors, and by the end of Chapter 2, you will have a working circuit that has useful applications.

I don't believe that hobby electronics has to be difficult to understand. Of course, if you want to study electronics more formally and do your own circuit design, this can be challenging. But in this book, the tools and supplies will be inexpensive, the objectives will be clearly defined, and the only math you'll need will be addition, subtraction, multiplication, division, and the ability to move decimal points from one position to another.

Moving Through This Book

Basically there are two ways to present information in a book of this kind: in tutorials and in reference sections. I'm going to use both of these methods. You'll find the tutorials in sections headed as follows:

- Shopping Lists
- Using Tools
- Experiments

You'll find reference sections under the following headings:

- Fundamentals
- Theory
- Background

How you use the sections is up to you. You can skip many of the reference sections and come back to them later. But if you skip many of the tutorials, this book won't be of much use to you. Learning by discovery means that you absolutely, positively have to do some hands-on work, and this in turn means that you have to buy some basic components and play with them. You will gain very little by merely imagining that you are doing this.

It's easy and inexpensive to buy what you need. In almost any urban or suburban area in the United States, chances are you live near a store that sells electronic components and some basic tools to work with them. I am referring, of course, to RadioShack franchises. Some Shacks have more components than others, but almost all of them have the basics that you'll need.

You can also visit auto supply stores such as AutoZone and Pep Boys for basics such as hookup wire, fuses, and switches, while stores such as Ace Hardware, Home Depot, and Lowe's will sell you tools.

If you prefer to buy via mail order, you can easily find everything you need by searching online. In each section of the book, I'll include the URLs of the most popular supply sources, and you'll find a complete list of URLs in the appendix.

Fundamentals

Mail-ordering components and tools

Here are the primary mail-order sources that I use myself online:

http://www.radioshack.com

RadioShack, a.k.a. The Shack. For tools and components. Not always the cheapest, but the site is easy and convenient, and some of the tools are exactly what you need.

http://www.mouser.com

Mouser Electronics.

http://www.digikey.com

Digi-Key Corporation.

http://www.newark.com

Newark.

Mouser, Digi-Key, and Newark are all good sources for components, usually requiring no minimum quantities.

http://www.allelectronics.com

All Electronics Corporation. A narrower range of components, but specifically aimed at the hobbyist, with kits available.

http://www.ebay.com

You can find surplus parts and bargains here, but you may have to try several eBay Stores to get what you want. Those based in Hong Kong are often very cheap, and I've found that they are reliable.

http://www.mcmaster.com

McMaster-Carr. Especially useful for high-quality tools.

Lowe's and Home Depot also allow you to shop online.

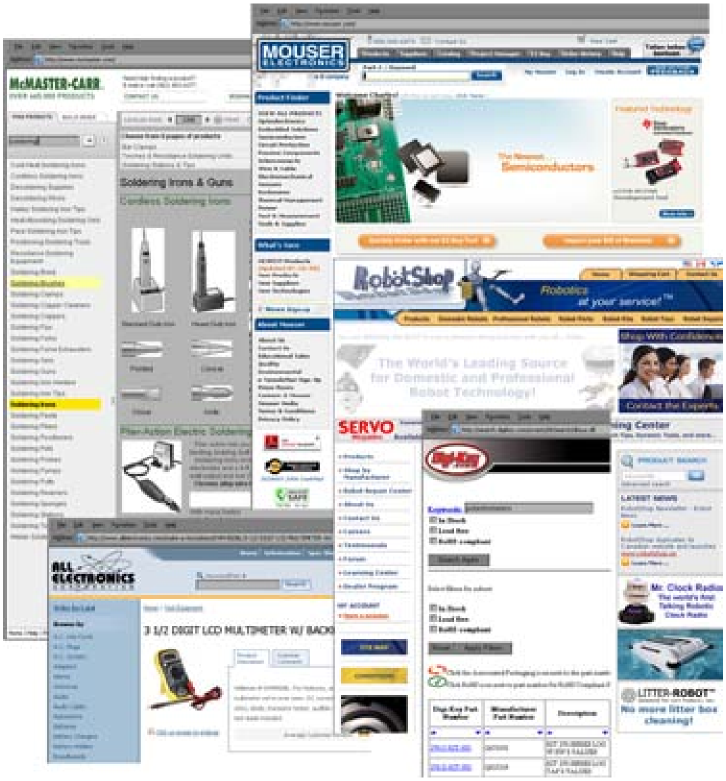
McMASTER-CARR
MOUSER ELECTRONICS
Soldering Irons & Guns
Cordless Soldering Irons
RobotShop
Robotics at your service!
The World's Leading Source for Domestic and Professional Robot Technology!
SERVO
Digi-Key
ALL ELECTRONICS
3 1/2 DIGIT LCD MULTIMETER W/ BACK
LITTER-ROBOT
Mr. Clock Radio

Figure P-2. *You'll find no shortage of parts, tools, kits, and gadgets online.*

Companion Kits

Maker Shed (*www.makershed.com*) offers a number of *Make: Electronics* companion kits, both toolkits and bundles of the various components used in the book's experiments. This is a simple, convenient, and cost-effective way of getting all the tools and materials you need to do the projects in this book.

Comments and Questions

Please address comments and questions concerning this book to the publisher:

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472
800-998-9938 (in the United States or Canada)
707-829-0515 (international or local)
707-829-0104 (fax)

We have a web page for this book, where we list errata, examples, larger versions of the book's figures, and any additional information. You can access this page at:

http://oreilly.com/catalog/9780596153748

To comment or ask technical questions about this book, send email to:

bookquestions@oreilly.com

For more information about our books, conferences, Resource Centers, and the O'Reilly Network, see our website at:

http://oreilly.com

Safari® Books Online

Safari Books Online is an on-demand digital library that lets you easily search over 7,500 technology and creative reference books and videos to find the answers you need quickly.

With a subscription, you can read any page and watch any video from our library online. Read books on your cell phone and mobile devices. Access new titles before they are available for print, and get exclusive access to manuscripts in development and post feedback for the authors. Copy and paste code samples, organize your favorites, download chapters, bookmark key sections, create notes, print out pages, and benefit from tons of other time-saving features.

O'Reilly Media has uploaded this book to the Safari Books Online service. To have full digital access to this book and others on similar topics from O'Reilly and other publishers, sign up for free at *http://my.safaribooksonline.com*.

Experiencing Electricity

1

I want you to get a taste for electricity—literally!—in the first experiment. This first chapter of the book will show you:

- How to understand and measure electricity and resistance
- How to handle and connect components without overloading, damaging, or destroying them

Even if you have some prior knowledge of electronics, you should try these experiments before you venture on to the rest of the book.

IN THIS CHAPTER

Shopping List: Experiments 1 Through 5
Experiment 1: Taste the Power!
Experiment 2: Let's Abuse a Battery!
Experiment 3: Your First Circuit
Experiment 4: Varying the Voltage
Experiment 5: Let's Make a Battery

Maker Shed (www.makershed.com) *has put together a series of* Make: Electronics *companion kits. These include all of the tools and components used in book's experiments. This is a quick, simple, and cost-effective way of getting everything you need to complete the projects in this book.*

Shopping List: Experiments 1 Through 5

If you want to limit your number of shopping trips or online purchases, look ahead in the book for additional shopping lists, and combine them to make one bulk purchase.

In this first chapter, I will give you part numbers and sources for every tool and component that we'll be using. Subsequently, I won't expect you to need such specific information, because you will have gained experience searching for items on your own.

Tools

Small pliers

RadioShack Kronus 4.5-inch, part number 64-2953 or Xcelite 4-inch mini long-nose pliers, model L4G.

Or similar. See Figures 1-1 through 1-3. Look for these tools in hardware stores and the sources listed in the preface. The brand is unimportant. After you use them for a while, you'll develop your own preferences. In particular, you have to decide whether you like spring-loaded handles. If you decide you don't, you'll need a second pair of pliers to pull the springs out of the first.

Wire cutters

RadioShack Kronus 4.5-inch, part number 64-2951, or Stanley 7-inch model 84-108.

Or similar. Use them for cutting copper wire, not harder metals (Figure 1-4).

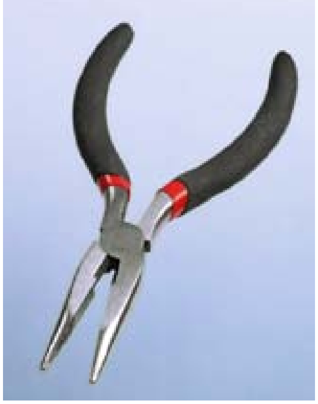

Figure 1-1. *Generic long-nosed pliers are your most fundamental tool for gripping, bending, and picking things up after you drop them.*

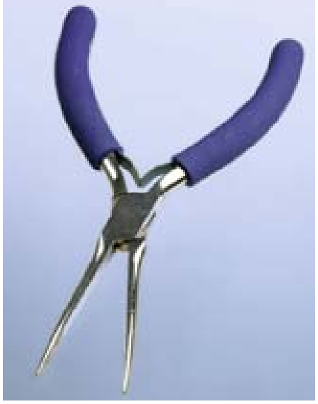

Figure 1-2. *Longer-nosed pliers: these are useful for reaching into tiny spaces.*

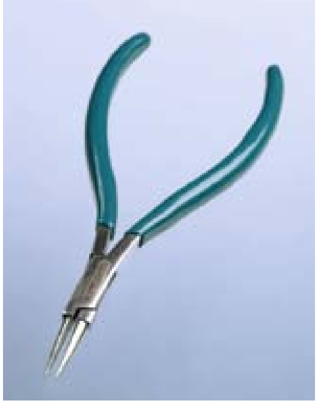

Figure 1-3. *Sharp-pointed pliers are designed for making jewelry, but are also useful for grabbing tiny components.*

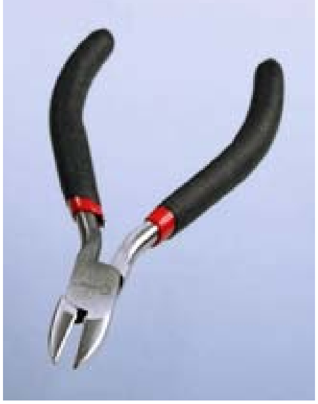

Figure 1-4. *Wire cutters, sometimes known as side cutters, are essential.*

Multimeter

Extech model EX410 or BK Precision model 2704-B or Amprobe model 5XP-A.

Or similar. Because electricity is invisible, we need a tool to visualize the pressure and flow, and a meter is the only way. A cheap meter will be sufficient for your initial experiments. If you buy online, try to check customer reviews, because reliability may be a problem for cheap meters. You can shop around for retailers offering the best price. Don't forget to search on eBay.

The meter must be digital—don't get the old-fashioned analog kind with a needle that moves across a set of printed scales. This book assumes that you are looking at a digital display.

I suggest that you do not buy an autoranging meter. "Autoranging" sounds useful—for example, when you want to check a 9-volt battery, the meter figures out for itself that you are not trying to measure hundreds of volts, nor fractions of a volt. The trouble is that this can trick you into making errors. What if the battery is almost dead? Then you may be measuring a fraction of a volt without realizing it. The only indication will be an easily overlooked "m" for "millivolts" beside the large numerals of the meter display.

On a manual-ranging meter, you select the range, and if the source that you are measuring is outside of that range, the meter tells you that you made an error. I prefer this. I also get impatient with the time it takes for the autoranging feature to figure out the appropriate range each time I make a measurement. But it's a matter of personal preference. See Figures 1-5 through 1-7 for some examples of multimeters.

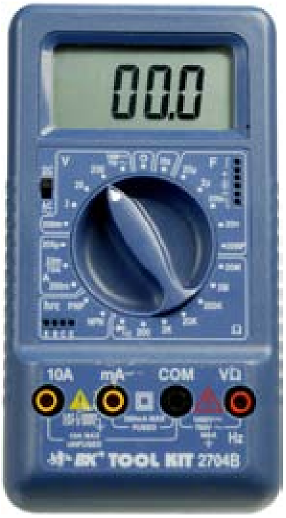

Figure 1-5. *You can see by the wear and tear that this is my own favorite meter. It has all the necessary basic features and can also measure capacitance (the F section, for Farads). It can also check transistors. You have to choose the ranges manually.*

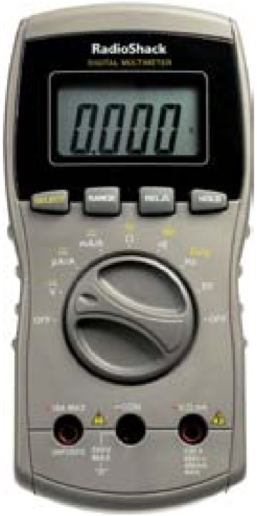

Figure 1-6. *Mid-priced RadioShack meter, which has the basic features; however, the dual purpose for each dial position, selected with the SELECT button, may be confusing. This is an autoranging meter.*

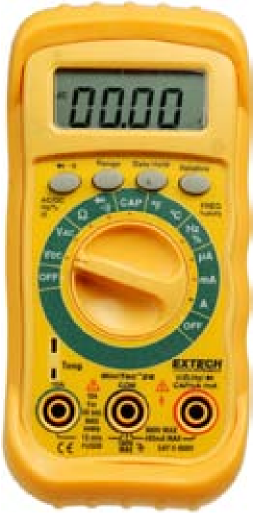

Figure 1-7. *An autoranging meter from Extech offers basic functions, plus a temperature probe, which may be useful to check whether components such as power supplies are running unduly hot.*

Supplies

Batteries

9-volt battery. Quantity: 1.

AA batteries, 1.5 volts each. Quantity: 6.

The batteries should be disposable alkaline, the cheapest available, because we may destroy some of them. You should *absolutely not* use rechargeable batteries in Experiments 1 and 2.

Battery holders and connectors

Snap connector for 9-volt battery, with wires attached (Figure 1-8). Quantity: 1. RadioShack part number 270-325 or similar. Any snap connector that has wires attached will do.

Battery holder for single AA cell, with wires attached (Figure 1-9). Quantity: 1. RadioShack part number 270-401 or Mouser.com catalog number 12BH410-GR, or similar; any single-battery holder that has thin wires attached will do.

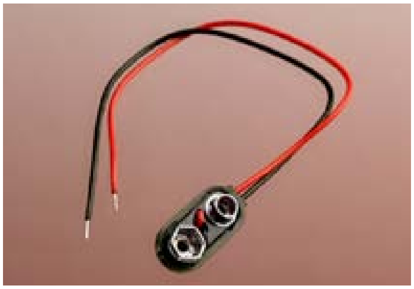

Figure 1-8. *Snap connector for a 9-volt battery.*

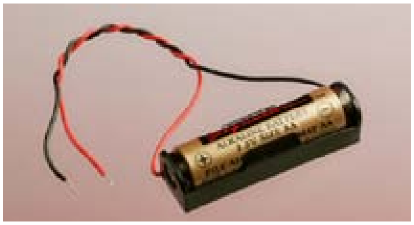

Figure 1-9. *Single AA-sized battery carrier with wires.*

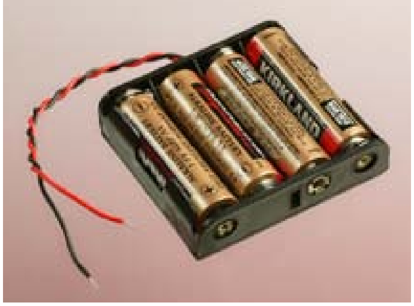

Figure 1-10. *Battery carrier for four AA cells, to be installed in series, delivering 6 volts.*

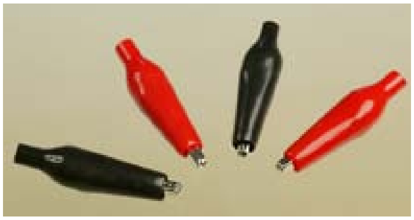

Figure 1-11. *Alligator clips inside vinyl sheaths, which reduce the chance of accidental short circuits.*

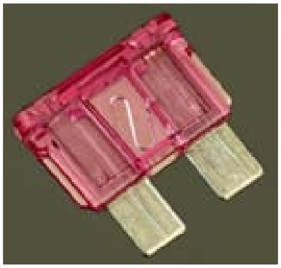

Figure 1-12. *A 3-amp fuse intended primarily for automotive use, shown here larger than actual size.*

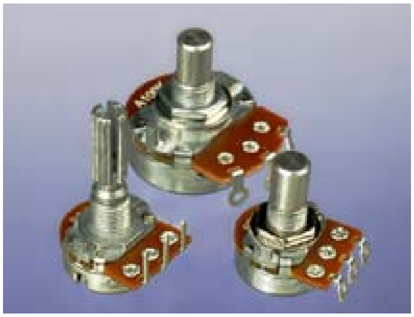

Figure 1-13. *Potentiometers come in many shapes and sizes, with different lengths of shafts intended for different types of knobs. For our purposes, any style will do, but the larger-sized ones are easier to play with.*

Battery holder for four AA cells, with wires attached (Figure 1-10). Quantity: 1. All Electronics catalog number BH-342 or RadioShack part 270-391 or similar.

Alligator clips

Vinyl-insulated. Quantity: at least 6. All Electronics catalog number ALG-28 or RadioShack part number 270-1545 or similar (Figure 1-11).

Components

You may not know what some of these items are, or what they do. Just look for the part numbers and descriptions, and match them with the photographs shown here. Very quickly, in the learning by discovery process, all will be revealed.

Fuses

Automotive-style, mini-blade type, 3 amps. Quantity: 3. RadioShack part number 270-1089, or Bussmann part ATM-3, available from automotive parts suppliers such as AutoZone (Figure 1-12).

Or similar. A blade-type fuse is easier to grip with alligator clips than a round cartridge fuse.

Potentiometers

Panel-mount, single-turn, 2K linear, 0.1 watt minimum. Quantity: 2. Alpha part RV170F-10-15R1-B23 or BI Technologies part P160KNPD-2QC25B2K, from Mouser.com or other component suppliers (Figure 1-13).

Or similar. The "watt" rating tells you how much power this component can handle. You don't need more than 0.5 watts.

Resistors

Assortment 1/4-watt minimum, various values but must include 470 ohms, 1K, and 2K or 2.2K. Quantity: at least 100. RadioShack part number 271-312.

Or search eBay for "resistor assorted."

Light-emitting diodes (LEDs)

Any size or color (Figures 1-14 and 1-15). Quantity: 10. RadioShack part number 276-1622 or All Spectrum Electronics part K/LED1 from Mouser.com.

Or similar. Just about any LEDs will do for these first experiments.

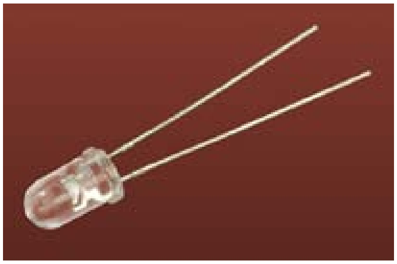

Figure 1-14. *Typical 5-mm diameter light-emitting diode (LED).*

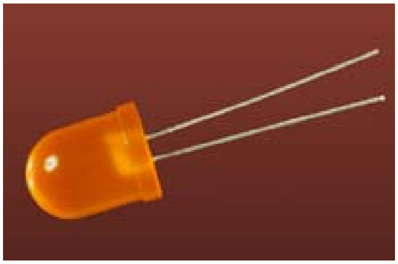

Figure 1-15. *Jumbo-sized LED (1 cm diameter) is not necessarily brighter or more expensive. For most of the experiments in this book, buy whatever LEDs you like the look of.*

Experiment 1: Taste the Power!

Can you taste electricity? Maybe not, but it feels as if you can.

You will need:

- 9-volt battery
- Snap connector for battery terminals
- Multimeter

No More Than 9 Volts

A 9-volt battery won't hurt you. But do not try this experiment with a higher-voltage battery or a larger battery that can deliver more current. Also, if you have metal braces on your teeth, be very careful not to touch them with the battery.

Procedure

Moisten your tongue and touch the tip of it to the metal terminals of a 9-volt battery. The sudden sharp tingle that you feel is caused by electricity flowing from one terminal of the battery (Figure 1-16), through the moisture on and in your tongue, to the other terminal. Because the skin of your tongue is very thin (it's actually a mucus membrane) and the nerves are close to the surface, you can feel the electricity very easily.

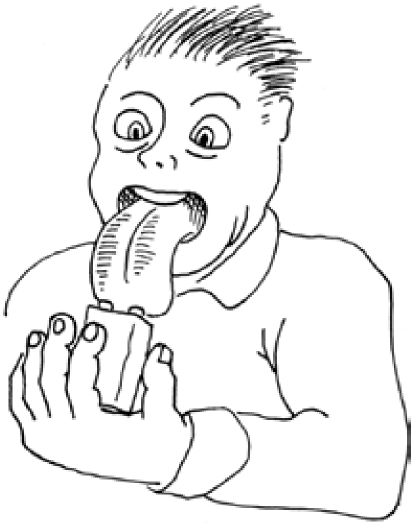

Figure 1-16. *Step 1 in the process of learning by discovery: the 9-volt tongue test.*

Now stick out your tongue, dry the tip of it very thoroughly with a tissue, and repeat the experiment without allowing your tongue to become moist again. You should feel less of a tingle.

What's happening here? We're going to need a meter to find out.

Tools

Setting up your meter

Check the instructions that came with the meter to find out whether you have to install a battery in it, or whether a battery is preinstalled.

Most meters have removable wires, known as *leads* (pronounced "leeds"). Most meters also have three sockets on the front, the leftmost one usually being reserved to measure high electrical currents (flows of electricity). We can ignore that one for now.

The leads will probably be black and red. The black wire plugs into a socket labeled "COM" or "Common." Plug the red one into the socket labeled "V" or "volts." See Figures 1-17 through 1-20.

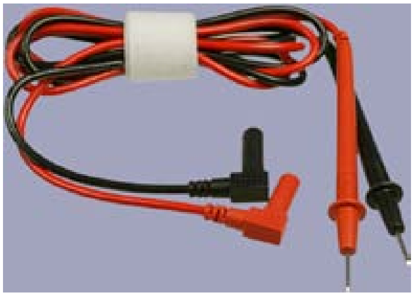

Figure 1-17. *The black lead plugs into the Common (COM) socket, and the red lead plugs into the red socket that's almost always on the righthand side of a multimeter.*

The other ends of the leads terminate in metal spikes known as *probes*, which you will be touching to components when you want to make electrical measurements. The probes detect electricity; they don't emit it in significant quantities. Therefore, they cannot hurt you unless you poke yourself with their sharp ends.

If your meter doesn't do autoranging, each position on the dial will have a number beside it. This number means "no higher than." For instance if you want to check a 6-volt battery, and one position on the voltage section of the dial is numbered 2 and the next position is numbered 20, position 2 means "no higher than 2 volts." You have to go to the next position, which means "no higher than 20 volts."

If you make a mistake and try to measure something inappropriate, the meter will show you an error message such as "E" or "L." Turn the dial and try again.

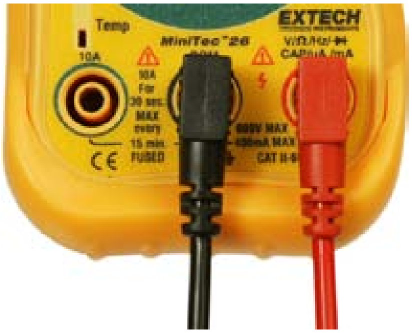
Figure 1-18

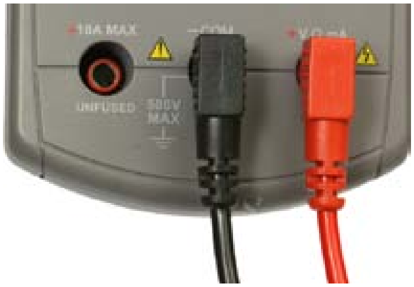
Figure 1-19

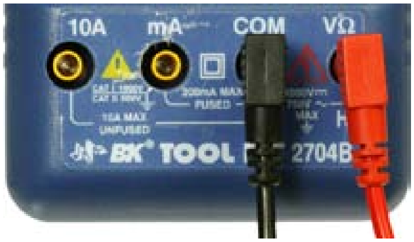
Figure 1-20. *To measure resistance and voltage, plug the black lead into the Common socket and the red lead into the Volts socket. Almost all meters have a separate socket where you must plug the red lead when you measure large currents in amps, but we'll be dealing with this later.*

FUNDAMENTALS

Ohms

We measure distance in miles or kilometers, weight in pounds or kilograms, temperature in Fahrenheit or Centigrade—and electrical resistance in ohms. The ohm is an international unit.

The Greek omega symbol (Ω) is used to indicate ohms, as shown in Figures 1-21 and 1-22. Letter K (or alternatively, KΩ) means a kilohm, which is 1,000 ohms. Letter M (or MΩ) means a megohm, which is 1,000,000 ohms.

A material that has very high resistance to electricity is known as an *insulator*. Most plastics, including the colored sheaths around wires, are insulators.

A material with very low resistance is a *conductor*. Metals such as copper, aluminum, silver, and gold are excellent conductors.

Number of ohms	Usually expressed as	Abbreviated as
1,000 ohms	1 kilohm	1KΩ or 1K
10,000 ohms	10 kilohms	10KΩ or 10K
100,000 ohms	100 kilohms	100KΩ or 100K
1,000,000 ohms	1 megohm	1MΩ or 1M
10,000,000 ohms	10 megohms	10MΩ or 10M

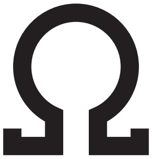
Figure 1-21. *The omega symbol is used internationally to indicate resistance on ohms.*

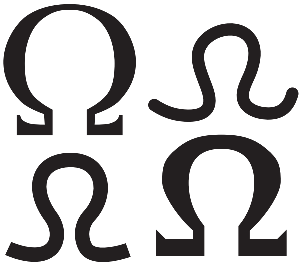
Figure 1-22. *You'll find it printed or written in a wide variety of styles.*

Procedure

We're going to use the meter to discover the electrical resistance of your tongue. First, set your meter to measure resistance. If it has autoranging, look to see whether it is displaying a K, meaning kilohms, or M, meaning megohms. If you have to set the range manually, begin with no less than 100,000 ohms (100K). See Figures 1-23 through 1-25.

Touch the probes to your tongue, about an inch apart. Note the reading, which should be around 50K. Now put aside the probes, stick out your tongue, and use a tissue to dry it very carefully and thoroughly. Without allowing your tongue to become moist again, repeat the test, and the reading should be higher. Finally, press the probes against the skin of your hand or arm: you may get no reading at all, until you moisten your skin.

> *When your skin is moist (for instance, if you perspire), its electrical resistance decreases. This principle is used in lie detectors, because someone who knowingly tells a lie, under conditions of stress, tends to perspire.*

A 9-volt battery contains chemicals that liberate electrons (particles of electricity), which want to flow from one terminal to the other as a result of a chemical reaction inside it. Think of the cells inside a battery as being like two water tanks—one of them full, the other empty. If they are connected with a pipe, water flows between them until their levels are equal. Figure 1-26 may help you visualize this. Similarly, when you open up an electrical pathway between the two sides of a battery, electrons flow between them, even if the pathway consists only of the moisture on your tongue.

Electrons flow more easily through some substances (such as a moist tongue) than others (such as a dry tongue).

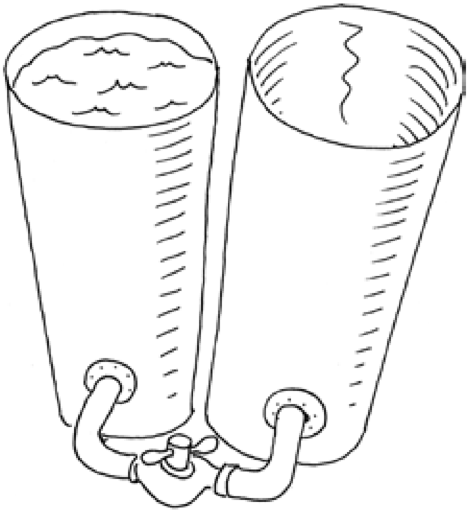

Figure 1-26. *Think of the cells in a battery as being like two cylinders: one full of water, the other empty. Open a connection between the cylinders, and the water will flow until the levels are equal on both sides. The less resistance in the connection, the faster the flow will be.*

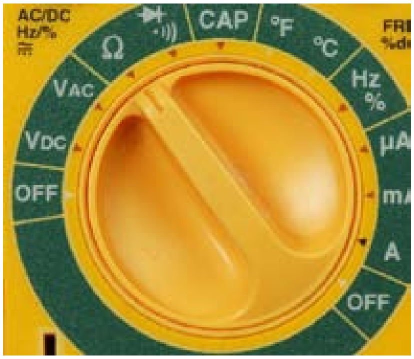

Figure 1-23

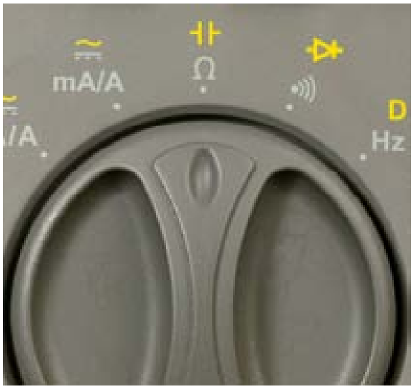

Figure 1-24

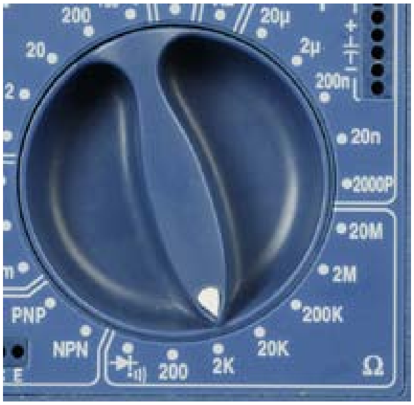

Figure 1-25. *To measure ohms, turn the dial to the ohm (omega) symbol. On an autoranging meter, you can then press the Range button repeatedly to display different ranges of resistance, or simply touch the probes to a resistance and wait for the meter to choose a range automatically. A manual meter requires you to select the range with the dial (you should set it to 100K or higher, to measure skin resistance). If you don't get a meaningful reading, try a different range.*

BACKGROUND

The man who discovered resistance

Georg Simon Ohm, pictured in Figure 1-27, was born in Bavaria in 1787 and worked in obscurity for much of his life, studying the nature of electricity using metal wire that he had to make for himself (you couldn't truck on down to Home Depot for a spool of hookup wire back in the early 1800s).

Despite his limited resources and inadequate mathematical abilities, Ohm was able to demonstrate in 1827 that the electrical resistance of a conductor such as copper varied in proportion with its area of cross-section, and the current flowing through it is proportional to the voltage applied to it, as long as temperature is held constant. Fourteen years later, the Royal Society in London finally recognized the significance of his contribution and awarded him the Copley Medal. Today, his discovery is known as Ohm's Law.

Figure 1-27. *Georg Simon Ohm, after being honored for his pioneering work, most of which he pursued in relative obscurity.*

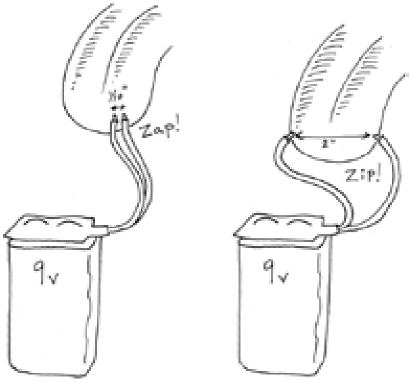

Figure 1-28. *Modifying the tongue test to show that a shorter distance, with lower resistance, allows greater flow of electricity, and a bigger zap.*

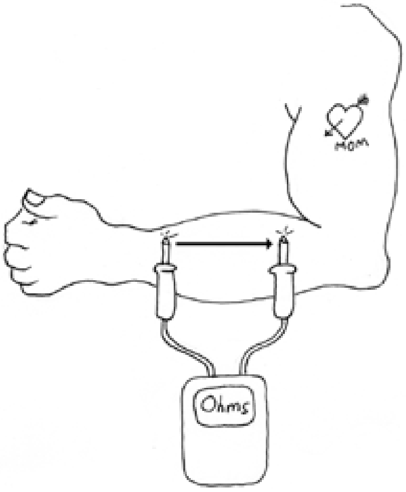

Figure 1-29. *Moisten your skin before trying to measure its resistance. You should find that the resistance goes up as you move the meter probes farther apart. The resistance is proportional to the distance.*

Further Investigation

Attach the snap-on terminal cap (shown earlier in Figure 1-8) to the 9-volt battery. Take the two wires that are attached to the cap and hold them so that the bare ends are just a few millimeters apart. Touch them to your tongue. Now separate the ends of the wires by a couple of inches, and touch them to your tongue again. (See Figure 1-28.) Notice any difference?

Use your meter to measure the electrical resistance of your tongue, this time varying the distance between the two probes. When electricity travels through a shorter distance, it encounters less total resistance. As a result, the current (the flow of electricity per second) increases. You can try a similar experiment on your arm, as shown in Figure 1-29.

Use your meter to test the electrical resistance of water. Dissolve some salt in the water, and test it again. Now try measuring the resistance of distilled water (in a clean glass).

The world around you is full of materials that conduct electricity with varying amounts of resistance.

Cleanup and Recycling

Your battery should not have been damaged or significantly discharged by this experiment. You'll be able to use it again.

Remember to switch off your meter before putting it away.

Experiment 2: Let's Abuse a Battery!

To get a better feeling for electrical power, you're going to do what most books tell you not to do. You're going to short out a battery. A short circuit is a direct connection between the two sides of a power source.

Short Circuits

Short circuits can be dangerous. Do not short out a power outlet in your home: there'll be a loud bang, a bright flash, and the wire or tool that you use will be partially melted, while flying particles of melted metal can burn you or blind you.

If you short out a car battery, the flow of current is so huge that the battery might even explode, drenching you in acid (Figure 1-30).

Lithium batteries are also dangerous. Never short-circuit a lithium battery: it can catch fire and burn you (Figure 1-31).

Use only an alkaline battery in this experiment, and only a single AA cell (Figure 1-32). You should also wear safety glasses in case you happen to have a defective battery.

You will need:

- 1.5-volt AA battery
- Single-battery carrier
- 3-amp fuse
- Safety glasses (regular eyeglasses or sunglasses will do)

Procedure

Use an alkaline battery. Do not use any kind of rechargeable battery.

Put the battery into a battery holder that's designed for a single battery and has two thin insulated wires emerging from it, as shown in Figure 1-32. Do not use any other kind of battery holder.

Use an alligator clip to connect the bare ends of the wires, as shown in Figure 1-32. There will be no spark, because you are using only 1.5 volts. Wait one minute, and you'll find that the wires are getting hot. Wait another minute, and the battery, too, will be hot.

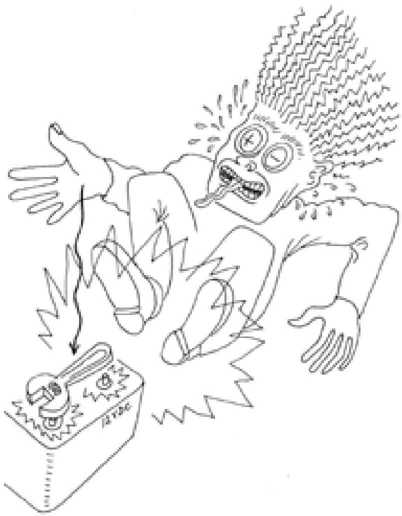

Figure 1-30. *Anyone who has dropped an adjustable wrench across the bare terminals of a car battery will tell you that short circuits can be dramatic at a "mere" 12 volts, if the battery is big enough.*

Figure 1-31. *The low internal resistance of lithium batteries (which are often used in laptop computers) allows high currents to flow, with unexpected results. Never fool around with lithium batteries!*

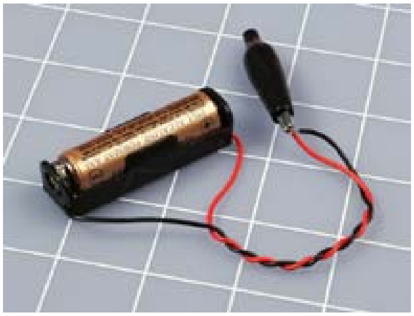

Figure 1-32. *Shorting out an alkaline battery can be safe if you follow the directions precisely. Even so, the battery is liable to become too hot to touch comfortably. Don't try this with any type of rechargeable battery.*

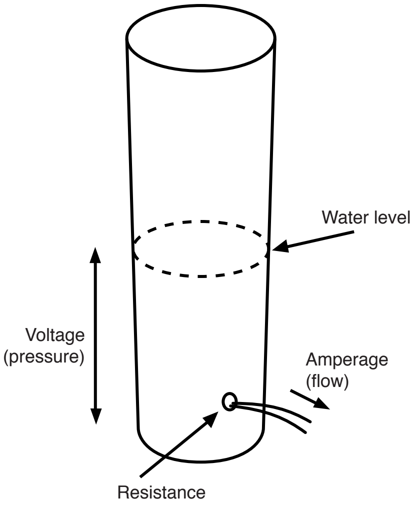

Figure 1-33. *Think of voltage as pressure, and amperes as flow.*

The heat is caused by electricity flowing through the wires and through the electrolyte (the conductive fluid) inside the battery. If you've ever used a hand pump to force air into a bicycle tire, you know that the pump gets warm. Electricity behaves in much the same way. You can imagine the electricity being composed of particles (electrons) that make the wire hot as they push through it. This isn't a perfect analogy, but it's close enough for our purposes.

Chemical reactions inside the battery create electrical pressure. The correct name for this pressure is *voltage*, which is measured in volts and is named after Alessandro Volta, an electrical pioneer.

Going back to the water analogy: the height of the water in a tank is proportionate to the pressure of the water, and comparable to voltage. Figure 1-33 may help you to visualize this.

But volts are only half of the story. When electrons flow through a wire, the flow is known as *amperage*, named after yet another electrical pioneer, André-Marie Ampère. The flow is also generally known as current. It's the current—the amperage—that generates the heat.

BACKGROUND

Why didn't your tongue get hot?

When you touched the 9-volt battery to your tongue, you felt a tingle, but no perceptible heat. When you shorted out a battery, you generated a noticeable amount of heat, even though you used a lower voltage. How can we explain this?

The electrical resistance of your tongue is very high, which reduces the flow of electrons. The resistance of a wire is very low, so if there's only a wire connecting the two terminals of the battery, more current will pass through it, creating more heat. If all other factors remain constant:

- Lower resistance allows more current to flow (Figure 1-34).
- The heat generated by electricity is proportional to the amount of electricity (the current) that flows.

Here are some other basic concepts:

- The flow of electricity per second is measured in amperes, or amps.
- The pressure of electricity causes the flow, measured in volts.
- The resistance to the flow is measured in ohms.
- A higher resistance restricts the current.
- A higher voltage overcomes resistance and increases the current.

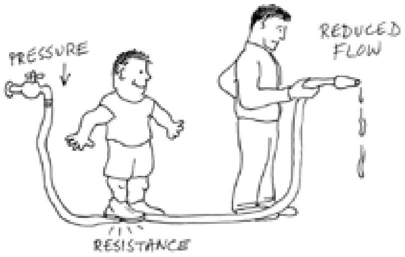

Figure 1-34. *Larger resistance results in smaller flow—but if you increase the pressure, it may overcome the resistance and increase the flow.*

If you're wondering exactly how much current flows between the terminals of a battery when you short it out, that's a difficult question to answer. If you try to use your multimeter to measure it, you're liable to blow the fuse inside the meter. Still, you can use your very own 3-amp fuse, which we can sacrifice because it didn't cost very much.

First inspect the fuse very carefully, using a magnifying glass if you have one. You should see a tiny S-shape in the transparent window at the center of the fuse. That S is a thin section of metal that melts easily.

Remove the battery that you short-circuited. It is no longer useful for anything, and should be recycled if possible. Put a fresh battery into the battery carrier, connect the fuse as shown in Figure 1-35, and take another look. You should see a break in the center of the S shape, where the metal melted almost instantly. Figure 1-36 shows the fuse before you connected it, and Figure 1-37 depicts a blown fuse. This is how a fuse works: it melts to protect the rest of the circuit. That tiny break inside the fuse stops any more current from flowing.

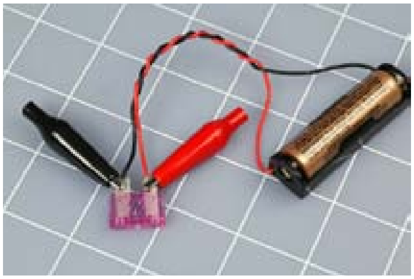

Figure 1-35. *When you attach both wires to the fuse, the little S-shaped element inside will melt almost instantly.*

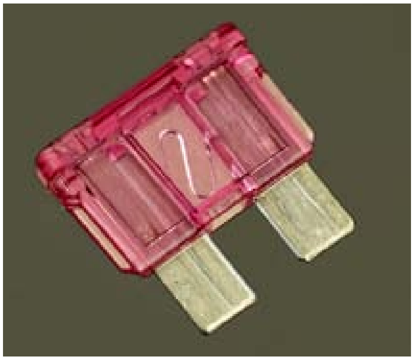

Figure 1-36. *A 3-amp fuse, before its element was melted by a single 1.5-volt battery.*

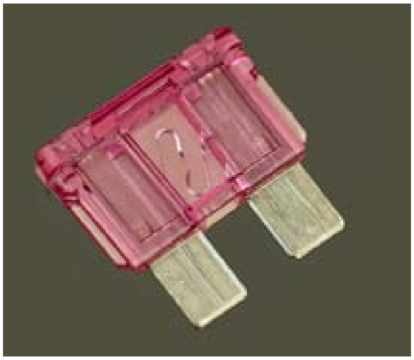

Figure 1-37. *The same fuse after being melted by electric current.*

FUNDAMENTALS

Volt basics

Electrical pressure is measured in volts. The volt is an international unit. A millivolt is 1/1,000 of a volt.

Number of volts	Usually expressed as	Abbreviated as
0.001 volts	1 millivolt	1 mV
0.01 volts	10 millivolts	10 mV
0.1 volts	100 millivolts	100 mV
1 volt	1,000 millivolts	1V

Ampere basics

We measure electrical flow in amperes, or amps. The ampere is an international unit, often referred to as an "amp." A milliamp is 1/1,000 of an ampere.

Number of amperes	Usually expressed as	Abbreviated as
0.001 amps	1 milliamp	1 mA
0.01 amps	10 milliamps	10 mA
0.1 amps	100 milliamps	100 mA
1 amp	1,000 milliamps	1A

BACKGROUND

Inventor of the battery

Alessandro Volta (Figure 1-38) was born in Italy in 1745, long before science was broken up into specialties. After studying chemistry (he discovered methane in 1776), he became a professor of physics and became interested in the so-called galvanic response, whereby a frog's leg will twitch in response to a jolt of static electricity.

Using a wine glass full of salt water, Volta demonstrated that the chemical reaction between two electrodes, one made of copper, the other of zinc, will generate a steady electric current. In 1800, he refined his apparatus by stacking plates of copper and zinc, separated by cardboard soaked in salt and water. This "voltaic pile" was the first electric battery.

Figure 1-38. *Alessandro Volta discovered that chemical reactions can create electricity.*

FUNDAMENTALS

Direct and alternating current

The flow of current that you get from a battery is known as *direct current*, or DC. Like the flow of water from a faucet, it is a steady stream, in one direction.

The flow of current that you get from the "hot" wire in a power outlet in your home is very different. It changes from positive to negative 50 times each second (in Great Britain and some other nations, 60 times per second). This is known as *alternating current*, or AC, which is more like the pulsatile flow you get from a power washer.

Alternating current is essential for some purposes, such as cranking up voltage so that electricity can be distributed over long distances. AC is also useful in motors and domestic appliances. The parts of an American power outlet are shown in Figure 1-39. A few other nations, such as Japan, also use American-style outlets.

For most of this book I'm going to be talking about DC, for two reasons: first, most simple electronic circuits are powered with DC, and second, the way it behaves is much easier to understand.

I won't bother to mention repeatedly that I'm dealing with DC. Just assume that everything is DC unless otherwise noted.

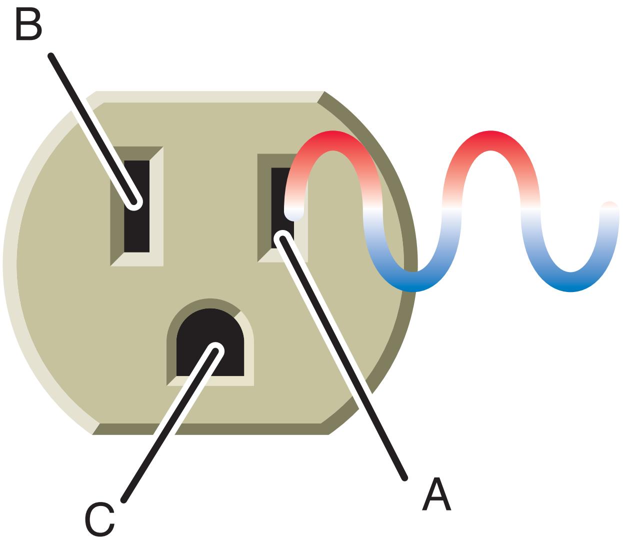

Figure 1-39. *This style of power outlet is found in North America, South America, Japan, and some other nations. European outlets look different, but the principle remains the same. Socket A is the "live" side of the outlet, supplying voltage that alternates between positive and negative, relative to socket B, which is called the "neutral" side. If an appliance develops a fault such as an internal loose wire, it should protect you by sinking the voltage through socket C, the ground.*

Cleanup and Recycling

The first AA battery that you shorted out is probably damaged beyond repair. You should dispose of it. Putting batteries in the trash is not a great idea, because they contain heavy metals that should be kept out of the ecosystem. Your state or town may include batteries in a local recycling scheme. (California requires that almost all batteries be recycled.) You'll have to check your local regulations for details.

The blown fuse is of no further use, and can be thrown away.

The second battery, which was protected by the fuse, should still be OK. The battery holder also can be reused later.

Experiment 3: Your First Circuit

Now it's time to make electricity do something that's at least slightly useful. For this purpose, you'll use components known as resistors, and a light-emitting diode, or LED.

You will need:

- 1.5-volt AA batteries. Quantity: 4.
- Four-battery holder. Quantity: 1.
- Resistors: 470Ω, 1K, and either 2K or 2.2K (the 2.2K value happens to be more common than 2K, but either will do in this experiment). Quantity: 1 of each resistor.
- An LED, any type. Quantity: 1.
- Alligator clips. Quantity: 3.

Setup

It's time to get acquainted with the most fundamental component we'll be using in electronic circuits: the humble resistor. As its name implies, it resists the flow of electricity. As you might expect, the value is measured in ohms.

If you bought a bargain-basement assortment package of resistors, you may find nothing that tells you their values. That's OK; we can find out easily enough. In fact, even if they are clearly labeled, I want you to check their values yourself. You can do it in two ways:

- Use your multimeter. This is excellent practice in learning to interpret the numbers that it displays.
- Learn the color codes that are printed on most resistors. See the following section, "Fundamentals: Decoding resistors," for instructions.

After you check them, it's a good idea to sort them into labeled compartments in a little plastic parts box. Personally, I like the boxes sold at the Michaels chain of crafts stores, but you can find them from many sources.

BACKGROUND

Father of electromagnetism

Born in 1775 in France, André-Marie Ampère (Figure 1-40) was a mathematical prodigy who became a science teacher, despite being largely self-educated in his father's library. His best-known work was to derive a theory of electromagnetism in 1820, describing the way that an electric current generates a magnetic field. He also built the first instrument to measure the flow of electricity (now known as a *galvanometer*), and discovered the element fluorine.

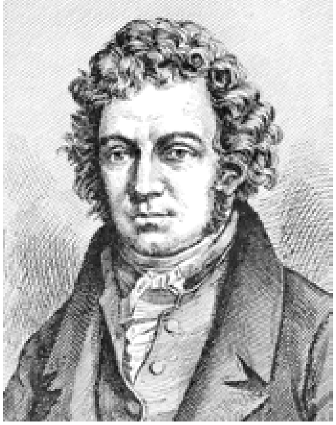

Figure 1-40. *Andre-Marie Ampere found that an electric current running through a wire creates a magnetic field around it. He used this principle to make the first reliable measurements of what came to be known as amperage.*

FUNDAMENTALS

Decoding resistors

Some resistors have their value clearly stated on them in microscopic print that you can read with a magnifying glass. Most, however, are color-coded with stripes. The code works like this: first, ignore the color of the body of the resistor. Second, look for a silver or gold stripe. If you find it, turn the resistor so that the stripe is on the righthand side. Silver means that the value of the resistor is accurate within 10%, while gold means that the value is accurate within 5%. If you don't find a silver or gold stripe, turn the resistor so that the stripes are clustered at the left end. You should now find yourself looking at three colored stripes on the left. Some resistors have more stripes, but we'll deal with those in a moment. See Figures 1-41 and 1-42.

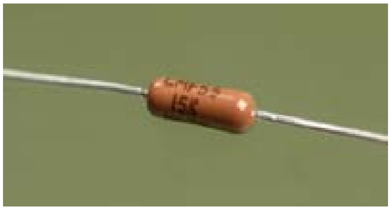

Figure 1-41. *Some modern resistors have their values printed on them, although you may need a magnifier to read them. This 15K resistor is less than half an inch long.*

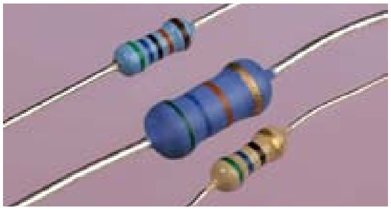

Figure 1-42. *From top to bottom, these resistor values are 56,000 ohms (56K), 5,600 ohms (5.6K), and 560 ohms. The size tells you how much power the resistor can handle; it has nothing to do with the resistance. The smaller components are rated at 1/4 watt; the larger one in the center can handle 1 watt of power.*

Starting from the left, the first and second stripes are coded according to this table:

Black	0
Brown	1
Red	2
Orange	3
Yellow	4
Green	5
Blue	6
Violet	7
Gray	8
White	9

The third stripe has a different meaning: It tells you how many zeros to add, like this:

Black	-	No zeros
Brown	0	1 zero
Red	00	2 zeros
Orange	000	3 zeros
Yellow	0000	4 zeros
Green	00000	5 zeros
Blue	000000	6 zeros
Violet	0000000	7 zeros
Gray	00000000	8 zeros
White	000000000	9 zeros

FUNDAMENTALS

Decoding resistors (continued)

Note that the color-coding is consistent, so that green, for instance, means either a value of 5 (for the first two stripes) or 5 zeros (for the third stripe). Also, the sequence of colors is the same as their sequence in a rainbow.

So, a resistor colored brown-red-green would have a value of 1-2 and five zeros, making 1,200,000 ohms, or 1.2MΩ. A resistor colored orange-orange-orange would have a value of 3-3 and three zeros, making 33,000 ohms, or 33KΩ. A resistor colored brown-black-red would have a value of 1-0 and two additional zeros, or 1KΩ. Figure 1-43 shows some other examples.

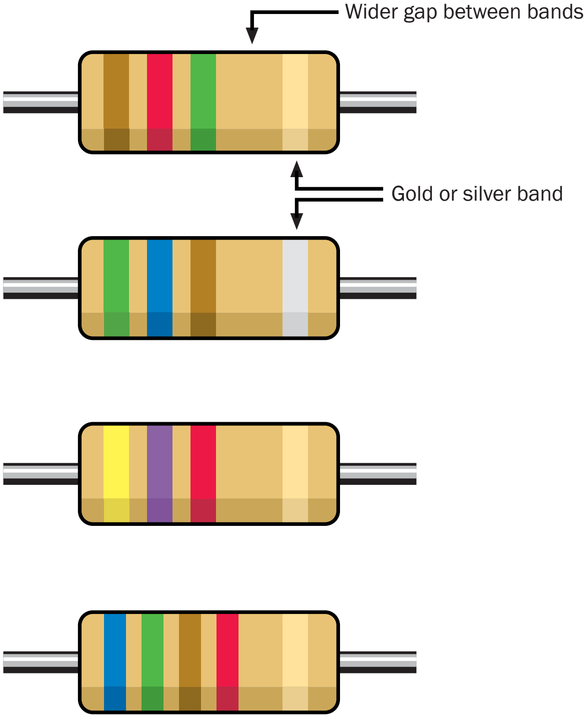

Figure 1-43. *To read the value of a resistor, first turn it so that the silver or gold stripe is on the right, or the other stripes are clustered on the left. From top to bottom: The first resistor has a value of 1-2 and five zeros, or 1,200,000, which is 1.2MΩ. The second is 5-6 and one zero, or 560Ω. The third is 4-7 and two zeros, or 4,700, which is 4.7KΩ. The last is 6-5-1 and two zeros, or 65,100Ω, which is 65.1KΩ.*

If you run across a resistor with four stripes instead of three, the first *three* stripes are digits and the *fourth* stripe is the number of zeros. The third numeric stripe allows the resistor to be calibrated to a finer tolerance.

Confusing? Absolutely. That's why it's easier to use your meter to check the values. Just be aware that the meter reading may be slightly different from the claimed value of the resistor. This can happen because your meter isn't absolutely accurate, or because the resistor is not absolutely accurate, or both. As long as you're within 5% of the claimed value, it doesn't matter for our purposes.

Lighting an LED

Now take a look at one of your LEDs. An old-fashioned lightbulb wastes a lot of power by converting it into heat. LEDs are much smarter: they convert almost all their power into light, and they last almost indefinitely—as long as you treat them right!

An LED is quite fussy about the amount of power it gets, and the way it gets it. Always follow these rules:

- The *longer* wire protruding from the LED must receive a *more positive* voltage than the shorter wire.
- The voltage difference between the long wire and the short wire must not exceed the limit stated by the manufacturer.
- The current passing through the LED must not exceed the limit stated by the manufacturer.

What happens if you break these rules? Well, we're going to find out!

Make sure you are using fresh batteries. You can check by setting your multimeter to measure volts DC, and touching the probes to the terminals of each battery. You should find that each of them generates a pressure of at least 1.5 volts. If they read slightly higher than this, it's normal. A battery starts out above its rated voltage, and delivers progressively less as you use it. Batteries also lose some voltage while they are sitting on the shelf doing nothing.

Load your battery holder (taking care that the batteries are the right way around, with the negative ends pressing against the springs in the carrier). Use your meter to check the voltage on the wires coming out of the battery carrier. You should have at least 6 volts.

Now select a 2KΩ resistor. Remember, "2KΩ" means "2,000 ohms." If it has colored stripes, they should be red-black-red, meaning 2-0 and two more zeros. Because 2.2K resistors are more common than 2K resistors, you can substitute one of them if necessary. It will be colored red-red-red.

Wire it into the circuit as shown in Figures 1-44 and 1-45, making the connections with alligator clips. You should see the LED glow very dimly.

Now swap out your 2K resistor and substitute a 1K resistor, which will have brown-black-red stripes, meaning 1-0 and two more zeros. The LED should glow more brightly.

Swap out the 1K resistor and substitute a 470Ω resistor, which will have yellow-violet-brown stripes, meaning 4-7 and one more zero. The LED should be brighter still.

This may seem very elementary, but it makes an important point. The resistor blocks a percentage of the voltage in the circuit. Think of it as being like a kink or constriction in a flexible hose. A higher-value resistor blocks more voltage, leaving less for the LED.

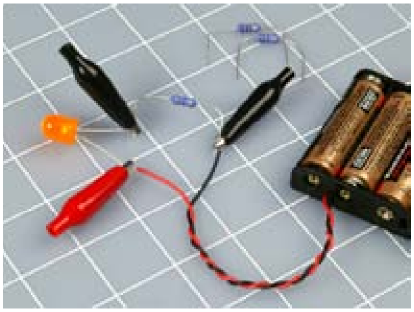

Figure 1-44. *The setup for Experiment 3, showing resistors of 470Ω, 1KΩ, and 2KΩ. Apply alligator clips where shown, to make a secure contact, and try each of the resistors one at a time at the same point in the circuit, while watching the LED.*

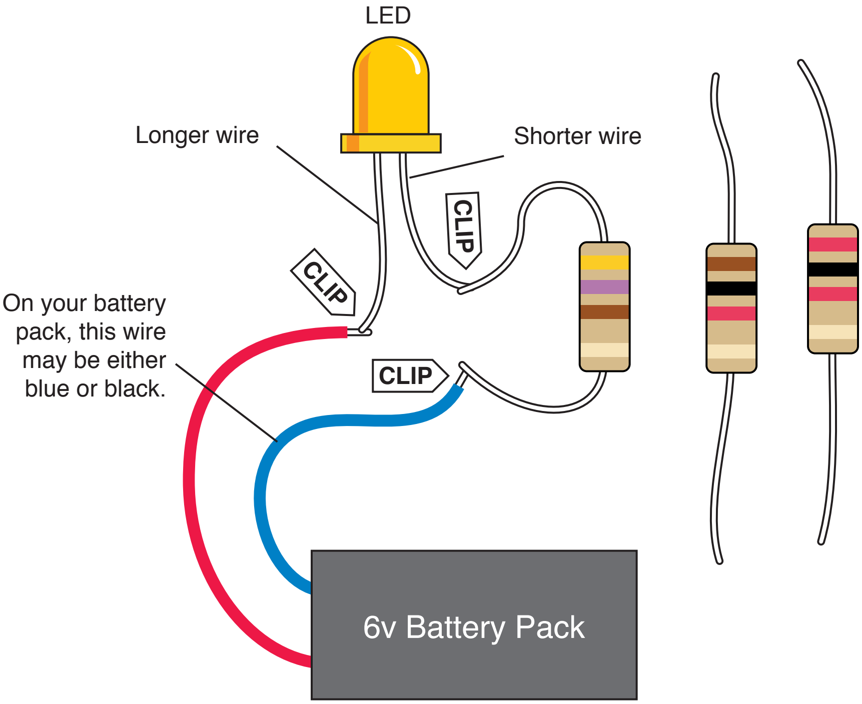
LED
Longer wire
Shorter wire
CLIP
CLIP
CLIP
On your battery pack, this wire may be either blue or black.
6v Battery Pack

Figure 1-45. *Here's how it actually looks, using a large LED. If you start with the highest value resistor, the LED will glow very dimly as you complete the circuit. The resistor drops most of the voltage, leaving the LED with insufficient current to make it shine brightly.*

Cleanup and Recycling

We'll use the batteries and the LED in the next experiment. The resistors can be reused in the future.

Experiment 4: Varying the Voltage

Potentiometers come in various shapes and sizes, but they all do the same thing: they allow you to vary voltage and current by varying resistance. This experiment will enable you to learn more about voltage, amperage, and the relationship between them. You'll also learn how to read a manufacturer's data sheet.

You will need the same batteries, battery carrier, alligator clips, and LED from the last experiment, plus:

- Potentiometer, 2KΩ linear. Quantity: 2. (See Figure 1-46.) Full-sized potentiometers that look like this are becoming less common, as miniature versions are taking their place. I'd like you to use a large one, though, because it's so much easier to work with.
- One extra LED.
- Multimeter.

Look Inside Your Potentiometer

The first thing I want you to do is find out how a potentiometer works. This means you'll have to open it, which is why your shopping list required you to buy two of them, in case you can't put the first one back together again.

Most potentiometers are held together with little metal tabs. You should be able to grab hold of the tabs with your wire cutters or pliers, and bend them up and outward. If you do this, the potentiometer should open up as shown in Figures 1-47 and 1-48.

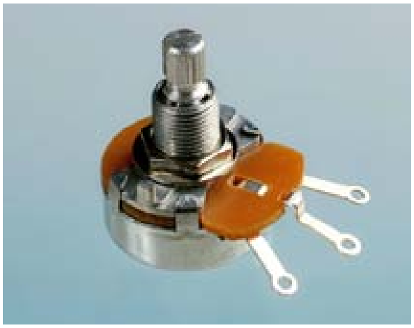

Figure 1-46

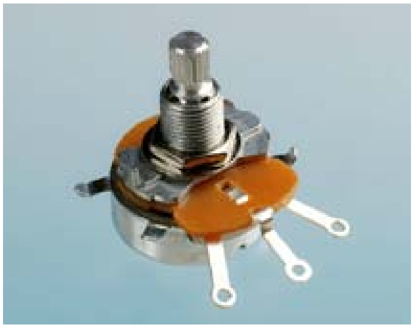

Figure 1-47

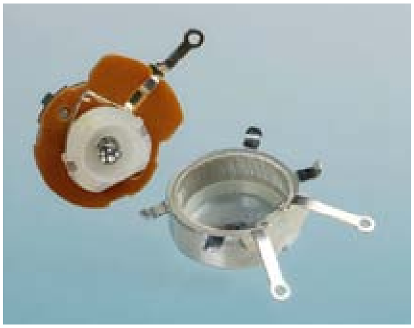

Figure 1-48. *To open the potentiometer, first pry up the four little metal tabs around the edge (you can see one sticking out at the left and another one sticking out at the right in Figure 1-47). Inside is a coil of wire around a flat plastic band, and a pair of springy contacts (the wiper), which conduct electricity to or from any point in the coil when you turn the shaft.*

Depending whether you have a really cheap potentiometer or a slightly more high-class version, you may find a circular track of conductive plastic or a loop of coiled wire. Either way, the principle is the same. The wire or the plastic possesses some resistance (a total of 2K in this instance), and as you turn the shaft of the potentiometer, a wiper rubs against the resistance, giving you a shortcut to any point from the center terminal.

You can try to put it back together, but if it doesn't work, use your backup potentiometer instead.

To test your potentiometer, set your meter to measure resistance (ohms) and touch the probes while turning the potentiometer shaft to and fro, as shown in Figure 1-49.

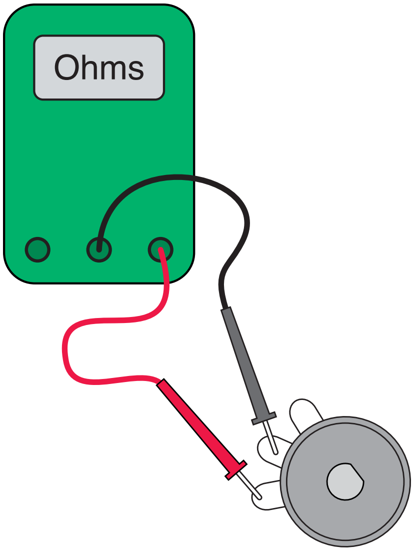

Figure 1-49. *Measure the resistance between these two terminals of the potentiometer while you turn its shaft to and fro.*

Dimming Your LED

Begin with the potentiometer turned all the way counterclockwise, otherwise you'll burn out the LED before we even get started. (A very, very small number of potentiometers increase and decrease resistance in the opposite way to which I'm describing here, but as long as your potentiometer looks like the one in Figure 1-48 after you open it up, my description should be accurate.)

Now connect everything as shown in Figures 1-50 and 1-51, taking care that you don't allow the metal parts of any of the alligator clips to touch each other. Now turn up the potentiometer *very* slowly. You'll notice the LED glowing brighter, and brighter, and brighter—until, oops, it goes dark. You see how easy it is to destroy modern electronics? Throw away that LED. It will never glow again. Substitute a new LED, and we'll be more careful this time.

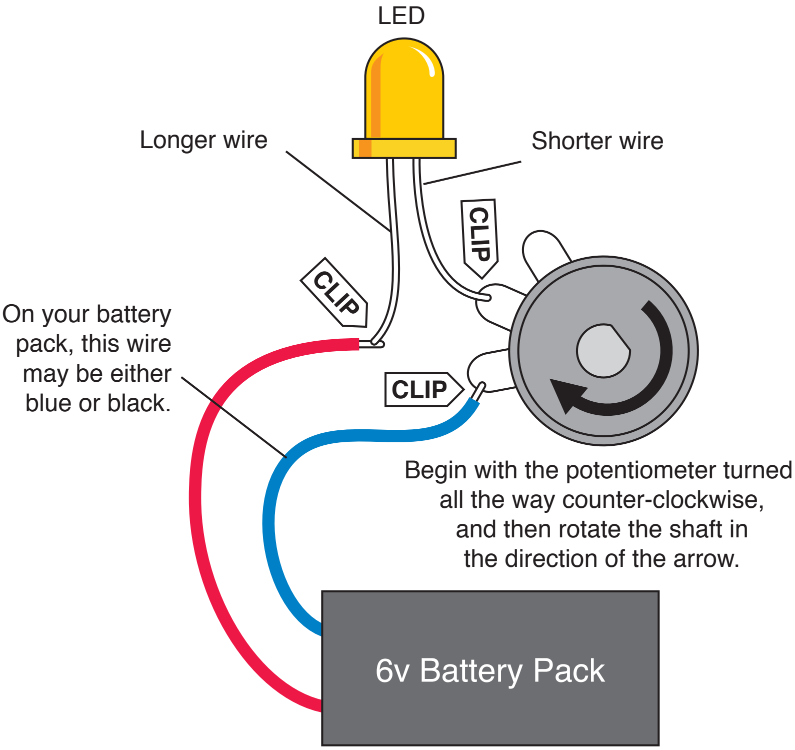
LED
Longer wire
Shorter wire
CLIP
CLIP
CLIP
On your battery pack, this wire may be either blue or black.
Begin with the potentiometer turned all the way counter-clockwise, and then rotate the shaft in the direction of the arrow.
6v Battery Pack

Figure 1-50. *The setup for Experiment 4. Rotating the shaft of the 2K potentiometer varies its resistance from 0 to 2,000Ω. This resistance protects the LED from the full 6 volts of the battery.*

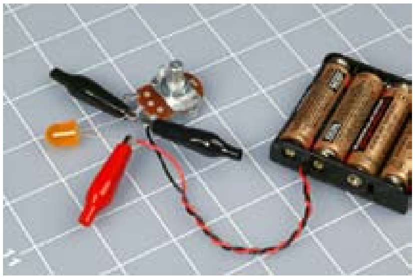

Figure 1-51. *The LED in this photo is dark because I turned the potentiometer up just a little bit too far.*

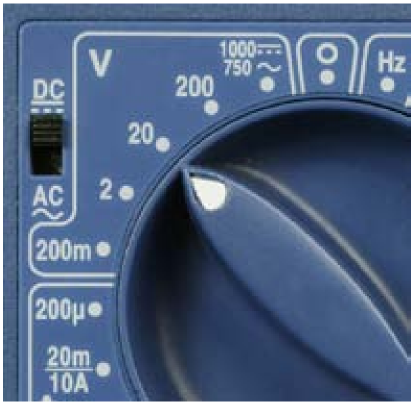

Figure 1-52

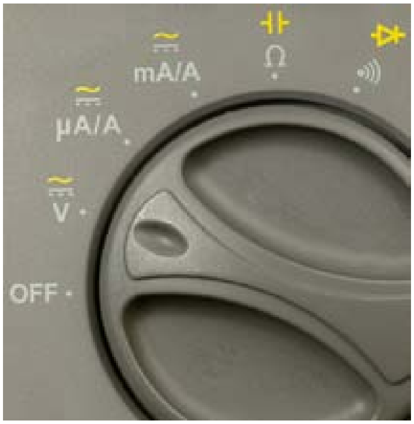

Figure 1-53

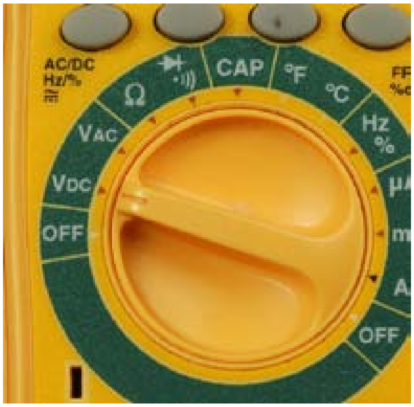

Figure 1-54. *Each meter has a different way to measure volts DC. The manually adjusted meter (top) requires you to move a slider switch to "DC" and then choose the highest voltage you want to measure: In this case, the selected voltage is 20 (because 2 would be too low). Using the autoranging RadioShack meter, you set it to "V" and the meter will figure out which range to use.*

While the batteries are connected to the circuit, set your meter to measure *volts DC* as shown in Figures 1-52 through 1-54. Now touch the probes either side of the LED. Try to hold the probes in place while you turn the potentiometer up a little, and down a little. You should see the voltage pressure around the LED changing accordingly. We call this the *potential difference* between the two wires of the LED.

If you were using a miniature old-fashioned lightbulb instead of an LED, you'd see the potential difference varying much more, because a lightbulb behaves like a "pure" resistor, whereas an LED self-adjusts to some extent, modifying its resistance as the voltage pressure changes.

Now touch the probes to the two terminals of the potentiometer that we're using, so that you can measure the potential difference between them. The potentiometer and the LED share the total available voltage, so when the potential difference (the voltage drop) around the potentiometer goes up, the potential difference around the LED goes down, and vice versa. See Figures 1-55 through 1-57. A few things to keep in mind:

- If you add the voltage drops across the devices in the circuit, the total is the same as the voltage supplied by the batteries.
- You measure voltage relatively, between two points in a circuit.
- Apply your meter like a stethoscope, without disturbing or breaking the connections in the circuit.

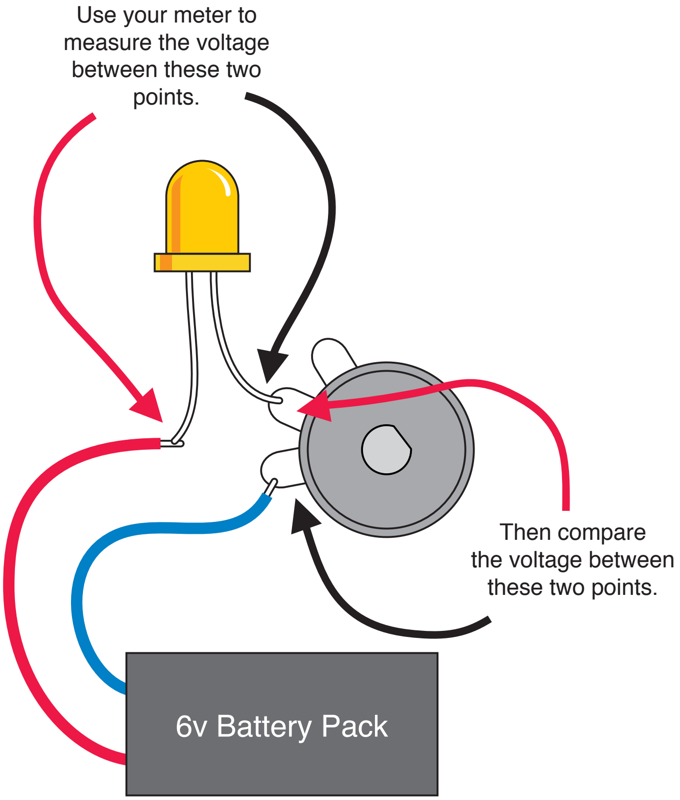

Use your meter to measure the voltage between these two points.
Then compare the voltage between these two points.
6v Battery Pack

Figure 1-55. *How to measure voltage in a simple circuit.*

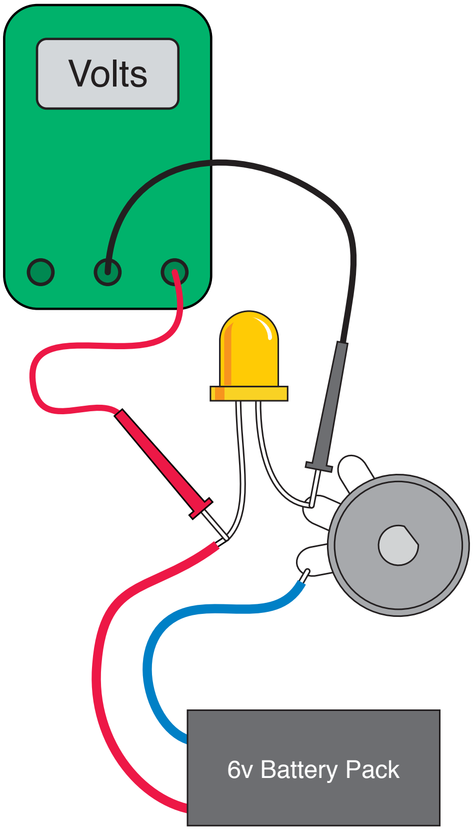

Figure 1-56. *The meter shows how much voltage the LED takes.*

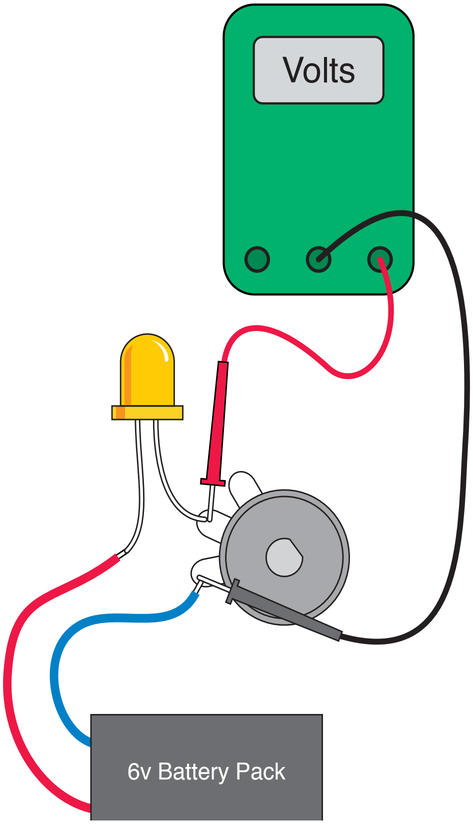

Figure 1-57. *The meter shows how much voltage the potentiometer takes.*

Checking the Flow

Now I want you to make a different measurement. I want you to measure the flow, or current, in the circuit, using your meter set to mA (milliamps). Remember, to measure current:

- You can only measure current when it passes *through* the meter.
- You have to insert your meter into the circuit.
- Too much current will blow the fuse inside your meter.

Make sure you set your meter to measure mA, not volts, before you try this. Some meters require you to move one of your leads to a different socket on the meter, to measure mA. See Figures 1-58 through 1-61.

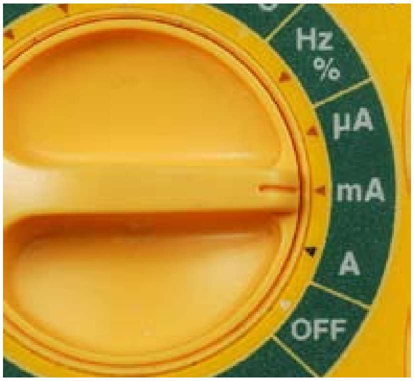

Figure 1-58. *Any meter will blow its internal fuse if you try to make it measure too high an amperage. In our circuit, this is not a risk as long as you keep the potentiometer in the middle of its range. Choose "mA" for milliamps and remember that the meter displays numbers that mean thousandths of an amp.*

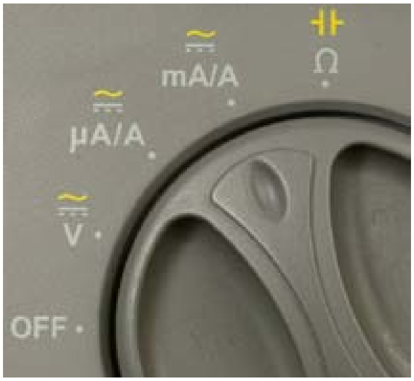

Figure 1-59

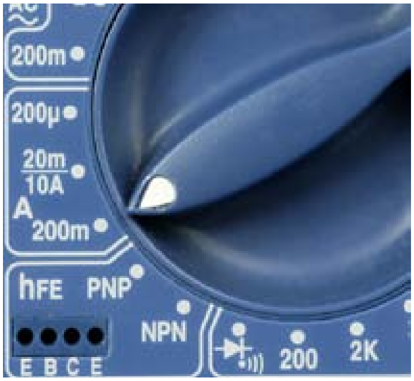

Figure 1-60

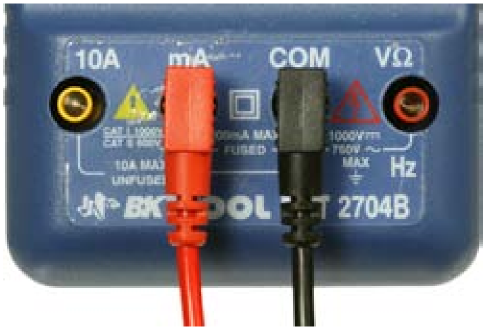

Figure 1-61. *A manual meter such as the one here may require you to shift the red lead to a different socket, to measure milliamps. Most modern meters don't require this until you are measuring higher currents.*

Insert your meter into the circuit, as shown in Figure 1-62. Don't turn the potentiometer more than halfway up. The resistance in the potentiometer will protect your meter, as well as the LED. If the meter gets too much current, you'll find yourself replacing its internal fuse.

As you adjust the potentiometer up and down a little, you should find that the varying resistance in the circuit changes the flow of current—the amperage. This is why the LED burned out in the previous experiment: too much current made it hot, and the heat melts it inside, just like the fuse in the previous experiment. *A higher resistance limits the flow of current, or amperage*.

Now insert the meter in another part of the circuit, as shown in Figure 1-63. As you turn the potentiometer up and down, you should get exactly the same results as with the configuration in Figure 1-64. This is because the current is the same at all points in a similar circuit. *It has to be, because the flow of electrons has no place else to go*.

It's time now to nail this down with some numbers. Here's one last thing to try. Set aside the LED and substitute a 1KΩ resistor, as shown in Figure 1-64. The total resistance in the circuit is now 1KΩ plus whatever the resistance the potentiometer provides, depending how you set it. (The meter also has some resistance, but it's so low, we can ignore it.)

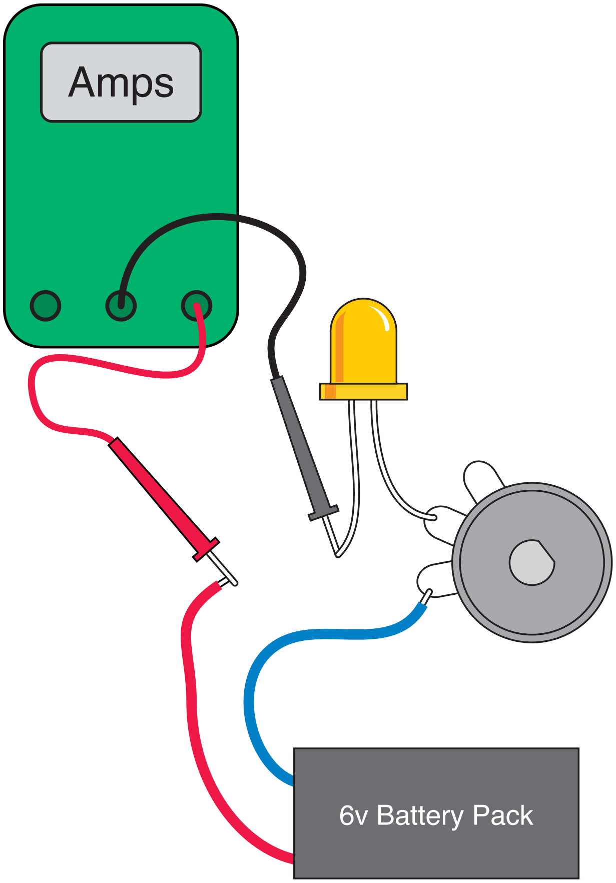

Figure 1-62. *To measure amps, as illustrated here and in Figure 1-63, the current has to pass through the meter. When you increase the resistance, you restrict the current flow, and the lower flow makes the LED glow less brightly.*

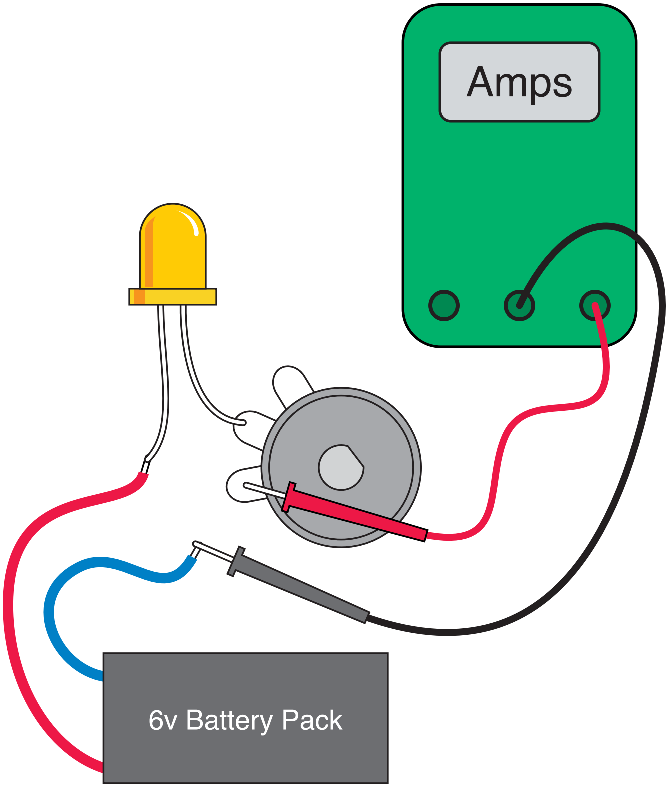

Figure 1-63

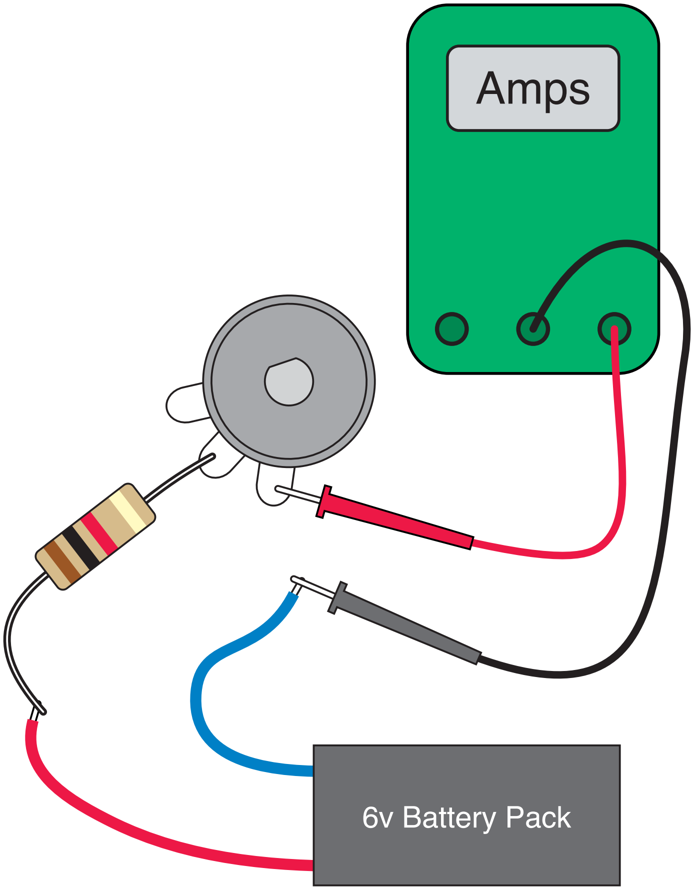

Figure 1-64. *If you substitute a resistor instead of the LED, you can confirm that the current flowing through the circuit varies with the total resistance in the circuit, if the voltage stays the same.*

Turn the potentiometer all the way counterclockwise, and you have a total of 3K resistance in the circuit. Your meter should show about 2 mA flowing. Now turn the potentiometer halfway, and you have about 2K total resistance. You should see about 3 mA flowing. Turn the potentiometer all the way clockwise, so there's a total of 1K, and you should see 6 mA flowing. You may notice that if we multiply the resistance by the amperage, we get 6 each time—which just happens to be the voltage being applied to the circuit. See the following table.

Total resistance	**Current**	**Voltage**
(KΩ)	(mA)	(Volts)
3	2	6
2	3	6
1	6	6

In fact, we could say:

voltage = kilohms × milliamps

But wait a minute: 1K is 1,000 ohms, and 1mA is 1/1,000 of an amp. Therefore, our formula should really look like this:

voltage = (ohms × 1,000) × (amps/1,000)

The two factors of 1,000 cancel out, so we get this:

volts = ohms × amps

This is known as Ohm's Law. See the section, "Fundamentals: Ohm's Law," on the following page.

FUNDAMENTALS

Series and parallel

Before we go any further, you should know how resistance in a circuit increases when you put resistors in series or in parallel. Figures 1-65 through 1-67 illustrate this. Remember:

- Resistors in series are oriented so that one follows the other.
- Resistors in parallel are oriented side by side.

When you put two equal-valued resistors in series, you double the total resistance, because electricity has to pass through two barriers in succession.

When you put two equal-valued resistors in parallel, you divide the total resistance by two, because you're giving the electricity two paths which it can take, instead of one.

In reality we don't normally need to put resistors in parallel, but we often put other types of components in parallel. Lightbulbs in your house, for instance, are all wired that way. So, it's useful to understand that resistance in a circuit goes down if you keep adding components in parallel.

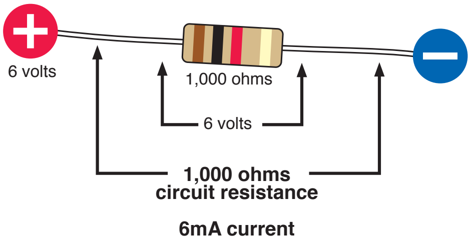
6 volts
1,000 ohms
6 volts
1,000 ohms
circuit resistance
6mA current

Figure 1-65. *One resistor takes the entire voltage, and according to Ohm's Law, it draws v/R = 6/1,000 = 0.006 amps = 6mA of current.*

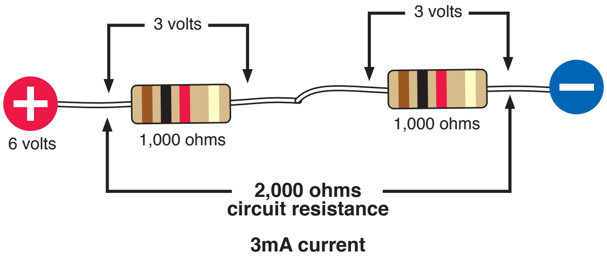
3 volts
3 volts
6 volts
1,000 ohms
1,000 ohms
2,000 ohms
circuit resistance
3mA current

Figure 1-66. *When two resistors are in series, the electricity has to pass through one to reach the other, and therefore each of them takes half the voltage. Total resistance is now 2,000 ohms, and according to Ohm's Law, the circuit draws v/R = 6/2,000 = 0.003 amps = 3mA of current.*

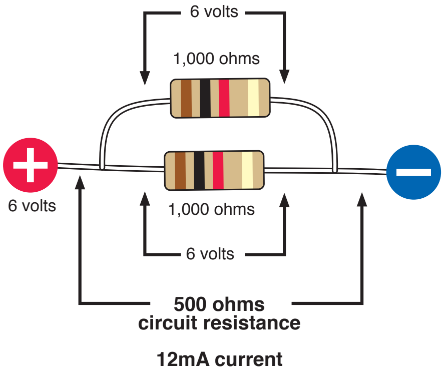
6 volts
1,000 ohms
6 volts
1,000 ohms
6 volts
500 ohms
circuit resistance
12mA current

Figure 1-67. *When two resistors are in parallel, each is exposed to the full voltage, so each of them takes 6 volts. The electricity can now flow through both at once, so the total resistance of the circuit is half as much as before. According to Ohm's Law, the circuit draws v/R = 6/500 = 0.012 amps = 12mA of current.*

FUNDAMENTALS

Ohm's Law

For reasons I'll explain in a moment, amps are normally abbreviated with the letter I. V stands for volts and R stands for resistance in ohms (because the omega symbol, Ω, is not easily generated from most keyboards). Using these symbols, you can write Ohm's Law in three different ways:

$$V = I \times R$$

$$I = V/R$$

$$R = V/I$$

Remember, V is a *difference* in voltage between two points in a simple circuit, R is the resistance in ohms *between* the same two points, and I is the current in amps flowing *through* the circuit between the two points.

Letter I is used because originally current was measured by its *inductance*, meaning the ability to induce magnetic effects. It would be much less confusing to use A for amps, but unfortunately it's too late for that to happen.

Using Ohm's Law

Ohm's Law is extremely useful. For example, it helps us to figure out whether a component can be used safely in a circuit. Instead of stressing the component until we burn it out, we can predict whether it will work.

For instance, the first time you turned the potentiometer, you didn't really know how far you could go until the LED burned out. Wouldn't it be useful to know precisely what resistance to put in series with an LED, to protect it adequately while providing as much light as possible?

How to Read a Data Sheet

Like most information, the answer to this question is available online.

Here's how you find a manufacturer's data sheet (Figure 1-68). First, find the component that you're interested in from a mail-order source. Next, Google the part number and manufacturer's name. Usually the data sheet will pop up as the first hit. A source such as *Mouser.com* makes it even easier by giving you a direct link to manufacturers' data sheets for many products.

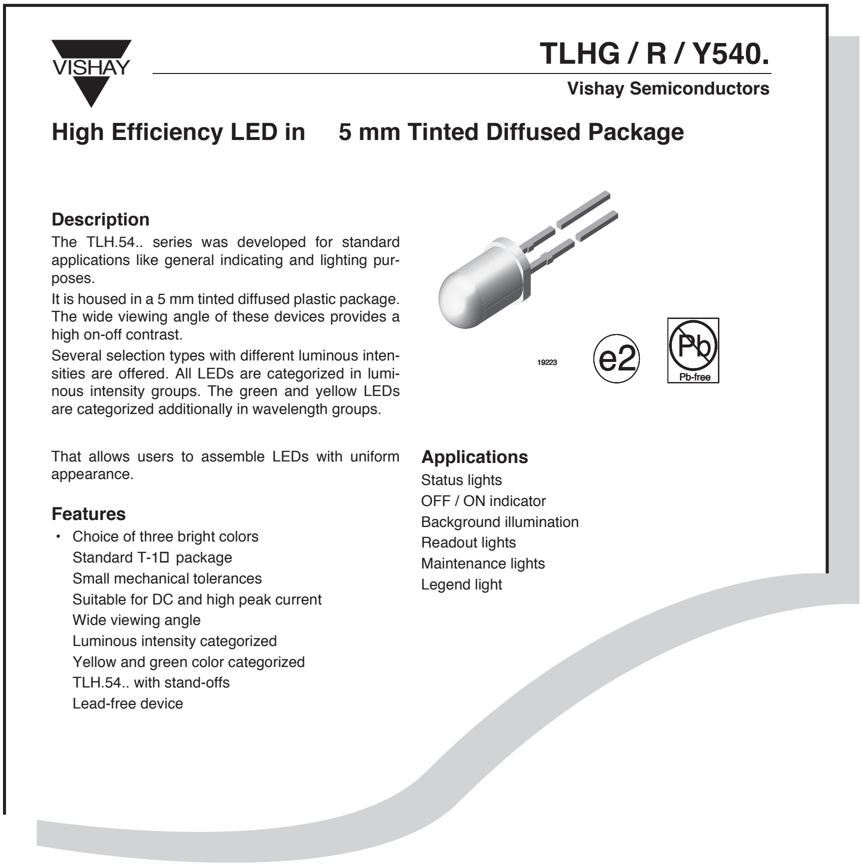

VISHAY

TLHG / R / Y540.

Vishay Semiconductors

High Efficiency LED in 5 mm Tinted Diffused Package

Description

The TLH.54.. series was developed for standard applications like general indicating and lighting purposes.

It is housed in a 5 mm tinted diffused plastic package. The wide viewing angle of these devices provides a high on-off contrast.

Several selection types with different luminous intensities are offered. All LEDs are categorized in luminous intensity groups. The green and yellow LEDs are categorized additionally in wavelength groups.

That allows users to assemble LEDs with uniform appearance.

Features

- Choice of three bright colors
  Standard T-1□ package
  Small mechanical tolerances
  Suitable for DC and high peak current
  Wide viewing angle
  Luminous intensity categorized
  Yellow and green color categorized
  TLH.54.. with stand-offs
  Lead-free device

Applications

Status lights
OFF / ON indicator
Background illumination
Readout lights
Maintenance lights
Legend light

Figure 1-68. *The beginning of a typical data sheet, which includes all relevant specifications for the product, freely available online.*

BACKGROUND

How much voltage does a wire consume?

Normally, we can ignore the resistance in electric wires, such as the little leads of wire that stick out of resistors, because it's trivial. However, if you try to force large amounts of current through long lengths of thin wire, the resistance of the wire can become important.

How important? Once again, we can use Ohm's Law to find out.

Suppose that a very long piece of wire has a resistance of 0.2Ω. And we want to run 15 amps through it. How much voltage will the wire steal from the circuit, because of its resistance?

Once again, you begin by writing down what you know:

$R = 0.2$

$I = 15$

We want to know V, the potential difference, for the wire, so we use the version of Ohm's Law that places V on the left side:

$V = I \times R$

Now plug in the values:

$V = 15 \times 0.2 = 3$ volts

Three volts is not a big deal if you have a high-voltage power supply, but if you are using a 12-volt car battery, this length of wire will take one-quarter of the available voltage.

Now you know why the wiring in automobiles is relatively thick—to reduce its resistance well below 0.2Ω. See Figure 1-69.

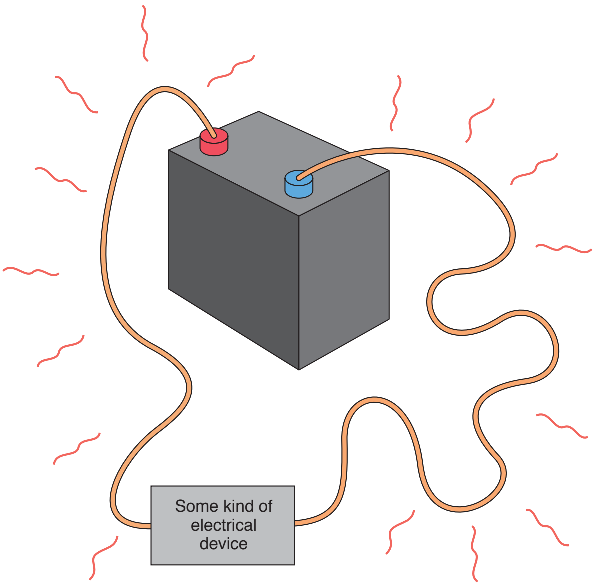
Some kind of electrical device

Figure 1-69. *When a 12-volt car battery runs some kind of electrical device through a long piece of thin wire, the resistance of the wire steals some of the voltage and dissipates it as heat.*

BACKGROUND

The origins of wattage

James Watt (Figure 1-70) is known as the inventor of the steam engine. Born in 1736 in Scotland, he set up a small workshop in the University of Glasgow, where he struggled to perfect an efficient design for using steam to move a piston in a cylinder. Financial problems and the primitive state of the art of metal working delayed practical applications until 1776.

Despite difficulties in obtaining patents (which could only be granted by an act of parliament in those times), Watt and his business partner eventually made a lot of money from his innovations. Although he predated the pioneers in electricity, in 1889 (70 years after his death), his name was assigned to the basic unit of electric power that can be defined by multiplying amperes by volts. See the Fundamentals section, "Watt Basics," on page 31.

Figure 1-70. *James Watt's development of steam power enabled the industrial revolution. After his death, he was honored by having his name applied to the basic unit of power in electricity.*

Here's an example. Suppose I want a red LED, such as the Vishay part TLHR5400, which has become such a common item that I can buy them individually for 9 cents apiece. I click the link to the data sheet maintained by the manufacturer, Vishay Semiconductor. Almost immediately I have a PDF page on my screen. This data sheet is for TLHR, TLHG, and TLHY types of LED, which are red, green, and yellow respectively, as suggested by the R, G, and Y in the product codes. I scroll down and look at the "Optical and Electrical Characteristics" section. It tells me that under conditions of drawing a current of 20 mA, the LED will enjoy a "Typ," meaning, typical, "forward voltage" of 2 volts. The "Max," meaning maximum, is 3 volts.

Let's look at one other data sheet, as not all of them are written the same way. I'll choose a different LED, the Kingbright part WP7113SGC. Click on the link to the manufacturer's site, and I find on the second page of the data sheet a typical forward voltage of 2.2, maximum 2.5, and a maximum forward current of 25 mA. I also find some additional information: a maximum reverse voltage of 5 and maximum reverse current of 10 uA (that's microamps, which are 1,000 times smaller than milliamps). This tells us that you should avoid applying excessive voltage to the LED the wrong way around. If you exceed the reverse voltage, you risk burning out the LED. Always observe polarity!

Kingbright also warns us how much heat the LED can stand: 260° C (500° F) for a few seconds. This is useful information, as we'll be putting aside our alligator clips and using hot molten solder to connect electrical parts in the near future. Because we have already destroyed a battery, a fuse, and an LED in just four experiments, maybe you won't be surprised when I tell you that we will destroy at least a couple more components as we test their limits with a soldering iron.

Anyway, now we know what an LED wants, we can figure out how to supply it. If you have any difficulties dealing with decimals, check the Fundamentals section "Decimals," on the next page, before continuing.

How Big a Resistor Does an LED Need?

Suppose that we're use the Vishay LED. Remember its requirements from the data sheet? Maximum of 3 volts, and a safe current of 20mA.

I'm going to limit it to 2.5 volts, to be on the safe side. We have 6 volts of battery power. Subtract 2.5 from 6 and we get 3.5. So we need a resistor that will take 3.5 volts from the circuit, leaving 2.5 for the LED.

The current flow is the same at all places in a simple circuit. If we want a maximum of 20mA to flow through the LED, the same amount of current will be flowing through the resistor.

Now we can write down what we know about the resistor in the circuit. Note that we have to convert all units to volts, amps, and ohms, so that 20mA should be written as 0.02 amps:

V = 3.5 (the potential drop across the resistor)
I = 0.02 (the current flowing through the resistor)

We want to know R, the resistance. So, we use the version of Ohm's Law that puts R on the left side:

$R = V/I$

Now plug in the values:

$R = 3.5/0.02$

Run this through your pocket calculator if you find decimals confusing. The answer is:

$R = 175\Omega$

It so happens that 175Ω isn't a standard value. You may have to settle for 180 or 220Ω, but that's close enough.

Evidently the 470Ω resistor that you used in Experiment 3 was a very conservative choice. I suggested it because I said originally that you could use any LED at all. I figured that no matter which one you picked, it should be safe with 470Ω to protect it.

Cleanup and Recycling

The dead LED can be thrown away. Everything else is reusable.

FUNDAMENTALS

Decimals

Legendary British politician Sir Winston Churchill is famous for complaining about "those damned dots." He was referring to decimal points. Because Churchill was Chancellor of the Exchequer at the time, and thus in charge of all government expenditures, his difficulty with decimals was a bit of a problem. Still, he muddled through in time-honored British fashion, and so can you.

You can also use a pocket calculator—or follow two basic rules.

Doing multiplication: combine the zeros

Suppose you need to multiply 0.04 by 0.005:

1. Count the total number of zeros following both of the decimal points. In this case, three zeros.
2. Multiply the numbers which follow the zeros. In this case, $4 \times 5 = 20$.
3. Write down the result as 0 followed by a decimal point, followed by the number of zeros, followed by the multiplication result. Like this: 0.00020, which is the same as 0.0002.

Doing division: cancel the zeros

Suppose you need to divide 0.006 by 0.0002:

1. Shift the decimal points to the right, in both the numbers, by the same number of steps, until both the numbers are greater than 1. In this case, shift the point four steps in each number, so you get 60 divided by 2.
2. Do the division. The result in this case is 30.

THEORY

Doing the math on your tongue

I'm going to go back to the question I asked in the previous experiment: why didn't your tongue get hot?

Now that you know Ohm's Law, you can figure out the answer in numbers. Let's suppose the battery delivered its rated 9 volts, and your tongue had a resistance of 50K, which is 50,000 ohms. Write down what you know:

$V = 9$

$R = 50{,}000$

We want to know the current, I, so we use the version of Ohm's Law that puts this on the left:

$I = V/R$

Plug in the numbers:

$I = 9/50{,}000 = 0.00018$ amps

Move the decimal point three places to convert to milliamps:

$I = 0.18$ mA

That's a tiny current that will not produce much heat at 9 volts.

What about when you shorted out the battery? How much current made the wires get hot? Well, suppose the wires had a resistance of 0.1 ohms (probably it's less, but I'll start with 0.1 as a guess). Write down what we know:

$V = 1.5$

$R = 0.1$

Once again we're trying to find I, the current, so we use:

$I = V/R$

Plug in the numbers:

$I = 1.5/0.1 = 15$ amps

That's 100,000 times the current that may have passed through your tongue, which would have generated much more heat, even though the voltage was lower.

Could that tiny little battery really pump out 15 amps? Remember that the battery got hot, as well as the wire. This tells us that the electrons may have met some resistance inside the battery, as well as in the wire. (Otherwise, where else did the heat come from?) Normally we can forget about the internal resistance of a battery, because it's so low. But at high currents, it becomes a factor.

I was reluctant to short-circuit the battery through a meter, to try to measure the current. My meter will fry if the current is greater than 10A. However I did try putting other fuses into the circuit, to see whether they would blow. When I tried a 10A fuse, it did not melt. Therefore, for the brand of battery I used, I'm fairly sure that the current in the short circuit was under 10A, but I know it was over 3A, because the 3A fuse blew right away.

The internal resistance of the 1.5-volt battery prevented the current in the short circuit from getting too high. This is why I cautioned against using a larger battery (especially a car battery). Larger batteries have a much lower internal resistance, allowing dangerously high currents which generate explosive amounts of heat. A car battery is designed to deliver literally hundreds of amps when it turns a starter motor. That's quite enough current to melt wires and cause nasty burns. In fact, you can weld metal using a car battery.

Lithium batteries also have low internal resistance, making them very dangerous when they're shorted out. High current can be just as dangerous as high voltage.

FUNDAMENTALS

Watt basics

So far I haven’t mentioned a unit that everyone is familiar with: watts.

A watt is a unit of work. Engineers have their own definition of work—they say that work is done when a person, an animal, or a machine pushes something to overcome mechanical resistance. Examples would be a steam engine pulling a train on a level track (overcoming friction and air resistance) or a person walking upstairs (overcoming the force of gravity).

When electrons push their way through a circuit, they are overcoming a kind of resistance, and so they are doing work, which can be measured in watts. The definition is easy:

watts = volts × amps

Or, using the symbols customarily assigned, these three formulas all mean the same thing:

$W = V \times I$

$V = W/I$

$I = W/V$

Watts can be preceded with an “m,” for “milli,” just like volts:

Number of watts	Usually expressed as	Abbreviated as
0.001 watts	1 milliwatt	1mW
0.01 watts	10 milliwatts	10 mW
0.1 watts	100 milliwatts	100 mW
1 watt	1,000 milliwatts	1W

Because power stations, solar installations, and wind farms deal with much larger numbers, you may also see references to kilowatts (using letter K) and megawatts (with a capital M, not to be confused with the lowercase m used to define milliwatts):

Number of watts	Usually expressed as	Abbreviated as
1,000 watts	1 kilowatt	1 KW
1,000,000 watts	1 megawatt	1 MW

Lightbulbs are calibrated in watts. So are stereo systems. The watt is named after James Watt, inventor of the steam engine. Incidentally, watts can be converted to horsepower, and vice versa.

THEORY

Power assessments

I mentioned earlier than resistors are commonly rated as being capable of dealing with 1/4 watt, 1/2 watt, 1 watt, and so on. I suggested that you should buy resistors of 1/4 watt or higher. How did I know this?

Go back to the LED circuit. Remember we wanted the resistor to drop the voltage by 3.5 volts, at a current of 20 mA. How many watts of power would this impose on the resistor?

Write down what you know:

$V = 3.5$ (the voltage drop imposed by the resistor)

$I = 20\text{mA} = 0.02$ amps (the current flowing through the resistor)

We want to know W, so we use this version of the formula:

$W = V \times I$

Plug in the values:

$W = 3.5 \times 0.02 = 0.07$ watts (the power being dissipated by the resistor)

Because 1/4 watt is 0.25 watts, obviously a 1/4 watt resistor will have about four times the necessary capacity. In fact you could have used a 1/8 watt resistor, but in future experiments we may need resistors that can handle 1/4 watt, and there’s no penalty for using a resistor that is rated for more watts than will actually pass through it.

Experiment 5: Let's Make a Battery

Long ago, before web surfing, file sharing, or cell phones, kids were so horribly deprived that they tried to amuse themselves with kitchen-table experiments such as making a primitive battery by pushing a nail and a penny into a lemon. Hard to believe, perhaps, but true!

This is seriously old-school—but I want you to try it anyway, because anyone who wants to get a feel for electricity should see how easy it is to extract it from everyday objects around us. Plus, if you use enough lemons, you just *might* generate enough voltage to power an LED.

The basic components of a battery are two metal electrodes immersed in an electrolyte. I won't define these terms here (they're explained in the following section "Theory: The nature of electricity"). Right now all you need to know is that lemon juice will be your electrolyte, and copper and zinc will be your electrodes. A penny provides the necessary copper, as long as it is fairly new and shiny. Pennies aren't solid copper anymore, but they are still copper-plated, which is good enough.

To find some metallic zinc, you will have to make a trip to a hardware store, where you should ask for roofing nails. The nails are zinc-plated to prevent them from rusting. Small metal brackets or mending plates also are usually zinc-plated. They should have a slightly dull, silvery look. If they have a mirror-bright finish, they're more likely to be nickel-plated.

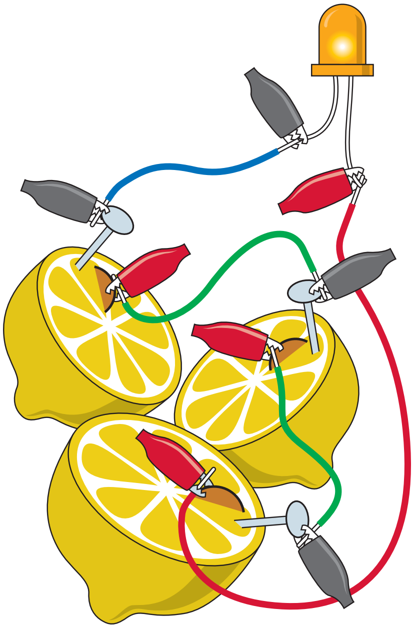

Figure 1-71. *A three-lemon battery. Don't be too disappointed if the LED fails to light up. The lemons have a high electrical resistance, so they can't deliver much current, especially through the relatively small surface area of the nails and the pennies. However, the lemon battery does generate voltage that you can measure with your meter.*

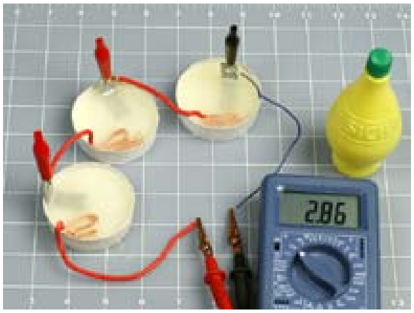

Figure 1-72. *Bottled lemon juice seems to work just as well as fresh lemon juice. I cut the bottoms off three paper cups, inserted a galvanized bracket into each, and used heavyweight stranded copper wire to make the positive electrodes*

Cut a lemon in half, set your multimeter so that it can measure up to 2 volts DC, and hold one probe against a penny while you hold the other probe against a roofing nail (or other zinc-plated object). Now force the penny and the nail into the exposed juicy interior of the lemon, as close to each other as possible, but not actually touching. You should find that your meter detects between 0.8 volts and 1 volt.

You can experiment with different items and liquids to see which works best. Immersing your nail and penny in lemon juice that you have squeezed into a shot glass or egg cup may enhance the efficiency of your battery, although you'll have a harder time holding everything in place. Grapefruit juice and vinegar will work as substitutes for lemon juice.

To drive a typical LED, you need more than 1 volt. How to generate the extra electrical pressure? By putting batteries in series, of course. In other words, more lemons! (Or more shot glasses or egg cups.) You'll also need lengths of wire to connect multiple electrodes, and this may entail skipping ahead to Chapter 2, where I describe how to strip insulation from hookup wire. Figures 1-71 and 1-72 show the configuration.

If you set things up carefully, making sure than none of the electrodes are touching, you may be able to illuminate your LED with two or three lemon-juice batteries in series. (Some LEDs are more sensitive to very low currents than others. Later in the book I'll be talking about very-low-current LEDs. If you want your lemon-juice battery to have the best chance of working, you can search online for low-current LEDs and buy a couple.)

THEORY

The nature of electricity

To understand electricity, you have to start with some basic information about atoms. Each atom consists of a nucleus at the center, containing protons, which have a positive charge. The nucleus is surrounded by electrons, which carry a negative charge.

Breaking up the nucleus of an atom requires a lot of energy, and can also liberate a lot of energy—as happens in a nuclear explosion. But persuading a couple of electrons to leave an atom (or join an atom) takes very little energy. For instance, when zinc reacts chemically with an acid, it can liberate electrons. This is what happens at the zinc electrode of the chemical battery in Experiment 5.

The reaction soon stops, as electrons accumulate on the zinc electrode. They feel a mutual force of repulsion, yet they have nowhere to go. You can imagine them like a crowd of hostile people, each one wanting the others to leave, and refusing to allow new ones to join them, as shown in Figure 1-73.

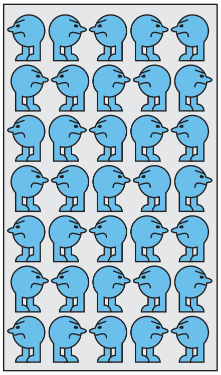

Figure 1-73. *Electrons on an electrode have a bad attitude known as mutual repulsion.*

Now consider what happens when a wire connects the zinc electrode, which has a surplus of electronics, to another electrode, made from a different material, that has a shortage of electrons. The electrons can pass through the wire very easily by jumping from one atom to the next, so they escape from the zinc electrode and run through the wire, propelled by their great desire to get away from each other. See Figure 1-74. This mutual force of propulsion is what creates an electrical current.

Now that the population of electrons on the zinc electrode has been reduced, the zinc-acid reaction can continue, replacing the missing electrodes with new ones—which promptly imitate their predecessors and try to get away from each other by running away down the wire. The process continues until the zinc-acid reaction grinds to a halt, usually because it creates a layer of a compound such as zinc oxide, which won't react with acid and prevents the acid from reacting with the zinc underneath. (This is why your zinc electrode may have looked sooty when you pulled it out of the acidic electrolyte.)

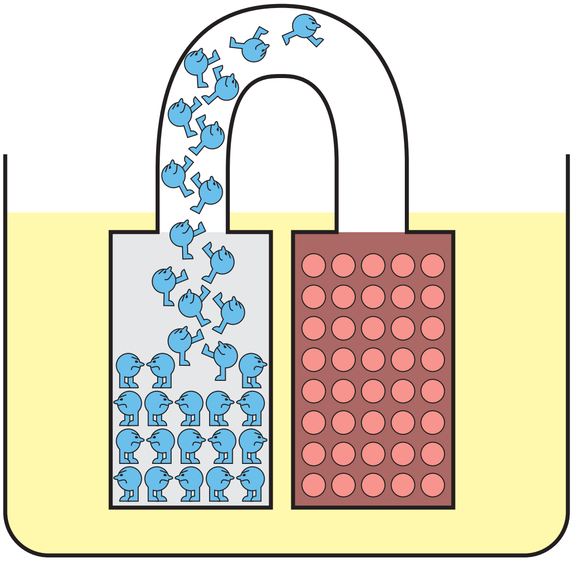

Figure 1-74. *As soon as we open up a pathway from a zinc electrode crowded with electrons to a copper electrode, which contains "holes" for the electrons, their mutual repulsion makes them try to escape from each other to their new home as quickly as possible.*

This description applies to a "primary battery," meaning one that is ready to generate electricity as soon as a connection between its terminals allows electrons to transfer from one electrode to the other. The amount of current that a primary battery can generate is determined by the speed at which chemical reactions inside the battery can liberate electrons. When the raw metal in the electrodes has all been used up in chemical reactions, the battery can't generate any more electricity and is dead. It cannot easily be recharged, because the chemical reactions are not easily reversible, and the electrodes may have oxidized.

In a rechargeable battery, also known as a secondary battery, a smarter choice of electrodes and electrolyte does allow the chemical reactions to be reversed.

How much current is being generated in your lemon battery? Set your meter to measure milliamps, and connect it between the nail and the penny. I measured about 2mA, but got 10mA when I used some #10 stranded copper wire instead of a penny and a large mending plate instead of a roofing nail, immersed in a cup of grapefruit juice. When a larger surface area of metal makes better contact with the electrolyte, you get a greater flow of current. (Don't ever connect your meter to measure amps directly between the terminals of a real battery. The current will be too high, and can blow the fuse inside your meter.)

What's the internal resistance of your lemon? Put aside the copper and zinc electrodes and insert your nickel-plated meter probes into the juice. I got a reading of around 30K when both probes were in the same segment of the lemon, but 40K or higher if the probes were in different segments. Is the resistance lower when you test liquid in a cup?

Here are a couple more questions that you may wish to investigate. For how long will your lemon battery generate electricity? And why do you think your zinc-plated electrode becomes discolored after it has been used for a while?

Electricity is generated in a battery by an exchange of ions, or free electrons, between metals. If you want to know more about this, check the section "Theory: The nature of electricity" on the previous page.

Cleanup and Recycling

The hardware that you immersed in lemons or lemon juice may be discolored, but it is reusable. Whether you eat the lemons is up to you.

BACKGROUND

Positive and negative

If electricity is a flow of electrons, which have a negative charge, why do people talk as if electricity flows from the positive terminal to the negative terminal of a battery?

The answer lies in a fundamental embarrassment in the history of research into electricity. For various reasons, when Benjamin Franklin was trying to understand the nature of electric current by studying phenomena such as lightning during thunderstorms, he believed he observed a flow of "electrical fluid" from positive to negative. He proposed this concept in 1747.

In fact, Franklin had made an unfortunate error that remained uncorrected until after physicist J. J. Thomson announced his discovery of the electron in 1897, 150 years later. Electricity actually flows from an area of greater negative charge, to some other location that is "less negative"—that is, "more positive." In other words, electricity is a flow of negatively charged particles. In a battery, they originate from the negative terminal and flow to the positive terminal.

You might think that when this fact was established, everyone should have discarded Franklin's idea of a flow from positive to negative. But when an electron moves through a wire, you can still think of an equal positive charge flowing in the opposite direction. When the electron leaves home, it takes a small negative charge with it; therefore, its home becomes a bit more positive. When the electron arrives at its destination, its negative charge makes the destination a bit less positive. This is pretty much what would happen if an imaginary positive particle traveled in the opposite direction. Moreover, all of the mathematics describing electrical behavior are still valid if you apply them to the imaginary flow of positive charges.

As a matter of tradition and convenience we still retain Ben Franklin's erroneous concept of flow from positive to negative, because it really makes no difference. In the symbols that represent components such as diodes and transistors, you will actually find arrows reminding you which way these components should be placed—and the arrows all point from positive to negative, even though that's not the way things really work at all! Ben Franklin would have been surprised to learn that although most lightning strikes occur when a negative charge in clouds discharges to neutralize a positive charge on the ground, some forms of lightning are actually a flow of electrons from the negatively charged surface of the earth, up to a positive charge in the clouds. That's right: someone who is "struck by lightning" may be hurt by *emitting* electrons rather than by receiving them, as shown in Figure 1-75.

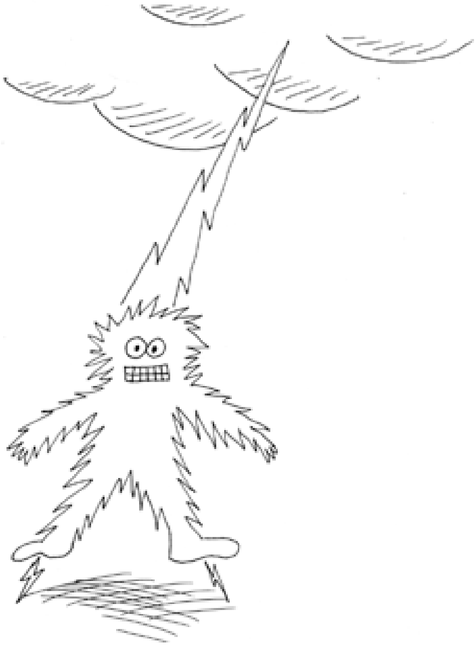

Figure 1-75. *In some weather conditions, the flow of electrons during a lightning strike can be from the ground, through your feet, out of the top of your head, and up to the clouds. Benjamin Franklin would have been surprised.*

THEORY

Basic measurements

Electrical potential is measured by adding up the charges on individual electrons. The basic unit is the *coulomb*, equal to the total charge on about 6,250,000,000,000,000,000 electrons.

If you know how many electrons pass through a piece of wire each second, this establishes the flow of electricity, which can be expressed in amperes. In fact 1 ampere can be defined as 1 coulomb per second. Thus:

1 ampere = 1 coulomb/second
= about 6.25 quintillion electrons/second

There's no way to "see" the number of electrons running through a conductor (Figure 1-76), but there are indirect ways of getting at this information. For instance, when an electron goes running through a wire, it creates a wave of electromagnetic force around it. This force can be measured, and we can calculate the amperage from that. The electric meter installed at your home by the utility company functions on this principle.

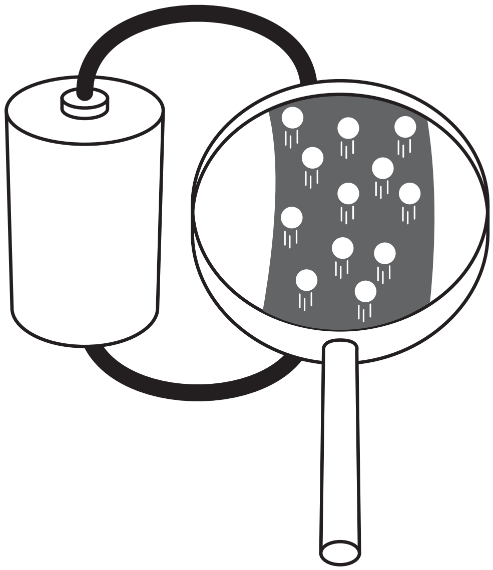

Figure 1-76. *If you could look inside an electric wire with a sufficiently powerful magnifying device, and the wire happened to be carrying 1 ampere of electron flow at the time, you might hope to see about 6.25 quintillion electrons speeding past each second.*

If electrons are just moving freely, they aren't doing any work. If you had a loop of wire of zero resistance, and you kick-started a flow of electrons somehow, they could just go buzzing around forever. (This is what happens inside a superconductor—almost.)

Under everyday conditions, even a copper wire has some resistance. The force that we need to push electrons through it is known as "voltage," and creates a flow that can create heat, as you saw when you shorted out a battery. (If the wire that you used had zero resistance, the electricity running through it would not have created any heat.) We can use the heat directly, as in an electric stove, or we can use the electrical energy in other ways—to run a motor, for instance. Either way, we are taking energy out of the electrons, to do some work.

One volt can be defined as the amount of pressure that you need to create a flow of 1 ampere, which does 1 watt of work. As previously defined, 1 watt = 1 volt × 1 ampere, but the definition actually originated the other way around:

1 volt = 1 watt/1 ampere

It's more meaningful this way, because a watt can be defined in nonelectrical terms. Just in case you're interested, we can work backward through the units of the metric system like this:

1 watt = 1 joule/second
1 joule = a force of 1 newton acting through 1 meter
1 newton = the force required to accelerate 1 kilogram by 1 meter per second, each second

On this basis, the electrical units can all be anchored with observations of mass, time, and the charge on electrons.

Practically Speaking

For practical purposes, an intuitive understanding of electricity can be more useful than the theory. Personally I like the water analogies that have been used for decades in guides to electricity. Figure 1-77 shows a tall tank half full of water, with a hole punched in it near the bottom. Think of the tank as being like a battery. The height of the water is comparable to voltage. The volume of flow through the hole, per second, is comparable to amperage. The smallness of the hole is comparable to resistance. See Figure 1-79 on the next page.

Where's the wattage in this picture? Suppose we place a little water wheel where it is hit by the flow from the hole. We can attach some machinery to the water wheel. Now the flow is doing some work. (Remember, wattage is a measurement of work.)

Maybe this looks as if we're getting something for nothing, extracting work from the water wheel without putting any energy back into the system. But remember, the water level in the tank is falling. As soon as I include some helpers hauling the waste water back up to the top of the tank (in Figure 1-78), you see that we have to put work in to get work out.

Similarly, a battery may seem to be giving power out without taking anything in, but the chemical reactions inside it are changing pure metals into metallic compounds, and the power we get out of a battery is enabled by this change of state. If it's a rechargeable battery, we have to push power back into it to reverse the chemical reactions.

Going back to the tank of water, suppose we can't get enough power out of it to turn the wheel. One answer could be to add more water. The height of the water will create more force. This would be the same as putting two batteries end to end, positive to negative, in series, to double the voltage. See Figure 1-80. As long as the resistance in the circuit remains the same, greater voltage will create more amperage, because amperage = voltage/resistance.

What if we want to run two wheels instead of one? We can punch a second hole in the tank, and the force (voltage) will be the same at each of them. However, the water level in the tank will drop twice as fast. Really, we'd do better to build a second tank, and here again the analogy with a battery is good. If you wire two batteries side by side, in parallel, you get the same voltage, but for twice as long. The two batteries may also be able to deliver more current than if you just used one. See Figure 1-81.

Summing up:

- Two batteries in series deliver twice the voltage.
- Two batteries in parallel can deliver twice the current.

All right, that's more than enough theory for now. In the next chapter, we'll continue with some experiments that will build on the foundations of knowledge about electricity, to take us gradually toward gadgets that can be fun and useful.

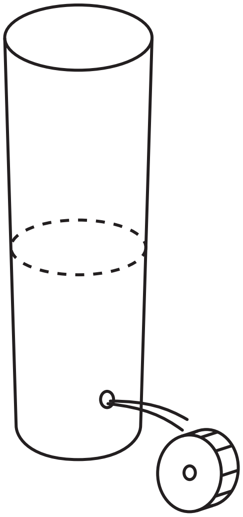

Figure 1-77. *If you want to get work out of a system...*

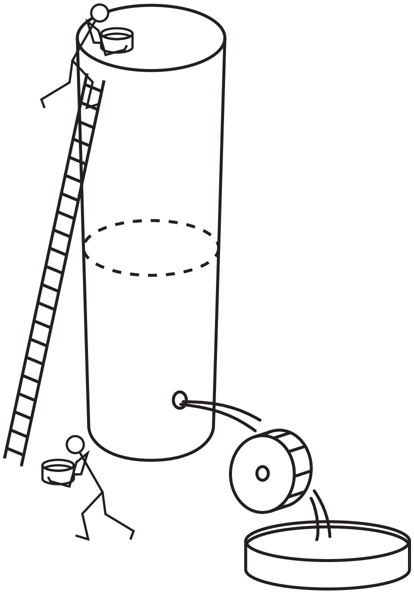

Figure 1-78. *. . . somehow or other you have to put work back into it.*

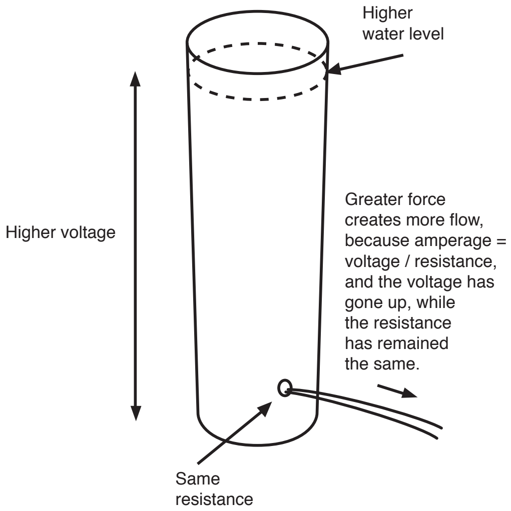
Higher water level
Higher voltage
Greater force creates more flow, because amperage = voltage / resistance, and the voltage has gone up, while the resistance has remained the same.
Same resistance

Figure 1-79. *Greater force generates more flow, as long as the resistance remains the same.*

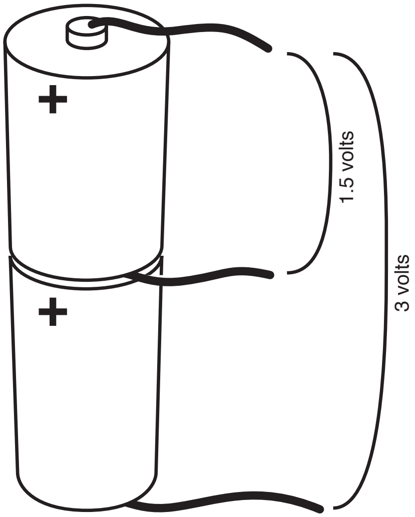
+
+
1.5 volts
3 volts

Figure 1-80. *When you place two equal batteries in series, you double the voltage.*

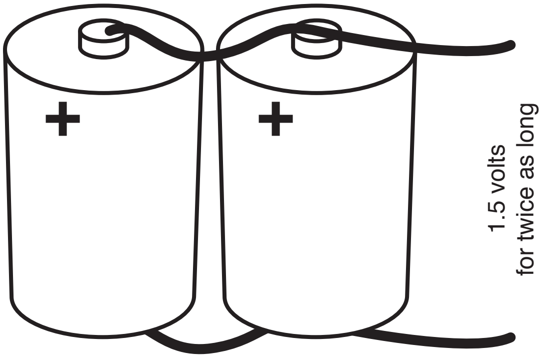
+
+
1.5 volts
for twice as long

Figure 1-81. *Two equal batteries that are wired in parallel will deliver the same voltage for twice as long as one.*

2

Switching Basics and More

The concept of switching is fundamental in electronics, and I'm not just talking about power switches. By "switching," I mean using one flow of electricity to switch, or control, another. This is such an important principle that no digital device can exist without it.

Today, switching is mostly done with semiconductors. Before I deal with them, I'll back up and illustrate the concept by introducing you to relays, which are easier to understand, because you can see what's going on inside them. And before I get to relays, I'll deal with everyday on/off switches, which may seem very simple—but we have to nail down the basics.

Also in this chapter, I'll deal with capacitance, because capacitance and resistance are fundamental to electronic circuits. By the end of the chapter, you should have a basic grounding in electronics and be able to build the noise-making section of a simple intrusion alarm. This will be your first circuit that does something genuinely useful!

IN THIS CHAPTER

Shopping List: Experiments 6 Through 11
Experiment 6: Very Simple Switching
Experiment 7: Relay-Driven LEDs
Experiment 8: A Relay Oscillator
Experiment 9: Time and Capacitors
Experiment 10: Transistor Switching
Experiment 11: A Modular Project

Shopping List: Experiments 6 Through 11

As in the previous shopping list, you should visit the various online suppliers for availability and pricing of components and devices. Manufacturers seldom sell small numbers of parts directly. Check the appendix for a complete list of URLs for all the companies mentioned here.

Devices

- Power supply/universal AC adapter, 3 to 12 volts at 1A (1,000 mA). See Figure 2-1. Part number 273-029 from RadioShack, part PH-62092 by Philips, or similar.
- Breadboard suitable for integrated circuits. Quantity: 1. See Figures 2-2 and 2-3. Part 276-002 from RadioShack, model 383-X1000 made by PSP, part 923252-I by 3M, or similar. A breadboard that has screw terminals mounted beside it will be a little easier to use but more expensive than one that does not have terminals.

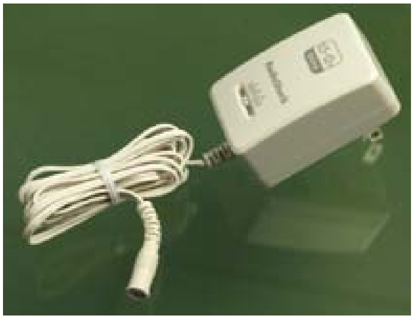

Figure 2-1. *This AC adapter delivers DC from 3 to 12 volts, and is ideal for electronics projects.*

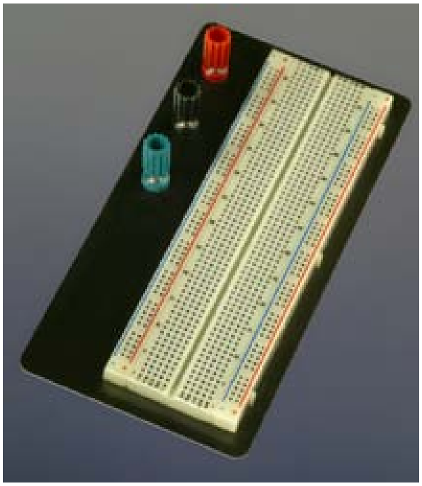

Figure 2-2. *This "breadboard" for quickly constructing electronic circuits has a metal base, and screw terminals for attaching wires from a power supply.*

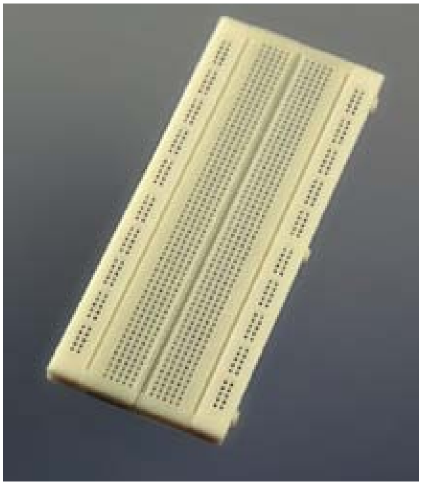

Figure 2-3. *A breadboard without screw terminals is almost as convenient, and is cheaper.*

Tools

Wire strippers

Ideal model 45-121 wire strippers for 16- to 26-gauge wire, or similar. See Figure 2-4. (The "gauge" of the wire tells you how thick it is. A higher gauge means a thinner wire. In this book, we will mainly be using thin wire of 20- to 24-gauge.)

You may also consider the Kronus Automatic Wire Strippers, part 64-2981 from RadioShack, or GB Automatic Wire Strippers, part SE-92 from Amazon.com. See Figure 2-5.

The Kronus and GB wire strippers are functionally identical. The advantage of their design is that it enables you to strip insulation from a wire with one hand. But they do not work well on really thin wire.

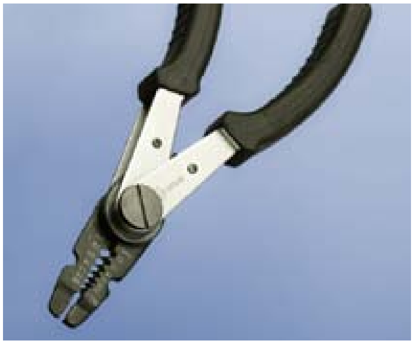

Figure 2-4. *To use these wire strippers, insert a piece of insulated wire in the appropriate-sized hole between the jaws, grip the handles, and pull a section of insulation away. See page 45.*

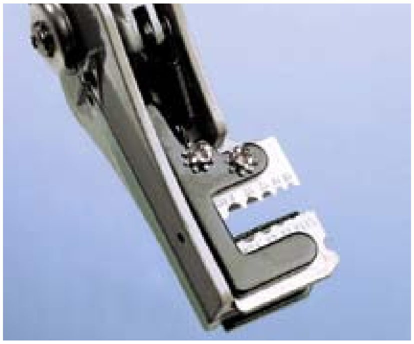

Figure 2-5. *These automatic wire strippers enable one-handed operation, but are not suitable for very small wire diameters. See page 44.*

Supplies

Hookup wire

Solid-conductor, 22-gauge, minimum 25 feet of each color. See Figure 2-6. Part 278-1222 from RadioShack, catalog item 9948T17 from McMaster-Carr, or check eBay for deals.

It's easy to buy the wrong kind of wire. You need solid-core wire, which has a single conductor inside the plastic insulation, not stranded, which has multiple, thinner conductors. See Figures 2-7 and 2-8. You're going to be pushing wires into little holes in a "breadboard," and stranded wire won't let you do this. You will also have problems if you buy wire thicker than 22-gauge. Remember: the lower the gauge number, the thicker the wire.

For a little extra money, you can buy an assortment of precut sections of wire, with ends stripped and ready for use. Try catalog item JW-140 (jumper wire assortment) from All Electronics or search eBay for "breadboard wire." See Figure 2-9.

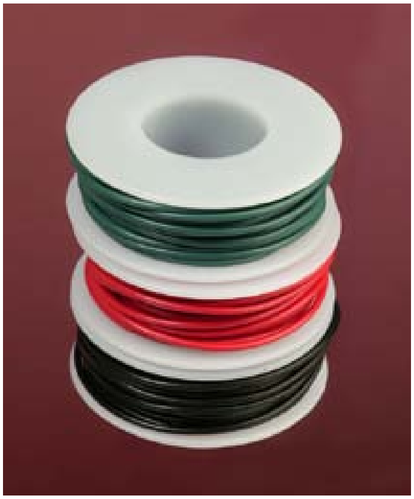

Figure 2-6. *Using hookup wire with different colors of plastic insulation will help you to distinguish one wire from another in your circuits.*

Patch cords

Patch cords are not strictly necessary but very convenient. You don't want audio or video patch cords, which have a plug on each end; you want wires with alligator clips on each end, also sometimes referred to as "test leads." Try catalog item 461-1176-ND from Digi-Key or catalog item MTL-10 from All Electronics. See Figure 2-10.

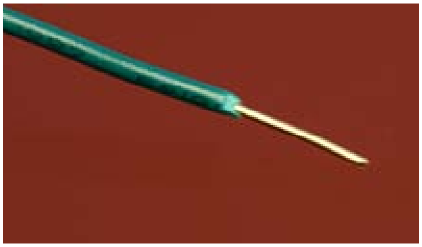

Figure 2-7. *Solid-conductor wire of 22 or 24 gauge is suitable for most of the experiments in this chapter.*

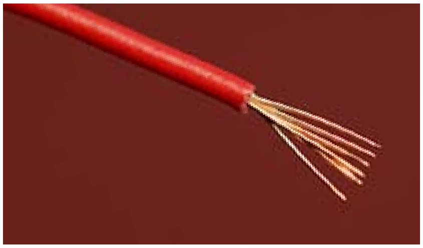

Figure 2-8. *Stranded is more flexible, but cannot be used easily with breadboards.*

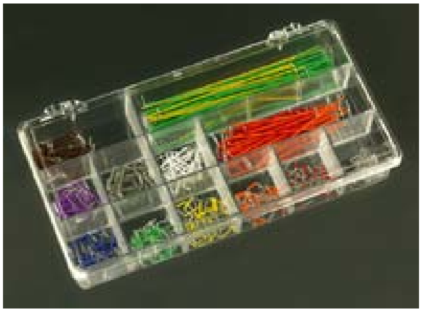

Figure 2-9. *Precut wires with stripped ends can save a lot of time and trouble—if you don't mind paying a little extra.*

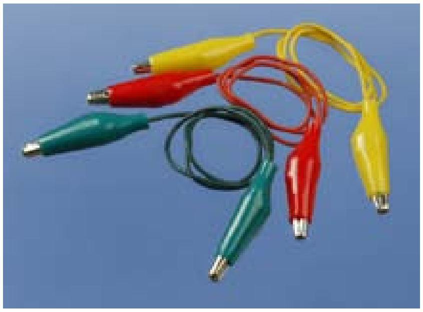

Figure 2-10. *Patch cords, sometimes known as test leads, consist of wires preattached to alligator clips. This is another of those little luxuries that reduces the hassle factor in hobby electronics.*

Components

Pushbutton

Momentary-on, SPST, sometimes referred to as OFF-(ON) or (ON)-OFF type. Must be PCB- or PC-mount, meaning is extremely small with thin spiky contacts on the bottom. Quantity: 1. See Figure 2-11.

Examples are part number AB11AP by NKK, part MPA103B04 by Alcoswitch, or part EP11SD1CBE by C&K. If you have a choice, buy the cheapest, as we're going to switch very low current.

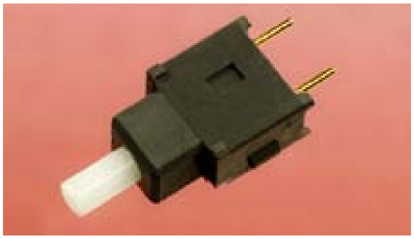

Figure 2-11. *The terminals protruding from this tiny pushbutton are spaced 0.2 inches apart, making it ideal for the "breadboard" that you'll be using.*

Switches

Toggle switch, single-pole, double-throw (SPDT), sometimes referred to as ON-ON type. Quantity: 2. See Figure 2-12.

Model S302T-RO by NKK is ideal; it has screw terminals that will eliminate the need for alligator clips. Other options are catalog item MTS-4PC from All Electronics or part 275-603 from RadioShack.

We won't be switching large currents or high voltages, so the exact type of switch is unimportant. However, the terminals on larger-size switches are spaced wider apart, which makes them easier to deal with.

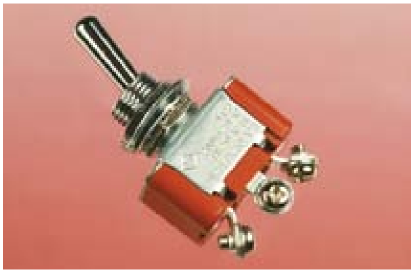

Figure 2-12. *This relatively large toggle switch made by NKK has screw terminals, which will reduce the inconvenience of attaching it to hookup wire.*

Relays

DPDT, nonlatching, 12v DC. Quantity: 2.

It's important to get the right kind of relay—one whose configuration matches the pictures I'll be using. Look for parts FTR-F1CA012V or FTR-F1CD012V by Fujitsu, G2RL-24-DC12 by Omron, or OMI-SH-212D by Tyco. Avoid substitutions.

Potentiometer

1 megohm linear potentiometer, Part number 271-211 from RadioShack, part number 24N-1M-15R-R from Jameco, or similar.

Transistors

NPN transistor, general-purpose, such as 2N2222 by STMicroelectronics, part PN2222 by Fairchild, or part 2N2222 from RadioShack. Quantity: 4. See Figure 2-13.

2N6027 programmable unijunction transistor manufactured by On Semiconductor or Motorola. Quantity: 4.

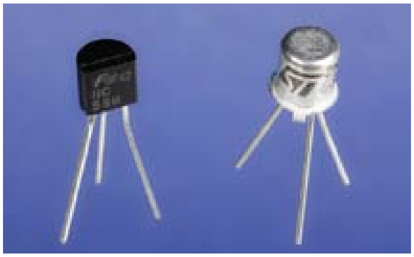

Figure 2-13. *Transistors are commonly sold either in little metal cans or sealed into little lumps of plastic. For our purposes, the packaging makes no difference.*

Capacitors

Electrolytic capacitors, assorted. Must be rated for a minimum of 25 volts and include at least one capacitor of 1,000 µF (microfarads). If you search on eBay, make sure you find *electrolytic* capacitors. If they're rated for higher voltages, that's OK, although they will be physically larger than you need. See Figure 2-14.

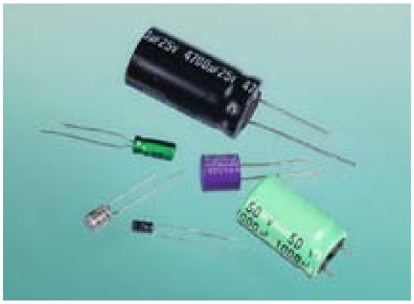

Figure 2-14. *An assortment of electrolytic capacitors.*

Ceramic capacitors, assorted. Make sure you get at least one rated at 0.0047 µF (which can also be written as 47 nF). See Figure 2-15.

Resistors

If you bought only a minimal selection for experiments 1 through 5, now's the time to buy a larger assortment, so that you won't be stuck needing the one value that you don't have. 1/4-watt minimum.

Loudspeaker

Any 8Ω, 1-inch miniature loudspeaker such as part 273-092 from RadioShack. See Figure 2-16.

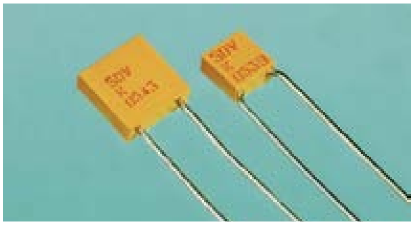

Figure 2-15. *Ceramic capacitors mostly look like this, although many of them are round or bead-shaped instead of square. The packaging shape is unimportant to us.*

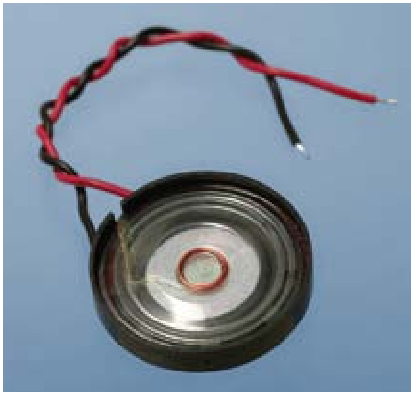

Figure 2-16. *This miniature loudspeaker, just over 1 inch in diameter, is useful for verifying audio output direct from transistor circuits.*

Experiment 6: Very Simple Switching

You will need:

- AA batteries. Quantity: 4.
- Battery carrier for 4 AA batteries. Quantity: 1.
- LED. Quantity: 1.
- Toggle switches, SPDT. Quantity: 2. See Figure 2-12.
- 220Ω or similar value resistor, 1/4-watt minimum. Quantity: 1.
- Alligator clips. Quantity: 8.
- Wire or patch cords. See Figure 2-10, shown previously.
- Wire cutters and wire strippers if you don't use patch cords. See Figure 2-4, shown previously.

In Experiment 3, you illuminated an LED by attaching a battery, and switched it off by removing the battery. For greater convenience our circuits should have proper switches to control power, and while I'm dealing with the general topic of switches, I'm going to explore all the varieties, using a circuit to suggest some possibilities.

Assemble the parts as shown in Figures 2-17 and 2-18. The long lead on the LED must connect with the resistor, because that is the more positive side of the circuit.

You'll notice that you have to include a couple lengths of wire. I suggest green wire to remind you that these sections are not connected directly to positive or to negative power. But you can use any color you like. You can also substitute patch cords, if you have them. However, learning to strip insulation from pieces of wire is a necessary skill, so let's deal with that now.

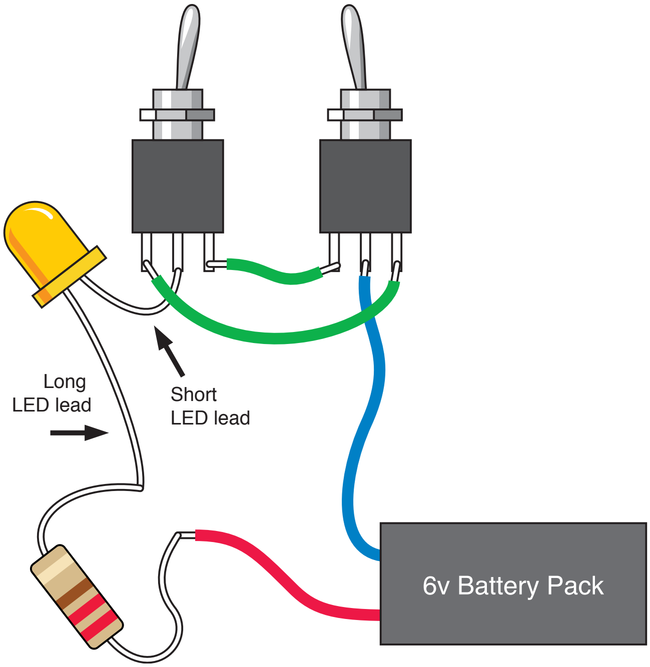

Figure 2-17. *If the LED is on, flipping either of the switches will turn it off. If the LED is off, either of the switches will turn it on. Use alligator clips to attach the wires to each other, and to the switches if your switches don't have screw terminals. Be careful that the clips don't touch each other.*

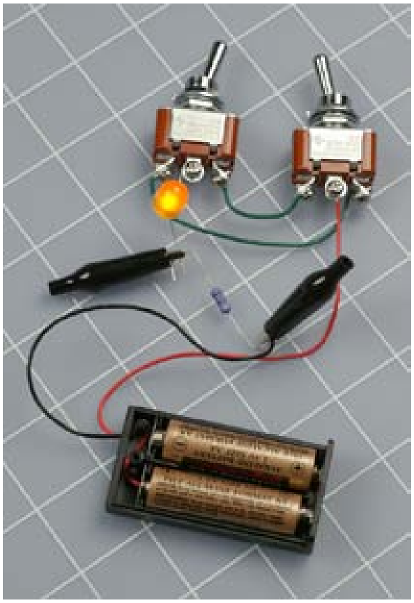

Figure 2-18. *Full-size toggle switches with screw terminals make it easy to hook up this simple circuit.*

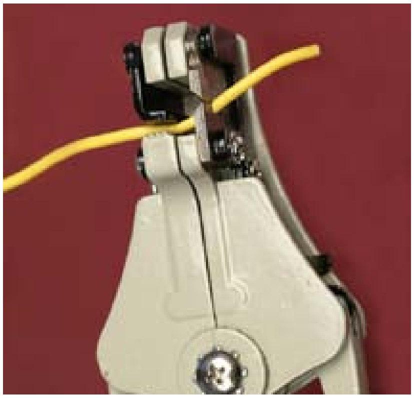

Figure 2-19. *Using automatic wire strippers, when you squeeze the handles the jaw on the left clamps the wire, the sharp grooves on the right bite into the insulation. Squeeze harder and the jaws pull away from each other, stripping the insulation from the wire.*

Tools

If automatic wire strippers (Figure 2-19) don't grip skinny 22-gauge wire very effectively, try the Ideal brand of wire strippers shown back in Figure 2-4, or use plain and simple wire cutters as shown in Figure 2-20. When using wire cutters, you hold the wire in one hand and apply the tool in your other hand, squeezing the handles with moderate pressure—just enough to bite into the insulation, but not so much that you chop the wire. Pull the wire down while you pull the cutters up, and with a little practice you can rip the insulation off to expose the end of the wire.

Macho hardware nerds may use their teeth to strip insulation from wires. When I was younger, I used to do this. I have two slightly chipped teeth to prove it. Really, it's better to use the right tool for the job.

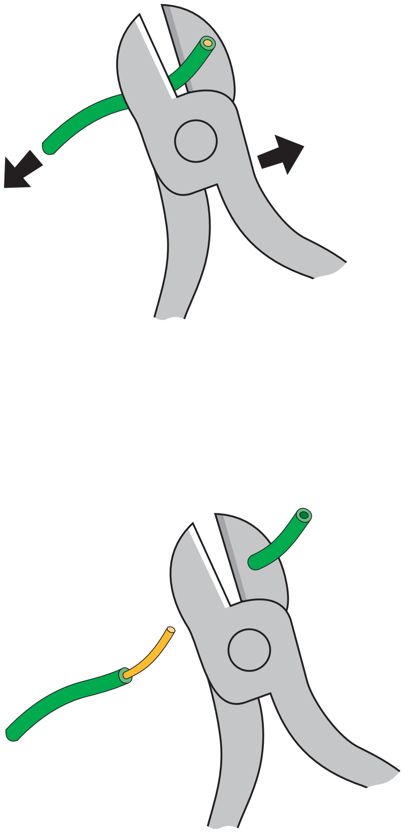

Figure 2-20. *To remove insulation from the end of a thin piece of wire, you can also use wire cutters. This takes a little practice.*

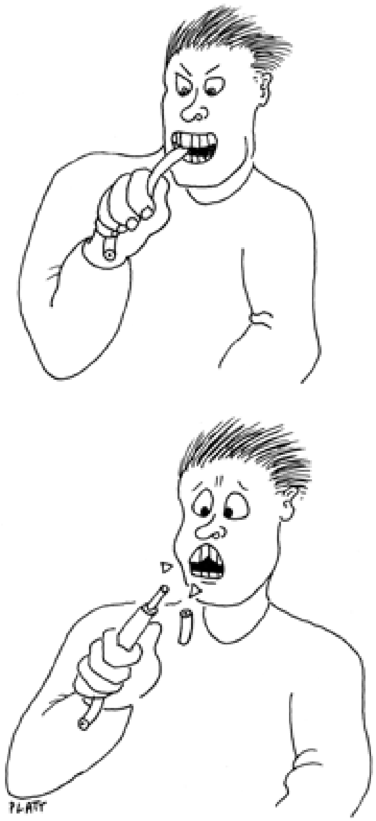

Figure 2-21. *Those who tend to misplace tools, and feel too impatient to search for them, may feel tempted to use their teeth to strip insulation from wire. This may not be such a good idea.*

Connection Problems

Depending on the size of toggle switches that you are using, you may have trouble fitting in all the alligator clips to hold the wires together. Miniature toggle switches, which are more common than the full-sized ones these days, can be especially troublesome (see Figure 2-22). Be patient: fairly soon we'll be using a breadboard, which will eliminate alligator clips almost completely.

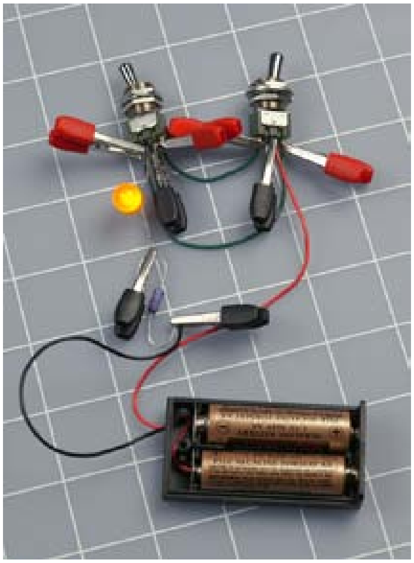

Figure 2-22. *Miniature toggle switches can be used—ideally, with miniature alligator clips—but watch out for short circuits.*

Testing

Make sure that you connect the LED with its long wire toward the positive source of power (the resistor, in this case). Now flip either of the toggle switches. If the LED was on, it will go off, and if it was off, it will go on. Flip the other toggle switch, and it will have the same effect. If the LED does not go on at all, you've probably connected it the wrong way around. Another possibility is that two of your alligator clips may have shorted out the battery.

Assuming your two switches do work as I described them, what's going on here? It's time to nail down some basic facts.

FUNDAMENTALS

All about switches

When you flip the type of toggle switch that you used in Experiment 6, it connects the center terminal with one of the outer terminals. Flip the switch back, and it connects the center terminal with the other outer terminal, as shown in Figure 2-23.

The center terminal is called the pole of the switch. Because you can flip, or throw, this switch to make two possible connections, it is called a *double-throw switch*. As mentioned earlier, a single-pole, double-throw switch is abbreviated *SPDT*.

Some switches are on/off, meaning that if you throw them in one direction they make a contact, but in the other direction, they make no contact at all. Most of the light switches in your house are like this. They are known as *single-throw switches*. A single-pole, single-throw switch is abbreviated *SPST*.

Some switches have two entirely separate poles, so you can make two separate connections simultaneously when you flip the switch. These are called *double-pole switches*. Check the photographs in Figures 2-24 through 2-26 of old-fashioned "knife" switches (which are still used to teach electronics to kids in school) and you'll see the simplest representation of single and double poles, and single and double throws. Various toggle switches that have contacts sealed inside them are shown in Figure 2-27.

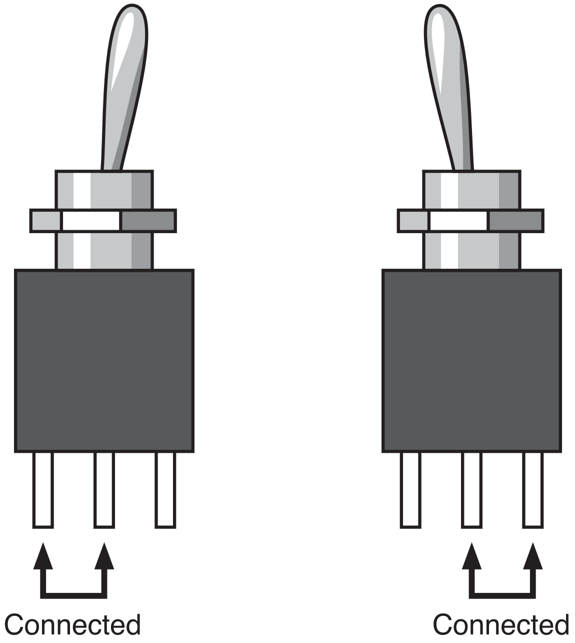

Figure 2-23. *The center terminal is the pole of the switch. When you flip the toggle, the pole changes its connection.*

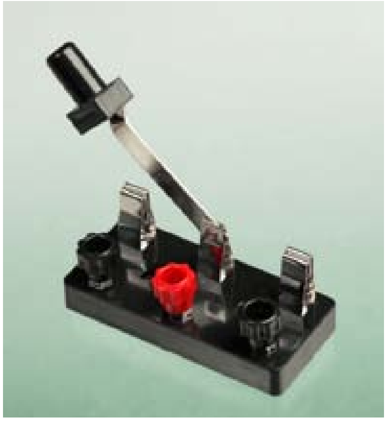

Figure 2-24. *This primitive-looking single-pole, double-throw switch does exactly the same thing as the toggle switches in Figures 2-23 and 2-27.*

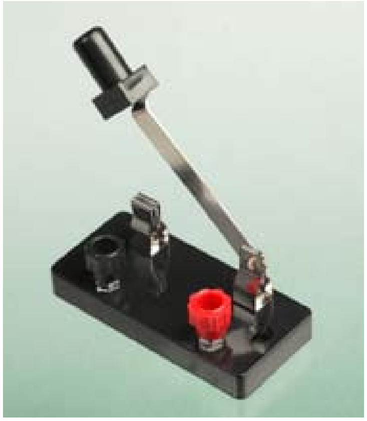

Figure 2-25. *A single-pole, single-throw switch makes only one connection with one pole. Its two states are simply open and closed, on and off.*

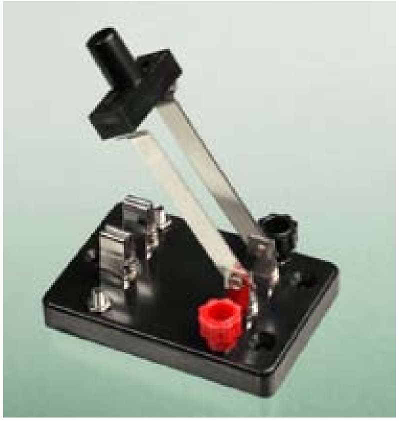

Figure 2-26. *A double-pole, single-throw switch makes two separate on/off connections.*

FUNDAMENTALS

All about switches (continued)

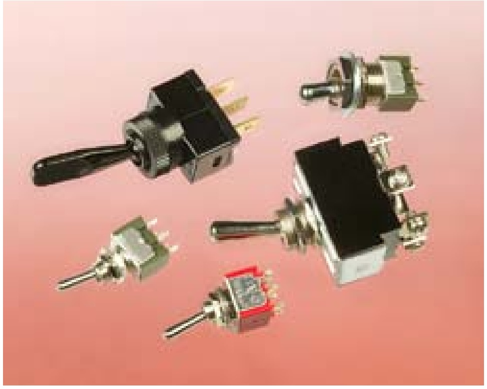

Figure 2-27. *These are all toggle switches. Generally, the larger the switch, the more current it can handle.*

To make things more interesting, you can also buy switches that have three or four poles. (Some rotary switches have even more, but we won't be using them.) Also, some double-throw switches have an additional "center off" position.

Putting all this together, I made a table of possible types of switches (Figure 2-28). When you're reading a parts catalog, you can check this table to remind yourself what the abbreviations mean.

	Single Pole	Double Pole	3-Pole	4-Pole
Single Throw	SPST ON-OFF	DPST ON-OFF	3PST ON-OFF	4PST ON-OFF
Double Throw	SPDT ON-ON	DPDT ON-ON	3PDT ON-ON	4PDT ON-ON
Double Throw with Center Off	SPDT ON-OFF-ON	DPDT ON-OFF-ON	3PDT ON-OFF-ON	4PDT ON-OFF-ON

Figure 2-28. *This table summarizes all the various options for toggle switches and pushbuttons.*

Now, what about pushbuttons? When you press a door bell, you're making an electrical contact, so this is a type of switch—and indeed the correct term for it is a momentary switch, because it makes only a momentary contact. Any spring-loaded switch or button that wants to jump back to its original position is known as a momentary switch. We indicate this by putting its momentary state in parentheses. Here are some examples:

- OFF-(ON): Because the ON state is in parentheses, it's the momentary state. Therefore, this is a single-pole switch that makes contact only when you push it, and flips back to make no contact when you let it go. It is also known as a "normally open" momentary switch, abbreviated "NO."
- ON-(OFF): The opposite kind of momentary single-pole switch. It's normally ON, but when you push it, you break the connection. So, the OFF state is momentary. It is known as a "normally closed" momentary switch, abbreviated "NC."
- (ON)-OFF-(ON): This switch has a center-off position. When you push it either way, it makes a momentary contact, and returns to the center when you let it go.

Other variations are possible, such as ON-OFF-(ON) or ON-(ON). As long as you remember that parentheses indicate the momentary state, you should be able to figure out what these switches are.

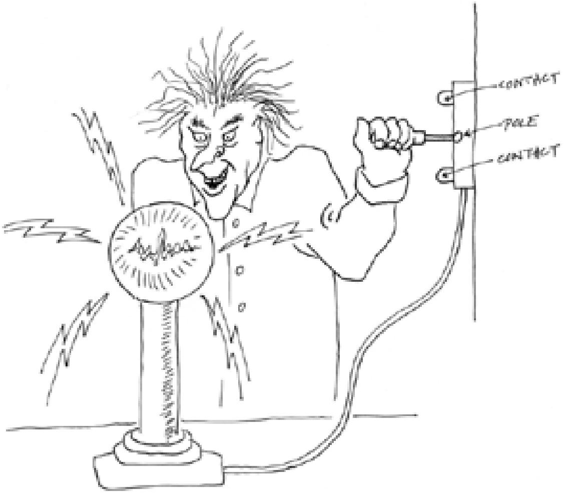

Figure 2-29. *This evil mad scientist is ready to apply power to his experiment. For this purpose, he is using a single-pole, double-throw knife switch, conveniently mounted on the wall of his basement laboratory.*

FUNDAMENTALS

All about switches (continued)

Sparking

When you make and break an electrical connection, it tends to create a spark. Sparking is bad for switch contacts. It eats them until the switch doesn't make a reliable connection anymore. For this reason, you must use a switch that is appropriate for the voltage and amperage that you are dealing with. Electronic circuits generally are low-current, and low-voltage, so you can use almost any switch, but if you are switching a motor, it will tend to suck an initial surge of current that is at least double the rating of the motor when it is running constantly. You should probably use a 4-amp switch to turn a 2-amp motor on and off.

Checking a switch

You can use your meter to check a switch. Doing this helps you find out which contacts are connected when you turn a switch one way or the other. It's also useful if you have a pushbutton and you can't remember whether it's the type that is normally open (you press it to make a connection) or normally closed (you press it to break the connection). Set your meter to measure ohms, and touch the probes to the switch terminals while you work the switch.

This is a hassle, though, because you have to wait while the meter makes an accurate measurement. When you just want to know whether there is a connection, your meter has a "continuity tester" setting. It beeps if it finds a connection, and stays silent if it doesn't. See Figures 2-30 through 2-32 for examples of meters set to test continuity. Figure 2-33 offers an example of a toggle switch being tested for continuity.

Use the continuity-testing feature on your meter only on circuits or components that have no power in them at the time.

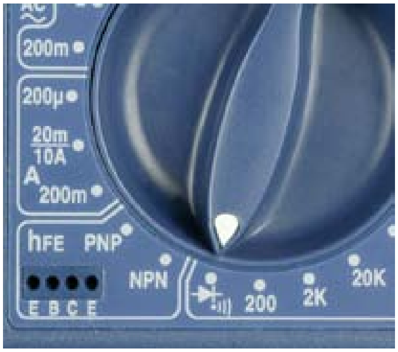

Figure 2-30

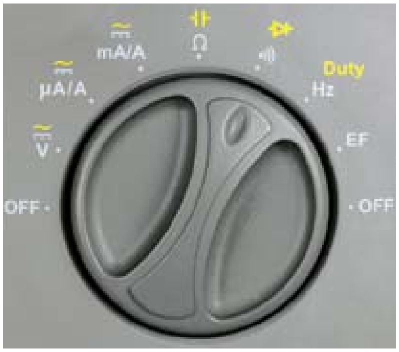

Figure 2-31

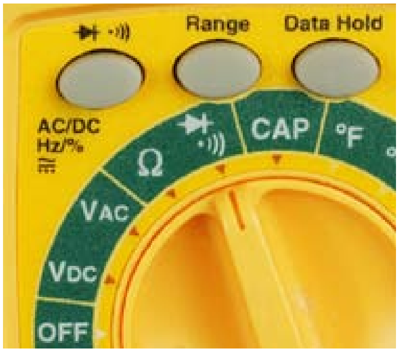

Figure 2-32. *To check a circuit for continuity, turn the dial of your meter to the symbol shown. Only use this feature when there is no power in the component or the circuit that you are testing.*

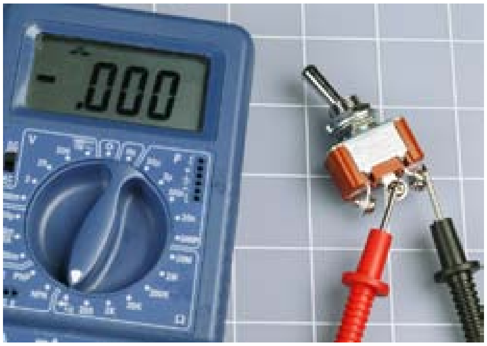

Figure 2-33. *When the switch connects two of its terminals, the meter shows zero resistance between them and will beep if you have set it to verify continuity.*

BACKGROUND

Early switching systems

Switches seem to be such a fundamental feature of our world, and their concept is so simple that it's easy to forget that they went through a gradual process of development and refinement. Primitive knife switches were quite adequate for pioneers of electricity who simply wanted to connect and disconnect electricity to some apparatus in a laboratory, but a more sophisticated approach was needed when telephone systems began to proliferate. Typically, an operator at a "switchboard" needed a way to connect any pair of 10,000 lines on the board. How could it be done?

In 1878, Charles E. Scribner (Figure 2-34) developed the "jack-knife switch," so called because the part of it that the operator held looked like the handle of a jackknife. Protruding from it was a plug, and when the plug was pushed into a socket, it made contact inside the socket. The socket, in fact, was the switch.

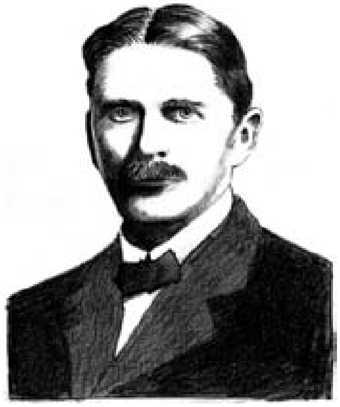

Figure 2-34. *Charles E. Scribner invented the "jack-knife switch" to satisfy the switching needs of telephone systems in the late 1800s. Today's audio jacks still work on the same basis.*[*]

Audio connectors on guitars and amplifiers still work on the same principle, and when we speak of them as being "jacks," the term dates back to Scribner's invention. Switch contacts still exist inside a jack socket.

Today, of course, telephone switchboards have become as rare as telephone operators. First they were replaced with relays—electrically operated switches, which I'll talk about later in this chapter. And then the relays were superceded by transistors, which made everything happen without any moving parts. Before the end of this chapter, you'll be switching current using transistors.

---

[*] The photo on which this drawing is based first appeared in *The History of the Telephone* by Herbert Newton Casson in 1910 (Chicago: A. C. McClurg & Co.).

Introducing Schematics

In Figure 2-35, I've redrawn the circuit from Experiment 6 in a simplified style known as a "schematic." From this point onward, I will be illustrating circuits with schematics, because they make circuits easier to understand. You just need to know a few symbols to interpret them.

Larger versions of all schematics and breadboard photos are available online at this book's website: http://oreilly.com/catalog/9780596153748.

Compare the schematic here with the drawing of the circuit in Figure 2-17. They both show exactly the same thing: Components, and connections between them. The gray rectangles are the switches, the zigzag thing is the resistor, and the symbol with two diagonal arrows is the LED.

The schematic LED symbol includes two arrows indicating that it emits light, because there are some kinds of diodes, which we'll get to later, that don't. The triangle inside the diode symbol always points from positive to negative.

Trace the path that electricity can take through the circuit and imagine the switches turning one way or the other. You should see clearly now why either switch will reverse the state of the LED from on to off or off to on.

This same circuit is used in houses where you have a switch at the bottom of a flight of stairs, and another one at the top, both controlling the same lightbulb. The wires in a house are much longer, and they snake around behind the walls, but because their connections are still the same, they could be represented with the same basic schematic. See Figure 2-36.

A schematic doesn't tell you exactly where to put the components. It just tells you how to join them together. One problem: Different people use slightly different schematic symbols to mean the same thing. Check the upcoming section, "Fundamentals: Basic schematic symbols," for the details.

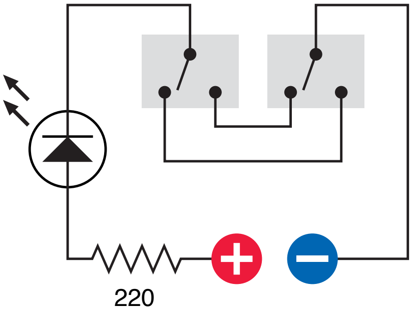

Figure 2-35. *This schematic shows the same circuit as in Figure 2-17 and makes it easier to see how the switches function.*

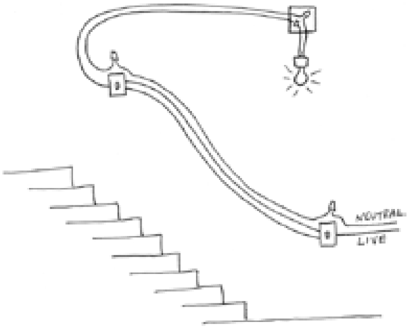

Figure 2-36. *The two-switch circuit shown in Figures 2-17 and 2-35 is often found in house wiring, especially where switches are located at the top and bottom of a flight of stairs. This sketch shows what you might find inside the walls. Wires are joined with "wire nuts" inside boxes that are hidden from everyday view.*

FUNDAMENTALS

Basic schematic symbols

Schematic symbols are like words in a language: they have mutated over the years into a confusing range of variations. A simple on/off (single-pole, single-throw, or SPST) switch, for instance, can be represented by any of the symbols shown in Figure 2-37. They all mean exactly the same thing.

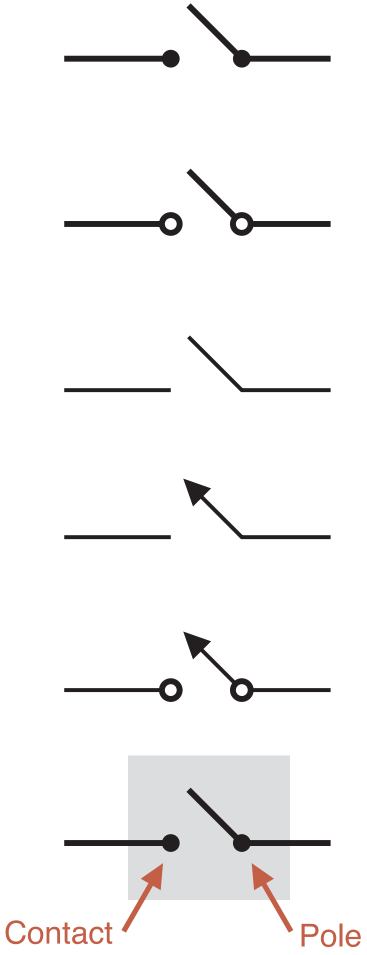

Figure 2-37. *Variations on a theme: Just some of the different styles used to depict a single-pole, single-throw switch in schematic diagrams. The bottom version is the style used in this book.*

Figure 2-38 shows double-pole, double-throw switches. A dotted line indicates a mechanical connection inside the switch, so that when you flip it, you affect both the poles simultaneously. Remember, the poles are electrically isolated from each other.

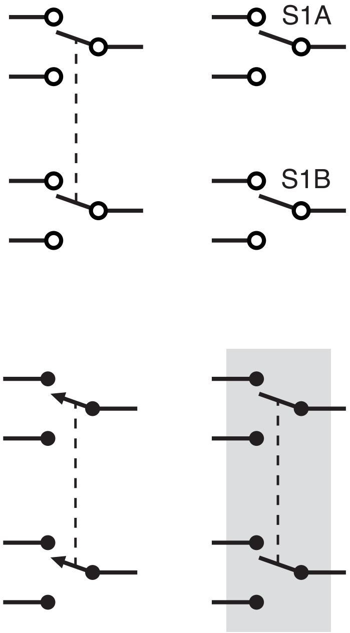

Figure 2-38. *More variations: some different styles for depicting a double-pole, double-throw switch. The style at bottom-right is used in this book.*

Once in a while, you may find a schematic in which switches seem to be scattered around, but the way they are identified (such as S1A, S1B, S1C, and so on) tells you that this is really all one switch with multiple poles.

In the schematics in this book, I'll place a gray rectangle behind each switch. This gray rectangle is not a standard symbol; you won't find it in other books. I'm just including it to remind you that the parts inside are all contained in one package.

A very important stylistic variation in schematics is the way they show whether wires make a connection with each other. Old schematics used to show a little semicircular bump in a wire if it crossed another wire without making a connection. Because modern circuit-drawing software doesn't create this style of schematic, it is no longer often used. The modern style, which you are likely to find if you browse through schematics online, can be summarized like this:

- A dot joining two wires indicates an electrical connection.
- No dot indicates no connection.

FUNDAMENTALS

Basic schematic symbols (continued)

The problem is that this is not very intuitive, especially when you're just beginning to use schematics. When you see two wires crossing, you can easily imagine that they are making a connection, even though there's no dot at the intersection. Therefore, in the interests of clarity, I've chosen to use the old "semicircular bump" style of schematic in this book (see Figure 2-39). It can be summarized like this:

- A dot joining two wires indicates an electrical connection.
- A bump in a wire that crosses another wire indicates no connection.

In this book, you won't find wires crossing each other without either a dot or a bump.

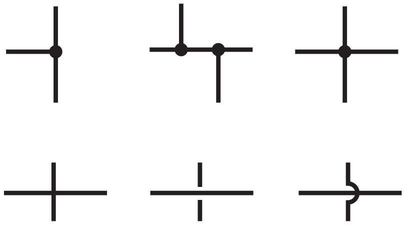

Figure 2-39. *In wiring schematics, a dot always indicates an electrical connection. However, the cross-shaped intersection of wires at top-right is considered bad style because if the dot is accidentally omitted or poorly printed, the intersection can be mistaken for the type shown at bottom left, in which the wires do not make a connection. All three of the configurations in the bottom row indicate no connection, the first example being the most common style, the center example being least common, and the third being the most old-fashioned—although for reasons of clarity, it is used in this book.*

In a battery-powered circuit, you may find a battery symbol, but more often you will find a little note indicating where positive voltage enters the system, while negative is indicated by a "ground" symbol. In fact there may be ground symbols all over the place. You have to remember that when you build a circuit, all the wires leading to grounds must actually be connected together, to the negative side of the voltage supply.

The idea of the ground symbol dates back to the time when electronic gadgets were mounted on a metal chassis, which was connected to the negative side of the power supply. The ground symbol really meant "connect to the chassis." Some variants in the ground symbol are shown in Figure 2-40.

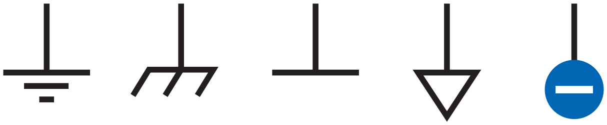

Figure 2-40. *All of these symbols are used to mean the same thing: connect the wire to "ground" or "chassis" or the negative side of the power source. The far-right symbol is used in this book.*

In this book, we have color throughout, so I'll show a red positive and blue negative to clarify where the power is connected, and I won't use ground symbols. Once again, my purpose is to minimize the risk of misunderstandings, because I know how frustrating it is to build a circuit that doesn't work.

A big inconsistency in schematics is the way in which they show resistors. The traditional zigzag symbol has been abandoned in Europe. Instead they use a rectangle with a number inside indicating the number of ohms. See Figure 2-41. The Europeans also changed the way in which decimal points are represented: they omit them as much as possible, because in badly printed schematics, the little dots tend to get lost (or can be confused with dust and dirt). So, a 4.7KΩ resistor will be listed as 4K7, and a 1.2MΩ resistor will be 1M2. I like this notation, so I'm going to be using it myself, but I'll be keeping the zigzag resistor symbol, which remains widely used in the United States.

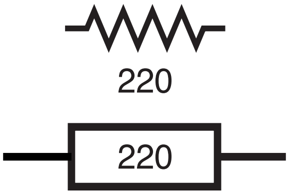

Figure 2-41. *Two styles for depicting a 220Ω resistor. The upper version is traditional, and still used in the United States. The lower version is European.*

FUNDAMENTALS

Basic schematic symbols (continued)

Potentiometers suffer from the same inconsistent style between the United States and Europe, but either way, you'll find an arrow showing where the wiper (usually, the center terminal) touches the resistance. See Figure 2-42. And sometimes LEDs are shown inside circles, and sometimes not. I prefer circles, myself. See Figure 2-46.

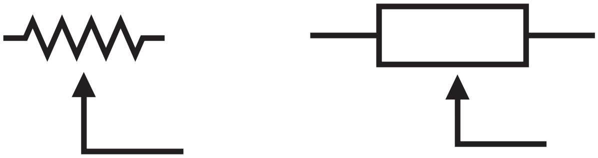

Figure 2-42. *Potentiometer symbols: the left is traditional and used in the United States, the right is European. In both cases the arrow indicates the wiper (usually the center terminal).*

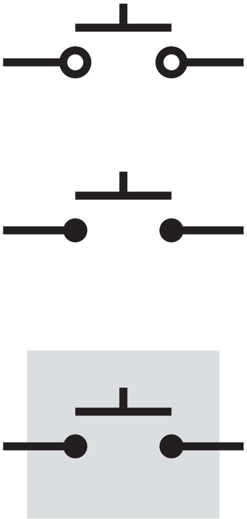

Figure 2-43. *Three ways of indicating a pushbutton switch.*

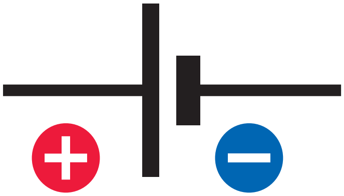

Figure 2-44. *The battery symbol is usually shown without + and – symbols. I've added them for clarity.*

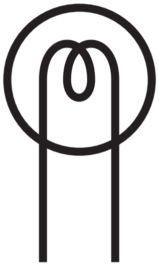

Figure 2-45. *Symbol for an incandescent lightbulb.*

I'll explore other symbol variants later in the book. Meanwhile, the most important things to remember are:

- The positions of components in a schematic are not important.
- The styles of symbols used in a schematic are not important.
- The connections between the components are extremely important.

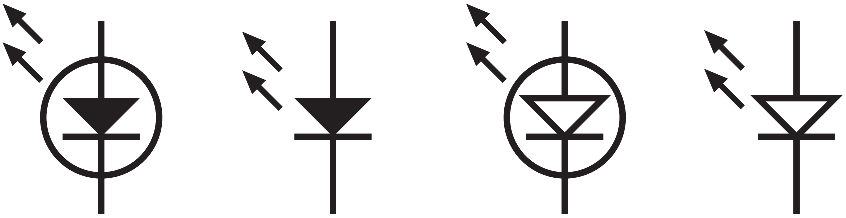

Figure 2-46. *Sometimes an LED is shown with a circle around it; sometimes not. In this book, I will include the circle. The arrows indicate emitted light.*

FUNDAMENTALS

Basic schematic symbols (continued)

For example, the three LED circuits that I have included in Figure 2-47 show components in different positions, using different symbols, but all three circuits function exactly the same way, because their connections are the same. In fact, they all depict the circuit that you built in Experiment 4, shown in Figure 1-50.

Often the symbols in a schematic are placed so that the circuit is most intuitively easy to understand, regardless of how you may build it with actual components. Compare the example in Figure 2-48, showing the two DPDT switches, with the version shown back in Figure 2-35. The previous one looked more like your bench-top version of it, but Figure 2-48 shows the flow of electricity more clearly.

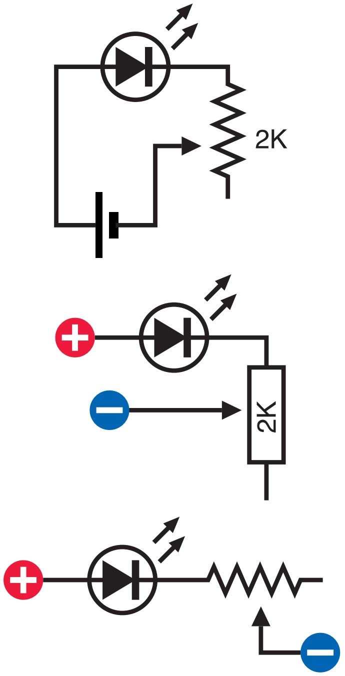

Figure 2-47. *These three schematics all depict the same basic circuit. It's the circuit that you built with the potentiometer in Experiment 4.*

In many schematics, the positive side of the power supply is shown at the top of the diagram, and negative or ground at the bottom. Many people also tend to draw a schematic with an input (such as an audio input, in an amplifier circuit) at the left side, and the output at the right. So, "positive voltage" flows from top to bottom while a signal tends to pass from left to right.

When I was planning this book, initially I drew the schematics to conform with this top-to-bottom, left-to-right convention, but as I started building and testing the circuits, I changed my mind. We use a device known as a "breadboard" to create circuits, and its internal connections require us to lay out components very differently from a typical schematic. When you're starting to learn electronics, it's very confusing to try to rearrange components from a schematic in the configuration that you need for a breadboard.

Therefore, throughout this book, you'll find that I have drawn the schematics to imitate the way you'll wire them on a breadboard. I believe the advantages of doing things this way outweigh the disadvantages of being a little different from the schematic styles that are used elsewhere.

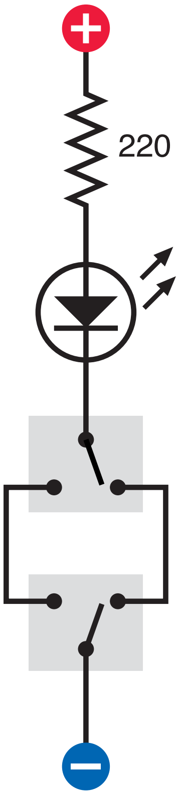

Figure 2-48. *This schematic is just another, clearer, simpler way of showing the circuit that appeared in Figure 2-35.*

Experiment 7: Relay-Driven LEDs

You will need:

- AC adapter, wire cutters and strippers.
- DPDT relay. Quantity: 2.
- LEDs. Quantity: 2.
- Resistor, 680Ω approx. Quantity: 1.
- Pushbutton, SPST. Quantity: 1.
- Hookup wire, 22 gauge, or patch cords.
- Alligator clips. Quantity: 8.
- Utility knife.

The next step in our exploration of switching is to use a remote-controlled switch. By "remote-controlled," I mean one to which you can send a signal to turn it on or off. This kind of switch is known as a *relay*, because it relays an instruction from one part of a circuit to another. Often a relay is controlled by a low voltage or small current, and switches a larger voltage or higher current.

This setup can be cost-effective. When you start your car, for instance, a relatively small, cheap switch sends a small signal down a relatively long, thin, inexpensive piece of wire, to a relay that is near the starter motor. The relay activates the motor through a shorter, thicker, more expensive piece of wire, capable of carrying as much as 100 amps.

Similarly, when you raise the lid on an a top-loading washing machine during its spin cycle, you close a lightweight switch that sends a small signal down a thin wire to a relay. The relay handles the bigger task of switching off the large motor spinning the drum full of wet clothes.

Before you begin this experiment, you need to upgrade your power supply. We're not going to use batteries anymore, because most relays require more than 6 volts, and in any case you should have a power supply that can give you a variety of voltages on demand. The simplest way to achieve this is by using an AC adapter.

First you'll set up the AC adapter. After you have it running, you'll use it to power the relay. Initially the relay will just switch between two LEDs, but then you'll modify the circuit to make the LEDs flash automatically. Finally you'll rebuild the circuit on a breadboard, and say goodbye to alligator clips, for most of the time at least.

Preparing Your AC Adapter

An AC adapter plugs into the wall and converts the high-voltage AC supply in your home into a safe, low DC voltage for electronic devices. Any charger that you use with your cell phone, or iPod, or laptop computer is a special-purpose AC adapter that delivers only one voltage via a specific type of plug. I've asked you to buy a general-purpose adapter that delivers many different voltages, and we're going to begin by getting rid of its plug.

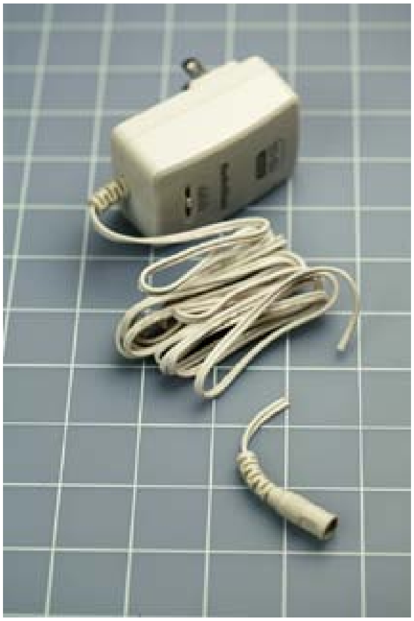

Figure 2-49. *Preparing an AC adapter. First, cut off the little low-voltage plug and throw it away.*

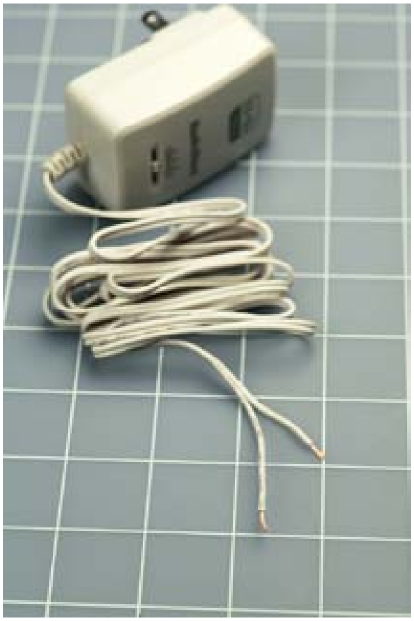

Figure 2-50. *Second, strip the wires, making one shorter than the other to reduce the risk of them touching. Color one of the adapter wires red with a marker, to identify it as the positive one.*

1. It's important to make sure that your AC adapter is not plugged into the wall!
2. Chop off the little plug at the end of its wire. See Figure 2-49.
3. Use a box cutter or utility knife or scissors to make a half-inch cut between the two conductors, and then pull the conductors apart a couple of inches.
4. Use wire cutters to trim one of the conductors shorter than the other, so that after you strip away a little of the insulation, the exposed copper wires cannot easily touch each other. This is a precaution against short-circuiting your AC adapter and burning it out.
5. Strip the two conductors using your wire strippers. Twist the copper strands between finger and thumb so that there are no loose strands sticking out. See Figure 2-50.
6. Make sure that the two wires are not touching each other, and plug your AC adapter into a wall outlet. Set your meter to DC volts and apply the meter probes to the wires from the adapter. If the voltage is preceded with a minus sign, you have the probes the wrong way around. Reverse them and the minus sign should go away. This tells you which wire is positive.
7. Mark the positive wire from the adapter. If the wire has white insulation, you can mark it with a red marker. If it has black insulation, you can tag it with a label. The positive wire will remain positive regardless of which way up you plug the AC adapter into a wall outlet.

The Relay

The type of relay that I want you to use has little spiky legs on the bottom, in a standardized layout. If you buy some other kind of relay, you will have to figure out for yourself which pins are connected to the coil inside, which pins go to the poles of the switch inside it, and which go to the normally closed and normally open contacts. You can check the manufacturer's data sheet for this purpose, but I strongly suggest you use one of the relays mentioned in the shopping list, so that you can follow the instructions here more easily.

I asked you to buy two relays so that you can use one for investigational purposes—meaning that you can break it open and take a look inside. If you do this very, very carefully, the relay should still be usable afterward. If not, well, you still have a spare.

The easiest way to open the relay is with a box cutter or utility knife. Figures 2-52, 2-53, and 2-54 show the technique. Shave the edges of the plastic shell containing the relay, beveling them until you see just a hair-thin opening. Don't go any farther; the parts inside are very, very close to its housing. Now pop the top off. You can use needle-nosed pliers to nibble the rest of the shell away. Read the following section, "Fundamentals: Inside a relay," and then apply power to the relay to see how it works.

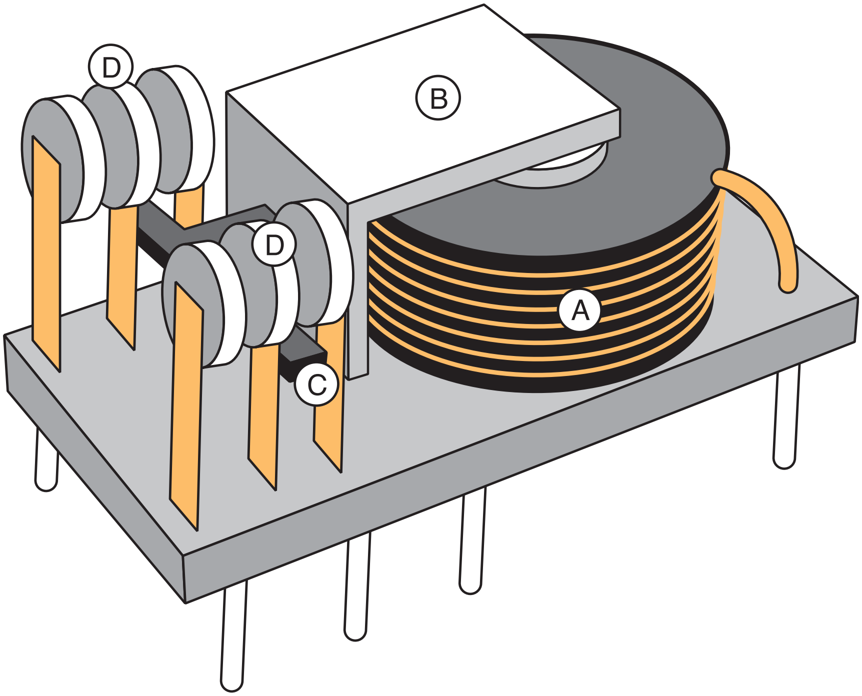

Figure 2-51. *This is one way that the parts inside a relay can be arranged. The coil, A, generates a magnetic attraction pulling lever B downward. A plastic extension, C, pushes outward against flexible metal strips and moves the poles of the relay, D, between the contacts.*

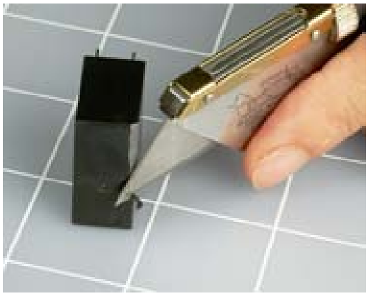

Figure 2-52. *To look inside a sealed relay, shave the top edges of the plastic package with a utility knife til you open a thin crack.*

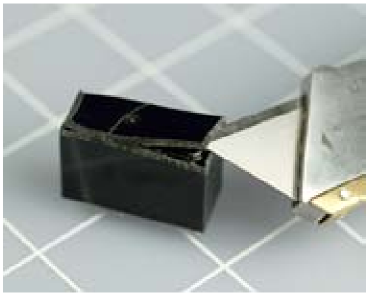

Figure 2-53. *Insert the blade of your knife to pry open the top, then repeat the procedure for the sides.*

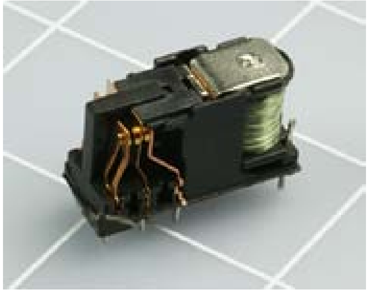

Figure 2-54. *If you are really, really careful, the relay should still work after you open it.*

Figure 2-55

Figure 2-56. *Patience is essential when carving the edges of a relay package in order to open it. Faster methods such as a tomahawk or a flamethrower will satisfy the emotional needs of those with a short attention span, but results may be unpredictable.*

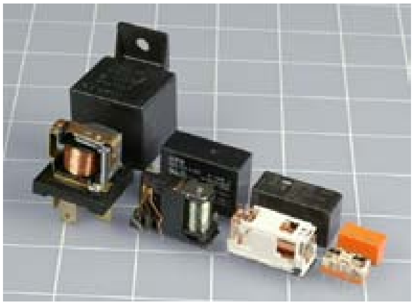

Figure 2-57. *Four assorted 12-volt relays, shown with and without their packages. The automotive relay (far left) is the simplest and easiest to understand, because it is designed without much concern for the size of the package. Smaller relays are more ingeniously designed, more complex, and more difficult to figure out. Usually, but not always, a smaller relay is designed to switch less current than a larger one.*

FUNDAMENTALS

Inside a relay

A relay contains a coil of wire wrapped around an iron core. When electricity runs through the coil, the iron core exerts a magnetic force, which pulls a lever, which pushes or pulls a springy strip of metal, closing two contacts. So as long as electricity runs through the coil, the relay is "energized" and its contacts remain closed.

When the power stops passing through the coil, the relay lets go and the springy strip of metal snaps back into its original position, opening the contacts. (The exception to this rule is a latching relay, which requires a second pulse through a separate coil to flip it back to its original position; but we won't be using latching relays until later in the book.)

Relays are categorized like switches. Thus, you have SPST relays, DPST, SPDT, and so on.

Compare the schematics in Figure 2-58 with the schematics of switches in Figure 2-38. The main difference is that the relay has a coil that activates the switch. The switch is shown in its "relaxed" mode, when no power flows through the coil.

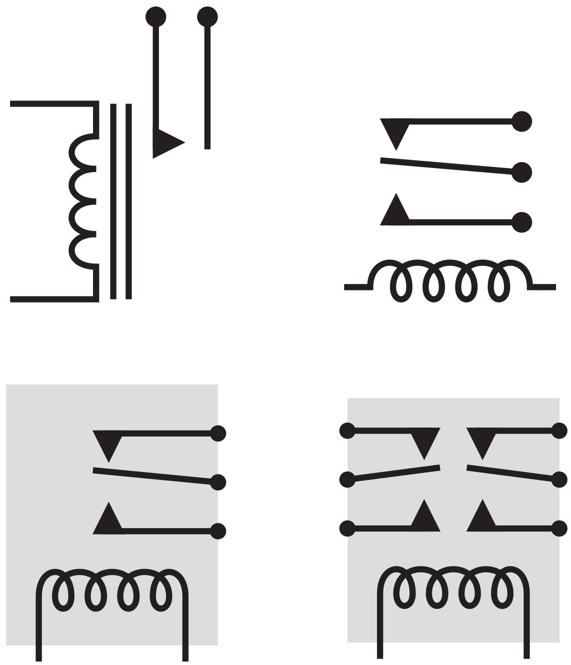

Figure 2-58. *Various ways to show a relay in a schematic. Top left: SPST. Top right and bottom left: SPDT. Bottom right: DPDT. The styles at bottom-left and bottom-right will be used in this book.*

The contacts are shown as little triangles. When there are two poles instead of one, the coil activates both switches simultaneously.

Most relays are nonpolarized, meaning that you can run electricity through the coil in either direction, and the relay doesn't care. You should check the data sheet to make sure, though. Some relay coils work on AC voltage, but almost all low-voltage relays use direct current—a steady flow of electricity, such as you would get from a battery. We'll be using DC relays in this book.

Relays suffer from the same limitations as switches: their contacts will be eroded by sparking if you try to switch too much voltage. It's not worth saving a few dollars by using a relay that is rated for less current or voltage than your application requires. The relay will fail you when you need it most, and may be inconvenient to replace.

Because there are so many different types of relays, read the specifications carefully before you buy one. Look for these basics:

Coil voltage

The voltage that the relay is supposed to receive when you energize it.

Set voltage

The minimum voltage that the relay needs to close its switch. This will be a bit less than the ideal coil voltage.

Operating current

The power consumption of the coil, usually in milliamps, when the relay is energized. Sometimes the power is expressed in milliwatts.

Switching capacity

The maximum amount of current that you can switch with contacts inside the relay. Usually this is for a "resistive load," meaning a passive device such as light bulb. When you use a relay to switch on a motor, the motor takes a big initial surge of current before it gets up to speed. In this case, you should choose a relay rated for double the current that the motor draws when it is running.

Procedure

Turn the relay with its legs in the air and attach wires and LEDs as shown in Figure 2-59, with a 680Ω resistor (a 1K resistor will be OK if you don't have the correct value). Also attach a pushbutton switch. (Your pushbutton switch may look different from the one shown, but as long as it is a SPST pushbutton with two contacts at the bottom, it will work the same way.) When you press the pushbutton, the relay will make the first LED go out and the second LED light up. When you release the pushbutton, the first LED lights up and the second one goes out.

How It Works

Check the schematic in Figure 2-60 and compare it with Figure 2-59. Also see Figure 2-62, which shows how the pins outside the relay make connections inside the relay when its coil is energized, and when it is not energized.

This is a DPDT relay, but we are only using one pole and ignoring the other. Why not buy a SPDT relay? Because I want the pins to be spaced the way they are when you will upgrade this circuit by transferring it onto a breadboard, which will happen very shortly.

On the schematic, I have shown the switch inside the relay in its relaxed state. When the coil is energized, the switch flips upward, which seems counterintuitive, but just happens to be the way that this particular relay is made.

When you're sure you understand how the circuit works, it's time to move on to the next step: making a small modification to get the relay to switch itself on and off, as we'll do in Experiment 8.

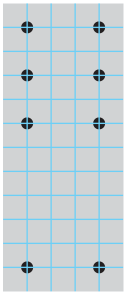

Figure 2-61. *The layout of the pins of the relay, superimposed on a grid of 1/10-inch squares. This is the type of relay that you will need in Experiment 8.*

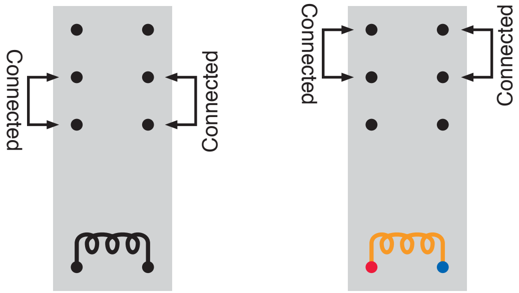

Figure 2-62. *How the relay connects the pins, when it is not energized (left) and when it is energized (right).*

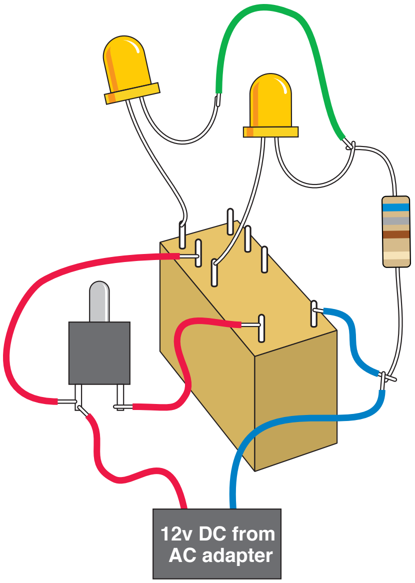

Figure 2-59. *As before, you can use patch cords, if you have them, instead of some of the wired connections shown here.*

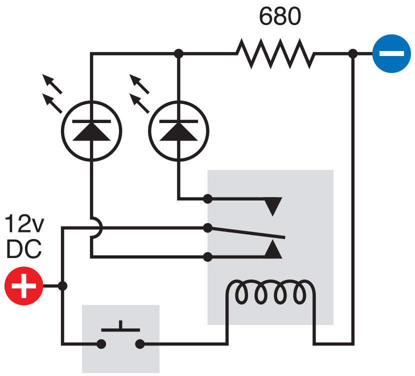

Figure 2-60. *Same circuit, shown in schematic form.*

Larger versions of all schematics and breadboard photos are available online at this book's website: http://oreilly.com/catalog/9780596153748.

Experiment 8: A Relay Oscillator

You will need:

- AC adapter, breadboard, wire, wire cutters and strippers.
- DPDT relay. Quantity: 1.
- LEDs. Quantity: 2.
- Pushbutton, SPST. Quantity: 1.
- Alligator clips. Quantity: 8.
- Resistor, approximately 680Ω. Quantity: 1.
- Capacitor, electrolytic, 1,000 µF. Quantity: 1.

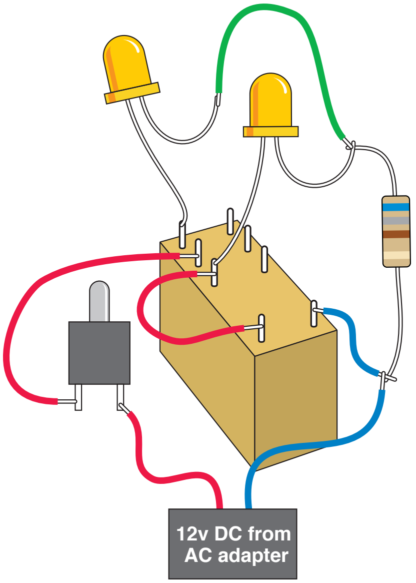

Figure 2-63. *A small revision to the previous circuit causes the relay to start oscillating when power is applied.*

Look at the revised drawing in Figure 2-63 and the revised schematic in Figure 2-64 and compare them with the previous ones. Originally, there was a direct connection from the pushbutton to the coil. In the new version, the power gets to the coil by going, first, through the contacts of the relay.

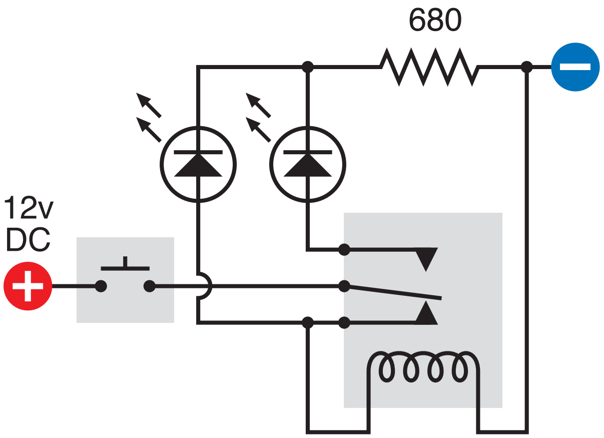

Figure 2-64. *The oscillator circuit shown in schematic form.*

Now, when you press the button, the contacts in their relaxed state feed power to the coil as well as to the lefthand LED. But as soon as the coil is energized, it opens the contacts. This interrupts the power to the coil—so the relay relaxes, and the contacts close again. They feed another pulse of power to the coil, which opens the contacts again, and the cycle repeats endlessly.

Because we're using a very small relay, it switches on and off extremely fast. In fact, it oscillates perhaps 50 times per second (too fast for the LEDs to show what's really happening). Make sure your circuit looks like the one in the diagram, and then press the pushbutton very briefly. You should hear the relay make a buzzing sound. If you have impaired hearing, touch the relay lightly with your finger, and you should feel the relay vibrating.

When you force a relay to oscillate like this, it's liable to burn itself out or destroy its contacts. That's why I asked you to press the pushbutton briefly. To make the circuit more practical, we need something to slow the relay down and prevent it from self-destructing. That necessary item is a capacitor.

Adding Capacitance

Add a 1,000 µF electrolytic capacitor in parallel with the coil of the relay as shown in the diagram in Figure 2-65 and the schematic in Figure 2-66. Check Figure 2-14 if you're not sure what a capacitor looks like. The 1,000 µF value will be printed on the side of it, and I'll explain what this means a little later.

Make sure the capacitor's *short* wire is connected to the *negative* side of the circuit; otherwise, it won't work. In addition to the short wire, you should find a minus sign on the body of the capacitor, which is there to remind you which side is negative. Electrolytic capacitors are fussy about this.

When you press the button now, the relay should click slowly instead of buzzing. What's happening here?

A capacitor is like a tiny rechargeable battery. It's so small that it charges in a fraction of a second, before the relay has time to open its lower pair of contacts. Then, when the contacts are open, the capacitor acts like a battery, providing power to the relay. It keeps the coil of the relay energized for about one second. After the capacitor exhausts its power reserve, the relay relaxes and the process repeats.

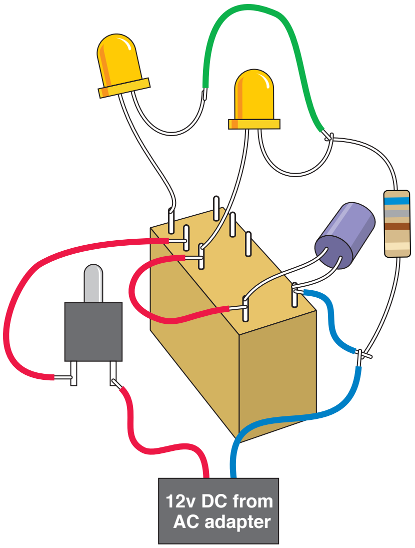
12v DC from AC adapter

Figure 2-65. *Adding a capacitor makes the relay oscillate more slowly.*

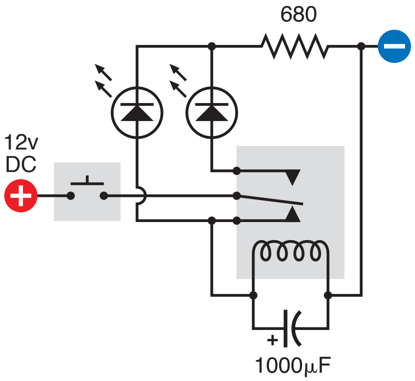

Figure 2-66. *The capacitor appears at the bottom of this schematic diagram.*

FUNDAMENTALS

Farad basics

The Farad is an international unit to measure capacitance. Modern circuits usually require small capacitors. Consequently it is common to find capacitors measured in microfarads (one-millionth of a farad) and even picofarads (one-trillionth of a farad). Nanofarads are also used, more often in Europe than in the United States. See the following conversion table.

0.001 nanofarad	1 picofarad	1 pF
0.01 nanofarad	10 picofarads	10 pF
0.1 nanofarad	100 picofarads	100 pF
1 nanofarad	1,000 picofarads	1,000 pF
0.001 microfarad	1 nanofarad	1 nF
0.01 microfarad	10 nanofarads	10 nF
0.1 microfarad	100 nanofarads	100 nF
1 microfarad	1,000 nanofarads	1,000 nF
0.000001 Farad	1 microfarad	1 µF
0.00001 Farad	10 microfarads	10 µF
0.0001 Farad	100 microfarads	100 µF
0.001 Farad	1,000 microfarads	1,000 µF

(You may encounter capacitances greater than 1,000 microfarads, but they are uncommon.)

FUNDAMENTALS

Capacitor basics

DC current does not flow through a capacitor, but voltage can accumulate very quickly inside it, and remains after the power supply is disconnected. Figures 2-67 and 2-68 may help to give you an idea of what happens inside a capacitor when it is fully charged.

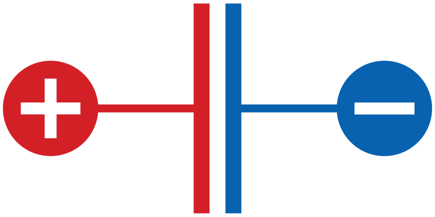

Figure 2-67. *When DC voltage reaches a capacitor, no current flows, but the capacitor charges itself like a little battery. The positive and negative charges are equal and opposite.*

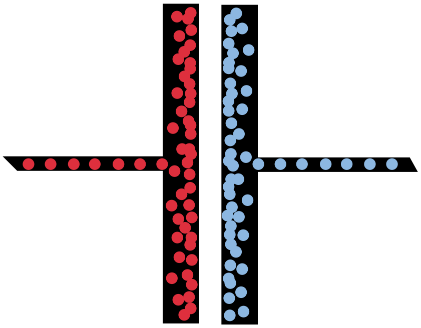

Figure 2-68. *You can imagine positive "charge particles" accumulating on one side of the capacitor and attracting negative "charge particles" to the opposite side.*

In most modern electrolytic capacitors, the plates have been reduced to two strips of very thin, flexible, metallic film, often wrapped around each other, separated by an equally thin insulator. Disc ceramic capacitors typically consist of just a single disc of nonconductive material with metal painted on both sides and leads soldered on.

The two most common varieties of capacitors are ceramic (capable of storing a relatively small charge) and electrolytic (which can be much larger). Ceramics are often disc-shaped and yellow in color; electrolytics are often shaped like miniature tin cans and may be just about any color. Refer back to Figures 2-13 and 2-14 for some examples.

Getting Zapped by Capacitors

If a large capacitor is charged with a high voltage, it can retain that voltage for a long time. Because the circuits in this book use low voltages, you don't have to be concerned about that danger here, but if you are reckless enough to open an old TV set and start digging around inside (which I do not recommend), you may have a nasty surprise. An undischarged capacitor can kill you as easily as if you stick your finger into an electrical outlet. Never touch a large capacitor unless you really know what you're doing.

FUNDAMENTALS

Capacitor basics (continued)

Ceramic capacitors have no polarity, meaning that you can apply negative voltage to either side of them. Electrolytics do have polarity, and won't work unless you connect them the right way around.

The schematic symbol for a capacitor has two significant variants: with two straight lines (symbolizing the plates inside a capacitor), or with one straight line and one curved line, as shown in Figure 2-69. When you see a curved line, that side of the capacitor should be more negative than the other. The schematic symbol may also include a + sign. Unfortunately, some people don't bother to draw a curved plate on a polarized capacitor, yet others draw a curved plate even on a nonpolarized capacitor.

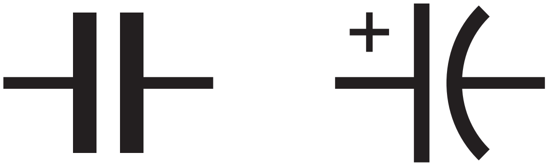

Figure 2-69. *The generic schematic for a capacitor is on the left. The version on the right indicates a polarized capacitor which requires its left plate to be "more positive" than its right plate. The plus sign is often omitted.*

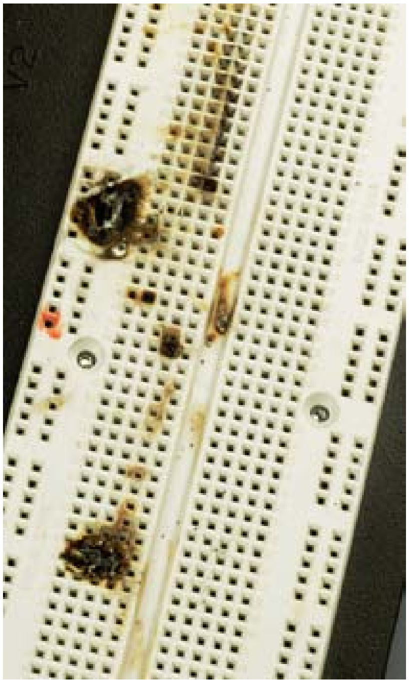

Figure 2-70. *A tantalum capacitor was plugged into this breadboard, accidentally connected the wrong way around to a power source capable of delivering a lot of current. After a minute or so of this abuse, the capacitor rebelled by popping open and scattering small flaming pieces, which burned their way into the plastic of the breadboard. Lesson learned: observe polarity!*

Capacitor Polarity

You must connect an electrolytic capacitor so that its longer wire is more positive than its shorter wire. The shell of the capacitor is usually marked with a negative sign near the shorter wire.

Some capacitors may behave badly if you don't observe their polarity. One time I connected a tantalum capacitor to a circuit, using a power supply able to deliver a lot of current, and was staring at the circuit and wondering why it wasn't working when the capacitor burst open and scattered little flaming fragments of itself in a 3-inch radius. I had forgotten that tantalum capacitors can be fussy about positive and negative connections. Figure 2-70 shows the aftermath.

BACKGROUND

Michael Faraday and capacitors

The earliest capacitors consisted of two metal plates with a very small gap between them. The principle of the thing was simple:

- If one plate was connected to a positive source, the positive charges attracted negative charges onto the other plate.
- If one plate was connected to a negative source, the negative charges attracted positive charges onto the other plate.

Figures 2-67 and 2-68, shown previously, convey the basic idea.

The electrical storage capacity of a capacitor is known as its *capacitance*, and is measured in farads, named after Michael Faraday (Figure 2-71), another of the pantheon of electrical pioneers. He was an English chemist and physicist who lived from 1791 to 1867.

Although Faraday was relatively uneducated and had little knowledge of mathematics, he had an opportunity to read a wide variety of books while working for seven years as a bookbinder's apprentice, and thus was able to educate himself. Also, he lived at a time when relatively simple experiments could reveal fundamental properties of electricity. Thus he made major discoveries including electromagnetic induction, which led to the development of electric motors. He also discovered that magnetism could affect rays of light.

His work earned him numerous honors, and his picture was printed on English 20-pound bank notes from 1991 through 2001.

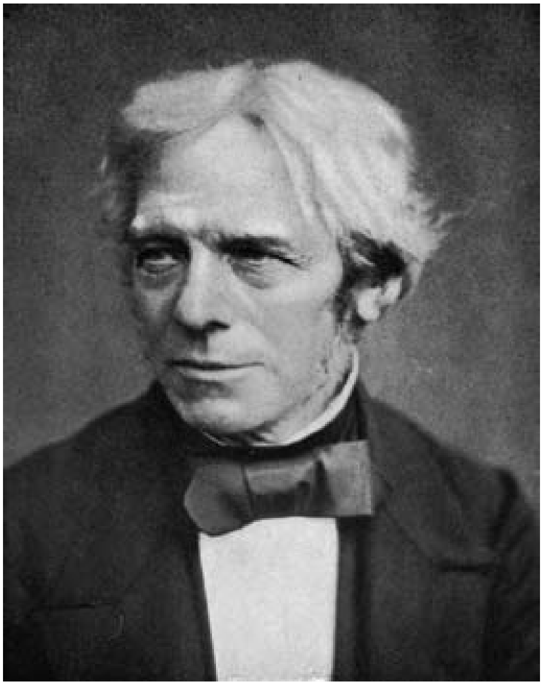

Figure 2-71. *Michael Faraday*

Breadboarding the Circuit

I promised to free you in time from the frustrations of alligator clips, and that time has come. Please turn your attention to the block of plastic with lots of little holes in it that I asked you to buy. For reasons that I do not know, this is called a *breadboard*. When you plug components into the holes, hidden metal strips inside the breadboard connect the components for you, allowing you to set up a circuit, test it, and modify it very easily. Afterward you can pull the components off the breadboard and put them away for future experiments.

Without a doubt, breadboarding is the most convenient way to test something before you decide whether you want to keep it.

Almost all breadboards are designed to be compatible with integrated circuit chips (which we will be using in Chapter 4 of this book). The chip straddles an empty channel in the center of the breadboard with rows of little holes either side—usually five holes per row. You insert other components into these holes.

In addition, the breadboard should have columns of holes running down each side. These are used to distribute positive and negative power.

Take a look at Figures 2-72 and 2-73, which show the upper part of a typical breadboard seen from above, and the same breadboard seen as if with X-ray vision, showing the metal strips that are embedded behind the holes.

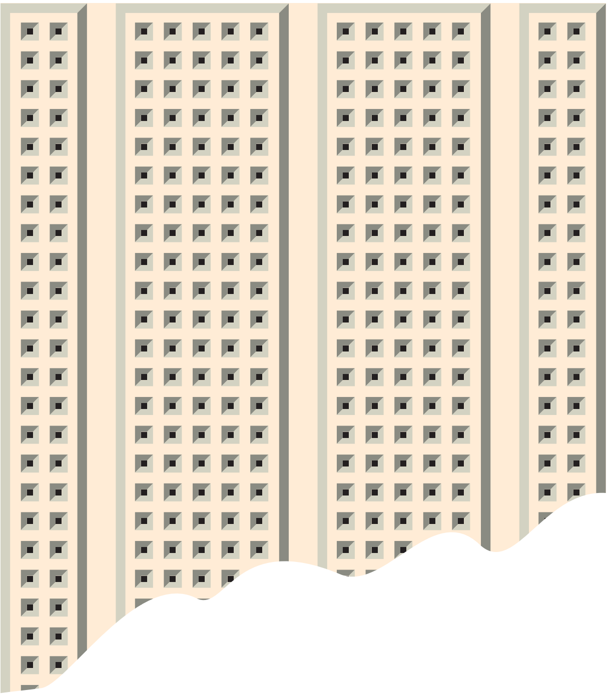

Figure 2-72. *A typical breadboard. You can plug components into the holes to test a circuit very quickly.*

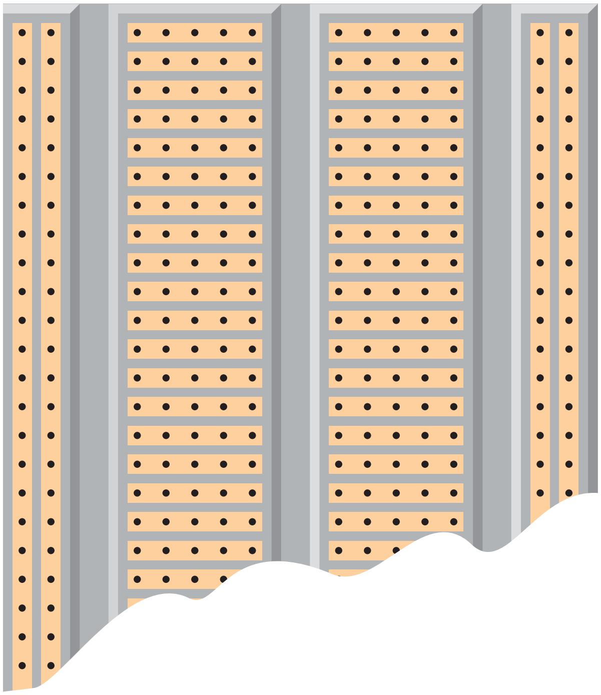

Figure 2-73. *This X-ray-vision view of the breadboard reveals the copper strips that are embedded in it. The strips conduct electricity from one component to another.*

Important note: some breadboards divide each vertical column of holes, on the left and the right, into two separate upper and lower sections. Use your meter's continuity testing feature to find out if your breadboard conducts power along its full length, and add jumper wires to link the upper and lower half of the breadboard if necessary.

Figure 2-74 shows how you can use the breadboard to replicate your oscillating relay circuit. To make this work, you need to apply the positive and negative power from your AC adapter. Because the wire from your AC adapter is almost certainly stranded, you'll have difficulty pushing it into the little holes. A way around this is to set up a couple of pieces of bare 22-gauge wire, and use them as terminals to which you clip the wire from the adapter, as in Figure 2-75. (Yes, you still need just a couple of alligator clips for this purpose.) Alternatively, you can use a breadboard with power terminals built into it, which is more convenient.

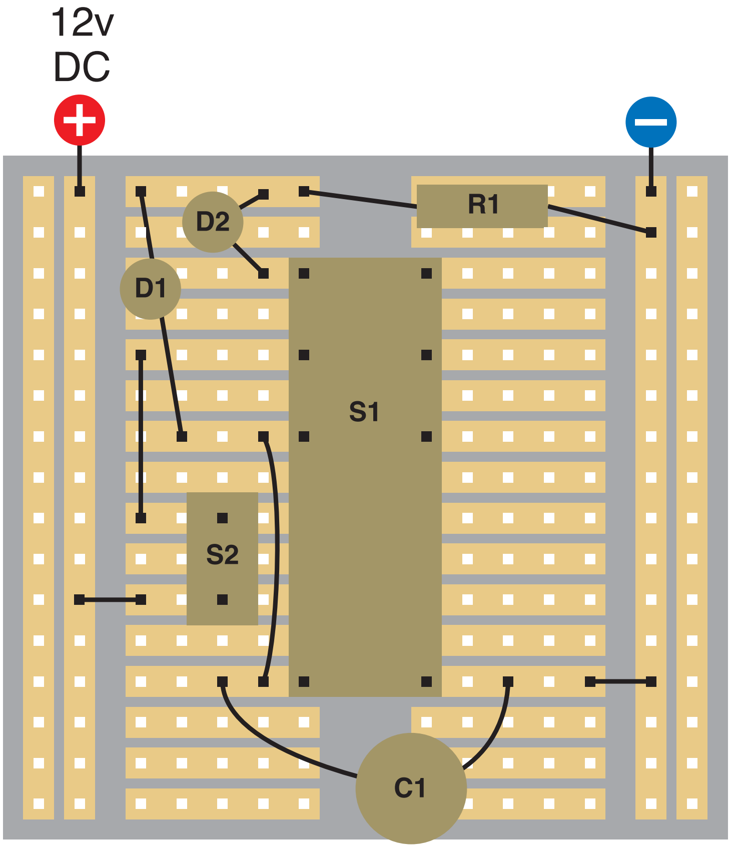

Figure 2-74. *If you place the components on your breadboard in the positions shown, they will create the same circuit that you built from wire and alligator clips in Experiment 8. Component values:*

D1, D2: Light-emitting diodes
S1: DPDT relay
S2: SPST momentary switch
C1: Electrolytic capacitor, 1,000 µF
R1: Resistor, 680Ω minimum

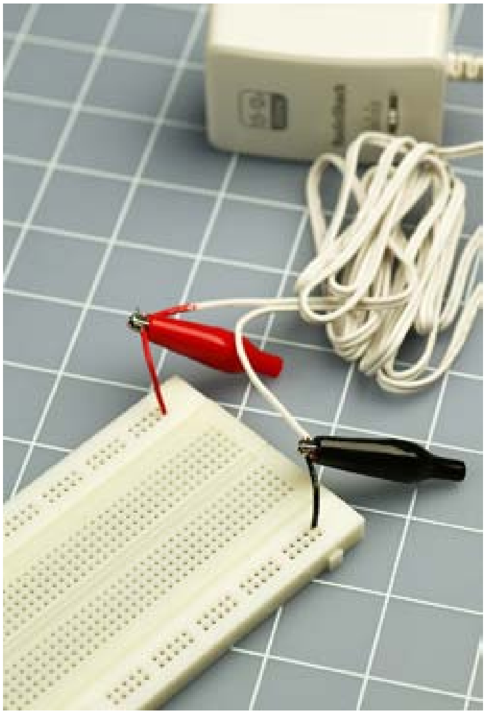

Figure 2-75. *If your breadboard doesn't have screw terminals, insert two short pieces of solid-core wire with stripped ends and then attach the stranded wires from the adapter using alligator clips.*

You'll need some more 22-gauge wire, or some precut hookup wire, to supply the power to your components, which are plugged into the breadboard as shown in Figures 2-76 and 2-77. If you get all the connections right, the circuit should function the same way as before.

The geometry of the metal connecting strips in the breadboard often forces you to connect components in a roundabout way. The pushbutton, for instance, supplies power to the pole of the relay but cannot be connected directly opposite, because there isn't room for it.

Remember that the strips inside the breadboard that don't have any wires or components plugged into them are irrelevant; they don't do anything.

I'll include some suggested breadboard layouts for circuits as you continue through this book, but eventually you'll have to start figuring out breadboard layouts for yourself, as this is an essential part of hobby electronics.

Larger versions of all schematics and breadboard photos are available online at this book's website: http://oreilly.com/catalog/9780596153748.

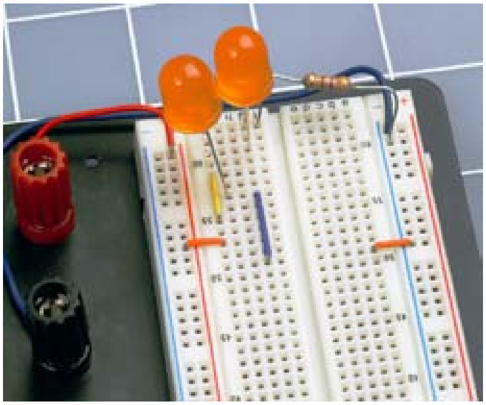

Figure 2-76. *Two oversized LEDs, one resistor, and the necessary jumper wires have been added to the breadboard.*

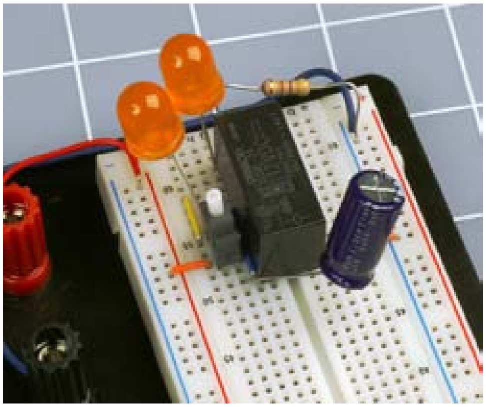

Figure 2-77. *Now the pushbutton, relay, and capacitor have been added to complete the circuit shown in the diagram and the schematic. When the pushbutton is pressed, the relay oscillates and the LEDs flash.*

Experiment 9: Time and Capacitors

You will need:

- AC adapter, breadboard, wire, wire cutters, and strippers.
- Multimeter.
- Pushbutton, SPST. Quantity: 1.
- Resistors and electrolytic capacitors, assorted.

In Experiment 8, when you put a capacitor in parallel with the coil of the relay, the capacitor charged almost instantly before discharging itself through the relay coil. If you add a resistor in series with a capacitor, the capacitor will take longer to charge. By making a capacitor take longer to charge, you can measure time, which is a very important concept.

Clean the components off your breadboard and use it to set up the very simple circuit shown in Figure 2-78, where C1 is a 1,000 µF capacitor, R1 is a 100K resistor, R2 is a 100Ω resistor, and S1 is the pushbutton that you used previously. Set your meter to measure volts DC, place the probes around the capacitor, and hold down the pushbutton. You should see the meter counting upward as the voltage accumulates on the capacitor. (This is easier with a meter that doesn't have autoranging, because you won't have to wait while the meter figures out which range to apply.) Resistor R1 slows the charging time for the capacitor.

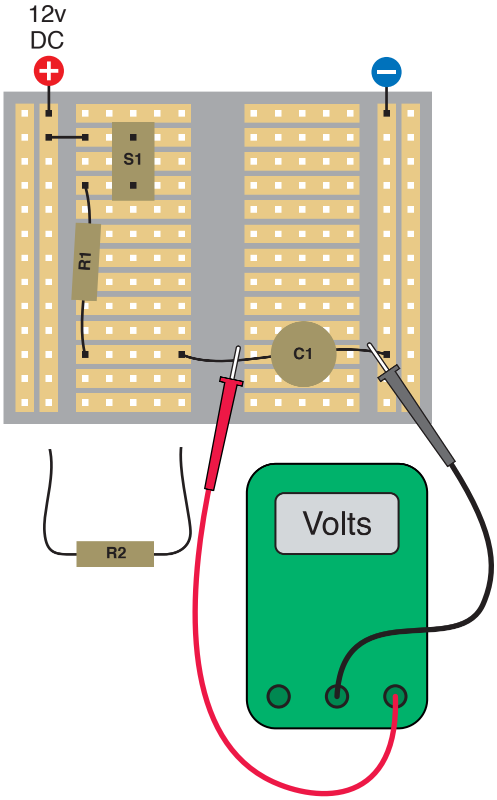

Figure 2-78. *Watch the voltage building up on the capacitor while you hold down the pushbutton. Substitute other values for R1, discharge the capacitor by touching R2 across it, and then repeat your measurement process.*

S1: Momentary pushbutton, OFF (ON)
R1: 100K initially
R2: 100Ω
C1: 1,000 µF

Release the pushbutton, set aside your meter, and discharge the capacitor by touching R2 across it for a second or two. Now substitute a 50K resistor for R1, and repeat the measurement. The meter should count upward almost twice as fast as before.

Voltage, Resistance, and Capacitance

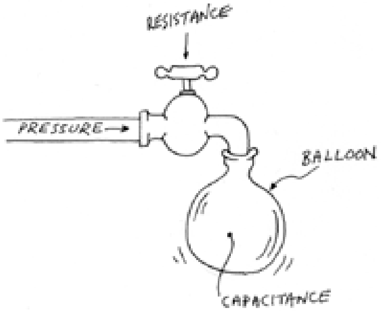

Figure 2-79. *When the faucet is closed half-way, the balloon will take longer to fill, but will still contain as much water and as much pressure in the end.*

Think of the resistor as a faucet, and the capacitor as a balloon that you are trying to fill with water. When you screw down the faucet until only a trickle comes through, the balloon will take longer to fill. But a slow flow of water will still fill the balloon completely if you wait long enough, and (assuming the balloon doesn't burst) the process ends when the pressure inside the balloon is equal to the water pressure in the pipe supplying the faucet. See Figure 2-79.

Similarly, in your circuit, if you wait long enough, eventually the voltage across the capacitor should reach the same value as the voltage of the power supply. In a 12-volt circuit, the capacitor should eventually acquire 12 volts (although "eventually" may take longer than you think).

This may seem confusing, because earlier you learned that when you apply voltage at one end of a resistor, you get less voltage coming out than you have going in. Why should a resistor deliver the full voltage when it is paired with a capacitor?

Forget the capacitor for a moment, and remember how you tested just two 1K resistors. In that situation, each resistor contained half the total resistance of the circuit, so each resistor dropped half the voltage. If you held the negative probe of your meter against the negative side of your power supply and touched the positive probe to the center point between the two resistors, you would measure 6 volts. Figure 2-80 illustrates this.

Now, suppose you remove one of the 1K resistors and substitute a 9K resistor. The total resistance in the circuit is now 10K, and therefore the 9K resistor drops 90% of the 12 volts. That's 10.8 volts. You should try this and check it with your meter. (You are unlikely to find a 9K resistor, because this is not a standard value. Substitute the nearest value you can find.)

Now suppose you remove the 9K resistor and substitute a 99K resistor. Its voltage drop will be 99% of the available voltage, or 11.88 volts. You can see where this is heading: the larger the resistor, the larger its contribution to voltage drop.

However, I noted previously that a capacitor blocks DC voltage completely. It can accumulate an electrical charge, but no current passes through it. Therefore, a capacitor behaves like a resistor that has infinite resistance to DC current.

(Actually the insulation inside the capacitor allows a little bit of "leakage," but a perfect capacitor would have infinite resistance.)

The value of any resistor that you put in series with the capacitor is trivial by comparison. No matter how high the value of the resistor is, the capacitor still provides much more resistance in the circuit. This means that the capacitor steals almost the complete voltage drop in the circuit, and the voltage difference between one end of the resistor and the other will be zero (assuming that we ignore little imperfections in the components). Figure 2-80 may help to clarify this concept.

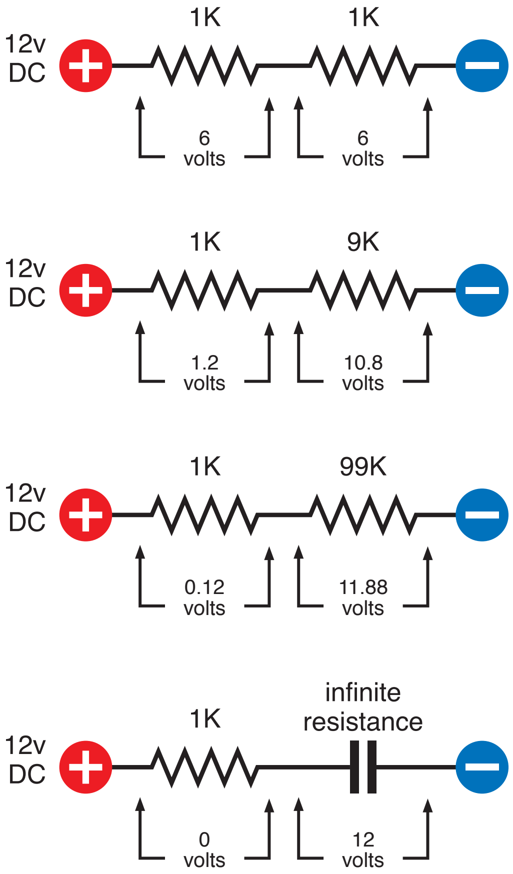

Figure 2-80. *When two resistances are in series, the larger one drops the voltage more than the smaller one. If the larger resistance becomes infinite (as in the case of a capacitor), the smaller one no longer has any measurable contribution to the voltage drop, and the voltage is almost exactly the same at both ends.*

You should try this using real resistors and capacitors—although if you do, you will run into a little problem. When you use your meter in its "DC volts" mode, it diverts a little of the current in the circuit—just a tiny taste—in the process of measuring it. The meter steals such a small amount, it doesn't affect the reading significantly when you are checking voltage across a resistor. The internal resistance of the meter is higher than the values of most resistors. However, remember that the internal resistance of a capacitor is almost infinite. Now the internal resistance of the meter becomes significant. Because you can never have an ideal meter, any more than you can have an ideal capacitor or resistor, your meter will always interfere with the circuit slightly, and you will get only an approximate indication.

If you try to measure the voltage on a capacitor that has been charged but is now not connected to anything else, you'll see the number slowly falling, as the capacitor discharges itself through the meter.

THEORY

The time constant

You may be wondering if there's a way to predict exactly how much time it takes for various capacitors to charge, when they are paired with various resistors. Is there a formula to calculate this?

Of course, the answer is yes, but the way we measure it is a bit tricky, because a capacitor doesn't charge at a constant rate. It accumulates the first volt very quickly, the second volt not quite as quickly, the third volt even less quickly—and so on. You can imagine the electrons accumulating on the plate of a capacitor like people walking into an auditorium and looking for a place to sit. The fewer seats that are left, the longer people take to find them.

The way we describe this is with something called a "time constant." The definition is very simple:

$$TC = R \times C$$

where TC is the time constant, and a capacitor of C farads is being charged through a resistor of R ohms.

Going back to the circuit you just tested, try using it again, this time with a 1K resistor and the 1,000 μF capacitor. We have to change those numbers to farads and ohms before we can put them in the formula. Well, 1,000 μF is 0.001 farads, and 1K is 1,000 ohms, so the formula looks like this:

$$TC = 1{,}000 \times 0.001$$

In other words, TC = 1—a lesson that could not be much easier to remember:

> ***A 1K resistor in series with a 1,000 μF capacitor has a time constant of 1.***

Does this mean that the capacitor will be fully charged in 1 second? No, it's not that simple. TC, the time constant, is the time it takes for a capacitor to acquire *63%* of the voltage being supplied to it, if it starts with zero volts.

(Why 63%? The answer to that question is too complicated for this book, and you'll have to read about time constants elsewhere if you want to know more. Be prepared for differential equations.) Here's a formal definition for future reference:

> TC, the time constant, is the time it takes for a capacitor to acquire 63% of the difference between its current charge and the voltage being applied to it. When TC=1, the capacitor acquires 63% of its full charge in 1 second. When TC=2, the capacitor acquires 63% of its full charge in 2 seconds. And so on.

What happens if you continue to apply the voltage? History repeats itself. The capacitor accumulates *another* 63% of the *remaining* difference between its current charge, and the voltage being applied to it.

Imagine someone eating a cake. In his first bite he's ravenously hungry, and eats 63% of the cake in one second. In his second bite, not wanting to seem too greedy, he takes just another 63% of the cake that is left—and because he's not feeling so hungry anymore, he requires the same time to eat it as he took to eat the first bite. In his third bite, he takes 63% of what still remains, and still takes the same amount of time. And so on. He is behaving like a capacitor eating electricity (Figure 2-81).

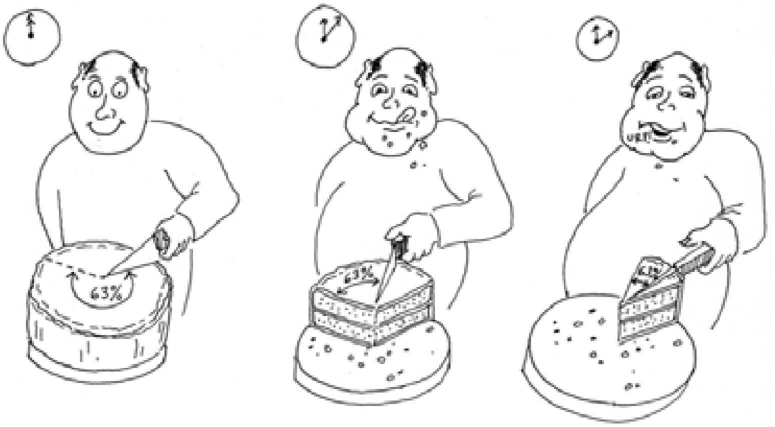

Figure 2-81. *If our gourmet always eats just 63% of the cake still on the plate, he "charges up" his stomach in the same way that a capacitor charges itself. No matter how long he keeps at it, his stomach is never completely filled.*

THEORY

The time constant (continued)

The cake eater will always have a few crumbs to eat, because he never takes 100% of the remainder. Likewise, the capacitor will never acquire a full charge. In a perfect world of perfect components, this process would continue for an infinite time.

In the real world, we say rather arbitrarily:

> After 5 × TC the capacitor will be so nearly fully charged, we won't care about the difference.

In the table is a calculation (rounded to two decimal places) showing the charge accumulating on a capacitor in a 12-volt circuit where the time constant is 1 second.

Here's how to understand the table. V1 is the current charge on the capacitor. Subtract this from the supply voltage (12 volts) to find the difference. Call the result V2. Now take 63% of V2, and add this to the current charge (V1) and call the result V4. This is the new charge that the capacitor will have after 1 second, so we copy it down to the next line in the table, and it becomes the new value for V1.

Now we repeat the same process all over again. Figure 2-82 shows this in graphical form. Note that after 5 seconds, the capacitor has acquired 11.92 volts, which is 99% of the power supply voltage. This should be close enough to satisfy anyone's real-world requirements.

If you try to verify these numbers by measuring the voltage across the capacitor as it charges, remember that because your meter steals a little current, there will be a small discrepancy that will increase as time passes. For practical purposes, the system works well enough.

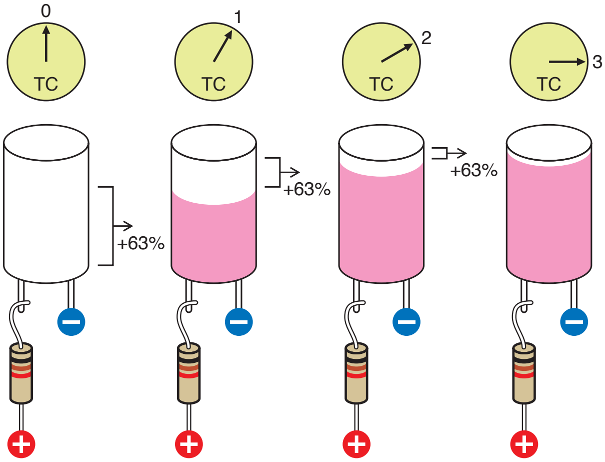

Figure 2-82. *A capacitor starts with 0 volts. After 1 time constant it adds 63% of the available voltage. After another time constant, it adds another 63% of the remaining voltage difference, and so on.*

Time in secs	V1: Charge on capacitor	V2: 12 – V1	V3: 63% of V2	V4: V1 + V3
0	0.00	12.00	7.56	7.56
1	7.56	4.44	2.80	10.36
2	10.36	1.64	1.03	11.39
3	11.39	0.61	0.38	11.77
4	11.77	0.23	0.15	11.92
5	11.92			

Experiment 10: Transistor Switching

You will need:

- AC adapter, breadboard, wire, and meter.
- LED. Quantity: 1.
- Resistors, various.
- Pushbutton, SPST. Quantity: 1.
- Transistor, 2N2222 or similar. Quantity: 1.

A transistor can switch a flow of electricity, just like a relay. But it's much more sensitive and versatile, as this first ultra-simple experiment will show.

We'll start with the 2N2222 transistor, which is the most widely used semiconductor of all time (it was introduced by Motorola in 1962 and has been in production ever since).

First, you should get acquainted with the transistor. Because Motorola's patents on the 2N2222 ran out long ago, any company can manufacture their own version of it. Some versions are packaged in a little piece of black plastic; others are enclosed in a little metal "can." (See Figure 2-83.) Either way, it contains a piece of silicon divided into three sections known as the collector, the base, and the emitter. I'll describe their function in more detail in a moment, but initially you just need to know that in this type of transistor, the collector receives current, the base controls it, and the emitter sends it out.

Use your breadboard to set up the circuit shown in Figure 2-85. Be careful to get the transistor the right way around! (See Figure 2-84.) For the three brands I have mentioned in the shopping list, the flat side should face right, if the transistor is packaged in black plastic, or the little tab should face toward the lower left, if the transistor is packaged in metal.

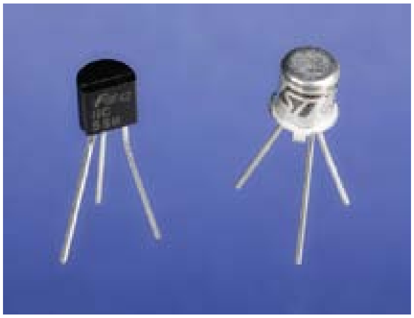

Figure 2-83. *A typical transistor is packaged either in a little metal can or a molded piece of black plastic. The manufacturer's data sheet tells you the identities of the three wire leads, relative to the flat side of a black plastic transistor or the tab that sticks out of a metal-can transistor.*

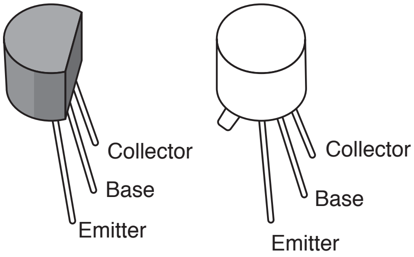

Figure 2-84. *The 2N2222 transistor may be packaged in either of these formats. Left: RadioShack or Fairchild. Right: STMicroelectronics (note the little tab sticking out at the lower-left side). If you use a different brand, you'll have to check the manufacturer's data sheet. Insert the transistor in your breadboard with the flat side facing right, as seen from above, or the tab pointing down and to the left, seen from above.*

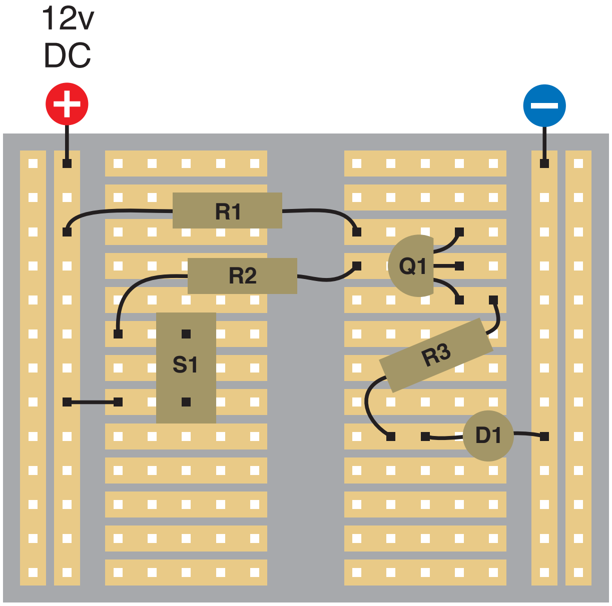

Figure 2-85. *The transistor blocks voltage that reaches it through R1. But when pushbutton S1 is pressed, this tells the transistor to allow current to pass through it. Note that transistors are always identified with letter Q in wiring diagrams and schematics.*

S1: Pushbutton, momentary, OFF (ON)
R1: 180Ω
R2: 10K
R3: 680Ω
Q1: 2N2222 or similar
D1: LED

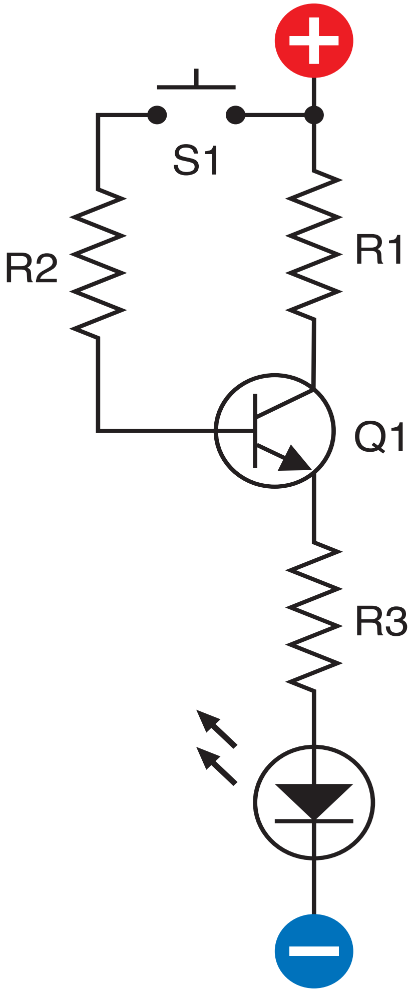

Figure 2-86. *This shows the same circuit as the breadboard diagram in Figure 2-85.*

Initially, the LED should be dark. Now press the pushbutton and the LED should glow brightly. Electricity is following two paths here. Look at the schematic in Figure 2-86, which shows the same circuit more clearly. I've shown positive at the top and negative at the bottom (the way most schematics do it) because it helps to clarify the function of this particular circuit. If you view the schematic from the side, the similarity with the breadboard layout is easier to see.

Through R1, voltage reaches the top pin (the collector) of the transistor. The transistor only lets a tiny trickle of it pass through, so the LED stays dark. When you press the button, voltage is also applied along a separate path, through R2 to the middle pin (the base) of the transistor. This tells the transistor to open its solid-state switch and allow current to flow out through its third pin (the emitter), and through R3, to the LED.

You can use your meter in volts DC mode to check the voltage at points in the circuit. Keep the negative probe from the meter touching the negative voltage source while you touch the positive probe to the top pin of the transistor, the middle pin, and the bottom pin. When you press the button, you should see the voltage change.

Fingertip Switching

Now here's something more remarkable. Remove R2 and the pushbutton, and insert two short pieces of of wire as shown in Figure 2-87. The upper piece of wire connects with the positive voltage supply; the lower piece connects with the middle pin of the transistor (its base). Now touch the tip of your finger to the two wires. Once again, the LED should glow, although not as brightly as before. Lick the tip of your finger, try again, and the LED should glow more brightly.

Never Use Two Hands

The fingertip switching demo is safe if the electricity passes just through your finger. You won't even feel it, because it's 12 volts DC from a power supply of 1 amp or less. But it's not a good idea to put the finger of one hand on one wire, and the finger of your other hand on the other wire. This would allow the electricity to pass through your body. Although the chance of hurting yourself this way is extremely small, you should never allow electricity to run through you from one hand to the other. Also, when touching the wires, don't allow them to penetrate your skin.

Your finger is conducting positive voltage to the base of the transistor. Even though your skin has a high resistance, the transistor still responds. It isn't just switching the LED on and off; it is *amplifying the current* applied to its base. This is an essential concept: *a transistor amplifies any changes in current that you apply to its base*.

Check Figure 2-88 to see more clearly what's happening.

If you studied the section "Background: Positive and negative" in Chapter 1, you learned that there is really no such thing as positive voltage. All we really have is negative voltage (created by the pressure of free electrons) and an absence of negative voltage (where there are fewer free electrons). But because the idea of a flow of electricity from positive to negative was so widely believed before the electron was discovered, and because the inner workings of a transistor involve "holes" which are an absence of electrons and can be thought of as positive, we can still pretend that electricity flows from positive to negative. See the following section, "Essentials: All about NPN and PNP transistors," for more details.

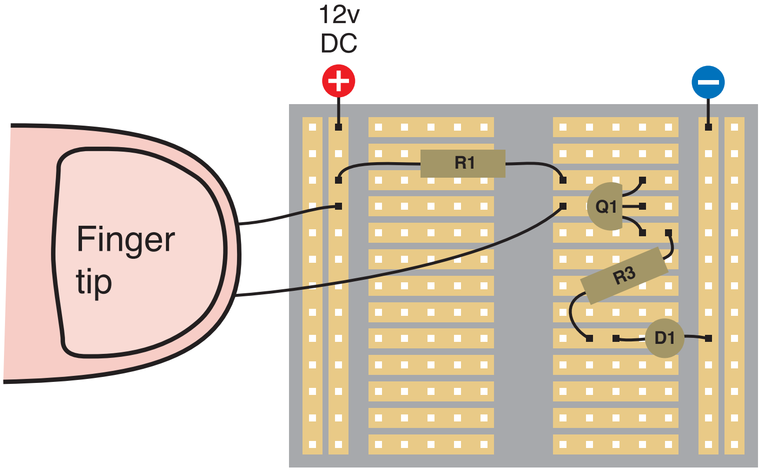

Figure 2-87

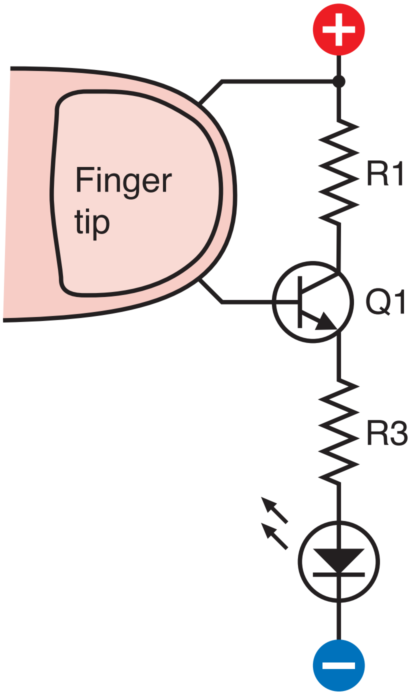

Figure 2-88. *These two diagrams show the same components as before, with a fingertip substituted for R2. Although only a trickle of voltage now reaches the base of the transistor, it's enough to make the transistor respond.*

ESSENTIALS

All about NPN and PNP transistors

A transistor is a semiconductor, meaning that sometimes it conducts electricity, and sometimes it doesn't. Its internal resistance varies, depending on the power that you apply to its base.

NPN and PNP transistors are bipolar semiconductors. They contain two slightly different variants of silicon, and conduct using both polarities of carriers—holes and electrons.

The NPN type is a sandwich with P-type silicon in the middle, and the PNP type is a sandwich with N-type silicon in the middle. If you want to know more about this terminology, and the behavior of electrons when they try to cross an NP junction or a PN junction, you'll have to read a separate source on this subject. It's too technical for this book. All you need to remember is:

- All bipolar transistors have three connections: Collector, Base, and Emitter, abbreviated as C, B, and E on the manufacturer's data sheet, which will identify the pins for you.
- NPN transistors are activated by *positive* voltage on the base relative to the emitter.
- PNP transistors are activated by *negative* voltage on the base relative to the emitter.

In their passive state, both types block the flow of electricity between the collector and emitter, just like an SPST relay in which the contacts are normally open. (Actually a transistor allows a tiny bit of current known as "leakage.")

You can think of a bipolar transistor as if it contains a little button inside, as shown in Figures 2-89 and 2-90. When the button is pressed, it allows a large current to flow. To press the button, you inject a much smaller current into the base by applying a small voltage to the base. In an NPN transistor, the control voltage is positive. In a PNP transistor, the control voltage is negative.

NPN transistor basics

- To start the flow of current from collector to emitter, apply a relatively positive voltage to the base.
- In the schematic symbol, the arrow points from base to emitter and shows the direction of positive current.
- The base must be at least 0.6 volts "more positive" than the emitter, to start the flow.
- The collector must be "more positive" than the emitter.

PNP transistor basics

- To start the flow of current from emitter to collector, apply a relatively negative voltage to the base.
- In the schematic symbol, the arrow points from emitter to base and shows the direction of positive current.
- The base must be at least 0.6 volts "more negative" than the emitter, to start the flow.
- The emitter must be "more positive" than the collector.

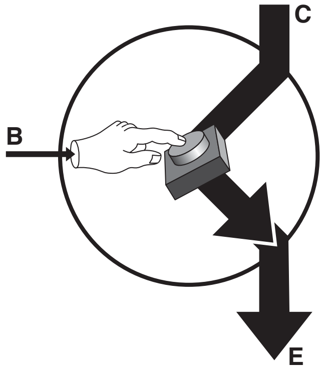

Figure 2-89. *You can think of a bipolar transistor as if it contains a button that can connect the collector and the emitter. In an NPN transistor, a small positive potential presses the button.*

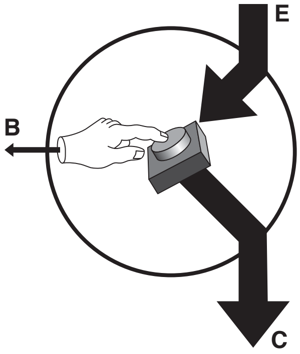

Figure 2-90. *In a PNP transistor, a small negative potential has the same effect. The arrows point in the direction of "positive current flow."*

ESSENTIALS

All about NPN and PNP transistors (continued)

All-transistor basics

- Never apply a power supply directly across a transistor. You can burn it out with too much current.
- Protect a transistor with a resistor, in the same way you would protect an LED.
- Avoid reversing the connection of a transistor between positive and negative voltages.
- Sometimes an NPN transistor is more convenient in a circuit; sometimes a PNP happens to fit more easily. They both function as switches and amplifiers, the only difference being that you apply a relatively positive voltage to the base of an NPN transistor, and a relatively negative voltage to the base of a PNP transistor.
- PNP transistors are used relatively seldom, mainly because they were more difficult to manufacture in the early days of semiconductors. People got into the habit of designing circuits around NPN transistors.
- Remember that bipolar transistors amplify current, not voltage. A small fluctuation of current through the base enables a large change in current between emitter and collector.
- Schematics sometimes show transistors with circles around them, and sometimes don't. In this book, I'll use circles to draw attention to them. See Figures 2-91 and 2-92.
- Schematics may show the emitter at the top and the collector at the bottom, or vice versa. The base may be on the left, or on the right, depending on what was most convenient for the person drawing the schematic. Be careful to look carefully at the arrow in the transistor to see which way up it is, and whether it is NPN or PNP. You can damage a transistor by connecting it incorrectly.
- Transistors come in various different sizes and configurations. In many of them, there is no way to tell which wires connect to the emitter, the collector, or the base, and some transistors have no part numbers on them. Before you throw away the packaging that came with a transistor, check to see whether it identifies the terminals.
- If you forget which wire is which, some multimeters have a function that will identify emitter, collector, and base for you. Check your multimeter instruction booklet for more details.

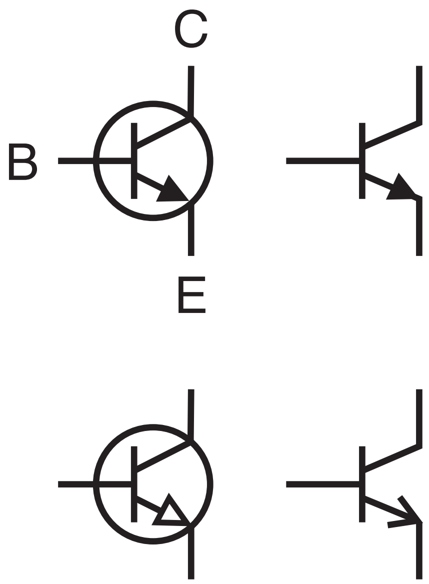

Figure 2-91. *The symbol for an NPN transistor always has an arrow pointing from its base to its emitter. Some people include a circle around the transistor; others don't bother. The style of the arrow may vary. But the meaning is always the same. The top-left version is the one I use in this book.*

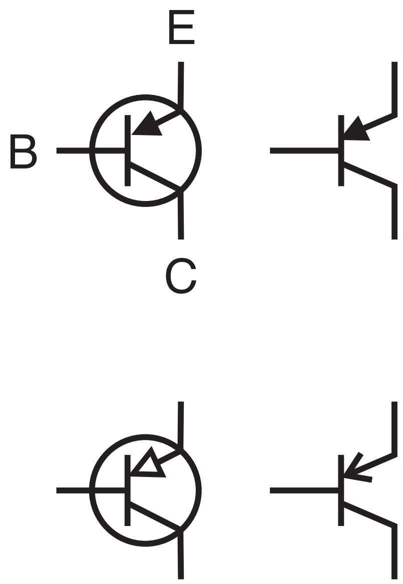

Figure 2-92. *The symbol for a PNP transistor always has an arrow pointing from its emitter to its base. Some people include a circle around the transistor; others don't bother. The style of the arrow may vary. But the meaning is always the same. The top-left version is the one I use in this book.*

BACKGROUND

Transistor origins

Though some historians trace the origins of the transistor back to the invention of diodes (which allow electricity to flow in one direction while preventing reversal of the flow), there's no dispute that the first working transistor was developed at Bell Laboratories in 1948 by John Bardeen, William Shockley, and Walter Brattain (Figure 2-93).

Shockley was the leader of the team, who had the foresight to see how potentially important a solid-state switch could be. Bardeen was the theorist, and Brattain actually made it work. This was a hugely productive collaboration—until it succeeded. At that point, Shockley started maneuvering to have the transistor patented exclusively under his own name. When he notified his collaborators, they were—naturally—unhappy about this idea.

A widely circulated publicity photograph didn't help, in that it showed Shockley sitting at the center in front of a microscope, as if he had done the hands-on work, while the other two stood behind him, implying that they had played a lesser role. In fact Shockley, as the supervisor, was seldom present in the laboratory where the real work was done.

The productive collaboration quickly disintegrated. Brattain asked to be transferred to a different lab at AT&T. Bardeen moved to the University of Illinois to pursue theoretical physics. Shockley eventually left Bell Labs and founded Shockley Semiconductor in what was later to become Silicon Valley, but his ambitions outstripped the capabilities of the technology of his time. His company never manufactured a profitable product.

Eight of Shockley's coworkers in his company eventually betrayed him by quitting and establishing their own business, Fairchild Semiconductor, which became hugely successful as a manufacturer of transistors and, later, integrated circuit chips.

Figure 2-93. *Photographs provided by the Nobel Foundation show, left to right, John Bardeen, William Shockley, and Walter Brattain. For their collaboration in development of the world's first working transistor in 1948, they shared a Nobel prize in 1956.*

ESSENTIALS

Transistors and relays

One limitation of NPN and PNP transistors is that they are naturally "off" until you turn them "on." They behave like a normally open pushbutton, which conducts electricity only for as long as you hold it down. They don't normally behave like a normal on switch, which stays on until you apply a signal to turn it off.

A relay offers more switching options. It can be normally open, normally closed, or it can contain a double-throw switch, which gives you a choice of two "on" positions. It can also contain a double-pole switch, which makes (or breaks) two entirely separate connections when you energize it. Single-transistor devices cannot provide the double-throw or double-pole features, although you can design more complex circuits that emulate this behavior.

Here's a list of transistor and relay characteristics.

	Transistor	Relay
Long-term reliability	Excellent	Limited
Configurable for DP and DT switching	No	Yes
Ability to switch large currents	Limited	Good
Able to switch alternating current (AC)	Usually not	Yes
Can be triggered by alternating current (AC)	Usually not	Optional
Suitability for miniaturization	Excellent	Very limited
Sensitive to heat	High	Moderate
Ability to switch at high speed	Excellent	Limited
Price advantage for low-voltage low-current	Yes	No
Price advantage for high-voltage high-current	No	Yes
Current leakage when "off"	Yes	No

The choice between relays or transistors will depend on each particular application.

THEORY

See the current

If you want to get a more precise understanding of how a transistor works, you should try this little test. It shows the precise behavior and limits of the 2N2222 transistor that you used in the previous experiment.

I've said that in an NPN transistor, the collector should always be more positive than the emitter and that the base should have a potential somewhere between those two voltages. Figure 2-94 shows this rather vague relationship. Now I want to substitute some numbers for these general statements.

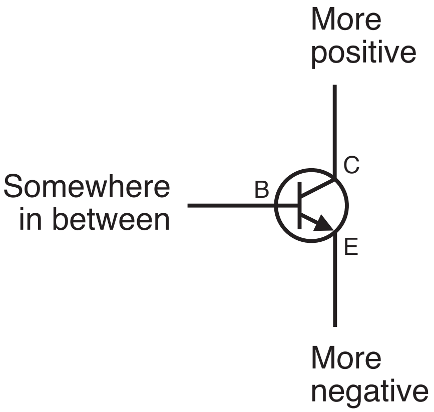

Figure 2-94. *The proper functioning of an NPN transistor requires you to maintain these voltage relationships.*

Take a look at the schematic in Figure 2-95, and check the component values. Notice that the total resistance above the transistor, from R1 + R2, is the same as the total resistance below it, from R3 + R4. Therefore the potential on the base of the transistor should be halfway between the two extremes—until you use potentiometer P1 to adjust the voltage of the base of the transistor up and down.

The two 180Ω resistors, R1 and R3, protect the transistor from passing excessive current. The two 10K resistors, R2 and R4, protect the base when the potentiometer is turned all the way up or all the way down.

I would like you to see what the transistor is doing by measuring the amperage flowing into the base at the position marked A1, and the total amperage flowing out through the emitter at the position marked A2. To do this, it would be really helpful if you had two meters. As that may be impractical, the breadboard diagrams in Figures 2-96 and 2-97 show how you can swap one meter between the two locations.

Remember that to measure milliamps, you have to pass electricity through the meter. This means that the meter must be inserted into the circuit, and whenever you remove the meter, you have to remake the connection where the meter was. The breadboard diagram shows how you can do this. Fortunately, it's very easy to remove and replace wires in a breadboard. Where wires are connected to the potentiometer, you may need to revert to using alligator clips.

Begin with the potentiometer turned about halfway through its range. Measure at A1 and A2. Turn the potentiometer up a bit, and measure current at the two locations again. Following is a table showing some actual readings I obtained at those two locations, using two digital meters simultaneously.

Milliamps passing through location A1	Milliamps passing through location A2
0.01	1.9
0.02	4.9
0.03	7.1
0.04	9.9
0.05	12.9
0.06	15.5
0.07	17.9
0.08	19.8
0.09	22.1
0.10	24.9
0.11	26.0
0.12	28.3

There's a very obvious relationship. The current emerging from the emitter of the transistor, through location A2, is about 24 times the current passing through location A1, into the base. The ratio of current coming out from the emitter of an NPN transistor to current going into the base is known as the *beta value* for a transistor. The beta value expresses the transistor's amplifying power.

It's a very constant ratio, until you push it a little too far. Above 0.12 mA, this particular transistor becomes "saturated," meaning that its internal resistance cannot go any lower.

THEORY

See the current (continued)

In my little experiment, I found that the maximum current at A2 was 33mA. A simple calculation using Ohm's Law showed me that this meant the transistor's internal resistance was near zero. This is why you should protect a transistor with some additional resistance in the circuit. If you don't, its low internal resistance would allow a huge current flow that would immediately burn it out.

What about the other end of its range? When it passes only 1.9 mA, the transistor has an internal resistance of around 6,000Ω. The conclusion is that depending how much current you apply to this transistor, its internal resistance varies between zero and 6,000Ω, approximately.

So much for the theory. Now what can we do with a transistor that's fun, or useful, or both? We can do Experiment 11!

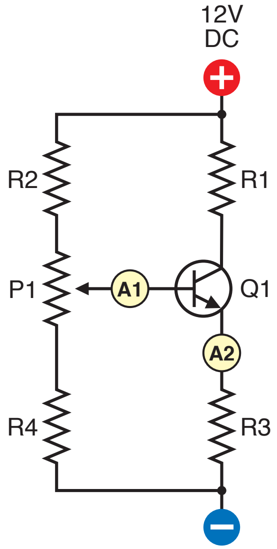

Figure 2-95. *This is basically the same as the previous circuit, with a potentiometer added and the LED removed. Component values:*

R1: 180Ω
R2: 10K
R3: 180Ω
R4: 10K
P1: 1M linear potentiometer
Q1: 2N2222 transistor

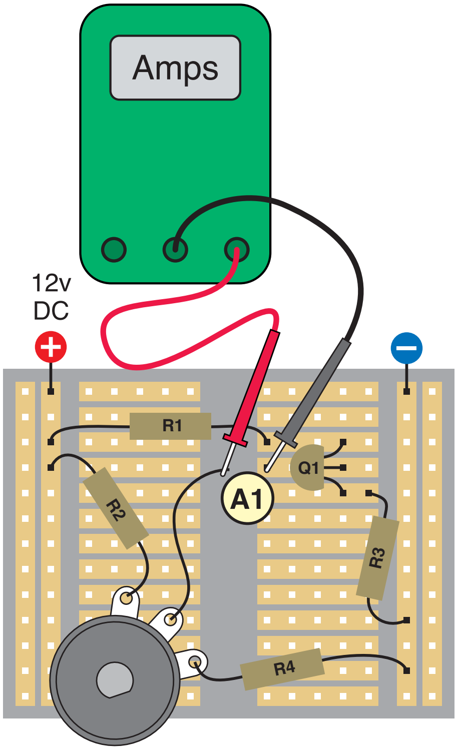

Figure 2-96. *The meter is measuring current flowing from the potentiometer into the base of the transistor at position A1 (see Figure 2-95).*

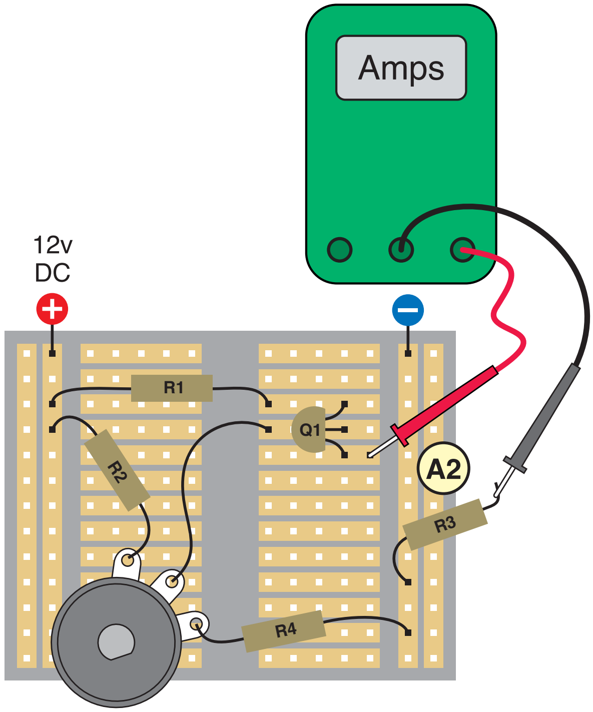

Figure 2-97. *One end of resistor R3 has been unplugged from the breadboard so that the meter now measures current flowing out through the emitter of the transistor, into R3, at position A2.*

Experiment 11: A Modular Project

You will need:

- AC adapter, breadboard, wire, and meter.
- LED. Quantity: 1.
- Resistors, various.
- Capacitors, various.
- Transistor, 2N2222 or similar. Quantity: 2.
- 2N6027 programmable unijunction transistor (PUT). Quantity: 2.
- Miniature 8Ω loudspeaker. Quantity: 1.

So far, I've described small circuits that perform very simple functions. Now it's time to show how modules can be combined to create a device that does a bit more.

The end product of this experiment will be a circuit that makes a noise like a small siren, which could be used in an intrusion alarm. You may or may not be interested in owning an alarm, but the four-step process of developing it is important, because it shows how individual clusters of components can be persuaded to communicate with each other.

I'll begin by showing how to use a transistor to make a solid-state version of the oscillating circuit that you built with a relay in Experiment 8. The relay, you may remember, was wired in such a way that the coil received power through the contacts of the relay. As soon as the coil was energized, it opened the contacts, thus cutting off its own power. As soon as the contacts relaxed they restored the power, and the process repeated itself.

There's no way to do this with a single bipolar transistor. You actually need two of them, switching each other on and off, and the way that this works is quite hard to understand. An easier option is to use a different thing known as a programmable unijunction transistor, or PUT.

Unijunction transistors were developed during the 1950s, but fell into disuse when simple silicon chips acquired the ability to perform the same kinds of functions, more accurately and more cheaply. However, the so-called programmable unijunction transistor is still widely available, often used in applications such as lamp dimmers and motor controllers. Because its primary use is in generating a stream of pulses, it's ideal for our purposes.

If you put together the components shown in Figure 2-98, the LED should start flashing as soon as you apply power.

Note that this circuit will work on 6 volts. You won't damage anything if you run it with 12 volts, but as we continue adding pieces to it, you'll find that it actually performs better at 6 volts than at 12. If you read the next section, "Essentials: All about programmable unijunction transistors," you'll find out how the circuit works.

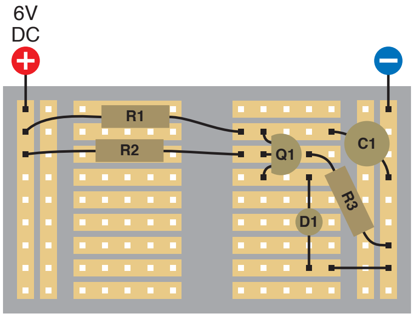

Figure 2-98. *Assemble these components, apply power, and the LED should start flashing.*

R1: 470K
R2: 15K
R3: 27K
C1: 2.2 µF electrolytic capacitor
D1: LED
Q1: 2N6027 programmable unijunction transistor

ESSENTIALS

All about programmable unijunction transistors

The schematic symbol for a programmable unijunction transistor, or PUT, looks very different from the symbol for a bipolar transistor, and its parts are named differently, too. Nevertheless, it does have a similar function as a solid-state switch. The symbol and the names of the three connections are shown in Figure 2-99.

Note that this is a rare case (maybe the only one in the whole of electronics!) in which you won't run into confusing variations of the basic schematic symbol. A PUT always seems to look the way I've drawn it here. Personally I think it would be clearer if we added a circle around it, but no one seems to do that, so I won't, either.

The 2N6027 is probably the most common PUT, and seems to be standardized in its packaging and pin-outs. I've only seen it in a plastic module rather than a little tin can. Figure 2-100 shows the functions of the leads if your 2N6027 is manufactured by Motorola or On Semiconductor. If you have one from another source, you should check the data sheet.

Note that the flat side of the plastic module faces the opposite way around compared with the 2N2222 bipolar transistor, when the two devices are functioning similarly.

The PUT blocks current until its internal resistance drops to allow flow from the "anode" to the "cathode." In this way, it seems very similar to an NPN transistor, but there's a big difference in the circumstances that cause the PUT to lower its resistance. The voltage at the anode determines when the PUT allows current to flow.

Suppose you start with, say, 1 volt at the anode. Slowly, you increase this voltage. The transistor blocks it until the anode gets close to 6 volts. Suddenly this pressure breaks down the resistance and current surges from the anode to the cathode. If the voltage goes back down again, the transistor reverts to its original state and blocks the flow.

I've included another version of the "finger on the button" drawing to convey this concept. The voltage on the anode is itself responsible for pushing the button that opens the pathway to the cathode. See Figure 2-101.

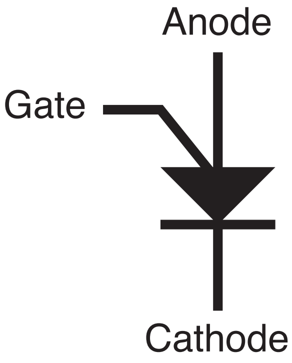

Figure 2-99. *The schematic symbol for a PUT.*

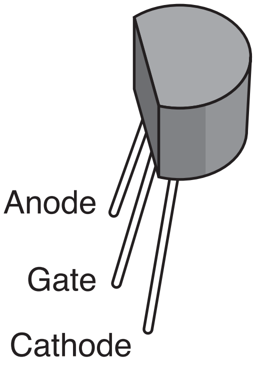

Figure 2-100. *In PUTs manufactured by On Semiconductor and Motorola, the leads have these functions.*

ESSENTIALS

All about programmable unijunction transistors (continued)

This may cause you to wonder what the function of the gate is. You can think of it as “assisting” the finger on the button. In fact, the gate is the “programmable” part of a PUT. By choosing a voltage for the gate, you establish the threshold point when current starts to flow.

Here’s a simple take-home summary:

- The anode has to be more positive than the cathode, and the gate should be between those two extremes.
- If anode voltage increases above a threshold point, current bursts through and flows from the anode to the cathode.
- If anode voltage drops back down below the threshold, the transistor stops the flow.
- The voltage you apply to the gate determines how high the threshold is.
- The gate voltage is adjusted with two resistors, shown as R1 and R2 in the simple schematic in Figure 2-102. Typically, each resistor is around 20K. The PUT is protected from full positive voltage by R3, which can have a high value, 100K or greater, because very little current is needed to bias the transistor.
- You add your input signal in the form of positive voltage at the anode. When it exceeds the threshold, it flows out of the cathode and can work some kind of output device.

The only remaining question is how we make a PUT oscillate, to create a stream of on/off pulses. The answer is the capacitor that you included in the circuit that you breadboarded at the beginning of Experiment 11.

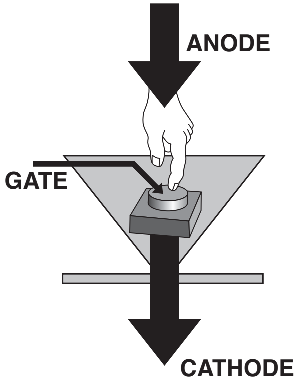

Figure 2-101. *When voltage at the anode of a PUT crosses a threshold (determined by a preset voltage at the gate), current breaks through and surges from the anode to the cathode. In this sense, the anode voltage acts as if it presses a button itself to open a connection inside the PUT, with some assistance from control voltage at the gate.*

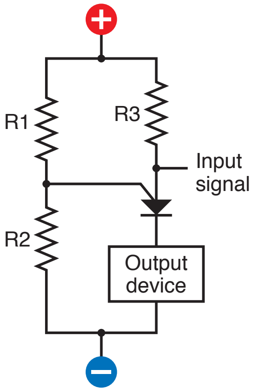

Figure 2-102. *This simple schematic shows how a PUT is used. R1 and R2 determine the voltage at the gate, which sets the threshold point for the input at the anode. Above the threshold, current flows from anode to cathode.*

Step 1: Slow-Speed Oscillation

Figure 2-103 is a schematic version of the previous PUT breadboard circuit shown in Figure 2-98, drawn so that the layout looks as much like the breadboard as possible.

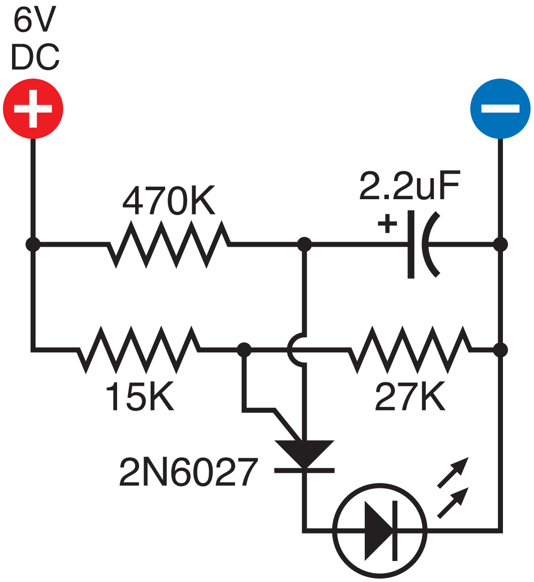

Figure 2-103. *This makes it easier to see what's happening in the breadboard version.*

The 15K resistor and 27K resistor establish the voltage at the gate. The 470K resistor supplies the anode of the PUT, but the PUT begins in its "off" condition, blocking the voltage. So the voltage starts to charge the 2.2 μF capacitor.

You may remember that a resistor slows the rate at which a capacitor accumulates voltage. The bigger the resistor and/or the larger the capacitor, the longer the capacitor takes to reach a full charge. In this circuit, the capacitor takes about half a second to get close to 6 volts.

But notice that the PUT is connected directly with the capacitor. Therefore, whatever voltage accumulates on the capacitor is also experienced by the PUT. As the voltage gradually increases, finally it reaches the threshold, which flips the PUT into its "on" state. The capacitor immediately discharges itself through the PUT, through the LED (which flashes), and from there to the negative side of the power supply.

The surge depletes the capacitor. The voltage drops back down, and the PUT returns to its original state. Now the capacitor has to recharge itself all over again, until the whole process repeats itself.

If you substitute a 22 μF capacitor, the charge/discharge cycle should take about 10 times as long, which will give you time to measure it. Set your meter to measure volts DC and place its probes on either side of the capacitor. You can actually watch the charge increasing until it reaches the threshold, at which point the capacitor discharges and the voltage drops back down again.

So now we have an oscillator. What's next?

Step 2: Beyond the Persistence of Vision

If you substitute a much smaller capacitor, it will charge much more quickly, and the LED will flash faster. Suppose you use a capacitor of 0.0047 µF (which can also be expressed as 47 nanofarads, or 47 nF). This seems like an odd number, but it's a standard value for a capacitor. This will reduce the capacitance by a factor of more than 500, and therefore the LED should flash about 500 times as fast, which should be about 1,000 times per second. The human eye cannot detect such rapid pulses. The human ear, however, can hear frequencies up to 10,000 per second and beyond. If we substitute a miniature loudspeaker for the LED, we should be able to hear the oscillations.

Figure 2-104 shows how I'd like you to make this happen. Please leave your original, slow-flashing circuit untouched, and make a duplicate of it farther down the breadboard, changing a couple of component values as indicated. In the schematic in Figure 2-105, the new part of the circuit is in solid black, while the previous section is in gray.

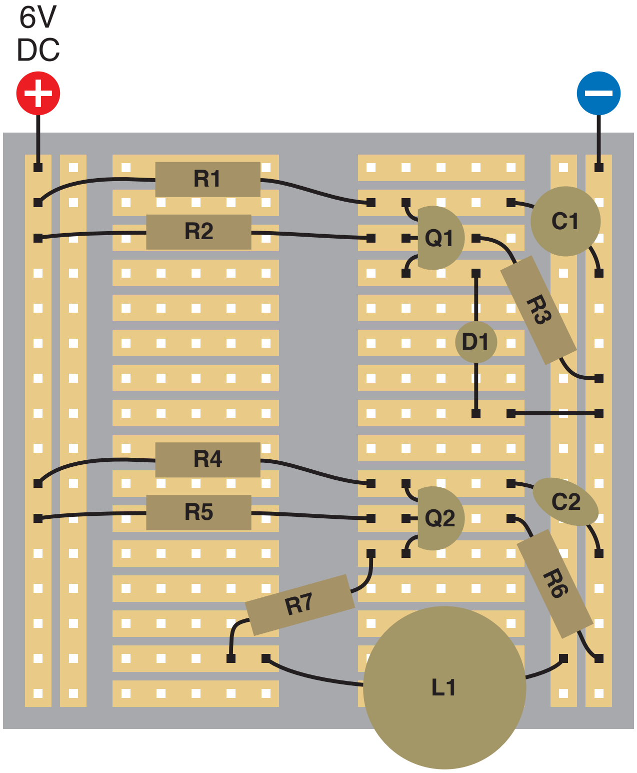

Figure 2-104. *The extra components which have been added at the lower half of the breadboard have the same functions as the components at the top, but some values are slightly different:*

R4: 470K
R5: 33K
R6: 27K
R7: 100Ω
C2: 0.0047 µF
Q2: 2N6027
L1: 8Ω 1-inch loudspeaker

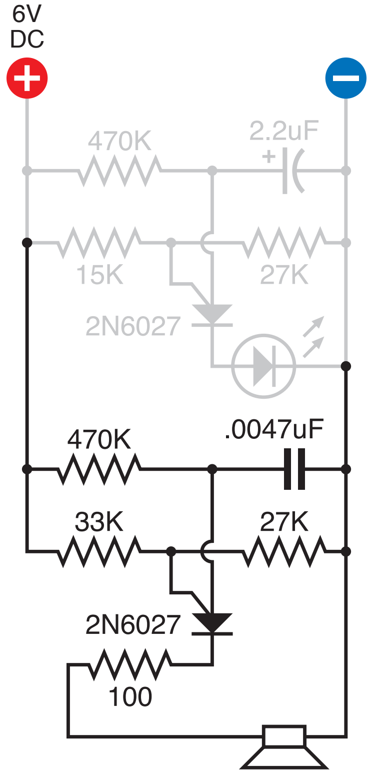

Figure 2-105. *The previous section that you built is shown in gray. Just add the new section in black.*

I want you to keep the slow-flashing circuit separately, untouched, because I have an idea to make use of it a little later. You can leave the LED blinking.

The loudspeaker should be wired in series with a 100Ω resistor to limit the current that flows out of the PUT. The loudspeaker doesn't have any polarity, even though it is fitted with a red wire and a black wire. You can connect it either way around.

Initially, you may be disappointed, because the circuit will not seem to be doing anything. However, if you place your ear very, very close to the loudspeaker, and if you wired the circuit correctly, you should hear a faint buzz, like a mosquito. Obviously, this isn't loud enough to serve any practical purpose. We need to make it louder. In other words, we need to amplify it.

Maybe you remember that the 2N2222, which you played with previously, can function as an amplifier. So let's try using that.

Step 3: Amplification

Disconnect the loudspeaker and its 100Ω series resistor. Then add the 2N2222, which is linked with the output from the PUT via a 1K resistor to protect it from excessive current. See Figure 2-107.

The emitter of the 2N2222 is connected to ground, and the collector is supplied through the loudspeaker and its 100Ω series resistor. This way, small fluctuations in the output from the PUT are sensed by the base of the 2N2222 which converts them into bigger fluctuations between the collector and the emitter, which draw current through the loudspeaker. Check the schematic in Figure 2-108.

Now the sound should be louder than an insect buzz, but still not really loud enough to be useful. What to do?

Well—how about if we add another 2N2222? Bipolar transistors can be placed in series, so that the output from the first one goes to the base of the second one. The 24:1 amplification of the first one is multiplied by another 24:1, giving a total amplification of more than 500:1.

There are limits to this technique. The 2N2222 can only conduct so much current before getting overloaded, and excess amplification can cause distortion. But when I built this circuit, I used a meter to verify that we're still within the design limits of a 2N2222, and for this project, I don't care whether the sound is slightly distorted.

BACKGROUND

Mounting a loudspeaker

The diaphragm or cone of a loudspeaker is designed to radiate sound, but as it oscillates to and fro, it emits sound from its back side as well as its front side. Because the sounds are opposite in phase, they tend to cancel each other out.

The perceived output from a loudspeaker can increase dramatically if you add a horn around it in the form of a tube to separate the output from the front and back of the speaker. For a miniature 1-inch loudspeaker, you can bend and tape a file card around it. See Figure 2-106.

Better still, mount it in a box so that the box absorbs the sound from the rear of the loudspeaker. For purposes of these simple experiments, I won't bother to go into the details of vented enclosures and bass-reflex designs.

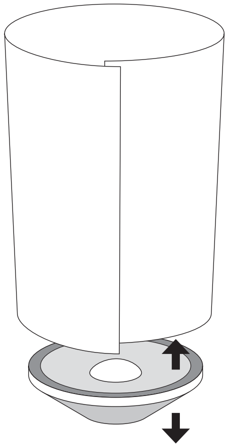

Figure 2-106. *A loudspeaker emits sound from its bottom surface as well as its top surface. To increase the perceived audio volume, use a cardboard tube to separate the two sound sources, or mount the speaker in a small box.*

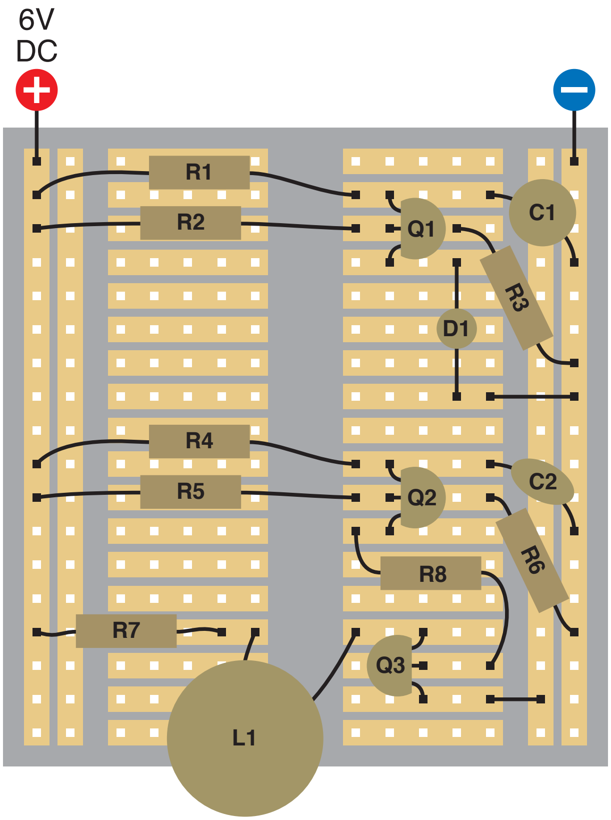

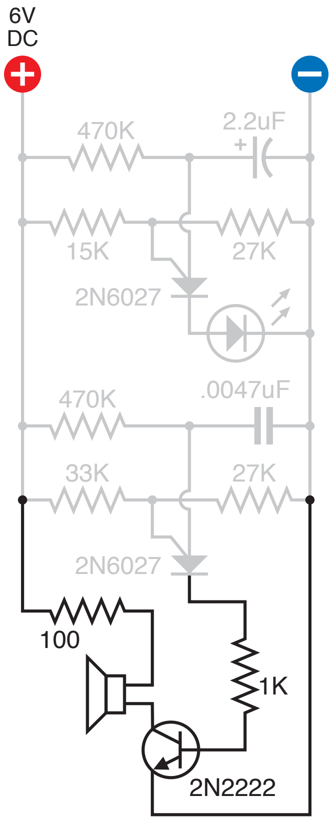

Figure 2-107. *By adding a 2N2222 general-purpose transistor, we amplify the signal from Q2:*

R8: 1K
Q3: 2N2222

Other components are the same as in the previous step in constructing this circuit.

Figure 2-108

Add the second 2N2222 as shown in Figure 2-109. In Figure 2-110, once again the previously wired section is in gray.

If the accumulation of electrical components is beginning to seem confusing, remember that each cluster of parts has a separate defined function. We can draw a block diagram to illustrate this, as in Figure 2-112.

Using the second 2N2222, you should find that the output is more clearly audible, at least within the limits of your tiny 1-inch loudspeaker. Cup your hands around it to direct the sound, and you'll find that the volume seems to increase. You can also try using a 3-inch loudspeaker, which will create a generally better audio output while still remaining within the limits of the little 2N2222 transistor. See Figure 2-106, shown previously, and Figure 2-111.

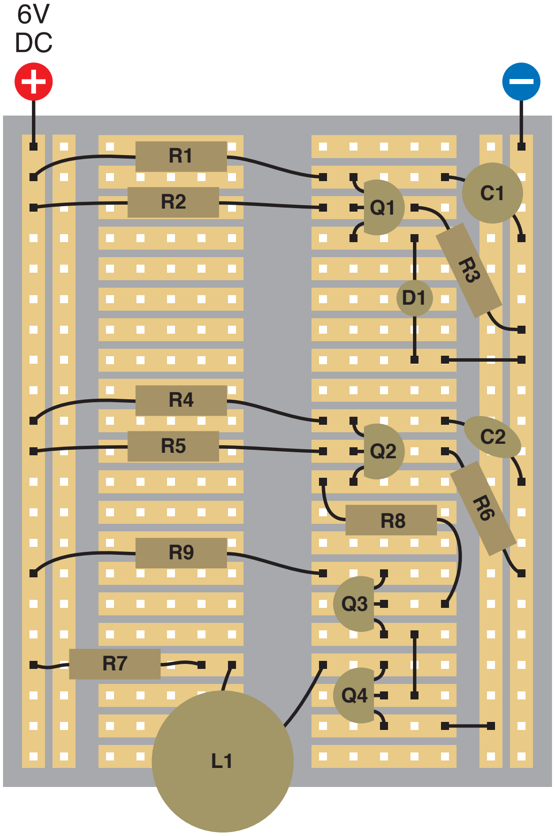
6V
DC
R1
R2
Q1
C1
R3
D1
R4
R5
Q2
C2
R6
R8
R9
Q3
R7
Q4
L1

Figure 2-109. *Q4 is another 2N2222 transistor that further amplifies the signal. It receives power through R9: 2.2K.*

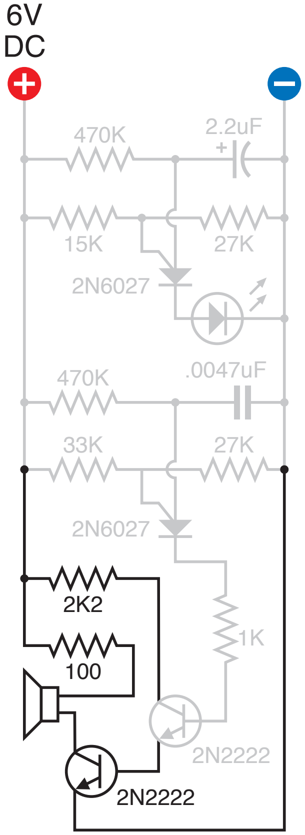
6V
DC
470K
2.2uF
+
15K
27K
2N6027
470K
.0047uF
33K
27K
2N6027
2K2
1K
100
2N2222
2N2222

Figure 2-110. *This schematic is comparable with the component layout in Figure 2-109.*

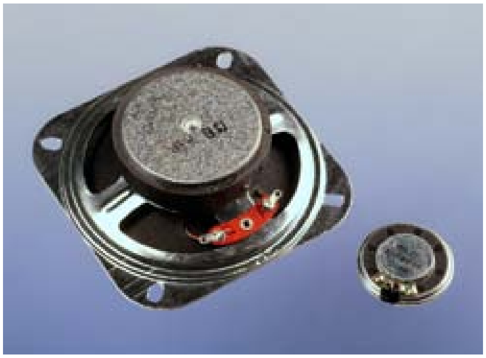

Figure 2-111. *The 2N2222 transistor is quite capable of driving a 3-inch loudspeaker, which will create much better sound than a 1-inch speaker.*

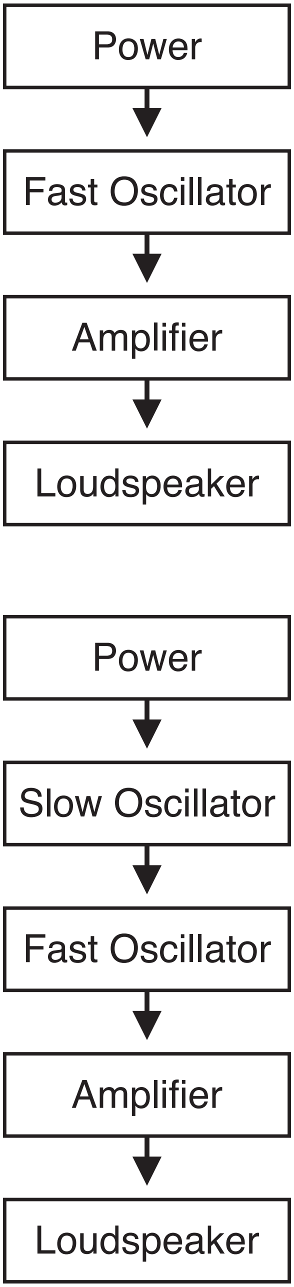

Figure 2-112. *Top: The basic functions of the noisemaking oscillator circuit shown as a block diagram. Bottom: The same functions with a slow oscillator added to control the fast oscillator.*

Step 4: Pulsed Output

If you wanted to use this audio signal as some kind of an alarm, a steady droning noise is not very satisfactory. A pulsing output would be a much better attention-getter.

Well, the first section of the circuit that you assembled created a pulsing signal about twice per second. You used it to flash an LED. Maybe we can get rid of the LED and feed the output from the first section to the second section. The lower block diagram in Figure 2-112 explains this concept.

Can it really be that simple? Well, yes and no. The trick is to make the output from the first section compatible with the input to the second section. If you simply connect a wire from the cathode of the first PUT to the anode of the second PUT, that's not going to work, because the second PUT is already oscillating nicely between low and high voltage, about 1,000 times each second. Add more voltage, and you will disrupt the balance that enables oscillation.

However, remember that the voltage on the gate of a PUT affects its threshold for conducting electricity. Maybe if we connect the output from Q1 to the gate of Q2, we'll be able to adjust that threshold automatically. The voltage still has to be in a range that the PUT finds acceptable, though. We can try various resistors to see which one works well.

This sounds like trial and error—and that's exactly what it is. Doing the math to predict the behavior of a circuit like this is far too complicated—for me, anyway. I just looked at the manufacturer's data sheet, saw the range of resistor values that the PUT would tolerate, and chose one that seemed as if it should work.

If you remove the LED and substitute R10 as shown in the breadboard diagram in Figure 2-113, you'll find that the fluctuating output from Q1 makes Q2 emit a two-tone signal. This is more interesting, but still not what I want. I'm thinking that if I make the pulses out of Q1 less abrupt, the result could be better, and the way to smooth a pulsing output is to hook up another capacitor that will charge at the beginning of each pulse and then release its charge at the end of each pulse. This is the function of C3 in Figure 2-114, and it completes the circuit so that it makes a whooping sound almost like a "real" alarm.

If you don't get any audio output, check your wiring very carefully. It's easy to make a wrong connection on the breadboard, especially between the three legs of each transistor. Use your meter, set to DC volts, to check that each section of the circuit has a positive voltage relative to the negative side of the power supply.

Figure 2-115 shows how your circuit should actually look on the breadboard.

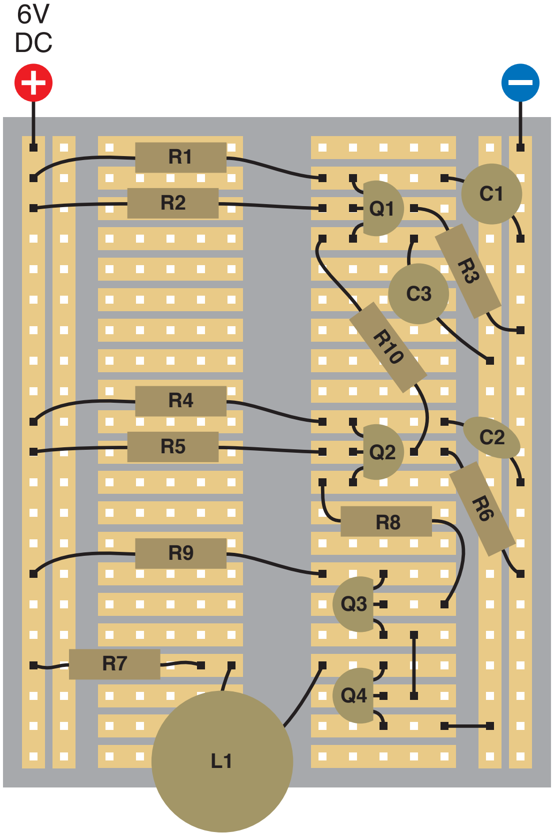
6V
DC
R1
R2
Q1
C1
R3
C3
R10
R4
R5
Q2
C2
R6
R8
R9
Q3
R7
Q4
L1

Figure 2-113. *R10 connects the slow-running oscillator at the top of the breadboard to the gate of Q2, the PUT in the middle of the breadboard. This modulates the audio oscillator, with addition of a smoothing capacitor.*

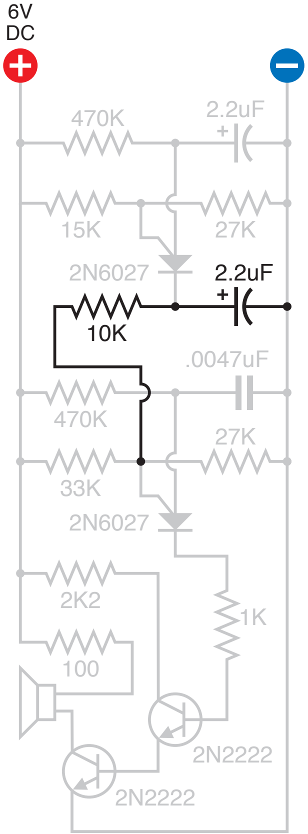
6V
DC
470K
2.2uF
+
15K
27K
2N6027
2.2uF
+
10K
.0047uF
470K
27K
33K
2N6027
2K2
1K
100
2N2222
2N2222

Figure 2-114. *This schematic shows the same circuit as in Figure 2-113:*

R10: 10K
C3: 2.2 µF

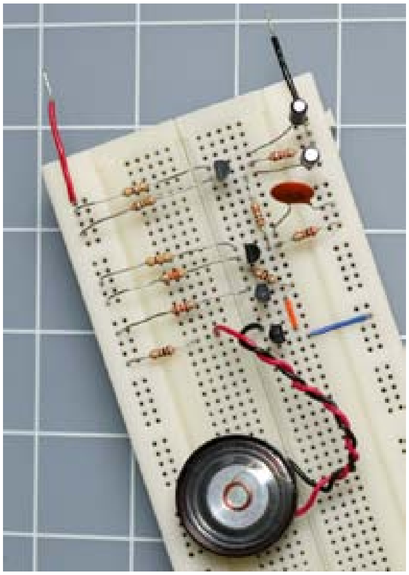

Figure 2-115. *This photograph shows the complete alarm-audio circuit on a breadboard.*

Tweaking it

There's still a lot of room for creativity here:

- Adjust the frequency of the sound: Use a smaller or larger capacitor instead of C2 (half or twice the current value). Use a smaller or larger value for R5.
- Adjust the pulsing feature: Use a smaller or larger capacitor instead of C1 (half or twice the current value). Use a smaller or larger value for R2.
- General performance adjustments: try a larger value for R1. Try smaller or larger values for C3.
- Try running the circuit at 7.5 volts, 10 volts, and 12 volts.

The circuits in this book are suggested as only a starting point. You should always try to tweak them to make them your own. As long as you follow the general rule of protecting transistors and LEDs with resistors, and respecting their requirements for positive and negative voltage, you're unlikely to burn them out. Of course, accidents will happen—I myself tend to be careless, and fried a couple of LEDs while working on this circuit, just because I connected them the wrong way around.

Step 5: Enhancements

A noisemaking circuit is just the output of an alarm. You would need several enhancements to make it useful:

1. Some kind of an intrusion sensor. Maybe magnetic switches for windows and doors?
2. A way to start the sound if any one of the sensors is triggered. The way this is usually done is to run a very small but constant current through all of the switches in series. If any one switch opens, or if the wire itself is broken, this interrupts the current, which starts the alarm. You could make this happen with a double-throw relay, keeping the relay energized all the time until the circuit is broken, at which point, the relay relaxes, opening one pair of contacts and closing the other pair, which can send power to the noisemaker.

   The trouble is that a relay draws significant power while it's energized, and it also tends to get hot. I want my alarm system to draw very little current while it's in "ready" mode, so that it can be powered by a battery. Alarm systems should never depend entirely on AC house current.

   If we don't use a relay, can we use a transistor to switch on the rest of the circuit when the power is interrupted? Absolutely; in fact, one transistor will do it.
3. But how do we arm the alarm in the first place? Really, we need a three-step procedure. First, check a little light that comes on when all the doors and windows are closed. Second, press a button that starts a 30-second countdown, giving you time to leave, if that's what you want to do. And third, after 30 seconds, the alarm arms itself.

4. If the alarm is triggered, what then? If someone forces open a window, should the alarm stop sounding as soon as the window is closed again? No, the alarm should lock itself on, until you turn it off.
5. How do you turn it off? Some kind of secret-code keypad would be good.
6. But to avoid driving everyone crazy if the alarm is triggered when you're not there, it should eventually stop itself, perhaps after about 10 minutes. At that point it should remain quiet but should light an LED to tell you what happened. You can then press a reset button to switch off the LED.

Implementing the Wish List

I've compiled a wish list that seems likely to make the project at least five times as complicated as it is already. Well, that's what tends to happen when you go beyond little demo circuits and try to design something that will be useful in everyday life. Suddenly you find yourself having to accommodate all kinds of circumstances and situations.

Actually, I can and will show you how to take care of all the enhancements on the wish list, but I'm thinking that they will require us to get a little more serious about electronics projects in general first. If you're going to build something ambitious, you'll want to make it more permanent, and probably more compact, than a breadboard with components pushed into it.

You will need to know how to connect everything permanently with solder, on a piece of perforated board that you can install in a neat little project box with switches and lights on the outside.

I'm going to deal with soldering and packaging in the next chapter. After that, we'll get back to the alarm project.

3

Getting Somewhat More Serious

IN THIS CHAPTER

Shopping List: Experiments 12 Through 15
Experiment 12: Joining Two Wires Together
Experiment 13: Broil an LED
Experiment 14: A Pulsing Glow
Experiment 15: Intrusion Alarm Revisited

I don't know how far you'll want to delve into electronics, but I do know that I've shown you about as much as I can with just a handful of components, some wires, a breadboard, and a few tools. To continue, you'll need:

- Some more tools and supplies
- Basic soldering skills
- Additional knowledge about:
  - Integrated circuits
  - Digital electronics
  - Microcontrollers
  - Motors

The tools are not particularly exotic or expensive, and the soldering skills are easily acquired. Learning to join wires with solder is far easier than mastering high-level crafts such as jewelry making or welding.

As for additional areas of knowledge about electronics, they are no more challenging than those that I have covered already.

By the end of this section, you should be able to transplant components from a breadboard onto perforated board, where you will solder everything together, and then mount the board in a little box with switches and warning lights on the front, for everyday use.

Shopping List: Experiments 12 Through 15

Tools

Each of the following tools is rated Essential, Recommended, or Optional. The Essential tools will take you through this chapter of the book. If you supplement them with the Recommended tools, they should be sufficient to get you to the end of the book. The Optional tools will make your work easier, but whether they're worth the money is for you to decide. Remember that URLs for manufacturers and sources of supply are all listed in the appendix.

I am assuming that you already have some commonly used workshop basics, such as an electric drill.

Essential: Pencil-type 15-watt soldering iron

Examples are RadioShack part 64-2051, McMaster-Carr catalog item 7016A34, or Xytronic model 252. See Figure 3-1. Soldering irons rated at 15 watts are less common than those that deliver 25 watts or more. Still, the 15-watt size is desirable for the small-scale work you'll be doing, and greatly reduces the risk of damaging components by inflicting excess heat.

When comparing prices, remember that a plated tip, which costs a little more, will last longer, will be easier to keep clean, and will conduct heat more reliably than a plain copper tip. If the manufacturer's specification doesn't mention a plated tip, the soldering iron probably doesn't have one.

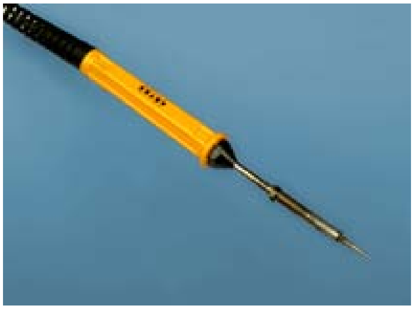

Figure 3-1. *The low wattage of this pencil-style soldering iron enables you to use it safely on sensitive components, and the sharp tip helps to apply heat selectively.*

Essential: General-duty soldering iron, 30 to 40 watts

Although most of the projects in this book entail small, heat-sensitive components and thin wire, at some point you're likely to want to make a solder joint with larger components and/or thicker wire. A 15-watt soldering iron will be unable to deliver enough heat. You should consider having a larger soldering iron in reserve, especially because they are relatively inexpensive.

Personally, I like the Weller Therma-Boost, shown in Figure 3-2, because it has an extra button that delivers more heat on demand. This is useful when you want the iron to get hot quickly, or if you are trying to solder very thick wire, which absorbs a lot of heat.

If you can't find or don't like the Weller, almost any 30-watt or 40-watt soldering iron will do. Check eBay or your local hardware store.

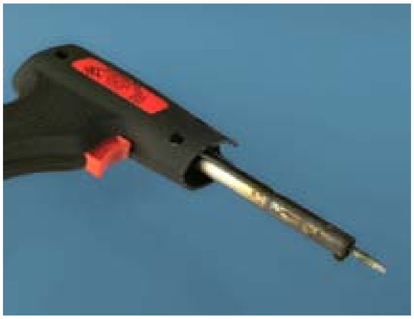

Figure 3-2. *This higher-wattage soldering iron delivers the additional heat necessary for thicker wire or larger components. The discoloration quickly occurs as a result of everyday use and has no effect on the capability of the iron, as long as the tip of it is clean.*

Essential: Helping hand

The so-called "helping hand" (or "third hand") has two alligator clips that hold components or pieces of wire precisely in position while you join them with solder. Some versions of the "helping hand" also feature a magnifying lens, a wire spiral in which you can rest your soldering iron, and a little sponge that you use to clean the tip of your iron when it becomes dirty. These additional features are desirable. Helping hands are available from all electronics hobby sources. Examples are the catalog item HH55 from Elenco or model 64-2991 from RadioShack. See Figure 3-3.

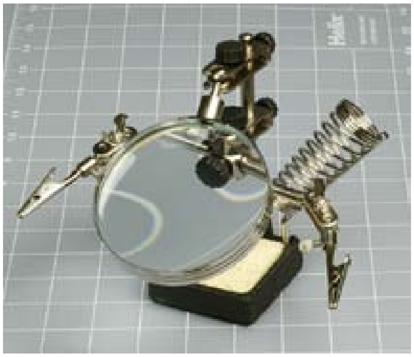

Figure 3-3. *The helping hand is fitted with two alligator clips to hold your work. The metal spiral is a safe place to holster a hot soldering iron, and you use the sponge to wipe its tip.*

Essential: Magnifying lens

No matter how good your eyes are, a small, handheld, powerful magnifying lens is essential when you are checking solder joints on perforated board. The three-lens set in Figure 3-4 is designed to be held close to your eye, and is more powerful than the large lens on a "helping hand." The folding lens in Figure 3-5 stands on your workbench for hands-free operation. Both are available from RadioShack and similar items are stocked by art supply stores and hobby shops. Plastic lenses are quite acceptable if you treat them carefully.

Figure 3-4. *As long as you treat it carefully, a cheap set of plastic magnifying lenses is perfectly acceptable. Handheld magnification is essential for inspecting the solder joints that you make on perforated board.*

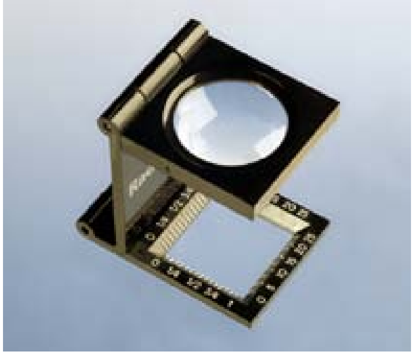

Figure 3-5. *This kind of folding magnifier can stand on your desktop and is useful for checking part numbers on tiny components.*

Essential: Clip-on meter test leads

The probes that came with your multimeter require you to hold them in contact while you make a reading. This requires both hands, preventing you from doing anything else at the same time.

When you use a pair of "minigrabber" probes with little spring-loaded clips at the end, you can attach the Common (negative) lead from your meter to the negative side of your circuit and leave it there, while you touch or attach the positive probe elsewhere.

The Pomona model 6244-48-0 (shown in Figure 3-6) from Meter Superstore and some other suppliers is what you need. If you have trouble finding it or you object to the cost, you may consider making your own by buying a couple of "banana plugs" (such as RadioShack part 274-721) that will fit the sockets on your meter, and then use 16-gauge or thicker stranded wire to connect the plugs with IC test clips, such as Kobiconn 13IC331 or RadioShack "mini hook clips," part number 270-372C. See Figures 3-7 and 3-8.

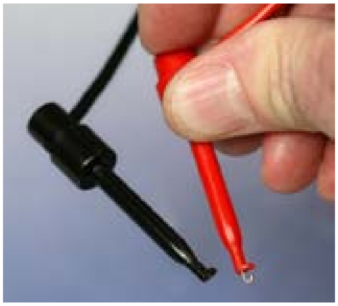

Figure 3-6. *These "minigrabber" add-ons for meter leads make it much easier to measure voltage or current. Push the spring-loaded button, and a little copper hook slides out. Attach it to a wire, release the button, and you have your hands free for other tasks. It's a mystery that meters are not supplied with these grabbers as standard equipment.*

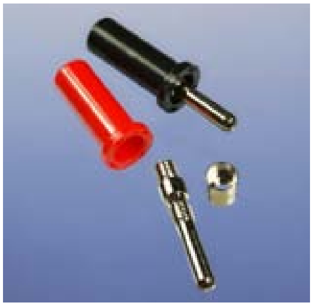

Figure 3-7. *To make your own minigrabber meter leads, first attach a banana plug to a wire by sliding the wire through the cap, into the plug, and out through a hole in the side.*

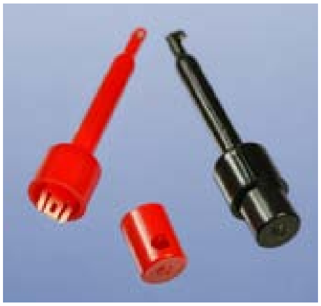

Figure 3-8. *Then screw a collar over the protruding piece of wire, and screw on the cap. The other end of the wire is soldered to a probe.*

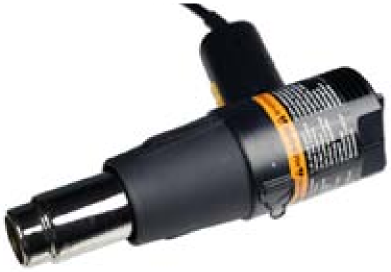

Figure 3-9. *Like an overpowered hair dryer, the heat gun is used with heat-shrink tubing to create a snug, insulated sheath around bare wire.*

Essential: Heat gun

After you join two wires with solder, you often need to insulate them. Electrical tape, sometimes called insulating tape, is messy and tends to come unstuck. You'll be using heat-shrink tube, which forms a safe, permanent sheath around a bare-metal joint. To make the tube shrink, use a heat gun, which is like a very powerful hair dryer. They're available from any hardware supply source, and I suggest you buy the cheapest one you can find. See Figure 3-9.

Essential: Solder pump

This little gadget sucks up hot, melted solder when you are trying to remove a solder joint that you made in the wrong place. Available from All Electronics (catalog item SSR-1) or RadioShack 64-2086. See Figure 3-10.

Essential: Desoldering wick

Also known as desoldering braid. See Figure 3-11. You use this to soak up solder, in conjunction with the Solder Pump. Available from All Electronics (catalog item SWK) or RadioShack (part 64-2090).

Essential: Miniature screwdriver set

Dinky little electronic parts often have dinky little screws in them, and if you try to use the wrong size of screwdriver, you'll tend to mash the heads of the screws. I like the Stanley precision set, part number 66-052, shown in Figure 3-12. But any set will do as long as it has both small Phillips and straight-blade screwdrivers.

Recommended: Soldering stand

Like a holster for a gun, you rest your soldering iron in this stand when the iron is hot but not on use. Examples are catalog item 50B-205 from All Electronics, RadioShack model 64-2078, or check eBay. See Figure 3-13. This item may be built into the helping hand, but you need an extra one for your second soldering iron.

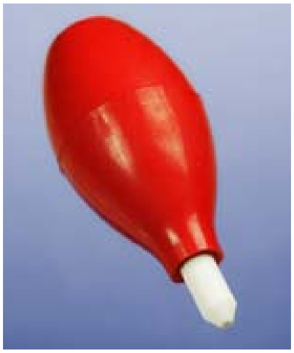

Figure 3-10. *To remove a solder joint, you can heat the solder until it's liquid, then suck it up into this squeezable rubber bulb.*

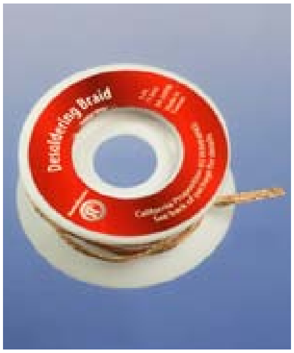

Figure 3-11. *An additional option for removing liquid solder is to soak it up in this copper braid.*

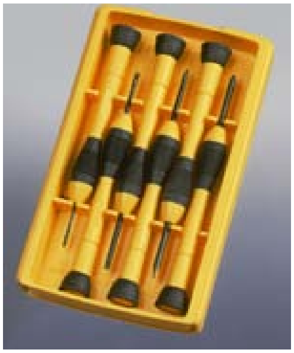

Figure 3-12. *A set of small screwdrivers is essential.*

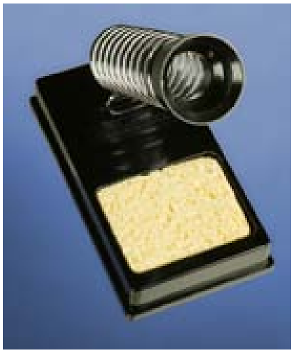

Figure 3-13. *A safe and simple additional stand for a hot soldering iron.*

Recommended: Miniature hand saw

I assume that you will want to mount a finished electronics project in a decent-looking enclosure. Consequently, you are likely to need tools to cut, shape, and trim thin plastic. For example, you may want to cut a square hole so that you can mount a square power switch in it.

Power tools are not suitable for this kind of delicate work. A miniature handsaw (a.k.a. a "hobby saw") is ideal for trimming things to fit. X-Acto makes a range of tiny saw blades. I suggest the #15 blade, plus the handle that it fits in, shown in Figure 3-14. Available online from Tower Hobbies, Hobbylinc, ArtCity, and many other arts/crafts sources. Also look for the larger X-Acto saw blade, #234 or #239, which you can use for cutting perforated board.

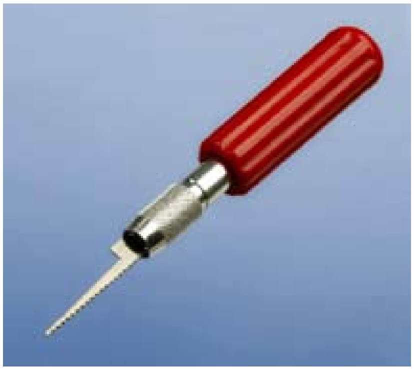

Figure 3-14. *X-Acto makes a range of small saw blades that are ideal for cutting square holes to mount components in plastic panels.*

Recommended: Miniature vise

A miniature vise can do things that the helping hand cannot. I use mine when I'm sawing small pieces of plastic and as a dead weight to anchor a piece of perforated board while I'm working on it. See Figure 3-15.

Look for a cast-iron vise that is listed as being 1 inch or slightly larger, available from Megahobby, eBay, and other arts/crafts sources. Also consider the PanaVise, which has a tilting head to allow you to turn your work to any angle.

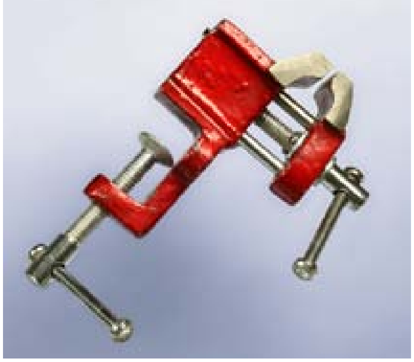

Figure 3-15. *This one-inch vise is available from the McMaster-Carr catalog.*

Recommended: Deburring tool

A deburring tool instantly smoothes and bevels any rough edge (when you have sawn or drilled a piece of plastic, for instance) and also can enlarge holes slightly. This may be necessary because some components are manufactured to metric sizes, which don't fit in the holes that you drill with American bits. Your small local hardware store may not stock deburring tools, but they are very inexpensively available from Sears, McMaster-Carr, KVM Tools, or Amazon. See Figure 3-16.

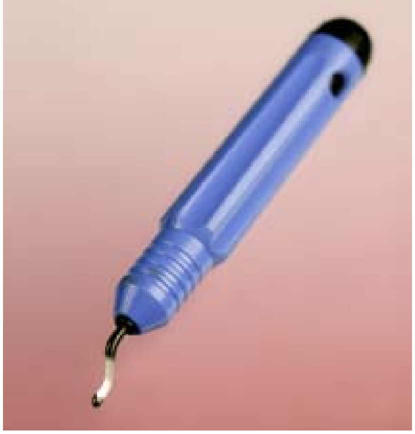

Figure 3-16. *This cunning little blade, safety-tipped with a round bump on the end, removes rough edges from saw cuts with a single stroke, and can enlarge holes that are almost big enough—but not quite.*

Optional: Hand-cranked countersink

You need a countersink to bevel the edges of screw holes to accept flat-headed screws. If you use a countersink bit in an electric drill, it won't give you precise control when you're working with thin, soft plastic.

Handheld countersinks that you grasp and turn like a screwdriver are easy to find, but McMaster-Carr (catalog item 28775A61) is the only source I've found for a hand-cranked tool that is much quicker to use. It comes with a set of bits, as shown in Figure 3-17.

Optional: Pick and hook set

Made by Stanley, part number 82-115, available from Amazon and hardware stores. You can find imported imitations for a few dollars less. See Figure 3-18.

Optional: Calipers

These may seem like a luxury, but are useful for measuring the external diameter of a round object (such as the screw thread on a switch or a potentiometer) or the internal diameter of a hole (into which you may want a switch or potentiometer to fit).

I like Mitutoyo calipers, and the low-end model 505-611 (shown in Figure 3-19) does everything I need. You can find cheaper brands, but economizing on precision measuring tools may not be a wise policy in the long term. The manufacturer's site will show you all their available models, after which you can Google "Mitutoyo" to find retail sources.

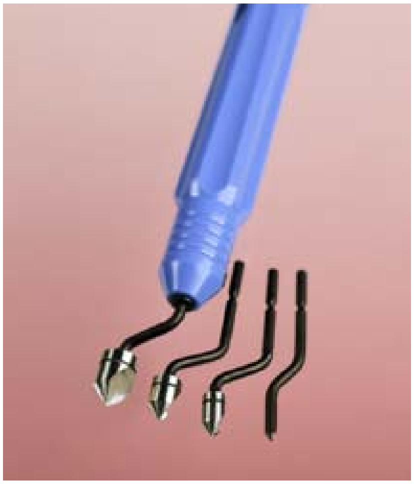

Figure 3-17. *You spin this countersink tool like a hand crank to add just the right amount of bevel to a hole, so that it will accommodate a flat-head screw.*

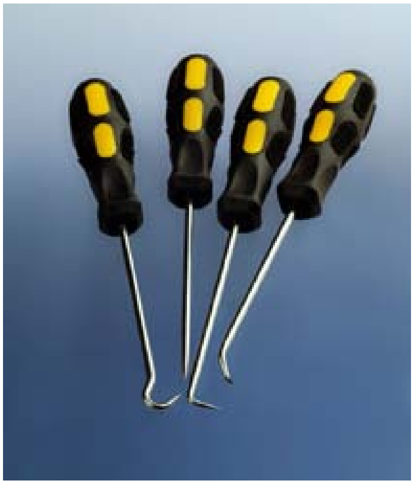

Figure 3-18. *This pick-and-hook set is useful in many unexpected ways.*

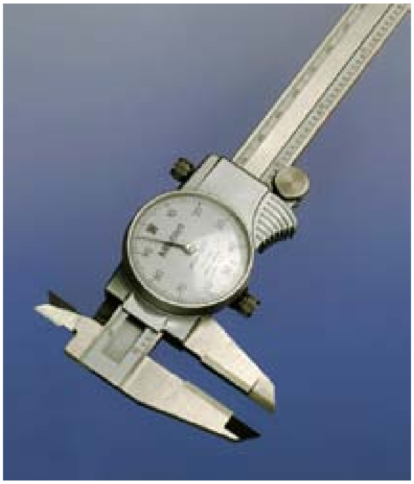

Figure 3-19. *Calipers can be digital (which automatically convert from millimeters to 1/64 inch to 1/1,000 inch), or analog like these (so you never need to worry about a dead battery).*

Supplies

Solder

This is the stuff that you will melt to join components together on a permanent (we hope) basis. You need some very thin solder, size 0.022 inches, for very small components, and thicker solder, 0.05 inches, for heavier items. Avoid buying solder that is intended for plumbers, or for craft purposes such as creating jewelry. A range of solder thicknesses is shown in Figure 3-20. You want to make sure to get lead-free solder.

Electronics solder has a nonacidic rosin core that is appropriate for electronic components. Rolls of solder are available from all hobby-electronics sources including All Electronics, RadioShack, and Jameco, or search for "electronic solder" on Amazon.

Figure 3-20. *Spools of solder in various thicknesses.*

Wire

You'll need some stranded wire to make flexible external connections with the circuit that you'll be building. Look for 22-gauge stranded hookup wire, in red, black, and green, 10 feet (minimum) of each.

If you want to install the intrusion alarm after completing that project in Experiment 15, you'll need white-insulated two-conductor wire of the type sold for doorbells or furnace controls. This is available by the foot from Lowe's, Home Depot, Ace Hardware, and similar stores. You'll decide how much to buy after you measure the distances between the magnetic sensor switches that you decide to install.

Heat-shrink tube

For use in conjunction with your heat gun, described previously. You'll need a range of sizes in any colors of your choice. See Figure 3-21. Check RadioShack part 278-1627, other electronics suppliers, or your local hardware store. Prices will vary widely. You can buy the cheapest.

Copper alligator clips

These absorb heat when you are soldering delicate components. The Mueller BU-30C is a full-size solid copper alligator clip for maximum heat absorption. RadioShack sells smaller clips (part number 270-373, shown in Figure 3-22) that are suitable for tiny components.

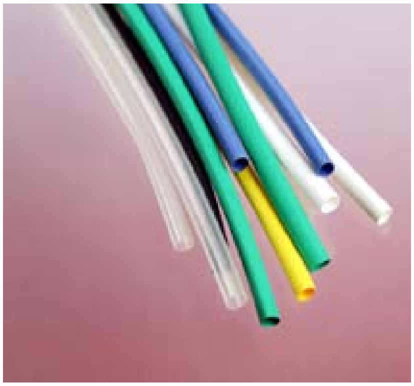

Figure 3-21. *Slide heat-shrink tubing over a bare joint and apply heat from a heat gun to make a tight insulating seal around the joint.*

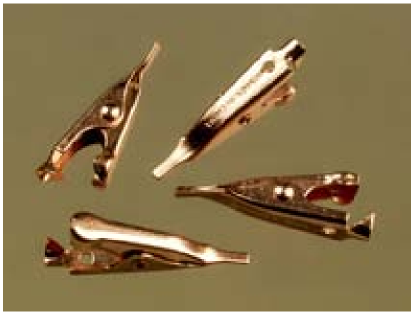

Figure 3-22. *These small clips absorb heat to protect components when you're soldering them.*

Perforated board

When you're ready to move your circuit from a breadboard to a more permanent location, you'll want to solder it to a piece of perforated board, often known as "prototyping board" but also called "perfboard."

You need the type that has copper strips etched onto the back, in exactly the same "breadboard layout" as the conductors hidden inside a breadboard, so that you can retain the same layout of your components when you solder them into place. Examples are RadioShack part 276-150 (shown in Figure 3-23) for small projects and part 276-170 (in Figure 3-24) for larger projects, such as Experiment 15.

For very small projects in which you will connect components using their wires alone, you need perfboard that isn't etched with copper strips connecting the holes. I like the Twin Industries 7100 range (available from Mouser.com) or Vectorboard from Newark Electronics, shown in Figure 3-25. You use a saw to cut out as small a piece as you need. Cheaper options are RadioShack part 276-147 (shown in Figure 3-26), or PC-1 from All Electronics. These have little copper circles around each hole that are not necessary for our purposes, but not a problem, either.

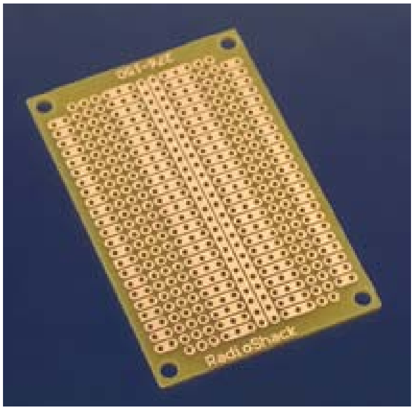

Figure 3-23. *This perforated board has a pattern of copper traces similar to the pattern inside a breadboard, so that you can lay out the components with minimal risk of wiring errors, when you're ready to create a permanently soldered version of your project.*

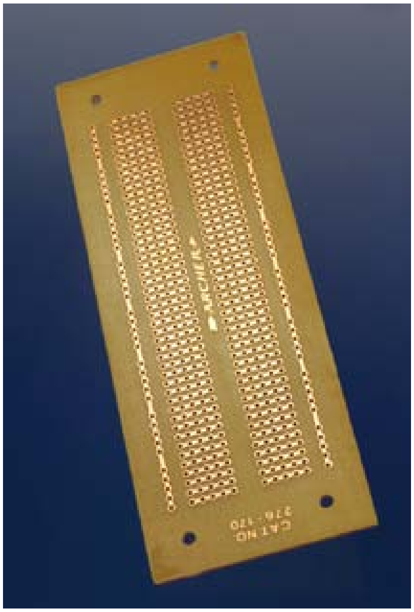

Figure 3-24. *A larger example of perforated board with breadboard geometry.*

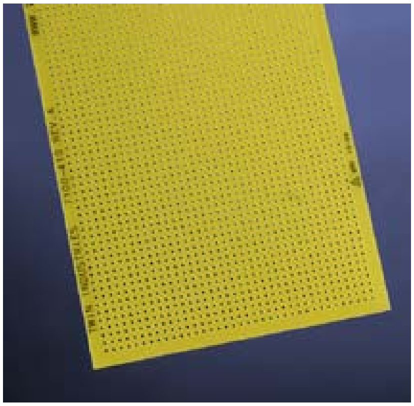

Figure 3-25. *Plain perforated board (with no copper traces) can be used for mounting components when you want to do point-to-point wiring.*

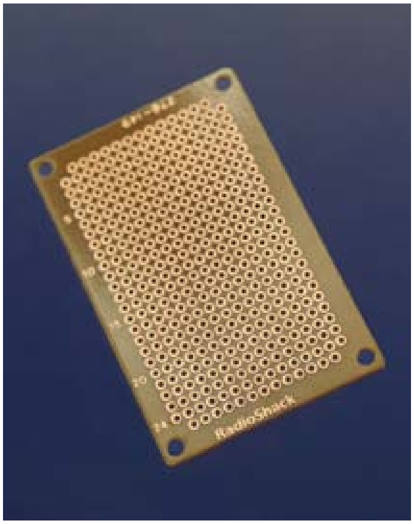

Figure 3-26. *A small piece of perforated board with individual copper solder pads to assist you in mounting components.*

Plywood

When you use a soldering iron, hot drops of solder tend to fall onto your table or workbench. The solder solidifies almost instantly, can be difficult to remove, and will leave a scar. Consider using a 2-foot square of half-inch plywood to provide disposable protection. You can buy it precut at Home Depot or Lowe's.

Machine screws

To mount components behind a panel, you need small machine screws (or "bolts"). They look nice if they have flat heads that fit flush against the panel. I suggest stainless-steel machine screws, #4 size, in 1/2-, 5/8-, 3/4-, and 1-inch lengths, 100 of each, plus 400 washers and 400 #4 locknuts of the type that have nylon inserts, so that they won't work loose. Check McMaster-Carr for a large and reasonably priced selection.

Project boxes

A project box is just a small box (usually plastic) with a removable lid. You mount your switches, potentiometers, and LEDs in holes that you drill through the box, and you attach your circuit on a perforated board that goes inside the box. Search All Electronics for "project box" or RadioShack for "project enclosure."

You need a box measuring approximately 6 inches long, 3 inches wide, and 2 inches high, such as RadioShack part 270-1805. Anything similar will do. I suggest you buy a couple other sizes as well, as they will be useful in the future.

Components

Power plugs, sockets, and binding posts

After you finish a project and put it in a box, you'll need a convenient way to supply it with power. Buy yourself a pair of insulated binding posts, such as RadioShack part 274-661, shown in Figure 3-27. Also obtain a panel-mounted power jack, size N, such as RadioShack part 274-1583, and DC power plug, size N, such as RadioShack 274-1573. The plug-and-socket pair is pictured in Figure 3-28.

Finally, you will need interconnects that are sized to fit a perforated board that is drilled at intervals of 1/10 inch. Sometimes known as "single inline sockets and headers," but also known as "boardmount sockets and pin-strip headers," they come in strips of 36 or more, and you can snip off as many as you need. Examples are Mill-Max part numbers 800-10-064-10-001000 and 801-93-050-10-001000, or 3M part numbers 929974-01-36-RK and 929834-01-36-RK. You can buy them from the usual electronics suppliers. Figure 3-29 shows headers before and after being snapped into small sections. Make sure that the interconnects have a terminal spacing of 0.1 inch.

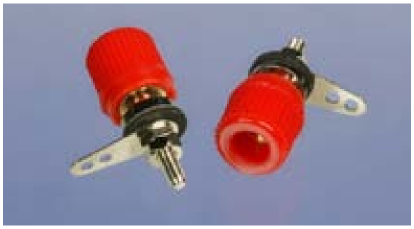

Figure 3-27. *These terminals, also known as binding posts, enable a solderless connection with wires that have stripped ends. Also available in black.*

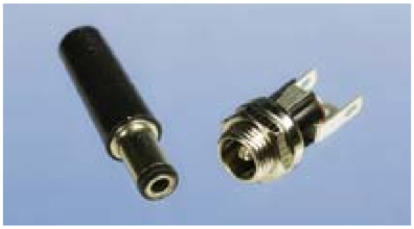

Figure 3-28. *The socket on the right can be mounted in a project box to receive power from the plug on the left.*

Battery

After you complete Experiment 15 at the end of this section of the book, if you want to use the project on a practical basis, you'll need a 12-volt battery. Search online for "12v battery" and you'll find many sealed, rechargeable lead-acid batteries that are designed for alarm systems, some measuring as small as 1×2×3-inch and costing under $10. You need a charger with it, which will probably cost you about $10.

Switches and relays

You will need the same DPDT relay and the same SPDT toggle switch that were mentioned in Chapter 2 shopping list.

For Experiment 15, you'll need magnetic switches that you can apply to doors or windows, such as the Directed model 8601, available from dozens of sources online.

Also you will need a DPDT pushbutton switch, ON-(ON) type, with solder terminals. Examples are model MPG206R04 by Tyco or model MB2061SS1W01-RO by NKK (with optional cap). Or search eBay for "DPDT pushbutton."

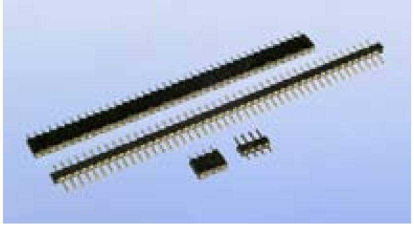

Figure 3-29. *Single inline sockets (top) and headers (middle) allow you to make very compact plug-and-socket connections to a PC board. They can be sawn, cut, or snapped into smaller sections (bottom). The terminals are 0.1 inch apart.*

Diodes

Buy at least half-a-dozen red 5 mm LEDs rated for approximately 2 volts, such as the Optek part number OVLFR3C7, Lumex part number SSL-LX-5093IT, or Avago part HLMP-D155. Buy half-a-dozen similar green LEDs at the same time.

In addition, you'll need a signal diode, type 1N4001 (any brand will do). Figure 3-30 shows an example, highly magnified. They're cheap, and likely to be useful in the future, so buy 10 of them.

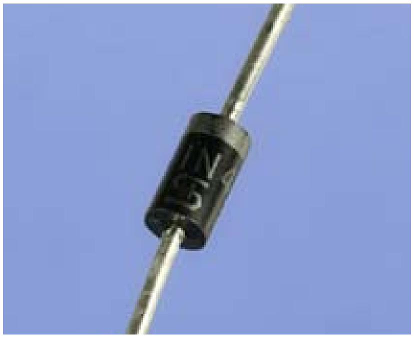

Figure 3-30. *This 1N4001 diode is about 1/4 inch long and can handle up to 50 volts.*

Soldering Irons Get Hot!

Please take these basic precautions:

Use a proper stand (such as the one incorporated in your helping hands) to hold your soldering iron. Don't leave it lying on a workbench.

If you have infants or pets, remember that they may play with, grab, or snag the wire to your soldering iron. They could injure themselves (or you).

Be careful never to rest the hot tip of the iron on the power cord that supplies electricity to the iron. It can melt the plastic in seconds and cause a dramatic short circuit.

If you drop a soldering iron, don't be a hero and try to catch it. Most likely you will grab the hot part, which hurts. (I speak from experience.) When you burn your hand, you will instinctively let go of the iron, so you may as well let it drop freely without the intermediate step of burning yourself while it's on its way to the floor. Naturally, you should pick it up quickly after it does hit the floor, but by then you will have gained the necessary time in which to make a sensible decision to grab it by the cool end.

Always bear in mind that others in your home are more at risk of hurting themselves on a soldering iron than you are, because they won't know that it's hot. Most soldering irons have no warning lights to tell you that they're plugged in. As a general rule, always assume that a soldering iron is hot, even if it's unplugged. It may retain sufficient heat to burn you for longer than you expect.

Loudspeaker

To complete the project in Experiment 15, you'll need a loudspeaker small enough to fit inside your project box but louder than the 1-inch speaker that you used previously. It should be 2 inches or 2.5 inches (50 to 60 mm) in diameter. If you can find a 100Ω speaker, it will give you more output, but an 8Ω speaker will be acceptable.

Experiment 12: Joining Two Wires Together

Your adventure into soldering begins with the prosaic task of joining one wire to another, but will lead quickly to creating a full electronic circuit on perforated board. So let's get started!

You will need:

- 30-watt or 40-watt soldering iron
- 15-watt pencil-type soldering iron
- Thin solder (0.022 inches or similar)
- Medium solder (0.05 inches or similar)
- Wire strippers and cutters
- "Helping hand" gadget to hold your work
- Shrink-wrap tubing, assorted
- Heat gun
- Something to protect your work area from drops of solder

Your First Solder Joint

We'll start with your general-duty soldering iron—the one rated for 30 or 40 watts. Plug it in, leave it safely in its holder, and find something else to do for five minutes. If you try to use a soldering iron without giving it time to get fully hot, you will *not* make good joints.

Strip the insulation from the ends of two pieces of 22-gauge solid wire and clamp them in your helping hand so that they cross each other and touch each other, as shown in Figure 3-31.

To make sure that the iron is ready, try to melt the end of a thin piece of solder on the tip of the iron. The solder should melt instantly. If it melts slowly, the iron isn't hot enough yet.

Now follow these steps (shown in Figures 3-32 through 3-36):

1. Make sure the tip of the soldering iron is clean (wipe it on the moistened sponge in the base of your helping hand if necessary), then touch it against the intersection of the wires steadily for three seconds to heat them. If you have hard tap water, use distilled water to wet the sponge to avoid a buildup of mineral deposits on the tip of your soldering iron.
2. While maintaining the iron in this position, feed a little solder onto the intersection of the wires, also touching the tip of the soldering iron. Thus, the two wires, the solder, and the tip of the iron should all come together at one point. The solder should spread over the wires within another two seconds.
3. Remove the iron and the solder. Blow on the joint to cool it. Within 10 seconds, it should be cool enough to touch.
4. Unclamp the wires and try to tug them apart. Tug hard! If they defeat your best attempts to separate them, the wires are electrically joined and should stay joined. If you didn't make a good joint, you will be able to separate the wires relatively easily, probably because you didn't apply enough heat or enough solder to connect them.

The reason I asked you to begin by using the higher-powered soldering iron is that it delivers more heat, which makes it easier to use.

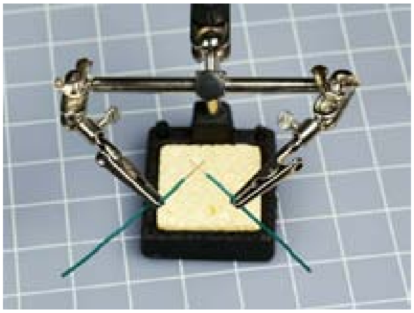

Figure 3-31. *A helping hand work aid is shown here holding two wires with their stripped ends touching. The magnifying glass has been hinged out of the way.*

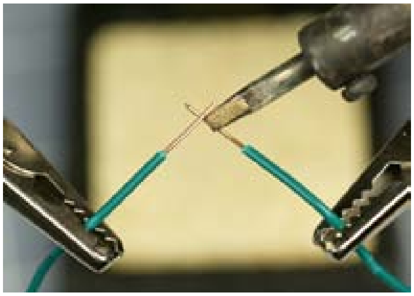

Figure 3-32

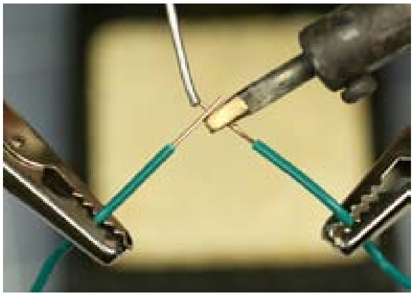

Figure 3-33

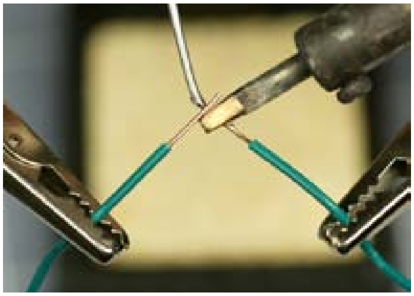

Figure 3-34

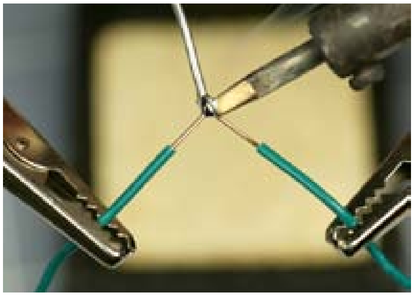

Figure 3-35. *This and the preceding three figures illustrate four steps to making a solder joint: apply heat to the wires, bring in the solder while maintaining the heat, wait for the solder to start to melt, and wait a moment longer for it to form a completely molten bead. The whole process should take between 4 and 6 seconds.*

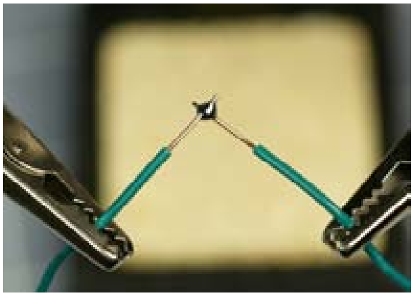

Figure 3-36. *The completed joint should be shiny, uniform, and rounded in shape.*

BACKGROUND

Soldering myths

Myth #1: Soldering is very difficult.

Millions of people have learned how to do it, and statistically, you are unlikely to be less coordinated than all of them. I have a lifelong problem with a tremor in my hands that makes it difficult for me to hold small things steadily. I also get impatient with repetitive detail work. If I can solder components, almost anyone should be able to.

Myth #2: Soldering involves poisonous chemicals.

Modern solder contains no lead. You should avoid inhaling the fumes for prolonged periods, but that also applies to everyday products such as bleach and paint. If soldering was a significant health hazard, we should have seen a high death rate among electronics hobbyists decades ago.

Myth #3: Soldering is hazardous.

A soldering iron is less hazardous than the kind of iron that you might use to iron a shirt, because it delivers less heat. In fact, in my experience, soldering is safer than most activities in a typical home or basement workshop. That doesn't mean you can be careless!

Soldering alternatives

As recently as the 1950s, connections inside electronic appliances such as radio sets were still being hand-soldered by workers on production lines. But the growth of telephone exchanges created a need for a faster way to make large numbers of rapid, reliable point-to-point wiring connections, and "wire wrap" became a viable alternative.

In a wire-wrapped electronics project, components are mounted on a circuit board that has long, gold-plated, sharp-cornered square pins sticking out of the rear. Special silver-plated wire is used, with an inch of insulation stripped from its ends. A manual or power-driven wire-wrap tool twirls the end of a wire around one of the pins, applying sufficient tension to "cold-weld" the soft silver plating of the wire to the pin. The wrapping process exerts sufficient pressure to make a very reliable joint, especially as 7 to 9 turns of wire are applied, each turn touching all four corners of the pin.

During the 1970s and 1980s, this system was adopted by hobbyists who built their own home computers. A wire-wrapped circuit board from a hand-built computer is shown in Figure 3-37. The technique was used by NASA to wire the computer in the Apollo spacecraft that went to the moon, but today, wire-wrapping has few commercial applications.

The widespread industrial use of "through-hole" components, such as the chips on early desktop computers, encouraged development of wave soldering, in which a wave or waterfall of molten solder is applied to the underside of a preheated circuit board where chips have been inserted. A masking technique prevents the solder from sticking where it isn't wanted.

Today, surface-mount components (which are significantly smaller than their through-hole counterparts) are glued to a circuit board with a solder paste, and the entire assembly is then heated, melting the paste to create a permanent connection.

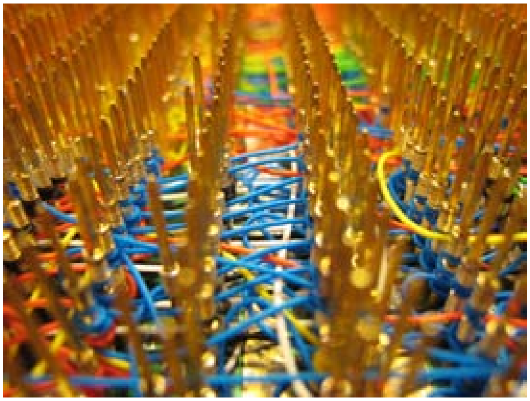

Figure 3-37. *This picture shows some of the wire-wrapping in Steve Chamberlin's custom-built, retro 8-bit CPU and computer. "Back in the day," connecting such a network of wires with solder joints would have been unduly time-consuming and prone to faults. Photo credit: Steve Chamberlin.*

TOOLS

Eight most common soldering errors

1. Not enough heat.

   The joint looks OK, but because you didn't apply quite enough heat, the solder didn't melt sufficiently to realign its internal molecular structure. It remained granular instead of becoming a solid, uniform blob, and you end up with a "dry joint," also known as a "cold joint," which will come apart when you pull the wires away from each other. Reheat the joint thoroughly and apply new solder.

   A leading cause of underheated solder is the temptation to use the soldering iron to carry solder to the joint. This results in the cold wires reducing the temperature of the solder. What you should do is touch the soldering iron to heat the wires first, and then apply the solder. This way, the wires are hot and help to melt the solder, which wants to stick to them.

   Because this is such a universal problem, I'll repeat myself: *Never melt solder on the tip of the iron and then use it to carry the solder to the joint.*

   You don't want to put hot solder on cold wires. You want to put cold solder on hot wires.

2. Too much heat.

   This may not hurt the joint, but can damage everything around it. Vinyl insulation will melt, exposing the wire and raising the risk of short circuits. You can easily damage semiconductors, and may even melt the internal plastic components of switches and connectors.

   Damaged components must be desoldered and replaced, which will take time and tends to be a big hassle (see "Tools: Desoldering" on page 109 for advice).

3. Not enough solder.

   A thin connection between two conductors may not be strong enough. When joining two wires, always check the underside of the joint to see whether the solder penetrated completely.

4. Moving the joint before the solder solidifies.

   You may create a fracture that you won't necessarily see. It may not stop your circuit from working, but at some point in the future, as a result of vibration or thermal stresses, the fracture can separate just enough to break electrical contact. Tracking it down will then be a chore. If you clamp components before you join them, or use perforated board to hold the components steady, you can avoid this problem.

5. Dirt or grease.

   Electrical solder contains rosin that cleans the metal that you're working with, but contaminants can still prevent solder from sticking. If any component looks dirty, clean it with fine sandpaper before joining it.

6. Carbon on the tip of your soldering iron.

   The iron gradually accumulates flecks of black carbon during use, and they can act as a barrier to heat transfer. Wipe the tip of the iron on the little sponge mounted in the base of your soldering iron stand or your helping hand.

7. Inappropriate materials.

   Electronic solder is designed for electronic components. It will not work with aluminum, stainless steel, or various other metals. You may be able to make it stick to chrome-plated items, but only with difficulty.

8. Failure to test the joint.

   Don't just assume that it's OK. Always test it, by applying manual force if you can (see Figures 3-38 and 3-39 for the ideal protocol) or, if you can't get a grip on the joint, slip a screwdriver blade under it and flex it just a little, or use small pliers to try to pull it apart. Don't be concerned about ruining your work. If your joint doesn't survive rough treatment, it wasn't a good joint.

Of the eight errors, dry/cold joints are by far the worst, because they are easy to make and can look OK.

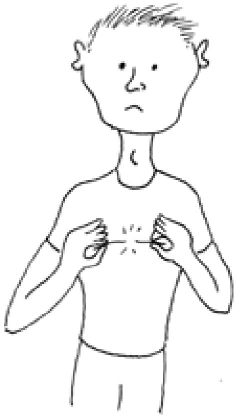

Figure 3-38. *Test result of a bad solder joint.*

Figure 3-39. *Test result of a good solder joint.*

Your Second Solder Joint

Time now to try your pencil-style soldering iron. Once again, you must leave it plugged in for a good five minutes to make sure it's hot enough. In the meantime, don't forget to unplug your other soldering iron, and put it somewhere safe while it cools.

This time I'd like you to align the wires parallel with each other. Joining them this way is a little more difficult than joining them when they cross each other, but it's a necessary skill. Otherwise, you won't be able to slide heat-shrink tubing over the finished joint to insulate it.

Figures 3-40 through 3-44 show a successful joint of this type. The two wires do not have to make perfect contact with each other; the solder will fill any small gaps. But the wires must be hot enough for the solder to flow, and this can take an extra few seconds when you use the low-wattage pencil-style iron.

Be sure to feed the solder in as shown in the pictures. Remember: don't try to carry the solder to the joint on the tip of the iron. Heat the wires first, and then touch the solder to the wires and the tip of the iron, while keeping it in contact with the wires. Wait until the solder liquifies, and you will see it running eagerly into the joint. If this doesn't happen, be more patient and apply the heat for a little longer.

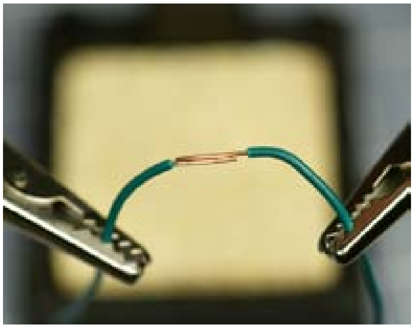

Figure 3-40

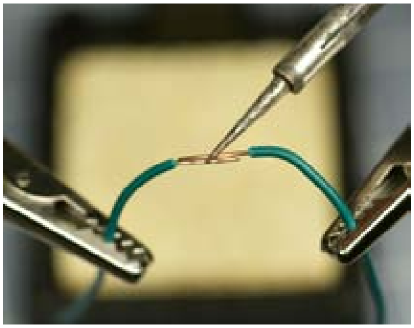

Figure 3-41

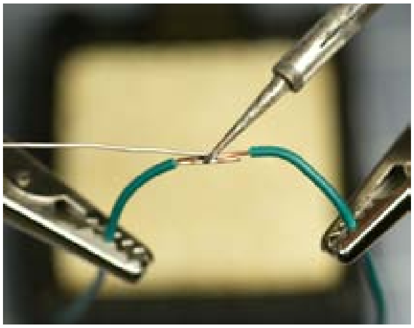

Figure 3-42

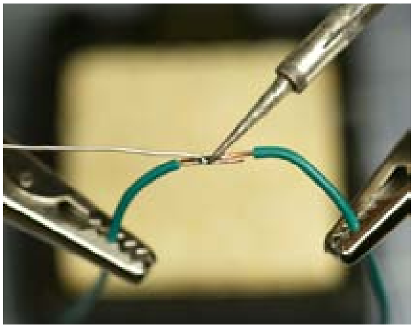

Figure 3-43. *This and the preceding three figures show how joining two wires that are parallel is more difficult, and the low-wattage, pencil-type soldering iron will require longer to heat them sufficiently for a good joint. Thinner solder can be used.*

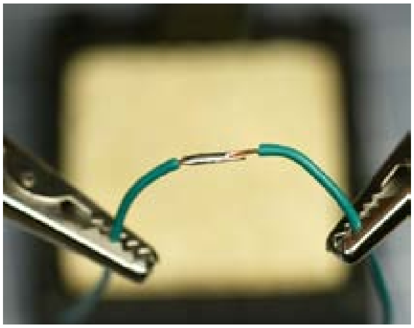

Figure 3-44. *The finished joint has enough solder for strength, but not so much solder that it will prevent heat-shrink tubing from sliding over it.*

THEORY

Soldering theory

The better you understand the process of soldering, the easier it should be for you to make good solder joints.

The tip of the soldering iron is hot, and you want to transfer that heat into the joint that you are trying to make. In this situation, you can think of the heat as being like a fluid. The larger the connection is between the soldering iron and the joint, the greater the quantity of heat, per second, that can flow through it.

For this reason, you should adjust the angle of the soldering iron so that it makes the widest possible contact. If it touches the wires only at a tiny point, you'll limit the amount of heat flow. Figures 3-45 and 3-46 illustrate this concept. Once the solder starts to melt, it broadens the area of contact, which helps to transfer more heat, so the process accelerates naturally. Initiating it is the tricky part.

The other aspect of heat flow that you should consider is that it can suck heat away from the places where you want it, and deliver it to places where you don't want it. If you're trying to solder a very heavy piece of copper wire, the joint may never get hot enough to melt the solder, because the heavy wire conducts heat away from the joint. You may find that even a 40-watt iron isn't powerful enough to overcome this problem, and if you are doing heavy work, you may need a more powerful iron.

As a general rule, if you can't complete a solder joint in 10 seconds, you aren't applying enough heat.

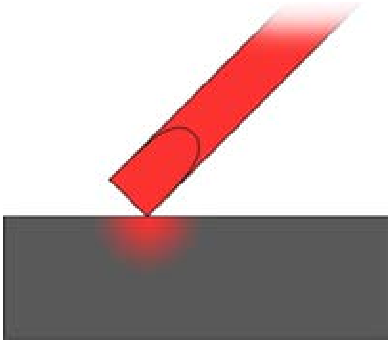

Figure 3-45. *With only a small surface area of contact between the iron and the working surface, an insufficient amount of heat is transferred.*

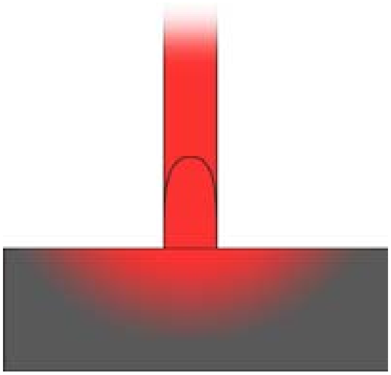

Figure 3-46. *A larger area of contact between the soldering iron and its target will greatly increase the heat transfer.*

TOOLS

Desoldering

Desoldering is much, much harder than soldering. Two simple tools are available:

- *Suction pump*. First, you apply the soldering iron to make the solder liquid. Then you use this simple gadget to try to suck up as much of the liquid as possible. Usually it won't remove enough metal to allow you to pull the joint apart, and you will have to try the next tool. Refer back to Figure 3-10.
- *Desoldering wick or braid*. Desoldering wick, also known as braid, is designed to soak up the solder from a joint, but again, it won't clean the joint entirely, and you will be in the awkward position of trying to use both hands to pull components apart while simultaneously applying heat to stop the solder from solidifying. Refer back to Figure 3-11.

I don't have much advice about desoldering. It's a frustrating experience (at least, I think so) and can damage components irrevocably.

Heat Guns Get Hot, Too!

Notice the chromed steel tube at the business end of your heat gun. Steel costs more than plastic, so the manufacturer must have put it there for a good reason—and the reason is that the air flowing through it becomes so hot that it would melt a plastic tube.

The metal tube stays hot enough to burn you for several minutes after you've used it. And, as in the case of soldering irons, other people (and pets) are vulnerable, because they won't necessarily know that the heat gun is hot. Most of all, make sure that no one in your home ever makes the mistake of using a heat gun as a hair dryer (Figure 3-47).

This tool is just a little more hazardous than it appears.

Adding Insulation

After you've succeeded in making a good inline solder connection between two wires, it's time for the easy part. Choose some heat-shrink tubing that is just big enough to slide over the joint with a little bit of room to spare.

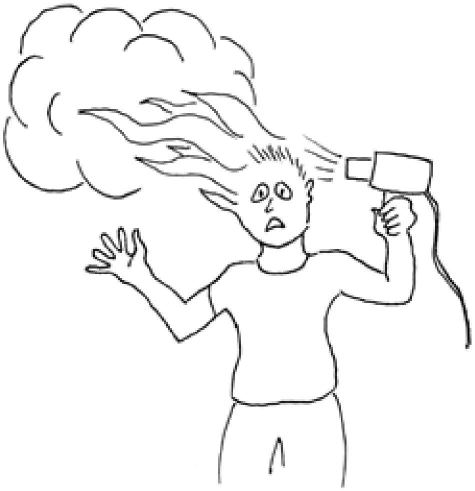

Figure 3-47. *Other members of your family should understand that although a heat gun looks like a hair dryer, appearances may be deceptive.*

Slide the tubing along until the joint is centered under it, hold it in front of your heat gun, and switch on the gun (keeping your fingers away from the blast of superheated air). Turn the wire so that you heat both sides. The tubing should shrink tight around the joint within half a minute. If you overheat the tubing, it may shrink so much that it splits, at which point you must remove it and start over. As soon as the tubing is tight around the wire, your job is done, and there's no point in making it any hotter. Figures 3-48 through 3-50 show the desired result. I used white tubing because it shows up well in photographs. Different colors of heat-shrink tubing all perform the same way.

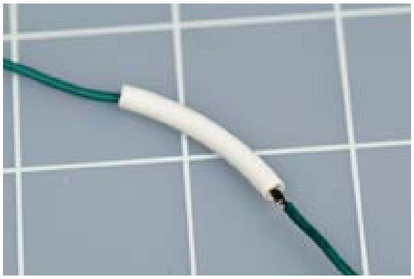

Figure 3-48. *Slip the tubing over your wire joint.*

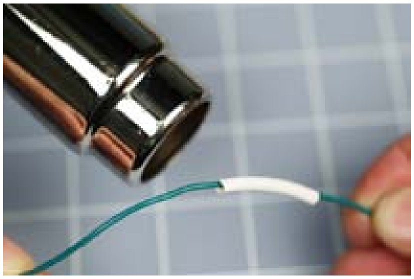

Figure 3-49. *Apply heat to the tubing.*

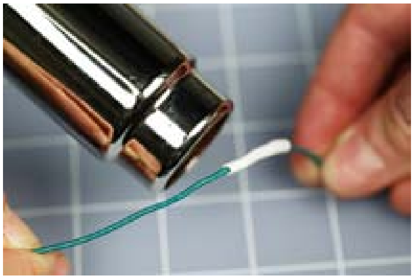

Figure 3-50. *Leave the heat on the tubing until it shrinks to firmly cover the joint.*

I suggest you next practice your soldering skills on a couple of practical projects. In the first one, you can add color-coded, solid-core wires to your AC adapter, and in the second one, you can shorten the power cord for a laptop power supply. You can use your larger soldering iron for both of these tasks, because neither of them involves any heat-sensitive components.

Modifying an AC Adapter

In the previous chapter, I mentioned the irritation of being unable to push the wires from your AC adapter into the holes of your breadboard. So, let's fix this right now:

1. Cut two pieces of solid-conductor 22-gauge wire—one of them red, the other black or blue. Each should be about 2 inches long. Strip a quarter-inch of insulation from both ends of each piece of wire.
2. Trim the wire from your AC adapter. You need to expose some fresh, clean copper to maximize your chance of getting the solder to stick.

   I suggest that you make one conductor longer than the other to minimize the chance of the bare ends touching and creating a short circuit. Use your meter, set to DC volts, if you have any doubt about which conductor is positive.

Solder the wires and add heat-shrink tubing as you did in the practice session. The result should look like Figure 3-51.

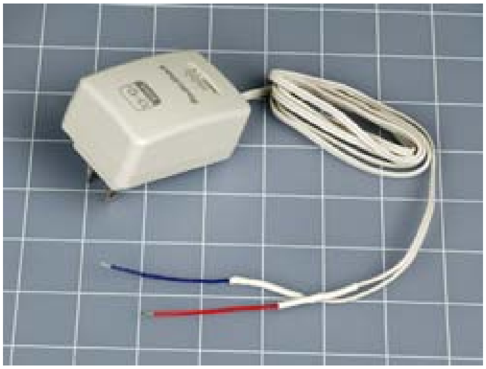

Figure 3-51. *Solid-core color-coded wires, soldered onto the wires from an AC adapter, provide a convenient way to feed power to a breadboard. Note that the wires are of differing lengths to reduce the risk of them touching each other.*

Choose the Right Tubing

If you use heat-shrink tubing on 110v AC cord, as is being done in this experiment, make sure you use tubing that's been rated for 110v use.

Shortening a Power Cord

When I travel, I like to minimize everything. It always annoys me that the power cord for the power supply of my laptop is 4 feet long. The thinner wire that connects the power supply to the computer is also 4 feet long, and I just don't need that much wire.

After searching exhaustively I couldn't find any laptop power cables shorter than 3 feet, so I decided to shorten one myself. If you feel no need to do this, you should try the following procedure on an old extension cord, just as an exercise. You do need to go through these steps to acquire some practice in soldering heavier, stranded wire and using heat-shrink tubing:

1. Use your wire cutters to chop the wire, and then a utility knife to split the two conductors, with one shorter than the other. When splicing a power cord or similar cable containing two or more conductors, it's good to avoid having the joints opposite each other. They fit more snugly if they are offset, and there's less risk of a short circuit if a joint fails.

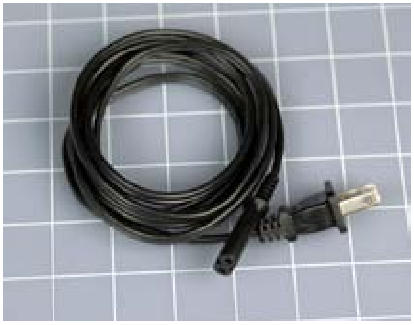

Figure 3-52

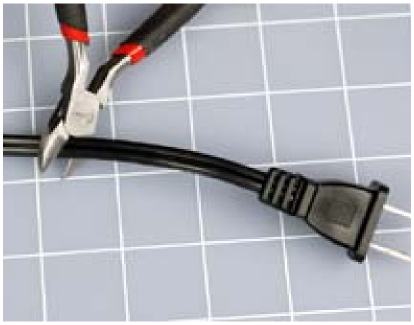

Figure 3-53

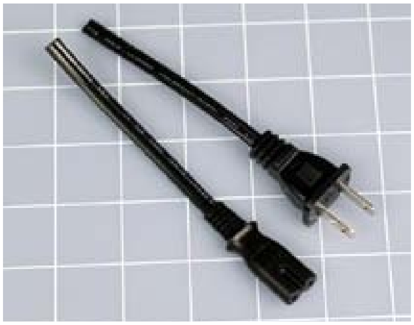

Figure 3-54

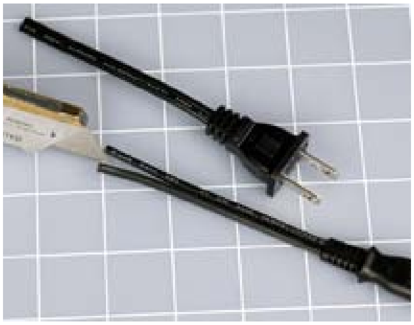

Figure 3-55

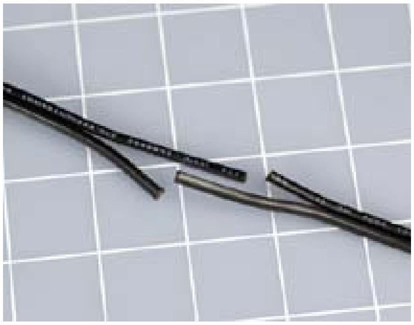

Figure 3-56

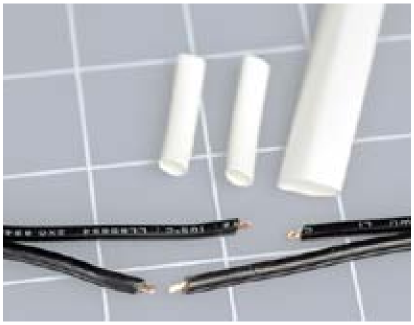

Figure 3-57

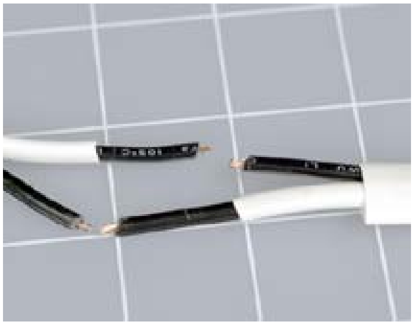

Figure 3-58. *Figures 3-52 through 3-58 illustrate the sequence of steps to prepare for making a shortened power cord for a laptop computer power supply.*

2. Strip off a minimal amount of insulation. One-eighth of an inch (3 mm) is sufficient. The automatic wire strippers that I mentioned in the shopping list in Chapter 1 are especially convenient, but regular wire strippers will do the job.

3. Cut two pieces of heat-shrink tubing, each 1 inch long, big enough to slide over the separate conductors in your cable. Cut a separate 2-inch piece of larger tubing that will slide over the entire joint when it's done. The steps described so far are illustrated in Figures 3-52 through 3-58.

4. Now for the most difficult part: activating your human memory. You have to remember to slide the tubing onto the wire *before* you make your solder joint, because the plugs on the ends of the wires will prevent you from adding any heat-shrink tubing later. If you're as impatient as I am, it's very difficult to remember to do this every time.

5. Use your helping hand to align the first joint. Push the two pieces of wire together so that the strands intermingle, and then squeeze them tight between finger and thumb, so that there are no little bits sticking out. A stray strand of wire can puncture heat-shrink tubing when the tubing is hot and soft and is shrinking around the joint.
6. The wire that you're joining is much heavier than the 22-gauge wire that you worked with previously, so it will suck up more heat, and you must touch the soldering iron to it for a longer time. Make sure that the solder flows all the way into the joint, and check the underside after the joint is cool. Most likely you'll find some bare copper strands there. The joint should become a nice solid, rounded, shiny blob. Keep the heat-shrink tubing as far away from the joint as possible while you're using the soldering iron, so that heat from the iron doesn't shrink the tubing prematurely, preventing you from sliding it over the joint later.

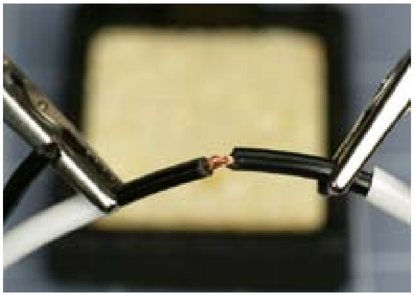

Figure 3-59

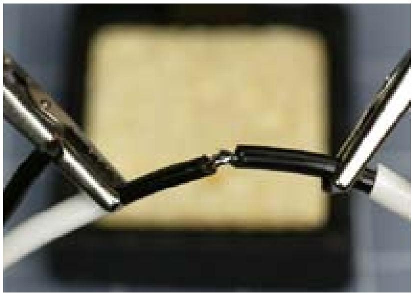

Figure 3-60

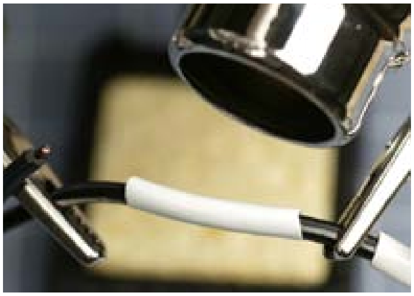

Figure 3-61

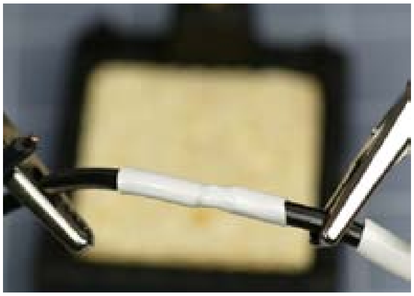

Figure 3-62

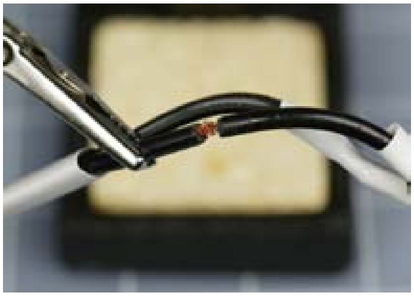

Figure 3-63

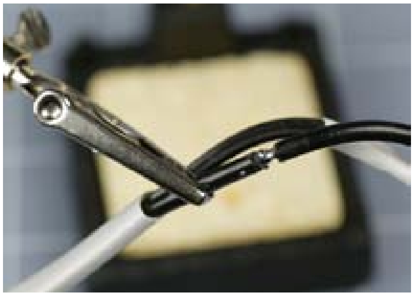

Figure 3-64

7. When the joint has cooled, slide the heat-shrink tubing over it, and apply the heat gun. Now repeat the process with the other conductor. Finally, slide the larger piece of tubing over the joint. You did remember to put the large tubing onto the wire at the beginning, didn't you?

Figures 3-59 through 3-65 show the steps all the way through to the end.

If you have completed the soldering exercises so far, you now have sufficient basic skills to solder your first electronic circuit. But first, I want you to verify the vulnerability of components to heat.

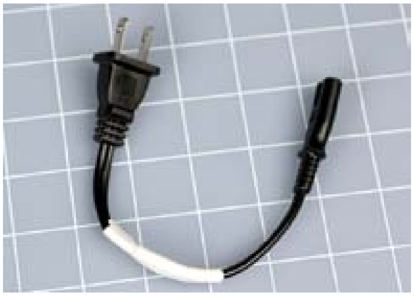

Figure 3-65. *Completion of the shortened power cord for a laptop power supply.*

Experiment 13: Broil an LED

In Chapter 1, you saw how an LED can be damaged if too much current flows through it. The electricity caused heat, which melted the LED. Unsurprisingly, you can just as easily melt it by applying too much heat to one of its leads with a soldering iron. The question is: how much heat is too much? Let's find out.

You will need:

- 30-watt or 40-watt soldering iron
- 15-watt pencil-type soldering iron
- A couple of LEDs (that are expendable)
- 680Ω resistor
- Wire cutters and sharp-nosed pliers
- "Helping hand" gadget to hold your work

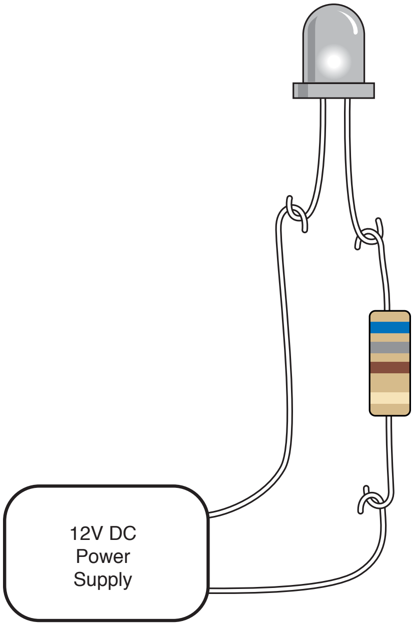

Figure 3-66. *By literally hooking together the leads from a resistor and a white-light LED, we minimize pathways for heat to escape during the subsequent test.*

I don't want you to use alligator clips to join the LED to a power supply, because the alligator clip will divert and absorb some of the heat from your soldering iron. Instead, please use some sharp-nosed pliers to bend each of the leads from an LED into little hooks, and do the same thing with the wires on a 680Ω load resistor. Finally bend the new wires on your AC adapter so that they, too, are tiny hooks. Now you can put the hooks together like links in a chain, as shown in Figure 3-66.

Grip the plastic body of the LED in your helping hand. Plastic is not a good thermal conductor, so the helping hand shouldn't siphon too much heat away from our target. The resistor can dangle from one of the leads on the LED, and the wire from the AC adapter can hang from that, a little farther down. Gravity should be sufficient to make this work. Set your AC adapter to deliver 12 volts as before, plug it in, and your LED should be shining brightly. I used a white LED in this experiment, because it's easier to photograph.

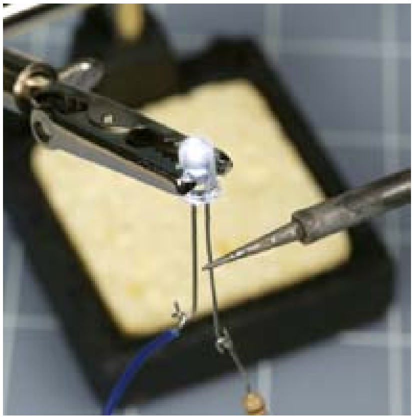

Figure 3-67. *Applying heat with a 15-watt soldering iron. A typical LED should withstand this treatment for two or three minutes, but if you substitute a 30-watt soldering iron, the LED is likely to burn out in under 15 seconds.*

Make sure your two soldering irons are really hot. They should have been plugged in for at least five minutes. Now take the pencil-style iron and hold its tip firmly against one of the leads on your glowing LED, while you check the time with a watch. Figure 3-67 shows the setup.

I'm betting that you can sustain this contact for a full three minutes without burning out the LED. This is why you use a 15-watt soldering iron for delicate electronics work—it doesn't endanger the components.

Allow your LED wire to cool, and then apply your more powerful soldering iron to the same piece of wire as before. Again, make sure it is completely hot, and I think you'll find that the LED will go dark after as little as 10 seconds (note, some LEDs can survive higher temperatures than others). This is why you *don't* use a 30-watt soldering iron for delicate electronics work.

The large iron doesn't necessarily reach a higher temperature than the small one. It just has a larger heat capacity. In other words, a greater quantity of heat can flow out of it, at a faster rate.

Throw away your burned-out LED. Substitute a new one, connected as before, but add a full-size copper alligator clip to one of the leads up near the body of the LED, as shown in Figure 3-68. Press the tip of your 30-watt or 40-watt soldering iron against the lead just *below* the alligator clip. This time, you should be able to hold the powerful soldering iron in place for a full two minutes without burning out the LED.

Imagine the heat flowing out through the tip of your soldering iron, into the wire that leads to the LED—except that the heat meets the alligator clip along the way, as shown in Figure 3-69. The clip is like an empty vessel waiting to be filled. It offers much less resistance to heat than the remainder of the wire leading to the LED, so the heat prefers to flow into the copper clip, leaving the LED unharmed. At the end of your experiment, if you touch the clip, you'll find that it's hot, while the LED remains relatively cooler.

The alligator clip is known as a heat sink, and it should be made of copper, because copper is one of the best conductors of heat.

Because the 15-watt soldering iron failed to harm the LED, you may conclude that the 15-watt iron is completely safe, eliminating all need for a heat sink. Well, this may be true. The problem is, you don't really know whether some semiconductors may be more heat-sensitive than LEDs. Because the consequences of burning out a component are so exasperating, I suggest you should play it safe and use a heat sink in these circumstances:

- If you apply 15-watt iron extremely close to a semiconductor for 20 seconds or more.
- If you apply a 30-watt iron near resistors or capacitors for 10 seconds or more. (Never use it near semiconductors.)
- If you apply a 30-watt iron near anything meltable for 20 seconds or more. Meltable items include insulation on wires, plastic connectors, and plastic components inside switches.

Rules for Heat Sinking

1. Full-size copper alligator clips do work better.
2. Clamp the alligator clip as close as possible to the component and as far as possible from the joint. (You don't want to suck too much heat away from the joint.)
3. Make sure there is a metal-to-metal connection between the alligator clip and the wire to promote good heat transfer.

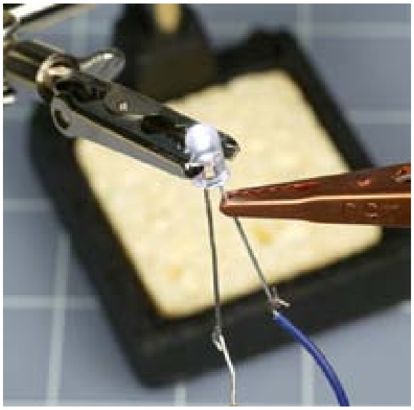

Figure 3-68. *When a copper alligator clip is used as a heat sink, you should be able to apply a 30-watt soldering iron (below the clip) without damaging the LED.*

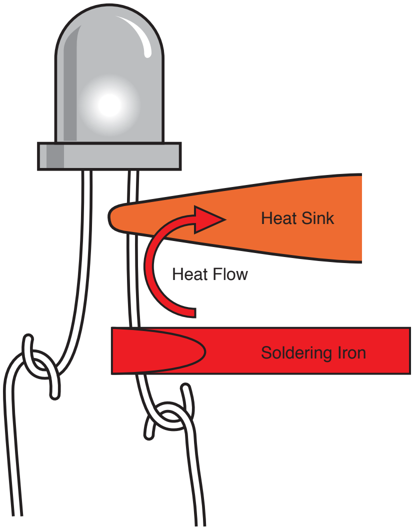

Figure 3-69. *The heat sink intercepts the heat, sucks it up, and protects the LED from damage.*

FUNDAMENTALS

All about perforated board

For the remainder of this book, you'll be using perforated board whenever you want to create permanent, soldered circuits. There are three ways to do this:

1. Point-to-point wiring. You use perforated board that has no connections behind the holes. Either the board has no copper traces on it at all, as in Figure 3-70, or you will find a little circular copper circle around each hole, as in Figure 3-71. These circles are not connected with each other and are used only to stabilize the components that you assemble.

   Point-to-point wiring allows you to place the components in a convenient, compact layout that can be very similar to a schematic. Under the board you bend the wires to link the components, and solder them together, adding extra lengths of wire if necessary. The advantage of this system is that it can be extremely compact. The disadvantage is that the layout can be confusing, leading to errors.

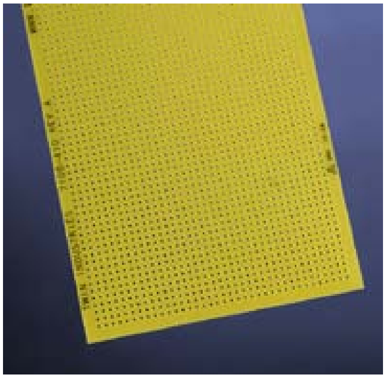

Figure 3-70

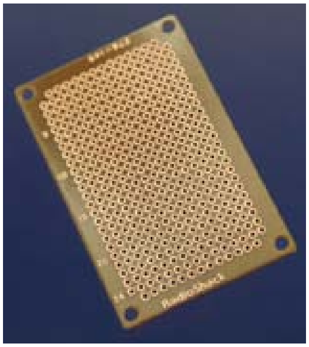

Figure 3-71. *Either this type of perforated board or the type in Figure 3-70 can be used for point-to-point wiring in Experiment 14.*

2. Breadboard-style wiring. Use perforated board that is printed with copper traces in exactly the same pattern as the conductors inside a breadboard. Once your circuit works on the breadboard, you move the components over to the perfboard one by one, maintaining their exact same positions relative to each other. You solder the "legs" of the components to the copper traces, which complete the circuit. Then you trim off the surplus wire. The advantage of this procedure is that it's quick, requires very little planning, and minimizes the possibility for errors. The disadvantage is that it tends to waste space. A cheap example is shown in Figure 3-72.
3. You can etch your own circuit board with customized copper traces that link your components in a point-to-point layout. This is the most professional way to complete a project, but it requires more time, trouble, and equipment than is practical in this book.

   Point-to-point wiring is like working with alligator clips, on a much smaller scale. The first soldered project will use this procedure.

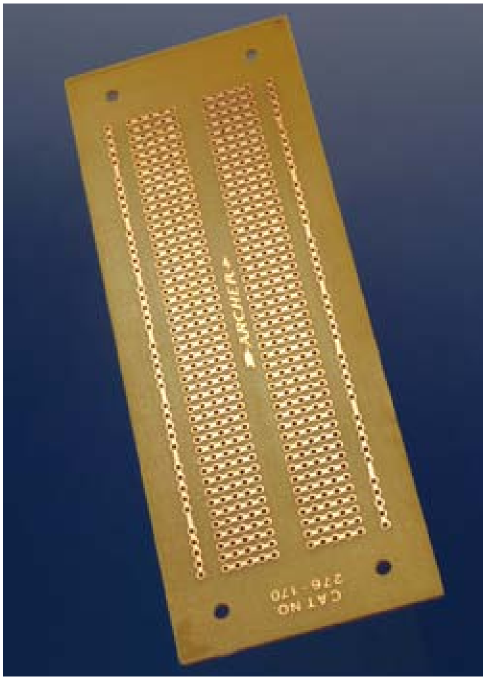

Figure 3-72. *Perforated board etched with copper in variants of a breadboard layout. This example is appropriate for Experiment 15.*

Experiment 14: A Pulsing Glow

You will need:

- Breadboard
- 15-watt pencil-type soldering iron
- Thin solder (0.022 inches or similar)
- Wire strippers and cutters
- Plain perforated board (no copper etching necessary)
- Small vise or clamp to hold your perforated board
- Resistors, various
- Capacitors, electrolytic, 100 µF and 220 µF, one of each
- Red LED, 5 mm, rated for 2 volts approximately
- 2N6027 programmable unijunction transistor

Your first circuit using a PUT was a slow-speed oscillator that made an LED flash about twice each second. The flashes looked very "electronic," by which I mean that the LED blinked on and off without a gradual transition between each state. I'm wondering if we can modify this circuit to make the LED pulse in a more gentle, interesting way, like the warning light on an Apple MacBook when it's in "sleep" mode. I'm thinking that something of this sort might be wearable as an ornament, if it's small enough and elegant enough.

I'm also thinking that this first soldering project will serve three purposes. It will test and refine your skill at joining wires together, will teach you point-to-point wiring with perfboard, and will give you some additional insight into the way that capacitors can be used to adjust timing.

Look back at the original schematic in Experiment 11, on page 82. Refresh your memory about the way it worked. The capacitor charges through a resistor until it has enough voltage to overcome the internal resistance in the PUT. Then the capacitor discharges through the PUT and flashes the LED.

If you drew a graph of the light coming out of the LED, it would be a thin, square-shaped pulse, as shown in Figure 3-73. How can we fill it out to make it more like the curve in Figure 3-74, so that the LED fades gently on and off, like a heartbeat?

One thing is obvious: the LED is going to be emitting a greater total amount of light in each cycle. Therefore, it's going to need more power. This means that C1, in Figure 3-75, must be a larger capacitor.

When we have a larger capacitor, it takes longer to charge. To keep the flashes reasonably frequent, we'll need a lower-value resistor for R1 to charge the capacitor quickly enough. In addition, reducing the values of R2 and R3 will program the PUT to allow a longer pulse.

Most important, I want to discharge the capacitor through a resistor to make the onset of the pulse gradual instead of sudden. Remember, when you have a resistor in series with a capacitor, the capacitor not only charges more slowly, but discharges more slowly.

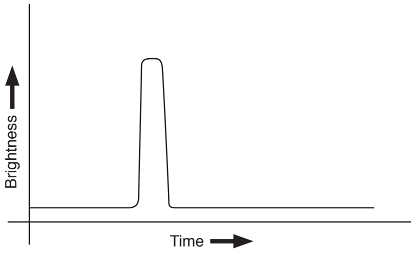

Figure 3-73

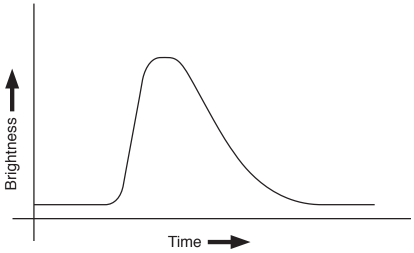

Figure 3-74. *The original PUT oscillator circuit in Experiment 11 made the LED emit sharp, short flashes. The upper graph shows what we might find if we measured light output over time. The second graph shows a gentler onset to each flash, followed by a slow fade-out. Capacitors can be used to create this effect.*

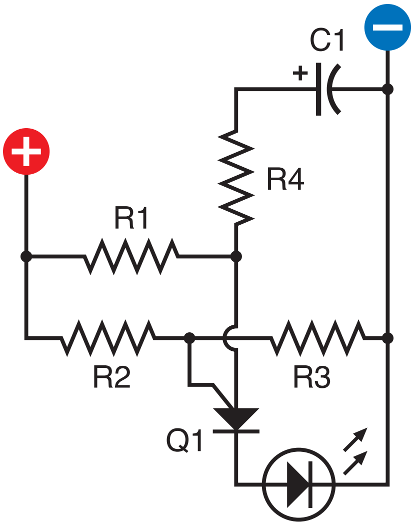

Figure 3-75. *The first step toward creating a gentler flashing effect is to use a larger capacitor for C1 and discharge it through a resistor, R4. Lower-value resistors are necessary to charge the capacitor rapidly enough.*

R1: 33K
R2: 1K
R3: 1K
R4: 1K
C1: 100 μF electrolytic
Q1: 2N6027

Figure 3-75 shows these features. Compare it with Figure 2-103 on page 85. R1 is now 33K instead of 470K. R2 and R3 are reduced to 1K. R4 also is 1K, to so that the capacitor takes longer to discharge through it. And C1 is now 100 μF instead of 2.2 μF.

Assemble this circuit on a breadboard, and compare the results when you include R4 or bypass it with a plain piece of jumper wire. It softens the pulse a bit, but we can work on it some more. On the output side of the PUT, we can add another capacitor. This will charge itself when the pulse comes out of the PUT, and then discharge itself gradually through another resistor, so that the light from the LED dies away more slowly.

Figure 3-76 shows the setup. C2 is large—220 μF—so it sucks up the pulse that comes out of the PUT, and then gradually releases it through 330Ω resistor R5 and the LED. You'll see that the LED behaves differently now, fading out inside of blinking off. But the resistances that I've added have dimmed the LED, and to brighten it, you should increase the power supply from 6 volts to 9 volts.

Remember that a capacitor imposes a smoothing effect only if one side of it is grounded to the negative side of the power supply. The presence of the negative charge on that side of the capacitor attracts the positive pulse to the other side.

I like the look of this heartbeat effect. I can imagine a piece of wearable electronic jewelry that pulses in this sensual way, very different from the hard-edged, sharp-on-and-off of a simple oscillator circuit. The only question is whether we can squeeze the components into a package that is small enough to wear.

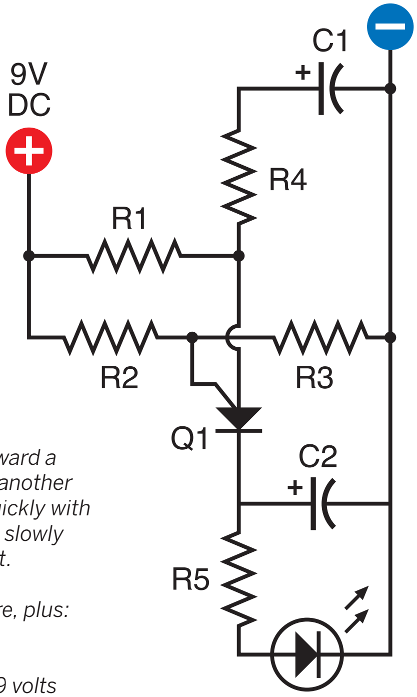

Figure 3-76. *The second step toward a gentler flashing effect is to add another capacitor, C2, which charges quickly with each pulse and then discharges slowly through R5 and the LED below it.*

Same components as before, plus:
R5: 330Ω
C2: 220 μF electrolytic
Power supply increased to 9 volts

Figure 3-77. *On a dark night in a rural area, the heartbeat flasher may be attractive in unexpected ways.*

Resizing the Circuit

The first step is to look at the physical components and imagine how to fit them into a small space. Figure 3-78 shows a 3D view of a compact arrangement. Check this carefully, tracing all the paths through the circuit, and you'll see that it's the same as the schematic. The trouble is that if we solder the components together like this, they won't have much strength. All the little wires can bend easily, and there's no easy way to mount the circuit in something or on something.

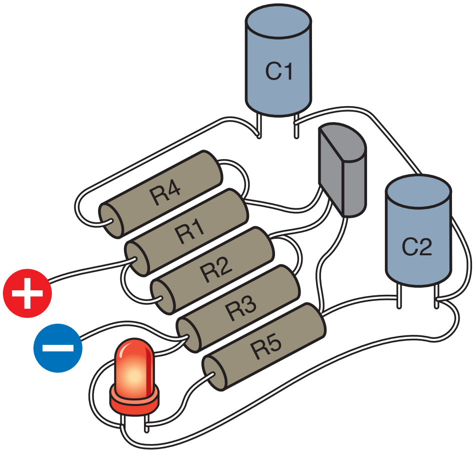

Figure 3-78. *This layout of components replicates their connections in the schematic diagram while squeezing them into a minimal amount of space.*

The answer is to put it on a substrate, which is one of those terms that people in the electronics field like to use, perhaps because it sounds more technical than "perfboard." But perforated board is what we need, and Figure 3-79 shows the components transferred onto a piece of board measuring just 1 inch by 0.8 inch.

The center version of this diagram uses dotted lines to show how the components will be connected with each other underneath the board. Mostly the leads that stick out from underneath the components will be long enough to make these connections.

Finally, the bottom version of the perfboard diagram shows the perfboard flipped left-to-right (notice the L and the R have been transposed to remind you, and I've used a darker color to indicate the underside of the board). Orange circles indicate where solder joints will be needed.

The LED should be unpluggable, because we may want to run it at some distance from the circuit. Likewise the power source should be unpluggable. Fortunately we can buy miniature connectors that fit right into the perforated board. You may have to go to large online retail suppliers such as Mouser.com for these. Some manufacturers call them "single inline sockets and headers," while others call them "boardmount sockets and pinstrip headers." Refer back to Figure 3-29 and check the shopping list for more details.

This is a very compact design that will require careful work with your pencil-style soldering iron. Because a piece of perforated board as small as this will tend to skitter around, I suggest that you apply your miniature vise to one end to anchor it with some weight while still allowing you to turn it easily.

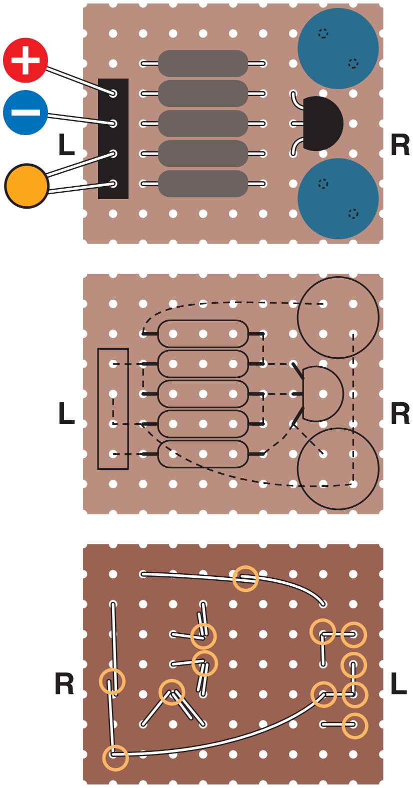

Figure 3-79. *Perforated board can be used to support the layout of components. Their leads are soldered together under the board to create the circuit. The middle diagram shows the wires under the board as dashed lines. The bottom diagram shows the board from underneath, flipped left to right. Orange circles indicate where solder joints will be necessary.*

When I'm working on this kind of project, I like to place it (with the vise attached) on a soft piece of polyurethane foam—the kind of slab that is normally used to make a chair cushion. The foam protects the components from damage when the board is upside-down, and again helps to prevent the work from sliding around unpredictably.

Step by Step

Here's the specific procedure for building this circuit:

1. Cut the small piece of perfboard out of a sheet that has no copper traces on it. You can cut the section using your miniature hobby saw, or you may be able to snap the board along its lines of holes, if you're careful. Alternatively, use a small ready-cut piece of perfboard with copper circles on it that are not connected to one another. You'll ignore the copper circles in this project. (In the next experiment, you'll deal with the additional challenge of making connections between components and copper traces on perforated board.)
2. Gather all the components and carefully insert them through holes in the board, counting the holes to make sure everything is in the right place. Flip the board over and bend the wires from the components to anchor them to the board and create connections as shown. If any of the wires isn't long enough, you'll have to supplement it with an extra piece of 22-gauge wire from your supply. You can remove all the insulation, as we'll be mounting the perfboard on a piece of insulating plastic.
3. Trim the wires approximately with your wire cutters.
4. Make the joints with your pencil soldering iron. Note that in this circuit, you are just joining wires to each other. The components are so close together that they'll prevent each other from wiggling around too much. If you are using board with copper pads (as I did), and some solder connects with them, that's OK—as long as it doesn't creep across to the neighboring component and create a short circuit.
5. Check each joint using a close-up magnifying glass, and wiggle it with pointed-nosed pliers. If there isn't enough solder for a really secure joint, reheat it and add more. If solder has created a connection that shouldn't be there, use a utility knife to make two parallel cuts in the solder, and scrape away the little section between them.

Generally, I insert three or four components, trim the wires approximately, solder them, trim their wires finally, then pause to check the joints and the placement. If I solder too many components in succession, there's a greater risk of missing a bad joint, and if I make an error in placing a component, undoing it will be much more problematic if I have already added a whole lot more components around it.

Figures 3-80 and 3-81 show the version of this project that I constructed, before I trimmed the board to the minimum size.

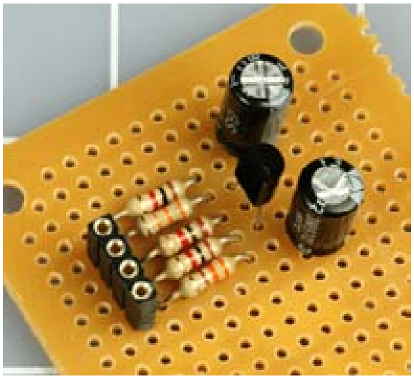

Figure 3-80. *Components mounted on a piece of perforated board.*

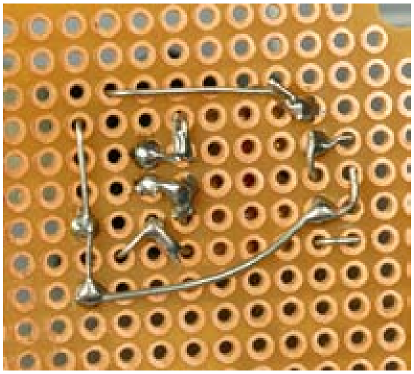

Figure 3-81. *The assembly seen from below. The copper circles around the holes are not necessary for this project. Some of them have picked up some solder, but this is irrelevant as long as no unintentional short circuits are created.*

Flying Wire Segments

The jaws of your wire cutters exert a powerful force that peaks and then is suddenly released when they cut through wire. This force can be translated into sudden motion of the snipped wire segment. Some wires are relatively soft, and don't pose a risk, but harder wires can fly in unpredictable directions at high speed, and may hit you in the eye. The leads of transistors are especially hazardous in this respect.

I think it's a good idea to wear safety glasses when trimming wires.

Finishing the Job

I always use bright illumination. This is not a luxury; it is a necessity. Buy a cheap desk lamp if you don't already have one. I use a daylight-spectrum fluorescent desk lamp, because it helps me identify the colored bands on resistors more reliably. Note that this type of fluorescent lamp emits quite a lot of ultraviolet light, which is not good for the lens in your eye. Avoid looking closely and directly at the tube in the lamp, and if you wear glasses, they will provide additional protection.

No matter how good your close-up vision is, you need to examine each joint with that close-up magnifier. You'll be surprised how imperfect some of them are. Hold the magnifier as close as possible to your eye, then pick up the thing that you want to examine and bring it closer until it comes into focus.

Finally, you should end up with a working circuit. You can insert the wires from your power supply into two of the tiny power sockets, and plug a red LED into the remaining two sockets. Remember that the two center sockets are negative, and the two outer sockets are positive, because it was easier to wire the circuit this way. You should color-code them to avoid mistakes.

So now you have a tiny circuit that pulses like a heartbeat. Or does it? If you have difficulty making it work, retrace every connection and compare it with the schematic. If you don't find an error, apply power to the circuit, attach the black lead from your meter to the negative side, and then go around the circuit with the red lead, checking the presence of voltage. Every part of this circuit should show at least some voltage while it's working. If you find a dead connection, you may have made a bad solder joint, or missed one entirely.

When you're done, now what? Well, now you can stop being an electronics hobbyist and become a crafts hobbyist. You can try to figure out a way to make this thing wearable.

First you have to consider the power supply. Because of the components that I used, we really need 9 volts to make this work well. How are you going to make this 9-volt circuit wearable, with a bulky 9-volt battery?

I can think of three answers:

1. You can put the battery in a pocket, and mount the flasher on the outside of the pocket, with a thin wire penetrating the fabric. Note that the tiny power connector on the perforated board will accept two 22-gauge wires if they are solid core, or if they are stranded (like the wires from a 9-volt battery connector) but have been thinly coated with solder.
2. You could mount the battery inside the crown of a baseball cap, with the flasher on the front.
3. You can put together three 3-volt button batteries in a stack, held in some kind of plastic clip. If you try this option, it may not be a good idea to try to solder wire to a battery. You will heat the liquid stuff inside the battery, which may not be good for it, and may not be good for you if the liquid starts boiling and the battery bursts open. Also, solder doesn't stick easily to the metallic finish on most battery terminals.

Most LEDs create a sharply defined beam of light, which you may want to diffuse to make it look nicer. One way to do this is to use a piece of transparent acrylic plastic, at least 1/4 inch thick, as shown in Figure 3-82. Sandpaper the front of the acrylic, ideally using an orbital sander that won't make an obvious pattern. Sanding will make the acrylic translucent rather than transparent.

Drill a hole slightly larger than the LED in the back of the acrylic. Don't drill all the way through the plastic. Remove all fragments and dust from the hole by blasting some compressed air into it, or by washing it if you don't have an air compressor. After the cavity is completely dry, get some transparent silicone caulking or mix some clear five-minute epoxy and put a drop in the bottom of the hole. Then insert the LED, pushing it in so that it forces the epoxy to ooze around it, making a tight seal. See Figure 3-82.

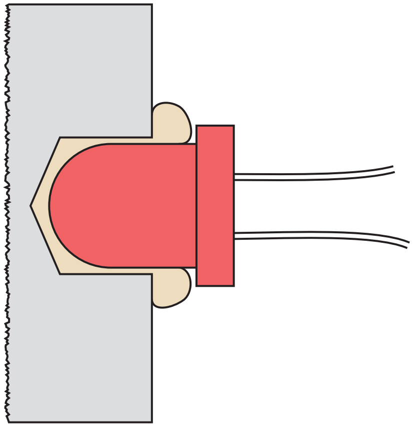

Figure 3-82. *This cross-sectional view shows a sheet of transparent acrylic in which a hole has been drilled part of the way from the back toward the front. Because a drill bit creates a hole with a conical shape at the bottom, and because the LED has rounded contours, transparent epoxy or silicone caulking can be injected into the hole before mounting the LED.*

Try illuminating the LED, and sand the acrylic some more if necessary. Finally, you can decide whether to mount the circuit on the back of the acrylic, or whether you want to run a wire to it elsewhere.

Because the LED will flash at about the speed of a human heart while the person is resting, it may look as if it's measuring your pulse, especially if you mount it on the center of your chest or in a strap around your wrist. If you enjoy hoaxing people, you can suggest that you're in such amazingly good shape, your pulse rate remains constant even when you're taking strenuous exercises.

To make a good-looking enclosure for the circuit, I can think of options ranging from embedding the whole thing in clear epoxy to finding a Victorian-style locket. I'll leave you to consider alternatives, because this is a book about electronics rather than handicrafts.

However, I will address one final issue: how long will this gadget continue flashing?

If you check the following section "Essentials: Battery life," you'll find that a regular alkaline 9-volt battery should keep the LED flashing for about 50 hours.

ESSENTIALS

Battery life

Any time you finish a circuit that you intend to run from a battery, you'll want to calculate the likely battery life. This is easily done, because manufacturers rate their batteries according to the "ampere hours" they can deliver. Keep the following in mind:

- The abbreviation for amp-hours is Ah, sometimes printed as AH. Milliampere-hours are abbreviated mAh.
- The rating of a battery in amp-hours is equal to the current, in amps, multiplied by the number of hours that the battery can deliver it.

Thus, in theory 1 amp-hour can mean 1 amp for 1 hour, or 0.1 amp for 10 hours, or 0.01 amp for 100 hours—and so on. In reality, it's not as simple as this, because the chemicals inside a battery become depleted more quickly when you draw a heavy current, especially if the battery gets hot. You have to stay within limits that are appropriate to the size of the battery.

For instance, if a small battery is rated for 0.5 amp-hours, you can't expect to draw 30 amperes from it for 1 minute. But you should be able to get 0.005 amps (i.e., 5 milliamps) for 100 hours without any trouble. Remember, though, that the voltage delivered by a battery will be greater than its rated voltage when the battery is fresh, and will diminish below its rated voltage while the battery is delivering power.

According to some test data that I trust (I think they are a little more realistic than the estimates supplied by battery manufacturers), here are some numbers for typical batteries:

- Typical 9 volt alkaline battery: 0.3 amp-hours, while delivering 100 mA.
- Typical AA size, 1.5-volt alkaline battery: 2.2 amp-hours, while delivering 100 mA.
- Rechargeable nickel-metal hydride battery: about twice the endurance of a comparably sized alkaline battery.
- Lithium battery: maybe three times the endurance of an alkaline battery.

BACKGROUND

Maddened by measurement

Throughout most of this book, I've mostly used measurements in inches, although sometimes I've digressed into the metric system, as when referring to "5-mm LEDs." This isn't inconsistency on my part; it reflects the conflicted state of the electronics industry, where you'll find inches and millimeters both in daily use, often in the very same data sheet.

The United States is the only major nation still using the old system of units that originated in England. (The other two holdouts are Liberia and Myanmar, according to the CIA's *World Factbook*.) Still, the United States has led many advances in electronics, especially the development of silicon chips, which have contacts spaced 1/10 inch apart. These standards became firmly established, and show no sign of disappearing.

To complicate matters further, even in the United States, you can encounter two incompatible systems for expressing fractions of an inch. Drill bits, for instance, are measured in multiples of 1/64 inch, while metal thicknesses may be measured in decimals such as 0.06 inch (which is approximately 1/16 inch).

The metric system is not necessarily more rational than the U.S. system. Originally, when the metric system was formally introduced in 1875, the meter was defined as being 1/10,000,000 of the distance between the North Pole and the equator, along a line passing through Paris—a quixotic, Francocentric conceit. Since then, the meter has been redefined three times, in a series of efforts to achieve greater accuracy in scientific applications.

As for the usefulness of a 10-based system, moving a decimal point is certainly simpler than doing calculations in 64ths of an inch, but the only reason we count in tens is because we happen to have evolved with that number of digits on our hands. A 12-based system would really be more convenient, as numbers would be evenly divisible by 2 and 3.

As we're stuck with the whimsical aspects of length measurement, I've created the charts in Figures 3-83 and 3-84 to assist you in going from one system to another. From these you will see that when you need to drill a hole for a 5 mm LED, a 3/16-inch drill bit is about right. (In fact, it results in a better, tighter fit than if you drill an actual 5 mm hole.)

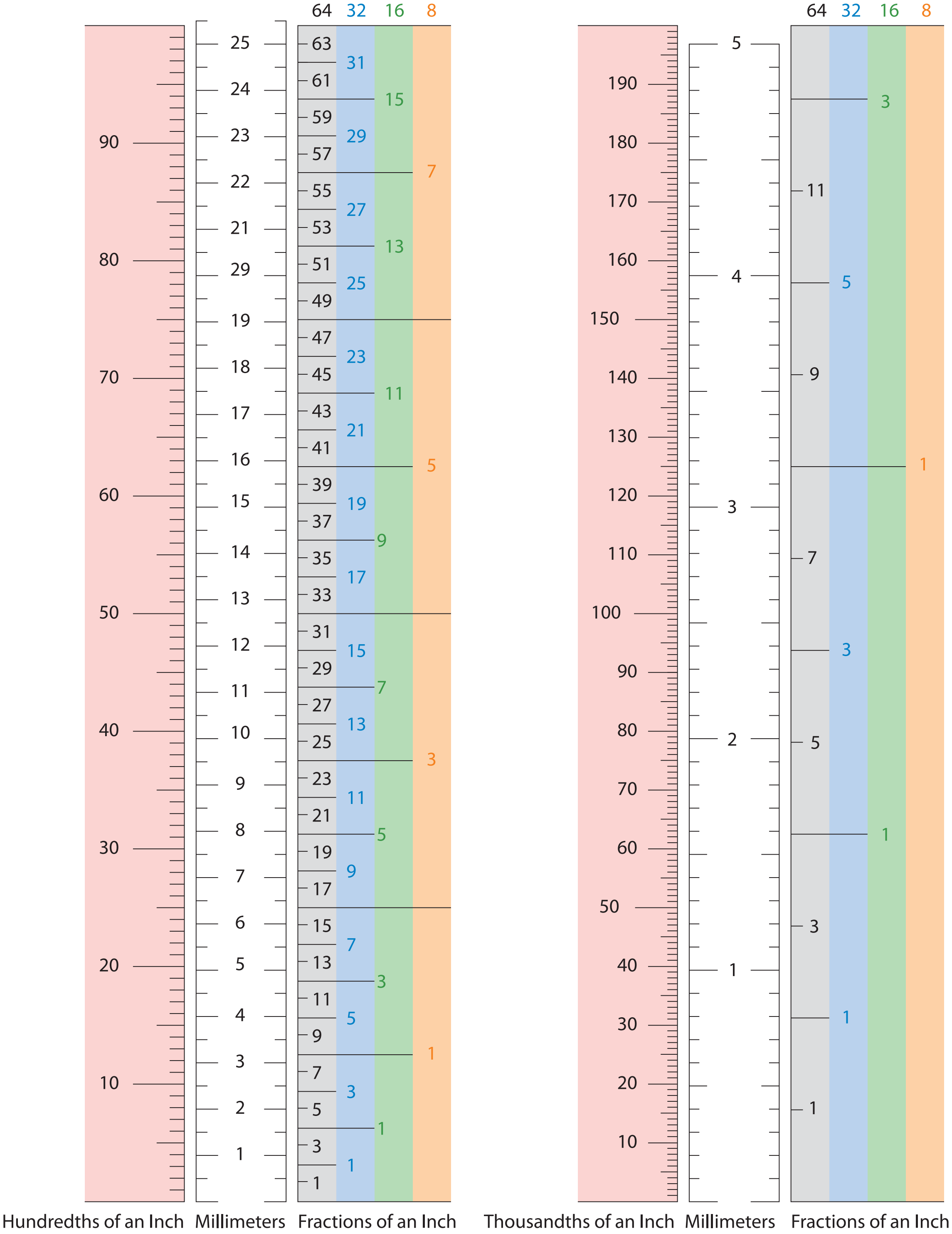

Figure 3-83. *Because units of measurement are not standardized in electronics, conversion is often necessary. The chart on the right is a 5x magnification of the bottom section of the chart on the left.*

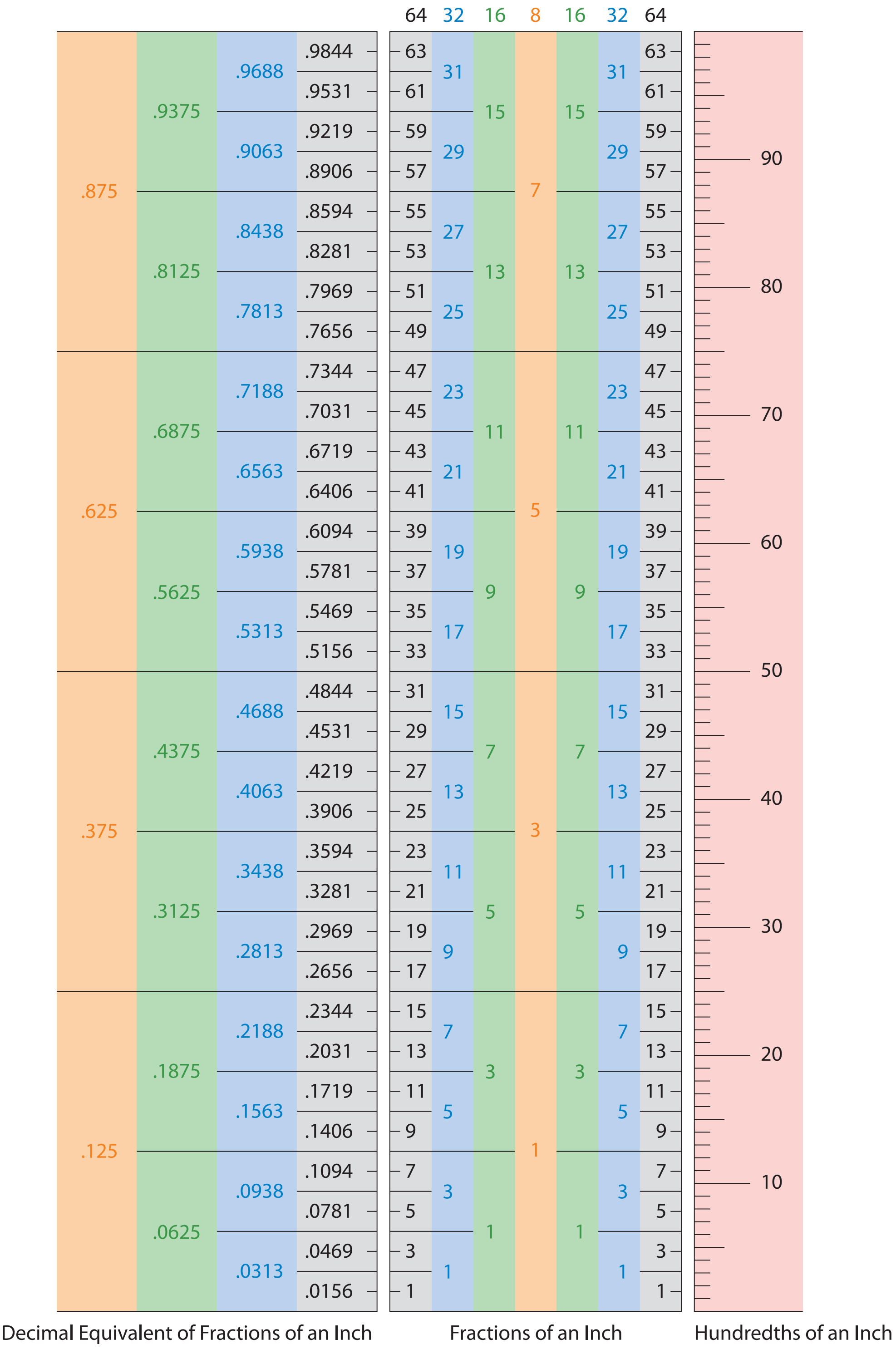

				64	32	16	8	16	32	64
.875	.9375	.9688	.9844	63	31	15	7	15	31	63
			.9531	61						61
		.9063	.9219	59	29				29	59
			.8906	57						57
	.8125	.8438	.8594	55	27	13		13	27	55
			.8281	53						53
		.7813	.7969	51	25				25	51
			.7656	49						49
.625	.6875	.7188	.7344	47	23	11	5	11	23	47
			.7031	45						45
		.6563	.6719	43	21				21	43
			.6406	41						41
	.5625	.5938	.6094	39	19	9		9	19	39
			.5781	37						37
		.5313	.5469	35	17				17	35
			.5156	33						33
.375	.4375	.4688	.4844	31	15	7	3	7	15	31
			.4531	29						29
		.4063	.4219	27	13				13	27
			.3906	25						25
	.3125	.3438	.3594	23	11	5		5	11	23
			.3281	21						21
		.2813	.2969	19	9				9	19
			.2656	17						17
.125	.1875	.2188	.2344	15	7	3	1	3	7	15
			.2031	13						13
		.1563	.1719	11	5				5	11
			.1406	9						9
	.0625	.0938	.1094	7	3	1		1	3	7
			.0781	5						5
		.0313	.0469	3	1				1	3
			.0156	1						1

90
80
70
60
50
40
30
20
10
Decimal Equivalent of Fractions of an Inch
Fractions of an Inch
Hundredths of an Inch

Figure 3-84. *This chart allows conversion between hundredths of an inch, conventional U. S. fractions of an inch, and fractions expressed in thousandths of an inch.*

Experiment 15: Intrusion Alarm Revisited

Time now to add some of the enhancements to the intrusion alarm that I discussed at the end of Experiment 11. I'm going to show you how the alarm can be triggered if you install various detectors on windows and doors in your home. I'll also show how the alarm can be wired so that it locks itself on and continues to make noise even after a door or window is reclosed.

This experiment will demonstrate the procedure for transferring a project from a breadboard to a piece of perforated board that has copper connections laid out identically to the ones inside the breadboard, as shown earlier in Figure 3-72. And you'll mount the finished circuit in a project box with switches and connectors on the front.

When all is said and done, you'll be ready for wholesale circuit building. The explanations in the rest of this book will get gradually briefer, and the pace will increase.

You will need:

- 15-watt pencil-type soldering iron
- Thin solder (0.022 inches or similar)
- Wire strippers and cutters
- Perforated board etched with copper in a breadboard layout
- Small vise or clamp to hold your perforated board
- The same components that you used in Experiment 11, plus:
    - 2N2222 NPN transistor. Quantity: 1.
    - DPDT relay. Quantity: 1.
    - SPDT toggle switch. Quantity: 1.
    - 1N4001 diode. Quantity: 1.
    - Red and green 5mm LEDs. Quantity: 1 each.
    - Project box, 6 × 3 × 2 inches.
    - Power jack, type N, and matching power socket, type N.
    - Binding posts.
    - Stranded 22-gauge wire, three different colors.
    - Magnetic sensor switches, sufficient for your home.
    - Alarm network wiring, sufficient for your home.

Magnetic Sensor Switches

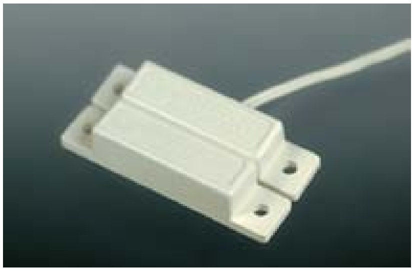

Figure 3-85. *In this simple alarm sensor switch, the lower module contains a magnet, which opens and closes a reed switch sealed into the upper module.*

A typical alarm sensor switch consists of two modules: the magnetic module and the switch module, as shown in Figures 3-85 and 3-86. The magnetic module contains a permanent magnet, and nothing else. The switch module contains a "reed switch," which makes or breaks a connection (like a contact inside a relay) under the influence of the magnet. When you bring the magnetic module close to the switch module, you may faintly hear the reed switch click as it flips from one state to the other.

Like all switches, reed switches can be normally open or normally closed. For this project, you want the kind of switch that is normally open, and closes when the magnetic module is close to it.

Attach the magnetic module to the moving part of a door or window, and attach the switch module to the window frame or door frame. When the window or door is closed, the magnetic module is almost touching the switch module. The magnet keeps the switch closed until the door or window is opened, at which point the switch opens.

The only question is: how do we use this component to trigger our alarm? As long as a small current flows through all our magnetic sensor switches, the alarm should be off, but if the flow of current stops, the alarm should switch on.

We could use a relay that is "always on" while the alarm is armed. When the circuit is interrupted, the relay relaxes and its other pair of contacts closes, which could power up the alarm noisemaker.

But I don't like this idea. Relays take significant power, and they can get hot. Most of them are not designed to be kept "always-on." I'd prefer to handle the task using a transistor.

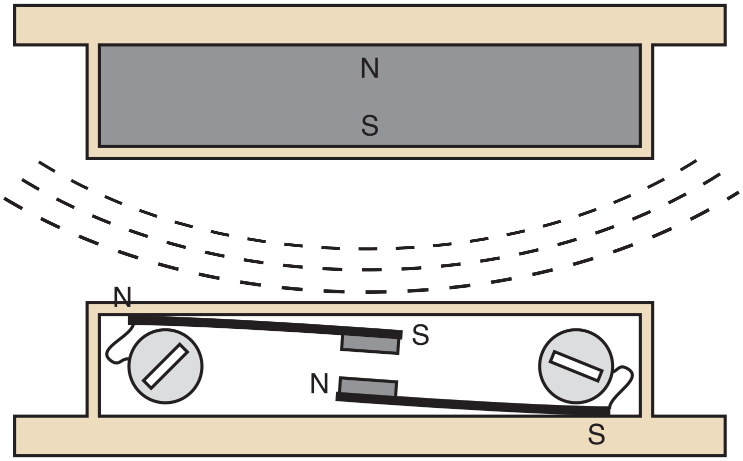

Figure 3-86. *This cutaway diagram shows a reed switch (bottom) and the magnet that activates it (top), inside an alarm sensor. The switch contains two flexible magnetized strips, the upper one with its south pole adjacent to an electrical contact, the lower one with its north pole adjacent to an electrical contact. When the south pole of the magnet approaches the switch, the magnetic force (shown as dashed lines) repels the south contact and attracts the north contact, causing them to snap together. Two screws on the outside of the casing are connected with the strips inside.*

A Break-to-Make Transistor Circuit

First, recall how an NPN transistor works. When the base is not sufficiently positive, the transistor blocks current between its collector and emitter, but when the base is relatively positive, the transistor passes current.

Take a look at the schematic in Figure 3-87, which is built around our old friend the 2N2222 NPN transistor. When the switch is closed, it connects the base of the transistor to the negative side of the power supply through a 1K resistor. At the same time, the base is connected with the positive side of the power supply through a 10K resistor. Because of the difference in resistances and the relatively high turn-on voltage for the LED, the base is forced below its turn-on threshold, and as a result, the transistor will not pass much current. The LED will glow dimly at best.

Now what happens when the switch is opened? The base of the transistor loses its negative power supply and has only its positive power supply. It becomes much more positive, above the turn-on threshold for the transistor, which tells the transistor to lower its resistance and pass more current. The LED now glows brightly. Thus, when the switch is turned off and breaks the connection, the LED is turned on.

This seems to be what we want. Imagine a whole series of switches instead of just one switch, as shown in Figure 3-88. The circuit will still work the same way, even if the switches are scattered all over your home, because the resistance in the wires connecting the switches will be trivial compared with the resistance of the 1K resistor.

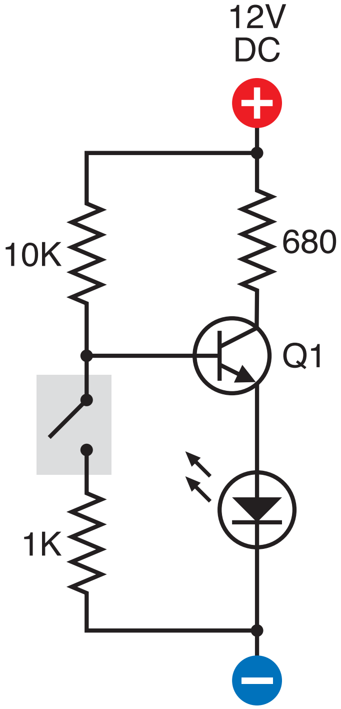

Figure 3-87. *In this demonstration circuit, when the switch is opened, it interrupts negative voltage to the base of the transistor, causing the transistor to lower its resistance, allowing current to reach the LED. Thus, when the switch is turned off, it turns on the LED.*

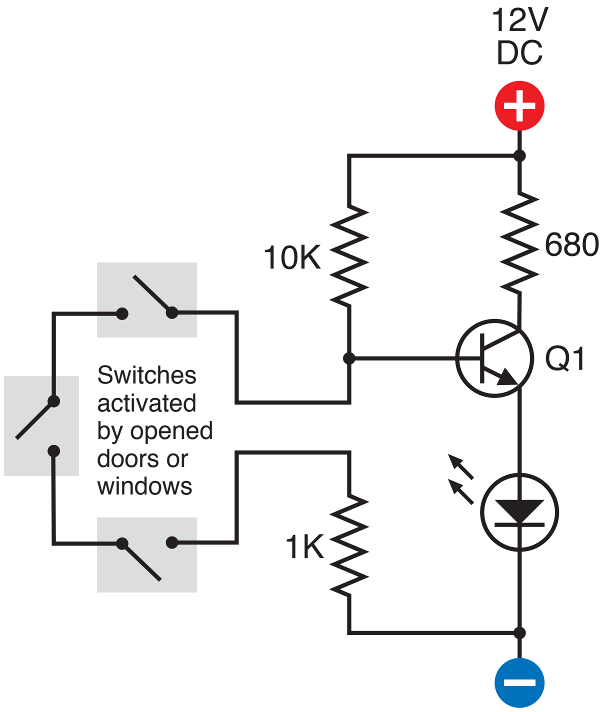

Figure 3-88. *A network of switches, wired in series, can be substituted for the single switch in Figure 3-87. Now any one switch will break continuity and trigger the transistor.*

I have shown the switches open, because that's the way the schematic for a switch is drawn, but imagine them all closed. The base of the transistor will now be supplied through the long piece of wire connecting all the closed switches, and the LED will stay dark. Now if just one switch is opened, or if anyone tampers with the wire linking them, the base of the transistor loses its connection to negative power, at which point the transistor conducts power and the LED lights up.

While all the switches remain closed, the circuit is drawing very little current—probably about 1.1 mA. So you could run it from a typical 12-volt alarm battery.

Now suppose we swap out the LED and put a relay in there instead, as shown in Figure 3-89. I don't mind using a relay in this location, because the relay will not be "always on." It will normally be off, and will draw power only when the alarm is triggered.

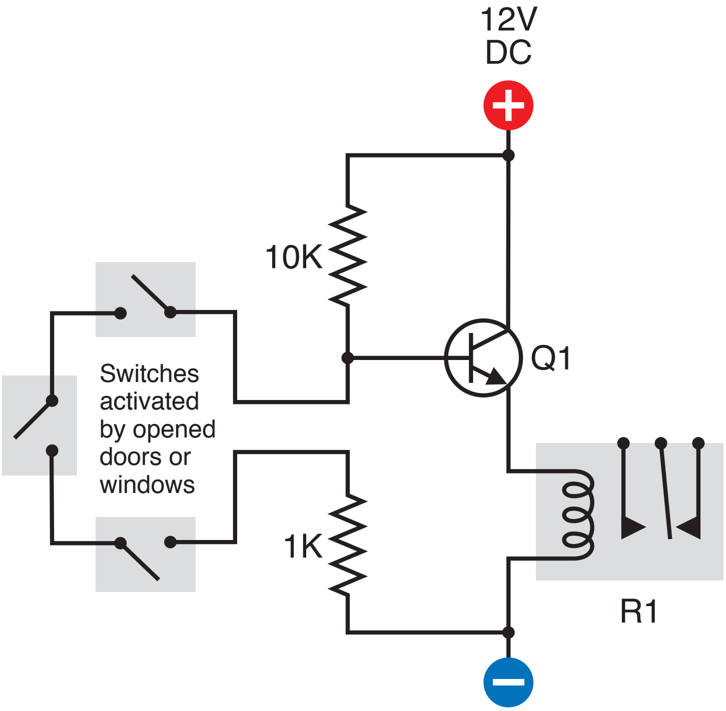

Figure 3-89. *If the LED and 680Ω resistor shown in are removed, and a relay takes their place, the relay will be activated when any switch in the sensor network is opened.*

Try one of the 12-volt relays that you used previously. You should find that when you open the switch, the relay is energized. When you close the switch, the relay goes back to sleep. Note that I eliminated the 680Ω resistor from the circuit, because the relay doesn't need any protection from the 12-volt power supply.

Self-Locking Relay

There's only one remaining problem: we want the alarm to continue making noise even after someone who has opened a door or window closes it again quickly. In other words, when the relay is activated, it must lock itself on.

One way to do this would be by using a latching relay. The only problem is that we would then need another piece of circuitry to unlatch it. I prefer to show you how you can make any relay keep itself switched on after it has received just one jolt of power. This idea will be useful to you later in the book as well.

The secret is to supply power to the relay coil through the two contacts inside the relay that are normally open. (Note that this is exactly opposite to the relay oscillator, which supplied power to its coil through the contacts that were normally closed. That setup caused the relay to switch itself off almost as soon as it switched itself on. This setup causes the relay to keep itself switched on, as soon as it has been activated.)

In Figure 3-90, the four schematics illustrate this. You can imagine them as being like frames in a movie, photographed microseconds apart. In the first picture, the switch is open, the relay is not energized, and nothing is happening. In the second, the switch has been closed to energize the coil. In the third, the coil has pulled the contact inside the relay, so that power now reaches the coil via two paths. In the fourth, the switch has been opened, but the relay is still powering its own coil through its contacts. It will remain locked in this state until the power is disconnected.

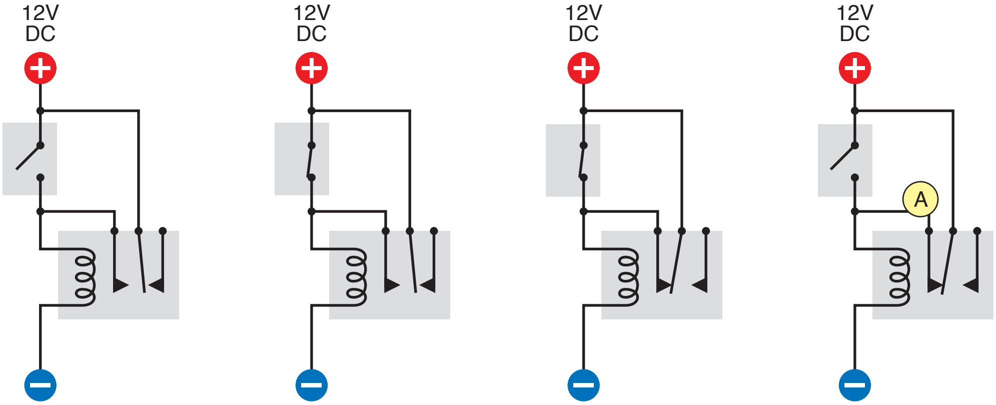

Figure 3-90. *This sequence of schematics shows the events that occur when a relay is energized. Initially, the switch is open. Then the switch is closed, activating the relay. The relay then powers itself through its own internal contacts. The relay remains energized even after the switch is opened again. Power switched by the relay can be taken from the circuit at point A.*

All we need to do, to make use of this idea, is to substitute the transistor for the on/off switch, and tap into the circuit at point A, running a wire from there to the noisemaking module.

Figure 3-91 shows how that would work. When the transistor is activated by any of the network of sensor switches, as previously explained, the transistor conducts power to the relay. The relay locks itself on, and the transistor becomes irrelevant.

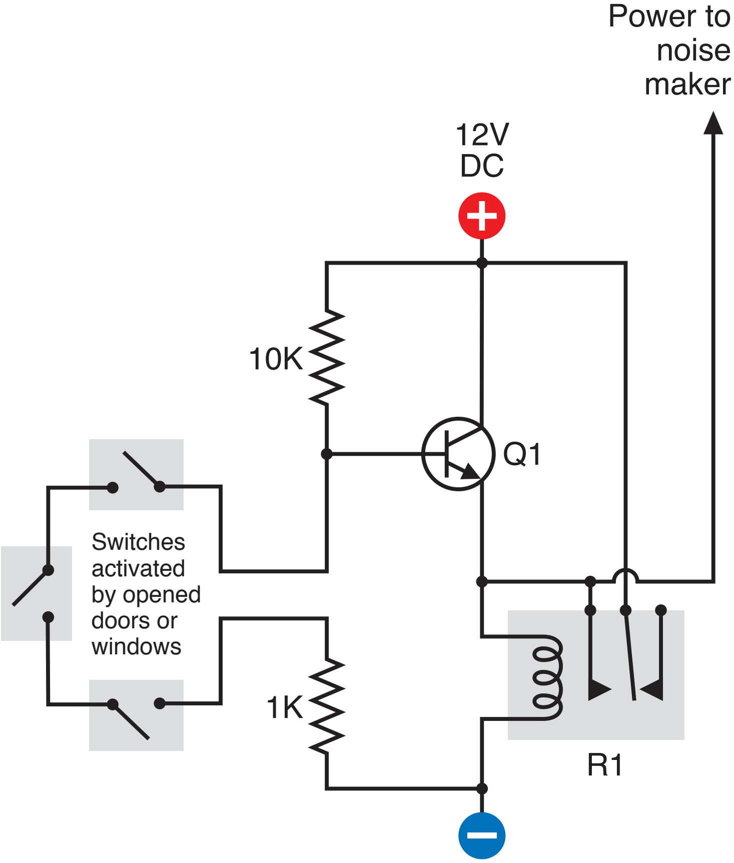
Power to
noise
maker
12V
DC
10K
Q1
Switches
activated
by opened
doors or
windows
1K
R1

Figure 3-91. *The self-locking relay depicted in has been incorporated in the alarm circuit, so that if any switch in the network is opened, the relay will continue to power the noise maker even if the switch is closed again.*

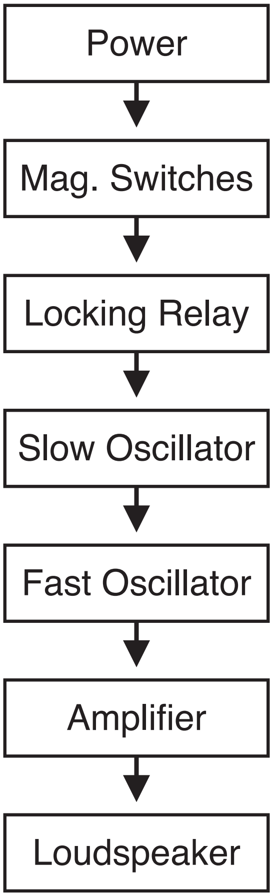
Power
Mag. Switches
Locking Relay
Slow Oscillator
Fast Oscillator
Amplifier
Loudspeaker

Figure 3-92. *This block diagram previously shown in Figure 2-112 on page 90 has been updated to include the magnetic-switch network and locking-relay control system.*

Because I've been adding pieces to the original alarm noisemaker circuit, I've updated the block diagram from Figure 2-112 to show that we can still break this down into modules with simple functions. The revised diagram is shown in Figure 3-92.

Blocking Bad Voltage

One little problem remains: in the new version of the circuit, if the transistor goes off while the relay is still on, current from the relay can flow back up the wire to the emitter of the transistor, where it will try to flow backward through the transistor to the base, which is "more negative," as it is linked through all the magnetic switches and the 1K resistor to the negative side of the power supply.

Applying power backward through a transistor is not a nice thing to do. Therefore the final schematic in this series shows one more component, which you have not seen before: a diode, labeled D1. See Figure 3-93. The diode looks like the heart of an LED, and indeed, that's pretty much what it is, although some diodes are much more robust. It allows electricity to flow in only one direction, from positive to negative, as shown by its arrow symbol. If current tries to flow in the opposite direction, the diode blocks it. The only price you pay for this service is that the diode imposes a small voltage drop on electricity flowing in the "OK" direction.

So now, positive flow can pass from the transistor, through the diode, to the relay coil, to get things started. The relay then supplies itself with power, but the diode prevents the positive voltage from getting back into the transistor the wrong way.

Perhaps a more elegant solution to the problem is to connect the NO leg of the relay via a 10k resistor to the base connection. When the relay is not energized, the NO leg is inert and simply behaves as a parasitic capacitance on the node. When the relay becomes energized, the NO leg shunts +12V through the common terminal via a 10k resistor into the base of the transistor. In this circuit configuration, the transistor is never exposed to a potentially harmful voltage and you are not depending on leakage currents of non-ideal elements to protect devices.

However, I needed an opportunity to introduce you to the concept of diodes. You can check the following section "Essentials: All about diodes" to learn more.

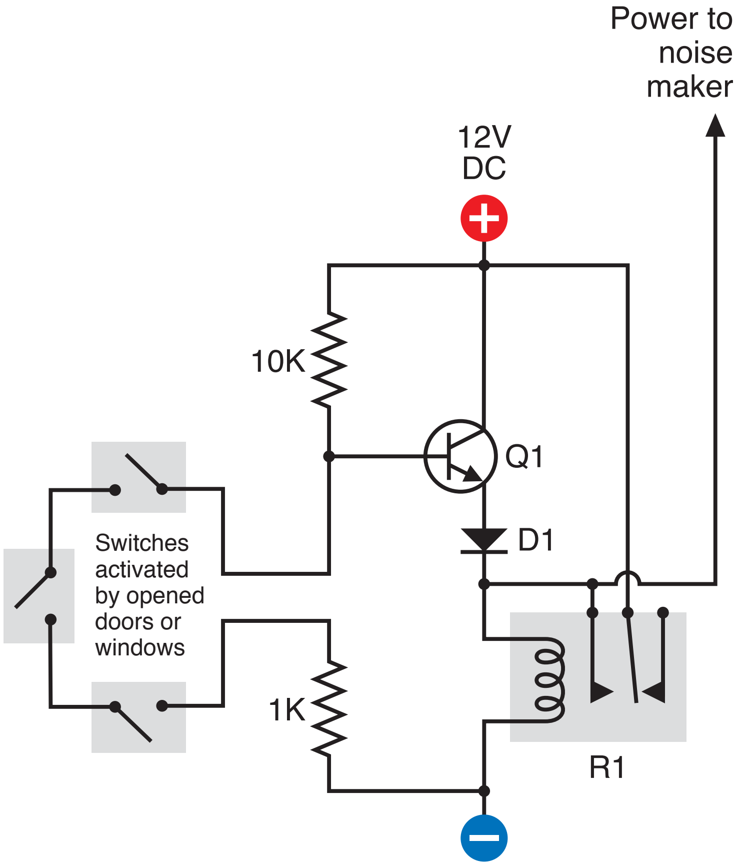

Figure 3-93. *Diode D1 has been added to protect the emitter of Q1 from positive voltage when the relay is energized.*

ESSENTIALS

All about diodes

A diode is a very early type of semiconductor. It allows electricity to flow in one direction, but blocks it in the opposite direction. (A light-emitting diode is a much more recent invention.) Like an LED, a diode can be damaged by reversing the voltage and applying excessive power, but most diodes generally have a much greater tolerance for this than LEDs. The end of the diode that blocks positive voltage is always marked, usually with a circular band, while the other end remains unmarked. Diodes are especially useful in logic circuits, and can also convert alternating current (AC) into direct current (DC).

A Zener diode is a special type that we won't be using in this book. It blocks current completely in one direction, and also blocks it in the other direction until a threshold voltage is reached—much like a PUT.

Signal diodes are available for various different voltages and wattages. The 1N4001 diode that I recommend for the alarm activation circuit is capable of handling a much greater load at a much higher voltage, but I used it because it has a low internal resistance. I wanted the diode to impose a minimal voltage drop, so that the relay would receive as much voltage as possible.

It's good practice to use diodes at less than their rated capacity. Like any semiconductor, they can overheat and burn out if they are subjected to mistreatment.

The schematic symbol for a diode has only one significant variant: sometimes the triangle is outlined instead of filled solid black (see Figure 3-94).

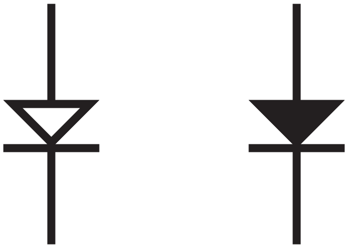

Figure 3-94. *Either of these schematic symbols may be used to represent a diode, but the one on the right is more common than the one on the left.*

Completing the Breadboard Alarm Circuit

It's time now to breadboard the control circuit for your alarm noisemaker. Figure 3-95 shows how this can be done. I am assuming that you still have the noisemaker, which functions as before. I'm assuming that you still have its relevant components mounted on the top half of the breadboard. To save space, I'm just going to show the additional components mounted on the bottom half of the same breadboard.

It's important to remember that you are not supplying power directly to the left and right "rails" on the breadboard anymore; you are supplying power to the relay-transistor section, and when the relay closes its contacts, the *relay* supplies power to the rails. These then feed the power up to the top half of the breadboard. So disconnect your power supply from the breadboard rails and reconnect it as shown in Figure 3-95.

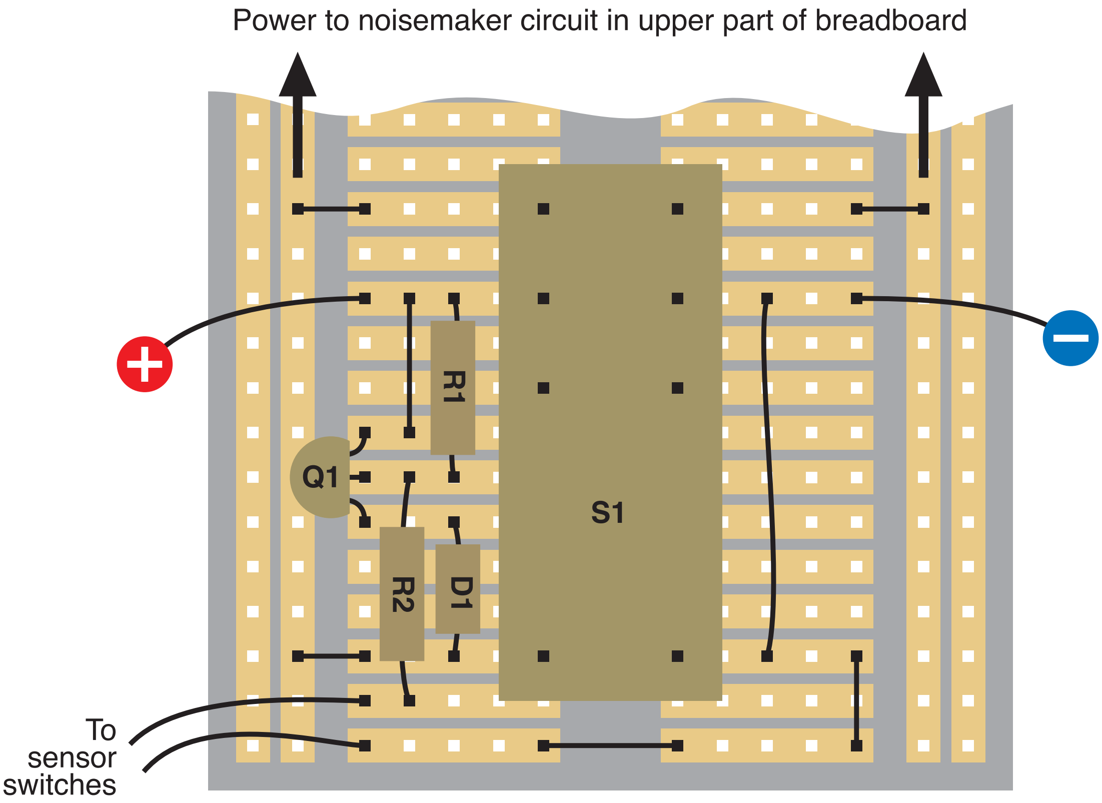

Figure 3-95. *The schematic that was developed in the previous pages can be emulated with components on a breadboard, as shown here. S1 is a DPDT relay. Wires to the sensor switch network and to the power supply must be added where shown.*

Because it's a double-pole relay, I am using it to switch negative as well as positive. This means that when the relay contacts are open, the noisemaking section of the circuit is completely isolated from the rest of the world.

The breadboarded relay circuit is exactly the same as the schematic in Figure 3-93. The components have just been rearranged and squeezed together so that they will fit alongside the relay. Two wires at the lower-left corner go to the network of magnetic sensor switches that will trip the alarm; for testing purposes, you can just hold the stripped ends of these two wires together to simulate all the switches being closed, and separate the wires to simulate a switch opening.

Two more wires bring power to the breadboard on either side of the relay. This is where you should connect your power supply during testing. The output from the relay, through its top pair of contacts, is connected with the rails of the breadboard by a little jumper wire at top left, and another at top right. Don't forget to include them! One more little wire at the lower-left corner (easily overlooked) connects the lefthand side rail to the lefthand coil terminal of the relay, so that when the relay is powering the noisemaker circuit, it powers itself as well.

When you mount the diode, remember that the end of it that is marked with a band around it is the end that blocks positive current. In this circuit, that's the lower end of the diode.

Try it to make sure that it works. Short the sensor wires together and then apply power. The alarm should remain silent. You can use your meter to check that no voltage exists between the side rails. Now separate the sensor wires, and the relay should click, supplying power to the side rails, which activates the noisemaker. Even if you bring the sensor wires back together, the relay should remain locked on. The only way to unlock it is to disconnect the power supply.

When the circuit is active, the transistor followed by the diode drops the voltage slightly, but the 12-volt relay should still work.

In my test circuit, trying three different relays, they drew between 27 and 40 milliamps at 9.6 volts. Some current still leaked through the transistor when it was in its "off" mode, but only a couple of milliamps at 0.5 volts. This low voltage was far below the threshold required to trip the relay.

Ready for Perfboarding

If the circuit works, the next step is to immortalize it on perforated board. Use the type of board that has a breadboard contact pattern etched on it in copper, as shown in Figure 3-72 on page 116. Check the following section, "Essentials: Perfboard procedure," for guidance on the best way to make this particular kind of solder joint—and the subsequent section for the most common problems.

ESSENTIALS

Perfboard soldering procedure

Carefully note the position of a component on your breadboard, and then move it to the same relative position on the perfboard, poking its wires through the little holes.

Turn the perfboard upside down, make sure that it's stable, and examine the hole where the wire is poking through, as shown in Figure 3-96. A copper trace surrounds this hole and links it with others. Your task is to melt solder so that it sticks to the copper and also to the wire, forming a solid, reliable connection between the two of them.

Take your pencil-style soldering iron in one hand and some solder in your other hand. Hold the tip of the iron against the wire and the copper, and feed some thin solder to their intersection. After two to four seconds, the solder should start flowing.

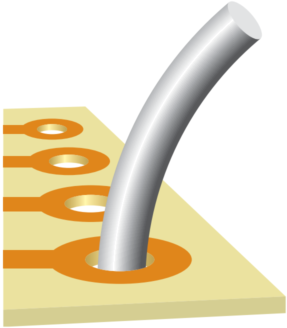

Figure 3-96

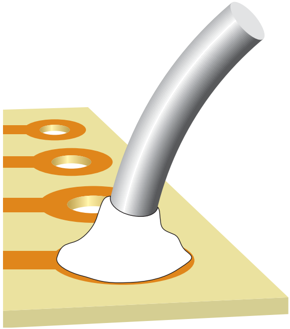

Figure 3-97

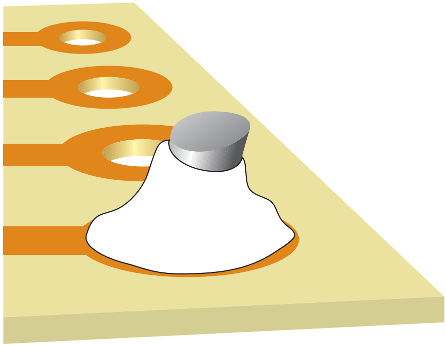

Figure 3-98. *To establish a connection between a section of wire and a copper trace on perforated board, the wire is pushed through the hole, and solder (shown in pure white for illustrative purposes) completes the connection. The wire can then be snipped short.*

Allow enough solder to form a rounded bump sealing the wire and the copper, as shown in Figure 3-97. Wait for the solder to harden thoroughly, and then grab the wire with pointed-nosed pliers and wiggle it to make sure you have a strong connection. If all is well, snip the protruding wire with your cutters. See Figure 3-98.

Because solder joints are difficult to photograph, I'm using drawings to show the wire before and after making a reasonably good joint, which is shown in pure white, outlined with a black line.

Actual soldered perfboard is shown in the photographs in Figures 3-99 and 3-100.

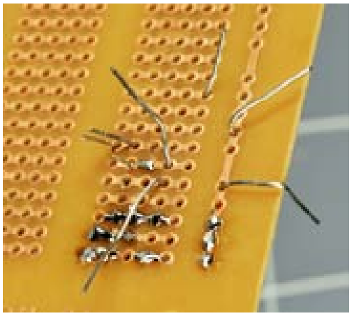

Figure 3-99. *This photograph was taken during the process of transferring components from breadboard to perforated board. Two or three components at a time are inserted from the other side of the board, and their leads are bent over to prevent them from falling out.*

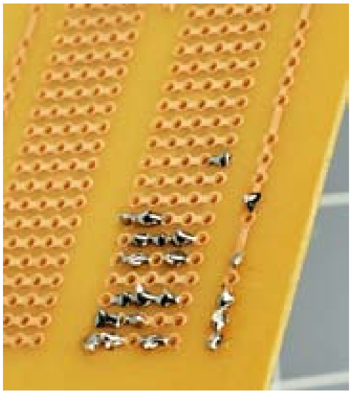

Figure 3-100. *After soldering, the leads are snipped short and the joints are inspected under a magnifying glass. Another two or three components can now be inserted, and the process can be repeated.*

TOOLS

Four most common perfboarding errors

1. Too much solder

Before you know it, solder creeps across the board, touches the next copper trace, and sticks to it, as depicted in Figure 3-101. When this happens, you have to wait for it to cool, and then cut it with a utility knife. You can also try to remove it with a rubber bulb and solder wick, but some of it will tend to remain.

Even a microscopic trace of solder is enough to create a short circuit. Check the wiring under a magnifying glass while turning the perfboard so that the light strikes it from different angles (or use your solder wick to suck it away).

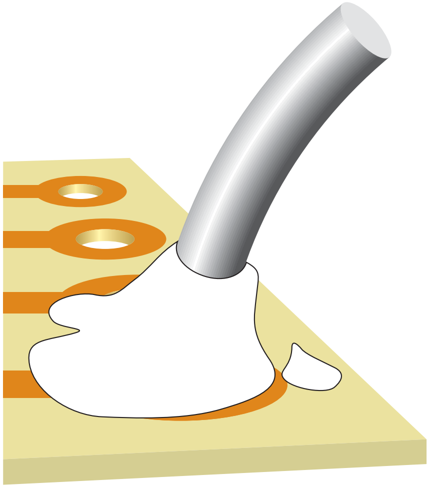

Figure 3-101. *If too much solder is used, it makes a mess and can create an unwanted connection with another conductor.*

2. Not enough solder

If the joint is thin, the wire can break free from the solder as it cools. Even a microscopic crack is sufficient to stop the circuit from working. In extreme cases, the solder sticks to the wire, and sticks to the copper trace around the wire, yet doesn't make a solid bridge connecting the two, leaving the wire encircled by solder yet untouched by it, as shown in Figure 3-102. You may find this undetectable unless you observe it with magnification.

You can add more solder to any joint that may have insufficient solder, but be sure to reheat the joint thoroughly.

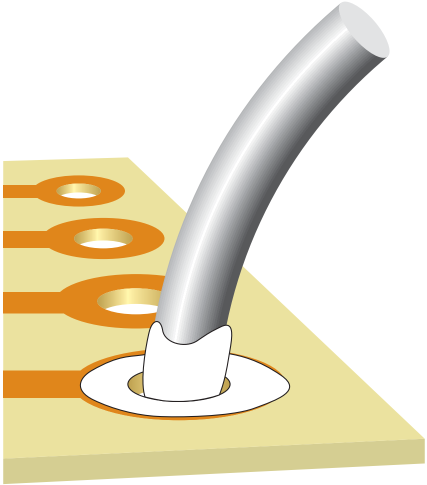

Figure 3-102. *Too little solder (or insufficient heat) can allow a soldered wire to remain separate from the soldered copper on the perforated board. Even a hair-thin gap is sufficient to prevent an electrical connection.*

3. Components incorrectly placed

It's very easy to put a component one hole away from the position where it should be. It's also easy to forget to make a connection.

I suggest that you print a copy of the schematic, and each time you make a connection on the perforated board, you eliminate that wire on your hardcopy, using a highlighter.

4. Debris

When you're trimming wires, the little fragments that you cut don't disappear. They start to clutter your work area, and one of them can easily get trapped under your perforated board, creating an electrical connection where you don't want it.

This is another reason for working with something soft, such as polyurethane foam, under your project. It tends to absorb or hold little pieces of debris, reducing the risk of you picking them up in your wiring.

Clean the underside of your board with an old (dry) toothbrush before you apply power to it, and keep your work area as neat as possible. The more meticulous you are, the fewer problems you'll have later.

Once again, be sure to check every joint with a magnifying glass.

Transferring components from the breadboard to the perforated board should be fairly simple, as long as you don't try to move too many at once. Follow the suggestions described in previous section "Essentials: Perfboard procedure," and pause frequently to check your connections. Impatience is almost always the cause of errors in this kind of work.

Figure 3-103 shows the noisemaker section of the circuit on perfboard, with the components positioned to minimize wasted space. Figure 3-104 shows the perfboard with the relay and its associated components added. The two black wires will go to the loudspeaker, the black-and-red pair of wires will bring power to the board, and the green wires will go to the magnetic sensor switches. Each wire penetrates the board, and its stripped end is soldered to the copper beneath.

Test it now, in the same way that you tested the same circuit on the breadboard. If it doesn't work, check the following section, "Essentials: Real-world fault tracing." If it does work, you're ready to trim the board and mount it in a project box.

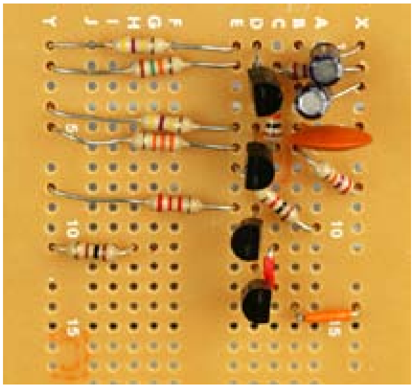

Figure 3-103. *The noisemaker circuit has been transplanted from breadboard to perforated board, with no additions or changes.*

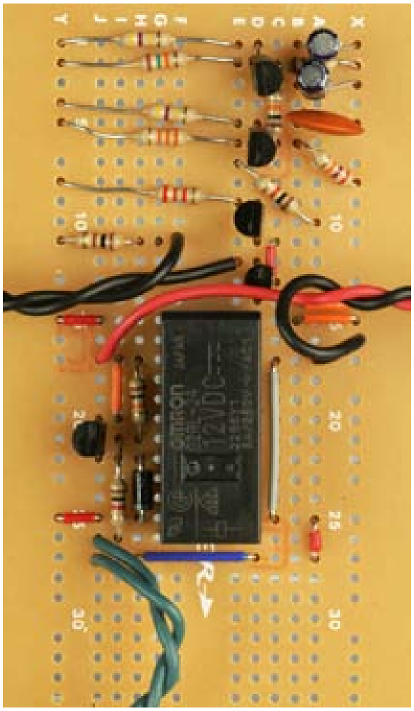

Figure 3-104. *The relay-transistor control circuit has been added. Wires to external devices have been stripped and poked into the perforated board, where they are soldered in place. The green wires connect with the sensor network, the black wires go to the loudspeaker, and the red-and-black wires supply power.*

ESSENTIALS

Real-world fault tracing

Here's a real-life description of the procedure for tracing a fault.

After I assembled the perfboard version of the combined noisemaker and relay circuit, I checked my work, applied power—and although the relay clicked, no sound came out of the loudspeaker. Of course, everything had worked fine on the breadboard.

First I looked at component placement, because this is the easiest thing to verify. I found no errors. Then I flexed the board gently while applying power—and the loudspeaker made a brief "beep." Any time this happens, you can be virtually certain that a solder joint has a tiny crack in it.

The next step was to anchor the black lead of my meter to the negative side of the power supply, and then switch on the power and go through the circuit point by point, from top to bottom, checking the voltage at each point with the red lead of the meter. In a simple circuit like this, every part should show at least some voltage.

But when I got to the second 2N2222 transistor, which powers the loudspeaker, its output was completely dead. Either I had melted the transistor while soldering it (unlikely), or there was a bad joint. I checked the perfboard beneath the transistor with the magnifying glass, and found that solder had flowed around one of the leads of the transistor without actually sticking to it. The gap must have been less than one-thousandth of an inch, but still, that was enough. Probably, the problem had been caused by dirt or grease.

This is the kind of patient inquiry you need to follow when a circuit doesn't work. Check whether your components are placed correctly, check your power supply, check the power on the board, check the voltage at each stage, and if you are persistent, you'll find the fault.

Switches and Inputs for the Alarm

Now you need to make the system easy to use. The block diagram in Figure 3-105 shows one additional box near the top of the sequence: User Controls. These will consist of switches, LEDs, and connections to the outside world. To plan this part of the job, first I have to summarize the way in which our alarm system works at this point in its evolution.

A full-featured home alarm system normally has two modes: in-home and away-from-home:

- Using the in-home mode, you switch on the alarm while you are at home so that it will alert you if an intruder opens a door or window.
- Using the away-from-home mode, typically you enter a code number, after which you have 30 seconds to leave and close the door behind you. When you return, you trigger the alarm by opening the door, but now you have 30 seconds to go to the control panel and enter your code number again to stop the alarm from sounding.

So far, the alarm system that you've been building has only an in-home mode. Still, many people find this function useful and reassuring. Later in the book I'll suggest a way in which you can modify it to incorporate an away-from-home mode, but for now, making it practical for in-home use is enough of a challenge.

Consider how it should be used on an everyday basis. It should have an on/off switch, naturally. When it's on, any of the magnetic sensor switches should trigger the alarm. But what if you switch it on without realizing that you've left a window open? At that time it won't be appropriate for the alarm to sound. What you really need is a circuit-test feature, to tell you if all the doors and windows are closed. Then you can switch on the alarm.

I think a pushbutton would be useful to test the alarm circuit. When you press it, a green LED should light up to indicate that the circuit is good. After you see the green light, you let go of the pushbutton and turn on the power switch, which illuminates a red LED, to remind you that the alarm is now armed and ready.

One additional feature would be useful: an alarm noisemaker test feature, so that you can be sure that the system is capable of sounding its alert when required to do so.

The circuit shown in Figure 3-106 incorporates all of these features. S1 is a SPDT switch; S2 is a DPDT momentary pushbutton of ON-(ON) type. The schematic shows it in its "relaxed" mode, when the button is not being pressed.

D1 is a red LED, D2 is a green LED, J1 is a power input jack (to be connected with an external 12-volt supply), and R1 is a 680Ω resistor to protect the LEDs. Note that J1 follows the usual practice of supplying positive voltage in its center contact, and negative in the circular shell around the center.

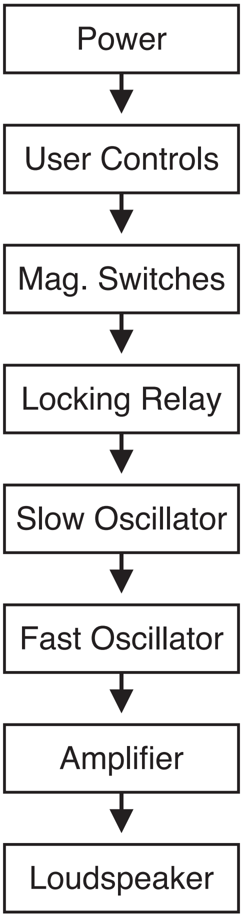

Figure 3-105. *The final block diagram for this phase of the project shows where user controls fit in the series of functions.*

When S1 is in its Off position, it still supplies positive power through its upper contact to S2, the pushbutton. When the pushbutton is pressed, so that it goes into its "Test" position, the pole of S2 connects with the power and sends it out through the sensor switches on doors and windows. The wires to these switches will be attached via a couple of binding posts, shown here as two circles. If the sensor switches are all closed, power returns through the second binding post, passes through the lower set of contacts in S2, and lights D2, the green LED. Because S1 is not supplying power to the alarm circuit board, the alarm does not sound at this time.

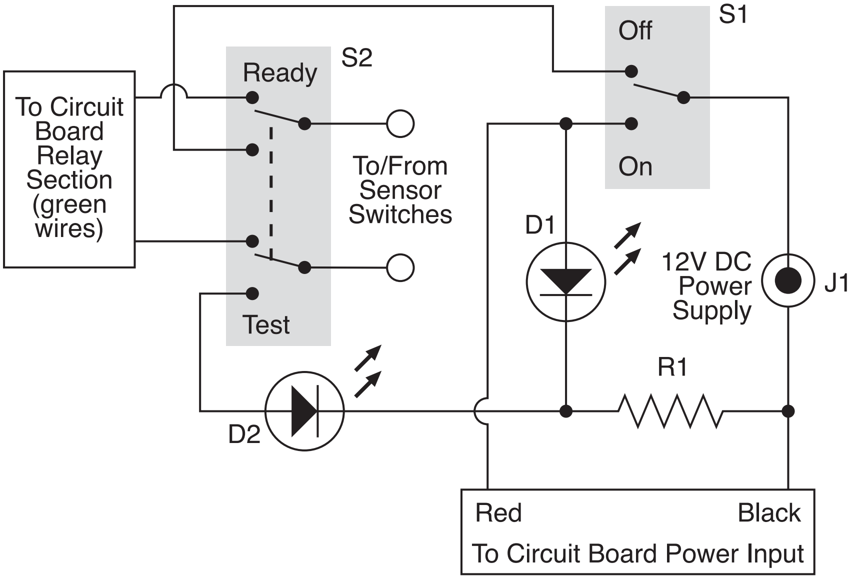

Figure 3-108. *This schematic suggests a convenient way to add an on/off switch, a continuity testing feature, and a noisemaker testing feature to the alarm.*

Now if S1 is turned to its On position, it sends power to the components on the perforated board. The transistor circuit sends power out along the green wires to S2, and as long as the button is not being pressed, the power goes out through the switch network and back through S2, to the transistor, keeping its base relatively negative. So the alarm remains silent. But as soon as a sensor switch is opened, the circuit is broken and the alarm sounds. The only way to stop it will be by switching S1 into its Off position.

Finally, if you press S2, the pushbutton, while S1 is in its On mode, you interrupt the network of switches and activate the noisemaker. In this way, S2 does dual duty: when S1 is off, pushbutton S2 tests the sensor switches for continuity. When S1 is on, S2 tests the noisemaker to make sure it creates a noise. I think this is the simplest possible way to implement these features.

Installing the Switches

If you bought a project box from RadioShack, it may have come with two optional top panels: one made of metal, the other made of plastic. I'll assume that you'll use the plastic one, as you'll have more trouble drilling holes in metal. The type of plastic used by RadioShack is ABS, which is very easy to shape with the tools that I have recommended.

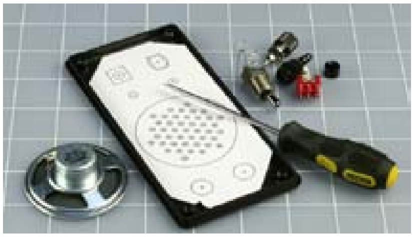

Figure 3-107. *A printed layout for the switches, LEDs, and other components has been taped to the underside of the lid of the project box. An awl is pressed through the paper to mark the center of each hole to be drilled in the lid.*

You have to choose a layout for the switches and other components that will share the top panel of the project box. I like a layout to be neat, so I take the trouble to draw it using illustration software, but a full-size pencil sketch is almost as good. Just make sure there's room for the components to fit together, and try to place them similarly to the schematic, to minimize the risk of confusion.

Tape your sketch to the inside of the top panel, as shown in Figure 3-107, and then use a sharp pointed tool, such as a pick, to press through and mark the plastic at the center of each hole. The indentations will help to center your bit when you drill the holes. Remember that you'll need to make multiple holes to vent the sound from the loudspeaker, which will be beneath the top panel of the box. The result is shown in Figure 3-108.

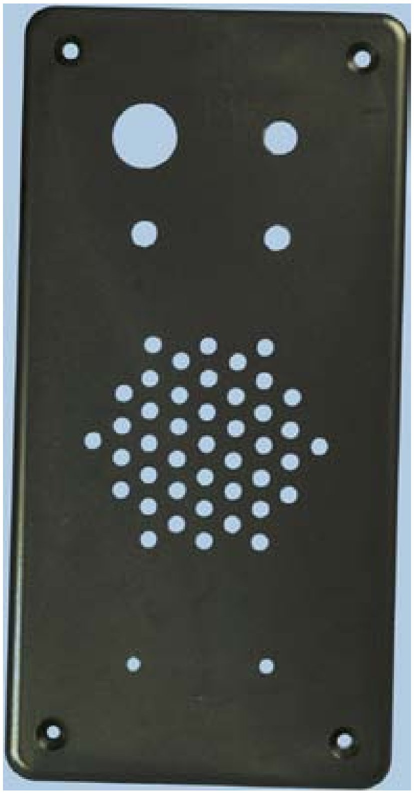

Figure 3-108. *The exterior of the panel after drilling. A small handheld cordless drill can create a neat result if the holes were marked carefully.*

I placed all the components on the top panel, with the exception of the power input jack, which I positioned at one end of the box. Naturally, each hole has to be sized to fit its component, and if you have calipers, they'll be very useful for taking measurements and selecting the right drill bit. Otherwise, make your best guess, too small being better than too large. A deburring tool is ideal for slightly enlarging a hole so that a component fits snugly. This may be necessary if you drill 3/16-inch holes for your 5-inch mm LEDs. Fractionally enlarge each hole, and the LEDs should push in very snugly.

If your loudspeaker lacks mounting holes, you'll have to glue it in place. I used five-minute epoxy to do this. Be careful not to use too much. You don't want any of the glue to touch the speaker cone.

Drilling large holes in the thin, soft plastic of a project box can be a problem. The drill bit tends to dig in and create a mess. You can approach this problem in one of three ways:

1. Use a Forstner drill bit if you have one. It creates a very clean hole.
2. Drill a series of holes of increasing size.
3. Drill a smaller hole than you need, and enlarge it with a deburring tool.

Regardless of which approach you use, you'll need to clamp or hold the top panel of the project box with its outside surface face-down on a piece of scrap wood. Then drill from the inside, so that your bit will pass through the plastic and into the wood.

Finally, mount the components in the panel, as shown in Figure 3-109, and turn your attention to the underneath part of the box.

The circuit board will sit on the bottom, held in place with four #4-size machine screws (bolts) with washers and nylon-insert locknuts. You need to use locknuts to eliminate the risk of a nut working loose and falling among components where it can cause a short circuit.

You'll have to cut the perfboard to fit, taking care not to damage any of the components on it. Also check the underside the board for loose fragments of copper traces after you finish cutting.

Drill bolt holes in the board, if necessary, taking care again not to damage any components. Then mark through the holes to the plastic bottom of the box, and drill the box. Countersink the holes (i.e., bevel the edges of a hole so that a flat-headed screw will fit into it flush with the surrounding surface), push the little bolts up from underneath, and install the circuit board. Be extremely careful not to attach the circuit board too tightly to the project box. This can impose bending stresses, which may break a joint or a copper trace on the board.

I like to include a soft piece of plastic under the board to absorb any stresses. Because you're using locknuts, which will not loosen, there's no need to make them especially tight.

Test the circuit again after mounting the circuit board, just in case.

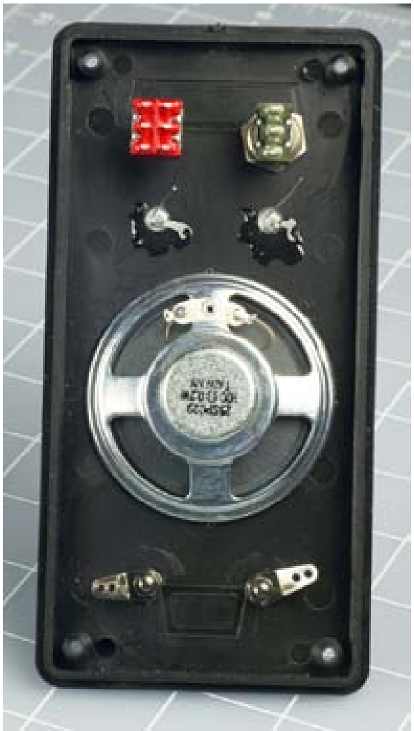

Figure 3-109. *Components have been added to the control panel of the project box (seen from the underside). The loudspeaker has been glued in place. Spare glue was dabbed onto the LEDs, just in case. The SPDT on/off switch is at the top right, the DPDT pushbutton is at top left, and the binding posts, which will connect with the network of magnetic sensor switches, are at the bottom.*

Soldering the Switches

Figure 3-110 shows how the physical switches should be wired together. Remember that S1 is a toggle switch and S2 is a DPDT pushbutton. Your first step is to decide which way up they should be. Use your meter to find out which terminals are connected when the switch is flipped, and when the button is pressed. You'll probably want the switch to be on when the toggle is flipped upward. Be especially careful with the orientation of the pushbutton, because if you wire it upside-down, it will constantly have the alarm in "test" mode, which is not what you want.

Remember, the center terminal of any double-throw switch is almost always the pole of the switch, connecting with the terminals immediately above it and below it.

Stranded wire is appropriate to connect the circuit board with the components in the top panel, because the strands flex easily and impose less stress on solder joints. Twisting each pair of wires together helps to minimize the mess.

Remember to install the LEDs with their short, negative wires connected with the resistor. This will entail some wire-to-wire soldering. You may want to protect some of these bare leads and joints with thin heat-shrink tubing, to minimize the risk of short circuits when you push all the parts into the box.

When you connect wires or components with the lugs on the switches, your pencil-style soldering iron probably won't deliver enough heat to make good joints. You can use your higher-powered soldering iron in these locations, but you absolutely must apply a good heat sink to protect the LEDs when you attach them, and don't allow the iron to remain in contact with anything for more than 10 seconds. It will quickly melt insulation, and may even damage the internal parts of the switches.

In projects that are more complex than this one, it would be good practice to link the top panel with the circuit board more neatly. Multicolored ribbon cable is ideal for this purpose, with plug-and-socket connectors that attach to the board. For this introductory project, I didn't bother. The wires just straggle around, as shown in Figure 3-111.

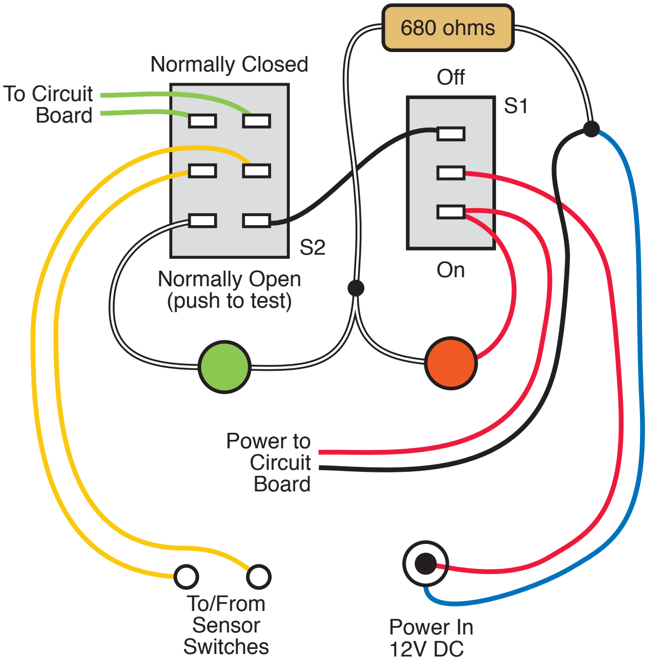

Figure 3-110. *The components can be wired together like this to replicate the circuit shown in The red and green circles are LEDs. Small, solid black circles indicate wire-to-wire solder joints.*

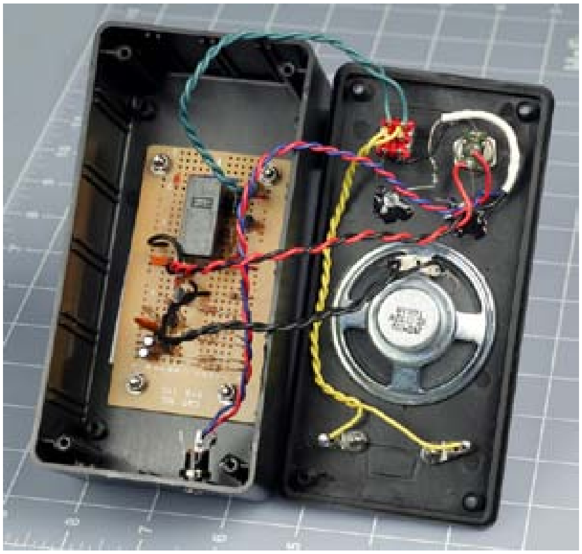

Figure 3-111. *The circuit board has been installed in the base of the project box, and the power input jack has been screwed into the end of the box. Twisted wire-pairs have been connected on a point-to-point basis, without much concern over neatness, as this is a relatively small project. The white insulation at the top-right corner of the front panel is heat-shrink tube that encloses a solder joint and the 680Ω load resistor. Soldering wires to the pushbutton switch requires care and precision, as the contacts are closely spaced.*

Final Test

When you've completed the circuit, test it! If you don't have your network of magnetic sensor switches set up yet, you can just use a piece of wire to connect the two binding posts. Make sure that S1 is in its Off position, then solder the appropriate plug to your 12-volt power source, and plug it into the power jack. When you press the button, the green LED should light up to show continuity between the two binding posts. Now disconnect the wire between the binding posts, press the button again, and the green LED should remain dark.

Reconnect the binding posts, flip S1 to its On position, and the red LED should light up. Press the button, and the alarm should start. Reset it by turning S1 off and then on again; then disconnect the wire between the binding posts. Again, the alarm should start, and it should continue even if you reconnect the wire.

If everything works the way it should, it's time to screw the top of the box in place, pushing the wires inside. Because you're using a large box, you should have no risk of metal parts touching each other accidentally, but still, proceed carefully.

Alarm Installation

Before you install your magnetic sensor switches, you should test each one by moving the magnetic module near the switch module and then away from it, while you use your meter to test continuity between the switch terminals. The switch should close when it's next to the magnet, and open when the magnet is removed.

Now draw a sketch of how you'll wire your switches together. Always remember that they have to be in series, not in parallel! Figure 3-112 shows the concept in theory. The two terminals are the binding posts on top of your control box (which is shown in green), and the dark red rectangles are the magnetic sensor switches on windows and doors. Because the wire for this kind of installation usually has two conductors, you can lay it as I've indicated but cut and solder it to create branches. The solder joints are shown as orange dots. Note how current flows through all the switches in series before it gets back to the control box.

Figure 3-113 shows the same network as you might actually install it in a situation where you have two windows and a door. The blue rectangles are the magnetic modules that activate the switch modules.

You'll need a large quantity of wire, obviously. The type of white, stranded wire that is sold for doorbells or furnace thermostats is good. Typically, it is 20-gauge or larger.

After you install all the switches, clip your meter leads to the wires that would normally attach to the alarm box. Set your meter to test continuity, and open each window or door, one at a time, to check whether you're breaking the continuity. If everything is OK, attach the alarm wires to the binding posts on your project box.

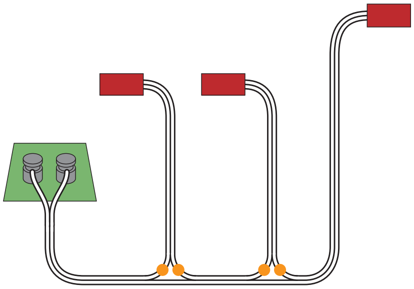

Figure 3-112. *Dual-conductor, white insulated wire can be used to connect the terminals on the alarm control box with magnetic sensors (shown in dark red). Because the sensors must be in series, the wire is cut and joined at positions marked with orange dots.*

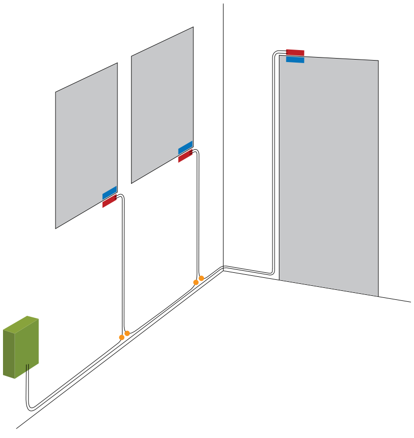

Figure 3-113. *In an installation involving two windows and a door, the magnetic components of the sensors (blue rectangles) could be placed as shown, while the switches (dark red) are located alongside them.*

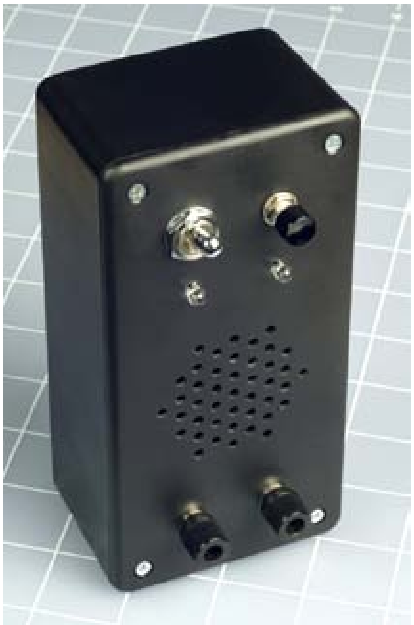

Figure 3-114. *The intrusion alarm completed and in its project box.*

Now deal with the power supply. Use your AC adapter, set to 12 volts, hooked up to your type N power plug, or attach the power plug to a 12-volt alarm battery.

If you use a battery, be especially careful that the wire leading to the center terminal of your power plug is positive! A 12-volt battery can deliver substantial current, which can fry your components if you connect it the wrong way around. It would be a shame to destroy your entire project at the very last step.

The only remaining task is to label the switch, button, power socket, and binding posts on the alarm box. You know that the switch turns the power on and off, and the button tests the circuit and the noisemaker, but no one else knows, and you might want to allow a guest to use your alarm while you're away. For that matter, months or years from now, you may forget some details. Will you remember that the power source for this unit should be 12 volts?

Labeling really is a good idea. But as you can see in Figure 3-114, I haven't quite gotten around to it for the box that I built.

Conclusion

The alarm project has taken you through the basic steps that you will usually follow any time you develop something:

1. Draw a schematic and make sure that you understand it.
2. Modify it to fit the pattern of conductors on a breadboard.
3. Install components on the breadboard and test the basic functions.
4. Modify or enhance the circuit, and retest.
5. Transfer to perforated board, test, and trace faults if necessary.
6. Add switches, buttons, power jack, and plugs or sockets to connect the circuit with the outside world.
7. Mount everything in a box (and add labeling).

While going through this sequence, I hope you've learned the basics of electricity, along with some simple electrical theory, and fundamentals about electronic components. This knowledge should enable you to move on to the much more powerful realm of integrated circuit—which I'll cover in Chapter 4.

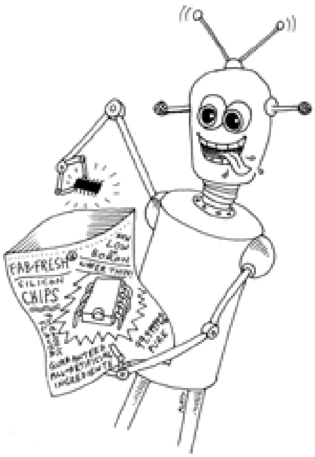

4

Chips, Ahoy!

IN THIS CHAPTER

Shopping List: Experiments 16 Through 24
Experiment 16: Emitting a Pulse
Experiment 17: Set Your Tone
Experiment 18: Reaction Timer
Experiment 19: Learning Logic
Experiment 20: A Powerful Combination
Experiment 21: Race to Place
Experiment 22: Flipping and Bouncing
Experiment 23: Nice Dice
Experiment 24: Intrusion Alarm Completed

Before I get into the fascinating topic of integrated circuit (IC) chips, I have to make a confession: some of the things I asked you to do in Chapter 3 could have been done a bit more simply. Does this mean you have been wasting your time? No, I firmly believe that by building circuits with old-fashioned components—capacitors, resistors, and transistors—you acquire the best possible understanding of the principles of electronics. Still, you are going to find that integrated circuit chips, containing dozens, hundreds, or even thousands of transistor junctions, will enable some shortcuts.

Shopping List: Experiments 16 Through 24

Tools

The only new tool that I recommend using in conjunction with chips is a logic probe. This tells you whether a single pin on a chip has a high or low voltage, which can be helpful in figuring out what your circuit is doing. The probe has a memory function so that it will light its LED, and keep it lit, in response to a pulse that may have been too quick for the eye to see.

Search online and buy the cheapest logic probe you can find. I don't have any specific brand recommendations. The one shown in Figure 4-1 is fairly typical.

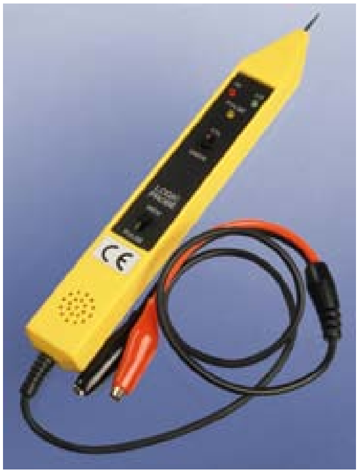

Figure 4-1. *A logic probe detects the high or low voltage on each pin of a chip, and reveals pulses that may occur too quickly for you to perceive them with the unaided eye.*

Supplies

Integrated circuit chips

If you buy everything on this shopping list, and you bought basic parts such as resistors and capacitors that were listed previously, you should have everything you need for all the projects in this chapter.

As chips are quite cheap (currently around 50 cents apiece), I suggest you buy extras. This way, if you damage one, you'll have some in reserve. You'll also have a stock for future projects.

Please read the next section, "Fundamentals: Choosing chips," before you begin chip-shopping. Chips should be easily obtainable from all the major electronics retail suppliers, and sometimes are found on eBay shops. Look in the appendix for a complete list of URLs.

FUNDAMENTALS

Choosing chips

Figure 4-2 shows what is often referred to as an integrated circuit (IC). The circuit is actually etched onto a tiny wafer or "chip" of silicon, embedded in a black plastic body, which is properly referred to as the "package." Tiny wires inside the package link the circuit with the two rows of pins on either side. Throughout this book, I will use the word "chip" to refer to the whole object, including its pins, as this is the most common usage.

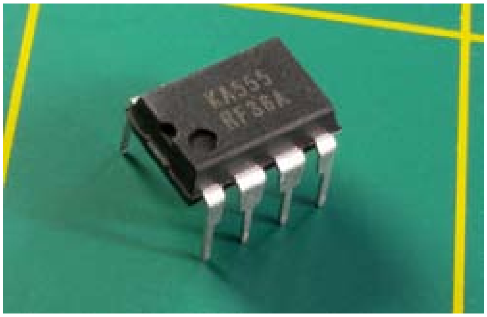

Figure 4-2. *An integrated circuit chip in Plastic Dual-Inline Pin package, abbreviated PDIP, or, more often, DIP.*

The pins are mounted at intervals of 1/10 inch in two rows spaced 3/10 inch apart. This format is known as a Plastic Dual Inline Package, abbreviated PDIP, or, more often, just DIP. The chip in the photograph has four pins in each row; others may have many more. The first thing you need to know, when shopping for chips, is that you'll only be using the DIP package. This book will not be featuring the more modern type, known as "surface-mount," because they're much smaller, more difficult to handle, and require special tools that are relatively expensive. Figure 4-3 shows a size comparison between a 14-pin DIP package and a 14-pin surface-mount package. Many surface-mount chips are even smaller than the one shown.

Just about every chip has a part number printed on it. In Figure 4-2, the part number is KA555. In Figure 4-3, the DIP chip's part number is M74HC00B1, and the surface-mount chip is a 74LVC07AD. You can ignore the second line of numbers and/or letters on each chip, as they are not part of the part number.

Notice in Figure 4-3 that even though the chips look quite different from each other, they both have "74" in their part numbers. This is because both of them are members of the "7400" family of logic chips, which originally had part numbers from 7400 and upward (7400, 7401, 7402, 7403, and so on). Often they are now referred to as "74xx" chips, where "xx" includes all the members of the family. I'll be using this family a lot, so you need to know how to buy them. I'll give you some advice on that without going into details yet about what the chips actually do.

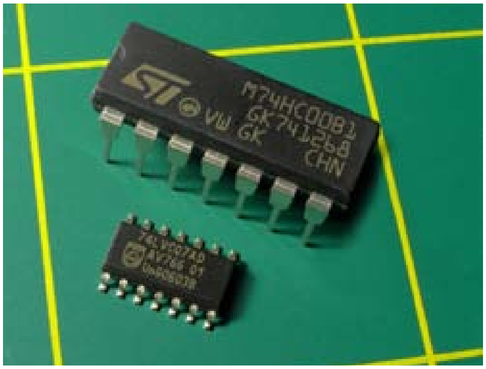

Figure 4-3. *The DIP chip, at the rear, has pins spaced 1/10 inch apart, suitable for insertion in a breadboard or perforated board. It can be soldered without special tools. The small-outline integrated circuit (SOIC) surface-mount chip (foreground) has solder tabs spaced at 1/20 inch. Other surface-mount chips have pins spaced at 1/40 inch or even less (these dimensions are often expressed in millimeters). Surface-mount chips are designed primarily for automated assembly and are difficult to work with manually. In this photo, the yellow lines are 1 inch apart to give you an idea of the scale.*

Take a look at Figure 4-4, which shows how to interpret a typical part number in a 74xx family member. The initial letters identify the manufacturer (which you can ignore, as it really makes no difference for our purposes). Skip the letters until you get to the "74." After that, you find two more letters, which are important. The 74xx family has evolved through many generations, and the letter(s) inserted after the "74" tell you which generation you're dealing with. Some generations have included:

- 74L
- 74LS
- 74C
- 74HC
- 74AHC

And there are more. Generally speaking, subsequent generations tend to be faster or more versatile than previous generations. In this book, for reasons I'll explain later, we are mostly using the HC generation.

FUNDAMENTALS

Choosing chips (continued)

After the letters identifying the generation, you'll find two (sometimes more) numerals. These identify the specific function of the chip. You can ignore any remaining letters and numerals. Looking back at Figure 4-3, the DIP chip part number, M74HC00B1 tells you that it is a chip in the 74xx family, HC generation, with its function identified by numerals 00. The surface-mount chip number, 74LVC07AD, tells you that it is in the 74xx family, LVC generation, with function identified by numerals 07. For convenience we could refer to the first chip as a "74HC00" and the second chip as a "74HC07" because, regardless of their different manufacturers and package sizes, the fundamental behavior of the circuit inside remains the same.

The purpose of this long explanation is to enable you to interpret catalog listings when you go chip shopping. You can search for "74HC00" and the online vendors are usually smart enough to show you appropriate chips from multiple manufacturers, even though there are letters preceding and following the term that you're searched for.

Suppose a circuit requires a 74HC04 chip. If you search for "74HC04" on the website of a parts supplier, you may find versions such as the CD74HC04M96 by Texas Instruments, the 74HC04N by NXP Semiconductors, or MM74HC04N by Fairchild Semiconductor. Because they all have "74HC04" in the middle, any of them will work.

Just be careful that you buy the larger DIP-style package, not the surface-mount package. If the part number has an "N" on the end, you can be sure that it's a DIP package. If there is no "N" on the end, it may or may not be a DIP package, and you will have to check a photo or additional description to make sure. If the part number begins with SS, SO, or TSS, it's absolutely definitely surface-mount, and you don't want it. Many catalogs show photographs of the chips to assist you in buying the right package style.

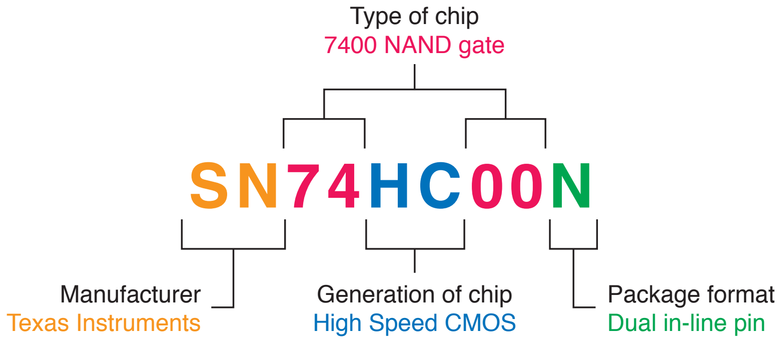

Figure 4-4. *Look for the chip family (74xx, in this case) with the correct generation (HC, in this case) embedded in the number. Make sure you buy the DIP version, not the surface-mount version. The manufacturer is unimportant.*

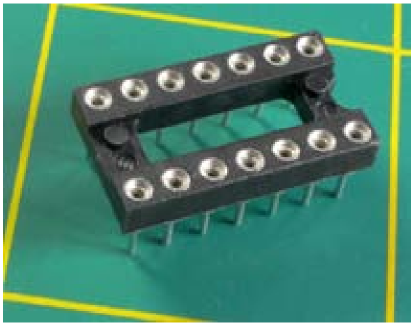

Figure 4-5. *When you're soldering a circuit onto perforated board, sockets eliminate the risk of overheating integrated circuit chips and reduce the risk of zapping them with static electricity, and enable easy replacement.*

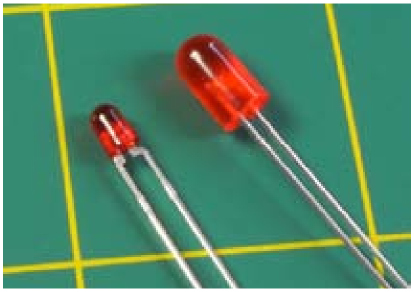

Figure 4-6. *An HC series logic chip is rated to deliver only 4mA at each pin. This is inadequate to drive a typical 5 mm LED (right), which is rated for 20mA forward current. Miniature, low-current LEDs (left) will use as little as 1mA in series with a suitable resistor, and are ideal for test circuits in which you want to see the output with a minimum of hassle.*

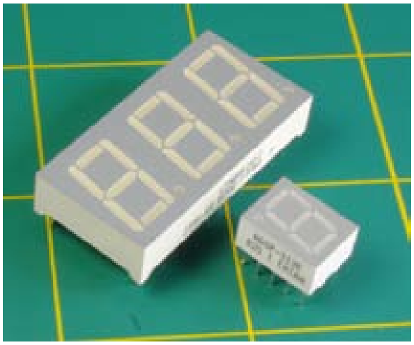

Figure 4-7. *Seven-segment displays are the simplest way to show a numeric output and can be driven directly by some CMOS chips. For finished projects, they are typically mounted behind transparent red acrylic plastic panels.*

Here's your chip list:

- 555 Timer. STMicroelectronics SA555N, Fairchild NE555D, RadioShack TLC555, or similar. Do not get the "CMOS" version of this chip, or any fancy versions such as those of high precision. Buy the cheapest you can find. Quantity: 10. The chip in Figure 4-2 is a 555 timer.
- Logic chips types 74HC00, 74HC02, 74HC04, 74HC08, 74HC32, and 74HC86. Actual part numbers could be M74HC00B1, M74HC02B1, M74HC04B1, and so on, by STMicroelectronics, or SN74HC00N, SN74HC02N, SN74HC04N, and so on, by Texas Instruments. Any other manufacturers are acceptable.
- Remember, each part number should have "HC" in the middle of it, and you want the DIP or PDIP package, not surface-mount. Quantity of each: At least 4.
- 4026 Decade Counter (a chip that counts in tens). Texas Instruments CD-4026BE or similar. Quantity: 4 (you'll need 3, but because this is a CMOS chip sensitive to static electricity, you should have at least one in reserve). Any chip with "4026" in its part number should be OK.
- 74LS92 counter chip, 74LS06 open-collector inverter chip, and 74LS27 triple-input NOR chip. Quantity: 2 of each. Note the "LS" in these part numbers! There will be one experiment in which I want you to use the LS generation instead of the HC generation.

IC sockets

I suggest that you avoid soldering chips directly onto perforated board. If you damage them, they're difficult to remove. Buy some DIP sockets, solder the sockets onto the board, and then plug the chips into the sockets. You can use the cheapest sockets you can find (you don't need gold-plated contacts for our purposes). You will need 8-pin, 14-pin, and 16-pin DIP sockets, such as parts 276-1995, 276-1999, and 276-1998 from RadioShack. See Figure 4-5. Quantity of each: 5 minimum.

Low-power LEDs

The logic chips that you'll be using are not designed to deliver significant power. You'll need to add transistors to amplify their output if you want to drive bright LEDs or relays. Because adding transistors is a hassle, I suggest an alternative: Special low-power LEDs that will draw as little as 1mA, such as the Everlight model T-100 Low Current Red, part number HLMPK150. Figure 4-6 shows a size comparison with a regular 5mm LED. Quantity: 10 (at a minimum).

LED numeric displays

In at least one of our projects, you'll want to illuminate some seven-segment LED numerals. You'll need either three individual numerals, or one package containing three numerals, such as the Kingbright High Efficiency Red Diffused, part number BC56-11EWA, which will be specifically referred to in schematics in this book. If you buy a different seven-segment display, it must be an LED with a "common cathode." (Don't buy liquid-crystal LCD numerals; they require different electronics to drive them.) If you have a choice of power consumption, buy whichever product consumes the least current. See Figure 4-7.

Latching relays

You're going to need a 5-volt latching relay that has two coils, instead of one. The first coil flips the relay one way; the second coil flips it back. The relay consumes no additional power while remaining passively in each state. I suggest the Panasonic DS2E-SL2-DC5V relay. If you buy a different relay, it must be dual-coil latching to run off 5 volts DC, switching at least 1 amp, in a "2 form C" package, to fit your breadboard.

Potentiometers

You'll need 5K, 10K, and 100K linear potentiometers (one of each). Also, a 10K trimmer potentiometer (which you may find described just as a "trimmer"). The manufacturer is unimportant.

Voltage regulators

Because many logic chips require precisely 5 volts DC, you need a voltage regulator to deliver this. The LM7805 does the job. Here again, the chip number will be preceded or followed with an abbreviation identifying the manufacturer and package style, as in the LM7805CT from Fairchild. Any manufacturer will do, but the package should look like the one in Figure 4-8, and if you have a choice, buy a regulator that can deliver at least 1 amp.

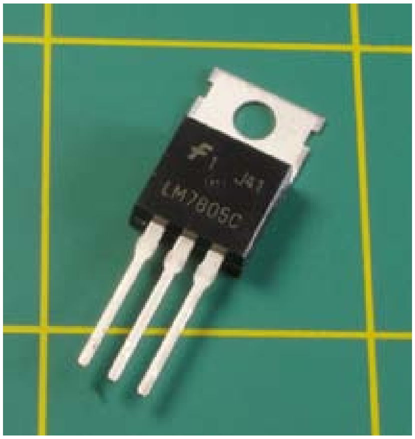

Figure 4-8. *Many integrated circuit chips require a controlled power supply of 5 volts, which can be delivered by this regulator when you apply 7.5 to 9 volts to it. The lefthand pin is for positive input, the center pin is a common ground, and the righthand pin is the 5V output. For currents exceeding 250mA, you should bolt the regulator to a metal heat sink using the hole at the top.*

Tactile switches

These are SPST pushbuttons (momentary switches), usually with four legs. Look for the ALPS part number SKHHAKA010 or any similar item that has pins to fit your breadboard or perforated board. See Figure 4-9.

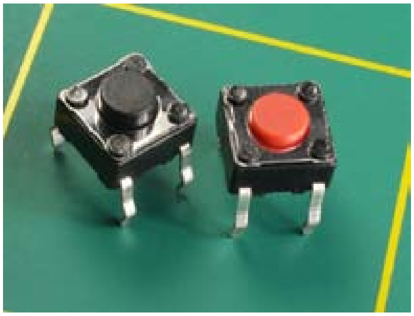

Figure 4-9. *A tactile switch delivers tactile feedback through your fingertip when you press it. They are almost always SPST pushbuttons designed for mounting in circuit boards with standard 1/10-inch hole spacing.*

12-key numeric keypad

Velleman "12 keys keyboard with common output" (no part number, but has been available through All Electronics under catalog code KP-12). Quantity: 1.

This type of keypad has the same layout as an old-fashioned touchtone phone. It should have at least 13 pins or contacts, 12 of which connect with individual pushbuttons, the thirteenth connecting with the other side of all the pushbuttons. In other words, the last pin is "common" to all of them, and this type of keypad is often described as having a "common output." The type of keypad that you don't want is "matrix-encoded," with fewer than 13 contacts, requiring additional external circuitry. See Figures 4-10 and 4-11. If you can't find the Velleman keypad that I suggest, look carefully at keypad descriptions and photographs to make sure that the one you buy is not matrix-encoded and has a common terminal.

Alternatively, you may substitute 12 cheap SPST NO pushbuttons and mount them in a small project box.

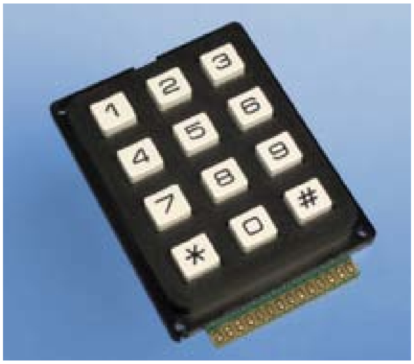

Figure 4-10. *When shopping for a numeric keypad, it should have 12 keys in "touch-tone phone" layout, and should have at least 13 contacts for input/output. The contacts are visible here along the front edge.*

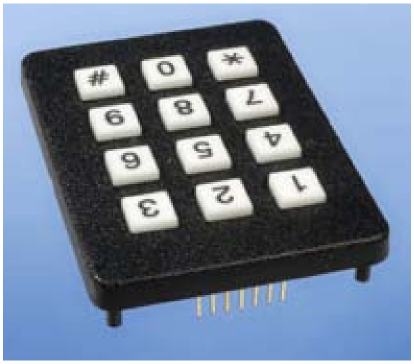

Figure 4-11. *This keypad has insufficient pins and will not work in the circuit in this book.*

BACKGROUND

How chips came to be

The concept of integrating solid-state components into one little package originated with British radar scientist Geoffrey W. A. Dummer, who talked about it for years before he attempted, unsuccessfully, to build one in 1956. The first true integrated circuit wasn't fabricated until 1958 by Jack Kilby, working at Texas Instruments. Kilby's version used germanium, as this element was already in use as a semiconductor. (You'll encounter a germanium diode when I deal with crystal radios in the next chapter of this book.) But Robert Noyce, pictured in Figure 4-12, had a better idea.

Born in 1927 in Iowa, in the 1950s Noyce moved to California, where he found a job working for William Shockley. This was shortly after Shockley had set up a business based around the transistor, which he had coinvented at Bell Labs.

Noyce was one of eight employees who became frustrated with Shockley's management and left to establish Fairchild Semiconductor. While he was the general manager of Fairchild, Noyce invented a silicon-based integrated circuit that avoided the manufacturing problems associated with germanium. He is generally credited as the man who made integrated circuits possible.

Early applications were for military use, as Minuteman missiles required small, light components in their guidance systems. These applications consumed almost all chips produced from 1960 through 1963, during which time the unit price fell from around $1,000 to $25 each, in 1963 dollars.

In the late 1960s, medium-scale integration chips emerged, each containing hundreds of transistors. Large-scale integration enabled tens of thousands of transistors on one chip by the mid-1970s, and today's chips can contain as many as several billion transistors.

Robert Noyce eventually cofounded Intel with Gordon Moore, but died unexpectedly of a heart attack in 1990. You can learn more about the fascinating early history of chip design and fabrication at *http://www.siliconvalleyhistorical.org*.

Figure 4-12. *This picture of Robert Noyce, late in his career, is from the Wikimedia Commons.*

Experiment 16: Emitting a Pulse

I'm going to introduce you to the most successful chip ever made: the 555 timer. As you can find numerous guides to it online, you might question the need to discuss it here, but I have three reasons for doing so:

1. It's unavoidable. You simply have to know about this chip. Some sources estimate that more than 1 billion are still being manufactured annually. It will be used in one way or another in most of the remaining circuits in this book.
2. It provides a perfect introduction to integrated circuits, because it's robust, versatile, and illustrates two functions that we'll be dealing with later: comparators and a flip-flop.
3. After reading all the guides to the 555 that I could find, beginning with the original Fairchild Semiconductor data sheet and making my way through various hobby texts, I concluded that its inner workings are seldom explained very clearly. I want to give you a graphic understanding of what's happening inside it, because if you don't have this, you won't be in a good position to use the chip creatively.

You will need:

- 9-volt power supply.
- Breadboard, jumper wires, and multimeter.
- 5K linear potentiometer. Quantity: 1.
- 555 timer chip. Quantity: 1.
- Assorted resistors and capacitors.
- SPST tactile switches. Quantity: 2.
- LED (any type). Quantity: 1.

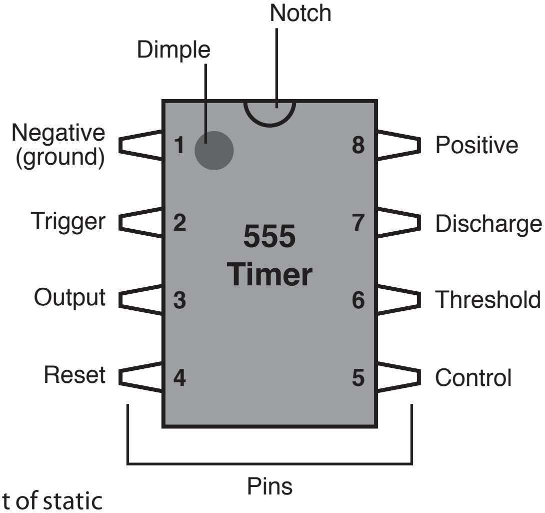

Figure 4-13. *The 555 timer chip, seen from above. Pins on chips are always numbered counterclockwise, from the top-left corner, with a notch in the body of the chip uppermost, or a circular indentation at top-left, to remind you which end is up.*

Procedure

The 555 chip is very robust, but still, in theory, you can zap it with a jolt of static electricity and kill it. Therefore, to be on the safe side, you should ground yourself before handling it. See the "Grounding yourself" warning on page 172 for details. Although this warning primarily refers to the type of chips known as CMOS, which are especially vulnerable, grounding yourself is always a sensible precaution.

Look for a small circular indentation, called the *dimple*, molded into the body of the chip, and turn the chip so that the indentation is at the top-left corner with the pins pointing down. Alternatively, if your chip is of the type with a notch at one end, turn the chip so that the notch is at the top.

The pins on chips are always numbered counterclockwise, starting from the top-left pin (next to the dimple). See Figure 4-13, which also shows the names of the pins on the 555 timer, although you don't need to know most of them just yet.

Insert the chip in your breadboard so that its pins straddle the channel down the center. Now you can easily feed voltages to the pins on either side, and read signals out of them. See Figure 4-14 for a precise guide to placement, in the first project. The timer is identified as "IC1," because "IC" is the customary abbreviation for "Integrated Circuit."

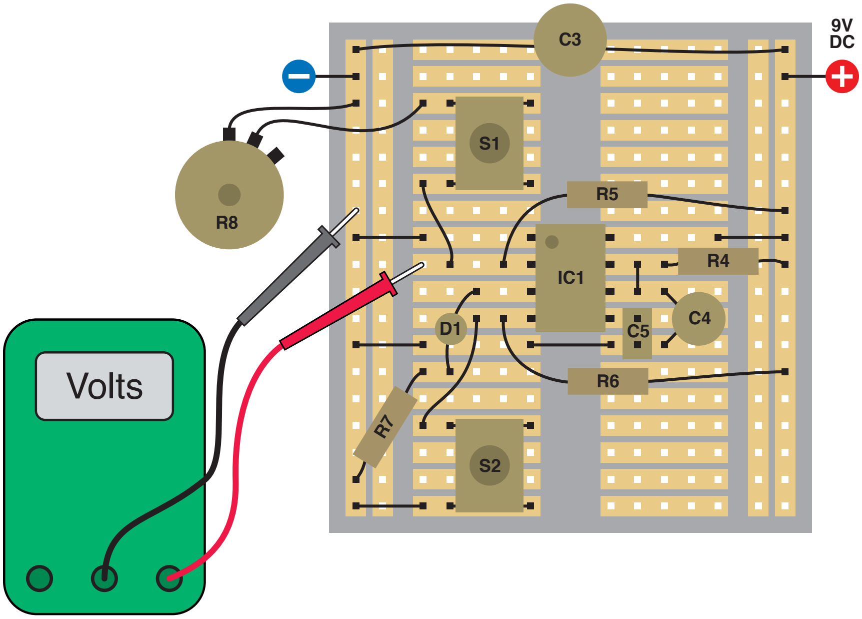

Figure 4-14. *This circuit allows you to explore the behavior of the 555 timer chip. Use your meter to monitor the voltage on pin 2 as shown. There are no resistors labeled R1, R2, or R3 and no capacitors labeled C1 or C2, because they'll be added in a later schematic. Component values in this schematic:*

R4: 100K
R5: 2K2
R6: 10K
R7: 1K
R8: 5K linear potentiometer
C3: 100 µF electrolytic
C4: 47 µF electrolytic
C5: 0.1 µF ceramic
IC1: 555 timer
S1, S2: SPST tactile switches (pushbuttons)
D1: Generic LED

R5 holds the trigger (pin 2) positive until S1 is pressed, which lowers the voltage depending on the setting of potentiometer R8. When the trigger voltage falls below 1/3 of the power supply, the chip's output (pin 3) goes high for a period determined by the values of R4 and C4. S2 resets (zeros) the timer, by reducing the voltage to pin 4, the Reset. C3 smoothes the power supply, and C5 isolates pin 5, the control, so that it won't interfere with the functioning of this test circuit. (We'll use the control pin in a future experiment.)

All integrated circuit chips require a power supply. The 555 is powered with negative voltage applied to pin 1 and positive to pin 8. If you reverse the voltage accidentally, this can permanently damage the chip, so place your jumper wires carefully.

Set your power supply to deliver 9 volts. It will be convenient for this experiment if you supply positive down the righthand side and negative down the lefthand side of the breadboard, as suggested in Figure 4-14. C3 is a large capacitor, at least 100 µF, which is placed across the power supply to smooth it out and provide a local store of charge to fuel fast-switching circuits, as well as to guard against other transient dips in voltage. Although the 555 isn't especially fast-switching, other chips are, and you should get into the habit of protecting them.

Begin with the potentiometer turned all the way counterclockwise to maximize the resistance between the two terminals that we're using, and when you apply the probe from your meter to pin 2, you should measure about 6 volts when you press S1.

Now rotate the potentiometer clockwise and press S1 again. If the LED doesn't light up, keep turning the potentiometer and pressing and releasing the button. When you've turned the potentiometer about two-thirds of the way, you should see the LED light up for just over 5 seconds when you press and release the button. Here are some facts that you should check for yourself:

- The LED will keep glowing after you release the button.
- You can press the button for any length of time (less than the timer's cycle time) and the LED always emits the same length of pulse.
- The timer is triggered by a fall in voltage on pin 2. You can verify this with your meter.
- The LED is either fully on or fully off. You can't see a faint glow when it's off, and the transition from off to on and on to off is very clean and precise.

Check Figure 4-16 to see how the components should look on your breadboard, and then look at the schematic in Figure 4-15 to understand what's happening. I will be adding more components later, which I will be labeling R1, R2, C1, and C2 to be consistent with data sheets that you may see for the 555 timer. Therefore, in this initial circuit the resistors are labeled R4 and up, and capacitors C3 and up.

When S1 (the tactile switch) is open, pin 2 of the 555 timer receives positive power through R5, which is 2K2. Because the input resistance of the timer is very high, the voltage on pin 2 is almost the full 9 volts.

When you press the button, it connects negative voltage through R8, the 5K potentiometer to pin 2. Thus, R8 and R5 form a voltage divider with pin 2 in the middle. You may remember this concept from when you were testing transistors. The voltage between the resistances will change, depending on the values of the resistances.

If R8 is turned up about halfway, it is approximately equal to R5, so the midpoint, connected to pin 2, has about half the 9-volt power supply. But when you turn the potentiometer so that its resistance falls farther, the negative voltage outweighs the positive voltage, so the voltage on pin 2 gradually drops.

If you have clips on your meter leads, you can hook them onto the nearest jumper wires and then watch the meter while you turn the potentiometer up and down and press the button.

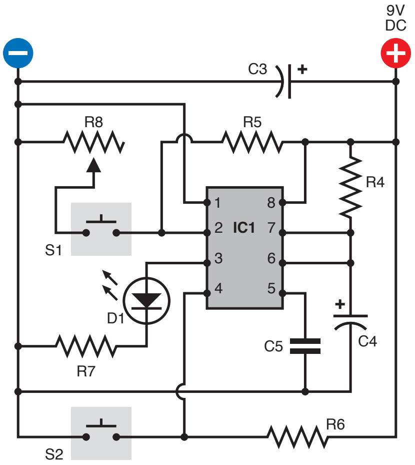

Figure 4-15. *A schematic view of the circuit shown in Figure 4-14. Throughout this chapter, the schematics will be laid out to emulate the most likely placement of components on a breadboard. This is not always the simplest layout, but will be easiest for you to build. Refer to Figure 4-14 for the values of the components.*

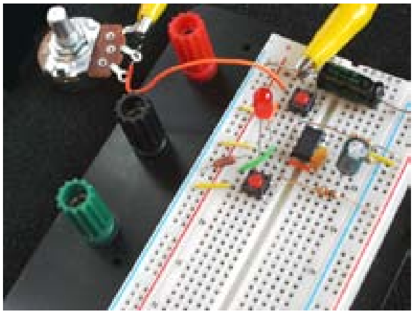

Figure 4-16. *This is how the components look when installed on the breadboard. The alligator clips are attached to a patch cord that links the 100 µF capacitor to the potentiometer. The power supply input is not shown.*

The graphs in Figure 4-17 illustrate what is happening. The upper graph shows the voltage applied to pin 2 by random button-presses, with the potentiometer turned to various values. The lower graph shows that the 555 is triggered if, and only if, the voltage on pin 2 actively drops from above 3 volts to below 3 volts. What's so special about 3 volts? It's one-third of our 9-volt power supply.

Here's the take-home message:

- The output of the 555 (pin 3) emits a *positive* pulse when the trigger (pin 2) drops *below* one-third of the supply voltage.
- The 555 delivers the *same duration* of positive pulse every time (so long as you don't supply a prolonged low voltage on pin 2).
- A *larger* value for R4 or for C4 will *lengthen* the pulse.
- When the output (pin 3) is high, the voltage is almost equal to the supply voltage. When the output goes low, it's almost zero.

The 555 converts the imperfect world around it into a precise and dependable output. It doesn't switch on and off absolutely instantly, but is fast enough to *appear* instant.

Now here's another thing to try. Trigger the timer so that the LED lights up. While it is illuminated, press S2, the second button, which grounds pin 4, the reset. The LED should go out immediately.

When the reset voltage is pulled *low*, the output goes *low*, regardless of what voltage you apply to the trigger.

There's one other thing I want you to notice before we start using the timer for more interesting purposes. I included R5 and R6 so that when you first switch on the timer, it is not emitting a pulse—but is ready to do so. These resistors apply a positive voltage to the trigger and the reset pin, to make sure that the 555 timer is ready to run when you first apply power to it.

As long as the *trigger* voltage is *high*, the timer *will not* emit a pulse. (It emits a pulse when the trigger voltage drops.)

As long as the *reset* voltage is *high*, the timer is *able* to emit a pulse. (It shuts down when the reset voltage drops.)

R5 and R6 are known as "pull-up resistors" because they pull the voltage up. You can easily overwhelm them by adding a direct connection to the negative side of the power supply. A typical pull-up resistor for the 555 timer is 10K. With a 9-volt power supply, it only passes 0.9mA (by Ohm's Law).

Finally, you may be wondering about the purpose of C5, attached to pin 5. This pin is known the "control" pin, which means that if you apply a voltage to it, you can control the sensitivity of the timer. I'll get to this in more detail a little later. Because we are not using this function right now, it's good practice to put a capacitor on pin 5 to protect it from voltage fluctuations and prevent it from interfering with normal functioning.

Make sure you become familiar with the basic functioning of the 555 timer before you continue.

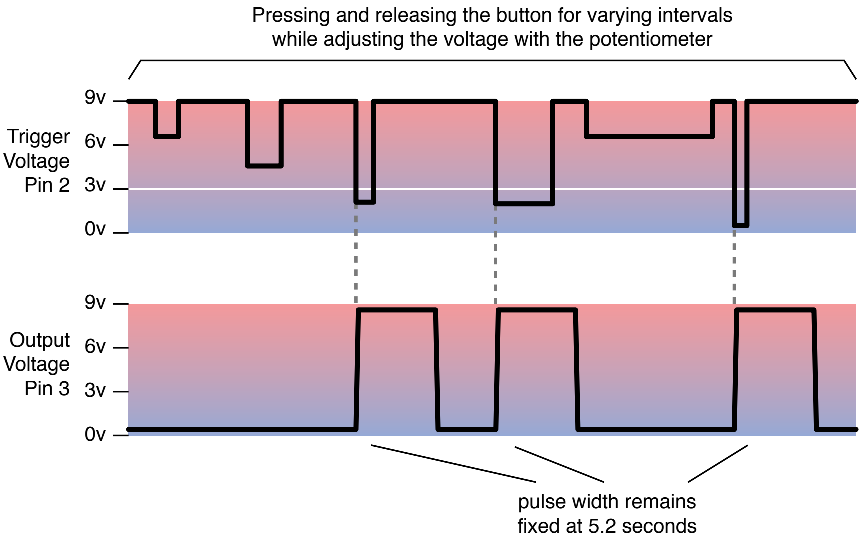

Figure 4-17. *The top graph shows voltage on the trigger (pin 2) when the pushbutton is pressed, for different intervals, at different settings of the potentiometer. The lower graph shows the output (pin 3), which rises until it is almost equal to the power supply, when the voltage on pin 2 drops below 1/3 the full supply voltage.*

FUNDAMENTALS

The following table shows 555 pulse duration in monostable mode:

- Duration is in seconds, rounded to two figures.
- The horizontal scale shows common resistor values between pin 7 and positive supply voltage.
- The vertical scale shows common capacitor values between pin 6 and negative supply voltage.

To calculate a different pulse duration, multiply resistance × capacitance × 0.0011 where resistance is in kilohms, capacitance is in microfarads, and duration is in seconds.

47 µF	0.05	0.11	0.24	0.52	1.1	2.4	5.2	11	24	52
22 µF	0.02	0.05	0.11	0.24	0.53	1.1	2.4	5.3	11	24
10 µF	0.01	0.02	0.05	0.11	0.24	0.52	1.1	2.4	5.2	11
4.7 µF		0.01	0.02	0.05	0.11	0.24	0.52	1.1	2.4	5.2
2.2 µF			0.01	0.02	0.05	0.11	0.24	0.53	1.1	2.4
1.0 µF				0.01	0.02	0.05	0.11	0.24	0.52	1.1
0.47 µF					0.01	0.02	0.05	0.11	0.24	0.52
0.22 µF						0.01	0.02	0.05	0.11	0.24
0.1 µF							0.01	0.02	0.05	0.11
0.047 µF								0.01	0.02	0.05
0.022 µF									0.01	0.02
0.01 µF										0.01
	1K	2K2	4K7	10K	22K	47K	100K	220K	470K	1M

THEORY

Inside the 555 timer: monostable mode

The plastic body of the 555 timer contains a wafer of silicon on which are etched dozens of transistor junctions in a pattern that is far too complex to be explained here. However, I can summarize their function by dividing them into groups, as shown in Figure 4-18. An external resistor and two external capacitors are also shown, labeled the same way as in Figure 4-15.

The negative and positive symbols inside the chip are power sources which actually come from pins 1 and 8, respectively. I omitted the internal connections to those pins for the sake of clarity.

The two yellow triangles are "comparators." Each comparator compares two inputs (at the base of the triangle) and delivers an output (from the apex of the triangle) depending on whether the inputs are similar or different. We'll be using comparators for other purposes later in this book.

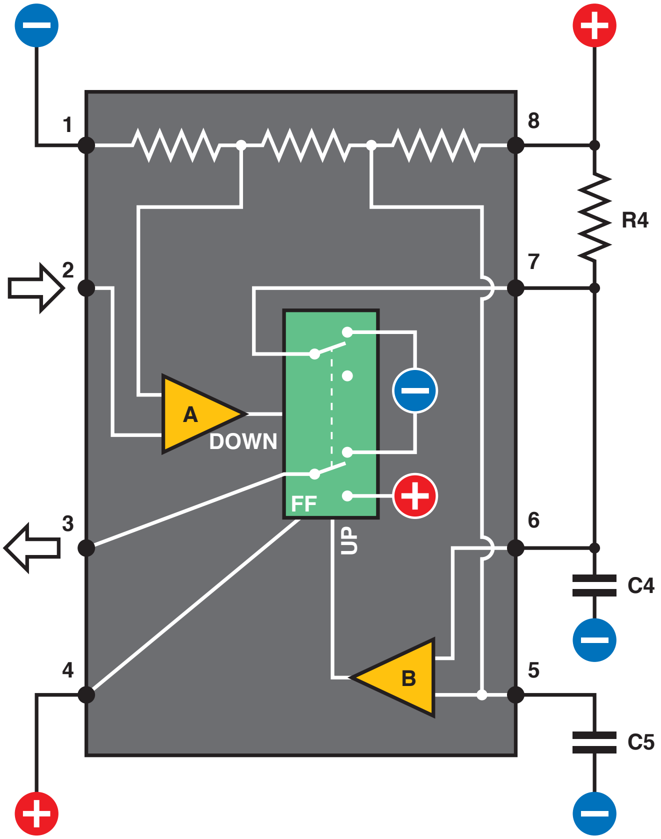

Figure 4-18. *Inside the 555 timer. White lines indicate connections inside the chip. A and B are comparators. FF is a flip-flop which can rest in one state or the other, like a double-throw switch. A drop in voltage on pin 2 is detected by comparator A, which triggers the flip-flop into its "down" position and sends a positive pulse out of pin 3. When C4 charges to 2/3 of supply voltage, this is detected by comparator B, which resets the flip-flop to its "up" position. This discharges C4 through pin 7.*

THEORY

Inside the 555 timer: monostable mode (continued)

The green rectangle, identified as "FF," is a "flip-flop." I have depicted it as a DPDT switch, because that's how it functions here, although of course it is really solid-state.

Initially when you power up the chip, the flip-flop is in its "up" position which delivers low voltage through the output, pin 3. If the flip-flop receives a signal from comparator A, it flips to its "down" state, and flops there. When it receives a signal from comparator B, it flips back to its "up" state, and flops there. The "UP" and "DOWN" labels on the comparators will remind you what each one does when it is activated.

Flip-flops are a fundamental concept in digital electronics. Computers couldn't function without them.

Notice the external wire that connects pin 7 with capacitor C4. As long as the flip-flop is "up," it sinks the positive voltage coming through R4 and prevents the capacitor from charging positively.

If the voltage on pin 2 drops to 1/3 of the supply, comparator A notices this, and flips the flip-flop. This sends a positive pulse out of pin 3, and also disconnects the negative power through pin 7. So now C4 can start charging through R4. While this is happening, the positive output from the timer continues.

As the voltage increases on the capacitor, comparator B monitors it through pin 6, known as the threshold. When the capacitor accumulates 2/3 of the supply voltage, comparator B sends a pulse to the flip-flop, flipping it back into its original state. This discharges the capacitor through pin 7, appropriately known as the discharge pin. Also, the flip-flop ends the positive output through pin 3 and replaces it with a negative voltage. This way, the 555 returns to its original state.

I'll sum up this sequence of events very simply:

1. Initially, the flip-flop grounds the capacitor and grounds the output (pin 3).
2. A drop in voltage on pin 2 to 1/3 the supply voltage or less makes the output (pin 3) positive and allows capacitor C4 to start charging through R4.
3. When the capacitor reaches 2/3 of supply voltage, the chip discharges the capacitor, and the output at pin 3 goes low again.

In this mode, the 555 timer is "monostable," meaning that it just gives one pulse, and you have to trigger it again to get another.

You adjust the length of each pulse by changing the values of R4 and C4. How do you know which values to choose? Check the table on page 157, which gives an approximate idea and also includes a formula so that you can calculate values of your own.

I didn't bother to include pulses shorter than 0.01 second in the table, because a single pulse of this length is usually not very useful. Also I rounded the numbers in the table to 2 significant figures, because capacitor values are seldom more accurate than that.

BACKGROUND

How the timer was born

Back in 1970, when barely a half-dozen corporate seedlings had taken root in the fertile ground of Silicon Valley, a company named Signetics bought an idea from an engineer named Hans Camenzind. It wasn't a huge breakthrough concept—just 23 transistors and a bunch of resistors that would function as a programmable timer. The timer would be versatile, stable, and simple, but these virtues paled in comparison to its primary selling point. Using the emerging technology of integrated circuits, Signetics could reproduce the whole thing on a silicon chip.

This entailed some trial and error. Camenzind worked alone, building the whole thing initially on a large scale, using off-the-shelf transistors, resistors, and diodes on a breadboard. It worked, so then he started substituting slightly different values for the various components to see whether the circuit would tolerate variations during production and other factors such as changes in temperature when the chip was in use. He made at least 10 different versions of the circuit. It took months.

Next came the crafts work. Camenzind sat at a drafting table and used a specially mounted X-Acto knife to scribe his circuit into a large sheet of plastic. Signetics then reduced this image photographically by a ratio of about 300:1. They etched it into tiny wafers, and embedded each of them in a half-inch rectangle of black plastic with the product number printed on top. Thus, the 555 timer was born.

It turned out to be the most successful chip in history, both in the number of units sold (tens of billions and counting) and the longevity of its design (unchanged in almost 40 years). The 555 has been used in everything from toys to spacecraft. It can make lights flash, activate alarm systems, put spaces between beeps, and create the beeps themselves.

Today, chips are designed by large teams and tested by simulating their behavior using computer software. Thus, chips inside a computer enable the design of more chips. The heyday of solo designers such as Hans Camenzind is long gone, but his genius lives inside every 555 timer that emerges from a fabrication facility. (If you'd like to know more about chip history, see *http://www.semiconductormuseum.com/Museum_Index.htm*.)

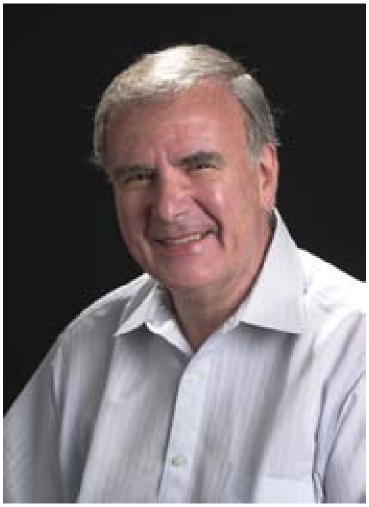

Figure 4-13. *Hans Camenzind, inventor and developer of the 555 timer chip for Signetics.*

FUNDAMENTALS

Why the 555 is useful

In its monostable mode (which is what you just saw), the 555 will emit a single pulse of fixed (but programmable) length. Can you imagine some applications? Think in terms of the pulse from the 555 controlling some other component. A motion sensor on an outdoor light, perhaps. When an infra-red detector "sees" something moving, the light comes on for a specific period—which can be controlled by a 555.

Another application could be a toaster. When someone lowers a slice of bread, a switch will close that triggers the toasting cycle. To change the length of the cycle, you could use a potentiometer instead of R4 and attach it to the external lever that determines how dark you want your toast. At the end of the toasting cycle, the output from the 555 would pass through a power transistor, to activate a solenoid (which is like a relay, except that it has no switch contacts) to release the toast.

Intermittent windshield wipers could be controlled by a 555 timer—and on earlier models of cars, they actually were.

And what about the burglar alarm that was described at the end of Chapter 3? One of the features that I listed, which has not been implemented yet, is that it should shut itself off after a fixed interval. We can use the change of output from a 555 timer to do that.

The experiment that you just performed seemed trivial, but really it implies all kinds of possibilities.

555 timer limits

1. The timer can run from a stable voltage source ranging from 5 to 15 volts.
2. Most manufacturers recommend a range from 1K to 1M for the resistor attached to pin 7.
3. The capacitor value can go as high as you like, if you want to time really long intervals, but the accuracy of the timer will diminish.
4. The output can deliver as much as 100mA at 9 volts. This is sufficient for a small relay or miniature loudspeaker, as you'll see in the next experiment.

Beware of Pin-Shuffling!

In all of the schematics in this book, I'll show chips as you'd see them from above, with pin 1 at top left. Other schematics that you may see, on websites or in other books, may do things differently. For convenience in drawing circuits, people shuffle the pin numbers on a chip so that pin 1 isn't necessarily shown adjacent to pin 2.

Look at the schematic in Figure 4-20 and compare it with the one in Figure 4-15. The connections are the same, but the one in Figure 4-20 groups pins to reduce the apparent complexity of the wiring.

"Pin shuffling" is common because circuit-drawing software tends to do it, and on larger chips, it is necessary for functional clarity of the schematic (i.e., logical groupings of pin names versus physical groupings on memory chips, for example). When you're first learning to use chips, I think it's easier to understand a schematic that shows the pins in their actual positions. So that's the practice I will be using here.

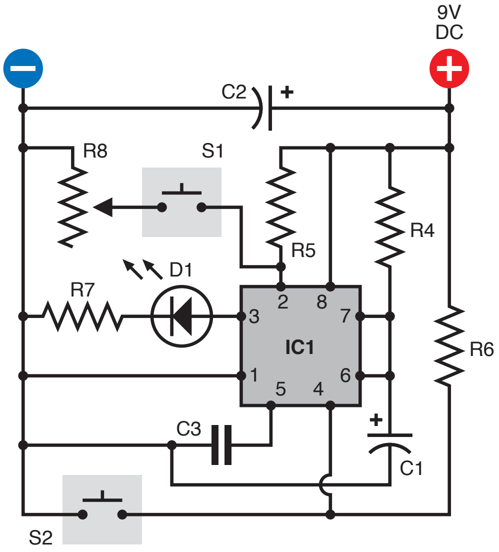

Figure 4-20. *Many people draw schematics in which the pin numbers on a chip are shuffled around to make the schematic smaller or simpler. This is not helpful when you try to build the circuit. The schematic here is for the same circuit as in Figure 4-15. This version would be harder to recreate on a breadboard.*

Experiment 17: Set Your Tone

I'm going to show you two other ways in which the 555 timer can be used.

You will need the same items as in Experiment 16, plus:

- Additional 555 timer chip. Quantity: 1.
- Miniature loudspeaker. Quantity: 1.
- 100K linear potentiometer. Quantity: 1.

Procedure

Leave the components from Experiment 16 where they are on the breadboard, and add the next section below them, as shown in Figures 4-21 and 4-22. Resistor R2 is inserted between pins 6 and 7, instead of the jumper wire that shorted the pins together in the previous circuit, and there's no external input to pin 2 anymore. Instead, pin 2 is connected via a jumper wire to pin 6. The easiest way to do this is by running the wire across the top of the chip.

I have omitted the smoothing capacitor from the schematic in Figure 4-22, because I'm assuming that you're running this circuit on the same breadboard as the first, where the previous smoothing capacitor is still active.

A loudspeaker in series with a 100Ω resistor (R3) has been substituted for the LED to show the output from the chip. Pin 4, the reset, is disabled by connecting it to the positive voltage supply, as I'm not expecting to use the reset function in this circuit.

Now what happens when you apply power? Immediately, you should hear noise through the loudspeaker. If you don't hear anything, you almost certainly made a wiring error.

Notice that you don't have to trigger the chip with a pushbutton anymore. The reason is that when C1 charges and discharges, its fluctuating voltage is connected via a jumper wire across the top of the chip to pin 2, the trigger. In this way, the 555 timer now triggers itself. I'll describe this in more detail in the next section "Theory: Inside the 555 timer: astable mode," if you want to see exactly what is going on.

In this mode, the chip is "astable," meaning that it is not stable, because it flips to and fro endlessly, sending a stream of pulses for as long as the power is connected. The pulses are so rapid that the loudspeaker reproduces them as noise.

In fact, with the component values that I specified for R1, R2, and C1, the 555 chip is emitting about 1,500 pulses per second. In other words, it creates a 1.5 KHz tone.

Check the table on page 166 to see how different values for R2 and C1 can create different pulse frequencies with the chip in this astable mode. Note that the table assumes a fixed value of 1K for R1!

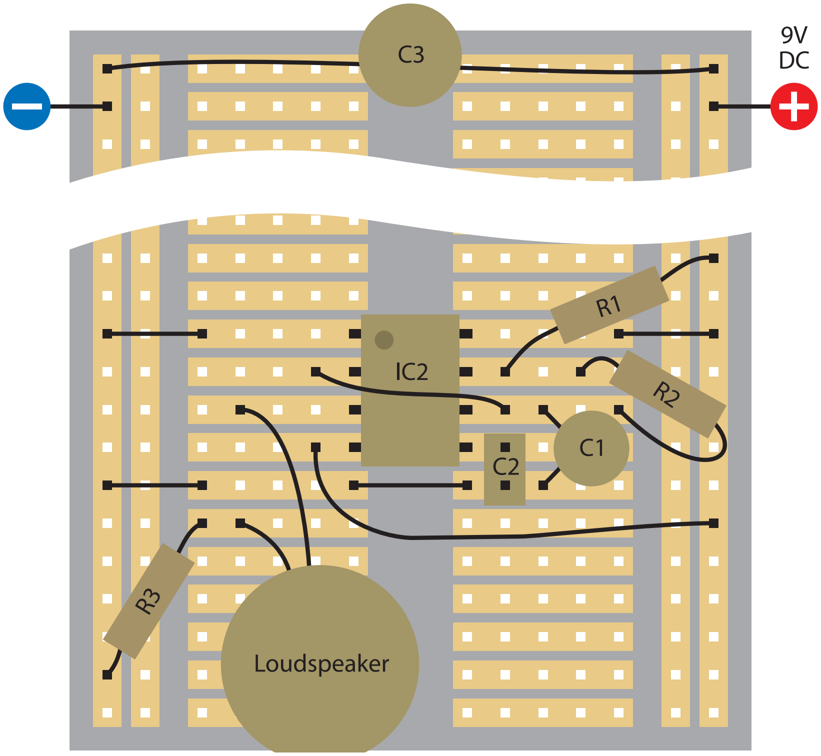
C3
9V
DC
IC2
R1
R2
C1
C2
R3
Loudspeaker

Figure 4-21. *These components should be added on the same breadboard below the components shown in Figure 4-14. Use the following values to test the 555 timer in its astable mode:*

R1: 1K
R2: 10K
R3: 100Ω
C1: 0.047 µF ceramic or electrolytic
C2: 0.1 µF ceramic
IC2: 555 timer

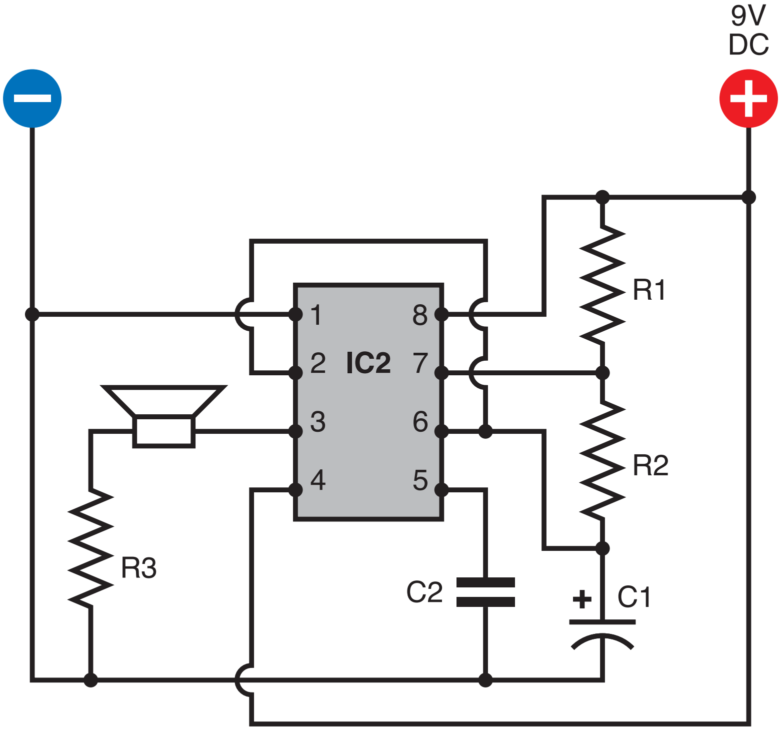
9V
DC
1
2
3
4
8
7
6
5
IC2
R1
R2
R3
C2
C1
+

Figure 4-22. *This is the schematic version of the circuit shown in Figure 4-21. The component values are the same.*

THEORY

Inside the 555 timer: astable mode

Here's what is happening now, illustrated in Figure 4-23. Initially, the flip-flop grounds C1 as before. But now the low voltage on the capacitor is connected from pin 7 to pin 2 through an external wire. The low voltage tells the chip to trigger itself. The flip-flop obediently flips to its "on" position and sends a positive pulse to the loudspeaker, while removing the negative voltage from pin 7.

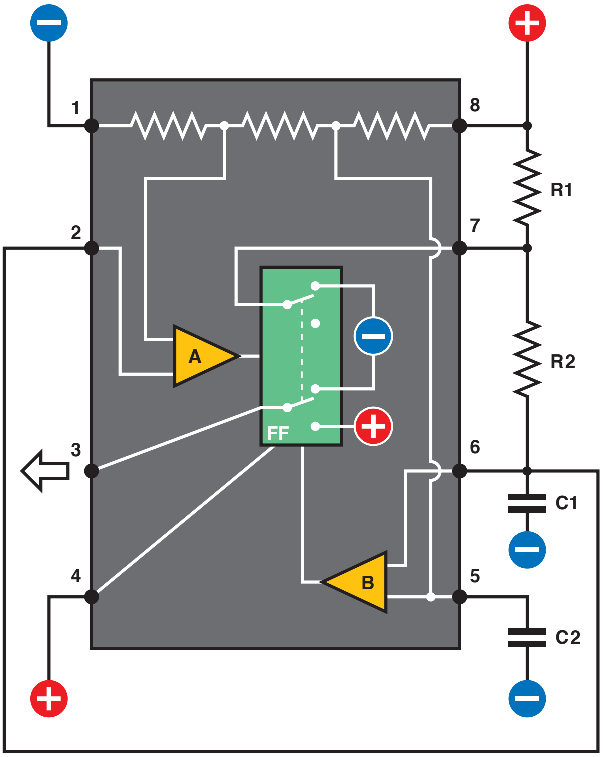

Figure 4-23. *When the 555 timer is used in astable mode, resistor R2 is placed between pin 6 and pin 7, and pin 6 is connected via an external wire to pin 2, so that the timer triggers itself.*

Now C1 starts charging, as it did when the timer was in monostable mode, except that it is being charged through R1 + R2 in series. Because the resistors have low values, and C1 is also small, C1 charges quickly. When it reaches 2/3 full voltage, comparator B takes action as before, discharging the capacitor and ending the output pulse from pin 3.

The capacitor takes longer to discharge than before, because R2 has been inserted between it and pin 7, the discharge pin. While the capacitor is discharging, its voltage diminishes, and is still linked to pin 2. When the voltage drops to 1/3 of full power or less, comparator A kicks in and sends another pulse to the flip-flop, starting the process all over again.

Summing up:

1. In astable mode, as soon as power is connected, the flip-flop pulls down the voltage on pin 2, triggering comparator A, which flips the flip-flop to its "down" position.
2. Pin 3, the output, goes high. The capacitor charges through R1 and R2 in series.
3. When the capacitor reaches 2/3 of supply voltage, the flip-flop goes "up" and the output at pin 3 goes low. The capacitor starts to discharge through R2.
4. When the charge on the capacitor diminishes to 1/3 of full voltage, the pull-down on pin 2 flips the flip-flop again and the cycle repeats.

Unequal on/off cycles

When the timer is running in astable mode, C1 charges through R1 and R2 in series. But when C1 discharges, it dumps its voltage through R2 only. This means that the capacitor charges more slowly than it discharges. While it is charging, the output on pin 3 is high; while it is discharging, the output on pin 3 is low. Consequently the "on" cycle is always longer than the "off" cycle. Figure 4-24 shows this as a simple graph.

If you want the on and off cycles to be equal, or if you want to adjust the on and off cycles independently (for example, because you want to send a very brief pulse to another chip, followed by a longer gap until the next pulse), all you need to do is add a diode, as shown in Figure 4-25.

Now when C1 charges, the electricity flows through R1 as before but takes a shortcut around R2, through diode D1. When C1 discharges, the diode blocks the flow of electricity in that direction, and so the discharge goes back through R2.

THEORY

Inside the 555 timer: astable mode (continued)

R1 now controls the charge time on its own, while R2 controls the discharge time. The formula for calculating the frequency is now:

Frequency = 1440 / ((R1 + R2) × C1)

If you set R1 = R2, you should get almost equal on/off cycles ("almost" because the diode itself imposes a small internal voltage drop of about 0.6V). The exact value depends primarily on the manufacturing process used to make the diode.

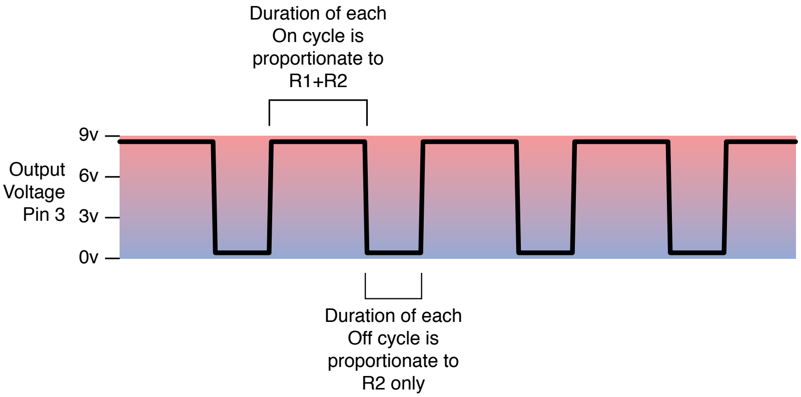

Figure 4-24. *In its usual astable configuration, the timer charges a capacitor through R1+R2 and discharges the capacitor through R2 only. Therefore its output on cycles are longer than its output off cycles.*

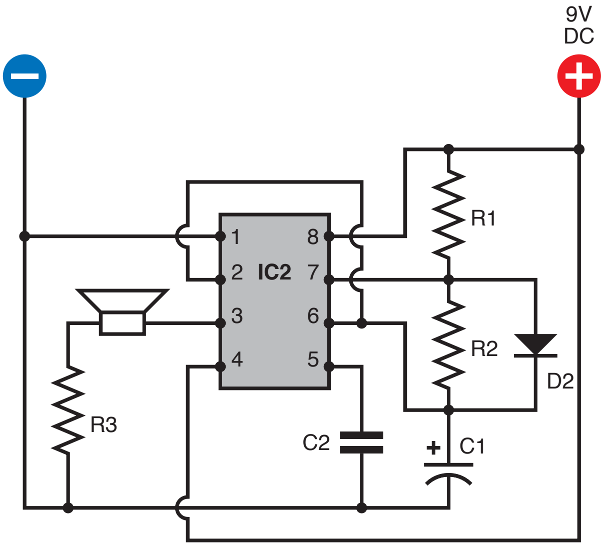

Figure 4-25. *This is a modification of the schematic shown in Figure 4-22. By adding a diode to a 555 timer running in astable mode, we eliminate R2 from the charging cycle of capacitor C1. Now we can adjust the output on cycle with the value of R1, and the output off cycle with the value of R2, so that the on and off durations are independent of each other.*

FUNDAMENTALS

The following table shows 555 timer frequency in astable mode:

- Frequency is in pulses per second, rounded to two figures.
- The horizontal scale shows common resistor values for R2.
- The vertical scale shows common capacitor values for C1. Resistor R1 is assumed to be 1K.
- Resistor R1 is assumed to be 1K.

To calculate a different frequency: double R2, add the product to R1, multiply the sum by C1, and divide the result into 1440. Like this:

Frequency = 1440 / ((R1 + 2R2) × C1) cycles per second

In this formula, R1 and R2 are in kilohms, C1 is in microfarads, and the frequency is in hertz (cycles per second). Note that the frequency is measured from the start of one pulse to the start of the next. The duration of each pulse is not the same as the length of time between each pulse. This issue is discussed in the previous section.

47 µF	10	5.7	3.0	1.5	0.7	0.3	0.2	0.1		
22 µF	22	12	6.3	3.1	1.5	0.7	0.3	0.2	0.1	
10 µF	48	27	14	6.9	3.2	1.5	0.7	0.3	0.2	0.1
4.7 µF	100	57	30	15	6.8	3.2	1.5	0.7	0.3	0.2
2.2 µF	220	120	63	31	15	6.9	3.3	1.5	0.7	0.3
1.0 µF	480	270	140	69	32	15	7.2	3.3	1.5	0.7
0.47 µF	1,000	570	300	150	68	32	15	7	3.3	1.5
0.22 µF	2,200	1,200	630	310	150	69	33	15	7	3.3
0.1 µF	4,800	2,700	1,400	690	320	150	72	33	15	7.2
0.047 µF	10,000	5,700	3,000	1,500	680	320	150	70	33	15
0.022 µF	22,000	12,000	6,300	3,100	1,500	690	330	150	70	33
0.01 µF	48,000	27,000	14,000	6,900	3,200	1,500	720	330	150	72
	1K	2K2	4K7	10K	22K	47K	100K	220K	470K	1M

Astable Modifications

In the circuits shown in Figures 4-22 or 4-25, if you substitute a 100K potentiometer for R2, you can adjust the frequency up and down by turning the shaft.

Another option is to “tune” the timer by using pin 5, the control, as shown in the Figure 4-26. Disconnect the capacitor that was attached to that pin and substitute the series of resistors shown. R9 and R11 are both 1K resistors, either side of R10, which is a 100K potentiometer. They ensure that pin 5 always has at least 1K between it and the positive and negative sides of the power supply. Connecting it directly to the power supply won’t damage the timer, but will prevent it from generating audible tones. As you turn the potentiometer to and fro, the frequency will vary over a wide range. If you want to generate a very specific frequency, a trimmer potentiometer can be used instead.

A primary advantage of using pin 5 to adjust frequency is that you can control it remotely. Take the output from pin 3 of another 555 timer running slowly in astable mode, and pipe it through a 2K2 resistor to pin 5. Now you get a two-tone siren effect, as one timer controls the other. If, in addition, you add a 100 µF capacitor between pin 5 and ground, the charging and discharging of the capacitor will make the tone slide up and down instead of switching abruptly. I'll describe this in more detail shortly. This leads me to the whole topic of one chip controlling another chip, which will be our last variation on this experiment.

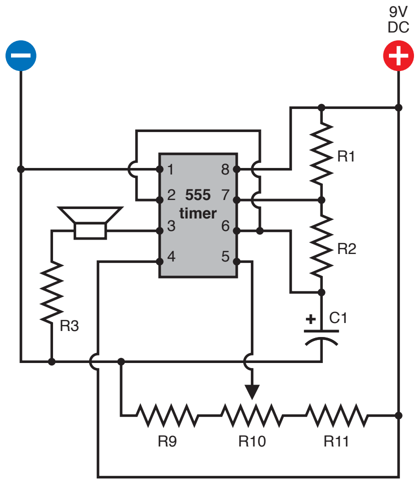
9V DC
1
2
3
4
8
7
6
5
555 timer
R1
R2
R3
+
C1
R9
R10
R11

Figure 4-28. *The control (pin 5) is seldom used but can be useful. Varying the voltage on it will adjust the speed of the timer. This circuit enables you to test the behavior of it. Component values:*

R1: 1K
R2: 10K
R3: 100 ohms
R9, R11: 1K
R10: 100K linear potentiometer
C1: 0.047 µF

Chaining Chips

Generally speaking, chips are designed so that they can talk to each other. The 555 couldn't be easier in this respect:

- Pin 3, the output, from one 555 can be connected directly to pin 2, the trigger, of a second 555.
- Alternatively, the output can be sufficient to provide power to pin 8 of a second 555.
- The output is appropriate to control or power other types of chips too.

Figure 4-27 shows these options.

When the output from the first 555 goes high, it is about 70 to 80% of its supply voltage. In other words, when you're using a 9V supply, the high output voltage is at least 6 volts. This is still above the minimum of 5V that the second chip needs to trigger its comparator, so there's no problem.

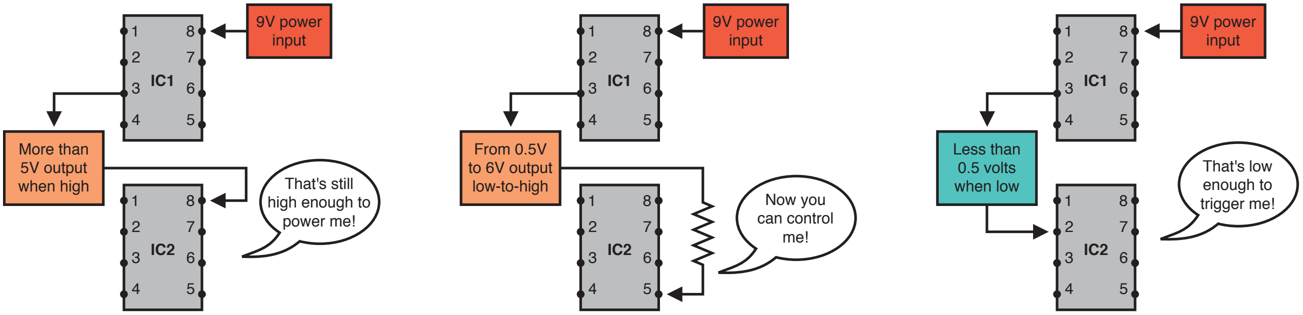
9V power input
IC1
1
2
3
4
8
7
6
5
More than 5V output when high
IC2
That's still high enough to power me!
9V power input
IC1
From 0.5V to 6V output low-to-high
IC2
Now you can control me!
9V power input
IC1
Less than 0.5 volts when low
IC2
That's low enough to trigger me!

Figure 4-27. *Three ways to chain 555 timers together. The output of IC1 can power a second timer, or adjust its control voltage, or activate its trigger pin.*

You can chain together the two 555 timers that you already have on your breadboard. Figure 4-28 shows how to connect the two circuits that were shown previously in Figures 4-15 and 4-22. Run a wire from pin 3 (the output) of the first chip to pin 8 (the positive power supply) of the second chip, and disconnect the existing wire connecting pin 8 to your power supply. The new wire is shown in red. Now when you press the button to activate the first chip, its output powers the second chip.

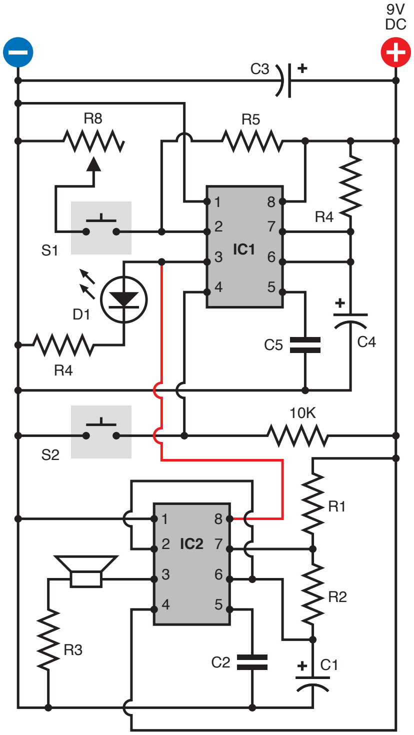

Figure 4-28. *You can combine the two circuits shown in Figures 4-15 and 4-22 simply by disconnecting the wire that provides power to pin 8 of the second timer, and running a substitute wire (shown in red).*

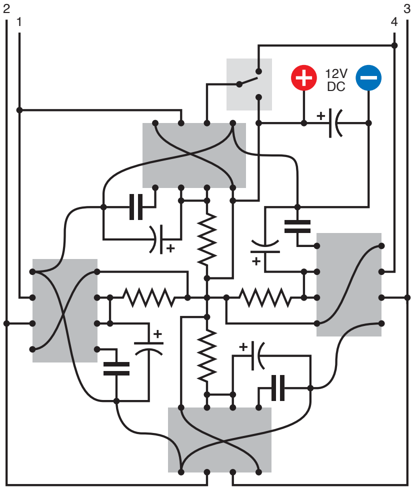

Figure 4-29. *Four 555 timers, chained together in a circle, can flash a series of four sets of LEDs in sequence, like Christmas lights or a movie marquee.*

You can also use the output from one chip to trigger another (i.e., you can connect pin 3 from the first chip to pin 2 of the second). When the output from the first chip is low, it's less than half a volt. This is well below the threshold that the second chip requires to be activated. Why would you want to do this? Well, you might want to have both timers running in monostable mode, so that the end of a high pulse from the first one triggers the start of a high pulse in the second one. In fact, you could chain together as many timers as you like in this way, with the last one feeding back and triggering the first one, and they could flash a series of LEDs in sequence, like Christmas lights. Figure 4-29 shows how four timers could be linked this way, in a configuration that would occupy minimal space (and would be wired point-to-point on perforated board, not on breadboard-format board). Each of the outputs numbered 1 through 4 would have about enough power to run maybe 10 LEDs, if you used relatively high load resistors to limit their current.

Incidentally, you can reduce the chip count (the number of chips) by using two 556 timers instead of four 555 timers. The 556 contains a pair of 555 timers in one package. But because you have to make the same number of external connections (other than the power supply), I haven't bothered to use this variant.

You can even get a 558 timer that contains four 555 circuits, all preset to function in astable mode. I decided not to use this chip, because its output behaves differently from a normal 555 timer. But you can buy a 558 timer and play with it if you wish. It is ideal for doing the "chain of four timers" that I suggested previously. The data sheet even suggests this.

Lastly, going back to the idea of modifying the frequency of a 555 timer in astable mode, you can chain two timers, as shown in Figure 4-30. The red wire shows the connection from the output of the first timer to the control pin of the second. The first timer has now been rewired in astable mode, so that it creates an oscillating on/off output around four times per second. This output flashes the LED (to give you a visual check of what's going on) and feeds through R7 to the control pin of the second timer.

But C2 is a large capacitor, which takes time to charge through R7. While this happens, the voltage detected by pin 5 slowly rises, so that the tone generated by IC2 gradually rises in pitch. Then IC1 reaches the end of its on cycle and switches itself off, at which point C2 discharges and the pitch of the sound generated by IC2 falls again.

You can tweak this circuit to create all kinds of sounds, much more controllably then when you were using PUT transistors to do the same kind of thing. Here are some options to try:

- Double or halve the value of C2.
- Omit C2 completely, and experiment with the value of R7.
- Substitute a 10K potentiometer for R7.
- Change C4 to increase or decrease the cycle time of IC1.

- Halve the value of R5 while doubling the value of C4, so that the cycle time of IC1 stays about the same, but the On time becomes significantly longer than the Off time.
- Change the supply voltage in the circuit from 9 volts to 6 volts or 12 volts.

Remember, you can't damage a 555 timer by making changes of this kind. Just make sure that the negative side of your power supply goes to pin 1 and the positive side to pin 8.

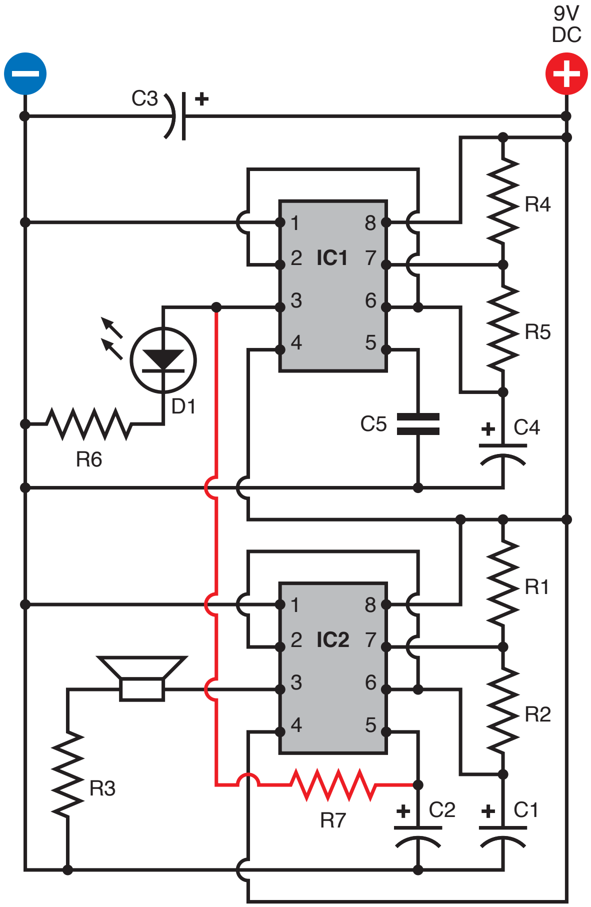

Figure 4-30. *When both timers are astable, but IC1 runs much more slowly than IC2, the output from IC1 can be used to modulate the tone generated by IC2. Note that as this is a substantial modification to the previous schematics, several components have been relabeled. To avoid errors, you may need to remove the old circuit from your breadboard and build this version from scratch. Try these values initially:*

R1, R4, R6, R7: 1K
R2, R5: 10K
R3: 100 ohms
C1: 0.047 µF
C2, C3: 100 µF
C4: 68 µF
C5: 0.1 µF

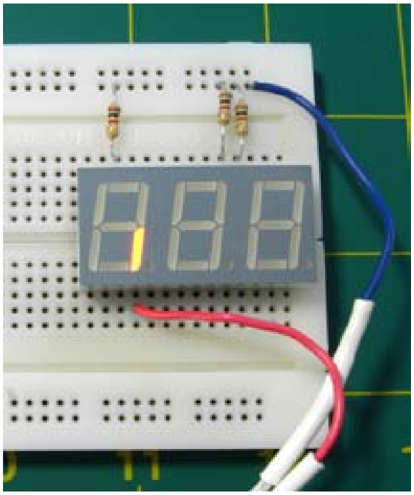

Figure 4-31. *After putting a 1K resistor between the common cathode of the display and the negative supply voltage, you can use the positive supply voltage to illuminate each segment in turn.*

Experiment 18: Reaction Timer

Because the 555 can easily run at thousands of cycles per second, we can use it to measure human reactions. You can compete with friends to see who has the fastest response—and note how your response changes depending on your mood, the time of day, or how much sleep you got last night.

Before going any further, I have to warn you that this circuit will have more connections than others you've tackled so far. It's not conceptually difficult, but requires a lot of wiring, and will only just fit on a breadboard that has 63 rows of holes. Still, we can build it in a series of phases, which should help you to detect any wiring errors as you go.

You will need:

- 4026 chip. Quantity: 4 (really you need only 3, but get another one in case you damage the others).
- 555 timers. Quantity: 3.
- Tactile switches (SPST momentary switches). Quantity: 3.
- Three numeric LEDs, or one 3-digit LED display (see the shopping list at the beginning of this chapter). Quantity: 1.
- Breadboard, resistors, capacitors, and meter, as usual.

Step 1: Display

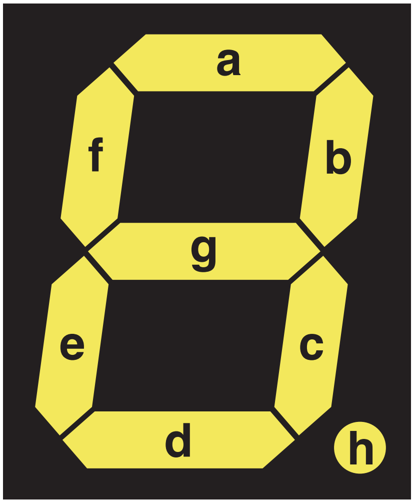

Figure 4-32. *The most basic and common digital numeral consists of seven LED segments identified by letters, as shown here, plus an optional decimal point.*

You can use three separate LED numerals for this project, but I suggest that you buy the Kingbright BC56-11EWA on the shopping list at the beginning of this chapter. It contains three numerals in one big package.

You should be able to plug it into your breadboard, straddling the center channel. Put it all the way down at the bottom of the breadboard, as shown in Figure 4-31. Don't put any other components on the breadboard yet.

Now set your power supply to 9 volts, and apply the *negative* side of it to the row of holes running up the breadboard on the *righthand* side. Insert a 1K resistor between that negative supply and each of pins 18, 19, and 26 of the Kingbright display, which are the "common cathode," meaning the negative connection shared by each set of LED segments in the display. (The pin numbers of the chip are shown in Figure 4-33. If you're using another model of display, you'll have to consult a data sheet to find which pin(s) are designed to receive negative voltage.)

Switch on the power supply and touch the free end of the positive wire to each row of holes serving the display on its left and right sides. You should see each segment light up, as shown in Figure 4-31.

Each numeral from 0 to 9 is represented by a group of these segments. The segments are always identified with lowercase letters *a* through *g*, as shown in Figure 4-32. In addition, there is often a decimal point, and although we won't be using it, I've identified it with the letter *h*.

Check Figure 4-33 showing the Kingbright display, and you'll see I have annotated each pin with its function. You can step down the display with the positive wire from your power supply, making sure that each pin lights an appropriate segment.

Incidentally, this display has two pins, numbered 3 and 26, both labeled to receive negative voltage for the first of the digits. Why two pins instead of one? I don't know. You need to use only one, and as this is a passive chip, it doesn't matter if you leave the unused one unconnected. Just take care not to apply positive voltage to it, which would create a short circuit.

A numeric display has no power or intelligence of its own. It's just a bunch of light-emitting diodes. It's not much use, really, until we can figure out a way to illuminate the LEDs in appropriate groups—which will be the next step.

Step 2: Counting

Fortunately, we have a chip known as the 4026, which receives pulses, counts them, and creates an output designed to work with a seven-segment display so that it shows numbers 0–9. The only problem is that this is a rather old-fashioned CMOS chip (meaning, Complementary Metal Oxide Semiconductor) and is thus sensitive to static electricity. Check the caution on page 172 before continuing.

Switch off your power supply and connect its wires to the top of the breadboard, noting that for this experiment, we're going to need positive and negative power on both sides. See Figure 4-34 for details. If your breadboard does not already have the columns of holes color-coded, I suggest you use Sharpie markers to identify them, to avoid polarity errors that can fry your components.

The 4026 counter chip is barely powerful enough to drive the LEDs in our display when powered by 9 volts. Make sure you have the chip the right way up, and insert it into the breadboard immediately above your three-digit display, leaving just one row of holes between them empty.

The schematic in Figure 4-35 shows how the pins of the 4026 chip should be connected. The arrows tell you which pins on the display should be connected with pins on the counter.

Figure 4-36 shows the "pinouts" (i.e., the functions of each pin) of a 4026 counter chip. You should compare this with the schematic in Figure 4-35.

Include a tactile switch between the positive supply and pin 1 of the 4026 counter, with a 10K resistor to keep the input to the 4026 counter negative until the button is pressed. Make sure all your positives and negatives are correct, and turn on the power. You should find that when you tap the tactile switch lightly, the counter advances the numeric display from 0 through 9 and then begins all over again from 0. You may also find that the chip sometimes misinterprets your button-presses, and counts two or even three digits at a time. I'll deal with this problem a little further on.

The LED segments will not be glowing very brightly, because the 1K series resistors deprive of them of the power they would really like to receive. Those resistors are necessary to avoid overloading the outputs from the counter.

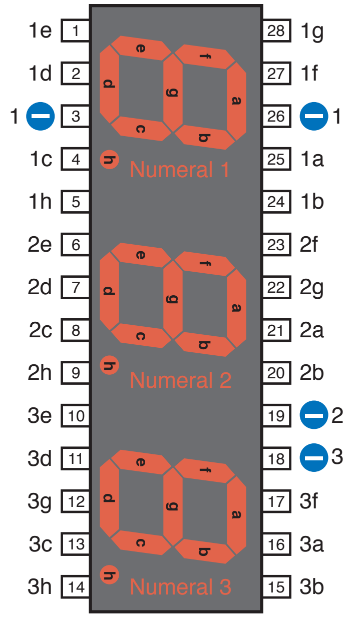

Figure 4-33. *This Kingbright unit incorporates three seven-segment numeric displays in one package, and can be driven by three chained 4026 decade counters. The pin numbers are shown close to the chip. Segments a through g of numeral 1 are identified as 1a through 1g. Segments a through g of numeral 2 are identified as 2a through 2g. Segments a through g of numeral 3 are identified as 3a through 3g.*

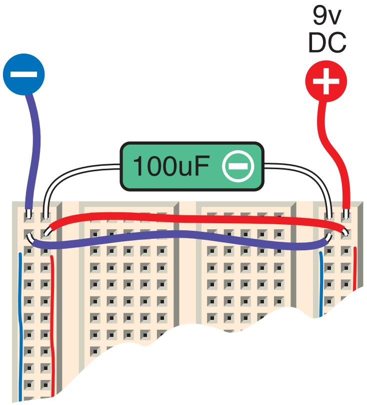

Figure 4-34. *When building circuits around chips, it's convenient to have a positive and negative power supply down each side of your breadboard. For the reaction timer circuit, a 9V supply with a 100 µF smoothing capacitor can be set up like this. If your breadboard doesn't color-code the columns of holes on the left and right sides, I suggest you do that yourself with a permanent marker.*

Grounding Yourself

To avoid the frustration that occurs when you power up a circuit and nothing happens, be sure to take these precautions when you use the older generation of CMOS chips (which often have part numbers from 4000 upward, such as 4002, 4020, and so on):

Chips are often shipped with their legs embedded in black foam. This is electrically conductive foam, and you should keep the chips embedded in it until you are ready to use them.

If the chips are supplied to you in plastic tubes, you can take them out and poke their legs into pieces of conductive foam or, if you don't have any, use aluminum foil. The idea is to avoid one pin on a chip acquiring an electric potential that is much higher than another pin.

While handling CMOS components, grounding yourself is important. I find that in dry weather, I accumulate a static charge merely by walking across a plastic floor-protecting mat in socks that contain some synthetic fibers. You can buy a wrist strap to keep yourself grounded, or simply touch a large metal object, such as a file cabinet, before you touch your circuit board. I am in the habit of working with my socked foot touching a file cabinet, which takes care of the problem.

Never solder a CMOS chip while there is power applied to it.

Grounding the tip of your soldering iron is a good idea.

Better still, don't solder CMOS chips at all. When you're ready to immortalize a project by moving it from a breadboard into perforated board, solder a socket into your perforated board, then push the chip into the socket. If there's a problem in the future, you can unplug the chip and plug in another.

Use a grounded, conductive surface on your workbench. The cheapest way to do this is to unroll some aluminum foil and ground it (with an alligator clip and a length of wire) to a radiator, a water pipe, or a large steel object. I like to use an area of conductive foam to cover my workbench—the same type of foam that is used for packaging chips. However, this foam is quite expensive.

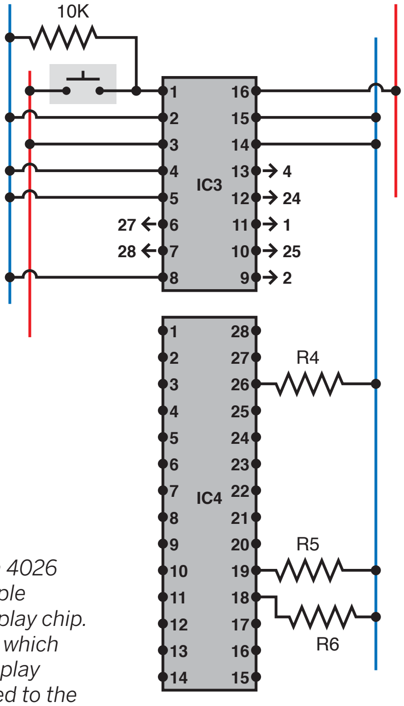
10K
IC3
1
2
3
4
5
6
7
8
16
15
14
13
12
11
10
9
4
24
1
25
2
27
28
IC4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
28
27
26
25
24
23
22
21
20
19
18
17
16
15
R4
R5
R6

Figure 4-35. *IC3 is a 4026 counter. IC4 is a triple seven-segment display chip. The arrows tell you which pins on the LED display should be connected to the pins on the counter.*

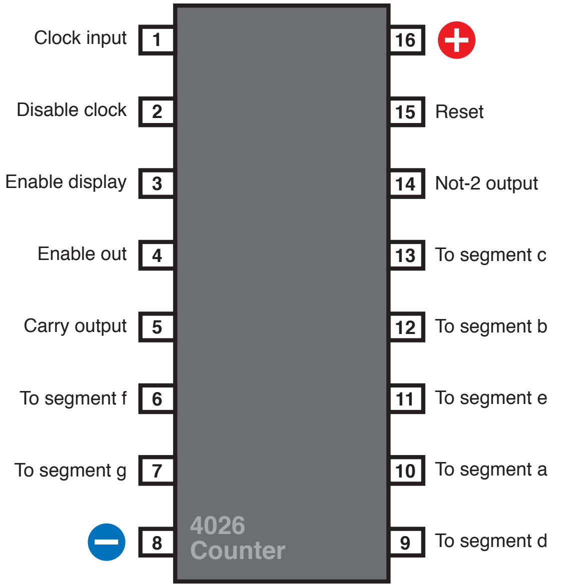
Clock input
1
Disable clock
2
Enable display
3
Enable out
4
Carry output
5
To segment f
6
To segment g
7
8
16
15
Reset
14
Not-2 output
13
To segment c
12
To segment b
11
To segment e
10
To segment a
9
To segment d
4026
Counter

Figure 4-36. *The 4026 decade counter is a CMOS chip that accepts clock pulses on pin 1, maintains a running total from 0 to 9, and outputs this total via pins designed to interface with a seven-segment LED numeric display.*

FUNDAMENTALS

Counters and seven-segment displays

Most counters accept a stream of pulses and distribute them to a series of pins in sequence. The 4026 decade counter is unusual in that it applies power to its output pins in a pattern that is just right to illuminate the segments of a 7-segment numeric display.

Some counters create positive outputs (they "source" current) while others create negative outputs (they "sink" current). Some seven-segment displays require positive input to light up the numbers. These are known as "common cathode" displays. Others require negative input and are known as "common anode" displays. The 4026 delivers positive outputs and requires a common cathode display.

Check the data sheet for any counter chip to find out how much power it requires, and how much it can deliver. CMOS chips are becoming dated, but they are very useful to hobbyists, because they will tolerate a wide range of supply voltage—from 5 to 15 volts in the case of the 4026. Other types of chips are much more limited.

Most counters can source or sink only a few milliamps of output power. When the 4026 is running on a 9-volt power supply, it can source about 4mA of power from each pin. This is barely enough to drive a seven-segment display.

You can insert a series resistor between each output pin of the counter and each input pin of the numeric display, but a simpler, quicker option is to use just one series resistor for each numeral, between the negative-power pin and ground. The experiment that I'm describing uses this shortcut. Its disadvantage is that digits that require only a couple of segments (such as numeral 1) will appear brighter than those that use many segments (such as numeral 8).

If you want your display to look bright and professional, you really need a transistor to drive each segment of each numeral. An alternative is to use a chip containing multiple "op amps" to amplify the current.

When a decade counter reaches 9 and rolls over to 0, it emits a pulse from its "carry" pin. This can drive another counter that will keep track of tens. The carry pin on that counter can be chained to a third counter that keeps track of hundreds, and so on. In addition to decade counters, there are hexadecimal counters (which count in 16s), octal counters (in 8s), and so on.

Why would you need to count in anything other than tens? Consider that the four numerals on a digital clock each count differently. The rightmost digit rolls over when it reaches 10. The next digit to the left counts in sixes. The first hours digit counts to 10, gives a carry signal, counts to 2, and gives another carry signal. The leftmost hours digit is either blank or 1, when displaying time in 12-hour format. Naturally there are counters specifically designed to do all this.

Counters have control pins such as "clock disable," which tells the counter to ignore its input pulses and freeze the display, "enable display," which enables the output from the chip, and "reset," which resets the count to zero.

The 4026 requires a positive input to activate each control pin. When the pins are grounded, their features are suppressed.

To make the 4026 count and display its running total you must ground the "clock disable" and "reset" pins (to suppress their function) and apply positive voltage to the "enable display" pin (to activate the output). See Figure 4-36 to see these pins identified.

Assuming that you succeed in getting your counter to drive the numeric display, you're ready to add two more counters, which will control the remaining two numerals. The first counter will count in ones, the second in tens, and the third in hundreds.

In Figure 4-37, I've continued to use arrows and numbers to tell you which pins of the counters should be connected to which pins of the numeric display. Otherwise, the schematic would be a confusing tangle of wires crossing each other.

At this point, you can give up in dismay at the number of connections—but really, using a breadboard, it shouldn't take you more than half an hour to complete this phase of the project. I suggest you give it a try, because there's

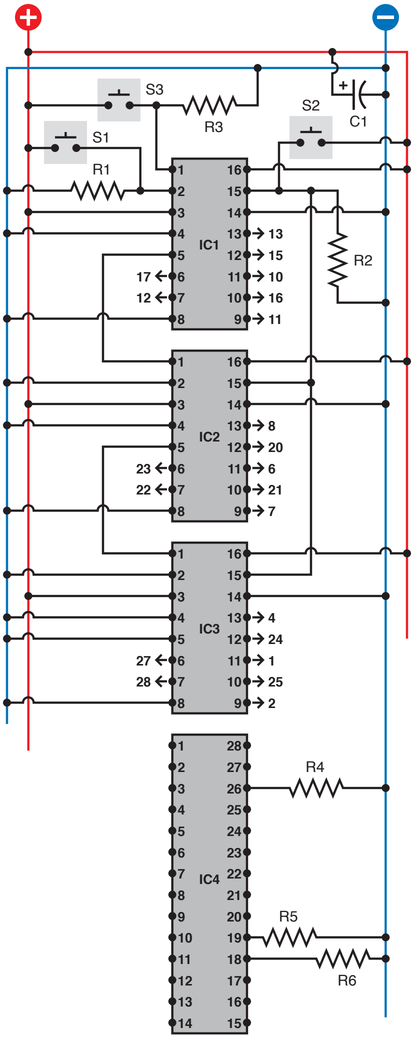

Figure 4-37. *This test circuit, laid out as you would be likely to place it on a breadboard, allows you to trigger a counter manually to verify that the display increments from 000 upward to 999.*

Component values:

All resistors are 1K.
S1, S2, S3: SPST tactile switches, normally open
IC1, IC2, IC3: 4026 decade counter chips
IC4: Kingbright 3-digit common-cathode display
C1: 100 µF (minimum) smoothing capacitor

Wire the output pins on IC1, IC2, and IC3 to the pins on IC4, according to the numbers preceded by arrows. The actual wires have been omitted for clarity. Check for the pinouts of IC4.

something magical about seeing a display count from 000 through 999 "all by itself," and I chose this project because it also has a lot of instructional value.

S1 is attached to the "clock disable" pin of IC1, so that when you hold down this button, it should stop that counter from counting. Because IC1 controls IC2, and IC2 controls IC3, if you freeze IC1, the other two will have to wait for it to resume. Therefore you won't need to make use of their "clock disable" features.

S2 is connected to the "reset" pins of all three counters, so that when you hold down this button, it should set them all to zero.

S3 sends positive pulses manually to the "clock input" pin of the first counter.

S1, S2, and S3 are all wired in parallel with 1K resistors connected to the negative side of the power supply. The idea is that when the buttons are not being pressed, the "pull-down" resistors keep the pins near ground (zero) voltage. When you press one of the buttons, it connects positive voltage directly to the chip, and easily overwhelms the negative voltage. This way, the pins remain either in a definitely positive or definitely negative state. If you disconnect one of these pull-down resistors you are likely to see the numeric display "flutter" erratically. (The numeric display chip has some unconnected pins, but this won't cause any problem, because it is a passive chip that is just a collection of LED segments.)

Always connect input pins of a CMOS chip so that they are either positive or negative. See the "No Floating Pins" warning on the next page.

I suggest that you connect all the wires shown in the schematic first. Then cut lengths of 22-gauge wire to join the remaining pins of the sockets from IC1, IC2, and IC3 to IC4.

Switch on the power and press S2. You'll see three zeros in your numeric display.

Each time you press S3, the count should advance by 1. If you press S2, the count should reset to three zeros. If you hold down S1 while you press S3 repeatedly, the counters should remain frozen, ignoring the pulses from S3.

FUNDAMENTALS

Switch bounce

When you hit S3, I think you'll find that the count sometimes increases by more than 1. This does not mean that there's something wrong with your circuit or your components; you are just observing a phenomenon known as "switch bounce."

On a microscopic level, the contacts inside a pushbutton switch do not close smoothly, firmly, and decisively. They vibrate for a few microseconds before settling; the counter chip detects this vibration as a series of pulses, not just one.

Various circuits are available to "debounce" a switch. The simplest option is to put a small capacitor in parallel with the switch, to absorb the fluctuations; but this is less than ideal. I'll come back to the topic of debouncing later in the book. Switch bounce is not a concern in this circuit, because we're about to get rid of S3 and substitute a 555 timer that generates nice clean bounceless pulses.

Pulse Generation

A 555 timer is ideal for driving a counter chip. You've already seen how to wire a 555 to create a stream of pulses that made noise through a loudspeaker. I'm reproducing the same circuit in Figure 4-38 in simplified form, using the positive and negative supply configuration in the current project. Also I'm showing the connection between pins 2 and 6 in the way that you're most likely to make it, via a wire that loops over the top of the chip.

For the current experiment, I'm suggesting initial component values that will generate only four pulses per second. Any faster than that, and you won't be able to verify that your counters are counting properly.

Install IC5 and its associated components on your breadboard immediately above IC1. Don't leave any gap between the chips. Disconnect S3 and R3 and connect a wire directly between pin 3 (output) of IC5 and pin 1 (clock) of IC1, the topmost counter. Power up again, and you should see the digits advancing rapidly in a smooth, regular fashion. Press S1, and while you hold it, the count should freeze. Release S1 and the count will resume. Press S2 and the counter should reset, even if you are pressing S1 at the same time.

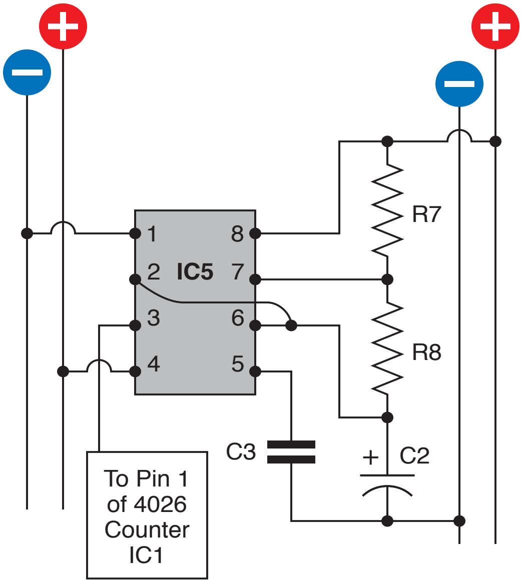

Figure 4-38. *A basic astable circuit to drive the decade counter in the previous schematic. Output is approximately 4 pulses per second.*

R7: 1K
R8: 2K2
C2: 68 µF
C3: 0.1 µF
IC5: 555 timer

No Floating Pins!

A CMOS chip is hypersensitive. Any pin that is not wired either to the supply voltage or to ground is said to be "floating" and may act like an antenna, sensitive to the smallest fluctuations in the world around it.

The 4026 counter chip has a pin labeled "clock disable." The manufacturer's data sheet helpfully tells you that if you give this pin a positive voltage, the chip stops counting and freezes its display. As you don't want to do that, you may just ignore that pin and leave it unconnected, at least while you test the chip. This is a very bad idea!

What the data sheet doesn't bother to tell you (presumably because "everyone knows" such things) is that if you want the clock to run normally, the clock-disable feature itself has to be disabled, by wiring it to negative (ground) voltage. If you leave the pin floating (and I speak from experience), the chip will behave erratically and uselessly.

All input pins must be either positively or negatively wired, unless otherwise specified.

Refinements

Now it's time to remember that what we really want this circuit to do is test a person's reflexes. When the user starts it, we want an initial delay, followed by a signal—probably an LED that comes on. The user responds to the signal by pressing a button as quickly as possible. During the time it takes for the person to respond, the counter will count milliseconds. When the person presses the button, the counter will stop. The display then remains frozen indefinitely, displaying the number of pulses that were counted before the person was able to react.

How to arrange this? I think we need a flip-flop. When the flip-flop gets a signal, it starts the counter running—and keeps it running. When the flip-flop gets another signal (from the user pressing a button), it stops the counter running, and keeps it stopped.

How do we build this flip-flop? Believe it or not, we can use yet another 555 timer, in a new manner known as bistable mode.

FUNDAMENTALS

The bistable 555 timer

Figure 4-39 shows the internal layout of a 555 timer, as before, but the external components on the righthand side have been eliminated. Instead, I'm applying a constant negative voltage to pin 6. Can you see the consequences? Suppose you apply a negative pulse to the trigger (pin 2). Normally when you do this and the 555 starts running, it generates a positive output while charging a capacitor attached to pin 6. When the capacitor reaches 2/3 of the full supply voltage, this tells the 555 to ends its positive output, and it flips back to negative.

Well, if there's no capacitor, there's nothing to stop the timer. Its positive output will just continue indefinitely. However, pin 4 (the reset) can still override everything, so if you apply negative voltage to pin 4, it flips the output to negative. After that, the output will stay negative indefinitely, as it usually does, until you trigger the timer by dropping the voltage to pin 2 again. This will flip the timer back to generating its positive output.

Here's a quick summary of the bistable configuration:

- A negative pulse to pin 2 turns the output positive.
- A negative pulse to pin 4 turns the output negative.
- The timer is stable in each of these states. Its run-time has become infinite.

It's OK to leave pins 5 and 7 of the timer unconnected, because we're pushing it into extreme states where any random signals from those pins will be ignored.

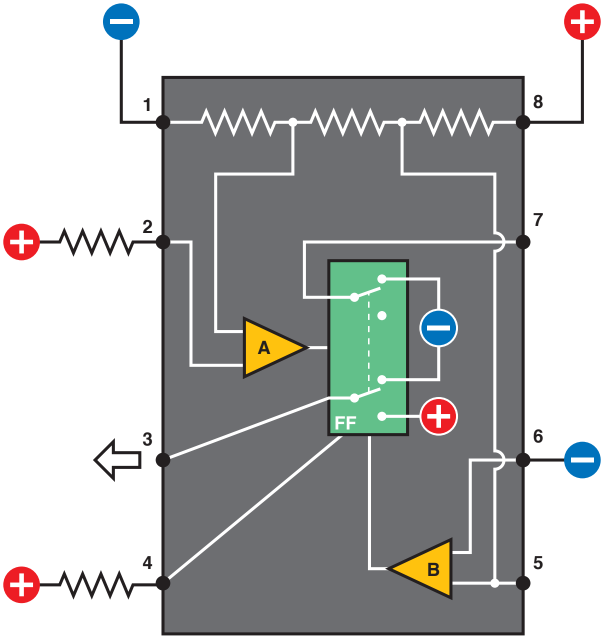

Figure 4-39. *In the bistable configuration, pin 6 of a 555 timer is perpetually negative, so the timer cycle never ends, unless you force it to do so by applying a negative pulse to pin 4 (the reset).*

In bistable mode, the 555 has turned into one big flip-flop. To avoid any uncertainty, we keep pins 2 and 4 normally positive via pull-up resistors, but negative pulses on those pins can overwhelm them when we want to flip the 555 into its opposite state. The schematic for running a 555 timer in bistable mode, controlled by two pushbuttons, is shown in Figure 4-40. You can add this above your existing circuit. Because you're going to attach the output from IC6 to pin 2 of IC1, the topmost counter, you can disconnect S1 and R1 from that pin. See Figure 4-41.

Now, power up the circuit again. You should find that it counts in the same way as before, but when you press S4, it freezes. This is because your bistable 555 timer is sending its positive output to the "clock disable" pin on the counter. The counter is still receiving a stream of pulses from the astable 555 timer, but as long as pin 2 is positive on the counter, the counter simply ignores the pulses.

Now press S5, which flips your bistable 555 back to delivering a negative output, at which point the count resumes.

We're getting close to a final working circuit here. We can reset the count to zero (with S3), start the count (with S5), and wait for the user to stop the count (with S4). The only thing missing is a way to start the count unexpectedly.

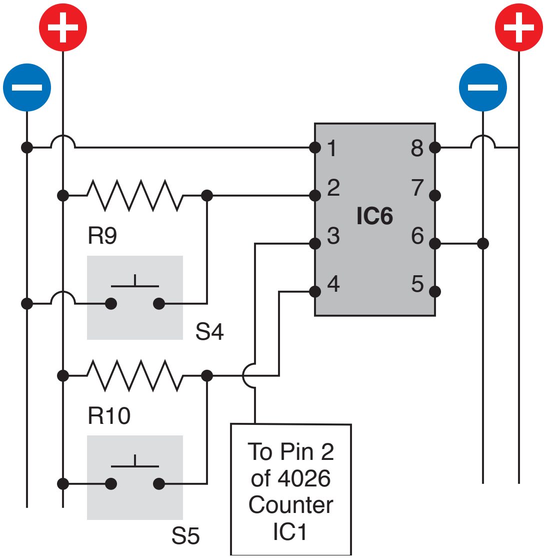

Figure 4-40. *Adding a bistable 555 timer to the reflex tester will stop the counter with a touch of a button, and keep it stopped.*

R9, R10: 1K
IC6: 555 timer

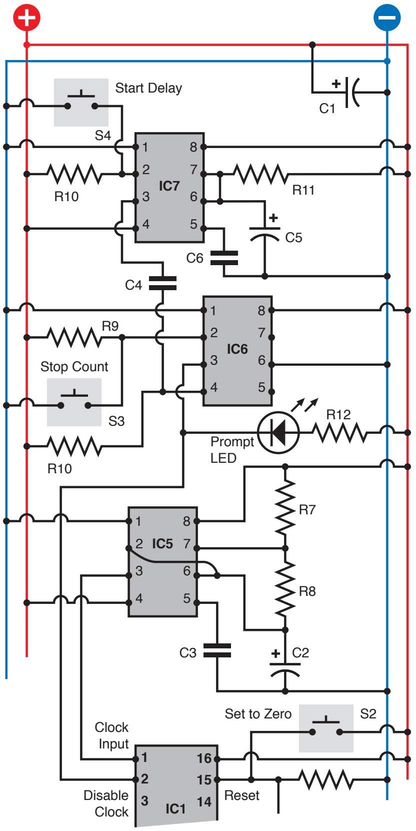

Figure 4-41. *The completed control section of the circuit, to be added above these timers.*

R7, R9, R10, R12: 1K
R8: 2K2
R11: 330K
C1: 100 µF
C2: 68 µF
C3, C4, C6: 0.1 µF
C5: 10 µF
S1, S2, S3: tactile switches
IC5, IC6, IC7: 555 timers

The Delay

Suppose we set up yet another 555 in monostable mode. Trigger its pin 2 with a negative pulse, and the timer delivers a positive output that lasts for, say, 4 seconds. At the end of that time, its output goes back to being negative. Maybe we can hook that positive-to-negative transition to pin 4 of IC6. We can use this instead of switch S5, which you were pressing previously to start the count.

Check the new schematic in Figure 4-41 which adds another 555 timer, IC7 above IC6. When the output from IC7 goes from positive to negative, it will trigger the reset of IC6, flipping its output negative, which allows the count to begin. So IC7 has taken the place of the start switch, S4. You can get rid of S4, but keep the pull-up resistor, R9, so that the reset of IC6 remains positive the rest of the time.

This arrangement works because I have used a capacitor, C4, to connect the output of IC7 to the reset of IC6. The capacitor communicates the sudden change from positive to negative, but the rest of the time it blocks the steady voltage from IC7 so that it won't interfere with IC6.

The final schematic in Figure 4-41 shows the three 555 timers all linked together, as you should insert them above the topmost counter, IC1. I also added an LED to signal the user. Figure 4-42 is a photograph of my working model of the circuit.

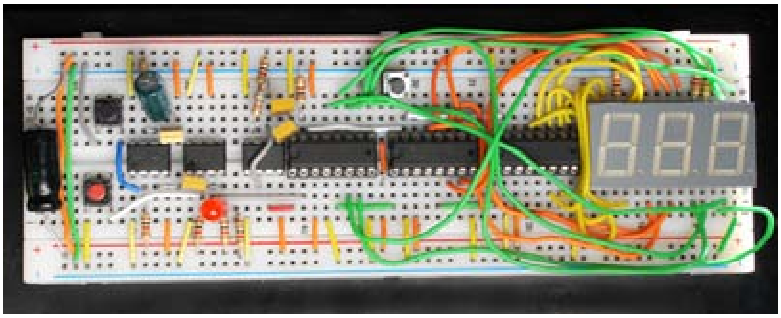

Figure 4-42. *The complete reaction-timer circuit barely fits on a 63-row breadboard.*

Because this circuit is complicated, I'll summarize the sequence of events when it's working. Refer to Figure 4-41 while following these steps:

1. User presses Start Delay button S4, which triggers IC7.
2. IC7 output goes high for a few seconds while C5 charges.
3. IC7 output drops back low.
4. IC7 communicates a pulse of low voltage through C4 to IC6, pin 4.
5. IC 6 output flips to low and flops there.
6. Low output from IC6 sinks current through LED and lights it.
7. Low output from IC6 also goes to pin 2 of IC1.

8. Low voltage on pin 2 of IC1 allows IC1 to start counting.
9. User presses S3, the "stop" button.
10. S3 connects pin 2 of IC6 to ground.
11. IC6 output flips to high and flops there.
12. High output from IC6 turns off the LED.
13. High output from IC6 also goes to pin 2 of IC1.
14. High voltage on pin 2 of IC1 stops it from counting.
15. After assessing the result, user presses S2.
16. S2 applies positive voltage to pin 15 of IC1, IC2, IC3.
17. Positive voltage resets counters to zero.
18. The user can now try again.
19. Meanwhile, IC5 is running continuously throughout.

In case you find a block diagram easier to understand, I've included that, too, in Figure 4-43.

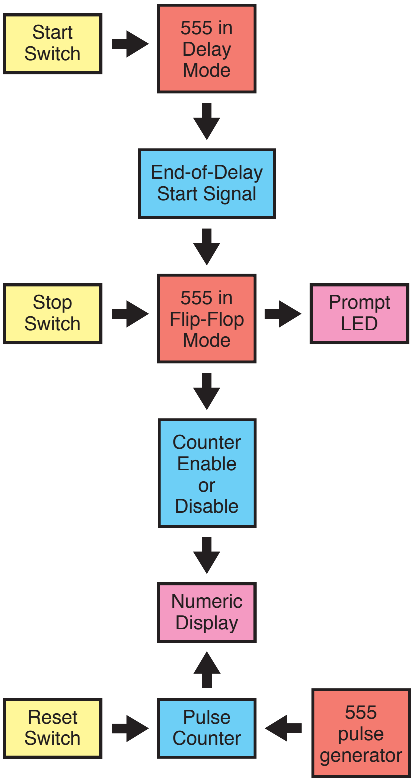

Figure 4-43. *The functions of the reflex tester, summarized as a block diagram.*

Using the Reflex Tester

At this point, you should be able to test the circuit fully. When you first switch it on, it will start counting, which is slightly annoying, but easily fixed. Press S3 to stop the count. Press S2 to reset to zero.

Now press S4. Nothing seems to happen—but that's the whole idea. The delay cycle has begun in stealth mode. After a few seconds, the delay cycle ends, and the LED lights up. Simultaneously, the count begins. As quickly as possible, the user presses S3 to stop the count. The numerals freeze, showing how much time elapsed.

There's only one problem—the system has not yet been calibrated. It is still running in slow-motion mode. You need to change the resistor and capacitor attached to IC5 to make it generate 1,000 pulses per second instead of just three or four.

Substitute a 10K trimmer potentiometer for R8 and a 1 μF capacitor for C2. This combination will generate about 690 pulses per second when the trimmer is presenting maximum resistance. When you turn the trimmer down to decrease its resistance, somewhere around its halfway mark the timer will be running at 1,000 pulses per second.

How will you know exactly where this point is? Ideally, you would attach an oscilloscope probe to the output from IC5. But, most likely, you don't have an oscilloscope, so here are a couple other suggestions.

First remove the 1 μF capacitor at C2 and substitute a 10 μF capacitor. Because you are multiplying the capacitance by 10, you will reduce the speed by 10. The leftmost digit in your display should now count in seconds, reaching 9 and rolling over to 0 every 10 seconds. You can adjust your trimmer potentiometer while timing the display with a stopwatch. When you have it right, remove the 10 μF capacitor and replace the 1 μF capacitor at C2.

The only problem is, the values of capacitors may be off by as much as 10%. If you want to fine-tune your reflex timer, you can proceed as follows.

Disconnect the wire from pin 5 of IC3, and substitute an LED with a 1K series resistor between pin 5 and ground. Pin 5 is the “carry” pin, which will emit a positive pulse whenever IC3 counts up to 9 and rolls over to start at 0 again. Because IC3 is counting tenths of a second, you want its carry output to occur once per second.

Now run the circuit for a full minute, using your stopwatch to see if the flashing LED drifts gradually faster or slower than once per second. If you have a camcorder that has a time display in its viewfinder, you can use that to observe the LED.

If the LED flashes too briefly to be easily visible, you can run a wire from pin 5 to another 555 timer that is set up in monostable mode to create an output lasting for around 1/10 of a second. The output from that timer can drive an LED.

Enhancements

It goes without saying that anytime you finish a project, you see some opportunities to improve it. Here are some suggestions:

1. No counting at power-up. It would be nice if the circuit begins in its “ready” state, rather than already counting. To achieve this you need to send a negative pulse to pin 2 of IC6, and maybe a positive pulse to pin 15 of IC1. Maybe an extra 555 timer could do this. I’m going to leave you to experiment with it.
2. Audible feedback when pressing the Start button. Currently, there’s no confirmation that the Start button has done anything. All you need to do is buy a piezoelectric beeper and wire it between the righthand side of the Start button and the positive side of the power supply.
3. A random delay interval before the count begins. Making electronic components behave randomly is very difficult, but one way to do it would be to require the user to hold his finger on a couple of metal contacts. The skin resistance of the finger would substitute for R11. Because the finger pressure would not be exactly the same each time, the delay would vary. You’d have to adjust the value of C5.

Summing Up

This project demonstrated how a counter chip can be controlled, how counter chips can be chained together, and three different functions for 555 timers. It also showed you how chips can communicate with each other, and introduced you to the business of calibrating a circuit after you've finished building it.

Naturally, if you want to get some practical use from the circuit, you should build it into an enclosure with heavier-duty pushbuttons—especially the button that stops the count. You'll find that when people's reflexes are being tested, they are liable to hit the stop button quite hard.

Because this was a major project, I'll follow it up here with some quicker, easier ones as we move into the fascinating world of another kind of integrated circuit: logic chips.

Experiment 19: Learning Logic

You will need:

- Assorted resistors and capacitors.
- 74HC00 quad 2-input NAND chip, 74HC08 quad 2-input AND chip, and LM7805 voltage regulator. Quantity: 1 of each.
- Signal diode, 1N4148 or similar. Quantity: 1.
- Low-current LED. Quantity: 1.
- SPST tactile switches. Quantity: 2.

You're going to be entering the realm of pure digital electronics, using "logic gates" that are fundamental in every electronic computing device. When you deal with them individually, they're extremely easy to understand. When you start chaining them together, they can seem intimidatingly complex. So let's start with them one at a time.

Logic gates are much fussier than the 555 timer or the 4026 counter that you used previously. They demand an absolutely precise 5 volts DC, with no fluctuations or "spikes" in the flow of current. Fortunately, this is easy to achieve: just set up your breadboard with an LM7805 voltage regulator, as shown in the schematic in Figure 4-44 and the photograph in Figure 4-45. The regulator receives 9 volts from your usual voltage supply, and reduces it to 5 volts, with the help of a couple of capacitors. You apply the 9 volts to the regulator, and distribute the 5 volts down the sides of your breadboard instead of the unregulated voltage that you used previously. Use your meter to verify the voltage, and make sure you have the polarity clearly marked.

After installing your regulator, take a couple of tactile switches, two 10K resistors, a low-current LED, and a 1K resistor, and set them around a 74HC00 logic chip as shown in Figure 4-46. You may notice that many of the pins of the chip are shorted together and connected to the negative side of the power supply. I'll explain that in a moment.

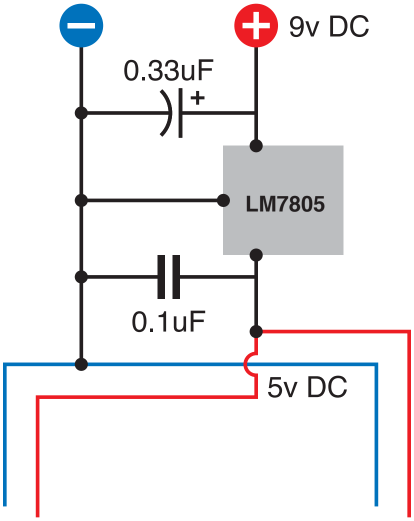

Figure 4-44. *This simple circuit is essential to provide a regulated 5V DC supply for logic chips.*

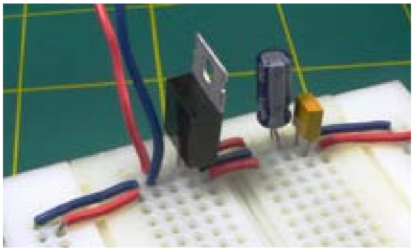

Figure 4-45. *The voltage regulator and its two capacitors can fit snugly at the top of a breadboard. Remember to apply the 9V input voltage at the left pin of the regulator, and distribute the 5V output down the sides of the breadboard.*

FUNDAMENTALS

Voltage regulators

The simplest versions of these little semiconductors accept a higher DC voltage on one pin and deliver a lower DC voltage on another pin, with a third pin (usually in the middle) serving as a common negative, or ground. You should also attach a couple of capacitors to smooth the current, as shown in Figure 4-46.

Typically you can put a 7.5-volt or 9-volt supply on the "input" side of a 5-volt regular, and draw a precise 5 volts from the "output" side. If you're wondering where the extra voltage goes, the answer is, the regulator turns the electricity into heat. For this reason, small regulators (such as the one in Figure 4-8) often have a metal back with a hole in the top. Its purpose is to radiate heat, which it will do more effectively if you bolt it to a piece of aluminum, since aluminum conducts heat very effectively. The aluminum is known as a heat sink, and you can buy fancy ones that have multiple cooling fins.

For our purposes, we won't be drawing enough current to require a heat sink.

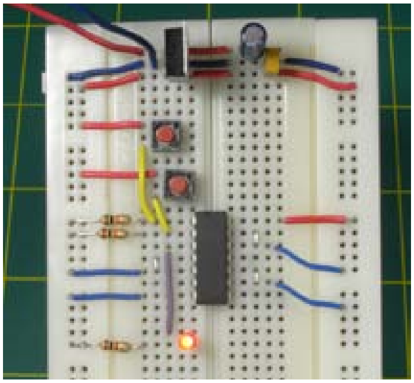

Figure 4-47. *This breadboard layout is exactly equivalent to the schematic in Figure 4-46.*

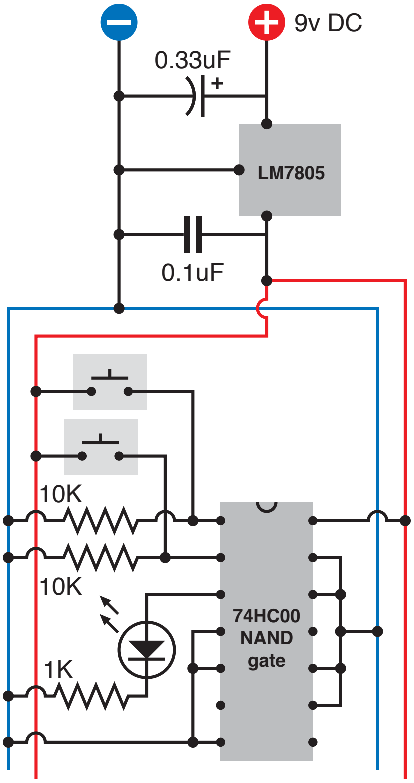

Figure 4-46. *By observing the LED when you press either, both, or neither of the buttons, you can easily figure out the logical function of the NAND gate.*

When you connect power, the LED should light up. Press one of the tactile switches, and the LED remains illuminated. Press the other tactile switch, and again the LED stays on. Now press both switches, and the light should go out.

Pins 1 and 2 are logic inputs for the 74HC00 chip. Initially they were held at negative voltage, being connected to the negative side of the power supply through 10K resistors. But each pushbutton overrides its pull-down resistor and forces the input pin to go positive.

The logic output from the chip, as you saw, is normally positive—but *not* if the first input *and* the second input are positive. Because the chip does a "Not AND" operation, it's known as a NAND logic gate. You can see the breadboard layout in Figure 4-47. Figure 4-48 is a simplified version of the circuit. The U-shaped thing with a circle at the bottom is the logic symbol for a NAND gate. No power supply is shown for it, but in fact all logic chips require a power supply, which enables them to put out more current than they take in. Anytime you see a symbol for a logic chip, try to remember that it has to have power to function.

The 74HC00 actually contains four NAND gates, each with two logical inputs and one output. They are arrayed as shown in Figure 4-49. Because only one gate was needed for the simple test, the input pins of the unused gates were shorted to the negative side of the power supply.

Pin 14 supplies positive power for the chip; pin 7 is its ground pin. Almost all the 7400 family of logic chips use the same pins for positive and negative power, so you can swap them easily.

In fact, let's do that right now. First, disconnect the power. Carefully pull out the 74HC00 and put it away with its legs embedded in conductive foam. Substitute a 74HC08 chip, which is an AND chip. Make sure you have it the right way up, with its notch at the top. Reconnect the power and use the pushbuttons as you did before. This time, you should find that the LED comes on if the first input AND the second input are both positive, but it remains dark otherwise. Thus, the AND chip functions exactly opposite to the NAND chip. Its pinouts are shown in Figure 4-50.

You may be wondering why these things are useful. Soon you'll see that we can put logic gates together to do things such as create an electronic combination lock, or a pair of electronic dice, or a computerized version of a TV quiz show where users compete to answer a question. And if you were really insanely ambitious, you could build an entire computer out of logic gates.

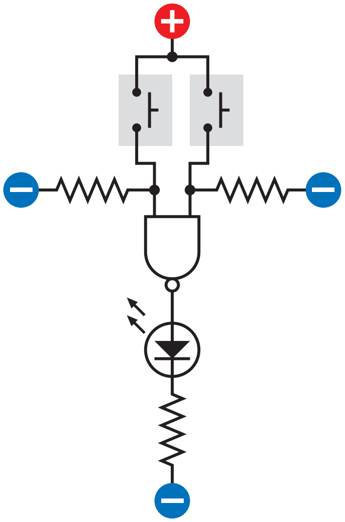

Figure 4-48. *The structure and function of the NAND gate is easier to visualize with this simplified schematic that omits the power supply for the chip and doesn't attempt to place the wires to fit a breadboard layout.*

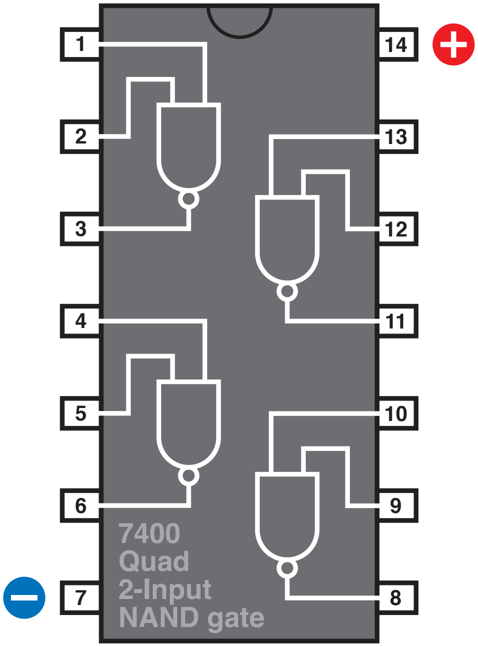

Figure 4-49. *The pinouts of the logic gates in a 74HC00chip.*

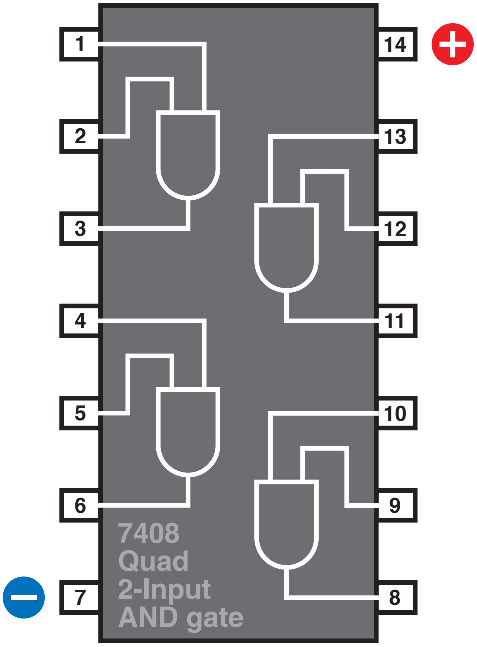

Figure 4-50. *The pinouts of the logic gates in a 74HC08chip.*

BACKGROUND

From Boole to Shannon

George Boole was a British mathematician, born in 1815, who did something that few people are ever lucky enough or smart enough to do: he invented an entirely new branch of mathematics.

Interestingly, it was not based on numbers. Boole had a relentlessly logical mind, and he wanted to reduce the world to a series of true-or-false statements which could overlap in interesting ways. For instance, suppose there is a couple named Ann and Bob who have so little money, they only own one hat. Clearly, if you happen to run into Ann and Bob walking down the street, there are four possibilities: neither of them may be wearing a hat, Ann may be wearing it, or Bob may be wearing it, but they cannot both be wearing it.

The diagram in Figure 4-51 illustrates this. All the states are possible except the one where the circles overlap. (This is known as a Venn diagram. I leave it to you to search for this term if it interests you and you'd like to learn more.) Boole took this concept much further, and showed how to create and simplify extremely complex arrays of logic.

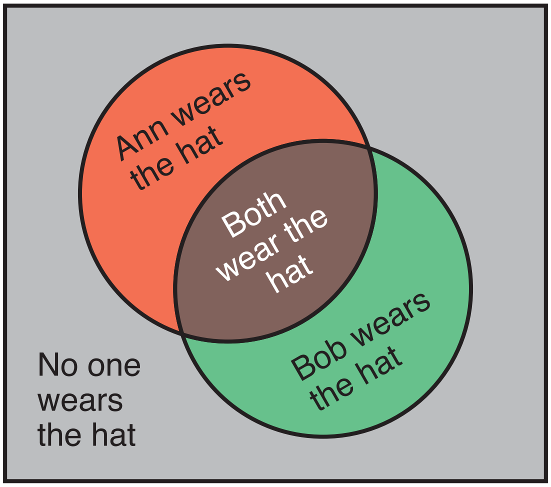

Figure 4-51. *This slightly frivolous Venn diagram illustrates the various possibilities for two people, Ann and Bob, who own only one hat.*

Another way to summarize the hat-wearing situation is to make the "truth table" shown in Figure 4-52. The rightmost column shows whether each combination of propositions can be true. Now check the table in Figure 4-53. It's the same table but uses different labels, which describe the pattern you have seen while using the NAND gate.

Boole published his treatise on logic in 1854, long before it could be applied to electrical or electronic devices. In fact, during his lifetime, his work seemed to have no practical applications at all. But a man named Claude Shannon encountered Boolean logic while studying at MIT in the 1930s, and in 1938 he published a paper describing how Boolean analysis could be applied to circuits using relays. This had immediate practical applications, as telephone networks were growing rapidly, creating complicated switching problems.

BACKGROUND

From Boole to Shannon (continued)

Ann is wearing the hat	Bob is wearing the hat	This combination can be
NO	NO	TRUE
NO	YES	TRUE
YES	NO	TRUE
YES	YES	FALSE

Figure 4-52. *The hat-wearing possibilities can be expressed in a "truth table."*

Input A	Input B	Output
OFF	OFF	ON
OFF	ON	ON
ON	OFF	ON
ON	ON	OFF

Figure 4-53. *The truth-table from can be relabeled to describe the inputs and outputs of a NAND gate.*

A very simple telephone problem could be expressed like this. Suppose two customers in a rural area share one telephone line. If one of them wants to use the line, or the other wants to use it, or neither of them wants to use it, there's no problem. But they cannot both use it at once. You may notice that this is exactly the same as the hat-wearing situation for Ann and Bob.

We can easily draw a circuit using two normally closed relays that creates the desired outcome (see Figure 4-54), but if you imagine a telephone exchange serving many thousands of customers, the situation becomes very complicated indeed. In fact, in Shannon's time, no logical process existed to find the best solution and verify that it used fewer components than some other solution.

Shannon saw that Boolean analysis could be used for this purpose. Also, if you used an "on" condition to represent numeral 1 and an "off" condition to represent numeral 0, you could build a system of relays that could count. And if it could count, it could do arithmetic.

When vacuum tubes were substituted for relays, the first practical digital computers were built. Transistors took the place of vacuum tubes, and integrated circuit chips replaced transistors, leading to the desktop computers that we now take for granted today. But deep down, at the lowest levels of these incredibly complex devices, they still use the laws of logic discovered by George Boole. Today, when you use a search engine online, if you use the words AND and OR to refine your search, you're actually using Boolean operators.

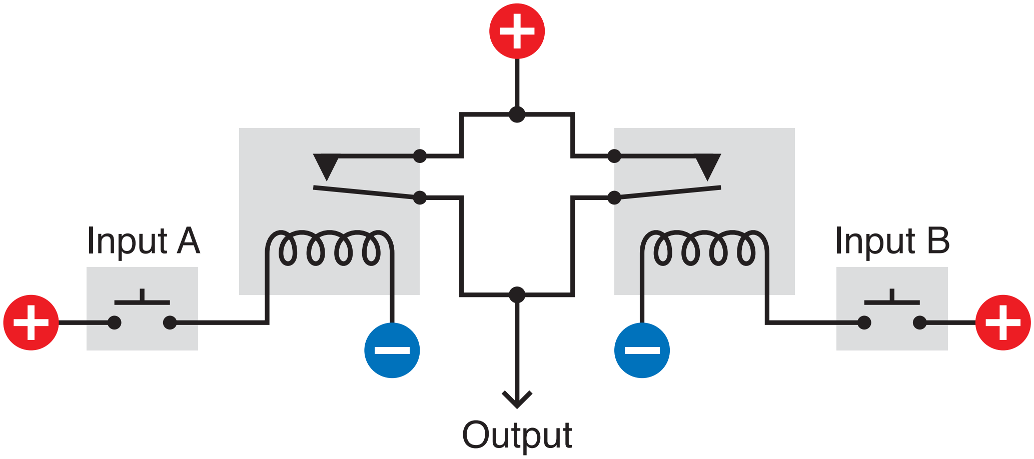

Figure 4-54. *This relay circuit could illustrate the desired logic for two telephone customers wanting to share one line, and its behavior is almost identical to that of the NAND schematic shown in Figure 4-48.*

Figure 4-55. *Ann and Bob attempt to overcome the limitations of Boolean logic.*

ESSENTIALS

Logic gate basics

The NAND gate is the most fundamental building block of digital computers, because (for reasons which I don't have space to explain here) it enables digital addition. If you want to explore more try searching online for topics such as "binary arithmetic" and "half-adder."

Generally, there are seven types of logic gates:

- AND
- NAND
- OR
- NOR
- XOR
- XNOR
- NOT

Of the six two-input gates, the XNOR is hardly ever used. The NOT gate has a single input, and simply gives a negative output when the input is positive or a positive output when the input is negative. The NOT is more often referred to as an "Inverter." The symbols for all seven gates are shown in Figure 4-56.

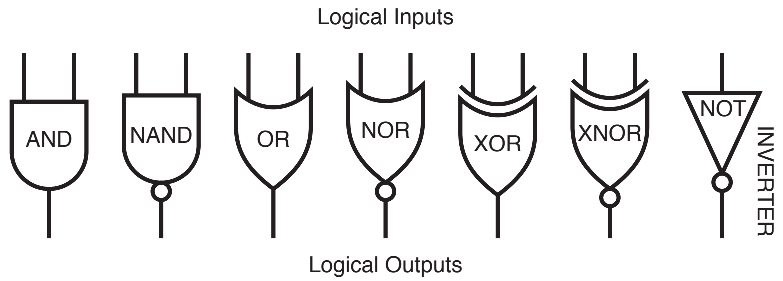

Figure 4-56. *American symbols for the six types of two-input logic gates, and the single-input inverter.*

I've shown the American symbols. Other symbols have been adopted in Europe, but the traditional symbols shown here are the ones that you will usually find, even being used by Europeans. I also show the truth tables, in Figure 4-57, illustrating the logical output (high or low) for each pair of inputs of each type of gate.

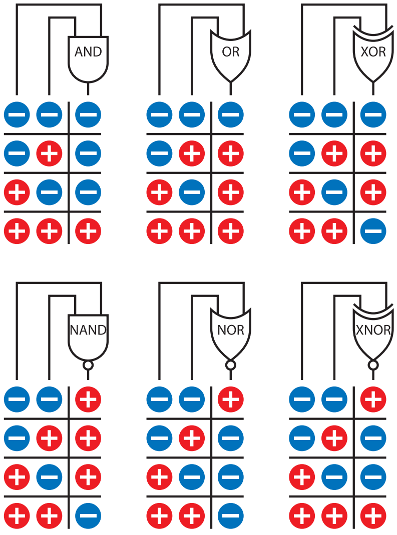
AND
OR
XOR
NAND
NOR
XNOR

		AND
−	−	−
−	+	−
+	−	−
+	+	+

		OR
−	−	−
−	+	+
+	−	+
+	+	+

		XOR
−	−	−
−	+	+
+	−	+
+	+	−

		NAND
−	−	+
−	+	+
+	−	+
+	+	−

		NOR
−	−	+
−	+	−
+	−	−
+	+	−

		XNOR
−	−	+
−	+	−
+	−	−
+	+	+

Figure 4-57. *Inputs and corresponding outputs for the six types of logic gates (note that the XNOR gate is seldom used). The minus signs indicate low voltage, close to ground potential. The plus signs indicate higher voltage, close to the positive potential of the power supply in the circuit. The exact voltages will vary depending on other components that may be actively connected.*

ESSENTIALS

Logic gate basics (continued)

If you have difficulty visualizing logic gates, a mechanical comparison may help. You can think of them as being like sliding plates with holes in them, in a bubblegum machine. Two people, A and B, can push the plates. These people are the two inputs, which are considered positive when they are doing something. (Negative logic systems also exist, but are uncommon, so I'm only going to talk about positive systems here.)

The flow of bubblegum represents a flow of positive current. The full set of possibilities is shown in Figures 4-58 through 4-63.

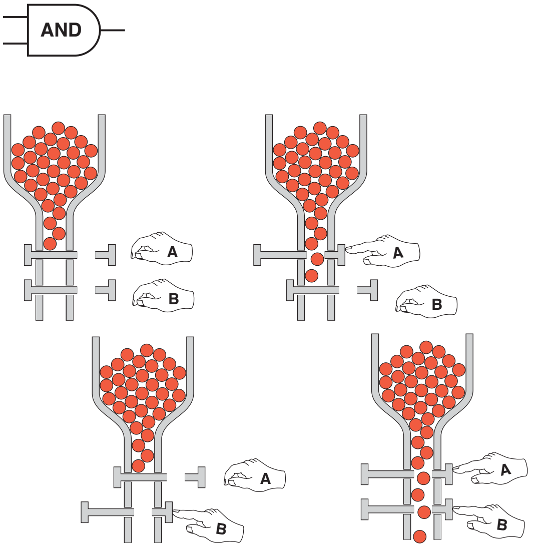

INPUT B	INPUT A	OUTPUT
NO	NO	NO
NO	YES	NO
YES	NO	NO
YES	YES	YES

Figure 4-58

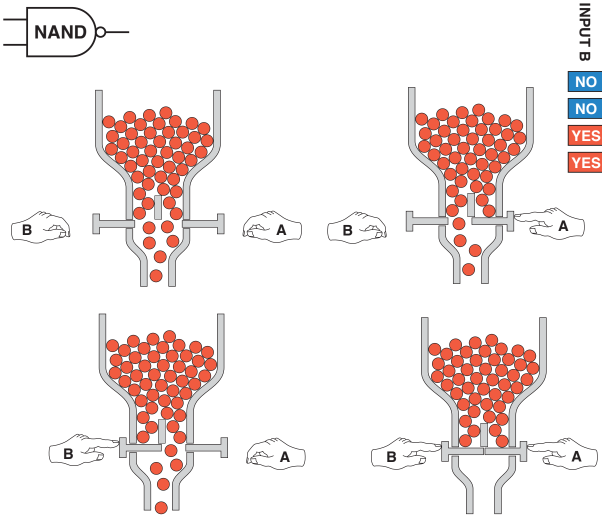

INPUT B	INPUT A	OUTPUT
NO	NO	YES
NO	YES	YES
YES	NO	YES
YES	YES	NO

Figure 4-59

ESSENTIALS

Logic gate basics (continued)

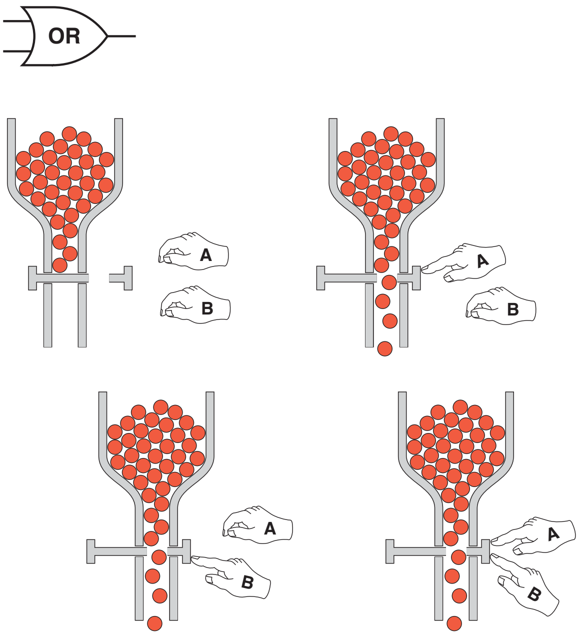
OR
A
B
A
B
A
B
A
B

Figure 4-80

INPUT B	INPUT A	OUTPUT
NO	NO	NO
NO	YES	YES
YES	NO	YES
YES	YES	YES

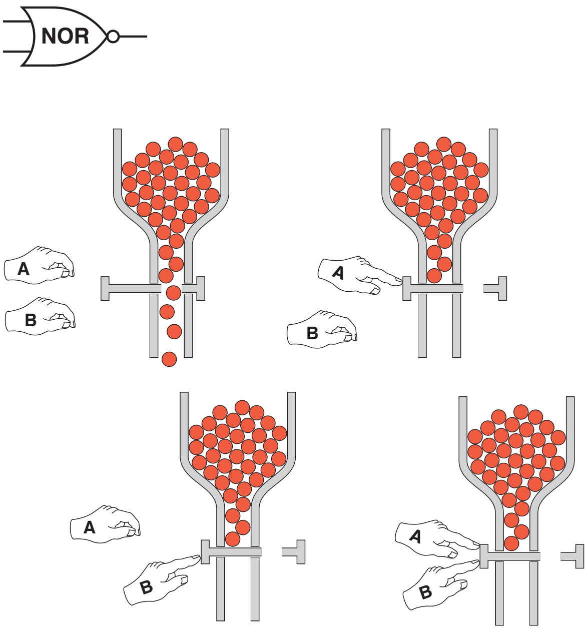
NOR
A
B
A
B
A
B
A
B

Figure 4-81

INPUT B	INPUT A	OUTPUT
NO	NO	YES
NO	YES	NO
YES	NO	NO
YES	YES	NO

ESSENTIALS

Logic gate basics (continued)

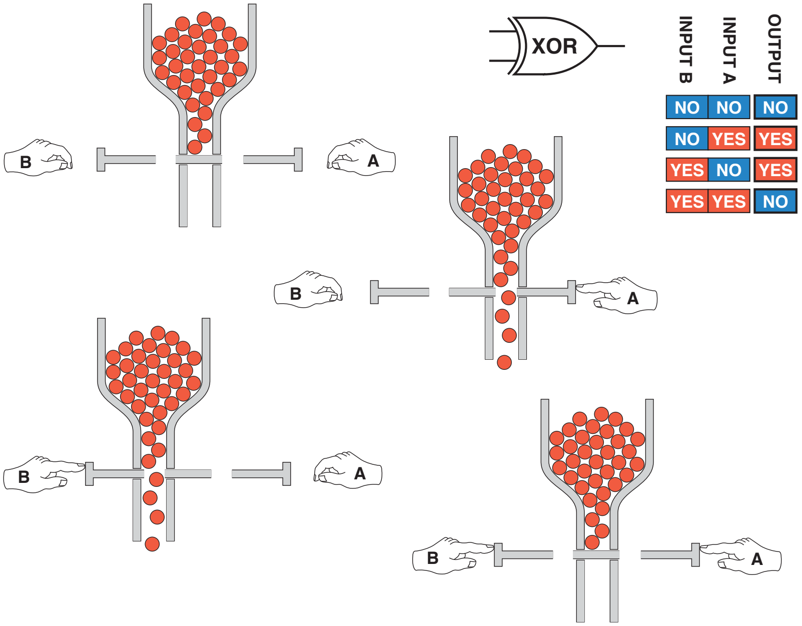
XOR
B
A
B
A
B
A
B
A

INPUT B	INPUT A	OUTPUT
NO	NO	NO
NO	YES	YES
YES	NO	YES
YES	YES	NO

Figure 4-62

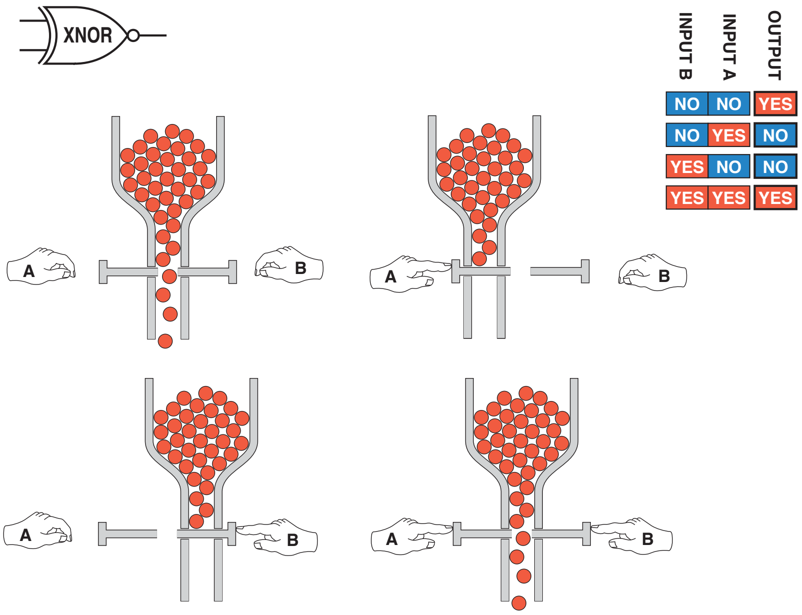
XNOR
A
B
A
B
A
B
A
B

INPUT B	INPUT A	OUTPUT
NO	NO	YES
NO	YES	NO
YES	NO	NO
YES	YES	YES

Figure 4-63

BACKGROUND

The confusing world of TTL and CMOS

Back in the 1960s, the first logic gates were built with Transistor-Transistor Logic, abbreviated TTL, meaning that tiny bipolar transistors were etched into a single wafer of silicon. Soon, these were followed by Complementary Metal Oxide Semiconductors, abbreviated CMOS. Each of these chips was a collection of metal-oxide field-effect transistors (known as MOSFETs). The 4026 chip that you used earlier is an old CMOS.

You may remember that bipolar transistors amplify current. TTL circuits are similar: they are sensitive to current, rather than voltage. Thus they require a significant flow of electricity, to function. But CMOS chips are like the programmable unijunction transistor that I featured previously. They are voltage-sensitive, enabling them to draw hardly any current while they are waiting for a signal, or pausing after emitting a signal.

The two families named TTL and CMOS still exist today. The table in Figure 4-64 summarizes their basic advantages and disadvantages. The CMOS series, with part numbers from 4000 upward, were easily damaged by static electricity but were valuable because of their meager power consumption. The TTL series, with part numbers from 7400 upward, used much more power but were less sensitive and very fast. So, if you wanted to build a computer, you used the TTL family, but if you wanted to build a little gizmo that would run for weeks on a small battery, you used the CMOS family.

From this point on everything became extremely confusing, because CMOS manufacturers wanted to grab market share by emulating the advantages of TTL chips. Newer generations of CMOS chips even changed their part numbers to begin with "74" to emphasize their compatibility, and the functions of pins on CMOS chips were swapped around to match the functions of pins on TTL chips. Consequently, the pinouts of CMOS and TTL chips are usually now identical, but the meaning of "high" and "low" states changed in each new generation, and the maximum supply voltages for CMOS chips were revised downward. Note I have included question marks beside two categories in the CMOS column, as modern CMOS chips have overcome those disadvantages—at least to some extent.

Here's a quick summary, which will be useful to you if you look at a circuit that you find online, and you wonder about the chips that have been specified.

Where you see a letter "x," it means that various numbers may appear in that location. Thus "74xx" includes the 7400 NAND gate, the 7402 NOR gate, the 74150 16-bit data selector, and so on. A combination of letters preceding the "74" identifies the chip manufacturer, while letters following the part number may identify the style of package, may indicate whether it contains heavy metals that are environmentally toxic, and other details.

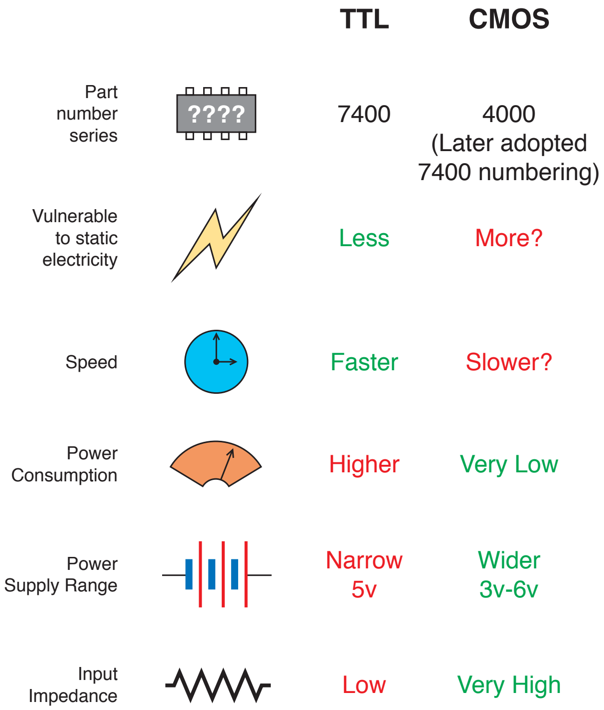

	TTL	CMOS
Part number series	7400	4000 (Later adopted 7400 numbering)
Vulnerable to static electricity	Less	More?
Speed	Faster	Slower?
Power Consumption	Higher	Very Low
Power Supply Range	Narrow 5v	Wider 3v-6v
Input Impedance	Low	Very High

Figure 4-64. *The basic differences between the two families of logic chips. In successive generations, these differences have gradually diminished.*

TTL family:

74xx
The old original generation, now obsolete.

74Sxx
Higher speed "Schottky" series, now obsolete.

74LSxx
Lower power Schottky series, still used occasionally.

74ALSxx
Advanced low-power Schottky.

74Fxx
Faster than the ALS series.

BACKGROUND

The confusing world of TTL and CMOS (continued)

CMOS family:

40xx
The old original generation, now obsolete.

40xxB
The 4000B series was improved but still susceptible to damage from static electricity. Many hobby circuits still use these chips because they will run from relatively high voltages and can power LEDs and even small relays directly.

74HCxx
Higher-speed CMOS, with part numbers matching the TTL family, and pinouts matching the TTL family, but input and output voltages not quite the same as the TTL family. I've used this generation extensively in this book, because it's widely available, and the circuits here have no need for greater speed or power.

74HCTxx
Like the HC series but matching the TTL voltages.

74ACxx
Advanced version of HC series. Faster, with higher output capacity.

74ACTxx
Like the AC series but with the same pin functions and voltages as TTL.

74AHCxx
Advanced higher-speed CMOS.

74AHCTxx
Like the AHC series but with the same pin functions and voltages as TTL.

74LVxx
Lower voltage (3.3v) versions, including LV, LVC, LVT, and ALVC series.

As you can see, these days you have to interpret part numbers very carefully. But which family and generation of chips should you use? Well, that depends! Following are some guidelines.

What you don't need:

1. Speed differences are irrelevant from our point of view, as we're not going to be building circuits running at millions of cycles per second.
2. Price differences are so small as to be inconsequential.
3. Lower-voltage (LV) CMOS chips are not very interesting for our small experimental circuits.
4. Try to avoid mixing different families, and different generations of the same family, in the same circuit. They may not be compatible.
5. Some modern chip varieties may be only available in the surface-mount package format. Because they're so much more difficult to deal with, and their only major advantage is miniaturization, I don't recommend them.
6. In the TTL family, the LS and ALS series cannot handle as much output current as the S series and the F series. You don't need them.

What you should use:

1. The old 74LSxx series of TTL chips was so popular, you'll still find schematics that specify these chips. You should still be able to buy them from sources online, but if not, you can substitute the 74HCTxx chips, which are designed to function identically.
2. The old 4000B series of CMOS chips are still used by hobbyists because their willingness to tolerate high voltages is convenient. While TTL or TTL-compatible chips require a carefully regulated 5 volts, the 4000B chips would handle 15 volts—and also delivered enough power to energize LEDs or even very small relays. Some hobbyists also have a nostalgic affection for the 74Cxx series of chips, which had the same pin connections as the TTL chips but could still tolerate higher voltages and higher output current. The trouble is, some of the 74Cxx chips are almost extinct, and while the 4000B chips are still available, they are considered almost obsolete.

BACKGROUND

The confusing world of TTL and CMOS (continued)

Bottom line: I suggest you use the 4000B chips only if you want to replicate an old circuit, or if a modern equivalent is unavailable (which is why I specified the 4026B chip for the reaction timer—I could not find a modern equivalent that will drive seven-segment numeric displays directly, and I didn't want you to have to deal with more parts than necessary).

If you check online suppliers such as *Mouser.com* you'll find that the HC family is by far the most popular right now. They are all available in through-hole format (to fit your breadboard and perforated board). They have the high input impedance of CMOS (which is useful) and they have the same pin identities as the old 74LSxx series.

Abbreviations

When looking at data sheets, you are likely to encounter some or all of these abbreviations:

- VOH min: Minimum output voltage in high state
- VOL max: Maximum output voltage in low state
- VIH min: Minimum input voltage to be recognized as high
- VIL max: Maximum input voltage recognized as low

BACKGROUND

Logic gate origins

The 7400 family of integrated circuits was introduced by Texas Instruments, beginning with the 7400 NAND gate in 1962. Other companies had sold logic chips previously, but the 7400 series came to dominate the market. The Apollo lunar missions used a computer built with 7400 chips, and they were a mainstay of minicomputers during the 1970s.

RCA introduced its 4000 series of logic chips in 1968, built around CMOS transistors; Texas Instruments had chosen TTL. The CMOS chips used less power, thus generating much less heat and enabling flexible circuit design, as each chip could power many others. CMOS was also tolerant of wide voltage ranges (from 3 to 15 volts) but prohibited switching speeds faster than around 1MHz. TTL was 10 times faster.

Design tweaks gradually eradicated the speed penalty for CMOS, and TTL chips have become relatively rare. Still, some people retain a special nostalgic loyalty to "the logic gates that went to the moon." A hardcore enthusiast named Bill Buzbee has built an entire web server from TTL-type 7400 chips, currently online at *http://magic-1.org*. Figure 4-65 shows just one of the handmade circuit boards that Bill assembled to run his computer.

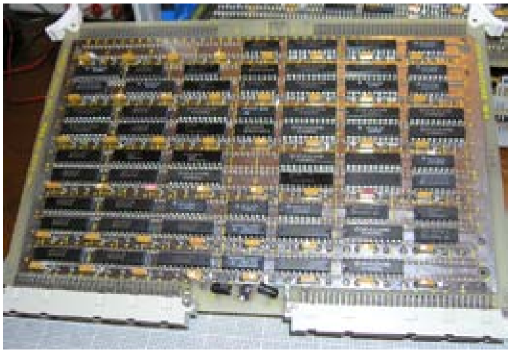

Figure 4-65. *Hobbyist Bill Buzbee built himself a web server entirely from 7400 series logic chips, the oldest of which was fabricated back in 1969. The web server can be found online at http://magic-1.org, displaying pictures of itself and details of its construction. The picture here that Bill took shows just one of the circuit boards of this remarkable project.*

FUNDAMENTALS

Common part numbers

Each 14-pin chip can contain four 2-input gates, three 3-input gates, two 4-input gates, one 8-input gate, or six single-input inverters, as shown in the following table.

	2 input	3 input	4 input	8 input
AND	7408	7411	7421	
NAND	7400	7410	7420	7430
OR	7432			744078*
NOR	7402	7427		744078*
XOR	7486			
XNOR	747266			
Inverter	(1 input) 7404			

*The 744078 has an OR output and a NOR output on the same chip.

Figures 4-66 through 4-74 show the internal connections of the logic chips that you are most likely to use. Note that the 7402 NOR gate has its logical inputs and outputs arranged differently from all the other chips.

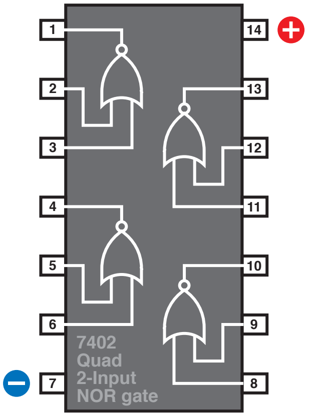
1
2
3
4
5
6
7
14
13
12
11
10
9
8
7402
Quad
2-Input
NOR gate

Figure 4-66. *Figures 4-66 through 4-74 show pinouts for some of the most widely used logic chips. Note that the inputs of the 7402 are reversed compared with the other chips.*

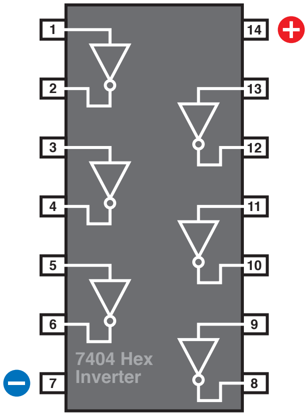
1
2
3
4
5
6
7
14
13
12
11
10
9
8
7404 Hex
Inverter

Figure 4-67

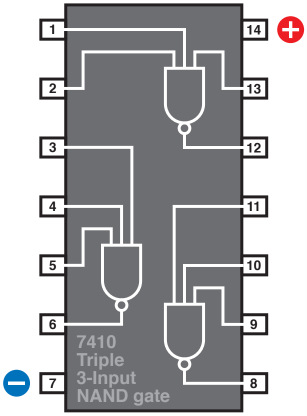
1
2
3
4
5
6
7
14
13
12
11
10
9
8
7410
Triple
3-Input
NAND gate

Figure 4-68

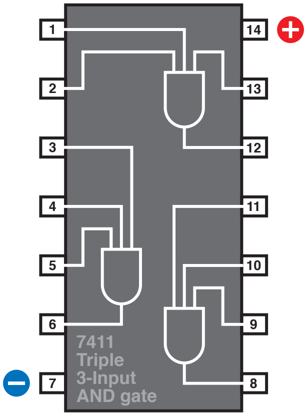
1
2
3
4
5
6
7
14
13
12
11
10
9
8
7411
Triple
3-Input
AND gate

Figure 4-69

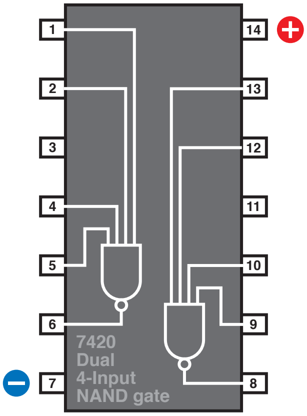
1
2
3
4
5
6
7
14
13
12
11
10
9
8
7420
Dual
4-Input
NAND gate

Figure 4-70

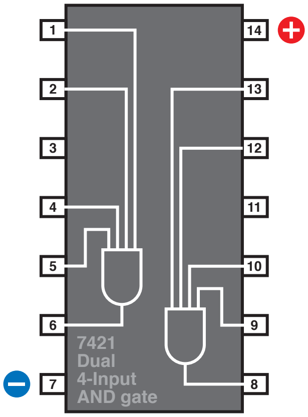
1
2
3
4
5
6
7
14
13
12
11
10
9
8
7421
Dual
4-Input
AND gate

Figure 4-71

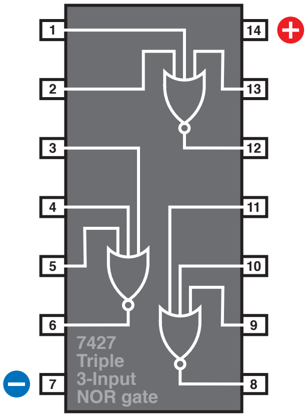
1
2
3
4
5
6
7
14
13
12
11
10
9
8
7427
Triple
3-Input
NOR gate

Figure 4-72

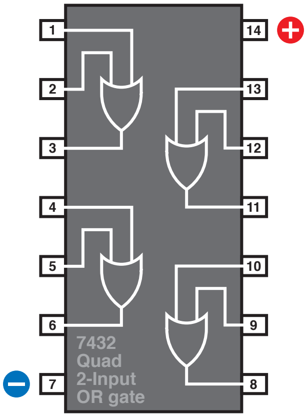
1
2
3
4
5
6
7
14
13
12
11
10
9
8
7432
Quad
2-Input
OR gate

Figure 4-73

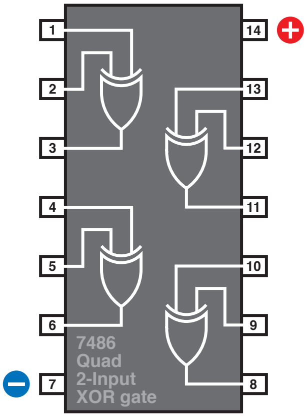
1
2
3
4
5
6
7
14
13
12
11
10
9
8
7486
Quad
2-Input
XOR gate

Figure 4-74

FUNDAMENTALS

Rules for connecting logic gates

Permitted:

- You can connect the input of a gate directly to your regulated power supply, either positive side or negative side.
- You can connect the output from one gate directly to the input of another gate.
- The output from one gate can power the inputs of many other gates (this is known as "fanout"). The exact ratio depends on the chip, but you can always power at least ten inputs with one logic output. The output from a logic chip can drive the trigger (pin 3) of a 555 timer. The output from the timer can then deliver 100mA, easily enough for half-a-dozen LEDs or a small relay.
- Low input doesn't have to be zero. A 74HCxx logic gate will recognize any voltage up to 1 volt as "low."
- High input doesn't have to be 5 volts. A 74HCxx logic gate will recognize any voltage above 3.5 volts as "high."

See Figures 4-75 and 4-76 for a comparison of permitted voltages on the input and output side of 74HCxx and 74LSxx chips.

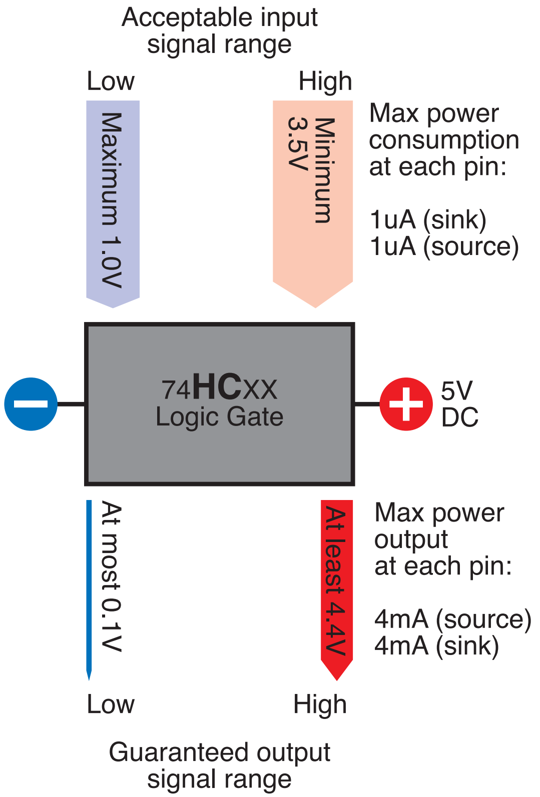
Acceptable input signal range
Low
High
Maximum 1.0V
Minimum 3.5V
Max power consumption at each pin:
1uA (sink)
1uA (source)
74HCxx Logic Gate
5V DC
At most 0.1V
At least 4.4V
Max power output at each pin:
4mA (source)
4mA (sink)
Low
High
Guaranteed output signal range

Figure 4-75. *Each family of logic chips, and each generation in each family, has different standards for input and output minimum and maximum voltages. This diagram shows the standards used by the HC generation of the CMOS family, which was chosen for most of the projects in this book. Note that the current required for input is minimal compared with the current available for output. The power supply to the chip makes up the difference.*

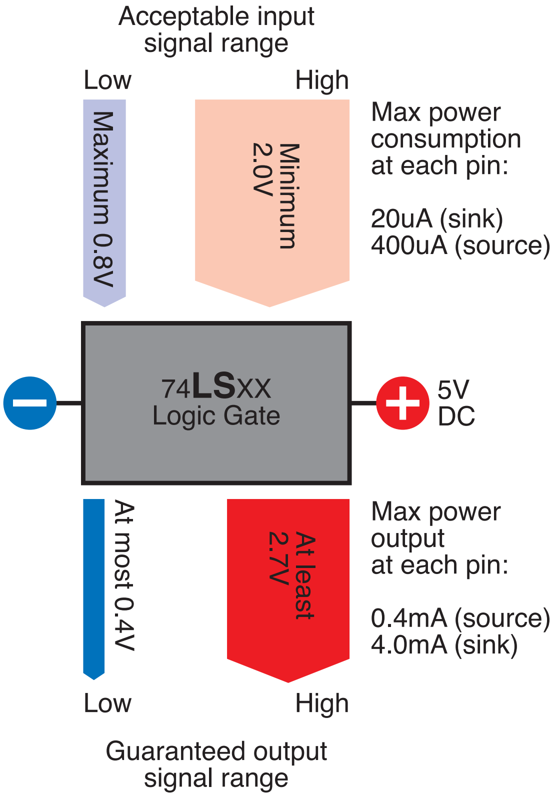
Acceptable input signal range
Low
High
Maximum 0.8V
Minimum 2.0V
Max power consumption at each pin:
20uA (sink)
400uA (source)
74LSxx Logic Gate
5V DC
At most 0.4V
At least 2.7V
Max power output at each pin:
0.4mA (source)
4.0mA (sink)
Low
High
Guaranteed output signal range

Figure 4-76. *Because the LS generation of the TTL family has such different tolerances for input voltages and different standards for output voltages, the LS generation of TTL chips should not be mixed in the same circuit as the HC generation of CMOS chips, unless pull-up resistors are used to bring the LS chips into conformance with standards expected by the HC chips. See Experiment 21 for a case study in using LS chips.*

FUNDAMENTALS

Rules for connecting logic gates (continued)

Not permitted:

- No floating-input pins! On CMOS chips such as the HC family, you must always connect all input pins with a known voltage, even if they supply a gate on the chip that you're not using. When you use a SPST switch to control an input, remember that in its "off" position, it leaves the input unconnected. Use a pull-up or pull-down resistor to prevent this situation. See Figure 4-77.
- Don't use an unregulated power supply, or more than 5 volts, to power 74HCxx or 74LSxx logic gates.
- Be careful when using the output from a logic gate to power even a low-current LED. Check how many milliamps are being drawn. Also be careful when "sharing" the output from a logic gate with the input of another gate, at the same time that it is driving an LED. The LED may pull down the output voltage, to a point where the other gate won't recognize it. Always check currents and voltages when modifying a circuit or designing a new one.
- Never apply a significant voltage or current to the output pin of a logic gate. In other words, don't force an input into an output.
- Never link the outputs from two or more logic gates. If they must share a common output wire, use diodes to protect them from each other. See Figure 4-78.

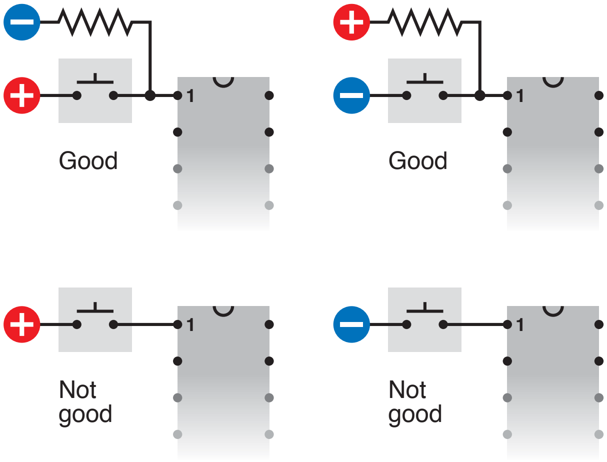

Figure 4-77. *Because a CMOS chip is so sensitive to input fluctuations, a logical input should never be left "floating," or unattached to a defined voltage source. This means that any single-throw switch or pushbutton should be used with a pull-up or pull-down resistor, so that when the contacts are open, the input is still defined.*

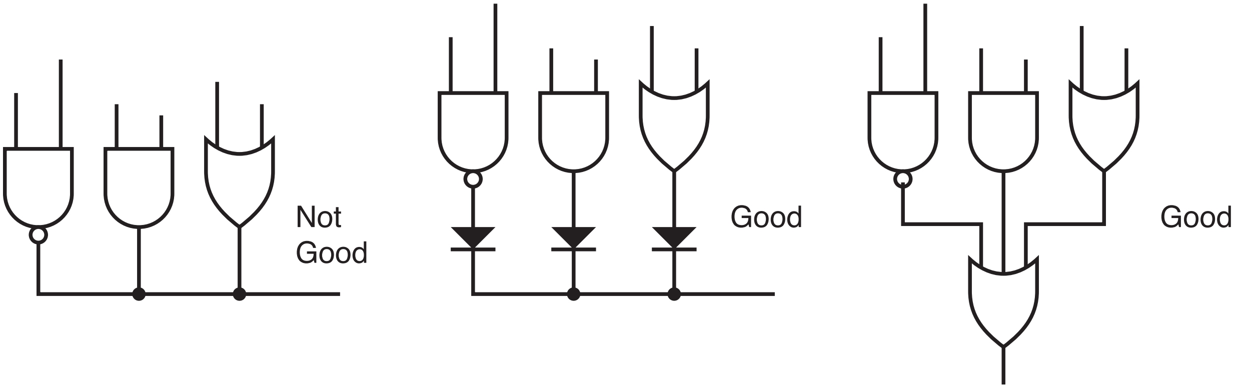

Figure 4-78. *The output from one logic gate must not be allowed to feed back into the output from another logic gate. Diodes can be used to isolate them, or they can be linked via another gate.*

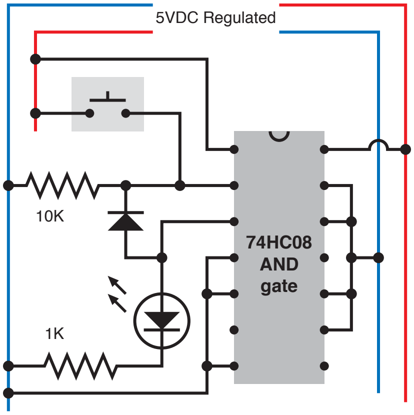

Figure 4-79. *Using a diode, the logical output from a gate can be allowed to feed back to one of its inputs, so that the gate latches after receiving a brief logical input pulse.*

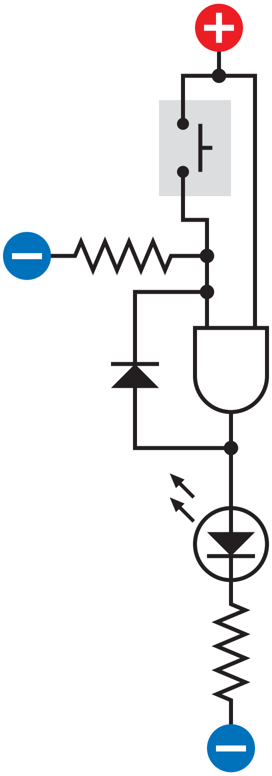

Figure 4-80. *The breadboard-format schematic in is simplified here to show more clearly the way in which a gate can latch itself after receiving an input pulse.*

In the 74HCxx logic family, each input of a logic gate consumes just a microamp, while the output can source 4 milliamps. This seems paradoxical: how can the chip give out more than it takes in? The answer is that it also consumes power from the power supply attached to pins 7 and 14. That's where the additional electricity comes from.

Because the logical output from a chip can be greater than the logical input, we can put the chip in a state where it keeps itself "switched on" in a way which is similar to the way the relay in the alarm project was wired to lock itself on. The simplest way to do this in a logic chip is by feeding some of the output back to one of the inputs.

Figure 4-79 shows an AND gate with one of its inputs wired to positive and its other input held low by a pull-down resistor, with a pushbutton that can make the input high. A signal diode connects the output of the chip back to the pushbutton-controlled input. Remember that the diode has a mark on it indicating the end which should be connected to the *negative* side of the power supply, which in this case will be the end of the 10K resistor.

The schematic in Figure 4-79 shows how the circuit should look in breadboard format. Figure 4-80 shows it in a simpler format.

From this point on, I won't bother to show the power regulator and the capacitors associated with it. Just remember to include them every time you see the power supply labeled as "5V DC Regulated."

When you switch on the power, the LED is dark, as before. The AND gate needs a positive voltage on both of its logical inputs, to create a positive output, but it now has positive voltage only on one of its inputs, while the other input is pulled down by the 10K resistor. Now touch the pushbutton, and the LED comes on. Let go of the pushbutton, and the LED stays on, because the positive output from the AND gate circulates back through the diode and is high enough to overcome the negative voltage coming through the pull-down resistor.

The output from the AND gate is powering its own input, so the LED will stay on until we disconnect it. This arrangement is a simple kind of "latch," and can be very useful when we want an output that continues after the user presses and releases a button.

You can't just connect the output from the gate to one of its inputs using an ordinary piece of wire, because this would allow positive voltage from the tactile switch to flow around and interfere with the output signal. Remember, you must never apply voltage to the output pin of a logic gate. The diode prevents this from happening.

If you've grasped the basics of logic gates, you're ready now to continue to our first real project, which will use all the information that I've set out so far.

Experiment 20: A Powerful Combination

Suppose you want to prevent other people from using your computer. I can think of two ways to do this: using software, or using hardware. The software would be some kind of startup program that intercepts the normal boot sequence and requests a password. You could certainly do it that way, but I think it would be more fun (and more relevant to this book) to do it with hardware. What I'm imagining is a numeric keypad requiring the user to enter a secret combination before the computer can be switched on.

You will need:

- Numeric keypad. As specified in the shopping list at the beginning of this chapter, it must have a "common terminal" or "common output." The schematic in Figure 4-82 shows what I mean. Inside the keypad, one conductor (which I have colored red to distinguish it from the others) connects with one side of every pushbutton. This conductor is "common" to all of them. It emerges from the keypad on an edge connector or set of pins at the bottom, which I've colored yellow.

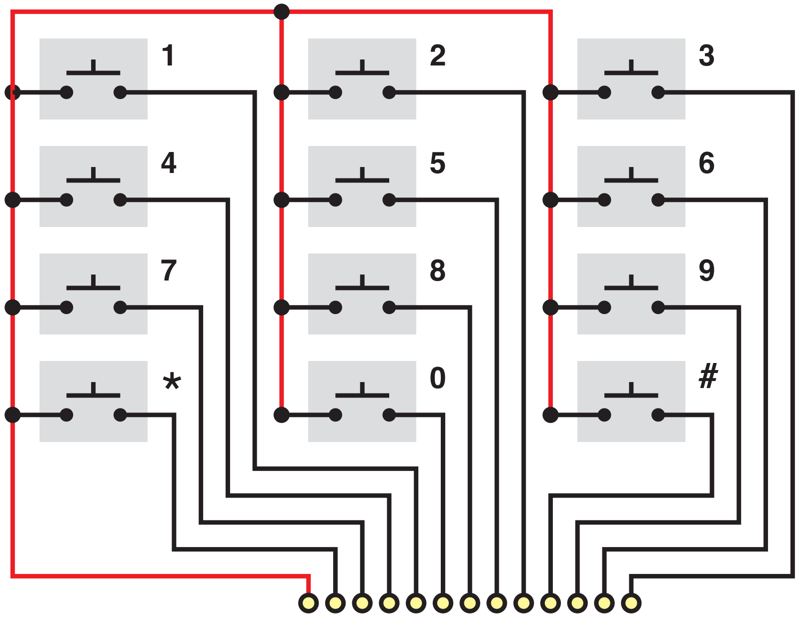

Figure 4-82. *A keypad of the type required for Experiment 20 incorporates a common terminal connected to one side of each of the 12 pushbuttons. The wire from the common terminal is shown red, here, to make it more easily identifiable.*

- Keypads that use "matrix encoding" won't work with the circuit that I'm going to describe. If the Velleman keypad, which I recommend, is unavailable, and you can't find another like it, you can use 12 separate SPST pushbuttons. Of course, that will cost a little more.
- 74HC08 logic chip containing four AND gates. Quantity: 1.
- 74HC04 logic chip containing six inverters. Quantity: 1.
- 555 timer chip. Quantity: 1.

The Warranty Issue

If you follow this project all the way to its conclusion, you'll open your desktop computer, cut a wire, and saw a hole in the cabinet. Without a doubt, this will void your warranty. If this makes you nervous, here are three options:

1. Breadboard the circuit for fun, and leave it at that.

2. Use the numeric keypad on some other device.

3. Use it on an old computer.

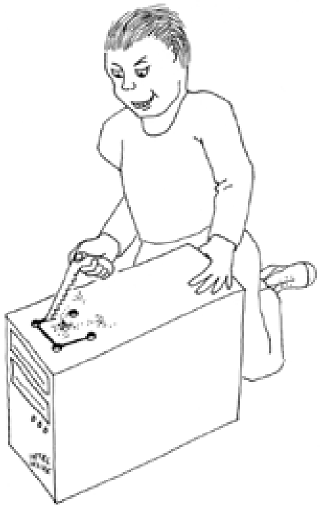

Figure 4-81. *Caution: This just might void your warranty.*

- Latching relay, 5 volt, DPST or DPDT, "2 form C" package, Panasonic DS2E-SL2-DC5V or similar. Must have two separate coils (one to latch, one to unlatch) with separate inputs. Quantity: 1.
- LEDs, 5mm generic, your choice of colors. Quantity: 3.
- Ribbon cable, with six conductors minimum, if you want to do a really neat job. You can use a cable of the type sold for hard drives, and split off the six conductors that you need, or shop around on eBay.
- Tools to open your computer, drill four holes, and make saw cuts between the holes, to create a rectangular opening for the keypad (if you want to take this project to its conclusion). Also, four small bolts to attach the keypad to the computer cabinet after you create the opening for it.

The Schematic

This time I'd like you to study the schematic before building anything. Let's start with the simplified version, shown in Figure 4-83.

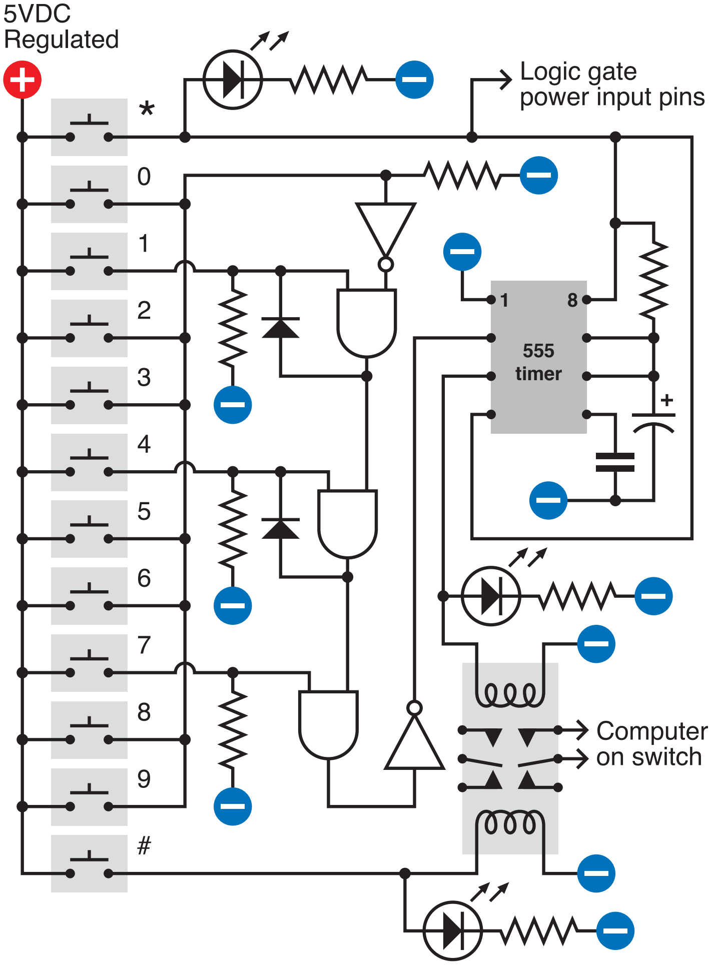

Figure 4-83. *A simplified schematic showing the basic structure of the combination lock circuit.*

I want this to be a battery-powered circuit, so that you don't have to run a separate power supply to it or (worse) try to tap into your computer's 5-volt bus. Battery power means that the circuit has to be "off" most of the time, to prevent the battery from running down. Because the keypad has two spare buttons (the asterisk and the pound sign), I'm going to use the asterisk as the "power on" button. When you press it, the LED at the top of the schematic lights up to confirm that everything's working, and the button sends power to the two logic chips and the 555 timer. You have to hold down the asterisk button while you punch in a three-digit code to unlock the computer.

Arbitrarily, I've chosen 1-4-7 as the three-digit code. Let's track what happens when you enter this sequence. (Naturally, if you build the circuit, you can wire it to choose any three digits you prefer.)

Pressing the 1 button sends positive power to one logical input of the first AND gate. The other logical input of this gate is also positive, because an inverter is supplying it, and the input of the inverter is being held negative by a pull-down resistor. When an inverter has a negative input, it gives a positive output, so pressing the 1 button activates the AND gate, and makes its output positive. The AND gate locks itself on, as its output cycles back to its switched input via a diode. So the gate output remains high even after you let go of the 1 button.

The output from the first AND gate also supplies one logical input of the second AND gate. When you press the 4 button, you send positive voltage to the other logical input of this AND gate, so its output goes high, and it locks itself on, just as the first gate did.

The second AND gate feeds the third AND gate, so when you press the 7 button, the third AND gate changes its output from low to high. This passes through an inverter, so the output from the inverter goes from high to low. This in turn goes to the trigger of a 555 timer wired in monostable mode.

When the trigger of a 555 timer goes from high to low, the timer emits a positive pulse through its output, pin 3. This runs down to the upper coil of the latching relay, and also flashes an LED to confirm that the code has been accepted and the relay has been activated.

Two of the contacts in the relay are wired into the power-up button of your computer. A little later in this description I'll explain why this should be safe with any modern computer.

Because we're using a latching relay, it flips into its "on" state and remains there, even when the power pulse from the timer ends. So now you can let go of the asterisk button to disconnect the battery power to your combination lock, and press the power-up button that switches on your computer.

At the end of your work session, you shut down your computer as usual, then press the pound button on your keypad, which flips the relay into its other position, reactivating the combination lock.

Incorrect Inputs

What happens if you enter the wrong code? If you press any button other than 1, 4, or 7, it sends positive voltage to the inverter near the top of the schematic. The positive voltage overwhelms the negative voltage being applied to the inverter through a pull-down resistor, and causes the inverter to output a negative voltage, which it applies to one of the logical inputs of the first AND gate. If the AND gate was locked on, the negative input will switch it off. If it was supplying the second AND gate, it'll switch that one off too.

Thus, any error when entering the first, second, or third digit of the secret code will reset the AND gates, forcing you to begin the sequence all over again.

What if you enter 1, 4, and 7 out of their correct sequence? The circuit won't respond. The third AND gate needs a high input supplied by the second AND gate, and the second AND gate needs a high input supplied by the first AND gate. So you have to activate the AND gates in the correct sequence.

Questions

Why did I use a 555 timer to deliver the pulse to the relay? Because the logical output from an AND gate cannot deliver sufficient power. I could have passed it through a transistor, but I liked the idea of a pulse of a fixed length to flip the relay and illuminate an LED for about 1 second, regardless of how briefly the user presses the 7 button.

Why do I need three LEDs? Because when you're punching buttons to unlock your computer, you need to know what's going on. The Power On LED reassures you that your battery isn't dead. The Relay Active LED tells you that the system is now unlocked, in case you are unable to hear the relay click. The System Relocked LED reassures you that you have secured your computer.

Because all the LEDs are driven either directly from the 5-volt supply or from the output of the 555 timer, they don't have to be low-current LEDs and can be used with 330Ω series resistors, so they'll be nice and bright.

How do you connect the keypad with the circuit? That's where your ribbon cable comes in. You carefully strip insulation from each of the conductors, and solder them to the contact strip or edge connector on your keypad. Push the conductors on the other end of the cable into your breadboard (when you're test-building the circuit) or solder them into perforated board (when you're building it permanently). Find a convenient spot inside your computer case where you can attach the perforated board, with double-sided adhesive or small bolts or whatever is convenient. Include a 9-volt battery carrier, and don't forget your power regulator to step the voltage down to 5 volts.

Breadboarding

No doubt you have realized by now that breadboards are very convenient as a quick way to push in some components and create connections, but the layout of their conductors forces you to put components in unintuitive configurations. Still, if you carefully compare the breadboard schematic in Figure 4-83 with the simplified schematic in Figure 4-84, you'll find that the connections are the same.

To help it make sense, I've shown the logic gates that exist inside the chips. I've also colored the power supply wires, as before, to reduce the risk of confusion. The positive side of the supply goes only to the common terminal on your keypad, and you have to press the asterisk key to send the power back down the ribbon cable, to supply the chips.

If you build the circuit and you can't understand why everything's dead, it's most likely because you forgot to hold down the asterisk button.

Note that the "wrong" numbers on the keypad are all shorted together. This will create some inconvenience if you want to change the combination in the future. I'll suggest a different option in the "enhancements" section that follows. For now, ideally, you should run a wire from every contact on your keypad, down to your circuit on its breadboard, and short the "wrong" keypad numbers together with jumper wires on the breadboard.

Also note that if you use a meter to test the inputs to the AND gates, and you touch your finger against the meter probe while doing so, this can be sufficient to trigger the sensitive CMOS inputs and give a false positive.

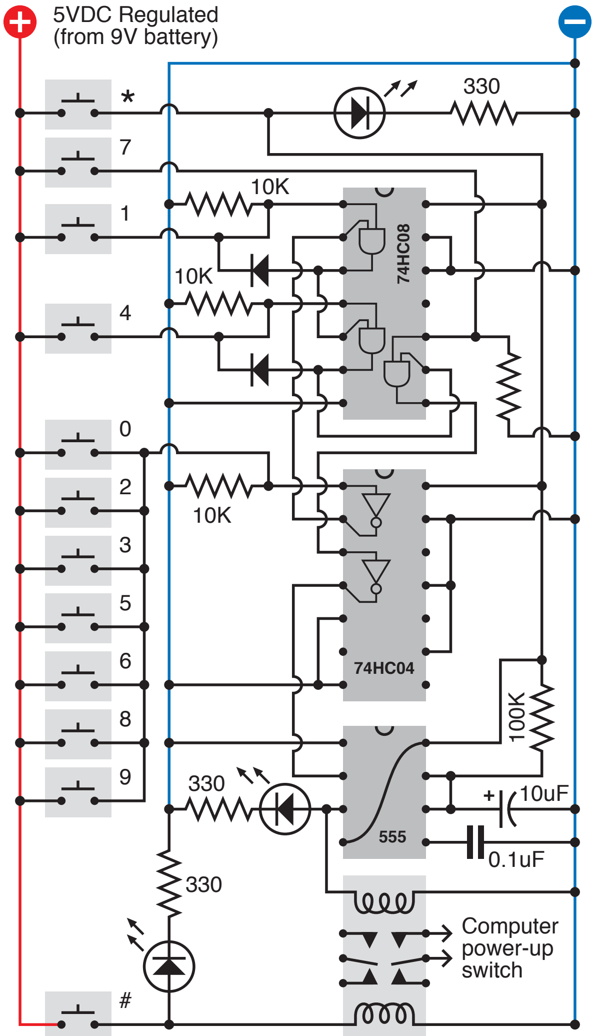

Figure 4-84. *The combination lock schematic redrawn to show how the components can be laid out on a breadboard.*

One Little Detail: The Computer Interface

Old computers used to have a big switch at the back, attached to the heavy metal box inside the computer, that transformed house current to regulated voltages that the computer needs. Most modern computers are not designed this way; you leave the computer plugged in, and you touch a little button on the box (if it's a Windows machine) or the keyboard (if it's a Mac), which sends a low-voltage pulse to the motherboard.

This is ideal from our point of view, because we don't have to mess with high voltages. Don't even think of opening that metal box with the fan mounted in it, containing the computer power supply. Just look for the wire (usually containing two conductors, on a Windows machine) that runs from the "power up" button to the motherboard.

To check that you found the right one, *make sure that your computer is unplugged*, ground yourself (because computers contain CMOS chips that are sensitive to static electricity) and very carefully snip just one of the two conductors in the wire. Now plug in your computer and try to use the "power up" button. If nothing happens, you've probably cut the right wire. Even if you cut the wrong wire, it still prevented your computer from booting, which is what you want, so you can use it anyway. Remember, we are not going to introduce any voltage to this wire. We're just going to use the relay as a switch to reconnect the conductor that you cut. You should have no problem if you maintain a cool and calm demeanor, and look for that single wire that starts everything. Check online for the maintenance manual for your computer if you're really concerned about making an error.

After you find the wire and cut just one of its conductors, unplug your computer again, and keep it unplugged during the next steps.

Find where the wire attaches to the motherboard. Usually there's a small unpluggable connector. First, mark it so that you know how to plug it back in the right way around, and then disconnect it while you follow the next couple of steps.

Strip insulation from the two ends of the wire that you cut, and solder an additional piece of two-conductor wire, as shown in Figure 4-85, with heat-shrink tube to protect the solder joints. (This is very important!)

Run your new piece of wire to the latching relay, making sure you attach it to the pair of contacts which close, inside the relay, when it is energized by the unlocking operation. You don't want to make the mistake of unlocking your computer when you think you're locking it, and vice versa.

Reconnect the connector that you disconnected from your motherboard, plug in your computer, and try to power it up. If nothing happens, this is probably good! Now enter the secret combination on your keypad (while holding down the asterisk button to provide battery power) and listen for the click as the relay latches. Now try the "power up" button again, and everything should work.

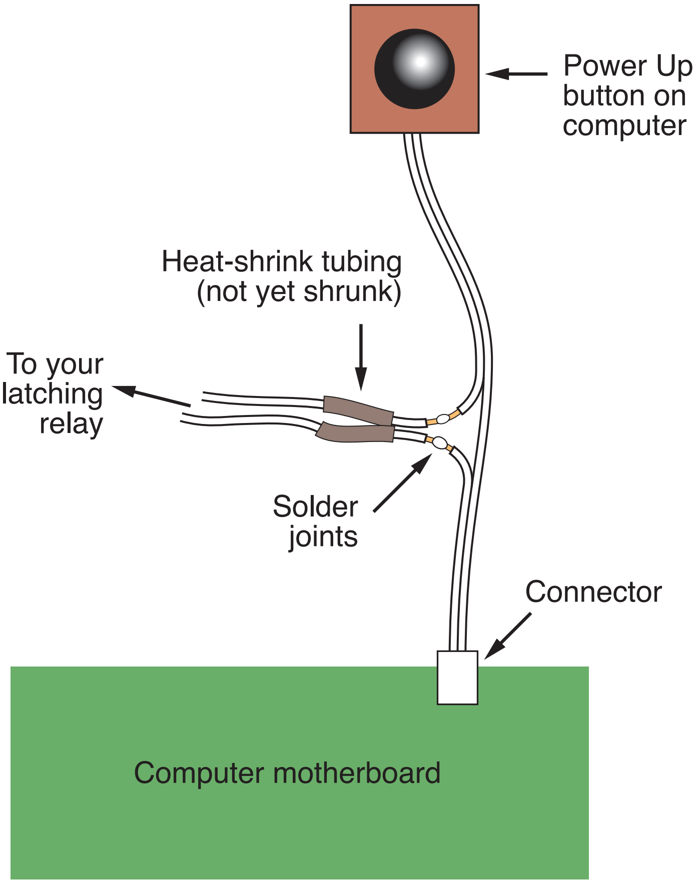
Power Up
button on
computer
Heat-shrink tubing
(not yet shrunk)
To your
latching
relay
Solder
joints
Connector
Computer motherboard

Figure 4-85. *The combination lock project can be interfaced with a typical desktop computer by cutting one conductor in the wire from the "power up" pushbutton, soldering an extension, and covering the joints with heat-shrink tube.*

Enhancements

At the end of any project, there's always more you can do.

To make this setup more secure, you could remove the usual screws that secure the case of the computer, and replace them with tamper-proof screws. Check any online source for "tamper-proof screws," such as *http://www.mcmaster.com*. Naturally, you will also need the special tool that fits the screws, so that you can install them (or remove them, if your security system malfunctions for any reason).

Another enhancement could be an additional 555 timer that is activated by the asterisk button, and delivers power to the other chips for, say, a limited period of 30 seconds, allowing you that much time to unlock the system. This would eliminate the need to hold down the asterisk button while you enter the unlocking code. A 555 timer can supply power to all the other chips, because they don't use very much. I omitted this feature for the sake of simplicity.

Yet another enhancement, if you are security-crazed, is to go for a four-button code. After all, the 74HC08 chip still has one unused AND gate. You could insert that into the chain of the existing AND gates and wire it to another keypad button of your choice.

Still another enhancement would be a way to change the code without unsoldering and resoldering wires. You can use the miniature sockets that I suggested in the heartbeat flasher project. These should enable you to swap around the ends of your wires from the keypad.

And for those who are absolutely, positively, totally paranoid, you could fix things so that entering a wrong code flips a second high-amperage relay which supplies a massive power overload, melting your CPU and sending a big pulse through a magnetic coil clamped to your hard drive, instantly turning the data to garbage (Figure 4-86). Really, if you want to protect information, messing up the hardware has major advantages compared with trying to erase data using software. It's faster, difficult to stop, and tends to be permanent. So, when the Record Industry Association of America comes to your home and asks to switch on your computer so that they can search for illegal file sharing, just accidentally give them an incorrect unlocking code, sit back, and wait for the pungent smell of melting insulation.

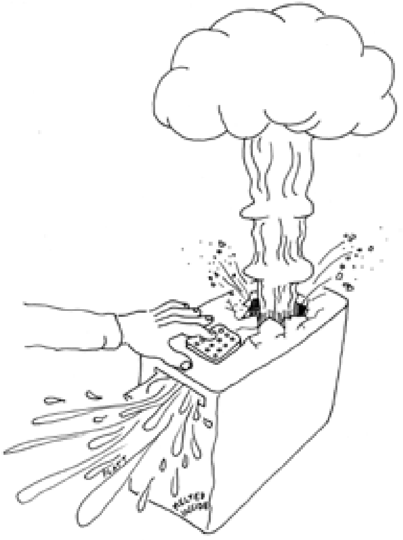

Figure 4-86. *For those who are absolutely, positively, totally paranoid: a meltdown/self-destruct system controlled by a secret key combination provides enhanced protection against data theft or intrusions by RIAA investigators asking annoying questions about file sharing.*

If you pursue this option, I *definitely* take no responsibility for the consequences.

On a more realistic level, no system is totally secure. The value of a hardware locking device is that if someone does defeat it (for instance, by figuring out how to unscrew your tamper-proof screws, or simply ripping your keypad out of the computer case with metal shears), at least you'll know that something happened—especially if you put little dabs of paint over the screws to reveal whether they've been messed with. By comparison, if you use password-protection software and someone defeats it, you may never know that your system has been compromised.

Experiment 21: Race to Place

The next project is going to get us deeper into the concept of feedback, where the output is piped back to affect the input—in this case, blocking it. It's a small project, but quite subtle, and the concepts will be useful to you in the future.

You will need:

- 74HC32 chip containing four OR gates. Quantity: 1.
- 555 timers. Quantity: 2.
- SPDT switch. Quantity: 1.
- SPST tactile switches. Quantity: 2.
- Various resistors.
- 5V supply using power regulator as before.

The Goal

On quiz shows such as *Jeopardy*, contestants race to answer each question. The first person who hits his answer button automatically locks out the other contestants, so that their buttons become inactive. How can we make a circuit that will do the same thing?

If you search online, you'll find several hobby sites where other people have suggested circuits to work this way, but they lack some features that I think are necessary. The approach I'm going to use here is both simpler and more elaborate. It's simpler because it has a very low chip count, but it's more elaborate in that it incorporates "quizmaster control" to make a more realistic game.

I'll suggest some initial ideas for a two-player version. After I develop that idea, I'll show how it could be expanded to four or even more players.

A Conceptual Experiment

I want to show how this kind of project grows from an idea to the finished version. By going through the steps of developing a circuit, I'm hoping I may inspire you to develop ideas of your own in the future, which is much more valuable than just replicating someone else's work. So join me in a conceptual experiment, thinking our way from a problem to a solution.

First consider the basic concept: two people have two buttons, and whoever goes first locks out the other person. I always find it helps me to visualize this kind of thing if I draw a sketch, so that's where I'll begin. In Figure 4-87, the signal from each button passes through a component that I'll call a "button blocker," activated by the other person's button. I'm not exactly sure what the button blocker will be or how it will work, yet.

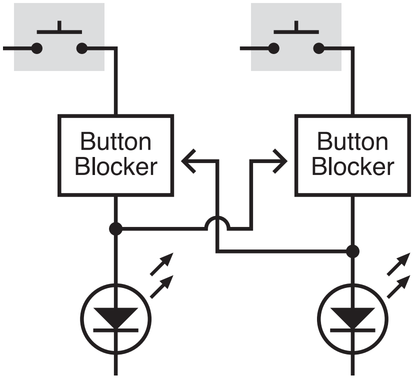

Figure 4-87. *The basic concept of the quiz project is that the output from one button should feed back to intercept the output of another button. At this point, the way in which the "button blocker" circuit works has not been figured out.*

Now that I'm looking at it, I see a problem here. If I want to expand this to three players, it will get complicated, because each player must activate the "button blockers" of *two* opponents. Figure 4-88 shows this. And if I have four players, it's going to get even more complicated.

Anytime I see this kind of complexity, I think there has to be a better way.

Also, there's another problem. After a player lets his finger off the button, the other players' buttons will be unblocked again. I need a latch to hold the signal from the first player's button and continue to block the other players.

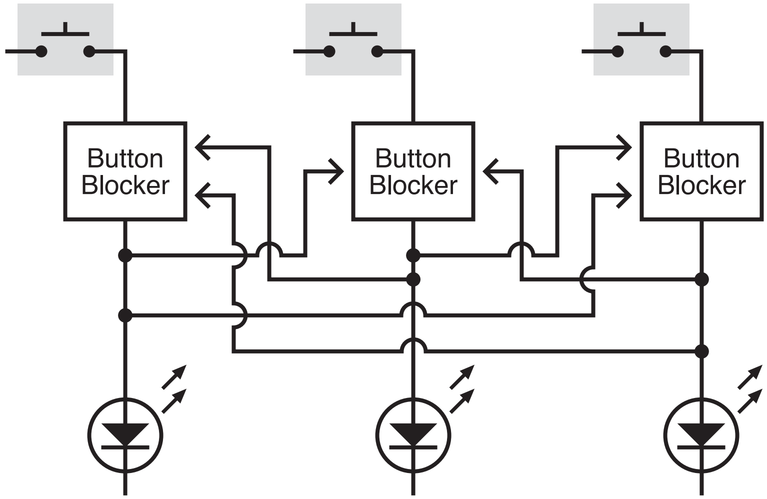

Figure 4-88. *The quiz concept becomes more complicated when an additional pushbutton is added. Now each button must block two other buttons. If a fourth button is added, the circuit will become unmanageably complex. There has to be a better way.*

This now sounds even more complicated. But wait a minute, if I have a latch which allows the winning player to take his finger off his button, I don't care if any of the buttons are being pressed anymore—including the button of the winning player. As soon as his signal is latched, *all* the buttons can be blocked. This makes things much simpler. I can summarize it as a sequence of events:

1. First player presses his button.
2. The signal is latched.
3. The latched signal feeds back and blocks all the buttons.

The new sketch in Figure 4-89 shows this. Now the configuration is modular, and can be expanded to almost any number of players, just by adding more modules.

There's something important missing, though: a reset switch, to put the system back to its starting mode after the players have had time to press their buttons and see who won. Also, I need a way to prevent players from pressing their buttons too soon, before the quizmaster has finished asking the question. Maybe I can combine this function in just one switch, which will be under the quizmaster's control. In its Reset position, the switch can reset the system and remove power to the buttons. In its Play position, the switch stops holding the system in reset mode, and provides power to the buttons. Figure 4-90 shows this. I've gone back to showing just two players, to minimize the clutter of lines and boxes, but the concept is still easily expandable.

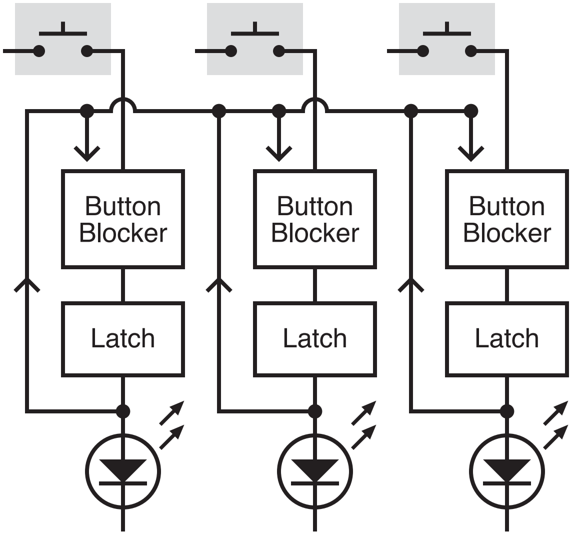

Figure 4-89. *If a latch is added below each button, it can retain one input and then block all inputs from all buttons. This simplifies the concept.*

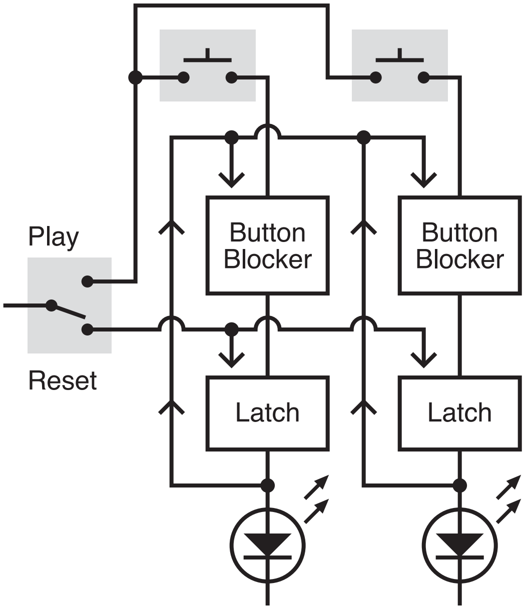

Figure 4-90. *A quizmaster switch will be needed to activate the buttons initially and then reset the circuit after a winning input has been recorded.*

Now I have to deal with a logic problem in the diagram. The way I've drawn it, after the output from the lefthand latch goes up to the "button blockers," it can also run down the wire to the other half of the circuit (against the direction to the arrows), because everything is joined together. In other words, if the lefthand LED lights up, the righthand LED will light up, too. How can I stop this from happening?

Well, I could put diodes in the "up" wires to block current from running down them. But I have a more elegant idea: I'll add an OR gate, because the inputs to an OR gate are separated from each other electrically. Figure 4-91 shows this.

Usually an OR gate has only two logical inputs. Will this prevent me from adding more players? No, because you can actually buy an OR that has eight inputs. If any one of them is high, the output is high. For fewer than eight players, I can short the unused inputs to ground, and ignore them.

Looking again at Figure 4-91, I'm getting a clearer idea of what the thing I've called a "button blocker" should actually be. I think it should be another logic gate. It should say, "If there's only one input, from a button, I'll let it through. But if there is a second input from the OR gate, I won't let it through."

That sounds like a NAND gate, but before I start choosing chips, I have to decide what the latch will be. I can buy an off-the-shelf flip-flop, which flips "on" if it gets one signal and "off" if it gets another, but the trouble is, chips containing flip-flops tend to have more features than I need for a simple circuit like this. Therefore I'm going to use 555 timers again, in flip-flop mode. They require very few connections, work very simply, and can deliver a good amount of current. The only problem with them is that they require a negative input at the trigger pin to create a positive output. But I think I can work with that.

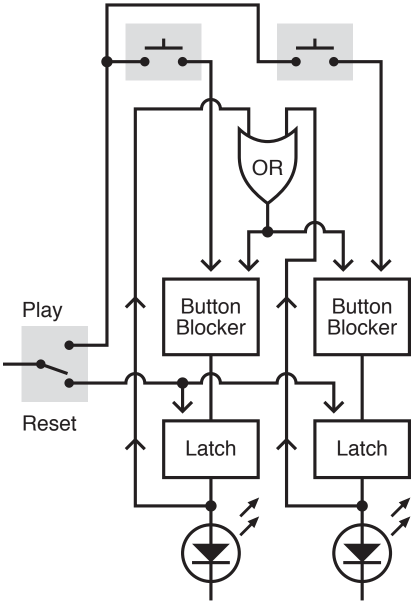

Figure 4-91. *To prevent the output from one latch feeding back around the circuit to the output from another latch, the outputs can be combined in an OR gate.*

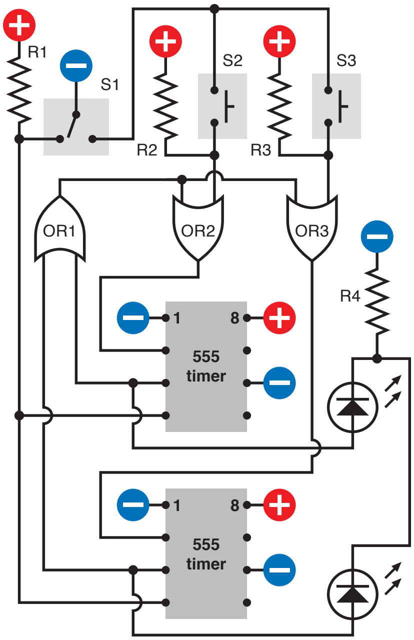

Figure 4-92. *Now that the basic concept of the quiz circuit has been roughed out, specific components can be inserted, with compatible inputs and outputs.*

So now, finally, here's a simplified schematic, in Figure 4-92. I like to show the pins of the 555 timers in their correct positions, so I had to move the components around a little to minimize wire crossovers, but you can see that logically, it's the same basic idea.

Before you try to build it, just run through the theory of it, because that's the final step, to make sure there are no mistakes. The important thing to bear in mind is that because the 555 needs a negative input on its trigger pin to create its output, when any of the players presses a button, the button has to create a negative "flow" through the circuit. This is a bit counterintuitive, so I'm including a three-step visualization in Figure 4-93, showing how it will work.

In Step 1, the quizmaster has asked a question and flipped his switch to the right, to supply (negative) power to the players' buttons. So long as no one presses a button, the pull-up resistors supply positive voltage to OR2 and OR3. An OR gate has a positive output if it has any positive input, so OR2 and OR3 keep the trigger inputs of the 555 timers positive. Their outputs remain low, and nothing is happening yet.

In Step 2, the lefthand player has pressed his button. Now OR2 has two negative inputs, so its output has gone low. But IC1 hasn't reacted yet.

In Step 3, just a microsecond later, IC1 has sensed the low voltage on its trigger, so its output from pin 3 has gone high, lighting the LED. Remember, this 555 timer is in flip-flop mode, so it locks itself into this state immediately. Meanwhile its high output also feeds back to OR1. Because OR1 is an OR gate, just one high input is enough to make a high output, so it feeds this back to OR2 and OR3. And now that they have high inputs, their outputs also go high, and will stay high, regardless of any future button-presses.

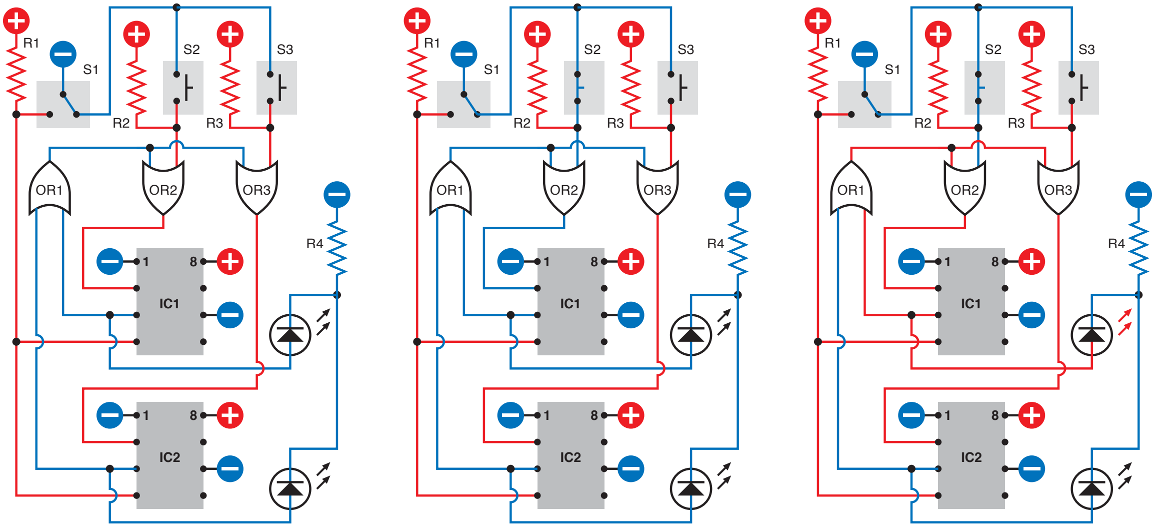

Figure 4-93. *These three schematics show the prevalence of higher and lower voltages (red and blue lines) through the quiz circuit when a pushbutton is pressed.*

Because OR2 and OR3 now have high inputs and outputs, IC1 and IC2 cannot be triggered. But IC1 is still locked into its "on" state, keeping the LED illuminated.

The only way to change IC1 is if the quizmaster flips his switch back to the left. That applies negative power to the reset pins of both the timers. Consequently their outputs go low, the LED goes out, and the circuit goes back into the same state as where it started. Having reset it, the quizmaster can ask another question, but the players' buttons are not activated until the quizmaster flips the switch back to the right again.

There's only one situation that I haven't addressed: what if both players press their buttons absolutely simultaneously? In the world of digital electronics, this is highly unlikely. Even a difference of a microsecond should be enough time for the circuit to react and block the second button. But if somehow both buttons are pressed at the same instant, both of the timers should react, and both of the LEDs will light up, showing that there has been a tie.

In case you feel a little uncertain about the way in which a two-player circuit can be upgraded to handle extra players, I've included a simplified three-player schematic in Figure 4-94.

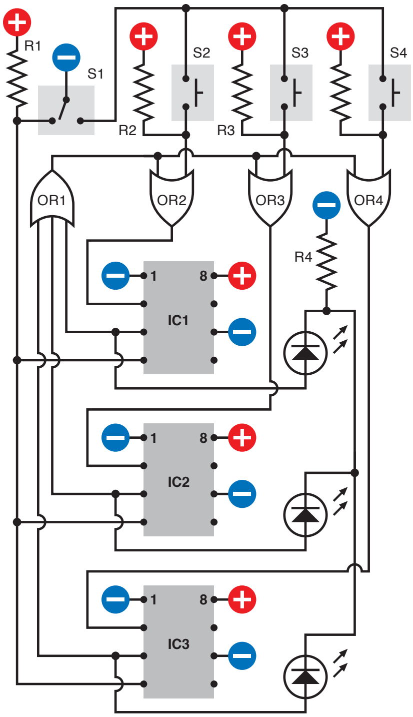

Figure 4-94. *The two-player schematic in can be easily upgraded to a three-player version, as shown here, provided the first OR gate can handle three inputs.*

Breadboarding It

Now it's time to create a schematic that's as close to the breadboard layout as possible, so that you can build this thing easily. The schematic is shown in Figure 4-95 and the actual components on a breadboard are in Figure 4-96. Because the only logic gates that I've used are OR gates, and there are only three of them, I just need one logic chip: the 74HC32, which contains four 2-input OR gates. (I've grounded the inputs to the fourth). The two OR gates on the left side of the chip have the same functions as OR2 and OR3 in my simplified schematic, and the OR gate at the bottom-right side of the chip works as OR1, receiving input from pin 3 of each 555 timer. If you have all the components, you should be able to put this together and test it quite quickly.

You may notice that I've made one modification of the previous schematic. A 0.01 µF capacitor has been added between pin 2 of each 555 timer (the Input) and negative ground. Why? Because when I tested the circuit without the capacitors, sometimes I found that one or both of the 555 timers would be triggered simply by flipping S1, the quizmaster switch, without anyone pressing a button.

At first this puzzled me. How were the timers getting triggered, without anyone doing anything? Maybe they were responding to "bounce" in the quizmaster switch. Sure enough, the small capacitors solved the problem. They may also slow the response of the 555 timers fractionally, but not enough to interfere with slow human reflexes.

As for the buttons, it doesn't matter if they "bounce," because each timer locks itself on at the very first impulse and ignores any hesitations that follow.

You can experiment building the circuit, disconnecting the 0.01 µF capacitors, and flipping S1 to and fro a dozen times. If you have a high-quality switch, you may not experience any problem. If you have a lower-quality switch, you may see a number of "false positives." I'm going to explain more about "bounce," and how to get rid of it, in the next experiment.

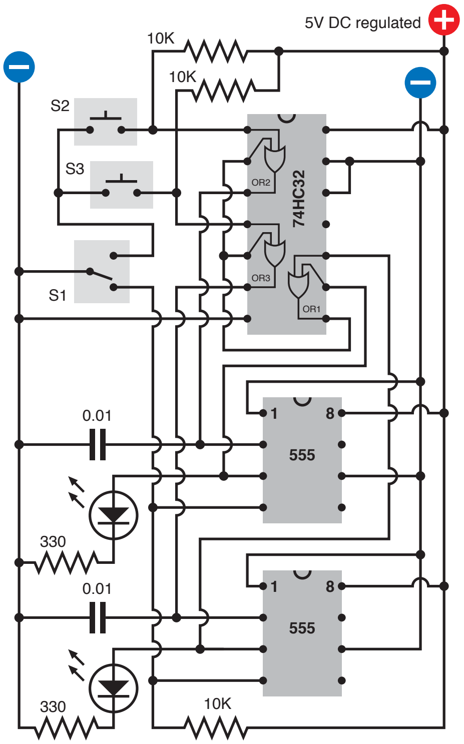

Figure 4-95. *Applying the simplified schematic to a breadboard inevitably entails a wiring layout that is less intuitively obvious and appears more complex. The connections are the same, though.*

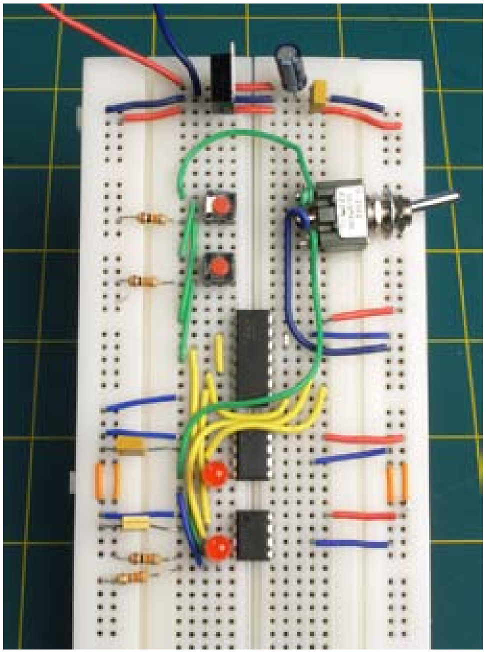

Figure 4-96. *The quiz schematic applied to a breadboard, to test the concept prior to full-scale implementation.*

Enhancements

After you breadboard the circuit, if you proceed to build a permanent version, I suggest that you expand it so that at least four players can participate. This will require an OR gate capable of receiving four inputs. The 74HC4078 is the obvious choice, as it allows up to eight. Just connect any unused inputs to negative ground.

Alternatively, if you already have a couple of 74HC32 chips and you don't want to bother ordering a 74HC4078, you can gang together three of the gates inside a single 74HC32 so that they function like a four-input OR. Look at the simple logic diagram in Figure 4-97 showing three ORs, and remember that the output from each OR will go high if at least one input is high.

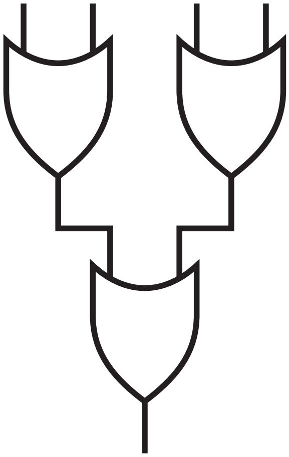

Figure 4-97. *Although a four-input OR gate is not manufactured, its functionality can be achieved easily by linking three 2-input OR gates together.*

And while you're thinking about this, see if you can figure out the inputs and output of three ANDs in the same configuration.

For a four-player game, you'll also need two additional 555 timers, of course, and two more LEDs, and two more pushbuttons.

As for creating a schematic for the four-player game—I'm going to leave that to you. Begin by sketching a simplified version, just showing the logic symbols. Then convert that to a breadboard layout. And here's a suggestion: pencil, paper, and an eraser can still be quicker, initially, than circuit-design software or graphic-design software, in my opinion.

Experiment 22: Flipping and Bouncing

I mentioned in the previous experiment that "bounce" from the buttons in the circuit wouldn't be a problem, because the buttons were activating 555 timers that were wired in bistable, flip-flop mode. As soon as the timer receives the very first pulse, it flips into its new state and flops there, ignoring any additional noise in the circuit. So can we debounce a switch or a button using a flip-flop? And as some 74HCxx chips are available containing flip-flops, can we use them?

The answers are yes, and yes, although it's not quite as simple as it sounds.

You will need:

- 74HC02 logic chip containing 4 NOR gates. 74HC00 logic chip containing 4 NAND gates. Quantity: 1 of each.
- SPDT switch. Quantity: 1.
- LEDs, low-current. Quantity: 2.
- 10K resistors and 1K resistors. Quantity: 2 of each.

Assemble the components on your breadboard, following the schematic shown in Figure 4-98. When you apply power (through your regulated 5-volt supply), one of the LEDs should be lit.

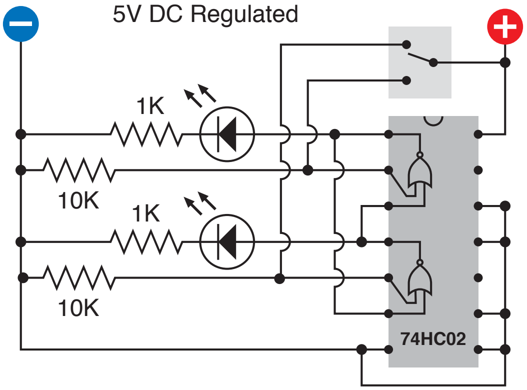

Figure 4-98. *A simple circuit to test the behavior of two NOR gates wired as a simple flip-flop that retains its state after an input pulse ceases.*

Now I want you to do something odd. Please disconnect the SPDT switch by taking hold of the wire that connects the positive power supply to the pole of the switch, and pulling the end of the wire out of the breadboard. When you do this, you may be surprised to find that the LED remains lit.

Push the wire back into the breadboard, flip the switch, and the first LED should go out, while the other LED should become lit. Once again, pull out the wire, and once again, the LED should remain lit.

Here's the take-home message:

- A flip-flop requires only an initial pulse.
- After that, it ignores its input.

How It Works

Two NOR gates or two NAND gates can function as a flip-flop:

- Use NOR gates when you have a positive input from a double-throw switch.
- Use NAND gates when you have a negative input from a double-throw switch.

Either way, you have to use a double-throw switch.

I've mentioned the double-throw switch three times (actually, four times if you count this sentence!) because for some strange reason, most introductory books fail to emphasize this point. When I first started learning electronics, I went crazy trying to understand how two NORs or two NANDs could debounce a simple SPST pushbutton—until finally I realized that they can't. The reason is that when you power up the circuit, the NOR gates (or NAND gates) need to be told in which state they should begin. They need an initial orientation, which comes from the switch being in one state or the other. So it has to be a double-throw switch. (Now I've mentioned it five times.)

I'm using another simplified multiple-step schematic, Figure 4-99, to show the changes that occur as the switch flips to and fro with two NOR gates. To refresh your memory, I've also included a truth table showing the logical outputs from NOR gates for each combination of inputs.

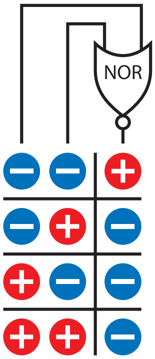

−	−	+
−	+	−
+	−	−
+	+	−

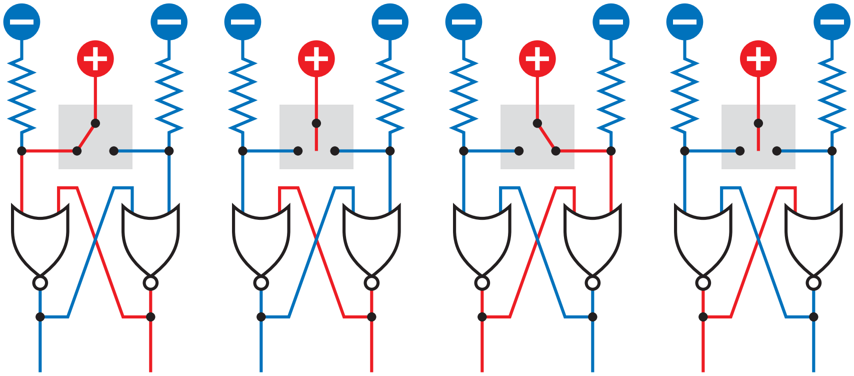

Figure 4-99. *Using two NOR gates in conjunction with a positive input through a SPDT switch, this sequence of four diagrams shows how a flip-flop circuit responds.*

Suppose that the switch is turned to the left. It sends positive current to the lefthand side of the circuit, overwhelming the negative supply from the pull-down resistor, so we can be sure that the NOR gate on the left has one positive logical input. Because any positive logical input will make the NOR give a negative output (as shown in the truth table), the negative output crosses over to the righthand NOR, so that it now has two negative inputs, which make it give a positive output. This crosses back to the lefthand NOR gate. So, in this configuration everything is stable.

Now comes the clever part. Suppose that you move the switch so that it doesn't touch either of its contacts. (Or suppose that the switch contacts are bouncing, and failing to make a good contact. Or suppose you disconnect the switch entirely.) Without a positive supply from the switch, the lefthand input of the left NOR gate goes from positive to negative, as a result of the pull-down resistor. But the righthand input of this gate is still positive, and one positive is all it takes to make the NOR maintain its negative output, so nothing changes. In other words, the circuit has "flopped" in this state.

Now if the switch turns fully to the right and supplies positive power to the righthand pin of the right NOR gate, quick as a flash, that NOR recognizes that it now has a positive logical input, so it changes its logical output to negative. That goes across to the other NOR gate, which now has two negative inputs, so its output goes positive, and runs back to the right NOR.

In this way, the output states of the two NOR gates change places. They flip, and then flop there, even if the switch breaks contact or is disconnected again. The second set of drawings in Figure 4-100 shows exactly the same logic, using a negatively powered switch and two NAND gates. You can use your 74HC00 chip, specified in the parts list for this experiment, to test this yourself.

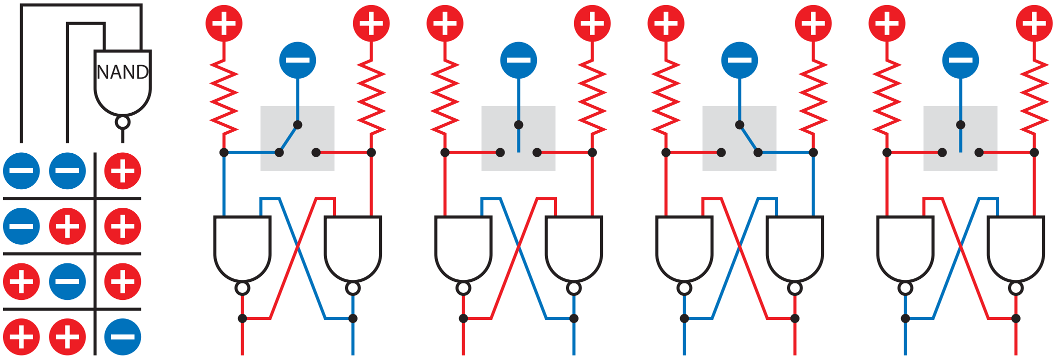

Figure 4-100. *The schematic from Figure 4-99 can be rewired with NAND gates and a negative switched input.*

Both of these configurations are examples of a jam-type flip-flop, so called because the switch forces it to respond immediately, and jams it into that state. You can use this circuit anytime you need to debounce a switch (as long as it's a double-throw switch).

A more sophisticated version is a clocked flip-flop, which requires you to set the state of each input first and then supply a clock pulse to make the flip-flop respond. The pulse has to be clean and precise, which means that if you supply it from a switch, the switch must be debounced—probably by using another jam-type flip-flop! Considerations of this type have made me reluctant to use clocked flip-flops in this book. They add a layer of complexity, which I prefer to avoid in an introductory text.

What if you want to debounce a single-throw button or switch? Well, you have a problem! One solution is to buy a special-purpose chip such as the 4490 "bounce eliminator," which contains digital delay circuitry. A specific part number is the MC14490PG from On Semiconductor. This contains six circuits for six separate inputs, each with an internal pull-up resistor. It's relatively expensive, however—more than 10 times the price of a 74HC02 containing NOR gates. Really, it may be simpler to use double-throw switches that are easily debounced as described previously.

Experiment 23: Nice Dice

This is the one experiment where I want you to use the 74LSxx generation of the TTL logic family, instead of the 74HCxx family of CMOS. Two reasons: first, I need to use the 7492 counter, which is unavailable in the HC family. And second, you should know the basic facts about the LS series of TTL chips, as they still crop up in circuits that you'll find in electronics books and online.

In addition, you'll learn about "open collector" TTL chips such as the 74LS06 inverter, which can be a convenient substitute for transistors when you want to deliver as much as 40mA of current.

The idea of this circuit is simple enough: run a 555 timer sending very fast pulses to a counter that counts in sixes, driving LEDs that are placed to imitate the spots on a die. (Note that the word "die" is the singular of "dice.") The counter runs so fast, the die-spots become a blur. When the user presses a button, the counter stops arbitrarily, displaying an unpredictable spot pattern.

Dice simulations have been around for many, many years, and you can even buy kits online. But this one will do something more: it will also demonstrate the principles of binary code.

So, if you're ready for the triple threat of TTL chips, open collectors, and binary, let's begin.

You will need:

- 74LS92 counter such as SN74LS92N by Texas Instruments. Quantity: 1 if you want to create one die, 2 to make two dice.
- 74LS27 three-input NOR gate such as SN74LS27N by Texas Instruments. Quantity: 1.
- 555 timers. Quantity: 1 if you want to make one die, 2 to make two dice.
- Signal diodes, 1N4148 or similar. Quantity: 4, or 8 to make two dice.

Seeing Binary

The counter that we dealt with before was unusual, in that its outputs were designed to drive seven-segment numerals. A more common type has outputs that count in binary code.

The 74LS92 pinouts are shown in Figure 4-101. Plug the chip into your breadboard and make connections as shown in Figure 4-102. Initially, the 555 timer will drive the counter in slow-motion, at around 1 step per second. Figure 4-103 shows the actual components on a breadboard.

Note that the counter has unusual power inputs, on pins 5 and 10 instead of at the corners. Also four of its pins are completely unused, and do not connect with anything inside the chip. Therefore, you don't need to attach any wire to them on the outside.

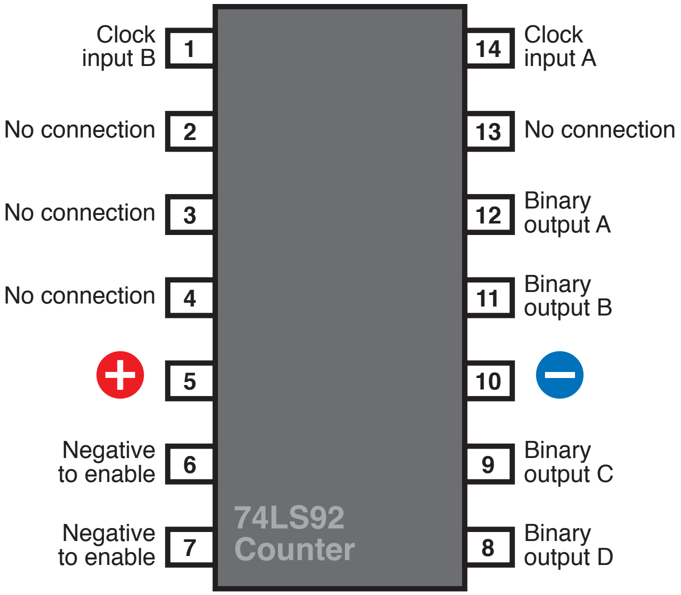
Clock input B
1
No connection
2
No connection
3
No connection
4
5
Negative to enable
6
Negative to enable
7
74LS92 Counter
14
Clock input A
13
No connection
12
Binary output A
11
Binary output B
10
9
Binary output C
8
Binary output D

Figure 4-101. *The unusual pin assignments include four that have no connection of any kind inside the chip, and can be left unattached.*

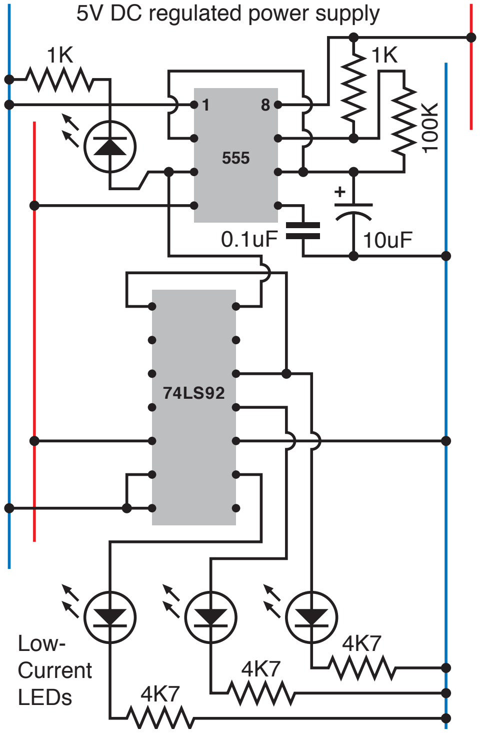
5V DC regulated power supply
1K
1K
1
8
555
100K
+
0.1uF
10uF
74LS92
Low-Current LEDs
4K7
4K7
4K7

Figure 4-102. *This simple circuit uses a 555 timer running slowly to control the 74LS92 binary counter and display the succession of high states from its outputs.*

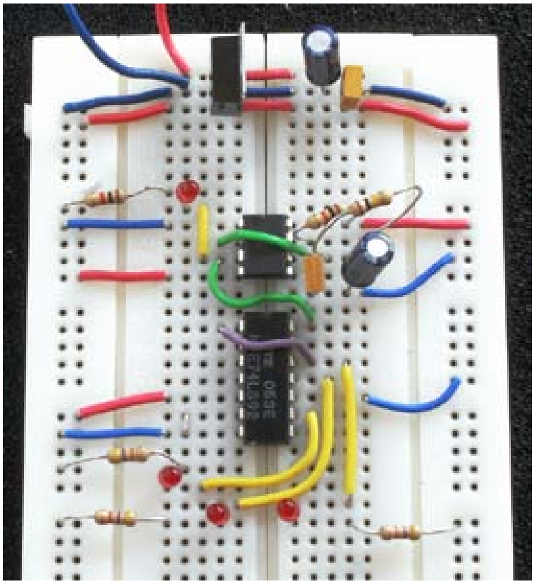

Figure 4-103. *The breadboard version of the schematic in Figure 4-102 to display the outputs from a 74LS92 counter.*

Now we come to the first new and difficult fact about the 74LSxx generation of TTL chips that makes them less desirable, for our purposes, than the 74HCxx generation of CMOS chips that I have recommended in previous projects. The modern and civilized HC chips will source 4mA or sink 4mA at each logical output, but the older LS generation is fussier. It will sink around 8mA into each output pin from a positive source, but when its output is high, it hardly gives you anything at all. This is a very basic principle:

- Outputs from TTL logic chips are designed to sink current.
- They are not designed to source significant current.

In fact, the 74LS92 is rated to deliver less than half a milliamp. This is quite acceptable when you're just connecting it with another logic chip, but if you want to drive an external device, it doesn't provide much to work with.

The proper solution is to say to the chip, "All right, we'll do it your way," and set things up with a positive source that flows through a load resistor to the LED that you want to use, and from there into the output from the chip. This is the "better" option shown in Figure 4-104.

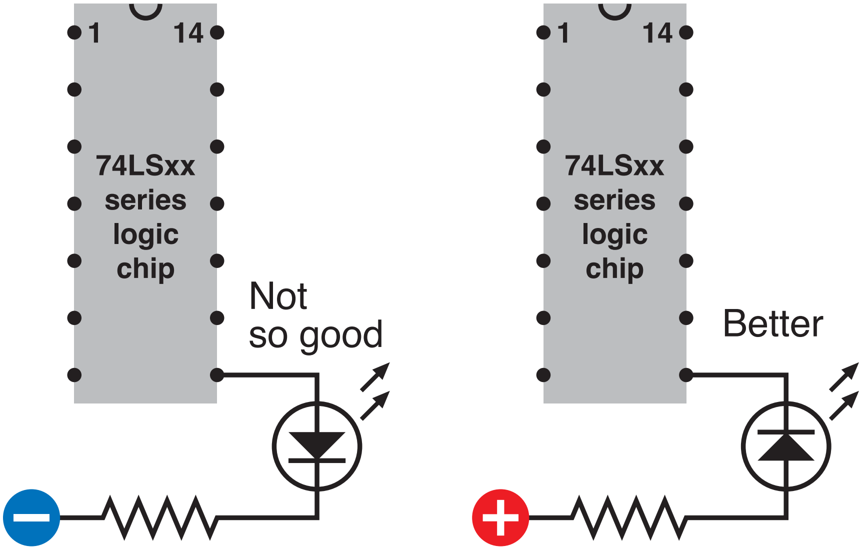

Figure 4-104. *Most TTL chips, including those in the LS generation, are unable to source much current from their logical output pins (left) and should usually be wired to sink current from a positive source (right).*

The only problem is that now the LED lights up when the counter's output is low. But the counter is designed to display its output in high pulses. So your LED is now off when it should be on, and on when it should be off.

You can fix this by passing the signal through an inverter, but already I'm getting impatient with this inconvenience. My way around the problem, at least for demo purposes, is to use the "Not so good" option in Figure 4-104 and make it work by connecting a very-low-current LED with a large 4K7 load resistor. This will enable us to "see" the output from the counter without asking it to give more than its rated limit, and if you want to create a more visibly powerful display for a finished version of the dice circuit, I'll deal with that later. According to my meter, the 4K7 resistor holds the current between 0.3mA and 0.4mA, which is the counter's rated maximum.

Set up your initial version of the circuit as shown in Figures 4-102 and 4-103. Be careful when you wire the positive and negative power supply to the counter chip, with its nonstandard pin assignments.

The 555 will run in astable mode, at about 1 pulse per second. This becomes the clock signal for the counter. The first three binary outputs from the counter then drive the three LEDs.

The counter advances when the input signal goes from *high* to *low*. So when the LED beside the 555 timer goes out, that's when the counter advances.

If you stare at the pattern generated by the outputs for long enough, you may be able to see the logic to it, bearing in mind that its zero state is when they are all off, and it counts up through five more steps before it repeats. The diagram in Figure 4-105 shows this sequence. If you want to know why the pattern works this way, check the following section, "Theory: Binary arithmetic."

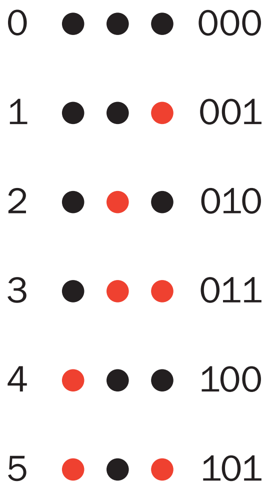

Figure 4-105. *The three output pins of the 74LS92 counter have high states shown by the red circles as the counter steps from 000 to 101 in binary notation.*

THEORY

Binary arithmetic

The rule for binary counting is just a variation of the rule that we normally use for everyday counting, probably without thinking much about it. In a 10-based system, we count from 0 to 9, then carry 1 over to the next position on the left, and go from 0 to 9 again in the right-most position. We repeat this procedure until we get to 99, then carry a 1 over to a new position to make 100, and continue counting.

In binary we do the same thing, except that we restrict ourselves to digits 0 and 1 only. So begin with 0 in the rightmost position, and count up to 1. As 1 is our limit, to continue counting we carry 1 over to the next place on the left, and start again from 0 in the right-most position. Count up to 1, then add 1 to the next place on the left—but, it already has a 1 in it, so it can't count any higher. So, carry 1 from there one space further, to the next place beside that—and so on.

If a glowing LED represents a 1, and a dark LED represents a 0, the diagram in Figure 4-105 shows how the 74LS92 counts up from 0 to (decimal) 6 or (binary) 101 in its inimitable fashion. I've also included a diagram in Figure 4-106 showing how a counter with four binary outputs would display decimal numbers from 0 through 15, again using the LEDs to represent 1s and 0s.

Here's a question for you: how many LEDs would you need to represent the decimal number 1024 in binary? And how many for 1023?

Obviously binary code is ideally suited to a machine full of logic components that either have a high or a low state. So it is that all digital computers use binary arithmetic (which they convert to decimal, just to please us).

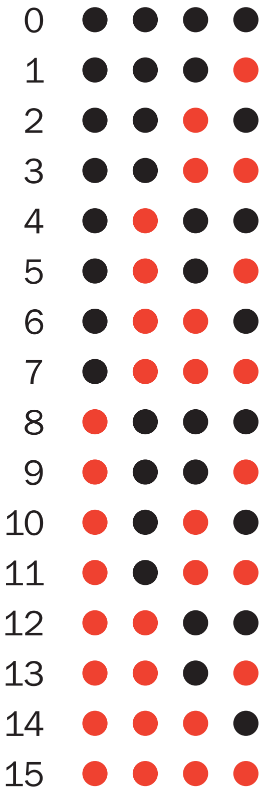

Figure 4-106. *A hexadecimal (16-based) binary counter would generate this succession of high states from its four output pins as it counts from 0 through 15 in decimal notation.*

Getting back to our project: I want to take the three binary outputs and make them create patterns like the spots on a die. How can I do this? Quite easily, as it turns out.

I'm assuming that I'll use seven LEDs to simulate the patterns of spots on a die. These patterns can be broken down into groups, which I have assigned to the three outputs from the counter in Figure 4-108. The first output (farthest to the right) can drive an LED representing the dot at the center of the die face. The second (middle) output can drive two more diagonal LEDs. The third output must switch on all four corner LEDs.

This will work for patterns 1 through 5, but won't display the die pattern for a 6. Suppose I tap into all three outputs from the counter with a three-input NOR gate. It has an output that goes high only when all three of its inputs are low, so it will only give a high output when the counter is beginning with all-low outputs. I can take advantage of this to make a 6 pattern.

Note that it's bad practice to mix the LS generation of TTL chips with the HC generation of CMOS chips, as their input and output ranges are different; so, the NOR chip has to be a 74LS27, not a 74HC27.

We're ready now for a simple schematic. In Figure 4-107 I've colored some of the wires just to make it easier for you to distinguish them. The colors have no other significance.

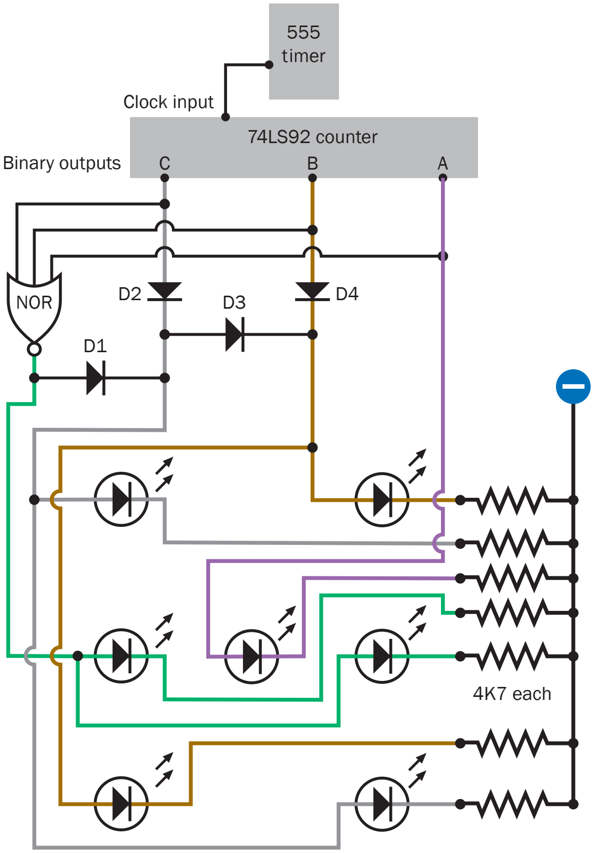
555
timer
Clock input
74LS92 counter
Binary outputs
C
B
A
NOR
D2
D3
D4
D1
4K7 each

Figure 4-107. *A simplified schematic shows how outputs from the 74LS92 counter can be combined, with signal diodes and a single three-input NOR gate, to generate the spot patterns on a die. The wire colors have no special meaning and are used just to distinguish them from each other.*

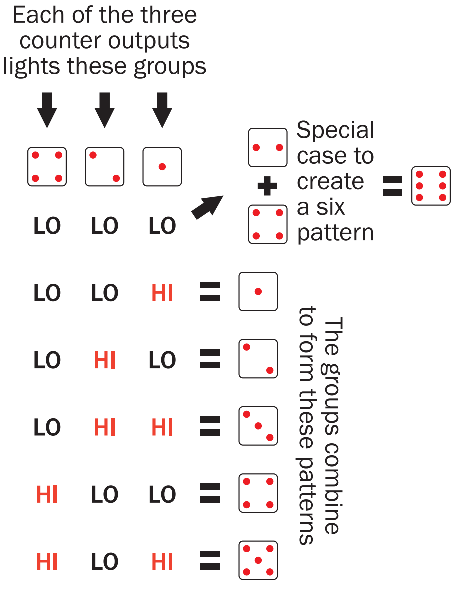
Each of the three
counter outputs
lights these groups
Special
case to
create
a six
pattern
LO LO LO
LO LO HI
LO HI LO
LO HI HI
HI LO LO
HI LO HI
The groups combine
to form these patterns

Figure 4-108. *Binary outputs from the 74LS92 counter can be used to power LEDs arrayed in groups to simulate the pattern of spots on a die.*

Each of the LEDs is grounded through a separate 4K7 load resistor. Unfortunately, this means that when they are displaying the pattern for a 6, all of them are running in parallel from the output of the NOR gate, which overloads it. As long as you don't leave the display in this mode for very long periods, it shouldn't cause a problem. You could compensate by increasing the load resistors, or by running pairs of the LEDs through one resistor, but this will make them so dim that they'll be difficult to see, as they're so close to their lower limit for current already.

Notice how I have added four signal diodes, D1 through D4. When Output C goes high, it has to illuminate all four corner LEDs, and so its power goes into the brown wire as well as the gray wire. But we must never allow one output to feed back into another, so D4 is needed to protect Output B when Output C is high.

Because there is now a connection between B and C, we need D2 to protect Output C when Output B is high. And because Output B must only feed two of the corner LEDs, we also need D3 to stop it from illuminating the other two. And, we have to protect the output from the NOR gate when either Output C or Output B is high. This requires D1.

Figure 4-109 shows everything that I've described so far assembled in breadboard format, while Figure 4-110 shows the test version that I built. Note that the unused logical inputs on the 74LS27 chip are shorted together and connected to the *positive* side of the power supply. Here's the rule:

- When using CMOS chips (such as the HC series), connect unused logical inputs to the negative side of the power supply.
- When using TTL chips (such as the LS series), connect unused logical inputs to the positive side of the power supply.

I assume that you have had enough fun watching the LEDs count slowly, so I've changed the capacitor and resistor values for the 555 to increase its speed from approximately 1 pulse per second to about 50,000 pulses per second. The counter could run much faster than this, but I just want it to cycle fast enough so that when the user presses and releases a button, the count will stop at an unforeseeable number.

The button starts and stops the 555 timer by applying and releasing power to the timing circuit only. This is the equivalent of shaking and then throwing the die.

While the counter is running fast, the LEDs are flashing so fast that all of them will seem to be on at once. At the same time, the circuit charges a new 68 µF capacitor, which I have added between the pushbutton and ground. When you release the button, this capacitor discharges itself through the 1K timing resistor. As the charge dissipates, the timing capacitor will take longer and longer to charge, and discharge, and the frequency of the 555 will gradually diminish. Consequently the LED display will also flash slower, like the reel on a Las Vegas slot machine gradually coming to a stop. This increases the tension as players can see the die display counting to the number that they're hoping for—and maybe going one step beyond it.

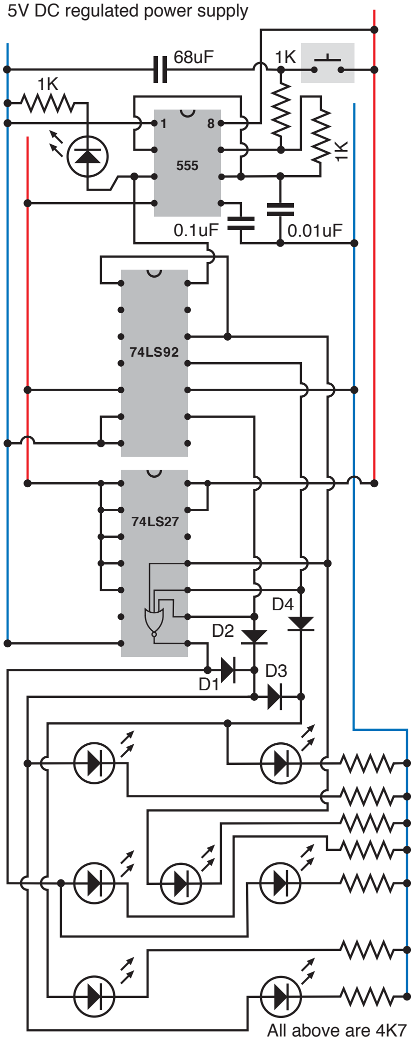

Figure 4-109. *With some extra components, the schematics from Figures 4-102 and 4-107 can be combined to make the working dice simulation.*

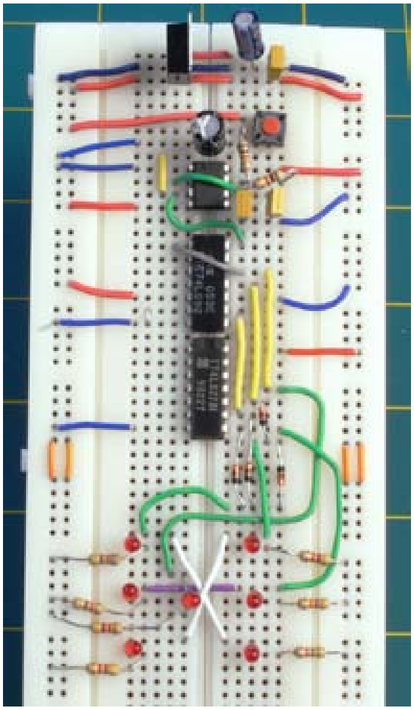

Figure 4-110. *The electronic dice schematic applied to a breadboard, with a pushbutton at the top to start and stop the counter, and 7 LEDs at the bottom to display the output.*

Note that to maximize this effect, the button has to be held down for a full second or more, so that the 68 μF capacitor becomes fully charged before the button is released.

So, this circuit now fulfills the original goal. But can it be better? Of course it can.

Enhancements

The main thing I want to improve is the brightness of the LEDs. I could add a transistor to amplify the current to each one, but there's a simpler alternative: a TTL "open collector" inverter.

I want to use an inverter because in the world of TTL, as I mentioned earlier, we can sink much more power into the output pin of a chip than we can source from it. So, I'm going to turn each LED the other way around and connect their load resistors to the positive side of the power supply. This way, they'll sink their power into the outputs of the inverter.

And the great advantage of an "open collector" version of the inverter chip is that it is designed to sink much more current than a normal TTL logic chip. It is rated for 40mA per pin. The only disadvantage is that it cannot source any current at all; instead of its output going high, it just behaves like an open switch. But that's OK for this circuit.

So the next and final schematic, in Figure 4-111, includes the 74LS06 inverter, which has also been added to the breadboarded version shown in Figure 4-112. I suggest that you put aside the little low-current LEDs and substitute some normal-size ones. Using Kingbright "standard" WP15031D 5mm LEDs, I find that each draws almost exactly 20mA with a voltage drop of about 2V with a 120 ohm series resistor. Because each output pin from the 74LS06 inverter powers no more than two LEDs at a time, this is exactly within its specification. I suggest that if you build this circuit, you check the consumption of your particular choice of LEDs and adjust the resistors if necessary.

Remember: to measure the voltage drop across an LED, simply touch the probes of your meter across it while it is illuminated. To measure the current, disconnect one side of the LED and insert the meter, in milliamp mode, between the leg of the LED and the contact that it normally makes in the circuit.

For a really dramatic display, you can get some 1 cm diameter LEDs (Figure 4-113). Check the specification, and you should find that many of these don't use more power than the usual 5 mm type. But whatever kind you use, don't forget to turn them around so that their negative sides face toward the inverter, and their positive sides face the resistors, which are connected to the positive side of the power supply.

One last detail: I had to add two 10K resistors to this version of the circuit. Can you see why? Diodes D1 through D4 are designed to transmit positive voltage through to the inverter when appropriate, but they prevent the inputs of the inverter from "seeing" the negative side of the power supply when the counter outputs are low. These inverter inputs require pull-down resistors to prevent them from "floating" and producing erroneous results.

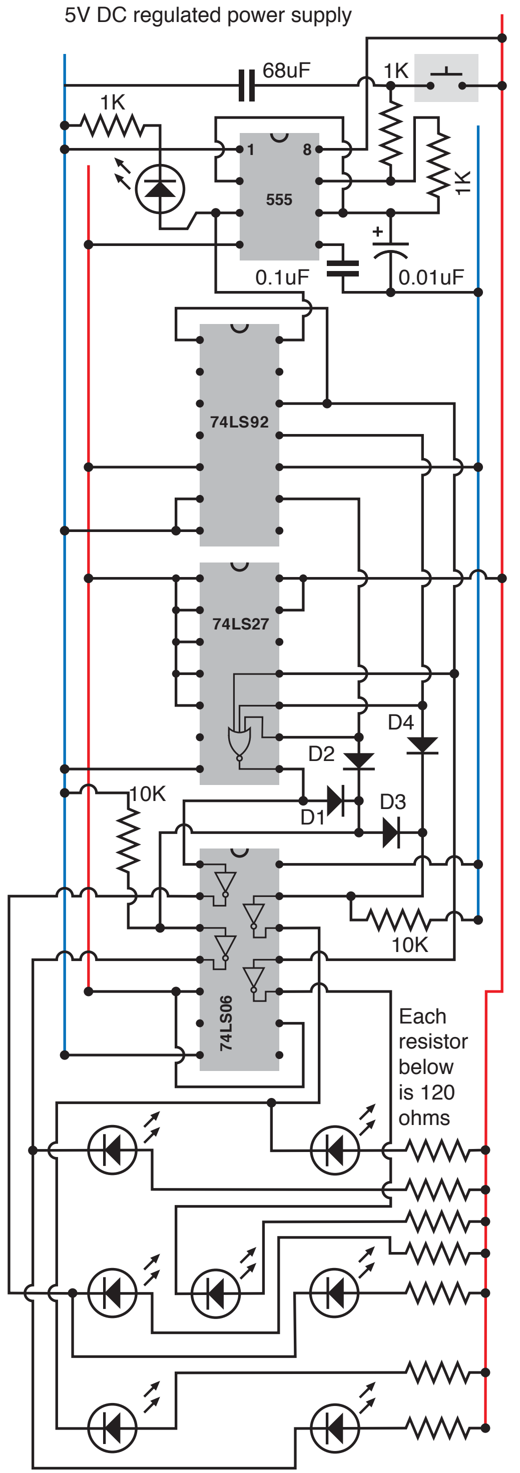
5V DC regulated power supply
68uF
1K
1K
1
8
555
1K
0.1uF
0.01uF
74LS92
74LS27
D4
D2
10K
D1
D3
74LS06
10K
Each resistor below is 120 ohms

Figure 4-111. *If open-collector inverters are added to the dice schematic, it can drive full-size LEDs with up to 40mA, as long as the LEDs are turned around to sink current into the TTL output stage instead of trying to source current from it.*

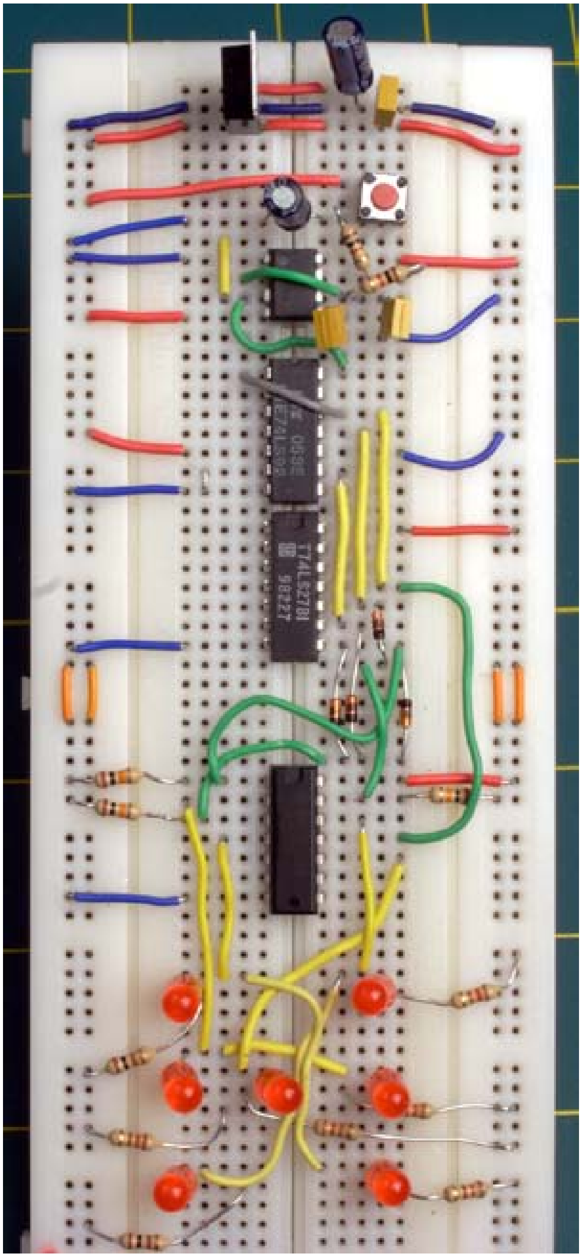

Figure 4-112. *The completed circuit using an open-collector inverter to drive full-size LEDs.*

The final enhancements are up to you. Most obviously, you can add a second die, as many games require two dice. The 74LS27 chip still has a couple of spare NOR gates in it, one of which you can make use of, but you will need an additional 555 timer, running at a significantly different speed to ensure randomness, and it will have to drive a second counter.

After you get your dice up and running, you may want to test them for randomness. Because the pulses from a 555 timer are of equal length, every number has an equal chance of coming up; but the longer you hold down the Start button, the better your odds are of interrupting the counting process at a truly random moment. Anyone using your electronic dice should be told that "shaking" them for a full second is mandatory.

Of course, I could have simulated dice more easily by writing a few lines of software to generate random numbers on a screen, but even a fancy screen image cannot have the same appeal as a well-made piece of hardware. Figure 4-113 shows white 1 cm LEDs mounted in a sanded polycarbonate enclosure for dramatic effect.

Most of all, I derived satisfaction from using simple, dedicated chips that demonstrate the binary arithmetic that is fundamental in every computer.

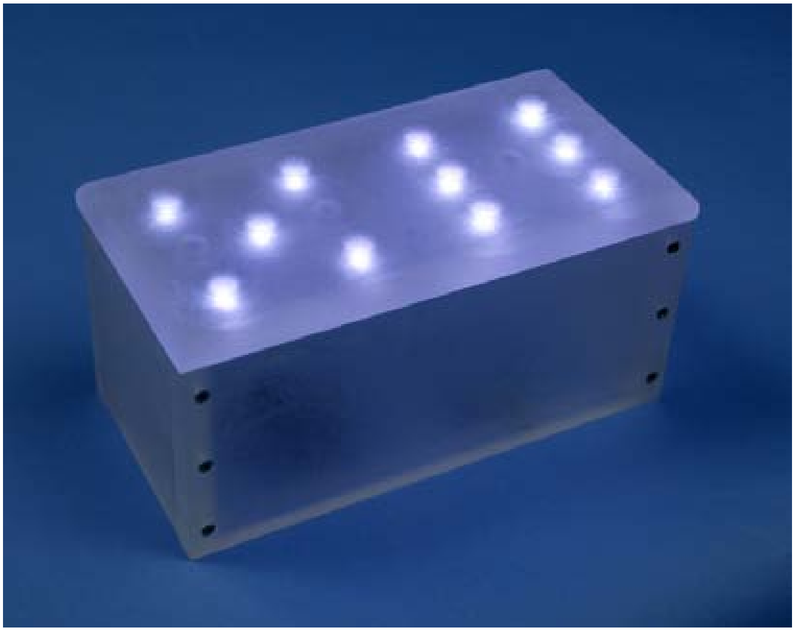

Figure 4-113. *The open-collector inverter chip in the dice circuit is sufficiently powerful to drive 1-cm white LEDs that draw about 20mA each, using a potential of 2V. In this finished version, the LEDs were embedded in cavities drilled from the underside of half-inch polycarbonate, which has been treated with an orbital sander to create a translucent finish.*

Experiment 24: Intrusion Alarm Completed

Now let me suggest how you can apply the knowledge from this chapter of the book to upgrade the burglar alarm project that was last modified in Experiment 15. You'll probably need to check Chapters 2 and 3 to refamiliarize yourself with some features of the alarm.

Upgrade 1: Delayed Activation

The biggest flaw in the alarm was that as soon as it was activated, it would immediately respond to any signal from the door and window sensors. It needed a feature to delay activation to give you a chance to exit from the building before the alarm armed itself. A 555 timer can provide this functionality, probably in conjunction with a relay. The power to the alarm should pass through the contacts of the relay, which are normally closed. When you press a button on the timer, it sends a positive pulse to the relay lasting for around 30 seconds, holding the relay open for that period. You could mount the timer in its own little box with a button on it, which you press when you're ready to leave the building. The 12-volt power supply to the burglar alarm passes through the box containing the delay circuit. For 30 seconds, the 555 interrupts power to the alarm, and then restores it, ready for action.

Upgrade 2: Keypad Deactivation

This is now really simple. You can substitute a latching relay instead of the switch, S1, on the alarm box (shown in Figure 3-110), and use a keypad to set and reset the relay in exactly the same way as in the combination lock in Experiment 20. You'll have to run an additional three wires from the relay, out of the alarm box, to the keypad (one supplying power to the "on" relay coil, another supplying power to the "off" coil, the third being a common ground). You can either use a 9-volt battery to power the electronics in association with the keypad, or run an additional fourth wire from the alarm box, to carry positive power to the logic chips, bearing in mind that you have to insert a voltage regulator at some point, to drop the 12 volts that the alarm uses to the 5 volts that the logic gates require. As the gates consume so little power, the 12-to-5 drop should be OK for the voltage regulator; it won't have to dissipate too much heat.

With this additional feature, you can use the alarm like this:

- Press the pound key on the keypad to flip the latching relay into its "on" mode, so that it passes power to the alarm box, which is now armed.
- If you want to leave the house, push the button on the delay unit to give you 30 seconds in which to do so.
- If the alarm is triggered, enter your secret code on the keypad to deactivate it by flipping the latching relay to its "off" position and cutting power to the alarm box.

These modifications are so simple that I think the block diagram in Figure 4-114 should be all you need. I don't think I need to give you any schematics. The only change you have to make to the existing alarm is to substitute the latching relay for the on/off switch.

But, there is still one obvious necessary enhancement needed: how can you get back into the house without instantly triggering the alarm?

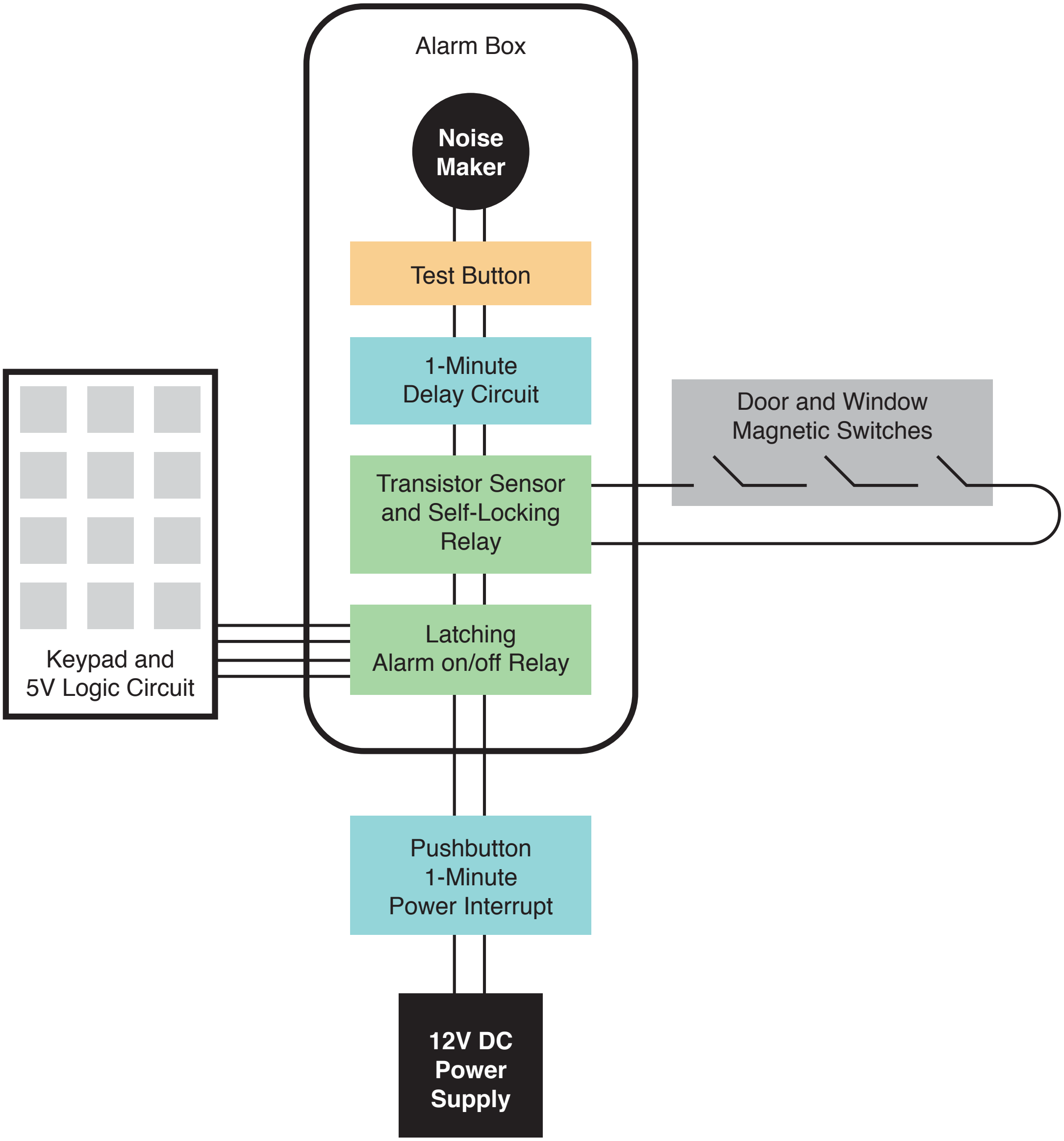
Alarm Box
Noise Maker
Test Button
1-Minute Delay Circuit
Door and Window Magnetic Switches
Transistor Sensor and Self-Locking Relay
Keypad and 5V Logic Circuit
Latching Alarm on/off Relay
Pushbutton 1-Minute Power Interrupt
12V DC Power Supply

Figure 4-114. *This block diagram shows the relative placement of the old and new components. The pushbutton power interrupt (which allows you to leave the house before the alarm switches itself back on) goes between the power supply and everything else. The latching relay substitutes for the DPDT switch on the previous version of the alarm. The transistor and self-locking relay, connected with the door and window magnetic switches, remain unchanged. The new delay circuit is inserted between the self-locking relay and the noisemaker. The test button is wired with the latching relay in the same way that it was wired previously with the DPDT switch.*

Upgrade 3: Delay Before Deactivation

Typically, alarms include another delay feature. When you open a door on your way into the building and it triggers the alarm, you have 30 seconds to deactivate it, before it starts making a noise.

How can we implement this delay feature? If I try to use another 555 timer to generate a pulse to inhibit the noise, that won't work, because the output from either the transistor or the relay can continue indefinitely. The relay locks itself on, and the transistor can continue passing voltage for as long as someone leaves a door open. If either of these signals activates a timer in monostable mode, the pulse from the timer will never end, and it will suppress the alarm indefinitely.

I think what I have to do is use a resistor and a capacitor to create a delay. I'll power them through the existing relay, so that I can be sure that they'll receive the full voltage of the power supply, after beginning from zero. Gradually the capacitor will acquire voltage—but I can't connect this directly to the noisemaker, because the noisemaker will gradually get louder as the voltage increases.

I have to insert a device that will be triggered to give full voltage when the input rises past a certain point. To do this, I'll use a 555 timer that's wired in bistable mode. This kind of modification is generally known as a "kludge," because it's not elegant, uses too many components, and does not use them appropriately. What I really need is a comparator, but I don't have space to get into that topic. So, using the knowledge that you have so far, the schematic in Figure 4-115 shows how a delay can be added to the alarm—not elegantly, but reliably.

The only problem is that if you power up a 555 timer in bistable mode, there's a 50-50 chance that the timer starts itself with its output high or low. So I need to pull the voltage low on the reset pin (to start the timer with its output inhibited) and gradually let it become positive (to permit the output). At the same time, I want to start with the voltage high on the trigger pin and gradually lower it, until it falls below 1/3 of the power supply and triggers the output.

So there are two timing circuits. The one for the reset pin works faster than the one on the trigger pin, so that at the point when the timer is triggered, the reset won't stop it.

The schematic shows component values that will do this. The 10 µF capacitor starts low but is charged through the 10K resistor in a couple of seconds. The timer is now ready to be triggered. But the 68 µF capacitor starts high (being connected with the positive side of the power supply) and takes a full minute to be pulled down to 1/3 of supply voltage through the 1M resistor. At that point, its voltage is low enough to trigger the 555. The timer output goes high and supplies the noisemaker.

You should be able to insert this little delay module in your alarm box, between the output from the relay and the input to the noisemaker without too much trouble. And if you want to adjust the delay, just use a higher or lower value resistor than 1M.

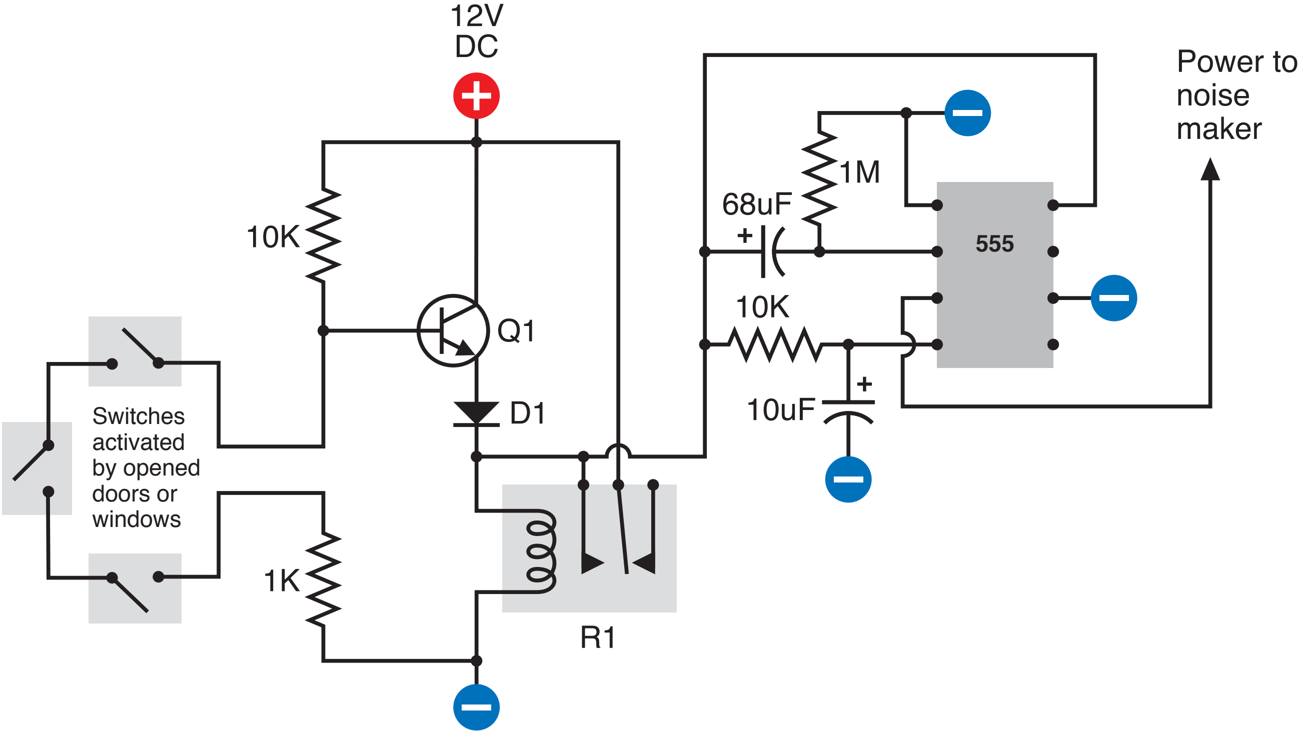

Figure 4-115. *This addition to the original alarm circuit imposes a one-minute delay before the alarm sounds. The 555 timer (wired in bistable mode) receives power through relay R1. The lower timing circuit initially applies negative voltage to the reset, ensuring that the 555 powers up with its output suppressed. This voltage quickly rises. Meanwhile the upper timing circuit applies a voltage to the trigger that gradually diminishes as the 68 μF capacitor equalizes its charge through the 1M resistor. When the voltage diminishes to 1/3 of the supply, the timer's output goes high and starts the noisemaker. If the power to the circuit is interrupted at any time before this, the relay relaxes, the capacitors gradually discharge, and the alarm does not sound.*

The Wrap-Up

If you add these three enhancements, your alarm will have all the features on my original wish list. Of course, if I were designing it from scratch, with all the information that has been added in this chapter of the book, it could be more elegant. But the modifications have not entailed making destructive changes to our original project, and all the design goals have been met.

What Next?

5

At this point, we can branch out in numerous directions. Here are some possibilities:

Audio electronics
This is a field in itself, including hobby projects, such as simple amplifiers and "stomp boxes," to modify guitar sound.

Radio-frequency devices
Anything that receives or transmits radio waves, from an ultra-simple AM radio to remote controllers.

Motors
The field of robotics has encouraged the growth of many online sites selling stepper motors, gear motors, synchronous motors, servo motors, and more.

Programmable microcontrollers
These are tiny computers on a single chip. You write a little program on your desktop computer, which will tell the chip to follow a series of procedures, such as receiving input from a sensor, waiting for a fixed period, and sending output to a motor. Then you download your program onto the chip, which stores it in nonvolatile memory. Popular controllers include the PICAXE, BASIC Stamp, Arduino, and many more. The cheapest ones retail for a mere $5 each.

Obviously, I don't have space to develop all of these topics fully, so what I'm going to do is introduce you to them by describing just one or two projects in each category. You can decide which interests you the most, and then proceed beyond this book by reading other guides that specialize in that interest.

I'm also going to make some suggestions about setting up a productive work area, reading relevant books, catalogs, and other printed sources, and generally proceeding further into hobby electronics.

IN THIS CHAPTER

Shopping List: Experiments 25 Through 36
Customizing Your Work Area
Reference Sources
Experiment 25: Magnetism
Experiment 26: Tabletop Power Generation
Experiment 27: Loudspeaker Destruction
Experiment 28: Making a Coil React
Experiment 29: Filtering Frequencies
Experiment 30: Fuzz
Experiment 31: One Radio, No Solder, No Power
Experiment 32: A Little Robot Cart
Experiment 33: Moving in Steps
Experiment 34: Hardware Meets Software
Experiment 35: Checking the Real World
Experiment 36: The Lock, Revisited
In Closing

Shopping List: Experiments 25 Through 36

Tools

You won't need any new tools for this section of the book.

Supplies and Components

As we have progressed to the point where you may want to pick and choose which projects you attempt, I will list the supplies and components at the beginning of each experiment.

Customizing Your Work Area

At this point, if you're getting hooked on the fun of creating hardware but haven't allocated a permanent corner to your new hobby, I have some suggestions. Having tried many different options over the years, my main piece of advice is this: don't build a workbench!

Many hobby electronics books want you to go shopping for 2×4s and plywood, as if a workbench has to be custom-fabricated to satisfy strict criteria about size and shape. I find this puzzling. To me, the exact size and shape of a bench is not very important. I think the most important issue is storage.

I want tools and parts to be easily accessible, whether they're tiny transistors or big spools of wire. I certainly don't want to go digging around on shelves that require me to get up and walk across the room.

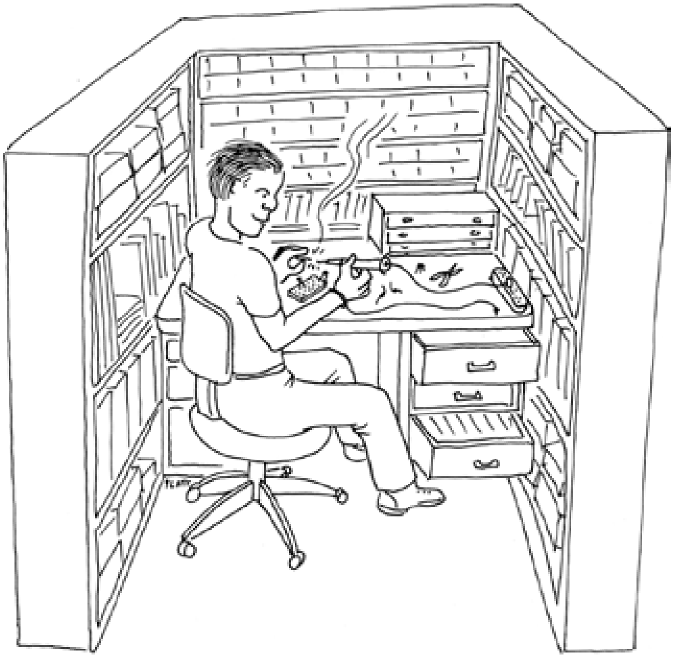

Figure 5-1. *The ideal work area: surrounded by storage. Never again will you need to get out of your chair!*

This leads me to two conclusions:

1. You need storage above the workbench.
2. You need storage below the workbench.

Many DIY workbench projects allow little or no storage underneath. Or, they suggest open shelves, which will be vulnerable to dust. My minimum configuration would be a pair of two-drawer file cabinets with a slab of 3/4-inch plywood or a Formica-clad kitchen countertop placed across them. File cabinets are ideal for storing all kinds of objects, not just files.

Of all the workbenches I've used, the one I liked best was an old-fashioned steel office desk—the kind of monster that dates back to the 1950s. They're difficult to move (because of their weight) and don't look beautiful, but you can buy them cheaply from used office furniture dealers, they're generous in size, they withstand abuse, and they last forever. The drawers are deep and usually slide in and out smoothly, like good file-cabinet drawers. Best of all, the desk has so much steel in it that you can use it to ground yourself before touching components that are sensitive to static electricity. If you use an anti-static wrist strap, you can simply attach it to a sheet-metal screw that you drive into one corner of the desk.

What will you put in the deep drawers of your desk or file cabinets? Some paperwork may be useful, perhaps including the following documents:

- Product data sheets
- Parts catalogs
- Sketches and plans that you draw yourself

The remaining capacity of each drawer can be filled with plastic storage boxes. The boxes can contain tools that you don't use so often (such as a heat gun or a high-capacity soldering iron), and larger-sized components (such as loudspeakers, AC adapters, project boxes, and circuit boards). You should look for storage boxes that measure around 11 inches long, 8 inches wide, and 5 inches deep, with straight sides. Boxes that you can buy at Wal-Mart will be cheaper, but they often have tapering sides (which are not space-efficient).

The boxes that I like best are Akro-Grids, made by Akro-Mils (see Figures 5-2 and 5-3). These are very rugged, straight-sided, with optional transparent snap-on lids. You can download the full Akro-Mills catalog from *http://www.akro-mils.com* and then search online for retail suppliers. You'll find that Akro-Mils also sells an incredible variety of parts bins, but I don't like open bins because their contents are vulnerable to dust and dirt.

For medium-size components, such as potentiometers, power connectors, control knobs, and toggle switches, I like storage containers measuring about 11 inches long, 8 inches wide, and 2 inches deep , divided into four to six sections. You can buy these from Michaels (the craft store), but I prefer to shop online for the Plano brand, as they seem more durably constructed. The Plano products that are most suitable for medium-size electronic parts are classified as fishing-tackle boxes, and you'll see them at *http://www.planomolding.com/tackle/products.asp*.

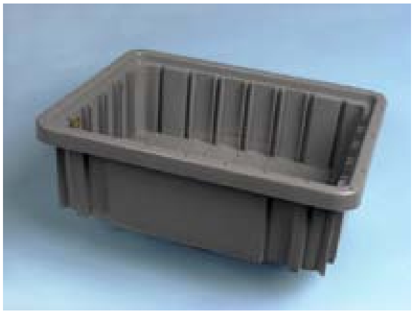

Figure 5-2. *Akro-Grid boxes contain grooves allowing them to be partitioned into numerous compartments for convenient parts storage.*

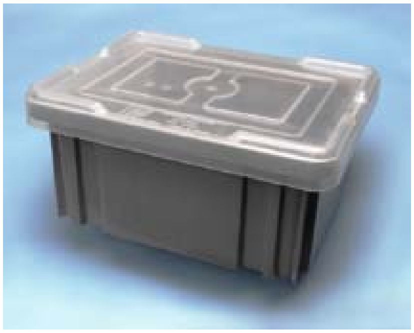

Figure 5-3. *Lids are sold separately for Akro-Grid boxes to keep the contents dust-free. The height of the box in Figure 5-2 allows three to be stacked in a typical file-cabinet drawer. The box shown here allows two to be stacked.*

For undivided, flat-format storage boxes, the Prolatch 23600-00 is ideally sized to fit a file-cabinet drawer, and the latches are sufficiently secure that you could stack a series of them on their long edges. See Figure 5-4.

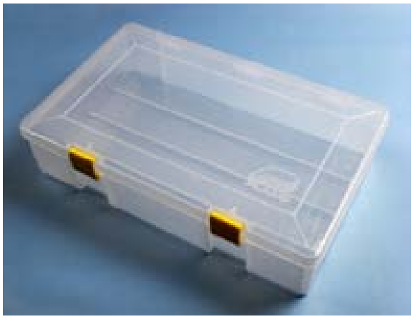

Figure 5-4. *This Plano brand box is undivided, making it useful for storing spools of wire or medium-size tools. When stacked upright on its long edge, three will fit precisely in a file-cabinet drawer.*

Plano also sells some really nicely designed toolboxes, one of which you can place on your desktop. It will have small drawers for easy access to screwdrivers, pliers, and other basics. Because you need a work area that's only about three feet square for most electronics projects, surrendering some desk space to a toolbox is not a big sacrifice.

If you have a steel desk with relatively shallow drawers, one of them can be allocated for printed catalogs. Don't underrate the usefulness of hard copy, just because you can buy everything online. The Mouser catalog, for instance, has an index, which is better in some ways than their online search feature, and the catalog is divided into helpful categories. Many times I've found useful parts that I never knew existed, just by browsing, which is much quicker than flipping through PDF pages online, even with a broadband connection. Currently, Mouser is still quite generous about sending out their catalogs, which contain over 2,000 pages. McMaster-Carr will also send you a catalog, but only after you've ordered from them, and only once a year.

Now, the big question: how to store all those dinky little parts, such as resistors, capacitors, and chips? I've tried various solutions to this problem. The most obvious is to buy a case of small drawers, each of which is removable, so you can place it on your desk while you access its contents. But I don't like this system, for two reasons. First, for very small components, you need to subdivide the drawers, and the dividers are never secure. And second, the removability of the drawers creates the risk of accidentally emptying the contents on the floor. Maybe you're too careful to allow this to happen, but I'm not!

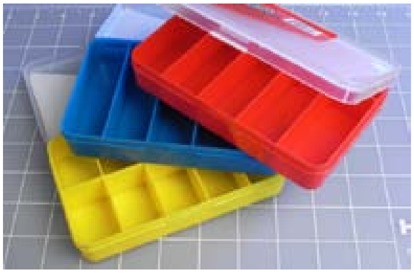

Figure 5-5. *Darice Mini-Storage boxes are ideal for components such as resistors, capacitors, and semiconductors. The boxes can be stacked stably or stored on shelves, with their ends labeled. The brand sticker is easily removed after being warmed with a heat gun.*

My personal preference is to use Darice Mini-Storage boxes, shown in Figure 5-5. You can find these at Michaels in small quantities, or buy them more economically in bulk online from suppliers such as *http://www.craftamerica.com*. The blue boxes are subdivided into five compartments that are exactly the right size and shape for resistors. The yellow boxes are subdivided into ten compartments, which are ideal for semiconductors. The purple boxes aren't divided at all, and the red boxes have a mix of divisions.

The dividers are molded into the boxes, so you don't have the annoyance associated with removable dividers that slip out of position, allowing components to mix together. The box lids fit tightly, so that even if you drop one of the boxes, it probably won't open. The lids have metal hinges, and a ridge around the edge that makes the boxes securely stackable.

I keep my little storage boxes on a set of shelves above the desk, with a gap of 3 inches between one shelf and the next, allowing two boxes to be stacked on each shelf. If I want to work with a particular subset of boxes, I shift them onto the desktop and stack them there.

Labeling

No matter which way you choose to store your parts, labeling them is essential. Any ink-jet printer will produce neat-looking labels, and if you use peelable (nonpermanent) labels, you'll be able to reorganize your parts in the future, as always seems to become necessary. I use color-coded labels for my collection of resistors, so that I can compare the stripes on a resistor with the code on the label, and see immediately if the resistor has been put in the wrong place. See Figure 5-6.

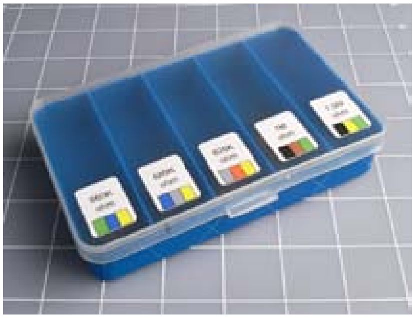

Figure 5-6. *To check that resistors are not placed in the wrong compartments, print the color code on each label.*

Even more important: you need to place a second (non-adhesive) label inside each compartment with the components. This label tells you the manufacturer's part number and the source, so that reordering is easy. I buy a lot of items from Mouser, and whenever I open their little plastic bags of parts, I snip out the section of the bag that has the identifying label on it, and slide it into the compartment of my parts box before I put the parts on top of it. This saves frustration later.

If I were *really* well organized, I would also keep a database on my computer listing everything that I buy, including the date, the source, the type of component, and the quantity. But I'm not that well organized.

On the Bench

Some items are so essential that they should sit on the bench or desktop on a permanent basis. These include your soldering iron(s), helping hands with magnifier, desk lamp, breadboard, power strip, and power supply. For a desk lamp, I prefer the type that has a daylight-spectrum fluorescent bulb, because it spreads a uniform light and helps me to distinguish colors of adjacent stripes on resistors.

The power supply is a matter of personal preference. If you're serious about electronics, you can buy a unit that delivers properly smoothed current at a variety of properly regulated and calibrated voltages. Your little wall-plug unit from RadioShack cannot do any of these things, and its output may vary depending on how heavily you load it. Still, as you've seen, it is sufficient for basic experiments, and when you're working with logic chips, you need to mount a 5-volt regulator on your breadboard anyway. Overall, I consider a good power supply optional.

Another optional item is an oscilloscope. This will show you, graphically, the electrical fluctuations inside your wires and components, and by applying probes at different points, you can track down errors in your circuit. It's a neat gadget to own, but it will cost a few hundred dollars, and for our tasks so far, it has not been necessary. If you plan to get seriously into audio circuits, an oscilloscope becomes far more important, because you'll want to see the shapes of the waveforms that you generate.

You can try to economize on an oscilloscope by buying a unit that plugs into the USB port of your computer and uses your computer monitor to display the signal. I have tried one of these, and was not entirely happy with the results. It worked, but did not seem accurate or reliable for low-frequency signals. Maybe I was unlucky; I decided not to try any other brands.

The surface of your desk or workbench will undoubtedly become scarred by random scuffs, cut marks, and drops of molten solder. I use a piece of half-inch plywood, two feet square, to protect my primary work area, and I clamp a miniature vise to its edge. To reduce the risk of static electricity when working with sensitive components, I cover the plywood with a square of conductive foam. This is not cheap, but offers advantages in addition to protecting chips from being zapped. Instead of scattering stray components, I can stick them into the foam, like plants growing in a garden. And like a garden, I can divide it into sections, with resistors on one side, capacitors on the other, and chips straight ahead.

Inevitably, during your work you'll create a mess. Little pieces of bent wire, stray screws, fasteners, and fragments of stripped insulation tend to accumulate, and can be a liability. If metal parts or fragments get into a project that you're building, they can cause short circuits. So you need a trash container. But it has to be easy to use. I use a full-size garbage pail, because it's so big that I can't miss it when I throw something toward it, and I can never forget that it's there.

Last, but most essential: a computer. Now that all data sheets are available online, and all components can be ordered online, and many sample circuits are placed online by hobbyists and educators, I don't think anyone can work efficiently without quick Internet access. To avoid wasting space, I suggest you use a small, cheap laptop that has a minimal footprint. A possible workbench configuration, using a steel desk, is shown in Figure 5-7.

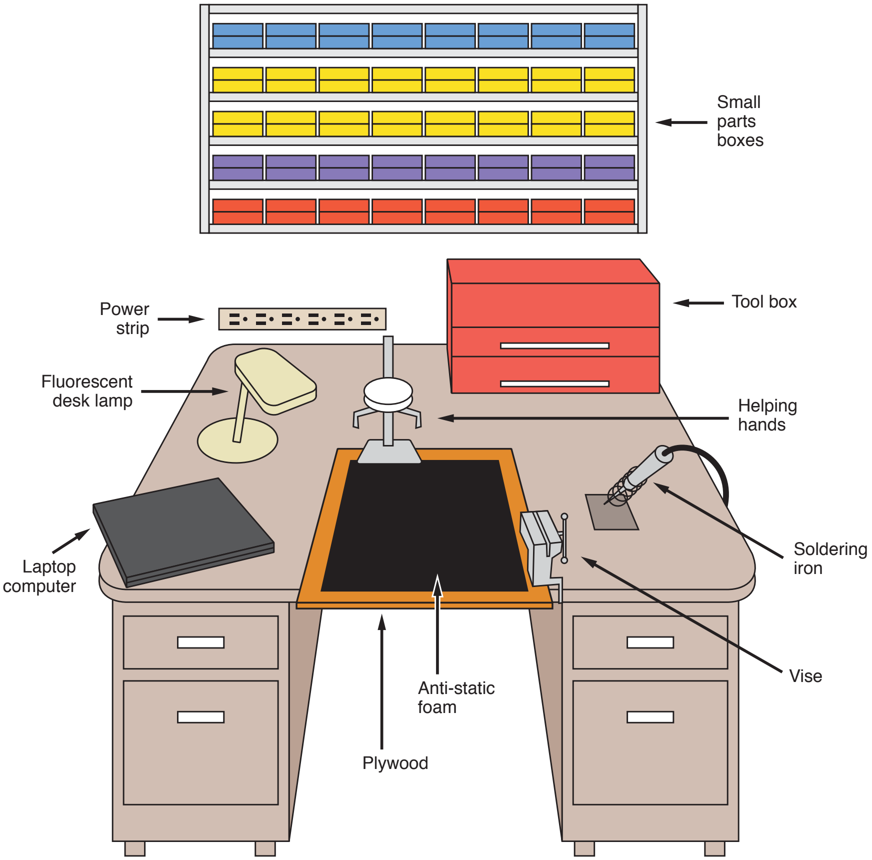

Figure 5-7. *An old steel office desk can be as good as, if not better than, a conventional workbench when building small electronics projects. It provides a large work area and ample storage, and has sufficient mass for you to ground yourself when dealing with components that are sensitive to static electricity.*

Reference Sources

Online

My favorite educational and reference site is Doctronics (*http://www.doctronics.co.uk)*. I like the way they draw their schematics, and I like the way they include many illustrations of circuits on breadboards (which most sites don't bother to do). They also sell kits, if you're willing to pay and wait for shipping from the UK. Part of a page from the doctronics website is reproduced in Figure 5-8.

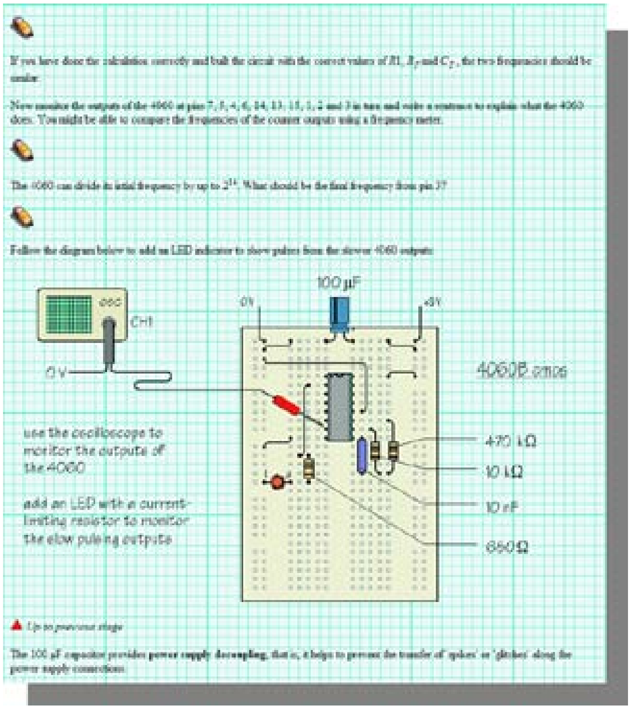
100 µF
0V
+9V
OSC
CH1
0 V
4060B CMOS
470 kΩ
10 kΩ
10 nF
680Ω
use the oscilloscope to monitor the outputs of the 4060
add an LED with a current-limiting resistor to monitor the slow pulsing outputs

Figure 5-8. *A sample page from http://www.doctronics.co.uk shows their detailed instructional approach. This is a valuable free online resource.*

My next favorite hobby site is also British-based: the Electronics Club (*http://www.kpsec.freeuk.com)*. It's not as comprehensive as Doctronics, but very friendly and easy to understand.

For a more theory-based approach, try *http://www.electronics-tutorials.ws*. This will go a little farther than the theory sections I've included here.

For an idiosyncratic selection of electronics topics, try Don Lancaster's Guru's Lair (*http://www.tinaja.com*). Lancaster wrote *The TTL Cookbook* more than 30 years ago, which opened up electronics to at least two generations of hobbyists and experimenters. He knows what he's talking about, and isn't afraid of getting into some fairly ambitious areas such as writing his own PostScript drivers and creating his own serial-port connections. You'll find a lot of ideas there.

Books

Yes, you do need books. As you're already reading this one, I won't recommend other beginners' guides. Instead, in keeping with the orientation of this chapter, I'll suggest some titles that will take you farther in various directions, and can be used for reference. I own all of these myself, and find them valuable:

Practical Electronics for Inventors, by Paul Scherz (McGraw-Hill, Second Edition, 2007)

This is a massive, comprehensive book, well worth the $40 cover price. Despite its title, you won't need to invent anything to find it useful. It's my primary reference source, covering a wide range of concepts, from the basic properties of resistors and capacitors all the way to some fairly high-end math. If you buy only one book (in addition to this one, of course!), this would be my recommendation.

Getting Started with Arduino, by Massimo Banzi (Make: Books, 2009)

If you enjoy the simplicity and convenience of the PICAXE programmable microcontroller that I describe later in this chapter, you'll find that the Arduino can do a lot more. *Getting Started* is the simplest introduction around, and will help to familiarize you with the Processing language used in Arduino (similar to the C language, not much like the version of BASIC used by the PICAXE).

Making Things Talk, by Tom Igoe (Make: Books, 2007)

This ambitious and comprehensive volume shows how to make the most of the Arduino's ability to communicate with its environment, even getting it to access sites on the Internet.

TTL Cookbook, by Don Lancaster (Howard W. Sams & Co, 1974)

The 1974 copyright date is not a misprint! You may be able to find some later editions, but whichever one you buy, it will be secondhand and possibly expensive, as this title now has collectible value. Lancaster wrote his guide before the 7400 series of chips was emulated on a pin-for-pin basis by CMOS versions, but it's still a good reference, because the concepts and part numbers haven't changed, and his writing is so accurate and concise.

CMOS Sourcebook, by Newton C. Braga (Sams Technical Publishing, 2001)

This book is entirely devoted to the 4000 series of CMOS chips, not the 74HC00 series that I've dealt with primarily here. The 4000 series is older and must be handled more carefully, because it's more vulnerable to static electricity than the generations that came later. Still, the chips remain widely available, and their great advantage is their willingness to tolerate a wide voltage range, typically from 5 to 15 volts. This means you can set up a 12-volt circuit that drives a 555 timer, and use output from the timer to go straight into CMOS chips (for example). The book is well organized in three sections: CMOS basics, functional diagrams (showing pinouts for all the main chips), and simple circuits showing how to make the chips perform basic functions.

The Encyclopedia of Electronic Circuits, by Rudolf F. Graf (Tab Books, 1985)

A totally miscellaneous collection of schematics, with minimal explanations. This is a useful book to have around if you have an idea and want to see how someone else approached the problem. Examples are often more valuable than general explanations, and this book is a massive compendium of examples. Many additional volumes in the series have been published, but start with this one, and you may find it has everything you need.

The Circuit Designer's Companion, by Tim Williams (Newnes, Second Edition, 2005)

Much useful information about making things work in practical applications, but the style is dry and fairly technical. May be useful if you're interested in moving your electronics projects into the real world.

The Art of Electronics, by Paul Horowitz and Winfield Hill (Cambridge University Press, Second Edition, 1989)

The fact that this book has been through 20 printings tells you two things: (1) Many people regard it as a fundamental resource; (2) Secondhand copies should be widely available, which is an important consideration, as the list price is over $100. It's written by two academics, and has a more technical approach than *Practical Electronics for Inventors*, but I find it useful when I'm looking for backup information.

Getting Started in Electronics, by Forrest M. Mims III (Master Publishing, Fourth Edition, 2007)

Although the original dates back to 1983, this is still a fun book to have. I think I have covered many of its topics here, but you may benefit by reading explanations and advice from a completely different source, and it goes a little farther than I have into some electrical theory, on an easy-to-understand basis, with cute drawings. Be warned that it's a brief book with eclectic coverage. Don't expect it to have all the answers.

Making Things
Getting Started with Arduino
Make:

Figure 5-9. *These books from MAKE provide guidance if you want to go beyond basic microcontrollers into the more exotic realms of the Arduino chip.*

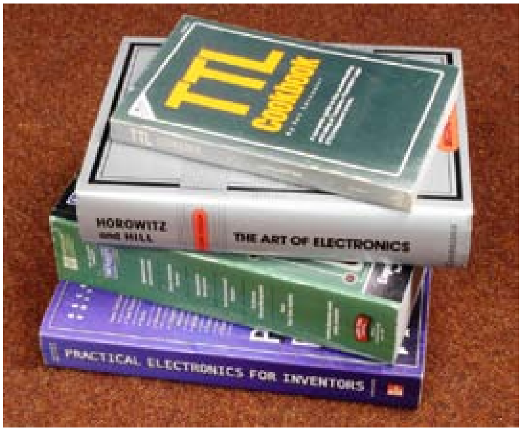

Figure 5-10. *A sun-damaged copy of the Don Lancaster's classic TTL Cookbook, a 2,000-page catalog from the Mouser Electronics supply company, and two comprehensive reference books that can provide years of additional guidance in all areas of electronics.*

Experiment 25: Magnetism

This experiment should be a part of any school science class, but even if you remember doing it, I suggest that you do it again, because setting it up takes only a matter of moments, and it's going to be our entry point to a whole new topic: the relationship between electricity and magnetism. Quickly this will lead us into audio reproduction and radio, and I'll describe the fundamentals of self-inductance, which is the third and final basic property of passive components (resistance and capacitance being the other two). I left self-inductance until last because it's not very relevant to the experiments that you've done so far. But as soon as we start dealing with analog signals that fluctuate, it becomes essential.

You will need:

- Large screwdriver.
- 22-gauge wire (or thinner). Quantity: 6 feet.
- AA battery.

Procedure

This couldn't be simpler. Wind the wire around the shaft of the screwdriver, near its tip. The turns should be neat and tight and closely spaced, and you'll need to make 100 of them, within a distance of no more than 2 inches. To fit them into this space, you'll have to make turns on top of previous turns. If the final turn tends to unwind itself (which will happen if you're using stranded wire), secure it with a piece of tape. See Figure 5-11.

Now apply a battery, as shown in Figure 5-12. At first sight, this looks like a very bad idea, because you're going to short out your battery just as you did in Experiment 2. But by passing the current through a wire that's coiled instead of straight, we'll get some work out of it before the battery expires.

Put a small paper clip near the screwdriver blade, on a soft, smooth surface that will not present much friction. A tissue works well. Because many screwdrivers are already magnetic, you may find that the paper clip is naturally attracted to the tip of the blade. If this happens, move the clip just outside the range of attraction. Now apply the 1.5 volts to the circuit, and the clip should jump to the tip of the screwdriver. Congratulations: you just made an electromagnet.

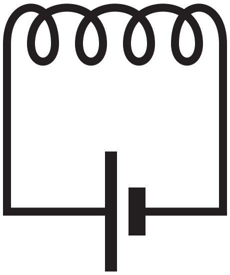

Figure 5-12. *A schematic can't get much simpler than this.*

FUNDAMENTALS

A two-way relationship

Every electric motor that was ever made uses some aspect of the relationship between electricity and magnetism. It's absolutely fundamental in the world around us. Remember that electricity can create magnetism:

When electricity flows through a wire, it creates a magnetic force around the wire.

The principle works in reverse: magnetism can create electricity.

When a wire moves through a magnetic field, it creates a flow of electricity in the wire.

This second principle is used in power generation. A diesel engine, or a water-powered turbine, or a windmill, or some other source of energy either turns coils of wire through a powerful magnetic field, or turns magnets amid some massive coils of wire. Electricity is induced in the coils. In the next experiment, you'll see a dramatic mini-demo of this effect.

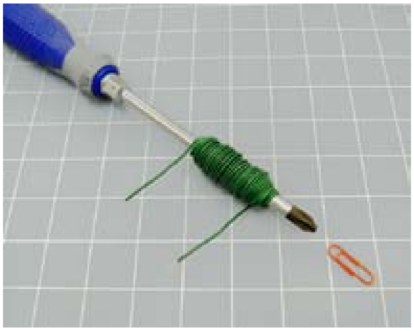

Figure 5-11. *Anyone who somehow missed this most basic childhood demo of electromagnetism should try it just for the fun of proving that a single AA battery can move a paper clip.*

THEORY

Inductance

When electricity flows through a wire, it creates a magnetic field around the wire. Because the electricity "induces" this effect, it is known as *inductance*. The effect is illustrated in Figure 5-13.

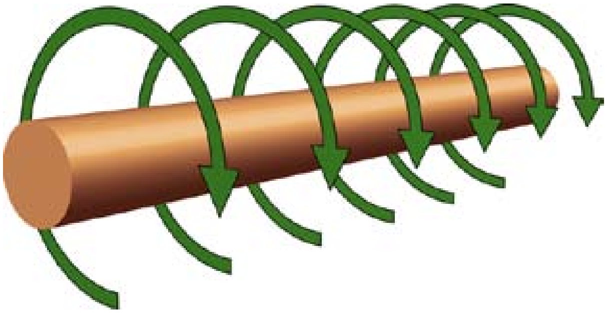

Figure 5-13. *When the flow of electricity is from left to right along this conductor, it induces a magnetic force shown by the green arrows.*

The field around a straight wire is very weak, but if we bend the wire into a circle, the magnetic force starts to accumulate, pointing through the center of the circle, as shown in Figure 5-14. If we add more circles, to form a coil, the force accumulates even more. And if we put a magnetic object (such as a screwdriver) in the center of the coil, the effectiveness increases further.

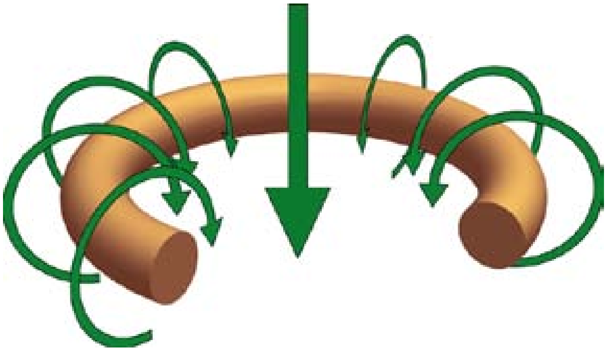

Figure 5-14. *When the conductor is bent to form a circle, the cumulative magnetic force acts through the center of the circle, as shown by the large arrow.*

Here's an approximated formula showing the relationship between the diameter of the coil, the width of the coil from end to end, the number of turns, and its inductance. The letter L is the symbol for inductance, even though the unit is the Henry, named after an American electrical pioneer named Joseph Henry:

L (in microHenrys) =
[(D × D) × (N x N)] / [(18 × D) + (40 × W)]
(Approximately)

In this formula, D is the diameter of the coil, N is the number of turns, and W is the width of the coil from end to end. See Figure 5-15. Here are three simple conclusions from this formula:

- Inductance increases with the diameter of the coil.
- Inductance increases with the square of the number of turns. (In other words, three times as many turns create nine times the inductance.)
- If the number of turns remains the same, inductance is lower if you wind the coil so that it's slender and long, but is higher if you wind it so that it's fat and short.

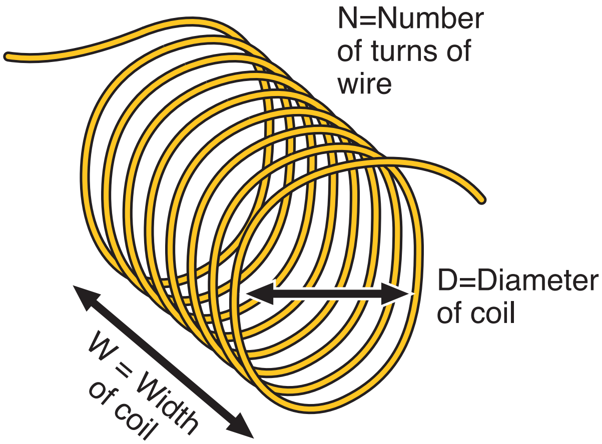

Figure 5-15. *The inductance of a coil increases with its diameter and with the square of its number of turns. If all other parameters remain the same, reducing the width (the distance from end to end) by packing the turns more tightly will increase the inductance.*

FUNDAMENTALS

Coil schematics and basics

Check the schematic symbols for coils in Figure 5-16. Note that if a coil has an iron core, this is shown with an extra couple of lines (sometimes only one line). If it has a ferrite core, the line is sometimes shown with dashes.

An iron core will add to the inductance of a coil, because it increases the magnetic effect.

A coil in isolation does not generally have any polarity. You can connect it either way around, but the magnetic force will be reversed accordingly (coils that interact with stuff—such as in transformers and solenoids—do have polarity).

Perhaps the most widespread application of coils is in transformers, where alternating current in one coil induces alternating current in another, often sharing the same iron core. If the primary (input) coil has half as many turns as the secondary (output) coil, the voltage will be doubled, at half the current—assuming hypothetically that the transformer is 100% efficient.

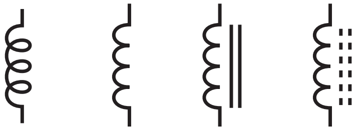

Figure 5-16. *Schematic diagrams represent coils. At far right is the older style. The third and fourth symbols indicate that the coil is wound around a solid or powdered magnetic core, respectively.*

BACKGROUND

Joseph Henry

Born in 1797, Joseph Henry was the first to develop and demonstrate powerful electromagnets. He also originated the concept of "self-inductance," meaning the "electrical inertia" that is a property of a coil of wire.

Henry started out as the son of a day laborer in Albany, New York. He worked in a general store before being apprenticed to a watchmaker, and was interested in becoming an actor. Friends persuaded him to enroll at the Albany Academy, where he turned out to have an aptitude for science. In 1826, he was appointed Professor of Mathematics and Natural Philosophy at the Academy, even though he was not a college graduate and described himself as being "principally self-educated." Michael Faraday was doing similar work in England, but Henry was unaware of it.

Henry was appointed to Princeton in 1832, where he received $1,000 per year and a free house. When Morse attempted to patent the telegraph, Henry testified that he was already aware of its concept, and indeed had rigged a system on similar principles to signal his wife, at home, when he was working in his laboratory at the Philosophical Hall.

Henry taught chemistry, astronomy, and architecture, in addition to physical science, and because science was not divided into strict specialties as it is now, he investigated phenomena such as phosphorescence, sound, capillary action, and ballistics. In 1846, he headed the newly founded Smithsonian Institution as its secretary.

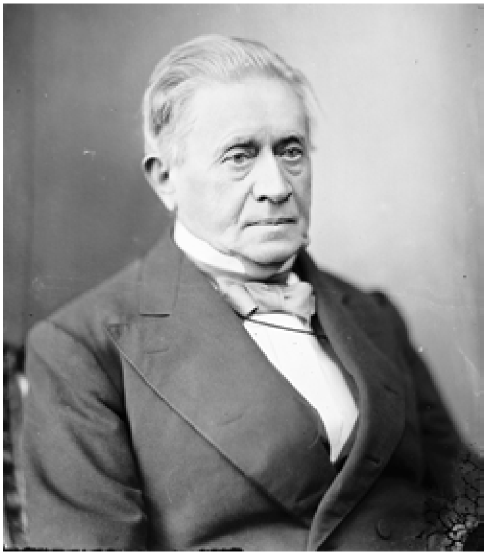

Figure 5-17. *Joseph Henry was an American experimenter who pioneered the investigation of electromagnetism. This photograph is archived in Wikimedia Commons.*

Experiment 26: Tabletop Power Generation

If you have just three components, you can see magnetism generating electricity right in front of you, right now.

You will need:

- Cylindrical neodymium magnet, 3/4-inch diameter, axially magnetized. Quantity: 1. (Obtainable online at sites such as *http://www.kjmagnetics.com*.)
- Spool of hookup wire, 26-gauge, 100 feet. Quantity: 1.
- Spool of magnet wire, quarter-pound, 26-gauge, about 350 feet. Quantity: 1. (Search online for sources for "magnet wire.")
- Generic LED. Quantity: 1.
- 100 µF electrolytic capacitor. Quantity: 1.
- Signal diode, 2N4001 or similar. Quantity: 1.
- Jumper wires with alligator clips on the ends. Quantity: 2.

Procedure

You may be able to make this experiment work with the spool of hookup wire, depending on the size of the spool relative to the size of your magnet, but as the results are more likely to be better with the magnet wire, I'll assume that you're using that—initially, at least. The advantage of the magnet wire is that its very thin insulation allows the coils to be closely packed, increasing their inductance.

First peek into the hollow center of the spool to see if the inner end of the wire has been left accessible, as is visible in Figures 5-18 and 5-19. If it hasn't, you have to unwind the wire onto any large-diameter cylindrical object, then rewind it back onto the spool, this time taking care to leave the inner end sticking out.

Scrape the transparent insulation off each end of the magnet wire with a utility knife or sandpaper, until bare copper is revealed. To check, attach your meter, set to measure ohms, to the free ends of the wire. If you make a good contact, you should measure a resistance of 30 ohms or less.

Place the spool on a nonmagnetic, nonconductive surface such as a wooden, plastic, or glass-topped table. Attach the LED between the ends of the wire using jumper wires. The polarity is not important. Now take a cylindrical neodymium magnet of the type shown in Figure 5-20 and push it quickly down into the hollow core, then pull it quickly back out. See Figure 5-21. You should see the LED blink, either on the down stroke or the up stroke.

The same thing may or may not happen if you use 100 feet of 26-gauge hookup wire. Ideally, your cylindrical magnet should fit fairly closely in the hollow center of the spool. If there's a big air gap, this will greatly reduce the effect of the magnet. Note that if you use a weaker, old-fashioned iron magnet instead of a neodymium magnet, you may get no result at all.

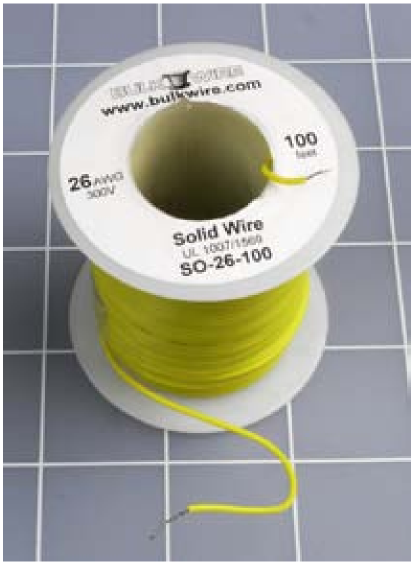

Figure 5-18. *An everyday 100-foot spool of hookup wire is capable of demonstrating the inductive power of a coil.*

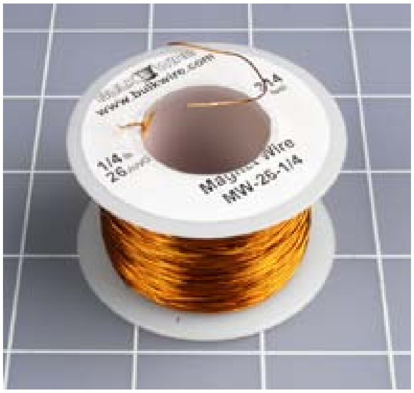

Figure 5-19. *Magnet wire has thinner insulation than hookup wire, allowing the turns to be more densely packed, inducing a more powerful magnetic field.*

Blood Blisters and Dead Media

Neodymium magnets can be hazardous. They're brittle and can shatter if they slam against a piece of magnetic metal (or another magnet). For this reason, many manufacturers advise you to wear eye protection.

Because a magnet pulls with increasing force as the distance between it and another object gets smaller, it closes the final gap very suddenly and powerfully. You can easily pinch your skin and get blood blisters.

If there's an object made of iron or steel anywhere near a neodymium magnet, the magnet will find it and grab it, with results that may be unpleasant, especially if the object has sharp edges and your hands are in the vicinity. When using a magnet, create a clear area on a nonmagnetic surface, and watch out for magnetic objects underneath the surface. My magnet sensed a steel screw embedded in the underside of a kitchen countertop, and slammed itself into contact with the countertop unexpectedly.

Be aware that magnets create magnets. When a magnetic field passes across an iron or steel object, the object picks up some magnetism of its own. Be careful not to magnetize your watch!

Don't use magnets anywhere near a computer, a disk drive, credit cards with magnetic stripes, cassettes of any type, and other media. Also keep magnets well away from TV screens and video monitors (especially cathode-ray tubes). Last but not least, powerful magnets can interfere with the normal operation of cardiac pacemakers.

Figure 5-20. *Three neodymium magnets, 1/4-, 1/2-, and 3/4-inch in diameter. I would have preferred to photograph them standing half-an-inch apart, but they refused to permit it.*

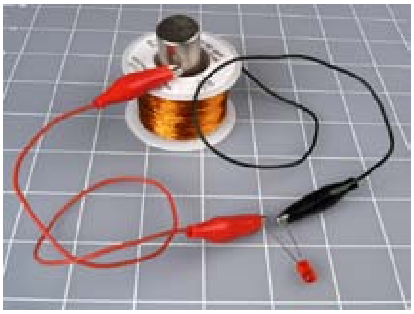

Figure 5-21. *By moving a magnet vigorously up and down through the center of a coil, you generate enough power to make the LED flash brightly.*

Here's another thing to try. Disconnect the LED and connect a 100 µF electrolytic capacitor in series with signal diode, as shown in Figure 5-23. Attach your meter, measuring volts, across the capacitor. If your meter has a manual setting for its range, set it to 20V DC. Make sure the positive (unmarked) side of the diode is attached to the negative (marked) side of the capacitor, so that positive voltage will pass through the capacitor and then through the diode.

Now move the magnet vigorously up and down in the coil. The meter should show that the capacitor is accumulating charge, up to about 10 volts. When you stop moving the magnet, the voltage reading will gradually decline, mostly because the capacitor discharges itself through the internal resistance of your meter.

This experiment is more important than it looks. Bear in mind that when you push the magnet into the coil, it induces current in one direction, and when you pull it back out again, it induces current in the opposite direction. You are actually generating alternating current.

The diode only allows current to flow one way through the circuit. It blocks the opposite flow, which is how the capacitor accumulates its charge. If you jump to the conclusion that diodes can be used to change alternating current to direct current, you're absolutely correct. We say that the diode is "rectifying" the AC power.

Experiment 24 showed that voltage can create a magnet. Experiment 25 has shown that a magnet can create voltage. We're now ready to apply these concepts to the detection and reproduction of sound.

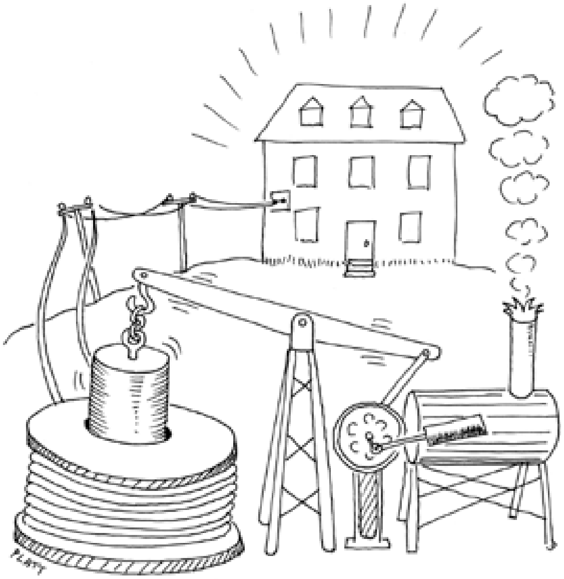

Figure 5-22. *Because inductance increases with the diameter of a coil and with the square of the number of turns, your power output from moving a magnet through the coil can increase dramatically with scale. Those wishing to live "off the grid" may consider this steam-powered configuration, suitable for powering a three-bedroom home.*

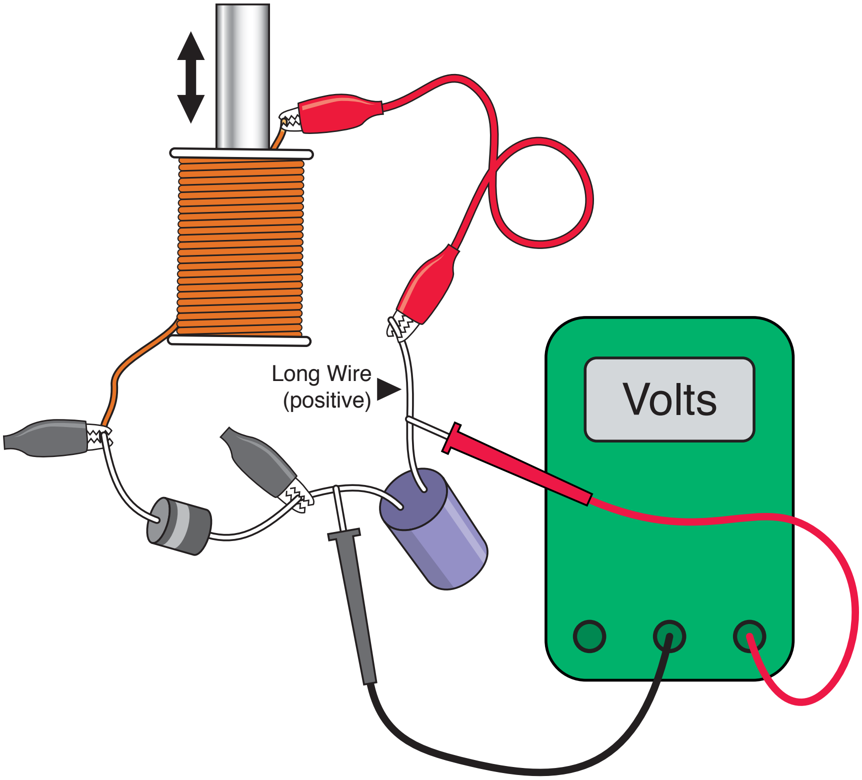

Figure 5-23. *Using a diode in series with a capacitor, you can charge the capacitor with the pulses of current that you generate by moving the magnet through the center of the coil. This demo illustrates the principle of rectifying alternating current.*

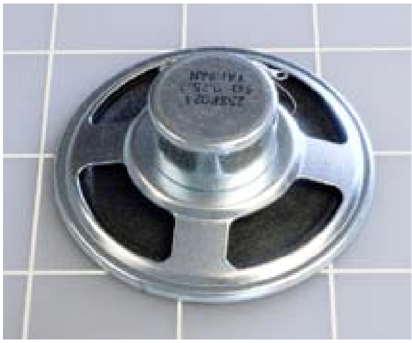

Figure 5-24. *A 2-inch loudspeaker can be instructively destroyed with a utility knife or X-Acto blade.*

Experiment 27: Loudspeaker Destruction

I'd like you to sacrifice a 2-inch loudspeaker, even though it means wasting the $5 or so that it probably costs. Actually, I don't consider this a waste, because if you want to learn how a component works, there's no substitute for actually seeing inside it. You might also have such a speaker already, part of a piece of cast-off personal electronics or toy you have in your basement.

You will need:

- Cheapest possible 2-inch loudspeaker. Quantity: 1. Figure 5-24 shows a typical example.

Procedure

Turn the loudspeaker face-up (as shown in Figure 5-25) and cut around the edge of its cone with a sharp utility knife or X-Acto blade. Then cut around the circular center and remove the ring of black paper that you've created. The result should look like Figure 5-26: you should see the flexible neck of the loudspeaker, which is usually made from a yellow weave. If you cut around its edge, you should be able to pull up the hidden paper cylinder, which has the copper coil of the loudspeaker wound around it. In Figure 5-27, I've turned it over so that it is easily visible. The two ends of this copper coil normally receive power through two terminals at the back of the speaker. When it sits in the groove visible between the inner magnet and the outer magnet, the coil reacts to voltage fluctuations by exerting an up-and-down force in reaction to the magnetic field. This vibrates the cone of the loudspeaker and creates sound waves.

Large loudspeakers in your stereo system work exactly the same way. They just have bigger magnets and coils that can handle more power (typically, as much as 100 watts).

Whenever I open up a small component like this, I'm impressed by the precision and delicacy of its parts, and the way it can be mass-produced for such a low cost. I imagine how astonished the pioneers of electrical theory (such as Faraday and Henry) would be, if they could see the components that we take for granted today. Henry spent days and weeks winding coils by hand to create electromagnets that were far less efficient than this cheap little loudspeaker.

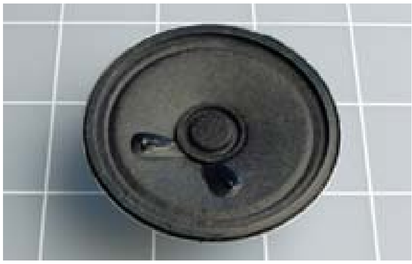

Figure 5-25. *Loudspeaker ready for creative destruction.*

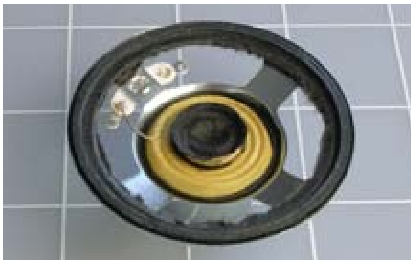

Figure 5-26. *The cone has been removed.*

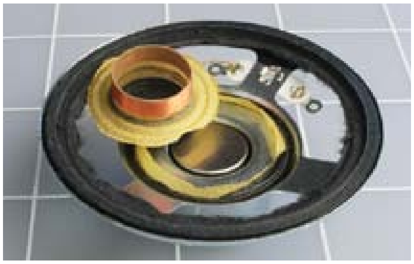

Figure 5-27. *The neck of the cone has been pulled out. Note the coil of copper wire, which fits precisely in the groove between two magnets in the base of the speaker.*

BACKGROUND

Origins of loudspeakers

Loudspeakers utilize the fact that if you run a varying electrical current through a coil situated in a magnetic field, the coil will move in response to the current. This idea was introduced in 1874 by Ernst Siemens, a prolific German inventor. (He also built the world's first electrically powered elevator in 1880.) Today, Siemens AG is one of the largest electronics companies in the world.

When Alexander Graham Bell patented the telephone in 1876, he used Siemen's concept to create audible frequencies in the earpiece. From that point on, sound-reproduction devices gradually increased in quality and power, until Chester Rice and Edward Kellogg at General Electric published a paper in 1925 establishing basic principles that are still used in loudspeaker design today.

At *http://www.radiolaguy.com/Showcase/Gallery-HornSpkr.htm* you'll find photographs of very beautiful early loudspeakers, which used a horn design to maximize efficiency. As sound amplifiers became more powerful, speaker efficiency became less important compared with quality reproduction and low manufacturing costs. Today's loudspeakers convert only about 1% of electrical energy into acoustical energy.

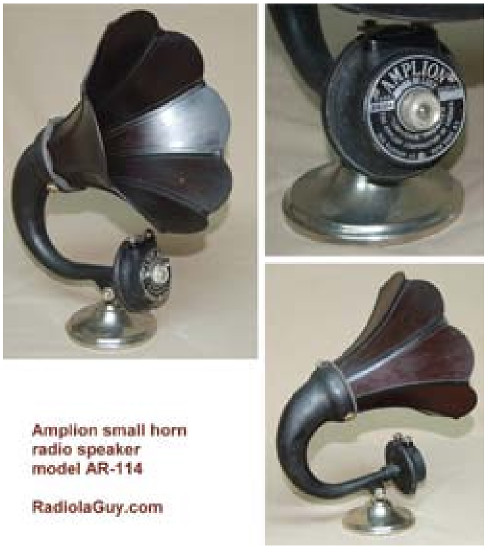

Figure 5-28. *This beautiful Amplion AR-114x illustrates the efforts of early designers to maximize efficiency in an era when the power of audio amplifiers was very limited. Photos by "Sonny, the RadiolaGuy." Many early speakers are illustrated at www.radiolaguy.com. Some are for sale.*

THEORY

Sound, electricity, and sound

Time now to establish a clear idea of how sound is transformed into electricity and back into sound again.

Suppose someone bangs a gong with a stick. The flat metal face of the gong vibrates in and out, creating sound waves. A sound wave is a peak of higher air pressure, followed by a trough of lower air pressure.

The wavelength of the sound is the distance (usually ranging from meters to millimeters) between one peak of pressure and the next peak.

The frequency of the sound is the number of waves per second, usually expressed as hertz.

Suppose we put a very sensitive little membrane of thin plastic in the path of the pressure waves. The plastic will flutter in response to the waves, like a leaf fluttering in the wind. Suppose we attach a tiny coil of very thin wire to the back of the membrane so that it moves with the membrane, and let's position a stationary magnet inside the coil of wire. This configuration is like a tiny, ultra-sensitive loudspeaker, except that instead of electricity producing sound, it is configured so that sound produces electricity. Sound pressure waves make the membrane move to and fro along the axis of the magnet, and the magnetic field creates a fluctuating voltage in the wire.

This is known as a *moving-coil* microphone. There are other ways to build a microphone, but this is the configuration that is easiest to understand. Of course, the voltage that it generates is very small, but we can amplify it using a transistor, or a series of transistors. Then we can feed the output through the coil around the neck of a loudspeaker, and the loudspeaker will recreate the pressure waves in the air. Figures 5-29 through 5-32 illustrate this sequence.

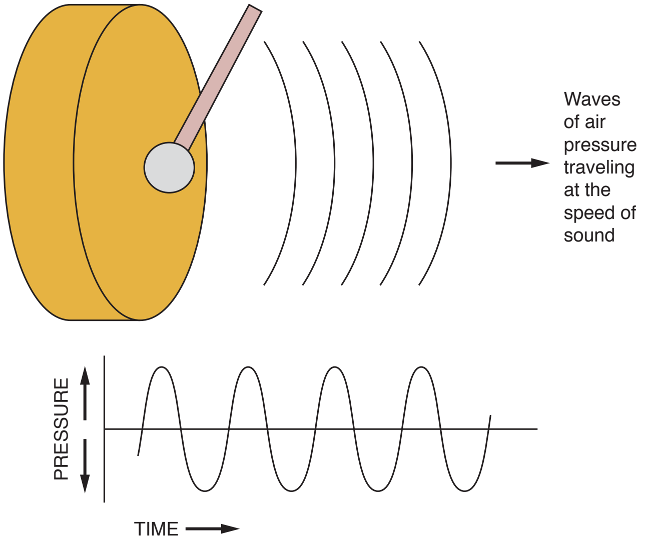

Figure 5-29. *Step 1 in the process of converting sound to electricity, and back again. When the hammer hits the gong, the face of the gong vibrates, creating pressure waves that travel through the air.*

THEORY

Sound, electricity, and sound (continued)

Somewhere along the way, we may want to record the sound and then replay it. But the principle remains the same. The hard part is designing the microphone, the amplifier, and the loudspeaker so that they reproduce the waveforms *accurately* at each step. It's a significant challenge, which is why accurate sound reproduction can be elusive.

Time now to think about what happens inside the wire when it generates a magnetic field. Obviously, some of the power in the wire is being transformed into magnetic force. But just what exactly is going on?

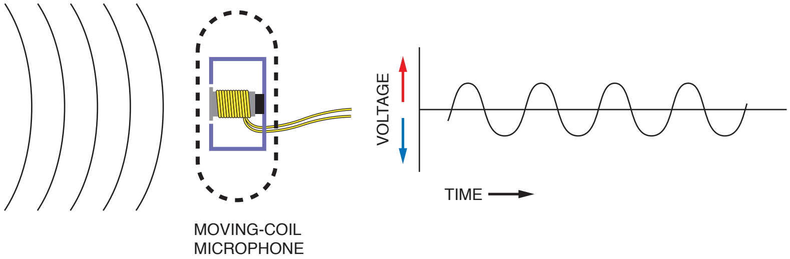

Figure 5-30. *Step 2: the pressure waves penetrate the perforated shell of a microphone and cause a diaphragm to vibrate in sympathy. The diaphragm has a coil attached to it. When the coil vibrates to and fro, a magnet at its center induces alternating current.*

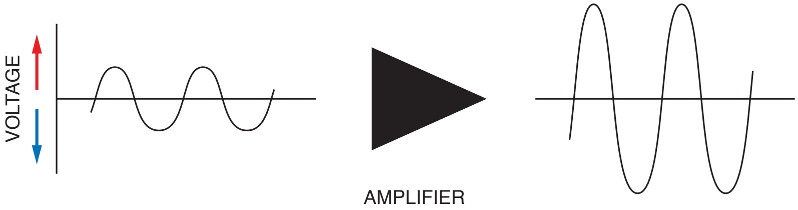

Figure 5-31. *Step 3: the tiny signals from the microphone pass through an amplifier, which enlarges their amplitude while retaining their frequency and the shape of their waveform.*

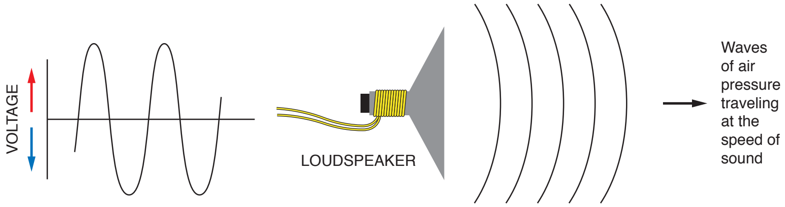

Figure 5-32. *Step 4: the amplified electrical signal is passed through a coil around the neck of a loudspeaker cone. The magnetic field induced by the current causes the cone to vibrate, reproducing the original sound.*

Experiment 28: Making a Coil React

A capacitor will absorb some DC current until it is fully charged, at which point it blocks the flow. There's another phenomenon that I haven't mentioned so far, which is the exact opposite of capacitance. It's known as *self-inductance*, and you find it in any coil of wire. Initially it blocks DC current (it reacts against it), but then its opposition gradually disappears. Here are a few definitions:

Resistance

Constrains current flow and drops voltage.

Capacitance

Allows current to flow initially and then blocks it. This behavior is properly known as capacitive reactance.

Self-Inductance

Blocks the flow of current initially and then allows it. This is also often referred to as *inductive reactance*. In fact, you may find the term "reactance" used as if it means the same thing, but since self-inductance is the correct term, I'll be using it here.

In this experiment, you'll see self-inductance in action.

You will need:

- LEDs, low-current type. Quantity: 2.
- Spool of hookup wire, 26-gauge, 100 feet. Quantity: 1.
- Resistor, 220Ω, rated 1/4 watt or higher. Quantity: 1.
- Capacitor, electrolytic, 2,000 µF or larger. Quantity: 1.
- SPST tactile switch. Quantity: 1.

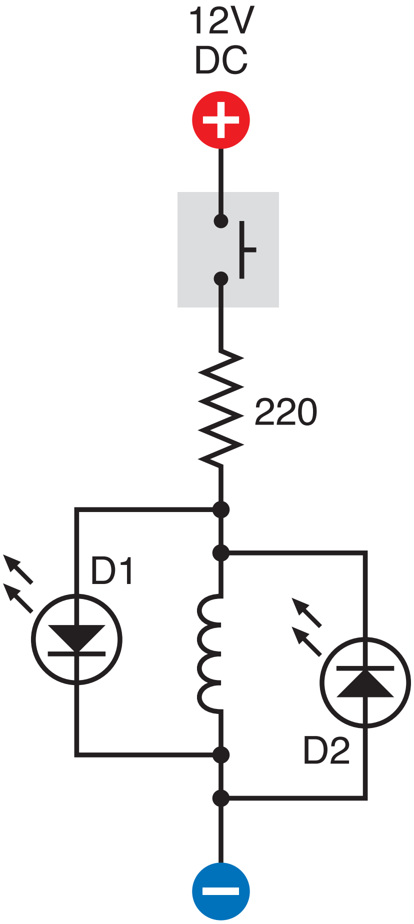

Figure 5-33. *In this demonstration of self-inductance, D1 and D2 are light-emitting diodes. When the switch is closed, D1 flashes briefly because the coil obstructs the initial flow of electricity. When the switch is opened, D2 flashes as the collapsing magnetic field induced by the coil releases another short burst of current.*

Procedure

Take a look at the schematic in Figure 5-33. At first it may not make much sense. The curly symbol is a coil of wire—nothing more than that. So apparently the voltage will pass through the 220Ω resistor, and then through the coil, ignoring the two LEDs, because the coil obviously has a much lower resistance than either of them (and one of them is upside-down anyway).

Is that what will happen? Let's find out. The coil can be a spool of 100 feet of 26-gauge (or smaller) hookup wire, although the spool of magnet wire listed in Experiment 25 will work better, if you have that. Once again, you will need access to both ends of the wire, and if the inner end is inaccessible, you'll need to rewind the coil, leaving the end sticking out.

Now that you have a coil, you can hook it up on your breadboard as shown in Figure 5-34, where the green circle is a tactile switch and the two circular red objects are LEDs. Make sure that you use low-current LEDs (otherwise, you may not see anything) and make sure that one of them is negative-side-up, positive-side-down and the other is positive-side-up, negative-side-down. Also, the 220Ω resistor should be rated at 1/4 watt or higher, if possible (see the following caution).

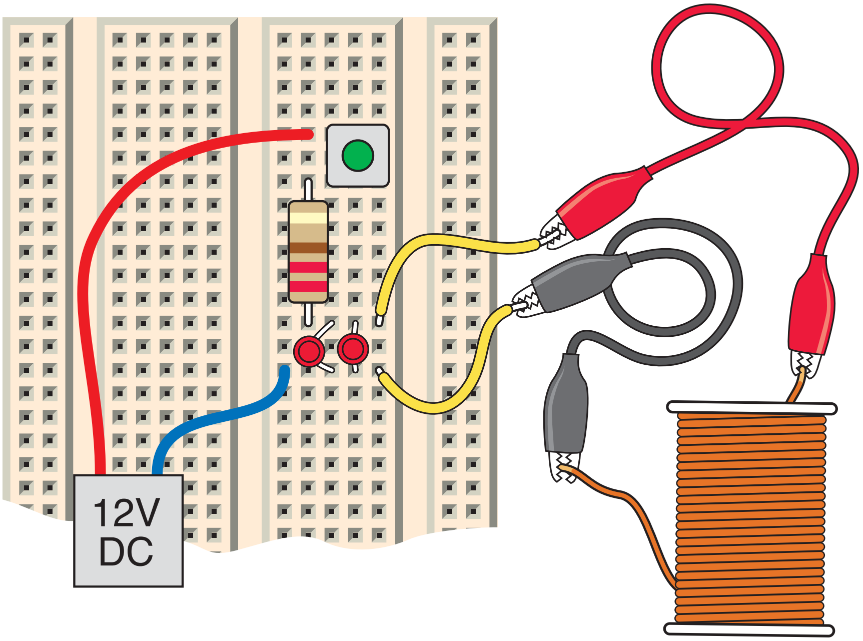

Figure 5-34. *The breadboarded version of the schematic in Figure 5-33 shows a simple way to set it up for a quick demo. The green button is a tactile switch. The two red LEDs should be placed so that the polarity of one is opposite to the polarity of the other.*

Hot Resistors

You'll be passing about 50mA through the 220Ω resistor, while the current is flowing. At 12 volts, this works out at 0.6 watts. If you use a 1/8-watt resistor, you will be overloading it, and it will get quite hot and may burn out. If you use a 1/4-watt resistor, it will still get hot, but is unlikely to burn out, as long as you don't press the button for more than a second or two.

Don't run the circuit without the coil of wire; you'll be trying to pass more than 50mA through the LEDs.

When you press the button, one LED should flash briefly. When you release the button, the other LED should flash.

What's happening here? The coil possesses self-inductance, which means that it reacts against any sudden change in the flow of electricity. First it fights it, and during that brief moment, it blocks most of the current. Consequently, the current looks for an alternative path and flows through D1, the lefthand LED in the schematic. (D2 doesn't respond, because it can pass current only in the opposite direction.)

Meanwhile, the voltage pressure overcomes the coil's self-inductance. When the self-inductance disappears, the resistance of the coil is no more than 10 ohms—so now the electricity flows mostly through the coil, and because the LED receives so little, it goes dark.

When you disconnect the power, the coil reacts again. It fights *any* sudden changes. After the flow of electricity stops, the coil stubbornly sustains it for a moment, because as the magnetic field collapses, it is turned back into electricity. This residual flow of current depletes itself through D2, the LED on the right.

In other words, the coil stores some energy in its magnetic field. This is similar to the way a capacitor stores energy between two metal plates, except that the coil blocks the current initially and then lets it build up, whereas the capacitor sucks up current initially, and then blocks it.

The more turns of wire you have in your coil, the more self-inductance the coil will have, causing your LEDs to flash more brightly.

Here's one last variation on this experiment to test your understanding of electrical fundamentals. Remove the 220Ω resistor, and substitute a 1K resistor (to protect your LED from sustained current). Remove the coil, and substitute a very large capacitor—ideally, about 4,700 μF. (Be careful to get its polarity the right way around.) What will you see when you press the button? Note that you will have to hold it down for a couple of seconds to get a result. And what will you see when you release the button? Remember: the behavior of capacitance is opposite to the behavior of self-inductance.

THEORY

Alternating current concepts

Here's a simple thought experiment. Suppose you set up a 555 timer to send a stream of pulses through a coil. This is a primitive form of alternating current.

We might imagine that the self-inductance of the coil will interfere with the stream of pulses, depending how long each pulse is, and how much inductance the coil has. If the pulses are too short, the self-inductance of the coil will tend to block them. Maybe if we can time the pulses exactly right, they'll synchronize with the time constant of the coil. In this way, we can "tune" a coil to allow a "frequency" to pass through it.

What happens if we substitute a capacitor? If the pulses are too long, compared with the time constant of the capacitor, it will tend to block them, because it will have enough time to become fully charged. But if the pulses are shorter, the capacitor can charge and discharge in rhythm with the pulses, and will seem to allow them through.

I don't have space in this book to get deeply into alternating current. It's a vast and complicated field where electricity behaves in strange and wonderful ways, and the mathematics that describe it can become quite challenging, involving differential equations and imaginary numbers. However, we can easily demonstrate the audio filtering effects of a loudspeaker and a coil.

Experiment 29: Filtering Frequencies

In this experiment, you'll see how self-inductance and capacitance can be used to filter audio frequencies. You're going to build a crossover network: a simple circuit that sends low frequencies to one place and high frequencies to another.

You will need:

- Loudspeaker, 8Ω, 5 inches in diameter. Quantity: 1. Figure 5-35 shows a typical example.
- Audio amplifier, STMicroelectronics TEA2025B or similar. Quantity: 1. See Figure 5-36.

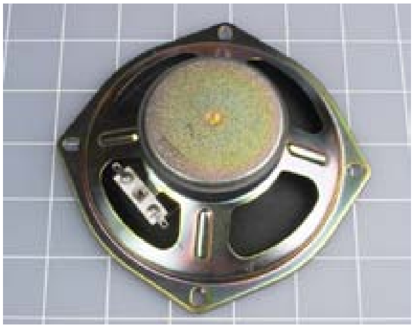

Figure 5-35. To hear the effects of audio filters using coils and capacitors, you'll need a loudspeaker capable of reproducing lower frequencies. This 5-inch model is the minimum required.

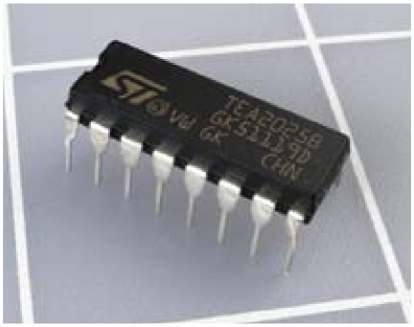

Figure 5-36. This single chip contains a stereo amplifier capable of delivering a total of 5 watts into an 8Ω speaker when the two channels are combined.

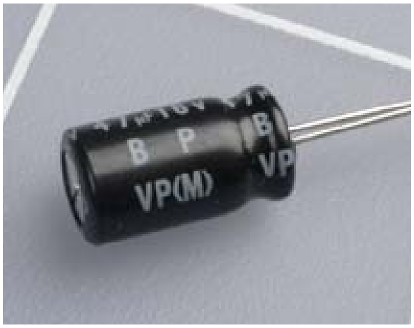

Figure 5-37. A nonpolarized electrolytic capacitor, also known as a bipolar capacitor, looks just like an electrolytic capacitor, except that it will have "NP" or "BP" printed on it.

- Nonpolarized electrolytic capacitors (also known as bipolar). 47 µF. Quantity: 2. A sample is shown in Figure 5-37. They should have "NP" or "BP" printed on them to indicate "nonpolarized" or "bipolar."
- Nonpolarized electrolytic capacitors (also known as bipolar). 100 µF. Quantity: 5. (Because you'll be working with audio signals that alternate between positive and negative, you can't use the usual polarized electrolytic capacitors. If you want to avoid the trouble and expense of ordering nonpolarized capacitors, you can substitute two regular electrolytics in series, facing in opposite directions, with their negative sides joined in the middle. Just remember that when you put capacitors in series, their total capacitance is half that of each individual component. Therefore, you would need two 220 µF electrolytics in series to create 110 µF of capacitance. See Figure 5-38.)
- Potentiometer, with audio taper if possible. 100K. Quantity: 1.
- Coil, for crossover network. Quantity: 1. You can search a source such as eBay for keywords "crossover" and "coil," but if you can't find one at a reasonable price, you can make do with a spool of 100 feet of 20-gauge hookup wire.
- Plastic shoebox. Quantity: 1.

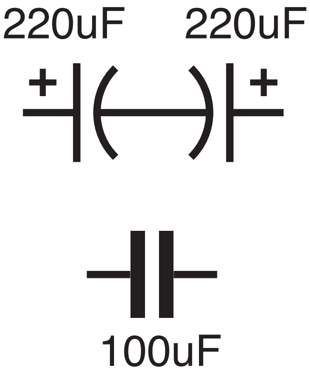

Figure 5-38. You can make a nonpolarized electrolytic capacitor by putting two regular electrolytics in series. (In fact, that's what you'd find if you opened a real nonpolarized capacitor.) The symbol at the bottom is roughly equivalent to the pair of symbols at the top; bear in mind that two capacitors in series have a total capacitance that is half that of each of them.

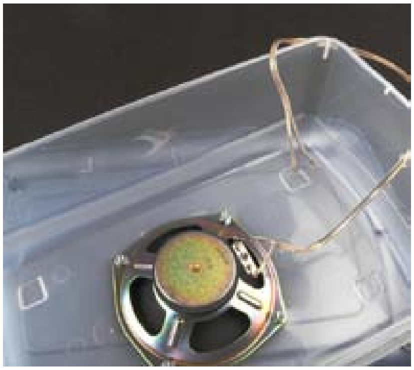

Figure 5-39. *A resonant enclosure is necessary if you want to hear some bass (lower frequencies) from your speaker. A cheap plastic shoebox is sufficient for demo purposes.*

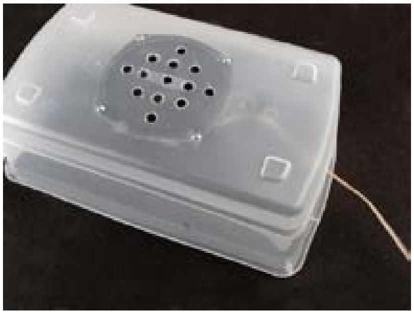

Figure 5-40. *Drill some half-inch holes in the bottom of the box, then bolt the speaker in place, running a wire out through a hole in one end. Snap on the lid, and you're ready for not-quite-high-fidelity audio.*

Procedure

The purpose of the audio amplifier chip is to provide enough power to get a decent amount of sound out of your loudspeaker. The purpose of using a 5-inch speaker is to enable you to hear lower-frequency sounds than the baby speakers that we have used previously. Bass notes have long wavelengths that small speakers are not able to generate effectively.

Maybe you remember from building the intrusion alarm that a speaker makes much more noise if you prevent the sound waves from the back of the cone from cancelling the sound waves from the front of the cone. The obvious way to achieve this is by enclosing the speaker in a box. I suggest a plastic box, because they're cheap, and we don't care too much about sound quality as long as we can hear at least some of the low frequencies. Figure 5-39 shows the speaker bolted into the bottom of a plastic box, and Figure 5-40 shows the box turned upside-down after snapping its lid into place.

Normally, a speaker should be mounted in a cabinet of heavy, thick material that has a very low resonant frequency—below the limits of human hearing. To minimize the resonance of the shoebox, you can put some soft, heavy fabric inside it before you snap the lid on. A hand towel or some socks should be sufficient to absorb some of the vibration.

Adding an Amplifier

Back in the 1950s, you needed vacuum tubes, transformers, and other power-hungry heavyweight components to build an audio amplifier. Today, you can buy a chip for about $1 that will do the job, if you add a few capacitors around it, and a volume control. The TEA2025B that I'm recommending is intended for use in cheap portable cassette players and CD players, and can work in stereo or mono mode, from a power supply ranging from 3 to 9 volts. With 9 volts and the two sides of the chip bridged together to drive one 8Ω speaker, it can generate 5 watts of audio power. That doesn't sound much compared with a typical home theater system rated at 100 watts per channel, but because loudness is a logarithmic scale, 5 watts will be quite enough to irritate any family members in the same room—and possibly even in other rooms.

If you can't find the TEA2025B chip, you can use any alternative listed as an audio amplifier. Try to find one that is designed to drive an 8Ω speaker with up to 5 watts in mono mode. Check the manufacturer's data sheet to see where you attach capacitors around it. Note carefully whether some of the capacitors have no polarity marked, even though they have fairly high values, such as 100 µF. These capacitors must function regardless of which way the alternating current is flowing, and I've marked them "NP" in my schematic in Figure 5-41, meaning "nonpolarized." (You may find them identified as "bipolar" or "BP" in parts catalogs.) As noted in the shopping list, you can put two 220 µF capacitors in series, negative-to-negative, to get the same effect as a single 100 µF nonpolarized capacitor.

For this project, it's essential to include the regular 100 µF electrolytic smoothing capacitor across the power supply. Otherwise, the amplifier will pick up and—yes, amplify—small voltage spikes in the circuit.

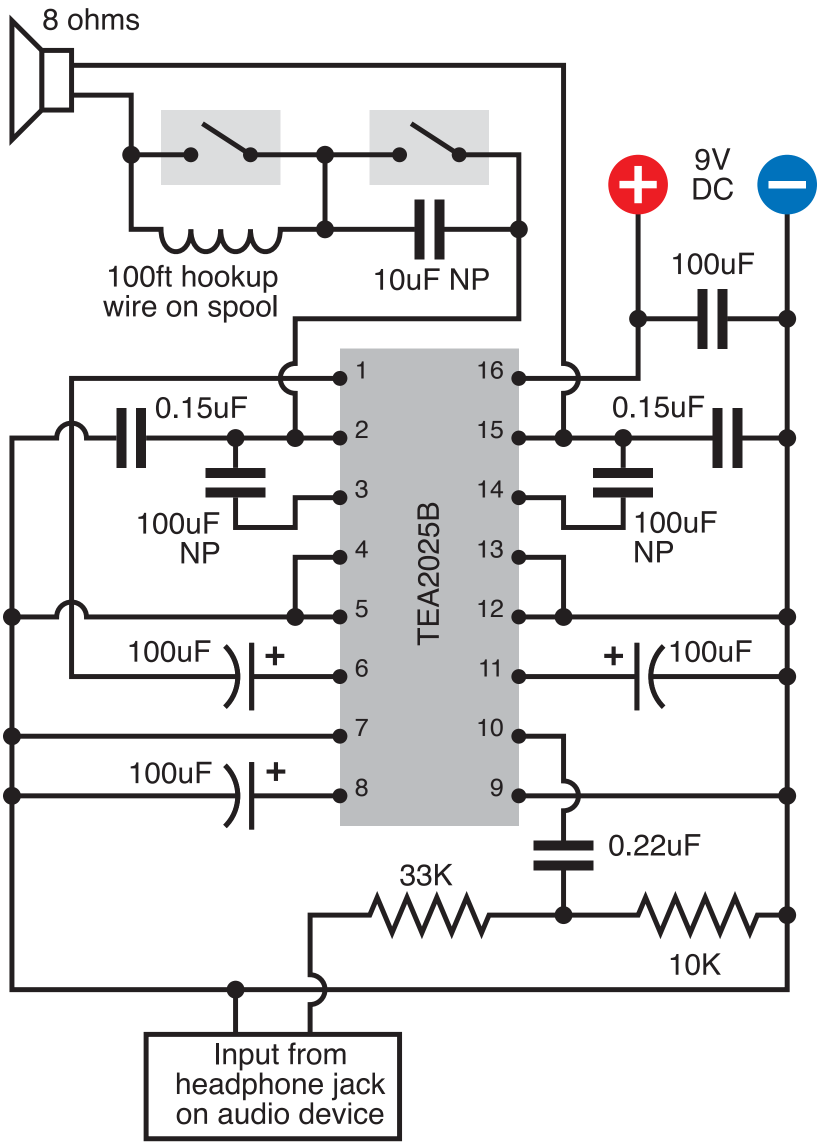
8 ohms
100ft hookup wire on spool
10uF NP
9V DC
100uF
0.15uF
0.15uF
100uF NP
100uF NP
TEA2025B
1
2
3
4
5
6
7
8
16
15
14
13
12
11
10
9
100uF
100uF
100uF
0.22uF
33K
10K
Input from headphone jack on audio device

Figure 5-41. *The audio amplifier chip should be wired with capacitors around it as shown, "NP" denoting the ones that are not polarized. The acronym "BP," meaning bipolar, is also often used to mean the same thing. The output from pins 2 and 15 of the chip can be passed through a coil or a 10 µF capacitor to demonstrate audio filtering.*

The input shown in the schematic can receive a signal from a typical media player, such as a portable MP3 player, CD player, or cassette player. To connect its headphone jack to the breadboard, you can use an adapter that converts it to a pair of RCA-type audio jacks, and then stick a wire into one of them as shown in Figure 5-42. The wire will connect to the 33K resistor on the breadboard circuit. The chromed neck of the RCA jack (which is sometimes gold-plated, or at least gold-colored) *must* be connected with the negative side of your power supply on the breadboard; otherwise, you won't hear anything. You can ignore the second output on the adapter, because we're working in mono, here, not stereo.

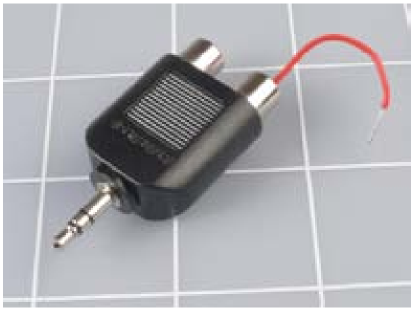

Figure 5-42. *To sample the output from the headphone socket of a music player, you can use this adapter and push the stripped end of a piece of hookup wire into one of the sockets. Then use alligator clips on a jumper wire to connect the audio to your breadboarded circuit. Don't forget to use an additional jumper wire to connect the outside of the socket to the negative side of the power supply on the breadboard. Because we're only using one speaker, the amplifier is connected to only one of the stereo outputs from your player. The other is ignored.*

The 33K resistor is necessary to protect the amplifier from being overdriven. If you don't get enough volume using your music player, decrease the 33K value. If the music is too loud and distorted, increase the value. You can also try omitting or increasing the 10K resistor next to it, which is included in an effort to reduce background hum noise.

I've shown two switches at the top of the schematic: one to bypass a coil, the other to bypass a capacitor. You can use alligator clips instead, as long as you can easily compare the sound when each of the components is inserted into the circuit.

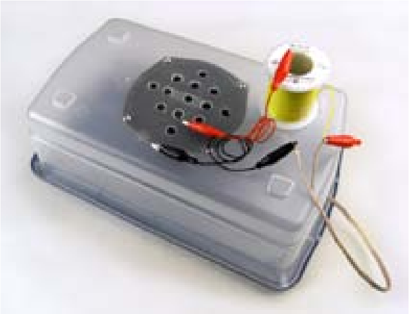

Figure 5-43. *The red and black alligator clips, lying on top of the shoebox, should connect with the output from your amplifier chip. The red jumper wire passes the signal through a coil of hookup wire on its way to the speaker. Note the change in sound when you short out the coil.*

Figure 5-43 shows a coil consisting of a spool of hookup wire being used. The red and black alligator clips resting loose on top of the shoebox will go to the output from the chip (on pins 2 and 15). There is no polarity; it doesn't matter which clip goes to which pin.

Begin by turning the volume control on your music source *all the way down* before you apply power. Don't be surprised if you hear humming or crackling noises when you activate the amplifier; it will pick up any stray voltages, because in this simple experiment, I haven't suggested that you should shield the input, and the amplifier circuit can pick up noise, as its wires can act like antennas.

Note that you may also get additional unwanted sound if you use the amplifier on a conductive desktop surface. Remove any aluminum foil or conductive foam for this project.

Make sure that your player is playing music, and slowly turn up its volume control until you hear it. If you don't hear anything, you'll have to check for circuit errors.

Now comes the interesting part. Insert the 100-foot spool of hookup wire between one output from the amplifier, and one input of the speaker (it doesn't matter which one), or if you used switches, open the switch that bypasses the coil. You should find that the music loses all its high-end response. By comparison, if you disconnect the coil and substitute a 10 µF capacitor, you should find that the music sounds "tinny," meaning that it loses all its low range, leaving only the high frequencies.

You've just tested two very simple filters. Here's what they are doing:

- The coil is a low-pass filter. It passes low frequencies but blocks high frequencies, because brief audio cycles don't have time to overcome the coil's self-inductance. A bigger coil eliminates a wider range of frequencies.
- The capacitor is a high-pass filter. It passes high frequencies and blocks low frequencies because longer audio cycles can fill the capacitance, at which point the capacitor stops passing current. A smaller capacitor eliminates a wider ranger of frequencies.

You can go a lot farther into filter design, using complex combinations of coils and capacitors to block frequencies at any point in the audible spectrum. Search online for audio filter schematics—you'll find hundreds of them.

Crossover Networks

In a traditional audio system, each speaker cabinet contains two drivers—one of them a small speaker called a *tweeter*, which reproduces high frequencies, the other a large speaker known as a *woofer*, which reproduces low frequencies. (Modern systems often remove the woofer and place it in a separate box of its own that can be positioned almost anywhere, because the human ear has difficulty sensing the direction of low-frequency sounds.)

The schematic that you just looked at and may have constructed is known as a "crossover network," and truly hardcore audiophiles have been known make their own (especially for use in car systems) to go with speakers of their choice in cabinets that they design and build themselves.

If you want to make a crossover network, you should use high-quality polyester capacitors (which have no polarity, last longer than electrolytics, and are better made) and a coil that has the right number of turns of wire and is the right size to cut high frequencies at the appropriate point. Figure 5-44 shows a polyester capacitor.

Figure 5-45 shows an audio crossover coil that I bought on eBay for $6. I was curious to find out what was inside it, so I bought two of them, and took one apart.

First I peeled away the black vinyl tape that enclosed the coil. Inside was some typical magnet wire—copper wire thinly coated with shellac or semitransparent plastic, as shown in Figure 5-46. I unwound the wire and counted the number of turns. Then I measured the length of the wire, and finally used a micrometer to measure the diameter of the wire, after which I checked online to find a conversion from the diameter in mils (1/1,000 of an inch) to American wire gauge.

As for the spool, it was plain plastic with an air core—no iron or ferrite rod in the center. Figure 5-47 shows the spool and the wire.

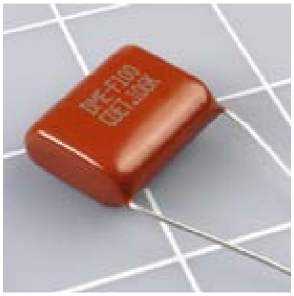

Figure 5-44. *Some nonelectrolytic capacitors have no polarity, such as this high-quality polyester film capacitor. However, they tend to be much more expensive, and are hard to find in values higher than 10 µF.*

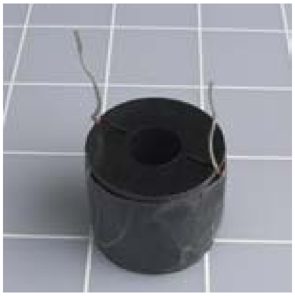
Figure 5-45. *What exotic components may we find inside this high-end audio component that's used with a subwoofer to block high frequencies?*

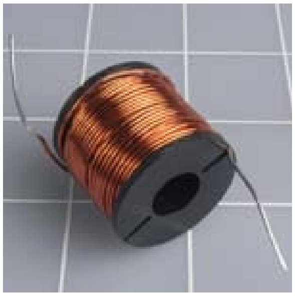
Figure 5-46. *The black tape is removed, revealing a coil of magnet wire.*

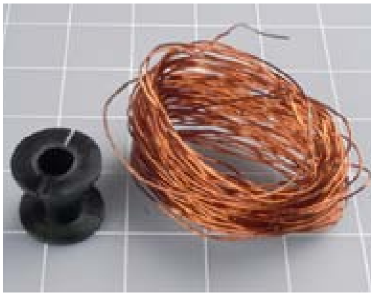
Figure 5-47. *The audio crossover coil consists of a plastic spool and some wire. Nothing more.*

So here's the specification for this particular coil in an audio crossover network. Forty feet of 20-gauge copper magnet wire, wrapped in 200 turns around a spool of 1/16–inch-thick plastic with a hub measuring 7/8 inch in length between the flanges and 1/2-inch external diameter. Total retail cost of materials if purchased separately: probably about $1, assuming you can find or make a spool of the appropriate size.

Conclusion: there's a lot of mystique attached to audio components. They are frequently overpriced, and you can make your own coil if you start with these parameters and adjust them to suit yourself.

Suppose you want to put some thumping bass speakers into your car. Could you build your own filter so that they only reproduce the low frequencies? Absolutely—you just need to wind a coil, adding more turns until it cuts as much of the high frequencies as you choose. Just make sure the wire is heavy enough so that it won't overheat when you push 100 or more audio watts through it.

Here's another project to think about: a color organ. You can tap into the output from your stereo and use filters to divide audio frequencies into three sections, each of which drives a separate set of colored LEDs. The red LEDs will flash in response to bass tones, yellow LEDs in response to the mid-range, and green LEDs in response to high frequencies (or whatever colors you prefer). You can put signal diodes in series with the LEDs to rectify the alternating current, and series resistors to limit the voltage across the LEDs to, say, 2.5 volts (when the music volume is turned all the way up). You'll use your meter to check the current passing through each resistor, and multiply that number by the voltage drop across the resistor, to find the wattage that it's handling, to make sure the resistor is capable of dissipating that much power without burning out.

Audio is a field offering all kinds of possibilities if you enjoy designing and building your own electronics.

THEORY

Waveforms

If you blow across the top of a bottle, the mellow sound that you hear is caused by the air vibrating inside the bottle, and if you could see the pressure waves, they would have a distinctive profile.

If you could slow down time and draw a graph of the alternating voltage in any power outlet in your house, it would have the same profile.

If you could measure the speed of a pendulum swinging slowly to and fro in a vacuum, and draw a graph of the speed relative to time, once again it would have the same profile.

That profile is a *sine wave*, so called because you can derive it from basic trigonometry. In a right-angled triangle, the sine of an angle is found by dividing the length of the side opposite the angle by the length of the hypoteneuse (the sloping side of the triangle).

To make this simpler, imagine a ball on a string rotating around a center point, as shown in Figure 5-48. Ignore the force of gravity, the resistance of air, and other annoying variables. Just measure the vertical height of the ball and divide it by the length of the string, at regular instants of time, as the ball moves around the circular path at a constant speed. Plot the result as a graph, and there's your sine wave, shown in Figure 5-49. Note that when the ball circles below its horizontal starting line, we consider its distance negative, so the sine wave becomes negative, too.

Why should this particular curve turn up in so many places and so many ways in nature? There are reasons for this rooted in physics, but I'll leave you to dig into that topic if it interests you. Getting back to the subject of audio reproduction, what matters is this:

- Any sound can be broken down into a mixture of sine waves of varying frequency and amplitude.

Or, conversely:

- If you put together the right mix of audio sine waves, you can create *any sound at all*.

Suppose that there are two sounds playing simultaneously. Figure 5-50 shows one sound as a red curve, and the other as pale blue. When the two sounds travel either as pressure waves through air or as alternating electric currents through a wire, the amplitudes of the waves are added together to make the more complex curve, which is shown in black. Now try to imagine dozens or even hundreds of different frequencies being added together, and you have an idea of the complex waveform of a piece of music.

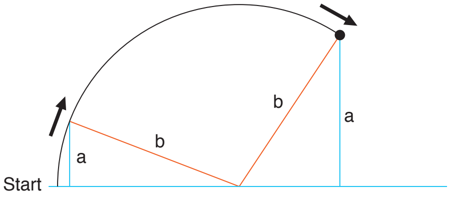

Figure 5-48. *If a weight on the end of a string (length b, in the diagram) follows a circular path at a steady speed, the distance of the weight from a horizontal center line (length a, in the diagram) can be plotted as a graph relative to time. The graph will be a sine wave, so called because in basic trigonometry, the ratio of a/b is the sine of the angle between line b and the horizontal baseline, measured at the center of rotation. Sinewaves occur naturally in the world around us, especially in audio reproduction and alternating current.*

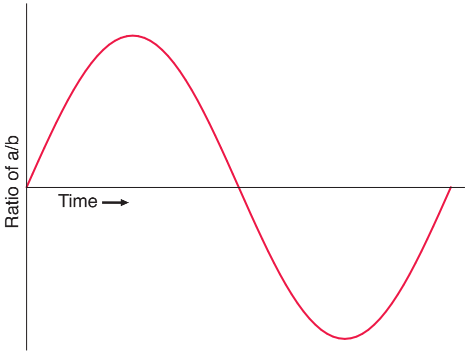

Figure 5-49. *This is what a "pure" sinewave looks like.*

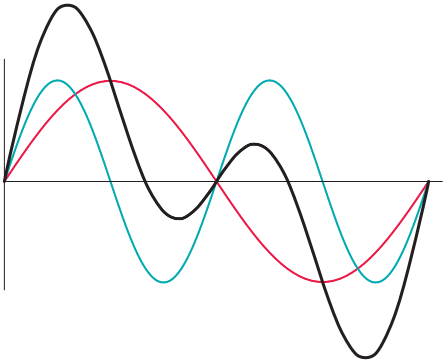

Figure 5-50. *When two sinewaves are generated at the same time (for instance, by two musicians, each playing a flute), the combined sound creates a compound curve. The blue sinewave is twice the frequency of the red sinewave. The compound curve (black line) is the sum of the distances of the sinewaves from the baseline of the graph.*

THEORY

Waveforms (continued)

You can create your own waveform as an input for your audio amplifier with the basic astable 555 timer circuit shown in Figure 5-51. You have to be careful, though, not to overload the amplifier input. Note the 680K series resistor on the output pin of the timer. Also note the 500Ω potentiometer.

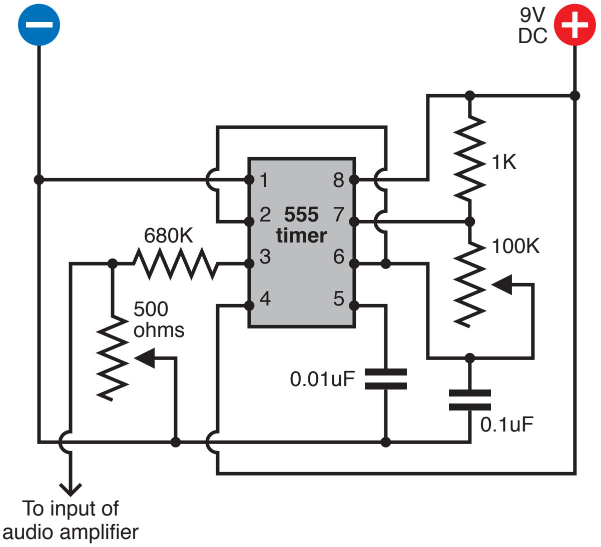

Figure 5-51. *A 555 timer is wired in astable mode using the component values shown here to generate a wide range of audible frequencies when the 100K potentiometer is adjusted. After the output is reduced in power, it can feed into the amplifier chip that was used previously.*

Disconnect your music player and connect the output from the 555 circuit to the input point (the 33K resistor) in the amplifier circuit shown earlier in Figure 5-41. You don't have to worry about a separate connection on the negative side as long as the 555 timer shares the same breadboard and the negative side of its power supply.

Make sure that the 500Ω potentiometer is turned all the way to short the output from the timer to the negative side of the power supply. This functions as your volume control. Also make sure the 100K potentiometer is in the middle of its range. Switch on the power and slowly turn up the 500Ω potentiometer until you hear a tone.

Now adjust the 100K potentiometer to create a low-pitched note. You'll find that it doesn't have a "pure" sound. There are some buzzing overtones. This is because the 555 timer is generating square waves such as those shown in Figure 5-52, not sine waves, and a square wave is actually a sum of many different sine waves, some of which have a high frequency. Your ear hears these harmonics, even though they are not obvious when you look at a square-shaped waveform.

Route one of the connections to your loudspeaker through your spool of hookup wire, and now you should hear a much purer tone, as the buzzing high frequencies are blocked by the self-inductance of the coil. Remove the coil and substitute the 10 μF capacitor, and now you hear more buzzing and less bass.

You've just taken a small step toward sound synthesis. If this subject interests you, you can go online and search for oscillator circuits. For a thorough understanding of the relationship between waveforms and the sounds you hear, you'll really need an oscilloscope, which will show you the shape of each waveform that you generate and modify.

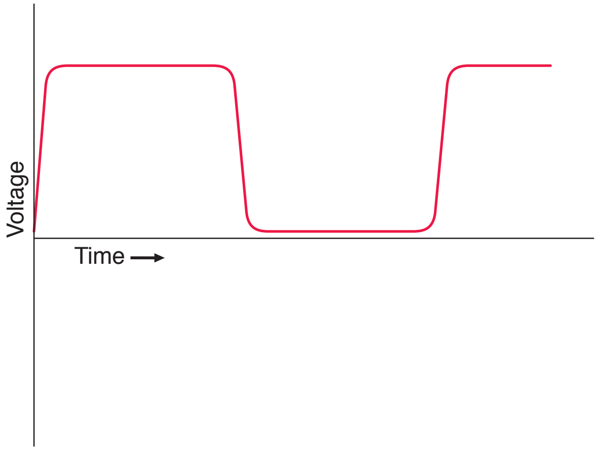

Figure 5-52. *The output from a 555 timer is either "on" or "off," with a very fast transition between those two states. The result is an almost perfect square wave. Theoretically, this can be disassembled into a complex set of sine waves that have many different frequencies. The human ear hears the high frequencies as harsh overtones.*

Experiment 30: Fuzz

Let's try one more variation on the circuit in Experiment 28. This will demonstrate another fundamental audio attribute: distortion.

You will need:

- One more 100K potentiometer.
- Generic NPN transistors: 2N2222 or similar. Quantity: 2.
- Various resistors and capacitors.

BACKGROUND

Clipping

In the early days of "hi-fi" sound, engineers labored mightily to perfect the process of sound reproduction. They wanted the waveform at the output end of the amplifier to look identical with the waveform at the input end, the only difference being that it should be bigger, so that it would be powerful enough to drive loudspeakers. Even a very slight distortion of the waveform was unacceptable.

Little did they realize that their beautifully designed tube amplifiers would be abused by a new generation of rock guitarists whose intention was to create as much distortion as possible.

The most common form of waveform abuse is technically known as "clipping." If you push a vacuum tube or a transistor to amplify a sine wave beyond the component's capabilities, it "clips" the top and bottom of the curve. This makes it look more like a square wave, and as I explained in the section on waveforms, a square wave has a harsh, buzzing quality. For rock guitarists trying to add an edge to their music, the harshness is actually a desirable feature.

Figure 5-53. *This Vox Wow-Fuzz pedal was one of the early stomp boxes, which deliberately induced the kind of distortion that audio engineers had been trying to get rid of for decades.*

The first gadget to offer this on a commercial basis was known as a "fuzz box," which deliberately clipped the input signal. An early fuzz box is shown in Figure 5-53. The clipping of a sine wave is shown in Figure 5-54.

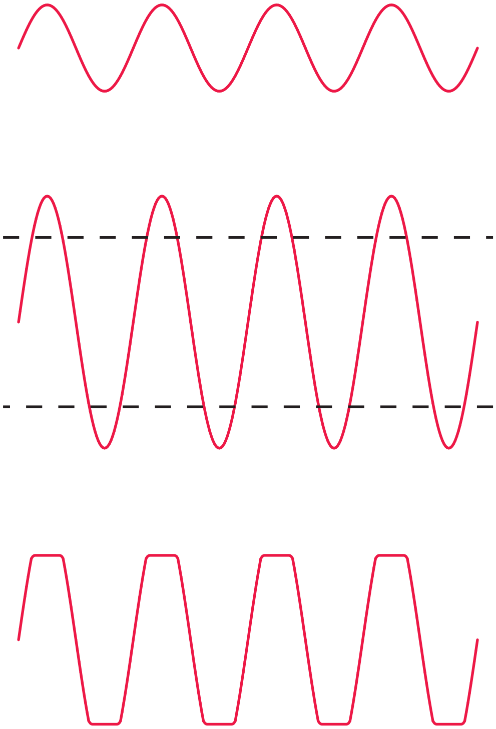

Figure 5-54. *When a sinewave (top) is passed through an amplifier which is turned up beyond the limit of its components (shown as dashed lines, center), the amplifier chops the wave (bottom) in a process known as "clipping." The result is close to a square wave and is the basic principle of a fuzz box commonly used to create a harsh guitar sound.*

Schematic

The output from the 555 timer is a square wave, so it already sounds quite "fuzzy," but we can make it more intense to demonstrate the clipping principle. I've redrawn the whole circuit in Figure 5-55, as several components have changed. The principal alteration is the addition of two NPN transistors.

If you assemble this circuit on your breadboard, note carefully that the 33K and 10K resistors at the bottom of the amplifier chip have been removed, and there's now just an 820Ω resistor in that location. The bottom of the adjacent 0.22 µF capacitor is still the input point for the amplifier, and if you follow the connection around to the middle of the schematic, you'll find it leading to a 100K potentiometer. This is your "fuzz adjuster."

The two NPN transistors are arranged so that the one on the left receives output from the 555 timer. This signal controls the flow of electricity through the transistor from a 33K resistor. This flow, in turn, controls the base of the righthand transistor, and the flow of current through that is what ultimately controls the amplifier.

When you power up the circuit, use the 100K potentiometer attached to the 555 timer to adjust the frequency (as before) and crank the "fuzz adjuster" potentiometer to hear how it adds increasing "bite" to the sound until ultimately it becomes pure noise.

The two transistors act as amplifiers. Of course, we didn't need them for that purpose—the input level for the amplifier chip was already more than adequate. The purpose of the lefthand transistor is simply to overload the righthand transistor, to create the "fuzz" effect. And when you turn up the output from the transistors with the "fuzz adjuster," eventually they overload the input of the amplifier chip, creating even more distortion.

If you want to tweak the output, try substituting different values for the 1K resistor and the 1 µF capacitor that are positioned between the emitter of the righthand transistor and the negative side of the power supply. A larger resistor should overload the transistor less. Different capacitor values should make the sound more or less harsh.

You can find literally thousands of schematics online for gadgets to modify guitar sound. The circuit I've included here is one of the most primitive. If you want something more versatile, you should search for "stomp box schematics" and see what you can find.

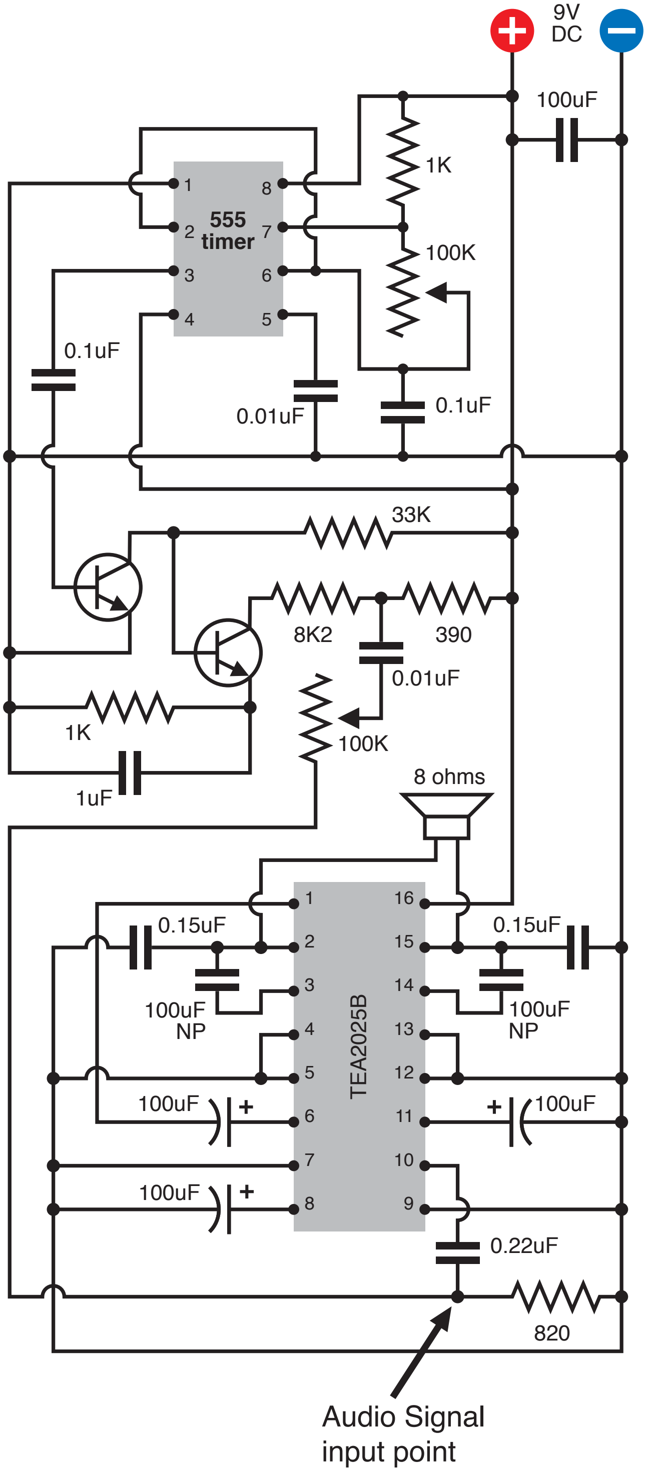
9V
DC
100uF
1K
100K
555
timer
1
2
3
4
8
7
6
5
0.1uF
0.01uF
0.1uF
33K
8K2
390
0.01uF
1K
1uF
100K
8 ohms
TEA2025B
0.15uF
100uF
NP
100uF
100uF
0.15uF
100uF
NP
100uF
0.22uF
820
Audio Signal
input point

Figure 5-55. *For a quick demo of clipping, insert a couple of transistors between the output from the 555 timer and the input of the amplifier chip. One transistor overdrives the other, so that when you adjust the potentiometer at the center of the circuit, you hear an increasingly harsh, distorted sound.*

BACKGROUND

Stomp-box origins

The Ventures recorded the first single to use a fuzz box, titled "The 2,000 Pound Bee," in 1962. Truly one of the most awful instrumentals ever made, it used distortion merely as a gimmick and must have discouraged other musicians from taking the concept seriously.

Ray Davies of the Kinks was the first to embody distortion as an integral part of his music. Davies did it initially by plugging the output from one amp into the input of another, supposedly during the recording of his hit "You Really Got Me." This overloaded the input and created clipping—the basic fuzz concept. From there it was a short step to Keith Richards using a Gibson Maestro Fuzz-Tone when the Rolling Stones recorded "(I Can't Get No) Satisfaction" in 1965.

Today, you can find thousands of advocates promoting as many different mythologies about "ideal" distortion. In Figure 5-56, I've included a schematic from Flavio Dellepiane, a circuit designer in Italy who gives away his work (with a little help from Google AdSense) at *http://www.redcircuits.com*. Dellepiane is self-taught, having gained much of his knowledge from electronics magazines such as the British *Wireless World*. In his fuzz circuit, he uses a very high-gain amplifier consisting of three field-effect transistors (FETs), which closely imitate the rounded square-wave typical of an overdriven tube amp.

Dellepiane offers dozens more schematics on his site, developed and tested with a dual-trace oscilloscope, low-distortion sinewave oscillator (so that he can give audio devices a "clean" input), distortion meter, and precision audio volt meter. This last item, and the oscillator, were built from his own designs, and he gives away their schematics, too. Thus his site provides one-stop shopping for home-audio electronics hobbyists in search of a self-administered education.

Before fuzz, there was tremolo. A lot of people confuse this with vibrato, so let's clarify that distinction right now:

- Vibrato applied to a note makes the frequency waver up and down, as if a guitarist is bending the strings.
- Tremolo applied to a note makes its volume fluctuate, as if someone is turning the volume control of a guitar up and down very quickly.

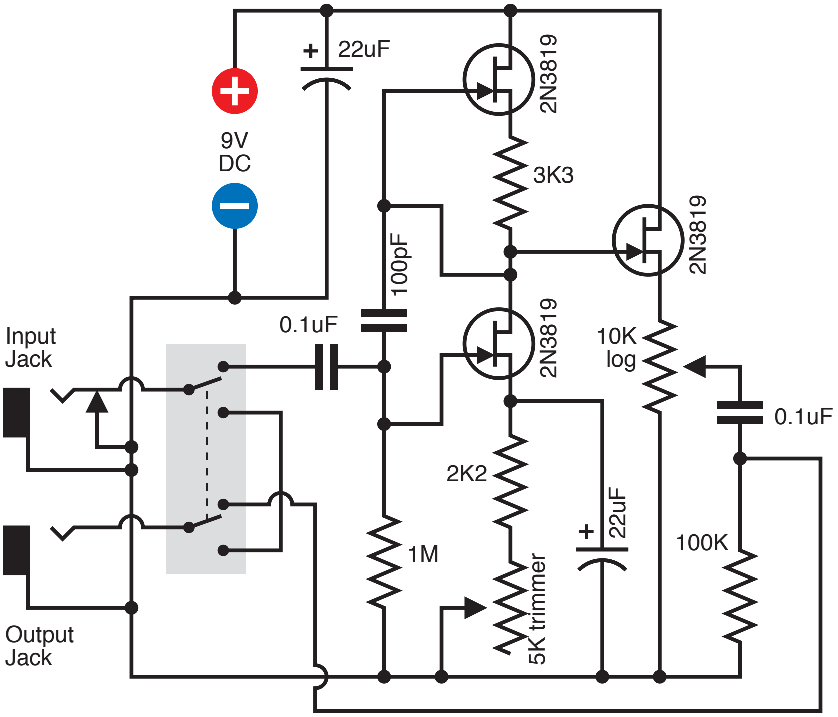

Figure 5-56. *This circuit designed by Flavio Dellepiane uses three transistors to simulate the kind of distortion that used to be created by overloading the input of a tube amplifier.*

BACKGROUND

Stomp-box origins (continued)

Harry DeArmond sold the first tremolo box, which he named the Trem-Trol. It looked like an antique portable radio, with two dials on the front and a carrying handle on top. Perhaps in an effort to cut costs, DeArmond didn't use any electronic components. His steam-punkish Trem-Trol contained a motor fitted with a tapered shaft, with a rubber wheel pressing against it. The speed of the wheel varied when you turned a knob to reposition the wheel up and down the shaft. The wheel, in turn, cranked a little capsule of "hydro-fluid," in which two wires were immersed, carrying the audio signal. As the capsule rocked to and fro, the fluid sloshed from side to side, and the resistance between the electrodes fluctuated. This modulated the audio output.

Today, Trem-Trols are an antique collectible. When industrial designer Dan Formosa acquired one, he put pictures online at *http://www.danformosa.com/dearmond.html*. And Johann Burkard has posted an MP3 of his DeArmond Trem-Trol so you can actually hear it: *http://johannburkard.de/blog/music/effects/DeArmond-Tremolo-Control-clip.html*.

The idea of a mechanical source for electronic sound mods didn't end there. The original Hammond organs derived their unique, rich sound from a set of toothed wheels turned by a motor. Each wheel created a fluctuating inductance in a sensor like the record head from a cassette player.

It's easy to think of other possibilities for motor-driven stomp boxes. Going back to tremolo: imagine a transparent disc masked with black paint, except for a circular stripe that tapers at each end. While the disc rotates, if you shine a bright LED through the transparent stripe toward a light-dependent resistor, you would have the basis for a tremolo device. You could even create never-before-heard tremolo effects by swapping discs with different stripe patterns. Figures 5-57 and 5-58 show the kind of thing I have in mind. For a real fabrication challenge, how about an automatic disc changer?

Today's guitarists can choose from a smorgasbord of effects, all of which can be home-built using plans available online. For reference, try these special-interest books:

- *Analog Man's Guide to Vintage Effects* by Tom Hughes (For Musicians Only Publishing, 2004). This is a guide to every vintage stomp box and pedal you can imagine.
- *How to Modify Effect Pedals for Guitar and Bass* by Brian Wampler (Custom Books Publishing, 2007). This is an extremely detailed guide for beginners with little or no prior knowledge. Currently it is available only by download, from sites such as *http://www.openlibrary.org*, but you may be able to find the previous printed edition from secondhand sellers, if you search for the title and the author.

Of course, you can always take a shortcut by laying down a couple hundred dollars for an off-the-shelf item such as a Boss ME-20, which uses digital processing to emulate distortion, metal, fuzz, chorus, phaser, flanger, tremolo, delay, reverb, and several more, all in one convenient multi-pedal package. Purists, of course, will claim that it "doesn't sound the same," but maybe that's not the point. Some of us simply can't get no satisfaction until we build our own stomp box and then tweak it, in search of a sound that doesn't come off-the-shelf and is wholly our own.

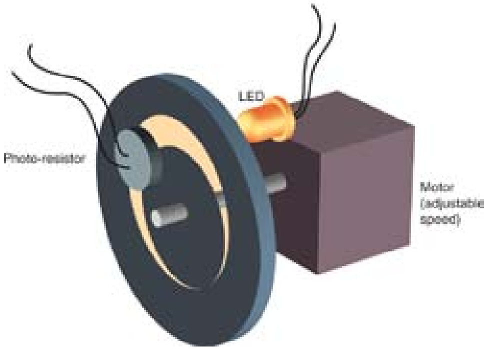

Figure 5-57. *Although electromechanical audio devices are obsolete now, some unexplored possibilities still exist. This design could create various tremolo effects, if anyone had the patience to build it.*

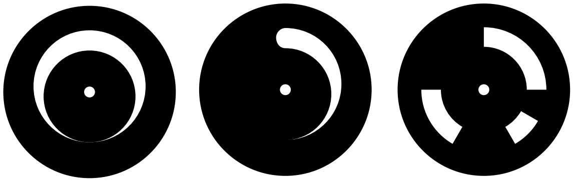

Figure 5-58. *Different stripe patterns could be used in conjunction with the imaginary electromechanical device in Figure 5-57 to create various tremolo effects.*

Experiment 31: One Radio, No Solder, No Power

Time now to go back one more time to inductance and capacitance, and demonstrate an application which also makes use of the way that waveforms can be added to each other. I want to show you how a simple circuit with no power supply at all can receive AM radio signals and make them audible. This is often known as a crystal radio, because the circuit includes a germanium diode, which has a crystal inside it. The idea dates back to the dawn of radio, but if you've never tried it, you've missed an experience that is truly magical.

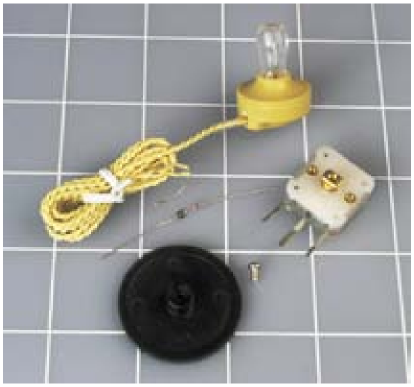

Figure 5-59. *Just add wire and a coil, and this is all you need to receive AM radio signals. The black disc becomes the tuning dial, after it is screwed onto the variable capacitor (right). This is actually an optional extra. A germanium diode (center) rectifies the radio signal. The high-impedance earphone (top) creates a barely audible sound.*

You will need:

- Rigid cylindrical object, such as a vitamin bottle. Quantity: 1.
- 22-gauge hookup wire, solid-core. Quantity: 60 feet.
- 16-gauge wire, stranded. Quantity: 100 feet.
- Polypropylene rope ("poly rope") or nylon rope. Quantity: 10 feet.
- Germanium diode. Quantity: 1.
- High-impedance headphone. Quantity: 1.

The diode and headphone can be ordered from *http://www.scitoyscatalog.com*. You cannot use a modern headphone of the type you wear with an MP3 player.

Some of these items are shown in Figure 5-59.

First, you need to make a coil. It should be about 3 inches in diameter, and you can wind it around any empty glass or plastic container of that size, so long as it's rigid. A soda bottle or water bottle isn't suitable, because the cumulative squeezing force of the turns of wire can deform the bottle so that it isn't circular anymore.

I chose a vitamin bottle that just happened to be exactly the right size. To remove the label, I softened its adhesive with a heat gun (lightly, to avoid melting the bottle) and then just peeled it off. The adhesive left a residue, which I removed with Xylol (also known as Xylene). This is a handy solvent to have around, as it can remove "permanent" marker stains as well as sticky residues, but you should always use latex or nitrile gloves to avoid getting it on your skin, and minimize your exposure the fumes. Because Xylol will dissolve some plastics, clearly it's not good for your lungs.

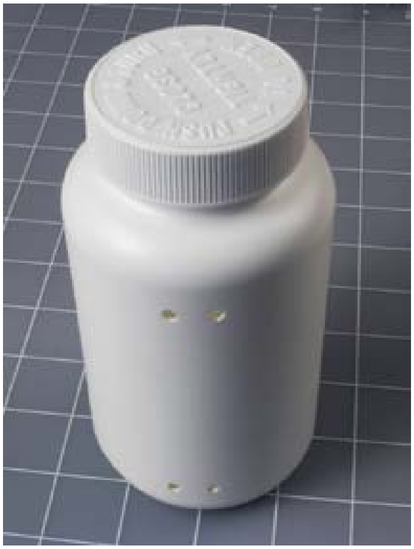

Figure 5-60. *A large, 3-inch diameter empty vitamin bottle makes an ideal core for a crystal radio coil. The drilled holes will anchor wire wrapped around the bottle.*

After you prepare a clean, rigid bottle, drill two pairs of holes in it, as shown in Figure 5-60. You'll use them to anchor the ends of the coil.

Now you need about 60 feet of 22-gauge solid-core wire. If you use magnet wire, its thin insulation will allow the turns of the coil to be more closely spaced, and the coil may be slightly more efficient. But everyday vinyl-insulated wire will do the job, and is much easier to work with.

Begin by stripping the first 6 inches of insulation from the end of the wire. Now measure 50 inches along the insulated remainder and apply your wire strippers at that point, just enough to cut the insulation without cutting the wire. Use your two thumb nails to pull the insulation apart, revealing about a half-inch of bare wire, as shown in Figure 5-61. Bend it at the center point and twist it into a loop, as shown in Figure 5-62.

You just made a "tap," meaning a point where you will be able to tap into the coil after you wind it. You'll need another 11 of these taps, all of them spaced 50 inches apart. (If the diameter of the bottle that you'll be using as the core of your coil is not 3 inches, multiply its diameter by 16 to get the approximate desired spacing of taps.)

After you have made 12 taps, cut the wire and strip 6 inches off that end. Now bend the end into a U shape about a half-inch in diameter, so that you can hook it through the pair of holes that you drilled at one end of the bottle. Pull the wire through, then loop it around again to make a secure anchor point.

Now wind the rest of the wire around the bottle, pulling it tightly so that the coils stay close together. When you get to the end of the wire, thread it through the remaining pair of holes to anchor it as shown in Figure 5-63. The completed coil is shown in Figure 5-64.

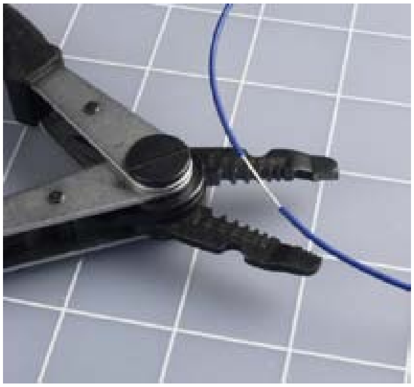

Figure 5-61. *Wire strippers expose the solid conductor at intervals along a 22-gauge wire.*

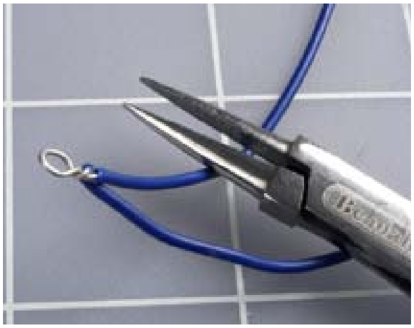

Figure 5-62. *Each exposed section of wire is twisted into a loop using sharp-nosed pliers.*

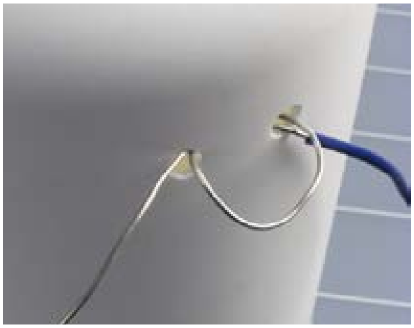

Figure 5-63. *The stripped end of the wire is secured through the holes drilled in the bottle.*

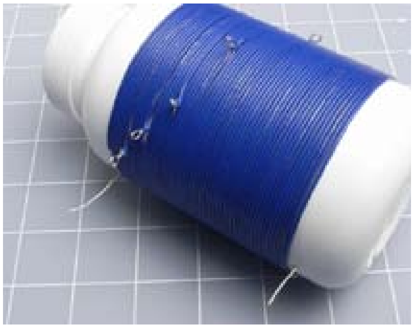

Figure 5-64. *The completed coil, wrapped tightly around the bottle.*

Your next step is to set up an antenna. If you live in a house with a yard outside, this is easy: just open a window, toss out a reel of 16-gauge wire while holding the free end, then go outside and string up your antenna by using polypropylene rope ("poly rope") or nylon rope, available from any hardware store, to hang the wire from any available trees, gutters, or poles. The total length of the wire should be about 100 feet. Where it comes in through the window, suspend it on another length of poly rope. The idea is to keep your antenna wire as far away from the ground or from any grounded objects as possible.

If you live in an apartment where you don't have access to a yard outside, you can try stringing your antenna around the room, hanging the wire from more pieces of poly rope. The antenna should still be about 100 feet long, but obviously it won't be in a straight line.

Hook the free end of your antenna to one end of your coil. At this point, you also need to add a germanium diode, which functions like a silicon-based diode but is better suited to the tiny voltages and currents that we'll be dealing with. The other end of the diode attaches to one of the wires leading to a high-impedance

High Voltage!

The world around us is full of electricity. Normally we're unaware of it, but a thunderstorm is a sudden reminder that there's a huge electrical potential between the ground below and the clouds above.

If you put up an outdoor antenna, never use it if there is any chance of a lightning strike. This can be extremely dangerous. Disconnect the indoor end of your antenna, drag it outside, and push the end of the wire into the ground to make it safe.

earphone. *A normal modern earphone or headphone will not work in this circuit.* The return wire from the headphone is connected to a jumper wire, the other end of which can be clipped to any of the taps in your coil.

One last modification, and you'll be ready to tune in. You have to ground the jumper wire. By this I mean connect it to something that literally goes into the ground. A cold-water pipe is the most commonly mentioned option, but (duh!) this will work only if the pipe is made of metal. Because a lot of plumbing these days is plastic, check under the sink to see if you have copper pipes before you try using a faucet for your ground.

Another option is to attach the wire to the screw in the cover plate of an electrical outlet, as the electrical system in your house is ultimately grounded. But the sure-fire way to get a good ground connection is to go outside and hammer a 4-foot copper-clad grounding stake into reasonably moist earth. Any wholesale electrical supply house should be able to sell you a stake. They're commonly used to ground welding equipment.

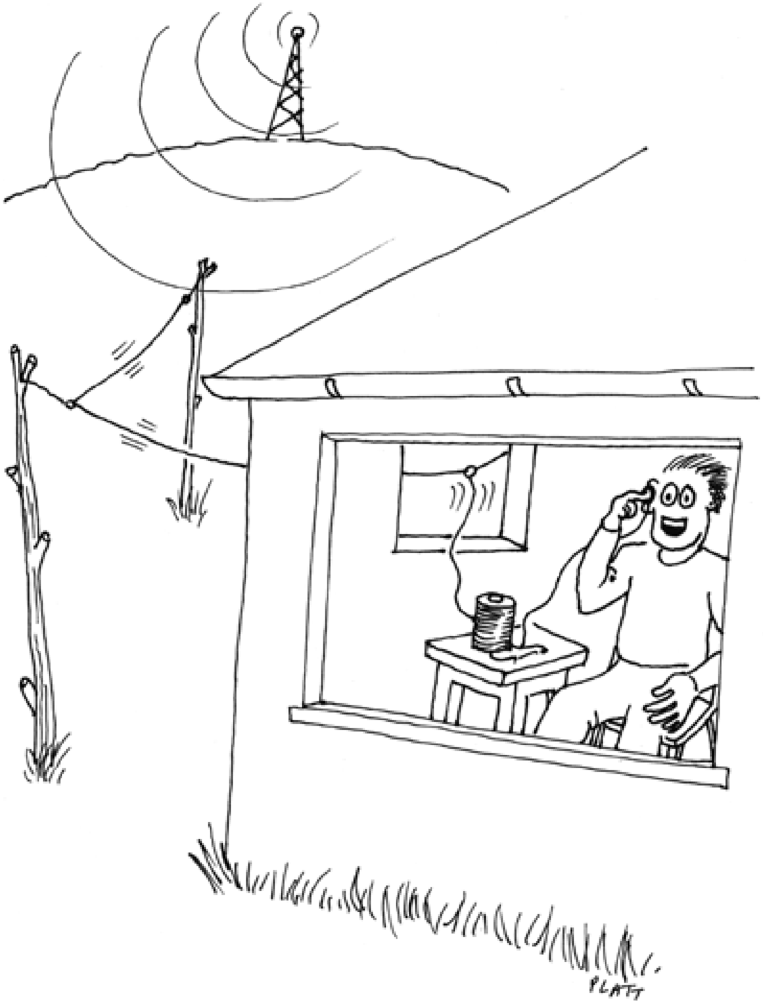

Figure 5-85. *The simple pleasure of picking up a radio signal with ultra-simple components and no additional power.*

Figures 5-66 and 5-67 show the completed radio.

If you've managed to follow these instructions (one way or another), it's time to tune your radio to the nearest station. Move the alligator clip at the end of your patch cord from one tap to another on your coil. Depending on where you live, you may pick up just one station, or several, some of them playing simultaneously.

It may seem that you're getting something for nothing here, as the earphone is making noise without any source of power. Really, though, there is a source of power: the transmitter located at the radio station. A large amplifier pumps power into the broadcasting tower, modulating a fixed frequency. When the combination of your coil and antenna resonates with that frequency, you're sucking in just enough voltage and current to energize a high-impedance headphone.

The reason you had to make a good ground connection is that the radio station broadcasts its signal at a voltage relative to ground. The earth completes the circuit between you and the transmitter. For more information on this and other concepts relating to radio, see the upcoming section "Theory: How radio works."

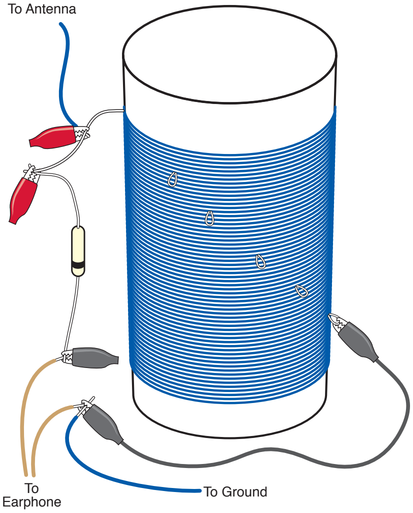

Figure 5-66. *A signal from the antenna can pass through the coil to ground. If the jumper wire is attached to an appropriate tap on the coil, it resonates with the radio signal, just powerfully enough to energize the earphone which is wired in series with a diode.*

Enhancements

The higher your antenna is, the better it should work. In my location, this is a major problem, as I live in a desert area without any trees. Still, just stringing the wire out of my window and tethering it (with rope) to the front bumper of my car enabled me to pick up a faint radio signal.

To improve the selectivity of your radio, you can add a variable capacitor, as shown in the following section. This allows you to "tune" the resonance of your circuit more precisely. Variable capacitors are uncommon today, but you can find one at the same specialty source that I recommended for the earphone and the germanium diode: the Scitoys Catalog (*http://www.scitoyscatalog.com*).

This source is affiliated with a smart man named Simon Quellan Field, whose site suggests many fun projects that you can pursue at home. One of his clever ideas is to remove the germanium diode from your radio circuit and substitute a low-power LED in series with a 1.5-volt battery. This didn't work for me, because I live 40 miles from the nearest AM broadcaster. If you're closer to a transmitter, you may be able to see the LED varying in intensity as the broadcast power runs through it.

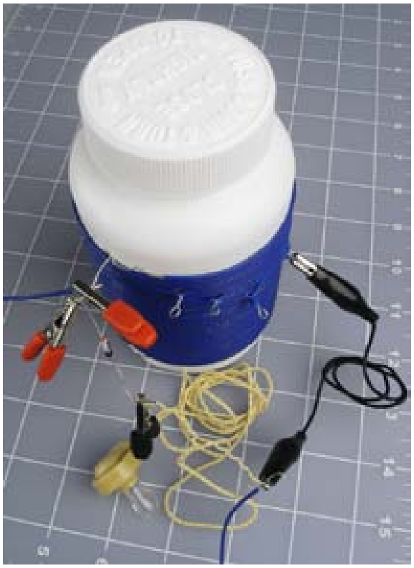

Figure 5-67. *The real-life version of Figure 5-66.*

THEORY

How radio works

When electrical frequencies are very high, the radiation they create has enough energy to travel for miles. This is the principle of radio transmission: A high-frequency voltage is applied to a broadcasting antenna, relative to the ground.

When I say "ground" in this instance, I literally mean the planet beneath us. If you set up a receiving antenna, it can pick up a faint trace of the transmission relative to the ground—as if the earth is one huge conductor. Actually the earth is so large and contains so many electrons, it can function as a common sink, like a gigantic version of the file cabinet that I suggested you should touch to get rid of static electricity in your body before touching a CMOS logic chip.

To make a radio transmitter, I could use a 555 timer chip running at, say, 850 kHz (850,000 cycles per second), and pass this stream of pulses through an amplifier to a transmission tower; if you had some way to block out all the other electromagnetic activity in the air, you could detect my signal and reamplify it.

This was more or less what Marconi (shown in Figure 5-68) was doing in 1901, after he had purchased rights to Edison's wireless telegraphy patent, although he had to use a primitive spark gap, rather than a 555 timer, to create the oscillations. His transmissions were of limited use, because they had only two states: on or off. You could send Morse code messages, and that was all.

Figure 5-68. *Marconi, the great pioneer of radio (photograph from Wikimedia Commons).*

Five years later, the first true audio signal was transmitted by imposing lower audio frequencies on the high-frequency carrier wave. In other words, the audio signal was "added" to the carrier frequency, so that the power of the carrier varied with the peaks and valleys of the audio.

At the receiving end, a very simple combination of a capacitor and a coil detected the carrier frequency out of all the other noise in the electromagnetic spectrum. The values of the capacitor and the coil were chosen so that their circuit would "resonate" at the same frequency as the carrier wave. Figures 5-69 and 5-70 illustrate these concepts.

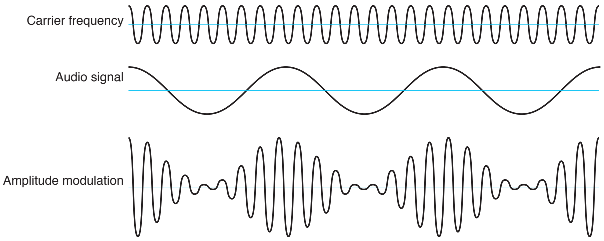

Figure 5-69. *When an audio signal (middle) is combined electronically with a high carrer frequency (top), the result looks something like the compound signal at the bottom. In actuality, the carrier frequency would be much higher compared with the audio frequency, by a radio of perhaps 1,000:1.*

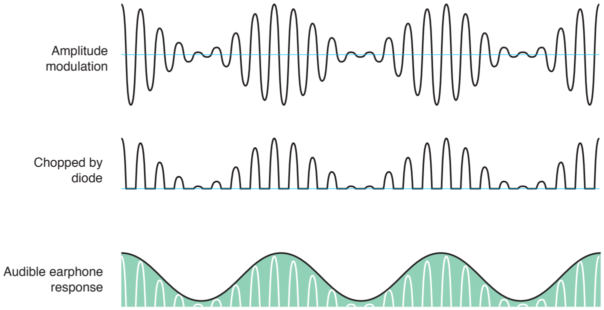

Figure 5-70. *When the compound signal is passed through a diode, only the upper half remains. An earphone cannot react fast enough to reproduce the high carrier frequency, so it "rides" the peaks and thus reproduces the audio frequency.*

The schematic in Figure 5-71 shows the simple circuit that you built by wrapping a coil around an empty vitamin bottle. When a positive pulse was received by the antenna, it resonated with the antenna and the coil, provided that the antenna was long enough and the coil was tapped at the appropriate number of turns.

THEORY

How radio works (continued)

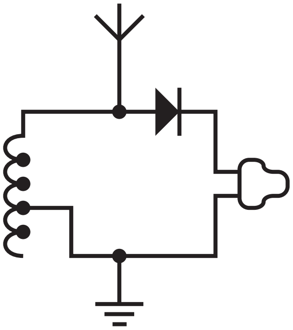

Figure 5-71. *An antenna at the top of the schematic picks up faint electromagnetic radiation from a distant transmitter. The coil at the left side is tapped at intervals so that its resonance can be adjusted to match the carrier frequency of the radio signal. Other frequencies are grounded (at the bottom of the schematic). The diode passes the "top half" of the signal to the earphone at the right, which is incapable of responding fast enough to reproduce the carrier frequency, and thus filters it out, leaving only the audio frequencies that were superimposed on it.*

By adding a capacitor, you can tune the circuit. Now an incoming pulse from the transmitter is initially blocked by the self-inductance of the coil, while it charges the capacitor. If an equally negative pulse is received after an interval that is properly synchronized with the values of the coil and the capacitor, it coincides with the capacitor discharging and the coil conducting. In this way, the right frequency of carrier wave makes the circuit resonate in sympathy. At the same time, audio-frequency fluctuations in the strength of the signal are translated into fluctuations in voltage in the circuit.

What happens to other frequencies pulled in by the antenna? The lower ones pass through the coil to ground; the higher ones pass through the capacitor to ground. They are just "thrown away."

The righthand half of the circuit samples the signal by passing it through a germanium diode and an earphone. The power from the transmitter is just sufficient to vibrate the diaphragm in the earphone, after the diode has subtracted the negative half of the signal.

Look back at the diagram of the amplitude-modulated signal. You'll see that it fluctuates up and down so rapidly, the earphone cannot possibly keep up with the positive-negative variations—hence the need for the diode. It will remain hesitating at the midpoint between the highs and lows, producing no sound at all. The diode solves this problem by subtracting the lower half of the "audio envelope," leaving just the positive spikes of voltage. Although these are still very small and rapid, they are now all pushing the diaphragm of the earphone in the same direction, so that it averages them out, approximately reconstructing the original sound wave.

Figure 5-72 shows how the circuit can be enhanced with a variable capacitor, to tune it without needing to tap the coil at intervals.

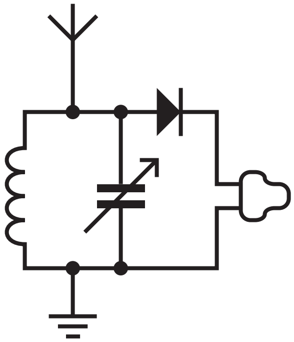

Figure 5-72. *By adding a capacitor to the circuit, its resonance can be tuned more precisely. The diagonal arrow indicates that a variable capacitor is used.*

The radio can pull in the stations on the AM (amplitude-modulated) waveband that happen to be most powerful in your area. The waveband ranges from 300 kHz to 3 MHz. If you find yourself interested in radio, your next step could be to build a powered radio using a couple of transistors. Alternatively you could build your own (legal) low-power AM transmitter. There's an ultra-simple kit available from *http://www.scitoys.com* consisting of just two principal components: a crystal oscillator, and a transformer, shown in Figure 5-73. That's all it takes.

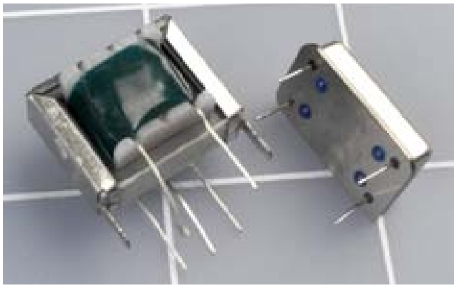

Figure 5-73. *An AM radio transmitter can be made from just two components: a transformer (left) and a crystal oscillator (right), available from http://www.scitoys.com.*

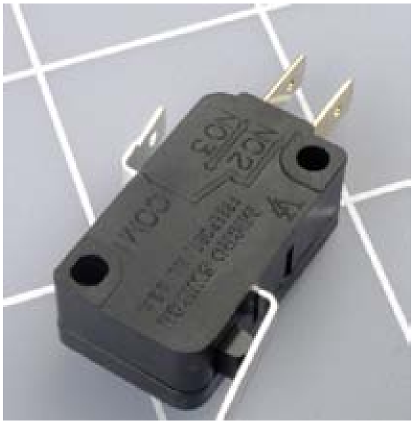

Figure 5-74. *A microswitch has a small button (at the front, righthand side in this picture) that is often actuated by a pivoted metal lever. The switch can respond to a very light pressure, but can handle relatively high currents.*

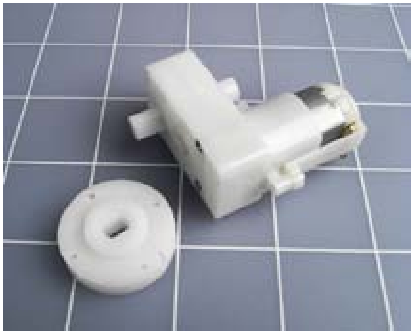

Figure 5-75. *For the Little Robot Cart, I found this 5-volt motor, which is supplied with a disc that fits its output shaft. The combination cost less than $10.*

Experiment 32: A Little Robot Cart

Robotics is another application of electronics that deserves a book in itself—or several books. So, once again, I'm going to give you an introduction followed by some points that you can follow if you want to go further. As always, I will start with the simplest possible device, which in the world of robotics is a cart that finds its way around your living room.

You will need:

- SPST or SPDT microswitches requiring minimal pressure to activate them. A force between 0.02 and 0.1 newtons would be ideal. Quantity: 2. See Figure 5-74.
- DC gear-motor, rated for any voltage between 5 and 12, drawing a maximum of 100mA in its free-running state, output shaft turning between 30 and 60 RPM. Quantity: 1. A motor is shown in Figure 5-75.
- Disc or arm that fits securely onto your motor shaft. Quantity: 1.
- 555 timer. Quantity: 1.
- DPDT nonlatching relay rated for the same voltage as your motor. Quantity: 1.
- 1/4-inch plywood or plastic, one piece about 2 feet square.
- #4 sheet-metal screws, 5/8 inch or 3/4 inch long. Quantity: 2 dozen.
- #6 bolts, 3/4 inch long, with nylon-insert lock nuts. Quantity: 2 dozen.
- 1/4-inch bolts, 1 inch long, with nuts, to mount the wheels. Quantity: 4.

I'm not specifying one particular motor, because if I did, it might not be available by the time you read this. Motors aren't like logic chips, which have retained their basic function throughout various improvements over a period of several decades. Motors come and go, and many that you may run across will be surplus parts that will never been seen again. Search online for "gearmotor" or "gearhead motor" and find one as close as possible to the specification that I have provided. The mechanical power output of the motor shouldn't be important, because we won't be requiring it to do much work.

The important consideration when you buy your motor is that you should also obtain something that fits onto its output shaft. Typically, this will be a disk or arm that can be screwed into place. To this you can then add a larger wheel of your own, which you can cut with a hole saw or make from the screw-on lid of a jar, or anything else circular that you may find in the house.

A larger wheel will make your cart move faster than a smaller wheel, but will reduce its torque, thus limiting its power to overcome obstacles.

This brings me to my next topic: fabrication. Although this is an electronics book, motors are electromechanical devices, and you have to be able to install them in some kind of a machine to get any interesting results. You can use plywood to complete the two little robotics projects here (ideally, the kind of thin, high-quality plywood sold by hobby stores) but I recommend something that looks better and is easier to work with: ABS plastic. Before you start on the robotic cart, you may want to check the section "Fundamentals: All about ABS."

FUNDAMENTALS

All about ABS

Unless you think the steampunk movement isn't going back far enough, you probably don't want your autonomous robot cart to resemble a relic from before the 1800s. Therefore, wood may not be the best construction material. Metal can look nice, but is not easy to work with. For quick results that have a twentieth-century look (maybe even a 21st-century look), plastic is the obvious choice, and I feel that ABS is the best type of plastic to use, because it provides such quick, easy results. ABS stands for "acrylonitrile butadiene styrene." Lego® blocks are made of ABS. Car-stereo installers and model-railroad buffs use it. You can use it, too. You can saw it, drill it, sand it, whittle it, and drive screws into it, and it won't warp, split, or splinter. It's washable, doesn't need to be painted, and will last almost forever.

Delrin is another type of plastic, but tends to cost more and is a little tougher to drill and cut. It's a matter of personal preference. ABS machines fairly well, but when you drill it, for example, it can "catch" on the bit and the piece will spin with the bit due to the way that plastic chips off with the bit. Delrin is self-lubricating and has better melting properties under the heat of machining, so it drills and cuts much more cleanly and easily than ABS.

Where to find ABS

Pieces of ABS a couple of feet square are available from online sources such as *http://hobbylinc.com* or estreetplastics (an eBay store), but you'll save money if you can truck on down to your nearest plastic supply house and buy it like plywood, in sheets measuring 4 by 8 feet. To discover whether you have a nearby plastic supply house, search for "plastic supply" in your yellow pages or Google Local.

Piedmont Regal Plastics has many supply centers around the nation, but you'll have to collect it yourself, and they may not be willing to cut small pieces. You can check online at *http://www.piedmontplastics.com* for their locations.

Stock colors of ABS include black, white, and "natural," which is beige. Sheets usually are textured on one side, which is the side that should face outward, as it is more scratch-resistant than the smooth side.

Because you won't be adding paint or other finishes, you'll have to be careful not to scuff the plastic or scratch it while working. Clean your bench thoroughly before you begin, taking special care to remove any metal particles, which tend to become embedded in the plastic. Use wooden shims in the jaws of your vise, and avoid resting the plastic accidentally on any sharp tools or screws. Working with ABS requires a clean environment and a very gentle touch.

Cut with Care

You can saw ABS, but if you use a table saw, the plastic will tend to melt and stick to the blade. These smears will get warm and sticky when you feed the next piece of plastic into the saw, and the result will be extremely unpleasant. The whirling blade will grab the plastic and hurl it at you powerfully enough to break bones. This is known as "kickback" and is a very serious risk when sawing plastic.

If you have extensive experience using a table saw, you are actually more vulnerable, because the reflexes and cautions you have developed while dealing with wood will not be adequate for working with plastic. Please take this warning seriously!

Your first and most obvious precaution is to use a plastic-cutting blade, which has a larger number of thicker teeth to absorb the heat. The blade I have used is a Freud 80T, but there are others. If you use a blade that is not suitable, you will see it starting to accumulate sticky smears. This is the only warning you will get. Clean that blade with a solvent such as acetone, and never use it for ABS again.

Regardless of other precautions, always wear gloves and eye protection when using a table saw, and stand to one side when feeding materials into it. Personally, after one episode of kickback that I thought had broken my arm, I prefer not to use a table saw on plastic at all.

For long, straight cuts, the alternatives include:

- *Panel saw (big and expensive, but safe and accurate).*
- *Miniature handheld circular saw with a blade around 4 inches in diameter, guided with a straight edge clamped to the sheet.*
- *Hand saw. This is my old-school preference. My favorite is a Japanese pull-to-cut saw, which makes very clean cuts: the Vaughan Extra-Fine Cross-Cut Bear Saw, 9-1/2 inches, 17 tpi (teeth per inch). If you use one of these, be careful to keep your free hand out of the way, as the saw can easily jump out of the cut. Because it is designed to cut hard materials such as wood, it has no difficulty cutting soft flesh. Gloves are strongly recommended.*

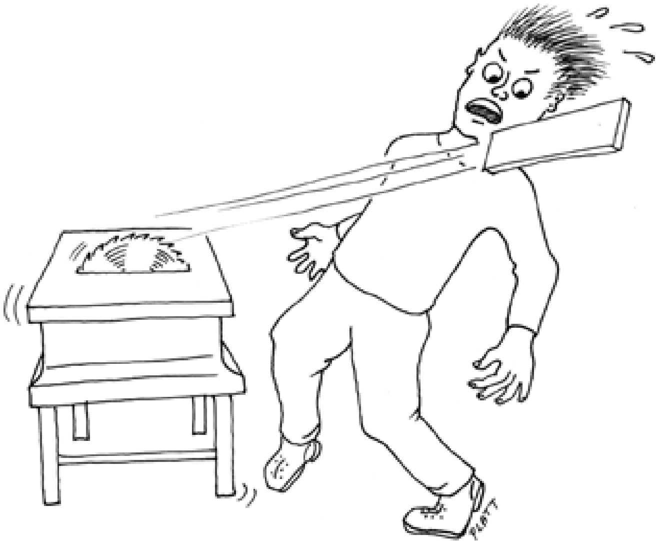

Figure 5-76. *The perils of kickback. Plastic easily sticks to the blade of a table saw, which will hurl it at you unexpectedly. Use other tools to cut plastic.*

Curving Cuts

Curving cuts involve relatively little danger, although eye protection and gloves are still advisable. My preferred tools:

- Band saw with a 3/8- or 1/4-inch blade designed for thin wood or plywood.
- Jigsaw. I have a special liking for the DeWalt XRP using Bosch blades that are designed for hardwood or plastic. This will cut complex curves in ABS as easily as scissors cutting paper.

No matter what type of saw you use, you'll have to clean ragged bits of plastic off the cut afterward, and the absolutely necessary item for this purpose is a deburring tool, available from *http://www.mcmaster.com* and most other online hardware sources. A belt sander or disc sander is ideal for rounding corners, and a metal file can be used to remove bumps from edges that are supposed to be straight.

Figures 5-77 through 5-80 show various cutting tools. Figure 5-81 shows a deburring tool, and Figure 5-82 shows a disc sander.

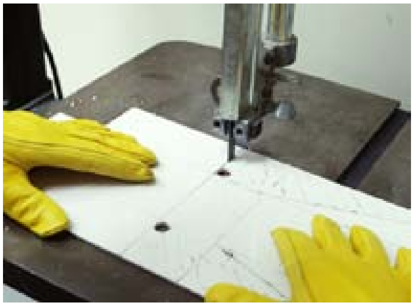

Figure 5-77. *A band saw is an ideal tool for cutting complex shapes out of ABS plastic. You can often find them secondhand for under $200.*

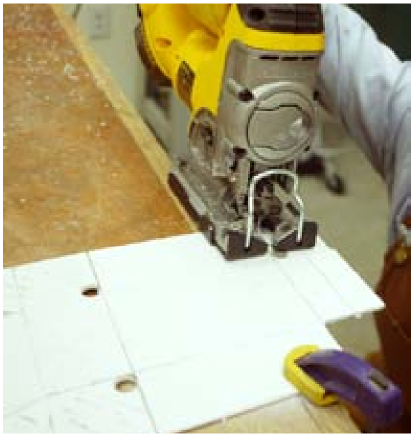

Figure 5-80. *This DeWalt jigsaw can run at very slow speeds, enabling precise and careful work with plastic.*

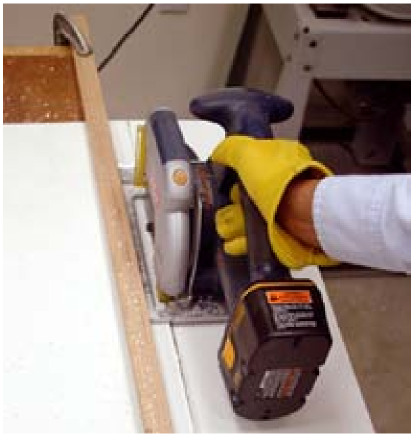

Figure 5-78. *A handheld circular saw, running along a straight-edge, is much safer than a table saw for cutting plastic, and can produce comparable results.*

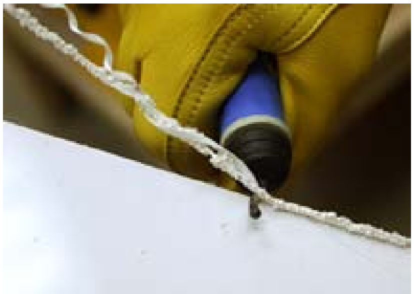

Figure 5-81. *A deburring tool will clean and bevel the sawn edge of a piece of plastic in just a couple of quick strokes.*

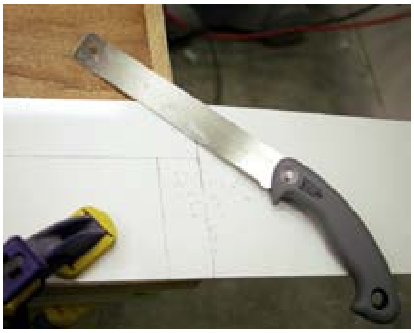

Figure 5-79. *This Japanese-style saw cuts when you pull it, rather than when you push it. After some practice, you can use it to make very accurate cuts. Because ABS is so soft, minimal muscle-power is required.*

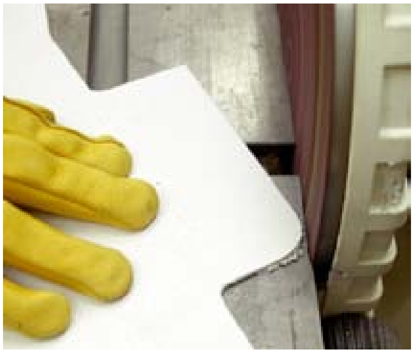

Figure 5-82. *A belt sander or disc sander is the ideal tool for rounding corners when working with ABS plastic.*

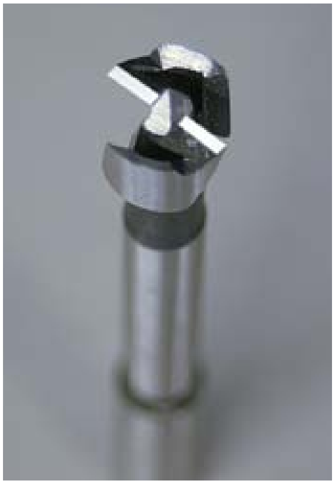

Figure 5-83. *A Forstner drill bit makes clean, precise holes; a large regular drill bit will chew up ABS plastic and make a mess.*

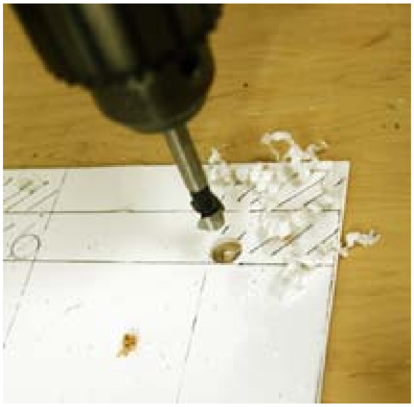

Figure 5-84. *By drilling holes at any location where two bends intersect, you reduce the risk of the plastic fissuring.*

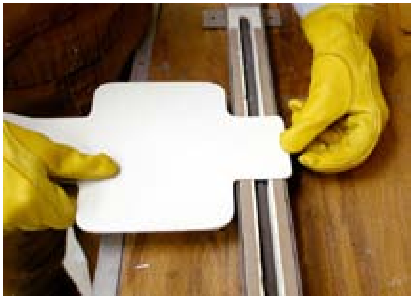

Figure 5-85. *Making clean, precise bends in ABS is simply a matter of resting the plastic over a bender that consists of an electric heating element.*

Making Plans

I like to use drawing software to create plans, and I try to print them at actual scale. I tape them to the smooth side of a piece of white or natural-color ABS, then use an awl to prick through the plan into the soft surface beneath. I remove the paper and connect the awl marks by drawing onto the plastic using a pencil or a fine-point water-soluble pen. Its lines can be wiped away later with a damp cloth. Don't use a permanent marker, as the solvents that you will need to clean it may dissolve the plastic.

Because ABS will tend to open a fissure when you bend it at any inside corner where you don't have a smooth radius, you need to drill holes at these locations, as shown in the cart plans in Figure 5-92 on page 275.

A regular half-inch drill bit is too aggressive; it will tend to jam itself into the plastic within one turn of the drill. Use Forstner bits (shown in Figures 5-83 and 5-84) to cut nice smooth circles.

Note that the heat from bending will tend to make any marks on the plastic permanent.

Bending It

A big advantage of plastic over wood is that you can make complex shapes by bending them, instead of cutting separate pieces and joining them with nails, screws, or glue. Unfortunately, bending does require an appropriate bender: an electric heating element mounted in a long, thin metal enclosure that you place on your workbench. The bender I use is made by FTM, a company that offers all kinds of neat gadgets for working with plastic. Their cheapest bender, shown in Figure 5-85, is just over $200 with a 2-foot element. You can get a 4-foot model for about $50 more. Check them out at *http://thefabricatorssource.com*.

Avoid Burns While Bending

A plastic bender will inflict serious burns if you happen to rest your hand on it accidentally, and because it has no warning light, you can easily forget that you have left it plugged in. Use gloves!

To bend plastic, lay it over the hot element of a plastic bender for a brief time (25 to 30 seconds for 1/8-inch ABS, 40 to 45 seconds for 3/16-inch, and up to a minute for 1/4-inch). If you overheat the plastic, you'll smell it, and when you turn it over you'll find it looks like brown melted cheese. Naturally you should learn to intervene before the plastic reaches that point.

ABS is ready to bend when it yields to gentle pressure. Take it off the bender and bend it *away* from the side that you heated. If you bend it toward the hot side, the softened plastic will bunch up inside the bend, which doesn't look nice.

You can work with it for about half a minute, and when you have it the way you want it, spray or sponge water onto it to make it set quickly. Alternatively, if you need more time, you can reheat it. The amount of force necessary to bend the sheet increases in proportion with the length of the bend, so a long bend can be difficult, and I usually insert it into a loose vise, push it a bit, move it along to the next spot, and push it again.

Because plastic bending is very similar to making shapes in origami, it's a good idea to model your projects in paper before you commit yourself to ABS.

If you decide that you don't want to spend money on a bender, don't abandon plastic just yet—you can use screws to assemble separate sections with greater ease and convenience than if you were working with wood.

Making 90-Degree Joints

Driving screws into the edge of a piece of plywood will almost always separate its layers, but ABS has no layers (or grain, either), and never splits or shatters. This means that you can easily join two pieces at 90° using small screws (#4 size, 5/8-inch long).

Figures 5-86 through 5-90 show the procedure for joining 1/8-inch (or thicker) ABS to 1/4-inch ABS, which I regard as the minimum thickness when you're inserting screws into its edge:

1. Mark a guideline on the thinner piece of plastic, 1/8 inch from its edge. For #4 screws, drill holes using a 7/64-inch bit. If you're using flat-headed screws, countersink the holes very gently.

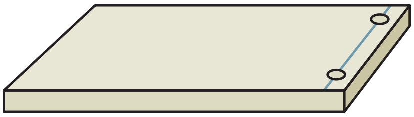

Figure 5-86

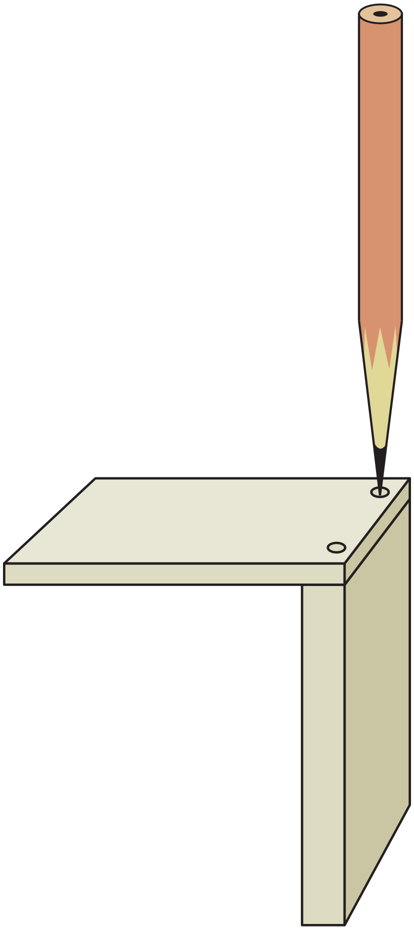

Figure 5-87

2. Hold or clamp the pieces in place and poke a pen or pencil through the holes to mark the edge of the 1/4-inch plastic beneath.
3. Remove the thin plastic, clamp the 1/4-inch plastic in a vise, and drill guide holes for the screws at each mark, centered within the thickness of the plastic. Because ABS does not compress like wood, the holes must be larger than you may expect; otherwise, the plastic will swell around the screw. A 3/32-inch bit is just right for a #4 screw.
4. Assemble the parts. Be careful not to overtighten the screws; it's easy to strip the threads that they cut in the soft plastic.

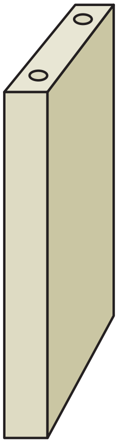

Figure 5-88

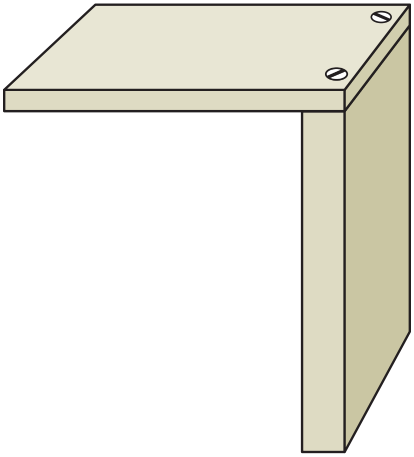

Figure 5-89. *Figures 5-86 through 5-89 illustrate four steps to join two pieces of ABS using #4 sheet-metal screws. Cut 7/64-inch holes on a line 1/8 inch from the edge of the first piece, then mark through the holes to the edge of the second piece. Drill 3/32-inch holes that are precisely centered in the edge, then screw the pieces together.*

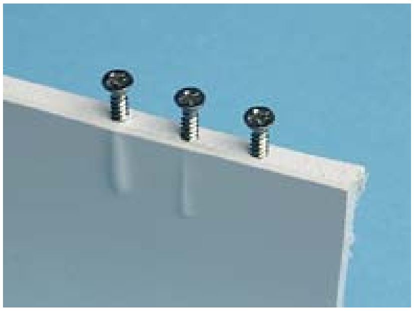

Figure 5-90. *Three #4 screws driven into the edge of ABS, using a 1/16-inch guide hole, a 5/64-inch guide hole, and a 3/32-inch guide hole. respectively Because the first two guides holes were too small, the plastic swelled around the screw (but did not break).*

Framing Your Cart

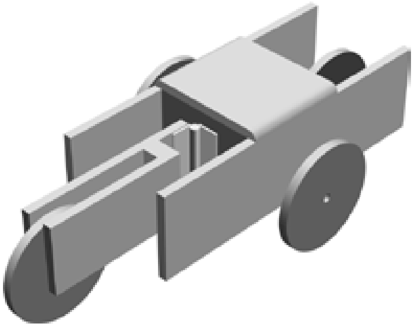

Figure 5-91. *If you have 3D rendering software, it can be a great way to test the feasibility of a construction project before you start cutting materials and trying to fit pieces together. This rendering was a proof-of-concept for the Little Robot Cart.*

For reasons that will soon be apparent, I've chosen an unusual diamond-shaped configuration of wheels. In the rendering shown in Figure 5-91, the front wheel (at the far end of the cart) applies power, the rear wheel (at the near end of the cart) steers the cart when backing up, and the side wheels prevent it from falling over.

Depending on the type of motor that you buy, you'll have to improvise a way to mount it in the front section of the cart. Don't be afraid to use kludges such as cable ties, duct tape, or even rubber bands to attach the motor to the frame. We're making a rough prototype, here, not a thing of beauty (although if you decide you like the cart, you can always rebuild it beautifully later).

The plan in Figure 5-92 shows the pieces that you will need. Part A is the body of the cart. If you're going to bend it from ABS, you should drill half-inch holes, with a forstner bit, at the four inside corners, so that these corners have rounded edges. If you simply saw the plastic to make sharp 90° corners, the plastic may develop fissures at the corners when you bend it. If you don't have a plastic bender and don't feel inclined to buy one, you can make Part A from three separate rectangles and then screw them together.

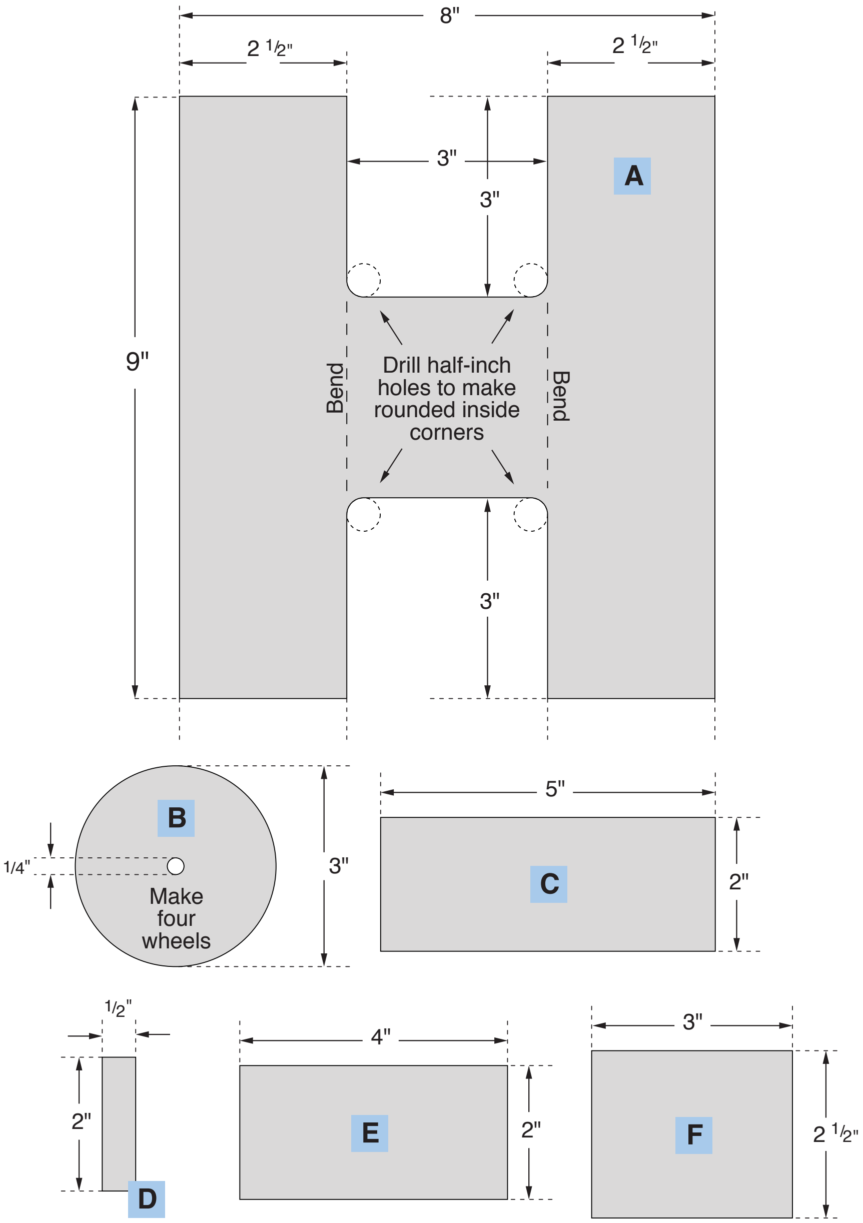
8"
2 1/2"
2 1/2"
3"
3"
A
9"
Bend
Drill half-inch holes to make rounded inside corners
Bend
3"
B
1/4"
Make four wheels
3"
5"
C
2"
1/2"
2"
D
4"
E
2"
3"
F
2 1/2"

Figure 5-92. *These sections of 1/4-inch plastic can be assembled to create the simple cart described in Experiment 31.*

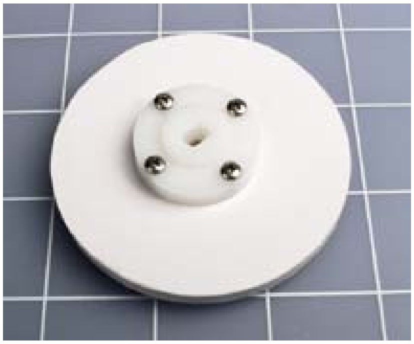

Figure 5-93. *A 3-inch wheel is screwed to the disc that mates with the drive shaft of the motor.*

Part B is a wheel, of which you will need four. I cut them using a 3-inch hole saw. The front wheel is screwed to whatever disc or arm you obtained to mate with the shaft of your motor. See Figure 5-93.

Parts C, D, and E assemble to form a yoke in which the rear wheel is mounted. I used a 2-inch hinge to pivot the yoke. The hinge is mounted on Part F, which is a partition located midway in the frame of the cart. The photographs in Figures

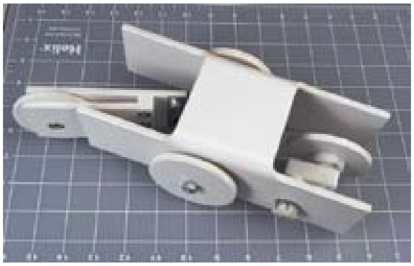

Figure 5-94. *The assembled body of the cart, before adding control electronics. The wheel at the righthand end will pull the cart from left to right. The hinged trailing wheel will allow the cart to move in a relatively straight line when it moves forward, but will tend to turn it when it backs up.*

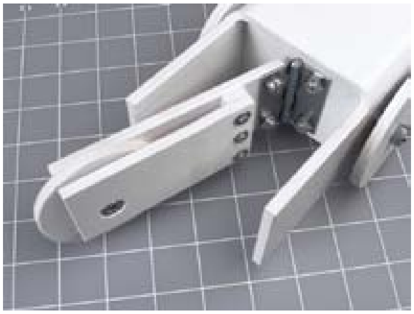

Figure 5-95. *A closeup of the hinged trailing wheel, which rotates freely and can flip from side to side with minimal friction.*

5-94 and 5-95 will help to make this clear. Initially, when you install Part F, use only two screws, one each side, so that you can adjust its angle a little. This will be necessary to optimize the contact of the wheels with the floor.

The side wheels and rear wheel must spin freely, but on the other hand, they shouldn't wobble. I simply tightened the nuts on the bolts that serve as axles for the wheels, until there was maybe half a millimeter of clearance. I added a drop of Loctite to stop the nuts from getting loose.

The plans don't show precisely where to drill holes for the axle bolts, because the location will depend on the size of your wheels. You can figure this out as you go along. Just make sure that the side wheels aren't mounted too low. We don't want them to lift the front wheel or the rear wheel off the floor. If the side wheels are a fraction higher off the ground than the front and rear wheels, that's good.

If you have tile or wood floors, your cart may acquire better traction if you wrap a thick rubber band around each disc that you use for the drive wheel and the steering wheel.

The most important aspect of the construction is to place microswitches where they'll be triggered when the cart runs into something. I placed mine at the front corners, as shown in Figures 5-96 and 5-97. And that brings me to the electronics.

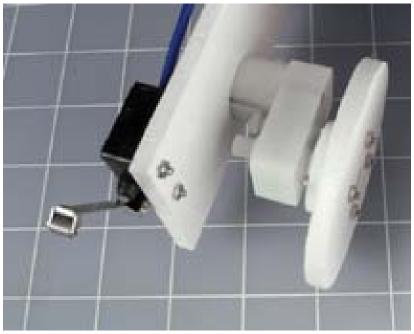

Figure 5-96

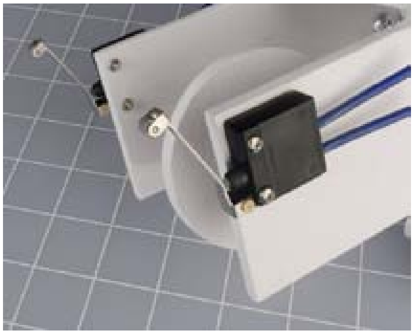

Figure 5-97. *Two microswitches with metal arms are mounted on each side of the cart, where they will sense any obstacle.*

The Circuit

The schematic is very, very simple, with only four principal components: two microswitches that sense obstacles in front of the cart, one relay, and one 555 timer. You will also need a small power switch, a battery or battery pack, and a resistor, and capacitors to go with the timer. A trimmer potentiometer will allow you to adjust the "on" time of the 555 timer, which will determine how long the cart takes to back up. See Figure 5-98.

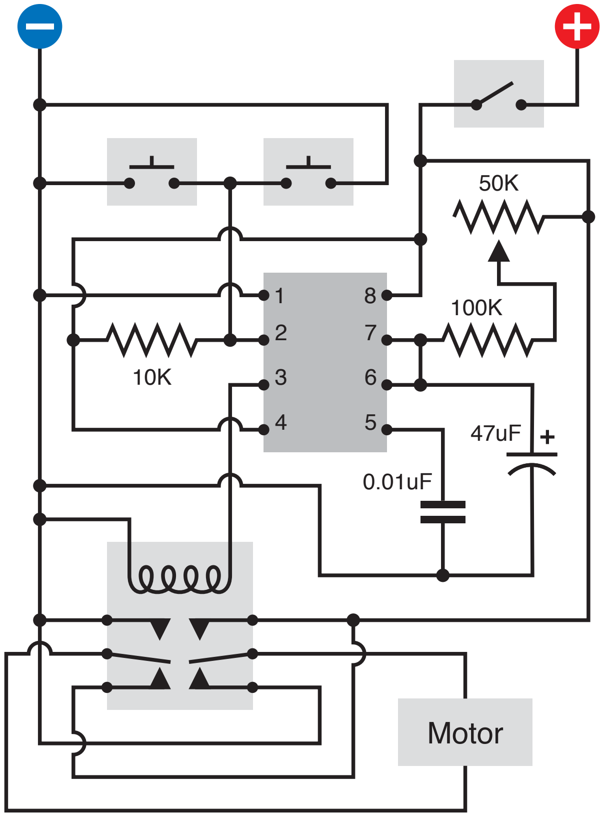

Figure 5-98. *This ultrasimple schematic is all the cart needs to enable it to back up when it hits an obstacle.*

The motor I chose requires 5 volts, so I had to use a voltage regulator with a 9-volt battery. If your motor uses 6 volts, you can wire four AA batteries to it directly. If you have a 12-volt motor, you can use two 9-volt batteries in series, supplying power through a 12-volt voltage regulator.

Assemble the components, mount them on the cart, and switch it on, and it should move forward slowly in a more-or-less straight line. If it moves backward, reverse your connection to the terminals on the motor.

When the cart bumps into something, either of the microswitches will connect negative voltage to the input pin of the 555 timer. This triggers the timer, which runs in monostable mode, generating a single pulse lasting about 5 seconds, which closes the relay, which is wired so that it reverses the voltage to the motor.

When the voltage is reversed to a simple DC motor, it runs backward. So the cart backs up. Because the rear wheel is mounted in a yoke that pivots, the yoke will tend to flip one way or the other, causing the cart to describe an arc as it moves backward. At the end of the timer cycle, the relay relaxes and the cart starts moving forward again. In forward mode, the rear wheel just follows along without applying any steering force, so the cart tends to follow a straight line—until it hits another obstacle, at which point it backs up, and tries another path.

FUNDAMENTALS

All about limit switches

The most obvious enhancement for your cart would be a better steering mechanism. You could use another motor to take care of this, with a pair of limit switches. Because limit switches are a basic, important idea in conjunction with motors, I'll explain them in detail.

Figure 5-99 shows three successive views of a motor with an arm attached to it, which can press either a lower pushbutton or an upper pushbutton. Both of the pushbuttons are normally closed, but will open when pressed by the motor arm. These buttons are the limit switches. Typically you would use microswitches for this purpose, just like the ones that I suggested as barrier-sensors at the front of the cart.

In addition, there's a DPDT relay that is activated by a simple on/off switch at the righthand side. On the cart, the 555 timer takes the place of the on/off switch, by feeding power to the relay.

Suppose that the motor begins with the arm pointing downward, as shown in the top view in Figure 5-99, and the motor is wired so that when it receives negative voltage at its lower terminal and positive at its upper terminal, it rotates counter-clockwise. This is what happens when the on/off switch closes and sends power to the DPDT relay. Positive voltage from the relay contacts cannot pass through the upper diode, but can pass through the upper limit switch, which is closed. Negative voltage cannot pass through the lower limit switch, because it's open, but can pass through the lower diode. So, the motor starts to turn counterclockwise. During the midpoint of its arc, it receives power through both of the limit switches.

Finally, the motor arm reaches the upper switch, and opens it. This prevents positive voltage from reaching the motor through that switch, and the positive voltage is also blocked by the upper diode. So, at this time, the motor stops.

Now suppose that the on/off switch is opened, as in the top view in Figure 5-100. The relay loses its power, so its contacts relax. The voltage to the motor is now reversed. Negative voltage passes through the upper diode, while positive voltage reaches the motor through the lower limit switch. The motor starts running clockwise, until its arm hits the lower switch, opening it and cutting off power to the motor.

Limit switches are necessary, because if you continue to apply voltage to a simple DC motor that is unable to turn, the motor sucks more current, gets hot, and may burn out.

You can easily see how this kind of system could be used to control the cart's steering. Even though the motor has only two positions, these are sufficient to make the cart turn when going backward, and proceed straight ahead when going forward.

To reduce power consumption, the DPDT relay could be replaced with a two-coil latching relay. The circuit would then have to be revised so that the relay is flipped to and fro by a pulse to each of its coils.

FUNDAMENTALS

All about limit switches (continued)

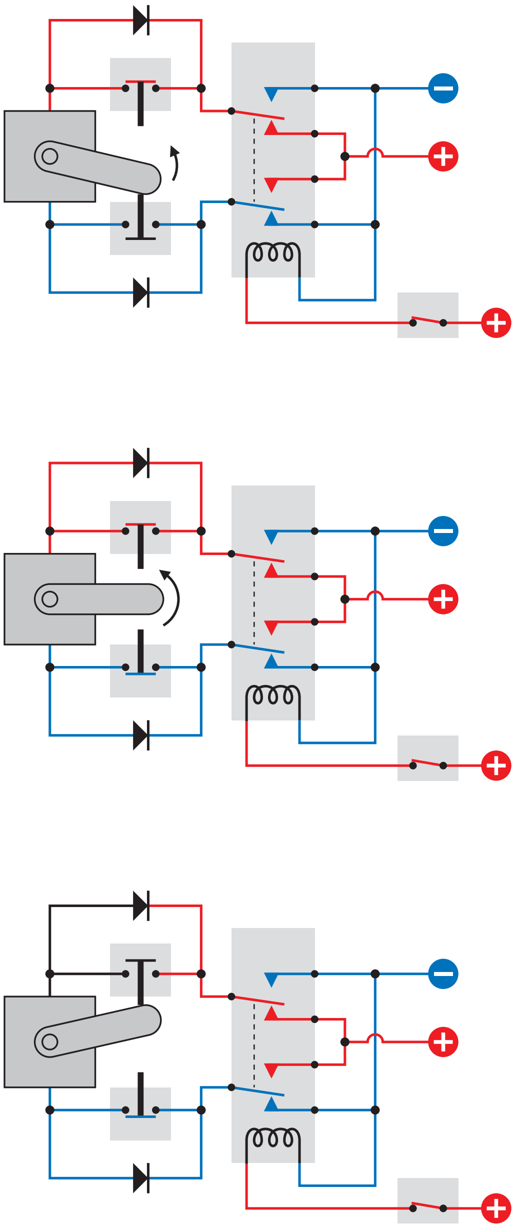

Figure 5-99. *The three diagrams, from top to bottom, show three snapshots of a motor controlled by a DPDT relay and two limit switches. When the on/off switch at bottom-right sends power to the relay, the lower relay contacts cause the motor to run counterclockwise until it stops itself as its arm opens the upper limit switch.*

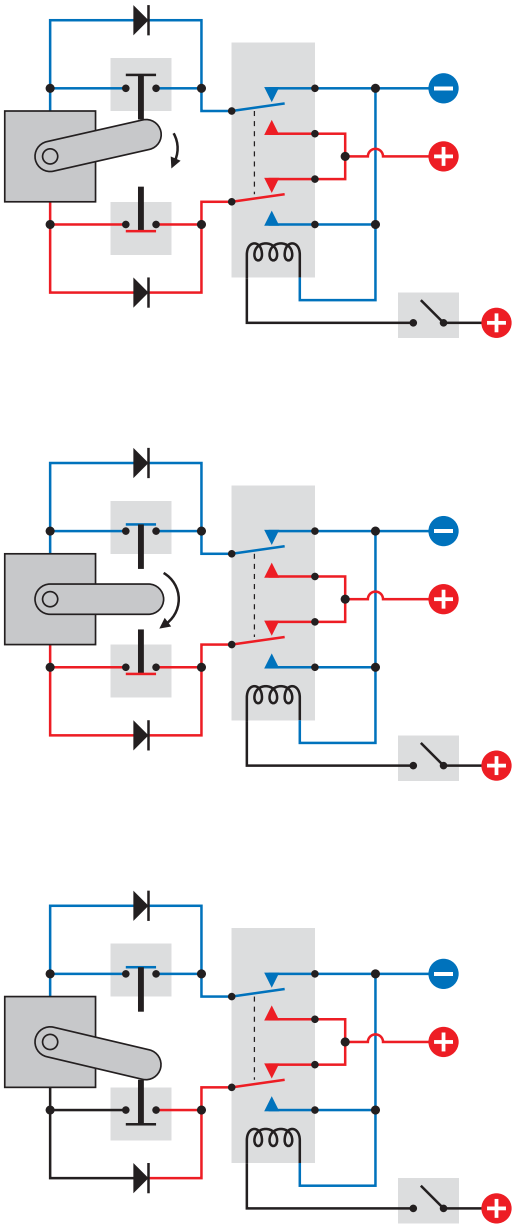

Figure 5-100. *When the on/off switch at bottom-right opens, the relay connects its upper contacts. This causes the motor to run clockwise until its arm opens the lower limit switch. Limit switches avoid the overheating and possible damage that are likely when power is delivered to a motor that is prevented from turning.*

FUNDAMENTALS

All about motors

Brushed DC motor

This is the oldest, simplest design for an electric motor, shown in very simplified form in Figure 5-101. Coils are attached to a shaft where they can interact with stationary magnets around them. The magnetic attraction turns the shaft a little, at which point the next coil on the shaft is energized to turn the shaft a little more, and then the next coil—and so on. To make this happen, electricity has to be fed into the coils by "brushes," often consisting of soft carbon pads that conduct power to a hub, known as a commutator, divided into sections, each of which is connected to a separate coil.

This basic design has several advantages if we want to build a small motorized gadget, such as a miniature robot or even a model airplane:

- Widely available
- Low cost
- Simple
- Reliable
- Will run in reverse when voltage reverses

In addition, brushed motors are often sold with reduction gearing built in. Such units are known as *gearhead motors* or *gear motors*. They free you from the need to use your own gears or belts to adjust the output speed yourself. You simply choose the motor that fits your specification.

DC stepper motor

This requires a controller, consisting of some electronics to tell the motor to rotate its shaft in small, discrete steps. The advantages of a stepper motor are:

- Precise positioning of the shaft
- Precise speed adjustment

Stepper motors are ideal for devices such as computer printers, where the paper has to roll up by a precise distance and the print head has to move laterally by an equally precise distance, but they are also useful in robots. If the motor is small enough to draw less than 200mA and will run on 12 volts or less, you can control it with pulses from a 555 timer. I'll describe stepper motors in more detail in Experiment 33.

Servo motor

This is generally used in conjunction with a programmable microcontroller, which sends instructors to rotate the motor shaft to a specific position and then hold it there. I'll mention servo motors when I introduce you to microcontrollers, but we won't be dealing with them in detail.

Other types of motors exist, including brushless DC motors (which require a different type of controller and are found in computer disk drives and CD players), and AC motors (including synchronous motors, which synchronize their rotation with the frequency of AC voltage, and were used extensively in clocks, before clocks mostly became digital).

In this book, I'll be talking mostly about brushed DC motors and DC stepper motors.

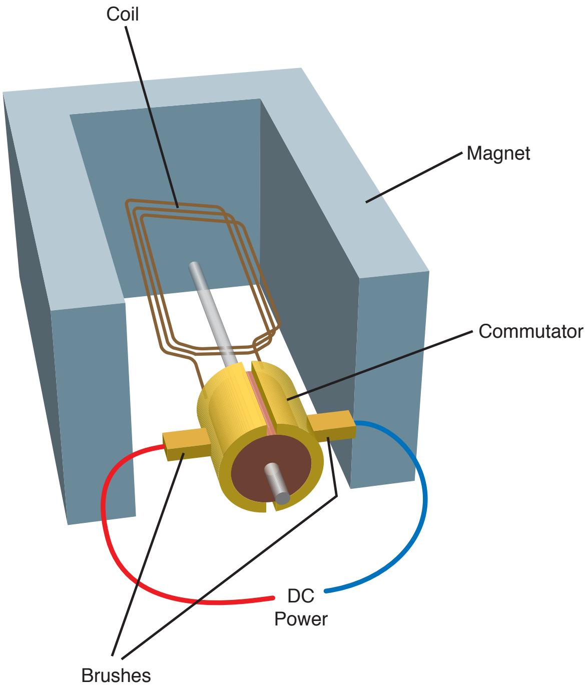

Figure 5-101. *The basic principle of a simple DC motor. The commutator passes electricity through a coil, creating a magnetic field that interacts with a magnet around the motor. The coil turns, and the commutator turns with it, until the electric field through the coil is reversed. This causes the process to repeat. In reality, a motor is likely to have a commutator formed from multiple segments, connected with multiple coils. The principle, however, remains the same.*

Take-home messages from this experiment include the following:

- You can buy simple DC motors with reduction gearing built in, providing your choice of RPM. Literally hundreds of websites will sell you small motors for robotics projects.
- When you reverse the voltage to a DC motor, the motor runs in reverse.
- A DPDT relay can be wired so that when it closes its contacts, it reverses a power supply to a motor.
- You can use two limit switches and a pair of diodes to stop a motor at two positions. In each of its stopping positions, the motor consumes no power and you won't have the risk of it burning out.

What other projects can you imagine using this simple set of techniques?

Mechanical Power

In the United States, the turning force, or torque, of a motor is usually measured in pound-feet or ounce-inches. In Europe, the metric system is used to measure torque in dynes.

A pound-foot is easy to understand. Imagine a lever pivoted at one end, as shown in Figure 5-102. If the lever is one foot long, and you hang a one-pound weight at the end of it, the turning force is one pound-foot.

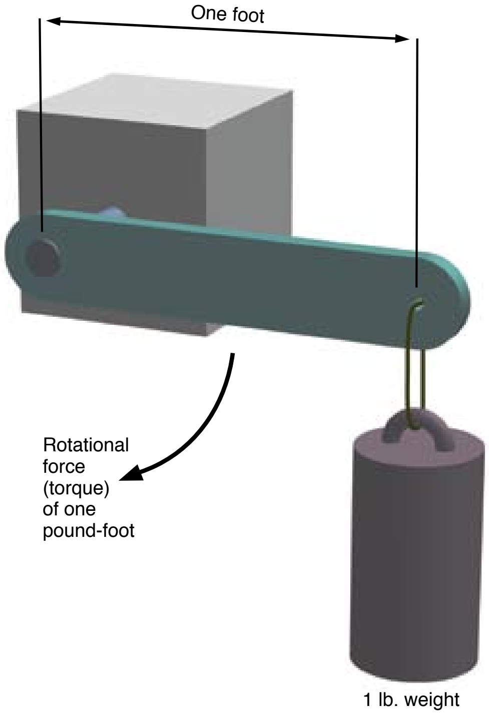

Figure 5-102. *The rotational force created by a motor is known as "torque," and in the United States it is measured in pound-feet (or ounce-inches, for small motors). In the metric system, torque is measured in dynes. Note that the torque created by a motor will vary according to the speed at which the motor is running.*

FUNDAMENTALS

Wire gauges

If you're going to power larger motors, or other components that take more current than LEDs or small relays, you really need to know about wire gauges. In particular, what's the relationship between wire thickness and AWG (American Wire Gauge)? And what gauge of wire should you use for any given current?

You can find numerous charts and tables if you go online, but many of these sources contradict each other, especially on the topic of how much current is safe to run through each gauge of wire.

After making several comparisons (and testing some wire samples myself), I've compiled the table in Figure 5-103, which I recommend as a compromise. Note the following:

- This table applies to solid-core copper wire.
- For stranded wire, or copper that has been tinned (giving it a silver appearance), the number of ohms per foot will *increase*, the number of feet per ohm will *decrease*, and the maximum amperage will *decrease*, probably by around 20%.

The maximum amperage assumes that the wire is insulated, preventing it from radiating heat as effectively as a bare conductor. I am also assuming that the wire is likely to be at least partially enclosed, inside a box or cabinet. At the amperages listed for each gauge of wire, you should expect the wire to become noticeably warm, and personally I would tend to use thicker wire instead of the maximums indicated in the table.

Most tables of this type only tell you the resistance of each gauge of wire in ohms per 1,000 feet. I have included that number but have also expressed the function the other way around, as the number of feet per ohm, as this doesn't require you to do so much arithmetic with decimals.

AWG		Diameter in inches	Ohms per 1,000 ft	Feet per ohm	Maximum amps (insulated)
0000	●	0.46	0.049	20,400	225
000	●	0.41	0.062	16,200	200
00	●	0.365	0.078	12,800	175
0	●	0.325	0.098	10,200	150
1	●	0.289	0.124	8,070	125
2	●	0.258	0.156	6,400	100
3	●	0.229	0.197	5,080	90
4	●	0.204	0.249	4,020	80
5	●	0.182	0.313	3,190	70
6	●	0.162	0.395	2,510	60
7	●	0.144	0.498	2,010	51
8	●	0.128	0.628	1,590	44
10	●	0.102	0.999	1,000	32
12	●	0.081	1.59	630	23
14	●	0.064	2.53	396	17
16	•	0.051	4.02	249	13
18	•	0.04	6.39	157	10
20	•	0.032	10.2	99	8
22	·	0.025	16.1	62	5
24	·	0.02	25.7	39	2.5
26	·	0.016	40.1	25	1.5
28	·	0.013	64.9	15	1.0
30	·	0.010	103.2	10	0.5

Figure 5-103. *American wire gauges (AWG) and their properties.*

THEORY

Calculating voltage drop

Another fact that you often need to know is how much of a voltage drop a particular length of wire will introduce in a circuit. If you want to get maximum power from a motor, you don't want to lose too much voltage in the wires that go to and from the motor.

Voltage drop is tricky, because it depends not only on the wire, but also on how heavily the circuit is loaded. Suppose that you are using 100 feet of 22-gauge wire, which has a resistance of about 1.5 ohms. If you attach it to a 12-volt battery and drive an LED and a series resistor offering a total effective resistance of about 1,200 ohms, the resistance of the wire is trivial by comparison. According to Ohm's Law:

amps = volts / ohms

so the current through the circuit is only about 10mA.

Again, by Ohm's Law:

volts = ohms × amps

so the wire with resistance of 1.5 ohms imposes a voltage drop of 1.5 × 0.01 = 0.015 volts.

Now suppose you're running a motor. The coils in the motor create impedance, rather than resistance, but still if we measure how much current is going through the circuit, we can establish its effective resistance. Suppose the current is 1 amp. Repeating the second calculation:

volts = ohms × amps

So the voltage drop in the wire is now 1.5 × 1 = 1.5 volts! This is illustrated in Figure 5-104.

Bearing these factors in mind, I have compiled a table for you. I've rounded the numbers to just two digits, as variations in the wire that you use make any pretense of greater accuracy unrealistic.

To use this table, you need to know how much current is passing through your circuit. You can calculate it (by adding up all the resistances and dividing it into the voltage that you are applying) or you can simply measure the current with a meter. Just make sure that your units are consistent (all in ohms, amps, and volts, or milliohms, milliamps, and millivolts).

In the table, I have arbitrarily assumed a length of 10 feet of wire. Naturally you will have to make allowances for the actual length of wire in your circuit. The shorter the wire, the less the loss will be. A circuit with only 5 feet of wire, and the same amperage and voltage, will suffer half of the percentage loss shown in the table. A circuit with 15 feet of wire, and the same amperage and voltage, will suffer 1.5 times the percentage loss. So, to use the table:

1. Divide your length of wire by 10. (Make sure that you measure the length in feet.)
2. Use the result to multiply the number in the table.

The table also arbitrarily assumes that you have a 12-volt supply. Again, you will have to make allowances if you are using a different voltage. So, to use the table:

1. Divide 12 by the actual voltage of your power supply.
2. Use the result to multiply the number in the table.

I can summarize those two steps like this:

Percent voltage lost = P × (12 / V) × (L / 10)

where P is the number from the table, V is your power-supply voltage, and L is the length of your wire.

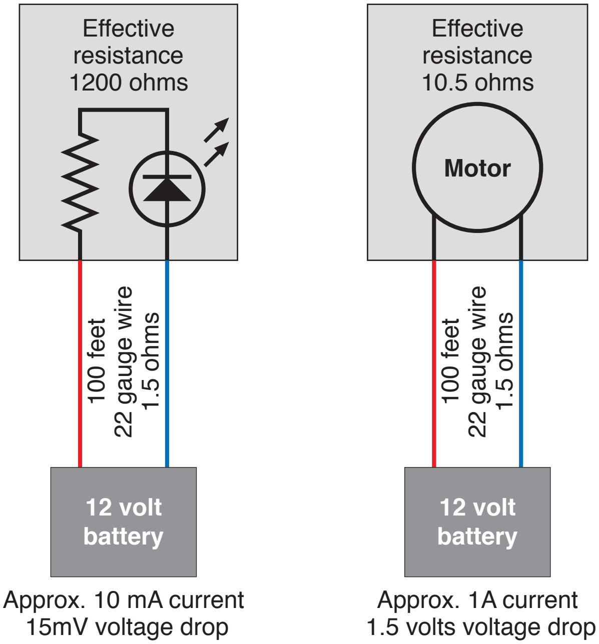

Figure 5-104. *The voltage drop imposed by wiring will depend on the current and the resistance in the circuit. The drop will be greatest when the resistance of the circuit is low and the amperage is high.*

THEORY

Calculating voltage drop (continued)

This table shows the percent voltage lost in a circuit with 10-foot wire at 12 volts.

Wire Gauge	Amperes									
	1	2	3	4	5	6	7	8	9	10
10	0.08	0.17	0.25	0.33	0.42	0.50	0.58	0.67	0.75	0.83
12	0.13	0.27	0.40	0.53	0.66	0.80	0.93	1.1	1.2	1.3
14	0.21	0.42	0.63	0.84	1.1	1.3	1.4	1.5	1.9	2.1
16	0.33	0.67	1.0	1.3	1.7	2.0	2.3	2.7	3.0	3.4
18	0.53	1.1	1.6	2.1	2.7	3.2	3.7	4.3	4.8	5.3
20	0.85	1.7	2.6	3.4	4.3	5.1	6.0	6.8	7.7	8.5
22	1.3	2.7	4.0	5.4	6.7	8.1	9.4	11	12	13
24	2.1	4.3	6.4	8.6	11	13	15	17	19	21
26	3.4	6.8	10	14	17	20	24	27	31	34
28	5.4	11	16	22	27	32	38	43	49	54
30	8.6	17	26	34	43	52	60	69	77	86

Remember, though, that the wire resistance will be higher if you are using stranded copper wire or tinned copper wire, and this will increase the percentage of voltage lost.

Experiment 33: Moving in Steps

Time now to build something more sophisticated: a cart that orients itself toward a light source. I'm going to tell you all you need to get started on this project, but this time I won't go all the way to the end in exhaustive detail. I want you to get into the habit of figuring out the details, improving on plans, and eventually inventing things for yourself.

You will need:

- 555 timers. Quantity: 8.
- Trimmer potentiometer, 2K linear. Quantity: 2.
- LEDs. Quantity: 4. If you get tired of using series resistors to protect LEDs in a 12-volt circuit, consider buying 12-volt LEDs such as Chicago Miniature 606-4302H1-12V, which contain their own resistors built in. However, the schematic in Figure 5-108 assumes that you will use regular 2V or 2.5V LEDs.
- Stepper motor: Unipolar, four-phase, 12-volt. Parallax 27964 or similar, consuming 100mA maximum. Quantity: 2.
- Photoresistors, ideally 500 to 3,000Ω range. Quantity: 2.

- ULN2001A or ULN2003A Darlington arrays by STMicroelectronics. Quantity: 2.
- CMOS octal or decade counter. Quantity: 2.
- Various resistors and capacitors.

Exploring Your Motor

I've specified a unipolar, four-phase, 12-volt motor because this is a very common type. A typical sample is shown in Figure 5-105. If you can't easily find the one that I've listed, you should feel safe in buying any other that has the same generic description. "Unipolar" means that you don't have to switch the power supply from positive to negative and back to positive again, to run the motor. Four-phase means that the pulses that run the motor must be applied in sequence to four separate wires. Because you will be running your motor directly from 555 timers, the lower its power consumption, the better.

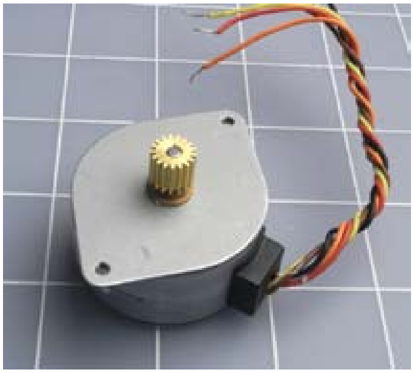

Figure 5-105. *A typical stepper motor. The shaft rotates in steps when negative pulses are applied to four of the wires in sequence, the fifth wire being common-positive.*

First, though, we can apply voltage to the motor without using any other components at all. Most likely it will have five wires already attached, with the ends stripped and tinned, so that you can easily insert them into holes in a breadboard, as shown in Figure 5-106. Check the data sheet for your motor; you should find that four of the wires are used to energize the motor and turn it in steps, while the fifth is the common connection. In many cases, the common connection should be hooked to the positive side of your power supply, while you apply negative voltage to the other four wires in sequence, one step at a time.

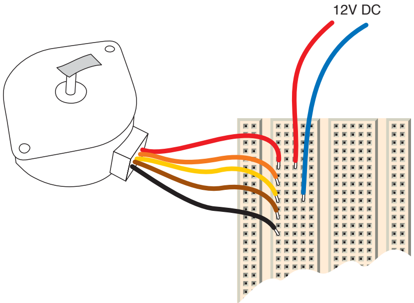

Figure 5-106. *The simplest test of a stepper motor is to apply voltage manually to each of its four control wires, while a piece of duct tape, attached to the output shaft, makes it easy to see how the motor responds.*

The data sheet will tell you in what sequence to apply power to the wires. You can figure this out by trial and error if necessary. One thing to bear in mind: a stepper motor is very tolerant. As long as you apply the correct voltage to it, you can't burn it out.

To see exactly what the motor is doing, stick a piece of duct tape to the end of the shaft. Then apply voltage to wires, one at a time, by moving your negative power connection from one to the next. You should see the shaft turning in little steps.

Inside the motor are coils and magnets, but they function differently from those in a DC motor. You can begin by imagining the configuration as being like the diagram in Figure 5-107. Each time you apply voltage to a different coil, the black quadrant of the shaft turns to face that coil. In reality, of course, the motor turns less than 90° from one coil to the next, but this simplified model is a good way to get a rough idea of what's happening. For a more precise explanation, see the upcoming section "Theory: Inside a stepper motor."

Bear in mind that as long as any of the wires of the motor are connected, it is constantly drawing power, even while sitting and doing nothing. Unlike a regular DC motor, a stepper motor is designed to do nothing for much of the time. When you apply voltage to a different wire, it steps to that position and then resumes doing nothing.

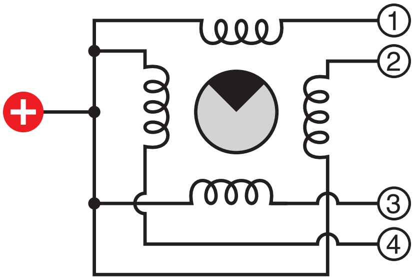

Figure 5-107. *This greatly simplified diagram helps in visualizing the way in which a stepper motor works. In reality, almost all motors rotate by less than 90° in response to each pulse.*

The coil inside the motor is holding the shaft in position, and the power that the motor draws will be dissipated as heat. It's quite normal for the motor to get warm while you're using it. The trouble is, if you use a battery to power it, and you forget that you have it connected, the battery will not hold its charge for long.

Quick Demo

Now that you've proved that your motor is functional, how can you actually run it? You need to send a pulse to each of the four wires in turn, in a rapidly repeating sequence. If you can also adjust the speed of the pulses, so much the better. I'm thinking that for a quick and simple demo, you can handle the challenge simply by using four 555 timers, all of them in monostable mode, with each one triggering the next.

The schematic in Figure 5-108 shows what I have in mind. It looks more complex than it really is. Each timer has the same pattern of components around it, so after you create the first module, you just make three copies of it.

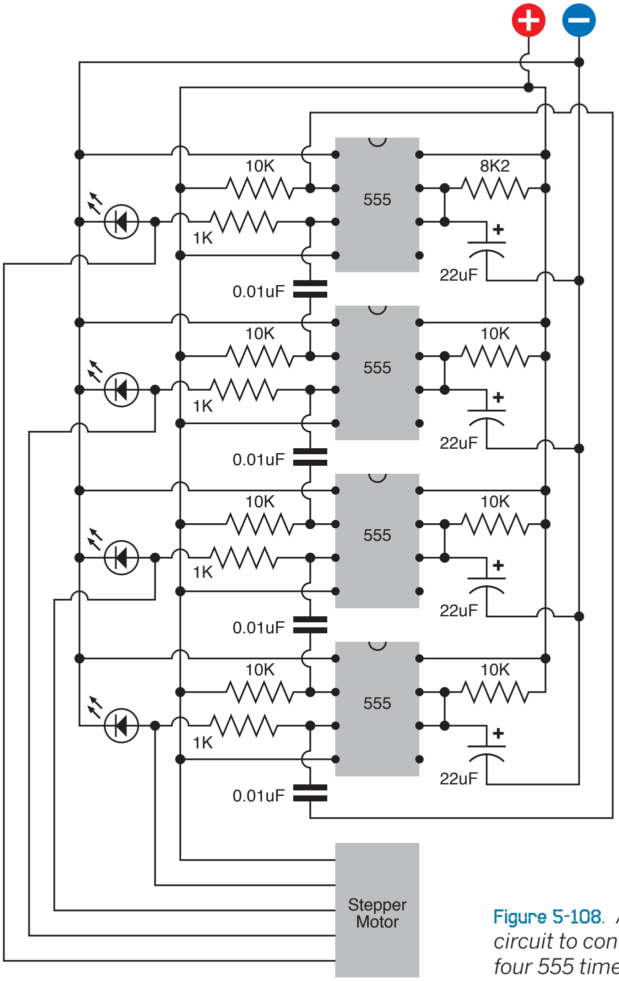

Figure 5-108. *A very quick and simple circuit to control a stepper motor uses four 555 timers, each in monostable mode, triggering each other in a repeating sequence.*

I've used a 10K resistor to pull up the input to each 555, so that the timers are naturally in their quiescent state. A 0.01 µF capacitor links the output from one timer to the input of the next so that they are electrically isolated from each other, and the capacitor just conveys a "spike" of voltage when one timer finishes its "on" cycle, and its output goes low, which triggers the next.

On the righthand side, I've used 10K resistors and 22 µF capacitors to generate a cycle of about a quarter of a second—except that the topmost timer has a 8K2 timing resistor. The reason for this is that when power is first applied, the timers will all be waiting for each other to begin, and timers 2 and 4 or 1 and 3 may fire together. By giving one timer a shorter cycle than the others, I minimize this problem.

The LEDs are included just to give you some visual verification of what's happening. Without them, if you make a wiring error, the motor may turn to and fro erratically, and you won't know why. Initially you can run your circuit with only the LEDs connected, just to make sure it works. Figure 5-109 shows the breadboarded circuit before the motor is plugged in. Then add the motor by plugging its wires into the breadboard, where you'll make connection with the outputs (pins 3) of the timers. See Figure 5-110.

Apply power, and you should see the motor turning in steps, in sequence with the LEDs. If the LED sequence isn't stable:

1. Connect a wire directly from the input (pin 2) of the topmost timer to the positive side of the voltage supply, and wait for the timers to calm down.
2. Restart the sequence by disconnecting the free end of this wire, or (if necessary) touch the free end of it briefly to the negative side of the supply, to trigger the first timer.

One thing you may have noticed, if you're paying very close attention: the common terminal of the motor is connected to positive. Therefore, when each timer flashes positive, that positive signal isn't actually powering the motor. The *low* outputs from the three timers that are *not* firing at any given moment are sinking current from the motor. It seems quite happy with this arrangement. You'll need some theory to understand why.

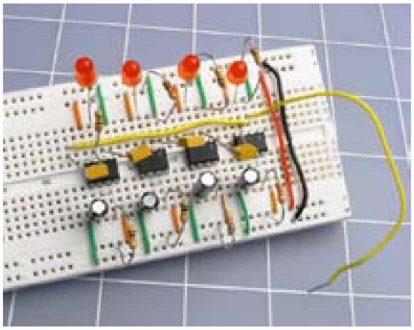

Figure 5-109. *To test the control circuit for errors, four LEDs show the outputs from the four 555 timers. The loose yellow wire at the righthand side connects to pin 2 of the first timer. Touch the free end of this wire to the positive side of the power supply to reset the timers, and then, if necessary, make a brief negative connection with the free end of the wire to restart their sequence.*

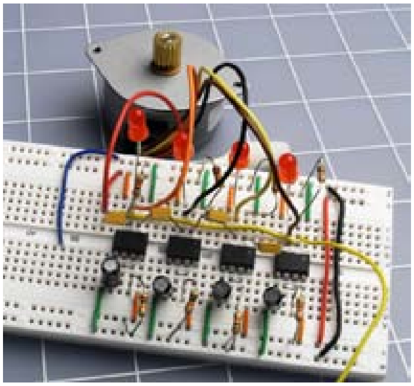

Figure 5-110. *After the circuit has been tested, the motor can be added by hooking its control wires to the outputs of the four 555 timers.*

THEORY

Inside a stepper motor

If you check the Wikipedia entry for stepper motors, you may see a very nice 3D rendering showing a toothed rotor and four coils arrayed around it. Maybe stepper motors used to be manufactured like this once upon a time, but not anymore.

Imagine two horizontal rows of coils. In the space between them is a series of little magnets, like a freight train, that can move left or right, as shown in Figures 5-111 and 5-112. Each coil has two windings, in opposite directions, so that current through one winding will create an upward magnetic force while current through the other will create a downward force. Each row of windings is connected in parallel, so that they switch on and off simultaneously.

In Step 1, the negative connection energizes the upper windings of the upper coils, which creates an upward magnetic force. I've shown this force using blue-green arrows so that you won't mistake it for a flow of electricity. It so happens that this force attracts the north poles of the magnets and repels the south poles, so if the magnets begin in the position shown in Step 1, they will want to move one step to the right.

This brings them to the position shown in Step 2. Now the upper windings of the lower coils are energized, and again, this produces an upward force, which again attracts the north poles and repels the south poles.

This advances the magnets to their location in Step 3. Now the lower windings of the upper coils are energized, producing a downward force. This repels the north poles of the magnets and attracts their south poles. So the magnets keep moving.

They reach the position shown in Step 4. The lower windings of the lower coils are energized, producing a downward force which continues to attract the south poles while repelling the north poles. So the magnets move a final step to the right—which leaves them in the same orientation shown in Step 1. And the process can repeat all over again.

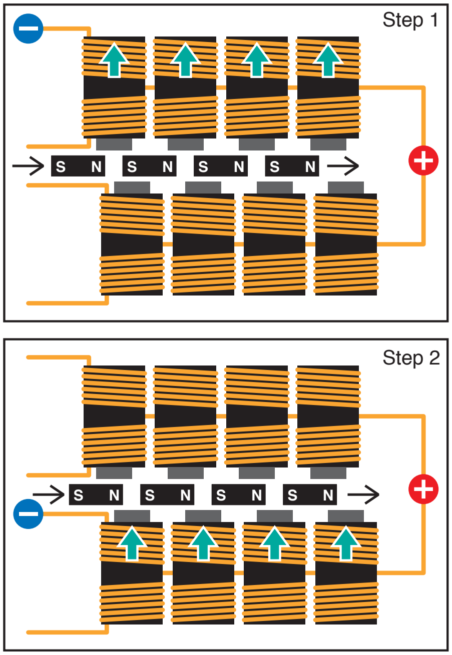

Figure 5-111. *This sequence shows the first two steps as the rotor of a stepper motor (shown as a series of north-south magnets) moves in response to pulses through electromagnets.*

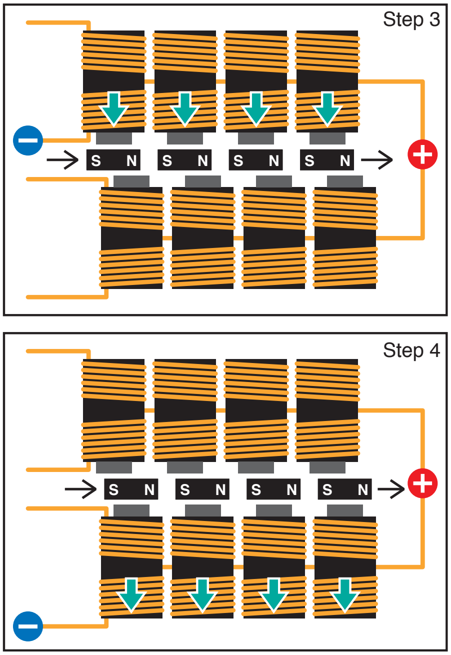

Figure 5-112. *After taking another two steps, the motor will be back where it started at Step 1 in Figure 5-111.*

THEORY

Inside a stepper motor (continued)

In reality, the magnets are not separate from each other. The edge of a rotor is magnetized in zones that alternate between south and north polarity. And instead of multiple coils, there are just four windings that go around all the magnetic cores. But the principle is exactly the same. The 3D rendering gives a general idea, and the photograph shows what I found when I cracked open a typical stepper motor.

Now bear in mind that when this device is driven by a set of 555 timers, we don't just connect negative to one wire at a time on the left, leaving the others floating. In reality, at any given moment, three of the timers have a negative output and the fourth has a positive output. The last diagram in Figure 5-112 shows this situation.

Suppose the top wire is positive while the other three are negative, as shown in Figure 5-113. The positive output does nothing, because it is balanced by the positive power on the other end of the coils. The two negatives attached to the bottom set of coils create equal and opposite forces that cancel each other out (while wasting some power). So the net result is the same as in Step 3.

In fact, you should find that you can disconnect the common wire completely while using the stepper motor with 555 timers, and the motor will still turn, because one of the timers is providing positive power while the others are negative. In fact, you'll be running them more efficiently this way.

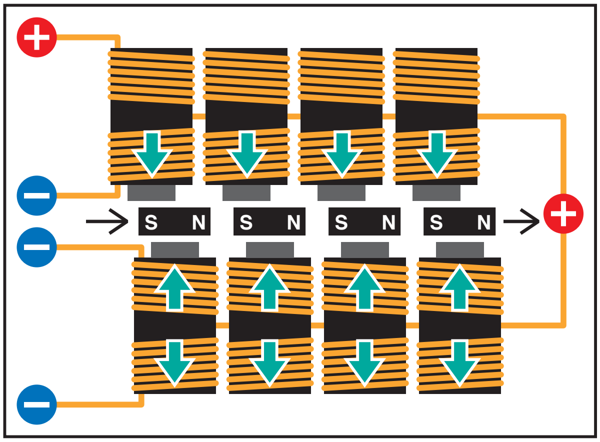

Figure 5-113. *When the motor is driven by four 555 timers, they are activating it by sinking positive voltage from it. The interior workings of the motor look something like this. It's not the most efficient way to do the job.*

Figures 5-114 and 5-115 may help to give you a clearer idea of what the motor actually looks like inside.

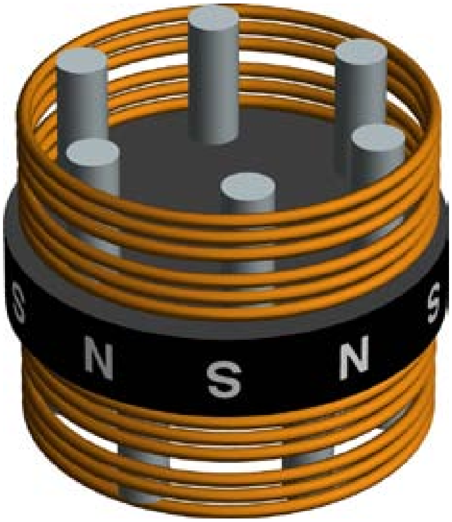

Figure 5-114. *This 3D rendering gives a better idea of what a typical stepper motor looks like inside. The copper coils and gray cylinders are stationary, while the black disc rotates between them.*

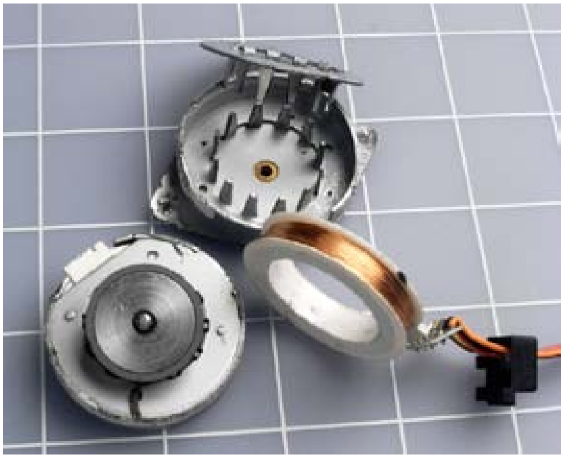

Figure 5-115. *When a stepper motor is broken open, this is what you're likely to find. On the left, the rotor of the motor, which has a magnetized band around its circumference, is still attached to the lower half of the casing. On the right, the upper half of the casing has been opened, and the coil has been removed (actually the winding you can see consists of two coils, wound in opposite directions). The spikes are the magnetic cores that exert force on the rotor.*

Speed Control

If you are a truly exceptionally observant, you may have noticed that I left pin 5 of each of the timers unconnected in the schematic for driving the stepper motor in Figure 5-108. Normally, pin 5 should be grounded through a capacitor to prevent it from picking up stray voltages which can affect the accuracy of the chip.

I left the pins unconnected because I had a plan for them. In fact, changing the timing of the chip is exactly what we want to do now, as a way to change the speed of the stepper motor.

If you tie pin 5 of all four timers together, as shown in Figure 5-116, and put a 2K trimmer potentiometer (shown in Figure 5-117) between them and the negative side of the power supply, you'll find that as you turn the trimmer to reduce its resistance, the timers start to run faster. Figure 5-118 shows the breadboard layout. Eventually, when the resistance goes below around 150 ohms, everything stops. The LEDs go dark, because you've reduced the voltage on pin 5 below the threshold level that the 555 timer finds acceptable.

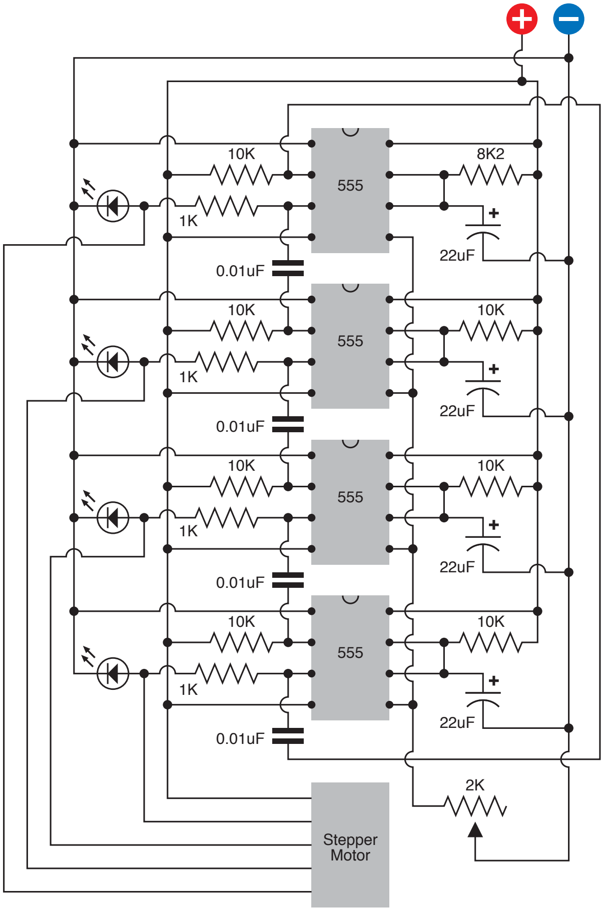

Figure 5-116. *To adjust the speed of the sequence of 555 timers, their control pins (pin 5 on each timer) are linked together and attached to a trimmer potentiometer that adjusts the resistance between the pins and the negative side of the power supply.*

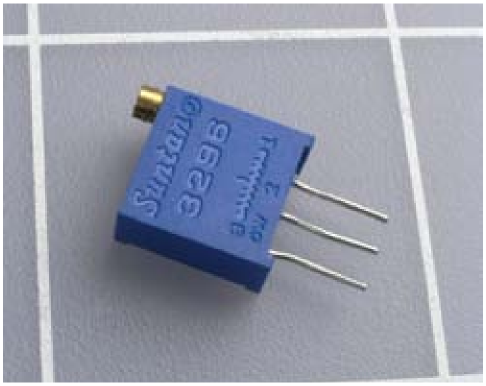

Figure 5-117. *Close-up of a trimmer potentiometer with pins spaced at 1/10 inch for insertion in a breadboard or perforated board. The brass screw, at top-left, turns a worm gear inside the unit, allowing precise adjustment of internal resistance.*

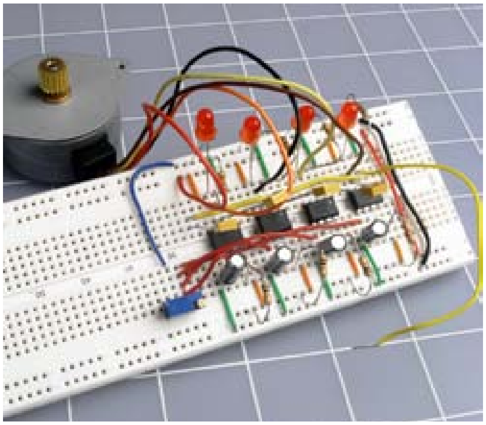

Figure 5-118. *The trimmer potentiometer has been added to the circuit, allowing motor speed control.*

Initially I suggested a step time of 1/4 second just so that you could see what was happening. When you're actually using this circuit, you'll never need it to run as slowly as that. So you can increase the entire range of speeds. Remove the 22 µF timing capacitors and substitute, say, 4.7 µF capacitors, or smaller. Now when you adjust the potentiometer, you'll get a useful range of speed.

Adding Autonomy

Currently, the circuit simply does what you tell it to do. The next step is to make it autonomous—in other words, give it the illusion of making up its own mind. I'm thinking that instead of a trimmer potentiometer, we could substitute a photocell, properly known as a photoresistor. Typically, the resistance of a cadmium sulfide photo resistor is highest in the dark, and lowest when light shines on it.

One problem with photoresistors is that they're not as widely available as many other types of electronic components. If you search Mouser.com, for instance, you'll find virtually nothing. Partly this is because the online search function at Mouser is the weakest feature of the site, and partly it's because Mouser is not oriented toward hobbyists. What you need to do is conduct a "product search." Go to *http://www.google.com/products*, enter the search terms "CdS" and "photocell," and you'll find a bunch of cheap cadmium sulfide components from places you may never have heard of.

Because photoresistors seem to come and go as erratically as DC motors, I am not offering any part numbers. You can buy any product that has an appropriate minimum resistance (in bright light) and maximum resistance (in the dark). If you find a component that ranges from 500 to 3,000Ω, that would be a good choice. If the only ones you can find have a higher minimum than 500Ω, you could consider putting a couple of them in parallel.

Setting Up Your Light Seeking Robot

Why would you want to control the speed of a stepper motor by using a photo resistor? Because the original objective was to build a robot that is attracted to light.

The idea is simple enough: use two stepper motors, each controlling the speed of one wheel of the cart. Use two photoresistors, each controlling the speed of the opposite stepper motor. When the righthand photoresistor picks up more light, its resistance lowers, causing the lefthand set of timers to run faster, which will make the lefthand wheel run faster. Thus, the cart will turn toward the light. Figure 5-119 illustrates the concept.

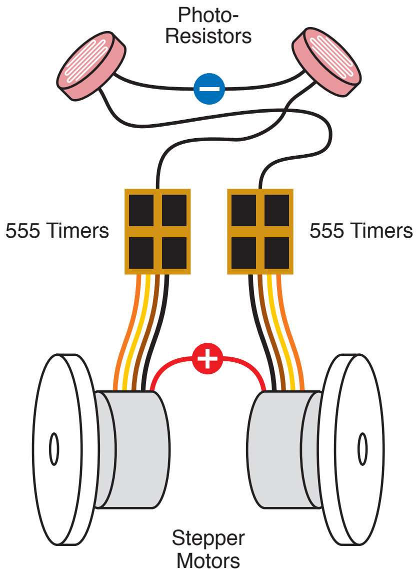

Figure 5-119. *If two photoresistors control the speed of two 555-timer arrays, the difference in speed between one wheel and the other can turn the cart toward a light source.*

Before you start wiring more 555 timers, though, you might consider doing the job with a more appropriate component. The ULN2001A and ULN2003A are chips containing Darlington amplifiers specifically designed to deliver current to inductive loads such as solenoids, relays, and (you guessed it) motors. Each chip has seven inputs that require very little current, and seven outputs that can deliver 500mA each. The inputs are TTL and CMOS compatible (the 2001A has a wider tolerance for voltages than the 2003A) and each channel of the chip functions as an inverter, so that when the input goes high, the output goes low and sinks current. This is of course just what we need for our stepper motor that has a common positive connection.

The ULN2001A is only an amplification device, so you have to precede it with a counter that runs from 1 to 4 and then repeats. You can stick with your 555 timers, as you've already assembled them, or substitute almost any CMOS octal or decade counter that sends its output pulses to a series of pins. Just use the output from the fifth pin as the "carry" output to restart the counting sequence. I suggest a CMOS counter simply because it will run on 12 volts, so you can use the same power supply that suits your stepper motors.

If you switch to CMOS counters, you will still need a pair of 555 timers sending pulses to the counters. The timers will be free-running in astable mode, and your photoresistors will control their speed. Figure 5-120 shows the configuration.

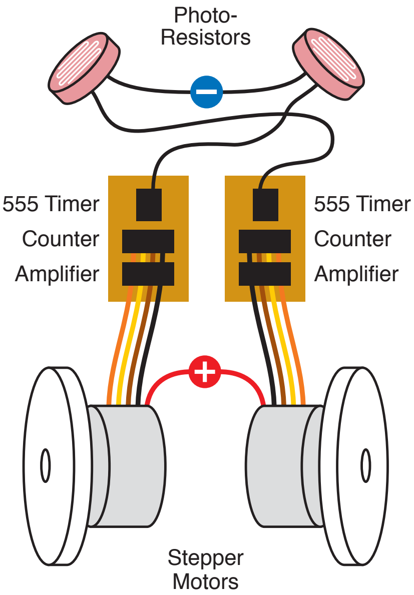

Figure 5-120. *A more efficient way to drive the motors is to use just one timer to set the speed of each, with a counter and amplifier (such as a Darlington array chip) sending the pulses down the wires. The principle is still the same, though.*

One last item: you'll need a 12-volt battery. You can of course put eight AA cells together, but I think you should consider a rechargeable pack from a source such as *http://www.all-battery.com*, which has a section entirely devoted to "robot batteries."

If you put it all together, you should find that when you place your robot cart in a very dimly lit room, it will turn toward the beam from a bright, well-focused flashlight. To get reliable results, you may have to recess each of the photoresistors in little tubes, so that they receive much more light when they face your flashlight than when they face away from it. Figure 5-121 is a 3D rendering of the concept.

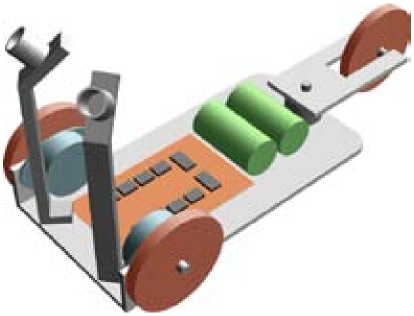

Figure 5-121. *This 3D rendering shows a possible configuration of the light-seeking cart, with two photoresistors enclosed in small tubes to restrict their response to light.*

Another idea is to rewire your cart so that it actually runs away from the light. Can you imagine how this might be done?

Just one more thought: if you use infrared photoresistors, you can control your cart with beams from infrared LEDs, in normal room lighting. If you and a couple of friends all have infrared transmitters, you can get your cart to run from one of you to the next, like an obedient dog.

This takes us about as far as I'm going to go into robotics. I urge you to check out the sites online if you want to pursue the topic further. You can also buy a wide variety of robot kits, although of course I feel that it's more fun to invent or develop things for yourself.

All that's left now is to perform one last introduction: to a device that should make your life much easier, even though the device is much more complicated than anything we have dealt with so far.

Experiment 34: Hardware Meets Software

Throughout this book, in accordance with the goal of *learning by discovery*, I have asked you to do an experiment first, after which I've suggested the general principles and ideas that we can learn from it. I now have to change that policy, because the next experiment involves so much setup that it's only fair to tell you what to expect before you begin the preparations.

We are about to enter the realm of controller chips, often known as MCUs, which is an acronym for micro controller unit. An MCU contains some flash memory, which stores a program that you write yourself. The flash memory is like the memory in a portable media player, or the memory card that you use in a digital camera. It needs no electricity to power it. In addition, the chip has a processor which carries out the instructions in your program. It has RAM to store the values of variables on a temporary basis, and ROM, which tells it how to perform tasks such as sensing a varying voltage input and converting it into digital form for internal use. It also contains an accurate oscillator, so that it can keep track of time. Put it all together, and it's a tiny computer that you can buy for under $5.

Let's suppose that you have a greenhouse where the temperature must never fall below freezing. You set up a temperature sensor, and you have two different heaters. You want to switch on the first heater if the temperature falls below 38° Fahrenheit. But if, for some reason, that heater is broken, you want to switch on the second, backup heater when the temperature goes below 36° Fahrenheit.

Programming a MCU to take care of this can be very simple indeed. You could even add extra features, such as a second temperature sensor, just in case the first one fails, and you could tell the chip to use whichever sensor gives a lower reading.

Another application for a MCU would be in a fairly elaborate security system. The chip could monitor the status of various intrusion sensors, and can take various preprogrammed steps, depending on the sensors' status. You could include delay intervals, too.

Many MCUs have additional useful features built in, such as the ability to control servo motors that turn to a specific angle in response to a stream of pulses. Servos are widely used in radio-controlled model boats, airplanes, and hobby robotics.

Perhaps you are now wondering why, if MCUs can do all this, haven't we been using them all along? Why did I spend so much time describing the development of an alarm system using discrete components, if one chip could have done everything?

There are three answers:

1. MCUs cannot do everything. They need other components to help them interact with the world, such as transistors, relays, sensors, and amplifiers. You need to know how those things work, so that you can make intelligent use of them.
2. MCUs can introduce their own kinds of problems and errors, associated with using software in addition to hardware. I'll have more to say about this later.
3. MCUs have limits and restrictions, most obviously their requirement for a 5-volt regulated power supply, and their inability to source or sink much current from each pin. They also demand that you learn a programming language (which differs from one brand of MCU to the next). And to get the program into the chip, you have to be able to plug it into a computer and do a download, which is not always convenient.

In this experiment, you'll learn how to write a program for a small and simple MCU, and you'll transfer the program into it and see how it works.

BACKGROUND

Origins of programmable chips

In factories and laboratories, many procedures are repetitive. A flow sensor may have to control a heating element. A motion sensor may have to adjust the speed of a motor. Microcontrollers are perfect for this kind of routine task.

A company named General Instrument introduced an early line of MCUs in 1976, and called them PICs, meaning Programmable Intelligent Computer—or Programmable Interface Controller, depending which source you believe. General Instrument sold the brand to another company named Microchip Technology, which owns it today.

"PIC" is trademarked, but is sometimes used as if it's a generic term, like Scotch tape. In this book, I've chosen a range of controllers based on the PIC architecture. They are licensed by a British company named Revolution Education Ltd., which calls its range of chips the PICAXE, for no apparent logical reason other than that it sounds cool.

I like these microcontrollers because they were developed originally as an educational tool and because they are very easy to use. They're cheap, and some of them are quite powerful. Despite their odd name, I think they're the best way to get acquainted with the core concepts of MCUs.

After you play with the PICAXE, if you want to go farther into MCUs, I suggest the BASIC Stamp (which uses a very similar language, but with additional powerful commands) and the very popular Arduino (which is a more recent design, packed with powerful features, but requires you to learn a variant of C language to program it). I'll say more about these chips later.

If you search for "picaxe" on Wikipedia, you'll find an excellent introduction to all the various features. In fact, I think it's a clearer overview than you'll get from the PICAXE website.

Supplies

Figure 5-122 shows some of the chips in the PICAXE range. I'll be telling you how to use the smallest—the 08M—which costs less than $5 and is cheaper than any other MCU that I've found. It has only 256 bytes of memory to store a program (not gigabytes, megabytes, or kilobytes, just 256 bytes!), but you'll be surprised how many possibilities this amount allows for. Figure 5-123 shows a closeup of the 08M with its legs safely embedded in a piece of conductive foam.

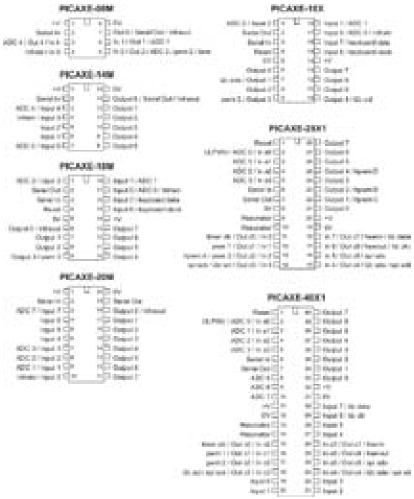

Figure 5-122. *A page from the PICAXE catalog lists only some of the chips that are available. What began as an educational aid has become a useful prototyping tool.*

In the United States, there are three distributors of this chip:

- *http://www.advancedmicrocircuits.com*
- *http://www.phanderson.com/picaxe*
- *http://www.sparkfun.com*

I like P. H. Anderson for its grass-roots hobbyist approach, and they have good prices if you want to buy multiple chips. But SparkFun Electronics offers other associated products that you may find interesting.

All the distributors will want to sell you "starter kits," such as the one in Figure 5-124, perhaps because the PICAXE itself is so cheap that it doesn't offer much of a profit margin. Still, for our purposes, you should buy the chip as a standalone item. And buy two of them, just in case you damage one (for example, by connecting voltage to it incorrectly).

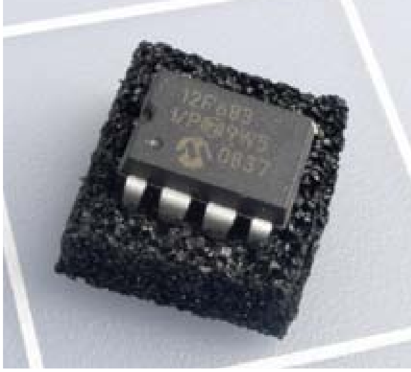

Figure 5-123. *When supplied by one of its American distributors, a PICAXE 08M arrives embedded in a little square of conductive foam. The chip is the same size as a 555 timer but has the power of a tiny computer.*

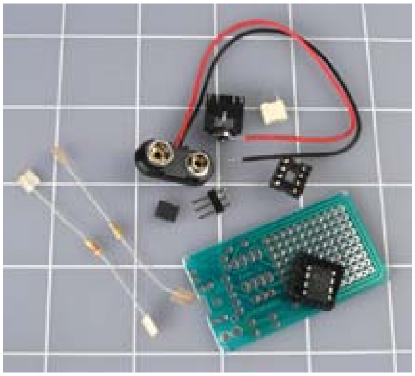

Figure 5-124. *A typical PICAXE kit includes a printed circuit board, which you may not really need, and some other not-entirely-essential items. But the 3.5-mm stereo jack socket (top, center) is absolutely necessary.*

To download your programming instructions into the chip, first you'll type the instructions on a computer, and then you'll feed them through a cable into the PICAXE memory. So you'll need to buy a cable, and you'll need software to help you to write the program.

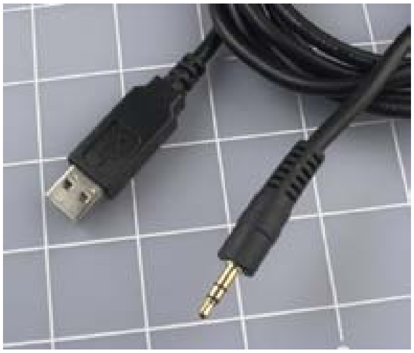

Figure 5-125. *The USB download cable made for use with the PICAXE terminates in a 3.5-mm audio plug. This should not be inserted in any audio device. It establishes a serial connection with a computer, allowing program code to be downloaded into the chip.*

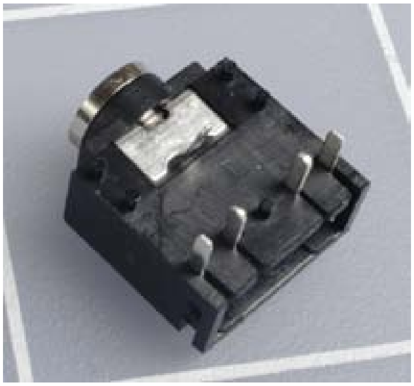

Figure 5-126. *Closeup of the 3.5-mm stereo socket that is used with the USB download cable.*

You can use the PICAXE with a serial cable, but I don't recommend it. The old RS-232 serial communications standard on PCs is pretty much obsolete, and PICAXE has recognized this by offering a USB cable (which contains a serial converter inside its plug). The USB cable is a little more expensive, but is also simpler and compatible with Apple computers. From any of the U.S. distributors, buy USB cable part AXE027, also sold as part PGM-08312 by *http://www.sparkfun.com* (quantity: 1). The cable is shown in Figure 5-125.

To write your software and send it down the wire to the chip, the PICAXE Programming Editor is the tool of choice. It comes in only a Windows version. For those who prefer Mac OS or Linux, you can get a free download of another piece of software known as AXEpad, which has fewer features, but will do the job. All the downloadable software is freely available from *http://www.rev-ed.co.uk/picaxe/software.htm*.

Finally, you need a 3.5-mm stereo audio socket with solder connections. The reason for this is that the manufacturers of the PICAXE have used a stereo audio plug on the free end of their USB cable, and you have to be able to plug it into something. The PICAXE breadboard adapter, SparkFun stock number DEV-08331, contains the necessary stereo socket in addition to a few other little items. Quantity: 1. See Figure 5-126.

Oddly enough, the USB cable is the most expensive item on the list, because of the electronics hidden inside it.

Software Installation and Setup

Now you have to go through a setup procedure. There is no way around this. Here is what you will be doing:

1. Install a driver so that your computer will recognize the special USB cable.
2. Install the Programming Editor software (or AXEpad for Mac/Linux) so that your computer will help you to write programs and then download them into the chip.
3. Mount the PICAXE on your breadboard and add the socket to receive downloads.

These steps are explained in the following sections.

The USB driver

Fair warning: If you go to the PICAXE website and try to use its search function, it probably won't find what you want. Search for "USB Driver," for instance, and it will pretend it has never heard of such a thing.

The PICAXE home page also has irritating drop-down menus that tend to disappear just when you're about to click on them, but at the time of writing, you can bypass these issues by going straight to the Software Downloads section at *http://www.rev-ed.co.uk/picaxe/software.htm*.

Scroll down past all the software until you get to Additional Resources. Look for the AXE027 PICAXE USB Download Cable. At first glance, it looks as if they want to sell you a cable, but in fact this is the list of drivers. Double-click the one appropriate to your computer, and choose a destination on your computer for the download—a place where you will find it easily, such as your desktop.

Be careful not to download the driver for the USB010 USB-Serial adapter by mistake. The USB-Serial adapter is something else entirely.

The download will leave you with a zipped file folder. You will have to unzip it. On Windows XP, right-click the folder and choose "Extract all." View the extracted files and you will find a PDF installation guide. Linux and Mac users can find instructions currently stashed at *http://www.rev-ed.co.uk/docs/AXE027.pdf*.

When installing the driver on a Windows platform, here are a few tips to minimize your exasperation level:

1. Remember, the special USB cable contains some electronics. It is not just a cable, but a device designed for interacting with a PICAXE chip. *Don't try to use it for anything else!*
2. You have to plug the cable into a USB port *before* you install the driver, because your computer will need to verify that the driver matches the cable.
3. You must not attach the PICAXE to the other end of the cable until *after* you have installed the driver.
4. Every USB port on your computer has a separate identity. Whichever one you choose when you first plug in the cable, you should use that port *every time in the future*. Otherwise, you will have to repeat the process of telling your computer what the cable is.
5. Bearing in mind Tip #4, you should avoid using the cable in a standalone USB hub.
6. The cable is fooling the PICAXE into thinking that it's talking to a serial port on your computer. Those "communication" ports are known as COM1, COM2, COM3, or COM4. When you install the driver, the installer will choose one of those COM ports for you, and later you will have to know which one it is. The PDF guide should help you through this procedure. Unfortunately, you cannot skip it.

The Programming Editor software

If you have come this far, you're ready for the next big step, which is much easier. You need the PICAXE Programming Editor, available for free on the Software Downloads web page where you found the USB driver. (If you are using a Mac or Linux, you will need AXEpad, which is on the same web page.)

Downloading and installing the Programming Editor should be simple and painless. Once you have done that, you should find that it has placed a shortcut on your desktop. Double-click it, go to View→Options, and in the window that opens (shown in Figure 5-127), click the Serial Port tab. You should see a dialog box like the one in Figure 5-128. Now make sure that the Programming Editor is looking at the same COM port that was chosen by the USB driver. Otherwise, the Programming Editor won't know where to find your PICAXE chip.

In the Programming Editor, go to View→Options and click the Mode tab, then click the button to select the 08M chip.

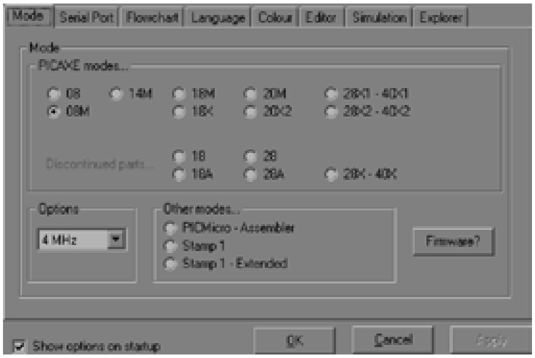

Mode

PICAXE modes...

○ 08 ○ 14M ○ 18M ○ 20M ○ 28X1 - 40X1
● 08M ○ 18X ○ 20X2 ○ 28X2 - 40X2

Discontinued parts... ○ 18 ○ 28
○ 18A ○ 28A ○ 28X - 40X

Options: 4 MHz

Other modes...
○ PICMicro - Assembler
○ Stamp 1
○ Stamp 1 - Extended

☑ Show options on startup

Figure 5-127. *This screenshot shows the options window of the PICAXE Program Editor, which you must use to select the type of chip that you intend to program (in our case, the 08M).*

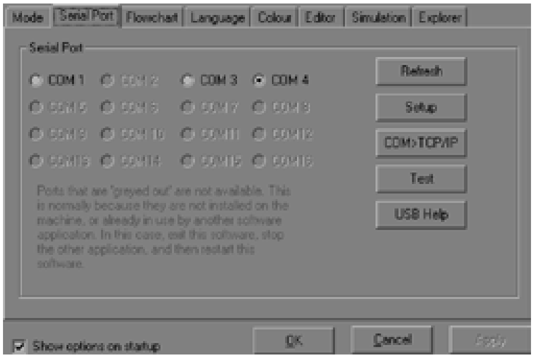

Serial Port

○ COM 1 ○ COM 2 ○ COM 3 ● COM 4
○ COM 5 ○ COM 6 ○ COM 7 ○ COM 8
○ COM 9 ○ COM 10 ○ COM11 ○ COM12
○ COM13 ○ COM14 ○ COM15 ○ COM16

Ports that are 'greyed out' are not available. This is normally because they are not installed on the machine, or already in use by another software application. In this case, exit this software, stop the other application, and then restart this software.

☑ Show options on startup

Figure 5-128. *Another screenshot of the options window shows the second essential choice that you must make: selection of the COM port that the installer chose on your computer.*

Are we having fun yet? Obviously not, but you're through with software hassles for the time being. The last step before you're ready to use the PICAXE is to mount it, and its socket, on your breadboard.

Setting up the hardware

The PICAXE 08M looks like a 555 timer. (Other chips in the PICAXE range have more pins and more features.) It requires a properly regulated 5 volts, just like the logic chips you dealt with previously. In fact, the PICAXE people are rather emphatic about protecting it from voltage spikes. They want you to use two capacitors (one 100 µF, one 0.1 µF) on either side of an LM7805 regulator. This seems like overkill, but the PICAXE is more inconvenient to replace than a 555 timer. You certainly can't run down to RadioShack to buy one. So let's do what the manufacturer says, just in case, and set up a breadboard as shown in Figures 5-129 and 5-130.

Now for the chip itself. Note that the pins for positive and negative power are *exactly opposite* to those for the 555 timer, so be careful!

Set up your breadboard following the schematic shown in Figure 5-131. Note that I am showing the stereo socket on its underside, because I think that's how you'll have to use it with the breadboard. If you try to stick its pins into the holes in the board, they will fit, but when you insert the plug into the socket, the thickness of the plug will tend to raise the socket up so that it loses contact. I really think the way to go is to solder wires to the pins on the socket and push the wires into the breadboard. See Figure 5-133.

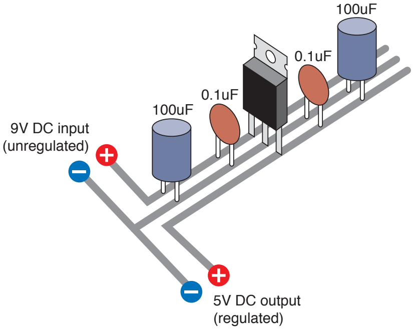

Figure 5-129. *PICAXE documentation specifies a 100 µF and 0.1 µF capacitor on the input side of a 5-volt regulator, and a similar pair of capacitors on its output side. On a breadboard, they can be arrayed like this.*

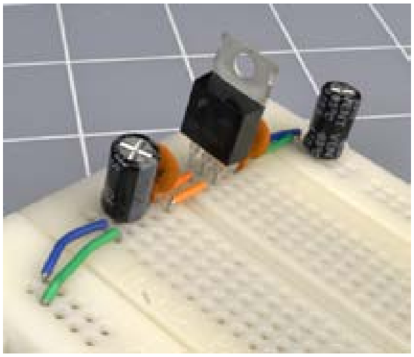

Figure 5-130. *The actual components for power regulation, applied to a breadboard, delivering 5 volts (positive and negative) down each side.*

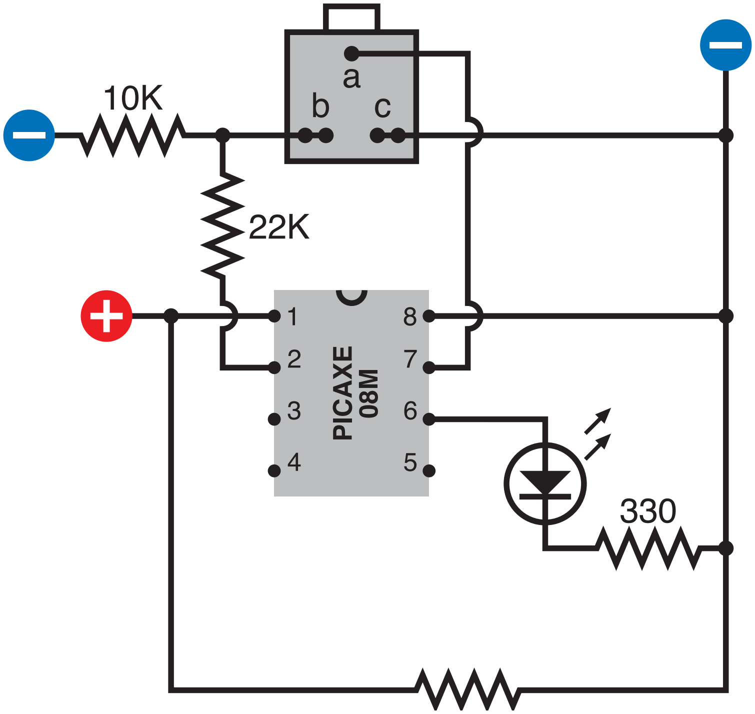

Figure 5-131. *The schematic of a test circuit for the PICAXE 08M shows the underside of the stereo socket, the essential 10K and 22K resistors on the input pin, and an LED to show an output from the chip.*

Be aware that the PICAXE manual shows things differently (although I have retained their labeling convention for the parts of the socket and the parts of the plug, identified as a, b, and c).

One little detail about the socket that is commonly supplied for use with the PICAXE: typically it has two *pairs* of contacts for the connections labeled b and c in the manual, and in my diagram. When you solder a connection, your solder joint should include *both* of the contacts in each pair, as shown in Figure 5-132.

Remember that the PICAXE must have 5 volts DC, and remember that your voltage regular will deliver this voltage reliably only if you give it extra voltage on its input side. If you provide it with 9 volts, that will provide a good amount of headroom.

The 22K and 10K resistors are essential for using the chip; see the following warning note for an explanation. My schematic also includes an LED and a 330Ω resistor, but they are needed only for the test that we'll be making momentarily.

Pin 2 Pull-Down

Always include the 22K resistor and the 10K resistor in the configuration shown in Figure 5-131. These resistors apply correct voltage to the serial connection, and when you're using the PICAXE on its own, they pull down the voltage on pin 2.

If pin 2 is left unconnected (floating), it may pick up random voltages, which the chip can misinterpret as a new program or other instructions, with unpredictable and undesirable results.

The 22K and 10K resistors should be regarded as permanent items accompanying your PICAXE regardless of whether you have it attached to your computer.

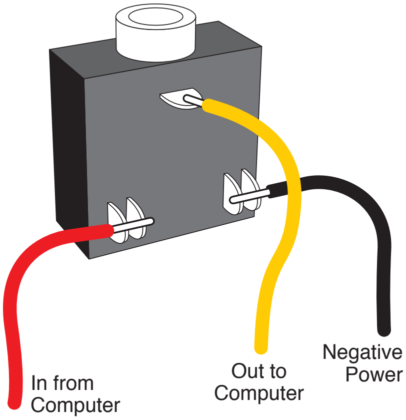

Figure 5-132. *Correct wiring of the socket is essential. When soldering wires to the lower terminals, make sure that you attach the wires to both of the terminals in each pair.*

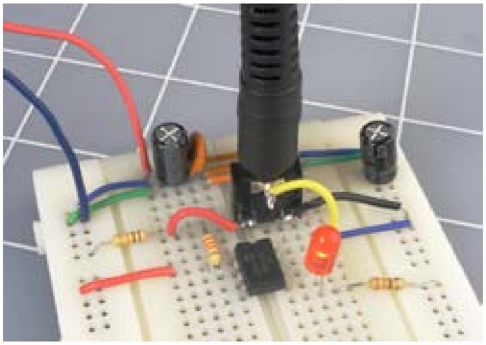

Figure 5-133. *The breadboard version of the test schematic, with the plug of the USB download cable inserted in the socket on the board. The PICAXE chip can now receive a downloaded program, and will immediately start to execute it.*

Verifying the Connection

Follow these steps carefully every time you want to program or reprogram your PICAXE chip:

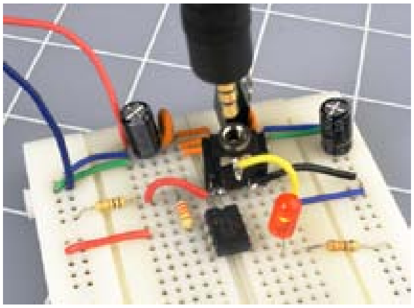

Figure 5-134. *After the program has been downloaded, the plug can be removed, and the program will continue to run, causing the LED to continue flashing.*

1. Insert the USB plug of your PICAXE cable into the same USB port that you used before.
2. Start the Programming Editor (or AXEpad if you are using a Mac OS or Linux).
3. In the Programming Editor, select View→Options to verify that the editor is using the right COM port and is expecting the 08M PICAXE chip.
4. Plug the stereo plug on the free end of the USB cable into the stereo socket that is now wired into your breadboard. See Figures 5-133 and 5-134.
5. Check your wiring, and then connect your power supply to the breadboard.
6. Click the button labeled “program” in the Program Editor window to tell the software to look for the PICAXE.

What If It Doesn’t Work?

The first thing to do is pull out the plug of the USB cable from the PICAXE breadboard, leaving the other end of the cable attached to your computer. Set your multimeter to measure DC volts, and attach its probes to sections b and c of the plug. See Figure 5-135. Now click the “program” button again, and your meter should show 5 volts briefly coming out of your computer to the plug on the end of the cable.

If you detect the voltage, the software is installed and working properly. In that case, there's a problem on your breadboard, either in the chip or in the wiring around it.

If you cannot detect any voltage, the software probably wasn't installed properly, or is looking for the wrong serial port. Try uninstalling it and reinstalling it.

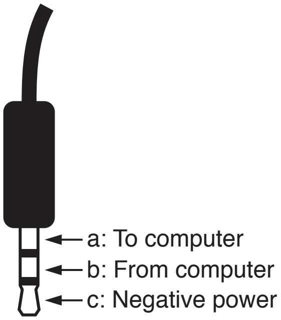
a: To computer
b: From computer
c: Negative power

Figure 5-135. *The stereo plug on the end of the USB download cable can be used for fault tracing. A multimeter set to measure DC volts can be attached to sections b and c of the plug to establish whether the Programming Editor is sending data through the serial connection.*

Your First Program

Finally you're ready to create your first program. Type the following code into the Programming Editor window:

```
main:
    high 1
    pause 1000
    low 1
    pause 1000
    goto main
```

Be sure to include the colon after the word "main" on the first line. See Figure 5-136 for a screenshot. The indents are created by pressing the Tab key. Their only purpose is to make program listings more legible. The software ignores them.

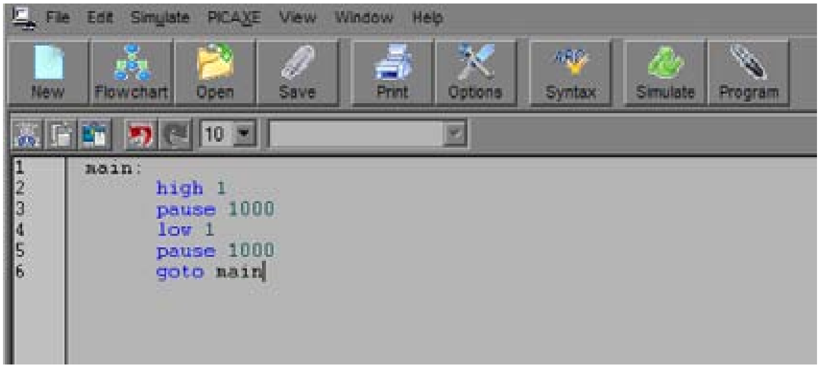
File
Edit
Simulate
PICAXE
View
Window
Help
New
Flowchart
Open
Save
Print
Options
Syntax
Simulate
Program
10

```
main:
        high 1
        pause 1000
        low 1
        pause 1000
        goto main
```

Figure 5-136. *This screenshot shows the first test program as it should be displayed by the Programming Editor (on a Windows computer).*

Click the Program button in the Programming Editor to download this program into the chip. As soon as the download is complete, the chip should start flashing the LED, lighting it for 1 second and then switching it off for 1 second. Figure 5-137 shows the steps that you should have followed to program the chip.

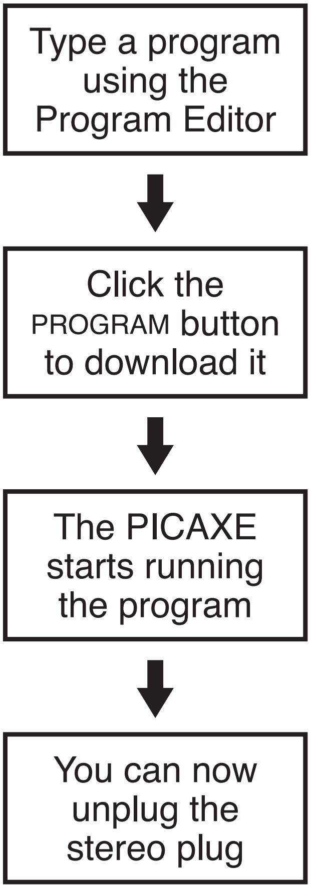
Type a program using the Program Editor
Click the PROGRAM button to download it
The PICAXE starts running the program
You can now unplug the stereo plug

Figure 5-137. *Four steps to create and run a program on the PICAXE controller chip.*

Now for the interesting part: disconnect the USB cable from the breadboard. The chip should continue flashing.

Disconnect the power supply from the breadboard and wait a minute or two for the capacitors to lose their charge. Reconnect the power, and the chip will start flashing again.

The program that you downloaded to the chip will remain in the memory inside the chip and will begin running every time power is applied to the chip.

The program doesn't use the usual pin numbers on the chip. It uses what I will call "logic pin numbers." Figure 5-138 shows how they are numbered. Figure 5-139 shows their multiple functions. I have put Logic Pin 0 in parentheses because its main purpose is to send data to the computer through the USB cable. It can do double duty as a digital output, but you have to disconnect it from the USB cable first. It's easy to forget to do this. It's a hassle that I prefer to avoid.

Decoding the Code

Let's take a look at the little program that you typed in. The first line identifies a section of the program. This program only has one section, and we're calling it "main." Any word with a colon after it is the name of a section of a program:

```
main:
```

The second line tells the chip to send a high output from Logic Pin 1:

```
high 1
```

The third line of the program tells the chip to wait for 1,000 milliseconds. This of course is the same as one second:

```
pause 1000
```

The fourth line tells the chip to change Logic Pin 1 back to its low state:

```
low 1
```

The fifth line tells the chip to wait for another 1,000 milliseconds:

```
pause 1000
```

The last line tells the chip to go back to the beginning of the "main" section:

```
goto main
```

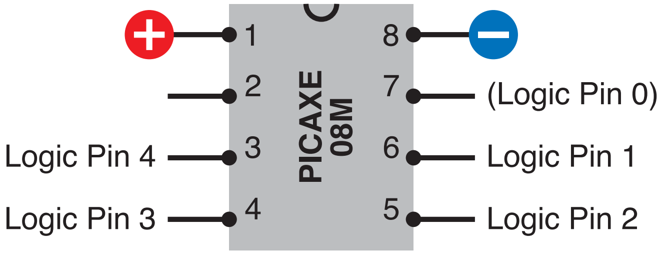
1
2
3
4
8
7
6
5
PICAXE
08M
Logic Pin 4
Logic Pin 3
(Logic Pin 0)
Logic Pin 1
Logic Pin 2

Figure 5-138. *The conventional pin numbers of the PICAXE chip are incompatible with the numbering system that is used in the PICAXE programming language. To minimize confusion, this guide refers to "Logic Pins" when using the numbering system that is required for programming the chip.*

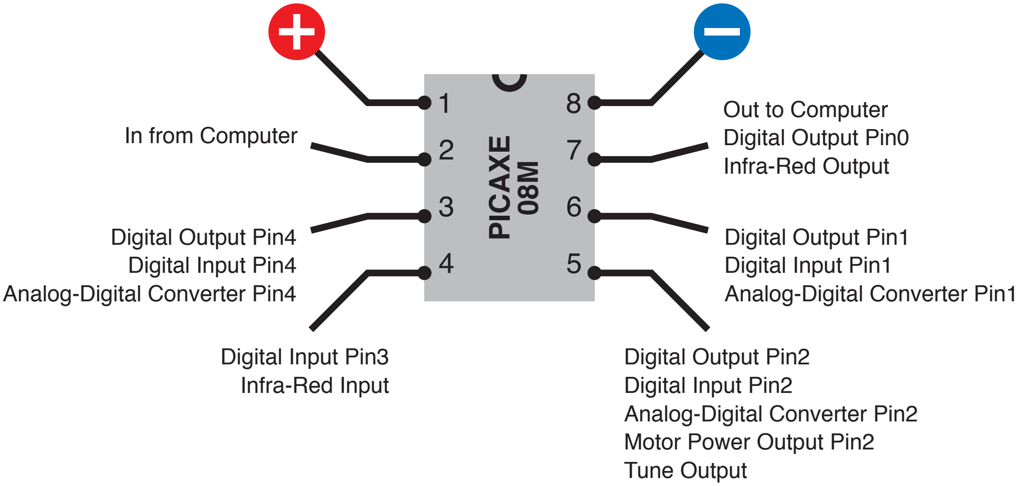
In from Computer
Digital Output Pin4
Digital Input Pin4
Analog-Digital Converter Pin4
Digital Input Pin3
Infra-Red Input
1
2
3
4
PICAXE
08M
8
7
6
5
Out to Computer
Digital Output Pin0
Infra-Red Output
Digital Output Pin1
Digital Input Pin1
Analog-Digital Converter Pin1
Digital Output Pin2
Digital Input Pin2
Analog-Digital Converter Pin2
Motor Power Output Pin2
Tune Output

Figure 5-139. *Many of the pins on the PICAXE 08M have multiple functions, which can be selected by appropriate program instructions.*

Editing

What if you want to change the program? No problem! Use the Programming Editor to change one of the lines in the program. Substitute 100 instead of 1000 milliseconds, for instance. (The `pause` command can be followed by any number up to 65535.) In your program, don't use a thousands comma in any of the numbers that you specify.

Plug the USB cable into the breadboard again, hit the Program button on the screen, and the new version of the program will be automatically downloaded to the chip, overwriting the old version.

What if you want to save the program for future use? Just go to the File menu in the Programming Editor and save the program onto your computer's hard drive. Because the PICAXE uses a variant of the BASIC computer language, it adds a *.bas* filename extension.

Simulation

If you make a simple typing error, the Programming Editor will find it and stop you from downloading your program. It will leave you to figure out how to fix the line that contains the error.

Even if all the statements in your program are correctly typed, it's a good idea to run a simulation of what they'll do, before you download them. This is easily done: click the "simulate" button on the menu bar of the Programming Editor. A new window will open, displaying a diagrammatic view of the PICAXE chip and showing you the states of its pins. (Note that if you use very short `pause` commands, the simulation won't run fast enough to display the time accurately.) A simulation screenshot is shown in Figure 5-140.

The >> button at the bottom-right corner of the simulation window will open up a list of all the variables in your program. So far, it doesn't have any variables, but it soon will. All the zeros on the righthand side are binary numbers, which you can ignore for now.

You'll need to check the second part of the PICAXE documentation, which contains all the programming statements and their correct syntax. At the time of writing, this is stored at http://www.rev-ed.co.uk/docs/picaxe_manual2.pdf.

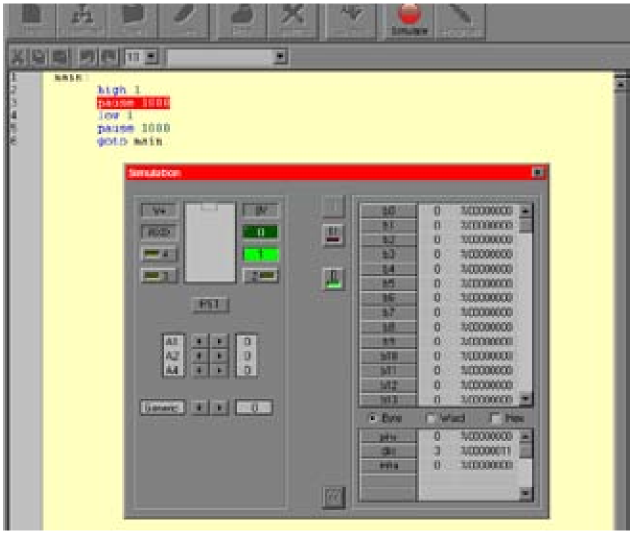

```
main:
	high 1
	pause 1000
	low 1
	pause 1000
	goto main
```

Figure 5-140. *This screenshot shows the simulation window that can be opened in the Program Editor to test program code before it is downloaded to the chip. The values of variables are shown in the section on the right. The pin states are shown on the left.*

Loops

Here's another thing I'd like you to try. Rewrite your program code as shown here and download it onto the PICAXE:

```
main:
    for b0 = 1 to 5
        high 1
        pause 200
        low 1
        pause 200
    next
    wait 2
    goto main
```

Note that `b0` is letter b followed by a zero, not letter b followed by letter O. The extra indents once again are added to make the listing easier to understand. The four lines beginning "high 1" and ending "pause 200" will be executed repeatedly. It's helpful to see them as a block.

Watch the light and see what happens. It should flash five times quickly, then wait for two seconds, and then repeat. You just added a *loop* to your program. You can use a loop if you want something to happen more than once.

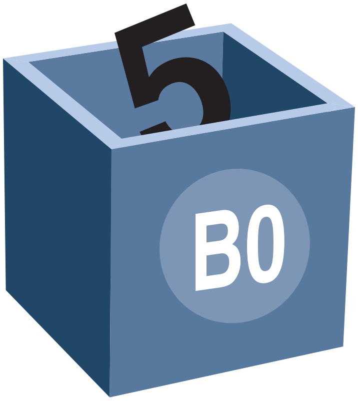

Figure 5-141. *To understand how a program works, visualize a variable as being like a "memory box" with its name on the outside and a number stored on the inside.*

b0 is known as a *variable*. Think of it as being like a little "memory box" with its name, b0, on a label on the outside. Figure 5-141 illustrates this concept. This particular memory box can contain any number from 0 through 255. The loop begins by telling the computer to put number 1 in the box, then process the remaining statements, until the word "next" sends the processor back to the first line, at which point it adds 1 to the contents of b0. If the value of b0 is 5 or less, the loop repeats. If the value is 6, the loop has run five times, so it's over, and the PICAXE skips down to the "wait 2" statement after "next." See Figure 5-142 for an annotated version of the program listing.

"Wait" is a PICAXE command that is measured in whole seconds, so "wait 2" waits for 2 seconds. Then "goto main" begins the procedure all over again.

If your flashing-light demo worked out as planned, it's time to take the next step and make the chip do something more useful.

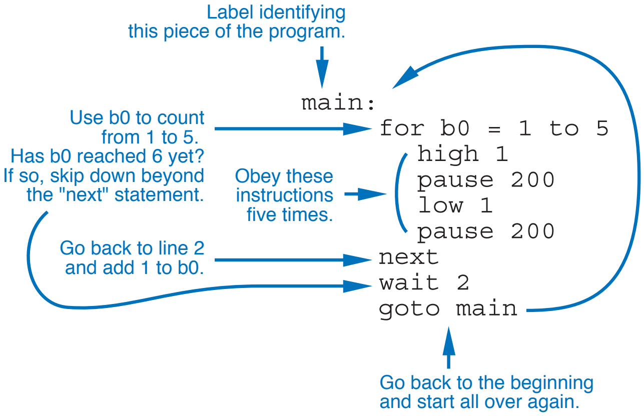

```
main:
for b0 = 1 to 5
  high 1
  pause 200
  low 1
  pause 200
next
wait 2
goto main
```

Label identifying this piece of the program.
Use b0 to count from 1 to 5. Has b0 reached 6 yet? If so, skip down beyond the "next" statement.
Obey these instructions five times.
Go back to line 2 and add 1 to b0.
Go back to the beginning and start all over again.

Figure 5-142. *The blue annotations explain what the program, on the right, is telling the PICAXE to do.*

FUNDAMENTALS

Basic PICAXE parameters

Here are some of the most useful parameters of the PICAXE:

- The PICAXE requires 5 volts DC, regulated.
- The inputs and outputs of the PICAXE are compatible with 5-volt logic chips. You can attach them directly.
- Each PICAXE pin can sink or source up to 20mA. The whole chip can deliver up to 90mA. This means that you can run LEDs directly from the pins, or a piezo noisemaker (which draws very little current), or a transistor.
- You can use a chip such as the ULN2001A Darlington array (mentioned in the previous experiment) to amplify the output from the PICAXE and drive something such as a relay or a motor.
- The chip executes each line of your program in about 0.1 milliseconds.
- The 08M chip has enough flash memory for about 80 lines of program code. Other PICAXE chips have more memory.
- The PICAXE provides 14 variables named b0 through b13. The "b" stands for "byte," as each variable occupies a single byte. Each can hold a value ranging from 0 through 255.
- No negative or fractional values are allowed in variables.
- You also have 7 double-byte variables, named w0 through w6. The "w" stands for "word." Each can hold a value ranging from 0 through 65535.
- The "b" variables share the same memory space as the "w" variables. Thus:
  - b0 and b1 use the same bytes as w0.
  - b2 and b3 use the same bytes as w1.
  - b3 and b4 use the same bytes as w2.
  - b5 and b6 use the same bytes as w3.
  - b7 and b8 use the same bytes as w4.
  - b9 and b10 use the same bytes as w5.
  - b11 and b12 use the same bytes as w6.
  - b13 and b14 use the same bytes as w7.

  Therefore, if you use w0 as a variable, do not use b0 or b1. If you use b6 as a variable, do not use w3, and so on.
- Variable values are stored in RAM, and disappear when the power is switched off.
- The program is stored in nonvolatile memory, and remains intact when the power is off.
- The manufacturer's specification claims that the nonvolatile memory is rewritable up to about 100,000 times.
- If you want to attach a switch or pushbutton to a pin and use it as an input, you should add a 10K pull-down resistor between the pin and the negative side of the power supply to hold the pin in a low state when the switch is open. Figure 5-143 shows how pull-down resistors should be used in conjunction with a SPST switch or a pushbutton.
- On the 08M chip, if you apply a varying resistance between Logic Pins 1, 2, or 4, and the negative side of the power supply, the chip can measure it and "decide" what to do. This is the "Analog-Digital Conversion" feature—which leads to our next experiment.

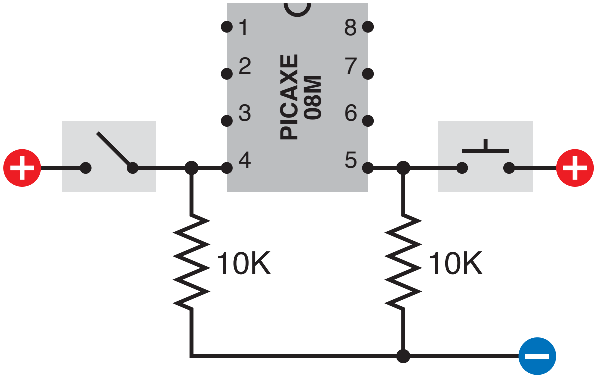

Figure 5-143. *The PICAXE can respond to the state of a switch or button attached to any of its input-capable pins. A 10K resistor must be used to pull down the state of the pin when the contact of the switch or button is open. Otherwise, you may get unpredictable results.*

Experiment 35: Checking the Real World

Often we want a microcontroller to measure something and respond in an appropriate way. For instance, it can measure a low temperature and sound an alarm, as I suggested in the example that I gave earlier.

The PICAXE has three analog-to-digital converters (ADCs) built in, accessible via logic pins 1, 2, and 4, as shown in Figure 5-139. The best way to use them is by applying a potential somewhere between 0 and 5 volts. In this experiment, I'll show you how to calibrate the response of the chip.

You will need:

- Trimmer potentiometer, 2K. Quantity: 1.
- PICAXE 08M chip and associated USB cable and socket. Quantity: 1 of each.

Procedure

Take the same trimmer potentiometer that you used in Experiment 32 and wire its center terminal to Logic Pin 2 of the PICAXE (which is hardware pin 5). The other two terminals of the 2K trimmer go to positive and to negative, respectively. So depending how you set the trimmer, the pin of the PICAXE is directly connected to positive (at one end of the scale), or directly connected to negative (at the other end of the scale), or somewhere in between. See Figure 5-144 for the revised schematic, and Figure 5-145 for a photograph of the breadboarded circuit.

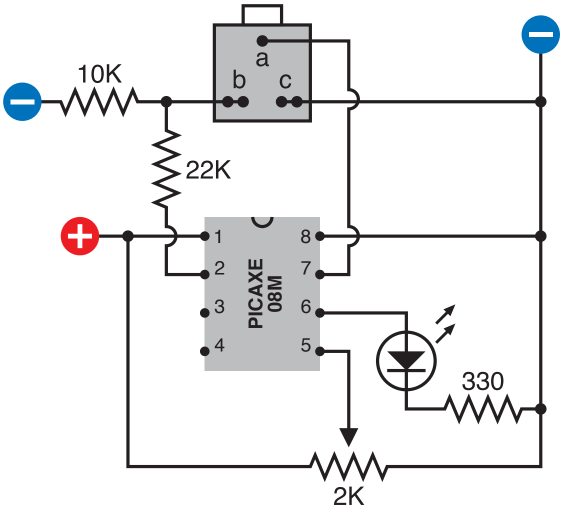

Figure 5-144. *This schematic, drawn in a layout suitable for breadboarding, shows how a 2K potentiometer can be used to apply a varying voltage to one of the pins of the PICAXE that is capable of converting an analog signal to a digital value.*

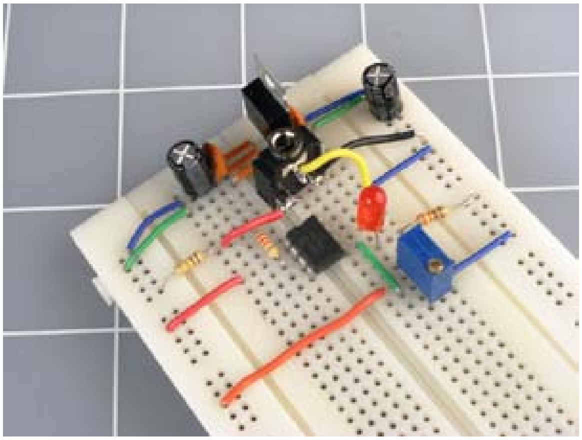

Figure 5-145. *The trimmer potentiometer added to the previously breadboarded circuit.*

Now we need a program to tell the chip what to do. Using the Programming Editor, start a new document. The code should look like this:

```
main:
    readadc 2,b0
    debug b0
    goto main
```

The command "readadc 2,b0" means "read the analog input on Logic Pin 2, convert from analog to digital, and store the result in b0."

The command "debug b0" tells the chip to go into program debugging mode, in which it uses its USB cable to tell the Programming Editor the values of all the variables while the program is running. The variables are displayed in a debugging window.

Download the program, and as the program starts to execute, the debugging window should open. Start adjusting the trimmer while looking at the value of b0, and you'll see b0 change its value.

You can make a table and draw a graph showing the relationship between the resistance between Logic Pin 2 and ground, and the value of b0. Just pull the trimmer off the breadboard, measure its resistance with a meter, then increase its resistance by, say, 200Ω, put it back into the breadboard, and look at the value of b0 again.

This is laborious, but calibrating equipment is always laborious—and in any case, I decided to do it for you. The graph is shown in Figure 5-146. You can also see the raw data numbers in the following table. I was pleased to find that the PICAXE gives a very precise, linear response to the input voltage. In other words, the graph is a straight line.

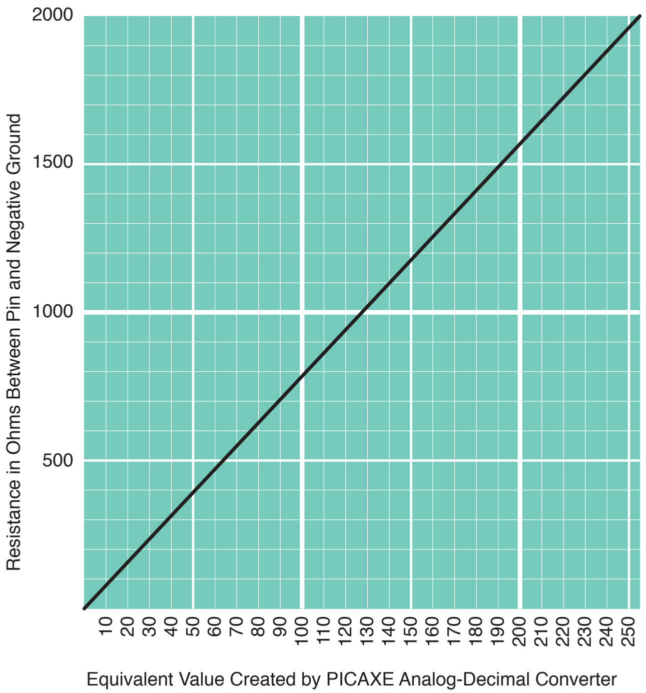

Figure 5-146. *When an ADC input pin is hooked up to a 2K potentiometer, which is connected across the same voltage that powers the chip, you should find that the resistance between the input pin and the negative side of the power supply generates the series of digital values shown on the graph. Note that the potentiometer must have a 2K value, and the power supply is assumed to be precisely 5 volts.*

This table shows measurements made with PICAXE 08M controller.

Resistance (in ohms) between the ADC pin and the negative supply	Equivalent digital value
2000	255
1900	243
1800	230
1700	218
1600	205
1500	192
1400	179
1300	166
1200	154
1100	141
1000	128
900	115
800	102
700	90
600	77
500	64
400	51
300	38
200	26
100	13
0	0

Now we can modify the program to make it do something with the information that it's taking in:

```
main:
    readadc 2,b0
    let w1 = 5 * b0
    high 1
    pause w1
    low 1
    pause w1
    goto main
```

Notice what's happening here. First we get a value in b0, and then on the next line, we do some arithmetic with it. The asterisk means "multiply." So the statement says, "Take whatever value is in b0, multiply by 5, and transfer it to another variable, w1." We have to use a w variable, because when we multiply the value of b0 by 5, we may get a number that is bigger than 255—too big to fit into a byte variable.

Finally, we take variable w1 and use it with a "pause" statement instead of a fixed number value. We're saying to the PICAXE, "pause for whatever number of microseconds you get by checking the value of w1."

So the software checks a variable resistance, turns it into a number, and applies that number to adjust the flashing speed of the LED.

Think back to the need of the cart powered by stepper motors. It was supposed to check two photoresistors, and adjust the speed of each motor accordingly. Well, this PICAXE program is a step in that direction. It can measure voltage on a pin and change the output frequency on another pin. If you had two PICAXE chips, you could wire each of them to a photoresistor and a motor. Then you could adjust the behavior of your cart by editing the second line in the program, where it converts the value of b0 to the value of w1 which will be used in the "pause" command to determine the number of pulses per second. Instead of multiplying by 5 you could multiply by 7 or whatever number gives you the result you need. This leads to an important conclusion: *a big advantage of a programmable chip is that you can make adjustments in software*.

Because the PICAXE 08M actually has more than one ADC input, and has three pins that can be used for output, you might wonder whether you could use just the one chip to control both motors in response to inputs from two sensors. The problem is that the three output pins on the 08M also function as the three ADC input pins. You'd do better to buy one of the more advanced PICAXE chips, such as the 18M, which has more pins to choose from. It uses the same basic set of programming instructions, and doesn't cost much more money.

Also, you should read the PICAXE documentation and look up the "pwmout" command, which is short for "pulse-width modulation output," but you can think of as meaning "power motor output." This is specifically intended to run stepper motors. It establishes an output frequency of pulses that will continue while the chip obeys other instructions in its program.

FUNDAMENTALS

Extra features

A complete guide to the 08M would fill a book of its own, and of course such books already exist (just search the books section of Amazon.com for keyword “picaxe”). But I’ll finish my introduction to the controller by listing some of its extra capabilities, leaving you to look them up and explore them. Then I’m going to suggest one last experiment.

Interrupts

The PICAXE 08M allows you to set one “interrupt.” This feature tells the chip to make a mental note that if a particular event occurs—such as a switch applying voltage to one pin—it should stop doing whatever else it was doing, and respond to the interruption.

Infrared

One pin on the PICAXE 08M can be used to receive infrared signals from a TV-style remote that you can buy from the same suppliers that sell the PICAXE itself. With an infrared sensor attached to the chip, you can issue commands remotely. If you want to build a remote-controlled robot, the chip is specifically designed with this in mind.

Servo motors

Every PICAXE chip has at least one pin that can send a stream of pulses to control a typical servo motor. On the 08M chip, it’s Logic Pin 2. The width of each pulse tells the motor how far to rotate from its center position before stopping. A 555 timer can send this stream, but the PICAXE makes it easier. You can search online for more information about servo motors, which are especially useful for applications such as steering model vehicles, adjusting the flaps on model airplanes, and actuating robots.

Music

The PICAXE has an onboard tone generator that can be programmed with a “tune” command to play tunes that you write using a simple code.

Alphanumeric input/output

The “kbin” programming command is available in the PICAXE models 20X2, 28X1 and 28X2, and 40X1 and 40X2. You can plug a standard computer keyboard into the chip, and it will read the keypresses. You can also attach alphanumeric displays, but these procedures are nontrivial. For instance, when you’re trying to figure out which key someone has pressed on a keyboard, your program has to contain a list of the special hexadecimal codes that the keyboard creates.

Pseudorandom number generation

All PICAXE models can generate pseudorandom numbers using a built-in algorithm. If you initialize the number generator by asking the user to press a button, and you measure the arbitrary time that this takes, you can seed the pseudorandom number generator with the result, and the pseudorandom number generator will have a less repeatable sequence.

Visit *http://www.rev-ed.co.uk/docs/picaxe_manual1.pdf* to learn more.

Experiment 36: The Lock, Revisited

The combination lock that I described in Experiment 20 is especially appropriate for a microcontroller, because it requires a series of operations that resemble a computer program. I'm going to show how this project can be redesigned using a PICAXE 08M, and then leave it to you to consider how some of the other projects in this book could be converted.

You will need:

- The same type of keypad and relay recommended in Experiment 20.
- A transistor or Darlington array to amplify the output from the PICAXE so that it can drive the relay.

Getting the User Input

Any of the input pins on the PICAXE can sense a switch closing. The trouble is that we only have three pins capable of doing this, and even the most advanced PICAXE chip has fewer than 10 such pins. So how can we attach a 10-key keypad to the 08M?

I have a suggestion: attach various resistors to the keypad, so that each key applies a different voltage to one of the ADC pins. Then use the ADC feature to convert the voltage to a number, and use a table of possible numbers to figure out which key is being pressed. This may not be the most elegant solution, but it works!

The keypad can be wired as shown in Figure 5-147. The asterisk key is still being used to supply power, as in the original experiment, while the pound key resets the relay at the end of your computing session, as before.

Current flows through a series of resistors, beginning with one that has a value of 500Ω. Because this is not a standard value, you will either have to make it by combining other resistors in series, or by presetting a trimmer potentiometer. After that, each button is separated from the next button by a 100Ω resistor. Finally, at the end of the chain, a 600Ω resistor separates the last button from the negative side of the power supply. Again, this is not a standard value, and you may have to use a trimmer.

Add up all the resistances and you have 2K, which is the range that the PICAXE wants us to use. When you press a button, you tap into the chain of resistances. Button 9 puts 600Ω between the PICAXE ADC pin and ground. Button 6 is 700Ω, button 3 is 800Ω, and so on. (You may prefer to lay out the buttons so that the resistance progresses in a more logical fashion. That's up to you. I chose to lay them out in the way that would be easiest to visualize on a keypad.)

Now look back at the ADC values that I supplied in the table on page 308. These are the values that you should get when you press various keypad buttons—but you cannot count on them being absolutely precise, because they may vary if your resistor values are not quite accurate, or if your power supply isn't exactly 5 volts. It's not safe to say, for example, that the PICAXE will deliver an ADC conversion value of precisely 77 when the resistance is 600Ω. It's safer to say that the value will be between 71 and 83. If we specify a range as shown in the following table, we have a much better chance of interpreting each button correctly.

Button number		Range
9	600	71–83
6	700	84–96
3	800	97–108
0	900	109–121
8	1000	122–134
5	1100	135–147
2	1200	148–160
7	1300	161–172
4	1400	173–185
1	1500	191–198

Suppose you attach the common pin of your keypad to ADC Logic Pin 2 of the PICAXE. You can now use the Program Editor to write a program that looks like this:

```
getkey:
    readadc 2,b0
    let b1 = 9
    if b0 < 84 then finish
    let b1 = 6
    if b0 < 97 then finish
    let b1 = 3
    if b0 < 109 then finish
    let b1 = 0
    if b0 < 122 then finish
    let b1 = 8
    if b0 < 135 then finish
    let b1 = 5
    if b0 < 148 then finish
    let b1 = 2
    if b0 < 161 then finish
    let b1 = 7
    if b0 < 173 then finish
    let b1 = 4
    if b0 < 186 then finish
    let b1 = 2

finish:
    return
```

What does the word "return" mean at the end? I'll get to that in a second. I want to explain the rest of the routine first.

b0 receives the value supplied by the analog-digital converter when it looks at the keypad. After storing the number in b0, the routine has to figure out which keypad key it matches. The key identity (0 through 9) will be stored in another variable, b1.

The program starts by assigning value 9 to b1. Then it checks to see whether b0 < 84. This means "if b0 is less than 84." If it is, then the routine tells the PICAXE to "finish," which means "jump to the finish: label." But if b0 is *not* less than 84, by default the PICAXE continues on to the next line, which makes a second attempt at guessing which key has been pressed. It assigns number 6 to b1. Now there's another if-then test—and so on. This process of reassigning values to b1 stops only when it gets to the point where b0 is greater than a number in the table.

If you're familiar with other dialects of BASIC, this may seem a bit laborious to you. You may wonder why we can't use a statement such as this:

```
    if b0 > 70 and b0 < 84 then b1 = 9
```

The answer is that PICAXE BASIC isn't sufficiently sophisticated to allow this. An if-then statement has to result in a jump to another section of the program. That's the only permitted outcome.

If you don't have any prior programming experience, the routine may still seem laborious to you, and perhaps a bit puzzling, too. This is understandable, because you're getting a crash course in software design without any formal preparation. Still, the PICAXE Programming Editor can be a big help, because it has its simulation feature. Before you can use this, though, you have to precede the routine that I just supplied with a control routine that you must type above it. The screenshot in Figure 5-148 shows you how it should look.

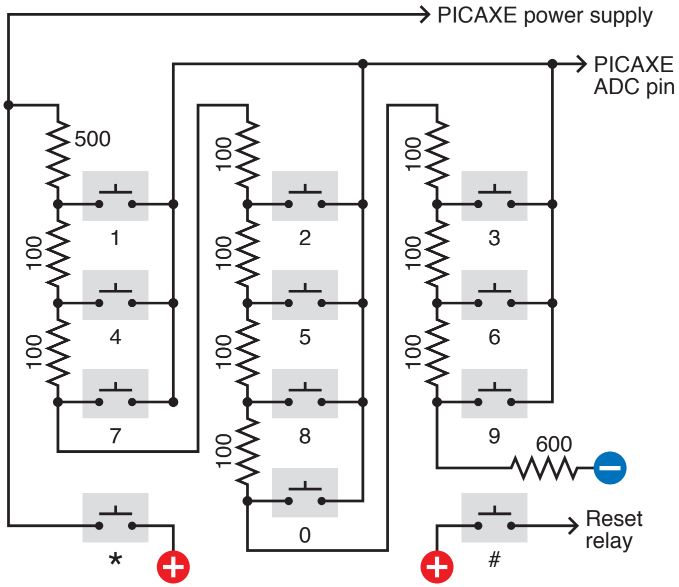

Figure 5-147. *A quick and simple way of attaching a keypad to provide numeric input to the PICAXE uses a chain of resistors totalling 2,000Ω. When a button is pressed, it connects the ADC input pin to a point in the chain. The resistance detected by the input pin can then be converted by the program in the chip to determine which key has been pressed.*

I have chosen an arbitrary combination of 7-4-1 for our combination lock. Using these numbers, the main section of the program looks like this:

```
main:
    low 1
    gosub getkey
    if b1<>7 then main
    gosub getkey
    if b1<>4 then main
    gosub getkey
    if b1<>1 then main
    high 1
    end
```

I should explain that the <> pair of symbols mean "is not equal to." So the fourth line of the program means, "if b1 is not equal to 7."

The value of b1 is supposed to be 7 if the user is putting in the correct combination. So if it's not 7, the user has entered the wrong value, and the if-then statement sends the PICAXE back to the beginning. In fact anytime the user inputs a number that is not in the correct 7-4-1 sequence, the program sends the PICAXE back to the beginning. This is the way the pure-hardware version of this experiment was set up.

But what is this word "gosub"? It means "go to a subroutine." A subroutine is any sequence of program statements that ends with the instruction to "return." So "gosub getkey" tells the PICAXE to mark its current place in the program while it skips to the `getkey:` section of code, which it obeys, until it finds the word "return," which returns it to the place from where it came.

The PICAXE continues in this fashion until it reaches the word "end." I had to insert the word "end" because otherwise the PICAXE will continue executing the program and will fall into the subroutine. "End" stops it from doing so. Figure 5-148 shows a screenshot of the complete listing.

So—is that all? Yes, that's it. If you enter the code into the Programming Editor exactly as I have supplied it, you should be able to run it in simulation mode, and in the simulation window, click the right-arrow beside Logical Pin A2 to increase its value in steps. Each time you pass one of the values in the `getkey:` subroutine, you should see the value for variable b1 change in the display.

This is really all you need to perform the functions of the combination lock. When the PICAXE runs this program, it waits for the correct combination. If it receives the combination, it sends the output from logical pin 1 high; otherwise, logical pin 1 stays low.

The only additional item you need is a transistor or CMOS gate between logical pin 1 and the relay that unlocks the computer, because the PICAXE cannot deliver enough current to operate the relay by itself.

Putting this procedure into a controller chip not only simplifies the circuit, but offers another advantage: you can change the combination simply by rewriting the program and downloading the new version into the chip.

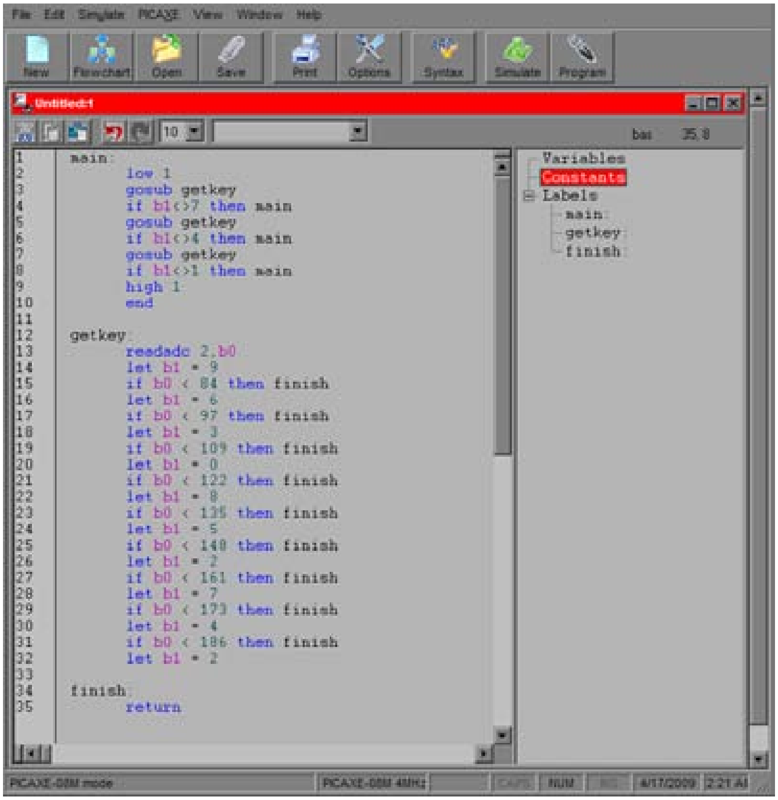

```
main:
        low 1
        gosub getkey
        if b1<>7 then main
        gosub getkey
        if b1<>4 then main
        gosub getkey
        if b1<>1 then main
        high 1
        end

getkey:
        readadc 2,b0
        let b1 = 9
        if b0 < 84 then finish
        let b1 = 6
        if b0 < 97 then finish
        let b1 = 3
        if b0 < 109 then finish
        let b1 = 0
        if b0 < 122 then finish
        let b1 = 8
        if b0 < 135 then finish
        let b1 = 5
        if b0 < 148 then finish
        let b1 = 2
        if b0 < 161 then finish
        let b1 = 7
        if b0 < 173 then finish
        let b1 = 4
        if b0 < 186 then finish
        let b1 = 2

finish:
        return
```

Figure 5-148. *This screenshot shows the complete listing of a program to read a sequence of three keypresses in conjunction with a combination lock. If the sequence is correct, the PICAXE sends a high output from one of its pins. If the sequence is incorrect, the program loops back to the beginning.*

FUNDAMENTALS

Limitations of MCUs

The PICAXE does have some disadvantages. Its voltage requirements alone restrict you from using it with the kind of freedom of a 555 timer.

Also, although I can get an instant result by plugging a 555 timer into a breadboard and adding a couple of resistors and a couple of capacitors, the PICAXE requires me to add a download socket, hook it up to my computer, write a program in the Programming Editor, and download the program.

Some people don't like writing software, or they have difficulty thinking in the relentlessly left-brain way that computer programming requires. They may prefer the hands-on process of assembling hardware.

Other people may have the opposite preference. This of course is a matter of taste, but one thing we know beyond all doubt is that computer programs often contain errors that may not reveal themselves until weeks or months later.

The PICAXE, for instance, doesn't protect you if a number is assigned to a variable that exceeds the limit for that type of variable. Suppose b1=200 and b2=60 and your program tells the PICAXE:

```
    let b3 = b1 + b2
```

The result should be 260, but byte-size variables can only count up to 255. What happens? You will find that b3 acquires a value of 4, without any warning or explanation. This is known as an "overflow error," which can be very difficult to predict, because it happens at runtime, when external factors are in control. The code looks perfectly good; the Programming Editor doesn't find any syntax errors; the simulation behaves properly. But in the real world, days or even months later, an unexpected set of circumstances results in an input that causes the overflow, and because the code is residing inside the chip at this point, you may have a hard time figuring out what on earth went wrong.

Software has its problems. Hardware has its advantages.

FUNDAMENTALS

Unexplored territory

If you've taken the time to complete most of the projects in this book with your own hands, you have gained a very rapid introduction to the most fundamental areas of electronics.

What have you missed along the way? Here are some topics that remain wide open for you to explore. Naturally you should search online if they interest you.

The informal, learning-by-discovery approach that I have used in this book tends to be light on theory. I've avoided most of the math that you'd be expected to learn in a more rigorous course on the subject. If you have mathematical aptitude, you can use it to gain a much deeper insight into the way in which circuits work.

I didn't deal much with computer architecture, either. We didn't go very far into binary code, and you didn't build a half-adder, which is a great way to learn how computers function on the most fundamental level. Perhaps you should think about assembling one.

I avoided going deeply into the fascinating and mysterious properties of alternating current. Here again, some math is involved, but just the behavior of current at high frequencies is an interesting topic in itself.

For reasons already stated, I avoided surface-mount components—but you can still go into this area yourself for a relatively small investment, if you like the idea of creating fascinatingly tiny devices. This may be the future of hobby electronics, so if you stick with it, you'll probably end up in the world of surface-mount.

Vacuum tubes were not mentioned, because at this point, they are mainly of historical interest. But there's something very special and beautiful about tubes, especially if you can enclose them in fancy cabinetwork. In the hands of a skilled craftsperson, tube amplifiers and radios become art objects.

I didn't show you how to etch your own printed circuit boards. This is a task that appeals to only certain people, and the preparation for it requires you to make very neat drawings or use computer software for that purpose. If you happen to have those resources, you might want to do your own etching. It could be a first step toward mass-producing your own devices.

I didn't cover static electricity at all. High-voltage sparks don't have any practical applications, and they entail some safety issues—but they are stunningly impressive, and you can easily obtain the necessary information to build the equipment. Maybe you should try.

Other Controllers

If you want something more powerful, a BASIC Stamp is the logical next step after a PICAXE. the BASIC Stamp is so called because it originally looked like a postage stamp. The BASIC Stamp has a larger vocabulary of commands and a bigger range of add-on devices (including displays with graphical capability, and a little keyboard that is specifically designed for use with the controller). The BASIC Stamp is shown in Figure 5-149.

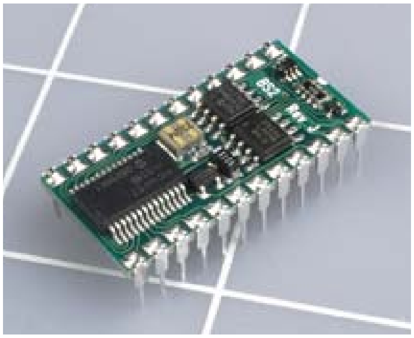

Figure 5-149. *The BASIC Stamp controller consists of surface-mounted components on a platform that has pins spaced at 1/10-inch intervals, for insertion in a breadboard or perforated board. This component uses a version of BASIC that is similar to the programming language of the PICAXE, but has many more extensions. The BASIC Stamp is available for use with a wide range of peripheral devices, including many alphanumeric dot-matrix displays.*

On the downside, you'll find that everything associated with the BASIC Stamp is a bit more expensive than in the PICAXE world, and the download procedure isn't quite as simple.

One of the more recent developments in the world of MCUs is the Arduino, which is both sophisticated and powerful. It does require programming in the C language. This language is a little more difficult to understand, and has only the vaguest similarity to the syntax that is used in the PICAXE and BASIC Stamp. On the other hand, because C dominates the larger world of computing, learning it might not be such a bad idea—and the Arduino offers some truly amazing capabilities. Because it is so popular, there are also many software tools, documentation, user forums, and many enthusatic hobbyists to help you. Two other Make: Books titles that I mentioned previously, *Getting Started with Arduino* and *Making Things Talk*, provide a great introduction.

In Closing

I believe that the purpose of an introductory book is to give you a taste of a wide range of possibilities, leaving you to decide for yourself what you want to explore next. Electronics is ideal for those of us who like to do things ourselves, because almost any application—from robotics, to radio-controlled aircraft, to telecommunications, to computing hardware—allows opportunities that we can explore at home, with limited resources.

As you delve deeper into the areas of electronics that interest you most, I trust you'll have a satisfying learning experience. But most of all, I hope you have lots of fun along the way.

Online Retail Sources and Manufacturers

A

This appendix contains URLs for companies mentioned as retail sources or manufacturers, along with the commonly used name of the source and the company name.

Colloquially used name	Actual corporate or company name	URL
3M	Minnesota Mining and Manufacturing Co.	*http://solutions.3m.com/en_US/*
Ace Hardware	Ace Hardware Corporation	*http://www.acehardware.com*
Advanced Micro Circuits	Advanced Micro Circuits Corp	*http://www.advancedmicrocircuits.com*
Akro-Mils	Myers Industries, Inc.	*http://www.akro-mils.com*
Alcoswitch	Division of Tyco Electronics Corporation	*http://www.tycoelectronics.com/catalog/menu/en/18025*
All Electronics	All Electronics Corporation	*http://www.allelectronics.com*
All Spectrum Electronics	All Spectrum Electronics	*http://www.allspectrum.com*
All-Battery.com	Tenergy Corporation	*http://www.all-battery.com*
Alpha potentiometers	Alpha Products Inc.	*http://www.alphapotentiometers.net*
ALPS pushbutton	ALPS Electric Co., Ltd.	*http://www.alps.com*
Amazon	Amazon.com, Inc.	*http://www.amazon.com*
Amprobe	Amprobe Test Tools	*http://www.amprobe.com*
Arduino	No corporate identity	*http://www.arduino.cc*
ArtCity	ArtCity	*http://www.artcity.com*
AutoZone	AutoZone, Inc.	*http://www.autozone.com*
Avago	Avago Technologies	*http://www.avagotech.com*
BASIC Stamp	Brand owned by Parallax, Inc.	*http://www.parallax.com*
BI Technologies	BI Technologies Corporation	*http://www.bitechnologies.com*
BK Precision	B&K Precision Corp.	*http://www.bkprecision.com*
Bussmann fuses	Cooper Bussman, Inc.	*http://www.cooperbussmann.com*
C&K switch	CoActive Technologies, Inc.	*http://www.ck-components.com*
Chicago lighting	CML Innovative Technologies	*http://www.cml-it.com*
CraftAmerica	Cardinal Enterprises	*http://www.craftamerica.com*
Darice	Darice Inc.	*http://www.darice.com*

Colloquially used name	Actual corporate or company name	URL
DeWalt	DeWalt Industrial Tool Company	*http://www.dewalt.com*
Digi-Key	Digi-Key Corporation	*http://www.digikey.com*
Directed switches	Directed Electronics Inc.	*http://www.directed.com*
Doctronics	Doctronics Educational Publishing	*http://www.doctronics.co.uk*
eBay	eBay Inc.	*http://www.ebay.com*
Elenco	Elenco Electronics Inc.	*http://www.elenco.com*
Everlight	Everlight Electronic Co. Ltd.	*http://www.everlight.com*
Extech	Extech Instruments Corporation	*http://www.extech.com*
Fairchild	Fairchild Semiconductor Incorporated	*http://www.fairchildsemi.com*
FTM	FTM Incorporated	*http://thefabricatorssource.com*
Fujitsu	Fujitsu America, Inc.	*http://www.fujitsu.com/us/*
GB wire strippers	Gardner Bender Inc.	*http://www.gardnerbender.com*
Hobbylinc	Hobbylinc Hobbies	*http://www.hobbylinc.com*
Home Depot	Homer TLC, Inc.	*http://www.homedepot.com*
Ideal wire strippers	Ideal Industries Inc.	*http://www.idealindustries.com*
Jameco	Jameco Electronics	*http://www.jameco.com*
K&J Magnetics	K&J Magnetics Inc.	*http://www.kjmagnetics.com*
Kingbright	Kingbright Corporation	*http://www.kingbrightusa.com*
Kobiconn	No web page found; use mouser.com	
KVM Tools	KVM Tools Inc.	*http://www.kvmtools.com*
Lowe's hardware	LF, LLC	*http://www.lowes.com*
Lumex	Lumex Inc.	*http://www.lumex.com*
McMaster-Carr	McMaster-Carr Supply Company	*http://www.mcmaster.com*
Megahobby	Megahobby.com	*http://www.megahobby.com*
Meter Superstore	Division of SRS Market Solutions Inc.	*http://www.metersuperstore.com*
Michaels craft stores	Michaels Stores, Inc.	*http://www.michaelscrafts.com*
Mill-Max	Mill-Max Manufacturing Corp.	*http://www.mill-max.com*
Mitutoyo	Mitutoyo America Corporation	*http://www.mitutoyo.com*
Motorola	Motorola, Inc.	*http://www.motorola.com/us*
Mouser electronics	Mouser Electronics, Inc.	*http://www.mouser.com*
Mueller alligator clip	Mueller Electric Company	*http://www.muellerelectric.com*
Newark	Subsidiary of Premier Farnell plc	*http://www.newark.com*
NKK switches	Nihon Kaiheiki Industry Co. Ltd.	*http://www.nkkswitches.com*
NXP semiconductors	NXP Semiconductors	*http://www.nxp.com*
Omron	Omron Corporation	*http://www.omron.com*
On Semiconductor	Semiconductor Components Industries, LLC	*http://www.onsemi.com*
Optek	Subsidiary of TT Electronics plc	*http://www.optekinc.com*
Panasonic	Panasonic Electric Works Corporation	*http://pewa.panasonic.com*

Colloquially used name	Actual corporate or company name	URL
PanaVise	Panavise Products, Inc.	*http://www.panavise.com*
Parallax	Parallax, Inc.	*http://www.parallax.com*
Pep Boys	Pep Boys-Manny, Moe and Jack	*http://www.pepboys.com*
Philips	Koninklijke Philips Electronics N.V.	*http://www.usa.philips.com*
PICAXE	Revolution Education Ltd.	*http://www.rev-ed.co.uk*
Piedmont Plastics	Piedmont Plastics, Inc.	*http://www.piedmontplastics.com*
Plano storages boxes	Plano Molding Company	*http://www.planomolding.com*
Pomona test equipment	Pomona Electronics Inc.	*http://www.pomonaelectronics.com*
RadioShack	RadioShack Corporation	*http://www.radioshack.com*
Sears	Sears Brands, LLC	*http://www.sears.com*
SparkFun Electronics	Sparkfun Electronics	*http://www.sparkfun.com*
Stanley tools	The Stanley Works	*http://www.stanleytools.com*
STMicroelectronics	STMicroelectronics Group	*http://www.st.com*
Texas Instruments	Texas Instruments Incorporated	*http://www.ti.com*
Tower Hobbies	Tower Hobbies	*http://www.towerhobbies.com*
Twin Industries	Twin Industries	*http://www.twinind.com*
Tyco	Tyco Electronics Corporation	*http://www.tycoelectronics.com*
Vaughan	Vaughan & Bushnell Mfg.	*http://hammernet.com/vaughan/*
Velleman keyboards	Velleman nv	*http://www.velleman.eu*
Vishay	Vishay Intertechnology Inc.	*http://www.vishay.com*
Wal-Mart	Wal-Mart Stores, Inc.	*http://www.walmart.com*
Weller	Division of Cooper Industries, LLC	*http://www.cooperhandtools.com/brands/weller/*
X-Acto	Division of Elmer's Products, Inc.	*http://www.xacto.com*
Xcelite	Division of Cooper Industries, LLC	*http://www.cooperhandtools.com/brands/xcelite/*
Xytronic	Xytronic Industries Ltd.	*http://www.xytronic-usa.com*

Acknowledgments

My association with MAKE magazine began when its editor, Mark Frauenfelder, asked me to write for it. I have always been very grateful to Mark for his support of my work. Through him I became acquainted with the exceptionally capable and motivated production staff at MAKE. Gareth Branwyn eventually suggested that I might like to write an introductory guide to electronics, so I am indebted to Gareth for initiating this project and supervising it as my editor. After I wrote an outline in which I described my idea for "Learning by Discovery" and the associated concept that cutting open components or burning them up can be an educational activity, MAKE's publisher, Dale Dougherty, uttered the memorable phrase, "I want this book!" Therefore I offer special thanks to Dale for his belief in my abilities. Dan Woods, the associate publisher, was also extremely supportive.

The production process was swift, competent, and painless. For this I thank my editor at O'Reilly, Brian Jepson; senior production editor Rachel Monaghan; copyeditor Nancy Kotary; proofreader Nancy Reinhardt; indexer Julie Hawks; designer Ron Bilodeau; and Robert Romano, who tweaked my illustrations. Most of all I am indebted to Bunnie Huang, my technical advisor, who reviewed the text in detail and knows a bunch of stuff that I don't know. Any residual errors are still my fault, even though I would prefer to blame them on Bunnie.

Thanks also to Matt Mets, Becky Stern, Collin Cunningham, Marc de Vinck, Phillip Torrone, Limor Fried, John Edgar Park, John Baichtal, and Jonathan Wolfe for helping out with some last-minute project testing.

Lastly I have to mention the genius of John Warnock and Charles Geschke, founders of Adobe Systems and creators of the very beautiful PostScript language, which revolutionized all of publishing. The horror of attempting to create this book using graphic-arts tools from…some other company…is almost unimaginable. In fact, without Illustrator, Photoshop, Acrobat, and InDesign, I doubt I would have attempted the task. I am also indebted to the Canon 1Ds with 100mm macro lens, which took many of the pictures in this book.

No free samples or other favors were received from any of the vendors mentioned herein, with the exception of two sample books from MAKE, which I read to ensure that I was not duplicating anything that had already been published.

Index

Numbers

2N2222 transistor, 73
  (see also Experiments 11 and 15)
2N6027 programmable unijunction transistor (see Experiment 14: A Pulsing Glow)
3M, 319
  breadboard, 39
  single inline sockets and headers, 103
9-volt battery, 5
12-key numeric keypad, 151
40xxB chip, 191
40xx chip, 191
74ACTxx chip, 191
74ACxx chip, 191
74AHCTxx chip, 191
74AHCxx chip, 191
74HC00 chip (see Experiment 22: Flipping and Bouncing)
74HC00 quad 2-input NAND chip (see Experiment 19: Learning Logic)
74HC02 chip (see Experiment 22: Flipping and Bouncing)
74HC04 chip, 149
  (see also Experiment 20: A Powerful Combination)
74HC08 quad 2-input AND chip (see Experiments 19 and 20)
74HC32 chip (see Experiment 21: Race to Place)
74HCTxx chip, 191
74HCxx chip, 191
74LS06 open-collector inverter chip, 150
74LS27 chip (see Experiment 23: Nice Dice)
74LS27 triple-input NOR chip, 150
74LS92 chip, 216
74LS92 chip (see Experiment 23: Nice Dice)
74LS92 counter chip, 150
74LVxx chip, 191
555 timer chip, 150
  alternating current, 248
  astable mode, 164–165
  audio frequencies, 256
  bistable configuration, 176–177
  frequency in astable mode, 166
  history, 160
  limits, 161
  monostable mode, 158–159
  powering up in bistable mode, 225
  pulse duration, 157
  reset voltage, 156
  square wave, 258
  trigger voltage, 156
  usefulness, 161
  (see also Experiments 16-24, 29, 32-33)
4026 chip, 150
  clock disable, 175
  (see also Experiment 18: Reaction Timer)
7400 family of integrated circuits, 192

A

ABS (acrylonitrile butadiene styrene)
  cutting, 270
  (see also Experiment 32: A Little Robot Cart)
AC adapter, 39
  modifying, 111
AC adapter plugs, 55–56
AC (alternating current), 12
Ace Hardware, 319
  wire, 101
Advanced Micro Circuits, 319
Akro-Grids, 229
Akro-Mils, 229, 319
alarms
  installation, 145–146
  intrusion (see Experiments 11, 15, and 24)
  magnetic sensor switches, 128
Alcoswitch, 319
  pushbutton, 42
All-Battery.com, 319
All Electronics, xii, 42, 319
  12-key numeric keypad, 151
  alligator clips, 4
  battery holders and connectors, 4
  desoldering wick, 98
  jumper wire assortment, 41
  patch cords, 41
  project boxes, 102
  solder, 100
  soldering stand, 98
  solder pump, 98
  switches, 42

alligator clips, 4
  copper, 101
All Spectrum Electronics, 319
  LEDs, 4
Alpha potentiometers, 4, 319
ALPS pushbutton, 319
alternating current, 12, 248
Amazon, 319
  deburring tool, 99
  electronic solder, 100
  GB Automatic Wire Strippers, 40
  pick and hook set, 99
amperage, 10
Ampère, André-Marie, 13
amperes, 36
  basics, 11
amplification, 87
amplifiers, 49
Amprobe, 319
amps, 11
Arduino, 319
  microcontrollers, 227
ArtCity, 319
  miniature hand saw, 99
The Art of Electronics, 235
audio amplifier (see Experiment 29: Filtering Frequencies)
audio distortion (see Experiment 30: Fuzz)
audio electronics, 227
AutoZone, 319
  fuses, 4
Avago, 319
  diodes, 103

B

Background
  Clipping, 257
  Early switching systems, 49
  Father of electromagnetism, 13
  From Boole to Shannon, 184–185
  How chips came to be, 152
  How much voltage does a wire consume?, 27
  How the timer was born, 160
  Inventor of the battery, 12
  Joseph Henry, 238
  Logic gate origins, 192
  Maddened by measurement, 124
  Michael Faraday and capacitors, 64
  Mounting a loudspeaker, 87
  Origins of loudspeakers, 243
  Origins of programmable chips, 294
  Positive and negative, 35
  Soldering myths, 106
  Stomp-box origins, 260–261
  The confusing world of TTL and CMOS, 190–192
  The man who discovered resistance, 8
  The origins of wattage, 28
  Transistor origins, 78
  Why didn't your tongue get hot?, 10
banana plugs, 97
Bardeen, John, 78
BASIC Stamp microcontrollers, 227, 294, 317, 319
batteries, 3, 103
  9-volt, 5
  abusing (see Experiment 2: Let's Abuse a Battery!)
  amperage, 10
  ampere basics, 11
  current, direct and alternating, 12
  heat, 10
  inventor, 12
  lithium, 9, 30
  making (see Experiment 5: Let's Make a Battery)
  schematics, 53
  voltage, 10
  volt basics, 11
  (see also Experiment 1: Taste the Power!)
battery holders and connectors, 3
battery life, 123
Bell, Alexander Graham, 243
benders, plastic, 272
binary arithmetic, 217
binary code, 214–220
BI Technologies, 319
  potentiometers, 4
BK Precision, 319
  multimeter, 2
boardmount sockets and pinstrip headers, 103
book sources, 234–235
Boolean logic, 184, 186
Boole, George, 184
Brattain, Walter, 78
breadboard alarm circuit, 135–136
breadboards, 39, 65–67
  (see also Experiments 8, 20, and 21)
break-to-make transistor circuit, 129–130
brushed DC motor, 280
Burkard, Johann, 261
Bussmann fuses, 319
Buzbee, Bill, 192

C

calipers, 99
Camenzind, Hans, 160
capacitance, 3, 39, 64, 86, 133, 157, 180, 236, 246–248, 249, 252, 262
  farad, 61
capacitors, 42, 61
  basics, 62–63
  diodes, 241
  Faraday, Michael, 64
  getting zapped by, 62
  nonpolarized electrolytic capacitors, 249
  polarity, 63
  time and (see Experiment 9: Time and Capacitors)
  time constant, 71–72
  (see also Experiment 8: A Relay Oscillator)
ceramic capacitors, 43
Chicago lighting, 319
chips, 147
  74HC00 quad 2-input NAND chip (see Experiment 19: Learning Logic)
  74HC00 (see Experiment 22: Flipping and Bouncing)
  74HC02 (see Experiment 22: Flipping and Bouncing)
  74HC04 (see Experiment 20: A Powerful Combination)
  74HC08 quad 2-input AND chip (see Experiments 19 and 20)

  - 74HC32 (see Experiment 21: Race to Place)
  - 74LS06 open-collector inverter chip, 150
  - 74LS27 (see Experiment 23: Nice Dice)
  - 74LS27 triple-input NOR chip, 150
  - 74LS92 counter chip, 150
  - 74LS92 (see Experiment 23: Nice Dice)
  - 555 timer (see 555 timer chip)
  - 4026, 150
    - (see also Experiment 18: Reaction Timer)
  - choosing, 148–149
  - common part numbers, 193
  - floating pins, 175
  - history, 152
  - logic chips, 150
  - power supplies, 154
  - programmable chips, origins, 294
- Churchill, Sir Winston, 29
- circuit chips (see chips)
- The Circuit Designer's Companion, 235
- C&K
  - pushbutton, 42
  - switch, 319
- clip-on meter test leads, 97
- clipping, 257
- CMOS chips, floating pins, 175
- CMOS Sourcebook, 234
- coil schematics, 238
- coil voltage, 58
- cold-weld, 106
- combination lock circuit schematic, 198, 201
- Common (COM) socket, 5, 6
- Complementary metal–oxide semiconductor (CMOS), 190–192
- computers, 232
  - magnets, 240
- conductor, 6
- control knobs, 229
- controller chips
  - limitations, 315
  - programmable chips, origins, 294
  - (see also Experiment 34: Hardware Meets Software)
- coulomb, 36
- countersink, hand-cranked, 99
- CraftAmerica, 319
- crossover networks, 253–257
- current
  - alternating, 248
  - checking flow, 21–25
  - direct and alternating, 12
  - transistors and, 80–81

D

- Darice, 319
- Darice Mini-Storage boxes, 230
- Davies, Ray, 260
- DC (direct current), 12
- DC power plug, 103
- DC stepper motor, 280
- DeArmond, Harry, 261
- deburring tool, 99, 271
- decimals, 29
- Dellepiane, Flavio, 260
- desoldering wick, 98
- DeWalt, 320
- DeWalt XRP jigsaw, 271
- diesel engine, 236
- Digi-Key, xi, 320
  - patch cords, 41
- diodes, 103, 134
  - capacitors, 241
  - schematics, 134
- direct current, 12
- Directed switches, 320
- distortion (see Experiment 30: Fuzz)
- Doctronics, 233, 320
- Doing the math on your tongue, 30
- double-throw switch, 42, 46
- DPDT relay, 103
- Dummer, Geoffrey W. A., 152

E

- eBay, xii, 320
- electrical potential, 36
- electricity
  - magnetism, relationship, 236
  - nature of, 33
  - sound into, 244–245
- electrolytic capacitors, 42, 61
- electromagnetism, father of, 13
- electromagnet (see Experiment 25: Magnetism)
- Electronics Club, 233
- electronic solder, 100
- electronics-tutorials.ws, 233
- Elenco, 320
  - helping hand, 96
- The Encyclopedia of Electronic Circuits, 235
- Essentials
  - All about diodes, 134
  - All about NPN and PNP transistors, 76–77
  - All about programmable unijunction transistors, 83–84
  - Battery life, 123
  - Logic gate basics, 186–189
  - Perfboard soldering procedure, 137
  - Real-world fault tracing, 139
  - Transistors and relays, 79
- Everlight, 320
  - low-power LEDs, 150
- Experiment 1: Taste the Power!, 5–9
  - cleanup and recycling, 9
  - further investigation, 8
  - procedure, 7
  - tools, 5
- Experiment 2: Let's Abuse a Battery!, 9–13
  - cleanup and recycling, 13
  - procedure, 9–12
- Experiment 3: Your First Circuit, 13–17
  - cleanup and recycling, 17
  - lighting LEDs, 16–17
  - setup, 13–15
- Experiment 4: Varying the Voltage, 18–31
  - checking flow, 21–25
  - cleanup and recycling, 29
  - dimming LEDs, 19–25
  - manufacturer's data sheet, 26
  - Ohm's Law, 26
  - potentiometers, 18
  - size of resistor required by LED, 28
- Experiment 5: Let's Make a Battery, 32–38
  - cleanup and recycling, 34
  - practical purposes, 37

Experiment 6: Very Simple Switching, 43–54
  connection problems, 45
  schematics, 50–54
  testing, 45–49
  tools, 44–45
Experiment 7: Relay-Driven LEDs, 55–59
  AC adapter plugs, 55–56
  how it works, 59
  operating current, 58
  procedure, 59
Experiment 8: A Relay Oscillator, 60–67
  adding capacitance, 61–64
  breadboarding circuits, 65–67
Experiment 9: Time and Capacitors, 68–72
  voltage, resistance, and capacitance, 69–72
Experiment 10: Transistor Switching, 73–81
  fingertip switching, 74–75
Experiment 11: A Modular Project, 82–93
  amplification, 87–89
  beyond persistence of vision, 86–87
  enhancements, 92–93
  pulsed output, 90–92
  slow-speed oscillation, 85
  wish list, 93
Experiment 12: Joining Two Wires Together, 104–113
  adding insulation, 110–111
  first solder join, 104–107
  modifying AC adapter, 111
  second solder join, 108
  shortening power cords, 111–113
Experiment 13: Broil an LED, 114–116
  heat-sinking, 115
Experiment 14: A Pulsing Glow, 117–126
  finishing, 121–126
  procedure for building this circuit, 120
  resizing the circuit, 119–120
Experiment 15: Intrusion Alarm Revisited, 127–146
  alarm installation, 145–146
  blocking bad voltage, 132–134
  breadboard alarm circuit, 135–136
  break-to-make transistor circuit, 129–130
  conclusion, 146
  final test, 145
  installing switches, 142–143
  magnetic sensor switches, 128
  perfboarding, 136–139
  self-locking relay, 131–132
  soldering switches, 143–144
  switches and inputs for alarm, 140–141
Experiment 16: Emitting a Pulse, 153–161
  procedure, 153–162
Experiment 17: Set Your Tone, 162–169
  astable modifications, 166–167
  chaining chips, 167–170
  procedure, 162–166
Experiment 18: Reaction Timer, 170–181
  counting, 171–174
  delay, 178–179
  display, 170–171
  enhancements, 180
  pulse generation, 175
  refinements, 176
  reflex tester, 179
  summary, 181
Experiment 19: Learning Logic, 181–196
Experiment 20: A Powerful Combination, 197–204
  breadboarding, 200–201
  computer interface, 202–203
  enhancements, 203
  incorrect inputs, 200
  questions, 200
  schematic, 198–199
Experiment 21: Race to Place, 205–210
  breadboarding, 209–210
  conceptual experiment, 205–209
  enhancements, 210
  goal, 205
  (see also Experiment 36: The Lock, Revisited)
Experiment 22: Flipping and Bouncing, 211–213
  how it works, 212–214
Experiment 23: Nice Dice, 214–222
  binary code, 214–220
  enhancements, 220–222
Experiment 24: Intrusion Alarm Completed, 223–226
  delay before deactivation, 225–226
  delayed activation, 223
  keypad deactivation, 223–224
  summary, 226
Experiment 25: Magnetism, 236–238
  procedure, 236
Experiment 26: Tabletop Power Generation, 239–241
  procedure, 239–242
Experiment 27: Loudspeaker Destruction, 242–245
  procedure, 242
Experiment 28: Making a Coil React, 246–248
  procedure, 246–248
Experiment 29: Filtering Frequencies, 249–256
  adding an amplifier, 250–252
  crossover networks, 253–257
  procedure, 250
Experiment 30: Fuzz, 257–261
  schematic, 258
Experiment 31: One Radio, No Solder, No Power, 262–267
  enhancements, 265
Experiment 32: A Little Robot Cart, 268–284
  bending, 272–273
  circuit, 276–281
  curving cuts, 271
  framing your cart, 274–276
  making 90 degree joints, 273–274
  making plans, 272
  mechanical power, 281–284

Experiment 33: Moving in Steps, 284–292
  adding autonomy, 291
  exploring your motor, 285–286
  quick demo, 286–289
  setting up light seeking robot, 291–292
  speed control, 290–291
Experiment 34: Hardware Meets Software, 293–305
  decoding the code, 302
  editing code, 303
  first program, 301
  loops, 304–305
  simulation, 303
  software installation and setup, 296–300
    Programming Editor software, 297–298
    setting up hardware, 298–300
    USB driver, 296–297
  troubleshooting, 300–301
  verifying connection, 300
Experiment 35: Checking the Real World, 306–310
  procedure, 306–310
Experiment 36: The Lock, Revisited, 311–317
  other controllers, 317–318
  user input, 311–315
  (see also Experiment 21: Race to Place)
Extech, 320
  multimeter, 2

F

Fairchild, 320
  integrated circuit chips, 150
  transistor, 42
  voltage regulators, 151
Faraday, Michael, 64
farad basics, 61
fault tracing, 139
Field, Simon Quellan, 265
flat-format storage boxes, 230
floating pins, 175
Formosa, Dan, 261
Franklin, Benjamin, 35
frequencies (see Experiment 29: Filtering Frequencies)
FTM, 320
Fujitsu, 320
  relays, 42
Fundamentals
  555 timer chip frequency in astable mod, 166
  555 timer chip pulse duration in monostable mode, 157
  555 timer limits, 161
  All about ABS, 269
  All about limit switches, 278
  All about motors, 280
  All about perforated board, 116
  All about switches, 46–48
  Ampere basics, 11
  A two-way relationship, 236
  Basic PICAXE parameters, 305
  Basic schematic symbols, 51–54
  Capacitor basics, 62–63
  Choosing chips, 148–149
  Coil schematics and basics, 238
  Common part numbers, 193
  Counters and seven-segment displays, 173
  Decimals, 29
  Decoding resistors, 14–15
  Direct and alternating current, 12
  Farad basics, 61
  Inside a relay, 58
  Limitations of MCUs, 315
  Ohms, 6
  Ohm's Law, 26
  PICAXE 08M extra features, 310
  Rules for connecting logic gates, 194–195
  Series and parallel, 25
  Switch bounce, 174
  The bistable 555 timer, 176–177
  Unexplored territory, 316
  Voltage regulators, 182
  Volt basics, 11
  Watt basics, 31
  Where to find ABS, 269
  Why the 555 is useful, 161
  Wire gauges, 282
fuses, 4
fuzz box, 257, 260

G

galvanometer, 13
GB Automatic Wire Strippers, 40
GB wire strippers, 320
general-duty soldering iron, 30 to 40 watts, 96
Getting Started in Electronic, 235
Getting Started with Arduino, 234
Gibson Maestro Fuzz-Tone, 260
grounding yourself, 172
Guru's Lair, 233

H

hand-cranked countersink, 99
hand saw, miniature, 99
headphone, high-impedance (see Experiment 31: One Radio, No Solder, No Power)
heat, 10
heat guns, 98
  warning, 110
heat-shrink tube, 101
heat-sinking, 115
helping hand, 96
Henry, Joseph, 237, 238
high-impedance headphone (see Experiment 31: One Radio, No Solder, No Power)
high voltage, 263
Hobbylinc, 320
  miniature hand saw, 99
Home Depot, xii, 320
  plywood, 102
  wire, 101
hookup wire, 41

I

IC sockets, 150
Ideal wire strippers, 320
incandescent lightbulb, 53
inductance, 26, 237, 241
  self-inductance, 246
inductive reactance, 246

insulated binding posts, 103
insulator, 6
integrated circuit chips (see chips)
intrusion alarm (see Experiments 11, 15, and 24)

J

jacks, 49
Jameco, 320
  potentiometer, 42
  solder, 100
joules, 36
jumper wire assortment, 41

K

Kellogg, Edward, 243
keypads
  (see Experiment 20: A Powerful Combination)
Kilby, Jack, 152
Kingbright, 320
Kinks, 260
K&J Magnetics, 320
Kobiconn, 320
  clip-on meter test leads, 97
Kronus Automatic Wire Strippers, 40
KVM Tools, 320
  deburring tool, 99

L

labeling, 231
Lancaster, Don, 233
latching relays, 151
leads, 5
LED numeric displays, 150
LEDs, 4
  dimming, 19–25
  lighting, 16–17
  low-powered, 150
  pulsing (see Experiment 14: A Pulsing Glow)
  relay-driven (see Experiment 7: Relay-Driven LEDs)
  schematics, 53, 54
  size of resistor required, 28
  too much heat (see Experiment 13: Broil an LED)
lemon battery, 32–38
lightbulb, 53
light-emitting diodes (see LEDs)
lightning, 35
limit switches, 278
lithium batteries, 9, 30
LM7805 voltage regulator (see Experiment 19: Learning Logic)
logic chips, 150
logic gates
  basics, 186–189
  origins, 192
  rules for connecting, 194–195
  (see also Experiments 19-24)
loudspeakers, 43, 104
  amplification, 87
  mounting, 87
  origins, 243
  sound into electricity, 244–245
  (see also Experiments 17, 27, and 29)
Lowe's hardware, xii, 320
  plywood, 102
  wire, 101
Lumex, 320
  diodes, 103

M

machine screws, 102
magnetic sensor switches, 128
magnetism and electricity, relationship, 236
magnetism generating electricity (see Experiment 26: Tabletop Power Generation)
magnets
  warnings, 240
  (see also Experiments 25 and 26)
magnifying lens, 96
mail-order sources, xi
Make: Electronic kits, xiii
Maker Shed, xiii
Making Things Talk, 234
manufacturer's data sheet, 26
Marconi, Guglielmo, 266
McMaster-Carr, xii, 230, 320
  deburring tool, 99
  hand-cranked countersink, 99
  hookup wire, 41
  soldering iron, 96
MCUs
  limitations, 315
  programmable chips, origins, 294
  (see Experiment 34: Hardware Meets Software)
measurements, 36
  conversion table, 124–126
  in inches, 124
Megahobby, 320
  miniature vise, 99
Meter Superstore, 97, 320
metric system, 124
Michaels craft stores, 229, 320
microcontrollers, 227
  PICAXE (see PICAXE)
microfarads, 61
Mill-Max, 320
  single inline sockets and headers, 103
miniature hand saw, 99
miniature screwdriver set, 98
miniature vise, 99
Mitutoyo, 320
  calipers, 100
Motorola, 320
  transistor, 42
motors, 227, 280
  stepper (see Experiment 33: Moving in Steps)
Mouser Electronics, xi, 320
  battery holders and connectors, 3
  potentiometers, 4
moving-coil microphone, 244
Mueller alligator clips, 101, 320
multimeter, 2
mutual repulsion, 33

N

NAND gates (see Experiments 19-24)
nanofarads, 61
negative charge, 35
neodymium magnet, 240
Newark, xii, 320

Newark Electronics
  Vectorboard, 101
newtons, 36
NKK switches, 320
noisemaking circuit, 92
nonpolarized electrolytic capacitors, 249
nonpolarized relays, 58
NOR gates, 212
Noyce, Robert, 152
NPN and PNP transistors, 76–77, 79, 258
numeric keypads
  (see Experiment 20: A Powerful Combination)
NXP semiconductors, 320

O

Ohm, Georg Simon, 8
ohms, 6
Ohm's Law, 25, 26, 30, 283
Omron, 320
  relays, 42
online sources, 233
On Semiconductor, 320
  transistor, 42
open-collector inverter chip, 150, 222
operating current, 58
Optek, 320
  diodes, 103
oscillator, relay (see Experiment 8: A Relay Oscillator)
oscilloscope, 231

P

Panasonic, 320
PanaVise, 321
panel-mounted power jack, 103
Parallax, 321
patch cords, 41
pencil-type 15-watt soldering iron, 96
Pep Boys, 321
perforated boards, 101, 116, 136–139
  common errors, 138
  soldering procedure, 137
P. H. Anderson, 295
Philips, 321
photoresistors (see Experiment 33: Moving in Steps)
PICAXE, 227, 321
  alphanumeric input/output, 310
  infrared, 310
  interrupts, 310
  limitations, 315
  origins of programmable chips, 294
  pseudorandom numbers, 310
  servo motor, 310
  tone generator, 310
  (see also Experiments 34-36)
pick and hook set, 99
picofarads, 61
Piedmont Plastics, 321
pin-shuffling, 161
Plano, 229
Plano storages boxes, 321
plastic benders, 272
pliers, 1
plywood, 102
PNP transistors (see NPN and PNP transistors)
point-to-point wiring connections, 106
Pomona
  clip-on meter test leads, 97
  test equipment, 321
positive charge, 35
potentiometers, 4, 42, 151, 229, 249
  schematics, 53
  (see also Experiments 4, 32, and 35)
power assessments, 31
power connectors, 229
power cords, shortening, 111–113
power outlet, 12
power supplies, 39, 231
  chips, 154
Practical Electronics for Inventors, 234
programmable chips, origins, 294
Programmable Intelligent Computer (PIC), 294
Programmable Interface Controller (PIC), 294
programmable microcontrollers, 227
project boxes, 102
Prolatch, 230
prototyping board, 101
PSP breadboard, 39
pushbuttons, 42
  options for toggle switches and pushbuttons, 47
PUTs (programmable unijunction transistors)
  2N6027 (see Experiment 14: A Pulsing Glow)
  schematics, 83
  (see also Experiments 11 and 14)

R

radio-frequency devices, 227
radio, how it works, 266–267
RadioShack, xi, 103, 321
  alligator clips, 4, 101
  battery holders and connectors, 3
  breadboard, 39
  clip-on meter test leads, 97
  desoldering wick, 98
  fuses, 4
  heat-shrink tube, 101
  helping hand, 96
  hookup wire, 41
  IC sockets, 150
  insulated binding posts, 103
  integrated circuit chips, 150
  Kronus Automatic Wire Strippers, 40
  LEDs, 4
  loudspeaker, 43
  magnifying lens, 96
  panel-mounted power jack, 103
  perforated board, 101
  pliers, 1
  potentiometer, 42
  power supply/universal AC adapter, 39
  project boxes, 102
  resistors, 4
  solder, 100
  soldering iron, 96
  soldering stand, 98
  transistor, 42
  wire cutters, 2

reactance, 246
reference sources, 233–235
  books, 234–235
  online, 233
relay-driven LEDs (see Experiment 7: Relay-Driven LEDs)
relay oscillator (see Experiment 8: A Relay Oscillator)
relays, 42, 56–58
  coil voltage, 58
  inside, 58
  latching, 151
  nonpolarized, 58
  self-locking relay, 131–132
  set voltage, 58
  switching capacity, 58
  transistors and, 79
reset voltage, 156
resistance, 25, 246
  discovery of, 8
  larger, 10
  measuring, 6
  (see also Experiment 1: Taste the Power!)
resistors, 4, 43
  decoding, 14–15
  hot, 247
  in parallel, 25
  in series, 25
  power assessments, 31
  schematics, 52
  (see also Experiment 3: Your First Current)
Rice, Chester, 243
Richards, Keith, 260
robotics, 227
  (see Experiments 32 and 33)
Rolling Stones, 260

S

saw, miniature, 99
schematics, 50–54
  4026 chip, 171
  batteries, 53
  coil, 238
  combination lock circuit, 198, 201
  diodes, 134
  dots, 51
  Experiment 20: A Powerful Combination, 198–199
  Experiment 30: Fuzz, 258
  incandescent lightbulb, 53
  LEDs, 53, 54
  potentiometer, 53
  PUTs, 83
  resistors, 52
  switches, 51
  symbols, 51
  wires, 52
Scitoys Catalog, 265
screwdriver set, miniature, 98
screws, 102
  tamper-proof, 203
Scribner, Charles E., 49
Sears, 321
  deburring tool, 99
self-inductance, 246
self-locking relay, 131–132
servo motor, 280
  PICAXE, 310
set voltage, 58
Shannon, Claude, 184
Shockley, William, 78, 152
short circuits, 9
Siemens, Ernst, 243
signal diodes, 134
sine wave, 255
single inline sockets and headers, 103
single-pole, double-throw switch (SPDT), 46
single-pole, single-throw switch (SPST), 46
single-pole switch, 42
single-throw switches, 46
sockets, 49
solder, 100
soldering
  alternatives, 106
  desoldering, 109
  errors, 107
  myths, 106
  switches, 143–144
  theory, 109
  (see also Experiment 12: Joining Two Wires Together)
soldering irons, 96
  warning, 104
soldering stand, 98
solder pump, 98
sound into electricity, 244–245
SparkFun Electronics, 321
sparking, 48
SPDT toggle switch, 103
square wave, 258
Stanley, 321
  pick and hook set, 99
  wire cutters, 2
stepper motors
  inside, 288–289
  (see also Experiment 33: Moving in Steps)
STMicroelectronics, 321
  integrated circuit chips, 150
  logic chips, 150
  transistor, 42
stomp-box origins, 260–261
storage containers, 229
switch bounce, 174
switches, 42, 46–48
  checking, 48
  double-pole switches, 46
  double-throw switch, 46
  early systems, 49
  installing for alarm, 142–143
  limit, 278
  magnetic sensor switches, 128
  options for toggle switches and pushbuttons, 47
  relays and, 103
  schematics, 51
  simple (see Experiment 6: Very Simple Switching)
  single-pole, double-throw switch (SPDT), 46
  single-pole, single-throw switch (SPST), 46
  single-throw switches, 46
  soldering, 143–144
  sparking, 48
  tactile, 151
switching capacity, 58
switching transistors (see Experiment 10: Transistor Switching)

T

tactile switches, 151
tamper-proof screws, 203
tantalum capacitor, 63
Texas Instruments, 321
  4026 Decade Counter, 150
  logic chips, 150
Theory
  Alternating current concepts, 248
  Basic measurements, 36
  Binary arithmetic, 217
  Calculating voltage drop, 283–284
  Doing the math on your tongue, 30
  How radio works, 266–267
  Inductance, 237
  Inside a stepper motor, 288–289
  Inside the 555 timer: astable mode, 164–165
  Inside the 555 timer: monostable mode, 158–159
  Power assessments, 31
  See the current, 80–81
  Soldering theory, 109
  Sound, electricity, and sound, 244–245
  The nature of electricity, 33
  The time constant, 71–72
  Waveforms, 255–256
third hand, 96
Thomson, J. J., 35
time constant, 71–72
toggle switches, 42, 229
  options for toggle switches and pushbuttons, 47
Tools
  Desoldering, 109
  Eight most common soldering errors, 107
  Four most common perfboarding errors, 138
torque, 281
Tower Hobbies, 321
  miniature hand saw, 99
tracing a fault, 139
transistors, 42
  2N2222 (see Experiments 11 and 15)
  2N6027 programmable unijunction transistor (see Experiment 14: A Pulsing Glow)
  basics, 77
  beta value, 80
  break-to-make transistor circuit, 129–130
  getting a better understanding, 80–81
  NPN and PNP, 76–77
  origins, 78
  relays and, 79
  switching (see Experiment 10: Transistor Switching)
  unijunction, 82–84
Transistor-Transistor Logic (TTL), 190–192
tremolo box, 261
Trem-Trol, 261
trigger voltage, 156
TTL Cookbook, 234
tweeter, 253
Twin Industries, 321
  perforated board, 101
Tyco, 321
  relays, 42

U

unijunction transistors, 82–84

V

Vaughan, 321
Vectorboard, 101
Velleman keyboards, 321
Ventures, 260
VIH min, 192
VIL max, 192
vise, miniature, 99
Vishay, 321
VOH min, 192
VOL max, 192
Volta, Alessandro, 12
voltage, 10
  blocking bad voltage, 132–134
  calculating drop, 283–284
  high, 263
  how much wire consumes, 27
  measuring, 6
voltage regulators, 151, 182
  LM7805 (see Experiment 19: Learning Logic)
volts, 36
  basics, 11
Vox wow-fuzz pedal, 257

W

Wal-Mart, 229, 321
Warnings
  Avoid Burns While Bending, 272
  Beware of Pin-Shuffling!, 161
  Blood Blisters and Dead Media, 240
  Capacitor Polarity, 63
  Cut with Care, 270
  Flying Wire Segments, 121
  Getting Zapped by Capacitors, 62
  Grounding Yourself, 172
  Heat Guns Get Hot, Too!, 110, 111
  High Voltage!, 263
  Hot Resistors, 247
  Never Use Two Hands, 74
  No Floating Pins!, 175
  No More Than 9 Volts, 5
  Pin 2 Pull-Down, 299
  Short Circuits, 9
  Soldering Irons Get Hot!, 104
  The Warranty Issue, 197
water-powered turbine, 236
wattage, history, 28
Watt, James, 28
watts, 36
  basics, 31
waveforms, 255–256
  clipping, 257
wavelength of the sound, 244
Weller, 321
Weller Therma-Boost, 96
windmill, 236
wire cutters, 2
  warning, 121
wire gauges, 282
wires, 100
  point-to-point wiring connections, 106
  schematics, 52
  soldering (see also Experiment 12: Joining Two Wires Together)
  voltage consumed by, 27

wire strippers, 40
wire wrap, 106
woofer, 253
work area, customizing, 228–232
  bench, 231
  computer, 232
  labeling, 231
  oscilloscope, 231

X

X-Acto, 321
  miniature hand saw, 99
Xcelite, 321
  pliers, 1
Xytronic, 321
  soldering iron, 96

Z

Zener diode, 134
zinc electrode, 33

Colophon

The heading and cover font are BentonSans, the text font is Myriad Pro, and the code font is TheSansMonoCondensed.

About the Author

Charles Platt became interested in computers when he acquired an Ohio Scientific C4P in 1979. After writing and selling software by mail order, he taught classes in BASIC programming, MS-DOS, and subsequently Adobe Illustrator and Photoshop. He wrote five computer books during the 1980s.

He has also written science-fiction novels such as *The Silicon Man* (published originally by Bantam and later by Wired Books) and *Protektor* (from Avon Books). He stopped writing science fiction when he started contributing to *Wired* in 1993, and became one of its three senior writers a couple of years later.

Charles began contributing to MAKE magazine in its third issue and is currently a contributing editor. *Make: Electronics* is his first title for Make Books. Currently he is designing and building prototypes of medical equipment in his workshop in the northern Arizona wilderness.

Radio Mathematics

Understanding radio and electronics beyond an intuitive or verbal level requires the use of some mathematics. None of the math is more advanced than trigonometry or advanced algebra, but if you don't use math regularly, it's quite easy to forget what you learned during your education. In fact, depending on your age and education, you may not have encountered some of these topics at all.

It is well beyond the scope of this book to be a math textbook, but this section touches some of topics that most need explanation for amateurs. For introductory-level tutorials and explanations of advanced topics, a list of on-line, no-cost tutorials is presented in the sidebar, "Online Math Resources". You can browse these tutorials whenever you need them in support of the information in this reference text.

Online Math Resources

The following Web links are a compilation of on-line resources organized by topic. Other resources are available online at **www.arrl.org/tech-prep-resource-library**. Look for the Math Tutorials section.

Many of the tutorials listed below are part of the *Interactive Mathematics* Web site (**www.intmath.com**), a free, online system of tutorials. The system begins with basic number concepts and progresses all the way through introductory calculus. The lessons referenced here are those of most use to a student of radio electronics.

Basic Numbers & Formulas

Order of Operations — **www.intmath.com/Numbers/3_Order-of-operations.php**
Powers, Roots, and Radicals — **www.intmath.com/Numbers/4_Powers-roots-radicals.php**
Introduction to Scientific Notation — **www.ieer.org/clssroom/scinote.html**
Scientific Notation — **www.intmath.com/Numbers/6_Scientific-notation.php**
Ratios and Proportions — **www.intmath.com/Numbers/7_Ratio-proportion.php**
Geometric Formulas — **www.equationsheet.com/sheets/Equations-4.html**
Tables of Conversion Factors — **en.wikipedia.org/wiki/Conversion_of_units**

Metric System

Metric System Overview — **en.wikipedia.org/wiki/Metric_system**
Metric System Tutorial —see Unit 3 at **www.arrl.org/radio-lab-handbook**

Conversion of Units

Metric-English — **www.m-w.com/mw/table/metricsy.htm**
Metric-English — **en.wikipedia.org/wiki/Metric_yardstick**
Conversion Factors — **oakroadsystems.com/math/convert.htm**

Fractions

Equivalent Fractions — **www.intmath.com/Factoring-fractions/5_Equivalent-fractions.php**
Multiplication and Division— **www.intmath.com/Factoring-fractions/6_Multiplication-division-fractions.php**
Adding and Subtracting — **www.intmath.com/Factoring-fractions/7_Addition-subtraction-fractions.php**
Equations Involving Fractions — **www.intmath.com/Factoring-fractions/8_Equations-involving-fractions.php**
Basic Algebra — **www.intmath.com/Basic-algebra/Basic-algebra-intro.php**

Graphs

Basic Graphs — **www.intmath.com/Functions-and-graphs/Functions-graphs-intro.php**
Polar Coordinates — **www.intmath.com/Plane-analytic-geometry/7_Polar-coordinates.php**
Exponents & Radicals — **www.intmath.com/Exponents-radicals/Exponent-radical.php**
Exponential & Logarithmic Functions — **www.intmath.com/Exponential-logarithmic-functions/Exponential-log-functionsintro.php**

Trigonometry

Basic Trig Functions — **www.intmath.com/Trigonometric-functions/Trig-functions-intro.php**
Graphs of Trig Functions — **www.intmath.com/Trigonometric-graphs/Trigo-graph-intro.php**

Complex Numbers

Complex Numbers — **www.intmath.com/Complex-numbers/imaginary-numbers-intro.php**

Working With Decibels

The *decibel* (*dB*) is a way of expressing a ratio *logarithmically*, meaning as a power of some *base* number, such as 10 or the base for *natural logarithms*, the number $e \approx 2.71828$. A decibel, using the metric prefix "deci" (d) for one-tenth, is one-tenth of a *Bel* (B), a unit used in acoustics representing a ratio of 10, and named for the telephone's inventor, Alexander Graham Bell. As it turns out, the Bel was too large a ratio for common use and the deci-Bel, or decibel, became the standard unit.

In radio, the logarithmic ratio is used, but it compares signal strengths — power, voltage, or current. The decibel is more useful than a *linear* ratio because it can represent a wider scale of ratios. The numeric values encountered in radio span a very wide range and so the dB is more suited to discuss large ratios. For example, a typical receiver encounters signals that have powers differing by a factor of 100,000,000,000,000 (10^{14} or 100 trillion!). Expressed in dB, that range is 140, which is a lot easier to work with than the preceding number, even in scientific notation!

The formula for computing the decibel equivalent of a power ratio is

$$\text{dB} = 10 \log (\text{power ratio}) = 10 \log (P1/P2) \tag{1}$$

or if voltage is used

$$\text{dB} = 20 \log (\text{voltage ratio}) = 20 \log (V1/V2) \tag{2}$$

For equations 1 and 2 to produce equivalent results, both of the voltages must be measured across equivalent impedances. Otherwise, impedance must be accounted for according to $P = V^2 / Z$.

Positive values of dB mean the ratio is greater than 1 and negative values of dB indicate a ratio of less than 1. For example, if an amplifier turns a 5 W signal into a 25 W signal, that's a gain of 10 log (25 / 5) = 10 log (5) = 7 dB. On the other hand, if by adjusting a receiver's volume control the audio output signal voltage is reduced from 2 V to 0.1 V, that's a loss: 20 log (0.1 / 2) = 20 log (0.05) = –26 dB.

If you are comparing a measured power or voltage (P_M or V_M) to some reference power (P_{REF} or V_{REF}) the formulas are:

$$\text{dB} = 10 \log (P_M / P_{REF})$$

$$dB = 20 \log (V_M / V_{REF})$$

There are several commonly-used reference powers and voltages, such as 1 V or 1 mW. When a dB value uses one of them as the references, dB is followed with a letter. Here are the most common:

- dBV means dB with respect to 1 V (V_{REF} = 1 V)
- dBμV means dB with respect to 1 μV (V_{REF} = 1 μV)
- dBm means dB with respect to 1 mW (P_{REF} = 1 mW)

If you are given a ratio in dB and asked to calculate the power or voltage ratio, use the following formulas:

$$\text{Power ratio} = \text{antilog}\ (dB / 10) \qquad (3)$$

$$\text{Voltage ratio} = \text{antilog}\ (dB / 20) \qquad (4)$$

Example: A power ratio of –9 dB = antilog –(9 / 10) = antilog (–0.9) = ⅛ = 0.125

Example: A voltage ratio of 32 dB = antilog (32 / 20) = antilog (1.6) = 40

Antilog is also written as $\log^{-1}$ and may be labeled that way on calculators.

CONVERTING BETWEEN DB AND PERCENTAGE

You may also have to convert back and forth between decibels and percentages. Here are the required formulas:

$$dB = 10 \log\ (\text{percentage power} / 100\%) \qquad (5)$$

$$dB = 20 \log\ (\text{percentage voltage} / 100\%) \qquad (6)$$

$$\text{Percentage power} = 100\% \times \text{antilog}\ (dB / 10) \qquad (7)$$

$$\text{Percentage voltage} = 100\% \times \text{antilog}\ (dB / 20) \qquad (8)$$

Example: A power ratio of 20% = 10 log (20% / 100%) = 10 log (0.2) = –7 dB

Example: A voltage ratio of 150% = 20 log (150% / 100%) = 20 log (1.5) = 3.52 dB

Example: –1 dB represents a percentage power = 100% × antilog (–1 / 10) = 79%

Example: 4 dB represents a percentage voltage = 100% × antilog (4 / 20) = 158%

Rectangular and Polar Coordinates

Graphs are drawings of what equations describe with symbols — they're both saying the same thing. Graphs are used to present a visual representation of what an equation expresses. The way in which mathematical quantities are positioned on the graph is called the *coordinate system*. *Coordinate* is another name for the numeric scales that divide the graph into regular units. The location of every point on the graph is described by a pair of *coordinates*.

The two most common coordinate systems used in radio are the *rectangular-coordinate* system shown in **Figure 1** (sometimes called *Cartesian coordinates*) and the *polar-coordinate* system shown in **Figure 2**.

The line that runs horizontally through the

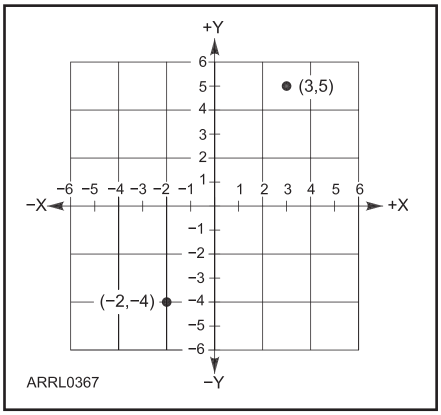

Figure 1 — Rectangular-coordinate graphs use a pair of axes at right angles to each other; each calibrate in numeric units. Any point on the resulting grid can be expressed in terms of its horizontal (X) and vertical (Y) values, called coordinates.

Computer Software and Calculators

Every version of the *Windows* operating system comes with a calculator program located in the Accessories program group and a number of free calculators are available on-line. Enter "online" and "calculator" into the search window of an Internet search engine for a list. If you can express your calculation as a mathematical expression, such as sin(45) or 10log(17.5), it can be entered directly into the search window at **www.google.com** and the Google calculator will attempt to solve for the answer.

Microsoft *Excel* (and similar spreadsheet programs) also make excellent calculators. If you are unfamiliar with the use of spreadsheets, here are some online tutorials and help programs:

homepage.cs.uri.edu/tutorials/csc101/pc/excel97/excel.html
www.baycongroup.com/el0.htm
www.bcschools.net/staff/ExcelHelp.htm

For assistance in converting units of measurement, the Web site **www.onlineconversion.com** is very useful. The Google online unit converter can also be used by entering the required conversion, such as "12 gauss in tesla", into the Google search window.

Decibels In Your Head

Every time you increase the power by a factor of 2 times, you have a 3-dB increase of power. Every 4 times increase of power is a 6-dB increase of power. When you increase the power by 10 times, you have a power increase of 10 dB. You can also use these same values for a decrease in power. Cut the power in half for a 3-dB loss of power. Reduce the power to ¼ the original value for a 6-dB loss in power. If you reduce the power to 1⁄10 of the original value you will have a 10-dB loss. The power-loss values are often written as negative values: –3, –6 or –10 dB. The following tables show these common decibel values and ratios.

Common dB Values and Power Ratios

P2/P1	*dB*
0.1	–10
0.25	–6
0.5	–3
1	0
2	3
4	6
10	10

Common dB Values and Voltage Ratios

V2/V1	*dB*
0.1	–20
0.25	–12
0.5	–6
0.707	–3
1.0	0
1.414	3
2	6
4	12
10	20

You can also derive all the dB equivalents of integer ratios by adding or subtracting dB values. For example, to calculate the dB for a power ratio of 10/4 (2.5), subtract the dB equivalents for 10 and 4: 10 – 6 = 4 dB. Similarly for a ratio of 10/2 (5), 10 – 3 = 7 dB. The ratio of 5/4 (1.25) is 7 – 6 = 1 dB and so forth. The same trick can be used with the voltage ratios.

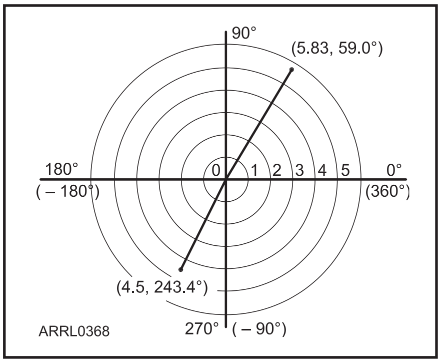

Figure 2 — Polar-coordinate graphs use a radius from the origin and an angle from the 0º axis to specify the location of a point. Thus, the location of any point can be specified in terms of a radius and an angle.

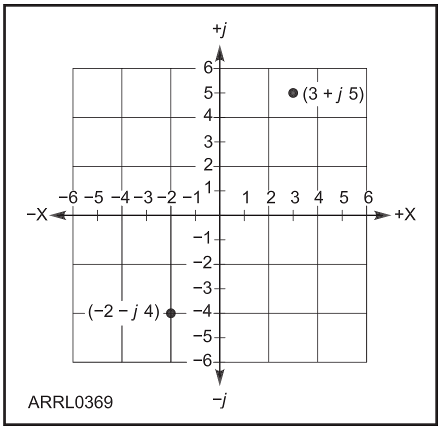

Figure 3 — The Y axis of a complex-coordinate graph represents the imaginary portion of complex numbers. This graph shows the same numbers as in Figure 1, graphed as complex numbers.

center of a rectangular coordinate graph is called the *X axis*. The line that runs vertically through the center of the graph is normally called the *Y axis*. Every point on a rectangular coordinate graph has two coordinates that identify its location, X and Y, also written as (X,Y). Every different pair of coordinate values describes a different point on the graph. The point at which the two axes cross — and where the numeric values on both axes are zero — is called the *origin*, written as (0,0).

In Figure 1, the point with coordinates (3,5) is located 3 units from the origin along the X axis and 5 units from the origin along the Y axis. Another point at (–2,–4) is found 2 units to the left of the origin along the X axis and 4 units below the origin along the Y axis. Do not confuse the X coordinate with the X representing reactance! Reactance is usually plotted on the Y-axis, creating occasional confusion.

In the polar-coordinate system, points on the graph are described by a pair of numeric values called *polar coordinates*. In this case, a length, or *radius*, is measured from the origin, and an *angle* is measured counter-clockwise from the 0º line as shown in Figure 2. The symbol *r* is used for the radius and θ for the angle. A number in polar coordinates is written $r\angle\theta$. So the two points described in the last paragraph could also be written as (5.83, ∠59.0°) and (4.5, ∠243.4°). Unlike geographic maps, 0º is always to the *right* along the X-axis and 90°at the top along the Y-axis!

Angles are specified counterclockwise from the 0º axis. A negative value in front of an angle specifies an angle measured clockwise from the 0º axis. (Angles can also be given with reference to some arbitrary line, as well.) For example, –270º is equivalent to 90º, –90º is equivalent to 270º, 0º and –360º are equivalent, and +180º and –180º are equivalent. With an angle measured clockwise from the 0º axis, the polar coordinates of the second point in Figure 2 would be (4.5, ∠–116.6°).

In some calculations the angle will be specified in *radians* (1 radian = 360 / 2π degrees = 57.3°), but you may assume that all angles are in degrees in this book. Radians are used when *angular frequency* (ω) is used instead of frequency (*f*). Angular frequency is measured in units of 2π radians/sec and so ω = 2πf.

In electronics, it's common to use both the rectangular and polar-coordinate systems when dealing with impedance problems. The examples in the next few sections of this book should help you become familiar with these coordinate systems and the techniques for changing between them.

Complex Numbers

Mathematical equations that describe phases and angles of electrical quantities use the symbol *j* to represent the square root of minus one ($\sqrt{-1}$). (Mathematicians use *i* for the same purpose, but *i* is used to represent current in electronics.) The number *j* and any real number multiplied by *j* are called *imaginary numbers* because no real number is the square root of minus one. For example, 2*j*, 0.1*j*, 7*j*/4, and 457.6*j* are all imaginary numbers. Imaginary numbers are used in electronics to represent reactance, such as *j*13 Ω of inductive reactance or –*j*25 Ω of capacitive reactance.

Real and imaginary numbers can be combined by using addition or subtraction. Adding a real number to an imaginary number creates a hybrid called a *complex number*, such as 1 + *j* or 6 – 7*j*. These numbers come in very handy in radio, describing impedances, relationships between voltage and current, and many other phenomena.

If the complex number is broken up into its real and imaginary parts, those two numbers can also be used as coordinates on a graph using *complex coordinates*. This is a special type of rectangular-coordinate graph that is also referred to as the *complex plane*. By convention, the X axis coordinates represent the real number portion of the complex number and Y axis coordinates the imaginary portion. For example, the complex number 6 – 7*j* would have the same location as the point (6,–7) on a rectangular-coordinate graph. **Figure 3** shows the same points as Figure 1.

The number *j* has a number of interesting properties:

$1/j = -j$

$j^2 = -1$

$j^3 = -j$

$j^4 = 1$

Multiplication by *j* can also represent a phase shift or rotation of +90°. In polar coordinates *j* = 1 ∠90°.

WORKING WITH COMPLEX NUMBERS

Complex numbers representing electrical quantities can be expressed in either rectangular form (a + *j*b) or polar form (r∠θ) as described above. Adding complex numbers is easiest in rectangular form:

$$(a + jb) + (c + jd) = (a+c) + j(b+d) \tag{9}$$

Multiplying and dividing complex numbers is easiest in polar form:

$$a\angle\theta_1 \times b\angle\theta_2 = \left(a \times b\right)\angle\left(\theta_1 + \theta_2\right) \tag{10}$$

and

$$\frac{a\angle\theta_1}{b\angle\theta_2} = \left(\frac{a}{b}\right)\angle\left(\theta_1 - \theta_2\right) \tag{11}$$

Converting from one form to another is useful in some kinds of calculations. For example, to calculate the value of two complex impedances in parallel you use the formula

$$Z = \frac{Z_1 Z_2}{Z_1 + Z_2}$$

To calculate the numerator (Z_1Z_2) you would write the impedances in polar form. To calculate the numerator ($Z_1 + Z_2$) you would write the impedances in rectangular form. So you need to be able to convert back and forth from one form to the other. Here is the procedure:

To convert from rectangular (a + *j*b) to polar form (r ∠ θ):

$$r = \sqrt{a^2 + b^2} \text{ and } \theta = \tan^{-1}(b / a) \quad (12)$$

To convert from polar to rectangular form:

$$a = r \cos\theta \text{ and } b = r \sin\theta \quad (13)$$

Many calculators have polar-rectangular conversion functions built-in and they are worth learning how to use. Be sure that your calculator is set to the correct units for angles, radians or degrees.

Example: Convert 3 ∠ 60° to rectangular form:

$$a = 3 \cos 60° = 3\,(0.5) = 1.5$$

$$b = 3 \sin 60° = 3\,(0.866) = 2.6$$

$$3\angle 60° = 1.5 + j2.6$$

Example: Convert 0.8 + *j*0.6 to polar form:

$$r = \sqrt{0.8^2 + 0.6^2} = 1$$

$$\theta = \tan^{-1}\left(0.6 / 0.8\right) = 36.8°$$

$$0.8 + j0.6 = 1\angle 36.8°$$

Accuracy and Significant Figures

The calculations you have encountered in this chapter and will find throughout this handbook follow the rules for accuracy of calculations. Accuracy is represented by the number of *significant digits* in a number. That is, the number of digits that carry numeric value information beyond order of magnitude. For example, the numbers 0.123, 1.23, 12.3, 123, and 1230 all have three significant digits.

The result of a calculation can only be as accurate as its *least* accurate measurement or known value. This is important because it is rare for measurements to be more accurate than a few percent. This limits the number of useful significant digits to two or three. Here's another example; what is the current through a 12-Ω resistor if 4.6 V is applied? Ohm's Law says I in amperes = 4.6 / 12 = 0.3833333... but because our most accurate numeric information only has two significant digits (12 and 4.6), strictly applying the significant digits calculation rule limits our answer to 0.38. One extra digit is often included, 0.383 in this case, to act as a guide in rounding off the answer.

Quite often a calculator is used, and the result of a calculation fills the numeric window. Just because the calculator shows 9 digits after the decimal point, this does not mean that is a more correct or even useful answer.

JON B. HAGEN

Radio-Frequency Electronics

Circuits and Applications

Second Edition

CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521889742

This page intentionally left blank

Radio-Frequency Electronics

Circuits and Applications

This second, much updated edition of the best-selling *Radio-Frequency Electronics* introduces the basic concepts and key circuits of radio-frequency systems. It covers the fundamental principles applying to all radio devices, from wireless single-chip data transceivers to high-power broadcast transmitters.

New to this edition:

- Extensively revised and expanded throughout, including new chapters on radar, digital modulation, GPS navigation, and *S*-parameter circuit analysis.
- New worked examples and end-of-chapter problems aid and test understanding of the topics covered.
- Numerous extra figures provide a visual aid to learning, with over 400 illustrations throughout the book.

Key topics covered include filters, amplifiers, oscillators, modulators, low-noise amplifiers, phase lock loops, transformers, waveguides, and antennas. Assuming no prior knowledge of radio electronics, this is a perfect introduction to the subject. It is an ideal textbook for junior or senior courses in electrical engineering, as well as an invaluable reference for professional engineers in this area.

Praise for the first edition:
This book is wonderfully informative, and refreshingly different from the usual rehash of standard engineering topics. Hagen has put his unique insights, gleaned from a lifetime of engineering and radio science, into this volume and it shows. There's an insight per page, at least for me, that makes it truly enjoyable reading, even for those of us who think we know something about the field! *Paul Horowitz, Harvard University*

Jon B. Hagen was awarded his Ph.D. from Cornell University in 1972, where he went on to gain 30 years' experience as an electronic design engineer, as well as establishing and teaching a Cornell electrical engineering course on RF electronics. Now retired, he has held positions as Principal Engineer at Raytheon, Electronics Department Head at the Arecibo Observatory in Puerto Rico, and Director of the NAIC Support Laboratory at Cornell.

Radio-Frequency Electronics

Circuits and Applications

Second Edition

Jon B. Hagen

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521889742

© Cambridge University Press 1996, 2009

This publication is in copyright. Subject to statutory exception and to the provision of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published in print format 2009

ISBN-13 978-0-511-58012-3 eBook (EBL)

ISBN-13 978-0-521-88974-2 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of urls for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

Preface *page* xiii

1 Introduction 1
1.1 RF circuits 2
1.2 Narrowband nature of RF signals 3
1.3 AC circuit analysis – a brief review 3
1.4 Impedance and admittance 4
1.5 Series resonance 4
1.6 Parallel resonance 5
1.7 Nonlinear circuits 5
Problems 5

2 Impedance matching 10
2.1 Transformer matching 11
2.2 L-networks 12
2.3 Higher Q – pi and T-networks 14
2.4 Lower Q – the double L-network 15
2.5 Equivalent series and parallel circuits 16
2.6 Lossy components and efficiency of matching networks 16
Problems 17

3 Linear power amplifiers 19
3.1 Single-loop amplifier 19
3.2 Drive circuitry: common-collector, common-emitter, and common-base 20
3.3 Shunt amplifier topology 22
3.4 Dual-polarity amplifiers 22
3.5 Push–pull amplifiers 23
3.6 Efficiency calculations 25
3.7 AC amplifiers 26

3.8 RF amplifiers 29
3.9 Matching a power amplifier to its load 31
Problems 31

4 Basic filters 34
4.1 Prototype lowpass filter designs 35
4.2 A lowpass filter example 36
4.3 Lowpass-to-bandpass conversion 38
Appendix 4.1 Component values for normalized lowpass filters 41
Problems 43
References 45

5 Frequency converters 46
5.1 Voltage multiplier as a mixer 46
5.2 Switching mixers 48
5.3 A simple nonlinear device as a mixer 51
Problems 53

6 Amplitude and frequency modulation 54
6.1 Amplitude modulation 55
6.2 Frequency and phase modulation 58
6.3 AM transmitters 62
6.4 FM transmitters 65
6.5 Current broadcasting practice 65
Problems 66

7 Radio receivers 67
7.1 Amplification 67
7.2 Crystal sets 68
7.3 TRF receivers 68
7.4 The superheterodyne receiver 69
7.5 Noise blankers 74
7.6 Digital signal processing in receivers 75
Problems 75
References 76

8 Suppressed-carrier AM and quadrature AM (QAM) 77
8.1 Double-sideband suppressed-carrier AM 77
8.2 Single-sideband AM 78
8.3 Product detector 80
8.4 Generation of SSB 81

8.5 Single-sideband with class C, D, or E amplifiers 83
8.6 Quadrature AM (QAM) 84
Problems 85
References 86

9 Class-C, D, and E Power RF amplifiers 87
9.1 The class-C amplifier 87
9.2 The class-D RF amplifier 92
9.3 The class-E amplifier 94
9.4 Which circuit to use: class-C, class-D, or class-E? 99
Problems 100
References 100

10 Transmission lines 101
10.1 Characteristic impedance 101
10.2 Waves and reflected waves on transmission lines 103
10.3 Modification of an impedance by a transmission line 106
10.4 Transmission line attenuation 107
10.5 Impedance specified by reflection coefficient 107
10.6 Transmission lines used to match impedances 111
Appendix 10.1. Coaxial cable – Electromagnetic analysis 114
Problems 116

11 Oscillators 120
11.1 Negative feedback (relaxation) oscillators 120
11.2 Positive feedback oscillators 121
11.3 Oscillator dynamics 128
11.4 Frequency stability 128
11.5 Colpitts oscillator theory 129
Problems 132

12 Phase lock loops and synthesizers 134
12.1 Phase locking 134
12.2 Frequency synthesizers 144
Problems 150
References 151

13 Coupled-resonator bandpass filters 152
13.1 Impedance inverters 152
13.2 Conversion of series resonators to parallel resonators and vice versa 155
13.3 Worked example: a 1% fractional bandwidth filter 156
13.4 Tubular bandpass filters 158
13.5 Effects of finite Q 160
13.6 Tuning procedures 161

13.7 Other filter types 161
Problems 162
References 163

14 Transformers and baluns 164

14.1 The "ideal transformer" 165
14.2 Transformer equivalent circuit 166
14.3 Power transformer operation 168
14.4 Mechanical analogue of a perfectly coupled transformer 169
14.5 Magnetizing inductance used in a transformer-coupled amplifier 170
14.6 Double-tuned transformer: making use of magnetization and leakage inductances 170
14.7 Loss in transformers 172
14.8 Design of iron-core transformers 172
14.9 Transmission line transformers 175
14.10 Baluns 176
Problems 178
References 180

15 Hybrid couplers 181

15.1 Directional coupling 182
15.2 Transformer hybrid 182
15.3 Quadrature hybrids 185
15.4 How to analyze circuits containing hybrids 186
15.5 Power combining and splitting 187
15.6 Other hybrids 189
Problems 192
Reference 194

16 Waveguide circuits 195

16.1 Simple picture of waveguide propagation 195
16.2 Exact solution: a plane wave interference pattern matches the waveguide boundary conditions 196
16.3 Waveguide vs. coax for low-loss power transmission 201
16.4 Waveguide impedance 201
16.5 Matching in waveguide circuits 202
16.6 Three-port waveguide junctions 202
16.7 Four-port waveguide junctions 203
Appendix 16.1 Lowest loss waveguide vs. lowest loss coaxial line 204
Appendix 16.2 Coax dimensions for lowest loss, highest power, and highest voltage 206
Problems 207
References 207

17 Small-signal RF amplifiers 208
17.1 Linear two-port networks 208
17.2 Amplifier specifications – gain, bandwidth, and impedances 210
17.3 Narrowband amplifier circuits 213
17.4 Wideband amplifier circuits 214
17.5 Transistor equivalent circuits 214
17.6 Amplifier design examples 215
17.7 Amplifier noise 219
17.8 Noise figure 220
17.9 Other noise parameters 222
17.10 Noise figure measurement 223
Problems 223
References 226

18 Demodulators and detectors 227
18.1 AM Detectors 227
18.2 FM demodulators 233
18.3 Power detectors 238
Problems 240
References 241

19 Television systems 242
19.1 The Nipkov system 242
19.2 The NTSC system 243
19.3 Digital television 251
Problems 257
References 258

20 Antennas and radio wave propagation 259
20.1 Electromagnetic waves 259
20.2 Radiation from a current element 261
20.3 Dipole antenna 262
20.4 Antenna directivity and gain 264
20.5 Effective capture area of an antenna 266
20.6 Reflector and horn antennas 267
20.7 Polarization 271
20.8 A spacecraft radio link 272
20.9 Terrestrial radio links 273
20.10 The ionosphere 273
20.11 Other modes of propagation 275
Problems 276
References 277

21 Radar 278
21.1 Some representative radar systems 278
21.2 Radar classification 281
21.3 Target characteristics and echo strengths 283
21.4 Pulse compression 285
21.5 Synthetic aperture radar 286
21.6 TR switches 288
21.7 Diode switches 291
21.8 Radar pulse modulators 293
Problems 297
References 298

22 Digital modulation techniques 300
22.1 Digital modulators 300
22.2 Pulse shaping 303
22.3 Root raised-cosine filter 307
22.4 8-VSB and GMSK modulation 308
22.5 Demodulation 309
22.6 Orthogonal frequency-division multiplexing – OFDM 310
22.7 Spread-spectrum and CDMA 315
Problems 318
Glossary 318
References 320

23 Modulation, noise, and information 321
23.1 Matched filtering 321
23.2 Analysis of a BPSK link 323
23.3 On–off keying with envelope detection 325
Problems 335
References 335

24 Amplifier and oscillator noise analysis 336
24.1 Amplifier noise analysis 336
24.2 Oscillator noise 342
24.3 Effect of nonlinearity 346
Problems 346
References 348

25 The GPS Navigation system 349
25.1 System description 349
25.2 GPS broadcast format and time encoding 350
25.3 GPS satellite transmitter 352
25.4 Signal tracking 353

25.5 Acquisition 356
25.6 Ionospheric delay 359
25.7 Differential GPS 360
25.8 Augmented GPS 361
25.9 Improvements to GPS 361
25.10 Other satellite navigation systems 362
Problems 362
References 363

26 Radio and radar astronomy 364
26.1 Radiometry 365
26.2 Spectrometry 366
26.3 Interferometry 366
26.4 Radar astronomy 368
Problems 374
References 374

27 Radio spectrometry 375
27.1 Filters and filterbanks 376
27.2 Autocorrelation spectrometry 376
27.3 Fourier transform spectrometry 381
27.4 I and Q mixing 384
27.5 Acousto-optical spectrometry 385
27.6 Chirp-z spectrometry 386
Problems 388
References 389

28 *S*-parameter circuit analysis 390
28.1 S-parameter definitions 390
28.2 Circuit analysis using S parameters 394
28.3 Stability of an active two-port (amplifier) 397
28.4 Cascaded two-ports 399
28.5 Reciprocity 400
28.6 Lossless networks 400
Problems 404
References 405

29 Power supplies 406
29.1 Full-wave rectifier 406
29.2 Half-wave rectifier 408
29.3 Electronically regulated power supplies 409
29.4 Three-phase rectifiers 410

29.5 Switching converters 411
Problems 419
References 421

30 RF test equipment 422

30.1 Power measurements 422
30.2 Voltage measurements 423
30.3 Spectrum analysis 424
30.4 Impedance measurements 425
30.5 Noise figure meter 432
Problems 432
References 433

Index 434

Preface

This book was written to help the reader to understand, analyze, and design RF circuits. Developed as a textbook for an RF engineering course at Cornell University, it can also be used for self-study and as a reference for practising engineers. The scope of topics is wide and the level of analysis ranges from introductory to advanced. In each chapter, I have tried to convey an intuitive "how things work" understanding from which the mathematical analysis follows. The initial chapters present the amplifiers, filters, modulators, and demodulators, which are the basic building blocks of radio systems, from AM and FM to the latest digital radio systems. Later chapters alternate between systems, such as television, and radio astronomy, and theoretical topics, such as noise analysis and radio spectrometry. The book provides the RF vocabulary that carries over into microwave engineering, and one chapter is devoted to waveguides and other microwave components.

In this second edition, many chapters have been expanded. Others have been rearranged and consolidated . New chapters have been added to cover radar, the GPS navigation system, digital modulation, information transmission, and *S*-parameter circuit analysis.

The reader is assumed to have a working knowledge of basic engineering mathematics and electronic circuit theory, particularly linear circuit analysis. Many students will have had only one course in electronics, so I have included some fundamental material on amplifier topologies, transformers, and power supplies. The reader is encouraged to augment reading with problem-solving and lab work, making use of mathematical spreadsheet and circuit simulation programs, which are excellent learning aids and confidence builders. Some references are provided for further reading, but whole trails of reference can be found using the internet.

For helpful comments, suggestions, and proofreading, I am grateful to many students and colleagues, especially Wesley Swartz, Dana Whitlow, Bill Sisk, Suman Ganguly, Paul Horowitz, Michael Davis, and Mario Ierkic.

Jon B. Hagen
Brooklyn, NY
July 2008

CHAPTER

1

Introduction

Consider the magic of radio. Portable, even hand-held, short-wave transmitters can reach thousands of miles beyond the horizon. Tiny microwave transmitters riding on spacecraft return data from across the solar system. And all at the speed of light. Yet, before the late 1800s, there was nothing to suggest that telegraphy through empty space would be possible even with mighty dynamos, much less with insignificantly small and inexpensive devices. The Victorians could extrapolate from experience to imagine flight aboard a steam-powered mechanical bird or space travel in a scaled-up Chinese skyrocket. But what experience would have even hinted at *wireless* communication? The key to radio came from theoretical physics. Maxwell consolidated the known laws of electricity and magnetism and added the famous displacement current term, $\partial D/\partial t$. By virtue of this term, a changing electric field produces a magnetic field, just as Faraday had discovered that a changing magnetic field produces an electric field. Maxwell's equations predicted that *electromagnetic waves* can break away from the electric currents that generate them and propagate independently through empty space with the electric and magnetic field components of the wave constantly regenerating each other.

Maxwell's equations predict the velocity of these waves to be $1/\sqrt{\varepsilon_0 \mu_0}$ where the constants, ε_0 and μ_0, can be determined by simple measurements of the forces between static electric charges and between current-carrying wires. The dramatic result is, of course, the experimentally-known speed of light, 3×10^8 m/sec. The electromagnetic nature of light is revealed. Hertz conducted a series of brilliant experiments in the 1880s in which he generated and detected electromagnetic waves with wavelengths very long compared to light. The utilization of Hertzian waves (electromagnetic waves) to transmit information developed hand-in-hand with the new science of electronics.

Where is radio today? AM radio, the pioneer broadcast service, still exists, along with FM, television and two-way communication. But radio now also includes digital broadcasting formats, radar, surveillance, navigation and broadcast satellites, cellular telephones, remote control devices, and wireless data communications. Applications of radio frequency (RF) technology outside

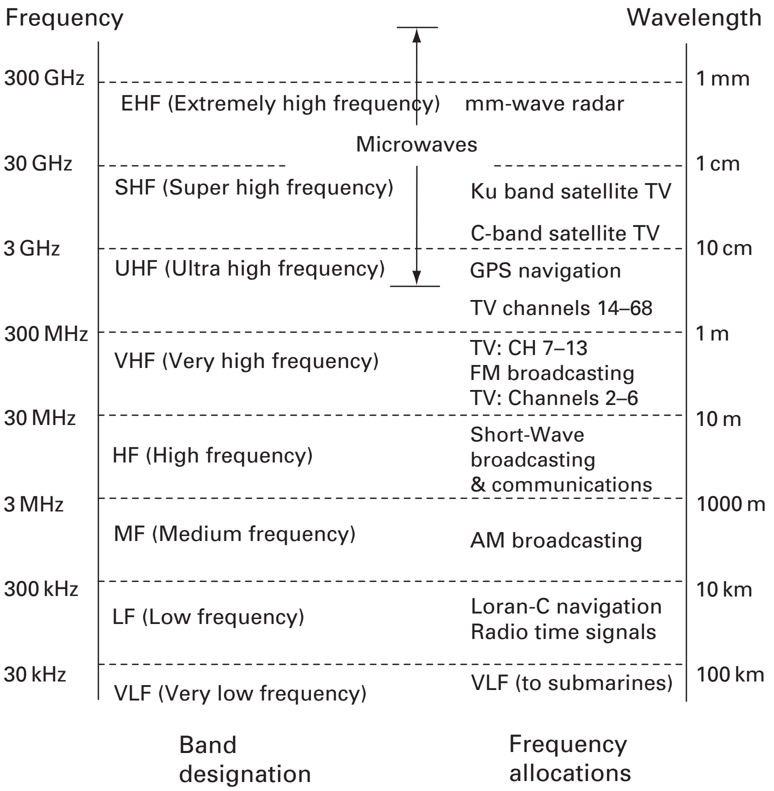
Frequency
Wavelength
300 GHz
1 mm
EHF (Extremely high frequency)
mm-wave radar
Microwaves
30 GHz
1 cm
SHF (Super high frequency)
Ku band satellite TV
C-band satellite TV
3 GHz
10 cm
UHF (Ultra high frequency)
GPS navigation
TV channels 14–68
300 MHz
1 m
VHF (Very high frequency)
TV: CH 7–13
FM broadcasting
TV: Channels 2–6
30 MHz
10 m
HF (High frequency)
Short-Wave broadcasting & communications
3 MHz
1000 m
MF (Medium frequency)
AM broadcasting
300 kHz
10 km
LF (Low frequency)
Loran-C navigation
Radio time signals
30 kHz
100 km
VLF (Very low frequency)
VLF (to submarines)
Band designation
Frequency allocations

Figure 1.1. The radio spectrum.

radio include microwave heaters, medical imaging systems, and cable television. Radio occupies about eight decades of the electromagnetic spectrum, as shown in Figure 1.1.

1.1 RF circuits

The circuits discussed in this book generate, amplify, modulate, filter, demodulate, detect, and measure ac voltages and currents at radio frequencies. They are the blocks from which RF systems are designed. They scale up and down in both power and frequency. A six-section bandpass filter with a given passband shape, for example, might be large and water-cooled in one application but subminiature in another. Depending on the frequency, this filter might be made of sheet metal boxes and pipes, of solenoidal coils and capacitors, or of piezoelectric mechanical resonators, yet the underlying circuit design remains the same. A class-C amplifier circuit might be a small section of an integrated circuit for a wireless data link or the largest part of a multi-megawatt broadcast transmitter. Again, the design principles are the same.

1.2 Narrowband nature of RF signals

Note that most frequency allocations have small fractional bandwidths, i.e., the bandwidths are small compared to the center frequencies. The fractional bandwidth of the signal from any given transmitter is less than 10 percent – usually much less. It follows that the RF voltages throughout a radio system are very nearly sinusoidal. An otherwise purely sinusoidal RF "carrier" voltage[1] must be *modulated* (varied in some way) to transmit information. Every type of modulation (audio, video, pulse, digital coding, etc.) works by varying the amplitude and/or the phase of the sinusoidal RF wave, called the "carrier" wave. An unmodulated carrier has only infinitesimal bandwidth; it is a pure spectral line. Modulation always broadens the line into a spectral band, but the energy clusters around the carrier frequency. Oscilloscope traces of the RF voltages in a transmitter or on a transmission line or antenna are therefore nearly sinusoidal. When modulation is present, the amplitude and/or phase of the sinusoid changes but only over many cycles. Because of this narrowband characteristic, elementary sine wave ac circuit analysis serves for most RF work.

1.3 AC circuit analysis – a brief review

The standard method for ac circuit analysis that treats voltages and currents in linear networks is based on the linearity of the circuit elements: inductors, capacitors, resistors, etc. When a sinusoidal voltage or current generator drives a circuit made of linear elements, the resulting steady-state voltages and currents will all be perfectly sinusoidal and will have the same frequency as the generator. Normally we find the response (voltage and current amplitudes and phases) of driven ac circuits by a mathematical artifice. We replace the given sinusoidal generator by a hypothetical generator whose time dependence is $e^{j\omega t}$ rather than $\cos(\omega t)$ or $\sin(\omega t)$. This source function has both a real and an imaginary part since $e^{j\omega t} = \cos(\omega t) + j\sin(\omega t)$. Such a nonphysical (because it is complex) source leads to a nonphysical (complex) solution. But the real and imaginary parts of the solution are separately good physical solutions that correspond respectively to the real and imaginary parts of the complex source. The value of this seemingly indirect method of solution is that the substitution of the complex source converts the set of linear *differential* equations into a set of easily solved linear *algebraic* equations. When the circuit has a simple topology, as is often the case, it can be reduced to a single loop by combining obvious series and parallel branches. Many computer programs are available to find the currents and voltages in complicated ac circuits. Most versions of SPICE will do this steady-state ac analysis (which is much simpler than the transient analysis which is their primary function). Special linear ac analysis programs for RF and microwave work such as Agilent's ADS and MMICAD include circuit models for strip lines, waveguides, and other RF components. You can write your own program to analyze ladder networks (see Problem 1.3) and to analyze most filters and matching networks.

---

[1] There is no low-frequency limit for radio waves but the wavelengths corresponding to audio frequencies, hundreds to thousands of kilometers, make it inefficient to connect an audio amplifier directly to an antenna of reasonable size. Instead, the information is impressed on a carrier wave whose wavelength is compatible with practical antennas.

1.4 Impedance and admittance

The coefficients in the algebraic circuit equations are functions of the complex *impedances* (V/I), or *admittances* (I/V), of the RLC elements. The voltage across an inductor is $L\mathrm{d}I/\mathrm{d}t$. If the current is $I_0\mathrm{e}^{\mathrm{j}\omega t}$, then the voltage is $(\mathrm{j}\omega L)I_0\mathrm{e}^{\mathrm{j}\omega t}$. The impedance and admittance of an inductor are therefore respectively $\mathrm{j}\omega L$ and $1/(\mathrm{j}\omega L)$. The current into a capacitor is $C\mathrm{d}V/\mathrm{d}t$, so its impedance and admittance are $1/(\mathrm{j}\omega C)$ and $\mathrm{j}\omega C$. The impedance and admittance of a resistor are just R and $1/R$. Elements in series have the same current so their total impedance is the sum of their separate impedances. Elements in parallel have the same voltage so their total admittance is the sum of their separate admittances. The real and imaginary parts of impedance are called *resistance* and *reactance* while the real and imaginary parts of admittance (the reciprocal of impedance) are called *conductance* and *susceptance*.

1.5 Series resonance

A capacitor and inductor in series have an impedance $Z_\mathrm{s} = \mathrm{j}\omega L+1/(\mathrm{j}\omega C)$. This can be written as $Z_\mathrm{s} = \mathrm{j}(L/\omega)(\omega^2 - 1/[LC])$, so the impedance is zero when the (angular) frequency is $1/\sqrt{LC}$. At this *resonant frequency*, the series LC circuit is a perfect *short circuit* (Figure 1.2). Equal voltages are developed across the inductor and capacitor but they have opposite signs and the net voltage drop is zero.

At resonance and in the steady state there is no transfer of energy in or out of this combination. (Since the overall voltage is always zero, the power, IV, is always zero.) However, the circuit does contain *stored* energy which simply sloshes back and forth between the inductor and the capacitor. Note that this circuit, by itself, is a simple bandpass filter.

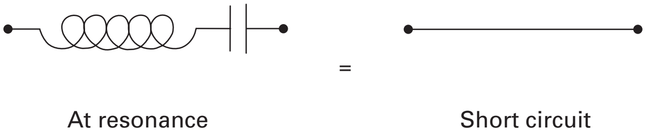

Figure 1.2. Series-resonant LC circuit.

1.6 Parallel resonance

A capacitor and an inductor in parallel have an admittance $Y_p = j\omega C + 1/(j\omega L)$ which is zero when the (angular) frequency is $1/\sqrt{LC}$. At this resonant frequency, the parallel *LC* circuit is a perfect *open circuit* (Figure 1.3) – a simple bandstop filter.

Like the series *LC* circuit, the parallel *LC* circuit stores a fixed quantity of energy for a given applied voltage. These two simple combinations are important building blocks in RF engineering.

Figure 1.3. Parallel-resonant *LC* circuit.

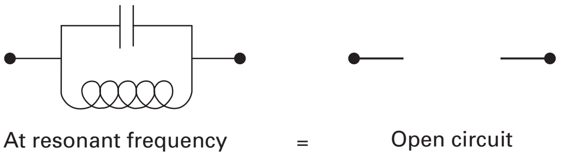

1.7 Nonlinear circuits

Many important RF circuits, including mixers, modulators, and detectors, are based on nonlinear circuit elements such as diodes and saturated transistors used as switches. Here we cannot use the linear $e^{j\omega t}$ analysis but must use time-domain analysis. Usually the nonlinear elements can be replaced by simple models to explain the circuit operation. Full computer modeling can be used for accurate circuit simulations.

Problems

Problem 1.1. A generator has a source resistance r_S and an open circuit rms voltage V_0. Show that the maximum power available from the generator is given by $P_{max} = V_0^2/(4r_S)$ and that this maximum power will be delivered when the load resistance, R_L, is equal to the source resistance, r_S.

Problem 1.2. A passive network, for example a circuit composed of resistors, inductors, and capacitors, is placed between a generator with source resistance r_S and a load resistor, R_L. The *power response* of the network (with respect to these resistances) is defined as the fraction of the generator's maximum available power that reaches the load. If the network is *lossless*, that is, contains no resistors or other dissipative elements, its power response function can be found in terms of the impedance, $Z_{in}(\omega) = R(\omega) + jX(\omega)$, seen looking into the network with the load connected. Show that the expression for the power response of the lossless network is given by

$$P(\omega) = \frac{4r_S R(\omega)}{(R(\omega) + r_S)^2 + X(\omega)^2}$$

where $R = \text{Re}(Z_{in})$ and $X = \text{Im}(Z_{in})$.

Problem 1.3. Most filters and matching networks take the form of the ladder network shown below.

Ladder network topology.

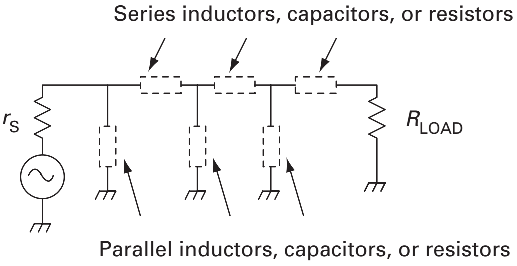
Series inductors, capacitors, or resistors
r_S
R_{LOAD}
Parallel inductors, capacitors, or resistors

Write a program whose input data is the series and shunt circuit elements and whose output is the power response as defined in Problem 1.2.

Hints: One approach is to begin from the load resistor and calculate the input impedance as the elements are added, one by one. When all the elements are in place, the formula in Problem 1.2 gives the power response – as long as the load resistor is the only resistor. The process is repeated for every desired frequency.

A better approach, which is no more complicated and which allows resistors, is the following: Assume a current of 1 + j0 ampere is flowing into the load resistor. The voltage at this point is therefore $R_L + j0$ volts. Move to the left one element. If this is a series element, the current is unchanged but the voltage is higher by IZ where Z is the impedance of the series element. If the element is a shunt element, the voltage remains the same but the input current is increased by VY where Y is the admittance of the shunt element. Continue adding elements, one at a time, updating the current and voltage. When all the elements are accounted for, you have the input voltage and current and could calculate the total input impedance of the network terminated by the load resistor. Instead, however, take one more step and treat the source resistance, r_S, as just another series impedance. This gives you the voltage of the source generator, from which you can calculate the maximum power available from the source. Since you already know the power delivered to the load, $(1)^2 R_L$, you can find the power response. Repeat this process for every desired frequency.

The ladder elements (and, optionally, the start frequency, stop frequency, frequency increment, and source and load resistances) can be treated as data, that is, they can be located together in a block of program statements or in a file so they can be changed easily. For now, the program only needs to deal with six element types: series and parallel inductors, capacitors, and resistors. Each element in the circuit file must therefore have an identifier such as “PL”, “SL”, “PC”, “SC”, “PR”, and “SR” or 1, 2, 3, 4, 5, 6, or whatever, plus the value of the component in henrys, farads, or ohms. Organize the circuit file so that it begins with the element closest to R_L and ends with some identifier such as “EOF” (for “End Of File”) or some distinctive number.

An example program, which produces both tabular and graphical output, is shown below, written in MATLAB, which produces particularly compact and readable code. The input data (included as program statements) is for the circuit shown below, of an LC network designed to connect a 50-ohm load to a 1000-ohm source. You will find this, or your own equivalent program, to be a useful tool when designing matching networks and filters. In the problems for Chapters 4, 10, 14, and 17, the program will be enhanced to plot phase response and to handle transmission lines, transformers, and transistors, making it a powerful design tool.

```
%MATLAB program to solve ladder networks
%Problem 1.3 in “Radio-Frequency Electronics”
%Save this file as “ladder.m” and run by typing “ladder” in command window
%INPUT DATA(circuit components from load end;‘SL’ is series inductor,%etc.)
%——————————————————————————
ckt={‘SL’,23.1e-6,‘PC’,463e-12,‘EOF’}; %‘EOF’ terminates list
Rload=50; Rsource=1000; startfreq=1e6; endfreq=2e6; freqstep= 5e4;
%——————————————————————————

f=(startfreq:freqstep:endfreq); % frequency loop
w=2*pi.*f; %w is angular frequency
I=ones(size(w));V=ones(size(w))*Rload;%set up arrays for inputI(f) and V(f)
ckt_index=0; morecompsflag=1;
while morecompsflag == 1 %loop through string of components
ckt_index=ckt_index+1; %ckt_index prepared for next item in list
component=ckt{ckt_index};
morecompsflag=1-strcmp(component,‘EOF’); %zero after last component

if strcmp(component,‘PC’)==1
ckt_index=ckt_index+1; capacitance=ckt{ckt_index};
I=I+V.*(1j.*w.*capacitance);
elseif strcmp(component,‘SC’)==1
ckt_index=ckt_index+1; capacitance=ckt{ckt_index};
V=V+I./(1j.*w.*capacitance);
elseif strcmp(component,‘PL’)==1
ckt_index=ckt_index+1; inductance=ckt{ckt_index};
I=I+V./(1j.*w.*inductance);
elseif strcmp(component,‘SL’)==1
ckt_index=ckt_index+1; inductance=ckt{ckt_index};
V=V+I.*(1j.*w.*inductance);
elseif strcmp(component,‘PR’)==1
ckt_index=ckt_index+1; resistance=ckt{ckt_index};
I=I+V/resistance;
elseif strcmp(component,‘SR’)==1
ckt_index=ckt_index+1; resistance=ckt{ckt_index};
V=V+I*resistance;
end %components loop

end %frequency loop

Z=V./I; V=V+I.*Rsource;
frac=Rload./((abs(V).^2)/(4.*Rsource));
db=10/log(10)*log(frac);
heading = ‘freq(MHz) frac dB’ %print heading in command window
A=[(1E-6*f)’ frac’ db’] %print table of data in command window
plot(f,db); %graph the data
grid;xlabel(‘Frequency’);ylabel(‘dB’);title(‘Frequency response’);
```

The circuit corresponding to the input data statements in the example program above is shown below, together with the analysis results produced by the program.

Example circuit.

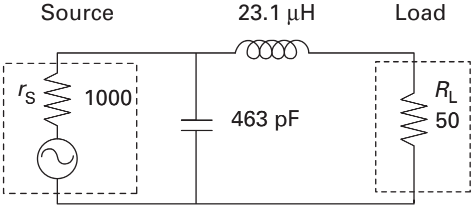

Analysis results.

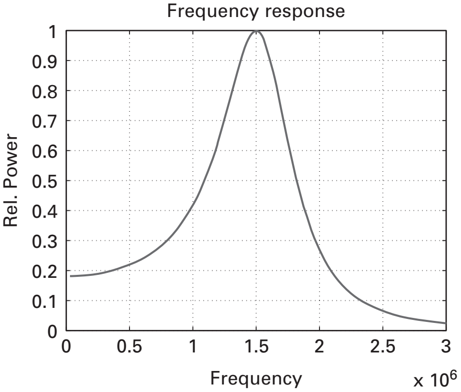

Problem 1.4. (AC circuit analysis review problem.) For the circuit in the figure, derive an expression for $I_R(t)$. Use a complex source voltage, $V_0e^{j\omega t}$, the real part of which is $V_0\cos(\omega t)$. The impedances of C, L, and R are $(j\omega C)^{-1}, j\omega L$, and R, respectively. Find the complex current through the resistor. $I_R(t)$ will be the real part of this complex current.

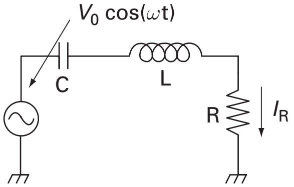

Problem 1.5. (More AC circuit analysis review.) For the circuit in the figure, derive an expression for $I_R(t)$. Note that the source voltage, $V_0 \sin(\omega t + \theta)$, is equal to the *imaginary* part of $V_0 e^{j(\omega t+\theta)}$. Therefore, if we take the complex voltage to be $V_0 e^{j(\omega t+\theta)}$, $I_R(t)$ will be the imaginary part of the complex current through R. Alternatively, you can let the complex voltage source have the value $-jV_0 e^{j(\omega t+\theta)}$, the real part of which is $V_0 \sin(\omega t)$. With this source, $I_R(t)$ is the real part of the complex current through R.

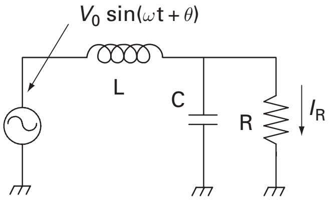

CHAPTER

2

Impedance matching

Matching normally means the use of a lossless (nonresistive) network between a signal source and a load in order to maximize the power transferred to the load. This presupposes that the source is not capable of supplying infinite power, i.e., that the source is not just an ideal voltage generator or an ideal current generator. Rather, the source is assumed to be an ideal voltage generator in series with a source impedance, i.e., a Thévenin equivalent circuit, or an ideal current generator in parallel with a source admittance, a Norton equivalent circuit. Note that these equivalent circuits are themselves equivalent; each can be converted into the form of the other. An antenna that is feeding a receiver is an example of an ac signal source connected to a load. Figure 2.1 shows the simplest situation, a dc generator driving a resistive load. The generator is represented in Thévenin style (a) and in Norton style (b).

Figure 2.1. DC generator driving a resistive load, Equivalent circuits.

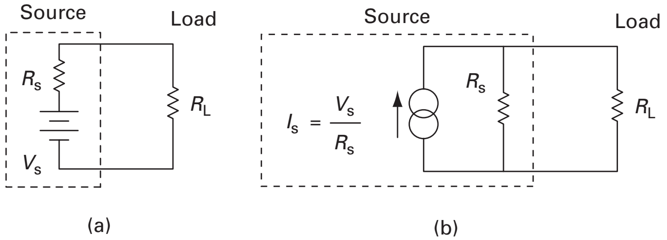

You can see the equivalence by inspection: the generators have the same open-circuit voltage and the same short-circuit current. Maximum power is transferred when the load resistance is made equal to the source resistance. You can show this by differentiating the expression for the power, $P_{\text{load}} = [V_S\, R_L/(R_L{+}R_S)]^2/R_L$. Figure 2.2 plots the relative transferred power (Pwr/MaxPwr) as a function of the normalized load resistance ($r = R_L/R_S$). In (a) the scales are linear and in (b) the scales are logarithmic so the relative power is expressed in dB. Note that R_L can differ by a factor of 10 from R_S and the power transferred is still 33% of the maximum value.

Figure 2.3. Transformer converts R_L to R_S for maximum power transfer.

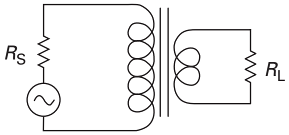

2.1 Transformer matching

Figure 2.2. Relative power transfer as a function of R_L/R_S, (=r)

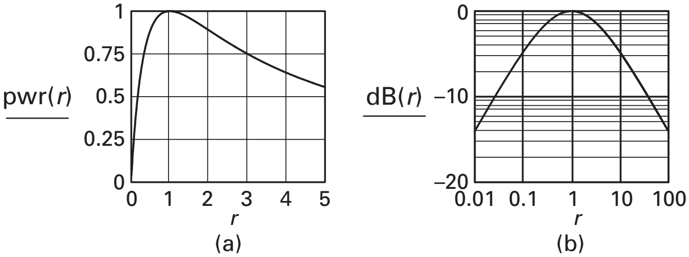

(a)

(b)

In the case of an ac source, a transformer can make the load resistance match (equal) the source resistance (and vice versa) as shown in Figure 2.3. The impedance is transformed by the square of the turns ratio.[1]

The ac situation often has a complication: the source and/or the load may be reactive, i.e., have an unavoidable built-in reactance. An example of a reactive load is an antenna; most antennas are purely resistive at only one frequency. Above this resonant frequency they usually look like a resistance in series with an inductor and below the resonant frequency they look like a resistance in series with a capacitor. An obvious way to deal with this is first to cancel the reactance to make the load and/or source impedance purely resistive and then use a transformer to match the resistances. In the circuit of Figure 2.4, an inductor cancels

Figure 2.4. A series reactor makes the load a pure resistance.

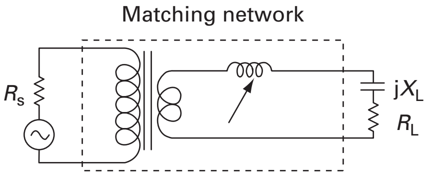

---

[1] Let the secondary winding be the load side. Then $V_{sec} = V_{pri}\, N_{sec}/N_{pri}$. For energy to be conserved, $V_{pri}I_{pri} = V_{sec}I_{sec}$. Therefore $I_{sec} = I_{pri}\, N_{pri}/N_{sec}$ and $V_{pri}/I_{pri} = (V_{sec}/I_{sec})\,(N_{pri}/N_{sec})^2$ or $Z_{pri} = Z_{sec}(N_{pri}/N_{sec})^2$.

the reactance of a capacitive (but not purely capacitive) load. If we are working at 60 Hz, we would say the inductor corrects the load's power factor.

From the standpoint of the load, the matching network converts the source impedance, $R_S + j0$, into the complex conjugate of the load impedance. When a matching network is used between two devices, each device will look into an impedance that is the complex conjugate of its own impedance. As a result, the reactances cancel and the resistances are equal. Whenever the source and/or load have a reactive component, the match will be frequency dependent, i.e., away from the design frequency the match will not be perfect. In fact, with reactive sources and/or reactive loads, *any* lossless matching circuit will be frequency dependent – a filter of some kind – whether we like it or not.

2.2 L-networks

More often than not, matching circuits use no transformers (i.e., no coupled inductors). Figure 2.5 shows a two-element L-network (in this figure, a rotated letter L) that will match a source to a load resistor whose resistance is smaller than the source resistance. The trick is to put a reactor, X_P, in *parallel* with the *larger* resistance. Consider a specific example: $R_S = 1000$ and $R_L = 50$.

The impedance of the left-hand side is given by

$$Z_{\text{left}} = R_{\text{left}} + jX_{\text{left}} = \frac{1000\,jX_P}{1000 + jX_P} = \frac{(1000jX_P)(1000 - jX_P)}{(1000 + jX_P)(1000 - jX_P)}$$

$$= \frac{1000^2 jX_P + 1000X_P^2}{1000^2 + X_P^2}. \tag{2.1}$$

We can pick the value of X_P so that the real part of Z_{left} will be 50 ohms, i.e., equal to the load resistance. Using Equation (2.1), we find that $X^2_P = 52\,628$ so we can pick either $X_P = 229$ (an inductor) or $X_P = -229$ (a capacitor). The left-hand side now has the correct equivalent series resistance, 50 ohms, but it is accompanied by an equivalent series reactance, X_{left}, given by the imaginary part of Equation (2.1). We can cancel X_{left} by inserting a series reactor, X_S, equal to $-X_{\text{left}}$. Figure 2.6 shows the matching circuits that result when X_P is an inductor and when X_P is a capacitor.

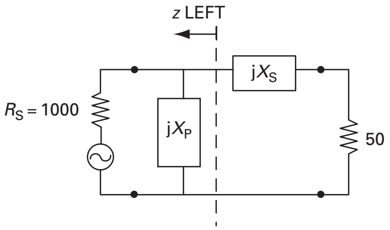

Figure 2.5. Two reactors in an L-network match R_L to R_S.

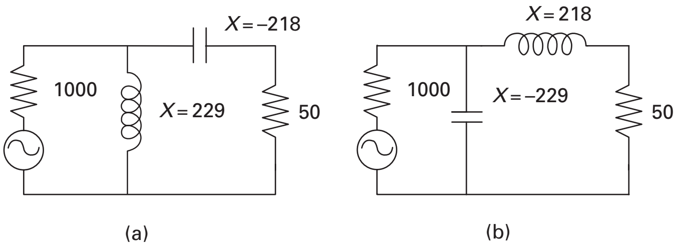

Figure 2.6. The two realizations for the L-network of Figure 2.5.

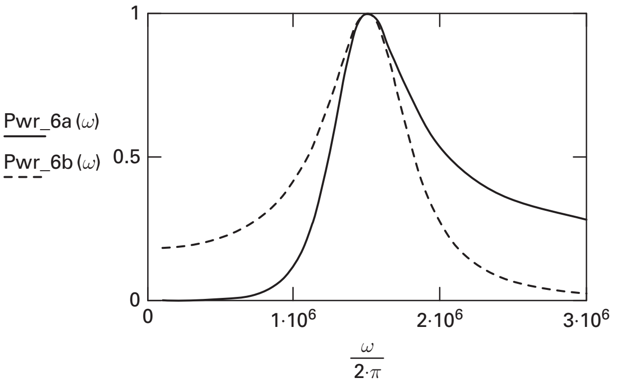

Figure 2.7. Frequency response (power vs. frequency) for the L-networks of Figure 2.6.

The final step is to find the values of L and C that produce the specified reactances at the given frequency. For the circuit of Figure 2.6(b), $\omega L = 218$. Suppose the design frequency is 1.5 MHz ($\omega = 2\pi \cdot 1.5 \cdot 10^6$), near the top of the AM broadcast band. Then $L = 23.1\ \mu\text{H}$ and $C = 462\ \text{pF}$. Note that the values of the two reactors are completely determined by the source and load resistances. Except for the choice of which element is to be an inductor and which is to be a capacitor, there are no free parameters in this two-element matching circuit. The match is perfect at the design frequency but, away from that frequency, we must accept the resulting frequency response. The frequency responses (fractional power reaching the load vs. frequency) for the two circuits of Figure 2.6 are plotted in Figure 2.7. Note that around the design frequency, i.e., around the resonant peak, the curves are virtually identical. Otherwise, the complete cutoff at very low frequencies of Figure 2.6a and the complete cutoff at very high frequencies of Figure 2.6b can be predicted from inspection of the circuits.

Quick design procedure for L-networks

If you remember only that the parallel reactance goes across the larger resistance you will be able to repeat the steps used above and design L-networks. But if you are doing these things often it may be worth memorizing the following "Q factor" for L-network design:

$$Q_{\mathrm{EL}} = \sqrt{\frac{R_{\mathrm{high}}}{R_{\mathrm{low}}} - 1}. \tag{2.2}$$

You can verify (Problem 2.6) that the ratios $R_{\mathrm{high}}/X_{\mathrm{parallel}}$ and $X_{\mathrm{series}}/R_{\mathrm{low}}$ are both equal to this factor, Q_{EL}. Remember the definition of Q_{EL} and these ratios immediately give you the L-network reactance values. You can also verify that, when Q_{EL} is large, the two elements in an L-network have nearly equal and opposite reactances, i.e., together they resonate at the design frequency. In this case the magnitude of the reactances is given by the geometric mean of R_{high} and R_{low} (especially easy to remember).

When the ratio of the source resistance to the load resistance is much different from unity, an L-network produces a narrowband match, i.e., the match will be good only very close to the design frequency. Conversely, when the impedance ratio is close to unity, the match is wide. The width of any resonance phenomenon is described by a factor, *the effective Q* (or *circuit Q* or just *Q*), which is equal to the center frequency divided by the two-sided 3-dB bandwidth (the difference between the half-power points). Equivalently, Q_{eff} is the reciprocal of the fractional bandwidth. When an ideal voltage generator drives a simple *RLC* series circuit, Q_{eff} is given by X/R where X is either X_L or X_C at the center frequency (since they are equal). The L-network matching circuit is equivalent to a simple series *RLC* circuit, but Q_{EL} is twice Q_{eff} because the nonzero source resistance is also in series; the matching circuit makes the effective source resistance equal to the load resistance so the loop's total series resistance is twice the load resistance. As a result, the fractional bandwidth is given by $1/Q_{\mathrm{eff}} = 2/Q_{\mathrm{EL}}$. In many applications the bandwidth of the match is important and the match provided by the L-network (which is completely determined by the source and load resistances) may be too narrow or too wide. When matching an antenna to a receiver, for example, one wants a narrow bandwidth so that signals from strong nearby stations won't overload the receiver. In another situation the signal produced by a modulated transmitter might have more bandwidth than the L-network would pass. Networks described below solve these problems.

2.3 Higher *Q* – pi and T-networks

Higher *Q* can be obtained with back-to-back L-networks, each one transforming down to a center impedance that is lower than either the generator or the source resistance. The resulting pi-network is shown in Figure 2.8.

With the simple L-networks we had $Q_{\mathrm{EL}} = \sqrt{19} = 4.4$. In this pi-network both the 1000-ohm source and the 50-ohm load are matched down to a center impedance of 10 ohms (a free parameter). The bandwidth is equivalent to that of an L-network with $Q_{\mathrm{EL}} = 11.95$. When $R_{\mathrm{HIGH}} \gg R_{\mathrm{LOW}}$, the pi-network has a bandwidth equivalent to that of an L-network with $Q_{\mathrm{EL}} = \sqrt{R_{\mathrm{HIGH}}/R_{\mathrm{CENTER}}}$. Again, the fractional bandwidth is given by $1/Q_{\mathrm{eff}} = 2/Q_{\mathrm{EL}}$.

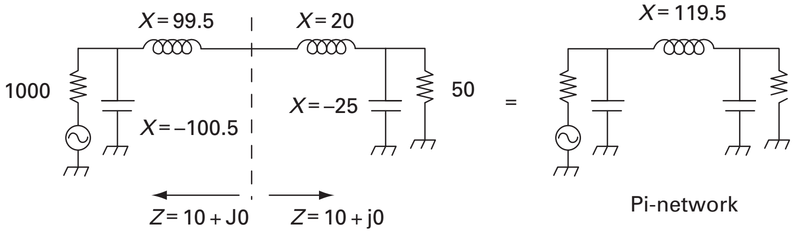

Figure 2.8. Pi-network (back-to-back L-networks) provides higher *Q*.

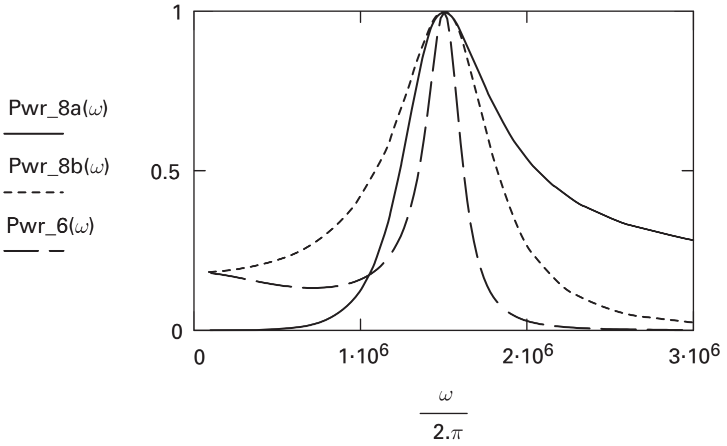

Figure 2.9. Response of the pi-network of Figure 2.8 compared with the L-networks of Figure 2.6.

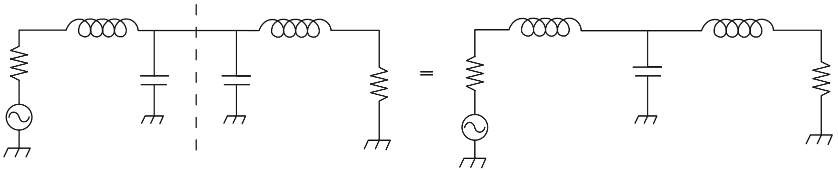

Figure 2.10. The T-network, like the pi-network, provides higher *Q*.

The response of this pi-network is shown in Figure 2.9 together with the responses of the L-networks of Figure 2.6.

You can guess that we could just as well have used "front-to-front" L-networks, each one transforming up to a center impedance that is higher than both the source and load impedances. This produces the T-network of Figure 2.10. Note that both the pi-network and the T-network have a free parameter (the center impedance) which gives us some control over the frequency response while still providing a perfect match at the center or design frequency.

2.4 Lower *Q* – the double L-network

In a double L-network (Figure 2.11) the first stage transforms to an impedance value between the source and load impedances. The second stage takes it the

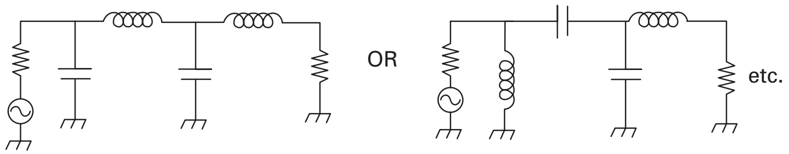

Figure 2.11. Double L-network for lower Q (wider bandwidth).

rest of the way. The process can, of course, be done in smaller steps with any number of cascaded networks. A long chain of L-networks forms an artificial transmission line that tapers in impedance to produce a frequency-independent match. Real transmission lines (i.e., lines with distributed L and C) are sometimes physically tapered to provide this kind of impedance transformation. A tapered transmission line is sometimes called a transformer, since, like the transformer in Figure 2.3, it provides frequency-independent matching.

2.5 Equivalent series and parallel circuits

To design the L-network we used the fact that a two-element parallel XR circuit, where $1/Z = 1/R_{\text{parallel}} + 1/\mathrm{j}X_{\text{parallel}}$, has an equivalent series circuit, where $Z = R_{\text{series}} + \mathrm{j}X_{\text{series}}$. Conversion between equivalent series and parallel representations is used so often it is worth a few more words. If you are given, for example, an antenna or a black box with two terminals and you make measurements at a *single frequency* you can only determine whether the box is "capacitive," i.e., equivalent to an RC combination, or is "inductive," i.e., equivalent to an RL combination. Suppose it is capacitive. Then you can represent it equally well as a series circuit where $Z = R_{\text{series}} + 1/\mathrm{j}\omega C_{\text{series}}$ or as a parallel circuit where $1/Z = 1/R_{\text{parallel}} + \mathrm{j}\omega C_{\text{parallel}}$. As long as you're working only at (or never very far from) the single frequency, either representation is equally valid, even if the box contains a complicated circuit with discrete resistors, capacitors, inductors, transmission lines, metallic and resistive structures, etc. If you measure the impedance at more than one frequency you might determine that the box does indeed contain a simple parallel RC or series RC circuit or that its impedance variation at least resembles that of a simple parallel circuit more than it resembles that of a simple series circuit.

2.6 Lossy components and efficiency of matching networks

So far we have considered networks made of ideal inductors and capacitors. Real components, however, are lossy due to the finite conductivity of metals, lossy dielectrics or magnetic materials, and even radiation. Power dissipated in nonideal components is power that does not reach the load so, with lossy components, we must consider a matching network's efficiency. As explained above, a lossy reactor can be modeled as an ideal L or C together with either a series or parallel resistor. Normally we can make the approximation that the values of L or C and the value of the associated resistor are constant throughout the band of interest. Let us consider the efficiency of the L-network that uses a series inductor and a parallel capacitor. We shall assume that the loss in the capacitor is negligible compared to the loss in the inductor. (This is very often the case with lumped components.) We shall model the lossy inductor as an ideal inductor in series with a resistor of value r_S. The ratio of the inductive reactance, X_L, to this resistance value is the *quality factor*, Q_U, where the subscript denotes "*unloaded Q*" or *component Q*. (Less series resistance certainly implies a higher quality component.) Note that this resistance, like the inductor, is in series with the load resistor so the same current, I, flows through both. The power delivered to the load is I^2R_L and the power dissipated in r_S is I^2r_S. Using the relations $X_S = Q_{EL}R_{load}$ and $Q_U = X_S/r_S$, we find the efficiency of the match is given by

$$\eta = \text{Efficiency} = \frac{\text{Power Out}}{\text{Power In}} = \frac{I^2R_L}{I^2R_L + I^2r_S} = \frac{1}{1 + Q_{EL}/Q_U}. \tag{2.3}$$

Efficiency is maximized by maximizing the ratio Q_U/Q_{EL}, i.e., the ratio of unloaded Q to loaded Q. If we model the lossy inductor as a parallel LR circuit and define the unloaded Q as r_P/X we would get the same expression for efficiency (Problem 2.7). Likewise, if the loss occurs in the capacitor we will also get this expression, as long as we define the unloaded Q of the capacitor again as parallel resistance over parallel reactance or as series reactance over series resistance. When the load resistance is very different from the source resistance, the effective Q of an L-network will be high so, for high efficiency, the unloaded Q of the components must be very high. The double L-network, with its lower loaded Q's, can be used to provide higher efficiency.

Q factor summary

Loaded Q, the Q factor associated with *circuits*, can be either high or low depending on the application. Narrowband filters have high loaded Q. Wideband matching circuits have low loaded Q. Loaded Q is therefore not a measure of quality. Unloaded Q, however, which specifies the losses in *components*, is indeed a measure of quality since lowering component losses always increases circuit efficiency.

Problems

Problem 2.1. A nominal 47-ohm, $\frac{1}{4}$-watt carbon resistor with 1.5 inch wire leads is measured at 100 MHz to have an impedance of 48+j39 ohms. Find the component values for (a) an equivalent series RL circuit, and (b) an equivalent parallel RL circuit.

Problem 2.2. (a) Design an L-network to match a 50-ohm generator to a 100-ohm load at a frequency of 1.5 MHz. Let the parallel element be an inductor. Use your circuit analysis program (Problem 1.3) to find the frequency response of this circuit from 1 MHz to 2 MHz in steps of 50 kHz.

(b) Same as (a), but let the parallel element be a capacitor.

Problem 2.3. Design a double L matching network for the generator, source, and frequency of Problem 2.2(a). For maximum bandwidth, let the intermediate impedance be the geometric mean of the source impedance and the load impedance, i.e., $\sqrt{50 \cdot 100}$. Use your circuit analysis program (Problem 1.3) to find the response as in Problem 2.2.

Problem 2.4. Suppose the only inductors available for building the networks of Problems 2.2(a) and 2.3 have a Q_U (unloaded Q) of 100 at 1.5 MHz. Assume the capacitors have no loss. Calculate the efficiencies of the matching networks at 1.5 MHz. Check your results using your circuit analysis program.

Problem 2.5. The diagram below shows a network that allows a 50-ohm generator to feed two loads (which might be antennas). The network divides the power such that the top load receives twice as much power as the bottom load. The generator is matched, i.e., it sees 50 ohms. Find the values of X_{L1} , X_{L2} and X_C. Hint: transform each load first with an L-section network and then combine the two networks into the circuit shown.

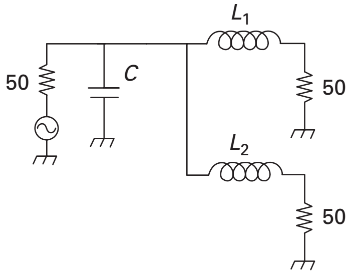

Problem 2.6. Verify the prescription given for calculating the values of an L-network: $X_P = \pm R/Q$ and $X_S = \mp rQ$ where $R > r$ and $Q = \sqrt{R/r - 1}$.

Problem 2.7. At a single frequency, a lossy inductor can be modeled as a lossless inductor in series with a resistance or as a lossless inductor in parallel with a resistance. Convert the series combination r_S, L_S to its equivalent parallel combination r_P, L_P and show that Q_U defined as X_S/r_S is equal to Q_U defined as r_P/X_P.

CHAPTER

3

Linear power amplifiers

An amplifier is a circuit designed to impose a specified voltage waveform, $V(t)$, or, sometimes, a specified current waveform, $I(t)$, upon the terminals of a device known as the "load." The specified waveform is often supplied in the form of an analog "input signal" such as the millivolt-level signal from a dynamic microphone. In a public address system, an audio amplifier produces a scaled-up copy of the microphone voltage (the input signal) and this amplified voltage (the output signal) is connected to a loudspeaker (the load). An audio amplifier generally supplies more than a watt to the loudspeaker. The microphone cannot supply more than milliwatts, so the audio amplifier is a power amplifier as well as a voltage amplifier. The ability to amplify power is really the defining characteristic of an amplifier. Of course energy must be conserved; amplifiers contain or are connected to power supplies, usually batteries or power line-driven ac-to-dc converter circuits, originally known as "battery eliminators" but long since simply called "power supplies" (see Chapter 29). The amplifiers discussed in this chapter are the basic "resistance-controlled"[1] circuits in which transistors (or vacuum tubes) are used as electronically variable resistors to control the current through the load. Such circuits span the range from monolithic op-amps to the output amplifiers in high-power microwave transmitters.

3.1 Single-loop amplifier

Figure 3.1 shows the simplest resistance-controlled amplifier. This circuit is just a resistive voltage divider. The manually variable resistor (rheostat) in (a) represents the electronically variable resistor (transistor) in (b). Remember that the main current path through the transistor is between the emitter and the

---

[1] Resistance-controlled amplifier are also called *linear* amplifiers, to distinguish them from *switching* amplifiers, which are discussed in Chapters 9 and 29. Note, however, that *linear amplifier* is also used to denote amplifiers whose output waveform is a faithful (linearly proportional) copy of the input.

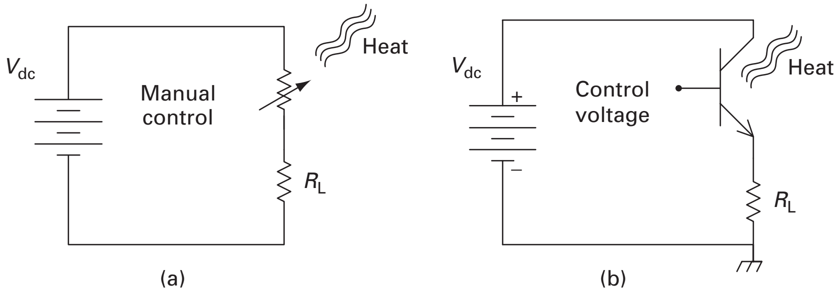

Figure 3.1. Basic single-loop amplifier.

collector; the current through the control terminal, the base, is typically less than 1% of the emitter-collector current. The base voltage can vary the transistor's resistance from infinity to almost zero so any arbitrary current waveform (within the range of zero to V_{dc}/R_L) can be obtained by using an appropriate corresponding control voltage waveform. The load, represented as a resistor, R_L, could be, for example a heating element, a loudspeaker, a servomotor, or a transmitting antenna.

3.2 Drive circuitry: common-collector, common-emitter, and common-base

We will concentrate on the topologies of the output circuit or "business end" of amplifiers, i.e., the high-current paths between the load and the power supply(s). But let us briefly discuss the drive circuitry that controls the resistance of the transistor(s). The discussion is illustrated with circuits using bipolar transistors, but the basic concepts apply also to FETs and tubes.

In Figure 3.1(b), only one terminal is shown for the control voltage (drive signal) input. When the return connection for the drive signal is made at the bottom of the load resistor, we get the amplifier shown in Figure 3.2. This circuit is called a *common-collector* amplifier because the collector, in common with the drive signal return, is a ground point with respect to *ac* signals, even though it does have a dc voltage.

This circuit is also called an *emitter follower*. The transistor will adjust its current flow to make the instantaneous emitter voltage almost equal to the instantaneous base voltage. Here "almost identical" means a small dc offset (the emitter voltage will be about 0.7 volts less than the base voltage) along with a one or two percent reduction in signal amplitude. The reason the emitter voltage closely follows the base voltage is that the emitter current is a rapidly increasing (exponential) function of the base-to-emitter voltage.

Figure 3.3 shows the *common-emitter* drive arrangement, in which the return for the drive signal is connected to the transistor's emitter. It is common to rearrange the circuit as in (b), so that the return connection for the drive signal and the negative terminal of the supply are at the same point (ground).

Figure 3.2. The emitter follower.

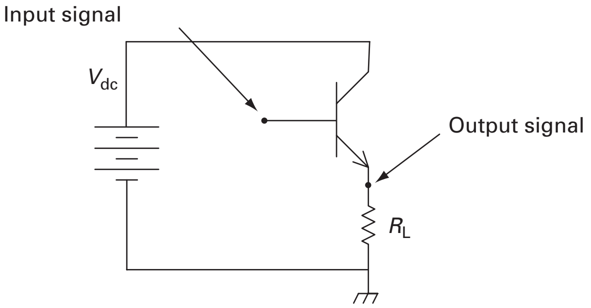

Figure 3.3. Common-emitter amplifier.

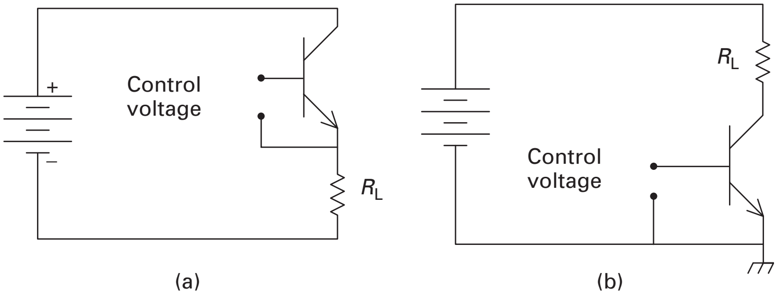

Figure 3.4. Common-base amplifier.

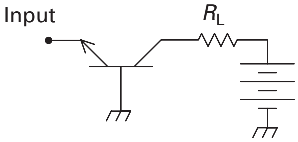

With the drive voltage placed directly across the base-emitter junction, the transistor current is a nonlinear (exponential) function of the drive voltage, and the output voltage (voltage across the load) will not be an accurate scaled version of the drive voltage. A special driver circuit can be used to generate an inverse exponential (logarithmic) drive signal to linearize the amplifier. This happens automatically if the base drive is a current waveform; the output current (and hence the voltage across the load) will have the same waveform shape as the base current. It is also common to use a negative feedback correction loop to force the output signal to follow the input signal. The common-emitter amplifier can supply voltage amplification as well as power amplification.

The third and final base drive arrangement, common-base, is shown in Figure 3.4. In this circuit the drive current flows through the main loop. Gain is obtained because, while the driver and load have essentially the same current (the base current might be only one percent as large as the collector-emitter current), the voltage swing at the collector, determined by the supply voltage, is much greater than the voltage swing at the emitter, determined by the near short-circuit base-emitter junction.

Note that, as in the common-emitter amplifier, the drive signal is applied across the transistor's base-to-emitter junction, so the signal developed across the load resistor will be a nonlinear function of the input voltage. But if the drive voltage is applied to the emitter through a series resistor, the drive current and, hence, the output current, will be essentially proportional to the drive voltage.

3.3 Shunt amplifier topology

In the amplifiers discussed above, the load current is controlled by a transistor in series with the load and the power supply. Another way to vary the current in the load is to divert current around it, as in the shunt circuit amplifier shown in Figure 3.5, where the supply and a series resistor form a (nonideal) current source.

Figure 3.5. Shunt amplifier.

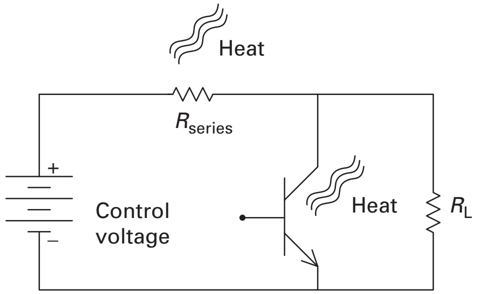

This shunt circuit appears inferior to the series circuit; power will be wasted in the series resistor and the full supply voltage is not available to the load. We will see later, however, that shunt circuits can be used to advantage in *ac* amplifiers which, unlike the general-purpose amplifiers above, are amplifiers designed for signals whose average dc value is zero, e.g., audio and RF signals.

3.4 Dual-polarity amplifiers

If the amplifier must supply output voltages of either polarity and also must handle arbitrary waveforms (as opposed to ac waveforms, whose average dc value is zero) it will require a circuit with two power supplies (or a single "floating" power supply, as we shall see later). The two-loop circuit shown in Figure 3.6 still uses only one transistor.

Here R_B pulls the output toward the negative supply as much as the transistor allows. The voltage on the load is determined by a tug of war between R_B and the transistor. Note that this circuit is a combination of the series and shunt amplifier arrangements. If we make $V_{NEG} = -2\ V_{POS}$ and $R_B = R_L$, you can see

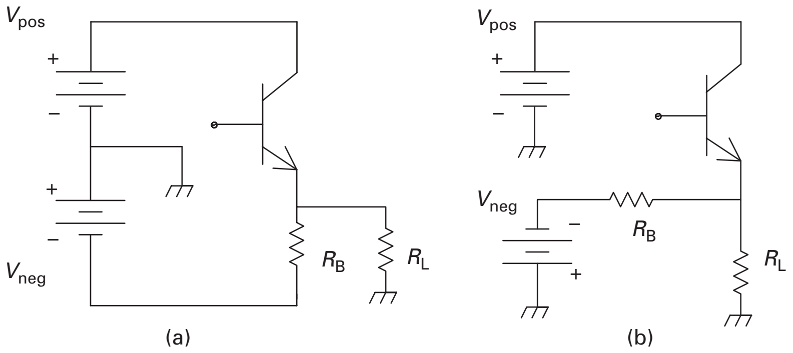

(a) (b)

Figure 3.6. A dual-supply, single-transistor amplifier (drawn in two ways).

that the maximum negative swing will be equal to the maximum positive swing. A constant *bias* current is maintained in the transistor to set the output voltage at zero when the input signal is zero. The maximum efficiency is calculated in the next section, and is only $\frac{1}{12}$. Biased amplifiers (this one and everything discussed so far), which draw current from the supply(s) even when the input signal is zero, are known as *class-A* amplifiers. They are commonly used where their low efficiency is not a problem.

3.5 Push–pull amplifiers

The two-transistor *push–pull* configuration shown in Figure 3.7 provides output voltage of both polarities and has high efficiency compared to the single-transistor amplifiers. (Note: non-push–pull amplifiers are often called “single-ended” amplifiers.)

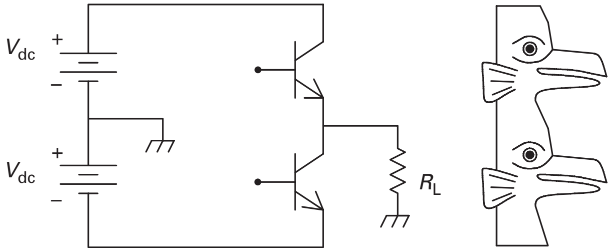

Figure 3.7. Totem pole push–pull amplifier.

The top transistor allows the top supply to “push” current into the load. The lower transistor lets the lower supply “pull” current from the load. The push–pull circuit is the circuit of choice for arbitrary waveforms. The efficiency, calculated in the next section, is $\pi/4$ (78%) for a sine wave of maximum amplitude. Since there are no series resistors, both positive and negative load currents are limited only by the size of the transistors and power supplies. By contrast, the single-transistor circuit of Figure 3.6 can deliver high positive

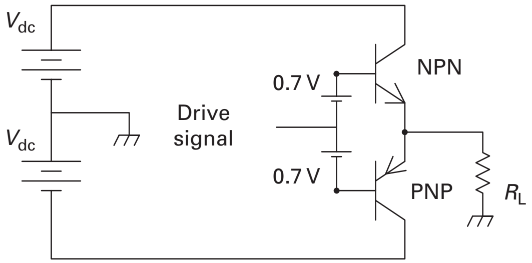

Figure 3.8. Complementary (PNP/NPN) push–pull amplifier.

current but the maximum negative current is limited by R_B.[2] Push–pull amplifiers are normally set up to run as *class-B* amplifiers, which means that, when the input voltage is zero, both transistors are just turned off and there is no power drawn from the supply(s). For low distortion, it is important that the crossover at $I = 0$ be continuous, so sometimes push–pull amplifiers are run class-AB which means that each transistor is given some bias current. Note that the amplifier of Figure 3.7 uses two NPN transistors, placed one above the other like faces on a totem pole. The top transistor acts as an emitter follower; when it is conducting, the output voltage will be almost equal to that transistor's base voltage. The bottom transistor, however, is driven in the common-emitter mode. The two transistors need drive signals of opposite polarities and present different drive impedances, requiring separate and different drive circuits. This unappealing asymmetry is eliminated in the complementary push–pull amplifier shown in Figure 3.8.

The complementary push–pull amplifier uses an NPN and a complementary (identical, except for polarity) PNP transistor. Both operate as emitter followers. Except for the 0.7 V offsets, their bases could be tied together. This can be taken care of, as shown in the figure, by using a pair of 0.7 V batteries. (In practice, this is done with some diode circuitry, rather than batteries.) There is no vacuum tube analog to this circuit, because there are no PNP tubes.[3] For completeness, we note that a third type of push–pull amplifier is obtained by interchanging the transistors in the complementary push–pull amplifier. This circuit, in which both transistors are driven in the common-emitter mode, is found in some switching (one transistor full-on while the other is full-off) servo amplifiers.

So-called bridge amplifiers are shown in Figure 3.9(a). They use a single power supply, but can supply the load with either polarity.

---

[2] As long as the load is a pure resistor, the pull-down resistor, R_B, is not a problem; any waveform not exceeding the power supply voltage limits can be faithfully amplified. But if the load contains an unavoidable capacitance C_L in parallel with R_L, the amplifier must be able to deliver current $V_{out}/R_L + C_L\, d/dt\, (V_{out})$. See Problem 3.6.

[3] The charge carriers in semiconductors are both electrons (negative) and "holes" (positive). Vacuum tubes use only electrons as charge carriers, although unwanted positive ions are sometimes produced by electron impact.

Figure 3.9. Bridge amplifiers: (a) full bridge (b) half bridge.

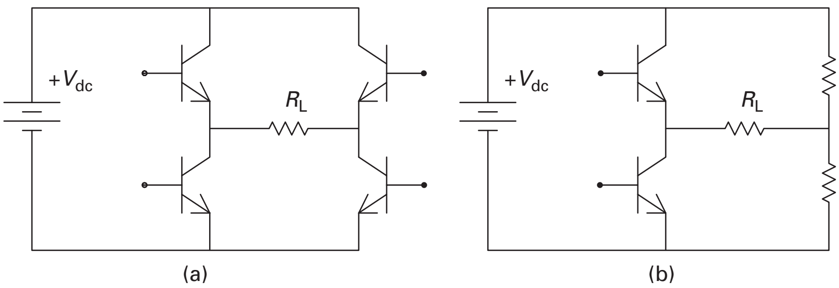

These circuits have no direct connection between the power supply and the load; either the supply or the load must “float,” i.e., have no ground connection. In the circuit of Figure 3.9(a), the top pair or bottom pair of transistors can operate as on–off switches (fully conducting or fully turned off). This circuit is a true push–pull amplifier, while the half bridge of Figure 3.9(b) is equivalent to a push-pull amplifier that has resistors in series with the power supples. This reduces the maximum voltage swing as well as the efficiency.

3.6 Efficiency calculations

As we are assuming that the drive power is small compared to the output power, we will calculate efficiency as the ratio of the average output power to the dc input power. Depending on the devices used, this ratio is known as the *collector efficiency*, *drain efficiency*, or *plate efficiency*. Most often, we want to compute the efficiency for the situation in which the amplifier is producing a sinusoidal output at full power (the condition for which the efficiency is usually a maximum).

When calculating efficiency, it is important to remember that average power is the time average of instantaneous power, i.e., the time average of voltage × current. Consider, for example, a power supply of voltage V_{dc} that is furnishing a current $I = I_0 + I_1\cos(\omega t)$. The average power is $\langle V_{dc}(I_0 + I_1\cos(\omega t))\rangle$, where the brackets $\langle\ldots\rangle$ indicate averaging. Since the average of a sum is equal to the sum of the averages, we can expand this expression.

$$\langle V_{dc}(I_0 + I_1\cos(\omega t))\rangle = \langle V_{dc}I_0\rangle + \langle V_{dc}I_1\cos(\omega t)\rangle = V_{dc}I_0, \tag{3.1}$$

since the average value of $\cos(\omega t)$ is zero. When a sine wave $V_0\sin(\omega t)$ is applied to a resistor, the instantaneous power is $VI = [V_0\sin(\omega t)]^2/R$ and the average power is $V_0^2\langle\sin^2(\omega t)\rangle/R = V_0^2/(2R)$.[4] It is also useful to remember that $\langle\sin(\theta)\cos(\theta)\rangle = 0$ and that the average value of $\sin(\theta)$ or $\cos(\theta)$ over one positive loop is equal to $2/\pi$.

---

[4] To see that $\langle\sin^2(\theta)\rangle = 1$, note that $\langle\sin^2(\theta)\rangle = \langle\cos^2(\theta)\rangle$, since the waveforms are identical, and use the identity $\sin^2(\theta) + \cos^2(\theta) = 1$.

We will first calculate the efficiency of the amplifier of Figure 3.6, under the conditions that $R_B = R_L$, $V_{pos} = V_{dc}$, and $V_{neg} = -2V_{dc}$. Let I_L denote the current downward into the load; I_B, the bias current leftward into R_B; and I_E, the current downward from the emitter. Note that $I_E = I_L + I_B$. Assume the maximum signal condition: $V_L = V_{dc} \cos(\omega t)$. This lets us write $I_B = (V_{dc} \cos(\omega t) + 2V_{dc})/R_B$. The power from the negative supply is therefore $P_{neg} = \langle I_B\ 2V_{dc} \rangle = 4V_{dc}^2/R_B$. The current into the load is just $I_L = V_{dc} \cos(\omega t)/R_L$. Adding I_L and I_B, we have $I_E = V_{dc} \cos(\omega t)/R_L + (V_{dc} \cos(\omega t) + 2V_{dc})/R_B$. Since this is the same as the current supplied by the positive supply (ignoring the transistor's small base current), we find that the power from the positive supply is given by $P_{pos} = 2V_{dc}^2/R_B$. The total dc power, P_{dc}, is the sum of P_{pos} and P_{neg}: $P_{dc} = 6V_{dc}^2/R_B$. The power into the load is $V_0^2/(2R_L)$, so the efficiency is given by

$$\eta = \frac{V_{dc}^2/(2R_L)}{6V_{dc}^2/R_B} \tag{3.2}$$

which reduces to $\eta = \frac{1}{12}$ when $R_B = R_L$.

Next we will find the efficiency of the push–pull amplifiers of Figures 3.7 and 3.8. Again, we assume the maximum signal condition, $V_L = V_{dc} \cos(\omega t)$. Consider the top transistor, which conducts during the positive half of the cycle. Assuming negligible base current, the current through this transistor will be the same as the current in the load, $I = (V_{dc} \cos(\omega t))/R_L$. During the positive half-cycle, the positive supply furnishes an average power given by $\langle V_{dc} \times (V_{dc}/R_L) \cos(\omega t) \rangle = (V_{dc}^2/R_L)\, 2/\pi$, since here the average is just over the positive loop. The negative supply, during its half-cycle, furnishes the same average power, since the circuit is symmetric. Thus, the efficiency is given by

$$\eta = \frac{V_{dc}^2/(2R_L)}{(V_{dc}^2/R_L)2/\pi} = \frac{\pi}{4} = 78\%. \tag{3.3}$$

3.7 AC amplifiers

All the amplifiers discussed above are known as *dc amplifiers* because they can handle signals of arbitrarily low frequency. (They might well be called universal amplifiers since they have no high-frequency limitations except those set by the transistors.) Audio and RF signals, however, are pure ac signals: their average value, i.e., their dc component, is zero. For these signals, special *ac amplifier* circuits provide simplicity and efficiency.

The circuit in Figure 3.10(a) is an ac version of the class-A amplifier of Figure 3.6. The drive signal at the base is given a positive offset (bias) which will create the same bias voltage at the emitter and a bias current through the pull-down resistor, R_E. The coupling capacitor (*dc blocking capacitor*) eliminates the dc bias from the load, and the output signal swings both positive and

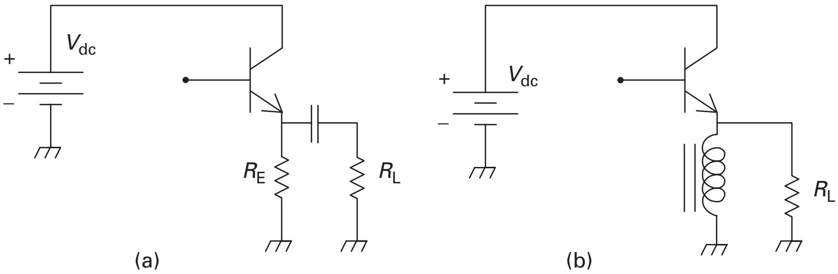

Figure 3.10. Common-collector single-ended ac amplifiers.

negative. The capacitance is chosen to be high enough that the full ac signal at the emitter will appear at the load. Only one power supply is required for this ac version. If $R_E = R_L$, the maximum peak-to-peak output swing is 2/3 V_{dc} and efficiency is again only $\frac{1}{12}$.

A major improvement is to replace the power-dissipating pull-down resistor with an inductor (ac *choke*) as shown in Figure 3.10(b). The inductor allows the output to go negative as well as positive[5] and makes possible a maximum output swing from $-V_{dc}$ to $+V_{dc}$. The inductance is chosen to be high enough to eliminate currents at the signal frequencies. No capacitor is needed; assuming the choke has negligible dc resistance, the average dc on the load will be zero. There must be sufficient bias current through the inductor to keep the transistor always on for the continuous control needed in linear operation. You can calculate (Problem 3.1) that the maximum efficiency of this circuit is 50%; the inductor improves the efficiency by a factor of 6 and the output swing by a factor of 3.

It might seem that the maximum efficiency of the class-B amplifier (78%) is only slightly better than the maximum efficiency of this class-A amplifier (50%). But these maximum efficiencies apply only when the amplifier is delivering a sine wave of maximum amplitude. For speech and music, the average power is much less than the maximum power. The class-B amplifier has little dissipation when the signal is low but a class-A amplifier, with its constant bias current, draws constant power equal to twice the maximum output power. A class-A audio amplifier rated for 25 watts output would consume a continuous 50 watts from its supply while a class-B amplifier of equal power rating would consume, on average, only a few watts, since the average power of audio signals is much lower than the peak power.

Common-emitter versions of this class-A amplifier are shown in Figure 3.11. The circuit of Figure 13.11(b) uses the shunt amplifier topology of Figure 3.5. Here, the inductor provides a wideband constant current source. (If the signal has a narrow bandwidth (RF), a parallel-resonant *LC* circuit will serve the same

---

[5] The bias current flowing downward in the inductor is essentially constant since the inductance is large. At the part(s) of the cycle when the current through the transistor becomes less than the inductor current, the inductor maintains its constant current by "sucking" current out of the load resistor and thus producing the negative output voltage.

Figure 3.11. Common-emitter single-ended ac amplifiers.

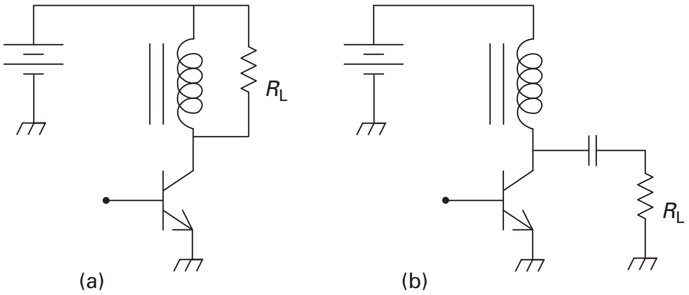

Figure 3.12. Transformer-coupled single-ended ac amplifier.

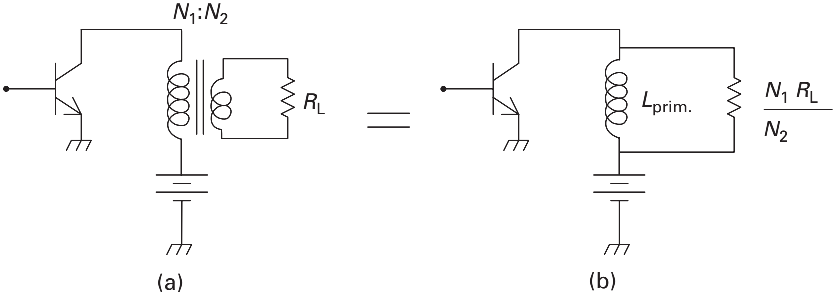

purpose.) A blocking capacitor allows one end of the load to be grounded, which is often a convenience. As we have seen, current drive is called for to achieve linearity if the emitter is tied directly to ground. At the expense of some efficiency, however, the emitter can be tied to ground through a resistor to allow the emitter voltage to follow the drive (base) voltage. Then the emitter current, collector current, and output voltage will all be linearly proportional to the input voltage. This technique of linearizing a common-emitter amplifier is known as "emitter degeneration" or "series feedback." To its credit, the common-emitter arrangement requires only a small and always positive drive signal, whereas the circuit of Figure 3.10(b) requires a drive voltage identical to the output signal, swinging both positive and negative.

Often the resistance of a given load is unsuitable for obtaining the desired power with the given power supply voltage. In this case, the choke and blocking capacitor can be replaced by a transformer as shown below in Figure 3.12.

This circuit is equivalent to that of Figure 13.11(a). The equivalence is shown in (b). Note how the load resistor is R_L, multiplied by the square of the transformer turns ratio. The transformer's primary winding provides the inductor. Again, these choke-coupled and transformer-coupled class-A amplifiers can provide a peak-to-peak collector swing of twice V_{dc}. Note that if we used only the simple "ideal transformer" model to replace the load resistor by its transformed value, we would have no inductance and would predict a maximum

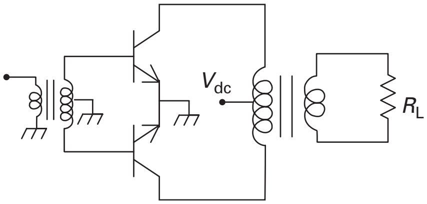

Figure 3.13. Symmetric transformer-coupled push–pull ac amplifier.

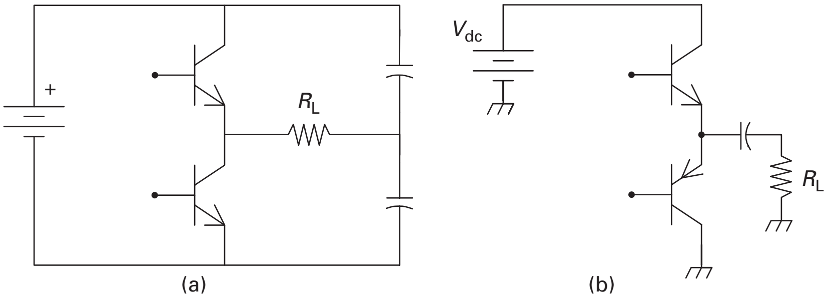

Figure 3.14. Half-bridge ac amplifier (a) and an equivalent version (b).

peak-to-peak swing of only V_{dc} (see Chapter 14). A center-tapped transformer can be used to make the symmetric push–pull amplifier shown in Figure 3.13.

This push–pull circuit, like the transformerless push–pull circuits, can be operated class B for high efficiency. Some high-power tube-type audio and RF amplifiers use this symmetric transformer circuit. Note the use of a center-tapped driver transformer – one way to supply the bases with the required opposite polarity signals.

Transistor audio amplifiers are usually push–pull and transformerless. They can be built with a single power supply by using capacitor coupling, as in the half-bridge circuit shown below in Figure 3.14(a). The circuit of Figure 3.14(b) is equivalent and uses only a single capacitor. The transistors can be in either the totem pole arrangement (a), or in the complementary NPN/PNP arrangement (b).

Since an audio or RF waveform has no dc component, the capacitors each charge to half the supply voltage and are equivalent to batteries. The capacitors thus form an artificial center tap for the power supply so this is an efficient push–pull amplifier, unlike the resistive half-bridge circuit of Figure 3.9(b).

3.8 RF amplifiers

RF amplifiers form a subset of ac amplifiers. RF signals are narrowband ac signals. Besides having no dc component, they have an almost constant waveform; while the amplitude and phase can vary, the shape remains sinusoidal.

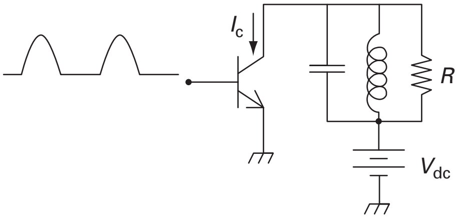

Figure 3.15. Single-ended class-B RF amplifier.

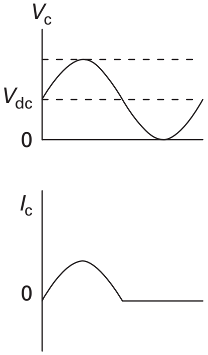

Figure 3.16. Collector voltage and current in the single-ended class-B RF amplifier.

This makes it possible to build a class-B RF amplifier with only a single transistor. The circuit (which looks no different from a class-A amplifier, except for the input waveform) is shown in Figure 3.15.

Here the transistor "pulls" current from the load, but there is no pusher transistor. Instead, a parallel resonant *LC* circuit provides a flywheel (energy storage) effect which maintains the sinusoidal waveform during the half-cycle in which the transistor is not conducting. The drive waveform consists only of positive loops. Between these loops, the transistor is nonconducting. This amplifier has the same 78% maximum efficiency of a push–pull class-B amplifier and also the class-B virtue of drawing no power when the signal level is zero. Let us analyze this circuit. Assume that, during the active half-cycle, the transistor current is $I_0 \sin(\theta)$, where $\theta = \omega t$ and I_0 is to be determined. Assume the resonant circuit provides enough energy storage (high enough Q) that the output voltage (the voltage across the load resistor) can be written as $A \sin(\theta)$. The output power is therefore $A^2/(2R)$. These voltage and current waveforms are shown in Figure 3.16.

During the active half-cycle, the power (*IV* product) into the *RCL* parallel circuit is $I_0 \sin(\theta) A \sin(\theta)$. Over a cycle, this averages to $I_0 A/4$, where one factor of $\frac{1}{2}$ is the time average of $\sin^2(\theta)$ and the other factor of $\frac{1}{2}$ comes from the transistor being turned off during half of every cycle. The power delivered to the parallel circuit must be equal to the power delivered to the load:

$$\frac{I_0 A}{4} = \frac{{V_{dc}}^2}{2R}. \tag{3.4}$$

From this we find $I_0 = 2A/R$. We can calculate the power delivered by the supply by noting that over the half-cycle, the instantaneous power is $V_{dc} I_0 \sin(\theta)$. The average over the half-cycle is $V_{dc}\, I_0\, 2/\pi$ and the average over the entire cycle is again half of this or $V_{dc}\, I_0/\pi = V_{dc}\, 2A/(\pi R)$. The efficiency, power out divided by power supplied by the supply, is therefore

$$\eta = \frac{A^2/(2R)}{V_{dc} 2A/(\pi R)} = \frac{A}{V_{dc}} \frac{\pi}{4}. \tag{3.5}$$

At the maximum output, where $A = V_{dc}$, the efficiency of this "single-ended" class-B amplifier is $\pi/4$, the same as the maximum efficiency for a push–pull class-B amplifier.

The circuit of Figure 3.15, if made to conduct throughout the complete cycle, would be a class-A amplifier. The *LC* would not be needed as a flywheel, but it does act as a bandpass filter. Moreover, the inductance, or part of it, serves to cancel out the unavoidable collector-to-emitter parasitic capacitance inherent in the transistor.

3.9 Matching a power amplifier to its load

In Chapter 2 we saw that, for maximum power transfer, a load should have the same impedance as the source that drives it. However, the power amplifiers discussed in this chapter all have essentially zero output impedances. Their Thévenin equivalent circuits are almost perfect voltage generators. Do we therefore try to make the load resistances as low as possible? No. The amplifiers are designed to deliver a specified power to a specified load. This determines the power supply voltages and current capacities and the required current-handling capacity of the transistor(s). Therefore, power amplifiers are deliberately mismatched to their loads. But a power amplifier does still have some very small output impedance. Won't it therefore supply the most power to a load of that impedance? The answer is yes, as long as the amplitude is kept very low. If the amplitude is turned up, such a load will simply "short out" the amplifier.

The output amplifier in a transmitter usually includes an impedance transforming network, often called an *antenna tuner*. The purpose of this network is *not* to make the antenna impedance equal to the amplifier's very low output impedance. Rather, the network transforms the antenna impedance to the impedance needed for the amplifier to produce its rated power.

Problems

Problem 3.1. Calculate efficiency of the class-A amplifier of Figure 3.10(b). Assume the output is a sine wave whose peak-to-peak amplitude is $2V_{\rm dc}$, symmetric about $V=0$.

Assume that the dc bias current in the inductor is just enough to allow the amplifier to produce the maximum signal.

Problem 3.2. The class-A amplifier shown below is operating at maximum power, applying a 24-volt peak-to-peak sine wave to the load resistor.

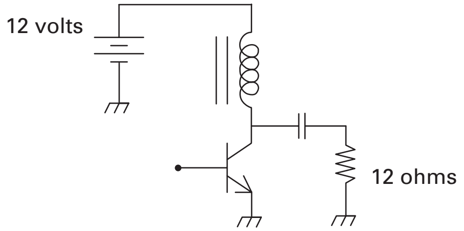

Assume the choke has zero dc resistance and enough inductance to block any ac current and that the capacitor has enough susceptance to prevent any ac voltage drop.

(a) Draw the waveform of the collector voltage. Hint: Remember that there can be no dc voltage drop across the choke.
(b) Draw the waveform of the collector current. Hint: Remember that there is no ac current through the choke.
(c) What power is drawn from the supply under the maximum signal sine wave condition?
(d) Show that the efficiency under this maximum signal sine wave condition is 50%.
(e) What power is drawn from the supply if the signal is zero?

Problem 3.3. An ideal push–pull amplifier does not have the current limitation of the emitter follower, so it can drive capacitive loads at high frequencies. But what about inductive loads? Suppose a load has an unavoidable series inductance but large voltages at high frequencies must be produced across the resistive part of the load. How does this impact the amplifier design?

Problem 3.4. Justify the statements made about the voltage gain (about 99%) and the offset (about 0.7 volts) of the emitter follower. Use the relation between the emitter current and base-to-emitter voltage of a (bipolar) transistor, $I \approx I_{\text{sat}} \exp([V_b - V_e]/0.026)$. To get a value for I_{sat}, assume that $I = 10$ ma when $V_b - V_e = 0.7$ volts. Remember that in the emitter follower, $V_e = I_e R$. Assume a reasonable value for R such as 1000 ohms and find V_e for several values of V_b.

Problem 3.5. Find the power gain of an emitter follower, i.e., the ratio of output signal power to input signal power. Use the fact that the input current (base current) is less than the emitter current by a factor $1/(\beta+1)$ where β is the transistor's current gain (typically on the order of 100). Remember that the output voltage is essentially the same (follows) the input voltage.

Problem 3.6. The emitter follower amplifier shown below has a load which includes an unavoidable parallel capacitance.

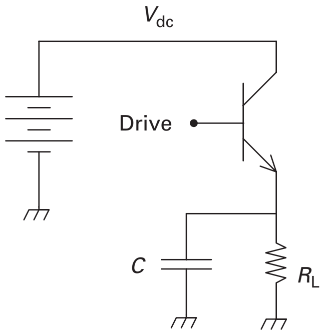

(a) What is the maximum peak-to-peak voltage that can be delivered to the load at low frequencies (where the capacitor can be neglected)?
(b) At what frequency will a sine-wave output signal of half the maximum amplitude become distorted? Hint: Express the emitter current as the sum of the resistor current

and capacitor current and note that distortion will occur if this current should ever be negative (the transistor can only supply positive current). Answer: $\omega = \sqrt{3(RC)}$.

Problem 3.7. Consider the push–pull amplifier of Figure 3.8 when it is being driven by a sine-wave signal, $V(t) = V_0\sin(\omega t)$, and is connected to a load that is an inductor, L, rather than a resistor. Draw a graph showing the current flowing into the inductor and the individual emitter currents flowing in the direction of the load. Based on your graph, would you agree with the statement: The top transistor applies positive voltage to the load and the bottom transistor applies negative voltage to the load?

Problem 3.8. The maximum efficiency (the efficiency when the signal is a maximum-amplitude sine wave) of the amplifier in Figure 3.6 is $\frac{1}{12}$ when $R_B = R_L$ and $V_{NEG} = -2V_{POS}$.

(a) Calculate the maximum efficiency when $R_B = \sqrt{2}R_L$ and $V_{NEG} - (1 + \sqrt{2})\ V_{POS}$.

(b) Show that this combination of R_B and V_{NEG} yields the greatest maximum efficiency.

Problem 3.9. Draw a circuit for a "double push–pull" amplifier with four transistors. Two of the transistors connect the load to supply voltages $V_{dc}/2$ and $-V_{dc}/2$. The other two transistors connect the load to a second pair of supplies with voltages V_{dc} and $-V_{dc}$. The transistors connecting the smaller power supplies are turned off (nonconducting) when $|V_{out}| > V_{dc}/2$ and the transistors connecting the larger power supplies are turned off when $|V_{out}| < V_{dc}/2$. Calculate the efficiency when the output is a sine wave swinging from $-V_{dc}$ to $+V_{dc}$. This circuit is sometimes called a class-K amplifier.

CHAPTER

4 Basic filters

Bandpass filters are key elements in radio circuits, for example, in radio receivers, to select the desired station. Here we will discuss lumped-element filters made of inductors and capacitors. We will first look at lowpass filters, and then see how they serve as prototypes for conversion to bandpass filters. We begin with the well-established lowpass filter prototypes – Butterworth, Chebyshev, Bessel, etc. These lowpass prototypes are simple *LC* ladder networks with series inductors and shunt capacitors, as shown in Figure 4.1.

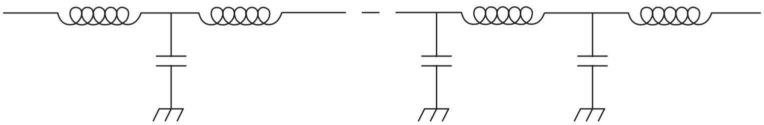

Figure 4.1. Lowpass ladder network.

An n-section lowpass filter has n components (capacitors plus inductors). The end components can be either series inductors, as shown above, or shunt capacitors, or one of each. Since they contain no (intentional) resistance, these filters are *reflective filters*; outside the passband, it is mismatch that keeps power from reaching the load. The ladder network can be redrawn as a cascade of voltage dividers as in Figure 4.2.

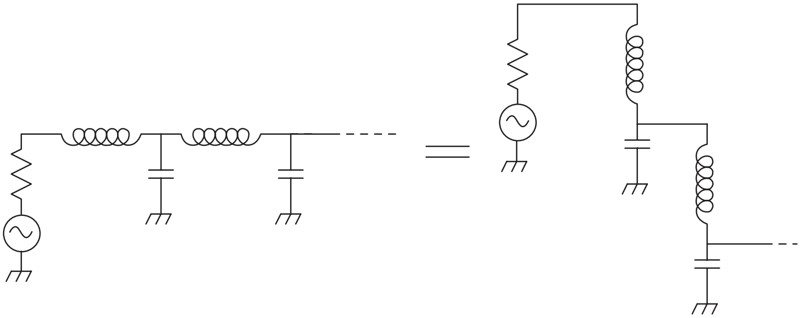

Figure 4.2. Ladder network as a cascade of voltage dividers.

At high frequencies the division ratio increases so the load is increasingly isolated from the source. For frequencies well above cutoff, each circuit element contributes 6 dB of attenuation per octave (20 dB per decade). Within the passband, an ideal lowpass filter provides a perfect match between the load and the source. Filters with many sections approach this ideal. When the source and load impedances have no reactance (either built-in or parasitic) it is theoretically possible to have a perfect match across a wide band.

4.1 Prototype lowpass filter designs

The Butterworth filter is maximally flat, that is, it is designed so that at zero frequency the first $2n-1$ derivatives with respect to frequency of the power transfer function are zero. The final condition (needed to determine the values of n elements) is the specification of the cutoff frequency, f_0, often specified as the 3-dB or half-power frequency. The frequency response of the Butterworth filter turns out to be

$$\left|\frac{V_\text{out}}{V_\text{in}}\right|^2 = \frac{1}{1+(f/f_0)^{2n}}. \tag{4.1}$$

While it is the flattest filter, the Butterworth filter does not have skirts as sharp as those of the Chebyshev filter. The trade-off is that the Chebyshev filters have some passband ripple. The design criterion for the Chebyshev filter is that these ripples all have equal depth. The response is given by

$$\left|\frac{V_\text{out}}{V_\text{in}}\right|^2 = \frac{1}{1+(V_\text{r}^{-2}-1)\cosh^2\left(n\cosh^{-1}(f/f_0)\right)}, \tag{4.2}$$

where V_r is the height above zero of the ripple valley (in voltage) relative to the height of the peaks.

You will find tables of filter element values in many handbooks and textbooks. Two tables from Matthaei, Young and Jones [2] are given in Appendix 4.1 at the end of this chapter. These tables are for normalized filters, i.e., the cutoff frequency[1] is 1 radian/sec ($1/2\pi$ Hz). The value of the n-th component is g_n farads or henrys, depending on whether the filter begins with a capacitor or with an inductor. The proper source impedance is $1+j0$ ohms. This is also the proper load impedance except for the even-order Chebyshev filters, where it is $1/g_{n+1}+j0$ ohms. Figure 4.3 shows plotted power responses of a Butterworth filter and several Chebyshev filters.

---

[1] The cutoff frequency for the Butterworth filters is the half-power (3 db) point. For an n-dB Chebyshev filter it is the highest frequency for which the response is down by n dB (see Figure 4.3).

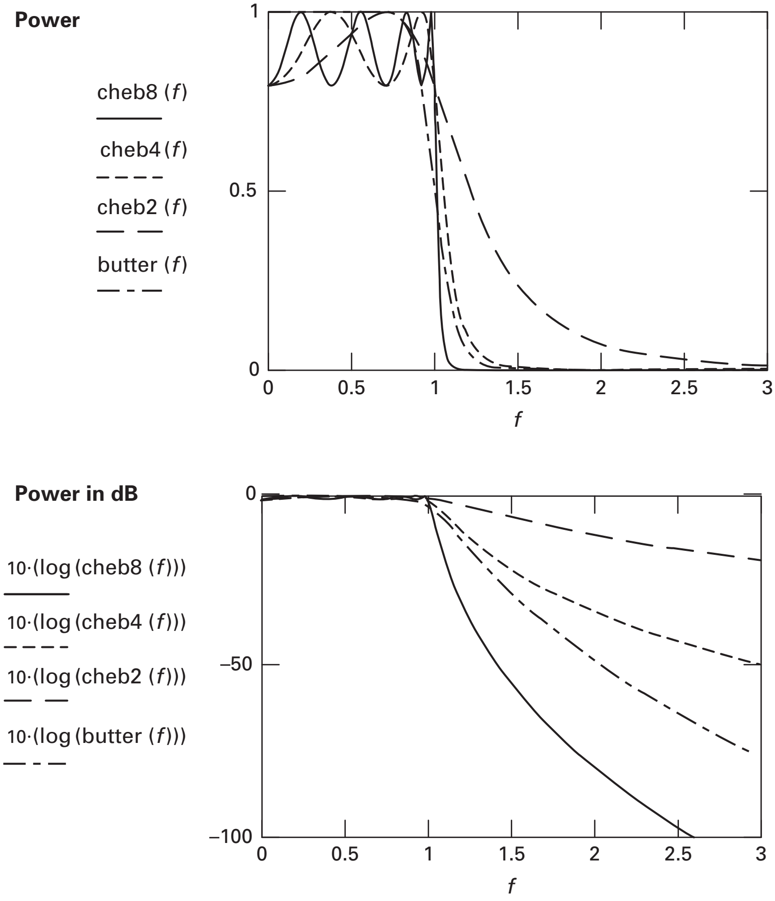

Figure 4.3. Butterworth and Chebyshev responses.

4.2 A lowpass filter example

As an example, we will look at the three-section Butterworth lowpass filter. From the table, the filter has values of 1 H, 2 F, and 1 H (Figure 4.4a) or 1F, 2H, and 1F (Figure 4.4b). The (identical) responses for these two filters are given in Table 4.1 and plotted in Figure 4.5. Note that they work as advertised; the 3-dB point is at 0.159 Hz.

Suppose we need a three-section Butterworth that is 5 kHz wide and works between a 50-ohm generator and a 50-ohm load. We can easily find the element values by scaling the prototype. The values of the inductors are just multiplied by 50 (we need 50 times the reactance) and divided by $2\pi \cdot 5000$ (we need to reach that reactance at 5 kHz, not 1 radian/sec). Similarly, the capacitor values are divided by 50 and divided by $2\pi \cdot 5000$. Figure 4.6 shows the circuit resulting from scaling the values of Figure 4.4b.

The response of the scaled filter is shown below in Table 4.2 and Figure 4.7.

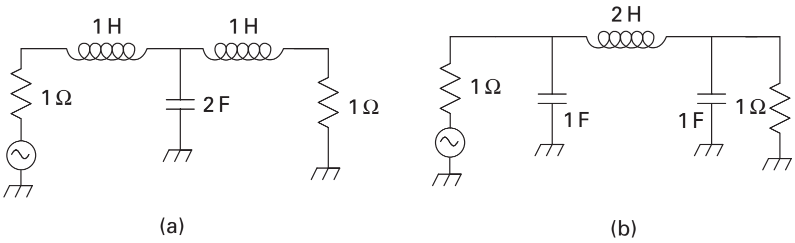

Figure 4.4. Equivalent three-section Butterworth lowpass filters.

Table 4.1 Frequency response for filters of Figure 4.4.

Frequency (Hz)	Power	Response (dB)
0.00	1.000	−0.0
0.0321	0.000	−0.0
0.0640	0.996	−0.02
0.095	0.955	−0.20
0.1270	0.792	−1.01
0.1590	0.500	−3.01
0.1910	0.251	−6.00
0.2230	0.117	−9.31
0.2540	0.056	−12.5
0.2860	0.029	−15.4
0.3180	0.015	−18.1

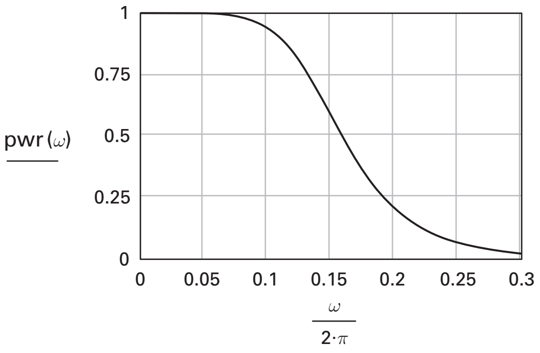

Figure 4.5. Plotted response of filters of Figure 4.4.

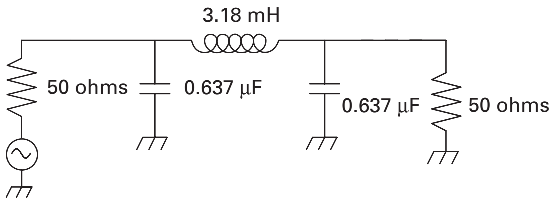

Figure 4.6. Filter of Figure 4.4(b), after conversion to 50 ohms and 5 kHz cutoff frequency.

Table 4.2 Response of the scaled lowpass filter of Figure 4.6.

Frequency (Hz)	Power	Response (dB)
0	1.000	−0.0
1000	1.000	−0.0
2000	0.996	−0.02
3000	0.956	−0.20
4000	0.793	−1.01
5000	0.500	−3.01
8000	0.056	−12.5
6000	0.251	−6.00
7000	0.117	−9.31
9000	0.029	−15.4
10000	0.015	−18.1

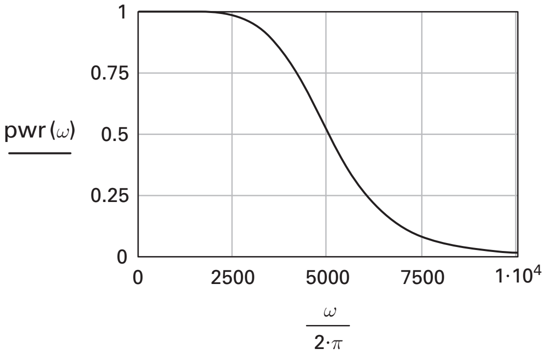

Figure 4.7. Plotted response of the scaled lowpass filter of Figure 4.6.

4.3 Lowpass-to-bandpass conversion

Here we will see how to convert lowpass filters into bandpass filters. Remember how the lowpass filters work: as frequency increases, the series arms (inductors), which are short circuits at dc, begin to pick up reactance. Likewise, the shunt arms (capacitors), which are open circuits at dc, begin to pick up susceptance. Both effects impede the signal transmission, as we have seen. To convert these lowpass filters in the most direct way to bandpass filters, we can replace the inductors by series *LC* combinations and the capacitors by parallel *LC* combinations. The series combinations are made to resonate (have zero impedance) at the center frequency of the desired bandpass filter, just as the inductors had zero impedance at dc, the "center frequency" of the prototype lowpass filter. It is important to note that as we move away from resonance, a series LC arm picks up reactance at twice the rate of the inductor alone. This is easy to see: The reactance of the series arm is given by

$$X_{\text{series}} = \omega L - \frac{1}{\omega C}. \tag{4.3}$$

Differentiating with respect to ω, we find

$$\frac{\mathrm{d}X}{\mathrm{d}\omega} = L + \frac{1}{\omega^2 C}. \tag{4.4}$$

At $\omega = \omega_0$,

$$\frac{\mathrm{d}X}{\mathrm{d}\omega} = L + \frac{1}{{\omega_0}^2 C} = 2L. \tag{4.5}$$

As we move off resonance, the inductor and the capacitor provide equal contributions to the reactance. Likewise, the parallel LC circuits, which replace the capacitors in the prototype lowpass filter, pick up susceptance at twice the rate of their capacitors. With this in mind, let us convert our 5-kHz lowpass filter into a bandpass filter. Suppose we want the center frequency to be 500 kHz and the bandwidth to be 10 kHz. As we move up from the center frequency, the series arms must pick up a reactance at the same rate the inductors picked up a reactance in the prototype lowpass filter. Similarly, the shunt arms must pick up susceptance at the same rate the capacitors picked up susceptance in the prototype. This will cause the bandpass filter to have the same shape above the center frequency as the prototype had above dc. If the 3-dB point of the prototype filter was 5 kHz, the upper 3-dB point of the bandpass filter will be at 5 kHz above the center frequency. The bandpass filter, however, will have a mirror-image response as we go below the center frequency. (Below center frequency the reactances and susceptances change sign but the response remains the same.)

Let us calculate the component values. As we leave center frequency, the series circuits will get equal amounts of reactance from the L and the C, as explained above. Therefore the series inductor values should be exactly half what they were in the low pass prototype. Note: no matter how high we make the center frequency, the values of the inductors are reduced only by a factor of 2 from the those of the scaled lowpass filter. The series capacitors are chosen to resonate at the center frequency with the new (half-value) series inductors. The values of the parallel arms are determined similarly; the parallel capacitors must have half the value they had in the prototype lowpass filter. Finally, the parallel inductors are chosen to resonate with the new (half-value) parallel capacitors. These simple conversions yield the bandpass filter shown in Figure 4.8.

The response of this bandpass filter is given below in Table 4.3 and Figure 4.9.

While this theoretical filter works perfectly (since its components are lossless), the component values are impractical; typical real components with these values would be too lossy to achieve the calculated filter shape. When a bandpass filter is

Figure 4.8. Bandpass filter.

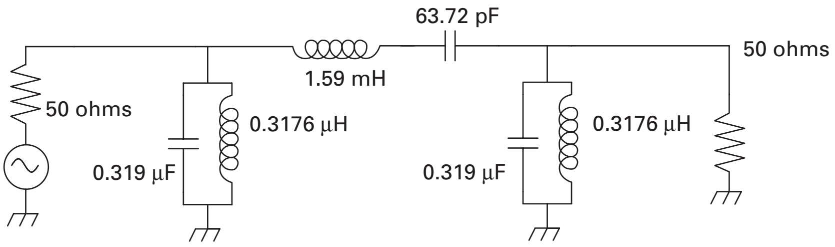

Table 4.3 Response of the bandpass filter of Figure 4.8.

Frequency (kHz)	Power	Response (dB)
490	0.014	− 18.1
492	0.053	− 12.8
494	0.241	− 6.19
496	0.785	− 1.05
498	0.996	− 0.18
500	1.000	− 0.00
502	0.996	− 1.16
504	0.801	− 0.966
506	0.260	− 5.84
508	0.059	− 12.9
510	0.016	− 17.9

Figure 4.9. Plotted response for Table 4.3.

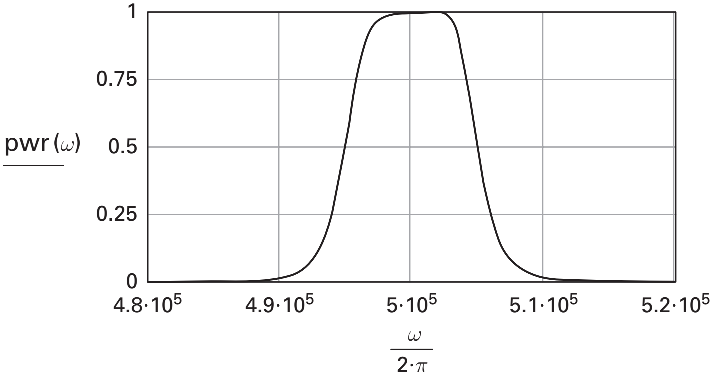

to have a large fractional bandwidth (bandwidth divided by center frequency) this direct conversion from lowpass to bandpass can be altogether satisfactory. It is when the fractional bandwidth is small, as in this example, that the direct conversion gets into trouble.[2] We will see later that the problem is solved by transforming the prototype lowpass filters into somewhat more complicated bandpass circuits known as *coupled resonator filters*. Those filters retain the desired shape (Butterworth, Chebyshev, etc.) and can serve, in turn, as prototypes for filters made from quartz or ceramic resonators and for filters made with resonant irises (thin aperture plates that partially block a waveguide).

---

[2] The component problem with the straightforward lowpass-to-bandpass conversion is that the values of the series inductors are very different from the values of the parallel inductors. (The same is true of the capacitors, but high-Q capacitors can usually be found.) In the above example, the inductors differ by a factor of about 5000 and it is normally impossible to find high-Q components over this range. (Low-Q inductors, of course, make the filter lossy and, if not accounted for, distort the bandpass

Appendix 4.1. Component values for normalized lowpass filters[3]

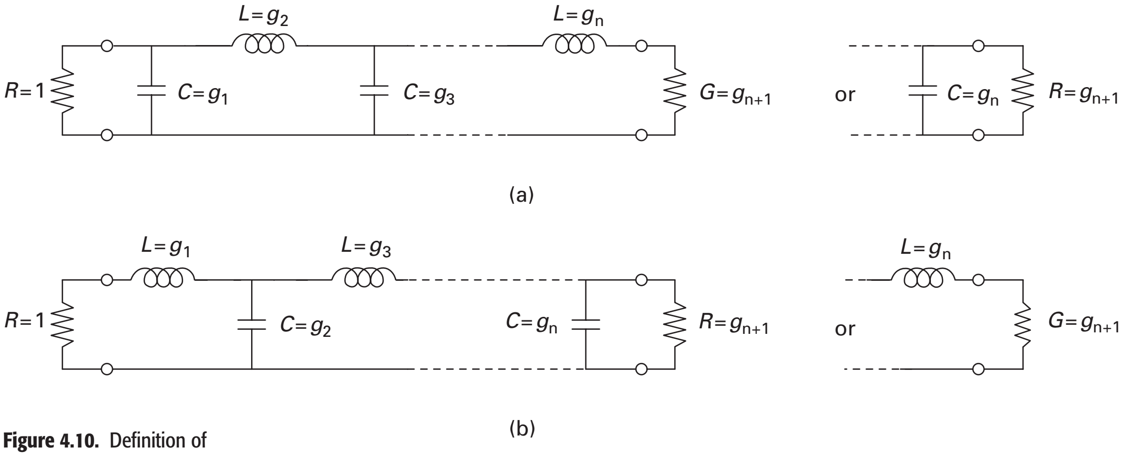

Figure 4.10. Definition of prototype filter parameters, g_1, g_2, …, g_n, g_{n+1}. The prototype circuit (a) and its dual (b) give the same response.

Table A4.1 Element values for Butterworth (maximally flat) lowpass filters (the 3-dB point is at $\omega = 1$ radian/sec).

Value of n	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}	g_{11}
1	2.000	1.000									
2	1.414	1.414	1.000								
3	1.000	2.000	1.000	1.000							
4	0.7654	1.848	1.848	0.7654	1.000						
5	0.6180	1.618	2.000	1.618	0.6180	1.000					
6	0.5176	1.414	1.932	1.932	1.414	0.5176	1.000				
7	0.4450	1.247	1.802	2.000	1.802	1.247	0.4450	1.000			
8	0.3902	1.111	1.663	1.962	1.962	1.663	1.111	0.3902	1.000		
9	0.3473	1.000	1.532	1.879	2.000	1.879	1.532	1.000	0.3473	1.000	
10	0.3129	0.9080	1.414	1.782	1.975	1.975	1.782	1.414	0.9080	0.3129	1.000

---

shape.) The inductors in coupled-resonator filters are all of about the same value. If a high-Q inductor can be found, the coupled resonator filter is designed for whatever impedance calls for that value of inductor and then transformers or matching sections are used at each end to convert to the desired impedance.

[3] From Matthaei, Young, and Jones [2].

Table A4.2 Element values for Chebyshev lowpass filters (for a filter with *N*-dB ripple, the last *N*-dB point is at $\omega = 1$ radian/sec).

Value of n	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}	g_{11}
0.01-dB ripple											
1	0.0960	1.0000									
2	0.4488	0.4077	1.1007								
3	0.6291	0.9702	0.6291	1.0000							
4	0.7128	1.2003	1.3212	0.6476	1.1007						
5	0.7563	1.3049	1.5773	1.3049	0.7563	1.0000					
6	0.7813	1.3600	1.6896	1.5350	1.4970	0.7098	1.1007				
7	0.7969	1.3924	1.7481	1.6331	1.7481	1.3924	0.7969	1.0000			
8	0.8072	1.4130	1.7824	1.6833	1.8529	1.6193	1.5554	0.7333	1.1007		
9	0.8144	1.4270	1.8043	1.7125	1.9057	1.7125	1.8043	1.4270	0.8144	1.0000	
10	0.8196	1.4369	1.8192	1.7311	1.9362	1.7590	1.9055	1.6527	1.5817	0.7446	1.1007
0.1-dB ripple											
1	0.3052	1.0000									
2	0.8430	0.6220	1.3554								
3	1.0315	1.1474	1.0315	1.0000							
4	1.1088	1.3061	1.7703	0.8180	1.3554						
5	1.1468	1.3712	1.9750	1.3712	1.1468	1.0000					
6	1.1681	1.4039	2.0562	1.5170	1.9029	0.8618	1.3554				
7	1.1811	1.4228	2.0966	1.5733	2.0966	1.4228	1.1811	1.0000			
8	1.1897	1.4346	2.1199	1.6010	2.1699	1.5640	1.9444	0.8778	1.3554		
9	1.1956	1.4425	2.1345	1.6167	2.2053	1.6167	2.1345	1.4425	1.1956	1.0000	
10	1.1999	1.4481	2.1444	1.6265	2.2253	1.6418	2.2046	1.5821	1.9628	0.8853	1.3554
0.2-dB ripple											
1	0.4342	1.0000									
2	1.0378	0.6745	1.5386								
3	1.2275	1.1525	1.2275	1.0000							
4	1.3028	1.2844	1.9761	0.8468	1.5386						
5	1.3394	1.3370	2.1660	1.3370	1.3394	1.0000					
6	1.3598	1.3632	2.2394	1.4555	2.0974	0.8838	1.5386				
7	1.3722	1.3781	2.2756	1.5001	2.2756	1.3781	1.3722	1.0000			
8	1.3804	1.3875	2.2963	1.5217	2.3413	1.4925	2.1349	0.8972	1.5386		
9	1.3860	1.3938	2.3093	1.5340	2.3728	1.5340	2.3093	1.3938	1.3860	1.0000	
10	1.3901	1.3983	2.3181	1.5417	2.3904	1.5536	2.3720	1.5066	2.1514	0.9034	1.5386
0.5-dB ripple											
1	0.6986	1.0000									
2	1.4029	0.7071	1.9841								
3	1.5963	1.0967	1.5963	1.0000							
4	1.6703	1.1926	2.3661	0.8419	1.9841						
5	1.7058	1.2296	2.5408	1.2296	1.7058	1.0000					
6	1.7254	1.2479	2.6064	1.3137	2.4758	0.8696	1.9841				
7	1.7372	1.2583	2.6381	1.3444	2.6381	1.2583	1.7372	1.0000			
8	1.7451	1.2647	2.6564	1.3590	2.6964	1.3389	2.5093	0.8796	1.9841		
9	1.7504	1.2690	2.6678	1.3673	2.7239	1.3673	2.6678	1.2690	1.7504	1.0000	
10	1.7543	1.2721	2.6754	1.3725	2.7392	1.3806	2.7231	1.3485	2.5239	0.8842	1.9841
1.0-dB ripple											
1	1.0177	1.0000									
2	1.8219	0.6850	2.6599								
3	2.0236	0.9941	2.0236	1.0000							

Table A4.2 (cont.)

Value of n	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	g_{10}	g_{11}
4	2.0991	1.0644	2.8311	0.7892	2.6599						
5	2.1349	1.0911	3.0009	1.0911	2.1349	1.0000					
6	2.1546	1.1041	3.0634	1.1518	2.9367	0.8101	2.6599				
7	2.1664	1.1116	3.0934	1.1736	3.0934	1.1116	2.1664	1.0000			
8	2.1744	1.1161	3.1107	1.1839	3.1488	1.1696	2.9685	0.8175	2.6599		
9	2.1797	1.1192	3.1215	1.1897	3.1747	1.1897	3.1215	1.1192	2.1797	1.0000	
10	2.1836	1.1213	3.1286	1.1933	3.1890	1.1990	3.1738	1.1763	2.9824	0.8210	2.6599
2.0-dB ripple											
1	1.5296	1.0000									
2	2.4881	0.6075	4.0957								
3	2.7107	0.8327	2.7107	1.0000							
4	2.7925	0.8806	3.6063	0.6819	4.0957						
5	2.8310	0.8985	3.7827	0.8985	2.8310	1.0000					
6	2.8521	0.9071	3.8467	0.9393	3.7151	0.6964	4.0957				
7	2.8655	0.9119	3.8780	0.9535	3.8780	0.9119	2.8655	1.0000			
8	2.8733	0.9151	3.8948	0.9605	3.9335	0.9510	3.7477	0.7016	4.0957		
9	2.8790	0.9171	3.9056	0.9643	3.9598	0.9643	3.9056	0.9171	2.8790	1.0000	
10	2.8831	0.9186	3.9128	0.9667	3.9743	0.9704	3.9589	0.9554	3.7619	0.7040	4.0957
3.0-dB ripple											
1	1.9953	1.0000									
2	3.1013	0.5339	5.8095								
3	3.3487	0.7117	3.3487	1.0000							
4	3.4389	0.7483	4.3471	0.5920	5.8095						
5	3.4817	0.7618	4.5381	0.7618	3.4817	1.0000					
6	3.5045	0.7685	4.6061	0.7929	4.4641	0.6033	5.8095				
7	3.5182	0.7723	4.6386	0.8039	4.6386	0.7723	3.5182	1.0000			
8	3.5277	0.7745	4.6575	0.8089	4.6990	0.8018	4.4990	0.6073	5.8095		
9	3.5340	0.7760	4.6692	0.8118	4.7272	0.8118	4.6692	0.7760	3.5340	1.0000	
10	3.5384	0.7771	4.6768	0.8136	4.7425	0.8164	4.7260	0.8051	4.5142	0.6091	5.8095

Problems

Problem 4.1. Design a five-element lowpass filter with a Chebyshev 0.5-dB ripple shape. Let the input and output impedances be 100 ohms. Use parallel capacitors at the ends. The bandwidth (from dc to the last 0.5-dB point) is to be 100 kHz. Use Table A4.2 to find the values of the prototype 1 ohm, 1 rad/sec filter and then alter these values for 100 ohms and 100 kHz.

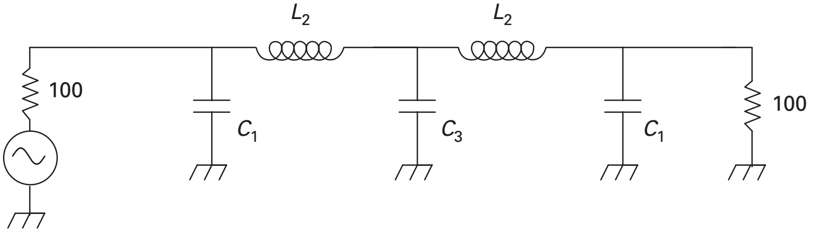

Problem 4.2. Use the results of Problem 4.1 to design a five-element bandpass filter with a Chebyshev 0.5-dB ripple shape. Let the input and output impedances remain at 100 ohms. The center frequency is to be 5 MHz and the total bandwidth (between outside 0.5-dB points) is to be 200 kHz.

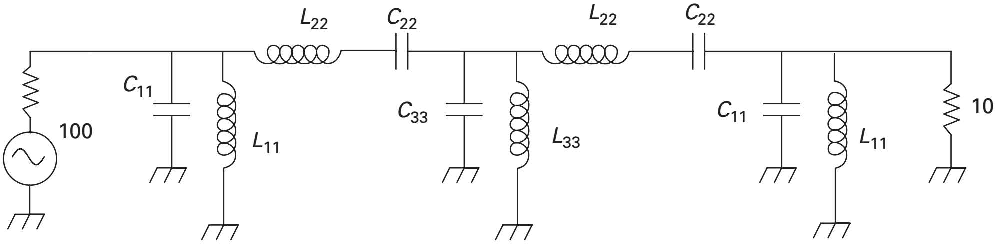

Problem 4.3. Convert the filter of Problem 4.2 to operate at 50 ohms by adding an L-section matching network at each end. Test the filter design using your ladder network analysis program, sweeping from 4.5 to 5.5 MHz in steps of 20 KHz.

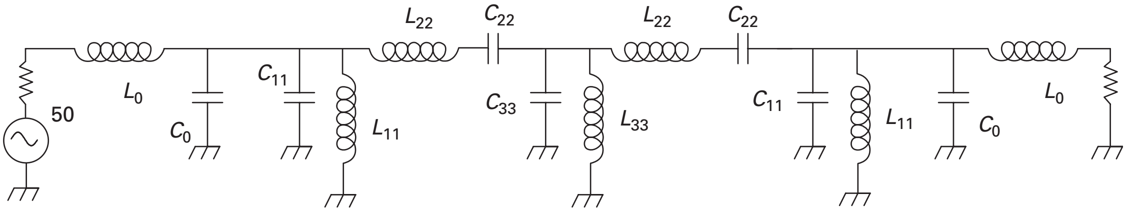

Problem 4.4. The one-section bandpass filter shown below uses a single parallel resonator. In its prototype lowpass filter, the resonator is a single shunt capacitor. Show that the frequency response of this filter is given by

$$\frac{P}{P_{\max}} = \frac{1}{1 + Q^2(f/f_0 - f_0/f)^2}$$

where f_0 is the resonant frequency of the LC combination and Q is defined as $R/(\omega_0 L)$, where R is the parallel combination of R_S and R_L.

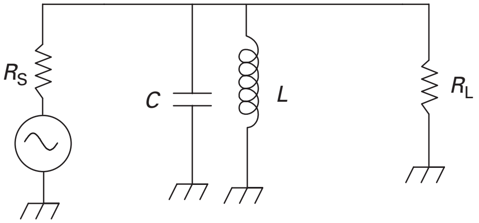

Problem 4.5. Highpass filters are derived from lowpass filters by changing inductors to capacitors and vice versa and replacing the component values in the prototype lowpass

tables by their reciprocals. (A 2-F capacitor, for example, would become a 0.5-H inductor.) The prototype highpass response at ω will be equal to the prototype lowpass response at 1/ω. Convert the lowpass filter of Figure 4.4(b) into a highpass filter. (Answer: 1 H, 0.5 F, 1 H.) Next, scale it to have a cutoff frequency of 5 kHz and to operate at 50 ohms. Finally, convert the scaled filter into a bandstop filter with a stopband 10 kHz wide, centered at 500 kHz.

Problem 4.6. Enhance your ladder network analysis program (Problem 1.3) to display not just the amplitude response of a network, but also the phase response (phase angle of the output voltage minus phase angle of the input voltage). Calculate the phase response of the Butterworth filter in Figure 4.4(a). Note: ladder networks belong to a class of networks ("minimum phase networks") for which the amplitude response uniquely determines the phase response and vice versa. In Chapter 12 we will encounter "allpass" filters which are not in this class; phase varies with frequency while amplitude remains constant.

Example answer: For the MATLAB program listing in Problem 1.3, simply insert the following two lines ahead of the last two lines in the original program.

```
figure(3);plot(-180/pi*angle(Vgen));
grid; xlabel('Frequency');ylabel('degrees');title('Phase response');
```

References

[1] Fink, D. G., *Electronic Engineers' Handbook*, New York: McGraw-Hill, 1975. See Section 12, Filters, Coupling Networks, and Attenuators by M. Dishal. Contains an extensive list references.

[2] Matthaei, G., Young, L. and Jones, E. M. T., *Microwave Filters, Impedance-Matching Networks, and Coupling Structures* New York: McGraw Hill, 1964, reprinted in 1980 by Artech House, Inc. Contains fully developed designs, comparing measured results with theory (spectacular fits, even at microwave frequencies) and has an excellent introduction and review of the theory.

CHAPTER

5

Frequency converters

A common operation in RF electronics is frequency translation, whereby all the signals in a given frequency band are shifted to a higher frequency band or to a lower frequency band. Every spectral component is shifted by the same amount. Cable television boxes, for example, shift the selected cable channel to a low VHF channel (normally channel 3 or 4). Nearly every receiver (radio, television, radar, cell phone, …) uses the *superheterodyne* principle, in which the desired channel is first shifted to an intermediate frequency or "IF" band. Most of the amplification and bandpass filtering is then done in the fixed IF band, with the advantage that nothing in this major portion of the receiver needs to be retuned when a different station or channel is selected. The same principle can be used in frequency-agile transmitters; it is often easier to shift an already modulated signal than to generate it from scratch at an arbitrary frequency. Frequency translation is also called *conversion* and is even more commonly called *mixing*.[1]

5.1 Voltage multiplier as a mixer

A mixer takes the input signal or band of signals (segment of spectrum), which is to be shifted, and combines it with a reference signal whose frequency is equal to the desired shift in frequency. In a radio receiver, the reference or "L.O." signal is a sine-wave, generated within the receiver by a *local oscillator*.[2] Mixers, in order to produce new frequencies, must necessarily be nonlinear since linear circuits can change only the amplitudes and phases of a set of superposed sine waves. Multiplication is the nonlinear operation used in mixers

---

[1] In audio work "mixing" means addition, a linear superposition that produces no new frequencies. In RF work, however, mixing means multiplication; an RF mixer either directly or indirectly forms the product of the input signal voltage and a sinusoidal "local oscillator" (L.O.) voltage. Multiplication produces new frequencies.

[2] The earliest radio receivers employed no frequency conversion, so they had no "local" oscillator; the only oscillator was remote – at the transmitter location.

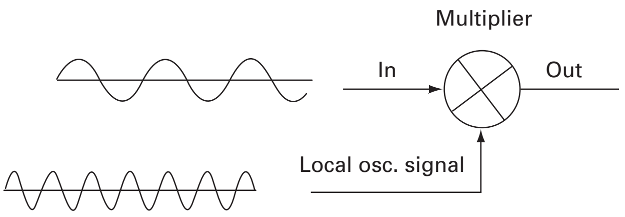

Figure 5.1. A voltage multiplier used as a frequency converter (mixer).

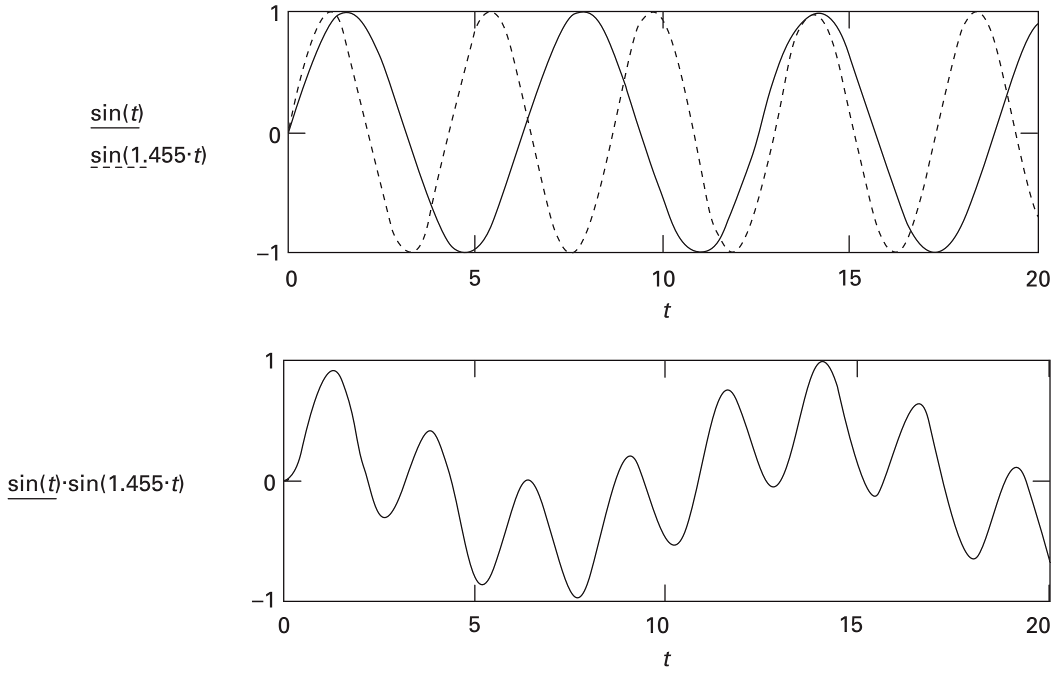

Figure 5.2. Multiplier input and output waveforms.

to produce new signals at the shifted frequencies. Figure 5.1 shows a voltage multiplier with its signal and L.O. inputs. A circumscribed *X* is the standard symbol for a mixer. The input signal port of the mixer is usually labeled "RF," while the other two ports are labeled "LO" and "OUT" (or "IF").

The output voltage from the multiplier is the product (or proportional to the product) of the two input voltages. In Figure 5.2 a sine-wave input signal is multiplied by an L.O. that is 1.455 times higher in frequency. These multiplicands are shown in the top graph. The bottom graph shows their product which can be seen to contain frequencies both higher and lower than the original frequencies.

The familiar "sin(*a*)sin(*b*)" trigonometric identity shows that, in this simple case, the multiplier output consists of just two frequencies: an up-shifted signal at $\omega_L + \omega_R$ and a down-shifted signal at $\omega_L - \omega_R$:

$$\sin(\omega_R t)\sin(\omega_L t) = {}^1\!/_2[\cos(\omega_R - \omega_L)t - \cos(\omega_R + \omega_L)t]. \qquad (5.1)$$

If we replace the single RF signal by $V_1\sin(\omega_{R1}t)+V_2\sin(\omega_{R2}t)$, a signal with two spectral components, you can confirm that the output will be

$$\begin{aligned}&{}^1\!/_2V_1\cos([\omega_{R1}-\omega_L]t)+{}^1\!/_2V_2\cos([\omega_{R2}-\omega_L]t)\\&-{}^1\!/_2V_1\cos([\omega_{R1}+\omega_L]t)-{}^1\!/_2V_2\cos([\omega_{R2}+\omega_L]t).\end{aligned}\tag{5.2}$$

Just as this linear combination of two signals is faithfully copied into both an up-shifted band and a down-shifted band, any linear combination, i.e., any spectral distribution of signals, will be faithfully copied into these shifted output bands. With respect to signals applied to the RF port, you can see that the mixer is a linear device; all the components are translated (both up and down) in frequency, but their relative amplitudes are left unchanged and there is no interaction between them. Usually one wants only the up-shifted band or only the down-shifted band; the other is eliminated with an appropriate bandpass filter. Ideal analog (or digital) multipliers are being used more commonly as mixers in RF electronics as their speeds increase with improving technology.

5.2 Switching mixers

If the L.O. waveform is square, rather than a sinusoidal, the mixer output will contain not only the fundamental up-shifted and down-shifted outputs but also components at offsets corresponding to the third, fifth, and all other odd harmonics of the L.O. frequency, i.e., at offsets corresponding to all the frequencies in the Fourier decomposition of the square wave. You can confirm this by simply multiplying the Fourier series for the square-wave L.O. by the superposition of signals in the input at the RF port. These new components are usually very easy to filter out so there is no disadvantage in using a square-wave L.O. In fact, there is an advantage. Consider an L.O. signal that is a square wave with values ±1. In this case, since the multiplier multiplies the input signal only by either +1 or − 1, it can be replaced by an electronic SPDT switch that connects the output alternately to the input signal and the negative of the input signal. This equivalence is shown in Figure 5.3.

The phase inversion needed for the bottom side of the switch can easily be done with a center-tapped transformer and the switching can be done with two

Figure 5.3. Switching mixer operation.

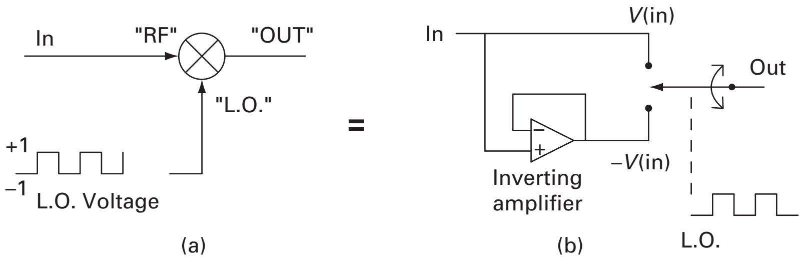

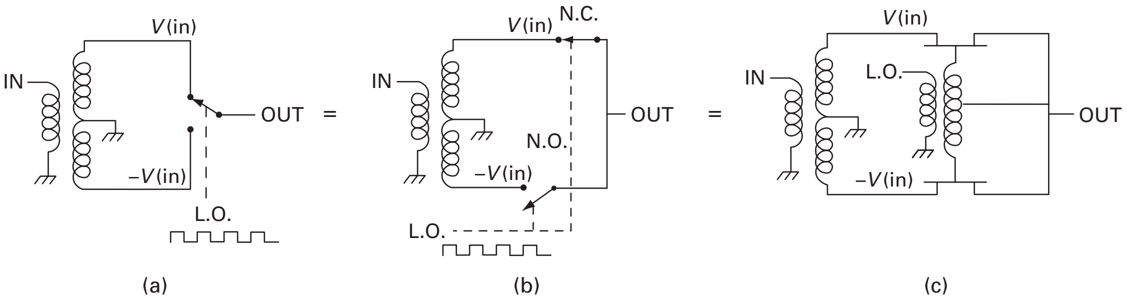

Figure 5.4. Active switching mixer using transistors.

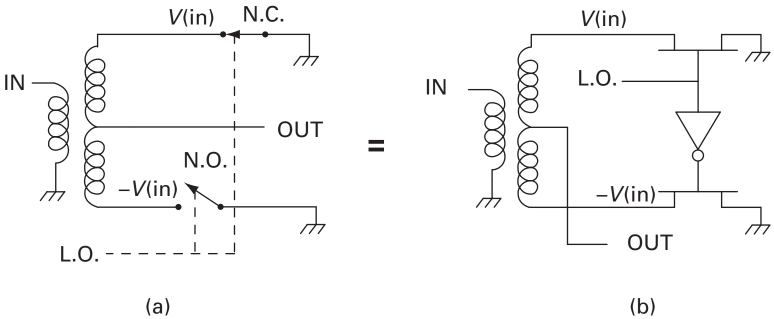

Figure 5.5. Alternate active switching mixer.

transistors, one for the high side and one for the low side. In the circuit of Figure 5.4 the switches are FETs. (Mixers based on transistors are called *active mixers*.) A second center-tapped transformer provides the L.O. phase inversion so that one FET is turned on while the other is turned off.

We could just as well have taken the signal from the center tap and used the FETs to ground one end of the secondary and then the other. With this arrangement, shown in Figure 5.5, it is easier to provide the drive signals to the transistors, since they are not floating.

Diodes are commonly used as the switching elements for the arrangement shown in Figure 5.5. This results in the *passive* switching mixer circuit shown in Figure 5.6.

Voltage from the L.O. transformer alternately drives the top diode pair and the bottom pair into conduction. The L.O. signal is made large enough that the conducting diodes have very low impedance (small depletion region) and the nonconducting diodes have a very large impedance (wide depletion region). The end of the input transformer connected between the turned-on diode pair is effectively connected to ground through the secondary of the L.O. transformer. Note that current uses both sides of the L.O. transformer on the way to ground, so no net flux is created in that transformer and it has zero impedance for this current. This circuit is usually drawn in the form shown in Figure 5.6(b) and is referred to as a *diode ring mixer*.

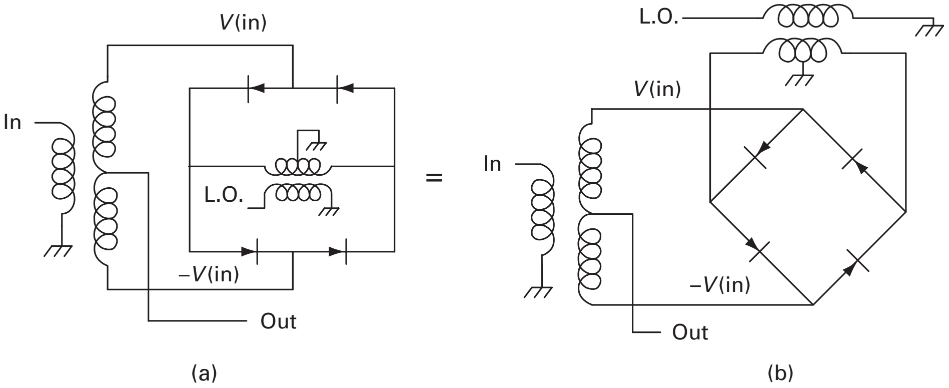

Figure 5.6. Diode ring mixer.

Figure 5.7. Unbalanced switching mixers.

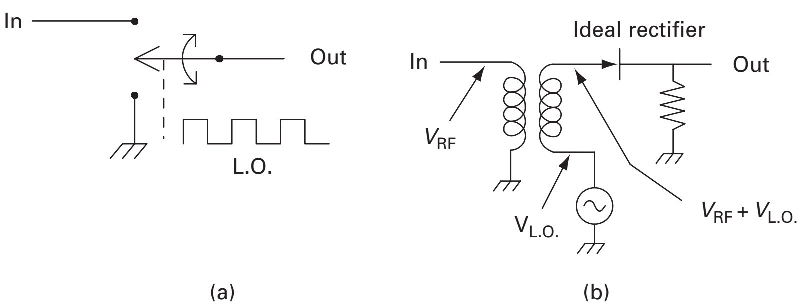

All the switching mixers shown above are "double-balanced" which means that no L.O. frequency energy appears at the RF or IF ports and no RF, except for the mixing products, appears at the OUT port. A balanced mixer is desirable, for example, when it is the first element in a receiver. An unbalanced mixer would allow L.O. energy to feed back into the antenna and the radiation could cause interference to other receivers (and could also reveal the position of the receiver). An unbalanced switching mixer is shown in Figure 5.7(a). It multiplies the signal by a square wave that goes from +1 to 0 (rather than +1 to − 1) which is just a $+\frac{1}{2}$ to $-\frac{1}{2}$ square wave together with a bias of $\frac{1}{2}$. The square-wave term produces the up-shifted and down-shifted bands as before, but the bias term allows one-quarter of the RF input power to get through, unshifted in frequency, to the output.

A simple version of this mixer, using an ideal rectifier, is shown in Figure 5.7 (b). The L.O. and RF voltages are added here, by means of a transformer, and their sum is rectified. The voltage at the output is equal to the sum voltage when the sum is positive and is equal to zero when the sum is negative. If the L.O. voltage is large compared to the RF voltage, the rectifier effectively conducts when the L.O. voltage is positive and disconnects when the L.O. voltage is negative, allowing the resistor to pull down the output voltage to zero. Thus, the RF signal is switched to the output at the L.O. rate. Note, however, that this mixer is totally unbalanced; both the L.O. and RF signals appear at the mixer, together with the sum and difference frequencies.

Switching mixer loss

Let us consider the loss in the switching mixers in Figures 5.4, 5.5, and 5.6. Refer to Figure 5.4(a) and assume that the OUT port is terminated by a load resistor R_L whose value equals R_S, the source impedance of a sine-wave signal source connected to the IN port. Note that the source has no way of knowing that the load resistor is being reversed on half of every cycle of the L.O. The source just sees a matched load and therefore delivers its maximum power. Some of the power on the load is at the desired sum or difference frequency. The ratio of this desired power to the power available from the source is known as the conversion gain. For the diode ring mixer, the ratio is less than unity, i.e., a conversion loss. We can easily calculate this loss. The reversing switch presents the load with a voltage that is half the source voltage (since R_S and R_L form a voltage divider), multiplied by a dimensionless ±1 square wave. Fourier analysis shows that the square wave is made of a sine wave at the square-wave frequency plus a sine wave at every odd multiple of this frequency. The amplitude of the fundamental sine wave is $4/\pi$. The amplitudes of the higher harmonics fall off as $1/n$. Evaluating the product of this square wave and the source voltage we have

$$\begin{aligned} V_{\mathrm{OUT}} &= \tfrac{1}{2}V_S\cos(\omega_R t)\cdot[4/\pi)(\cos(\omega_L t)+3^{-1}\cos(3\omega_L t)+5^{-1}\cos(5\omega_L t)+\ldots] \\ &= \tfrac{1}{2}V_S(4/\pi)\cos(\omega_R t)\cos(\omega_L t)+\ldots \\ &= \tfrac{1}{2}V_S(4/\pi)[(1/2)\cos(\omega_R-\omega_L)t+(1/2)\cos(\omega_R+\omega_L)t]+\ldots \end{aligned} \tag{5.3}$$

We see that the amplitude of the desired sum or difference frequency component is $\tfrac{1}{2}V_S(4/\pi)\cdot(1/2)$. The amplitude available from the source is $\tfrac{1}{2}V_S$, so the conversion gain is the ratio of the squares of these amplitudes:

$$\text{Conversion gain} = \left(\frac{1}{2}\frac{4}{\pi}\frac{1}{2}\right)^2 \Bigg/ \left(\left(\frac{1}{2}\frac{4}{\pi}\frac{1}{2}\right)^2\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\right)^2 = \left(\frac{4}{\pi}\frac{1}{2}\right)^2 = 0.4052 \tag{5.4}$$

or, in dB, 10 log (0.4052) = −3.92 dB. In practice, the loss is typically greater than this by a dB or so, due to loss in the diodes and in the transformers.

5.3 A simple nonlinear device as a mixer

Finally, let us consider a mixer that uses a single nonlinear device, but not as a switch. Figure 5.8 shows a single-diode mixer. The first op-amp is used to sum the RF and L.O. voltages. The sum is applied to the diode. The input of the second op-amp is a virtual ground so the full sum voltage is applied across the

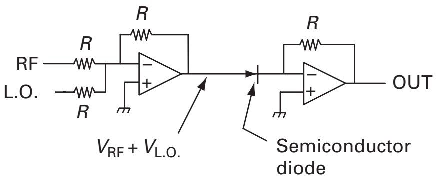

Figure 5.8. Hypothetical single-diode mixer circuit.

diode. With its feedback resistor, the second op-amp acts as a current-to-voltage converter; it produces a voltage proportional to the current in the diode. The current, a nonlinear (exponential) function of the applied voltage, will contain mixing products at frequencies $N\omega_{\mathrm{RF}} \pm M\omega_{\mathrm{L.O.}}$ where N and M are simple integers. Note that this circuit is essentially the same as that of Figure 5.7(b), except that here we are considering low-level signals, where the diode cannot be treated as an ideal rectifier.

This op-amp circuit is intended to emphasize that the diode's nonlinearity operates on the *sum* of the RF and L.O. Commonly used circuits use passive components and the summing is not always obvious. Diodes are exponential devices; the current vs. applied voltage is given by

$$I = I_{\mathrm{s}}(\exp(V/V_{\mathrm{th}}) - 1), \tag{5.5}$$

where $V_{\mathrm{th}} = V_{\mathrm{thermal}} = kT/e$ (Boltzmann's constant × absolute temperature / electron charge) $= 26\,\mathrm{mV}$. The term I_{s} is a temperature-dependent "saturation current." In a small-signal situation, i.e., when $V \ll 26\,\mathrm{mV}$, we can expand the exponential to find the output of the above mixer:

$$V_{\mathrm{out}} = I_{\mathrm{s}}R[V/V_{\mathrm{th}} + (V/V_{\mathrm{th}})^2/2! + (V/V_{\mathrm{th}})^3/3! + \ldots \tag{5.6}$$

Since $V = V_{\mathrm{RF}} + V_{\mathrm{L.O.}}$, the first term will give feedthrough (no balance) at both the RF and L.O. frequencies. The second term (the square law term) will produce the desired up-shifted and down-shifted sidebands since the square of $V_{\mathrm{L.O.}} + V_{\mathrm{RF}}$ contains the cross-product, $2V_{\mathrm{RF}}\, V_{\mathrm{L.O.}}$. This term also produces bias terms and double frequency components. The third-order term will give outputs at the third harmonics of the RF and L.O. frequencies and at $2\omega_{\mathrm{RF}}+\omega_{\mathrm{L.O.}}$, $2\omega_{\mathrm{RF}}-\omega_{\mathrm{L.O.}}$, $2\omega_{\mathrm{L.O.}}+\omega_{\mathrm{RF}}$, and $2\omega_{\mathrm{L.O.}}-\omega_{\mathrm{RF}}$. Normally these products are far removed from the desired output band and can be filtered out. If the input voltage is small enough, we do not have to continue the expansion. For larger signals, however, the next term (fourth-order) gives undesirable products within the desired output band. To see how this happens, consider an input signal with two components, $A_1\cos(\omega_1 \mathrm{t})$ and $A_2\cos(\omega_2 \mathrm{t})$. One of the fourth-order output terms will be, except for a constant,

$$\cos(\omega_{\mathrm{L}}t)[\mathrm{A}_1 \cos(\omega_1 t) + \mathrm{A}_2 \cos(\omega_2 t)]^3. \tag{5.7}$$

You can expand this expression to show that it contains components with frequencies $\omega_{\mathrm{L}}+2\omega_1+\omega_2$ and $\omega_{\mathrm{L}}-2\omega_1-\omega_2$. When ω_1 and ω_2 are close to each other, $2\omega_1+\omega_2$ and $2\omega_1-\omega_2$ are nearby and can lie within the desired output band. In a radio receiver, this means that two strong signals will create an objectionable mixing product at a nearby, i.e., inband, frequency which will impede the reception of a weak signal at that frequency.

Simple mixers can also be made with a transistor. A bipolar transistor, if driven by the sum of the RF and L.O. voltages, will have a collector current containing the same set of frequency components as the diode mixer discussed above. Sometimes a dual-gate FET is used as a mixer; the L.O. voltage is applied to one gate and the RF voltage is applied to the other. This provides some isolation between L.O. and RF (which is provided automatically in a balanced mixer such as the diode ring mixer).

We will see later that multiplication, the basis of mixing, is also the operation needed to modulate the amplitude of a *carrier* sine wave, i.e., to produce amplitude modulation (AM). Multiplication, mixing, and AM modulation are all the same basic operation.

Problems

Problem 5.1. Sometimes two multipliers, two phase shifters, and an adder are used to build a mixer that has only one output band (a so-called single-sideband mixer). The design for an *upper* sideband mixer, for example, follows directly from the identity:

$$\cos([\omega_{\mathrm{RF}}+\omega_{\mathrm{L.O.}}]t)=\cos(\omega_{\mathrm{RF}}t)\cos(\omega_{\mathrm{L.O.}}t)-\sin(\omega_{\mathrm{RF}}t)\sin(\omega_{\mathrm{L.O.}}t).$$

Draw a block diagram for a circuit that carries out this operation.

Problem 5.2. The diode ring switching mixer also works when the L.O. and RF ports are interchanged. Explain the operation in this case.

Problem 5.3. Show that the diode ring switching mixer will work if the L.O. frequency is one-third of the nominal L.O. frequency. This is sometimes done for convenience if this $\frac{1}{3}\omega_{\mathrm{L.O.}}$ frequency is readily available. Find the conversion gain (loss) for this situation. Why would this scheme not work if the L.O. frequency is half the nominal L.O. frequency?

Problem 5.4. Consider a situation where two signals of the same frequency but with a phase difference, θ, are separately mixed to a new frequency. Suppose identical mixers are used and that they are driven with the same L.O. signal. Show that the phase difference of the shifted signals is still θ.

Problem 5.5. In RF engineering, considerable use is made of the trigonometry identities $\cos(a+b)=\cos(a)\cos(b)-\sin(a)\sin(b)$ and $\sin(a+b)=\sin(a)\cos(b)+\cos(a)\sin(b)$. Prove these identities, either using geometric constructions or using the identity $e^{jx}=\cos(x)+j\sin(x)$.

CHAPTER

6

Amplitude and frequency modulation

Modulation means *variation* of the amplitude or the phase (or both) of an otherwise constant sinusoidal RF *carrier wave* in order that the signal carry information: digital data or analog waveforms such as audio or video. In this chapter we look at pure amplitude modulation (AM) and pure frequency modulation (FM). Historically, these were the first methods to be used for communications and broadcasting. While still used extensively, they are giving way to modulation schemes, mostly digital, some of which amount to simultaneous AM and FM.

The simplest form of AM is *on/off keying*. This binary digital AM (full on is a data “1” and full off is a data “0”) can be produced with a simple switch, originally a telegraph key in series with the power source or the antenna. The first voice transmissions used a carbon microphone as a variable resistor in series with the antenna to provide a continuous range of amplitudes. With AM, the frequency of the carrier wave is constant, so the zero crossings of the RF signal are equally spaced, just as they are for an unmodulated carrier. The simplest FM uses just two frequencies; the carrier has frequency f_0 for data “zero” and $f_0 + \Delta f$ for data “one.” FM is usually generated by a VCO (voltage-controlled oscillator). For binary FSK (frequency shift keying), the control voltage has only two values: one produces f_0 and the other produces $f_0 + \Delta f$. FM broadcasting uses a continuous range of frequencies; the instantaneous frequency is determined by the amplitude of the audio signal. With FM, the amplitude of the carrier wave signal is constant. Figure 6.1 shows an unmodulated carrier wave, an AM-modulated wave, an FM-modulated wave, and a wave with simultaneous AM and FM modulation.

Phase modulation is a type of frequency modulation, since frequency is the time derivative of phase.

Amplitude and frequency (or phase) exhaust the list of carrier wave properties that can be modulated. The fractional bandwidth of RF signals is low enough that if one zooms in on a stretch of several cycles, the waveform is essentially sinusoidal and can be described by just its amplitude and frequency. Schemes such as single-sideband suppressed carrier (SSBSC), double-sideband

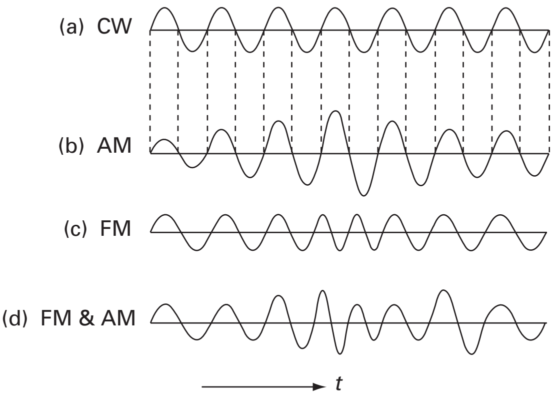

Figure 6.1. (a) Unmodulated wave; (b) with AM modulation; (c) with FM modulation; (d) with simultaneous AM and FM.

suppressed carrier AM (DSBSC), quadrature amplitude modulation (QAM), and others, produce simultaneous amplitude and frequency modulation. A sine-wave generator with provisions for both amplitude and phase modulation can, in principle, produce a radio signal with *any* specified type of modulation, for example, the comb of 52 individually modulated subcarriers used in Wi-Fi data links. Note: the term *digital modulation* means that the modulation uses a finite number of modulation states (often just two) rather than a continuum of states and that the time spent in any state is an integral multiple of a fundamental symbol time (baud).

6.1 Amplitude modulation

Amplitude modulation, the first scheme used for radio broadcasting, is still used in the long-wave, middle-wave, and short-wave broadcast bands.[1] Let us examine AM, first in the time domain and then in the frequency domain.

6.1.1 AM in the time domain

Without modulation (when the speech or music is silent) the voltage applied to the antenna is a pure sine wave at the carrier frequency. The power of a radio

---

[1] The long-wave (LW) band, used only outside the Western Hemisphere, extends from 153 to 179 kHz. The middle-wave (MW) band extends from 520 to 1700 kHz. Short-wave bands are usually identified by wavelength: 75 m, 60 m, 49 m, 41 m, 31 m, 25 m, 19 m, 16 m, 13 m and 11 m. The spacing between LW frequency assignments is 9 kHz. MW spacings are 10 kHz in the Western Hemisphere and 9 kHz elsewhere. Short-wave frequency assignments are less coordinated but almost all short-wave stations operate on frequencies which are an integral number of kHz.

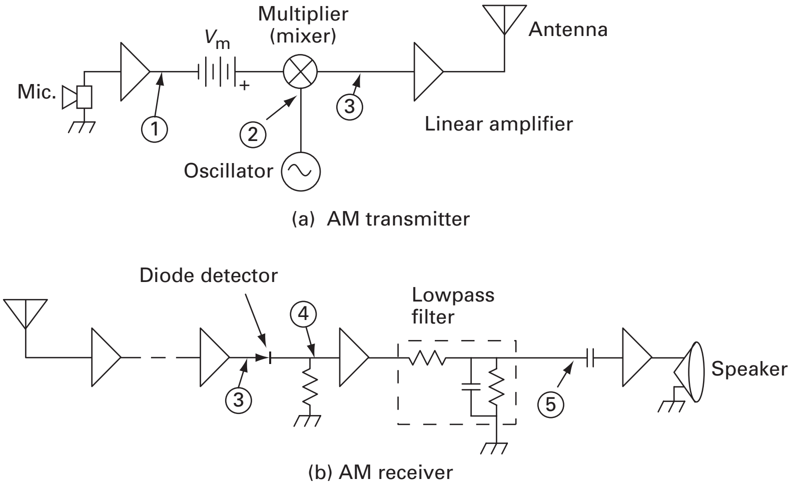

Figure 6.2. Hypothetical AM transmitter and receiver.

station is defined as the transmitter output power when the modulation is zero. The presence of an audio signal changes the amplitude of the carrier. Figure 6.2 shows a hypothetical AM transmitter and receiver.

The audio signal (amplified microphone voltage) has positive and negative excursions but its average value is zero. Suppose the audio voltage is bounded by $+V_m$ and $-V_m$. A dc bias voltage of V_m volts is added to the audio voltage. The sum, $V_m + V_{audio}$, is always positive and is used to multiply the carrier wave, $\sin(\omega_c t)$. The resulting product is the AM signal; the amplitude of the RF sine wave is proportional to the biased audio signal. The simulation in Figure 6.3 shows the various waveforms in the transmitter and receiver of Figure 6.2 corresponding to a random segment of an audio waveform. The biased audio waveform is called the *modulation envelope*. Note that at full modulation where $V_{audio} = +V_m$, the carrier is multiplied by $2V_m$ whereas, at zero modulation, the carrier is multiplied by V_m (bias only). This factor of 2 in amplitude means the fully (100%) modulated signal has four times the power of the unmodulated signal (the carrier wave alone). It follows that the antenna system for a 50 000 W AM transmitter must be capable of handling 200 000 W peaks without breakdown. The average power of the modulated signal is determined by the average square of the modulation envelope. For example, in the case of 100% modulation by a single audio tone, the average power of the modulated signal is greater than the carrier by a factor of $\langle(1+\cos\theta)^2\rangle = \langle 1+2\cos(\theta) + \cos(\theta)^2\rangle = 1 + \langle\cos(\theta)^2\rangle = 3/2$.

Figure 6.2 also shows how the receiver demodulates the signal, i.e., how it recovers the modulation envelope from the carrier wave. The detector is just an ideal rectifier, which eliminates the negative cycles of the modulated RF signal. A simple lowpass filter then produces the average voltage of the positive loops. (Usually this is a simple *RC* filter, but the lowpass filter used in Figure 6.3 is a

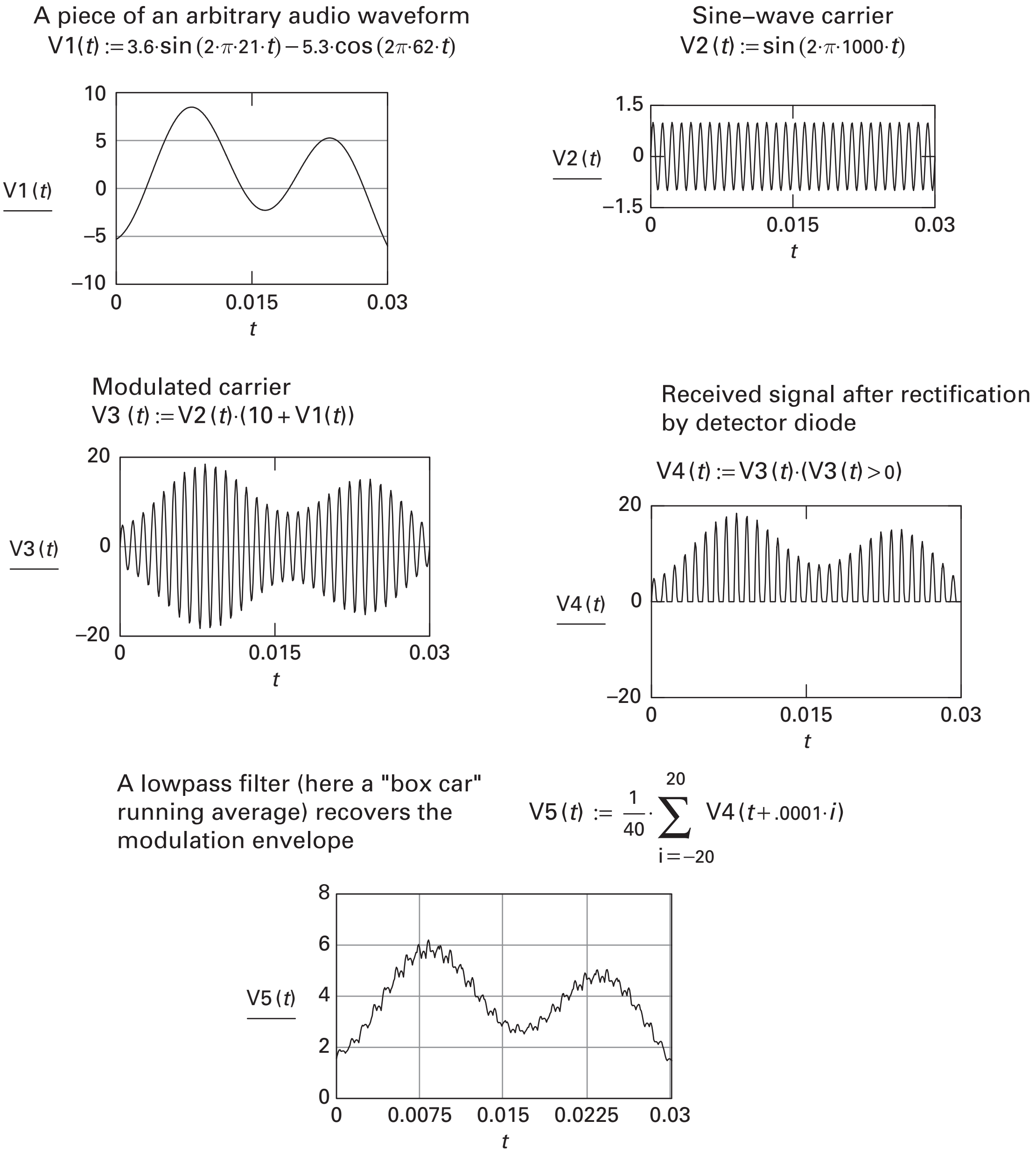

Figure 6.3. Waveforms in the AM system of Figure 6.2.

"boxcar" integrator that forms a running average.) Finally, ac coupling removes the bias, leaving an audio signal identical to the signal from the microphone.

6.1.2 AM in the frequency domain

Let us look at the spectrum of the AM signal to see how the power is distributed in frequency. This is easy when the audio signal is a simple sine wave, say V_a sin

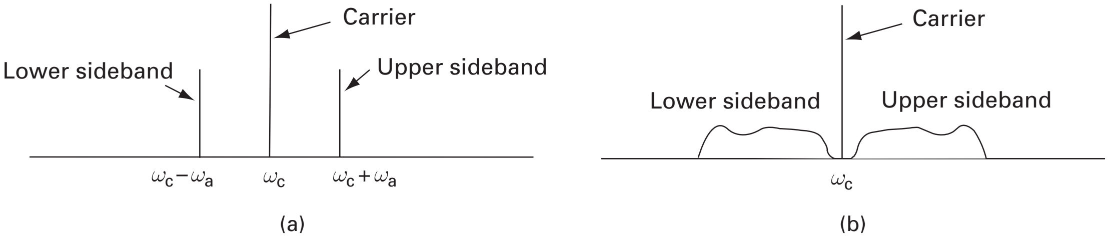

Figure 6.4. Spectrum of an AM power spectrum: (a) modulation by a single-tone modulation; (b) modulation by an audio spectrum.

$(\omega_a t)$. The voltage at point 3 in Figure 6.2 is then $\sin(\omega_c t)(V_m + V_a \sin(\omega_a t)) = V_m \sin(\omega_c t) + 1/2\ V_a \cos([\omega_c - \omega_a]t) - 1/2\ V_a \cos([\omega_c + \omega_a]t)$. These three terms correspond to the *carrier* at ω_c with power $V_m^2/2$, a *lower sideband* at $\omega_c - \omega_a$ with power $V_a^2/8$, and an *upper sideband* at $\omega_c + \omega_a$, also with power $V_a^2/8$. This spectrum is shown in Figure 6.4(a).

Note that the carrier, the component at ω_c, is always present and that its amplitude is independent of the modulation level. At 100% sine-wave modulation $V_a = V_m$ and the average power in each sideband is $\frac{1}{4}$ of the carrier power. The maximum average power is therefore $\frac{3}{2}$ times the carrier power, as we saw earlier from the time domain description. The sidebands, which carry all the information, account here for only $\frac{1}{3}$ of the total power transmitted. With speech waveforms the sidebands may contain even less power.

If, instead of a single sine wave, the audio signal is a superposition of sine waves with different frequencies (and amplitudes), the above analysis is readily extended to show that the upper and lower sidebands become continuous bands of frequency components straddling the carrier symmetrically, as shown in Figure 6.4(b). As the audio signal changes, the shape and size of the sidebands change. A typical audio signal has components at frequencies as high as 10 KHz. AM broadcast stations restrict their audio to not exceed 10 KHz but this still produces a 20 KHz-wide spectrum. The spacing between frequency assignments in the AM broadcast band is only 10 KHz. To prevent overlap, no two stations in a given listening area are assigned the same frequency or adjacent frequencies.

6.2 Frequency and phase modulation

In frequency modulation (FM) and phase modulation (PM) the audio voltage controls the phase angle of the sinusoidal carrier wave while the amplitude remains constant. Since frequency is the time derivative of phase, the two quantities cannot be varied independently; FM and PM are only slightly different methods of *angle modulation*. The advantage of FM (or PM) broadcasting over AM broadcasting is that, under strong signal conditions, the audio signal-to-noise ratio at the output of the receiver can be much higher for FM than for AM. The only price paid for this improvement is increased bandwidth (this topic is discussed in detail in Chapter 23).

6.2.1 FM in the time domain

A simple unmodulated carrier wave is given by $V(t)=\cos(\omega_c t+\phi_0)$ where ω_c and ϕ_0 are constants. The phase, $\phi(t)=(\omega_c t+\phi_0)$, of this cw signal increases linearly with time at a rate ω_c radians per second. In FM, the instantaneous frequency is made to shift away from ω_c by an amount proportional to the modulating audio voltage, i.e., $\omega(t)=\omega_0+k_{osc}V_a(t)$. A linear voltage-controlled oscillator (VCO) driven by the audio signal, as shown in Figure 6.5a, makes an FM modulator.

The excursion from the center frequency, $k_{osc}V_a$, is known as the (radian/s) *deviation*. In FM broadcasting the maximum deviation, defined as 100% modulation, is 75 kHz, i.e., $k_{osc}V_{a_{max}}/2\pi=75\cdot10^3$. Consider the case of modulation by a single audio tone, $V_a(t)=A\cos(\omega_a t)$. The instantaneous frequency of the VCO is then $\omega(t)=\omega_c+k_{osc}A\cos(\omega_a t)$. And, since the instantaneous frequency is the time derivative of phase, the phase is given by the integral

$$\phi(t)=\int\omega(t)\mathrm{d}t=\int(\omega_c+k_{osc}A\cos(\omega_a t))\mathrm{d}t=\omega_c t+\frac{k_{osc}A}{\omega_a}\sin(\omega_a t)+\phi_0. \tag{6.1}$$

From Equation (6.1), we see that, in addition to the linearly increasing phase, $\omega_c t$, of the unmodulated carrier, there is a phase term that varies sinusoidally at the audio frequency. The amplitude of this sinusoidal phase term is $k_{osc}A/\omega_a$. This maximum phase excursion is known as the *modulation index*. The inverse dependence on the modulation frequency, ω_a, results from phase being the integral of frequency. Phase modulation (PM) differs from FM only in that it does not have this ω_a denominator; it could be produced by a fixed-frequency oscillator followed by a voltage-controlled phase shifter. An indirect way to produce a PM signal is to use a standard FM modulator (i.e., a VCO), after first passing the audio signal through a differentiator. The differentiator produces a factor, ω_a, which cancels the ω_a denominator produced by the VCO. This scheme is shown in Figure 6.6.

Similarly, a PM transmitter can be adapted to produce FM. A possible receiver for PM would be the receiver of Figure 6.5(b) with an integrator after

Figure 6.5. Basic FM system: (a) transmitter; (b) receiver.

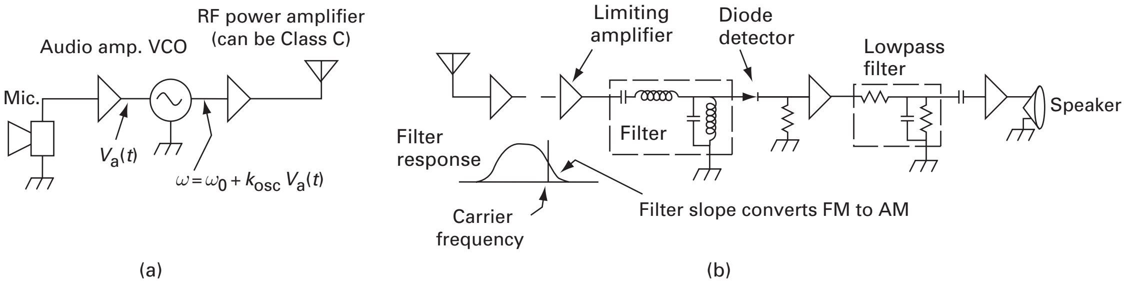

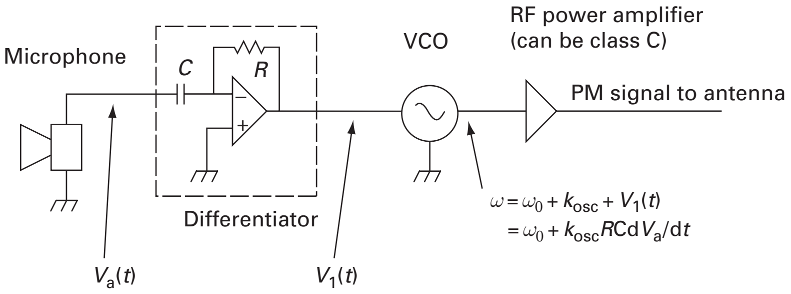
Microphone
C
R
−
+
Differentiator
VCO
RF power amplifier
(can be class C)
PM signal to antenna
$\omega = \omega_0 + k_{\mathrm{osc}} + V_1(t)$
$= \omega_0 + k_{\mathrm{osc}} RC \mathrm{d} V_{\mathrm{a}}/\mathrm{d}t$
$V_{\mathrm{a}}(t)$
$V_1(t)$

Figure 6.6. Phase modulation via frequency modulation.

the diode detector to undo the differentiation done in the transmitter of Figure 6.6. (See Chapter 18 for more FM and PM detectors.) In the simple FM receiver of Figure 6.5(b), a bandpass filter converts FM into AM because the carrier frequency is placed on the slope of the filter. When the carrier frequency increases, the filter response decreases the amplitude. After this filter, the rest of the receiver is an AM receiver. A limiting amplifier ahead of the filter is a refinement to eliminate amplitude noise, as well as to make the audio volume independent of signal strength. More refined AM and FM detectors are presented in Chapter 18.

6.2.2 Frequency spectrum of FM

Using the expression for phase from Equation (6.1) (neglecting ϕ_0) and using the usual expansion for $\cos(a+b)$, the complete FM signal (VCO output) for single-tone modulation becomes

$$V(t) = \cos(\phi(t)) = \cos(\phi_{\mathrm{m}} \sin(\omega_{\mathrm{a}} t)) \cos(\omega_{\mathrm{c}} t) - \sin(\phi_{\mathrm{m}} \sin(\omega_{\mathrm{a}} t)) \sin(\omega_{\mathrm{c}} t) \tag{6.2}$$

where $\phi_{\mathrm{m}} = k_{\mathrm{osc}} A/\omega_{\mathrm{a}}$ is the modulation index. Let us first consider the case where the modulation index is small.

6.2.3 Narrowband FM or PM

At very low modulation levels, both FM and PM produce a power spectrum similar to AM, i.e., a carrier with an upper sideband ω_{a} above the carrier and a lower sideband ω_{a} below the carrier. To see this, we expand Equation (6.2) for very small ϕ_{m}, obtaining

$$V(t) \approx \left(1 - \frac{\phi_{\mathrm{m}}^2 \sin^2(\omega_{\mathrm{a}} t)}{2}\right) \cos(\omega_{\mathrm{c}} t) - \phi_{\mathrm{m}} \sin(\omega_{\mathrm{a}} t) \sin(\omega_{\mathrm{c}} t). \tag{6.3}$$

The first term is an almost constant amplitude spike at the carrier frequency, much like the carrier of an AM signal. You can verify directly that the amplitude of $V(t)$ remains constant by summing the squares of the coefficients of $\cos(\omega_c t)$ and $\sin(\omega_c t)$. Expanding the right-hand term gives the two sidebands,

$$V(t) \approx \cos(\omega_c t)\left(1 - \frac{\phi_m^2 \sin^2(\omega_a t)}{2}\right) + (\phi_m/2)\cos[(\omega_c + \omega_a)t] - (\phi_m/2)\cos[(\omega_c - \omega_a)t]. \tag{6.4}$$

This differs from an AM signal in that the sidebands are equivalent to having multiplied the audio signal not by $\cos(\omega_c t)$, the carrier, but by $\sin(\omega_c t)$. Note also that, in this limit of small modulation index, the amplitude of the sidebands is small and the width of the spectrum stays fixed at $2\omega_a$, just as with AM.

6.2.4 Wideband FM spectrum

Looking at Equation (6.2), we see that the signal consists of the products of $\cos(\omega_c t)$ and $\sin(\omega_c t)$, multiplied respectively by the baseband signals $\cos[\phi_m \sin(\omega_a t)]$ and $\sin[\phi_m \sin(\omega_a t)]$. When ϕ_m is not small, these products are similar and each consists of upper and lower sidebands around ω_c, with frequency components spaced from the carrier frequency by integral multiples of ω_a. The $\cos(\omega_c t)$ and $\sin(\omega_c t)$ carriers are suppressed and the spectrum has no distinct central carrier spike. An exact analysis uses Fourier analysis of the cos(sin) and sin(sin) terms[2] to find the comb of sidebands, but a simpler argument can give an estimate of the bandwidth of the FM signal. Referring again to Equation (6.2), the baseband modulating signals produce sidebands. In a time equal to one-quarter of an audio cycle, the expression $\phi_m \sin(\omega_a t)$ changes by ϕ_m radians, so the phase of the baseband modulating signals changes by ϕ_m radians. Dividing this phase change by the corresponding time interval, $1/(4f_a) = \pi/(2\omega_a)$, the average sideband frequency is given by $2\phi_m \omega_a/\pi$. Substituting $\phi_m = k_{osc} A/\omega_a$, the average frequency is $2k_{osc}A/\pi = (2/\pi) \times$ deviation. Therefore, the one-sided bandwidth of the FM signal, when the modulation index is large, is roughly equal to the maximum deviation and the full bandwidth is about twice the maximum deviation. If the deviation is reduced, the bandwidth goes down proportionally at first but then, in the narrowband regime, stays constant at twice the audio bandwidth, as shown above.

---

[2] The terms in Equation (6.2) can be expanded into harmonics of the audio frequency by using the Bessel function identities:

$$\cos(\phi \sin(\omega t)) = J_0(\phi) + 2\sum_{n=1}^{n=\infty} J_{2n}(\phi)\cos(2n\omega t)$$

and

$$\sin(\phi \sin(\omega t)) = 2\sum_{n=0}^{n=\infty} J_{2n+1}(\phi)\sin((2n+1)\omega t).$$

6.2.5 Frequency multiplication of an FM signal

When a signal is passed through a times-N frequency multiplier its phase is multiplied by N. A square-law device, for example, can serve as a frequency doubler since $\cos^2(\phi) = 1/2 + 1/2\cos(2\phi)$. So if the phase of an FM signal, ϕ, (the right-hand expression in Equation 6.1), is multiplied by 2 by using a frequency doubler, both ω_c and kA/ω_a are multiplied by 2, i.e., the frequency and the modulation index are doubled. When a given VCO cannot be linearized well enough within the intended operating range, it can be operated with a low deviation and its signal can be frequency-multiplied to increase the deviation. If the resulting center frequency is too high, an ordinary mixer can shift it back down, preserving the increased deviation.

6.3 AM transmitters

The simple AM transmitter shown in Figure 6.2 is entirely practical. (Of course we would fix it up so that a battery would not be needed to supply the bias voltage.) The linear RF power amplifier would be the major part of this transmitter. We saw in Chapter 3 that a class-B linear amplifier has relatively high efficiency, π/4 or 78.5%, but that is for a maximum-amplitude sine wave. Let us find the efficiency for the transmitter of Figure 6.2, assuming the average modulation power is 10% of the peak modulation power. (The "crest" factor for speech is often taken to be 16 (12 dB), but broadcasters normally use dynamic range compression to increase "loudness.") Let the output signal be given by

$$V_{\text{out}} = V_{\text{dc}}\tfrac{1}{2}[1 + V_{\text{m}}(t)]\cos(\omega t) \tag{6.5}$$

where $|V_{\text{m}}(t)| \leq 1$. Note that the maximum output voltage is V_{dc}, the power supply voltage. (Refer to the amplifier circuit of Figures 3.7 or 3.8.) Remembering that $\langle\cos^2\rangle = 1/2$ and assuming that the audio signal is an ac signal, i.e., $\langle V_{\text{m}}(t)\rangle = 0$, we can express the average output power:

$$\begin{aligned} P_{\text{out}} &= \langle V_{\text{out}}{}^2\rangle/R = \tfrac{1}{4}V_{\text{dc}}{}^2\langle[1 + V_{\text{m}}(t)]^2\rangle/(2R) \\ &= \tfrac{1}{4}V_{\text{dc}}{}^2(1 + 0.1)/(2R) = 1.1V_{\text{dc}}{}^2/(8R), \end{aligned} \tag{6.6}$$

where R is the load (antenna) resistance and where we have used the fact that $\langle\cos(\omega t)\rangle = 0$. The average power delivered by the dc supply is given by

$$\begin{aligned} \langle V_{\text{dc}}I_{\text{supply}}\rangle &= V_{\text{dc}}\langle|V_{\text{out}}(t)/R|\rangle \\ &= V_{\text{dc}} \cdot \tfrac{1}{2}V_{\text{dc}}\langle[1 + V_{\text{m}}(t)/R\rangle 2/\pi = V^2_{\text{dc}}/(\pi R), \end{aligned} \tag{6.7}$$

where we have used the fact that the average of a positive loop sin(x) is equal to 2/π. Calculating the efficiency, we find

$$\eta = \frac{\text{power out}}{\text{power in}} = (1.1)/V_{\text{dc}}{}^{2}/(8R) \div V_{\text{dc}}^{2}/\pi = 1.1\pi/8 = 43.2\%. \qquad (6.8)$$

Therefore, when running this AM transmitter, only 43.2% of the prime power gets to the antenna; the rest is dissipated as heat in the class-B amplifier.

Almost all AM transmitters obtain higher efficiency by using class-C, D, or E RF amplifiers. These amplifiers are not linear in the normal sense, that is, the output signal amplitude is not a constant multiple of the input signal amplitude. But, for a fixed input RF amplitude, the output amplitude *is* proportional to the supply voltage. These amplifiers can therefore be used as high-power *multipliers* that form the product of the power supply voltage times a unit sine wave at the RF frequency. Furthermore, the efficiency of these amplifiers, which is high, is essentially independent of the supply voltage. (These amplifiers are discussed in detail in Chapter 9.) Let us look at the overall efficiency of transmitters using these amplifiers.

6.3.1 AM transmitter using a class-C RF amplifier and a class-B modulator

In the traditional AM transmitter, audio voltage developed by a high-power class-B audio amplifier (the modulator) is added to a dc bias and the sum of these voltages powers a class-C RF amplifier. As explained above, a class-C RF amplifier acts as a high-power multiplier. In the traditional tube-type circuit of Figure 6.7, the dc bias is fed through the secondary of the modulation transformer. Audio voltage produced by the modulator appears across the secondary winding and adds to the bias voltage. With no audio present, the class-B audio amplifier consumes negligible power and the bias voltage supply provides power for the carrier. At 100% sine-wave modulation, the modulator must supply audio power equal to half the bias supply power. A 50 000 watt transmitter thus requires a modulator that can supply 25 000 W of audio power. (Again, this result is for a single tone, but is essentially the same for speech or music.)

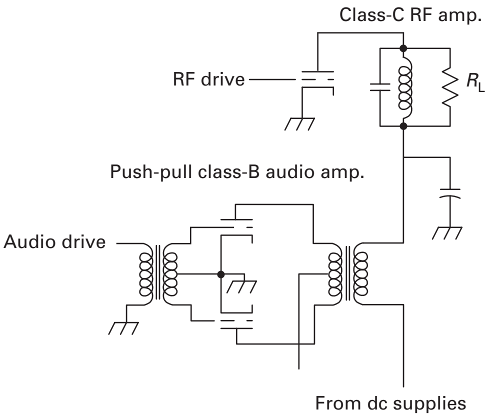

Figure 6.7. Class-B plate-modulated AM transmitter.

Let us find the efficiency of this transmitter. We will assume the class-C amplifier has an efficiency of 80% and that, as before, the average modulation power is 10% of the maximum possible modulation power. If the normalized output carrier power is 1 W, the dc bias supply must provide 1/0.8 W and the modulator must supply 0.1/0.08 W. To handle peaks, the maximum power from the modulator must be 1/0.08 W. To find the efficiency of the class-B modulator, we must know not only its peak output voltage, but also the mean of the absolute value of its output voltage. (Note that this last piece of information was not needed in the analysis of the class-B RF amplifier transmitter discussed above.) Let us just assume the modulating signal is a single audio sine wave whose power is 0.1/0.8 W. With that assumption, the modulator will have an efficiency of 0.35. The total input power, carrier plus modulation, will therefore be 1/0.8 +(0.1/0.35)/0.8 = 1.61 W. The output power will be 1 + 0.1, so the overall efficiency of this transmitter is 68%, a significant improvement over the transmitter using a linear class-B RF amplifier.

6.3.2 Class-C RF amplifier with a switching modulator

There are several newer methods to produce AM with even higher efficiency. All use switching techniques. The modulator shown in Figure 6.8(a) is just a high-power digital-to-analog converter. It uses solid-state switches to add the voltage of many separate low-voltage power supplies, rather than tubes or transistors to resistively drop the voltage of a single high-voltage supply. Modulators of this type can, in principle, be 100% efficient, since the internal switch transistors are either fully on or fully off. The modulator of Figure 6.8(b) is a pulse rate modulator whose output voltage is equal to the supply voltage multiplied by the duty factor of the switch tube.

This type of circuit, whose efficiency can also approach 100%, is discussed in detail with switching power supplies in Chapter 29. Again specifying an average modulation power of 10%, the combination of an 80% efficient class-C RF amplifier together with a switching modulator, makes an AM transmitter with an overall efficiency of 80%.

Figure 6.8. Class-C RF amplifier with (a) high-power D-to-A modulator and (b) duty cycle switching modulator.

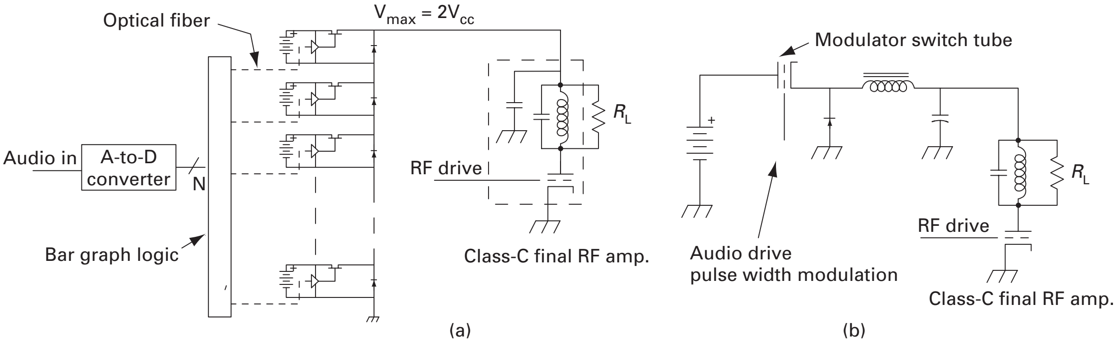

6.3.3 Class-D or E amplifiers with switching modulators

Class D and class-E RF amplifiers can approach 100% efficiency. Like the class-C amplifier, the amplitude of their output sine wave is proportional to the dc supply voltage. When one of these amplifiers is coupled with a switching modulator, the resulting AM transmitters can approach an overall efficiency of 100% for any modulation level.

6.4 FM transmitters

Since FM modulation is generated by a VCO at a low level, and the signal has a constant amplitude, there is no other modulator nor is there any requirement that the final power amplifier have a linear amplitude response. Class-C amplifiers have been used most often, with efficiencies around 80%.

6.5 Current broadcasting practice

Many $\frac{1}{2}$ MW to 2 MW AM stations operate in the long-wave, medium-wave, and short-wave bands. These superpower transmitters use vacuum tubes. Standard AM broadcast band transmitters in the U.S. are limited to 50 kW. For this power and lower powers, new AM transmitters manufactured in the U.S. use transistors. These transmitters combine power from a number of modular amplifiers in the 1 kW range. The advantages of solid state include lower (safer) voltages, indefinite transistor life rather than expensive tube changes every year or two, and better “availability;” a defective module only lowers the power slightly and can be replaced while the transmitter remains on the air. Most FM transmitters over about 10 kW still use vacuum tubes, but solid-state FM transmitters are available up to about 40 kW.

Stereophonic sound was added to FM broadcasting around 1960. Existing monophonic receivers operated as they had, receiving the left + right (L + R) audio signal. Thus the system is (backwardly) *compatible*, just as color television was compatible with existing black and white television receivers. The L − R signal information is in the demodulated signal, but clustered around 38 kHz, above the audible range. The L − R signal uses double-sideband suppressed carrier AM modulation, which is explained in Chapter 8. The demodulator to recover the L-R signal is discussed in Chapter 18. Stereo was added to AM radio broadcasting around 1980. Several systems competed to become the standard, but the public was largely indifferent, having already opted for FM stereo. Compatible digital broadcasting has recently been introduced, with versions for both the AM and FM bands. In this IBOC (In-Band On-Channel) system, a station’s digital simulcast signal (which can contain different program material) shares the same channel with the standard AM or FM signal, mostly using the relatively empty spectral regions near (and somewhat past) the nominal edges of the channel. The digital signal, which can be produced by a separate transmitter and may even use a separate antenna, uses COFDM (Coded Orthogonal Frequency-Division Multiplexing), a modulation system, described in Chapter 22. This system is intended to be a stepping-stone to all-digital radio broadcasting in the traditional AM and FM bands. IBOC broadcasting equipment is equipped to transmit an all-digital mode, to be used if and when the traditional analog broadcasting is discontinued.

Problems

Problem 6.1. An AM transmitter, 100% modulated with an audio sine wave, has sidebands whose total power is equal to half the carrier power. Consider 100% modulation by an audio square wave. What is the ratio of sideband power to carrier power?

Problem 6.2. Suppose you are trying to listen to a distant AM station, but another station on the same frequency is coming in at about the same strength. Will you hear both programs clearly? If not, how will they interfere with each other?

Problem 6.3. During periods where the audio signal level is low, the amplitude of an AM signal varies only slightly from the carrier level. The modulation envelope, which carries all the information, rides on top of the high-power carrier. If the average amplitude could be decreased without decreasing the amplitude of the modulation, power could be saved. Discuss how this might be accomplished at the transmitter and what consequences, if any, it might have at the receiver.

Problem 6.4. Show how a PM transmitter can be used to generate FM.

Problem 6.5. Consider an FM transmitter modulated by a single audio tone. As the modulation level is increased, the spectral line at the carrier frequency decreases. Find the value of the modulation index that makes the carrier disappear completely.

Problem 6.6. One of the methods used for compatible AM stereo was a combined AM/PM system. The left + right (L + R) signal AM-modulated the carrier in the usual way, while the L − R signal was used to phase-modulate the carrier. Draw block diagrams of a transmitter and receiver for this system. Why would PM be preferable to FM for the L − R signal?

CHAPTER

7

Radio receivers

In this chapter we will be mostly concerned with the sections of the receiver that come before the detector, sections that are common to nearly all receivers: AM, FM, television, cell phones, etc. Basic specifications for any kind of radio receiver are *gain, dynamic range, sensitivity* and *selectivity*, i.e., does a weak signal at the selected frequency produce a sufficiently strong and uncorrupted output (audio, video, or data) and does this output remain satisfactory in the presence of strong signals at nearby frequencies?

Sensitivity is determined by the noise power contributed by the receiver itself. Usually this is specified as an equivalent noise power at the antenna terminals. Selectivity is determined by a bandpass-limiting filter and might be specified as “3 dB down at 2 kHz from center frequency and 20 dB down at 10 kHz from center frequency.” (Receiver manufacturers usually do not specify the exact bandpass shape.)

7.1 Amplification

Let us consider how much amplification is needed in ordinary AM receivers. One milliwatt of audio power into a typical earphone produces a sound level some 100 dB above the threshold of hearing. A barely discernable audio signal can therefore be produced by −100 dBm (100 dB below 1 mW or 10^{-13} watts). Let us specify that a receiver, for comfortable earphone listening, must provide 50 dB more than this threshold of hearing, or 10^{-8} watts. You can see that, with efficient circuitry, the batteries in a portable receiver could last a very long time! (Sound power levels are surprisingly small; you radiate only about 1 mW of acoustic power when shouting and about 1 nW when whispering.) How much signal power arrives at a receiver? A simple wire antenna could intercept 10^{-8} watts of RF power at a distance of about 20 000 km from a 10 kW radio station at 1 MHz having an omnidirectional transmitting antenna, so let us first consider “self-powered” receivers.

7.2 Crystal sets

The earliest radios, crystal sets, were self-powered. A crystal diode rectifier recovered the modulation envelope, converting enough of the incoming RF power into audio power to drive the earphone. A simple *LC* tuned circuit served as a bandpass filter to select the desired station and could also serve as an antenna matching network. The basic crystal set receiver is shown in Figure 7.1.

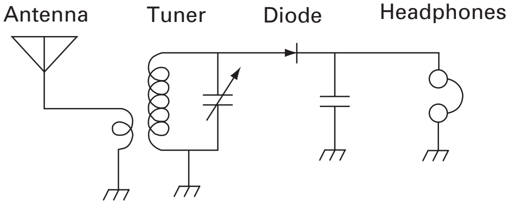

Figure 7.1. Self-powered crystal set receiver.

The considerations given above show that a self-powered receiver can have considerable range. But when the long-wire antenna is replaced by a compact but very inefficient loop antenna and the earphone is replaced by a loudspeaker, amplification is needed. In addition, we will see later that the diode detector, when operated at low signal levels, has a square-law characteristic, which causes the audio to be distorted. For proper envelope detection of an AM signal, the signal applied to a diode detector must have a high level, several milliwatts. The invention of the vacuum tube, followed by the transistor, provided the needed amplification. Receivers normally contain both RF and audio amplifiers. RF amplification provides enough power for proper detector operation, while subsequent audio amplification provides the power to operate loudspeakers.

7.3 TRF receivers

The first vacuum tube radios used a vacuum tube detector instead of a crystal, and added RF preamplification and audio post-amplification, as described above. These *TRF* (Tuned Radio Frequency) sets[1] had individual tuning adjustments for each of several cascaded RF amplifier stages. Changing stations

---

[1] Early radios were called “radio sets” because they were literally a set of parts including one or more tubes and batteries (or maybe just a crystal detector), inductors, “condensers,” resistors, and headphones. Many of these parts were individually mounted on wooden bases, and, together with “hook-up” wires, would spread out over a table top.

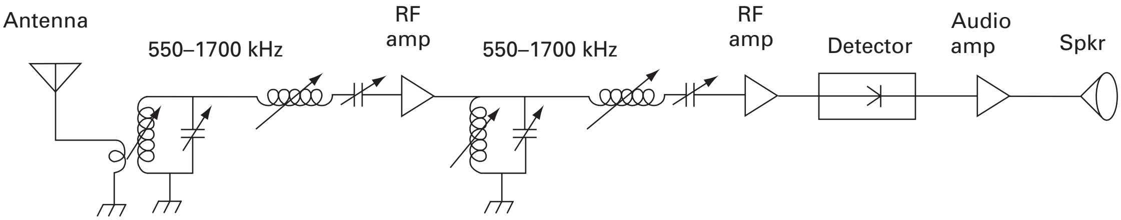

Figure 7.2. TRF receiver.

required the user to adjust several dials (often with the aid of a tuning chart or graph).

Figure 7.2 shows a hypothetical TRF receiver with cascaded amplifiers and bandpass filters.

Note that all the inductors and capacitors should be variable in order to tune the center frequency of the bandpass filters and also maintain the proper bandwidth, which is about 10 kHz for AM. (In a practical circuit, the bandpass filters would use a coupled-resonator design rather than the straightforward lowpass-to-bandpass conversion design shown here.)

7.4 The superheterodyne receiver

The disadvantages of TRF sets were the cost and inconvenience of having many tuning adjustments. Most of these adjustments were eliminated with the invention of the superheterodyne circuit by Edwin H. Armstrong in 1917. Armstrong's circuit consists of a fixed-tuned, i.e., single-frequency, TRF back-end receiver, preceded by a frequency converter front-end (mixer and local oscillator) so that the signal from any desired station can be shifted to the frequency of the TRF back-end. This frequency is known as the intermediate frequency or IF. The superheterodyne is still the circuit used in nearly every radio, television, and radar receiver. Among the few exceptions are some toy walkie-talkies, garage-door openers, microwave receivers used in radar-controlled business place door openers, and highway speed trap radar detectors. Figure 7.3 shows the classic broadcast band "superhet."

Selectivity is provided by fix-tuned bandpass filters in the IF amplifier section. The AM detector here is still a diode, i.e., the basic envelope detector. (In Chapter 18 we will analyze this detector, among others.) All the RF gain can be contained in the fixed-tuned IF amplifier, although we will see later that there are sometimes reasons for having some amplification ahead of the mixer as well. Figure 7.3 also serves as the block diagram for FM broadcast receivers, where the IF frequency is usually 10.7 MHZ, and for television receivers, where the IF center frequency is commonly around 45 MHZ.

Note: There was indeed a *heterodyne* receiver that preceded the superheterodyne. Invented by radio pioneer Reginald Fessenden, the heterodnye receiver

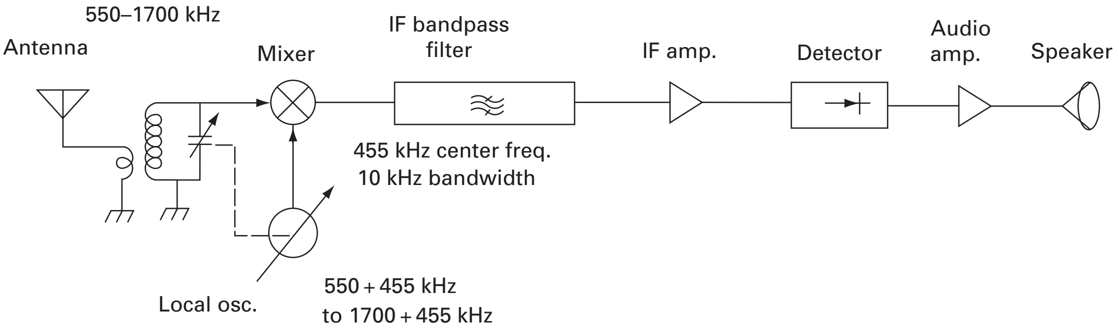

Figure 7.3. Standard superheterodyne receiver for the AM broadcast band.

converted the incoming RF signal directly to audio. This design, known as a "direct-conversion receiver," is disucussed below.

7.4.1 Image rejection

The superhet has some disadvantages of its own. With respect to signals at the input to the mixer, the receiver will simultaneously detect signals at the desired frequency and also any signals present at an undesired frequency known as the *image frequency*. Suppose we have a conventional AM receiver with an IF frequency of 455 kHz and suppose the local oscillator is set at 1015 kHz in order to receive a station broadcasting at $1015 - 455 = 560$ kHz (560 kHz is near the lower edge of the AM broadcast band). All the mixers we have considered will also produce a 455-kHz IF signal from any input signal present at 1470 kHz, i.e., 455 kHz above the local oscillator. If the receiver has no RF filtering before the mixer and if there happens to be a signal at 1470 kHz, it will be detected along with the desired 560 kHz signal. A bandpass filter ahead of the mixer is needed to pass the desired frequency and greatly suppress signals at the image frequency. Note that in this example (the most common AM receiver design), this anti-image bandpass filter must be tunable and, for the receiver to have single-dial tuning, the tuned filter must always "track" 455 kHz below the L.O. frequency. In this example, the tracking requirement is not difficult to satisfy; since the image frequency is more than an octave above the desired frequency, the simple one-section filter shown in Figure 7.2 can be fairly broad and still provide adequate image rejection, maybe 20 dB. (Note, though, that 20 dB is not adequate if a signal at the image frequency is 20 dB stronger than the signal at the desired frequency.)

What if a receiver with the same 455 kHz IF frequency is also to cover the short-wave bands? The worst image situation occurs at the highest frequency, 30 MHz, where the image is only about 3% higher in frequency than the desired frequency. A filter 20 dB down at only 3% from its center frequency will need to have many sections, all of which must be tuned simultaneously with a mechanical multisection variable capacitor or voltage-controlled varicaps. As explained above, the center frequency of the filter must be track with a 455 kHz offset from the L.O. frequency in order that the desired signal fall within the narrow IF passband. Image rejection is not simple when the IF frequency is much lower than the input frequency.

Solving the image problem

A much higher IF frequency can solve the image problem. If the AM broadcast band receiver discussed above were to have an IF of 10 MHz rather than 455 kHz, the L.O. could be tuned to 10.560 MHz to tune in a station at 560 kHz. The image frequency would be 20.560 MHz. As the radio is tuned up to the top end of the AM broadcast band, 1700 kHz, the image frequency increases to 21.700 MHz. In this case, a fixed-tuned bandpass filter, wide enough to cover the entire broadcast band, can be placed ahead of the receiver to render the receiver insensitive to images. This system is shown in Figure 7.4. Only the local oscillator needs to be changed to tune this receiver. Of course, the 10 MHz IF filter must still have a narrow 10 kHz passband to establish the receiver's basic selectivity.

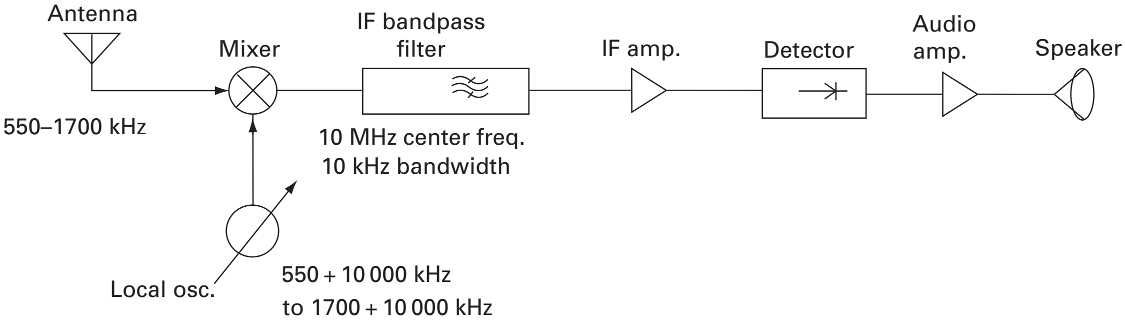

Figure 7.4. Image-free broadcast receiver using a 10 MHz IF.

The circuit of Figure 7.4 is entirely practical although it is more expensive to make the necessary narrowband filters at higher frequencies; quartz crystal resonator elements must be used in place of lumped *LC* elements. If the input band is wider, e.g., 3–30 MHz for a short-wave receiver, the IF frequency would have to be much higher, and narrowband filters become impractical. (Even at 10 MHz, a bandwidth of 10 kHz implies a very narrow fractional bandwidth, 0.1%.) A solution to both the image problem and the narrow fractional bandwidth problem is provided by the *double-conversion superhet*.

Double conversion superhet

Figure 7.5 shows how a second frequency converter takes the first IF signal at 10 MHz and converts it down to 455 kHz, where it can be processed by the standard IF section of the receiver of Figure 7.3. The 10-MHz first IF filter can be wider than the ultimate passband.

Suppose, for example, that the first IF section has a bandwidth of 500 kHz. The second L.O. frequency is at 10.455 MHz, so the second mixer would

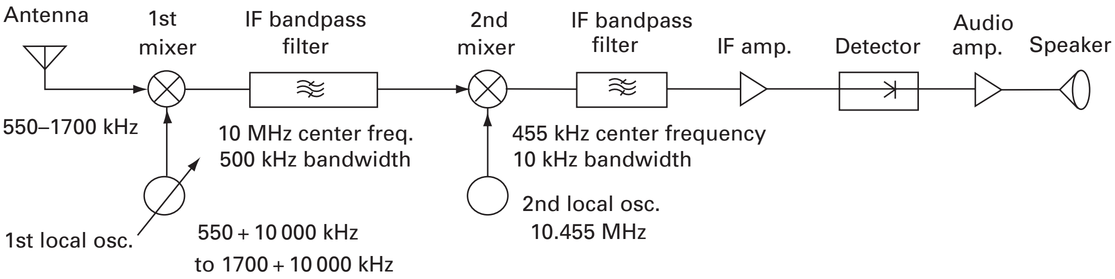

Figure 7.5. Double conversion superhet.

produce an image from a signal at 10.910 MHz. But note that our first IF filter cuts off at 10 MHz + 0.500 MHz /2 = 10.25 MHz, so there will be no signals at 10.910. This system has its own special disadvantages: the receiver usually cannot be used to receive signals in the vicinity of its first IF, since it is difficult to avoid direct feedthrough into the IF amplifier. Multiple conversions require multiple local oscillators and various sum and difference frequency combinations inevitably are produced by nonlinearities and show up as spurious signals known as “birdies.”

Present practice for communications receivers is to use double or triple conversion with a first IF frequency at, say, 40 MHz. The front-end image filter is usually a 30 MHz lowpass filter. In as much as modern crystal filters can have a fairly small bandwidth even at 40 MHz, the output of the first IF section can be mixed down to a second IF with a much lower frequency. Sometimes triple conversion is necessary when the final IF frequency is very low, e.g., 50 kHz. The use of first IF frequencies in the VHF region requires very stable local oscillators but crystal oscillators and frequency synthesizers provide the necessary stability. (Oscillator phase noise was a problem in the first generation of receivers with synthesized local oscillators; the oscillator sideband noise was shifted into the passband by strong signals near the desired signal but outside the nominal passband.)

Image rejection mixer

Another method of solving the image problem is to use an image rejection mixer, such as the circuit shown in Figure 7.6.

This circuit uses two ordinary mixers (multipliers). The lower multiplier forms the product of $\cos(\omega_{\mathrm{L.O}}.t)$ and an input signal, $\cos[(\omega_{\mathrm{L.O}}.\pm\omega_{\mathrm{IF}})\ \mathrm{t}]$, depending on whether the input signal is above or below the L.O. frequency. The upper multiplier has a 90° delay in its connection to the L.O., so it forms the product of $\sin(\omega_{\mathrm{L.O}}.t)$ and $\cos[(\omega_{\mathrm{L.O}}.\pm\omega_{\mathrm{IF}})\mathrm{t}]$. Neglecting the sum frequency terms, the outputs of the upper and lower multipliers are, respectively, $\mp\sin(\omega_{\mathrm{IF}}\mathrm{t})$ and $\cos(\omega_{\mathrm{IF}}t)$. The output from the lower multiplier is delayed by 90°, so the upper and lower signals at the input to the adder are $\mp\sin(\omega_{\mathrm{IF}}t)$ and $\sin(\omega_{\mathrm{IF}}t)$. Thus, for an input frequency of $\omega_{\mathrm{L.O}}. - \omega_{\mathrm{IF}}$, the output of the adder is 2 $\sin(\omega_{\mathrm{IF}}t)$, but when the input frequency is $\omega_{\mathrm{L.O}}. + \omega_{\mathrm{IF}}$, the output of the adder is

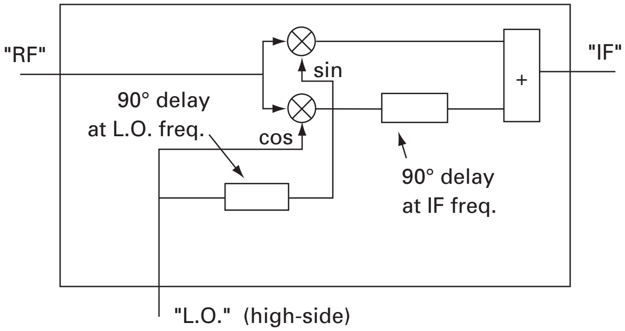

Figure 7.6. Image rejection mixer.

zero. Thus, this mixer rejects signals above the L.O. frequency. The same circuit crops up in Chapter 8, as a *single-sideband generator*. In practice this circuit might provide 20–40 dB of image rejection. It can be used together with the standard filtering techniques to get further rejection.

Zero IF frequency – direct conversion receivers

The evolution of the superhet, which was always toward higher IF frequencies and multiple conversions, has taken a new twist with the advent of the nearly limitless signal processing power available from DSP chips. Direct conversion receivers, in a reversion to the heterodyne architecture, shift the center frequency of the desired signal, ω_C, all the way to zero Hz ("baseband"). Generally this requires that the signal be mixed separately with local oscillator signals $\cos(\omega_C t)$ and $\sin(\omega_C t)$ to preserve all the signal information. To see this, note that the input signal, which is sinusoidal, will, in general, be out of phase at times with $\cos(\omega_C t)$ and at other times with $\sin(\omega_C t)$. Thus, the outputs of the "cosine mixer" or "sine mixer" can go to zero but, together, these "I" (in-phase) and "Q" (quadrature) signals contain all the signal information. To see this, note that the original signal could be easily reconstructed from the I and Q signals. Lowpass filtering of the I and Q signals determines the passband of the receiver, e.g., 5-kHz rectangular lowpass filters on the I and Q signals would give the receiver a flat passband extending 5 kHz above and 5 kHz below the L.O. frequency. The classic image problem, severe at low IF frequencies, disappears when the IF frequency is zero. The low-frequency I and Q signals can be digitized directly for subsequent digital processing and demodulation (see Problem 7.8). No bulky IF bandpass filters are required. Even tiny surface acoustic wave (SAW) bandpass filters are huge, compared to the real estate on DSP chips. With this architecture, nearly 100% of a receiver can be incorporated on a chip, including a tunable frequency synthesizer to produce the L.O. signals. Figure 7.7 shows a block diagram of a direct-conversion receiver. Everything on this diagram plus an L.O. synthesizer is available on a single chip intended for use in HD television receivers.

Figure 7.7. Direct conversion receiver.

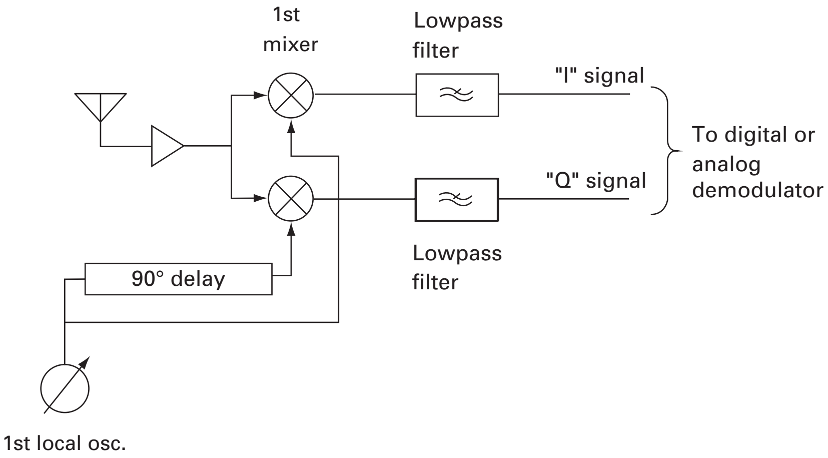

7.4.2 Automatic gain control

Nearly every receiver has some kind of automatic gain control (AGC) to adjust the gain of the RF and/or IF amplifiers according to the strength of the input signal. Without this feature the receiver will overload; overdriven amplifiers go nonlinear (“clip”) and the output will be distorted as well as too loud. The output sound level of an FM receiver, depending on the design of the demodulator, may not vary with signal level but overloading the IF amplifier stages will still produce distortion, so FM receivers also need AGC. Television receivers need accurate AGC to maintain the correct contrast level (analog TV) or threshold levels (digital TV). Any AGC circuit is a feedback control system. In simple AM receivers the diode detector provides a convenient dc output voltage that can control the bias current (and hence gain) of the RF amplifiers. The controlled bias current can also be used to drive a signal strength indicator.

7.5 Noise blankers

Many receivers, including most television receivers, have a noise blanker circuit to reduce the effects of impulse noise such as the spiky noise produced by automobile ignition systems. Here the interfering pulses are of such short duration that the IF stages can be gated off briefly while the interference is present. The duty cycle of the receiver remains high and the glitch is all but inaudible (or invisible). An important consideration is that the gating must be done before the bandwidth is made very narrow since narrow filters elongate pulses.

7.6 Digital signal processing in receivers

Advances in digital electronics, notably fast A-to-D conversion and processors, make it possible to do all-digital filtering and detection in a receiver. Any desired filter amplitude and phase response can be realized.

Besides direct-conversion receivers on a chip, there are many single-chip superhet chips, usually using image cancelling mixers followed by digital bandpass filters operating at low IF frequencies. Adaptive digital filters can correct for propagation problems such as multipath signals. Digital demodulators allow the use of elaborate signal encoding which provides high spectral efficiency (bits/sec/Hz) as well as error correction through the use of redundant bits. Digital modulation techniques are discussed in Chapter 22.

Problems

Problem 7.1. The FM broadcast band extends from 88 to 108 MHz. Standard FM receivers use an IF frequency of 10.7 MHz. What is the required tuning range of the local oscillator?

Problem 7.2. Why are airplane passengers asked not to use radio receivers while in flight?

Problem 7.3. Two sinusoidal signals that are different in frequency, if simply added together, will appear to be a signal at a single frequency but amplitude modulated. This "beat" phenomenon is used, for example, to tune two guitar strings to the same frequency. When they are still at slightly different frequencies, the sound seems to pulsate slowly at a rate equal to their frequency difference. Show that

$$\sin([\omega_0 - \delta\omega]t) + \sin([\omega_0 + \delta\omega]t) = A(t)\sin(\omega_0 t)$$

where

$$A(t) = 2\cos(\delta\omega t).$$

(Note that this addition is a linear process; no new frequencies are generated.)

Problem 7.4. Using an AM receiver in an environment crowded with many stations, you will sometimes hear an annoying high-pitched 10 kHz tone together with the desired audio. If you rock the tuning back and forth the pitch of this tone does not change. What causes this?

Problem 7.5. When tuning an AM receiver, especially at night, you may hear "heterodynes" or whistling audio tones that change frequency as you slowly tune the dial. What causes this? Can it be blamed on the receiver? (Answer: Yes.)

Problem 7.6. Using modern components and digital control, we could build good TRF radios. What advantages would such a radio have over a superheterodyne? What disadvantages?

Problem 7.7. You may have observed someone listening to distorted sound from an AM radio whose tuning is not centered on the station. Often this mistuning is done deliberately when the listener has impaired high-frequency hearing and/or the radio has insufficient bandwidth. What is going on here? Why would a radio not have sufficient bandwidth and why would insufficient bandwidth cause some listeners to tune slightly off station?

Problem 7.8. Design a direct-conversion AM broadcast receiver using digital processing of the I and Q signals. Assume that the L.O. frequency may not be set exactly equal to the frequency of the desired station. Hint: compute the input signal amplitude from the digitized I and Q signals. Then feed this stream of amplitudes into a D-to-A converter.

References

[1] The American Radio Relay League, *The ARRL Handbook for Radio Communications*, 2008 Edition. Almost five pounds of practical circuits, explanations, and construction information.

[2] Gosling, W., *Radio Receivers*, London: Peter Peregrinus, 1986. Good concise discussion of receivers.

[3] Rohde, U. and Bucher, T., *Communications Receivers, Principles & Design*, Second Edition, New York: McGraw-Hill, 1988. A whole course in itself.

[4] Rohde, U. and Whitaker, J., *Communications Receivers*, Third Edition, New York: McGraw-Hill, 2001. The third edition includes material on digital processing in receivers.

CHAPTER

8

Suppressed-carrier AM and quadrature AM (QAM)

Viewed just in the time domain, ordinary AM, as used in broadcasting, seems so obvious that one would scarcely imagine how it might be done otherwise. But viewed in the frequency domain, as in Chapter 6, this system shows some obvious inefficiencies. First, most of the average transmitted power (about 95% when transmitting typical audio material) is in the carrier, which is a spike or "delta function" in the frequency domain. Since its amplitude and frequency are constant, it carries virtually no information. The information is in the sidebands. Could broadcasters just suppress the carrier to reduce their electric power costs by 95%? Second, the upper and lower sidebands are mirror images of each other, so they contain the same information. Could they not suppress (filter away) one sideband, making room for twice as many stations on the AM band? The answer to both questions is yes but, in both cases, the simple AM receiver, with its envelope detector, will no longer work properly. Economics favored the simplicity of the traditional AM receiver until it because possible to put all the receiver signal processing on an integrated-circuit chip, where the additional complexity can have negligible cost. In this chapter we examine alternate AM systems that remove the carrier and then at AM systems that reduce the signal bandwidth or double the information carried in the original bandwidth.

8.1 Double-sideband suppressed-carrier AM

Let us look at a system that removes the carrier at the transmitter and regenerates it at the receiver. It is easy enough to modify the transmitter to eliminate the carrier. Review the circuit diagram given previously for an AM transmitter (Figure 6.2a). If we replace the bias battery by a zero-volt battery (a wire), the carrier disappears. The resulting signal is known as *Double-Sideband Suppressed-Carrier* (DSBSC). To restore the missing carrier we might try the receiver circuit shown in Figure 8.1. The locally generated carrier, from a *beat frequency oscillator* (BFO), is simply added back into the IF signal, just ahead

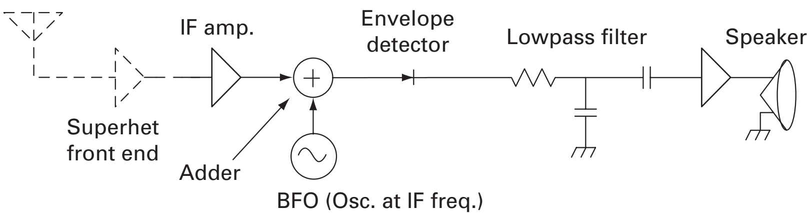

Figure 8.1. Hypothetical DSBSC receiver: additive carrier reinsertion and envelope detection.

of the envelope detector. The energy saved by suppressing the carrier can increase battery lifetime in walkie-talkies by a factor of maybe 20. The modulation schemes used in many cell phones (after the first generation) likewise do away with battery-draining constant carriers.

In a superheterodyne receiver, we only have to generate this carrier at one frequency, the intermediate frequency (IF). As you would expect, the added carrier must have the right frequency. But it must also have the right phase. Suppose, for example, that the modulation is a single audio tone at 400 Hz. If the replacement carrier phase is off by 45°, the output of the envelope detector will be severely distorted. And if the phase is off by 90° the output of the detector will be a tone at 800 Hz (100% distortion!). These example waveforms are shown in Figure 8.2.

Correct carrier reinsertion is fairly easy if the transmitter provides a low-amplitude “pilot” carrier to provide a phase reference. In the model transmitter of Figure 6.2(a) we would go back to using a bias battery, but it would have a low voltage. In principle, the receiver could have a tuned IF amplifier with a narrowband gain peak at the center frequency to bring the pilot carrier up to full level. The most common method, however, is to use a VCO for the BFO in Figure 8.1, together with a feedback circuit to phase lock it to the pilot carrier.

8.2 Single-sideband AM

The requirement for correct BFO phasing is not critical if only one of the two sidebands is transmitted. The sidebands are mirror images of each other so they carry the same information and one will do. We will discuss later three methods to eliminate one of the sidebands from a DSBSC signal, but the first and most obvious method is to use a bandpass filter to select the desired sideband. The resulting signal is known as *Single-Sideband Suppressed-Carrier* (SSBSC) or simply as *Single-Sideband* (SSB). Take the previous example of a single 400-Hz audio tone. The SSB transmitter will put out a single frequency, 400 Hz above the frequency of the (suppressed) carrier if we have selected the upper sideband. This signal will appear in the receiver 400 Hz away from the IF center frequency. Suppose we are still using the BFO and envelope detector. The

Demonstration that Detection of Double Sideband A.M. Requires the Correct Carrier Phase

$t := 0, 0.1 .. 300$

carrier(t) := sin(t)

audio(t) := $0.8 \cdot \sin(0.05 \cdot t)$ Audio is sine wave at 1/20 carrier frequency

dssc(t) := audio(t)·carrier(t) Double sideband suppressed carrier (sidebands alone)

Note that dssc(t) = $1/2(\cos(0.95t) - \cos(1.05)t)$

signal(t) := carrier(t) + dssc(t) Carrier plus sidebands = (1 + audio)(carrier) = normal AM

(a) Sidebands plus correct carrier. (note that the envelope is the sinewave audio)

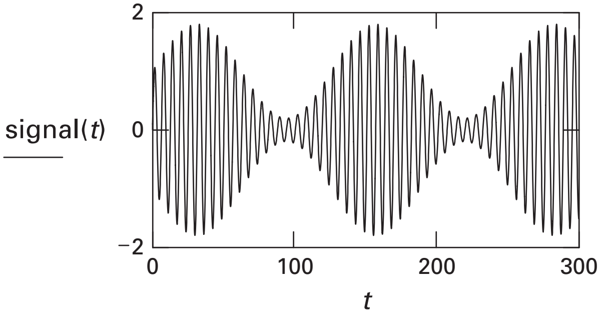

alt(t) := dssc(t) + $\cos(t + \frac{\pi}{4})$

Give the re-injected carrier a 45° phase error.

(b) Sidebands with carrier 45° out of phase

(Note envelope distortion)

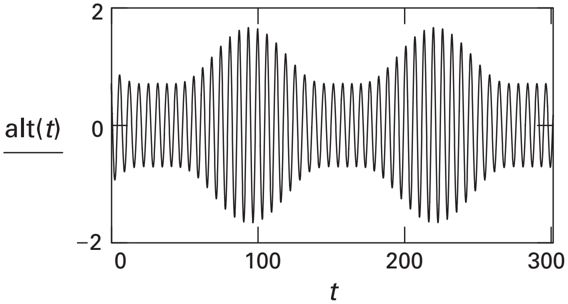

alt2(t) := dssc(t) + cos(t)

Here the re-injected carrier has a 90° phase error.

(c) Sidebands with carrier 90° out of phase

(Note that the envelope has twice the audio frequency – total distortion)

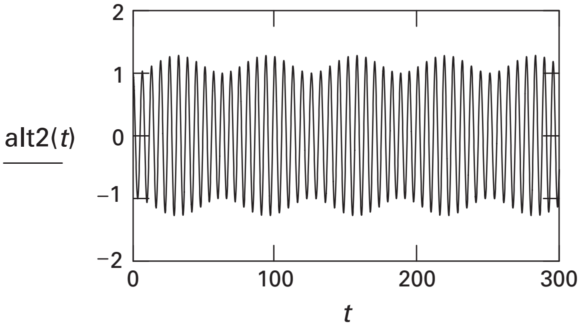

Figure 8.2. Detection of DSSC by adding a local carrier.

envelope of the signal-plus-BFO will indeed be an undistorted sine wave at 400 Hz. So far, so good. But suppose the signal had been two tones, one at 400 Hz and one at 600 Hz. When the BFO carrier is added, the resulting envelope will have tones at 400 and 600 Hz but it will also have a component at 200 Hz (600 − 400) which is distortion. Unwanted signals are produced by the IF components beating with each other; their strength is proportional to the product of the two IF signals. The strength of each wanted component, on the other hand, is proportional to the product of its amplitude times the amplitude of the BFO. If the BFO amplitude is much greater than the IF signal the distortion can be reduced.

8.3 Product detector

The distortion described above can be avoided by using a *product detector* instead of an envelope detector. In the receiver shown in Figure 8.3, the product detector is just the familiar multiplier (a.k.a. mixer).

The output of this detector is, by the distributive law of multiplication, the sum of the products of each IF signal component times the BFO signal. There are no cross-products of the various IF signal components. Product detectors are also called "baseband mixers" or "baseband detectors," because they translate the sideband components all the way down to their original audio frequencies.

While a product detector does not produce cross-products (intermodulation distortion) of the IF signal components, the injected carrier should ideally have the correct phase. The wrong phase is of no consequence in our single-tone example. But when the signal has many components, their relative phases are important. A waveform will become distorted if all its spectral components are given an identical phase shift (see Problem 8.3). It is only when every component is given a phase shift proportional to its frequency that a waveform is not distorted but only delayed in time. Therefore a single-sideband transmitter, like the double-sideband transmitter, should really transmit a pilot carrier and the receiver should lock its BFO phase to this pilot. Some SSB systems do just that. But for voice communications it is common to use no pilot. Speech remains intelligible and almost natural-sounding even when the BFO phase is

Figure 8.3. SSBSC receiver using a product detector.

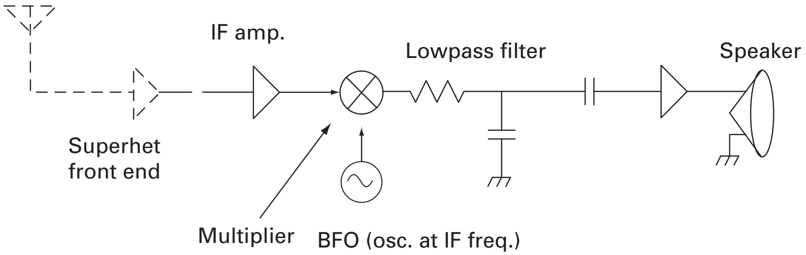

wrong.[1] The BFO can even be slightly off frequency. When the frequency is too high, a demodulated upper sideband (USB) voice signal has a lower than natural pitch. When it is too low, the pitch is higher than natural.

Finally, what happens if we try to use a product detector with a free-running (not phase locked) BFO to receive DSBSC? If the BFO phase is off by 90°, the detector produces no output. If the BFO phase happens to be correct, the output will have maximum amplitude. For an intermediate-phase error the amplitude will be reduced.

Advantages of SSB

Single-sideband, besides not wasting power on a carrier, uses only half the bandwidth, so a given band can hold twice as many channels. Halving the receiver bandwidth also halves the background noise so there is a 3-dB improvement over conventional AM in signal-to-noise ratio. When a spectrum is crowded with conventional AM signals, their carriers produce annoying beat notes in a receiver (a carrier anywhere within the receiver passband appears to be a sideband belonging to the spectrum of the desired signal). With single-sideband transmitters there are no carriers and no beat notes. Radio amateurs, the military, and aircraft flying over oceans use SSB for short-wave (1.8–30 MHz) voice communication.

8.4 Generation of SSB

There are at least three well-known methods to generate single-sideband. In the filter method of Figure 8.4, a sharp bandpass filter removes the unwanted sideband. This filter usually uses crystal or other high-Q mechanical resonators and has many sections. It is never tunable, so the SSB signal is generated at a single frequency (a transmitter IF frequency) and then mixed up or down to the desired frequency. The filter, in as much as it does not have infinitely steep skirts, will cut off some of the low-frequency end of the voice channel.

The phasing method uses two multipliers and two phase-shift networks to implement the trigonometric identity:

$$\cos(\omega_{\mathrm{c}}t)\cos(\omega_{\mathrm{a}}t) \pm \sin(\omega_{\mathrm{c}}t)\sin(\omega_{\mathrm{a}}t) = \cos([\omega_{\mathrm{c}} \mp \omega_{\mathrm{a}}]t). \qquad (8.1)$$

The subscripts “a” and “c” denote the audio and (suppressed) carrier frequencies. If the audio signal is $\cos(\omega_{\mathrm{a}}t)$, we have to generate the $\sin(\omega_{\mathrm{a}}t)$ needed for the second term. A 90° audio phase shift network can be built with passive components or digital signal processing techniques. The audio phase-shift network is usually implemented with two networks, one ahead of each mixer. Their phase *difference* is close to 90° throughout the audio band. The second term also

---

[1] Human perception of speech and even music seems to be remarkably independent of phase. When an adjustable allpass filter is used to produce arbitrary phase vs. frequency characteristics (while not affecting the amplitude), test subjects find it difficult or impossible to distinguish phase-modified audio from the original source.

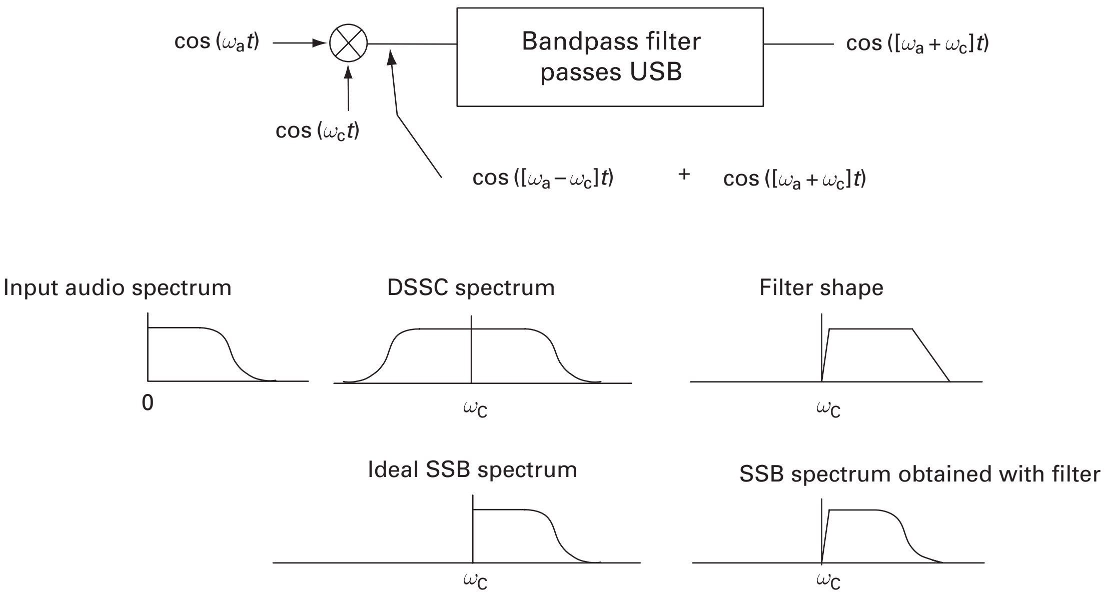
cos ($\omega_a t$)
cos ($\omega_c t$)
Bandpass filter passes USB
cos ([$\omega_a + \omega_c$]t)
cos ([$\omega_a - \omega_c$]t) + cos ([$\omega_a + \omega_c$]t)
Input audio spectrum
0
DSSC spectrum
ω_C
Filter shape
ω_C
Ideal SSB spectrum
ω_C
SSB spectrum obtained with filter
ω_C

Figure 8.4. SSB generator – filter method.

Figure 8.5. SSB Generator – phasing method.

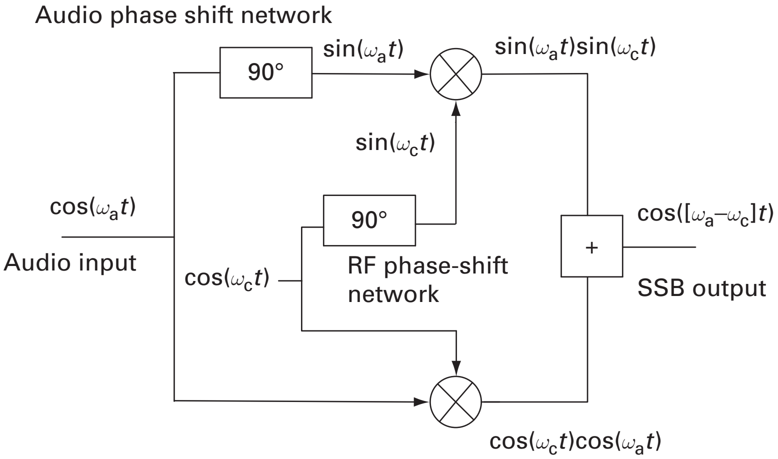
Audio phase shift network
90°
sin($\omega_a t$)
sin($\omega_a t$)sin($\omega_c t$)
sin($\omega_c t$)
cos($\omega_a t$)
Audio input
90°
cos($\omega_c t$)
RF phase-shift network
+
cos([ω_a–ω_c]t)
SSB output
cos($\omega_c t$)cos($\omega_a t$)

requires that we supply sin($\omega_c t$) but, since ω_c is fixed, anything that provides a 90° shift at this one frequency will suffice. Figure 8.5 illustrates the phasing method used to generate an upper sideband (USB) signal.

If the adder is changed to a subtractor (by inverting the polarity of one input), the output will be the upper rather than the lower sideband.

A third method [2] needs neither a sharp bandpass filter nor phase-shift networks. Weaver's method, shown in Figure 8.6, uses four multipliers and two lowpass filters. The trick is to mix the audio signal with a first oscillator whose frequency, ω_0, is in the center of the audio band. The outputs of the first

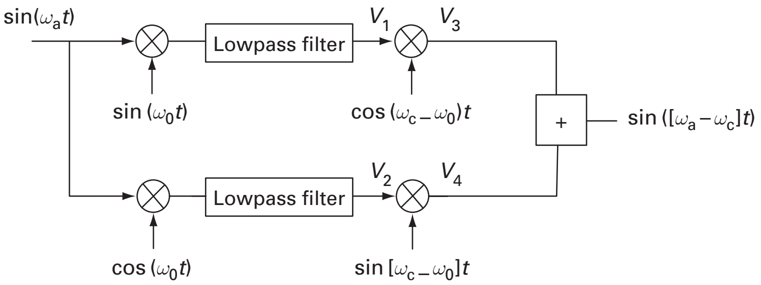

Figure 8.6. SSB generator – Weaver method.

set of mixers have the desired 90° phase difference. The second set of mixers then works the same way as the two mixers in the phasing method.

Referring to Figure 8.6,

$$V_1(t) = \text{lowpass }[\sin(\omega_a t)\sin(\omega_0 t)] = {}^1\!/_2 \cos([\omega_a - \omega_0]t) \tag{8.2}$$

$$V_2(t) = \text{lowpass }[\sin(\omega_a t)\cos(\omega_0 t)] = {}^1\!/_2 \sin([\omega_a - \omega_0]t) \tag{8.3}$$

$$\begin{aligned} V_3(t) &= {}^1\!/_2 \cos[(\omega_c - \omega_0)t]\cos[(\omega_a - \omega_0)t] \\ &= {}^1\!/_4 \cos([\omega_a - \omega_c]t) + {}^1\!/_4 \cos([\omega_a + \omega_c - 2\omega_0]t) \end{aligned} \tag{8.4}$$

$$\begin{aligned} V_4(t) &= {}^1\!/_2 \sin[(\omega_c - \omega_0)t]\sin[(\omega_a - \omega_0)t] \\ &= {}^1\!/_4 \cos[(\omega_a - \omega_c)t] - {}^1\!/_4 \cos[(\omega_a + \omega_c - 2\omega_0)t] \end{aligned} \tag{8.5}$$

$$V_{\mathrm{ou}}\mathrm{t}(t) = V_3(t) + V_4(t) = {}^1\!/_2 \cos\,[(\omega_c - \omega_a)t]. \tag{8.6}$$

We see from Equation (8.6) that this particular arrangement generates the lower sideband.

8.5 Single-sideband with class C, D, or E amplifiers

The three methods described above for generating a single-sideband signal are done in low-level circuitry. Linear amplifiers then produce the required power for the antenna. There is, however, a way to use a class-C or class-D amplifier with simultaneous phase modulation and amplitude modulation to produce SSB. This follows from the fact that any narrowband signal (CW, AM, FM, PM, SSB, DSB, narrowband noise, etc.) is essentially a sinusoid at the center frequency, ω_0, whose phase and amplitude vary on a time-scale longer than $(\omega_0)^{-1}$. Suppose we have generated SSB at low level. We can envelope-detect it, amplify it, and use the amplified envelope to AM-modulate the class-C or class-D amplifier. At the same time we can phase-modulate the amplifier with the phase determined from the low-level SSB signal. The low-level signal is simply amplitude-limited and then used to drive the modulated amplifier. The point is that the high-power SSB signal is produced with an amplifier that has close to 100% efficiency. Moreover, an existing AM transmitter can be converted to single-sideband by making only minor modifications.

8.6 Quadrature AM (QAM)

We noted above that a product detector used for DSBSC will produce no output if the BFO phase is wrong by 90°. It follows that two independent DSBSC signals can be transmitted over the same channel. At the receiver, the phase of the BFO selects one or the other. This is known as quadrature AM (QAM). Note that the carrier can be suppressed, but retaining at least a low-power pilot carrier provides a phase reference for the receiver. Figure 8.7 shows the transmitter (a) and receiver (b) for a QAM system. Here, the receiver uses a narrowband filter to isolate the pilot carrier and then amplifies it to obtain a reconstituted BFO or LO, depending on whether the receiver is a superheterodyne or uses direct conversion to baseband. In practice, carrier regeneration is almost always done with a phase lock loop.

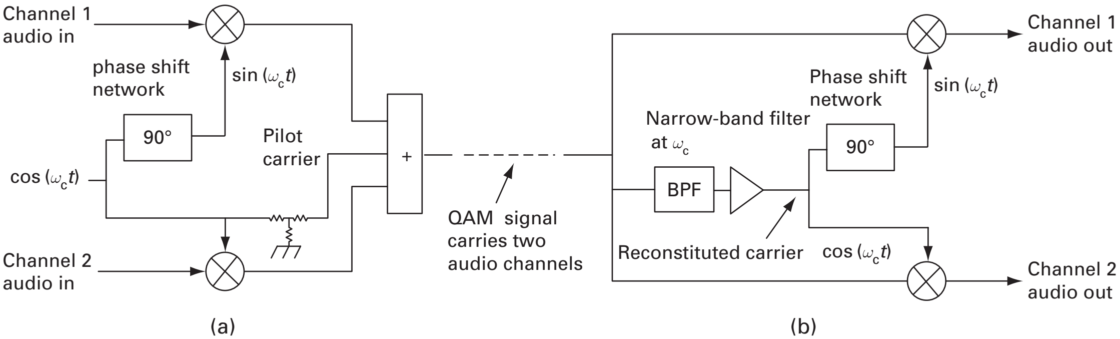
Channel 1 audio in
phase shift network
sin ($\omega_c t$)
90°
Pilot carrier
cos ($\omega_c t$)
Channel 2 audio in
+
QAM signal carries two audio channels
(a)
Channel 1 audio out
Phase shift network
sin ($\omega_c t$)
Narrow-band filter at ω_c
90°
BPF
Reconstituted carrier
cos ($\omega_c t$)
Channel 2 audio out
(b)

Figure 8.7. QAM:
(a) transmitter; (b) receiver.

The U.S. NTSC and European PAL standards for color television use QAM to transmit a pair of video color difference signals. In the NTSC system, these so-called *I* and *Q* signals (In-phase and Quadrature) are multiplied by sine and cosine versions of a 3.579-MHZ oscillator signal and the resulting DSBSC signals are added to the normal video ("luminance") signal. Other uses of QAM include delivery of digital TV signals to set-top converters in some cable television systems and data distribution in some LANs.

The QAM system of Figure 8.7 could also be used to broadcast AM stereo sound, with both the L + R (left plus right) and L − R signals fitting into the channel used traditionally for L + R alone. However, such a QAM stereo signal would not be compatible with the envelope detector in standard AM receivers. Several compatible analog systems were developed and used for AM stereo, but at this writing, both AM and FM stations are adopting compatible *hybrid digital* (HD) systems in which the digital signals occupy spectral space on either side of the conventional analog signal. This system, which uses *COFDM* modulation, is discussed in Chapter 22.

Problems

Problem 8.1. Demonstrate for yourself the kind of phase distortion that will occur when the BFO in a product detector does not have the same phase as the suppressed carrier. Use the Fourier decomposition of a square wave: $V(t) = \sin(\omega t) + 1/3 \sin(3\omega t) + 1/5 \sin(5\omega t) + \ldots$. Have your computer plot $V(\theta) = \sin(\theta) + 1/3 \sin(3\theta) + 1/5 \sin(5\theta) + \cdots + 1/9 \sin(9\theta)$ for θ from 0 to 2π. Then plot $V'(\theta) = \sin(\theta+1) + 1/3 \sin(3\theta+1) + 1/5 \sin(5\theta+1) + \cdots + 1/9 \sin(9\theta+1)$, i.e., the same function but with every Fourier component given an equal phase shift (here 1 radian). Would you expect these waveforms to sound the same? Consider the case in which the phase shift is 180°. This just inverts the waveform. Is this a special case or is an inverted audio waveform actually distorted?

Problem 8.2. The phase-shift networks used in the phasing method of single-sideband generation have a flat amplitude vs. frequency response – they are known as *allpass* filters. Allpass filters have equal numbers of poles (all in the left-hand plane) and zeros (all in the right-hand plane). Find the amplitude and phase response of the network shown below, which is a first-order allpass filter. (Note that the inverting op-amp is used only to provide an inverted version of the input.)

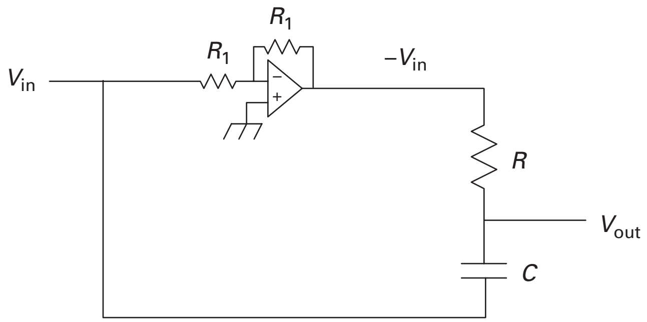

Problem 8.3. The most general allpass filter can be obtained by cascading first-order allpass sections (Problem 8.2) with second-order allpass sections. Find the amplitude and phase response of the network shown below, a second-order allpass filter.

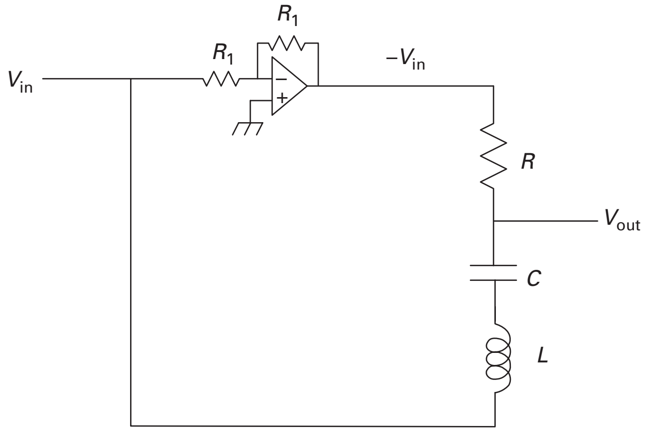

References

[1] Sabin, W. E., and Schoenike, E. O., Editors, *Single-Sideband Systems and Circuits*, New York: McGraw- Hill, 1987.

[2] Weaver, D. K. Jr., A third method of generation and detection of single-sideband signals, *Proceedings of the IRE*, pp. 1703–1706, December 1956.

CHAPTER

9

Class-C, D, and E power RF amplifiers

Class-C, D, and E RF power amplifiers are all about high efficiency. They are used in large transmitters and industrial induction heaters, where high efficiency reduces the power bill and saves on cooling equipment, and also in the smallest transmitters, such as cell phones, where high efficiency increases battery life. These amplifiers are so nonlinear (the output signal amplitude is not proportional to the input signal amplitude), they might better be called synchronized sine wave generators. They consist of a power supply, at least one switching element (a transistor or vacuum tube), and an *LCR* circuit. The "*R*" is the load, R_{L}, often the radiation resistance of an antenna, equivalent to a resistor. The *LC* network is resonant at the operating frequency. The output sine-wave amplitude, while not a linear function of the input signal amplitude, *is* proportional to the power supply voltage. Thus, these amplifiers can be amplitude modulated by varying the supply voltage. Of course they can also be frequency modulated by varying the drive frequency (within a restricted bandwidth, determined by the *Q* of the *LC* circuit). Finally, they can be used as frequency multipliers by driving them at a subharmonic of the operating frequency.

9.1 The class-C amplifier

Figure 9.1 shows a class-C amplifier (a), together with an equivalent circuit (b). The circuit looks no different from the class-B amplifier of Figure 3.15 or a small-signal class-A amplifier. But here the active device (transistor or tube) is used not as a continuously variable resistor, but as a switch. To simplify the analysis, we consider the switch to have a constant on-resistance, *r*, and infinite off-resistance. This model is a fairly good representation of a power FET, when used as a switch.

The switch is closed for less than half the RF cycle, during which time the power supply essentially "tops off" the capacitor, restoring energy that the load has sapped from the resonant circuit during the cycle. The switch has internal resistance (i.e., loss), so in recharging the capacitor, some energy is lost in the

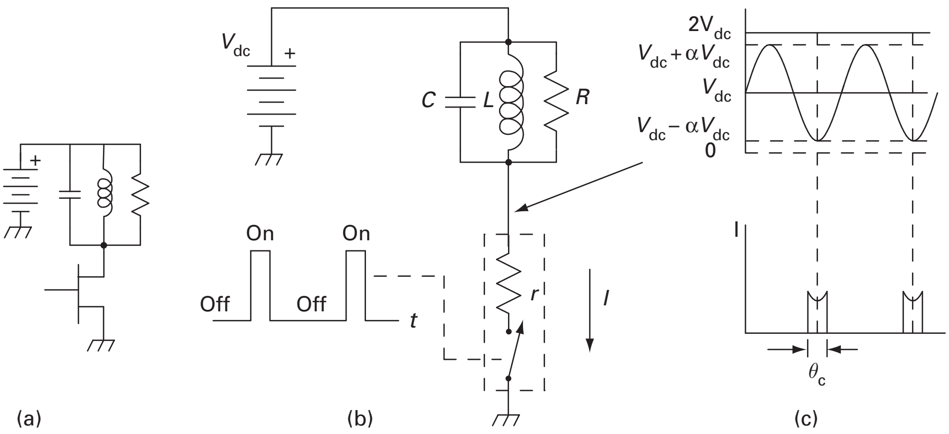

Figure 9.1. Class-C amplifier operation.

switch. Normally the LC circuit has a high Q (at least 5) so its flywheel action minimizes distortion of the sine wave caused by the abrupt pull-down of the switch and by the damping caused by the load. The drive is shown as a rectangular pulse but is often a sine wave, biased so that conduction takes place just around the positive tips. Class-C amplifiers are normally run in saturation, meaning that the switch, when on, always has its lowest possible resistance, ideally much less than the load resistance.

9.1.1 Simplified analysis of class-C operation

The simplified analysis, which we will also use later to analyze the class-E amplifier, is based on the assumption that the LC circuit provides enough flywheel effect to maintain a perfect sine wave throughout the cycle, including the interval when the resistive switch is closed. With this assumption, let us analyze the circuit of Figure 9.1 to find the output voltage and, thereby, the power and the efficiency. Referring to the figure, θ_c is the conduction angle, V_{dc} is the supply voltage, r is the on-resistance of the switch, and αV_{dc} is the peak voltage of the sine wave. We can find α as follows. The input power (the power supplied by the battery) must be equal to the sum of the output power (the power dissipated in R) plus the power dissipated in the switch resistance r. These terms are given respectively by the average of the battery voltage times the current, $(\alpha V_{dc})^2/(2R)$, and the average of I^2/r. The power equation becomes

$$\frac{1}{2\pi}\int_{\theta_c/2}^{\theta_c/2} V_{dc}\left(\frac{V_{dc}-\alpha V_{dc}\cos\theta}{r}\right)\mathrm{d}\theta = \frac{(\alpha V_{dc})^2}{2R} + \frac{1}{2\pi}\int_{-\theta_c/2}^{\theta_c/2}\left(\frac{V_{dc}-\alpha V_{dc}\cos\theta}{r}\right)^2 r\mathrm{d}\theta. \tag{9.1}$$

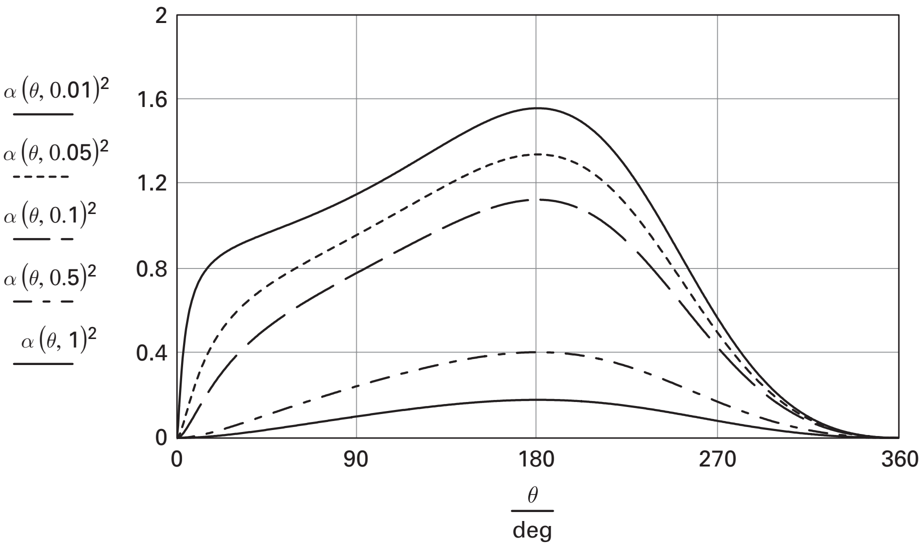

Figure 9. 2. Class-C output power ($\alpha 2$) vs. conduction angle for five values of r/R. The efficiency is given by the output power divided by the power supplied by the battery.

Rearranging this equation, we get

$$\frac{1}{\pi r/R}\int_{0}^{\theta_c/2}(1-\alpha\cos\theta)(\alpha\cos\theta)\mathrm{d}\theta-\frac{\alpha^2}{2}=0. \tag{9.2}$$

Carrying out the integral in Equation (9.2) and solving for α, we find

$$\alpha(\theta_c, r/R)=\frac{2\sin(\theta_c/2)}{\theta_c/2+\sin(\theta_c)/2+\pi r/R}. \tag{9.3}$$

The quantity α^2, proportional to the output power, is plotted in Figure 9.2 for five values of r/R, the ratio of the switch resistance to the load resistance.

The middle value, $r/R = 0.1$, is typical in actual practice. Maximum power is produced for a conduction angle of 180°, i.e., if the switch is closed during the entire negative voltage loop. If the conduction angle exceeds 180°, the incursions into the positive loop extract energy from the tuned circuit and the power is reduced. Note that α^2, and therefore α, can be greater than unity, especially when the conduction angle is 180°. We will see below, however, that much higher efficiency is obtained for conduction angles substantially less than 180°. The efficiency is given by the output power divided by the power supplied by the battery:

$$\eta(\theta_c, r/R)=\frac{(\alpha V_{\mathrm{dc}})^2/2R}{\frac{1}{2\pi}\int_{-\theta_c/2}^{\theta_c/2} V_{\mathrm{dc}}(V_{\mathrm{dc}}-\alpha V_{\mathrm{dc}}\cos\theta'/r)\mathrm{d}\theta'}. \tag{9.4}$$

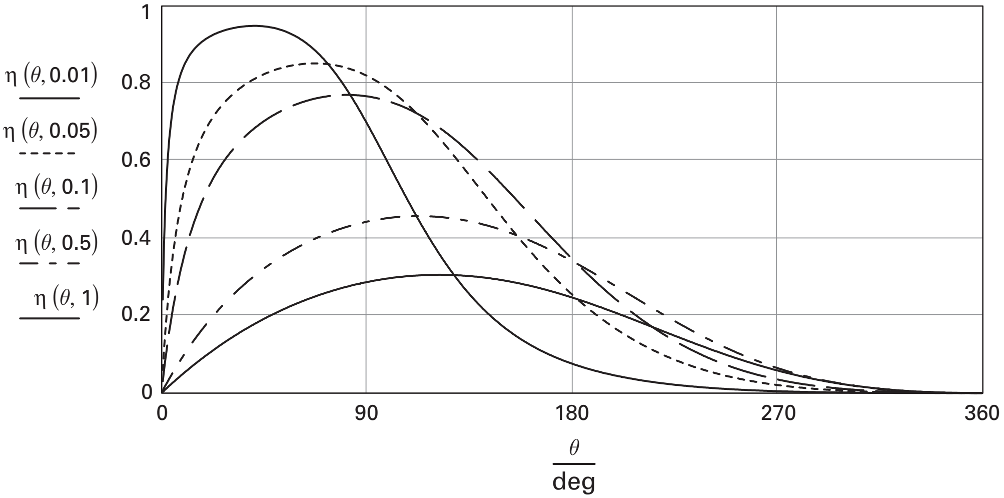

Figure 9.3. Class-C efficiency vs. conduction angle for five values of *r*/*R*.

Evaluating the integral in Equation (9.4) we find

$$\eta(\theta_c, r/R) = \frac{(\pi r/R)\alpha^2}{\theta_c - 2\alpha \sin(\theta_c/2)}, \tag{9.5}$$

where α is given by Equation (9.3). This expression for efficiency is plotted in Figure 9.3 for the same five values of r/R. For $r/R = 0.1$, the efficiency is a maximum at about 90°.

Note that this amplifier model, assuming a constant resistance in the switch, can be solved exactly by finding the general solution for the transient waveform during the switch-off period, as well as the transient waveform during the switch-on period. Because the differential equations are of second order, these general solutions will each have two adjustable parameters. The parameters are found by imposing the boundary conditions that, across the switch openings and closings, the voltage on the capacitor is continuous and the current through the inductor is continuous.

9.1.2 General analysis of a class-C operation with a nonideal tube or transistor

The above analysis will not give accurate results for class-C amplifiers made with tubes or bipolar transistors, because these devices do not have the simple constant on-resistance characteristic of a FET. Nevertheless, their nonlinear characteristics are specified graphically on data sheets, and accurate class-C analyses can be done numerically. The method is basically the same as the simplified analysis; one assumes the resonant *LC* circuits at the input and output have enough *Q* to force the input and output waveforms to be sinusoidal. For this analysis, the device characteristics are plotted in a "constant current" format: in the case of a tube, curves of constant plate current are plotted on a graph whose axes are plate voltage and grid voltage. Sinusoidal plate and grid voltages are assumed and a numerical integration of plate current × plate voltage, averaged through one complete cycle, gives the power dissipated in the device. The power supply voltage times the *average* current gives the total input power. The difference is the power delivered to the load. The designer selects a device and a power supply voltage and assumes trial waveforms with different bias points and sine-wave amplitudes. Usually several trial designs are needed to maximize output power with the given device or to maximize efficiency for a specified output power. A class-C amplifier can approach 100% efficiency, but only in the limit that the output power goes to zero. We will see below that class-D and E amplifiers can approach 100% efficiency and still produce considerable power.

9.1.3 Drive considerations

Class-C amplifiers using vacuum tubes[1] nearly always drive the control grid positive when the tube is conducting. The grid draws current and dissipates power. Data sheets include grid current curves so that the designer can use the procedure outlined above to verify that the chosen operating cycle stays within both the maximum plate dissipation rating and the maximum grid dissipation rating. When tetrodes are used, a third analysis must be done to calculate the screen grid dissipation.

9.1.4 Shunt-fed class-C amplifier

In Figure 9.1, since the switch and the *LCR* tank circuit are in series, they can be interchanged, allowing the bottom of the tank to be at ground potential, which is often a convenience. But the usual way to put the tank at dc ground is to use the shunt-fed configuration shown in Figure 9.4(c).

In the shunt-fed circuit, an RF choke connects the dc supply to the transistor and a blocking capacitor keeps dc voltage off the tank circuit. The RF choke has

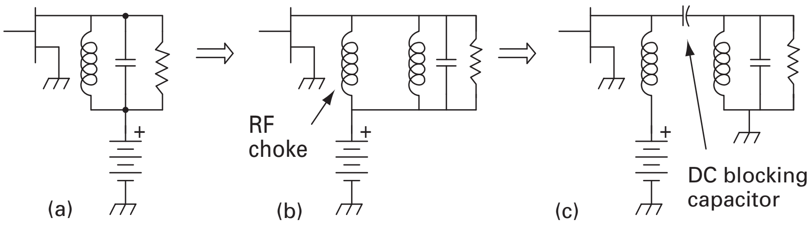

Figure 9.4. Equivalence of series-fed and shunt-fed circuits.

---

[1] A triode vacuum tube is analogous to an NPN transistor. The tube's plate, control grid, and cathode correspond respectively to the transistor's collector, base, and emitter. Tetrode tubes have an additional grid, the screen grid, between the control grid and the plate. The screen grid is usually run at a fixed bias voltage and forms an electrostatic shield between the control grid and the plate [5].

a large inductance, so the current through it is essentially constant. The switch pulls pulses of charge from the blocking capacitor. This charge is replenished by the current through the choke. Figure 9.4 shows, from right to left, the equivalence of the shunt-fed and series-fed circuits. Note that the series-fed and shunt-fed equivalence applies as well to amplifiers of class-A, B, or C and for large-signal or small-signal operation.

9.1.5 The class-C amplifier as a voltage multiplier

An important property of the saturated class-C amplifier is that the peak voltage of the output RF sine wave is directly proportional to the supply voltage ($V_{\text{pk}} = \alpha V_{\text{dc}}$). In the standard saturated operation, the proportionality constant, α, is about 0.9. The class-C amplifier is therefore equivalent to a voltage multiplier which forms the product of a nearly unit-amplitude sine wave times the power supply voltage. Modulating (varying) the power supply voltage of a class-C amplifier is the classic method used in AM transmitters. (Note that this useful property of a class-C (or D or E) amplifier would be considered a defect for an op-amp circuit – poor power supply rejection.)

9.1.6 The class-C amplifier as a frequency multiplier

If the class-C amplifier drive circuit furnishes a turn-on pulse only every other cycle, you can see that the decaying oscillation of the tank circuit voltage will execute two cycles between refreshes. Of course there will be a greater voltage droop, producing less output power, but note that the circuit becomes a frequency doubler. If the circuit is pulsed only on every third cycle, it becomes a tripler, etc. It is common to use a cascade of frequency multipliers to produce a high RF frequency that is a multiple of the frequency of a stable low-frequency oscillator. Class-D and class-E amplifiers can also be used as frequency multipliers.

9.2 The class-D RF amplifier

Class-D amplifiers can, in principle, achieve 100% efficiency. At least two switches are required, but neither is forced to support simultaneous voltage and current. The *class-D series resonant amplifier* is shown in Figure 9.5(a).

A single-pole double-throw switch produces a square-wave voltage. The series LC filter lets the fundamental sine-wave component reach the load, R_{L}. The bottom of the switch could be connected to a negative supply but ground will work since the resonating capacitor provides ac coupling to the load resistor. A real circuit is shown in Figure 9.5(b); two transistors form the switch. This is a push–pull circuit, i.e., the transistors are driven out-of-phase so that when one is on the other is off. Let us find the voltage on the load. Since the capacitor also acts as a dc block, we can consider this square wave to be

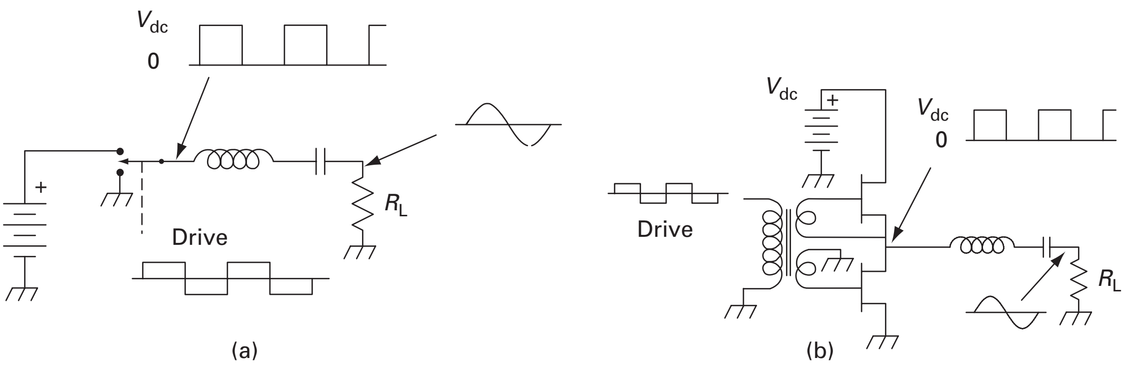

Figure 9.5. Class-D series resonant amplifier.

symmetric about zero, swinging from $-V_{dc}/2$ to $+V_{dc}/2$. The square wave is equivalent to a Fourier series, i.e., a sum of sine waves whose frequencies are the fundamental, and odd multiples of the fundamental. The series resonant *LC* filter passes the fundamental sinusoidal component of the square wave. We can find the amplitude of the output sine wave by equating the dc input power to the sine-wave output power. They must be equal, since the circuit has no lossy element other than R_L, the load. Since both the voltage and current change sign during the negative half of the cycle, the power is the same in the negative half-cycle as in the positive half-cycle. The power delivered by the dc supply is therefore the product of the square-wave voltage, $V_{dc}/2$, times the average of the positive current in the series *LCR*. The current can be written as $I(t) = I_{pk}\sin(\omega t)$ and the average positive current is $\langle|I|\rangle = (2/\pi)\ I_{pk} = (2/\pi)\ V_{pk}/R_L$, where V_{pk} is the peak value of the sine-wave voltage on the load. Equating the power from the supply to the sine-wave power on the load, we have $(V_{dc}/2)\cdot(2/\pi)(V_{pk}/R_L) = 2V_{dc}^2/(R_L)$. Solving for V_{pk}, we find $V_{pk} = 2V_{dc}/\pi$.

To estimate the loss, we will assume the FETs have constant on-resistance, *r*. The current through the load passes through one or the other of the switches, so the ratio of output power to switch power is $(I^2R_L)/(I^2\,r)$ and the efficiency is $\eta = R_L/(R_L{+}r)$. You can see that, inas much as $r \ll R_L$, the efficiency can approach 100%. However, there is another, more important, source of loss in this class-D amplifier. Each switching transistor has parasitic capacitance which is abruptly charged and discharged through the transistor's resistance once per cycle. The energy lost is $\frac{1}{2}CV^2$ per transition so, for the circuit of Figure 9.5, the switching losses would be $4 \times \frac{1}{2}CV_{dc}^2 \times f$. Suppose the frequency, f, is 10 MHz and the FET switches each have, say, 200 pF of parallel capacitance. With a supply voltage of 200 volts, this would produce a loss of 160 watts! This loss can be avoided by using an alternate topology, the parallel resonant class-D amplifier.

9.2.1 Parallel-resonant class-D RF amplifiers

The class-D amplifier of Figure 9.6 consists of a square-wave *current source* driving a parallel *LCR* circuit. In this circuit, a large inductor (RF choke)

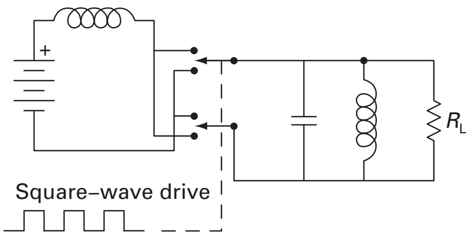

Figure 9.6. Class-D parallel-resonant RF amplifier operation.

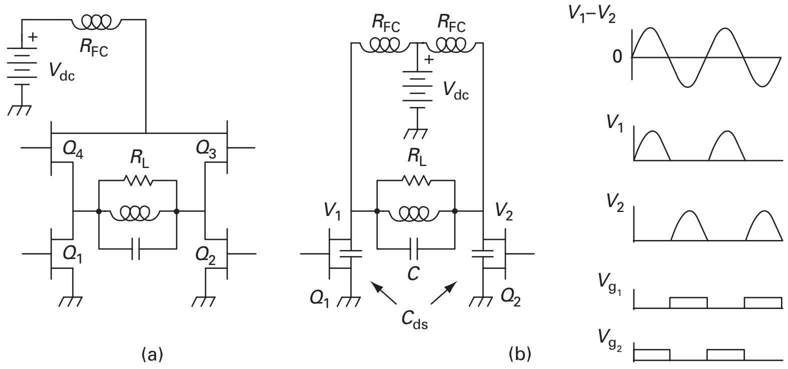

Figure 9.7. Practical class-D parallel-resonant RF amplifiers.

provides the constant current. A DPDT switch commutates the load, effectively forming a square-wave current source.

Two practical versions of this circuit are shown in Figure 9.7. In (a) the peak sine-wave voltage on the load, R_L, will be $\pi V_{\mathrm{dc}}/2$ and the dc supply current will be $\pi^2 V_{\mathrm{dc}}/(8R_{\mathrm{L}})$. For the appealing circuit in (b), these quantities are πV_{dc} and $\pi^2 V_{\mathrm{dc}}/(2R_{\mathrm{L}})$. (See problem 9.4.) But the beauty of the parallel class-D amplifier is that the voltage on the parasitic capacitances of the transistors (drain-to-source capacitance, shown in dotted lines in Figure 9.7b) is zero at the instants the switches open or close. Thus, there is no lossy abrupt charging or discharging of these parasitic capacitors. (Refer to the waveforms in Figure 9.7b. Note that the value of C is effectively increased by C_{ds}, the parasitic capacitance of one transistor – first one, then the other.)

9.3 The class-E amplifier

We have seen that a class-C amplifier can be built with a single transistor and that its efficiency is high, relative to class-B and class-A amplifiers. The class-D switching amplifier, on the other hand can, in principle, achieve 100% efficiency, but requires two transistors. The *class-E amplifier* [4], has both virtues: single transistor operation and up to 100% efficiency. Figure 9.8 shows the circuit.

Figure 9.8. Class-E amplifiers.

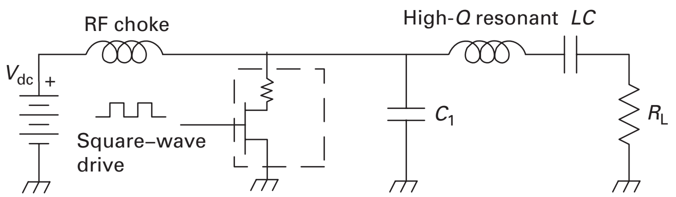

Figure 9.9. Class-E amplifier equivalent circuit.

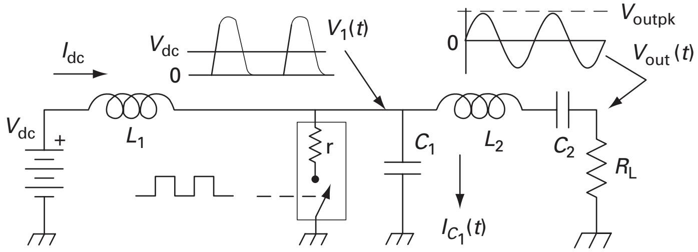

The transistor is used as a switch: fully on or fully off. An equivalent circuit is shown in Figure 9.9, where the transistor is represented as a switch. The switch operates with a 50% duty factor, so the voltage at the switch, $V_1(t)$, is forced to be zero for half the cycle (or almost zero if we consider the switch to have a small series resistance, r).

During the other half-cycle, $V_1(t)$ consists of a rounded positive pulse. The power supply is connected through L_1, an RF choke (a high-value inductor). The choke and the switch form a *flyback* circuit in which the power supply pumps energy into the inductor while the switch is closed and the inductor pumps energy into the rest of the circuit while the switch is open. Since there can be no dc drop across the choke, you can see that the pulses in $V_1(t)$ at the switch must have an average amplitude equal to twice the power supply voltage. The key to the efficiency of this circuit is that it can be designed so that the voltage $V_1(t)$ has fallen to zero at precisely the instant the switch closes, i.e., the capacitor C_1 is shorted out without a sudden lossy discharge. Note that C_1 includes the transistor's parasitic capacitance. The RF choke, L_1, is large enough to ensure that the dc supply current, I_{dc}, has essentially no ac component, i.e., the inductor current's increase and decrease during the flyback cycle are much smaller than the average current.

A simplified description of the circuit operation is as follows. The pulses at C_1 are not square, but must have an average voltage of $2V_{dc}$. The waveform is a rough approximation to a sine wave with a peak voltage of V_{dc} plus an equal dc offset. The bandpass filter formed by L_2 and C_2 passes a good sine wave to the load, R_L. Let us now look in detail at the circuit operation and design.

9.3.1 Class-E amplifier design procedure

We have seen that we can expect this amplifier to furnish the load with a sine wave whose amplitude is something like V_{dc}, so the RF power output will be approximately ${V_{\mathrm{dc}}}^2/(2R_{\mathrm{L}})$. If a different power is desired, the resistance of the load, R_{L}, can be transformed to a different value using any of the techniques of Chapter 2. Next, we can simply pick a value for L_2 that will give a reasonably high Q, maybe 10, so that the waveform at the load will be a good sine wave, i.e., have minimal harmonic content. We are left with finding values for C_1 and C_2. So far, the only constraint on the circuit is the condition that $V_1(t)$ have fallen to zero at $\tau/2$, the instant of switch closure. It turns out that this can be satisfied over a range of combinations of C_1 and C_2 and that this provides a way to set the output power. Combinations with lower values of C_1 reduce the output power. However, it is possible and beneficial to impose a second constraint which will require a unique combination of C_1 and C_2. This constraint is that $\mathrm{d}V_1/\mathrm{d}t$, as well as $V_1(t)$, be zero at the instant the switch is closed. If this condition is met, the frequency of the amplifier can be shifted somewhat without seriously violating the condition that $V_1(\tau/2)$ be zero. Moreover, at high frequencies, where the transistor's switching time is relevant, $V_1(t)$ will at least remain close to zero during the switching process. Before outlining the analysis, we summarize the resulting design procedures as follows:

1. Pick a value for Q, say $Q = 10$. Then $\omega L_2 = QR_{\mathrm{L}}$.

$$\text{Pick } \omega \mathrm{L}_1 \text{ to be, say, } 100R_{\mathrm{L}}. \tag{9.6}$$

2. Calculate C_1 as follows:

$$\frac{1}{\omega C_1} = \frac{\pi}{2}\left(1 + \frac{\pi^2}{4}\right)R_{\mathrm{L}} = 5.45R_{\mathrm{L}}. \tag{9.7}$$

3. Calculate C_2 as follows:

$$\frac{1}{\omega C_2} = \omega L_2 - \frac{\left(\frac{\pi^2}{8} - \frac{1}{2}\right)}{\left(1 + \frac{\pi^2}{4}\right)}\frac{1}{\omega C_1} = \omega L_2 - 0.212\frac{1}{\omega C_1}. \tag{9.8}$$

4. The amplitude of the output sine wave (the voltage on R_{L}) is

$$V_{\mathrm{out_{pk}}} = V_{\mathrm{dc}}\frac{2}{\sqrt{1 + \frac{\pi^2}{4}}} = 1.07V_{\mathrm{dc}}. \tag{9.9}$$

5. The output power is given by

$$P_{\mathrm{out}} = \frac{(V_{\mathrm{out_{pk}}})^2}{2R_{\mathrm{L}}} = 1.154\frac{V_{\mathrm{dc}}^2}{2R_{\mathrm{L}}}. \tag{9.10}$$

6. The average current drawn from the power supply is

$$I_{\mathrm{dc}} = \frac{V_{\mathrm{out_{pk}}}}{2R_{\mathrm{L}}} = 0.577\frac{V_{\mathrm{dc}}}{R_{\mathrm{L}}}. \tag{9.11}$$

9.3.2 Class-E circuit analysis

This section is included for the reader who wants to understand the derivation of the above design formulas. We will analyze the circuit assuming $r = 0$ and afterward estimate the loss when r is small compared with R_{L}. The switch will be open whenever $\sin(\omega t)$ is positive, i.e., it opens at $t = 0$, closes at $\omega t = \pi$, etc. Because of the high Q of the series combination, L_2C_2, we will again use the simplifying assumption that the load voltage is a perfect sine wave, $V_{\mathrm{RL}} = V_{\mathrm{out_{pk}}}(t)\sin(\omega t+\phi)$ where ϕ is a phase shift to be determined [3]. As with the class-C amplifier, an exact analysis requires finding the properly connected switch-open and switch-closed transient solutions. Assuming the sine wave, the current into the load is given by $I_{\mathrm{RL}}(t) = (V_{\mathrm{out_{pk}}}/R_{\mathrm{L}})\sin(\omega t + \phi)$.

By inspection of Figure 9.9, we can immediately write an equation for the current through the capacitor C_1 while the switch is open:

$$I_{C_1}(t) = C_1\frac{\mathrm{d}V_1}{\mathrm{d}t} = I_{\mathrm{dc}} - \frac{V_{\mathrm{out_{pk}}}}{R_{\mathrm{L}}}\sin(\omega t + \phi). \tag{9.12}$$

Imposing the condition that I_{C_1} be zero at $\omega t = \pi$, we find that

$$\sin(\phi) = \frac{-I_{\mathrm{dc}}R_{\mathrm{L}}}{V_{\mathrm{out_{pk}}}}. \tag{9.13}$$

We can integrate Equation (9.12) to get $V_1(t)$, the voltage on the capacitor while the switch is open:

$$V_1(t) = \frac{1}{C_1}\left(I_{\mathrm{dc}}t + \frac{V_{\mathrm{out_{pk}}}}{\omega R_{\mathrm{L}}}\cos(\omega t + \phi) - \frac{V_{\mathrm{out_{pk}}}}{\omega R_{\mathrm{L}}}\cos(\phi)\right), \tag{9.14}$$

where the last term is a constant of integration, added to satisfy the condition $V_1(0) = 0$, imposed by the switch having been in its closed state before $t = 0$. Next, imposing the condition that V_1 be zero also at $\omega t = \pi$, Equation (9.14) gives us

$$\cos(\phi) = \frac{\pi}{2}\frac{I_{\mathrm{dc}}R_{\mathrm{L}}}{V_{\mathrm{out_{pk}}}}. \tag{9.15}$$

We can combine Equations (9.13) and (9.15), using $\cos^2\theta + \sin^2\theta = 1$, to find

$$\frac{V_{\mathrm{out_{pk}}}}{I_{\mathrm{dc}}R_{\mathrm{L}}} = a \quad \text{and} \quad \sin(\phi) = \frac{-1}{a}, \quad \text{where} \quad a = \sqrt{\frac{\pi^2}{4} + 1}. \tag{9.16}$$

From this point, it remains to express $V_1(t)$ and the load current as complex constants times $\mathrm{e}^{\mathrm{j}\omega t}$ (the actual voltage and current are, of course, the real parts) and calculate the ratio, which must be equal to the complex impedance looking into the $L_2C_2R_L$ series circuit, $Z = R_L + j(\omega L_2 - 1/(\omega C_2))$.

The load current, $(V_{pk}/R_L)\sin(\omega t+\phi) = a\ I_{dc}\ \sin(\omega t+\phi)$, corresponds to the complex current $I(t) = -j\ a\ I_{dc}\ e^{j(\omega t+\phi)}$, since $\mathrm{Re}(-j\ a\ I_{dc}\ e^{j(\omega t+\phi)}) = a\ I_{dc}\ \sin(\omega t+\phi)$. Since the voltage $V_1(t)$ is not a sine wave, we have to find the complex representation of its fundamental Fourier component by evaluating the integral

$$\text{complex } V(t) = e^{j\omega t}\frac{2}{\tau}\int_0^{\pi/\omega} V(t)e^{-j\omega t}dt. \tag{9.17}$$

Note that the upper limit of the integral would normally be $2\pi/4$ but by integrating over only the first half-cycle, we account for the fact that the closed switch forces $V_1(t)$ to be zero during the second half-cycle. We can rewrite Equation (9.14) in terms of a and I_{dc}:

$$V_1(t) = \frac{I_{dc}}{C_1}\left(t + \frac{a}{\omega}\cos(\omega t + \phi) - \frac{a}{\omega}\cos(\phi)\right). \tag{9.18}$$

Evaluating the integral in Equation (9.17), and then setting the ratio *complex* $V_1(t)$ / *complex* $I(t)$ equal to the impedance, $R_L + j(\omega L - 1/(\omega C_2))$, results in the formulas for C_1 and C_2 (Equations 9.7 and 9.8).

9.3.3 Efficiency of the class-E amplifier

We will now assume that r is not zero, but still small enough that we can use the currents derived above, where r was taken as zero. Power is dissipated in r when the switch is closed. The current through r is just

$$I_r(t) = I_{dc} - \frac{V_{out_{pk}}}{R_L}\sin(\omega t + \phi), \tag{9.19}$$

which is the same as Equation (9.12), except that now we assume any current flowing in the capacitor is negligible compared with the current flowing through the closed switch.

The instantaneous power dissipated in r is

$$(I_r(t))^2 r = \left(I_{dc} - \frac{V_{out_{pk}}}{R_L}\sin(\omega t + \phi)\right)^2 r \tag{9.20}$$

and the average power dissipation in r is, therefore, given by

$$\begin{aligned}\langle (I_r(t))^2 r\rangle &= \left\langle \left(I_{dc} - \frac{V_{out_{pk}}}{R_L}\sin(\omega t + \phi)\right)^2\right\rangle r \\ &= \frac{1}{\tau}\int_{\tau/2}^{\tau}\left(I_{dc} - \frac{V_{out_{pk}}}{R_L}\sin(\omega t + \phi)\right)^2 r dt,\end{aligned} \tag{9.21}$$

where the period, τ, is just $2\pi/\omega$. Evaluating this integral and using Equations (9.15) and (9.16), we find

$$\langle (I_r(t))^2 r \rangle = \frac{r}{R_\mathrm{L}^2} \left(\frac{3 {V_{\mathrm{out}_\mathrm{pk}}}^4}{8 {V_\mathrm{dc}}^2} + \frac{{V_{\mathrm{out}_\mathrm{pk}}}^2}{4} \right). \tag{9.22}$$

To get the fractional power loss, just divide this by the power out, $V_{\mathrm{out}_\mathrm{pk}}/(2R_\mathrm{L})$.

9.3.4 Class-E design example

The formulas above were used to design a class-E amplifier designed to produce a 1-MHz, 12-V peak sine wave on a 50-ohm load, using a 12-V power supply. The component values were as follows: $L_1 = 1$ mH, $C_1 = 584$ pF, $L_2 = 79.6$ µH, $C_2 = 360$ pF, and $R_\mathrm{L} = 50$ ohms. If the transistor has an on-resistance of 0.5 ohms, the fractional loss will be 1.9% (98% efficiency). A SPICE analysis of this circuit shows the second harmonic power at the load to be 26 dB below the fundamental power. The 3-dB bandwidth is about 15%.

9.4 Which circuit to use: class-C, class-D, or class-E?

Remember that the high-efficiency amplifiers discussed in this chapter are narrowband and nonlinear. They are ideal for producing conventional signals with AM, pulse, or phase modulation, such as radio broadcast signals and signals from cell phones. Lower efficiency linear class-A, AB, or B amplifiers are used with more complicated signals such as the superposition of many individual signals from a cellular base station or from a direct broadcast television satellite.[2] AM and FM broadcast transmitters have historically used class-C vacuum tube amplifiers. This type of transmitter is still widely used. Tubes are available with maximum plate dissipations of up to more than 1 MW. Typical class-C amplifiers have efficiencies of 75–85%, so a single large tube can produce in excess of 6 MW of RF power. The highest power used in radio broadcasting is about 2 MW, but amplifiers (or class-C oscillators) for much higher power are found in industrial heating applications such as curing plywood and welding. Many new designs use class-D and class-E solid-state amplifier modules, with the power from multiple modules being combined to achieve the desired total power. The complexity of using many modules is sometimes offset by the ability to “hot swap” defective modules without interrupting operations. The switching amplifiers, class-D and class-E, benefit from advances in transistor technology and have reached GHz frequencies.

---

[2] In principle, class-C, D, and E amplifiers can reproduce any type of band-limited signal by using simultaneous amplitude and phase modulation, as in the efficient amplification of single-sideband signals discussed in Chapter 8.

Problems

Problem 9.1. For the shunt-fed class-C amplifier of Figure 9.4(c), sketch the waveforms of the voltage on each side of the blocking capacitor and of the current through the blocking capacitor.

Problem 9.2. Suppose the simple series class-D amplifier of Figure 9.5(b) is driving a 50-ohm resistive load at a frequency of 1 MHz. The values of the inductor and capacitor are 9.49 μH and 2.67 nF (equal and opposite reactances at 1 MHz). This resonant circuit passes the 1-MHz component of the square wave and greatly reduces the harmonics. By what factor is the 3-MHz (i.e., third harmonic) power delivered to the load lower than the fundamental (1-MHz) power? Hints: in a square-wave, the amplitude of the third harmonic is one-third of the amplitude of the fundamental. Remember that at 1 MHz, $X_L = X_C$ while at 3 MHz, X_L is increased by a factor of 3 and X_C is decreased by a factor of 3.

Problem 9.3. A single-tube class-C amplifier with 75% efficiency is providing 500 kW of continuous cw output power. The supply voltage is 60 kV and the conduction angle is 90°. What is the average current drawn from the power supply? Answer: 11.1 amperes. What is the average current when the tube is on? Answer: 44.4 amperes.

Problem 9.4. (a) Consider the class-D amplifier of Figure 9.7(a). Show that the peak sine-wave voltage on the load, R_L, will be $\pi V_{dc}/2$ and the dc supply current will be $\pi^2 V_{dc}/(8R_L)$. Hint: since the efficiency is 100% (there are no lossy components except the load), the power supplied by the dc source will be $I \cdot V_{dc}$ where I is the constant current flowing through the RF choke. But the power from the supply can also be written as V_{dc} times the average of the absolute value of the sine wave on the load. (The average of $|A\sin(\theta)|$ is $2A/\pi$.)

(b) For the amplifier of Figure 9.7(b), show that the peak sine-wave voltage on the load will be πV_{dc} and the dc supply current will be $\pi^2 \mathrm{V}_{dc}/(2R_L)$.

References

[1] Krauss, H. L., Bostian, C. W. and Raab, F. H. *Solid State Radio Engineering*, New York: John Wiley, 1980.

[2] Raab, F. H., High efficiency amplification techniques, *IEEE Circuits and Systems*, Vol. 7, No. 10, pp. 3–11, December 1975.

[3] Raab. F. H., *Idealized operation of the class-E tuned power amplifier*, *IEEE Trans. Circuits Syst*., Vol. CAS-25, pp. 725–735, Dec. 1977.

[4] Sokal, N. O. and Sokal, A. D., *Class-E – A new class of high-efficiency tuned single-ended switching power amplifiers*, *IEEE J. Solid-State Circuits*, vol. 10, no. 3, pp. 168–176, 1975.

[5] Eimac division of CPI, Inc., *Care and Feeding of Power Grid Tubes*, 5th edn, 2003, CPI, Inc., 301 Industrial Rd., San Carlos, CA. PDF: http://www.cpii.com/docs/related/22/C&F1Web.pdf

CHAPTER

10

Transmission lines

We draw circuit diagrams with "lumped"components: ideal *R*'s, *C*'s, *L'*s, transistors, etc., connected by lines that represent zero-length wires. But all real wires, if not much shorter than the shortest relevant wavelength, are themselves complicated circuit elements; the current is not the same everywhere along such a wire, nor is voltage uniform, even if the wire has no resistance. On the other hand, when interconnections are made with transmission lines, which are well-understood circuit elements, we can accurately predict circuit behavior. In this section we will consider two-conductor lines such as coaxial cables and open parallel wire lines. "Microstrip lines" (conducting metal traces on an insulation layer over a metal ground plane) behave essentially in the same way, but they have some subtle complications, which are mentioned in Appendix 10.1.

10.1 Characteristic impedance

The first thing one learns about transmission lines is that they have a parameter known as *characteristic impedance*, denoted Z_0. How "real" is characteristic impedance? If we connect an ordinary dc ohmmeter to the end of a 50-ohm cable will it indicate 50 ohms? Yes, if the cable is *very* long, so that a reflection from the far end does not arrive back at the meter before we finish the measurement. Otherwise, the meter will simply measure whatever is connected to the far end, which could be short, an open circuit, or a resistance. However, using a pulse generator and an oscilloscope, you can easily make an ohmmeter set-up that is fast enough that, even for a short cable, you can determine V_{in} and I_{in} and then calculate $V_{\text{in}}/I_{\text{in}} = Z_0$.

To make a theoretical determination of Z_0, we first model the transmission line as a ladder network made of shunt capacitors and series inductors, as shown in Figure 10.1.

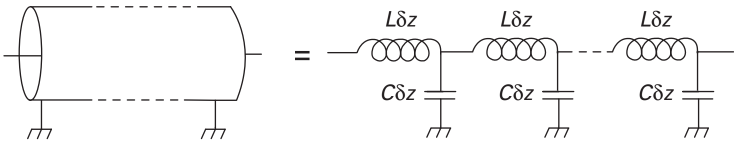

Figure 10.1. Transmission line model – a ladder network of infinitesimal *LC* sections.

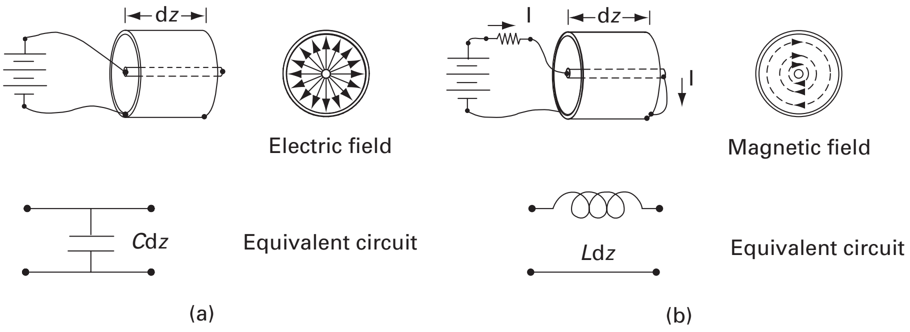

Figure 10.2. Capacitance and inductance per unit length.

To see that this model is reasonable, consider Figure 10.2(a), which shows the electric field lines in a length of coaxial cable connected to a voltage source. The field lines are radial and their number is obviously proportional to the length of the cable, so that capacitance per unit length is a constant. Likewise, a current through the cable (b) sets up a magnetic field, so another characteristic of the cable is its inductance per unit length. We will follow common convention and use the symbols C and L to denote capacitance and inductance per unit length. That convention is obvious when capacitors and inductors are labeled, respectively, $C\delta z$ and $L\delta z$, where δz is a short increment of length along the z-axis, i.e., parallel to the cable.

Every increment of a transmission line contributes series inductance and shunt capacitance; the ladder network shown in Figure 10.1 models a real transmission line in the limit that δz goes to zero. For some situations, e.g., baseband telephony and digital data transmission through long cables, the model must also include series and shunt resistance. At radio frequencies, however, the series reactance is usually much greater than the series resistance and the shunt reactance is usually much less than the shunt resistance so both resistances can be neglected. (See Problem 10.3.)

To see that $Z_0{}^2 = L/C$, consider the circuit of Figure 10.3, where we have added another infinitesimal LC section to the model transmission line, which is either infinitely long or terminated with a resistance equal to the characteristic impedance, so as to appear infinitely long. After adding the section, the line is still infinitely long and the impedance looking into it must still be Z_0. If the voltage and current at the input of the line were V and I, they will be modified to become $V+\delta V$ and $I+\delta I$ at the input to the new section. (This does not imply an increase in power; $V+\delta V$ and $I+\delta I$ are merely phase-shifted versions of V and I.)

Since the impedance looking into the line must stay the same, we have

$$\frac{V + \delta V}{I + \delta I} = \frac{V}{I}, \tag{10.1}$$

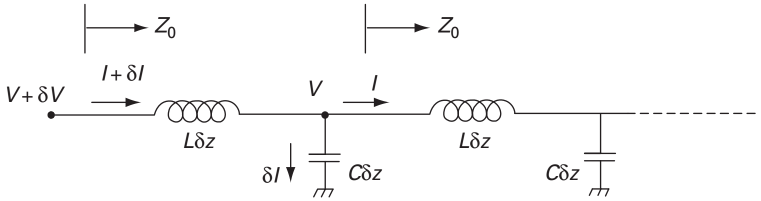

Figure 10.3. Adding another infinitesimal section must leave Z_0 unchanged.

from which $\delta V/\delta I = V/I = Z_0$.

Using this, and substituting $\delta I = (C\delta z)\,\mathrm{d}V/\mathrm{d}t$ and $\delta V = (L\delta z)\,\mathrm{d}/\mathrm{d}t(I+\delta I)$ and ignoring the vanishingly small $\delta z\,\delta I$ term, we have

$$Z_0 = \frac{\delta V}{\delta I} = \frac{L\delta z\,(\mathrm{j}\omega I)}{C\delta z\,(\mathrm{j}\omega V)} = \frac{L}{C}\,Z_0^{-1}. \tag{10.2}$$

Looking at the first and last terms of this equation, we see that $Z_0 = (L/C)^{1/2}$. Note: you can verify that, because δV and δI are infinitesimal, the above equations are the same if the network starts with a capacitor instead an inductor.

To evaluate Z_0, it is sufficient to know either L or C, since it follows from electrodynamics that they are related by $LC = \varepsilon_r/c^2$ where ε_r is the dielectric constant (relative to vacuum), and c is the speed of light. This relation between L and C holds for any two-conductor structure with translational symmetry such as an unlikely transmission line consisting of a square inner conductor inside a triangular outer conductor.

For a coaxial transmission line, $C = 2\pi\varepsilon_r\varepsilon_0/\ln(b/a)$ farads/meter, where a and b are the inner and outer radii and ε_0, the "permittivity of free space," is equal to $(4\pi\times 10^{-7}c^2)^{-1}$. Using this, together with the relation $LC = \varepsilon_r/c^2$, gives us $Z_0 = (\varepsilon_r)^{-1/2}\,60\ln(b/a)$. Note that Z_0 depends on the ratio a/b, but not on the size of the cable.

10.2 Waves and reflected waves on transmission lines

We will use a simple ac analysis to show that an applied sinusoidal voltage causes a spatial voltage sine wave to propagate down the line: Let us apply a voltage $e^{j\omega t}$ and find the voltage drop across an incremental length of line (see Figure 10.4).

Since we already know the input impedance is Z_0, the input current must be V/Z_0 and the voltage across the inductor can be written $\delta V = -(V/Z_0)(\mathrm{j}\omega L\delta z)$. But this is just the differential equation

$$\frac{\mathrm{d}V}{\mathrm{d}z} = -\mathrm{j}\omega\frac{L}{Z_0}V = -\mathrm{j}\omega\sqrt{LC}\,V. \tag{10.3}$$

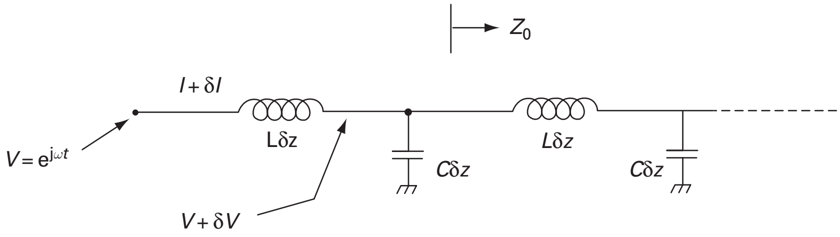

Figure 10.4. Finding the change in voltage, δV, over a distance δz.

The solution to this familiar equation is

$$V = V_f e^{-j\omega\sqrt{LC}z} = V_f e^{-jkz} \quad \text{where} \quad k = \omega\sqrt{LC}, \tag{10.4}$$

where V_f is a constant, the amplitude. The constant k is known as the *propagation constant* and is the number of radians the wave progresses per unit length. The wave therefore repeats in a distance (the wavelength) given by $\lambda = 2\pi/k$. Since $V/I = Z_0$, the current along the line is also a wave: $I = (V_f/Z_0)e^{-jkz}$. If we include the otherwise implicit multiplicative time dependence factor $e^{j\omega t}$, the voltage is

$$V = V_f e^{j\omega t} e^{-jkz} = V_f e^{j(\omega t - kz)}. \tag{10.5}$$

This is just a sine wave running in the forward z-direction. The complex exponential now contains space as well as time but, as always, the physical voltage is the real part, i.e., Re[$V_f e^{j(\omega t - kx)}$] which is a weighted superposition of $\sin(\omega t - kz)$ and $\cos(\omega t - kz)$. For a point of constant phase, $\omega t - kz = \text{constant}$, we have $\delta z/\delta t = \omega/k$. This velocity, $\omega/k = c/\sqrt{\varepsilon_r}$, is known as the *phase velocity*, v_{phase}. Figure 10.5 shows a forward-running wave on a coaxial cable. The electric and magnetic field lines are drawn only at the points where they reach their peak values. A graph shows the spatial distribution. Everything has the same phase, i.e., the voltage, current, and charge density all rise and fall together along the z-axis. Note that a wave of amplitude V transfers power at a rate $|V|^2/(2Z_0)$.

A transmission line can equally well support waves running in the negative z-direction. If we had assumed a current in the $(-z)$-direction, the phase would progress as $\omega t + kz$. A transmission line in a circuit operating at a frequency ω will, in general, have both a forward wave and a reverse wave. The waves have complex amplitudes, V_f and V_r, each containing magnitude and phase. Of course both waves have the same frequency and propagation constant. We regard current as positive when it is in the (+z)-direction, so the current of a forward wave is $I_f(z,t) = V_f(z,t)/Z_0$, but the current of a reverse wave is $I_r(z,t) = -V_r(z,t)/Z_0$, since the reverse wave is traveling in the $(-z)$-

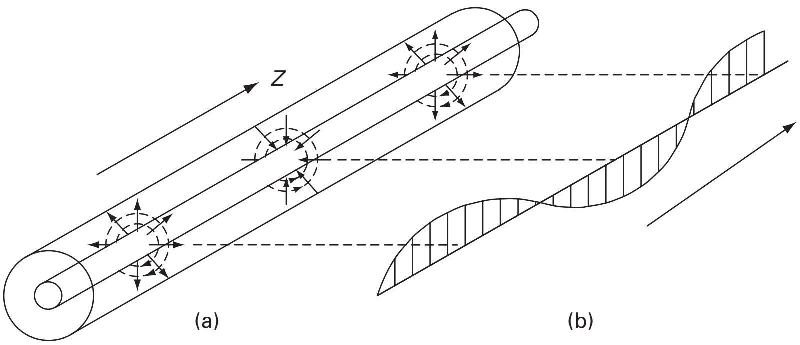

Figure 10.5. Forward wave on a transmission line.

direction. Together, the forward and reverse waves are, in general, equivalent to a stationary (*standing*) wave plus a single propagating wave.

Note also that the phase velocity is independent of ω; there is no dispersion in this kind of lossless transmission line. Therefore, if we apply an arbitrary voltage waveform, $V_{\text{arb}}(t)$, at the input to the line, this waveform, considered as a Fourier superposition of sine waves, will propagate down the line without distortion. At any point z, the voltage will be $V_{\text{arb}}(t - z/v_{\text{phase}})$, a delayed but undistorted version of the input signal. For example, if, at $t = 0$, we connect a dc voltage to the line, a step function propagates down the line.

The *electrical length* of a line is the phase change imparted by the line. For example, a "quarter wave line" imparts a 90° phase shift, $kl = \pi/2$, and therefore $l = \pi/(2k) = \pi c/(2\omega\sqrt{\varepsilon_r}) = (c/f)/(4\sqrt{\varepsilon_r}) = 1/4(\lambda_0/\sqrt{\varepsilon_r})$, where λ_0 is the wavelength in free space.

Standing waves

When both a forward and a reverse wave are present on a transmission line the voltage along the line, which is the sum of the contributions from the two waves, forms an interference pattern or standing wave. To see this, let $V(z,t) = V_f e^{j(\omega t - kz)} + V_r e^{j(\omega t + kz)}$. The real parts of these two rotating phasors will be in phase at points along the transmission line which are separated by $\lambda/2$. At these points, the magnitude of the sum will be $|V_f| + |V_r|$. Halfway between these points, the real parts of the phasors will be out of phase and the magnitude of the sum will be $||V_f| - |V_r||$. The ratio of these maximum and minimum voltage magnitudes is called the *voltage standing wave ratio*: $\text{VSWR} = (|V_f| + |V_r|)/||V_f| - |V_r||$. If $|V_f| = |V_r|$ there is only a standing wave and the VSWR is infinite. When $|V_f| \neq |V_r|$, the weaker one, along with an equal portion of the stronger one, form a standing wave, leaving the remainder of the stronger one as a travelling wave.

10.3 Modification of an impedance by a transmission line

From the discussion above, you can see that a transmission line terminated by a resistor of value Z_0 will always present an input impedance of Z_0. But a piece of transmission line that is terminated with an arbitrary impedance, $Z \neq Z_0$, as shown in Figure 10.6, will produce a modified ("transformed") impedance, Z'.

This figure shows a line of length l whose right-hand end ($z=0$) is connected to some impedance Z_L (L denotes "load"). Assume that some constant ac source produces a constant incident wave traveling to the right, $V_f e^{-jkz}$ (we will not bother writing the always present factor $e^{j\omega t}$), and that Z_L causes a constant reflected wave, $\Gamma\, V_f e^{jkz}$, to travel to the left.[1] The factor Γ is known as the *reflection coefficient*. At any point, z, the voltage on the line is $V(z)=V_f e^{-jkz}+\Gamma\, V_f e^{jkz}$. The corresponding current is $I(z)=(V_f/Z_0)(e^{-jkz}-\Gamma e^{jkz})$. The minus sign occurs because the current in the reflected wave flows in the negative z-direction. At the right-hand end ($z=0$), the load ensures that $V(0)/I(0)=Z_L$. This will give us Γ:

$$\frac{V(0)}{I(0)}=Z_L=\frac{(1+\Gamma)}{(1-\Gamma)/Z_0} \quad \text{so} \quad \frac{Z_L}{Z_0}=\frac{(1+\Gamma)}{(1-\Gamma)} \quad \text{and} \quad \Gamma=\frac{(Z_L-Z_0)}{(Z_L+Z_0)}. \tag{10.6}$$

Putting this expression in Equation (10.6) for Γ, together with the expressions for $V(z)$ and $I(z)$, we can immediately find $V(-l)/I(-l)$ which is what we are after, i.e., Z', the input impedance at a point l to the left of the load:

$$\begin{aligned} Z' &= \frac{V(-l)}{I(-l)} = \frac{e^{-jk(-l)}+\Gamma e^{jk(-l)}}{e^{-jk(-l)}/Z_0-\Gamma e^{jk(-l)}/Z_0} \\ &= Z_0\frac{(Z_L+Z_0)e^{jkl}+(Z_L-Z_0)e^{-jkl}}{(Z_L+Z_0)e^{jkl}-(Z_L-Z_0)e^{-jkl}} \end{aligned}$$

or

$$Z' = Z_0\frac{Z_L+jZ_0\tan(kl)}{Z_0+jZ_L\tan(kl)}. \tag{10.7}$$

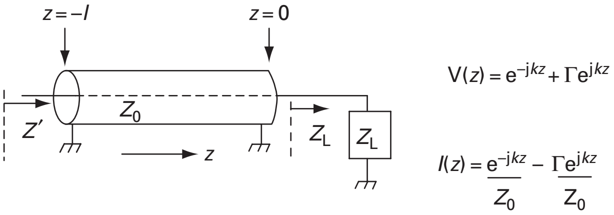
$z=-l$
$z=0$
Z_0
Z'
z
Z_L
Z_L
$V(z)=e^{-jkz}+\Gamma e^{jkz}$
$I(z)=\frac{e^{-jkz}}{Z_0}-\frac{\Gamma e^{jkz}}{Z_0}$

Figure 10.6. An impedance is modified when seen through a transmission line.

---

[1] Since everything is linear, superposition holds and the incident and reflected waves do not collide or interact in any way. They simply pass through one another unaltered. At any point, the current is the sum of their currents and the voltage is the sum of their voltages.

This important result, the modification of an impedance Z_L by a length l of transmission line, is not hard to remember; it has no minus signs and is symmetric. Just remember (1+j tan)/(1+j tan). Once you have written this framework, you will remember how to put in the coefficients. Some important special cases are listed below:

- If $Z_L = Z_0$, then $Z' = Z_0$ for any length of line.
- If $Z_L = 0$ (a short) then $Z' = jZ_0\tan(kl)$, a pure reactance, which is inductive[2] for $kl < \pi/2$, then capacitive, etc.
- If Z_L = infinity (an open circuit) then $Z' = Z_0/j\tan(kl)$ which is capacitive for $kl < \pi/2$, then inductive, etc.
- An impedance is left unchanged by a line of arbitrary Z_0 whose length is a half-wave ($kz = \pi$) or any integral multiple of a half-wave.
- A quarter-wave line ($kz = \pi/2$) or an odd multiple of a quarter-wave line, *inverts* an impedance: $Z' = Z_0^2/Z_L$. A short is transformed into an open and an open into a short, an inductor is transformed into a capacitor and vice versa, etc.

10.4 Transmission line attenuation

In a lossy transmission line, i.e., a line that causes attenuation of the signal, the e^{-jkz} or e^{jkz} spatial dependence of the wave is replaced by $e^{-jkz}\, e^{-\alpha z} = e^{-j(k-j\alpha)z}$ (forward wave) or $e^{jkz}\, e^{\alpha z} = e^{j(k-j\alpha)z}$ (reverse wave), where α is the attenuation constant. In a distance $1/\alpha$, the amplitude falls by a factor 1/e and the power falls by a factor $(1/e)^2$. Note that k for a lossless line is simply replaced by $k - j\alpha$, i.e., the propagation constant becomes complex. You can put this complex k into the "tan tan" formula to see how an impedance is modified by a lossy cable.

Transmission line attenuation is usually expressed in units of dB/meter. To find α for a line whose loss is A_L dB/m, note that, since the amplitude falls by a factor $e^{-\alpha\cdot 1}$ in 1 meter, we can write $-A_L = 10\log(e^{-\alpha\cdot 1})^2 = -20\alpha\log(e)$ from which $\alpha = A_L/(20\log(e))$.

10.5 Impedance specified by reflection coefficient

We have seen that an impedance Z produces a reflection coefficient given by $\Gamma = (Z - Z_0)\,/\,(Z + Z_0)$. This relation is easily inverted, $Z = Z_0(1+\Gamma)/(1-\Gamma)$, so there is a one-to-one mapping between Z and Γ. In antenna and microwave work, especially when using S-parameter analysis (Chapter 28), it is customary to think in terms of Γ, rather than Z.

One big advantage of working in the complex Γ-plane is that the modification of an impedance (represented by its equivalent Γ) is extremely simple. The reflection coefficient for the given impedance as seen through a length l of transmission line is just

---

[2] Note that "inductive" does not mean equivalent to a lumped inductor since $Z_0\tan(kl) = Z_0\tan(\omega l/v_{\text{phase}})$ is not proportional to ω, except for small kl. Likewise, a short open-ended line is not equivalent to a lumped capacitor, except for small $\omega l/v_{\text{phase}}$.

$$\Gamma' = \Gamma e^{-j2kl}, \tag{10.8}$$

which means we simply rotate the point clockwise around the origin ($\Gamma = 0$) by an angle $2kl$ to give Γ', the modified reflection coefficient. This is easy to see: when we add a length of cable, the incident wave's phase is delayed by kl getting to the end of the cable and the reflected wave is delayed by the same kl getting back again. The effect of a cable is therefore to rotate the complex number Γ clockwise by an angle $2kl$.[3] (Since the time dependence is $e^{j\omega t}$, the round-trip time delay is a clockwise displacement.) Keep in mind that the Γ-plane is a complex plane but that it is *not* the $R + jX$ plane. Let us look at a few special points in the Γ-plane.

1. The center of the plane, $\Gamma = 0$, corresponds to a reflected wave of zero amplitude, so this point represents the impedance $Z_0 + j0$.
2. The magnitude of Γ (radius from the origin) must be less than or equal to unity for passive impedances. Otherwise the reflected wave would have more power than the incident wave.
3. The point $\Gamma = -1 + j0$ corresponds to $Z = 0$, a short circuit.
4. The point $\Gamma = 1 + j0$ corresponds to $Z = \infty$, an open circuit.
5. Points on the circle $|\Gamma| = 1$ correspond to pure reactances, $Z = 0 + jX$. All points inside this circle map to impedances with positive nonzero R.
6. The point $\Gamma = 0 + j1$ corresponds to an inductance, $Z = 0 + jZ_0$. All points in the top half of the Γ-plane are "inductive," i.e., $Z = R + j|X|$ or, equivalently, $Y = G - j|B|$.
7. The point $\Gamma = 0 - j1$ corresponds to a capacitance, $Z = 0 - jZ_0$. All points in the bottom half of the Γ-plane are "capacitive," i.e., $Z = R - j|X|$ or, equivalently, $Y = G + j|B|$.

These special cases of mapping of Z into Γ are shown in Figure 10.7.

In the Γ-plane, if you plot $\Gamma = R + jX$, where R is a constant and X varies, you will get a circle centered on the real axis and tangent to the line $\mathrm{Re}(\Gamma) = 1$. For every value of R there is one of these "resistance circles." The resistance circle for $R = 0$ is the unit circle in the Γ-plane. The resistance circle for $R = \infty$ is a circle of zero radius at the point $\Gamma = 1 + j0$. Likewise, if you plot $\Gamma(R + jX)$ where X is a constant and R varies, you will get "reactance circles" centered on the line $\mathrm{Re}(\Gamma) = 1$ and tangent to the line $\mathrm{Im}(\Gamma) = 0$. These circles are shown in Figure 10.8.

If you now trim the circles to leave only the portions within the $|\Gamma| = 1$ circle (corresponding to passive impedances, i.e., impedances whose real part is

---

[3] If the line is lossy, the magnitude of Γ decreases as it rotates around the origin, forming a spiral. For a long enough length of lossy line, Γ spirals all the way into the origin producing $Z = Z_0$, no matter what value of Z terminates the far end of the cable.

Figure 10.7. Impedances mapped into the reflection plane.

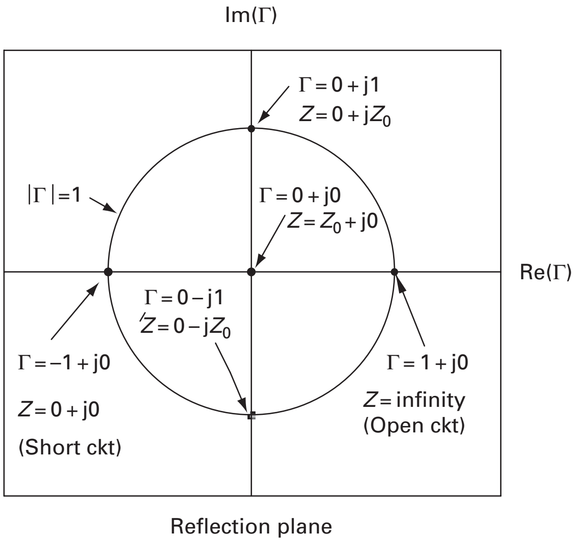
Im(Γ)
Γ = 0 + j1
Z = 0 + jZ0
|Γ| = 1
Γ = 0 + j0
Z = Z0 + j0
Re(Γ)
Γ = 0 − j1
Z = 0 − jZ0
Γ = −1 + j0
Z = 0 + j0
(Short ckt)
Γ = 1 + j0
Z = infinity
(Open ckt)
Reflection plane

Figure 10.8. Loci of constant resistance and of constant reactance – circles in the Γ-plane.

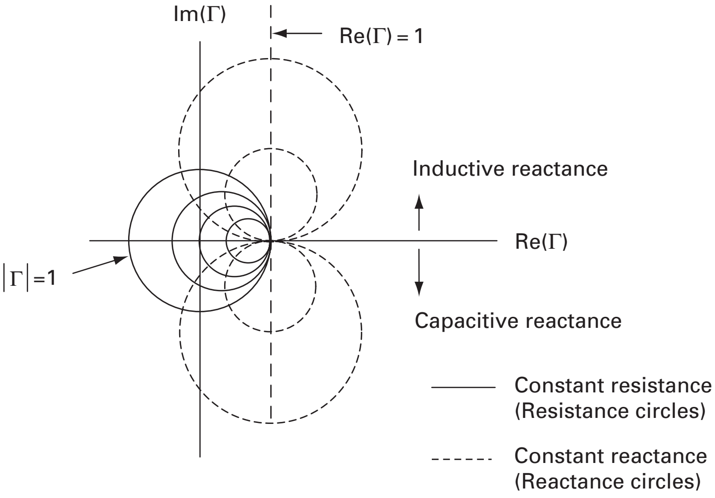
Im(Γ)
Re(Γ) = 1
Inductive reactance
Re(Γ)
|Γ| = 1
Capacitive reactance
Constant resistance
(Resistance circles)
Constant reactance
(Reactance circles)

positive) you are left with a useful piece of graph paper, the famous *Smith chart*, shown in Figure 10.9.

The circular R and X "axes" on the Smith chart allow you to locate the Γ-point that corresponds to $Z = R + jX$. We have already seen that when we have located an impedance on the Γ-plane, we can find how that impedance is modified by a length of transmission line (whose Z_0 is the same as the Z_0 used to draw the chart) by rotating the point clockwise around the origin. We simply rotate the point clockwise around the origin by an angle equal to twice the electrical length of the line. The values of R and X corresponding to the rotated point can be read

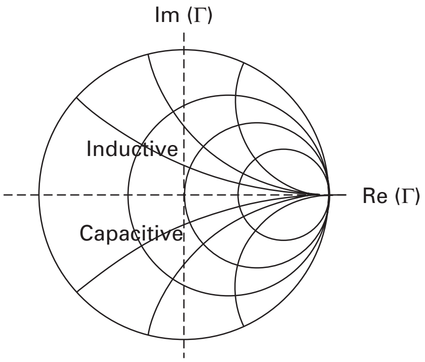
Im (Γ)
Inductive
Re (Γ)
Capacitive

Figure 10.9. The Smith chart – resistance and reactance circles on the Γ-plane.

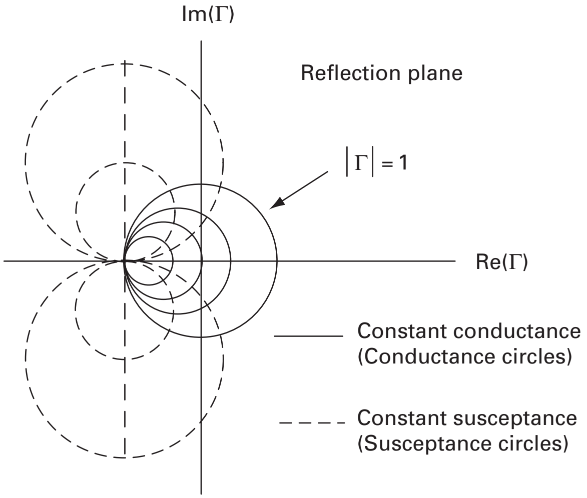
Im(Γ)
Reflection plane
|Γ| = 1
Re(Γ)
Constant conductance
(Conductance circles)
Constant susceptance
(Susceptance circles)

Figure 10.10. Conductance and susceptance circles.

off the chart's R and X "axes." We can also use the chart to find how an impedance is modified by adding a series R or series X. In this operations, the Smith chart can be considered something of a calculator. Note that the Smith chart can also be made with G and B "axes". As you might guess, these produce "G circles" and "B circles" as shown in Figure 10.10.

Sometimes the Smith chart contains G and B circles as well as R and X circles. This full-blown chart, which can be quite dense, is shown in Figure 10.11. Again, remember the Smith chart is actually a rectangular graph of Γ; the x-axis is Re(Γ) and the y-axis is Im(Γ). Because only the area inside the circle $|\Gamma| = 1$, i.e., $x^2+y^2 = 1$, is used, the Smith chart resembles a polar graph. And, indeed, when we rotate a point around the origin to how a transmission line modifies an impedance, we are using it in a polar fashion. Sometimes the Smith chart is scaled for a specific Z_0 (usually 50 ohms or 75 ohms). Other charts are normalized; the $R = 1$ circle would be the 50-ohm circle if we are dealing with 50-ohm cable, etc.

Figure 10.11. Smith chart with *R*, *X*, *G*, and *B* circles.

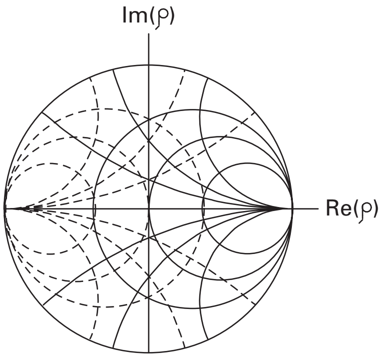

10.6 Transmission lines used to match impedances

Designing a matching network becomes an exercise in moving from a given Γ to a desired Γ′ in the reflection plane. Working graphically, it is often easy to find a matching strategy. Let us use the Smith chart and revisit the 1000-ohm-to-50-ohm matching circuit example of Chapter 2.

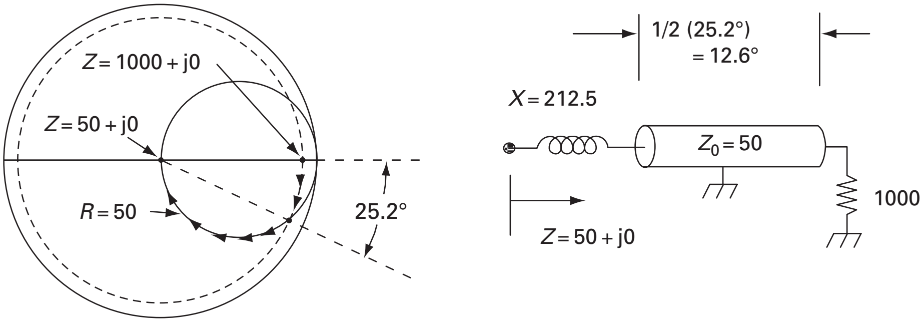

Figure 10.12. Conversion from 1000 ohms to 50 ohms – transmission line and inductor circuit.

I. The starting impedance, 1000 ohms, and the final target impedance, 50 ohms, are indicated on the chart in Figure 10.12. Also shown is the 50-ohm circle. We can use a (50-ohm) transmission line to move along the dashed circle until we reach the 50-ohm circle. Now we have $R = 50$ plus a capacitive reactance. A series inductor will cancel the capacitive reactance, taking us to $Z = 50+j0$ (the center of the chart).

II. Another solution (Figure 10.13) would be to use a longer piece of cable to circle most of the chart, hitting the 50-ohm circle in the top half of the plane. At this point we have $Z = 50 + jX$ where X is positive (inductive). We can add a series capacitor to cancel this X and again arrive at $Z = 50 + j0$.

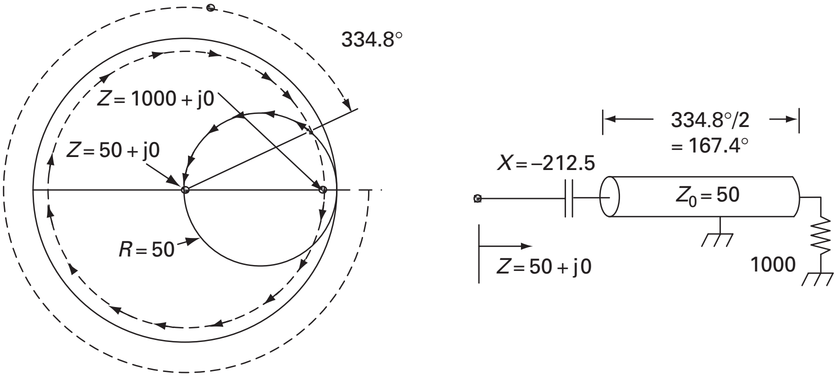

Figure 10.13. Transmission line and capacitor matching circuit.

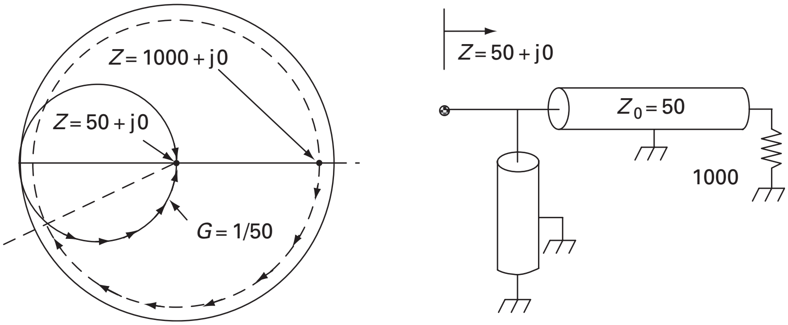

Figure 10.14. Series and shunt transmission line matching circuit.

III. So far we have only used series elements. Let us now start by traveling around to the $G = 1/50$ circle. Then we can add a shunt element to reach the center of the chart. The first intersection of the $G = 1/50$ circle is in the lower half-plane (capacitive) so, to get from this point to the center, we need a shunt inductor. Instead of a lumped inductor we might use a shorted length of transmission, as shown in Figure 10.14, to make a matching circuit using only transmission line elements.

IV. Figure 10.15 shows a solution that uses no transmission line. We start on the $G = 1/1000$ circle, at $G = 0$. If we apply shunt reactance we can move along this circle. Let us pick shunt inductance which will move us upward along the G circle to the 50-ohm circle. We now have $R = 50$, but there is inductive reactance. As in the above example, we can now cancel the inductance reactance with a series capacitor. This is just the L-network found in Chapter 2.

V. If we had used shunt capacitance rather than shunt inductance, we would have moved downward to the 50-ohm circle, as shown in Figure 10.16. The remaining series capacitance can be cancelled with an inductor. This produces an L-network where the positions of the L and C are reversed.

Figure 10.15. *LC* matching network.

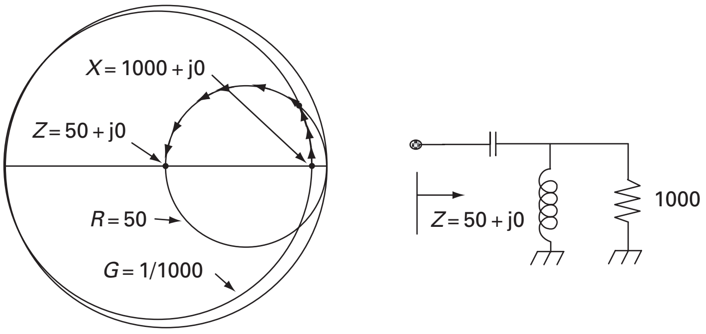
X = 1000 + j0
Z = 50 + j0
R = 50
G = 1/1000
Z = 50 + j0
1000

Figure 10.16. *CL* matching network.

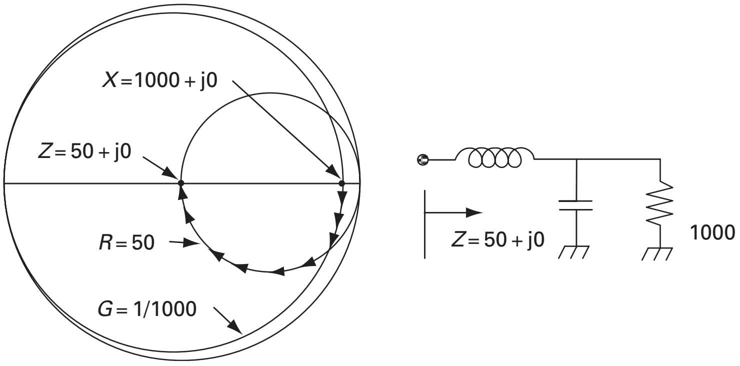
X = 1000 + j0
Z = 50 + j0
R = 50
G = 1/1000
Z = 50 + j0
1000

Figure 10.17. An impedance-admittance chart.

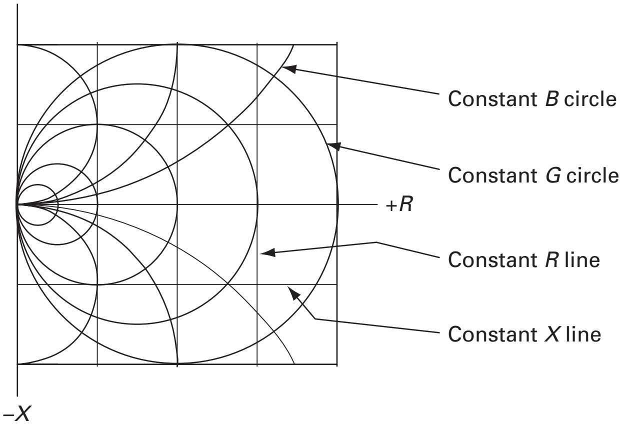
Constant B circle
Constant G circle
+R
Constant R line
Constant X line
−X

In these examples, our final impedance was at the center of the chart ($Z = 50 + j0$), but you can see that these techniques allow us to transform any point on the chart (i.e., any impedance) into any other point on the chart (any other impedance).

The Smith chart is a favorite because it handles networks that include transmission lines as well as inductors and capacitors. If we did not care about transmission lines, then any chart that maps R, X into G, B would do. For example, take the R, X plane (half-plane, since we will exclude negative R). Draw in the curves for $G =$ constant and $B =$ constant. The resulting chart, shown in Figure 10.17, can be used to design lumped element L, C,R ladder networks, such as the networks of Figures 10.15 and 10.16.

Appendix 10.1. Coaxial cable – Electromagnetic analysis

This chapter began with a derivation of Z_0 based on an equivalent lumped-element circuit model of a transmission line. That derivation required only elementary ac circuit theory, but is a rather indirect approach to what is really a problem in electromagnetics. Even then, some electromagnetic theory is needed to derive the expressions for capacitance and inductance per unit length.

An electromagnetic analysis of a coaxial transmission line is presented here for the reader who has some familiarity with Maxwell's equations. We make use of the fact that the propagation velocity of a *TEM wave*[4] is given by $v = (\mu\varepsilon)^{-1/2}$, where μ is the magnetic permeability and ε is the electrical permitivity of the material through which the fields propagate.[5]

To find the impedance of the coaxial line, we will first assume that the current on the inner conductor is given by $I = I_0 \cos(\omega t - kz)$, which is a wave traveling in the $(+z)$-direction. This is illustrated in Figure 10.18.

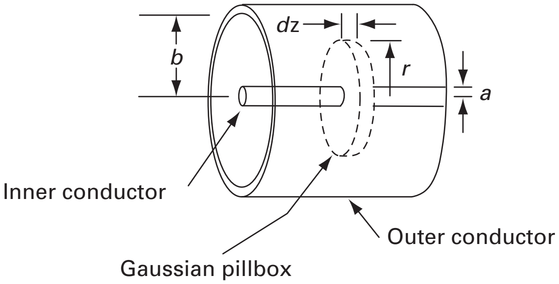

Figure 10.18. Transmission line element.

We will then proceed to find the charge density, the electric field, and then the voltage, which will have the form $V_0 \cos(\omega t - kz)$. Once we have the voltage, the characteristic impedance is simply given by V_0/I_0. (Note that the current in the outer conductor is just the negative of the current in the inner conductor.)

---

[4] In a TEM wave, by definition, both the electric field and the magnetic field are transverse, i.e., perpendicular to the direction along which the wave propagates. In most applications of coaxial cables and parallel-wire transmission lines, the wavelength is much greater than the transverse dimensions of the line and only TEM waves can propagate. Waves in free space are also TEM waves.

[5] To find the propagation velocity of a TEM wave: The variables t and z appear in E_x, E_y, B_x, and B_y only in the factor $e^{j(\omega t - kz)}$. Using the condition $E_z = 0$, the x-component of the Maxwell equation $\mathrm{curl}(E) = -\partial B/\partial t$ gives us $B_x = -(k/\omega)E_y$. Likewise, using the condition $B_z = 0$, the y-component of the Maxwell equation $\mathrm{curl}(B/\mu) = \partial(\varepsilon E)/\partial t$ gives us $B_x = -(\omega\mu\varepsilon/k)E_y$. Equating these two expressions for B_x gives $k^2 = \omega^2\mu\varepsilon$.

Consider an incremental segment of the inner conductor from z to $z + \delta z$. The rate at which charge accumulates on this element is $\partial/\partial t\ (\rho_L)\delta z$, where ρ_L is the charge per unit length. But the rate at which charge accumulates in δz is nothing more than the difference between the current flowing into δz and the current flowing out of δz. Therefore, we can write

$$\frac{\partial \rho_L}{\partial t} = \frac{-\partial I}{\partial z} = -k\, I_0 \sin(\omega t - kz). \tag{10.9}$$

Integrating $d\rho_L/dt$ with respect to time gives us the linear charge density (coulombs/meter):

$$\rho_l(z,t) = \frac{1}{\omega}(kI_0 \cos(\omega t - kz)) = \frac{kI_0}{\omega}\cos(\omega t - kz). \tag{10.10}$$

Now that we know the charge density, we can find the electric field. The field is radial with field lines like spokes of a wheel. Imagining a Gaussian “pillbox” of radius r and height δz around the center conductor, we use Gauss’s law: the integral of the E field over the sidewall surface must be equal to the enclosed charge divided by ε:

$$E(r,z,t)(2\pi r\delta z) = \frac{1}{\varepsilon}\rho_l(z,t)\delta z. \tag{10.11}$$

Substituting for λ and solving for E, we have

$$E(r,z,t) = \frac{kI_0 \cos(\omega t - kz)}{2\pi r\varepsilon\omega}. \tag{10.12}$$

Integrating this electric field from $r = a$ to $r = b$ gives us the voltage between the inner and outer conductors:

$$V(z,t) = \int_a^b E(r,z,t)dr = \frac{kI_0 \cos(\omega t - kz)\ln(b/a)}{2\pi\varepsilon\omega}. \tag{10.13}$$

Finally, we divide $V(z,t)$ by $I(z,t)$ to get the characteristic impedance:

$$Z_0 = \frac{V(z,t)}{I(z,t)} = \frac{k\ln(b/a)}{2\pi\varepsilon\omega} = \frac{\ln(b/a)}{2\pi\varepsilon v} = \frac{1}{2\pi}\left(\frac{\mu}{\varepsilon}\right)^{1/2}\ln(b/a), \tag{10.14}$$

which is the same as the result we obtained using the δL δC ladder network equivalent circuit.

This derivation (as well as the LC derivation) for Z_0 is for TEM waves, where both $\boldsymbol{E}$ and $\boldsymbol{H}$ are perpendicular to z. For TEM solutions to exist, the line must be uniformly filled with homogenous dielectric material or vacuum. The dielectric can be lossy, but the metal conductors must, strictly speaking, have no resistance. In practice, these conditions are usually not satisfied perfectly, and the waves will be slightly different from the TEM waves corresponding to ideal conditions. In particular, the waves will have a small E_z or H_z field, or both. Microstrip lines are a case of nonuniform dielectric; some of the E-field lines arch through the air above the conductor, before plunging through the dielectric to the ground plane. The wave must have a unique phase velocity, but $(\mu\varepsilon)^{-1/2}$ has one value in the air and another value in the dielectric. The waves, therefore, cannot be TEM. They turn out to have both E_z and H_z components. Known as quasi-TEM waves, they show some frequency dependence in both Z_0 and v_{phase}, which can be important at millimeter-wave frequencies. Closed form expressions have not been derived for a microstrip; designers find Z_0 and v_{phase} vs. frequency by using graphs or approximate formulas based on numerical solutions of Maxwell's equations.

Problems

Problem 10.1. A common 50-ohm coaxial cable, RG214, has a shunt capacitance of 30.8 pF/ft. Calculate the series inductance per ft and the propagation velocity.

Problem 10.2. (a) Use the "tan tan" formula to show that a short length, δz, of transmission line, open-circuited at the far end, behaves as a capacitor, i.e., that it has a positive susceptance, directly proportional to frequency. Express the value of this capacitor in terms of the cable's capacitance per unit length. (Hint: $\tan(\theta) \approx \theta$ for small θ.)

(b) Show that a short length, δz, of transmission line, short-circuited at the far end, acts as an inductor, i.e., that it has a negative susceptance inversely proportional to frequency. Express the value of this inductor in terms of the cable's inductance/unit length.

Problem 10.3. (a) Find a formula for the characteristic impedance of a lossy cable where the loss can be due to a series resistance per unit length, R, as well as a parallel conductance per unit length, G. R represents the ohmic loss of the metal conductors while G represents dielectric loss.

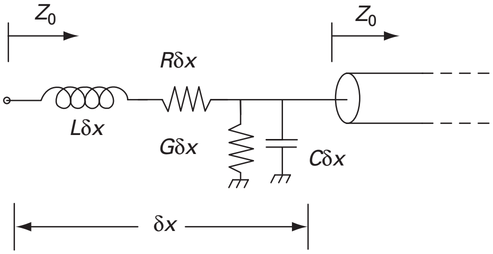

Hint: You can generalize the result for the lossless cable by simply replacing L by $L+R/(j\omega)$ and C by $C+G/(j\omega)$.

(b) Find the formula for the propagation constant k of this lossy cable. Hint: apply the substitutions given above to the formula $k = \omega\sqrt{LC}$. What distance (in wavelengths) is required to reduce by 1/e the power of a signal at frequency ω_1 if $R/(\omega_1) = 0.01L$?

Problem 10.4. If the (sinusoidal) voltage, V, and current, I, at the right-hand end of a transmission line are given, find the corresponding voltage, V', and current, I', at the left-hand end.

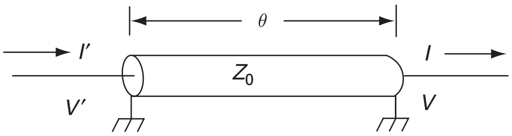

Hint: assume the (complex) voltage on the line is given by $V(\phi) = V_F e^{-j\phi} + V_R e^{j\phi}$. The corresponding current is given by $Z_0 I(\phi) = V_F e^{-j\phi} - V_R e^{j\phi}$. Let $\phi = 0$ at the right-hand end. Show that $V_F = (V + IZ_0)/2$ and $V_R = (V - IZ_0)/2$. Then show that, at the left-hand end, where $\phi = -\theta$, that $V' = V\cos\theta + IZ_0\, j\sin\theta$ and $I' = I\cos\theta + j\sin\theta\, V/Z_0$.

Problem 10.5. Use the results of Problem 10.4 to upgrade your ladder network analysis program (Problem 1.3) to handle another type of element, a series lossless transmission line. Three parameters are necessary to specify the line. These could be the characteristic impedance, the physical length, and the velocity of propagation. For convenience in later problems, however, let the three parameters be the characteristic impedance (Z_0), the electrical length (θ_0) in degrees for a particular frequency, and that frequency (f_0). A 50-ohm cable that has an electrical length of 80° at 10 MHZ would appear in the circuit file as "TL, 50, 80, 10E6." For any frequency, f, the electrical length is then $\theta = \theta_0 f/f_0$.

Example answer: For the MATLAB program shown in Problem 1.3, insert the following lines of code in "elseif chain":

```
elseif strcmp(component,'TL')==1
ckt_index=ckt_index+1; Z0=ckt{ckt_index}; %characteristic impedance
ckt_index=ckt_index+1; refdegrees=ckt{ckt_index};%electrical length
ckt_index=ckt_index+1; reffreq=ckt{ckt_index};   %at ref. frequency
eleclength= pi/180*f(i)*(refdegrees/reffreq);
Iold=I; I=I*(cos(eleclength))+V*(1j/Z0*sin(eleclength));
V=V*(cos(eleclength))+1j*Z0*Iold*(sin(eleclength));
```

Problem 10.6. Use your program to analyze the circuit of Figure 10.13. Assume a design frequency, say 1 MHZ, in order to determine the value of the capacitor. Run the analysis from 0 to 2 MHz. Then make the transmission line 360° longer and repeat the analysis. What form will the response take if the transmission line is made very long?

Problem 10.7. A 50-ohm transmission line is connected in parallel with an equal length transmission line of 75 ohms, i.e., at each end the inner conductors are connected and the outer conductors are connected. The cables have equal phase velocities. Show that the characteristic impedance of this composite transmission line is given by (50·75)/(50+75), i.e., the characteristic impedances add like parallel resistors.

Problem 10.8. In the circuit shown below, the impedance, Z, is modified by a transmission line in parallel with a lumped impedance, Z_1, which could be an R, C, or L or a network.

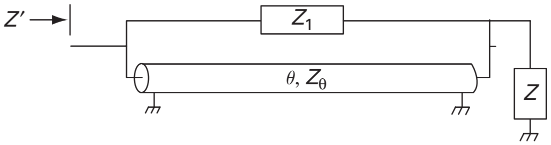

Show that the admittance looking in from the left, $Y' = 1/Z'$, is given by

$$\frac{1}{Z'} = Y' = Y_0 \frac{Y + \mathrm{j}Y_0 \tan\theta + \left(2Y_1 - \dfrac{2Y_1}{\cos\theta} + \mathrm{j}\dfrac{YY_1}{Y_0}\tan\theta\right)}{Y_0 + \mathrm{j}Y\tan\theta + (\mathrm{j}Y_1\tan\theta)}.$$

Hint: extend the argument used in the text to find Z' for a cable without a bridging lumped element. Assume a forward and reverse wave in the cable with amplitudes 1 and Γ. The voltage on the cable is then $V(z) = e^{j\omega t}\,(e^{-jkz} + \Gamma e^{jkz})$ and the current is $I(z) = Z_0^{-1} e^{j\omega t}\,(e^{-jkz} - \Gamma e^{jkz})$. The current into Z is the sum of the current from the cable and the current from Z_1 while the current into the circuit is the sum of the current into the cable and the current into Z_1.

Problem 10.9. Using a 50-ohm network analyzer, it is found that a certain device, when tested at 1 GHz, has a (complex) reflection coefficient of 0.6 at an angle of −22° (standard polar coordinates: the positive x-axis is at 0° and angles increase in the counterclockwise direction).

(a) Calculate the impedance, R+jX.

(b) Find the component values for both the equivalent series R_sC_s circuit and the equivalent parallel R_pC_p circuit that, at 1 GHz, represent the device.

Problem 10.10. The circuit below matches a 1000-ohm load to a 50-ohm source at a frequency of 10 MHz. The characteristic impedance of the cable is 50 ohms.

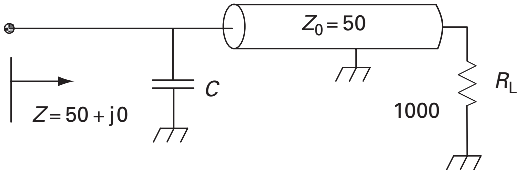

(a) Make a Smith chart sketch that shows the derivation of this circuit.

(b) Find the length of the (shortest) cable and the value of the capacitor. Specify the length in degrees and the capacitance in picofarads. Calculate these values rather than reading them from an accurately drawn Smith chart.

(c) Use your ladder network analysis program (Problems 1.3 and 10.5) to find the transmission from 9 MHz to 11 MHz in steps of 0.1 MHz.

Problem 10.11. Find a transmission line element to replace the capacitor in the circuit of Problem 10.9.

Problem 10.12. Suppose that a transmission line has small shunt susceptance (capacitive or inductive) at a point z. By itself, this will cause a small reflection. If an identical shunt reactance is placed one quarter-wave from the first, its reflection will compensate the first and the cable will have essentially perfect transmission. Show that this is the case (a) analytically, using the "tan tan" formula for Z' and B', and (b) graphically, using the Smith chart (the area around the center of the chart).

Problem 10.13. Find the size and position of the constant resistance circles on the normalized Smith chart. Use the following procedure:

We have $z(x)=r+\mathrm{j}x$ where x is a variable and r is a constant. This vertical line in the z-plane maps into the ρ-plane via the equation $\rho(x)=[z(x)-1]/[z(x)+1]$. We want to show that the locus of points in the ρ-plane is a circle with radius $1/(r+1)$ centered at $[r/(r+1), 0]$.

Assume that the locus will be a circle centered on the real axis at $[a,0]$. Write the equation $|\rho(x)-a|=\mathit{radius}$. This equation has the form

$$|[N_{\mathrm{Re}}(x)+\mathrm{j}N_{\mathrm{Im}}(x)]/[D_{\mathrm{Re}}(x)+\mathrm{j}\,D_{\mathrm{Im}}(x)]|=\mathit{radius}, \tag{1}$$

where $N_{\mathrm{Re}}(x)$ and $N_{\mathrm{Im}}(x)$ are the real and imaginary parts of the numerator and $D_{\mathrm{Re}}(x)$ and $D_{\mathrm{Im}}(x)$ are the real and imaginary parts of the denominator. If every point on the circle is to have the same value of r, the radius of the circle must be independent of x.

$$\begin{aligned}|\rho(x)-a|^2 &= [(N_{\mathrm{Re}}(x))^2+(N_{\mathrm{Im}}(x))^2]/[(D_{\mathrm{Re}}(x))^2+(D_{\mathrm{Im}}(x))^2]=\mathit{radius}^2\\ &= \text{function only of } r.\end{aligned} \tag{2}$$

In this case, the way to satisfy Equation (2) is to set $N_{\mathrm{Re}}(x)/D_{\mathrm{Re}}(x)=-N_{\mathrm{Im}}(x)/D_{\mathrm{Im}}(x)$.

This will let us find a and *radius*. Other ways to make the radius constant will produce circles on which both r and x vary.

CHAPTER

11 Oscillators

Oscillators are autonomous dc-to-ac converters. They are used as the frequency-determining elements of transmitters and receivers and as master clocks in computers, frequency synthesizers, wristwatches, etc. Their function is to divide time into regular intervals. The invention of mechanical oscillators (clocks) made it possible to divide time into intervals much smaller than the Earth's rotation period and much more regular than a human pulse rate. Electronic oscillators are analogs of mechanical clocks.

11.1 Negative feedback (relaxation) oscillators

The earliest clocks used a "verge and foliot" mechanism which resembled a torsional pendulum but was not a pendulum at all. These clocks operated as follows: torque derived from a weight or a wound spring was applied to a pivoted mass. The mass accelerated according to Torque $= I\,\mathrm{d}^2\theta/\mathrm{d}t^2$ (the angular version of $F = ma$). When θ reached a threshold, θ_0, the mechanism reversed the torque, causing the mass to accelerate in the opposite direction. When it reached $-\theta_0$ the torque reversed again, and so on. The period was a function of the moment of inertia of the mass, the magnitude of the torque, and the threshold setting. These clocks employed negative feedback; when the controlled variable had gone too far in either direction, the action was reversed. Most home heating systems are negative feedback oscillators; the temperature cycles between the turn on and turn off points of the thermostat. Negative feedback electronic oscillators are called "relaxation oscillators." Most of these circuits operate by charging a capacitor until its voltage reaches an upper threshold and then discharging it until the voltage reaches a lower threshold voltage. In Figure 11.1(a), when the voltage on the capacitor builds up to about 85 V, the neon bulb fires. The capacitor then discharges quickly through the ionized gas (relaxes) until the voltage decays to about 40 V. The bulb then extinguishes and the cycle begins anew.

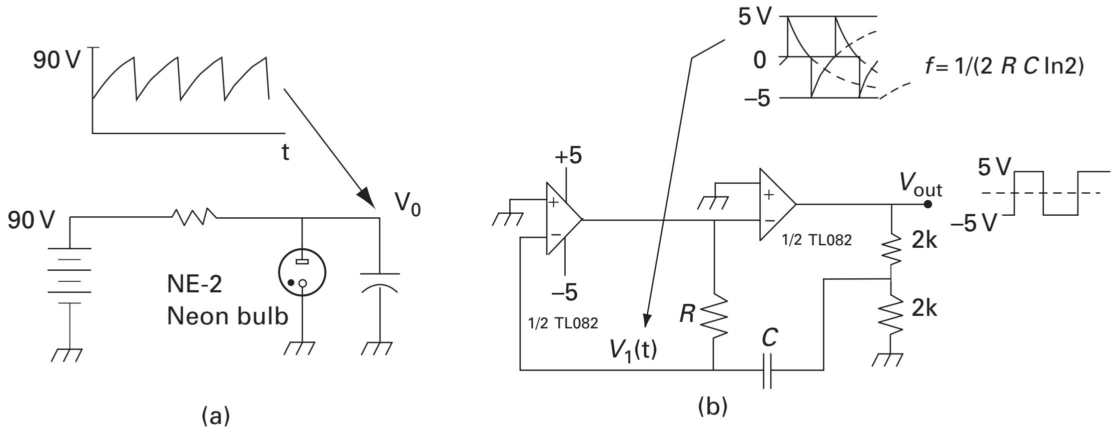

Figure 11.1. Relaxation (negative feedback) oscillators.

The circuit of Figure 11.1(b) alternately charges the capacitor, *C*, until its voltage reaches 2.5 V, and then discharges it until the voltage has fallen to −2.5 V. ($V_1(t)$ decays alternately toward +5 or −5 volts. When it reaches zero volts, the left-hand op-amp abruptly saturates in the opposite direction, kicking $V_1(t)$ to the voltage it had been approaching. The voltage then begins to decay in the opposite direction, and so forth.) Voltage-to-frequency converters are usually relaxation oscillators in which the control voltage determines the slope, and hence the oscillation period, of a fixed-amplitude sawtooth wave. Relaxation oscillators typically contain waveforms that are ramps or exponential decays. In the verge and foliot clock, the angle $\theta(t)$ consists of a sequence of parabolic arcs. Note that relaxation oscillators are nonlinear circuits which switch alternately between a charge mode and a discharge mode. Positive feedback oscillators, the main subject of this chapter, are nominally linear circuits. They generate sine waves.

11.2 Positive feedback oscillators

Clock makers improved frequency stability dramatically by using a true pendulum, a moving mass with a restoring force supplied by a hair spring or gravity.[1] As first observed by Galileo, a pendulum has its own natural frequency, independent of amplitude. It moves sinusoidally in simple harmonic motion. A pendulum clock uses *positive* feedback to push the pendulum in the direction of its motion, just as one pushes a swing to restore energy lost to friction.

[1] The Salisbury Cathedral clock, when installed around 1386, used a verge and foliot mechanism. Some 300 years later, after Christian Huygens invented the pendulum clock based on Galileo's observations of pendulum behavior, the Salisbury clock was converted into a positive feedback pendulum clock, its present form.

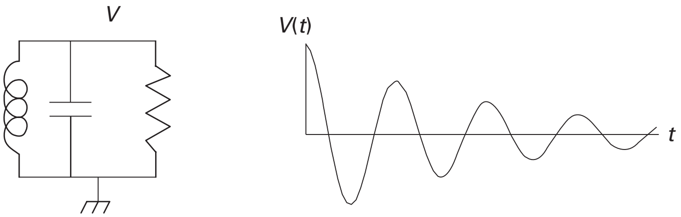

Figure 11.2. Damped oscillation in a parallel *LCR* circuit.

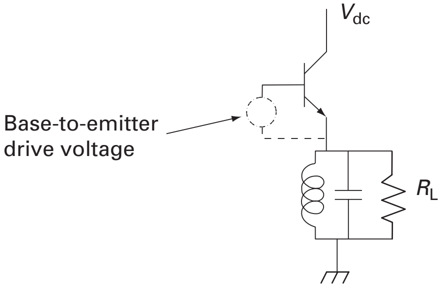

Figure 11.3. Transistor and dc supply replace energy lost to damping.

Electronic versions of the pendulum clock are usually based on resonators such as parallel or series *LC* circuits or electromechanical resonators such as quartz crystals. They use positive feedback to maintain the oscillation. A resonator with some initial energy (inductor current, capacitor charge, or mechanical kinetic energy) will oscillate sinusoidally with an exponentially decaying amplitude as shown in Figure 11.2. The decay is due to energy loss in the load and in the internal loss of the finite-Q resonator. In Figure 11.2 the resonator is a parallel *LCR* circuit.

To counteract the exponential decay, a circuit pumps current into the resonator when its voltage is positive and/or pulls current out when its voltage is negative. Figure 11.3 shows how a transistor and a dc supply can provide this energy. In this example circuit, the transistor is shown in the emitter-follower configuration simply because it is so easy to analyze; the emitter voltage tracks the base voltage and the base draws negligible current. The single transistor cannot supply negative current but we can set it up with a dc bias as a class-A amplifier so that current values less than the bias current are equivalent to negative current.

All that remains to complete this oscillator circuit is to provide the transistor drive, i.e., the base-to-emitter voltage. We want to increase the transistor's conduction when the output voltage (emitter voltage) increases, and we see that the emitter voltage has the correct polarity to be the drive signal. Since an emitter follower's voltage gain is slightly less than unity, the base needs a drive

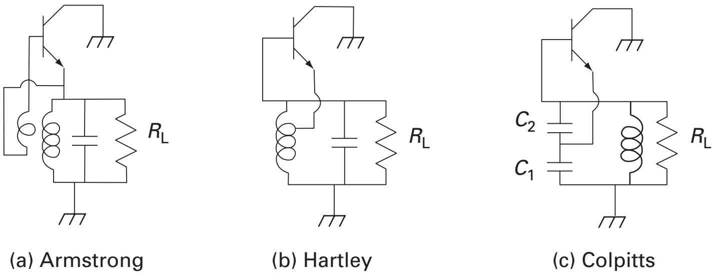

(a) Armstrong (b) Hartley (c) Colpitts

Figure 11.4. Feedback loop details define (a) Armstrong; (b) Hartley; (c) Colpitts oscillators.

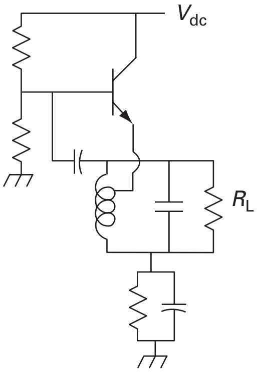

Figure 11.5. Hartley oscillator circuit including bias circuitry.

signal with slightly more amplitude than the sine wave on the emitter. Figure 11.4 shows three methods to provide this drive signal. The Armstrong oscillator adds a small secondary winding to the inductor. The voltage induced in the secondary adds to the emitter voltage. The Hartley oscillator accomplishes the same thing by connecting the emitter to a tap slightly below the top of the inductor. This is just an autotransformer version of the Armstrong oscillator if the magnetic flux links all the turns of the inductor. But the top and bottom portions of the inductor do not really have to be magnetically coupled at all; most of the current in the inductor(s) is from energy stored in the high-Q resonant circuit. This current is common to the two inductors so they essentially form a voltage divider. (Note, though, that the ratio of voltages on the top and bottom portions of the inductor ranges from the turns ratio, when they are fully coupled, to the square of the turns ratio, when they have no coupling.) If we consider the totally decoupled Hartley oscillator – no mutual inductance – and then replace the inductors by capacitors of equal (but opposite) reactance and replace the capacitor by an inductor, we get the Colpitts oscillator. Note that each oscillator in Figure 11.4 is an amplifier with a positive feedback loop. No power supply or biasing circuitry is shown in these figures; they simply indicate the ac signal paths.

Using the Hartley circuit as an example, Figure 11.5 shows a practical circuit. It includes the standard biasing arrangement to set the transistor's operating point. (A resistor voltage divider determines the base voltage and an emitter resistor then determines the emitter current, since V_{be} will be very close to 0.7 V.) A blocking capacitor allows the base to be dc biased with respect to the emitter. A bypass capacitor puts the bottom of the resonant circuit at RF ground.

In practice, one usually finds oscillators in grounded emitter circuits, as shown in Figure 11.6. The amplitude of the base drive signal must be much smaller than the sine wave on the resonant circuit. Moreover, the polarity of the base drive signal must be inverted with respect to the sine wave on the collector. You can inspect these circuits to see that they do satisfy these conditions. But, on closer inspection, you can note the circuits are *identical* to the circuits of Figure 11.4, except that the ground point has been moved from the collector to

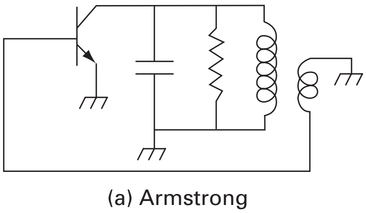

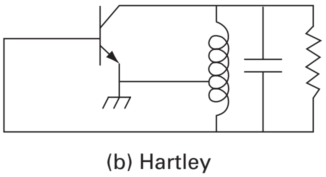

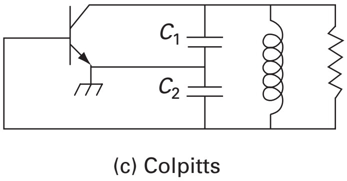

Figure 11.6. Grounded-emitter oscillator circuits.

the emitter. In these oscillators, an amplifier is enclosed in a positive feedback loop. But, because there is no input signal to have a terminal in common with the output signal, oscillators, unlike amplifiers, do not have common-emitter, common-collector, and common-base versions.

The Colpitts oscillator, needing no tap or secondary winding on the inductor, is the most commonly used circuit. Sometimes the transistor's parasitic collector-to-emitter capacitance is, by itself, the top capacitor, C_1, so this capacitor may appear to be missing in a circuit diagram. A practical design example for the Colpitts circuit of Figure 11.6(c) is presented later in this chapter.

11.2.1 Unintentional oscillators

In RF work it is common for a casually designed amplifier to break into oscillation. One way this happens is shown in Figure 11.7. The circuit is a basic common-emitter amplifier with parallel resonant circuits on the input and output (as bandpass filters and/or to cancel the input and output capacitances of the transistor). When the transistor's parasitic base-to-collector capacitance is included, the circuit has the topology of the decoupled Hartley oscillator. If the feedback is sufficient, it will oscillate. The frequency will be somewhat lower than that of the input and output circuits so that they look inductive as shown in the center figure. This circuit known as a TPTG oscillator, form Tuned-Plate Tuned-Grid, in the days of the vacuum tube.

Figure 11.7. Tuned amplifier as an oscillator

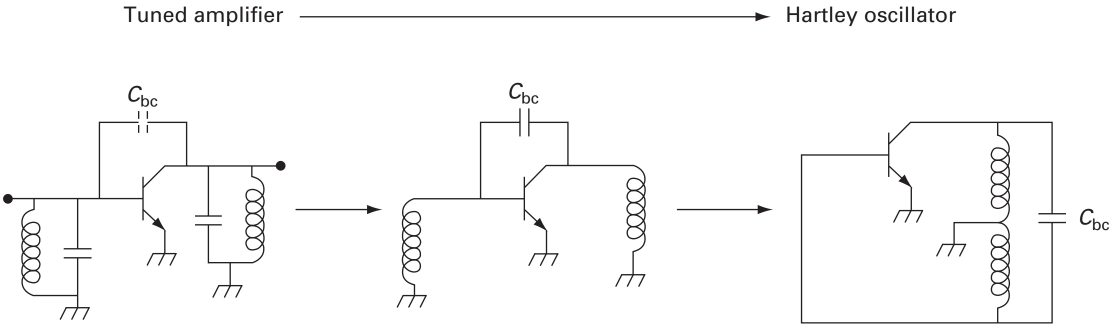

With luck, the loop gain of any amplifier will be less than unity at any frequency for which the total loop phase shift is 360° and an amplifier will be stable. If not, it can be *neutralized* to avoid oscillation. Two methods of neutralization are shown in Figure 11.8.

Figure 11.8. Amplifier neutralization.

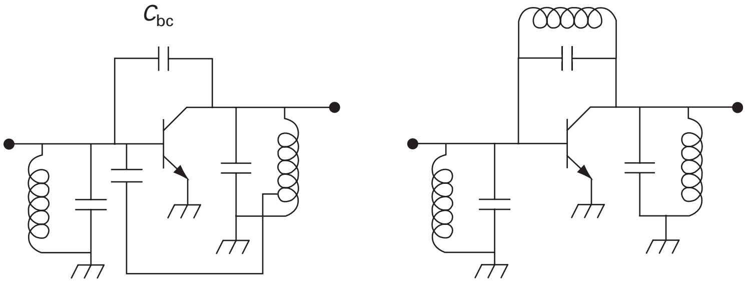

In Figure 11.8(a), a secondary winding is added to provide an out-of-phase voltage which is capacitor-coupled to the base to cancel the in-phase voltage coupled through C_{bc}. In Figure 11.8(b), an inductor from collector to base resonates C_{bc} to effectively remove it (a dc blocking capacitor would be placed in series with this inductor). In grounded-base transistor amplifiers and grounded-grid vacuum tube amplifiers the input circuit is shielded from the output circuit. These are stable without neutralization (but provide less power gain than their common-emitter and common-cathode-counterparts).

11.2.2 Series resonant oscillators

The oscillators discussed above were all derived from the parallel resonant circuit shown in Figure 11.2. We could just as well have started with a series *LCR* circuit. Like the open parallel circuit, a shorted series *LCR* circuit executes an exponentially damped oscillation unless we can replenish the dissipated energy. In this case we need to put a voltage source in the loop which will be positive when the current is positive and negative when the current is negative, as shown in Figure 11.9.

Figure 11.9. Series-mode oscillator operation.

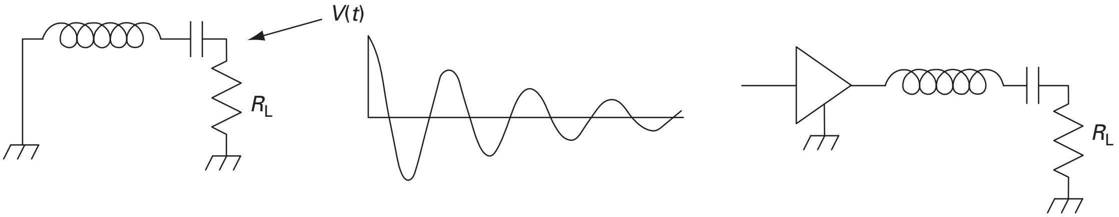

While a bare transistor with base-to-emitter voltage drive makes a good current source for a parallel-mode oscillator, a low-impedance voltage source is needed for a series-mode oscillator. In the series-mode oscillator shown in Figure 11.10, an op-amp with feedback is such a voltage source.

Figure 11.10. An op-amp series-mode oscillator.

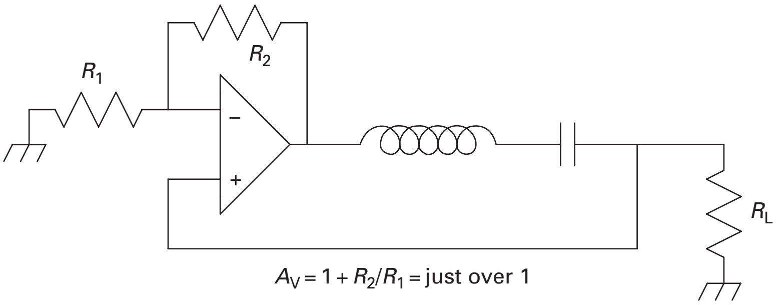

Since no phase inversion is provided by the tank circuit, the amplifier is connected to be noninverting. An emitter-follower has a low output impedance and can be used in a series-mode oscillator (see Problem 11.4). When the series *LC* circuit is replaced by a multisection RC network, the resulting oscillator is commonly known as a *phase-shift oscillator* (even though every feedback oscillator oscillates at the frequency at which the overall loop phase shift is 360°). An *RC* phase-shift oscillator circuit is shown in Figure 11.11. Op-amp voltages followers make the circuit easy to analyze.

For the three cascaded *RC* units, the transfer function is given by $V_2(t)/V_1(t) = 1/(\omega RC{+}1)^3$. The inverting amplifier at the left provides a voltage gain of $-16/2 = -8$, so $V_1(t)/V_2(t) = -8$. Combining these two equations yields a cubic equation with three roots: $\omega RC = 3\mathrm{j}$, $\sqrt{3}$, and $-\sqrt{3}$. The first root corresponds to an exponential decay of any initial charges on the capacitors while the two imaginary roots indicate that the circuit will produce a steady sine-wave oscillation whose frequency is given by $\omega RC = \sqrt{3}$. In practice, the 16k resistor would be increased to perhaps twice that value to ensure oscillation. Note that

Figure 11.11. An *RC* phase-shift oscillator.

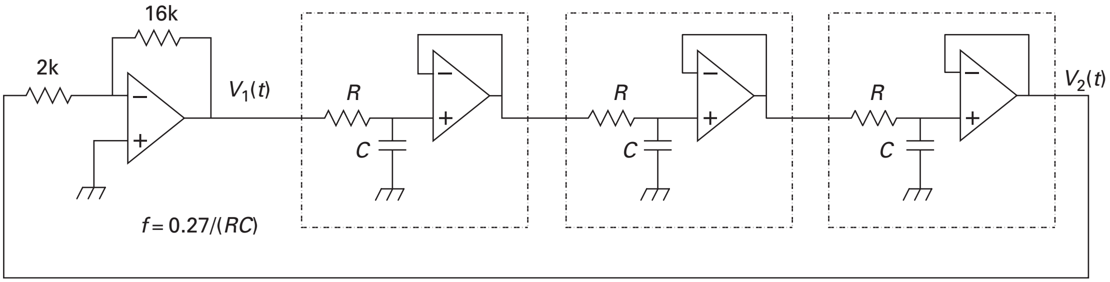

this circuit is a positive-feedback sine-wave oscillator even though it does not contain a resonator. When the 16k resistor value is increased, the loop gain for the original frequency becomes greater than unity, but for the new gain, there will be a nearby complex frequency, $\omega - \mathrm{j}\alpha$, for which the loop gain is unity. The time dependence therefore becomes $\mathrm{e}^{\mathrm{j}(\omega-\mathrm{j}\alpha)t} = \mathrm{e}^{\mathrm{j}\omega t}\mathrm{e}^{\alpha t}$, showing that the oscillation amplitude grows as $\mathrm{e}^{\alpha t}$. This circuit illustrates how any linear circuit with feedback will produce sine-wave oscillations if there is a (complex) frequency for which the overall loop gain is unity and the overall phase shift is 360°. (Of course α must be positive, or the oscillation dies out exponentially.)

11.2.3 Negative-resistance oscillators

In the circuits described above, a transistor provides current to an *RLC* circuit when the voltage on this circuit is positive, i.e., the transistor behaves as a negative resistance. But the transistor is a three-terminal device and the third terminal is provided with a drive signal derived from the *LCR* tank. Figure 11.12 shows how two transistors can be used to make a two-terminal negative resistance that is simply paralleled with the *LCR* tank to make a linear sine wave oscillator that has no feedback loop.

The two transistors form an emitter-coupled differential amplifier in which the resistor to $-V_{\mathrm{ee}}$ acts as a constant current source, supplying a bias current, I_0. The input to the amplifier is the base voltage of the right-hand transistor. The output is the collector current of the left-hand transistor. The ratio of input to output is $-4V_{\mathrm{T}}/I_0$, where V_{T} is the thermal voltage, 26 mV. This ratio is just the negative resistance, since the input and output are tied together. This negative-resistance oscillator uses a parallel-resonant circuit, but a series-resonant version is certainly possible as well.

Any circuit element or device that has a negative slope on at least some portion of its I–V curve can, in principle, be used as a negative resistance. Tunnel diodes can be used to build oscillators up into the microwave frequency range. At microwave frequencies, single-transistor negative-resistance oscillators are common. A plasma discharge exhibits negative resistance and provided a pre-vacuum tube method to generate coherent sine waves. High-efficiency

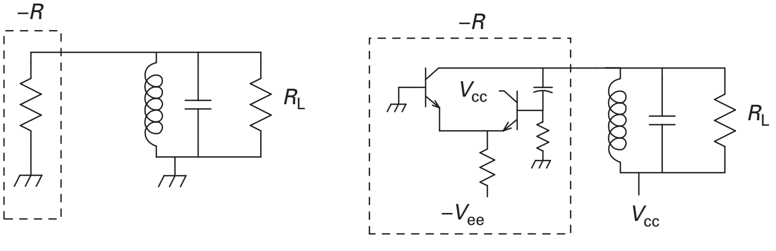

(a) (b)

Figure 11.12. A negative-resistance oscillator.

Poulsen arc transmitters, circa World War I, provided low-frequency RF power exceeding 100 kW.

11.3 Oscillator dynamics

These resonant oscillators are basically linear amplifiers with positive feedback. At turn-on they can get started by virtue of their own noise if they run class A. The tiny amount of noise power at the oscillation frequency will grow exponentially into the full-power sine wave. Once running, the signal level is ultimately limited by some nonlinearity. This could be a small-signal nonlinearity in the transistor characteristics. Otherwise, the finite voltage of the dc power provides a severe large-signal nonlinearity, and the operation will shift toward class-C conditions. The fact that amplitude cannot increase indefinitely shows that some nonlinearity is operative in every real oscillator. Any nonlinearity causes the transistor's low- frequency $1/f$ noise to mix with the RF signal, producing more noise close to the carrier than would exist for linear operation. An obvious way to mitigate large-signal nonlinearity is to detect the oscillator's output power and use the detector voltage in a negative feedback arrangement to control the gain. This can maintain an amplitude considerably lower than the power supply voltage. Alternatively, if the oscillator uses a device (transistor or op-amp circuit) with a soft saturation characteristic, the amplitude will reach a limit while the operation is still nearly linear. For example, the amplifier in the oscillator of Figure 11.10 might have a small cubic term, i.e., $V_{\mathrm{OUT}} = AV_{\mathrm{IN}} - BV_{\mathrm{IN}}^3$, where B/A is very small (see Problem 11.5).

11.4 Frequency stability

Long-term (seconds to years) frequency fluctuations are due to component aging and changes in ambient temperature and are called *drift*. Short-term fluctuations, known as oscillator noise, are caused by the noise produced in the active device, the finite loaded Q of the resonant circuit, and nonlinearity in the operating cycle. The higher the Q, the faster the loop phase-shift changes with frequency. Any disturbances (transistor fluctuations, power supply variations changing the transistor's parasitic capacitances, etc.) that tend to change the phase shift will cause the frequency to move slightly to reestablish the overall 360° shift. The higher the resonator Q, the smaller the frequency shift. Note that this is the *loaded Q*, so the most stable oscillators, besides having the highest Q resonators, are loaded as lightly as possible. In LC oscillators, losses in the inductor almost always determine the resonator Q. A shorted piece of transmission line is sometimes used as a high-Q inductor. Chapter 24 treats oscillator noise in detail.

11.5 Colpitts oscillator theory

Let us look in some detail at the operation of the Colpitts oscillator. Figure 11.13 shows the Colpitts oscillator of Figure 11.6(c) redrawn as a small-signal equivalent circuit (compare the figures). The still-to-be-biased transistor is represented as a voltage-controlled current source. The resistor r_{be} represents the small-signal base-to-emitter resistance of the transistor.

The parallel combination of L and the load resistor, R, is denoted as Z, i.e., $Z = \mathrm{j}\omega LR/(\mathrm{j}\omega L+R) = j\omega L_{\mathrm{S}} + R_{\mathrm{S}}$, where L_{S} and R_{S} are the component values for the equivalent series network. Likewise, it is convenient to denote ${r_{\mathrm{be}}}^{-1}$ as g. The voltage V_{be}, a phasor, is produced by the current I (a phasor) from the current source. This is a linear circuit, so V_{be} can be written as $V_{\mathrm{be}} = I\,Z_{\mathrm{T}}$, where Z_{T} is a function of ω. We will calculate this "transfer impedance" using standard circuit analysis. Since the current I is proportional to V_{be}, we can write an equation expressing that, in going around the loop, the voltage V_{be} exactly reproduces itself :

$$-g_{\mathrm{m}}V_{\mathrm{be}}Z_{\mathrm{T}} = V_{\mathrm{be}} \quad \text{or} \quad \frac{1}{Z_{\mathrm{T}}} = -g_{\mathrm{m}}. \tag{11.1}$$

This equation will let us find the component values needed for the circuit to oscillate at the desired frequency, i.e., the values that will make the loop gain equal to unity and the phase shift equal 360°.

We can arbitrarily select L, choosing an inductor whose Q is high at the desired frequency. Equation (11.1), really two equations (real and imaginary parts), will then provide values for C_1 and C_2. To derive an expression for Z_{T}, we will assume that $V_{\mathrm{be}} = 1$ and work backward to find the corresponding value of I. With this assumption, inspection of Figure 11.13 shows that the current I_1 is given by

$$I_1 = \mathrm{j}\omega C_2 + g. \tag{11.2}$$

Now the voltage V_{c} is just the 1 volt assumed for V_{be} plus I_1Z, the voltage developed across Z:

$$V_{\mathrm{c}} = 1 + (\mathrm{j}\omega C_2 + g)Z. \tag{11.3}$$

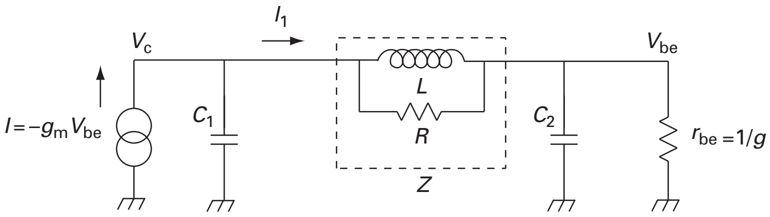

Figure 11.13. Colpitts oscillator small-signal equivalent circuit.

Finally, the current I is just the sum of I_1 plus $V_c \, j\omega C_1$, the current going into C_1:

$$I = (j\omega C_2 + g) + [1 + (j\omega C_2 + g)Z]\, j\omega C_1. \tag{11.4}$$

Since we had assumed that $V_{be} = 1$, we have $Z_T = 1/I$ or

$$\frac{1}{Z_T} = j\omega C_2 + g + [1 + (j\omega C_2 + g)Z]\, j\omega C_1. \tag{11.5}$$

Using this, the condition for oscillation, Equation (11.1) becomes

$$g_m + j\omega C_2 + g + [1 + (j\omega C_2 + g)Z]\, j\omega C_1 = 0. \tag{11.6}$$

The job now is to solve Equation (11.6) for C_1 and C_2. If we assume that ω is real i.e., that the oscillation neither grows nor decays, we find from the imaginary part of this equation, that

$$\frac{C_2 + C_1 + gR_S C_1}{L_S C_1 C_2} = \omega^2 \tag{11.7}$$

and, from the real part, that

$$\omega^2 C_1 C_2 R_S + g(\omega^2 L_S C_1 - 1) = g_m. \tag{11.8}$$

Solving Equations (11.7) and (11.8) simultaneously for C_2 and C_1 produces

$$C_2 = \frac{g_m L_S}{2R_S}\left(1 + \sqrt{\frac{4R_S(1 + gR_S)(g_m + g)}{{g_m}^2 L_S \omega^2}}\right) \tag{11.9}$$

and

$$C_1 = \frac{C_2}{\omega^2 L_S C_2 - 1 - gR_S}. \tag{11.10}$$

Normally C_2 will have a much larger value than C_1 and $\omega \approx 1/\sqrt{LC_1}$. Moreover, the second term in the square root of Equation (11.9) is usually much less than unity so $C_2 \approx g_m L_S / R_S$.

11.5.1 Colpitts oscillator design example

Let us design a practical grounded-emitter Colpitts oscillator. Suppose this oscillator is to supply 1 mW at 5 MHz and that it will be powered by a 6 V dc supply. Assuming full swing, the peak output sine wave voltage will be 6 V. The output power is given by $0.001\ \text{W} = (6\ \text{V})^2/(2R_L)$ so the value of the load resistor, R_L, will be 18 k ohms. Assuming class-A operation, the bias current in the transistor is made equal to the peak current in the load: $I = I_{pk} = 6\ \text{V}/18\ \text{k} = 0.33\ \text{mA}$. If we let the emitter biasing resistor be 1.5 k, the emitter bias voltage will be $1500 \times 0.33\ \text{mA} = 0.5\ \text{V}$. Assuming the typical 0.7 V offset between the base and emitter, the base voltage needs to be 1.2 V. A voltage divider using a 40 k resistor and a 10 k resistor will produce 1.2 V from the 6 V

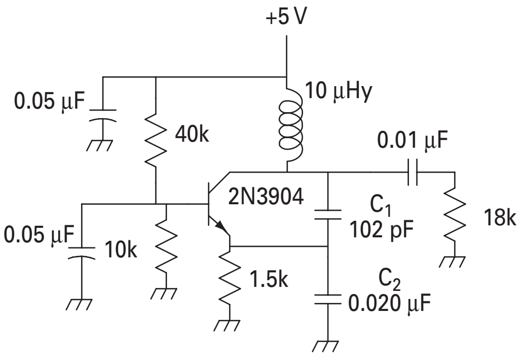

Figure 11.14. Colpitts oscillator: 5 MHz, 1 mW.

supply. These bias components are shown in the schematic diagram of Figure 11.14.

A 0.05 μF bypass capacitor pins the base to ac ground and another bypass capacitor ensures that the dc input is held at a firm RF ground. Note that the 1.5 k emitter bias resistor provides an unwanted signal path to ground. This path could be eliminated by putting an inductor in series with the bias resistor as an RF choke, but this is not really necessary; the1.5k resistor is in parallel with C_2, which will have such a low reactance that the resistor will divert almost no current from it.

With the biasing out of the way, we now deal with the signal components. The transconductance of the transistor is found by dividing the bias current, I_0, by 26 mV, the so-called thermal voltage,[2] i.e., $g_m = 0.33$ mA/ 26 mV = 0.013 mhos. The small-signal base-to-emitter resistance, r, is given by $r = \beta V_{\text{thermal}}/I_0$. For a typical small-signal transistor, such as a 2N3904, β is about 100, so r_{in} = 100·0.026 V/0.33 mA = 8000 ohms.

Using Equation (11.9) and (11.10), the values of C_1 and C_2 are 102 pF and 0.023 μF, respectively. These are the values for which the oscillator theoretically will maintain a constant amplitude. In practice, we increase the feedback by decreasing the value of C_2 to ensure oscillation. This produces a waveform that grows exponentially until it reaches a limit imposed by circuit nonlinearity. The frequency becomes complex, i.e., ω becomes $\omega - \mathrm{j}\alpha$ and the time dependence therefore becomes $\mathrm{e}^{\mathrm{j}(\omega - \mathrm{j}\alpha)t} = \mathrm{e}^{\mathrm{j}\omega t}\mathrm{e}^{\alpha t}$. Suppose we want α to be, say 10^5, which will cause oscillation to grow by a factor e every 10 μsec. (Fast growth would be important if, for example, the oscillator is to be rapidly pulsed on and off.) How do we find the value of C_2 to produce the desired α? To avoid doing more analysis, it is convenient to use a standard computer program such as Mathcad to find the root(s) of Equation (11.6) for trial values of C_2. In this example, if we decrease C_2 to 0.020 μF, we obtain the desired α.

---

[2] The thermal voltage is given by $V_{\text{thermal}} = 0.026\text{V} = kT/e$, where k is Boltzman's constant, T is the absolute ambient temperature, and e is the charge of an electron.

Problems

Problem 11.1. Draw a schematic diagram (without component values) for a bipolar transistor Colpitts oscillator with the collector at ground for both dc RF. Include the biasing circuit. The oscillator is to run from a positive dc supply.

Problem 11.2. Design (without specifying component values) a single-transistor series-mode oscillator based on the emitter follower circuit.

Problem 11.3. A simple computer simulation can illustrate how an oscillator builds up to an amplitude determined by the nonlinearity of its active element. The program shown below models the negative-resistance oscillator of Figure 11.12(a). The *LC* resonant frequency is 1 Hz. This network is in parallel with a negative-resistance element whose voltage vs. current relation is given by $I = -(1/R_n)*(V - \varepsilon V^3)$, to model the circuit of Figure 11.12. The small-signal (negative) resistance is just $-R_n$. The term $-\varepsilon V^3$ makes the resistance become less negative for large signals. The program integrates the second-order differential equation for $V(t)$ and plots the voltage versus time from an arbitrary initial condition, $V = 1$ volt.

Run this or an equivalent program. Change the value of the load resistor *R*. Find the minimum value of *R* for sustained oscillation. Experiment with the values of *R* and R_n. You will find that when the loaded *Q* of the *RLC* circuit is high, the oscillation will be sinusoidal even when the value of the negative resistor is only a fraction of *R*. When *Q* is low (as it is for $R = 1$), a low value of R_n such as $R_n = 0.2$ will produce a distinctly distorted waveform.

```
'QBasic simulation of negative-resistance oscillator of Figure 11.12a.
SCREEN 2
R = 1: L = 1 / 6.2832: C = L 'the parallel RLC circuit: 1 'ohms, 1/2pi henries, 1/2pi farads
RN = .9 'run program also with RN=.2 to see non-'sinusoidal waveform
E = .01 'negative resistance: I= (1/RN)*(V-EV^3)
V = 1: U = 0 'initial conditions, V is voltage, U is 'dV/dt
DT = .005 'step size in seconds
FOR I = 1 TO 3000
T = T + DT 'increment the time
VNEW = V + U * DT
U = U + (DT / C) * ((1 / RN) * (U - 3 * E * V * V * U) - V / L - U / R)
V = VNEW
PSET (40 * T, 100 + 5 * V) 'plot the point
NEXT I
```

Problem 11.4. In the oscillator shown below, the voltage gain of the amplifier decreases with amplitude. The voltage transfer function is $V_{\text{out}} = 2\,V_{\text{in}} - 0.5V_{\text{in}}^3$. This characteristic will limit the amplitude of the oscillation.

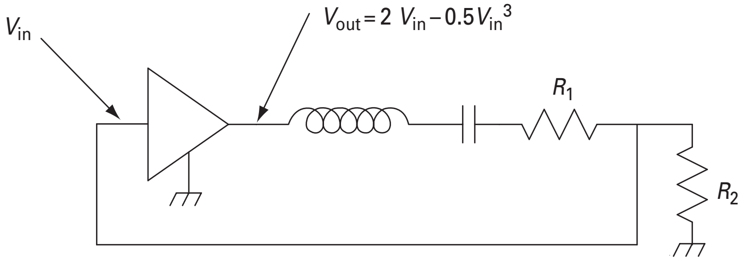

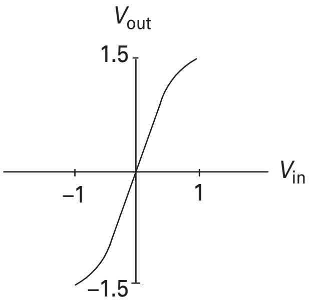

Find the ratio R_2/R_1 in order that the peak value of the sine wave V_{in} will be one volt. Hint: assume $V_{in} = \sin(\omega t)$. The amplifier output is then $2\ \sin(\omega t)\ -0.5\sin^3(\omega t)$. The second term resembles the sine wave but is more peaked. The *LC* filter will pass the fundamental Fourier component of this second term. Find this term and add it to $2\sin(\omega t)$. Then calculate the ratio R_2/R_1 so that the voltage divider output is $\sin(\omega t)$.

CHAPTER

12

Phase lock loops and synthesizers

Oscillators whose frequencies are derived from a stable reference source are used in transmitters and receivers as L.O.'s for accurate digital tuning. A VCO in a loop that tracks the frequency excursions of an incoming FM signal is a commonly used FM detector. The first widespread use of phase lock loops was in television receivers where they provide noise-resistant locking of the horizontal sweep frequency to the synchronization pulses in the signal.

12.1 Phase locking

A phase lock loop (PLL) circuit forces the phase of a voltage-controlled oscillator (VCO) to follow the phase of a reference signal. Once lock is achieved, i.e., once the phases stay close to each other, the frequency of the VCO will be equal to the frequency of the reference. In one class of applications, the PLL is used to generate a stable signal whose frequency is determined by an unstable (noisy) reference signal. Here the PLL is, in effect, a narrow bandpass filter that passes a carrier, while rejecting its noise sidebands. Examples include telemetry receivers that lock onto weak pilot signals from spacecraft and various "clock smoother" circuits. It is often necessary to lock an oscillator to the suppressed carrier of a modulated signal, an operation known as carrier recovery. In yet another class of applications the PLL is designed to detect all the phase fluctuations of the reference. An example is the PLL-based FM demodulator where the VCO reproduces the input signal, which is usually in the IF band. The voltage applied by the loop to the VCO is proportional to the instantaneous frequency (in as much as the VCO has a linear voltage-to-frequency characteristic), and this voltage is the audio output.

12.1.1 Phase adjustment by means of frequency control

In a PLL, the VCO frequency, rather than phase, is determined by the control voltage, but it is easy to see how frequency control provides phase adjustment.

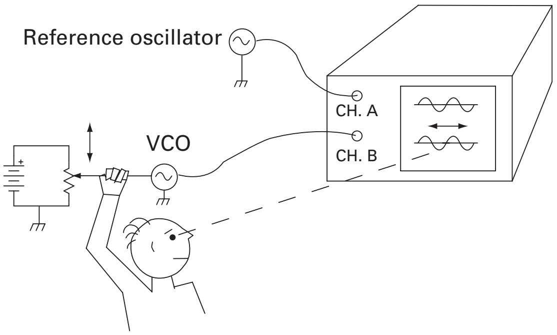

Figure 12.1. Phase lock loop concept.

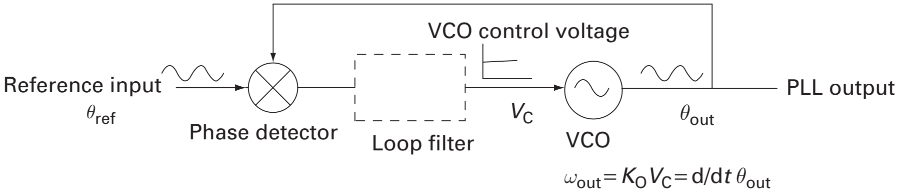

Figure 12.2. Phase lock loop block diagram.

Suppose you have two mechanical clocks. You want to make Clock *B* agree with Clock *A*. You notice that Clock *B* is consistently five minutes behind Clock *A*. You could exert direct phase control, using the time adjustment knob to set Clock *B* ahead 5 min to agree with Clock *A*. Or you could use frequency control, regulating the speed of Clock *B* to run somewhat faster. Once Clock *B* catches up with Clock *A*, you reset the frequency control to its original value. You may have done an electronic version of this in the lab (Figure 12.1). Suppose you have a two-channel oscilloscope. The first channel displays a sine wave from a fixed-frequency reference oscillator. The scope is synchronized to this reference oscillator so Channel *A* displays a motionless sine wave. Channel *B* is connected to a variable frequency oscillator which might be a laboratory instrument with a frequency knob or a VCO fitted with a potentiometer. Suppose the frequencies are the same. Then the sine wave seen on Channel *B* is also motionless. If you lower the VCO frequency slightly, the Channel *B* trace will drift to the right. And if you raise the VCO frequency, the trace will drift to the left. To align the two traces, you can shift the frequency slightly to let the Channel *B* trace drift into position, and then return the VCO frequency to the reference frequency value to stop the drift. The operator keeps the traces aligned and is therefore an element of this phase lock loop.

To automate the loop we use an electronic phase detector, i.e., an element that produces a voltage proportional to the phase difference between the reference and the VCO. Figure 12.2 shows the block diagram for a PLL.

If the VCO phase gets behind (lags) the reference phase, the phase detector will produce a positive "error voltage" which will speed up the VCO. As the phase error then decreases, the error voltage also decreases and the VCO slows back down. The exact way in which equilibrium is established, i.e., the mathematical form of the VCO phase, $\theta_{\text{out}}(t)$, depends on the yet-unspecified loop filter or "compensation network," shown as a dotted box in Figure 12.2. We will examine phase detectors and loop filters below.

12.1.2 Mechanical analog of a PLL

The PLL is a feedback control system – a servo. The electromechanical system shown in Figure 12.3 is a positioning servo. If an operator adjusts the input crank angle, the output shaft automatically turns so that the output angle, θ_{out}, tracks the input shaft angle, θ_{ref}. The gear teeth are in constant mesh but the top gear can slide axially. A dc motor provides torque to turn the output shaft. The input shaft might be connected to the steering wheel of a ship while the output shaft drives the rudder.

Consider the operation of this servo system. Assume for now that the voltage, V_{bias}, is set to zero, and that the system is at rest. When the input and output angles are equal, the slider on the potentiometer is centered and produces zero error voltage. But suppose the reference phase gets slightly ahead of the output phase – maybe the input crank is abruptly moved clockwise. During this motion, the threads on the input shaft have caused the top gear to slide to the left, producing a positive voltage at the output of the potentiometer. This voltage is passed on by the loop filter to the power amplifier and drives the motor clockwise, increasing θ_{out}. This rotates the top gear clockwise. The input shaft is now stationary, so the top gear moves along the threads to the right, reducing the potentiometer output voltage back to zero. The output angle (output phase) is again in agreement with the input or "reference" angle (input phase). In this system the gear/screw thread mechanism is the phase detector. Note that the

Figure 12.3. Positioning system with feedback control – a mechanical phase lock loop.

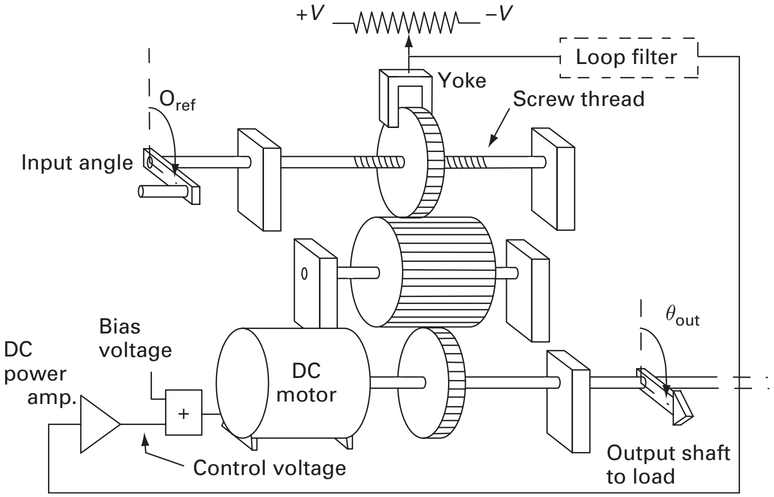

range of this phase detector can be many turns – the number of threads on the shaft. The negative feedback error signal provides the drive to force the (high-power) output to agree with the (low-power) input command.

We see how this mechanical system operates as a power-steering servo. To see that it can also operate as a PLL, suppose that an operator is cranking the input shaft at a constant or approximately constant rate (input frequency). We will now set the bias voltage so that, with the bias voltage alone, the motor turns at the nominal input frequency. (A VCO can be regarded as having an internal bias that determines its nominal frequency, i.e., its frequency when the control voltage is zero.) At equilibrium, the top gear is being turned by the middle gear but the threaded shaft is turning just as fast. The top gear therefore does not move along the threads, but stays at an equilibrium position. The output frequency is locked to the input frequency. How about the phases; will they also agree exactly? They did in the power steering model,[1] but now, with continuous rotation, they may not. The key is the loop filter. Suppose the loop filter is just a piece of wire (no filter at all). If the input frequency changes from its nominal value, there will have to be a constant phase error in order to keep the potentiometer off-center and produce a control voltage which will add to or subtract from the bias voltage. How much phase error (*tracking* error) is necessary depends on the gain of the power amplifier. Increasing the gain of the amplifier/motor and/or the gain (sensitivity) of the phase detector will reduce the error. But, if the loop filter contains an integrator, the control voltage will be the integral of the error voltage history. The steady-state error voltage can be zero. In fact, it *must* be zero, on average, or the integrator output, and hence the output frequency, would increase indefinitely.

Note: a servo's *type* is given by the number of integrators contained in its loop. Even without an integrator in the loop filter block, a PLL contains one integrator, the VCO, since phase is the time integral of frequency. In the mechanical analog, rotation angle is the integral of angular velocity.

12.1.3 Loop dynamics

Let us now look at how a PLL responds to disturbances in the input phase. We have already described qualitatively the response to a step function; the loop catches up with the reference. The way in which it catches up, i.e., quickly, slowly, with overshoot, or with no overshoot, depends on the loop filter and the characteristics of the phase detector, amplifier and motor. An exact analysis is straightforward if the entire system is *linear*, i.e., if the system can be described with linear differential equations. In this case it can be analyzed with the standard techniques applied to linear electronic circuits, i.e., complex numbers, Fourier and Laplace transforms, superposition, etc. In the PLL, the loop filter is linear since it consists of passive components: resistors, capacitors, and op-amps. The VCO is linear if its frequency is strictly proportional to the control voltage. Of course, over a small operating region, any smooth voltage–frequency characteristic is approximately linear. The most commonly used phase detector is a simple multiplier (mixer). You can show[2] that it produces an error voltage proportional to $\sin(\theta_{\text{out}}(t)-\theta_{\text{ref}}(t)+\pi/2)$. For small x, $\sin(x) \approx x$, so a multiplier is a fairly linear phase detector over a restricted region around $\pi/2$. The phase detector in our mechanical analog is linear over a range limited only by the length of the screw thread.

---

[1] We will assume the motor speed is proportional to the applied voltage, independent of load.

To describe the loop dynamics, we will find its response to an input phase disturbance, $\theta_{\text{dist\_in}} = e^{j\omega t}$. The complete input function is therefore $\theta_{\text{ref}}(t) = \omega_0 t + e^{j\omega t}$. Since the system is linear, we invoke superposition to note that the output will consist of a response to the $\omega_0 t$ term plus an output response to the $e^{j\omega t}$ disturbance term. The input term, $\omega_0 t$, results in an output contribution $\omega_0 t + \Theta$, where the constant Θ might be 90°, if the phase detector is a mixer, and might include another constant, if the nominal frequency of the VCO is not quite ω_0 and the loop filter includes no integrator. Since the input disturbance function is sinusoidal, the output response to it will also be sinusoidal, $A(\omega)e^{j\omega t}$, where the complex amplitude $A(\omega)$ is known as the input-to-output transfer function. Before we can calculate the transfer function, we must specify the loop filter.

12.1.4 Loop filter

A standard filter for type-II loops uses an op-amp to produce a weighted sum of the phase detector output plus the integral of the phase detector output. Figure 12.4 shows the circuit commonly used.

Note that if we let C go to infinity and set $R_2 = R_1$, the response of this filter is just unity (except for a minus sign) so this circuit will serve to analyze the type-I PLL as well as the type-II PLL. Remembering that the negative input of the op-amp is a virtual ground, we can write the transfer function of this filter.

$$\frac{-V_2}{V_1} = \frac{R_2 + 1/j\omega C}{R_1} = \frac{R_2}{R_1} + \frac{1}{j\omega C R_1} = \frac{\tau_2}{\tau_1} - \frac{j}{\omega\tau_1}. \tag{12.1}$$

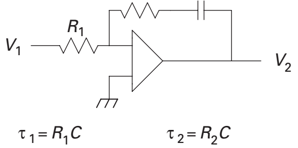

Figure 12.4. Loop filter circuit.

---

[2] Just expand the product $\cos(\omega_0 t + \theta_{\text{out}}(t))\cos(\omega_0 t + \theta_{\text{ref}}(t))$ and ignore (filter off) components at $2\omega_0$.

In the time domain, the relation between V_2 and V_1 is given by

$$\frac{dV_2}{dt} = \frac{-R_2}{R_1}\frac{dV_1}{dt} - \frac{V_1}{R_1 C}. \tag{12.2}$$

12.1.5 Linear analysis of the PLL

Figure 12.5 is a block diagram of the loop, including the loop filter of Figure 12.4.

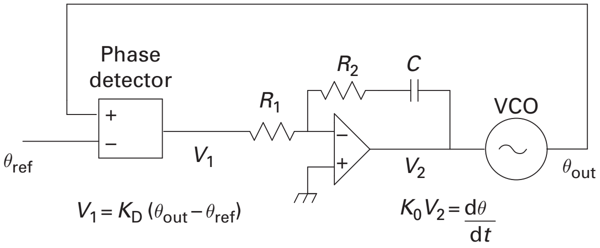
Phase detector
+
−
R_1
R_2
C
VCO
θ_{ref}
V_1
V_2
θ_{out}
$V_1 = K_D\,(\theta_{out} - \theta_{ref})$
$K_0 V_2 = \frac{d\theta}{dt}$

Figure 12.5. Type-II, second-order PLL block diagram.

For this linear system, let us find the frequency response of the output phase to a sinusoidal disturbance of the reference phase, $\theta_{dist}e^{j\omega t}$, where ω is the disturbing frequency. From inspection of Figure 12.5 we can write

$$-K_D(\theta_{out} - \theta_{dist})\left(\frac{\tau_2}{\tau_1} - \frac{j}{\omega\tau_1}\right)K_O = j\omega\theta_{out}. \tag{12.3}$$

Solving for θ, we have the frequency response

$$\frac{\theta_{out}(\omega)}{\theta_{dist}(\omega)} = \frac{1}{1 - \left[K_D K_O/(\omega^2\tau_1) + jK_D K_O \tau_2/(\omega\tau_1)\right]^{-1}}. \tag{12.4}$$

12.1.5.1 Frequency response of the type-I loop

We will look at the frequency response for the type-I loop by letting C go to infinity and letting $R_1 = R_2$, effectively eliminating the loop filter. The frequency response, Equation (12.4), becomes

$$\frac{\theta_{out}(\omega)}{\theta_{dist}(\omega)} \rightarrow \frac{1}{1 + j\omega/K}, \tag{12.5}$$

where $K = K_D K_O$, and is called the *loop gain*. The frequency response is identical to that of a simple RC lowpass filter with a time constant $RC = 1/K$. We saw earlier that a type-I PLL (a PLL with no integrator in the loop filter) needs high loop gain to have good tracking accuracy (the ability to follow a changing reference frequency without incurring a large phase error). Here we see that high gain implies a large bandwidth; the type-I PLL cannot have both high gain (for good tracking) and a narrow bandwidth (to filter out high-frequency reference phase noise).

12.1.5.2 Frequency response of the type-II loop

To deal with the type-II loop (the most common PLL) it is standard to define two constants, the *natural frequency*, ω_n, and the *damping coefficient*, ζ, as follows:

$$\omega_\text{n} = \sqrt{\frac{K_\text{O}K_\text{D}}{\tau_1}} \quad \zeta = \frac{\tau_2\omega_\text{n}}{2}. \tag{12.6}$$

We will see soon that these names relate to the transient response of the loop but, for now, we will substitute them into Equation (12.4), the expression for the frequency response, to get

$$\frac{\theta_\text{out}(\omega)}{\theta_\text{dist}(\omega)} = \frac{{\omega_\text{n}}^2 + 2\text{j}\omega\zeta\omega_\text{n}}{{\omega_\text{n}}^2 + 2\text{j}\omega\zeta\omega_\text{n} - \omega^2}. \tag{12.7}$$

This transfer function is plotted vs. ω/ω_n in Figure 12.6 for several different damping coefficients.

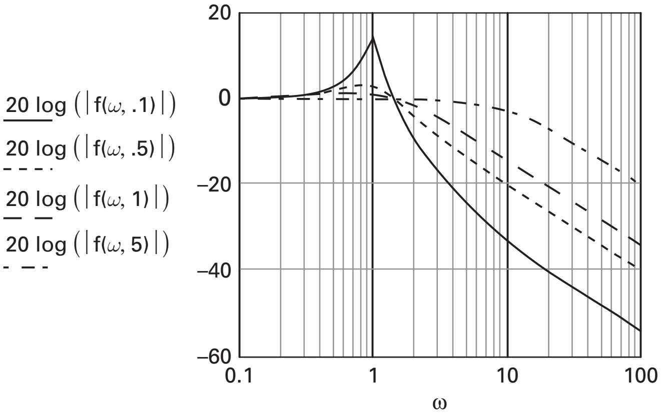

Figure 12.6. Frequency response vs. ω/ω_n for a second-order type-II loop with various damping coefficients.

12.1.5.3 Transient response

The transient response can be determined from the frequency response by using Fourier or Laplace transforms, but this loop is simple enough that we can work directly in the time domain. Equation (12.2) gives the time response for the loop filter. We assume the loop has been disturbed but the reference is now constant. By inspection of Figure 12.5 we can write

$$\frac{\text{d}V_1}{\text{d}t} = K_\text{D}K_\text{O}V_2 = KV_2. \tag{12.8}$$

Combining this with Equation (12.2) we get

$$\frac{\text{d}^2V_1}{\text{d}t^2} + K\frac{\tau_2}{\tau_1}\frac{\text{d}V_1}{\text{d}t} + \frac{K}{\tau_1}V_1 = 0 \quad \text{or} \quad \frac{\text{d}^2V_1}{\text{d}t^2} + 2\omega_\text{n}\zeta\frac{\text{d}V_1}{\text{d}t} + {\omega_\text{n}}^2V_1 = 0. \tag{12.9}$$

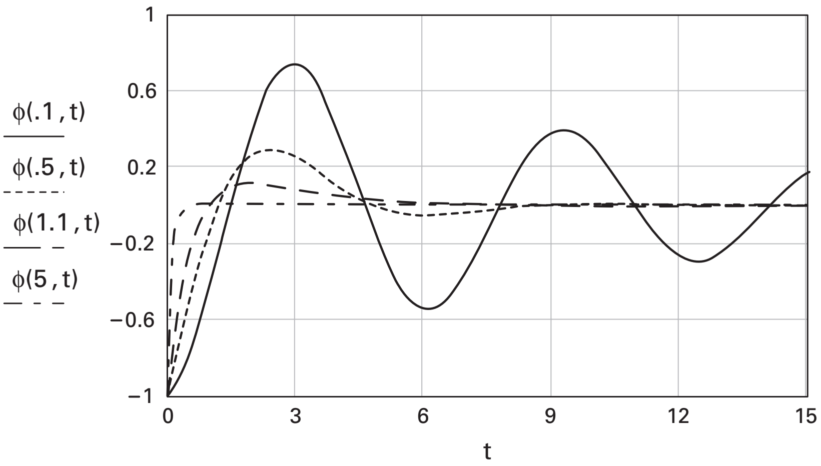

Figure 12.7. Phase error for a step of −1 radian in the reference phase at $t = 0$.

Assuming a solution of the form $e^{j\omega t}$, we find an equation for ω:

$$-\omega^2 + 2j\omega_n\zeta\omega + \omega_n^2 = 0. \tag{12.10}$$

The roots of this equation are

$$\omega = -\omega_n\zeta \pm \omega_n\sqrt{\zeta^2 - 1}. \tag{12.11}$$

When ζ is less than unity, the two solutions for ω are complex conjugates, giving sine and cosine oscillations with damped exponential envelopes. This is the underdamped situation. When ζ is greater than unity, the two solutions for ω are both damped exponentials, but with different time constants. If ζ is exactly unity, the two solutions are $e^{-\omega t}$ and $te^{-\omega t}$. In every case, a linear combination of the two appropriate solutions can match any given set of initial conditions, i.e., the $t = 0$ values of V_1 and $\mathrm{d}/\mathrm{d}t\, V_1$. Figure 12.7 shows the transient phase response to a step function in the reference phase of one radian at $t = 0$. This initial condition, $V_1(0) = -1$, determines the other initial condition, $\mathrm{d}V_1/\mathrm{d}t = 2\zeta\omega_n$, through the action of the loop filter circuit. Curves are shown for the recovery from this transient for damping coefficients of 0.1, 0.5, 1.1 and 5. In each case, $\omega_n = 1$. The underdamped cases show the trademark oscillatory behavior. When the damping is 5, the fast exponential recovery is essentially that of a wideband type-I loop.

Note: Equation (12.10) (which is the same as the denominator of the transfer function) is known as the *characteristic equation* of the system. Here it is a second-order equation and this type-II loop is therefore also a second-order loop. Equivalently, the order of a given system is the order of the differential equation describing its transient response.

12.1.6 A multiplier as a phase detector

A multiplier (mixer) is the most commonly used phase detector, especially at higher frequencies. When the reference sine wave is multiplied by the VCO sine wave, the usual sum and difference frequencies are generated. The desired baseband output is the difference component. If the VCO and reference frequencies are equal, this baseband component will be a dc voltage, zero when the VCO and reference phases differ by 90°, and linearly proportional to the phase difference in the vicinity of 90°. Therefore, when a multiplier is used as a phase detector, the loop will lock with the output phase 90° away from the reference phase. If this is a problem (in most applications it is not), a 90° phase shift network can be put at one of the multiplier inputs or in the output line. Note that a multiplier phase detector puts out zero volts when the phase shifts are different by 90° in either direction. One of these is a point of metastable equilibrium. As soon as it "falls off," the feedback will be positive, rather than negative, and the loop will rush to the stable equilibrium point.

As a phase detector, a mixer has a linear response of about $\pm \pi/4$, where $\sin(x) \approx x$. For somewhat larger phase differences, the sine curve flattens out and K_D begins to decrease. The system becomes nonlinear and the techniques of linear analysis fail. What is more important, if $\theta_{out} - \theta_{dist}$ needs to increase beyond $\pm \pi/2$, the $\sin(\theta_{out} - \theta_{dist})$ response of the mixer produces alternating positive and negative error signals, which will push the VCO frequency one way and then the other, and the loop will become unlocked. This is in contrast to the mechanical analog of Figure 12.3, whose phase detector range is 2π times the number of threads on the phase detector shaft. Electronic *phase/frequency detectors* (PFDs) are digital circuits that operate as linear phase detectors while the phase difference remains small, but then produce an output voltage proportional to the frequency difference, driving the loop in the correct direction to achieve lock. How a simple mixer-based PLL ever manages to lock without assistance is discussed below.

12.1.7 Frequency range and stability

The type-II loop can operate over the full range of the VCO since its integrator can build up as much bias as needed. The type-I loop, if $K_D K_O$ is small and if the phase detector has a limited range, may not be able to track over the full range of the VCO. Both the loops discussed above are unconditionally stable since the transient responses are decaying exponentials for any combination of the parameters K_D, K_O, R_1, R_2, and C. When you build one and it oscillates it is usually because you have high-frequency poles you did not consider (maybe in your op-amp or circuit parasitics) and you have actually built a higher-order loop.

12.1.8 Acquisition time

A high-gain type-I loop can achieve lock very quickly. A type-II loop with a small bandwidth can be very slow. Acquisition depends on some of the beat note from the phase detector getting through the filter to FM modulate the VCO. That beat note puts a pair of small sidebands on the VCO output. One of these sidebands will be at the reference frequency and will mix with the reference to produce dc with the correct polarity to push the VCO toward the reference frequency. The integrator gradually builds up this dc until the beat frequency comes within the loop bandwidth. From that point the acquisition is very fast. This reasoning, applied to the type-II loop, predicts an acquisition time of about $4f^2/B^3$ seconds where f is the initial frequency error and B is the bandwidth of the loop (see Problem 12.6). If the bandwidth is 10 Hz and the initial frequency error is 1 kHz, the predicted time is about one hour. (In practice, unavoidable offsets in the op-amp would make the integrator drift to one of the power supply rails and never come loose.) Obviously in such a case some assist is needed for a lockup. One common method is to add search capability, i.e., circuitry that causes the VCO voltage to sweep up and down until some significant dc component comes out of the phase detector. At that point the search circuit turns off and the loop locks itself up. The integrator can form part of the sweep circuit. Another way to aid acquisition is to use a frequency/phase detector instead of a simple phase detector. This is a digital phase detector which, when the input frequencies are different, puts out dc whose polarity indicates whether the VCO is above or below the reference. This dc quickly pumps the integrator up or down to the correct voltage for lock to occur. You will see these circuits described in Motorola literature. In digitally controlled PLLs, it is common for a microprocessor with a lookup table in ROM to pretune the VCO to the commanded frequency. This pretuning allows the loop to acquire lock quickly.

12.1.9 PLL receiver

There are many inventive circuits and applications in the literature on PLLs. Here is an example: deep space probes usually provide very weak telemetry signals. The receiver bandwidth must be made very small to reject noise that is outside the narrow band containing the modulated signal. A PLL can be used to do the narrowband filtering and to do the detection as well. One simple modulation scheme uses slow-frequency shift keying (FSK) where the signal is at one frequency for a “0” and then slides over to a second frequency for a “1.” The loop bandwidth of the PLL is made just wide enough to follow the keying. The control voltage on the VCO is used as the detected signal output. (With a wide loop bandwidth, such a circuit can demodulate ordinary FM audio broadcast signals.) In this narrowband example, the PLL circuit has the advantage that it will track the signal automatically when the transmitter frequency drifts or is Doppler shifted. Another modulation scheme uses phase shift keying (PSK). Suppose the transmitter phase is shifted 180° to distinguish a “1” from a “0.” This would confuse the PLL receiver because, for random data, the average output from the phase detector would be zero and the PLL would be unable to lock. A simple cure is to use the “doubling loop” shown in Figure 12.8. The incoming signal is first sent through a doubler to produce an output that looks like the output of a full-wave rectifier. (The waveform loops are positive no

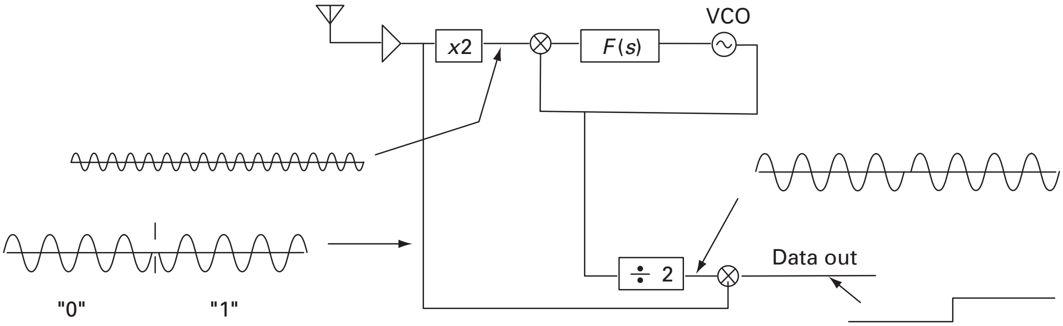

Figure 12.8. Doubling loop for detection of binary phase-shift modulation.

matter what the modulation does.) The VCO output frequency is then divided by two to produce a phase-locked reference. This reference is used as one input of a mixer. The other input of the mixer is the raw signal and the output is the recovered modulation.

12.2 Frequency synthesizers

By *frequency synthesizer* we usually mean a signal generator which can be switched to put out any one of a discrete set of frequencies and whose frequency stability is derived from a standard oscillator, either a built-in crystal oscillator or an external "station standard." Most laboratory synthesizers generate sine waves but some low-frequency synthesized function generators also generate square and triangular waves. General-purpose synthesizers have high resolution; the step between frequencies is usually less than 1 Hz and may be millihertz or even microhertz. Many television receivers and communications receivers have synthesized local oscillators. Special-purpose synthesizers may generate only a single frequency.

At least three general techniques are used for frequency synthesis. *Direct synthesizers* use frequency multipliers, frequency dividers, and mixers. *Indirect synthesizers* use phase lock loops. *Direct digital synthesizers* use a digital accumulator to produce a staircase sawtooth. A lookup table then changes the sawtooth to a staircase sinusoid, and a D-to-A converter provides the analog output. Some designs combine these three techniques.

12.2.1 Direct synthesis

The building blocks for direct synthesis are already familiar. Frequency multiplication can be done with almost any nonlinear element. A limiting amplifier (limiter) or a diode clipper circuit will convert sine waves into square waves, which include all the odd harmonics of the fundamental frequency. A delta function pulse train (in practice, a train of narrow pulses) contains all harmonics. When frequencies in only a narrow range are to be multiplied, class-C amplifiers can be used. (A child's swing does not need a push on each and every

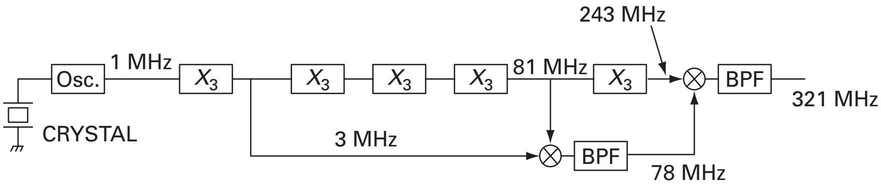
243 MHz
1 MHz
Osc.
X_3
X_3
X_3
X_3
81 MHz
X_3
BPF
321 MHz
CRYSTAL
3 MHz
BPF
78 MHz

Figure 12.9. A direct synthesizer that produces 321 MHz from a 1-MHz reference.

cycle.) Frequency division, used in all three types of synthesizers, is almost always done digitally; the input frequency is used as the clock for a digital counter made of flip-flops and logic gates. Frequency translation can be done with any of the standard mixer circuits.

Single-frequency synthesizers are usually ad hoc designs; the arrangement of mixers, multipliers, and dividers depends on the ratio of the desired frequency and the reference frequency. As an example of direct synthesis, the circuit of Figure 12.9 generates 321 MHz from a 1-MHz standard. A prime factor of 321 is 107. It would be difficult to build a times-107 multiplier. This design uses only triplers and mixers. Laboratory instruments that cover an entire range of frequencies must, of course, use some general scheme of operation.

12.2.2 Mix and divide direct synthesis

Some laboratory synthesizers use an interesting mix and divide module. An *n*-digit synthesizer would use *n* identical modules. An example is shown in Figure 12.10. Each module has access to a 16-MHz source and ten other

Figure 12.10. A mix-and-divide direct synthesizer.

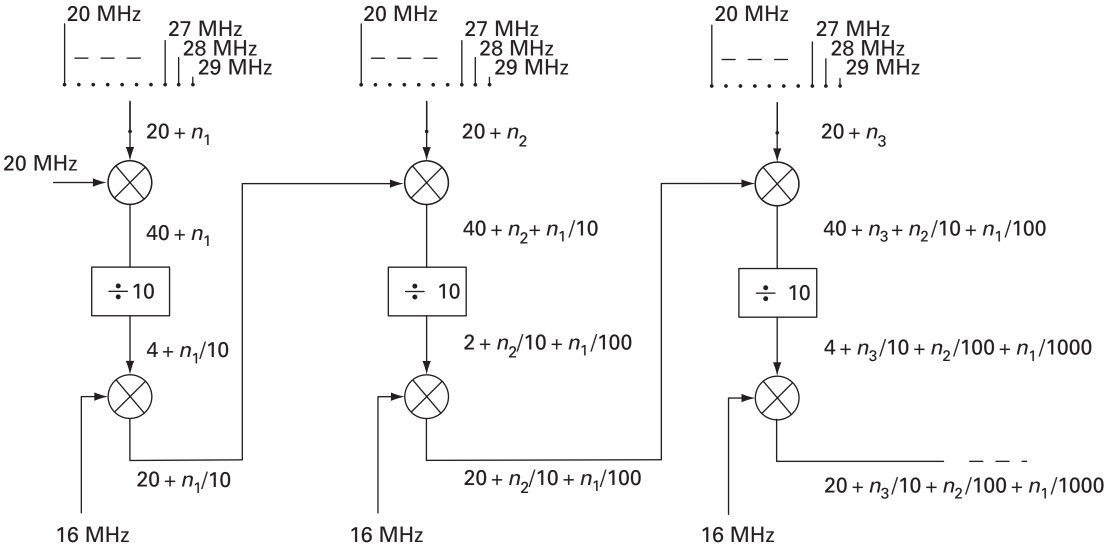
20 MHz
27 MHz
28 MHz
29 MHz
$20 + n_1$
20 MHz
$40 + n_1$
÷10
$4 + n_1/10$
$20 + n_1/10$
16 MHz
20 MHz
27 MHz
28 MHz
29 MHz
$20 + n_2$
$40 + n_2 + n_1/10$
÷ 10
$2 + n_2/10 + n_1/100$
$20 + n_2/10 + n_1/100$
16 MHz
20 MHz
27 MHz
28 MHz
29 MHz
$20 + n_3$
$40 + n_3 + n_2/10 + n_1/100$
÷ 10
$4 + n_3/10 + n_2/100 + n_1/1000$
$20 + n_3/10 + n_2/100 + n_1/1000$
16 MHz

reference sources from 20 to 29 MHz. (These references are derived from an internal or external standard, often at 5 MHz). In this kind of design the internal frequencies must be chosen carefully so that after each mixer the undesired sideband can be filtered out easily.

12.2.3 Indirect synthesis

The phase lock loop circuit of Figure 12.11 is an indirect synthesizer to generate the frequency Nf_{std}/M where N and M are integers. If the $\div N$ and/or the $\div M$ blocks are variable modulus counters, the synthesizer frequency is adjustable.

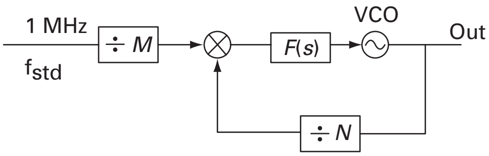

Figure 12.11. Basic indirect (PLL) synthesizer.

The 321-MHz synthesizer of Figure 12.9 could be built as the indirect synthesizer of Figure 12.11 by using a divide-by-321 counter (most likely a divide-by-3 counter followed by a divide-by-107 counter). Circuits like that of Figure 12.11 are used as local oscillators for digitally tuned radio and television receivers. For an AM broadcast band receiver, the frequency steps would be 10 kHz (the spacing between assigned frequencies). This requires that the phase detector reference frequency, f_{std}/M, be 10 kHz so only the modulus N would be adjustable. What about a synthesized local oscillator for a short-wave radio? We would probably want a tuning resolution of, say, 10 Hz. The reference frequency in the simple circuit of Figure 12.11 would have to be 10 Hz and the loop bandwidth would have to be about 2 Hz. This low bandwidth would make fast switching difficult. Moreover, with such a narrow loop bandwidth, the close-in noise of the VCO would not be cleaned up by the loop and the performance of the radio would suffer. For this application a more sophisticated circuit is needed. One method is to synthesize a VHF or UHF frequency with steps of 1 or 10 kHz, divide it to produce the necessary smaller steps, and then mix it to a higher frequency. Other circuits use complicated multiple loops. The newest receivers use L.O. synthesizers based on the principle of direct digital synthesis.

12.2.4 Direct digital synthesis (DDS)

This technique, illustrated in Figure 12.12, uses an adder with two n-bit inputs, A and B, together with an n-bit register to accumulate phase. The output of the adder is latched into the register on every cycle of a high-frequency clock. The inputs to the adder are the current register contents (the phase) and an adjustable addend (the phase increment). On every clock cycle, the accumulated phase increases by an amount given by the addend. Since the accumulator has a finite

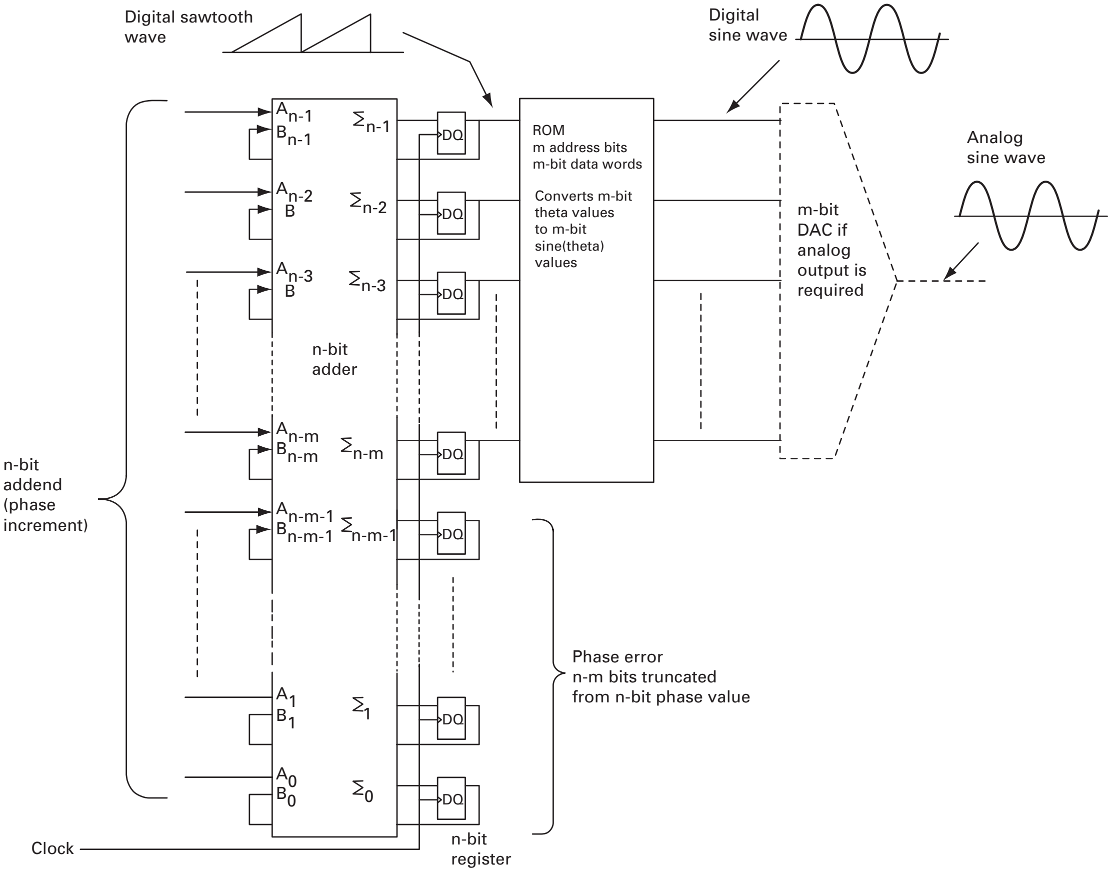
Digital sawtooth wave
Digital sine wave
Analog sine wave
A_{n-1} B_{n-1} Σ_{n-1}
A_{n-2} B Σ_{n-2}
A_{n-3} B Σ_{n-3}
n-bit adder
A_{n-m} B_{n-m} Σ_{n-m}
A_{n-m-1} B_{n-m-1} Σ_{n-m-1}
A_1 B_1 Σ_1
A_0 B_0 Σ_0
DQ
ROM
m address bits
m-bit data words
Converts m-bit theta values to m-bit sine(theta) values
m-bit DAC if analog output is required
n-bit addend (phase increment)
Phase error n-m bits truncated from n-bit phase value
Clock
n-bit register

Figure 12.12. Direct digital synthesizer (DDS).

length, it rolls over, like an automobile odometer, and the sequence of its output values forms a sawtooth wave. The maximum addend value is 2^{n-1}, where n is the number of bits in the adder/accumulator. Therefore, the output frequency of the DDS extends up to one-half the clock frequency. At this maximum frequency, the MSB toggles on every clock pulse and the lower bits remain constant. The output of the accumulator is used to address a ROM (read-only memory) that converts the stairstep digital sawtooth into a stairstep digital sine wave.

Let us consider the operation of the DDS in some detail. Suppose the selected addend is the integer A. Let θ_i denote the number held in the register after the i-th clock pulse. Because the accumulator rolls over when the output value would have reached or exceeded 2^n, we can write $\theta_i = (\theta_{i-1}+A) \bmod 2^n$, a formula which lets us easily simulate the DDS. On average, it will take $2^n/A$ clock pulses to fill the accumulator,[3] so the average period of the sawtooth wave will be $f_{\text{clk}}^{-1}\ 2^n/A$ and the output frequency will be $f_{\text{clk}}\ A/2^n$. This frequency can be changed in very fine increments if the number of bits in the accumulator is large. With a 32-bit accumulator, for example, the frequency resolution will be $f_{\text{clock}}/2^{32}$. For a clock frequency of 100 MHz the resolution will be $10^8/2^{32} = 0.023$ Hz.

---

[3] Consider that, after every 2^nA clock pulses, the accumulator must return to the same value. During that period, the value that would be accumulated without rollover is $A(2^nA)$. Thus the number of rollovers would be A^2. The average time per rollover is therefore $f_{\text{clk}}^{-1}\ (2^nA)/A^2 = f_{\text{clk}}^{-1}\ 2^n/A$.

One might think that, since the most significant bit (MSB) of the accumulator toggles at the desired output frequency, we could use it alone, simply filtering away its harmonics to obtain a sine wave at the desired frequency. However, while the MSB has the right *average* frequency it can be quite irregular. If the addend is a power of 2, the MSB will toggle at uniform time intervals but otherwise the MSB will have a jitter which depends on the value of the addend. A sine wave made from just the MSB would, therefore, contain too much phase noise (FM noise) for most RF applications.

The solution is to use more than just the top bit. If we use the top m bits, then the phase error is never greater than $360/2^m$ degrees. A table-lookup ROM uses these m-bit phase values, θm_i, as address bits to produce an output sequence, $\sin(\theta m_i)$, which has the desired wave shape (except for the roundoff error), together with low phase noise. An analog output can be provided by adding an m-bit D-to-A converter but, in digital radio applications, the digital sine wave is used directly. The more bits used to form the output sine wave, the lower the phase noise. Note that the DDS can incorporate the technique of digital pipelining in the adder/accumulator to achieve simultaneously high output frequencies and fine frequency resolution. The pipe delay is of no consequence for most system applications. Note also that the DDS can easily be modified to produce a chirped frequency by adding another accumulator to produce a sawtooth addend value. A precise FM modulator consists of a DDS whose addend is the digitized audio signal. Clock rates in DDS ICs have reached several GHz.

Output spectrum of the DDS

The exact spectrum of the DDS output can be calculated directly using Fourier analysis. For any value of the addend, the sequence of accumulator values will always repeat after at most 2^n clock pulses. Therefore, the sequence of digital values produced by the lookup ROM also repeats after at most 2^n cycles. This repetitive waveform can be represented as a Fourier series whose components are spaced in frequency by at least $f_{\text{clk}}/2^n$. The squares of the Fourier coefficients at each frequency are proportional to power. A simpler and more instructive approach is to note that the number contained in truncated accumulator bits (the bits below $n - m$) is just the phase error, $\delta\theta_i = \theta_i - \theta m_i$. The output sequence formed by the m-bit numbers can therefore be written as

$$\sin(\theta m_i) = \sin(\theta_i - \delta\theta_i) \approx \sin(\theta_i) - \delta\theta_i \cos(\theta_i), \tag{12.12}$$

where we have assumed $\delta\theta_i$ is small. The term $\sin(\theta_i)$ is the desired output, while the term $\delta\theta_i \cos(\theta_i)$ is the noise. The error sequence, $\delta\theta_i \cos(\theta_i)$, will repeat after, at most, 2^{n-m} clock pulses, so the noise forms a Fourier series with components separated in frequency by at least $f_{\text{clk}}/2^{n-m}$. The nature of the noise will depend on the value of the bottom part of the addend, i.e., the lowest $n - m$ bits of the addend. If that number is a power of 2, the noise components will fall at harmonics of the output frequency, distorting the output waveform slightly, but contributing no phase noise. If the bottom part of the addend contains no factors of 2, the noise may be a grass of components at all multiples of $f_{\text{clk}}/2^{n-m}$. The more factors of 2 in the bottom part of the addend, the wider the spacing between noise components. Thus, the nature of the noise can change dramatically from one output frequency to the next.

12.2.5 Switching speed and phase continuity

Indirect synthesizers cannot change frequency faster than the time needed for their phase lock loops to capture and settle. Direct synthesizers and direct digital synthesizers can switch almost instantly. Sometimes it is necessary to switch frequencies without losing phase continuity. The DDS is perfect for this since the addend is changed and the phase rate changes but there is no sudden phase jump. Other times, when the synthesizer is retuned to a previously selected frequency, the phase must take on the value it would have had if the frequency had never been changed. This second kind of continuity might be called phase memory. A frequency synthesizer that provides the first kind of phase continuity clearly will not provide the second and vice versa. Continuity of the second kind can be obtained with a direct synthesizer that uses only mixers and multipliers (no dividers – which can begin in an arbitrary state).

12.2.6 Phase noise from multipliers and dividers

It is important to see how any noise on the reference signal of a synthesizer determines the noise on its output signal. Let us examine how noise is affected by the operations of frequency multiplication and division. We will assume the input noise sidebands are much weaker than the carrier. Suppose the input signal has a discrete sideband at 60 Hz. (Such a sideband would normally be one of a pair but for this argument we can consider them one at a time.) Let this sideband have a level of −40 dBc, i.e., its power is 40 dB below the carrier power. If this signal drives a times-N frequency multiplier, it turns out that the output signal will also have a sideband at 60 Hz but its level will be $-40 + 20\log(N)$ dBc. The relative sideband power increases by the *square* of the multiplication factor.

Let us verify this, at least for a specific case – a particular frequency tripler. Let the input signal be $\cos(\omega t) + \alpha\cos([\omega+\delta\omega]t)$, i.e., a carrier at ω having a sideband at $\omega+\delta\omega$ with relative power of α^2 or $20\log(\alpha)$ dBc. We assume that $\alpha \ll 1$. Here the tripler will be a circuit whose output voltage is the cube of the input voltage. Expanding the output and keeping only terms of order α or higher whose frequencies are at or near $3f$ we have

$$[\cos(\omega t) + \alpha\cos[(\omega+\delta\omega)t]]^3 \rightarrow \cos^3(\omega t) + 3\alpha\cos^2(\omega t)cos[(\omega+\delta\omega)t] \tag{12.13}$$

$$\rightarrow 1/2\ \cos(\omega t)\cos(2\omega t) + 3/2\ \alpha\cos(2\omega t)\cos[(\omega + \delta\omega)t]$$

$$\begin{aligned} &\rightarrow 1/4\ \cos(3\omega t) + 3/4\ \alpha\cos[(3\omega + \delta\omega)t] \\ &\ = 1/4\ [\cos(3\omega t) + 3\alpha\cos[(3\omega + \delta\omega)t]]. \end{aligned} \tag{12.14}$$

Note that the carrier-to-sideband spacing is still $\delta\omega$ but the relative amplitude of the sideband has gone up by 3, the multiplication factor. The relative power of the sideband has therefore increased by 3^2, the square of the multiplication factor. A continuous distribution of phase noise $S(\delta\omega)$ is like a continuous set of discrete sidebands so, if the noise spectrum of a multiplier input is $S(\delta\omega)$, the noise spectrum of the output will be $n^2 S(\delta\omega)$. Sideband noise enhancement is a direct consequence of multiplication. If the multiplier circuit itself is noisy, the output phase noise will increase by more than n^2. Fortunately, most multipliers contribute negligible additive noise.

Division, the inverse of multiplication, reduces the phase noise power by the square of the division factor. Mixers just translate the spectrum of signals; they do not have a fundamental effect on noise. Additive noise from mixers is usually negligible. If a direct synthesizer is built with ideal components, the relation between the output phase noise and the phase noise of the standard will be as if the synthesizer were just a multiplier or divider, no matter what internal operations are used. The phase noise produced by indirect synthesizers depends on the quality of the internal VCOs and the bandwidths of the loop filters.

Problems

*These more difficult problems could be used as projects.

Problem 12.1. How would you modify the gear train in Figure 12.3 so that the output shaft would turn *N/M* times faster than the input shaft?

Problem 12.2. Show that Equation (12.2) describes the time domain input/output relation for the loop filter circuit of Figure 12.4.

Problem 12.3. Suppose the VCO in the block diagram of Figure 12.5 is noisy and that its noise can be represented as an equivalent noise voltage added at the control input of the VCO. Redraw the block diagram showing a summing block just in front of the VCO. Find the frequency transfer function for this noise input. Show that the loop is able to "clean up" the low-frequency noise of the VCO.

Problem 12.4.* Write a computer program to numerically simulate the PLL shown in Figure 12.5. Use a multiplier type phase detector and investigate the process of lock-up. Let the phase detector output be proportional to $\cos(\theta - \theta_{\text{ref}})$. Use numerical integration on the simultaneous first-order differential equations for $V_1(t)$ and $V_2(t)$.

Problem 12.5.* Invent a phase detector circuit that would have a range of many multiples of 2π rather than the restricted range of the multiplier phase detector.

Problem 12.6.* Derive the formula given for lock-up time, $\tau_{\text{LOCKUP}} \approx 4(\Delta\omega)^2/B^3$, where $\Delta\omega$ is the initial frequency error and B is the loop bandwidth (in radians/sec). Consider the type-II loop of Figure 12.5. Find the ac component on the VCO control input. Assume that $\Delta\omega$ is high enough that the gain of the loop filter for this ac voltage is just $-R_2/R_1$. Find the amplitude of the sidebands caused at the output of the VCO from this ac control voltage component. Use this amplitude to find the dc component at the output of the phase detector caused by product of the reference signal and the VCO sideband at the reference frequency. This dc component will be integrated by the loop filter. Find $\mathrm{d}V/\mathrm{d}t$ at the output of the loop filter. Find an expression for τ_{LOCKUP} by estimating the time for the dc control voltage to change by the amount necessary to eliminate the initial error.

Problem 12.7. Design a direct synthesizer (ad hoc combination of mixers, dividers, multipliers) to produce a frequency of 105.3 MHz from a 10-MHz reference. Avoid multipliers higher than ×5 and do not let the two inputs of any mixer have a frequency ratio higher than 5:1 (or lower than 1:5).

Problem 12.8. Design a direct digital synthesizer with 10-kHz steps for use as the tunable local oscillator in a middle-wave broadcast band (530–1700 kHz) AM receiver with a 455 kHz IF frequency. Assume a high-side L.O., i.e., the synthesizer frequencies range from 530 + 455 to 17800 + 455 An accurate reference frequency is available at 10.24 MHz.

Problem 12.9. Design a synthesizer with the range of 1–2 MHz that has phase memory, i.e., when the synthesizer is reset to an earlier frequency, its phase will be the same as if it had been left set to the earlier frequency. The required step size is 50 kHz and the available frequency reference is 5 MHz. Hint: one approach is to generate a frequency comb with a spacing of 50 kHz and then phase-lock a tunable oscillator to the desired tooth of the comb.

Problem 12.10. Explain why a direct synthesizer that includes one or more dividers will not have phase memory.

Problem 12.11. Draw a block diagram for an FM transmitter in which a phase locked loop keeps the average frequency of the VCO equal to a stable reference frequency.

Problem 12.12. Draw block diagrams for PM and FM generators based on the direct digital synthesizer (DDS) principle.

References

[1] Crawford, James A. "*Frequency Synthesizer Design Handbook*," Boston: Artech House, 1994.

[2] Gardner, Floyd M., *Phaselock Techniques*, 2nd edn. New York: John Wiley, 1979.

[3] Kuo, Benjamin, *Automatic Control Systems*, 5th edn. Englewood Cliffs: Prentice Hall, 1987.

[4] Manassewitsch, V., *Frequency Synthesizers Theory and Design*, 2nd edn. New York: John Wiley, 1980.

CHAPTER

13 Coupled-resonator bandpass filters

We saw in Chapter 4 that the straightforward transformation of a prototype lowpass filter to a bandpass filter yields a circuit with alternating parallel resonant circuits and series resonant circuits as shown in Figure 13.1.

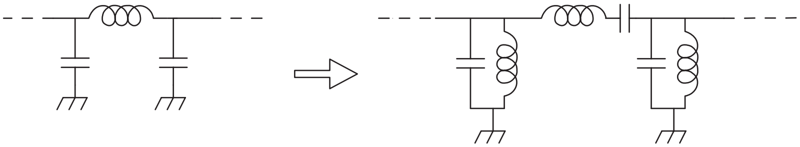

Figure 13.1. Conversion of a lowpass filter to a canonical bandpass filter.

If the prototype lowpass filter has a response $F_{\mathrm{LP}}(\omega)$, the corresponding bandpass filter will have the response $F_{\mathrm{BP}}(\omega)=F_{\mathrm{LP}}(\,|\omega-\omega_{\mathrm{C}}|)$, where ω_{C} is the center frequency. These canonical bandpass filters work perfectly – when simulated with a network analysis program. But usually they call for impractical component values. The inductors in the shunt branches must be smaller than the inductors in the series branches by a factor on the order of the square of the fractional bandwidth. For a 5% bandwidth filter, the ratio of the inductor values would be of the order of 1:400. For a given center frequency we might be lucky to find a high-Q inductor of any value, let alone high-Q inductors with such different values. Low-Q (resistive) inductors make a filter lossy and change its nominal passband shape. The series and shunt capacitor values have the same ratio. Generally Q is not a problem with capacitors, but very small values are impractical when they become comparable to the stray wiring capacitances.

13.1 Impedance inverters

This component value problem can be solved by transforming canonical bandpass filters into *coupled-resonator bandpass filters*, which can be built with identical or almost identical *LC* resonant circuits. The coupled-resonator filters have the same filter shapes, based on prototype lowpass designs, such as Butterworth or Chebyshev. Figure 13.2 shows some coupled-resonator fiter designs.

These filters are based on *impedance inverters*. Three examples of impedance inverters are shown in Figure 13.3, a 90° length of transmission line and two lumped *LC* circuits.

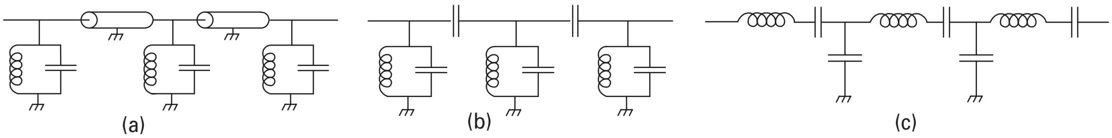

Figure 13.2. Three examples of coupled-resonator bandpass filter circuits.

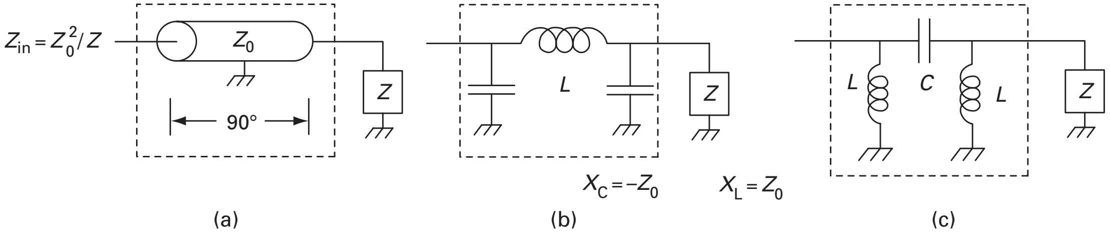

Figure 13.3. Three impedance inverter circuits

In every case, an impedance Z, when seen through the inverter, becomes $Z_0{}^2/Z$ where Z_0 can be called the characteristic impedance of the inverter. For the transmission line inverter, a 90° length of line, Z_0 is just the characteristic impedance of the line. For the *LC* inverters, both the inductor's reactance, X_L, and the capacitor's reactance, X_C, must be equal to the desired Z_0. Like the 90° cable, the lumped element circuits are perfect inverters only at one frequency but, in practice, they are adequate over a considerable range. An inverted capacitor is an inductor. An inverted inductor is a capacitor. Figure 13.4 shows an inverter (in this example, a 90° transmission line) used to invert a parallel circuit, making an equivalent series circuit.

The mathematics of this inversion is just

$$Z_{in} = Z_0{}^2 Y = Z_0{}^2\left(\frac{1}{j\omega L_p} + j\omega C_p + \frac{1}{R_p}\right) = \frac{1}{j\omega(L_p/Z_0{}^2)} + j\omega(Z_0{}^2 C_p) + \frac{Z_0{}^2}{R_p}. \tag{13.1}$$

Let us look at four inverters which include inductors or capacitors with negative values. For these inverters, shown in Figure 13.5, $X_C = Z_0$ or $X_L = Z_0$.

Figure 13.6 verifies the inverter action of the all-capacitor T-section inverter. You can use this kind of reasoning to verify the inverter action of the other circuits:

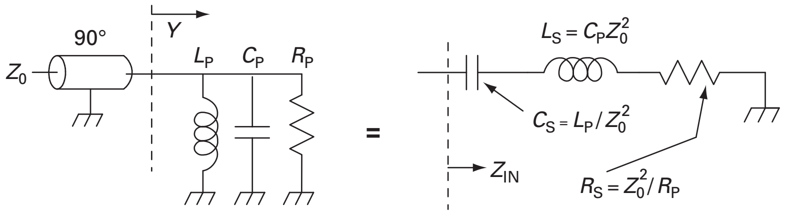
Y
90°
Z_0
L_P
C_P
R_P
=
$L_S = C_P Z_0^2$
$C_S = L_P / Z_0^2$
Z_{IN}
$R_S = Z_0^2 / R_P$

Figure 13.4. Impedance inverter makes a parallel circuit appear as a series circuit.

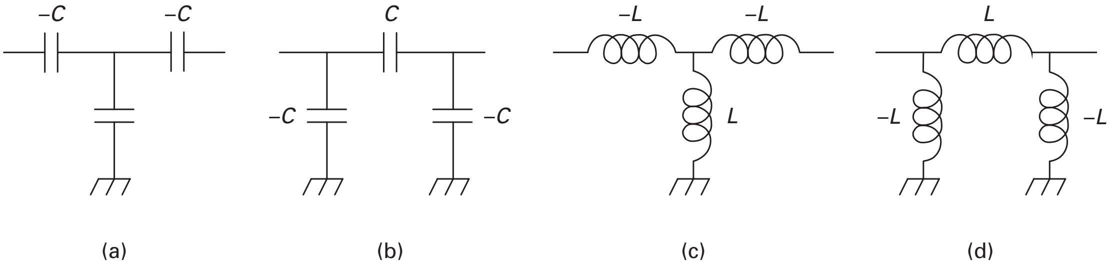
−C
−C
C
−C
−C
−L
−L
L
L
−L
−L
(a)
(b)
(c)
(d)

Figure 13.5. Impedance inverters based on negative value components.

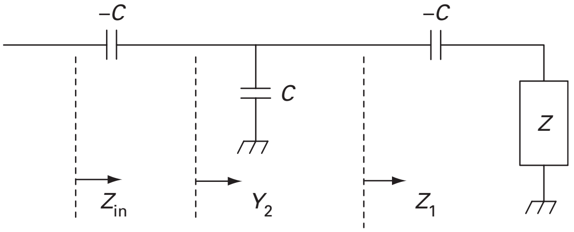
−C
−C
C
Z
Z_{in}
Y_2
Z_1

Figure 13.6. Operation of the T-network negative capacitor inverter.

$$Z_1 = \frac{1}{j\omega(-C)} + Z \tag{13.2}$$

$$Y_2 = j\omega C + \frac{1}{Z_1} = \frac{j\omega CZ}{Z - 1/j\omega C} \tag{13.3}$$

$$Z_{in} = \frac{1}{j\omega(-C)} + \frac{1}{Y_2} = \frac{1}{\omega^2 C^2 Z} = \frac{Z_0^2}{Z}. \tag{13.4}$$

Because they contain negative capacitances or negative inductances, the four inverters in Figure 13.5 might seem to be only mathematical curiosities. Not at

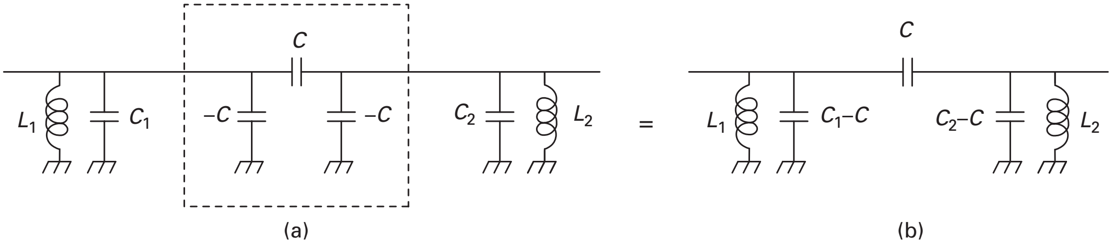

Figure 13.7. Negative capacitors absorbed into adjacent positive capacitors.

all; the negative elements can be absorbed by positive elements in the adjacent circuitry as shown in Figure 13.7, where a π-section capacitor inverter is placed between two parallel *LC* "tanks."

13.2 Conversion of series resonators to parallel resonators and vice versa

Ladder network filters have alternating series and shunt branches. Let us see how inverter *pairs* are used in ladder filters. Suppose we embed a series capacitor between a pair of inverters at some point along a ladder network.

Figure 13.8. A series capacitor between inverters is equivalent to a shunt inductor.

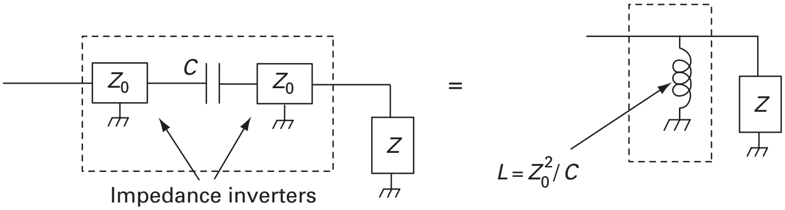

The combination of the capacitor and the inverter pair is equivalent to a shunt inductor, as shown in Figure 13.8.

You can show just as easily that any series impedance, Z_s, together with a pair of bracketing inverters of characteristic impedance Z_0 is equivalent to a shunt admittance $Y_p = Z_0^{-2} Z_s$. Likewise, the combination of any shunt admittance Y and a pair of bracketing inverters is equivalent to a series impedance $Z = Z_0^2 Y$. Figure 13.9 illustrates this, showing how a series resonant series branch in an ordinary bandpass filter can be replaced by a parallel resonant shunt branch imbedded between a pair of inverters.

Likewise, a parallel resonant shunt branch can be realized as a series resonant series branch imbedded between a pair of inverters, as shown in Figure 13.10.

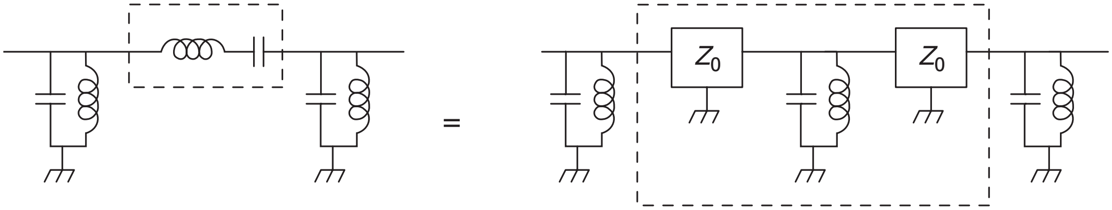

Figure 13.9. A shunt resonator between inverters is equivalent to a series resonator.

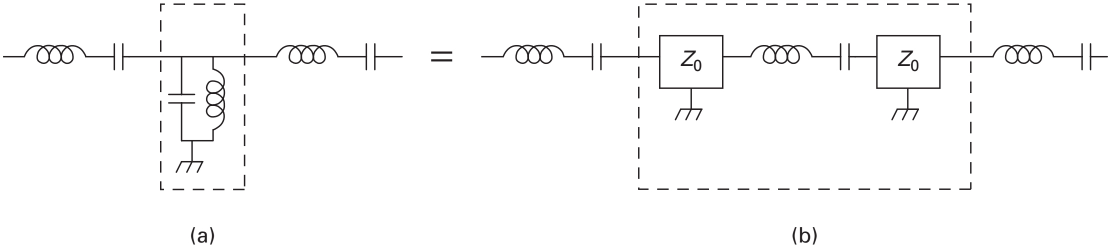

Figure 13.10. A series resonator between inverters is equivalent to a shunt resonator.

13.3 Worked example: a 1% fractional bandwidth filter

Consider a 50-ohm, 1-dB Chebyshev filter with a 10-MHz center frequency and a bandwidth of 100 KHz between the 1-dB points. The filter, which results from the straightforward lowpass to bandpass transformation (Chapter 4) is shown in Figure 13.11 and its response is shown in Figure 13.12.

We might find 86-μH inductors with high Q at 10 MHz but the 3.728 nH and 2.645 nH inductors would be tiny single turns of wire with very poor Q. To get around these component limitations, we will convert this filter into a coupled-resonator filter. Suppose we have in hand some adjustable 0.3 to 0.5 microhenry

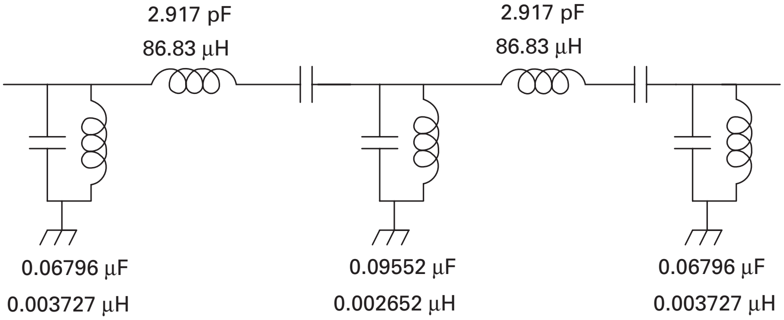

Figure 13.11. A straightforward (but impractical) bandpass filter. The calculated response of this filter is shown in Figure 13.12.

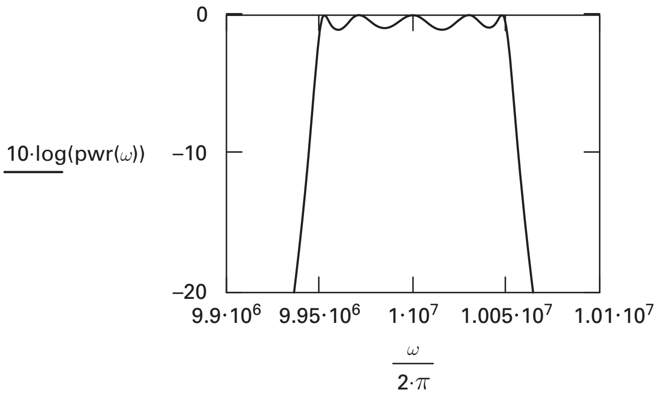

Figure 13.12. Calculated response for filter of Figure 13.11.

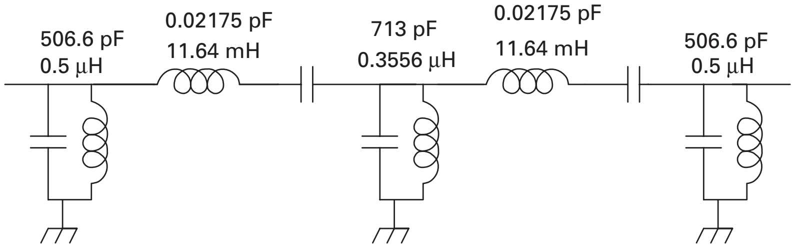

Figure 13.13. Filter of Figure 13.11, scaled from 50 to 6705 ohms.

inductors which, at 10 MHz, have very high Q (we will see later just how much Q is required). Let us first change the working impedance of the filter so that the parallel resonators at the end will use 0.5 μH, which is 134.1 times the original end inductors and implies that the filter will be scaled to 50 × 134.1 = 6705 ohms. We multiply the other inductors by 134.1 and divide the capacitors by 134.1 to get the circuit of Figure 13.13.

The parallel resonators now use the desired inductors but the series resonators call for inductors of 11.6 mH, a very large value for which we surely will not find high Q components. Moreover, the series capacitors are only 0.02 pF, a value far too small to be practical. We can solve this problem by using impedance inverters to convert the series resonators into parallel resonators. Let us use the all-capacitor π-section inverters of Figure 13.5(b) and the same parallel resonators we used for the end sections. Figure 13.14 shows how two inverters and the parallel resonator replace each series resonator.

We can calculate the inverter's characteristic impedance, Z_0, as follows:

$$Z_0{}^2 Y = Z;\; Z_0{}^2(\mathrm{j}\omega C_\mathrm{p} + 1/\mathrm{j}\omega L_\mathrm{p}) = \mathrm{j}\omega L_\mathrm{S} + 1/\mathrm{j}\omega C_\mathrm{S} \qquad (13.5)$$

$$Z_0{}^2 = L_\mathrm{p}/C_\mathrm{S} = 0.5\times10^{-6}/0.02175\times10^{-12} = 4796^2. \qquad (13.6)$$

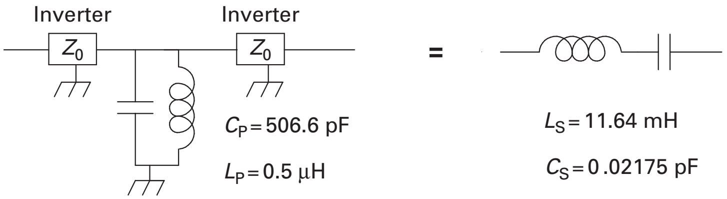
Inverter
Z_0
Inverter
Z_0
$C_P = 506.6$ pF
$L_P = 0.5$ µH
=
$L_S = 11.64$ mH
$C_S = 0.02175$ pF

Figure 13.14. Inverters transform a 0.5-μH shunt inductor into a 11.644-mH series inductor.

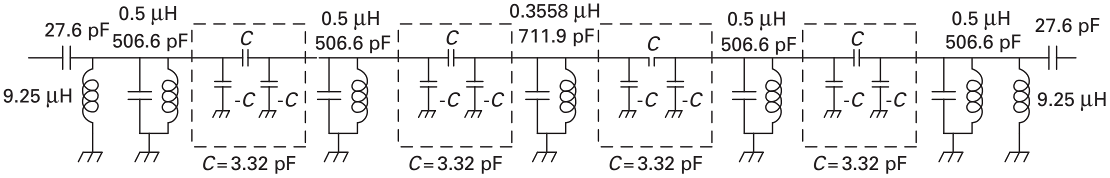
27.6 pF
0.5 µH
506.6 pF
9.25 µH
C
-C
-C
C = 3.32 pF
0.5 µH
506.6 pF
C
-C
-C
C = 3.32 pF
0.3558 µH
711.9 pF
C
-C
-C
C = 3.32 pF
0.5 µH
506.6 pF
C
-C
-C
C = 3.32 pF
0.5 µH
506.6 pF
27.6 pF
9.25 µH

Figure 13.15. Coupled-resonator version of previous bandpass filter.

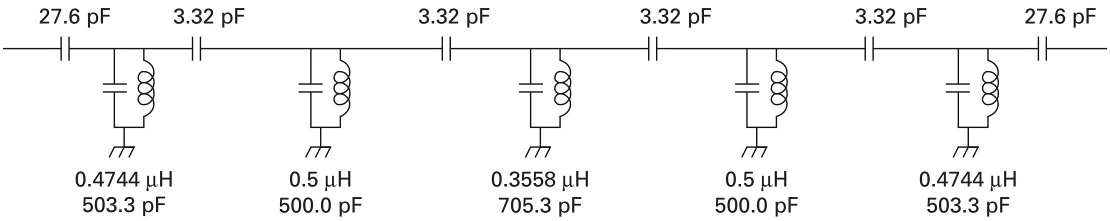
27.6 pF
3.32 pF
3.32 pF
3.32 pF
3.32 pF
27.6 pF
0.4744 µH
503.3 pF
0.5 µH
500.0 pF
0.3558 µH
705.3 pF
0.5 µH
500.0 pF
0.4744 µH
503.3 pF

Figure 13.16. Finished coupled-resonator filter.

For this type of inverter, we had seen that $Z_0 = X_C$, so $C = 3.32$ pF. We now have our coupled-resonator filter but since it works at 6705 ohms we will add L-section matching networks at each end to convert it back to 50 ohms. The filter, at this point, is shown in Figure 13.15. All the resonators are now parallel resonators. (In other situations we might use inverters to convert series resonators into equivalent parallel resonators to make an all-series-resonator filter – see Figure 13.1.)

The final clean-up step is to absorb the -3.32 pF capacitors into the resonator capacitors and combine the matching section inductors with the end-section resonator inductors as shown in Figure 13.16.

The response of the finished filter is shown in Figure 13.17 and is almost identical to the response of the prototype filter of Figure 13.11. The difference, a fraction of a dB, occurs because the inverters work perfectly only at the center frequency.

13.4 Tubular bandpass filters

A popular bandpass filter design, the “tubular filter” is produced by many filter manufacturers. Figure 13.18 shows the construction of a three-resonator tubular filter.

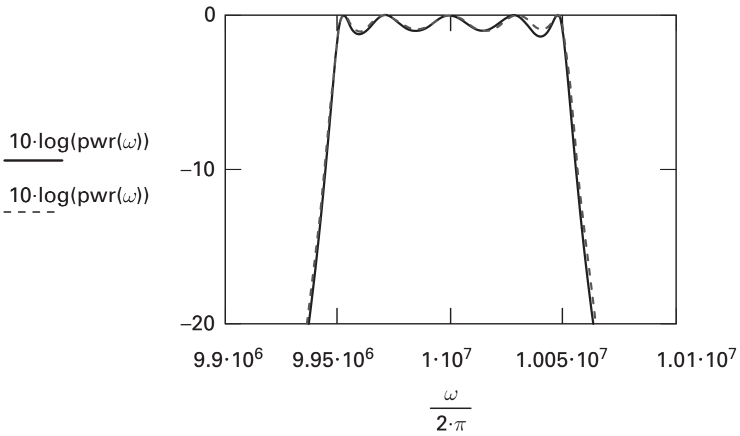

Figure 13.17. Calculated response of the filters of Figure 13.16 (pwr) and Figure 13.11 (pwr).

The only standard electronic components are the coaxial connectors at the ends. There are also (in this example) three inductors (wire coils), four metal cylinders, two dielectric spacers, two (or one long) dielectric sleeves, and a tubular metal body. Figure 13.19 shows how a coupled-resonator filter design, of the type we have discussed, is transformed into the tubular filter design. You can verify that Figure 13.19(d) is the circuit diagram of the tubular filter. The three-capacitor π-sections are formed by the capacitance between the adjacent faces of the metal cylinders and the capacitors are formed between the outside surfaces of the cylinders and the tubular body.

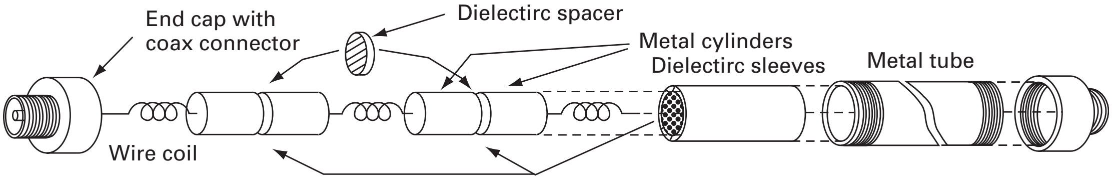

Figure 13.18. Tubular bandpass filter.

Beginning with Figure 13.19(a), we have a standard coupled-resonator bandpass filter using series resonators. In the canonical prototype for this filter, the middle section is a parallel resonator, but this has been replaced by a series resonator sandwiched between two impedance inverters. In (b), the center capacitor has been replaced by two capacitors (each of twice the value of the original capacitor so that, in series, the total series capacitance is the same). The capacitors have been shifted slightly in (c) to identify a T-section capacitor network at each side of the central inductor. Finally, in going from (c) to (d), these T-networks are replaced by equivalent π-networks, to arrive at the circuit of the tubular filter. Any

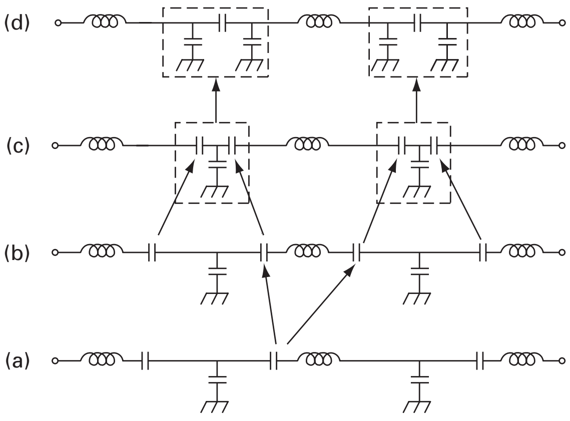

Figure 13.19. Tubular bandpass filter evolution.

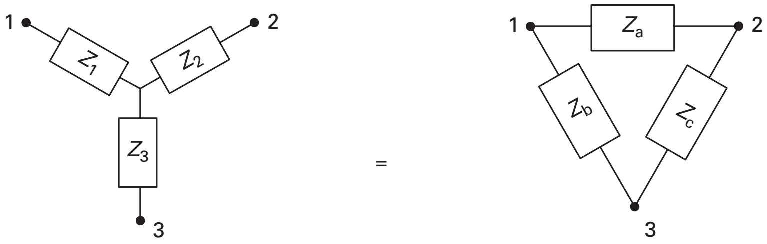

$$Z_1 = \frac{Z_a Z_b}{Z_a + Z_b + Z_c} \qquad Z_a = \frac{Z_1 Z_2 + Z_2 Z_3 + Z_3 Z_1}{Z_3}$$

Figure 13.20. Equivalent π-section and T-section networks.

T-network has an equivalent π-network and vice versa (Problem 13.5). These transformations, also known as T–π and π–T are shown in Figure 13.20. Formulas are given for one element in each network; the others follow from symmetry.

13.5 Effects of finite *Q*

These calculated filter responses assume components of infinitely high *Q*. We can calculate the effects of finite *Q* by paralleling the (lossless) inductors in our model with resistors equal to *Q* times the inductor reactances at the center frequency. If, for example, the *Q* is 500 (quite a high value for a coil), we would parallel the inductors in the filter of Figure 13.15 with resistors of about 15 000 ohms. Reanalyzing the circuit response, we would find that the filter will have a midband insertion loss of 7 dB and that the flat (within 1 dB) passband response becomes rounded. The effect will be somewhat less for a filter with more gradual skirts, e.g., a 0.01 dB Chebyshev or a Butterworth filter. But the real problem is still the small fractional bandwidth. For a filter with small fractional bandwidth to have the ideal shape of Figure 13.17, the resonators must be quartz or ceramic or other resonators with Qs in the thousands. An approximate analysis predicts that the midband loss per section in a bandpass filter will be on the order of

$$\frac{\text{power transmitted}}{\text{power incident}} = \left(1 - \frac{L_0/2}{Q \cdot \text{fractional bandwidth}}\right) \tag{13.7}$$

where L_0 represents the inductor value in the normalized lowpass prototype filter. For our five-section filter we can take L_0 to be about 1.5 henrys. If the inductor Q is 500, the predicted transmission of the five-section filter is $5 \times 10 \log[1-(1.5\ /2)/(500\cdot(1/100))] = -10$ dB, which is roughly equal to the actual value of −7 dB.

13.6 Tuning procedures

Filters with small fractional bandwidths and sharp skirts are extremely sensitive to component values. In the filter of Figure 13.16, for example, the resonators must be tuned very precisely or the shape will be distorted and the overall transmission will be lowered. (The values of the small coupling capacitors – all that remains of the impedance inverters – are not as critical.) Usually each resonator is adjustable by means of a variable capacitor or variable inductor. All the adjustments interact and, if the filter is totally out of tune, it may be hard to detect any transmission at all. A standard tuning procedure is to monitor the input impedance of the filter while tuning the resonators, one-by-one, beginning at input end. While resonator N is being adjusted, resonator N+1 is short circuited. The tuning of one resonator is done to produce a maximum input impedance while the tuning of the next is done to produce a minimum input impedance. The procedure must sometimes be customized to account for matching sections at the ends.

13.7 Other filter types

The coupled-resonator technique is used from HF through microwaves. Not all RF bandpass filters, however, use the coupled-resonator technique. The IF bandpass shape in television receivers is usually determined by a SAW (surface acoustic wave) bandpass filter. SAW filters are FIR (finite impulse response) filters, whereas all the *LC* filters we have discussed are IIR (infinite impulse response) networks. This classification is made according to the behavior of the output voltage following a delta function (infinitely sharp impulse) excitation. Digital filters can be designed to be either FIR or IIR filters.

Problems

Problem 13.1. Use your network analysis program to verify that the filter of Figure 13.16 does indeed give the response shown in Figure 13.17.

Problem 13.2. Verify that the two *LC* circuits in Figure 13.3 are impedance inverters.

Problem 13.3. The filter shown below was developed in Chapter 4 as an example of the straightforward conversion from a prototype lowpass filter to a bandpass filter. This Butterworth (maximally flat) filter has a bandwidth of 10 kHz and a center frequency of 500 kHz. Suppose you have available some 30 μH inductors with a *Q* of 100 at 500 kHz. Convert the filter into a coupled-resonator filter that uses these inductors. Use your ladder network analysis program to verify the performance of your filter.

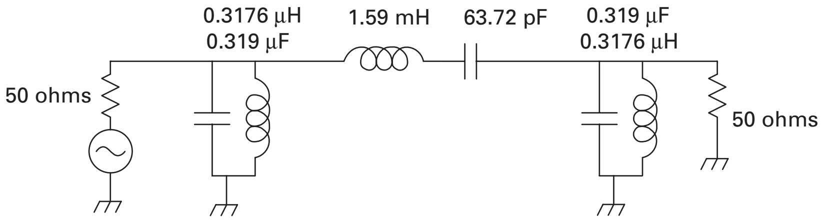

Problem 13.4. A bandpass filter is to have the following specifications:

Center frequency: 10 MHz; shape: three-section 1-dB Chebyshev; bandwidth: 3 KHz (between outermost 1-dB points); source and load Impedances: 50 ohms. Since the loaded *Q* of this filter is very high, $10^6/3000 = 333$, it is important to use very high-*Q* resonators. Suppose you have located some resonators (cavities, crystals, or whatever) with adequate *Q*. These resonators are all identical. At 10 MHz they exhibit a parallel resonance, equivalent to a parallel *LC* circuit. At 10 MHz, they have a susceptance slope of 10^{-6} (1 mho/MHz).

(a) Find the *LC* equivalent circuit for these resonators (in the vicinity of 10 MHz).
(b) Design the filter shown below around these resonators.
(c) Use your ladder network analysis program to verify the frequency response of your design.

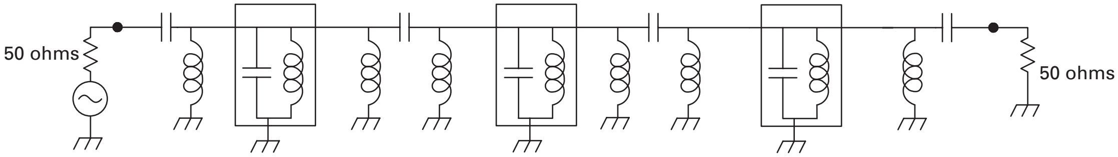

Problem 13.5. Derive expressions for Z_a, Z_b, and Z_c in terms of Z_1, Z_2, and Z_3 for the equivalent *T* and *π* networks shown in Figure 13.19. Hint: consider the connections

shown below. The sketched-in wires show that $Y_A + Y_B = (Z_3 + Z_1 \| Z_2)^{-1}$. If you write the corresponding $Y_B + Y_C$ and $Y_C + Y_A$ equations, then add the first two and subtract the third, you will have the formula for Y_B. A similar technique yields the expressions for Z_1, Z_2, and Z_3.

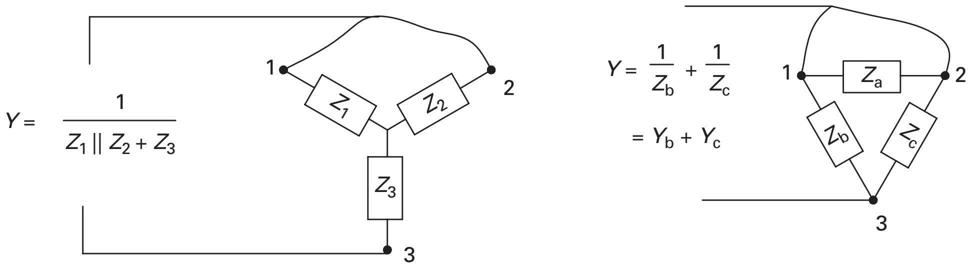
$Y = \dfrac{1}{Z_1 \| Z_2 + Z_3}$
1
2
3
Z_1
Z_2
Z_3
$Y = \dfrac{1}{Z_b} + \dfrac{1}{Z_c}$
$= Y_b + Y_c$
1
2
3
Z_a
Z_b
Z_c

(a) (b)

Problem 13.6. The bridge circuit shown below in (a) is the simplest network whose resistance cannot be found immediately by series and parallel reduction. Rather than resorting to loop or node equations, note that the circuit contains two πs and two Ts. Replace a π by its equivalent T or a T by its equivalent π. Now find the resistance of the network by simple reduction. The circuit at the right shows how one of the π's can be replaced by a T.

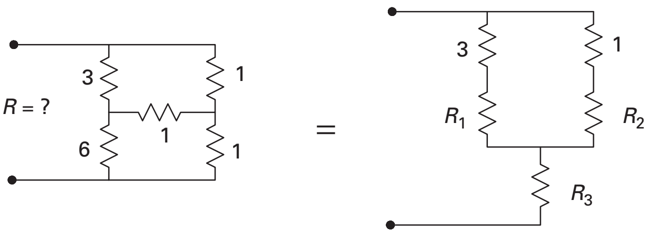

(a) (b)

References

[1] Christiansen, D., Alexander, C., Jurgen, R. K. *Standard Handbook of Electronic Engineering*, 5th edn, New York: McGraw-Hill, 2004.

[2] Matthaei, G., Young, L. and Jones, E. M. T. *Microwave Filters, Impedance-Matching Networks, and Coupling Structures*, New York: McGraw Hill, 1964, reprinted, Boston: Artech House, Inc. 1980.

CHAPTER

14 Transformers and baluns

Transformers are harder to understand than resistors, capacitors, and single inductors. First, transformers have two terminal pairs rather than one, so we must deal with two voltages and two currents. Second, we can be misled by the deceptive simplicity of the simplest mathematical model, the "ideal transformer." In this chapter we discuss the conventional transformers used in power supplies, switching power supplies, amplifiers, and RF matching networks. We will then examine transmission line transformers, which work to higher frequencies, and *baluns*, which are devices used to connect *bal*anced circuits to *un*balanced circuits.

Figure 14.1(a) shows two inductors (here, wire coils) with arbitrary placement. The region does not have to be otherwise empty; it can contain any distribution of clumped and/or continuous magnetic materials.

The inductance values ("self-inductances") of these coils are L_1 and L_2, each measured with the other coil open circuited, so that it carries no current. We will refer to these coils as L_1 and L_2. Two representative magnetic flux lines are shown, corresponding to a current in L_1. Note that one of these flux lines is encircled by three turns of L_2. Therefore, when L_1 carries an ac current, by Faraday's law there will be an ac voltage induced in L_2, proportional to the time derivative of the encircled magnetic flux. We can write Faraday's law for this situation as

$$V_2 = \mathrm{j}\omega t M I_1 + \mathrm{j}\omega t L_2 I_2, \tag{14.1}$$

where the constant M is known as the "mutual inductance." The jω factor represents the time derivative since we are using standard ac circuit analysis, where the time dependence is contained in an implicit factor $\mathrm{e}^{\mathrm{j}\omega t}$. The second term, containing the self-inductance, L_2, is a voltage induced by the current, if any, flowing in L_2 itself. Note that we have assumed that the wires have negligible resistance – no IR voltage drop.[1] The directions of the currents are

---

[1] To take account of the winding resistances, we would add a term I_2R_2 to the right-hand side of Equation (14.1) and a term I_1R_1 to the right-hand side of Equation (14.2). The effect of the resistance distributed in the windings is the same as if the resistance were consolidated into two external resistors, R_1 and R_2, in series with the windings.

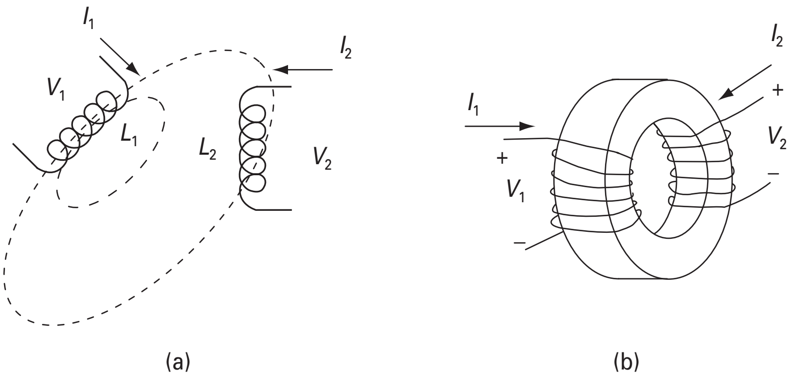

Figure 14.1. (a) Coupled inductors; (b) prototype transformer.

defined by arrows in Figure 14.1 as entering the positive end of each coil. This symmetric assignment produces a symmetric equation for V_1:

$$V_1 = \mathrm{j}\omega t L_1 I_1 + \mathrm{j}\omega t M I_2. \tag{14.2}$$

Note that both equations contain the same constant M. There is no need to write M_{12} and M_{21}, since the mutual inductances are always equal. This can be seen quite easily for the arrangement of coupled coils shown in Figure 14.1(b). This a conventional transformer, in which a toroid of iron or other magnetic material effectively contains all the field lines, forcing them to thread through every turn of both L_1 and L_2. An ac current in L_1 produces a flux proportional to N_1, the number of turns in L_1. The ac voltage induced in L_2 is proportional to N_2 times the flux. Hence M_{21} is proportional to the product N_1N_2. Likewise the ac current in L_2 produces an ac voltage in L_1 with the same proportionality to N_1N_2, so M_{12} and M_{21} are equal.[2]

If all the flux lines thread both windings, the transformer is said to be *perfectly coupled*. A *coupling coefficient*, k, is defined by $M = k(L_1\ L_2)^{1/2}$. The value of k ranges from zero, for inductors with no coupling, to unity, for perfect coupling. In the transformer of Figure 14.1(b), perfect coupling is approached by using a core material of extreme magnetic permeability. For such a transformer, it makes no difference whether the windings are side-by-side, as shown, or wound one on top of the other. Nor is it necessary that they be wound tightly around the core; loose windings can be used to allow circulation in an oil-cooled power transformer.

14.1 The "ideal transformer"

If a transformer is perfectly coupled and the windings have negligible resistance, then the ratio of primary-to-secondary[3] voltages is equal to the turns ratio and these voltages have the same phase. This strict proportionality of voltages follows directly from Faraday's law, since the time derivative of the magnetic flux is equal in both windings. The situation with the currents is not as simple. The primary and secondary currents are not strictly proportional. This follows from the transformer's ability to store magnetic energy. If the transformer could *not* store energy, the instantaneous net power into the transformer would have to be zero. The proportionality of primary and secondary voltages would demand that the currents be proportional, i.e., setting $V_pI_p = V_sI_s$, we would have $I_s/I_p = V_p/V_s = \text{constant} = n_P/n_S$, the effective turns ratio. This hypothetical transformer, which stores no energy, is known as the *ideal transformer* and is a useful abstraction. We will discuss below a circuit with two inductors and an ideal transformer which together, are equivalent to a real transformer. But first let us emphasize how the ideal transformer, by itself, is an unrealistic model.

---

[2] The equality of M_{12} and M_{21} is a general reciprocity relation that holds true for any passive two-port network, e.g., any network made from resistors, capacitors, inductors, and transformers.

[3] The names "primary" and "secondary" are arbitrary and refer only to the way the transformer is used; power usually flows into the primary and out of the secondary.

Consider an ideal transformer with an effective turns ratio n_1/n_2. If an impedance Z_{load} is connected to the secondary, the ratio of primary voltage to primary current will be $(n_1/n_2)\ V_2/[(n_2/n_1)I_s] = (n_1/n_2)^2\ V_s/I_s = (n_1/n_2)^2Z_{\text{load}}$ times the secondary current. The primary will therefore present an impedance of $(n_1/n_2)^2Z_{\text{load}}$, a simple impedance multiplication. If Z_{load} is infinite (an open circuit), the impedance looking into the primary of the ideal transformer is also infinite. But inspection of Equation (14.2) shows that, in this case, the impedance looking into a real transformer is $j\omega L_1$. Another "unreal" feature of an ideal transformer is that it contains no magnetic field! The field (or flux) from the primary winding is exactly cancelled by the field from the secondary winding. And with no magnetic field there would be no $d\Phi/dt$ and therefore no voltage across either winding. A real transformer approaches the ideal transformer model only when the number of turns approaches infinity and the magnetic coupling approaches 100%. Transformers used in practice are usually far from ideal. This is in contrast to resistors and capacitors which, at least at low frequencies, are almost ideal components obeying the relations $Z_R = R$, $Z_C = 1/j\omega C$, and $Z_L = j\omega L$.

14.2 Transformer equivalent circuit

An equivalent circuit for a real transformer is shown in Figure 14.2(b).

This model circuit consists of two inductors plus an ideal transformer having an equivalent turns ratio n_1/n_2. The shunt inductor at the left is known as the *magnetizing inductance* and its value is L_1, the inductance of the primary. Note that, even if the secondary is left open, a voltage V applied to the primary will produce a primary current $I = V/(j\omega L_1)$. The series inductor at the right is known as the *leakage inductance*. Its value is $L_2\ (1-\ k^2)$, so a perfectly coupled transformer ($k = 1$) has no leakage inductance. In practice, maximum coupling is limited to maybe 98% at low frequencies and less at RF frequencies. The useful frequency range of a transformer is determined by these two inductances. Suppose we put a transformer between a resistive load and a signal generator.

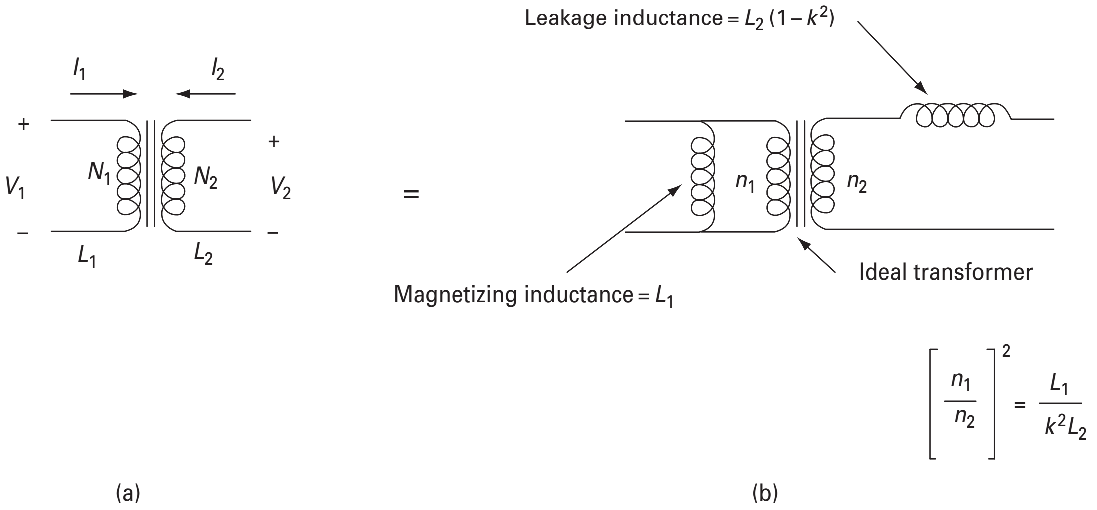

Figure 14.2. (a.) Transformer symbol with voltage and current assignments; (b.) an equivalent circuit.

Below the useful frequency range, the magnetizing inductance becomes a short circuit across the generator. Above the useful range, the leakage inductance becomes a high impedance in series with the load. In both extremes, the power delivered to the load becomes negligible.

Let us demonstrate that the model circuit of Figure 14.2(b) does, indeed, agree with the fundamental equations (14.1) and (14.2), i.e., that it really is an equivalent circuit. We have already seen that the value of the magnetizing inductance must be L_1, the inductance value of the primary winding. Leaving the secondary open, so that $I_2 = 0$, the fundamental equations (14.1) and (14.2) produce the relation $V_2/V_1 = M/L_1 = k\sqrt{L_1 L_2}/L_1 = k\sqrt{L_2/L_1}$. In this situation, the model circuit gives $V_2/V_1 = n_2/n_1$. For the model to agree with the fundamental equations, $n_2/n_1 = k\sqrt{L_2/L_1}$. Finally, consider the situation in which the primary is shorted, so that $V_1 = 0$. The secondary current given by Equations (14.1) and (14.2) must be the same as the current predicted by the model. Equation (14.1) produces $I_1 = -I_2 M/L_1$. Putting this into Equation (14.2) gives $V_2/I_2 = \mathrm{j}\omega\ (L_2/M - M^2\ /L_1) = \mathrm{j}\omega L_2(1-k^2)$. Looking at the model, the impedance at the secondary, with the primary shorted, is just $\mathrm{j}\omega$ time the leakage inductance, so the leakage inductance must be assigned the value $L_2(1- k^2)$. With these assignments for the values of the magnetization and leakage inductances, the model correctly reproduces Equations (14.1) and (14.2).

This is not the only possible equivalent circuit. We could just as well have constructed this equivalent circuit with the magnetizing inductance on the right side and the leakage inductance on the left side. Or we could "push" either the magnetizing inductance or the leakage inductance through the transformer, correcting the inductance by a factor $(n_1/n_2)^2$ or $(n_2/n_1)^2$, so that they are both on the same side. Another equivalent circuit, is shown in Figure 14.3. Using arguments like those presented above, you can show that this circuit also

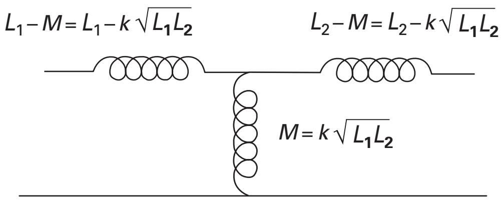
$L_1 - M = L_1 - k\sqrt{L_1L_2}$
$L_2 - M = L_2 - k\sqrt{L_1L_2}$
$M = k\sqrt{L_1L_2}$

Figure 14.3. An all-inductor equivalent circuit.

satisfies Equations (14.1) and (14.2). This circuit contains no unphysical ideal transformer (and therefore has no dc isolation between primary and secondary, making it not quite as equivalent). But you can see from the labels in the figure that, in general, one of the inductors must have an unphysical negative inductance. However, note that if $k < L_1/L_2$ and $k < L_2/L_1$ all the inductors are positive. As the turns ratio becomes close to unity, the inductors all remain positive as the coupling is increased. In the case of a 1:1 transformer with perfect coupling, the values of the series inductors go to zero and the equivalent circuit is just a single shunt inductor, the magnetizing inductance. (You could put a 1:1 ideal transformer on each side of this inductor to produce a symmetric equivalent circuit that preserves dc isolation.) And, of course, the circuit of Figure 14.3 could be converted from the T configuration to an equivalent pi configuration.

To approach the ideal, a transformer must have a very high magnetizing inductance and a very small leakage inductance. You can increase the magnetizing inductance by increasing the number of turns (keeping the turns ratio constant) but, in practice, this only increases the leakage inductance and the ohmic resistance of the windings. You can decrease the leakage inductance by using fewer turns, but this lowers the magnetizing inductance. A compromise is generally needed. However, we will see that there are applications in which the leakage and/or magnetizing inductances become useful circuit components.

14.3 Power transformer operation

Power transformers are usually iron-core transformers with high coupling. The best power transformer would be an ideal transformer; its stored energy, excitation current and leakage inductance would all be negligibly small. The primary is connected to the ac line, which can be regarded as a perfect voltage source with negligible source impedance. If the coupling is high enough to make the leakage inductance negligible and the winding resistances are low, the secondary voltage will be constant, $V_2 = (n_2/n_1)V_1$, independent of the load. The magnetizing inductance will draw a constant "magnetizing" current from the line, $I_M = V/(j\omega L_1)$. Since this current is 90° out of phase with respect to the primary voltage, it consumes no average power, but does cause "excitation" energy to slosh in and out of the transformer. When a resistive load is connected to the secondary, additional "working" currents flow in both the primary and the secondary. These currents have the ratio n_2/n_1. They are in phase with the primary voltage and transfer power from the source to the load. Sometimes a capacitor is placed across the primary to resonate with the magnetizing inductance. The excitation energy will then slosh back and forth between the capacitor and the magnetizing inductance. This corrects the power factor; the power line now only has to supply the component of current that is in phase with the voltage. In practice, the magnetizing current may be comparable to the maximum working current. In a power transformer, the magnetic core is used close to saturation. When the magnetizing current is at its maximum, the inductance of the core is reduced. This nonlinear behavior of the core distorts the otherwise sinusoidal waveform of the magnetizing current. Nevertheless, the voltages on the primary and secondary remain proportional and sinusoidal, because of the low source impedance of the power line, low *IR* drops in the windings, and negligible leakage inductance.

14.4 Mechanical analogue of a perfectly coupled transformer

A transformer transfers ac power, usually with a step-up or step-down in voltage. Figure 14.4 shows how a lever could be used to step down the velocity of a sinusoidally reciprocating arm. The resistive load on the right-hand side is a dashpot (damper), which produces a reaction force proportional to velocity. A voltage step-down transformer increases current. This lever steps down velocity (and amplitude) and provides increased force. For an ideal transformer (infinite magnetizing inductance) or an ideal lever (zero mass) the input power (primary voltage times primary current or primary velocity times primary force) is equal to the output power at every instant. But, for a real transformer with finite magnetizing inductance, there is also the "excitation" current, lagging the voltage by 90°, pumping energy in and out of the core. Likewise, for a real lever, with nonzero mass, there is an additional component of input force, leading the velocity by 90°, that pumps mechanical kinetic energy in and out of the lever. For both the transformer and the lever the average reactive power is zero but the excitation current or force can be considerable.

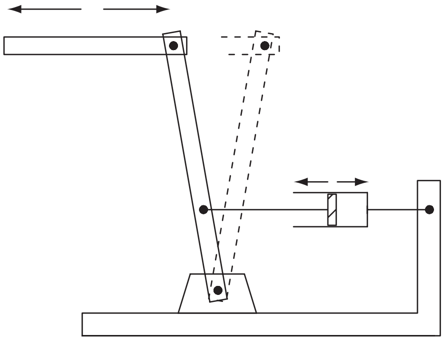

Figure 14.4. Mechanical analogue of a transformer.

14.5 Magnetizing inductance used in a transformer-coupled amplifier

In Chapter 3 we saw a circuit whose operation cannot be explained if its transformer is modeled as an ideal transformer. That circuit, a transformer-coupled class-A amplifier, is shown in Figure 14.5. Since the transformer windings have almost no dc resistance, the average voltage at the collector must be V_{cc}. Under maximum signal conditions the collector voltage swings between 0 and $2V_{dc}$, applying a peak-to-peak voltage of $2V_{dc}$ to the transformer primary.

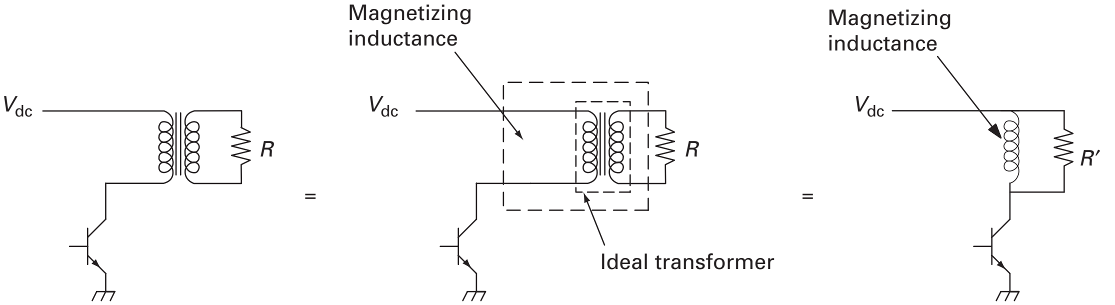

Figure 14.5. Transformer-coupled amplifier.

If we do not include the magnetizing inductance, the transformed load is a pure resistance. We would mistakenly conclude that the quiescent collector voltage must be $V_{cc}/2$ rather than V_{cc} and that the largest peak-to-peak collector signal would be V_{cc} rather than $2V_{cc}$. We would also conclude incorrectly that the frequency response would be unlimited, rather than being limited at low frequencies by the magnetizing inductance and limited at high frequencies by leakage inductance.

14.6 Double-tuned transformer: making use of magnetization and leakage inductances

Leakage inductance and magnetizing inductance limit the performance of transformers used in audio and other baseband applications. But in RF work these parasitic inductances can be tuned out with capacitors. Sometimes the leakage and magnetizing inductance can be intentionally used as in the band-pass filter of Figure 14.6(a).

To see how this circuit works, consider Figure 14.6(b), where the transformer has been replaced by its equivalent circuit. Only the leakage and magnetizing inductances are shown; the ideal transformer in the equivalent circuit of the transformer is either one-to-one or the resistor and capacitor on the right-hand side have been multiplied by its ratio. This equivalent circuit, with its vertical

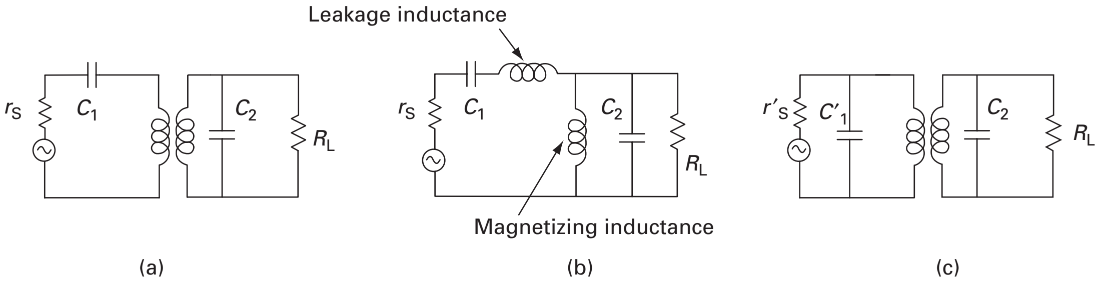

Figure 14.6. (a) Bandpass filter made with a loosely-coupled transformer; (b) equivalent circuit; (c) alternate circuit with two shunt capacitors.

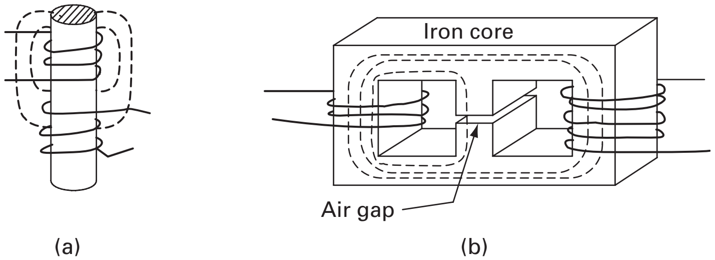

Figure 14.7. Loosely coupled transformers: (a) with powdered iron core; (b) with iron core for low frequencies.

parallel resonator and its horizontal series resonator, is a canonical two-section bandpass filter, as discussed in Chapter 4. The transformer, to have enough intentional leakage inductance, may be air-wound or may be wound on a permeable rod as shown in Figure 14.7(a). The alternate circuit of Figure 14.6(c) uses a parallel capacitor on each side of the transformer. The values of r'_S and C'_1 can be determined from the values of r_S and C_1 by noting that the equivalent Thévenin generator containing r'_S and C'_1 must have an impedance equal to the impedance of $r_S + jC_1$. (Note: whenever you see capacitors across both windings of a transformer, you can guess that the coupling is less than unity – otherwise the two capacitors would be effectively in parallel, and a single capacitor could be used.)

Power transformers are sometimes designed to have intentional leakage inductance to provide short-circuit protection (the leakage inductance limits the current). One way to build an iron core transformer with leakage inductance is shown in Figure 14.7(b). The magnetic path containing the air gap effectively shunts some of the flux generated by one winding from reaching the other winding. This design also provides magnetic shielding, in that the fringing fields are contained within the body of the transformer.

14.7 Loss in transformers

Large power distribution transformers are designed to have efficiencies around 99%. Inexpensive transformers are designed with enough efficiency to avoid premature burn-out. (Plug-in “wall-wart” transformers can be hot to the touch even without a load.) High-frequency transformers have efficiency limits set by core materials and the “skin effect” that excludes high-frequency currents from the interior of conductors, thus increasing the effective resistance of the windings.

Resistance of the windings (“copper loss”) is an obvious loss mechanism. We have seen that this can be included easily in a transformer equivalent circuit by simply putting a resistor in series with the primary and another in series with the secondary.

A magnetic core made of a conductive material such as iron will dissipate energy as ordinary I^2R loss since closed paths in the iron core will act as shorted turns around magnetic flux lines. To minimize these *eddy current* losses, any such closed paths are kept short by making the core a stack of thin sheet iron laminations separated by insulating varnish or oxide. The core of a toroidal transformer can be a bundle of insulated iron wire rings, a stack of varnished sheet metal toroids, or a toroid wound of varnished thin metal tape. High-frequency transformers use cores made of magnetic particles, held together in an inert binder material. Eddy current losses can be included in the transformer equivalent circuit as a resistor in parallel with the ideal transformer.

The final loss mechanism comes from magnetic hysteresis in the core. Ideally, the magnetic flux density is proportional to the magnetic force, $B = \mu H$, where μ is the permeability. But B often lags H, as magnetic domains exhibit a kind of static friction before they break loose and reverse their direction. As a result, B vs. H through the ac cycle forms a closed curve whose included area is the energy loss per unit volume per cycle. Hysteresis loss is associated with the magnetizing current, since it is the magnetizing current that produces H, which induces B. We can therefore include hysteresis loss in the equivalent circuit as a resistor in parallel with the magnetizing inductance. This is satisfactory for a power transformer, where the magnetizing current is constant and the frequency is constant. But note that the hysteresis loss is proportional to frequency, since the B-H loop is traversed once per cycle. It also has a nonlinear amplitude dependence. Thus a simple resistor is not adequate to model hysteresis loss in a wideband transformer or a transformer operating over a range of input voltages.

14.8 Design of iron-core transformers

A transformer designer usually strives to find the smallest, lightest, and least expensive (usually synonymous) transformer that conforms to a set of electrical specifications. To see the issues involved, let us consider the design of a 60-Hz

Figure 14.8. Iron-core transformer geometry.

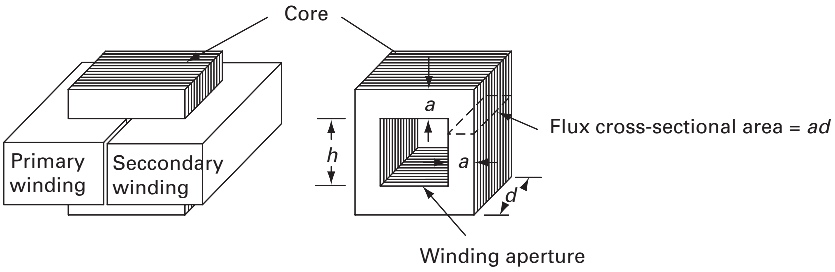

power transformer. Suppose the primary voltage is 220 volts rms and the power delivered to the load is 500 watts. The efficiency is to be 96% and the magnetizing current must be no greater than the "working" (in-phase) current.

We will pick a silicon-steel core material for which the maximum flux density before saturation, $B_{\max}$, is 1.5 webers/m^2. The core, a square toroid, is shown in Figure 14.8. For minimum copper loss (neglecting the excitation current) the primary and secondary windings will have equal loss and will each occupy half of the winding aperture.

The transformer will be specified by four parameters: the number of turns on the primary, N, and the three linear dimensions of the core, a, h, and d. To determine these four parameters we must write equations for the maximum B field, the loss, and the inductance of the primary winding. Faraday's law of induction gives us the maximum B field:

$$V_{\max} = N\frac{\mathrm{d}\phi}{\mathrm{d}t} = N\omega a d B_{\max}. \tag{14.3}$$

Since the rms primary voltage is 220, $V_{\max} = 220\sqrt{2}$ and we have

$$B_{\max} = \frac{220\sqrt{2}}{\omega N a d} < 1.5\,\text{webers/m}^2. \tag{14.4}$$

The copper loss in the primary winding will be I^2R_{p} where I is the rms current in the primary and R_{p} is the resistance of the winding. The number of turns on the primary, N, is given by

$$N = \frac{(h^2/2)}{\sigma} \tag{14.5}$$

where $h^2/2$ is the winding area for the primary and σ is the cross-sectional area of the wire. The mean length per turn is given by $2(a+d+h)$ so the primary resistance is found to be

$$R_{\mathrm{p}} = \rho\frac{\text{length}}{\sigma} = \frac{4\rho N^2(a+d+h)}{h^2}. \tag{14.6}$$

As for core losses, the 60-Hz loss for the selected core material (when $B_{\max}$= 1.5) is 0.6 watts/lb = 11 000 watts/m^3. The overall loss is the sum of the winding losses and the core loss. Since this loss is to be 500(1−0.96) = 20 watts, we can write

$$\text{Loss}_{\text{watts}} = 11\,000(4ad(a+h)) + \left(\frac{500}{220}\right)^2 \frac{4\rho N^2(a+d+h)}{h^2} = 20. \quad (14.7)$$

Finally, the specification on the magnetizing current is equivalent to specifying that the reactance of the primary, ωL, is greater than the equivalent input load resistance, or

$$\omega L \geq \frac{220^2}{500}. \quad (14.8)$$

The inductance of the primary, L, can be written as

$$L = \frac{\mu N^2(\text{flux area})}{\text{mean flux path length}}. \quad (14.9)$$

The mean flux path length, from Figure 14.8, is $4(h+a)$, so

$$L = \frac{\mu N^2 ad}{4(h+a)}. \quad (14.10)$$

We must use Equations 14.4, 14.7, and 14.8 to find transformer parameters, a, d, h, and N that will satisfy the given specifications and minimize the size of the transformer. This is not quite as simple as solving four equations in four unknowns. The equations are really inequalities and, in general, there will not be a solution that simultaneously produces the maximum allowable flux density, the maximum allowable loss, and the minimum allowable inductance. Instead, this problem in linear programming is most often solved by cut-and-try iterative methods, conveniently done using a spreadsheet program. In this particular example, such a procedure led to the set of parameters: $d = 5$ cm, $a = 2$ cm, $h = 5$ cm and $N = 580$ turns. These dimensions give a core weight of 5.1 lbs, a loss of 19.3 watts and B_{max} of 1.42. The reactance of the primary, assuming a relative permeability of 1000, is 5.9 times the input load resistance – five times more than the minimum required reactance. Note: in the equations presented above, no consideration was made for the space occupied by wire insulation and lamination stacking but these can be accounted for by simply increasing the value of the winding wire resistivity, ρ, and decreasing the permeability.

14.8.1 Maximum temperature and transformer size

The heat generated by a transformer makes its way to the outside surface to be radiated or conducted away. The interior temperature buildup must not damage the insulation or reach the Curie temperature where the ferromagnetism quits (a consideration with high-frequency ferrite cores). “Class-A” insulation materials

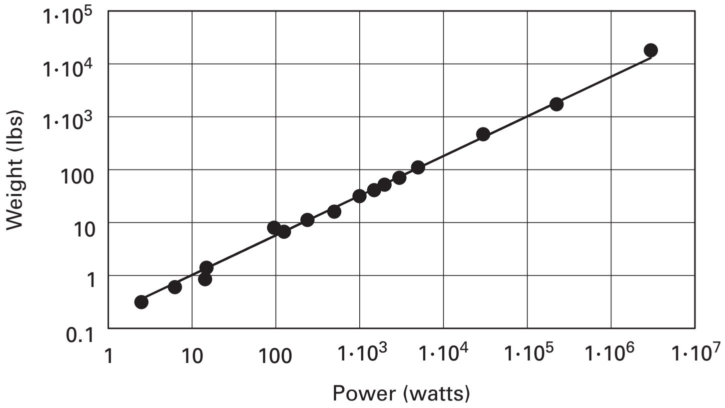

Figure 14.9. Weight vs. power rating for 16 commercial 60-Hz power transformers. The data points fit the solid line, for which (weight in lbs) = 0.18 (power in watts)$^{3/4}$.

(cotton, silk, paper, phenolics, varnishes) are limited to a maximum temperature of 105 °C. If reliable theoretical or empirically determined equations are available to predict internal temperatures, they can be included in the iterative design procedure described above. Rules of thumb can be used, at least as a starting point to determine core sizes for conventional transformers. One such rule for transformers up to, say, 1 kW is that the flux cross-sectional area of the core ($a \times d$ in Figure 14.8) in square inches should be about $0.25\sqrt{\text{Power in watts}}$. Transformer weight vs. power rating (from catalog specifications) is plotted in Figure 14.9 for sixteen 60-Hz power transformers, ranging from 2.5 watts to 3 megawatts.

The solid line, which fits the data, shows that the weight is proportional to the power raised to the exponent 3/4. Transformer manufacturers seem to use the rule-of-thumb that makes core area proportional to power$^{1/2}$ since this results in the volume and weight being proportional to power$^{3/4}$.

14.9 Transmission line transformers

Leakage inductance and distributed capacitance eventually determine the high-frequency limit of conventional transformers. For wideband applications we cannot simply resonate away these parasitics. Wideband *transmission line transformers* [3] are built like ordinary core-type transformers except that the windings are made with transmission line – either a coaxial cable, as shown in Figure 14.10, or a bifilar winding. The core must have high permeability but modest loss is acceptable (cores in chokes and transformers store little energy so high Q is not necessary). The effect of the core is to choke off any common-mode current in the transmission line, leaving only differential currents. When the transformer is wound with a piece of coaxial cable, as shown in Figure 14.10, the core suppresses current flowing on the outside of the shield,

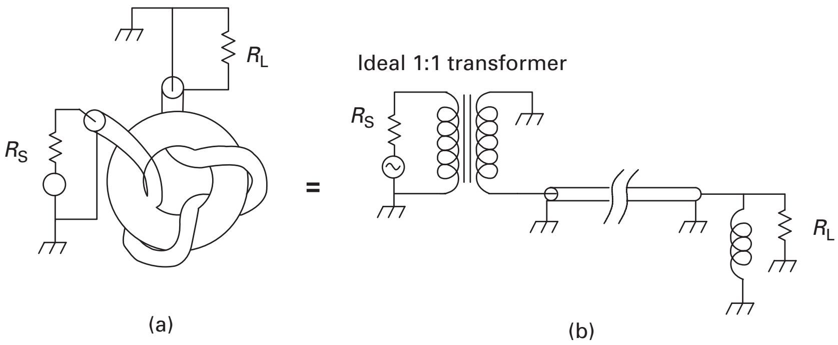

Figure 14.10. Transmission line reversing transformer.

leaving only the equal and opposite currents on the inside of the shield and the inner conductor.

This circuit is a reversing transformer, i.e., $V_{\text{out}} = -V_{\text{in}}$. The polarity flip is achieved by reversing the transmission line connections at the load end where the center conductor is grounded. Normally this would simply short the generator, but the inductance provided by the magnetic core chokes off the otherwise short-circuit current. An equivalent circuit model is shown in (b). Here the reversal is done with an ideal transformer. The length of the coax is the same, to duplicate the additional phase shift between the generator and the load. The inductor in parallel with the load represents the inductance of the cable winding around the core. At the lowest frequencies, this inductor diverts current from the load, just as the magnetization inductance limits the low-frequency response of a conventional transformer. At high frequencies, however, the circuit becomes just a piece of transmission line and its response does not fall off. There is effectively no leakage inductance nor stray capacitances. The time lag through the transmission line, however, will shift the phase from the nominal 180° as the frequency increases. Nevertheless, if an application calls for a pair of signals, identical except for polarity, the "reference" signal can be provided by using an identical piece of transmission line to provide an identical delay.

Transmission line transformers extend the range of ordinary transformers by two octaves or more. In addition to this reversing transformer, many other transformers can be made with the transmission line technique [3, 4]. Commercial hybrids good from 0.1 MHz to 1000 MHz use transmission line transformers. Miniature transmission line transformers are commercially available as standard components.

14.10 Baluns

A balun is any device that converts a *bal*anced (double-ended symmetric) signal into an *un*balanced (single-ended) signal. Baluns are commonly used to feed

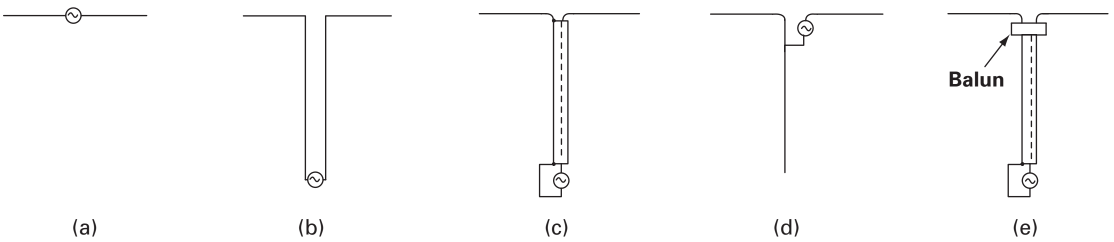

Figure 14.11. A symmetric dipole antenna fed (a) at the antenna feed point, (b) with a balanced feedline, (c) with an unbalanced coaxial feedline, (d) equivalent circuit for (c) showing how the antenna is modified by current on the feedline, (e) Balun provides symmetric feed.

symmetric antennas (e.g., dipoles) from unbalanced coaxial feed lines. Figure 14.11 shows what happens if we feed a dipole directly with a coaxial transmission line. In (a), the generator is at the feed point of the dipole so there is no question of balance or imbalance. In (b), a balanced feedline is used. Everything is still symmetric. At any point along the feedline, the current in one side is equal and opposite to the current in the other side. The spacing between the conductors is very small compared to the wavelength, so “cancellation” assures there is negligible radiation from the line. In (c) a coaxial line feeds the dipole improperly; the shield of the coax tied to the left-hand element of the dipole. An equivalent circuit (d) shows how the outer conductor of the coax becomes part of the left-hand dipole element. The antenna now has one straight element and one L-shaped element. The radiation pattern will not be the intended dipole pattern and there will be RF current flowing on the outside of the feedline. In (e) a balun at the end of the coaxial feedline provides equal and opposite voltages to each side of the dipole and eliminates any current from the outside of the feedline.

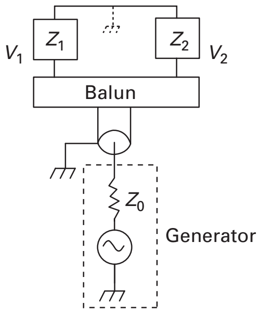

Figure 14.12. Balun operation: unbalanced-to-balanced.

Figure 14.12 illustrates the requirement for a balun; with equal Z_1 and Z_2, i.e., a load structure symmetric with respect to ground, we want V_1 and V_2 to be equal and opposite with respect to ground.

The dotted ground symbol indicates that this point of symmetry will have zero voltage (when $Z_1 = Z_2$) and can be grounded if necessary or desirable. Figure 14.13 shows the equivalent situation with the load at the unbalanced side. When V_1 and V_2 are in phase (common mode) there must be no excitation of Z. But when V_1 and V_2 are 180° apart (differential mode) the load, Z, is fully excited. Baluns are normally reciprocal devices so the name “Unbal” is not needed.

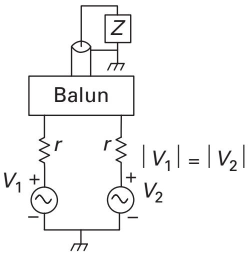

Figure 14.13. Balun operation: balenced-to-unbalanced.

From a transmitting standpoint, the balun eliminates common mode current on the feedline which otherwise would radiate and affect the pattern of the antenna. Figure 14.14(a) shows a reversing transformer used as a balun for this application. Note that this balun is also a 4-to-1 impedance transformer since the voltage across the dipole is $2V$. In Figure 14.14(b), the reversing transformer is replaced by a half-wave length of transmission line. The phase shift through this piece of line transforms V into $-V$ just as the transformer did, but the reversal is only correct over a narrow frequency band. (Note that the half-wave line can have any value for Z_0.)

The simple reversing transformer in Figure 14.14(a) can be replaced by the wideband transmission line transformer of Figure 14.10. The phase shift this transformer picks up at increasing frequencies is compensated by using a

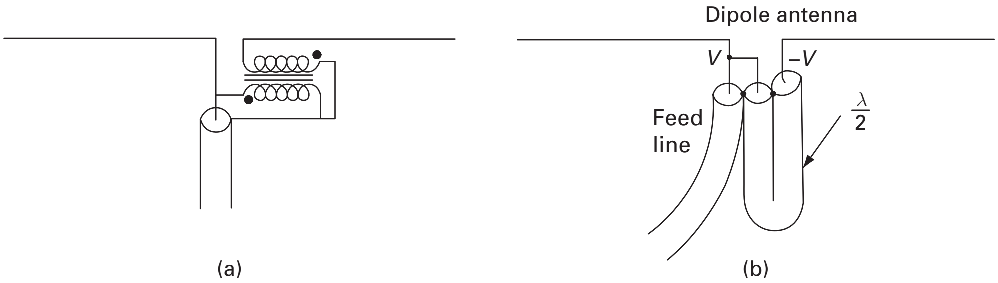

Figure 14.14. 4:1 baluns.

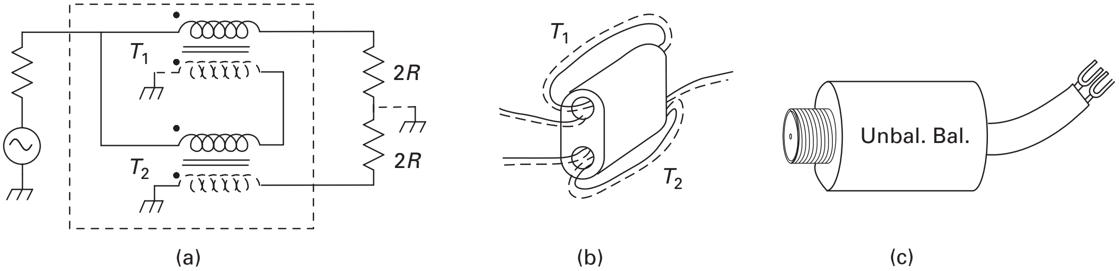

Figure 14.15. Wideband 4:1 balun made from two 1:1 transmission line transformers.

second transformer, identical except with no reversal, to provide an equal frequency-dependent phase shift. The combination of these transformers, connected in parallel at one end and in series at the other end, as shown in Figure 14.15(a), makes a very wideband 4-to-1 balun.

This is the circuit most often found in the television balun of Figure 14.15(c). The transformers are often wound on a “binocular core” (Figure 14.15b). This core operates as two separate cores, i.e., there is nominally no magnetic coupling between the two transformers, T_1 and T_2. For clarity, the figure shows the transformers wound with only two turns; in practice several turns are used.

Problems

Problem 14.1. Use Equations (14.1) and (14.2) to show that when the primary and secondary windings of a transformer are connected in series the total inductance is given by $L = L_1 + L_2 \pm 2M$ where M, the mutual inductance, is given by $M = k\sqrt{L_1 L_2}$ and the ± changes when one of the windings is reversed. (This is a standard method for measuring mutual inductance.)

Problem 14.2. Consider the following transformer: $L_1 = 0.81\,\text{H}$ (inductance of the primary winding), $L_2 = 1\,\text{H}$ (inductance of the secondary winding), $k = 0.9$ (coupling coefficient).

(a) If the secondary is open circuited and 1 volt (ac, of course) is applied to primary, show that the secondary voltage is 1 V. (b) If the primary is open circuited and 1 volt is applied to the secondary, show that the primary voltage is 0.81 V.

Problem 14.3. Calculate the low-frequency cutoff (half-power frequency) for a resistive load coupled by a particular transformer to a generator. The transformer has perfect coupling and a 1:1 turns ratio. The source and load impedances are both 100 ohms and the reactance of the transformer primary is 50 ohms at 20 Hz.

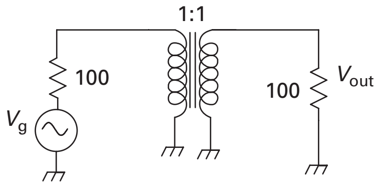

Problem 14.4. Upgrade your ladder network analysis program (Problem 1.3) to handle conventional transformers. Let the transformer be specified by its primary inductance, secondary inductance, and coupling coefficient.

Example answer: For the MATLAB example solution given in Problem 1.3, add the element, “XFRMR” by inserting the following sequence of statements in the “elseif chain”:

```
elseif strcmp(component,‘XFRMR’)==1
ckt_index=ckt_index+1; Lpri=ckt{ckt_index}; %primary inductance
ckt_index=ckt_index+1; Lsec=ckt{ckt_index};% secondary inductance
ckt_index=ckt_index+1; k =ckt{ckt_index}; %coupling coefficient
V=V+I*(1j*w*Lsec*(1-k^2));
ratio= sqrt(Lpri/(k^2*Lsec));
V=V*ratio; I=I/ratio;
I=I+V/(1j*w*Lpri);
```

Problem 14.5. A lossless transformer is placed between a 50-ohm signal generator and a 4.5-ohm load. (a) Use your ladder network analysis program (or an equivalent program) to plot the relative power at the load vs. frequency. Use the following parameters: primary inductance=100 μH, secondary inductance=10 μH, coupling coefficient k=0.9. (b) Find the values of a capacitor to be shunted across the primary (i.e., in parallel with the magnetizing inductance) and another capacitor to be placed in series with the secondary (i.e., in series with the leakage inductance) so that the magnetizing and leakage inductances will be cancelled (resonated out) at 0.5 MHz. Plot the resulting frequency response to verify that the transmission is now perfect at 0.5 MHz.

Problem 14.6. When a power transformer is first turned on, i.e., connected to the line, there is sometimes an initial inrush of current strong enough to dim lights on the same circuit and produce an audible “grunt” from the transformer itself. Decide whether this effect is strongest when the circuit is closed at a zero crossing of the line voltage or at a maximum of the line voltage. (This involves the magnetizing inductance of the transformer so simply analyze the transient when an inductor is connected to an ac line.)

Problem 14.7. (a) Suppose you have a power transformer designed to be fed from 220 V, 60 Hz but you want to use it in a country where the power line supplies 220 V, 50 Hz. Why is the transformer likely to overheat when fed with 50 Hz power? Consider the magnetizing current, copper losses, and core losses. (b) Consider the reverse situation. Would there be any mechanism to cause extra power dissipation if a 50 Hz transformer is used on a 60 Hz line?

Problem 14.8. Two identical perfectly coupled 1:1 transformers are connected in series, i.e., the secondary of the first is connected to the secondary. The primary and secondary inductances of each transformer are L. Show that this combination is equivalent to a single transformer and find its magnetizing inductance L'. If you enjoy algebra, assume the transformers are not perfectly coupled and find L' and k'.

Problem 14.9. Consider a lossless transformer with primary and secondary inductances L_1 and L_2. Suppose the coupling coefficient has a value that results in a 1:1 ideal transformer in the transformer's equivalent circuit. Find the values of the magnetizing inductance and the leakage inductance.

Problem 14.10. The transformer in the figure has a turns ratio of 1:1. The primary and secondary inductances are both L. The amplitude of the sine wave from the generator is V_0. Assume the transformer has no leakage inductance. Find an expression for the current in the resistor. Hint: use the equivalent circuit for the transformer: an inductor together with an ideal transformer.

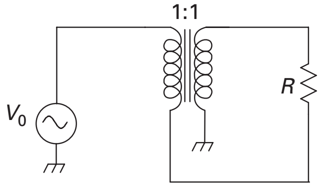

References

[1] Flanagan, W. M. *Handbook of Transformer Design & Applications*, 2nd edn. New York: McGraw-Hill, 1993.

[2] McLyman, Col. W. T. *Transformer and Inductor Design Handbook*, 3rd edn, Bora Rotan: CRC Press, 2004.

[3] Ruthroff, C. L., Some broadband transformers, *Proceedings of the IRE*, August 1968, pp 1357–1342.

[4] Sevick, J., *Transmission Line Transformers*, Newington CT: American Radio Relay League, 1987.

CHAPTER

15

Hybrid couplers

Hybrid couplers, also known as hybrid junctions or simply "hybrids," are lossless passive four-port devices used to make interconnections between circuit elements. Hybrids are used as power dividers ("signal splitters") and combiners. They are also used in mixers and TR (transmit/receive) switches. A useful schematic representation for a hybrid, Figure 15.1, shows the four connection points (ports).

RF hybrids usually have unbalanced ports designed for coaxial transmission lines, so all four ports share a common ground, indicated in Figure 15.1 by a dotted ground symbol (usually not shown). Each port has a characteristic impedance. Most packaged RF hybrids with coaxial ports are made so that the characteristic impedance of all four ports is 50 or 75 ohms. The symbol in Figure 15.1 shows signal flow paths; power incident on Port 1 splits and exits through Ports 2 and 3. If both of these ports are properly terminated there will be no reflections and the impedance seen looking into Port 1 will be equal to the characteristic impedance of that port. In this case no power will reach Port 4 as opposite ports are isolated. But if Port 2 and/or Port 3 are not terminated in their own characteristic impedances, the power exiting these ports will be partially or completely reflected back into the hybrid. The reflection, which depends on the mismatch, is calculated exactly as if the power had exited from a transmission line whose impedance is equal to that of the respective port. Any power reflected back into the hybrid splits and follows the signal paths, just as if it had come from an external source. You can see that, with arbitrary terminations and arbitrary signals, the situation could become complicated. But usually we deal with continuous wave (cw) sinusoidal signals so, rather than analyze multiple reflections in the time domain, we only have to solve for the forward and reverse wave amplitudes on each of the four internal paths. In most applications, things are even simpler; hybrids usually have proper terminations and the signal flows are simple and can be determined by inspection of the signal flow diagram.

15.1 Directional coupling

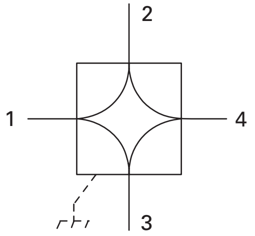

Figure 15.1. Schematic symbol for a hybrid coupler.

From inspection of the signal paths in Figure 15.1, we see that with the ports matched and with power flowing from Port 1 to Port 2 there will also be power flowing out of Port 3 but none out of Port 4. If the power is now reversed, to flow from Port 2 to Port 1, there will be power flowing out of Port 4 but none from Port 3. Port 3 is therefore coupled to power flowing from 1 to 2. Likewise, Port 4 is coupled to power flowing from 2 to 1. Therefore, a hybrid is a *directional coupler*, and can be used to determine how much power is flowing in each direction on a transmission line. Here we will use the term *hybrid* only for 3-dB directional couplers, i.e., directional couplers that split the incident power in half.

15.2 Transformer hybrid

The name *hybrid transformer* was first applied around 1920[1] to the simple center-tapped transformer shown in Figure 15.2. The primary winding has N turns while each half of the secondary winding has $N/\sqrt{2}$ turns. For this transformer hybrid (hybrid composed of a transformer), the characteristic impedances of Ports 1, 2, and 3 are equal and are twice the impedance of Port 4. (Here the port impedances are R, R, R, and $R/2$, where the value of R is arbitrary, as long as the transformer behaves as an ideal transformer, i.e., its magnetizing inductance has a reactance substantially larger than R.)

Let us confirm that this circuit has the power splitting and isolation characteristics of a hybrid. First consider a signal connected to Port 1. If Ports 2 and 3 have identical terminations, they will have equal and opposite voltages. The voltage at Port 4, since it is midway between the voltages at Ports 2 and 3, must be zero and Port 4 is indeed isolated from Port 1. Next note that a signal applied to Port 4 will appear unchanged at Ports 2 and 3 but will not appear at Port 1. (The currents to Ports 2 and 3 are in opposite directions so there is no net flux in the transformer to provide a voltage at Port 1 or produce an IX_L drop.) You can verify that Ports 2 and 3 are also isolated from each other (see Problem 15.1). Figure 15.2 also shows the symbol appropriate

Figure 15.2. Transformer hybrid.

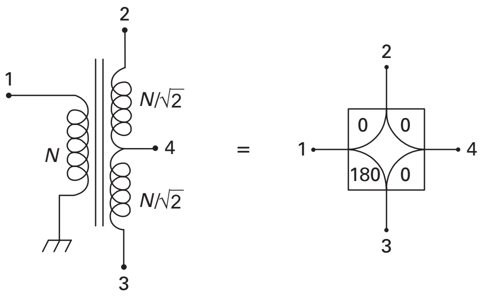

---

[1] The origin of the term seems to be lost – a hybrid of what? – but it came from the telephone industry, where the terms *hybrid transformer* and *hybrid coil* were both common.

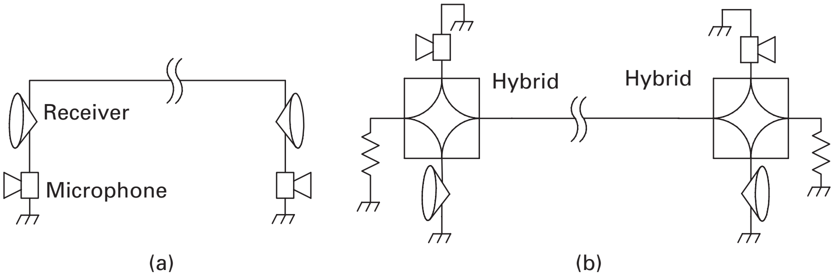

Figure 15.3. Hybrids allow full-duplex communication over a single line.

for this hybrid. The labels 0, 0, 0, and 180 indicate the phase shifts through the respective branches. A signal incident on Port 1, for example, appears at Port 2 with the phase unchanged (shifted 0°) and at Port 3 with its polarity inverted (shifted 180°). Any hybrid with these four phase shifts is called a *180° hybrid*.

15.2.1 Applications of the transformer hybrid

In telephony or other wired communication, hybrids allow a transmission line to carry independent signals in each direction. Figure 15.3 shows two telephone circuits. In the simple series circuit of Figure 15.3(a), each user hears his own voice as well as the voice from the other end. In the circuit of Figure 15.3(b), hybrids isolate each receiver from its own microphone.

If we are using the transformer hybrid of Figure 15.2, we would terminate Port 4 with a resistor of value $Z_0/2$, where Z_0 is the characteristic impedance of the phone line. The microphones and receivers must have impedances equal to Z_0. This arrangement provides two-way signaling , "full duplex," over a single cable.[2]

The circuit of Figure 15.4 uses two hybrids and two amplifiers to make a bidirectional repeater for a long (lossy) line. Here the hybrids let the signals in each direction be independently amplified without feedback and consequent oscillation.

It is convenient to make the characteristic impedances be the same for all four ports of a general-purpose hybrid. The transformer hybrid, fixed up to have equal impedances, is shown in Figure 15.5. This is the kind of circuit found inside an off-the-shelf wideband 3-dB hybrid. Transformer hybrids made with toroidal cores (ferrite beads) can work over large bandwidths, e.g., 10 KHz to 20 MHz and 1 MHz to 500 MHz.

---

[2] In telephony, the circuit is deliberately unbalanced – just enough so that the users, hearing their own voices or ambient noise, will sense that the call is connected, but not enough that the users hold the receiver (and hence the microphone) away from their heads. Of course cancellation should be as great as possible when this kind of full-duplex circuit carries two-way digital data.

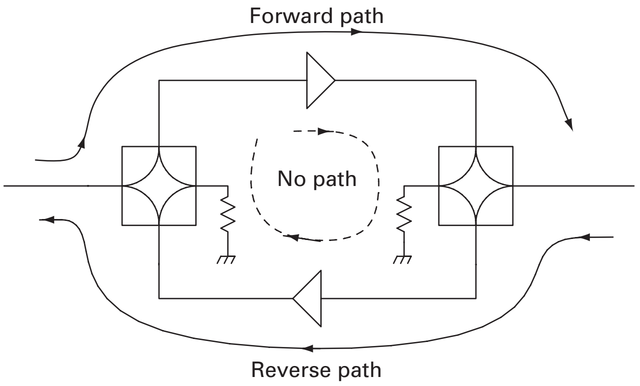
Forward path
No path
Reverse path

Figure 15.4. Two-way telephone repeater for long lines.

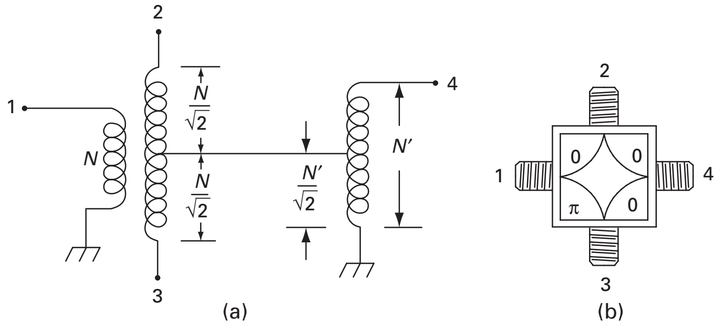
2
1
N
$\frac{N}{\sqrt{2}}$
$\frac{N}{\sqrt{2}}$
$\frac{N'}{\sqrt{2}}$
N′
4
3
2
1
0
0
π
0
4
3

(a) (b)

Figure 15.5. Two-transformers make a hybrid with the same impedance at all four ports.

Hybrids are often used in circuits like those illustrated above, as well as signal splitting and combining, where one port is terminated in its characteristic impedance. An easy way to terminate Port 1 of the hybrid of Figure 15.5 is to put a resistor of value $2Z_0$ between Ports 2 and 3. When this is done, the Port 1 winding on the hybrid transformer can be eliminated. The resulting circuit, shown in Figure 15.6a, is

Figure 15.6. Internally terminated hybrid is a two-way splitter/combiner.

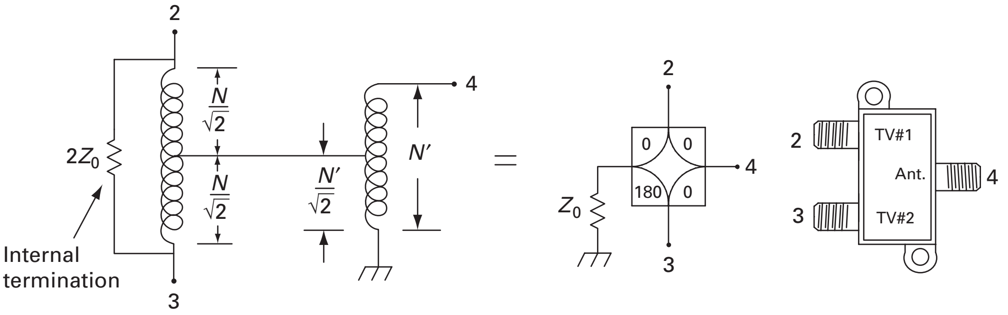
2
$2Z_0$
Internal termination
3
$\frac{N}{\sqrt{2}}$
$\frac{N}{\sqrt{2}}$
$\frac{N'}{\sqrt{2}}$
N′
4
=
2
0
0
180
0
4
Z_0
3
2
TV#1
Ant.
4
3
TV#2

(a) (b) (c)

commonly found in the signal splitters used to connect two receivers to a single antenna (c) and in other packaged 2:1 splitter/combiners. When it is acceptable for the impedance of Port 4 to be $Z_0/2$, the right-hand transformer can be omitted, and the hybrid consists only of the center-tapped *hybrid coil*.

15.3 Quadrature hybrids

The transformer hybrid is naturally a 180° hybrid. Other circuits are natural 90° hybrids, the symbol for which is shown in Figure 15.7.

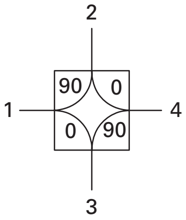

Figure 15.7. Symbol for a 90° hybrid.

Let us look at an interesting application of the 90° hybrid (often called a *quadrature hybrid*). Here the internal phase paths are zero and 90°. A 90° path means a phase shift equal to that produced by a quarter-wave length of cable. For circuit analysis, one deals with hybrids in terms of voltages. We will consider the hybrid to be connected to transmission lines (of the same impedance as the hybrid) in order to describe the signals in terms of incident and reflected waves. A signal incident at Port 1 will be split equally into signals exiting Ports 2 and 3. Since the power division is equal, the magnitudes of the voltages of the signals exiting Ports 2 and 3 will be $1/\sqrt{2}$ times the magnitude of the incident voltage. The phases of the exiting signals will be delayed as indicated on the symbol for the hybrid. For the hybrid of Figure 15.7, the signal exiting Port 3 has no additional phase shift but the signal exiting Port 2 is multiplied by $e^{-j\pi/2}$. Suppose a signal is also incident at Port 4. It will also split into signals exiting from Ports 2 and 3. The total voltage of the waves exiting Ports 2 and 3 is just the superposition of the waves originating from Ports 1 and 4.

15.3.1 Balanced amplifier

A common application for 90° hybrids is the balanced amplifier circuit shown in Figure 15.8. As long as the two amplifiers are identical they can have arbitrary input and output impedances but the overall circuit will have input and output impedances of Z_0.

To see how this happens, suppose that the hybrids are 50-ohm devices but that the input impedance of the amplifiers is not 50 ohms. Imagine that the interconnections are made using 50-ohm transmission line. The input lines have equal lengths and the output lines have equal lengths. The amplifiers are identical so the two signals have equal phase changes upon reflection. An input signal is split by the input hybrid; half the power will be incident on the top amplifier and half on the bottom amplifier. But reflections from the amplifiers will be out of phase by 180° when they arrive back at the input of the hybrid because the signal on the upper path will have made a round trip through the 90° arm of the hybrid. The two reflections therefore cancel and there is no net

Figure 15.8. A balanced amplifier has constant input and output impedances.

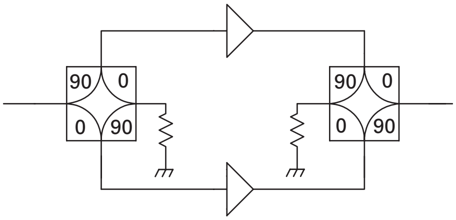

reflection. The input impedance of the overall amplifier will be the characteristic impedance of the hybrid (the value of the resistors, if transformer hybrids are used). The output side works the same way, and this combination of two arbitrary but identical amplifiers produces an amplifier with ideal constant input and output impedances.

15.4 How to analyze circuits containing hybrids

Let us work out an example problem to illustrate the way voltages are added in the manner of transmission line analysis. In Figure 15.9, a 50-ohm hybrid has arbitrary impedances, Z_2 and Z_3 terminating Ports 2 and 3. Port 4 is terminated in 50 ohms, so no power exiting Port 4 will be reflected back into the hybrid. We want to find the impedance seen looking into Port 1. Rather than work directly with impedances, we work with the equivalent reflection coefficients, $\Gamma = (Z - Z_0)/(Z+Z_0)$. Once we have found Γ_{in}, it can be converted to Z_{in}. The input signal is denoted by the forward arrow at Port 1. We will give it unity amplitude.

Figure 15.9. An example circuit problem: finding the impedance looking into Port 1.

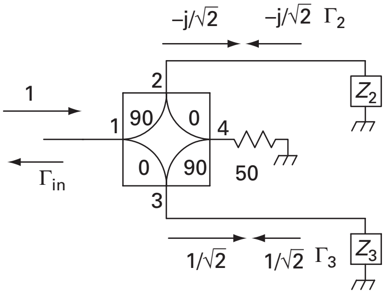

The signal incident on Z_2 is therefore $(1/\sqrt{2})e^{-\mathrm{j}\pi/2} = -\mathrm{j}/\sqrt{2}$ where $1/\sqrt{2}$ is the reduction in amplitude due to the equal power split and $-\pi/2$ is the phase shift of the 90° path between Ports 1 and 2. The signal reflected back into Port 2 is $[-\mathrm{j}\sqrt{2}]\Gamma_2$, i.e., the signal incident on Z_2 multiplied by the reflection coefficient of Z_2. This reflected signal will split as it enters Port 2 and its contribution to the wave leaving Port 1 will be its amplitude multiplied by $(1/\sqrt{2})e^{-\mathrm{j}\pi/2}$ as it is split and phase shifted by the 90° path or $[\,-\mathrm{j}(\sqrt{2})]\Gamma_2 \times -\mathrm{j}(\sqrt{2}) = -\Gamma_2/2$.

Figure 15.10. A balanced amplifier built with 180° hybrids.

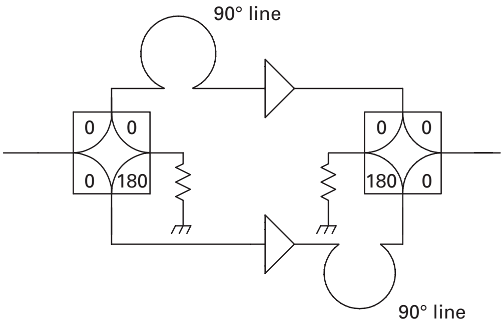

Figure 15.11. Conversions between 90° and 180° hybrids

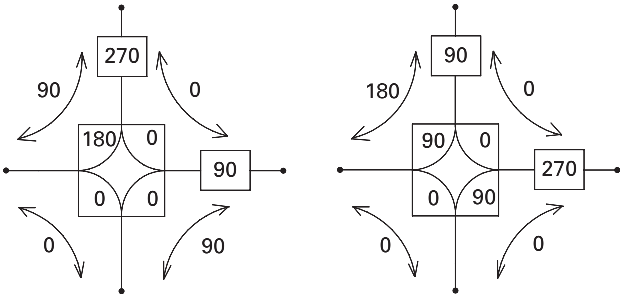

Similarly, the contribution to Γ from the reflection at Z_3 is Γ_3 /2, and the overall reflection is given by $\Gamma = -\Gamma_2/2 + \Gamma_3/2$, which solves the problem. The reader can work out the general case where Port 4 is also terminated by an arbitrary impedance, Z_4. Hint: let V_2, V_3, and V_4 be the wave amplitudes flowing out of Ports 2, 3, and 4. The waves flowing into these ports will therefore have amplitudes $\Gamma_2 V_2$, $\Gamma_3 V_3$, and $\Gamma_4 V_4$. By examination of the circuit, write three equations for V_2, V_3, and V_4, the three unknowns.

A balanced amplifier can also be built with 180° hybrids if two 90° lines are added as shown in Figure 15.10.

This use of cables to allow substitution of 180° hybrids for 90° hybrids in this circuit can be taken much farther; any hybrid can be converted to any other hybrid by adding lengths of transmission line to the ports. Figure 15.11 shows how to convert a 180° hybrid into a 90° hybrid and vice versa.

15.5 Power combining and splitting

An obvious way to power N loads from a single 50-ohm source is to transform the impedance of each load to $50N$ +j0 ohms and then connect the transformed

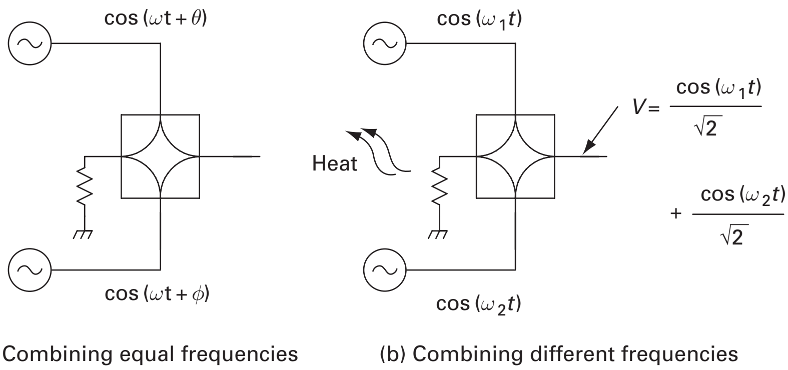
cos (ωt + θ)
cos (ωt + ϕ)
(a) Combining equal frequencies
cos (ω₁t)
cos (ω₂t)
Heat
V = cos (ω₁t)/√2 + cos (ω₂t)/√2
(b) Combining different frequencies

Figure 15.12. Hybrids used as power combiners.

loads in parallel across the generator. However, if any load changes, the power delivered to the other loads will change. A similar argument applies to combining the power from several sources: if any source changes amplitude or phase, the combined output will change. Hybrids provide a way to combine or split power without using impedance transformation and in a way that isolates multiple sources or multiple loads from each other.

To use a hybrid as a two-input power combiner, the unused port is terminated, either externally, as shown in Figure 15.12 or internally, as shown in Figure 15.6. The two signals to be combined should have equal amplitudes, as well as the correct phase relationship, to steer the total available power into the desired port. If the phase difference is changed by 180°, all the power will flow into the terminated port. Note that if one source fails, the other source will not know it; it still sees a matched load. This provides a fail-safe circuit, although the power output will drop by 75%.

When a hybrid is used to combine two signals of different frequencies, as in Figure 15.12(b), half the power of each signal will always be lost in the terminated port. Circuits to combine signals of different frequencies without loss are known as diplexers. (How would you make one?)

Power splitting is, of course, just time-reversed power combining. When hybrids are used as splitters and the source impedance is equal to the port impedance, the signal at any output port will remain constant when the loads on the other output ports are removed, shorted, or changed in any way.

15.5.1 Hybrid trees

Trees of hybrids are often used to make multi-input combiners or multiple-output splitters, as shown in Figure 15.13. High-power transmitters often use such tree structures to combine the power of many low-power solid-state amplifiers.

These trees have the same advantages as single hybrids, when used for power combining or splitting. A 2^N-to-1 combiner/splitter has $2^N - 1$ internal hybrids.

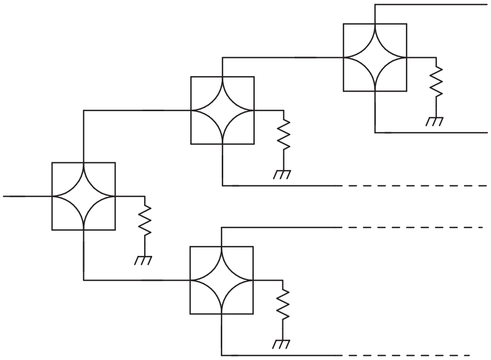

Figure 15.13. A tree of hybrids.

If you buy a three-way TV antenna splitter, you may find that each output provides only one quarter, rather than one third of the input power. What would have been a fourth output port is internally terminated. Or, if the splitter contains two hybrids, one output port can supply half the input power while the other two output ports can each supply one quarter of the input power.

15.6 Other hybrids

There are many ways to make hybrids without transformers. Most of them are circuits whose elements are transmission lines or capacitors and inductors. Unlike the ideal transformer hybrid, these hybrids are all frequency dependent – perfect hybrids only at their center frequency. However, most have a useful bandwidth of about an octave, which is usually sufficient for RF and microwave applications. You can use straightforward circuit analysis to analyze these hybrids.

15.6.1 Wilkinson power divider (or combiner)

This hybrid, shown in Figure 15.14 (in a 50-ohm version), uses two quarter-wave pieces of 70.7 $(50\sqrt{2})$ ohm transmission line. It has only three external ports; the fourth port is internally terminated, that is, connected to a load equal to its characteristic impedance.

It is easy to see that power applied to Port 1 will divide between Ports 2 and 3. By symmetry, the voltages at Ports 2 and 3 must be identical so no power is dissipated in the internal termination. Fifty-ohm loads at Ports 2 and 3 are transformed by the 90° cables to 100 ohms. The parallel connection of 100 ohms and 100 ohms at Port 1 produces the desired 50-ohm input impedance. The Wilkinson divider is usually classified as a 180° hybrid since its outputs have the same phase, even though this phase is 90° rather than 0°.

Figure 15.14. Wilkinson power divider.

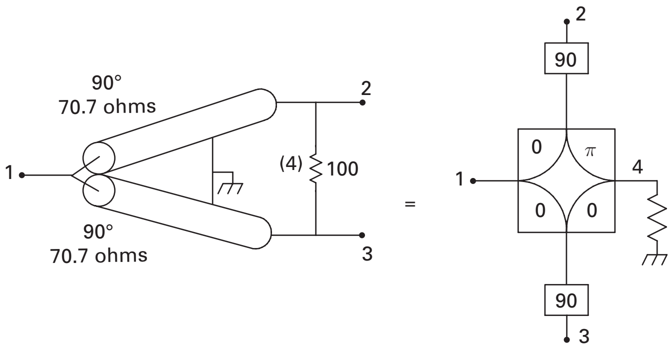
90°
70.7 ohms
90°
70.7 ohms
1
2
3
(4)
100
=
2
90
0
π
0
0
1
4
90
3

15.6.2 Ring hybrid

The ring hybrid, shown in Figure 15.15, uses four pieces of transmission line. To have 50-ohm ports, the hybrid must be built from 70.7-ohm transmission line.

Figure 15.15. Ring hybrid.

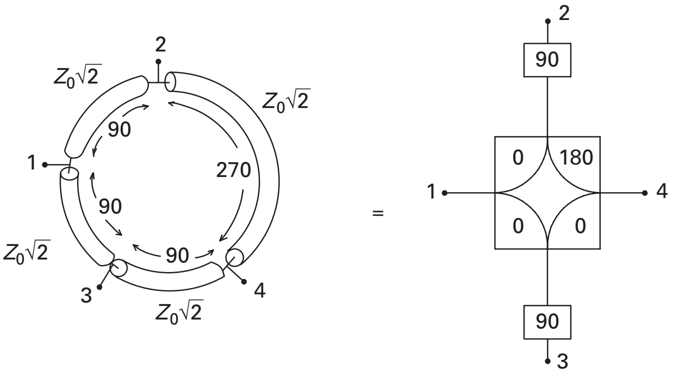
2
$Z_0\sqrt{2}$
$Z_0\sqrt{2}$
90
1
270
90
$Z_0\sqrt{2}$
90
3
4
$Z_0\sqrt{2}$
=
2
90
0
180
1
4
0
0
90
3

Branch-line hybrids

These are ladder networks made of quarter-wave lengths of transmission line. The simplest is shown in Figure 15.16. More complicated versions have more branches and provide more bandwidth.

Figure 15.16. Branch line hybrid.

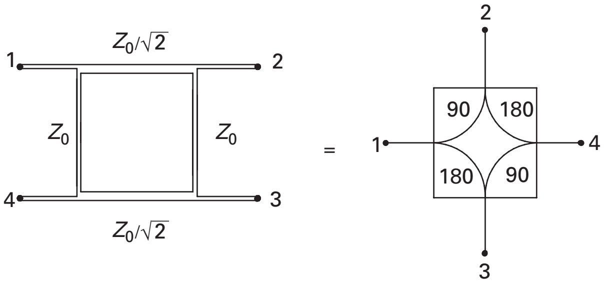
$Z_0/\sqrt{2}$
1
2
Z_0
Z_0
4
3
$Z_0/\sqrt{2}$
=
2
90
180
1
4
180
90
3

15.6.3 Lumped element hybrids

Two examples of 50-ohm lumped element hybrids are shown in Figure 15.17. These two circuits are obtained by replacing the Z_0 and $Z_0/\sqrt{2}$ arms of the simple branch line hybrid with the pi (or T) lumped circuit impedance inverters discussed in Chapter 13.

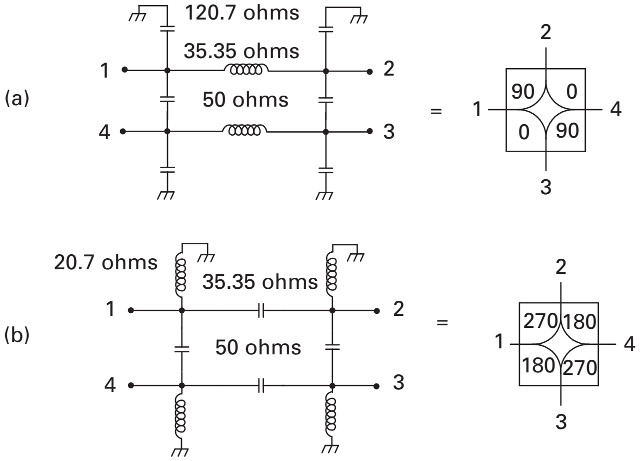

Figure 15.17. Two lumped-element hybrids.

15.6.4 Backward coupler

The *backward coupler* is shown in Figure 15.18 in shielded pair and stripline versions. The coupled transmission lines have both electric and magnetic coupling. When power flows from left to right, between Ports 1 and 2, coupled power flows left (backwards) out of Port 3. This type of coupler can be designed so that the voltage coupling coefficient, c, between Ports 1 and 3 is anywhere between 0 and 1. If designed for $c = 1/\sqrt{2}$, the coupler is a hybrid (a 3-dB directional coupler). This coupler can be analyzed in terms of its common mode and differential characteristic impedances, Z_{CM} and Z_{DIFF}, which are defined as follows. If the two inner conductors are driven together as a single (though split) center conductor, the characteristic impedance of the line is Z_{CM}. When the two inner conductors are regarded as a balanced shielded transmission line, its characteristic impedance is Z_{DIFF}. Two equivalent parameters are defined: $Z_{\mathrm{EVEN}} = 2\ Z_{\mathrm{CM}}$ and $Z_{\mathrm{ODD}} = Z_{\mathrm{DIFF}}/2$. Analysis using simultaneous forward and reflected common mode and differential mode signals (see Problem 15.8) produces the formulas:

$$Z_{\mathrm{ODD}} = Z_0([1-c]/[1+c])^{1/2} \text{ and } Z_{\mathrm{EVEN}} = Z_0([1+c]/[1-c])^{1/2} \quad (15.1)$$

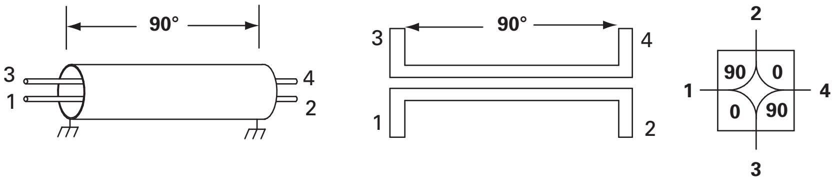

Figure 15.18. Backward couplers (coupled transmission line hybrids).

where Z_0 is the desired port impedance. However, these are only indirect design formulas, since electromagnetic analysis is needed to find physical dimensions to produce the required values for Z_{ODD} and Z_{EVEN}.

The *magic-T* of Figure 15.19 is a four-port waveguide junction that combines an *E*-plane tee junction with an *H*-plane tee junction. This waveguide hybrid, whose operation is surprisingly simple, is discussed in Chapter 16.

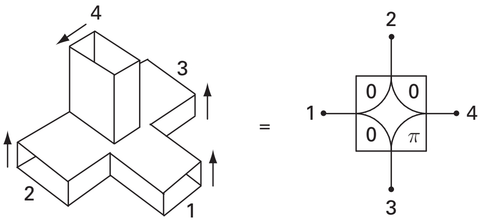

Figure 15.19. Waveguide magic-T hybrid.

Problems

Problem 15.1. Verify for the transformer hybrid of Figure 15.2 that Port 2 is isolated from Port 3, i.e., show that with a generator connected to Port 3, the voltage at Port 2 will be zero provided the impedance terminating Port 1 is twice the impedance terminating Port 4.

Problem 15.2. Explain why the duplex telephone circuit and the two-way telephone repeater could be built with either 90° or 180° or any other hybrids.

Problem 15.3. (a) Calculate the overall gain of the balanced amplifier of Figure 15.8 if the (identical) individual amplifiers each have gain G_0. (Answer: G_0)

(b) Calculate the overall gain if one of the interior amplifiers is dead, i.e., has zero gain. (Answer: $G_0/4$.)

Problem 15.4. The 50-ohm hybrid shown below is properly terminated by 50-ohm resistors on two of its ports. The third port is terminated by a 25-ohm resistor. If a 50-ohm generator is connected to the remaining port, what fraction of the incident power will be reflected back into the generator? (Answer: 1/36.)

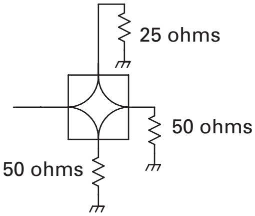

Problem 15.5. In the figure below, a 50-ohm hybrid is fed power from two amplifiers. The signals from these amplifiers have the same frequency and the same phase but the upper amplifier supplies 1 watt while the lower amplifier supplies 2 watts. In this seemingly unbalanced configuration, how much power reaches the load resistor? (Answer: 2.91 watts.)

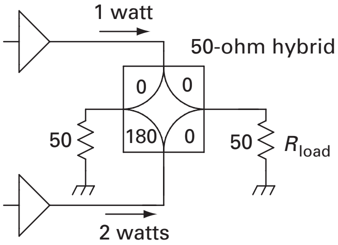

Problem 15.6. Four identical 1-watt amplifiers and six hybrids are used to make a 4-watt amplifier. If one of the interior amplifiers fails, how much power will be delivered to the load? (Answer: 2.25 watts.)

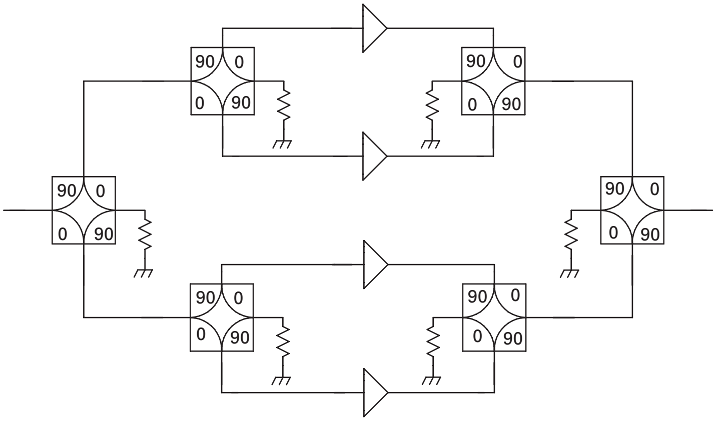

Problem 15.7. Design an op-amp circuit to replace the telephone hybrids in Figure 15.3(b). Put a low-value resistor in series with the line so that a differential

amplifier will produce a voltage proportional to the current × Z_0. Use op-amps to produce signals that are the sum and difference of this voltage and the line voltage.

Problem 15.8. Analyze the coupled-line hybrid to derive Equations (15.1) and (15.2). Let $V_a(z)$ and $V_b(z)$ denote respectively the voltages on the bottom conductor and top conductors. You can write these voltages as the superposition of a common-mode ("even") wave and its reflection plus a differential-mode ("odd") wave and its reflection. Putting $z = 0$ at the right-hand end, confirm that these voltages take the form

$$V_a(x) = \mathrm{V_e e^{-j\mathit{kx}} + V_e \Gamma_e e^{j\mathit{kx}} + V_0 e^{-j\mathit{kx}} + V_0 \Gamma_0 e^{j\mathit{kx}}}$$

$$V_b(x) = V_e \mathrm{e}^{-\mathrm{j}kx} + V_e \Gamma_e \mathrm{e}^{\mathrm{j}kx} - V_0 \mathrm{e}^{-\mathrm{j}kx} - V_0 \Gamma_0 \mathrm{e}^{\mathrm{j}kx}$$

where, as usual, $k = 2\pi/\lambda$. The reflection coefficients, Γ_e and Γ_0, are calculated in terms of the even and odd characteristic impedances, $Z_e = 2\ Z_{CM}$ and $Z_0 = Z_{DIFF}/2$, in the usual way, i.e., $\Gamma_{e,0} = (Z_0 - Z_{e,0})/(Z_0 + Z_{e,0})$ (see Chapter 10). Note that Z_0 is the *load* impedance for the even and odd waves.

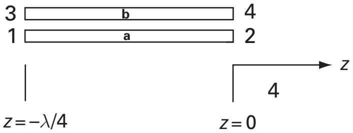

Assume a unity amplitude wave incident at Port 1, i.e., $V_a(-\lambda/4) = 1$. The voltage at the isolated port must be zero, i.e., $V_b(0) = 0$. Use these two equations to find V_e and V_0. Then use V_e and V_0 to show that the coupled voltage, $V_3 = V_b(-\lambda/4)$ is given by $c = (Z_e - Z_o)/(Z_e + Z_o)$. Finally, impose the (match) condition that $I_1 = 1/Z_0$, i.e., $I_a(-\lambda/4)) = 1/Z_0$, to show that $Z_e Z_o = Z_0{}^2$.

Reference

Montgomery, C. G., Dicke, R. H. and Purcell, E. M., *Principles of Microwave Circuits*, Volume 8 of the MIT Radiation Laboratory Series, New York: McGraw Hill, 1948. Reprinted London: Peter Peregrinus, 1987.

CHAPTER

16 Waveguide circuits

In this chapter we examine rectangular metal waveguides and, in particular, their most common mode of operation, the fundamental "TE_{10}" mode. We will also see how the concepts developed for two-conductor transmission lines apply to waveguides and look at waveguide versions of some low-frequency components.

The ability of a hollow metal pipe to transmit electromagnetic waves can be demonstrated by holding it in front of your eye. You can see through it, so, at least, it passes electromagnetic waves of extremely short wavelengths. From a purely dimensional analysis, you would guess correctly that the longest wavelength a pipe could transmit must be of the order of the pipe's transverse dimensions. It turns out that, for propagation in a rectangular pipe, the free-space wavelength, c/f, must be less than twice the longer transverse dimension and, for a circular pipe, less than 1.706[1] times the diameter. Waveguides have less loss and more power handling capacity than coaxial lines of the same size and they need no center conductor nor insulating structures to support a center conductor. Metal waveguides are used most often in the range from 1000 MHz to 100 GHz, where they have practical dimensions. Waveguides for optical frequencies are coated glass fibers.

16.1 Simple picture of waveguide propagation

A common RF engineering argument for the plausibility of transmitting electromagnetic waves through a hollow metal pipe is shown in Figure 16.1, where a two-conductor transmission line evolves into a waveguide. Quarter-wave shorted stubs are added to the line. Since a shorted quarter-wave line presents an open circuit, these stubs do not short the line. More stubs are added to both sides until a rectangular pipe is formed.

This plausibility argument, while not rigorous, does illustrate some important properties of waveguide propagation in the fundamental mode (the

---

[1] The factor 1.706 is $\pi/1.841$, where 1.841 is the smallest root of the equation $\mathrm{d}/\mathrm{d}x\, J_1(x) = 0$ and $J_1(x)$ is the first-order Bessel function of the first kind, a function whose shape resembles $\sin(x)/(x+1)^{1/2}$.

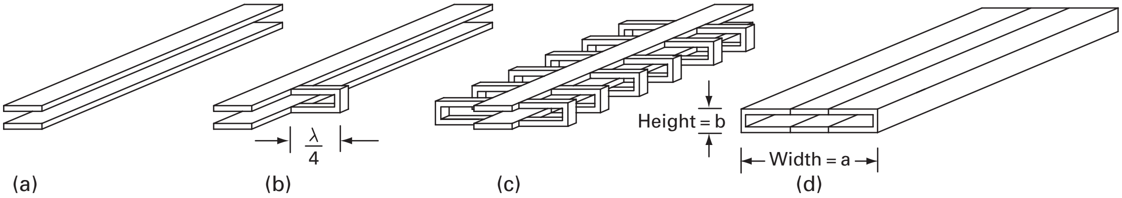

Figure 16.1. Transmission line-to-waveguide evolution. Shorted quarter-wave stubs do not short the transmission line.

simplest and lowest frequency mode): the electric field, which is essentially vertical between the conductors in Figure 16.1(a), becomes perfectly vertical in the waveguide, though its magnitude must fall to zero at the waveguide's sides, since the metallic walls short out any tangential electric field. And, just as the conductors of the transmission line of Figure 16.1(a) can have any separation and still support wave propagation, the waveguide of Figure 16.1(d) can have any height. The width, however, is critical. The total width of the guide must be at least $\lambda/2$ to accommodate a quarter-wave stub on each side and still have nonvanishing conductor strips, as shown in Figure 16.1(c). This means that there is a low-frequency cutoff; wave propagation is not possible if the wavelength is greater than $2a$ where a is the waveguide width (the longer dimension).

Standard waveguide designations indicate the shape and size of the guide. WR430, for example, denotes "Waveguide, Rectangular," 4.3 inches (10.9 cm) wide. The standard width-to-height ratio is two-to-one. (While the height of the guide can be made arbitrarily small, the waveguide will become increasingly lossy because, for a given power, the currents increase.) One of the largest standard waveguide sizes, WR2300, with a width of 23 inches (58.4 cm), has a low-frequency cutoff of 257 MHz. One of the smallest, WR3, with a width of 0.03 inches (0.076 cm), has a low-frequency cutoff of 197 GHz.

For a standard (width = 2 × height) waveguide, the fundamental mode, called the *TE_{10} mode*, is the only possible mode for frequencies above the low-frequency cutoff and below twice the low-frequency cutoff. Other modes exist above this one-octave range. At frequencies where higher modes are possible, these modes can be unintentionally excited at sharp bends, robbing power from the desired mode. This power does not couple properly to circuit elements designed for the fundamental mode and dissipates in the walls. Whenever possible, a microwave system designer therefore tries to use only the fundamental mode.

The essential details of this most important mode are derived and discussed below.

16.2 Exact solution: a plane wave interference pattern matches the waveguide boundary conditions

Exact solutions for the E and B fields within waveguides of arbitrary shape are normally deduced through a head-on assault using Maxwell's equations.

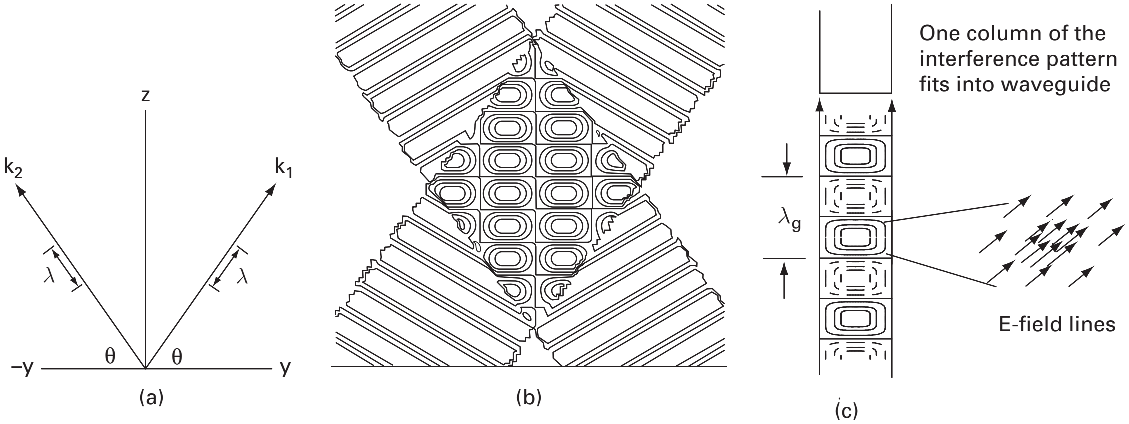

Figure 16.2. Superposition of two plane waves (a) produces an interference pattern (b) streaming in the *z*-direction. One (or more) columns of that pattern satisfies waveguide boundary conditions (c). *E* is in the *x*-direction (coming out of the page).

However, for rectangular waveguides, an exact solution can be obtained indirectly by setting up two plane waves in empty space.

A single plane wave satisfies Maxwell's equations inside a waveguide, but cannot satisfy the boundary conditions at the metal walls. However, the superposition of two properly chosen plane waves forms a traveling interference pattern which does satisfy the boundary conditions and is therefore a valid solution to the waveguide problem. This solution technique can be compared with the "image charge" method in electrostatics, often introduced as a technique to solve for the electric field when a point charge sits alone above an infinite metal sheet. An equal but opposite charge is placed at the mirror image point behind the sheet and the sheet is removed. The electric field lines connecting the charges pass perpendicularly through the x–y plane, satisfying the boundary condition that the E field must be perpendicular when intersecting a conducting surface. The superposition of the two fields, that of the actual charge and that of the image charge, is the solution to the original problem.

Here we construct a solution by superposing two plane waves, identical except for their propagation directions. Both plane waves will be polarized in the x-direction, i.e., their electric fields are in the x-direction. Since electromagnetic waves in free space are transverse waves, the propagation vectors, $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$,[2] corresponding to these waves must both lie in the y–z plane, as shown in Figure 16.2(a).

The first wave, with propagation vector $\boldsymbol{k}_1$, travels in the NNE direction, while the other, $\boldsymbol{k}_2$, travels NNW. In (b), the plane waves are drawn as streams with finite width. Contour lines of the electric field are perpendicular to the directions of propagation. Figure 16.2(b) shows that, in the area where the streams overlap, the sum of the individual electric fields produces an *interference pattern* consisting of columns of cells which stream northward in the z-direction. If we could watch two waves come together in the ocean, we would see them produce this interference pattern. As the streams leave the overlap region, they recover their original plane wave form. In the interference pattern, the contour lines of constant E resemble squared-off ovals. (Remember that $\boldsymbol{E}$ is always perpendicular to the page; the oval-like figures are contours of field strength; they are not field lines.)

---

[2] By definition, the propagation vector, $\boldsymbol{k}$, is in the direction of travel, i.e., perpendicular to the wavefront, and has magnitude $|\boldsymbol{k}| = 2\pi/\lambda$.

The pattern formed by a column of cells (Figure 16.2c) solves the waveguide problem if the width of the cells is equal to a, the width of the waveguide. Why is this a solution? First, the electric field at the side walls is zero at all times, satisfying the boundary condition that, at a conducting wall, there can no parallel electric field. Second, the electric field is always in the x-direction, so it is perpendicular where it intersects the top and bottom walls, satisfying the boundary on those walls as well. Third, each plane wave and therefore their sum, is a solution to Maxwell's equations in empty space, i.e., the interior of the waveguide. What about all the columns of cells outside the boundary of the waveguide? We can forget them, just as we ignore the electric field on the image charge side of the x–y plane in the electrostatic example.

Let us apply a little algebra to find the wave's propagation vector, cutoff frequency and phase velocity. Let k denote the magnitude of $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$ so that $\boldsymbol{k}_1 = k\cos(\theta)\,\hat{\boldsymbol{y}} + k\sin(\theta)\,\hat{\boldsymbol{z}}$ and $\boldsymbol{k}_2 = -k\cos(\theta)\,\hat{\boldsymbol{y}} + k\sin(\theta)\,\hat{\boldsymbol{z}}$. The electric field is the sum of the fields of the two waves, i.e.,

$$E(r,t) = \frac{-E}{2\mathrm{j}}\mathrm{e}^{\mathrm{j}(\omega t - k_1\cdot r)} + \frac{E}{2\mathrm{j}}\mathrm{e}^{\mathrm{j}(\omega t - k_2\cdot r)}, \tag{16.1}$$

where the vector $\boldsymbol{r}$ denotes position in the y–z plane and E is a constant equal to the twice maximum electric field of each wave. The amplitudes, $-E/(2\mathrm{j})$ and $E/(2\mathrm{j})$, have been chosen so that $y=0$ will be a column boundary and also to phase the E field to be maximum at $z=0$ when $t=0$. Substituting the expressions for $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$, we have

$$\begin{aligned} E &= E\Big(-\mathrm{e}^{(\mathrm{j}\omega t - k(z\cos\theta + y\sin\theta))} + \mathrm{e}^{\mathrm{j}(\omega t - k(z\cos\theta - y\sin\theta))}\Big)/(2\mathrm{j}) \\ &= E\sin(ky\sin\theta)\mathrm{e}^{\mathrm{j}(\omega t - kz\cos\theta)}. \end{aligned} \tag{16.2}$$

As always, it is the real part of $\boldsymbol{E}$ that is the actual electric field. Note that all the y dependence is contained in the expression $\sin(k\sin(\theta)\,y)$. This is independent of t, so, in the y-direction, the diffraction pattern is a standing wave. The z and t dependence, however, are contained in the wave factor $\mathrm{e}^{\mathrm{j}(\omega t - kz\cos\theta)}$, so the entire diffraction pattern propagates in the $(+z)$-direction with an effective wavevector $k_{\text{guide}} = k\cos\theta$. For a wave of a given frequency, we can find the value of θ that satisfies the side-wall boundary condition. Suppose that the bottom wall of the waveguide extends from $y=0$ to $y=a$, i.e., the guide width is a. The boundary condition at $y=0$ is satisfied for any value of θ, since $\sin(0)=0$. But the boundary condition at $y=a$ demands that $\sin(k\sin(\theta)\ a)=0$. This will be satisfied if $k\sin(\theta)\ a = n\pi$, where n is an integer. Here we will let $n=1$, so that $\sin(\theta) = \pi/(ka)$ and $\cos(\theta) = (1-[\pi/(ka)]^2)^{1/2}$. Thus the E field in the waveguide is

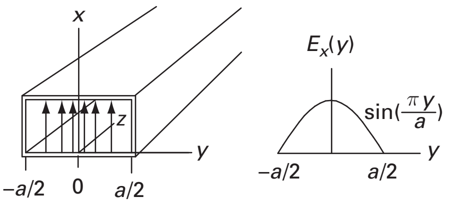

Figure 16.3. Electric field configuration in the TE_{10} mode.

given by (the real part of) $E = E\sin(\pi y/a)e^{j(\omega t-k_g z))}$ where = $k\cos(\theta)=k(1-[\pi/(ka)]^2)^{1/2}$. Figure 16.3 shows how the magnitude of the E field is a maximum at the center line of the waveguide and falls to zero at the side walls.

16.2.1 Guide wavelength

From the expression for k_g, we see that the spatial period along the waveguide will be $2\pi/k_g$. Solving for this length, known as the *guide wavelength*, we find

$$\lambda_{\text{guide}} = \frac{\lambda}{\sqrt{1-\left(\frac{\lambda}{2a}\right)^2}}. \tag{16.3}$$

16.2.2 Magnetic field

Just as the electric field in the guide is the superposition of the electric fields of two plane waves, the magnetic field is the superposition of their magnetic fields. For a plane wave, the magnetic field is perpendicular to both the electric field and the propagation vector,

$$\boldsymbol{B} = \hat{k} \times \mathbf{E}/c, \tag{16.4}$$

where c is the speed of light. The magnetic fields of our two plane waves have z-components as well as y-components, so the magnetic field in the waveguide is not purely transverse with respect to the direction of propagation. In this TE_{10} mode and all other TE modes, only the electric field is purely transverse. There are also TM modes, in which only the magnetic field is purely transverse. Waveguides, unlike coaxial cable, have no TEM modes, in which both E and H fields are transverse.

We can use Equation (16.4) to find the magnetic of field each plane wave and then sum them to get the field in the waveguide. The result is

$$\frac{B_y}{E} = \frac{k_g}{\omega}\sin\left(\frac{\pi y}{a}\right)e^{j(\omega t-k_g z)} \tag{16.5}$$

and

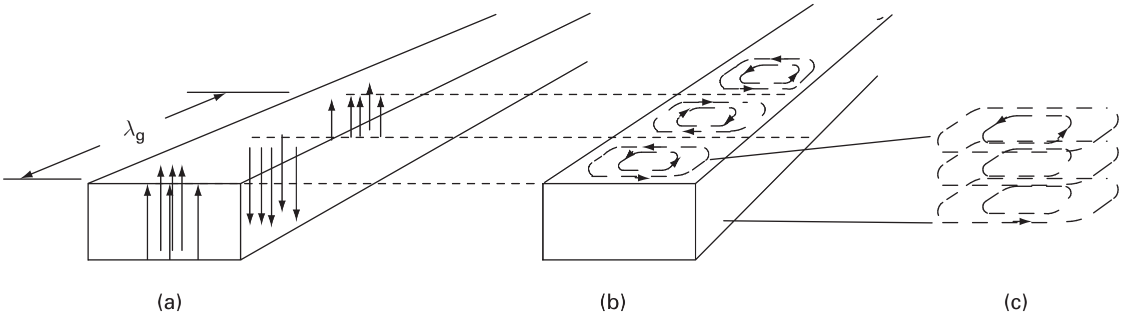

Figure 16.4. Electric field lines (a) and magnetic field Lines (b) The electric lines are bundles of vertical vectors while the magnetic lines are stacks of concentric loops (c.).

$$\frac{B_z}{E} = \frac{\pi}{\mathrm{j}a\omega}\cos\left(\frac{\pi y}{a}\right)\mathrm{e}^{\mathrm{j}(\omega t - k - gz)}. \tag{16.6}$$

The form of this B field is shown in Figure 16.4(b).[3]

Note that the magnetic field lines are stacked concentric loops in the y-z plane with no component normal to the walls. You can use Equations (16.5) and (16.6) to find the exact shape of these loops (see Problem 16.4).

16.2.3 Wall currents

Wall currents, which flow on the inside surfaces of the waveguide, are determined by the tangential magnetic field. The currents are perpendicular to the B field and their magnitude (in amperes/meter) is given by B/μ_0 (the permeability of free space, μ_0, is equal to $4\pi{\cdot}10^{-7}$). The wall currents are indicated in Figure 16.5. These currents converge or diverge from areas on the broad wall

Figure 16.5. Wall currents (solid lines) in relation to the magnetic field (dashed lines).

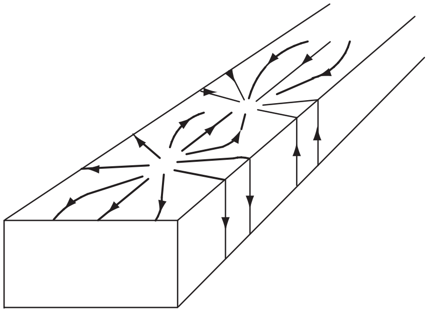

---

[3] The reader familiar with Maxwell's equations can quickly derive Equations (16.5) and (16.6) from Equation (16.2) by applying the curl E equation, which here becomes $\mathrm{j}\omega B_y = -\partial E_x/\partial z$ and $\mathrm{j}\omega B_x = \partial E_x/\partial y$.

where positive charge is arriving or leaving. The E-field lines start and end at these charge patches. Note that the currents on the narrow walls are perfectly vertical because the tangential magnetic field has no x-component.

The fields and currents shown in Figures 16.2–16.5 are, of course, snapshots at an instant in time. As the wave propagates, these patterns move uniformly along the z-axis with a (phase) velocity given by $v_{\text{ph}} = \omega/\beta$.

16.3 Waveguide vs. coax for low-loss power transmission

Consider a situation requiring a low-loss transmission line. Let us compare a standard 2:1 aspect ratio waveguide to a cylindrical coaxial line. To minimize the loss we will make both as large as possible, but here we will impose the restriction that they are also small enough so that modes higher than the fundamental mode cannot propagate. Appendix 16.1 shows that the diameter of this lowest-loss coaxial line and the height of the lowest-loss waveguide are very close to $\lambda/2$ and that the coaxial line will have 2.4 times the loss of the waveguide and will carry only 23% as much power before breakdown.

16.4 Waveguide impedance

There are several ways to define an impedance for a waveguide. One way is to define the voltage to be the potential difference between the top and bottom walls at the middle of the guide and the current to be the integrated current across the top wall. The ratio of voltage to current gives an impedance. Another definition uses voltage and power flow. Still another method uses the ratio of electric field to magnetic field at the center of the guide. The various definitions give $Z_0 = 377$ ohms (impedance of free space) within a factor of 2. But regardless of how impedance is defined, there is no ambiguity in the concept of reflection coefficient. Recall that a shunt capacitance on an ordinary (TEM) transmission line produces a reflection coefficient on the negative j-axis of the Smith chart. The same kind of reflection is produced in a waveguide by a short vertical post or a horizontal iris. These obstructions are therefore called "capacitive posts" or "capacitive irises." An iris across the narrow dimension of the guide causes a reflection on the positive j-axis so is called an "inductive iris." Figure 16.6 shows examples of inductive and capacitive irises. (The equivalent circuit for a thin iris is just a single shunt susceptance.)

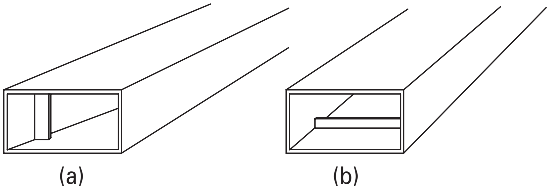
(a) (b)

Figure 16.6. Waveguide irises: (a) inductive iris; (b) capacitive iris.

The combination of an inductive and a capacitive iris (a thin wall with a hole) is equivalent to a parallel resonant circuit. You can see how these resonant irises could be spaced at quarter-wave intervals in a waveguide to make a coupled-resonator filter.

16.5 Matching in waveguide circuits

Impedance matching in waveguide circuits can be done with the same techniques used for ordinary transmission lines. Suppose we are using a waveguide to supply power to some device, maybe a horn antenna, and that we have an instrument – a reflectometer or network analyzer – that can measure the reflection coefficient looking into the waveguide. We can locate the reflection coefficient on the complex reflection plane (Smith chart) as shown in Figure 16.7.

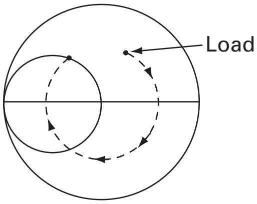

Figure 16.7. A load's reflection coefficient located on the Smith chart.

As we move down the guide, away from the load, the reflection coefficient circles the center of the chart and eventually arrives at the unity conductance circle. We locate this position on the guide and install the appropriate inductive or capacitive iris. In practice the tuning process is sometimes very simple: we find the point at which we need to add shunt capacitance. If the reflected wave is small (not a severe mismatch) we do not have to add much capacitance so we get out the ball-peen hammer and dent the broad side of the guide. An expert learns just how hard to swing the hammer.

A note on matching: suppose we join two dissimilar waveguides (perhaps of different sizes) at a junction, which could be some kind of elbow, coupling, butt joint, etc. Assume that the system is nominally lossless, i.e., all metal. We want to match the junction so that a wave coming from either direction will suffer no reflection. We carry out the above procedure on one side of the junction. Do we have to then match the other side? No, the job has been done. Time-reversal produces an equally good solution to Maxwell's equations in which all the power flows in the opposite direction. Of course this applies just as well to ordinary (TEM) transmission lines as it does to waveguides. This simple argument fails for lossy junctions because the time-reversed solution requires the absorptive material to *produce* power, but a stronger argument, based on the reciprocity theorem leads to the same conclusion.

16.6 Three-port waveguide junctions

Two kinds of waveguide T-junctions (three-port junctions) are shown in Figure 16.8.

The *series-T* gets its name from the fact that the voltage of the input guide divides between the two output guides. This works out well because the half-height output guides have half the impedance of the input guide and the junction is inherently matched. (The half-height guides could be increased to full-width in a gradual taper that would not cause much reflection.) The *shunt-T* applies the full input voltage across each of the output arms – not such a natural as the series-T.

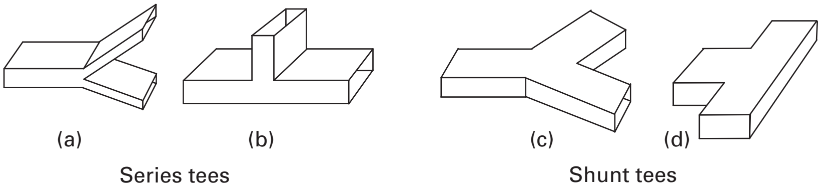

Figure 16.8. Waveguide T-junctions.

16.7 Four-port waveguide junctions

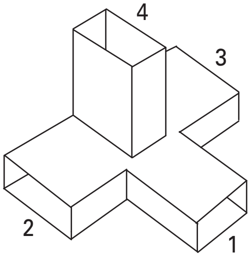

Figure 16.9. Waveguide Magic T.

The *Magic T* hybrid can be built using a procedure that itself seems like magic. We start with the bare waveguide junction (nothing hidden inside) as shown in Figure 16.9.

First we match Port 1, i.e., eliminate reflections from Port 2 when the other ports are feeding matched, i.e., reflectionless, loads. To do this we start by putting matched loads on Ports 2 and 3. (We do not have to put a load on Port 4 since, by symmetry, it is isolated from Port 1.[4]) With the loads in place we measure the reflection at Port 1 and install the necessary iris (or dent) somewhere down line 1. Then we do the same process on Port 4. That's it. The two matches and the isolation by virtue of symmetry are sufficient. We now have a perfectly matched Magic T hybrid.

Simple narrowband transitions from coax to waveguide have mostly been built with empirical methods. With the aid of three-dimensional finite element simulation programs, wideband transitions have been designed. In general, the designer first looks at the fields on both sides and finds a mechanical structure that causes the main features of the fields to line up. The remaining reflection should be small and can be tuned out with a small iris or other structure whose complexity depends on the desired bandwidth.

Rectangular waveguides, like TEM lines, can carry only one signal in each direction. But square or round guides can have two independent waves; they are both fundamental mode waves but they have different polarizations. To launch or recover these two waves independently requires an *orthomode coupler*, which has no TEM counterpart. The simplest orthomode couplers use coaxial or waveguide connections mounted at right angles on the sides of the square or round guide. Some couplers produce circular rather than rectangular polarizations. Wideband orthomode transitions are always needed for radio astronomy and their development is an active field.

---

[4] The *E* field is vertical as a wave enters Port 1. Would it point left or right as it emerged from Port 4? Since the geometry is symmetric, there is no reason to favor right or left. Hence, no wave emerges from Port 4.

Appendix 16.1 Lowest loss waveguide vs. lowest loss coaxial line

For lowest loss we will make the waveguide and the coaxial line as large as possible, but, as explained above, with the restriction that each be capable of supporting only its fundamental mode. Our lowest loss $TE_{1,0}$ waveguide will be made with its width equal to the wavelength. (If it is any wider, the second mode, $TE_{2,0}$, becomes possible.) We will make the height equal to half the width, i.e., the usual aspect ratio. For air-filled coaxial line at the frequency where non-TEM modes become possible, the inner and outer radii, r_i and r_o, satisfy the inequality $(r_o + r_i)\,\pi \geqq 1.03\lambda$.[5] The equal sign applies when $r_i / r_o = 1/3.6$. This ratio also provides the lowest loss air-filled coaxial line for a given outer diameter (see Appendix 16.2). Note that the characteristic impedance, $Z_0 = 60 \ln (r_o / r_i)$, will be 77 ohms for this lowest-loss coaxial cable. Using this ratio of diameters, the maximum outer diameter is given by $r_o = 1.03\lambda\pi^{-1}/(1+1/3.6) = 0.26\lambda$. These relative waveguide and coax cross-sections are shown in Figure 16.10.

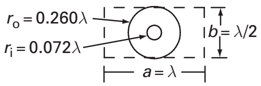

Figure 16.10. Relative cross-sections of lowest-loss waveguide and coaxial cable.

Let us compare the losses. The amplitude of a wave propagating in the $(+x)$-direction on any lossy line is proportional to $\exp(-\alpha x)$ where α, the loss factor, has units of inverse meters. The power is therefore proportional to $\exp(-2\alpha x)$ and the fractional power loss per meter is 2α. Note that $20 \log(e)\alpha = 8.686\alpha$ dB/meter. Because of the skin effect, the loss of a line is proportional to its surface resistance which is given by

$$R_S = \sqrt{\frac{\omega\mu}{2\sigma}} \quad \text{ohms per square} \tag{16.7}$$

where σ is the bulk dc conductivity and ω is the (angular) frequency. (For copper, $R_s = 2.61 \times 10^{-7}$ ohms/$\sqrt{\text{Hz}}$; for aluminum, $R_s = 3.26 \times 10^{-7}$ ohms/$\sqrt{\text{Hz}}$). The loss factor for air-filled coax line is given by

$$\alpha_{\text{coax}} = \frac{R_s}{2Z_0}\left(\frac{1}{2\pi r_o} + \frac{1}{2\pi r_i}\right). \tag{16.8}$$

Our lowest loss coax line has $Z_o = 77$, $r_o = 0.26\lambda$ and $r_i = 0.072\lambda$ so its loss becomes

$$(\alpha_{\text{coax}})_{\min} = 0.0183\frac{R_s}{\lambda}. \tag{16.9}$$

For air-filled rectangular waveguide in the fundamental mode the loss factor is given by

$$\alpha_{WG} = \frac{R_s}{377}\frac{\left(1 + \dfrac{2b}{a}\left(\dfrac{\lambda}{2a}\right)^2\right)}{b\sqrt{1 - \left(\dfrac{\lambda}{2a}\right)^2}} \tag{16.10}$$

---

[5] Reference [2], p. 42.

where 377 ohms is $\sqrt{\mu_0/\epsilon_0}$, the "impedance of free space." In our case $a = 2b = \lambda$ so

$$(\alpha_{\mathrm{WG}})_{\min} = \frac{R_{\mathrm{s}}}{\lambda}\frac{5}{337\sqrt{3}} = 0.0076\frac{R_{\mathrm{s}}}{\lambda}. \tag{16.11}$$

The loss of the coax is therefore higher than that of the waveguide by a factor of 0.0183/0.0076 or about 2.4. What about power handling capacity? The breakdown of either the waveguide or the coax depends on the maximum E field, $E_{\max}$. (For air at sea-level pressure, $E_{\max}$ is about 30 000 volts/cm.) For rectangular waveguide in the fundamental mode the power is related to the maximum E field by

$$\frac{\mathrm{Pwr}}{{E_{\max}}^2} = \frac{1}{4\cdot 377}ab\frac{\lambda_0}{\lambda_g} = 6.63 \times 10^{-4}ab\frac{\lambda_0}{\lambda_g} \tag{16.12}$$

where Pwr is in watts, $E_{\max}$ is in volts/cm, a and b are in cm, λ_0 is the free-space wavelength, and λ_g, the guide wavelength, is given by

$$\lambda_g = \frac{\lambda_0}{\sqrt{\left(1-\left(\frac{\lambda_0}{\lambda_{\mathrm{cutoff}}}\right)^2\right)}}. \tag{16.13}$$

For our waveguide $\lambda_0/\lambda_{\mathrm{cutoff}} = 1/2$ so $\lambda_g = 2\lambda_0/\sqrt{3}$ and

$$\frac{\mathrm{Pwr}}{{E_{\max}}^2} = 5.74 \times 10^{-4}ab = 5.74 \times 10^{-4}\lambda\,\lambda/2 = 2.37 \times 10^{-4}\lambda^2. \tag{16.14}$$

Turning to the coax, the ln(r) dependence of voltage and the characteristic impedance, $Z_0 = (377/2\pi)\ln(r_{\mathrm{o}}/r_{\mathrm{i}}) = 60\ln(r_{\mathrm{o}}/r_{\mathrm{i}})$ allow us to find the power in terms of the maximum E field:

$$\frac{\mathrm{Pwr}}{{E_{\max}}^2} = \frac{Z_0 r_{\mathrm{i}}^2}{2\cdot 60^2}. \tag{16.15}$$

In our case $Z_0 = 77$ and $r_{\mathrm{i}} = 0.072\lambda$ so

$$\frac{\mathrm{Pwr}}{{E_{\max}}^2} = \frac{77}{2\cdot 60^2}(0.072\lambda)^2 = 0.55 \times 10^{-4}\,\lambda^2. \tag{16.16}$$

We see that the waveguide can handle 2.37/0.55 = 4.3 times the power of the minimum loss coaxial line. The waveguide is clearly better both for loss and power handling capacity. In high-power applications the waveguide has the additional advantage that there are no interior surfaces needing cooling and no mechanical spacers to center an inner conductor. (Insulating spacers in high-power coaxial lines must fit tightly; high voltage develops across any gap. This problem generally reduces the power-handling capacity of the coaxial line by something like an order of magnitude.)

Appendix 16.2 Coax dimensions for lowest loss, highest power, and highest voltage

Lowest loss

For a given outer diameter, the characteristic impedance of a coaxial line is increased by making the inner diameter smaller. For a given power, the current is decreased. But the smaller inner conductor has more resistance. The I^2R product, i.e., the dissipation, has a minimum when the ratio of diameters is 3.6. This follows from Equation (16.8) which can be rewritten as

$$\alpha_{\text{coax}} = \frac{R_{\text{s}}}{2 \cdot 60 \cdot 2\pi r_{\text{o}}} \frac{1}{\ln(x)} (1 + x) \tag{16.17}$$

where $x = r_o/r_i$. The minimum of $(1+x)/\ln(x)$ occurs at $x = 3.6$ so the characteristic impedance of lowest-loss air-filled coaxial line is $Z_0 = 60 \ln(3.6) = 77$ ohms.

Highest power

From Equation (16.15) we see that to maximize the power-handling capability of the coaxial line we must maximize the expression $Z_0 \, r_i^2$, i.e., we must maximize $\ln(x)/x^2$. The maximum occurs when $\ln(x) = 1/2$ so the characteristic impedance of the maximum power line is $Z_0 = 60/2 = 30$ ohms.

Maximum voltage

If the line is to withstand maximum voltage the optimum value of $\ln(x)$ is 1 and the characteristic impedance is 60 ohms. This also follows from Equation (16.15): if we express power as $V_{\max}{}^2/(2Z_0)$ then $V_{\max}$ is proportional to $Z_0 \, r_i$ or $\ln(x)/x$, which reaches a maximum at $x = e$.

Relative performance of 50-ohm coaxial line

The 50-ohm line commonly used in RF work ($x = r_o/r_i = 2.3$) strikes a compromise between lowest loss, highest power and highest voltage. For loss, we compare $(1+x)/\ln(x)$ for $x = 2.3$ and $x = 3.6$ to see that the 50-ohm line will have only 10% more loss than a 77-ohm line with the same outer diameter. For power handling, we compare $\ln(x)/x^2$ and find that the 50-ohm line can carry 62% as much power as a 30-ohm line with the same outer diameter. Finally, for voltage we compare $\ln(x)/x$ and find that the 50-ohm line can handle 98% as much voltage as a 60-ohm cable with the same outer diameter.

Problems

Problem 16.1. Suppose a car enters a long tunnel which is essentially a rectangular metal tube 10 meters wide by 5 meters high. The car radio becomes silent inside the tunnel. Was the radio more likely tuned to an AM station or an FM station?

Problem 16.2. Examine the waveguide current distribution shown in Figure 16.5 (for the fundamental mode) and draw a sketch showing the position(s) in which a narrow slot could be cut through the waveguide wall without affecting its operation.

Problem 16.3. Describe an experimental setup that could be used to demonstrate the waveguide E-field and B-field distributions shown in Figure 16.4.

Problem 16.4. In the discussion just above Equation (16.2), let $n = 2$ instead of 1. For this choice of n (the TE_{20} mode), find the cutoff wavelength and the guide wavelength. Sketch the electric and magnetic field lines.

Problem 16.5. Use Equations (16.5) and (16.6) to find the mathematical shape of the magnetic field loops. Hint: the slope of a field line, $\mathrm{d}z/\mathrm{d}y$, is given by B_z/B_y. Set up the equation $\mathrm{d}z/B_z = \mathrm{d}y/B_y$ and note that the left side contains only z while the right side contains only y. They can therefore be integrated separately. Remember to add a constant of integration.

References

[1] Collin, R. E., *Foundations for Microwave Engineering*, New York: McGraw Hill, 1992.

[2] Montgomery, C. G., Dicke, R. H. and Purcell, E. M., *Principles of Microwave Circuits*, London: Peter Peregrinus, 1987 (originally Volume 8 of the MIT Radiation Laboratory Series, New York: McGraw Hill, 1948).

[3] Moreno, T., *Microwave Transmission Design Data*, Sperry Gyroscope Corp, 1948, reprinted by Dover Publications, 1958.

[4] Ramo, S., Whinnery, S. R. and Van Duzer, T. *Fields and Waves in Communication Electronics*, 3rd edn., New York: John Wiley, 1994. (Original edition was Ramo, S. and Whinnery, S. R., *Fields and Waves in Modern Radio*, New York: John Wiley, 1944).

CHAPTER

17

Small-signal RF amplifiers

In this chapter we discuss the amplifiers used commonly in the front-end and IF stages of receivers and in antenna-mounted preamplifiers. The maximum output power of these amplifiers is typically from 0.01 W to 0.1 W (10–20 dBm). The power amplifiers discussed in Chapter 3 use the full range of the transistor conductance to “push” or “pull” the output voltage to any value from zero to the dc supply voltage(s). Small-signal amplifiers, on the other hand, are class-A amplifiers in which the signal voltages are small, compared with the dc bias voltages. The small ac signals add to dc bias voltages, so the output signal, δV_{out}, produced by an input signal δV_{in} is given by $\delta V_{out} = [dV_{out}/dV_{in}]\ \delta V_{in} + 1/2\ [d^2 V_{out}/dV_{in}^2]\ (\delta V_{in})^2 + \cdots$ The ac voltage gain is, therefore, $dV_{ou}t/dV_{in}$, evaluated at the quiescent bias point. When operated over only a small range of δV_{in}, the higher derivatives of V_{out} versus V_{in} make only small contributions, and the amplifier is essentially linear. Key characteristics of these amplifiers are gain, bandwidth, input and output impedances, linearity (those higher derivatives), and internally generated noise.

17.1 Linear two-port networks

Small-signal amplifiers are linear amplifiers; the output signal should be a faithful reproduction of the input signal.[1] A general definition of small-signal amplifiers could be that they are amplifiers built entirely of nominally linear elements (which include resistors, capacitors, inductors, transmission lines, and transistors operated over a small differential range), from which it follows that the overall circuit will also be nominally linear. An amplifier, being an example of a two-port network (or simply a “two-port”), has an input terminal, an output terminal, and a common terminal (ground). The operation of any linear two-port can be described by four variables: the input and output voltages and currents. Any two of these variables can be considered independent variables ("input" or "cause"). The remaining two variables are then dependent variables ("output" or "effect"). If, for example, V_1 and V_2 are chosen as the dependent variables, the two-port is described by the equations:

---

[1] While small-signal amplifiers are linear almost by definition, an important exception is the limiting amplifier or *limiter*. In these amplifiers, the gain decreases for increasing signal levels. A cascade of limiters can have an output level almost independent of input level. A limiter is used ahead of an FM detector if the particular FM detector is sensitive to amplitude variations as well as frequency variations.

$$V_1 = Z_{11}I_1 + Z_{12}I_2 \tag{17.1}$$

$$V_2 = Z_{21}I_1 + Z_{22}I_2. \tag{17.2}$$

For this choice of dependent variables, the four coefficients are known as the Z *parameters*. We are implicitly dealing with ac circuit analysis, so these four parameters generally are complex and are functions of frequency. The important point to note is that, for a given frequency, *any* two-port network (amplifier, filter, transmission line, etc.) can be completely described by just four complex numbers. By convention, the current at either terminal is positive when it flows *into* the terminal. Note that the output variables are linear functions of the input variables, since the input variables appear raised only to the first power.[2] By inspection of Equation (17.1) we see that Z_{11} is the network's input impedance when the output current is zero, i.e., when the output is open circuited. The *forward transfer impedance*, Z_{21}, is the open-circuit output voltage divided by the input current – a "transimpedance." If we are given the load impedance, we can use Equations (17.1) and (17.2) to calculate the power gain of an amplifier (Problem 17.1). The *reverse transfer impedance*, Z_{12}, is a measure of reverse feedthrough. If the RF amplifier preceding an unbalanced mixer in a superheterodyne receiver has reverse feedthrough, some power from the local oscillator will get to the antenna and be radiated. But what is more important, reverse feedthrough can cause an amplifier to oscillate for certain combinations of input and output terminations. This two-port formalism provides more than just a top-level description of an amplifier. It is the basis for amplifier circuit analysis and design, since the active devices (transistors) inside the amplifier can themselves be represented as two-port networks, whose parameters are furnished by the manufacturer on data sheets. Another equivalent set of parameters is the Y-parameters, defined by

$$I_1 = Y_{11}V_1 + Y_{12}V_2 \tag{17.3}$$

$$I_2 = Y_{21}V_1 + Y_{22}V_2. \tag{17.4}$$

Conversion formulas between parameter sets are easily derived. For example,

---

[2] The dependence of the output variables on only the first power of the input variables follows from a general definition of linearity: If an input *a* causes an output *A* and an input *b* causes an output *B*, then an input C_1a+C_2b, where C_1+C_2 are constants, will result in an output C_1A+C_2B.

$$Y_{11} = \left.\frac{I_1}{V_1}\right|_{V_2=0.} = \left[Z_{11} - \frac{Z_{21}Z_{12}}{Z_{22}} \right]^{-1}. \tag{17.5}$$

The widely used *S-parameters*, which are the subject of Chapter 28, form another equivalent four-parameter set, for which the variables are linear combinations of the voltages and currents, and correspond to input and output waves at each port. Two of the parameters, S_{11} and S_{22}, are reflection coefficients, while the other two are transmission coefficients. A characteristic impedance, Z_0, usually 50 ohms, is implicit. Again, conversion formulas are readily derived, for example,

$$S_{11} = \frac{Y_{22} + Z_0^{-1} - Z_0(Y_{11}Y_{22} + Y_{11}Z_0^{-1} - Y_{12}Y_{21})}{Y_{22} + Z_0^{-1} + Z_0(Y_{11}Y_{22} + Y_{11}Z_0^{-1} - Y_{12}Y_{21}}. \tag{17.6}$$

In this introductory chapter, we discuss amplifiers in terms of voltages and currents, in the interest of presenting the basic concepts in terms totally familiar to the reader.

17.2 Amplifier specifications – gain, bandwidth, and impedances

The small-signal gain (forward and reverse), bandwidth, input impedance and output impedances could be called “linear specifications” because they can all be calculated from the amplifier’s Z-parameters or from any of the other equivalent sets of parameters. The gain and bandwidth of an amplifier are ultimately limited by the characteristics of the transistor(s). Transistors have unavoidable built-in reactances: there are at least two capacitors in even the simplest transistor circuit models (approximate equivalent circuits). Elaborate models for microwave transistors can contain a dozen capacitors and inductors. Amplifiers designed for narrowband use (fractional bandwidths of 20% or less) use input and output matching networks to absorb or “cancel” these reactances. At higher frequencies, the shunt capacitive reactances become lower. The matching networks must then necessarily have higher loaded Qs which means that bandwidth decreases. This limitation is fundamental; no matter how complicated the matching network, gain must be traded for bandwidth. Negative feedback around a transistor will lessen the effect of its reactances. But feedback decreases the gain so again there is a tradeoff between gain and bandwidth. In some applications the input and output impedances of an amplifier are critical. For example, if a narrowband filter is placed between two amplifiers, the amplifiers must present the proper impedances to the filter if the intended passband shape is to be realized. The frequency dependences of the input and output impedances of an amplifier are, of course, related to the bandwidth, since the frequency response is normally determined by mismatch (i.e., reflection).

17.2.1 Amplifier stability

An amplifier is required to be stable (not oscillate) in its working environment. A 100-MHz amplifier, for example, will not be satisfactory if it oscillates, even at a very different frequency, say 1 GHz. Oscillation invariably takes the circuit into large-amplitude excursions and the combination of amplification and oscillation is highly nonlinear. An amplifier that remains stable when presented with *any* combination of (passive) source and load impedances (but no external feedback paths) is said to be *unconditionally stable*. Unconditional stability is not always necessary. An IF amplifier in a receiver needs only to be stable in its never-changing working environment. The input RF amplifier in a short-wave radio, however, might be connected to any random arbitrary antenna so it should be unconditionally stable, at least with respect to input impedance. General-purpose commercial modular amplifiers are usually designed to be unconditionally stable. Using these, a system designer can realize a needed transfer function by cascading amplifiers, filters, etc., and know that the combination will be stable. Stability, like gain and input and output impedances, is predictable from the two-port parameters. To find whether an amplifier will be unconditionally stable, it is necessary and sufficient to show that the real parts of the input and output impedances for any frequency are positive for any passive load and source impedances. Suppose that an analysis shows that for some combinations of load impedance and frequency, the real part of the input impedance is negative, but never more negative than −5 ohms. Then adding a series resistor of more than 5 ohms to the input of the amplifier would make it unconditionally stable. Such resistive remedies, however, always decrease gain and increase internally generated noise. The reverse transfer parameter (Z_{12}, Y_{12}, S_{12}, …) plays a key role in stability. For example, a sufficient (but not necessary) condition for unconditional stability is that its reverse transfer parameter be equal to zero. It is important to note, however, "unconditional stability" simply means that the two-port cannot be provoked into oscillation by varying its termination impedances. A multistage amplifier circuit could contain an oscillating internal stage and still have input and output impedances with positive real parts for all frequencies and arbitrary terminations. (Of course a multistage amplifier will be unconditionally stable if every stage is unconditionally stable.)

17.2.2 Overload characteristics

Any amplifier will become nonlinear at high enough signal levels, if only because the output runs up against the "rail" of the dc power supply. But before this occurs, transistor nonlinearity comes into play. A straightforward specification of an amplifier's upper power limit is the *1-dB compression point*. This is the value of the output power at which the gain has dropped by 1 dB, i.e., the point at which the output power is 79.4% of what would be predicted on the basis of low-power gain measurements.

Intermodulation

Small departures from linearity, even when the amplifier is far below compression, become a concern in a receiver when the passband of an RF amplifier contains two or more signals at frequencies f_a, f_b, f_c, … that are much stronger than the desired signal (the signal that will be isolated downstream by a narrow bandpass filter). Nonlinearity can produce mixing products at frequencies of $n_a f_a + n_b f_b + n_c f_c + \cdots$ where $n_i = 0, \pm1, \pm2, \ldots$ In receivers, the most troublesome of these products are the third-order products $2f_a - f_b$ and $2f_b - f_a$. (Third-order products will be inevitably produced if the output voltage of an amplifier contains even a small term proportional to the cube of the input voltage.) The special problem with these particular products is that they can fall within the IF passband. To see this, suppose f_2 and f_3 are the frequencies of signals close enough to a desired frequency, f_1, that they will pass through the broadband front-end of a receiver. The local oscillator is tuned to convert f_1 to the center of the IF passband. But the third-order products $2f_3 - f_2$ and/or $2f_2 - f_3$, being very close to f_1, can also fall within the narrow IF passband and interfere with the desired signal. Second-order intermodulation is not so troublesome since the products have frequencies far outside the IF passband.

A standard measurement of intermodulation is the *two-tone test*, which uses two closely spaced signals of equal amplitude, A. On a log-log plot of output power versus input power each of these fundamental signals will fall on a 45° line, with slope = 1. The third-order products, however, will fall on a line with slope = 3 because the power in the third-order products is proportional to the cube of the power in each of the input signals. The *third-order intercept* is the point at which the third-order product would have as much power as each of the fundamental signals. Usually the number given for the intercept point is the *output* signal strength. The second-order intercept is defined the same way. Figure 17.1 shows a third-order intercept point of about +37 dBm. Generally an

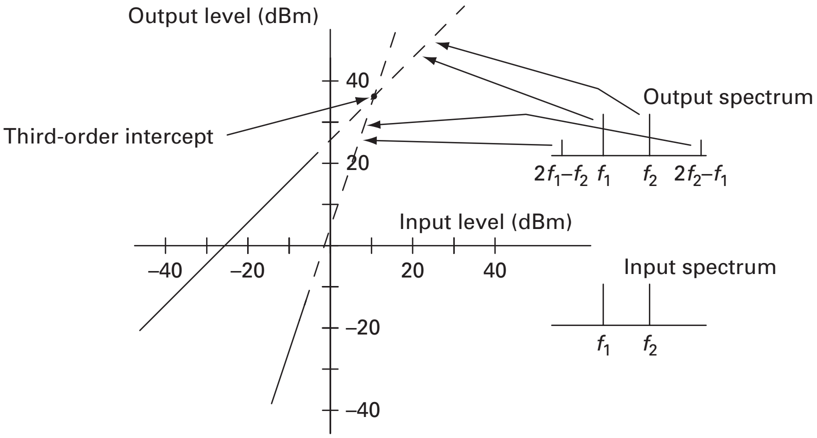

Figure 17.1. Two-tone test to specify amplifier linearity.

amplifier cannot be driven all the way to the intercept points; they are extrapolations from measurements made at much lower input levels. (The output strengths of the fundamental and the second- or third-order product need only be measured at one input level. Lines with slopes of one, two, or three are then drawn through them to locate the intercepts.)

Dynamic range

Every amplifier adds some noise to the signal. (Later we will discuss amplifier noise in some detail.) Very weak signals will be buried in this noise and lost. The *dynamic range* of an amplifier is therefore determined at the low end by the added noise and at the high end by nonlinearity. In order to handle strong signals, a receiver should keep mixing products small by having as little amplification as possible prior to the narrowest bandpass filter. We will see, on the other hand, that if a receiver begins with a mixer or with a narrowband filter, the loss in these elements adds noise and will render the receiver less sensitive than if the first element after the antenna had been a low-noise amplifier. A trade-off must often be made between sensitivity and dynamic range. Power dissipation is obviously important for battery-operated equipment where milliwatts may count. But to achieve high dynamic range, a small-signal amplifier may have a fairly high-power quiescent point and have to dissipate as much as several watts of power.

17.3 Narrowband amplifier circuits

Amplifiers for frequencies below about 30 MHz look very much like resistance-coupled audio amplifiers. The load resistors are replaced by shunt inductors which cancel the transistor capacitances, which would otherwise tend to be short circuits at RF. These resonant circuits make a narrowband amplifier. Often an even narrower bandpass is desired; the inductors are given smaller values and are shunted with external capacitors (effectively increasing the transistor capacitances). Focusing on one stage of an amplifier (or an amplifier of one stage), the fundamental design decisions are transistor selection and circuit configuration, i.e., common-emitter, common-base, or common-collector. (Here, and usually elsewhere, *emitter, base*, and *collector* can also mean *source, gate, and drain*.) The choice of a transistor will be based on the ability to provide gain at the desired frequency, noise, and perhaps linearity. The orientation of the transistor might be common-emitter when maximum gain is required, common-base when the device is being pushed near its upper frequency limit or when the isolation between input and output is critical, or common-collector when very low output impedance is needed. As far as noise goes, it turns out that the three orientations are equivalent when used in a high-gain multistage amplifier.

17.4 Wideband amplifier circuits

Most wideband amplifiers use feedback. An unbypassed emitter impedance provides series feedback. An impedance between collector and base provides shunt feedback. Commercial modular general-purpose amplifiers use resistive series and shunt feedback. These amplifiers are quite flat up to one or two GHz and have input and output impedances close to 50 ohms over the whole range. Resistive feedback is simple but degrades the noise performance of an amplifier. Wideband low-noise amplifiers often use feedback networks made only of lossless elements, i.e., reactors. The Miller[3] effect multiplies the effective input capacitance in a common-emitter amplifier. This capacitance can be neutralized, at least in a narrowband amplifier. Wideband amplifiers often use the *cascode* circuit in which a common-emitter input stage drives a (low impedance) common base stage. Another good high-frequency circuit, the differential pair, uses an emitter follower stage (high input impedance and low output impedance) to drive a common-base stage.

17.5 Transistor equivalent circuits

An amplifier designer needs a precise electrical description of the transistor(s). For analysis, it is sufficient to have tables of the small-signal parameters of the transistor(s). These tables are usually given in data sheets from the manufacturer; they can be produced using a vector network analyzer. A table of numbers, however, is an awkward representation for design (synthesis) and a common tactic is to represent the transistor by an (approximately) equivalent model circuit of resistors, capacitors, inductors, voltage-controlled voltage generators, voltage-controlled current generators, etc. An exact equivalent circuit for a single frequency can be constructed directly from the small-signal parameters corresponding to that frequency (Problem 17.4). This might be an adequate model for the design of a narrowband amplifier but remember that even an amplifier intended for only a narrow frequency range must be stable at *all* frequencies. For this reason, and also to aid in the design of wideband amplifiers, models are constructed to represent the transistor over a wide frequency range. Normally the topology of an equivalent circuit is based on the construction and physics of the transistor. The element values are determined by least-square fitting programs that make the small-signal parameters of the model agree as closely as possible with the measured small-signal

---

[3] The collector signal in a common-emitter amplifier has a larger magnitude (due to amplification) than the base signal. It also has the opposite sign. The voltage across the transistor's inherent base-to-collector capacitance is therefore larger than the base voltage. As a result, more current flows in this capacitance than if the collector were grounded. The value of the capacitance is, in effect (Miller effect), multiplied.

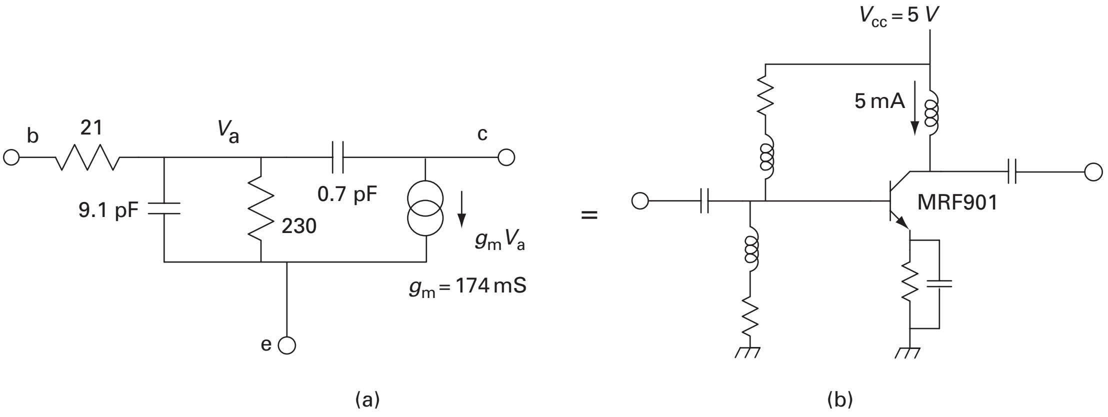

(a) (b)

Figure 17.2. Simple equivalent circuit model for the MRF901 bipolar transistor (100 MHz-2 GHz).

parameters of the actual transistor over the desired frequency range. Agreement can always be improved by adding more elements to the model, but an overly-complicated model will block the intuition of the designer. Equivalent circuits have from one to perhaps twenty parameters. Figure 17.2(b) shows a simple "hybrid-π" equivalent circuit model for a common high-frequency transistor. The component values in the model circuit were determined by least-square fitting to the data sheet values over the range from 100 to 2000 MHz. Since the equivalent circuit models a *biased* transistor, it is actually equivalent to the circuit like that of Figure 17.2(a), which includes a power supply and dc biasing components. The biasing components include three resistors to set the dc collector current. A bypass capacitor grounds the emitter at RF frequencies. Blocking capacitors keep the dc bias voltages from interacting with circuitry outside the transistor and vice versa. An RF choke (inductor) provides enough reactance that practically no signal current can flow from the collector to V_{cc} (RF ground). Two more two chokes prevent RF currents from flowing in the base bias resistors. (Usually these resistors have high enough values that these chokes are unnecessary.)

17.6 Amplifier design examples

Designing a small-signal amplifier for microwave frequencies can be difficult, especially when the design must meet specifications for frequency response, gain, stability, input and output impedances, and noise. Computer-aided design is often used. But, to provide an example, let us design a simple amplifier using the transistor model of Figure 17.2. The amplifier is to be driven from a 50-ohm source and is to drive a 50-ohm load. The only other specifications imposed are that the power gain be at least 10 dB at 430 MHz, and that the amplifier not oscillate while connected to the specified source and load. Let us try a very simple common-emitter circuit consisting of the transistor with a matching inductor in series with the collector (rather than going with a two-element

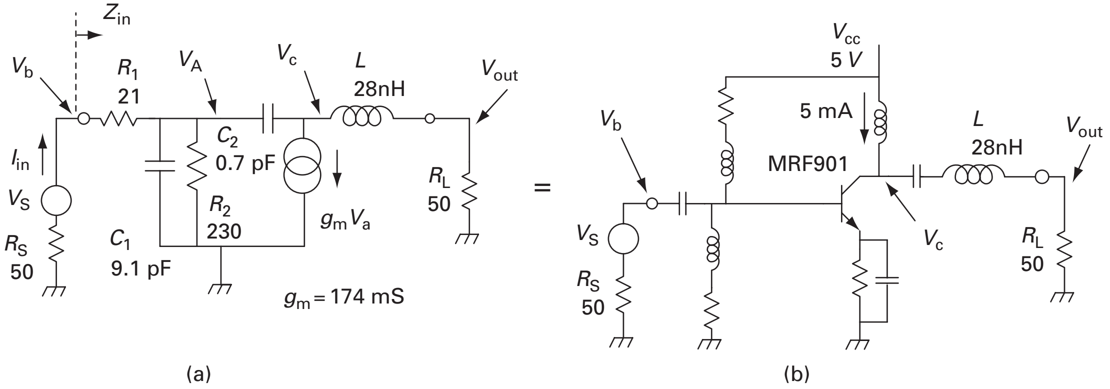

Figure 17.3. Common-emitter amplifier design example: (a) small signal equivalent circuit; (b) full circuit.

matching circuit on each side). The small-signal equivalent circuit of the amplifier is shown in Figure 17.3(a), while the complete amplifier, including bias circuitry, is shown in Figure 17.3(b).

The circuit of Figure 17.3(a) can be analyzed to find V_c in terms of V_b. The output voltage, V_{out}, is then simply $V_c R_L/(R_L + j\omega L)$. Note first that the currents to ground must sum to zero or

$$\frac{V_A - V_b}{R_1} + \frac{V_A}{R_2} + V_A j\omega C_1 + V_A g_m + \frac{V_c}{j\omega L + R_L} = 0. \tag{17.7}$$

Next, consider the current flowing to the left through C_2. This gives us the equation

$$(V_c - V_A)j\omega C_2 = \frac{V_A - V_b}{R_1} + \frac{V_A}{R_2} + V_A j\omega C_1. \tag{17.8}$$

Solving these two equations for V_c as a function of V_b, we find

$$V_c = \frac{-V_b(g_m - j\omega C_2)/R_1}{j\omega C_2(1/R_1 + 1/R_2 + j\omega C_1 + g_m) + (1/R_1 + 1/R_2 + j\omega(C_1 + C_2))/(R_L + j\omega L)}. \tag{17.9}$$

From inspection of Figure 17.3, we see that $V_{out} = V_c R_L/(R_L + j\omega L)$ and that the input current, I_b, is given by $I_b = (V_b - V_A)/R_1$. Since we have already found V_c as a function of V_b, Equation (17.7) (or 17.8) gives us V_A, which lets us calculate I_b and also the input impedance, $Z_{in} = V_b/I_b$. The input power (drive power delivered to the amplifier) is given by $|I_b|^2$ Re(Z_{in}). The power gain of the amplifier is then calculated from

$$\text{Power Gain} = \frac{\text{power out}}{\text{power in}} = \frac{|V_{out}|^2/R_L}{|I_b|^2 \mathrm{Re}(Z_{in})}. \tag{17.10}$$

These expressions are easy to evaluate using a program such as Mathcad or MATLAB. Using Equation (17.9), we find that the power gain at 430 MHz reaches a maximum of 17.25 dB for $L = 28$ nH. This is within 0.25 dB of the

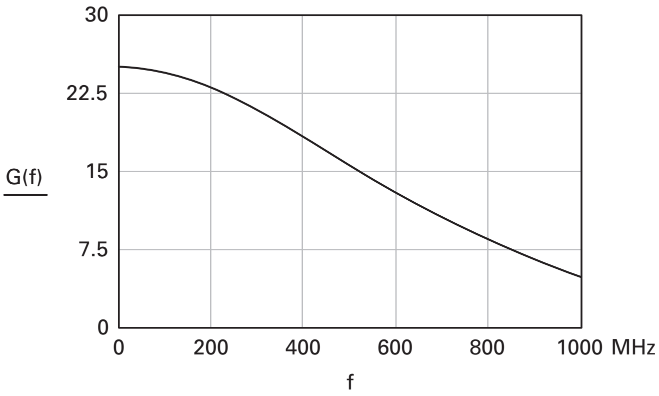

Figure 17.4. Power gain vs. frequency for common-emitter amplifier design example.

maximum power gain available from this transistor (which could be obtained by using a two-element output matching network rather than just a series inductor). Figure 17.4 shows the power gain versus frequency. The input impedance is 13.3–17.0j, from which you can calculate that the output power can be increased by 2.18 dB by the addition of an input matching network to transform the 50-ohm source impedance to 13.3+17.0j.

Analysis of even this simple amplifier requires a fair amount of algebraic effort. Problem 17.5 shows you how to add the hybrid-π transistor model to the circuit analysis program of Problem 1.3. This addition will let you plot the frequency response and find input and output impedances of this amplifier or any arbitrary cascade of transistors and other two-port devices.

We can check the stability of this amplifier when connected to the specified 50-ohm source and load impedances by verifying that any arbitrary set of initial conditions (two capacitor voltages and one inductor current, at $t=0$) results in transient currents that decay exponentially rather than grow exponentially. Because the circuit is linear, we know that there will be transient solutions in which the time dependence goes as $e^{j\omega t}$, where ω is a *characteristic frequency.* If ω has a positive imaginary part, the solution decays exponentially. To find the characteristic frequencies, we set the source voltage, V_s, equal to zero, and assume that V_A and V_B in Equations (17.7) and (17.8) are proportional to $e^{j\omega t}$. Solving these equations results in a third-order polynomial in ω which must be equal to zero. The three roots of the polynomial turn out to be $\omega_1 = j\,2.64E9$, $\omega_2 = j2.16E9$, and $\omega_3 = j18.1E9$. This shows that all three of these transient solutions will decay exponentially, since their imaginary parts of the three frequencies are all positive.[4] For example, $e^{j\omega_1 t} = e^{-2.64E9\ t}$, which is an exponential decay with a time constant of 0.38 nsec. A superposition of these three particular solutions that satisfies an arbitrary initial set of initial conditions will, therefore, produce a decaying transient and this circuit will be stable with the 50-ohm source and load impedances.

---

[4] Here, the three roots are purely imaginary. In general, they also have real parts, corresponding to decaying *oscillations*. For example, if the load resistance is changed from 50 to 500 ohms, the three roots become j2.44e8 and ±6.89e9 + j1.94e10.

Let us check stability if we add an input matching network consisting of a series inductor with 17 ohms of reactance to cancel the −j17.0 input reactance and a transformer to step the 50-ohm source down to 13.3-ohm input resistance. With the addition of a fourth reactive element, the circuit will now have four characteristic frequencies. Straightforward algebraic manipulations produce a fourth-order polynomial in ω whose four roots are ±4.18E9 +j7.83E8, j2.28E10, and j2.48E10. Since the imaginary parts are positive, the amplifier remains stable with this input matching network and the amplifier's *available gain* (power available from the amplifier divided by power available from the source) becomes equal to the power gain. Chapter 28 discusses how stability is usually evaluated in the context of *S* parameters.

Figure 17.5(a) shows an amplifier using the same transistor model arranged in a common-base configuration; the base is grounded and the input signal is applied to the emitter. The equivalent small-signal circuit is shown in (b).

You can analyze this circuit with the methods used above for the common-emitter configuration. Write one equation that sets to zero the sum of all the currents away from node "A." Do the same for node "c" (the collector terminal) and solve the two equations simultaneously for V_c as a function of V_e (the emitter voltage).

Figure 17.6 shows a model for a microwave GaAS FET, which can be used in place of the BJT model for the same kind of amplifier analysis. Manufacturers'

Figure 17.5. A common-base RF amplifier.

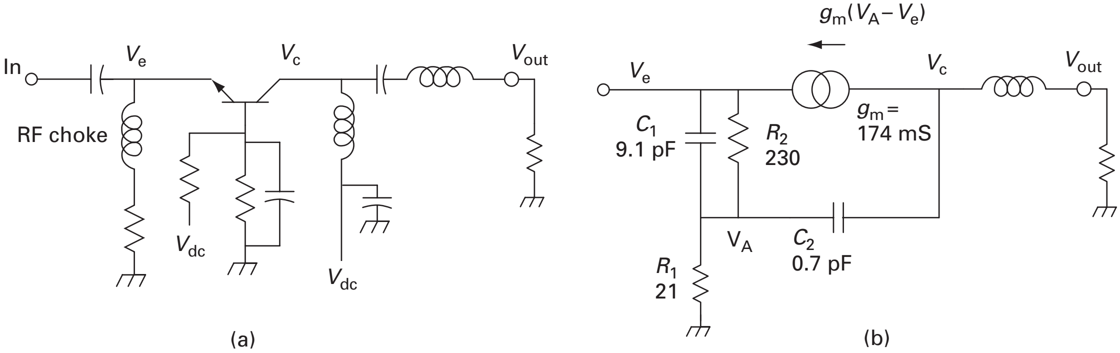

Figure 17.6. High-frequency GaAs FET model.

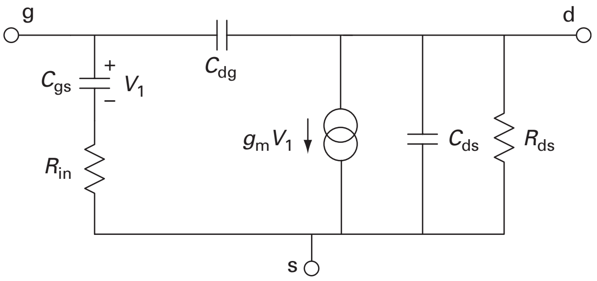

data sheets often include model circuits along with two-port data. There are many excellent textbooks devoted entirely to the amplifier design.

17.7 Amplifier noise

The output signal from any amplifier will always include some random noise generated within the amplifier itself. Most of the hiss from a radio receiver is due to noise generated by atmospheric electricity. But if the antenna is disconnected, the noise does not entirely disappear. The remainder is being generated within the receiver. Physical mechanisms that cause this noise include thermal noise (discussed below) and shot noise, which is noise due to the randomness in the flow of discrete charges – electrons and holes in transistors. The first stage in most receivers is an RF amplifier, and its noise usually dominates any other receiver-generated noise. This is easy to see; since this stage usually has considerable gain, its output power will be much greater than the noise power contributed by the second stage, so the second stage will hardly change the signal-to-noise ratio. In the same way, noise contributed by the third stage is even less important, and so forth.

Thermal noise

Thermal noise is such a universal phenomenon that it provides the very vocabulary for the definition of terms such as *receiver noise temperature* and *antenna noise temperature*. Let us examine the fundamentals of thermal noise. Any object, being hotter than absolute zero, converts thermal energy (molecular vibrational energy) into electromagnetic radiation: at infrared radiation for ordinary temperatures, but also visible radiation, if the object is red hot. Likewise, a resistor can convert thermal energy into electrical power. If two resistors are connected in parallel, each one delivers a tiny amount of electrical power (in the form of a random voltage waveform) to the other. If they are at equal temperatures, the power flow is equal in both directions.

How much power can a resistor generate by virtue of being hot? Answer: Any resistor can deliver kT watts per hertz, that is, kT is the spectral density of the power that a resistor of R ohms will deliver to a matched load (a load of impedance R+j0). Here k is Boltzmann's constant (1.38×10^{-23} joule/kelvin $= 1.38 \times 10^{-23}$ watt seconds/kelvin) and T is the absolute temperature. It is useful to remember that for $T_0 = 290$ K, which is universally taken as a standard reference temperature, the power available from a resistor, referred to 1 mW, is $= -114$ dBm/MHz, since $10 \log (1.38 \times 10^{-23} \times 10^3 \times 290 \times 10^6) = -114$.

To demonstrate that a resistor should produce this power, kT watts/Hz, an argument appropriate to radio engineering considers an antenna surrounded by a blackbody, i.e., an antenna within a cavity whose walls are at temperature T_1. We will be concerned with the spectral density at a particular spot frequency so

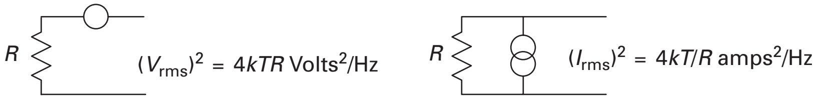
R
$(V_{rms})^2 = 4kTR$ Volts^2/Hz
R
$(I_{rms})^2 = 4kT/R$ amps^2/Hz

Figure 17.7. Equivalent circuits for a resistor as a noise source.

we can specify that the antenna be resonant, i.e., that it have a purely resistive impedance, R+j 0, at that frequency. Let a transmission line connect the antenna to a resistor R which is outside the blackbody but also at temperature T_1. We know the antenna will intercept blackbody radiation and that power will be transmitted through the line to the external resistor. We also know from thermodynamics that, in this isolated system, it is impossible for the resistor to get hotter than T_1; heat cannot flow from a colder to a hotter object. The only way to resolve this is for the resistor to produce an equal amount of power, which travels back to the antenna and is radiated back into the cavity. We can use some antenna theory to calculate the power. All antennas are directive; when used to receive, they have more effective area to intercept power from some directions than from other directions. But for any antenna, the average area turns out to be $\lambda^2/4\pi$ where λ is the wavelength. Blackbody radiation flux at long wavelengths is given by the Rayleigh–Jeans law, brightness = $2kT/\lambda^2$ watts/m^2/Hz/steradian. This includes power in two polarizations. Since any antenna can respond to only one polarization, we use half the Rayleigh–Jeans brightness to calculate the power the antenna puts on the transmission line:

$$P = \int \frac{B(\theta,\phi)}{2} A(\theta,\phi)\mathrm{d}\Omega = \frac{kT}{\lambda^2}\frac{\lambda^2}{4\pi}4\pi = kT. \qquad (17.11)$$

This value, kT, is then also the power a resistor of R ohms will deliver to another resistor of R ohms. It follows that the open-circuit noise voltage from a resistor is therefore $\langle V_n^2 \rangle = 4kTR$ volt^2/Hz. Figure 17.7 shows the Thévenin and Norton equivalent circuits of a resistor as a noise generator.

17.8 Noise figure

At any given frequency, a figure of merit for a receiver, an amplifier, a mixer, etc., is its *noise figure*, whose definition is as follows:

> Noise figure is the ratio of the total output noise power density to the portion of that power density engendered by the resistive part of the source impedance, with the condition that the temperature of the input termination be 290 K.

Noise figure is a function of frequency and of source impedance but (as we will see later) is independent of output termination. Consider Figure 17.8.

The voltage source V_n represents the thermal noise voltage from R_s, the resistive part of the source impedance. The source, V_s, is the actual signal voltage, if any. The internal noise of the amplifier can be considered to result

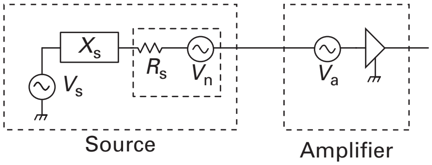

Figure 17.8. Equivalent circuit of an amplifier and signal source.

from another equivalent input noise source, V_a, at the input of the amplifier. With this model, the noise figure, as defined above, can be written in terms of V_n and V_a: $F = (V_n{}^2 + V_a{}^2)/V_n{}^2$. Note that, because the amplifier noise is represented by an equivalent generator at the input side, this expression does not contain G, the gain. Since $V_n{}^2$ is proportional to the source temperature, T_0, it is natural to assign the amplifier an equivalent noise temperature by writing the noise figure as $F = (T_0 + T_a)/T_0$. This amplifier noise temperature is just $T_a = (F-1)T_0 = (F-1) \times 290$ K. Conversely, the noise figure is given by $F = (290 + T_a)/290$. An amplifier can have a noise temperature less than its physical temperature. The dish-mounted amplifiers used for home satellite reception have typical noise temperatures of 30 K. (Refrigeration, however, can help; FET amplifiers for radio astronomy are often physically cooled to about 10 K and produce noise temperatures of only a few kelvins.)

So far we have not mentioned the signal voltage, V_s. Equation (17.12) shows that the noise figure also specifies the ratio of the input signal-to-noise ratio to the output signal-to-noise ratio:

$$\frac{\text{Input SNR}}{\text{Output SNR}} = \frac{V_s^2/V_n^2}{V_s^2/(V_n^2 + V_a^2)} = (V_n^2 + V_a^2)/V_n^2 = \mathrm{F}. \tag{17.12}$$

Cascaded amplifiers

The noise from an amplifier of only modest gain will not totally dominate the noise added downstream, so it is useful to know how noise figures add. Suppose amplifier 1, with noise figure F_1 and gain G_1, is followed by amplifier 2 with F_2 and gain G_2. Suppose further that they are matched at their interface so that $G = G_1G_2$ and that the output impedance of amplifier 1 is equal to the source impedance corresponding to the specified F_2. Figure 17.9 shows how to compute the overall noise figure.

The noise figure of the cascade is $F_{12} = F_1 + (F_2 - 1)/G_1$. It is interesting to calculate the noise figure of an infinite cascade of identical amplifiers as it is a lower limit to the noise we would get from any shorter cascade. You can verify that T_{infinity}, the equivalent noise temperature of the infinite chain, is given by $T_a/(1 - 1/G)$ where T_a and G refer to the individual identical amplifiers.

Finally, let us look at the overall noise figure of an amplifier preceded by an attenuator, as shown in Figure 17.10. Suppose the gain of the attenuator is G_{attn}. (The gain of an attenuator is less than unity; the gain of a 6-db attenuator, for example, is 1/4.) Referred to the amplifier input, the noise power engendered by

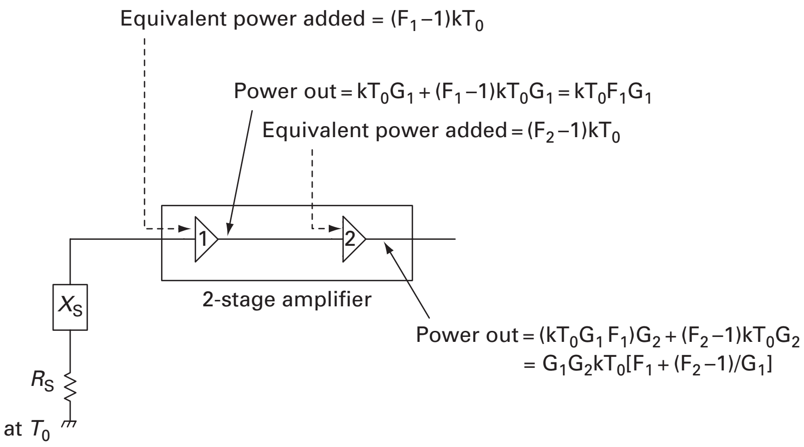
Equivalent power added = $(F_1-1)kT_0$
Power out = $kT_0G_1+(F_1-1)kT_0G_1=kT_0F_1G_1$
Equivalent power added = $(F_2-1)kT_0$
1
2
X_S
2-stage amplifier
Power out = $(kT_0G_1F_1)G_2+(F_2-1)kT_0G_2$
$= G_1G_2kT_0[F_1+(F_2-1)/G_1]$
R_S
at T_0

Figure 17.9. Overall noise figure of cascaded amplifiers.

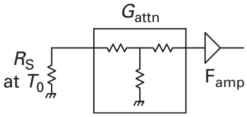

Figure 17.10. An amplifier preceded by an attenuator.

the source resistance is T_0G_{attn}. Since the amplifier still sees its standard source impedance, its total noise, referred to the input, is still $(F_{amp})T_0$.

The overall noise figure is therefore

$$F_{tot}=\frac{F_{amp}T_0}{G_{attn}T_0}=\frac{F_{amp}}{G_{attn}} \tag{17.13}$$

so, if an amplifier is preceded by an M-dB attenuator ($G_{attn}=10^{-M/10}$), the noise figure of the combination is M dB higher than the noise figure of the amplifier alone. We could just as well have derived this result by using the relation for cascaded devices. The noise figure of the attenuator, from the definition of noise figure, is $F_{attn}=kT_0/(G_{attn}kT_0)=1/G_{attn}$. The noise figure of the cascade becomes $F_{tot}=1/G_{attn}+(F-1)/G_{attn}=F/G_{attn}$ as before.

17.9 Other noise parameters

In what we have considered so far, the noise produced by a device, a transistor, amplifier, etc. is specified by a single parameter, its noise figure. But the noise figure depends on the source impedance from which the device is fed, which makes this parameter something less than a complete noise description of the device. We will see in Chapter 24 that a total of *four* noise parameters are sufficient to describe a device. The noise figure for any given source impedance can then be calculated from these four parameters which are R_{opt}, X_{opt}, F_{min}, and R_n. The (complex) impedance $Z_{opt}=R_{opt}+X_{opt}$ is the source impedance that yields the minimum noise figure, F_{min}. The "noise resistance," R_n, is a parameter that determines how fast the noise figure increases as the source impedance departs from Z_{opt}. We will see in Chapter 24 that the noise figure for an arbitrary source impedance is given by

$$F=F_{min}+(R_n/G_s)|Y_{source}-Y_{opt}|^2. \tag{17.14}$$

(Here Y_{opt} is just $1/Z_{\mathrm{opt}}$ and G_s is the real part of the source admittance.) We will also see that noise figure is somewhat deficient as a figure of merit. A piece of wire has $F = 1$ but is not a valuable amplifier since it has no gain. With a given transistor, circuit A might produce a lower noise figure than circuit B, but circuit A may have less gain. We will see that T_{infinity}, defined above, is the proper figure of merit.

17.10 Noise figure measurement

A straightforward determination of an amplifier's noise figure is possible if one knows its gain and has a spectrum analyzer suitable for measuring noise power density. Consider the common situation where we have an amplifier to be used in a 50-ohm environment. We connect a 50-ohm load to its input and use the spectrum analyzer to measure the output power density, S_{out}(watts/Hz), at the frequency of interest. We know that the portion of this power density engendered by the input load is kTG, where G is the amplifier gain. The noise figure is therefore given by $F = S_{\mathrm{out}}/(kTG)$. This assumes we have done the measurement at $T = 290\,\mathrm{K}$. If T was not 290, you can verify that $F = [S_{\mathrm{out}} - Gk(\mathrm{T} - 290)]/(290Gk)$. For low-noise amplifiers, a comparison method is used. This method requires a cold load and a hot load, i.e., two input loads at different temperatures, T_{hot} and T_{cold}. The amplifier is connected to a bandpass filter (whose shape is not critical) and then to a power meter (which needs to have only relative, not absolute accuracy). The ratio of power meter readings, hot to cold, is called the Y-factor. The noise temperature of the amplifier is then given by:

$$T_a = (T_{\mathrm{hot}} - YT_{\mathrm{cold}})/(Y - 1). \qquad (17.15)$$

Problems

Problem 17.1. Derive an expression for the power gain (output power/input power) for the two-port network described by Equations (17.1) and (17.2) when the load impedance is Z_L.

Problem 17.2. For a general two-port network, derive expressions for the Y parameters in terms of the Z parameters.

Problem 17.3. (a) A certain amplifier with 20 dB of gain has a third-order intercept of 30 dBm (one watt at the output). If the input consists of 0 dBm (0.001watt) signal at 100 MHz and another 0 dBm signal at 101 MHz, what will be the output power of the third-order products at 102 MHz and 99 MHz?

(b) Same as (a) except that the input signal at 100 MHz increases in power to 10 dBm (0.1 watt) while the input signal at 101 MHz remains at 0 dBm.

Problem 17.4. The Z-parameter description of a two-port corresponds in a one-to-one fashion to the equivalent circuit shown below in (a). Another circuit is shown in (b). Find expressions for Z_A, Z_B , Z_C and V in terms of the Z parameters to make the two circuits equivalent.

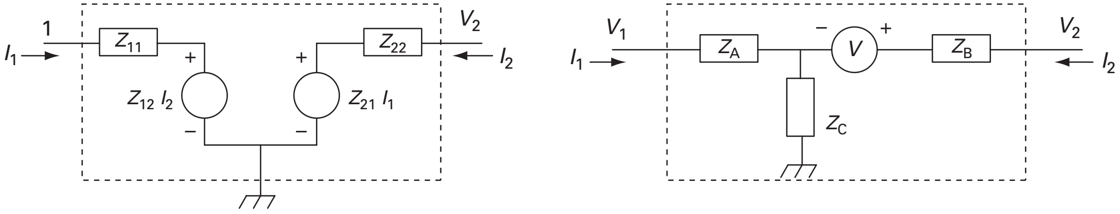

(a) (b)

Problem 17.5. The figure below at (a) shows a small-signal hybrid-π model for a common-emitter BJT transistor. (This model contains one more component than the model of Figure 17.2.)

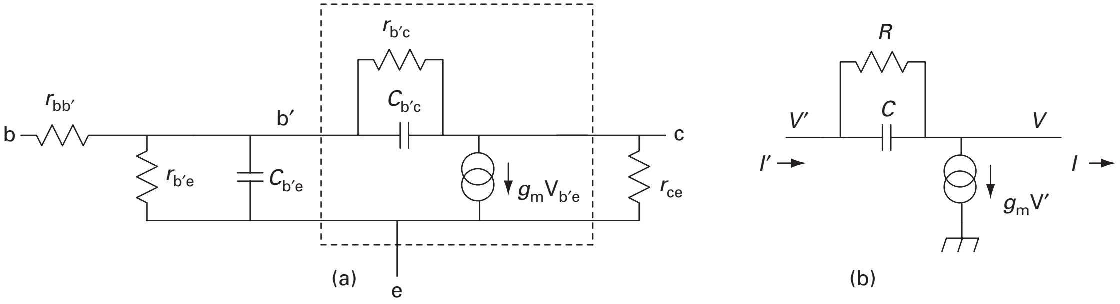

(a) (b)

The components form a simple ladder network, except for the portion inside the dashed line box. Show, for this box, the circuit in (b), that the relations between the input voltage and current and the output voltage and current are given by

$$V' = \frac{V}{1 - g_m Z} + \frac{IZ}{1 - g_m Z}$$

and

$$I' = \frac{V g_m}{1 - g_m Z} + \frac{I}{1 - g_m Z}.$$

Use these equations to make this box a new circuit element in the analysis program of Problem 1.3. The program will then be able to analyze BJT common-emitter amplifiers. (Note that the other two resistors and the other capacitor in this hybrid-π model can be included in a circuit as if they were external components.)

Example answer: For the MATLAB example given in Problem 1.3, add the element, “HY_PI” by inserting the following sequence of statements in the “elseif chain”:

```
elseif strcmp(component,HY_PI)==1
ckt_index=ckt_index+1; gm=ckt{ckt_index}; %gm
ckt_index=ckt_index+1; R=ckt{ckt_index};%R
ckt_index=ckt_index+1; C=ckt{ckt_index}; %C
Z=R/(1+1j*w*R*C);
Iold=I; I=(I+gm*V)/(1-gm*Z);V=(V+Iold*Z)/(1-gm*Z);
```

Problem 17.6. Show that the noise figure of an infinite cascade of identical amplifiers is given by $F_\infty = (F - 1/G)/(1 - 1/G)$. Assume that the amplifiers have a standard input and output impedance such as 50 ohms, that G is the gain corresponding to this impedance, that F is the noise figure corresponding to a source of this impedance, and that F_∞ is also to be with respect to this standard impedance. Hint: use the formula for a cascade of two amplifiers and the standard "infinite-chain-of-anything" technique – adding another link does not change the answer.

(This problem is not just academic. With only a few amplifiers the gain will be high enough to make the noise figure very close to F_∞ which is the best possible combination of the given amplifiers.)

Problem 17.7. Consider the balanced amplifier circuit shown below. The 3-dB, 90° hybrids are ideal. The amplifiers are identical and all impedances are matched. The individual amplifiers have power gain G and noise figure F_0. The hybrids are perfect, i.e., they have no internal loss and are perfectly matched.

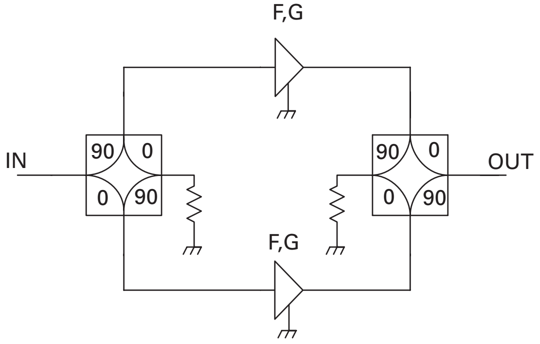

a. Show that the overall noise figure of this circuit is equal to the noise figure of the individual amplifiers.
b. If one amplifier dies, i.e., provides zero output, what is the overall noise figure?

Problem 17.8. We derived the overall noise figure, $F_{\text{tot}} = F\,(G_{\text{attn}})^{-1}$, for an amplifier preceded by an attenuator when the physical temperature of the attenuator is T_0, the standard 290 K reference temperature. Assuming now that the attenuator is at some different physical temperature, T_1, show that the overall noise figure is given by $(G_{\text{attn}})^{-1}$ $[G_{\text{attn}} + (T_1/T_0)(1 - G_{\text{attn}}) + \mathrm{F} - 1]$.

References

[1] Carson, R. S., *High Frequency Amplifiers*, New York: John Wiley, 1975.

[2] Gonzalez, G., *Microwave Transistor Amplifier Analysis and Design*, Englewood Cliffs, N.J.: Prentice Hall, 1984.

[3] Krauss, H. L., Bostian, C. W. and Raab, F. H., *Solid State Radio Engineering*, New York: John Wiley, 1980.

[4] Vendelin, G. D., Pavio, A. M. and Rohde, U. L., *Microwave Circuit Design Using Linear and Nonlinear Techniques*, New York: John Wiley, 1990.

CHAPTER

18 Demodulators and detectors

In communications equipment, "detection" is synonymous with demodulation, the process of recovering information from the received signal. The term detector is also used for circuits designed to measure power, such as square-law microwave power detectors. In this chapter we discuss various AM, FM, and power detector circuits. The demodulator is the last module in the cascade of circuits that form a receiver and at this stage the frequencies (IF or baseband) are relatively low. For this reason the detector (and, sometimes, the final IF bandpass filter) was the first receiver section to evolve from analog to digital processing, notably in broadcast television receivers. Receivers for ordinary AM and FM have, for the most part, continued to use traditional analog detectors, but receivers for the newer digital radio formats can also use their digital signal processors to demodulate traditional AM and FM broadcasts. Demodulators for OFDM and CDMA digital modulation formats are discussed in Chapter 22.

18.1 AM detectors

There are two basic types of AM detector. An envelope detector uses rectification to produce a voltage proportional to the amplitude of the IF voltage. A "product" detector multiplies the IF signal by a reconstituted version of the carrier. This detector is a mixer (see Chapter 5), producing sum and difference frequencies. The sum component is filtered away. The difference frequency component, at $f = 0$ (baseband), is proportional to the amplitude of the IF voltage.

18.1.1 AM diode detector

The classic diode envelope detector circuit for AM is shown in Figure 18.1. The input signal voltage, $V_{\text{sig}}\cos(\omega t)$, is usually provided by a tuned transformer at the output of the IF amplifier. This tuned circuit forms part of the IF bandpass filter.

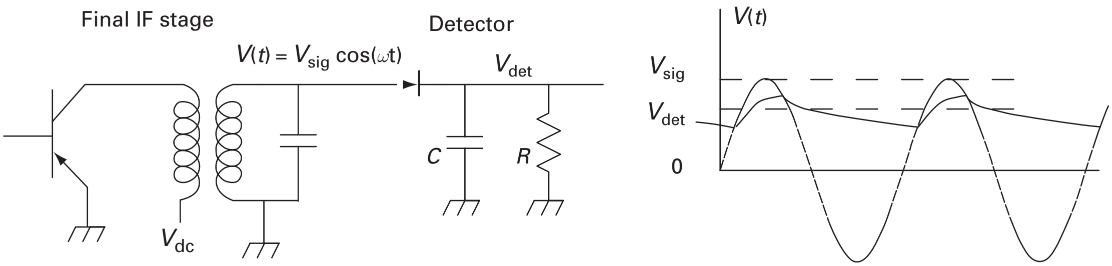

Figure 18.1. Diode envelope detector.

The diode and the parallel *RC* form a fading-memory peak detector. Except for the resistor, *R*, the output capacitor would remain charged to the maximum peak voltage of the input sine wave. The resistor provides a discharge path so that the detector output, V_{det}, can follow a changing peak voltage (AM). Since the input sine wave, an RF signal, has a much higher frequency than the amplitude modulation frequencies, the *RC* time constant can be made large enough so that the droop between charge pulses is much less than indicated in Figure 18.1. (Of course the time constant must also be small enough that output voltage can accurately follow a rapidly changing modulation envelope.) This detector, or any other envelope detector, is known as a *linear detector* since its output voltage is linearly proportional to the amplitude of the input sine wave.

Analysis assuming an ideal rectifier

Note that the detector of Figure 18.1 is identical to a simple half-wave rectifier capacitor-input power supply. As with the power supply, this circuit has poor regulation with respect to a changing load. But here we have a constant load resistance *R* (which we will assume includes the parallel input resistance of the subsequent audio amplifier). The equivalent circuit is shown in Figure 18.2. If the diode is modeled as a perfect rectifier (zero forward resistance and infinite reverse resistance) the analysis of this circuit is straightforward. The value of the diode's forward resistance, R_{d}, can be increased to account for source resistance. But here the high-*Q* resonant circuit at the detector input forces the waveform to remain sinusoidal. If we assume that *C* is large enough to make the output ripple negligible compared to the output voltage, the output voltage, V_{det}, can be calculated by noting that V_{det}/R must equal the average current through the diode, which is the average of $(V-V_{\mathrm{det}})/R_{\mathrm{d}}$ during the part of the input cycle

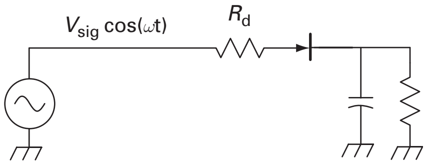

Figure 18.2. Diode envelope detector equivalent circuit.

when this expression is positive. The result is that V_{det} is proportional to the source voltage. (Curves showing output voltage vs. ωCR for various values of R_{d}/R are found in the power supply chapters of many handbooks.) The ratio of V_{det} to the peak source voltage is known as the detector efficiency. For a typical AM detector, $R \geq 10R_{\mathrm{d}}$ and $\omega CR \geq 100$. This gives a detector efficiency greater than 65% and an rms ripple less than about 1% of the dc output. With the assumed ideal rectifier, however, R could be any value. The analysis in the following section shows the limitations imposed on R by a real diode.

Analysis with a real diode

Here we will discard the perfect rectifier in favor of the standard diode for which $I_{\mathrm{diode}} = I_{\mathrm{s}} \exp(V_{\mathrm{diode}}/V_{\mathrm{T}} - 1)$, where I_{s} is the reverse saturation current and V_{T} is the so-called thermal voltage, 0.026 volts.[1] In the equivalent circuit of Figure 18.2, we will now let R_{d} be zero, i.e., we will assume that any voltage drop across the diode's bulk resistance is negligible compared to the drop across the junction. As before, the analysis to find V_{det} consists in equating V_{det}/R to the average current through the diode:

$$\frac{V_{\mathrm{det}}}{R} = \langle I_{\mathrm{diode}} \rangle = \frac{1}{2\pi} \int_0^{2\pi} I(\theta)\mathrm{d}\theta \tag{18.1}$$

where

$$I(\theta) = I_s\{\exp[(V\cos\theta - V_{\mathrm{dc}})/V_T] - 1\}. \tag{18.2}$$

This pair of equations is equivalent to

$$\frac{V_{\mathrm{det}}}{0.026} = \ln\left(\frac{V_s I(V)}{2\pi(V_{\mathrm{det}} + V_{\mathrm{s}})}\right), \tag{18.3}$$

where V_{s}, a "saturation voltage," is defined by $V_{\mathrm{s}} = I_{\mathrm{s}} R$ and

$$I(V) = \int_0^{2\pi} \exp(V\cos\theta/0.026)\mathrm{d}\theta. \tag{18.4}$$

Using a desk-top computer math utility to solve Equation (18.3) results in a set of curves (Figure 18.3) showing V_{det} vs. V for various values of V_{s}. Note that the detector output is very nonlinear for low-amplitude input signals when V_{s} is less than about 0.01 V. Suppose, then, that we pick $V_{\mathrm{sat}} = 0.01$ V. A germanium diode or zero-bias Schottky diode might have a saturation current of 10^{-6} A, which would

---

[1] The thermal voltage is given by kT/e, where e is the charge of an electron, k is Boltzmann's constant, and T is an assumed temperature, 300 K.

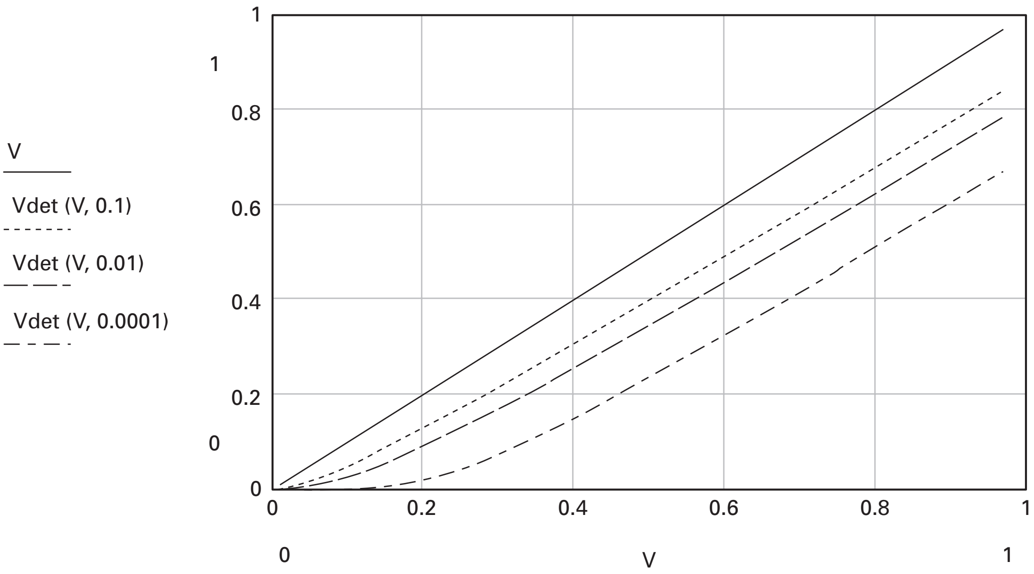

Figure 18.3. Detector output vs. input voltage for several values of $V_s = I_s R$.

then require that $R = 0.01/10^{-6} = 10$ k ohms, a convenient value. The low I_s of an ordinary silicon diode might require that R be more than 10^7 ohms, which would require the audio amplifier to have an inconveniently high input impedance. The power dissipated in the detector is V_{det}^2/R. This detector would typically produce, say, 2 V (to operate up in the linear range), which corresponds to a power dissipation of $2^2/10^4 = 0.4$ mW. For a given signal strength at the receiver input, the RF stages must have enough gain to produce 0.4 mW.

AC-coupled diode detector

Sometimes circuit considerations require that the detector input be ac coupled. In this case a dc return must be furnished for the detector diode. Such a circuit is shown in Figure 18.4 where an RF choke (a large value inductor) provides this

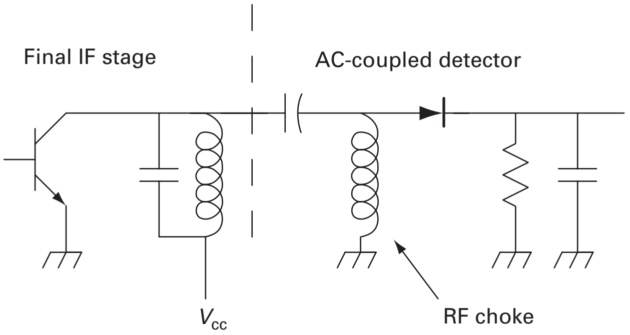

Figure 18.4. AC-coupled diode detector.

dc return path, forcing the average voltage at the left side of the diode to be zero, as in the circuits of Figures 18.1 and 18.2.

18.1.2 Product detectors

Demodulation of SSB and Morse code (cw) signals is usually done by a *product detector* which mixes (multiplies) the IF signal with a locally generated carrier. The resulting difference frequency components become the demodulated signal, while the sum frequency components are discarded. This is shown in Figure 18.5. The free-running oscillator is historically known as a *beat frequency oscillator* (BFO).

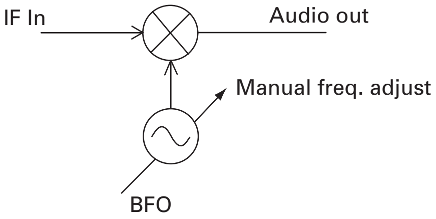

Figure 18.5. Product detector for SSB and CW.

For SSB reception of voice signals, the frequency of the BFO is manually adjusted until the audio sounds approximately natural. For Morse code reception, the BFO is deliberately offset to produce an audible tone, the "beat note." This was first done by the radio pioneer, Reginald Fessenden. In his heterodyne detector, the predecessor to the Edwin Armstrong's superheterodyne, the incoming signal voltage was combined with the voltage from a local oscillator,[2] which was a small arc source, an early negative resistance oscillator.

A product detector can also be used for AM demodulation. Again the signal is *multiplied* by a locally generated carrier. Here the local carrier must have the frequency and phase of the received carrier. Any error in frequency creates a strong audio beat note with the carrier of the received signal and an error in phase reduces the amplitude. Nevertheless, the product detector overcomes the limitations of diode envelope detectors; the input signal levels do not have to be as high and there is no low-signal threshold below which the detector is useless. When the AM signal is consistently strong (usually the case for most broadcast listeners) the local carrier can be a hard-limited version of the input signal. This works because, in double sideband AM, the modulated signal has the same zero crossings as the unmodulated carrier. The product detector shown in Figure 18.6 uses this method. This detector is commonly used for the video detector in analog television receivers.

[2] In his heterodyne patent of 1902, Fessenden proposed that the transmitting station send *two* signals, closely spaced in frequency. At the receiver, the nonlinear detector would produce an audible beat note. One of these signals could be continuous (not keyed). Instead, it became practical to produce the continuous signal at the receiver site using a *local* oscillator.

Figure 18.6. Product AM detector.

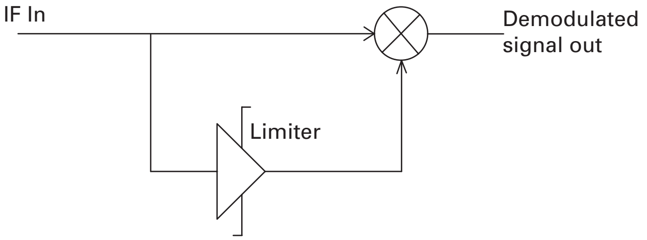

Figure 18.7. Synchronous AM detector.

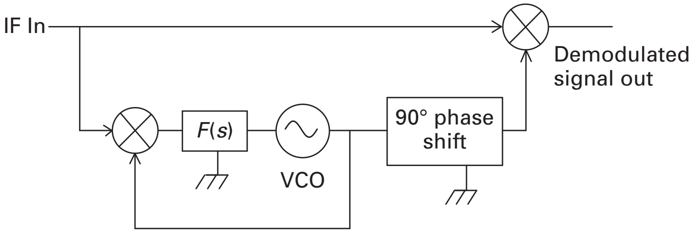

The *synchronous detector*, shown in Figure 18.7, is an improved product detector circuit in which a phase lock loop is used to generate the local carrier. The carrier of the input signal provides the reference signal for the loop. In a practical circuit, a limiting amplifier can be used at the reference input to make the loop dynamics independent of the signal level.

The PLL gives the synchronous detector a flywheel effect: the narrowband loop maintains the regenerated carrier during abrupt selective fades (common in short-wave listening). This prevents distortion, common in short-wave receivers, caused by momentary dropouts of the carrier. The PLL provides, in effect, a bandpass filter so narrow that its output cannot change quickly. Note the 90° phase shift network; if the phase detector is the standard multiplier (mixer), the VCO output phase differs from the reference phase by 90°. Without the network to bring the phase back to 0°, the output of the detector would be zero.

18.1.3 Digital demodulation of AM

The analog AM demodulators discussed above can be implemented as well in digital circuitry. Envelope detection via full-wave rectification of the IF signal can be done by simply taking the absolute value of the numbers produced by the A-to-D converter. For digital processing, the IF signal is often converted to two baseband signals, I and Q, which result from mixing the IF signal with $\cos(\omega_{\mathrm{IF}}t)$ and $\sin(\omega_{\mathrm{IF}}t)$. In this case, the amplitude of the signal is given by $(I^2 + Q^2)^{1/2}$, which can also be computed digitally, though not as easily as $|V|$. Synchronous detection can be done.

18.2 FM demodulators

A variety of circuits, often called *discriminators*, have been used to demodulate FM. Most of these circuits are sensitive to amplitude variations as well as frequency variations, so the signal is usually amplitude limited before it arrives at the FM detector. This reduces the noise output, since amplitude noise is eliminated (leaving only phase noise). In addition, it ensures that the audio volume is independent of signal strength.

18.2.1 Phase lock loop FM demodulator

We have already pointed out that a phase lock loop may be used as an FM discriminator. As the loop operates, the instantaneous voltage it applies to the voltage-controlled-oscillator (VCO) is determined by the reference frequency, which here is the signal frequency. The linearity of the VCO determines the linearity of this detector, shown in Figure 18.8.

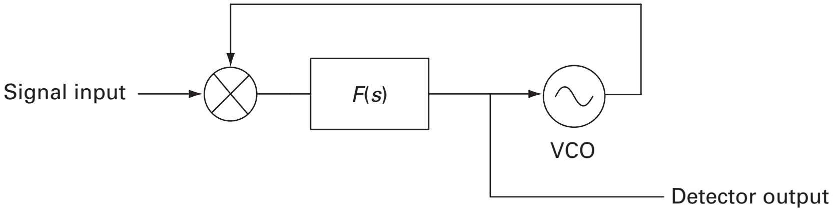

18.8. Phase lock loop FM detector.

18.2.2 Tachometer FM detector

A tachometer FM detector or "pulse counting detector", shown in Figure 18.9, is just a one-shot multivibrator that fires on the zero crossings of the signal. Each positive zero crossing produces a constant-width output pulse. The duty cycle of the one-shot output therefore varies linearly with input frequency so, by integrating the output of the one-shot, we get an output voltage that varies linearly with frequency.

Figure 18.9. Tachometer FM detector.

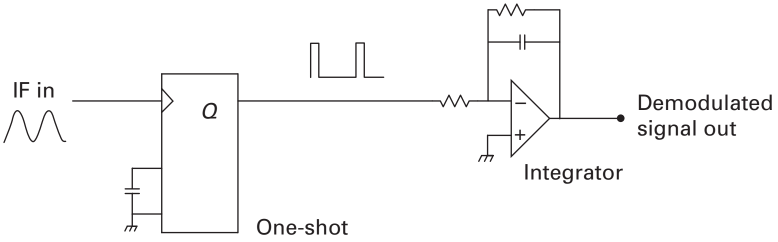

18.2.3 Delay line FM detector

The *delay line discriminator* is often used in C-band satellite television receivers to demodulate the FM-modulated video and sound. The IF frequency in these receivers is typically 70 MHz. Figures 18.10 shows a quarter-wave delay line (which could be a piece of ordinary transmission line) which delays the signal at one input of the multiplier. If the input signal is $\cos((\omega_0+\delta\omega)t)$, then the signal at the output of the delay line is $\cos((\omega_0+\delta\omega)(t-\tau)) = \sin((\omega_0+\delta\omega)t-\tau\delta\omega)$ and the baseband component at the output of the multiplier is $-\sin(\tau\delta\omega)$. For small $\tau\delta\omega$ this is just $-\tau\delta\omega$. The output voltage is thus proportional to the frequency offset, $\delta\omega$. If the delay line is lengthened by an integral number of half-wave lengths, the sensitivity of the detector is increased, i.e., a given shift from center frequency produces a greater output voltage.

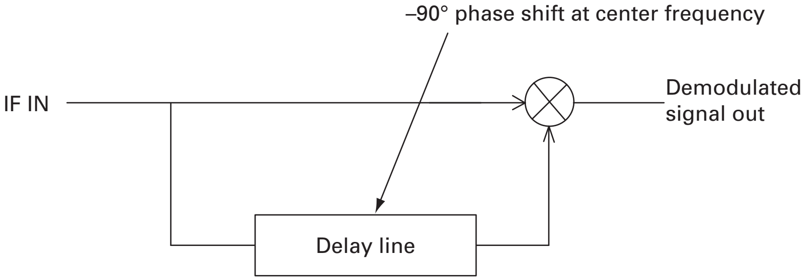
–90° phase shift at center frequency
IF IN
Demodulated signal out
Delay line

Figure 18.10. Delay line FM detector.

FM quadrature demodulator

The *quadrature FM demodulator*, shown in Figure 18.11, is the same as the delay line discriminator except that an *LC* network is used to provide the delay, i.e., a phase shift that varies linearly with frequency. These circuits are commonly used in integrated circuits for FM radios and television sound; the *LC* networks or an equivalent resonator is normally an off-chip component.

Figure 18.11. Quadrature FM detectors. (a) An LC network is used as the delay element. The multiplier provides the necessary resistive termination, (b) A voltage divider provides the necessary phase shift (see Problem 18.3).

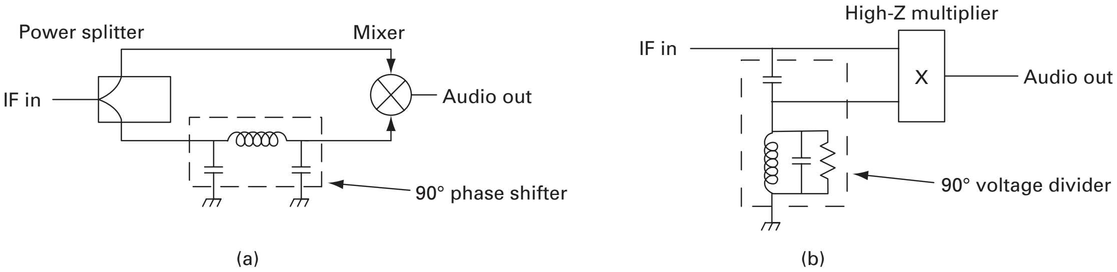
Power splitter
Mixer
IF in
Audio out
90° phase shifter
(a)
High-Z multiplier
IF in
X
Audio out
90° voltage divider
(b)

18.2.4 Slope FM detector

Slope detection, in which FM modulation is converted to AM modulation, is the original method to demodulate FM. The amplitude response of the IF bandpass filter is made to have a constant slope at the nominal signal frequency. An input signal of constant amplitude will then produce an output signal whose amplitude depends linearly on frequency. A simple envelope detector can detect this amplitude variation. An AM receiver can slope-detect an FM signal if detuned slightly to put the FM signal on the upper or lower sloping skirt of the IF passband filter. A refined balanced slope detector (Figure 18.12) uses two filters with equal but opposite slopes. The filter outputs are individually envelope-detected and the detector voltages are subtracted. This makes the output voltage zero when the input signal is on center frequency, f_0, and also linearizes the detector by cancelling even-order curvature, such as an $(f - f_0)^2$ term, in the filter shape.

The Foster–Seeley "discriminator," a classic FM detector, is an example of a balanced slope detector, though this is hardly obvious from the circuit, shown in Figure 18.13.

Note that the transformer has a capacitor across both its primary and its secondary. If the transformer had unity coupling, a single capacitor on one side or the other would suffice. The fact that there are two capacitors tells us the

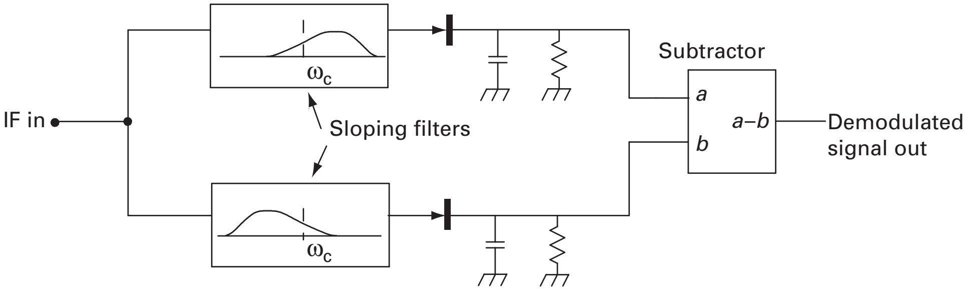

Figure 18.12. Balanced slope FM detector.

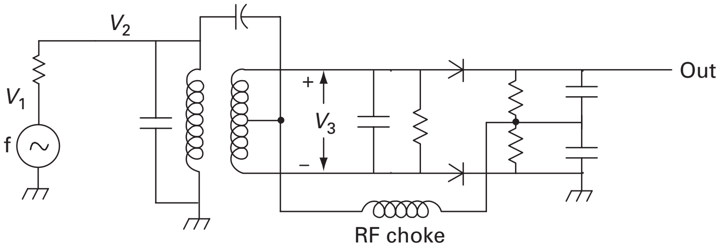

Figure 18.13. Foster–Seeley FM detector.

coupling is not unity; leakage inductance is an element in this circuit. In fact, the leakage inductance, and the magnetizing inductance, are used in a phase shift network, shown in Figure 18.14(a), used to produce the FM-to-AM conversion.

In this circuit, L_1 and L_2 are, respectively, the leakage and magnetizing inductances of the transformer. At the center frequency, f_c, V_2 lags V_3 by 90°. As the frequency increases from f_c, this lag increases. Mostly because of the change in relative phase over the operating region, the magnitude of the vector sum $V_2 + 0.5V_3$ decreases with frequency, while the magnitude of the vector sum $V_2 - 0.5V_3$ increases with frequency. These two combinations correspond to the outputs of the sloping filters in Figure 18.12.

Figure 18.15 is an equivalent circuit of the Foster–Seeley detector using an equivalent circuit for a transformer made up of the leakage inductance, the magnetizing inductance, and an ideal transformer. We will assume for convenience that the primary and secondary inductances and the coupling coefficient have values that give the ideal transformer a 1:1 ratio (see Problem 14.9).

Comparing Figures 18.15 and 18.13, you can see how the top and bottom halves of the transformer secondary are used to add $0.5V_3$ and $-0.5V_3$ to V_2. The

Figure 18.14. (a) Phase shift network: (b) magnitude vs. frequency of V_2+0.5V_3 and V_2−0.5 V_3.

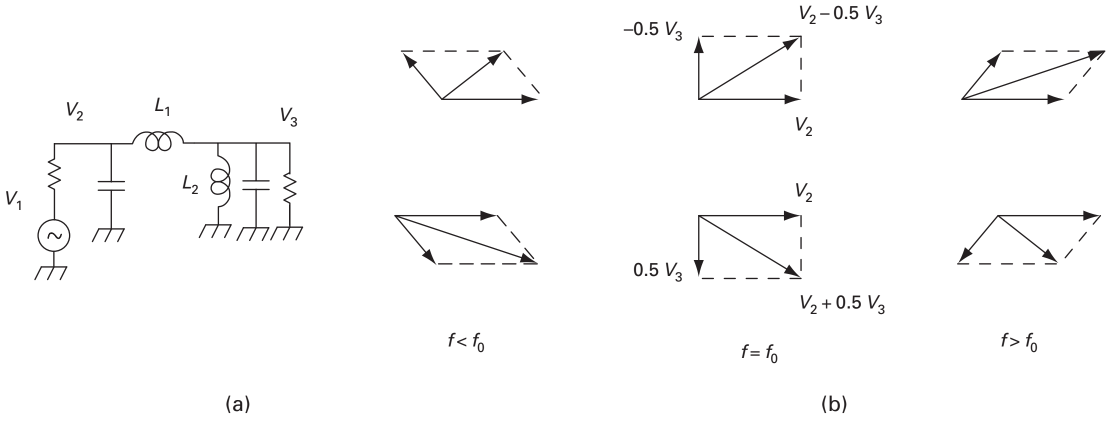
V_2
L_1
V_3
V_1
L_2
−0.5 V_3
$V_2 - 0.5\ V_3$
V_2
V_2
0.5 V_3
$V_2 + 0.5\ V_3$
$f < f_0$
$f = f_0$
$f > f_0$
(a)
(b)

Figure 18.15. Foster–Seeley detector equivalent circuit.

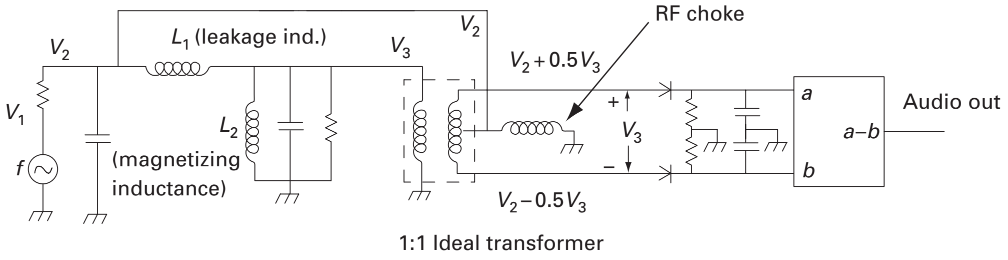
V_2
L_1 (leakage ind.)
V_3
V_2
RF choke
V_1
f
L_2
(magnetizing inductance)
$V_2 + 0.5V_3$
+
V_3
−
$V_2 - 0.5V_3$
1:1 Ideal transformer
a
b
a–b
Audio out

diode detectors produce voltages equal to the magnitudes of $V_2+0.5V_3$ and $V_2-0.5V_3$ and these magnitude voltages are subtracted, as in Figure 18.12, to produce the output. The RF choke provides a dc return for the diodes. Note that the subtractor can be eliminated by moving the grounding point in the secondary circuit to the bottom of the transformer's secondary winding. The dc blocking capacitor that bridges the transformer is only needed if there is a dc voltage on the primary winding, as when the signal source is the collector of a transistor biased through the primary winding. Finally, note that the capacitor and resistor in parallel with the magnetizing inductance L_2 in the equivalent circuit are actually located on the secondary side of the transformer which is equivalent to being on the left-hand side of the 1:1 ideal transformer.

18.2.5 FM stereo demodulator

Stereo FM, a compatible add-on dating to the 1960s, transmits an L+R (left plus right) audio signal in the normal fashion, and this signal is used by monaural receivers (or stereo receivers switched to "mono"). At the transmitter, The L−R audio is multiplied by a 38 kHz sine wave to produce a DSBSC (double sideband suppressed carrier) signal. This signal, well above the audio range, is added to the L+R audio signal, together with a weak 19 kHz sine wave "pilot" signal. The sum of these three signals, an example of *frequency division multiplexing* (FDM), drives the VCO (or equivalent) to produce the FM signal. At the receiver, the sum signal is demodulated by any ordinary FM demodulator. After demodulation, the L−R signal is brought back down to baseband by a product detector, i.e., a multiplier with a 38 kHz L.O. This L.O., which must have the correct phase (and therefore also the correct frequency), is derived by putting the pilot signal through a frequency doubler. This is shown in the block diagram of Figure 18.16.

Figure 18.16. FM stereo broadcast receiver broadcasting block diagram.

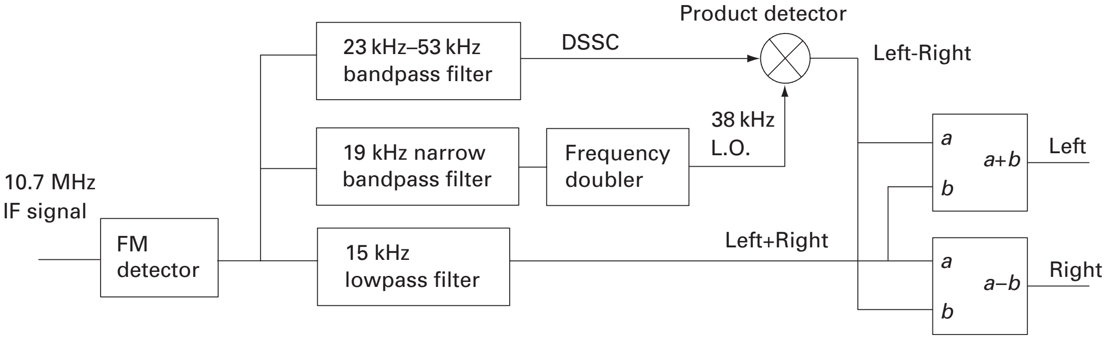

Note that the L−R signal has been doubly demodulated: first by the FM detector and then by the AM product detector. In this stereo system, the L−R signal is more susceptible to noise than the L+R signal. For “full quieting” stereo reception, about 20 dB more signal strength is required than for equivalent monaural reception. For this reason, FM stereo receivers are provided with manual or automatic switches to select monaural (L+R only) operation.

18.2.6 Digital demodulation of FM

When the IF signal has been converted into baseband I and Q signals, the phase of the IF sample is given by $\tan(\theta(t)) = Q(t)/I(t)$. Since frequency is the time derivative of phase, we take the differential of this expression: $(1+\tan^2 \theta)\delta\theta = \delta Q/I - Q\delta I/I^2$. Substituting Q/I for $\tan \theta$ yields

$$\delta\theta = \frac{I\delta Q - Q\delta I}{I^2 + Q^2}. \tag{18.5}$$

For each sampled I, Q pair, $\delta\theta$ is calculated from this pair and the previous pair. The resulting values of $\delta\theta$ are proportional to $\mathrm{d}\theta/\mathrm{d}t = \omega$. (Note that this could also be done with analog circuitry. Only two analog multipliers would be needed if the IF signal has passed through a limiter, making $(I^2 + Q^2)$ a constant.)

18.3 Power detectors

Square-law detectors are not used as demodulators but are used in laboratory instruments that measure power (wattmeters and rms voltmeters). If we are measuring a sine wave we know that the rms voltage is equal to the peak voltage divided by $\sqrt{2}$. When we know the shape of the waveform, a true square-law meter is not necessary. Even noise power can sometimes be estimated with other than a square-law device. The power of a Gaussian random noise source can be measured by averaging the output of a V^{2n} law device or a $|V|^n$ law device where $n = 1,2, \ldots$ The square-law device, however, is always the optimum detector in that it provides the most accurate power estimate for a fixed averaging time. When we need the optimum power measuring strategy or if we need to measure the rms voltage of an unknown waveform, we must average the output from a true square-law device. “True rms voltmeters” built for this purpose use a variety of techniques to form the square of the input voltage. Some instruments use a network of diodes and resistors to form a piecewise approximation of a square-law transfer function. Other instruments use a thermal method where the unknown voltage heats a resistor. The temperature of the resistor is monitored by a thermistor while a servo circuit removes or adds dc (or sine wave ac) current to the resistor to keep it at a constant temperature. The diode network

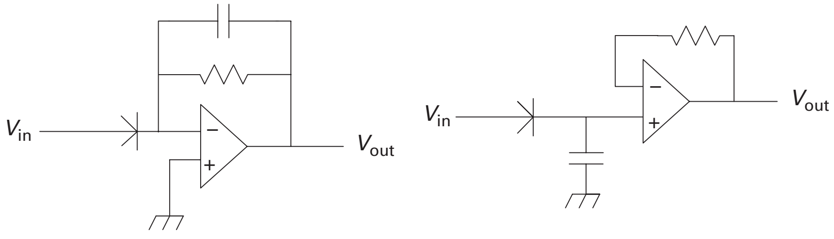

(a) (b)

Figure 18.17. (a) Simple diode power detector; (b) preferred detector.

requires large signals and the thermal method has a very slow response. Gilbert cell analog multipliers can be used to square the input voltage, but they are limited to relatively low-frequency signals. Generally, when the power is very low and/or the frequency is very high, and/or a very wide dynamic range is needed, the square-law detector uses a semiconductor diode. A simple (but not particularly recommended) circuit is shown in Figure 18.17. In this circuit the average current through the diode is converted to a voltage by the op-amp. The virtual ground at the op-amp input ensures that V_{in} is applied in full to the diode.

It is important to remember that the I vs. V curve for a diode does not follow a square law, even in a limited region. Rather, the law of the diode junction is exponential: $I = I_s[\exp(V/V_T) - 1]$ where V_T, the thermal voltage, is 26 mV. The diode is used at voltages much smaller than V_T so it is permissible to write $I = I_s[\,V/V_T + 1/2(V/V_T)^2 + 1/6(V/V_T)^3 + 1/24(V/V_T)^4]$. Obviously we can restrict the input voltage enough to neglect the last term. But what about the first term, the linear term? This dominant term will provide a current component that has the same frequency spectrum as the input signal; if that spectrum extends down to zero the output of the detector will be corrupted with extra noise. The third-order term will also have a baseband component. But if the input signal is a bandpass signal, the first-order and third-order terms are high-frequency signals that will be eliminated by whatever lowpass filter is applied to do the averaging (the capacitor in the above circuit). The simple square-law diode detector, then, is appropriate for measuring signals whose frequency components do not extend down into the baseband output spectrum of the detector circuit. (You can think of various two-diode balanced circuits to cancel the linear component but remember that the two diodes must be matched very closely to start with and then maintained at the same temperature.) The diode and op-amp circuit shown above serves to explain why diode detectors are used for bandpass signals but not for baseband signals. A better circuit – the preferred circuit – is shown in Figure 18.17(b). Here the sensitivity is not dependent on I_s and the circuit has a much better temperature coefficient. In this circuit no dc current can flow in the diode. The capacitor, however, ensures that the full ac signal voltage is applied to the diode. Expanding the exponential relation for the diode current and taking only the dc components we have

$$0 = \frac{-V_{\mathrm{dc}}}{V_{\mathrm{T}}} + \frac{\frac{1}{2}\langle {V_{\mathrm{in}}}^2 \rangle}{{V_{\mathrm{T}}}^2} \tag{18.6}$$

and

$$V_{\mathrm{dc}} = \frac{\frac{1}{2}\langle {V_{\mathrm{in}}}^2 \rangle}{V_{\mathrm{T}}} \tag{18.7}$$

which shows that the dc output voltage is indeed proportional to the average square of the input voltage.

Problems

Problem 18.1. Assume that in the envelope detector circuit shown below, the diode is a perfect rectifier (zero forward resistance and infinite reverse resistance) and that the op-amps are ideal.

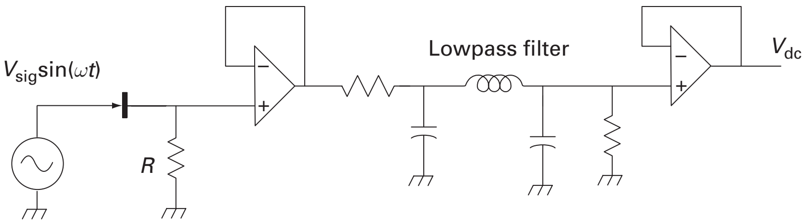

(**a**) Calculate the efficiency, V_{dc}/V.

(**b**) Calculate the effective load presented to the generator (the value of a resistor that would draw the same average power from the source).

Problem 18.2. Draw a vector diagram to show why an envelope detector will produce an audio tone when it is fed with the sum of an IF signal and a (much stronger) BFO signal.

Problem 18.3. Show that the voltage divider network in Figure 18.11(b) can produce an output voltage shifted 90° from the input voltage when the input voltage is at a specified center frequency, ω_0. Determine the position of the output phase when the input signal is slightly higher or slightly lower than ω_0.

Problem 18.4. When an interfering AM station is close in frequency to a desired AM station, an audio tone "beat note" is produced, no matter whether the receiver uses an envelope detector or a product detector. (In the case of an envelope detector, the beat note is produced because the amplitude of the vector sum of the two carriers is effectively modulated by an audio envelope. In the case of the product detector, the carrier of the undesired station acts as a modulation sideband and beats with the BFO.) Will the same thing happen with FM? Suppose two carriers (i.e., cw signals), separated by say, 1 kHz, appear in the IF passband of an FM receiver. Let their amplitudes be in the ratio of, say, 1:10. Draw a phasor diagram of the sum of these two signals. Does the vector sum have phase modulation? Will the receiver produce an audio tone? What happens in the case when the amplitudes of the two signals are equal?

Problem 18.5. Try the following experiment with two FM receivers. Tune one receiver to a moderately strong station near the low-frequency end of the band. Use the other receiver (the local oscillator) as a signal generator. (This receiver must have continuous rather than digital tuning.) Turn its volume down and hold it close to the first receiver so there will be local oscillator pickup. Carefully tune the second receiver 10.7 MHz higher in frequency until an effect is produced in the sound from the first receiver. What is the effect? Can you use this experiment to confirm your answer to Problem 18.4?

References

[1] Gosling, W., *Radio Receivers*, London: Peter Peregrinus, 1986.

[2] Landee, R. W., Davis, D. C. and Albrecht, A. P., *Electronics Designers' Handbook*, New York: McGraw-Hill, 1957.

[3] Rohde, U. and Whitaker, J., *Communication Receivers – DSP, Software Radios, and Design*, 3rd edn, New York: McGraw Hill, 2000.

CHAPTER

19

Television systems

Television system dissect the image and transmit the pixel information serially. The image is divided into a stack of horizontal stripes (“lines”) which are scanned left to right, producing a sequence of pixel (picture element) brightness values. The lines are scanned in order, one after the other, from top to bottom. Brightness values for each pixel are transmitted to the receiver(s). The image is reconstructed by a display device, whose pixels are illuminated according to the received brightness values. This chapter presents television technology in historical order: (1) the electromechanical system that Nipkov patented in 1884 but which was not demonstrated until 1923; (2) all-electronic television, made possible by the development of cathode ray picture tubes and camera tubes; and (3) digital television, which uses data storage and processing in the receiver, allowing the station to update the changing parts of the image, rather than retransmit the entire image for every frame. With the lowered data rate, the bandwidth needed previously to transmit one analog television program can now hold multiple programs.

19.1 The Nipkov system

Electronic image dissection and reconstruction were first proposed in the Nipkov disk system, patented in 1884, which used a pair of rotating disks, as shown in Figure 19.1. The camera disk dissected the image while the receiver disk reconstructed it. The receiver screen, a rectangular aperture mask, was covered by an opaque curtain containing a pin hole, illuminated from behind by an intensity-modulated gas discharge lamp. The position of the pin hole was analogous to the position of the illuminated spot on a CRT. This scanning pinhole was actually a set of N pinholes, arranged in a spiral on an opaque disk that rotated behind the aperture mask. Only one hole at a time was uncovered by the aperture. As the active hole rotated off the right-hand side of the aperture, the next hole arrived at the left-hand side, displaced downward by one scanning line. Identical holes in the transmitter disk

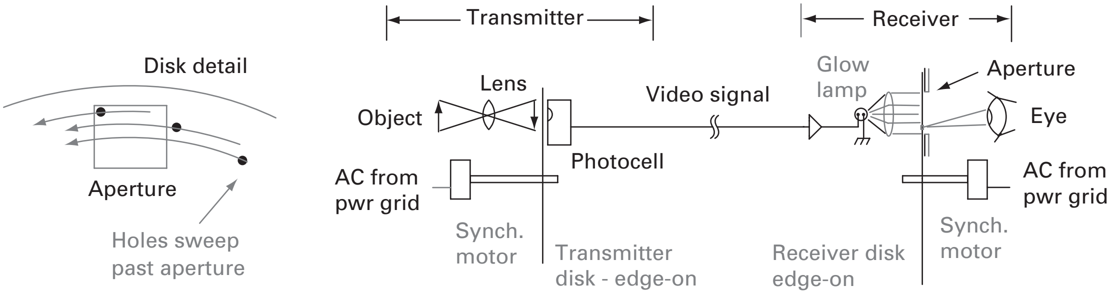
Transmitter
Receiver
Disk detail
Aperture
Holes sweep past aperture
Lens
Object
Photocell
AC from pwr grid
Synch. motor
Transmitter disk - edge-on
Video signal
Glow lamp
Aperture
Eye
Receiver disk edge-on
AC from pwr grid
Synch. motor

Figure 19.1. Nipkov rotating disk television system.

allowed light from the original image to hit a photocell. Of course the rotating disks had to be synchronized, but when the disks were driven by synchronous motors on the same ac power grid, it was only necessary to find the correct phase.

This primitive low-resolution system was finally demonstrated in 1923 after the invention of the photoelectric cell, vacuum tube amplifier, and neon glow lamp. While the eye views only a single illuminated spot, the persistence of human vision retains an image on the retina of the eye long enough to make the image appear complete if the entire screen is scanned at a rate more than about 20 times per second.[1] Some early experimental broadcasts were made with this very low resolution system in the U.S. on 100-kHz wide channels in the 2–3 MHz range.

19.2 The NTSC system

All-electronic television broadcasts were first made in Germany in 1935 and in England in 1936. The system used in the United States was proposed by the National Television Standards Committee (NTSC) of the Radio Manufacturers Association (RMA). Commercial broadcasting using the NTSC began on July 1, 1941. NBC and CBS both started television service that day in New York City. These stations and four others (Philadelphia, Schenectady, Los Angeles, Chicago) maintained broadcasts throughout World War II to some 10–20 thousand installed receivers. Compatible color broadcasting was added to the NTSC standard in the early 1950s. An engineering tour de force, this system effectively transmits simultaneous red, green, and blue images through the original 6-MHz channels in such a way that monochrome receivers are unaffected.

The NTSC standard specifies a horizontal-to-vertical aspect ratio of four-to-three for the raster (German for screen) with 525 horizontal lines, each scanned in 62.5 microseconds. About 40 of these lines occur during the vertical retrace interval, so there are some 525−40 or 485 lines in the picture. If the horizontal resolution were equal to the vertical resolution, the number of horizontal picture elements would be 4/3 × 485 = 646. The NTSC standard specifies somewhat less horizontal resolution: 440 picture elements. The horizontal retrace of a CRT requires about ten microseconds, so the active portion of each line is 62.5−10 = 52.5 microseconds. The maximum video frequency is therefore given by ½ × 440/52.5 = 4.2 MHz. (A video sine wave at 4.2 MHz would produce 220 white stripes and 220 black stripes.)

---

[1] The florescent “phosphorous” material on a television CRT faceplate provides addition persistence.

19.2.1 Interlace

The NTSC frame rate is 30 Hz, i.e., the entire image is scanned 30 times each second. The line rate is therefore 525 × 30 = 15 750 Hz. Interlaced scanning is specified. This is shown in Figure 19.2. Lines 1–262 and the first half of line 263 make up the first field. The second half of line 263 plus lines 264–525 make up the second field. The lines in the second field fit between the lines of the first field. Each field takes 1/60 sec, fast enough that the viewer perceives no flicker. If all 525 lines were scanned in 1/60 sec, the signal would require twice the bandwidth and the CRT beam deflection circuitry would require more power. Interlacing provides full resolution for fixed scenes but creates artifacts on moving objects (see Problem 19.5).

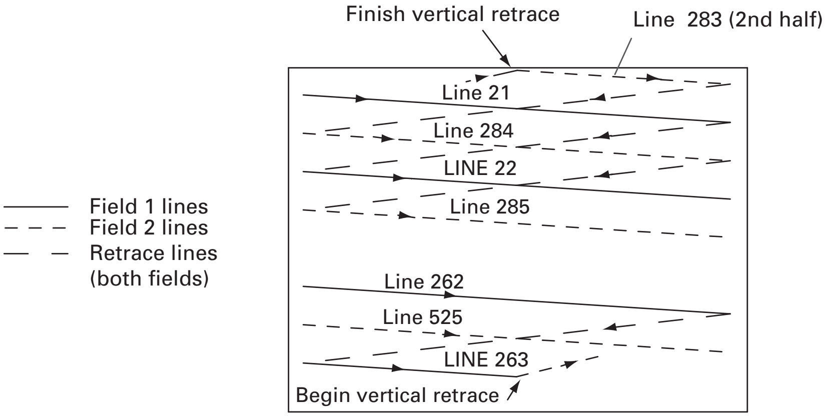

Figure 19.2. Interlaced scanning.

19.2.2 The video signal

The video signal amplitude-modulates the video transmitter. Synchronization pulses are inserted between every line of picture information. The NTSC system uses negative video modulation so that less amplitude denotes more brightness. The principal reason for using this polarity is that impulse interference creates black dots rather than more visible white dots on the screen of the receiver.

19.2.3 Synchronization

A horizontal sync pulse is inserted in the retrace interval between each scanning line and is distinguished by having a higher amplitude than the highest amplitude picture information, i.e., the sync pulse is “blacker” than the black level already blanking the beam during the retrace. The composite video, picture information plus synchronizing pulses, is shown in Figure 19.3. This waveform, which would be observed at the output of the video detector after lowpass filtering to remove the 4.5 MHz sound, shows three successive scan lines.

Television receivers have a threshold detector in the synchronization circuitry in order to look only at the tips of the sync pulses, i.e., the portion that is above the black level, and therefore totally independent of the video information. The burst of eight sinusoidal cycles at 3.579 545 MHz on the “back porch” of each synch pulse provides a reference for the color demodulator, described later. A vertical sync reference is provided by a series of wider sync pulses that occur near the beginning of the vertical blanking period, i.e., every 1/60 second at the end of every field.

Figure 19.3. NTSC video waveforms.

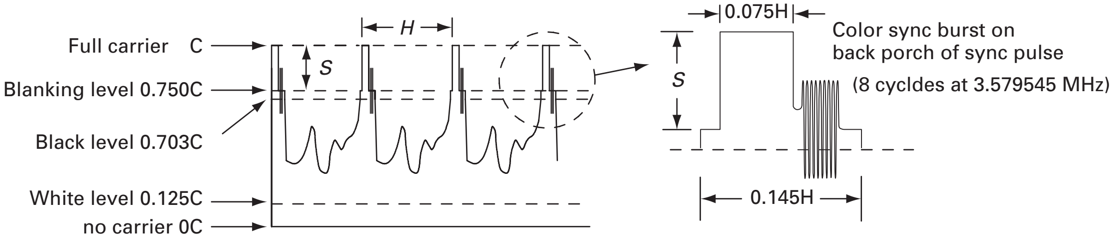

19.2.4 Modulation

Radio transmission of video information (television) requires that we modulate a carrier wave with the composite video signal. The NTSC system uses full-carrier AM modulation for the video. Since the NTSC video signal extends to 4.2 MHz, ordinary double-sideband AM would require a bandwidth of 8.4 MHz. To save bandwidth, the lower part of the lower sideband is removed at the transmitter by filtering, allowing a 6-MHz channel spacing. The resulting *vestigial sideband* signal consists of the entire upper sideband, the carrier, and a vestige of the lower sideband, as shown in Figure 19.4.

At the receiver, a low-frequency video component, because it is present in both sidebands, would produce twice the voltage that would be produced by a high-frequency video signal, present only in the upper sideband. This problem is corrected by using an IF bandpass shape that slopes off at the lower end such as shown in Figure 19.5. At the video carrier frequency the amplitude response is ½. (This response is as good as and simpler to obtain than the dotted curve where all of the double sideband region is reduced to ½.)

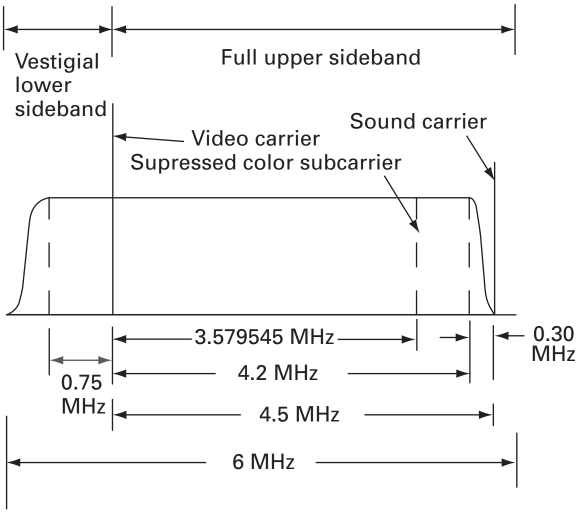
Vestigial lower sideband
Full upper sideband
Video carrier
Sound carrier
Supressed color subcarrier
3.579545 MHz
0.30 MHz
0.75 MHz
4.2 MHz
4.5 MHz
6 MHz

Figure 19.4. NTSC 6-MHz channel allocation.

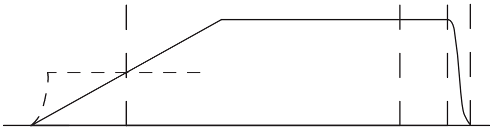

Figure 19.5. IF response to equalize the vestigial sideband.

For this IF equalization to work correctly, this IF filter should also have linear phase response. When surface acoustic wave (SAW) filters became available to determine the IF bandpass shape, this requirement was easily satisfied.

19.2.5 Sound

The audio or "aural" signal is transmitted on a separate carrier, 4.5 MHz higher in frequency than the video carrier. The NTSC system uses FM modulation for the audio component. The maximum deviation is 25 kHz, i.e., the maximum audio amplitude shifts the audio carrier 25 kHz. Normally the audio transmitter is separate from the visual transmitter and their signals are combined with a diplexer to feed a common antenna.

19.2.6 NTSC color

The NTSC standard for compatible color television was adopted in 1953. Like color photography, color television is based on a tricolor system. When two or more colors land at the same spot on the retina, their relative intensities determine the perceived hue. When three complete images in three suitably chosen primary colors are superimposed, the eye perceives a full-color image. The particular red, green, and blue standards specified in the NTSC standard were based on practical phosphors used for the dots on the faceplate of the CRT. Since color television requires the simultaneous transmission of three images, it is interesting to see how the color system can shoehorn three images into the 6-MHz channel originally allocated for a single monochrome image and do it in a way that made color broadcasts compatible with existing monochrome receivers. The solution to the bandwidth problem takes advantage of the fact that the monochrome video signal leaves empty gaps across the 6-MHz band. There is considerable redundancy in a typical picture. In particular, any given line is usually very much like the line preceding it (producing strong correlation at 62.5 microseconds) so the video signal is similar to a repetitive waveform with a frequency of 15 750 Hz, the horizontal scan frequency. If the lines were truly identical, the spectrum would be a comb of delta functions at 15 750 Hz and its harmonics. Since one line differs somewhat from the next, these delta functions are broadened, but the spectral energy is still clumped around the harmonics of the horizontal scanning frequency, leaving relatively empty windows which can be used to transmit color information. (Note that if the entire picture is stationary then the spectrum is a comb of delta functions spaced every 30 Hz and essentially all the bandwidth is unused.)

Instead of transmitting the red, blue, and green (RGB) signals on an equal basis, three linear combinations are used. One of these, the *luminance* signal, Y, is chosen to be the brightness signal that would have been produced by a monochrome camera: $Y = 0.299R + 0.587G + 0.114B$. The other two linear combinations are $I = 0.74(R-Y) - 0.27(B-Y)$ and $Q = 0.48(R-Y) + 0.41(B-Y)$.

The luminance signal directly modulates the carrier, just as in monochrome television, and monochrome receivers respond to it in the standard fashion. Each of the other two signals, I and Q, modulates a subcarrier (just as the Right-minus-Left audio signal modulates a 38-kHz subcarrier in FM stereo). The color subcarriers have the same frequency, 3.579 545 MHz, but they differ in phase by 90°. Figure 19.6 shows how two independent signals are transmitted and recovered using this quadrature AM modulation (QAM).

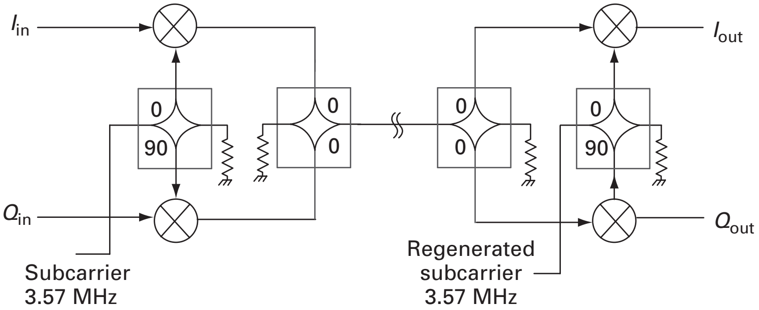

Figure 19.6. Quadrature AM modulator/demodulator lets the two color signals occupy one band.

The reference signal necessary to regenerate these local carriers is sent as a burst of about eight sinusoidal cycles at 3.579 545 MHz on the back porch of each horizontal sync pulse as shown in a previous figure. This "color burst" is used in the receiver as the reference for a phase lock loop. The color information, like the luminance information, is similar from line to line so its power spectrum is also a comb whose components have a spacing equal to the horizontal scanning frequency. The subcarrier frequency is chosen at the middle of one of the spectral slots left by the luminance signal so the comb of color sidebands interleaves with the comb of luminance sidebands – a frequency multiplexing technique known as spectral interlacing.

Compatibility is achieved because the spectral interlacing greatly reduces visible cross-talk between chrominance and luminance information. To see this, consider a very simple signal, a uniform color field such as an all-yellow screen. Since this field has a color, i.e., it is not black, white, or gray, there will be nonzero I and Q signals. In this example, since the color information is constant, the I and Q signals together are just a sine wave at the color subcarrier frequency. Their relative amplitudes determine the hue while their absolute amplitudes determine the saturation. One would expect this 3.58-MHz video component to produce vertical stripes. And the beam, as it sweeps across the screen, does indeed get brighter and dimmer at a 3.58-MHz rate, trying to make some 186 stripes in the 52.5 microsecond scan. But, on the next line, these stripes are displaced by exactly one half-cycle. The result is that the entire screen, rather than having 186 vertical stripes, has a fine-gridded checkerboard or "low visibility" pattern. Colored objects viewed on a monochrome receiver can be seen to have this low visibility checkerboard pattern. Note that there are some unusual situations where spectral interleaving does not work. If the image itself is like a checkerboard with just the right grid spacing the luminance signal will fall into the spectral slots allocated for the chrominance signal and vice versa. A herringbone suit for example, will often have a gaudy sparkling appearance when viewed on a color receiver. NTSC receivers eventually were equipped with comb filters to separate the chrominance and luminance signals, but at the expense of some vertical resolution.

The low-visibility principle is applied not only to avoid luminance–chrominance cross-talk but also to reduce the effect of the beat between the 4.5-MHz sound carrier and the color subcarrier. To take advantage of the low-visibility principle, television standards were modified slightly when color television was introduced. The sound carrier remained the same at 4.5 MHz above the video carrier. The relation between the horizontal scanning frequency, f_h, and the color subcarrier frequency, f_{SC}, was picked to be $f_{SC} = 227.5\, f_h$. Then the sound subcarrier-minus-color subcarrier beat was likewise made an odd number of half-multiples of the horizontal scanning frequency: $4.5\,\text{MHz} - f_{SC} = 58.5 f_h$. Putting these two relations together determines the horizontal frequency, 15 734.264 Hz, and the color subcarrier frequency, 3.579 545 MHz. The number of scanning lines remained at 525 so the vertical frequency changed from 60 Hz to $262.5f_{\mathrm{h}} = 59.940$ Hz. With these choices, the sound carrier is at 286 times the horizontal frequency. This would produce a high visibility pattern but the sound carrier is above the nominal video band and can be filtered out easily.

Color television receiver

The block diagram of Figures 19.7 shows the overall organization of an NTSC color television receiver. The first block is simply an AM radio receiver for VHF, UHF, and cable frequencies (about 52–400 MHz).

The carrier of the selected channel is translated to an IF frequency of 45.75 MHz and the IF bandwidth is about 6 MHz. The shape of the IF passband must be quite accurately set to equalize the vestigial sideband, to provide full video bandwidth, and to reject adjacent channels. This originally required careful factory adjustment of many *LC* tuned circuits but was later determined by the geometry in a single SAW filter. The IF signal contains the composite video (luminance and sync) plus the sound and color information. The sound and color signals, which are essentially narrowband signals around 4.5 MHz and 3.57 MHz, ride on top of the luminance signal. A 4.5-MHz bandpass filter isolates the sound signal in the block labeled "4.5-MHz FM Receiver." FM stereo sound uses the demodulator described in Chapter 18. The sound subcarrier is at 31.5 kHz so that the oscillator in the demodulator can be synchronized to the horizontal sweep trigger at 15.75 kHz. A 4-MHz lowpass filter eliminates the sound signal from the video. The resulting video signal, the "*Y*" signal, contains the brightness information and would produce the correct picture if sent to a monochrome picture tube.

Chrominance processor

The 3.57-MHz color burst on the back porch of the horizontal sync pulses provides a reference for the phase locked L.O. in this QAM demodulator. Not shown on the diagram is an electronic switch, the burst gate, which is controlled by the synchronization circuitry to apply the reference signal only during the burst period in order to improve the signal-to-noise ratio of the loop. The local carrier is fed to a product detector (multiplier) to demodulate the *I* signal. A 90° phase shift network proves a second local carrier, shifted in phase for a second product detector to demodulate the *Q* signal.

Sync processor

A comparator, with its threshold set at the black level, strips away the video signal, producing a clean train of sync pulses. As explained above, a simple RC differentiator then provides horizontal reference pulses and an RC integrator provides vertical reference pulses. A VCO, phase-locked to the horizontal reference pulses, provides a flywheel stabilized horizontal time base. The VCO operates at twice the horizontal frequency and is divided by 2 to provide 15 734 Hz for the PLL phase detector and to drive the horizontal deflection

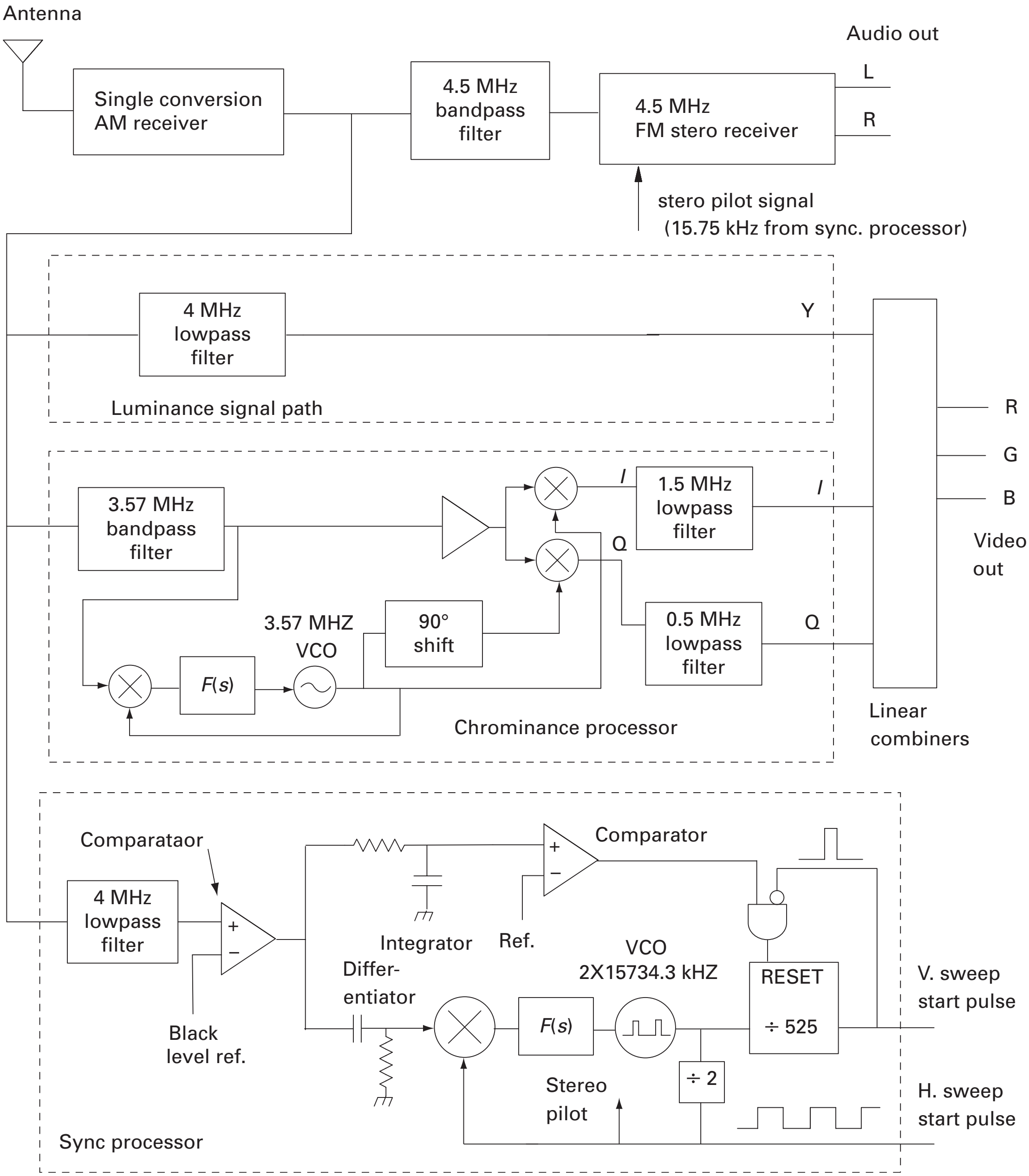
Antenna
Single conversion AM receiver
4.5 MHz bandpass filter
4.5 MHz FM stero receiver
Audio out
L
R
stero pilot signal (15.75 kHz from sync. processor)
4 MHz lowpass filter
Y
Luminance signal path
3.57 MHz bandpass filter
I
1.5 MHz lowpass filter
I
Q
3.57 MHZ VCO
90° shift
F(s)
0.5 MHz lowpass filter
Q
Chrominance processor
Linear combiners
R
G
B
Video out
Comparataor
4 MHz lowpass filter
+
−
Black level ref.
Integrator
Differ- entiator
Comparator
+
−
Ref.
VCO 2X15734.3 kHZ
F(s)
RESET
÷ 525
÷ 2
Stereo pilot
V. sweep start pulse
H. sweep start pulse
Sync processor

Figure 19.7. NTSC color television receiver block diagram.

circuitry. The VCO is also divided by 525 to provide an equally stable vertical time base. This divider must operate with the right phase for the picture to have the correct vertical alignment so the divide-by-525 counter has a reset input which will be triggered when the counter output has failed to coincide with several consecutive vertical reference pulses.

19.3 Digital television

The first digital television broadcasting in the U.S. began in 1998, following the 1996 adoption by the Federal Communications Commission (FCC) of the system developed by the Advanced Television Standards Committee (ATSC). Television stations were assigned new channels for digital broadcasting but have also continued their NTSC analog broadcasts during the digital phase-in period which, at this writing, is scheduled to end in 2009.

The ATSC standard incorporates MPEG-2 program compression (itself a standard, ISO/IEC 13818), which is based on the temporal and spatial redundancy of the video program material.[2] MPEG-2 is also used for digital television in Europe (DVB) and Japan (ISDB), for direct television broadcasts from satellites, and for DVDs. Digital processing and storage make it possible to exploit the redundancies of the program material to compress up to six standard-definition (480-line) programs or at least two high-definition (720 or 1080-line) programs into the same 6-MHz wide channel needed for a single NTSC program. (The same redundancies are exploited to a much smaller degree in the NTSC system with its temporally-interlaced fields and its frequency-interleaved luminance and chrominance signals.) The net data rate for the ATSC system is 19.3 Mbits/sec through a 6-MHz channel, after overhead for error correction. Let us estimate how much compression is involved when four standard-definition programs with typical frame-to-frame motion are compressed down to a total bit rate of 19.3 Mbits/sec. We will assume these programs are in the "480i" format, where 480×704 pixels are transmitted at a 30-Hz rate. Let us further assume there are three 8-bit numbers, a luminance value and two chrominance values, for each pixel. The overall data rate would be four channels $\times 480 \times 704 \times 8$ bits $\times$ 3(colors) $\times$ 30 Hz = 973 Mbits/sec. Dividing by 19.3 Mbits/sec, the compression factor is 50.4. When audio is included in the calculation, the factor increases slightly.

Digital television uses a three-step process. First, the video and audio program material, in raw digital bit streams, is compressed to packet streams having much lower bit rates. The video and audio packets, together with packets of ancillary data and null packets (padding) are merged together to form a stream of *transport packets*, each one containing 187 data bytes. The packets have headers to identify the program (since a stream may contain several independent programs) and the type of packet: video, audio, or, ancillary information, or null. Second, the stream of packets is run through a two-stage forward error correction (FEC) encoder which adds redundant bits to each packet so that the receiver can detect and correct transmission errors caused by noise and interference. Third, the bits in the stream of expanded packets produce the RF signal for transmission. This signal is single-sideband suppressed-carrier AM (SSBSC) except that a vestige of the lower sideband remains (as in the NTSC standard) and the carrier is not completely suppressed; a pilot carrier is inserted as a reference for synchronous detection in the receiver. At the receiver, the three steps occur in the opposite order. The bit stream is demodulated from the RF signal. The FEC-encoded packets are then decoded, correcting transmission errors and producing the original packets of compressed data. These packets are then separated, according to program, and each program is decompressed into the original audio and video bit streams.

---

[2] The ATSC system does not use the audio part of MPEG-2 for audio, but instead uses AC-3 (Dolby digital) compression, which provides each program with up to five full audio channels plus a low-frequency subwoofer channel for "surround sound."

19.3.1 MPEG encoding

The picture is encoded (compressed) at the lowest level in blocks of eight pixels by eight pixels. (Note that "compression" and "decompression" of the program data are usually called coding and encoding, the same terms used for the processes used to anticipate and correct transmission errors.) Video usually has a high degree of spatial correlation, i.e., adjacent pixels tend to have similar brightness and color. When an $(n \times n)$-pixel picture is encoded using a two-dimensional discrete Fourier transform (DCT), the resulting coefficient matrix can be inverse-tranformed to exactly reconstruct the picture. But if the coefficients are coarsely quantized, i.e., rounded to use fewer bits before the inverse transformation, the reconstructed picture is found to retain most of its original quality.[3] Moreover, because of spatial correlation, these coefficients tend to concentrate in one corner of the coefficient matrix. Away from this corner, the coefficients have low values. Thus, a great many of the coefficients can be represented as zeros and the rest by numbers of only a few bits, so DCT encoding providing a significant amount of compression.

Except for scene changes, the differences from one frame to the next are mostly due to motion of elements within the picture (subject motion) or motion of the picture as a whole (camera motion). From frame to frame, a given block will therefore mostly just shift its position somewhat. In the ATSC system, for each 16 × 16 block (*macroblock*) in a new frame, the MPEG encoder determines the displaced macroblock in the previous frame that provides the best match by minimizing the sum of the absolute values of the differences of the pixel brightness values. The position of the displaced macroblock is specified by a *motion vector*, e.g., 1,1 would indicate that the best match macroblock is shifted one pixel up and one pixel to the right. These motion vectors are part of the video update information. The rest of the information consists of DCT-tranformed *differences* between the pixel values in these displaced "prediction" macroblocks and the new pixel values in the block being updated. At the encoder, the macroblock is subdivided into four 8 × 8 blocks, and the 64 pixel differences for each block which, again, have considerable spatial correlation, are transformed using an 8 × 8 DCT.[4] (It is computationally more efficient to transform small blocks, and the 8 × 8 size was found adequate to provide the desired compression.) When most of the coefficients in a transformed block are negligible, the resulting stream of digital numbers consists mostly of zeros, and run-length encoding (specification of the number of consecutive zeros) is used to increase the degree of compression. The number of consecutive zeros is increased by using a certain zigzag readout ordering of the coefficient differences. In addition, variable length coding is used for the nonzero transform differences; the viewer can tolerate coarser quantization for the high spatial frequencies.

---

[3] The same principle is used for audio compression. Blocks of audio voltage samples are transfomed and the resulting coefficients are compressed and transmitted.

Note that not all the frames can be predicted from previous frames; channel surfers need the picture to change promptly and, of course, a change of scene requires all-new data. Occasional *refresh frames* (tagged as such) are therefore inserted into the stream of prediction/correction frames. In a refresh frame, every block is transmitted. Between refresh frames, only changing macroblocks need to be transmitted. As in the NTSC system, the two color signals can have lower spatial resolution than the luminance signal, just as a black and white photograph can be "colorized" using a relatively broad brush. In the ATSC system, each of the two color elements has half the resolution (1/4 the number of pixels) as the luminance element, further compressing the video signal. The output from the MPEG encoder, together with the audio material, is a string of 188-byte packets: 187 data bytes plus one sync byte. Note that the packets are asynchronous; their lengths and mix is program dependent. Video and audio are both time-stamped so that they can be aligned by the decoder. After leaving the MPEG encoder, the bits in the 187 data bytes are "pseudo-randomized" by multiplication with the bits from a pseudo-random shift register sequence. This step, which is undone in the receiver, just ahead of its MPEG decoder, ensures that the spectrum is flat, even if the output of the encoder is interrupted, and makes it easier for the receiver to find the transmission packet sync byte which is added during the forward error correction encoding.

19.3.2 Forward error correction

The ATSC system applies two levels of forward error correction. "Forward" means that bits added in the encoding process make it possible to correct errors rather than simply detect them. Of course, this works only up to a limit; when the signal-to-noise ratio drops below a threshold, uncorrectable errors begin to produce visible artifacts and beyond this, the signal soon suffers total

---

[4] The DCT can be done by multiplying the (8 × 8) pixel or pixel difference matrix by an (8 × 8) transform matrix. The inverse of this matrix provides the inverse transformation. However, there are faster methods, analogous to the FFT algorithm for calculating the DFT.

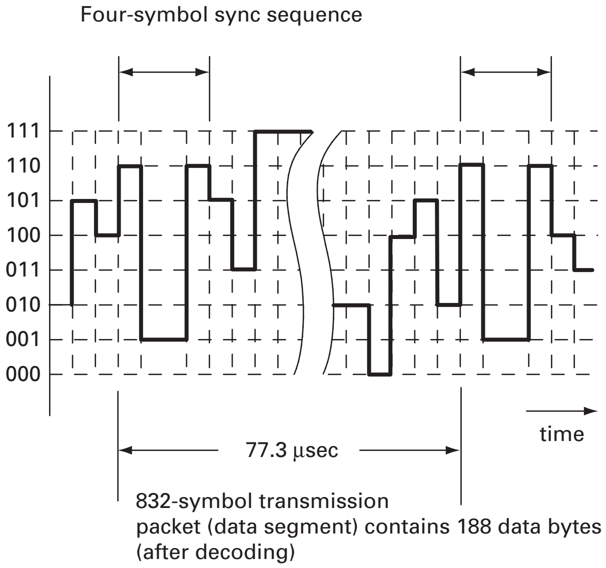
Four-symbol sync sequence
111
110
101
100
011
010
001
000
time
77.3 µsec
832-symbol transmission packet (data segment) contains 188 data bytes (after decoding)

Figure 19.8. ATSC transmission packet.

degradation – the well known "threshold effect" in digital communication links. The bit-randomized MPEG packets are first processed by a Reed–Solomon encoder which adds 20 parity bytes to the 187 data bytes increasing each packet to 207 bytes. This Reed–Solomon encoding makes it possible to correct up to 10 defective bytes per packet, no matter how many bit errors these bytes contain. The next stage is an interleaving (shuffling) of bytes within blocks of 52 of the 207-byte packets. This interleaving makes it less likely that a wide noise burst will leave more than 10 bad bytes in any packet after de-interleaving has been done at the receiver. The second level of FEC encoding is a convolutional or trellis encoding, which is effective against white noise. The trellis encoder produces three output bits for every two input bits, bringing each packet up to $207 \times 3/2 \times 8 = 2484$ bits, or $2484/3 = 828$ three-bit (eight-level 8-VSB) symbols. The trellis encoding is a flow process; each bit triplet leaving the trellis encoder is calculated from only the last four pairs of input bits. Finally, a four-symbol sync sequence is added for a grand total of 832 8-VSB symbols per transmitted packet, as shown in Figure 19.8.

With the concatenated Reed–Solomon and Trellis error correction coding, the threshold of visibility for the ATSC system occurs at a S/N ratio of 14.9 dB. At this threshold, the data segment (832-symbol transmission packet) error rate is about 1.93×10^{-4} or 2.5 segment errors per second.[5] The FCC requires that over the area of service, a broadcaster's ATSC signal must exceed the threshold by 7 dB. Compared to NTSC, ATSC broadcasting requires less than 1/10 as much transmitter power. (Nearly all the ATSC transmitted power bears information, while much of the NTSC power goes into the carrier and horizontal sync pulses.)

---

[5] These numbers are taken from ATSC Document A/54A, "Recommended Practice Guide to the Use of the ATSC Digital Television Standard."

19.3.3 RF modulation

The three-bit symbols are converted by an A-to-D converter into an eight-level analog signal which amplitude modulates the carrier wave. The carrier is suppressed except for a low-amplitude pilot. Figure 19.9 shows a modulator circuit whose input is the stream of transmission packets, each containing 832 three-bit symbols, including the four-symbol segment sync. A filter eliminates all but a vestige of the lower sideband.

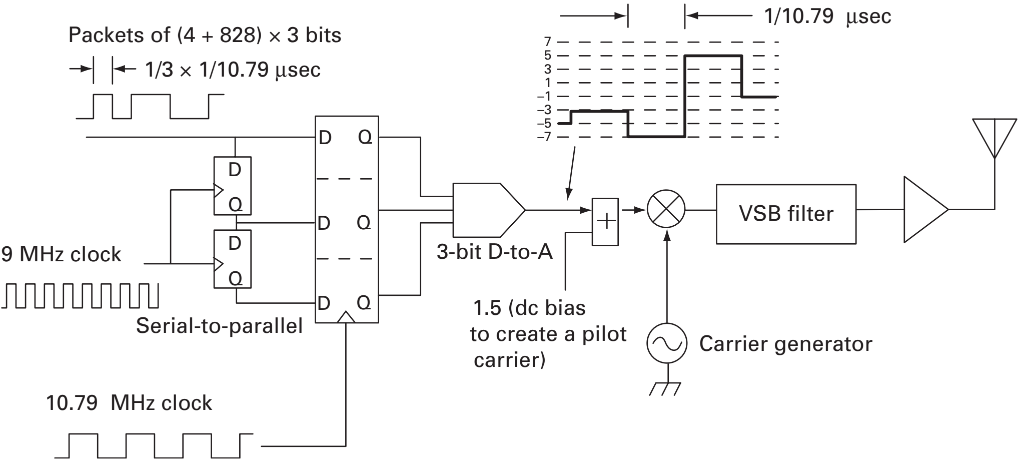
Packets of (4 + 828) × 3 bits
1/3 × 1/10.79 μsec
1/10.79 μsec
7
5
3
1
−1
−3
−5
−7
D
Q
D
Q
D
Q
D
Q
D
Q
2/3 × 10.79 MHz clock
Serial-to-parallel
3-bit D-to-A
+
VSB filter
1.5 (dc bias to create a pilot carrier)
Carrier generator
10.79 MHz clock

Figure 19.9. ATSC 8-VSB modulator.

Figure 19.10 compares the spectrum of an ATSC signal with that of an NTSC spectrum.

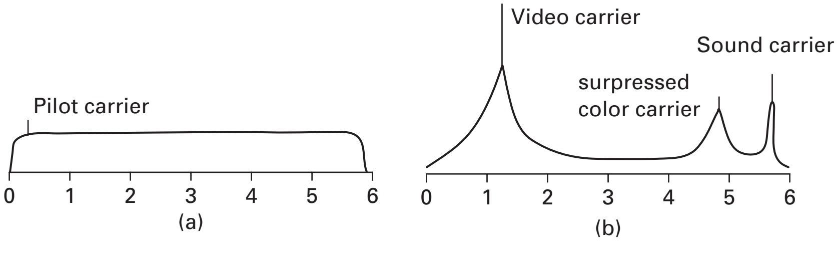

Figure 19.10. Comparison of the ATSC spectrum (a) and the NTSC spectrum (b).

19.3.4 HDTV receiver

The ATSC standard provides 18 different video formats, ranging from high-definition 1920 × 1080 pixels, interlaced with 60 fields/second or noninterlaced with 30 or 24 frames/second, down to 640 × 480 pixels, interlaced with 60 fields/second or noninterlaced with 60, 30, or 24 frames/second. The receiver identifies the video format being used and decodes the video. It then scales and interpolates the decoded image to produce a picture that fits the native resolution of the display screen, unless the program happens to match the native resolution. When the screen does not have the same aspect ratio (16:9 or 4:3) as the program, the receiver will produce a "letter box" display, with the picture between black horizontal strips, or a "keyhole" display, with the picture between black vertical strips, or the view may be able to opt for a distorted picture that fills the screen. Figure 19.11 is a simplified block diagram for an ATSC television receiver.

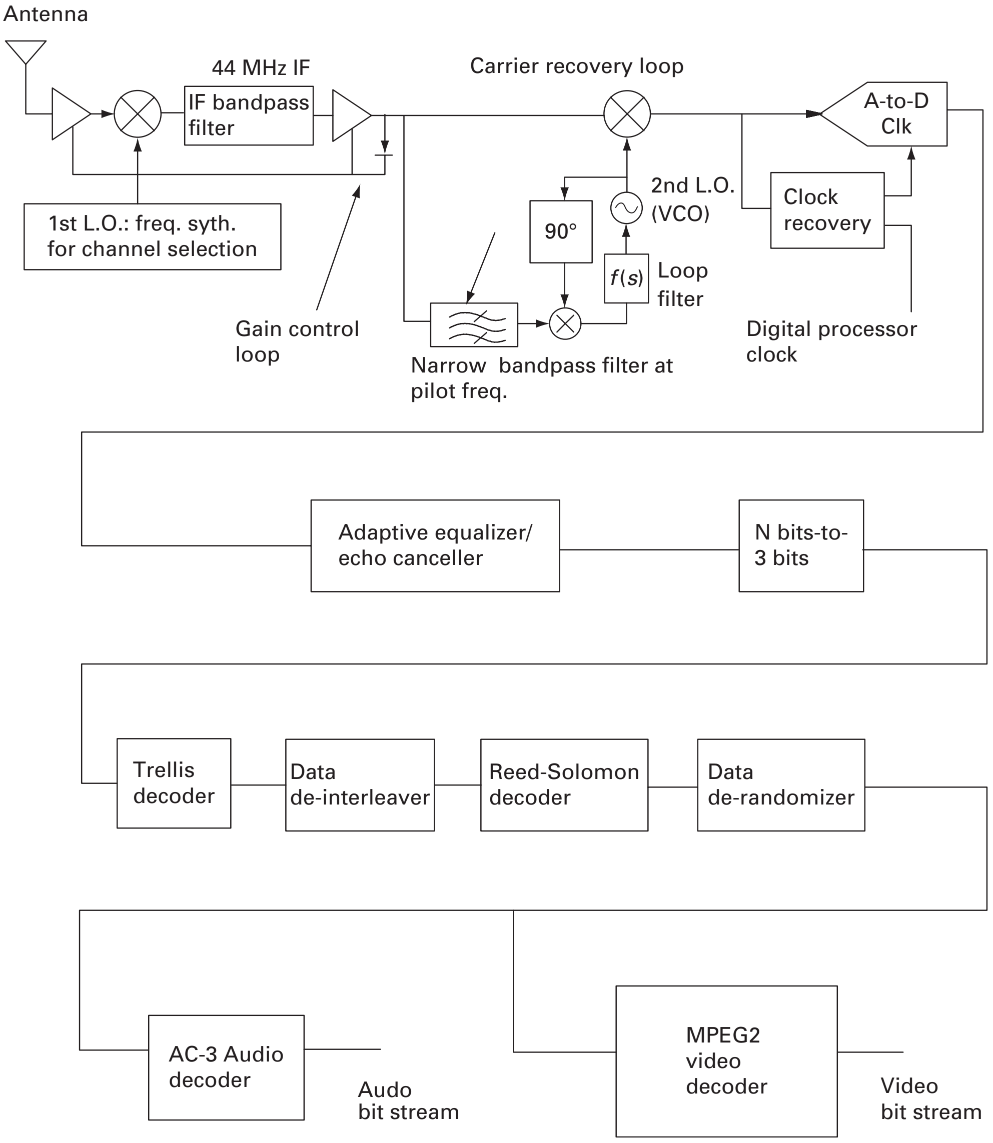
Antenna
44 MHz IF
IF bandpass filter
Carrier recovery loop
A-to-D Clk
1st L.O.: freq. syth. for channel selection
2nd L.O. (VCO)
90°
Clock recovery
f(s)
Loop filter
Gain control loop
Narrow bandpass filter at pilot freq.
Digital processor clock
Adaptive equalizer/ echo canceller
N bits-to-3 bits
Trellis decoder
Data de-interleaver
Reed-Solomon decoder
Data de-randomizer
AC-3 Audio decoder
Audo bit stream
MPEG2 video decoder
Video bit stream

Figure 19.11. An ATSC television receiver block diagram.

The conventional RF sections are all in the first row of blocks. The signal is first mixed down to a first-IF frequency, often around 44MHz. Just as in an NTSC receiver, the L.O. for the first mixer selects the channel. The amplified IF signal is synchronously detected by multiplying it by a sine wave at the carrier frequency. This sine wave is produced by a VCO referenced to the pilot carrier in a phase lock loop. Without the 90° phase shifter in the loop, the VCO output would be 90° out of phase with the pilot carrier and there would be no output from the synchronous detector. (If the consortium that developed the ATSC standard had opted for quadrature AM (QAM), there would have been two final mixers producing I and Q signals, rather than just a single mixer producing the I signal.) The detected signal, which is a zero-frequency IF "baseband" signal, is fed to an A-to-D converter. While the amplitude of the signal has nominally only eight levels (three bits), at this point it is still dithered by multipath echos and noise, so the A-to-D converter may have eight or ten bits. The clock rate of this converter may be a multiple of the symbol rate in order to recover the symbol clock with high precision. The clock recovery block finds the four-symbol segment sync sequence and phase locks the recovered symbol clock to it.

The equalizer is an adaptive filter which serves two purposes. It corrects the overall frequency response (transfer function) to minimize intersymbol interference (see Chapter 22) and it cancels echos – multipath interference. Both of these functions are essential in order to reproduce the transmitted signal well enough to distinguish one of eight possible amplitude values for each symbol. The usual form of the equalizer is a transversal FIR filter with programmable tap weights (amplitude and phase). The weighted outputs from the taps are summed to form the filter output. Every 313th transmission packet contains a "training" sequence, which is a reference waveform used by the receiver to determine the proper weights for the equalizer filter taps. Once the equalizer and the overall gain are correctly set, the signal can be sampled at the symbol rate × 1 to reproduce the data stream shown in Figure 19.8. Subsequent processing steps then undo the various stages of encoding. A trellis decoder (Viterbi decoder) is followed by a block that de-interleaves the bytes. Then the Reed–Solomon encoding is decoded. Finally, the pseudorandom coding is undone and the data can proceed to the MPEG decoder, which outputs synchronized video and audio.

Problems

Problem 19.1. Suppose the NTSC signal from a television receiving antenna consists of the direct signal (via the line-of-sight path to the transmitting station) and a weak secondary signal (via reflection from a metal tower off the line-of-sight). If the path taken by the secondary signal is 1 km longer than the path of the direct signal, what will be the position of the "ghost" image on the screen of the receiver?

Problem 19.2. Motion pictures from film shot at 24 frames/second are transmitted as 60 field/second television images using a technique called *3:2 pull down*. A film frame is held in place while two television fields are transmitted. The subsequent film frame is then pulled down and held in place while three television fields are transmitted. Show that this scheme results in an average film rate of 24 frames/second.

Problem 19.3. Consider a high-definition television monochrome signal with 1920×1080 pixels and progressive scan (no interlace) at 60 frames/sec. If the brightness of each pixel is specified by an 8-bit number, and if no compression is used, what is the data rate? Answer: 995 Mbps. What compression ratio is needed if this signal is to be transmitted over a standard width television channel at a rate of 19.3 Mbits/second?

Problem 19.4. Suppose an NTSC test pattern consists of five vertical bars of equal width but different colors. Let the bars all have the same luminance (brightness) and be only lightly colored (unsaturated). Sketch the waveform for one line of video.

Problem 19.5. Interlaced scanning provides more resolution for fixed scenes but can produce artifacts with moving objects. Think of a situation where interlaced scanning could make a moving object vanish.

Problem 19.6. Shannon's channel capacity theorem states that data, if suitably encoded for error correction, can be sent with an arbitrarily low error rate if the net data rate (bits/sec.) does not exceed the channel capacity, given by $C = B\log_2(1+S/N)$. Compare the ATSC bit rate, 19.3 Mbits/second with the theoretical capacity of the 6-MHz channel at the threshold S/N ratio of 14.9 dB.

References

[1] A/53: ATSC Digital Television Standard, Parts 1–6, 2007 Advanced Television Systems Committee, 1750 K Street, N.W., Suite 1200 Washington, D.C. 20006 (available at www.atsc.org).

[2] Benson, K. B., Editor, *Television Engineering Handbook*, New York: McGraw-Hill, 1986. (Updated version of Fink, D. G., Editor, "*Television Engineering Handbook*," New York: McGraw-Hill, 1957.)

[3] Jackson, K. G. and Townsend, G. B. Editor, *TV & Video Engineer's Reference Book*, London: Butterworth Heineman, 1991. Modern handbook from Britain.

[4] *NAB Engineering Handbook*, 10th edn, 2007, National Association of Broadcasters. National Association of Broadcasters, *Engineering Handbook*, 8th edn., 1993.

[5] Recommended Practice: Guide to the Use of the ATSC Digital Television Standard, including Corrigendum No. 1, Advanced Television Systems Committee, 1750 K Street, N.W., Suite 1200 Washington, D.C. 20006 (available at www.atsc.org).

CHAPTER

20

Antennas and radio wave propagation

While discussing transmitter and receiver circuitry we did not have to know much about antennas or propagation. It sufficed to know only that a voltage applied to the terminals of a transmitting antenna causes a proportional voltage to appear very shortly thereafter at the terminals of a receiving antenna. To be more exact, it was sufficient to know that everything between the terminals of the two antennas is equivalent to a linear two-port network. Here we will consider the transmission through this propagation link.

When an ac source (transmitter) is connected to an antenna (practically any metal structure) the resulting current has a component that is in phase with the applied voltage. The impedance of the antenna therefore has a real part, a resistance, and draws power from the source. If the antenna is efficient, most of the power flows away from the antenna in the form of (energy-bearing) electromagnetic waves and only a small fraction of the power will be dissipated by ohmic heating of the antenna itself. The impedance will also generally have a nonzero imaginary part, a reactance. If the reactance is zero at the operating frequency the antenna is said to be resonant, just as an *RLC* circuit is purely resistive at its resonant frequency. An external tuning network (an *antenna tuner*) can be used to cancel the reactance and also transform the resistance to a value that matches a receiver's input impedance or to a value that draws a desired amount of power from a transmitter.

20.1 Electromagnetic waves

As an electromagnetic wave propagates away from the transmitting antenna, it takes on a spherical wavefront. By the time it reaches a distant receiving antenna, the wavefront has a very large radius of curvature and is essentially a plane wave. The ***E*** and ***H*** vectors (electric and magnetic fields, measured respectively in volts/m and amperes/m) both lie in the plane of the wavefront, i.e., they are transverse to the direction of propagation as shown in Figure 20.1. The fields are in phase with each other; they rise and fall together in space and time.

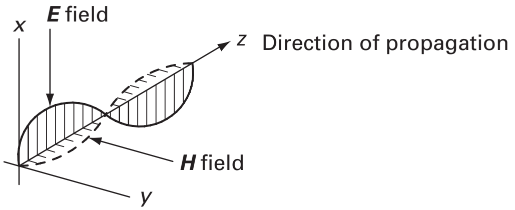

Figure 20.1. An electromagnetic wave – the ***E*** and ***H*** fields are transverse to the direction of propagation.

The ratio of the electric field strength to the magnetic field strength in free space is given by $E/H = (\mu_0/\varepsilon_0)^{1/2} = 120\pi = 377$ ohms. This ratio is known as the "impedance of free space." These fields are perpendicular to each other and their vector cross-product, $\boldsymbol{E} \times \boldsymbol{H}$, points in the direction of propagation. A continuous monochromatic wave has sinusoidal spatial components, $E_0 e^{-jkz}$ and $H_0 e^{-jkz}$, as shown in the figure. The wavenumber, k, is defined as $2\pi/\lambda$ radians per meter. The phase velocity, ω/k, is given by $c = (\mu_0\varepsilon_0)^{-1/2} = 3 \times 10^8$ m/s (the speed of light).

20.1.1 Propagation in a vacuum

Static electric and magnetic fields are always associated with sources, i.e., electric field lines terminate on electric charges and magnetic field lines encircle current filaments. But when an antenna launches an electromagnetic wave, the field lines break away from the sources, reconnecting into closed loops, and the wave becomes autonomous. For a vacuum, Maxwell's curl equations are

$$\nabla \times \boldsymbol{E} = -\mu_0 j\omega \boldsymbol{H}, \tag{20.1a}$$

$$\nabla \times \boldsymbol{H} = \varepsilon_0 j\omega \boldsymbol{E}. \tag{20.1b}$$

The right-hand sides of these equations are like sources; the E-field is produced by the changing H-field and the H-field is produced by the changing E-field. You can verify the statements made above about plane waves by writing the E-field shown in Figure 20.1 as $\boldsymbol{E} = \hat{x}E_0 e^{j(\omega t - kz)}$ and substituting it into Equation (20.1a) to get $\boldsymbol{H}$. You can then verify that $\boldsymbol{H}$ satisfies Equation (20.1b).

Electromagnetic waves are produced by electric charges undergoing acceleration. Time-varying currents contain accelerating charges. A sinusoidally time-varying current distribution on an antenna launches sinusoidal electromagnetic waves. The time-averaged power density or "energy flux" of the waves is given by $S = ½(E \times H) = ½E^2/377$ watts/m^2, where E and H are the (peak) field amplitudes and the factor ½ is the time average of $\cos^2(\omega t)$. At a receiving antenna, the fields from an incident wave produce currents that result in a voltage at the antenna terminals.

20.2 Radiation from a current element

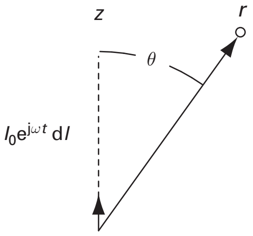

Figure 20.2. An elemental sinusoidal ac current element.

Just as Maxwell's equations can be manipulated to yield back Coulomb's law (which gives the E-field from an elementary point charge) and Ampere's law (which gives the H-field from a constant current element), they can also be stirred together to yield the E and H fields produced by a *time-varying* current element. The E and H fields produced by an antenna are superpositions of the E and H fields from every elemental ac current element on the antenna. Consider an ac current element in the z-direction, as shown in Figure 20.2.

This element has length δl and carries a current $I_0 \cos(\omega t)$ which, as usual, we express as $I_0 e^{j\omega t}$. At the observation point, $\boldsymbol{r}$, the magnetic field is given by

$$\delta H = \frac{I_0 \delta l}{4\pi}\left(\frac{jk\sin(\theta)e^{j(\omega t - kr)}}{r} + \frac{\sin(\theta)e^{j(\omega t - kr)}}{r^2}\right), \tag{20.2}$$

where $k = \omega/c$ and θ is the angle between the current element $\boldsymbol{\delta l}$ and the vector $\boldsymbol{r}$. Equation (20.2) is Ampere's law, generalized for an elemental current with a sinusoidal time variation. The direction of $\boldsymbol{\delta H}$ is given by $\boldsymbol{\delta l} \times \boldsymbol{r}$ (into the page). The first term in the bracket falls off slowly as $1/r$ and is the radiation term. The second term falls off quickly as $1/r^2$ and corresponds to near-field stored energy; j is just a phase factor since $j = e^{j\pi/2}$. The normally implicit $e^{j\omega t}$ term is included here to highlight the wave term, $e^{j\omega t - kr}$. If we let the frequency approach zero, the radiation term vanishes (since k goes to zero), the second term becomes $\sin(\theta)/r^2$ and Equation (20.2) reduces to Ampere's law (also known as the Biot – Savart law) for calculating the magnetic field produced by a dc current element.

Figure 20.3 shows the radiation pattern (*antenna pattern*) obtained from the radiation term in Equation (20.2) (the first term in the bracket). This is a surface plot where the distance from the origin to any point on the surface is proportional to the power radiated in that direction, i.e., proportional to $|\boldsymbol{H}|^2$ at a large constant value of r. We do not have to work out the E-field separately, since we already know that, in the far field, $\boldsymbol{E}$ is perpendicular to both $\boldsymbol{H}$ and $\boldsymbol{k}$ and that its magnitude (in volts/m) is 377 times the magnitude of $\boldsymbol{H}$ (in amps/m).[1]

In this chapter we are interested in radiation far from the antenna, so we can ignore the near-field (second) term in Equation (20.2). At any observation point, the H-field is the integral of the contributions from every current element in any given antenna. For a wire antenna, we would evaluate a line integral. For a reflector antenna, we would have a surface integral, summing up the contributions from the current flowing on each element of the surface.

---

[1] If we do need to know E in the $1/r^2$(near field) region, we can calculate the curl of H from Equation (20.3) and plug that into Equation (20.2) to get E.

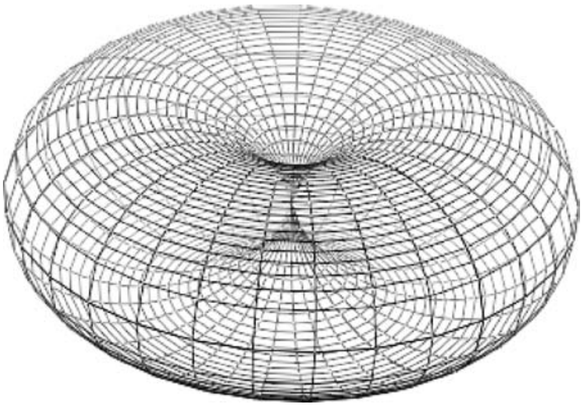

Figure 20.3. Elementary dipole radiation pattern (surface of constant field strength).

20.3 Dipole antenna

A short dipole antenna (shorter than, say, $\lambda/20$) will have the same radiation pattern as the elemental dipole since we can consider it to be a string of elemental dipoles, short enough that the phase paths from every one to the observation point are essentially the same. Thus, the string is equivalent to a single elemental dipole of strength $I\delta\, l = \sum I_i \delta l_i$. We can calculate the power radiated by a single elemental dipole by integrating the average energy flux density, ½EH, over the surface of a bounding sphere to include the power radiated in every direction. Using the radiation term in Equation (20.2) for H, we have

$$\delta H = \frac{I_0 \delta l}{4\pi}\left(\frac{jk\sin(\theta)\mathrm{e}^{\mathrm{j}(\omega t - kr)}}{r}\right), \tag{20.3}$$

$$\begin{aligned}
\text{Total Pwr} &= \int (1/2) EH \mathrm{d}S = \int (1/2)(\mu_0/\epsilon_0)^{1/2} H^2 r^2 \mathrm{d}\boldsymbol{\Omega} \\
&= (1/2)(\mu_0/\epsilon_0)^{1/2}(I_0 k\delta l/(4\pi))^2 \int \frac{\sin^2(\theta)}{r^2} r^2 \mathrm{d}\boldsymbol{\Omega} \\
&= (1/2)(\mu_0/\epsilon_0)^{1/2}(I_0 k\delta l/(4\pi))^2 \int_0^{\pi} 2\pi \sin^3(\theta)\mathrm{d}\theta \qquad (20.4) \\
&= (1/2)120\pi(I_0(2\pi/\lambda)\delta l/(4\pi))^2 8\pi/3 \\
&= 40\pi^2 (I_0 \delta l/\lambda)^2.
\end{aligned}$$

It turns out that the current distribution for a short dipole is a triangle function (actually the almost linear ramp portion of a sine function near zero), as shown in Figure 20.4.

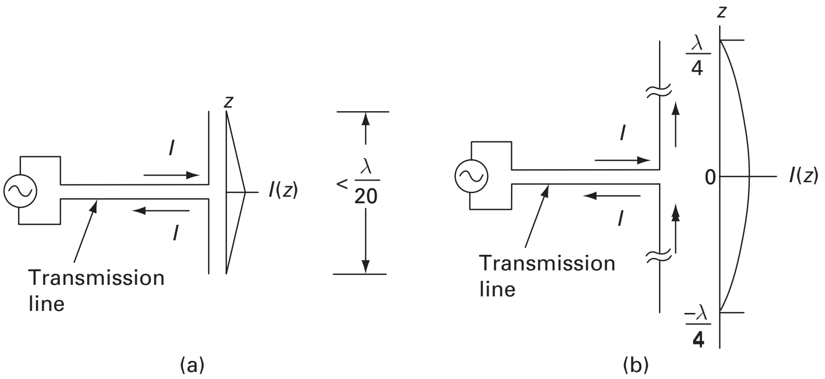

Figure 20.4. (a) A short dipole has a triangular current distribution. (b) A half-wave dipole has a sinusoidal current distribution.

The average current for a short dipole is therefore half the input current, so the power radiated is

$$P_{\text{short dipole}} = 40\pi^2 \left(\frac{I_{\text{in}}}{2} l/\lambda\right)^2, \tag{20.5}$$

where I_{in} is the peak input current and l is the length of the dipole. The input power must be equal to ½ $I_{\text{in}}{}^2 R$, where R is the real part of the antenna's input impedance, known as the *radiation resistance*. For the short dipole, we can therefore write

$$R_{\text{short dipole}} = 2P_{\text{short dipole}}/I_{\text{in}}{}^2 = 20\pi^2(l/\lambda)^2. \tag{20.6}$$

For a short dipole with $l = 0.05\lambda$, $R = 0.49$ ohms. The imaginary part of the input impedance is negative and large, corresponding to a series capacitance with a low value. An antenna tuning network that increases the low value of R and tunes out the high value of capacitive reactance will need at least one high-value inductor. However, the limited Q of practical inductors will introduce losses that result in very low efficiency. A counterpart to the elemental dipole is the elemental current loop or *elemental magnetic dipole* (as opposed to *elemental electric dipole*). The loop is assumed to have constant current and its radiation pattern is the same as the pattern of the elemental dipole. A very small loop antenna has an extremely low radiation resistance and a high positive (inductive) reactance. It is an inefficient antenna in that its ohmic resistance is usually higher than its radiation resistance.

In practice, most dipoles have an overall length of $\lambda/2$. At this resonant length, the impedance of a thin wire dipole seen at the center gap is 73.1 + j0 ohms, an entirely practical value. Resonant loops, with a circumference of the order of a wavelength, are also practical antennas. While transmission line arguments can

Figure 20.5. Fat monopole geometry.

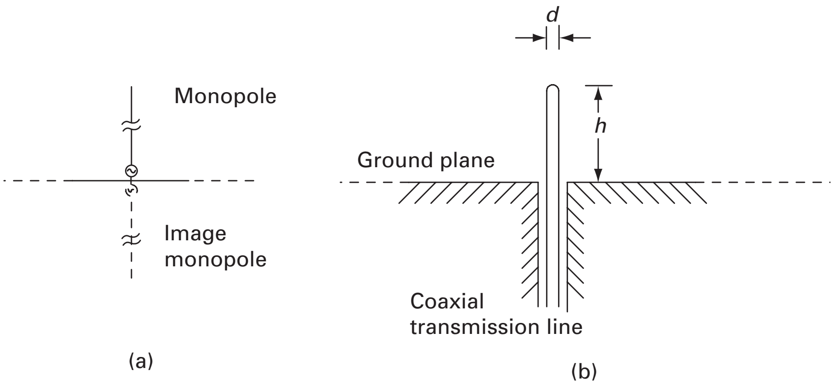

be used to show that the current distribution on a wire dipole is nearly sinusoidal, going to zero at the open ends, rigorous calculations of current distributions on arbitrary antennas require elaborate self-consistent treatments of the currents and fields. A benchmark problem is the fat dipole antenna shown in Figure 20.5(b). This "monopole" is really a dipole, since the ground plane induces an image monopole as shown in Figure 20.5(a).

This geometry is appealingly simple; the radiating element, as it enters the hole in the ground plane, becomes the center conductor of a coaxial line. You can find theoretical and experimental plots of input impedance of this antenna in most standard antenna textbooks. You can also model this dipole if you have access to a program for three-dimensional electromagnetic field simulation, such as Ansoft's HFSS.

20.4 Antenna directivity and gain

The radiation from any antenna is always stronger in some directions than in others; no antenna can be an isotropic radiator.[2] Let the energy flux (W/m^2) produced in a given direction be denoted as $S(\theta,\phi,r)$, where we are using standard polar coordinates. As before, the integral of S over an enclosing sphere will be the total radiated power transmitted, P_{RAD}:

$$\int S(\theta,\varphi,r)r^2\,\mathrm{d}\boldsymbol{\Omega} = P_{\mathrm{RAD}}. \tag{20.7}$$

---

[2] It is topologically impossible to have a constant tangential vector field on the surface of a sphere, a theorem sometimes expressed as "You cannot comb the hair smooth on a billiard ball." Thus, in the far field of an antenna (at a distance greater than the square of the largest physical dimension of the antenna divided by the wavelength), where the E and B fields are transverse (tangent to the surface of a sphere surrounding the antenna), the fields cannot be everywhere constant. The radiation pattern of an antenna, therefore, cannot be isotropic.

For any given direction, the ratio of the flux to the average flux is defined as the *directivity* of the antenna

$$D(\theta, \Phi, r) = \frac{S(\theta, \Phi, r)}{P_{\mathrm{RAD}}/(4\pi r^2)}. \tag{20.8}$$

An equivalent statement is that the directivity is the factor by which the flux in the strongest direction exceeds the flux from a hypothetical isotropic antenna radiating the same total power. We know that the radiation flux eventually falls off as r^{-2}. For large r, then, the far-field directivity, D, is a function only of the polar coordinates θ and ϕ. Combining Equations (20.7) and (20.8), we see that the average directivity of an antenna is unity:

$$\frac{1}{4\pi}\int D(\theta, \Phi)\mathrm{d}\boldsymbol{\Omega} = 1. \tag{20.9}$$

Directive gain, $G(\theta,\phi,\mathrm{r})$, has the same definition as the directivity except that the radiated power is replaced by P_{inc}, the power incident on the antenna terminals:

$$G(\theta, \Phi, r) = \frac{S(\theta, \Phi, r)}{P_{\mathrm{inc}}/(4\pi r^2)}. \tag{20.10}$$

If an antenna has no ohmic losses and its feedpoint impedance matches the transmission line impedance (so that no power is reflected), all the incident power will be radiated and the directive gain will be equal to the directivity. In most antennas used for transmitting, the losses are no more than a few percent and the distinction between directivity and gain is unimportant.[3] The maximum value of an antenna's directive gain is simply called the gain. A transmitter connected to an antenna having a gain of 20 dB will produce a directed signal 100 times more powerful than if it were connected to a hypothetical lossless isotropic radiator.

Since we have already calculated the total power radiated by an elemental dipole (Equation 20.4), we can easily calculate its directivity. The maximum flux density, calculated from Equation (20.3), is $S_{\max} = (60/\pi)[\,I_0 k\,\delta\,l\,/(4\pi r)]^2$, where we have substituted 120π for $(\mu_0/\varepsilon_0)^{1/2}$. Dividing $\mathrm{S}_{\max}$ by the average flux density, S_{avg} = total power/$(4\pi r^2)$, yields a directivity of 1.5. As explained above, this will also be the directivity of a short dipole. The directivity of a half-wave dipole is 1.64 and its radiation pattern differs little from that of the elemental dipole.

---

[3] Small inefficient antennas are adequate for low-frequency receivers. Even though antenna losses add noise at the receiver input, signal strengths in the AM broadcast band and short-wave bands must already be considerably higher than this added noise in order to exceed the noise power from static (atmospheric electricity). Even with an inefficient antenna, the total power delivered to the receiver is much greater than the thermal noise added by the antenna and by the receiver itself.

20.5 Effective capture area of an antenna

The distance between transmitting and receiving antennas is generally so large that a plane wave can be assumed incident on the receiving antenna. The energy extracted from the incident wave is, of course, proportional to the average energy flux density, $\frac{1}{2}E \times H$ W/m^2. The proportionality constant is called the *effective area* (capture area) of the receiving antenna. It turns out, as shown below, that the effective area is proportional to gain:

$$A_{\text{eff.}} = G\lambda^2/4\pi. \tag{20.11}$$

Since gain and effective area are proportional, there is really no distinction between transmitting antennas and receiving antennas; the best transmitting antenna (most gain) is also the best receiving antenna (most capture area). A standard derivation of Equation (20.11) begins by applying the reciprocity theorem[4] to a system of two arbitrary antennas. The two antennas, as shown in Figure 20.6, need not be identical. We can suppose they are both matched to the same impedance value and that we have both a generator and a receiver that match this impedance. First we connect the generator to Antenna 1 and measure the power from Antenna 2.

If we now interchange the generator and load, the reciprocity theorem states that the power delivered to the load will be unchanged. Expressing this in terms of gain and effective area, we have $G_1A_{\text{eff }2} = G_2A_{\text{eff }1}$ or $G_1/A_{\text{eff }1} = G_2/A_{\text{eff }2}$, from which we see that the ratio of gain to effective area has the same constant value for any and all antennas. We can pick any conveniently simple antenna

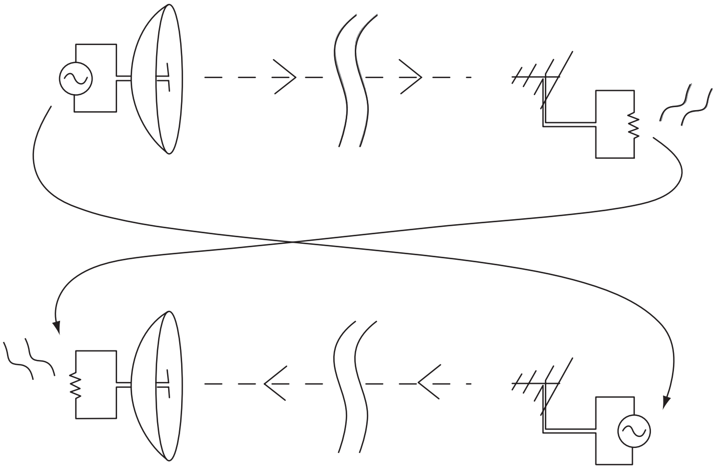

Figure 20.6. Reciprocity: received power is unchanged when source and load are interchanged.

---

[4] The remarkable reciprocity theorem finds application in mechanics, optics, and acoustics as well as electrical engineering.

and use electromagnetic theory to calculate its gain and effective area. When this is done, it is found that the ratio of gain to effective area is $4\pi/\lambda^2$. A half-wave dipole antenna has a maximum gain of 1.64. Its effective area is therefore $1.64\,\lambda^2/4\pi$. If the dipole is made of thin wire, it has no real physical area, only a length, yet it has a nonzero effective area and can extract energy from an incident electromagnetic wave. Note that even a very short dipole has nearly the same effective area as a half-wave dipole. The effective area of a microwave dish antenna can approach its physical area, as we will see below, but normally it is from 50 to 80 percent of the physical area. This fraction, known as *aperture efficiency*, is usually set not by ohmic losses but rather by the illumination pattern of the primary feed antenna. A perfectly uniform illumination pattern (and no ohmic loss) produces an aperture efficiency of unity. The Arecibo radio telescope dish uses an aperture 700 ft in diameter (area = 35 800 m^2). Its aperture efficiency is about 70% at $\lambda = 12$ cm (the wavelength used there for radar astronomy) so its gain, using Equation (20.11), is $G = 4\pi(0.70 \times 35\,800)/(0.12)^2 = 22 \times 10^6$ or 73 dB.

20.6 Reflector and horn antennas

Radar antennas and satellite TV dishes are familiar examples of reflector antennas. From a transmitting standpoint, a primary or "feed" antenna illuminates the reflector. Currents induced in the metallic surface of the reflector become secondary radiators and their radiation forms the beam of the antenna. The larger the dish, the more directive the antenna, as we will see. These antennas are examples of "aperture antennas;" one can readily identify the aperture (usually circular) from which the radiation emanates, as if from a searchlight.

Let us use Equation (20.3) to calculate the radiation from a large flat rectangular metal plate on which there is a surface current density, J_S, which has the same amplitude and phase at every point.[5] For surface current, $J_S\mathrm{d}x\mathrm{d}y$ replaces $I_0\mathrm{d}l$ in Equation (20.3). To find the antenna pattern we integrate over the plate, summing up the contributions from each element of area. These are phasor contributions and the far field is actually an interference pattern. Once we have calculated the antenna pattern, we will integrate the power over a bounding sphere, as we did for the elemental dipole, to determine the total power radiated. With this, we can then calculate the gain.

Figure 20.7 shows the geometry for this antenna. The size of the plate is $2a$ by $2b$. Radiation will be strongest in the z-direction, perpendicular to the plate. We

---

[5] Such a current distribution could be established by illuminating the plate with radiation from a dipole far out in front of the plate, though this would not be an efficient feed antenna, since the plate would intercept only a small fraction of the dipole's radiation.

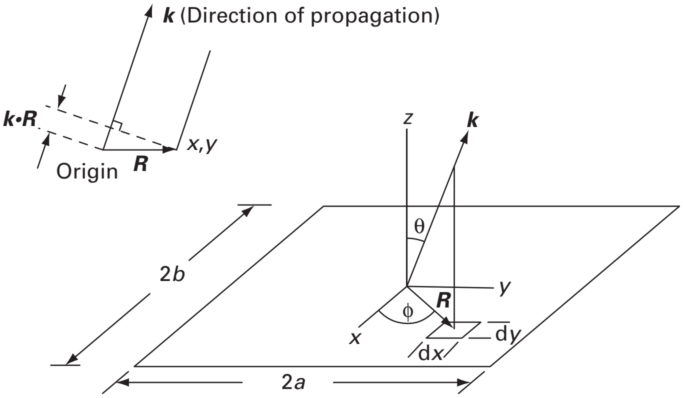

Figure 20.7. Rectangular aperture antenna geometry.

will assume that both a and b are much greater than λ, anticipating that this will form a very concentrated beam.[6]

The vector from the center of the plate to a distant observation point is denoted by $\boldsymbol{r}$, and the vector from the center of the aperture to an arbitrary point x,y on the aperture is denoted by $\boldsymbol{R}$. Again using the first term (the radiation term) of Equation (20.3), we see that the contribution to the distant H field from an element of area at the center of the plate is given by $\delta H = J_S\,\mathrm{d}x\mathrm{d}y\ \mathrm{j}k(4\pi r)^{-1}\cos(\theta)\mathrm{e}^{\mathrm{j}(\omega t-kr)}$ where, as always, $k = 2\pi/\lambda$ and $\boldsymbol{k}$ is directed from the plate to the point $\boldsymbol{r}$. Note that the $\sin(\theta)$ factor in Equation (20.3) is here $\cos(\theta)$ because the current is perpendicular, not parallel, to the z-axis. We will replace this $\cos(\theta)$ by unity, anticipating that, when both a and b are much greater than λ, the pattern will be highly concentrated around the $\theta = 0$ direction. Away from the origin, radiation to the point $\boldsymbol{r}$ from any other point on the plate will be shifted in phase by $\boldsymbol{k}\cdot\boldsymbol{R}$ radians, as shown by the inset in Figure 20.7. Using standard polar coordinates, this phase shift is given by

$$\boldsymbol{k}\cdot\boldsymbol{R} = xk_2 + yk_y = k\ \sin(\theta)[x\cos(\phi) + y\sin(\phi)]. \tag{20.12}$$

If we assume a uniform field over the aperture, the field at r, θ, ϕ can now be written as

$$\begin{aligned} H(r,\theta,\phi) &= \frac{jkJ_S\mathrm{e}^{\mathrm{j}\omega(t-kr)}}{4\pi r}\int_{-b}^{b}\int_{-a}^{a}\mathrm{e}^{\,\mathrm{j}k\cdot R}\mathrm{d}x\mathrm{d}y \\ &= \frac{jkJ_S\mathrm{e}^{j\omega(t-kr)}}{4\pi r}\int_{-b}^{b}\int_{-a}^{a}\mathrm{e}^{\,\mathrm{j}k\theta(x\cos\phi+y\sin\phi)}\mathrm{d}x\mathrm{d}y, \end{aligned} \tag{20.13}$$

---

[6] When used at a wavelength of 12 cm, the 210-m diameter illumination on the dish at Arecibo forms a beam whose width between half-power points is only 12/21 000 radians or 0.03 degree, a "pencil beam."

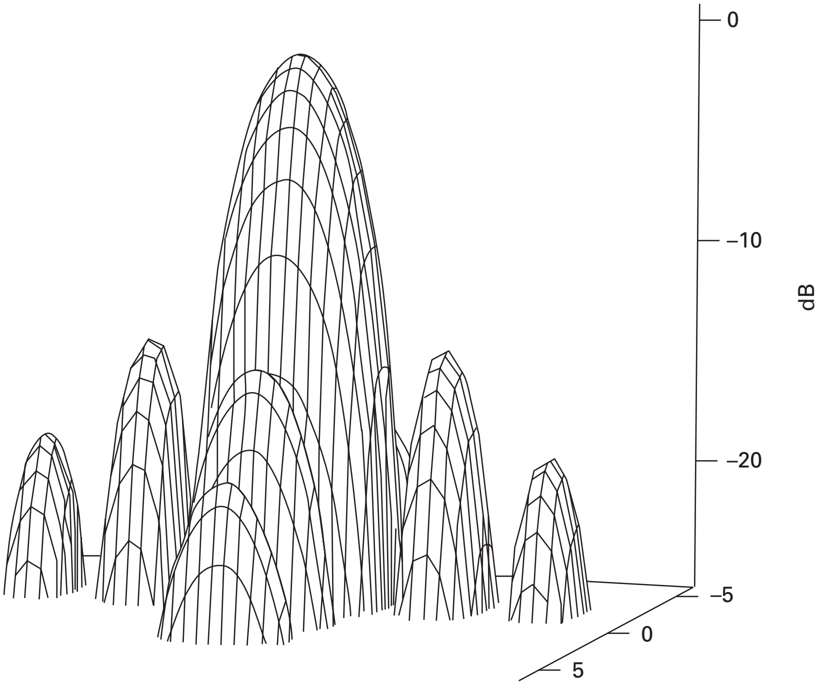

Figure 20.8. Radiation pattern (power in dB) from a rectangular aperture.

where we have replaced $\sin(\theta)$ by θ, again anticipating that the power will be negligible except for the region close to $\theta = 0$. Evaluating the integral we find

$$H(r,\theta,\phi) = \frac{\mathrm{j}kJ_{\mathrm{S}}ab\mathrm{e}^{\mathrm{j}\omega(t-kr)}}{4\pi r}\left(\frac{\sin(ka\theta\cos(\phi))}{ka\theta\cos(\phi)}\frac{\sin(kb\theta\sin(\phi))}{kb\theta\sin(\phi)}\right). \tag{20.14}$$

Since the average power density (watts/m^2) is given by $EH = ½(\mu_0/\varepsilon_0)^{1/2}H^2$, the antenna pattern is proportional to the square of the term in brackets. Figure 20.8 shows the pattern, which has a strong central *main lobe* surrounded by *sidelobes*, which for this $(\sin(x)/x)^2$ power pattern are 13.3 dB below the peak of the main beam.

This plot uses rectangular "sky coordinates," u and v, measured in radians, where $u = \theta\sin(\phi)$ and $v = \theta\sin(\phi)$. The term in brackets is, therefore, $\sin(kau)/(kau)\ \sin(kbv)/(kbv)$. This plot was made for a square aperture; a rectangular aperture produces an elongated beam. For $a < b$, the beam is broader in the u-direction than in the v-direction. We can integrate the power density of the pattern to get the total average power:

$$\begin{aligned}\mathrm{Pwr} &= 1/2\sqrt{\frac{\mu_0}{\varepsilon_0}}\left|\frac{jJ_{\mathrm{S}}kab\mathrm{e}^{\mathrm{j}\omega(t-kr)}}{4\pi r}\right|^2\iint r^2\left(\frac{\sin(kau)}{kau}\frac{\sin(kbv)}{kbv}\right)^2 \mathrm{d}u\mathrm{d}v\\ &= 1/2\sqrt{\frac{\mu_0}{\varepsilon_0}}\left(\frac{J_{\mathrm{s}}kab}{4\pi}\right)^2\frac{\pi^2}{k^2ab}.\end{aligned} \tag{20.15}$$

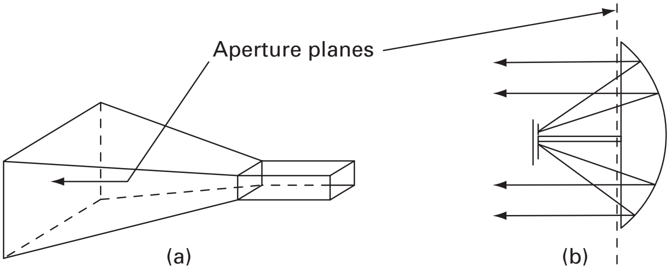

Figure 20.9. Horn and dish antenna aperture planes.

This let us calculate the gain:

$$G = 1/2\sqrt{\frac{\mu_0}{\epsilon_0}}|H(r,0,0)|^2 \div \frac{\text{Pwr}}{4\pi r^2} = \frac{4\pi(4ab)}{\lambda^2} = \frac{4\pi A}{\lambda^2}. \tag{20.16}$$

This example antenna serves to establish the relationship between gain and effective area, $G = 4\pi\ A_{\text{eff}}/\lambda^2$, as long as we can argue that the effective area (capture area) is equal to the actual physical area. To argue this case, consider that a wave transmitted by this antenna is essentially a plane wave as it leaves the large aperture (as contrasted with the spherical wave emitted by a small antenna). Therefore, if we invoke time reversal, the wave entering the antenna would be a plane wave, just as if it had come from a distant transmitting antenna. Since the time-reversed wave is completely "absorbed" by the antenna, the effective area is equal to the physical area, $4ab$.

The integral in Equation (20.13) is just a two-dimensional Fourier transform of the aperture illumination. In this example, we used uniform illumination, so we have transformed the shape of the aperture with uniform weighting. The aperture can have any shape. For example, a circular plate of radius $a \gg \lambda$ produces a field pattern proportional to $J_1(ka\theta)/(ka\theta)$, where J_1 is the first-order Bessel function of the first kind. (The function $J_1(x)$ resembles $\sin(x)[x+1]^{-1/2}$). For any given aperture, uniform illumination produces the highest gain. But "tapered" illumination, where the current density is less near the edges of the aperture, is often used because it reduces the amplitude of the sidelobes.

A simple *microwave horn antenna* is essentially a waveguide funnel, fed at the small end, as shown in Figure 20.9. The other end fans out, usually in both transverse directions, to form a large aperture. The beam radiated from the aperture is comparable to the beam radiated from a dish antenna having equal area. The field distribution in the aperture plane is the same as the field at any cross-section in the waveguide, expanded to fill the aperture.

The horn antenna has no secondary radiators in its aperture plane. Of course there are currents on the inside walls of the waveguide and, in principle, Equation (20.3) can be used to calculate the radiation pattern far out in front of the horn. A much simpler way to find the radiation pattern is to use Huygen's principle, where the aperture *field* is regarded as a *source*, a two-dimensional array of wavelets, each of which re-radiates as if it were a current element.[7] If you know the *E*-field (or *H*-field) in the aperture plane, the squared magnitude of its two-dimensional Fourier transform yields the radiation pattern, just as the squared magnitude of the transform of the currents on a metal plate yielded the radiation pattern in the above example. This analysis technique is also applied to dish antennas. Ray-tracing methods (geometric optics) are used to find the phase and amplitude of the field at an aperture plane – usually just in front of the dish. This field is then regarded as a Huygen's source; its Fourier transform gives the far-field pattern without requiring an intermediate calculation of the currents on the reflector surface. This Fourier transform method is also applied to systems having multiple antennas, such as arrays of dish antennas, used together as an *interferometer* to form an extremely narrow beam.

20.7 Polarization

The elementary dipole of Figure 20.2 produces radiation with *linear polarization*; at any observation point the electric field vector has a fixed direction and an amplitude that oscillates sinusoidally. A dipole receiving antenna would have to be placed parallel to the incoming polarization. If placed perpendicular to the incoming *E* vector, it would pick up no signal. In all the preceding discussions, we have implicitly assumed that the polarization of a receiving antenna was matched to that of a transmitting antenna. *Circular polarization* can be produced with an antenna made of crossed dipoles. The dipoles are fed by signals that are equal in amplitude but 90° out of phase. This results in a transmitted *E* vector that rotates in the plane perpendicular to the direction of propagation. At any observation point, the magnitude of the *E* vector remains constant, but the vector rotates one turn for every period of the wave. A 90° hybrid can be used as a power splitter to feed the two dipoles. An identical antenna (a pair of crossed dipoles with a 90° hybrid as a combiner) makes a polarization-matched receiving antenna. Note that a single dipole can serve as a receiving antenna for circular polarization, but it collects only half the available power. If the incoming *E* field, viewed from the receiving position, rotates in the counterclockwise sense, the radiation has, by definition, *right-hand circular polarization (RCP)*. If the cables from the receiving antennas were to be interchanged at the hybrid, the resulting *LCP* antenna would have the wrong handedness and could receive no power from an RCP transmitting antenna. The most general polarization is elliptical, which is an admixture of linear and circular. Linear and circular polarizations are just special cases of elliptical polarization.

---

[7] See Reference [1] for a theoretical justification of Huygen's principle.

20.8 A spacecraft radio link

Consider the following example of a spacecraft telemetry link for which we wish to find the maximum range. Suppose we have a 1-W telemetry transmitter aboard a spacecraft and that the data rate requires a channel capacity corresponding to a signal-to-noise ratio of at least unity in a 1-Hz bandwidth. This link uses a frequency of 3 GHz (10 cm wavelength). The transmitting antenna on the spacecraft is a 2-m diameter dish. The ground station antenna is a 10-m diameter dish, as shown in Figure 20.10. Assume that both these dish antennas have an effective area equal to 60% of their physical apertures. Assume also that there is no pointing error, i.e., the antennas always point directly at each other and that the system temperature of the ground station receiver is 25 K. (The system temperature is the sum of the equivalent receiver noise temperature, the antenna noise temperature, and the sky noise temperature.)

1. What is the equivalent input noise power of the receiver? Boltzmann's constant, k, is 1.38×10^{-23} W/Hz/K, so the equivalent input noise power, kTB, is $1.38 \times 10^{-23} \times 25 \times 1 = 3.45 \times 10^{-22}$ W.
2. What is the effective area of the receiving antenna? The physical aperture is $\pi \mathrm{R}^2$ so the effective area is $0.60\ \pi R^2 = 0.60\pi 5^2 = 47.1\ \mathrm{m}^2$.
3. What is the gain of the transmitting antenna? The physical aperture is πR^2 and the aperture efficiency is 0.60 so the effective aperture is $0.60\ \pi(1^2) = 1.88\ \mathrm{m}^2$. The gain is $4\pi A_{\mathrm{eff}}/\lambda^2 = 4\pi(1.88)/(0.1^2) = 2369$.
4. What is the maximum range, R, in kilometers for the spacecraft to maintain the required signal-to-noise ratio? Here we simply set $P_{\mathrm{noise}} = P_{\mathrm{received}} = A_{\mathrm{rcvr}}\ P_{\mathrm{trans.}} G_{\mathrm{trans}} / (4\pi R^2)$ from which we have $R^2 = (4\pi)^{-1} P_{\mathrm{trans.}}\ G_{\mathrm{trans}}/P_{\mathrm{noise}}$. Using the parameters calculated above, we have $R = \sqrt{(4\pi)^{-1} \times 1 \times 2369/(3.45 \times 10^{-22})} = 7.4 \times 10^{11}\mathrm{m} = 740 \times 10^6\mathrm{km}$. This is roughly the mean distance to Jupiter.

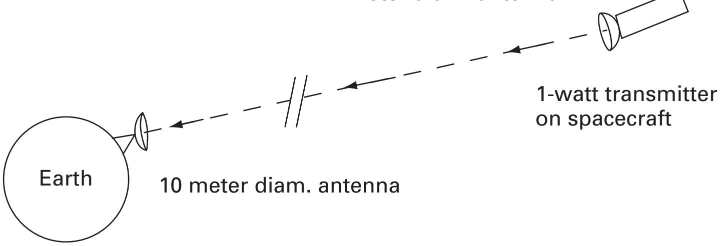

Figure 20.10. A spacecraft telemetry link.

20.9 Terrestrial radio links

VHF and UHF two-way radios used by cellular phones, emergency vehicles, etc. have transmitters with several watts of power but their range is limited by the curvature of the Earth to only a few miles or tens of miles. (Radio waves do not propagate through the highly conductive Earth though they do diffract slightly, so the radio horizon is somewhat beyond the optical horizon.) Mobile radios also use antennas that have gains only of the order of unity so they do not have to be pointed accurately – or even at all. Finally, in ground-to-ground radio links, the signal usually arrives from an angle near the horizon; the receiving antenna will pick up noise from the ground (thermal radiation). In these systems, then, extremely low-noise receivers are of no benefit. Broadcasting stations for FM and television also use VHF and UHF frequencies so their range is also essentially line-of-sight. Long-distance propagation in AM and short-wave broadcasting depends on reflection from the ionosphere.

20.10 The ionosphere

At altitudes above about 60 km the atmosphere is ionized by ultraviolet radiation from the Sun; electrons are stripped from the neutral particles (mostly oxygen atoms and O_2 and N_2 molecules) to produce a mixed electron and ion gas. During the day the density of this ionized gas is highest at around 250 km, the peak of the "F-region." Above the peak the ionization is less because the thinner atmosphere presents fewer particles to be ionized. Below the peak the ionization is less because the denser atmosphere exhausts the supply of ultraviolet photons; the electrons they produce quickly encounter nearby ions and recombine. At night, without sunlight, the ionization rate is zero. Recombination quickly neutralizes the ionization at the lowest altitudes, around 100 km, and depletes the F-region until sunrise.

20.10.1 Wave propagation in the ionosphere

An electromagnetic wave induces electric currents in the electron gas of the ionosphere. (The electrons, by virtue of their low mass, are accelerated by the incident wave to much higher velocities than the ions, so the ion contribution to the current is negligible.) The effective dielectric constant of an electron gas is not difficult to calculate from Maxwell's equations. (We will see that this dielectric constant becomes imaginary below a certain critical frequency which depends on the electron density; below this frequency, then, electromagnetic waves cannot propagate through the plasma.) No longer in a vacuum, we must use the general curl H equation which includes the real electric current, J, in addition to the displacement current:

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \epsilon_0 \frac{\partial \boldsymbol{E}}{\partial t}. \tag{20.17}$$

The electrons are in rapid thermal motion, but this motion is random and contributes nothing to the current. Yet all the electrons near any given point are accelerated equally and move together to produce a net current, $\boldsymbol{J} = Nev$, where N is the electron density, e is the electron charge, and v is the component of velocity imparted by the electric field. We will neglect the weak $\boldsymbol{v} \times \boldsymbol{B}$ force from the magnetic field so Newton's second law of motion is just $m\mathrm{d}\boldsymbol{v}/\mathrm{d}t = \boldsymbol{F} = e\mathbf{E}$. For a sinusoidal time dependence, $\mathrm{e}^{\mathrm{j}\omega t}$, we can write this equation of motion as $m(\mathrm{j}\omega\boldsymbol{v}) = e\mathbf{E}$. Substituting $\boldsymbol{J} = Ne\mathbf{v} = Ne^2\boldsymbol{E}/(\mathrm{j}\omega m)$ in Equation (20.17) gives

$$\nabla \times \boldsymbol{H} = \frac{Ne^2\boldsymbol{E}}{\mathrm{j}\omega m} + j\omega\epsilon_0\boldsymbol{E} = \mathrm{j}\omega\epsilon_0\left(1 - \frac{Ne^2}{\epsilon_0 m\omega^2}\right)\boldsymbol{E}. \tag{20.18}$$

Note that the term in brackets is the relative dielectric constant and that it becomes negative for low frequencies, in particular for $\omega^2 < \omega_\mathrm{p}^2$ where $\omega_\mathrm{p}^2 = Ne^2/(\varepsilon_0 m)$. This happens because the conduction current (the electron current) becomes greater than the displacement current. The total current (conduction current plus displacement current) changes sign and has the wrong polarity to source the H field of a traveling wave. This critical frequency, ω_p, is known as the *plasma frequency*. (If the local charge neutrality of an electron–ion gas is disturbed, the densities will oscillate at this frequency the way a spring and mass system oscillates at its resonant frequency.) For a wave to propagate in the plasma, the dielectric constant must be positive; only waves with frequencies lower than the plasma frequency will be reflected. The free-electron gas that gives metals their conductivity is dense enough to reflect visible light but the alkali metals (lithium, sodium, etc.) have relatively lower electron densities and are transparent in the ultraviolet.

20.10.2 Reflection of waves from the ionosphere

The reflection of radio waves is normally a process of refraction because the waves are not vertically incident on the ionosphere. As they travel obliquely upward, the dielectric constant decreases so the phase velocity increases, causing the propagation vector, which is perpendicular to the wavefront, to turn around gradually. This is shown in Figure 20.11.

For signal strength calculations over an ionospheric path the ionosphere can be considered a specular mirror and field strengths can be calculated by taking the inverse square of the total path length of the ray.

20.10.3 Daytime vs. nighttime propagation

Short-wave broadcasts from distant transmitters on the higher frequency bands are not heard at night because the ionospheric electron density is too low for

Figure 20.11. Ionospheric refraction.

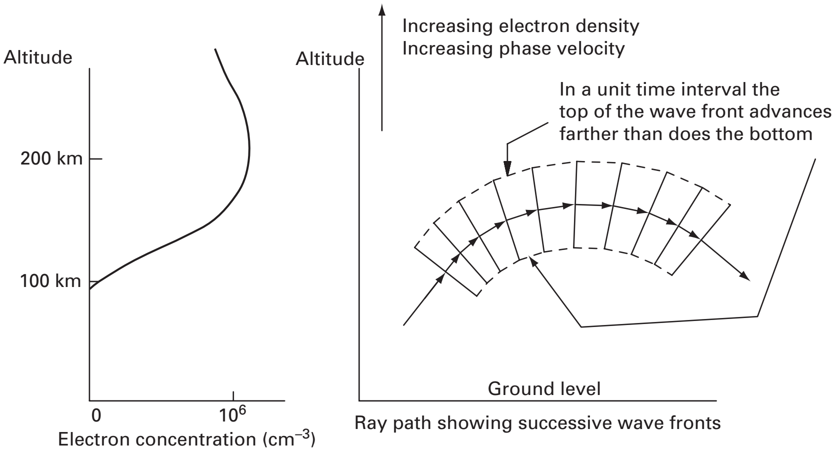
Altitude
200 km
100 km
0
10^6
Electron concentration (cm^{-3})
Altitude
Increasing electron density
Increasing phase velocity
In a unit time interval the top of the wave front advances farther than does the bottom
Ground level
Ray path showing successive wave fronts

reflection, i.e., the waves cannot get turned around sufficiently. On the other hand, low-frequency stations, such as those in the AM broadcasting band, are received from great distances *only* at night. During the day their energy is dissipated in the lower ionosphere through collisions between the accelerated electrons and neutral atoms and molecules. Why don't higher frequency waves suffer this daytime attenuation in the lower ionosphere? Consider a low-frequency and a high-frequency wave of equal power, that is, equal field strengths. Both cause the ionospheric electrons to execute synchronized sinusoidal motion as described above. But the low-frequency wave, because its period is long, will produce higher electron velocities; we saw above that $v = Ee/(j\omega m)$. The average kinetic energy of each electron is therefore $m\langle v^2\rangle/2 = m[Ee/(\omega m)]^2/4$ where E is the amplitude of the electric field. All the electrons in the lower ionosphere suffer collisions with the neutral particles (which are present since the ionization is not 100%). The average collision leaves the electron with a random velocity, i.e., its share of the synchronized sinusoidal motion is lost. The frequency of collisions does not depend on the frequency of the electromagnetic wave (or on there being any electromagnetic wave present at all) so the rate of energy loss is inversely proportional to the square of ω, the wave frequency, and long-distance AM listeners have to wait for nighttime.

20.11 Other modes of propagation

Besides reflection by the ionosphere, there are a number of other ways that an electromagnetic wave can get around the curvature of the Earth. These include scattering from the ionized trails of meteors entering the Earth's atmosphere, scattering from irregularities in the ionosphere even when the ionosphere is otherwise not dense enough to turn the waves around by refraction, scattering from density irregularities in the neutral atmosphere (i.e., fluctuations in the index of propagation), and ducting beneath atmospheric temperature inversion layers.

Problems

Problem 20.1. The voltage at the terminals of a receiving antenna is proportional to the E field of an incident electromagnetic wave. The “effective length” (or effective height if the antenna is vertical) is defined as the open-circuit voltage at the terminals divided by the incident E field: volts / (volts/meter) = meters. Show that the effective length is given by

$$\text{effective length} = (4RA_{\text{eff}}/Z_0)^{1/2},$$

where R, the radiation resistance, is the real part of the antenna impedance, A_{eff} is the effective area ($A_{\text{eff}} = G\cdot\lambda^2/(4\pi)$), and $Z_0 = 377$ ohms = $(\mu_0/\epsilon_0)^{1/2}$, the impedance of free space. Find the effective length of a half-wave dipole ($G = 1.6$ and $R = 73$ ohms).

Problem 20.2. Suppose we have a 1-W transmitter connected to a dipole antenna which is aligned to provide the maximum signal strength at a distant receiving antenna. Needing more signal strength, we obtain a second, identical dipole and, using a power splitter, feed each dipole with ½W. We space the second far enough from each other so that they do not interact. We make sure that both antennas are aligned toward the receiver and we also make sure that the cables from the power splitter have equal length. At the receiving antenna, each transmitting antenna provides a field amplitude that is less than the original field by $1/\sqrt{2}$. But the two signals are in phase so the total amplitude is increased by $2/\sqrt{2} = \sqrt{2}$. Squaring this we see that the received signal strength is doubled. Have we gotten something for nothing? Could we repeat this process to increase the received power even more?

Problem 20.3. Let the individual antennas of Problem 20.2 be AM broadcast towers with omnidirectional patterns and vertical polarization. Suppose the spacing between these antennas is $\lambda/2$. As before, they are fed symmetrically, that is, with the same power and same the phase. Find the radiation pattern in the horizontal plane: make a polar plot of the relative field strength vs. azimuth angle for a distance far from the antennas. Hint: at any observation point in the horizontal plane at a distance r from the center of the line joining the two antennas, the total voltage is the sum of the contributions from the two antennas, $e^{j\phi_1}$ and $e^{j\phi_2}$. The phases π_1 and π_2 are the phase path lengths corresponding to r_1 and r_2, the distances from the observation point to the respective antennas. These phase paths are just $2\pi r_1/\lambda$ and $2\pi r_2/\lambda$. The field strength is given by $(|e^{j\phi_1} + e^{j\phi_2}|)$.

Problem 20.4. Consider a pair of crossed dipoles. The first dipole points in the z-direction and carries a current $\cos(\omega t)$. The second dipole points in the x-direction and carries a current $\sin(\omega t)$. Find the type of polarization and the relative power density of the radiation in the $+z$, $-z$, $+x$, $-x$, $+y$, and $-y$ directions. If you have a program like MATLAB or Mathcad, make a three-dimensional surface plot, like Figure 20.3, for which the distance from any point on the surface to the origin is proportional to the radiation power in that direction.

References

[1] Collin, R. E., *Antennas and Radiowave Propagation*, New York: McGraw-Hill, 1985.

[2] Davies, K., *Ionospheric Radio Propagation*, New York: Dover Publications, 1966.

[3] Kelley, M. C., *The Earth's Ionosphere: Plasma Physics and Electrodynamics*, New York: Academic Press, 1989.

CHAPTER

21 Radar

The idea that radio waves could be used to detect the presence of stationary or moving objects emerged around 1900, almost as soon as radio itself. Christian Hueslmeyer, a German inventor, demonstrated an apparatus in 1904 which, when mounted on a bridge above the Rhine, rang a bell when a ship passed beneath. He used a (now) primitive spark gap RF source and coherer detector. The system may have shown only marginal potential for collision avoidance, as the German Navy demonstrated no interest. Sir Robert Watson-Watt developed meteorological radar in Britain in the 1930s and then a chain of air defense radars during World War II. In the U.S., the MIT Radiation Laboratory, set up to develop military microwave radar systems, had nearly 4000 employees between 1940 and 1945. The acronym RADAR, for Radio Detection And Ranging, has been attributed to U.S. Navy officers F. R. Furth and S. M. Tucker, who introduced it in 1940, though the term remained classified throughout the war.[1]

Today, radar goes beyond aircraft tracking to applications as diverse as space object monitoring, storm tracking, detection of clear air turbulence, and velocity measurements of speeding automobiles and baseballs. In this chapter, we look at some commonly used radars, some general system aspects of radar, and, finally, some RF components and techniques developed specifically for radar.

21.1 Some representative radar systems

Classic surveillance radar

Figure 21.1 is a block diagram of the classic radar system used to monitor air traffic. A rotating antenna sweeps continuously in azimuth while it repeatedly transmits short pulses and receives subsequent echos from targets. As each pulse is transmitted, the beam of the CRT display begins to sweep from the center in a radial direction. The detected signal from the receiver modulates the

---

[1] Butrica, A. J. *To See the Unseen: A History of Planetary Radar Astronomy*, Diane Publishing Co. 1997 (also available on the Web).

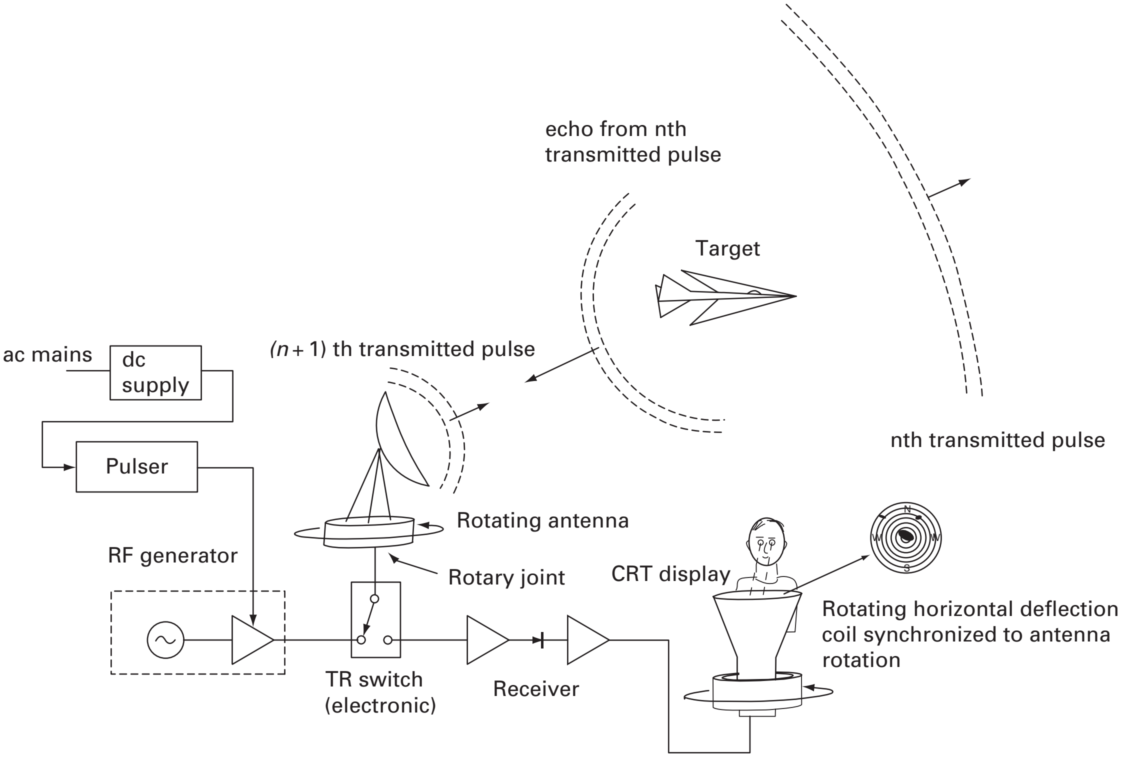
echo from nth
transmitted pulse
Target
ac mains
dc supply
(n + 1) th transmitted pulse
Pulser
nth transmitted pulse
Rotating antenna
RF generator
Rotary joint
CRT display
Rotating horizontal deflection coil synchronized to antenna rotation
TR switch (electronic)
Receiver
N
S
W

Figure 21.1. Classic surveillance radar.

intensity of the spot, rather than causing a deflection perpendicular to the sweep direction. The beam deflection coil is made to rotate continuously around the neck of the CRT, synchronized to the rotation of the antenna, so that the radial sweep direction at any instant is the azimuthal pointing direction of the antenna.

Generally there is a bright area near the center of the CRT, produced by reflections from nearby hills, buildings and towers. The receiver may incorporate an electronic gain control, slaved to the CRT sweep to compensate for the decrease in echo strength with increasing target range. Dashed lines on the figure show the spherical wavefronts for two transmitted pulses and for the reflection (echo) from one pulse, but do not indicate that the intensity of the transmitted waves is sharply concentrated in the forward direction by the high directivity of the antenna.

21.1.1 CW Doppler radar

The classic radar described above uses short pulses for good range resolution. It detects the amplitude of each returning pulse, paying no attention to the phase. Quite the opposite is the radar speed gun, which transmits cw and therefore has no range resolution, but does detect the phase to measure speed. A functional block diagram of the speed gun is shown in Figure 21.2.

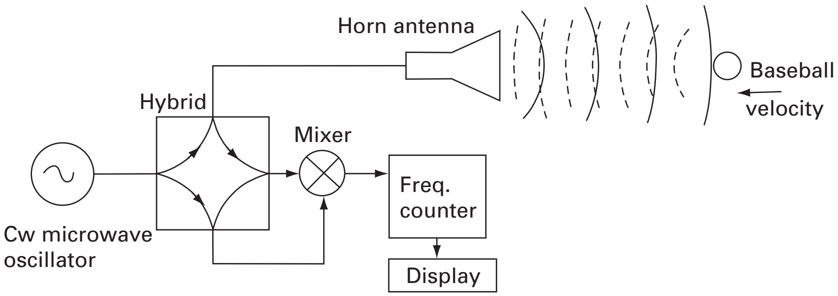
Horn antenna
Baseball
velocity
Hybrid
Mixer
Freq. counter
Display
Cw microwave oscillator

Figure 21.2 Radar speed gun.

The signal from a cw microwave oscillator in the 10–20 GHz range (usually a Gunn diode negative resistance oscillator) is split to feed the horn antenna and a mixer (multiplier). The echo from the target also provides a signal to the mixer. A moving target causes a Doppler shift, so the echo has a higher frequency if the target is approaching the radar. The output from the mixer contains a signal whose frequency is equal to the Doppler shift. This frequency is counted and multiplied by the appropriate factor so that the display reads line-of-sight velocity. Note that the Doppler shift depends only on the line-of-sight velocity of the target. This radar is totally insensitive to range.

21.1.2 Chirped cw radar

The radar in Figure 21.3 differs from the radar speed gun only in that its oscillator is a "chirped" VCO, whose output frequency follows a sawtooth control voltage. Here this radar is used to determine the position of a stationary (or very nearly stationary) fluid level in a storage tank.

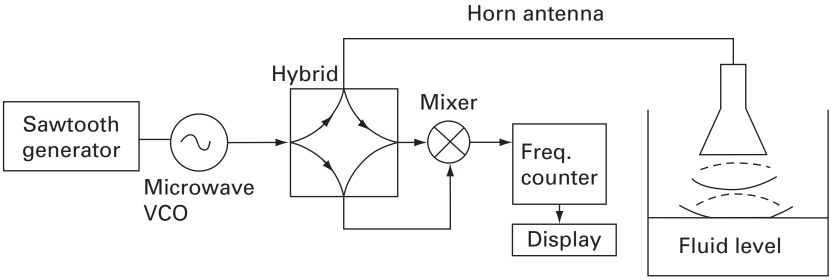
Horn antenna
Hybrid
Mixer
Sawtooth generator
Microwave VCO
Freq. counter
Display
Fluid level

Figure 21.3 Radar fluid level detector.

The frequency of the tone at the mixer output is given by the frequency difference of the two signals arriving at the mixer inputs. One of these signals is delayed by the round-trip time to the fluid surface, so the lower the level in the tank, the higher the output frequency of the mixer. Note that this radar cannot determine velocity; if the fluid level were changing rapidly, it would create a

Doppler shift that could not be distinguished from a change in fluid level. (Range and velocity are ambiguous for this radar.)

21.1.3 Pulse-Doppler radar

The radar of Figure 21.1 can be modified to sense velocity as well as position. The resulting *pulse-Doppler radar*, when mounted on an aircraft, can distinguish moving targets on the ground from the much stronger "ground clutter" produced by stationary objects. The earliest form of pulse-Doppler radar, called "MTI", for *moving target indicator*, used a mercury acoustic delay line as analog storage for the train of echos received after each pulse. The stored signal was subtracted from the echoes produced by the next pulse to remove the echoes from fixed targets. (This type of delay line, when configured as a recirculating shift register, was used around 1950 as the memory in some of the earliest digital computers.)

Pulse-Doppler radar really became practical with the advent of frequency-stable microwave signal sources (klystron amplifiers, rather than high-power magnetron oscillators), digital storage capability, and digital signal processing. These advances allow coherent signal averaging for increased sensitivity, as well as actual determination of target velocity. Successive received echoes, after being mixed to a low IF frequency, are stored, preserving both amplitude and phase, and then the stored echos are analyzed to see how the phase changes from one pulse to the next. A moving target causes a progressive phase change (Doppler shift) from which the line-of-sight velocity can be determined. When used as a search radar, successive received signals are stored separately for every range. Fourier analysis of the sequence of sampled IF voltages for each range determines whether there is a target at that range and, if so, determines its velocity. A problem that often arises with pulse-Doppler systems involves a range-velocity ambiguity. If target velocities are in the range $\pm v_{\max}$, the corresponding Doppler shift range is $\pm 2f\ v_{\max}/c$, where f is the radar frequency. This produces a total received bandwidth of $4fv_{\max}/c$. If we are to unambiguously determine the frequencies present in this band, the sampling rate, which is the pulse rate, must be at least $8f\ v_{\max}/c$. A high radar frequency and a high maximum velocity may dictate a pulse rate that restricts the maximum unambiguous range. (The echo from a target beyond the maximum unambiguous range arrives after the next pulse has been transmitted, as if it had been reflected from a close-in target.) This problem is discussed in Chapter 26 in the context of planetary radar astronomy.

21.2 Radar classification

Many radar systems are quite different from the classic configuration of Figure 21.1. Radar systems are conveniently classified according to several key system characteristics:

(a) Monostatic or bistatic:
A monostatic radar uses a single antenna for both transmitting and receiving, while a bistatic radar uses two separate antennas. An obvious advantage of the bistatic configuration is that it requires no TR (transmit/receive) switch. A military advantage is that only the transmitting site, which can be unmanned, is vulnerable to attack by a radiation-seeking missile. Bistatic and multistatic radar systems can also be passive, using "transmitters of opportunity," such as television broadcast transmitters.

(b) Coherent (Doppler) or incoherent:
Radars with stable frequency sources for the transmitter and the receiver L.O. can use coherent integration to extract a signal from noise. They can also use the Doppler effect to measure the line-of-sight velocity of a moving target. Incoherent radars detect the power of each pulse, making signal averaging less efficient. They cannot directly detect target velocity though they can, of course, infer velocity by observing successive changes in range.

(c) Pulse or cw:
Short pulses can distinguish closely separated targets, providing range resolution. Long pulses, whose limit is cw (continuous wave), can distinguish target velocity using the Doppler shift. Nevertheless, coherent trains of short pulses and phase-modulated long pulses can provide both range and Doppler resolution.

(d) Search or track:
Aircraft traffic monitoring is normally done with a monostatic search radar, using a rotating antenna that produces a beam pattern that is tall in elevation but narrow in azimuth. Echos from aircraft produce "blips" on a PPI (plan-position indicator) screen which is refreshed on every rotation of the antenna. A tracking radar uses servo motors to keep the antenna pointed at a single, usually fast-moving target.

(e) Mechanical or electronic scanning:
A *phased array* uses a closely spaced two-dimensional array of antenna elements, usually dipoles. Each dipole is equipped with a programmable phase shifter, allowing the array to form a beam in an arbitrary direction. If the signals received from the individual dipoles are made available to the signal processor, the processor can simultaneously form several independent receiver beams. This allows tracking of multiple targets or searching of multiple zones. Phased arrays with three or more faces of elements can be completely stationary.

(f) Detection or imaging:
Conventional radars simply detect the position and/or velocity of targets. Imaging radars make photographic-like reflectivity maps. Imaging systems include airborne side-looking (*synthetic aperture*) radars that map a strip of land parallel to the flight path and "planetary radar" (delay-Doppler) systems that image planetary objects and low-orbit artificial satellites.

21.3 Target characteristics and echo strengths

21.3.1 Radar cross-section

The strength of an echo depends on the nature of the target and on its range from the radar. In general, metal targets with dimensions commensurate with the wave length are good reflectors, especially when their geometry includes linear features. For example, airplane propellers act as dipole re-radiators. A flat metal plate produces a strong specular reflection, if perpendicular to the direction of the radar, but edge currents also produce other, less directional, echoes. Radar engineers define the *radar cross-section* (RCS) of a target, σ, as the collecting area of an object that would produce the target's observed echo strength while *isotropically* re-radiating the intercepted incident radiation. Radar cross-sections for practical targets, such as aircraft, depend on the aspect angle of the target as well as on the target's usually complicated geometry. RCS calculations for such targets are carried out using finite-element electromagnetics modeling programs. Exact expressions have long since been derived for the radar cross-sections of simple objects. A sphere, since it has no aspect angle dependence, makes a good calibration target for measuring radar sensitivity. The RCS of a metal sphere of radius a is just πa^2 (its geometric cross-sectional area) when a is greater than about $5\lambda/\pi$. But when a is less than about $\lambda/(4\pi)$, the RCS is given by the formula: $\sigma = \pi a^2\ [9\ (2\pi a/\lambda)^4]$. Note that the cross-section is proportional to a^6 and to λ^{-4} in this small-target or *Rayleigh scattering* regime.[2]

21.3.2 The radar equation

The definition of radar cross-section lets us write a simple expression for the flux density of a backscattered echo as it returns to the radar antenna

$$S_{\text{received}} = S_{\text{incident}}\sigma/(4\pi R^2), \tag{21.1}$$

where R is the distance to the target. This equation just says that the supposed isotropically scattered power, which is given by $S_{\text{incident}}\ \sigma$, produces a flux density determined by the area of a target-surrounding sphere whose radius is the distance to the radar. From antenna considerations discussed in Chapter 20, we can write the incident flux in terms of the transmitted power as $P_{\text{trans}}\ G/(4\pi\ R^2)$, where G is the gain of the antenna. Of course, this assumes the antenna is pointed directly at the target. Likewise, we can express the capture area of the antenna by $A_{\text{eff}} = \lambda^2 G/(4\pi)$. Putting this together, we obtain an expression for the received power, P_{rec}, which is just $S_{\text{incident}}\ A_{\text{eff}}$ or

---

[2] The reason the sky is not black is that the molecules that make up the atmosphere scatter direct sunlight to produce diffuse sky light. The λ^{-4} dependence of Rayleigh scattering causes sky light to be blue and sunsets to be red.

$$P_{\text{rec}} = [P_{\text{trans}}G/(4\pi R^2)][\sigma/(4\pi R^2)][\lambda^2 G/(4\pi)] = P_{\text{trans}}G^2\sigma\lambda^2/[(4\pi)^3 R^4)]. \quad (21.2)$$

This relation, however written, is known as the *radar equation*, and shows the inverse fourth-power dependence on range. For an aperture antenna, e.g., a dish antenna, the effective aperture is given by $A_{\text{eff}} = \eta_{\text{ap}}\,\pi(D/2)^2$ where D is the dish diameter and η_{a}, the aperture efficiency, is usually around 0.5. Using this, we can write the radar equation as

$$P_{\text{rec}} = P_{\text{trans}}\frac{\pi}{64}{\eta_{\text{a}}}^2\left(\frac{D}{R}\right)^4\frac{\sigma}{\lambda^2}. \quad (21.3)$$

21.3.3 Distributed targets

Targets such as aircraft subtend a solid angle much smaller than the radar beam, but distributed targets, such as rain in the atmosphere or free electrons in the ionosphere, can be much larger than the beam. In these cases the beam actually defines the extent of the target, in as much as there will be radar echoes from throughout the entire volume of the beam. When the individual scattering objects (raindrops and electrons in the above examples) are randomly distributed in space (*incoherent scattering*), the power received by the radar is the sum of the powers of the individual scatterers.[3] For this kind of target, the radar equation must be modified. We first define σ_{V}, the radar cross-section per unit volume of the target, to be $\sigma_{\text{V}} = n\sigma$, where n is the volume density of scattering particles and σ is the radar cross-section of an individual particle. For raindrops, σ is about 41% of the Rayleigh scattering cross-section given above for metal spheres. For individual electrons, $\sigma = [e^2/(4\pi\epsilon_0\, m_{\text{e}}c^2)]$, where e and m are the charge and mass of an electron and c is the speed of light. If we extend the arguments we used for a single particle centered in the beam to include each volume element throughout the beam, we find that

$$P_{\text{rec}} = \frac{P_{\text{trans}}\lambda^2\sigma_{\text{V}}c\tau}{(4\pi)^2 R^2}\frac{1}{4\pi}\int G^2(\phi,\theta)\text{d}\Omega. \quad (21.4)$$

In this radar equation τ is the pulse length, so $c\tau$ is the range depth of the target volume. The cross-range area of the beam is taken into account by integrating over solid angle times R^2, the square of the range. Note that, because the beam determines the target size, the return echo of a distributed target is proportional to R^{-2}, rather than R^{-4}. The last term in Equation (21.4), $(4\pi)^{-1}\int G^2\,\text{d}\Omega$, is the mean square gain, also called the *backscatter gain* of the antenna.

---

[3] The electric field of the radar echo is the sum of the field contributions from the N individual scattering particles illuminated by the beam. If the scatterers have random positions, their E-field contributions have random phases and add in the fashion of a random walk, causing the total E-field echo to be proportional to $\sqrt{N}$. The received power, proportional to the square of the received E field, is therefore proportional to N.

21.4 Pulse compression

Short pulses produce good range resolution, but shortening the pulses reduces the sensitivity of the radar unless the peak power (pulse power) can be increased to maintain the same average power. However, the peak power of any transmitter is eventually limited by voltage breakdown or other device limitations. This constraint led to the development of *pulse compression* schemes in which the transmitted pulses are modulated in such a way that the echoes can be "bunched up" by the receiver as if they had started out as short pulses. The received echo is passed through an appropriate matched filter (often a digital processor) whose output, for a point target, will be a narrow pulse.[4]

One method for pulse compression uses phase coding. The transmitted pulse is divided into N contiguous equal intervals. Each interval or *baud* is assigned a phase value of zero or 180 degrees, which will be the relative phase (and therefore polarity) of the transmitted signal during that interval. These assignments specify the code. Normally every transmitted pulse will have the same code. Figure 21.4 shows how such a code is used for pulse compression. At the receiver, the IF signal from a point target will be a replica of the transmitted pulse. This IF signal travels down a delay line. Taps along the delay line have a delay spacing equal to the baud length. Signals extracted at these taps are weighted (multiplied) by coefficients with values ± 1, whose order duplicates the code. The weighted signals are added together to form the pulse compressor output.

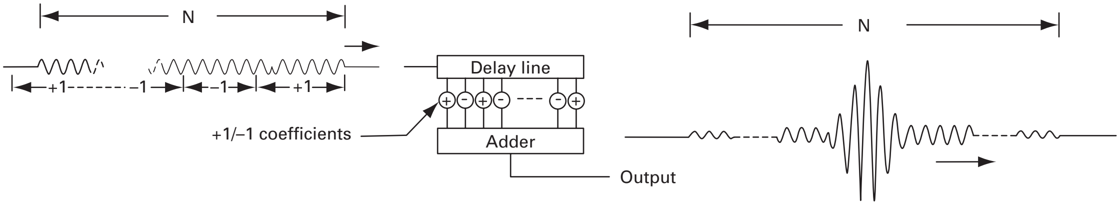

Figure 21.4. A compressor for phase-coded pulses.

Note that, for one position of the signal in the delay line, the multipliers will exactly "undo" the coding, giving the same phase to each of the N input signals to the adder. This alignment will peak at the middle of one baud, as the signal propagates down the delay line, attaining a peak amplitude N times greater than the amplitude of the uncompressed input signal. When the signal in the delay line is not in alignment with the code, there will be some output from the adder unless the code is such that the addends sum to exactly zero. There is no code with this property, though the *13-baud Barker code*, $+ + + + + - - + + - + - +$, produces

---

[4] In general, high range resolution requires that the transmitted signal have a large bandwidth. This bandwidth can be produced by transmitting either short pulses or appropriately modulated long pulses.

("sidelobe") values of only +1 or 0 when the code is not aligned and +13 when it is aligned. When longer codes are needed, it is common to use pseudorandom codes. In this case, the sidelobe levels are determined by random walk statistics; the close-in sidelobes will be less than the main lobe by a factor on the order of $\sqrt{N}$.

An analog surface acoustic wave (SAW) processor can do all the operations shown in Figure 21.4, but it is now more common to use digital processing. Typically, baseband signals are produced by mixing the IF signal, centered at ω_{IF} with $\cos(\omega_{\mathrm{IF}}\ t)$ and $\sin(\omega_{\mathrm{IF}}\ t)$. The resulting I and Q signals are each furnished with a tapped digital delay line/adder pulse compressor. The outputs from the two compressors can be squared and added together for immediate detection, or they can first be Doppler processed.

Pulse compression has other benefits. Because the rapidly changing code spreads the signal's spectrum, it lowers the spectral density. This gives a military radar signal an element of stealth; being less noticeable, it is less likely to provoke a hostile reaction, such as an antiradiation missile. By using different codes with low cross-correlation properties, multiple radars can share the same frequency band. This code-division multiple access property is also used in CDMA cellular telephone systems and in the GPS satellite system, where multiple users (the callers or satellites) share the same band. The GPS system also makes use of the pulse compression property to achieve the high time resolution needed for accurate range resolution (see Chapter 25).

21.5 Synthetic aperture radar

As described above, a phased-array antenna combines the signal from many individual elements to form a beam pattern. Consider a phased array directed toward a point source of cw radiation whose frequency is precisely known. The analog voltage from each element is given a programmed analog phase shift. These voltages then combine to form a vector sum. Note, however, that it is also possible to digitize the voltage (phase and amplitude) from each element and then *compute* the vector sum. If we do this, no hardware phase shifters are needed since, before summing, we can multiply each voltage, V_i by a phase factor, $\mathrm{e}^{\mathrm{j}\phi_i}$, to point the beam in any desired direction. This can be repeated with other sets of phase factors to form a host of beams from the one set of digitized voltages. The echo powers in these beams then form an image if there are multiple targets (assuming they all radiate at the same frequency). Finally, note that we could do all of this with only a *single* array element by moving the element successively from one position to another, recording the voltage (amplitude and phase) at each position. If the set of positions is identical to the element positions in the original phased array, we have effectively used a single small element to *synthesize* a steerable large-aperture antenna.

You can see that this all works out in a radar situation, where all the targets are illuminated by a coherent transmitter signal. The transmitter also provides a

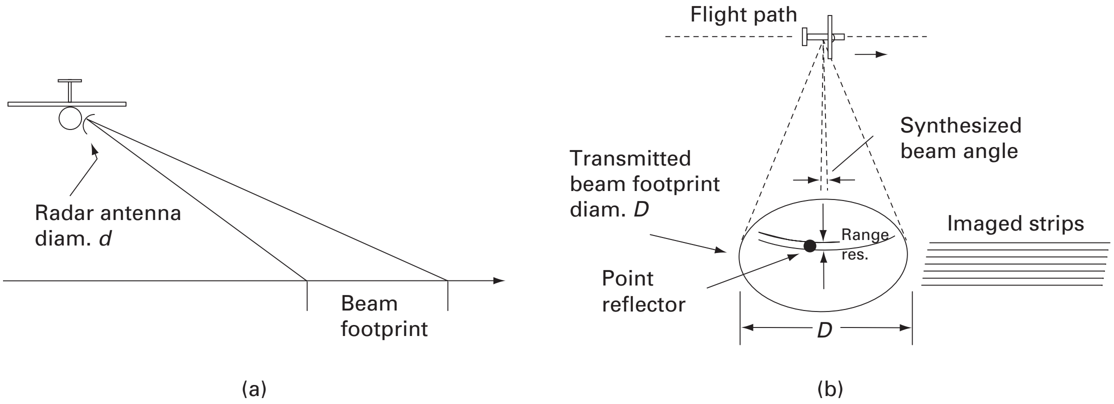

Figure 21.5. Side-looking synthetic aperture radar: (a) elevation view; (b) plan view.

phase reference for the receiver so it is practical to measure both the amplitude and the phase of the return signal at the antenna element. This is the general idea of *synthetic aperture radar* (SAR).[5] Here we assumed that the target was stationary and that the element was stationary, except while being relocated after each voltage measurement. In practice, most SAR techniques involve continuous motion of the elemental antenna, relative to the target. This continuous motion makes it possible to explain the synthesis procedure in terms of Doppler shifts. In addition, most SAR systems transmit short pulses to provide resolution in the down-range (pointing) direction, and use one-dimensional aperture synthesis to provide resolution in the cross-range direction. One of the most common SAR systems is the side-looking radar shown in Figure 21.5.

A small radar antenna of diameter d points outward and downward from the aircraft, which flies a straight-line course. The illuminated patch on the ground, called the *radar footprint*, has a large diameter, D. Since the beamwidth from the radar antenna is given by λ/d radians, we see that D is given by $D = R\lambda/d$, where R is the distance of the footprint from the aircraft. Range resolution is obtained by simply using short radar pulses. In the cross-range direction, the system synthesizes an antenna whose extent in the direction parallel to the flight path is D, since that is the dimension of the footprint that passes over any target feature. The angular size of the resulting synthesized beam is given by $\delta\theta = \lambda/D$. Substituting $D = R\lambda/d$ produces $\delta\theta = d/R$, and the cross-range resolution is therefore $R\delta\theta = d$, the diameter of the small radar antenna. It is interesting to note that the resolution is independent of λ and that making d smaller improves rather than degrades the resolution. Many range strips are observed simultaneously. You can find fascinating side-looking radar (and side-looking sonar) images by searching the internet. A "delay-Doppler" SAR system to image planetary surfaces is discussed in Chapter 26. Another interesting but difficult technique is *inverse synthetic aperture radar* (ISAR) where the aspect of the target (which might be a ship bobbing on the ocean) is variable, not known a priori, and must be inferred from the radar returns as part of the data processing.

---

[5] Synthetic aperture mapping is also possible when the target is not illuminated by coherent radiation, but emits its own wideband radio "noise." In this case, aperture synthesis requires at least two antenna elements. This interferometry technique, used by astronomers to make maps of the radio sky, is discussed in Chapter 26.

21.6 TR switches

Monostatic radars, which transmit and receive with the same antenna, require a TR (transmit-receive) switch. In most radar applications the desired echo arrives so soon after the pulse is transmitted that the TR switch (also known as a *duplexer*) must be electronic rather than a mechanical. Here we will first look at self-duplexing radar techniques based on the use of circular polarization or circulators, then at standard TR switch circuits, and finally at RF electronic switches in general.

21.6.1 Self-duplexing radar techniques

If a radar transmits a circularly polarized signal, reflection by the target changes the sense of polarization from left-hand to right-hand or vice versa. Circular polarization can be produced by transmitting simultaneous crossed linear polarizations 90° out of phase. Figure 21.6 shows how a 90° hybrid not only produces circular polarization of one sense but also routes received circular polarization of the other sense (the return signal) into the receiver.

Note that the x-dipole and the y-dipole, together with the hybrid, are really just equivalent to two separate antennas having opposite circular polarizations. In practice, the isolation between the transmitter and receiver in this scheme is usually no better than about 30 or 40 dB, so a limiter or SPST electronic switch (a *monoplexer*) is installed at the receiver input to protect it from burnout. A waveguide version of this circuit uses a *turnstile junction*, the microwave component shown in Figure 21.7. It is classified as a six-port junction because the round waveguide supports two independent modes: x and y or RCP and LCP or any other pair of orthogonal elliptical polarizations. When two opposite

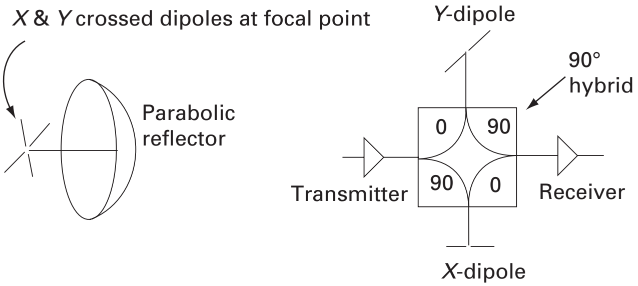

Figure 21.6. A self-duplexing radar using circular polarization.

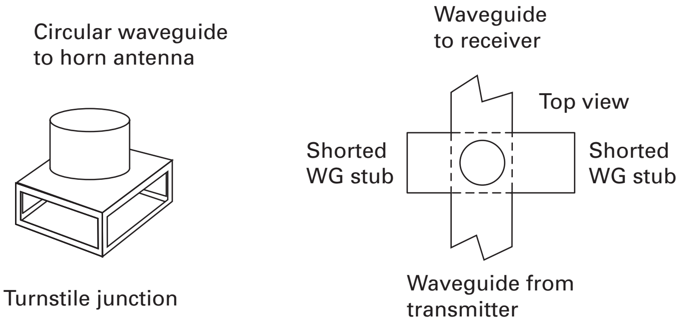

Figure 21.7. A waveguide turnstile junction combines the functions of the hybrid and crossed dipoles.

rectangular ports are fitted with shorts of appropriate lengths, the resulting four-port network is equivalent to the pair of dipoles and hybrid of Figure 21.6. The turnstile junction is described in Volumes 8 and 9 of the Rad. Lab. Series. The transmitter and receiver are connected to the remaining rectangular ports (the pair without shorts) while the antenna, usually a feed horn, is connected to the round waveguide.

A true self-duplexing circuit, shown in Figure 21.8, uses a circulator. The circulator has the property that a signal injected at the transmitter port will exit via the antenna port while a signal injected at the antenna port will exit through the receiver port. (If a signal were injected at the receiver port, it would exit through the transmitter port.) This nonreciprocal action depends on transmission through a nonreciprocal medium which, for the circulator, is a ferrite material, biased by the field of a permanent magnet. This elegant TR system is limited by available circulators to powers of tens of kilowatts.

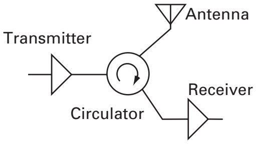

Figure 21.8. A circulator used as a TR switch.

21.6.2 TR switching devices and circuits

The classic TR circuits developed during WWII use gas discharge tubes or spark gaps and are self-activated by high-power RF on the transmission line. Lower power versions use PIN diode switches, turned on by an external bias circuit. (The radar has a timing generator providing pulses that (1) turn on the TR switch for transmitting, (2) pulse the transmitter, and (3) turn off the TR switch for receiving.) Gas discharge tube switches are usually built into a short piece of waveguide and come in two types: TR and ATR (anti-TR). The distinction is as follows: on transmit a TR device ionizes and presents a low-impedance shunt across the line. The ATR device also ionizes on transmit but it presents a low impedance in series with the line. (Reference [3], Volume 14 of the Rad. Lab. Series, is devoted mostly to these tubes.) There are two general classes of TR circuits, branch line TR switches and balanced TR switches. The former uses segments of transmission line while the latter uses hybrids.

21.6.2.1 Branch line TR switches

Figure 21.9 shows some standard branch line TR Switch circuits using TR, ATR, or both TR and ATR tubes (or PIN diode equivalents).

In the circuit of Figure 21.9(a), the switches are shown in the nonconducting (transmitter off) position. The open ATR switch is connected to the antenna by a half-wave line so it presents the same open circuit to the antenna-to-receiver line. Likewise, the open TR switch does not disturb the antenna-to-receiver line. On transmit, the TR switch places a protective short circuit at the receiver input. This short circuit is transformed by the quarter-wave line into an open circuit which does not affect the connection between the transmitter and the antenna. At high frequencies the switches contain some nonzero path lengths which form part of the half-wave or quarter-wave lines. For low-frequency designs the half-wave lines can be reduced to zero length and the quarter-wave lines can be replaced by lumped-element impedance inverters. You can see from the ways that the TR and ATR elements are used that their names are somewhat arbitrary.

Figure 21.9. Branch line TR switches.

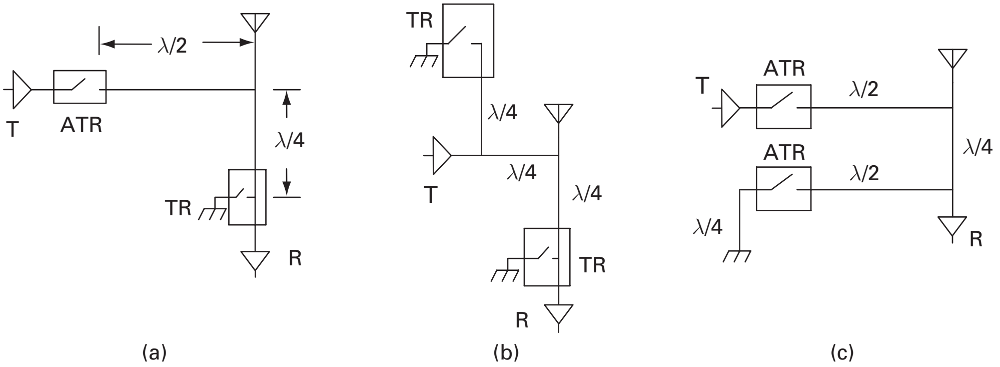

21.6.2.2 Balanced duplexers

Balanced duplexers use hybrids and can have wider bandwidths than the branch line circuits shown above (though more elaborate branch line circuits can have wider bandwidths). Both balanced duplexer circuits shown in Figure 21.10 use two 90° hybrids.

Figure 21.10. Balanced TR switches.

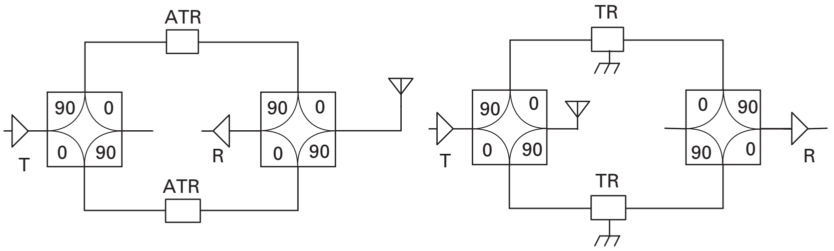

21.7 Diode switches

A single diode shunt switch circuit is shown in Figure 21.11.

Positive control voltage gives the diode a forward bias current to produce a low dynamic resistance, dV/dI. Negative control voltage turns the diode off, making its dynamic resistance very high. Let these dynamic resistance values be denoted respectively by r and R. You can verify (Problem 21.5) that the transmission values (power out/power available) for the switch of Figure 21.11 are given by:

$$\text{Isolation (forward biased state)} = 4r^2/Z_0^2$$
$$\text{Transmission (reverse biased state)} = 1/(1 + Z_0/R).$$

Note from these expressions that you could favor better isolation or lower insertion loss by transforming the line to have a larger or smaller Z_0 at the diode location. Better performance can be obtained with ladder networks analogous to multisection filters. Figure 21.12 shows how a shunt switch can be combined with a series switch to form a two-element ladder network. Isolation is improved since any signal leakage across the open series switch is shorted to ground by the closed shunt switch.

It is often more convenient to use shunt diodes than series diodes, since shunt diodes are easier to bias and to heat sink. Impedance inverters can transform series elements into shunt elements, as we saw when designing coupled

Figure 21.11. Shunt diode switch.

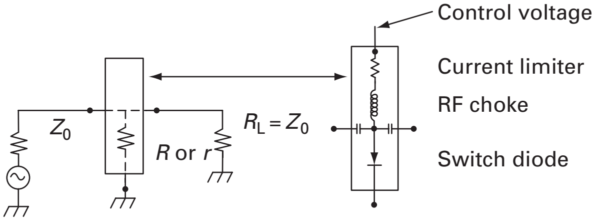
Control voltage
Current limiter
RF choke
Switch diode
Z_0
R or r
$R_L = Z_0$

Figure 21.12. Series-shunt diode switch.

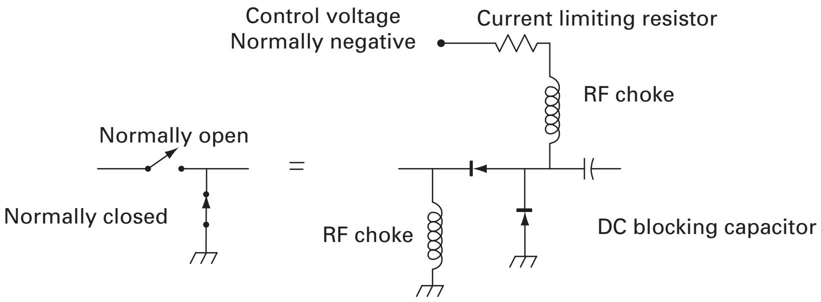
Control voltage
Normally negative
Current limiting resistor
RF choke
Normally open
Normally closed
=
RF choke
DC blocking capacitor

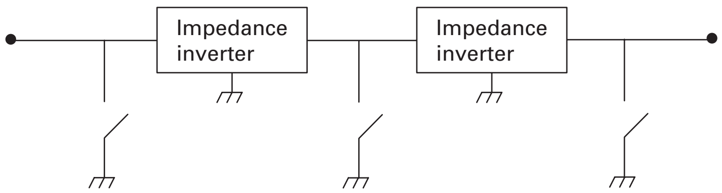

Figure 21.13. Three-section switch using only shunt sections.

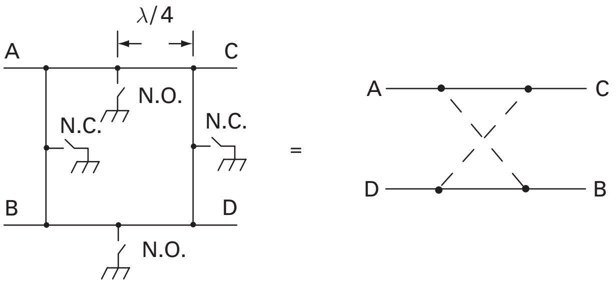

Figure 21.14. Transfer switch using quarter-wave transmission lines.

resonator filters. The switch circuit of Figure 21.13 uses two impedance inverters and three shunt switches.

The switches discussed above are all SPST switches. More complicated switches can be built up from the elementary SPST switch, but special designs can often be made such as the four-diode transfer switch shown in Figure 21.14. When the switches are in the indicated normal positions, the shorted quarter-wave lines appear as infinite-impedance shunts and do not disturb the transmission paths from A to C and from B to D. When the switches are reversed, transmission is from A to B and from C to *D*.

21.7.1 Diodes for RF switching

When ordinary diodes are used in these switching applications, the biases must be enough to keep the diode in the desired state. In particular, when the diode is off, the reverse bias voltage must be greater than the peak RF voltage and, when the diode is on, the forward bias current must be greater than the peak RF current. However, the *PIN diode*, a sandwich of p-type, intrinsic, and n-type semiconductor material, has the remarkable property that, for RF switching, the bias current and bias voltage values can be less than the corresponding RF current and voltage by perhaps an order of magnitude. The operation of the PIN diode (see reference 5) depends on having a large (small) stored charge in the intrinsic region when the diode is on (off). At high frequencies, the time between electric field reversals is much less than the transit time through the intrinsic region, determined by diffusion and drift, so the diode remains in the on (off) state. Finally, a word of caution: diodes, since they are nonlinear circuit elements, have the potential to distort a signal. In particular, they can create intermodulation products between the various signals in a complicated spectrum. For critical applications, e.g., receiver band-changing, diode switches must be turned off and on hard enough to keep any generated intermodulation products at a negligible level.

21.8 Radar pulse modulators

Any cw transmitter can generally be used as a pulse transmitter if a *pulse modulator* is added to provide the rapid turn on and turn off. Tube-type amplifiers can be operated with much higher instantaneous powers when they are pulsed. Tubes are primarily limited by their maximum anode dissipation (heat removal); the dissipation can be the result of either modest cw operation or high-power pulse operation. A cw amplifier can be converted for pulse operation by changing the output matching circuit in order to present a lower load resistance to the tube. Some tubes are available in special pulse-rated versions; they are fitted with high-emission cathodes. Gridded tubes (triodes, tetrodes, and pentodes) can be pulsed by switching the grid bias from negative, for pulse-off, to positive, for pulse-on. The negative bias keeps the tube completely turned off between pulses. Since the grid voltage and current are much smaller than the plate voltage and current, grid control requires only low-power circuitry compared to anode control. At microwave frequencies, magnetrons and klystrons replace gridded tubes. Magnetrons have no control element and therefore require high-power anode pulsers. Klystrons may or may not have a modulating anode ("mod anode") by which the beam current can be cut off. If not, they need high-power pulsers.[6]

Transistor amplifiers, unlike tube amplifiers, can make very little trade-off between duty cycle and peak power. Transistors suffer one type of breakdown or another when operated beyond their maximum continuous ratings. A high-power transistor amplifier for pulse service might differ from a cw amplifier only in that it will dissipate less heat (from the reduced duty cycle) and can therefore get by with a smaller heat sink.

No matter how an amplifier is pulsed, the power supply must furnish high-power pulses with minimum voltage droop. Duty cycles of pulsed transmitters are usually much less than unity so, in addition to at least one switching element, pulse modulators (pulsers) contain some form of energy storage element(s). The simple pulser circuit shown in Figure 21.15(a) stores energy in a capacitor.

---

[6] An air traffic control radar might have a peak power output of 2 MW and an efficiency of 50%. A klystron tube in this service could require 50 kV pulses at 80 amperes.

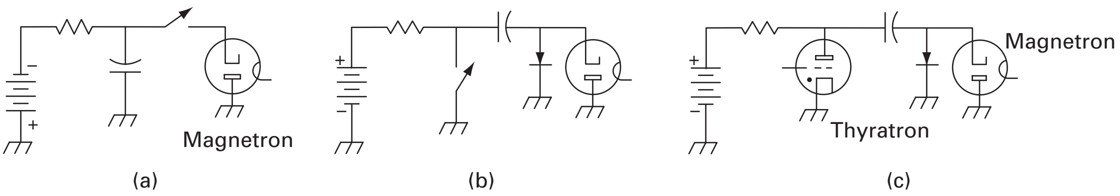

Figure 21.15. Capacitor discharge pulser.

In this circuit the tube (magnetron, klystron, or whatever) is shown as requiring negative voltage. Microwave tubes often use negative supply voltage applied to their cathodes because it is convenient to ground the external heat-dissipating anode. The version of the circuit in (b) allows one side of the switch to be grounded, which is another convenience. The diode provides a charging path for the energy storage capacitor. The circuit of Figure 21.15(c) uses a thyratron (vacuum tube version of the SCR) as the switch.

The simple capacitor discharge modulators in Figure 21.15 have several disadvantages: The voltage droops during the pulse. The droop can be reduced by increasing the size (weight, and cost) of the capacitor. Not much of the stored energy is used. Even if 10% voltage droop is permitted, only 20% of the stored energy is used for each pulse. This might be compared to a car which would not run well if the fuel tank was less than 80% full. They are expensive and heavy, a particular disadvantage for airborne equipment. Despite these disadvantages, capacitor banks are often used, as in the 430-MHZ pulse transmitter used for ionospheric research at the Arecibo Observatory, because a more efficient circuit, the line modulator discussed below, does not easily provide the flexibility needed to change the pulse width.

21.8.1 Line modulators

A length of transmission line (with the far end open) has capacitance and can therefore store electrostatic energy. When the line is discharged into a resistive load equal to its characteristic impedance, it will supply a perfect rectangular pulse rather than a drooping exponential pulse. The constant pulse amplitude during discharge is maintained by the distributed inductance of the line acting together with the distributed capacitance. In Figure 21.16(a), the line is a piece of coaxial cable, replacing the energy storage capacitor. As before, the tube is supplied with a negative pulse. A diode provides a path to recharge the line. Often the load has a higher impedance than the characteristic impedance of the line, and a pulse transformer is required.

The line supplies a pulse at half the charging voltage since, during the pulse, the charging voltage evenly divides between the load and the equivalent source resistance. The duration of the pulse is the time taken for the current to make a round trip through the line. At the end of the pulse the line is totally discharged; all the stored energy is delivered on every pulse. Figure 21.16(b) shows successive plots of the line's voltage and current distributions. In order to store more

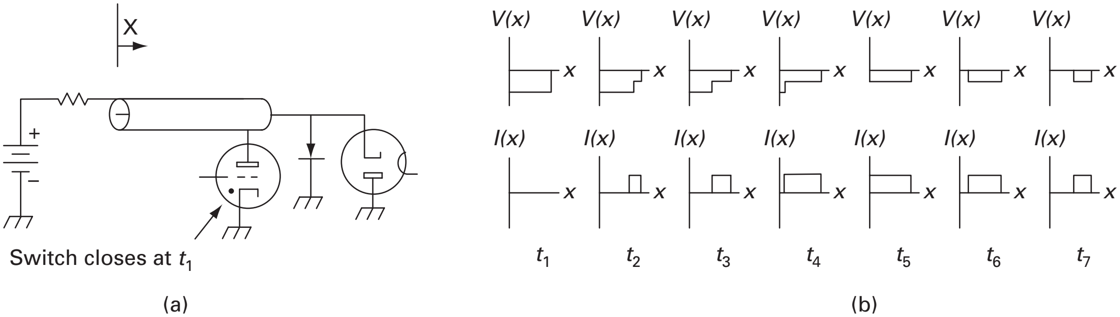

(a) (b)

Figure 21.16. (a) Line-type modulator; (b) line voltage and currents.

Figure 21.17. Pulser using an artificial transmission line (pulse-forming network).

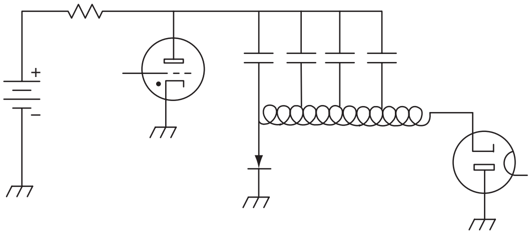

energy, it is common to use an "artificial transmission line" or pulse forming network (PFN) which is a ladder network of inductances and capacitances. A four-section network is shown in the modulator circuit of Figure 21.17.

The network looks like a lowpass filter and it is. Its cutoff frequency is given by $\omega^2 = 4/[LC]$. For frequencies well below cutoff, the network behaves like a transmission line with $Z_0 = \sqrt{L/C}$. Here L and C are in henries and farads rather than henries/meter and farads/meter as in the distributed element transmission line. The one-way time delay on this lumped line is $\sqrt{LC}$ seconds/section.

Let us consider a numerical example: Let us use four sections as in the figure above. Suppose we need a 1 microsecond pulse at 10 kV and 10 amperes. The voltage and current require that $Z_0 = \sqrt{L/C} = 1000$ ohms. We will use four sections which, for the desired one-microsecond round-trip delay, requires that $8\sqrt{LC}$ be equal to $10^{-6.}$ These impedance and time delay equations are satisfied by $L = 125$ μH and $C = 125$ pF. We can verify that the energy stored in the line is indeed equal to the energy delivered by the pulse. The latter is just $(IV)\tau = 10 \times 10\,000 \times 10^{-6} = 0.1$ joule. The former, remembering that we must charge the line to 20 000 V, is $CV^2/2 = 4(125 \times 10^{-12}) \times 20\,000^2/2$ which is also 0.1 joule. As often happens in filter design, these are not particularly practical values; real inductors of 125 μH may well have distributed capacitances that are not negligible compared with 125 pF. We can build the line for a lower

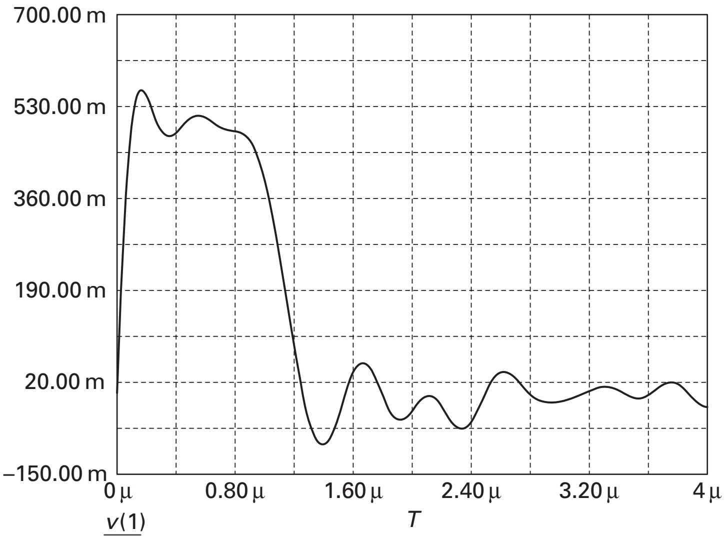

Figure 21.18. Waveform produced by a four-section pulse forming network.

impedance and use a pulse transformer between the line and the magnetron. If we lower the line impedance to 100 ohms, the *L* and *C* values become 12.5 μH and 1250 pF, values that are more practical. Using these values, a SPICE simulation of the discharge produced the voltage waveform shown in Figure 21.18. The voltage scale is normalized, i.e., the capacitors were charged to one volt so the nominal pulse voltage is 0.5 volts. Most pulse forming networks can be tuned slightly to improve the pulse shape; the artificial transmission line is, after all, only an approximation to an ideal transmission line.

The line modulator uses all the stored energy on each pulse but, precisely because of this virtue, deserves a more sophisticated charging circuit than the resistor shown in the circuits above. Remember that when a capacitor is charged through any resistive path from empty (no energy) to $CV^2/2$, the resistor will dissipate this same amount of energy, $CV^2/2$. Here the charging resistor, no matter what value, would dissipate half the power consumed by the radar. The solution to this problem is to charge the line through an inductor instead of a resistor. Figure 21.19(a) shows the voltage waveform on a capacitor as it is *resonantly charged* through an inductor.

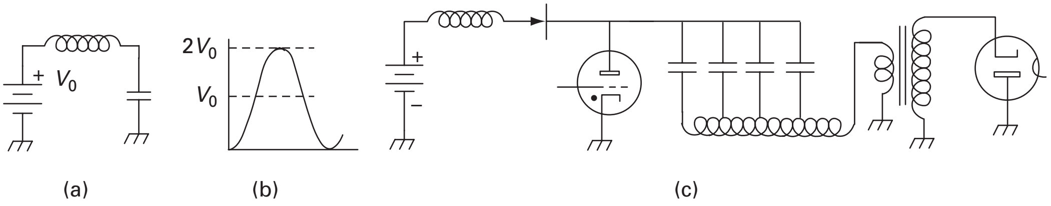

Figure 21.19. Resonant charging.

The voltage is a sinusoid, building up to a maximum of twice the supply voltage. The modulator can be triggered just as the voltage reaches this maximum. The brief pulse discharges the line and the charging curve begins anew. It would seem that the PRF is therefore determined rigidly by the charging time but, if a diode is put in series with the inductor, the charging stops at the maximum voltage and the next pulse can occur anytime. The resonantly charged modulator, with the diode and a pulse transformer, is shown in Figure 21.19(b). Note that the primary of the pulse transformer provides a charging path, eliminating the diode originally in parallel with the magnetron. Also remember that, because of the resonant charging, the supply voltage needs only to be half of the line charging voltage. Line modulators present less risk to tubes than partial-discharge capacitor modulators because there is less stored energy available when an arc occurs in the tube.

Problems

Problem 21.1. Find the radar cross-section of a flat metal plate of area A that is exactly perpendicular to the radar beam. Assume that $A \gg \lambda^2$. Hint: treat the plate as an aperture antenna with gain $4\pi A/\lambda^2$.

Problem 21.2. Find the lobe pattern produced when a point target is observed by a pulse compression radar using the 13-bit Barker code of Section 21.4. (Convolve the code with itself after padding both ends with zeros.)

Problem 21.3. Why is the side-looking radar antenna positioned to look to one side of the aircraft rather than straight down?

Problem 21.4. If we try to use the hybrid of the circular polarization duplexer as a circulator, we might consider the TR circuit shown below.

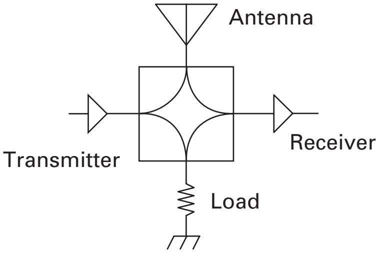

Assume the antenna, transmitter, receiver, load, and hybrid all have the same characteristic impedance. This circuit at least protects the receiver from the transmitter. What are its disadvantages (a) when transmitting, and (b) when receiving?

Problem 21.5. Verify the expressions below for $P_{out}/P_{available}$ for the circuit of Figure 21.11.

Diode state	Switch state	$P_{out}/P_{available}$
Foward biased	Isolation	$4r^2/Z_0^2$
Reversed biased	Transmission	$1/(1+Z_0/R)$

Problem 21.6. Apply your circuit analysis program (Problem 1-3) to the transfer switch circuit of Figure 21.4. Assume a 50-ohm load is connected to Port C, a 50-ohm generator is connected to Port A, and the transmission line sections have a 50-ohm characteristic impedance. Assume also that the internal switches are ideal. Find the transmission coefficient (in dB) over the frequency range from half the design frequency to twice the design frequency. Hint: the closed switches divide the circuit in two so you can ignore the bottom half.

Problem 21.7. Explain the operation of the balanced duplexers shown in Figure 21.10. What restrictions, if any, are there on the lengths of the interconnecting transmission lines?

Problem 21.8. (a) Show that when an uncharged capacitor is brought to potential V by connecting it through a resistor to a voltage source V, the energy supplied by the source is twice the energy deposited in the capacitor (CV^2 rather than $CV^2/2$).

(b) The charging efficiency in (a) is only 50%. Find the efficiency when the capacitor initially has a partial charge, i.e., when the capacitor is initially charged to a voltage αV, where $\alpha < 1$.

Problem 21.9. (a) Find the characteristic impedance of the artificial transmission line shown below. This impedance, Z_0 (which is complex), can be found by adding another LC section to the properly terminated line and noting that the new impedance must still be Z_0.

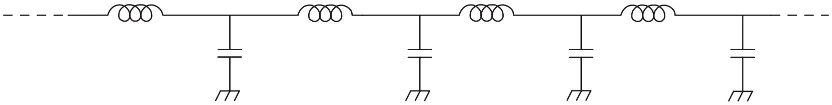

(b) Use the expression for Z_0 to show that the line has a cutoff frequency, $\omega_c = 2/\sqrt{LC}$, above which signals are reflected rather than transmitted.

Problem 21.10. Show that when $\omega \ll \omega_c$, the propagation delay per section for the artificial transmission line of Problem 21.9 is given by $\tau = \sqrt{LC}$.

Problem 21.11. (a) In a *conial scan* tracking radar, the antenna's feed horn is tilted slightly off center and mechanically rotates around the axis of the parabolic reflector. The antenna beam therefore executes a continuous tight conical scan, centered on the target. Many pulses are transmitted during the course of each scan. Draw a block diagram of circuitry to furnish x and y cross-range position error signals to the antenna's drive system in order to keep the antenna centered on the target. (b) Consider a *monopulse* tracking radar in which the antenna is effectively five antennas: one pointed directly at the target, one slightly above, one slightly below, one slightly to the left, and one slightly to the right. Draw a block diagram of circuitry to combine these five signals to provide x and y error signals.

References

[1] North, W., "High-Power Microwave-Tube Transmitters" Los Alamos Nation Laboratory, LA-12687, 1994. Available from U.S. Dept. of Energy, Office of

Scientific and Technical Information, Post Office Box 62, Oak Ridge, Tennessee 37831.

[2] Skolnik, M. I., *Radar Handbook*, 3rd edn. New York: McGraw-Hill, 2008.

[3] Smullin, L. D. and Montgomery, C. G., *Microwave Duplexers*, Rad. Lab. Series Vol. 14, New York: McGraw-Hill, 1948.

[4] Wehner, D. R., *High Resolution Radar*, Boston: Artech House, 1987.

[5] White, J. F., *Microwave Semiconductor Engineering*, New York: Van Nostrand Reinhold, 1982.

CHAPTER

22 Digital modulation techniques

Digital modulation is both the newest and the oldest radio technique. Morse code transmissions were strictly binary, with "key down" and "key up" equivalent to multiplying the carrier by one or zero. Many modern systems also use binary keying, but the zero state is usually signaled by reversing the polarity of the signal (*binary phase-shift keying*, BPSK) or by shifting the frequency (*binary frequency-shift keying*, BFSK). This improves the probability of distinguishing zeros from ones in the presence of noise.

In this chapter we look first at some of the methods used for binary and "*m*-ary" modulation. We then see how specially shaped pulses can be used with these methods in order to avoid intersymbol interference when the pulses, dispersed in time, partially overlap at the receiver. The "8-VSB" system used for digital television in the U.S. (see Chapter 19) provides an example of pulse amplitude modulation (PAM). Finally, we discuss two newer digital modulation systems: multicarrier and spread spectrum. A glossary is provided at the end of this chapter, listing the many common abbreviations used (BPSK, BFSK, 8-VSB, PAM, etc.).

22.1 Digital modulators

Digital modulation differs from analog modulation in that only a discrete set of states (in the space of amplitudes, phases, and frequencies) is used, and that the time devoted to any state is always an integral multiple of a basic time-step. The state during this time period constitutes a transmitted "symbol," and the symbol rate is one of the parameters defining a modulation system. Figure 22.1 shows the simple and widely used binary phase shift keying (BPSK) modulation technique. Instead of turning off the carrier to indicate a zero, the carrier phase is flipped 180°. This is equivalent to multiplying the carrier signal by minus one, as shown in the figure. Binary values of 0 or 1 are handled by sending −1 or +1, respectively, to the multiplier. As shown in Chapter 8, the multiplier (mixer) produces a double-sideband suppressed-carrier RF signal.

Figure 22.1. BPSK modulator.

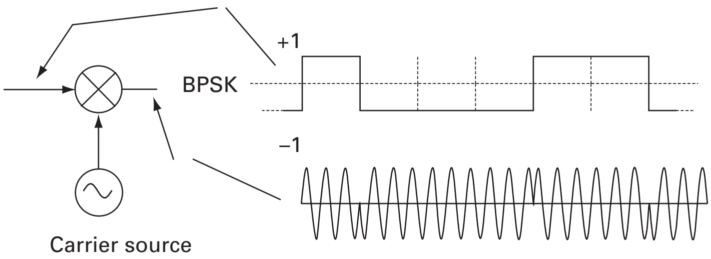

Figure 22.2. *N*-level pulse amplitude modulation (PAM).

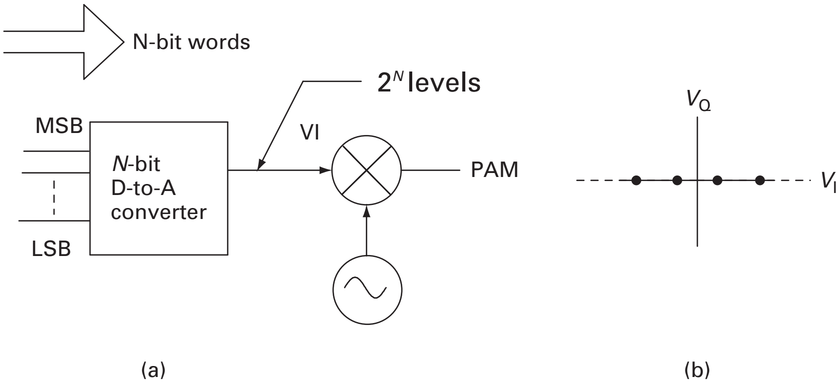

Since the BPSK signal has a nominally constant envelope, the receiver needs a phase-sensitive or "coherent" detector. In contrast, on–off keying (OOK) or frequency shift keying (FSK) can be detected incoherently. The advantage in using coherent systems is one of sensitivity; for a given situation, coherent detection provides a higher signal-to-noise ratio at the detector output. The one-bit BPSK system can be generalized to more than two amplitude levels. We saw in Chapter 21 that the U.S. ATSC digital television system uses eight (2^3) modulation levels in order to send three bits per symbol. Figure 22.2(a) shows an *N*-bit (2^N-level) pulse amplitude (PAM) modulator. The "constellation" shown in (b) shows the signed amplitudes as discrete points on a linear scale. Note that, because the negative amplitudes are produced by changing the phase 180°, PAM requires coherent demodulation.

We saw in Chapter 8 that quadrature AM (QAM) systems can transmit independent information, modulated on a *Q*-carrier which is in quadrature, i.e., 90° out of phase, with respect to the *I* ("in-phase") carrier. This "phase multiplexing" method is commonly used in digital communications. Figure 22.3 shows a "4-QAM" modulator.

The binary (two-level) *I* and *Q* signals result in an *IQ* constellation with four points. In this case, the points on the constellation have four possible phase values, but the amplitudes are equal, since each of the four points lies at an equal distance from the origin. You can see why this scheme is also known as QPSK (quadrature phase shift keying) and 4-PSK (four-state phase-shift keying). Figure 22.4 shows a QAM modulator with four-level modulation on both *I* and *Q* channels, producing a total of 16 states.

The constellation of output signals for this 16-QAM modulator is shown in (b). The 16 binary numbers, 0000 through 1111, could be assigned arbitrarily to the constellation points, but the assignments shown in the figure have the property that nearest neighbors differ by only one bit. As a result, in an environment with only modest noise, errors in transmission will nearly all be single-bit errors. Note that next-nearest neighboring points on the constellation differ by

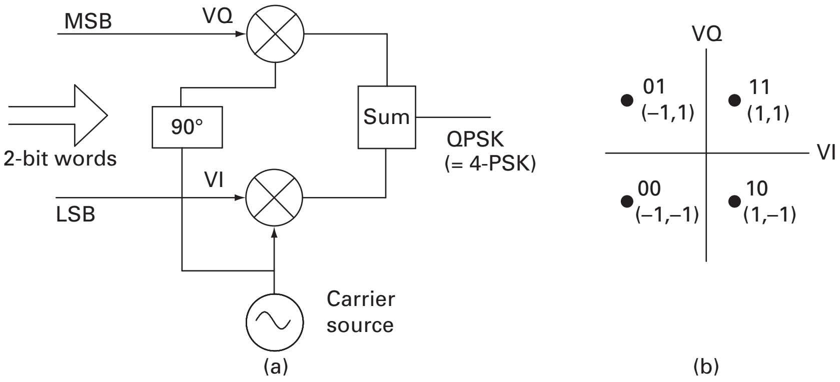
MSB
VQ
2-bit words
90°
Sum
QPSK
(= 4-PSK)
VI
LSB
Carrier
source
(a)
VQ
01
(−1,1)
11
(1,1)
VI
00
(−1,−1)
10
(1,−1)
(b)

Figure 22.3. (a) A 4-QAM (QPSK) digital modulator; (b) QPSK constellation.

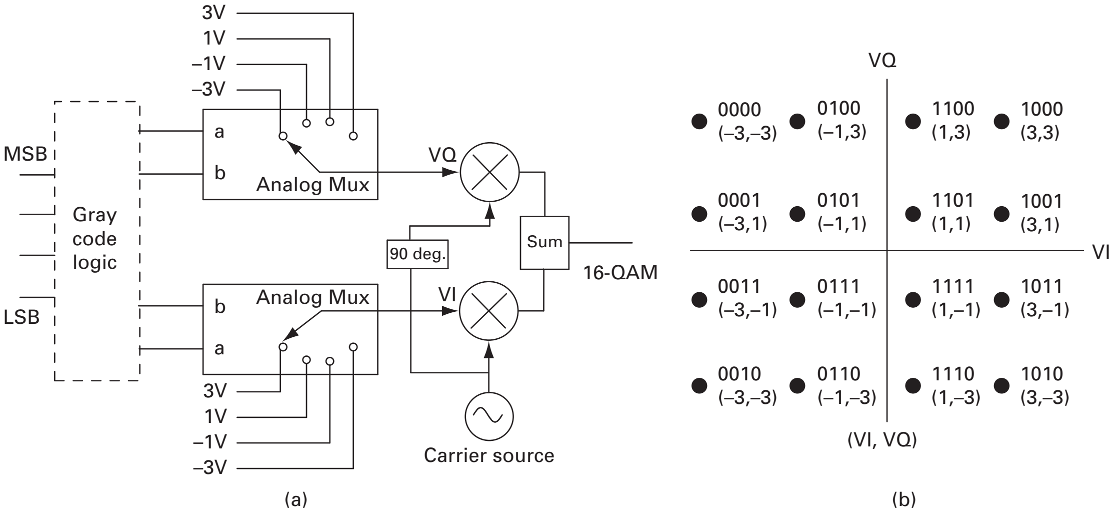
3V
1V
−1V
−3V
MSB
LSB
Gray
code
logic
a
b
Analog Mux
VQ
90 deg.
Sum
16-QAM
Analog Mux
b
a
VI
3V
1V
−1V
−3V
Carrier source
(a)
VQ
0000
(−3,−3)
0100
(−1,3)
1100
(1,3)
1000
(3,3)
0001
(−3,1)
0101
(−1,1)
1101
(1,1)
1001
(3,1)
VI
0011
(−3,−1)
0111
(−1,−1)
1111
(1,−1)
1011
(3,−1)
0010
(−3,−3)
0110
(−1,−3)
1110
(1,−3)
1010
(3,−3)
(VI, VQ)
(b)

Figure 22.4. A 16-QAM digital modulator.

just two bits. QAM systems can also have constellations in which the points are arranged in polar fashion; a constellation with eight points arrayed on a circle in the I-Q plane is known as 8-PSK. Note that BPSK could just as well be called 2-PSK.

22.2 Pulse shaping

The abrupt reversals in the BPSK waveform of Figure 22.1 result in a signal with an objectionably large bandwidth. The signal is a continuous stream of rectangular RF pulses whose polarity is determined by the data. If the data bits are random, with equal probabilities of being a one or a zero, these rectangular pulses produce a power spectrum proportional to $[\sin(\pi f T)/(\pi f T)]^2$, where T is the duration of a symbol, the "baud" length. This spectrum is plotted in Figure 22.5.[1] On this graph, frequency is in units of $1/T$, and the zero is at the nominal carrier frequency.

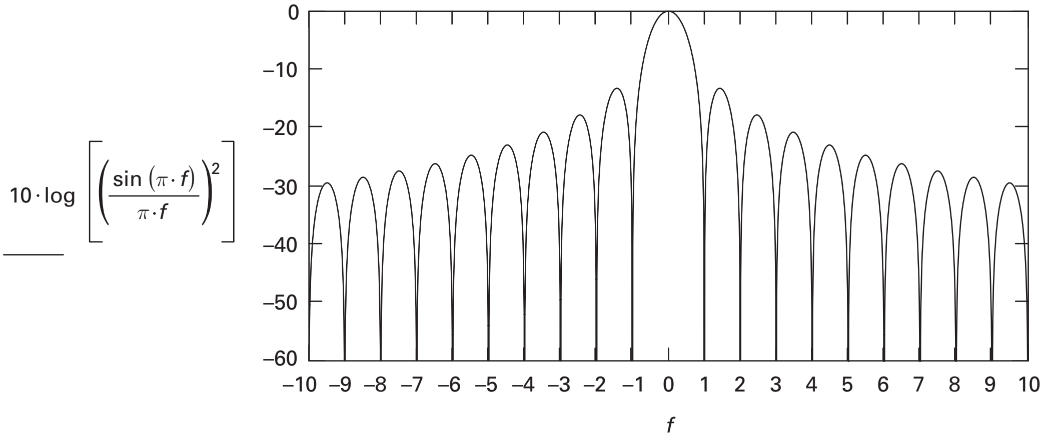

Figure 22.5. Power spectrum of a BPSK signal with random data and rectangular pulses. Frequency is in units of $1/T$, and the zero is at the nominal carrier frequency.

This wide spectrum applies to all the digital modulators discussed above, and is clearly unacceptable if other users are to use nearby frequency channels. Bandpass filtering at the transmitter, or otherwise transmitting a pulse shape whose spectrum is band-limited, solves this problem, but there is a complication – limiting the bandwith causes a widening of each pulse in the sequence so that, at the receiver, a currently arriving pulse will be contaminated by remnants or precursors of nearby pulses. This effect is known as *intersymbol interference* (ISI). However, there exists a class of pulse shapes, *Nyquist pulses*, for which (a) the bandwidth is limited and (b) there is no ISI, despite the fact that the pulses do spread into each other. One such pulse shape is $V(t) = \sin(\pi t/T)/(\pi t/T) = \mathrm{sinc}(t/T)$, the normalized sinc function. This pulse shape is the impulse response of a

---

[1] The power spectrum is calculated from the signal's average autocorrelation function (see Chapter 27).

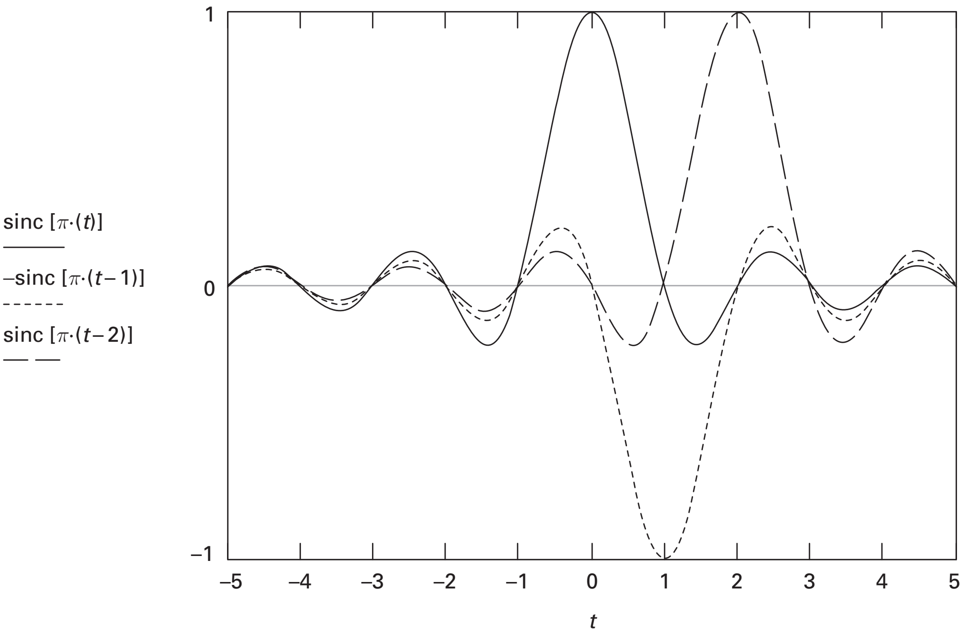

Figure 22.6. Successive $\sin(\pi t)/(\pi t)$ signaling pulses.

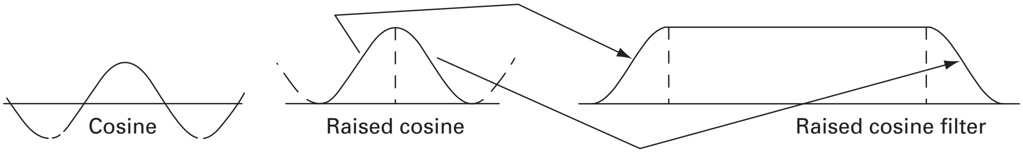

Figure 22.7. Construction of the raised cosine filter; $H(\omega)$ is the voltage response.

rectangular lowpass filter with a cutoff frequency of 1/(2T)Hz. This filter confines all the power to a bandwidth of 1/(2T). Successive output pulses for a 101 data string are plotted in Figure 22.6. The third pulse reaches a maximum at $t = 2$, the instant at which we sample it. Neither of the two previous pulses has yet died out but, at $t = 2$, both are crossing zero and thus contribute no interference to the sampled value of the desired pulse.

The same holds true for the contributions from all the other pulses, of any amplitude and either polarity, allowing us to transmit data pulses at a rate 1/T without intersymbol interference.

Pulses with this property are said to be *orthogonal* to one another. While they do invade each other's space, they do not create intersymbol interference. The $\sin(\pi t/T)/(\pi t/T)$ pulse is an extreme case; bandwidth is minimized, but the pulse is widely spread out in time. In practice, other orthogonal pulse shapes are chosen, which have somewhat greater bandwidth but are only modestly spread in time. One frequently used pulse shape is the impulse response of the *raised-cosine* filter. Figure 22.7 shows how this filter shape is produced by replacing the vertical sides of the rectangular filter with smooth roll-offs formed from quarter cycles of the cosine function, raised by 1/2 in the y-direction.

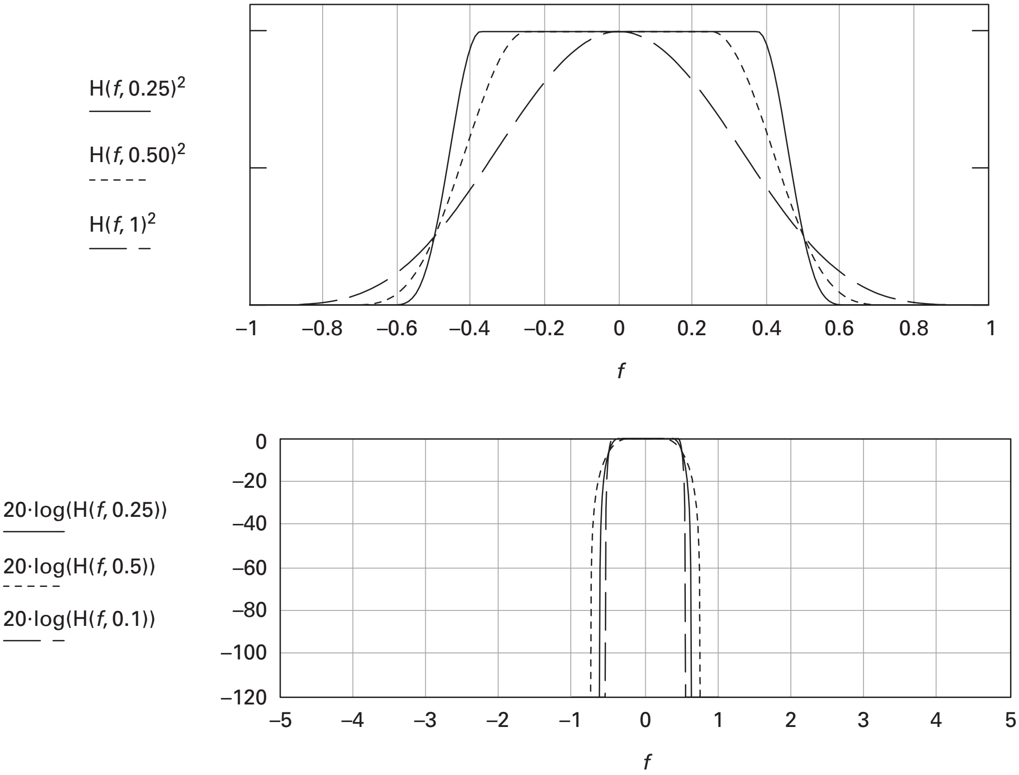

Figure 22.8. Power response, |*H* (*ω*)|2, of raised cosine filter for *α* = 0.25, 0.5, and 1.0.

We can express this filter shape as

$$H(f,\alpha)=\begin{cases} T: & 2|f|T\le 1-\alpha \\ \dfrac{T}{2}\left[1+\cos\left(\dfrac{\pi T}{\alpha}\left(|f|-\dfrac{1-\alpha}{2T}\right)\right)\right]: & 1-\alpha\le 2|f|T\le 1+\alpha \\ 0: & 2|f|T1+\alpha, \end{cases} \tag{22.1}$$

where α, a roll-off factor between zero and one, determines the width of the raised cosine edge sections. Note that the $\alpha = 0$ case is just the rectangular filter discussed above. This filter's power response, $|H(f,\alpha)|^2$, for $\alpha = 0.25$, $\alpha = 0.5$ and $\alpha = 1$ (pure raised cosine with no flat center section) is plotted in Figure 22.8(a) with a linear power scale and Figure 22.8(b) with a dB power scale.

The impulse response of this filter, i.e., the Fourier transform of $H(f,\alpha)$, often called by extension a "raised-cosine pulse," is given by

$$h(t,\alpha)=\frac{\sin(\pi t/T)}{\pi t/T}\cdot\frac{\cos(\pi\alpha t/T)}{1-(2\alpha t/T)^2}. \tag{22.2}$$

The zeros of the first term, $\sin(\pi t/T)\,/(\pi t/T)$, give this pulse shape the same orthogonality property as $\sin(\pi t/T)\,/(\pi t/T)$ alone. Figure 22.9 shows three

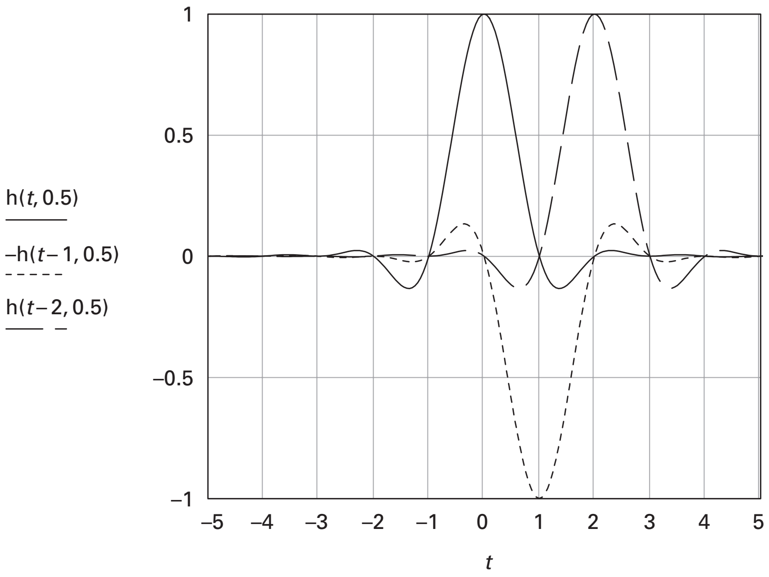

Figure 22.9. Successive raised-cosine signaling pulses.

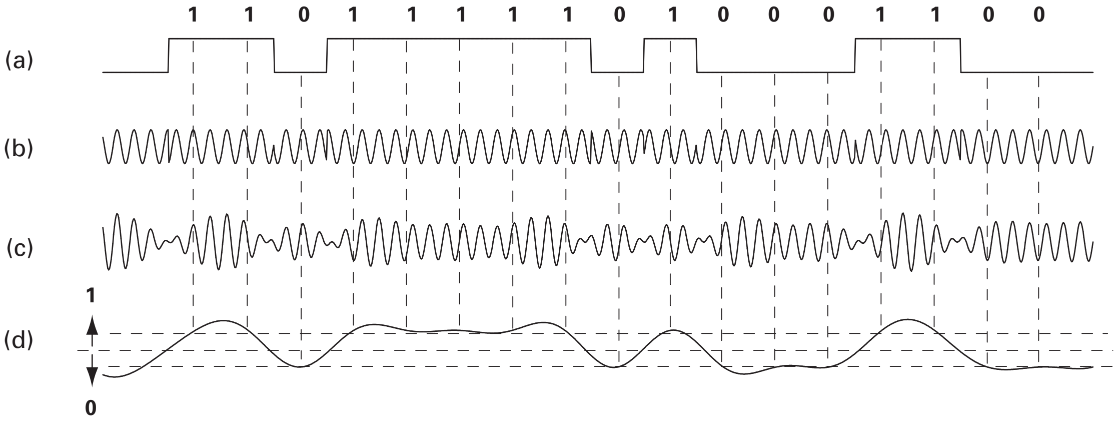

Figure 22.10. BPSK waveforms: (a) input data; (b) bi-phase modulated carrier; (c) modulated signal after raised-cosine filtering ($\alpha = 0.22$); (d) coherently detected signal.

consecutive $\alpha = 0.5$ raised cosine pulses. Note that these pulses, like the sin $(\pi t/T)/(\pi t/T)$ pulses of Figure 22.6, produce no intersymbol interference, but unlike those pulses, these are only modestly spread out in time.

Despite the bandwidth advantage, there are two penalties associated with choosing a low value for α. First, the impulse response duration of the filter is longer, requiring more computations per symbol when the filter is implemented digitally. Second, the increased ringing means that the received signal must be sampled with higher timing accuracy to avoid ISI.

The successive processes of BPSK modulation, pulse shaping, and detection are shown in Figure 22.10. The data string to be transmitted is shown in (a). In (b), the sinusoidal carrier has been multiplied by the data, producing phase

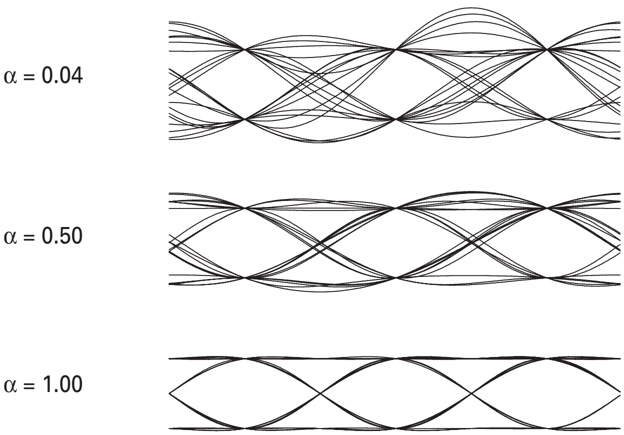

Figure 22.11. BPSK eye diagram.

reversals. In (c), the modulated signal has been band-limited by a raised-cosine filter of $\alpha = 0.22$. In this simulation the modulation rate is high, causing obvious variations in the envelope of the filtered signal. Finally, in (d), the filtered signal has been coherently detected, i.e., multiplied by a replica of the original sine wave carrier and then lowpass filtered to remove the components at twice the carrier frequency. Note that, at the sampling instants, the sampled values are exactly the same 0's and 1's of the original data, i.e., there is no intersymbol interference. Note: for clarity in alignment with the original data, no filter delay has been added to the waveforms in (c) and (d), so they appear noncausal.

If the detected data is displayed versus time for many sweeps on a storage oscilloscope that is synchronized to the data, we obtain the well-known "eye" diagram, shown in Figure 22.11. You can see that, at the sampling instants, the value of the waveform is either zero or one; there is no intersymbol interference. Eye diagrams are presented for three values of α. You can see that, for smaller values of α (narrower bandwidths), the sampling instants must be increasingly accurate to obtain only the values zero and one, whereas larger values of α make the system more tolerant of sampling "jitter."

22.3 Root raised-cosine filter

Any receiver must contain a bandpass filter to reject unwanted signals (i.e., signals as well as noise on nearby frequencies). If the receiver filter has a rectangular shape whose bandwidth is equal to or greater than the signal bandwidth, Nyquist pulses will pass through undistorted, maintaining their orthogonality, i.e., their freedom from intersymbol interference. However, a basic result from signal processing theory (see Chapter 23) is that the overall signal-to-noise ratio of a communication link is maximized when the receiver uses a *matched* *filter*, which is a filter whose impulse response is proportional to the time-reversed waveform of the incoming signal pulse. Let us apply this consideration to the two pulse shapes discussed above. For the $\sin(\pi t/T)/(\pi t/T)$ signaling pulse, the matched filter is just the rectangular filter which might have been used to produce the pulse. The cascade of two rectangular filters is equivalent to a single rectangular filter, and the pulse shape is preserved. For a raised-cosine signaling pulse, however, the matched filter is the raised-cosine filter of Equation 22.1. If we put this filter at the input of the receiver, the overall response will be the *square* of the raised-cosine shape. The signaling pulses will take on the shape of a raised-cosine squared filter – a shape that is not free of intersymbol interference. To avoid this, it is common to transmit "root raised-cosine" signaling pulses, i.e., pulses whose shape is the impulse response of a filter whose amplitude response is the square root of the raised-cosine filter. A root raised-cosine filter is used at the input of the receiver as a matched filter. The pulses exiting the receiver filter will therefore have the "raised-cosine pulse" shape shown in Figure 22.9, and there will be no intersymbol interference. The frequency response of the root raised-cosine filter is given by Figure 22.8(b), if you relabel the y-scale to go from 0 to – 60 dB instead of 0 to −120 dB.

Let us summarize this section. So-called Nyquist pulses can be transmitted without intersymbol interference up to a maximum rate of B pulses per second using double-sideband AM modulation within an RF bandwidth of at least B Hz. A second carrier, at the same frequency, but phase shifted by 90°, can be used simultaneously to transmit independent pulses. When the pulses are identical except for 0° or 180° phase changes, the system is called BPSK (one carrier only) or QPSK (pulses on both I and Q carriers). Synchronous detection, i.e., phase-sensitive detection, makes this 0° and 180° phase modulation equivalent to amplitude modulation in which the sign of the signal is preserved. If the I and Q components of a QPSK signal are each allowed to take on four values, the modulation forms a 4×4 constellation and is referred to as 16-QAM. (It would be convenient if the term "AM" always referred to this kind of sign and magnitude modulation. BPSK could be called "sign modulation." Traditional AM could be called "magnitude modulation".)

Note that in the above analysis we have not needed to bother with including a phase factor in the filter amplitude response functions. These particular filter functions are purely real, except for this omitted factor, $e^{-j\omega T_{\text{delay}}}$, where T_{delay} is a frequency-independent time delay. Neglecting this factor is equivalent to setting the delay equal to zero, which causes the impulse responses to begin at non-causal negative times.

22.4 8-VSB and GMSK modulation

The U.S. digital television system, discussed in Chapter 19, modulates an I-carrier with three-bit digital data in the form of eight evenly spaced amplitude levels whose nominal values are 7, 5, 3, 1, −1, −3, −5, and −7. A bias of 1.5 is added to these levels to create a small pilot carrier (a phase reference for the receiver) resulting in five transmitted levels that are positive and three that are negative. This is a PAM system (see Figure 22.2). The resulting double-sideband suppressed carrier signal is bandpass filtered to eliminate all but a vestige of the lower sideband, reducing the bandwidth by factor of almost 2. This modulation is called 8-VSB (eight levels, vestigial sideband). Because the signal had no Q-component before filtering, the lower sideband was the mirror image of the upper sideband and contained redundant information. With a general QAM signal (using both I and Q), the lower and upper sidebands are different and both are needed. But, as a result of filtering away the lower sideband, a transmitted VSB signal acquires a Q component, making the "Q-space" unavailable for the transmission of a second television signal.

The modulation method used in GSM cellular telephones is known as *GMSK* (*Gaussian minimum shift keying*), and works as follows. A standard QPSK modulator is used, summing the signal from an I mixer with the signal from a Q mixer. But the input signals to these mixers are such as to cause the sum signal to rotate smoothly from one point on the constellation to an adjacent point during the baud time. The rotation is 90°, clockwise or counterclockwise, depending on the binary data bit. The rotation takes place at a uniform rate. Since frequency is the time derivative of phase, this modulation is actually BFSK. In addition, smoothing, derived from a Gaussian filter shape, is applied to the phase command, so that the time derivative of the frequency is continuous during the transitions between the two nominal frequencies. This system is designed to minimize bandwidth, allowing closer channel spacing to accommodate more users. Bandwidth reduction comes from using both I and Q, from changing the phase slowly and smoothly, and from allowing some intersymbol interference (which is almost all from the immediately preceding baud and can be compensated for by the receiver).

22.5 Demodulation

Synchronous demodulation is required for the BPSK and QAM systems discussed above. This requires that the receiver be able to synthesize a replica of the carrier. This carrier recovery, which takes place in the presence of data, can be done, for example, using a squaring PLL (see Chapter 12). A standard PLL suffices if a pilot carrier is transmitted. The receiver contains a matched filter, followed by a mixer whose L.O. signal is the recovered carrier. The baseband output of the mixer is sampled at the signal baud rate. The signal may include a periodic synchronizing pattern to assist the receiver in establishing the correct phase for the sampler. A digital matched filter is often used. When realized as a weighted sum of signals from a tapped delay line (a transversal FIR filter), such a filter illustrates how a matched filter is equivalent to a correlator, finding the best alignment (match) of the signal to a replica of the signaling pulse.

Sometimes a free-running L.O. will be close enough to the carrier frequency that the dephasing over a data word or packet is significantly less than $\pi/2$. In this situation, differential modulation can be used whereby, for example, a “one” is signaled by flipping the phase of a BPSK signal, while a “zero” is signaled by leaving the phase at its prior value. If both I and Q channels are implemented, the phase changes can be reliably sensed and the data can be reconstructed. In a QPSK system, both I and Q channels are always implemented and the receiver can sense differential changes between points on the modulation constellation. Differential modulation, however, results in a higher error rate than coherent demodulation.

22.6 Orthogonal frequency-division multiplexing – OFDM

OFDM is a relatively new modulation technique in which the digital data stream to be transmitted is divided into N separate parallel data substreams, each having a data rate that is $1/N$ times the data rate of the original stream. Each substream independently modulates a separate RF carrier. The modulation scheme can be any of the schemes described above, except for FSK and MSK, where the phase does not remain constant throughout the duration of a symbol. At the receiver, the N substreams are demodulated in parallel and then combined to produce the original data stream. Applications of OFDM include 802.11*a* , *g*, and *n* Wi-Fi modems, DAB (digital audio broadcasting) in Europe and Canada, Digital Radio Mondiale short-wave broadcasting, HD (hybrid digital) radio broadcasting on the U.S. AM and FM bands, and European DVB-T (terrestrial digital television broadcasting). The number of carriers used in these systems varies from tens to thousands.

22.6.1 Advantages of OFDM

The system is bandwidth neutral, occupying about the same bandwidth as the traditional modulation systems it replaces. Its main advantage comes from its ability to deal with multipath propagation and channel equalization. Consider the problem of equalization. The propagation channel for a digital signal should have a specific amplitude and phase response, such as the raised-cosine shape discussed above. There are several propagation effects that can distort the shape. Receiving antennas are not always flat in frequency and linear in phase, so they can modify the channel response. Constructive and destructive interference of multipath components are frequency dependent and modify the channel response. Television receivers for 8-VSB use time correlation to detect multipath components and then subtract or realign them. This process, while essential, is not simple, and is particularly difficult in a mobile environment where the multipath situation is constantly changing. At fast baud rates, new multipath components are still arriving at the time the baud ends. With the slow data rates on the individual OFDM carriers, the baud can extend far beyond the time needed for all the multipath components to arrive. Dispersion is also a classic problem in wired data links.[2] Fast DSL (*digital subscriber line*) data links over ordinary telephone circuits have to deal with cable dispersion, poorly controlled impedances, and multipath echos caused by haphazard terminations and discontinuities. An early application of OFDM was in "discrete multi-tone" DSL modems of the early 1990s. While the principles of OFDM date back to the 1960s, practical implementations had to await the advent of fast digital signal processors. Most OFDM system incorporate coding and decoding for forward error correction and are known as COFDM (coded OFDM). As in digital TV transmission, the coding usually encloses the inner modulation/demodulation "physical layer."

22.6.2 Single-frequency broadcasting networks

The essentially perfect ability of OFDM to deal with multipath propagation can be exploited in single-frequency networks, in which broadcasting transmitters, all on the same frequency, serve overlapping areas of coverage. When the identical transmitted signals are synchronized in time and frequency (which can be done using GPS satellite signals), the signals arriving at any given site are indistinguishable from multipath signals originating from a single transmitter. At any given site, the receiving antenna is aimed at the strongest signal. The weaker signals, depending on their phase, increase or decrease the signal strength. They also shift the phases of the recovered Fourier coefficients, but this is not a problem because OFDM channels can be individually equalized via an occasional "training" sequence sent by the transmitter or by using a set of pilot carriers. Single-frequency networks allow efficient use of the spectrum, since the same television or radio channels can be used in adjacent areas – something like solving the classic map coloring problem with a single color. It is assumed that 8-VSB broadcasting could also use single-frequency networking, given echo cancellers of sufficiently high performance. Same-frequency fill-in or "booster" stations with well-synchronized frequencies have been used occasionally for conventional analog AM, FM, and television, but, without echo cancellation, they produce video echos ("ghosts").

---

[2] The earliest trans-Atlantic telegraph cable, over its great length, accumulated so much dispersion that baud rate had to be slowed down to where the received signal was traced out by a chart recorder. Likewise, early overland telegraph lines were too dispersive to support baseband telephony. The first solution, series loading coils, was a product of the then recent theory of waves on transmission lines, which showed that the reactance per unit length had to be increased, relative to the resistance per unit length.

22.6.3 Implementation of OFDM

The key to practical implementation of OFDM is the discrete Fourier transform (DFT) and its inverse (IDFT), which, in practice are implemented using FFT algorithms. At the transmitter, the data is arranged as a stream of data blocks, each containing N pairs of digital numbers. Each pair of numbers can be considered as the real and imaginary parts of a complex number. Let the complex data numbers in a given block be denoted as D_n, for $n = 0$ to $N-1$.

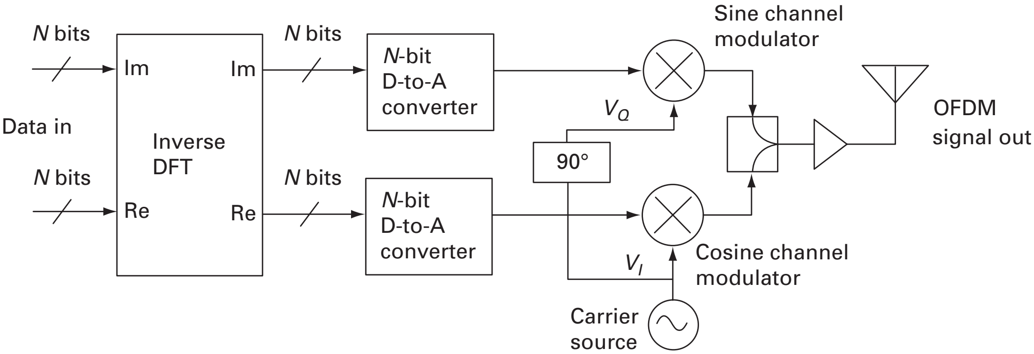
N bits
Im
Im
N bits
N-bit D-to-A converter
Sine channel modulator
Data in
Inverse DFT
V_Q
OFDM signal out
90°
N bits
Re
Re
N bits
N-bit D-to-A converter
V_I
Cosine channel modulator
Carrier source

Figure 22.12. OFDM transmitter block diagram.

Figure 22.12, a block diagram for an OFDM transmitter, shows that these numbers are first inverse Fourier-transformed as follows:

$$V_m = \sum_{n=0}^{N-1} D_n e^{j2\pi mn/N}. \tag{22.3}$$

The resulting set of N complex numbers, V_m, for $m = 0$ to $N - 1$, is read out serially at a rate N/T, where T must be no greater than the spacing between blocks of input data. If V_0 appears at $t = 0$, we can express the output numbers as a function of time:

$$V(t_m) = \sum_{n=0}^{N-1} D_n e^{j2\pi n t_m/T} = \sum_{n=0}^{N-1} D_n e^{j\omega_n t_m}, \tag{22.4}$$

where ω_n is $2\pi n/T$. Note that this is just the sum of N carriers at equally spaced frequencies, ω_n, and that each carrier is multiplied (amplitude modulated) by a coefficient, D_n.

This set of N modulated carriers is converted from baseband to a QAM signal at RF by using a “cosine mixer” for the real part and a “sine mixer” for the imaginary part. Of course the real and imaginary outputs of the IDFT must be converted to analog voltages before they are applied to the mixers. These (suppressed) carriers are often called subcarriers, although they are not part of the sideband structure of some main carrier. Figure 22.13 shows an OFDM

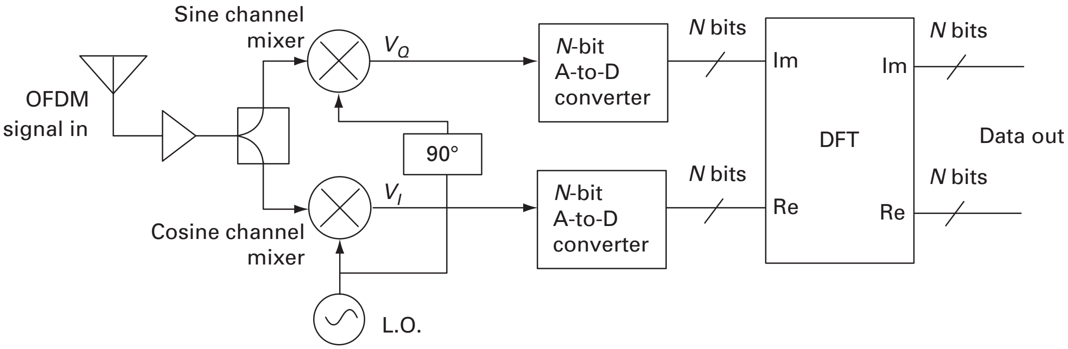

Figure 22.13. OFDM receiver block diagram.

receiver. Note that it is a mirror image of the transmitter. The operations done by the transmitter are undone in the reverse order by the receiver. I and Q samples from the cosine and sine mixers are digitized and combined to from the sequence of complex numbers, V_m, which is Fourier transformed to recover the date (the complex amplitudes of the subcarriers),

$$\frac{1}{N}\sum_{m=0}^{N-1} V_m\, \mathrm{e}^{-\mathrm{j}2\pi nm/N} = D_n, \qquad (22.5)$$

and we see that the data has indeed been recovered from the separate carriers. Thus the inverse Fourier transform effectively acts as N carrier oscillators and N mixers to create N modulated carriers while the Fourier transform at the receiver acts as N L.O.'s and N mixers to detect the (complex) amplitude of each carrier. During the interval T, the n-th carrier has constant amplitude; it is an RF pulse with a rectangular envelope. The Fourier transform of this pulse has the shape $\sin(\Delta\omega T/2)/(\Delta\omega T/2)$, where $\Delta\omega$ is the offset from ω_n. The shape plotted in Figure 22.5. We see that the spectrum for any one of the carriers has considerable overlap with the spectra of nearby carriers. Nevertheless, at the nominal frequency of any one of the carriers, the power spectrum of any other carrier is crossing through zero. For this reason, the carriers are said to be orthogonal in the frequency domain, analogous to Nyquist pulses in the time domain, whose maximum value occurs at an instant when the voltage from any earlier pulse is crossing zero.

22.6.4 OFDM guard interval

The key to dealing with multipath in OFDM is the use of a guard interval after every T-second transmission of the group of multicarrier signals. To see that this is necessary, consider the situation shown in Figure 22.14, where no guard interval is used. In the figure, the data is advancing in time toward the right. Ideally we would receive N signal samples from a

Figure 22.14. OFDM without a guard interval.

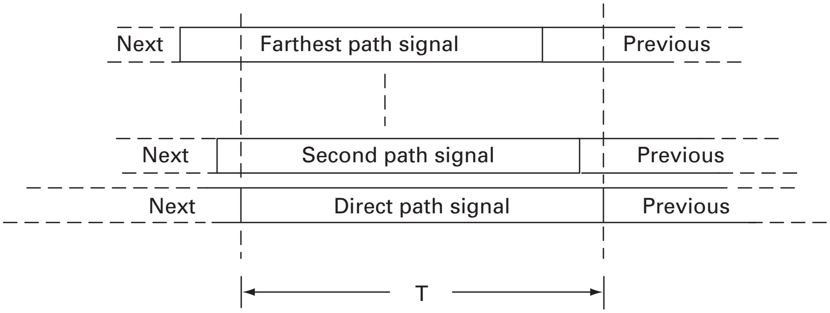
Next
Farthest path signal
Previous
Next
Second path signal
Previous
Next
Direct path signal
Previous
T

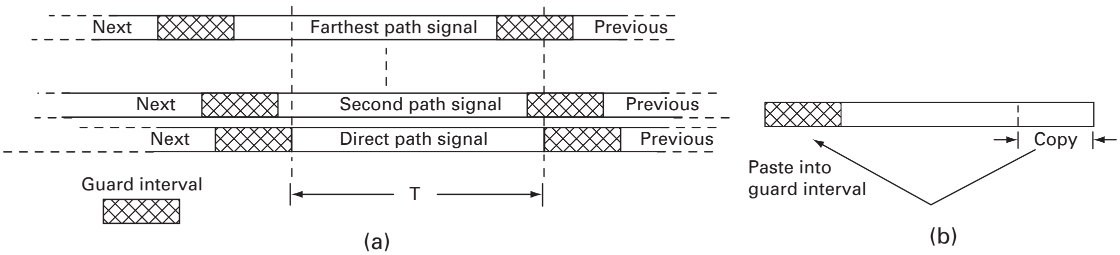
Next
Farthest path signal
Previous
Next
Second path signal
Previous
Next
Direct path signal
Previous
Guard interval
T
(a)
Copy
Paste into guard interval
(b)

Figure 22.15. OFDM with a guard interval.

direct signal during the time interval 0 to t and then apply the Fourier transform to this set of numbers. But signals arriving on paths other than the direct path will arrive somewhat later, as shown in the figure, where the signal from the farthest path arrives late by about $0.2T$. A set of voltage samples taken from $t=0$ to T will be the sum of the signal voltages from the various paths so the first 20% of the samples will be contaminated by data from the previous signal interval, arriving via the farthest path and the first 5% or so will be contaminated by the previous signal arriving over all the delayed paths.

Now let us add a guard interval of length αT, as shown in Figure 22.15(a). The contributions from the delayed paths during the interval $t=0$ to T will now be from whatever has been placed in the guard intervals. You might think of zeros as a possible choice; at least they contain no data from the previous signal. However, they would compromise the orthogonality property. A much better scheme is to extend the data cyclically into the guard intervals, as shown in Figure 22.15(b). Data from the first part of the sequence is pasted into the guard interval so that when the guard interval advances into the sampler, the data sequence for the data interval begins over again. The result is that each multipath signal is properly demodulated over a complete interval and orthogonality is maintained. The overall system works as follows. Blocks of $2N$ real numbers are considered as blocks of N complex numbers. Each arriving block is subjected to an inverse DFT, whose output is also a block of N complex numbers. These output blocks are cyclically extended from length N to length $N(1+\alpha)$.

The numbers are converted to analog voltages and fed to a QAM modulator at a rate $N(1+\alpha)/T(1+\alpha) = N/T$. The receiver takes a burst of N samples every $T(1+\alpha)$ seconds, nominally aligned with the frame of data arriving over the direct path. Sampling bursts are separated by a time interval αT. The frames of N samples are Fourier-transformed to yield the complex amplitude of each carrier. This set of amplitudes is the original data. Many methods are possible to get the proper framing at the receiver and to regenerate the carrier for coherent detection. DAB audio broadcasting uses differential PSK modulation in order to tolerate carrier phase errors. DVB-T television, however, uses PSK with coherent detection.

22.6.5 Disadvantages of OFDM

One disadvantage of OFDM is that the signal has a high dynamic range, requiring that the transmitter have excellent linearity. (The same is true for any transmitter that transmits the sum of many independent signals, e.g., a cellular telephone base station transmitter.) Another potential problem is use in a mobile environment where different multipath signal components will generally have different Doppler shifts, as when the receiver is moving away from the transmitter but is moving toward a reflective object (maybe a building or tower) which provides a strong additional signal. The spacing between OFDM subcarriers would have to be considerably greater than the Doppler spread of the multipath signals. But, if there are appreciable differences in the time delays of the multipath components, the carriers must be closely spaced. This conflict is less severe at low radio frequencies, since Doppler shifts are proportional to frequency.

22.7 Spread-spectrum and CDMA

Best known now for its application in CDMA cellular telephony, spread spectrum was originally developed to provide secure communications links by using signals spread out in frequency and, therefore, reduced in spectral amplitude below the noise level of common intercept receivers. Besides being hard to detect, the signals cannot be demodulated without finding the long code word used in the spreading process. In addition, spreading makes the signals hard to jam.

22.7.1 Direct sequence spreading

Figure 22.16 shows how a rapidly changing binary pseudorandom code sequence multiplies a data stream to produce a signal whose bandwidth is determined by the "chip"[3] length of the code, rather than by the relatively slow data baud.

---

[3] Code elements, unlike signal elements, carry no information and are called "chips" rather than "bauds."

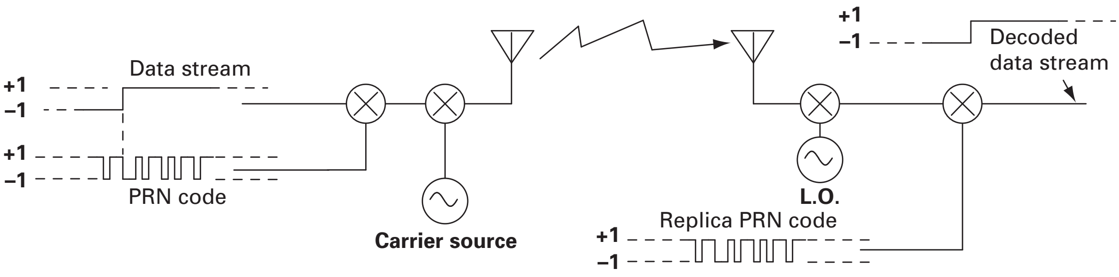

Figure 22.16. Pseudonoise spread spectrum.

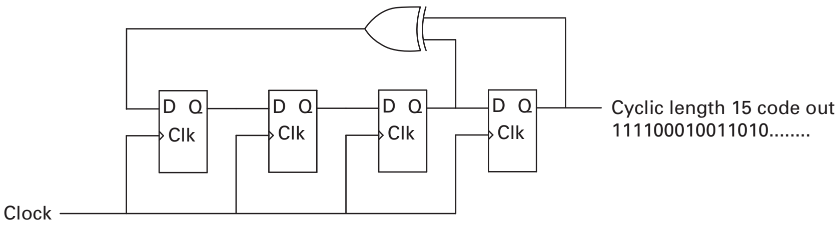

Figure 22.17. Four-stage maximum length shift register code generator.

This system is known as a *direct-sequence* spread spectrum (DSSS). We saw the same circuitry in Chapter 21, where its purpose was to produce a wide radar pulse with modest power that could be decoded (decompressed) to form a narrow high-power pulse. The purpose there was to get around the peak power limitation of the transmitter, but the system also provides the radar with a signal that is harder to intercept and jam. We will encounter the circuit again in Chapter 25, where its use in GPS satellites provides the high range resolution of pulse compression radar as well as multi-access – allowing all the satellites to transmit on the same frequency.

The pseudonoise or "direct sequence" spreading used in Figure 22.16 often uses *maximum length shift register sequences*, which are pseudorandom codes of length $2^N - 1$ generated by a circuit made of an N-stage shift register and a two (or more)-input modulus 2 adder. Figure 22.17 shows a four-stage generator and its 15-chip sequence. For this four-stage generator, only two feedback taps are needed, so the modulus-2 adder is a simple exclusive-OR gate. You can see that, for this kind of code generator, the shift register will never contain all zeros, as that state would then repeat endlessly. All other states do occur, so the length of the sequence is 2^N-1. A sequence of length $2^{32}-1$ can be generated by a 32-stage shift register with taps at stages 32, 31, 30, and 10. These taps can be added with a cascade of three ex-OR gates. With a 1-MHz clock, this generator would run for more than an hour before the sequence starts to repeat. For a given value of N, there are usually several sets of taps, each of which will yield a different maximum length sequence. The key property of the maximum length sequences is that they are noise-like. There is essentially zero correlation between different codes or between a code and a time-shifted version of itself.

The power spectrum of the signal transmitted by this system has a $[\sin(\pi f T)/(\pi f T)]^2$ shape, where ω is the offset from the carrier frequency and T is the chip length. Thus, the bandwidth is about $1/T$, independent of the code length. Longer codes, however, provide better security against interception. This system uses coherent detection; there must be provision for the receiver L.O. to acquire the correct phase (and, therefore, frequency) and for the replica code generator to acquire the phase for alignment with the incoming signal.

Wide frequency spreading requires a short chip time which, in turn, requires a high-precision time alignment of the replica code. Another spreading technique, *frequency hopping*, requires less alignment precision.

22.7.2 Frequency-hopping spread spectrum (FHSS)

In this system, the transmitter carrier frequency “hops” pseudorandomly from one frequency to another throughout the band. The timing of the hops is usually regular, though it, too, could be pseudorandom. An advantage of the hopping scheme is that the bandwidth is determined primarily by the assignment of hopping frequencies and not by the chip time – the time between hops. A convenient scheme is to hop to a new frequency for each data baud. These systems usually use incoherent detection since, until the advent of the direct digital synthesizer (DDS), it was difficult to achieve phase-coherent changes in frequency. One modulation scheme uses frequency shift keying. At the transmitter, the FSK signal is up-converted to the RF band by a frequency hopping carrier generator. At the receiver, a synchronized hopping L.O. converts the signal down to a constant IF frequency. The FSK IF signal can then be demodulated by an ordinary incoherent FM discriminator.

22.7.3 CDMA in cellular telephone systems

By itself, a direct sequence spread spectrum link is as power-efficient as a narrowband link, but deliberately has low spectral efficiency (data rate/RF bandwidth). However, the low cross-correlation property of the codes makes it possible for many users to have spread spectrum links operating in the same band. To each user, the signals from the others appear as random noise. This is known as *code-division multiple access* (CDMA). One advantage of CDMA is that no synchronization is required between users, in contrast to conventional time division and frequency multiplex systems where users must occupy scheduled frequencies and time slots. Another advantage is that the CDMA system quality degrades gracefully; as users are added, each user sees the background noise increase somewhat and experiences a lower signal-to-noise ratio. In time division and frequency division multiplexing, the maximum number of users is strictly limited by the number of frequency and time slots. For the same reason, in a CDMA cellular telephone system, adjacent cells can use the same band. Signals from adjacent cells will be, on average, weaker, and contribute only modestly to the background noise. Moreover, since a mobile user moving from one cell to another stays on the same frequency, there is a "soft handoff" from one base station to the next, rather than a momentary disconnection while new time or frequency slots are allocated.

Problems

Problem 22.1. Show that the impulse response of the filter function of Equation (22.1) is indeed the pulse shape of Equation (22.2).

Problem 22.2. Show that the filter whose frequency response is the *square* of that of the raised-cosine filter will produce intersymbol interference, i.e., show that such a filter, when excited by delta function at $t = 0$, will have an impulse response that is not identically zero at $t = T$.

Problem 22.3. Consider the simple single-frequency "network" for traditional AM broadcasting in which a single "booster" transmitter site is located at the fringe of the coverage area of the main transmitter. Describe what a listener, midway between the sites, could experience if (a) the carrier frequencies differ by 400 Hz; (b) the carrier frequencies differ by 0.1 Hz; (c) the carriers are locked to the same frequency.

Problem 22.4. Write a program to simulate the pseudonoise code generator of Figure 22.15 but use 10 stages, rather than four. One input to the ex-OR gate is the output of the tenth stage. Find which stage should be connected to the other input of this gate to produce a sequence of length 1023.

Hint: Start with the shift register in the all-ones state. Count the clock pulses needed to bring it back to this state. Do this for different taps until you find one that requires 1023 pulses.

Glossary

ATSC	Advanced Television Systems Committee: developed the ATV standard for US digital television broadcasting.
BER	Bit Error Rate: probability that a received "one" is deemed a zero or vice versa
BFSK	Binary Frequency-Shift Keying.
BPSK	Binary Phase-Shift Keying.

CDMA	Code Division Multiple Access: spread spectrum modulation that allows multiple users to share the same nominal frequency.
COFDM	Coded Orthogonal Frequency Division Multiplexing: multicarrier modulation system suited to multipath propagation environments. Includes coding for forward error correction.
DAB	Digital Audio Broadcasting standard used in Europe and Canada.
DDS	Direct Digital Synthesizer: frequency synthesizer with phase-continuous frequency transitions (see Chapter 12).
DFT	Discrete Fourier Transform, equivalent to a matrix multiplication that converts a set of N complex numbers to another set of N complex numbers.
DSL	Digital Subscriber Line: Telephone company high-speed digital service
DTV	Digital Television
DSSS	Direct Sequence Spread Spectrum: frequency spreading done via multiplication with pseudorandom ±1 sequence.
DVB-T	Digital Video Broadcasting – Terrestrial: European over-the-air digital television.
FEC	Forward Error Correction
FDM	Frequency Division Multiplexing: users are assigned different frequencies
FHSS	Frequency Hopping Spread Spectrum: transmitted signal "hops" rapidly between pseudorandomly chosen frequencies.
FIR	Finite Impulse Response: FIR filters are often based on a weighted sum of the signals from a tapped delay line.
GMSK	Gaussian Minimum Shift Keying: a particular frequency shift keying with smooth transitions between ones and zeros.
GPS	Global Positioning System: satellite navigation system.
GSM	Global System for Mobile Communications: the dominant cellular phone system.
HD radio	Hybrid Digital radio: a transition U.S. radio broadcasting standard in which COFDM digital signals share frequency channel assignments with AM and FM signals.
IDFT	Inverse Discrete Fourier Transform: equivalent to a matrix multiplication that inverts (undoes) a discrete Fourier transform.
ISI	Intersymbol Interference: self-interference caused by spreading of nearby pulses.
PAM	Pulse amplitude modulation
OFDM	Orthogonal Frequency Division Multiplexing: multi-carrier modulation system suited to multipath propagation environments.
OOK	On–Off Keying
QAM	Quadrature Amplitude Modulation: system in which two independent baseband signals are transmitted at the same frequency. One is up-converted to an "in-phase" carrier $\cos(2\pi f_{\mathrm{RF}}\, t)$, while the other is up-converted to a "quadrature" carrier $\sin(2\pi f_{\mathrm{RF}}\, t)$.
QPSK	Quadrature Phase Shift Keying: QAM using BPSK on both I and Q carriers.
TDM	Time Division Multiplexing: users on a single frequency channel have assigned time slots.
VSB	Vestigial Sideband: The remaining vestige of the lower sideband of an AM signal after asymmetric bandpass filtering.

References

[1] Federal Communications Commission *DTV Report on COFDM and 8-VSB Performance*, Office of Engineering and Technology, September 30, 1999.

[2] IEEE Std 802.11a-1999(R2003) (Supplement to IEEE Std 802.11–1999) [Adopted by ISO/IEC and redesignated as ISO/IEC 8802-11:1999/Amd 1:2000(E)]. This standard describes the Wi-Fi COFDM format.

[3] Lahti, B. P., *Modern Digital and Analog Communication Systems*, 3rd edn, Oxford: Oxford University Press, 1998.

[4] Proakis, J. B., *Digital Communications*, 4th edn, New York: McGraw-Hill, 2000.

[5] Schulze, H. and Lüders, C., *Theory and Applications of OFDM and CDMA Wideband Wireless Communications*, New York: John Wiley, 2005.

CHAPTER

23

Modulation, noise, and information

In this chapter we examine how noise degrades the accuracy of digital data transmission and the fidelity of analog transmission. We begin with an explanation of matched filtering, a subject which came up in Chapter 22. We show that, for binary data links using matched filtering and coherent detection, the probability of error depends only on the noise level at the input of the receiver and on the energy, but not the shape, of the transmitted pulses. We look at two example systems: BPSK (binary phase-shift keying) with coherent detection and OOK (on–off keying) with envelope detection. The error rates and channel capacities (maximum error-free data rates when using forward error correction coding) are calculated and compared with Shannon's expression for the capacity of a bandlimited channel. Finally, traditional AM and FM are examined with respect to their noise characteristics.

23.1 Matched filtering

We stated in Chapter 22 that, in the presence of noise, the post-detection signal-to-noise ratio is maximized when the predetection bandpass shape of the receiver is that of a *matched filter*.

A matched filter is one whose impulse response is proportional to the time-reversed waveform of the incoming signal pulse, as will be shown below. For example, if the input signal is a pulse whose shape is the same as the symmetric impulse response of a root raised-cosine filter, then the receiver should use a root raised-cosine filter or an equivalent cascade of filters. Sometimes we deal with complicated pulses, such as the biphase coded pulses used in pulse compression radar. In these cases, it is common to use a front-end bandpass filter which is a matched filter for the individual subpulses, followed by coherent detection and then a decoder which undoes the plus/minus phase coding. Here the cascade of the front-end filter and the decoder is equivalent to a matched filter for the coded pulses. Note that coherent detection is down-conversion to baseband, a linear operation. We can think of the entire matched filter as the actual detector, whose output samples are the received symbols.

Let D denote the output of the receiver's filter (which we have just defined as the "detector" output) and let $h(t)$ be the impulse response of this filter, a real function. For a signal pulse unaccompanied by noise, the output from the filter is given by

$$D_S = \int_{-\infty}^{\infty} h(t) V_S(-t)\mathrm{d}t, \tag{23.1}$$

where we have assumed the pulse position to be such that the filter output should be sampled at $t=0$. The output signal power, in units of volts^2, is given by $P_S = D_S{}^2$. The noise output of the filter can be written in terms of N_0, the standard one-sided noise density, in units of volts^2/Hz, as

$$P_N = \frac{N_0}{2} \int_{-\infty}^{\infty} |H(\omega)|^2 d\omega, \tag{23.2}$$

where $H(\omega)$ is the filter transfer function, i.e., the Fourier transform of $h(t)$. Note that we have assumed that N_0 is a constant (white noise). Using Parseval's theorem, we can write P_N in terms of $h(t)$:

$$P_N = \frac{N_0}{2} \int_{-\infty}^{\infty} h(t)^2 dt. \tag{23.3}$$

With these expressions for P_S and P_N, we can write the output signal-to-noise ratio as

$$\mathrm{SNR} = \frac{P_S}{P_N} = \frac{\left(\int_{-\infty}^{\infty} h(t) V_S(-t)\mathrm{d}t\right)^2}{\frac{N_0}{2}\int_{-\infty}^{\infty} h(t)^2 \mathrm{d}t}. \tag{23.4}$$

At this point, we invoke Schwarz's inequality[1] which gives us

$$\mathrm{SNR} \le \frac{\int_{-\infty}^{\infty} h(t)^2 \mathrm{d}t \int_{-\infty}^{\infty} V_S(-t)^2 \mathrm{d}t}{\frac{N_0}{2}\int_{-\infty}^{\infty} h(t)^2 \mathrm{d}t} = \frac{1}{N_0}\int_{-\infty}^{\infty} V_S(-t)^2 \mathrm{d}t = \frac{1}{\frac{N_0}{2}}\int_{-\infty}^{\infty} V_S(t)^2 \mathrm{d}t. \tag{23.5}$$

---

[1] Schwarz's inequality is written as $|\int f(x)g(x)\mathrm{d}x|^2 \le \int |f(x)|^2 \mathrm{d}x \int |g(x)|^2 \mathrm{d}x$. This can be seen by considering the integral of fg to be the dot product of a multidimensional vector, while the integral $\int |f(x)|^2 \mathrm{d}x$ is the length of the vector f and $\int |g(x)|^2 \mathrm{d}x$ is the length of the vector g. For two vectors $\boldsymbol{A}$ and $\boldsymbol{B}$, we know that $|\boldsymbol{A}\bullet\boldsymbol{B}|^2 \le |\boldsymbol{A}\bullet\boldsymbol{A}||\boldsymbol{B}\bullet\boldsymbol{B}|$.

Looking at Equation (23.4), we see that if $h(t) = \beta V_S(-t)$, where β is any constant, then the SNR will be equal to its maximum possible value, the right-hand expression in Equation (23.5). This is the matched filter. Note that, with a matched filter, the signal-to-noise ratio for a pulse is independent of the pulse shape and is simply the time integral of $V_S(t)^2$, i.e., the energy of the pulse, $\mathscr{E}_S$, divided by the noise spectral density N_0. (N_0 has units of volts^2/Hz, which is the same as volts^2 sec.) Thus, when a matched filter is used,

$$\mathrm{SNR} = \frac{1}{N_0/2}\int_{-\infty}^{\infty} V_S(t)^2 \mathrm{d}t = \frac{2\mathscr{E}_s}{N_0}. \tag{23.6}$$

23.2 Analysis of a BPSK link

Coherent detection makes the BPSK link especially easy to analyze. The signal at the output of the matched filter is just the sum of the signal voltage and the noise voltage, which we will assume to have a Gaussian distribution, characteristic of thermal noise. The detected signal voltage will be either A or $-A$, so the Gaussian noise distribution will be centered at either A or $-A$, as shown in Figure 23.1. The one/zero decision threshold will, of course, be set at $V = 0$ for this symmetric situation.

The probability distribution for a zero (the solid curve) is the normal Gaussian distribution function, $p_0 = (2\pi\sigma^2)^{-1/2} \exp[-(V+A)^2/(2\sigma^2)]$. By inspection of the figure we can write

$$p_\mathrm{e} = \int_0^{\infty} \frac{\mathrm{e}^{-(V+A)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} dU = \int_A^{\infty} \frac{\mathrm{e}^{-V^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \mathrm{d}U = \frac{1}{\sqrt{2\pi}}\int_{\frac{A}{\sigma}}^{\infty} \mathrm{e}^{-u^2/2}\mathrm{d}U. \tag{23.7}$$

Now let us relate this probability of error to the signal-to-noise ratio. We saw in the previous section that, at the output of a matched, the ratio of the square of the signal portion of the sampled output to the average square of the noise portion is $2\mathscr{E}_S/N_0$, where $\mathscr{E}_S$ is the energy of the pulse, i.e., the time integral of the V_S^2, and

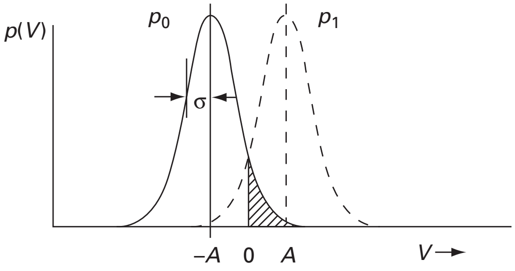

Figure 23.1. Gaussian voltage probability distributions for received BPSK ones and zeros for $A = 1.5\sigma$ (SNR = 1.5^2). The shaded area in the figure is the probability p_e of a transmission error, i.e., the probability that a received zero will be interpreted as a one or that a received one will be interpreted as a zero.

N_0 is the noise power density. Therefore $A^2/\sigma^2 = 2\mathscr{E}_S/N_0$, and we can rewrite Equation (23.7) as

$$p_e = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{2\,\mathscr{E}_S/N_0}}^{\infty} e^{\frac{-u^2}{2}} du. \tag{23.8}$$

We have been implicitly working at baseband, i.e., assuming that the signal (and the noise) have been converted down from an RF carrier frequency, ω_0, through multiplying the modulated RF signal by a sine wave in phase with the (normally suppressed) carrier. This, as we have seen, constitutes synchronous detection. In this common situation, the matched filter is a baseband filter.

It is interesting to note that we could alternately have used a matched filter at the RF (or some IF) frequency. In this case, the output of the filter is an RF sine wave, multiplied by the baseband pulse. We can sample this directly, but the sampling must be very precise, so that the samples are taken at points that, with a dc pulse, would be at the peaks of the RF sine wave. The SNR will be the same as for the synchronous conversion to baseband because even though the RF bandwidth is twice the baseband bandwidth and contains twice as much total noise power, the synchronous sampling will respond to only half this noise, e.g., the "cosine" component.

If symbols are arriving at the rate $1/T$ in the minimum baseband bandwidth of $1/(2T)$ that eliminates intersymbol interference (see Chapter 22), then $2\mathscr{E}_S/N_0 = (\mathscr{E}_S/T)/(N_0/[2T]) = P_S/P_N = SNR$ the signal-to-noise ratio. This probability of error from Equation 23.8 is plotted vs. SNR in Figure 23.2.

This figure displays the well-known "dropout" effect in digital communications systems. If, for example, the input SNR drops from ten to five, a factor of only 2, the error rate increases by a factor of 100. The probability of error for the BPSK distributions shown in Figure 23.1, where A/σ = 1.5, is 0.067.

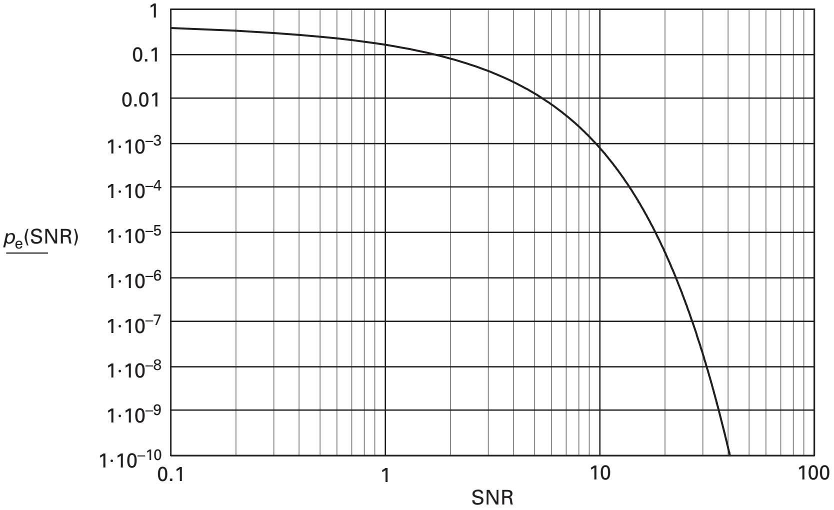

Figure 23.2 Envelope probability distributions $P(E, A, \sigma)$, *for* $\sigma = \sqrt{2}$ and $A = 0$ *or* $1.5\sqrt{2}$.

23.3 On–off keying with envelope detection

The BPSK system discussed above, using coherent detection, is suited for data transmission under low signal conditions. (When there is ample signal power, multilevel "M-ary" modulation can obviously transmit data at a faster bit rate.) But let us now consider simple on–off keying with envelope, i.e., noncoherent detection. Before we can calculate the probability of error, we must find the distribution function of the output voltage from the envelope detector.

23.3.1 Envelope probability distributions

As we have seen, the IF signal, containing random noise, is a random variable with Gaussian probability distribution. The presence of a signal offsets the Gaussian curve. The output of the envelope detector is also a random variable. Let us find the probability distribution of the envelope, first without the presence of a signal. Figure 23.3(a) shows the in-phase and quadrature components, $\boldsymbol{N}_I$ and $\boldsymbol{N}_Q$. of the IF noise voltage. The length of their vector sum is $\boldsymbol{E}$, the envelope voltage.

Since the I and Q noise components are uncorrelated, their joint probability distribution is the product of their individual Gaussian distributions.

$$p(V_I, V_Q) = \frac{\mathrm{e}^{-V_I^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \times \frac{\mathrm{e}^{-V_Q^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} = \frac{\mathrm{e}^{-(V_I^2+V_Q^2)/2\sigma^2}}{2\pi\sigma^2} = \frac{\mathrm{e}^{-E^2/2\sigma^2}}{2\pi\sigma^2}. \tag{23.9}$$

To obtain the distribution in terms of E and θ, we note that

$$p(E,\theta)E\mathrm{d}\theta\mathrm{d}E = \frac{\mathrm{e}^{-E^2/2\sigma^2}}{2\pi\sigma^2}E\mathrm{d}\theta\mathrm{d}E, \tag{23.10}$$

from which we identify

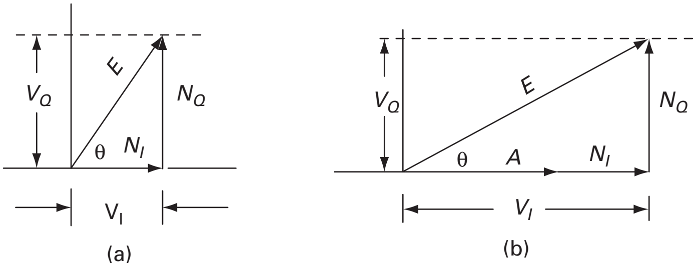

Figure 23.3 (a) *I* and *Q* noise phasors alone; (b) noise phasors together with a sine wave of amplitude *A*.

$$p(E,\theta)=\frac{E\mathrm{e}^{-E^2/2\sigma^2}}{2\pi\sigma^2}. \tag{23.11}$$

Integrating over θ, from 0 to 2π, we find $p(E)$, the envelope probability distribution:

$$p(E)=\int_0^{2\pi}P(E,\theta)\mathrm{d}\theta=\frac{E\mathrm{e}^{-E^2/2\sigma^2}}{\sigma^2}. \tag{23.12}$$

This function is known as the Rayleigh probability distribution. Note: for both the I and the Q noise components, the variance is σ^2. Therefore, the total noise power, which is the expectation of E^2, is $2\sigma^2$.

Now we must find the envelope distribution function when the IF signal is the superposition of noise plus a sinusoidal carrier. This is shown in Figure 23.3(b). Again, both noise voltages, N_I and N_Q, have Gaussian distributions, but now the distribution for N_I is centered at A, the amplitude of the carrier, so we can write

$$\begin{aligned}p(V_I,V_Q)&=\left[\sqrt{2\pi\sigma^2}\right]^{-2}\mathrm{e}^{[-(V_I-A)^2+V_Q^2]/2\sigma^2}\\&=(2\pi\sigma^2)^{-1}\mathrm{e}^{[-(E\cos(\theta)-A)^2+E\sin(\theta)^2]/2\sigma^2}.\end{aligned} \tag{23.13}$$

Expanding the argument of the exponential on the right-hand side we have

$$p(V_I,V_Q)=(2\pi\sigma^2)^{-1}\mathrm{e}^{-(E^2+A^2)/2\sigma^2}\mathrm{e}^{-2AE\cos(\theta)/2\sigma^2}. \tag{23.14}$$

Following the steps used in the carrier-free case, and integrating over θ, we find

$$\begin{aligned}p(E,A,\sigma)&=\frac{E\mathrm{e}^{-(E^2+A^2)/2\sigma^2}}{\sigma^2}\frac{1}{2\pi}\int_0^{2\pi}\mathrm{e}^{-AE\cos(\theta)/\sigma^2}\mathrm{d}\theta\\&=\frac{E\mathrm{e}^{-(E^2+A^2)/2\sigma^2}}{\sigma^2}I_0(AE/\sigma^2),\end{aligned} \tag{23.15}$$

where I_0 is the modified Bessel function of order zero. This function, $p(E,A,\sigma)$, known as the *Rician* distribution, is plotted in Figure 23.4 for $A=0$ (the envelope distribution of noise alone) and for $A=1.5\sqrt{2}$, for the same noise power and average transmitted power (assuming equal probabilities for transmitted ones and zeros) used for the BPSK example of Figure 23.1.

Let us calculate the probability p_e that a bit is received incorrectly when we have (arbitrarily) set the decision threshold at the intersection of the ON and OFF envelope distribution functions, which we will denote as E_t. We will also assume that we transmit, on average, equal numbers of ones and zeros. The probability of error is, therefore, $\frac{1}{2}p_{\mathrm{e}0}+\frac{1}{2}p_{\mathrm{e}1}$, where $p_{\mathrm{e}0}$ is the probability that a transmitted zero is received incorrectly and $p_{\mathrm{e}1}$ is the probability that a transmitted one is received incorrectly or

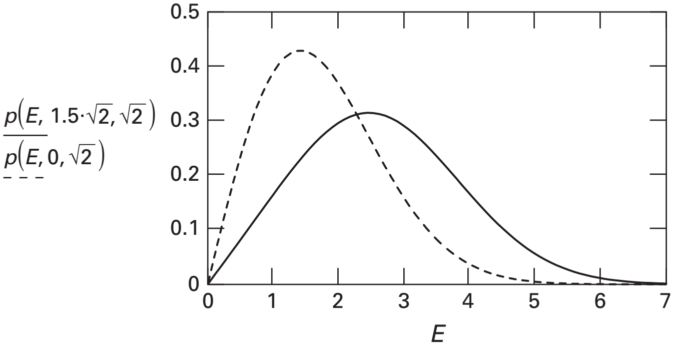

Figure 23.4 Probability distribution $p(E,A,\sigma)$, for the envelope of a signal that is the sum of a sine wave of amplitude $A = 1.5\sqrt{2}$ plus noise with $\sigma = 2$.

$$p_{\mathrm{eOOK}}(A,\sigma) = 1/2 \int_{E_t}^{\infty} p_0(E,A,\sigma)\mathrm{d}E + 1/2 \int_{0}^{E_t} p_1(E,A,\sigma)\mathrm{d}E. \tag{23.16}$$

Evaluating P_{eOOK} for A $= 1.5\sqrt{2}$ and $\sigma = \sqrt{2}$ for comparison with the BPSK example of Figure 23.1 (same average transmitted power, but twice the noise power (quadrature as well as in-phase components), yields $P_{\mathrm{eOOK}} = 0.34$ for a threshold $E_{\mathrm{t}} = 2.3$, compared to only .067 for P_{eBPSK}.

Channel capacity

We have seen how to calculate the expected bit error rate as a function of the signal and noise powers for two example situations, coherently-detected PSK and envelope-detected on–off keying. We also know that data error rates can be reduced by expanding (encoding) the data so that it contains redundancies.[2] Of course, the net data rate slows when redundant bits are transmitted. These redundancies may be crude, such as transmitting a packet of data several times, or elegant, such as the nested combination of block codes and convolutional codes used in data links and digital broadcasting systems. Given a particular communications link with an optimal encoding scheme, it is obvious to ask how many bits must be transmitted, on average, for each data bit, if we are willing to tolerate a certain average data error rate. One might expect that, as the error tolerance is reduced to zero, the optimal transmission efficiency would also go to zero, i.e., that an infinite number of bits would have to be transmitted for each recovered data bit, as indicated by the crossed out curve in Figure 23.5. Shannon's remarkable "noisy channel theorem" showed that this is not so. For any given communications link, there must exist coding schemes that will permit data transmission with an arbitrarily low error tolerance while still

---

[2] Making the data redundant enough so that errors can be detected and corrected at the receiver is known as forward error correction (FEC). In reverse error correction, the receiver can only detect errors. To make corrections, it must request repeat transmissions.

Figure 23.5 Transmission efficiency vs. tolerated output error rate when using optimum channel encoding.

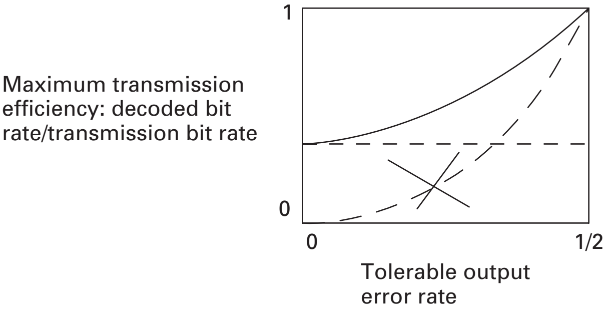

achieving a non-zero efficiency. This situation is shown by the solid curve Figure 23.5, which shows the efficiency for a link using an optimum error coding scheme.

The maximum efficiency is unity if we are willing to tolerate an output error rate of 1/2, but this is the point at which no net information is transmitted since a one is randomly declared a one or a zero and vice versa. But as the error rate tolerance is lowered to zero, the maximum efficiency approaches an asymptotic *non*zero value. This asymptotic maximum efficiency value is called the *channel capacity* (bits/bit) and, when multiplied by the bit rate, in bits/sec, the result, in units of bits/sec, is a rate and is also referred to as the channel capacity. (Context usually resolves any confusion between the two.) The ususal statement of Shannon's theorem is that it is possible to transmit data at a rate equal to or less than the channel capacity, with an arbitrarily low data error rate. While Shannon's work shows that optimum codes must exist, it does not show how to construct then. However, it does show how to calculate the channel capacity for a given link, the highest standard against which we can judge the efficiency of any coding scheme proposed for the link. If the arbitrarily low data error rate is set too low, the encoded data must be in the form of extremely long blocks. Fortunately, acceptably low error rates can be achieved with acceptably short data blocks.

Binary symmetric channel

The coherent BPSK link discussed above is an example of a *binary symmetric channel*. The transmitted symbols are either one or zero, and every received signal is deemed either a one or a zero. From the symmetry of Figure 23.1, it is obvious that the probability p_e that a transmitted one will be received in error as a zero is equal to the probability that a transmitted zero will be received in error as a one, as indicated in Figure 23.6. This figure, a sort of probability flow graph for a binary channel, is equivalent to a 2×2 matrix whose coefficients are probabilities. The binary symmetric channel corresponds to a symmetric channel matrix.

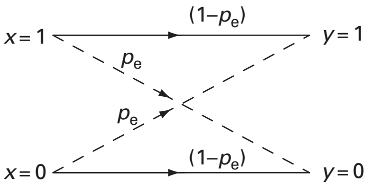

Figure 23.6. Binary symmetric channel: p_e is the probability that a symbol x_i is wrongly received as a symbol y_j.

The channel capacity for this link, which depends only on p_e and on p_i, the probability that the i-th transmitted bit is a one, is given by

$$\begin{aligned} C(p_e, p_i) &= p_e \log_2(p_e) + (1 - p_e)\log_2(1 - p_e) \\ &\quad - p_i \log_2(p_i) - (1 - p_i)\log_2(1 - p_i). \end{aligned} \tag{23.17}$$

This expression is always greatest for $p_i = 1/2$ (equal probability of transmitted ones and zeros). Figure 23.7 shows C(p_e,1/2) plotted versus p_e.

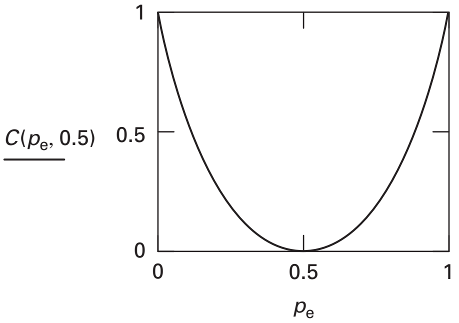

Figure 23.7 Channel capacity of a binary symmetric channel vs. transmission error rate, assuming ones and zeros are transmitted with equal probability.

Note that the channel capacity falls to zero at $p_e = 1/2$, where the receiver is equally likely to interpret a transmitted one as a zero and vice versa. Then C rises again to unity for $p_e = 1$, where the receiver always mistakes a one for a zero and zero for a one. In this case, the message is transmitted faithfully, assuming the receiver realizes that it should complement the bits. (It takes only a single transmitted bit to resolve this ambiguity.)

The reader will find excellent treatments of information theory, channel capacity, and coding in the texts listed at the end of this chapter but the origin of Equation (23.17) warrants at least a brief discussion, in the limited context of a binary channel. The discussion should at least serve to show that, while the concepts are subtle, the mathematics is simple. In the binary channel, the transmitter sends a symbol which is denoted either x_1 (a one) or x_2 (a zero). The *information* in a transmitted symbol is defined as $I(x_i) = \log_2 (p(x_i)^{-1})$, where $p(x_i)$ is the probability of x_i. Note that if $p(x_i)$ is nearly one, the information contained in x_i is nearly zero. If the transmitted symbols are almost always ones (or zeros), very little information is transmitted by sending a one (or zero), as we would have expected the one (or the zero). But if $p(x_i)$ is a very small number, the occurrence of x_i is highly informative, as we would not have expected it. A second definition is that of *mutual information*,

$I(x_i; y_j) = \log_2[p(x_i|y_j)|p(x_i)]$, where $p(x_i|y_j)$ is the conditional probability that x_i was transmitted, given that y_j was received. Note that while the distribution of transmitted symbols is generally flat (uniform), the distribution of $I(x_i; y_j)$ for a given j is not uniform, but peaked around the value of i most likely to have resulted in y_j.

The expected value of the mutual information is the sum of the $I(x_i; y_j)$'s, weighted by their probabilities, i.e.,

$$\langle I(x_i; y_j)\rangle = \sum_i \sum_j p(x_i, y_j) I(x_i; y_j), \tag{23.18}$$

where $p(x_i, y_j)$ is the joint probability of x_i and y_j, i.e., $p(x_i\, y_j) = p(x_i)\, p(y_j|x_i)$. For the binary symmetric channel discussed above, $p(y_j, x_i) = p_e$ for $i \neq j$ and $(1 - p_e)$ for $i = j$. The channel capacity is given by the maximum value of $\langle I(x_i; y_j)\rangle$ with respect to the set of values $p(x_\text{i})$, by

$$C = \max\langle I(x_i; y_j)\rangle \text{ w.r.t } \{p(x_i)\}. \tag{23.19}$$

For the binary symmetric channel, the maximum occurs when equal numbers of ones and zeros are transmitted, i.e., when $p(x_1) = p(x_2) = 1/2$. Evaluation of Equation (23.19) yields Equation (23.17), the expression for the channel capacity. Note that the channel capacity is based on the relative frequency of transmitted ones and zeros and on the values of $\text{p}(y_j \mid x_i)$, which is the conditional probability that the receiver produces y_j when the transmitter sent x_i. However, to calculate the values of $I(x_i; y_j)$, we need to know $\text{p}(x_i \mid y_j)$, the conditional probability that when the receiver produces y_j, the transmitter had sent x_i. The two conditional probabilities are related through Bayes' theorem: $p(x_i, y_j) = p(x_i)\, p(y_j \mid x_i) = p(y_j)\, p(x_i \mid y_j)$.

Channel capacity of the BPSK and OOK example channels

For the BPSK example shown in Figure 23.1, with $A/\sigma = 1.5$, we found the probability of error to be $p_\text{e} = .067$. Using this value in Equation 23.17, the channel capacity is 0.645 data bits/transmitted bit. For the comparable asymmetric OOK channel of Figure 23.4 the two integrals in Equation 23.16 are, respectively, the probability that a transmitted zero is received as a one and vice versa. If we calculate the channel capacity using Equation 23.19, we find that it maximizes at about $\text{C} = 0.08$, if the threshold is set at $E_\text{t} = 2.5$ and the proportion of transmitted zeros is 52%. In addition to low channel capacity, another disadvantage of the OOK system is that the threshold must be changed when either the signal or noise level changes.

Channel capacity of a bandpass channel

Shannon also presented a formula for the *maximum* channel capacity for a band-limited channel with added Gaussian white noise – the channel most amenable to analysis, and a channel often encountered in practice, for example, in spacecraft telemetry links. This formula states that the maximum channel capacity is given by

$$C = B\log_2[1 + S/N] = B\log_2[1 + S/(BN_0)] \tag{23.20}$$

where S is the received signal power, N is the noise power which is equal (for white noise) to BN_0, where B is the channel bandwidth, and N_0 is the noise spectral density at the receiver in Watts/Hz. The figure of merit for any particular modulation scheme is how close its channel capacity approaches this ideal maximum channel capacity. To see that Equation 23.20 is reasonable note that, when S/N is high, it is essentially the practical number of quantization levels, n, and $\log_2(n)$ converts this into bits. The bandwidth factor, B, will be the symbol rate or close to it.

For illustration, suppose the signal power available from a spacecraft is 10^{-16} Watts, contained within a bandwidth of 100Hz, and the noise density at the receiver input is 4×10^{-19} Watts/Hz. With these numbers, Equation 23.20 produces $C = 181$, which is the maximum rate (bits/sec) at which information could be transmitted over the channel at an arbitrarily small error rate, if the modulation scheme and signal coding are optimum.

Let us look at our example BPSK channel in the light of Equation 23.20. In that example, the SNR at the output of the detector was $(A/\sigma)^2 = 1.5^2$. If we are using a matched filter, we saw that $(A/\sigma)^2$ be equal to $2\mathscr{E}/N_0$. Let us assume we transmit bits at a rate of $1/T$ and use receiver bandwidth of $1/(2T)$. The received power is $S = \mathscr{E}/T$ and the noise at the receiver is $N_0/(2T)$, from which we have $S/N = 2\mathscr{E}/N_0$. Using Equation 23.20, we see that the maximum channel capacity will be $C = 1/(2T)^{-1}\log_2(1 + 2\mathscr{E}/N_0) = .850/T$ bits/sec. But, from the bit error probability, a function of A/σ, we had calculated the actual channel capacity of this BPSK link to be .645 bits/bit. If we multiply this by the bit rate, $1/T$, the channel capacity becomes $.645/T$ bits/second. This BPSK link, therefore, has $.645/.850 = 76\%$ of the maximum possible channel capacity.

For much larger values of S/N, we would find that, while the ideal channel capacity increases, our BPSK channel capacity saturates at the value $1/T$. This is simply due to staying with two-level binary modulation, as opposed to N-level PAM modulation.

Noise in analog FM and AM systems

Let us first look at the noise produced at the output of an FM receiver. The instantaneous received signal voltage, V_{SIG}, is accompanied by noise, V_N, as shown in Figure 23.8.

This is the same as the diagram in Figure 23.3(b), except that now we are interested in the phase angle rather than the magnitude of the vector sum of the signal plus noise. This "phase noise," ϕ_N, will cause noise in the detected output of an FM or PM detector. Clearly the angle ϕ_N becomes smaller if the signal strength is increased. But there is another way to defeat the noise. The signal-to-noise ratio at the detector output depends on the ratio of the signal's modulation phase excursions to the phase noise. If the modulation level is increased, even without increasing the signal strength, the output signal-to-noise ratio will be improved.

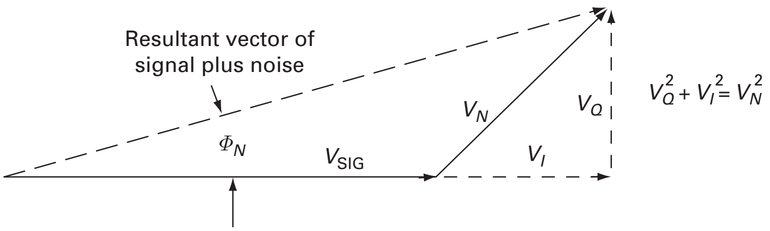

Figure 23.8. Signal and noise voltages.

If, for example, the rms phase noise is 1/10 radian and the modulation index is 1 radian, the phase SNR is 100. If the deviation is increased to produce a modulation index of 5 radians, the phase SNR increases to 2500. The improvement has been obtained not by increasing the signal power but by increasing the signal bandwidth. In the case of amplitude modulation, the SNR depends on the noise modulating the length of the vector $V_{\mathrm{SIG}} + V_N$. Since V_{N} is fixed in any given situation, the only way to improve the SNR in AM is to increase V_{SIG}, i.e., increase the transmitted power.

Analysis of the SNR improvement in FM

A quantitative analysis of the SNR improvement in FM is simpler if we take the noise to be the background noise produced by the detector when the signal is unmodulated, i.e., the total power in the hiss coming from the loudspeaker or other output device. For the signal we will take an audio sine wave with 100% modulation (maximum deviation). We can represent the noise, V_N, by in-phase (I) and quadrature (Q) noise components, V_I and V_Q where $V_I^2 + V_Q^2 = V_N^2$ as shown in Figure 23.8. Both V_Q, and V_I are phasors rotating at ω_0, the frequency of the unmodulated carrier. Their amplitudes are random and independent. The I component of the noise is the most effective in causing amplitude fluctuations and therefore contributes noise in AM demodulation. But it is mostly the Q component, since it is perpendicular to V_{SIG}, that causes angle fluctuations and therefore contributes noise in the FM demodulation. For $V_N \ll V_{\mathrm{SIG}}$, the instantaneous angle noise, $\phi_N(t)$, is just $V_Q(t)\,/V_{\mathrm{SIG}}$ radians. Since V_{SIG}, the carrier amplitude, is constant, the power spectrum of ϕ_n (call it S_ϕ) is proportional to the power spectrum of V_Q. The spectral distribution of V_Q can be assumed uniform (white) so S_ϕ is also uniform. The integral of S_ϕ over the IF band gives the mean square phase fluctuation, $\langle(\phi_n(t))^2\rangle$, so we can write $S_\phi = \langle(\phi_n(t))^2\rangle/B_{\mathrm{IF}} = \langle V^2{}_Q(t)\rangle/(V^2{}_{\mathrm{SIG}}B_{\mathrm{IF}}) = \langle V^2{}_N(t)/2\rangle/(V^2{}_{\mathrm{SIG}}\,B_{\mathrm{IF}})$ where B_{IF}, the IF bandwidth, is twice the maximum deviation. An FM demodulator produces an output spectral density proportional to the time derivative of the phase, $\langle(\mathrm{d}\phi_n(\mathrm{t})/\mathrm{d}t)^2\rangle$. The spectral density of the noise in the (one-sided) audio band at the detector output is therefore given by $2\omega_a{}^2 S_\phi(t)$, and the total noise power is the integral of this spectral density over the output bandwidth of the detector, i.e., the audio band (0 to B_a radians):

$$\text{Output noise power} = \int_0^{B_\mathrm{a}} 2\omega^2 S_\phi \mathrm{d}\omega = S_\phi \int_0^{B_\mathrm{a}} 2\omega^2 \mathrm{d}\omega = \frac{2}{3} S_\phi {B_\mathrm{a}}^3 \\ = \frac{{B_\mathrm{a}}^3}{6k_\mathrm{osc}A_\mathrm{MAX}} \frac{\langle {V_\mathrm{N}}^2 \rangle}{V_\mathrm{SIG}^2}. \tag{23.21}$$

The maximum amplitude of the sine-wave signal at the output of the detector is $k_\mathrm{osc}A_\mathrm{MAX}$, the maximum deviation, so the maximum signal power is just $P_\mathrm{SIG} = {k_\mathrm{osc}}^2{A_\mathrm{MAX}}^2/2$. Taking the noise power from Equation (23.21), the signal-to-noise ratio at the detector output is:

$$\text{Maximum output SNR} = 3\left(\frac{k_\mathrm{osc}A_\mathrm{MAX}}{B_\mathrm{a}}\right)^3 \frac{V_\mathrm{SIG}^2}{\langle {V_\mathrm{N}}^2 \rangle}. \tag{23.22}$$

Note that the output SNR improves as the cube of the ratio of the maximum deviation to the full audio bandwidth.

Output signal-to-noise ratio for an AM signal with the same carrier power

Let us consider an AM system with the same carrier power and the same audio bandwidth, in order to compare its output SNR to that of the FM system. Again the modulation will be a single sine-wave tone and the amplitude of the carrier will be V_SIG. At 100% modulation, the amplitude of the sine wave modulation envelope will also be V_SIG so the audio signal power at the AM detector output will be ${V_\mathrm{SIG}}^2/2$. For the AM system, the IF bandwidth needs to be only $2B$, twice the audio bandwidth (wide enough to accommodate the highest audio frequency but, to minimize noise, no wider). The noise voltage at the detector output will be V'_I, the in-phase component of V'_N, where the primes distinguish the noise voltages in the AM IF band from the noise voltages in the wider FM IF band. The noise power will be given by $\langle {V'_\mathrm{I}}^2 \rangle = \langle {V'_\mathrm{N}}^2 \rangle/2$. The SNR at the AM detector output is therefore

$$\text{AM output SNR} = \frac{{V_\mathrm{SIG}}^2}{\langle {V'_N}^2 \rangle}. \tag{23.23}$$

Comparison of noise, FM vs. AM under strong signal conditions

All that remains in order to compare the FM and AM systems is to note that the ratio of the IF noise powers is just the ratio of the IF bandwidths, i.e.,

$$\frac{\langle {V_N}^2 \rangle}{\langle {V'_N}^2 \rangle} = \frac{2kA_\mathrm{max}}{2B_\mathrm{a}} = \frac{kA_\mathrm{max}}{B_\mathrm{a}}. \tag{23.24}$$

Using Equations (23.22)–(23.24), we can write

$$\frac{\text{FM SNR}}{\text{AM SNR}} = \frac{3{k_\mathrm{osc}}^2 {A_\mathrm{MAX}}^2}{{B_\mathrm{a}}^2} = 3(\text{Deviation ratio})^2. \tag{23.25}$$

In FM broadcasting the deviation ratio is about five so the output SNR for a maximum-amplitude audio tone will be higher with FM than with AM by a factor of $3{\cdot}5^2 = 75$ or about 19 dB. The FM sound in television uses a deviation ratio only 1/3 as large as that of FM broadcasting so the signal-to-noise improvement is correspondingly less. Remember that the above analysis is only for the situations of high signal-to-noise. When the signal is comparable to or lower than the noise, the phase is almost totally determined by the noise and the FM system is useless.

FM, AM and channel capacity

The improvement in a signal-to-noise ratio that is possible with wideband FM is an example of increasing the channel capacity of a communications channel by increasing the bandwidth.

Note that S/N_O has units of bandwidth and can be considered a "natural" bandwidth for a given N_0 and S. N_0 is determined by the noise added along the channel, such as atmospheric noise and noise added by the receiver. The signal power, S, is determined by the transmitter power, transmitter and receiver antenna gains, and propagation loss. Channel capacity, from Equation (23.20), vs. bandwidth is plotted in Figure 28.9. Both are normalized to S/N_0.

Note that, for $b > 1$, the channel capacity has essentially reached an asymptotic value of 1.44 S/N_0. If we are below the knee of the curve, we can increase the channel capacity significantly at no cost in transmitter power by (somehow) using more bandwidth. We have seen that FM broadcasting does just this. On the other hand, Equation (23.20) shows that it is expensive (and ultimately impractical) to increase channel capacity by increasing power since the log term increases slowly.

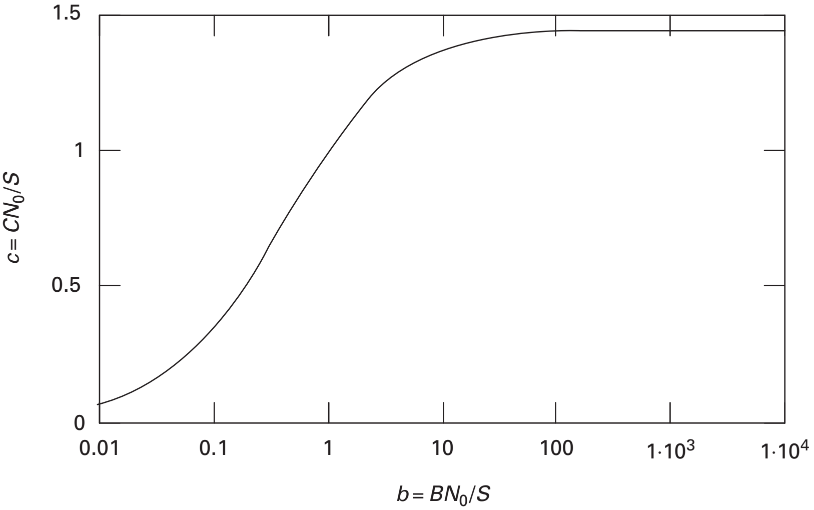

Figure 23.9. Channel capacity vs. bandwidth (both are normalized to S/N_0, the "natural bandwidth").

In AM, the bandwidth is fixed by the highest modulation frequency; the total bandwidth in standard full-carrier DSB AM is twice the highest modulation frequency. When a weak AM station is received, the signal-to-noise ratio is low enough to put us beyond the knee of the channel capacity vs. bandwidth curve and there would be no gain in going to a modulation scheme that increases the bandwidth. But if the station is strong, we are probably far below the knee of the curve. In this case, changing to FM modulation (without changing transmitter power) can bring us up the curve where the higher channel capacity allows a higher signal-to-noise ratio.

Problems

Problem 23.1. Show that the differential equation $x^2y'' + xy' - x^2y = 0$ (modified Bessel's equation of order zero) is satisfied by the function

$$I_0(x) = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{e}^{x\cos(\theta)}\mathrm{d}\theta.$$

Problem 23.2. Consider an asymmetric binary channel in which the probability that a one is correctly received is 0.9 but the probability that a zero is correctly received is 1/2. (A transmitted zero is equally likely to be declared a zero or a one.) Suggest a simple coding method to transmit data reliably through this channel.

Problem 23.3. Consider the situation where a rectangular pulse of length T is transmitted, but the receiver instead of using a $\sin(1/2\ \omega T)/(1/2\ \omega T)$ matched filter, uses a rectangular filter of bandwidth β/T. The noise at the input is Gaussian. Find the value of β that maximizes the SNR at the output of the filter. Compare this to the value that would have been obtained if the matched filter had been used.

References

[1] Lahti, B. P., *Modern Digital and Analog Communication Systems*, 3rd edn, Oxford: Oxford University Press, 1998.

[2] Proakis, J. B., *Digital Communications* – 4th edn, New York: McGraw-Hill, 2000.

CHAPTER

24 Amplifier and oscillator noise analysis

A fundamental limitation in amplifiers and oscillators is the internally generated random noise whose physical causes are shot noise in semiconductors and thermal noise (Johnson noise) in resistors. Amplifier noise is a practical concern whenever this noise level is comparable to the signal being amplified. Oscillator noise, characterized as phase or frequency fluctuations, adds noise to FM systems, limits how closely carriers may be spaced in frequency division multiplexing (FDM), and limits the length of signal averaging times in coherent communication systems. In this chapter we discuss the way transistor noise is characterized, treating the transistor itself as a two-port network, and how these characteristics are taken into account to design low-noise circuits.

24.1 Amplifier noise analysis

In Chapter 17 we quoted a formula showing how the noise figure of any active two-port device depends on the impedance of the signal source:

$$F = F_{\min} + (R_n/G_s)|Y_s - Y_{\text{opt}}|^2. \tag{24.1}$$

Here $F_{\min}$ is the minimum noise figure; $Y_s = G_s + jB_s$ is the source admittance, $Y_{\text{opt}} = G_{\text{opt}} + jB_{\text{opt}}$ is the source admittance that produces $F_{\min}$, and R_n is a real parameter known as the *noise resistance*. Equation (24.1) applies as well to audio and instrumentation amplifiers as to radio and microwave amplifiers, as long as the four noise parameters of the device, $F_{\min}$, R_n, G_{opt}, and B_{opt}, are known for the frequency in question. While Equation (24.1) requires some effort to derive, its use is simple, at least with RF (narrowband) amplifiers.

24.1.1 Noise matching

To design a low-noise RF amplifier, we get out the data sheet for the transistor, find the value of Y_{opt} for the desired frequency, and design an input network to convert the impedance of the intended source (probably 50 ohms) into $1/Y_{\text{opt}}$.

Figure 24.1. Input network to convert a 50-ohm source impedance to $G_{OPT} + jB_{OPT}$.

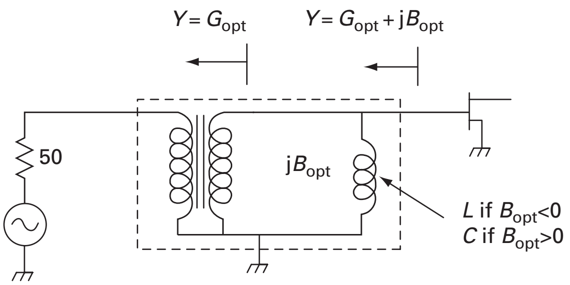

Figure 24.1 shows how a transformer and a reactor can form the matching network. (Of course there are many equivalent networks.)

With this input network, the transistor is said to be "noise matched," because the noise figure of the overall amplifier will be F_{min}. The input network that provides noise matching generally does not provide impedance matching, i.e., it does not convert the source impedance into the complex conjugate of the amplifier's input impedance. The resulting power mismatch means we will have less signal at the amplifier output than with a conjugate impedance match. But converting the source impedance into $1/Y_{opt}$, rather than into $1/Y^*_{in}$, will reduce the transistor noise more than the mismatch reduces the signal power. The net effect is to maximize the signal-to-noise ratio, i.e., provide the minimum possible noise figure.

(Note: In some cases simultaneous noise matching and impedance matching can be achieved or at least approached. If the device has a nonzero reverse transfer coefficient, the amplifier's input impedance is not fully determined by the input network. Depending on the transistor characteristics, a designer might find an output network to make the input impedance of the amplifier approach Z_0 while using an input network that provides noise matching. The use of feedback can provide additional degrees of freedom with which to pursue simultaneous matching.)

24.1.2 Equivalent circuits for noisy two-port networks

Equation 24.1 was derived by Rothe and Dahlke [4] who pointed out that a noisy two-port network (transistor, amplifier, etc.) can be modeled, for small signals, as an ordinary passive linear two-port with two external frequency-dependent noise sources. These noise sources can be series voltage generators at the input and output, shunt current generators at the input and output, a current source at one end and a voltage source at the other end, or, finally, both a voltage source and a current source at the same end. Figure 24.2 shows these equivalent circuits.

Note that when a current generator and a voltage generator appear on the same side they can be interchanged without changing their parameters, so Figures 24.2(e) and (f) (or g and h) are actually the same circuit.

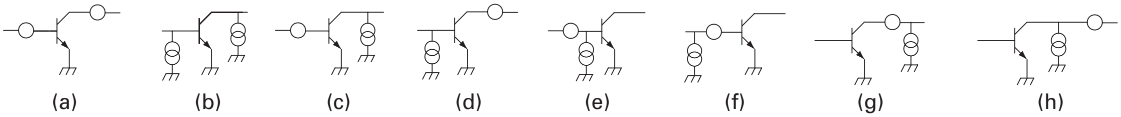

Figure 24.2. Equivalent representations of a noisy transistor as a noiseless transistor with two external noise generators.

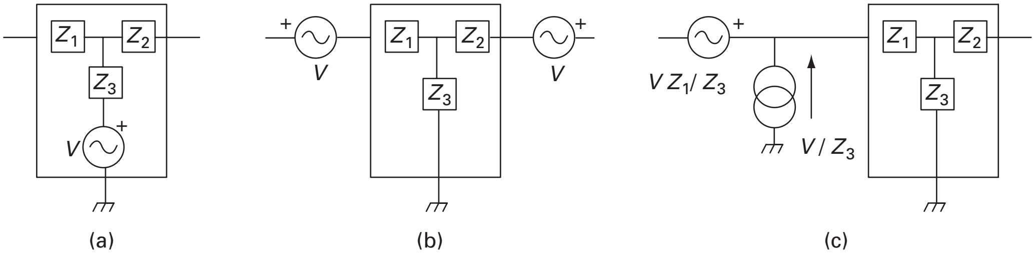

Figure 24.3. (a) An arbitrary two-port network with an internal voltage source; (b) and (c) networks equivalent to (a), but with external sources.

Note: Replacing internal sources with equivalent external sources is a sort of generalization of Thévenin's theorem and Norton's theorem, from one-port networks to two-port networks. Figure 24.3(a) shows an arbitrary two-port network with an internal voltage generator. Both (b) and (c) are equivalent to (a). In this example, the network is passive, but it can just as well be active, containing dependent voltage and current sources. When the internal source is a sine-wave source, the external generators are specified by complex amplitudes. When the internal source is a noise source, the external sources are specified by their rms values and the correlation between them. For example, if the external sources are a voltage source and a current source, the correlation is the complex number $\langle iv^* \rangle$, the time average of the product of the complex amplitude i and the complex conjugate of the complex amplitude v. When a two-port has multiple internal noise sources,[1] the resulting external sources can be consolidated into two external sources.

24.1.3 Noise figure of the equivalent circuit

Figure 24.2(e) (or 24.2f) is the easiest equivalent circuit to analyze for noise figure. Since both noise sources are placed at the input, the transfer characteristics (Z, Y, or S parameters) of the two-port will not come into play. In Figure 24.4, the circuit to be analyzed, the amplifier block can represent a transistor, a tube, or a complete amplifier. The two external noise sources, v and i, are equivalent to the device's internal sources of noise, whatever their number and physical nature. The other current source, i_T, represents the thermal

---

[1] A circuit model for a transistor usually contains several noise sources.

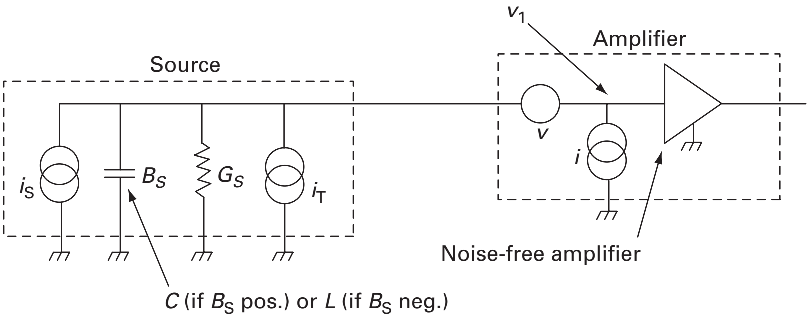

Figure 24.4. Signal source and noisy amplifier – equivalent circuit.

noise of the source, i.e., $\langle i_T\rangle^2 = 4kTG_s$. Since there are four noise parameters to be determined, we need more than just $\langle i^2\rangle$ and $\langle v^2\rangle$. The other two parameters come from the fact that i and v are generally not independent but are correlated. The part of i that is correlated with v defines the "correlation admittance," Y_c, through the relation $i_c = Y_c\, v$. Thus

$$\langle iv^*\rangle = \langle (i_c + i_u)v^*\rangle = \langle i_c v^*\rangle = \langle Y_c|v|^2\rangle = Y_c\langle |v|^2\rangle (G_c + jB_c)\langle |v|^2\rangle. \quad (24.2)$$

To find an expression for the noise figure, we must find the voltage v_1 at the input of the device and use the definition of noise figure to find the ratio of the average of $\langle |v_1|^2\rangle$ to the part of $\langle |v_1|^2\rangle$ attributable to the source. The result is independent of the input impedance of the device since, using superposition, we can find v_1, considering the noise sources one at a time. They all have the same effective source impedance, $1/Y_s$. For convenience, then, we can take the device to have an infinite input impedance. It follows immediately that

$$v_1 = (i + i_T)/Y_s + v. \quad (24.3)$$

To get the equivalent input *power* we use the average squared magnitude of v_1:

$$\langle |v_1|^2\rangle = (\langle |i|^2\rangle + \langle |i_T|^2\rangle/|Y_s|^2 + \langle |v|^2\rangle + (\langle iv^*\rangle/Y_s) + (\langle i^*v\rangle/Y_s^*). \quad (24.4)$$

Here we have used the fact that there is no correlation between i and i_T or between i_T and v. The noise figure is given by this total power divided by the part due to the noise current of the source:

$$F = \frac{((\langle |i|^2\rangle + \langle |i_T|^2\rangle))/|Y_s|^2 + \langle |v|^2\rangle + ((\langle iv^*\rangle/Y_s + \langle i^*v\rangle/Y_s{}^*)}{\langle |i_T|^2\rangle/|Y_s|^2} \quad (24.5)$$

or

$$F = 1 + \langle |i|^2\rangle/\langle |i_T|^2\rangle + ((\langle |v|^2\rangle/\langle |i_T|^2\rangle))|Y_s|^2(1 + Y_c/Y_s + Y_c^*/Y_s^*). \quad (24.6)$$

The noise current of the source is given by $\langle |i_T|^2\rangle = 4kT_0G_s$. It is convenient now to introduce the terms "noise conductance," g_n, and "noise resistance," R_n, which are defined by

$$\langle |i|^2 \rangle = 4kT_0 g_n \quad \text{and} \quad \langle |v|^2 \rangle = 4kT_0 R_n. \tag{24.7}$$

(The lower-case g_n follows the original notation of Rothe and Dahlke who reserved G_n for the part of i that is correlated with v.) In terms of these parameters we have

$$F = 1 + (g_n/G_\mathrm{s}) + (R_n/G_\mathrm{s})|Y_\mathrm{s}|^2(1 + Y_\mathrm{c}/Y_\mathrm{s} + Y_\mathrm{c}^*/Y_\mathrm{s}^*). \tag{24.8}$$

Replacing Y_s by $G_\mathrm{s} + \mathrm{j}B_\mathrm{s}$ and Y_c by $G_\mathrm{c} + \mathrm{j}B_\mathrm{c}$, this becomes

$$F = 1 + (g_n/G_\mathrm{s}) + (R_n/G_\mathrm{s})(G_\mathrm{s}^2 + B_\mathrm{s}^2) + 2(R_n/G_\mathrm{s})(G_\mathrm{c}G_\mathrm{s} + B_\mathrm{c}B_\mathrm{s}). \tag{24.9}$$

Differentiating with respect to B_s and setting the derivative to zero, we find the optimum source susceptance,

$$B_\mathrm{opt} = -B_\mathrm{c}. \tag{24.10}$$

Differentiating next with respect to G_s and setting that derivative equal to zero gives the optimum source conductance:

$$G_\mathrm{opt} = \sqrt{g_n/R_n - B_\mathrm{c}^2}. \tag{24.11}$$

After a few more steps we find

$$F = 1 + 2R_n(G_\mathrm{c} + G_\mathrm{opt}) + (R_n/G_{\mathrm{s})})([G_\mathrm{s} - G_\mathrm{opt}]^2 + [B_\mathrm{s} - B_\mathrm{opt}]^2), \tag{24.12}$$

which is the desired relation, the Rothe–Dahlke formula.The first term on the right is just T_min. Summarizing these relations between the equivalent external noise generators and the standard noise parameters we have

$$R_n = \frac{\langle |v|^2 \rangle}{4kT_0} \tag{24.13a}$$

$$B_\mathrm{opt} = \frac{-\mathrm{Im}\langle iv^* \rangle}{\langle v^2 \rangle} \tag{24.13b}$$

$$G_\mathrm{opt} = \sqrt{\frac{\langle |i|^2 \rangle}{\langle |v|^2 \rangle} - \left(\frac{\mathrm{Im}\langle iv^* \rangle}{\langle |v|^2 \rangle}\right)^2} \tag{24.13c}$$

$$F_\mathrm{min} = 1 + \frac{2\langle |v|^2 \rangle}{4kT_0}\left(\frac{\mathrm{Re}\langle iv^* \rangle}{\langle |v|^2 \rangle} + G_{opt}\right). \tag{24.13d}$$

The inverse relations are

$$\langle |v|^2 \rangle = 4kT_0 R_n \tag{24.14a}$$

$$\langle |i|^2 \rangle = 4kT_0({G_\mathrm{opt}}^2 + {B_\mathrm{opt}}^2) \tag{24.14b}$$

$$\mathrm{Im}\langle iv^*\rangle = -B_{\mathrm{opt}}\langle |v|^2\rangle \tag{24.14c}$$

$$\mathrm{Re}\langle iv^*\rangle = 4kT_0R_n\left(\frac{F_{\mathrm{min}}-1}{2R_n} - G_{\mathrm{opt}}\right). \tag{24.14d}$$

24.1.4 Devices in parallel

Noise matching allows us to realize the best noise figure, F_{min}, from a transistor. We saw earlier that multiple devices in series (an amplifier cascade) result in an unavoidable increase in noise figure. What about amplifiers in parallel? Identical transistors might be literally paralleled, emitter-to-emitter, base-to-base, and collector-to-collector. Or amplifiers might have their inputs directly connected and their outputs combined with a hybrid coupler. The result for two devices in parallel (see Problem 24.4) is that the resulting amplifier, compared to the individual amplifiers, has half as much noise resistance, twice as much noise conductance, and twice as much correlation admittance. This leaves F_{min} unchanged. It does, however, double Y_{opt}, making it easier to noise match to a low-impedance source, and can be a useful technique, for example, at very low frequencies where transformer matching is impractical.

24.1.5 Noise measure

We pointed out earlier that an amplifier with a good noise figure is of little value if its gain is low (a piece of wire has a perfect noise figure, unity, but has no gain). The application of negative feedback may lower the noise figure of a given transistor but at the same time it will reduce the gain. To recover the lost gain we can cascade two or more amplifiers. We saw in Chapter 17 the expression for F_∞, the noise figure of a cascade of identical amplifiers:

$$F_\infty - 1 = \frac{F-1}{1-1/G}. \tag{24.15}$$

This quantity, called *noise measure*, is a better figure of merit than noise figure in that a low-gain amplifier scores low, even if its noise figure is good. The optimum input network should, therefore, minimize F_∞ rather than F. But the gain, G, depends also on the output network. Haus and Adler [1] showed that the *minimum noise measure* is a fundamental invariant when the gain, G, is taken to be the "available gain" of the device (the maximum gain that can be obtained by varying the output match while leaving the input untouched). Minimum noise measure, like minimum noise figure, is therefore determined by the input network. Haus and Adler proved that the minimum noise measure of a device is left unchanged when the device is embedded in an arbitrary network of lossless reactances. It had been noticed earlier that the noise figure of a device could sometimes be lowered by using feedback. Haus and Adlers' results showed that any improvement in noise figure must be accompanied by a decrease in gain such that the minimum noise measure remains constant. A corollary is that changing the orientation of the device (common-emitter, common-collector, common-base) does not change the minimum noise measure. Circuit techniques, then, cannot produce a breakthrough low-noise amplifier. Any standard configuration can have the best possible noise performance. Circuit techniques, however, can improve bandwidth and stability. If we know the characteristics of a given device, we can calculate the minimum noise measure (best possible noise performance) at every frequency. There is, however, no theory available that tells us how well we can do over a given frequency band. R. N. Fano showed how well impedance matching can be done over a given band but no one has done the same for noise matching.

24.2 Oscillator noise

The frequency stability of an oscillator is determined by changes in ambient temperature and humidity, mechanical vibration, component aging, power supply variations, load variations, and finally, random noise generated by the circuit components – primarily the active element (transistor, tube, op-amp, etc). Except for noise, these effects are systematic, i.e., nonrandom, and can, in principle, be eliminated by using constant-temperature chambers, thermal compensation schemes, and shock mounting. But even when the systematic variations have been reduced to an acceptable level, random variations remain and determine the power spectrum of an oscillator (called "line shape" in spectroscopy). An ideal oscillator would have a power spectrum that is a delta function, i.e., an infinitely narrow unmodulated carrier. Noise broadens the spectrum of any real oscillator. In some applications the concentration of power is important; a specification might state, for example, that at least 90% of the power must be contained in a band within 0.01 Hz of the nominal frequency. More often it is only necessary that the spectral density beyond some distance from the center frequency be very low. In such a case it might be specified that "the noise power in a 1-Hz bandwidth centered 1 kHz from the carrier must be at least 60 dB below the total power." Spectral energy outside the narrow band containing most of the power is referred to as "sideband noise."

In a superheterodyne receiver the noise sideband power of the local oscillator can mix with an incoming signal that is nominally outside the IF passband to produce noise that falls within the passband, as shown in Figure 24.5. An undesired signal that is translated to a frequency outside the IF passband will normally be rejected by the IF bandpass filter. But a noisy L.O. will add noise sidebands to this signal that extend into the bandpass. If the undesired signal is near the bandpass and if it is much stronger than the desired signal, this noise can override the desired signal.

Figure 24.5. Translation of phase noise into the passband of a receiver.

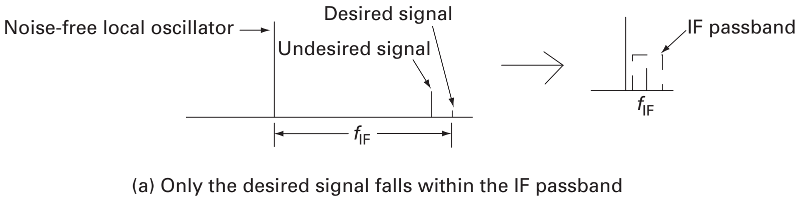

(a) Only the desired signal falls within the IF passband

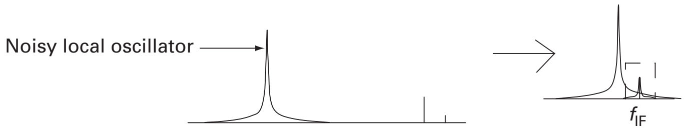

(b) Oscillator noise sidebands on the undesired signal fall within the IF passband

A similar situation occurs in MTI (moving target indicator) radar where the Doppler effect is used to distinguish moving objects from background "clutter" (the ground and stationary objects). Moving targets can be identified as spectral components shifted away from the frequency of the strong clutter echo. But, in as much as the transmitter's master oscillator or the receiver's local oscillator is noisy, moving targets with small velocities will be buried in the broadened clutter return. Good subclutter visibility therefore requires low-noise oscillators.

24.2.1 Power spectrum of a linear oscillator

In the simplest analysis, an oscillator is considered to be a linear system with just enough positive feedback to sustain oscillation. We have already used this approach to design a feedback oscillator. Here we will present a simple analysis [2, 3] that predicts the shape of the spectral line. This shape does not depend on the circuit configuration so here we will consider the simple series-mode oscillator presented in Chapter 11. This circuit, shown in Figure 24.6, uses a noninverting voltage amplifier having voltage gain, *A* (independent of frequency), infinite input impedance, and zero output impedance.

Figure 24.6. Prototype oscillator for phase noise analysis.

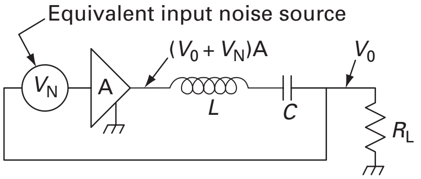

We will put an equivalent generator at the input to account for the noise produced by the amplifier. As long as the amplifier is linear, we need only consider the amplifier's noise spectrum in the vicinity of the oscillation frequency. The feedback voltage is taken from the load resistor forming the bottom leg of a voltage divider whose top leg is the series *LC* circuit. By inspection of Figure 24.6, we can write the loop equation:

$$V_0 = \frac{(V_0 + V_N)AR}{R + \mathrm{j}X} \tag{24.16}$$

where A is the voltage gain of the amplifier and $X = \omega L - 1/(\omega C)$. Note that V_0 and V_N are spectral densities, i.e., their units are volts $/\sqrt{Hz}$. Solving for V_0 we have

$$V_0 = \frac{V_N AR}{R(1 - A) + jX}. \tag{24.17}$$

Expanding X around the resonant frequency, we let $\omega = \omega_0 + \omega'$, where $\omega_0{}^{-2} = LC$. The reactance, X, then simplifies to $X \approx 2\omega' L$ and Equation (24.17) becomes

$$V_0(\omega') = \frac{-V_N A}{(A - 1) - 2\mathrm{j}\omega' L/R}. \tag{24.18}$$

The value of A is determined by the condition $\int |V_0{}^2(\omega')|/R \, \mathrm{d}\omega' = P$, where P is the total output power:

$$P = \frac{1}{R}\int V_0^2(\omega')\mathrm{d}\omega' = \frac{1}{R}\int_{-\infty}^{\infty} \frac{A^2 V_N^2 \mathrm{d}\omega'}{(A-1)^2 + 4\omega'^2 L^2/R^2} = \frac{\pi A^2 V_N^2}{2(A-1)L}. \tag{24.19}$$

Solving for $(A-1)$ we find

$$\frac{A - 1}{A^2} = \frac{\pi V_N^2}{2PL}. \tag{24.20}$$

Note that $A-1$ can be very small; the voltage gain is only slightly greater than unity. Using Equation (24.20) and the fact that A is approximately unity, Equation (24.18) gives the power spectrum:

$$S(\omega) = \frac{|V_0(\omega')|^2}{R} = \frac{4P^2 Q^2 R^2}{\pi^2 V_N^2 \omega_0^2} \frac{1}{1 + (\omega'/\Delta)^2} \tag{24.21}$$

where Δ, the half-width of the line, is given by

$$\Delta = \frac{\pi V_N^2 \omega_0^2}{4PQ^2 R}. \tag{24.22}$$

The $1/(1+x^2)$ "Lorenztian" line shape of Equation (24.21) gives the bell-shaped curve shown in Figure 24.7.

Figure 24.7. Lorentzian line shape.

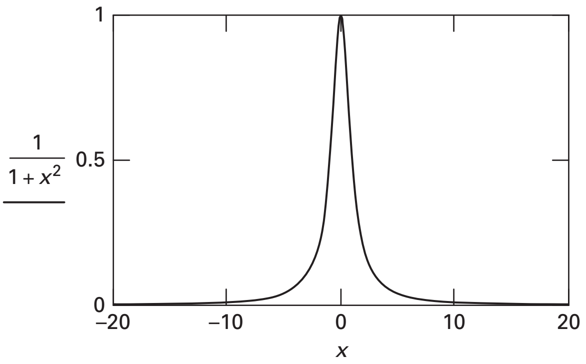

24.2.2 Sideband shape

As explained earlier, it is often important that the far-out sideband noise be low compared to the carrier power. If we assume $\omega' \gg \Delta$ then Equations (24.21) and (24.22) reduce to

$$\frac{S(\omega')}{P} = \frac{\omega_0^2 V_N^2}{\omega'^2 4Q^2 RP}\ \mathrm{Hz}^{-1}, \tag{24.23}$$

and we see that the far-out sideband noise falls off as $1/\omega'^2$.

Phase noise

So far, the oscillator noise has the same character as if it had been produced by amplifying white noise and passing it through a very narrow filter. The oscillator output voltage is equivalent to the sum of a noiseless phasor, i.e., the carrier, and a random noise phasor, as shown in Figure 24.8.

Figure 24.8. Vector diagram showing oscillator noise.

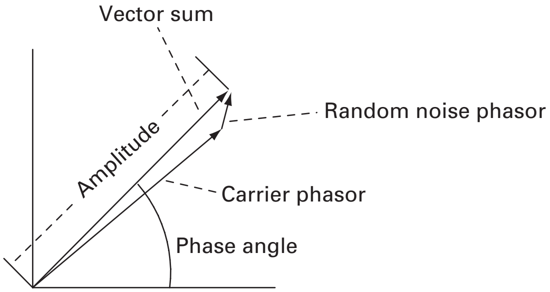

When the noise component is parallel to the carrier, the vector sum has the correct phase but an altered amplitude. When the noise is perpendicular to the carrier, the result is mostly just a phase error. The signal can be passed through an amplitude limiter to remove the amplitude fluctuations. This leaves only the perpendicular noise components, i.e., the phase noise. The parallel and perpendicular noise signals are uncorrelated, so if the amplitude variations are removed, the total noise power is cut in half. The symbo $\mathscr{L}(\omega')$ is often used to represent the phase noise power in a 1-Hz bandwidth at an offset of ω' from the nominal carrier frequency:

$$\mathscr{L}(\omega') = \frac{\omega_0^2 V_N^2}{\omega'^2 8Q^2 RP}\ \mathrm{Hz}^{-1}. \tag{24.24}$$

We can express the noise power $V_N^2(\omega')/R$ as FkT where F is the noise figure of the amplifier, k is Boltzmann's constant, and T is the reference temperature, 290 K:

$$\mathscr{L}(\omega') = \omega_0^2 \frac{FkT}{\omega'^2 8Q^2 P}\ \mathrm{Hz}^{-1}. \tag{24.25}$$

This result shows that best performance is obtained by using a low-noise device, by running it at high power, and by maximizing the loaded Q (by using a high-Q resonator with light loading).

24.3 Effect of nonlinearity

The analysis given above actually gives a lower limit for the noise. Very close to the carrier, oscillators depart from the predicted Lorenztian line shape and, as the offset frequency approaches the carrier, phase noise increases faster than ω'^{-2}. This additional noise is usually attributed to nonlinearity in the active device. We have made the assumption that the device is linear but some nonlinearity *must* be present in any actual oscillator or its amplitude would either decay or increase without limit. The nonlinearity can translate baseband $1/f$ noise up to the radio frequencies.

Problems

Problem 24.1. A transistor with noise parameters, F_{min}, G_{opt}, B_{opt}, and R_{n}, is provided with a noise matching network to produce an amplifier with $F'_{\mathrm{min}} = F_{\mathrm{min}}$, $B'_{\mathrm{opt}} = 0$, and $G'_{\mathrm{opt}} = 1/50$. What is the value of R'_n?

Problem 24.2. Can a transistor be fitted with an input network such that the correlation admittance, Y_{c}, is zero for the overall amplifier?

Problem 24.3. The portion of the noise current, i_{c}, that is correlated with the noise voltage cannot exceed the total noise current, i.e., $\langle |\, i_{\mathrm{c}}|^2 \rangle \leq \langle |\, i|^2 \rangle$. Use this to prove the following inequality:

$$F_{\mathrm{min}} - 1 \leq 4 R_n G_{\mathrm{opt}}.$$

Problem 24.4. Show that when two identical transistors are paralleled, the minimum noise figure of the combination is the same as the minimum noise figure of the individual transistors.

Problem 24.5. A standard amplifier for use in 50-ohm systems has a gain G and a noise figure F. This amplifier is combined with a 50-ohm directional coupler to form the feedback amplifier shown below. The 180° coupler has a power gain of α in its "main channel" and therefore a power gains of $1-\alpha$ in the adjacent arms. The 180° path of the coupler provides the negative feedback. (Assume that the amplifier has zero phase shift.)

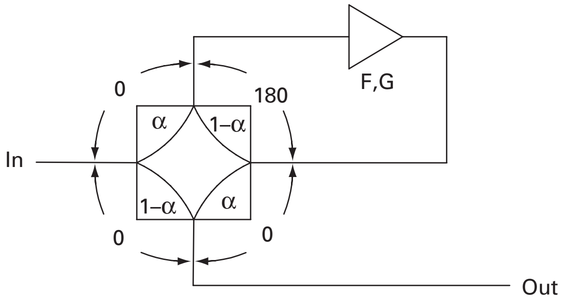

(a) Find the gain, G', of the feedback amplifier.
(b) Find the noise figure, F', of the feedback amplifier.
(c) Combine the results of (a) and (b) to show that the noise measure of the feedback amplifier is the same as the noise measure of the imbedded amplifier, i.e., show that

$$\frac{F'-1}{1-1/G'}=\frac{F-1}{1-1/G}.$$

This problem provides a simple demonstration of noise measure remaining invariant when an amplifier is imbedded in a lossless network. The negative feedback reduces the noise figure but it also reduces the gain and the noise measure stays the same. (What would positive feedback do − short of causing oscillation?)

Problem 24.6. It is certainly possible (and common) for an amplifier to have a noise temperature less than its physical temperature. Is it possible to build an *electronic cold load*, i.e., an active one-port circuit whose impedance is resistive but which generates less noise power than a resistor at ambient temperature?

Problem 24.7. A differential amplifier, e.g., an op-amp, is a three-port device (output and two inputs). Find an equivalent circuit which represents an op-amp as a noiseless device with three external noise sources. Arrange for these noise sources to be on the input side of the amplifier.

Problem 24.8. Which of the following statements best describes the operation of an oscillator?

(a) Noise is necessary to get an oscillator started. Once running, the oscillation is self-sustaining and noise plays no role.
(b) The output of an oscillator consists of amplified noise; the very narrow line shape derives from a resonator whose high loaded Q is effectively increased by the circuit's positive feedback.

Problem 24.9. Discuss the consequences (sound or picture quality) of a noisy oscillator used as a local oscillator in (a) an AM receiver; (b) an FM receiver; (c) an NTSC television receiver; (d) an ATSC television receiver.

Problem 24.10. Sketch a design for an oscillator that somehow regulates its own gain in order to stay as linear as possible, i.e., runs with just enough gain to maintain oscillation.

References

[1] Haus, H. A. and Adler, R. B., Optimum noise performance of linear amplifiers, *Proc. IRE*, Vol **46**, pp. 1517–1533, 1957.

[2] Leeson, D. B., *A simple model of feedback oscillator noise spectrum*, Proc. IEEE, Vol. **54**, pp. 329–330, February 1966.

[3] Robins, W. P., *Phase Noise in Signal Sources*, London: Peter Peregrinus on behalf of the Institution of Electrical Engineers, London, 1984.

[4] Rothe, H. and Dahlke, W., *Theory of noisy fourpoles*, *Proc. IRE*, Vol. **44**, pp. 811–818, June 1956.

[5] Vendelin, G. D., Pavio, A. M. and Rohde, U. L. *Microwave Circuit Design Using Linear and Nonlinear Techniques*, New York: John Wiley 1990.

CHAPTER

25

The GPS navigation system

The basic idea of GPS navigation is simple. A GPS receiver simultaneously monitors ("tracks") the signals from four GPS satellites. Using a highly accurate onboard clock, each satellite effectively stamps its signals with the time at which they are transmitted. The receiver has an internal clock and can, therefore, determine the apparent travel time of the signals from each satellite. Let τ_i denote the travel time of the signal from the i-th satellite. The receiver contains a database and can look up the position of the i-th satellite, x_i, y_i, and z_i, corresponding to the time of transmission. Let x, y, and z represent the (unknown) position of the receiver. For each of the four satellites, *distance = rate × time*, so we can immediately write four equations:

$$\begin{aligned}(x-x_1)^2+(y-y_1)^2+(z-z_1)^2 &= c^2(\tau_1-\Delta)^2\\ (x-x_2)^2+(y-y_2)^2+(z-z_2)^2 &= c^2(\tau_2-\Delta)^2\\ (x-x_3)^2+(y-y_3)^2+(z-z_3)^2 &= c^2(\tau_3-\Delta)^2\\ (x-x_4)^2+(y-y_4)^2+(z-z_4)^2 &= c^2(\tau_4-\Delta)^2,\end{aligned} \tag{25.1}$$

where c is the speed of light and Δ is the offset (error) in the receiver's clock (the same error that made determination of longitude a challenge for early navigators). Because there are four equations and four unknowns, a unique solution can be found for x, y, z, and Δ, thus determining both position and time from scratch. (The receiver's clock error does not matter.) While these equations are nonlinear, the ample processing power available, even in hand-held receivers, makes the solution process easy. An obvious method is to begin with an estimated position, e.g., the last determined position, and then linearize and solve Equations (25.1) in the vicinity of this position.

25.1 System description

The GPS system is operated by the U.S. Department of Defense. In 1993 the system began running a full complement of 24 active satellites, with four satellites

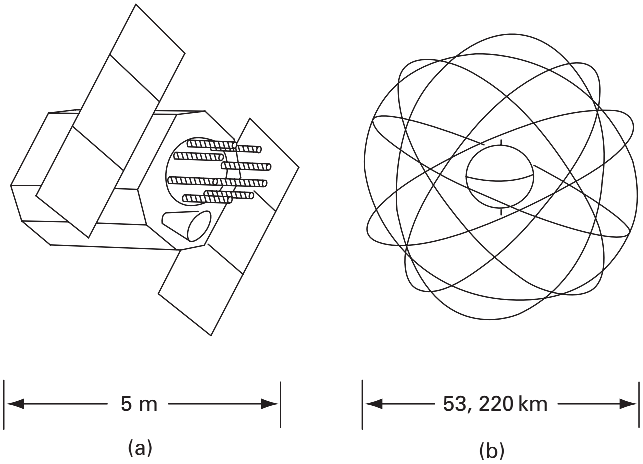

Figure 25.1. (a) Block IIa GPS satellite, 1816 kg; (b) GPS satellite orbits.

in each of six orbital planes, inclined at 55° with respect to the equatorial plane. The satellites were in 12-hour circular Earth-centric orbits, 20 200 km (3.17 Earth radii) above the surface of the Earth as shown in Figure 25.1.

At any point on the Earth, from six to twelve satellites will be above the horizon at any time. As the number of active satellites was increased to 31, the constellation was modified to provide better availability in the event of satellite failures. Orbits are adjusted once a year. The satellites have had operational lifetimes of approximately 10 years.

25.2 GPS broadcast format and time encoding

The system uses two L-band frequencies, 1575.42 MHz ("L1") and 1227.6 MHz ("L2"). Each satellite transmits on both of these frequencies using CDMA (code division multiple access) modulation. The L1 signal contains a "navigation message," described below, which is transmitted repeatedly as a continuous 50-Hz bit stream. Before transmission, this slow bit stream is exclusive-ORed (modulus-2 added) with a one-bit pseudorandom noise (PRN) sequence (the CDMA code) having a high clock rate. This produces a fast bit stream, spread out in bandwidth to approximately the clock rate ("chip" rate) of the PRN code. The L2 signal may also be modulated by the data stream, but is usually modulated only by the PRN code, equivalent to sending a data stream of all 1's. Each satellite is assigned a unique PRN code. In order to receive data from the *i*-th satellite the receiver converts the incoming signal L1 to baseband and multiplies it with a replica of that satellite's PRN code. When the replica, whose binary levels are ±1, is precisely positioned (phased), in order to decode or "strip" the code from the signal, it reproduces the original 50-Hz bit stream, as shown in Figure 25.2.

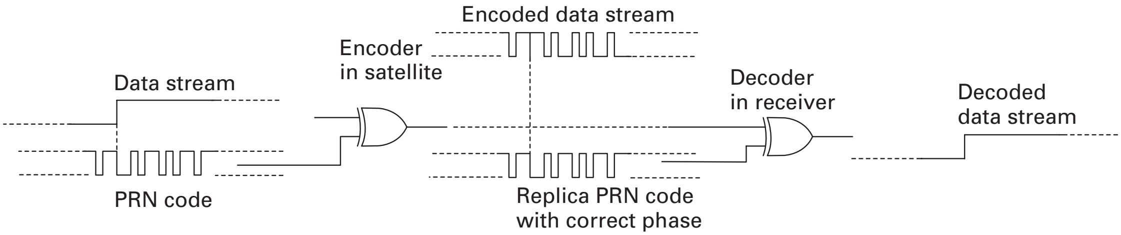

Figure 25.2. CDMA coding (spreading) and decoding (de-spreading).

If the phase of the replica code is shifted off by as little as one chip, the decoder output will be as random and as spread in frequency as its input. This decoding process is the way the receiver selects the desired signal. Signals from other GPS satellites will also be superposed in the received signal, but they have different PRN codes which, when multiplied by the selected code, average to zero for any value of phase shift. Thus, other signals appear as added noise. This added noise is insignificant, in that the received signal is already less than the thermal background noise by some 20–30 dB.[1]

The 50-Hz data streams contain the GPS time,[2] orbital corrections, satellite clock correction, propagation delay effects, and satellite condition, as well as an almanac of long-term satellite data and an auxiliary service, the current difference between GPS time and UTC time. But the precise time of transmission is inferred not from the data but from the phase of the PRN code. We saw in Chapter 21 that just such codes are used for pulse compression in radar to achieve high time resolution (and hence high range resolution) without using narrow high-power pulses. When PRN codes are correlated (convolved) with themselves, a sharp peak is formed, allowing precise time resolution. Thus, PRN coding does double duty in the GPS system; it provides multiple access (many satellites using the same frequency) and it also provides pulse compression for timing accuracy. The *P-code* (*P*recision), a high rate code transmitted on both L1 and L2, has a seven-day period. Therefore, a receiver that has successfully delayed the output of its replica code generator to match the incoming signal, i.e., to acquire the signal, can use that delay value to estimate the signal travel time. The P-code was originally intended only for military use.[3] The other transmitted code is the *C/A-code* (Coarse ranging and Acquisition). This code, always present, along with the P-code, on L1, has a 1.023-MHz chip rate vs. 10.23 MHz for the P-code. It therefore nominally provides only one-tenth the range resolution of the P-code. The C/A code is intended for civilian use; consumer GPS receivers use only the C/A code on L1. The C/A-code repeats every millisecond. This makes acquisition easier, but introduces a

---

[1] Assuming a receiving antenna with a gain of 3dB, the received power of a GPS signal is -160 dBW(10^{-16} watts).

[2] GPS time increases continuously, unlike Universal Coordinated Time (UTC) which is occasionally adjusted by a leap second. GPS and UTC coincided at the start of Jan. 6, 1980, UTC.

[3] Usually the P-code is referred to as the *P*(*Y*)-code because the published P-code can be replaced by a classified *Y-code* for military security.

1-msec ambiguity in the time, which must be resolved using time information transmitted in the data stream.

25.3 GPS satellite transmitter

Figure 25.3 is a functional block diagram of the transmitter aboard a GPS satellite and shows the signals that modulate L1 and L2. All signals are derived from an onboard 10.23-MHz atomic frequency standard. Newer satellites carry two rubidium and two cesium frequency standards. A data formatter produces the continuous one-bit data stream at a 50-Hz baud rate (20 msec/baud). This stream is exclusive-ORed separately with the C/A code and the P-code to create two-bit streams, with respective (main lobe) bandwidths of 1.023 MHz and

Figure 25.3. GPS satellite transmitter block diagram.

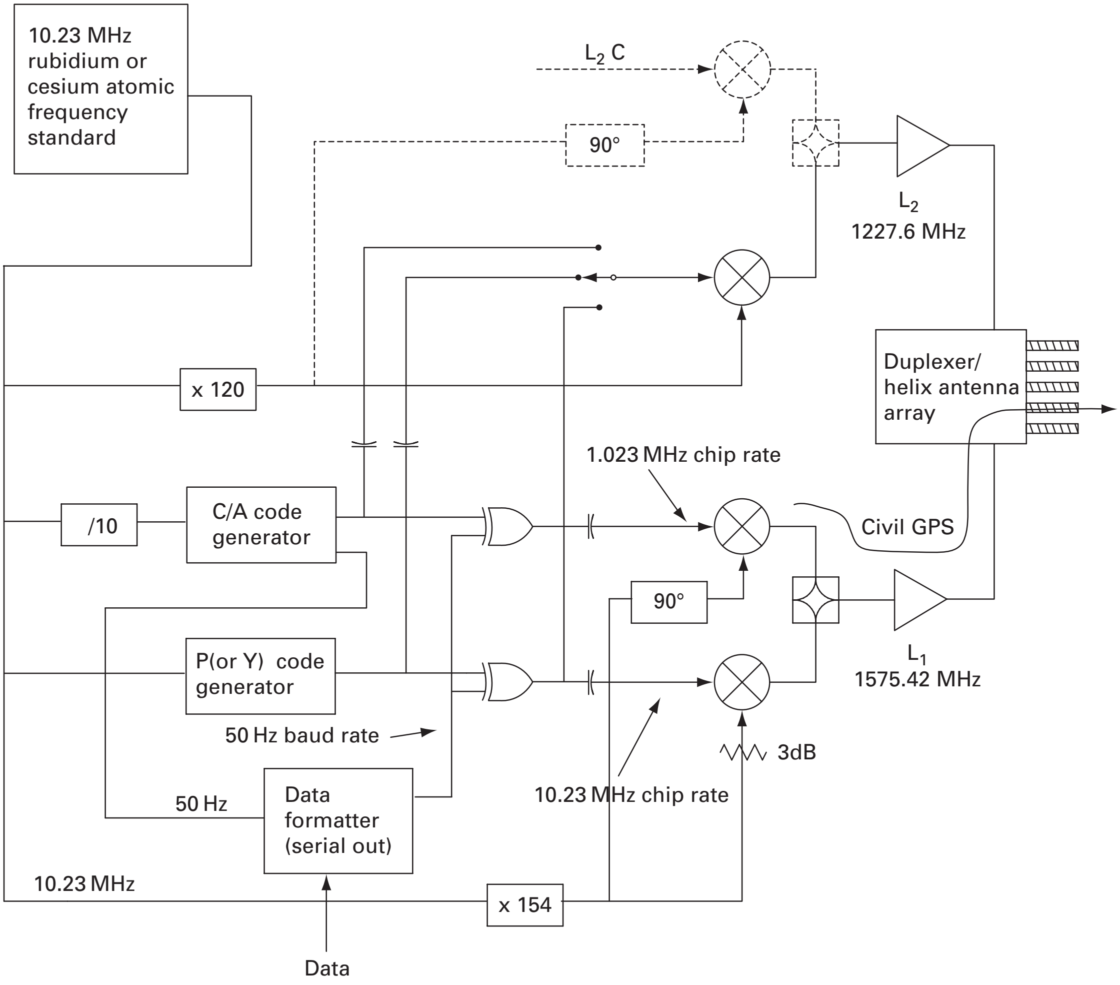
10.23 MHz rubidium or cesium atomic frequency standard
L_2 C
90°
L_2
1227.6 MHz
x 120
Duplexer/ helix antenna array
1.023 MHz chip rate
/10
C/A code generator
Civil GPS
90°
L_1
1575.42 MHz
P(or Y) code generator
50 Hz baud rate
3dB
50 Hz
Data formatter (serial out)
10.23 MHz chip rate
10.23 MHz
x 154
Data

10.23 MHz. The streams are offset to be symmetric about zero volts (indicated by a coupling capacitor on the diagram) and then multiplied by sine waves at 154×10.23 MHz =1575.42 MHz, one shifted 90° with respect to the other. These products are summed to form the L1 signal. The output of each multiplier is a double-sideband suppressed carrier (DSSC) signal. The sum of these two signals, as we have seen, is known as quadrature AM modulation (QAM).[4] The modulating signal fed to each mixer has only two values, $+V$ or $-V$, which determine the polarity of the mixer output, so these mixers can also be called BPSK (binary 0 and 180° phase-shift keying) modulators. The L1 C/A "Civil GPS" signal path is indicated on the diagram.

The original L2 signal is produced by a single mixer, whose signal input can set by the ground control station to be the C/A code, the P-code, or the exclusive-OR of the data and the P-code. Normally the P-code is selected and the L2 signal provides dual-frequency operation for ionospheric delay correction (described below). Newer satellites also provide an "L2C" civil L2 signal, as shown in dotted lines. This signal will provide civil users with dual-frequency capability for high-precision surveying.

25.4 Signal tracking

To see how tracking can be done, we will first consider the case of the L2 signal carrying only the long PRN code and no navigation data. We will then describe the somewhat more complicated case of tracking a signal containing the navigation data, exclusively-ORed with the short C/A code.

25.4.1 Tracking the L2 signal

Figure 25.4 is a block diagram of a hypothetical receiver set up to track this signal. The IF band, which is 20.46 MHz wide to contain the P-coded signal, is established by a bandpass filter which precedes the mixer but follows the low-noise amplifier (LNA), in order that the filter loss not degrade the receiver's noise figure.

Surface acoustic wave (SAW) filters are common in this very small fractional bandwidth application. A 20-MHz lowpass filter (LPF), probably not extending down quite all the way to dc, further defines the bandpass and eliminates the sum frequency component from the mixer. The IF signal is sampled at 41 MHz

---

[4] QAM is normally used to transmit two independent signals which will be independently recovered. In this case, however, the two signals have been made independent by the coding, so they could be algebraically added, with the sum feeding a single mixer. This would introduce "self noise," but a GPS receiver is dominated by thermal noise, so the system performance would not suffer. The QAM system has the advantage of producing a constant-amplitude signal which does not require a linear power amplifier.

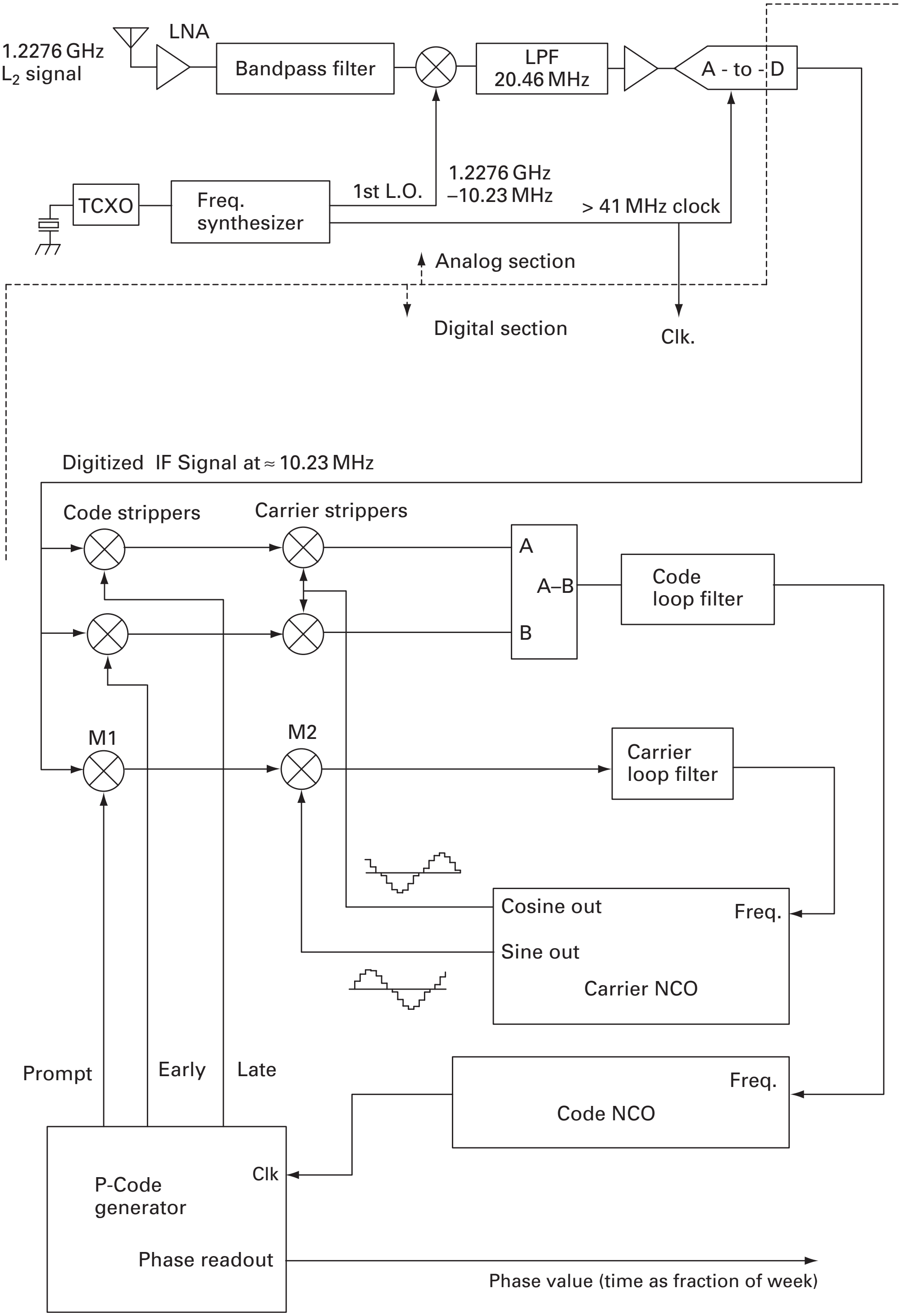
1.2276 GHz
L_2 signal
LNA
Bandpass filter
LPF
20.46 MHz
A - to - D
TCXO
Freq.
synthesizer
1st L.O.
1.2276 GHz
−10.23 MHz
> 41 MHz clock
Analog section
Digital section
Clk.
Digitized IF Signal at ≈ 10.23 MHz
Code strippers
Carrier strippers
A
A–B
B
Code
loop filter
M1
M2
Carrier
loop filter
Cosine out
Freq.
Sine out
Carrier NCO
Prompt
Early
Late
Freq.
Code NCO
Clk
P-Code
generator
Phase readout
Phase value (time as fraction of week)

Figure 25.4. A GPS navigation receiver tracking the L2 signal.

(Nyquist sampling at twice the signal bandwidth to avoid aliasing). From this point on, the receiver uses all-digital processing.

The digital portion of the block diagram includes the blocks necessary to track the L2 signal from one satellite. This entire section would be duplicated n times in a receiver that receives signals simultaneously from n satellites. The receiver synchronously (coherently) detects the signal by stripping off the PRN code and also mixing down to dc (stripping off the carrier). These operations can be done in either order. Once these stripping operations are working properly, the phase of the replica code will be aligned with the code on the incoming signal, allowing the time of transmission to be inferred, as described above. The L.O. used to strip the carrier is the *carrier NCO*, a numerically controlled oscillator in a digital phase lock loop. To convert the signal to precisely zero frequency, the receiver must tune the carrier NCO to compensate for Doppler shift, caused by line-sight-velocity, and by inevitable error in the first L.O. frequency.

Mixer M1 (a digital multiplier) multiplies the digital IF signal by a correctly phased replica of the ±1 PRN code for the desired satellite. This strips the code from the signal, leaving just the $\approx$ 10.23 MHz carrier signal. This carrier signal feeds mixer M2, as the reference signal for the PLL. Two additional code stripper mixers are provided. One is fed by the replica code, advanced by half a chip (*early* code), while the other is fed by the replica code, delayed by half a chip (*late* code). The outputs of these code strippers are mixed down to dc. The carrier NCO is part of a conventional PLL, so its phase is in quadrature with the phase of the reference. Therefore, a version shifted by 90° (the "cosine" output) is provided to synchronously demodulate the code-stripped early and late signals. When the code is in perfect alignment with the satellite signal, the demodulated early and late signals will have equal average values, half the value they would have if they had been on time ("prompt"). But if the alignment shifts slightly, the demodulated early and late signals will be different. A subtractor forms this difference and uses it to increase or decrease the frequency of the *code NCO*, which is the time base for the code generator. Thus, the code loop is also a conventional PLL.

Note that the two loops, while they can run independently, are related. Suppose that, while both loops are tracking a satellite signal, the line-of-sight velocity suddenly increases. The increase in Doppler shift causes the carrier loop to shift the L.O. by an amount $\Delta\omega/\omega_0 = \Delta v/c$. Since range = velocity × time, we can expect that the code tracking loop will shift the code NCO by the same fractional amount. In fact, the code loop can be assisted by properly scaling the carrier NCO frequency control value and adding it directly to the code NCO's frequency command. The scaling is chosen to make the fractional frequency shifts equal. Using this *assisted tracking* technique, the bandwidth of the code loop can be narrowed, for more accurate tracking. Note that it does not matter that the carrier NCO frequency is determined partly by frequency error in the first L.O., because the code NCO loop will pump up a bias to counteract this non-Doppler component.

25.4.2 Tracking the C/A code on the L1 signal

Figure 25.5 shows a hypothetical consumer GPS receiver tracking the L1 short C/A code-plus-data signal on L1. The RF section is tuned to L1, and, for the slower chip rate code, the IF bandwidth is now just 2.046 MHz. The presence of the 50-Hz data stream will cause the polarity of the carrier stripper outputs to change sign, according to whether a given data bit is a one or a zero. This will defeat the operation of the conventional PLLs used above for a signal with no data. A remedy for this is to use loops that detect the signals before subtracting them to form the loop error signals.

Consider the loop controlling the code NCO. The detected early and late signals are lowpass filtered, i.e., time averaged, over an interval substantially shorter than one data baud. Then they are squared (detected) and the difference of these power values forms the loop's error signal. The hypothetical carrier loop shown here works in the same way. Two carrier stripping mixers are fed quadrature L.O. signals (−45° and +45°). When the loop is locked, these mixers will produce equal powers. The carrier NCO also provides a 0° output, in phase with the original carrier, which feeds mixer M2 which strips the carrier to produce the data stream. The 0° carrier is also used for the code loop.

Since the C/A code repeats every millisecond, rather than every week, the phase of the code generator provides only the LSBs of the time. The MSBs come from data contained in the navigation message. The 50-baud data is divided into 300-bit subframes, so a new subframe begins every six seconds. Every subframe contains a TOW (time of week) word, which is the GPS elapsed time, measured in units of six seconds (6000 msec) from the start of the current week, that corresponds to the transmission of the first bit in the next subframe. A distinctive header word identifies the beginning of each subframe. (Every subframe also contains the PRN code i.d. number that identifies the satellite[5].) The receiver has a millisecond counter synchronized to the code generator, i.e., the counter is clocked at the each time the code cycles through its starting point. The receiver's control block checks to verify that the MSBs of the millisecond counter are correct in relation to the TOW word. If incorrect, the control block enables the counter to synchronously set itself to the TOW × 6000 at the beginning of the next 1-msec C/A code cycle. The complete time value from the receiver is, therefore, the time, rounded to milliseconds, from the counter plus the code generator phase, in fractions of a millisecond with submicrosecond precision.

25.5 Acquisition

From the discussion of tracking, you can see that neither loop can lock unless the other is also locked; they track simultaneously or not at all. Acquisition of

---

[5] Reference [1] contains a detailed format for a complete 12.5-minute navigation message, made of 25 frames, each having five subframes made of ten 30-bit words.

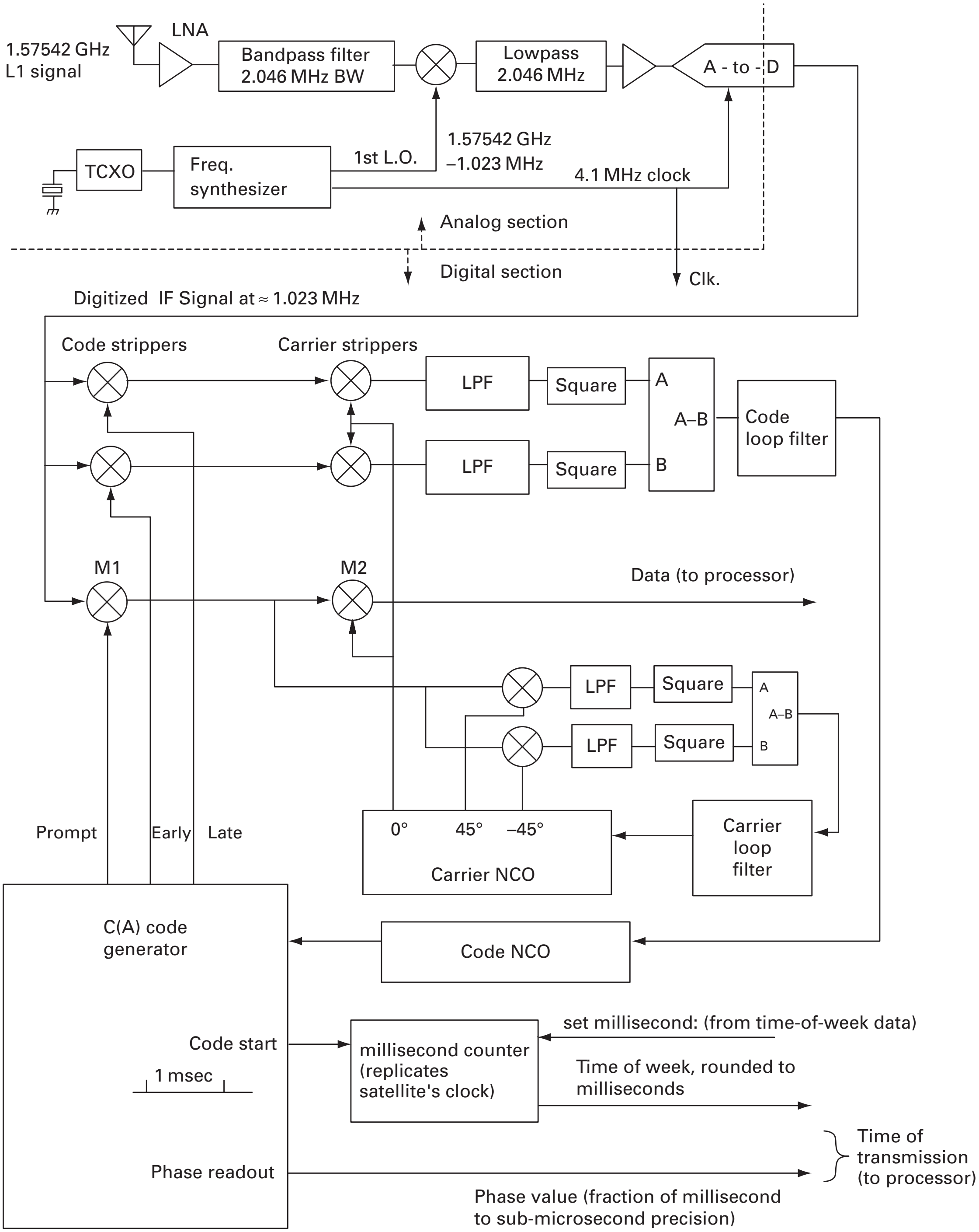
1.57542 GHz
L1 signal
LNA
Bandpass filter
2.046 MHz BW
Lowpass
2.046 MHz
A - to - D
TCXO
Freq.
synthesizer
1st L.O.
1.57542 GHz
–1.023 MHz
4.1 MHz clock
Analog section
Digital section
Clk.
Digitized IF Signal at ≈ 1.023 MHz
Code strippers
Carrier strippers
LPF
Square
A
A–B
B
Code
loop filter
LPF
Square
M1
M2
Data (to processor)
LPF
Square
A
A–B
B
LPF
Square
Prompt
Early
Late
0°
45°
–45°
Carrier NCO
Carrier
loop
filter
C(A) code
generator
Code NCO
Code start
1 msec
Phase readout
millisecond counter
(replicates
satellite's clock)
set millisecond: (from time-of-week data)
Time of week, rounded to
milliseconds
Time of
transmission
(to processor)
Phase value (fraction of millisecond
to sub-microsecond precision)

Figure 25.5. A GPS navigation receiver tracking the L1 C/A signal.

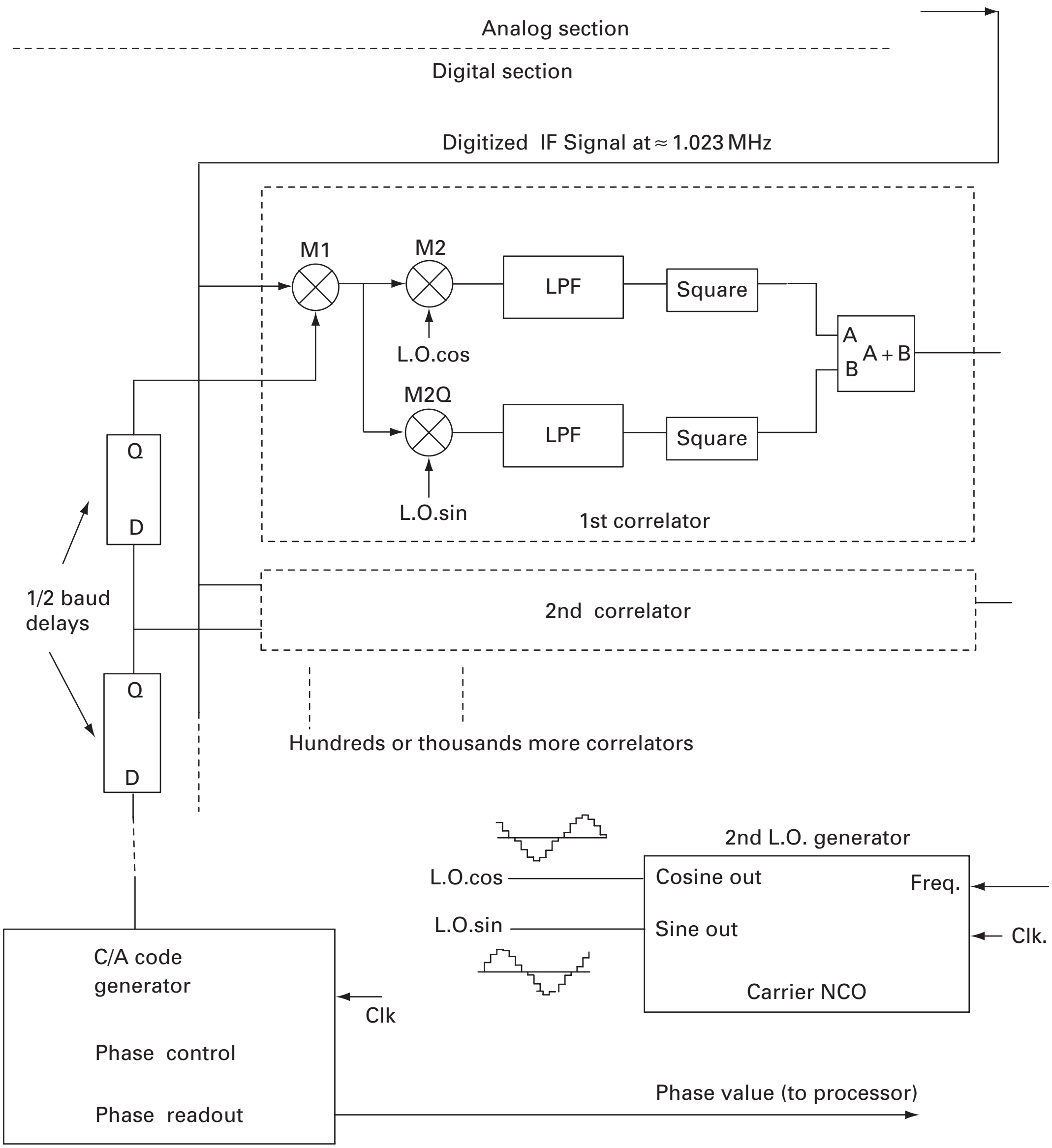
Analog section
Digital section
Digitized IF Signal at ≈ 1.023 MHz
M1
M2
LPF
Square
L.O.cos
M2Q
LPF
Square
A
B
A + B
L.O.sin
1st correlator
Q
D
1/2 baud delays
2nd correlator
Q
D
Hundreds or thousands more correlators
2nd L.O. generator
L.O.cos
Cosine out
Freq.
L.O.sin
Sine out
Clk.
Carrier NCO
C/A code generator
Phase control
Phase readout
Clk
Phase value (to processor)

Figure 25.6. GPS navigation receiver in acquisition mode.

the signal therefore requires that the receiver perform a two-dimensional search over both frequency space and code phase. Enough digital processing power is usually available so that the search can be done in a parallel fashion. Figure 25.6 shows a setup for acquisition processing.

One search process is as follows: Mixers (multipliers) M1, M2 and M2Q are provided for every code phase in increments of half a baud. Successively delayed versions of the replica code are fed to each of the M1 mixers. The M2 and M2Q mixers are all fed from a common L.O. generator. The L.O. frequency is stepped in equal frequency increments. At each frequency, the M2 mixer output signals are averaged (lowpass filtered) over an interval, a "dwell time," and the search ends when the detected power (the square of the sum of the two mixer outputs) exceeds a detection threshold. At that point, the L.O. frequency and carrier phase are close enough for a handover to the tracking mode. Longer dwell times require smaller frequency steps, so it takes more time to search for a weak signal. The changing data bits do not spoil the correlation as long as the dwell times are short compared to the data baud length (unless a dwell happens to straddle a data transition). The circuitry or processing resources for one Doppler/code phase cell is called a correlator. With enough correlators running in parallel, the GPS receiver can simultaneously search multiple frequency shifts, multiple code delays, and multiple satellites. Once the signals have been acquired, relatively little processing power is required for tracking.

The C/A signal is easier to acquire than the P-signal, as it has a higher signal-to-noise ratio and a shorter code (less space to search). Receivers that use the P-code usually acquire the C/A code as a first step in acquiring the P-code. (Hence the designation "C/A:" Civil and Acquisition.)

25.6 Ionospheric delay

GPS signals, at 1.5 GHz and 1.2 GHz, interact only slightly with the Earth's lower atmosphere, but the interaction with the ionosphere contributes the largest term to the measurement error budget. We saw in Chapter 20 that the ionosphere's free-electron gas, produced by ionizing ultraviolet solar radiation, changes the dielectric constant from ϵ_0, its vacuum value, to become

$$\epsilon = \epsilon_0 \left(1 - \frac{Ne^2}{\epsilon_0 m \omega^2} \right) = \epsilon_0 \epsilon_{\mathrm{r}}, \tag{25.2}$$

where N is the electron density, e is the electronic charge, m is the electron mass, and ϵ_{r} is the relative dielectric constant (relative to ϵ_0). The phase velocity of the signal is given by $v_{\mathrm{phase}} = \omega/k = (\mu_0 \epsilon)^{-1/2}$. From Equation (25.2) we see that, in the ionosphere, ϵ is less than ϵ_0. Therefore, v_{phase} is greater than the speed of light and the wavelength, $\lambda = 2\pi/k$, is longer than its vacuum value. Information, however travels at the group velocity,[6] $v_{\mathrm{group}} = \mathrm{d}\omega/\mathrm{d}k$, which, as you can verify, turns out here to be

$$v_{\mathrm{group}} = \frac{c^2}{v_{\mathrm{phase}}} = c\sqrt{1 - Ne^2/(\epsilon_0 m \omega^2)}. \tag{25.3}$$

---

[6] Group velocity: A narrowband signal around $\omega = \omega_0$ can be represented by a Fourier integral as $V(t) = \int F(\omega')\mathrm{e}^{\mathrm{j}(\omega_0+\omega')t}\mathrm{d}\omega'$, where the function $F(\omega')$ is concentrated around zero. When this signal is launched as a wave into a dispersive medium where the wave number, k, can be approximated as $k \approx k_0 + \omega' \mathrm{d}k/\mathrm{d}\omega$, the wave at t, z becomes $V(t,z) = \mathrm{e}^{\mathrm{j}(\omega_0 \mathrm{t} - k_0 z)} \int F(\omega')\mathrm{e}^{\mathrm{j}\omega'(t - z\,\mathrm{d}k/\mathrm{d}\omega)}\mathrm{d}\omega'$. Except for a phase factor (the first term on the right), $V(t,z)$ is identical to $V(t=0, z=0)$ when $z = t\mathrm{d}\omega/\mathrm{d}k = tv_{\mathrm{group}}$.

Thus v_{group} is less than c, and the ionosphere increases the satellite-to-ground propagation delay. The propagation time is just $\int (v_{\mathrm{group}})^{-1}\, \mathrm{d}s$ over the signal path or

$$T_{\mathrm{prop}} = c^{-1} \int \left[1 - \frac{Ne^2}{(\epsilon_0 m\omega^2)}\right]^{-1/2} \mathrm{d}s \approx c^{-1} \int \left[1 + \frac{Ne^2}{2\epsilon_0 m\omega^2}\right] \mathrm{d}s, \tag{25.4}$$

where we have assumed that the second term in the brackets is much less than unity. We see that the additional delay caused by the ionosphere is given by

$$\Delta T_{Iono} = \approx c^{-1} \int \frac{Ne^2 \mathrm{d}s}{2\epsilon_0 m\omega^2} = \frac{cr_{\mathrm{e}}}{2\omega^2} \int N \mathrm{d}s = \frac{cr_e \mathrm{TEC}}{2\omega^2}, \tag{25.5}$$

where $r_{\mathrm{e}} = \mathrm{e}^2/(\epsilon_0\, m\, c^2) = 2.82 \times 10^{-15}$ m (the "classical electron radius"), and TEC = $\int N\, \mathrm{d}s$, the total electron content, i.e., the number of electrons in a one square meter column along the signal path.

Under worst case conditions (the ionosphere enhanced by a solar storm and the satellite near the horizon), the value of TEC might reach 10^{19}, in which case ΔT_{iono} would be 43 nsec, corresponding to a position error of 14 m. Normally the ionospheric error would be less than one-tenth of this.

If the receiver is able to measure the delay on both frequencies, L1 and L2, the delay caused by the ionosphere can be calculated and subtracted from the measured delay. From Equation (25.5), the difference between the measured propagation delays will be $T_{\mathrm{L2}} - T_{\mathrm{L1}} = \frac{1}{2}\, c\, r_{\mathrm{e}}\, \mathrm{TEC}\, (1/\omega_2^{\,2} - 1/\omega_1^{\,2})$. This provides the value of TEC, from which we can calculate the true delay,

$$T = T_{L1} - \frac{\omega_2^2 (T_{\mathrm{L2}} - T_{\mathrm{L1}})}{2(\omega_2^2 - \omega_1^2)}. \tag{25.6}$$

25.7 Differential GPS

Over a limited area, GPS receivers tracking the same satellites will be subject to essentially the same systematic errors – satellite clock errors, ephemeris errors, ionospheric delay, and even intentional systematic errors. ("Selective availability" limited the accuracy for civil users by introducing pseudorandom errors in the satellite clocks which could be corrected only by authorized users. However, this was discontinued in 2000.) Thus, over the limited area, individual determinations of absolute position will be biased by the same amount, and calculated differential positions between them will be accurate. The U.S. Coast Guard maintains differential GPS reference stations at well-surveyed positions near harbors. They continuously broadcast the current correction to their GPS reading so that users, nominally aboard ships, can apply that correction to their own GPS reading. This system requires a means of transmission – dedicated beacon transmitters or piggybacking data on subcarriers of local FM stations. It also requires that the users be able to receive these signals and have GPS receivers equipped to apply the corrections.

In an even more accurate form of differential GPS used for precise surveying, the phase of the carrier itself is used to get extremely fine time resolution. This interferometric technique can produce accuracy better than 1 cm. This carrier phase tracking mode requires dual-frequency receivers for ionospheric correction. The maximum practical distance from the reference station is about 30 km.

25.8 Augmented GPS

The Federal Aviation Administration (FAA) has set up a large-scale “wide area augmentation system” (WAAS) with about 25 ground reference stations spread through the U.S. Each of these stations monitors all the GPS satellites within view, on both frequencies, L1 and L2. The data are consolidated at a central location where correction data are generated. The measured TEC values are used to create a model ionosphere in the form of a set of vertical TEC values, from which the TEC for an arbitrary path can be estimated. Correction data is also produced for the satellite clock and ephemerides errors. These data are sent to two geostationary satellites, which broadcast them back down to GPS users. The down-link is on the GPS L1 frequency, and provides the same signal strength as the GPS signals. The WAAS signals are spread with the same kind of C/A codes used by GPS, but with different PRN sequences. This allows a GPS receiver, equipped with suitable firmware, to receive correction data from a WAAS satellite as if it were just another GPS satellite.[7] The WAAS data is transmitted at a net rate of 250 bits/sec. Convolutional encoding for forward error correction is used at a rate of 1/2, so the data stream is 500 bits/sec. In practice, the use of WAAS data reduces C/A code (“Standard Positioning Service”) GPS position errors to about 2 m.

25.9 Improvements to GPS

Replacement GPS satellites include the new L2C (Civil) signal, whose signal path is shown at the top of Figure 25.3. This signal is actually two time-multiplexed signals, a medium-length code, CM, and a long-length code, CL. The CL code is transmitted without data. It can, therefore, be tracked with the system shown in Figure 25.4, which uses conventional phase lock loops which provide better tracking than the loops used in Figure 25.5, for a signal containing the data stream. The CL2 signal also provides dual-frequency operation so that civil GPS receivers can correct for ionospheric delay. The CM code contains the data, but convolutionally encoded for forward error correction.

[7] The WAAS satellites contain atomic frequency standards and can, in fact, be used as additional GPS navigation satellites.

This slows the net data rate to 25 bits/sec. Both the CM and CL codes have a chip rate of 511.5 kHz. The CM code repeats every 20 msec, while the CL code repeats every 1.5 sec. The L2C multiplexes CL with CM+Data by alternating between them every 1/1.023 microseconds.

25.10 Other satellite navigation systems

The Russian *Glonass s*ystem and the European *Galileo* system (scheduled for operation in 2013) are very similar to GPS. Both are L-band systems, using PRN coding for time resolution. The Glonass system, however, does not use the PRN coding for multiple access. Instead, each satellite transmits on its own assigned frequency, which requires that receivers be tunable. The Chinese *Beidou* system began with two geostationary satellites, but is expected to eventually have five geostationary and 30 medium Earth orbit satellites. All of these systems were preceded by the interesting *Transit* or NAVSAT system, which operated from 1964 until 1991. A ship at sea could locate its position within 200 m by analyzing the curve of Doppler shift vs. time from a single satellite pass. With only five satellites, a given user could get a position fix only once every several hours.

Problems

Problem 25.1. The nominal GPS signal power is −160 dBW (10^{-16} W) at the output of the receiver antenna (at the L1 frequency assuming the antenna gain is 3 dB). Assume an antenna temperature of 290 K, since the low directivity of the antenna causes ground noise to dominate. (a) Assume that the receiver noise figure is 4 dB and calculate the signal-to-noise ratio for a bandwidth of 2.046 MHz (the nominal bandwidth of the C/A signal). (b) Use the Shannon formula, $C = B \log_2 (1+S/N)$, to calculate the channel capacity in bits/sec. Compare this with the 50-Hz data rate of the navigation signal.

Problem 25.2. Estimate the gain of the GPS satellite antenna, assuming the beam angle just encloses the Earth. Then use the information given in Problem 25.1 to estimate the power of the satellite transmitter.

Problem 25.3. "Selective Availability" (SA), a now-discontinued GPS security feature, degraded the navigation accuracy for civil users. Suggest some ways to implement this feature.

Problem 25.4. Describe how you would use off-the-shelf RF components, a deep-memory sampling oscilloscope, and a PC to make an "off-line" detection of a GPS satellite signal.

Problem 25.5. Devise an algorithm and write a computer program to solve the set of equations (25.1) for x, y, z, and Δ.

Problem 25.6. Derive an expression for the Doppler shift of the GPS signal, assuming the orbit passes directly overhead.

References

[1] Navstar global positioning system interface specification IS-GPS-200, Revision D, 7 December 2004 Navstar GPS Space Segment/Navigation User Interfaces. This 193-page document provides a full description of the GPS satellite signals including navigation message formats, code generator details, and algorithms to make corrections using the data in the navigation message. You can find this document in PDF format by searching for "IS-GPS-200."

[2] Global positioning system standard positioning service Performance standard October 2001, Assistant secretary of defense for command, control, communications, and intelligence. This document provides general information on the system architecture and performance. To find this document in PDF format, search on "GPS SPS performance."

[3] Kaplan, E. D. and Hegarty, C. J. Ed., *Understanding GPS* 2nd edn; Boston: Artech House, 2006.

[4] Misra, P. and Enge, P., *Global Positioning System*, Ganga-Jamuna Press 2001.

[5] Prasad, R. and Ruieri, M., *Applied Satellite Navigation using GPS, Galileo and Augmentation Systems*; Boston: Artech House, 2005.

CHAPTER

26

Radio and radar astronomy

Radio astronomy was discovered accidentally in 1931 by Karl Jansky, a physicist at Bell Telephone Laboratories. Jansky had been assigned to identify the sources of noise encountered in a newly installed transatlantic short-wave radiotelephone service. Using a directional receiving antenna on 20.5 MHz, he observed that one component of the noise, a wideband hiss, had a diurnal variation that reached a maximum intensity on average four minutes earlier each day. Jansky knew that the stars advance in just this way (in siderial time) and deduced that the source of the hiss must be outside the solar system. His observations showed that this "cosmic noise" came from the galactic plane and was strongest from the direction of the galactic center (in the constellation Sagittarius).

After Jansky, the second pioneer of radio astronomy was a radio engineer, Grote Reber, who in 1937, built a 9-m (30-ft) parabolic reflector beside his house in Wheaton, Illinois. This was maybe the first modern dish antenna. Reber began his observations using a receiver at 3 GHz, which pushed the high-frequency state of the art, because he assumed that cosmic radio noise was the low-frequency tail of the thermally generated radiation spectrum from white-hot stars. The intensity of this radiation would increase as the square of the frequency, so using the highest practical frequency would make detection easier and would also make his antenna more directive. Detecting nothing at 3 GHz, he worked his way down finally to 160 MHz, where he was able to make contour maps of the cosmic noise intensity. The radiation he and Jansky observed is now known to be *synchrotron radiation*, caused by the centripetal acceleration of fast, i.e., non-thermal, electrons spiraling in a magnetic field. By the end of World War II, the Sun (an ordinary thermal source under low sunspot conditions) had been detected at microwave frequencies. After the war, a previously-predicted spectral line at 21 cm (1420 MHz) was quickly detected. This famous neutral hydrogen line corresponds to the energy difference between the parallel and antiparallel orientations of the magnetic moment of the nucleus (the proton) with respect to the magnetic moment of the spinning electron. Many radio telescopes have been built in the decades following the war, the largest being the

305-m (1000-ft) diameter dish built by Cornell University at Arecibo, Puerto Rico. Discoveries in radio astronomy include some one hundred atomic recombination and molecular lines, pulsars, natural masers, and the isotropic 3K blackbody cosmic background radiation, a remnant left from the Big Bang.

26.1 Radiometry

Most of the radio sources found in nature emit wideband noise; their radiation comes from a great number of individual radiators whose contributions add randomly to produce *Gaussian noise*. (A histogram of voltage samples from the antenna terminals forms a Gaussian curve centered on zero.) Since such a signal is itself a random process, we can only measure its average properties. The most important of these is the average of the square of the voltage, i.e., the power. If we take, say, several n-minute averages of the power, these averages will be scattered around the true average. If n is made larger, the averages will be distributed more tightly about the true average. It might not take long to measure the receiver output power to a precision of, say, 10% or even 1%. But that is almost never long enough to measure the power of a radio source to the same precision or, for that matter, even to detect a source. The problem is that the power from the source is masked by other sources of noise including receiver noise, antenna noise, and cosmic background noise. When an astronomer is trying to detect a source in a certain direction, the first step is to average the power received from that direction, the on-source direction. The next step is to measure the power from a nearby, but off-source, direction. Finally, the latter "off" power is subtracted from the former "on" power. For weak sources (most sources), these powers are almost identical and might correspond to a system temperature[1] of, say, 100 K. Yet the astronomer may need to detect a source that raises the system temperature by only 10 mK. This requires that the "on" and the "off" powers both be measured to an accuracy of, say, $3.3/\sqrt{2} = 2.3$ mK for a 3-sigma detection. The fractional accuracy of the "on" and "off" power measurements must therefore be $0.0023/100 = 2.3 \cdot 10^{-5}$ or about one part in 50 000 (see Problem 26.1).

If we average N samples of the squared voltage, the relative standard deviation, $\delta P/P$, will be $\sqrt{2/N}$ if the voltage has a Gaussian distribution and if all the samples are independent (see Problem 26.2). A signal from a channel of bandwidth, B, can furnish $2B$ independent samples every second. Integrating for a time T we can therefore collect $2BT$ independent samples and the relative standard deviation of the power measurement will therefore be

---

[1] A *system temperature* of 100 K, for example, means that the equivalent noise power at the receiver input is the same as the noise power from a resistor at 100 K. This equivalent noise power is the sum of the actual noise power (sky noise from the main lobe, ground noise picked up from back-lobes, some thermal noise if the antenna is lossy, plus a contribution representing the noise generated in the receiver.

$$\delta P/P = (BT)^{-1/2}. \tag{26.1}$$

Usually a radio astronomer expresses power in terms of antenna temperature and would write $\delta T/T_{\text{system}}$ rather than $\delta P/P$. Equation (26.1) is commonly known as the "radiometer equation." Note that sensitivity increases with bandwidth, contrary to many communication situations where increasing the bandwidth just increases the noise and decreases the signal-to-noise ratio. We arrived at Equation (26.1) by considering digital (discrete) signal processing. It is also common to use an analog square-law detector followed by an analog lowpass filter, which does the averaging. When the lowpass filter is just a simple RC integrator, an analog signal analysis shows that the radiometer equation can be written as

$$\delta T/T_{\text{sys}} = (2BRC)^{-1/2}. \tag{26.2}$$

26.2 Spectrometry

Many sources produce "colored" noise rather than white noise, i.e., the flux density from these sources varies with frequency. Often these variations reveal characteristic shapes of atomic or molecular lines. Instead of just measuring the total power, the signal is divided into adjacent frequency "bins" and the power in each bin is measured. When a spectrum of bandwidth B is divided into n frequency bins, each of bandwidth B/n, the radiometer equation shows that the integration time must be increased by n. It is therefore especially important in radio astronomy, where weak sources may require many hours of integration, to measure the n individual spectra simultaneously rather than sequentially. Simultaneous analysis is done with multichannel radiometers ("multiplex" spectrum analyzers). The simplest multiplex spectrometer is just a bank of n filters, each followed by its own square-law detector and averager. Such an instrument is called a filterbank, but might better be called a radiometer bank. Today, most radio spectrometry is done digitally, often using digital *autocorrelators* (see Chapter 27).

26.3 Interferometry

The classic single-dish radio telescope such as Reber's backyard dish or the Arecibo dish has an angular resolution which is diffraction limited; the angular size of the beam (between half-power points) is about λ/D radians where D is the diameter of the dish. Such a telescope can make maps by scanning the vicinity of a source and doing radiometric averages at each point, but the resolution of the map is limited to λ/D. Better resolution requires a larger antenna. Interferometry uses more than one antenna to form a beam whose size corresponds to the *spacing* between the antennas rather than their diameters. The simplest interferometer has just two elements. The beam formed by the two antennas together has a series of narrow lobes in one dimension. If a point source moves across these lobes, the radiometer will trace out the lobes, just as a point source moving across the beam of a single dish antenna traces out the beam pattern. (Of course the interferometer's fine lobe structure is multiplied by the broader beam pattern of the individual antennas.) The multilobed pattern of a two-element interferometer is unsuitable for the intensity mapping described above but it is often used to set an upper limit on the angular size of a source.

26.3.1 Imaging interferometry

By using data from multiple antennas, it is possible to synthesize a beam that is small in both dimensions, i.e., a beam that would correspond to a filled-aperture antenna whose diameter is the size of the interferometer array. The VLA (Very Large Array) at Socorro, New Mexico, is an example of such a system. Signals from 27 antennas are transmitted up to 21 km through optical fibers (originally through low-loss circular waveguides) to the central processing laboratory. The VLBA (Very Long Baseline Array) has 10 antennas with spacings as large as the distance from Hawaii to St. Croix in the Virgin Islands. Its signals, together with timing information, are recorded on magnetic media at each station and sent, physically or electronically, to New Mexico for after-the-fact combining and processing. Imaging interferometers (phased arrays) work essentially as follows. Suppose we have an array of antennas such as shown in Figure 26.1. Assume the antennas are all pointed in the same direction, say up. If the voltages from the antennas are all added together in phase, squared, and averaged, the result is a narrow beam, pointing straight up. This process can be written as follows:

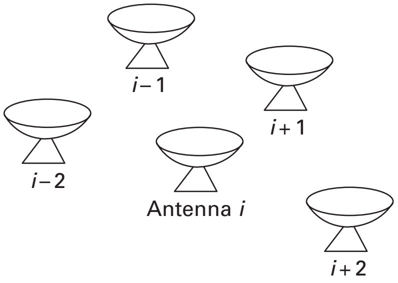

Figure 26.1. An interferometer antenna array.

$$P_0 = \left\langle \sum_n V_n^2 \right\rangle = \sum_{n,m} \langle V_n V_m^* \rangle. \tag{26.3}$$

Now suppose we add the voltages, multiplied by a progressive phase shift, to form a beam tilted slightly off the vertical. If we do this addition, followed by squaring and averaging, we get the *k*-th beam:

$$P_k = \left\langle \sum_{n,m} V_n \mathrm{e}^{\mathrm{j}\phi_{n,k}} V^* \mathrm{e}^{-\mathrm{j}\phi_{m,k}} \right\rangle = \sum_{n,m} \mathrm{e}^{\mathrm{i}(\phi_{n,k}-\phi_{m,k})} \langle V_n V_m^* \rangle. \tag{26.4}$$

This is the general case; for the straight-up beam, $\phi_{n,k}$ and $\phi_{m,k}$ are both zero. The important thing to notice is that we can multiply the signals from every possible antenna pair and average the products first, independent of what beam we wish to form. After averaging these pairs we can *then* form each beam by weighting the averaged products with the appropriate phase factors and performing the double sums. Note also that the averaged products do not have to be measured simultaneously; we could even use one fixed antenna and one mobile antenna and measure the average product for one baseline after another. The VLA has 27 antennas so it can measure (27 × 26)/2 baselines simultaneously. But, as the Earth turns, these baselines become different baselines so, over time, it is possible to collect averaged products from a huge number of baselines. This set of baselines is virtually the same as all the baselines one could form between pairs of points on a dish with a radius of 21 km, and the final synthesized beam can be as sharp and clean as if it had come from a filled dish. An array can thus have the resolution of an enormous antenna even if it does not have the corresponding sensitivity (collecting area). On the other hand, because the individual antenna elements have relatively wide beams, an array has a relatively wide field of view. Radio astronomers hope to have, by the year 2020, a super large array, the SKA (see reference 5) which would have a total collecting area of one square kilometer and provide baselines as long as 3000 km.

26.4 Radar astronomy

Ionized meteor trails produced echoes on WWII radars, but echoes from the Moon were not observed until after the war. In January 1946, a U.S. Army Signal Corps group headed by John H. DeWitt, Jr. observed real-time Moon echos using modified WWII radar equipment delivering 3 kW of power at 110 MHz to a 64-element dipole antenna array. In Hungary, one month later, Zoltán Bay used similar equipment, together with electrolytic cells (“coulombmeters”) to do signal integration. By the end of his experiment, the on-target cell had accumulated significantly more gas than the off-target cell. Many large radars were built in the late 1950s, and radar echo detections from the Sun, Venus, Mercury and Mars were made in the 1960s. Radar was used to measure the rotation rates of Venus and Mercury. Saturn’s rings were detected in the 1970s, as were the large moons of Jupiter, many asteroids and several comets.

26.4.1 The Moon

The Moon’s distance from the Earth, R, is 3.8×10^5 km, and it has a *radar cross-section,* σ, of about $6.6 \times 10^5\ \mathrm{km}^2$. This means that the power density of the echo's SR, in watts per square meter, received back on the radar antenna will be given by $\sigma/(4\pi R^2)$ times the power density incident on the target, S_{inc}. (Note that this defines *radar cross-section* as the collecting area of an equivalent target that isotropically scatters the intercepted incident power.) The incident power density is just $S_{\text{inc}} = P_T G/(4\pi R^2)$, where P_T is the transmitter power and G is the antenna gain. The power into the receiver, P_R, is equal to $S_R A_{\text{eff}}$ where A_{eff} is the effective collecting area of the antenna. We have seen earlier that G is equal to $4\pi A_{\text{eff}}/\lambda^2$. The Moon's radar cross-section is about 7% of its geometric cross-section, i.e., it is 7% as reflective as a metal sphere of the same size. An echo from the Moon will therefore produce a power at the receiver of

$$P_R = P_T \frac{4(\pi A_{\text{eff}}/\lambda^2)}{4\pi R^2}\frac{\sigma}{4\pi R^2}A_{\text{eff}} = \frac{P_T A_{\text{eff}}^2 \sigma}{4\pi R^4 \lambda^2}. \qquad (26.5)$$

Suppose we have a radar with a 1-kW transmitter and a modest 5-m diameter dish antenna with an aperture efficiency of 50%. The effective area of this antenna is therefore half of its geometric cross-section or $A_{\text{eff}} = 0\ .5 \times \pi (5/2)^2 = 9.8\ \text{m}^2$. It is an experimental fact that σ for a planet is approximately independent of wavelength, so we see from Equation (26.5) that the received power is inversely proportional to λ^2 (because of the fundamental antenna relation between gain and effective area). Suppose that the wavelength is 30 cm, i.e., the frequency is 1 GHz. With this radar system, the power received from the Moon would be

$$P_R = \frac{1000\ \text{W}(9.8\ \text{m}^2)^2 6.6 \times 10^{11}\ \text{m}^2}{4\pi(3.8 \times 10^8\ \text{m})^4 (0.30\ \text{m})^2} = 2.7 \times 10^{-18}\ \text{W}. \qquad (26.6)$$

Whether this amount of power is easy to detect or not depends on the noise it competes with. Let us assume that the noise from the antenna (sky noise from background cosmic radio sources plus some "spill-over" noise from the surrounding ground) has an equivalent noise temperature, $T_{\text{ant}} = 50$ K. Let us also assume that our receiver has an equivalent noise temperature, $T_{\text{rcvr}} = 35$ K. The total equivalent input temperature is therefore given by $T_{\text{system}} = 85$ K and the noise power will be $kT_{\text{sys}}B$. If we make the bandwidth, B, very small we decrease the noise power and hence detection becomes easier. But if we make B very small we will need a very accurate prediction of the Doppler shift caused by the relative motion of the Earth and Moon in order to tune the return echo into the passband of the filter. Finally, the rotation of the Moon causes a Doppler *broadening* of the return echo; if we make B less than 25 Hz we will begin to exclude some of the broadened signal. Let us compromise and use a bandwidth of 100 Hz. The signal-to-noise ratio at the receiver output will then be given by

$$S/N = \frac{P_R}{kT_{\text{sys}}B} = \frac{2.7 \times 10^{-18}\text{W}}{1.38 \times 10^{-23}\ \text{W/Hz/K} \times 85\ \text{K} \times 100\ \text{Hz}} = 23.0. \qquad (26.7)$$

This SNR of 23 is large enough that the signal would be visible on an oscilloscope connected at the intermediate frequency (IF) voltage output; no signal averaging would be needed. The modest radar system assumed here could be assembled for a few thousand U.S. dollars.

26.4.2 Venus

Venus is not nearly so easy to detect, being some 280 times more distant than the Moon. Its radar cross-section is about 20 times that of the Moon, so with the radar system described above, the SNR of the return echo would be lower by a factor $280^4/20$ or 307 million! This requires a *much* larger radar. The Arecibo radar, however, has an effective antenna diameter of 200 meters and an average power of 1 MW at 2.38 GHz. It would out-perform our Moon example radar by a factor of $(10^6/10^3)\ (2.38/1)^2\ (200^2/5^2)^2$ or 1.4×10^{10}. Thus, with the same assumed bandwidth and system temperature, the Venus echo from the Arecibo radar would have an SNR of $23\ (1.4 \times 10^{10})/(307 \times 10^6) = 1090$. This is overkill for a simple detection but is needed for high-resolution mapping.

26.4.3 Delay-Doppler mapping

Except for the Moon, planetary targets have angular sizes much smaller than the radar beam. Nevertheless, images of photographic quality can be made by a technique known as delay-Doppler mapping. This method uses short pulses (or pulse compression) to obtain adequate range resolution. The relative motion between the radar and the target provides resolution in the transverse direction via the Doppler effect. This was the method used to get the first surface images of Venus, whose cloud cover kept its features hidden to optical telescopes. The technique is essentially the same as side-looking or synthetic aperture radar by which photographic-like images of ground targets are made from an airborne radar with only a small dish antenna. Figure 26.2 shows how delay-Doppler mapping works. This is a view of the planet as seen from the radar. When a short pulse is transmitted, the first echo to return comes from the front cap or sub-Earth point. At subsequent times the echo signal

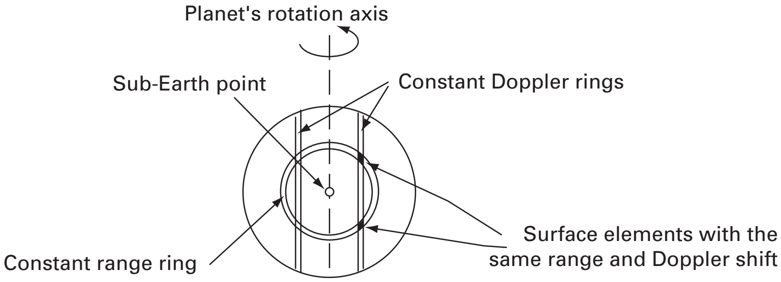

Figure 26.2. Geometry for delay-Doppler radar imaging.

corresponds to loci of equal range that are rings on the planetary surface, centered on the sub-Earth point. Because the planet is rotating, surface points on the right-hand side are moving away from the radar and their echoes are Doppler-shifted to lower frequencies. Likewise, points on the left-hand side are up-shifted. The magnitude of the of Doppler shift is proportional to the cross-range distance, i.e., the distance in the direction perpendicular to both the planet's axis of rotation and the line between the radar and the planet. Loci of constant Doppler shift are rings parallel to the rotation axis of the planet and to the line from the planet to the radar.

The two shaded surface elements have the same range and the same Doppler shift. This fundamental ambiguity can be removed, at least for Venus maps made at Arecibo, by tilting the radar beam slightly to illuminate only the Northern or only the Southern hemisphere of Venus. Obviously this technique requires the narrow beam provided by a very large antenna. Data taking is done as follows: a series of IF voltage samples is read and stored separately as each return pulse arrives. After many pulses have been received, we have a time series for each range ring. Each series is Fourier analyzed to get the Doppler spectrum. The magnitude of a given Doppler component corresponds to the reflectivity of the surface element at the intersection of the range ring and the Doppler ring.

26.4.3.1 Overspreading

The standard delay-Doppler[2] method will not work when the planet is too large and/or spins too fast. This is because the Nyquist sampling theorem requires that the (complex) sampling rate be at least equal to the bandwidth of the signal so that high-frequency Doppler components do not "alias" and appear as lower frequency components (the stroboscope effect). This establishes a minimum pulse repetition frequency (PRF). But if the depth of the planet is too large, even the minimum PRF will result in having more than one pulse on the planet at a time, causing the echoes from different ranges to "fold" together.

The bandwidth of the return echo (twice the highest Doppler shift) is given by $B = 2f_{\text{radar}}\, 2v_{\max}/\text{c} = 4f_{\text{rad}} r_{\text{planer}}\, \Omega\, /c$ where r_{planet} is the radius of the planet, Ω is its apparent angular velocity,[3] f_{radar} is the frequency of the radar, and α is the angle between the radar beam from the Earth and the planet's rotation axis. Therefore, the Nyquist sampling requirement is $f_{\text{sample}} > 4f_{\text{radar}}\, r_{\text{planet}}\, \Omega/c$. The condition to avoid range folding is $1/f_{\text{sample}} > 2\text{r}_{\text{planet}}/\text{c}$. Multiplying these two inequalities gives us $1 \geq 8\, f_{\text{radar}}\, r^2{}_{\text{planet}}\, \Omega/\text{c}^2$, which imposes an upper limit on the frequency of the radar:

---

[2] Note that delay-Doppler mapping of space objects is the same technique as pulse-Doppler radar, discussed in Chapter 21. But while the planet presents a "target" in every range-Doppler cell, a pulse-Doppler surveillance radar is likely to have a target (often an airplane) in only a single range-Doppler cell.

[3] The *apparent* angular velocity of the planet corrects for the planet's tilt and the relative translational (orbital) motion of the Earth and the planet.

$$f_{\mathrm{radar}} < \frac{c^2}{8r^2\Omega}. \tag{26.8}$$

Using Equation (26.8), the radar frequency should be less than about 1.5 GHz for Venus and less than about 1.8 GHz for Mercury. These are practical frequencies for conventional microwave radar technology. For Mars, however, the radar frequency would have to be lower than 14 MHz, because of the high rotation rate. Equation (26.5) shows that such a low frequency would require an extremely large antenna. To make matters worse, sky noise and atmospheric noise at 14 Mz are both much higher than at microwave frequencies.

While standard delay-Doppler does not work for Mars and other overspread targets, Mars images were made in 1988 using the VLA in New Mexico. During the observations, Mars was illuminated with a 10-GHz signal from the JPL NASA Solar System Radar transmitter at Goldstone, California. The resolution of these images was only about 100 km.

A modified delay-Doppler technique, developed around 1986 for radar probing of the ionosphere, was first used at Arecibo in 1990 to solve the overspreading problem for Mars. This "long code" technique is used with the 2.38-GHz transmitter, which puts out a continuous wave (cw) rather than pulsed power, but is biphase-modulated with a long pseudorandom code – for Mars, a sequence of 10-μs bauds. The code elements shift the RF phase by zero or 180°, equivalent to leaving the signal unchanged or changing its polarity. The length of the sequence must be greater than 0.33 seconds (the length of a coherent integration or "look"), though the sequence used at Arecibo repeats only every 3054 hours and can be considered totally random. To see how this technique works, imagine first a simple point target. The echo from such an object would be a delayed version of the code. At the receiver we sample the return echo at the baud rate and multiply successive samples of the echo signal by successive bauds of an identical "replica" code to undo the echo's phase reversals. When the replica code has the correct alignment, it undoes the phase reversals and produces a cw signal. We can separately multiply the echo samples by the replica code for all alignments. Identifying the alignment that produces a cw signal gives us the distance (range) to the point target. If the point target has a velocity component parallel to the radar beam, the recovered cw signal will be Doppler shifted and we can determine the velocity component. Next, imagine that the echo comes from *two* point reflectors, separated in time by at least one baud (1500 m in range). Again multiplying the echo by the replica code, we would find two alignments that produce cw signals and could thereby deduce the range and line-of-sight velocity of each target. Note, however, that the "decoded" signal from either target contains white noise that is the signal from the other target, randomized because the replica code is not aligned with its echo. This "self-noise" has power equal to that of the desired signal, assuming the targets have equal radar cross-sections. We can consider the range rings on a planet to be a set of N such targets, separated in range delay by the

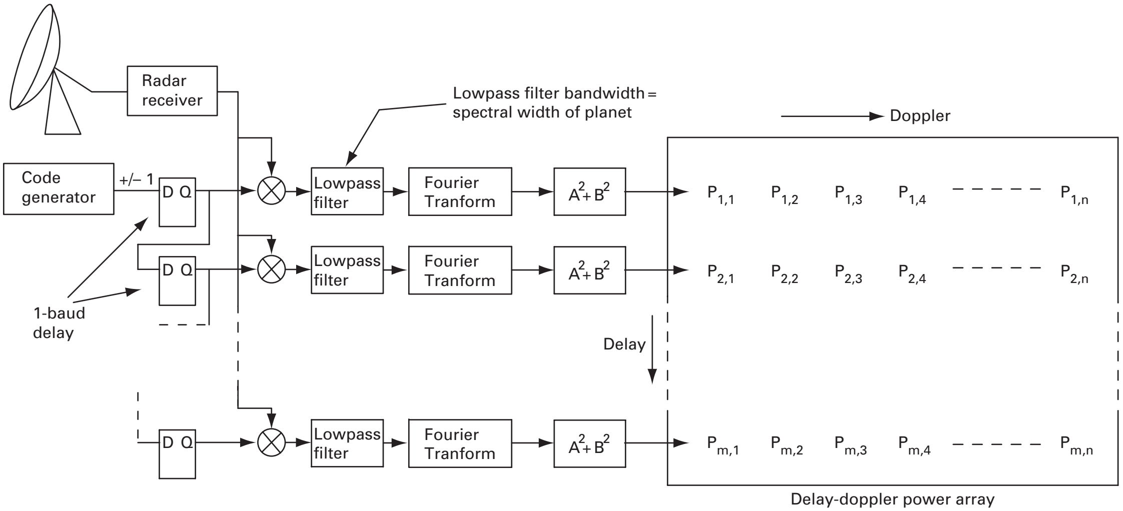
Radar receiver
Code generator
+/– 1
D Q
D Q
D Q
1-baud delay
Lowpass filter bandwidth = spectral width of planet
Lowpass filter
Fourier Tranform
A^2+B^2
Doppler
Delay
$P_{1,1}$ $P_{1,2}$ $P_{1,3}$ $P_{1,4}$ $P_{1,n}$
$P_{2,1}$ $P_{2,2}$ $P_{2,3}$ $P_{2,4}$ $P_{2,n}$
$P_{m,1}$ $P_{m,2}$ $P_{m,3}$ $P_{m,4}$ $P_{m,n}$
Delay-doppler power array

Figure 26.3. Long-code delay-Doppler method.

baud length. (Each range ring has a Doppler spectrum.) The self-noise voltage is therefore the sum of N voltages, which add as a random walk. The delay time depth of Mars is 23 ms so, for a baud length of 10 μs, $N=2300$. If we assume that the signals from the rings are approximately equal in strength, the self-noise power will be N times the signal power. Here, this factor can be reduced to 2300 (7.5/100) = 177 by lowpass filtering the data sequences before doing the Doppler analysis, since the self-noise is spread out over a much larger bandwidth (100 kHz) than the Mars echo bandwidth (7.5 kHz). This still appears disastrous, until we remember that the self-noise is in addition to "real" noise, i.e., cosmic noise, antenna noise, and receiver noise. Under the worst weak-signal conditions, the self-noise is insignificant compared to the real noise so, while the system performance will be poor, the self-noise is not to blame. Under the best conditions, the self-noise dominates the real noise and it will determine the integration time needed to pull the signal out of the total noise. When the long-code technique is used with the Arecibo S-band radar system to observe Mars, the effective signal-to-noise ratio is decreased by a factor of 5.2 or 1.7, depending on the polarization of the echo.[4] Figure 26.3 shows a signal processing arrangement that produces the decoded signals for each range, does a Fourier transform on each decoded range signal, and stores the resulting delay-Doppler data points in an array.

---

[4] The radar transmits a signal with circular polarization. Smooth portions of the planetary surface create specular (mirror-like) reflections, which reverse the sense of the circular polarization. Rough portions of the surface create multipath reflections, which cause some of the echo power (usually less than half) to return with the same circular sense as the transmitted signal. The radar is equipped to receive both polarizations.

Problems

Problem 26.1. Use the radiometer equation to find how much time is needed to make a 3-sigma detection of a 10 mK radio source if the radio telescope's system temperature is 100 K. Assume the predetection bandwidth, B, is 10 MHz. A "3-sigma detection" requires that the 0.01 K contribution from the source be $3\sqrt{2}$ times the δT fluctuation predicted by the radiometer equation. (The factor $\sqrt{2}$ takes account of the increased fluctuation when the "off" is subtracted from the "on", assuming an equal time T is spent measuring each.)

Problem 26.2. Verify that if a noise power estimator is defined as the average of N independent samples of the squared noise voltage, the standard deviation of this estimator will be $\sigma^2\sqrt{2/N}$. Assume that the probability distribution of the noise voltage is Gaussian with zero mean and variance σ^2. Hint: for a zero-mean Gaussian distribution, the expected value of V^2 is σ^2 and the expected value of V^4 is $3\sigma^4$.

Problem 26.3. Pulsars (rotating neutron stars) are radio sources that turn on and off with very regular periods ranging from about 2 milliseconds to about 2 seconds. Given a radio telescope at your disposal, how would you go about searching for pulsars, i.e., what kind of data processing scheme would you use to find these periodic sources?

References

[1] Butrica, A. J., *To See the Unseen: A History of Planetary Radar Astronomy*, Diane Publishing Co., 1997 (also available on the Web).
[2] Goldsmith, P. ed. *Instrumentation and Techniques for Radio Astronomy*, IEEE Press,1988. (This book is prefaced with 12-page historical overview.)
[3] Kraus, J. D., *Radio Astronomy*, 2nd edn, Cygnus-Quasar Books, 1986.
[4] Ostro, S. J., Planetary radar astronomy, *Rev Mod Phys*, Vol. **65**, No. 4, pp. 1235–1279, October 1993.
[5] http://www.skatelescope.org/PDF/SKABrochure_2007.pdf

CHAPTER

27

Radio spectrometry

Spectrometry or spectral analysis is the statistical characterization of random (*stochastic*) signals such as the IF voltage in a radio astronomy receiver, as described in Chapter 26. The spectrometers discussed in this chapter are all *multiplex* spectrometers, meaning that they measure N points on the spectrum simultaneously. This is to be distinguished from swept-frequency spectrum analyzers, which measure spectral points sequentially. Multiplex spectrometers are used when long integration times are needed to pull a signal out of the noise, as in radio astronomy. They are also used for low-frequency spectrum analysis, where narrow channel bandwidths require long measurement times to process a sufficient number of independent samples. Most often the signal is Gaussian; if a fine-grained histogram of samples of the signal's amplitude is scaled to make the area below the curve equal to unity, the average curve will be the Gaussian probability density function, $f(V) = (2\pi\sigma^2)^{-1/2}\ \exp(-V^2/2\sigma^2)$. You can verify (Problem 27.1) that σ is the rms value of V, i.e., σ^2 is the power, $\langle V^2 \rangle$. But total power does not completely characterize the signal. A complete description is contained in the *power spectral density function*, $S(\omega)$, the distribution of power vs. frequency. A set of bandpass filters and power meters, i.e., a set of radiometers, serves to measure points on the spectral density function (usually called the PSD or simply the power spectrum). However, the simplest mathematical definition of the power spectrum uses the signal's *autocorrelation function*, $R(\tau)$, a function of time delay. The value of the autocorrelation function for a given time delay, τ, is defined as the average value of the "lagged product"

$$R(\tau) = \langle V(t) \cdot V(t + \tau) \rangle. \tag{27.1}$$

For a signal whose characteristics are unchanging (a *stationary process*), $R(\tau) = R(-\tau)$. We will see that the Fourier transform of the autocorrelation function is the power spectral density so, directly or indirectly, we describe the signal in terms of sine-wave basis functions, i.e., we find the magnitudes (but here not the phases) of a set of sine waves whose superposition has the same power spectrum as the original signal. Why are sine waves preferred to other sets of basis functions? Often the process under study is a spectral line which shows up clearly in a few adjacent Fourier coefficients (frequency bins) or a Doppler shift which just displaces the spectrum. The frequency spectrum is clearly a natural way to represent such signals. Of course we deal with sine waves all the time as the characteristic functions of linear systems; a sinusoidal signal of a given frequency remains a sinusoid with the same frequency after passing through any arbitrary chain or network of linear elements such as amplifiers, filters, and transmission lines. A variety of instruments have been developed for spectrometry. In approximate historical order, these include analog filterbanks, autocorrelators, chirp-z spectrometers, Fourier transform spectrometers, and acousto-optical spectrometers.

27.1 Filters and filterbanks

We have already seen the analog filterbank in Chapter 26, as part of a radiometer bank. Usually the signal is converted down to a convenient intermediate frequency, e.g., tens of MHz, and the filters are the various *LC* bandpass filters discussed earlier or crystal filters or digital filters. A traditional disadvantage of an analog filterbank spectrometer has been that it is not "elastic," i.e., the filters have a fixed width; if narrower or wider filters are needed, another filterbank must be built. With high-speed digital-to-analog conversion and fast memory, this restriction can be lifted by using an interesting spectral expansion technique. Sequences of data are read into the memory at whatever slower rate is needed and then played back at a higher rate into a single wideband filterbank.

27.2 Autocorrelation spectrometry

Autocorrelation spectrometers estimate the power spectrum by taking pairs of samples with a given time separation, multiplying them, and averaging these "lagged products." (In Fourier transform spectral analysis the samples are multiplied by sine waves but the resulting coefficients are squared before averaging. In both cases powers, rather than voltages, are averaged.) When the time separation is zero, the average lagged product is just the total power. But if we measure other lagged products as well (where the time separation is not zero), we can indirectly find the power spectrum because the autocorrelation function (ACF) and the power spectrum form a Fourier transform pair:

$$S(\omega) = \int_{-\infty}^{\infty} R(\tau) \mathrm{e}^{-\mathrm{j}\omega\tau} \mathrm{d}\tau \tag{27.2}$$

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) \mathrm{e}^{\mathrm{j}\omega\tau} \mathrm{d}\omega. \tag{27.3}$$

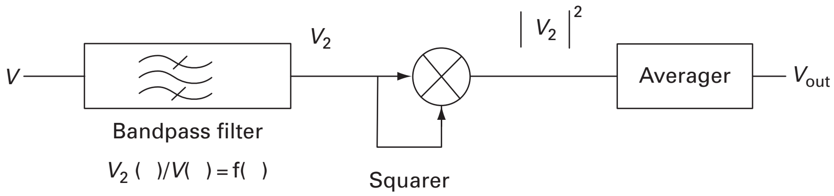

Figure 27.1. Single-channel radiometer.

This relation, the Wiener–Khinchin theorem, is often used to define the power spectrum. To see that S(ω) as defined by Equation (27.2) agrees with our intuitive ideas of the power spectrum, consider the single-channel radiometer shown in Figure 27.1. The (in situ) voltage transfer function of the bandpass filter is defined in the frequency domain by $H(\omega) = V_2(\omega)/V(\omega)$. The power response function or filter shape is therefore given by $|H(\omega)|^2$, so we expect that the time average of $(V_2)^2$ will be given by

$$\left\langle |V_2(t)|^2 \right\rangle = \int_{-\infty}^{\infty} S(\omega)|H(\omega)|^2(2\pi)^{-1}\mathrm{d}\omega \qquad (27.4)$$

where $S(\omega)$ is the spectral power density function. Note that $(2\pi)^{-1}\mathrm{d}\omega = \mathrm{d}f$; the power spectral density has units of watts/Hz.

Let us now show that the output of this radiometer is indeed given by Equation (27.4), and that $S(\omega)$ is given by Equation (27.2). We can express V_2 in the time domain as

$$V_2(t) = \int_{-\infty}^{t} V(t')h(t-t')\mathrm{d}t', \qquad (27.5)$$

where $h(\tau)$, the impulse response function, is the inverse Fourier transform of $H(\omega)$.[1] The output of the squarer is then given by

$$(V_2(t))^2 = \int_{-\infty}^{t}\int_{-\infty}^{t} V(t')V(t'')h(t-t')h(t-t'')\mathrm{d}t'\mathrm{d}t''. \qquad (27.6)$$

Since we assume $V_{\mathrm{in}}(t)$ is a stationary process, the value of t is arbitrary. Let us pick $t = 0$, which gives

$$(V_2(t))^2 = \int_{-\infty}^{0}\int_{-\infty}^{0} V(t')V(t'')h(-t')h(-t'')\mathrm{d}t'\mathrm{d}t''. \qquad (27.7)$$

---

[1] The impulse function, $h(t)$, is the filter output voltage at t, in response to a delta function input at $t = 0$. Note that $h(t)$ is real and that $h(t) = 0$ for $t < 0$.

From causality, $h(t) = 0$ for $|t| < 0$, so the upper limits of integration can be changed from 0 to infinity. Doing this, making a change of variable, $t'' = t' + \tau$, and taking the time average, we find

$$\left\langle (V_2(t))^2 \right\rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \langle V(t')V(t'+\tau)\rangle h(-t')\, h(-t'-\tau) \mathrm{d}t' \mathrm{d}\tau. \tag{27.8}$$

Using the definition of the ACF, $R(\tau) = \langle V(t)V(t+\tau)\rangle$, we find

$$\left\langle (V_2(t))^2 \right\rangle = \int_{-\infty}^{\infty} R(\tau) \left[\int_{-\infty}^{\infty} h(-t')h(-t'-\tau)dt' \right] \mathrm{d}\tau. \tag{27.9}$$

Applying the convolution theorem,[2] the Fourier transform of the term in square brackets is $|H(\omega)|^2$. Therefore, the bracketed term is the inverse Fourier transform of $|H(\omega)|^2$ or

$$\left[\int_{-\infty}^{\infty} h(-t')h(-t'-\tau)\mathrm{d}t' \right] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{j}\omega\tau} |H(\omega)|^2 \mathrm{d}\tau. \tag{27.10}$$

Substituting this expression into Equation (27.9) produces

$$\left\langle (V_2(t))^2 \right\rangle = \int_{-\infty}^{\infty} R(\tau) \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{j}\omega\tau} |H(\omega)|^2 \mathrm{d}\omega \right] \mathrm{d}\tau. \tag{27.11}$$

Finally, interchanging the order of integration gives us

$$\left\langle (V_2(t))^2 \right\rangle = \int_{-\infty}^{\infty} |H(\omega)|^2 \left[\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{j}\omega\tau} R(\tau)\mathrm{d}\tau \right] (2\pi)^{-1} \mathrm{d}\omega. \tag{27.12}$$

After all this effort, we can now compare Equation (27.12) with Equation (27.4) and confirm that the power spectrum defined by Equation (27.2) agrees with what we would expect to measure with a bandpass filter followed by a squarer and an averager. (Don't worry about the sign discrepancy in the $\mathrm{e}^{\mathrm{j}\omega t}$ term in Equation (27.12); since $V_2(t)$ and $R(\tau)$ are real, we can replace this term by $\mathrm{e}^{-\mathrm{j}\omega t}$.)

27.2.1 Hardware autocorrelators

A hardware autocorrelator is a special-purpose parallel signal processor that calculates an averaged ACF, usually in real time. A single Fourier transform operation in a computer then turns the ACF into the power spectrum with the same number of frequency bins (points on the spectrum) as the number of points

---

[2] If $H(\omega)$ is the Fourier transform of $h(\tau)$, then $|H(\omega)|^2$ is the Fourier transform of $\int h(t)h(t+\tau)\mathrm{d}t$.

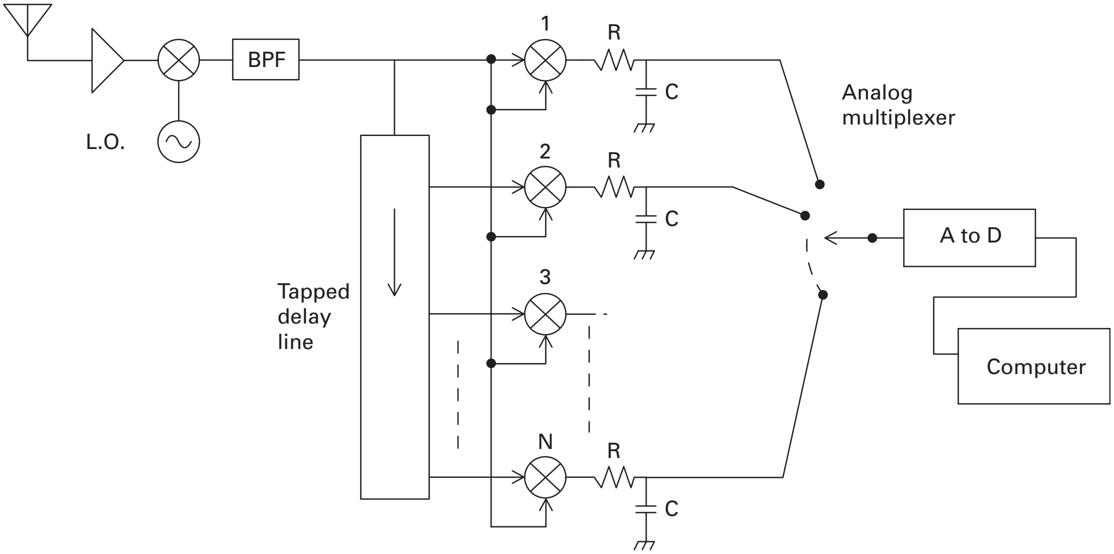
L.O.
BPF
1
2
3
N
R
C
Tapped
delay
line
Analog
multiplexer
A to D
Computer

Figure 27.2. Analog autocorrelator.

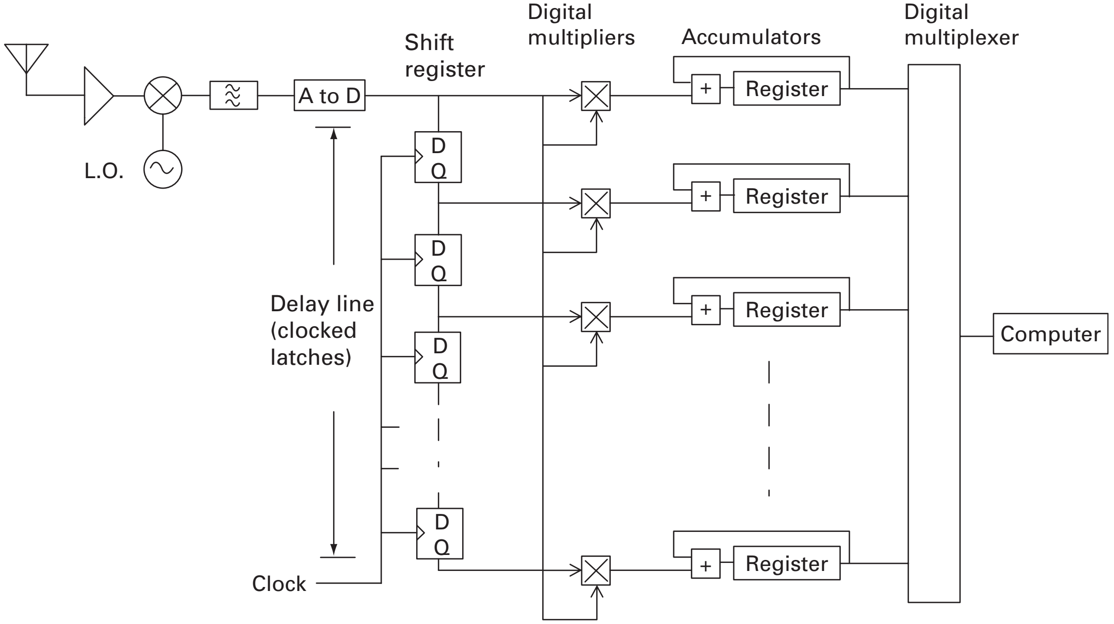
L.O.
A to D
Shift
register
Digital
multipliers
Accumulators
Digital
multiplexer
D
Q
+
Register
Delay line
(clocked
latches)
Clock
Computer

Figure 27.3. Digital autocorrelator.

on the ACF. Autocorrelators use a simple expandable architecture. Figure 27.2 shows an analog version of the autocorrelator.

Figure 27.3 shows a digital autocorrelator. The analog delay line is replaced with a digital shift register. The analog multipliers and averagers are replaced by digital multipliers and accumulators.

The autocorrelators of Figures 27.2 and 27.3 produce values only for discrete points on the ACF. But, as long as the input signal is band-limited (usually by a bandpass "anti-aliasing" filter), the spectrum can be represented as a Fourier *series*, so sampled ACF points are sufficient to compute the continuous power spectrum. In practice, however, we cannot measure an infinite number of points on the ACF. The spectral estimate obtained by transforming ACF points that extend only to $\tau_{\max}$, i.e., by transforming an ACF "windowed" with a rectangular function that is unity for $\tau_{\max} < \tau < \tau_{\max}$ and zero outside this range, produces a function which is the true spectrum convolved by the function $\sin(\omega \tau_{\max})/(\omega \tau_{\max})$. A sharp line in the spectrum therefore appears as a broadened line, surrounded by sidelobes. It is common to apply a smoother windowing function such as $\cos(½\pi\ \tau/\tau_{\max})$ to reduce the sidelobes even though the main load will be broadened – a trade-off between spectral resolution and leakage between channels.

27.2.2 One-bit autocorrelation

One-bit autocorrelation is a technique that greatly reduces the complexity of the digital circuitry. The input analog voltage is fed to a comparator whose output (one bit) indicates the sign, i.e., the polarity, of the voltage. The continuous range of analog voltages is compressed down to just two values: plus one and minus one. (In the digital hardware, a digital "1" indicates a value of +1 and a digital "0" indicates a value of −1.) The correlator's delay line shift register needs to be only one bit wide and the one-bit × one-bit multipliers can be exclusive NOR gates, since the output of an exclusive NOR gate is + 1 when the inputs are the same and −1 when they are different. The integrators that follow the multipliers are simple counters. Although the output of a one-bit correlator is a distorted version of the actual correlation function, the distortion can be entirely corrected. Because the input signal has Gaussian amplitude statistics, the output of the one-bit correlator turns out to be:

$$\rho_{1-\text{bit}}(\tau) = (2/\pi)\sin^{-1}[\rho(\tau)] \tag{27.13}$$

where ρ is the normalized autocorrelation function: $\rho(\tau) = R(\tau)/R(0) = R(\tau)/\sigma^2$. This functional relation was derived during World War II by J. H. van Vleck while studying the spectrum of clipped noise. (Jammers can have much better power efficiency if jamming with clipped noise works as well as jamming with noise having Gaussian amplitude statistics.)

The averaged values from a one-bit autocorrelator can, therefore, be inverted to get the true correlation function.

$$\rho(\tau) = \sin[(\pi/2)\rho_{1-\text{bit}}(\tau)]. \tag{27.14}$$

A straightforward proof of the van Vleck relation (Problem 27.4) proceeds from the Gaussian bivariate probability density,

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}} \exp\left(\frac{-1}{2(1-\rho^2)}\left(\frac{x_1{}^2}{\sigma_1{}^2} - 2\rho\frac{x_1 x_2}{\sigma_1\sigma_2} + \frac{x_2{}^2}{\sigma_2{}^2}\right)\right). \tag{27.15}$$

Here x_1 and x_2 are the voltages $V(\mathrm{t})$ and $V(\mathrm{t}+\tau)$, so $\sigma_1 = \sigma_2$. One gives up some radiometric sensitivity for the simplification of one-bit processing; the integration time has to be increased by a factor of $\pi^2/4$ compared to a correlator using multi-bit quantization.

Several generations of digital correlators have been used in radio astronomy. Custom CMOS correlator chips have thousands of lag channels and clock rates above 100 MHZ. A variety of serial and parallel multiplexing techniques are used to combine these chips to increase the number of lag channels and to increase input data rates to several hundred MHZ.

27.3 Fourier transform spectrometry

Fourier transform methods are more efficient than autocorrelation spectrometry simply because the FFT (Fast Fourier Transform algorithm) is an efficient way to compute the DFT (Discrete Fourier Transform). Processing a sequence of N data points requires on the order of N log(N) operations for an FFT vs. N^2 operations for either autocorrelation or a brute force DFT.

27.3.1 DFT spectrometer

The Fourier transform is, by itself, a spectrometer. The DFT of a sequence of voltages, V_n, is given by

$$u_k = \sum_{n=0}^{N-1} V_n \mathrm{e}^{-\mathrm{j}2\pi nk/N}. \tag{27.16}$$

We can write this DFT in terms of time and frequency as

$$u_{\omega_k} = \sum_{n=0}^{N-1} V_{t_n} \mathrm{e}^{-\mathrm{j}\omega_k t_n}, \tag{27.17}$$

where $t_n = n/f_\mathrm{sample}$ and $\omega_k = 2\pi k f_\mathrm{sample}/N$. Note that the transform is equivalent to a bank of N pairs of frequency converters with sine and cosine L.O.s. The k-th pair multiplies the input signal by $\cos(\omega_k t)$ and $\sin(\omega_k t)$, thus shifting the input band to place ω_k at zero frequency. For every converter, N consecutive multiplier output values are summed so one output value is produced every N/f_sample seconds. This summation is equivalent to lowpass filtering and decimation (reducing the sample rate). Successive values of u_k are, therefore, equivalent to (complex) voltages from the signal at and around ω_k, after being shifted to zero frequency and then lowpass filtered. When the input voltage is a random

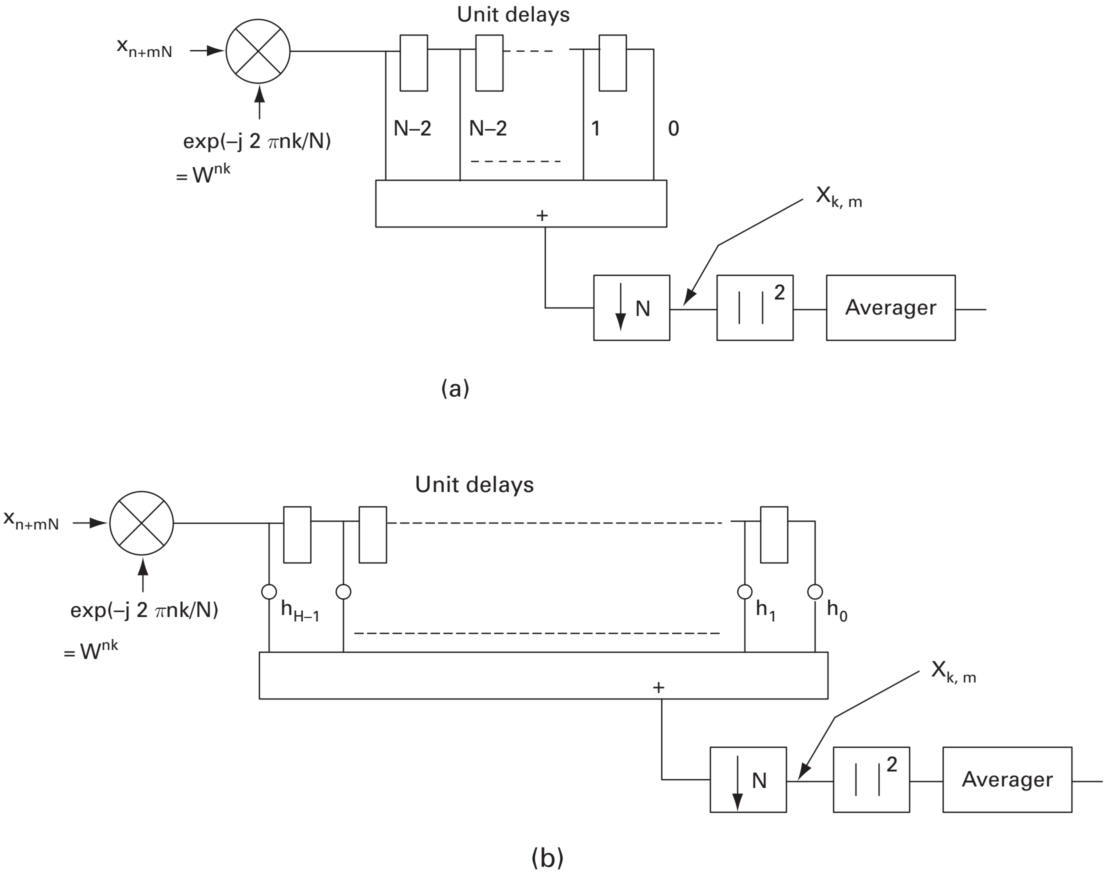

Figure 27.4. (a) One channel of a straight DFT spectrometer (periodogram averager);
(b) filter channel using a longer sum and with arbitrary weights on the inputs to the summer.

process, u_k is also a random process. Its mean is zero. The squared magnitude of u_k represents power, and the set of points, $\{|u_k|^2\}$, is known as a *periodogram*. To estimate the power spectrum of a random process, many periodograms must be averaged together. Figure 27.4(a) shows one channel of a straight periodogram spectrometer.

A shortcoming of the DFT spectrometer (periodogram averager) is that the equivalent filter shape of each channel is not rectangular but, instead, has a $\sin(x)/x$ form, due to the uniform boxcar lowpass filter weights inherent in the DFT. It therefore suffers "leakage" from adjacent and even not-so-adjacent channels. We saw the same problem with the autocorrelation spectrometer. There are obvious remedies for this problem. The application of a weighting function to the input data blocks to give relatively low weight to the data points near the beginning and end of the blocks is equivalent to tapering the weights of the low-pass filter and therefore reduces the amplitude of the filter sidelobes. Such weighting, however, broadens the main lobe and increases the variance of the spectral estimates, since the spectrum is estimated mostly from the data points around the center of the data blocks. This loss of sensitivity can be remedied by computing periodograms from data blocks that overlap each other and including these periodograms in the average. (Even with uniform data weighting, there is a loss of sensitivity when periodograms are computed only from nonoverlapping blocks of data.)

27.3.2 Combined FIR filter/DFT spectrometer

An "FIR/Fourier transform" spectrometer design that fixes the problems of the DFT spectrometer begins with the filter channel shown in Figure 27.4(b). Like the DFT filter, this filter channel will produce one output value for every N input samples. But now, the mixer is followed by a lowpass FIR (finite impulse response) filter which forms a running weighted sum of H successive mixer output values by summing the numbers, weighted by factors, h_i, at the taps of a digital delay line.[3] This would be the same as the DFT filter of the periodogram if all the weights, h_i, were equal to unity and the number of taps, H, were equal to N. Here, however, the number of taps will be an integral multiple of N, $H = RN$. This can be quite a large number, e.g., $H = 4 \times 1024 = 4096$, and the FIR filter can therefore produce a nearly ideal rectangular channel response. As with the straight DFT spectrometer, the bandwidth reduction from lowpass filtering lets us decimate the output of the summer, keeping only one sum per N-point block of input data. Again, the magnitudes of the output of the summer must be squared and averaged to form an estimate of the power spectrum. Note that, because the filter window, H, is longer than the block length, N, the blocks processed by the filter are overlapped; while one output value is produced for every N input samples, the output values are produced from H input samples, where $H > N$. We will designate the input voltage samples as x_n. They enter in blocks of N samples. The output of the summer is $X_{k,m}$, where m is the block number and k, which runs from zero to N, designates the point on the spectrum. Defining W as $W = \exp(-2\pi\,\mathrm{j}\,/N)$, inspection of Figure 27.4 allows us to write

$$X_{k,m} = \sum_{n=0}^{H-1} h_n W^{-k(mN+n)} x_{mN+n}. \tag{27.18}$$

Now, since $W^{-kmM} = 1$, we can rewrite this equation as

$$X_{k,m} = \sum_{n=0}^{H-1} h_n W^{-kn} x_{mN+n}. \tag{27.19}$$

Let us next express n as $rN + \rho$ where the index r goes from 0 to $R-1$, where $R = H/M$, and the index ρ goes from 0 to $M-1$. Using this notation, we have

---

[3] The finite (duration) impulse response is obvious; when an input sample makes its way down the line and falls off at the end, it no longer contributes the output. A FIR filter can be given a desired frequency response function by selecting the tap weights to form the desired impulse function, i.e., the Fourier transform of the desired frequency response.

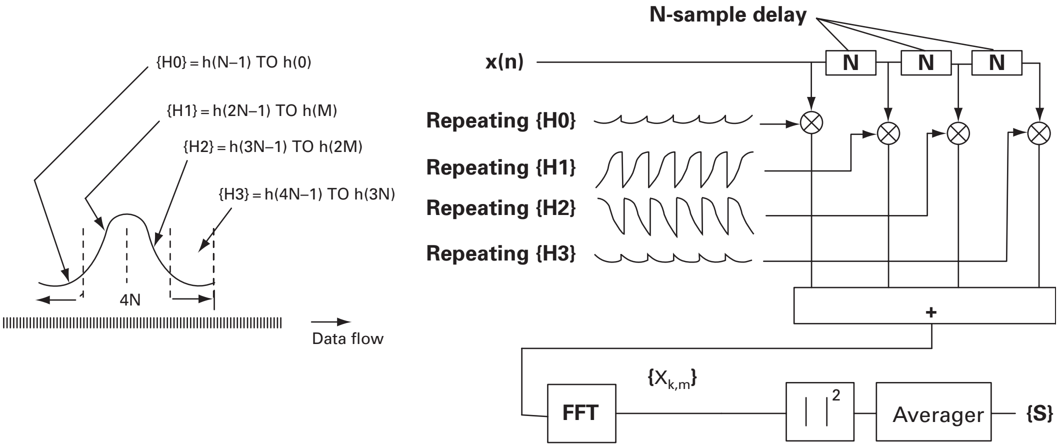

Figure 27.5. Weighted overlap-add filterbank.

$$X_{k,m} = \sum_{\rho=0}^{N-1} \left(\sum_{r=0}^{H/N-1} h_{rN+\rho} x_{mN+n} \right) W^{-k\rho}. \tag{27.20}$$

Note that Equation (27.20) is just an N-point DFT of the term in brackets, and that this term is independent of k, the frequency index. Consider this term for the case where $H = 4N$. The bracketed term is given by

$$(\ ) = h_\rho X_{Nm+\rho} + h_{N+\rho} X_{N(m+1)+\rho} + h_{2N+\rho} X_{N(m+2)+\rho} + h_{3N+\rho} X_{N(m+3)+\rho}. \tag{27.21}$$

In the first term on the right-hand side, successive samples are multiplied by h_0 through h_{N-1}. In the second term, delayed samples are multiplied by h_N through h_{2n-1}, and so on. An architecture that forms this bracketed term is shown in Figure 27.5. Four multipliers are used, together with three delay lines, each of length N. The set of H FIR coefficients is split into four sections, each of which feeds one of the four multipliers in a continuous cycle. An adder sums the products flowing from the multipliers. This preprocessor section feeds a standard FFT block. This architecture is called a *weighted overlap-add* (WOLA) filterbank. Another way of producing the bracketed term yields the *polyphase filter bank* [3].

27.4 *I* and *Q* mixing

In the discussions above, it was implicit that the spectrum of the band-limited signal voltage, $V(t)$, extends from zero to some cutoff frequency, B. This baseband signal can be produced from an IF signal that has been bandpass filtered and then down-converted to put the lower edge of the passband at zero frequency. For digital spectrum analysis, the sampling frequency must be at least $2B$ to avoid aliasing (the stroboscope effect). In the case of the DFT spectrometer and the improved WOLA version, the N channels extend in frequency from

0 to the sampling rate, i.e., from 0 to $2B$. Therefore, only the first $N/2$ output channels can be used, since those channels correspond to the frequency range 0 to B. (The second $N/2$ channels will contain the same information; channel $N-k$ will have the same data as channel k.) This effective waste of half the DFT channels can be avoided with a technique known as *complex sampling*. In this technique, the IF signal of bandwidth B is down-converted into baseband "*I*" and "*Q*" (in-phase and quadrature) channels by using two mixers having L.O. signals $\cos(\omega_c t)$ and $\sin(\omega_c t)$, where ω_c is the center of the IF band. We can express the original IF signal in terms of the mixer outputs, $I(t)$ and $Q(t)$, as

$$V_{\mathrm{IF}}(t) = I(t)\cos(\omega_c t) + Q(t)\sin(\omega_c t). \tag{27.22}$$

To find the spectral value at the point $\omega_c + \Delta$, we must multiply (mix) the IF voltage by $e^{-j(\omega c + \Delta) t}$. Carrying out this multiplication and ignoring terms of frequency $2\omega_c$, we find

$$V_{\mathrm{IF}}(t)e^{-j(\omega_c+\Delta)t} = [I(t) - jQ(t)]e^{-j\Delta t}. \tag{27.23}$$

The term on the right-hand side is equivalent to mixing a complex-valued signal, $I(t)-jQ(t)$, so that the spectral point at $\omega_c + \rho$ is shifted to zero frequency. The beauty of this is that both $I(t)$ and $Q(t)$ have bandwidth $B/2$, since $B/2$ is the greatest separation between ω_c and any point in the IF band. Therefore we can feed $I(t)-jQ(t)$ as a sequence of complex numbers to the straight DFT spectrometer or the improved WOLA version and the N output values will span the entire bandwidth B. The spectrum from $\Delta = 0$ to $B/2$ will appear in the first $N/2$ channels, while the spectrum from $\Delta = -B/2$ to 0 will appear in the second $N/2$ channels. When I and Q mixing is used in autocorrelation spectrometry, both autocorrelation functions and cross-correlation functions are computed from the I and Q signals.

27.5 Acousto-optical spectrometry

The acousto-optical spectrometer (AOS) uses an optical crystal which becomes a diffraction grating by virtue of mechanical waves (sound waves at radio frequencies) propagating through it. These waves produce corrugations in the refractive index along the length of the crystal. The waves are launched into this *Bragg cell* by an electrical-to-acoustic transducer (usually a piezoelectric crystal), driven by the IF signal to be analyzed. Figure 27.6 shows how the crystal is illuminated by a laser and how a multi-element CCD array detects (and averages) the diffraction pattern.

The linear CCD array, like those used in supermarket checkout scanners, accumulates charge along its length at a rate given by the incident light intensity. For read-out, the charge packets are clocked down the length of the CCD in a bucket brigade fashion. The voltage at the end of the array, which is proportional to the charge at that position, is digitized and made available to the data-taking computer.

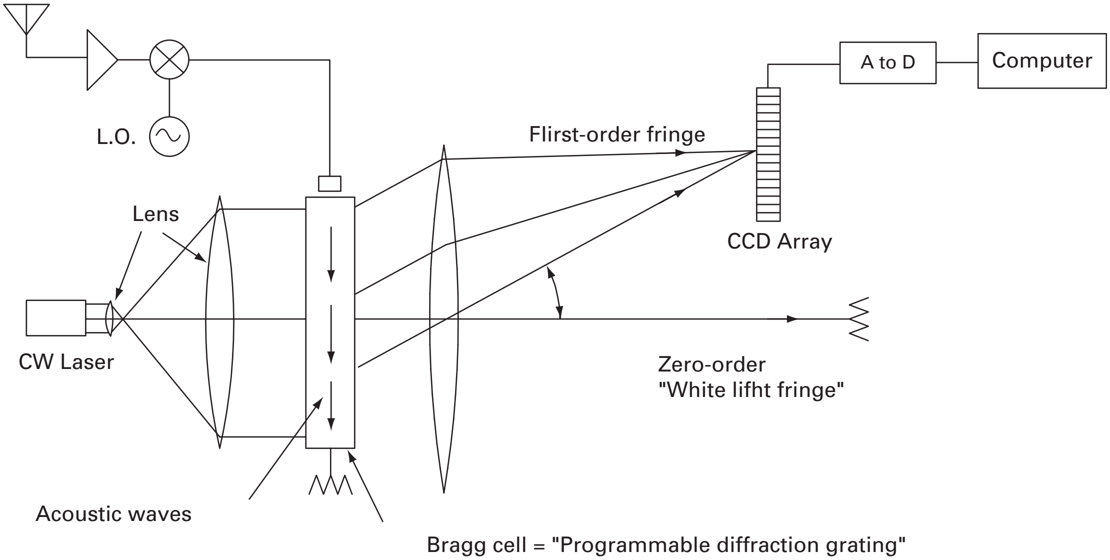

Figure 27.6. Acousto-optical spectrometer (AOS).

A figure of merit for any spectrometer is its number of frequency bins, i.e., its analysis range divided by its resolution. For the AOS, the effective number of channels is given by the time–bandwidth product of the Bragg cell. To see this we note that the angular size (in radians) of the light incident on the CCD is given by $\lambda_{\text{laser}}/L_{\text{cell}}$, the ratio of the laser wavelength to the length of the cell, just as the beam size for an aperture antenna with diameter D is given by λ/D. The first-order diffraction condition (so that all the rays arrive in phase at the CCD) is given by $\lambda_{\text{grating}}\sin\theta = 1 \times \lambda_{\text{laser}}$ where λ_{grating} is the corrugation wavelength in the cell. Now λ_{grating} can be written as v_{cell}/f where v_{cell} is the sound velocity and f is the input RF frequency. Approximating $\sin\theta$ by θ, we find the total angular range over the total frequency range given by $\Delta\theta/\Delta f = \lambda_{\text{laser}}/\text{v}_{\text{cell}}$. The number of channels is then given by the range divided by resolution,

$$N = \frac{\lambda_{\text{laser}}\Delta f L_{\text{cell}}}{v_{\text{cell}}\lambda_{\text{laser}}} = \frac{L_{\text{cell}}}{v_{\text{cell}}}\Delta f = T\Delta f, \tag{27.24}$$

where T is the propagation time through the length of the cell. The quantity $T\,\Delta f$ is known as the *time–bandwidth product* of the cell. Materials used for Bragg cells include quartz, lithium niobate, and glass. A spectrometer for radio astronomy might use cells with a 200-MHz bandwidth and a time–bandwidth product of one thousand.

27.6 Chirp-z spectrometry

The chirp-z spectrometer is shown in Figure 27.7. Also known as the microscan or compressive receiver, it uses a dispersive filter and a swept L.O. While it appears to be a swept spectrum analyzer, it is actually a multiplex analyzer. The

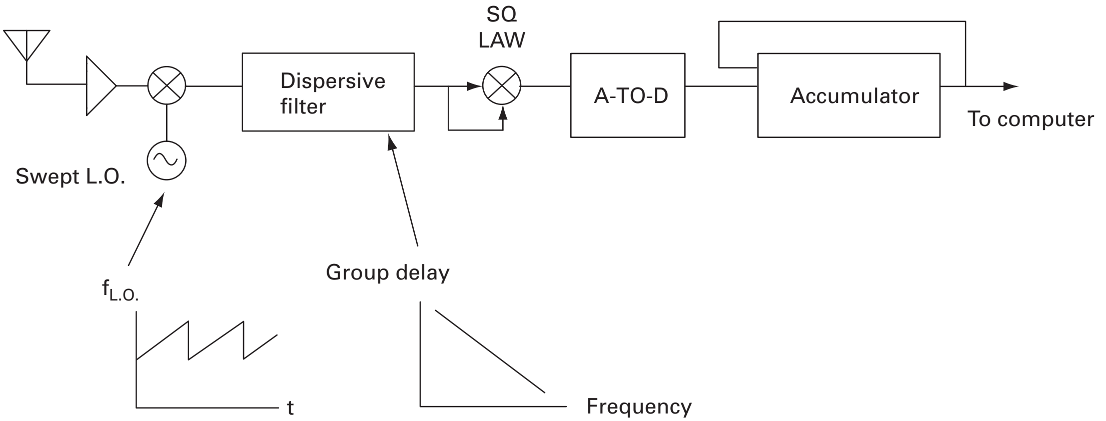

Figure 27.7. Chirp-z spectrometer.

filter has a group delay characteristic that is linear with respect to frequency. In Figure 27.7, the delay is greatest for frequencies at the low end of the filter band. The local oscillator frequency is given a linear sweep or *chirp*, repeated in a sawtooth fashion with period *T*. The ramp rate of the L.O. (MHz/sec) is made to be the reciprocal of the group delay slope (seconds per megahertz) of the filter.

Consider a continuous wave (cw) input signal. Since the L.O. is swept, the mixer output will also sweep in frequency. In the case of a high-side mixer (L.O. frequency above the input signal frequency), the mixer output sweeps up in frequency at the same rate as the L.O. sweep. As the L.O. sweeps upward, the first signals entering the filter are at the filter's low-frequency end. These signals, as they travel through the filter, are delayed the most. Frequencies entering the filter farther up the sweep are delayed less. But all frequencies arrive simultaneously at the filter output, producing a single sharp output pulse. If the input signal consists of two cw tones, there will be two output pulses per sweep. The position of the pulses in time will indicate the frequencies of the cw input signals. Of course this linear system can handle any combination, i.e., spectrum, of input signals. The output data rate from the squarer is equal to the input data rate so additional circuitry (the digital accumulator in Figure 27.7) is required for signal averaging.

The principal application of chirp-z spectrometers has been in electronic warfare – as surveillance receivers to sense radar or communication signals over a wide spectrum and with a high probability of intercept. To find the effective number of channels in this spectrometer we can first consider the output pulse produced by a single cw input signal. This output pulse is produced by a coherent superposition of frequencies over the entire bandwidth, *B*, of the filter. The width of this pulse is therefore given by $1/B$. Dividing the total output time *T* by the pulse width, the effective number of channels is given by *TB*, the time–bandwidth product, just as with the Bragg cell spectrometer.

Dispersive filters used for chirp-z spectrometry are usually surface acoustic wave devices. Piezoelectric transducers convert the input signal into mechanical surface waves and then back again to an electrical output signal. In one design, diffraction gratings on the surface of the crystal are arranged to make the input-to-output path length longer for low-frequency waves than for higher frequency waves. Surface-wave dispersive filters can have bandwidths of hundreds of MHZ and time–bandwidth products of several hundred. Dispersive filters are also made from charge-coupled devices (CCDs). Generically these filters are allpass networks and the first dispersive filters were cascades of many second-order allpass sections. (The number of sections required is equal to the time–bandwidth product.) Note: the filters are not just allpass networks; they are usually given some amplitude "taper," i.e., the amplitude response is made to roll off at the band edges, in order to eliminate the "sidelobes" of the instrumental function that result from uniform amplitude response. The same is done with the AOS; the transmission of the Bragg cell is reduced near the ends.

Problems

Problem 27.1. For the Gaussian probability density, $f(V)=(2\pi\sigma^2)^{-1/2}\exp(-V^2/2\sigma^2)$, verify that

$$\int_{-\infty}^{\infty} f(V)\mathrm{d}V = 1 \quad \text{and} \quad \int_{-\infty}^{\infty} f(V)V^2\mathrm{d}V = \sigma^2.$$

Problem 27.2. (a) The noise-like signals observed in radio astronomy can be modeled as a comb of delta functions in frequency space – a picket fence of sine waves. Use your computer to generate 50 sine waves spanning the angular frequency range 1.00–1.10. Let these frequencies be somewhat random, e.g., $\omega_n = 2\pi(1+0.002n+0.05\ \mathrm{rnd}(1))$ where rnd(1) is a random variable in the range zero to one. Give each sine wave a random starting phase, $\phi_n = 10\ \mathrm{rnd}(1)$. Plot the sum of the sine waves as a function of time for $t=0$ to $t=20$ using a time interval of 0.1. This narrowband spectrum should look quite sinusoidal but with a slowly varying phase and amplitude.

(b) Repeat this exercise but with the angular frequency range extending from 1 to 2, i.e., $\omega_n = 2\pi(1+0.05n+0.05\ \mathrm{rnd}(1))$. This wideband spectrum should look like random noise.

(c) In either case, (a) or (b), the resulting voltage (sum) behaves like a random variable with Gaussian statistics. Plot a histogram of voltage samples to verify this.

Problem 27.3. Suppose a random signal is produced by applying an ideal lowpass filter to white noise. The resulting power spectrum is flat from dc to the filter cutoff frequency ω_c. Show that the normalized autocorrelation function is given by $\rho(\tau)=\sin(\omega_c\tau)/(\omega_c\tau)$.

Problem 27.4. Derive the VanVleck relation, Equation (27.13), that gives the one-bit autocorrelation function in terms of the (normalized) autocorrelation function. Use the bivariate Gaussian probability density function, Equation (27.15), where x_1 and x_2 represent the voltages $V(t)$ and $V(t+\tau)$, and $\sigma_1=\sigma_2$. Since the product $\mathrm{sign}(x_1)\,\mathrm{sign}(x_2)$ is equal to +1 in the first and third quadrants and −1 in the second and fourth quadrants, the expectation of sign(x_1) sign(x_2) is the integral of $f(x_1,x_2)$ over x_1 and x_2, weighted by +1 or −1 according to the quadrant. Hint: change to polar coordinates and keep a table of integrals handy.

Problem 27.5. If the dispersive filter used in the chirp-z spectrometer or pulse compression radar is given an impulse (delta function in time), what sort of output waveform does it produce?

References

[1] Blackman R. B. and Tukey, J. W., *The Measurement of Power Spectra*, New York: Dover, 1959.

[2] Childers, D. B. ed., *Modern Spectrum Analysis*, New York: IEEE Press, 1978.

[3] Crochiere, R. E. and Rabiner, L. R., *Multirate Digital Signal Processing*, Englewood Cliffs: Prentice-Hall, 1983.

[4] Jack, M. A., Grant P. M. and Collins, J. H., The theory, design, and applications of surface acoustic wave Fourier-transform processors. *Proc. IEEE*, vol. 68, pp. 450–468, 1980.

[5] Kesler, S. B. ed. *Modern Spectrum Analysis, II*, New York: IEEE Press, 1986.

[6] Thomas, J. B., *An Introduction to Statistical Communication Theory*, New York: John Wiley, 1969.

[7] Thompson, A. R., Moran, S. M. and Swenson, Jr. G. W., *Interferometry and Synthesis in Radio Astronomy*, Malabar, Florida: Krieger Publishing Company, 1991.

CHAPTER

28

S-parameter circuit analysis

The *S*-parameter (scattering parameter) analysis technique for linear circuits has become the de facto standard in RF engineering, especially for microwave work, where it had its origins in WWII radar development. Physicists-turned-engineers, who were accustomed to analyzing the scattering (collisions) of atomic particles, used that term to describe the reflection and transmission of electromagnetic waves in electrical circuits. The resulting version of circuit theory, in terms of waves rather than currents and voltages, is convenient when working with circuits whose interconnections are transmission lines, especially when those lines are waveguides, where the definitions of voltage and current are somewhat arbitrary. The Smith chart, introduced in 1939, provided a familiar connection to the *S*-parameter method. In the 1960s, Hewlett Packard introduced the first network analyzers for direct *S*-parameter measurements.

28.1 *S*-parameter definitions

Keep in mind that the *S* parameters are just another set of transfer parameters, like the *Y* and *Z* parameters, to describe a linear circuit or circuit element at a given frequency. A two-port network is described equally well by its four complex *Y* parameters, its four complex *Z* parameters, or its four complex *S* parameters. Knowing one set of parameters, we can calculate any other set.

28.1.1 *S* parameter of a one-port network

A resistor, the simplest circuit element, can be described by a single parameter, its resistance, *R*, the ratio of voltage to current. It can be equally well described by its conductance, *G*, the ratio of current to voltage. The conversion between these parameter choices is particularly simple: $G = 1/R$. When a resistor is mounted in a coaxial connector as a dummy load, we regard it as a one-port RF device and often describe it by its reflection coefficient, $\Gamma = (R - Z_0)/(R + Z_0)$, where Z_0, the reference impedance, is often 50 ohms. We saw in Chapter 10 that Γ is the ratio of the reflected wave amplitude to the incident wave amplitude when a wave arrives at the resistor by way of a transmission line of characteristic impedance Z_0. This ratio can be the ratio of either the voltages or currents of the reflected and incident waves, since these ratios are equal. The reflection coefficient is denoted as S_{11}, and is the one and only S parameter for a one-port device. The subscripts refer to a wave leaving port 1 as a result of a wave entering port 1. Using the familiar formula for reflection coefficient, the S parameter of an arbitrary one-port is given by $S_{11} = (Z - Z_0)/(Z + Z_0)$, where Z is the impedance upon which the wave is incident and Z_0 is a reference impedance.

28.1.2 *S* parameters for a general *n*-port network

Let us now define the S parameters of a general n-port device, examples of which include attenuators, filters and amplifiers (two-port devices), two-way power splitters (three-port devices), and directional couplers (four-port devices). For any *properly terminated* n-port device, S_{ij} is the ratio of the (complex) amplitude of the wave exiting port i to the (complex) amplitude of a wave incident on port j. Note that, when $i = j$, the S parameter is a *reflection* coefficient and, when $i \neq j$, the S parameter is a *transmission* coefficient. An n-port device will have n^2 S parameters. Figure 28.1, shows a two-port network reviewing how the input and output currents are written in terms of the amplitudes of forward and reverse waves at each port.

The S parameters for a two-port network are defined as follows:

$$\begin{aligned} S_{11} &= \left.\frac{V_{1\text{rev}}}{V_{1\text{fwd}}}\right|_{V_{2\text{fwd}}=0} & S_{12} &= \left.\frac{V_{1\text{rev}}}{V_{2\text{fwd}}}\right|_{V_{1\text{fwd}}=0} \\ S_{21} &= \left.\frac{V_{2\text{rev}}}{V_{1\text{fwd}}}\right|_{V_{2\text{fwd}}=0} & S_{22} &= \left.\frac{V_{2\text{rev}}}{V_{2\text{fwd}}}\right|_{V_{1\text{fwd}}=0}. \end{aligned} \tag{28.1}$$

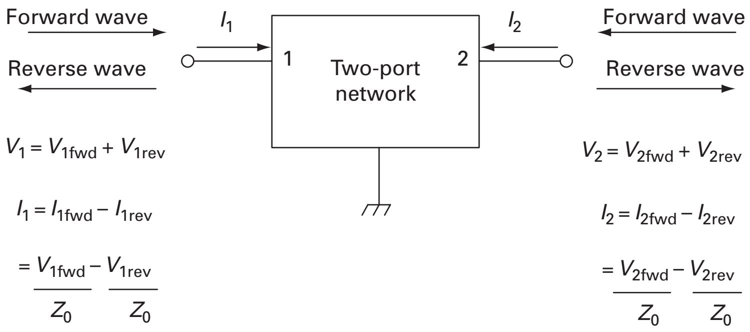
Forward wave
Reverse wave
I_1
1
Two-port network
2
I_2
Forward wave
Reverse wave
$V_1 = V_{1\text{fwd}} + V_{1\text{rev}}$
$I_1 = I_{1\text{fwd}} - I_{1\text{rev}} = \frac{V_{1\text{fwd}}}{Z_0} - \frac{V_{1\text{rev}}}{Z_0}$
$V_2 = V_{2\text{fwd}} + V_{2\text{rev}}$
$I_2 = I_{2\text{fwd}} - I_{2\text{rev}} = \frac{V_{2\text{fwd}}}{Z_0} - \frac{V_{2\text{rev}}}{Z_0}$

Figure 28.1. Forward and reverse waves at the ports of a network.

The condition that the device be properly terminated means that when the *S*-parameters are measured or calculated, each port must be connected to a load impedance equal to the value assigned to that port. Port impedance assignments are usually obvious. If an amplifier is designed to be driven by a 50-ohm source and to feed a 50-ohm load, the input and output ports will both be designated as 50-ohm ports. The one-port dummy load discussed above, if it contains a 50-ohm resistor, would surely be specified by the supplier as having $\Gamma = 0$ (or some very small value). Note, however, that the same 50-ohm device would have a reflection coefficient of (50−100)/(50+100)=−1/3, at a port with an assigned impedance of 100 ohms. When port impedances have been assigned and the corresponding *S* parameters are measured or calculated, the device is totally characterized; if necessary, one can use this set of *S* parameters to calculate a new set of *S* parameters, corresponding to different port impedances.

28.1.3 *S* parameters for an example network

Let us calculate the S parameters for the network shown in Figure 28.2, which contains two passive elements Z_1 and Z_2, which can be resistances or reactances or both, i.e., $Z_1 = R_1 + jX_1$ and $Z_2 = R_2 + jX_2$. Interconnection wires are considered to be infinitesimally short. We will assume 50-ohm port assignments.

To calculate S_{11}, we apply the proper termination, 50 ohms, to port 2 as shown in the figure. S_{11} is then the reflection coefficient at port 1, i.e., $S_{11} = (Z_{in} - Z_0)/(Z_{in} + Z_0)$, where Z_{in} is the impedance looking into port 1. By inspection we see that $Z_{in} = Z_1 + Z_2 \,||\, 50 = Z_1 + (50\, Z_2) /(Z_2 + 50)$. Substituting this expression into the expression for Γ_{in}, we find

$$S_{11} = \frac{Z_1 - 50 + 50Z_2/(Z_2 + 50)}{Z_1 + 50 + 50Z_2/(Z_2 + 50)}. \tag{28.2}$$

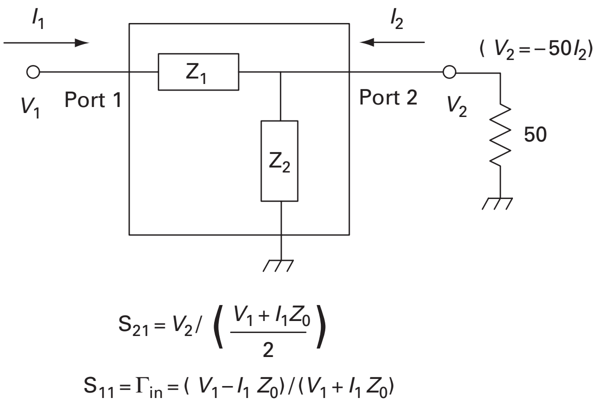

I_1
I_2
$(V_2 = -50I_2)$
Z_1
V_1
Port 1
Port 2
V_2
50
Z_2
$S_{21} = V_2 / \left(\frac{V_1 + I_1 Z_0}{2}\right)$
$S_{11} = \Gamma_{in} = (V_1 - I_1 Z_0)/(V_1 + I_1 Z_0)$

Figure 28.2. Example network whose S parameters are calculated.

We calculate S_{22} with the same method, but now terminating port 1 with 50 ohms. Then S_{22} is the reflection coefficient looking into port 2, and we find

$$S_{22} = \frac{Z_2(Z_1 + 50) - 50(Z_1 + Z_2 + 50)}{Z_2(Z_1 + 50) + 50(Z_1 + Z_2 + 50)}. \tag{28.3}$$

To calculate S_{21} (out 2, in 1) we put a 50-ohm load on port 2 and apply an arbitrary voltage V_1 to Port 1. We can then calculate the amplitude of the wave leaving port 2. Because port 2 is properly terminated, there will be no wave reflected back into it. Therefore, the amplitude of the voltage appearing at port 2, V_2, is the amplitude of the wave leaving that port. However, since S_{11} is not zero, V_1 is the sum of an incident wave and a reflected wave. Recalling the transmission line theory of Chapter 10, $V_1 = V_{\text{fwd}} + V_{\text{rev}}$ and $I_1 = I_{\text{fwd}} - I_{\text{rev}} = V_{\text{fwd}}/Z_0 - V_{\text{rev}}/Z_0$. Combining these two equations gives us the amplitude of the forward wave: $V_{\text{fwd}} = ½(V_1 + Z_0 I_1)$. Summing up, a procedure to calculate S_{21} is to apply a voltage V_1 to Port 1, solve for V_2 and I_1, and then calculate $S_{21} = V_2/[½(V_1 + Z_0 I_1)]$. Carrying out this procedure, we find, after a few steps of algebra,

$$S_{21} = \frac{2(50Z_2)}{50Z_2 + (Z_1 + 50)(Z_2 + 50)}. \tag{28.4}$$

If we do the same thing from the other side, to calculate S_{12}, we find that $S_{12} = S_{21}$, which, as we will see later, is not a coincidence.

If, instead of directly calculating S_{11} from the reflection coefficient formula, we had applied a voltage V_1 to port 1, as above, we would have used $S_{11} = (V_1 - I_1 Z_0)/(V_1 + I_1 Z_0)$. Note, then, that we can write general expressions for both S_{21} and S_{11} in terms of conditional ratios of linear combinations of voltage and current, without reference to transmission lines or reflections (or the internals of the two-port), just as we define the *Y* or *Z* parameters:

$$S_{11} = \left.\frac{V_1 - Z_0 I_1}{V_1 + Z_0 I_1}\right|_{V_2 = -Z_0 I_2} \tag{28.5}$$

$$S_{21} = \left.\frac{V_2}{0.5(V_1 + Z_0 I_1)}\right|_{V_2 = -Z_0 I_2}. \tag{28.6}$$

To get the expressions for S_{22} and S_{12}, simply change ones to twos and twos to ones in Equations (28.5) and (28.6).

Before leaving this example, let us look at an alternative way to calculate S_{12} or S_{21}. Rather than connecting an arbitrary voltage generator directly to port 1, we use a 2-V generator and make the connection through a 50-ohm resistor. If the resulting source (the 2-V generator and the 50-ohm series resistor) were connected to a properly terminated 50-ohm cable (or just to another 50 ohm resistor), it would produce a forward wave with an amplitude of 1 V. And if it is connected to a mismatched load, it will still produce a 1-V forward wave. A mismatched load will produce a reflection, but since the source has a 50-ohm

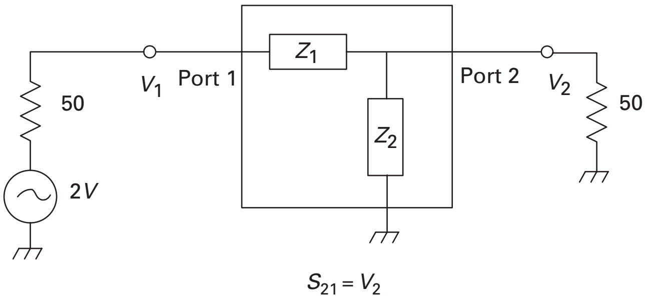

Figure 28.3. Alternate circuit for calculation of S_{21}.

impedance, the reflected wave is totally absorbed and the forward wave maintains its 1-V amplitude. Thus, to calculate S_{21} we simply excite the network with this source and calculate V_2, since $S_{21} \equiv V_{2\text{fwd}}/V_{1\text{fwd}} = V_2/V_{1\text{fwd}} = V_2$. This is shown in Figure 28.3.

28.1.4 Normalization

When the ports of a device are not all assigned the same impedance, it is common to normalize S-parameter magnitudes, i.e., the magnitude of S_{ij} is taken as the square root of the ratio of power emerging from port *i* to power entering port *j* when all other ports are terminated in their assigned impedances.

28.2 Circuit analysis using *S* parameters

Writing the relations between the independent variables (ingoing or "forward" waves) and dependent variables (outgoing or "reverse" waves) for a two-port network in matrix-like form, we have

$$V_{1\text{rev}} = S_{11}V_{1\text{fwd}} + S_{12}V_{2\text{fwd}} \tag{28.7}$$

$$V_{2\text{rev}} = S_{21}V_{1\text{fwd}} + S_{22}V_{2\text{fwd}}. \tag{28.8}$$

Let us use these two equations to solve a basic network problem: finding the reflection coefficient at port 1 of a two-port network when port 2 is connected to an arbitrary load, as shown in Figure 28.4.

As we have noted, the load can be specified by its reflection coefficient, Γ_L. That immediately requires that $V_{2\text{fwd}} = \Gamma_\text{L} V_{2\text{rev}}$, i.e., a fraction Γ_L of the wave leaving port 2 is reflected back into port 2. We will set $V_{1\text{fwd}}$ to unity so that the reflection coefficient at port 1 will be $V_{1\text{rev}}$. Equations (28.7) and (28.8) then become

$$V_{1\text{rev}} = S_{11} \cdot 1 + S_{12}\Gamma_L V_{2\text{rev}} \tag{28.9}$$

$$V_{2\text{rev}} = S_{21} \cdot 1 + S_{22}\Gamma_L V_{2\text{rev}}. \tag{28.10}$$

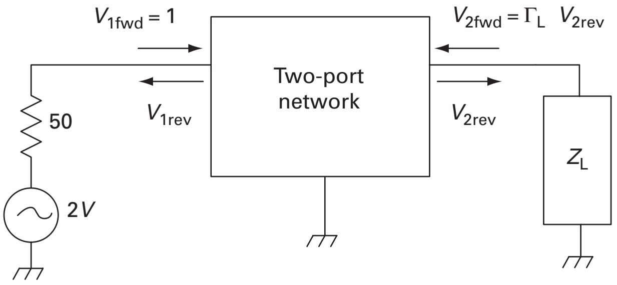

Figure 28.4. A two-port network with an arbitrary load.

Solving Equation (28.10) for V_{2rev} and substituting into Equation (28.9), we find

$$\Gamma_{in} = V_{1rev} = S_{11} + \frac{S_{12}\Gamma_L S_{21}}{1 - S_{22}\Gamma_L}. \tag{28.11}$$

If we simply turn things around, we see that the reflection coefficient seen at port 2 is obtained by interchanging 1's and 2's and *S*'s and *L*'s in this expression, i.e.,

$$\Gamma_{out} = V_{2rev} = S_{22} + \frac{S_{21}\Gamma_S S_{12}}{1 - S_{11}\Gamma_S}. \tag{28.12}$$

28.2.1 Signal flow graphs

A convenient aid to S-parameter circuit analysis is the signal flow graph – a directed graph that represents a set of linear equations. A signal flow graph for the circuit discussed above is shown in Figure 28.5(a).

The nodes (dots) on the graph correspond to the magnitudes of the forward and reverse waves. The branches (lines) are labeled with the transfer coefficients (*S* parameters) that determine the node values. Equations (28.7) and (28.8) are contained in this graph. Consider for example the node at the upper right, V_{2rev}. Only two branches are directed toward this node, one from V_{1fwd} and one from V_{2fwd}. Summing the contributions that arrive through these branches we write $\mathrm{V}_{2rev} = S_{21}\ V_{1fwd} + S_{22}\ V_{2fwd}$, which is just Equation (28.8). On this graph, the load is represented by a branch which indicates that $V_{2fwd} = \Gamma_L\ V_{2rev}$. The source, which could be the 2-V generator and series 50-ohm resistor, produces a forward wave with a magnitude of one, which is represented on the graph by assigning the value unity to the left-most node, V_{1fwd}. Let us now see how we can use the graph to arrive at our previous result for the reflection coefficient, V_{1rev}. Inspecting the graph, we see that if we first find the value of V_{2rev}, we will immediately be able to write

$$V_{1rev} = V_{2rev}\Gamma_L S_{12} + 1 \cdot S_{11} \tag{28.13}$$

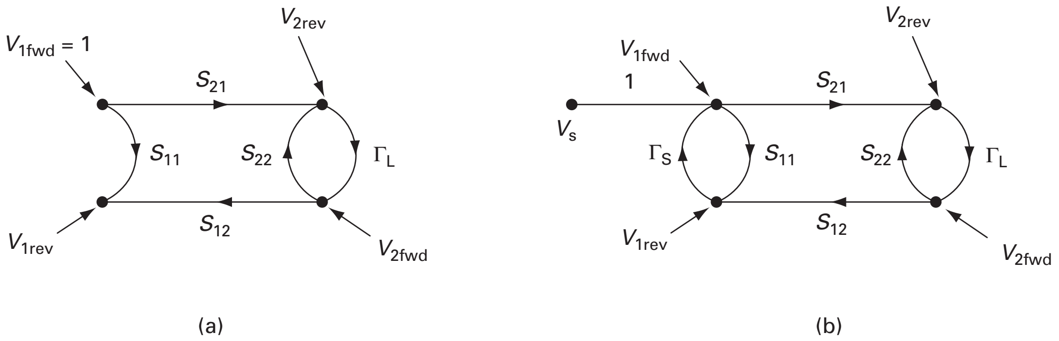

Figure 28.5. (a) Signal flow graph for the circuit of Figure 28.4; (b) same, but for a source whose impedance is not Z_0.

by following the two paths into the node V_{1rev}. Focusing on the node V_{2rev}, see that two branches are directed into it, so we write $V_{2rev} = 1 \cdot S_{21} + V_{2rev}\,\Gamma_L\,S_{22}$, from which $V_{2rev} = 1 \cdot S_{21}/(1-\Gamma_L)$. Substituting this into Equation (28.13) produces $V_{1rev} = S_{11} + \Gamma_L\,S_{12}\,S_{21}/(1-\Gamma_L)$, as before.

28.2.2 Two-port transmission and reflection with arbitrary source and load impedances

Let us now look at the same circuit, but with an imperfect source, i.e., a source having a reflection coefficient $\Gamma_S \neq 0$. The signal flow graph, Figure 28.5(b), acquires two extra branches. One is Γ_S, which steers some of the reflection from the two-port around, contributing to V_{1fwd}. The other branch comes from a node, V_S, which is the forward wave the source would produce at a matched input port. Our goal is to solve for V_{2rev}, the wave transmitted to the load. Examining the graph, we write simultaneous equations for V_{2rev} and V_{1fwd}:

$$V_{2rev} = V_{1fwd}S_{21} + V_{2rev}\Gamma_L S_{22} \tag{28.14}$$

$$V_{1fwd} = V_S + V_{1fwd}S_{11}\Gamma_S + V_{2rev}\Gamma_L S_{12}\Gamma_S. \tag{28.15}$$

Eliminating V_{1fwd}, we have our desired result:

$$V_{2rev} = \frac{V_S S_{21}}{(1-\Gamma_L S_{22})(1-\Gamma_S S_{11}) - S_{12}S_{21}\Gamma_L\Gamma_S}. \tag{28.16}$$

Note: if the identification of the equations for V_{2rev} and V_{1fwd} is too great a reach, the reader can readily identify equations for all four variables:

$$\begin{aligned} V_{2rev} &= V_{1fwd}S_{21} + V_{2fwd}S_{22} \\ V_{2fwd} &= V_{2rev}\Gamma_L \\ V_{1rev} &= V_{2fwd}S_{12} + V_{1fwd}S_{11} \\ V_{1fwd} &= V_S + V_{1rev}\Gamma_S \end{aligned}$$

and then eliminate V_{2fwd} and V_{1rev} to obtain Equations (28.14) and (28.15).

When the two-port is an amplifier, we often want to find the gain. There are several definitions of gain. The *operating power gain* is just the ratio of the net power delivered to the load to the net power delivered by the source. The *transducer gain* is the ratio of the net power delivered to the load to the power available from the source, and the *available power gain* is the ratio of the power available from the amplifier divided by the power available from the source. Let us calculate the operating power gain. The net power delivered to the load is the power of the wave into the load, $|V_{2\text{rev}}|^2$, minus the power of the wave leaving the load, $|\Gamma_L V_{2\text{rev}}|^2$, i.e., $P_{\text{load}} = |V_{2\text{rev}}|^2 (1 - |\Gamma_L|^2)$. The net power delivered by the source is the power of the wave from the source minus the power reflected back into the source, i.e., $P_{\text{source}} = |V_{1\text{rev}}|^2(1 - |\Gamma_{\text{in}}|^2)$, where Γ_{in} is the reflection coefficient already calculated (Equation 28.11). To find $V_{1\text{fwd}}$, we can substitute the expression for $V_{2\text{rev}}$ back into Equation (28.13). Stirring this all together, we find

$$G_{\text{operating}} = \frac{|S_{21}|^2(1 - |\Gamma_L|^2)}{|1 - \Gamma_L S_{22}|^2(1 - |\Gamma_{\text{in}}|^2)}. \tag{28.17}$$

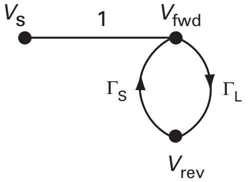

Figure 28.6. Signal flow graph for a source driving a load. The source would supply a forward wave of amplitude V_S to a load for which $\Gamma_L = 0$.

To calculate the other gain expressions, we must find expressions for the available powers. Figure 28.6 shows a load connected to a generator. Again, this is a generator that would supply V_S volts to a resistive load of Z_0 ohms.

From this signal flow graph we see that $V_{\text{fwd}} = V_S + V_{\text{fwd}} \Gamma_L \Gamma_S$ or $V_{\text{fwd}} = V_S/(1-\Gamma_L \Gamma_S)$. The net power to the load is $|V_{\text{fwd}}|^2/Z_0 - |V_{\text{rev}}|^2/Z_0 = |V_{\text{fwd}}|^2 (1-|\Gamma_L|^2)$. We know that maximum power is transferred to the load when $Z_L = Z_S^*$, from which $\Gamma_L = \Gamma_S^*$. Putting this together, the maximum available power from the source becomes $|V_S|^2/[Z_0 (1-|\Gamma_S|^2)]$. In the same way, the maximum available power from the network is $|V_{2\text{rev}}|^2/[Z_0 (1-|\Gamma_{\text{out}}|^2)]$. Finally, we can also find the voltage amplification factor $A_V = V_2/V_1$, since we know that $V_1 = V_{1\text{fwd}} - V_{1\text{rev}}$ and $V_2 = V_{2\text{fwd}} - V_{2\text{rev}}$.

28.3 Stability of an active two-port (amplifier)

If the input impedance measured at any port of a device has a negative real part – a negative resistance – the device can supply power. Obviously this can only occur with active devices, such as amplifiers, which are connected to an external source of power. An impedance with a negative resistance value produces a reflection coefficient, Γ, whose magnitude is greater than unity (Problem 28.2). Thus, if $|S_{ii}| > 1$, the device has the potential for oscillation. It will oscillate unless the port is terminated with an impedance whose real part is larger than the negative resistance of the port. To see if an amplifier is potentially unstable, we check whether there are any values of load impedance that will make $|\Gamma_{\text{in}}| > 1$ and also whether there are any values of source impedance that will make $|\Gamma_{\text{out}}| > 1$. Consider first the load impedances. We use the expression for Γ_{in}

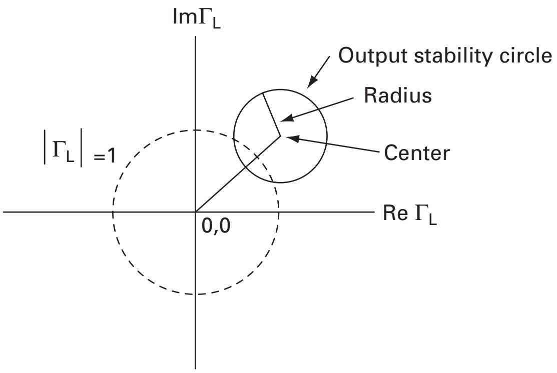

Figure 28.7. Output stability circle in the Γ_L plane.

given in Equation (28.11) and find all the values of Γ_L that will make $|\Gamma_{in}| = 1$. These values of Γ_L will form a closed curve, always a circle,[1] that divides the Γ_L plane into two regions. The values of Γ_L in one region result in $|\Gamma_{in}| < 1$, while the values of Γ_L in the other region result in $|\Gamma_{in}| > 1$ (potential instability). You can use Equation (28.11) to show that the circle in the Γ_L plane, called the *output stability circle*, will be centered at the point

$$\Gamma_L(\text{center point}) = \left| \frac{S_{22}^* - \Delta^* S_{21}}{|S_{22}|^2 - |\Delta|^2} \right|, \tag{28.18}$$

where $\Delta = S_{11}S_{22} - S_{12}S_{21}$, and will have a radius given by

$$\text{rad}_L = \left| \frac{S_{12}S_{21}}{|S_{22}|^2 - |\Delta|^2} \right|. \tag{28.19}$$

Figure 28.7 shows an example of an output stability circle. Suppose we have found that the “good” values of Γ_L are in the region outside the output stability circle (for example, by verifying that $|\Gamma_{in}| < 1$ for $\Gamma_L = 0$). For this example, the output stability circle partially overlaps the $|\Gamma_L| = 1$ circle. Values of Γ_L in the overlap region are “bad” values; they will give the amplifier a negative input resistance, even though these Γ_L values correspond to passive loads. The “bad” values outside the overlap region all have $|\Gamma_L| > 1$ so they represent loads with negative resistance. These values could be important if the load is not passive, for example, the input of an amplifier which might have $|\Gamma_{in}| > 1$.

---

[1] The locus of points $|\Gamma_{in}| = 1$ is a circle in the Γ_{in} plane. The expression for Γ_{in} (Equation 28.11) can be written in the form $\Gamma_{in}(z) = (a + b\,\Gamma_L)/(c + d\,\Gamma_L)$, where a, b, c, and d are constants. This form, a *bilinear transformation*, maps circles to circles. Thus the values of Γ_L that map to a circle in the Γ_{in} plane must form a circle in the Γ_L plane. In the same way, the constant resistance and constant reactance circles on the Smith chart (Γ plane) are mapped from the Z-plane, where they are segments of infinite-radius circles (lines).

As you can imagine, *input stability circles* are found in the same way. Interchange twos and ones in Equations (28.18) and (28.19) to get the radius and center point for the input stability circle. If both the output and input stability circles lie totally outside the $\Gamma < 1$ circle, an amplifier is unconditionally stable, i.e., no combination of passive terminations will cause it to oscillate.

28.4 Cascaded two-ports

Two-port devices are often connected in series, as in cascades of amplifiers and filters or in amplifiers, as cascades of components. Obviously a cascade of two two-port devices must be equivalent to a single two-port device, and the *S* parameters of the equivalent device must be functions of the *S* parameters of the individual devices. In Figure 28.8, a cascade is fed by a reflectionless source providing a wave of amplitude unity. We place a reflectionless load at the far end. Therefore, the wave reflected back into the source will be S_{11}, while the wave incident on the load will be S_{21}.

Because of the matched load and matched source, this circuit is quite simple to analyze. Let *x* represent the value of the signal at the top center, the forward wave into device B. By inspection of the graph we can immediately write an equation for *x*: $x = S^A{}_{21} + x\, S^B{}_{11} S^A{}_{22}$ or $x = S^A{}_{21}/(1\text{-}S^B{}_{11} S^A{}_{22})$. Therefore $S_{11} = S^A{}_{11} + x S^B{}_{11} S^A{}_{12}$ or

$$S_{11} = S^A_{11} + S^A_{21} S^B_{11} S^A_{12}/(1 - S^B_{11} S^A_{22}). \qquad (28.20)$$

Note that we had earlier found the reflection coefficient of a two-port terminated by an arbitrary load. In this case the arbitrary load is just $S^B{}_{11}$. Again by inspection, S_{21} is just $x S^B{}_{21}$ or

$$S_{21} = S^B_{21} S^A_{21}/(1 - S^B_{11} S^A_{22}). \qquad (28.21)$$

Interchanging S^A's with S^B's and ones with twos gives us the remaining two S parameters:

$$S_{22} = S^B_{22} + S^B_{12} S^A_{22} S^B_{21}/(1 - S^A_{22} S^B_{11}) \qquad (28.22)$$

$$S_{12} = S^A_{12} S^B_{12}/(1 - S^A_{22} S^B_{11}). \qquad (28.23)$$

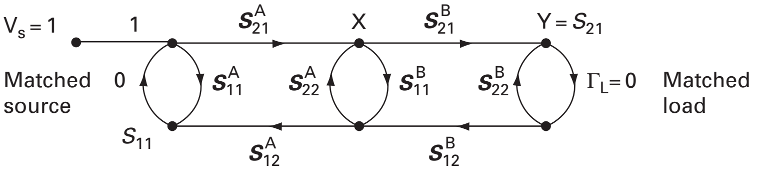

Figure 28.8. Circuit for calculation of overall *S* parameters for a cascade of two two-port devices having *S*-parameter sets $\{S^A\}$ and $\{S^B\}$.

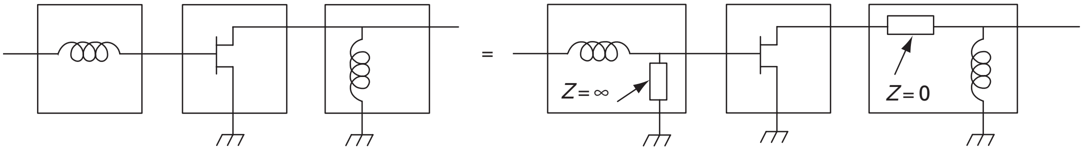

Figure 28.9. An amplifier drawn as a cascade of two-port networks.

In an amplifier, transistors are obvious two-port devices. Shunt and series impedances can also be treated as two ports, as shown in Figure 28.9, which shows that they are just special cases of the network of Figure 28.2.

28.5 Reciprocity

In circuit theory, a common statement of the reciprocity theorem is that if a voltage source V is inserted into any branch of a circuit, branch n, say, and it causes a current of I amps to flow in branch m, then the same current, I, would flow in branch n if the voltage source had been inserted in branch m. This is a conditional theorem; the circuit must contain only passive linear elements (no tubes or transistors or other dependent sources). Moreover, the circuit must not contain nonreciprocal elements, i.e., *circulators* or *isolators* which are microwave components that contain ferrites – magnetic materials in which the permeability is a tensor ($\boldsymbol{H}$ and $\boldsymbol{B}$ are not in the same direction). We used this theorem in Chapter 20 to find the relation between an antenna's receiving performance (effective capture area) and its transmitting performance (directional gain). In the context of S-parameter analysis, the theorem requires that $S_{ij} = S_{ji}$, i.e., the S matrix must be symmetric, even when the circuit itself is not symmetric, as in Figure 28.2. (Reciprocity takes the same form with the Z and Y parameters; $Z_{ij} = Z_{ji}$ and $Y_{ij} = Y_{ji}$.)

28.6 Lossless networks

A lossless network contains no resistive elements. For such networks, conservation of energy requires that the average total power going into the network be equal to the average total power leaving the network. (In S-parameter analysis we are dealing with ac, so the averaging time, strictly speaking, should be an integral number of cycles.) Let us see how the constraint of energy conservation puts constraints on the S parameters.

Suppose we connect a generator of output impedance Z_0 to port i of an N-port network while all the other ports are properly terminated. Let this generator supply a forward wave whose amplitude is unity. No power will flow into the network from the other ports, since they all have nonreflective terminations. Assuming normalized impedances, the total power flowing into the network is therefore 1. The total power flowing out will be the sum of $|S_{ii}|^2$, the power reflected back into the generator plus $\sum|S_{ji}|^2$, summed over the other ports. Putting this into a single sum and equating the powers, we have

$$\sum_{j=1}^{N}\left|S_{ji}\right|^2 = 1. \tag{28.24}$$

This says that the sum of the squared magnitudes of any column of the S matrix must be unity, if the network is lossless. And, since $S_{ij}=S_{ji}$, the same is true for any row. Next, consider the situation where power is fed simultaneously into *two* ports. Let us apply a forward wave of amplitude V_i to port i and a forward wave of amplitude V_k to port k. Again, we use reflectionless generators ($Z_{\text{out}}=Z_0$) and terminate the other $N-2$ ports in Z_0. The total input power will be $|V_i|^2+|V_k|^2$. This must equal the total output power, which is the sum of the squares of the magnitudes of the waves leaving the ports, i.e.,

$$|V_i|^2+|V_k|^2=\sum_{j=1}^{N}\left|(V_iS_{ji}+V_kS_{jk})\right|^2. \tag{28.25}$$

Expanding the right-hand side we find

$$|V_i|^2+|V_k|^2=|V_i|^2\sum_{j=1}^{N}\left|S_{ji}\right|^2+|V_k|^2\sum_{j=1}^{N}\left|S_{jk}\right|^2+2\mathrm{Re}\left(V_iV_k^*\sum_{j=1}^{N}S_{ji}S_{jk}^*\right). \tag{28.26}$$

From Equation (28.24), we see that the first two terms on the right are equal to the first two terms on the left, leaving us with

$$0=2\mathrm{Re}\left(V_iV_k^*\sum_{j=1}^{N}S_{ji}S_{jk}^*\right). \tag{28.27}$$

Since V_i and V_k are arbitrary, their product could be entirely real or entirely imaginary which requires that, for $i\neq k$,

$$\sum_{j=1}^{N}S_{ji}S_{jk}^*=0. \tag{28.28}$$

Thus, for a lossless network, the "dot product" of one column of the S matrix times the complex conjugate of any different column, must be zero. The same is true for rows since $S_{ij}=S_{ji}$.

Let us apply Equations (28.24) and (28.28) to a lossless two-port network. A general two-port is described by four complex S parameters, i.e., eight real numbers. If the network is reciprocal, $S_{ij}=S_{ji}$, reducing the eight real numbers to six. If the network is lossless, Equation (28.24) gives us $|S_{21}|^2+|S_{11}|^2=1^2$. Thus the magnitude of the transmitted wave is determined completely by the magnitude of the reflected wave, $|S_{21}|^2=1-|S_{11}|^2$, i.e., we could determine the bandpass shape of a lossless filter by just measuring its input reflection coefficient (with the output correctly terminated). Equation (28.24) also gives us $|S_{12}|^2 + |S_{22}|^2 = 1$. With these two equations, the number of real parameters drops to four. Note that, from these two equations, it also follows that $|S_{22}|^2 = |S_{11}|^2$. Next we apply Equation (28.28), giving us

$$S_{11}S_{12}^* + S_{21}S_{22}^* = 0. \tag{28.29}$$

This appears to be two equations (real and imaginary parts), but actually amounts to only one equation [1]. To see this we express the S parameters in polar form as $S_{11} = |S_{11}|e^{j\theta_1}$, $S_{22} = |S_{11}|e^{j\theta_2}$, and $S_{21} = (1 - |S_{11}|^2)^{½} e^{j\phi}$. Putting these expressions, which contain four parameters, into Equation (28.28) will give us $[e^{j(\theta_1-\phi)} + e^{j(\phi-\theta_2)}] = 0$, from which $\theta_2 = 2\phi - \theta_1 + (2n+1)\pi$ where n is an integer. Any odd multiple of π will be equivalent, so we can set $\theta_2 = 2\phi - \theta_1 + \pi$. Thus we see that a lossless two-port network can be completely specified by only *three* real numbers, $|S_{11}|$, θ_1 and ϕ.

If we have a vector reflectometer (a one-port network analyzer) and measure S_{11} and S_{22} for a lossless two-port, we will have values for $|S_{11}|$, θ_1, $|S_{22}|$ and θ_2. Writing ϕ in terms of θ_1 and θ_2 gives $\phi = [\theta_2 + \theta_1 - (2n+1)\pi]/2$. If n is increased by one, ϕ changes by π. Therefore, while measurements of either S_{11} or S_{22} let us calculate the magnitude of S_{12}, measurements of both S_{11} and S_{22} allow us also to calculate the phase of S_{12}, but with a 180° ambiguity. (To see that this must be the case, suppose we attach a half-wave cable to one side of a two-port network. Then S_{11} and S_{22} will remain unchanged, but S_{12} will change by 180°.)

Since it can be described by three real numbers, a lossless two-port can also be modeled (for one frequency) as a network containing three parameters. As an example of such a network, suppose we start with the capacitor, which, if paralleled with a 50-ohm resistor, would produce an impedance Z, such that $|(1-Z)/(1+Z)| = |S_{11}|$. Let this capacitor be a shunt element to ground. To its left, install a 50-ohm cable with the length needed to get the correct phase for S_{11}. To its right, install a 50-ohm cable whose length produces the specified phase for S_{22}. One capacitor value and two lengths make up the three parameters.

Finally, since a lossless two-port must contain only reactors (no resistors), Z or Y-parameter analysis will show that the Z or Y parameters must be completely imaginary. Add in reciprocity, and four imaginary numbers become three imaginary numbers, showing us again that a lossless two-port can be specified by three parameters.

Three-port lossless networks

We saw in Chapter 15 that hybrids are four-port networks that are completely matched, that is, the impedance looking into any port will be equal to Z_0 if the other three ports are terminated in Z_0. Let us show by contradiction that a lossless reciprocal three-port network cannot have this property. Suppose, then, that we have a matched three-port, i.e.,

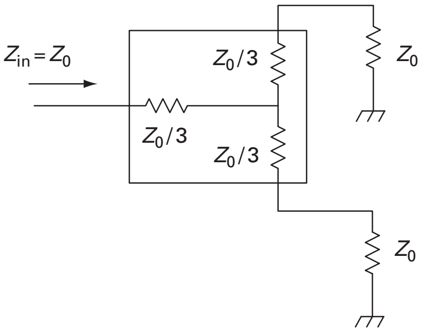

Figure 28.10. 6-dB matched power divider.

$$S = \begin{bmatrix} 0 & S_{12} & S_{13} \\ S_{12} & 0 & S_{23} \\ S_{13} & S_{23} & 0 \end{bmatrix}, \qquad (28.30)$$

where we have also assumed reciprocity, i.e., $S_{21}=S_{12}$, $S_{23}=S_{32}$, and $S_{13}=S_{31}$. Applying Equation (28.28) to rows one and two, we find $S_{13}\,S_{23}=0$. One of these parameters must be zero, say, S_{13}. The S matrix then becomes

$$S = \begin{bmatrix} 0 & S_{12} & 0 \\ S_{12} & 0 & S_{23} \\ 0 & S_{23} & 0 \end{bmatrix}. \qquad (28.31)$$

Finally, applying Equation (28.24) to row one yields $S_{12}=0$ and applying it to row three yields $S_{23}=0$.

Thus, if a lossless reciprocal three-port network is matched, all of its elements are zero. Power going in any port is totally absorbed – the opposite of lossless. Three resistors, each of value $Z_0/3$ in “T” connection, as shown in Figure 28.10, form a matched but lossy three-port network that is often used as a power divider.

A circulator is a matched nonreciprocal device; it contains ferrite material. You can show (Problem 28.4) that any lossless matched three-port must be a circulator. Power entering port 1 exits from port 2; power entering port 2 exits from port 3; and power entering port 3 exits from port 1.

Directional couplers

In Chapter 15 we looked at hybrids, which are 3-dB directional couplers. From Equations (28.24) and (28.28) it can be shown that any lossless matched ($S_{ii}=0$ for all i) reciprocal four-port network must be a directional coupler and that, conversely, and directional coupler will be matched. Figure 28.11 is a signal flow graph of the general directional coupler. Opposite ports are isolated from each other, i.e., $S_{14}=0$ and $S_{23}=0$. The real number c is the coupling coefficient. If $c=0.1$ for example, the power coupling is $c^2=0.01$ and the device is a 20-dB directional coupler. For hybrids, $c^2=1/2$.

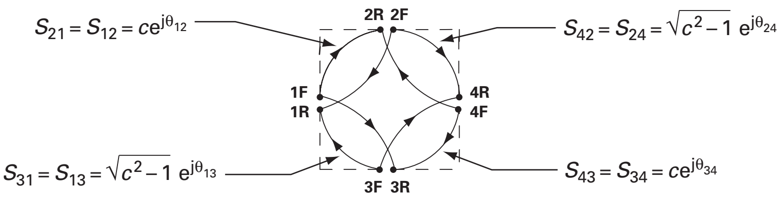
$S_{21} = S_{12} = ce^{j\theta_{12}}$
2R 2F
$S_{42} = S_{24} = \sqrt{c^2 - 1}\, e^{j\theta_{24}}$
1F
1R
4R
4F
$S_{31} = S_{13} = \sqrt{c^2 - 1}\, e^{j\theta_{13}}$
3F 3R
$S_{43} = S_{34} = ce^{j\theta_{34}}$

Figure 28.11. Directional coupler signal flow graph.

From Equation (28.28) you can show that

$$e^{j(\theta_{12}-\theta_{24})} + e^{j(\theta_{13}-\theta_{34})} = 0, \tag{28.32}$$

which imposes the condition

$$\theta_{12} - \theta_{24} = \theta_{13} - \theta_{34} \pm \pi. \tag{28.33}$$

Problems

Problem 28.1. The net power absorbed by a load is $P = (|V_{\text{fwd}}|^2/Z_0) - 1/2\,(|V_{\text{rev}}|^2/Z_0)$, i.e., the net power delivered to the load. Show that this is equal to the power expressed by the current voltage product at the load, $P = \text{Re}(VI^*)$.

Problem 28.2. Show that if the real part of Z is negative, the reflection coefficient, $(Z-Z_0)/(Z+Z_0)$, will have a magnitude greater than unity.

Problem 28.3. Derive an expression for S_{11} of a two-port in terms of its Y parameters. Hint: use Equation (28.5), together with the definition of the Y parameters: $I_1 = V_1Y_{11} + V_2\,Y_{12}$ and $I_2 = V_1Y_{21} + V_2Y_{22}$. Remember that port 2 is terminated in Z_0.

Problem 28.4. Three 50-ohm transmission lines are connected as shown in the figure to form a three-port network. The electrical lengths of these lines are θ_1, θ_2, and θ_3. Find the S parameters of this network. Hint: first find the S parameters when lengths are zero and then determine the appropriate phase factors.

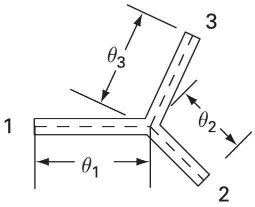

Problem 28.5. Draw a circuit for a three-port network whose S parameters are all zero.

Problem 28.6. Show that a matched three-port device must be a circulator. Hint: write down the S matrix with zeros on the diagonal (the matching condition) but do not assume $S_{ij} = S_{ji}$ for the off-diagonal elements. Use Equations (28.24) and (28.28).

Problem 28.7. A linear three-port network has ports 2 and 3 terminated with Γ_2 and Γ_3, respectively. Find the reflection coefficient, Γ, looking into port 1. Hint: the signal flow graph is shown below. A forward wave of magnitude 1 is assigned to port 1. Select the two intermediate points, y and z, and find their values by first writing an equation for each one by inspection of the graph. Then, by inspection of the graph, $\Gamma = S_{11} + y\,\Gamma_3\,S_{13} + z\,\Gamma_2\,S_{12}$.

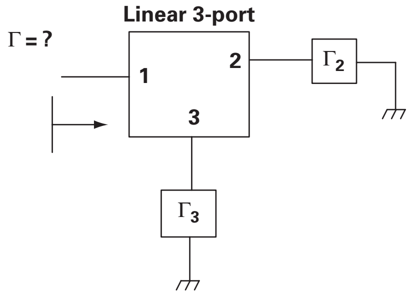

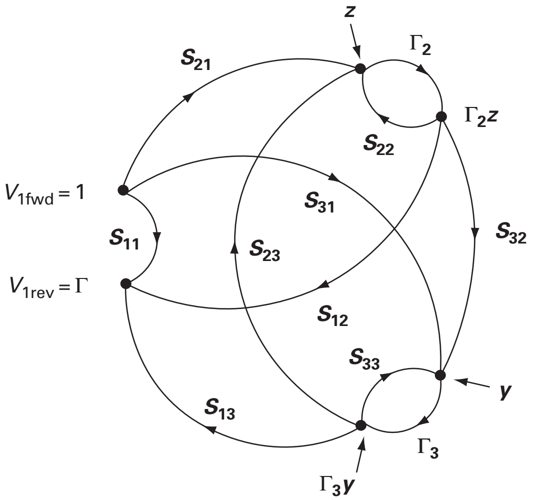

References

[1] Collin, R. E., *Foundations for Microwave Engineering*, New York: McGraw-Hill, 1966.

[2] Gonzalez, G., *Microwave Transistor Amplifiers – Analysis and Design*, Englewood Cliffs: Prentice-Hall, 1984.

[3] Montgomery, C. G., Dicke, R. H. and Purcell, E. M., Principles of Microwave Circuits, Volume 8 of the MIT Radiation Laboratory Series, New York: McGraw Hill 1948. (Reprinted by various publishers.)

[4] Pozar, D. M., *Microwave Engineering*, 2nd edn, New York: Wiley, 1998.

CHAPTER

29 Power supplies

Power supplies provide the constant dc voltage(s) used to power amplifiers, oscillators, and logic circuitry. As switching frequencies increase, RF engineers need to understand switching power supply techniques, which are often incorporated in the designs of high-efficiency RF amplifiers, and power supply designers need to learn RF techniques. Conventional power supplies, originally known as "battery eliminators," operate from the ac power line and are usually transformer/rectifier/filter combinations. Dc-to-dc converters are called switching power supplies. Both conventional and switching power supplies are discussed in this chapter.

29.1 Full-wave rectifier

Probably the most straightforward circuit combines a full-wave rectifier with a choke-input filter. A "choke" is an inductor which ideally has infinite inductance and no dc resistance. Two such power supplies are shown in Figure 29.1.

Figure 29.1. Choke-input power supplies: (a) full-wave; (b) full-wave bridge.

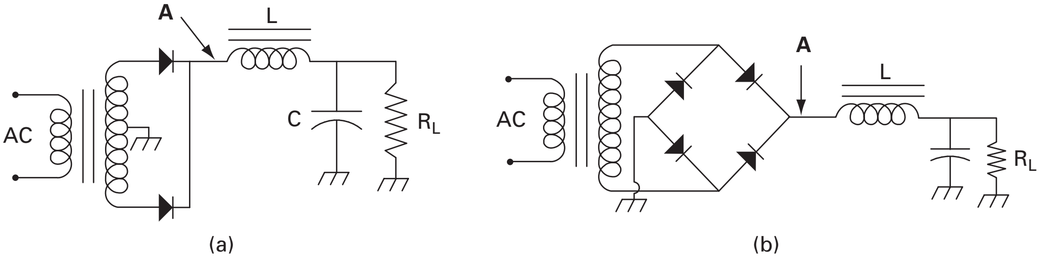

The full-wave rectifier delivers power on both the positive and negative halves of the input ac waveform. The bridge version does not require a center-tapped transformer and needs only half as much copper in the secondary winding, since it uses the entire secondary winding on both halves of the cycle. The two-rectifier circuit, on the other hand, has only half as much rectifier voltage drop and, for that reason, is usually chosen for low-voltage supplies (five volts

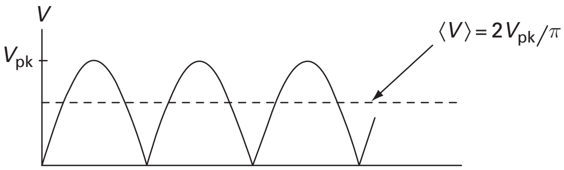

Figure 29.2. Voltage waveform at the filter input. The average value of this waveform is the dc voltage at the output of the filter.

or less) in order to maintain high efficiency. For either circuit, the voltage at point "A" should be as shown in Figure 29.2.

29.1.1 Inherent regulation of the choke-input power supply

Here regulation means voltage stability with respect to changes in the load, R_L. (A separate consideration is regulation with respect to variations in the input ac voltage.) It is important to see that a choke-input power supply provides constant output voltage as long as the current through the choke is never zero. Here is why: If the current through the choke is not zero, then either one rectifier (or rectifier pair) or the other is forward biased and has nearly zero voltage drop. The left side of the choke is therefore always connected to the transformer secondary and the voltage will be as shown in Figure 29.2, complete half-cycle cusps from zero to V_{pk}, the peak secondary voltage. This waveform, in as much as it derives from an almost ideal voltage source (the power line), is practically independent of load current. Its average value is $V_{pk}(2/\pi)$. Since there can be no dc voltage across an (ideal) inductor, the dc component on R_L is the same, V_{pk} $(2/\pi)$, and the regulation is perfect except for the small ohmic losses in the rectifiers, transformer windings and choke, and any leakage inductance in the transformer.

Let us see what is necessary to maintain a nonzero current in the choke. The L-section *LC* filter suppresses ripple, so the ac voltage at the right-hand side of the inductor is negligible compared to the ac voltage at the left-hand side. At the left-hand side, most of the ac voltage is the 120-Hz component of the rectifier output. Since the peak rectifier voltage is V_{pk}, the peak voltage of the 120-Hz Fourier component is $4V_{pk}/3\pi = 0.42V_{pk}$. The peak ac current through the choke will be $4V_{pk}/(3\pi\omega L)$. To keep the current from turning off the rectifiers and bottoming out, the dc current must be greater than this peak ac current. Therefore we have $2V_{pk}/(\pi R_L) > 4V_{pk}/(3\pi\omega L)$ or $L > 2R_L/3\omega$. This minimum inductance is known as the critical inductance,

$$L_{\text{critical}} = 2R_L/3\omega. \tag{29.1}$$

In some applications the load may vary such that the current drops to very low values, even zero. A minimum current can be guaranteed by paralleling a "bleeder" resistor across the load. Unfortunately, the resistor will consume

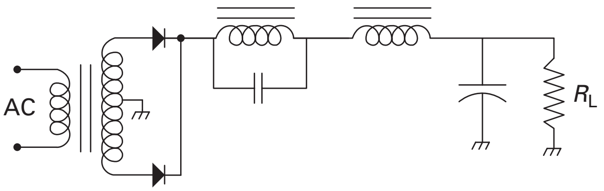

Figure 29.3. Power supply with a resonant filter.

some power. High-voltage power supplies often include a bleeder for safety; it discharges the filter capacitor when the supply is turned off. Another method to maintain constant current is to use a "swinging" choke which has high inductance when needed (i.e., at low current) but presents lower, though still sufficient, inductance at high current, where the core goes into partial saturation.

Ripple

The L and C form a simple L-section lowpass filter. In a full-wave power supply the peak value of the "ripple," the 120-Hz sine wave at the input of the filter (left-hand side of the choke) is $0.42V_{\text{pk}}$. The capacitor will normally have a reactance much smaller than the load resistance, R_L, so we can find the output ripple voltage by considering just the L and C as a voltage divider:

$$V_{\text{ripple(peak)}} = \frac{0.42V_{\text{pk}}X_C}{X_L - X_C}. \tag{29.2}$$

Sometimes a resonant LC circuit is used, as in Figure 29.3, as a bandpass filter to suppress the primary ripple frequency even more.

29.2 Half-wave rectifier

The simplest rectifier circuit, shown in Figure 29.4, uses a single diode. With just a load resistor (no filtering), the output waveform is zero during the negative half-cycles of the 60-Hz line voltage.

Obviously this arrangement does not make good use of the transformer secondary. Moreover, if we add the series L, shunt C filter to get a smooth dc output, the voltage waveform at a point "A" will float upward by an amount depending on the load and there will be no inherent voltage regulation. The solution is to add a "freewheeling" diode, as shown in Figure 29.5.

When the voltage at the top of the secondary becomes negative and "disconnects" the top diode, the freewheeling diode conducts and continues to supply current (from ground) to the inductor. The voltage at the left-hand side of the choke is therefore always the positive secondary voltage or zero volts, just as in Figure 29.4(b), and the circuit can have the inherent regulation of the full-wave choke-input power supply.

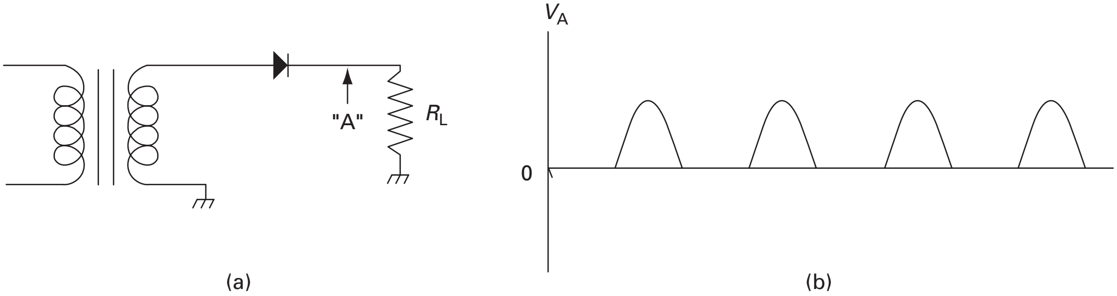

Figure 29.4. Simple half-wave rectifier circuit.

Figure 29.5. A half-wave circuit can also have choke-input regulation if a freewheeling diode is included.

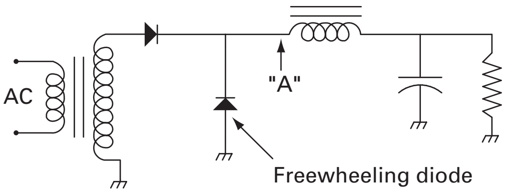

The simplest power supplies use an *RC* filter, substituting a resistor for the inductor. Power is wasted in the resistor and regulation is poor; the average voltage at the rectifier output depends on the load. Of course, for many low-power devices that draw constant current or can tolerate voltage variations, these disadvantages are slight, and this is the kind of circuit you will find in "wall wart" dc power adapters.

29.3 Electronically regulated power supplies

If a basic power supply, such as any of those described above, is followed by a series "pass transistor," a feedback circuit can control the pass transistor's resistance to maintain an extremely constant output voltage. A typical circuit is shown in Figure 29.6.

Note that the pass transistor is used as a series "dropping resistor." An alternate regulator circuit uses a series resistor followed by a parallel transistor as an active bleeder resistor to make a shunt-regulated power supply. These regulator circuits, really just class-A dc amplifiers with negative feedback, operate by dissipating power. Laboratory power supplies use this kind of circuitry to achieve precise regulation and to allow the user to set the desired voltage, as well as a protective current limit. If the load tries to exceed the current limit, the circuitry switches from a voltage regulation mode to a current regulation mode.

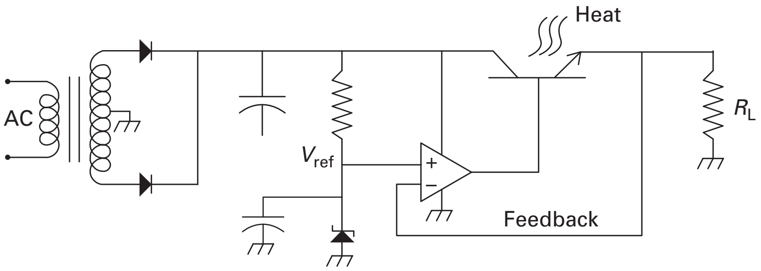

Figure 29.6. Electronically regulated power supply.

29.4 Three-phase rectifiers

Big power supplies, like big motors, benefit from three-phase power which allows more efficient power transmission and lighter transformers. Let us examine a few common circuits. Figure 29.7 shows a three-phase half-wave supply.

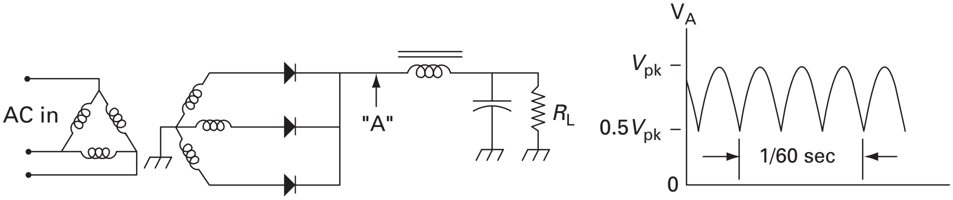

Figure 29.7. Three-phase half-wave power supply.

This is already an improvement over the full-wave single phase supply. The voltage at the input to the choke has a higher ripple frequency (180 rather than 120 Hz) and much less ripple, since the voltage at point “A” is always greater than $V_{pk}/2$. Choke-input filters can be used without the need for freewheeling diode(s). A three-phase full-wave rectifier is shown in Figure 29.8.

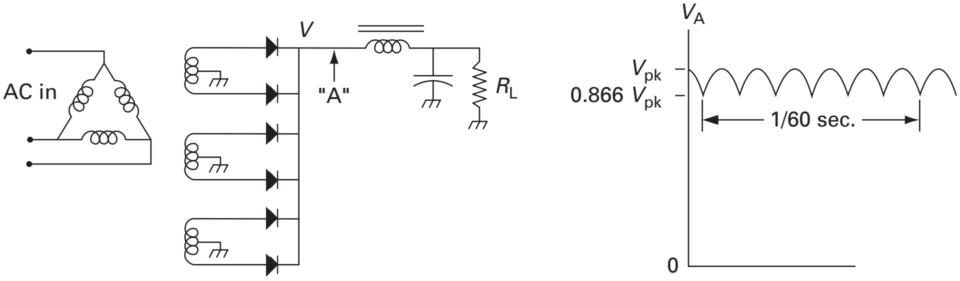

Figure 29.8. Three-phase full-wave power supply.

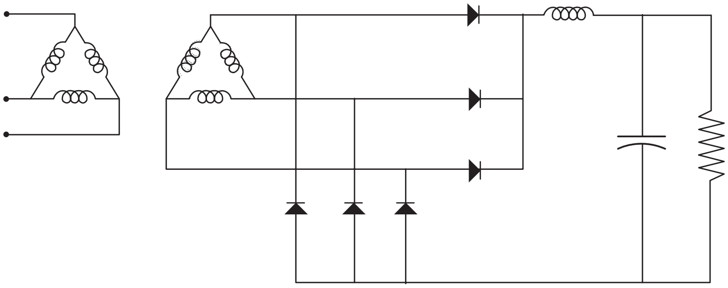

Figure 29.9. Three-phase full-wave bridge power supply.

This circuit uses center-tapped secondary windings. The voltage at the input to the choke now has very low ripple and the fundamental ripple frequency is 360 Hz (which allows even smaller values of L and C). Note that this is really a six-phase circuit; each rectifier connects to one of the six phases. Note again that the voltage from a three-phase or six-phase rectifier set never goes to zero. There is power available from a multiphase ac line at every instant. By comparison, an ordinary single-phase ac line suffers 120 power interruptions every second. A dc power supply operating off a single-phase line must include at least one energy storage element (C or L) to ride through these "power failures." On the other hand, an electronic regulator circuit like the one shown in Figure 29.6 could be powered by the three-phase rectifier of Figure 29.8 to make a dc supply having no energy storage elements. Such a supply would dissipate some of its input power in removing the ripples. When delivering maximum voltage, 0.866 V_{pk}, the efficiency would be 91%. A three-phase full-wave bridge circuit is shown in Figure 29.9.

This bridge rectifier has the same relation to the previous circuit as the four-diode full-wave bridge has to the two-diode ordinary full-wave circuit. Again there are six equivalent phases. You may recognize this circuit as the one used in automobile alternators.

29.5 Switching converters

Switching converters are dc-to-dc power converter circuits based on switching elements, often one transistor and one diode, and an inductor (sometimes the magnetizing inductance of a transformer). They are often used as local "board-level" power converters in systems with a single power bus. For example, a bus voltage of 20 volts dc can be converted wherever needed to power 5-volt logic circuitry, to furnish various positive and negative voltages for analog circuitry, and to drive high-voltage display systems. A switching converter can be preceded by a rectifier and filter capacitor to derive its input power directly from the ac power line. This combination, a *switching power supply*, is found, for example, in every line-powered computer. By varying the duty cycle of the

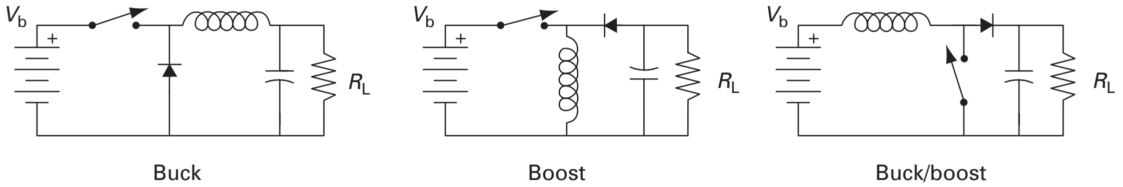

Figure 29.10. Three common dc-to-dc converters.

switching element, the dc output can be precisely regulated. If the duty cycle is deliberately modulated, the switching converter becomes a high-efficiency "class-D" power amplifier used, for example, as an AM modulator for a class-C RF amplifier or as a high-efficiency audio amplifier in a battery-operated MP3 player or hearing aid. The virtues of switching converters are their high efficiency and light weight. Three common converter circuits are shown in Figure 29.10. All include a switch, an inductor, a diode, and a filter capacitor. (The dc input supply is shown as a battery.)

In all these circuits an inductor provides the means by which energy is transferred from the supply side to the load side.

29.5.1 Buck circuit – continuous mode

The buck circuit is the same as a choke-input power supply fed by a half-wave rectifier with a freewheeling diode, except that here the filter sees rectangular voltage pulses rather than the positive halves of a sine wave. When the switch is open, the diode provides the left side of the inductor with a current path from ground. If the choke inductance is large enough to maintain nonzero positive current in the inductor, the average voltage at the input of the choke will be the supply voltage, V_b, times the duty factor of the switch (the fraction of the time the switch is on). The output voltage is equal to this average voltage, so it will remain constant with respect to load changes. (Perfect inherent regulation with respect to load changes requires ideal components, e.g., no dc resistance in the choke.) The duty factor may, of course, be changed to control the output voltage. Feedback control of the duty factor can be used to obtain regulation with respect variations of the input voltage. Feedback control will also provide very precise regulatizon with respect to load changes but, due to the inherent regulation, it will not have to make more than slight adjustments.

The current through the choke ramps up while the switch is closed and ramps down when it is open. Since $V = L\mathrm{d}I/\mathrm{d}t$, the slope of the current up-ramp is a constant, equal to the supply voltage minus the output voltage divided by the inductance. The slope of the down-ramp is also constant, equal to the output voltage divided by the inductance. This operation is shown in Figure 29.11.

When $V_{out} < V_b/2$ (i.e., the duty factor is $< 1/2$) the charging slope is steeper than the discharging slope and vice versa.

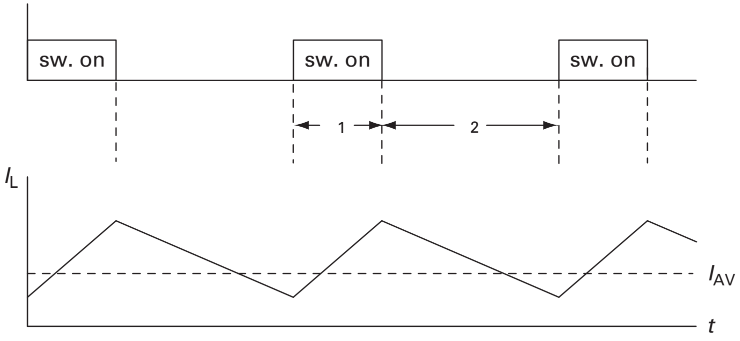

Figure 29.11. Buck supply operation cycle (continuous mode).

Buck circuit – discontinuous mode

If the inductor value is too low, the down-ramp will drain the inductor dry (zero current). The freewheeling diode will stop conducting and the left-hand side of the choke will float. The inherent regulation will be lost and output voltage will now be a function of the load. Figure 29.12 shows the operation cycle in this discontinuous mode. We will assume the capacitor is large enough to allow us to treat the output voltage as a constant.

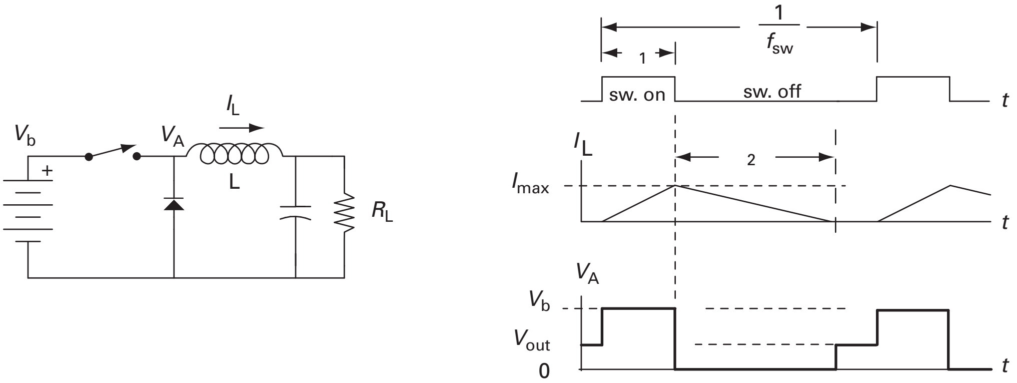

Figure 29.12. Buck supply operation cycle (discontinuous mode).

The length of the up-ramp is τ_1, the *on* time of the switch. In the discontinuous mode, the length of the down-ramp is the time needed for the inductor current to reach zero. To find the output voltage, we first note that the average current through the choke must be equal to the current in the load, V_{out}/R. From Figure 29.12, the average current through the choke is the area under the triangle divided by the length of the cycle, $1/f_{sw}$ where f_{sw} is the switching frequency. The duration of the down-ramp, τ_2, is given by $L\,I_{max}/V_{out}$. Finally, the up-ramp shows that I_{max} is equal to $(V_b-V_{out})\tau_1/L$. Putting all this together yields a quadratic equation for the output voltage, V_{out}.

$$\left(\frac{V_{\text{out}}}{V_{\text{b}}}\right)^2 + \frac{f_{\text{sw}}\tau_1^2 R}{2L}\left(\frac{V_{\text{out}}}{V_{\text{b}}}\right) - \frac{f_{\text{sw}}\tau_1^2 R}{2L} = 0. \tag{29.3}$$

But while this lets us find V_{out} in the discontinuous mode, we would prefer to avoid this mode altogether. To find the necessary critical inductance, we note that the discontinuous mode becomes continuous when there is no space between the triangular current pulses in Figure 12.12. The average current is then just half the height of the triangle, $I_{\text{max}}/2$ or $\frac{1}{2}\tau_1(V_{\text{b}} - V_{\text{out}})/L$. Since this average current has to be equal to V_{out}/R, we have

$$L_{\text{critical}} = \frac{\tau_1 R}{2}\left(\frac{V_{\text{b}}}{V_{\text{out}}} - 1\right). \tag{29.4}$$

29.5.2 Buck/boost circuit – continuous mode

The buck/boost circuit provides negative voltage, whose magnitude can be greater than the supply voltage. In this circuit, energy is pumped into the inductor while the switch is closed and is then transferred to the load when the switch is open. The inductor current vs. time, shown in Figure 29.11, also applies for the buck/boost circuit in the continuous mode. While the switch is closed, the inductor current ramps up linearly. The diode is back-biased during this time so the load side is disconnected. When the switch opens, the inductor current remains continuous, supplied now by the load via the diode. Since current is drawn from the load, the output voltage is negative. For the up-ramp we have $\Delta I = V_{\text{b}}\tau_1/L$. For the down-ramp we have $\Delta I = -V_{\text{out}}\tau_1/L$. The up and down ramps have equal current excursions, ΔI, so we have

$$V_{\text{out}} = -V_{\text{b}}\frac{\tau_1}{\tau_2}. \tag{29.5}$$

Note that, in the continuous mode, the output voltage is also inherently regulated, i.e., independent of the load.

Buck/boost circuit – discontinuous mode

Figure 29.13 shows the voltage and current waveforms for this circuit in the discontinuous mode. The analysis is not much different from that of the buck circuit. The maximum inductor current is given by $I_{\text{max}} = \tau_1 V_{\text{b}}/L$ and also by $I_{\text{max}} = -\tau_2 V_{\text{out}}/L$ so $\tau_2 = \tau_1 V_{\text{b}}/|V_{\text{out}}|$. Again, the load current is given by $|V_{\text{out}}|/R$ and must equal the average current furnished by the inductor. The latter is the area under the τ_2 triangle divided by the cycle length, $1/f_{\text{sw}}$, so we have

$$\frac{|V_{\text{out}}|}{R} = \frac{1}{2}\tau_2 \mathrm{I}_{\text{max}} f_{\text{sw}} = \frac{1}{2}\frac{V_{\text{b}}\tau_1}{|V_{\text{out}}|}\frac{V_{\text{b}}\tau_1}{L} f_{\text{sw}}. \tag{29.6}$$

Solving for V_{out} gives

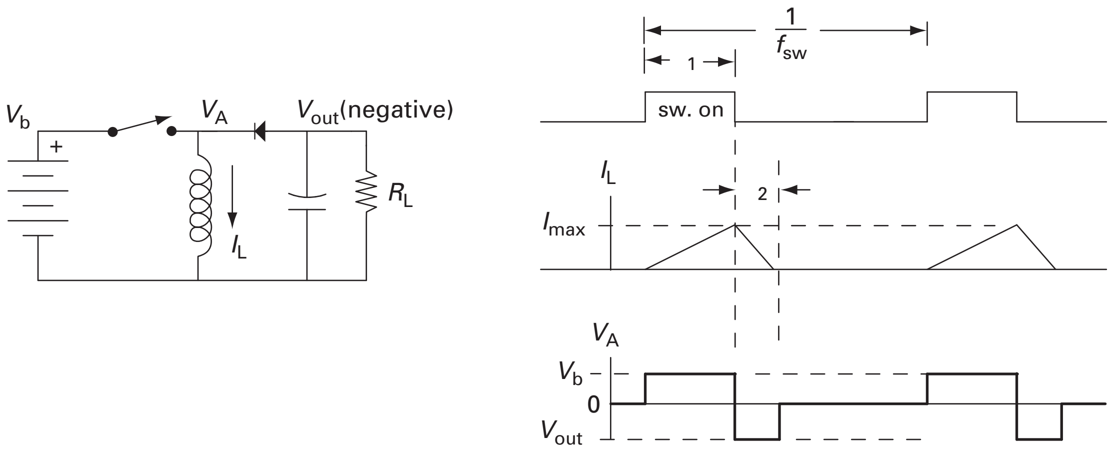

Figure 29.13. Buck/boost operation (discontinuous mode).

$$V_{out} = V_b \tau_1 \sqrt{\frac{f_{sw} R}{2L}}. \tag{29.7}$$

As before, the output voltage in this mode is a function of the load, so inherent regulation is lost. The critical inductance needed to stay out of this discontinuous mode is the subject of Problem 29.6(a).

29.5.3 Boost circuit – continuous mode

The boost circuit, like the buck/boost circuit, uses the inductor for energy storage. For the boost circuit, Figure 20.11 again represents the current in the inductor for the continuous mode. The inductor's charge and discharge become $\Delta I/\tau_1 = V_b/L$ and $\Delta I/\tau_1 = (V_{out} - V_b)/L$. The boosted voltage is therefore

$$V_{out} = V_b \left(1 + \frac{\tau_1}{\tau_2}\right). \tag{29.8}$$

Once again the continuous mode provides inherent regulation.

Boost circuit – discontinuous mode

Finally, we have the case of the discontinuous mode for the boost circuit. Figure 29.14 shows I_L and V_A vs. time. Again the current ramps up from zero when the switch is closed and ramps back down to zero when it is opened. As before, we assume the filter capacitor, C, is large so that V_{out} is constant. In this case the maximum inductor current is given by $I_{max} = \tau_1 V_b/L$ and also by $I_{max} = \tau_2 (V_{out} - V_b)/L$ so $\tau_2 = \tau_1 V_b/(V_{out} - V_b)$. Again, the current into the load is given by V_{out}/R and must equal the average current furnished by the inductor. The latter is the area under the τ_2 triangle divided by the cycle length, $1/f_{sw}$, so we have

$$\frac{V_{out}}{R} = \frac{1}{2}\tau_2 I_{max} f_{sw} = \frac{1}{2}\frac{V_b \tau_1}{V_{out} - V_b}\frac{V_b \tau_1}{L} f_{sw}. \tag{29.9}$$

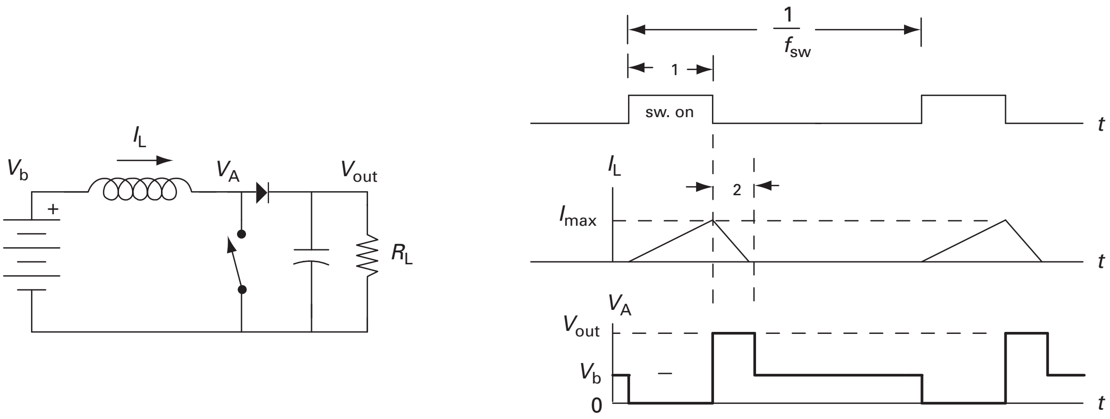

Figure 29.14. Boost supply operation cycle (discontinuous mode).

Rearranging gives us a quadratic equation for V_{out},

$$V_{out}(V_{out} - V_b) = V_b^2 \tau_1^2 \frac{f_{sw} R}{2L}. \tag{29.10}$$

29.5.4 Ćuk converter

In all three converter designs discussed above, an inductor is used as an energy storage element, to pass power from the input side to the output side. The Ćuk converter circuit, shown in Figure 29.15, uses a capacitor for this purpose.

The input and output inductors filter out ac currents (ripple). Their inductances can be large and, hence, they will store little pulsating energy. Unlike the previous converters, this converter will have negligible ripple currents at both input and output sides. Note also that the capacitor provides dc isolation between the input and output circuits, a useful feature for some applications. Like the buck/boost converter, the output voltage is negative and, for the continuous mode, is given by $V_{out} = -V_{in}\ \tau_1/\tau_2$. This can be seen as follows: Assume that the current in each inductor is essentially just a dc current, i.e., that the inductances are large enough to make any ripple currents very small. When the switch is off (open), current I_1 (from L_1) and current I_2 (from L_2) converge and flow through the diode to ground. During this time, V_A will, therefore, be zero, except for a small voltage drop across the diode. During this *off* period, of duration τ_2, the capacitor will receive an amount of charge $Q_{charge} = (I_1 + I_2)\tau_2$. When the switch is closed, V_A will go negative. The diode therefore stops conducting and I_2 flows into the capacitor. During this *on* period, of duration τ_1, the capacitor will therefore be (dis)charged by an amount $Q_{discharge} = I_2\tau_1$. In steady state, the capacitor must have same charge at the end of every cycle, so we can equate $Q_{discharge}$ and Q_{charge} to find

$$(I_1 + I_2)\tau_2 = I_2\tau_1. \tag{29.11}$$

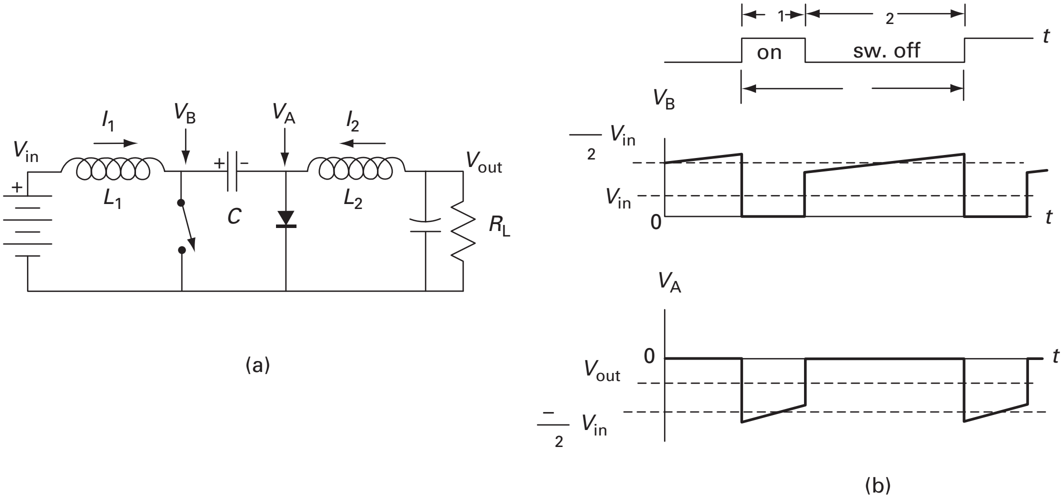

(a)

(b)

Figure 29.15. Ćuk converter uses a capacitor to transfer energy from input to output.

Next, consider that the input power from the dc supply must equal the power delivered to the load, since there are nominally no lossy components in the circuit. Setting $I_{in}V_{in}=I_{out}\,V_{out}$, we have

$$I_1 V_{in} = -I_2 V_{out}. \tag{29.12}$$

Combining these two equations and using the fact that $I_2=-V_{out}/R$, we confirm that

$$V_{out} = -V_{in}(\tau_2/\tau_1). \tag{29.13}$$

The waveforms are shown in Figure 29.15(b), for the continuous mode, which will be the operating mode unless the inductor values are too small. To verify these waveforms, remember that during the *on* period, $V_A=-V_{CAP}$ and during the *off* period, $V_B=V_{CAP}$. Moreover, there can be no average dc across the inductors, so $\langle V_A\rangle\tau_1/\tau= V_{out}$ and $\langle V_B\rangle\tau_2/\tau=V_{in}$, where $\tau= \tau_{1+}\tau_2$.

29.5.5 Transformer-coupled converters

When the dc input is derived from a rectifier connected directly to the ac power line, the three basic converters are modified by incorporating an isolation transformer. The transformer provides isolation from the power line, required for safety whenever external connections will be made to the circuitry powered by the converter. Operating at a high frequency, this transformer can be much smaller and lighter than a 60-Hz transformer of equivalent power. Transformer versions of the buck/boost and boost converters are known as *flyback* converters and can provide voltage of either polarity. Flyback converters require no inductor apart from the transformer. The inductance needed for energy storage is provided by the magnetizing inductance of the transformer. Positive and negative output flyback circuits are shown in Figure 29.16.

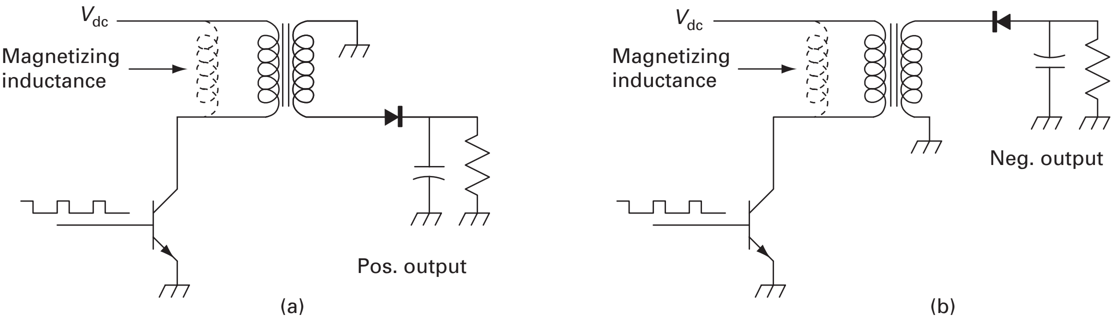

Figure 29.16. Flyback converters.

Figure 29.17 Forward converter.

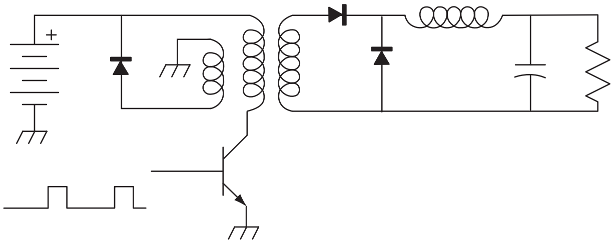

When the switch is closed, the diode disconnects the secondary and a ramp current flows in the primary, storing magnetic energy. When the switch is opened the transformer discharges, just as in the simple buck/boost circuit, except that now the discharge is through the secondary winding. (The magnetic energy does not care which winding carries the discharge current.)

The transformer version of the buck converter is the *forward converter* shown in Figure 29.17. Here the original choke remains; the transformer is an additional component.

While the switch is pulsed on, the transformer secondary supplies power to the load. Here the transformer is used as a real transformer rather than as an energy storage device. The turns ratio of the transformer, as well as the duty factor of the switch and the primary supply voltage, determine the output voltage. There is a problem, however, with opening the switch since there will be current flowing in the magnetizing inductance. (Remember that the magnetizing inductance is an equivalent shunt inductance on the primary side equal to the inductance of the primary winding or, just as well, a shunt inductance on the secondary side equal to the inductance of the secondary winding.) Even if the transformer has a high magnetizing inductance, there will be some stored energy when the switch is opened. A third winding can be provided, as shown above, to discharge this energy back into the primary supply, so that it is not wasted. This winding is sometimes called a *reset winding* since it resets the transformer to zero energy, readying it for the next pulse. In the forward converter, the transformer secondary delivers a stiff voltage pulse, i.e., a pulse from an equivalent low-impedance voltage generator. This circuit therefore needs a choke and a freewheeling diode to provide inherent voltage regulation.

Problems

Problem 29.1. The half-wave power supply shown below supplies 25 volts dc at 0.2 amps to the load (shown as a resistor). Assume the components are ideal and that the capacitor is large enough so that the ac voltage at the output (ripple) is very small.

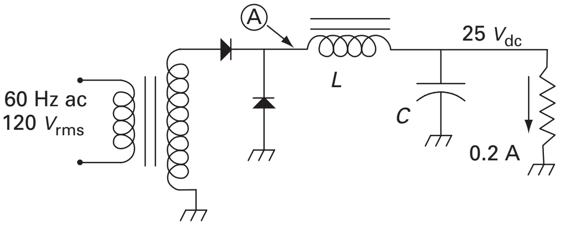

(a) Assuming the inductance of the choke is large enough to maintain positive current, sketch the voltage waveform at point "A."

(b) The value of the average voltage at "A" must be 25 volts since there is no dc drop across the choke. Find the rms voltage, $V_0/\sqrt{2}$, across the transformer secondary. Answer: 55.5 V rms.

(c) Calculate the peak value of the dominant ac component (the 60-Hz component) of the voltage at "A," i.e., find its Fourier amplitude. If you have trouble with this, estimate the magnitude of this component from your sketch. Answer: 39.3 V peak.

(d) Use the result from (c) to find the critical inductance of the choke, i.e., the inductance just large enough to keep the current in the choke always positive. Answer: 0.52 H.

(e) What value should the capacitor have to make the 60-Hz ripple at the output be less than 100 mV rms? Answer: 3700 μF.

Problem 29.2. Consider the common half-wave capacitor-input power supply shown below.

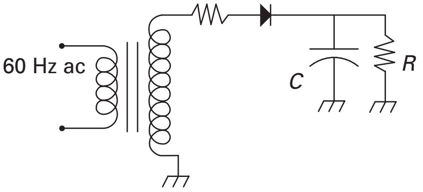

The resistors r and R represent, respectively, the ohmic resistance of the diode and the load resistance. Assume that the transformer is perfect and that C is large enough to make any ac voltage at the output negligible. Suppose that the peak voltage on the transformer secondary is 10 volts, and that the values of r and R are 1 ohm and 20 ohms. Calculate the output voltage, V_{dc}. Hint: the average current through the diode is equal to the current in the load, $V_{dc}/20$. Current flows in the diode only when the voltage at the top of the transformer secondary is greater than V_{dc}; the instantaneous value of the current is $(10 \cos \theta - V_{dc})/1$. Integrate this current over one cycle to find its average and set this expression equal to $V_{dc}/20$. You willl need a calculator to solve for the angle at which the current stops flowing. Answer: 7.51 V.

Problem 29.3. A not-so-common choke-input half-wave power supply is shown below. This circuit, without the free-wheeling diode, will not provide the inherent regulation of the circuit of Problem 19.1. If the load is constant, however, regulation is not important and with this circuit the voltage can be lowered (with no power loss) by increasing the inductance of the choke. Assume as before that the value of the capacitor is large enough to make any ac voltage at the output negligible.

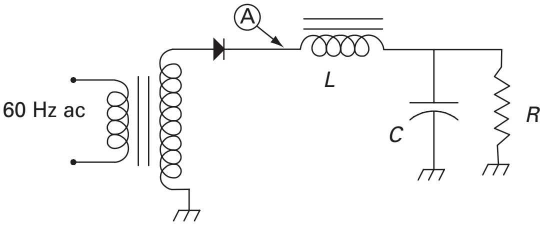

(a) Draw a sketch showing the voltage at point "A" and the current in the inductor. Hint: remember that $V = L\mathrm{d}I/\mathrm{d}t$. It may help you to consider the cases where R is very large and where R is very small.

(b) If the 60-Hz voltage at the transformer secondary is $10 \sin(\omega t)$ and the output voltage is 5 volts, find duty cycle of the diode, i.e., the angle at which the diode goes into conduction (where the diode becomes forward biased) and the angle at which it goes out of conduction (where the choke goes dry, i.e., choke current goes to zero).

(c) If the output current is 1 ampere, find the value of L.

Problem 29.4. Show that when an uncharged capacitor is brought to potential V by connecting it through a resistor to a battery of voltage V, the energy supplied by battery is twice the energy deposited in the capacitor (CV^2 versus $CV^2/2$), i.e., the transfer is only 50% efficient even if the value of the resistor is made very small.

Problem 29.5. Draw the voltage waveform at point "A" for continuous mode operation of the buck, buck/boost, and boost converters (see Figures 29.12, 29.13, and 29.14).

Problem 29.6. Find the critical inductance needed to stay out of the discontinuous mode (a) for the buck/boost converter of Figure 29.13, and (b) for the boost converter of Figure 29.14.

Problem 29.7. The diagram below shows a circuit for a class-S amplifier (switching amplifier). This could be an audio amplifier, in which case the load resistor represents a loudspeaker. The transistors operate as switches, i.e., fully on or fully off. The voltage on the left-hand side of the inductor is always either zero or V_{cc}. The average value of this voltage appears at the right-hand side of the inductor and also on the load (except that C_2 blocks dc).

(a) When the signal is zero, the duty cycle must be 50%. For this case, draw the waveforms of the currents in the inductor, Q_1, D_1, Q_2, and D_2.

(b) For this zero signal case, explain why there is no net energy flow from the power supply.

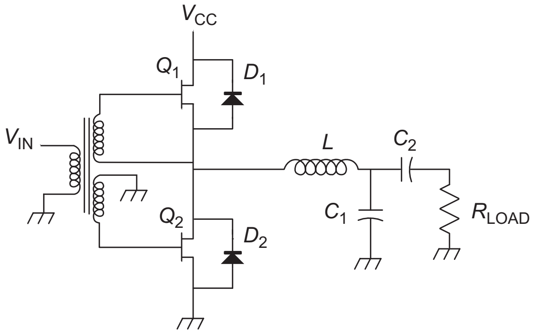

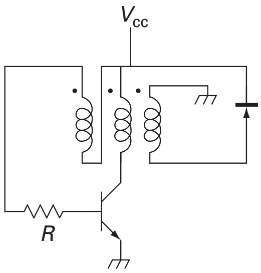

Problem 29.8. The circuit shown below, a blocking oscillator, is essentially a self-excited switching converter. The transformer has three windings. Assume they are perfectly coupled and have equal inductance, L. The current from the dc supply is a positive ramp during the first half of the cycle while the transistor is on. If the magnetic core does not saturate, this ramp ends when the transistor goes out of saturation, i.e., when $I_c = \beta I_b$. The transistor quickly turns off and the stored energy in the transformer is returned to the power supply via the diode.

(a) Show that the frequency is given by $f = R/(4\beta L)$. (Assume the base-to-emitter junction is a short circuit.) Sketch the waveforms of the collector voltage and of the collector and diode currents.

(b) Show that, if the magnetic core saturates before the transistor goes out of saturation, this circuit is a VCO or voltage-to-frequency converter.

References

[1] Kassakian, J. G., Schlect, M. F., and Verghese, G. C. *Principles of Power Electronics*, Addison Wesley, 1991.

[2] Mohan, N., Undeland, T. M. and Robbins, W. P. *Power Electronics*, 2nd edn, New York: John Wiley, 1995.

CHAPTER

30

RF test equipment

In RF work, as in low-frequency work, test instruments generally are designed either to analyze signals, or to determine parameters of components and systems. Signal analysis ranges from simple voltage, current, and power measurements to measurements of frequency, relative phase, frequency stability, modulation characteristics, RF field strength, and spectrum analysis. Parameter measurements range from simple resistance and reactance measurements through measurements of transfer characteristics of linear multiport devices (filter shapes, amplifier frequency response, etc.), and distortion (dynamic range, intermodulation). Parameter measurements require a signal generator to provide the stimulus whose response is measured. It is now common for measurement instruments to include built-in signal generators, but free-standing signal generators are widely used in specialized test setups.

30.1 Power measurements

Power measurements, as pointed out in Chapter 18, are best made with a square-law detector. If the waveform is unknown or noisy (i.e., has a random component), a square-law detector *must* be used, as power, by definition, is proportional to the average square of voltage. If the signal waveform is known, e.g., a sine wave, the power can also be calculated, for example, from a measurement of $\langle |V(t)| \rangle$. Commercial RF power meters normally present a 50-ohm impedance to source being measured, allowing a reflectionless connection to a 50-ohm cable.

Thermistor (temperature-dependent resistor) power meters are true square-law instruments and have a dynamic range typically from 1 μW (−30 dBm) to 100 mW (20 dBm). The power to be measured heats a thermistor which is one leg of a Wheatstone bridge circuit.[1] A feedback circuit supplies dc or low-frequency ac power to heat the a second *reference* thermistor in an adjacent leg in the bridge. When the heater power is equal to the input RF power, the two thermistors come to equal resistance values and the bridge is balanced. A feedback control circuit keeps the bridge balanced. The voltage × current product of the reference thermistor is calculated and this power is displayed. These instruments use small "bead"-type thermistors with very small heat capacities, so that when the power changes, the reading will come to equilibrium within less than one second. More refined circuits provide better compensation for changes in ambient temperature. Similar instruments use thermocouples instead of thermistors.

---

[1] With one end of the thermistor at ground, the RF signal is coupled to the other end through a dc blocking capacitor. A series RF choke isolates the input signal from the rest of the bridge circuit.

Powers down to about 10^{-3} μW (−60 dBm) can be measured with carefully calibrated *diode-based power meters*. An upper limit results from the need to keep the signal low enough that the fourth and higher even power terms in the diode I vs. V curve will be negligible (see Chapter 18). A typical upper limit is 100 mW (20 dBm). Thus, the nominal dynamic range of a diode-based power meter is three orders of magnitude greater than that of the thermistor-based power meter. However, when used to measure signals with high crest factors (signals such as noise or multiple carriers, whose peak powers greatly exceed their average power), the upper limit of the diode meters must be derated in order to preserve measurement accuracy.

Power meters measure the power a source would deliver to a standard impedance, usually 50+j0. Directional power meters measure the power traveling in each direction along a transmission line. This is best done with two power meters connected to a directional coupler with a low coupling factor. Negligible power is diverted from the main line and, in principle, the device introduces no reflections. Directional power meters are often installed permanently in the transmission line connecting an antenna to a transmitter as they show both that the transmitter is putting out its rated power and that the antenna is operating properly. Many times the reflected power detector is used to turn the transmitter off automatically or reduce power when a disturbance to the antenna has caused abnormal reflected power.

Flow calorimeters are used for high-power measurements and sometimes for low-power mm-wave signals. These contain a liquid-cooled load resistor ("dummy load"). The fluid temperature is measured on both sides of the load to get the temperature difference (a thermocouple pair can be used). The flow rate must also be measured. Note that all the power measurements discussed above are "true rms" measurements; they do not require that the unknown voltage be sinusoidal.

30.2 Voltage measurements

Probing any circuit with a voltmeter always inevitably "loads down" the circuit to some degree, so it is necessary that the impedance of the voltmeter be much

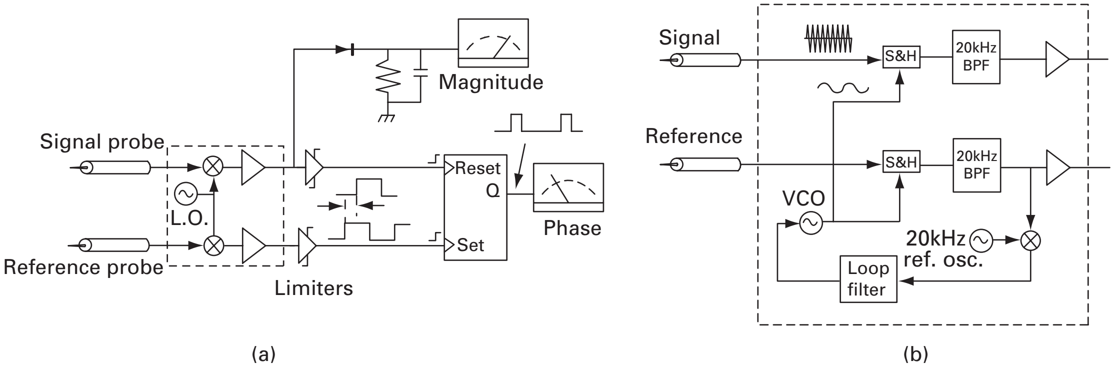

Figure 30.1. (a) Superheterodyne vector voltmeter; (b) phase-locked low-frequency oscillator drives sampling mixers.

greater than the equivalent Thévenin source impedance of the node being probed. For low-frequency work, it is sufficient that the meter probe present high shunt resistance. For RF work, the meter must present high shunt reactance as well.

An early innovative instrument was the *vector voltmeter* that displayed phase as well as amplitude. Since phase is relative, the instrument had an auxiliary "reference" input. A simplified circuit is shown in Figure 30.1(a). The signal and reference signal are both converted down to a low IF frequency, 20 kHz, and amplified. By mixing with identical L.O. signals, the relative phases are left unchanged. One channel is fed to a diode peak detector, to measure the amplitude. Both channels are fed to limiting amplifiers. The outputs of the limiters are used to produce a pulse whose duty factor is proportional to the phase difference or to operate a start/stop counter.

Figure 30.1(b) shows how the IF frequency was kept constant over the 1 MHz to 1 GHz range of the instrument. Sample and hold (S&H) circuits were used as mixers: an integral multiple of the sampling frequency differed by 20 kHz from the input frequency, causing the outputs of the S&H units to form a staircase version of a 20 kHz sine wave.[2] A phase lock loop controlled the sampling to keep the IF signal at 20 kHz. This scheme was also used in the receiver sections of vector network analyzers (VNAs), discussed below, instruments which have superseded the vector voltmeter.

30.3 Spectrum analysis

Spectrum analyzers, once called "panoramic receivers", are superheterodyne receivers with swept local oscillators. A nonlinear IF amplifier provides a logarithmic output voltage, i.e., the detector output is proportional to the log of the input signal power. This output is displayed on the *Y*-axis of a display

---

[2] With the sampling frequency low compared to the input frequency, the output of the sampler is an "aliased" version of the input, like the stroboscopic effect in movies or television that makes fast turning wheels slow down or appear to rotate backward.

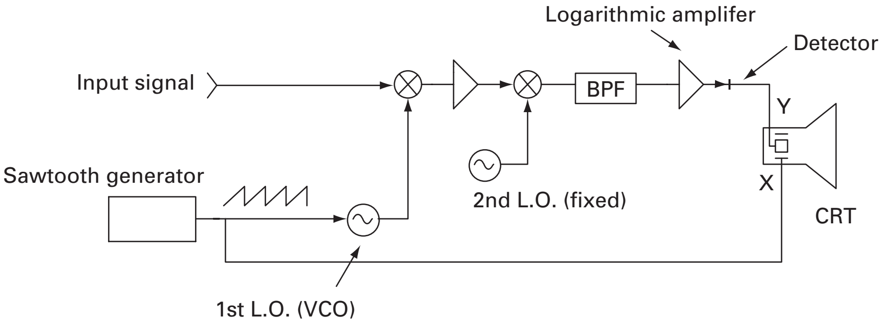

Figure 30.2. Spectrum analyzer.

while the *X*-axis is swept in sync with the local oscillator. The display is therefore a graph of signal strength vs. frequency. RF spectrum analyzers usually have a wide range such as 10 MHz to 20 GHz. A block diagram for such an instrument is shown in Figure 30.2.

A swept-frequency spectrum analyzer using a CRT has objectionable flicker in situations where the sweep time must be greater than about 1/20 second. The sweeping filter must spend a time at least equal to the reciprocal of its bandwidth on each frequency to get the desired resolution. The total sweep time must therefore be greater than the effective number of points divided by the resolution bandwidth which is the same as the frequency span divided by the square of the resolution bandwidth. For example, if it is necessary to analyze a span of 20 kHz with 200 effective points, the resolution bandwidth would be 20 000/200 = 100 Hz and the sweep time would have to be two seconds. Digital storage is a solution to the flicker problem when the unknown spectrum is stable (not noise-like) and when the total sweep time (time to acquire the spectrum) is not objectionable.

Multiplex spectrum analyzers, as opposed to swept-frequency spectrum analyzers, evaluate all the spectral points simultaneously. The probability of intercept is 100%, i.e., a signal that pops up momentarily will not be lost because the instrument is off scanning a different part of the band. Moreover, at low frequencies, parallel processors vastly increase the measurement speed. DSP chips calculate fast Fourier transforms in low-frequency spectrum analyzers and can be built into swept spectrum analyzers to extend the low-frequency range. FFT processors are even included as a feature in many oscilloscopes.

30.4 Impedance measurements

The simplest instrument for impedance measurement was the so-called *Q meter*, shown in Figure 30.3, an instrument for testing inductors. It provided the values of both the reactive part and the resistive part of the impedance. The circuit, shown in Figure 30.3, is very simple, and can be set up quickly in the lab in lieu

Figure 30.3. Q meter.

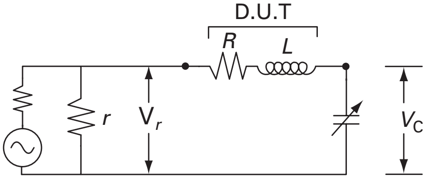

of a suitable instrument. A signal generator provides a fairly constant sine wave voltage V_r across r, a low-value resistor. The variable capacitor is tuned to maximize V_C, the voltage across the capacitor. This maximum occurs when $X_C = X_L$ (series resonance). The operator can read the capacitance dial and, knowing the generator frequency, calculate the unknown inductance. At series resonance, the current through the device-under-test (D.U.T.) is just V_r/r so $V_c = X_C I = (X_C/r)V_r = Q_{\mathrm{DUT}}V_r$ and, therefore, $Q = V_C/V_r$.

The operation of the Q meter assumes that the Q of the variable capacitor is very much higher than the Q of the inductor to be tested. This is almost always true in RF work, and for this reason, there was no similar instrument to measure capacitors.

30.4.1 Impedance bridges

At low frequencies (where lumped components are practical) impedance can be measured directly with a straightforward bridge such as shown in Figure 30.4.

Figure 30.4. Impedance bridge.

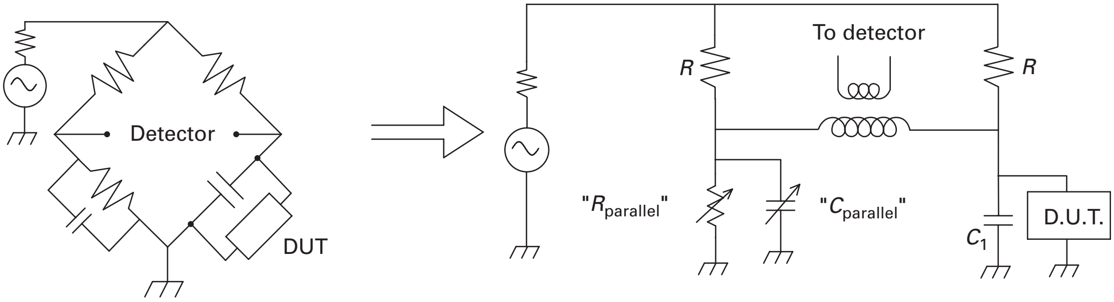

If the device under test (D.U.T.) is purely resistive, the variable capacitor, C, will be set to the same value as the fixed capacitor C_1. It is convenient to use a variable capacitor with a midpoint capacitance equal to C_1. The front panel “$\mathrm{C}_{\text{PARALLEL}}$” dial is offset so that this point is marked zero. If the load has an inductive component, the bridge can be balanced by setting C to less than C_1. The dial then indicates a negative capacitance which the operator interprets as a parallel inductance. The value of the variable resistor is indicated on its dial as “R_{PARALLEL}.” This kind of bridge was made famous by the Boonton Electronics Corporation’s “RX meter.” The RX meter used an amplified detector, actually a superheterodyne circuit.

Amateur radio operators use a version of this bridge, the “noise bridge,” in which the signal generator is replaced with a wideband noise generator, usually a Zener diode followed by a wideband amplifier, and the null detector is replaced with a tuned receiver, e.g., a communications receiver.

30.4.2 Reflectometers

A directional coupler, e.g., a hybrid, allows the measurement of the reflection coefficient magnitude for one-port devices and of gain, as well as input and output reflections of two-port devices. In Figure 30.5, any power reflected by the D.U.T. is divided between the generator port, where it is absorbed by the matched generator and the cross-port, where it can be measured.

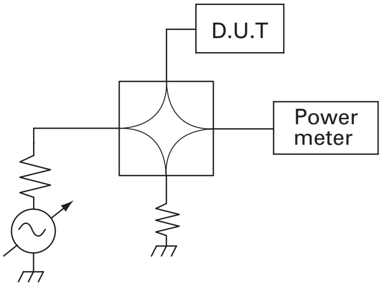

Figure 30.5. Reflectometer.

To calibrate the instrument, the D.U.T. is temporarily replaced by a perfect reflector (usually a short) and the power meter gain is adjusted for a reading of unity (0 dB). This kind of circuit, when used with a swept-frequency RF source (sweep generator) and power display, makes a *scalar network analyzer*, as shown in Figure 30.6, where a switch is provided for two-port devices, such as amplifiers, to measure either reflected power or transmitted power.

Figure 30.6. Scalar network analyzer.

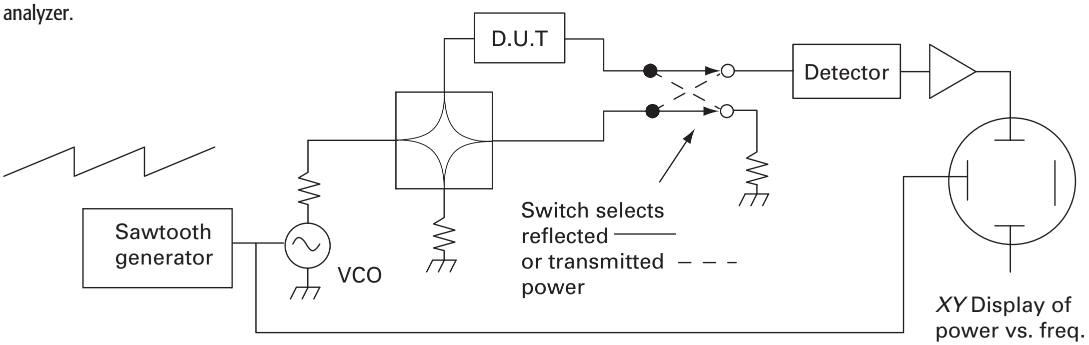

A *vector reflectometer* measures the phase of the reflection coefficient as well as its magnitude. This can be done by combining a vector voltmeter with the above circuit. But it is easy to build a circuit that produces a polar display, i.e., a Smith chart on a CRT screen. Such a circuit is shown in Figure 30.7.

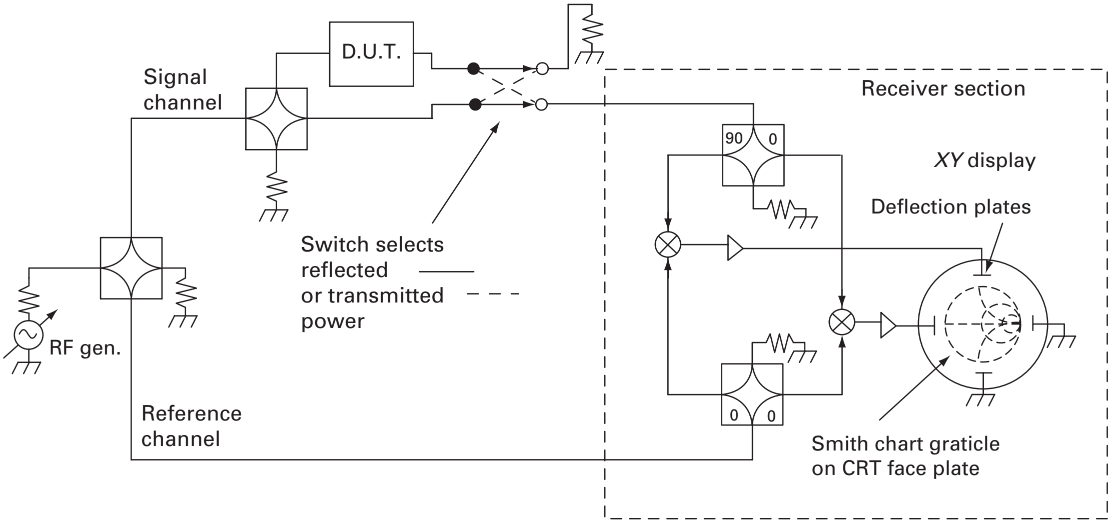
D.U.T.
Signal channel
Receiver section
90
0
XY display
Deflection plates
Switch selects reflected or transmitted power
RF gen.
0
0
Reference channel
Smith chart graticle on CRT face plate

Figure 30.7. Vector reflectometer.

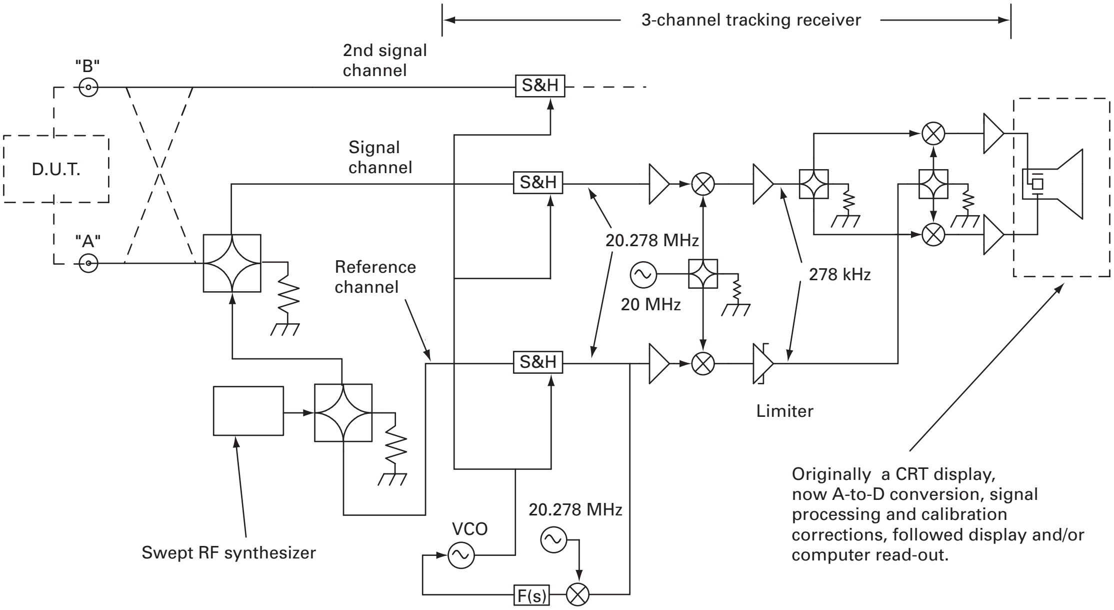
3-channel tracking receiver
"B"
2nd signal channel
S&H
D.U.T.
Signal channel
S&H
"A"
Reference channel
20.278 MHz
20 MHz
278 kHz
S&H
Limiter
Originally a CRT display, now A-to-D conversion, signal processing and calibration corrections, followed display and/or computer read-out.
Swept RF synthesizer
VCO
20.278 MHz
F(s)

Figure 30.8. Vector network analyzer with a three-channel tracking receiver.

A swept-frequency vector reflectometer is called a *vector network analyzer* (VNA) and is the ultimate RF test instrument. It uses a superheterodyne receiver section. Figure 30.8 shows how one commercial network analyzer uses two RF conversions before the final conversion to baseband. This design uses sample-and-hold mixers with the same phase lock L.O. circuit used in the vector voltmeter (Figure 30.1b). Other VNAs use straightforward fundamental mixing. Only one complete signal channel is shown, but usually two are provided, together with a dual-trace display. Often a reversing switch is included to interchange ports “A” and “B” so that all four *S* parameters of a two-port device can be measured without having to turn the device around. All newer vector network analyzers digitize the signals after the final conversion and apply calibration corrections before presenting the data on the display, sending it to a computer, or storing it for subsequent retrieval.

30.4.3 VNA calibration

When using a VNA, coaxial cables are normally used to connect the device under test to the analyzer. Reflection and transmission coefficients will therefore be rotated on the display screen, the Γ plane, by an amount proportional to the length of the connecting cables. Moreover, these coefficients will appear too small if the cables have nonnegligible loss. Unless told otherwise, the analyzer knows nothing of these cables, and will assume the D.U.T. is right up against its front panel. Knowing the propagation constant, loss factor, and length of the cables, the operator could calculate the necessary corrections. But modern analyzers have built-in data processors to make these corrections automatically, once the operator has gone through a calibration procedure. At its simplest, this procedure simply lets the analyzer determine electrical length and loss of the cables.

30.4.3.1 One-port calibration procedures

If only reflection coefficients are to be measured, the analyzer could prompt the operator with a message such as “Substitute a short circuit for the device under test. Then press the CAL button and wait for the CALIBRATION COMPLETED prompt.” When the operator presses the CAL button, the analyzer steps through the selected set of frequencies and reads the phase and amplitude of the reflected or transmitted signal at each frequency. It stores these numbers and uses them to correct each data point when the operator replaces the short with the D.U.T. Most network analyzers provide this calibration as an option. However, this simple procedure is often not adequate. Consider the situation in which the D.U.T. does not have a 50-ohm port. An adapter must be used to make the transition from the end of the 50-ohm cable to the device port. Examples of adapters are coax-to-waveguide transitions and microprobes, used to probe RF integrated circuit chips. Adapters are seldom perfect; they nearly always introduce a reflection of their own. In addition, an adapter will introduce a phase shift, since its length is not zero. The 50-ohm test cable may not be perfect; it might have worn or otherwise imperfect connectors or it may have been kinked, leaving it with an internal point where some power is reflected. The instrument can correct errors caused by all these defects, as well as by the line loss and delay, if it has first been calibrated with a procedure that uses a sequence of *three* calibration standards rather than just one. Often the three standards are chosen to be a short, an open, and a load. In Figure 30.9, the D.U.T. is a microwave horn antenna and

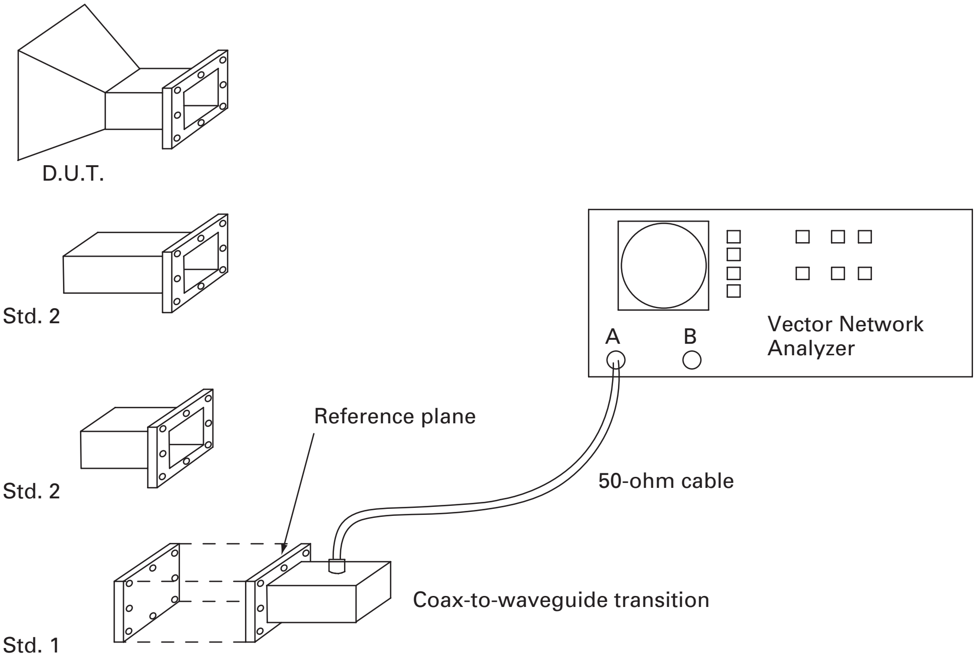

Figure 30.9. VNA setup to measure reflection coefficient vs. frequency of a microwave horn antenna.

the test setup includes a coax-to-waveguide transition. In this example, the three standards comprising the calibration kit are three waveguide shorts of different lengths. Let us look in detail to see why an accurate calibration requires three standards.

Network analyzer calibration for S_{11} *measurements using three standards*

In Figure 30.10, a signal flow diagram, the two-port network represents everything between the analyzer and the device to be tested, usually a cable plus an adapter. Either a standard or the D.U.T. is put at the right-hand side of this hypothetical two-port network. The S parameters of this network will be designated A_{11} , A_{12}, A_{22}, and A_{21} to distinguish them from the S parameters of the D.U.T., which are S_{11}, S_{12}, S_{22}, and S_{21} or just S_{11}, if the D.U.T. is a one-port device.

For the setup of Figure 30.9, for example, the two-port network is the combination of the cable and the coax-to-waveguide transition. Again, neither the cable nor the transition needs to be perfect, nor do they need to be previously characterized. The analyzer outputs a wave of unit amplitude. The wave Γ_{OBS} is the signal reflected back into the analyzer. The external device, either one of the three standards or the device under test, has reflection coefficient Γ_{EXT}. If the external device is the D.U.T., then Γ_{EXT} will be equal to S_{11} if the D.U.T. is a one-port device or if it is a multiport device with the other ports terminated in matched loads. Let us first find Γ_{OBS} as a function of Γ_{EXT}.

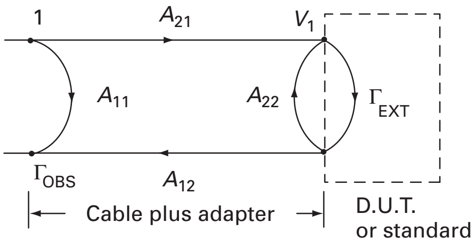

Figure 30.10. Signal flow diagram for reflection measurement setup.

By inspection of the diagram, we see that the forward wave into the external device, V_1, can be expressed as

$$V_1 = A_{21} + V_1\Gamma_{\mathrm{EXT}}A_{22} \quad \text{or} \quad V_1 = \frac{A_{21}}{1 - \Gamma_{\mathrm{EXT}}A_{22}}. \tag{30.1}$$

It follows that

$$\Gamma_{\mathrm{OBS}} = A_{11} + \frac{\Gamma_{\mathrm{EXT}}A_{21}A_{12}}{1 - \Gamma_{\mathrm{EXT}}A_{22}}. \tag{30.2}$$

Solving for Γ_{EXT}, we find

$$\Gamma_{\mathrm{EXT}} = \frac{\Gamma_{\mathrm{OBS}} - A_{11}}{A_{21}A_{12} - A_{22}A_{11} + \Gamma_{\mathrm{OBS}}A_{22}} \quad \text{or} \quad \Gamma_{\mathrm{EXT}} = \frac{\Gamma_{\mathrm{OBS}} - A}{B + \Gamma_{\mathrm{OBS}}C} \tag{30.3}$$

where A, B and C are three constants ($A = A_{11}$, $B = A_{21}A_{12} - A_{22}A_{11}$, and $C = A_{22}$). Measuring Γ_{OBS} for each of the three known values of Γ_{EXT} provided in the CAL kit provides three simultaneous linear equations that can be solved for the values of A, B, and C. This is done automatically by the analyzer when one performs a full S_{11} calibration. Moreover, it is done at each discrete frequency contained between the start and stop frequencies. Once the calibration procedure has been run, the device under test is attached and $\Gamma_{\mathrm{EXT}} = \Gamma_{\mathrm{DUT}}$ is calculated from Γ_{OBS} using Equation (30.3).

In the discussion above, we assumed the two-port network of Figure 30.9 included just the test cable and any adapter. But this network can represent everything between a hypothetically perfect network analyzer and the D.U.T; the three-standard calibration procedure corrects not only for the cable and adapter, but also for deficiencies of the analyzer itself: imperfect directivity and port reflection. You can see that none of these corrections are possible with the scalar network analyzer of Figure 30.6.

Full calibration for two-port measurements requires that we consider the second cable and any adapter as another two-port network. Both networks must be characterized using three standards. In addition, a “through” standard is required. In a coaxial cable set-up, the “through” might be a coaxial male-to-male “bullet.” In the waveguide setup, no actual “through” device is needed; the coax-to-waveguide adapters connected by cables to the analyzer’s “A” and “B” ports are simply placed face to face.

30.5 Noise figure meter

The automatic noise figure meter, Figure 30.11, consists of a tuneable receiver, power meter, and calibrated noise source.

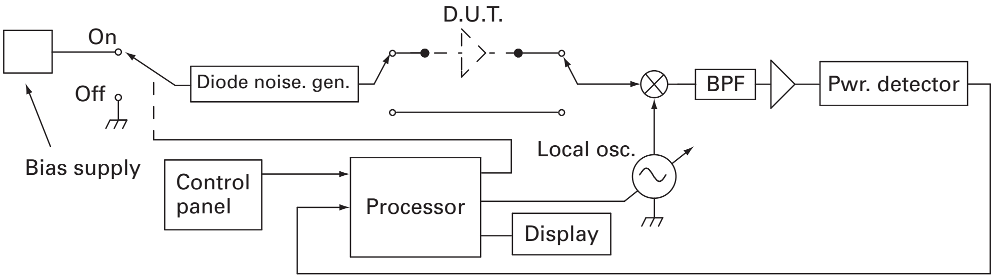

Figure 30.11. Automatic noise figure meter.

In principle, the noise figure of a D.U.T., often an amplifier, can be determined without the use of an external noise source. The noise spectral density at the D.U.T. output can be measured using a power meter, a filter whose bandpass shape is known, and, usually, an auxiliary high-gain amplifier whose gain and noise figure are known. The D.U.T. gain can be measured if a calibrated sine wave source is also available. But when a calibrated noise source is used, the shape of the filter and the gain and noise figure of the auxiliary amplifier need not be known. We define T_{sys} as the equivalent system noise temperature at the input of the D.U.T. when the noise source is turned off and equivalent to a 50-ohm resistor at the ambient temperature, T_0. When turned on, the calibrated noise generator has an equivalent noise temperature T_{hot}. The instrument turns the noise generator on and off and calculates the power ratio $Y = P_{\text{on}}/P_{\text{off}}$. As explained in Chapter 17, the system temperature is given by $T_{\text{sys}} = (T_{\text{hot}} - YT_0)/(Y-1)$. If the D.U.T. has high gain, its equivalent noise temperature, T_{DUT}, is essentially equal to T_{sys}. But if the D.U.T. has low gain, correction must be made for the noise in the rest of the system, i.e., the rest of the noise meter. Note that, by operating the switch that bypasses the D.U.T, the noise meter can measure its own noise temperature and then calculate the noise figure of the D.U.T.

Problems

Problem 30.1. In testing a high-power transmitter with a water-cooled dummy load it is found that the water leaving the dummy load is hotter by 10 degrees Celsius than the water entering the load. If the flow rate is 10 liters/second, what is the output power of the transmitter? (1 calorie = 4.18 joules.)

Problem 30.2. Sketch a digital circuit to provide a numerical phase readout for the vector voltmeter of Figure 30.1. Hint: use a start/stop counter in place of the S/R flipflop. Include provision to count multiple turns, i.e., give the readout a range of 360 degrees × *n*.

Problem 30.3. Design a three-short vector network analyzer calibration kit for WR430 waveguide (width = 4.30 inches). Assume that the midband frequency is 2.38 GHz. Determine the lengths of the three shorts so that, at midband, their reflection coefficients are as different as they can be, in order to reduce systematic errors.

Problem 30.4. After building the calibration kit of Problem 30.3, how would you measure the reflection coefficient of the waveguide-to-coax transition?

References

Hewlett Packard Application Note 1287–1 "Understanding the fundamental principles of vector network analyzers."

Hewlett Packard Application Note 1287–2 "Exploring the architecture of vector network analyzers."

Hewlett Packard Application Note 1287–3 "Applying error corrections to vector network analyzer measurements."

Index

Adler, R. B., 341
admittance, 4
Ampere’s law, 261
amplifier, 19
  ac, 26
  balanced, 185
  bridge,
  cascade of, 221
  characteristic frequencies, 217
  class-A, 23
  class-AB, 24
  class-B, 24
  class-C, 87–92
  class-D, 92–94
  class-E, 94–99
  common-base, 21
  common-collector, 20
  common-emitter, 20
  dc, 26
  dynamic range, 213
  efficiency, 25
  gain, 397
  intermodulation, 212
  limiting, 60
  narrowband, 213
  noise, 219, 336–342
  push–pull, 23
  RF, 29
  single-ended, 23
  small-signal, 208–223
  stability, 211, 397
  totem pole, 24
  wideband, 214
amplitude modulation (AM),
  envelope detector, 56
  in the frequency domain, 57
  in the time domain, 55
  suppressed carrier, 54, 77
  single sideband, 78
  synchronous detector, 232, 309
  transmitter, 62
amplitude modulator
  class-B, 63
  digital-to-analog, 64
  switching, 64
antenna, 259–271
  capture area, 266
  directivity and gain, 264
  pattern, 261
  phased array, 282, 367
  polarization, 271
  tuner, 31, 259
Armstrong, E. H., 69

balun, 164, 170, 176–178
boost converter, 415
boost/buck converter, 414
binary phase-shift keying (BPSK), 300

cascode amplifier, 214
carrier wave, 54
channel capacity, 327
characteristic impedance, 101
circuit analysis
  ac, 3–5
  nonlinear, 5
  program for, 7
  using *S* parameters, 394
coding, 252, 253, 285
Colpitts oscillator, 124
conductance, 4
Coulomb’s law, 261

Dalke, W., 337
delay-Doppler mapping, 282, 370–373
detector, 227
  balanced slope, 235
  delay line, 234
  digital, 232, 238
  diode, 69, 227
  envelope, 56, 69
  Foster–Seeley discriminator, 235
  linear, 228
  phase, 135, 138, 141
  phase lock loop, 233
  product, 80, 231
  quadrature, 234
  slope, 235
  square-law, 238
  synchronous, 232, 309
  tachometer, 233
digital modulation, 55, 300–318
digital signal processing, 73, 75
diode
  as a switch, 49, 291
  free-wheeling, 408
  ideal, 228
  mixer, 49
  PIN, 292
  real, 229
  square-law detector, 238
diplexer, 188, 246
directional coupler, 182, 403
  in a network analyzer, 427
double-sideband, *see* amplitude modulation
duplexer,
  balanced, 290
  branch line, 290
  circulator, 289
  using circular polarization, 288

eddy currents, 172
electromagnetic waves, 1, 259
  propagation of, 260
emitter follower, 20

Fano, R. N., 342
filter,
  allpass, 85
  bandpass, 34, 38, 152–161
  Butterworth, 34
  Chebyshev, 34
  coupled resonator, 35, 41, 152–161
  lowpass, 34, 35, 41
  matched, 321
  transversal, 309
  tubular bandpass, 158
  tuning procedures, 161
flyback converter, 417
FM stereo, 65
forward converter, 418
Foster–Seeley discriminator, 235
frequency
  allocations, 2
  converter,
  hopping, 317
  frequency modulation (FM), 58
    demodulator, 60, 233–238
    deviation, 59
    frequency spectrum, 60
    in the time domain, 59
    modulation index, 59
    modulator, 59, 148, 209
    noise in, 331
    transmitters, 65
    versus AM, 333
  multiplication, 62, 92
  synthesizer, 144–150
    direct, 144
    direct digital, 146–149
    indirect, 146
    mix-and-divide, 145
    phase continuity of, 149
  frequency-shift keying, FSK, 301

GPS system, 311, 349–362
  acquisition, 356
  differential, 360
  ionospheric delay, 359
  satellite transmitter, 352
  signal format, 350
  tracking, 353
  WAAS, 361
group velocity, 359

Haus, H. A., 341
Hertz, H., 1
hybrid coupler , 181–192
  as a power combiner or splitter, 184, 187
  backward coupler, 191
  branch line, 190
  in-phase (180°), 183
  lumped element, 191
  magic T, 203
  quadrature (90°), 185

hybrid coupler (cont.)
  ring, 190
  transformer, 182
  Wilkinson, 189

image rejection, 70
impedance, 4
  bridge, 426
  inverter, 152
  matching, 10–17
  measurement, 425
  specified by reflection coefficient, 107
  waveguide, 201
inductance
  leakage, 166
  magnetizing, 166
interferometry, 366
  imaging, 367
interferometer, 367
intermediate frequency (IF), 46, 69
intermodulation, 212, 293
ionosphere, 273
  day/night, 274
  reflection of waves by, 274
  wave propagation in, 273
intersymbol interference (ISI), 303

Jansky, Karl, 364

local oscillator (L.O.), 46

matched filter, 321
matching
  for best noise figure, 336
  impedance, 10–17
  in waveguide circuits, 202
  networks, 12
  of power amplifiers, 31
  transformer, 11
  using transmission line circuits, 111
Maxwell, 1
  equations, 1, 196, 260
mixer, 46
  active, 49
  diode ring, 49
  diode, 49
  double-balanced, 50
  image-rejection, 72
  passive, 49, 51
  single-sideband, 53
  switching, 48
modulation, 54
  8-VSB, 254, 300, 308
  AM, *see* amplitude modulation
  CDMA, 315, 317
  COFDM, 311
  digital, 55
  FM *see* frequency modulation
  frequency-shift (FSK), 54
  GMSK, 308
  index (FM), 59
  OFDM, 310–315
  on–off keying (OOK), 54, 301, 325
  phase (PM), 54, 58
  pulse amplitude (PAM), 301
  spread-spectrum, 315
  QAM, 55
Moon, 368
MPEG encoding, 252
multiplier
  as a mixer, 47
  frequency, 62
  voltage, 47

network
  analyzer, 428
  L-network, 12
  ladder, 6, 34
  linear two-port, 208
  matching, 12
  minimum phase, 45, 221
  pi-network, 14
  pulse forming, 295
  T-network, 14
noise
  cosmic, 364
  figure, 220
  in oscillators, 128, 149, 336, 342–346
  in amplifiers, 219, 336–342
  matching, 336
  measure, 341
  measurement, 223, 432
  parameters, 222
  temperature, 221
  thermal, 219
Norton equivalent circuit, 10, 338
Nyquist
  pulse shapes, 303
  sampling theorem, 355

oscillator, 120–131
  Armstrong, 123

  Colpitts, 124
  Hartley, 123
  negative resistance, 127, 129
  noise in, 128, 336, 342–346
  numerically controlled (NCO), 355
  phase-shift, 126
  relaxation, 120
  sine wave, 121
  voltage-controlled (VCO), 54
overspreading, 371

pulse amplitude modulation (PAM) 301
phase detector, 135, 138, 141
phase lock loop, 134–144
  acquisition, 142
  as an FM demodulator, 233
  dynamics, 136, 137
  frequency response, 139
  mechanical analogue, 136, 137
  receiver, 143
  stability, 142
  ransient response, 140
phase modulation (PM), 58
phase noise, 149, 331
phase velocity, 104
phased array, 282, 367
polarization, 271
power combiners and splitters, 184, 187
power measurements, 422
power supply, 406–419
  electronically regulated, 409
  inherently regulated, 407
  switching, 411–419
pulse shaping, 303

Q meter, 425
Q-factor, 14, 17
quadrature AM (QAM), 84, 301
quadrature phase-shift keying (QPSK), 302

Raab, F. H., 97
radar, 278–297
  astronomy, 368–373
  cross-section, 283
  cw-Doppler, 279
  equation, 283
  MTI, 281, 343
  planetary imaging, 368–373
  pulse compression, 285
  pulse modulator, 293
  pulse-Doppler, 281
  side-looking, 287
  surveillance (search), 278
  synthetic aperture (SAR), 286
  targets, 283
  TR switches, 288
  tracking, 282
radiation from a current element, 261
radio
  astronomy, 364–368
  communication link, 272
  propagation, 260
  receiver, 67–76
  spectrometry, 375–388
radiometry, 365–366
reactance, 4
Reber, Grote, 364
receiver,
  crystal set, 68
  digital, 73, 75
  direct conversion, 73, 75
  double conversion, 71
  phase lock, 143
  superheterodyne, 69
  television, 249, 255
  TRF, 68
reciprocity, 202, 266, 400
rectifier
  full-wave, 406
  half-wave, 408
  three-phase, 410
reflection coefficient, 106
reflectometer, 427
resonance, series and parallel, 4
resonant charging, 297
Rothe, H., 337

S parameters, 210, 390–403
  circuit analysis, 394
  for cascaded two-ports, 399
  for lossless networks, 400
  for *n*-port devices, 391
  for one-port devices, 390
  stability analysis, 397
scalar network analyzer, 427
Shannon, C., 327
signal flow graphs, 395
single-sideband, 78
  advantages of, 81
  class-C amplification of, 83
  demodulation, 231
  generation, 81

Smith chart, 109
Sokal, N. O. and A.D., 94
spectrometer, 366
  acousto-optical, 384, 385
  autocorrelation, 366
  chirp-z, 386
  filter bank, 376
  Fourier transform, 381
  multiplex, 366, 425
  polyphase, 384, 385
  swept analyzer, 424
  WOLA, 384
spectrum analyzer, 424
standing waves, 105
superheterodyne, 46, 69, 342, 428
suppressed carrier AM, 54, 77
susceptance, 4
switching converter, 411–419
  boost,
  buck, 412
  buck/boost, 414
  Ćuk, 416
  flyback, 417
  forward, 418
  transformer-coupled, 417
synthesizer, *see* frequency synthesizer

television,
  ATSC standard, 251, 301
  color, 246
  digital, 251
  first broadcasts, 243
  NTSC standard, 243
Thévenin equivalent circuit, 10, 338, 424
TEM wave, 115
thyratron, 294
TR switch, *see* duplexer
TRF receiver, 68
transformer, 164–176
  balun, 164, 170, 176–178
  double-tuned, 170, 176–178
  equivalent circuit, 166
  for matching, 11
  ideal, 165
  iron core, 165, 168
  leakage inductance, 166
  magnetizing inductance, 166
  mechanical analogue, 169
  transmission line, 175
transistor
  equivalent circuits, 214
  hybrid-π model, 215
  noise in, 338
transmission line, 101–116
  characteristic impedance, 101
  coaxial, 114
  microstrip, 101, 116
  modification of impedance by, 106
  phase velocity, 104
  vs. waveguide, 201

vacuum tube, 293
vector network analyzer (VNA), 428
vector voltmeter, 424
Venus, 370
video compression, 251
VSWR, 105

waveguide, 195–203
  impedance, 201
  junctions, 202, 203
  loss, 204
  matching, 202
  phase velocity, 198
  rectangular, fundamental mode, 196
  wall currents, 200
  wavelength in, 199
waves, electromagnetic, 1, 259
Weaver, D. K. Jr., 82

Y parameters, 148, 209

Z parameters, 209

advertisement

Need more WE time?

K-Y Ultragel Personal Water Based Lubricant helps couples reconnect

Buy Now on Amazon

FREE ENERGY
FROM THIN AIR!
*INVOLVES CRYSTALS AND DANGEROUS VOLTAGES!
Copyright 2016 Drew Paul Designs. All Rights Reserved.

(https://cdn.instructables.com/FAT/Z4PR/IOKBVYWL/FATZ4PRIOKBVYWL.MEDIUM.jpg)

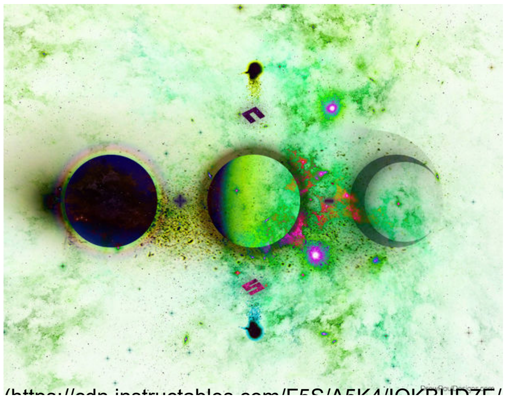
DrewPaulDesigns.com

(https://cdn.instructables.com/F5S/A5K4/IOKBUDZE/

"Dendera Light"
circa 1500 BC
Image: creative commons, Wikipedia

(https://cdn.instructables.com/FOH/L0OE/IOK

advertisement

iRobot

ROOMBA®
THE ROBOT VACUUM

GET YOURS STARTING AT $299

LEARN MORE

About This Instructable

75,101 views

1,071 favorites

License: BY-NC-SA

DrewPaulDesigns

DrewPaulDesigns (/member/DrewPaulDesi

DrewPaulDesigns.com (http://DrewPaulDesigns.com)

(/member/DrewPaulDesigns/) Follow 272

Bio: I am a tech maniac; from media, marketing and design to alternative energy and more. Check out my website for links to all my projects.

More by DrewPaulDesigns:

TRIPLE
YOUR SOLAR ARRAY'S OUTPUT

(/id/TRIPLE-YOUR-SOLAR-ARRAYS-OUTPUT/)

FREE ENERGY
FROM THIN AIR!

(/id/Free-Energy-From-Thin-Air/)

DIY
ELECTRO-MAGNETIC LEVITATION

(/id/DIY-Electro-Magnetic-Levitation/)

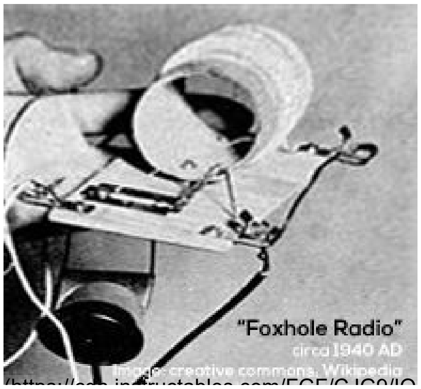

3.74
CEN-TECH

"Free energy from the air?, Yea, right!" Sardonic skepticism was my first reaction to this unusual concept, as well.

Though, its not so far out there, in fact. Light can be converted to DC current with solar panels, electricity can be converted to magnetism as I did in my last article (https://www.instructables.com/id/DIY-Electro-Magnetic-Levitation/), in a microphone sound waves are converted to an electrical signal (by vibrating a magnet near a coil (http://hyperphysics.phy-astr.gsu.edu/hbase/audio/mic.html)), solar rays can even be focused and converted to heat in awesome devices like this! (http://www.gosunstove.com/) When we think about it, energy is all around us and can be harvested in an enumerable many of ways.

Today, we are going to take a rather novel approach. We are going to build a device specifically designed to sense and capture a particular band of energy which is all around us.

The earth is magnetic and anyone who has ever used a compass knows this. Magnetic bodies in motion produce electricity, we can see this in any alternator, like the one in your car. So, therefore the earth is electric as well as magnetic, by definition.

Can we detect this energy? Yes, we sure can! Ever turn on a radio in the middle of nowhere and heard static? That is your radio picking up naturally occurring energy in the RF spectrum!

Can we use this energy to do work? Absolutely! This has been known for a long time. Crystal radios (https://en.wikipedia.org/wiki/Crystal_radio) have been around since before the 1930's and can run with no input energy other than the radio signal. Even when completely isolated, but from the atmosphere, a crystal radio will produce a voltage in the earpiece resulting in a sound (albeit and undesirable one).

Well, this is where it gets interesting...

Can we replicate this effect? Yea, and with modern components like the high quality crystals found in germanium diodes, we can even increase efficiency. By applying this concept as a Crystal Energy Receiver we can take advantage of a wide range of energetic frequencies rather than tuning in to just one.

Can we scale it up? Definitely. Things like micro germanium diodes, high efficiency antennas and compact contemporary capacitors make the components that are required to build a crystal receiver fit in the palm of your hand. While there may or may not be a more efficient way, this renewable energy solution is simple to employ and can be scaled up, or down, indefinitely.

It sounds like we can build a Crystal Energy Receiver. Let's give it a shot...

Step 1: What You'll Need

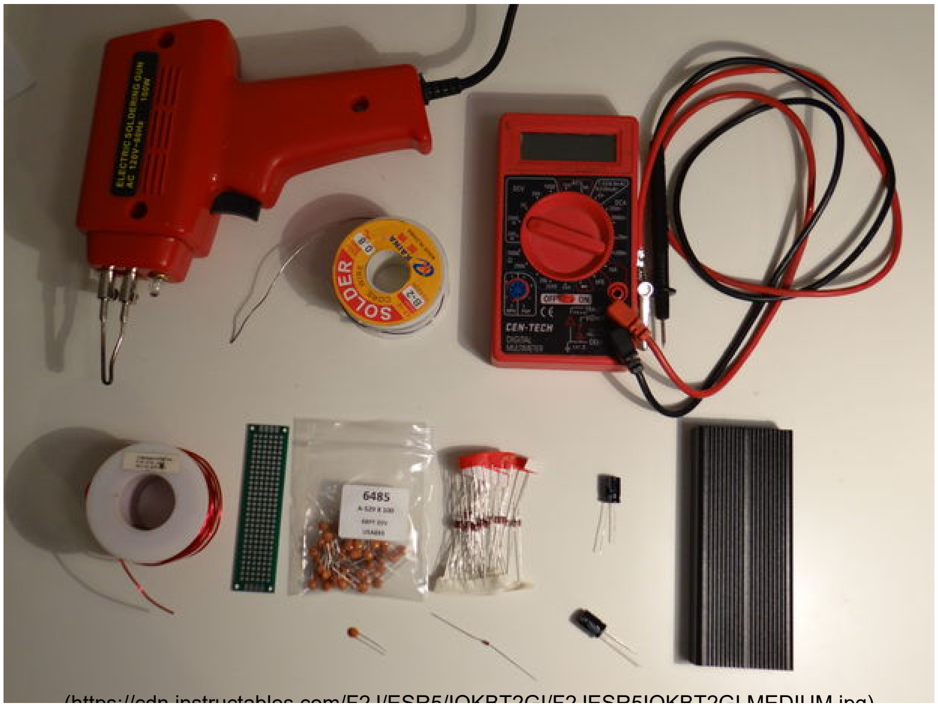

(https://cdn.instructables.com/F2J/ESP5/IOKBT2GI/F2JESP5IOKBT2GI.MEDIUM.jpg)

(https://cdn.instructables.com/F53/EHON/IOKBT2

(https://cdn.instructables.com/FP2/G266/IOKBT2

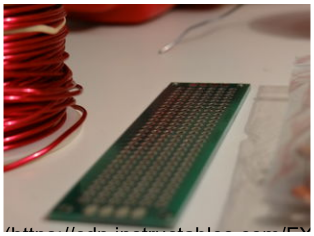

(https://cdn.instructables.com/FX

(https://cdn.instructables.com/FX

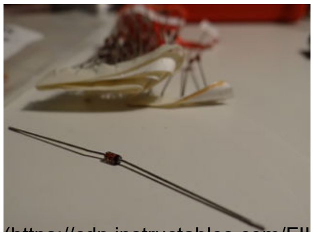

(https://cdn.instructables.com/FII

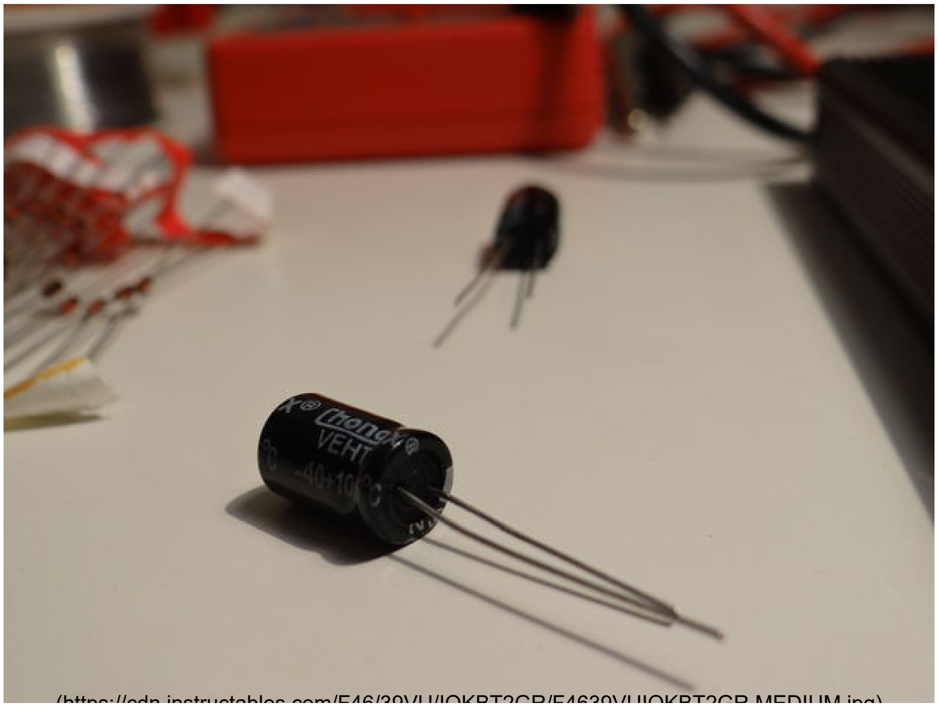

(https://cdn.instructables.com/F46/39VU/IQKBT2GP/F4639VUIQKBT2GP.MEDIUM.jpg)

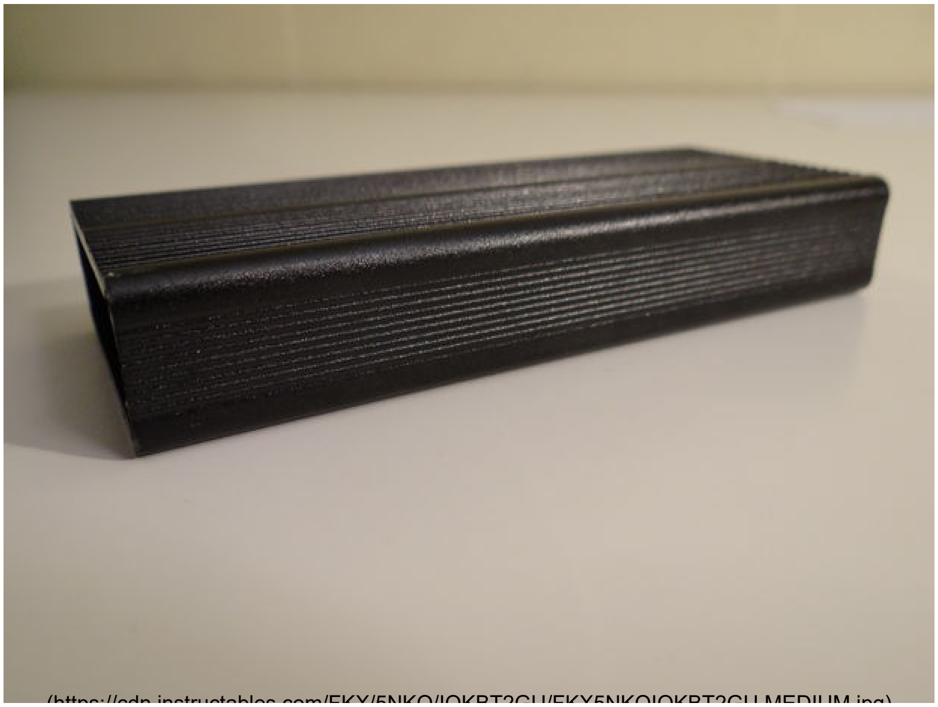

(https://cdn.instructables.com/FKY/5NKO/IQKBT2GU/FKY5NKOIQKBT2GU.MEDIUM.jpg)

One of the reasons this particular renewable energy harvesting method is so viable is the relatively few and easy to obtain materials required.

The simplest crystal receiver design needs no power and can be built with only three parts: a coil, a crystal and a resistor. We're going to optimize that design in order to produce a cleaner and more reliable output signal by first polarizing the input amplitude, then rectifying and filtering the signal. Then we'll add an antenna, case and connections.

Get the circuit diagram here (http://www.drewpauldesigns.com/crystal-energy-receiver-kit.html)
Get the kit here (http://www.drewpauldesigns.com/crystal-energy-receiver-kit.html)

The parts for the circuit include:
(1) Circuit Board (http://www.drewpauldesigns.com/crystal-energy-receiver-kit.html)
(1) 10-18 gauge Copper Wire
(2-12+) Ceramic Capacitors (matched)
(2-6+) Electrolytic Capacitors (matched)
*note various types of capacitors can be used
(4) Germanium Crystal Diodes (1A+)

Total Unit Cost: +/- **$0.40** (USD, scaled for volume of 1,000+ units)

In addition, you'll probably want to get:
(1) Project box (optional)
(1) Antenna (a loop antenna or elevated antenna is recommended and can be made with copper wire)

The tools you'll need are:
Soldering Iron/ Solder (optional)
Multimeter
Oscilloscope (http://www.seeedstudio.com/depot/DSO-Nano-v3-p-1358.html?cPath=63_65) (optional)

That's it. Yup, that's all. Once we've got it all, let's begin.

Step 2: Build the Circuit

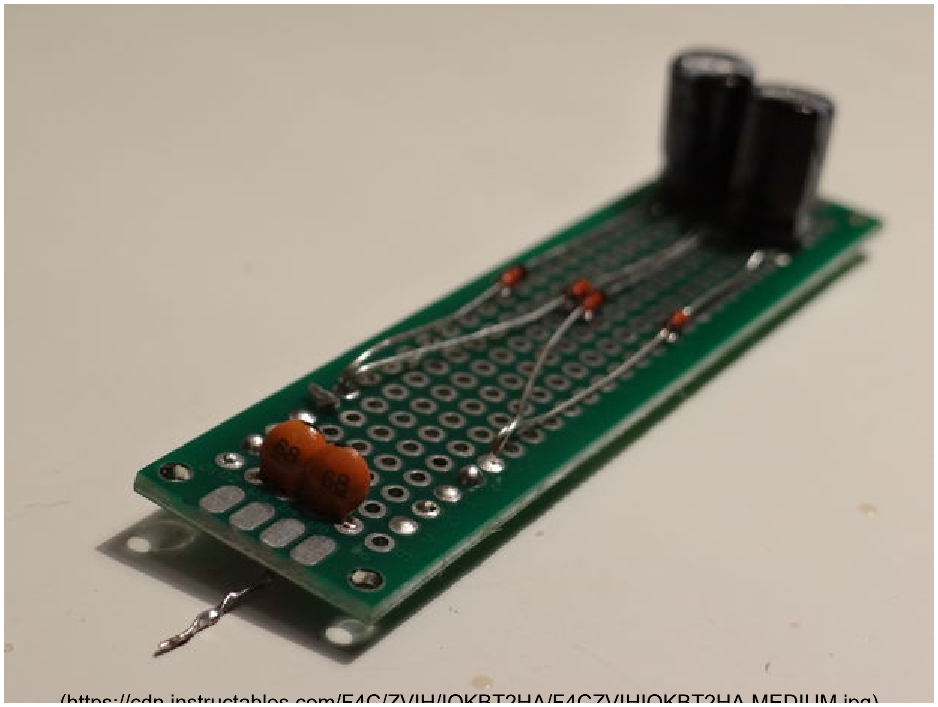

(https://cdn.instructables.com/F4C/ZVIH/IOKBT2HA/F4CZVIHIOKBT2HA.MEDIUM.jpg)

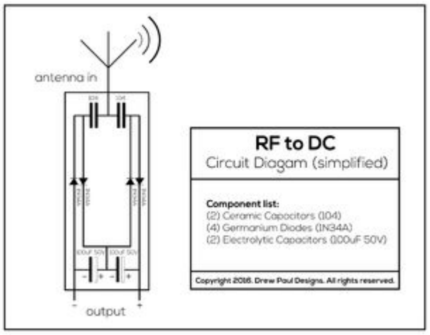
antenna in
104
104
IN34A
IN34A
IN34A
IN34A
100uF 50V
100uF 50V
-
+
-
+
-
output
+
RF to DC
Circuit Diagam (simplified)
Component list:
(2) Ceramic Capacitors (104)
(4) Germanium Diodes (IN34A)
(2) Electrolytic Capacitors (100uF 50V)
Copyright 2016. Drew Paul Designs. All rights reserved.

(https://cdn.instructables.com/FAW/ARS2/IOKBI

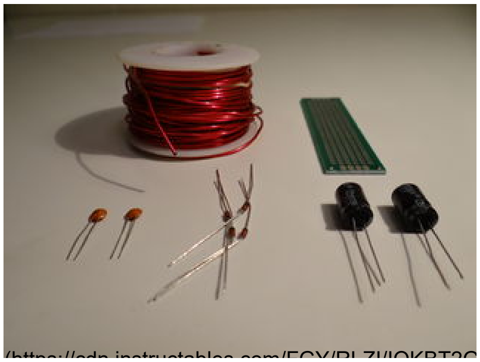

(https://cdn.instructables.com/FCY/BLZI/IOKBT2C

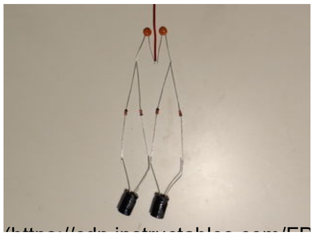

(https://cdn.instructables.com/FE

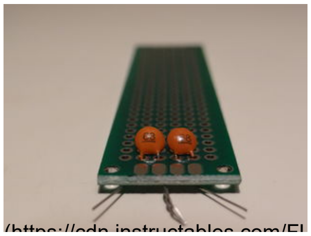

(https://cdn.instructables.com/FI

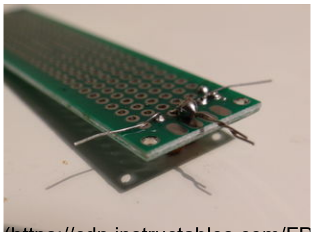

(https://cdn.instructables.com/FP

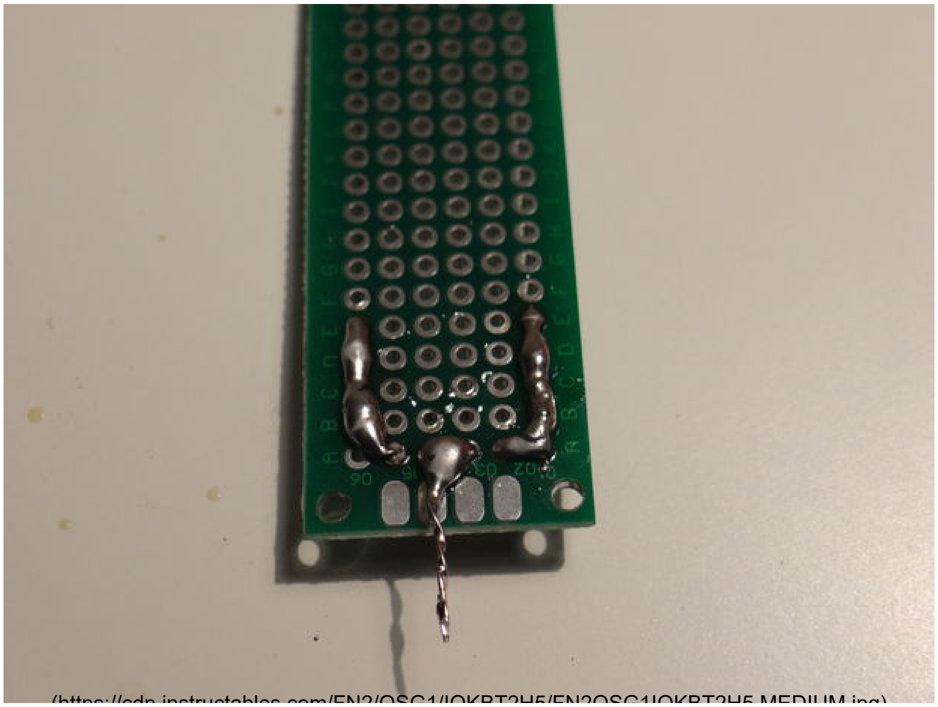

(https://cdn.instructables.com/FN2/QSC1/JOKBT2H5/FN2QSC1JOKBT2H5.MEDIUM.jpg)

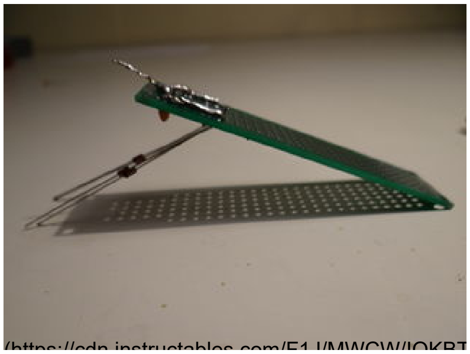

(https://cdn.instructables.com/F1J/MWCW/JOKBT

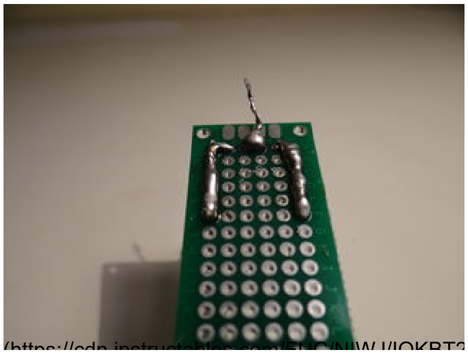

(https://cdn.instructables.com/FUC/NIWJ/JOKBT2

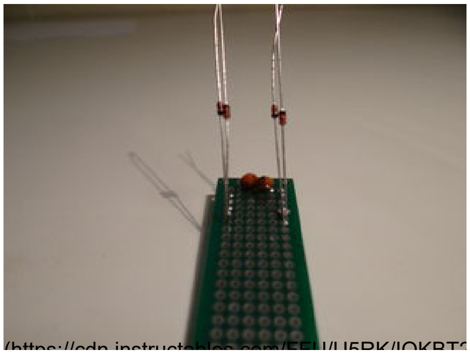

(https://cdn.instructables.com/FFU/U5RK/JOKBT2

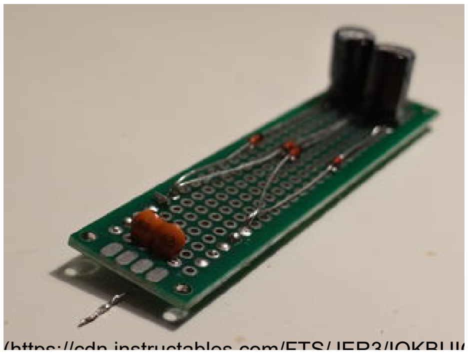

(https://cdn.instructables.com/FTS/JEB3/JOKBUI

We're going to build the simplest version of this circuit variation in order to understand how each component interacts and as a proof of concept.

There are three simple systems at work in the circuit that are composed of capacitors, which store energy, and diodes that direct it.

Energy in the band of radio waves, among others, will vibrate a wire antenna on an atomic level, sending a discernible signal to its lead. This signal will then meet the junction between two ceramic capacitors wired in series. This junction will force positive charge from the wave to travel in one direction and negative charge in the other direction which, when collected again, makes the signal uniform and polar. Connecting the two capacitors in series creates leads on each end; the now positively charged side of one and the now negatively charged side of the other creates a two cell battery.

This next stage of the circuit takes a signal with a net value of zero, adds the absolute values of the positive and negative amplitudes with respect to the origin and produces a positive integer. This concept can be thought of as taking:

(+1) + (-1) + (+1) + (-1) = **0**

and converting it to:

| [(+1) + (+1)] | = **2**
+
| [(-1) + (-1)] | = **2**
= 4

Isn't math fun?

To each of these leads from our two capacitors in series, we will connect two crystal diodes, one facing each direction, to form what is called a bridge rectifier. A bridge rectifier is a configuration which will convert an alternating current to a direct one by cleverly rerouting the signal.

By connecting the bridge rectifier as shown in the circuit diagram, this direct current from the diodes then charges the electrolytic capacitors. This stage normalizes the amplitude, making the current constant and usable.

Components can easily be twisted together for testing and then soldered to a circuit board to secure.

Step 3: Test and Optimize Your Circuit

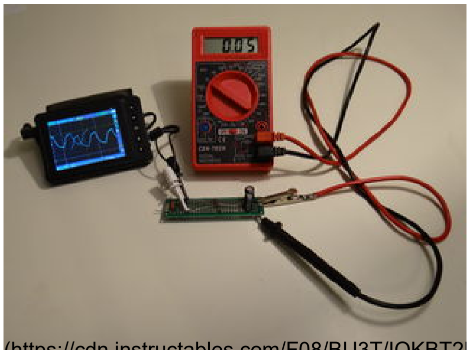

(https://cdn.instructables.com/F08/BU3T/IQKBT2

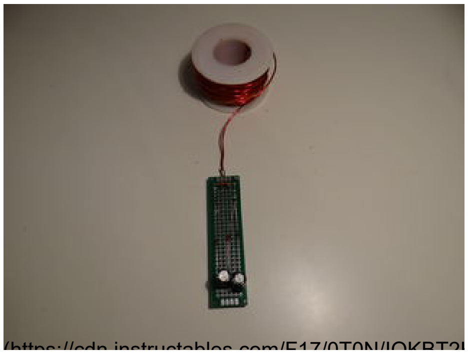

(https://cdn.instructables.com/F17/0T0N/IQKBT2

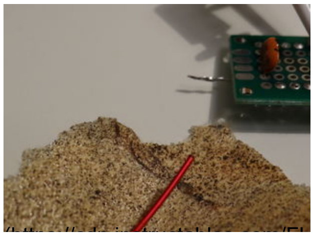

(https://cdn.instructables.com/FK

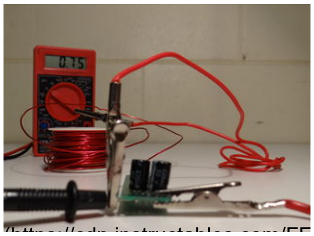

(https://cdn.instructables.com/FF

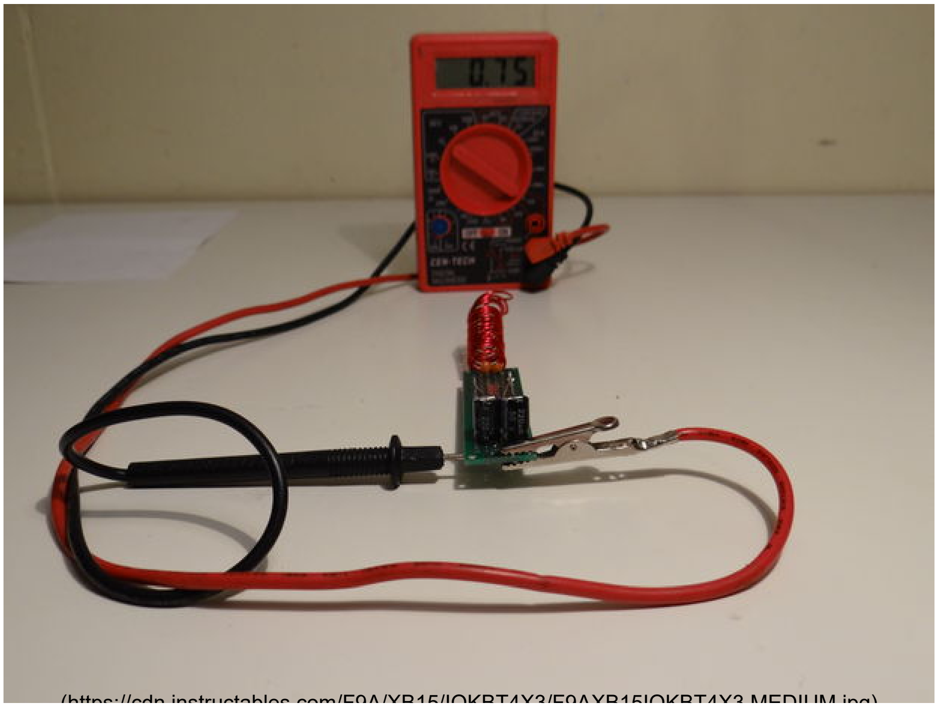

(https://cdn.instructables.com/F9A/YB15/IQKBT4X3/F9AYB15IQKBT4X3.MEDIUM.jpg)

(https://cdn.instructables.com/FOD/30OB/IQKBT

(https://cdn.instructables.com/F6T/0GE2/IQKBT4

To test and analyze our circuit, we'll be using a digital voltmeter and oscilloscope.

By connecting a voltmeter to the output, we'll immediately begin to see a small voltage climbing in the 10-100mV range. If not, we'll want to check our connections and make sure the circuit is not isolated from the environment by taking it outside to a clear area.

Then, by connecting an oscilloscope to the outside leads of our two ceramic capacitor bank, we will see the the polarized signal being captured from the air around us. We can then connect after the diodes to see our varying direct current and then to after the electrolytic capacitors to see a normalized, usable direct current at our output.

We can then optimize the input resistance in two ways. Firstly, we can add additional ceramic capacitors in parallel to our original two and make sure our soldered connections are consistent and thick in this area.

We can optimize the circuit's capacity by adding electrolytic capacitors in parallel to our original two which will allow this circuit to charge slightly when not in use. For this purpose, a charging circuit can also be added here in order to incorporate an optional battery bank.

We can optimize the antenna by attaching loops and coils of copper wire in various positions, store-bought antennas or by stringing some wire up to the highest point you can reach.

We don't have to stop there, either. We can now connect multiple circuits in series to increase voltage or in parallel to increase current. This can be done indefinitely.

Step 4: Add a Case and Antenna

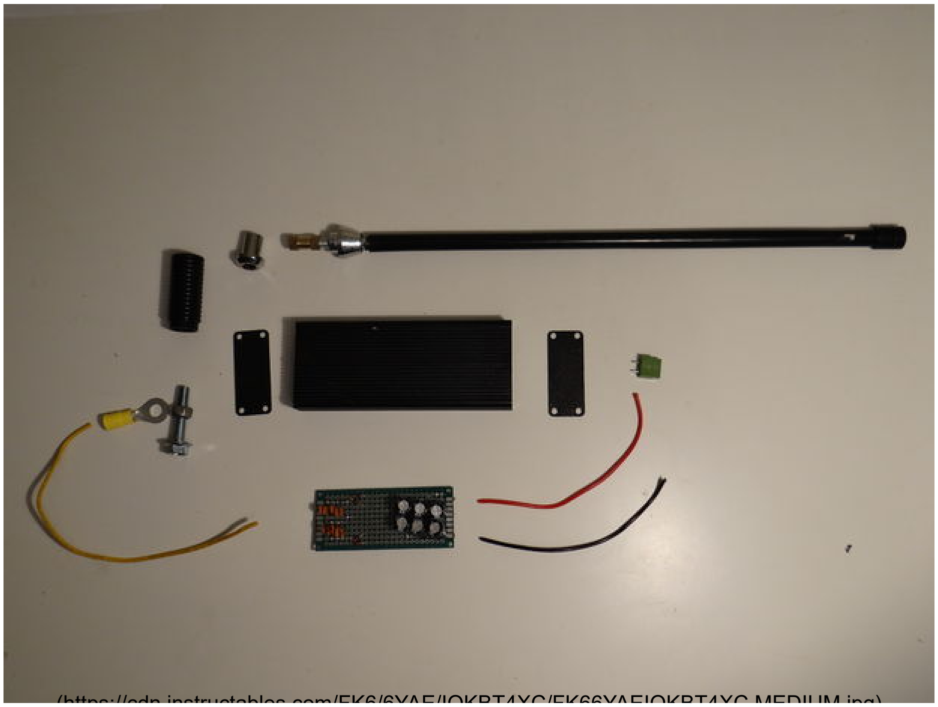

(https://cdn.instructables.com/FK6/6YAF/IOKRT4XC/FK66YAFIOKRT4XC.MEDIUM.jpg)

(https://cdn.instructables.com/F62/QRPB/IOKRT4

(https://cdn.instructables.com/F9JW9R4/IOKRT4

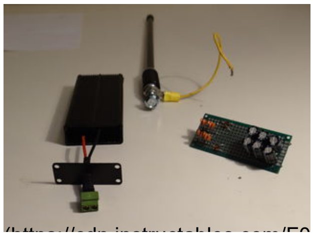
(https://cdn.instructables.com/F9

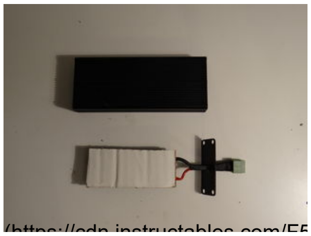
(https://cdn.instructables.com/F5

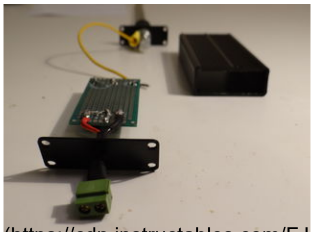
(https://cdn.instructables.com/F

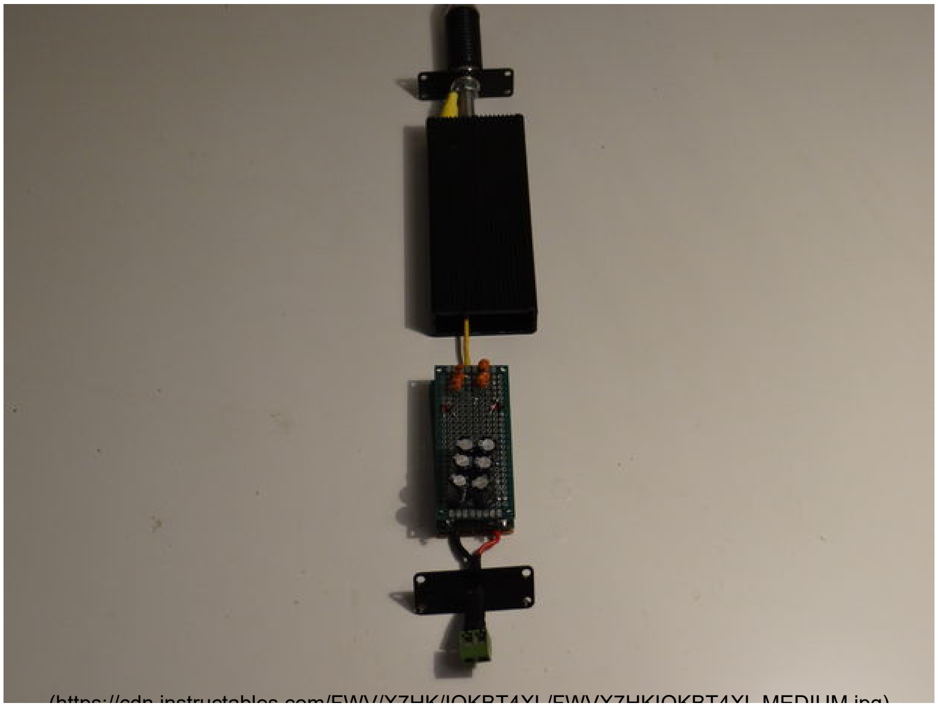
(https://cdn.instructables.com/FWV/XZHK/IOKRT4XL/FWVXZHKIOKRT4XL.MEDIUM.jpg)

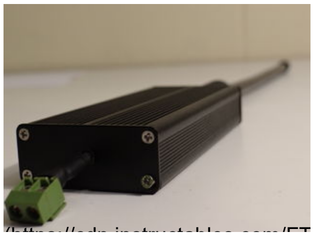
(https://cdn.instructables.com/FT

(https://cdn.instructables.com/FN

(https://cdn.instructables.com/FR

After choosing an antenna in the last step we'll now want to permanently wire it. Whether you choose a compact antenna for portability or a tall fixed antenna for power and range, we will wire it in the same manner according to the diagram in the previous step. Note that the input on the configuration here is grounded to the metallic case, and thus the users hand, and incorporation of longer antennas will require proportionally more substantial grounding.

We will then attach a terminal to the output to allow us to connect this circuit to an electrical device or charging circuit and battery bank.

Next, we will add a case, making sure to isolate exposed leads with non-conductive material especially if mounting in a conductive case. A piece of cardboard secured with glue is sufficient for the circuit's bottom and shrink wrap or electrical tape can be used in the case of any additional exposed leads. Drill two holes in your enclosure, one for the antenna or antenna lead and another for your output terminals. You can then insert your components, fasten the enclosure and your device is ready to use!

Step 5: Your Crystal Energy Receiver Is Complete!

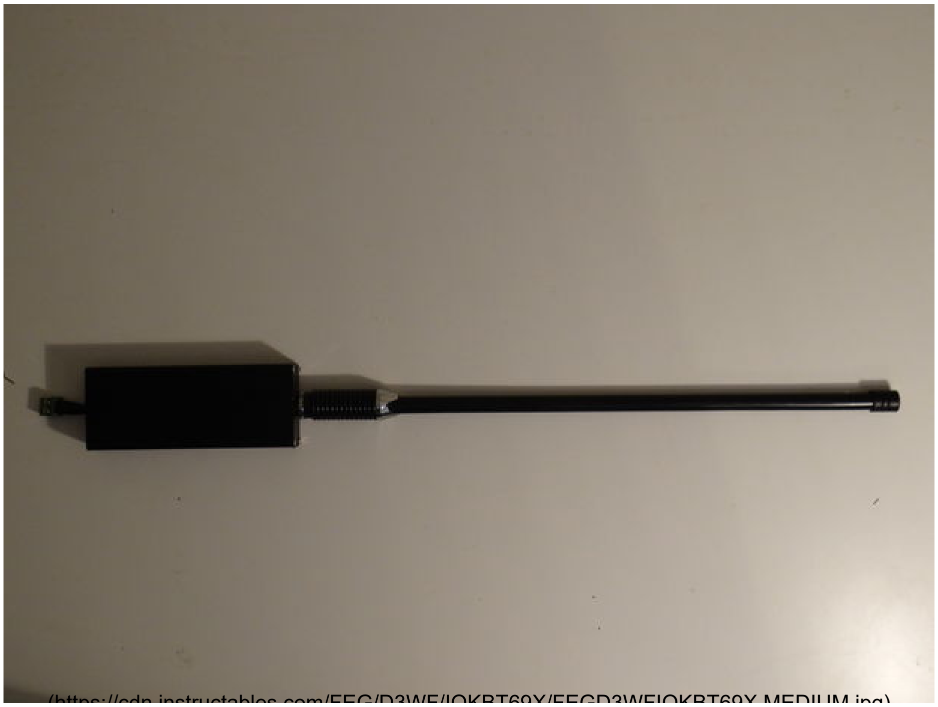

(https://cdn.instructables.com/FEG/D3WE/IOKBT69X/FEGD3WEIOKBT69X.MEDIUM.jpg)

(https://cdn.instructables.com/FY4/8V5I/IOKBT4X

(https://cdn.instructables.com/FLD/DTVI/IOKBT4X

(https://cdn.instructables.com/F1J/6B39/IOKBUZ

(https://cdn.instructables.com/FXB/6LU4/IOKBTC

Your Crystal Energy Receiver is now complete and ready to use!

I built a portable version, for proof of concept and demonstration purposes. However, you can go as big as you want- to passively charge batteries or run equipment remotely; or go as small as you want- to power sensors, RFID devices, small electronics and more.

I used this harvested energy to easily power a low-consumption quartz clock, a digital chronograph with integrated circuits and LCD and was even able to momentarily rotate a small dc motor.

Because of its simplicity this device is a durable, efficient and reasonably effective method of harvesting radiant energy in a simple, replicable and sustainable way. I humbly hope that the contributions made here, and by those reading, can be one day used by people worldwide to conveniently capture free energy.

Thanks for checking out my project and I look forward to seeing everyone's variations, suggestions and improvements!

advertisement

BEST BUY
EXPERT SERVICE. UNBEATABLE PRICE.*
UP TO 25% OFF
SELECT SONY 4K TVs
Upgrade Now
SONY
BRAVIA OLED
See Details

AdChoices
Tackles the Toughest Dried-On Stains!
OXI CLEAN
LEARN MORE

Comments

We have a be nice comment policy.
Please be positive and constructive.

I Made it!
Add Images
Post Comment

TzefaiO (/member/TzefaiO/) 2017-09-30 Reply

Can you ship to Hong Kong when I bought it? How can I contact you?

VinyasiQ (/member/VinyasiQ/) 2017-08-30 Reply

http://is.gd/battchar2

http://is.gd/portablebatterycharger

Dornier (/member/Dornier/) 2017-07-29 Reply

1: what rating capacitors (ceramic)?

2: what rating capacitor (electrolytic)?

3. why both types of capacitors and not one type?

4. You've mentioned grounding to earth, but where is the grounding line?

5. Looking to make this a kids' project to charge a phone so expanding these size isn't a problem at all.

DrewPaulDesigns

DrewPaulDesigns (/member/DrewPaulDesigns/) ▸ Dornier (/member/Dornier/)

2017-07-29 Reply

Thanks for checking out my project. A lot of these questions are covered in the Instructable but I will answer them briefly for you.
The ceramic capacitors used are 104s and their purpose is to polarize the charge. Electrolytic caps are 50v rated and are used to hold the charge in order to provide a usable current. In this example the ground is connected to the case, and thus grounded through the users hand. Scaling this up for that purpose will require a more significant antennae, ground and wiring up multiple circuits in series which is covered in more detail in the article.
Let me know if you have any other questions and be sure to post your project pictures here when you're done for everyone to see!

Loveforcats (/member/Loveforcats/) 2017-07-17 Reply

Hey,how much maximum voltage can be obtained at the output by adiing capacitors in parallel....????

DrewPaulDesigns (/member/DrewPaulDesigns/) ▸ Loveforcats (/member/Loveforcats/) 2017-07-17 Reply

Thanks for checking out the project. Great question. Adding capacitors will only increase the capacity (mAh). If you want to increase voltage you can build multiple circuits in parallel and make sure your antennae is as tall as possible.

Loveforcats (/member/Loveforcats/) ▸ DrewPaulDesigns (/member/DrewPaulDesigns/) 2017-07-26 Reply

I dont want to place my antenna on roof top.I want a min of 5 volt and max of 20 volts and a portable circuit.So what shud i add to the circuit for getting it.??? and what kind of antenna shud i make..??

Loveforcats (/member/Loveforcats/) ▸ Loveforcats (/member/Loveforcats/) Reply 2017-07-26

Can u please send me a video of how your project is working..??

Loveforcats (/member/Loveforcats/) ▸ Loveforcats (/member/Loveforcats/) Reply 2017-07-17

adding*

---

galihprog (/member/galihprog/) 2017-07-14 Reply

sir, why the first kit show value 0.75v and the second kit show value 3.50v,

what is different make? (coil turn, coil component, total component, etc..)

Raghdado (/member/Raghdado/) 2017-07-04 Reply

What a nice instructable! But I have few questions though, if you don't mind.
How long did it take you to complete the project?
How much did it cost you?
Would it be realizable for the duration of one month only (for a beginner)?
And most of all, Could you tell me about the output (after testing)?

Thank you :)

DrewPaulDesigns (/member/DrewPaulDesigns/) ▸ Raghdado (/member/Raghdado/)

2017-07-04 Reply

Thank for checking out my instructable! This project can be done in an afternoon and the kit can be found here: http://www.drewpauldesigns.com/crystal-energy-receiver-kit.html

Before getting started you will also want to get a soldering iron and take some time to practice with it.

Just follow the instructable carefully and you will have no problem.

You can get everything you need here: http://www.drewpauldesigns.com/crystal-energy-receiver-kit.html

Raghdado (/member/Raghdado/) ▸ DrewPaulDesigns (/member/DrewPaulDesigns/)

2017-07-04 Reply

Thank you, I can't wait to start it :)

But before, I just want to know the testing result (voltage, current...)

alaafprojs (/member/alaafprojs/)

2017-01-22 Reply

Cool :-)

DrewPaulDesigns (/member/DrewPaulDesigns/) ▸ alaafprojs (/member/alaafprojs/)

2017-01-22 Reply

So cool, right? Give it a try! Its easy, fun and the results are astounding!

alaafprojs (/member/alaafprojs/) ▸ DrewPaulDesigns (/member/DrewPaulDesigns/)

2017-05-30 Reply

The problem is I don't have any germanium diodes

DrewPaulDesigns (/member/DrewPaulDesigns/) ▸ alaafprojs (/member/alaafprojs/)

2017-05-30 Reply

You can get the kit here:

http://www.drewpauldesigns.com/crystal-energy-receiver-kit.html

alaafprojs (/member/alaafprojs/) ▸ DrewPaulDesigns (/member/DrewPaulDesigns/)

2017-06-01 Reply

Yes, I know.

But I just got some germanium diodes and I have all of the other parts

alaafprojs (/member/alaafprojs/) ▸ alaafprojs (/member/alaafprojs/)

Reply

2017-06-01

I just made it and I tried it with germanium diodes and 1N4001 diodes and the 1N4001 diodes worked better

PS: I have found some other circuits that work better

Raghdado (/member/Raghdado/) ▸ alaafprojs (/member/alaafprojs/) Reply

2017-07-03

Could you please post the image or the link of the other circuits that work better ?

---

frisbrob (/member/frisbrob/) 2017-03-30 Reply

I am always looking for ways to generate power, mostly because it's fun but most of our renewable energies have their down sides. Solar only works while it's day time, wind only works while there is wind. I don't know about where you live but here in Sandy Utah there is no sun at night and at night during winter or summer there is no wind at night, so that takes care of generating power at night. I have looked at several of these pulling power out of the air things, I have even looked at finding Tesla's plans and there is nothing usable. I don't know how to do the math to figure out what parts are needed to build these diy collectors and generators. To me it sounds like you would need an entire back yard to hold anything big enough to do any good, most of the things people build might charge a cell phone but that's about it. If I was to build this unit but big enough to constantly run an 800 watt unit, how big would the antenna need to be and what parts would I need? Just curious.

---

medicinet1 (/member/medicinet1/) 2017-03-26 Reply

Free Energy Generators

medicine turtle cherokee it a great ideal

http://www.hal-lo.at/

PhillipH50 (/member/PhillipH50/) 2017-03-12 Reply

I think what you should look into is tuning this simular curcuit to recieve in the frequency range of 7.83hz to 8.4hz. Because this is the earths magnetical resonance frequency. Tesla's paper state that 1 meter = 100v on a horizontal wire. So at 3 feet tuned into this frequency can produce 100v and at 6 feet produce 200v, ect..
About 10 years ago there was a video that was taken down of a Russian man who was demostrating his setup with a hand full of scientists. I don't remember all the details but one thing I do remember was that his antenna was 68 feet of 14 gauge copper wire hanging 6 feet above the ground horizontally. The ground stake was 1/2" rebar 6' into the ground.

You can look into these frequencies by searching "Schumann resonances are global electromagnetic resonances in the cavity formed between the earth’s surface and the ionosphere. "

Just some of my thoughts..

---

tatebullrider (/member/tatebullrider/) 2017-02-09 Reply

Great instructable!! Just a quick question though. So the higher the antenna is in the atmosphere the better? And if so, you could (in theory) use a long thin wire attached to a kite or a weather balloon to temporarily generate more power, correct?

DrewPaulDesigns (/member/DrewPaulDesigns/) ▸ tatebullrider (/member/tatebullrider/) 2017-02-09 Reply

You got it. You could use a kite or even a drone or quadcopter to take measurements at different altitudes. Post your data for us to see!

16me121lucifer (/member/16me121lucifer/) 2017-01-07 Reply

Sir i am getting 0.06v but negligible current. I am unable to light a LED. I wanted to know how can we connect similar circuits in series or parallel. I would be grateful if you upload a schematic diagram for the same.
Thank you.

16me121lucifer (/member/16me121lucifer/) 2017-01-02 Reply

Dear sir,
Can i use antenna of my old radio set to complete the given project?

DrewPaulDesigns (/member/DrewPaulDesigns/) ▸ 16me121lucifer (/member/16me121lucifer/) 2017-01-02 Reply

Yes, you may. In fact, that is what it was designed for; you'll just using that energy for a different purpose. From there, the taller the antennae you use, the better the results.

Alijo0nn (/member/Alijo0nn/) 2016-12-30 Reply

Dear Paul

Is it possible that this circuit generates 1 ampera current with highest place of Antenna and suitable Earth?

DrewPaulDesigns (/member/DrewPaulDesigns/) ▸ Alijo0nn (/member/Alijo0nn/) 2016-12-30 Reply

A single circuit, no. However, by running these circuits in parallel current can be increased to suit your output needs. Thanks for checking out my Instructable and if you have any other questions feel free to let me know.

16me121lucifer (/member/16me121lucifer/) 2016-12-21 Reply

Sir can we light led with help of this project? & Can we do something so that the received energy is used only when we require? ie. something like switch?

DrewPaulDesigns (/member/DrewPaulDesigns/) ▸ 16me121lucifer (/member/16me121lucifer/) 2016-12-21 Reply

Yes and yes. Give it a try and let us know how it works out.

Alijo0nn (/member/Alijo0nn/) 2016-12-16 Reply

Dear Pual

How can i connect some these circuits together?

is it work, if i connect 10 matched circuits to an antenna?

Best Regards

DrewPaulDesigns (/member/DrewPaulDesigns/) ▸ Alijo0nn (/member/Alijo0nn/)

2016-12-16 Reply

You sure can. Give it a try and post your pictures for us to see.

cmlucht (/member/cmlucht/) 2016-11-23 Reply

what if you hooked up some joule thief's low input higher output?

DrewPaulDesigns (/member/DrewPaulDesigns/) ▸ cmlucht (/member/cmlucht/)

2016-11-24 Reply

Yes, try it.

vvince (/member/vvince/) 2016-11-13 Reply

p.s

good ible all we need now is a nano amp amplifier

vvince (/member/vvince/) 2016-11-13 Reply

seems to me that thomask19 works for some big energy company who have suppressed ALL FREE ENERGy

egads man we cant have the minions getting ANYTHING free

there's no PROFIT in that

SHOE0007 (/member/SHOE0007/) 2016-11-07 Reply

I not trying to be rude but how do you get free energy from nothing? Unless you do this under a high voltage power line (very dangerous) i am not sure if this would work?

In state matter is neither created or destroyed so were is this energy coming from??

DrewPaulDesigns (/member/DrewPaulDesigns/) ▸ SHOE0007 (/member/SHOE0007/)

2016-11-07 Reply

Not at all and this is a very good question.

As Faraday and others have demonstrated, magnetic bodies in motion produce an electric charge. If you've ever used a magnetic compass, you know that the earth, along with orbiting bodies, are magnetic with a north and south pole. These moving bodies induce a charge in our atmosphere. This charge, with our new Crystal Energy Receiver, can be captured and harvested as usable energy.

Thanks for reading. Give it a try and be sure to follow and visit www.DrewPaulDesigns.com for more projects.

Brandonoutdoors (/member/Brandonoutdoors/) 2016-11-04 Reply

i like this "ible". I'm seeing a lot of negative comments though so I want lend my support to the author.
In no way is converting radio waves into electricity stealing. I mean think about it. When you turn on a radio to listen to music nobody bills you for the radio waves. They are paid for by ads.
With a homesteader in mind who may not live anywhere near transmission lines could use somthing like this at a larger scale (our planet is saturated in rf). when radio waves cross your property line they become yours to do with what you will (in my opinion). I say build (or buy) a large antenna and start experimenting with a larger scale.

cloethix (/member/cloethix/) 2016-08-02 Reply

good project

rah187 (/member/rah187/) 2016-07-22 Reply

Cool project! I have favorited this instructable and it is possible that I will eventually construct it. This circuit is actually very simaller to a few others I have seen on the internet, all of which produce around three volts. I have even seen people use crystal radio sets to generate electricity for LEDs. Of course, there are limitations to this setup, including the very low current and the fact that making this unit much larger than it already is may be considered illegal. The truth is, nobody would probably care about one device running off of the RF, seeing as seemingly innoccuous grounded objects such as flagpoles can also "steal" radio waves. However, I am glad I read through the comments before making a gigantic fractal antennae and getting myself in trouble.

İlhamiA1 (/member/%C4%B0lhamiA1/) 2016-07-18 Reply

Lolz its ??? Nano amperes and 0.06 v

DrewPaulDesigns (/member/DrewPaulDesigns/) ▸ İlhamiA1 (/member/%C4%B0lhamiA1/) 2016-07-18 Reply

Eureka! You got it working! You now have proof of concept. Now it's time to increase your current.

First, try extending the antennae all the way to the highest point you can find (try your roof). Then connect a ground lead to the junction between your output terminals (between the two electrolytic caps in series) and run that to a piece of steel rod pressed into the dirt.

Once that is complete, you can optimize it even further! Now, with your antennae and ground set up start wiring in more circuits. Wire them in series, parallel or both according to your needs. Put all of the components on the same circuit board for a supersized version.

Thanks for checking out my 'ible! I hope this helps!

unbottled19208306 (/member/unbottled19208306/) 2016-07-10 Reply

When you say the 2 12 and 6 matched caps, do you mean 12 and 6 volt caps?

DrewPaulDesigns (/member/DrewPaulDesigns/) ▸ unbottled19208306 (/member/unbottled19208306/) 2016-07-11 Reply

The voltage capacity is variable, however it is important to use matching pairs of both ceramic and electrolytic capacitors. Thanks for checking out my project and let us know how yours comes out!

Arghus (/member/Arghus/) 2016-06-23 Reply

can u get it to 5v? for a usb port?

RowanS1 (/member/RowanS1/) ▸ Arghus (/member/Arghus/) 2016-07-01 Reply

yes its possible but current would be too low to be usable

RobertH7 (/member/RobertH7/) 2016-06-13 Reply

Congratulations! You built a full-wave rectifier! You passed the first chapter of basic electronics.

Unfortunately, this particular rectifier is an extraordinary waste of time. It produces a negligible amount of power. Your oscilloscope voltage reading is meaningless. Power is voltage TIMES current, and something like this will produce current in the range of nano-Amperes.

Just get some solar cells and call it a day. They absorb energy from the peak of the sun's blackbody radiation spectrum (which happens to be visible light), not in the negligible radio frequency range. The cheapest, least efficient solar cells out there will produce orders of magnitude more power.

⬇ More Comments

Free Energy From Thin Air! by

DrewPaulDesigns (/member/DrewPaulDesigns/) in science (/technology/science/)

Download | (/id/Free-Energy-From-Thin-Air/) 5 Steps ▸ | + Collection | I Made it! | ♥ Favorite | Share ▾ | ⚑

advertisement

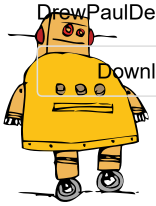

Newsletter

Let your inbox help you discover our best projects, classes, and contests. Instructables will help you learn how to make anything!

enter email | I'm in!

About Us

Who We Are (/about/)

Advertise (/advertise/)

Contact (/about/contact.jsp)

Jobs (/community/Positions-available-at-Instructables/)

Help (/id/how-to-write-a-great-instructable/)

Find Us

Facebook (http://www.facebook.com/instructables)

Youtube (http://www.youtube.com/user/instructablestv)

Twitter (http://www.twitter.com/instructables)

Pinterest (http://www.pinterest.com/instructables)

Google+ (https://plus.google.com/+instructables)

Resources

For Teachers (/teachers/)

Residency Program (/pier9residency)

Gift Premium Account (/account/give?sourcea=footer)

Forums (/community/?categoryGroup=all&category=all)

Answers (/tag/type-question/?sort=RECENT)

Sitemap (/sitemap/)

Terms of Service (http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=21959721) |
Privacy Statement (http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=21292079) |
Legal Notices & Trademarks (http://usa.autodesk.com/legal-notices-trademarks/) | Mobile Site (https://www.instructables.com)
(http://usa.autodesk.com/adsk/servlet/pc/index?id=20781545&siteID=123112)

© 2017 Autodesk, Inc.

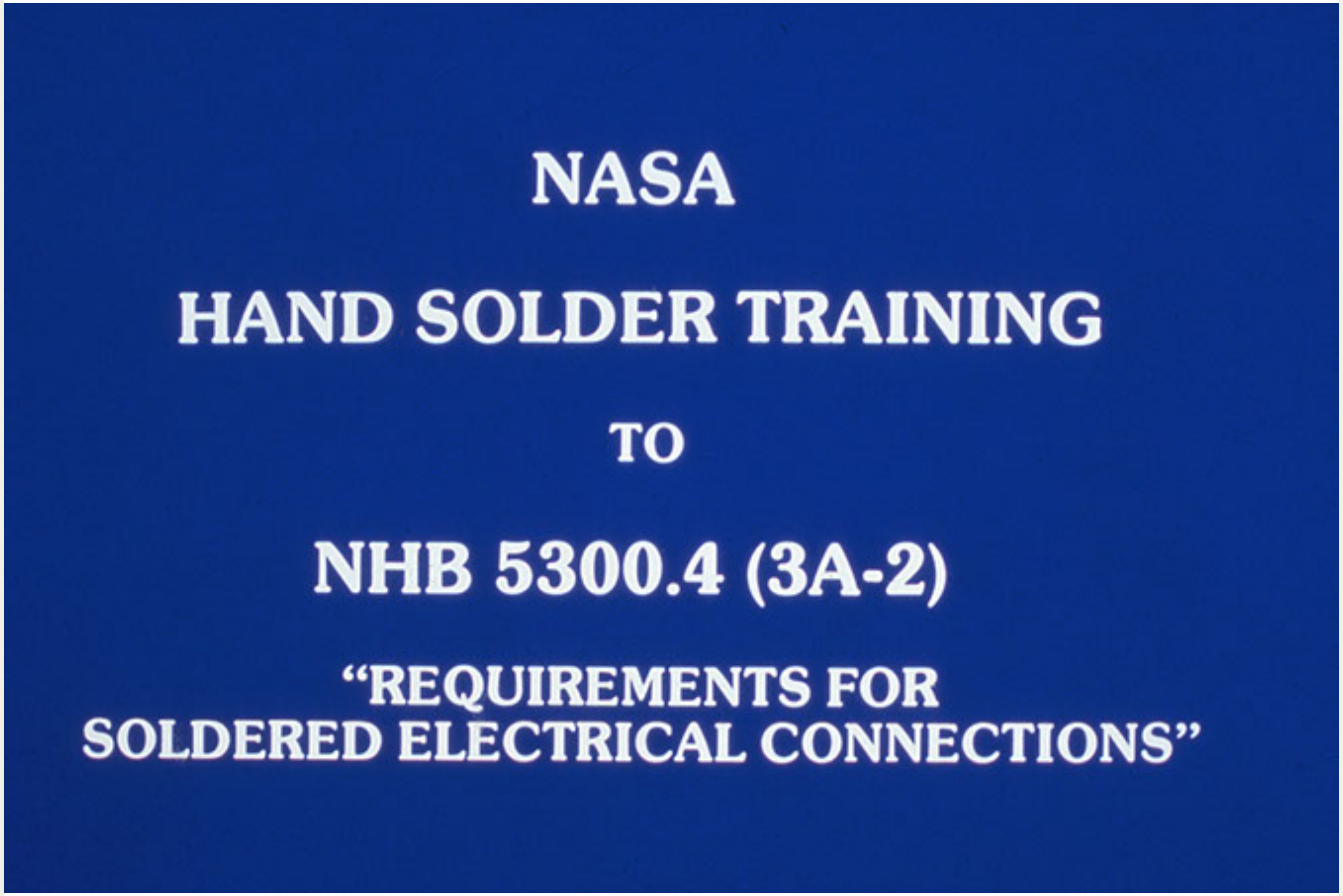

NASA

HAND SOLDER TRAINING

TO

NHB 5300.4 (3A-2)

“REQUIREMENTS FOR
SOLDERED ELECTRICAL CONNECTIONS”

NASA soldering

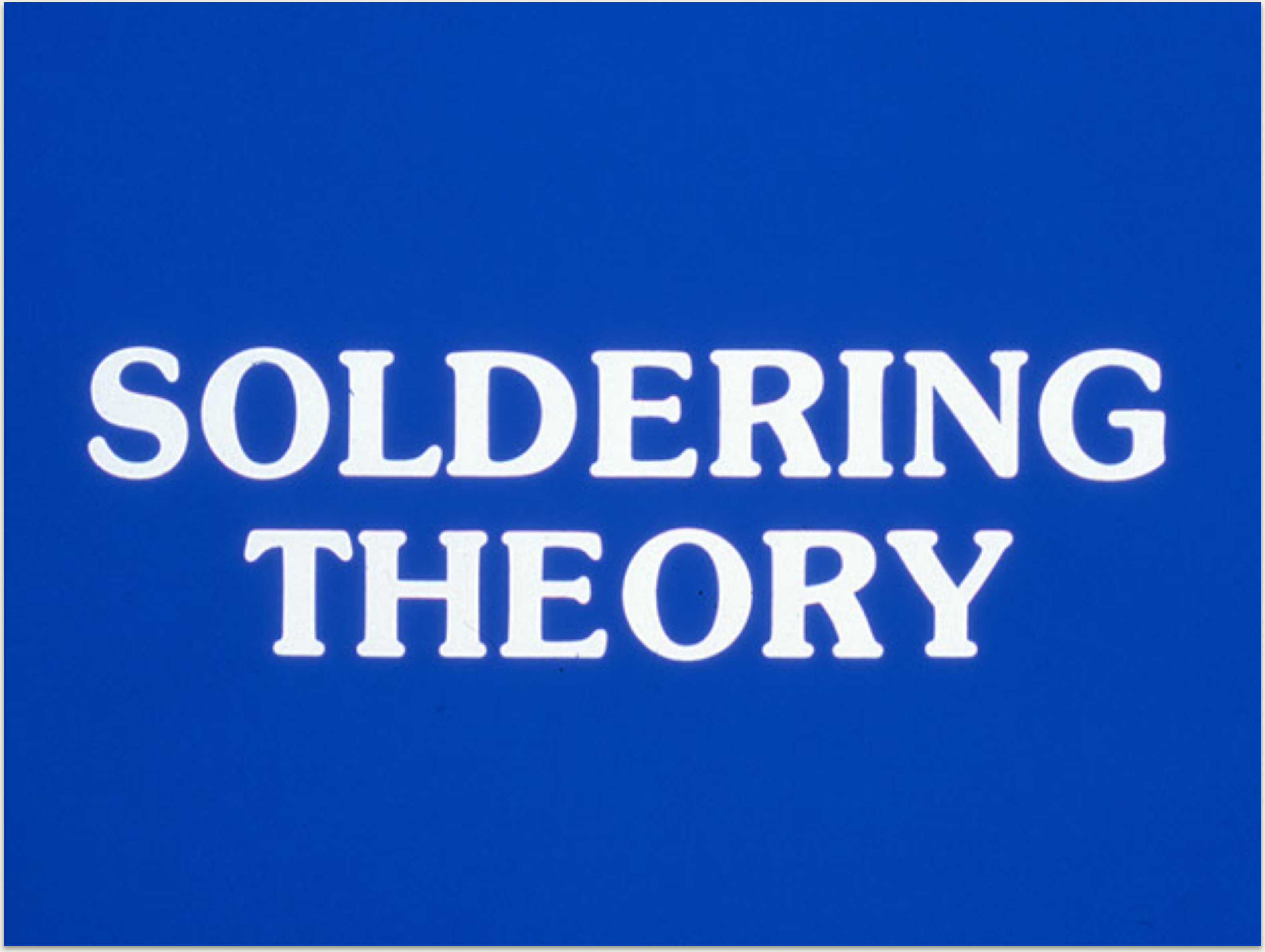

SOLDERING THEORY

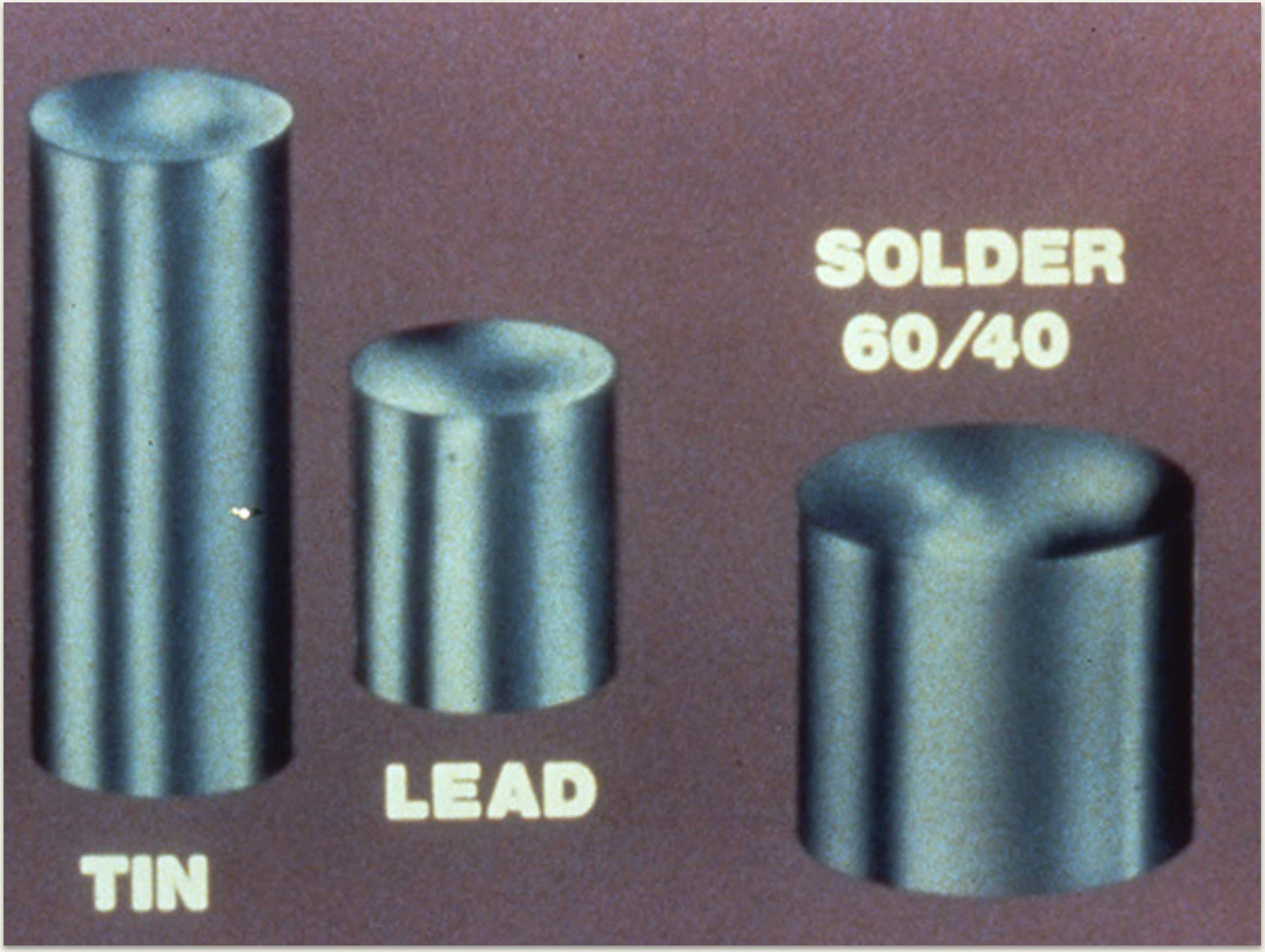
SOLDER
60/40
LEAD
TIN

Tin and Lead = 60/40 solder

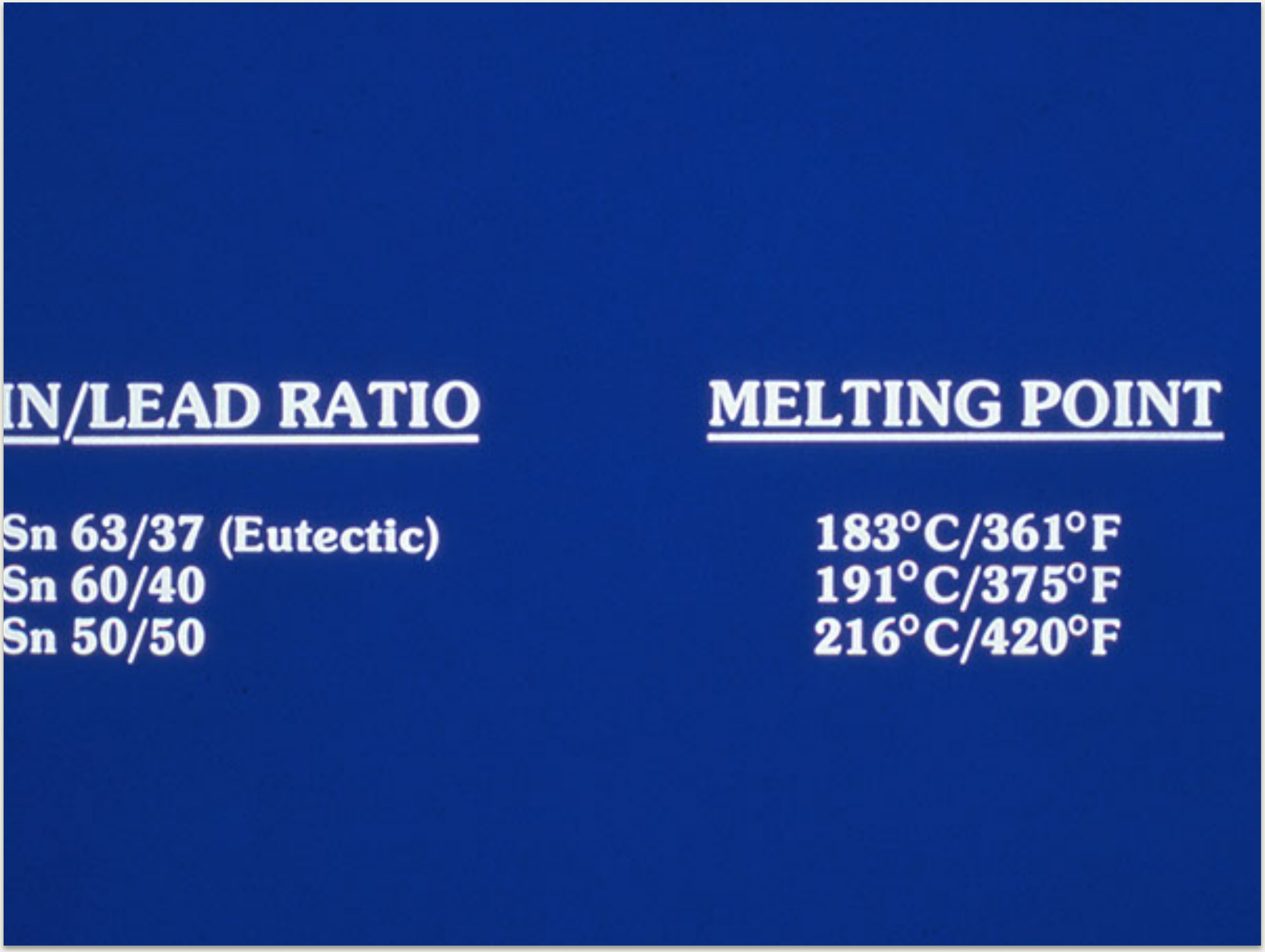

IN/LEAD RATIO	MELTING POINT
Sn 63/37 (Eutectic)	183°C/361°F
Sn 60/40	191°C/375°F
Sn 50/50	216°C/420°F

Tin/Lead ratio/melting point

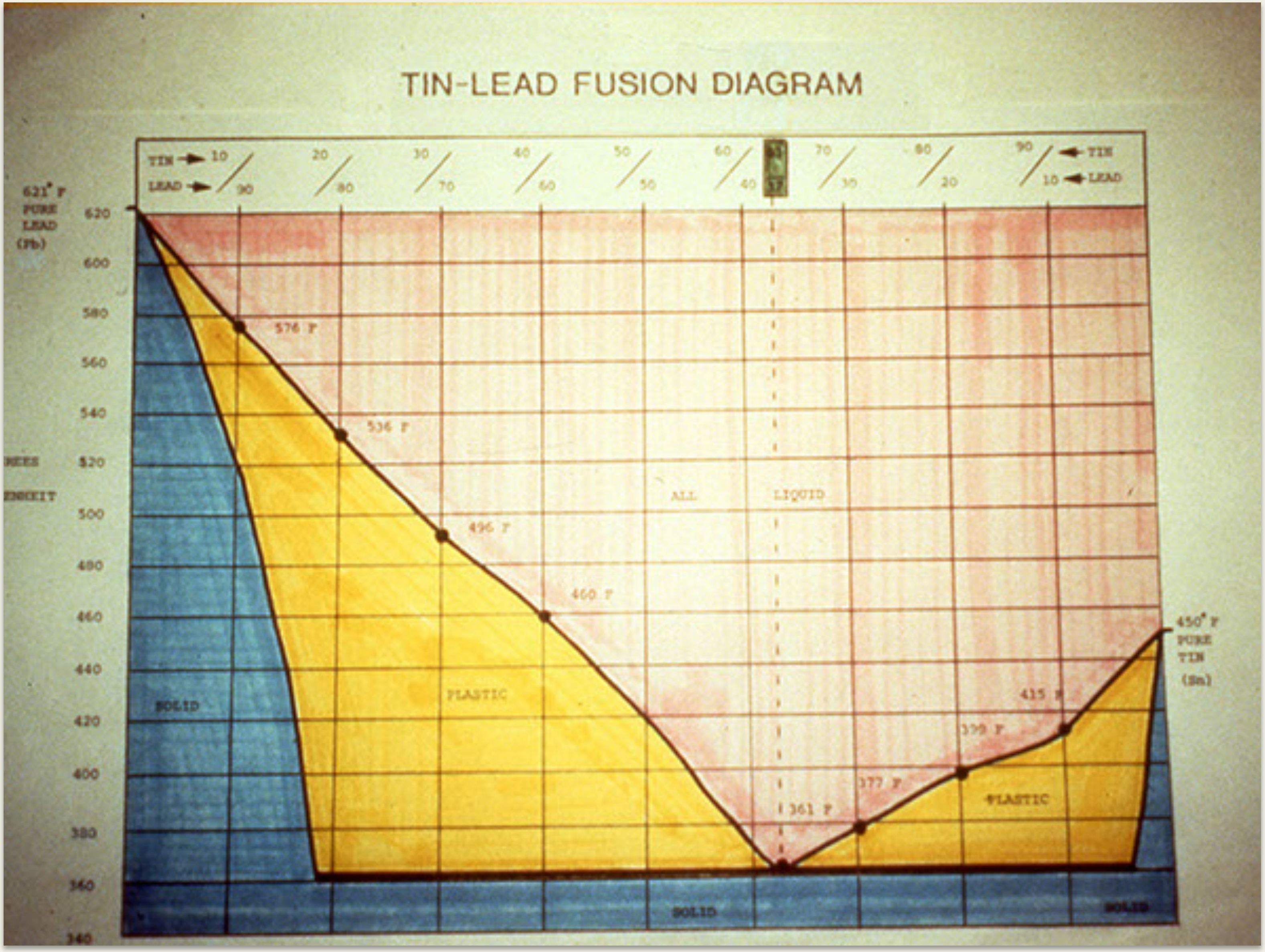

Tin/Lead Fusion

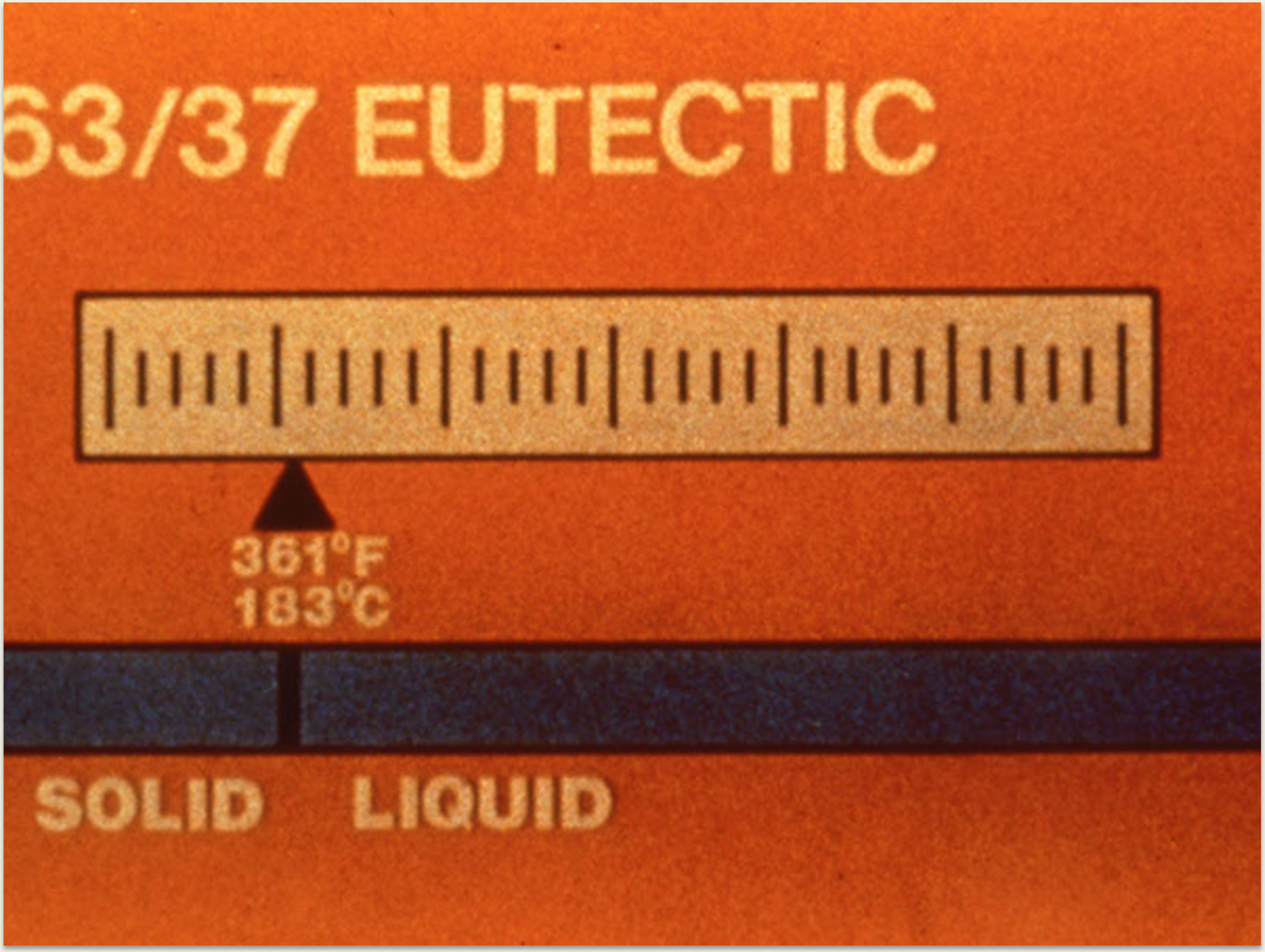
63/37 EUTECTIC
361°F
183°C
SOLID
LIQUID

63/37 Eutectic solder liquid temperature

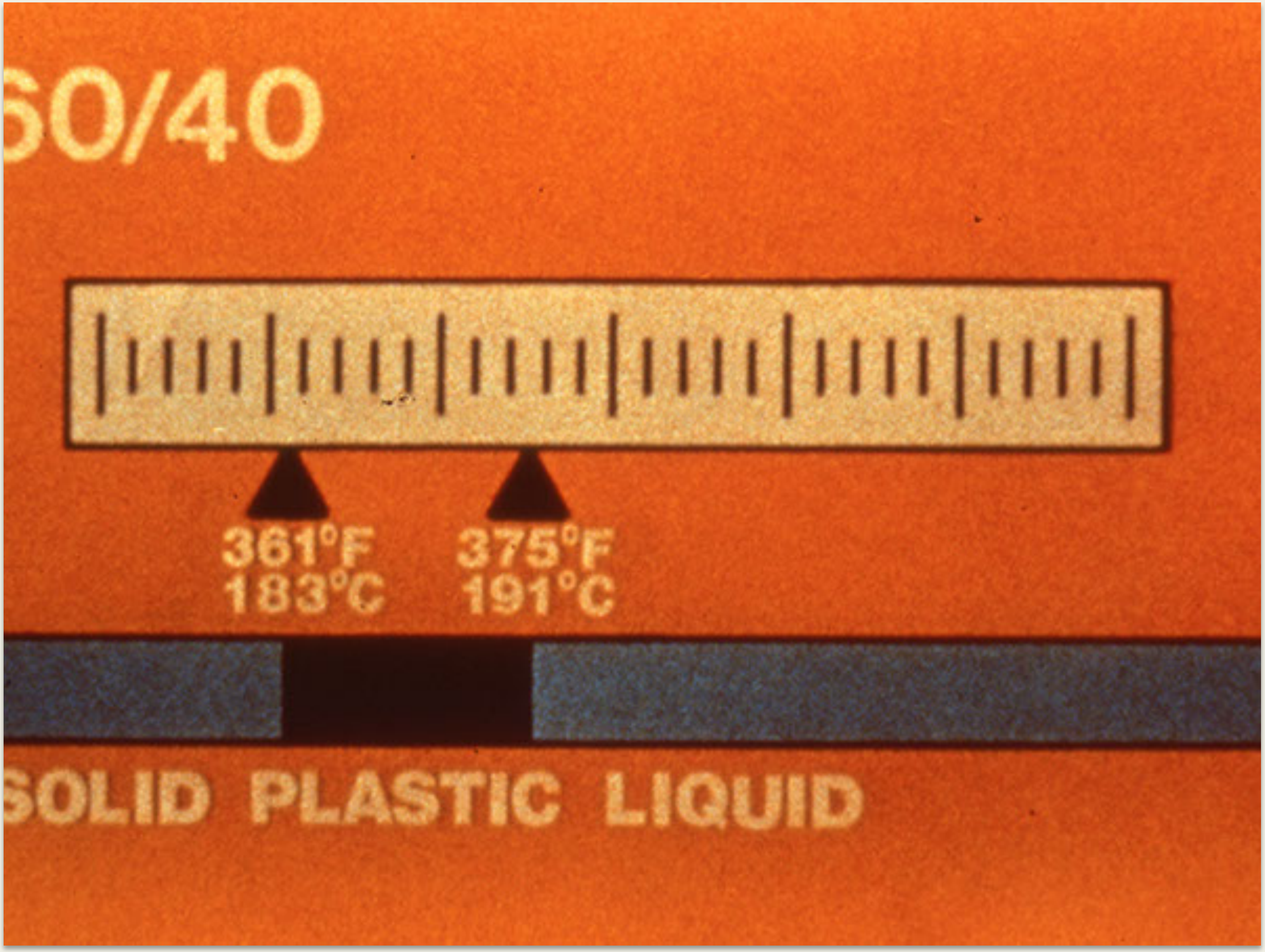
0/40
361°F
183°C
375°F
191°C
OLID
PLASTIC
LIQUID

60/40 solder with solid/plastic/liquid temperatures

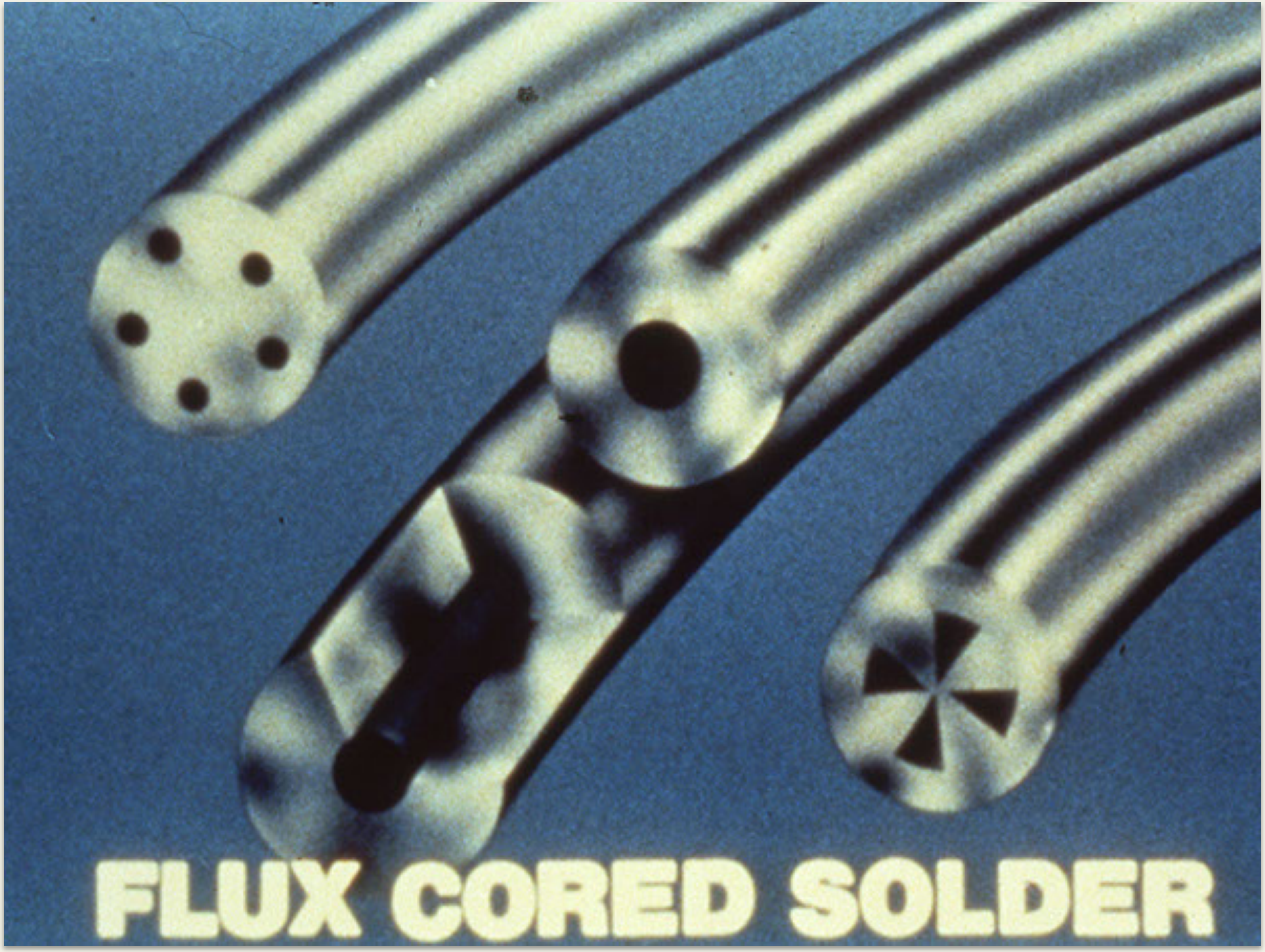
FLUX CORED SOLDER

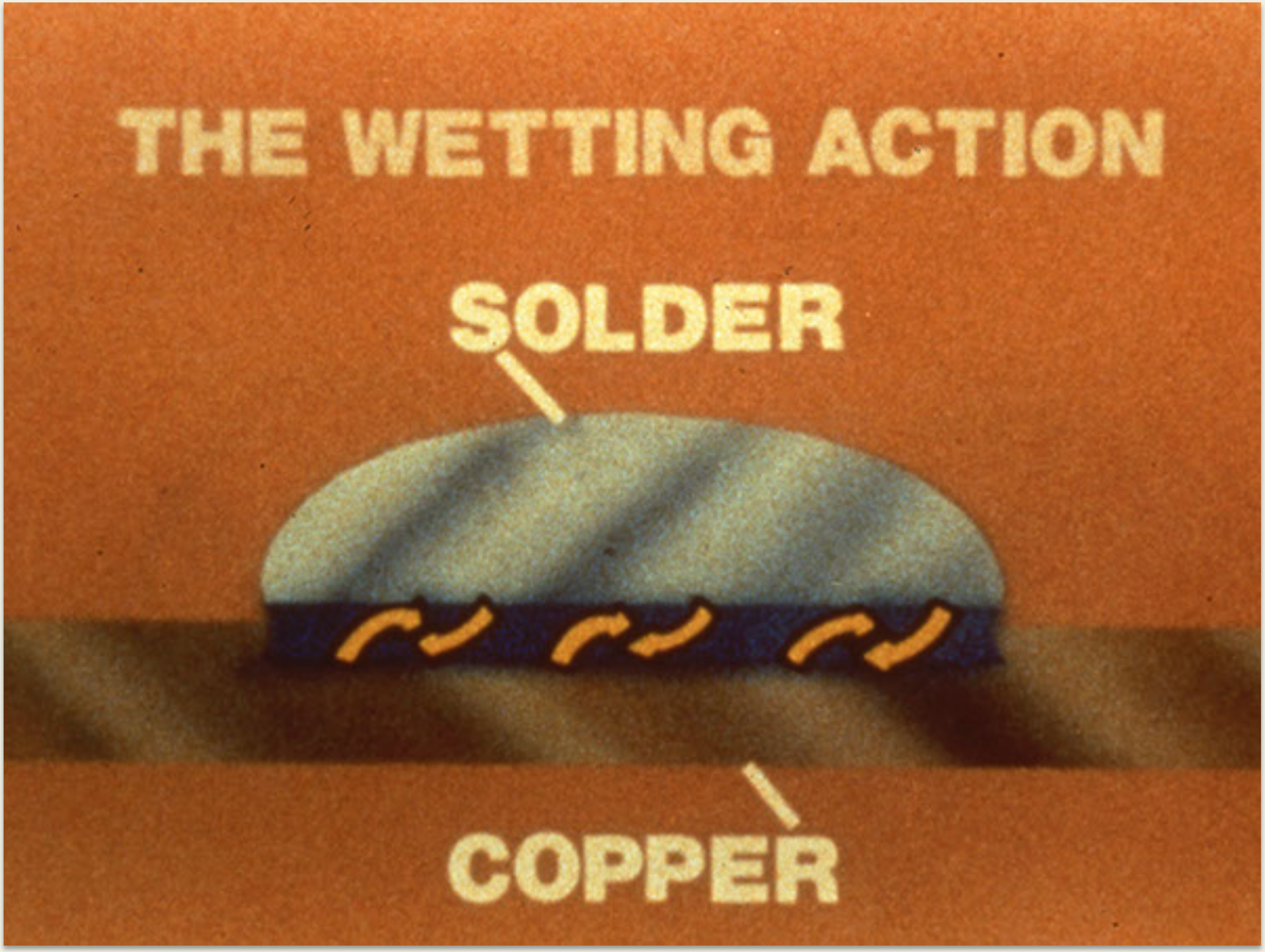
THE WETTING ACTION
SOLDER
COPPER

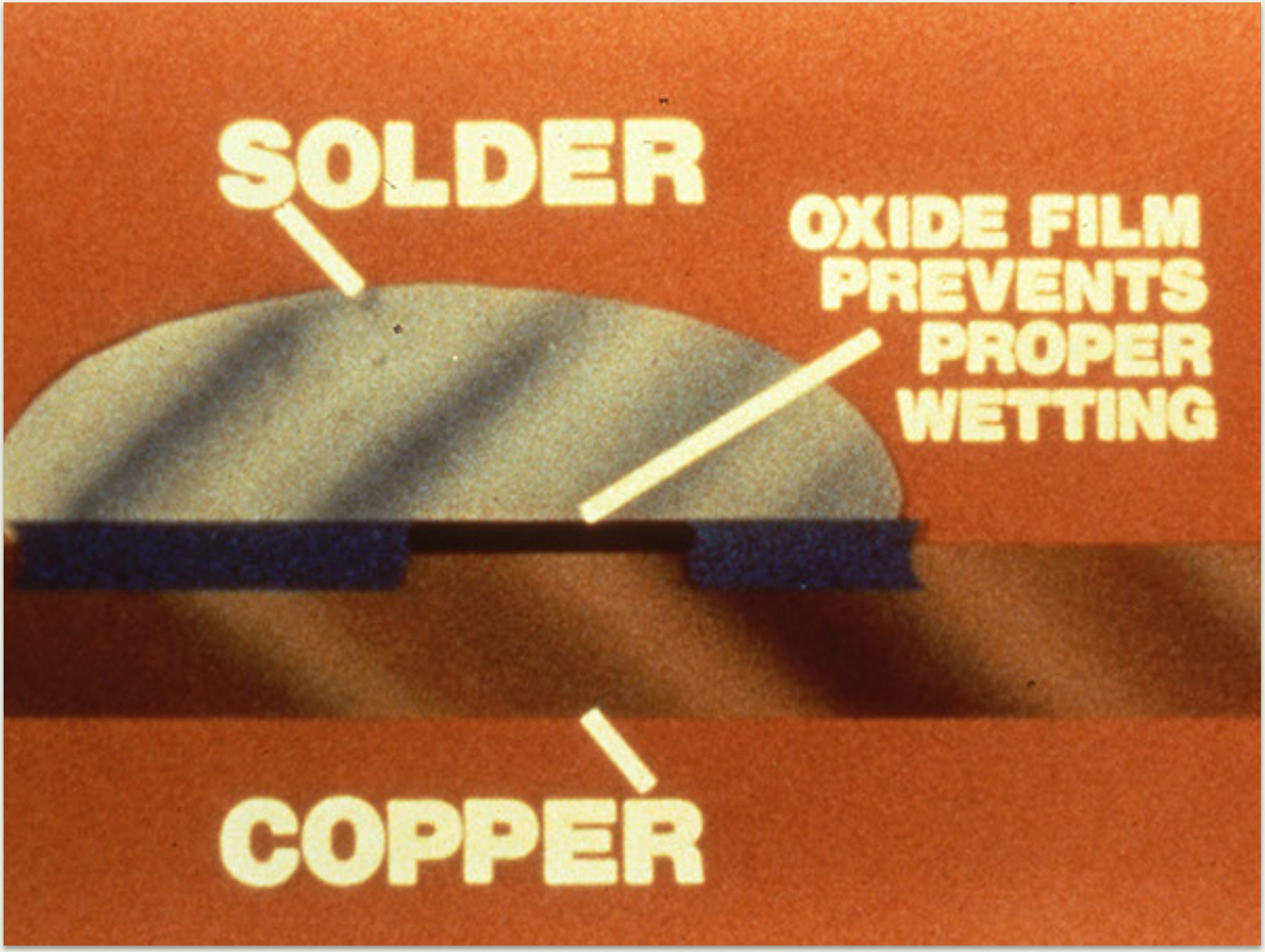
SOLDER
OXIDE FILM PREVENTS PROPER WETTING
COPPER

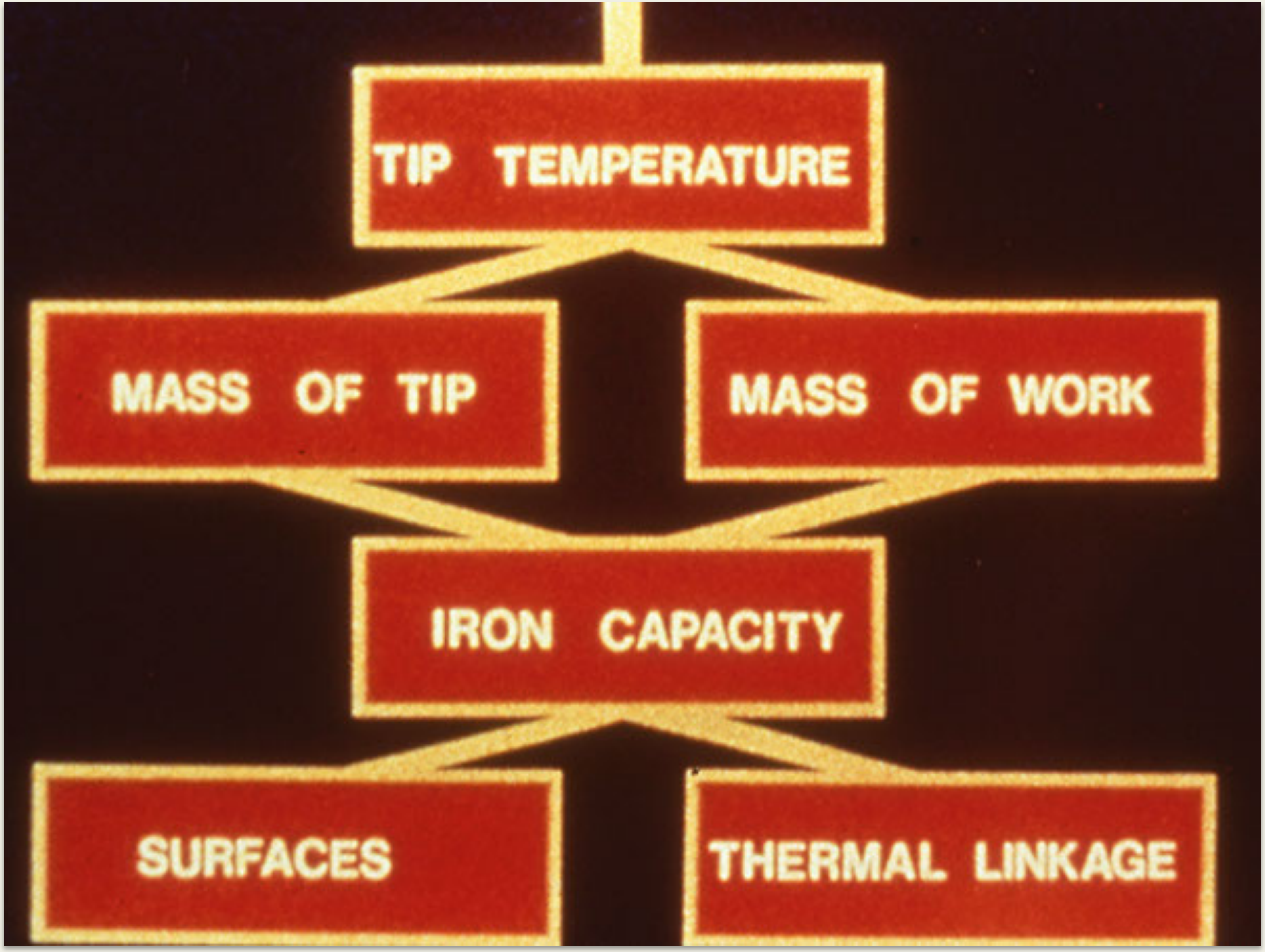
TIP TEMPERATURE
MASS OF TIP
MASS OF WORK
IRON CAPACITY
SURFACES
THERMAL LINKAGE

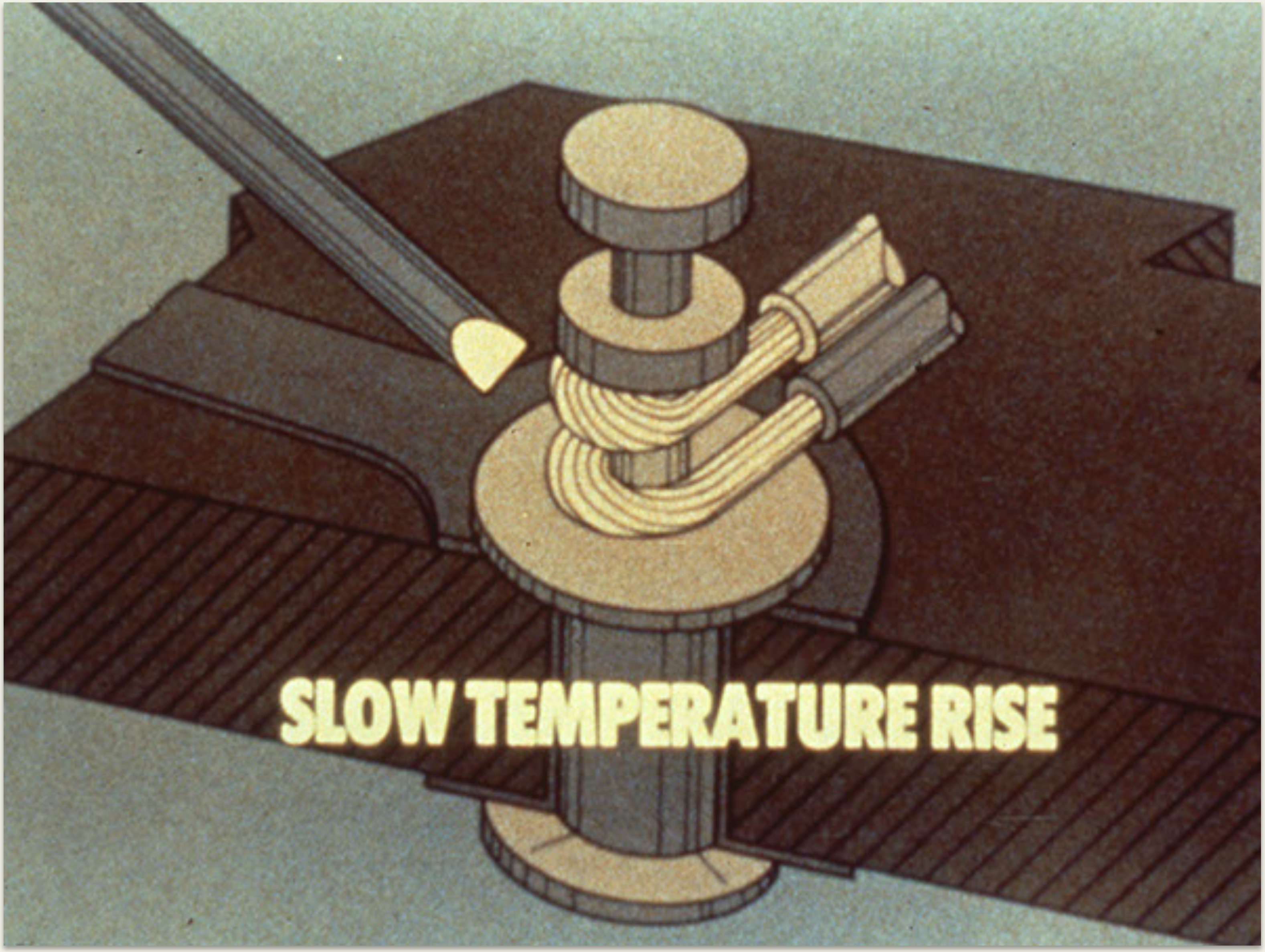
SLOW TEMPERATURE RISE

Small tip takes longer to heat the parts to the melting temperature

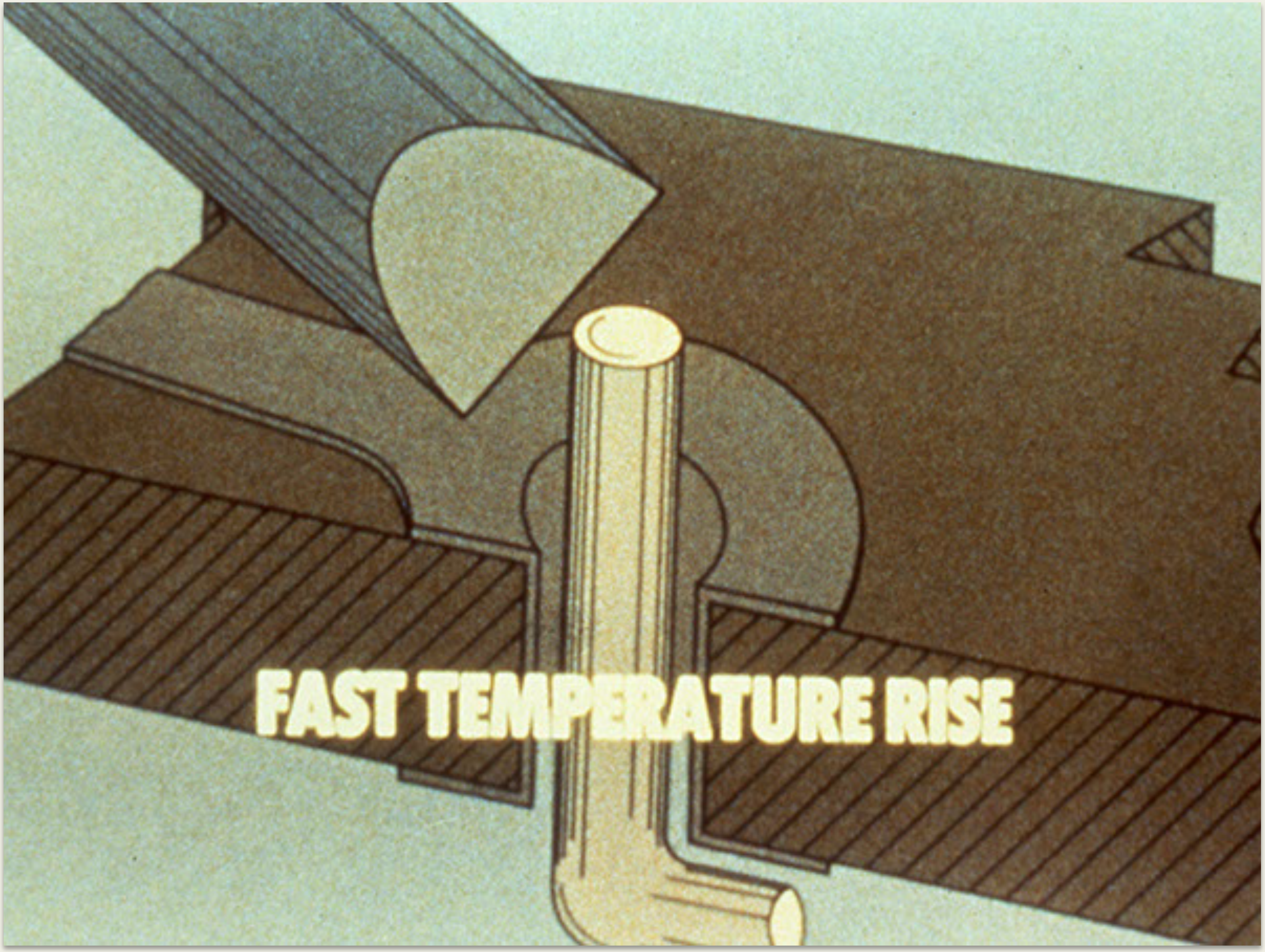
FAST TEMPERATURE RISE

Larger tip takes less time to heat the same part to the melting temperature

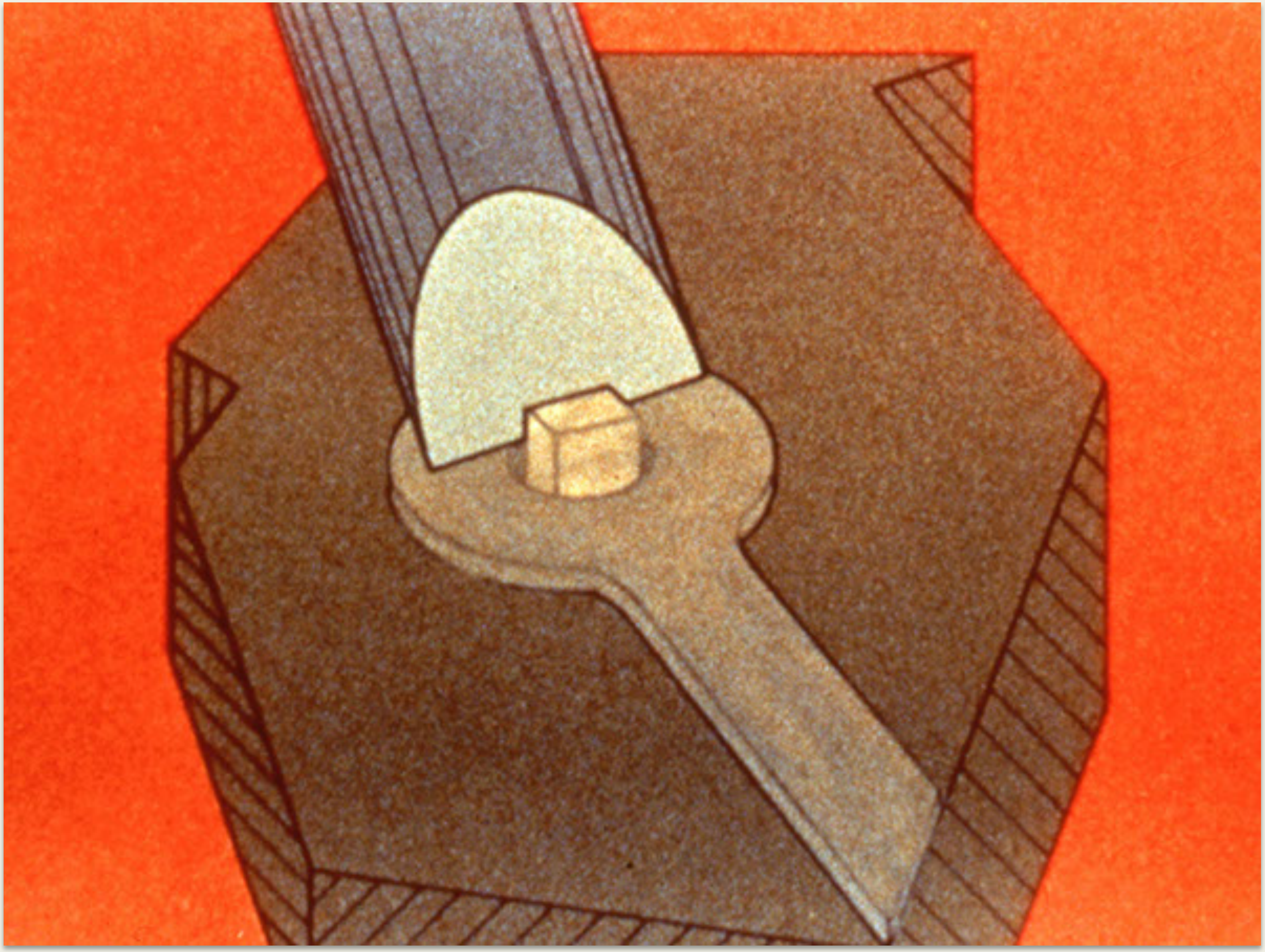

Proper sized tip

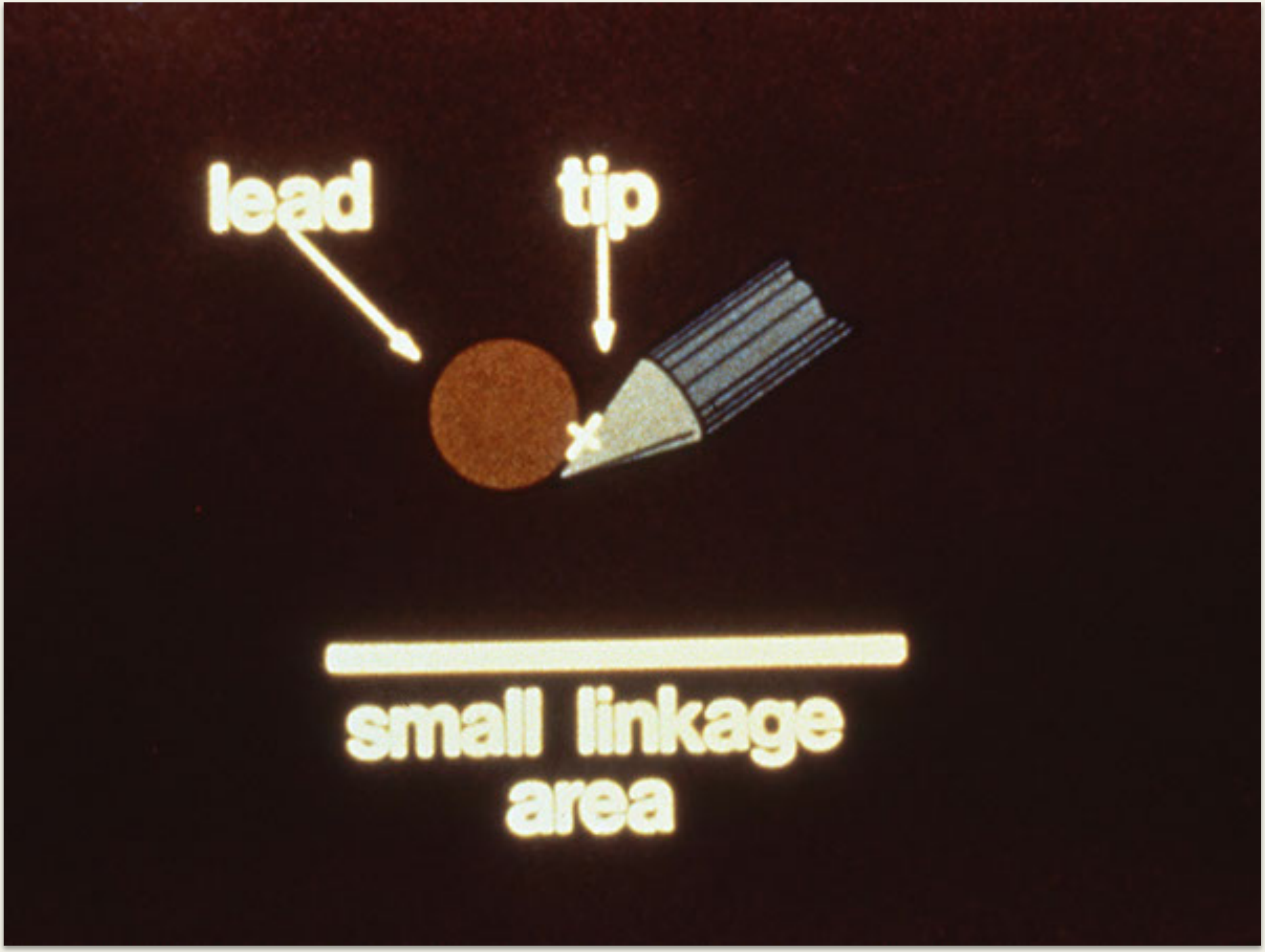
lead
tip
small linkage area

Small linkage

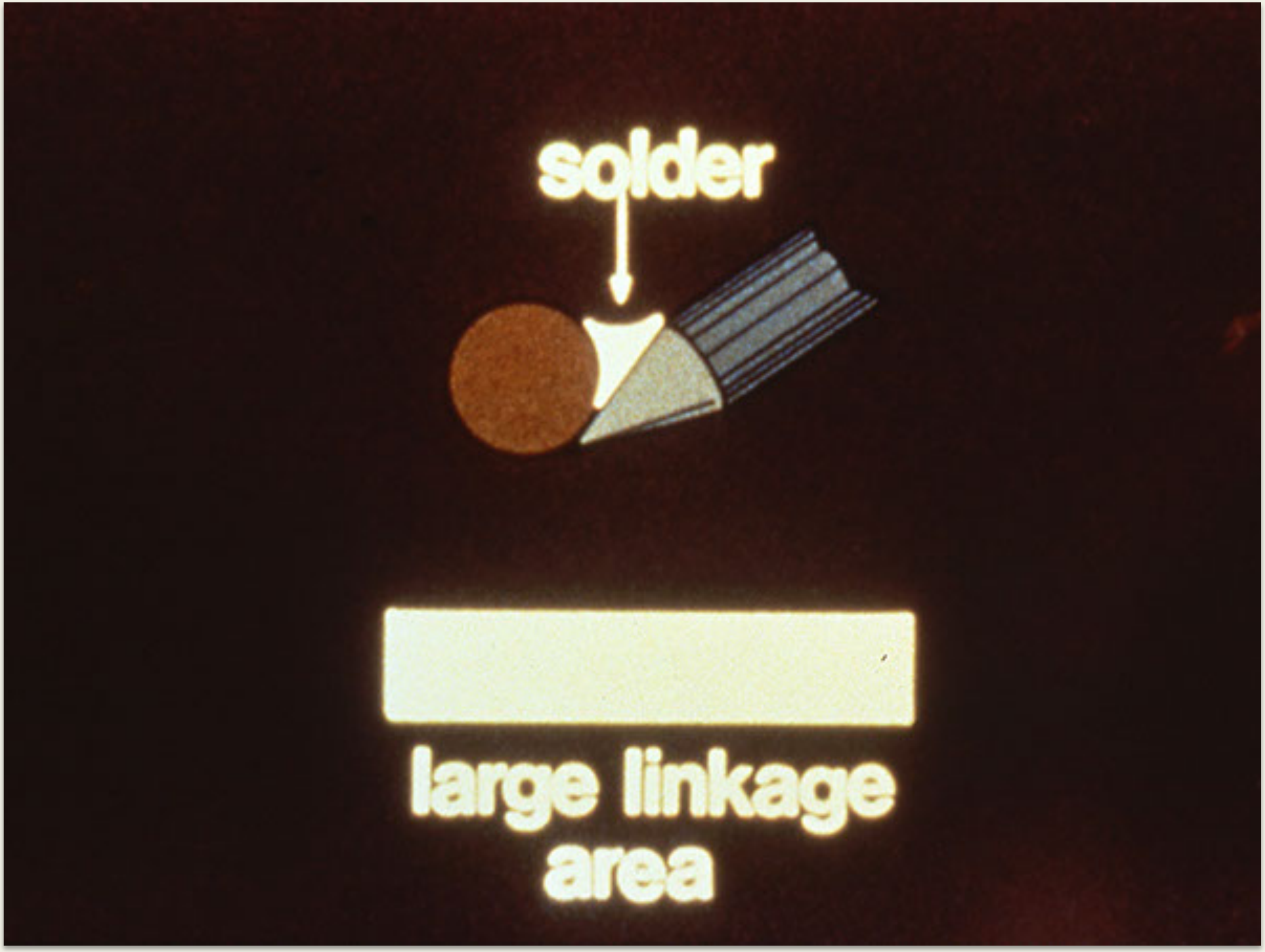
solder
large linkage area

Large linkage

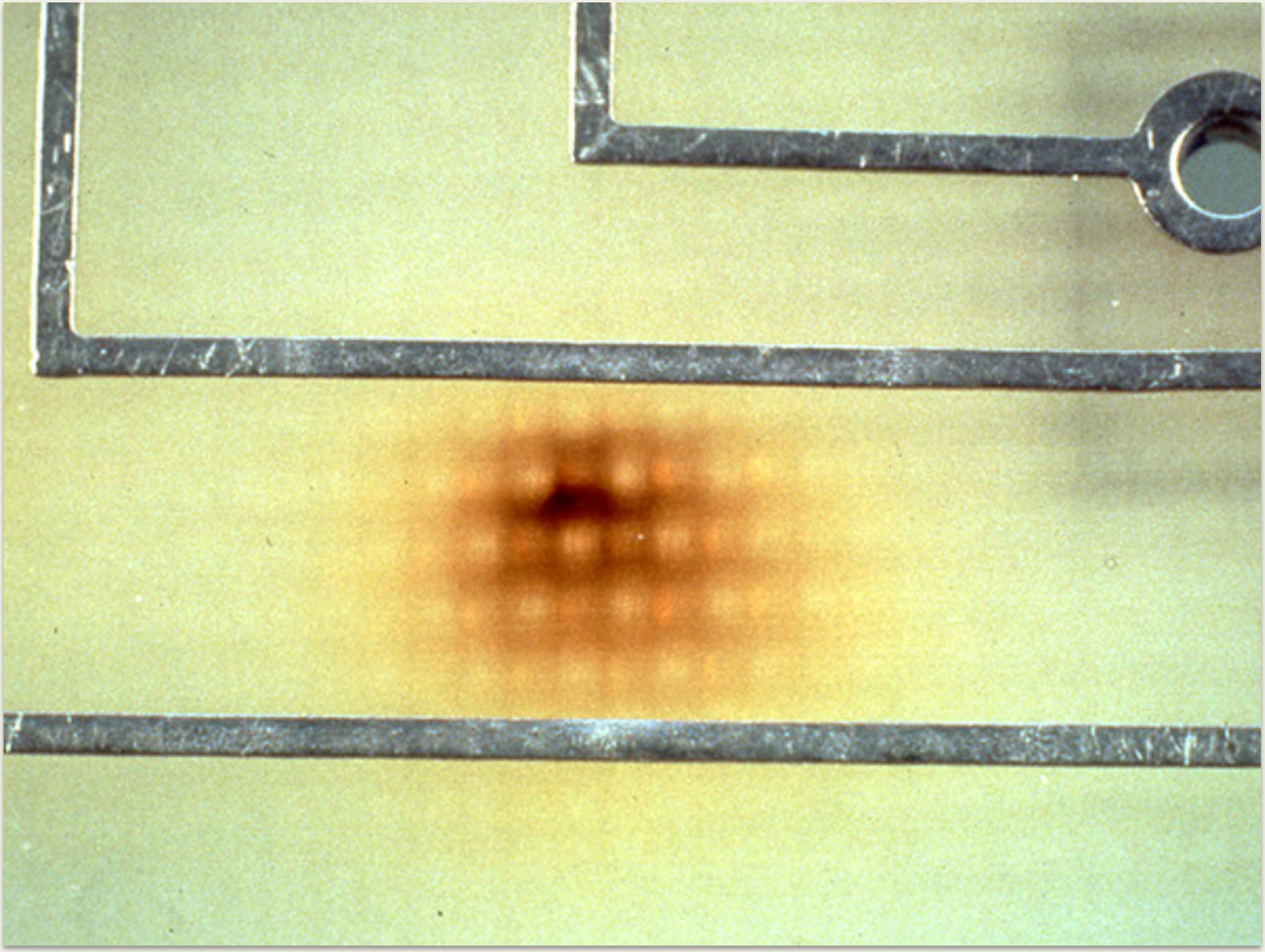

Burned printed circuit board

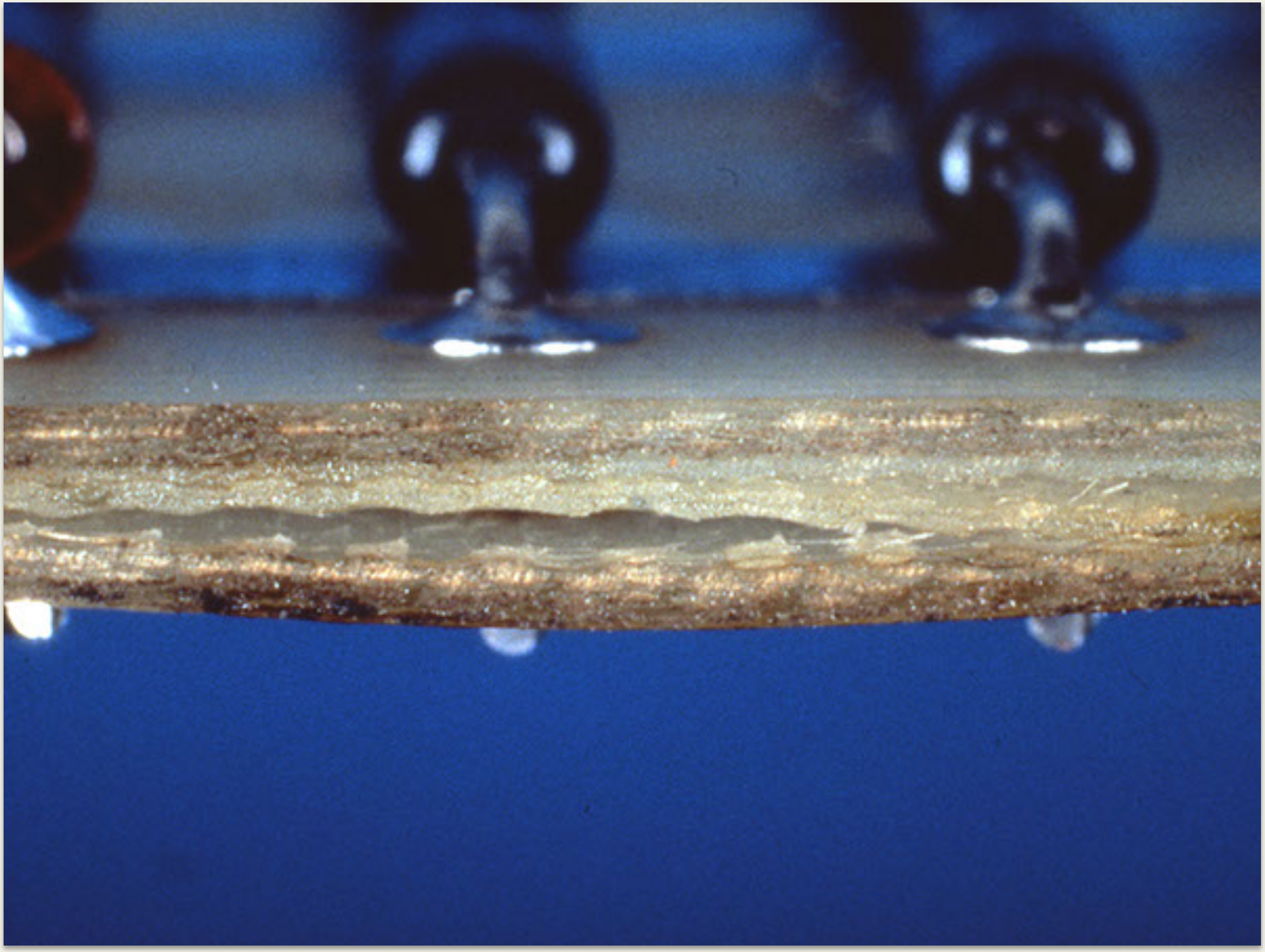

Delaminated printed circuit board

WIRE STRIPPING

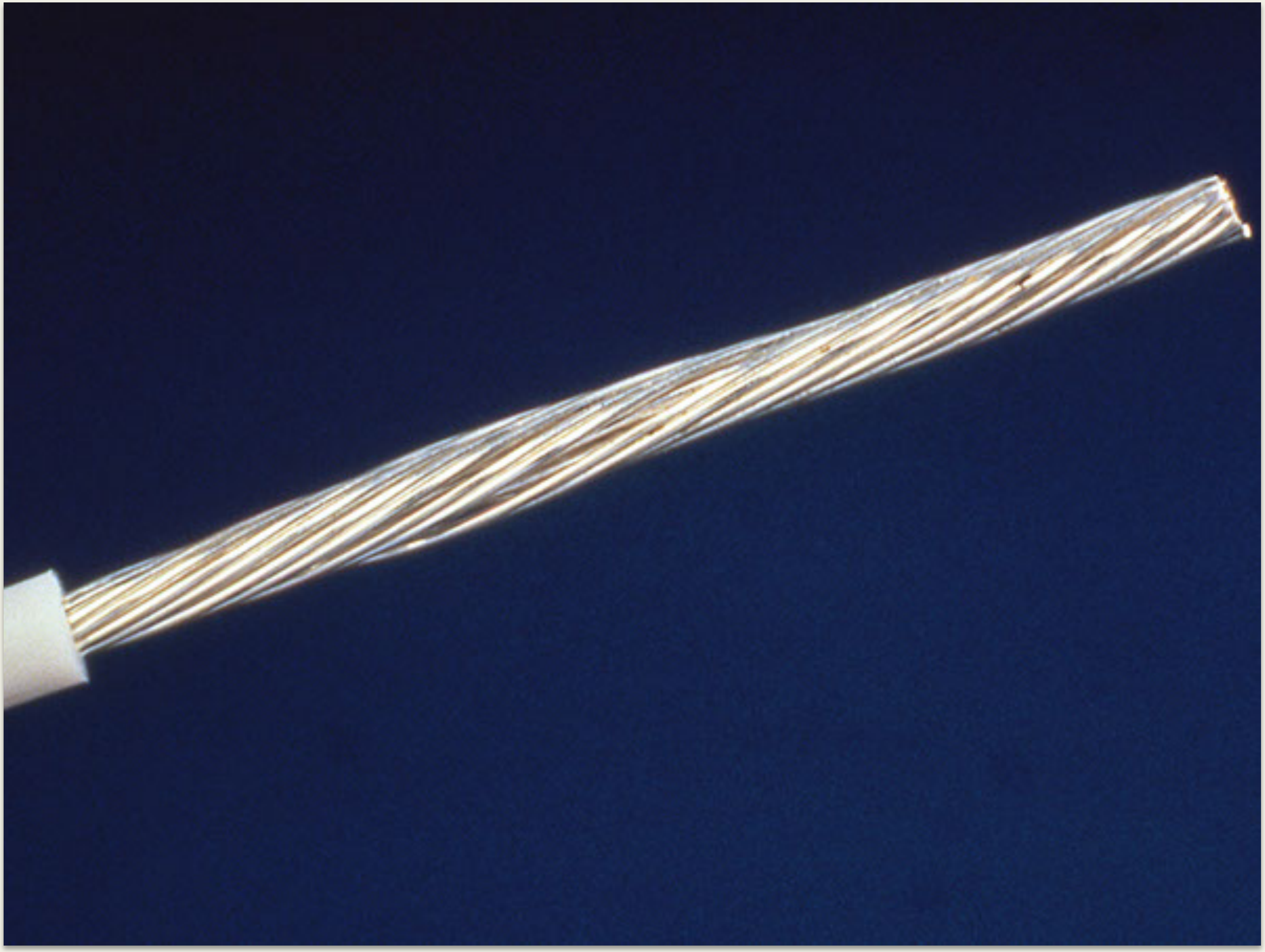

Disturbed lay of stranded wire - Reject

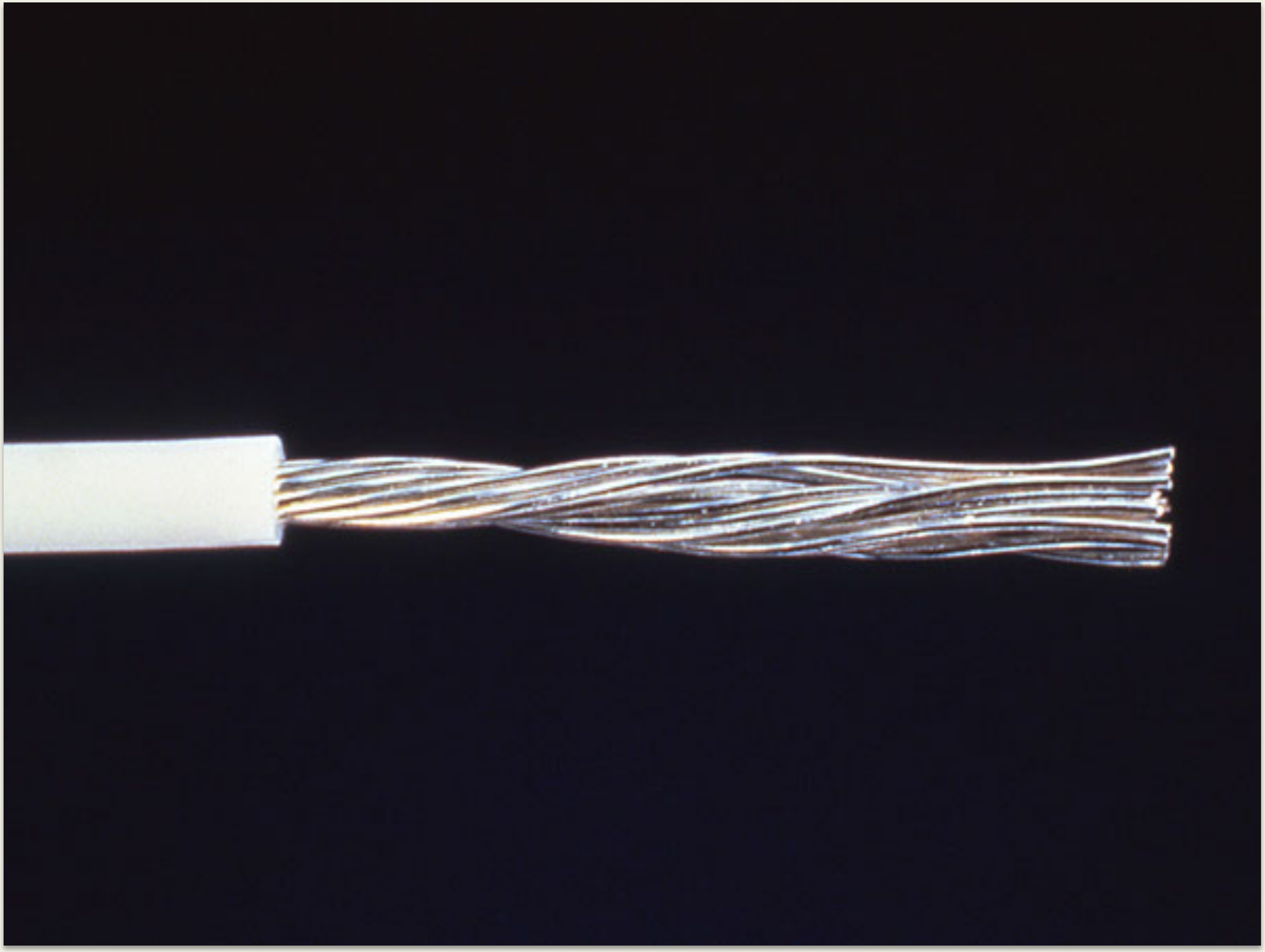

Disturbed lay of stranded wire - Reject

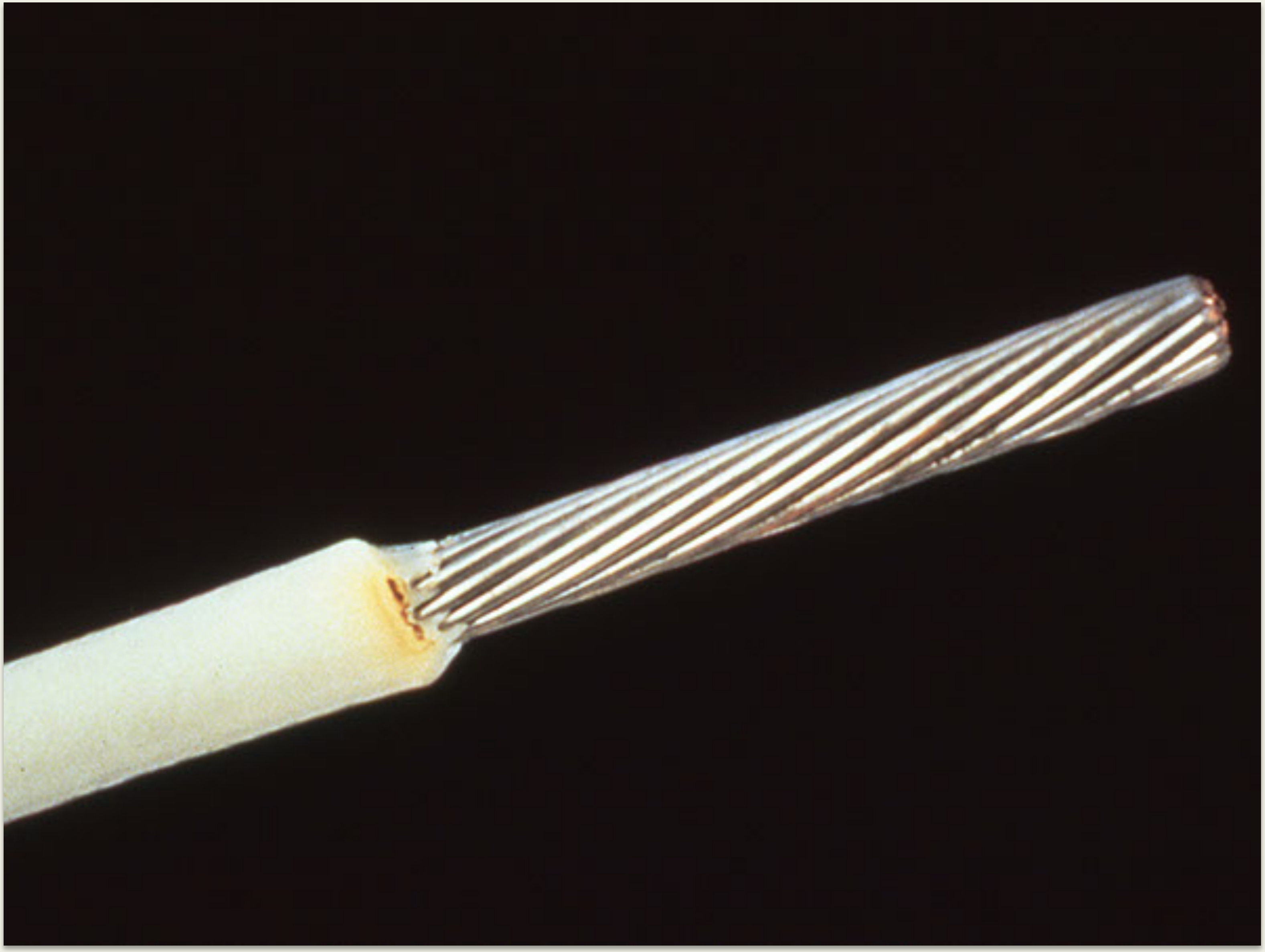

Discoloration burn to insulation - Accept

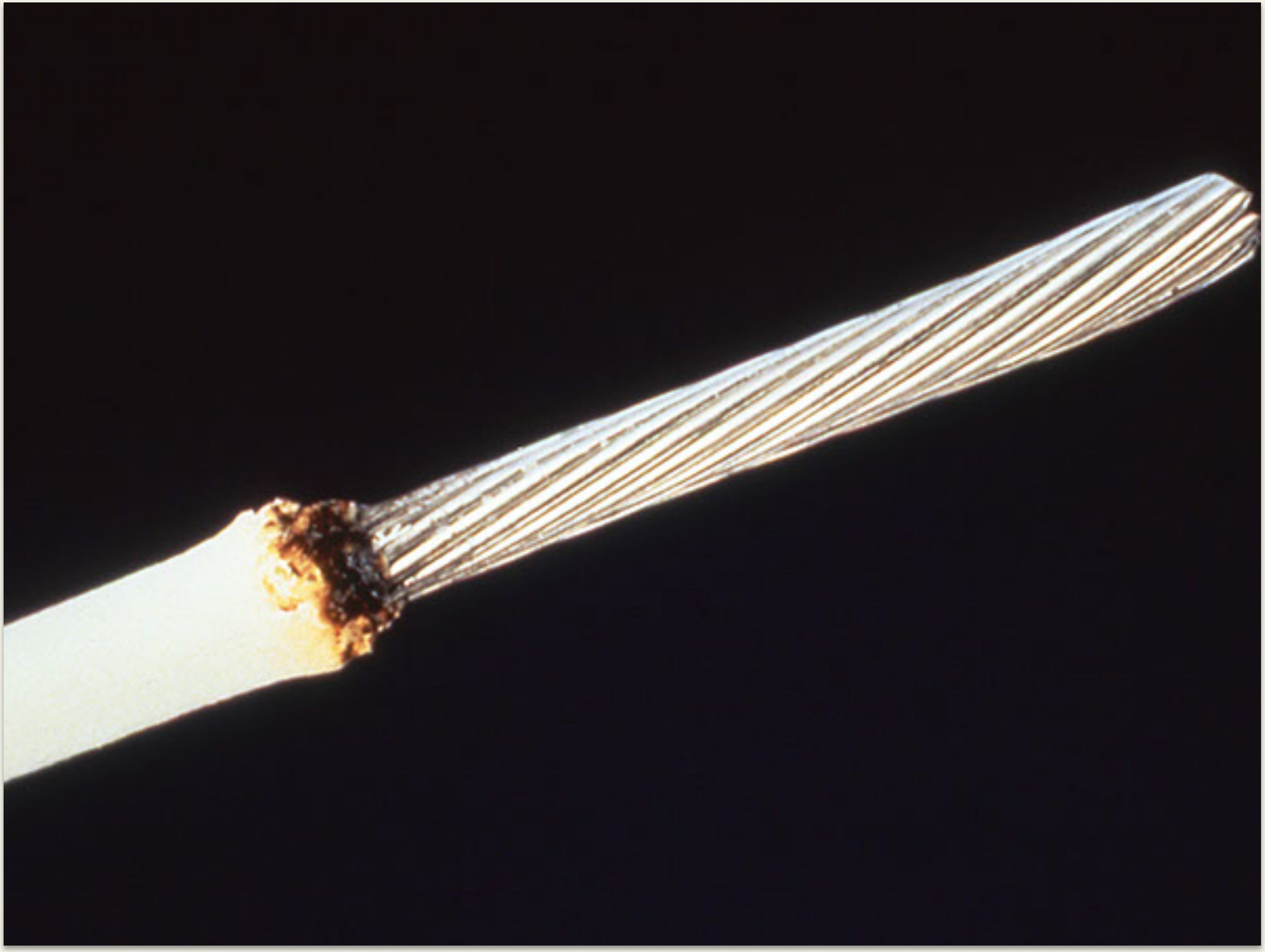

Charring: burning or damage to insulation - Reject

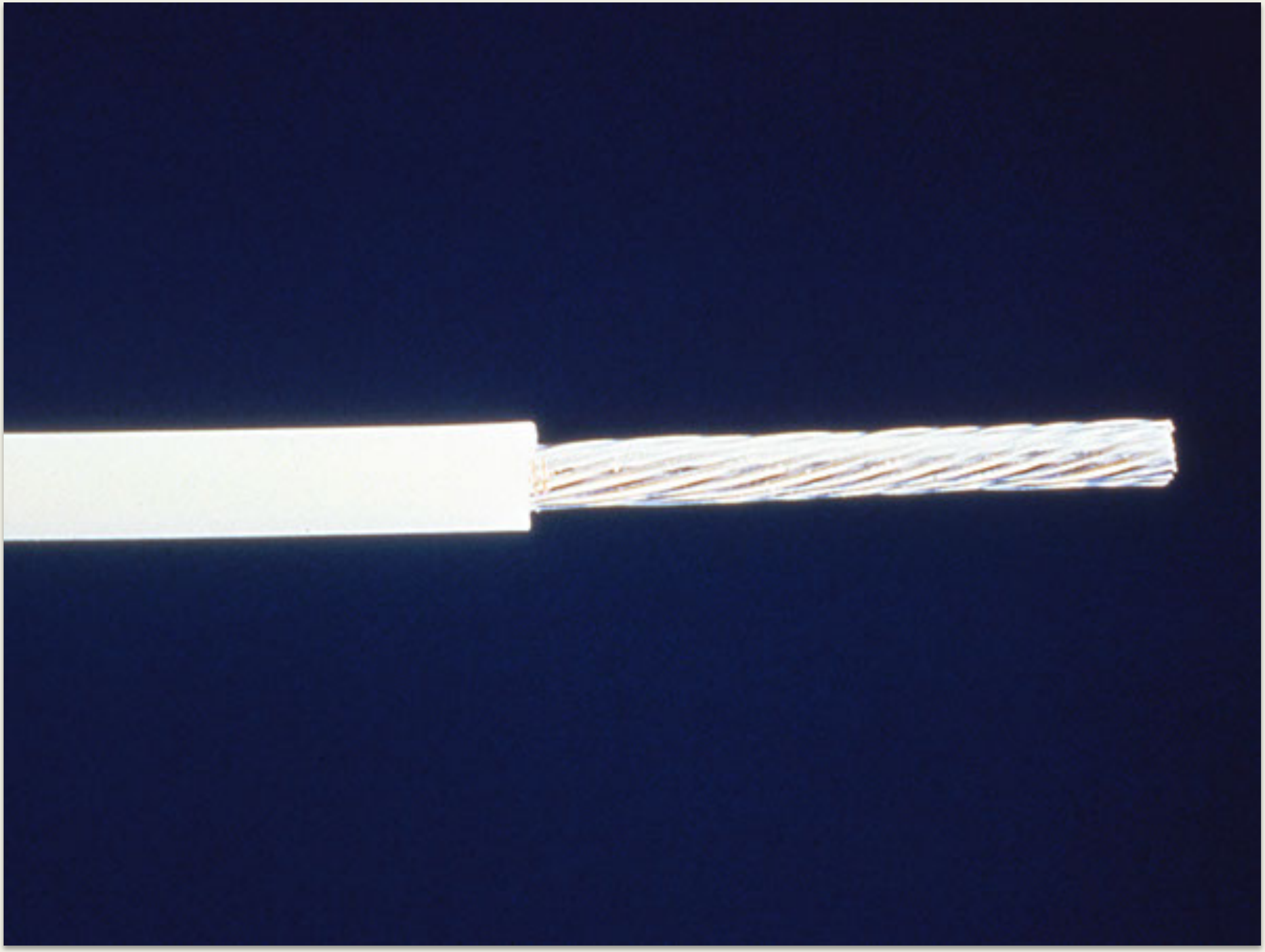

Cut or nicked leads or wires - Reject

WIRE
TINNING

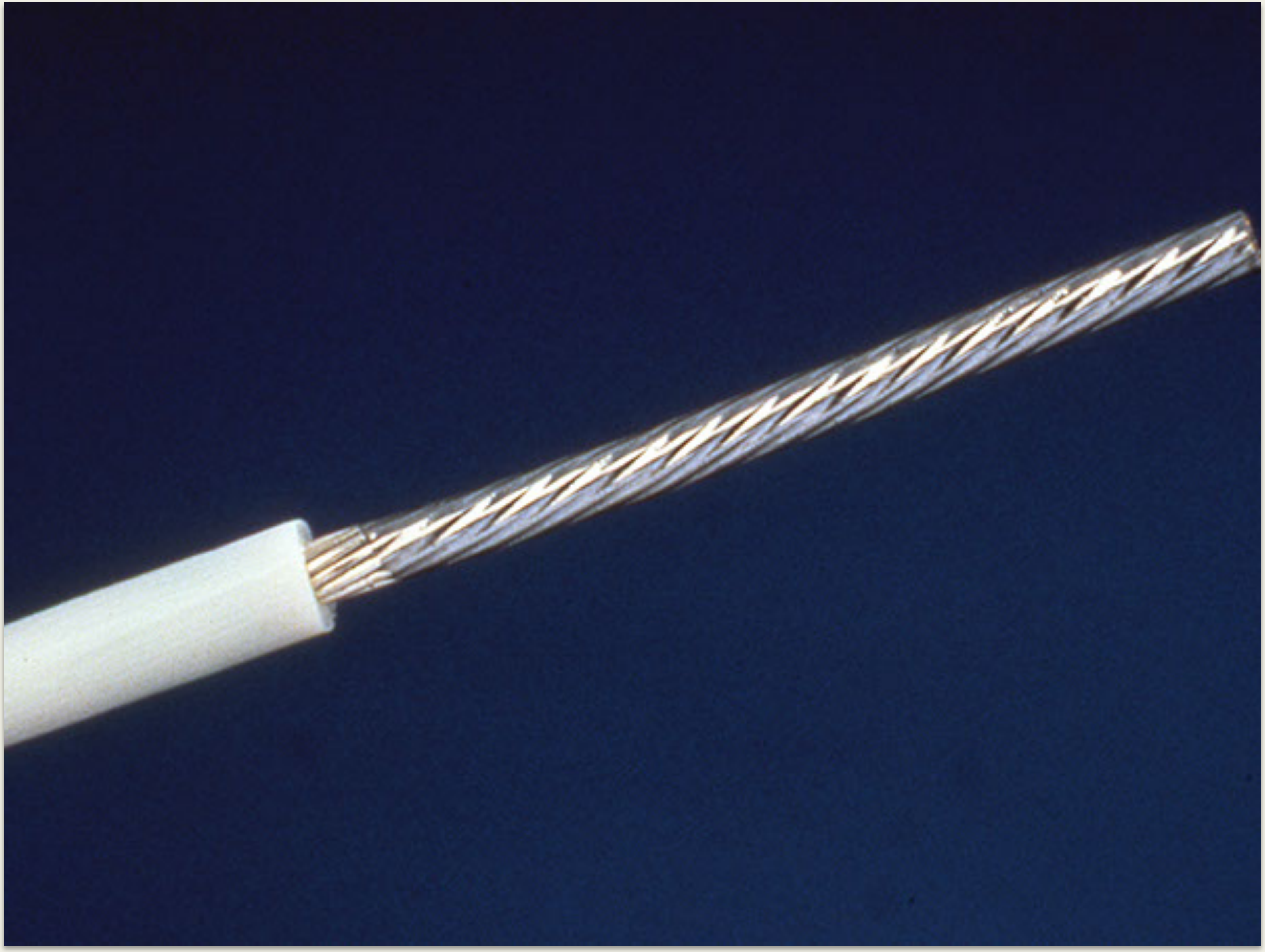

Preferred tinning of stranded wire with heat sink - Accept

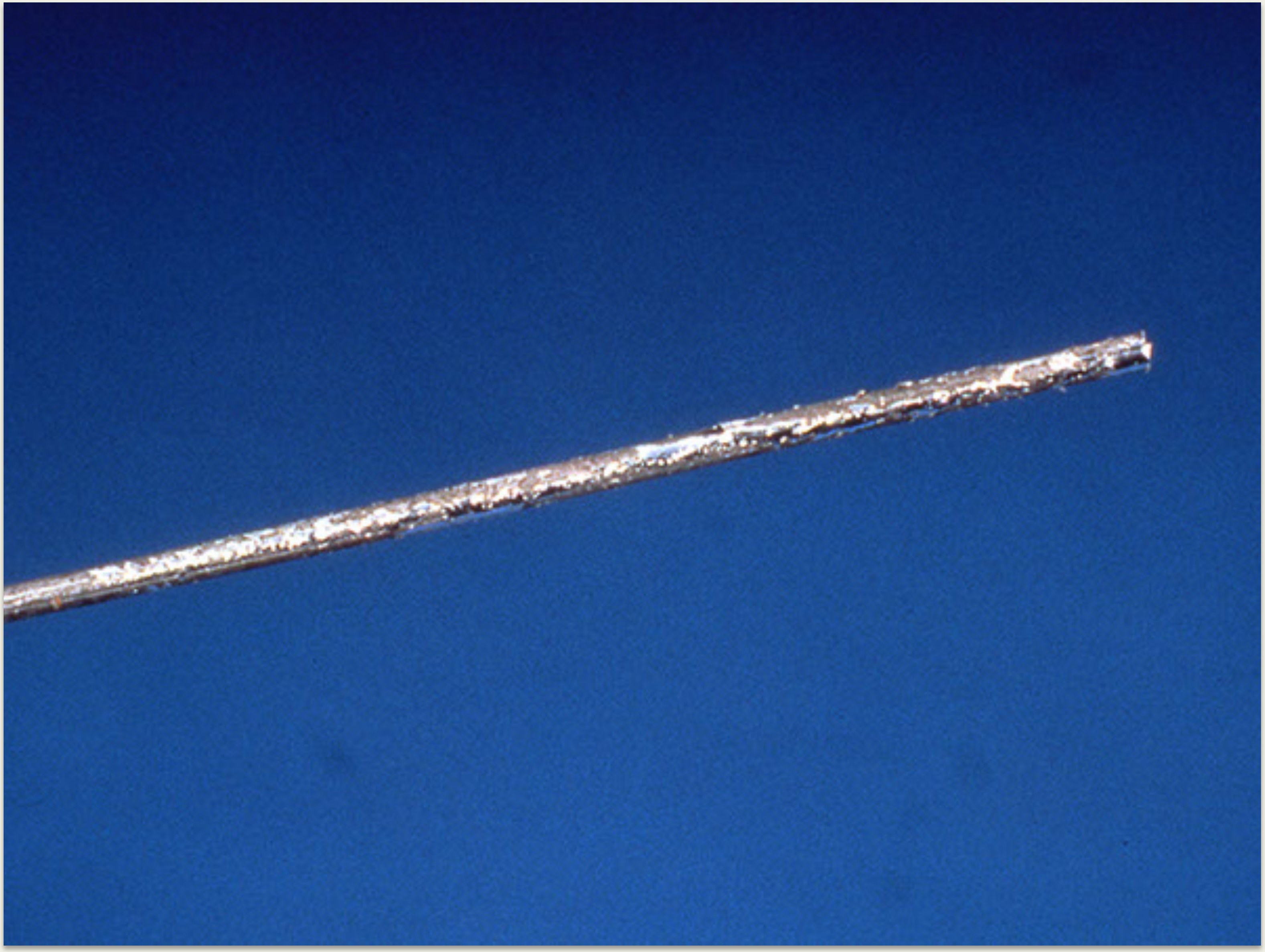

Poor tinning, no wetting - Reject

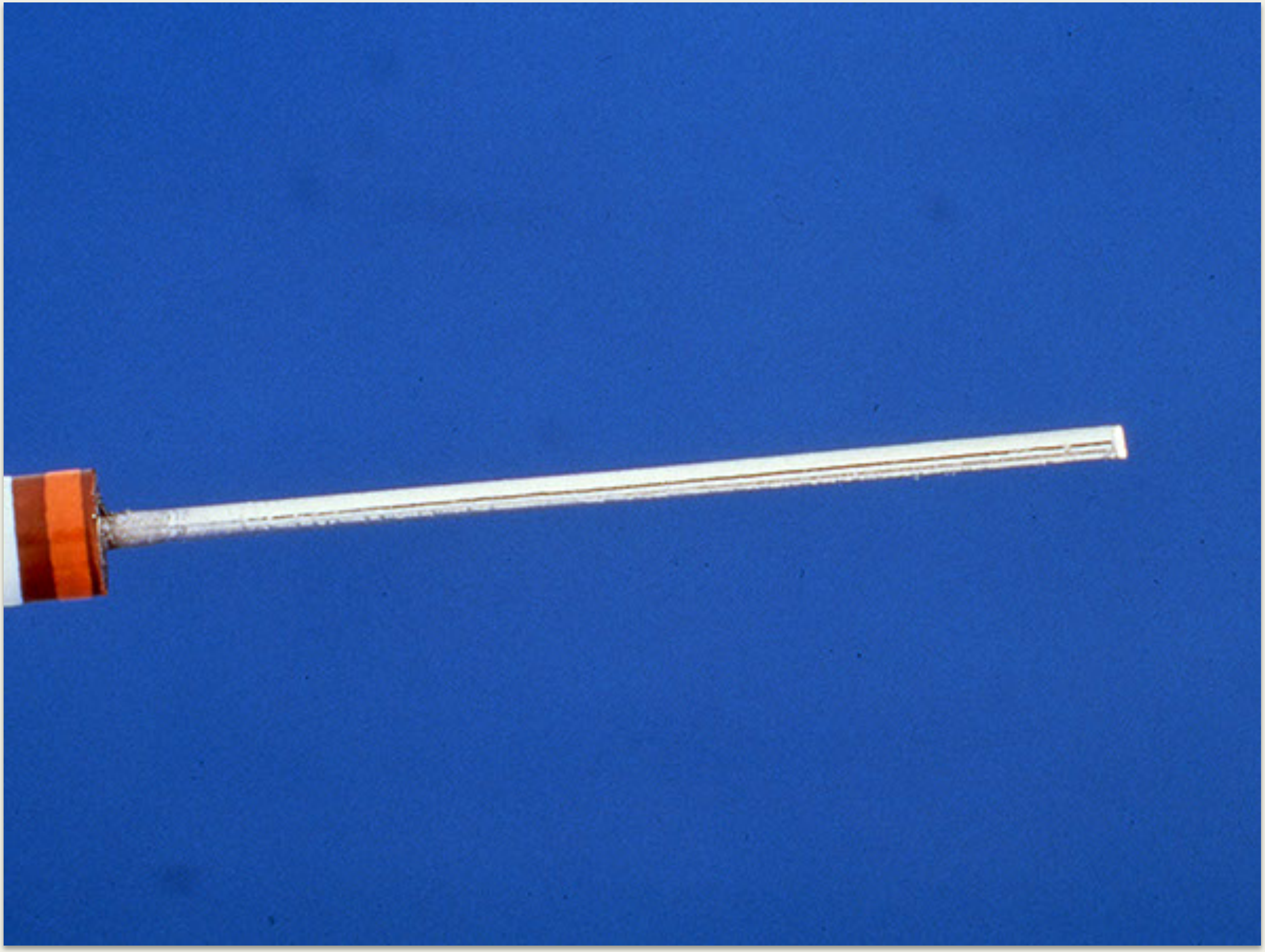

Tinning, good wetting - Accept

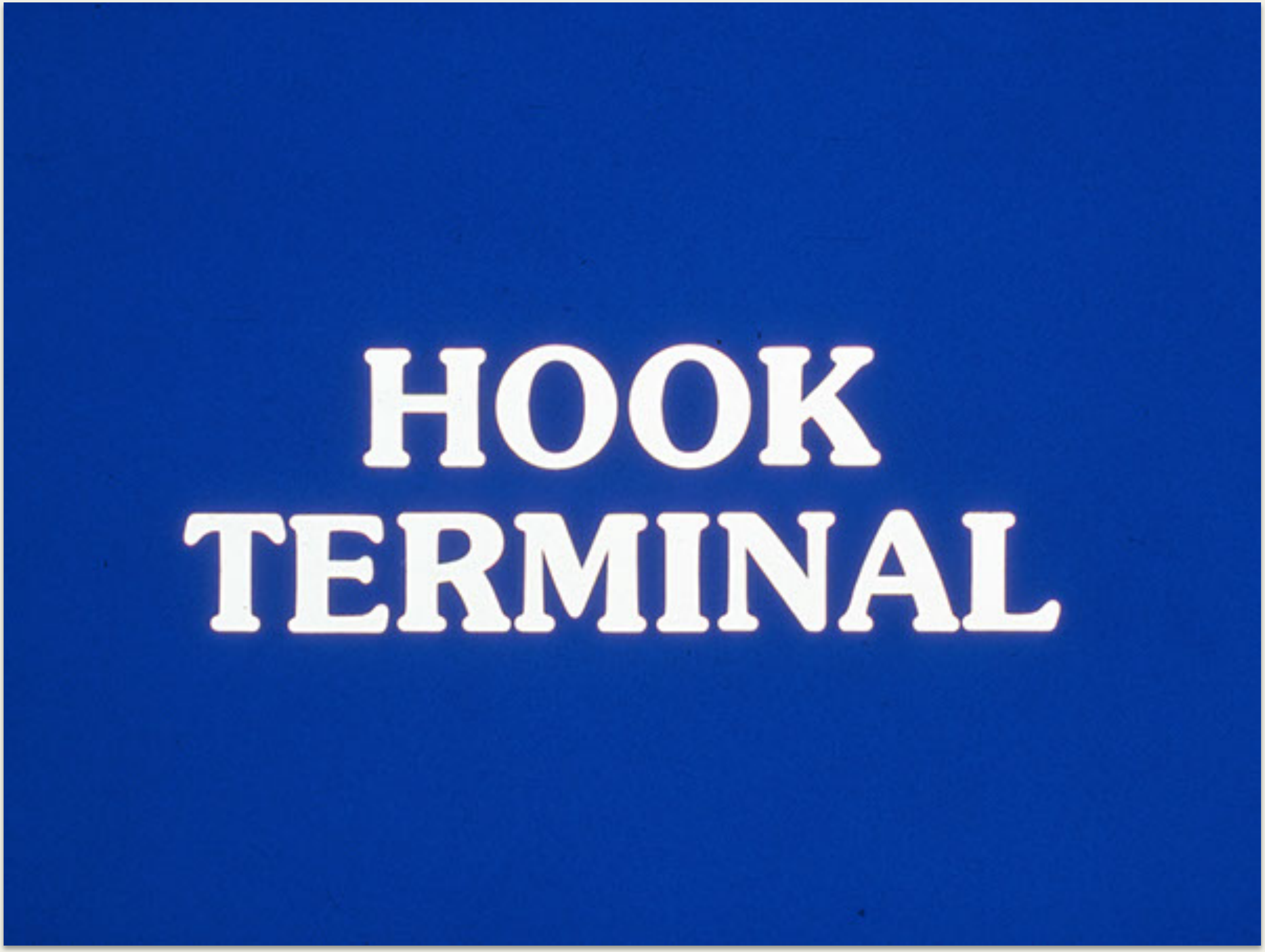

HOOK TERMINAL

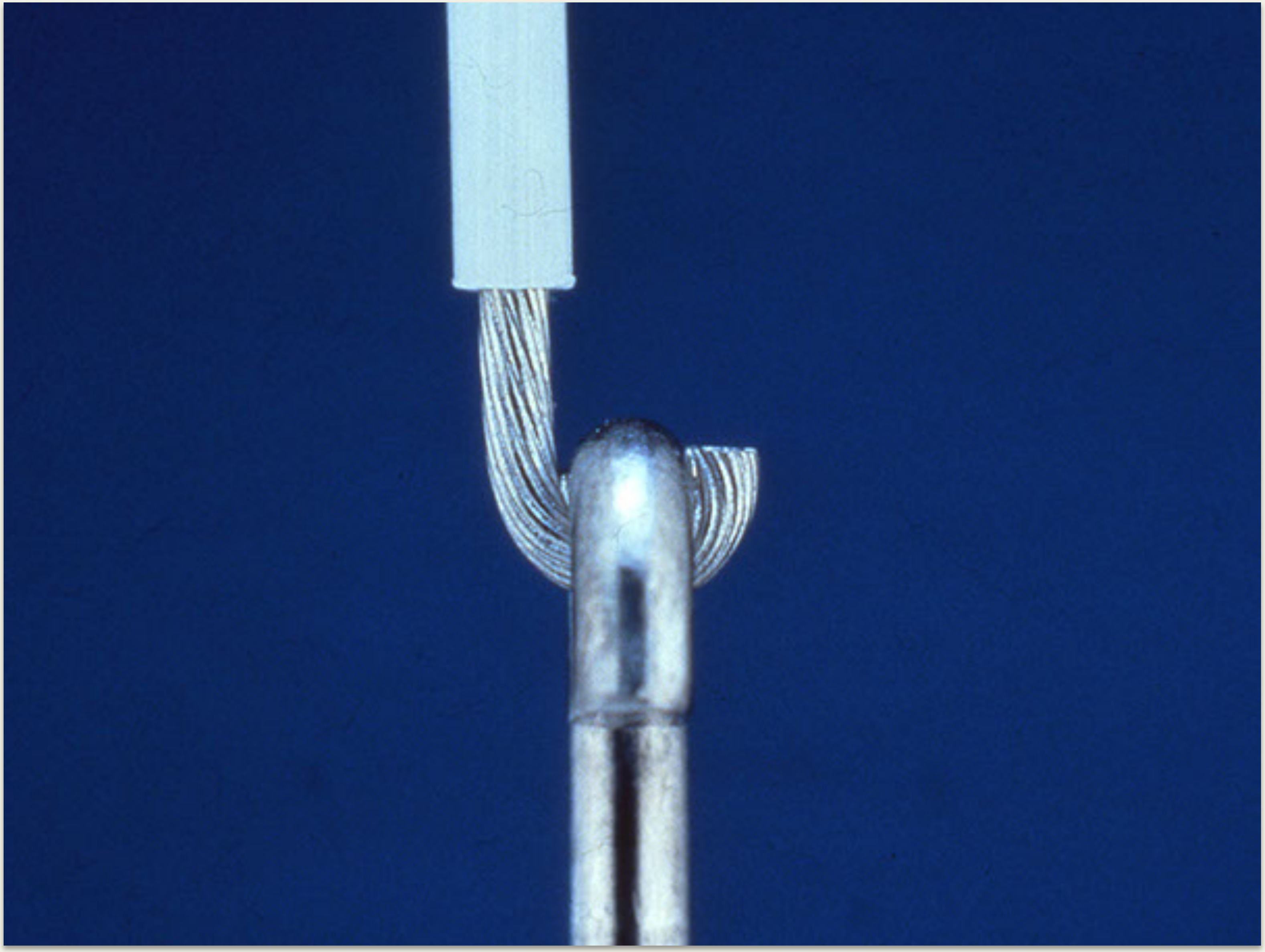

Preferred wrap, 180 degrees - Accept

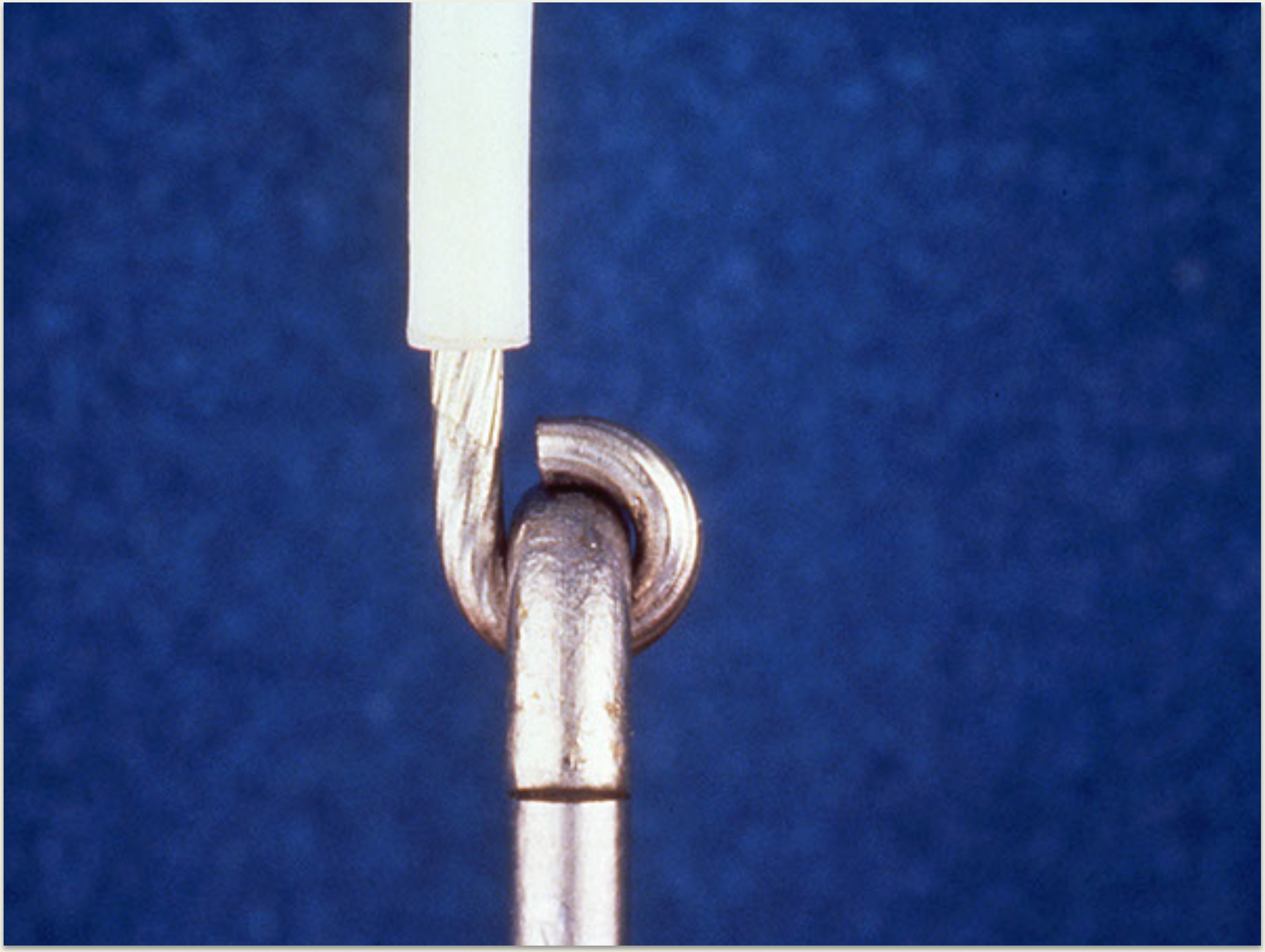

Maximum wrap, 270 degrees - Accept

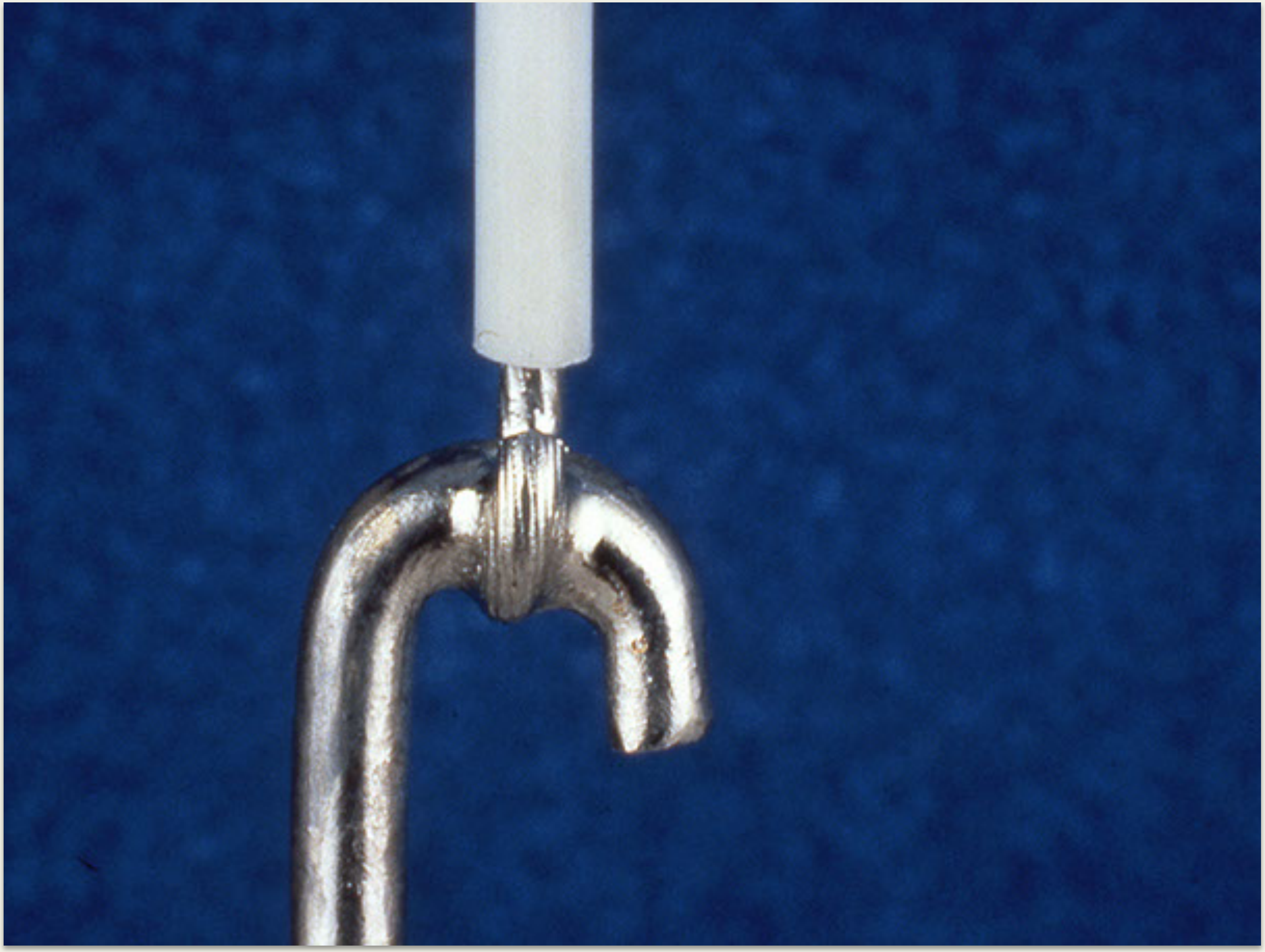

Preferred solder - Accept

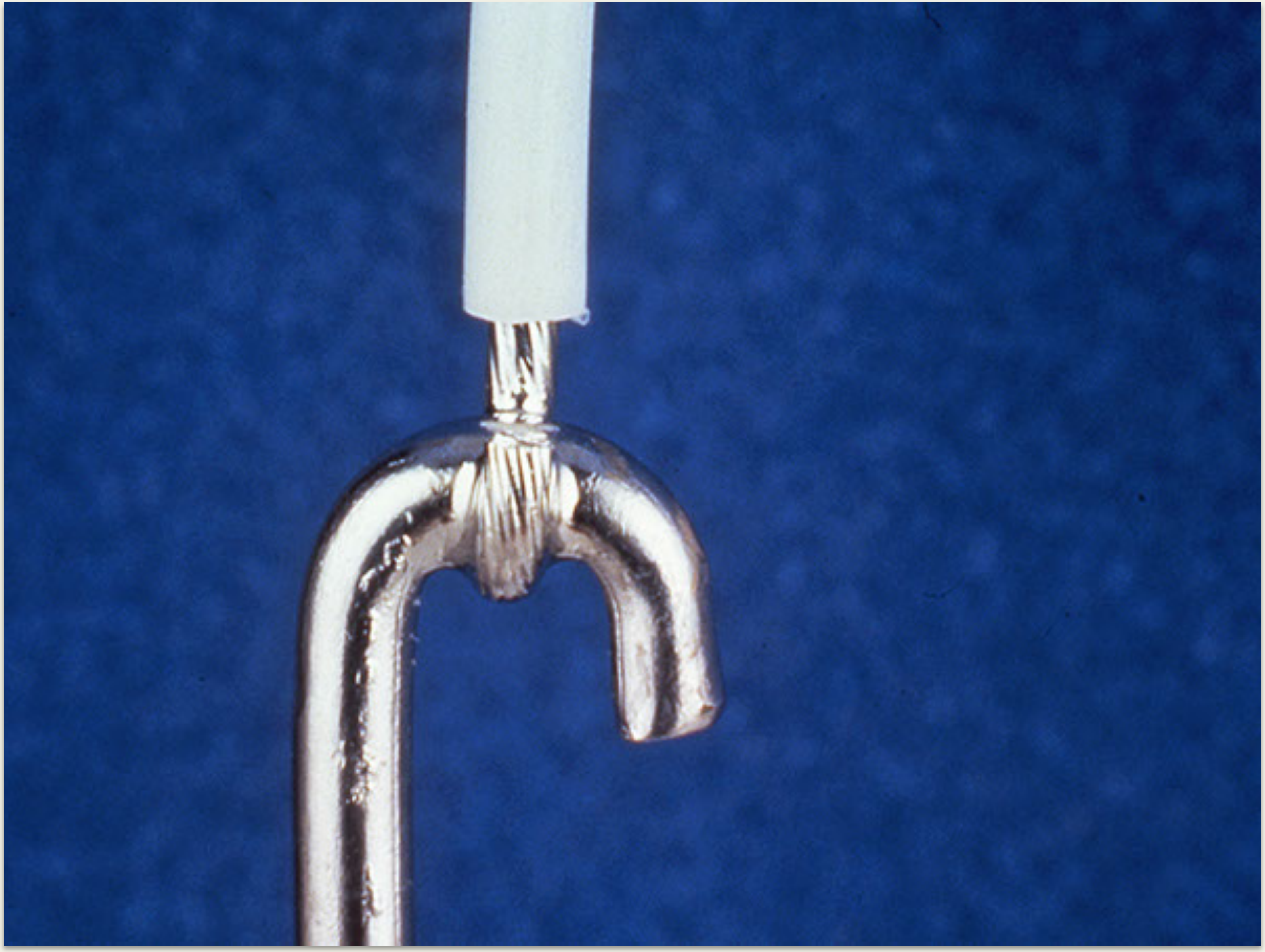

Minimum solder - Accept

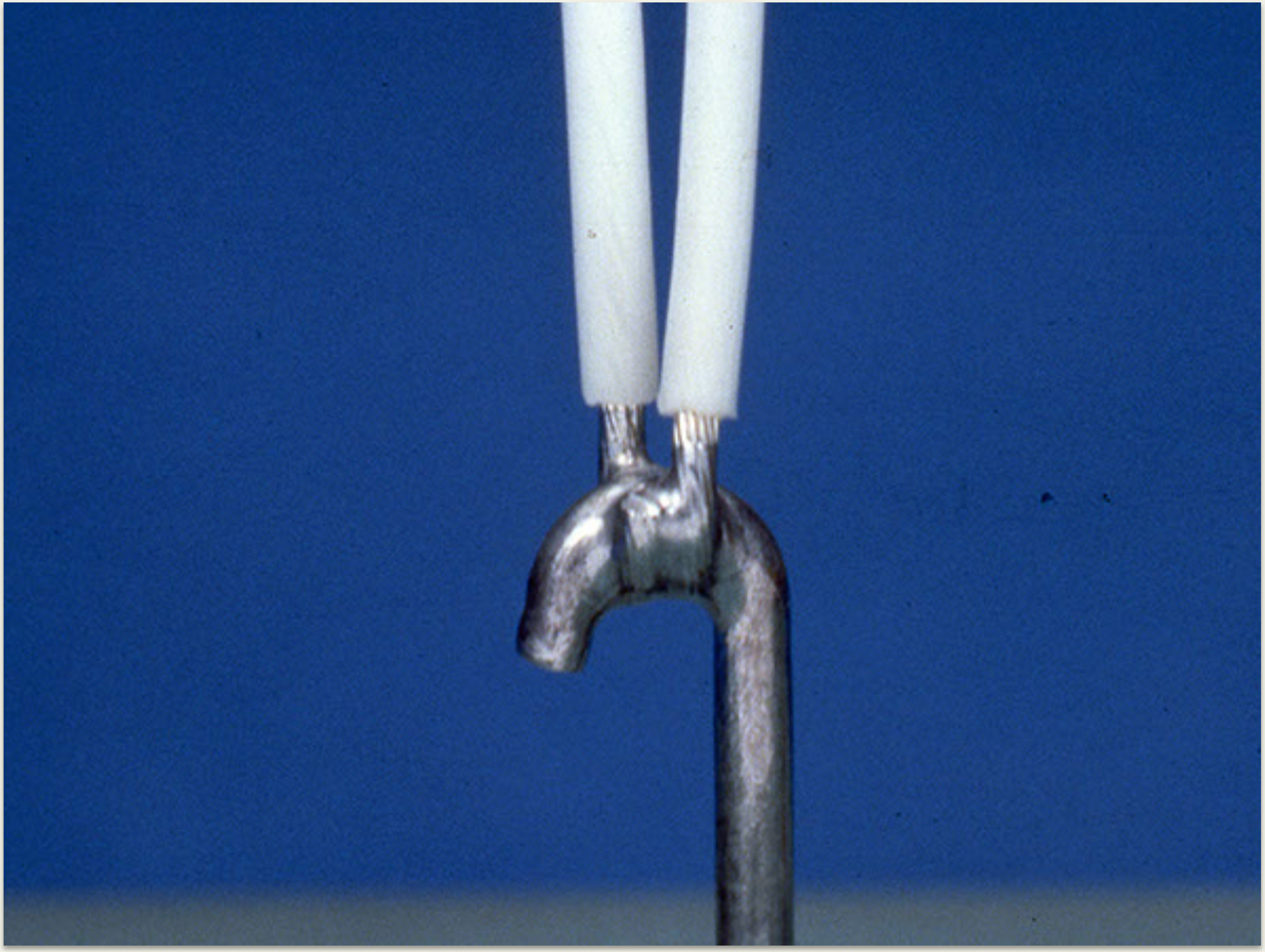

Double wrap - Accept

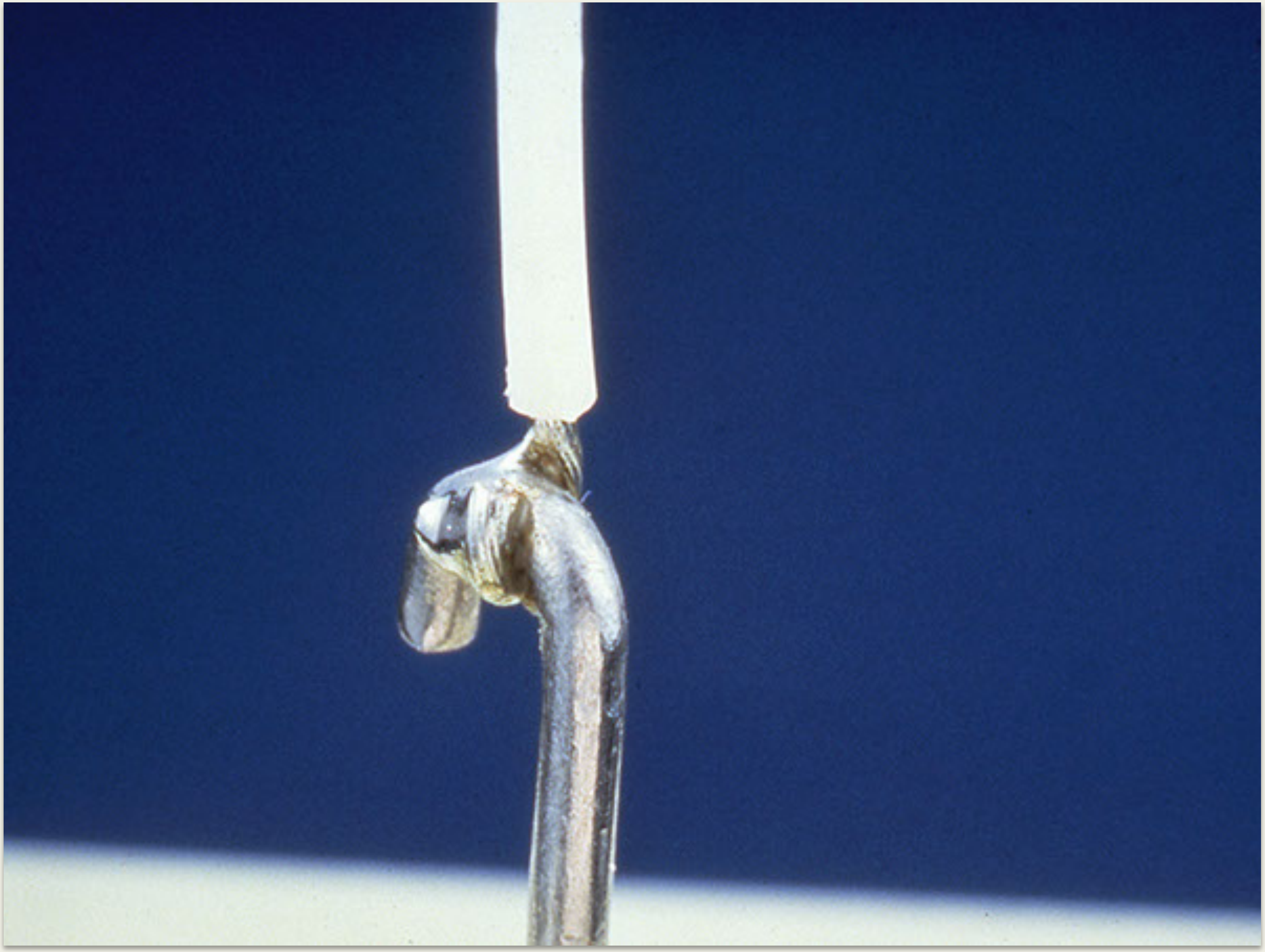

Rosin connection – Reject

PIERCED TERMINAL

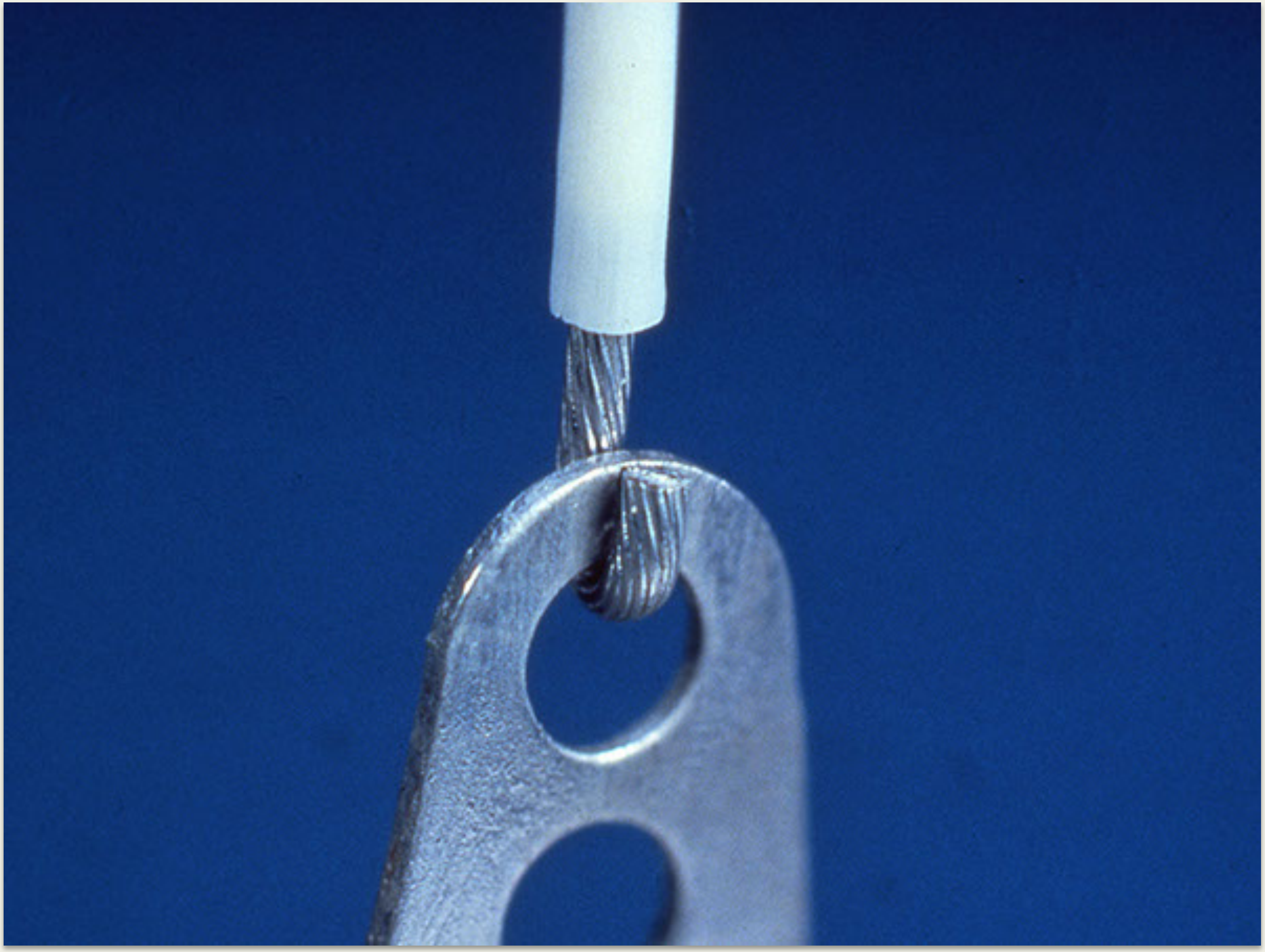

Preferred wrap - Accept

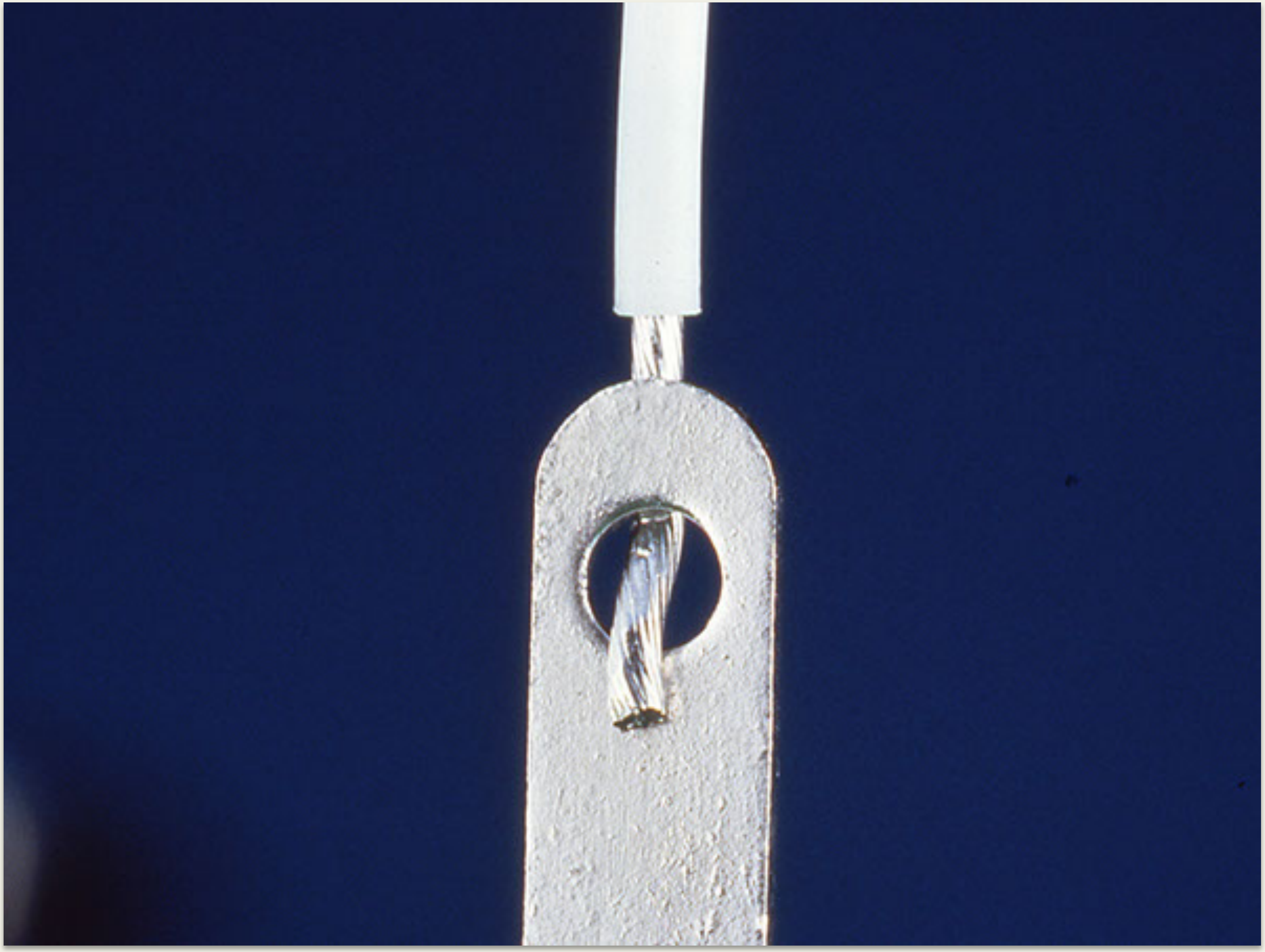

Preferred wrap “Z bend” - Accept

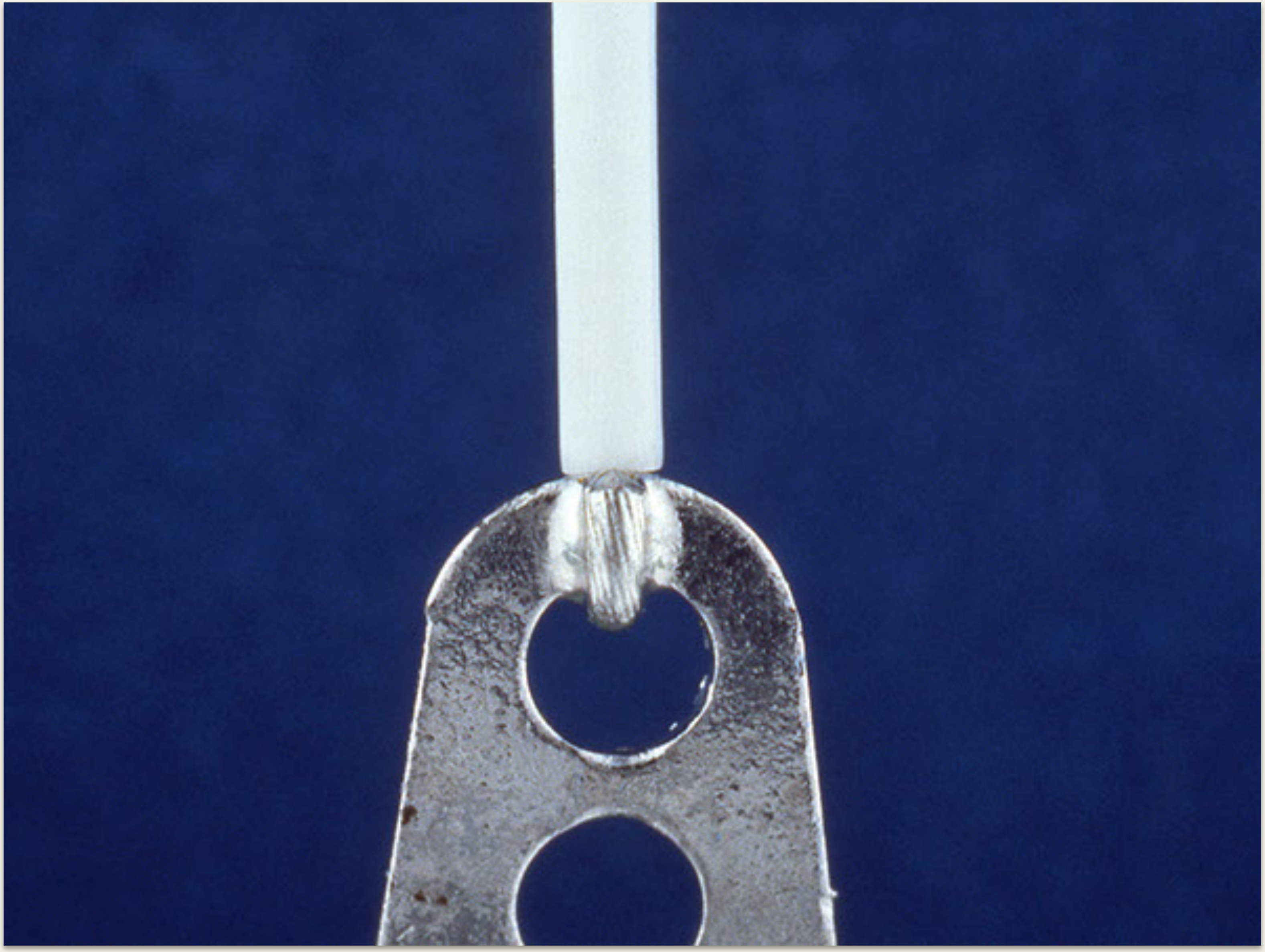

Minimum insulation clearance - Accept

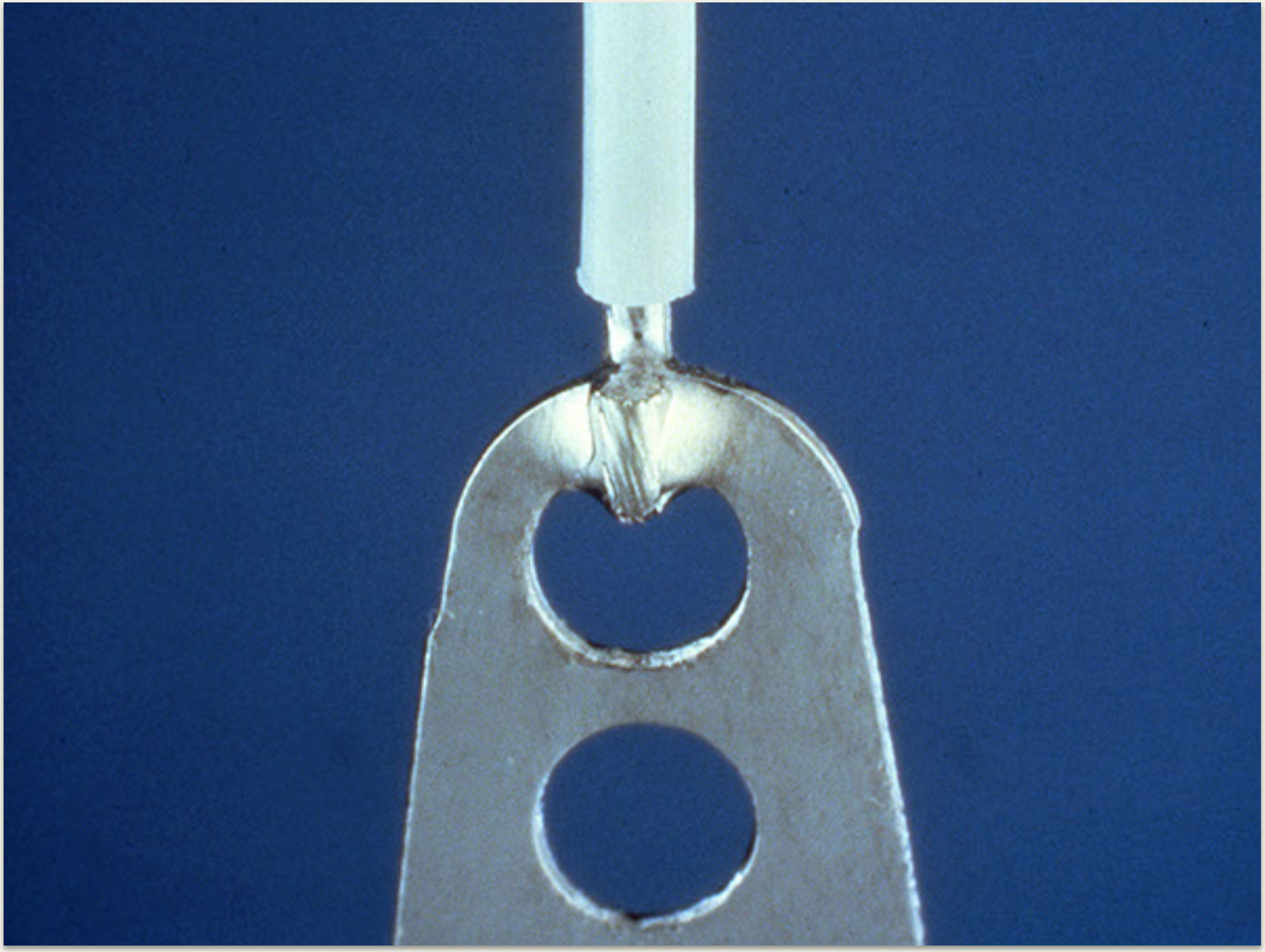

Preferred insulation clearance - Accept

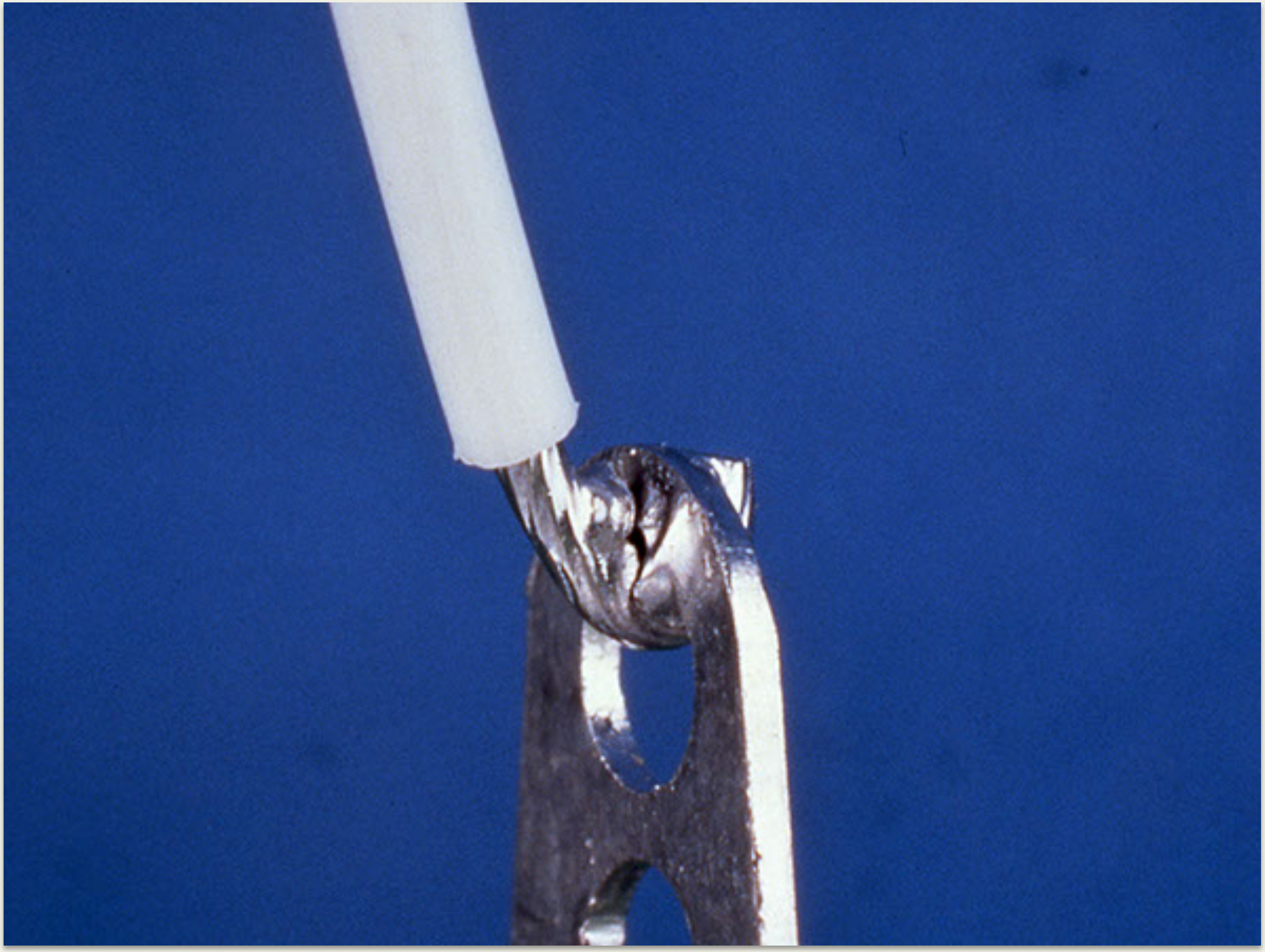

Disturbed/Fractured connection - Reject

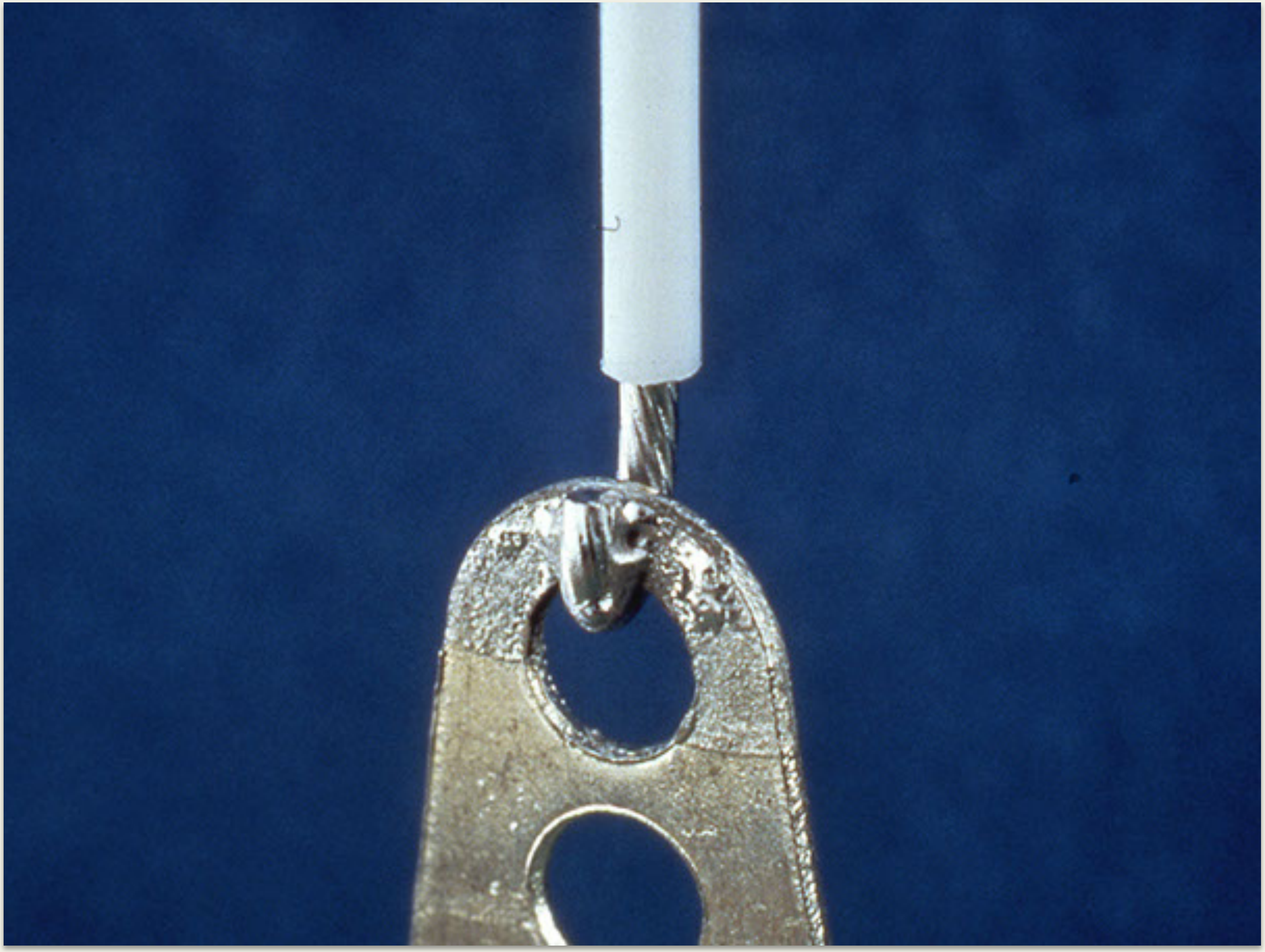

Dewetting - Reject

TURRET TERMINAL

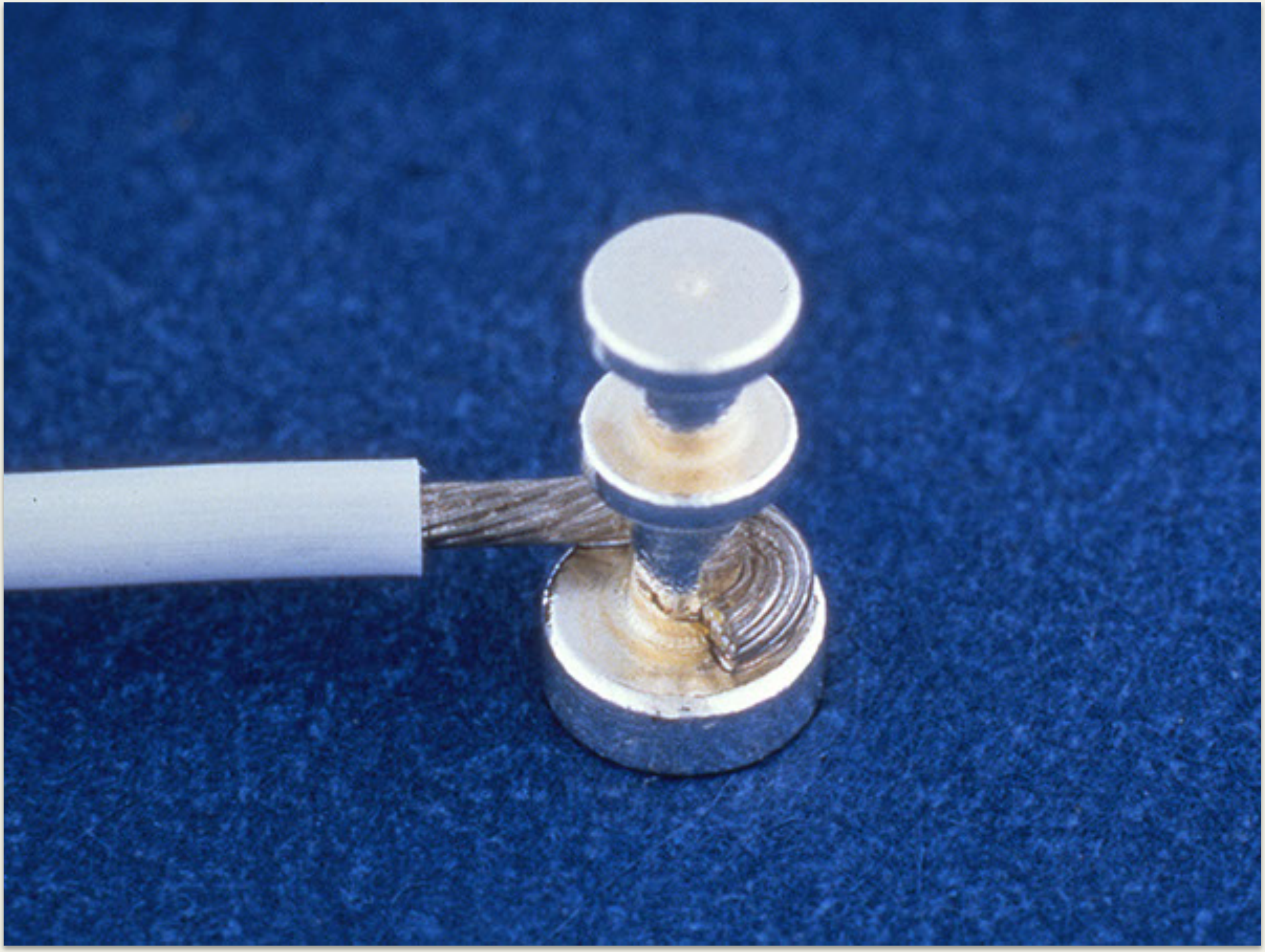

Insufficient wrap, less than 180 degrees - Reject

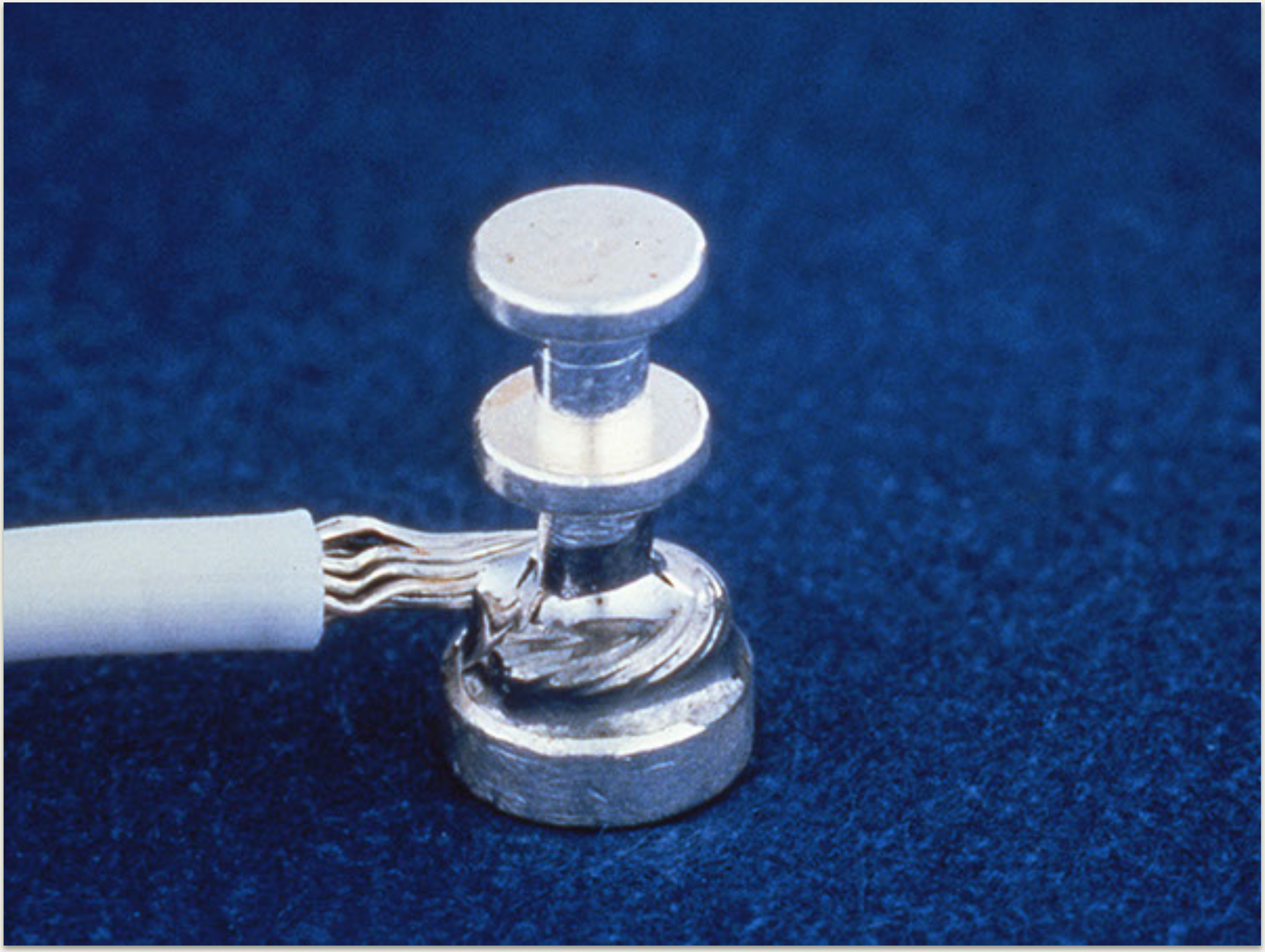

Birdcaged wire strands - Reject

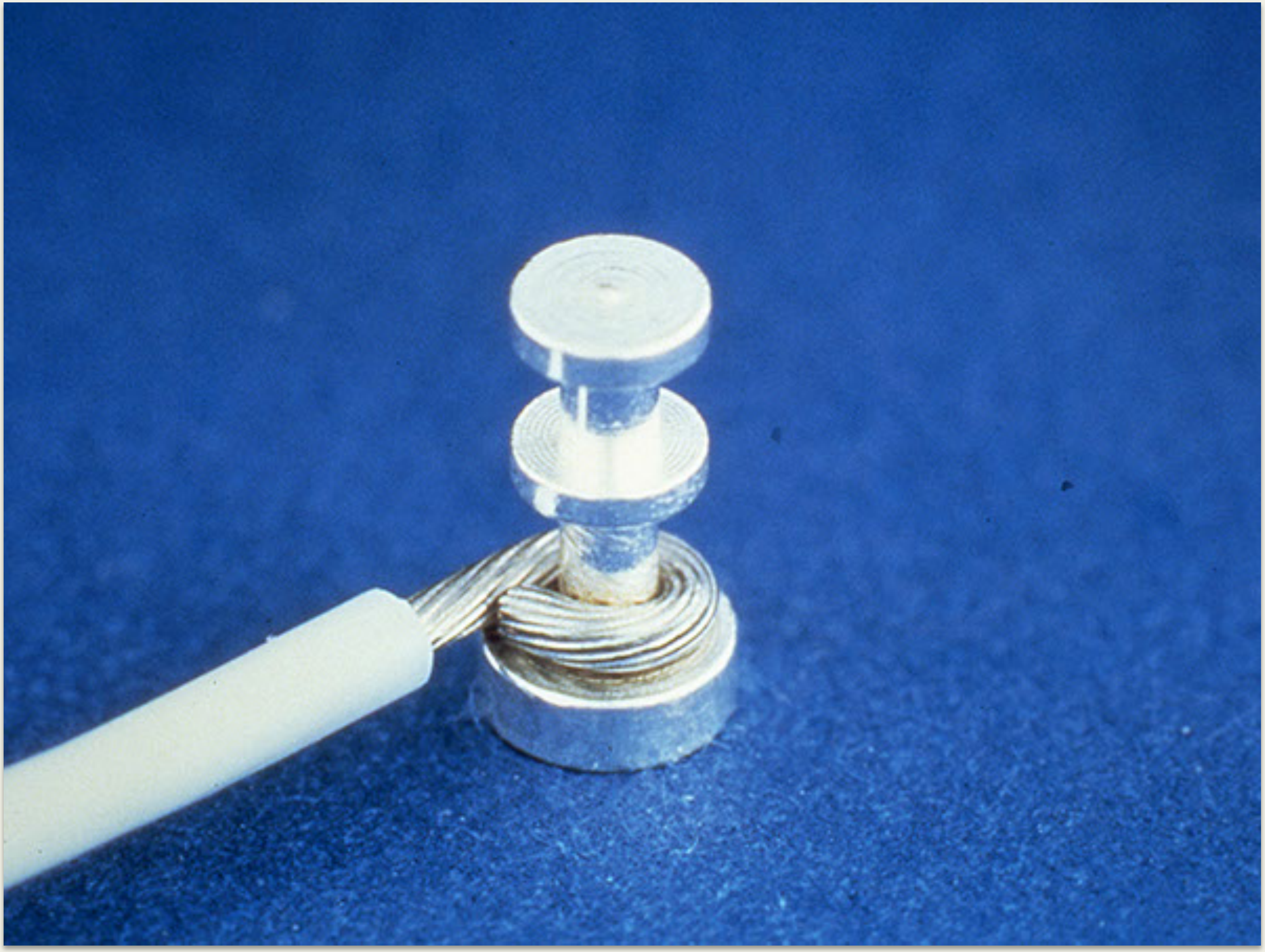

Loose/Excessive lead wrap - Reject

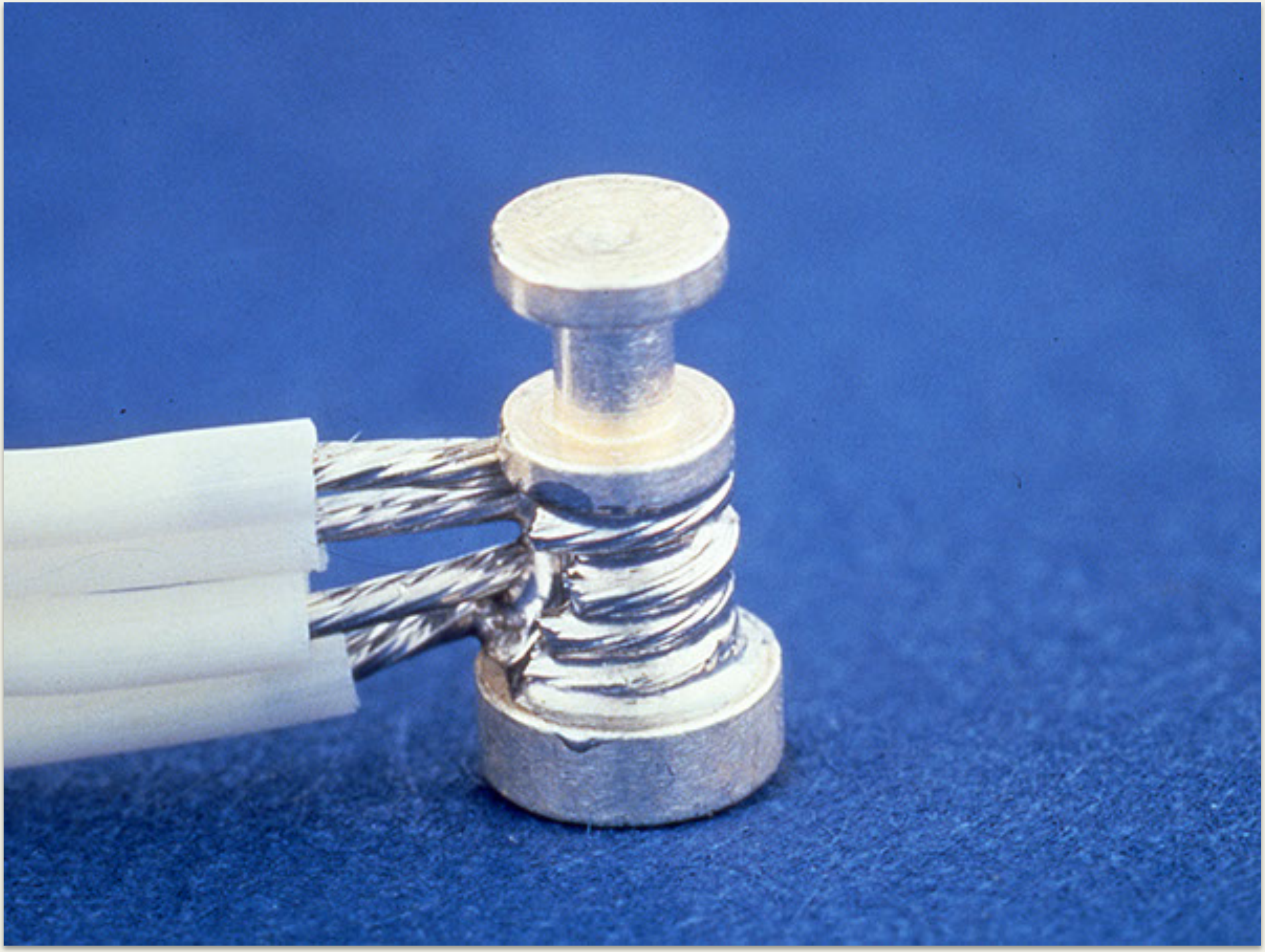

Terminal fill, all wires touching post - Accept

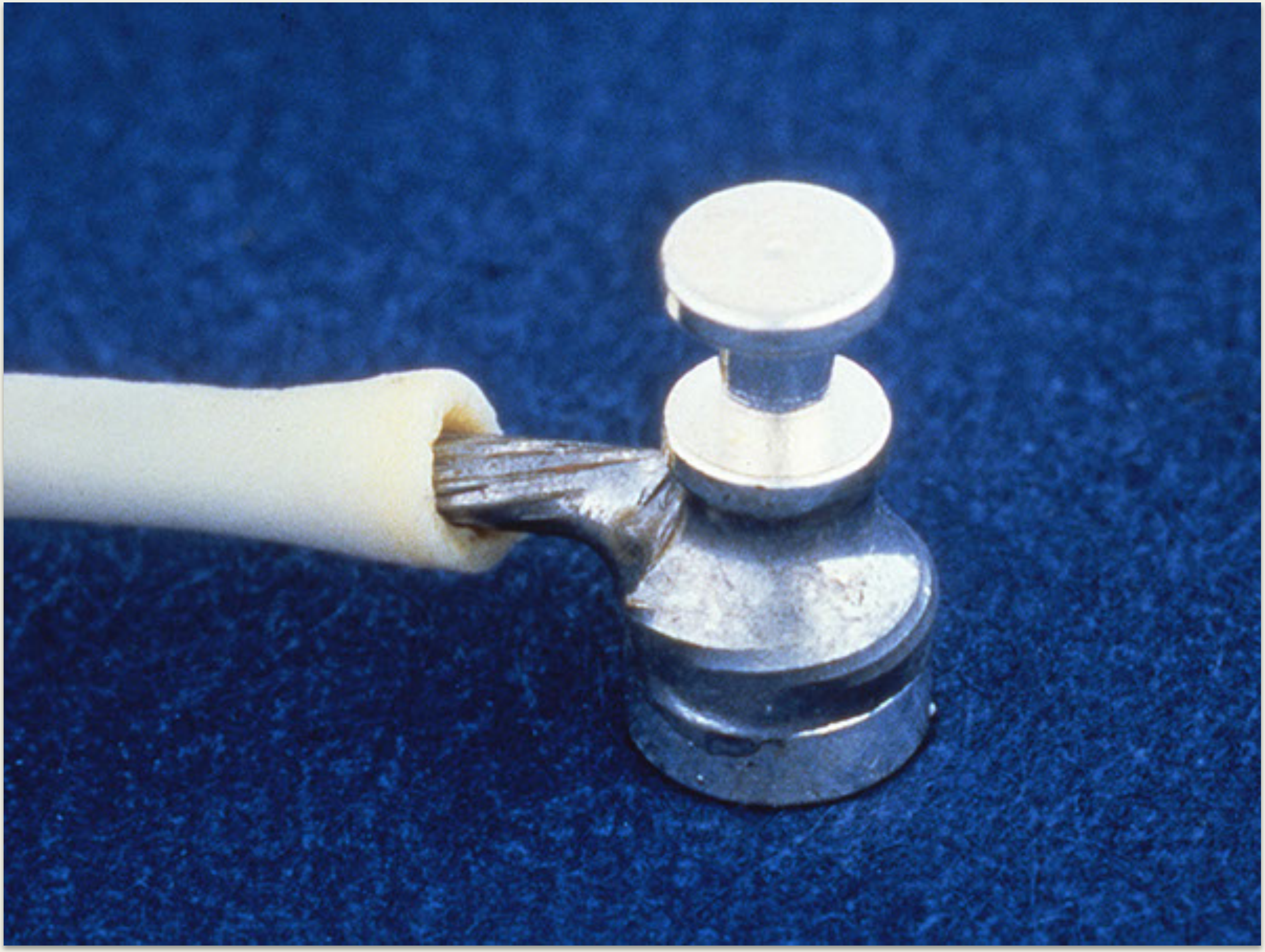

Excessive wicking - Reject

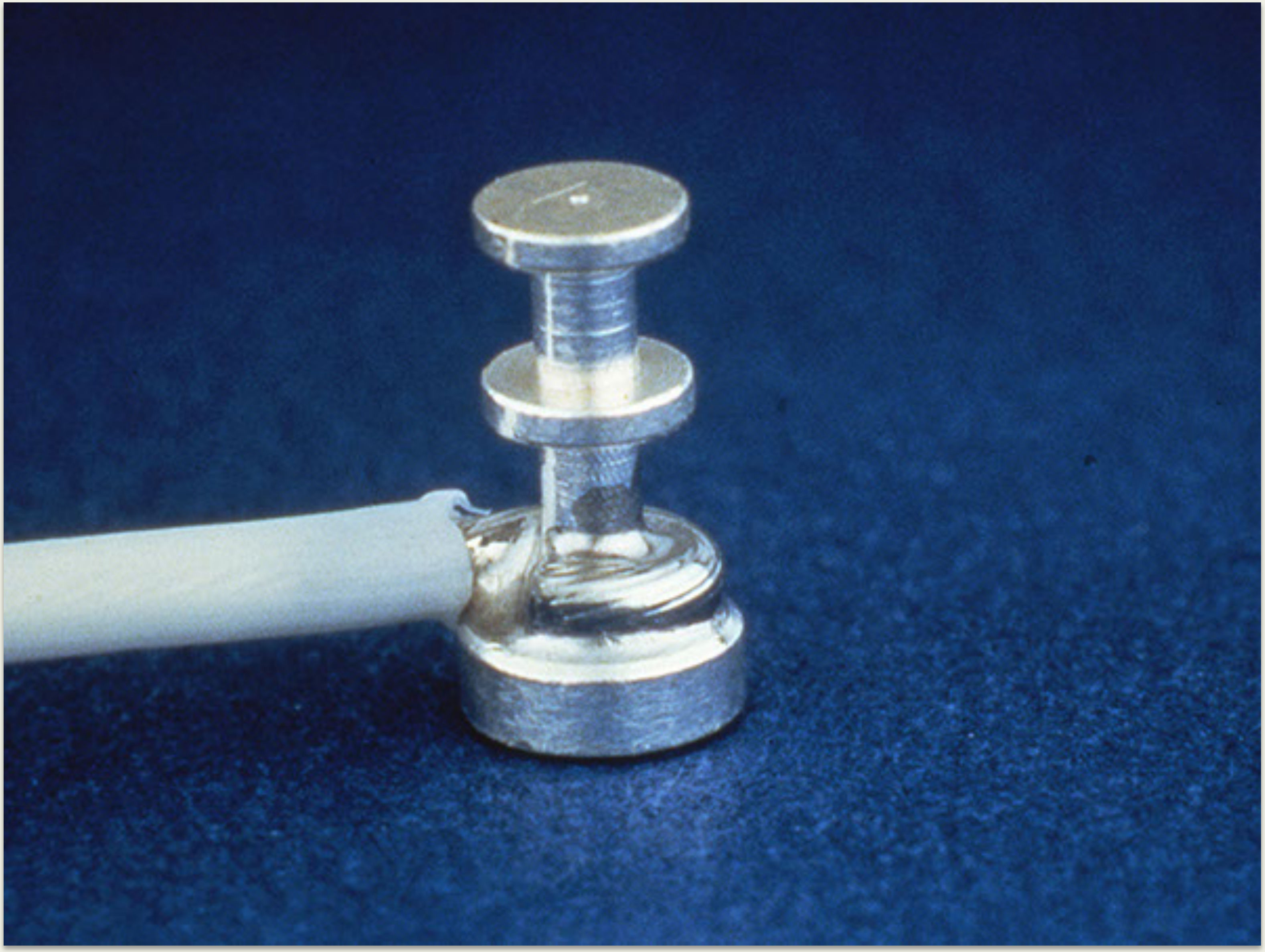

Minimum insulation clearance, possible contamination

Reject

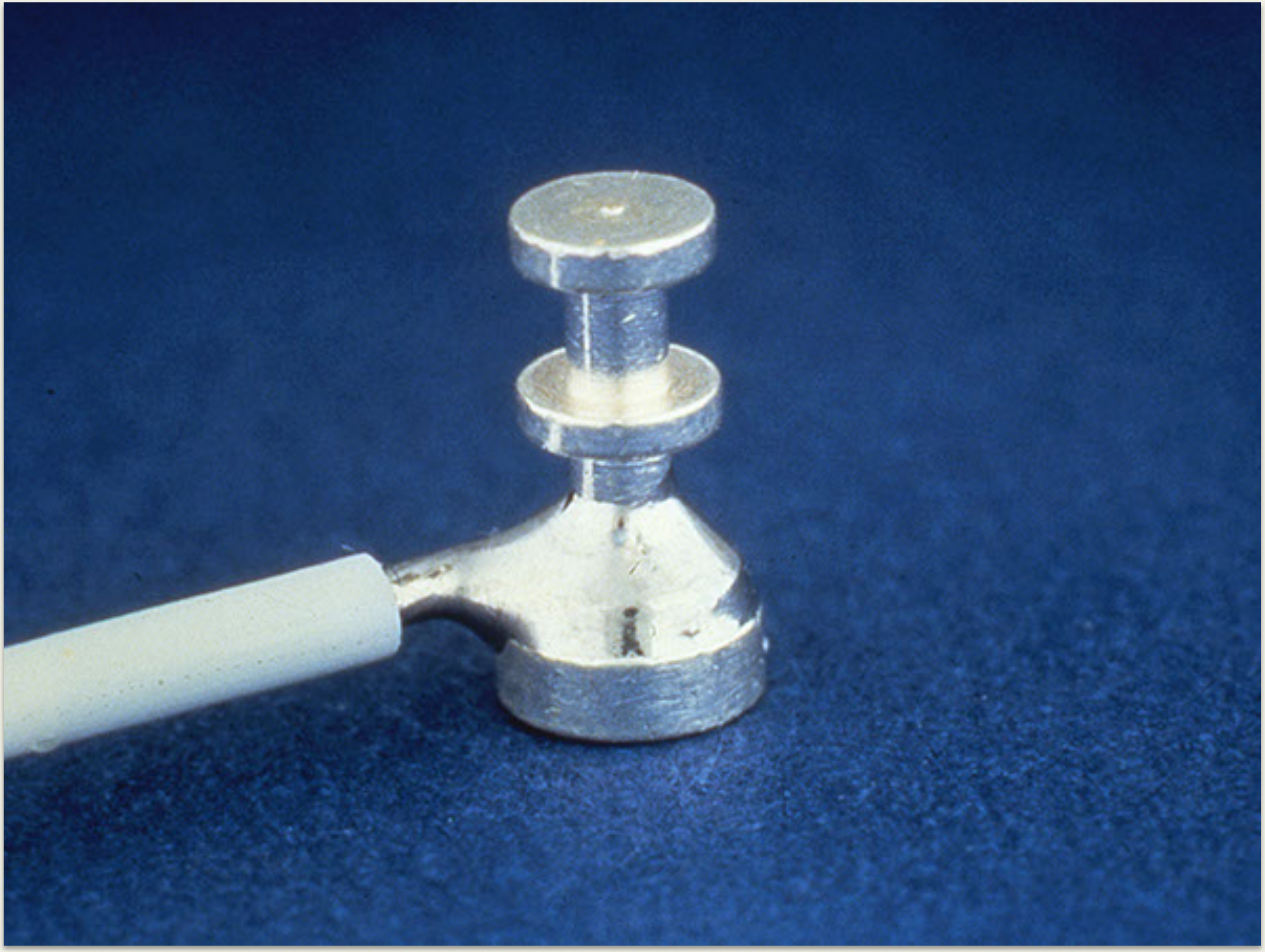

Excessive solder - Reject

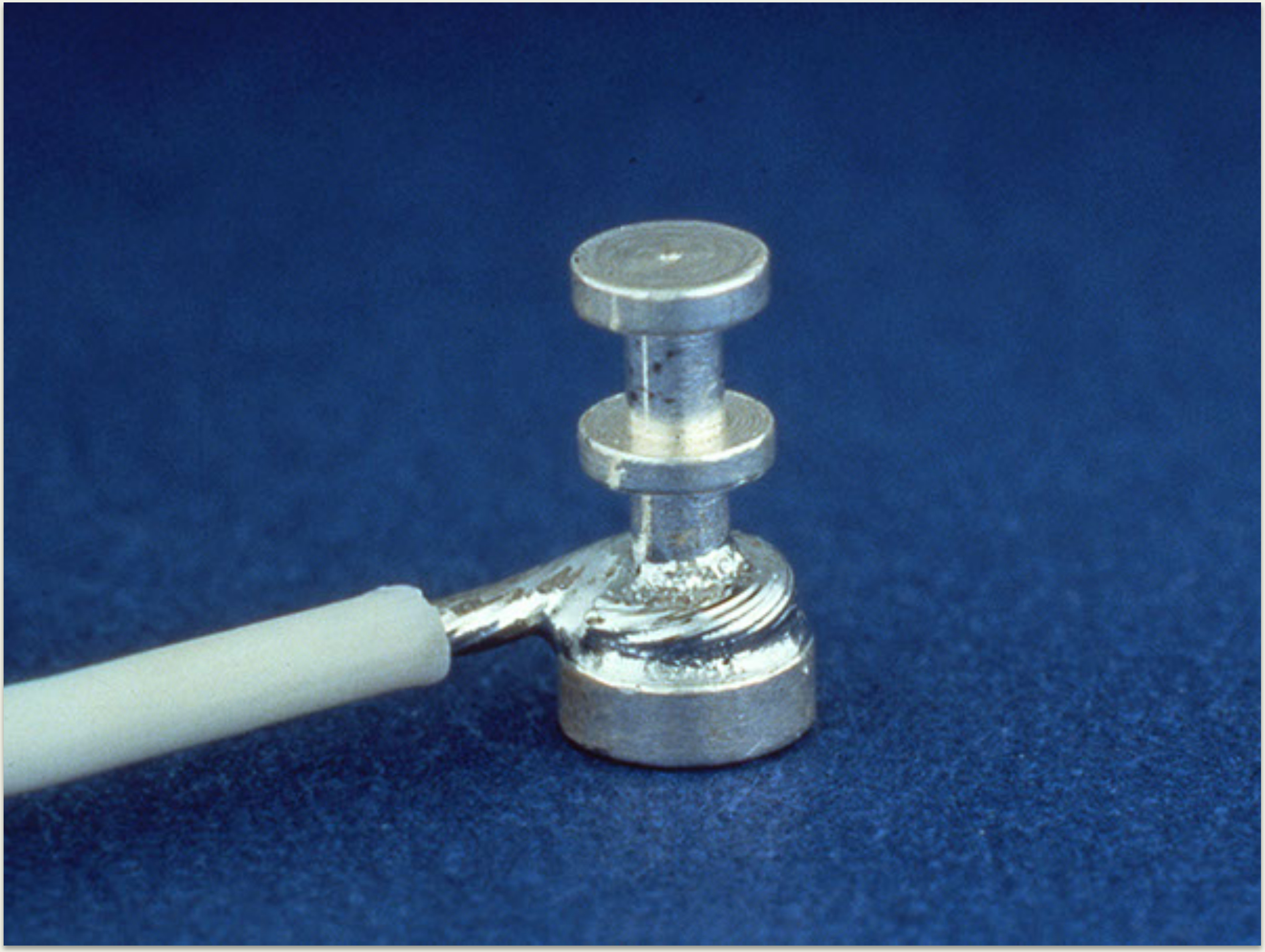

Grainy/Overheated - Reject

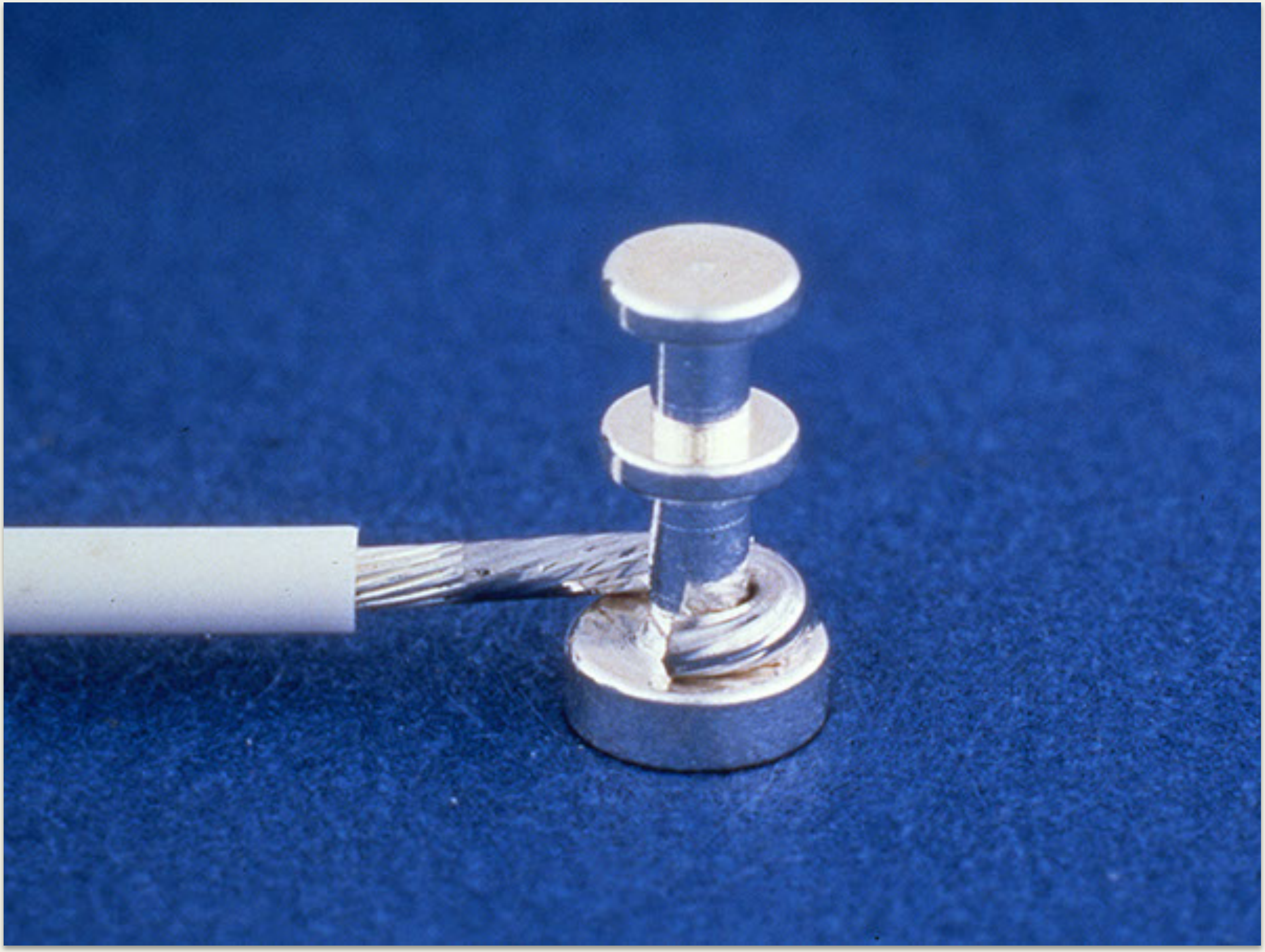

Dewetted - Reject

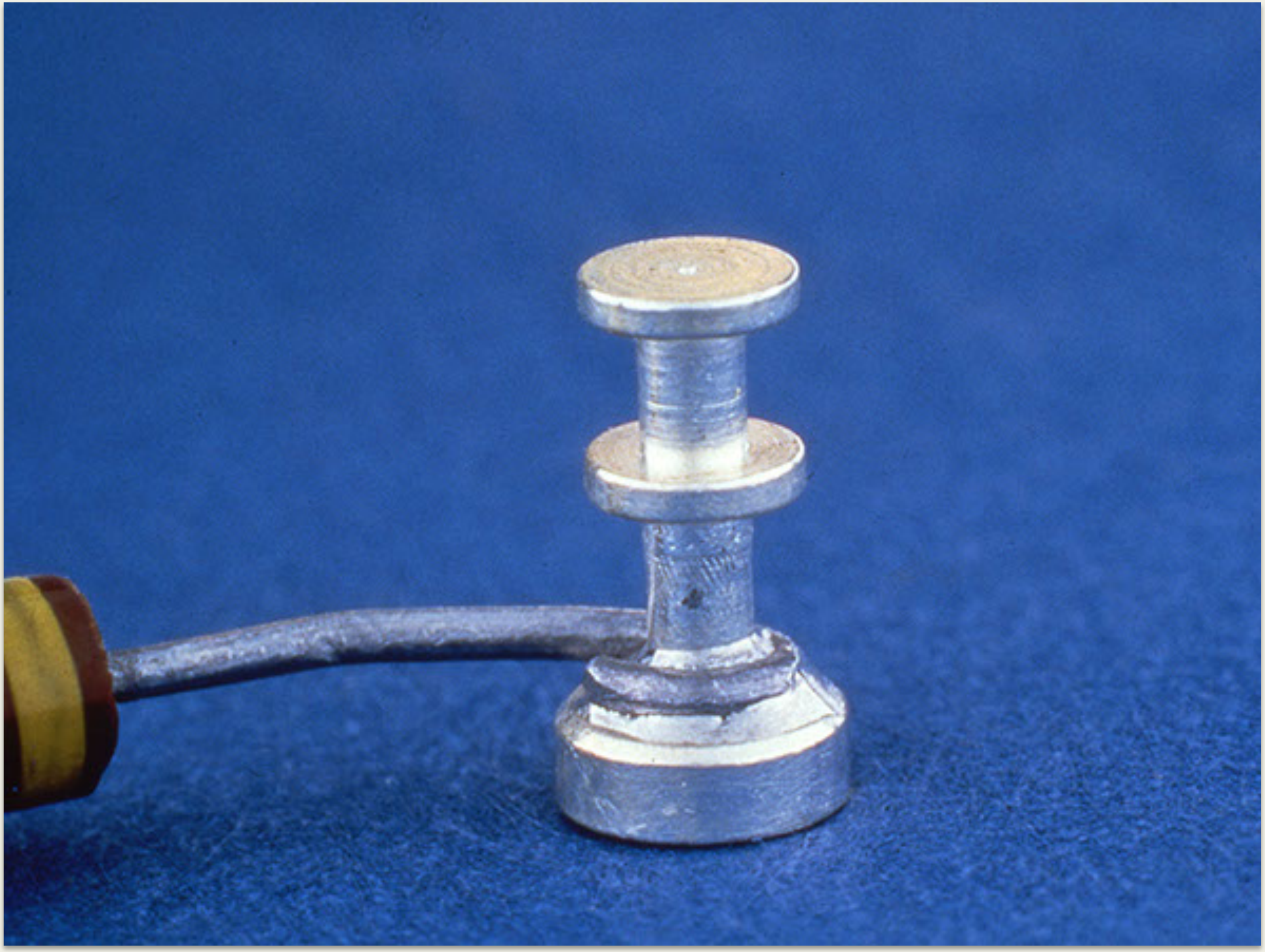

Dewetted - Reject

BIFURCATED TERMINAL

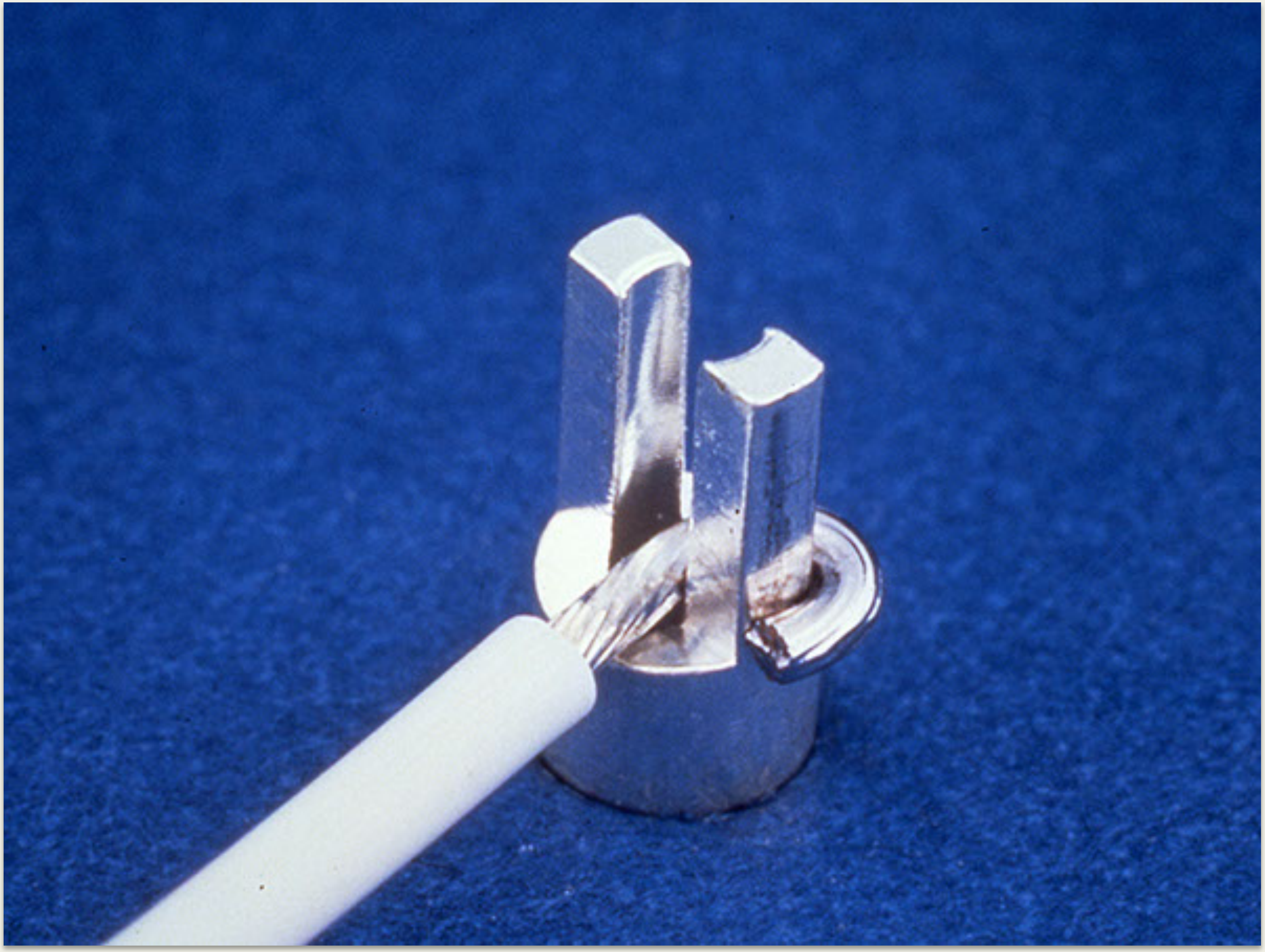

Preferred wrap (single) - Accept

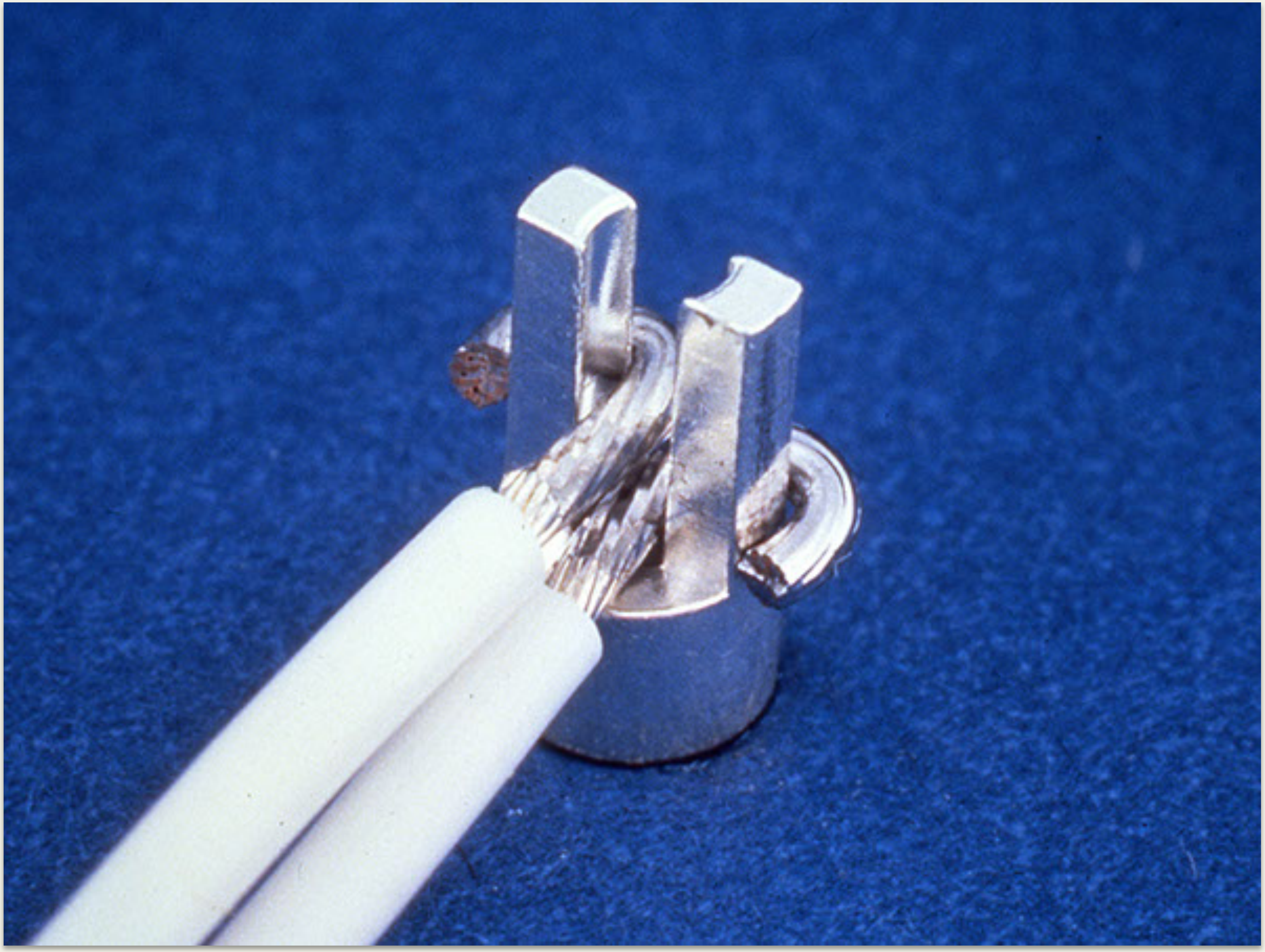

Preferred Wrap (double) - Accept

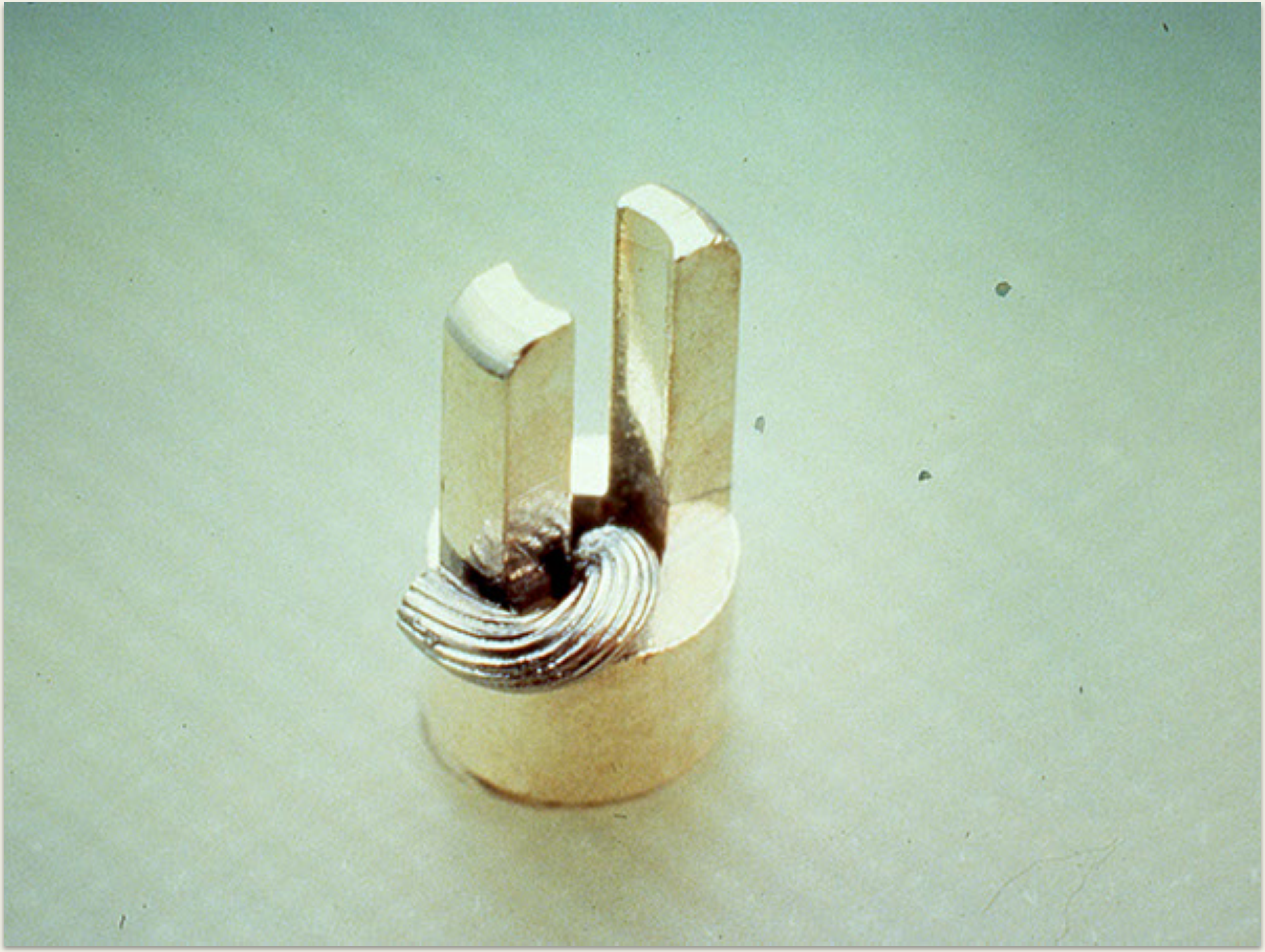

Preferred wrap (bottom route) - Accept

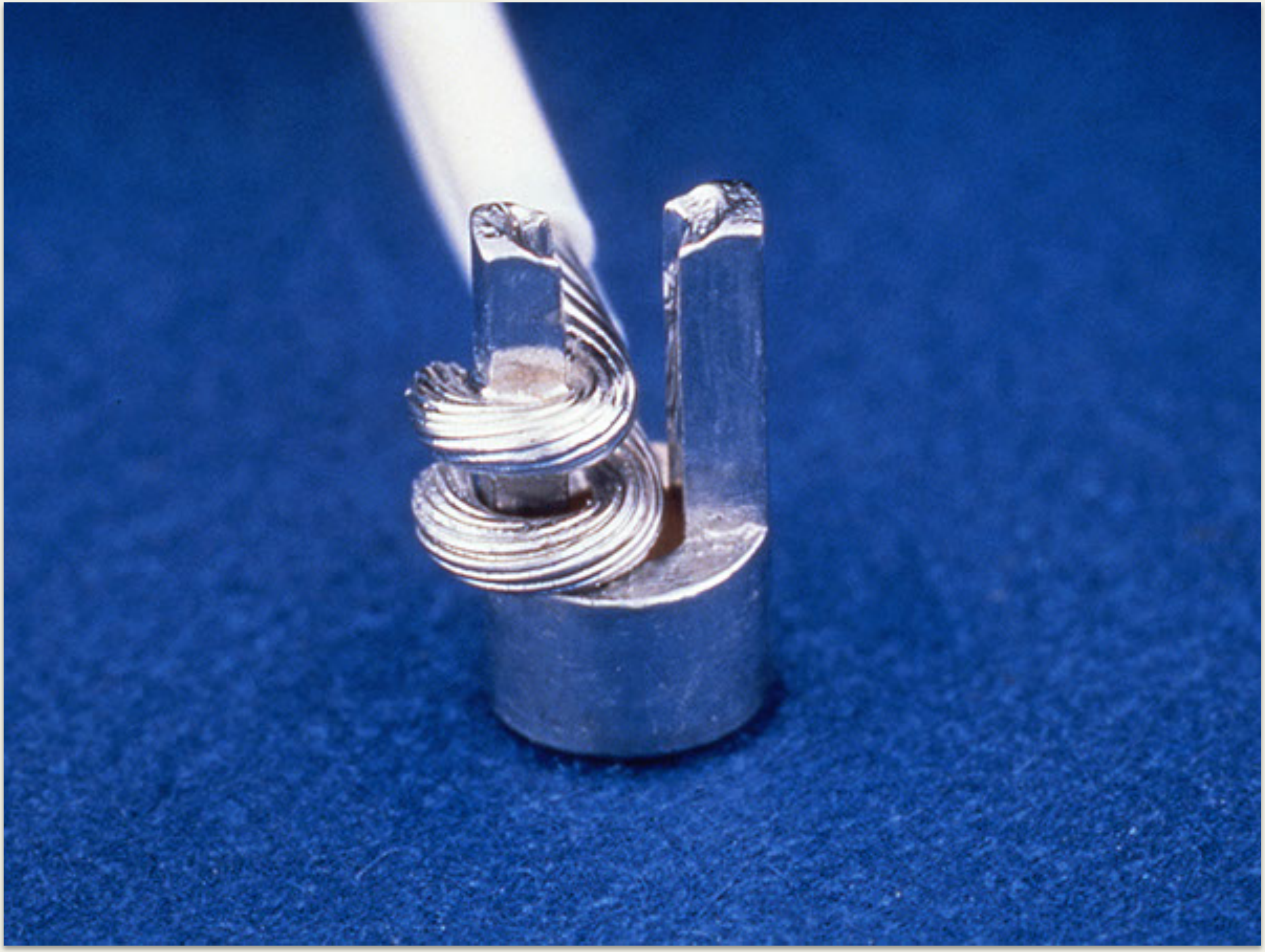

Improper mechanical wrap - Reject

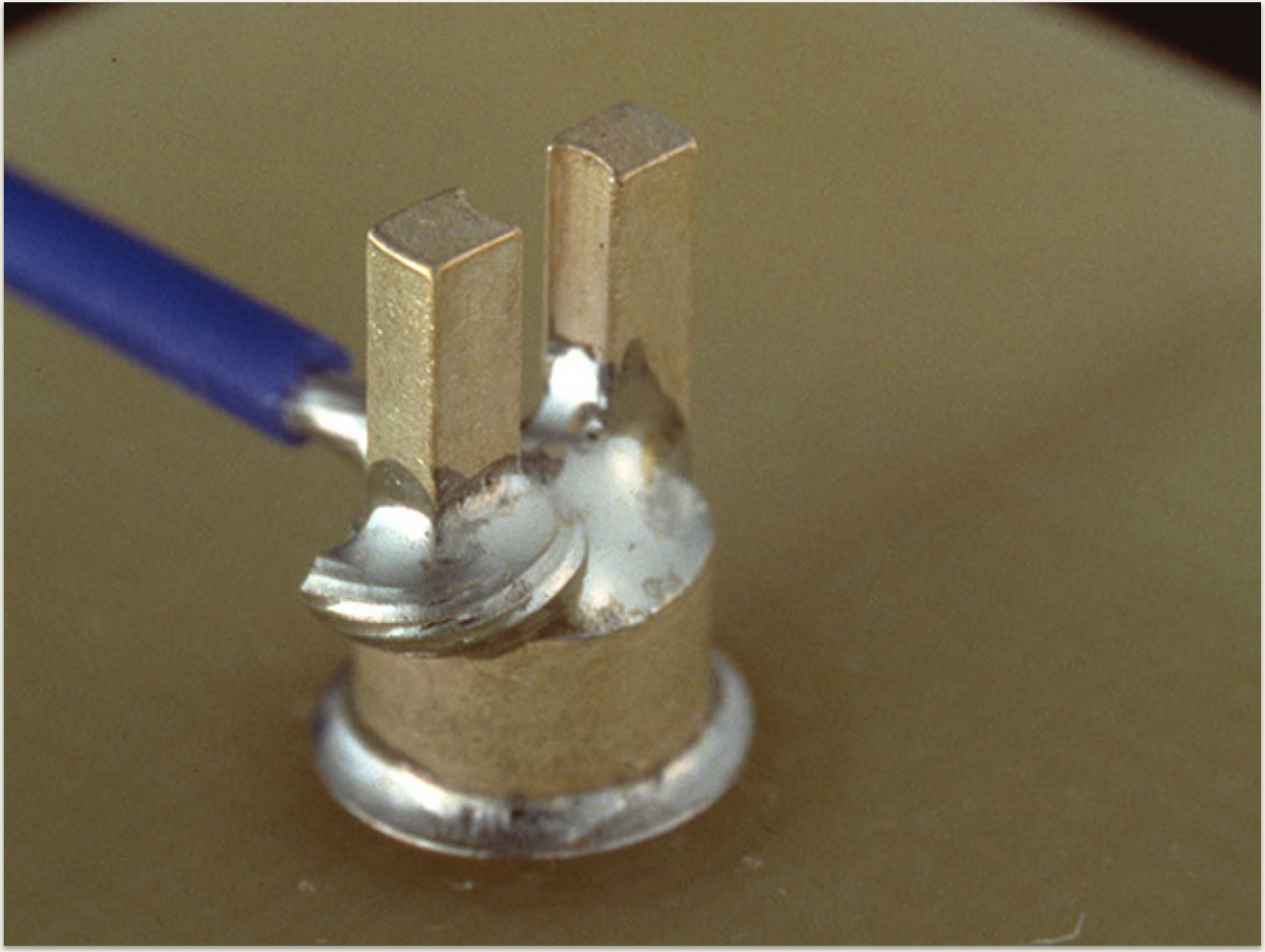

Excessive solder (should see strand contour) - Reject

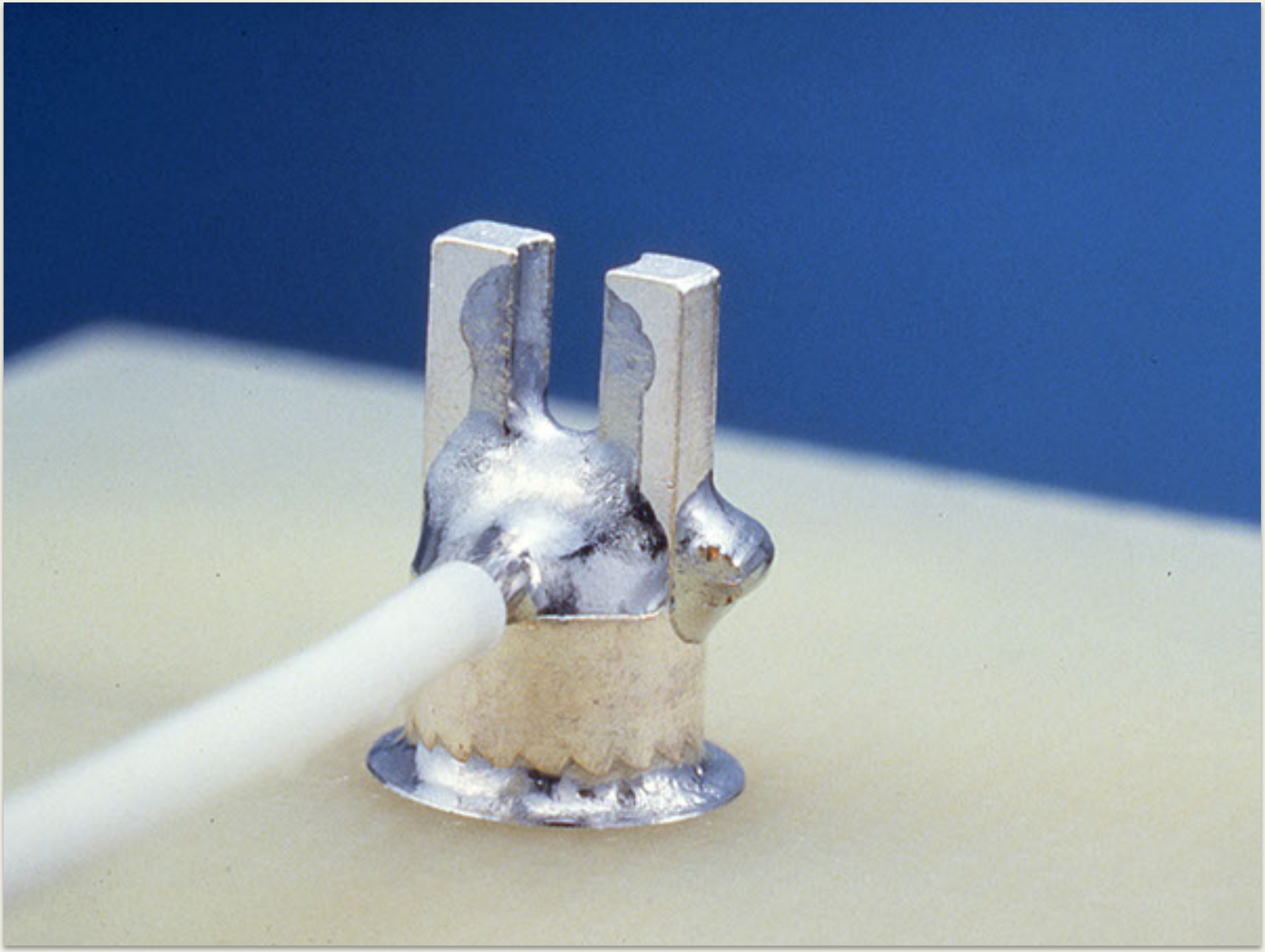

Excessive solder - Reject

SOLDER CUP TERMINAL

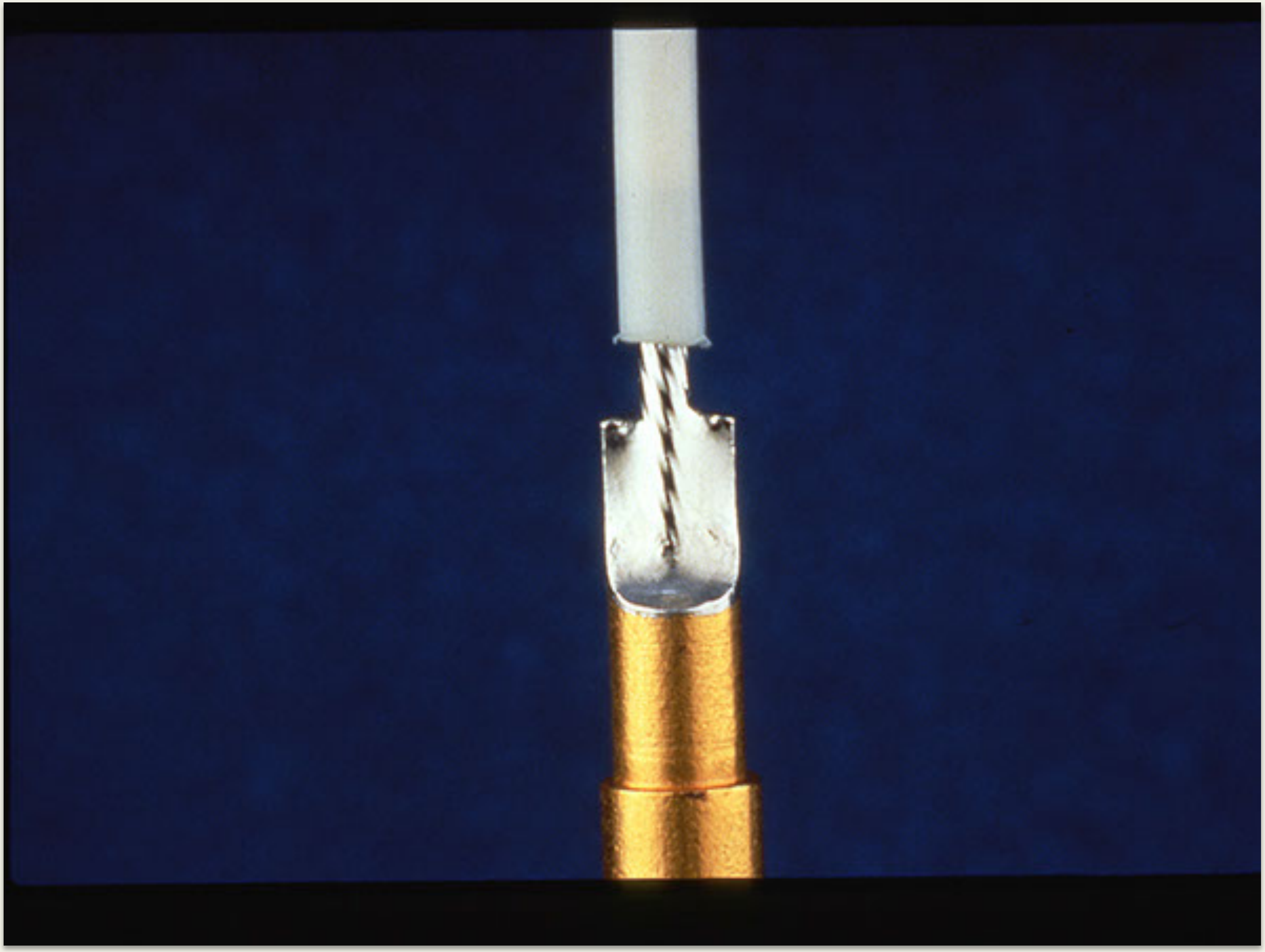

Preferred solder - Accept

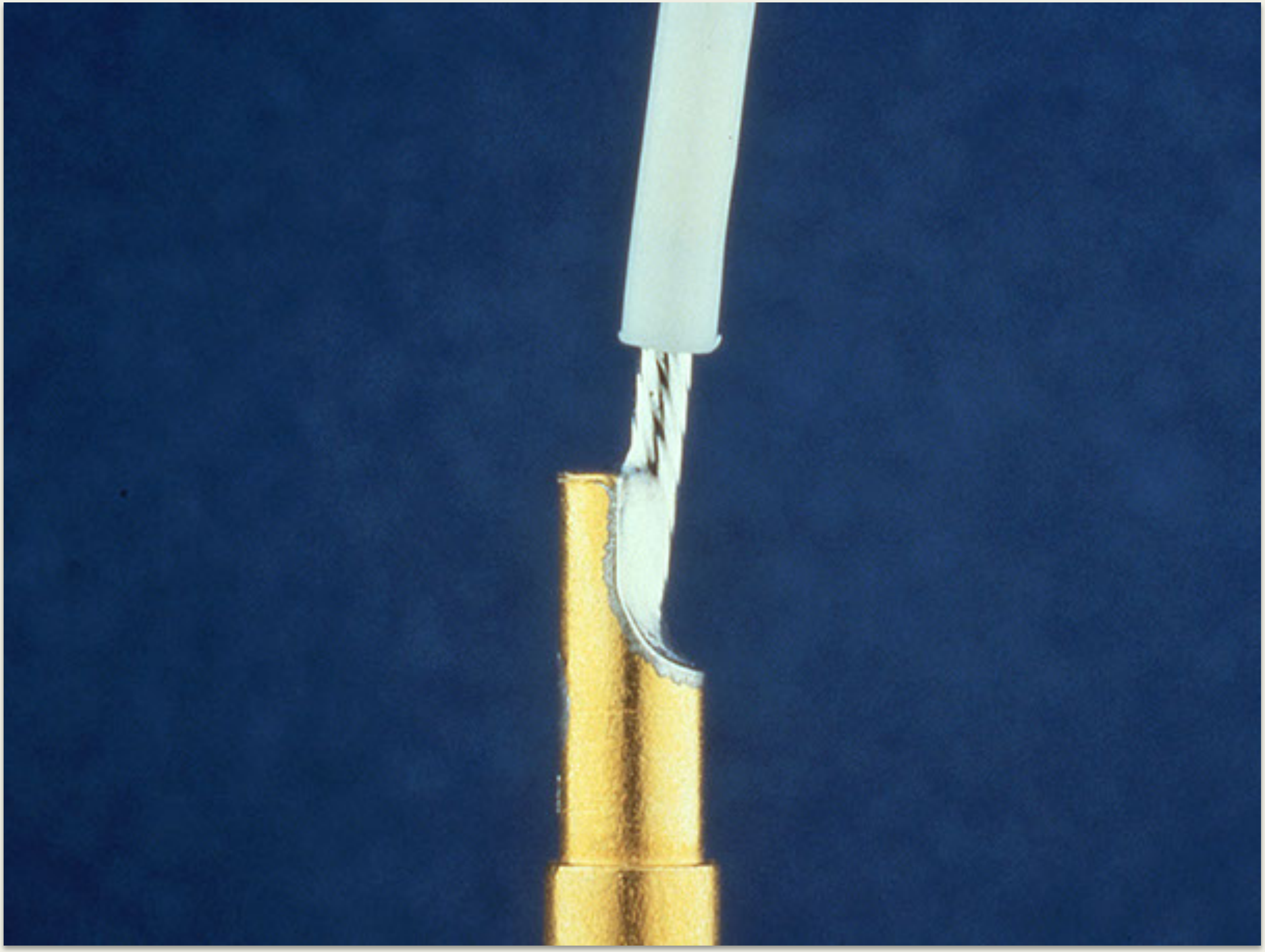

Improper seating to back of cup - Reject

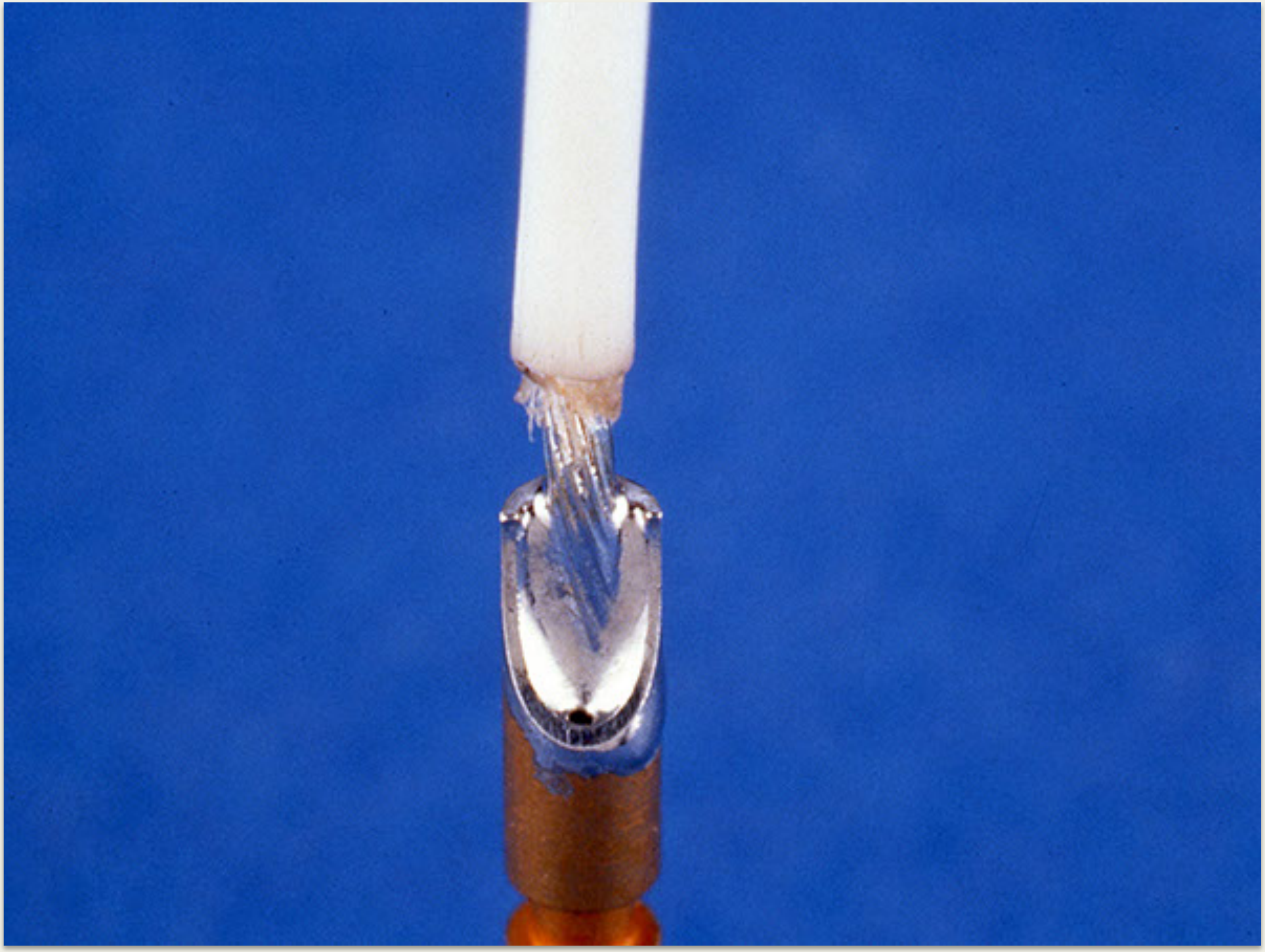

Void, pin hole - Reject

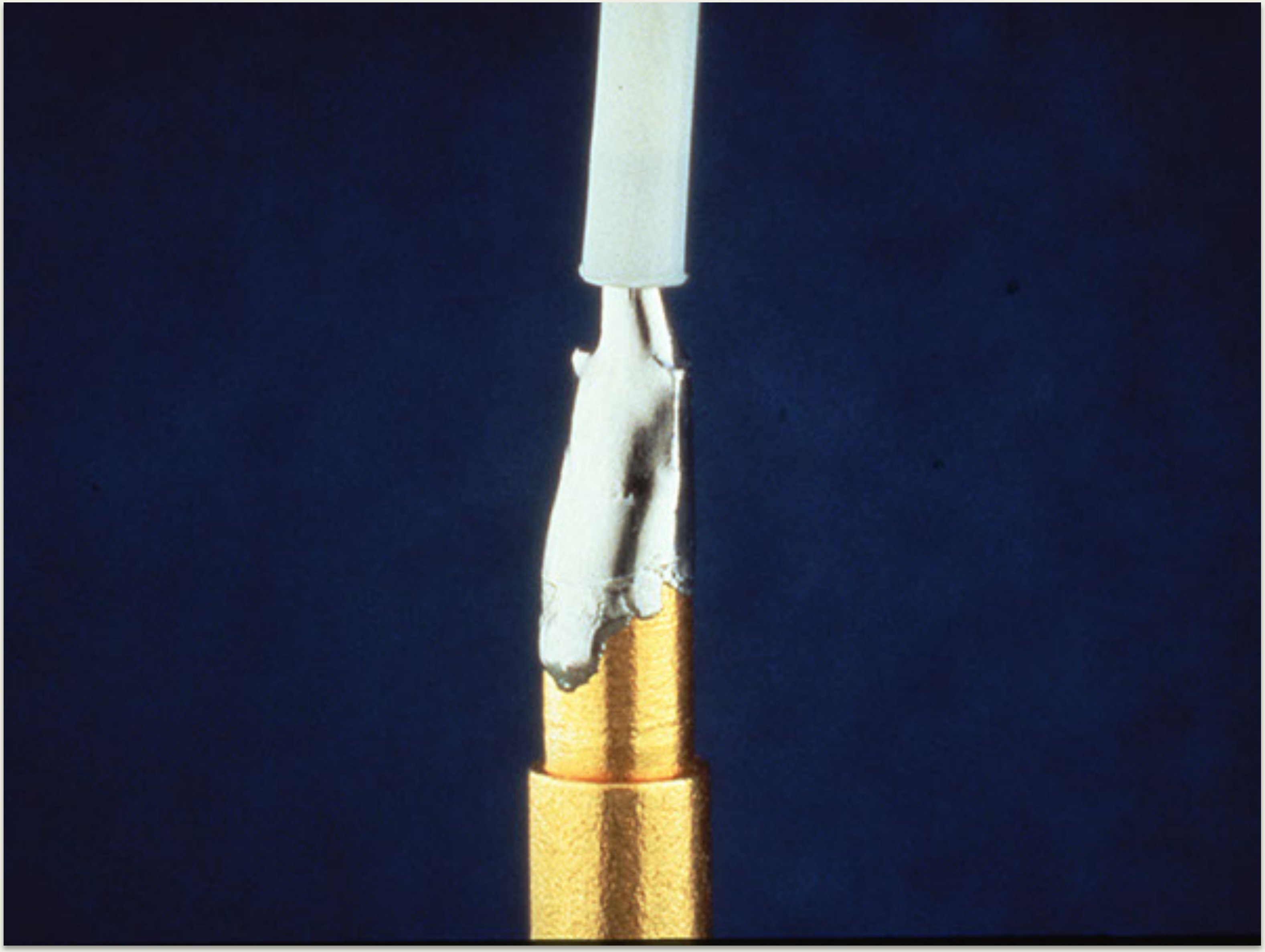

Excessive solder/spillage - Reject

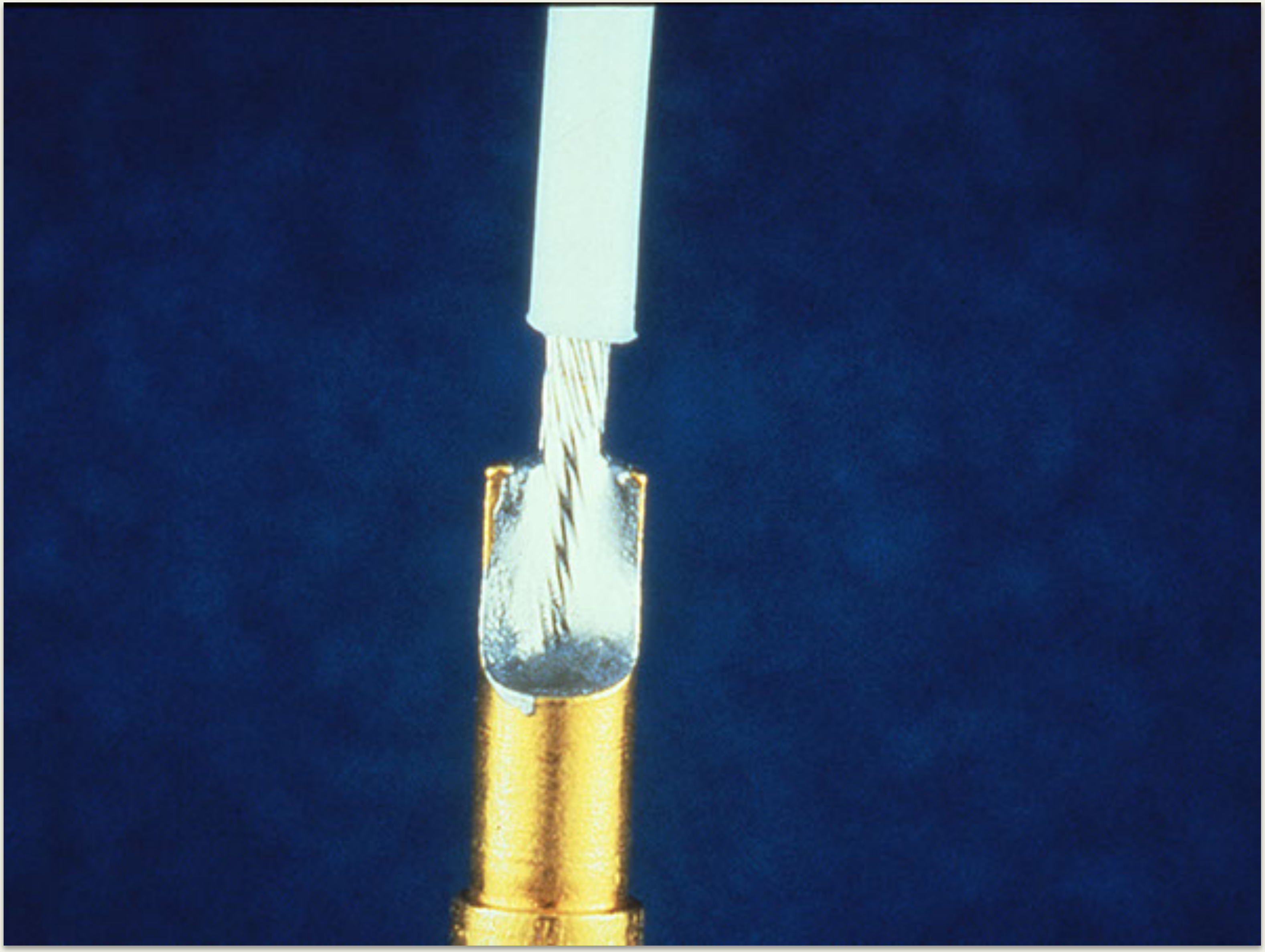

Gold embrittlement (amalgamation) - Reject

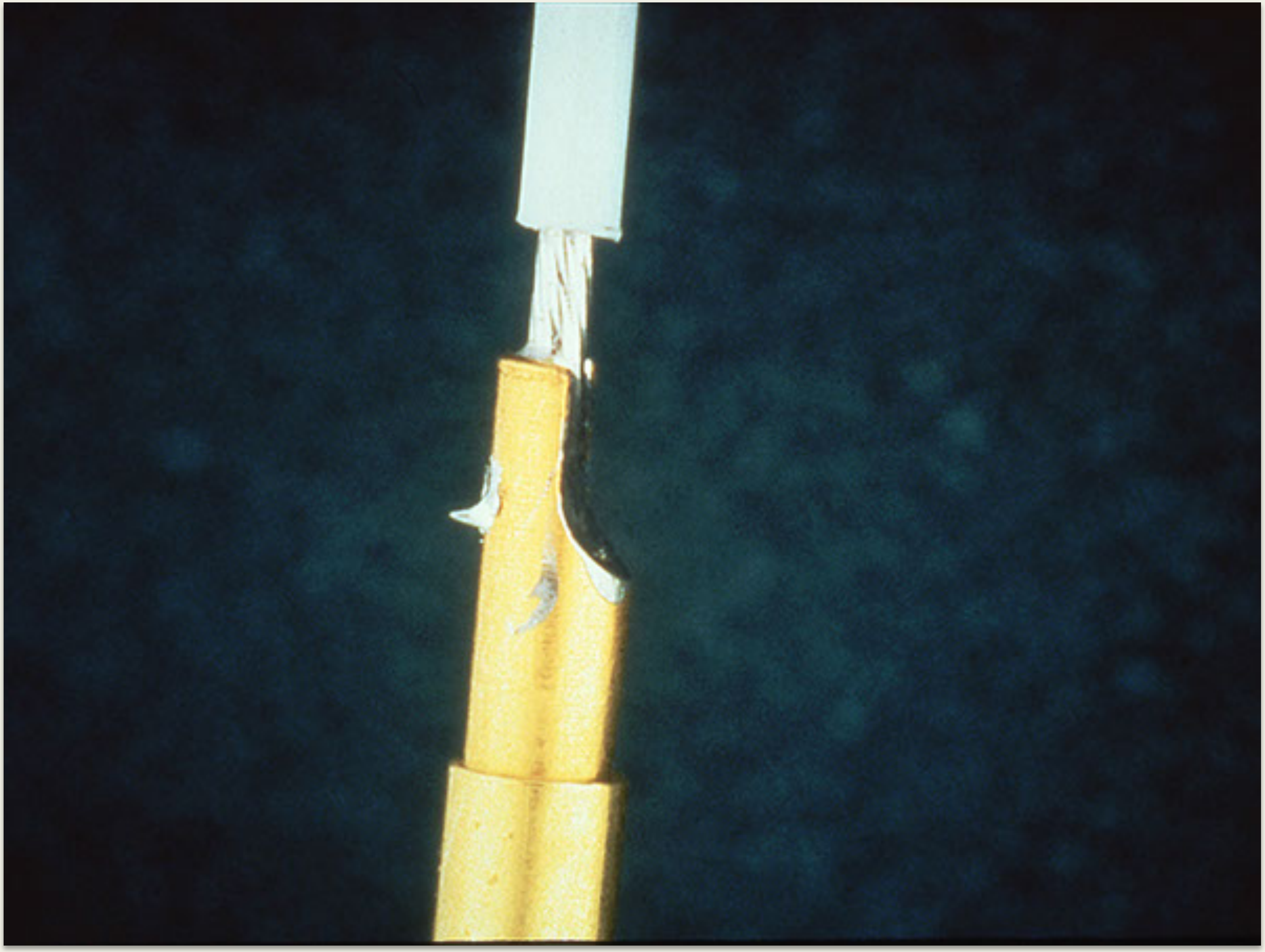

Solder spike - Reject

PRINTED
WIRING BOARD

TERMINAL SWAGING

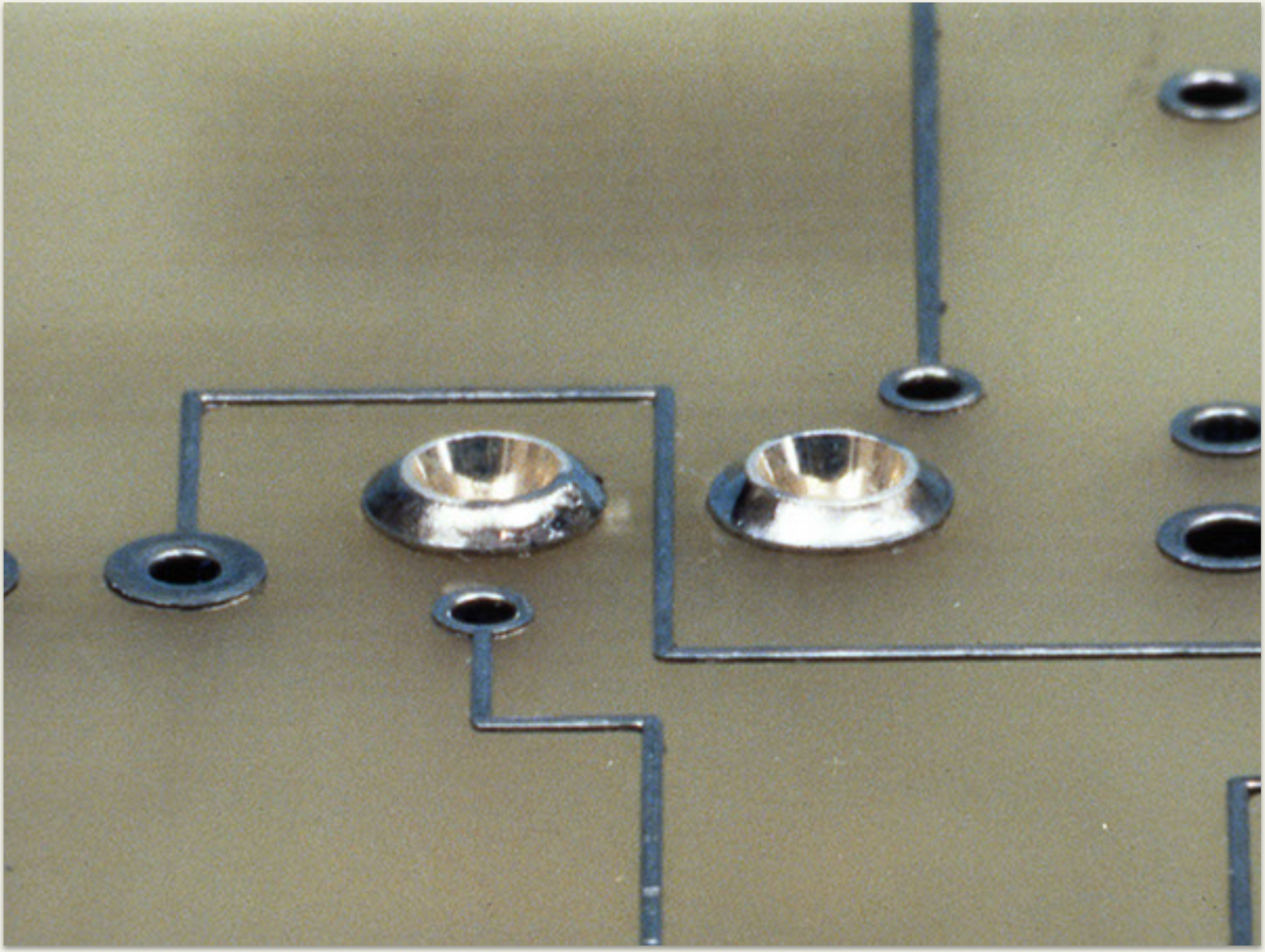

Left side, excessive solder/void/measle – Reject
Right side, preferred solder - Accept

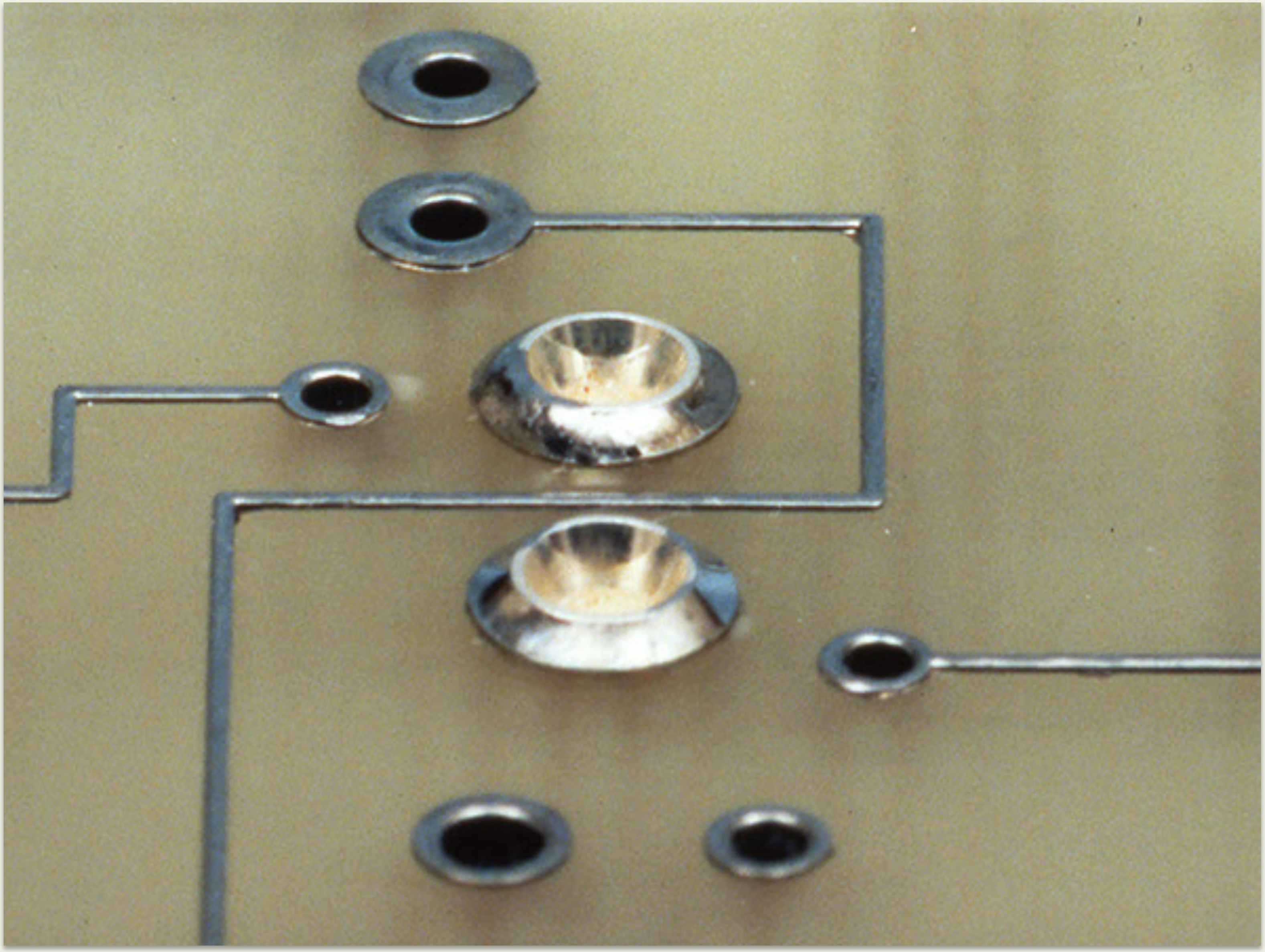

Top, excessive solder/measle – Reject
Bottom, preferred solder - Accept

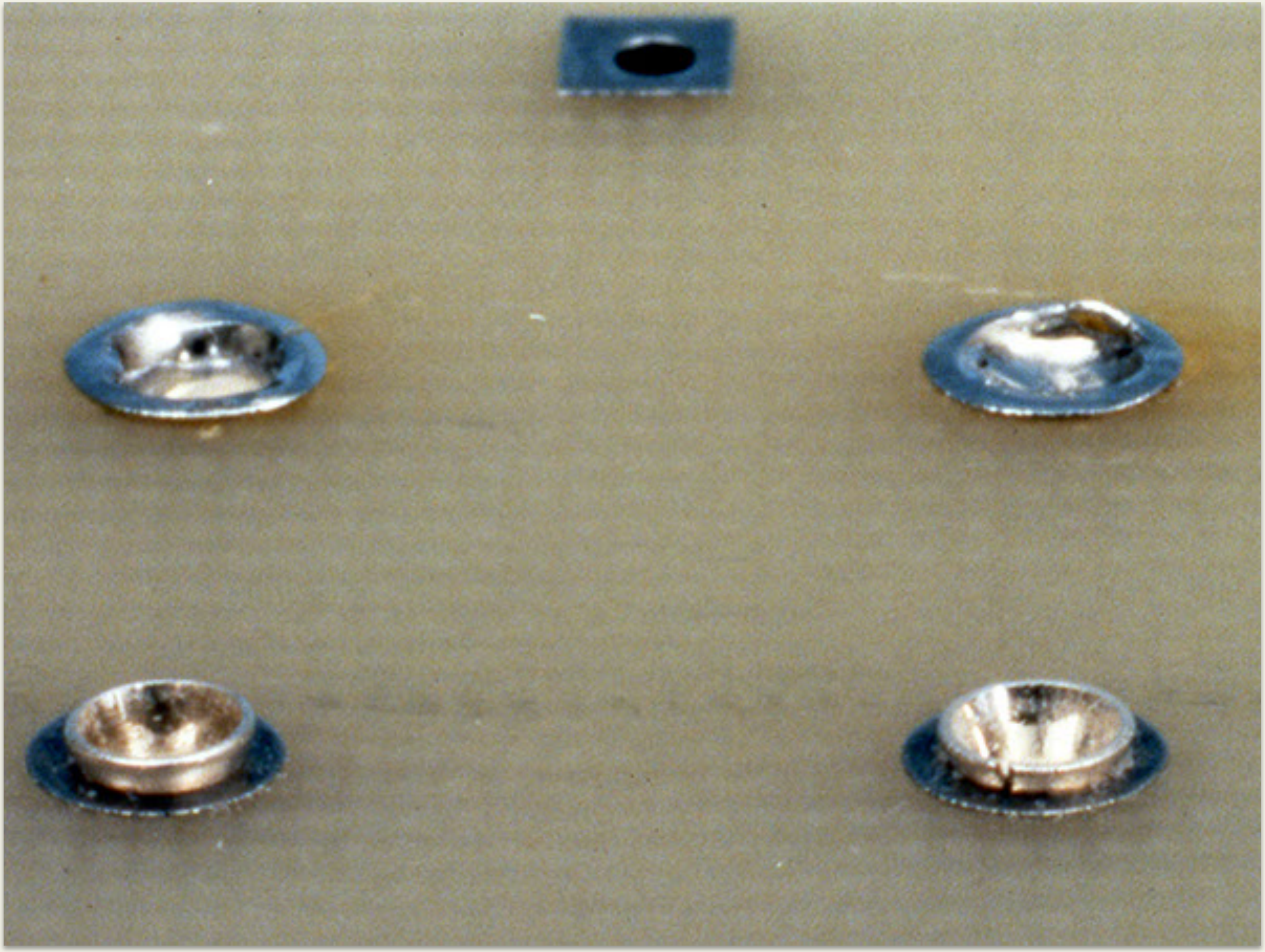

Top left, bad swage, top right, bad swage, burned – Reject
Bottom left, good swage, bottom right, radial split - Accept

AXIAL LEAD MOUNTING

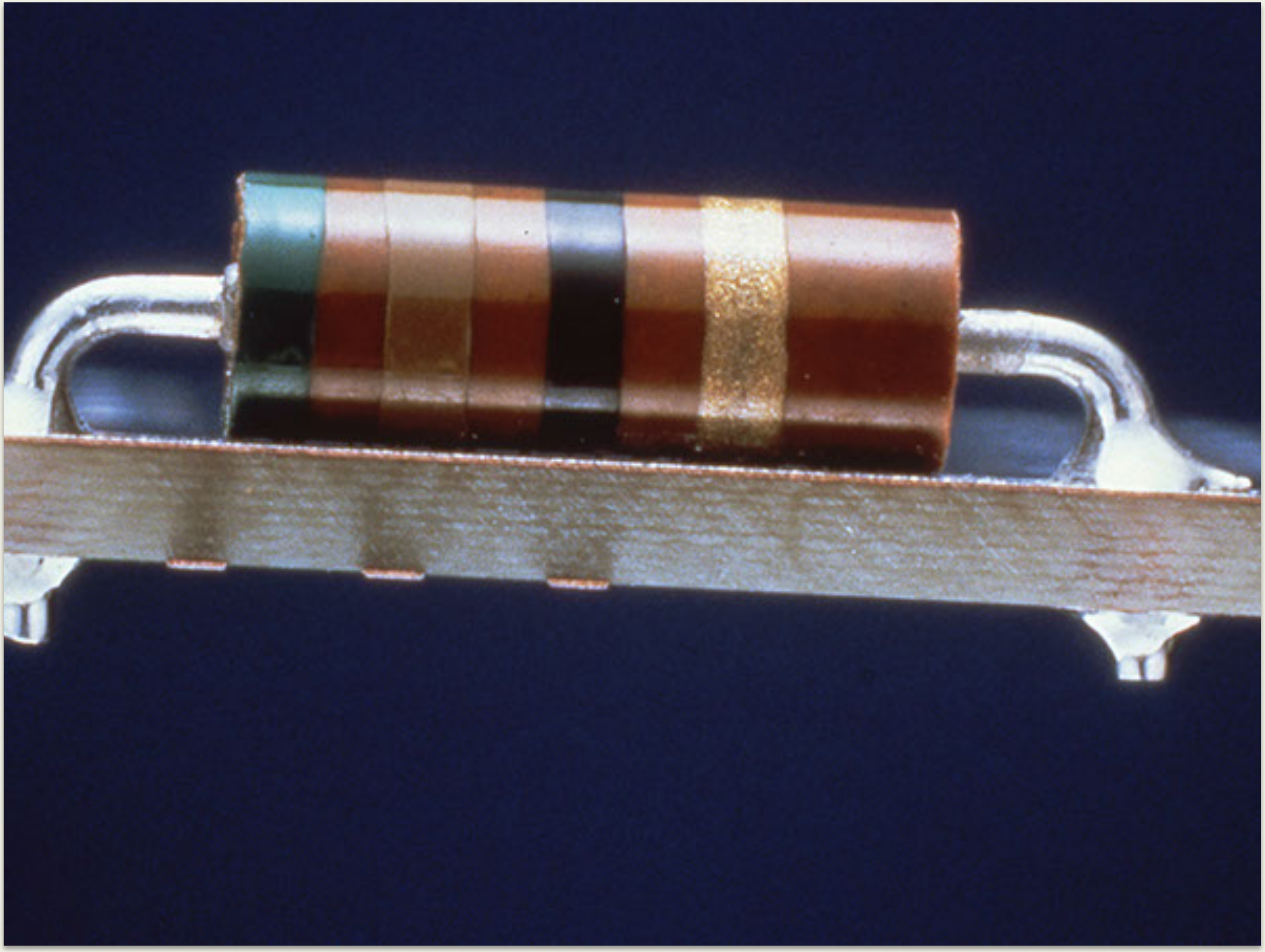

Axial component mounting - Accept

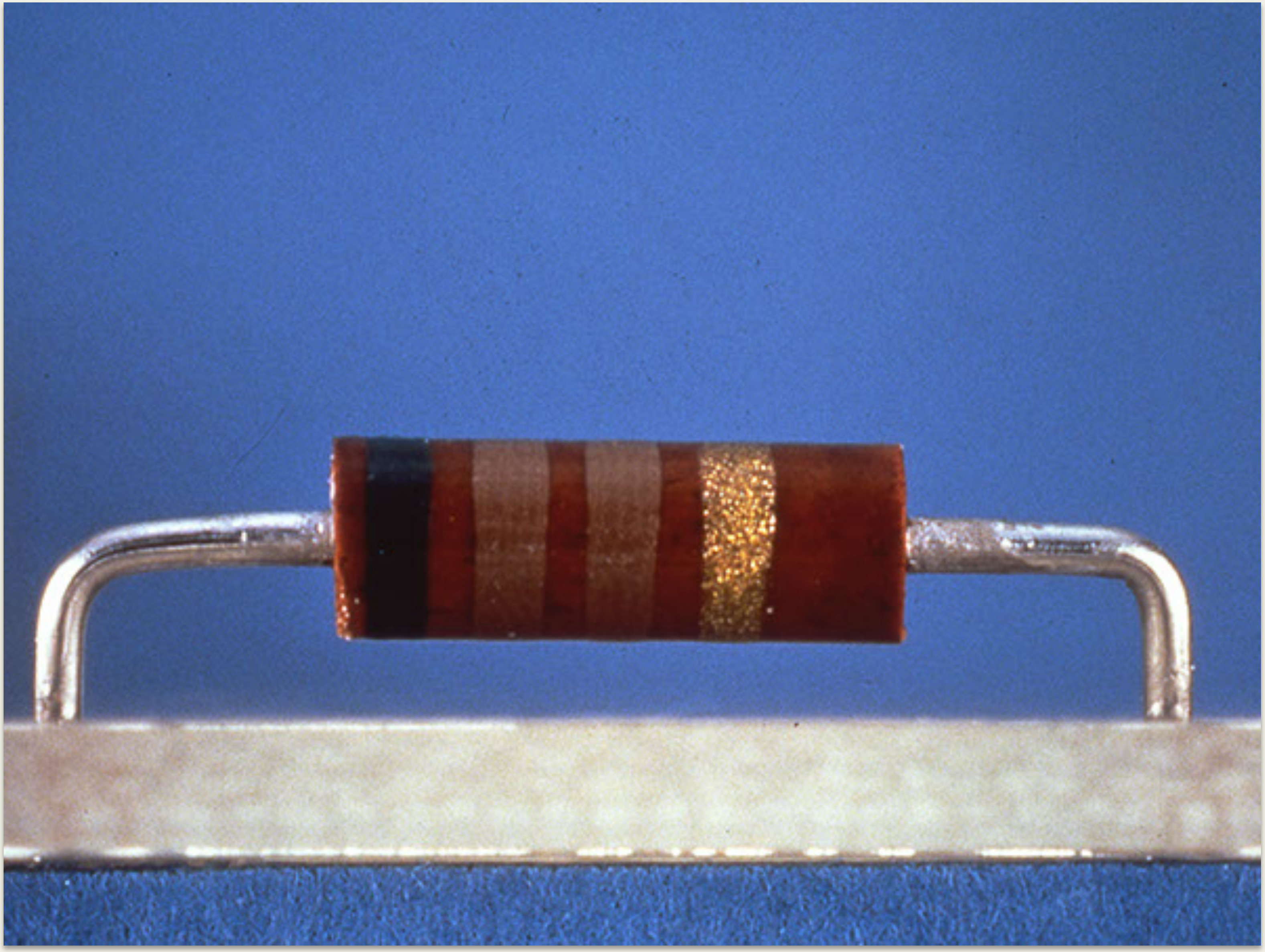

Axial component not seated against board - Reject

STUD TERMINATION

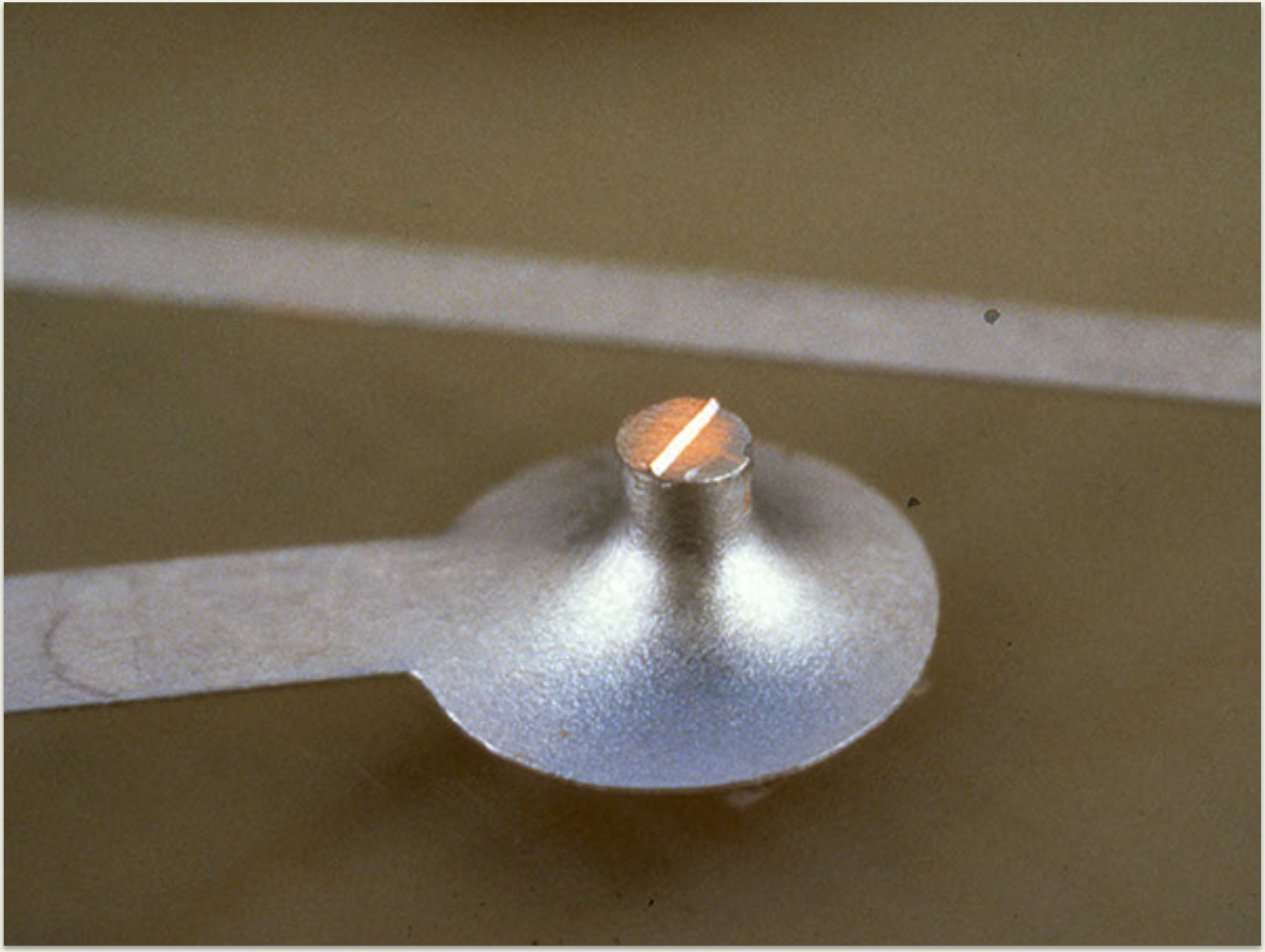

Exposed bare copper - Reject

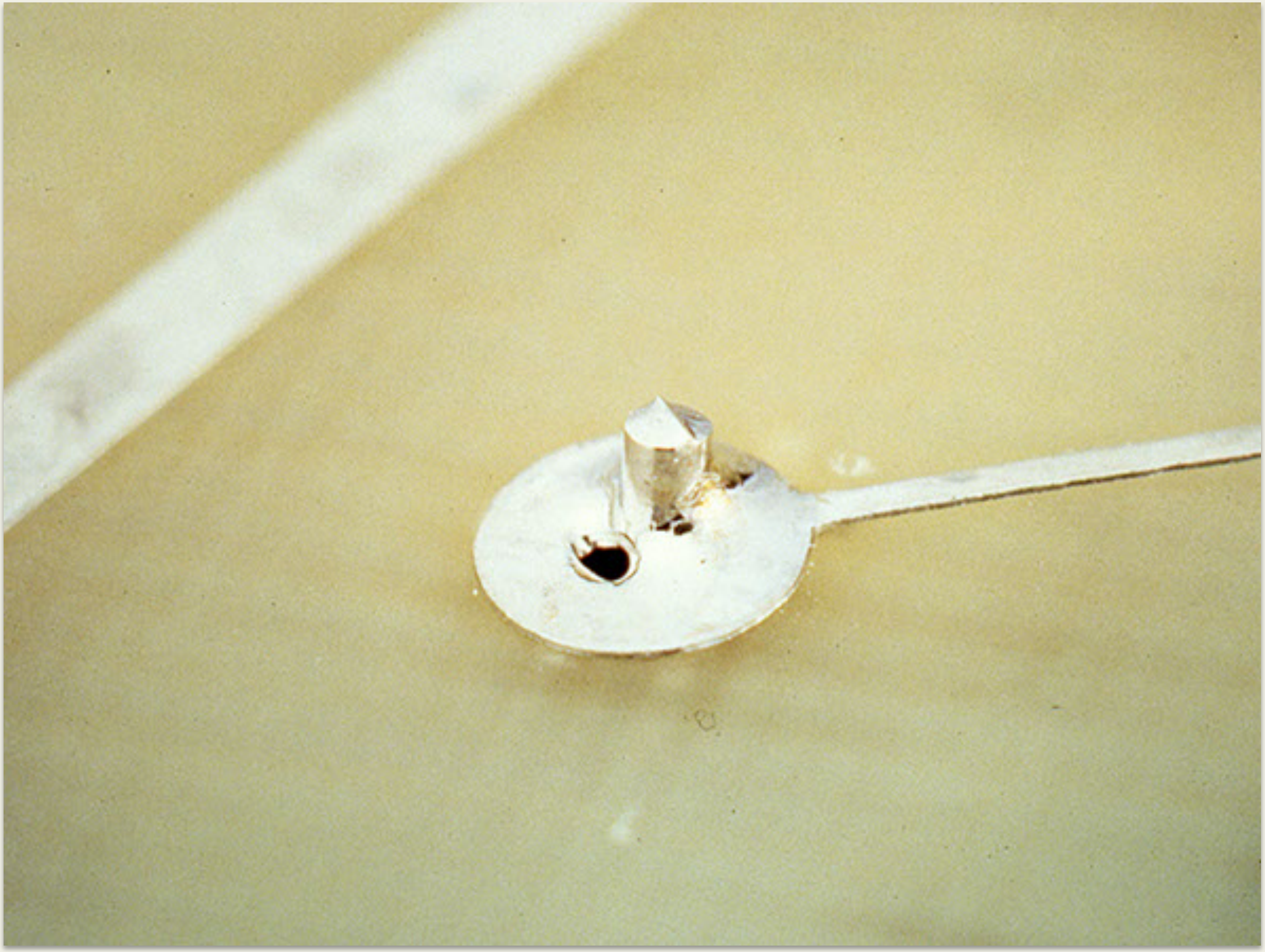

Blowhole - Reject

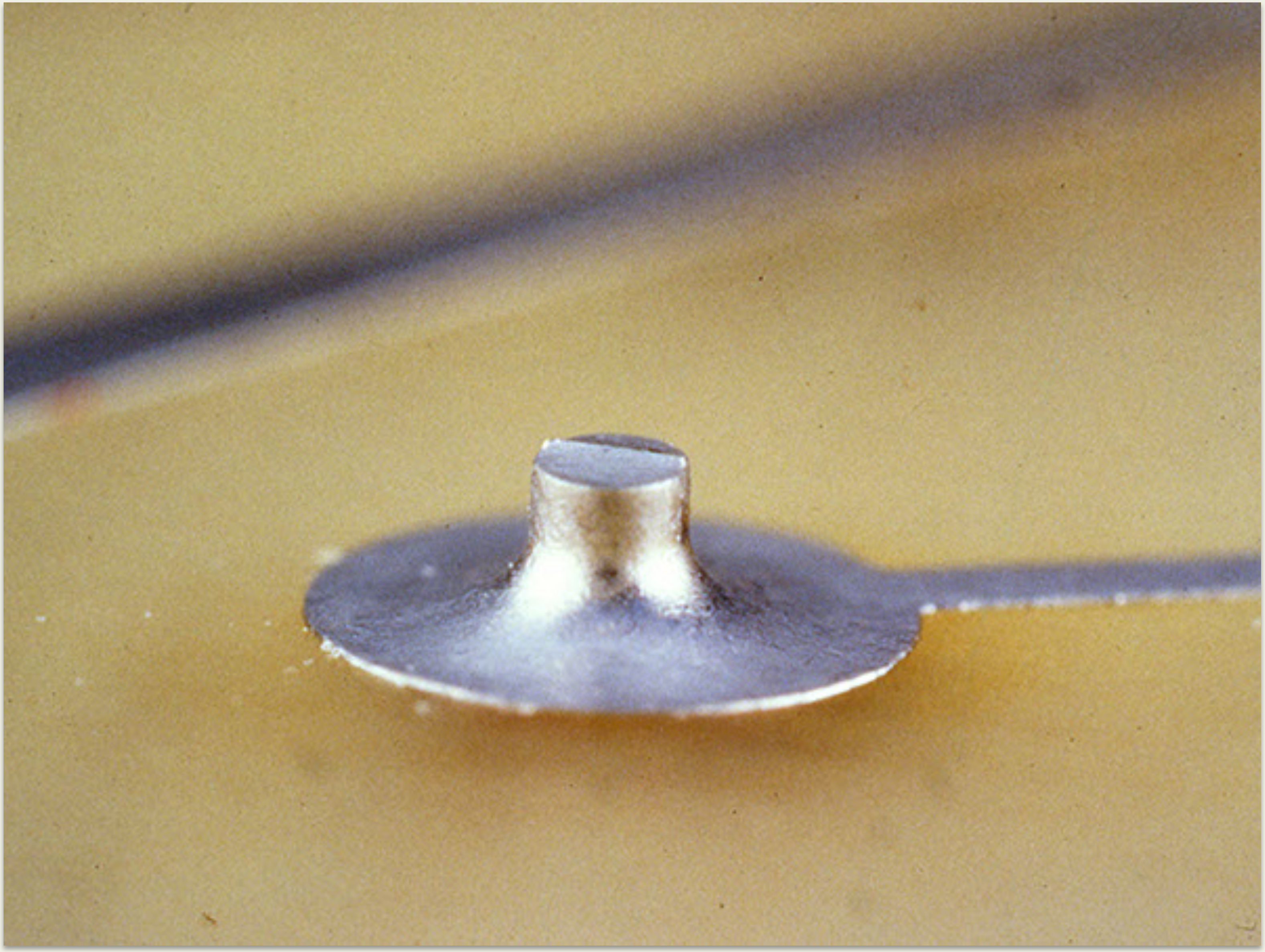

Preferred solder - Accept

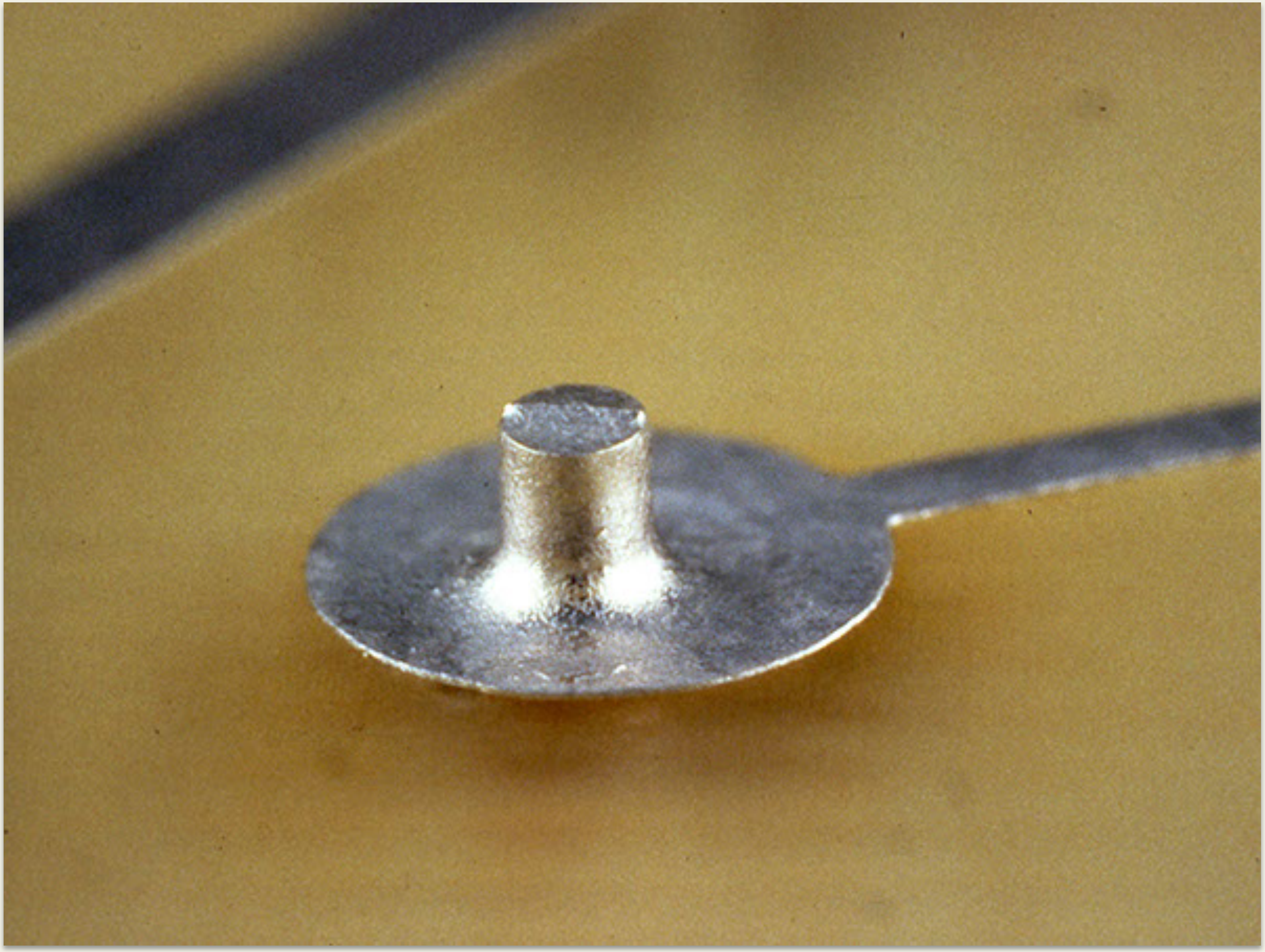

Minimum solder - Accept

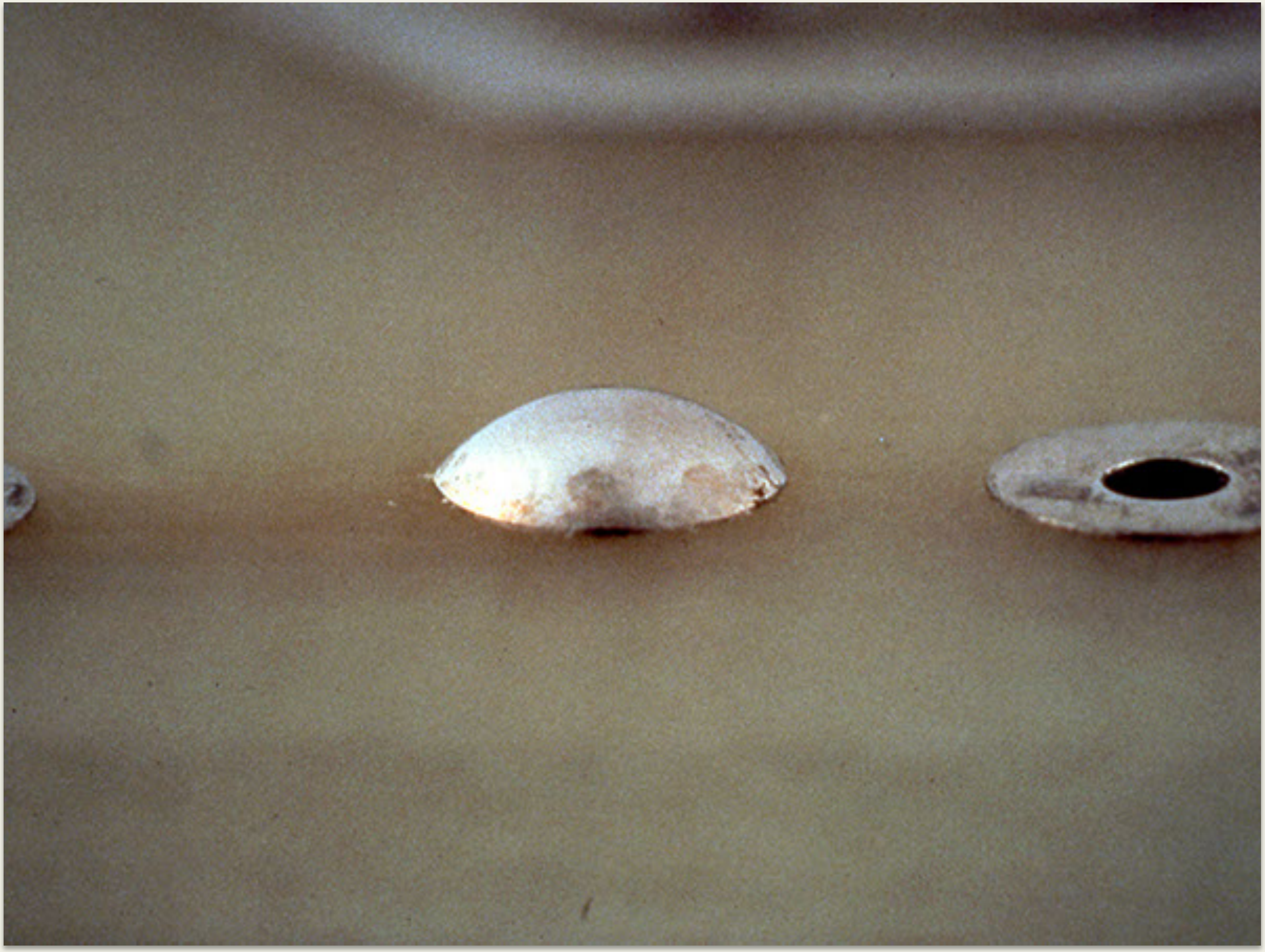

Excessive solder - Reject

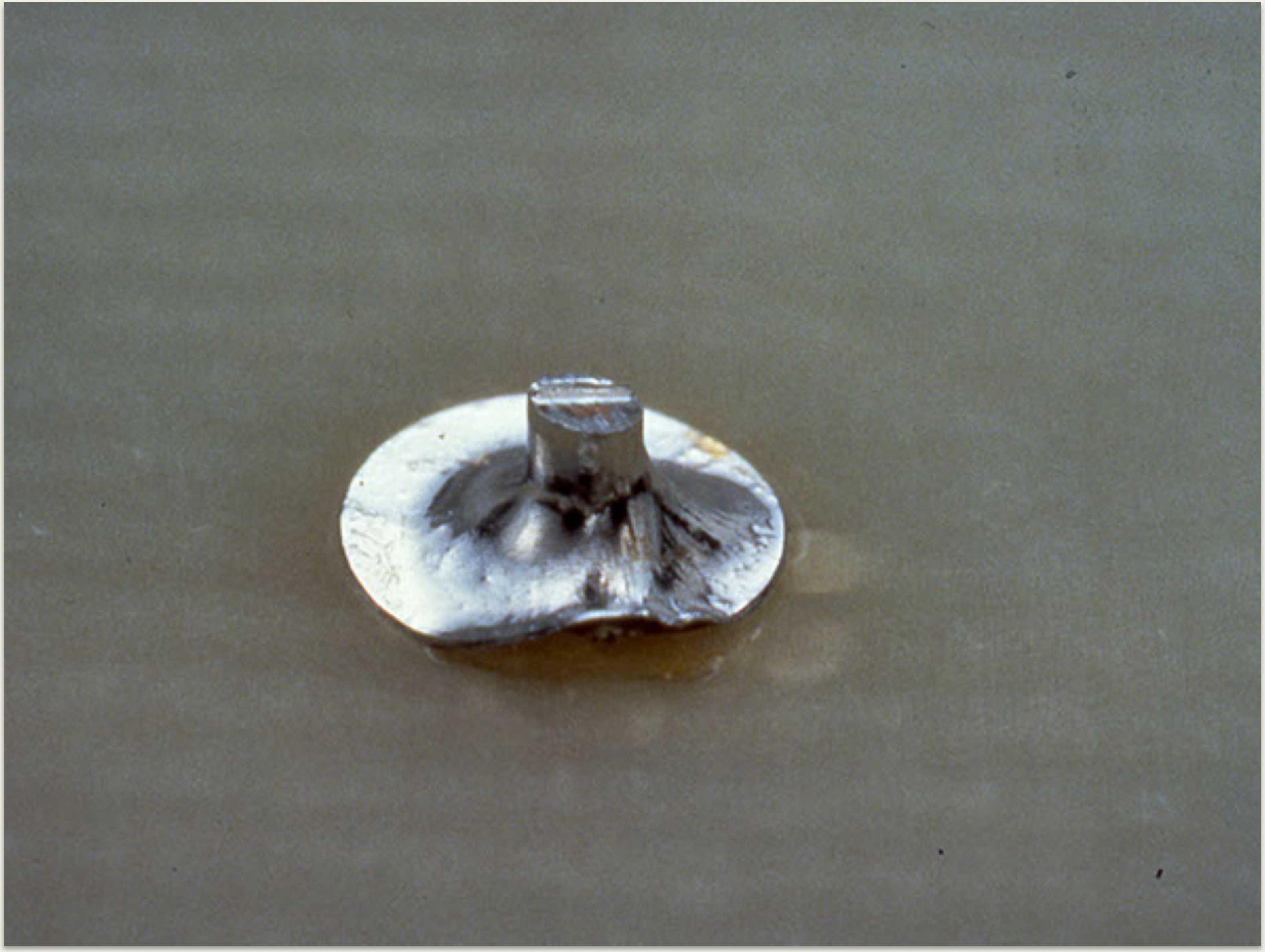

Lifted pad - Reject

CLINCHED TERMINATION

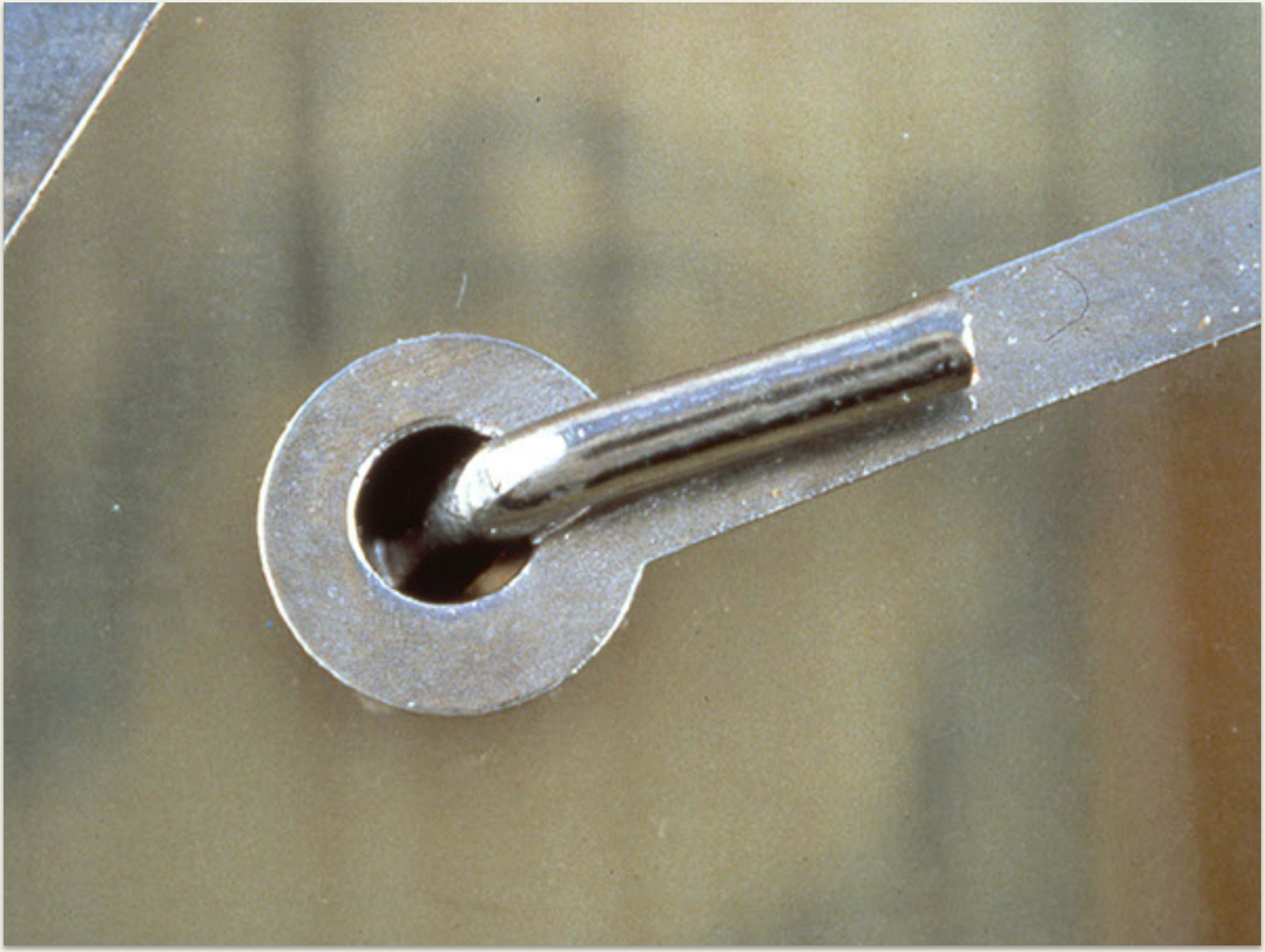

Clinched lead - Accept

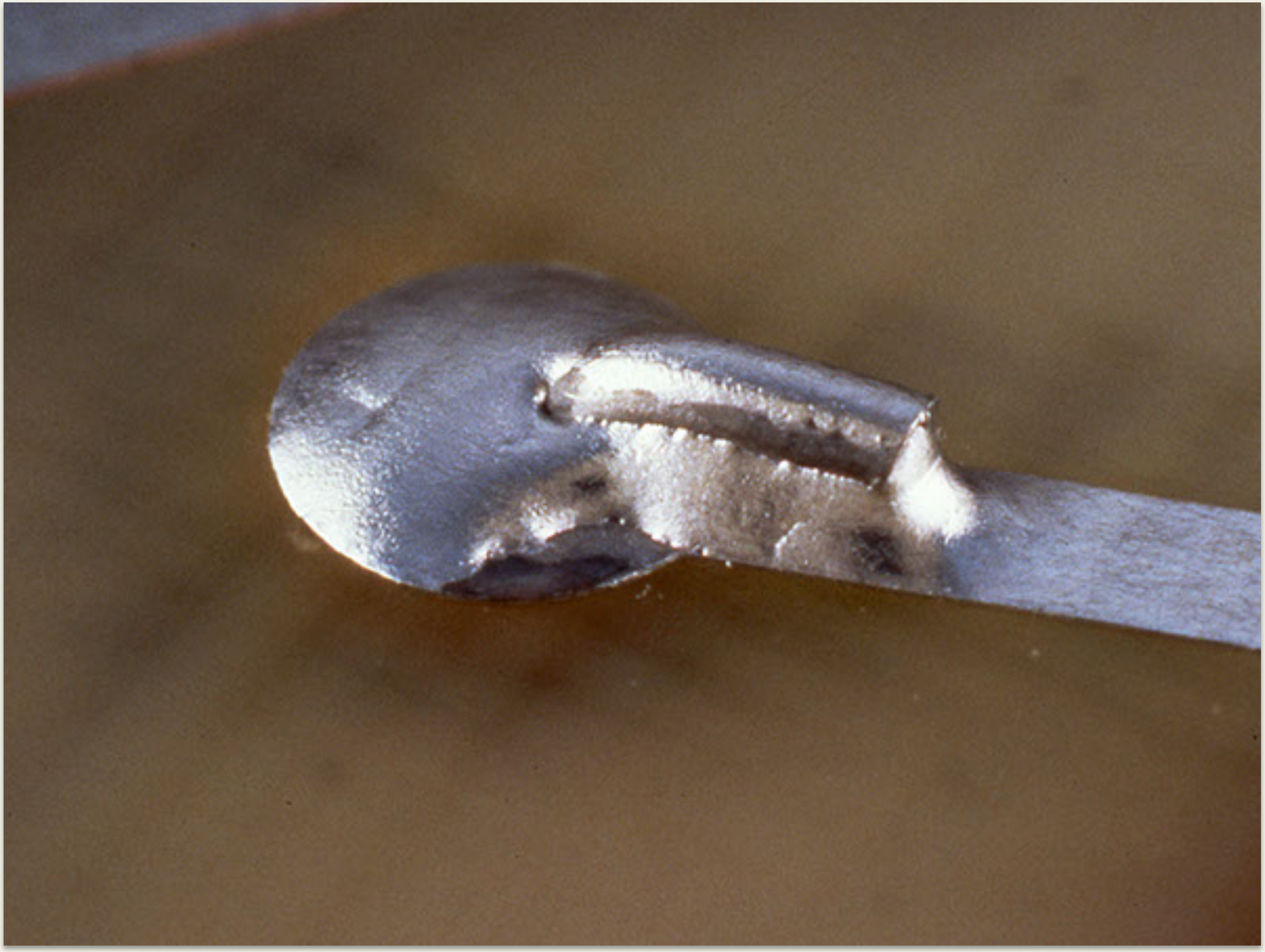

Cold solder, non wetting - Reject

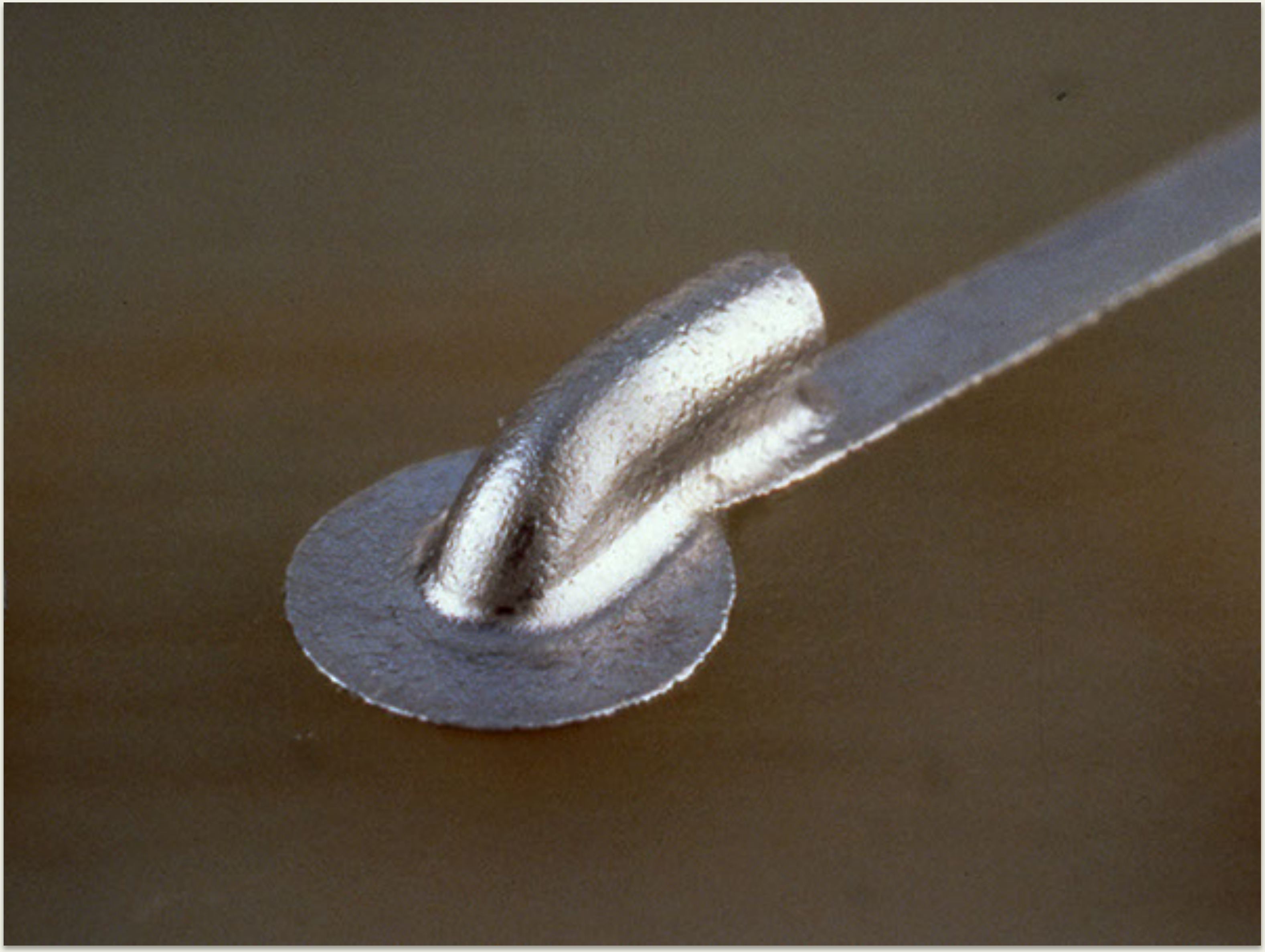

Minimum solder (should cover pad) - Accept

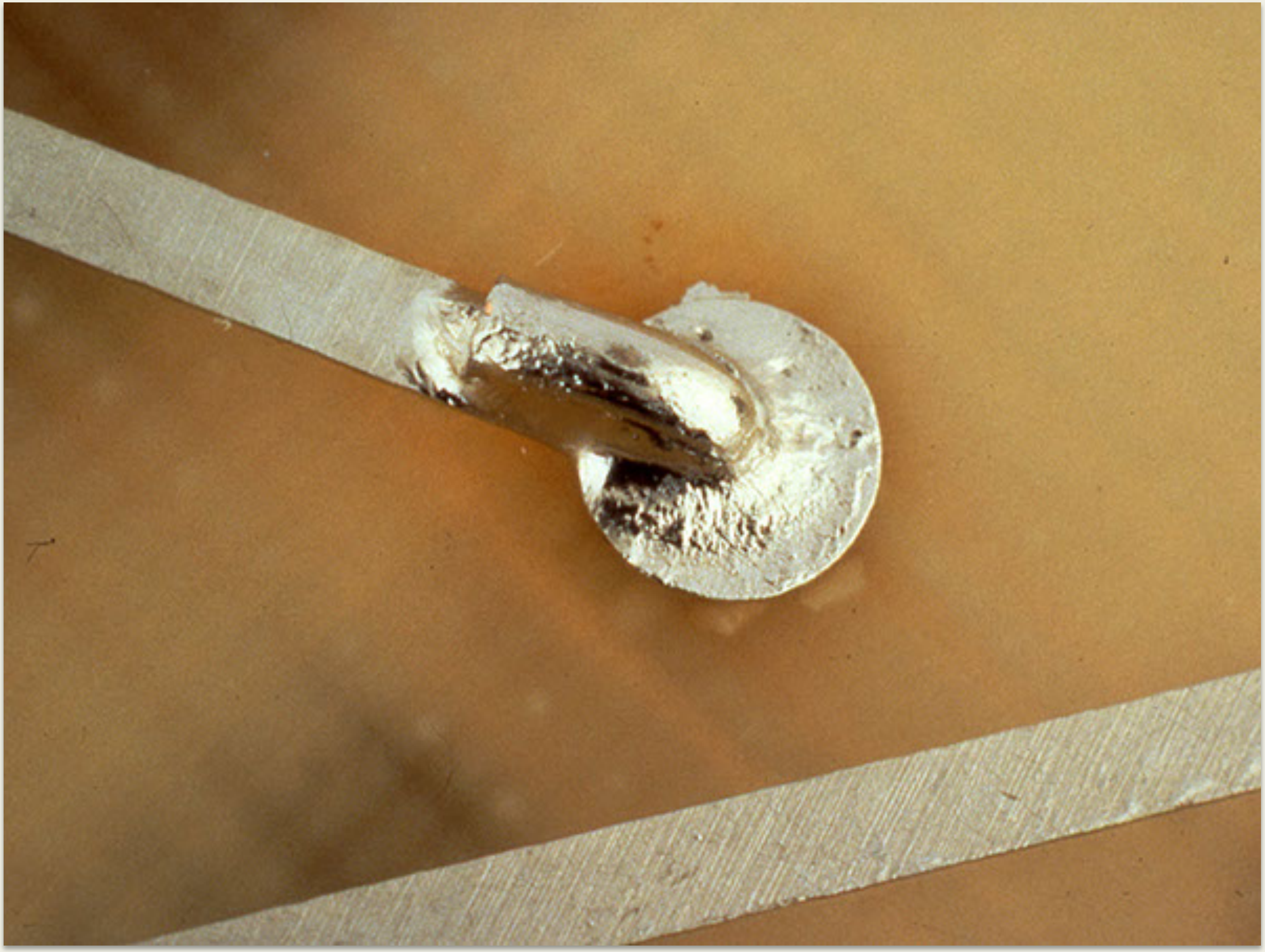

Grainy - Reject

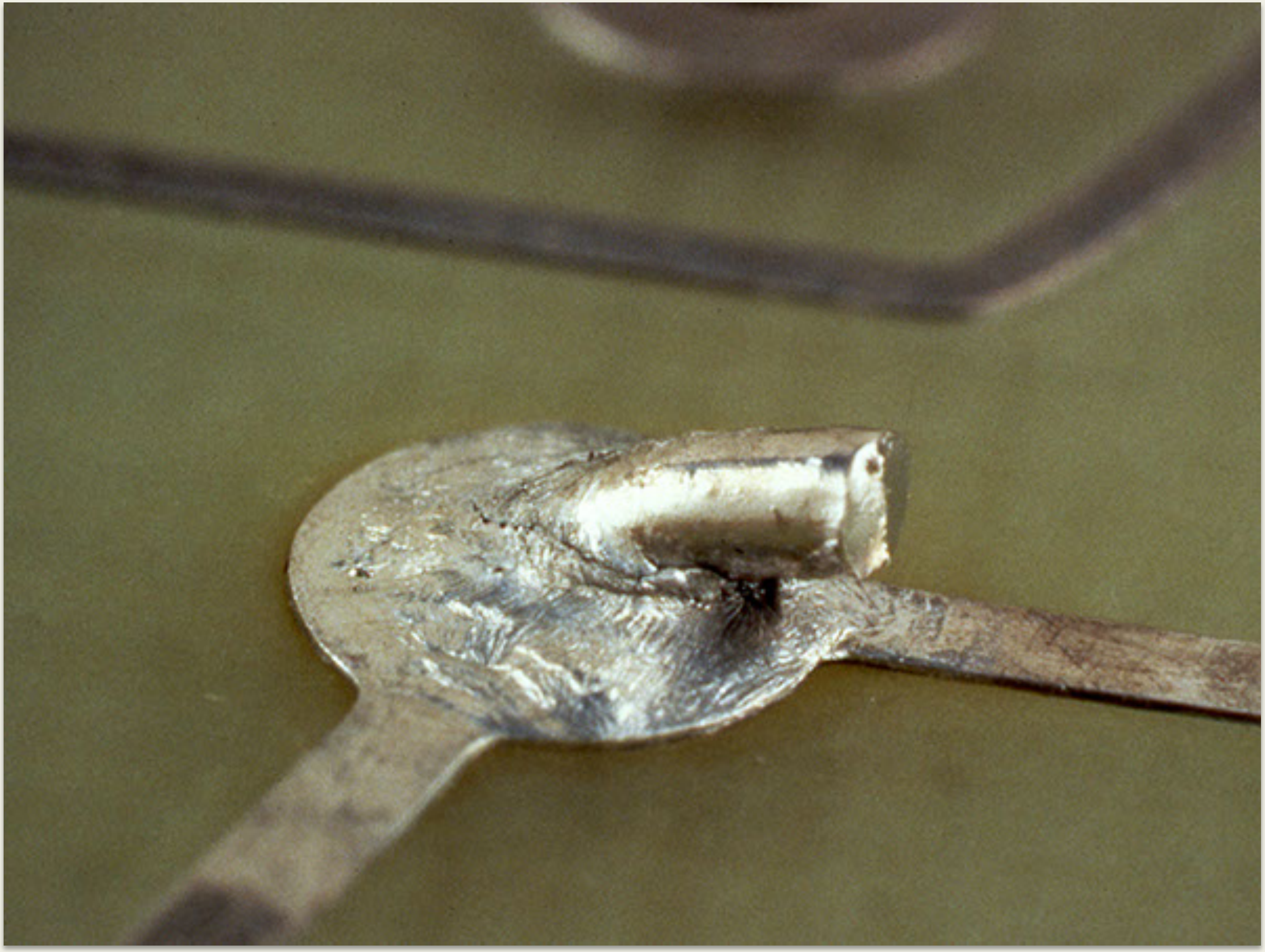

Disturbed solder - Reject

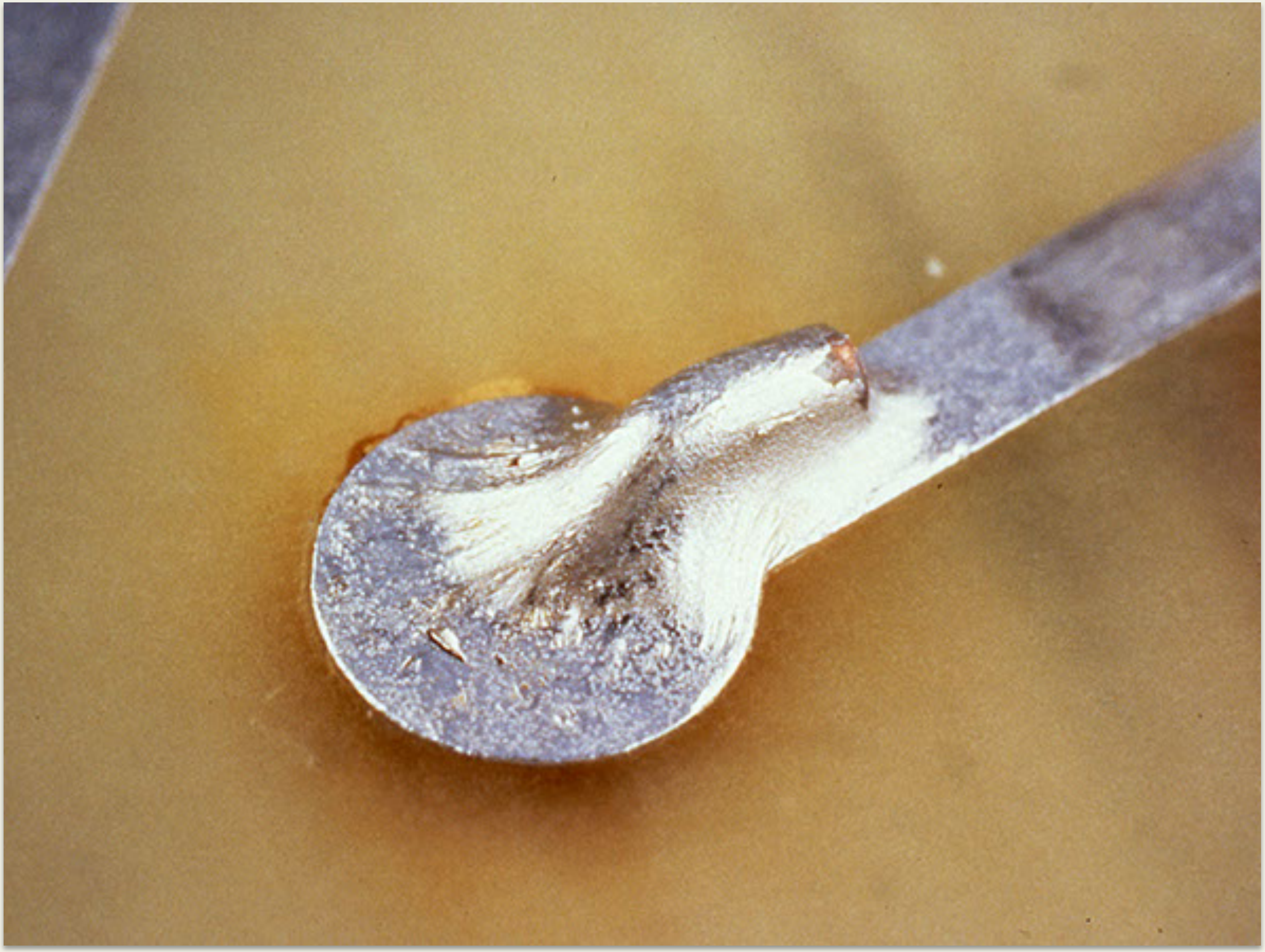

Overheated - Reject

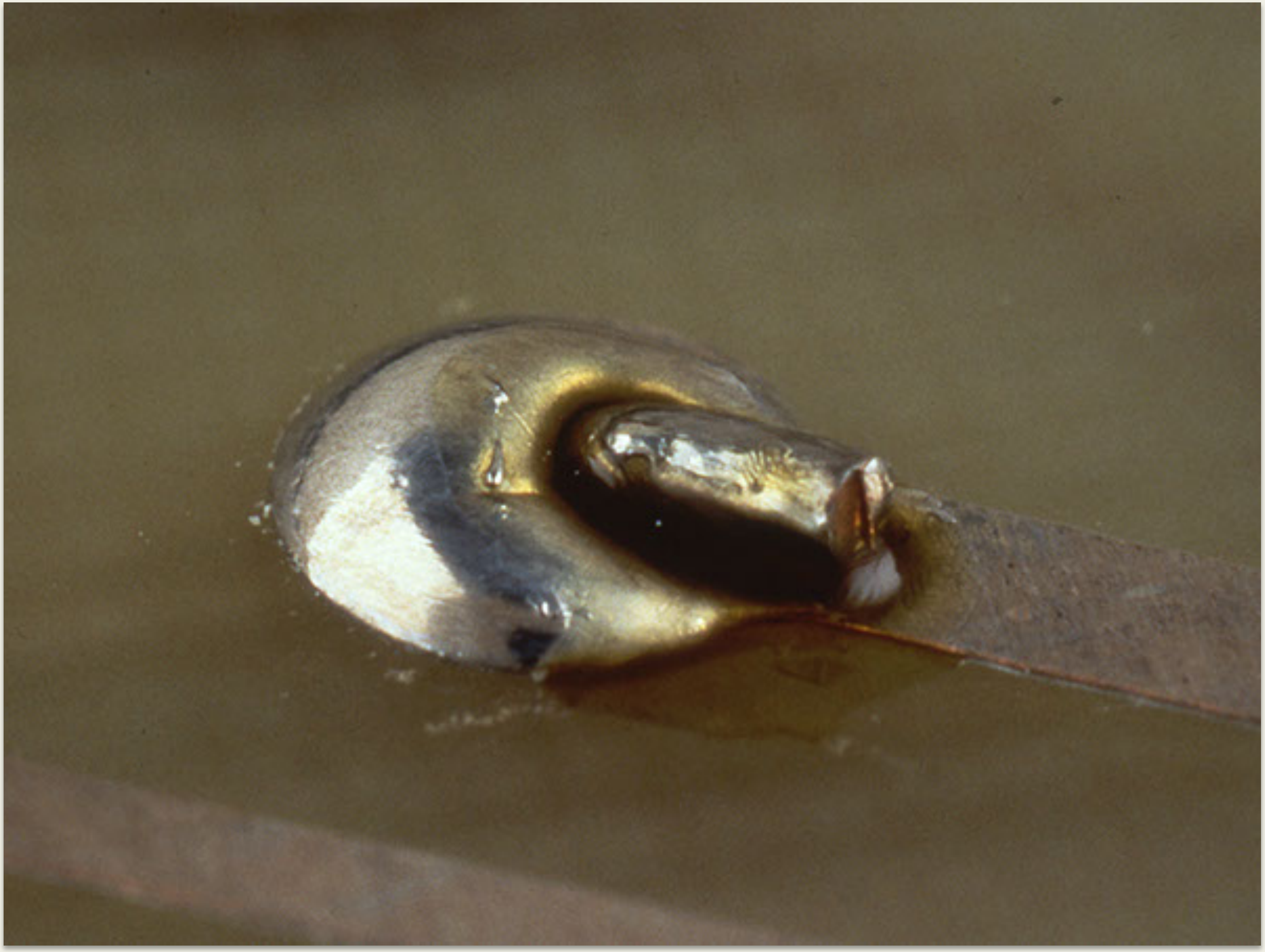

Rosin - Reject

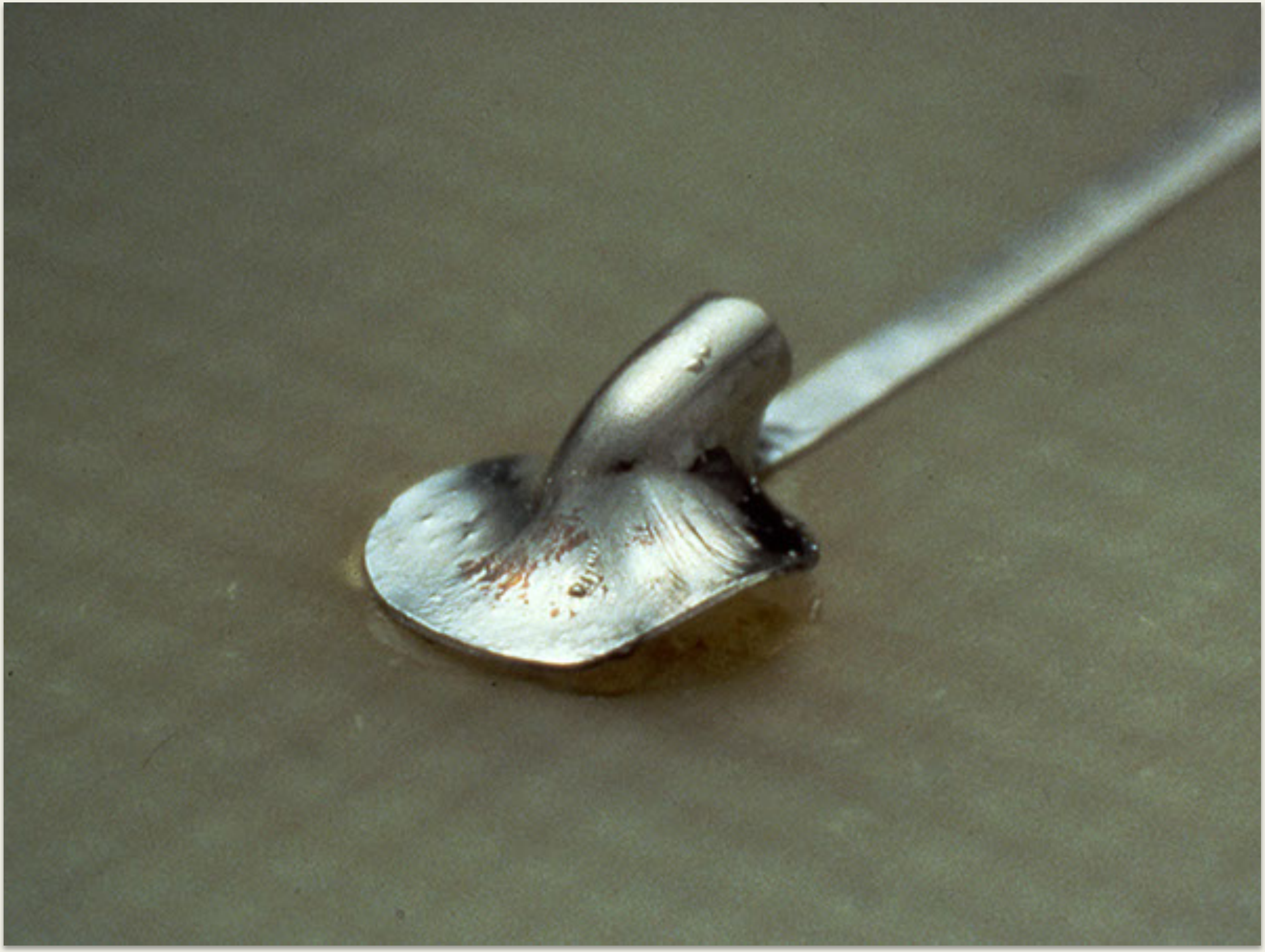

Lifted pad - Reject

VERTICAL MOUNTING

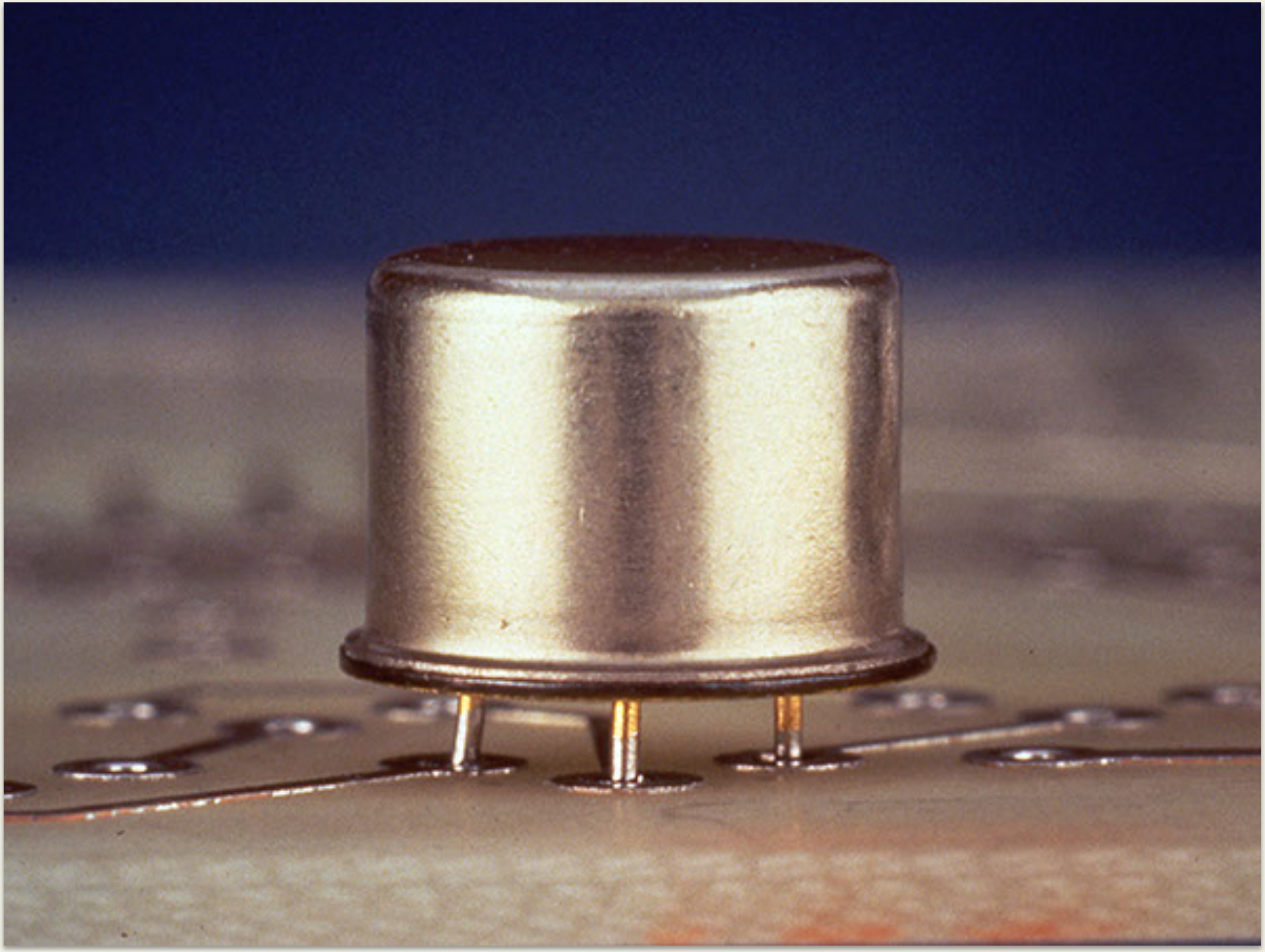

T05 transistor component mounting - Accept

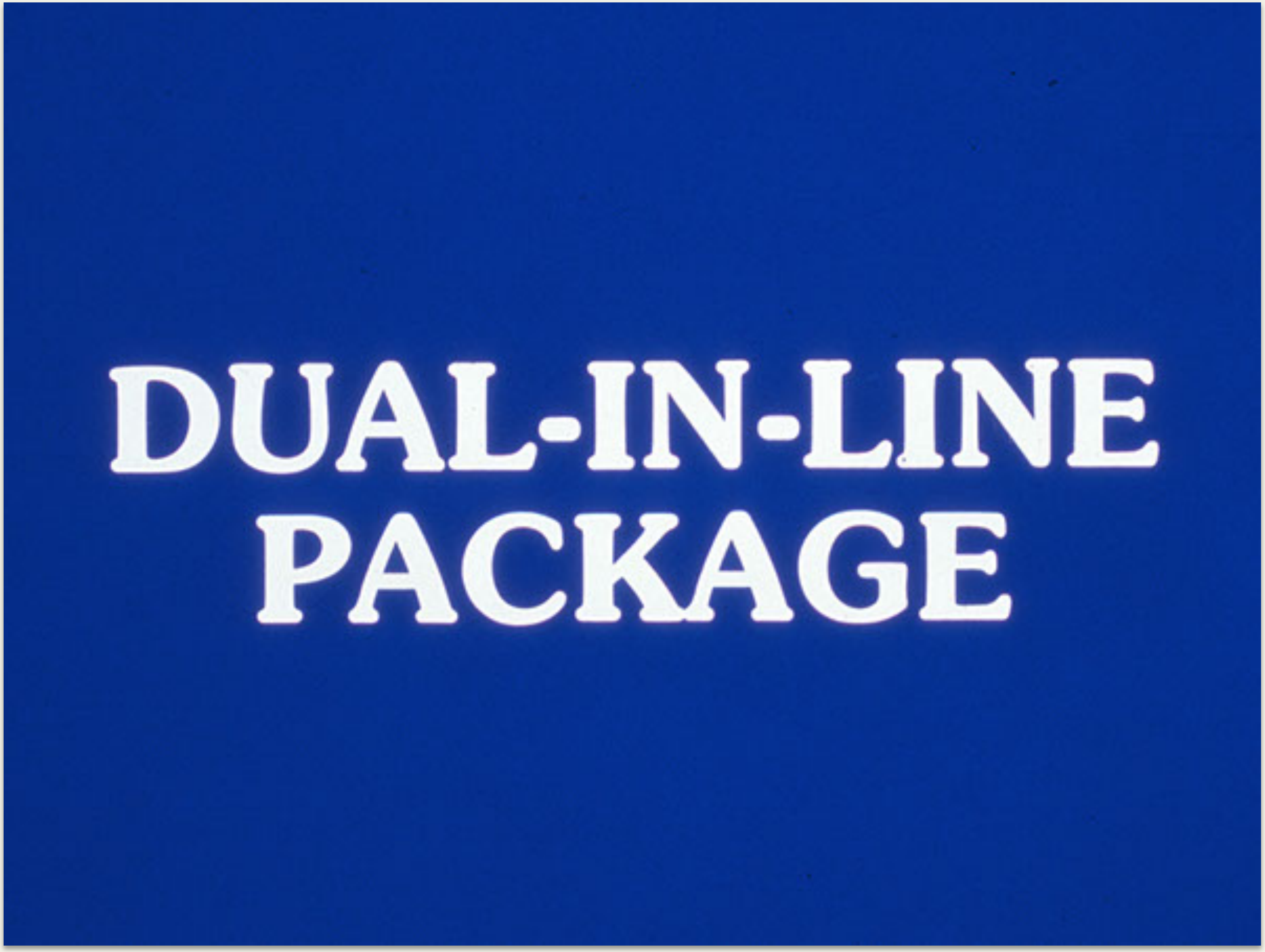

DUAL-IN-LINE PACKAGE

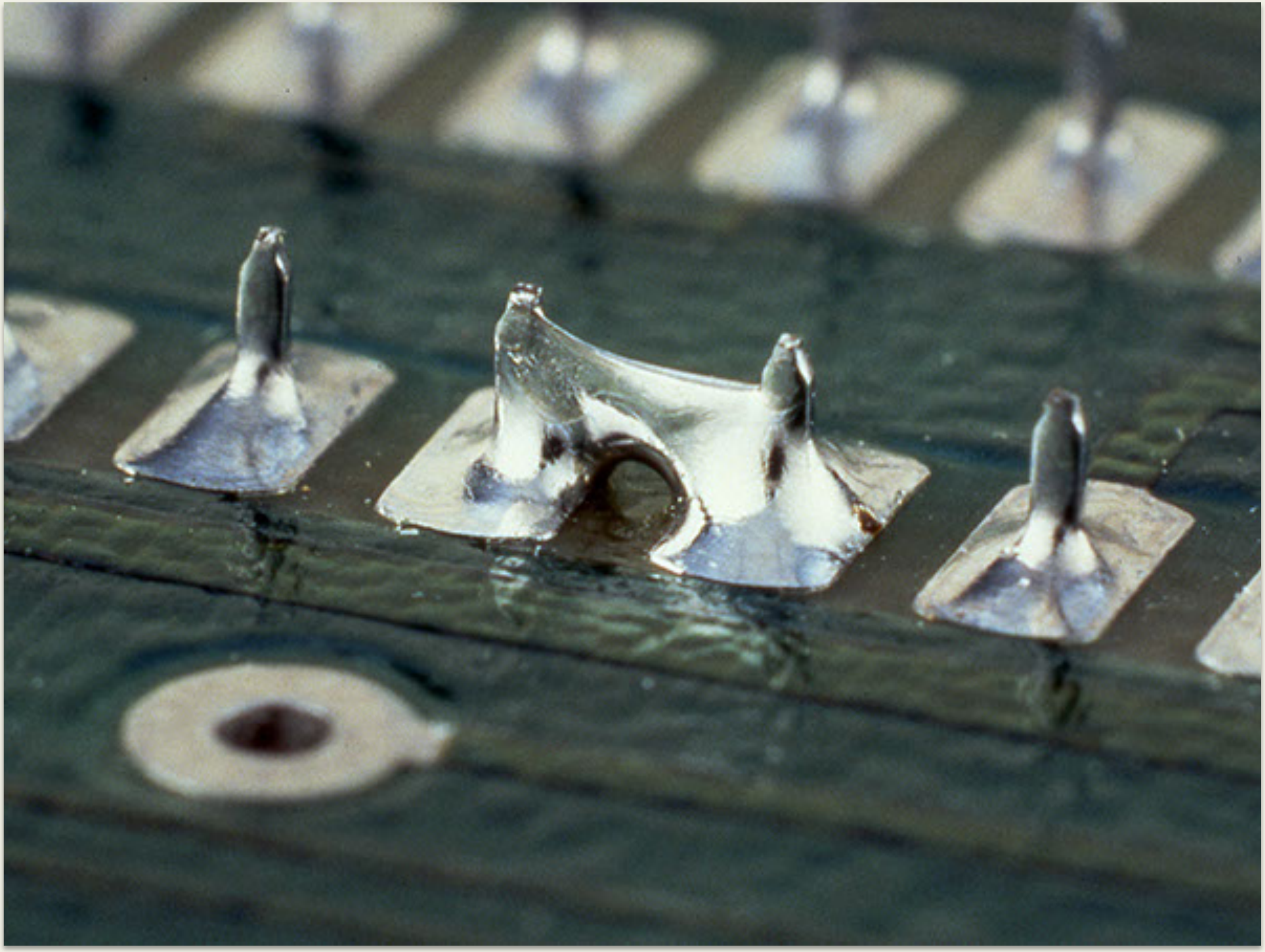

Solder bridge between leads - Reject

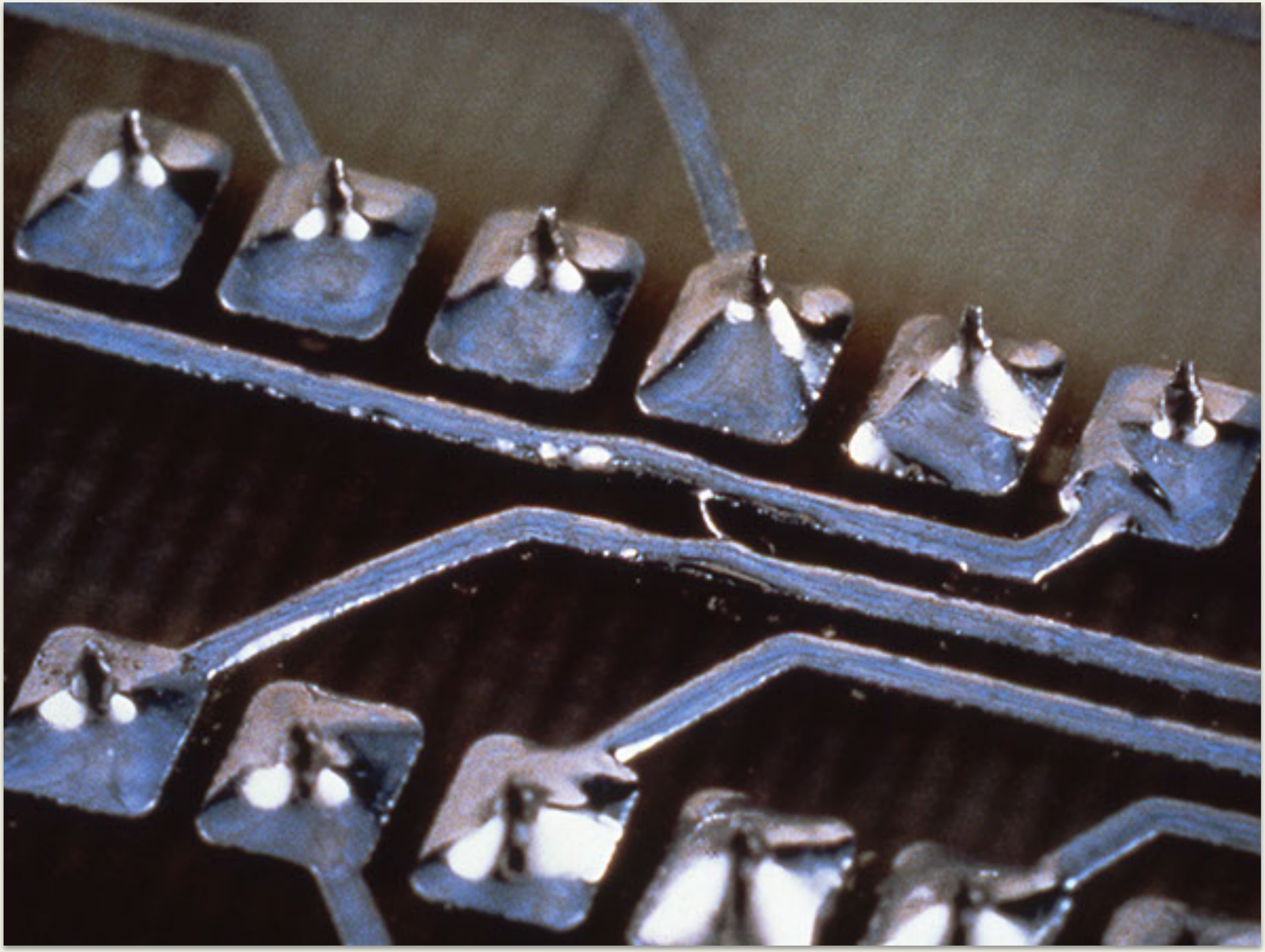

Solder bridge between leads - Reject

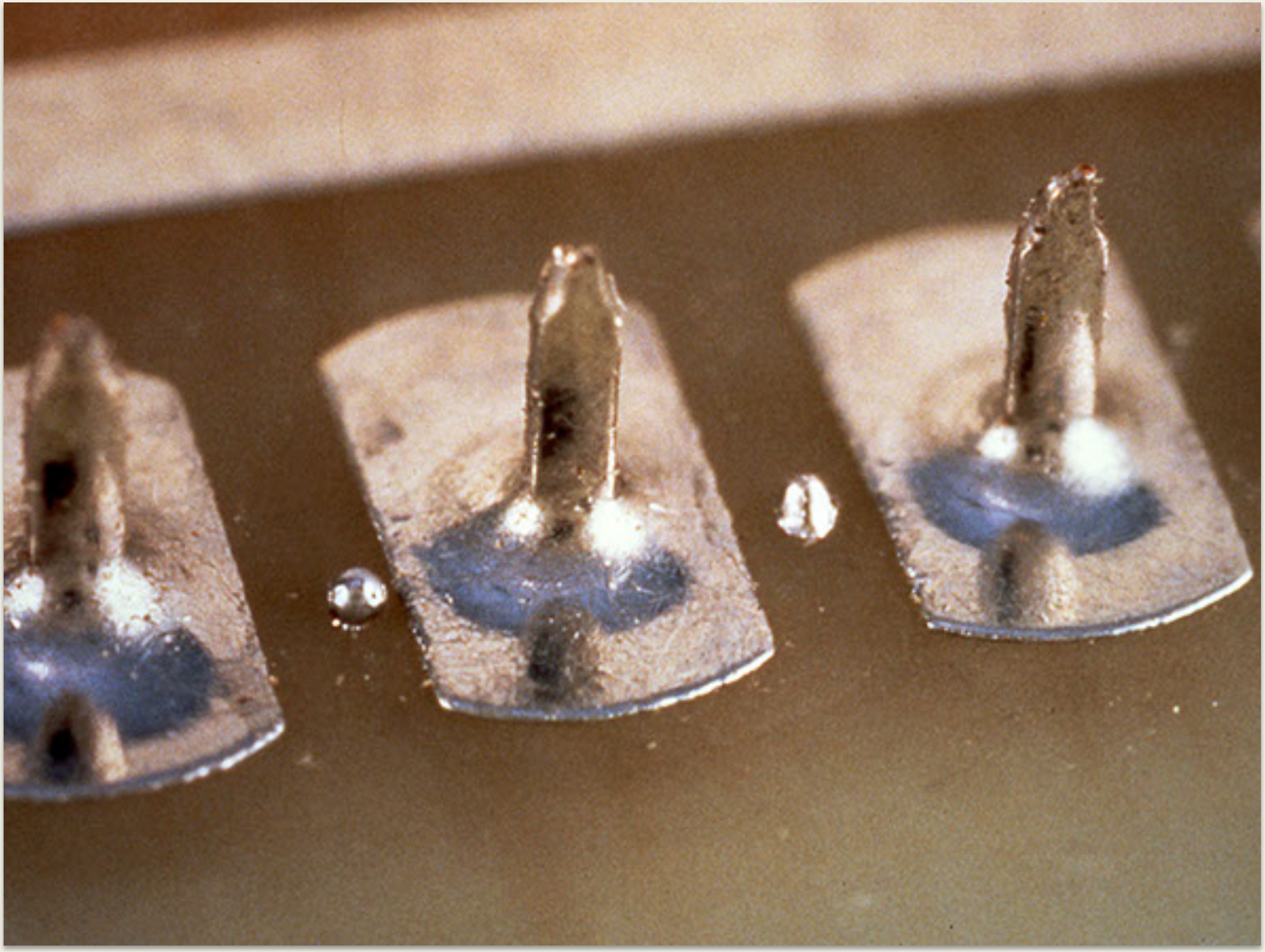

Solder splashes/splatters/balls - Reject

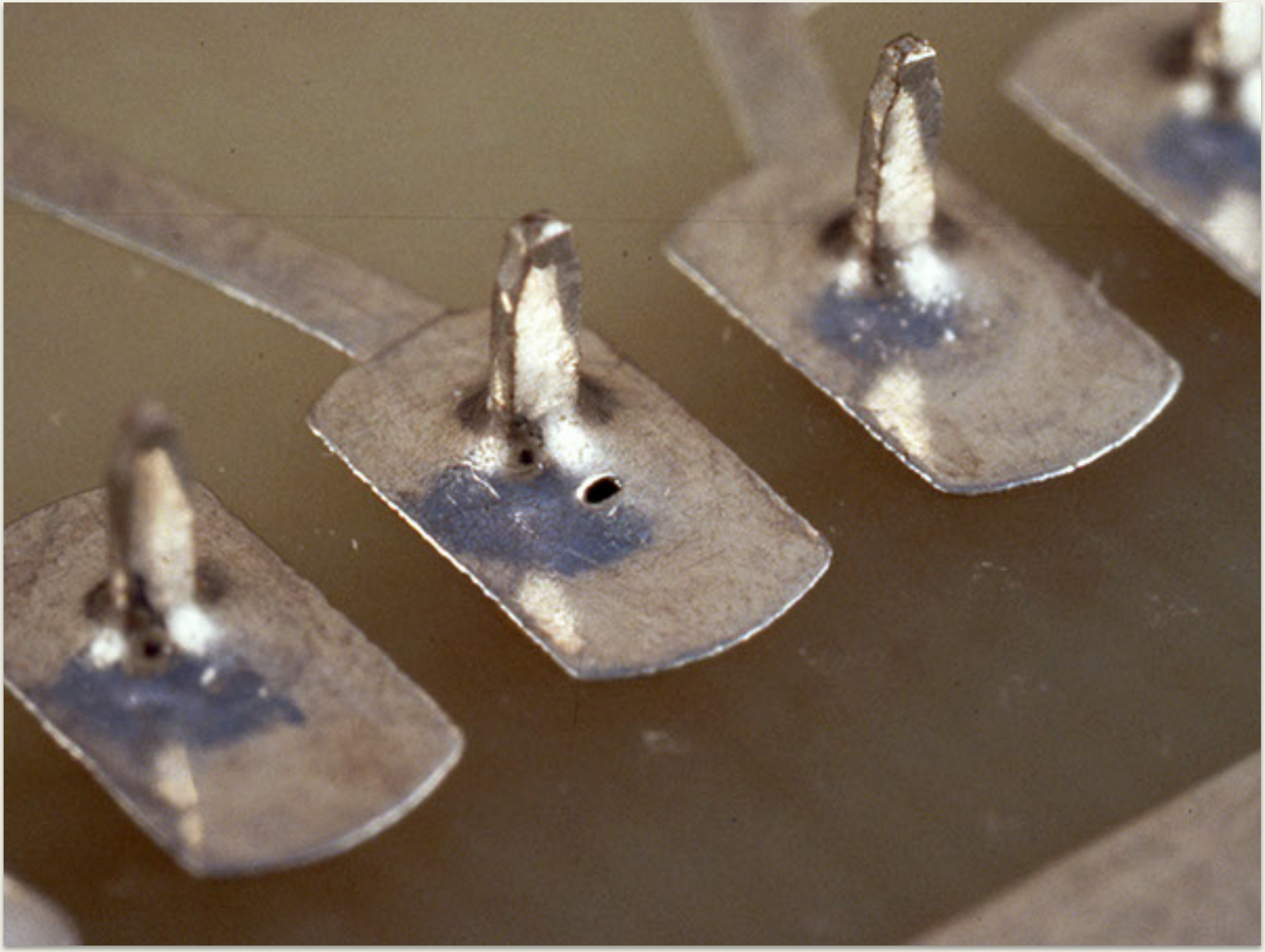

Pin hole - Reject

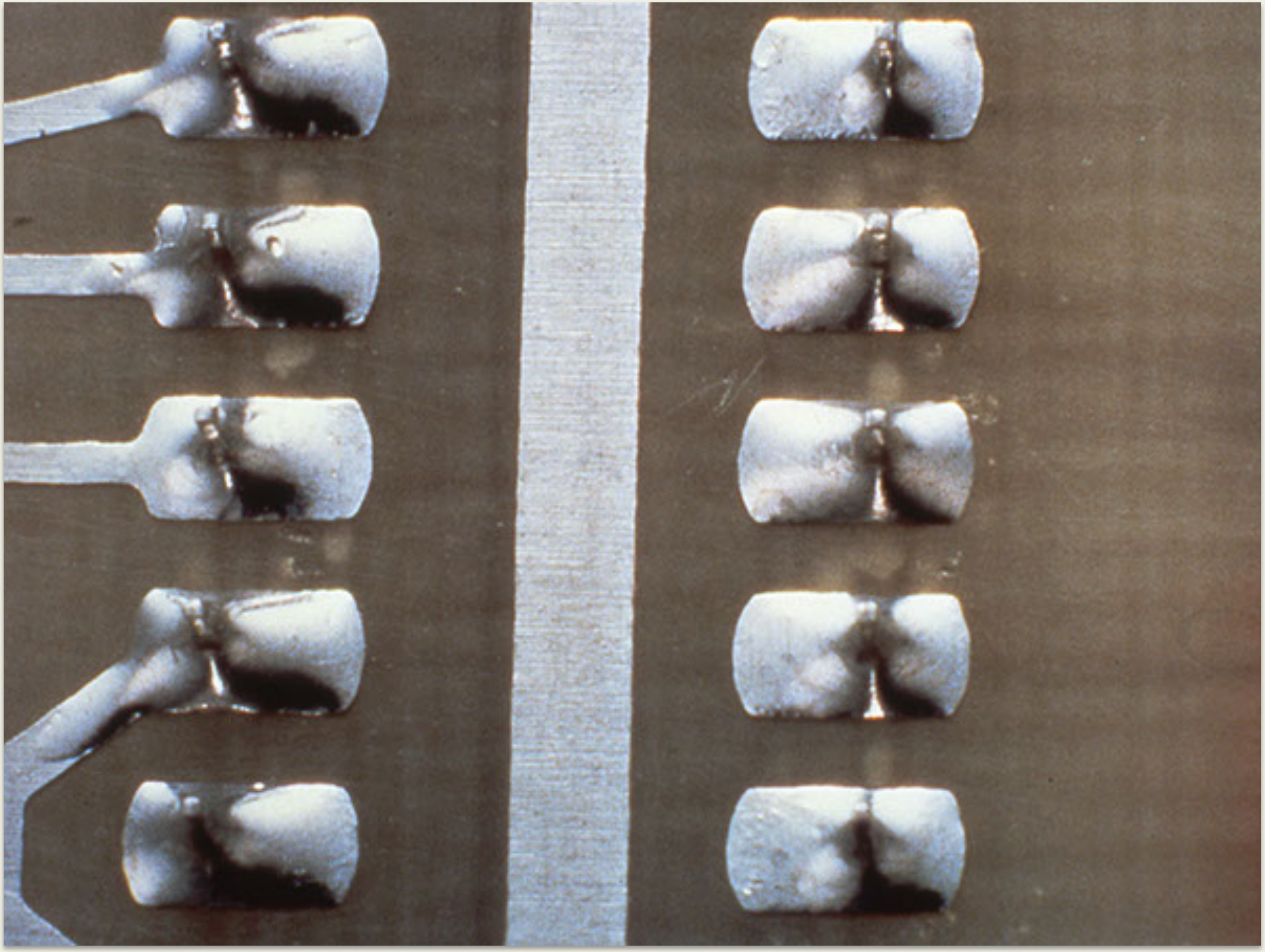

Measling - Reject

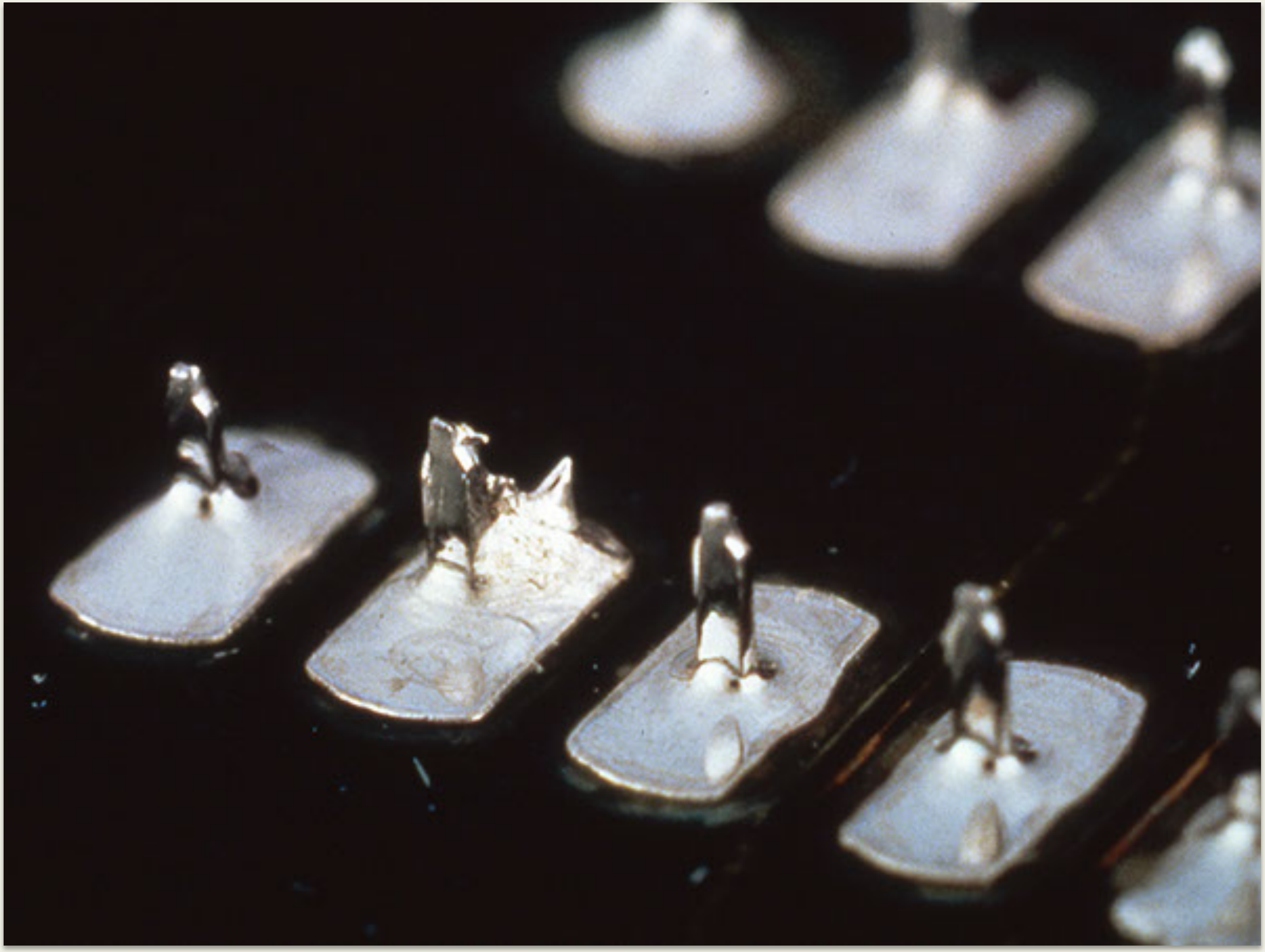

Solder spikes, peaks and icicles - Reject

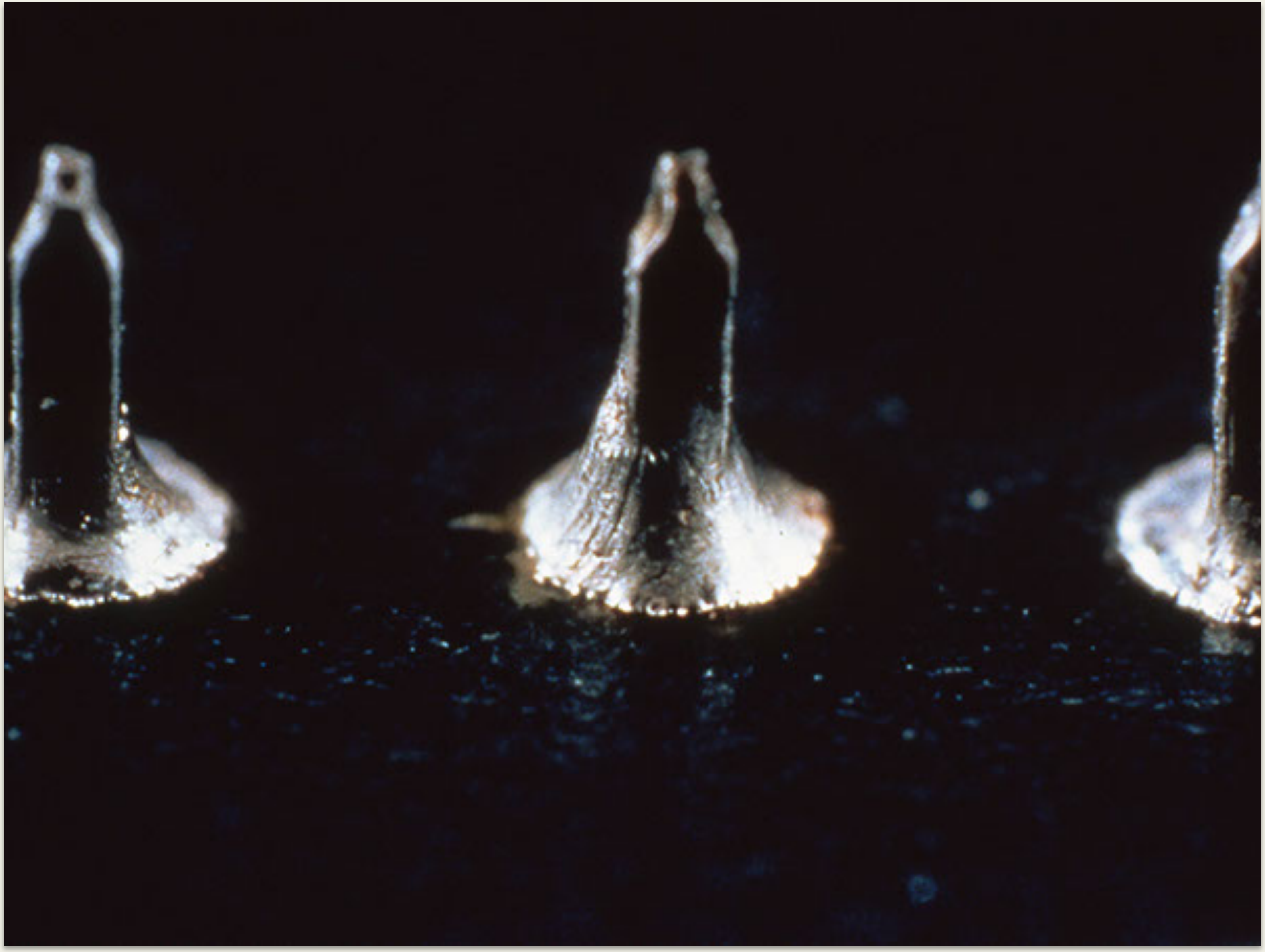

Non wetting/stress lines - Reject

INTERFACIAL CONNECTIONS

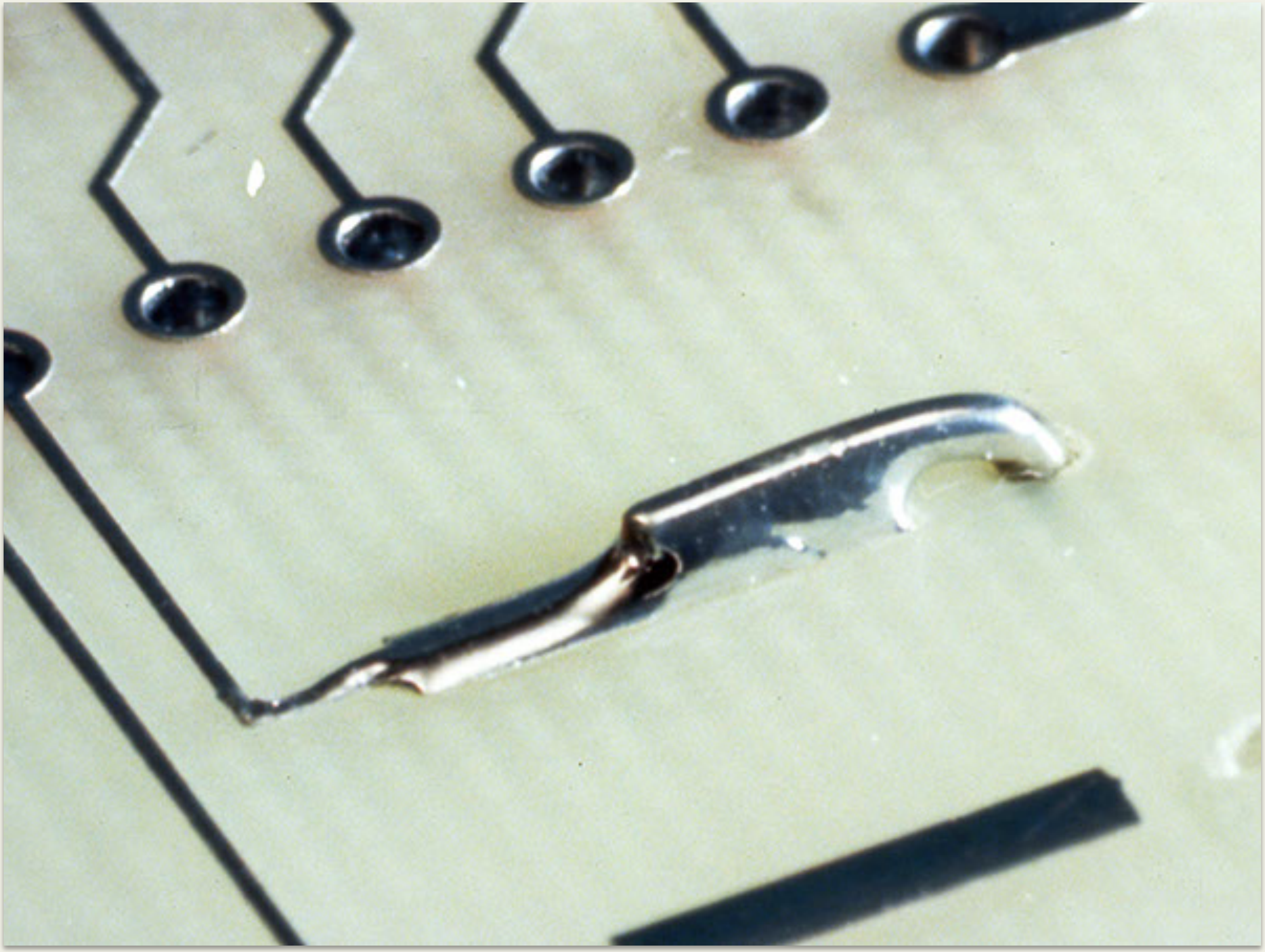

Void under end of wire - Reject

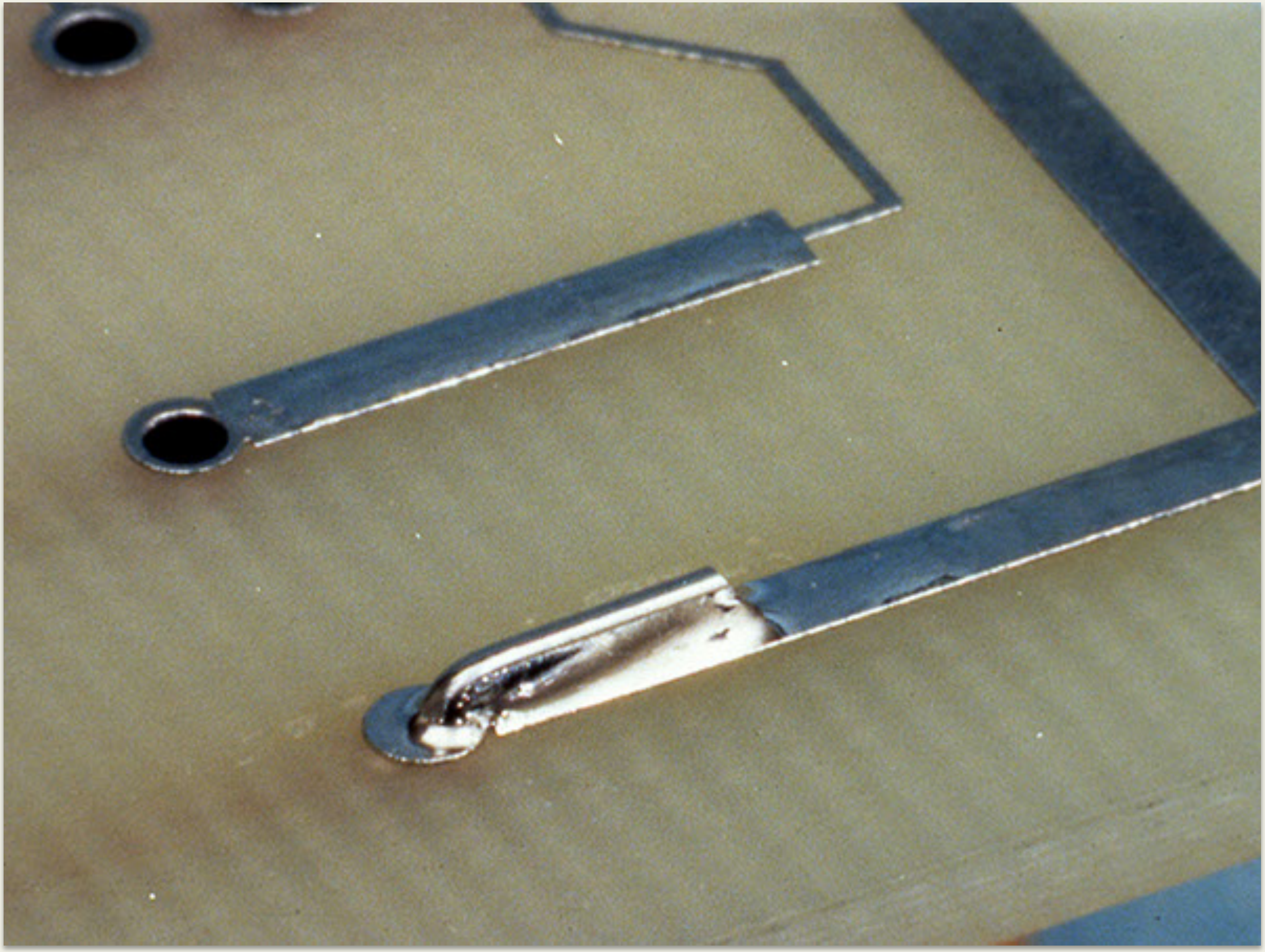

Insufficient solder - Reject

FLAT
PACK

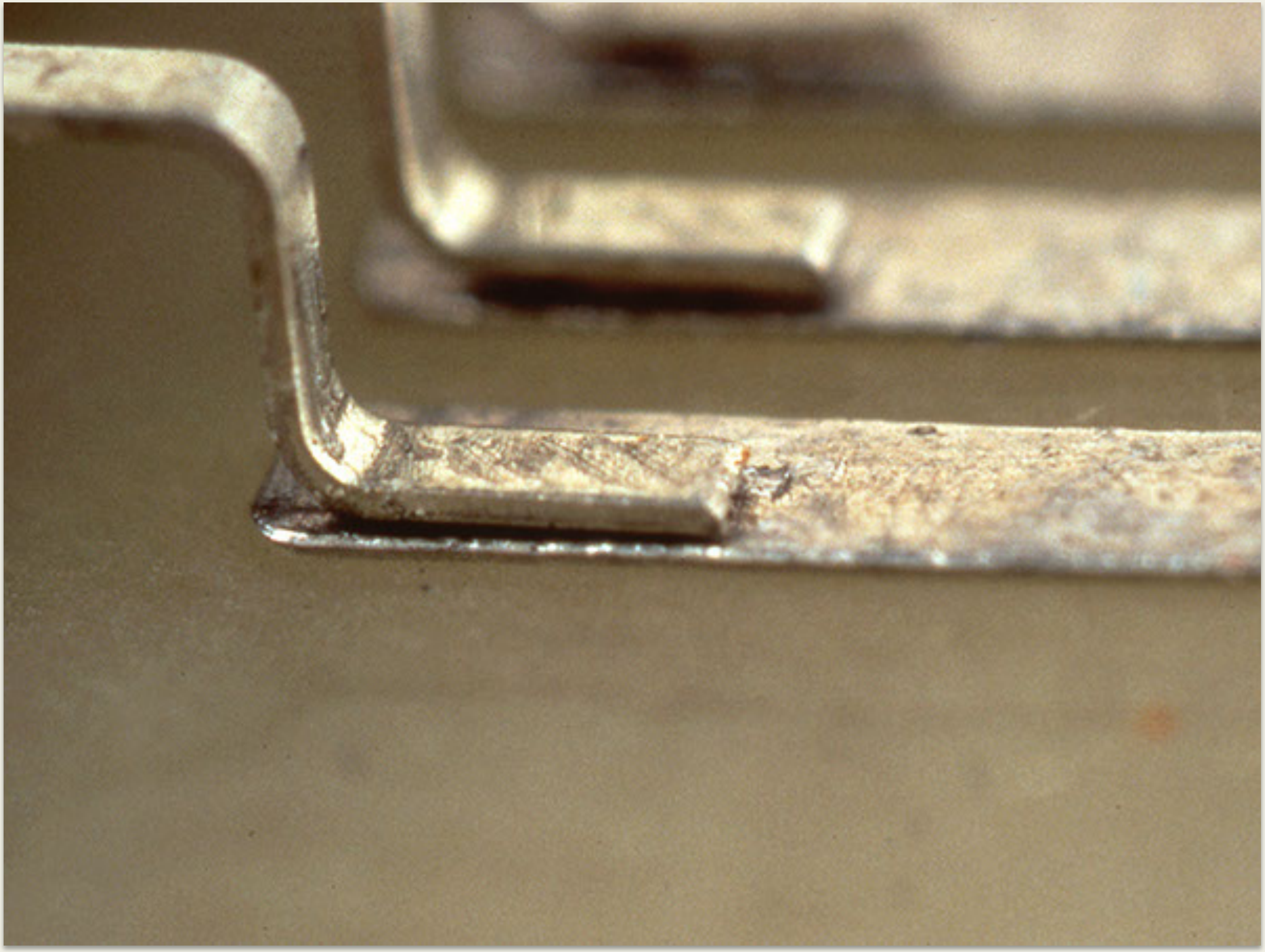

Preferred mounting - Accept

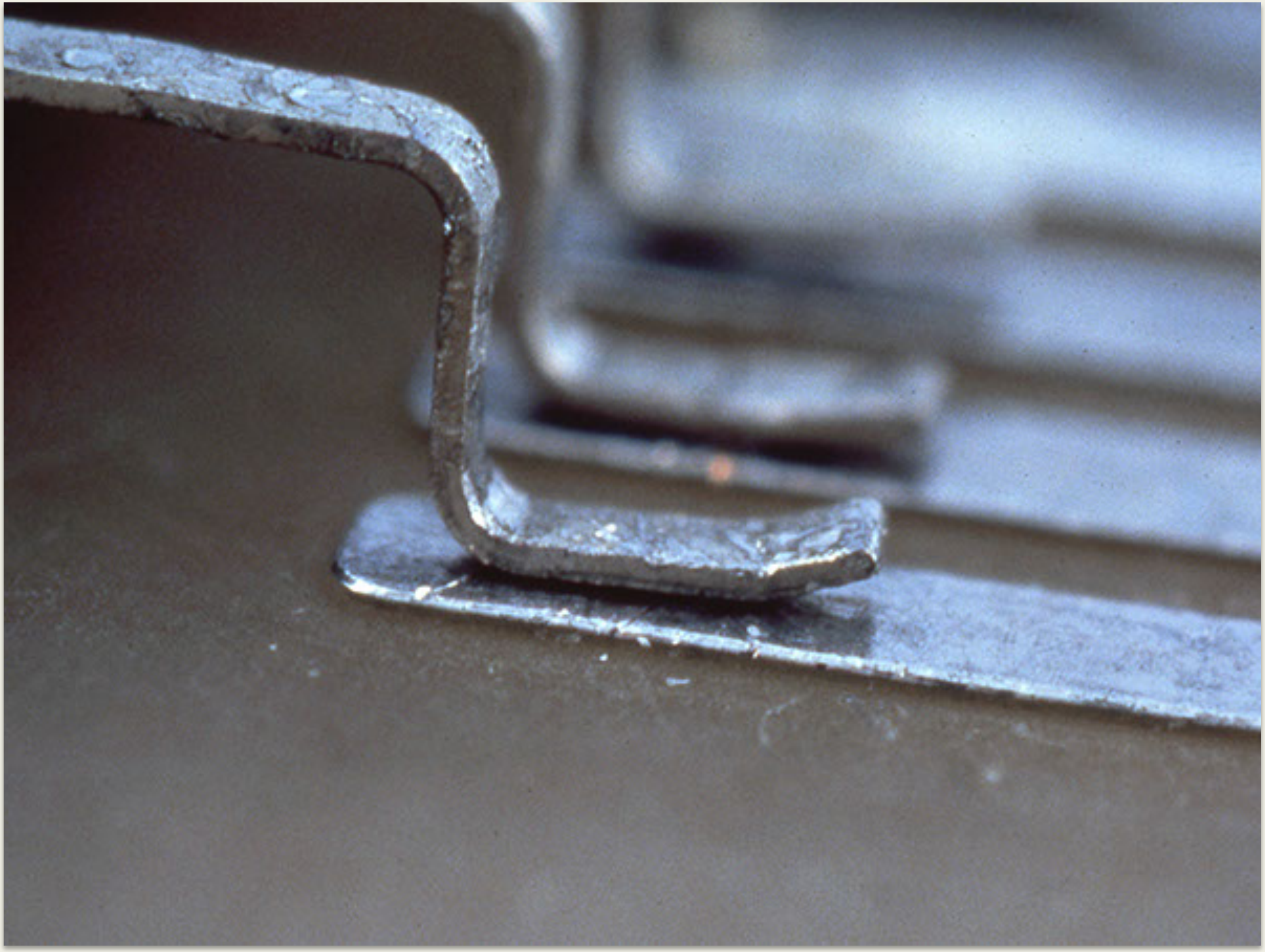

Toe up/curl - Reject

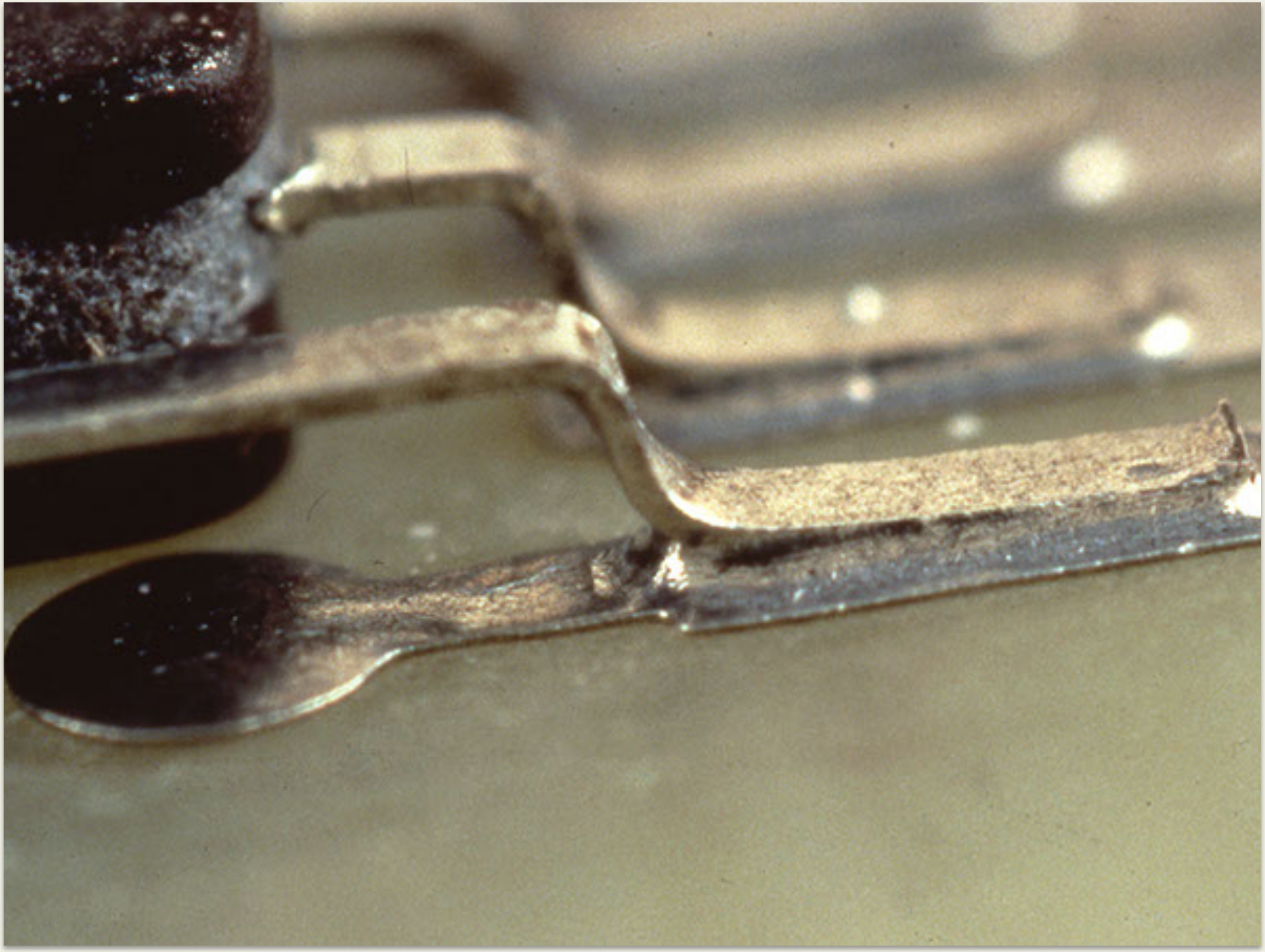

Heel fillet, lack of solder - Reject

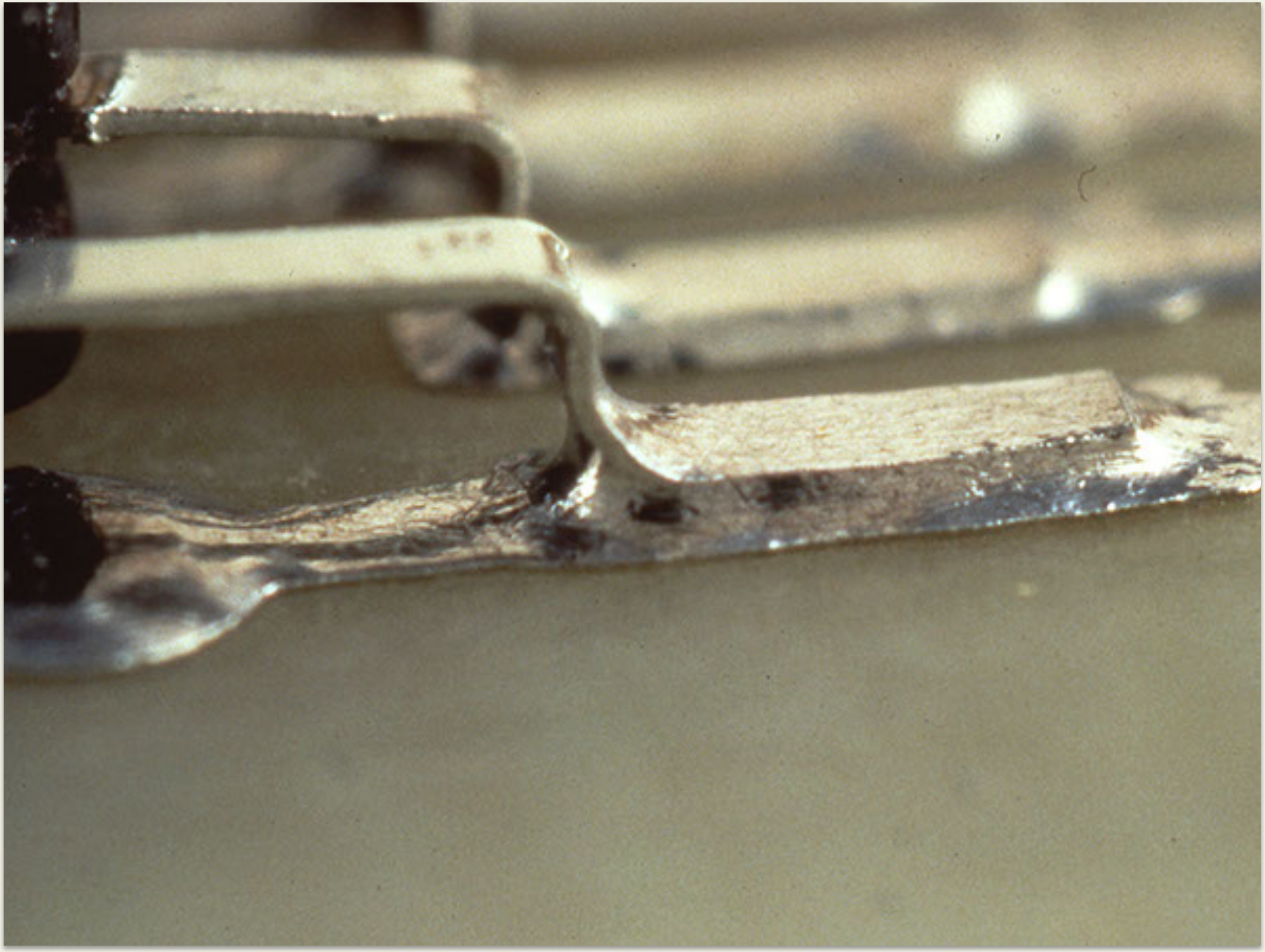

Heel fillet, not smooth - Reject

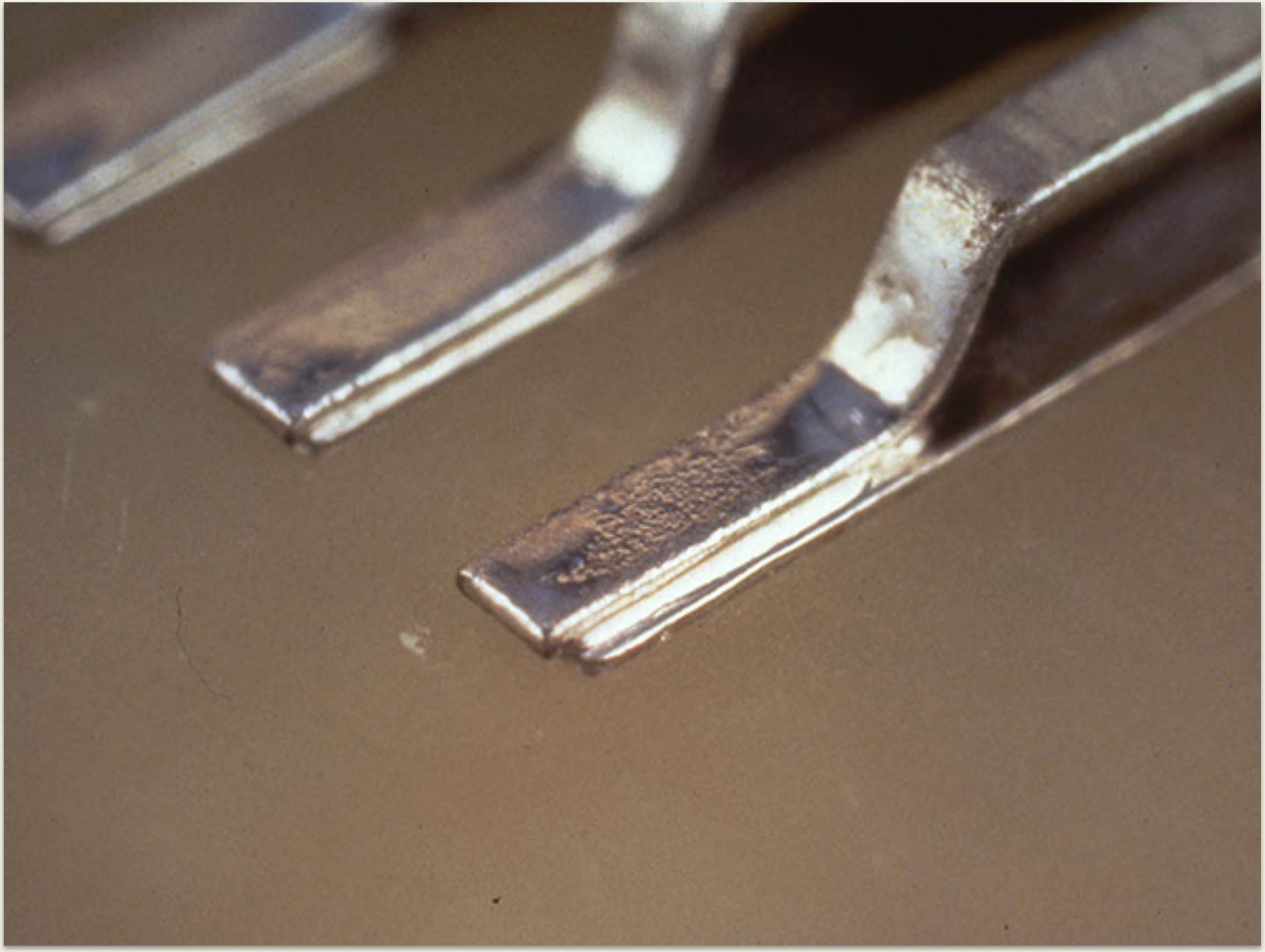

Toe overhang, excessive - Reject

CONTINUOUS WRAP

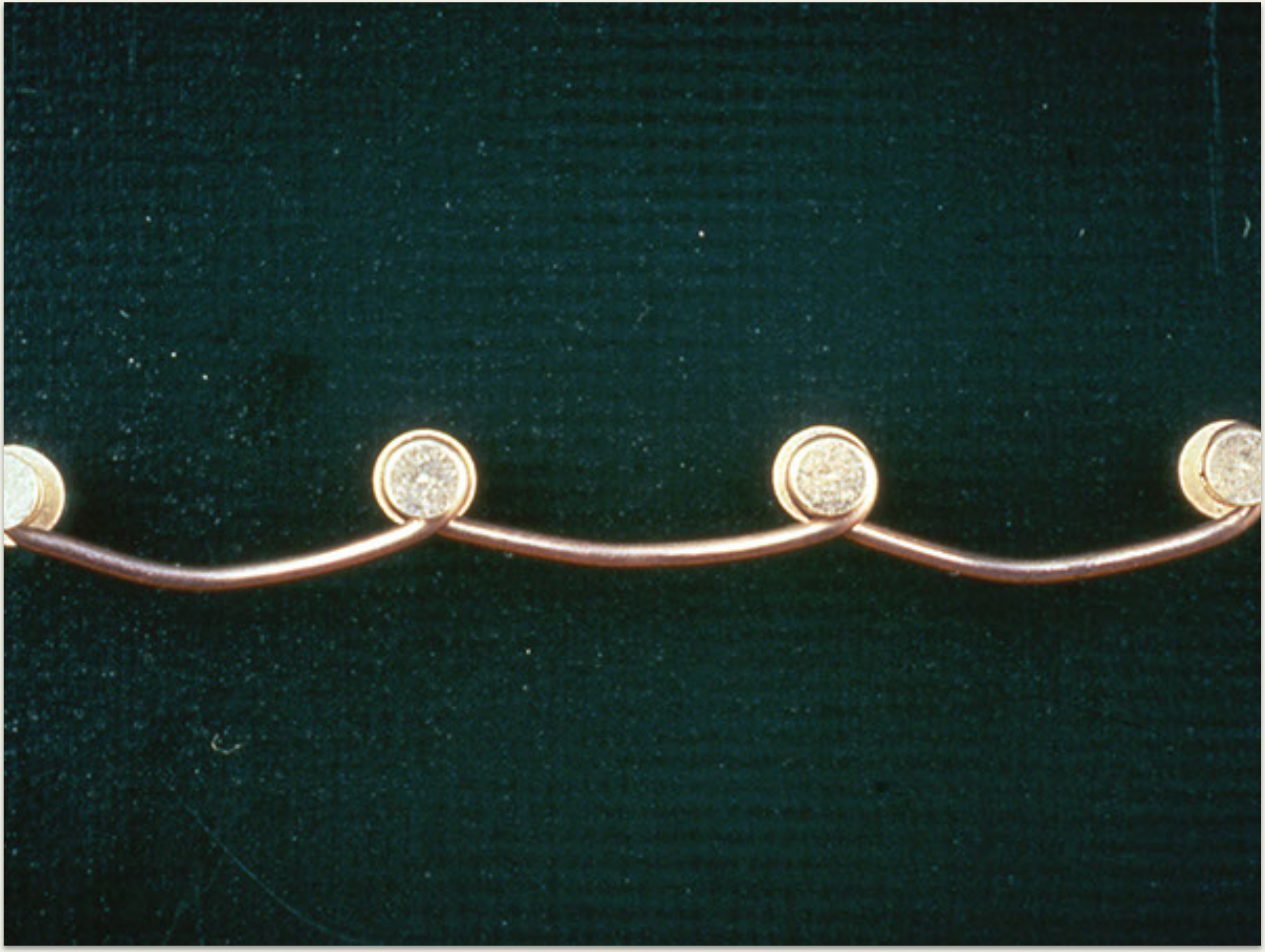

Turret terminals - Accept

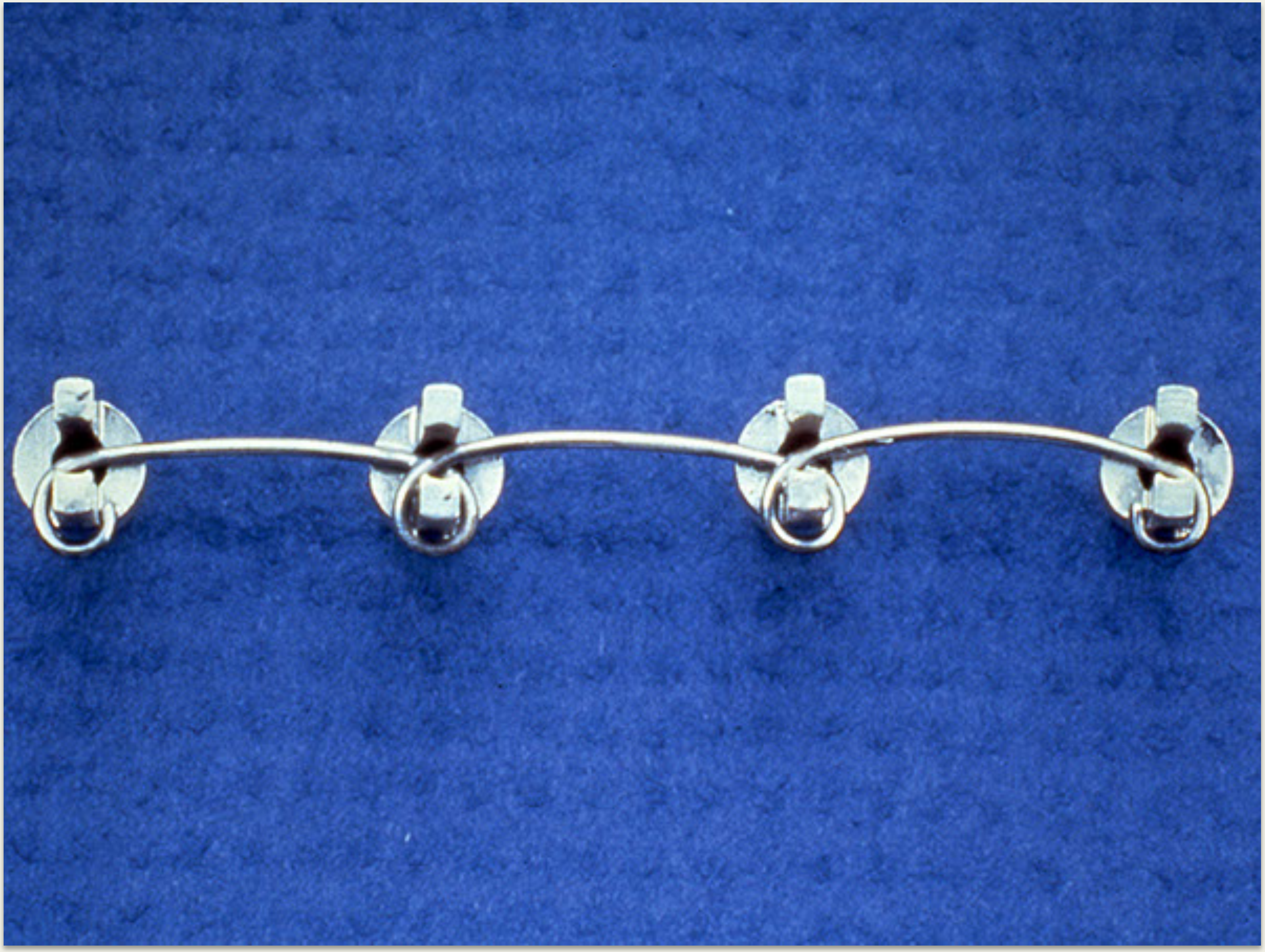

Loose wrap/greater than 180 degrees end wraps - Reject

HIGH VOLTAGE TERMINATION

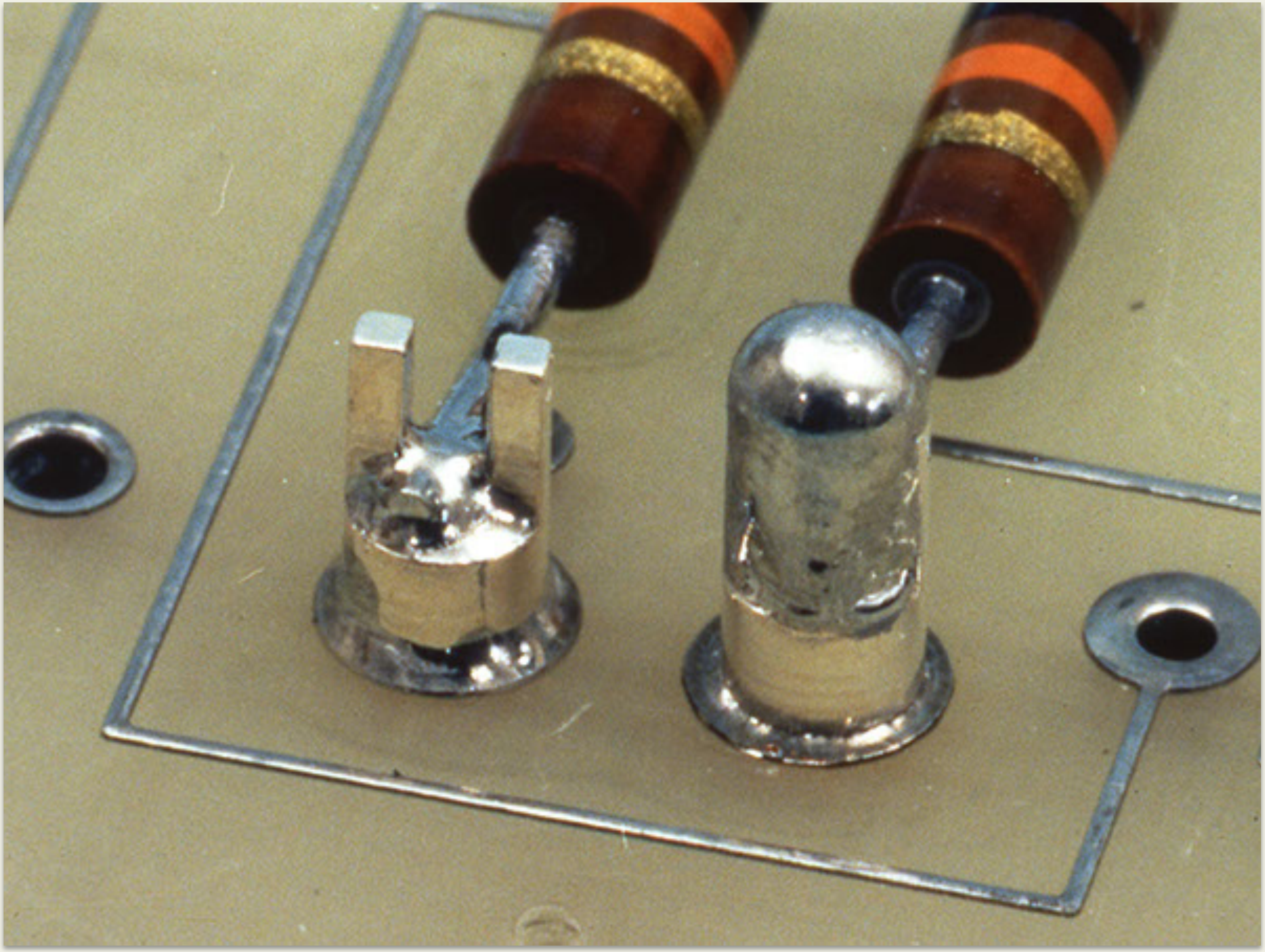

Left, solder lead/contamination – Reject
Right, projection, voids - Reject

INSPECTION

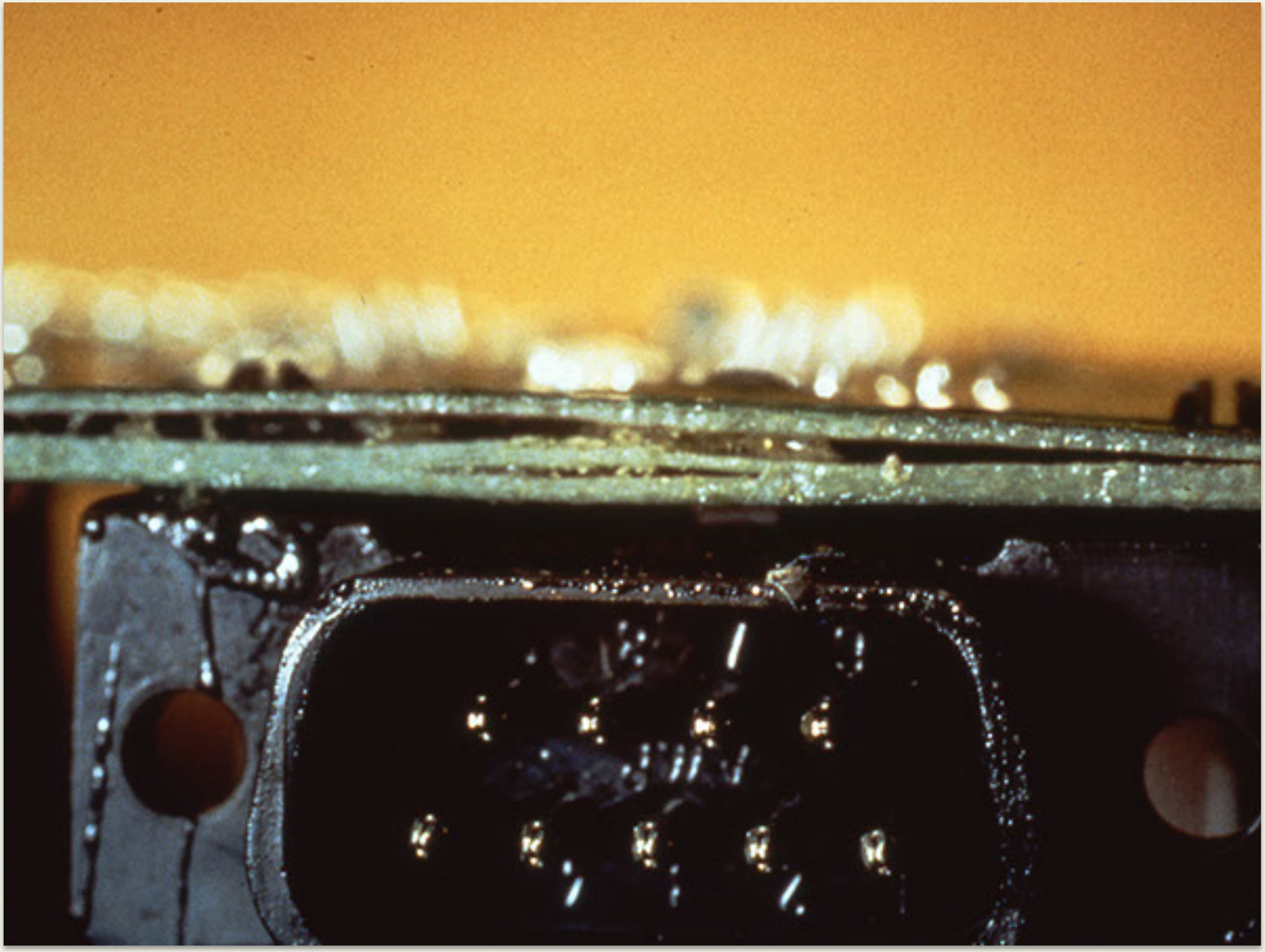

Delamination, base material - Reject

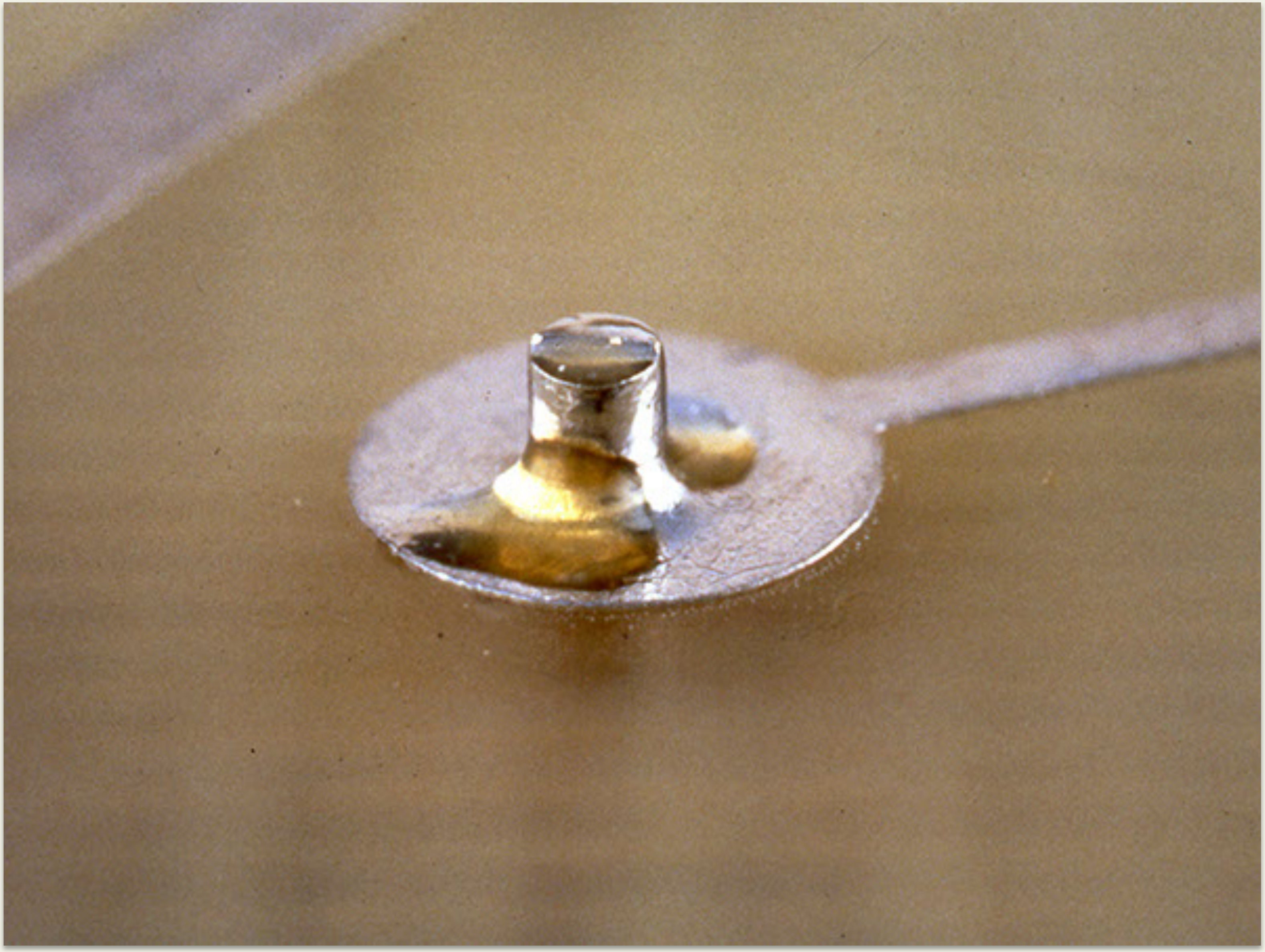

Flux residue - Reject

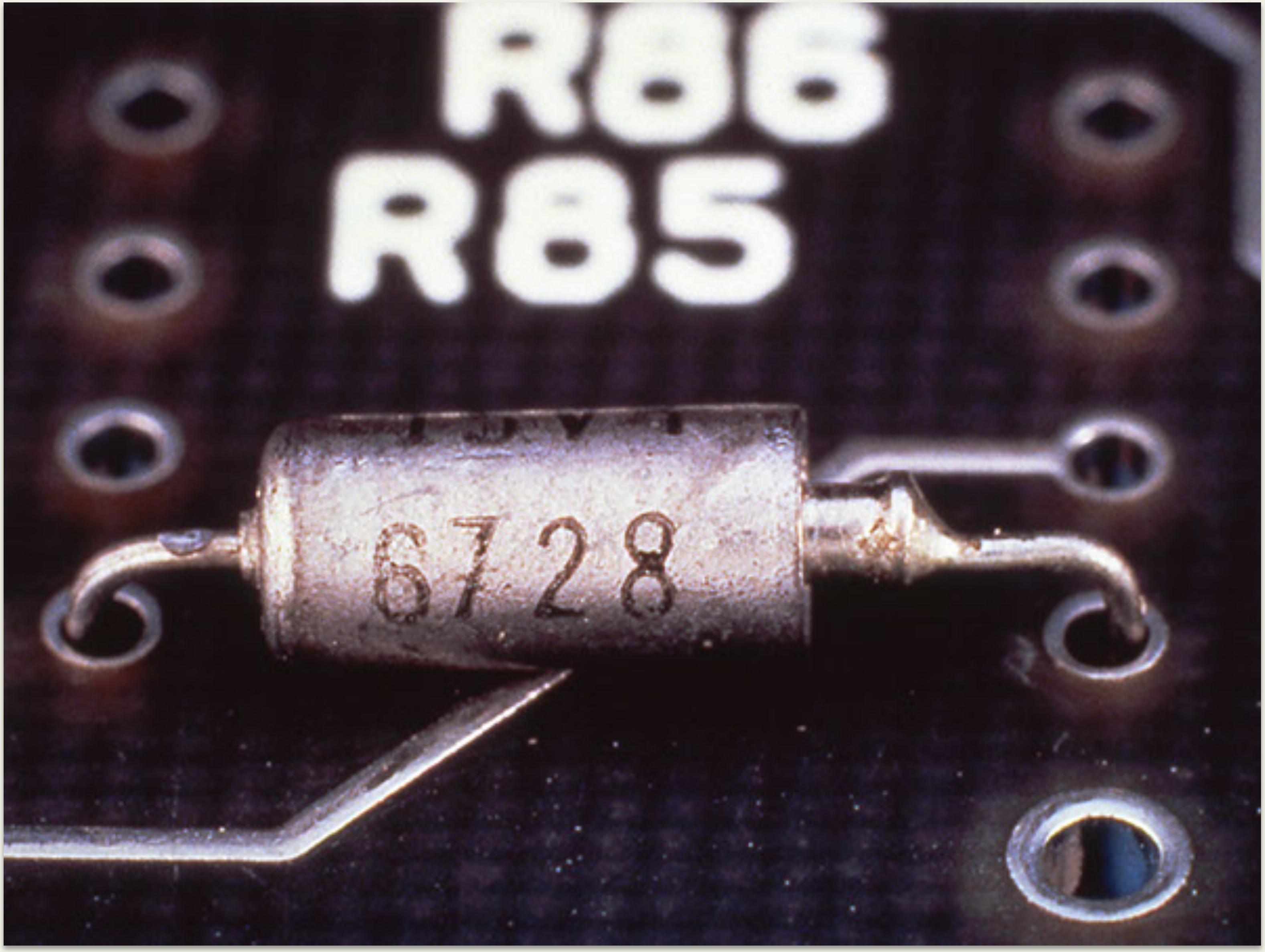
R86
R85
6728

Metal encased component mounted over circuitry - Reject

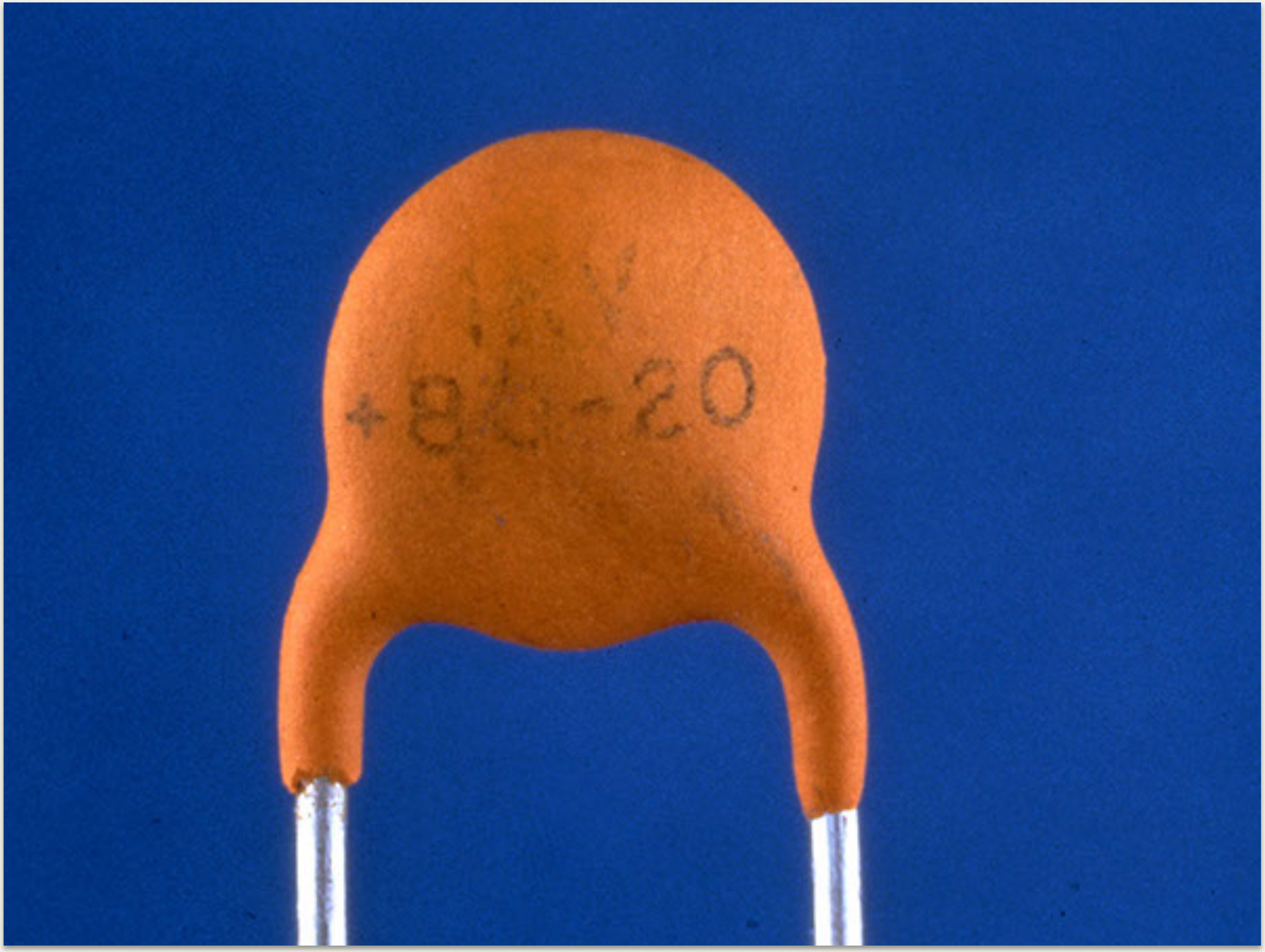

Markings not discernible - Reject

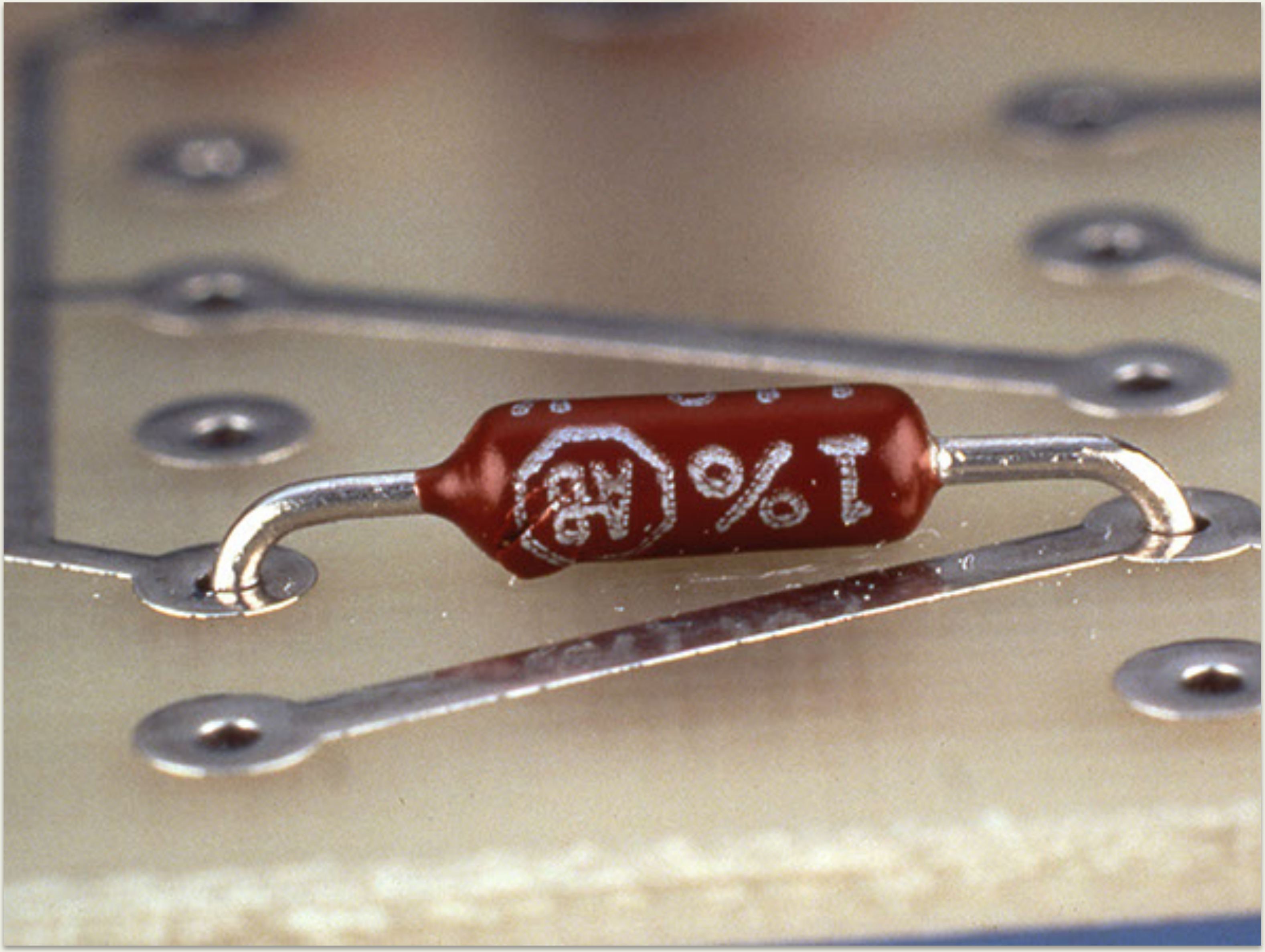

Damaged part - Reject

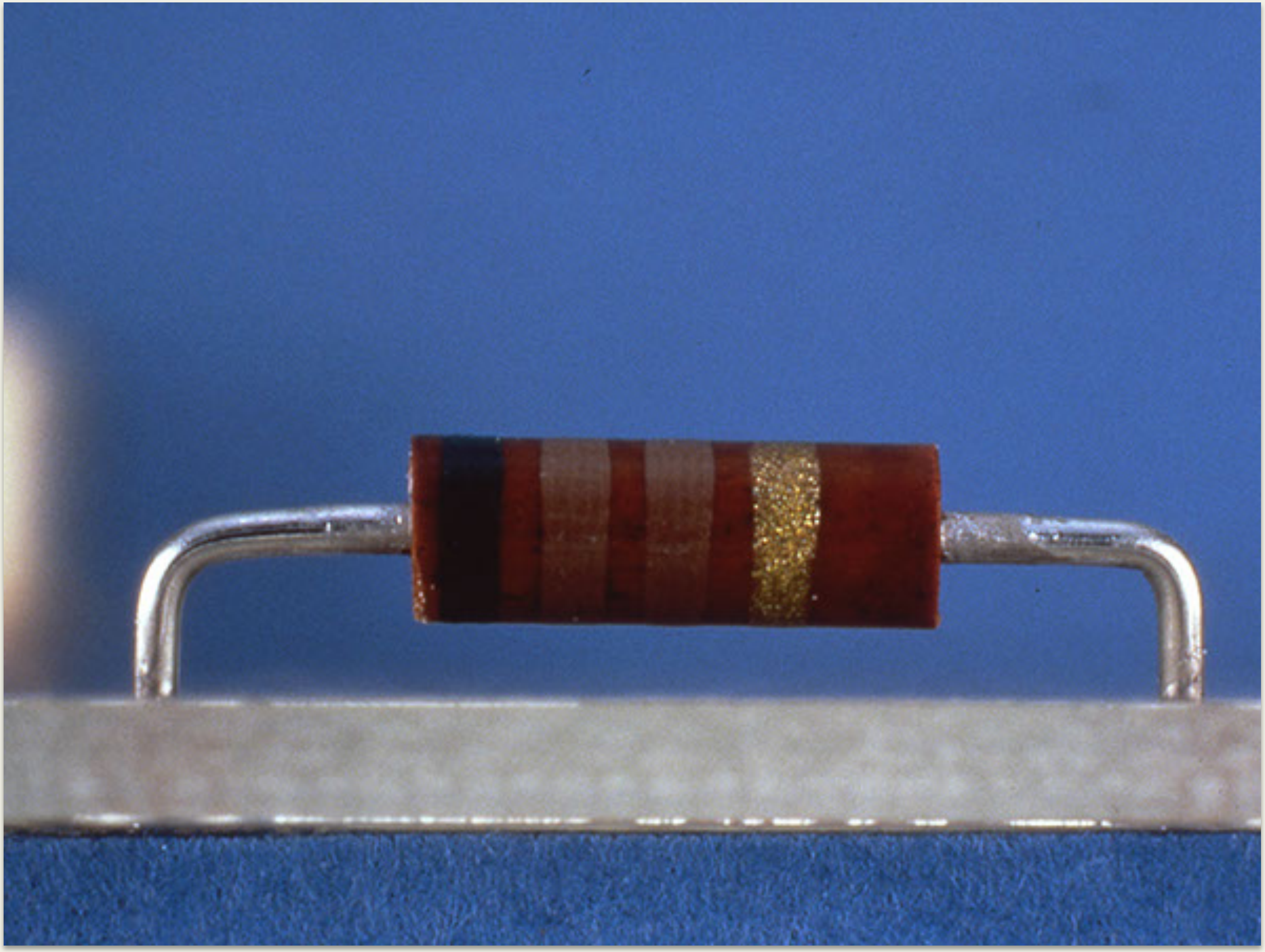

Component mounting - Reject

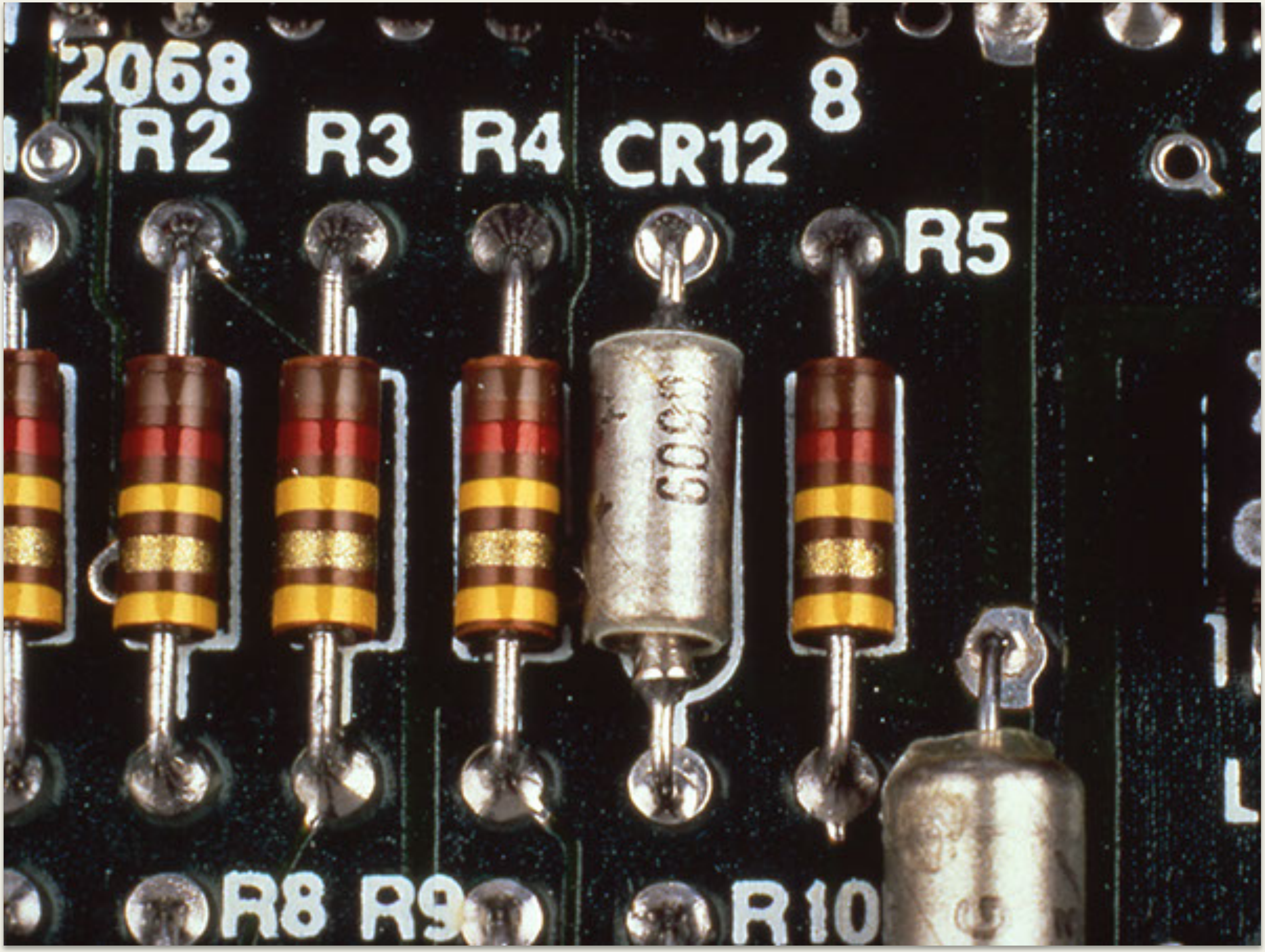
2068
R2
R3
R4
CR12
8
R5
R8
R9
R10

Incorrect component - Reject

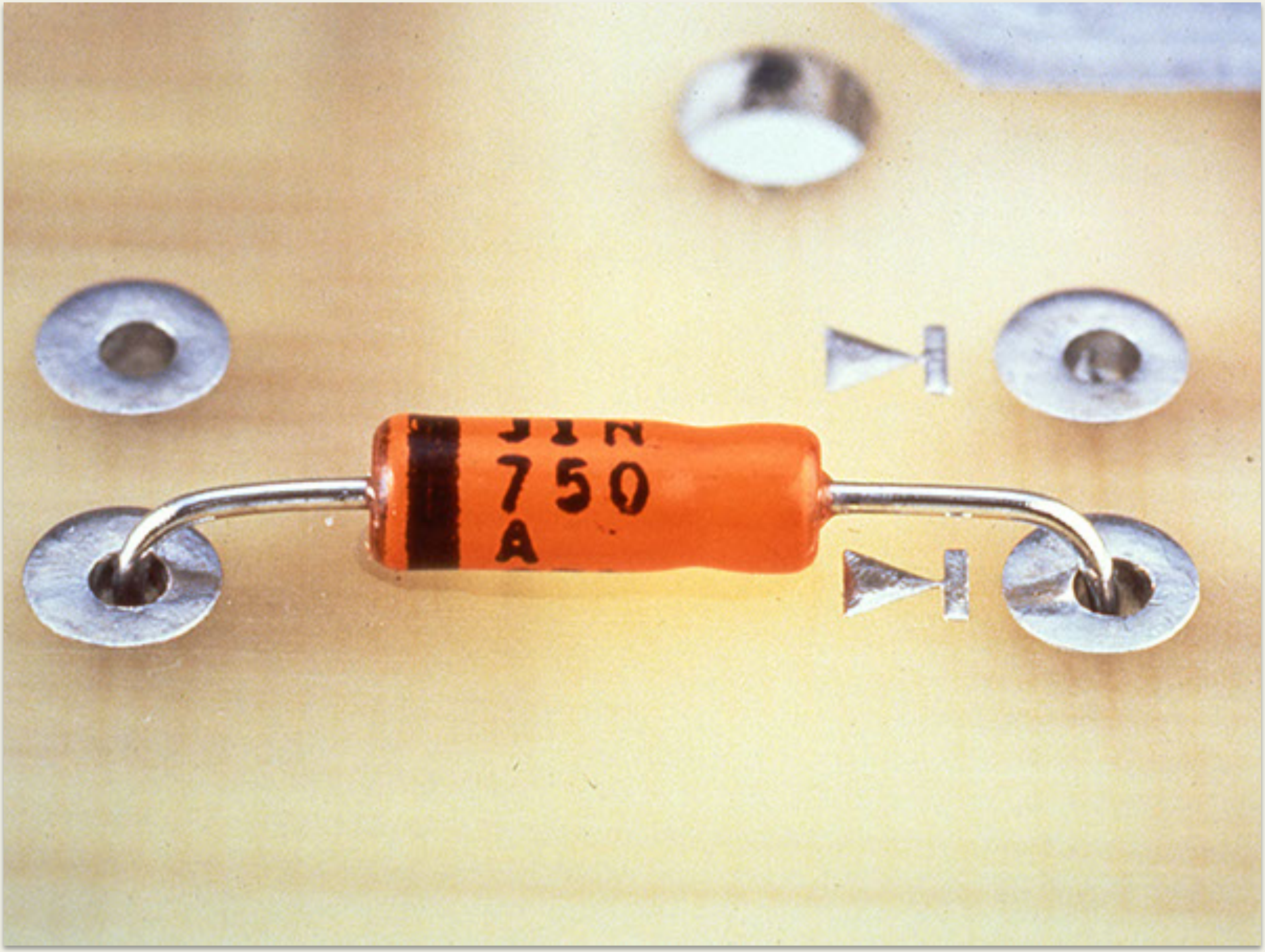

Reversed polarity - Reject

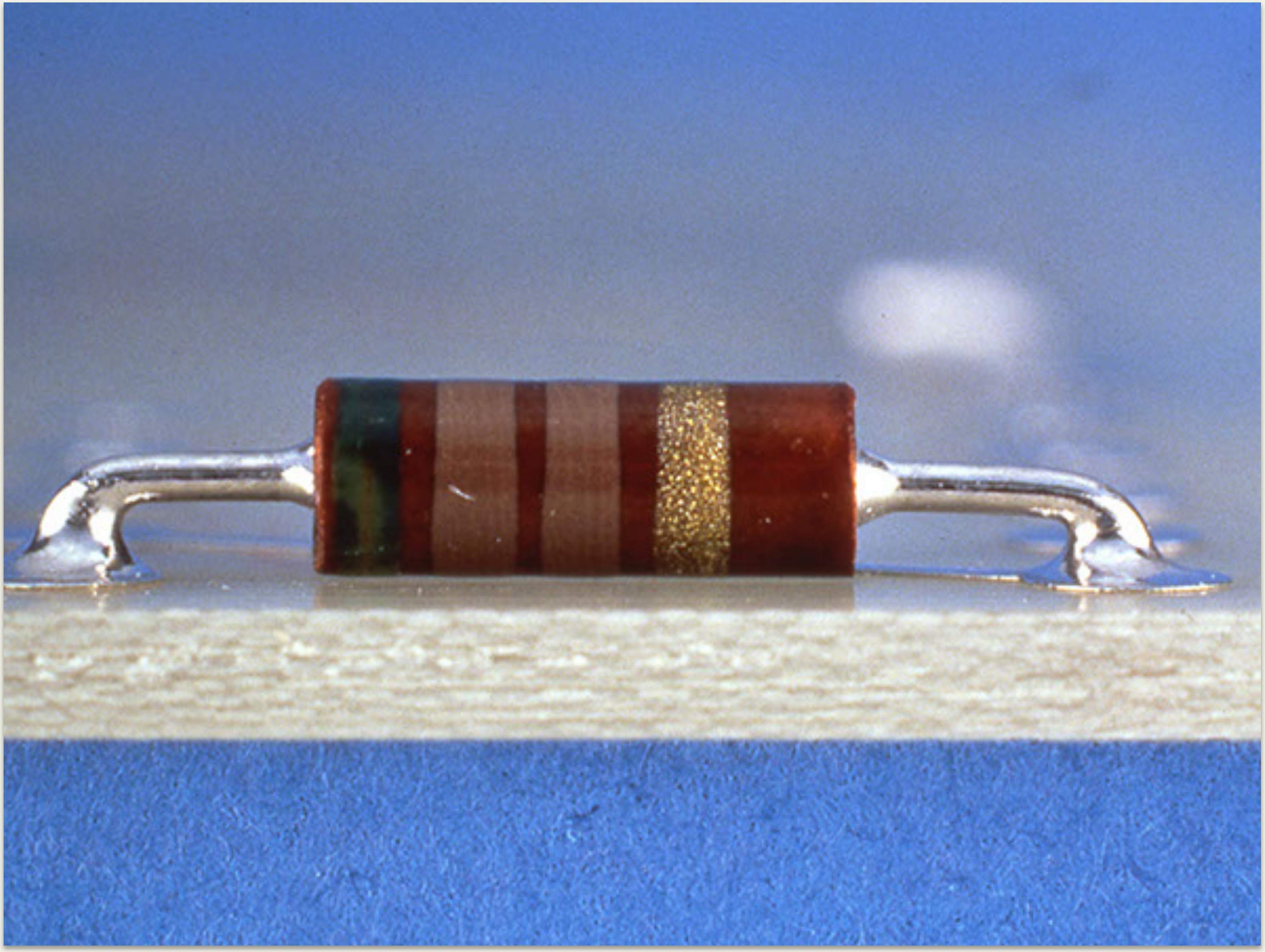

Solder in bend radii – ‘’’Reject

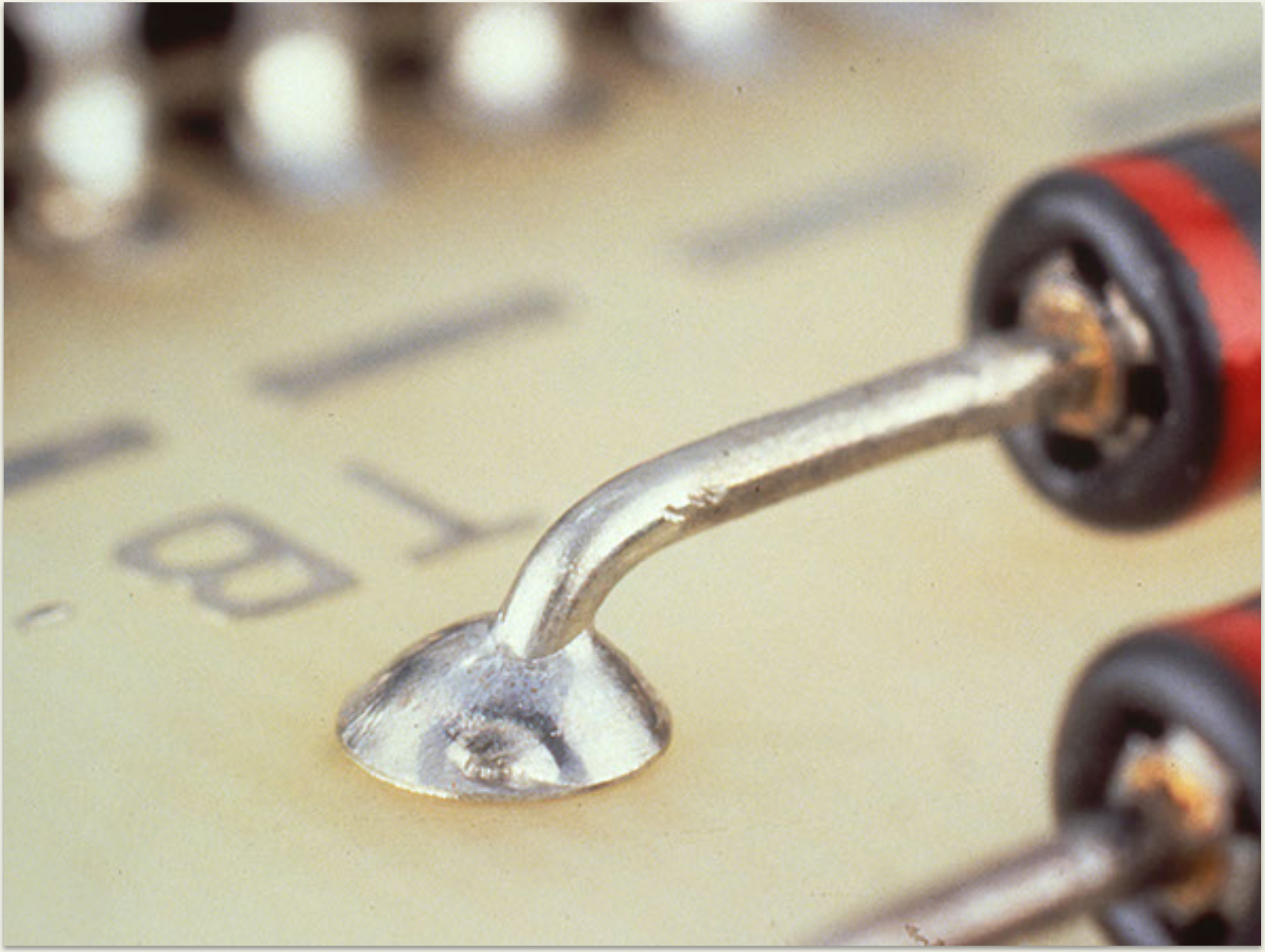

Poor wetting/reflow stress lines - Reject

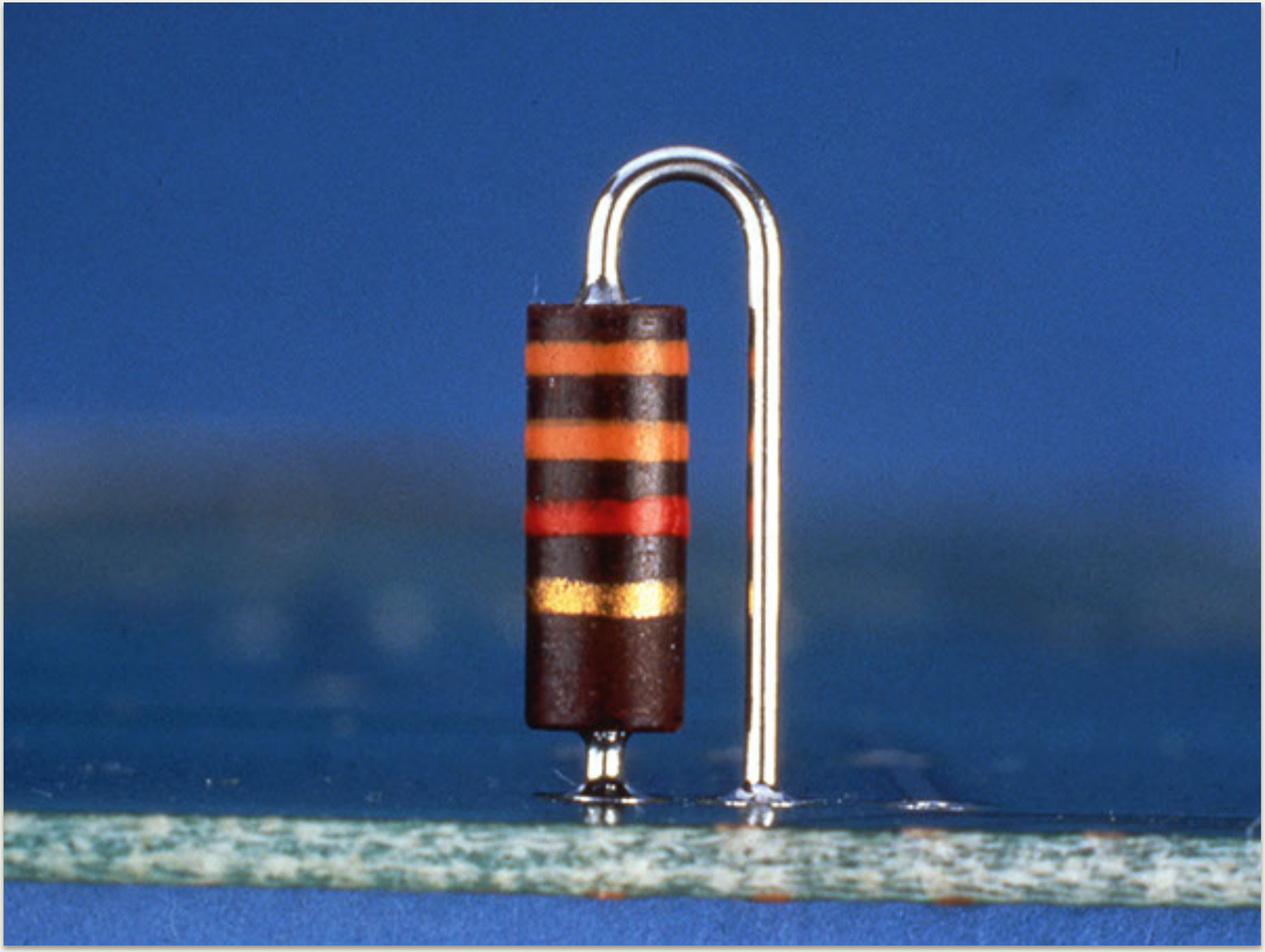

Vertical mounted axial lead component - Accept

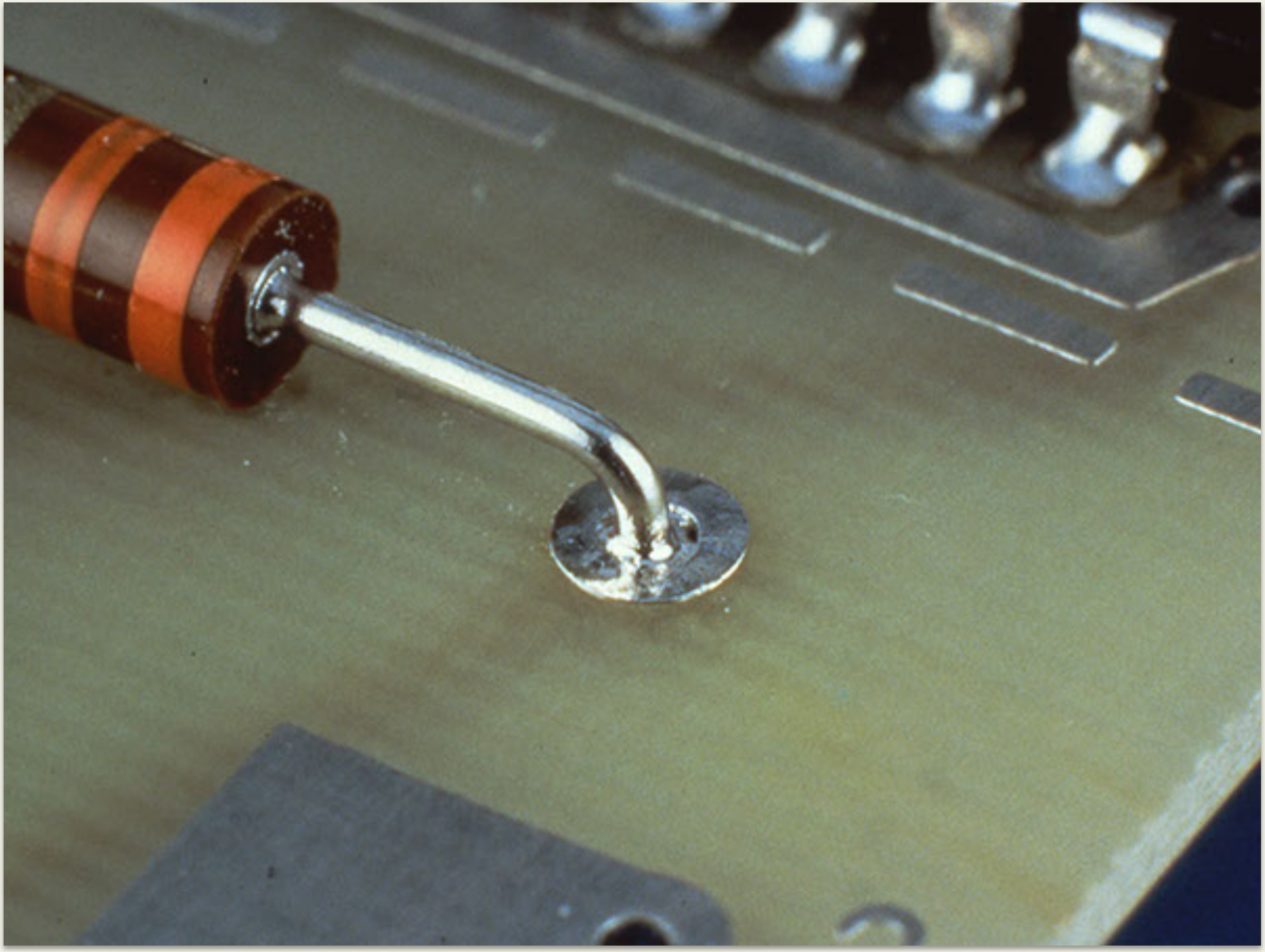

Poor solder flow through plated through hole - Reject

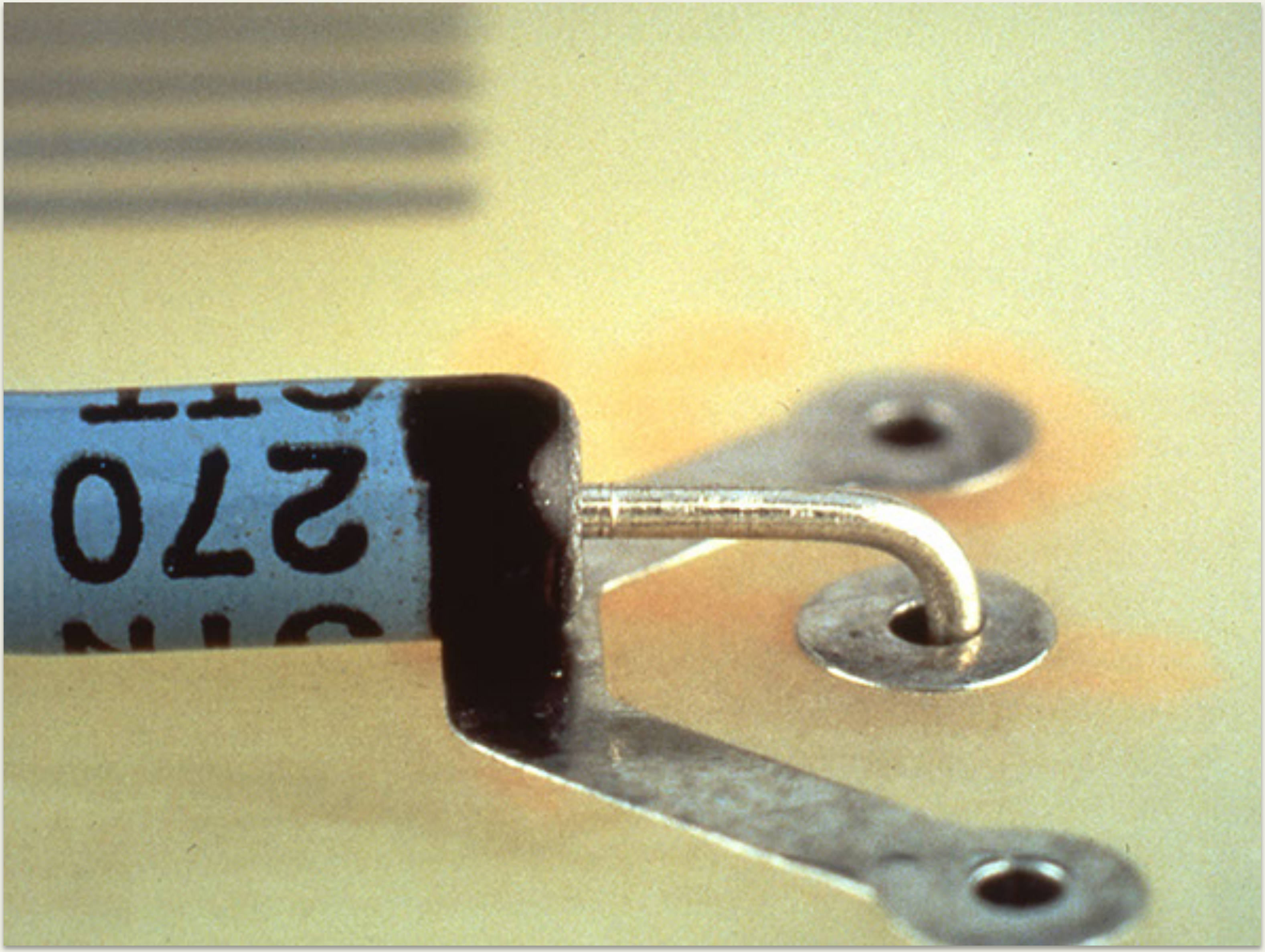

Glass body component not sleeved - Reject

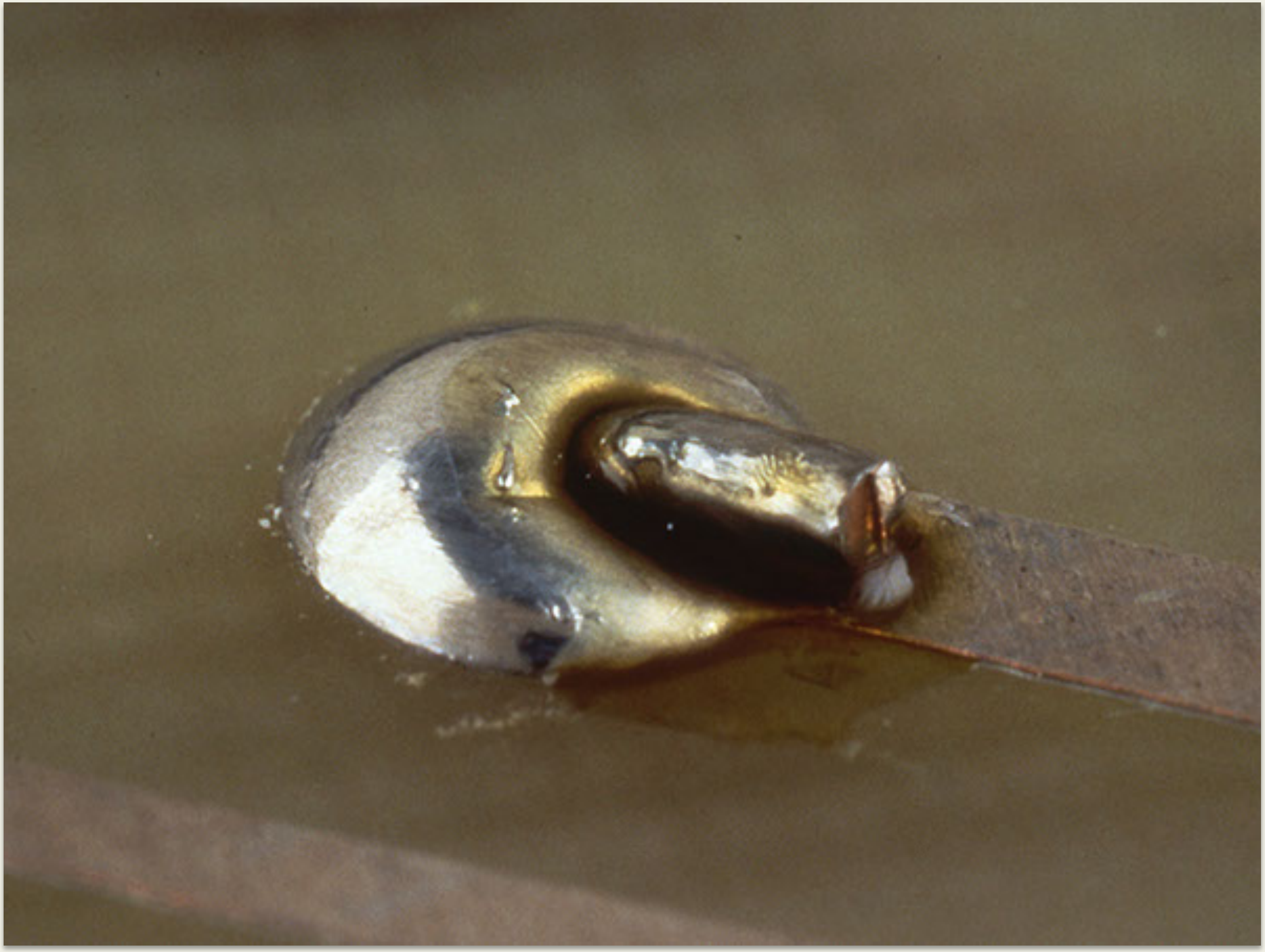

Rosin - Reject

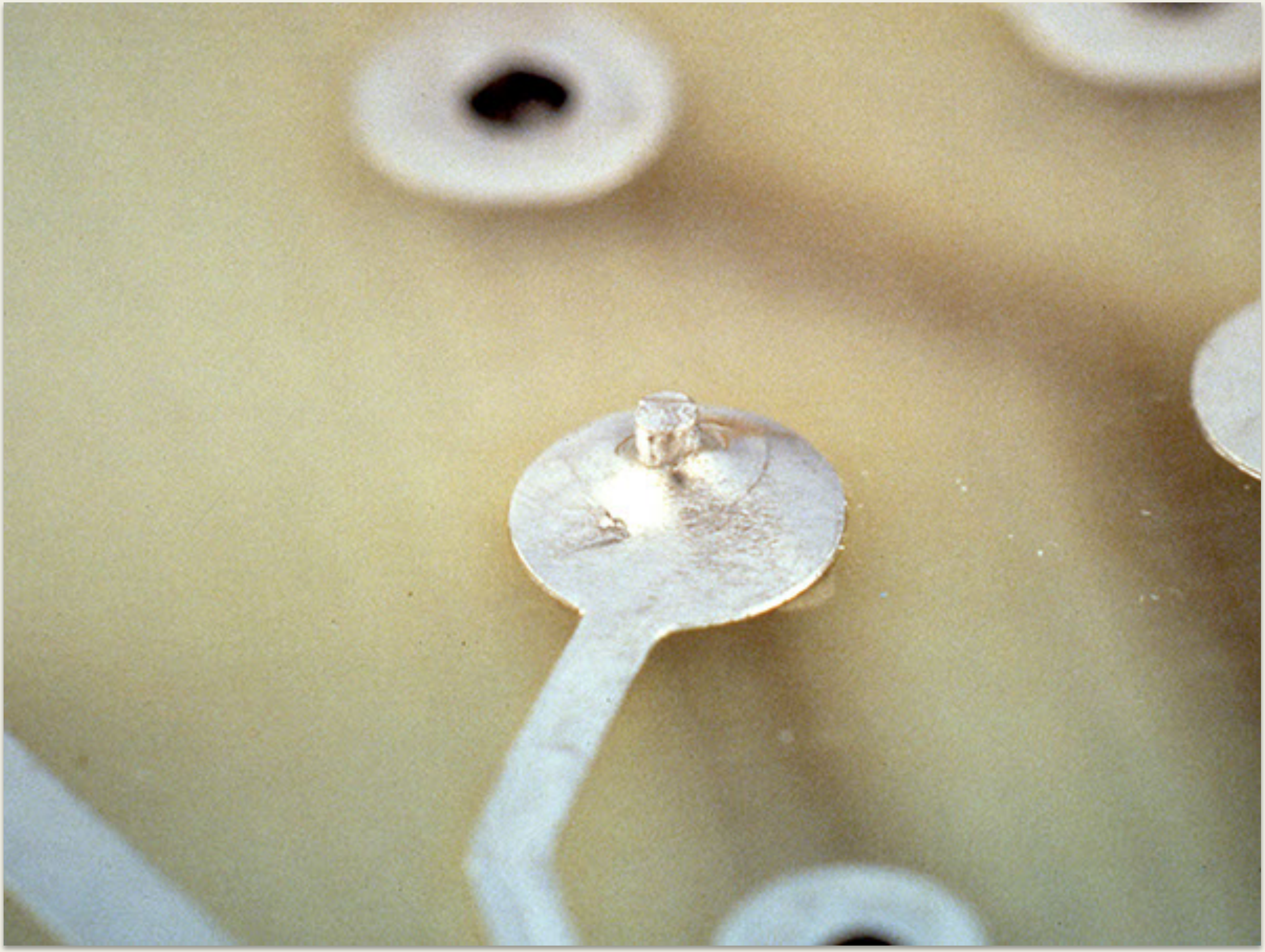

Pit - Reject

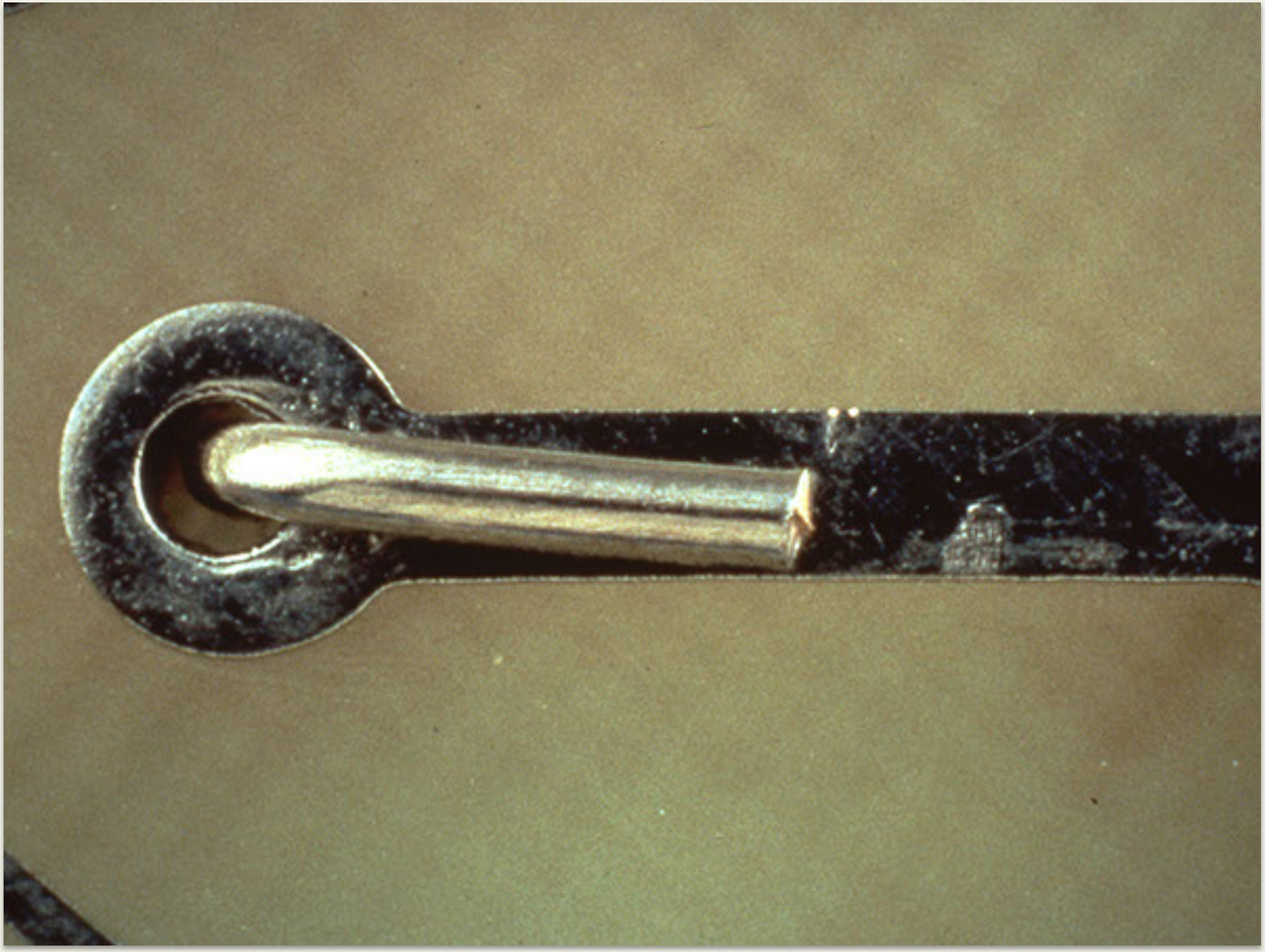

Excessive lead length, damaged trace - Reject

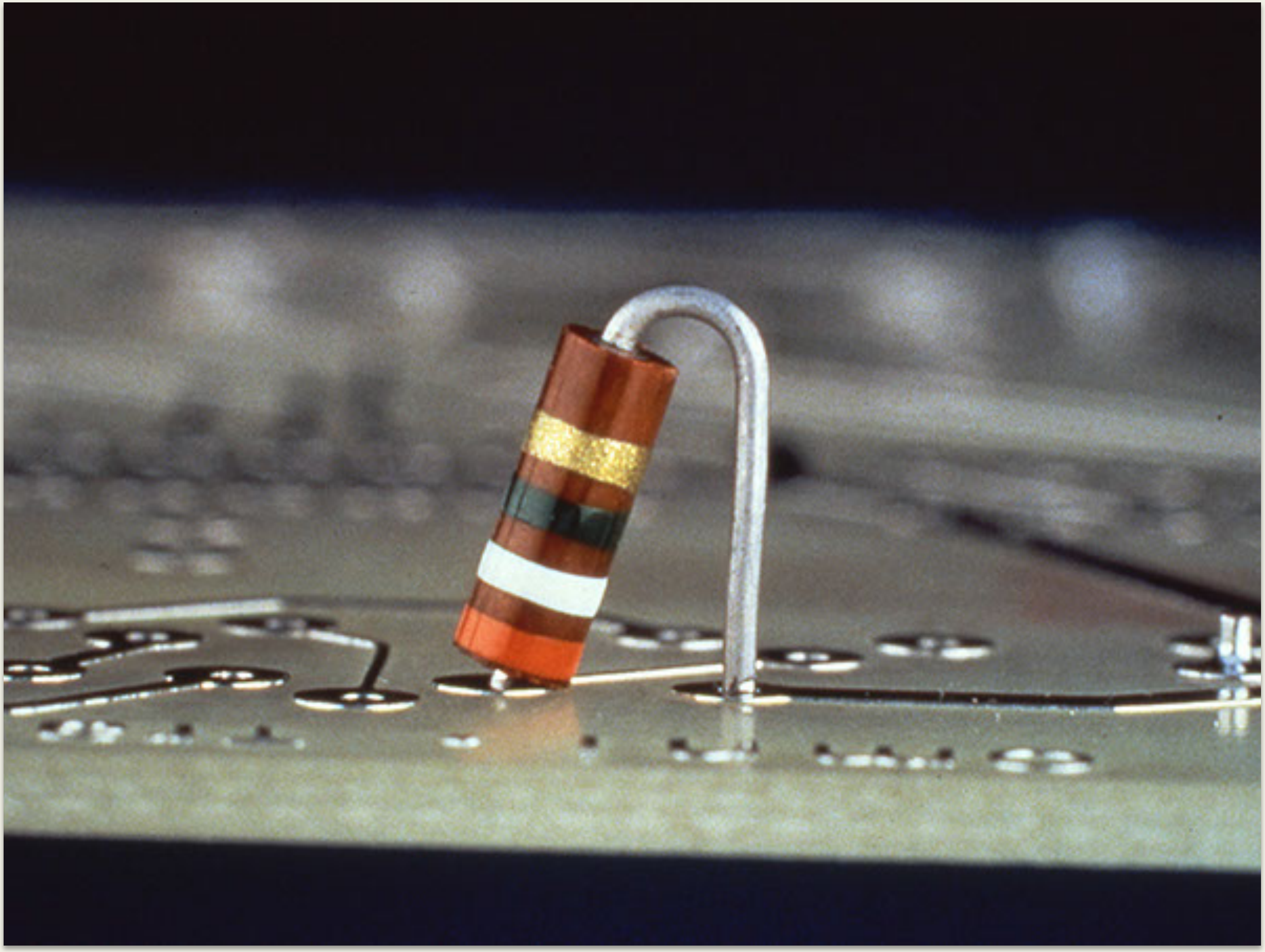

Improper component mounting - Reject

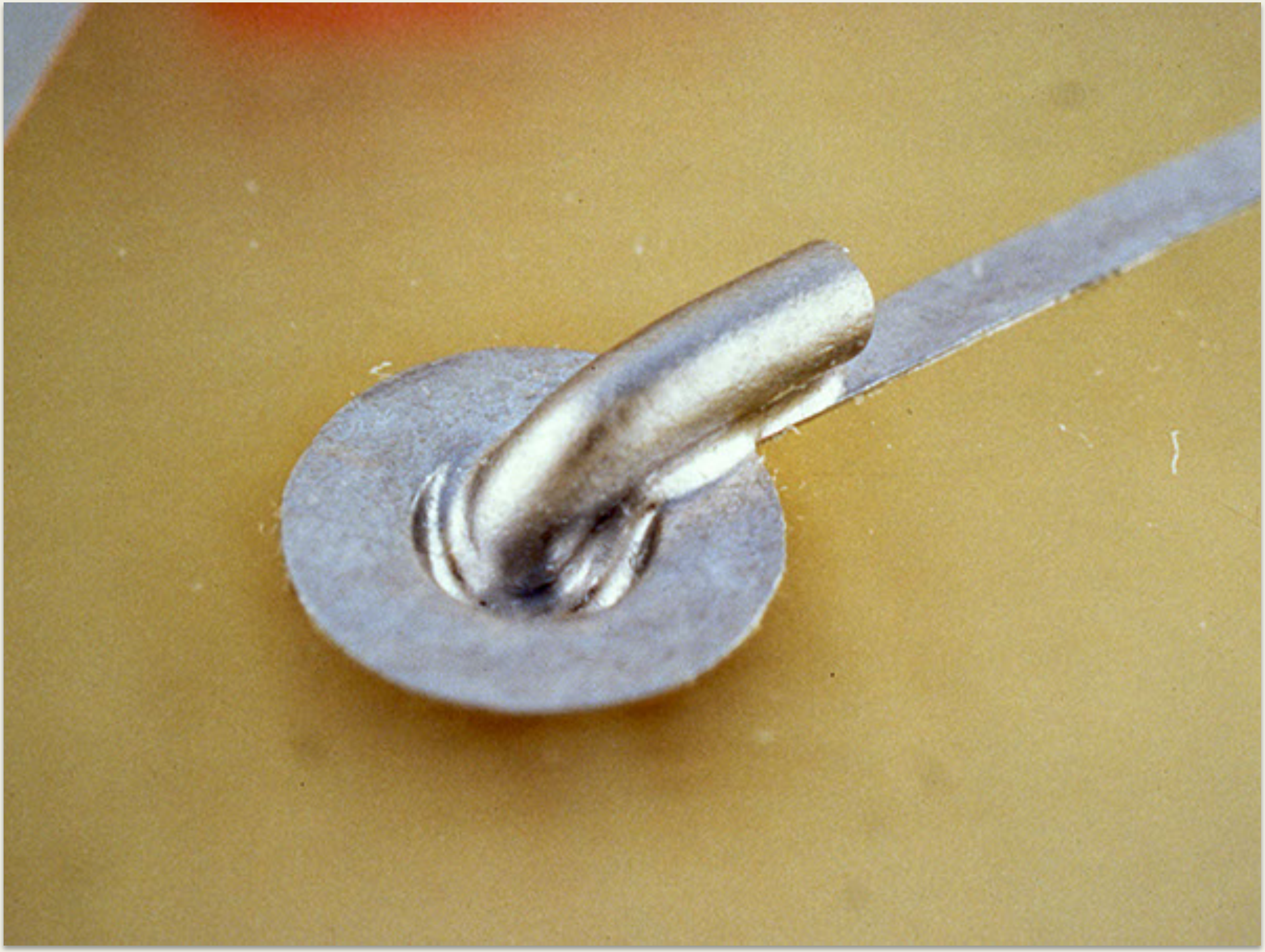

Insufficient solder - Reject

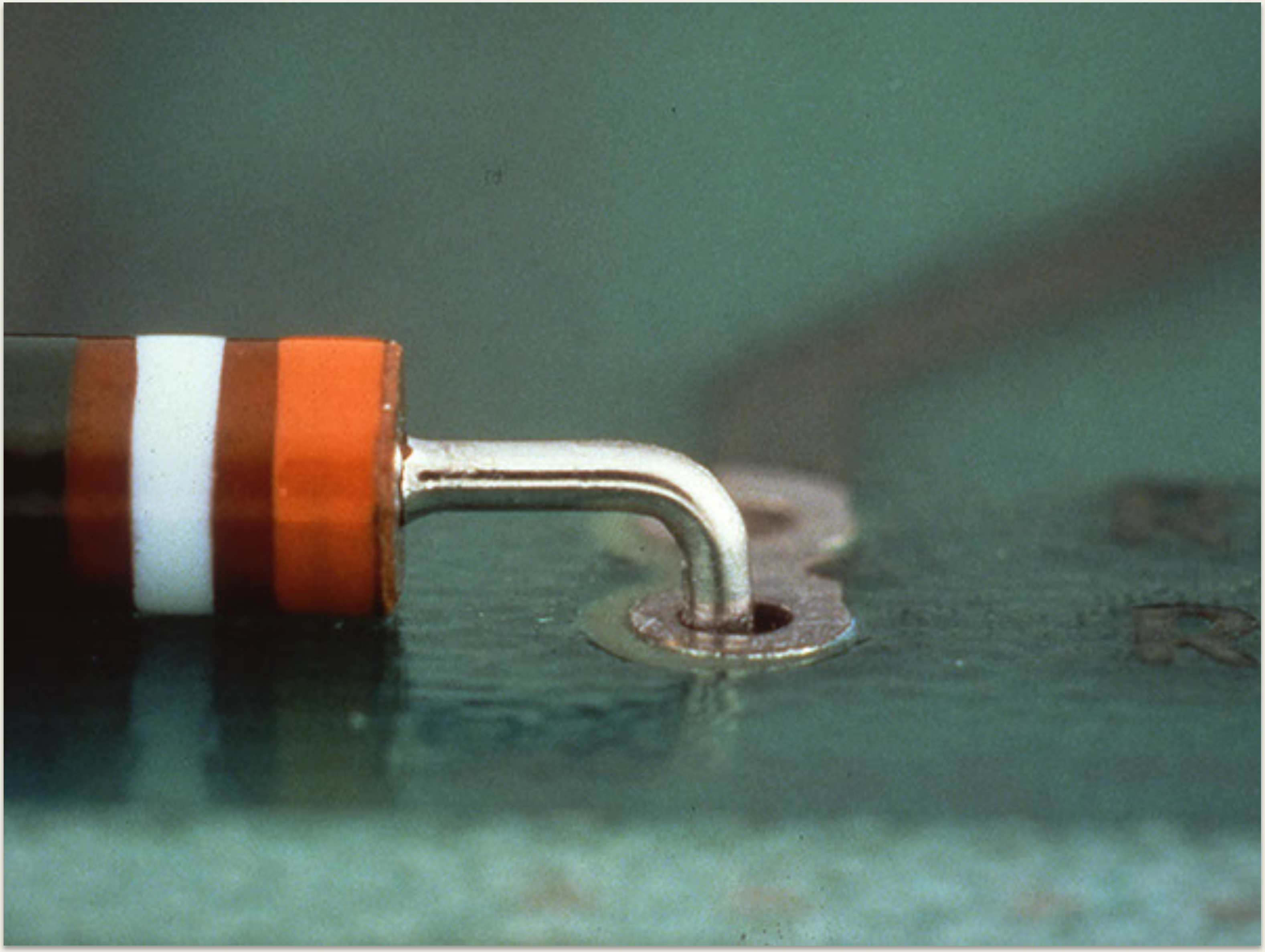

Correct lead bend radius - Accept

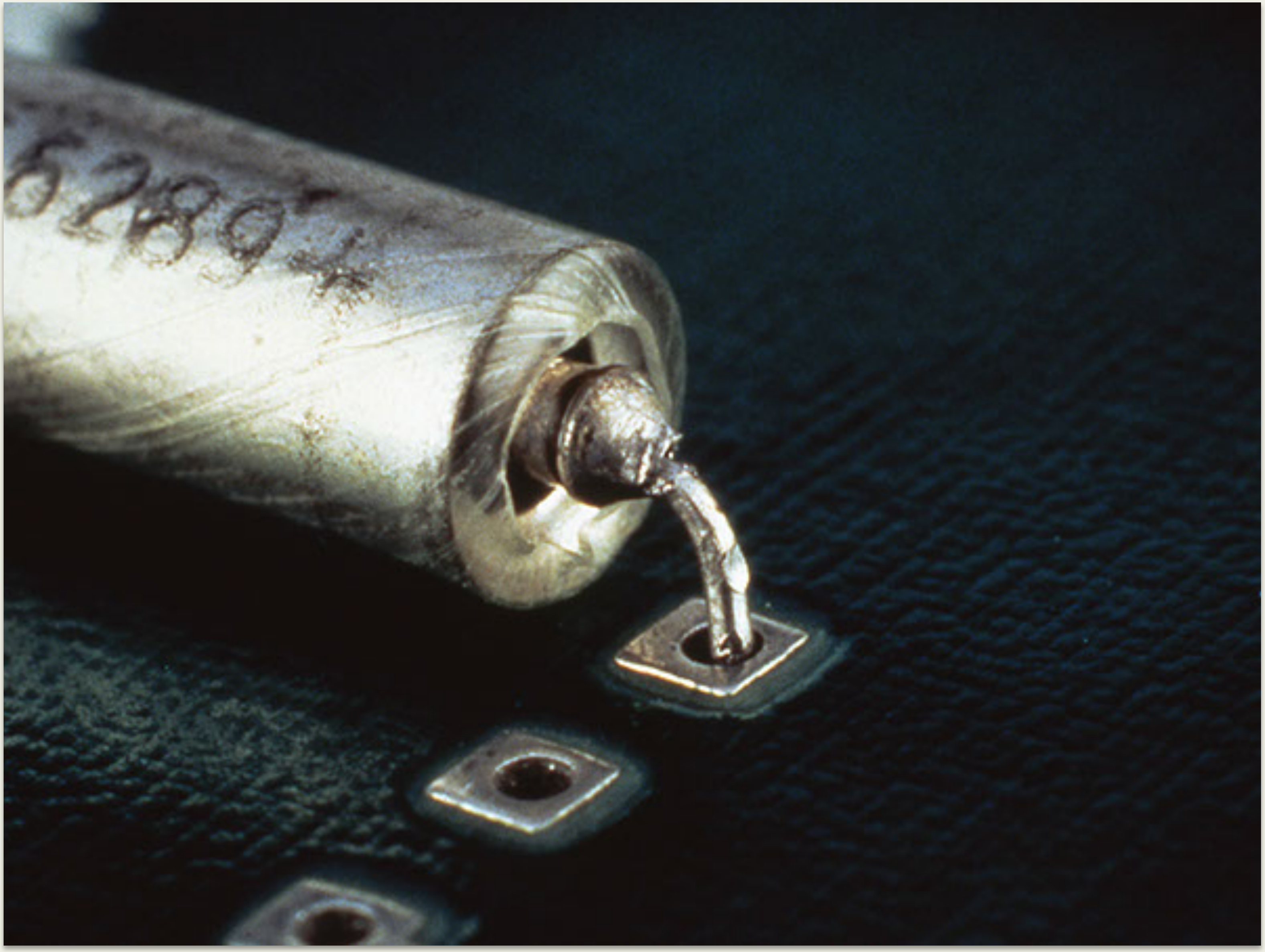

Distance from weld bead to bend - Reject

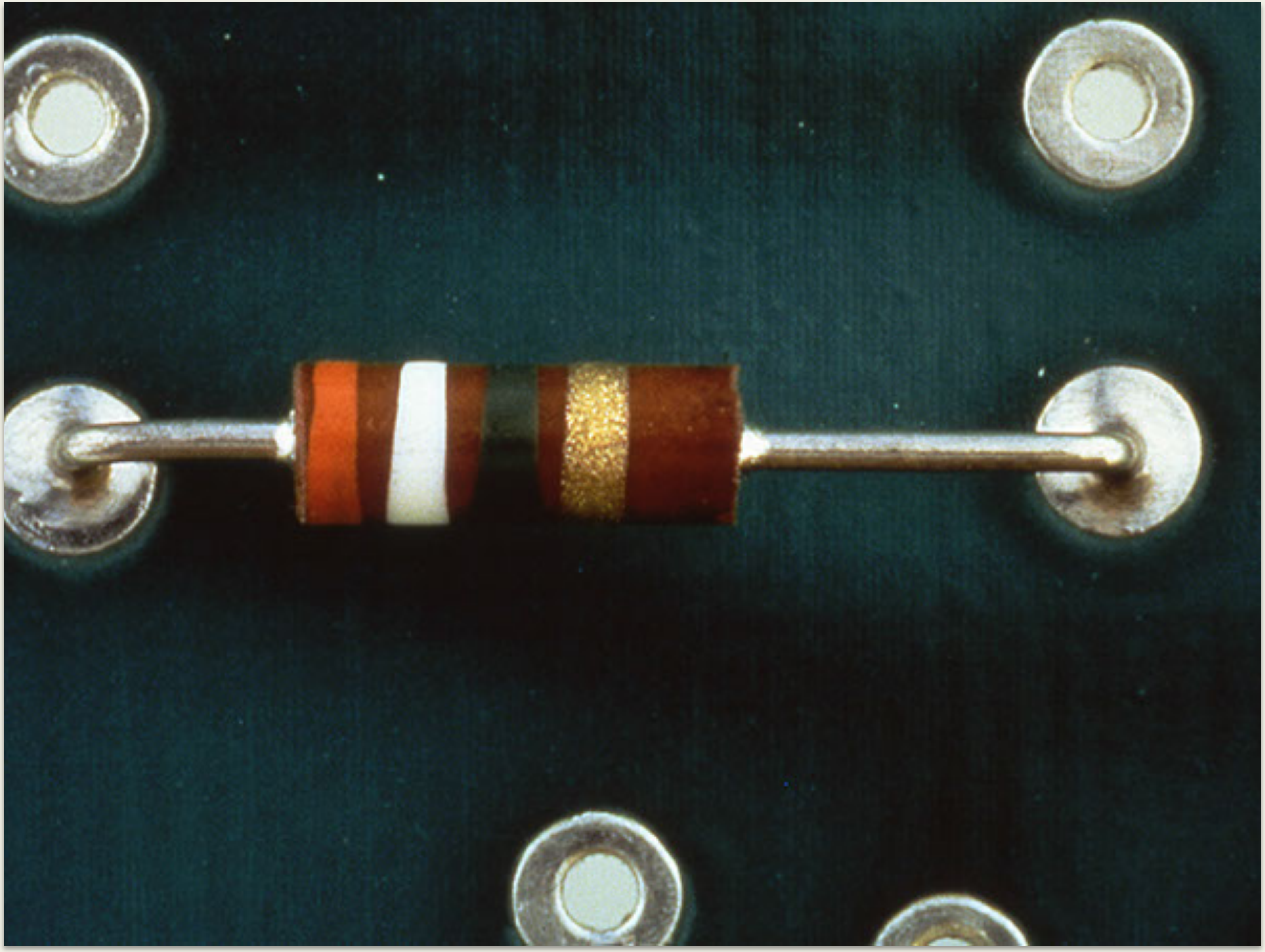

Component not centered - Reject

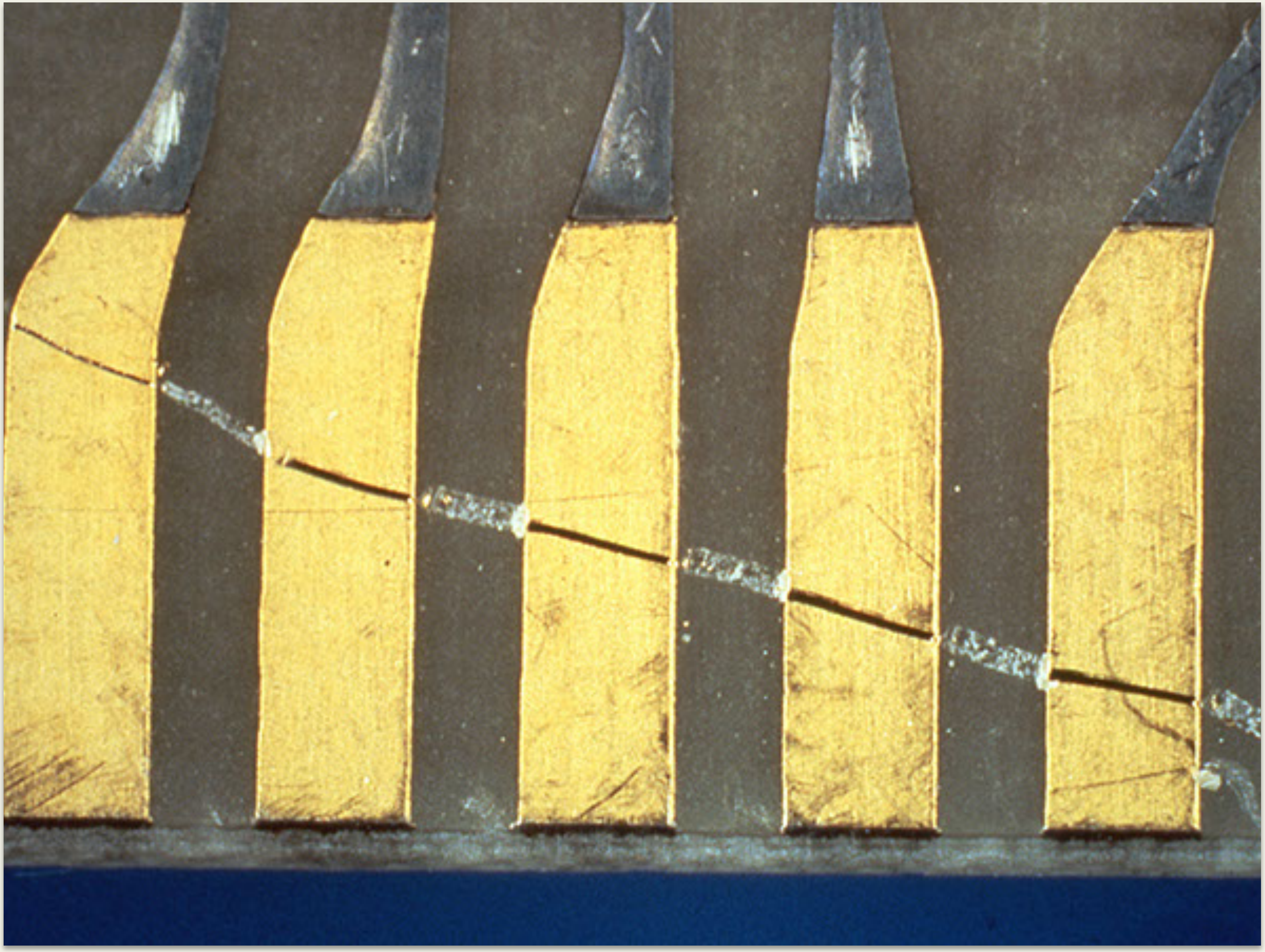

Scratches on pads - Reject

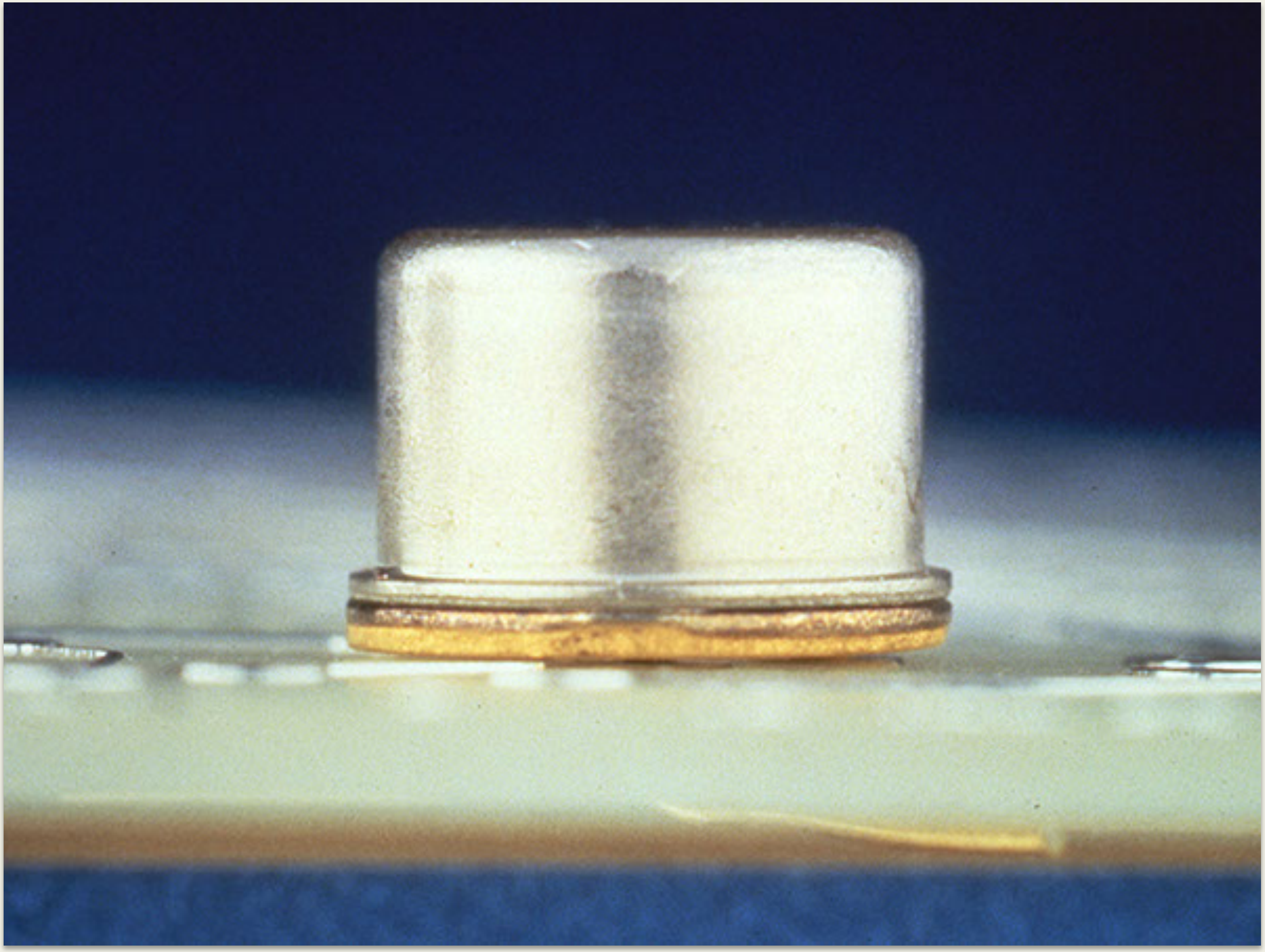

Component mounting - Reject

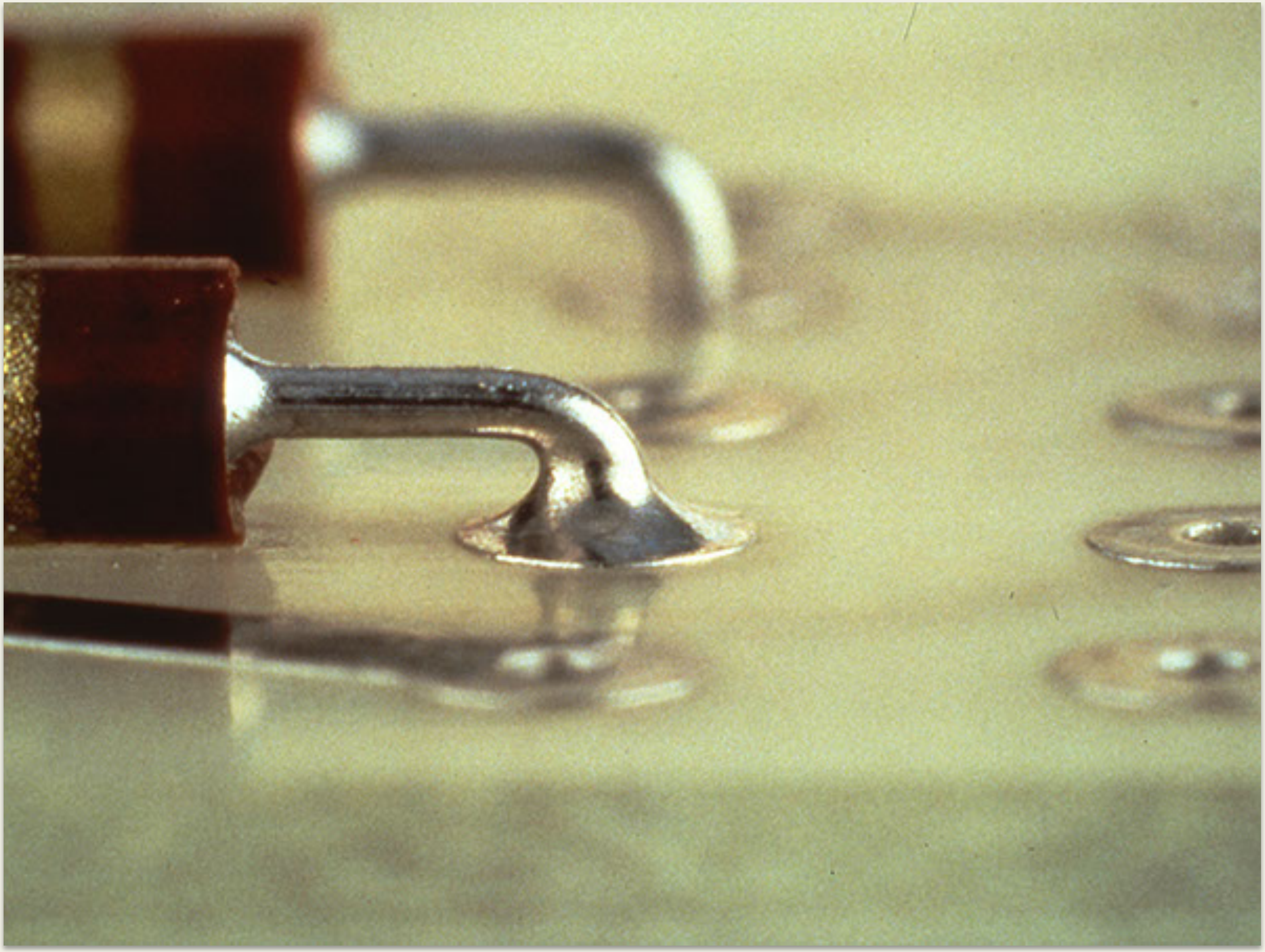

Solder in bend radius, minimum - Accept

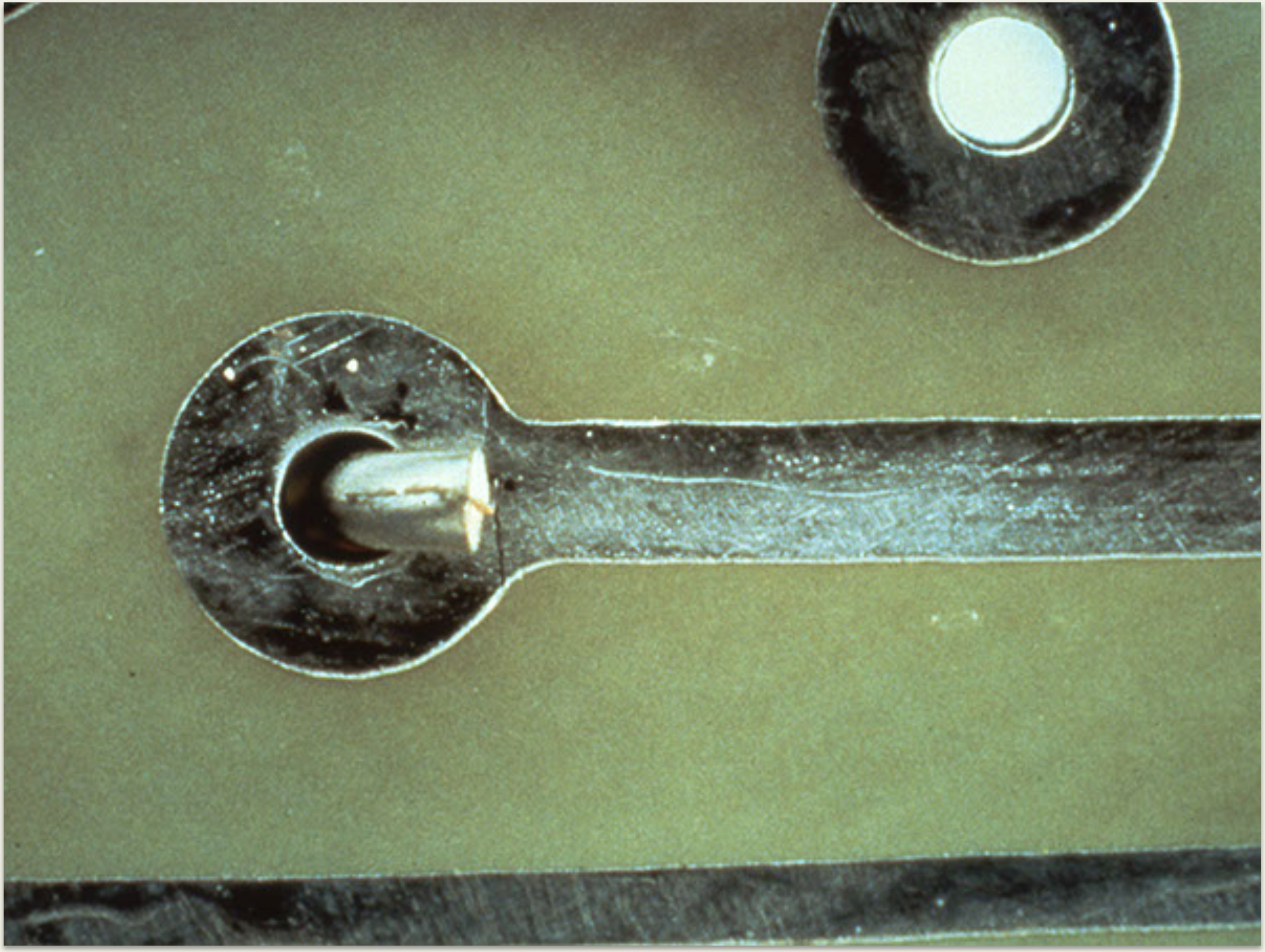

Lead length, insufficient - Reject

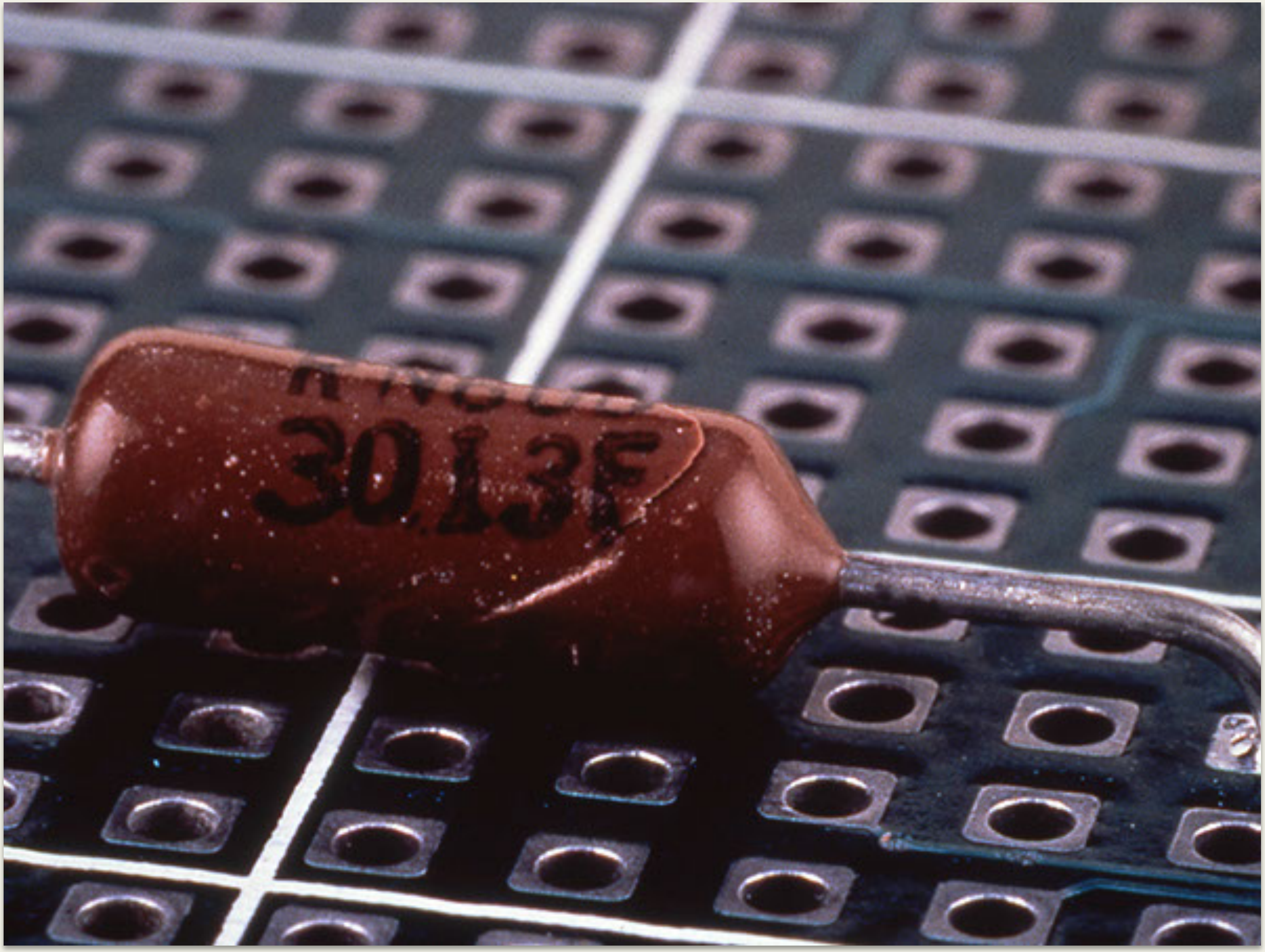

Broken, damaged part - Reject

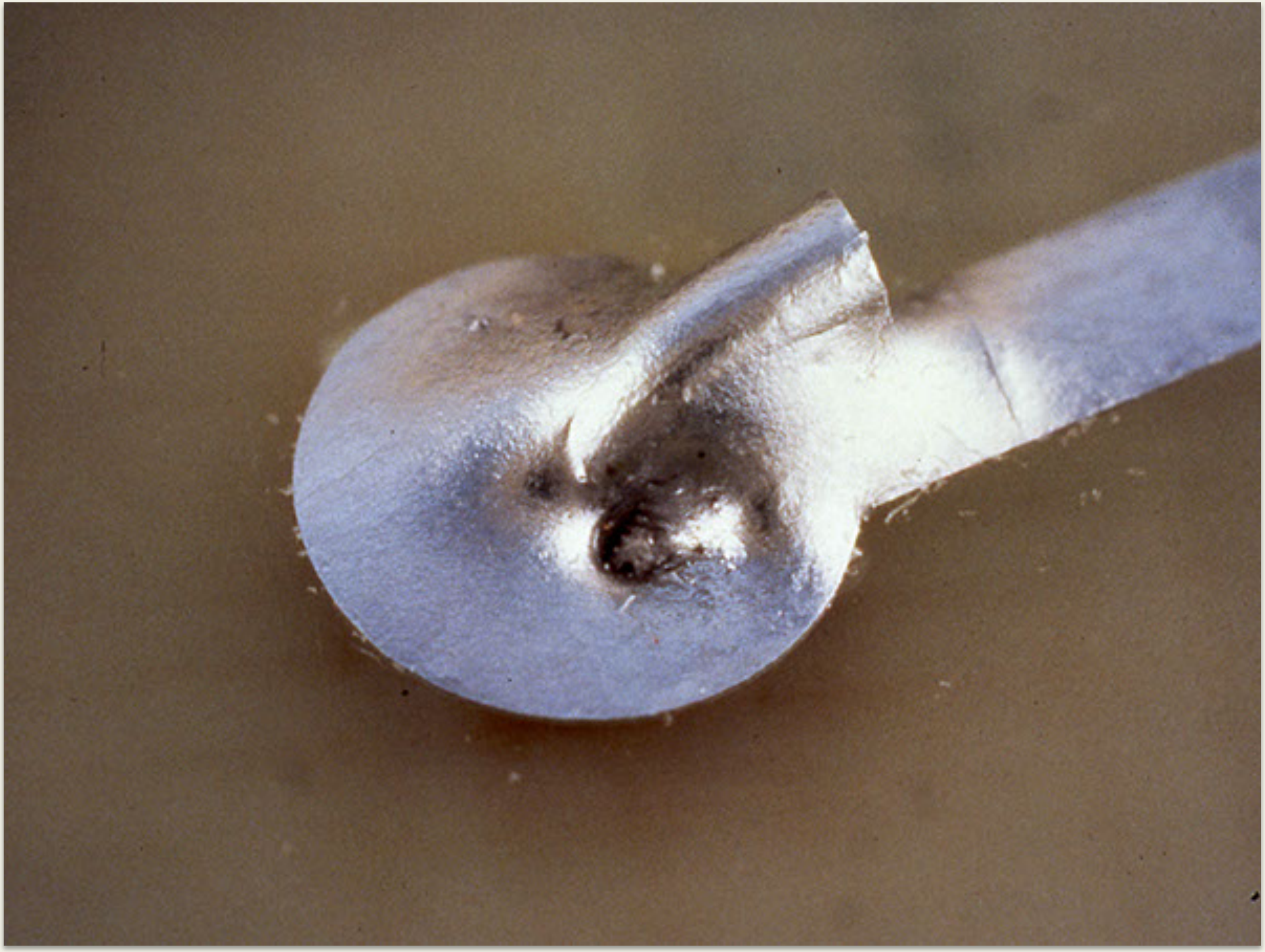

Pit, void - Reject

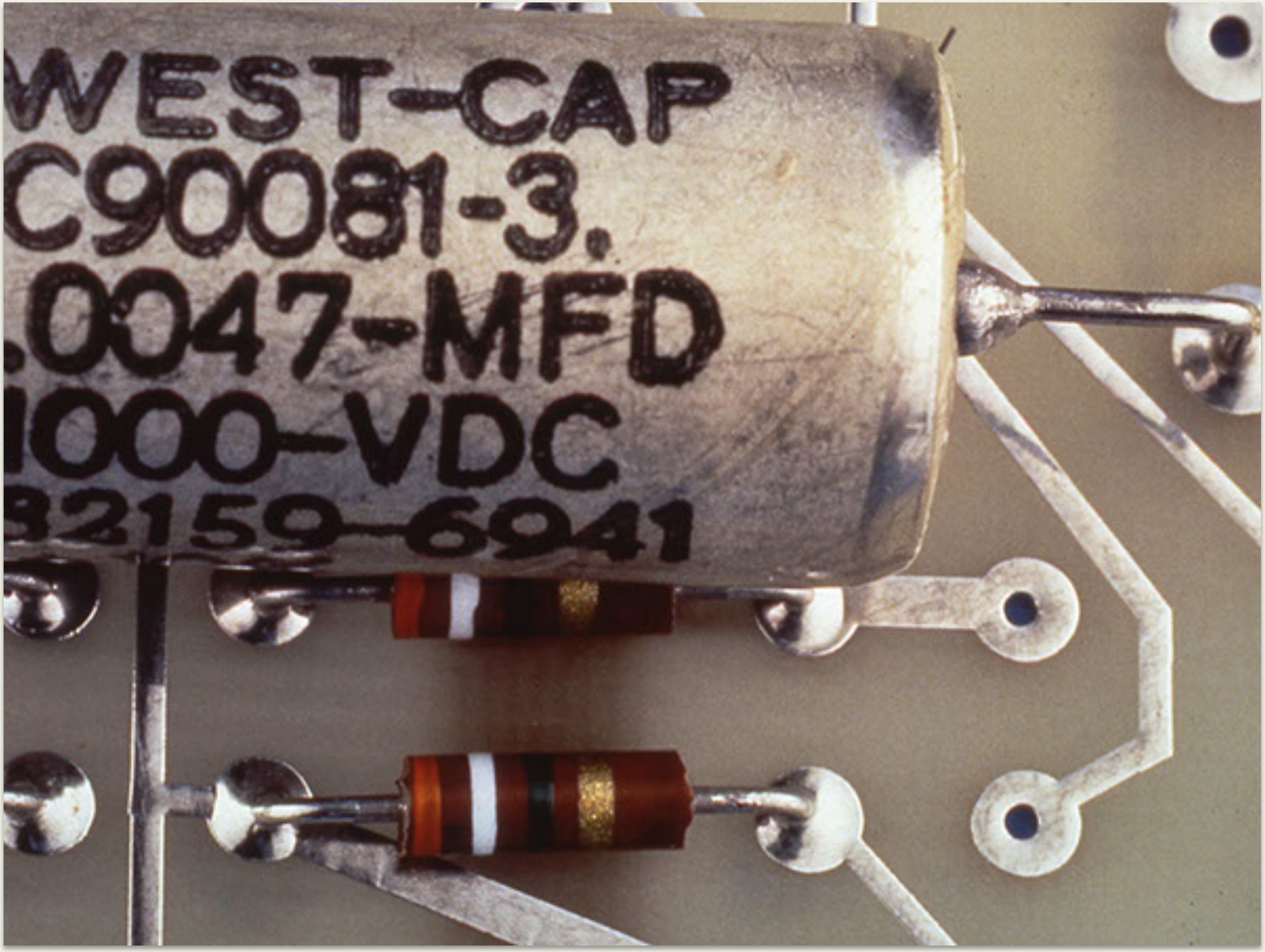
WEST-CAP
C90081-3.
.0047-MFD
1000-VDC
82159-6941

Large component obscures termination of another

Part - Reject

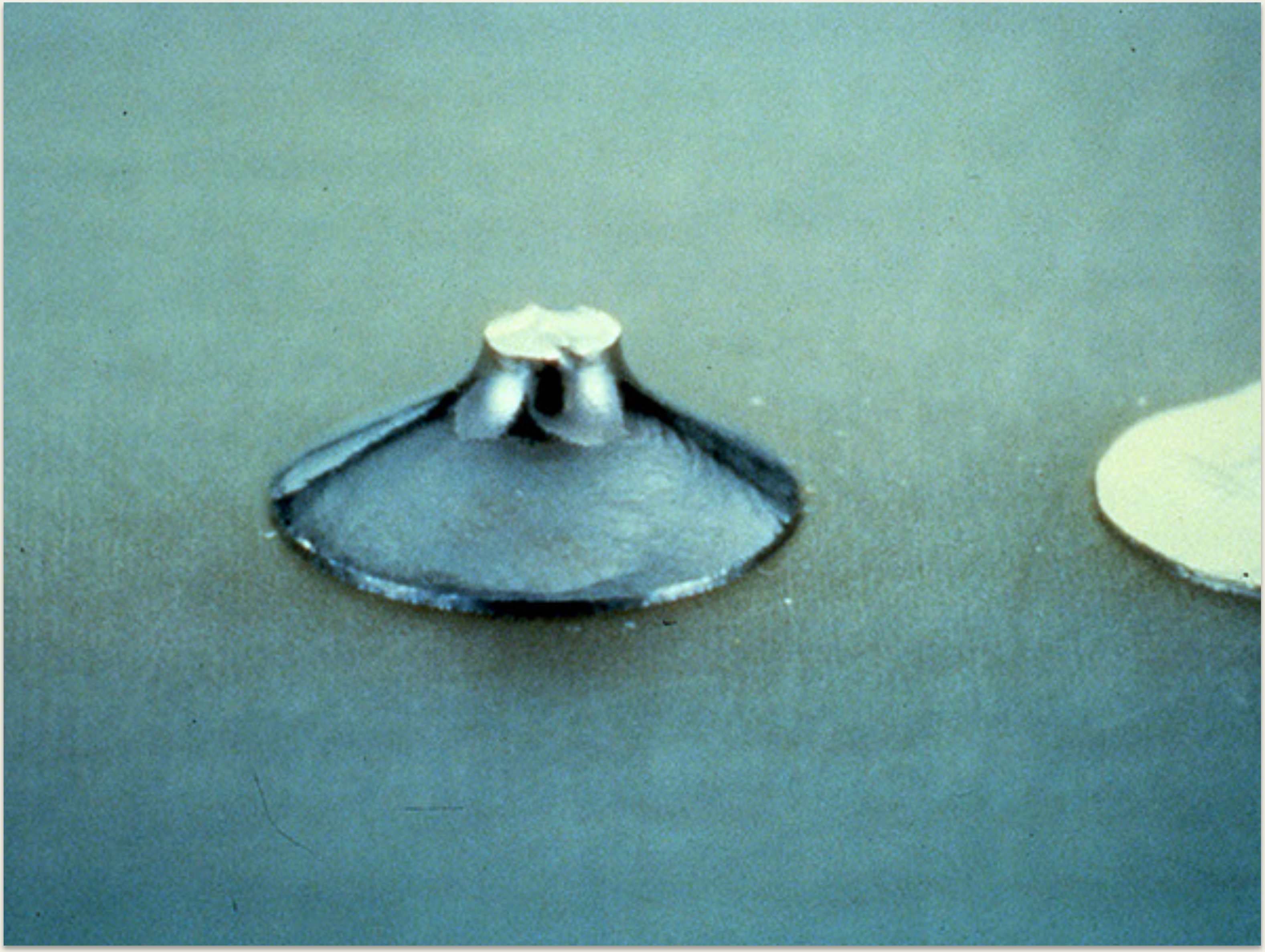

Maximum solder (stud) - Accept

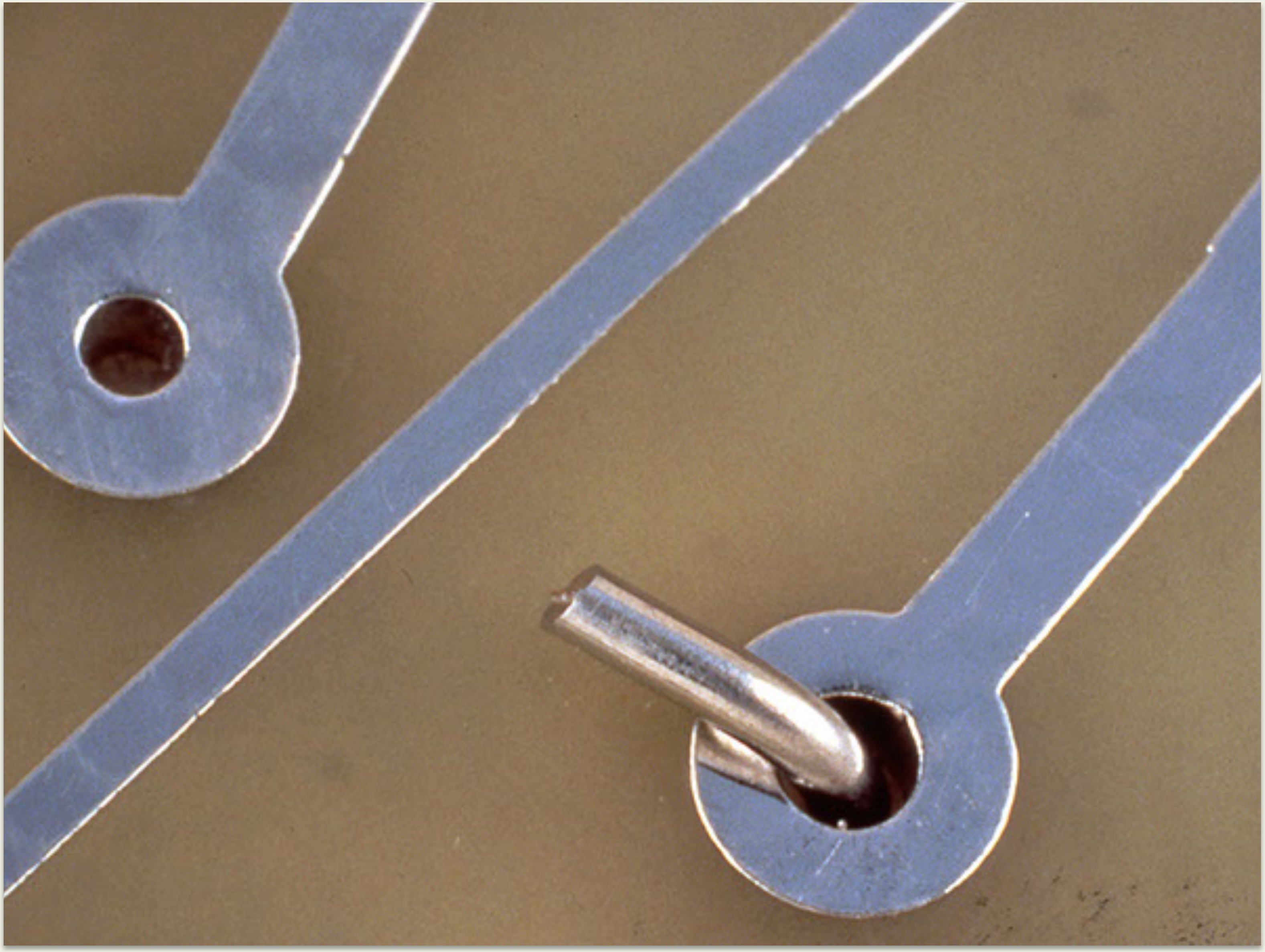

Improper lead clinched - Reject

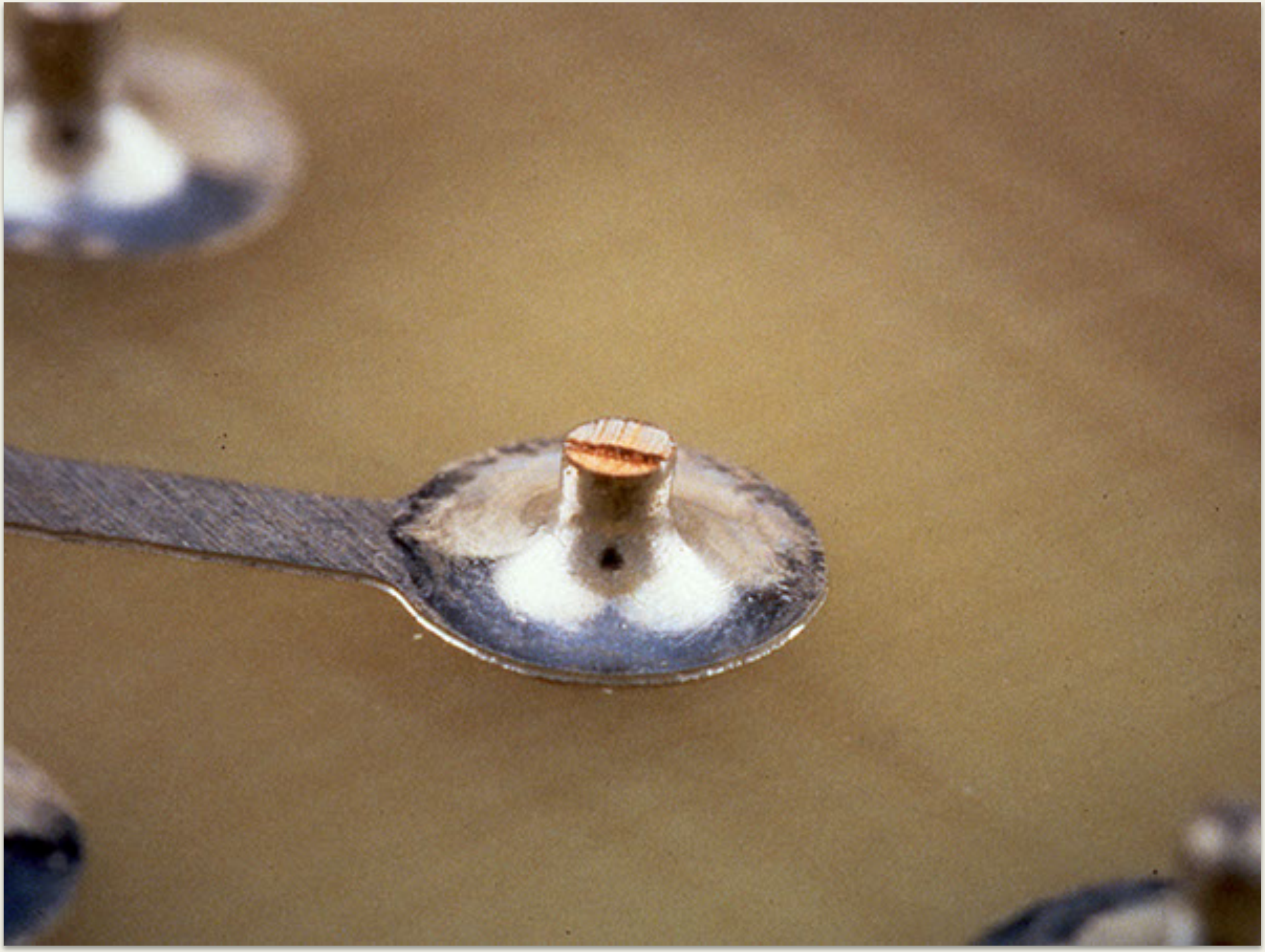

Exposed copper - Reject

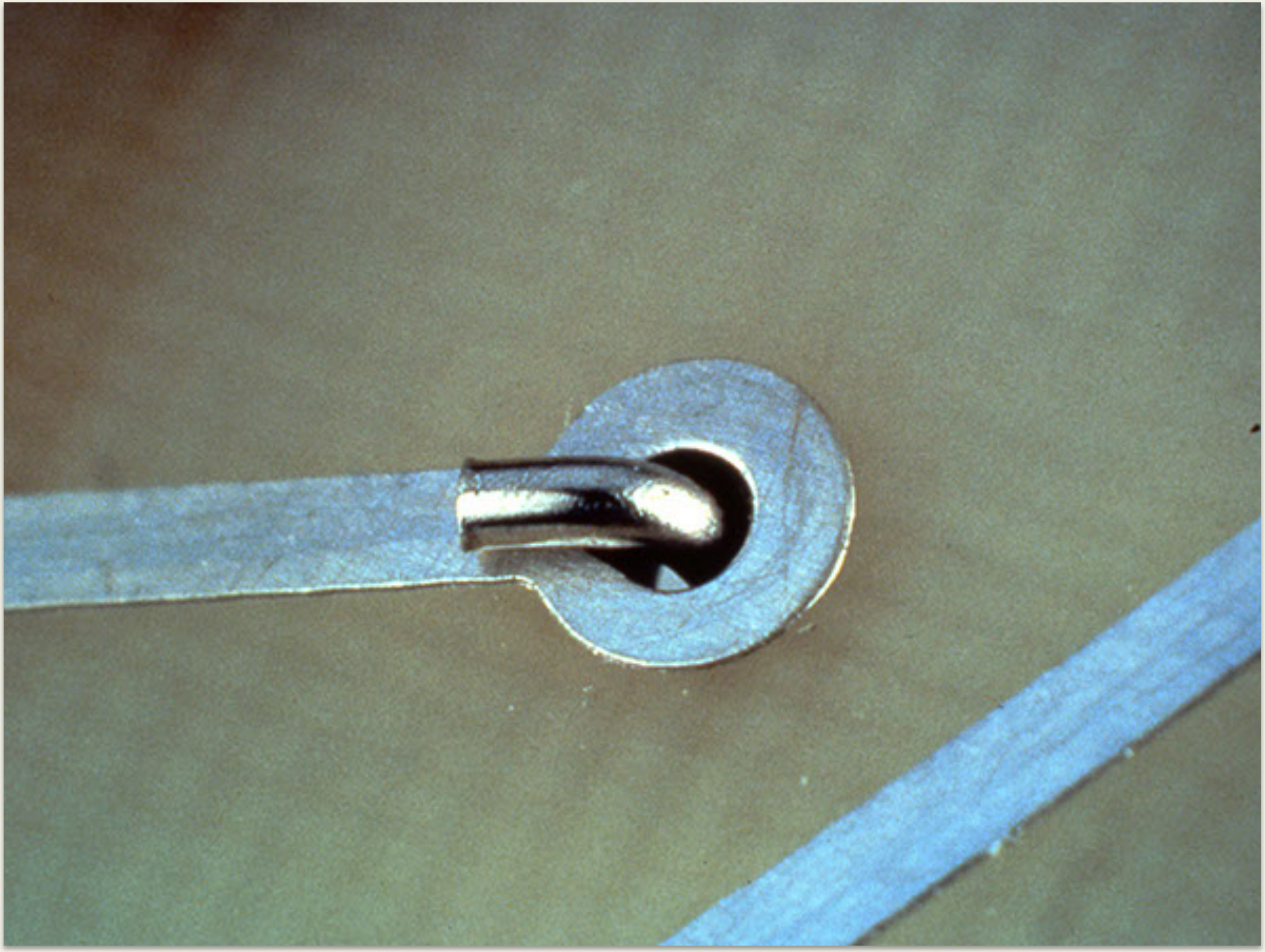

Preferred clinched lead - Accept

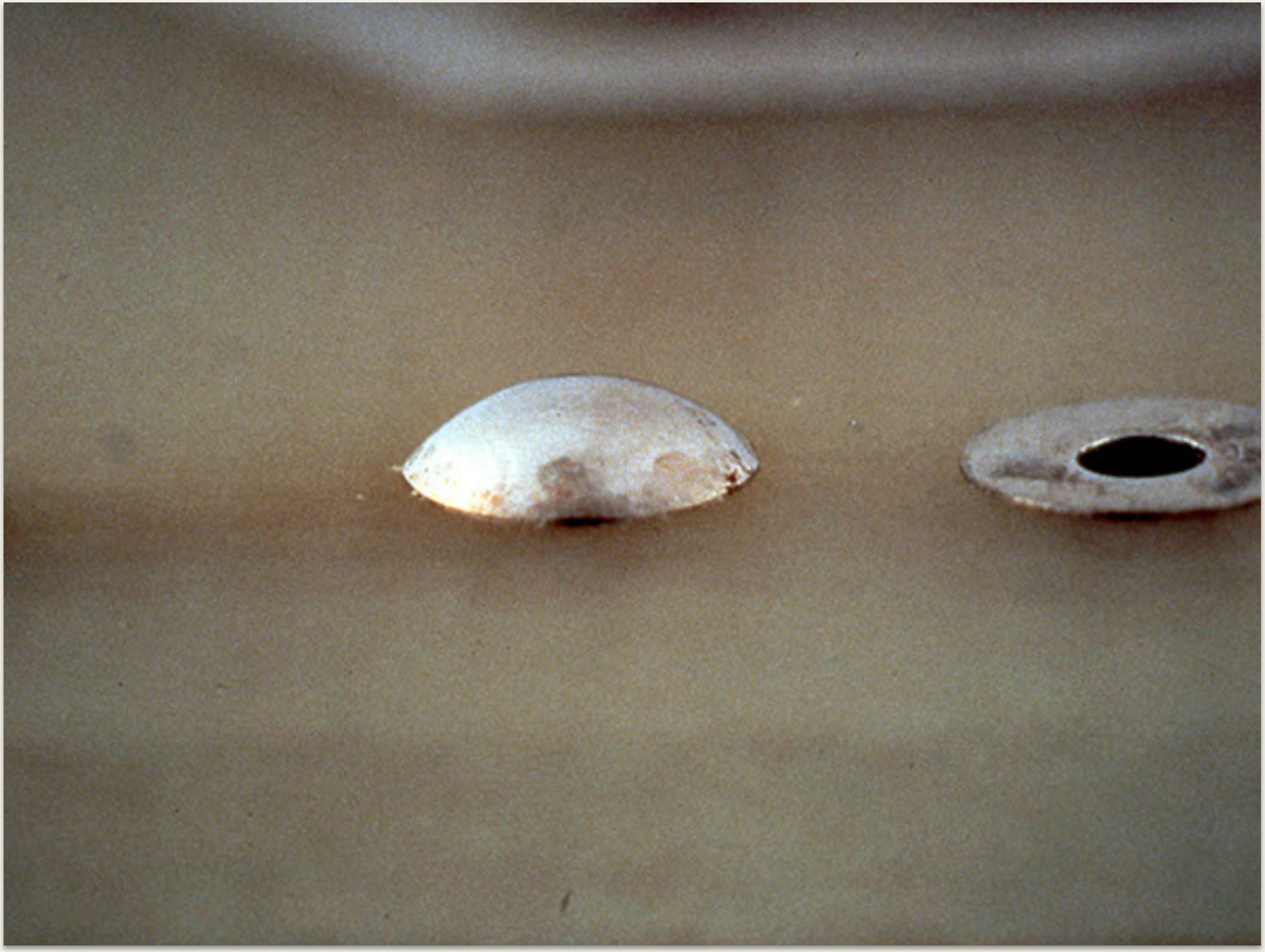

Excessive solder, lead not discernible - Reject

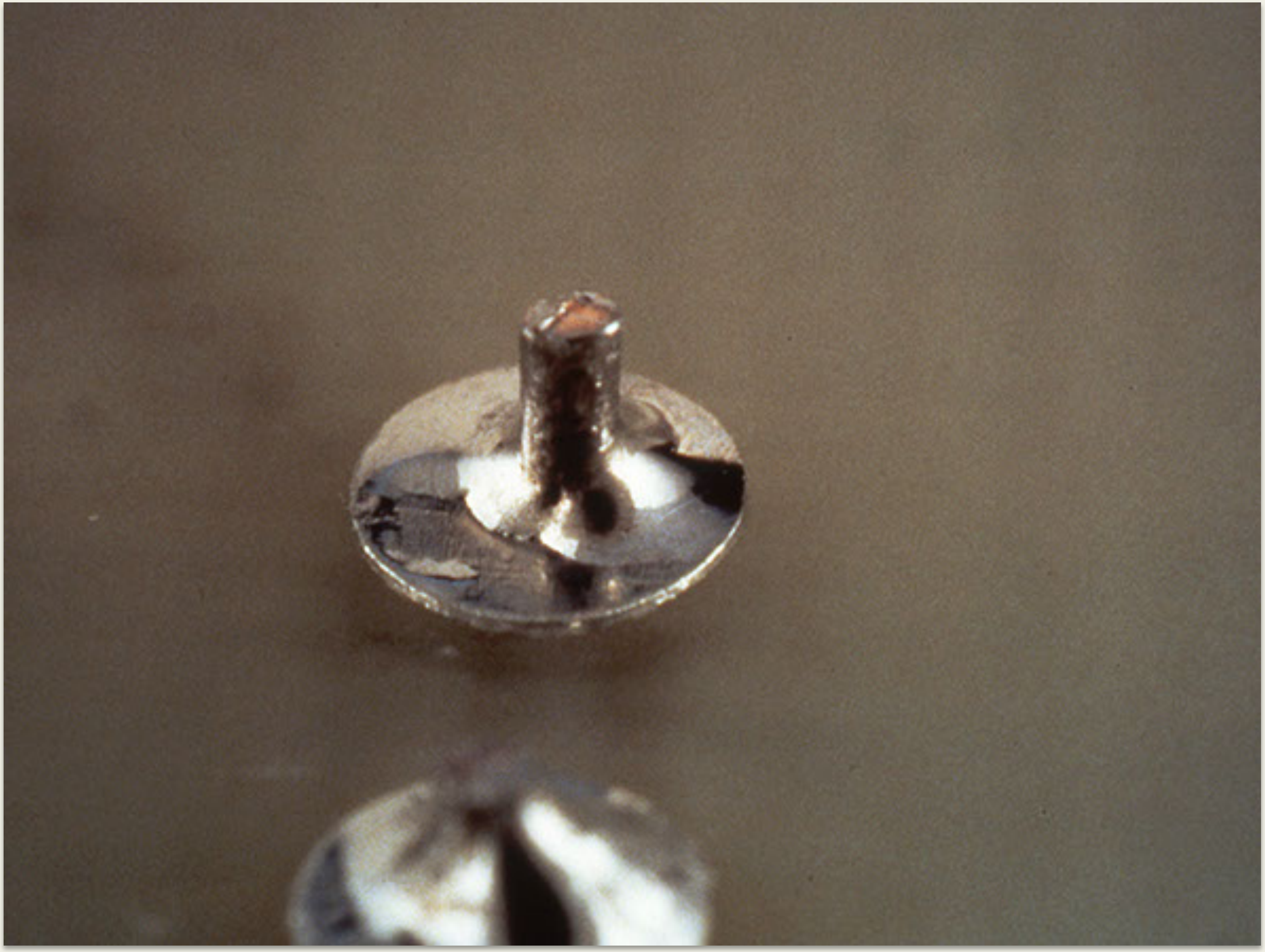

Poor wetting and exposed copper - Reject

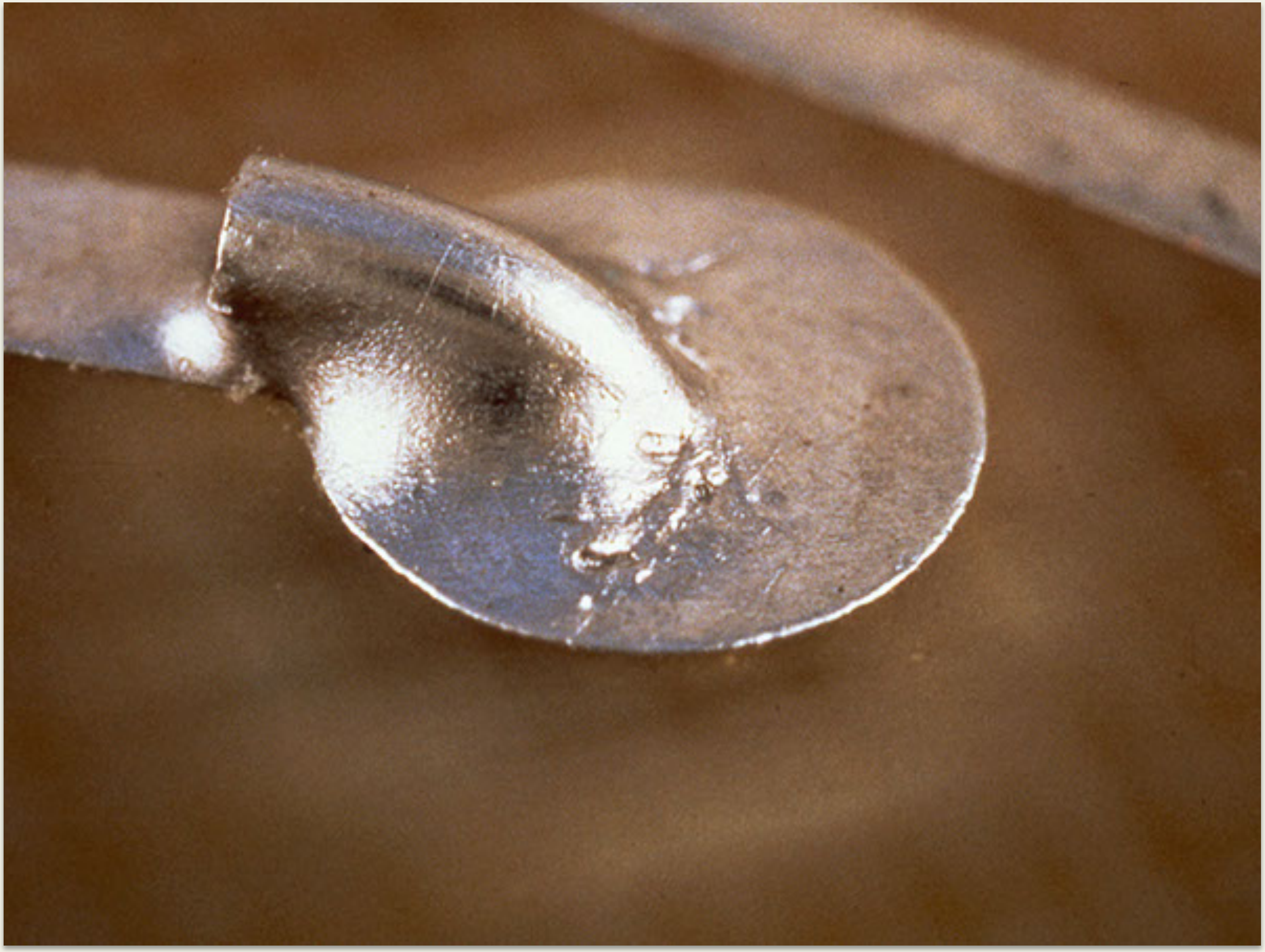

Pits, grainy - Reject

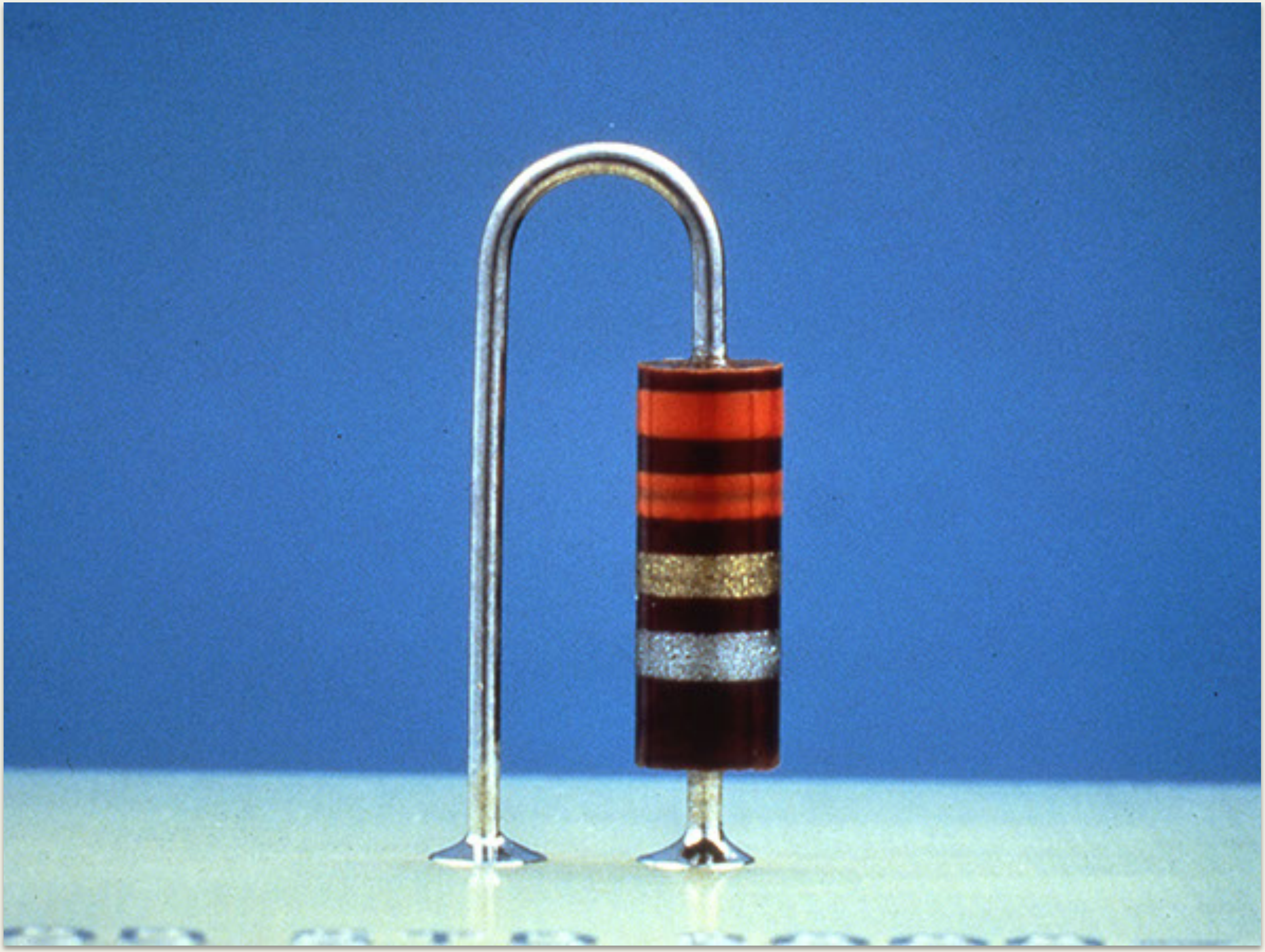

Vertical component mounting - Accept

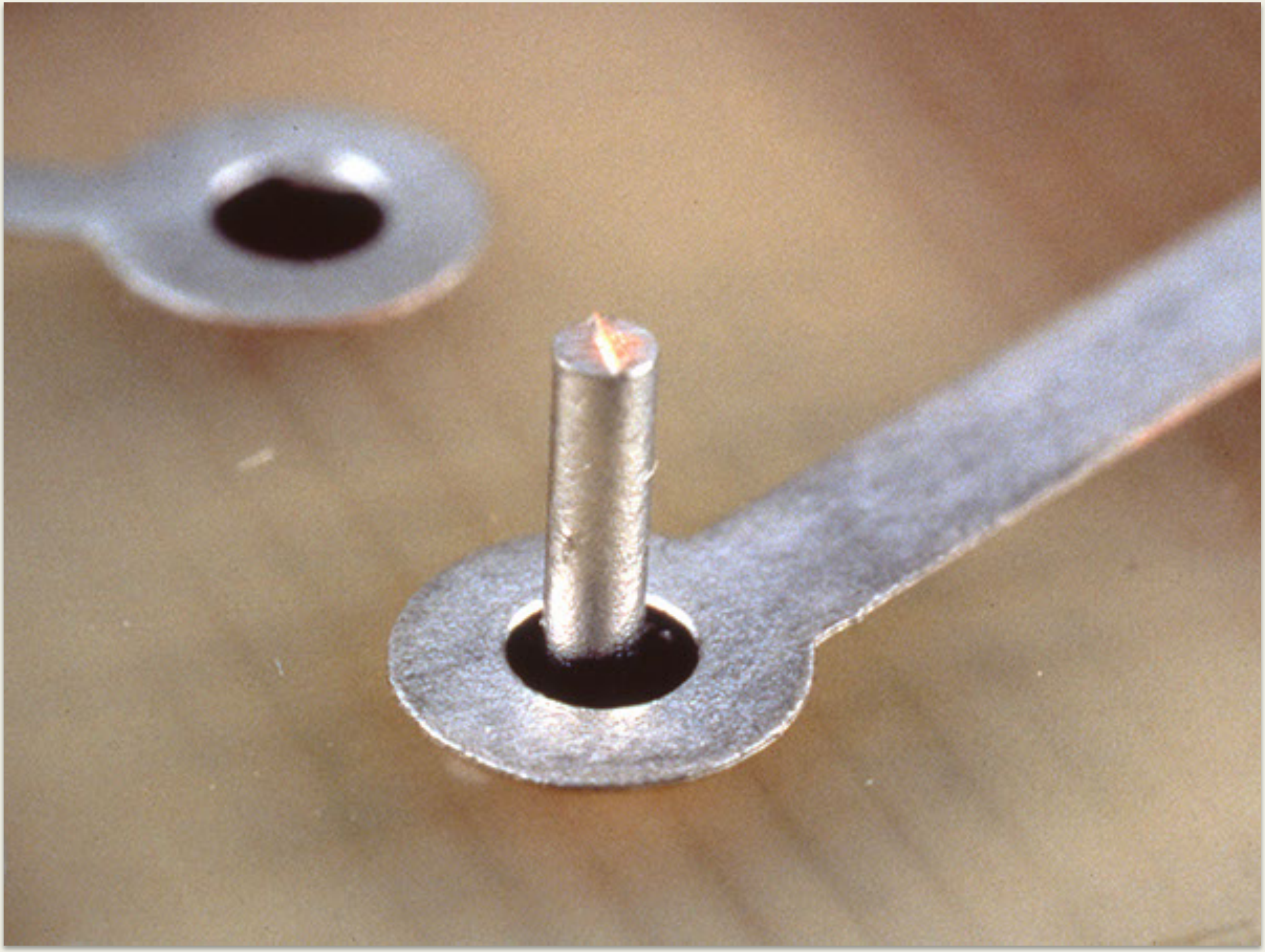

Optimum lead protrusion, stud mount - Accept

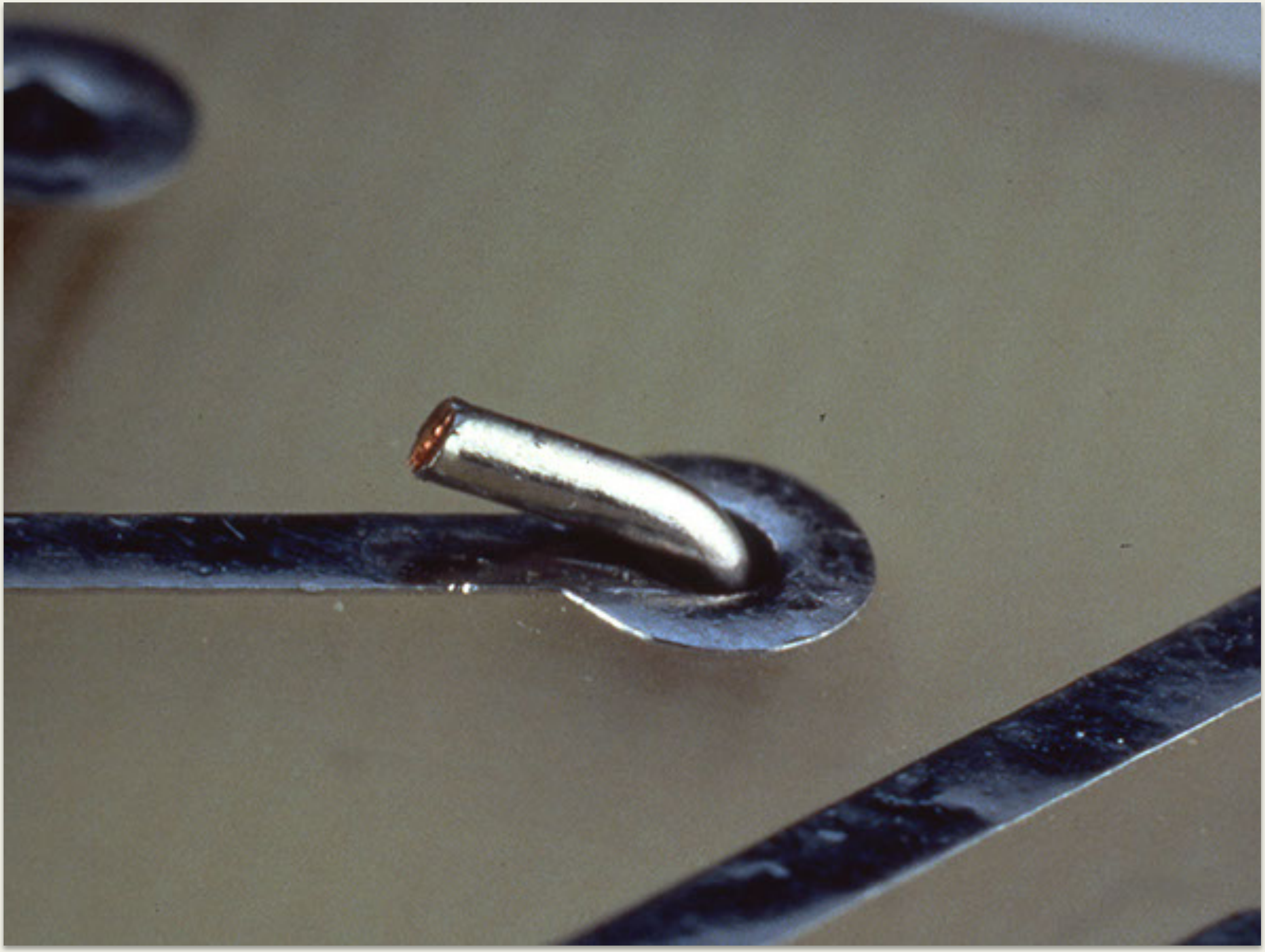

Excessive lead spring back - Reject

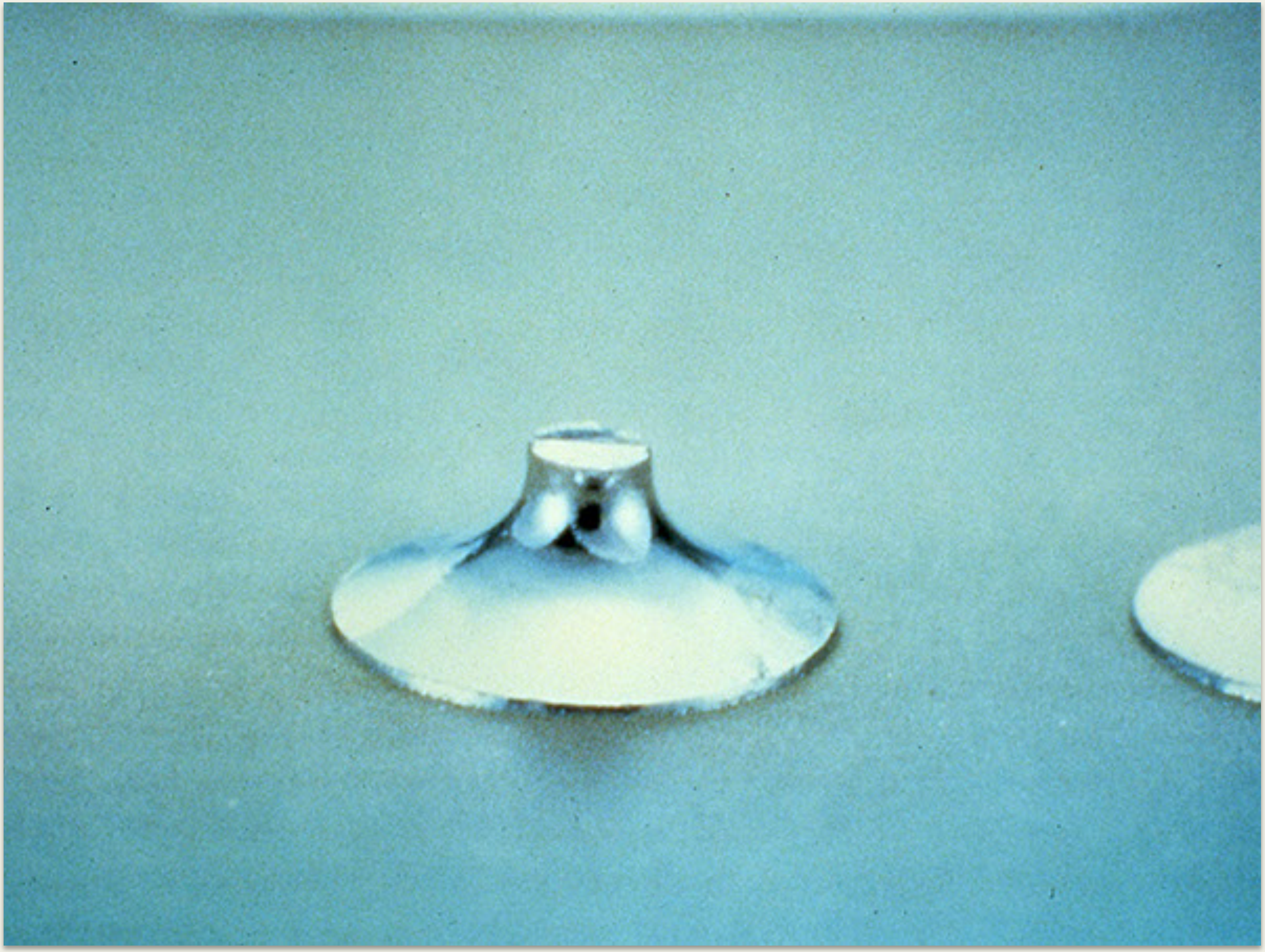

Solder fillet - Accept

Semi-Automated Diagnosis, Repair, and Rework of Spacecraft Electronics

Peter M. Struk[1]
Richard C. Oeftering[2]
NASA Glenn Research Center, Cleveland, OH 44135

John W. Easton[3]
National Center for Space Exploration Research, Cleveland, OH 44135

Eric E. Anderson[4]
Arctic Slopes Research Corporation, Cleveland, OH 44135

NASA's Constellation Program for Exploration of the Moon and Mars places human crews in extreme isolation in resource scarce environments. Near Earth, the discontinuation of Space Shuttle flights after 2010 will alter the up- and down-mass capacity for the International Space Station (ISS). NASA is considering new options for logistics support strategies for future missions. Aerospace systems are often composed of replaceable modular blocks that minimize the need for complex service operations in the field. Such a strategy however, implies a robust and responsive logistics infrastructure with relatively low transportation costs. The modular Orbital Replacement Units (ORU) used for ISS requires relatively large blocks of replacement hardware even though the actual failed component may really be three orders of magnitude smaller.

The ability to perform in-situ repair of electronics circuits at the component level can dramatically reduce the scale of spares and related logistics cost. This ability also reduces mission risk, increases crew independence and improves the overall "supportability" of the program. The Component-Level Electronics Assembly Repair (CLEAR) task under the NASA Supportability program was established to demonstrate the practicality of repair by first investigating widely used soldering "materials and processes" (M&P) performed by modest manual means. The work will result in program guidelines for performing manual repairs along with design guidance for circuit reparability.

The next phase of CLEAR recognizes that manual repair has its limitations and some highly integrated devices are extremely difficult to handle and demand semi-automated equipment. Further, electronics repairs require a broad range of diagnostic capability to isolate the faulty components. Finally repairs must pass functional tests to determine that the repairs are successful and the circuit can be returned to service. To prevent equipment demands from exceeding spacecraft volume capacity and skill demands from exceeding crew time and training limits, the CLEAR project is examining options provided by non-real time tele-operations, robotics, and a new generation of diagnostic equipment. This paper outlines a strategy to create an effective repair environment where, with the support of ground based engineers, crewmembers can diagnose, repair and test flight electronics in-situ. This paper also discusses the implications of successful tele-robotic repairs when expanded to rework and reconfiguration of used flight assets for building Constellation infrastructure elements.

I. Introduction

NASA's plans for long duration missions to the Moon and, especially, to Mars require a much greater degree of self sufficiency on the part of the crew than ever before. Such missions will have greatly reduced logistic support from Earth, compared to International Space Station (ISS) operations. Returning to Earth in the event of an emergency may not be an option either, as a lunar return flight could require two to three days, and a Martian return flight will require much longer.

One area of mission support that NASA must plan for is electronic repairs. Despite the rigorous testing required by NASA, electronics faults have already occurred in both Space Shuttle and ISS operations, leading to the use of backup systems or loss of capability to some degree[1,2]. While the electronics and other systems used in a long duration space mission will undergo rigorous testing, the crew of such missions will most likely encounter an electronics failure at some point in the mission. With the design of the Crew Exploration Vehicle (CEV) already beginning and likely influencing the design of future spacecraft and hardware, it is not too early for NASA to begin exploring and designing techniques and tools for crew members conducting electronics repair during long duration space mission. These considerations include system design (for accessibility, parts type and sizes, and board complexity), repair infrastructure (including diagnostic capabilities, tools, and other needed equipment), and logistics constraints. The decisions on how to approach each of these considerations depends on the overall repair strategy chosen.

---

[1] Aerospace Engineer, 21000 Brookpark Rd., MS 110-3, AIAA Member
[2] Electrical Engineer, 21000 Brookpark Rd, MS 86-5
[3] Associate Staff Scientist, 21000 Brookpark Rd., MS 110-3, AIAA Member
[4] Electrical Engineer, 21000 Brookpark Rd., MS ASRC

The historical approach to electronics repair on manned spaceflight missions (particularly those aboard the International Space Station, or ISS) has been to replace modular subassemblies called Orbital Replacement Units (ORU's). The astronaut simply replaces the malfunctioning ORU with a spare, and then returns the faulty ORU to Earth, where it is diagnosed and repaired, then returned to a pool of spares awaiting re-use in space. This approach allows for simplicity in the diagnosis of the problem, as the fault only has to be isolated to a given ORU, rather than determining the fault down to the component level. Crew training and experience in performing repairs are minimized. Also there is a high degree of confidence in the repair, as the ORU can be tested extensively on the ground prior to being re-launched to the ISS.

Conversely, this approach carries with it a severe penalty in terms of the logistics support required (and overall cost to the program). Conceptually, it is easy to imagine the penalty in mass and volume that is levied when an entire ORU (which may weigh ~25 lbs or more on Earth) is launched to remedy a failure of a single, small component such as a resistor, transistor, or other electrical component. While this approach is expensive, it may be acceptable for low Earth orbit (LEO) missions if they are re-supplied regularly since large quantities of spares would not need to be stored on orbit.

Alternatives to the ORU are lower-level repairs which can include shop-replaceable units (SRU) such as circuit board-level swap-outs and / or a component-level repair strategy. This could serve to greatly reduce the planned-spare ORU requirements, while providing a capability to handle unforeseen repair contingencies. The need to handle such unexpected events has been illustrated by several events in NASA flight history. The successful conclusion of the Apollo 13 mission, after an explosion severely damaged the service module, highlights the benefits of having the capability to implement on-the-spot repairs. The consequence of failing to allow for off-nominal repairs was demonstrated in the tragic loss of the shuttle Columbia; even if the crew had been aware of the problems with the leading edge tiles, they were not equipped to make repairs in-flight. Since that incident, NASA has made considerable efforts to allow for contingency repairs of problems of that nature[3].

The Component-Level Electronics-Assembly Repair (CLEAR) project under the NASA Supportability program was established to develop and demonstrate the technology necessary to allow crew-member to effectively perform electronic repair down to the component level. CLEAR involves collaborative efforts between NASA's Glenn Research Center, Langley Research Center, Johnson Space Center, the National Center for Space Exploration Research, and the U.S. Navy. The overall processes involved with repair which are the objectives of the CLEAR task are composed of four primary elements:

1. Capability to diagnose an electronics assembly and identify the faulty component(s) with equipment that fits within the mass-volume and power constraints of spacecraft.
2. Capability to repair electronics down to the component-level on-orbit with processes and materials that are safe and compatible with the space environment which allows the crew to make the necessary repair.
3. Capability to evaluate and determine that the repaired circuit is safe to return to service.
4. Capability to augment the flight crew with knowledge and skills to diagnose faults and perform repairs without expanding crew size.

This multi-faceted program utilizes a cross-disciplinary approach to examine pre- and post-repair diagnostics and functional test; material and process for repair such as component soldering, conformal coatings removal and replacement; and electronics design for supportability. These areas are investigated by a combination of trade studies, analysis, ground based testing, reduced gravity aircraft testing, and actual spaceflight testing aboard the ISS in multiple experiments. This paper provides an update to this task since previously reported at this meeting in the previous year.[4-5]

II. CLEAR Task Overview

A. Manual Electronics Repair

All repair scenarios for the Constellation Program (CxP) will include some manual repair capability in which the crew can make electronic repairs. Currently, the ISS has a soldering kit aboard which offers a very basic capability and is a logical baseline for future capabilities. The soldering kit and tools available on-orbit are capable of basic repairs (e.g. wire splicing and some limited component-level repair). However, component-level repair has not been actually demonstrated using this equipment aboard ISS. The experience of the US Space Program with repairs is limited to a very few cases (e.g. ARCTIC-1 repair discussed below) and some basic experiments[6]. Before mission planners can expect to have a repair capability, it must be demonstrated in the field (i.e. to raise the technology readiness level, or TRL, to a level six[7]).

In a series of 2 flight experiments, the CLEAR project will demonstrate manual component-level repair using an augmented version of the soldering capability currently on ISS. The first experiment, called SoRGE, uses the soldering kit to investigate the basic materials and processes (M&P) involved with soldering in a microgravity environment using a through-hole configuration. The second experiment, called Component Repair Exp-1, intends to demonstrate all of the physical processes of end-to-end component-level repair of a circuit board. These experiments, which are discussed further below, utilize the Space-station Development Test Objective (SDTO) process to get aboard ISS (ref).

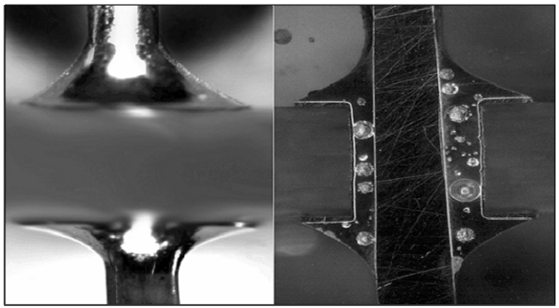

Figure 1: External profile of solder joint (left), and cross section of joint (right) showing internal voids.

1. *Soldering in Reduced Gravity Experiment (SoRGE)*

SoRGE is an experiment which is investigating the formation of solder joints in a microgravity environment (Figure 1). Results from earlier aircraft testing showed that void defects (bubbles /voids trapped within the solder) increased in joints formed in reduced gravity as those formed in normal gravity[8-12]. Joints were formed in various reduced-gravity environments including (nominally) Martian, Lunar, and zero gravity levels[*]. In these tests, the void fraction increased as gravity level was decreased. For joints formed in a nominally zero-gravity environment, the void fraction was 3-times higher than those joints formed in normal gravity. Changes in joint geometry were also observed, with the reduced-gravity joints being more symmetric.

The ISS experiment SoRGE is examining several techniques which may serve to mitigate the formation of voids in the solder joints. This experiment is comprised of a series of small circuit cards, with small electrical components (resistors) attached to the surface (Figure 2). The experiment has several kits, each of which is looking at various solder and flux combinations, to quantify and minimize the amount of void formation in solder joints. Also, this experiment will yield valuable information regarding the physical process of soldering in low-g with a crew-member who received only limited training.

This experiment was brought to the ISS by the Shuttle Atlantis during STS-115 and soldering was accomplished by Astronaut Sunita Williams (Figure 3) during February and May of 2007. Half of the available 12 test kits were completed and returned to Earth aboard STS-118 in August 2007. The samples were returned to GRC in October 2007 and are currently undergoing analysis. The planned analysis consists of a visual examination for solder process quality as well as non-destructive analysis to look at solder joint voiding. The results of this research will help specify the soldering process used in future in-space repairs.

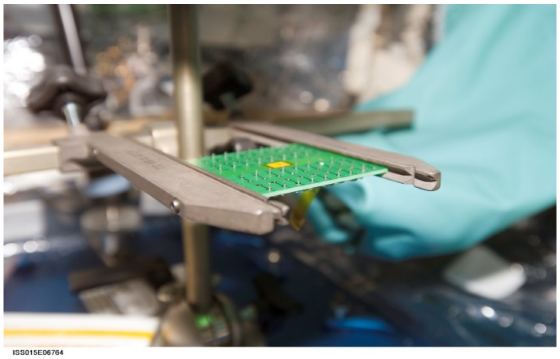

Figure 2: An image of the test card used during the SoRGE experiment.

ISS015E06767

Figure 3: Astronaut Sunita Williams conducting the SoRGE experiment during Expedition 14 aboard the ISS.

---

[*] Brief periods (~25 seconds) of reduced-gravity are experienced aboard these aircraft as they fly parabolic trajectories which are calculated to provide the desired relative acceleration levels.

2. SDTO2: Component Repair Exp-1 (CRE-1)

CRE-1 will demonstrate the physical processes of an end-to-end manual electronics repair in a microgravity environment. The physical processes include:

- Conformal coating removal
- Component removal
- Board cleaning
- Soldering a new component in place
- Reapplying conformal coating

Crew members will work on functional circuit boards designed to mimic those already in use. These board designs use standard circuit components and component spacing. These standard circuit components include through-hole parts, standard-pitch surface mount devices (SMD), and fine pitch SMDs with typical component spacing (Figure 4). The circuit boards also have three thicknesses of silicone room-temperature vulcanizing (RTV) conformal coating: a primed, 4-mil coating; an unprimed 15-mil coating; and no coating, allowing for continued operation should the crew encounter difficulties removing conformal coating. The repairs will be conducted within the Maintenance Work Area (MWA) aboard the ISS (Figure 5). The experiment will utilize the soldering kit available aboard the ISS as well as an augment tool set provided by the CRE-1 experiment. The goal of the tests is to evaluate the feasibility of performing component level repair and evaluate processes so recommendations to procedures, tool selection, crew training, as well as board design may be made. CRE-1 is currently at Kennedy Space Center awaiting a manifest opportunity to fly to the ISS.

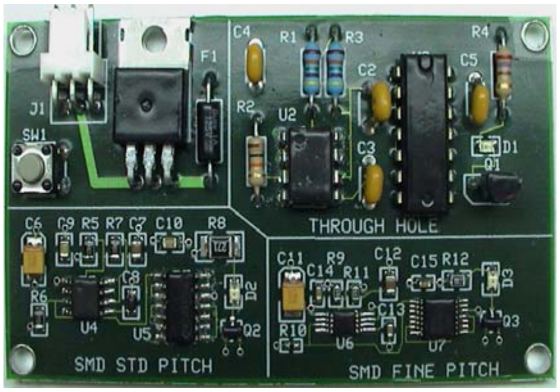

Figure 4: Image of CRE-1 circuit board

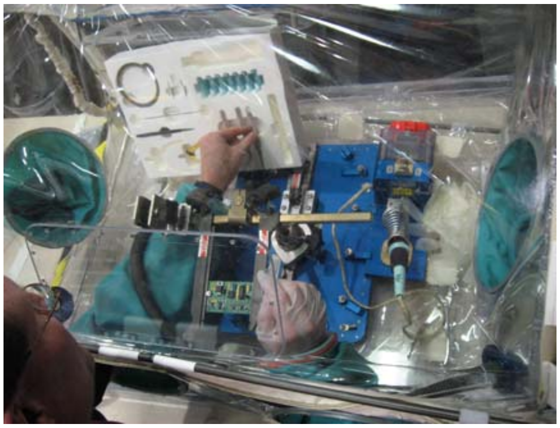

Figure 5: Testing of the CRE-1 experiment in the MWA mock-up at Johnson Space Center. Shown in the image are the auxiliary tools used to conduct manual repair of electronics.

3. Manual Repair Process Recommendation

The primary goal of the manual repair efforts is to develop a series of recommendations to the NASA CxP for processes, materials, tools, and training for crew members to perform manual electronics repairs during space missions. These recommendations will be built from the results of the on-orbit testing, ground based testing on the C-9 reduced gravity aircraft, and work in consultation with the U.S. Navy at the Naval Surface Warfare Center in Crane, IN. Additional work expected to begin in 2008 includes examination of possible conformal coating selection, both in manufacture and after a repair, that would help enable manual repair. It is anticipated that this recommendation will include future development areas needed to refine manual repair.

Analysis of Arctic Freezer Samples

Studying the results of electronics repairs actually performed on orbit will provide an insight into the actual, current capabilities and constraints of crew members performing these repairs. During Expedition-6, Astronaut Don Pettit attempted the repair of ARCTIC-1 (or the Arctic Freezer), a module used for the cold storage of experiment samples. This work included some aspects of electronics repair, including joining wires and connecting wires to component tabs (Figure 6), with the use of the soldering kit on the ISS. Nondestructive analysis and visual inspection of these joints, as well as, similar joints formed in normal gravity for comparison, is on-going and will aid in

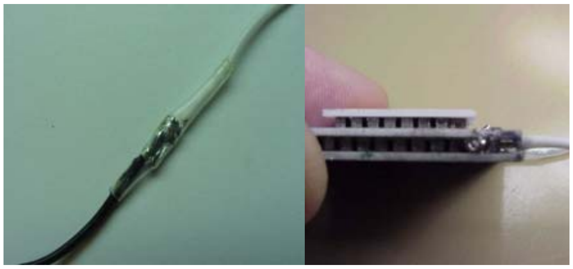

Figure 6: Typical solder joints examined from the repair of ARCTIC-1 freezer aboard the ISS.

evaluating the effectiveness of soldering in reduced gravity, especially in the mitigation of voids within solder joints.

C-9 Test Analysis

In February 2007, a series of solder tests were conducted on the C-9 reduced gravity aircraft which included the examination of various solder and flux combinations, most focusing on the use of no-clean or water-clean fluxes with eutectic solders. This work builds on previous aircraft tests[8-12], and will provide insight into general soldering and void mitigation techniques for reduced gravity operations. The analysis includes discriminating between samples based on the quality of the given in-flight performance (the aircraft parabola) as well as acceptance by a NASA flight qualified technician, based on NASA standards. Acceptable samples are nondestructively analyzed for void formation. Results from this analysis of a subset of these tests are presented in a complimentary paper at this conference[13]. These results will help develop criteria for selecting solder, flux, and techniques to be used for in-flight repair of electronics.

Crew Training Requirements (with U.S. Navy)

Crew training prior to a mission is vital for electronics repair. Crew members are presumed to have little or no prior experience in performing repairs, so training materials must prepare and reinforce the repair process. The team, working with a group at the Naval Surface Warfare Center in Crane, IN, responsible for preparing training materials for the U.S. Navy, will prepare and test crew training courses of 4, 8, and 16 hours, allowing NASA to tailor and choose a course suitable for electronics repairs. These materials are also planned to be available during flight as a refresher course and to prepare a crew member immediately prior to performing a repair. The team, together the U.S. Navy, will propose a tool set tailored to the level of crew training to allow crew members to effectively conduct the repairs they have been trained to perform.

B. Semi-Automated Repair

The previous section discussed hand or manually operated repairs used in the soldering and desoldering process. Manual repair will always provide an important repair capability, but it requires training and periodic practice to maintain proficiency. Manual repair also requires direct crew involvement throughout the repair process. A fully automated process such as in manufacturing requires special fixtures and equipment, along with a great deal of experimenting or system "tuning." For single unit repairs a mix of manual and automated operations is more practical and is most likely to achieve successful repairs within the spacecraft weight and volume constraints. Certain operations, such as wiping away process residues, are simple if performed manually but can be very difficult to accomplish by robotic means. Heating devices to narrow temperature bands and then quickly and precisely placing them demand automated operations. Semi-Automated repairs reduce the demand for high skill while increasing productivity.

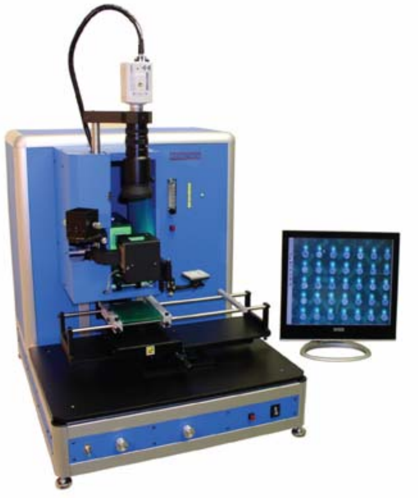

Figure 7: An ATCO AT-707 is typical of a commercial rework station capable of replacing high density integrated circuits. Photo from http://atco-us.com/.

Rework Stations (Figure 7) that melt or "reflow" solder with Hot Gas or Infrared heat evolved to handle complex devices. Such systems are semi-automated where manual operations involve alignment with optical aids but employ automated heating systems and mechanically guided placement. In this way, Semi-Automated rework stations extend repair capabilities to circuits not amenable to manual repair.

Figure 8 is a qualitative representation of how semi-automated technology can extend a repair capability. The vertical axis of Figure 8 qualitative represents how increased PCB layering increases board thermal conductivity. The increased conductivity increases heat transfer away from the solder joint(s) and may exceed the manual soldering iron's capacity. More advanced manual techniques (more advanced soldering tools than available on ISS) may include custom tips and board heaters to preheat the entire board to temperatures just below the reflow solder temperature. However, the high heat loads and long heating cycles require to solder the conductively cooled circuit boards, common in spacecraft, are more likely to be successfully repaired by semi-automatic systems.

The horizontal axis of Figure 8 shows increasing complexity in device packaging and lead count. At the extreme right is the Ball Grid Array (BGA) where lead counts may exceed 1000 and are impossible to solder by hand. Large integrated circuits like BGAs require simultaneous solder or desolder of many pins or solder balls and are best handled by semi-

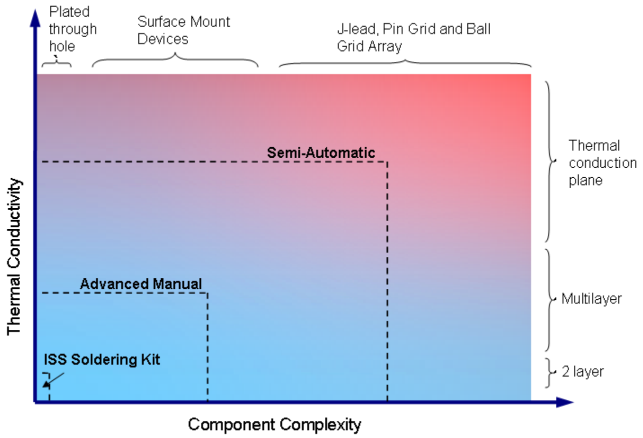

Figure 8: Graphic showing the increased repair capability offered by semi-automated repair technology.

automatic equipment to provide precise parameter control, such as, placement, coordinated motion, temperature ramp and soak profiles, and critical timing.

It is likely that the first operational need for repair capabilities is during long-duration outpost stays on the Moon and especially Mars. However, the ISS is a prime location to evaluate such technology and, because of the retirement of the space-shuttle in 2010, may benefit from a repair capability to help sustain systems with reduced logistical support. While the ISS and outposts share many similarities in terms of the environment, the prime difference is that there is significant gravitational force for the outpost whereas for ISS there is not. The semi-automated system is expected to employ or accommodate a variety of functions that include machine set-up, optical examination, process monitoring, a variety of mechanical repair operations (including soldering and de-soldering), handling and application of a variety of process fluids, and containment of solid debris and volatile products within the tight confines of a spacecraft or habitat. Before investing in specialized equipment, studies need to determine if the processes have any significant reduced gravity dependence, outgas, or other space environment sensitivity. The CLEAR task is in an early evaluation phase examining the general operation and capability of these stations, as well as the benefits and challenges for adapting them for use in space flight.

C. Diagnostics and Functional Test

While the physical repair of electronics in spacecraft and habitats is challenging, diagnosing faults down to the component level represent a bigger challenge. Furthermore, confidence of a successful repair must be accomplished via a functional test of the repaired board and/or subsystem. On the ground, NASA, the military, and industry all have practices, technicians, and equipment in place to perform diagnostics and repairs of electronics. Each group has adapted standard test equipment as well as customized equipment to perform the diagnostics particular to their requirements. None of these cases, though, have the volume, mass, power, and user experience level limitations imposed on them as for a long duration space mission.

Currently, ISS diagnostic and test equipment is limited to a portable Fluke Scope Meter®, a laptop logic analyzer and a limited output power supply and a special test set up to specifically test on-orbit multiplexer/de-multiplexer units. Such hardware is far from a complete diagnostic instrument set suited to test the array of systems including: electrical power, command & data handling, communications & tracking, guidance & navigation, and controls & display systems throughout spacecraft. Even if provided detailed information, it is beyond the most capable crews to fully understand diagnostic measurements of every spacecraft component. The next sections describe the strategy and technology that CLEAR is evaluating for diagnostics and test capability in future manned space missions.

NATIONAL INSTRUMENTS
NI PXI-1045
PXI

Figure 9: A representative National Instruments PXI chassis populated with a variety of instruments from different manufacturers. Photo from http://www.pxisa.org/.

Diagnostics

Repairs of circuit board assembles requires isolating faults to the component level. There is no single technology solution but success lies in focusing on the fundamentals of measurements rather than stacks of instruments. The challenge is selecting equipment and/or developing hardware that can cover the majority of diagnostic needs (as well as functional test discussed below) within the allowable mass and volume constraints of a spacecraft or habitat. The PXI (PCI eXtensions for Instrumentation) based suite of instruments (Figure 9) may be one potential solution. The advantage of such devices is the space saved by combining many PXI-based instruments into one bus. The drawbacks include requirements for instruments to be redesign into card format which is compatible with the bus architecture. Further, bus architecture often becomes obsolete. For example, the PXI architecture is currently being displaced by the new PXI Express.

Another approach would be to use Local Area Network (LAN) technology such as LXI (LAN Extension for Instrumentation). LXI instruments can be used in a coordinated suite of instruments or broken out as independent and portable instruments. High speed LAN based instruments are expected to resist obsolescence much longer. LXI instruments however, are not as compact as PXI. Recent industry trends indicate that a mixed system of LAN and Bus based instruments provide the best solution.

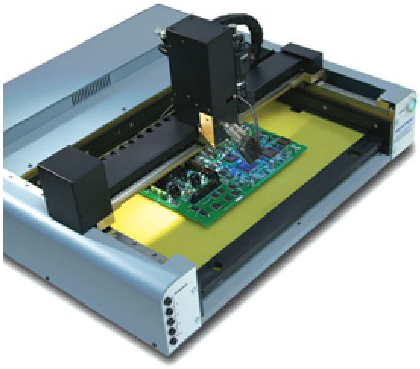

Figure 10: Huntron® ACCESS unit is an automated diagnostic system that employs circuit computer aided design (CAD) data and imaging to help guide the probing process. Photo courtesy of Huntron. Inc.

Diagnostics are often comprised of instruments that are used manually or programmed to automatically make measurements in a prescribed sequence. The instruments usually send stimulus signals and capture responses from the target circuit. For automated diagnosis, the same motion control capability used in repair can provide probing for diagnostics. One promising technology is the Analog Signal Analysis (ASA) technique used by the Huntron® ACCESS instrument to probe circuit nodes with a very low power signal that will not harm the circuit (Figure 10). This signal energizes single nets on the card to determine a characteristic waveform. A net is a common connection where multiple components are connected. This waveform is then compared with a known good board signature. If there is significant difference between the measured waveform and the known good waveform, one or more of the components attached to the net are defective. This approach reduces the number of components required to be replaced and minimizes the operator's skill requirements. However, ASA may not always isolate a fault to individual components, particularly, if they are not directly accessible. Complex Signature Analysis (CSA), being developed by CLEAR, is an alternative approach that relies on characterizing a circuit's self resonant and network resonant behaviors. CSA is expected to detect faulty devices that are inaccessible to ASA techniques. ASA and CSA both provide the ability to diagnose a circuit in a "power OFF" condition. Should ASA/CSA fail to isolate faults using a "known good board" technique, ASA/CSA can still capture measurements that could be further examined by knowledgeable ground based engineering. ASA/CSA operates on circuits in a "Power Off" condition which affords extra safety. However, it is often necessary to run tests on a circuit in a "Power On" condition to isolate functional problems such as software faults.

Functional Test Equipment

Functional testing is used to determine if a circuit is capable of performing its intended function before it is integrated into a system. It is often used to verify software functions that cannot be measured by fundamental diagnostic measurements. The actual functions of an ORU electronic circuit may be simple and repetitive. Most ORUs are part of a larger system and thus functionality may require interaction through multiple external interfaces. Functional test equipment often emulates the external system and operational conditions and includes additional "breakout" instrumentation to monitor interactions. Automatic Test Equipment (ATE) is employed when the number of channels and execution speed exceeds the capability of human operators. Typically, ATE systems are a combination of custom and off-the-shelf test equipment and software, with a software executive that governs the test process.

It is not uncommon for a single avionics box to require multiple racks of test equipment. In their terrestrial form, the weight and volume required for functional testers prohibits their use aboard spacecraft. However, functional test is an important part of any repair process and must be available. The ability to perform functional testing at the board level enables the generation of engineering data during the debug process. It also allows for the functional testing at the board level to minimize the risk of a circuit card damaging the system. Therefore, strategies for functional testing in spacecraft or habitat volumes require significant reduction of both mass and volume of terrestrial systems or newer technology. One such technology is offered by synthetic instruments.

Synthetic Instruments

Synthetic Instruments (SI) refers to an approach that exploits software and reconfigurable hardware to maximize the flexibility of test systems. Synthetic Instruments are implemented on generic hardware, typically, Field

Programmable Gate Arrays, (FPGAs). FPGAs are designed to provide a vast array of logic gates that can be configured to provide distinct data and control functions. Unlike software executed by microprocessors, the "Synthetic Instrument" programmed into the FPGA executes at extremely high speeds and with high levels of parallel processing. The FPGA can be reconfigured, on demand, as a completely new Synthetic Instrument. The latest generation of FPGAs can provide real-time signal processing which was once the domain of Digital Signal Processors. The main advantage of SI is that it allows us synthesize high performance instruments on an "as needed" basis.

The Department of Defense regards SI as a major advance in reducing (a) life cycle costs (b) time to develop and field new or upgraded test equipment, (c) test system logistics footprint, and (d) test systems physical footprint[14]. The strategy behind SI is to allocate any digitized functions to software. Innately analog functions, such as sensors, analog signal devices, and power sources are handled by modules that can be reconfigured electronically to provide wide operating ranges. In addition, to handle the vast variety of connector configurations, SI provides an upfront analog switching matrix that routes signals from a target test circuit to a selected analog instrument module. The CLEAR project is evaluating SI as part of a strategy to provide both diagnostic and functional test capability for a wide range of spacecraft electronics using only a small amount of physical hardware.

Both repair and diagnostics require significant skill and knowledge related to the circuit under test. Therefore, it is anticipated that the crewmember will require significant ground based support to debug issues. Furthermore, many of the process described above can be automatically controlled from the ground in a non-real time fashion. To keep the crew workload to a minimum, the semi-automated diagnostic and repair system should be amenable to tele-operations. Such operations also will allow data from the circuit under test to be available to ground personnel.

D. Tele-Operations: Linking the Crew to Knowledgeable and Skilled Support

In light of limitations on available technician-level skill and designer knowledge of the electronic systems, the key to success is to link the process to knowledgeable and skilled engineering staff though tele-operations. Other tele-robotic experiments focus on real-time control and real-time feed back while other robotic efforts focus on achieving total robot autonomy. Although crew autonomy is the desired goal, a likely strategy for a semi-automated diagnostic and rework station in spacecraft and habitats is to provide timely but not real-time interaction. The preferred mode is to avoid real-time "joystick" interaction with the hardware because the current infrastructure is prone to interruptions and signal delays.

For Lunar operations, the 4 second round trip transmission and system delays make real-time control awkward and even dangerous. The vast range of delay times for Mars makes real-time control impossible. Even the ISS has delays but also frequent "loss of signal" events which makes real-time control unreliable. Avoiding real-time interaction simplifies the communications and interaction and also provides an opportunity to "validate" commands prior to sending them to prevent costly and possible hazardous mistakes. Pre-scripted repair routines for anticipated needs will be developed. A potential solution to avoid the cost of a of support staff on standby, repair depots and manufacturers could support repairs on an "as needed" basis. The use of such a tele-operations capability is best captured by a Semi-Automated Diagnostic and Repair Operations scenario we describe below. This scenario is intended to illustrate how a diagnostic and repair tele-operation might proceed with a repair capability as shown in Figure 11.

Semi-Automated Diagnostic and Repair Operational Scenario

A vehicle health monitoring system reports a fault in a system ORU package. The faulty ORU is removed and replaced with a spare ORU which restores functionality to the system. Subsequently (as crew time permits and criticality requires), the faulty ORU is further diagnosed with a computer connected via an ORU test port. An internal Built-In-Tests (BIT) checks the diagnostic codes on all circuit boards. Ground support observes the diagnostics data via tele-operations. One circuit board fails to respond to BIT initiation. The subject board is removed from the ORU and interconnections were found to be good. Ground support asks the crew to move the board to the Semi-Automated Diagnostic and Repair Station and set up for further diagnostics and possible repair.

The board's failure to execute the BIT implies that the main processor is non-functional due to internal fault or failed supporting component. The ground support engineers examine high-resolution camera images down-linked from workstation but no visible damage is found. The circuit is equipped with a JTAG port and supports "boundary scan techniques[5]". The JTAG Boundary Scan Analyzer finds no problems at the I/O ports but reveals that several

---

[5] Boundary scan, also known as JTAG 1149, was developed by the Joint Test Action Group to allow board makers to check individual devices while isolating them from the external circuit[15].

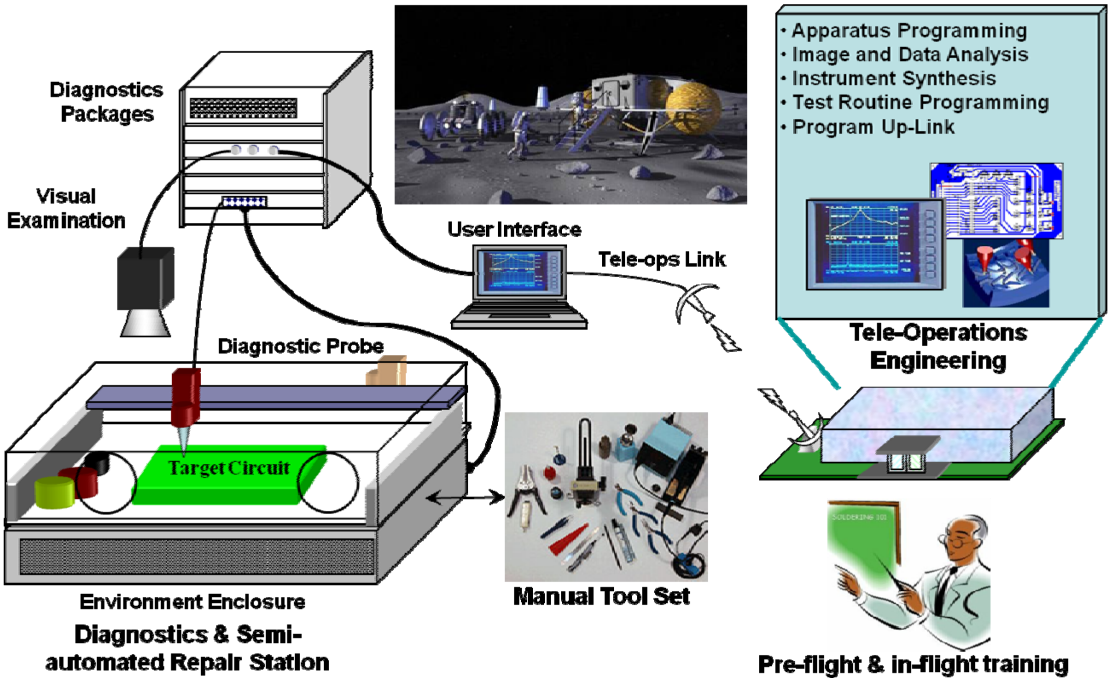
Diagnostics Packages
Visual Examination
User Interface
Tele-ops Link
• Apparatus Programming
• Image and Data Analysis
• Instrument Synthesis
• Test Routine Programming
• Program Up-Link
Tele-Operations Engineering
Diagnostic Probe
Target Circuit
Environment Enclosure
Diagnostics & Semi-automated Repair Station
Manual Tool Set
Pre-flight & in-flight training

Figure 11. Diagnostic and Repair concept (including both manual and semi-automated processes) being evaluated by the CLEAR team.

registers are irreparably damaged, possibly due to radiation. Having isolated the component, the repair workstation is set up to remove and replace the device.

The ground crew determines that no prior repair routine exists so they go “offline” to create a repair program while the crewmember resumes other activity. Assisted with information from the vendor, the ground crew loads the circuit’s Computer Aided Design (CAD) data into a Repair Process Modeling Tool that aids in programming robotic operations of the rework station. The circuit’s registration points and exclusion zones are defined to prevent collisions with clamps and some taller board components. The Repair Process Modeling Tool defines motion paths, tool change points, including the component pick & place operations. Reflow heater position, heating profile are calculated using a reflow thermal simulation tool developed as part of the rework station. When the control code script is complete, it is run through a verification tool that simulates the process and shows a 3D animation of the process. Tool collisions discovered in the simulation are corrected and the code is determined to be error free.

On-orbit a “touch-off” of circuit registration points assures the machine motion is calibrated. The ground crew has determined that the conformal coating around the component should be removed manually because jumper wires interfered with mechanical methods. The workstation enclosure circulation filters capture any stray debris as the crewmember strips away the coating.

The control is then switched to automatic mode and the process is initiated. After the reflow heat cycle, a robotic pick & place tool removes the component. A quick manual wipe of the site removes any residue and a solder paste dispenser applies new solder. Another reflow cycle melts the paste while a component pre-heater prepares the new device. The robotic suction tools picks up and positions the new device and holds it until the last reflow heat cycle is complete.

An imaging camera scans the board for completeness and looks for debris. The tool is changed to an Analog Signature Analysis Probe and the preprogrammed probing sequence reveals that the board’s signatures are normal. Subsequent JTAG boundary scan and register checks are normal. Power up BIT tests also return normal results. Finally, the ground team has come up with code to program the SI system available on orbit to do a function test on the card. The board is connected via a cable assembly and the instruments go through a functional checkout of the circuit board with normal results. The board is then re-integrated into the ORU and BIT tests indicate functionality. The final assembly is complete and any configuration information is noted. Since the original ORU was replaced with a spare, this unit now becomes the next serviceable spare until needed. The ground updates any lessons learned and refines the repair program. After further simulations, the repair script is “canned” as a ready-to-run routine for future use.

Impact on the Future of Constellation

Extrapolating a future capability based on Earth based practices is not enough. On Earth, the economy of mass markets, low cost labor, low cost transportation, and very high production rates results in circuits that have short service life and are un-repairable "throw-away" electronics. Further, the design, materials, and processing choices are based on meeting economical production needs and often involve waste, aggressive chemistry and materials that are unsuited for space.

Understanding the behavior of basic materials in low-g and high vacuum and making the effort to fully exploit processes that are innately suited for space will lead to in-situ repairable electronic circuits. The extremely high cost of launching replacement hardware drives us toward electronic designs that provide long service life, but are also repairable, and can be reworked and reconfigured to serve alternate uses.

The spacecraft constraints imposed on the addition of new equipment are formidable. The development of several key capabilities will enable us to repair and rework spacecraft electronics without relying on Earth launched logistics. The CLEAR task is considering the strategy of compressing large test systems to small scale because of the huge cost of redesigning every known instrument into a small form factor. We are seeking alternate ways of diagnosing circuits by exploiting built in features where possible such as the JTAG Boundary Scan, and by looking at potential technology solutions such as Synthetic Instruments. Such concepts can be the first steps for a fabrication capability to help build infrastructure to support future missions.

The tele-operations developed for robotic repair can evolve into a fabrication capability as well. The experience learned from conducting repairs as well as the underlying materials and processes directly translates to operations for manufacturing with in-situ resources. Another possibility is to build additional infrastructure by reworking spent flight hardware for different functional use. This strategy need not wait for Mars – it can be employed and perfected in the Lunar Outpost phase. The near term need not be overlooked as the impact on logistics is already being felt as Space Shuttle support for ISS begins to draw to a close. By delivering capabilities starting with the ISS and then the Lunar Outpost, we can benefit from cost savings and early return on our development investments as well as gain confidence in their capability by the time we send humans to Mars.

Summary

NASA's historical solution to the problem of in-flight electronics repair has been the replacement of sub-units called Orbital Replacement Units. This approach requires significant logistical support through regular re-supply flights. For long-duration space missions, resupply flights are limited. Therefore, NASA is investigating component-level repair of electronics as a potential way to reduce the logistical footprint required to support future missions.

The Component-Level Electronics-Assembly Repair (CLEAR) project under the NASA Supportability program was established to develop and demonstrate the technology necessary to allow crew-member to effectively perform electronic repair down to the component level. CLEAR is looking at ways future exploration crew can, within the constraints of a spacecraft or habitat, (a) diagnose electronic assemblies and identify the faulty components, (b) repair electronics down to the component-level, (c) evaluate the circuit post repair via a functional test, and (d) augment the flight crew with knowledge and skills to diagnose faults and perform repairs without expanding crew size. These areas are investigated by a combination of trade studies, analysis, ground based testing, reduced gravity aircraft testing, and actual spaceflight testing aboard the ISS in multiple experiments. This paper provides an update to this task since previously reported at this meeting in the previous year.

Disclaimer

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Aeronautics and Space Administration, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

Acknowledgments

The authors wish to acknowledge Dr. Richard Pettegrew for his valuable technical contribution and leadership in the development of the CLEAR task. Also, the authors wish to thank Mr. Duc Truong of the NASA Glenn Research Center for his work and dedication. This work is funded by the Supportability Project in NASA's Exploration Technology Development Program Office (Mr. Barmac Taleghani, Program Manager, NASA Langley Research Center).

References

[1] Beddingfield, K.L., Leach, R.D., and Alexander, M.B. (editor), “Spacecraft System Failures and Anomalies Attributed to the Natural Space Environment”, NASA Reference Publication 1390, August, 1996.

[2] Rutledge, P.J., Mosleh, A., “Dependent-Failures in Spacecraft: Root Causes, Coupling Factors, Defenses, and Design Implications”, *Proceedings of the IEEE Annual Reliability and Maintainability Symposium*, 0149-144X/95, (1995), pp. 337-342.

[3] Columbia Accident Investigation Board, “Columbia Accident Investigation Board Report” [online report], Vol. 1, August, 2003 URL: http://caib.nasa.gov/news/report/volume1/default.html [cited 26 December 2007].

[4] Pettegrew, R.D., Easton, J.W., and Struk, P.M., “Repair of Electronics for Long Duration Spaceflight”, *45th AIAA Aerospace Sciences Meeting and Exhibit*, Reno, NV, January 8-11, 2007, AIAA-2007-1364.

[5] Easton, J.W., Pettegrew, R.D., and Struk, P.M., “Electronic Repair Concepts for Long-Duration Spaceflight”, *45th AIAA Aerospace Sciences Meeting and Exhibit*, Reno, NV, January 8-11, 2007, AIAA-2007-545.

[6] Grugel, R.N., Cotton, L.J., Segre, P.N., Ogle, J.A., Funkhouser, G., Parris, F., Murphy, L., Gillies, D., Hua, F., and Anilkumar, A.V., “The In-Space Soldering Investigation (ISSI): Melting and Solidification Experiment Aboard the International Space Station”, *44th AIAA Aerospace Sciences Meeting and Exhibit*, Reno, NV, January 8-12, 2006, AIAA 2006-521.

[7] Mankins, J.C., “Technology Readiness Levels: A White Paper” April 6, 1995, Advanced Concepts Office, Office of Space Access and Technology, URL: http://www.hq.nasa.gov/office/codeq/trl/trl.pdf [cited 26 December 2007].

[8] Watson, J.K., Struk, P.M., Pettegrew, R.D., and Downs, R.S., Experimental Investigation of Solder Joint Defect Formation and Mitigation in Reduced-Gravity Environments, *AIAA Journal of Spacecraft and Rockets*, Vol. 44, No. 1 (Jan-Feb 2007), pp. 174-182.

[9] Pettegrew, R.D., Struk, P.M., Watson, J.K., Haylett, D.R., and Downs, R.S. Gravitational Effects on Solder Joints, *Welding Journal*, 2003, pp. 44-48.

[10] Struk, P.M., Pettegrew R.D., Downs, R.S., and Watson, J.K., The Influence Of Gravity on Joint Shape for Through-Hole Soldering, *43rd AIAA Aerospace Sciences Meeting and Exhibit*, Reno, NV, January 10-13, 2005, AIAA-2005-0541 and NASA TM-213589, 2005.

[11] Struk, P.M., Pettegrew, R.D., Downs, R.S., and Watson, J.K., The Effects of an Unsteady Reduced Gravity Environment on the Soldering Process, *42nd AIAA Aerospace Sciences Meeting and Exhibit*, Reno, NV, January 5-8, 2004, AIAA-2004-1311 and NASA-TM 212946, 2004.

[12] Pettegrew, R.D., Struk, P.M., Watson, J.K., and Haylett, D.R., Experimental Methods in Reduced Gravity Research, NASA TM-211993, 2002.

[13] Easton, J.W., Struk, P.M., and Rotella, A., Imaging and Analysis of Void-Defects in Solder Joints Formed in Reduced Gravity Using High-Resolution Computed Tomography, *46th AIAA Aerospace Sciences Meeting and Exhibit*, Reno, NV, January 7-10, 2008, AIAA-2007-0824.

[14] Humprey, R., Addressing future test challenges via synthetic instrumentation, *Military Embedded Systems*, May 2006, URL: http://www.mil-embedded.com/articles/authors/humphrey/ [cited 26 December 2007].

[15] Wikipedia contributors, “Joint Test Action Group”, *Wikipedia, The Free Encyclopedia*, 16 December 2007, URL: http://en.wikipedia.org/w/index.php?title=Joint_Test_Action_Group&oldid=178335316 [cited 26 December 2007].

Studying Electronics

From: Electronics Club Home Page

The Contents:

- Block Diagrams
- Circuit Diagrams
- Circuit Symbols - including the functions of components
- Electricity and the Electron
- Series and Parallel Connections
- Voltage and Current
- Meters - voltmeters, ammeters, galvanometers and ohmmeters
- Multimeters - choosing and using
- Resistance - resistors in series and parallel, conductors and insulators
- Ohm's Law - including the VIR triangle and calculations
- Power and Energy
- AC, DC and Electrical Signals
- Oscilloscopes (CRO) - setting up; measuring voltage and time
- Power Supplies
- Transducers
- Voltage Dividers
- Transistor Circuits
- Analogue and Digital Systems
- Logic Gates
- Capacitance and Uses of Capacitors
- Impedance and Reactance
- 555 and 556 Timer Circuits
- Counting Circuits
- Quantities and Units used in Electronics

- Capacitors
- Connectors and Cables
- Diodes including zener diodes
- Integrated Circuits (Chips)
    - 4000 series logic ICs (pin connections etc)
    - 74 series logic ICs (pin connections etc)
- Lamps
- LEDs (Light Emitting Diodes)
- Relays
- Resistors
- Switches
- Transistors
- Heat sinks for transistors
- Variable Resistors
- Other components including LDR and Thermistor
- Soldering Guide
- Tools required for electronics
- Starter kit of components
- Frequently Asked Questions

Block Diagrams

Block diagrams are used to understand (and design) complete circuits by breaking them down into smaller sections or blocks. Each block performs a particular function and the block diagram shows how they are connected together. No attempt is made to show the components used within a block, only the inputs and outputs are shown. This way of looking at circuits is called the **systems** approach.

Power supply (or battery) connections are usually not shown on block diagrams.

Audio Amplifier System

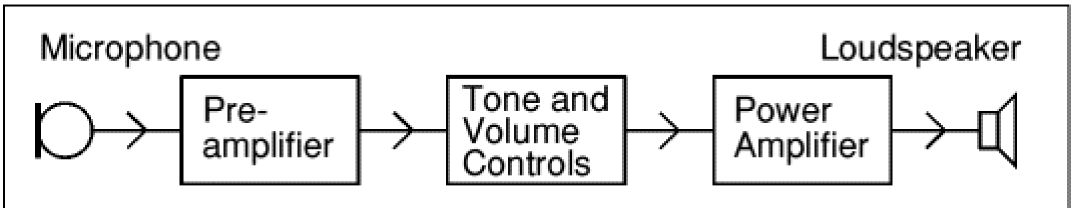

Block Diagram of an Audio Amplifier System

The power supply (not shown) is connected to the pre-amplifier and power amplifier blocks.

- **Microphone** - a transducer which converts sound to voltage.
- **Pre-Amplifier** - amplifies the small audio signal (voltage) from the microphone.
- **Tone and Volume Controls** - adjust the nature of the audio signal.
  The tone control adjusts the balance of high and low frequencies.
  The volume control adjusts the strength of the signal.
- **Power Amplifier** - increases the strength (power) of the audio signal.
- **Loudspeaker** - a transducer which converts the audio signal to sound.

Radio Receiver System

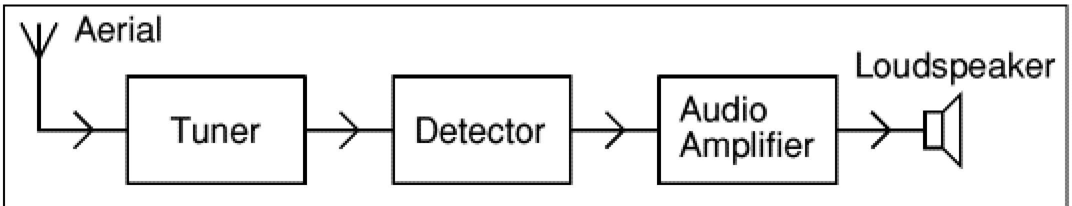

Block Diagram of a Radio Receiver System

The power supply (not shown) is connected to the audio amplifier block.

- **Aerial** - picks up radio signals from many stations.
- **Tuner** - selects the signal from just one radio station.
- **Detector** - extracts the audio signal carried by the radio signal.
- **Audio Amplifier** - increases the strength (power) of the audio signal.
  This could be broken down into the blocks like the Audio Amplifier System shown above.
- **Loudspeaker** - a transducer which converts the audio signal to sound.

Regulated Power Supply System

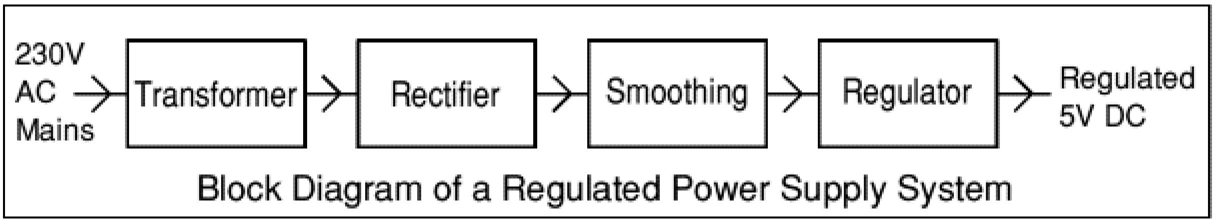
230V
AC
Mains
Transformer
Rectifier
Smoothing
Regulator
Regulated
5V DC
Block Diagram of a Regulated Power Supply System

- **Transformer** - steps down 230V AC mains to low voltage AC.
- **Rectifier** - converts AC to DC, but the DC output is varying.
- **Smoothing** - smooth the DC from varying greatly to a small ripple.
- **Regulator** - eliminates ripple by setting DC output to a fixed voltage.

Feedback Control System

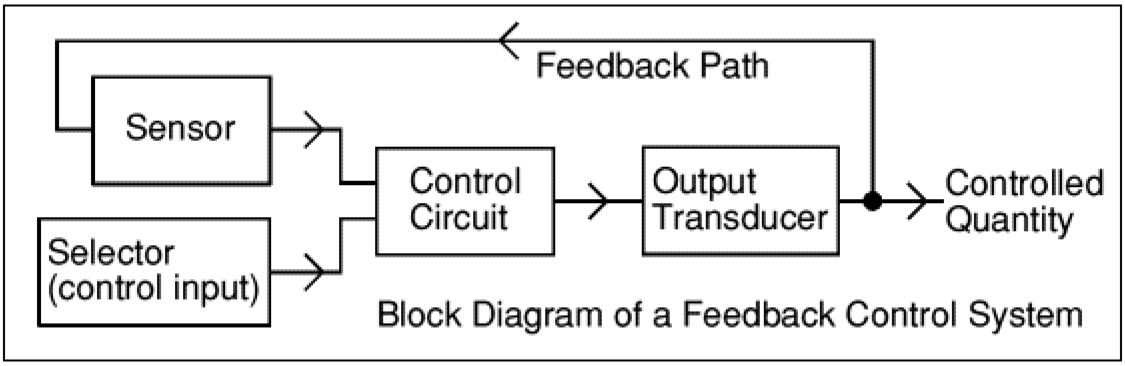
Feedback Path
Sensor
Control
Circuit
Output
Transducer
Controlled
Quantity
Selector
(control input)
Block Diagram of a Feedback Control System

The power supply (not shown) is connected to the control circuit block.

- **Sensor** - a transducer which converts the state of the controlled quantity to an electrical signal.
- **Selector (control input)** - selects the desired state of the output. Usually it is a variable resistor.
- **Control Circuit** - compares the desired state (control input) with the actual state (sensor) of the controlled quantity and sends an appropriate signal to the output transducer.
- **Output Transducer** - converts the electrical signal to the controlled quantity.
- **Controlled Quantity** - usually not an electrical quantity, e.g. motor speed.
- **Feedback Path** - usually not electrical, the Sensor detects the state of the controlled quantity.

Circuit Diagrams

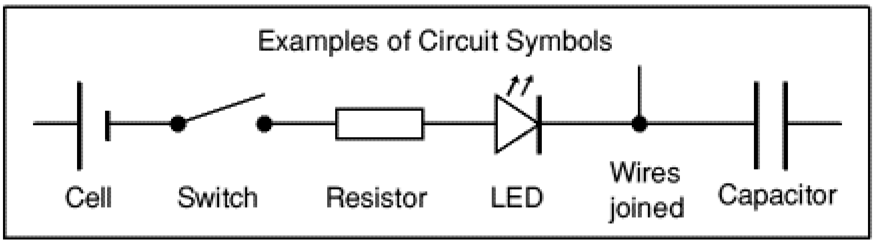
Examples of Circuit Symbols
Cell
Switch
Resistor
LED
Wires joined
Capacitor

Circuit diagrams show how electronic components are connected together. Each component is represented by a symbol and a few are shown here, for other symbols please see the Circuit Symbols page.

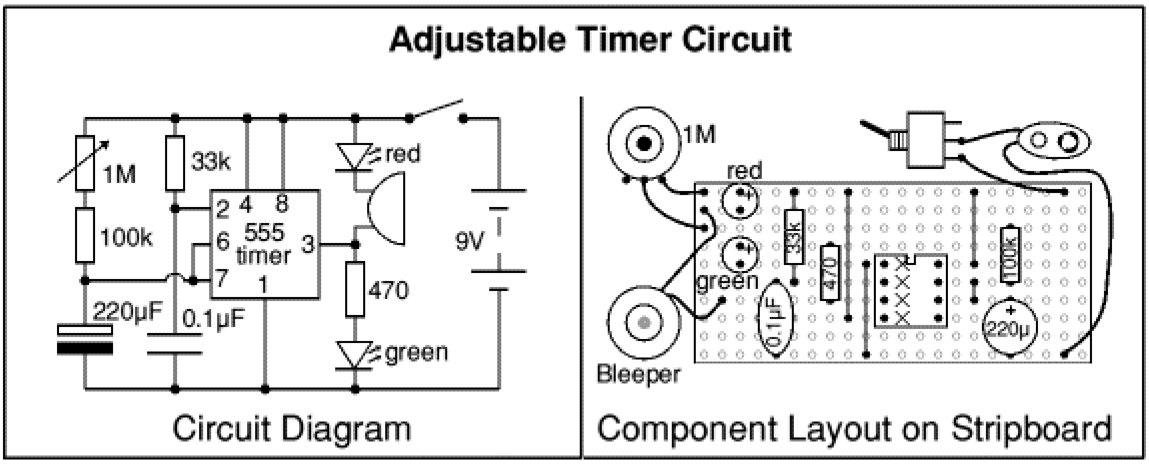
Adjustable Timer Circuit
1M
33k
100k
2 4 8
6 555 timer 3
7 1
red
9V
470
220µF
0.1µF
green
Circuit Diagram
1M
red
green
33k
470
100k
0.1µF
220µ
Bleeper
Component Layout on Stripboard

Circuit diagrams and component layouts

Circuit diagrams show the connections as clearly as possible with all wires drawn neatly as straight lines. The actual layout of the components is usually quite different from the circuit diagram and this can be confusing for the beginner. The secret is to concentrate on the *connections*, not the actual positions of components.

The circuit diagram and stripboard layout for the Adjustable Timer project are shown here so you can see the difference.

A circuit diagram is useful when testing a circuit and for understanding how it works. This is why the instructions for projects include a circuit diagram as well as the stripboard or printed circuit board layout which you need to build the circuit.

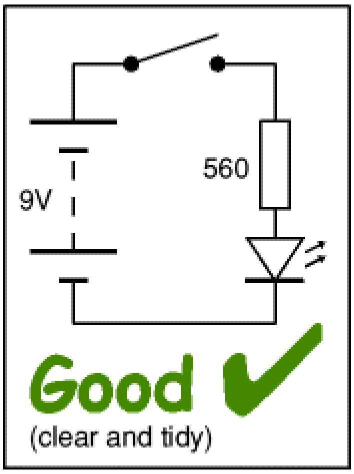
9V
560
Good
(clear and tidy)

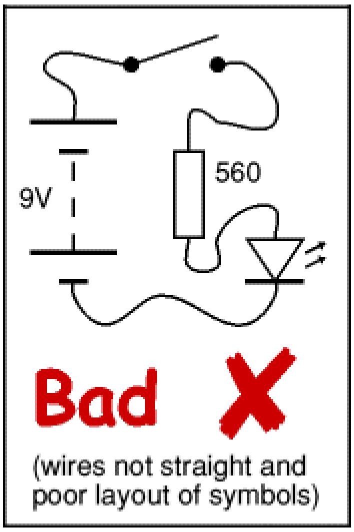
9V
560
Bad
(wires not straight and poor layout of symbols)

Drawing circuit diagrams

Drawing circuit diagrams is not difficult but it takes a little practice to draw neat, clear diagrams. This is a useful skill for science as well as for electronics. You will certainly need to draw circuit diagrams if you design your own circuits.

Follow these tips for best results:

- Make sure you use the correct symbol for each component.
- Draw connecting wires as straight lines (use a ruler).
- Put a 'blob' (●) at each junction between wires.
- Label components such as resistors and capacitors with their values.
- The positive (+) supply should be at the top and the negative (-) supply at the bottom. The negative supply is usually labeled 0V, zero volts.
  If you are drawing the circuit diagram for science please see the section about drawing diagrams the 'electronics way'.

If the circuit is complex:

- Try to arrange the diagram so that signals flow from left to right: inputs and controls should be on the left, outputs on the right.
- You may omit the battery or power supply symbols, but you must include (and label) the supply lines at the top and bottom.

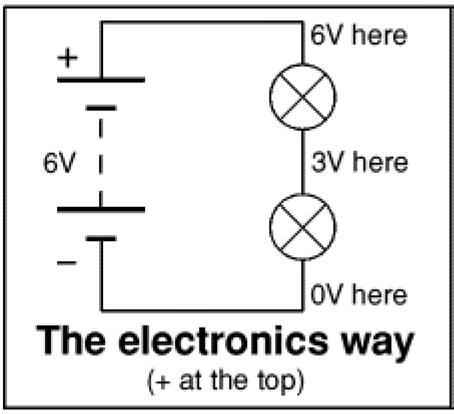
+
6V
–
6V here
3V here
0V here
The electronics way
(+ at the top)

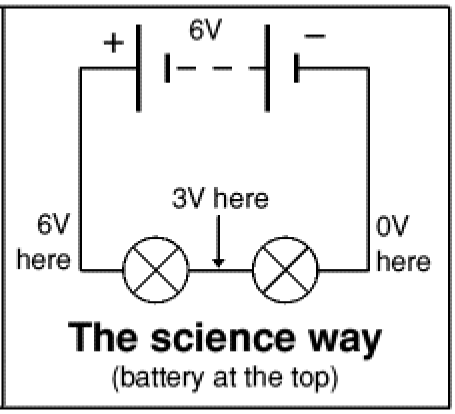
+
6V
–
6V here
3V here
0V here
The science way
(battery at the top)

Drawing circuit diagrams the 'electronics way'

Circuit diagrams for **electronics** are drawn with the positive (+) supply at the top and the negative (-) supply at the bottom. This can be helpful in understanding the operation of the circuit because the voltage decreases as you move down the circuit diagram.

Circuit diagrams for **science** are traditionally drawn with the battery or power supply at the top. This is not wrong, but there is usually no advantage in drawing them this way and I think it is less helpful for understanding the circuit.

I suggest that you always draw your circuit diagrams the 'electronics way', even for science!

Note that the negative supply is usually called 0V (zero volts). This is explained on the Voltage and Current page.

Circuit Symbols

Circuit symbols are used in circuit diagrams which show how a circuit is connected together. The actual layout of the components is usually quite different from the circuit diagram. To build a circuit you need a different diagram showing the layout of the parts on stripboard or printed circuit board.

Wires and connections

Component	Circuit Symbol	Function of Component
Wire	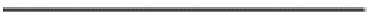	To pass current very easily from one part of a circuit to another.
Wires joined	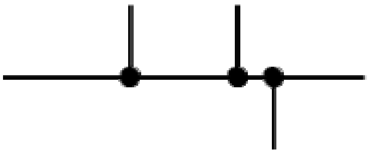	A 'blob' should be drawn where wires are connected (joined), but it is sometimes omitted. Wires connected at 'crossroads' should be staggered slightly to form two T-junctions, as shown on the right.
Wires not joined	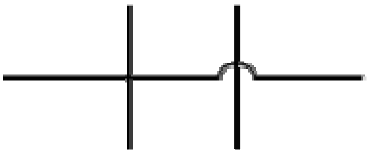	In complex diagrams it is often necessary to draw wires crossing even though they are not connected. I prefer the 'bridge' symbol shown on the right because the simple crossing on the left may be misread as a join where you have forgotten to add a 'blob'!

Power Supplies

Component	Circuit Symbol	Function of Component
Cell	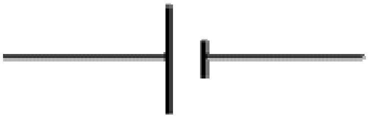	Supplies electrical energy. The larger terminal (on the left) is positive (+). A single cell is often called a battery, but strictly a battery is two or more cells joined together.
Battery	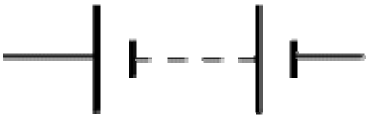	Supplies electrical energy. A battery is more than one cell. The larger terminal (on the left) is positive (+).
DC supply	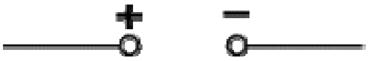	Supplies electrical energy. DC = Direct Current, always flowing in one direction.
AC supply	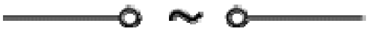	Supplies electrical energy. AC = Alternating Current, continually changing direction.
Fuse	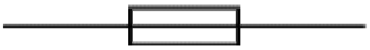	A safety device which will 'blow' (melt) if the current flowing through it exceeds a specified value.
Transformer	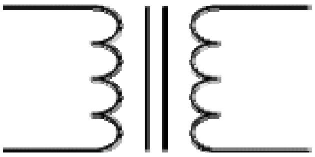	Two coils of wire linked by an iron core. Transformers are used to step up (increase) and step down (decrease) AC voltages. Energy is transferred between the coils by the magnetic field in the core. There is no electrical connection between the coils.
Earth (Ground)	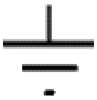	A connection to earth. For many electronic circuits this is the 0V (zero volts) of the power supply, but for mains electricity and some radio circuits it really means the earth. It is also known as ground.

Output Devices: Lamps, Heater, Motor, etc.

Component	Circuit Symbol	Function of Component
Lamp (lighting)	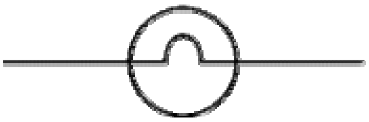	A transducer which converts electrical energy to light. This symbol is used for a lamp providing illumination, for example a car headlamp or torch bulb.
Lamp (indicator)	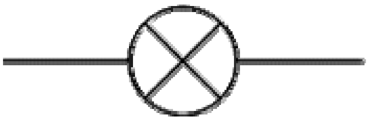	A transducer which converts electrical energy to light. This symbol is used for a lamp which is an indicator, for example a warning light on a car dashboard.
Heater	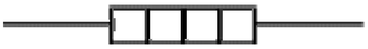	A transducer which converts electrical energy to heat.
Motor	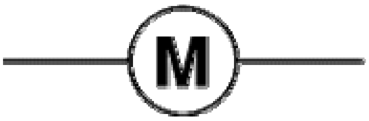	A transducer which converts electrical energy to kinetic energy (motion).

Bell	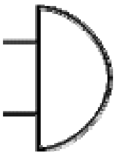	A transducer which converts electrical energy to sound.
Buzzer	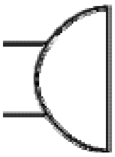	A transducer which converts electrical energy to sound.
Inductor (Coil, Solenoid)	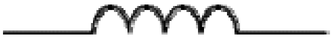	A coil of wire which creates a magnetic field when current passes through it. It may have an iron core inside the coil. It can be used as a transducer converting electrical energy to mechanical energy by pulling on something.

Switches

Component	**Circuit Symbol**	**Function of Component**
Push Switch (push-to-make)	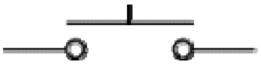	A push switch allows current to flow only when the button is pressed. This is the switch used to operate a doorbell.
Push-to-Break Switch	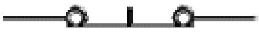	This type of push switch is normally closed (on), it is open (off) only when the button is pressed.
On-Off Switch (SPST)	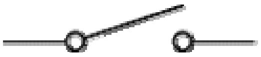	SPST = Single Pole, Single Throw. An on-off switch allows current to flow only when it is in the closed (on) position.
2-way Switch (SPDT)	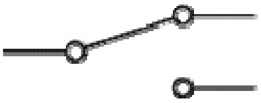	SPDT = Single Pole, Double Throw. A 2-way changeover switch directs the flow of current to one of two routes according to its position. Some SPDT switches have a central off position and are described as 'on-off-on'.
Dual On-Off Switch (DPST)	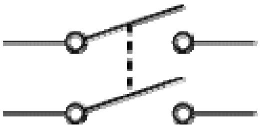	DPST = Double Pole, Single Throw. A dual on-off switch which is often used to switch mains electricity because it can isolate both the live and neutral connections.
Reversing Switch (DPDT)	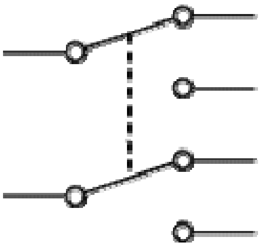	DPDT = Double Pole, Double Throw. This switch can be wired up as a reversing switch for a motor. Some DPDT switches have a central off position.
Relay	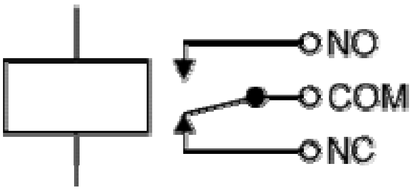	An electrically operated switch, for example a 9V battery circuit connected to the coil can switch a 230V AC mains circuit. NO = Normally Open, COM = Common, NC = Normally Closed.

Resistors

Component	Circuit Symbol	Function of Component
Resistor	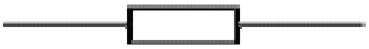	A resistor restricts the flow of current, for example to limit the current passing through an LED. A resistor is used with a capacitor in a timing circuit. Some publications still use the old resistor symbol: 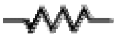
Variable Resistor (Rheostat)	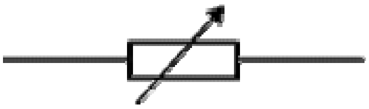	This type of variable resistor with 2 contacts (a rheostat) is usually used to control current. Examples include: adjusting lamp brightness, adjusting motor speed, and adjusting the rate of flow of charge into a capacitor in a timing circuit.
Variable Resistor (Potentiometer)	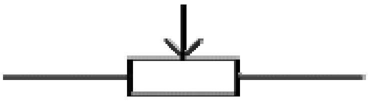	This type of variable resistor with 3 contacts (a potentiometer) is usually used to control voltage. It can be used like this as a transducer converting position (angle of the control spindle) to an electrical signal.
Variable Resistor (Preset)	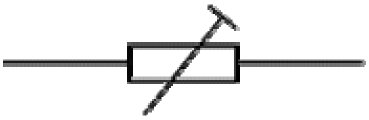	This type of variable resistor (a preset) is operated with a small screwdriver or similar tool. It is designed to be set when the circuit is made and then left without further adjustment. Presets are cheaper than normal variable resistors so they are often used in projects to reduce the cost.

Capacitors

Component	Circuit Symbol	Function of Component
Capacitor	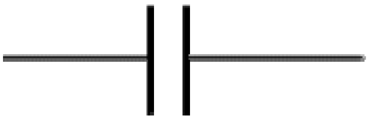	A capacitor stores electric charge. A capacitor is used with a resistor in a timing circuit. It can also be used as a filter, to block DC signals but pass AC signals.
Capacitor, polarized	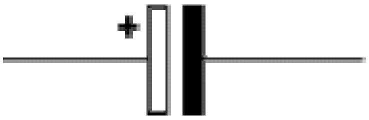	A capacitor stores electric charge. This type must be connected the correct way round. A capacitor is used with a resistor in a timing circuit. It can also be used as a filter, to block DC signals but pass AC signals.
Variable Capacitor	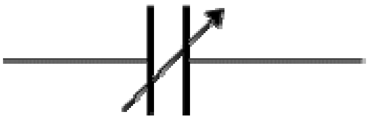	A variable capacitor is used in a radio tuner.
Trimmer Capacitor	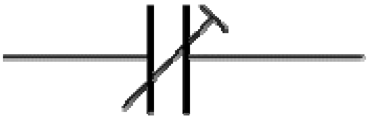	This type of variable capacitor (a trimmer) is operated with a small screwdriver or similar tool. It is designed to be set when the circuit is made and then left without further adjustment.

Diodes

Component	Circuit Symbol	Function of Component
Diode	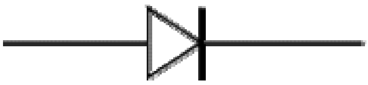	A device which only allows current to flow in one direction.
LED Light Emitting Diode	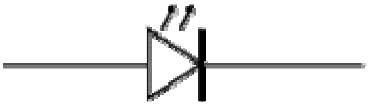	A transducer which converts electrical energy to light.
Zener Diode	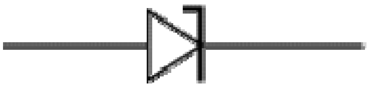	A special diode which is used to maintain a fixed voltage across its terminals.
Photodiode	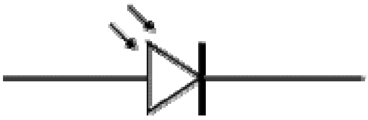	A light-sensitive diode.

Transistors

Component	Circuit Symbol	Function of Component
Transistor NPN	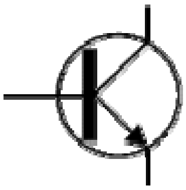	A transistor amplifies current. It can be used with other components to make an amplifier or switching circuit.
Transistor PNP	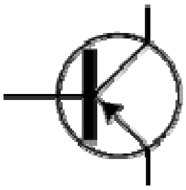	A transistor amplifies current. It can be used with other components to make an amplifier or switching circuit.
Phototransistor	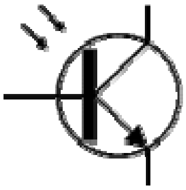	A light-sensitive transistor.

Audio and Radio Devices

Component	Circuit Symbol	Function of Component
Microphone	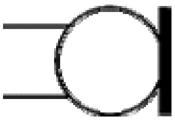	A transducer which converts sound to electrical energy.
Earphone	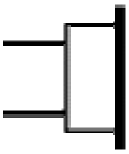	A transducer which converts electrical energy to sound.

Loudspeaker	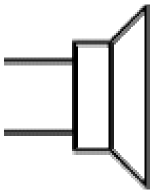	A transducer which converts electrical energy to sound.
Piezo Transducer	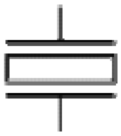	A transducer which converts electrical energy to sound.
Amplifier (general symbol)	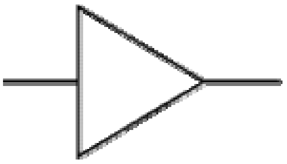	An amplifier circuit with one input. Really it is a block diagram symbol because it represents a circuit rather than just one component.
Aerial (Antenna)	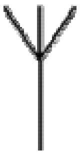	A device which is designed to receive or transmit radio signals. It is also known as an antenna.

Meters and Oscilloscope

Component	Circuit Symbol	Function of Component
Voltmeter	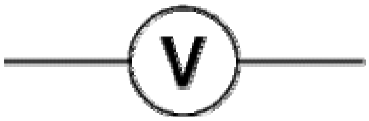	A voltmeter is used to measure voltage. The proper name for voltage is 'potential difference', but most people prefer to say voltage!
Ammeter	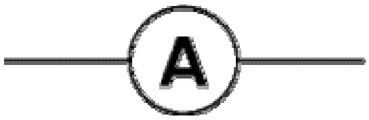	An ammeter is used to measure current.
Galvanometer	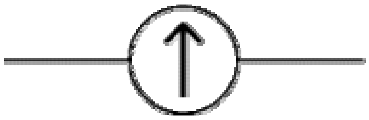	A galvanometer is a very sensitive meter which is used to measure tiny currents, usually 1mA or less.
Ohmmeter	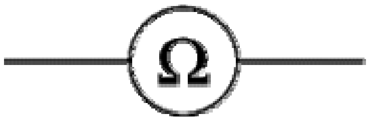	An ohmmeter is used to measure resistance. Most multimeters have an ohmmeter setting.
Oscilloscope	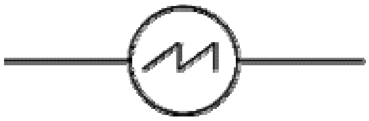	An oscilloscope is used to display the shape of electrical signals and it can be used to measure their voltage and time period.

Sensors (input devices)

Component	Circuit Symbol	Function of Component
LDR	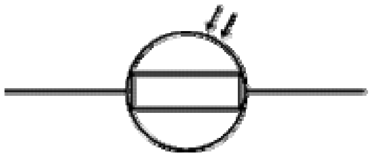	A transducer which converts brightness (light) to resistance (an electrical property). LDR = Light Dependent Resistor
Thermistor	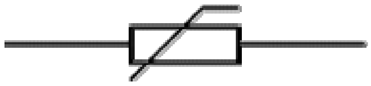	A transducer which converts temperature (heat) to resistance (an electrical property).

Logic Gates

Logic gates process signals which represent **true** (1, high, +Vs, on) or **false** (0, low, 0V, off). There are two sets of symbols: traditional and IEC (International Electrotechnical Commission).

Gate Type	Traditional Symbol	IEC Symbol	Function of Gate
NOT	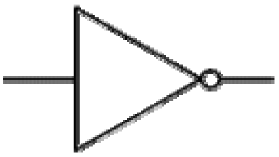	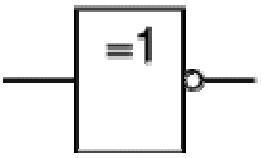	A NOT gate can only have one input. The 'o' on the output means 'not'. The output of a NOT gate is the inverse (opposite) of its input, so the output is true when the input is false. A NOT gate is also called an inverter.
AND	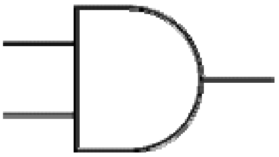	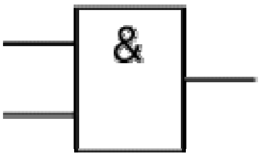	An AND gate can have two or more inputs. The output of an AND gate is true when all its inputs are true.
NAND	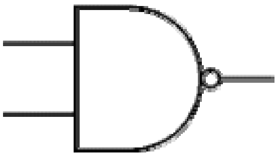	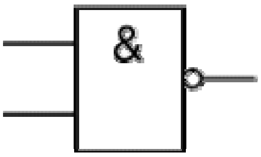	A NAND gate can have two or more inputs. The 'o' on the output means 'not' showing that it is a Not AND gate. The output of a NAND gate is true unless all its inputs are true.
OR	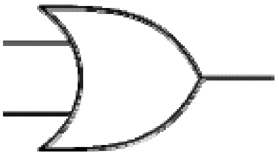	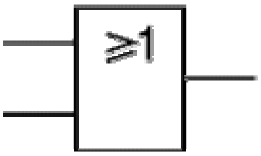	An OR gate can have two or more inputs. The output of an OR gate is true when at least one of its inputs is true.
NOR	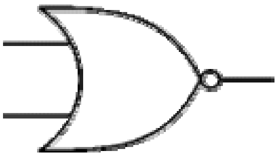	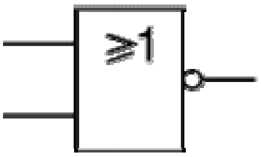	A NOR gate can have two or more inputs. The 'o' on the output means 'not' showing that it is a Not OR gate. The output of a NOR gate is true when none of its inputs are true.
EX-OR	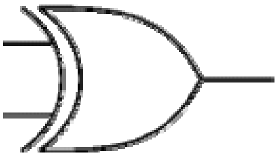	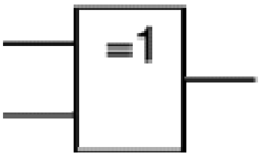	An EX-OR gate can only have two inputs. The output of an EX-OR gate is true when its inputs are different (one true, one false).
EX-NOR	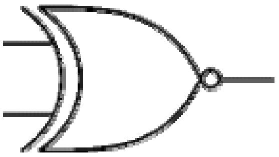	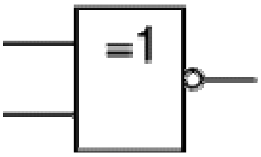	An EX-NOR gate can only have two inputs. The 'o' on the output means 'not' showing that it is a Not EX-OR gate. The output of an EX-NOR gate is true when its inputs are the same (both true or both false).

Sets of circuit symbols to download

You can download complete sets of all the circuit symbols shown above. The sets are 'zipped' for convenience and they are provided in three formats:

- WMF circuit symbols (32K) - Windows Metafiles.
  These vector drawings are the best format for printed documents on most computer systems, including Windows where they can be used in Word documents for example.

They can be enlarged without loss of quality. If you are not sure which format is best for you I suggest you try this one first.

- GIF circuit symbols (43K) - Graphics Interchange Format.
  These bitmap images are the best format for web pages but they print poorly and their bitmap nature will become obvious if they are enlarged. You can download individual symbols by saving the images used above on this page.
- Drawfile circuit symbols (29K) - for RISC OS (Acorn) computers.
  These high quality vector drawings are suitable for almost all documents on a RISC OS computer. All the symbols were originally drawn in this format. They print perfectly and can be enlarged without loss of quality. Sorry, this format is NOT suitable for Windows computers.

Electricity and the Electron

What is electricity?

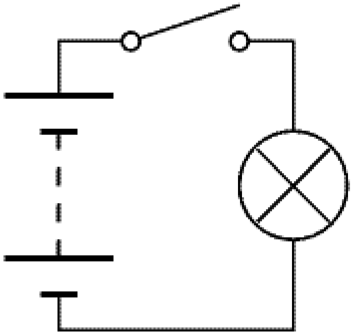

Electricity is the **flow of charge** around a circuit **carrying energy** from the battery (or power supply) to components such as lamps and motors.

Electricity can flow only if there is a **complete circuit** from the battery through wires to components and back to the battery again.

The diagram shows a simple circuit of a battery, wires, a switch and a lamp. The switch works by breaking the circuit.

With the **switch open** the circuit is broken - so electricity cannot flow and the lamp is off.

With the **switch closed** the circuit is complete - allowing electricity to flow and the lamp is on. The electricity is carrying energy from the battery to the lamp.

We can see, hear or feel the effects of electricity flowing such as lamp lighting, a bell ringing, or a motor turning - but we cannot see the electricity itself, so which way is it flowing?

Picture below shows an Imaginary positive particles moving in the direction of the conventional current.

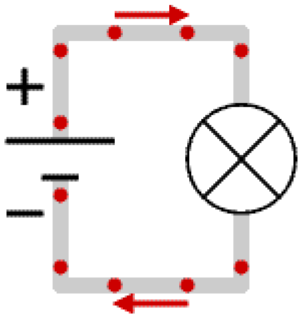

Which way does electricity flow?

We say that electricity flows from the positive (+) terminal of a battery to the negative (-) terminal of the battery. We can imagine particles with positive electric charge flowing in this direction around the circuit, like the red dots in the diagram.

This flow of electric charge is called **conventional current**.

This direction of flow is used throughout electronics and it is the one you should remember and use to understand the operation of circuits.

However this is not the whole answer because the particles that move in fact have negative charge! And they flow in the opposite direction!

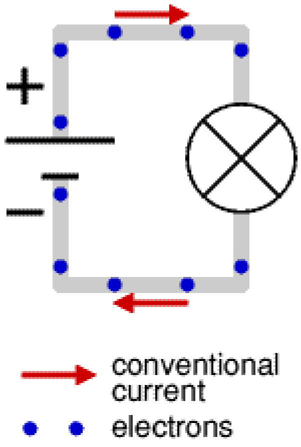

The Electron

When electricity was discovered scientists tried many experiments to find out which way the electricity was flowing around circuits, but in those early days they found it was impossible to find the direction of flow.

They knew there were two types of electric charge, positive (+) and negative (-), and they decided to say that electricity was a flow of positive charge from + to -. They knew this was a guess, but a decision had to be made! Everything known at that time could also be explained if electricity was negative charge flowing the other way, from - to +.

The electron was discovered in 1897 and it was found to have a negative charge. The guess made in the early days of electricity was wrong! Electricity in almost all conductors is really the flow of electrons (negative charge) from - to +.

By the time the electron was discovered the idea of electricity flowing from + to - (conventional current) was firmly established. Luckily it is not a problem to think of electricity in this way because positive charge flowing forwards is equivalent to negative charge flowing backwards.

To prevent confusion you should **always use conventional current** when trying to understand how circuits work, imagine positively charged particles flowing from + to -.

Series and Parallel Connections

Connecting Components

There are two ways of connecting components:

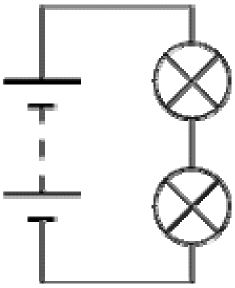

In series

so that each component has the **same current**.

The battery voltage is divided between the two lamps
Each lamp will have half the battery voltage if the lamps are identical.

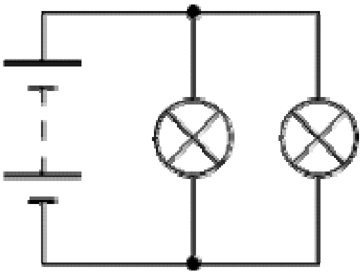

In parallel

so that each component has the **same voltage**.

Both lamps have the full battery voltage across them.
The battery current is divided between the two lamps.

Most circuits contain a mixture of series and parallel connections

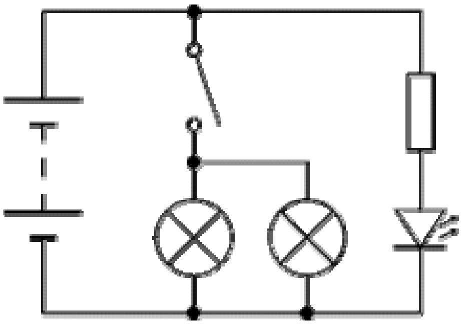

The terms *series circuit* and *parallel circuit* are sometimes used, but only the simplest of circuits are entirely one type or the other. It is better to refer to specific components and say they are *connected in series* or *connected in parallel*.

For example: the circuit on the right shows a **resistor and LED connected in series** (on the right) and **two lamps connected in parallel** (in the centre). The switch is connected in series with the two lamps.

Lamps in Series

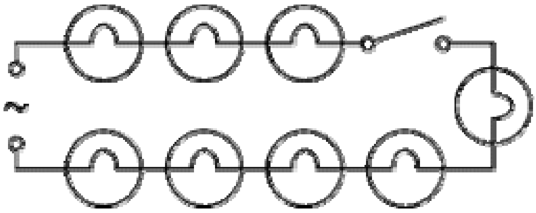

If several lamps are connected in series they will all be switched on and off together by a switch connected anywhere in the circuit. The supply voltage is divided equally between the lamps (assuming they are all identical). If one lamp blows, all the lamps will go out because the circuit is broken.

Christmas Tree Lights

The lamps on a Christmas tree are connected in series. Normally you would expect all the lamps to go out if one blew, but Christmas tree lamps are special! They are designed to short circuit (conduct like a wire link) when they blow, so the circuit is not broken and the other lamps remain lit, making it easier to locate the faulty lamp. Sets also include one 'fuse' lamp which blows normally.

If there are 20 lamps and the mains electricity voltage is 240V, each lamp must be suitable for a 12V supply because the 240V is divided equally between the 20 lamps: 240V ÷ 20 = 12V.

WARNING! The Christmas tree lamps may seem safe because they use only 12V but they are connected to the mains supply which can be lethal. Always unplug from the mains before changing lamps. The voltage across the holder of a missing lamp is the full 240V of the mains supply! (Yes, it really is!)

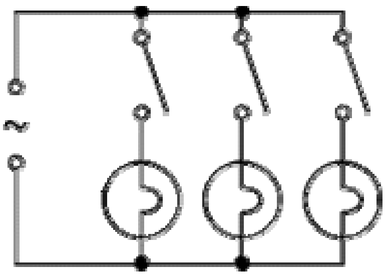

Lamps in Parallel

If several lamps are connected in parallel each one has the full supply voltage across it. The lamps may be switched on and off independently by connecting a **switch in series with each lamp** as shown in the circuit diagram. This arrangement is used to control the lamps in buildings.

This type of circuit is often called a *parallel circuit* but you can see that it is not really so simple - the switches are in series with the lamps, and it is these **switch and lamp pairs** that are connected in parallel.

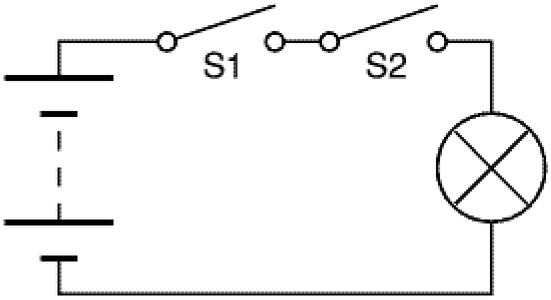

Switches in Series

If several on-off switches are connected in series they must all be closed (on) to complete the circuit.

The diagram shows a simple circuit with two switches connected in series to control a lamp.

Switch S1 **AND** Switch S2 must be closed to light the lamp.

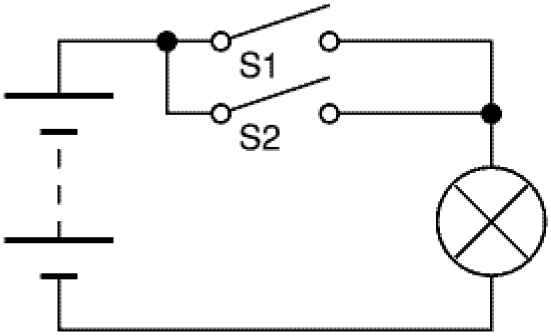

Switches in Parallel

If several on-off switches are connected in parallel only one needs to be closed (on) to complete the circuit.

The diagram shows a simple circuit with two switches connected in parallel to control a lamp.

Switch S1 **OR** Switch S2 (or both of them) must be closed to light the lamp.

Voltage and Current

Voltage and Current are vital to understanding electronics, but they are quite hard to grasp because we can't see them directly.

Voltage is the Cause, Current is the Effect

Voltage attempts to make a current flow, and current will flow if the circuit is complete. Voltage is sometimes described as the 'push' or 'force' of the electricity, it isn't really a force but this may help you to imagine what is happening. It is possible to have voltage without current, but current cannot flow without voltage.

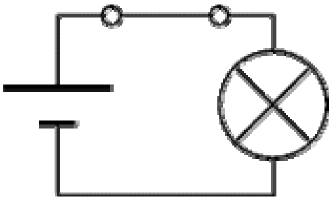

Voltage and Current
The switch is closed making a complete circuit so current can flow.

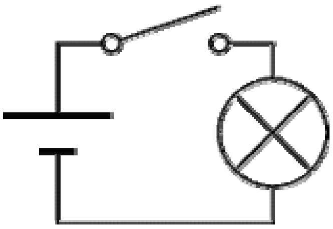

Voltage but No Current
The switch is open so the circuit is broken and current cannot flow.

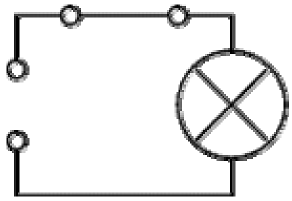

No Voltage and No Current
Without the cell there is no source of voltage so current cannot flow.

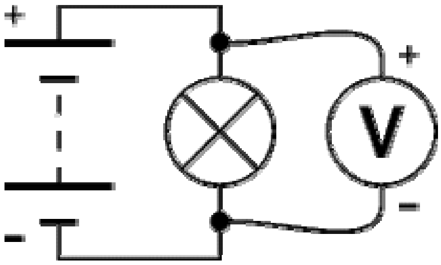

Connecting a voltmeter in parallel

Voltage, V

- Voltage is a measure of the **energy carried by the charge**.
  Strictly: voltage is the "energy per unit charge".
- The proper name for voltage is **potential difference** or p.d. for short, but this term is rarely used in electronics.
- Voltage is **supplied by the battery** (or power supply).
- Voltage is **used up in components**, but not in wires.
- We say **voltage across** a component.
- Voltage is measured in **volts**, **V**.
- Voltage is measured with a **voltmeter**, connected in **parallel**.
- The symbol **V** is used for voltage in equations.

Voltage at a point and 0V (zero volts)

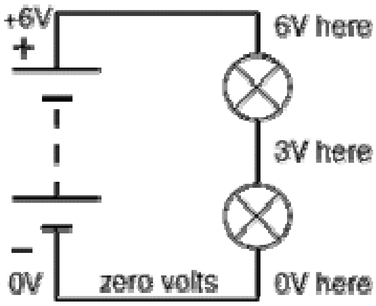

Voltage is a **difference between two points**, but in electronics we often refer to **voltage at a point** meaning the voltage difference between that point and a reference point of 0V (zero volts).

Zero volts could be any point in the circuit, but to be consistent it is normally the **negative terminal of the battery or power supply**. You will often see circuit diagrams labeled with 0V as a reminder.

You may find it helpful to think of voltage like height in geography. The reference point of zero height is the mean (average) sea level and all heights are measured from that point. The zero volts in an electronic circuit is like the mean sea level in geography.

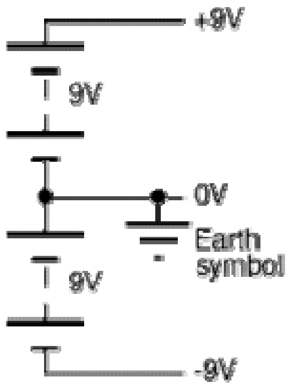

Zero volts for circuits with a dual supply

Some circuits require a dual supply with **three** supply connections as shown in the diagram. For these circuits the zero volts reference point is the **middle terminal** between the two parts of the supply.

On complex circuit diagrams using a dual supply the earth symbol is often used to indicate a connection to 0V, this helps to reduce the number of wires drawn on the diagram.

The diagram shows a ±9V dual supply, the positive terminal is +9V, the negative terminal is -9V and the middle terminal is 0V.

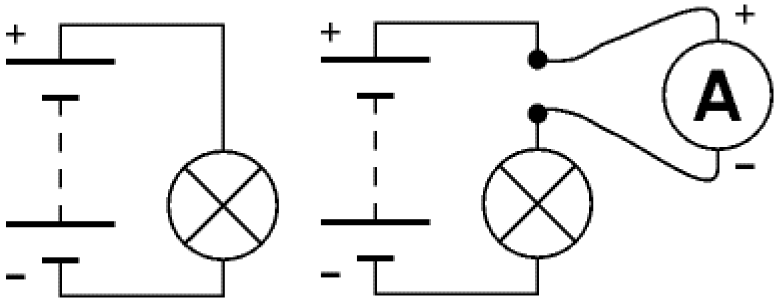

Connecting an ammeter in series

Current, I

- Current is the **rate of flow of charge**.
- Current **is not used up**, what flows into a component must flow out.
- We say **current through** a component.
- Current is measured in **amps (amperes)**, **A**.
- Current is measured with an **ammeter**, connected in **series**.
  To connect in series you must break the circuit and put the ammeter across the gap, as shown in the diagram.
- The symbol **I** is used for current in equations.
  Why is the letter I used for current? ... see Frequently Ask Questions.

1A (1 amp) is quite a large current for electronics, so mA (milliamps) are often used. m (milli) means "thousandth":

1mA = 0.001A, or 1000mA = 1A

The need to break the circuit to connect in series means that ammeters are difficult to use on soldered circuits. Most testing in electronics is done with voltmeters which can be easily connected without disturbing circuits.

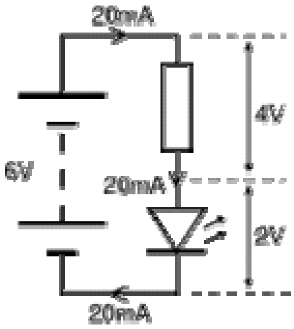

Voltage and Current for components in Series

Voltages add up for components connected in series.
Currents are the same through all components connected in series.

In this circuit the 4V across the resistor and the 2V across the LED add up to the battery voltage: 2V + 4V = 6V.

The current through all parts (battery, resistor and LED) is 20mA.

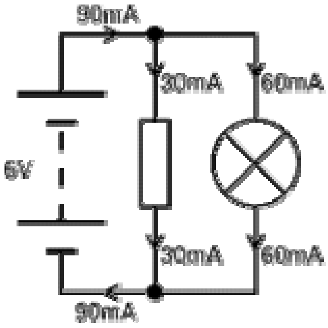

Voltage and Current for components in Parallel

Voltages are the same across all components connected in parallel.
Currents add up for components connected in parallel.

In this circuit the battery, resistor and lamp all have 6V across them.

The 30mA current through the resistor and the 60mA current through the lamp add up to the 90mA current through the battery.

Meters

Analogue display

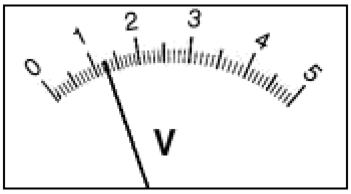

Analogue displays have a pointer which moves over a graduated scale. They can be difficult to read because of the need to work out the value of the smallest scale division. For example the scale in the picture has 10 small divisions between 0 and 1 so each small division represents 0.1. The reading is therefore **1.25V** (the pointer is estimated to be half way between 1.2 and 1.3).

The maximum reading of an analogue meter is called **full-scale deflection** or **FSD** (it is 5V in the example shown).

Analogue meters must be connected the correct way round to prevent them being damaged when the pointer tries to move in the wrong direction. They are useful for monitoring continuously changing values (such as the voltage across a capacitor discharging) and they can be good for quick rough readings because the movement of the pointer can be seen without looking away from the circuit under test.

Taking accurate readings

To take an accurate reading from an analogue scale you must have your eye in line with the pointer. Avoid looking at an angle from the left or right because you will see a reading which is a little too high or too low. Many analogue meters have a small strip of **mirror** along the scale to help you. When your eye is in the correct position the reflection of the pointer is hidden behind the pointer itself. If you can see the reflection you are looking at an angle.

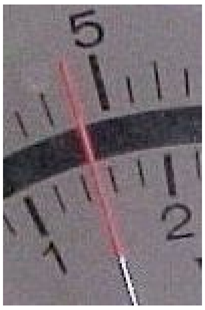

Correct
reflection hidden

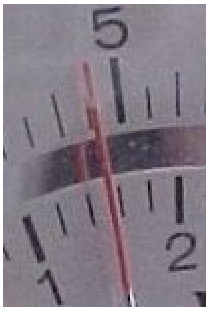

Wrong
reflection visible

Instead of a mirror, some meters have a **twisted pointer** to aid accurate readings. The end of the pointer is turned through 90° so it appears very thin when viewed correctly. The meter shown in the galvanometers section has a twisted pointer although it is too small to see in the picture.

Digital display

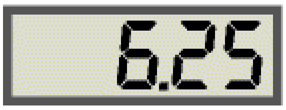

Values can be read directly from digital displays so they are easy to read accurately. It is normal for the least significant digit (on the right) to continually change between two or three values, this is a feature of the way digital meters work, not an error! Normally you will not need great precision and the least significant digit can be ignored or rounded up.

Digital meters may be connected either way round without damage, they will show a minus sign (-) when connected in reverse. If you exceed the maximum reading most digital meters show an almost blank display with just a 1 on the left-hand side.

All digital meters contain a battery to power the display so they use virtually no power from the circuit under test. This means that digital voltmeters have a very high resistance (usually called input impedance) of 1MΩ or more, usually 10MΩ, and they are very unlikely to affect the circuit under test.

For general use digital meters are the best type. They are easy to read, they may be connected in reverse and they are unlikely to affect the circuit under test.

Connecting meters

It is important to connect meters the correct way round:

- The **positive terminal** of the meter, marked + or colored **red** should be connected nearest to + on the battery or power supply.
- The **negative terminal** of the meter, marked - or colored **black** should be connected nearest to - on the battery or power supply.

Voltmeters

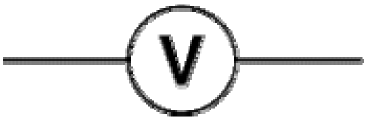

- Voltmeters measure **voltage**.
- Voltage is measured in **volts**, **V**.
- Voltmeters are connected in **parallel** across components.
- Voltmeters have a **very high resistance**.

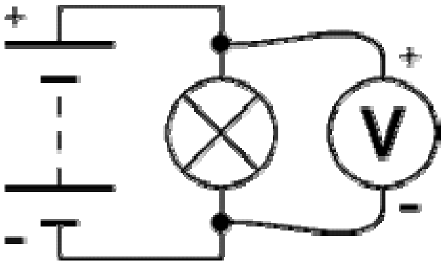

Connecting a voltmeter in parallel

Measuring voltage at a point

When testing circuits you often need to find the voltages at various points, for example the voltage at pin 2 of a 555 timer chip. This can seem confusing - where should you connect the second voltmeter lead?

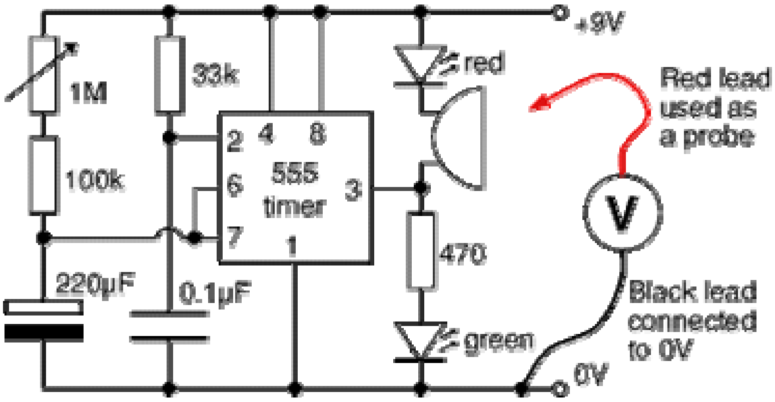

- Connect the **black** (negative -) voltmeter lead to 0V, normally the negative terminal of the battery or power supply.
- Connect the **red** (positive +) voltmeter lead to the point you where you need to measure the voltage.
- The **black** lead can be left permanently connected to 0V while you use the **red** lead as a probe to measure voltages at various points.
- You may wish to use a crocodile clip on the **black** lead to hold it in place.

Voltage at a point really means the voltage difference between that point and 0V (zero volts) which is normally the negative terminal of the battery or power supply. Usually 0V will be labelled on the circuit diagram as a reminder.

Analogue meters take a little power from the circuit under test to operate their pointer. This may upset the circuit and give an incorrect reading. To avoid this voltmeters should have a resistance of at least 10 times the circuit resistance (take this to be the highest resistor value near where the meter is connected).

Most analogue voltmeters used in school science are not suitable for electronics because their resistance is too low, typically a few kΩ. 100kΩ or more is required for most electronics circuits.

Ammeters

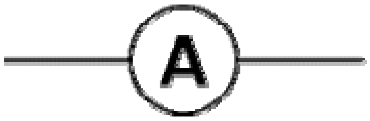

- Ammeters measure **current**.
- Current is measured in **amps (amperes)**, **A**.
  1A is quite large, so mA (milliamps) and µA (microamps) are often used. 1000mA = 1A, 1000µA = 1mA, 1000000µA = 1A.
- Ammeters are connected in **series**.
  To connect in series you must break the circuit and put the ammeter across the gap, as shown in the diagram.
- Ammeters have a **very low resistance**.

The need to break the circuit to connect in series means that ammeters are difficult to use on soldered circuits. Most testing in electronics is done with voltmeters which can be easily connected without disturbing circuits.

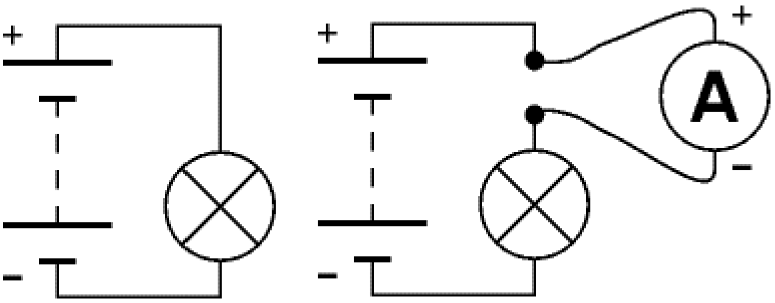

Connecting an ammeter in series

Galvanometers

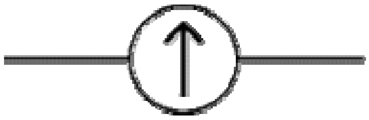

Galvanometers are very sensitive meters which are used to measure tiny currents, usually 1mA or less. They are used to make all types of analogue meters by adding suitable resistors as shown in the diagrams below. The photograph shows an educational 100µA galvanometer for which various multipliers and shunts are available.

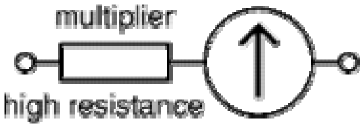
multiplier
high resistance

Making a Voltmeter
A galvanometer with a high resistance **multiplier** in series to make a voltmeter.

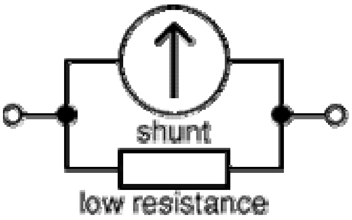
shunt
low resistance

Making an Ammeter
A galvanometer with a low resistance **shunt** in parallel to make an ammeter.

UNILAB
V
UNILAB
A
20-0-100
µA=
1000Ω

Galvanometer with multiplier and shunt
Maximum meter current 100µA (or 20µA reverse).
This meter is unusual in allowing small reverse readings to be shown.

Ohmmeters

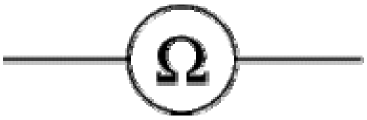

An ohmmeter is used to measure resistance in ohms (Ω). Ohmmeters are rarely found as separate meters but all standard multimeters have an ohmmeter setting.
1Ω is quite small so kΩ and MΩ are often used.

1kΩ = 1000Ω, 1MΩ = 1000kΩ = 1000000Ω.

Multimeters

Multimeters are very useful test instruments. By operating a multi-position switch on the meter they can be quickly and easily set to be a **voltmeter**, an **ammeter** or an **ohmmeter**. They have several settings (called 'ranges') for each type of meter and the choice of AC or DC.

Some multimeters have additional features such as transistor testing and ranges for measuring capacitance and frequency.

Analogue multimeters consist of a galvanometer with various resistors which can be switched in as multipliers (voltmeter ranges) and shunts (ammeter ranges).

Choosing a multimeter

The photographs below show modestly priced multimeters which are suitable for general electronics use, you should be able to buy meters like these for less than £15. A **digital multimeter** is the best choice for your first multimeter, even the cheapest will be suitable for testing simple projects.

If you are buying an **analogue multimeter** make sure it has a **high sensitivity** of 20kΩ/V or greater on DC voltage ranges, anything less is not suitable for electronics. The sensitivity is normally marked in a corner of the scale, ignore the lower AC value (sensitivity on AC ranges is less important), the higher DC value is the critical one. Beware of cheap analogue multimeters sold for electrical work on cars because their sensitivity is likely to be too low.

Digital multimeters

Digital Multimeter

All digital meters contain a battery to power the display so they use virtually no power from the circuit under test. This means that on their DC voltage ranges they have a very high resistance (usually called input impedance) of 1MΩ or more, usually 10MΩ, and they are very unlikely to affect the circuit under test.

Typical ranges for digital multimeters like the one illustrated:
(the values given are the maximum reading on each range)

- DC Voltage: 200mV, 2000mV, 20V, 200V, 600V.
- AC Voltage: 200V, 600V.
- DC Current: 200µA, 2000µA, 20mA, 200mA, 10A*.
  *The 10A range is usually unfused and connected via a special socket.
- AC Current: None. (You are unlikely to need to measure this).
- Resistance: 200Ω, 2000Ω, 20kΩ, 200kΩ, 2000kΩ, Diode Test.

Digital meters have a special diode test setting because their resistance ranges cannot be used to test diodes and other semiconductors.

Analogue multimeters

Analogue Multimeter

Analogue meters take a little power from the circuit under test to operate their pointer. They must have a high sensitivity of at least 20kΩ/V or they may upset the circuit under test and give an incorrect reading. See the section below on sensitivity for more details.

Batteries inside the meter provide power for the resistance ranges, they will last several years but you should avoid leaving the meter set to a resistance range in case the leads touch accidentally and run the battery flat.

Typical ranges for analogue multimeters like the one illustrated:
(the voltage and current values given are the maximum reading on each range)

- DC Voltage: 0.5V, 2.5V, 10V, 50V, 250V, 1000V.
- AC Voltage: 10V, 50V, 250V, 1000V.
- DC Current: 50µA, 2.5mA, 25mA, 250mA.
  A high current range is often missing from this type of meter.
- AC Current: None. (You are unlikely to need to measure this).
- Resistance: 20Ω, 200Ω, 2kΩ, 20kΩ, 200kΩ.
  These resistance values are in the middle of the scale for each range.

It is a good idea to leave an analogue multimeter set to a DC voltage range such as 10V when not in use. It is less likely to be damaged by careless use on this range, and there is a good chance that it will be the range you need to use next anyway!

Sensitivity of an analogue multimeter

Multimeters must have a high sensitivity of at least 20kΩ/V otherwise their resistance on DC voltage ranges may be too low to avoid upsetting the circuit under test and giving an incorrect reading. To obtain valid readings the meter resistance should be at least 10 times the circuit resistance (take this to be the highest resistor value near where the meter is connected). You can increase the meter resistance by selecting a higher voltage range, but this may give a reading which is too small to read accurately!

On any DC voltage range:

Analogue Meter Resistance = Sensitivity × Max. reading of range

e.g. a meter with 20kΩ/V sensitivity on its 10V range has a resistance of 20kΩ/V × 10V = 200kΩ.

By contrast, **digital multimeters** have a constant resistance of at least 1MΩ (often 10MΩ) on all their DC voltage ranges. This is more than enough for almost all circuits.

Measuring voltage and current with a multimeter

1. **Select a range** with a maximum greater than you expect the reading to be.
2. **Connect the meter**, making sure the leads are the correct way round.
   Digital meters can be safely connected in reverse, but an analogue meter may be damaged.
3. **If the reading goes off the scale:** immediately disconnect and select a higher range.

Multimeters are easily damaged by careless use so please take these precautions:

- Always disconnect the multimeter before adjusting the range switch.
- Always check the setting of the range switch **before** you connect to a circuit.
- Never leave a multimeter set to a current range (except when actually taking a reading). The greatest risk of damage is on the current ranges because the meter has a low resistance.

Measuring voltage at a point

When testing circuits you often need to find the voltages at various points, for example the voltage at pin 2 of a 555 timer chip. This can seem confusing - where should you connect the second multimeter lead?

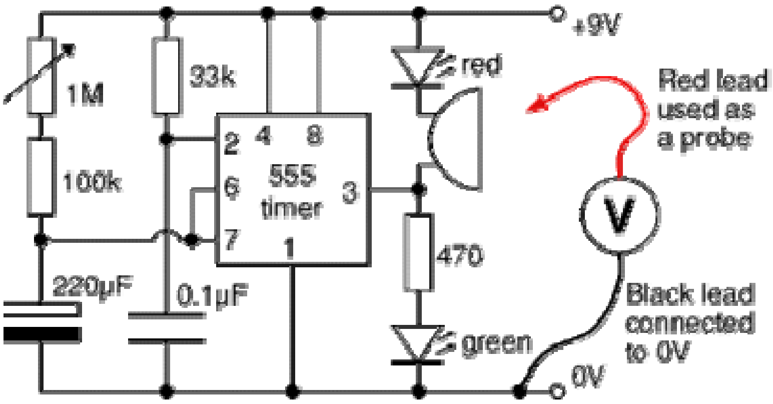

Measuring voltage at a point

- Connect the **black** (negative -) lead to 0V, normally the negative terminal of the battery or power supply.
- Connect the **red** (positive +) lead to the point you where you need to measure the voltage.
- The **black** lead can be left permanently connected to 0V while you use the **red** lead as a probe to measure voltages at various points.
- You may wish to fit a crocodile clip to the **black** lead of your multimeter to hold it in place while doing testing like this.

Voltage at a point really means the voltage difference between that point and 0V (zero volts) which is normally the negative terminal of the battery or power supply. Usually 0V will be labeled on the circuit diagram as a reminder.

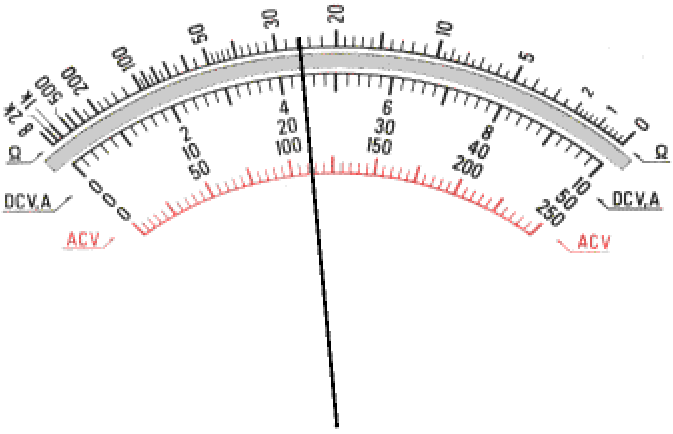

Analogue Multimeter Scales
These can appear daunting at first but remember
that you only need to read one scale at a time!
The top scale is used when measuring resistance.

Reading analogue scales

Check the setting of the range switch and **choose an appropriate scale**. For some ranges you may need to multiply or divide by 10 or 100 as shown in the sample readings below. For AC voltage ranges use the **red** markings because the calibration of the scale is slightly different.

Sample readings on the scales shown:
DC 10V range: 4.4V (read 0-10 scale directly)
DC 50V range: 22V (read 0-50 scale directly)
DC 25mA range: 11mA (read 0-250 and divide by 10)
AC 10V range: 4.45V (use the red scale, reading 0-10)

If you are not familiar with reading analogue scales generally you may wish to see the analogue display section on the general meters page.

Measuring resistance with a multimeter

To measure the resistance of a component it must not be connected in a circuit. If you try to measure resistance of components in a circuit you will obtain false readings (even if the supply is disconnected) and you may damage the multimeter.

The techniques used for each type of meter are very different so they are treated separately:

Measuring resistance with a DIGITAL multimeter

1. **Set the meter to a resistance range greater than you expect the resistance to be.**
   Notice that the meter display shows "off the scale" (usually blank except for a 1 on the left). Don't worry, this is not a fault, it is correct - the resistance of air is very high!
2. **Touch the meter probes together and check that the meter reads zero.**
   If it doesn't read zero, turn the switch to 'Set Zero' if your meter has this and try again.
3. **Put the probes across the component.**
   Avoid touching more than one contact at a time or **your** resistance will upset the reading!

Measuring resistance with an ANALOGUE multimeter

The resistance scale on an analogue meter is normally at the top, it is an unusual scale because it reads **backwards** and is **not linear** (evenly spaced). This is unfortunate, but it is due to the way the meter works.

1. **Set the meter to a suitable resistance range.**
   Choose a range so that the resistance you expect will be near the middle of the scale. For example: with the scale shown below and an expected resistance of about 50kΩ choose the × 1kΩ range.
2. **Hold the meter probes together and adjust the control on the front of the meter which is usually labeled** "0Ω ADJ" **until the pointer reads zero (on the RIGHT remember!).**
   If you can't adjust it to read zero, the battery inside the meter needs replacing.
3. **Put the probes across the component.**
   Avoid touching more than one contact at a time or **your** resistance will upset the reading!

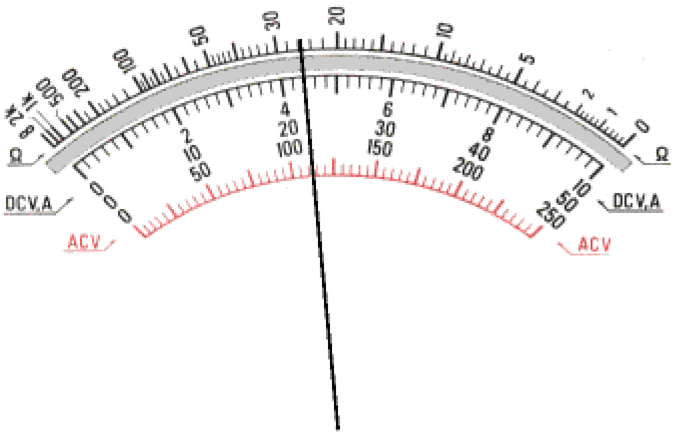

Analogue Multimeter Scales
The resistance scale is at the top, note that it reads
backwards and is not linear (evenly spaced).

Reading analogue resistance scales

For **resistance** use the **upper scale**, noting that it reads backwards and is not linear (evenly spaced).

Check the setting of the range switch so that you know by how much to multiply the reading.

Sample readings on the scales shown:
× 10Ω range: 260Ω
× 1kΩ range: 26kΩ

If you are not familiar with reading analogue scales generally you may wish to see the analogue display section on the general meters page.

Testing a diode with a multimeter

The techniques used for each type of meter are very different so they are treated separately:

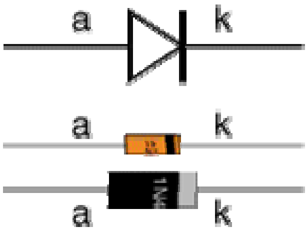

Diodes
a = anode
k = cathode

Testing a diode with a DIGITAL multimeter

- Digital multimeters have a special setting for testing a diode, usually labeled with the diode symbol.
- Connect the **red** (+) lead to the anode and the **black** (-) to the cathode. The diode should conduct and the meter will display a value (usually the voltage across the diode in mV, 1000mV = 1V).
- Reverse the connections. The diode should NOT conduct this way so the meter will display "off the scale" (usually blank except for a 1 on the left).

Testing a diode with an ANALOGUE multimeter

- Set the analogue multimeter to a low value resistance range such as × 10.
- It is essential to note that the polarity of analogue multimeter leads is reversed on the resistance ranges, so the **black** lead is positive (+) and the **red** lead is negative (-)! This is unfortunate, but it is due to the way the meter works.
- Connect the **black** (+) lead to anode and the **red** (-) to the cathode. The diode should conduct and the meter will display a low resistance (the exact value is not relevant).
- Reverse the connections. The diode should NOT conduct this way so the meter will show infinite resistance (on the left of the scale).

Testing a transistor with a multimeter

Set a digital multimeter to diode test and an analogue multimeter to a low resistance range such as × 10, as described above for testing a diode.

Test each pair of leads both ways (six tests in total):

- The **base-emitter (BE)** junction should behave like a diode and conduct one way only.
- The **base-collector (BC)** junction should behave like a diode and conduct one way only.
- The **collector-emitter (CE)** should not conduct either way.

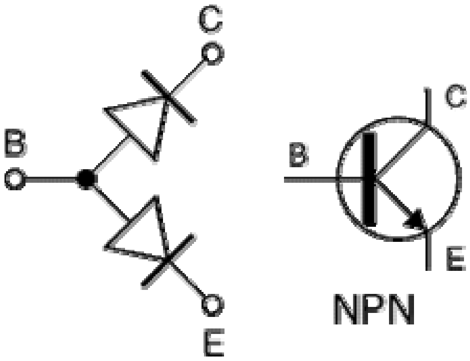

Testing an NPN transistor

The diagram shows how the junctions behave in an NPN transistor. The diodes are reversed in a PNP transistor but the same test procedure can be used.

Some multimeters have a 'transistor test' function; please refer to the instructions supplied with the meter for details.

Resistance

Resistance is the property of a component which **restricts the flow of electric current**. Energy is used up as the voltage across the component drives the current through it and this energy appears as heat in the component.

Resistance is measured in ohms, the symbol for ohm is an omega Ω.
1 Ωis quite small for electronics so resistances are often given in kΩ and MΩ.
1 kΩ = 1000 Ω   1 MΩ = 1000000 Ω.

Resistors used in electronics can have resistances as low as 0.1 Ωor as high as 10 MΩ.

Resistors connected in Series

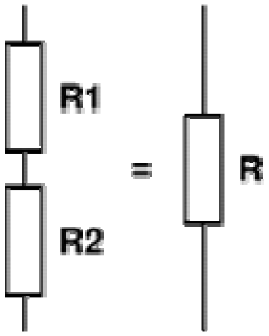

When resistors are connected in series their combined resistance is equal to the individual resistances added together. For example if resistors R1 and R2 are connected in series their combined resistance, R, is given by:

Combined resistance in **series**: **R = R1 + R2**

This can be extended for more resistors: **R = R1 + R2 + R3 + R4 + ...**

Note that the **combined resistance in series** will always be **greater** than any of the individual resistances.

Resistors connected in Parallel

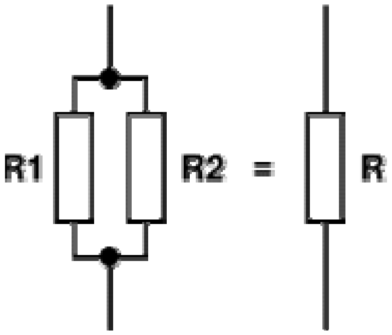

When resistors are connected in parallel their combined resistance is less than any of the individual resistances. There is a special equation for the combined resistance of **two** resistors R1 and R2:

Combined resistance of
two resistors in parallel: $$R = \frac{R1 \times R2}{R1 + R2}$$

For more than two resistors connected in parallel a more difficult equation must be used. This adds up the **reciprocal** ("one over") of each resistance to give the **reciprocal** of the combined resistance, R:

$$\frac{1}{R} = \frac{1}{R1} + \frac{1}{R2} + \frac{1}{R3} + \ldots$$

The simpler equation for **two** resistors in parallel is much easier to use!

Note that the **combined resistance in parallel** will always be **less** than any of the individual resistances.

Conductors, Semiconductors and Insulators

The resistance of an object depends on its **shape** and the **material** from which it is made. For a given material, objects with a smaller cross-section or longer length will have a greater resistance.

Materials can be divided into three groups:

- **Conductors** which have **low resistance**.
  Examples: metals (aluminum, copper, silver etc.) and carbon.
  Metals are used to make connecting wires, switch contacts and lamp filaments.
  Resistors are made from carbon or long coils of thin wire.
- **Semiconductors** which have **moderate resistance**.
  Examples: germanium, silicon.
  Semiconductors are used to make diodes, LEDs, transistors and integrated circuits (chips).
- **Insulators** which have **high resistance**.
  Examples: most plastics such as polythene and PVC (polyvinyl chloride), paper, glass.
  PVC is used as an outer covering for wires to prevent them making contact.

Ohm's Law

To make a current flow through a resistance there must be a voltage across that resistance. Ohm's Law shows the relationship between the voltage (V), current (I) and resistance (R). It can be written in three ways:

$$V = I \times R \quad \text{or} \quad I = \frac{V}{R} \quad \text{or} \quad R = \frac{V}{I}$$

where:
V = voltage in volts (V)
I = current in amps (A)
R = resistance in ohms (Ω)

or:
V = voltage in volts (V)
I = current in milliamps (mA)
R = resistance in kilohms (kΩ)

For most electronic circuits the amp is too large and the ohm is too small, so we often measure current in milliamps (mA) and resistance in kilohms (kΩ). 1 mA = 0.001 A and 1 kΩ = 1000 Ω.

The Ohm's Law equations work if you use V, A and Ω, or if you use V, mA and kΩ. You must not mix these sets of units in the equations so you may need to convert between mA and A or kΩ and Ω.

The VIR Triangle

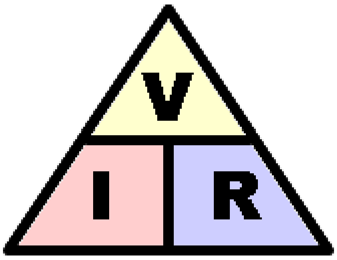

Ohm's Law Triangle

You can use the VIR triangle to help you remember the three versions of Ohm's Law.
Write down V, I and R in a triangle like the one in the yellow box on the right.

- To calculate **voltage, V**: put your finger over V,
  this leaves you with I R, so the equation is V = I × R
- To calculate **current, I**: put your finger over I,
  this leaves you with V over R, so the equation is I = $^{V}/_{R}$
- To calculate **resistance, R**: put your finger over R,
  this leaves you with V over I, so the equation is R = $^{V}/_{I}$

Ohm's Law Calculations

Use this method to guide you through calculations:

1. Write down the **Values**, converting units if necessary.
2. Select the **Equation** you need (use the VIR triangle).
3. Put the **Numbers** into the equation and calculate the answer.

It should be **V**ery **E**asy **N**ow!

- 3 V is applied across a 6 Ωresistor, what is the current?
    - **V**alues: V = 3 V, I = ?, R = 6 Ω
    - **E**quation: I = $^{V}/_{R}$
    - **N**umbers: Current, I = $^{3}/_{6}$ = 0.5 A
- A lamp connected to a 6 V battery passes a current of 60 mA, what is the lamp's resistance?
    - **V**alues: V = 6 V, I = 60 mA, R = ?
    - **E**quation: R = $^{V}/_{I}$
    - **N**umbers: Resistance, R = $^{6}/_{60}$ = 0.1 kΩ = 100 Ω
      (using mA for current means the calculation gives the resistance in kΩ)
- A 1.2 kΩ resistor passes a current of 0.2 A, what is the voltage across it?
    - **V**alues: V = ?, I = 0.2 A, R = 1.2 kΩ = 1200 Ω
      (1.2 kΩ is converted to 1200 Ωbecause A and kΩ must not be used together)
    - **E**quation: V = I × R
    - **N**umbers: V = 0.2 × 1200 = 240 V

Power and Energy

What is power?

Power is the rate of using or supplying energy:

$$\textbf{Power} = \frac{\textbf{Energy}}{\textbf{Time}}$$

Power is measured in watts (**W**)
Energy is measured in joules (**J**)
Time is measured in seconds (**s**)

Electronics is mostly concerned with small quantities of power, so the power is often measured in milliwatts (mW), 1mW = 0.001W. For example an LED uses about 40mW and a bleeper uses about 100mW, even a lamp such as a torch bulb only uses about 1W.

The typical power used in mains electrical circuits is much larger, so this power may be measured in kilowatts (kW), 1kW = 1000W. For example a typical mains lamp uses 60W and a kettle uses about 3kW.

Calculating power using current and voltage

There are three ways of writing an equation for power, current and voltage:

Power = Current × Voltage so **$P = I \times V$** or **$I = \frac{P}{V}$** or **$V = \frac{P}{I}$**

where:
P = power in watts (W)
V = voltage in volts (V)
I = current in amps (A)

or:
P = power in milliwatts (mW)
V = voltage in volts (V)
I = current in milliamps (mA)

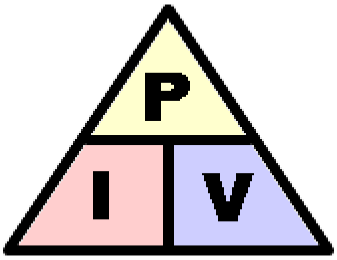

You can use the **PIV triangle** to help you remember the three versions of the power equations. Use it in the same way as the Ohm's Law triangle. For most electronic circuits the amp is too large, so we often measure current in milliamps (mA) and power in milliwatts (mW). 1mA = 0.001A and 1mW = 0.001W.

Calculating power using resistance and current or voltage

Using Ohm's Law $V = I \times R$ we can convert $P = I \times V$ to:

$P = I^2 \times R$ where:
or
$P = V^2 / R$

P = power in watts (W)
I = current in amps (A)
R = resistance in ohms (Ω)
V = voltage in volts (V)

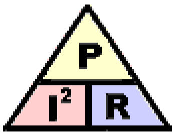

PI²R triangle

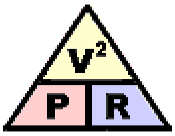

V²PR triangle

Wasted power and overheating

Normally electric power is useful, making a lamp light or a motor turn for example. However, electrical energy is converted to heat whenever a current flow through a resistance and this can be problems if it makes a device or wire overheat. In electronics the effect is usually negligible, but if the resistance is low (a wire or low value resistor for example) the current can be sufficiently large to cause a problem.

You can see from the equation $P = I^2 \times R$ that for a given resistance the power depends on the current **squared**, so doubling the current will give 4 times the power.

Resistors are rated by the maximum power they can have developed in them without damage, but power ratings are rarely quoted in parts lists because the standard ratings of 0.25W or 0.5W are suitable for most circuits. Further information is available on the Resistors page.

Wires and cables are rated by the maximum current they can pass without overheating. They have a very low resistance so the maximum current is relatively large. For further information about current rating please see the Connectors and Cables page.

Energy

The amount of energy used (or supplied) depends on the power and the time for which it is used:

Energy = Power × Time

A low power device operating for a long time can use more energy than a high power device operating for a short time. For example:

- A 60W lamp switched on for 8 hours uses 60W × 8 × 3600s = 1728kJ.
- A 3kW kettle switched on for 5 minutes uses 3000W × 5 × 60s = 900kJ.

The standard unit for energy is the joule (J), but 1J is a very small amount of energy for mains electricity so kilojoule (kJ) or megajoule (MJ) are sometimes used in scientific work. In the home we measure electrical energy in kilowatt-hours (kWh). 1kWh is the energy used by a 1kW power appliance when it is switched on for 1 hour:

1kWh = 1kW × 1 hour = 1000W × 3600s = 3.6MJ

For example:

- A 60W lamp switched on for 8 hours uses 0.06kW × 8 = 0.48kWh.
- A 3kW kettle switched on for 5 minutes uses 3kW × $^5/_{60}$ = 0.25kWh.

AC, DC and Electrical Signals

AC means Alternating Current and DC means Direct Current. AC and DC are also used when referring to voltages and electrical signals which are not currents! For example: a 12V AC power supply has an alternating voltage (which will make an alternating current flow). An electrical signal is a voltage or current which conveys information, usually it means a voltage. The term can be used for any voltage or current in a circuit.

Alternating Current (AC)

Alternating Current (AC) flows one way, then the other way, continually reversing direction.

An AC voltage is continually changing between positive (+) and negative (-).

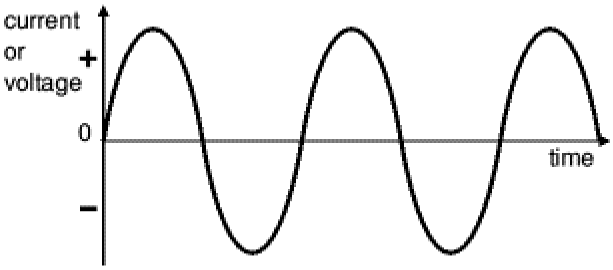

AC from a power supply
This shape is called a sine wave

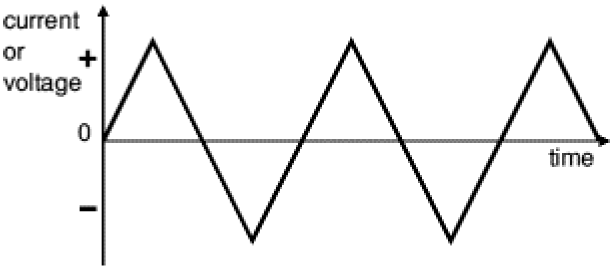

This triangular signal is AC
because it changes between
positive (+) and negative (-)

The rate of changing direction is called the **frequency** of the AC and it is measured in **hertz (Hz)** which is the number of forwards-backwards **cycles per second**.

An AC supply is suitable for powering some devices such as lamps and heaters but almost all electronic circuits require a steady DC supply (see below).

Direct Current (DC)

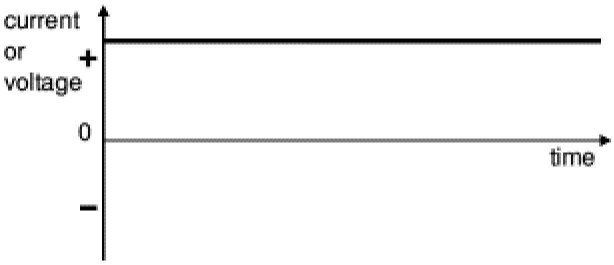

Steady DC
from a battery or regulated power supply,
this is ideal for electronic circuits

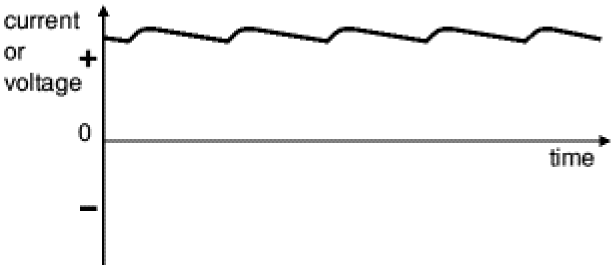

Smooth DC
from a smoothed power supply,
this is suitable for some electronics

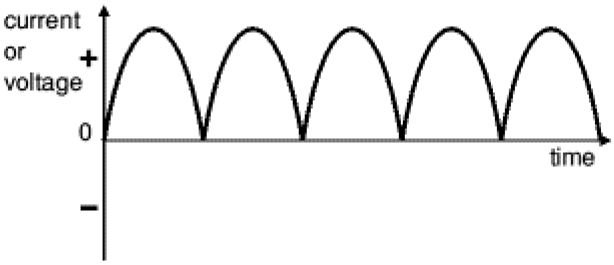

Varying DC
from a power supply without smoothing,
this is not suitable for electronics

Direct Current (DC) always flows in the same direction, but it may increase and decrease.

A DC voltage is always positive (or always negative), but it may increase and decrease.

Electronic circuits normally require a **steady DC** supply which is constant at one value or a **smooth DC** supply which has a small variation called **ripple**.

Cells, batteries and regulated power supplies provide **steady DC** which is ideal for electronic circuits.

Power supplies contain a transformer which converts the mains AC supply to a safe low voltage AC. Then the AC is converted to DC by a bridge rectifier but the output is **varying DC** which is unsuitable for electronic circuits.

Some power supplies include a capacitor to provide **smooth DC** which is suitable for less-sensitive electronic circuits, including most of the projects on this website. Lamps, heaters and motors will work with any DC supply.

Properties of electrical signals

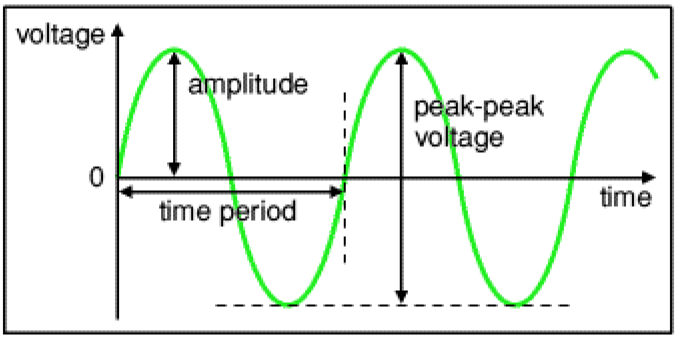

An electrical signal is a voltage or current which conveys information, usually it means a voltage. The term can be used for any voltage or current in a circuit.

The voltage-time graph on the right shows various properties of an electrical signal. In addition to the properties labeled on the graph, there is frequency which is the number of cycles per second.

The diagram shows a **sine wave** but these properties apply to any signal with a constant shape.

- **Amplitude** is the maximum voltage reached by the signal.
  It is measured in **volts**, **V**.
- **Peak voltage** is another name for amplitude.
- **Peak-peak voltage** is twice the peak voltage (amplitude). When reading an oscilloscope trace it is usual to measure peak-peak voltage.
- **Time period** is the time taken for the signal to complete one cycle.
  It is measured in **seconds (s)**, but time periods tend to be short so **milliseconds (ms)** and **microseconds (µs)** are often used. 1ms = 0.001s and 1µs = 0.000001s.
- **Frequency** is the number of cycles per second.
  It is measured in **hertz (Hz)**, but frequencies tend to be high so **kilohertz (kHz)** and **megahertz (MHz)** are often used. 1kHz = 1000Hz and 1MHz = 1000000Hz.

$$\textbf{frequency} = \frac{\textbf{1}}{\textbf{time period}} \quad \text{and} \quad \textbf{time period} = \frac{\textbf{1}}{\textbf{frequency}}$$

- Mains electricity in the UK has a frequency of **50Hz**, so it has a time period of $^1/_{50}$ = 0.02s = **20ms**.

Root Mean Square (RMS) Values

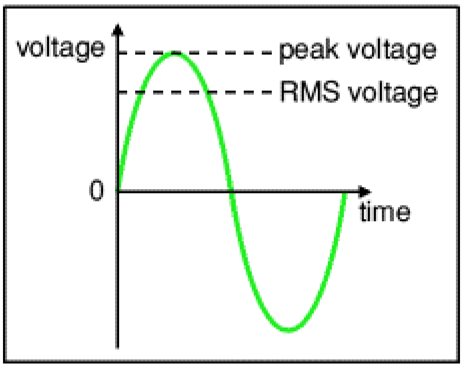

The value of an AC voltage is continually changing from zero up to the positive peak, through zero to the negative peak and back to zero again. Clearly for most of the time it is less than the peak voltage, so this is not a good measure of its real effect.

Instead we use the **root mean square voltage** (V_{RMS}) which is 0.7 of the **peak voltage** (V_{peak}):

$$V_{RMS} = 0.7 \times V_{peak} \quad \text{and} \quad V_{peak} = 1.4 \times V_{RMS}$$

These equations also apply to **current**.
They are only true for sine waves (the most common type of AC) because the 0.7 and 1.4 are different values for other shapes.

The **RMS value** is the **effective value** of a varying voltage or current. It is the equivalent steady DC (constant) value which gives the same effect.

For example a lamp connected to a **6V RMS AC** supply will light with the same brightness when connected to a **steady 6V DC** supply. However, the lamp will be dimmer if connected to a **6V peak AC** supply because the RMS value of this is only 4.2V (it is equivalent to a steady 4.2V DC).

You may find it helps to think of the RMS value as a sort of average, but please remember that it is NOT really the average! In fact the average voltage (or current) of an AC signal is zero because the positive and negative parts exactly cancel out!

What do AC meters show, is it the RMS or peak voltage?

AC voltmeters and ammeters show the **RMS value** of the voltage or current. DC meters also show the RMS value when connected to varying DC providing the DC is varying quickly, if the frequency is less than about 10Hz you will see the meter reading fluctuating instead.

What does '6V AC' really mean, is it the RMS or peak voltage?

If the peak value is meant it should be clearly stated, otherwise assume it is the **RMS value**. In everyday use AC voltages (and currents) are always given as **RMS values** because this allows a sensible comparison to be made with steady DC voltages (and currents), such as from a battery.

For example a '6V AC supply' means 6V RMS, the peak voltage is 8.6V. The UK mains supply is 230V AC, this means 230V RMS so the peak voltage of the mains is about 320V!

So what does root mean square (RMS) really mean?

First square all the values, then find the average (mean) of these square values over a complete cycle, and find the square root of this average. That is the RMS value. Confused? Ignore the math (it looks more complicated than it really is), just accept that RMS values for voltage and current are a much more useful quantity than peak values.

Oscilloscopes (CRO)

An oscilloscope is a test instrument which allows you to look at the 'shape' of electrical signals by displaying a graph of voltage against time on its screen. It is like a voltmeter with the valuable extra function of showing how the voltage varies with time. A graticule with a 1cm grid enables you to take measurements of voltage and time from the screen.

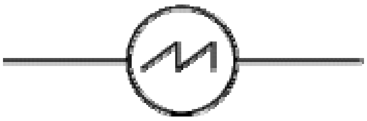

Circuit symbol for an oscilloscope

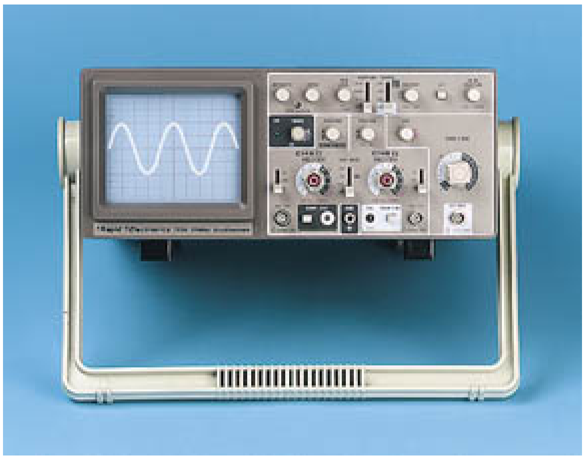

Cathode Ray Oscilloscope (CRO)

The graph, usually called the **trace**, is drawn by a beam of electrons striking the phosphor coating of the screen making it emit light, usually green or blue. This is similar to the way a television picture is produced.

Oscilloscopes contain a vacuum tube with a **cathode** (negative electrode) at one end to emit electrons and an **anode** (positive electrode) to accelerate them so they move rapidly down the tube to the screen. This arrangement is called an electron gun. The tube also contains electrodes to deflect the electron beam up/down and left/right.

The electrons are called cathode rays because they are emitted by the cathode and this gives the oscilloscope its full name of **cathode ray oscilloscope** or CRO.

A **dual trace** oscilloscope can display two traces on the screen, allowing you to easily compare the input and output of an amplifier for example. It is well worth paying the modest extra cost to have this facility.

Precautions

- An oscilloscope should be handled gently to protect its fragile (and expensive) vacuum tube.
- Oscilloscopes use high voltages to create the electron beam and these remain for some time after switching off - for your own safety do not attempt to examine the inside of an oscilloscope!

Setting up an oscilloscope

Oscilloscopes are complex instruments with many controls and they require some care to set up and use successfully. It is quite easy to 'lose' the trace off the screen if controls are set wrongly!

There is some variation in the arrangement and labeling of the many controls so the following instructions may need to be adapted for your instrument.

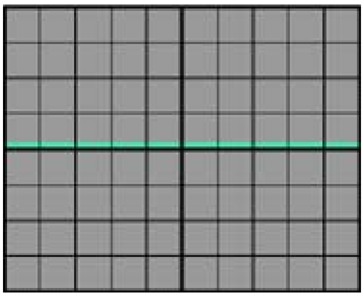
This is what you should see after setting up, when there is no input signal connected

- **Switch on** the oscilloscope to warm up (it takes a minute or two).
- Do **not** connect the input lead at this stage.
- Set the **AC/GND/DC** switch (by the Y INPUT) to **DC**.
- Set the **SWP/X-Y** switch to **SWP** (sweep).
- Set **Trigger Level** to **AUTO**.
- Set **Trigger Source** to **INT** (internal, the y input).
- Set the **Y AMPLIFIER** to **5V/cm** (a moderate value).
- Set the **TIMEBASE** to **10ms/cm** (a moderate speed).
- Turn the timebase **VARIABLE** control to **1** or **CAL**.
- Adjust **Y SHIFT** (up/down) and **X SHIFT** (left/right) to give a **trace across the middle of the screen**, like the picture.
- Adjust **INTENSITY** (brightness) and **FOCUS** to give a bright, sharp trace.
- The oscilloscope is now ready to use!
  Connecting the input lead is described in the next section.

Further information on the controls: Timebase | Y amplifier | AC/GND/DC switch

---

Connecting an Oscilloscope

Oscilloscope lead and probes kit

The Y INPUT lead to an oscilloscope should be a **co-axial lead** and the diagram shows its construction. The central wire carries the signal and the screen is connected to earth (0V) to shield the signal from electrical interference (usually called noise).

Most oscilloscopes have a **BNC socket** for the y input and the lead is connected with a **push and twist action**, to disconnect you need to twist and pull. Oscilloscopes used in schools may have red and black 4mm sockets so that ordinary, unscreened, 4mm plug leads can be used if necessary.

Professionals use a specially designed lead and probes kit for best results with high frequency signals and when testing high resistance circuits, but this is not essential for simpler work at audio frequencies (up to 20kHz).

An oscilloscope is connected like a voltmeter but you must be aware that the screen (black) connection of the input lead is connected to mains earth at the oscilloscope! This means it **must** be connected to earth or 0V on the circuit being tested.

Obtaining a clear and stable trace

Once you have connected the oscilloscope to the circuit you wish to test you will need to **adjust the controls** to obtain a clear and stable trace on the screen:

- The **Y AMPLIFIER (VOLTS/CM)** control determines the height of the trace. Choose a setting so the trace occupies at least half the screen height, but does not disappear off the screen.
- The **TIMEBASE (TIME/CM)** control determines the rate at which the dot sweeps across the screen. Choose a setting so the trace shows at least one cycle of the signal across the screen.
  Note that a steady DC input signal gives a horizontal line trace for which the timebase setting is not critical.
- The **TRIGGER** control is usually best left set to AUTO.

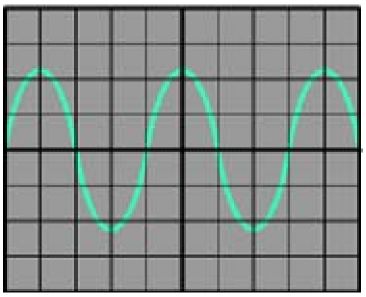

The trace of an AC signal with the oscilloscope controls correctly set

If you are using an oscilloscope for the first time it is best to start with an easy signal such as the output from an AC power pack set to about 4V.

Further information on the controls: Timebase | Y amplifier | AC/GND/DC switch

---

Measuring voltage and time period

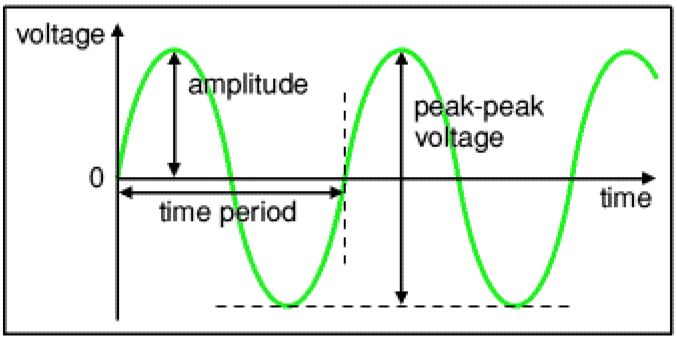

The trace on an oscilloscope screen is a **graph of voltage against time**. The shape of this graph is determined by the nature of the input signal.

In addition to the properties labeled on the graph, there is frequency which is the number of cycles per second.

The diagram shows a **sine wave** but these properties apply to any signal with a constant shape.

- **Amplitude** is the maximum voltage reached by the signal.
  It is measured in **volts**, **V**.
- **Peak voltage** is another name for amplitude.

- **Peak-peak voltage** is twice the peak voltage (amplitude). When reading an oscilloscope trace it is usual to measure peak-peak voltage.
- **Time period** is the time taken for the signal to complete one cycle.
  It is measured in **seconds (s)**, but time periods tend to be short so **milliseconds (ms)** and **microseconds (µs)** are often used. 1ms = 0.001s and 1µs = 0.000001s.
- **Frequency** is the number of cycles per second.
  It is measured in **hertz (Hz)**, but frequencies tend to be high so **kilohertz (kHz)** and **megahertz (MHz)** are often used. 1kHz = 1000Hz and 1MHz = 1000000Hz.

$$\textbf{frequency} = \frac{\textbf{1}}{\textbf{time period}} \quad \text{and} \quad \textbf{time period} = \frac{\textbf{1}}{\textbf{frequency}}$$

Voltage

Voltage is shown on the **vertical y-axis** and the scale is determined by the Y AMPLIFIER (VOLTS/CM) control. Usually **peak-peak voltage** is measured because it can be read correctly even if the position of 0V is not known. The **amplitude** is half the peak-peak voltage.

If you wish to read the amplitude voltage directly you must check the position of 0V (normally halfway up the screen): move the AC/GND/DC switch to GND (0V) and use Y-SHIFT (up/down) to adjust the position of the trace if necessary, switch back to DC afterwards so you can see the signal again.

Voltage = distance in cm × volts/cm
Example: peak-peak voltage = 4.2cm × 2V/cm = 8.4V
amplitude (peak voltage) = ½ × peak-peak voltage = 4.2V

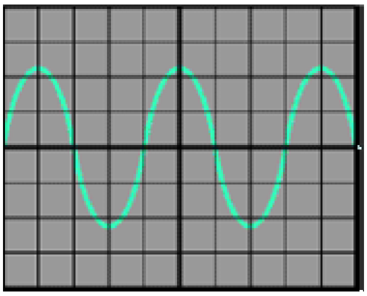

The trace of an AC signal

Y AMPLIFIER: 2V/cm
TIMEBASE: 5ms/cm

Example measurements:

peak-peak voltage = 8.4V
amplitude voltage = 4.2V

time period = 20ms
frequency = 50Hz

Time Period

Time is shown on the **horizontal x-axis** and the scale is determined by the TIMEBASE (TIME/CM) control. The **time period** (often just called **period**) is the time for one cycle of the signal. The **frequency** is the number of cycles per second, frequency = 1/time period

Ensure that the variable timebase control is set to 1 or CAL (calibrated) before attempting to take a time reading.

Time = distance in cm × time/cm
Example: time period = 4.0cm × 5ms/cm = 20ms
and frequency = $^{1}/_{\text{time period}} = {}^{1}/_{20\text{ms}}$ *= 50Hz*

Timebase (time/cm) and trigger controls

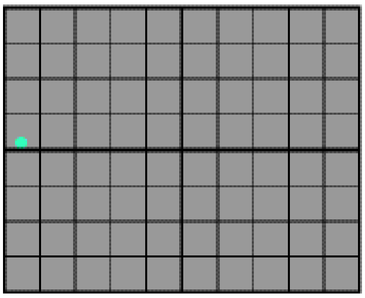

Slow timebase, no input
You can see the dot moving

The oscilloscope sweeps the electron beam across the screen from left to right at a steady speed set by the TIMEBASE control. Each setting is labeled with the time the dot takes to move 1cm, effectively it is setting the scale on the x-axis. The timebase control may be labeled TIME/CM.

At slow timebase settings (such as 50ms/cm) you can see a dot moving across the screen but at faster settings (such as 1ms/cm) the dot is moving so fast that it appears to be a line.

The VARIABLE timebase control can be turned to make a fine adjustment to the speed, but it must be left at the position labeled 1 or CAL (calibrated) if you wish to take time readings from the trace drawn on the screen.

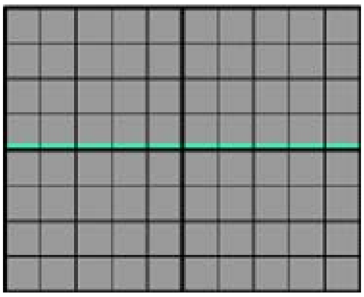

Fast timebase, no input
The dot is too fast to see
so it appears to be a line

The TRIGGER controls are used to maintain a steady trace on the screen. If they are set wrongly you may see a trace drifting sideways, a confusing 'scribble' on the screen, or no trace at all! The trigger maintains a steady trace by starting the dot sweeping across the screen when the input signal reaches the same point in its cycle each time.

For straightforward use it is best to leave the trigger level set to AUTO, but if you have difficulty obtaining a steady trace try adjusting this control to set the level manually.

Y amplifier (volts/cm) control

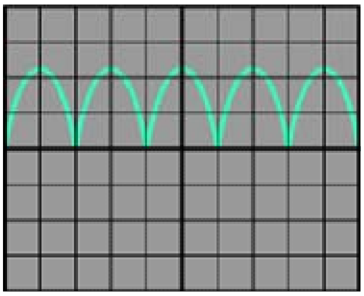

Varying DC (always positive)

The oscilloscope moves the trace up and down in proportion to the voltage at the Y INPUT and the setting of the Y AMPLIFIER control. This control sets the voltage represented by each centimeter (cm) on the screen; effectively it is setting the scale on the y-axis. Positive voltages make the trace move up, negative voltages make it move down.

The y amplifier control may be labeled Y-GAIN or VOLTS/CM.

The input voltage moving the dot up and down at the same time as the dot is swept across the screen means that the trace on the screen is a graph of **voltage (y-axis)** against **time (x-axis)** for the input signal.

The AC/GND/DC switch

The normal setting for this switch is **DC** for all signals, including AC!

Switching to **GND** (ground) connects the y input to 0V and allows you to quickly check the position of 0V on the screen (normally halfway up). There is no need to disconnect the input lead while you do this because it is disconnected internally.

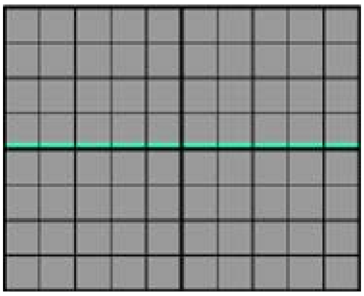

Switching to GND allows you to quickly check the position of 0V (normally halfway up).

Switching to **AC** inserts a capacitor in series with the input to block out any DC signal present and pass **only AC** signals. This is used to examine signals showing a small variation around one constant value, such as the ripple on the output of a smooth DC supply. Reducing the VOLTS/CM to see more detail of the ripple would normally take the trace off the screen! The **AC** setting removes the constant (DC) part of the signal, allowing you to view just the varying (AC) part which can now be examined more closely by reducing the VOLTS/CM. This is shown in the diagrams below:

Displaying a ripple signal using the AC switch

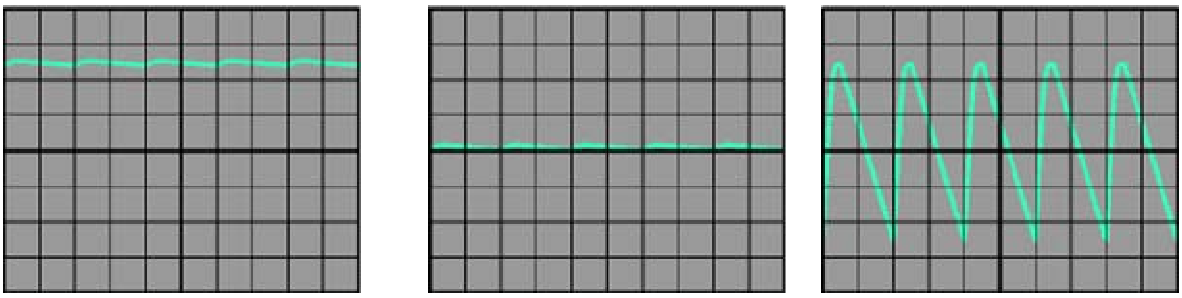

Switch in normal DC position. The ripple is difficult to see, but if VOLTS/CM is reduced to enlarge it the trace will disappear off the screen!

Switch moved to AC position. The constant (DC) part of the signal is removed, leaving just the ripple (AC) part.

VOLTS/CM reduced to enlarge the ripple. The ripple can now be examined more closely.

Power Supplies

There are many types of power supply. Most are designed to convert high voltage AC mains electricity to a suitable low voltage supply for electronics circuits and other devices. A power supply can by broken down into a series of blocks, each of which performs a particular function.

For example a 5V regulated supply:

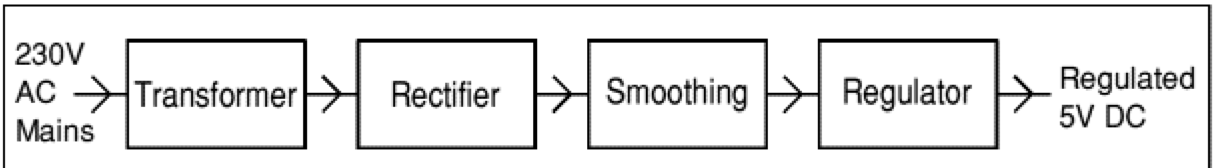

Block Diagram of a Regulated Power Supply System

Each of the blocks is described in more detail below:

- Transformer - steps down high voltage AC mains to low voltage AC.
- Rectifier - converts AC to DC, but the DC output is varying.
- Smoothing - smooth the DC from varying greatly to a small ripple.
- Regulator - eliminates ripple by setting DC output to a fixed voltage.

Power supplies made from these blocks are described below with a circuit diagram and a graph of their output:

- Transformer only
- Transformer + Rectifier
- Transformer + Rectifier + Smoothing
- Transformer + Rectifier + Smoothing + Regulator

Dual Supplies

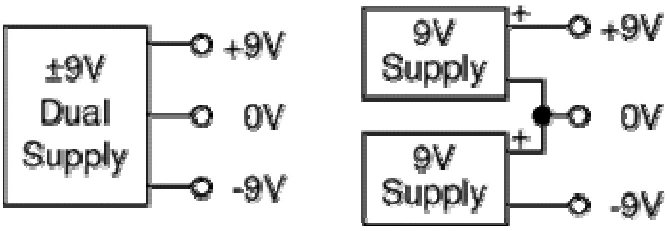

Some electronic circuits require a power supply with positive and negative outputs as well as zero volts (0V). This is called a 'dual supply' because it is like two ordinary supplies connected together as shown in the diagram.

Dual supplies have three outputs, for example a ±9V supply has +9V, 0V and -9V outputs.

Transformer only

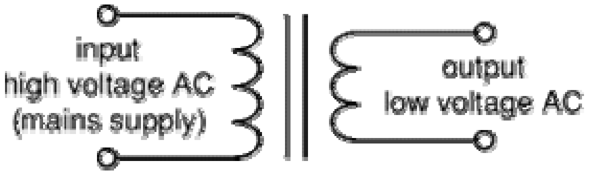

Transformer

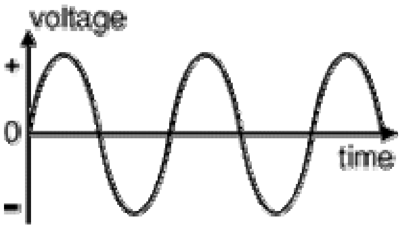

Output: low voltage AC

The **low voltage AC** output is suitable for lamps, heaters and special AC motors. It is **not** suitable for electronic circuits unless they include a rectifier and a smoothing capacitor.

Transformer + Rectifier

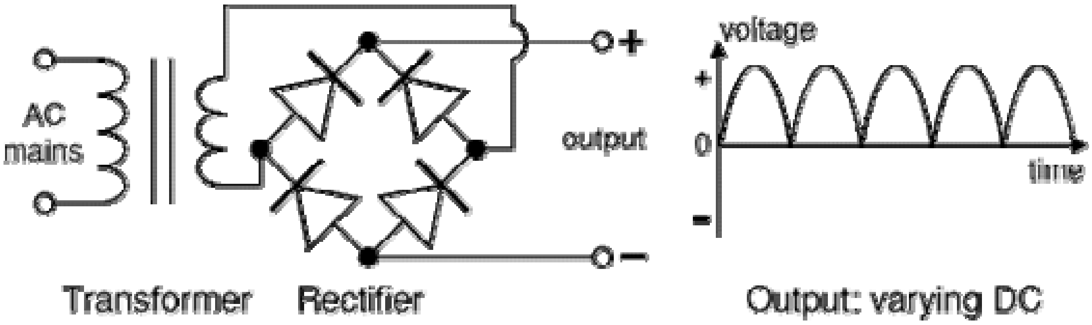
AC
mains
+
output
−
voltage
+
0
time
-
Transformer
Rectifier
Output: varying DC

The **varying DC** output is suitable for lamps, heaters and standard motors. It is **not** suitable for electronic circuits unless they include a smoothing capacitor.

Transformer + Rectifier + Smoothing

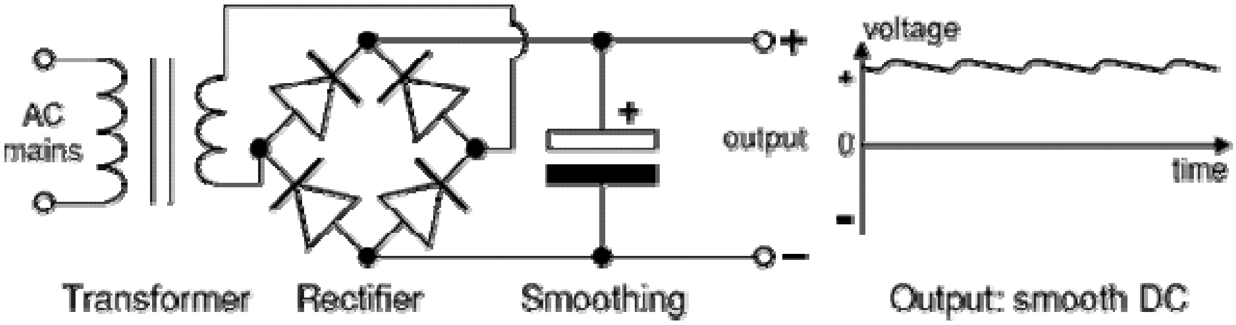
AC
mains
+
output
−
voltage
+
0
time
-
Transformer
Rectifier
Smoothing
Output: smooth DC

The **smooth DC** output has a small ripple. It is suitable for most electronic circuits.

Transformer + Rectifier + Smoothing + Regulator

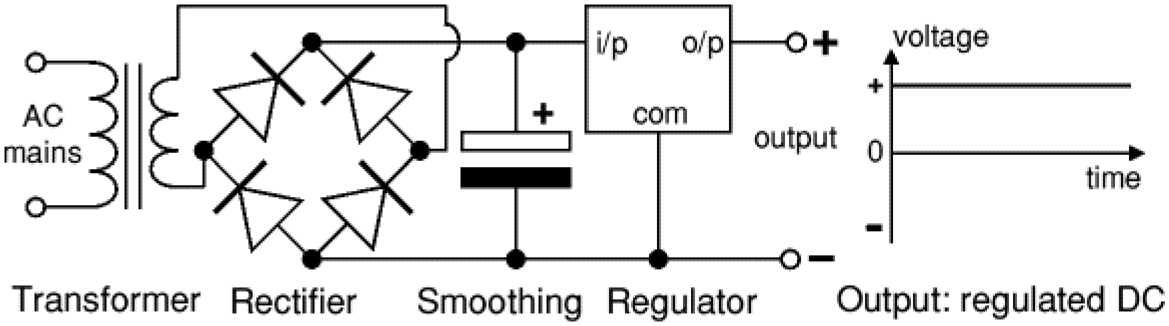
AC
mains
+
i/p
o/p
com
+
output
−
voltage
+
0
time
-
Transformer
Rectifier
Smoothing
Regulator
Output: regulated DC

The **regulated DC** output is very smooth with no ripple. It is suitable for all electronic circuits.

Transformer

Transformers convert AC electricity from one voltage to another with little loss of power. Transformers work only with AC and this is one of the reasons why mains electricity is AC.

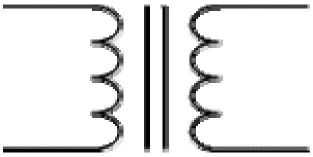

Transformer circuit symbol

Step-up transformers increase voltage, step-down transformers reduce voltage. Most power supplies use a step-down transformer to reduce the dangerously high mains voltage (230V in UK) to a safer low voltage.

The input coil is called the **primary** and the output coil is called the **secondary**. There is no electrical connection between the two coils, instead they are linked by an alternating magnetic field created in the soft-iron core of the transformer. The two lines in the middle of the circuit symbol represent the core.

Transformer

Transformers waste very little power so the power out is (almost) equal to the power in. Note that as voltage is stepped down current is stepped up.

The ratio of the number of turns on each coil, called the **turns ratio**, determines the ratio of the voltages. A step-down transformer has a large number of turns on its primary (input) coil which is connected to the high voltage mains supply, and a small number of turns on its secondary (output) coil to give a low output voltage.

$$\text{turns ratio} = \frac{V_p}{V_s} = \frac{N_p}{N_s} \quad \text{and} \quad \text{power out} = \text{power in} \quad V_s \times I_s = V_p \times I_p$$

Vp = primary (input) voltage
Np = number of turns on primary coil
Ip = primary (input) current

Vs = secondary (output) voltage
Ns = number of turns on secondary coil
Is = secondary (output) current

Rectifier

There are several ways of connecting diodes to make a rectifier to convert AC to DC. The bridge rectifier is the most important and it produces **full-wave** varying DC. A full-wave rectifier can also be made from just two diodes if a centre-tap transformer is used, but this method is rarely used now that diodes are cheaper. A single diode can be used as a rectifier but it only uses the positive (+) parts of the AC wave to produce **half-wave** varying DC.

Bridge rectifier

A bridge rectifier can be made using four individual diodes, but it is also available in special packages containing the four diodes required. It is called a full-wave rectifier because it uses all the AC wave (both positive and negative sections). 1.4V is used up in the bridge rectifier because each diode uses 0.7V when conducting and there are always two diodes conducting, as shown in the diagram below. Bridge rectifiers are rated by the maximum current they can pass and the maximum reverse voltage they can withstand (this must be at least three times the supply RMS voltage so the rectifier can withstand the peak voltages). Please see the Diodes page for more details, including pictures of bridge rectifiers.

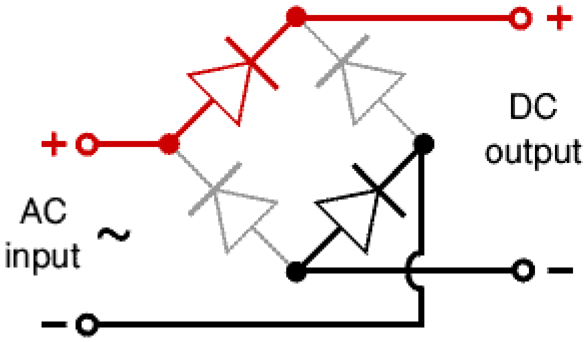

Bridge rectifier
Alternate pairs of diodes conduct, changing over the connections so the alternating directions of AC are converted to the one direction of DC.

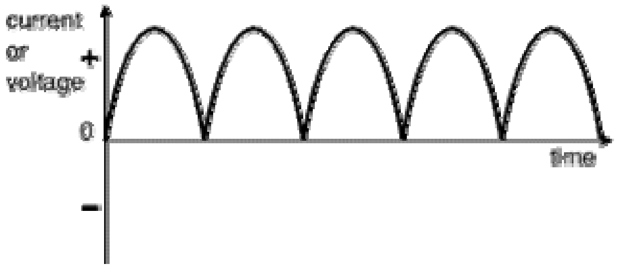

Output: full-wave varying DC
(using all the AC wave)

Single diode rectifier

A single diode can be used as a rectifier but this produces **half-wave** varying DC which has gaps when the AC is negative. It is hard to smooth this sufficiently well to supply electronic circuits unless they require a very small current so the smoothing capacitor does not significantly discharge during the gaps. Please see the Diodes page for some examples of rectifier diodes.

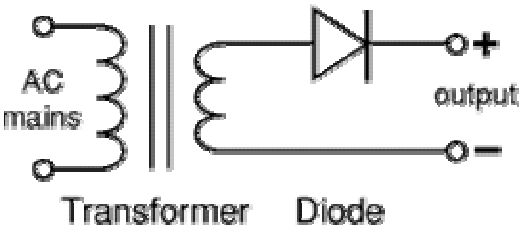

Single diode rectifier

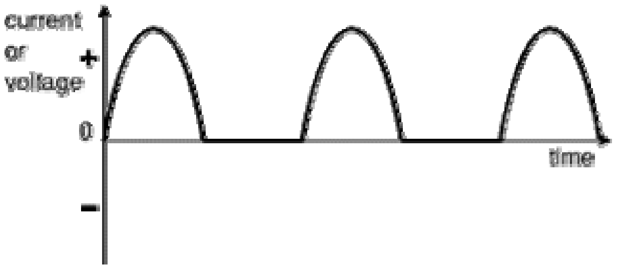

Output: half-wave varying DC
(using only half the AC wave)

Smoothing

Smoothing is performed by a large value electrolytic capacitor connected across the DC supply to act as a reservoir, supplying current to the output when the varying DC voltage from the rectifier is falling. The diagram shows the unsmoothed varying DC (dotted line) and the smoothed DC (solid line). The capacitor charges quickly near the peak of the varying DC, and then discharges as it supplies current to the output.

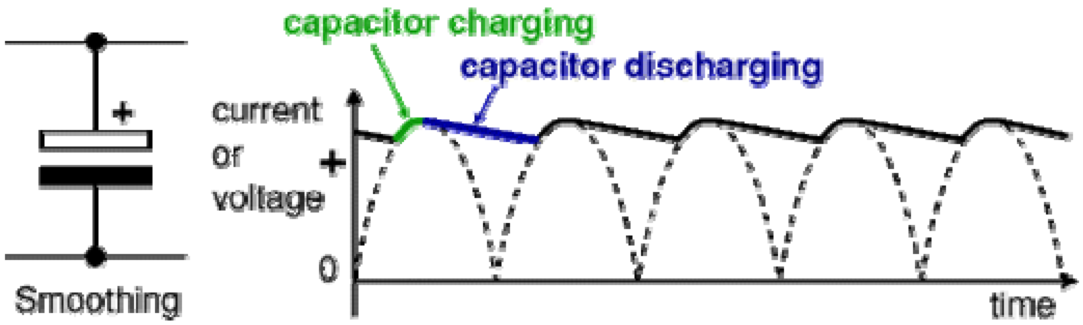

Note that smoothing significantly increases the average DC voltage to almost the peak value (1.4 × RMS value). For example 6V RMS AC is rectified to full wave DC of about 4.6V RMS (1.4V is lost in the bridge rectifier), with smoothing this increases to almost the peak value giving 1.4 × 4.6 = 6.4V smooth DC.

Smoothing is not perfect due to the capacitor voltage falling a little as it discharges, giving a small **ripple voltage**. For many circuits a ripple which is 10% of the supply voltage is satisfactory and the equation below gives the required value for the smoothing capacitor. A larger capacitor will give less ripple. The capacitor value must be doubled when smoothing half-wave DC.

Smoothing capacitor for 10% ripple, $\mathbf{C} = \dfrac{\mathbf{5 \times Io}}{\mathbf{Vs \times f}}$

C = smoothing capacitance in farads (F)
Io = output current from the supply in amps (A)
Vs = supply voltage in volts (V), this is the peak value of the unsmoothed DC
f = frequency of the AC supply in hertz (Hz)

Regulator

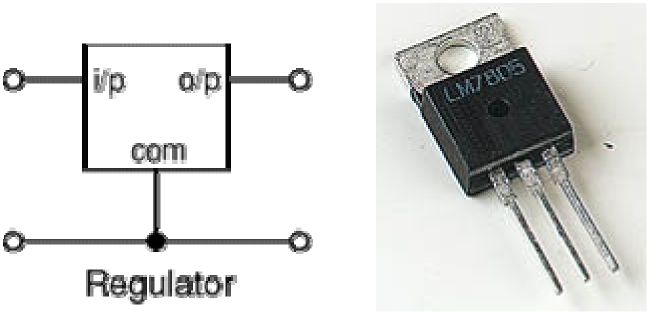

Voltage regulator

Voltage regulator ICs are available with fixed (typically 5, 12 and 15V) or variable output voltages. They are also rated by the maximum current they can pass. Negative voltage regulators are available, mainly for use in dual supplies. Most regulators include some automatic protection from excessive current ('overload protection') and overheating ('thermal protection').

Many of the fixed voltage regulator ICs have 3 leads and look like power transistors, such as the 7805 +5V 1A regulator shown on the right. They include a hole for attaching a heatsink if necessary.

Zener Diode Regulator

Zener diode
a = anode, k = cathode

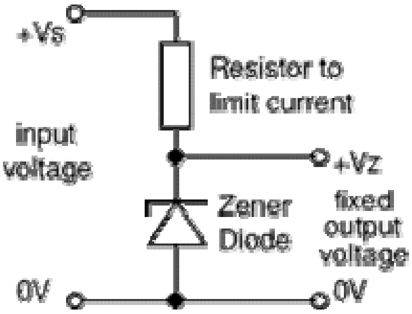

For low current power supplies a simple voltage regulator can be made with a resistor and a zener diode connected **in reverse** as shown in the diagram. Zener diodes are rated by their **breakdown voltage Vz** and **maximum power Pz** (typically 400mW or 1.3W).

The resistor limits the current (like an LED resistor). The current through the resistor is constant, so when there is no output current all the current flows through the zener diode and its power rating Pz must be large enough to withstand this.

Choosing a zener diode and resistor:

1. The **zener voltage Vz** is the output voltage required
2. The **input voltage Vs** must be a few volts greater than Vz
   (this is to allow for small fluctuations in Vs due to ripple)
3. The **maximum current Imax** is the output current required plus 10%
4. The **zener power Pz** is determined by the maximum current: Pz > Vz × Imax
5. The **resistor resistance**: R = (Vs - Vz) / Imax
6. The **resistor power rating**: P > (Vs - Vz) × Imax

Example: output voltage required is 5V, output current required is 60mA.

1. **Vz = 4.7V** (nearest value available)
2. **Vs = 8V** (it must be a few volts greater than Vz)
3. **Imax = 66mA** (output current plus 10%)
4. Pz > 4.7V × 66mA = 310mW, choose **Pz = 400mW**
5. R = (8V - 4.7V) / 66mA = 0.05kΩ = 50Ω, choose **R = 47Ω**
6. Resistor power rating P > (8V - 4.7V) × 66mA = 218mW, choose **P = 0.5W**

Transducers

A transducer is a device which converts a signal from one form to another.

Most electronics circuits use both input and output transducers:

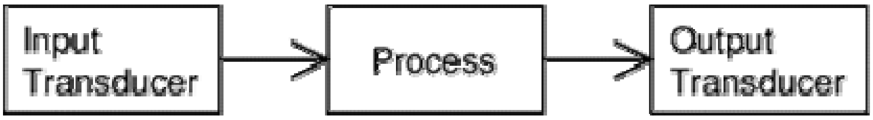

Input Transducers

Input Transducers convert a quantity to an electrical signal (voltage) or to resistance (which can be converted to voltage). Input transducers are also called **sensors**.

Examples:

- LDR converts brightness (of light) to resistance.
- Thermistor converts temperature to resistance.
- Microphone converts sound to voltage.
- Variable resistor converts position (angle) to resistance.

LDR

Output Transducers

Output Transducers convert an electrical signal to another quantity.

Examples:

- Lamp converts electricity to light.
- LED converts electricity to light.
- Loudspeaker converts electricity to sound.
- Motor converts electricity to motion.
- Heater converts electricity to heat.

Loudspeaker

Using input transducers (sensors)

Most input transducers (sensors) vary their **resistance** and this can be used directly in some circuits but it is usually converted to an electrical signal in the form of a **voltage**.

The voltage signal can be fed to other parts of the circuit, such as the input to a chip or a transistor switch.

The conversion of varying resistance to varying voltage is performed by a simple circuit called a **voltage divider**.

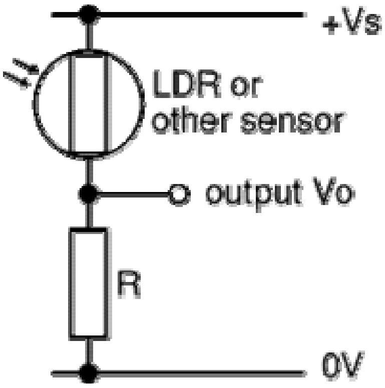

Voltage divider circuit

Voltage Dividers

Voltage divider (also called potential divider)

A voltage divider consists of two resistances R1 and R2 connected in series across a supply voltage Vs. The supply voltage is divided up between the two resistances to give an output voltage Vo which is the voltage across R2. This depends on the size of R2 relative to R1:

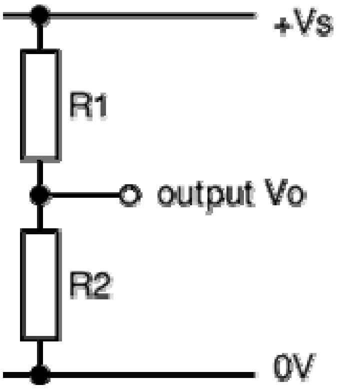

- **If R2 is much smaller** than R1, **Vo is small** (low, almost 0V) (because most of the voltage is across R1)
- **If R2 is about the same** as R1, **Vo is about half Vs** (because the voltage is shared about equally between R1 and R2)
- **If R2 is much larger** than R1, **Vo is large** (high, almost Vs) (because most of the voltage is across R2)

$$\mathbf{Vo} = \frac{\mathbf{Vs \times R2}}{\mathbf{R1 + R2}}$$

If you need a precise value for the output voltage Vo you can use Ohm's law and a little algebra to work out the formula for Vo shown on the right. The formula and the approximate rules given above assume that negligible current flows from the output. This is true if Vo is connected to a device with a high resistance such as voltmeter or a chip input. For further information please see the page on impedance. If the output is connected to a transistor Vo cannot become much greater than 0.7V because the transistor's base-emitter junction behaves like a diode.

Voltage dividers are also called **potential dividers**, a name which comes from potential difference (the proper name for voltage).

One of the main uses of voltage dividers is to connect input transducers into circuits...

Using an input transducer (sensor) in a voltage divider

Most input transducers (sensors) vary their **resistance** and usually a voltage divider is used to convert this to a varying **voltage** which is more useful. The voltage signal can be fed to other parts of the circuit, such as the input to a chip or a transistor switch.

The **sensor** is one of the resistances in the voltage divider. It can be at the top (R1) or at the bottom (R2), the choice is determined by when you want a large value for the output voltage Vo:

- Put the sensor at the **top** (R1) if you want a **large Vo** when the sensor has a **small resistance**.
- Put the sensor at the **bottom** (R2) if you want a **large Vo** when the sensor has a **large resistance**.

Then you need to choose a value for the resistor...

Choosing a resistor value

The value of the resistor R will determine the range of the output voltage Vo. For best results you need a large 'swing' (range) for Vo and this is achieved if the resistor is much larger than the sensor's minimum resistance Rmin, but much smaller than the sensor's maximum resistance Rmax.

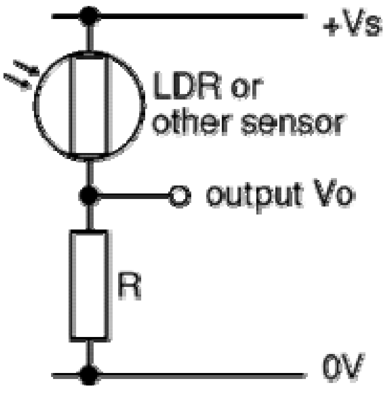

or

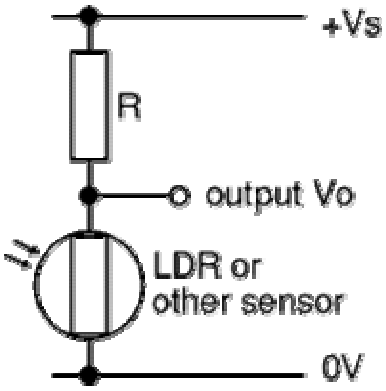

You can use a multimeter to help you find the minimum and maximum values of the sensor's resistance (Rmin and Rmax). There is no need to be precise, approximate values will do.

Then choose resistor value: **R = square root of (Rmin × Rmax)**
Choose a standard value which is close to this calculated value.

For example:
An LDR: Rmin = 100Ω, Rmax = 1MΩ,
so R = square root of (100 × 1M) = 10kΩ.

Swapping over the resistor and sensor

The resistor and sensor can be swapped over to invert the action of the voltage divider. For example an LDR has a high resistance when dark and a low resistance when brightly lit, so:

- **If the LDR is at the top (near +Vs)**,
  Vo will be **low in the dark** and **high in bright light**.
- **If the LDR is at the bottom (near 0V)**,
  Vo will be **high in the dark** and **low in bright light**.

Using a variable resistor

A variable resistor may be used in place of the fixed resistor R. It will enable you to adjust the output voltage Vo for a given resistance of the sensor. For example you can use a variable resistor to set the exact brightness level which makes a chip change state.

The variable resistor value should be larger than the fixed resistor value. For finer control you can use a fixed resistor in series with the variable resistor. For example if a 10kΩ fixed resistor is suitable you could replace it with a fixed 4.7kΩ resistor in series with a 10kΩ variable resistor, allowing you to adjust the resistance from 4.7k to 14.7kΩ.

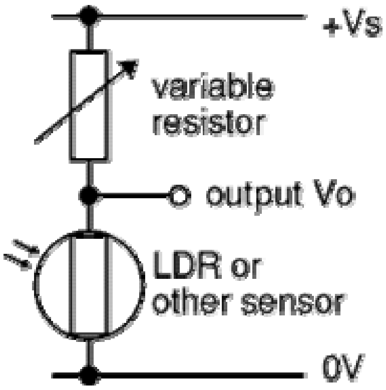

The sensor and variable resistor can be swapped over if necessary

If you are planning to use a variable resistor connected between the +Vs supply and the base of a transistor you must include a resistor in series with the variable resistor. This is to prevent excessive base current destroying the transistor when the variable resistor is reduced to zero. For further information please see the page on Transistor Circuits.

Transistor Circuits

This page explains the operation of transistors in circuits. Practical matters such as testing, precautions when soldering and identifying leads are covered by the Transistors page.

Types of transistor

There are two types of standard transistors, **NPN** and **PNP**, with different circuit symbols. The letters refer to the layers of semiconductor material used to make the transistor. Most transistors used today are NPN because this is the easiest type to make from silicon. This page is mostly about NPN transistors and if you are new to electronics it is best to start by learning how to use these first.

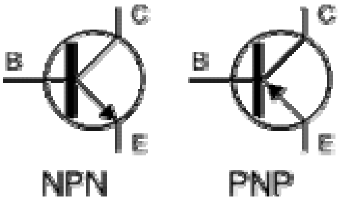

Transistor circuit symbols

The leads are labeled **base** (B), **collector** (C) and **emitter** (E).
These terms refer to the internal operation of a transistor but they are not much help in understanding how a transistor is used, so just treat them as labels!

A Darlington pair is two transistors connected together to give a very high current gain.

In addition to standard (bipolar junction) transistors, there are **field-effect transistors** which are usually referred to as **FET**. They have different circuit symbols and properties and they are not (yet) covered by this page.

Transistor Currents

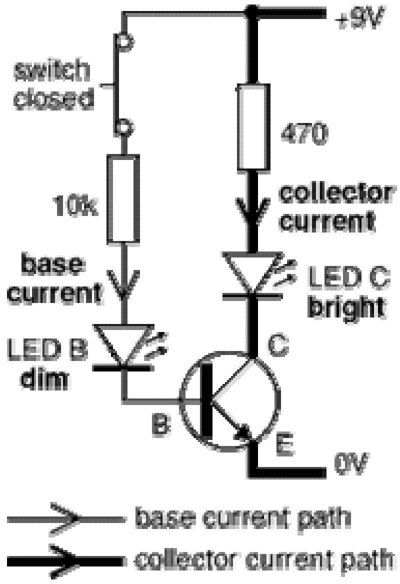

The diagram shows the two current paths through a transistor. You can build this circuit with two standard 5mm red LEDs and any general purpose low power NPN transistor (BC108, BC182 or BC548 for example).

The small **base current** controls the larger **collector current**.

When the switch is closed a small current flows into the base (B) of the transistor. It is just enough to make LED B glow dimly. The transistor amplifies this small current to allow a larger current to flow through from its collector (C) to its emitter (E). This collector current is large enough to make LED C light brightly.

When the switch is open no base current flows, so the transistor switches off the collector current. Both LEDs are off.

A transistor amplifies current and can be used as a switch.

This arrangement where the emitter (E) is in the controlling circuit (base current) and in the controlled circuit (collector current) is called **common emitter mode**. It is the most widely used arrangement for transistors so it is the one to learn first.

Functional model of an NPN transistor

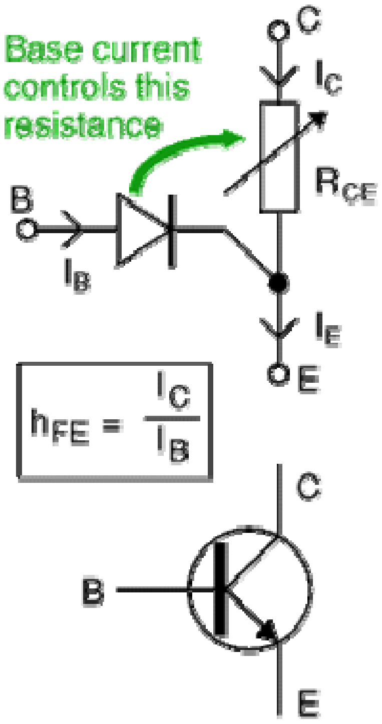
Base current controls this resistance
C
I_C
R_{CE}
B
I_B
I_E
E
$h_{FE} = \frac{I_C}{I_B}$
C
B
E

The operation of a transistor is difficult to explain and understand in terms of its internal structure. It is more helpful to use this functional model:

- The base-emitter junction behaves like a diode.
- A base current I_B flows only when the voltage V_{BE} across the base-emitter junction is 0.7V or more.
- The small base current I_B controls the large collector current Ic.
- $Ic = h_{FE} \times I_B$ (unless the transistor is full on and saturated)
  h_{FE} is the current gain (strictly the DC current gain), a typical value for h_{FE} is 100 (it has no units because it is a ratio)
- The collector-emitter resistance R_{CE} is controlled by the base current I_B:
  - $I_B = 0$ R_{CE} = infinity transistor off
  - I_B small R_{CE} reduced transistor partly on
  - I_B increased $R_{CE} = 0$ transistor full on ('saturated')

Additional notes:

- A resistor is often needed in series with the base connection to limit the base current I_B and prevent the transistor being damaged.
- Transistors have a maximum collector current Ic rating.
- The **current gain h_{FE} can vary widely**, even for transistors of the same type!
- A transistor that is **full on** (with $R_{CE} = 0$) is said to be '**saturated**'.
- When a transistor is saturated the collector-emitter voltage V_{CE} is reduced to almost 0V.
- When a transistor is saturated the collector current Ic is determined by the supply voltage and the external resistance in the collector circuit, not by the transistor's current gain. As a result the ratio Ic/I_B for a saturated transistor is less than the current gain h_{FE}.
- The emitter current $I_E = Ic + I_B$, but Ic is much larger than I_B, so roughly $I_E = Ic$.

There is a table showing technical data for some popular transistors on the transistors page.

Darlington pair

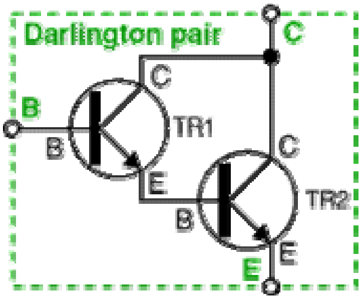

This is two transistors connected together so that the current amplified by the first is amplified further by the second transistor. The overall current gain is equal to the two individual gains multiplied together:

Darlington pair current gain, $h_{FE} = h_{FE1} \times h_{FE2}$
(h_{FE1} and h_{FE2} are the gains of the individual transistors)

This gives the Darlington pair a very high current gain, such as 10000, so that only a tiny base current is required to make the pair switch on.

A Darlington pair behaves like a single transistor with a very high current gain. It has three leads (**B**, **C** and **E**) which are equivalent to the leads of a standard individual transistor. To turn on there must be 0.7V across both the base-emitter junctions which are connected in series inside the Darlington pair, therefore it requires 1.4V to turn on.

Darlington pairs are available as complete packages but you can make up your own from two transistors; TR1 can be a low power type, but normally TR2 will need to be high power. The maximum collector current Ic(max) for the pair is the same as Ic(max) for TR2.

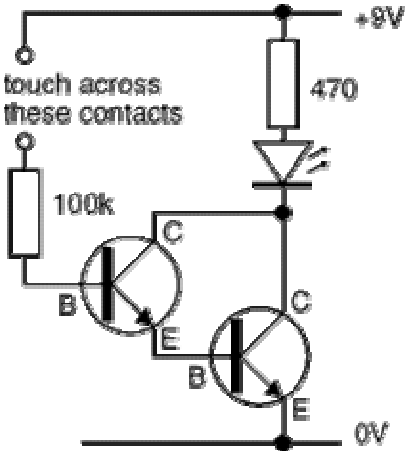

Touch Switch Circuit

A Darlington pair is sufficiently sensitive to respond to the small current passed by your skin and it can be used to make a **touch-switch** as shown in the diagram. For this circuit which just lights an LED the two transistors can be any general purpose low power transistors. The 100kΩ resistor protects the transistors if the contacts are linked with a piece of wire.

Using a Transistor as a Switch

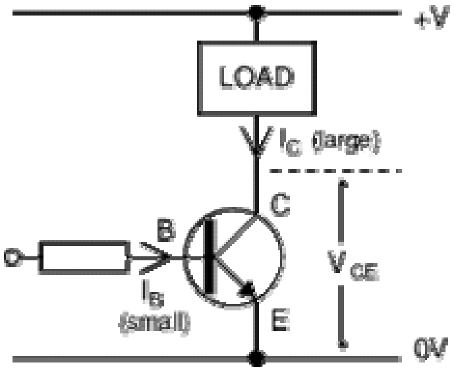

When a transistor is used as a switch it must be either **OFF** or **fully ON**. In the fully ON state the voltage V_{CE} across the transistor is almost zero and the transistor is said to be **saturated** because it cannot pass any more collector current Ic. The output device switched by the transistor is usually called the 'load'.

The power developed in a switching transistor is very small:

- In the **OFF** state: power = Ic × V_{CE}, but Ic = 0, so the power is zero.
- In the **full ON** state: power = Ic × V_{CE}, but V_{CE} = 0 (almost), so the power is very small.

This means that the transistor should not become hot in use and you do not need to consider its maximum power rating. The important ratings in switching circuits are the **maximum collector current Ic(max)** and the **minimum current gain h_{FE}(min)**. The transistor's voltage ratings may be ignored unless you are using a supply voltage of more than about 15V. There is a table showing technical data for some popular transistors on the transistors page.

For information about the operation of a transistor please see the functional model above.

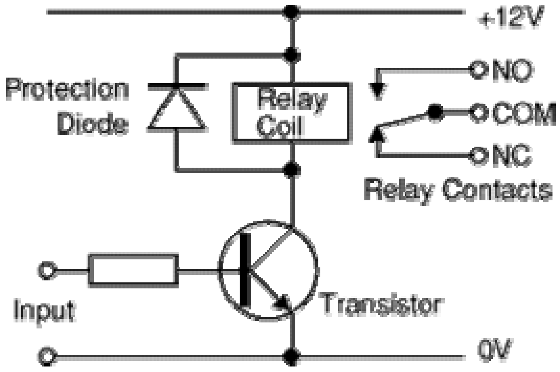

Protection Diode

If the load is a **motor**, **relay** or **solenoid** (or any other device with a coil) a diode must be connected across the load to protect the transistor (and chip) from damage when the load is switched off. The diagram shows how this is connected 'backwards' so that it will normally NOT conduct. Conduction only occurs when the load is switched off, at this moment current tries to continue flowing through the coil and it is harmlessly diverted through the diode. Without the diode no current could flow and the coil would produce a damaging high voltage 'spike' in its attempt to keep the current flowing.

When to use a Relay

Transistors cannot switch AC or high voltages (such as mains electricity) and they are not usually a good choice for switching large currents (> 5A). In these cases a relay will be needed, but note that a low power transistor may still be needed to switch the current for the relay's coil!

Relays

Advantages of relays:

- Relays can switch **AC and DC**, transistors can only switch DC.
- Relays can switch **high voltages**, transistors cannot.
- Relays are a better choice for switching **large currents** (> 5A).
- Relays can switch **many contacts** at once.

Disadvantages of relays:

- Relays are **bulkier** than transistors for switching small currents.
- Relays **cannot switch rapidly**; transistors can switch many times per second.
- Relays **use more power** due to the current flowing through their coil.
- Relays **require more current than many chips can provide**, so a low power transistor may be needed to switch the current for the relay's coil.

Connecting a Transistor to the Output from a Chip

Most chips cannot supply large output currents so it may be necessary to use a transistor to switch the larger current required for output devices such as lamps, motors and relays. The 555 timer chip is unusual because it can supply a relatively large current of up to 200mA which is sufficient for some output devices such as low current lamps, buzzers and many relay coils without needing to use a transistor.

A transistor can also be used to enable a chip connected to a low voltage supply (such as 5V) to switch the current for an output device with a separate higher voltage supply (such as 12V). The two power supplies must be linked, normally this is done by linking their 0V connections. In this case you should use an NPN transistor.

A resistor R_B is required to limit the current flowing into the base of the transistor and prevent it being damaged. However, R_B must be sufficiently low to ensure that the transistor is thoroughly saturated to prevent it overheating, this is particularly important if the transistor is switching a large current (> 100mA). A safe rule is to make the base current I_B about five times larger than the value which should just saturate the transistor.

Choosing a suitable NPN Transistor

The circuit diagram shows how to connect an **NPN transistor**; this will switch on the load when the chip output is **high**. If you need the opposite action, with the load switched on when the chip output is **low** (0V) please see the circuit for a PNP transistor below.

The procedure below explains how to choose a suitable switching transistor.

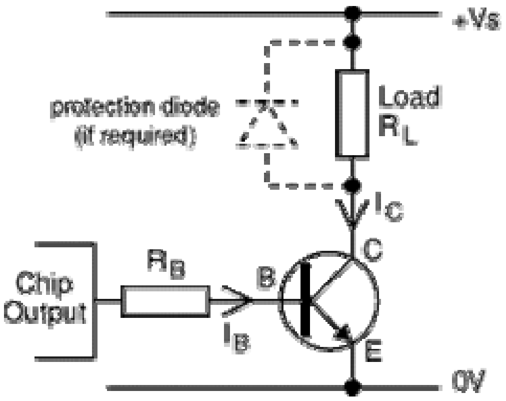

NPN transistor switch
(load is on when chip output is high)

Using units in calculations
Remember to use V, A and Ω or V, mA and kΩ. For more details please see the Ohm's Law page.

1. The transistor's maximum collector current Ic(max) must be greater than the load current Ic.

$$\textbf{load current Ic} = \frac{\textbf{supply voltage Vs}}{\textbf{load resistance } R_L}$$

2. The transistor's minimum current gain h_{FE}(min) must be at least **five** times the load current Ic divided by the maximum output current from the chip.

$$h_{FE}(\text{min}) > 5 \times \frac{\textbf{load current Ic}}{\textbf{max. chip current}}$$

3. Choose a transistor which meets these requirements and make a note of its properties: Ic(max) and h_{FE}(min).
   There is a table showing technical data for some popular transistors on the transistors page.
4. Calculate an approximate value for the base resistor:

$$R_B = \frac{Vc \times h_{FE}}{5 \times Ic}$$ where Vc = chip supply voltage
(in a simple circuit with one supply this is Vs)

5. For a simple circuit where the chip and the load share the same power supply (Vc = Vs) you may prefer to use: $R_B = 0.2 \times R_L \times h_{FE}$
6. Then choose the nearest standard value for the base resistor.
7. Finally, remember that if the load is a motor or relay coil a protection diode is required.

Example:

The output from a 4000 series CMOS chip is required to operate a relay with a 100Ω coil.
The supply voltage is 6V for both the chip and load. The chip can supply a maximum current of 5mA.

1. Load current = Vs/R_L = 6/100 = 0.06A = 60mA, so transistor must have Ic(max) > 60mA.
2. The maximum current from the chip is 5mA, so transistor must have h_{FE}(min) > 60 (5 × 60mA/5mA).
3. Choose general purpose low power transistor BC182 with Ic(max) = 100mA and h_{FE}(min) = 100.
4. $R_B = 0.2 \times R_L \times h_{FE} = 0.2 \times 100 \times 100 = 2000\Omega$. so choose R_B = 1k8 or 2k2.
5. The relay coil requires a protection diode.

Choosing a suitable PNP Transistor

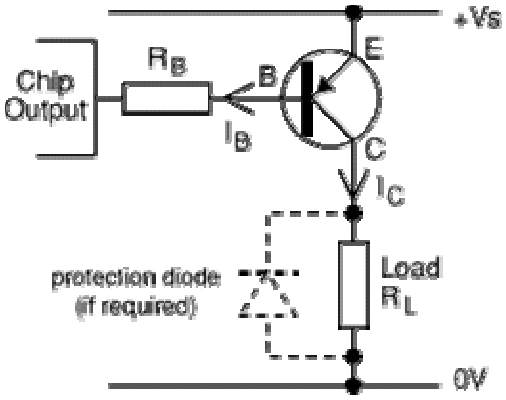

PNP Transistor Switch
(load is on when chip output is low)

The circuit diagram shows how to connect a **PNP transistor**, this will switch on the load when the chip output is **low** (0V). If you need the opposite action, with the load switched on when the chip output is **high** please see the circuit for an NPN transistor above.

The procedure for choosing a suitable PNP transistor is exactly the same as that for an NPN transistor described above.

Using a Transistor Switch with Sensors

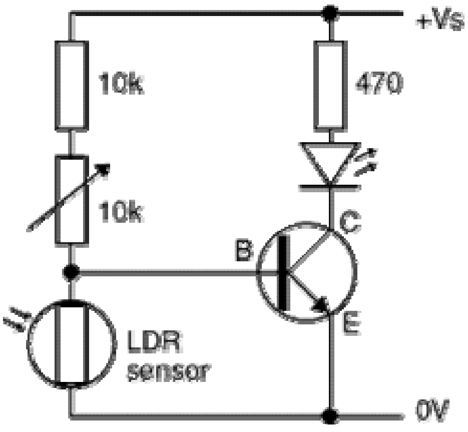

LED lights when the LDR is **dark**

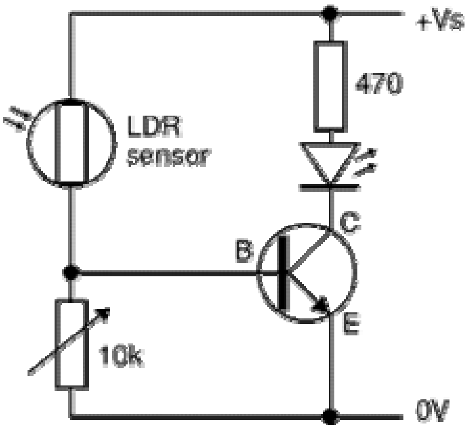

LED lights when the LDR is **bright**

The top circuit diagram shows an LDR (light sensor) connected so that the LED lights when the LDR is in darkness. The variable resistor adjusts the brightness at which the transistor switches on and off. Any general purpose low power transistor can be used in this circuit.

The 10kΩ fixed resistor protects the transistor from excessive base current (which will destroy it) when the variable resistor is reduced to zero. To make this circuit switch at a suitable brightness you may need to experiment with different values for the fixed resistor, but it must not be less than 1kΩ.

If the transistor is switching a load with a coil, such as a motor or relay, remember to add a protection diode across the load.

The switching action can be inverted, so the LED lights when the LDR is brightly lit, by swapping the LDR and variable resistor. In this case the fixed resistor can be omitted because the LDR resistance cannot be reduced to zero.

Note that the switching action of this circuit is not particularly good because there will be an intermediate brightness when the transistor will be **partly on** (not saturated). In this state the transistor is in danger of overheating unless it is switching a small current. There is no problem with the small LED current, but the larger current for a lamp, motor or relay is likely to cause overheating.

Other sensors, such as a thermistor, can be used with this circuit, but they may require a different variable resistor. You can calculate an approximate value for the variable resistor (Rv) by using a multimeter to find the minimum and maximum values of the sensor's resistance (Rmin and Rmax):

Variable resistor, Rv = square root of (Rmin × Rmax)

For example an LDR: Rmin = 100Ω, Rmax = 1MΩ, so Rv = square root of (100 × 1M) = 10kΩ.

You can make a much better switching circuit with sensors connected to a suitable IC (chip). The switching action will be much sharper with no partly on state.

A Transistor Inverter (NOT gate)

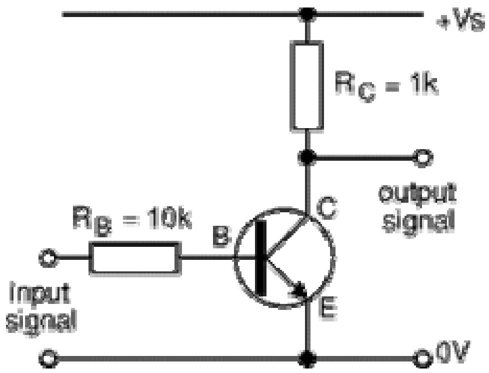

Inverters (NOT gates) are available on logic chips but if you only require one inverter it is usually better to use this circuit. The output signal (voltage) is the inverse of the input signal:

- When the input is high (+Vs) the output is low (0V).
- When the input is low (0V) the output is high (+Vs).

Any general purpose low power NPN transistor can be used. For general use R_B = 10kΩ and R_C = 1kΩ, then the inverter output can be connected to a device with an input impedance (resistance) of at least 10kΩ such as a logic chip or a 555 timer (trigger and reset inputs).

If you are connecting the inverter to a CMOS logic chip input (very high impedance) you can increase R_B to 100kΩ and R_C to 10kΩ, this will reduce the current used by the inverter.

Analogue and Digital Systems

Analogue Systems

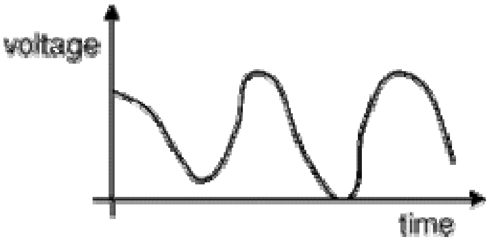

Analogue signal

Analogue systems process analogue signals which can take any value within a range, for example the output from an LDR (light sensor) or a microphone.

An audio amplifier is an example of an analogue system. The amplifier produces an output voltage which can be any value within the range of its power supply.

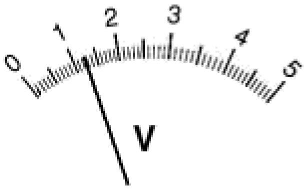

Analogue meter display

An analogue meter can display any value within the range available on its scale. However, the precision of readings is limited by our ability to read them. For example the meter on the right shows 1.25V because the pointer is estimated to be half way between 1.2 and 1.3. The analogue meter can show any value between 1.2 and 1.3 but we are unable to read the scale more precisely than about half a division.

All electronic circuits suffer from 'noise' which is unwanted signal mixed in with the desired signal, for example an audio amplifier may pick up some mains 'hum' (the 50Hz frequency of the UK mains electricity supply). Noise can be difficult to eliminate from analogue signals because it may be hard to distinguish from the desired signal.

Digital Systems

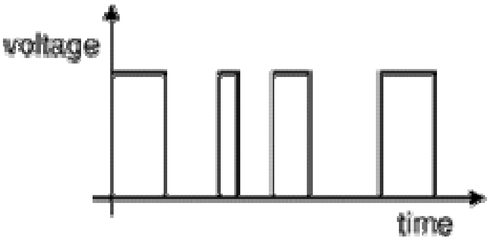

Digital (logic) signal

Digital systems process digital signals which can take only a limited number of values (discrete steps), usually just two values are used: the positive supply voltage (+Vs) and zero volts (0V).

Digital systems contain devices such as logic gates, flip-flops, shift registers and counters. A computer is an example of a digital system.

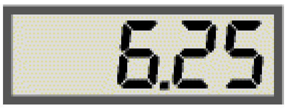

Digital meter display

A digital meter can display many values, but not every value within its range. For example the display on the right can show 6.25 and 6.26 but not a value between them. This is not a problem because digital meters normally have sufficient digits to show values more precisely than it is possible to read an analogue display.

Logic Signals

Most digital systems use the simplest possible type of signal which has just two values. This type of signal is called a **logic signal** because the two values (or states) can be called true and false. Normally the positive supply voltage +Vs represent true and 0V represents false. Other labels for the true and false states are shown in the table below.

Logic states	
True	**False**
1	0
High	Low
+Vs	0V
On	Off

Noise is relatively easy to eliminate from digital signals because it is easy to distinguish from the desired signal which can only have particular values. For example: if the signal is meant to be +5V (true) or 0V (false), noise of up to 2.5V can be eliminated by treating all voltages greater than 2.5V as true and all voltages less than 2.5V as false.

Logic Gates

Introduction

Logic gates process signals which represent **true** or **false**. Normally the positive supply voltage +Vs represent true and 0V represents false. Other terms which are used for the true and false states are shown in the table on the right. It is best to be familiar with them all.

Gates are identified by their function: NOT, AND, NAND, OR, NOR, EX-OR and EX-NOR. Capital letters are normally used to make it clear that the term refers to a logic gate.

Logic states	
True	**False**
1	0
High	Low
+Vs	0V
On	Off

Note that logic gates are not always required because simple logic functions can be performed with switches or diodes:

- Switches in series (AND function)
- Switches in parallel (OR function)
- Combining chip outputs with diodes (OR function)

Logic Gate Symbols

There are two series of symbols for logic gates:

- The **traditional symbols** have distinctive shapes making them easy to recognize so they are widely used in industry and education.

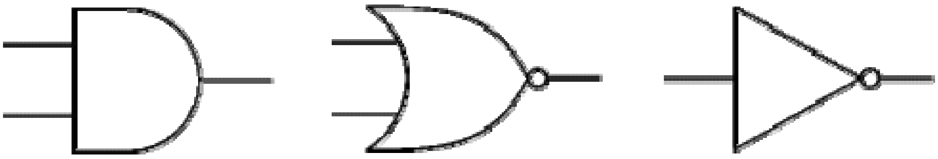

- The **IEC (International Electrotechnical Commission) symbols** are rectangles with a symbol inside to show the gate function. They are rarely used despite their official status, but you may need to know them for an examination.

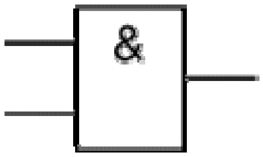

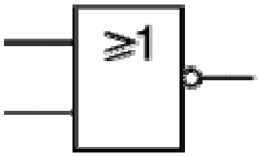

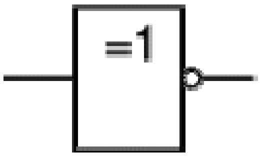

Inputs and Outputs

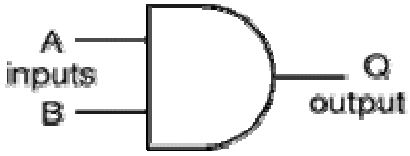

Gates have two or more inputs, except a NOT gate which has only one input. All gates have only one output. Usually the letters A, B, C and so on are used to label inputs, and Q is used to label the output. On this page the inputs are shown on the left and the output on the right.

The Inverting Circle (o)

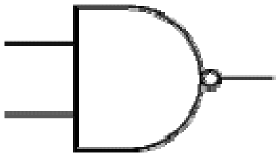

Some gate symbols have a circle on their output which means that their function includes **inverting** of the output. It is equivalent to feeding the output through a NOT gate. For example the NAND (Not AND) gate symbol shown on the right is the same as an AND gate symbol but with the addition of an inverting circle on the output.

Truth Tables

Input A	Input B	Output Q
0	0	0
0	1	0
1	0	0
1	1	1

A truth table is a good way to show the function of a logic gate. It shows the output states for every possible combination of input states. The symbols 0 (false) and 1 (true) are usually used in truth tables. The example truth table on the right shows the inputs and output of an AND gate.

There are summary truth tables below showing the output states for all types of 2-input and 3-input gates. These can be helpful if you are trying to select a suitable gate.

Logic ICs

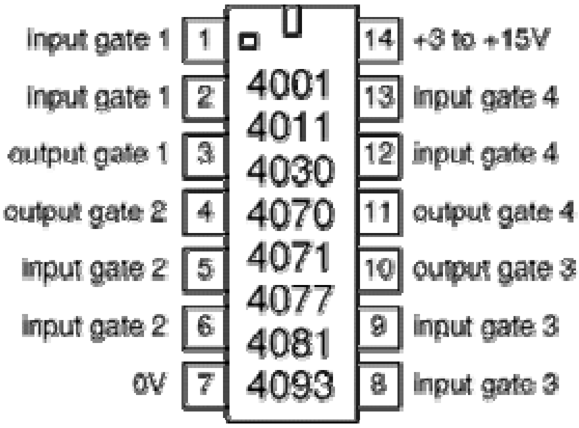
Input gate 1
Input gate 1
output gate 1
output gate 2
input gate 2
input gate 2
0V
1
2
3
4
5
6
7
4001
4011
4030
4070
4071
4077
4081
4093
14
13
12
11
10
9
8
+3 to +15V
Input gate 4
input gate 4
output gate 4
output gate 3
Input gate 3
Input gate 3

Logic gates are available on special ICs (chips) which usually contain several gates of the same type, for example the 4001 IC contains four 2-input NOR gates. There are several families of logic ICs and they can be split into two groups:

- 4000 Series
- 74 Series

To quickly compare the different families please see:

- Summary table of logic families

The **4000** and **74HC** families are the best for battery powered projects because they will work with a good range of supply voltages and they use very little power. However, if you are using them to design circuits and investigate logic gates please remember that all unused inputs MUST be connected to the power supply (either +Vs or 0V), this applies even if that part of the IC is not being used in the circuit!

NOT gate (Inverter)

The output Q is true when the input A is NOT true, the output is the inverse of the input:
Q = NOT A
A NOT gate can only have one input. A NOT gate is also called an inverter.

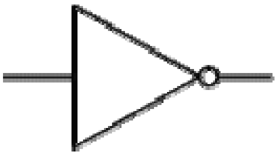

Traditional symbol

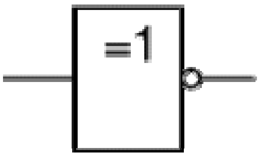

IEC symbol

Input A	Output Q
0	1
1	0

Truth Table

AND gate

The output Q is true if input A AND input B are both true: **Q = A AND B**
An AND gate can have two or more inputs, its output is true if all inputs are true.

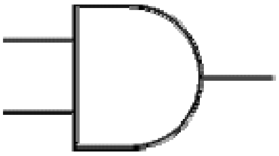

Traditional symbol

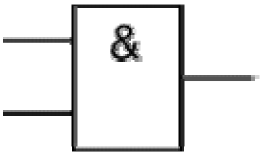

IEC symbol

Input A	Input B	Output Q
0	0	0
0	1	0
1	0	0
1	1	1

Truth Table

NAND gate (NAND = Not AND)

This is an AND gate with the output inverted, as shown by the 'o' on the output.
The output is true if input A AND input B are NOT both true: **Q = NOT (A AND B)**
A NAND gate can have two or more inputs, its output is true if NOT all inputs are true.

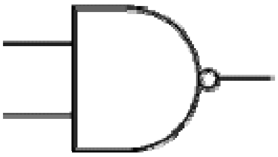

Traditional symbol

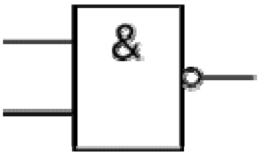

IEC symbol

Input A	Input B	Output Q
0	0	1
0	1	1
1	0	1
1	1	0

Truth Table

OR gate

The output Q is true if input A OR input B is true (or both of them are true): **Q = A OR B**
An OR gate can have two or more inputs, its output is true if at least one input is true.

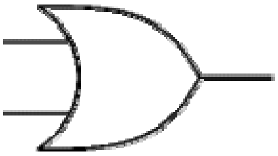

Traditional symbol

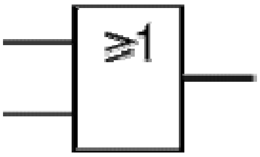

IEC symbol

Input A	Input B	Output Q
0	0	0
0	1	1
1	0	1
1	1	1

Truth Table

NOR gate (NOR = Not OR)

This is an OR gate with the output inverted, as shown by the 'o' on the output.
The output Q is true if NOT inputs A OR B are true: **Q = NOT (A OR B)**
A NOR gate can have two or more inputs, its output is true if no inputs are true.

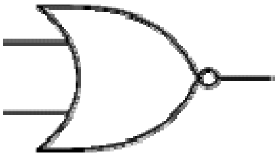

Traditional symbol

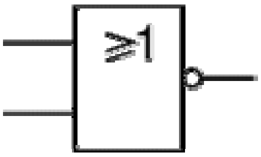

IEC symbol

Input A	Input B	Output Q
0	0	1
0	1	0
1	0	0
1	1	0

Truth Table

EX-OR (EXclusive-OR) gate

The output Q is true if either input A is true OR input B is true, **but not when both of them are true**: **Q = (A AND NOT B) OR (B AND NOT A)**
This is like an OR gate but excluding both inputs being true.
The output is true if inputs A and B are **DIFFERENT**.
EX-OR gates can only have 2 inputs.

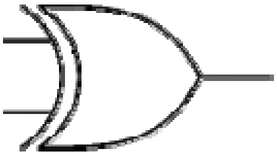

Traditional symbol

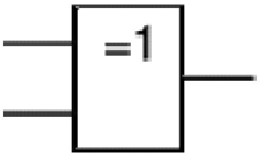

IEC symbol

Input A	Input B	Output Q
0	0	0
0	1	1
1	0	1
1	1	0

Truth Table

EX-NOR (EXclusive-NOR) gate

This is an EX-OR gate with the output inverted, as shown by the 'o' on the output.
The output Q is true if inputs A and B are the **SAME** (both true or both false):
Q = (A AND B) OR (NOT A AND NOT B)
EX-NOR gates can only have 2 inputs.

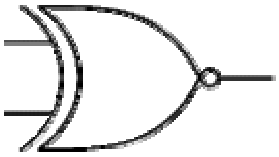

Traditional symbol

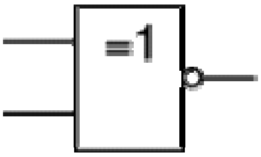

IEC symbol

Input A	Input B	Output Q
0	0	1
0	1	0
1	0	0
1	1	1

Truth Table

Summary Truth Tables

The summary truth tables below show the output states for all types of 2-input and 3-input gates.

Summary for all 2-input Gates							
Inputs		Output of each Gate					
A	B	AND	NAND	OR	NOR	EX-OR	EX-NOR
0	0	0	1	0	1	0	1
0	1	0	1	1	0	1	0
1	0	0	1	1	0	1	0
1	1	1	0	1	0	0	1

Note that EX-OR and EX-NOR gates can only have 2 inputs

Summary for all 3-input Gates						
Inputs			Output of each Gate			
A	B	C	AND	NAND	OR	NOR
0	0	0	0	1	0	1
0	0	1	0	1	1	0
0	1	0	0	1	1	0
0	1	1	0	1	1	0
1	0	0	0	1	1	0
1	0	1	0	1	1	0
1	1	0	0	1	1	0
1	1	1	1	0	1	0

Combinations of Logic Gates

Logic gates can be combined to produce more complex functions. They can also be combined to substitute one type of gate for another.

For example to produce an output Q which is true only when input A is true and input B is false, as shown in the truth table on the right, we can combine a NOT gate and an AND gate like this:

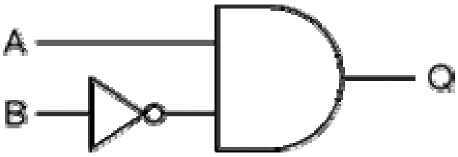

Q = A AND NOT B

Input A	Input B	Output Q
0	0	0
0	1	0
1	0	1
1	1	0

Working out the function of a combination of gates

Truth tables can be used to work out the function of a combination of gates.

For example the truth table on the right show the intermediate outputs D and E as well as the final output Q for the system shown below.

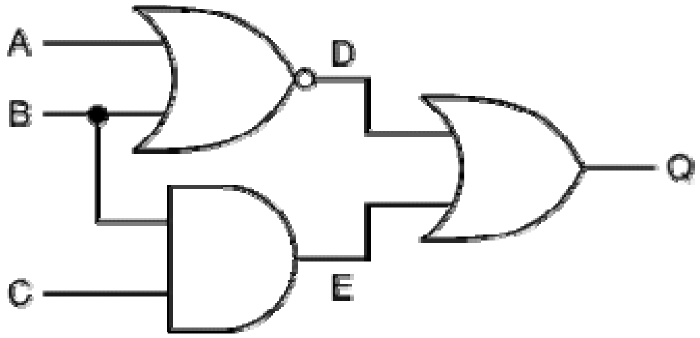

Inputs			Outputs		
A	**B**	**C**	**D**	**E**	**Q**
0	0	0	1	0	1
0	0	1	1	0	1
0	1	0	0	0	0
0	1	1	0	1	1
1	0	0	0	0	0
1	0	1	0	0	0
1	1	0	0	0	0
1	1	1	0	1	1

D = NOT (A OR B)
E = B AND C
Q = D OR E = (NOT (A OR B)) OR (B AND C)

Substituting one type of gate for another

Logic gates are available on ICs which usually contain several gates of the same type, for example four 2-input NAND gates or three 3-input NAND gates. This can be wasteful if only a few gates are required unless they are all the same type. To avoid using too many ICs you can reduce the number of gate inputs or substitute one type of gate for another.

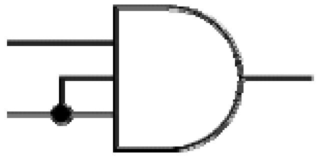

Reducing the number of inputs

The number of inputs to a gate can be reduced by connecting two (or more) inputs together. The diagram shows a 3-input AND gate operating as a 2-input AND gate.

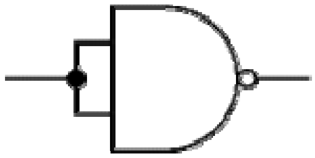

Making a NOT gate from a NAND or NOR gate

Reducing a NAND or NOR gate to just one input creates a NOT gate. The diagram shows this for a 2-input NAND gate.

Any gate can be built from NAND or NOR gates

As well as making a NOT gate, NAND or NOR gates can be combined to create any type of gate! This enables a circuit to be built from just one type of gate, either NAND or NOR. For example an AND gate is a NAND gate then a NOT gate (to undo the inverting function). Note that AND and OR gates cannot be used to create other gates because they lack the inverting (NOT) function.

To change the type of gate, such as changing OR to AND, you must do three things:

- Invert (NOT) each input.
- Change the gate type (OR to AND, or AND to OR)
- Invert (NOT) the output.

For example an OR gate can be built from NOTed inputs fed into a NAND (AND + NOT) gate.

NAND gate equivalents

The table below shows the NAND gate equivalents of NOT, AND, OR and NOR gates:

Gate		Equivalent in NAND gates
NOT	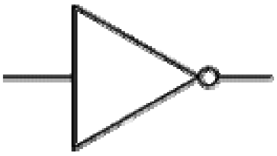	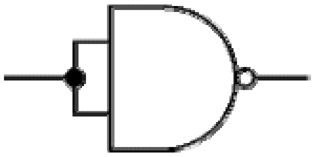
AND	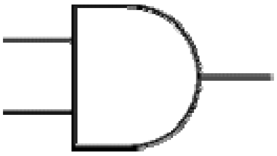	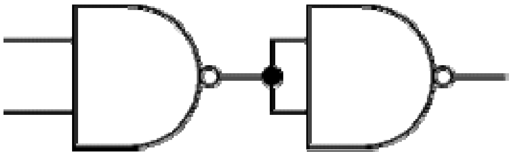
OR	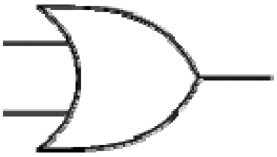	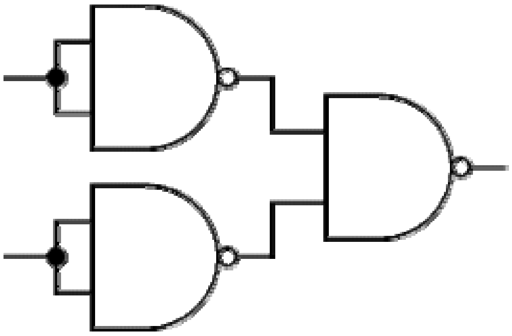
NOR	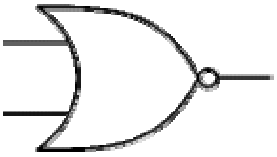	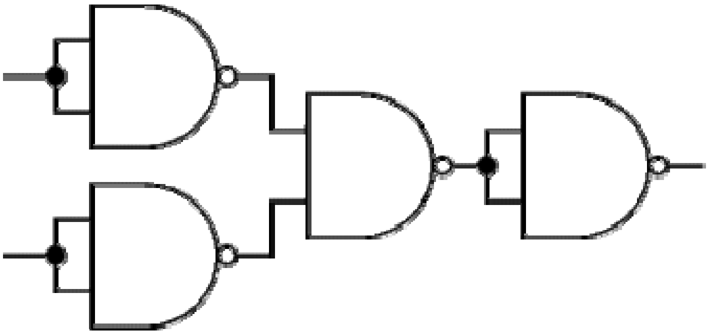

Substituting gates in an example logic system

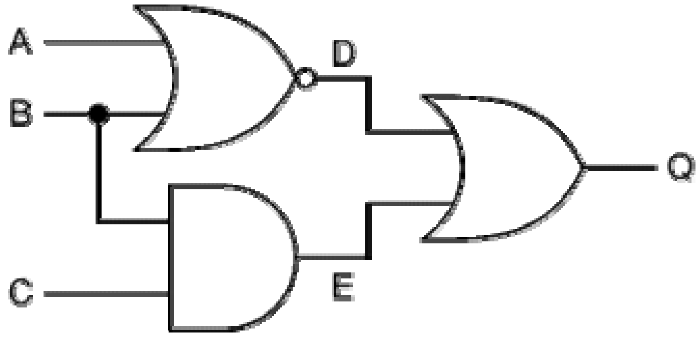

The original system has 3 different gates: NOR, AND and OR. This requires three ICs (one for each type of gate).

To re-design this system using NAND gates only begin by replacing each gate with its NAND gate equivalent, as shown in the diagram below.

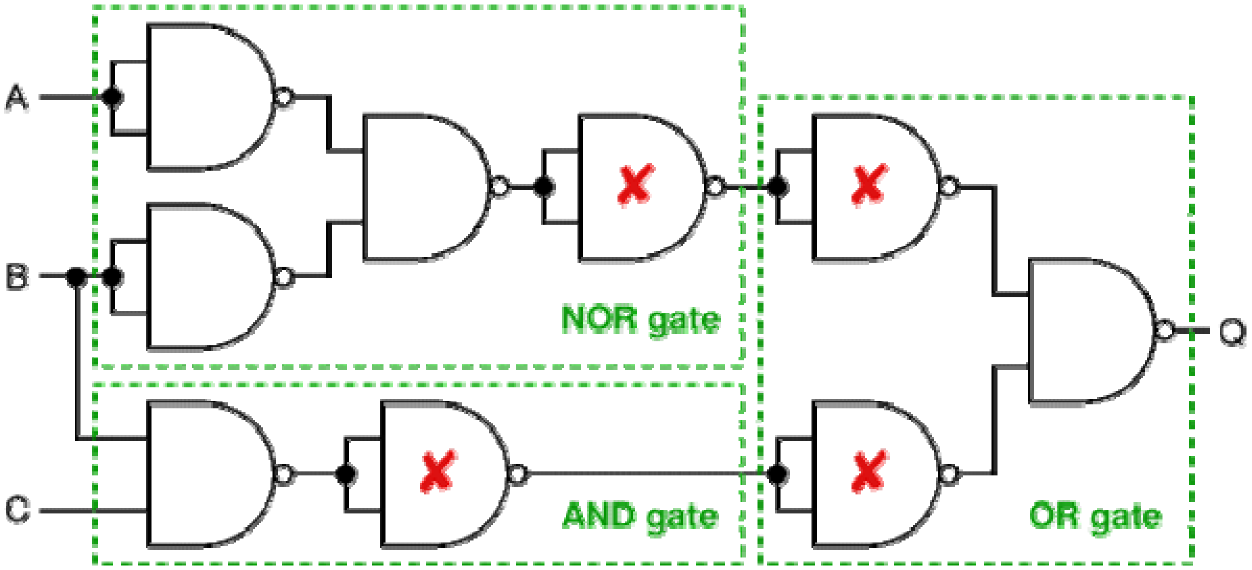
A
B
C
NOR gate
AND gate
OR gate
Q

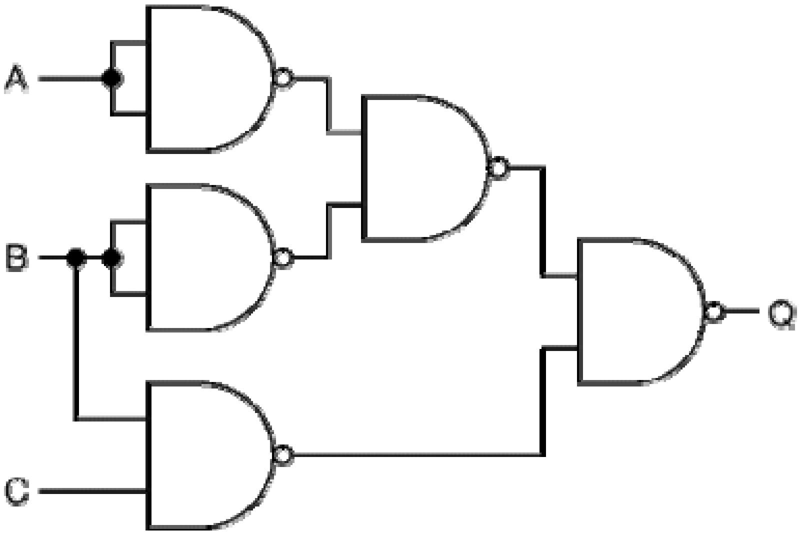

Then simplify the system by deleting adjacent pairs of NOT gates (marked X above). This can be done because the second NOT gate cancels the action of the first.

The final system is shown on the right. It has five NAND gates and requires two ICs (with four gates on each IC). This is better than the original system which required three ICs (one for each type of gate).

Substituting NAND (or NOR) gates does not always increase the number of gates, but when it does (as in this example) the increase is usually only one or two gates. The real benefit is reducing the number of ICs required by using just one type of gate.

Capacitance and Uses of Capacitors

Capacitance

Capacitance (symbol C) is a measure of a capacitor's ability to **store charge**. A large capacitance means that more charge can be stored. Capacitance is measured in farads, symbol F. However 1F is very large, so prefixes (multipliers) are used to show the smaller values:

- µ (micro) means 10^{-6} (millionth),
  so 1000000µF = 1F
- n (nano) means 10^{-9} (thousand-millionth),
  so 1000nF = 1µF
- p (pico) means 10^{-12} (million-millionth),
  so 1000pF = 1nF

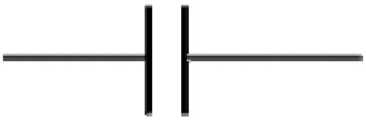

unpolarized capacitor symbol

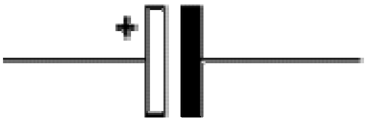

polarized capacitor symbol

Charge and Energy Stored

The amount of charge (symbol Q) stored by a capacitor is given by:

Charge, **Q = C × V** where:
Q = charge in coulombs (C)
C = capacitance in farads (F)
V = voltage in volts (V)

When they store charge, capacitors are also storing energy:

Energy, **E = ½QV = ½CV²** where E = energy in joules (J).

Note that capacitors return their stored energy to the circuit. They do not 'use up' electrical energy by converting it to heat as a resistor does. The energy stored by a capacitor is much smaller than the energy stored by a battery so they cannot be used as a practical source of energy for most purposes.

Capacitive Reactance Xc

Capacitive reactance (symbol Xc) is a measure of a capacitor's opposition to AC (alternating current). Like resistance it is measured in ohms, Ω, but reactance is more complex than resistance because its value depends on the frequency (f) of the electrical signal passing through the capacitor as well as on the capacitance, C.

Capacitive reactance, **Xc =** $\frac{1}{2\pi fC}$ where:

Xc = reactance in ohms (Ω)
f = frequency in hertz (**Hz**)
C = capacitance in farads (**F**)

The reactance Xc is large at low frequencies and small at high frequencies. For steady DC which is zero frequency, Xc is infinite (total opposition), hence the rule that **capacitors pass AC but block DC**.

For example a 1µF capacitor has a reactance of 3.2kΩ for a 50Hz signal, but when the frequency is higher at 10kHz its reactance is only 16Ω.

Note: the symbol Xc is used to distinguish capacitive reactance from inductive reactance X_L which is a property of inductors. The distinction is important because X_L increases with frequency (the opposite of Xc) and if both X_L and Xc are present in a circuit the combined reactance (X) is the *difference* between them. See the page on Impedance.

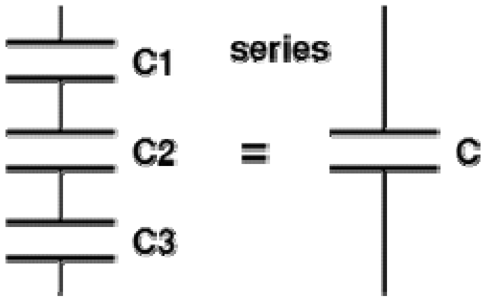

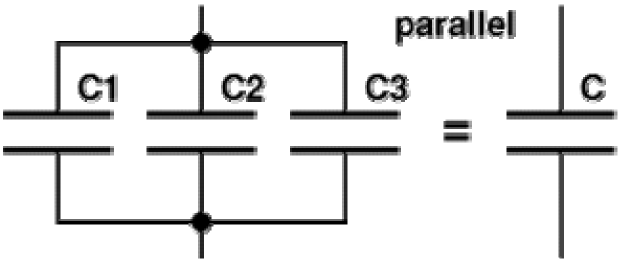

Capacitors in Series and Parallel

Combined capacitance (C) of capacitors connected in **series**: $\frac{1}{C} = \frac{1}{C1} + \frac{1}{C2} + \frac{1}{C3} + \ldots$

Combined capacitance (C) of capacitors connected in **parallel**: **C = C1 + C2 + C3 + ...**

Two or more capacitors are rarely deliberately connected in series in real circuits, but it can be useful to connect capacitors in parallel to obtain a very large capacitance, for example to smooth a power supply.

Note that these equations are the opposite way round for resistors in series and parallel.

Charging a Capacitor

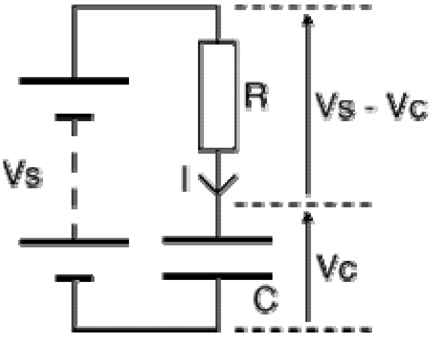

The capacitor (C) in the circuit diagram is being charged from a supply voltage (Vs) with the current passing through a resistor (R). The voltage across the capacitor (Vc) is initially zero but it increases as the capacitor charges. The capacitor is fully charged when Vc = Vs. The charging current (I) is determined by the voltage across the resistor (Vs - Vc):

Charging current, **I = (Vs - Vc) / R** *(note that Vc is increasing)*

At first Vc = 0V so the **initial current, Io = Vs / R**

Vc increases as soon as charge (Q) starts to build up (Vc = Q/C), this reduces the voltage across the resistor and therefore reduces the charging current. This means that the rate of charging becomes progressively slower.

time constant = R × C	where:	time constant is in seconds (s) R = resistance in ohms (Ω) C = capacitance in farads (F)

For example:
If R = 47kΩ and C = 22µF, then the time constant, RC = 47kΩ × 22µF = 1.0s.
If R = 33kΩ and C = 1µF, then the time constant, RC = 33kΩ × 1µF = 33ms.

A large time constant means the capacitor charges slowly. *Note that the time constant is a property of the **circuit** containing the capacitance and resistance, it is not a property of a capacitor alone.*

The time constant is the time taken for the charging (or discharging) current (I) to fall to $^{1}/e$ of its initial value (Io). 'e' is the base of natural logarithms, an important number in mathematics (like π). e = 2.71828 (to 6 significant figures) so we can roughly say that the time constant is the time taken for the current to fall to $^{1}/_{3}$ of its initial value.

After each time constant the current falls by ^{1}/e (about $^1/_3$). **After 5 time constants** (5RC) the current has fallen to less than 1% of its initial value and we can reasonably say that the capacitor is **fully charged**, but in fact the capacitor takes for ever to charge fully!

The **bottom graph** shows how the **voltage** (V) increases as the capacitor charges. At first the voltage changes rapidly because the current is large; but as the current decreases, the charge builds up more slowly and the voltage increases more slowly.

After 5 time constants (5RC) the capacitor is almost fully charged with its voltage almost equal to the supply voltage. We can reasonably say that the capacitor is fully charged after 5RC, although really charging continues for ever (or until the circuit is changed).

Time	Voltage	Charge
0RC	0.0V	0%
1RC	5.7V	63%
2RC	7.8V	86%
3RC	8.6V	95%
4RC	8.8V	98%
5RC	8.9V	99%

Graphs showing the current and voltage for a capacitor charging
time constant = RC

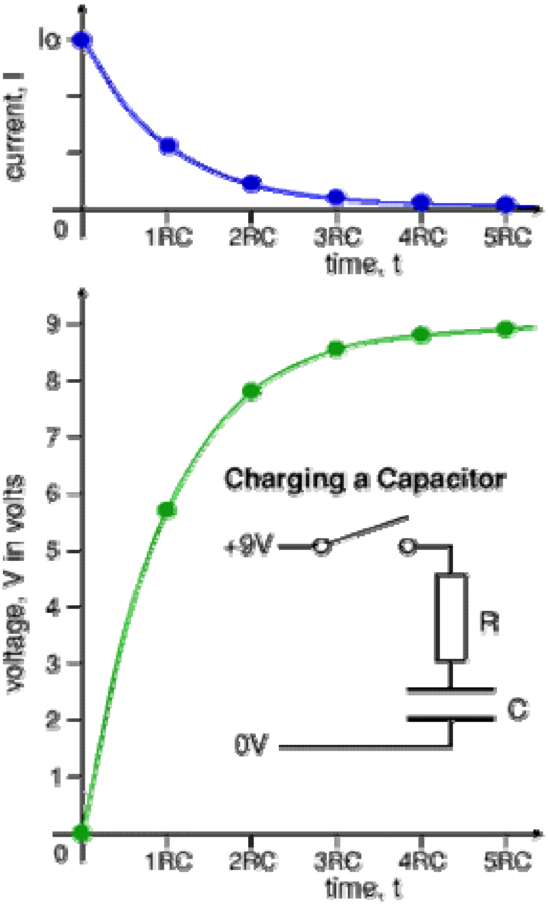

Discharging a capacitor

The **top graph** shows how the **current** (I) decreases as the capacitor discharges. The initial current (Io) is determined by the initial voltage across the capacitor (Vo) and resistance (R):

Initial current, **Io = Vo / R**.

Note that the current graphs are the same shape for both charging and discharging a capacitor. This type of graph is an example of exponential decay.

The **bottom graph** shows how the **voltage** (V) decreases as the capacitor discharges.

At first the current is large because the voltage is large, so charge is lost quickly and the voltage decreases rapidly. As charge is lost the voltage is reduced making the current smaller so the rate of discharging becomes progressively slower.

After 5 time constants (5RC) the voltage across the capacitor is almost zero and we can reasonably say that the capacitor is fully discharged, although really discharging continues for ever (or until the circuit is changed).

Graphs showing the current and voltage for a capacitor discharging
time constant = RC

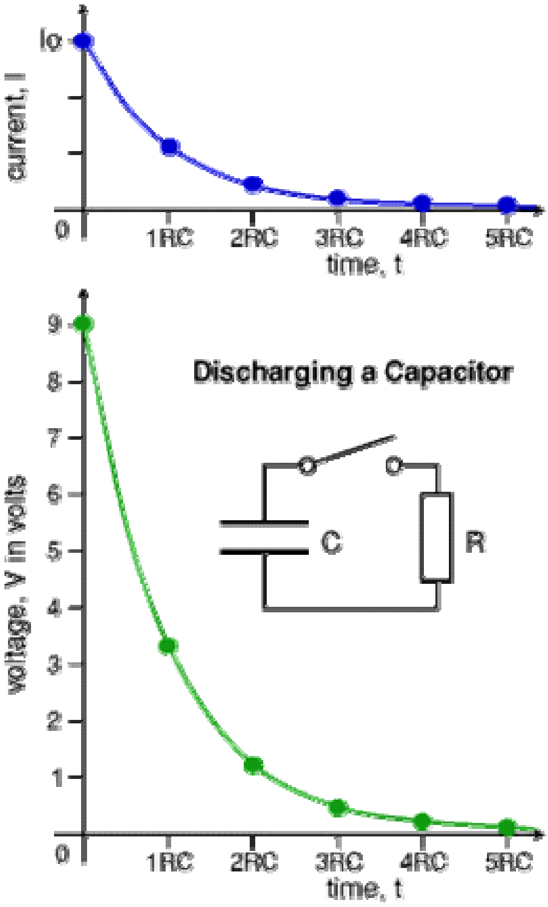

Time	Voltage	Charge
0RC	9.0V	100%
1RC	3.3V	37%
2RC	1.2V	14%
3RC	0.4V	5%
4RC	0.2V	2%
5RC	0.1V	1%

Uses of Capacitors

Capacitors are used for several purposes:

- **Timing** - for example with a 555 timer IC controlling the charging and discharging.
- **Smoothing** - for example in a power supply.
- **Coupling** - for example between stages of an audio system and to connect a loudspeaker.
- **Filtering** - for example in the tone control of an audio system.
- **Tuning** - for example in a radio system.
- **Storing energy** - for example in a camera flash circuit.

Capacitor Coupling (CR-coupling)

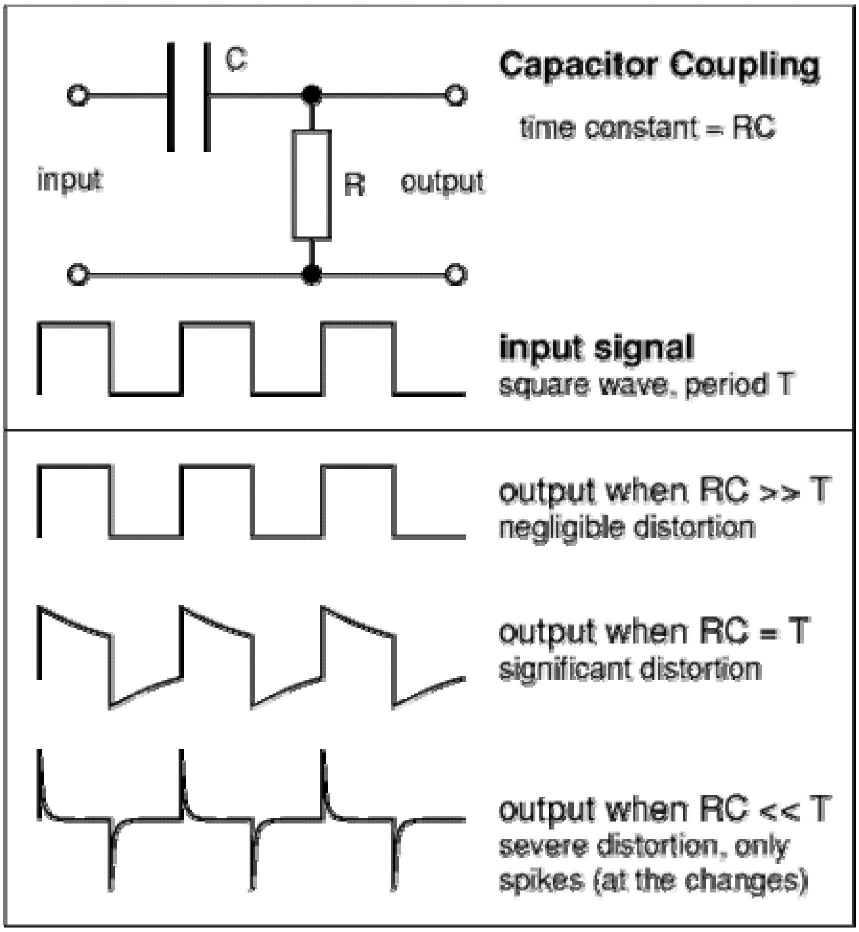

C
Capacitor Coupling
time constant = RC
input
R
output
input signal
square wave, period T
output when RC >> T
negligible distortion
output when RC = T
significant distortion
output when RC << T
severe distortion, only
spikes (at the changes)

Sections of electronic circuits may be linked with a capacitor because capacitors **pass AC** (changing) signals but **block DC** (steady) signals. This is called **capacitor coupling** or **CR-coupling**. It is used between the stages of an audio system to pass on the audio signal (AC) without any steady voltage (DC) which may be present, for example to connect a loudspeaker. It is also used for the 'AC' switch setting on an oscilloscope.

The precise behavior of a capacitor coupling is determined by its time constant (RC). Note that the resistance (R) may be inside the next circuit section rather than a separate resistor.

For successful capacitor coupling in an **audio system** the signals must pass through with little or no distortion. This is achieved if the time constant (RC) is larger than the time period (T) of the lowest frequency audio signals required (typically 20Hz, T = 50ms).

Output when RC >> T

When the time constant is much larger than the time period of the input signal the capacitor does not have sufficient time to significantly charge or discharge, so the signal passes through with negligible distortion.

Output when RC = T

When the time constant is equal to the time period you can see that the capacitor has time to partly charge and discharge before the signal changes. As a result there is significant distortion of the signal as it passes through the CR-coupling. Notice how the sudden changes of the input signal pass straight through the capacitor to the output.

Output when RC << T

When the time constant is much smaller than the time period the capacitor has time to fully charge or discharge after each sudden change in the input signal. Effectively only the sudden changes pass through to the output and they appear as 'spikes', alternately positive and negative. This can be useful in a system which must detect when a signal changes suddenly, but must ignore slow changes.

Impedance and Reactance

Impedance

Impedance (symbol Z) is a measure of the overall opposition of a circuit to current, in other words: how much the circuit **impedes** the flow of current. It is like resistance, but it also takes into account the effects of capacitance and inductance. Impedance is measured in ohms, symbol Ω.

Impedance, **Z** = $\frac{V}{I}$

Resistance, **R** = $\frac{V}{I}$

V = voltage in volts (**V**)
I = current in amps (**A**)
Z = impedance in ohms (**Ω**)
R = resistance in ohms (**Ω**)

Impedance is more complex than resistance because the effects of capacitance and inductance vary with the frequency of the current passing through the circuit and this means **impedance varies with frequency**! The effect of resistance is constant regardless of frequency.

The term 'impedance' is often used (quite correctly) for simple circuits which have no capacitance or inductance - for example to refer to their 'input impedance' or 'output impedance'. This can seem confusing if you are learning electronics, but for these simple circuits you can assume that it is just another word for resistance.

Four electrical quantities determine the impedance (Z) of a circuit: resistance (R), capacitance (C), inductance (L) and frequency (f).

Impedance can be split into two parts:

- **Resistance R** (the part which is constant regardless of frequency)
- **Reactance X** (the part which varies with frequency due to capacitance and inductance)

For further details please see the section on Reactance below.

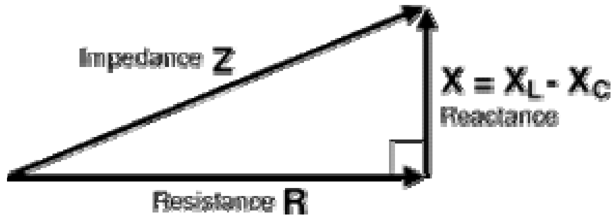

$$\text{Impedance,}\quad Z = \sqrt{R^2 + X^2}$$

The capacitance and inductance cause a **phase shift*** between the current and voltage which means that the resistance and reactance cannot be simply added up to give impedance. Instead they must be **added as vectors** with reactance at right angles to resistance as shown in the diagram.

* **Phase shift** means that the current and voltage are out of step with each other. Think of charging a capacitor. When the voltage across the capacitor is zero, the current is at a maximum; when the capacitor has charged and the voltage is at a maximum, the current is at a minimum. The charging and discharging occur continually with AC and the current reaches its maximum shortly before the voltage reaches its maximum: so we say the current leads the voltage.

Reactance, X

Reactance (symbol X) is a measure of the opposition of capacitance and inductance to current. Reactance varies with the frequency of the electrical signal. Reactance is measured in ohms, symbol Ω.

There are two types of reactance: capacitive reactance (Xc) and inductive reactance (X_L).

The **total reactance (X)** is the *difference* between the two: $\mathbf{X = X_L - Xc}$

- **Capacitive reactance, Xc**

$$Xc = \frac{1}{2\pi fC}$$

where:
Xc = reactance in ohms (Ω)
f = frequency in hertz (**Hz**)
C = capacitance in farads (**F**)

- Xc is large at low frequencies and small at high frequencies.
  For steady DC which is zero frequency, Xc is infinite (total opposition),
  hence the rule that **capacitors pass AC but block DC**.
- For example: a 1µF capacitor has a reactance of 3.2kΩ for a 50Hz signal,
  but when the frequency is higher at 10kHz its reactance is only 16Ω.
- **Inductive reactance, X_L**

$$X_L = 2\pi fL$$

where:
X_L = reactance in ohms (Ω)
f = frequency in hertz (**Hz**)
L = inductance in henrys (**H**)

- X_L is small at low frequencies and large at high frequencies.
  For steady DC (frequency zero), X_L is zero (no opposition),
  hence the rule that **inductors pass DC but block high frequency AC**.
- For example: a 1mH inductor has a reactance of only 0.3Ω for a 50Hz signal,
  but when the frequency is higher at 10kHz its reactance is 63Ω.

Input Impedance Z_{IN}

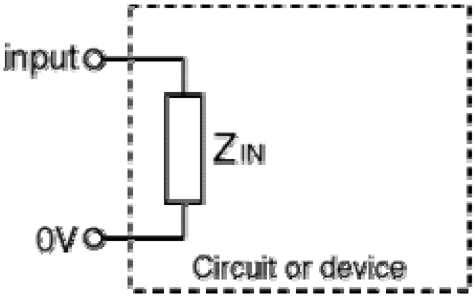

Input impedance (Z_{IN}) is the impedance 'seen' by anything connected to the input of a circuit or device (such as an amplifier). It is the combined effect of all the resistance, capacitance and inductance connected to the input inside the circuit or device.

It is normal to use the term 'input impedance' even for simple cases where there is only resistance and the term 'input resistance' could be used instead. In fact it is usually reasonable to assume that input impedance is just resistance providing the input signal has a low frequency (less than 1kHz say).

The effects of capacitance and inductance vary with frequency, so if these are present the input impedance will vary with frequency. The effects of capacitance and inductance are generally most significant at high frequencies.

Usually **input impedances should be high**, at least ten times the output impedance of the circuit (or component) supplying a signal to the input. This ensures that the input will not 'overload' the source of the signal and reduce the strength (voltage) of the signal by a substantial amount.

Output Impedance Z_{OUT}

The output of any circuit or device is equivalent to output impedance (Z_{OUT}) in series with a perfect voltage source (V_{SOURCE}). This is called the **equivalent circuit** and it represents the combined effect of all the voltage sources, resistance, capacitance and inductance connected to the output inside the circuit or device. Note that V_{SOURCE} is usually **not** the same as the supply voltage Vs.

It is normal to use the term 'output impedance' even for simple cases where there is only resistance and the term 'output resistance' could be used instead. In fact it is usually reasonable to assume that an output impedance is just resistance providing the output signal has a low frequency (less than 1kHz say).

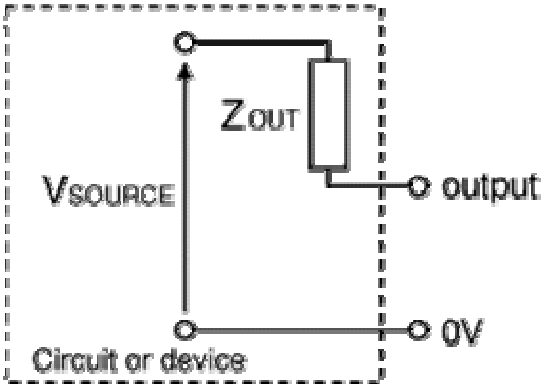

The equivalent circuit of any output

The effects of capacitance and inductance vary with frequency, so if these are present the output impedance will vary with frequency. The effects of capacitance and inductance are generally most significant at high frequencies.

Usually **output impedances should be low**, less than a tenth of the load impedance connected to the output. If an output impedance is too high it will be unable to supply a sufficiently strong signal to the load because most of the signal's voltage will be 'lost' inside the circuit driving current through the output impedance Z_{OUT}. The load could be a single component or the input impedance of another circuit.

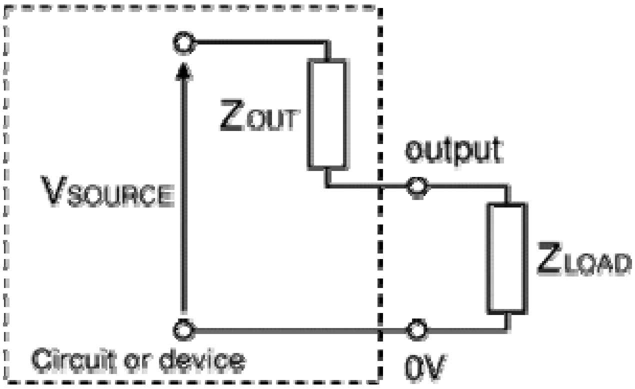

The load can be a single component or the input impedance of another circuit

- **Low output impedance**, $Z_{OUT} << Z_{LOAD}$
  Most of V_{SOURCE} appears across the load, very little voltage is 'lost' driving the output current through the output impedance. Usually this is the best arrangement.
- **Matched impedances**, $Z_{OUT} = Z_{LOAD}$
  Half of V_{SOURCE} appears across the load, the other half is 'lost' driving the output current through the output impedance. This arrangement is useful in some situations (such as an amplifier driving a loudspeaker) because it delivers **maximum power to the load**. Note that an equal amount of power is wasted driving the output current through Z_{OUT}, an efficiency of 50%.
- **High output impedance**, $Z_{OUT} >> Z_{LOAD}$
  Only a small portion of appears across the load, most is 'lost' driving the output current through the output impedance. This arrangement is unsatisfactory.

The output resistance of a voltage divider

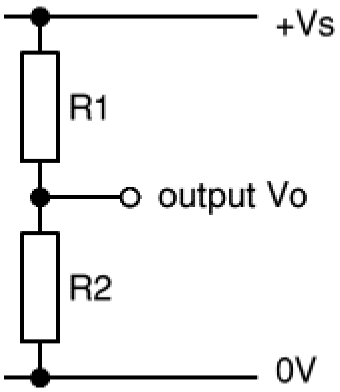

Voltage divider

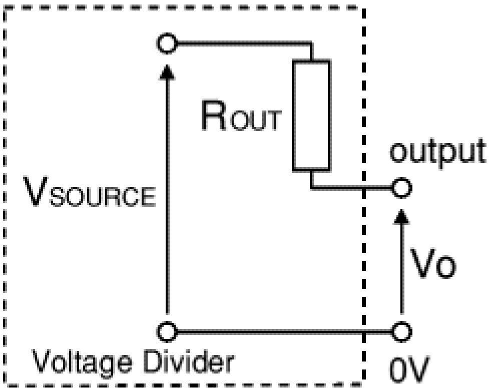

Equivalent circuit of a voltage divider

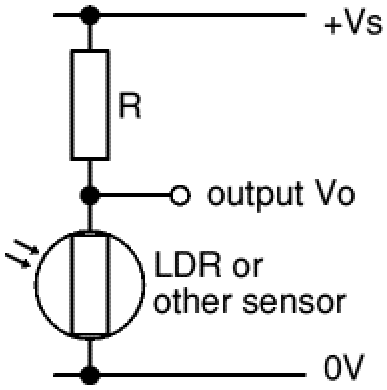

Voltage divider with LDR

Voltage dividers are widely used in electronics, for example to connect an input transducer such as an LDR to a circuit input.

For successful use the output impedance of the voltage divider should be much smaller than the input impedance of the circuit it is connected to. Ideally the output impedance should be less than a tenth of the input impedance.

In the equivalent circuit of a voltage divider the output impedance is just a resistance and the term 'output resistance' could be used. R_{OUT} is equal to the two resistances (R1 and R2) connected in parallel:

$$\text{Output impedance, } \mathbf{R_{OUT}} = \frac{\mathbf{R1 \times R2}}{\mathbf{R1 + R2}}$$

The voltage source V_{SOURCE} in the equivalent circuit is the value of the output voltage Vo when there is nothing connected to the output (and therefore no output current). It is sometimes called the 'open circuit' voltage.

$$\text{Voltage source, } \mathbf{V_{SOURCE}} = \frac{\mathbf{Vs \times R2}}{\mathbf{R1 + R2}}$$

In most voltage dividers one of the resistors will be an input transducer such as an LDR. The transducer's resistance varies and this will make both V_{SOURCE} and R_{OUT} vary too. To check that R_{OUT} is sufficiently low you should work out its highest value which will occur when the transducer has its maximum resistance (this applies wherever the transducer is connected in the voltage divider).

For example: If R1 = 10kΩ and R2 is an LDR with maximum resistance 1MΩ,
R_{OUT} = 10k × 1M / (10k + 1M) = 9.9kΩ (say 10kΩ). This means it should be connected to a load or input resistance of at least 100kΩ.

555 and 556 Timer Circuits

Introduction

The 8-pin 555 timer must be one of the most useful chips ever made and it is used in many projects. With just a few external components it can be used to build many circuits, not all of them involve timing!

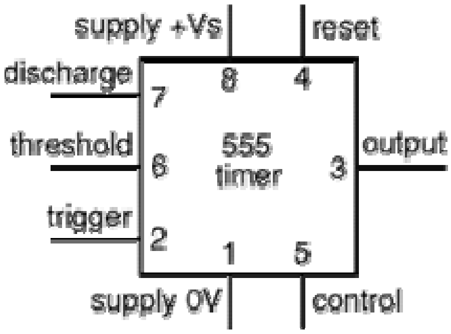
supply +Vs
reset
discharge
7
8
4
threshold
6
555
timer
3
output
trigger
2
1
5
supply 0V
control

Example circuit symbol (above)

Actual pin arrangements (below)

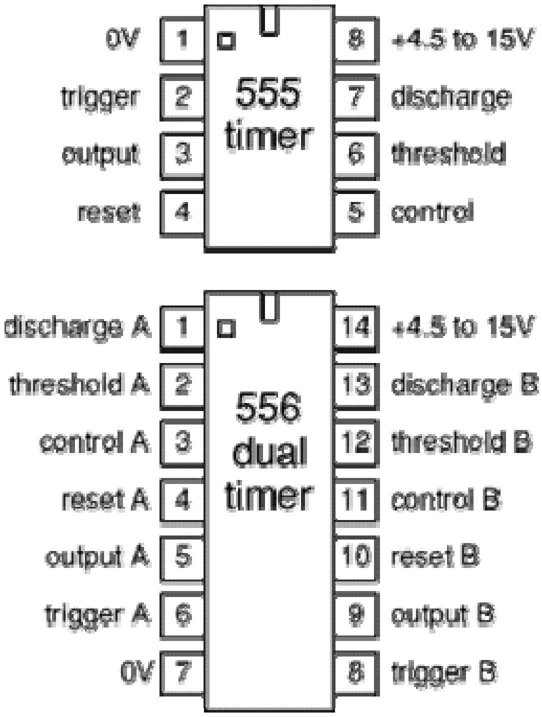
0V
1
trigger
2
output
3
reset
4
555
timer
8
+4.5 to 15V
7
discharge
6
threshold
5
control
discharge A
1
threshold A
2
control A
3
reset A
4
output A
5
trigger A
6
0V
7
556
dual
timer
14
+4.5 to 15V
13
discharge B
12
threshold B
11
control B
10
reset B
9
output B
8
trigger B

A popular version is the NE555 and this is suitable in most cases where a '555 timer' is specified. The 556 is a dual version of the 555 housed in a 14-pin package, the two timers (A and B) share the same power supply pins. The circuit diagrams on this page show a 555, but they could all be adapted to use one half of a 556.

Low power versions of the 555 are made, such as the ICM7555, but these should only be used when specified (to increase battery life) because their maximum output current of about 20mA (with a 9V supply) is too low for many standard 555 circuits. The ICM7555 has the same pin arrangement as a standard 555.

The circuit symbol for a 555 (and 556) is a box with the pins arranged to suit the circuit diagram: for example 555 pin 8 at the top for the +Vs supply, 555 pin 3 output on the right. Usually just the pin numbers are used and they are not labeled with their function.

The 555 and 556 can be used with a supply voltage (Vs) in the range 4.5 to 15V (18V absolute maximum).

Standard 555 and 556 chips create a significant 'glitch' on the supply when their output changes state. This is rarely a problem in simple circuits with no other ICs, but in more complex circuits a **smoothing capacitor** (e.g. 100µF) should be connected across the +Vs and 0V supply near the 555 or 556.

The input and output pin functions are described briefly below and there are fuller explanations covering the various circuits:

- Astable - producing a square wave
- Monostable - producing a single pulse when triggered
- Bistable - a simple memory which can be set and reset
- Buffer - an inverting buffer (Schmitt trigger)

Inputs of 555 / 556

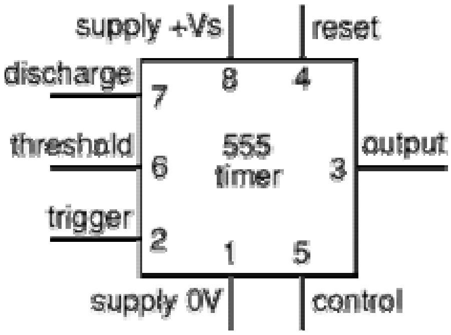

Trigger input: when < $^1/_3$ Vs ('active low') this makes the output high (+Vs). It monitors the discharging of the timing capacitor in an astable circuit. It has a high input impedance > 2MΩ.

Threshold input: when > $^2/_3$ Vs ('active high') this makes the output low (0V)*. It monitors the charging of the timing capacitor in astable and monostable circuits. It has a high input impedance > 10MΩ.
* providing the trigger input is > $^1/_3$ Vs, otherwise the trigger input will override the threshold input and hold the output high (+Vs).

Reset input: when less than about 0.7V ('active low') this makes the output low (0V), overriding other inputs. When not required it should be connected to +Vs. It has an input impedance of about 10kΩ.

Control input: this can be used to adjust the threshold voltage which is set internally to be $^2/_3$ Vs. Usually this function is not required and the control input is connected to 0V with a 0.01µF capacitor to eliminate electrical noise. It can be left unconnected if noise is not a problem.

The **discharge pin** is not an input, but it is listed here for convenience. It is connected to 0V when the timer output is low and is used to discharge the timing capacitor in astable and monostable circuits.

Output of 555 / 556

The output of a standard 555 or 556 can sink and source up to 200mA. This is more than most chips and it is sufficient to supply many output transducers directly, including LEDs (with a resistor in series), low current lamps, piezo transducers, loudspeakers (with a capacitor in series), relay coils (with diode protection) and some motors (with diode protection). The output voltage does not quite reach 0V and +Vs, especially if a large current is flowing.

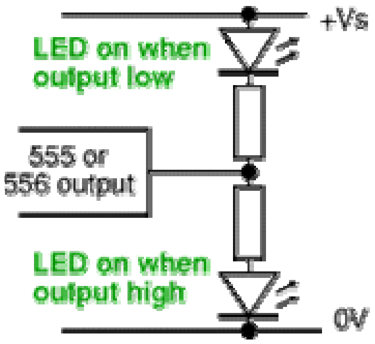
+Vs
LED on when output low
555 or 556 output
LED on when output high
0V

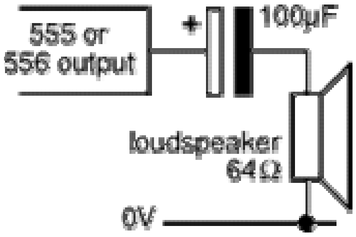
555 or 556 output
+
100µF
loudspeaker 64Ω
0V

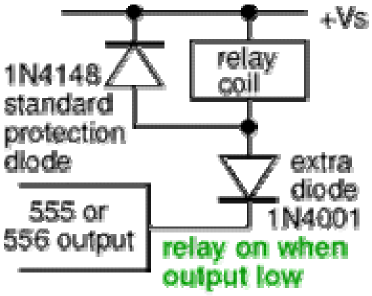
+Vs
1N4148 standard protection diode
relay coil
extra diode 1N4001
555 or 556 output
relay on when output low

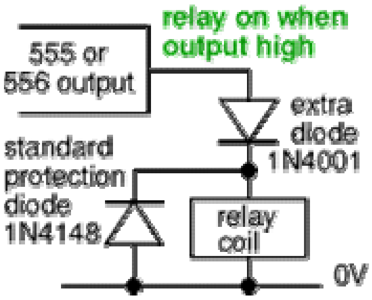
relay on when output high
555 or 556 output
extra diode 1N4001
standard protection diode 1N4148
relay coil
0V

To switch larger currents you can connect a transistor.

The ability to both sink and source current means that two devices can be connected to the output so that one is on when the output is low and the other is on when the output is high. The top diagram shows two LEDs connected in this way. This arrangement is used in the Level Crossing project to make the red LEDs flash alternately.

Loudspeakers

A loudspeaker (minimum resistance 64Ω) may be connected to the output of a 555 or 556 astable circuit but a capacitor (about 100µF) must be connected in series. The output is equivalent to a steady DC of about ½Vs combined with a square wave AC (audio) signal. The capacitor blocks the DC, but allows the AC to pass as explained in capacitor coupling.

Piezo transducers may be connected directly to the output and do not require a capacitor in series.

Relay coils and other inductive loads

Like all ICs, the 555 and 556 must be protected from the brief high voltage 'spike' produced when an inductive load such as a relay coil is switched off. The standard protection diode must be connected 'backwards' across the relay coil as shown in the diagram.

However, **the 555 and 556 require an extra diode** connected in series with the coil to ensure that a small 'glitch' cannot be fed back into the IC. Without this extra diode monostable circuits may re-trigger themselves as the coil is switched off! The coil current passes through the extra diode so it must be a 1N4001 or similar rectifier diode capable of passing the current, a signal diode such as a 1N4148 is usually **not** suitable.

---

555 / 556 Astable

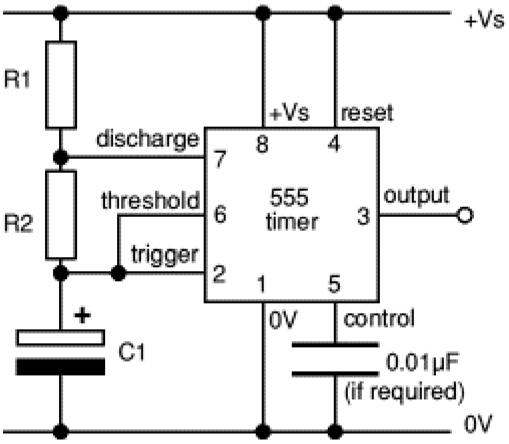

555 Astable Circuit

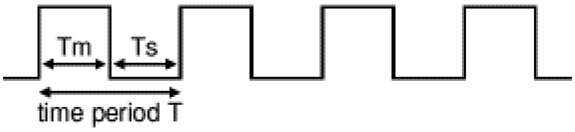

555 Astable Output, a Square Wave
(Tm and Ts may be different)

An astable circuit produces a 'square wave', this is a digital waveform with sharp transitions between low (0V) and high (+Vs). Note that the durations of the low and high states may be different. The circuit is called an astable because it is not stable in any state: the output is continually changing between 'low' and 'high'.

The time period (T) of the square wave is the time for one complete cycle, but it is usually better to consider frequency (f) which is the number of cycles per second.

$$\mathbf{T = 0.7 \times (R1 + 2R2) \times C1} \quad \text{and} \quad \mathbf{f = \frac{1.4}{(R1 + 2R2) \times C1}}$$

T = time period in seconds (**s**)
f = frequency in hertz (**Hz**)
R1 = resistance in ohms (**Ω**)
R2 = resistance in ohms (**Ω**)
C1 = capacitance in farads (F)

The time period can be split into two parts: **T = Tm + Ts**
Mark time (output high): **Tm = 0.7 × (R1 + R2) × C1**
Space time (output low): **Ts = 0.7 × R2 × C1**

Many circuits require Tm and Ts to be almost equal; this is achieved if R2 is much larger than R1.

For a standard astable circuit Tm cannot be less than Ts, but this is not too restricting because the output can both sink and source current. For example an LED can be made to flash briefly with long gaps by connecting it (with its resistor) between +Vs and the output. This way the LED is on during Ts, so brief flashes are achieved with R1 larger than R2, making Ts short and Tm long. If Tm must be less than Ts a diode can be added to the circuit as explained under duty cycle below.

Choosing R1, R2 and C1

R1 and R2 should be in the range 1kΩ to 1MΩ. It is best to choose C1 first because capacitors are available in just a few values.

555 astable frequencies			
C1	**R2 = 10kΩ** **R1 = 1kΩ**	**R2 = 100kΩ** **R1 = 10kΩ**	**R2 = 1MΩ** **R1 = 100kΩ**
0.001µF	68kHz	6.8kHz	680Hz
0.01µF	6.8kHz	680Hz	68Hz
0.1µF	680Hz	68Hz	6.8Hz
1µF	68Hz	6.8Hz	0.68Hz
10µF	6.8Hz	0.68Hz (41 per min.)	0.068Hz (4 per min.)

- **Choose C1** to suit the frequency range you require (use the table as a guide).
- **Choose R2** to give the frequency (f) you requires. Assume that R1 is much smaller than R2 (so that Tm and Ts are almost equal), then you can use:

$$\mathbf{R2 = \frac{0.7}{f \times C1}}$$

- **Choose R1** to be about a tenth of R2 (1kΩ min.) unless you want the mark time Tm to be significantly longer than the space time Ts.
- If you wish to use a **variable resistor** it is best to make it R2.
- If R1 is variable it must have a fixed resistor of at least 1kΩ in series (this is not required for R2 if it is variable).

Astable operation

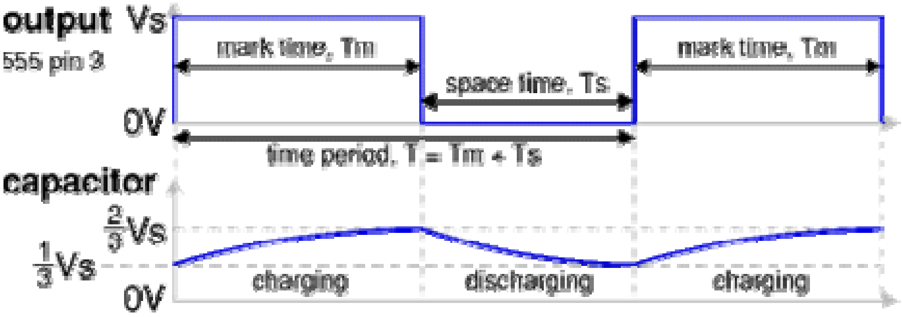

With the output high (+Vs) the capacitor C1 is charged by current flowing through R1 and R2. The threshold and trigger inputs monitor the capacitor voltage and when it reaches $^{2}/_{3}$Vs (threshold voltage) the output becomes low and the discharge pin is connected to 0V.

The capacitor now discharges with current flowing through R2 into the discharge pin. When the voltage falls to $^{1}/_{3}$Vs (trigger voltage) the output becomes high again and the discharge pin is disconnected, allowing the capacitor to start charging again.

This cycle repeats continuously unless the reset input is connected to 0V which forces the output low while reset is 0V.

An astable can be used to provide the **clock signal** for circuits such as counters.

A **low frequency** astable (< 10Hz) can be used to flash an LED on and off, higher frequency flashes are too fast to be seen clearly. Driving a loudspeaker or piezo transducer with a low frequency of less than 20Hz will produce a series of 'clicks' (one for each low/high transition) and this can be used to make a simple metronome.

An **audio frequency** astable (20Hz to 20kHz) can be used to produce a sound from a loudspeaker or piezo transducer. The sound is suitable for buzzes and beeps. The natural (resonant) frequency of most piezo transducers is about 3kHz and this will make them produce a particularly loud sound.

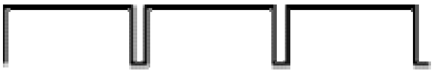

90% duty cycle (Tm = 9Ts)

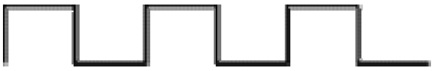

50% duty cycle (Tm = Ts)

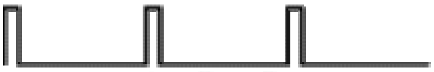

10% duty cycle (9Tm = Ts)

Duty Cycle

The duty cycle of an astable circuit is the proportion of the complete cycle for which the output is high (the mark time). It is usually given as a percentage.

For a standard 555/556 astable circuit the mark time (Tm) must be greater than the space time (Ts), so the duty cycle must be at least 50%:

$$\text{Duty cycle} = \frac{\mathbf{Tm}}{\mathbf{Tm + Ts}} = \frac{\mathbf{R1 + R2}}{\mathbf{R1 + 2R2}}$$

To achieve a duty cycle of less than 50% a diode can be added in parallel with R2 as shown in the diagram below. This bypasses R2 during the charging (mark) part of the cycle so that Tm depends only on R1 and C1:

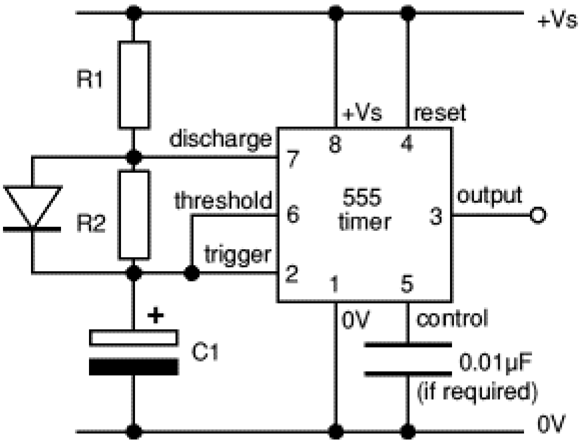

555 Astable Circuit with Diode across R2

Tm = 0.7 × R1 × C1 (ignoring 0.7V across diode)
Ts = 0.7 × R2 × C1 (unchanged)

Duty cycle with diode = $\frac{\mathbf{Tm}}{\mathbf{Tm + Ts}} = \frac{\mathbf{R1}}{\mathbf{R1 + R2}}$

---

555 / 556 Monostable

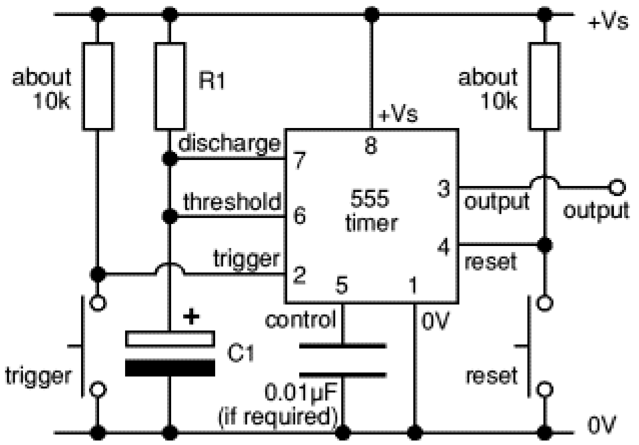
+Vs
about 10k
R1
+Vs
about 10k
discharge
8
7
threshold
6
555 timer
3
output
output
trigger
2
4
reset
5
1
+
control
0V
C1
trigger
reset
0.01µF (if required)
0V

555 Monostable Circuit with manual trigger

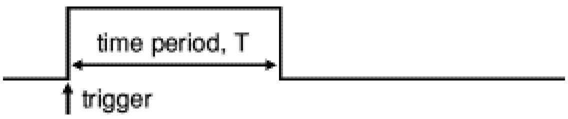
time period, T
trigger

555 Monostable Output, a single pulse

A monostable circuit produces a single output pulse when triggered. It is called a monostable because it is stable in just **one** state: 'output low'. The 'output high' state is temporary.

The duration of the pulse is called the **time period** (T) and this is determined by resistor R1 and capacitor C1:

time period, **T = 1.1 × R1 × C1**

T = time period in seconds (s)
R1 = resistance in ohms (Ω)
C1 = capacitance in farads (F)
The maximum reliable time period is about 10 minutes.

Why 1.1? The capacitor charges to $^{2}/_{3}$ = 67% so it is a bit longer than the time constant (R1 × C1) which is the time taken to charge to 63%.

- **Choose C1 first** (there are relatively few values available).
- **Choose R1** to give the time period you need. R1 should be in the range 1kΩ to 1MΩ, so use a fixed resistor of at least 1kΩ in series if R1 is variable.
- **Beware** that electrolytic capacitor values are not accurate, errors of at least 20% are common.
- **Beware** that electrolytic capacitors leak charge which substantially increases the time period if you are using a high value resistor - use the formula as only a very rough guide! *For example the Timer Project should have a maximum time period of 266s (about 4½ minutes), but many electrolytic capacitors extend this to about 10 minutes!*

Monostable Operation

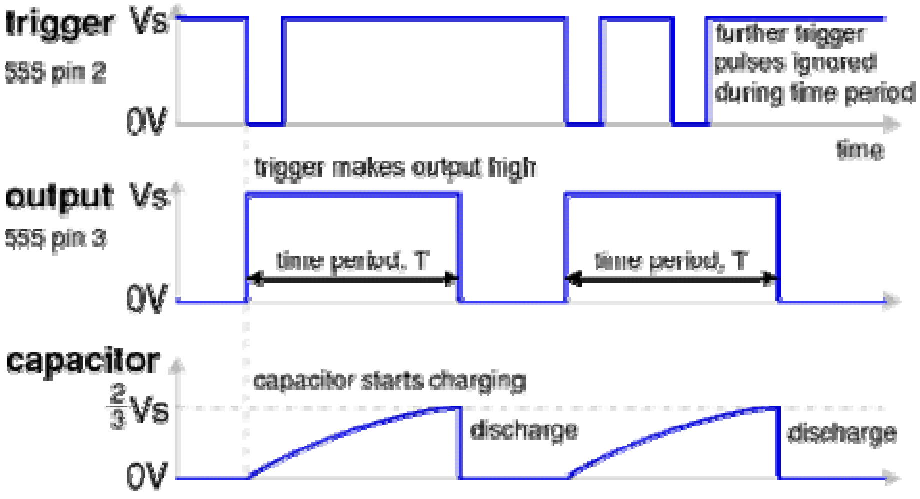

The timing period is triggered (started) when the **trigger** input (555 pin 2) is less than $^{1}/_{3}$ Vs, this makes the **output** high (+Vs) and the capacitor C1 starts to charge through resistor R1. Once the time period has started further trigger pulses are ignored.

The **threshold** input (555 pin 6) monitors the voltage across C1 and when this reaches $^{2}/_{3}$ Vs the time period is over and the **output** becomes low. At the same time **discharge** (555 pin 7) is connected to 0V, discharging the capacitor ready for the next trigger.

The **reset** input (555 pin 4) overrides all other inputs and the timing may be cancelled at any time by connecting reset to 0V, this instantly makes the output low and discharges the capacitor. If the reset function is not required the reset pin should be connected to +Vs.

Power-On Reset or Trigger

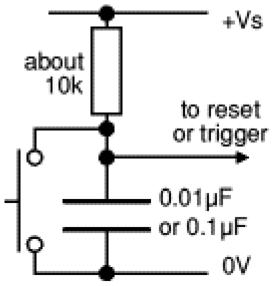

Power-On Reset or Trigger Circuit

It may be useful to ensure that a monostable circuit is reset or triggered automatically when the power supply is connected or switched on. This is achieved by using a capacitor instead of (or in addition to) a push switch as shown in the diagram.

The capacitor takes a short time to charge, briefly holding the input close to 0V when the circuit is switched on. A switch may be connected in parallel with the capacitor if manual operation is also required.

Edge-Triggering

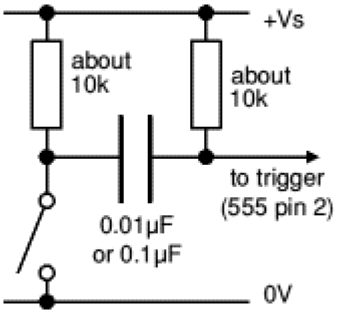
+Vs
about
10k
about
10k
to trigger
(555 pin 2)
0.01μF
or 0.1μF
0V

Edge-Triggering Circuit

If the trigger input is still less than $^1/_3$ Vs at the end of the time period the output will remain high until the trigger is greater than $^1/_3$ Vs. This situation can occur if the input signal is from an on-off switch or sensor.

The monostable can be made **edge triggered**, responding only to **changes** of an input signal, by connecting the trigger signal through a capacitor to the trigger input. The capacitor passes sudden changes (AC) but blocks a constant (DC) signal. For further information please see the page on capacitance. The circuit is 'negative edge triggered' because it responds to a sudden fall in the input signal.

The resistor between the trigger (555 pin 2) and +Vs ensures that the trigger is normally high (+Vs).

---

555 / 556 Bistable (flip-flop) - a Memory Circuit

The circuit is called a bistable because it is stable in **two** states: output high and output low. It is also known as a 'flip-flop'.

It has two inputs:

- **Trigger** (555 pin 2) makes the **output high**.
  Trigger is 'active low', it functions when < $^1/_3$ Vs.
- **Reset** (555 pin 4) makes the **output low**.
  Reset is 'active low', it resets when < 0.7V.

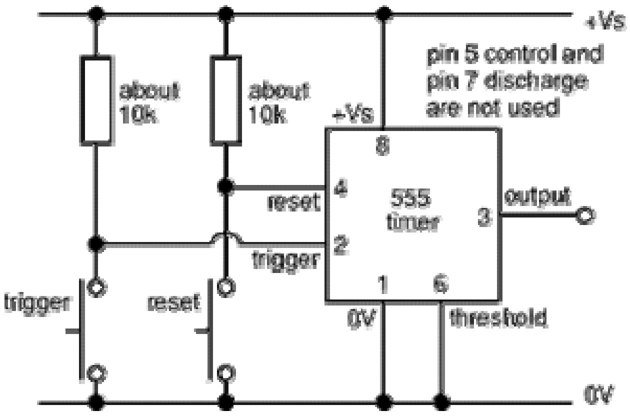
+Vs
pin 5 control and
pin 7 discharge
are not used
about
10k
about
10k
+Vs
8
reset
4
555
timer
3
output
trigger
2
1
6
trigger
reset
0V
threshold
0V

555 Bistable Circuit

The power-on reset, power-on trigger and edge-triggering circuits can all be used as described above for the monostable.

---

555 / 556 Inverting Buffer (Schmitt trigger) or NOT gate

The buffer circuit's input has a very high impedance (about 1MΩ) so it requires only a few µA, but the output can sink or source up to 200mA. This enables a high impedance signal source (such as an LDR) to switch a low impedance output transducer (such as a lamp).

It is an **inverting buffer** or NOT gate because the output logic state (low/high) is the inverse of the input state:

- **Input low** (< $^1/_3$ Vs) makes **output high**, +Vs
- **Input high** (> $^2/_3$ Vs) makes **output low**, 0V

When the input voltage is between $^1/_3$ and $^2/_3$ Vs the output remains in its present state. This intermediate input region is a deadspace where there is no response, a property called **hysteresis**, it is like backlash in a mechanical linkage. This type of circuit is called a **Schmitt trigger**.

If high sensitivity is required the hysteresis is a problem, but in many circuits it is a helpful property. It gives the input a high immunity to noise because once the circuit output has switched high or low the input must change back by at least $^1/_3$ Vs to make the output switch back.

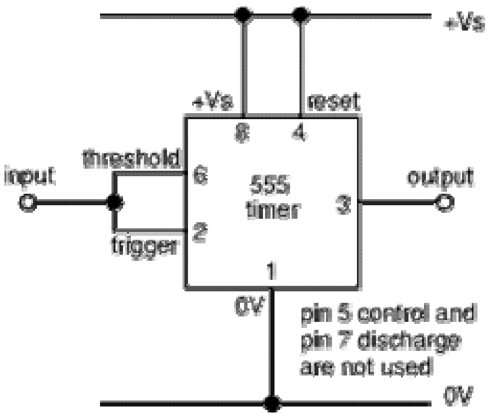

555 inverting buffer circuit (NOT gate)

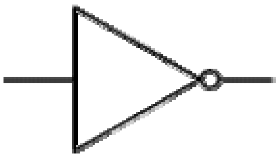

NOT gate symbol

Counting Circuits

Binary Numbers

Electronic circuits count in binary. This is the simplest possible counting system because it uses just two digits, 0 and 1, exactly like logic signals where 0 represents false and 1 represents true. The terms low and high are also used for 0 and 1 respectively as shown in the table.

Logic states	
True	**False**
1	0
High	Low
+Vs	0V
On	Off

Counting **one**, **two**, **three**, **four**, **five** in binary: 1, 10, 11, 100, 101.

Binary numbers rapidly become very long as the count increases and this makes them difficult for us to read at a glance. Fortunately it is rarely necessary to read more than 4 binary digits at a time in counting circuits.

In a binary number each digit represents a multiple of two (1, 2, 4, 8, 16 etc), in the same way that each digit in decimal represents a multiple of ten (1, 10, 100, 1000 etc).
For example 10110110 in binary equals 182 in decimal:

Digit value:	128	64	32	16	8	4	2	1
Binary number:	**1**	**0**	**1**	**1**	**0**	**1**	**1**	**0**
Decimal value:	128 +	0	+ 32 +	16 +	0 +	4 +	2 +	0 = 182

Bits, Bytes and Nibbles

Each binary digit is called a **bit**, so 10110110 is an 8-bit number.

A block of **8 bits** is called a **byte** and it can hold a maximum number of 11111111 = 255 in decimal. Computers and PIC microcontrollers work with blocks of 8 bits. Two (or more) bytes make a **word**, for example PICs work with a 16-bit word (two bytes) which can hold a maximum number of 65535.

A block of **4 bits** is called a **nibble** (half a byte!) and it can hold a maximum number of 1111 = 15 in decimal. Many counting circuits work with blocks of 4 bits because this number of bits is required to count up to 9 in decimal. (The maximum number with 3 bits is only 7).

Hexadecimal (base 16)

Hexadecimal (often just called 'hex') is base 16 counting with 16 digits. It starts with the decimal digits 0-9, then continues with letters A (10), B (11), C (12), D (13), E (14) and F (15). Each hexadecimal digit is equivalent to 4 binary digits, making conversion between the two systems relatively easy. You may find hexadecimal used with PICs and computer systems but it is not generally used in simple counting circuits.

Example: **10110110** binary = **B6** hexadecimal = **182** decimal.

4-bit Numbers

The table on the right shows the 4-bit numbers and their decimal values.

The labels A, B, C, D are widely used in electronics to represent the four bits:

- A = 1, the 'least significant bit' (LSB)
- B = 2
- C = 4
- D = 8, the 'most significant bit' (MSB)

Binary D C B A	**Decimal**	**Hex** base 16
0 0 0 0	0	0
0 0 0 1	1	1
0 0 1 0	2	2
0 0 1 1	3	3
0 1 0 0	4	4
0 1 0 1	5	5
0 1 1 0	6	6
0 1 1 1	7	7
1 0 0 0	8	8
1 0 0 1	9	9
1 0 1 0	10	A
1 0 1 1	11	B
1 1 0 0	12	C
1 1 0 1	13	D
1 1 1 0	14	E
1 1 1 1	15	F

Binary Coded Decimal, BCD

Binary Coded Decimal, BCD, is a special version of 4-bit binary where the count resets to zero (0000) after the ninth count (1001). It is used by **decade** counters and is easily converted to display the decimal digits 0-9 on a 7-segment display.

Several decade counters using BCD can be linked together to separately count the decimal ones, tens, hundreds, and so on. This is much easier than attempting to convert large binary numbers (such as 10110110) to display their decimal value.

Do not confuse BCD which stands for Binary Coded Decimal with the labels A, B, C, D used to represent the four binary digits; it is an unfortunate coincidence that the letters BCD occur in both!

Counters

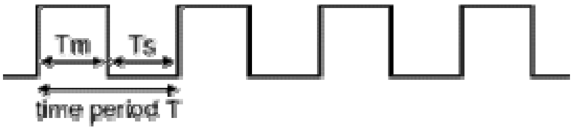

A square wave clock signal

All counters require a 'square wave' **clock** signal to make them count. This is a digital waveform with sharp transitions between low (0V) and high (+Vs), such as the output from a 555 astable circuit.

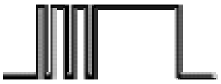

The bouncing output from a switch

Most switches **bounce** when the contacts close giving a rapid series of pulses. Connecting a switch directly to a clock input will usually give several counts when the switch is operated once! One way to 'debounce' the switch is to make it trigger a 555 monostable circuit with a short time period (such as 0.1s) and use the monostable output to drive the clock input.

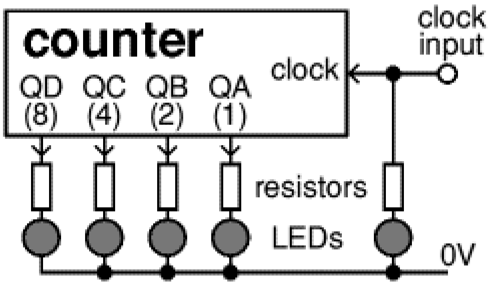

A 4-bit counter and clock input.
In this example counting advances on the falling-edge of the clock signal
LED on = 1 LED off = 0

The animated block diagram shows a clock signal driving a 4-bit (0-15) counter with LEDs connected to show the state of the clock and counter outputs QA-QD (Q indicates an output).

The LED on the first output QA flashes at half the frequency of the clock LED. In fact the frequency of each stage of the counter is half the frequency of the previous stage. You can see this pattern too in the table above showing the 4-bit numbers.

Notice how output QA changes state every time the clock input changes from high to low (that is when the clock LED turns off), this is called the **falling-edge**. If you watch the counting closely you can see that QB changes on the falling-edge of QA, QC on the falling-edge of QB and so on.

You may be surprised to see the diagram drawn with the input on the right and signals flowing from right to left, the opposite way to the usual convention in electronics! Drawing counter circuits like this means that the outputs are in the correct binary order for us to read easily and I think this is more helpful than rigidly sticking to the usual 'left to right' convention.

Ripple and Synchronous Counters

There are two main types of counter: **ripple** and **synchronous**. In simple circuits their behavior appears almost identical, but their internal structure is very different.

A ripple counter contains a chain of **flip-flops** with the output of each one feeding the input of the next. A flip-flop output changes state every time the input changes from high to low (on the falling-edge). This simple arrangement works well, but there is a slight delay as the effect of the clock 'ripples' through the chain of flip-flops.

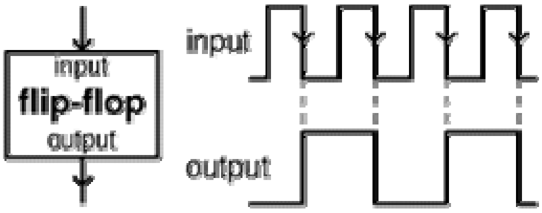

The operation of a flip-flop.
Notice how the output frequency is half the input frequency

In most circuits the **ripple delay** is not a problem because it is far too short to be seen on a display. However, a logic system connected to ripple counter outputs will briefly see false counts which may produce 'glitches' in the logic system and may disrupt its operation. For example a ripple counter changing from 0111 (7) to 1000 (8) will very briefly show 0110, 0100 and 0000 before 1000!

A synchronous counter has a more complex internal structure to ensure that all its outputs change precisely together on each clock pulse, avoiding the brief false counts which occur with ripple counters.

Rising-edge and Falling-edge clock inputs

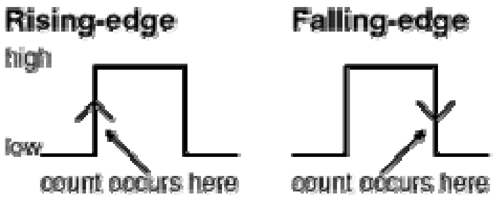
Rising-edge
high
low
count occurs here
Falling-edge
count occurs here

Counting occurs when the clock input changes state.

- Most **synchronous counters** count on the **rising-edge** which is the low to high transition of the clock signal.
- Most **ripple counters** count on the **falling-edge** which is the high to low transition of the clock signal.

It may seem odd that ripple counters use the falling-edge, but in fact this makes it easy to link counters because the most significant bit (MSB) of one counter can drive the clock input of the next. This works because the next bit must change state when the previous bit changes from high to low - the point at which a carry must occur to the next bit. Synchronous counters usually have **carry out** and **carry in** pins for linking counters without introducing any ripple delays.

Resetting a Counter

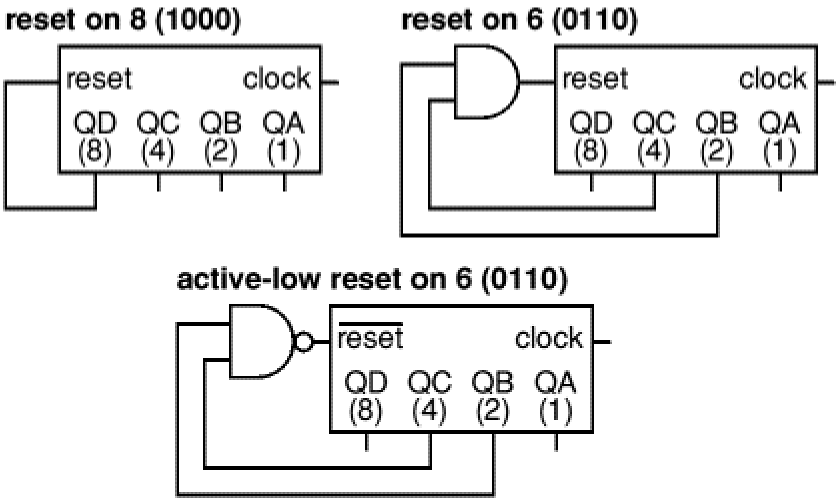
reset on 8 (1000)
reset
clock
QD (8)
QC (4)
QB (2)
QA (1)
reset on 6 (0110)
reset
clock
QD (8)
QC (4)
QB (2)
QA (1)
active-low reset on 6 (0110)
reset
clock
QD (8)
QC (4)
QB (2)
QA (1)

Counters can be reset to zero before their maximum count by connecting one (or more) of their outputs to their reset input, using an AND gate to combine outputs if necessary.

If the reset input is 'active-low' a NOT or NAND gate will be required to produce a low output at the desired count. If you see a line drawn above reset it means it is active low, for example: $\overline{\text{reset}}$ (say 'reset-bar').

The reset function normally occurs immediately and you should reset on the next count above the maximum you require. For example to count 0-5 (0000-0101) you should reset on 6 (0110).

Some synchronous counters have a **synchronous reset** which occurs on the next clock pulse rather than immediately. This is important because you must reset on the maximum count you require. For example to count 0-5 (0000-0101), reset on 5 (0101).

Presetting

Some counters can be preset by presenting a number to their inputs A-D and activating a preset input to load the number into the counter. By making inputs A-D all low you can also use this to reset the counter to zero.

Frequency division

Counters can be used to reduce the frequency of an input (clock) signal. Each stage of a counter halves the frequency, so for a 4-bit (0-15) counter QA is $^1/_2$, QB is $^1/_4$, QC is $^1/_8$ and QD is $^1/_{16}$ of the clock frequency. Division by numbers that are not powers of 2 is possible by resetting counters.

Frequency division is one of the main purposes of counters with more than 4 bits and their outputs are usually labeled Q1, Q2 and so on. **Qn** is the nth stage of the counter, representing 2^n. For example Q4 is $2^4 = 16$ ($^1/_{16}$ of clock frequency) and Q12 is $2^{12} = 4096$ ($^1/_{4096}$ of clock frequency).

Decoders

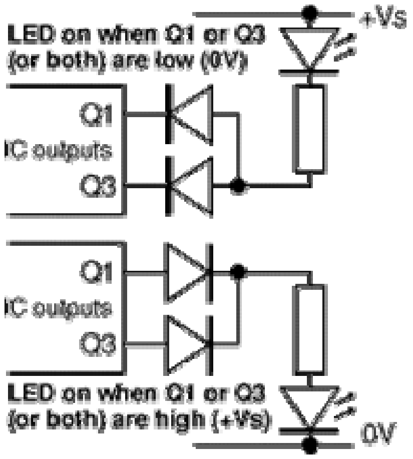

The most popular type is a **1-of-10 decoder** which contains a network of logic gates to make one of its ten **outputs Q0-9** become high (or low) in response to the BCD (binary coded decimal) **inputs A-D**. For example an input of binary 0101 (=5) will activate output Q5.

Decoders can be used for a simple counting display and for switching LEDs in sequences. The outputs must never be directly connected together, but diodes can be used to combine them as shown in the diagram.

For example using diodes to combine the 2nd (Q1) and 4th (Q3) outputs will make an LED flash twice followed by a longer gap. The top diagram shows this for a decoder where the outputs become **low** when activated (such as the 7442), and the bottom diagram for a decoder where the outputs become **high** when activated (such as the 4028).

7-Segment Display Drivers

The **inputs A-D** of a display driver are connected to the BCD (binary coded decimal) outputs QA-D from a decade counter. A network of logic gates inside the display driver makes its **outputs a-g** become high or low as appropriate to light the required segments a-g of a 7-segment display. A resistor is required in series with each segment to protect the LEDs, 330Ω is a suitable value for many displays with a 4.5V to 6V supply. *Beware that these resistors are sometimes omitted from circuit diagrams!*

There are two types of 7-segment displays:

- **Common Anode** (CA or SA) with all the LED anodes connected together. These need a display driver with outputs which become **low** to light each segment, for example the 7447. Connect the common anode to +Vs.
- **Common Cathode** (CC or SC) with all the cathodes connected together. These need a display driver with outputs which become **high** to light each segment, for example the 4511. Connect the common cathode to 0V.

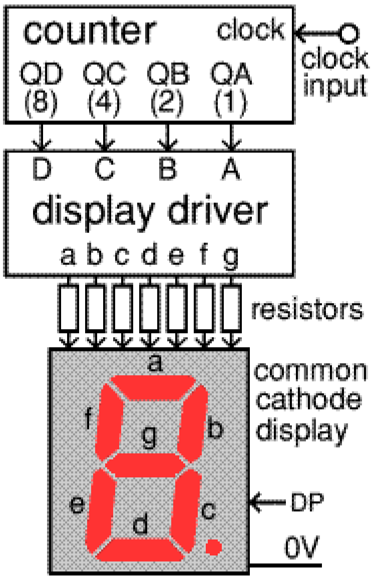

Decade counter with display driver and 7-segment display

The common anode/cathode is often available on 2 pins. Displays also have a decimal point (DP) but this is not controlled by the display driver. The segments of larger displays have two LEDs in series. For display connections please see your supplier's catalogue or manufacturer's datasheet.

Multiplexing

If there are many 7-segment display digits **multiplexing** is usually used. This is a system of switching so that of all the decade counters share a single display driver which is connected to all of the displays. The output of each counter is connected in turn to the inputs of the display driver and at the same time the common anode/cathode of the corresponding 7-segment display is connected so that only one display lights at a time. The switching is done very rapidly (typically 400 - 1000Hz) and the segment current is larger than normal so the display appears continuous and of normal brightness. Multiplexing requires ICs to do the switching, but the complete circuit has fewer ICs than having one display driver for each display.

Linking Counters

Counters may be linked together in a chain to count larger numbers. It may seem tempting to use a 12-bit or 14-bit counter, but it is not practical to convert their large binary numbers to decimal. You should use a chain of **decade (0-9)** counters which use BCD (binary coded decimal) to make the conversion to decimal very easy: the first counts the units, the second counts the tens, the third the hundreds and so on.

Some **dual counter** ICs are available with two separate counters on the same IC, the two counters must be linked externally if required (there is no internal link).

The way that counters are linked depends on the nature of the counter. The diagrams below show the general arrangements for standard ripple and synchronous counters but it is important to read the detailed information for particular counters, consulting a datasheet if necessary.

Linking Ripple Counters

The diagram below shows how to link standard ripple counters. Notice how the highest output QD of each counter drives the **clock (CK)** input of the next counter. This works because ripple counters have clock inputs that are 'active-low' which means that the count advances as the clock input becomes low, on the falling-edge.

Remember that with all ripple counters there will be a slight delay before the later outputs respond to the clock signal, especially with a long counter chain. This is not a problem in simple circuits driving displays, but it may cause glitches in logic systems connected to the counter outputs.

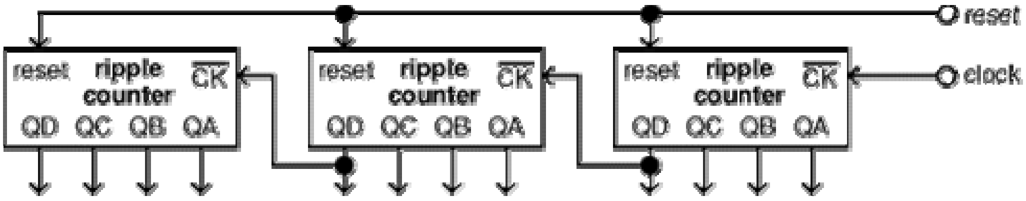
reset
reset
ripple
counter
$\overline{CK}$
QD QC QB QA
reset
ripple
counter
$\overline{CK}$
QD QC QB QA
reset
ripple
counter
$\overline{CK}$
QD QC QB QA
clock

Linking Synchronous Counters

The diagram below shows how to link standard synchronous counters. Notice how all the **clock (CK)** inputs are linked, and **carry out (CO)** is used to feed the **carry in (CI)** of the next counter. This ensures that the entire counter chain is synchronous, with every output changing at the same time. **Carry in (CI)** of the first counter should be made low or high to suit the particular counter IC being used.

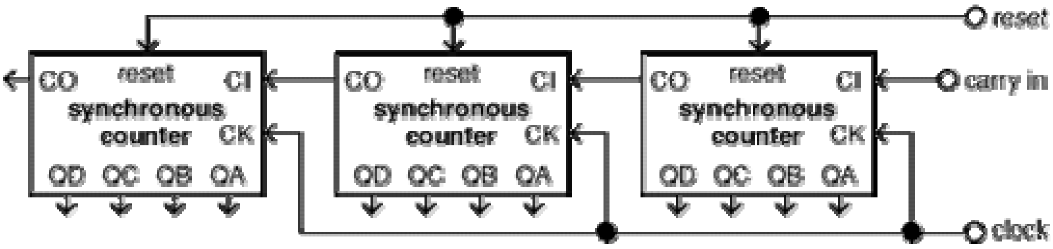
reset
CO
reset
CI
synchronous
counter
CK
QD QC QB QA
CO
reset
CI
synchronous
counter
CK
QD QC QB QA
CO
reset
CI
synchronous
counter
CK
QD QC QB QA
carry in
clock

Quantities and Units in Electronics

Quantities

The table shows electrical quantities which are used in electronics.

The relationship between quantities can be written using words or symbols (letters), but symbols are normally used because they are much shorter; for example V is used for voltage, I for current and R for resistance:

As a word equation:

voltage = current × resistance

The same equation using symbols: **V = I × R**

To prevent confusion we normally use the same symbol (letter) for each quantity and these symbols are shown in the second column of the table.

Quantity	Usual Symbol	Unit	Unit Symbol
Voltage	V	volt	V
Current	I	amp*	A
Charge	Q	coulomb	C
Resistance	R	ohm	Ω
Capacitance	C	farad	F
Inductance	L	henry	H
Reactance	X	ohm	Ω
Impedance	Z	ohm	Ω
Power	P	watt	W
Energy	E	joule	J
Time	t	second	s
Frequency	f	hertz	Hz

* strictly the unit is ampere, but this is almost always shortened to amp.

Units

The first table shows the unit (and unit symbol) which is used to measure each quantity. For example: Charge is measured in coulombs and the symbol for a coulomb is C.

Some of the units have a convenient size for electronics, but most are either too large or too small to be used directly so they are used with the prefixes shown in the second table. The prefixes make the unit larger or smaller by the value shown.

Prefix	Prefix Symbol	Value	
milli	m	10^{-3}	= 0.001
micro	µ	10^{-6}	= 0.000 001
nano	n	10^{-9}	= 0.000 000 001
pico	p	10^{-12}	= 0.000 000 000 001
kilo	k	10^{3}	= 1000
mega	M	10^{6}	= 1000 000
giga	G	10^{9}	= 1000 000 000
tera	T	10^{12}	= 1000 000 000 000

Some examples:
25 mA = 25 × 10^{-3} A = 25 × 0.001 A = 0.025 A
47µF = 47 × 10^{-6} F = 47 × 0.000 001 F = 0.000 047 F
270kΩ = 270 × 10^{3} Ω= 270 × 1000 Ω= 270 000 Ω

Why not change the units to be better sizes?

It might seem a good idea to make the farad (F) much smaller to avoid having to use µF, nF and pF, but if we did this most of the equations in electronics would have to have factors of 1000000 or more included as well as the quantities. Overall it is much better to have the units with their present sizes which are defined logically from the equations.

In fact if you use an equation frequently you can use special sets of prefixed units which are more convenient...

For example: **Ohm's Law, V = I × R**
  the standard units are volt (V), amp (A) and ohm (Ω),
  but you could use volt (V), milliamp (mA) and kilo-ohm (**kΩ**) if you prefer.
Take care though; you must never mix sets of units: using **V**, **A** and **kΩ** in Ohm's Law would give you wrong values.

Capacitors

Function

Capacitors store electric charge. They are used with resistors in timing circuits because it takes time for a capacitor to fill with charge. They are used to smooth varying DC supplies by acting as a reservoir of charge. They are also used in filter circuits because capacitors easily pass AC (changing) signals but they block DC (constant) signals.

Capacitance

This is a measure of a capacitor's ability to store charge. A large capacitance means that more charge can be stored. Capacitance is measured in farads, symbol F. However 1F is very large, so prefixes are used to show the smaller values.

Three prefixes (multipliers) are used, µ (micro), n (nano) and p (pico):

- µ means 10^{-6} (millionth), so 1000000µF = 1F
- n means 10^{-9} (thousand-millionth), so 1000nF = 1µF
- p means 10^{-12} (million-millionth), so 1000pF = 1nF

Capacitor values can be very difficult to find because there are many types of capacitor with different labeling systems!

There are many types of capacitor but they can be split into two groups, **polarized** and **unpolarized**. Each group has its own circuit symbol.

Polarized Capacitors (large values, 1µF +)

Examples:

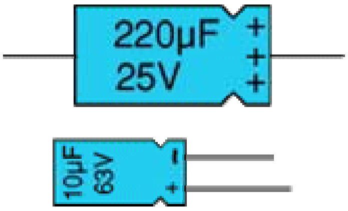

Circuit symbol:

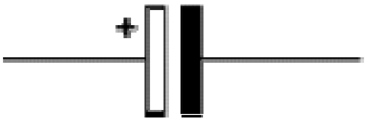

Electrolytic Capacitors

Electrolytic capacitors are polarized and **they must be connected the correct way round**, at least one of their leads will be marked **+** or **-**. They are not damaged by heat when soldering.

There are two designs of electrolytic capacitors; **axial** where the leads are attached to each end (220µF in picture) and **radial** where both leads are at the same end (10µF in picture). Radial capacitors tend to be a little smaller and they stand upright on the circuit board.

It is easy to find the value of electrolytic capacitors because they are clearly printed with their capacitance and voltage rating. The voltage rating can be quite low (6V for example) and it should always be checked when selecting an electrolytic capacitor. It the project parts list does not specify a voltage; choose a capacitor with a rating which is greater than the project's power supply voltage. 25V is a sensible minimum for most battery circuits.

Tantalum Bead Capacitors

Tantalum bead capacitors are polarized and have low voltage ratings like electrolytic capacitors. They are expensive but very small, so they are used where a large capacitance is needed in a small size.

Modern tantalum bead capacitors are printed with their capacitance, voltage and polarity in full. However older ones use a color-code system which has two stripes (for the two digits) and a spot of color for the number of zeros to give the value in µF. The standard color code is used, but for the spot, **grey** is used to mean × 0.01 and **white** means × 0.1 so that values of less than 10µF can be shown. A third color stripe near the leads shows the voltage (yellow 6.3V, black 10V, green 16V, blue 20V, grey 25V, white 30V, pink 35V). The positive (+) lead is to the right when the spot is facing you: '**when the spot is in sight, the positive is to the right**'.

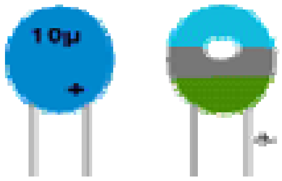

For example: **blue, grey, black spot** means 68µF
For example: **blue, grey, white spot** means 6.8µF
For example: **blue, grey, grey spot** means 0.68µF

Unpolarized Capacitors (small values, up to 1µF)

Examples: 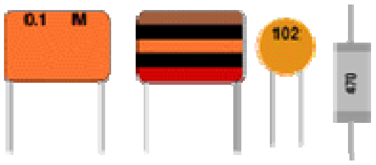

Circuit symbol: 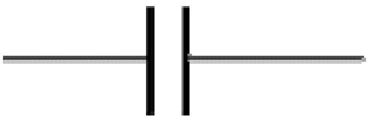

Small value capacitors are unpolarized and may be connected either way round. They are not damaged by heat when soldering, except for one unusual type (polystyrene). They have high voltage ratings of at least 50V, usually 250V or so. It can be difficult to find the values of these small capacitors because there are many types of them and several different labeling systems!

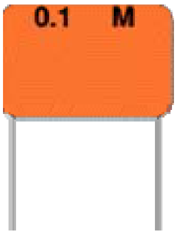

Many small value capacitors have their value printed but without a multiplier, so you need to use experience to work out what the multiplier should be!

For example **0.1** means 0.1µF = 100nF.

Sometimes the multiplier is used in place of the decimal point:
For example: **4n7** means 4.7nF.

Capacitor Number Code

A number code is often used on small capacitors where printing is difficult:

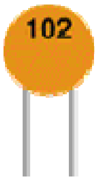

- the 1st number is the 1st digit,
- the 2nd number is the 2nd digit,
- the 3rd number is the number of zeros to give the capacitance in pF.
- Ignore any letters - they just indicate tolerance and voltage rating.

For example: **102** means 1000pF = 1nF *(not 102pF)*

For example: **472J** means 4700pF = 4.7nF (J means 5% tolerance).

Capacitor Color Code

A color code was used on polyester capacitors for many years. It is now obsolete, but of course there are many still around. The colors should be read like the resistor code, the top three color bands giving the value in pF. Ignore the 4th band (tolerance) and 5th band (voltage rating).

Color Code	
Color	Number
Black	0
Brown	1
Red	2
Orange	3
Yellow	4
Green	5
Blue	6
Violet	7
Grey	8
White	9

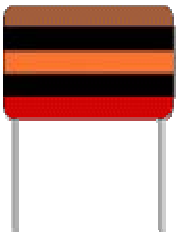

For example:

brown, black, orange means 10000pF = 10nF = 0.01µF.

Note that there are no gaps between the colors bands, so 2 identical bands actually appear as a wide band.

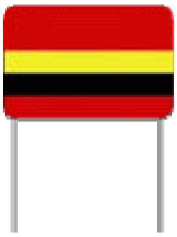

For example:

wide red, yellow means 220nF = 0.22µF.

Polystyrene Capacitors

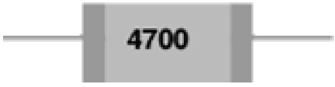

This type is rarely used now. Their value (in pF) is normally printed without units. Polystyrene capacitors can be damaged by heat when soldering (it melts the polystyrene!) so you should use a heat sink (such as a crocodile clip). Clip the heat sink to the lead between the capacitor and the joint.

Real capacitor values (the E3 and E6 series)

You may have noticed that capacitors are not available with every possible value, for example 22µF and 47µF are readily available, but 25µF and 50µF are not!

Why is this? Imagine that you decided to make capacitors every 10µF giving 10, 20, 30, 40, 50 and so on. That seems fine, but what happens when you reach 1000? It would be pointless to make 1000, 1010, 1020, 1030 and so on because for these values 10 is a very small difference, too small to be noticeable in most circuits and capacitors cannot be made with that accuracy.

To produce a sensible range of capacitor values you need to increase the size of the 'step' as the value increases. The standard capacitor values are based on this idea and they form a series which follows the same pattern for every multiple of ten.

The E3 series (3 values for each multiple of ten)
10, 22, 47, ... then it continues 100, 220, 470, 1000, 2200, 4700, 10000 etc.
Notice how the step size increases as the value increases (values roughly double each time).

The E6 series (6 values for each multiple of ten)
10, 15, 22, 33, 47, 68, ... then it continues 100, 150, 220, 330, 470, 680, 1000 etc.
Notice how this is the E3 series with an extra value in the gaps.

The E3 series is the one most frequently used for capacitors because many types cannot be made with very accurate values.

Variable capacitors

Variable capacitors are mostly used in radio tuning circuits and they are sometimes called 'tuning capacitors'. They have very small capacitance values, typically between 100pF and 500pF (100pF = 0.0001µF). The type illustrated usually has trimmers built in (for making small adjustments - see below) as well as the main variable capacitor.

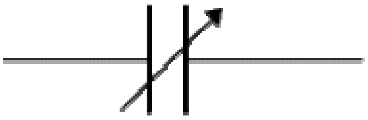

Variable Capacitor Symbol

Variable Capacitor

Many variable capacitors have very short spindles which are not suitable for the standard knobs used for variable resistors and rotary switches. It would be wise to check that a suitable knob is available before ordering a variable capacitor.

Variable capacitors are **not** normally used in timing circuits because their capacitance is too small to be practical and the range of values available is very limited. Instead timing circuits use a fixed capacitor and a variable resistor if it is necessary to vary the time period.

Trimmer Capacitors

Trimmer capacitors (trimmers) are miniature variable capacitors. They are designed to be mounted directly onto the circuit board and adjusted only when the circuit is built.

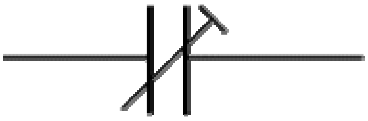

Trimmer Capacitor Symbol

Trimmer Capacitor

A small screwdriver or similar tool is required to adjust trimmers. The process of adjusting them requires patience because the presence of your hand and the tool will slightly change the capacitance of the circuit in the region of the trimmer!

Trimmer capacitors are only available with very small capacitances, normally less than 100pF. It is impossible to reduce their capacitance to zero, so they are usually specified by their minimum and maximum values, for example 2-10pF.

Trimmers are the capacitor equivalent of presets which are miniature variable resistors.

Connectors and Cables

Battery clips and holders

The standard battery clip fits a 9V PP3 battery and many battery holders such as the 6 × AA cell holder shown. Battery holders are also available with wires attached, with pins for PCB mounting, or as a complete box with lid, switch and wires.

Many small electronic projects use a 9V PP3 battery but if you wish to use the project for long periods a better choice is a battery holder with 6 AA cells. This has the same voltage but a much longer battery life and it will work out cheaper in the long run.

Larger battery clips fit 9V PP9 batteries but these are rarely used now.

Terminal blocks and PCB terminals

Terminal blocks are usually supplied in 12-way lengths but they can be cut into smaller blocks with a sharp knife, large wire cutters or a junior hacksaw. They are sometimes called 'chocolate blocks' because of the way they can be easily cut to size.

PCB terminal block

Terminal Block

PCB mounting terminal blocks provide an easy way of making semi-permanent connections to PCBs. Many are designed to interlock to provide more connections.

Crocodile clips

The 'standard' crocodile clip has no cover and a screw contact. However, miniature insulated crocodile clips are more suitable for many purposes including test leads. They have a solder contact and lugs which fold down to grip the cable's insulation, increasing the strength of the joint. Remember to feed the cable through the plastic cover *before* soldering! Add and remove the cover by fully opening the clip, a piece of wood can be used to hold the jaws open.

Crocodile Clips

4mm Plugs, Sockets and Terminals

These are the standard single pole connectors used on meters and other electronic equipment. They are capable of passing high currents (typically 10A) and most designs are very robust. Shrouded plugs and sockets are available for use with high voltages where there is a risk of electric shock. A wide variety of colors is available from most suppliers.

Plugs

Plugs may have a screw or solder terminal to hold the cable. Check if you need to thread the cable through the cover before connecting it. Some plugs, such as those illustrated, are 'stackable' which means that they include a socket to accept another plug, allowing several plugs to be connected to the same point - a very useful feature for test leads.

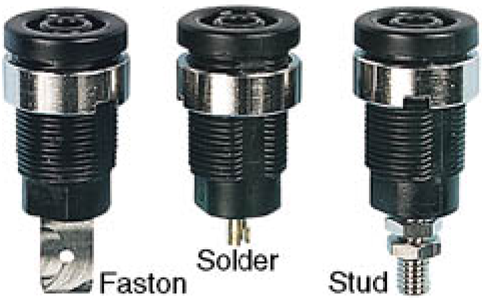

Sockets

These are usually described as 'panel mounting' because they are designed to be fitted to a case. Most sockets have a solder contact but the picture shows other options. Fit the socket in the case before attaching the wire otherwise you will be unable to add the mounting nut.

4mm terminal and solder tag

Terminals

In addition to a socket these have provision for attaching a wire by threading it through a hole (or wrapping it around the post) and tightening the top nut by hand. They usually have a threaded stud to fit a solder tag inside the case.

2mm Plugs and Sockets

These are smaller versions of the 4mm plugs and sockets described above, but terminals are not readily available. The plugs illustrated are stackable. Despite their small size these connectors can pass large currents and some are rated at 10A.

DC Power Plugs and Sockets

These 2-pole plugs and sockets ensure that the polarity of a DC supply cannot be accidentally reversed. The standard sizes are 2.1 and 2.5mm plug diameter. Standard plugs have a 10mm shaft, 'long' plugs have a 14mm shaft. Sockets are available for PCB or chassis mounting and most include a switch on the outer contact which is normally used to disconnect an internal battery when a plug is inserted.

Miniature versions with a 1.3mm diameter plug are used where small size is essential, such as for personal cassette players.

Jack Plugs and Sockets

These are intended for audio signals so mono and stereo versions are available. The sizes are determined by the plug diameter: ¼" (6.3mm), 3.5mm and 2.5mm. The 2.5mm size is only available for mono.

Screened plugs have metal bodies connected to the COM contact. Most connections are soldered, remember to thread cables through plug covers *before* soldering! Sockets are designed for PCB or chassis mounting.

¼" plug connections are similar to those for 3.5mm plugs shown below. ¼" socket connections are COM, R and L in that order from the mounting nut, ignore R for mono use. Most ¼" sockets have switches on all contacts which open as the plug is inserted so they can be used to isolate internal speakers for example.

¼" (6.3mm) jack plug and socket

3.5mm jack plug and socket

3.5mm jack line socket
(for fitting to a cable)

The connections for 3.5mm plugs and sockets are shown below. Plugs have a lug which should be folded down to grip the cable's insulation and increase the strength of the joint. 3.5mm mono sockets have a switch contact which can be used to switch off an internal speaker as the plug is inserted. Ignore this contact if you do not require the switching action.

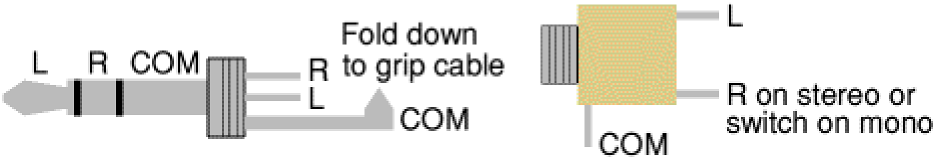

3.5mm jack plug and socket connections
(the R connection is not present on mono plugs)

L = left channel signal
R = right channel signal
COM = common (0V, screen)

Do not use jack plugs for power supply connections because the contacts may be briefly shorted as the plug is inserted. Use DC power connectors for this.

Phono Plugs and Sockets

These are used for screened cables carrying audio and video signals. Stereo connections are made using a pair of phono plugs and sockets. The centre contact is for the signal and the outer contact for the screen (0V, common). Screened plugs have metal bodies connected to the outer contact to give the signal additional protection from electrical noise. Sockets are available for PCB or chassis mounting, singly for mono, or in pairs for stereo. Line sockets are available for making extension leads.

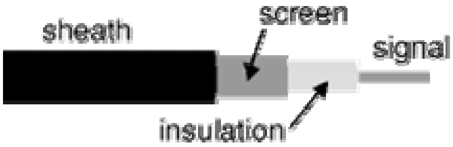

Construction of a screened cable

Coax Plugs and Sockets

These are similar to the phono plugs and sockets described above but they are designed for use with screened cables carrying much higher frequency signals, such as TV aerial leads. They provide better screening because at high frequencies this is essential to reduce electrical noise.

BNC Plugs and Sockets

These are designed for screened cables carrying high frequency signals where an undistorted and noise free signal is essential, for example oscilloscope leads. BNC plugs are connected with a **push and twist action**, to disconnect you need to twist and pull.

Plugs and sockets are rated by their impedance (50Ω or 75Ω) which must be the same as the cable's impedance. If the connector and cable impedances are not matched the signal will be distorted because it will be partly reflected at the connection, this is the electrical equivalent of the weak reflection which occurs when light passes through a glass window.

DIN Plugs and Sockets

DIN plug

These are intended for audio signals but they can be used for other low-current purposes where a multi-way connector is required. They are available from 3 way to 8 way. 5 way is used for stereo audio connections. The contacts are numbered on the connector, but they are not in numerical order! For audio use the 'common' (0V) wire is connected to contact 2. 5 way plugs and sockets are available in two versions: 180° and 270° (the angle refers to the arc formed by the contacts).

5 way 180° DIN socket
(chassis mounting)

Plastic covers of DIN plugs (and line sockets) are removed by depressing the retaining lug with a small screwdriver. You may also need small pliers to extract the body from the cover but do not pull on the pins themselves to avoid damage. Remember to thread the cable through the cover *before* starting to solder the connections!

Soldering DIN plugs is easier if you clamp the insert with the pins. Wires should be pushed into the hollow pins - first 'tin' the wires (coat them with a thin layer of solder) then melt a little solder into the hollow pin and insert the wire while keeping the solder molten. Take care to avoid melting the plastic base, stop and allow the pin to cool if necessary.

Mini-DIN connectors are used for computer equipment such as keyboards and mice but they are not a good choice for general use unless small size is essential.

D connectors

These are multi-pole connectors with provision for screw fittings to make semi-permanent connections, for example on computer equipment. The D shape prevents incorrect connection. Standard D-connectors have 2 rows of contacts (top picture); 9, 15 and 25-way versions are the most popular. High Density D-connectors have 3 rows of contacts (bottom picture); a 15-way version is used to connect computer monitors for example.

Note that covers (middle picture) are usually sold separately because both plugs and sockets can be fitted to cables by fitting a cover to a chassis mounted connector. PCB mounting versions of plugs and sockets are also available. The contacts are usually numbered on the body of the connector, although you may need a magnifying glass to see the very small markings. Soldering D-connectors requires a steady hand due to the closeness of the contacts, it is easy to accidentally unsolder a contact you have just completed while attempting to solder the next one!

IDC Communication Connectors

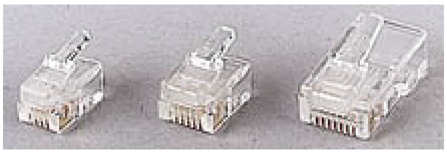

These multi-pole insulation displacement connectors are used for computer and telecommunications equipment. They automatically cut through the insulation on wires when installed and special tools are required to fit them. They are available as 4, 6 and 8-way versions.

The 8-way RJ45 is the standard connector for modern computer networks.

Standard UK telephone connectors are similar in style but a slightly different shape. They are called BT (British Telecom) connectors.

Cables

Cable... flex... lead... wire... what do all these terms mean?

- A **cable** is an assembly of one or more conductors (wires) with some flexibility.
- A **flex** is the proper name for the flexible cable fitted to mains electrical appliances.
- A **lead** is a complete assembly of cable and connectors.
- A **wire** is a single conductor which may have an outer layer of insulation (usually plastic).

Single core equipment wire

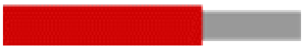

This is one solid wire with a plastic coating available in a wide variety of colors. It can be bent to shape but will break if repeatedly flexed. Use it for connections which will not be disturbed, for example links between points of a circuit board.

Typical specification: **1/0.6mm** (1 strand of 0.6mm diameter), maximum current 1.8A.

Stranded wire

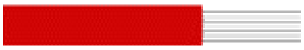

This consists of many fine strands of wire covered by an outer plastic coating. It is flexible and can withstand repeated bending without breaking. Use it for connections which may be disturbed, for example wires outside cases to sensors and switches. A very flexible version ('extra-flex') is used for test leads.

Typical specifications:

10/0.1mm (10 strands of 0.1mm diameter), maximum current 0.5A.
7/0.2mm (7 strands of 0.2mm diameter), maximum current 1.4A.
16/0.2mm (16 strands of 0.2mm diameter), maximum current 3A.
24/0.2mm (24 strands of 0.2mm diameter), maximum current 4.5A.
55/0.1mm (55 strands of 0.1mm diameter), maximum current 6A, used for test leads.

'Figure 8' (speaker) Cable

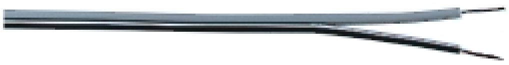

'Figure 8' cable consists of two stranded wires arranged in a figure of 8 shape. One wire is usually marked with a line. It is suitable for low voltage, low current (maximum 1A) signals where screening from electrical interference is not required. It is a popular choice for connecting loudspeakers and is often called 'speaker cable'.

Signal Cable

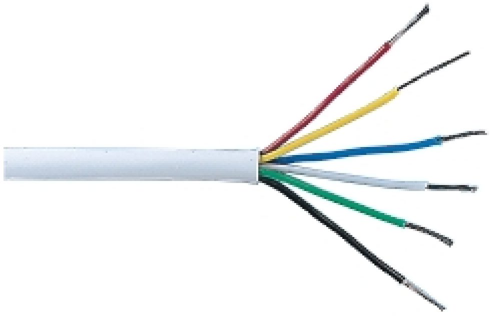

Signal cable consists of several color-coded cores of stranded wire housed within an outer plastic sheath. With a typical maximum current of 1A per core it is suitable for low voltage, low current signals where screening from electrical interference is not required.

The picture shows 6-core cable, but 4-core and 8-core are also readily available.

Screened Cable

The diagram shows the construction of screened cable. The central wire carries the signal and the screen is connected to 0V (common) to shield the signal from electrical interference. Screened cable is used for audio signals and dual versions are available for stereo.

Screened cable (mono)

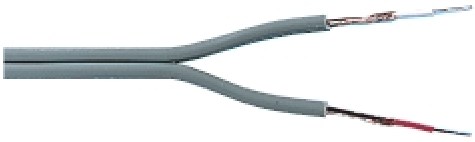

Screened cable (stereo)

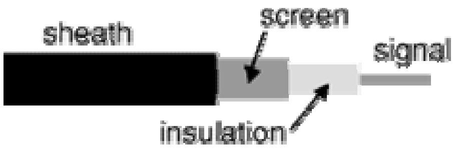

Construction of a screened cable

Screened cable (stereo)

Co-axial Cable

This type of screened cable (see above) is designed to carry high frequency signals such as those found in TV aerials and oscilloscope leads.

Mains flex

Flex is the proper name for the flexible cable used to connect appliances to the mains supply. It contains 2 cores (for live and neutral) or 3 cores (for live, neutral and earth). Mains flex has thick insulation for the high voltage (230V in UK) and it is available with various current ratings: 3A, 6A and 13A are popular sizes in the UK.

Mains flex is sometimes used for low voltage circuits which pass a high current, but please think carefully before using it in this way. The distinctive colors of mains flex should act as a warning of the mains high voltage which can be lethal; using mains flex for low voltage circuits can undermine this warning.

Diodes

Example:

Circuit symbol: 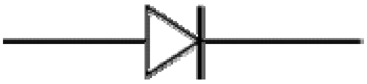

Diodes allow electricity to flow in only one direction. The arrow of the circuit symbol shows the direction in which the current can flow. Diodes are the electrical version of a valve and early diodes were actually called valves.

Function

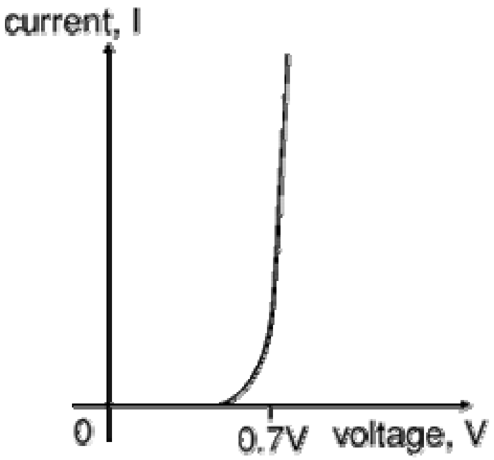

Characteristic of a Silicon Diode

Forward Voltage Drop

Electricity uses up a little energy pushing its way through the diode, rather like a person pushing through a door with a spring. This means that there is a small voltage across a conducting diode, it is called the **forward voltage drop** and is about 0.7V for all normal diodes which are made from silicon. The forward voltage drop of a diode is almost constant whatever the current passing through the diode so they have a very steep characteristic (current-voltage graph).

Reverse Voltage

When a reverse voltage is applied a perfect diode does not conduct, but all real diodes leak a very tiny current of a few µA or less. This can be ignored in most circuits because it will be very much smaller than the current flowing in the forward direction. However, all diodes have a **maximum reverse voltage** (usually 50V or more) and if this is exceeded the diode will fail and pass a large current in the reverse direction, this is called **breakdown**.

Ordinary diodes can be split into two types: Signal diodes which pass small currents of 100mA or less and Rectifier diodes which can pass large currents. In addition there are LEDs (which have their own page) and Zener diodes (at the bottom of this page).

Connecting and Soldering

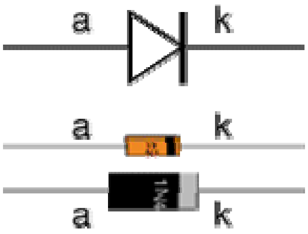

Diodes must be connected the correct way round, the diagram may be labeled **a** or **+** for anode and **k** or **-** for cathode (yes, it really is k, not c, for cathode!). The cathode is marked by a line painted on the body. Diodes are labeled with their code in small print; you may need a magnifying glass to read this on small signal diodes!

Small **signal diodes** can be damaged by heat when soldering, but the risk is small unless you are using a **germanium diode** (codes beginning OA...) in which case you should use a heat sink clipped to the lead between the joint and the diode body. A standard crocodile clip can be used as a heat sink. **Rectifier diodes** are quite robust and no special precautions are needed for soldering them.

Testing Diodes

You can use a multimeter or a simple tester (battery, resistor and LED) to check that a diode conducts in one direction but not the other. A lamp may be used to test a rectifier diode, but do NOT use a lamp to test a signal diode because the large current passed by the lamp will destroy the diode!

Signal Diodes (small current)

Signal diodes are used to process information (electrical signals) in circuits, so they are only required to pass small currents of up to 100mA.

General purpose signal diodes such as the 1N4148 are made from silicon and have a forward voltage drop of 0.7V.

Germanium diodes such as the OA90 have a lower forward voltage drop of 0.2V and this makes them suitable to use in radio circuits as detectors which extract the audio signal from the weak radio signal.

For general use, where the size of the forward voltage drop is less important, silicon diodes are better because they are less easily damaged by heat when soldering, they have a lower resistance when conducting, and they have very low leakage currents when a reverse voltage is applied.

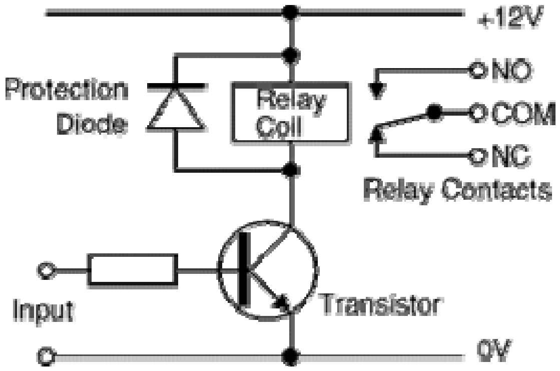

Protection Diodes for Relays

Signal diodes are also used with relays to protect transistors and integrated circuits from the brief high voltage produced when the relay coil is switched off. The diagram shows how a protection diode is connected across the relay coil, note that the diode is connected 'backwards' so that it will normally NOT conduct. Conduction only occurs when the relay coil is switched off, at this moment current tries to continue flowing through the coil and it is harmlessly diverted through the diode. Without the diode no current could flow and the coil would produce a damaging high voltage 'spike' in its attempt to keep the current flowing.

Rectifier Diodes (large current)

Rectifier diodes are used in power supplies to convert alternating current (AC) to direct current (DC), a process called rectification. They are also used elsewhere in circuits where a large current must pass through the diode.

All rectifier diodes are made from silicon and therefore have a forward voltage drop of 0.7V. The table shows maximum current and maximum reverse voltage for some popular rectifier diodes. The 1N4001 is suitable for most low voltage circuits with a current of less than 1A.

Diode	Maximum Current	Maximum Reverse Voltage
1N4001	1A	50V
1N4002	1A	100V
1N4007	1A	1000V
1N5401	3A	100V
1N5408	3A	1000V

Bridge Rectifiers

There are several ways of connecting diodes to make a rectifier to convert AC to DC. The bridge rectifier is one of them and it is available in special packages containing the four diodes required. Bridge rectifiers are rated by their maximum current and maximum reverse voltage. They have four leads or terminals: the two DC outputs are labeled **+** and **-**, the two AC inputs are labeled **~**.

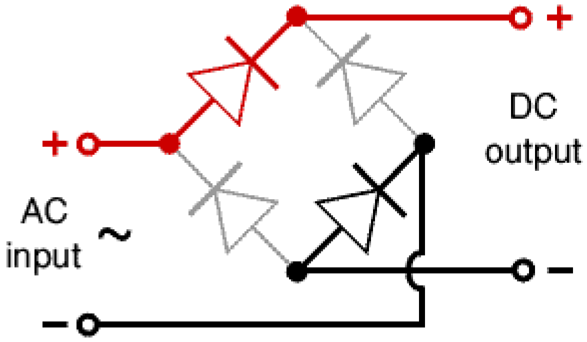

The diagram shows the operation of a bridge rectifier as it converts AC to DC. Notice how alternate pairs of diodes conduct.

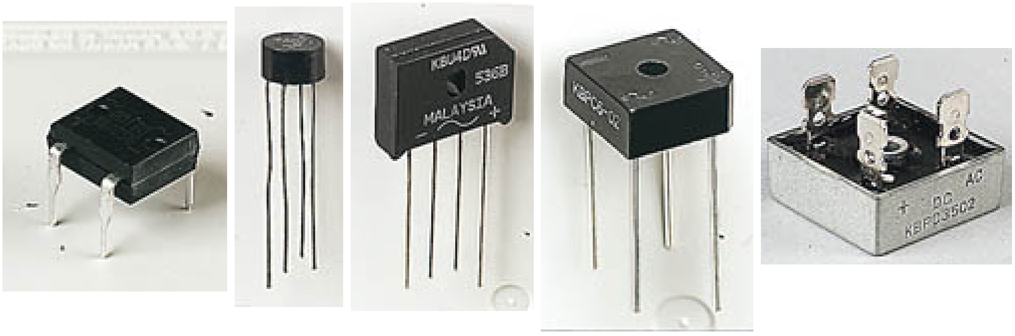

Various types of Bridge Rectifiers
Note that some have a hole through their centre for attaching to a heat sink

Zener Diodes

Example:

a = anode, k = cathode

Circuit symbol: 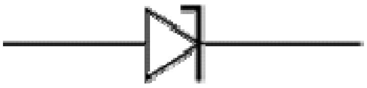

Zener diodes are used to maintain a fixed voltage. They are designed to 'breakdown' in a reliable and non-destructive way so that they can be used **in reverse** to maintain a fixed voltage across their terminals. The diagram below shows how they are connected, with a resistor in series to limit the current.

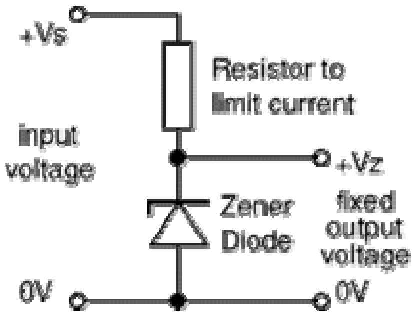

Zener diodes can be distinguished from ordinary diodes by their code and breakdown voltage which are printed on them. Zener diode codes begin BZX... or BZY... Their breakdown voltage is printed with V in place of a decimal point, so 4V7 means 4.7V for example.

Zener diodes are rated by their breakdown voltage and maximum power:

- The minimum voltage available is 2.7V.
- Power ratings of 400mW and 1.3W are common.

Integrated Circuits (Chips)

Integrated Circuits are usually called ICs or chips. They are complex circuits which have been etched onto tiny chips of semiconductor (silicon). The chip is packaged in a plastic holder with pins spaced on a 0.1" (2.54mm) grid which will fit the holes on stripboard and breadboards. Very fine wires inside the package link the chip to the pins.

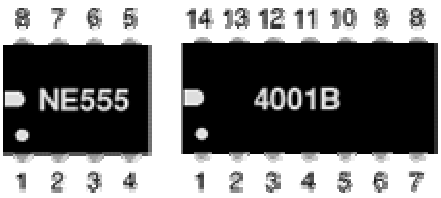

Pin Numbers

The pins are numbered anti-clockwise around the IC (chip) starting near the notch or dot. The diagram shows the numbering for 8-pin and 14-pin ICs, but the principle is the same for all sizes.

Chip Holders (DIL Sockets)

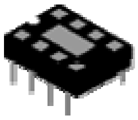

ICs (chips) are easily damaged by heat when soldering and their short pins cannot be protected with a heat sink. Instead we use a chip holder, strictly called a DIL socket (DIL = Dual In-Line), which can be safely soldered onto the circuit board. The chip is pushed into the holder when all soldering is complete.

Chip holders are only needed when soldering so they are not used on breadboards.

Commercially produced circuit boards often have chips soldered directly to the board without a chip holder, usually this is done by a machine which is able to work very quickly. Please don't attempt to do this yourself because you are likely to destroy the chip and it will be difficult to remove without damage by de-soldering.

Removing a Chip from its Holder

If you need to remove a chip it can be gently pull out of the holder with a small flat-blade screwdriver. Carefully lever up each end by inserting the screwdriver blade between the chip and its holder and gently twisting the screwdriver. Take care to start lifting at both ends before you attempt to remove the chip, otherwise you will bend and possibly break the pins.

Static Precautions

Antistatic bags for ICs

Many ICs are **static sensitive** and can be damaged when you touch them because your body may have become charged with static electricity, from your clothes for example. Static sensitive ICs will be supplied in antistatic packaging with a warning label and they should be left in this packaging until you are ready to use them.

It is usually adequate to earth your hands by touching a metal water pipe or window frame before handling the IC but for the more sensitive (and expensive!) ICs special equipment is available, including earthed wrist straps and earthed work surfaces. You can make an earthed work surface with a sheet of aluminum kitchen foil and using a crocodile clip to connect the foil to a metal water pipe or window frame with a 10kΩ resistor in series.

Datasheets

Datasheets are available for most ICs giving detailed information about their ratings and functions. In some cases example circuits are shown. The large amount of information with symbols and abbreviations can make datasheets seem overwhelming to a beginner, but they are worth reading as you become more confident because they contain a great deal of useful information for more experienced users designing and testing circuits.

Datasheets are available as PDF files from:

- DatasheetArchive.com
- Datasheets.org.uk
- DatasheetCatalog.com

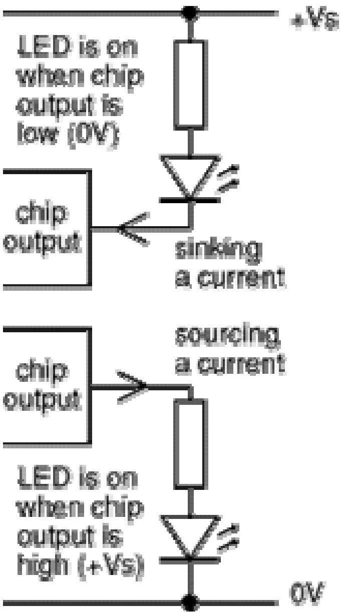
+Vs
LED is on when chip output is low (0V)
chip output
sinking a current
sourcing a current
chip output
LED is on when chip output is high (+Vs)
0V

Sinking and Sourcing Current

Chip outputs are often said to 'sink' or 'source' current. The terms refer to the direction of the current at the chip's output.

If the chip is **sinking current** it is flowing **into the output**. This means that a device connected between the positive supply (+Vs) and the chip output will be switched **on when the output is low (0V)**.

If the chip is **sourcing current** it is flowing **out of the output**. This means that a device connected between the chip output and the negative supply (0V) will be switched **on when the output is high (+Vs)**.

It is possible to connect two devices to a chip output so that one is on when the output is low and the other is on when the output is high. The maximum sinking and sourcing currents for a chip output are usually the same but there are some exceptions, for example 74LS TTL logic chips can sink up to 16mA but only source 2mA.

Using Diodes to Combine Outputs

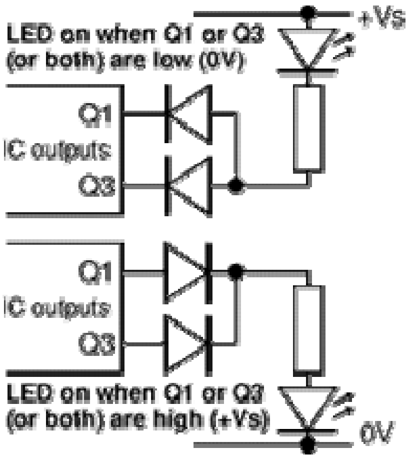
LED on when Q1 or Q3 (or both) are low (0V)
+Vs
Q1
IC outputs
Q3
Q1
IC outputs
Q3
LED on when Q1 or Q3 (or both) are high (+Vs)
0V

The outputs of chips (ICs) must never be directly connected together. However, diodes can be used to combine two or more digital (high/low) outputs from a chip such as a counter. This can be a useful way of producing simple logic functions without using logic gates!

The diagram shows two ways of combining outputs using diodes. The diodes must be capable of passing the output current. 1N4148 signal diodes are suitable for low current devices such as LED.

For example the outputs Q0 - Q9 of a 4017 1-of-10 counter go high in turn. Using diodes to combine the 2nd (Q1) and 4th (Q3) outputs as shown in the bottom diagram will make the LED flash twice followed by a longer gap. The diodes are performing the function of an OR gate.

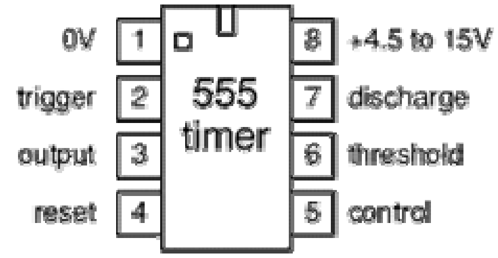
0V
1
trigger
2
output
3
reset
4
555
timer
8
+4.5 to 15V
7
discharge
6
threshold
5
control

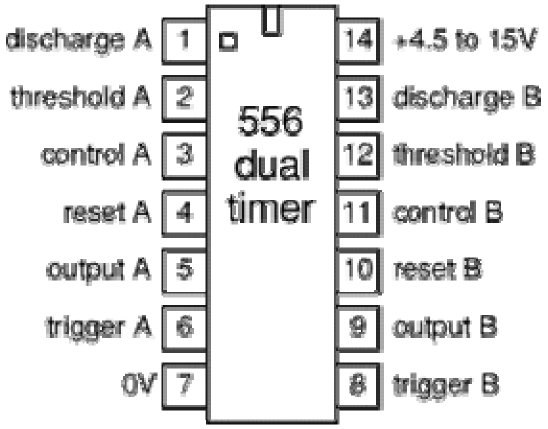
discharge A
1
threshold A
2
control A
3
reset A
4
output A
5
trigger A
6
0V
7
556
dual
timer
14
+4.5 to 15V
13
discharge B
12
threshold B
11
control B
10
reset B
9
output B
8
trigger B

The 555 and 556 Timers

The 8-pin 555 timer chip is used in many projects, a popular version is the NE555. Most circuits will just specify '555 timer IC' and the NE555 is suitable for these. The 555 output (pin 3) can sink and source up to 200mA. This is more than most chips and it is sufficient to supply LED, relay coils and low current lamps. To switch larger currents you can connect a transistor.

The 556 is a dual version of the 555 housed in a 14-pin package. The two timers (A and B) share the same power supply pins.

Low power versions of the 555 are made, such as the ICM7555, but these should only be used when specified (to increase battery life) because their maximum output current of about 20mA (with 9V supply) is too low for many standard 555 circuits. The ICM7555 has the same pin arrangement as a standard 555.

Logic ICs (chips)

Logic ICs process digital signals and there are many devices, including logic gates, flip-flops, shift registers, counters and display drivers. They can be split into two groups according to their pin arrangements: the 4000 series and the 74 series which consists of various families such as the 74HC, 74HCT and 74LS.

For most new projects the 74HC family is the best choice. The older 4000 series is the only family which works with a supply voltage of more than 6V. The 74LS and 74HCT families require a 5V supply so they are not convenient for battery operation.

The table below summarizes the important properties of the most popular logic families:

Property	**4000 Series**	**74 Series 74HC**	**74 Series 74HCT**	**74 Series 74LS**
Technology	CMOS	High-speed CMOS	High-speed CMOS TTL compatible	TTL Low-power Schottky
Power Supply	3 to 15V	2 to 6V	5V ±0.5V	5V ±0.25V
Inputs	Very high impedance. Unused inputs must be connected to +Vs or 0V. Inputs cannot be reliably driven by 74LS outputs unless a 'pull-up' resistor is used (see below).		Very high impedance. Unused inputs must be connected to +Vs or 0V. Compatible with 74LS (TTL) outputs.	'Float' high to logic 1 if unconnected. 1mA must be drawn out to hold them at logic 0.
Outputs	Can sink and source about 5mA (10mA with 9V supply), enough to light an LED. To switch larger currents use a transistor.	Can sink and source about 20mA, enough to light an LED. To switch larger currents use a transistor.	Can sink and source about 20mA, enough to light an LED. To switch larger currents use a transistor.	Can sink up to 16mA (enough to light an LED), but source only about 2mA. To switch larger currents use a transistor.
Fan-out	One output can drive up to 50 CMOS, 74HC or 74HCT inputs, but only one 74LS input.	One output can drive up to 50 CMOS, 74HC or 74HCT inputs, but only 10 74LS inputs.		One output can drive up to 10 74LS inputs or 50 74HCT inputs.
Maximum Frequency	about 1MHz	about 25MHz	about 25MHz	about 35MHz
Power consumption of the IC itself	A few µW.	A few µW.	A few µW.	A few mW.

Mixing Logic Families

It is best to build a circuit using just one logic family, but if necessary the different families may be mixed providing the power supply is suitable for all of them. For example mixing 4000 and 74HC requires the power supply to be in the range 3 to 6V. A circuit which includes 74LS or 74HCT ICs must have a 5V supply.

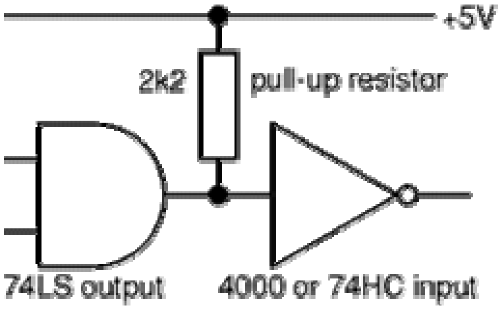

Driving 4000 or 74HC inputs from a 74LS output using a pull-up resistor.

A 74LS output cannot reliably drive a 4000 or 74HC input unless a 'pull-up' resistor of 2.2kΩ is connected between the +5V supply and the input to correct the slightly different logic voltage ranges used.

Note that a 4000 series output can drive only one 74LS input.

---

74 Series Logic ICs

General characteristics

There are several families of logic chips numbered from 74xx00 onwards with letters (xx) in the middle of the number to indicate the type of circuitry, e.g. 74LS00 and 74HC00. The original family (now obsolete) had no letters, e.g. 7400.

The **74LS** (Low-power Schottky) family (like the original) uses TTL (Transistor-Transistor Logic) circuitry which is fast but requires more power than later families. The 74 series is often still called the 'TTL series' even though the latest chips do not use TTL!

The **74HC** family has High-speed CMOS circuitry, combining the speed of TTL with the very low power consumption of the 4000 series. They are CMOS chips with the same pin arrangements as the older 74LS family. Note that 74HC inputs cannot be reliably driven by 74LS outputs because the voltage ranges used for logic 0 are not quite compatible, use 74HCT instead.

The **74HCT** family is a special version of 74HC with 74LS TTL-compatible inputs so 74HCT can be safely mixed with 74LS in the same system. In fact 74HCT can be used as low-power direct replacements for the older 74LS ICs in most circuits. The minor disadvantage of 74HCT is a lower immunity to noise, but this is unlikely to be a problem in most situations.

The CMOS circuitry used in the **74HC** and **74HCT** series ICs means that they are static sensitive. Touching a pin while charged with static electricity (from your clothes for example) may damage the IC. In fact most ICs in regular use are quite tolerant and earthing your hands by touching a metal water pipe or window frame before handling them will be adequate. ICs should be left in their protective packaging until you are ready to use them.

To compare the different logic families please see the Summary table of logic families

For most new projects the 74HC family is the best choice.
The 74LS and 74HCT families require a 5V supply so they are not convenient for battery operation.

74HC and 74HCT family characteristics:

- **74HC Supply:** 2 to 6V, small fluctuations are tolerated.
- **74HCT Supply:** 5V ±0.5V, a regulated supply is best.
- **Inputs** have very high impedance (resistance), this is good because it means they will not affect the part of the circuit where they are connected. However, it also means that unconnected inputs can easily pick up electrical noise and rapidly change between high and low states in an unpredictable way. This is likely to make the chip behave erratically and it will significantly increase the supply current. To prevent problems all unused inputs MUST be connected to the supply (either +Vs or 0V), this applies even if that part of the chip is not being used in the circuit!
  Note that 74HC inputs cannot be reliably driven by 74LS outputs because the voltage ranges used for logic 0 are not quite compatible. For reliability use **74HCT** if the system includes some 74LS chips.
- **Outputs** can sink and source about 4mA if you wish to maintain the correct output voltage to drive logic inputs, but if there is no need to drive any inputs the maximum current is about 20mA. To switch larger currents you can connect a transistor.
- **Fan-out:** one output can drive many inputs (50+), except 74LS inputs because these require a higher current and only 10 can be driven.
- **Gate propagation time**: about 10ns for a signal to travel through a gate.
- **Frequency**: up to 25MHz.
- **Power consumption** (of the chip itself) is very low, a few µW. It is much greater at high frequencies, a few mW at 1MHz for example.

74LS family TTL characteristics:

- **Supply:** 5V ±0.25V, it must be very smooth, a regulated supply is best. In addition to the normal supply smoothing, a 0.1µF capacitor should be connected across the supply near the chip to remove the 'spikes' generated as it switches state, one capacitor is needed for every 4 chips.
- **Inputs** 'float' high to logic 1 if unconnected, but do not rely on this in a permanent (soldered) circuit because the inputs may pick up electrical noise. 1mA must be drawn out to hold inputs at logic 0. In a permanent circuit it is wise to connect any unused inputs to +Vs to ensure good immunity to noise.
- **Outputs** can sink up to 16mA (enough to light an LED), but they can source only about 2mA. To switch larger currents you can connect a transistor.
- **Fan-out**: one output can drive up to 10 74LS inputs, but many more 74HCT inputs.
- **Gate propagation time**: about 10ns for a signal to travel through a gate.
- **Frequency**: up to about 35MHz (under the right conditions).
- **Power consumption** (of the chip itself) is a few mW.

Open Collector Outputs

Some 74 series ICs have open collector outputs, this means they can sink current but they cannot source current. They behave like an NPN transistor switch.

The diagram below shows how an open collector output can be connected to sink current from a supply which has a higher voltage than the logic IC supply. The maximum load supply is 15V for most open collector ICs.

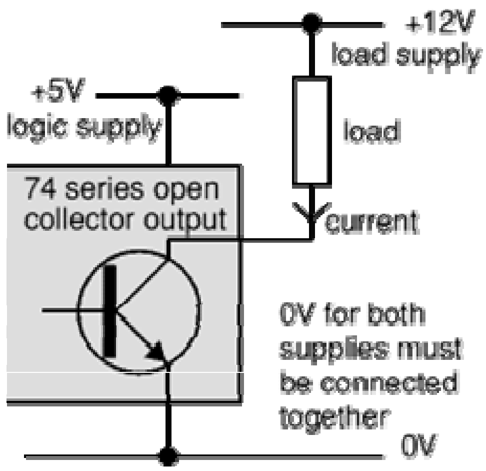

Open collector outputs can be safely connected together to switch on a load when any one of them is low; unlike normal outputs which must be combined using diodes.

There are many ICs in the 74 series and this page only covers a selection, concentrating on the most useful gates, counters, decoders and display drivers. For each IC there is a diagram showing the pin arrangement and brief notes explain the function of the pins where necessary. For simplicity the family letters after the 74 are omitted in the diagrams below because the pin connections apply to all 74 series ICs with the same number. For example 7400 NAND gates are available as 74HC00, 74HCT00 and 74LS00.

If you are using another reference please be aware that there is some variation in the terms used to describe pin functions, for example **reset** is also called **clear**. Some inputs are 'active low' which means they perform their function when low. If you see a line drawn above a label it means it is active low, for example: $\overline{\text{reset}}$ (say 'reset-bar').

Gates

Quad 2-Input Gates

- 7400 quad 2-input NAND
- 7403 quad 2-input NAND with open collector outputs
- 7408 quad 2-input AND
- 7409 quad 2-input AND with open collector outputs
- 7432 quad 2-input OR
- 7486 quad 2-input EX-OR
- 74132 quad 2-input NAND with Schmitt trigger inputs

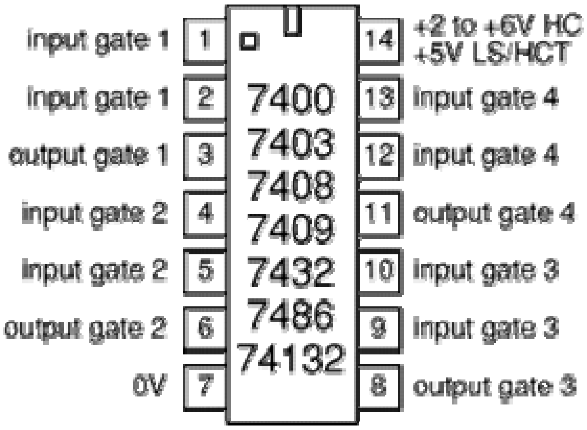
input gate 1
input gate 1
output gate 1
input gate 2
input gate 2
output gate 2
0V
1
2
3
4
5
6
7
7400
7403
7408
7409
7432
7486
74132
14
13
12
11
10
9
8
+2 to +6V HC
+5V LS/HCT
input gate 4
input gate 4
output gate 4
input gate 3
input gate 3
output gate 3

The 74132 has Schmitt trigger inputs to provide good noise immunity. They are ideal for slowly changing or noisy signals.

- 7402 quad 2-input NOR (**Note the unusual gate layout**)

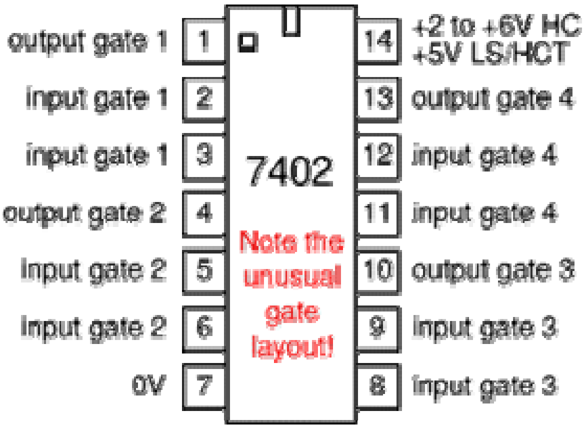
output gate 1
input gate 1
input gate 1
output gate 2
input gate 2
input gate 2
0V
1
2
3
4
5
6
7
7402
Note the unusual gate layout!
14
13
12
11
10
9
8
+2 to +6V HC
+5V LS/HCT
output gate 4
input gate 4
input gate 4
output gate 3
input gate 3
input gate 3

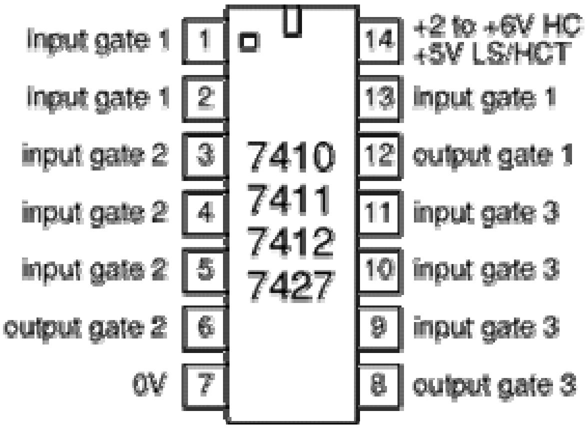
input gate 1
input gate 1
input gate 2
input gate 2
input gate 2
output gate 2
0V
1
2
3
4
5
6
7
7410
7411
7412
7427
14
13
12
11
10
9
8
+2 to +6V HC
+5V LS/HCT
input gate 1
output gate 1
input gate 3
input gate 3
input gate 3
output gate 3

Triple 3-Input Gates

- 7410 triple 3-input NAND
- 7411 triple 3-input AND
- 7412 triple 3-input NAND with open collector outputs

- 7427 triple 3-input NOR

Notice how gate 1 is spread across the two sides of the package.

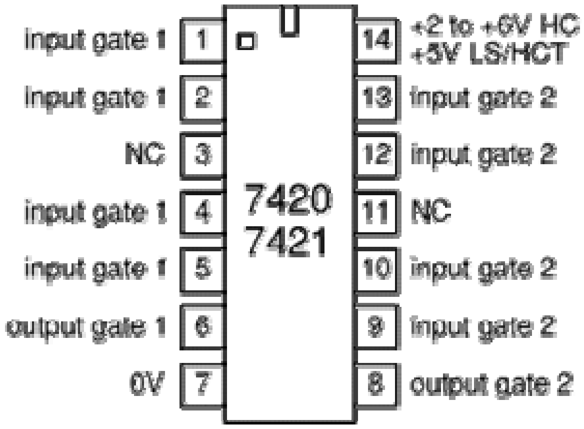
input gate 1
1
input gate 1
2
NC
3
input gate 1
4
input gate 1
5
output gate 1
6
0V
7
7420
7421
14
+2 to +6V HC
+5V LS/HCT
13
input gate 2
12
input gate 2
11
NC
10
input gate 2
9
input gate 2
8
output gate 2

Dual 4-Input Gates

- 7420 dual 4-input NAND
- 7421 dual 4-input AND

NC = **N**o **C**onnection (a pin that is not used).

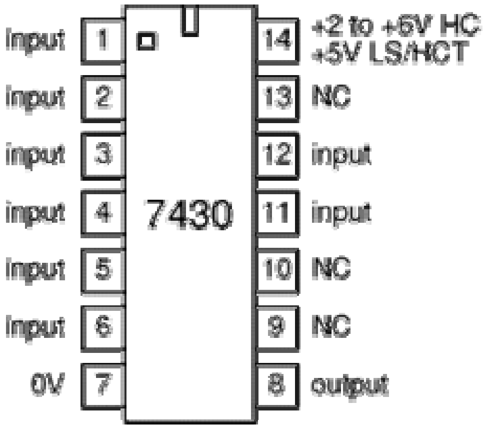
input
1
input
2
input
3
input
4
input
5
input
6
0V
7
7430
14
+2 to +6V HC
+5V LS/HCT
13
NC
12
input
11
input
10
NC
9
NC
8
output

7430 8-Input NAND Gate

NC = **N**o **C**onnection (a pin that is not used).

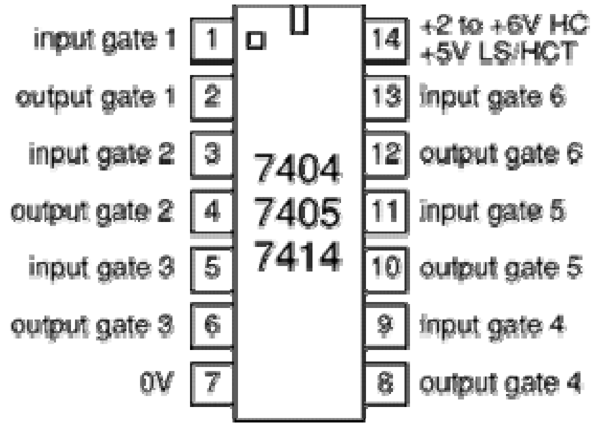
input gate 1
output gate 1
input gate 2
output gate 2
input gate 3
output gate 3
0V
1
2
3
4
5
6
7
7404
7405
7414
14
13
12
11
10
9
8
+2 to +6V HC
+5V LS/HCT
input gate 6
output gate 6
input gate 5
output gate 5
input gate 4
output gate 4

Hex NOT Gates (also called Inverter)

- 7404 hex NOT
- 7405 hex NOT with open collector outputs
- 7414 hex NOT with Schmitt trigger inputs

The 7414 has Schmitt trigger inputs to provide good noise immunity. They are ideal for slowly changing or noisy signals.

Counters

7490 decade (0-9) ripple counter
7493 4-bit (0-15) ripple counter

These are **ripple** counters so beware that glitches may occur in any logic gate systems connected to their outputs due to the slight delay before the later counter outputs respond to a clock pulse.

The count advances as the **clock** input becomes low (on the falling-edge), this is indicated by the bar over the clock label. This is the usual clock behavior of ripple counters and it means a counter output can directly drive the clock input of the next counter in a chain.

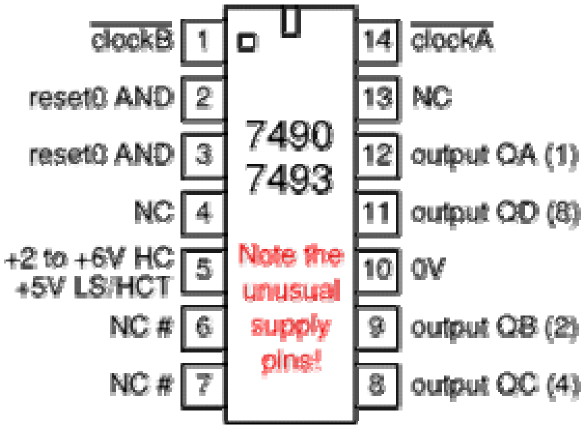
clockB
reset0 AND
reset0 AND
NC
+2 to +6V HC
+5V LS/HCT
NC #
NC #
1
2
3
4
5
6
7
7490
7493
Note the unusual supply pins!
14
13
12
11
10
9
8
clockA
NC
output QA (1)
output QD (8)
0V
output QB (2)
output QC (4)

Note:

NC = No Connection (a pin that is not used). # on the 7490 pins 6 and 7 connect to an internal AND gate for resetting to 9.

The counter is in two sections: clockA-QA and clockB-QB-QC-QD. For normal use connect QA to clockB to link the two sections, and connect the external clock signal to clockA.

For normal operation at least one **reset0** input should be low, making both high resets the counter to zero (0000, QA-QD low). Note that the 7490 has a pair of **reset9** inputs on pins 6 and 7, these reset the counter to nine (1001) so at least one of them must be low for counting to occur.

Counting to less than the maximum (9 or 15) can be achieved by connecting the appropriate output(s) to the two **reset0** inputs. If only one reset input is required the two inputs can be connected together. For example: to count 0 to 8 connect QA (1) and QD (8) to the reset inputs.

74390 Dual Decade (0-9) Ripple Counter

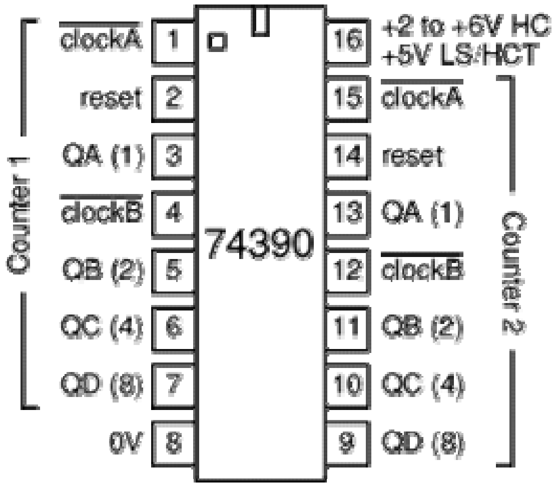

The 74390 contains two separate decade (0 to 9) counters, one on each side of the chip. They are **ripple** counters so beware that glitches may occur in any logic gate systems connected to their outputs due to the slight delay before the later counter outputs respond to a clock pulse.

The count advances as the **clock** input becomes low (on the falling-edge), this is indicated by the bar over the clock label. This is the usual clock behavior of ripple counters and it means a counter output can directly drive the clock input of the next counter in a chain.

Each counter is in two sections: clockA-QA and clockB-QB-QC-QD. For normal use connect QA to clockB to link the two sections, and connect the external clock signal to clockA. The **reset** input should be low; making it high resets the counter to zero (0000, QA-QD low).

Counting to less than 9 can be achieved by connecting the appropriate output(s) to the reset input, using an AND gate if necessary. For example: to count 0 to 7 connect QD (8) to reset, to count 0 to 8 connect QA (1) and QD (8) to reset using an AND gate.

74393 Dual 4-Bit (0-15) Ripple Counter

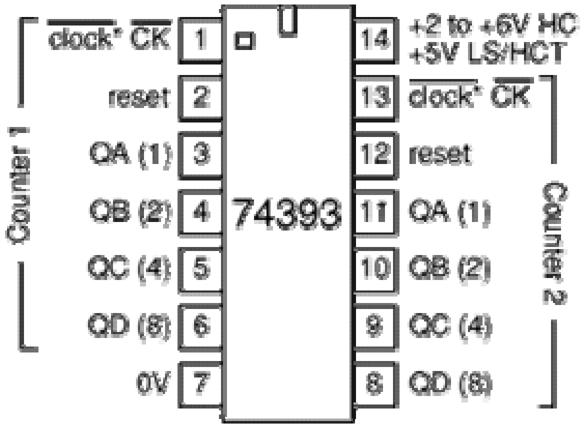
clock* CK
reset
QA (1)
QB (2)
QC (4)
QD (8)
0V
1
2
3
4
5
6
7
Counter 1
74393
14
13
12
11
10
9
8
+2 to +6V HC
+5V LS/HCT
clock* CK
reset
QA (1)
QB (2)
QC (4)
QD (8)
Counter 2

The 74393 contains two separate 4-bit (0 to 15) counters, one on each side of the chip. They are **ripple** counters so beware that glitches may occur in logic systems connected to their outputs due to the slight delay before the later outputs respond to a clock pulse.

The count advances as the **clock** input becomes low (on the falling-edge), this is indicated by the bar over the clock label. This is the usual clock behavior of ripple counters and it means a counter output can directly drive the clock input of the next counter in a chain.

For normal operation the **reset** input should be low, making it high resets the counter to zero (0000, QA-QD low).

Counting to less than 15 can be achieved by connecting the appropriate output(s) to the reset input, using an AND gate if necessary. For example to count 0 to 8 connect QA (1) and QD (8) to reset using an AND gate.

Connecting ripple counters in a chain

The diagram below shows how to link ripple counters in a chain, notice how the highest output QD of each counter drives the **clock** input of the next counter.

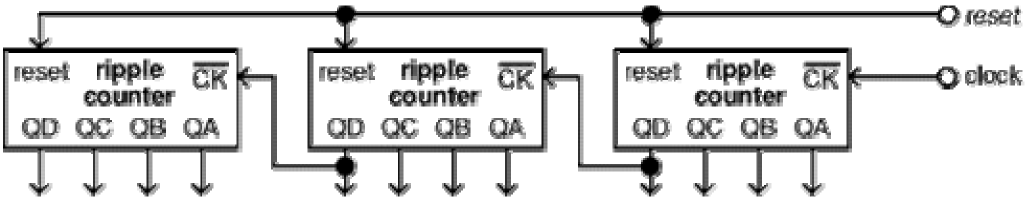
reset
ripple counter
CK
QD QC QB QA
reset
ripple counter
CK
QD QC QB QA
reset
ripple counter
CK
QD QC QB QA
reset
clock

74160-3 Synchronous Counters

- 74160 synchronous decade counter (standard reset)
- 74161 synchronous 4-bit counter (standard reset)
- 74162 synchronous decade counter (synchronous reset)
- 74163 synchronous 4-bit counter (synchronous reset)

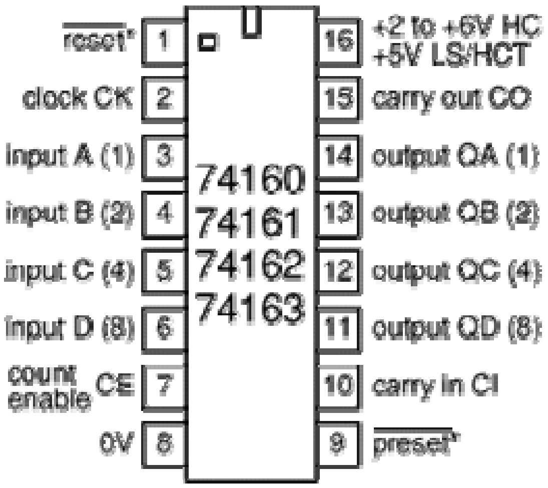
reset*
1
clock CK
2
input A (1)
3
input B (2)
4
input C (4)
5
input D (8)
6
count enable CE
7
0V
8
74160
74161
74162
74163
16
+2 to +6V HC
+5V LS/HCT
15
carry out CO
14
output QA (1)
13
output QB (2)
12
output QC (4)
11
output QD (8)
10
carry in CI
9
preset*

These are **synchronous** counters so their outputs change precisely together on each clock pulses. This is helpful if you need to connect their outputs to logic gates because it avoids the glitches which occur with ripple counters.

The count advances as the **clock** input becomes high (on the rising-edge). The **decade** counters count from 0 to 9 (0000 to 1001 in binary). The **4-bit** counters count from 0 to 15 (0000 to 1111 in binary).

For normal operation (counting) the **reset**, **preset**, **count enable** and **carry in** inputs should all be high. When **count enable** is low the clock input is ignored and counting stops.

The counter may be **preset** by placing the desired binary number on the **inputs A-D**, making the **preset** input low, and applying a positive pulse to the **clock** input. The **inputs A-D** may be left unconnected if not required; **preset** is also known as parallel enable (PE)

The **reset** input is active-low so it should be high (+Vs) for normal operation (counting). When low it resets the count to zero (0000, QA-QD low), this happens immediately with the 74160 and 74161 (**standard reset**), but with the 74162 and 74163 (**synchronous reset**) the reset occurs on the rising-edge of the clock input.

Counting to less than the maximum (15 or 9) can be achieved by connecting the appropriate output(s) through a NOT or NAND gate to the reset input. For the 74162 and 74163 (**synchronous reset**) you must use the output(s) representing **one less** than the reset count you require, e.g. to reset on 7 (counting 0 to 6) use QB (2) and QC (4).

Connecting synchronous counters in a chain

The diagram below shows how to link synchronous counters such as 74160-3, notice how all the **clock (CK)** inputs are linked. **Carry out (CO)** is used to feed the **carry in (CI)** of the next counter. **Carry in (CI)** of the first 74160-3 counter should be high.

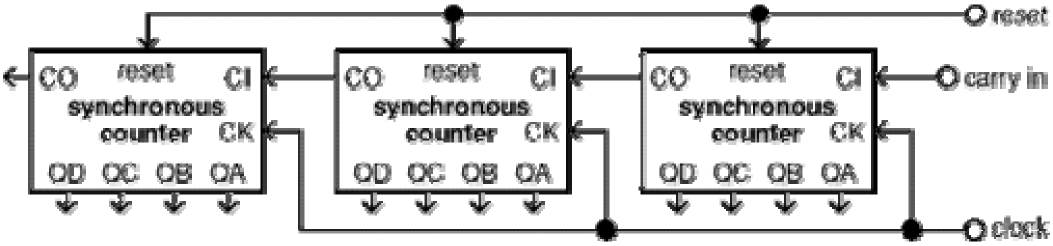
reset
carry in
clock
CO
reset
CI
synchronous counter
CK
QD
QC
QB
QA

74192 up/down decade (0-9) counter
74193 up/down 4-bit (0-15) counter

These are **synchronous** counters so their outputs change precisely together on each clock pulse. This is helpful if you need to connect their outputs to logic gates because it avoids the glitches which occur with ripple counters.

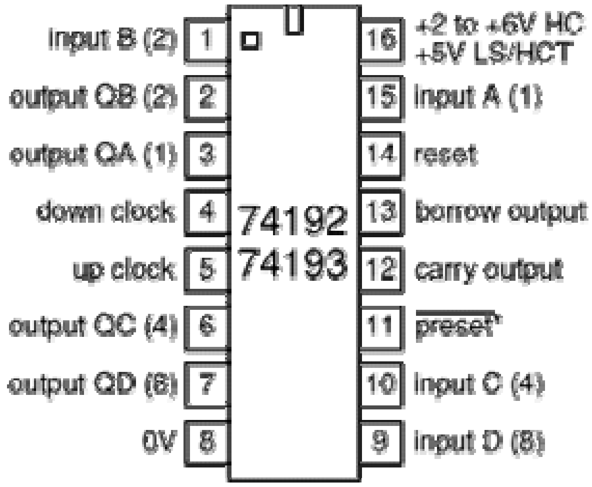
input B (2)
1
output QB (2)
2
output QA (1)
3
down clock
4
up clock
5
output QC (4)
6
output QD (8)
7
0V
8
74192
74193
16
+2 to +6V HC
+5V LS/HCT
15
input A (1)
14
reset
13
borrow output
12
carry output
11
preset
10
input C (4)
9
input D (8)

These counters have separate clock inputs for counting up and down. The count increases as the **up clock** input becomes high (on the rising-edge). The count decreases as the **down clock** input becomes high (on the rising-edge). In both cases the other clock input should be high.

For normal operation (counting) the **preset** input should be high and the **reset** input low. When the **reset** input is high it resets the count to zero (0000, QA-QD low)

The counter may be **preset** by placing the desired binary number on the **inputs A-D** and briefly making the **preset** input low. Note that a clock pulse is not required to preset, unlike the 74160-3 counters. The **inputs A-D** may be left unconnected if not required.

Connecting counters with separate up and down clock inputs in a chain

The diagram below shows how to link 74192-3 up/down counters with separate up and down clock inputs, notice how **carry** and **borrow** are connected to the **up clock** and **down clock** inputs respectively of the next counter.

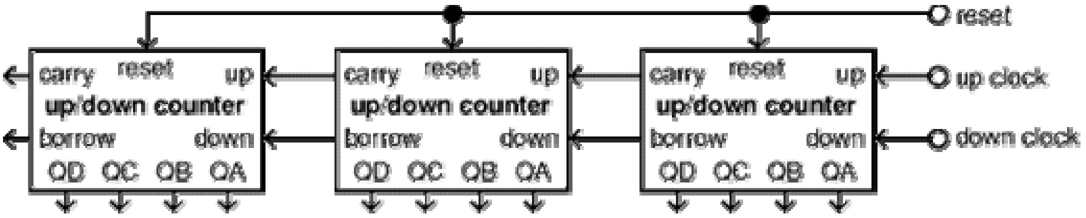
reset
up clock
down clock
carry
reset
up
up/down counter
borrow
down
QD QC QB QA

74HC4017 decade counter (1-of-10)
74HC4020 14-bit ripple counter
74HC4040 12-bit ripple counter
74HC4060 14-bit ripple counter with internal oscillator

These are the 74HC equivalents of 4000 series CMOS counters. Like all 74HC ICs they need a power supply of 2 to 6V. For pin connections and functions see: 4017, 4020, 4040, 4060

Decoders

7442 BCD to Decimal (1 of 10) Decoder

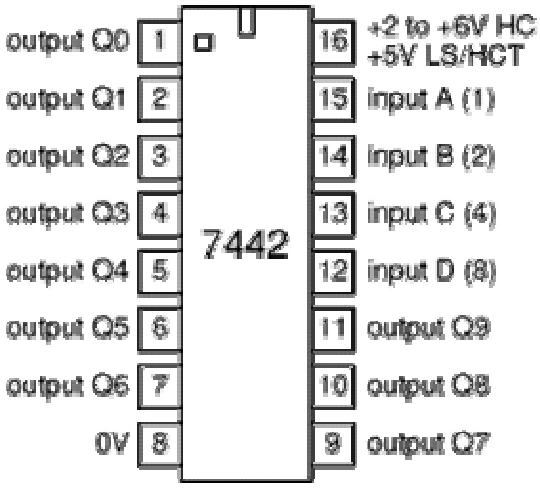
output Q0
1
output Q1
2
output Q2
3
output Q3
4
output Q4
5
output Q5
6
output Q6
7
0V
8
7442
16
+2 to +6V HC
+5V LS/HCT
15
input A (1)
14
input B (2)
13
input C (4)
12
input D (8)
11
output Q9
10
output Q8
9
output Q7

The 7442 **outputs** are **active-low** which means they become low when selected but are high at other times. They can sink up to about 20mA.

The appropriate output becomes low in response to the BCD (binary coded decimal) input. For example an input of binary 0101 (=5) will make output Q5 low and all other outputs high.

The 7442 is a BCD (binary coded decimal) decoder intended for input values 0 to 9 (0000 to 1001 in binary). With inputs from 10 to 15 (1010 to 1111 in binary) all outputs are high.

Note that the 7442 can be used as a **1-of-8 decoder** if input D is held low.

Also see: 74HC4017 and 4017 both are a decade counter and 1-of-10 decoder in a single IC.

7-Segment Display Drivers

7447 BCD to 7-Segment Display Driver

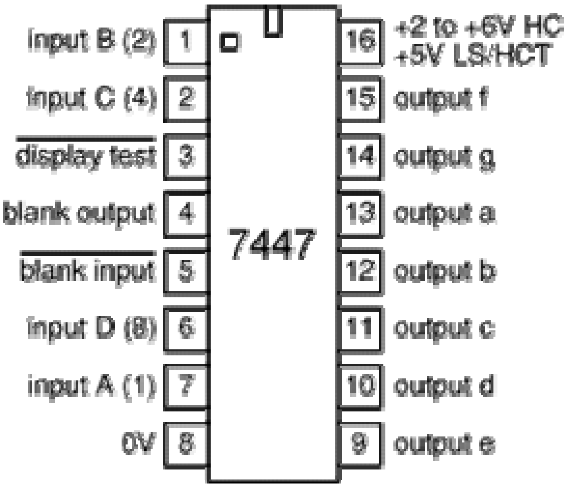
input B (2)
1
input C (4)
2
display test
3
blank output
4
blank input
5
input D (8)
6
input A (1)
7
0V
8
7447
16
+2 to +6V HC
+5V LS/HCT
15
output f
14
output g
13
output a
12
output b
11
output c
10
output d
9
output e

The appropriate **outputs a-g** becomes low to display the BCD (binary coded decimal) number supplied on **inputs A-D**. The 7447 has **open collector** outputs a-g which can sink up to 40mA. The 7-segment display segments must be connected between +Vs and the outputs with a resistor in series (330Ω with a 5V supply). A **common anode** display is required.

Display test and **blank input** are active-low so they should be high for normal operation. When **display test** is low all the display segments should light (showing number 8).

If the **blank input** is low the display will be blank when the count input is zero (0000). This can be used to blank leading zeros when there are several display digits driven by a chain of counters. To achieve this **blank output** should be connected to **blank input** of the next display down the chain (the next most significant digit).

The 7447 is intended for BCD (binary coded decimal) which is input values 0 to 9 (0000 to 1001 in binary). Inputs from 10 to 15 (1010 to 1111 in binary) will light odd display segments but will do no harm.

74HC4511 BCD to 7-Segment Display Driver

This is the 74HC equivalent of the CMOS 4511 display driver. Like all 74HC ICs it needs a power supply of 2 to 6V. For pin connections and functions see 4511.

4000 Series CMOS Logic ICs

This family of logic ICs is numbered from 4000 onwards, and from 4500 onwards. They have a B at the end of the number (e.g. 4001B) which refers to an improved design introduced some years ago. Most of them are in 14-pin or 16-pin packages. They use **CMOS circuitry** which means they use very little power and can tolerate a wide range of power supply voltages (3 to 15V) making them **ideal for battery powered projects**. CMOS is pronounced 'see-moss' and stands for Complementary Metal Oxide Semiconductor.

General characteristics

- **Supply:** 3 to 15V, small fluctuations are tolerated.
- **Inputs** have very high impedance (resistance), this is good because it means they will not affect the part of the circuit where they are connected. However, it also means that unconnected inputs can easily pick up electrical noise and rapidly change between high and low states in an unpredictable way. This is likely to make the chip behave erratically and it will significantly increase the supply current. To prevent problems all unused inputs MUST be connected to the supply (either +Vs or 0V), this applies even if that part of the chip is not being used in the circuit!
- **Outputs** can sink and source only about 1mA if you wish to maintain the correct output voltage to drive CMOS inputs. If there is no need to drive any inputs the maximum current is about 5mA with a 6V supply, or 10mA with a 9V supply (just enough to light an LED). To switch larger currents you can connect a transistor.
- **Fan-out:** one output can drive up to 50 inputs.
- **Gate propagation time**: typically 30ns for a signal to travel through a gate with a 9V supply, it takes a longer time at lower supply voltages.
- **Frequency**: up to 1MHz, above that the 74 series is a better choice.
- **Power consumption** (of the chip itself) is very low, a few µW. It is much greater at high frequencies, a few mW at 1MHz for example.

There are many ICs in the 4000 series and this page only covers a selection, concentrating on the most useful gates, counters, decoders and display drivers. For each IC there is a diagram showing the pin arrangement and brief notes explain the function of the pins where necessary. The notes also explain if the IC's properties differ substantially from the standard characteristics listed above.

If you are using another reference please be aware that there is some variation in the terms used to describe input pins. I have tried to be logically consistent so the term I have used describes the pin's function when high (true). For example 'disable clock' on the 4026 is often labeled 'clock enable' but this can be confusing because it enables the clock when low (false). An input described as 'active low' is like this, it performs its function when low. If you see a line drawn above a label it means it is active low, for example: $\overline{\text{reset}}$ (say 'reset-bar').

Static Precautions

The CMOS circuitry means that 4000 series ICs are static sensitive. Touching a pin while charged with static electricity (from your clothes for example) may damage the IC. In fact most ICs in regular use are quite tolerant and earthing your hands by touching a metal water pipe or window frame before handling them will be adequate. ICs should be left in their protective packaging until you are ready to use them.

Gates

Quad 2-Input Gates

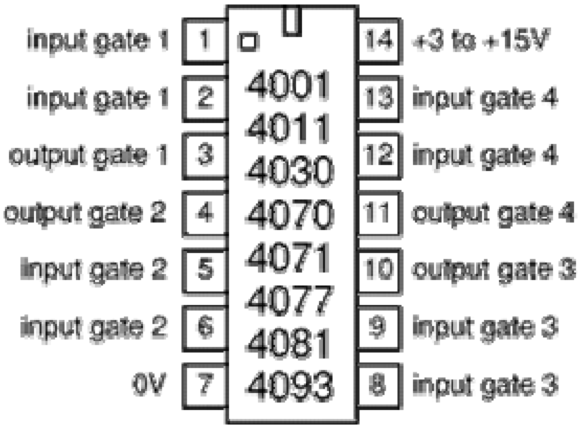

input gate 1
1
input gate 1
2
output gate 1
3
output gate 2
4
input gate 2
5
input gate 2
6
0V
7
4001
4011
4030
4070
4071
4077
4081
4093
14
+3 to +15V
13
input gate 4
12
input gate 4
11
output gate 4
10
output gate 3
9
input gate 3
8
input gate 3

- 4001 quad 2-input NOR
- 4011 quad 2-input NAND
- 4030 quad 2-input EX-OR (now obsolete)
- 4070 quad 2-input EX-OR
- 4071 quad 2-input OR
- 4077 quad 2-input EX-NOR
- 4081 quad 2-input AND
- 4093 quad 2-input NAND with Schmitt trigger inputs

The 4093 has Schmitt trigger inputs to provide good noise immunity. They are ideal for slowly changing or noisy signals. The hysteresis is about 0.5V with a 4.5V supply and almost 2V with a 9V supply.

Triple 3-Input Gates

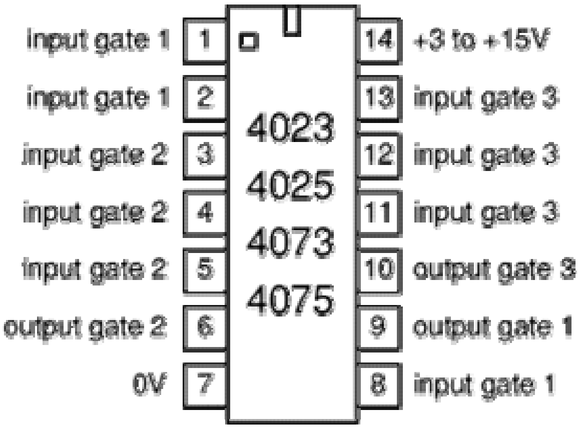

input gate 1
1
input gate 1
2
input gate 2
3
input gate 2
4
input gate 2
5
output gate 2
6
0V
7
4023
4025
4073
4075
14
+3 to +15V
13
input gate 3
12
input gate 3
11
input gate 3
10
output gate 3
9
output gate 1
8
input gate 1

- 4023 triple 3-input NAND
- 4025 triple 3-input NOR
- 4073 triple 3-input AND
- 4075 triple 3-input OR

Dual 4-Input Gates

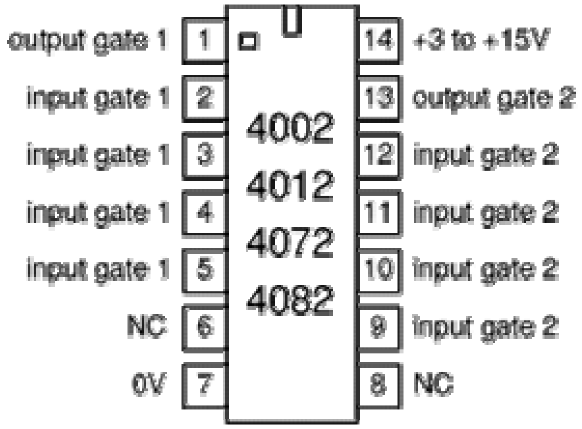
output gate 1
1
input gate 1
2
input gate 1
3
input gate 1
4
input gate 1
5
NC
6
0V
7
4002
4012
4072
4082
14
+3 to +15V
13
output gate 2
12
input gate 2
11
input gate 2
10
input gate 2
9
input gate 2
8
NC

- 4002 dual 4-input NOR
- 4012 dual 4-input NAND
- 4072 dual 4-input OR
- 4082 dual 4-input AND

NC = **N**o **C**onnection (a pin that is not used).

4068 8-Input NAND/AND* Gate

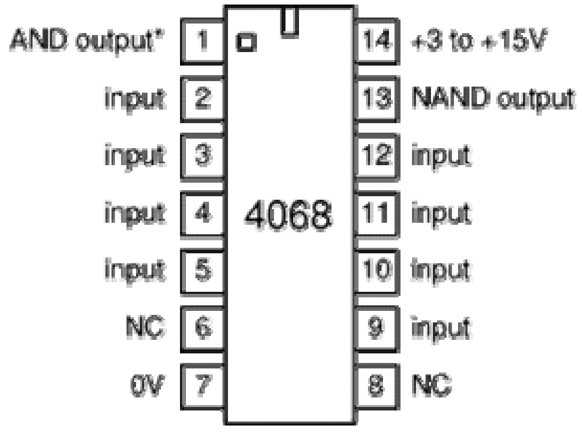
AND output*
1
input
2
input
3
input
4
input
5
NC
6
0V
7
4068
14
+3 to +15V
13
NAND output
12
input
11
input
10
input
9
input
8
NC

This gate has a propagation time which is about 10 times longer than normal so it is not suitable for high speed circuits.

NC = **N**o **C**onnection (a pin that is not used).

***** = The AND output (pin 1) is not available on some versions of the 4068.

4069 Hex NOT (Inverting Buffer)

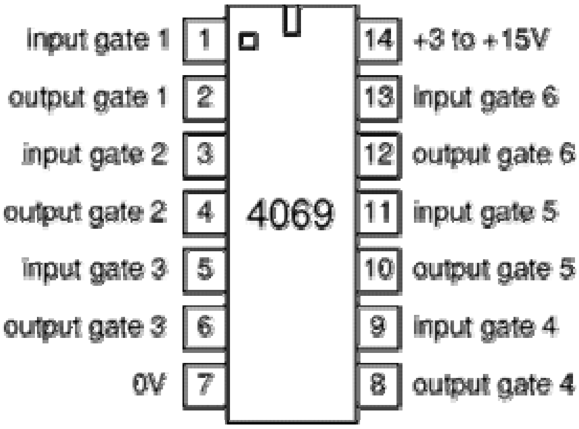
input gate 1
output gate 1
input gate 2
output gate 2
input gate 3
output gate 3
0V
1
2
3
4
5
6
7
4069
14
13
12
11
10
9
8
+3 to +15V
input gate 6
output gate 6
input gate 5
output gate 5
input gate 4
output gate 4

4049 Hex NOT and 4050 Hex Buffer

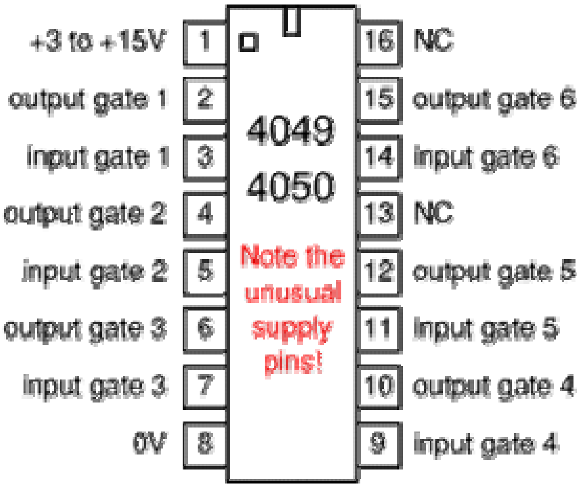
+3 to +15V
output gate 1
input gate 1
output gate 2
input gate 2
output gate 3
input gate 3
0V
1
2
3
4
5
6
7
8
4049
4050
Note the unusual supply pins!
16
15
14
13
12
11
10
9
NC
output gate 6
input gate 6
NC
output gate 5
input gate 5
output gate 4
input gate 4

- 4049 hex NOT (inverting buffer)
- 4050 hex non-inverting buffer

Inputs: These ICs are unusual because their gate inputs can withstand up to +15V even if the power supply is a lower voltage.

Outputs: These ICs are unusual because they are capable of driving 74LS gate inputs directly. To do this they must have a +5V supply (74LS supply voltage). The gate output is sufficient to drive four 74LS inputs.

NC = **N**o **C**onnection (a pin that is not used).

Note the unusual arrangement of the power supply pins for these ICs!

4000 Dual 3-Input NOR Gate and NOT Gate

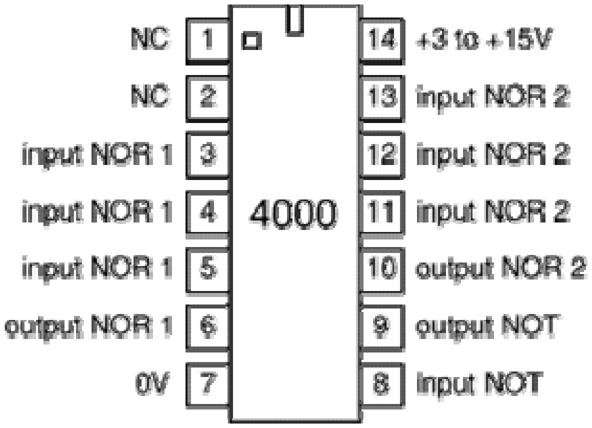

NC 1
NC 2
input NOR 1 3
input NOR 1 4
input NOR 1 5
output NOR 1 6
0V 7
4000
14 +3 to +15V
13 input NOR 2
12 input NOR 2
11 input NOR 2
10 output NOR 2
9 output NOT
8 input NOT

Two 3-input NOR gates and a single NOT gate in one package.

NC = **N**o **C**onnection (a pin that is not used).

Decade and 4-Bit Counters

4017 Decade Counter (1-of-10)

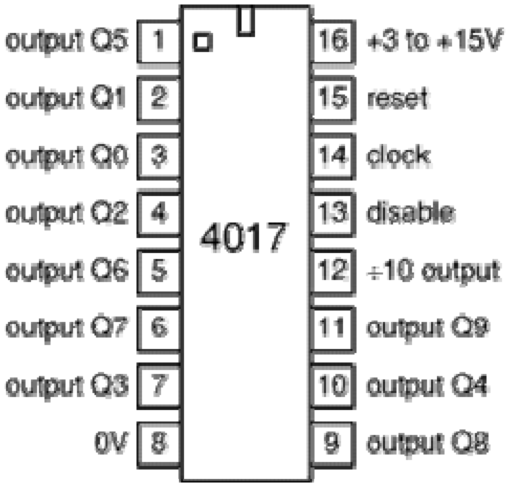

output Q5 1
output Q1 2
output Q0 3
output Q2 4
output Q6 5
output Q7 6
output Q3 7
0V 8
4017
16 +3 to +15V
15 reset
14 clock
13 disable
12 ÷10 output
11 output Q9
10 output Q4
9 output Q8

The count advances as the **clock** input becomes high (on the rising-edge). Each **output** Q0-Q9 goes high in turn as counting advances. For some functions (such as flash sequences) outputs may be combined using diodes.

The **reset** input should be low (0V) for normal operation (counting 0-9). When high it resets the count to zero (Q0 high). This can be done manually with a switch between reset and +Vs and a 10k resistor between reset and 0V. **Counting to less than 9** is achieved by connecting the relevant output (Q0-Q9) to reset, for example to count 0,1,2,3 connect Q4 to reset.

The **disable** input should be low (0V) for normal operation. When high it disables counting so that clock pulses are ignored and the count is kept constant.

The **÷10 output** is high for counts 0-4 and low for 5-9, so it provides an output at $^1/_{10}$ of the clock frequency. It can be used to drive the clock input of another 4017 (to count the tens).

4026 Decade Counter and 7-Segment Display Driver

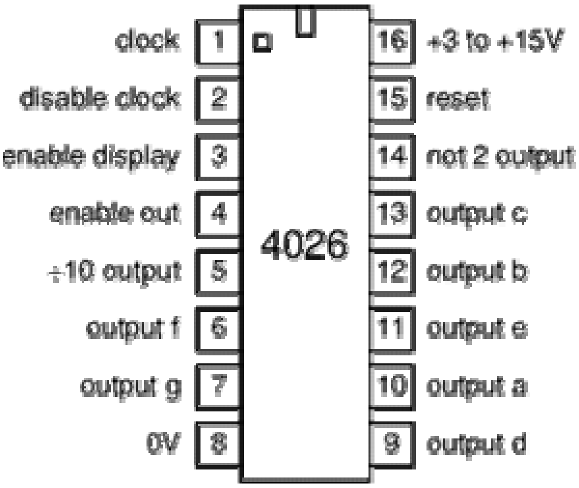
clock
1
disable clock
2
enable display
3
enable out
4
÷10 output
5
output f
6
output g
7
0V
8
4026
16
+3 to +15V
15
reset
14
not 2 output
13
output c
12
output b
11
output e
10
output a
9
output d

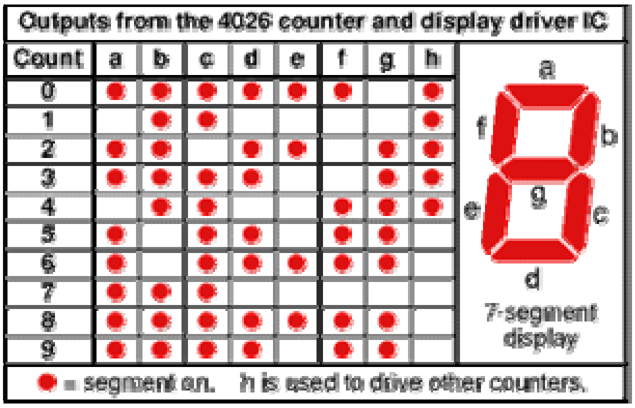
Outputs from the 4026 counter and display driver IC

Count	a	b	c	d	e	f	g	h
0	●	●	●	●	●	●		●
1		●	●					●
2	●	●		●	●		●	●
3	●	●	●	●			●	●
4		●	●			●	●	●
5	●		●	●		●	●	
6	●		●	●	●	●	●	
7	●	●	●					
8	●	●	●	●	●	●	●	
9	●	●	●	●		●	●	

● = segment on. h is used to drive other counters.

a
f
b
g
e
c
d
7-segment display

The count advances as the **clock** input becomes high (on the rising-edge). The **outputs a-g** go high to light the appropriate segments of a common-cathode 7-segment display as the count advances. The maximum output current is about 1mA with a 4.5V supply and 4mA with a 9V supply. This is sufficient to directly drive many 7-segment LED displays. The table below shows the segment sequence in detail.

The **reset** input should be low (0V) for normal operation (counting 0-9). When high it resets the count to zero.

The **disable clock** input should be low (0V) for normal operation. When high it disables counting so that clock pulses are ignored and the count is kept constant.

The **enable display** input should be high (+Vs) for normal operation. When low it makes outputs a-g low, giving a blank display. The **enable out** follows this input but with a brief delay.

The **÷10 output** (h in table) is high for counts 0-4 and low for 5-9, so it provides an output at $^1/_{10}$ of the clock frequency. It can be used to drive the clock input of another 4026 to provide multi-digit counting.

The **not 2 output** is high unless the count is 2 when it goes low.

4029 Up/Down Synchronous Counter with Preset

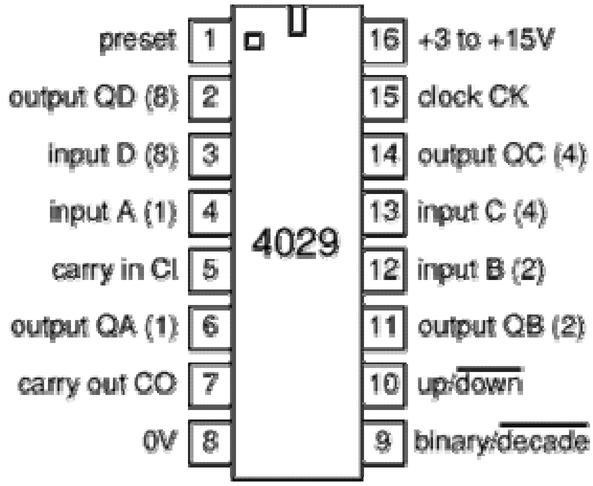
preset 1
output QD (8) 2
input D (8) 3
input A (1) 4
carry in CI 5
output QA (1) 6
carry out CO 7
0V 8
4029
16 +3 to +15V
15 clock CK
14 output QC (4)
13 input C (4)
12 input B (2)
11 output QB (2)
10 up/down
9 binary/decade

The 4029 is a **synchronous** counter so its outputs change precisely together on each clock pulse. This is helpful if you need to connect the outputs to logic gates because it avoids the glitches which occur with ripple counters.

The count occurs as the **clock** input becomes high (on the rising-edge). The **up/down** input determines the direction of counting: high for up, low for down. The state of **up/down** should be changed when the clock is high.

For normal operation (counting) **preset**, and **carry in** should be low.

The **binary/decade** input selects the type of counter: 4-bit binary (0-15) when high; decade (0-9) when low.

The counter may be **preset** by placing the desired binary number on the **inputs A-D** and briefly making the **preset** input high. There is no reset input, but **preset** can be used to reset the count to zero if **inputs A-D** are all low.

4510 Up/Down Decade (0-9) Counter with Preset
4516 Up/Down 4-Bit (0-15) Counter with Preset

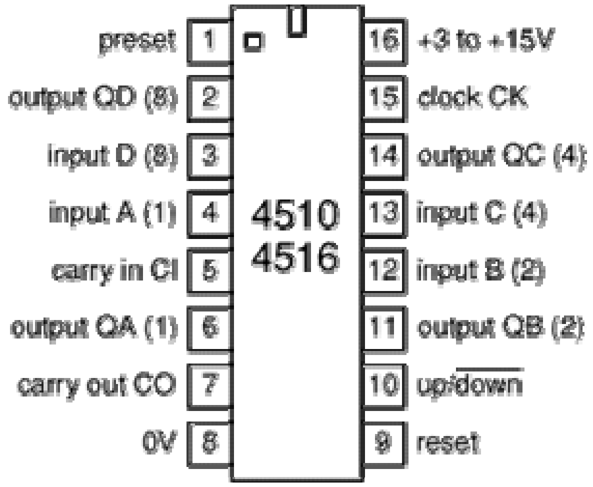
preset 1
output QD (8) 2
input D (8) 3
input A (1) 4
carry in CI 5
output QA (1) 6
carry out CO 7
0V 8
4510
4516
16 +3 to +15V
15 clock CK
14 output QC (4)
13 input C (4)
12 input B (2)
11 output QB (2)
10 up/down
9 reset

These are **synchronous** counters so their outputs change precisely together on each clock pulse. This is helpful if you need to connect their outputs to logic gates because it avoids the glitches which occur with ripple counters.

The count occurs as the **clock** input becomes high (on the rising-edge). The **up/down** input determines the direction of counting: high for up, low for down. The state of **up/down** should be changed when the clock is high.

For normal operation (counting) **preset**, **reset** and **carry in** should be low. When **reset** is high it resets the count to zero (0000, QA-QD low). The **clock** input should be low when resetting.

The counter may be **preset** by placing the desired binary number on the **inputs A-D** and briefly making the **preset** input high, the **clock** input should be low when this happens.

Connecting synchronous counters in a chain

The diagram below shows how to link synchronous counters, notice how all the **clock (CK)** inputs are linked. **Carry out (CO)** feeds **carry in (CI)** of the next counter. **Carry in (CI)** of the first counter should be low for 4029, 4510 and 4516 counters.

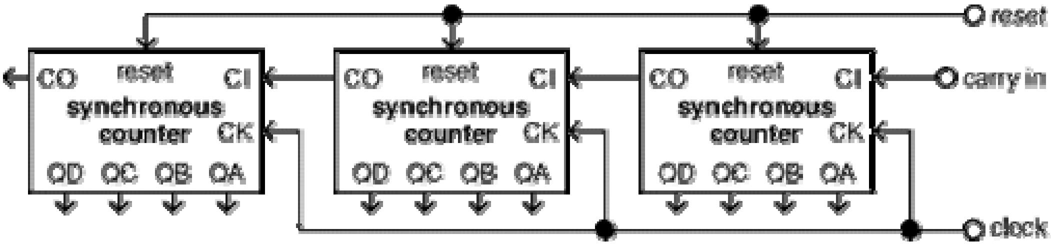
reset
carry in
clock
CO
reset
CI
synchronous
counter
CK
QD
QC
QB
QA
CO
reset
CI
synchronous
counter
CK
QD
QC
QB
QA
CO
reset
CI
synchronous
counter
CK
QD
QC
QB
QA

4518 Dual Decade (0-9) Counter
4520 Dual 4-Bit (0-15) Counter

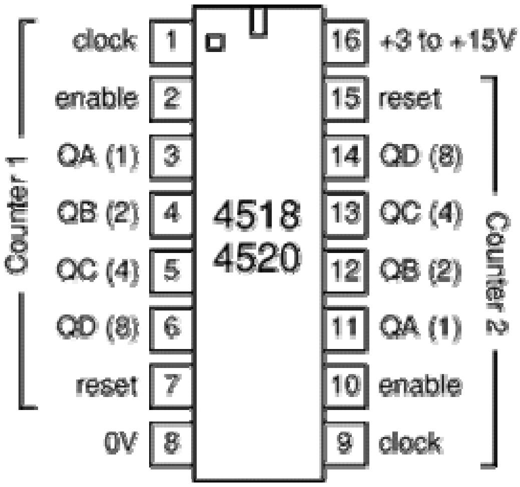
Counter 1
clock
1
enable
2
QA (1)
3
QB (2)
4
QC (4)
5
QD (8)
6
reset
7
0V
8
4518
4520
16
+3 to +15V
15
reset
14
QD (8)
13
QC (4)
12
QB (2)
11
QA (1)
10
enable
9
clock
Counter 2

These contain two separate **synchronous** counters, one on each side of the chip.

Normally a clock signal is connected to the **clock** input, with the **enable** input held high. Counting advances as the clock signal becomes high (on the rising-edge). Special arrangements are used if the 4518/20 counters are linked in a chain, as explained below.

For normal operation the **reset** input should be low, making it high resets the counter to zero (0000, QA-QD low).

Counting to less than the maximum (9 or 15) can be achieved by connecting the appropriate output(s) to the reset input, using an AND gate if necessary. For example to count 0 to 8 connect QA (1) and QD (8) to reset using an AND gate.

Connecting 4518 and 4520 counters in a chain

The diagram below shows how to link 4518 and 4520 counters. Notice how the normal **clock** inputs are held low, with the **enable** inputs being used instead. With this arrangement counting advances as the enable input becomes low (on the falling-edge) allowing **output QD** to supply a clock signal to the next counter. The complete chain is a ripple counter, although the individual counters are synchronous! If it is essential to have truly synchronous counting a system of logic gates is required, please see a 4518/20 datasheet for further details.

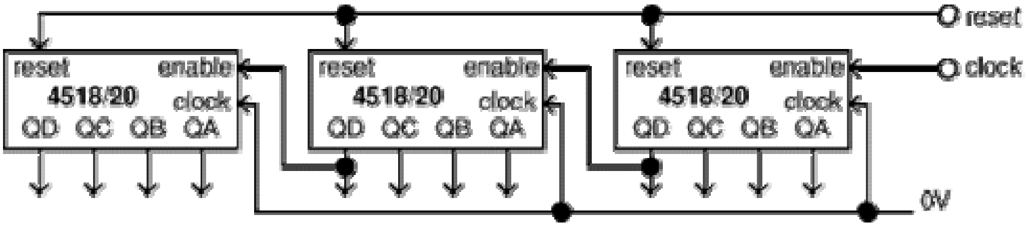
reset
clock
0V
reset
enable
4518/20
clock
QD
QC
QB
QA
reset
enable
4518/20
clock
QD
QC
QB
QA
reset
enable
4518/20
clock
QD
QC
QB
QA

7-Bit, 12-Bit and 14-Bit Counters

4020 14-Bit (÷16,384) Ripple Counter

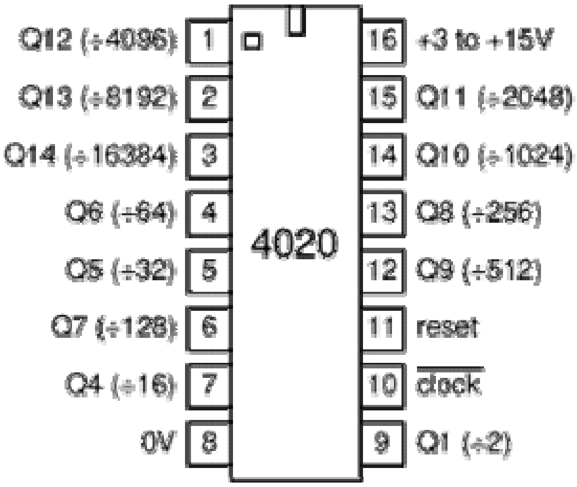
Q12 (÷4096)
1
Q13 (÷8192)
2
Q14 (÷16384)
3
Q6 (÷64)
4
Q5 (÷32)
5
Q7 (÷128)
6
Q4 (÷16)
7
0V
8
4020
16
+3 to +15V
15
Q11 (÷2048)
14
Q10 (÷1024)
13
Q8 (÷256)
12
Q9 (÷512)
11
reset
10
$\overline{\text{clock}}$
9
Q1 (÷2)

The 4020 is a **ripple** counter so beware that glitches may occur in any logic gate systems connected to its outputs due to the slight delay before the later counter outputs respond to a clock pulse.

The count advances as the **clock** input becomes low (on the falling-edge), this is indicated by the bar over the clock label. This is the usual clock behavior of ripple counters and it means a counter output can directly drive the clock input of the next counter in a chain.

Output Qn is the nth stage of the counter, representing 2^n, for example Q4 is 2^4 = 16 ($^1/_{16}$ of clock frequency) and Q14 is 2^{14} = 16384 ($^1/_{16384}$ of clock frequency). Note that Q2 and Q3 are not available.

The **reset** input should be low for normal operation (counting). When high it resets the count to zero (all outputs low).

Also see: 4040 (12-bit) and 4060 (14-bit with internal oscillator).

4024 7-Bit (÷128) Ripple Counter

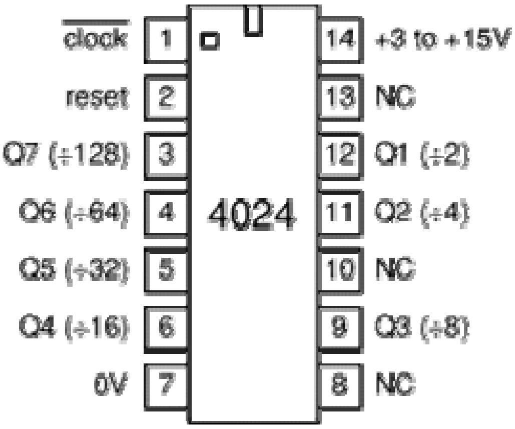

The 4024 is a **ripple** counter so beware that glitches may occur in any logic gate systems connected to its outputs due to the slight delay before the later counter outputs respond to a clock pulse.

The count advances as the **clock** input becomes low (on the falling-edge), this is indicated by the bar over the clock label. This is the usual clock behavior of ripple counters and it means a counter output can directly drive the clock input of the next counter in a chain.

Output Qn is the nth stage of the counter, representing 2^n, for example Q4 is 2^4 = 16 ($^1/_{16}$ of clock frequency) and Q7 is 2^7 = 128 ($^1/_{128}$ of clock frequency).

The **reset** input should be low for normal operation (counting). When high it resets the count to zero (all outputs low).

4040 12-Bit (÷4096) Ripple Counter

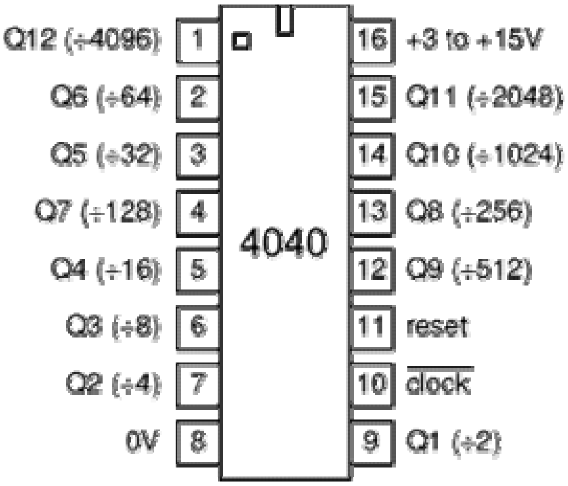

Q12 (÷4096)
Q6 (÷64)
Q5 (÷32)
Q7 (÷128)
Q4 (÷16)
Q3 (÷8)
Q2 (÷4)
0V
1
2
3
4
5
6
7
8
4040
16
15
14
13
12
11
10
9
+3 to +15V
Q11 (÷2048)
Q10 (÷1024)
Q8 (÷256)
Q9 (÷512)
reset
clock
Q1 (÷2)

The 4040 is a **ripple** counter so beware that glitches may occur in any logic gate systems connected to its outputs due to the slight delay before the later counter outputs respond to a clock pulse.

The count advances as the **clock** input becomes low (on the falling-edge), this is indicated by the bar over the clock label. This is the usual clock behavior of ripple counters and it means a counter output can directly drive the clock input of the next counter in a chain.

Output Qn is the nth stage of the counter, representing 2^n, for example Q4 is $2^4 = 16$ ($^1/_{16}$ of clock frequency) and Q12 is $2^{12} = 4096$ ($^1/_{4096}$ of clock frequency).

The **reset** input should be low for normal operation (counting). When high it resets the count to zero (all outputs low).

4060 14-Bit (÷16,384) Ripple Counter with Internal Oscillator

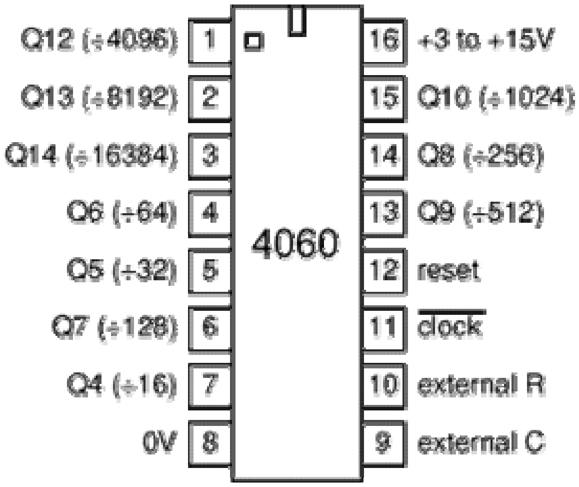

Q12 (÷4096)
Q13 (÷8192)
Q14 (÷16384)
Q6 (÷64)
Q5 (÷32)
Q7 (÷128)
Q4 (÷16)
0V
1
2
3
4
5
6
7
8
4060
16
15
14
13
12
11
10
9
+3 to +15V
Q10 (÷1024)
Q8 (÷256)
Q9 (÷512)
reset
clock
external R
external C

The 4060 is a **ripple** counter so beware that glitches may occur in any logic gate systems connected to its outputs due to the slight delay before the later counter outputs respond to a clock pulse.

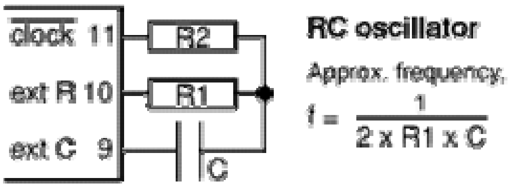
clock 11
ext R 10
ext C 9
R2
R1
C
RC oscillator
Approx. frequency,
f = 1 / 2 x R1 x C

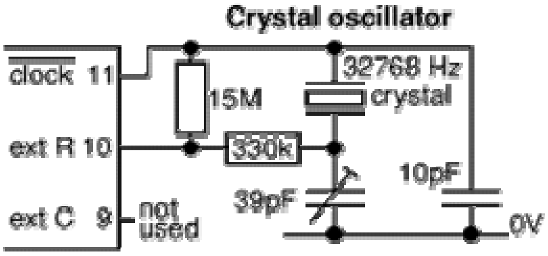
Crystal oscillator
clock 11
ext R 10
ext C 9
not used
15M
330k
32768 Hz crystal
10pF
39pF
0V

The count advances as the **clock** input becomes low (on the falling-edge), this is indicated by the bar over the clock label. This is the usual clock behavior of ripple counters and it means a counter output can directly drive the clock input of the next counter in a chain. The clock can be driven directly, or connected to the **internal oscillator** (see below).

Output Qn is the nth stage of the counter, representing 2^n, for example Q4 is $2^4 = 16$ ($^1/_{16}$ of clock frequency) and Q14 is $2^{14} = 16384$ ($^1/_{16384}$ of clock frequency). Note that Q1-3 and Q11 are not available.

The **reset** input should be low for normal operation (counting). When high it resets the count to zero (all outputs low).

The 4060 includes an **internal oscillator**. The **clock** signal may be supplied in three ways:

- **From an external source** to the clock input, as for a normal counter. In this case there should be no connections to **external C** and **external R** (pins 9 and 10).
- **RC oscillator** as shown in the diagram. The oscillator drives the clock input with an approximate frequency f = $^1/_{(2\times R1\times C)}$ (it partly depends on the supply voltage). R1 should be at least 50kΩ if the supply voltage is less than 7V. R2 should be between 2 and 10 times R1.
- **Crystal oscillator** as shown in the diagram, note that there is no connection to pin 9. The 32768 Hz crystal will give a 2Hz signal at the last output, Q14.

Decoders

4028 BCD to Decimal (1 of 10) Decoder

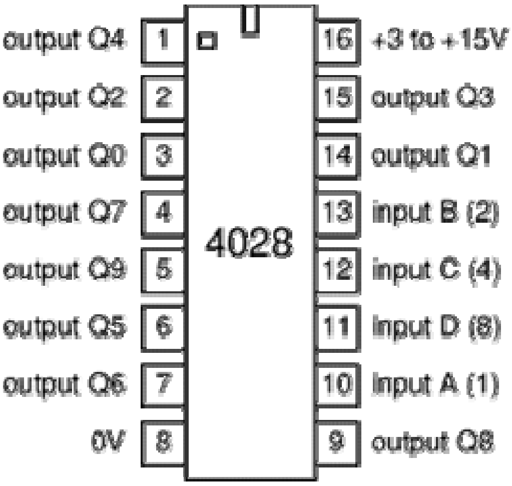
output Q4 1
output Q2 2
output Q0 3
output Q7 4
output Q9 5
output Q5 6
output Q6 7
0V 8
4028
16 +3 to +15V
15 output Q3
14 output Q1
13 input B (2)
12 input C (4)
11 input D (8)
10 input A (1)
9 output Q8

The appropriate **output Q0-9** becomes high in response to the BCD (binary coded decimal) input. For example an input of binary 0101 (=5) will make output Q5 high and all other outputs low.

The 4028 is a BCD (binary coded decimal) decoder intended for input values 0 to 9 (0000 to 1001 in binary). With inputs from 10 to 15 (1010 to 1111 in binary) all outputs are low.

Note that the 4028 can be used as a **1-of-8 decoder** if input D is held low.

Also see: 4017 (a decade counter and 1-of-10 decoder in a single chip).

7-Segment Display Drivers

4511 BCD to 7-Segment Display Driver

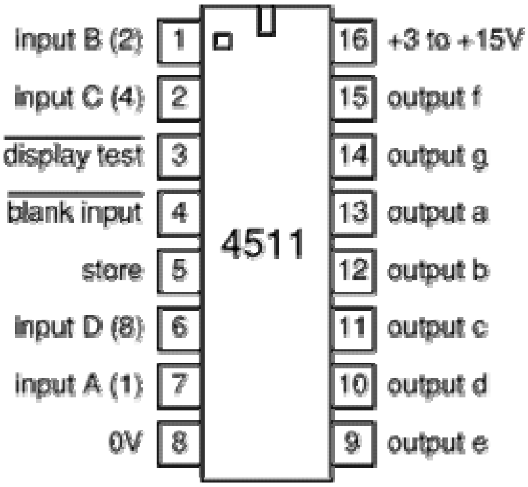
input B (2)
1
16
+3 to +15V
input C (4)
2
15
output f
display test
3
14
output g
blank input
4
13
output a
4511
store
5
12
output b
input D (8)
6
11
output c
input A (1)
7
10
output d
0V
8
9
output e

The appropriate **outputs a-g** becomes high to display the BCD (binary coded decimal) number supplied on **inputs A-D**. The **outputs a-g** can source up to 25mA. The 7-segment display segments must be connected between the outputs and 0V with a resistor in series (330Ω with a 5V supply). A **common cathode** display is required.

Display test and **blank input** are active-low so they should be high for normal operation. When **display test** is low all the display segments should light (showing number 8). When **blank input** is low the display will be blank (all segments off).

The **store** input should be low for normal operation. When **store** is high the displayed number is stored internally to give a constant display regardless of any changes which may occur to the **inputs A-D**.

The 4511 is intended for BCD (binary coded decimal). Inputs values from 10 to 15 (1010 to 1111 in binary) will give a blank display (all segments off).

Lamps

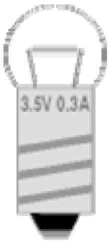

Function and Construction

Lamps emit light when an electric current passes through them. All of the lamps shown on this page have a thin wire **filament** which becomes very hot and glows brightly when a current passes through it. The filament is made from a metal with a high melting point such as tungsten and it is usually wound into a small coil. Filament lamps have a shorter lifetime than most electronic components because eventually the filament 'blows' (melts) at a weak point.

Circuit Symbols

There are two circuit symbols for a lamp, one for a lamp used to provide illumination and another for a lamp used as an indicator. Small lamps such as torch bulbs can be used for both purposes so either circuit symbol may use in simple educational circuits.

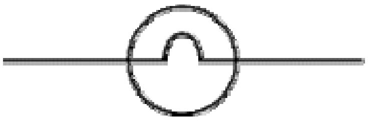

Lamp used for lighting
(for example a car headlamp or torch bulb)

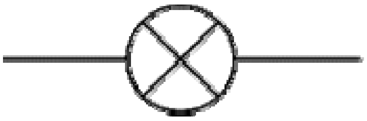

Lamp used as an indicator
(for example a warning light on a car dashboard)

Selecting a Lamp

There are three important features to consider when selecting a lamp:

- **Voltage rating** - the supply voltage for normal brightness.
- **Power or current rating** - small lamps are usually rated by current.
- **Lamp type** - please see the table below.

The voltage and power (or current) ratings are usually printed or embossed on the body of a lamp.

Voltage Rating

This is the supply voltage required for normal brightness. If a slightly higher voltage is used the lamp will be brighter but its lifetime will be shorter. With a lower supply voltage the lamp will be dimmer and its lifetime will be longer. The light from dim lamps has a yellow-orange color.

Torch lamps pass a relatively large current and this significantly reduces the output voltage of the battery. Some voltage is used up inside the battery driving the large current through the small resistance of the battery itself (its 'internal resistance'). As a result the correct voltage rating for a torch lamp is lower than the normal voltage of the battery which lights it!

For example: a lamp rated 3.5V 0.3A is correct for a 4.5V battery (three 1.5V cells) because when the lamp is connected the voltage across the battery falls to about 3.5V.

Power or Current Rating

This is the power or current for the lamp when connected to its rated voltage. Low power lamps are usually rated by their current and high power lamps by their power. It is easy to convert between the two ratings:

P = I × V where: P = power in watts (W)
or I = current in amps (A)
I = P / V V = voltage in volts (V)

Examples:

- A lamp rated **3.5V 0.3A** has a power rating P = I × V = 0.3 × 3.5 = **1.05W**
- A lamp rated **6V 0.06A** has a power rating P = I × V = 0.06 × 6 = **0.36W**
- A lamp rated **12W 2.4W** has a current rating I = P / V = 2.4 / 12 = **0.2A**

Lamp Type

Type of Lamp	**Example**
MES Miniature Edison Screw These are the standard small lamps. The bulb diameter is usually about 10mm, but tubular bulbs are also available. MES lamps have one contact on the base and the body forms the other contact. They are available with a good range of voltage and power (or current) ratings. Lens ended versions are available to produce a focused beam of light. **LES** Lilliput Edison Screw Smaller than MES, these have a bulb diameter of about 5mm.	

MCC Miniature Centre Contact These have a bayonet style fitting, like a standard mains lamp in the UK. They have one contact on the base and the body forms the other contact. The bulb diameter is about 10mm.	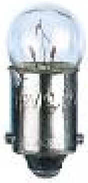
SBC Small Bayonet Cap These have a bayonet style fitting, like a standard mains lamp in the UK. They have two contacts on the base so the metal body is not connected in the circuit. SBC lamps have high power ratings (24W for example) and their bulbs are large with a diameter of up to about 40mm. Note the two filament arrangements in the lamps shown, horizontal on the left, vertical on the right.	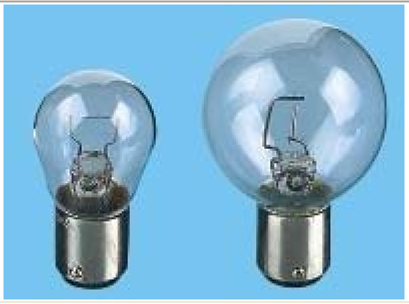
Pre-focus This type of lamp is used in torches and lanterns. The flange at the top of the metal body is used to hold the lamp in place. Lampholders are not readily available so this type is unsuitable for most projects.	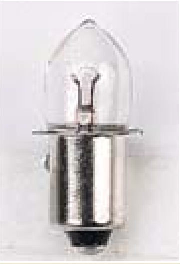
Wire ended These are very small lamps with a bulb about 3mm diameter and 6mm long. Take care to avoid snapping the wires where they enter the glass bulb.	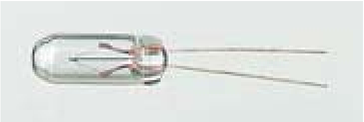
Grain of Wheat These are similar to the wire ended lamps above but they have stranded wire leads usually about 150mm long. The bulb is about 3mm diameter and 6mm long - the size of a grain of wheat!	

Connecting and Soldering

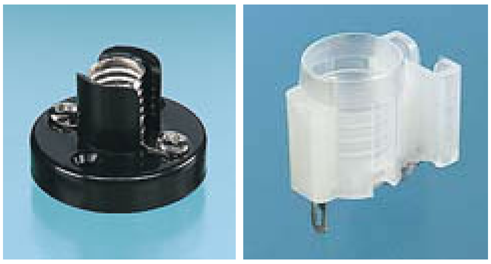

screw terminals solder tags

Lampholders

Lamps may be connected either way round in a circuit and the supply may be AC or DC.

Most lamps are designed to be used in a lampholder but the small 'wire ended' and 'grain of wheat' lamps have wires which may be soldered directly onto a circuit board.

Lampholders usually have screw terminals or solder tags to attach wires. Some small holders have contacts which may be soldered directly to a circuit board.

Lamps in Series

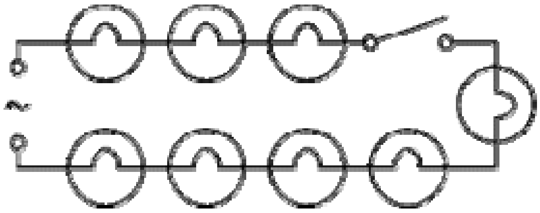

Several lamps can be successfully connected in series provided they all have identical voltage and power (or current) ratings. The supply voltage is divided equally between identical lamps so their voltage rating must be suitable for this. For example Christmas tree lights may have 20 lamps connected in series to a 240V supply, so each lamp will have 240V ÷ 20 = 12V across it.

A disadvantage of connecting lamps in series is that if one lamp blows all of them will go out because the circuit is broken. Christmas tree lamps have a special feature to overcome this problem; they are designed to short circuit (conduct like a wire link) when they blow, so the circuit is not broken and the other lamps remain lit, making it easier to locate the faulty lamp. Sets also include one 'fuse' lamp which blows normally.

WARNING! The Christmas tree lamps may seem safe because they use only 12V but they are connected to the mains supply which can be lethal. Always unplug from the mains before changing lamps. The voltage across the holder of a missing lamp is the full 240V of the mains supply!

Light Emitting Diodes (LEDs)

Example:

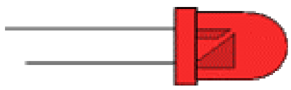

Circuit symbol:

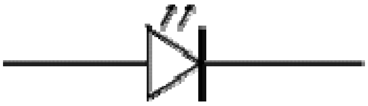

Function

LEDs emit light when an electric current passes through them.

Connecting and Soldering

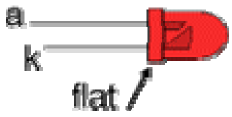

LEDs must be connected the correct way round, the diagram may be labeled **a** or **+** for anode and **k** or **-** for cathode (yes, it really is k, not c, for cathode!). The cathode is the short lead and there may be a slight flat on the body of round LEDs. If you can see inside the LED the cathode is the larger electrode (but this is not an official identification method).

LEDs can be damaged by heat when soldering, but the risk is small unless you are very slow. No special precautions are needed for soldering most LEDs.

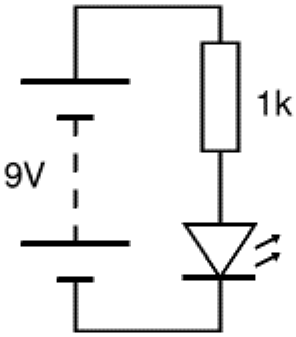

Testing an LED

Never connect an LED directly to a battery or power supply!
It will be destroyed almost instantly because too much current will pass through and burn it out.

LEDs must have a resistor in series to limit the current to a safe value, for quick testing purposes a 1kΩ resistor is suitable for most LEDs if your supply voltage is 12V or less.
Remember to connect the LED the correct way round!

Colors of LEDs

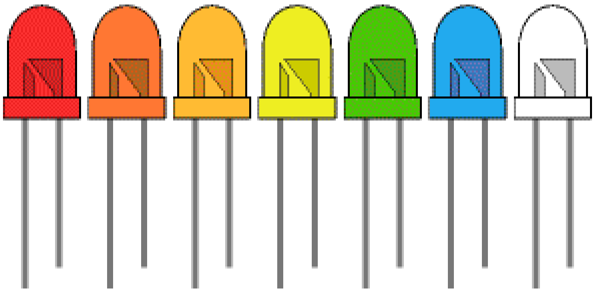

LEDs are available in red, orange, amber, yellow, green, blue and white. Blue and white LEDs are much more expensive than the other colors.

The color of an LED is determined by the semiconductor material, not by the coloring of the 'package' (the plastic body). LEDs of all colors are available in uncolored packages which may be diffused (milky) or clear (often described as 'water clear'). The colored packages are also available as diffused (the standard type) or transparent.

Tri-Color LEDs

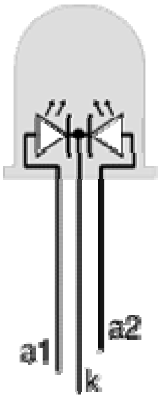

The most popular type of tri-color LED has a red and a green LED combined in one package with three leads. They are called tri-color because mixed red and green light appears to be yellow and this is produced when both the red and green LEDs are on.

The diagram shows the construction of a tri-color LED. Note the different lengths of the three leads. The centre lead (k) is the common cathode for both LEDs, the outer leads (a1 and a2) are the anodes to the LEDs allowing each one to be lit separately, or both together to give the third color.

Bi-Color LEDs

A bi-color LED has two LEDs wired in 'inverse parallel' (one forwards, one backwards) combined in one package with two leads. Only one of the LEDs can be lit at one time and they are less useful than the tri-color LEDs described above.

Sizes, Shapes and Viewing Angles of LEDs

LED Clip

LEDs are available in a wide variety of sizes and shapes. The 'standard' LED has a round cross-section of 5mm diameter and this is probably the best type for general use, but 3mm round LEDs are also popular.

Round cross-section LEDs are frequently used and they are very easy to install on boxes by drilling a hole of the LED diameter, adding a spot of glue will help to hold the LED if necessary. LED clips are also available to secure LEDs in holes. Other cross-section shapes include square, rectangular and triangular.

As well as a variety of colors, sizes and shapes, LEDs also vary in their viewing angle. This tells you how much the beam of light spreads out. Standard LEDs have a viewing angle of 60° but others have a narrow beam of 30° or less.

Calculating an LED Resistor Value

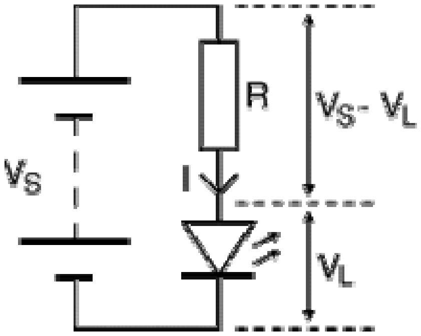

An LED must have a resistor connected in series to limit the current through the LED, otherwise it will burn out almost instantly.

The resistor value, R is given by:

$$\boxed{R = (V_S - V_L) / I}$$

V_S = supply voltage
V_L = LED voltage (usually 2V, but 4V for blue and white LEDs)
I = LED current (e.g. 20mA), this must be less than the maximum permitted

If the calculated value is not available choose the nearest standard resistor value which is **greater**, so that the current will be a little less than you chose. In fact you may wish to choose a greater resistor value to reduce the current (to increase battery life for example) but this will make the LED less bright.

For example

If the supply voltage V_S = 9V, and you have a red LED (V_L = 2V), requiring a current I = 20mA = 0.020A,
R = (9V - 2V) / 0.02A = 350Ω, so choose 390Ω (the nearest standard value which is greater).

Working out the LED resistor formula using Ohm's law

Ohm's law says that the resistance of the resistor, R = V/I, where:
V = voltage across the resistor (= $V_S - V_L$ in this case)
I = the current through the resistor

So $R = (V_S - V_L) / I$

Connecting LEDs in Series

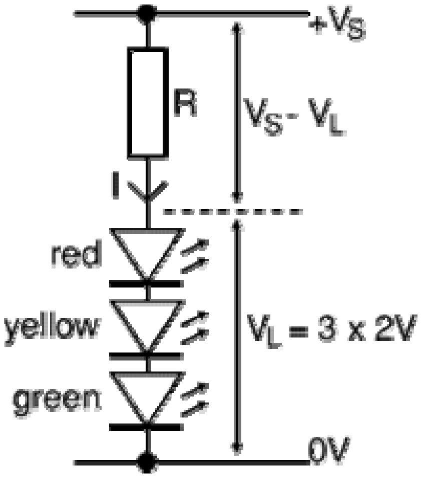

If you wish to have several LEDs on at the same time it may be possible to connect them in series. This prolongs battery life by lighting several LEDs with the same current as just one LED.

All the LEDs connected in series pass the **same current** so it is best if they are all the same type. The power supply must have sufficient voltage to provide about 2V for each LED (4V for blue and white) plus at least another 2V for the resistor. To work out a value for the resistor you must add up all the LED voltages and use this for V_L.

Example calculations:
A red, a yellow and a green LED in series need a supply voltage of at least 3 × 2V + 2V = 8V, so a **9V battery** would be ideal.
V_L = 2V + 2V + 2V = 6V (the three LED voltages added up).
If the supply voltage V_S is 9V and the current I must be 15mA = 0.015A,
Resistor R = (V_S - V_L) / I = (9 - 6) / 0.015 = 3 / 0.015 = 200Ω,
so choose R = 220Ω (the nearest standard value which is greater).

Avoid Connecting LEDs in Parallel

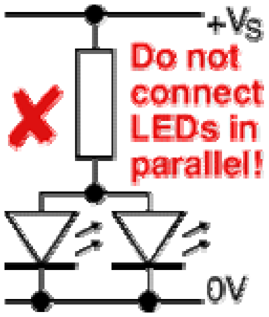

Connecting several LEDs in parallel with just one resistor shared between them is generally not a good idea.

If the LEDs require slightly different voltages only the lowest voltage LED will light and it may be destroyed by the larger current flowing through it. Although identical LEDs can be successfully connected in parallel with one resistor this rarely offers any useful benefit because resistors are very cheap and the current used is the same as connecting the LEDs individually. **If LEDs are in parallel each one should have its own resistor.**

Reading a table of technical data for LEDs

Suppliers' catalogues usually include tables of technical data for components such as LEDs. These tables contain a good deal of useful information in a compact form but they can be difficult to understand if you are not familiar with the abbreviations used.

The table below shows typical technical data for some 5mm diameter round LEDs with diffused packages (plastic bodies). Only three columns are important and these are shown in bold. Please see below for explanations of the quantities.

Type	Color	I_F max.	V_F typ.	V_F max.	V_R max.	Luminous intensity	Viewing angle	Wavelength
Standard	**Red**	**30mA**	**1.7V**	2.1V	5V	5mcd @ 10mA	60°	660nm
Standard	**Bright red**	**30mA**	**2.0V**	2.5V	5V	80mcd @ 10mA	60°	625nm
Standard	**Yellow**	**30mA**	**2.1V**	2.5V	5V	32mcd @ 10mA	60°	590nm
Standard	**Green**	**25mA**	**2.2V**	2.5V	5V	32mcd @ 10mA	60°	565nm
High intensity	**Blue**	**30mA**	**4.5V**	5.5V	5V	60mcd @ 20mA	50°	430nm
Super bright	**Red**	**30mA**	**1.85V**	2.5V	5V	500mcd @ 20mA	60°	660nm
Low current	**Red**	**30mA**	**1.7V**	2.0V	5V	5mcd @ 2mA	60°	625nm

I_F max.	Maximum forward current, forward just means with the LED connected correctly.
V_F typ.	Typical forward voltage, V_L in the LED resistor calculation. This is about 2V, except for blue and white LEDs for which it is about 4V.
V_F max.	Maximum forward voltage.
V_R max.	Maximum reverse voltage You can ignore this for LEDs connected the correct way round.
Luminous intensity	Brightness of the LED at the given current, mcd = millicandela.
Viewing angle	Standard LEDs have a viewing angle of 60°, others emit a narrower beam of about 30°.
Wavelength	The peak wavelength of the light emitted, this determines the color of the LED. nm = nanometer.

Flashing LEDs

Flashing LEDs look like ordinary LEDs but they contain an integrated circuit (IC) as well as the LED itself. The IC flashes the LED at a low frequency, typically 3Hz (3 flashes per second).

They are designed to be connected directly to a supply, usually 9 - 12V, and no series resistor is required. Their flash frequency is fixed so their use is limited and you may prefer to build your own circuit to flash an ordinary LED.

LED Displays

LED displays are packages of many LEDs arranged in a pattern, the most familiar pattern being the 7-segment displays for showing numbers (digits 0-9). The pictures below illustrate some of the popular designs:

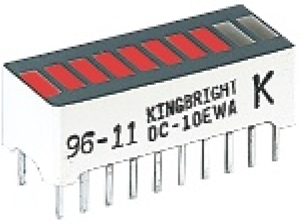
96-11
KINGBRIGHT
DC-10EWA
K

Bargraph

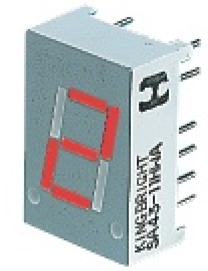
KINGBRIGHT
SA43-11HWA

7-segment

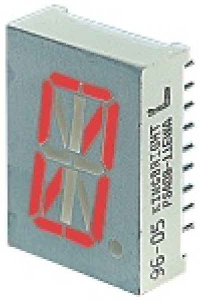
KINGBRIGHT
96-05

Starburst

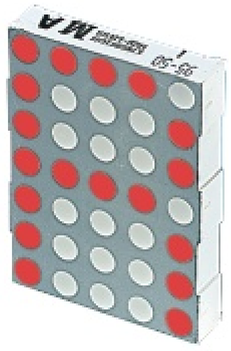

Dot matrix

Pin connections of LED displays

There are many types of LED display and a supplier's catalogue should be consulted for the pin connections. The diagram on the right shows an example. Like many 7-segment displays, this example is available in two versions: Common Anode (SA) with all the LED anodes connected together and Common Cathode (SC) with all the cathodes connected together. Letters a-g refers to the 7 segments, A/C is the common anode or cathode as appropriate (on 2 pins). Note that some pins are not present (NP) but their position is still numbered.

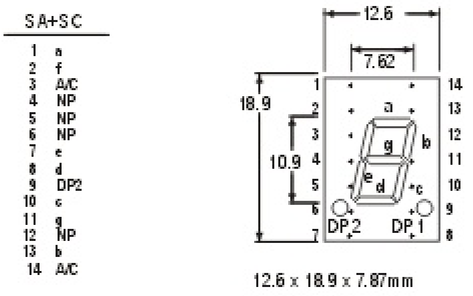
SA+SC
1 a
2 f
3 A/C
4 NP
5 NP
6 NP
7 e
8 d
9 DP2
10 c
11 g
12 NP
13 b
14 A/C
12.6
7.62
18.9
10.9
1
2
3
4
5
6
7
14
13
12
11
10
9
8
a
b
c
d
e
g
DP2
DP1
12.6 x 18.9 x 7.87mm

Pin connections diagram

Relays

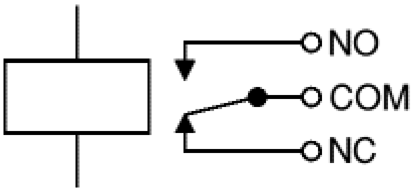

Circuit symbol for a Relay

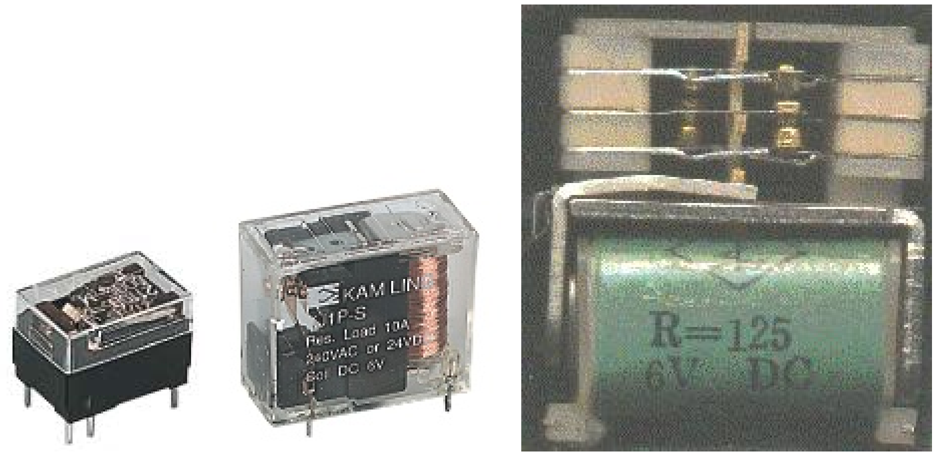

Different kind Picture of Relays

A relay is an **electrically operated switch**. Current flowing through the coil of the relay creates a magnetic field which attracts a lever and changes the switch contacts. The coil current can be on or off so relays have two switch positions and they are **double throw** (**changeover**) switches.

Relays allow one circuit to switch a second circuit which can be completely separate from the first. For example a low voltage battery circuit can use a relay to switch a 230V AC mains circuit. There is no electrical connection inside the relay between the two circuits, the link is magnetic and mechanical.

The coil of a relay passes a relatively large current, typically 30mA for a 12V relay, but it can be as much as 100mA for relays designed to operate from lower voltages. Most ICs (chips) cannot provide this current and a transistor is usually used to amplify the small IC current to the larger value required for the relay coil. The maximum output current for the popular 555 timer IC is 200mA so these devices can supply relay coils directly without amplification.

Relays are usually SPDT or DPDT but they can have many more sets of switch contacts, for example relays with 4 sets of changeover contacts are readily available. For further information about switch contacts and the terms used to describe them please see the page on switches.

Most relays are designed for PCB mounting but you can solder wires directly to the pins providing you take care to avoid melting the plastic case of the relay.

The supplier's catalogue should show you the relay's connections. The coil will be obvious and it may be connected either way round. Relay coils produce brief high voltage 'spikes' when they are switched off and this can destroy transistors and ICs in the circuit. To prevent damage, you must connect a protection diode across the relay coil.

The animated picture shows a working relay with its coil and switch contacts. You can see a lever on the left being attracted by magnetism when the coil is switched on. This lever moves the switch contacts. There is one set of contacts (SPDT) in the foreground and another behind them, making the relay DPDT.

The relay's switch connections are usually labeled COM, NC and NO:

- **COM** = Common, always connect to this, it is the moving part of the switch.
- **NC** = Normally Closed, COM is connected to this when the relay coil is **off**.
- **NO** = Normally Open, COM is connected to this when the relay coil is **on**.

- Connect to COM and NO if you want the switched circuit to be **on when the relay coil is on**.
- Connect to COM and NC if you want the switched circuit to be **on when the relay coil is off**.

Choosing a relay

You need to consider several features when choosing a relay:

1. **Physical size and pin arrangement**
   If you are choosing a relay for an existing PCB you will need to ensure that its dimensions and pin arrangement are suitable. You should find this information in the supplier's catalogue.
2. **Coil voltage**
   The relay's coil voltage rating and resistance must suit the circuit powering the relay coil. Many relays have a coil rated for a 12V supply but 5V and 24V relays are also readily available. Some relays operate perfectly well with a supply voltage which is a little lower than their rated value.
3. **Coil resistance**
   The circuit must be able to supply the current required by the relay coil. You can use Ohm's law to calculate the current:

$$\textbf{Relay coil current} = \frac{\textbf{supply voltage}}{\textbf{coil resistance}}$$

4. For example: A 12V supply relay with a coil resistance of 400Ω passes a current of 30mA. This is OK for a 555 timer IC (maximum output current 200mA), but it is too much for most ICs and they will require a transistor to amplify the current.
5. **Switch ratings (voltage and current)**
   The relay's switch contacts must be suitable for the circuit they are to control. You will need to check the voltage and current ratings. Note that the voltage rating is usually higher for AC, for example: "5A at 24V DC or 125V AC".
6. **Switch contact arrangement (SPDT, DPDT etc)**
   Most relays are SPDT or DPDT which are often described as "single pole changeover" (SPCO) or "double pole changeover" (DPCO). For further information please see the page on switches.

Protection Diodes for Relays

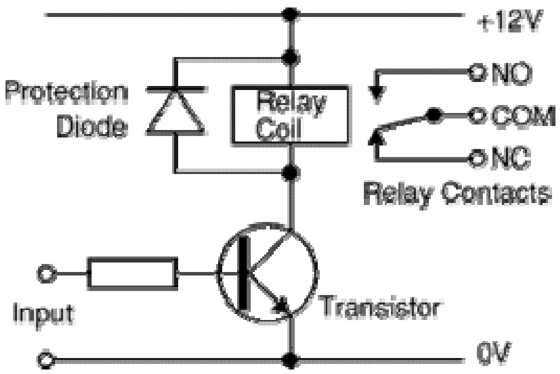

Transistors and ICs (chips) must be protected from the brief high voltage 'spike' produced when the relay coil is switched off. The diagram shows how a signal diode (e.g. 1N4148) is connected across the relay coil to provide this protection. Note that the diode is connected 'backwards' so that it will normally **not** conduct. Conduction only occurs when the relay coil is switched off, at this moment current tries to continue flowing through the coil and it is harmlessly diverted through the diode. Without the diode no current could flow and the coil would produce a damaging high voltage 'spike' in its attempt to keep the current flowing.

Reed Relays

Reed relays consist of a coil surrounding a reed switch. Reed switches are normally operated with a magnet, but in a reed relay current flows through the coil to create a magnetic field and close the reed switch.

Reed relays generally have higher coil resistances than standard relays (1000Ω for example) and a wide range of supply voltages (9-20V for example). They are capable of switching much more rapidly than standard relays, up to several hundred times per second; but they can only switch low currents (500mA maximum for example).

EDR201A0500

Reed Relay

The reed relay shown in the photograph will plug into a standard 14-pin DIL socket ('chip holder').

Relays and Transistors Compared

Like relays, transistors can be used as an electrically operated switch. For switching small DC currents (< 1A) at low voltage they are usually a better choice than a relay. However transistors cannot switch AC or high voltages (such as mains electricity) and they are not usually a good choice for switching large currents (> 5A). In these cases a relay will be needed, but note that a low power transistor may still be needed to switch the current for the relay's coil! The main advantages and disadvantages of relays are listed below:

Advantages of relays:

- Relays can switch **AC and DC**, transistors can only switch DC.
- Relays can switch **high voltages**, transistors cannot.
- Relays are a better choice for switching **large currents** (> 5A).
- Relays can switch **many contacts** at once.

Disadvantages of relays:

- Relays are **bulkier** than transistors for switching small currents.
- Relays **cannot switch rapidly** (except reed relays), transistors can switch many times per second.
- Relays **use more power** due to the current flowing through their coil.
- Relays **require more current than many chips can provide**, so a low power transistor may be needed to switch the current for the relay's coil.

Resistors

Example: 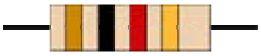 Circuit symbol: 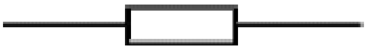

Function

Resistors restrict the flow of electric current, for example a resistor is placed in series with a light-emitting diode (LED) to limit the current passing through the LED.

Connecting and Soldering

Resistors may be connected either way round. They are not damaged by heat when soldering.

Resistor Values - the Resistor Color Code

Resistance is measured in ohms, the symbol for ohm is an omega Ω.
1 Ωis quite small so resistor values are often given in kΩ and MΩ.
1 kΩ = 1000 Ω   1 MΩ = 1000000 Ω.

Resistor values are normally shown using colored bands.
Each color represents a number as shown in the table.

Most resistors have 4 bands:

- The **first band** gives the **first digit**.
- The **second band** gives the **second digit**.
- The **third band** indicates the **number of zeros**.
- *The fourth band is used to shows the tolerance (precision) of the resistor, this may be ignored for almost all circuits but further details are given below.*

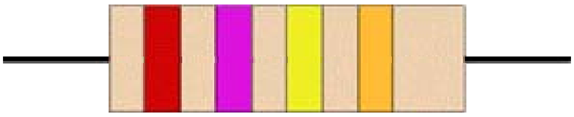

The Resistor Color Code

Color	Number
Black	0
Brown	1
Red	2
Orange	3
Yellow	4
Green	5
Blue	6
Violet	7
Grey	8
White	9

This resistor has red (2), violet (7), yellow (4 zeros) and gold bands.
So its value is 270000 Ω= 270 kΩ.
On circuit diagrams the Ω is usually omitted and the value is written 270K.

Small Value Resistors (less than 10 ohm)

The standard color code cannot show values of less than 10Ω. To show these small values two special colors are used for the **third band**: gold which means × 0.1 and silver which means × 0.01. The first and second bands represent the digits as normal.

For example:
red, **violet**, gold bands represent 27 × 0.1 = 2.7Ω
green, **blue**, silver bands represent 56 × 0.01 = 0.56Ω

Tolerance of Resistors (fourth band of color code)

The tolerance of a resistor is shown by the **fourth band** of the color code. Tolerance is the **precision** of the resistor and it is given as a percentage. For example a 390Ω resistor with a tolerance of ±10% will have a value within 10% of 390Ω, between 390 - 39 = 351Ω and 390 + 39 = 429Ω (39 is 10% of 390).

A special color code is used for the **fourth band** tolerance:
silver ±10%, **gold** ±5%, **red** ±2%, **brown** ±1%.
If no fourth band is shown the tolerance is ±20%.

Tolerance may be ignored for almost all circuits because precise resistor values are rarely required.

Resistor Shorthand

Resistor values are often written on circuit diagrams using a code system which avoids using a decimal point because it is easy to miss the small dot. Instead the letters R, K and M are used in place of the decimal point. To read the code: replace the letter with a decimal point, then multiply the value by 1000 if the letter was K, or 1000000 if the letter was M. The letter R means multiply by 1.

For example:

560R means 560 Ω
2K7 means 2.7 kΩ = 2700 Ω
39K means 39 kΩ
1M0 means 1.0 MΩ = 1000 kΩ

Real Resistor Values (the E6 and E12 series)

You may have noticed that resistors are not available with every possible value, for example 22kΩ and 47kΩ are readily available, but 25kΩ and 50kΩ are not!

Why is this? Imagine that you decided to make resistors every 10Ω giving 10, 20, 30, 40, 50 and so on. That seems fine, but what happens when you reach 1000? It would be pointless to make 1000, 1010, 1020, 1030 and so on because for these values 10 is a very small difference, too small to be noticeable in most circuits. In fact it would be difficult to make resistors sufficiently accurate.

To produce a sensible range of resistor values you need to increase the size of the 'step' as the value increases. The standard resistor values are based on this idea and they form a series which follows the same pattern for every multiple of ten.

The E6 series (6 values for each multiple of ten, for resistors with 20% tolerance)
10, 15, 22, 33, 47, 68, ... then it continues 100, 150, 220, 330, 470, 680, 1000 etc.
Notice how the step size increases as the value increases. For this series the step (to the next value) is roughly half the value.

The E12 series (12 values for each multiple of ten, for resistors with 10% tolerance)
10, 12, 15, 18, 22, 27, 33, 39, 47, 56, 68, 82, ... then it continues 100, 120, 150 etc.
Notice how this is the E6 series with an extra value in the gaps.

The E12 series is the one most frequently used for resistors. It allows you to choose a value within 10% of the precise value you need. This is sufficiently accurate for almost all projects and it is sensible because most resistors are only accurate to ±10% (called their 'tolerance'). For example a resistor marked 390Ω could vary by ±10% × 390Ω = ±39Ω, so it could be any value between 351Ω and 429Ω.

Resistors in Series and Parallel

For information on resistors connected in series and parallel, see the Resistance page,

---

Power Ratings of Resistors

Electrical energy is converted to heat when current flows through a resistor. Usually the effect is negligible, but if the resistance is low (or the voltage across the resistor high) a large current may pass making the resistor become noticeably warm. The resistor must be able to withstand the heating effect and resistors have power ratings to show this.

Power ratings of resistors are rarely quoted in parts lists because for most circuits the standard power ratings of 0.25W or 0.5W are suitable. For the rare cases where a higher power is required it should be clearly specified in the parts list, these will be circuits using **low value resistors** (less than about 300Ω) or **high voltages** (more than 15V).

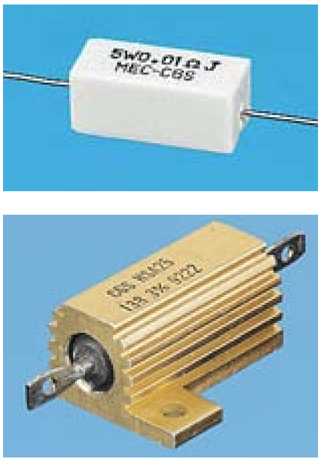

High power Resistors (5W top, 25W bottom)

The power, P, developed in a resistor is given by:

$P = I^2 \times R$ where: P = power developed in the resistor in watts (W)
or I = current through the resistor in amps (A)
$P = V^2 / R$ R = resistance of the resistor in ohms (Ω)
V = voltage across the resistor in volts (V)

Examples:

- A 470Ω resistor with 10V across it, needs a power rating $P = V^2/R = 10^2/470 = 0.21$W.
  In this case a standard 0.25W resistor would be suitable.
- A 27Ω resistor with 10V across it, needs a power rating $P = V^2/R = 10^2/27 = 3.7$W.
  A high power resistor with a rating of 5W would be suitable.

Switches

Selecting a Switch

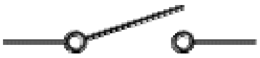

Circuit Symbol for a simple ON-OFF Switch

There are three important features to consider when selecting a switch:

- **Contacts** (e.g. single pole, double throw)
- **Ratings** (maximum voltage and current)
- **Method of Operation** (toggle, slide, key etc.)

Switch Contacts

Several terms are used to describe switch contacts:

- **Pole** - number of switch contact sets.
- **Throw** - number of conducting positions, single or double.
- **Way** - number of conducting positions, three or more.
- **Momentary** - switch returns to its normal position when released.
- **Open** - off position, contacts not conducting.
- **Closed** - on position, contacts conducting, there may be several on positions.

For example: the simplest on-off switch has one set of contacts (single pole) and one switching position which conducts (single throw). The switch mechanism has two positions: open (off) and closed (on), but it is called 'single throw' because only one position conducts.

Switch Contact Ratings

Switch contacts are rated with a maximum voltage and current, and there may be different ratings for AC and DC. The AC values are higher because the current falls to zero many times each second and an arc is less likely to form across the switch contacts.

For low voltage electronics projects the voltage rating will not matter, but you may need to check the current rating. The maximum current is less for inductive loads (coils and motors) because they cause more sparking at the contacts when switched off.

Standard Switches

Type of Switch	Circuit Symbol	Example
ON-OFF **Single Pole, Single Throw = SPST** A simple on-off switch. This type can be used to switch the power supply to a circuit. When used with mains electricity this type of switch *must* be in the live wire, but it is better to use a DPST switch to isolate both live and neutral.	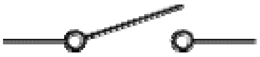	 SPST toggle switch

(ON)-OFF **Push-to-make = SPST Momentary** A push-to-make switch returns to its normally open (off) position when you release the button, this is shown by the brackets around ON. This is the standard doorbell switch.	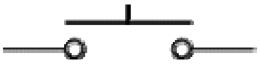	 Push-to-make switch
ON-(OFF) **Push-to-break = SPST Momentary** A push-to-break switch returns to its normally closed (on) position when you release the button.	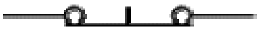	 Push-to-break switch
ON-ON **Single Pole, Double Throw = SPDT** This switch can be on in both positions, switching on a separate device in each case. It is often called a **changeover switch.** For example, a SPDT switch can be used to switch on a red lamp in one position and a green lamp in the other position. A SPDT toggle switch may be used as a simple on-off switch by connecting to COM and one of the A or B terminals shown in the diagram. A and B are interchangeable so switches are usually not labeled. **ON-OFF-ON** **SPDT Centre Off** A special version of the standard SPDT switch. It has a third switching position in the centre which is off. Momentary (ON)-OFF-(ON) versions are also available where the switch returns to the central off position when released.	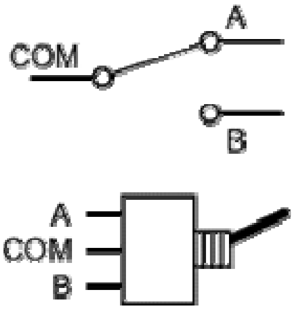	 SPDT toggle switch SPDT slide switch (PCB mounting) SPDT rocker switch
Dual ON-OFF **Double Pole, Single Throw = DPST** A pair of on-off switches which operate together (shown by the dotted line in the circuit symbol). A DPST switch is often used to switch mains electricity because it can isolate both the live and neutral connections.	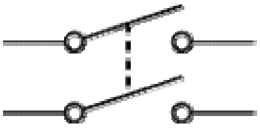	 DPST rocker switch

Dual ON-ON **Double Pole, Double Throw = DPDT** A pair of on-on switches which operate together (shown by the dotted line in the circuit symbol). A DPDT switch can be wired up as a **reversing switch** for a motor as shown in the diagram. **ON-OFF-ON** **DPDT Centre Off** A special version of the standard SPDT switch. It has a third switching position in the centre which is off. This can be very useful for motor control because you have forward, off and reverse positions. Momentary (ON)-OFF-(ON) versions are also available where the switch returns to the central off position when released.	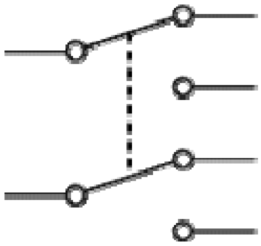	 DPDT slide switch 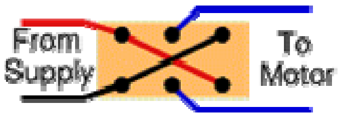 Wiring for Reversing Switch

Special Switches

Type of Switch	Example
Push-Push Switch (e.g. SPST = ON-OFF) This looks like a momentary action push switch but it is a standard on-off switch: push once to switch on, push again to switch off. This is called a **latching action**.	
Microswitch (usually SPDT = ON-ON) Microswitches are designed to switch fully open or closed in response to small movements. They are available with levers and rollers attached. Also called Limit Switch	
Keyswitch A key operated switch. The example shown is SPST.	
Tilt Switch (SPST) Tilt switches contain a conductive liquid and when tilted this bridges the contacts inside, closing the switch. They can be used as a sensor to detect the position of an object. Some tilt switches contain mercury which is poisonous.	

Reed Switch (usually SPST) The contacts of a reed switch are closed by bringing a small magnet near the switch. They are used in security circuits, for example to check that doors are closed. Standard reed switches are SPST (simple on-off) but SPDT (changeover) versions are also available. **Warning:** reed switches have a glass body which is easily broken! Handling precaution needed.	
DIP Switch (DIP = Dual In-line Parallel) This is a set of miniature SPST on-off switches, the example shown has 8 switches. The package is the same size as a standard DIL (Dual In-Line) integrated circuit. This type of switch is used to set up circuits, e.g. setting the code of a remote control.	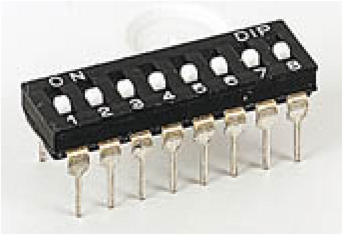
Multi-pole Switch The picture shows a 6-pole double throw switch, also known as a 6-pole changeover switch. It can be set to have momentary or latching action. Latching action means it behaves as a push-push switch, push once for the first position, push again for the second position etc.	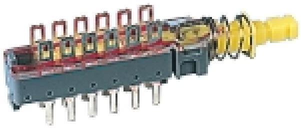
Multi-way Switch Multi-way switches have 3 or more conducting positions. They may have several poles (contact sets). A popular type has a rotary action and it is available with a range of contact arrangements from 1-pole 12-way to 4-pole 3 way. The number of ways (switch positions) may be reduced by adjusting a stop under the fixing nut. For example if you need a 2-pole 5-way switch you can buy the 2-pole 6-way version and adjust the stop. Contrast this multi-way switch (many switch positions) with the multi-pole switch (many contact sets) described above.	 Multi-way rotary switch 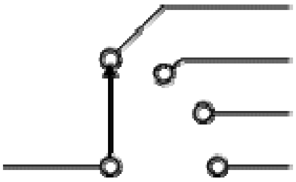 1-pole 4-way switch symbol

Transistors

This page covers practical matters such as precautions when soldering and identifying leads. The operation and use of transistors is covered by the Transistor Circuits page.

Function

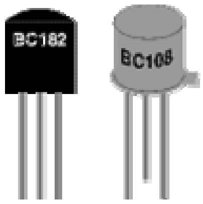

Transistors **amplify current**, for example they can be used to amplify the small output current from a logic chip so that it can operate a lamp, relay or other high current device. In many circuits a resistor is used to convert the changing current to a changing voltage, so the transistor is being used to **amplify voltage**.

A transistor may be used as a **switch** (either fully on with maximum current, or fully off with no current) and as an **amplifier** (always partly on).

The amount of current amplification is called the **current gain**, symbol h_{FE}.

Types of Transistor

There are two types of standard transistors, **NPN** and **PNP**, with different circuit symbols. The letters refer to the layers of semiconductor material used to make the transistor. Most transistors used today are NPN because this is the easiest type to make from silicon. If you are new to electronics it is best to start by learning how to use NPN transistors.

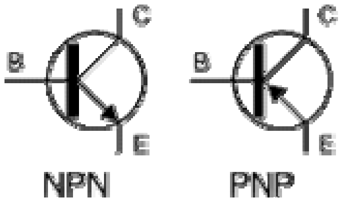

Transistor Symbols

The leads are labeled **base** (B), **collector** (C) and **emitter** (E). These terms refer to the internal operation of a transistor but they are not much help in understanding how a transistor is used, so just treat them as labels!

A Darlington pair is two transistors connected together to give a very high current gain.

In addition to standard (bipolar junction) transistors, there are **field-effect transistors** which are usually referred to as **FET**s. They have different circuit symbols and properties and they are not (yet) covered by this page.

Connecting

Transistors have three leads which must be connected the correct way round. Please take care with this because a wrongly connected transistor may be damaged instantly when you switch on.

If you are lucky the orientation of the transistor will be clear from the PCB or stripboard layout diagram, otherwise you will need to refer to a supplier's catalogue to identify the leads.

The drawings on the right show the leads for some of the most common case styles.

Please note that transistor lead diagrams show the view from **bottom** with the leads towards you. This is the opposite of IC (chip) pin diagrams which show the view from above.

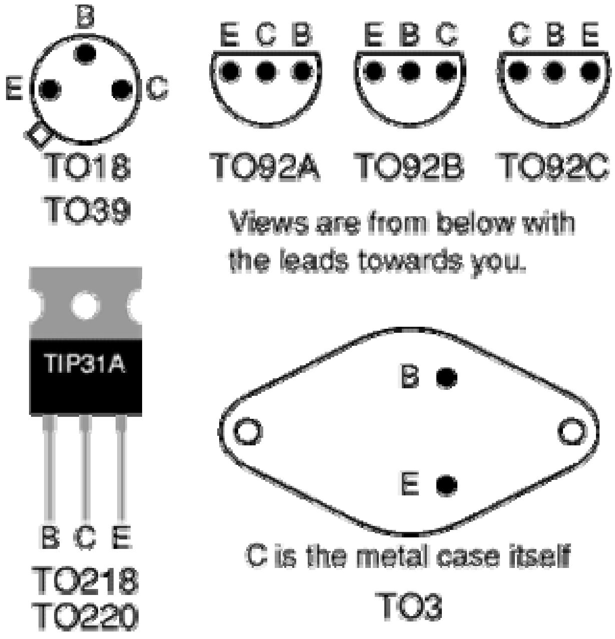
B
E
C
TO18
TO39
E C B
E B C
C B E
TO92A
TO92B
TO92C
Views are from below with the leads towards you.
TIP31A
B C E
TO218
TO220
B
E
C is the metal case itself
TO3

Transistor leads for some common case styles.

Soldering

Transistors can be damaged by heat when soldering so if you are not an expert it is wise to use a heat sink clipped to the lead between the joint and the transistor body. A standard crocodile clip can be used as a heat sink.

Do not confuse this temporary heat sink with the permanent heat sink (described below) which may be required for a power transistor to prevent it overheating during operation.

Crocodile Clip

Testing a transistor

Transistors can be damaged by heat when soldering or by misuse in a circuit. If you suspect that a transistor may be damaged there are two easy ways to test it:

1. Testing with a multimeter

Use a multimeter or a simple tester (battery, resistor and LED) to check each pair of leads for conduction. Set a digital multimeter to diode test and an analogue multimeter to a low resistance range.

Test each pair of leads both ways (six tests in total):

- The **base-emitter (BE)** junction should behave like a diode and conduct one way only.
- The **base-collector (BC)** junction should behave like a diode and conduct one way only.
- The **collector-emitter (CE)** should not conduct either way.

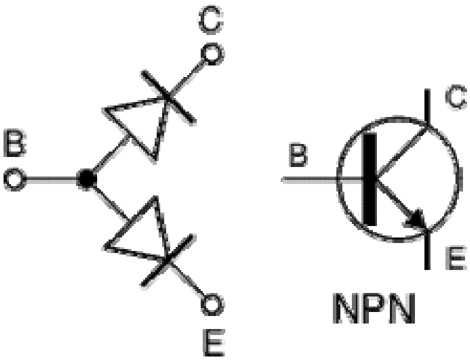

Testing an NPN transistor

The diagram shows how the junctions behave in an NPN transistor. The diodes are reversed in a PNP transistor but the same test procedure can be used.

2. Testing in a simple switching circuit

Connect the transistor into the circuit shown on the right which uses the transistor as a switch. The supply voltage is not critical, anything between 5 and 12V is suitable. This circuit can be quickly built on breadboard for example. Take care to include the 10kΩ resistor in the base connection or you will destroy the transistor as you test it!

If the transistor is OK the LED should light when the switch is pressed and not light when the switch is released.

To test a PNP transistor use the same circuit but reverse the LED and the supply voltage.

Some multimeters have a 'transistor test' function which provides a known base current and measures the collector current so as to display the transistor's DC current gain h_{FE}.

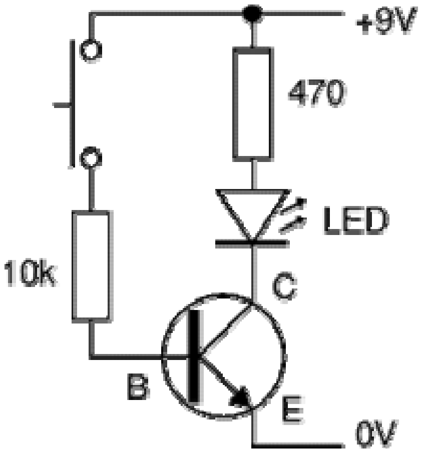

A simple switching circuit to test an NPN transistor

Transistor Codes

There are three main series of transistor codes used in the UK:

- **Codes beginning with B (or A), for example BC108, BC478**
  The first letter B is for silicon, A is for germanium (rarely used now). The second letter indicates the type; for example C means low power audio frequency; D means high power audio frequency; F means low power high frequency. The rest of the code identifies the particular transistor. There is no obvious logic to the numbering system. Sometimes a letter is added to the end (e.g. BC108C) to identify a special version of the main type, for example a higher current gain or a different case style. If a project specifies a higher gain version (BC108C) it must be used, but if the general code is given (BC108) any transistor with that code is suitable.

- **Codes beginning with TIP, for example TIP31A**
  TIP refers to the manufacturer: Texas Instruments Power transistor. The letter at the end identifies versions with different voltage ratings.
- **Codes beginning with 2N, for example 2N3053**
  The initial '2N' identifies the part as a transistor and the rest of the code identifies the particular transistor. There is no obvious logic to the numbering system.

Choosing a Transistor

Most projects will specify a particular transistor, but if necessary you can usually substitute an equivalent transistor from the wide range available. The most important properties to look for are the maximum collector current I_C and the current gain h_{FE}. To make selection easier most suppliers group their transistors in categories determined either by their **typical use** or **maximum power** rating.

To make a final choice you will need to consult the tables of technical data which are normally provided in catalogues. They contain a great deal of useful information but they can be difficult to understand if you are not familiar with the abbreviations used. The table below shows the most important technical data for some popular transistors, tables in catalogues and reference books will usually show additional information but this is unlikely to be useful unless you are experienced. The quantities shown in the table are explained below.

NPN Transistors

Code	Structure	Case style	I_C max.	V_{CE} max.	h_{FE} min.	P_{tot} max.	Category (typical use)	Possible substitutes
BC107	NPN	TO18	100mA	45V	110	300mW	Audio, low power	BC182 BC547
BC108	NPN	TO18	100mA	20V	110	300mW	General purpose, low power	BC108C BC183 BC548
BC108C	NPN	TO18	100mA	20V	420	600mW	General purpose, low power	
BC109	NPN	TO18	200mA	20V	200	300mW	Audio (low noise), low power	BC184 BC549
BC182	NPN	TO92C	100mA	50V	100	350mW	General purpose, low power	BC107 BC182L
BC182L	NPN	TO92A	100mA	50V	100	350mW	General purpose, low power	BC107 BC182
BC547B	NPN	TO92C	100mA	45V	200	500mW	Audio, low power	BC107B
BC548B	NPN	TO92C	100mA	30V	220	500mW	General purpose, low power	BC108B
BC549B	NPN	TO92C	100mA	30V	240	625mW	Audio (low noise), low power	BC109
2N3053	NPN	TO39	700mA	40V	50	500mW	General purpose, low power	BFY51
BFY51	NPN	TO39	1A	30V	40	800mW	General purpose, medium power	BC639
BC639	NPN	TO92A	1A	80V	40	800mW	General purpose, medium power	BFY51

TIP29A	NPN	TO220	1A	60V	40	30W	General purpose, high power	
TIP31A	NPN	TO220	3A	60V	10	40W	General purpose, high power	TIP31C TIP41A
TIP31C	NPN	TO220	3A	100V	10	40W	General purpose, high power	TIP31A TIP41A
TIP41A	NPN	TO220	6A	60V	15	65W	General purpose, high power	
2N3055	NPN	TO3	15A	60V	20	117W	General purpose, high power	

Please note: the data in this table was compiled from several sources which are not entirely consistent! Most of the discrepancies are minor, but please consult information from your supplier if you require precise data.

PNP Transistors

Code	Structure	Case style	I_C max.	V_{CE} max.	h_{FE} min.	P_{tot} max.	Category (typical use)	Possible substitutes
BC177	PNP	TO18	100mA	45V	125	300mW	Audio, low power	BC477
BC178	PNP	TO18	200mA	25V	120	600mW	General purpose, low power	BC478
BC179	PNP	TO18	200mA	20V	180	600mW	Audio (low noise), low power	
BC477	PNP	TO18	150mA	80V	125	360mW	Audio, low power	BC177
BC478	PNP	TO18	150mA	40V	125	360mW	General purpose, low power	BC178
TIP32A	PNP	TO220	3A	60V	25	40W	General purpose, high power	TIP32C
TIP32C	PNP	TO220	3A	100V	10	40W	General purpose, high power	TIP32A

Please note: the data in this table was compiled from several sources which are not entirely consistent! Most of the discrepancies are minor, but please consult information from your supplier if you require precise data.

Structure	This shows the type of transistor, NPN or PNP. The polarities of the two types are different, so if you are looking for a substitute it must be the same type.
Case style	There is a diagram showing the leads for some of the most common case styles in the Connecting section above. This information is also available in suppliers' catalogues.
I_C max.	Maximum collector current.
V_{CE} max.	Maximum voltage across the collector-emitter junction. You can ignore this rating in low voltage circuits.
h_{FE}	This is the **current gain** (strictly the DC current gain). The guaranteed minimum value is given because the actual value varies from transistor to transistor - even for those of the same type! Note that current gain is just a number so it has no units. The gain is often quoted at a particular collector current I_C which is usually in the middle of the transistor's range, for example '100@20mA' means the gain is at least 100 at 20mA. Sometimes minimum and

	maximum values are given. Since the gain is roughly constant for various currents but it varies from transistor to transistor this detail is only really of interest to experts. **Why h_{FE}?** It is one of a whole series of parameters for transistors, each with their own symbol. There are too many to explain here.
P_{tot} max.	Maximum total power which can be developed in the transistor, note that a heat sink will be required to achieve the maximum rating. This rating is important for transistors operating as amplifiers, the power is roughly I_C × V_{CE}. For transistors operating as switches the maximum collector current (I_C max.) is more important.
Category	This shows the typical use for the transistor, it is a good starting point when looking for a substitute. Catalogues may have separate tables for different categories.
Possible Substitutes	These are transistors with similar electrical properties which will be suitable substitutes in most circuits. However, they may have a different case style so you will need to take care when placing them on the circuit board.

Darlington Pair

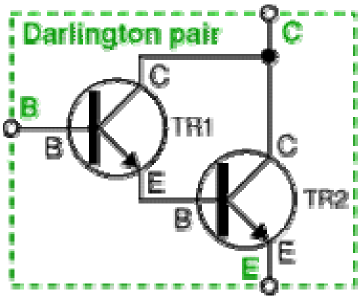

This is two transistors connected together so that the amplified current from the first is amplified further by the second transistor. This gives the Darlington pair a very high current gain such as 10000. Darlington pairs are sold as complete packages containing the two transistors. They have three leads (**B**, **C** and **E**) which are equivalent to the leads of a standard individual transistor.

You can make up your own Darlington pair from two transistors.

For example:

- For TR1 use BC548B with h_{FE1} = 220.
- For TR2 use BC639 with h_{FE2} = 40.

The overall gain of this pair is $h_{FE1} \times h_{FE2} = 220 \times 40 = 8800$.
The pair's maximum collector current I_C(max) is the same as TR2.

Heatsinks for Transistors

Heat sinks are needed for transistors passing large currents.

Why is a Heatsink Needed?

Waste heat is produced in transistors due to the current flowing through them. If you find that a transistor is becoming too hot to touch it certainly needs a heat sink! The heat sink helps to dissipate (remove) the heat by transferring it to the surrounding air.

Heatsink

The rate of producing waste heat is called the thermal power, P. Usually the base current I_B is too small to contribute much heat, so the thermal power is determined by the collector current I_C and the voltage V_{CE} across the transistor:

$P = I_C \times V_{CE}$ (see diagram below)

The heat is not a problem if I_C is small or if the transistor is used as a switch because when 'full on' V_{CE} is almost zero. However, power transistors used in circuits such as an audio amplifier or a motor speed controller will be partly on most of the time and V_{CE} may be about half the supply voltage. These power transistors will almost certainly need a heat sink to prevent them overheating.

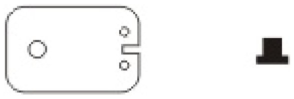

Insulation kit

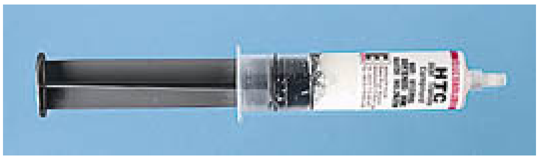

Heat-conducting paste

Power transistors usually have bolt holes for attaching heat sinks, but clip-on heat sinks are also available. Make sure you use the right type for your transistor. Many transistors have metal cases which are connected to one of their leads so it may be necessary to insulate the heat sink from the transistor. Insulating kits are available with a mica sheet and a plastic sleeve for the bolt. Heat-conducting paste can be used to improve heat flow from the transistor to the heat sink, this is especially important if an insulation kit is used.

Heatsink Ratings

Heat sinks are rated by their thermal resistance (Rth) in °C/W. For example 2°C/W means the heat sink (and therefore the component attached to it) will be 2°C hotter than the surrounding air for every 1W of heat it is dissipating. Note that a **lower** thermal resistance means a **better** heatsink.

This is how you work out the required heat sink rating:

1. Work out thermal power to be dissipated, $P = I_C \times V_{CE}$
   If in doubt use the largest likely value for I_C and assume that V_{CE} is half the supply voltage.
   For example if a power transistor is passing 1A and connected to a 12V supply, the power P is about 1 × ½ × 12 = 6W.
2. Find the maximum operating temperature (Tmax) for the transistor if you can, otherwise assume Tmax = 100°C.

3. Estimate the maximum ambient (surrounding air) temperature (Tair). If the heat sink is going to be outside the case Tair = 25°C is reasonable, but inside it will be higher (perhaps 40°C) allowing for everything to warm up in operation.
4. Work out the maximum thermal resistance (Rth) for the heat sink using: Rth = (Tmax - Tair) / P

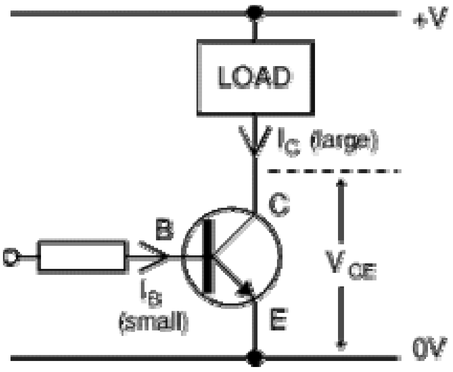

With the example values given above: Rth = (100-25)/6 = 12.5°C/W.

5. Choose a heat sink with a thermal resistance which is **less** than the value calculated above (remember lower value means better heat sinking!) for example 5°C/W would be a sensible choice to allow a safety margin. A 5°C/W heat sink dissipating 6W will have a temperature difference of 5 × 6 = 30°C so the transistor temperature will rise to 25 + 30 = 55°C (safely less than the 100°C maximum).
6. All the above assumes the transistor is at the same temperature as the heat sink. This is a reasonable assumption if they are firmly bolted or clipped together. However, you may have to put a mica sheet or similar between them to provide electrical insulation, then the transistor will be hotter than the heat sink and the calculation becomes more difficult. For typical mica sheets you should subtract 2°C/W from the thermal resistance (Rth) value calculated in step 4 above.

If this all seems too complex you can try attaching a moderately large heat sink and hope for the best. Cautiously monitor the transistor temperature with your finger, if it becomes painfully hot switch off immediately and use a larger heat sink!

Why Thermal Resistance?

The term '**thermal resistance**' is used because it is analogous to electrical resistance:

- The temperature difference across the heat sink (between the transistor and air) is like voltage (potential difference) across a resistor.
- The thermal power (rate of heat) flowing through the heat sink from transistor to air is like current flowing through a resistor.
- So R = V/I becomes Rth = (Tmax - Tair)/P
- Just as you need a voltage difference to make current flow, you need a temperature difference to make heat flow.

Variable Resistors

Construction

Variable resistors consist of a resistance **track** with connections at both ends and a **wiper** which moves along the track as you turn the spindle. The track may be made from carbon, cermet (ceramic and metal mixture) or a coil of wire (for low resistances). The track is usually rotary but straight track versions, usually called sliders, are also available.

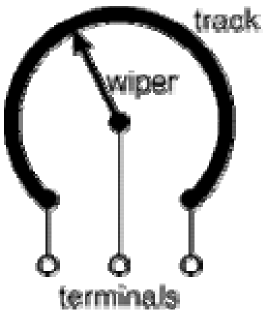

Standard Variable Resistor

Variable resistors may be used as a rheostat with **two** connections (the wiper and just one end of the track) or as a potentiometer with all **three** connections in use. Miniature versions called presets are made for setting up circuits which will not require further adjustment.

Variable resistors are often called **potentiometers** in books and catalogues. They are specified by their maximum resistance, linear or logarithmic track, and their physical size. The standard spindle diameter is 6mm.

The resistance and type of track are marked on the body:
4K7 LIN means 4.7 kΩ linear track.
1M LOG means 1 MΩ logarithmic track.

Some variable resistors are designed to be mounted directly on the circuit board, but most are for mounting through a hole drilled in the case containing the circuit with stranded wire connecting their terminals to the circuit board.

Linear (LIN) and Logarithmic (LOG) tracks

Linear (LIN) track means that the resistance changes at a constant rate as you move the wiper. This is the standard arrangement and you should assume this type is required if a project does not specify the type of track. Presets always have linear tracks.

Logarithmic (LOG) track means that the resistance changes slowly at one end of the track and rapidly at the other end, so halfway along the track is **not** half the total resistance! This arrangement is used for volume (loudness) controls because the human ear has a logarithmic response to loudness so fine control (slow change) is required at low volumes and coarser control (rapid change) at high volumes. It is important to connect the ends of the track the correct way round, if you find that turning the spindle increases the volume rapidly followed by little further change you should swap the connections to the ends of the track.

Rheostat

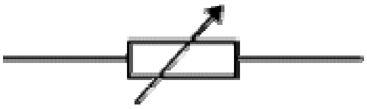

Rheostat Symbol

This is the simplest way of using a variable resistor. **Two terminals** are used: one connected to an end of the track, the other to the moveable wiper. Turning the spindle changes the resistance between the two terminals from zero up to the maximum resistance.

Rheostats are often used to **vary current**, for example to control the brightness of a lamp or the rate at which a capacitor charges.

If the rheostat is mounted on a printed circuit board you may find that all three terminals are connected! However, one of them will be linked to the wiper terminal. This improves the mechanical strength of the mounting but it serves no function electrically.

Potentiometer

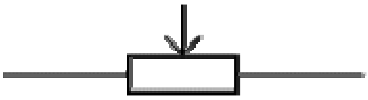

Potentiometer Symbol

Variable resistors used as potentiometers have all **three terminals** connected.

This arrangement is normally used to **vary voltage**, for example to set the switching point of a circuit with a sensor, or control the volume (loudness) in an amplifier circuit. If the terminals at the ends of the track are connected across the power supply then the wiper terminal will provide a voltage which can be varied from zero up to the maximum of the supply.

Presets

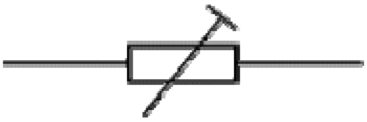

Preset Symbol

These are miniature versions of the standard variable resistor. They are designed to be mounted directly onto the circuit board and adjusted only when the circuit is built. For example to set the frequency of an alarm tone or the sensitivity of a light-sensitive circuit. A small screwdriver or similar tool is required to adjust presets.

Presets are much cheaper than standard variable resistors so they are sometimes used in projects where a standard variable resistor would normally be used.

Multiturn presets are used where very precise adjustments must be made. The screw must be turned many times (10+) to move the slider from one end of the track to the other, giving very fine control.

Preset
(open style)

Presets
(closed style)

Multiturn preset

Other Components

Light Dependent Resistor (LDR)

An LDR is an **input transducer** (sensor) which converts brightness (light) to resistance. It is made from cadmium sulphide (CdS) and the resistance decreases as the brightness of light falling on the LDR increases.

A multimeter can be used to find the resistance in darkness and bright light, these are the typical results for a standard LDR:

- **Darkness**: maximum resistance, about 1MΩ.
- **Very bright light**: minimum resistance, about 100Ω.

For many years the standard LDR has been the ORP12, now the NORPS12, which is about 13mm diameter. Miniature LDRs are also available and their diameter is about 5mm.

An LDR may be connected either way round and no special precautions are required when soldering.

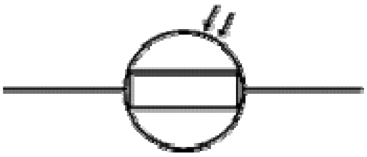

circuit symbol

Thermistor

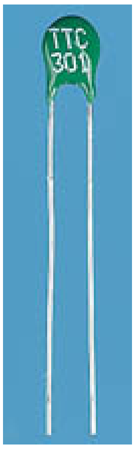

A thermistor is an **input transducer** (sensor) which converts temperature (heat) to resistance. Almost all thermistors have a negative temperature coefficient (NTC) which means their resistance decreases as their temperature increases. It is possible to make thermistors with a positive temperature coefficient (resistance increases as temperature increases) but these are rarely used. Always assume NTC if no information is given.

A multimeter can be used to find the resistance at various temperatures, these are some typical readings for example:

- **Icy water 0°C**: high resistance, about 12kΩ.
- **Room temperature 25°C**: medium resistance, about 5kΩ.
- **Boiling water 100°C**: low resistance, about 400Ω.

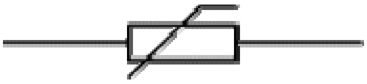

circuit symbol

Suppliers usually specify thermistors by their resistance at 25°C (room temperature). Thermistors take several seconds to respond to a sudden temperature change, small thermistors respond more rapidly.

A thermistor may be connected either way round and no special precautions are required when soldering. If it is going to be immersed in water the thermistor and its connections should be insulated because water is a weak conductor; for example they could be coated with polyurethane varnish.

Piezo Transducer

Piezo transducers are **output transducers** which convert an electrical signal to sound. They require a driver circuit (such as a 555 astable) to provide a signal and if this is near their natural (resonant) frequency of about 3kHz they will produce a particularly loud sound.

Piezo transducers require a small current, usually less than 10mA, so they can be connected directly to the outputs of most ICs. They are ideal for buzzes and beeps, but are not suitable for speech or music because they distort the sound. They are sometimes supplied with red and black leads, but they may be connected either way round. PCB-mounting versions are also available.

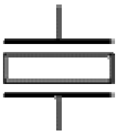

circuit symbol

Piezo transducers can also be used as **input transducers** for detecting sudden loud noises or impacts, effectively behaving as a crude microphone.

Loudspeaker

Loudspeakers are **output transducers** which convert an electrical signal to sound. Usually they are called 'speakers'. They require a driver circuit, such as a 555 astable or an audio amplifier, to provide a signal. There is a wide range available, but for many electronics projects a 300mW miniature loudspeaker is ideal. This type is about 70mm diameter and it is usually available with resistances of 8Ω and 64Ω. If a project specifies a 64Ω speaker you must use this higher resistance to prevent damage to the driving circuit.

Most circuits used to drive loudspeakers produce an audio (AC) signal which is combined with a constant DC signal. The DC will make a large current flow through the speaker due to its low resistance, possibly damaging both the speaker and the driving circuit. To prevent this happening a large value electrolytic capacitor is connected in series with the speaker, this blocks DC but passes audio (AC) signals. See capacitor coupling.

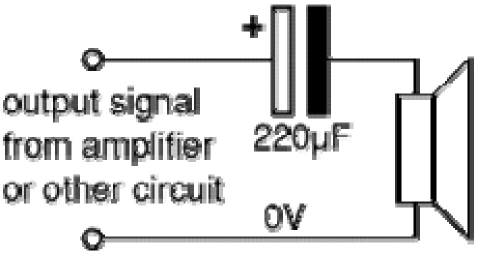

capacitor in series to block DC

Loudspeakers may be connected either way round except in stereo circuits when the + and - markings on their terminals must be observed to ensure the two speakers are in phase.

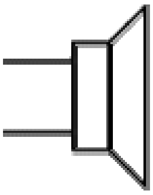

circuit symbol

Correct polarity must always be observed for large speakers in cabinets because the cabinet may contain a small circuit (a 'crossover network') which diverts the high frequency signals to a small speaker (a 'tweeter') because the large main speaker is poor at reproducing them.

Miniature loudspeakers can also be used as a microphone and they work surprisingly well, certainly good enough for speech in an intercom system for example.

Buzzer and Bleeper

These devices are **output transducers** converting electrical energy to sound. They contain an internal oscillator to produce the sound which is set at about 400Hz for buzzers and about 3kHz for bleeper.

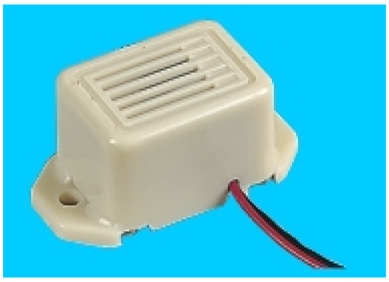

Buzzer (about 400Hz)

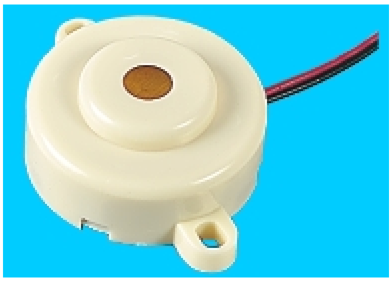

Bleeper (about 3kHz)

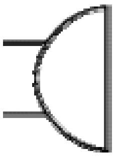

circuit symbol

Buzzers have a voltage rating but it is only approximate, for example 6V and 12V buzzers can be used with a 9V supply. Their typical current is about 25mA.

Bleepers have wide voltage

ranges, such as 3-30V, and they pass a low current of about 10mA.

Buzzers and bleepers must be connected the right way round, their red lead is positive (+).

Inductor (coil)

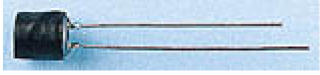

Inductor (miniature)

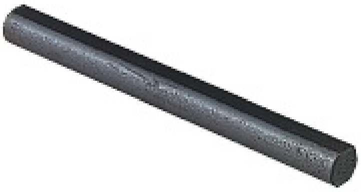

Ferrite rod

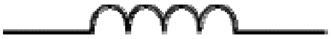

circuit symbol

An inductor is a coil of wire which may have a core of air, iron or ferrite (a brittle material made from iron). Its electrical property is called inductance and the unit for this is the Henry, symbol H. 1H is very large so mH and µH are used, 1000µH = 1mH and 1000mH = 1H. Iron and ferrite cores increase the inductance. Inductors are mainly used in tuned circuits and to block high frequency AC signals (they are sometimes called chokes). They pass DC easily, but block AC signals; this is the opposite of capacitors.

Inductors are rarely found in simple projects, but one exception is the tuning coil of a radio receiver. This is an inductor which you may have to make yourself by neatly winding enameled copper wire around a ferrite rod. Enameled copper wire has very thin insulation, allowing the turns of the coil to be close together, but this makes it impossible to strip in the usual way - the best method is to gently pull the ends of the wire through folded emery paper.
Warning: a ferrite rod is brittle so treat it like glass, not iron!

An inductor may be connected either way round and no special precautions are required when soldering.

Soldering Guide

How to Solder

First a few safety precautions:

- **Never touch the element or tip of the soldering iron.**
  They are very hot (about 400°C) and will give you a nasty burn.
- **Take great care to avoid touching the mains flex with the tip of the iron.**
  The iron should have a heatproof flex for extra protection. An ordinary plastic flex will melt immediately if touched by a hot iron and there is a serious risk of burns and electric shock.
- **Always return the soldering iron to its stand when not in use.**
  Never put it down on your workbench, even for a moment!
- **Work in a well-ventilated area.**
  The smoke formed as you melt solder is mostly from the flux and quite irritating. Avoid breathing it by keeping you head to the side of, not above, your work.
- **Wash your hands after using solder.**
  Solder contains lead which is a poisonous metal.

Preparing the soldering iron:

- **Place the soldering iron in its stand and plug in.**
  The iron will take a few minutes to reach its operating temperature of about 400°C.
- **Dampen the sponge in the stand.**
  The best way to do this is to lift it out the stand and hold it under a cold tap for a moment, then squeeze to remove excess water. It should be damp, not dripping wet.
- **Wait a few minutes for the soldering iron to warm up.**
  You can check if it is ready by trying to melt a little solder on the tip.
- **Wipe the tip of the iron on the damp sponge.**
  This will clean the tip.
- **Melt a little solder on the tip of the iron.**
  This is called 'tinning' and it will help the heat to flow from the iron's tip to the joint. It only needs to be done when you plug in the iron, and occasionally while soldering if you need to wipe the tip clean on the sponge.

You are now ready to start soldering:

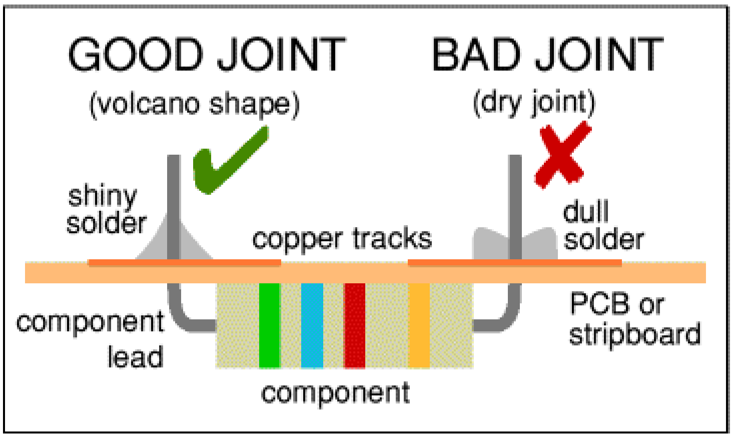

- **Hold the soldering iron like a pen, near the base of the handle.**
  Imagine you are going to write your name! Remember to never touch the hot element or tip.
- **Touch the soldering iron onto the joint to be made.**
  Make sure it touches both the component lead and the track. Hold the tip there for a few seconds and...
- **Feed a little solder onto the joint.**
  It should flow smoothly onto the lead and track to form a volcano shape as shown in the diagram. Apply the solder to the joint, not the iron.
- **Remove the solder, then the iron, while keeping the joint still.**
  Allow the joint a few seconds to cool before you move the circuit board.
- **Inspect the joint closely.**
  It should look shiny and have a 'volcano' shape. If not, you will need to reheat it and feed in a little more solder. This time ensure that **both** the lead and track are heated fully before applying solder.

Using a Heatsink

Some components, such as transistors, can be damaged by heat when soldering so if you are not an expert it is wise to use a heat sink clipped to the lead between the joint and the component body. You can buy a special tool, but a standard crocodile clip works just as well and is cheaper.

Crocodile clip

Further information

For a much more detailed guide to soldering, including troubleshooting, please see the Basic Soldering Guide on the Everyday Practical Electronics Magazine website.

Soldering Advice for Components

It is very tempting to start soldering components onto the circuit board straight away, but please take time to identify all the parts first. You are much less likely to make a mistake if you do this!

1. **Stick all the components onto a sheet of paper using sticky tape.**
2. **Identify each component** and write its name or value beside it.
3. **Add the code (R1, R2, C1 etc.) if necessary.**
   Many projects from books and magazines label the components with codes (R1, R2, C1, D1 etc.) and you should use the project's parts list to find these codes if they are given.
4. **Resistor values** can be found using the resistor color code which is explained on our Resistors page.
5. **Capacitor values** can be difficult to find because there are many types with different labeling systems! The various systems are explained on our Capacitors page.

Some components require special care when soldering. Many must be placed the correct way round and a few are easily damaged by the heat from soldering. Appropriate warnings are given in the table below, together with other advice which may be useful when soldering.

For most projects it is best to put the components onto the board in the order given below:

Components	Pictures	Reminders and Warnings
Chip Holders (DIL sockets)	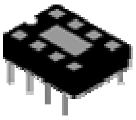	**Connect the correct way round** by making sure the notch is at the correct end. Do NOT put the ICs (chips) in yet.
Resistors	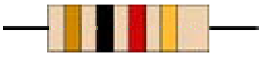	No special precautions are needed with resistors.
Small value capacitors (usually less than 1µF)	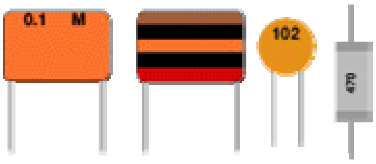	These may be connected either way round. Take care with polystyrene capacitors because they are easily damaged by heat.
Electrolytic capacitors (1µF and greater)	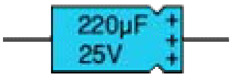	**Connect the correct way round.** They will be marked with a + or - near one lead.
Diodes	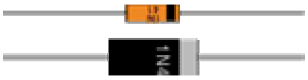	**Connect the correct way round.** Take care with germanium diodes (e.g. OA91) because they are easily damaged by heat.

LEDs	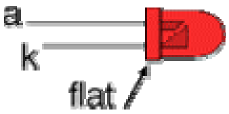	**Connect the correct way round.** The diagram may be labeled **a** or **+** for anode and **k** or **-** for cathode; yes, it really is k, not c, for cathode! The cathode is the short lead and there may be a slight flat on the body of round LEDs.
Transistors	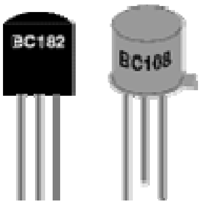	**Connect the correct way round.** Transistors have 3 'legs' (leads) so extra care is needed to ensure the connections are correct. Easily damaged by heat.
Wire Links between points on the circuit board.	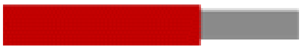 single core wire	Use single core wire, this is one solid wire which is plastic-coated. If there is no danger of touching other parts you can use tinned copper wire, this has no plastic coating and looks just like solder but it is stiffer.
Battery clips, **buzzers** and other parts with their own wires		**Connect the correct way round.**
Wires to parts off the circuit board, including **switches**, **relays**, **variable resistors** and **loudspeakers**.	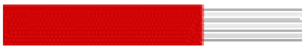 stranded wire	You should use stranded wire which is flexible and plastic-coated. Do not use single core wire because this will break when it is repeatedly flexed.
ICs (chips)	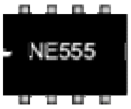	**Connect the correct way round.** **Many ICs are static sensitive.** Leave ICs in their antistatic packaging until you need them, then earth your hands by touching a metal water pipe or window frame before touching the ICs. **Carefully insert ICs in their holders**: make sure all the pins are lined up with the socket then push down firmly with your thumb.

What is solder?

Solder is an alloy (mixture) of tin and lead, typically 60% tin and 40% lead. It melts at a temperature of about 200°C. Coating a surface with solder is called 'tinning' because of the tin content of solder. Lead is poisonous and you should always wash your hands after using solder.

Solder for electronics use contains tiny cores of flux, like the wires inside a mains flex. The flux is corrosive, like an acid, and it cleans the metal surfaces as the solder melts. This is why you must melt the solder actually on the joint, not on the iron tip. Without flux most joints would fail because metals quickly oxidize and the solder itself will not flow properly onto a dirty, oxidized, metal surface.

Reels of solder

The best size of solder for electronics is 22swg (**SWG** = **S**tandard **W**ire **G**auge).

Desoldering

At some stage you will probably need to desolder a joint to remove or re-position a wire or component. There are two ways to remove the solder:

1. With a desoldering pump (solder sucker)

- Set the pump by pushing the spring-loaded plunger down until it locks.
- Apply both the pump nozzle and the tip of your soldering iron to the joint.
- Wait a second or two for the solder to melt.
- Then press the button on the pump to release the plunger and suck the molten solder into the tool.
- Repeat if necessary to remove as much solder as possible.
- The pump will need emptying occasionally by unscrewing the nozzle.

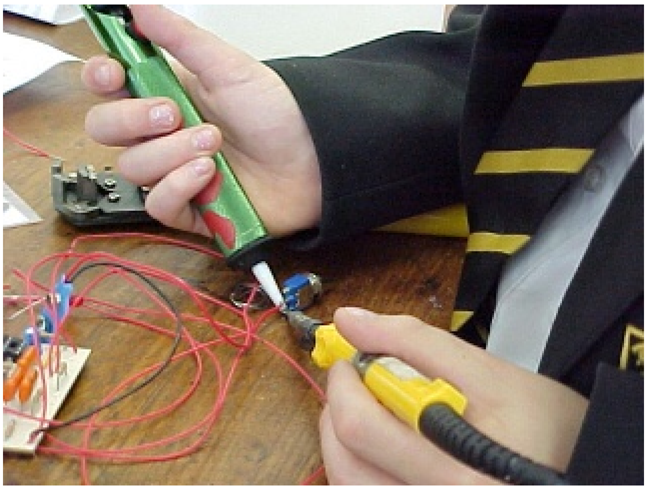

Using a desoldering pump (solder sucker)

2. With solder remover wick (copper braid)

- Apply both the end of the wick and the tip of your soldering iron to the joint.
- As the solder melts most of it will flow onto the wick, away from the joint.
- Remove the wick first, then the soldering iron.
- Cut off and discard the end of the wick coated with solder.

XYTRO WICK
SOLDER REMOVER
SW205

Solder remover wick

After removing most of the solder from the joint(s) you may be able to remove the wire or component lead straight away (allow a few seconds for it to cool). If the joint will not come apart easily apply your soldering iron to melt the remaining traces of solder at the same time as pulling the joint apart, taking care to avoid burning yourself.

First Aid for Burns

Most burns from soldering are likely to be minor and treatment is simple:

- **Immediately cool the affected area under gently running cold water.**
  Keep the burn in the cold water for at least 5 minutes (15 minutes is recommended). If ice is readily available this can be helpful too, but do not delay the initial cooling with cold water.
- **Do not apply any creams or ointments.**
  The burn will heal better without them. A dry dressing, such as a clean handkerchief, may be applied if you wish to protect the area from dirt.
- **Seek medical attention if the burn covers an area bigger than your hand.**

To reduce the risk of burns:

- Always return your soldering iron to its stand immediately after use.
- Allow joints and components a minute or so to cool down before you touch them.
- Never touch the element or tip of a soldering iron unless you are certain it is cold.

Tools Required for Electronics

Soldering iron

For electronics work the best type is one powered by mains electricity (230V in the UK), it should have a heatproof cable for safety. The iron's power rating should be 15 to 25W and it should be fitted with a small bit of 2 to 3mm diameter.

Other types of soldering iron:

Low voltage soldering irons are available, but their extra safety is undermined if you have a mains lead to their power supply! **Temperature controlled** irons are excellent for frequent use, but not worth the extra expense if you are a beginner. **Gas-powered** irons are designed for use where no mains supply is available and are not suitable for everyday use. Pistol shaped **solder guns** are far too powerful and cumbersome for normal electronics use.

Soldering iron stand

You must have a safe place to put the iron when you are not holding it. The stand should include a sponge which can be dampened for cleaning the tip of the iron.

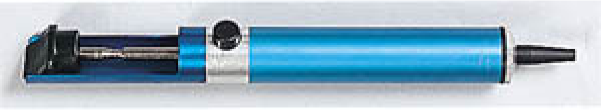

Desoldering pump (solder sucker)

A tool for removing solder when desoldering a joint to correct a mistake or replace a component.

XYTRO WICK
SOLDER REMOVER
SW205

Solder remover wick (copper braid)

This is an alternative to the desoldering pump shown above.

Reel of solder

The best size for electronics is 22SWG (SWG = Standard Wire Gauge).

Side Cutters

For trimming component leads close to the circuit board.

Wire Strippers

Most designs include a cutter as well, but they are not suitable for trimming component leads.

Small Pliers

Usually called 'snipe nose' pliers, these are for bending component leads etc. If you put a strong rubber band across the handles the pliers make a convenient holder for parts such as switches while you solder the contacts.

Rapid Electronics

Small flat-blade screwdriver

For scraping away excess flux and dirt between tracks, as well as driving screws!

Heat sink

You can buy a special tool, but a standard crocodile clip works just as well and is cheaper.

The following tool is only required if you are using stripboard:

Track Cutter

A 3mm drill bit can be used instead; in fact the tool is usually just a 3mm drill bit with a proper handle fitted.

The following tools are only required if you make your own PCBs:

PCB Rubber

This is an abrasive rubber for cleaning PCBs. It can also be used to clean stripboard where the copper tracks have become dull and tarnished.

Small Electric Drill

Ideally this should be mounted in a drill stand. You will need a range of small drill bits, but for most holes a 1mm bit is suitable. Larger holes can be drilled with a hand drill but 1mm bits are too fragile to use reliably in a hand drill.

Starter Kit of Components

If you are new to electronics and would like to try adapting published projects, or designing and building your own circuits, you need to have a small stock of components available. However, there is a very wide range of components and it can be difficult to know which ones you really need! I hope the list below will help you choose a sensible selection which is within your budget.

Remember that circuits built on breadboard can be dismantled after use and the components re-used. Remember that you will need to organize storage of the components!

Essential components

These are the components used in most projects. The individual components are quite cheap, but the total cost of the set will be significant! One way to spread the cost is to add a few items from this list every time you buy the components for a particular project. Click on the titles for further information.

Resistors

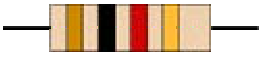

0.25W carbon film resistors are the cheapest type. Choose ones with 4-band colour codes because these are easier to read (the precision of 5-band codes is unnecessary).
Ideally you need a good selection of values over the range 100Ω to 1MΩ such as the E6 or E12 series, but that is a large number of resistors! As a minimum I suggest: 470*, 1k*, 2k2, 4k7, 10k*, 22k, 33k, 47k, 100k, 220k, 470k and 1MΩ. Buy at least 10 of each value and 20 of those marked *. The 470Ω resistors are for use with LEDs, even if a project specifies a slightly different value.

Resistors may be combined in series and parallel to obtain extra values, for example 100kΩ and 220kΩ in series is 320kΩ which is close enough to 330kΩ.

Capacitors

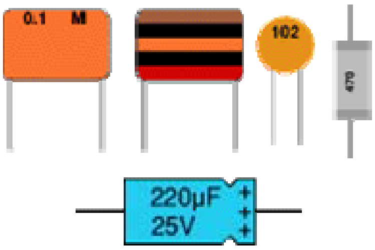

Low values: 0.01µF and 0.1µF metallized polyester, 10 of each.

High values: 1µF 63V, 10µF 25V, and 100µF 25V electrolytic with radial leads, 10 of each; 220µF 25V and 470µF 25V electrolytic with axial leads, 3 of each.

Diodes

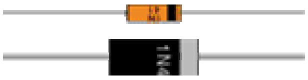

1N4148 signal diode and 1N4001 rectifier diode, 5 of each.

LEDs

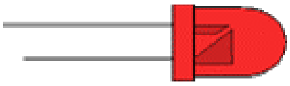

Red, yellow and green 5mm standard LEDs, 10 of each.

Transistors

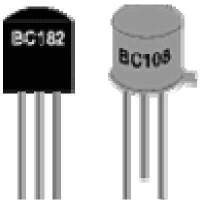

About 5 **general purpose, low power, NPN transistors**. These should have a maximum collector current (Ic max) of 100mA, and a minimum current gain (h_{FE} min) of 200.
For example: BC548B (BC108 equivalent).

About 5 **general purpose, medium power, NPN transistors**. These should have a maximum collector current (Ic max) of 1A, and a minimum current gain (h_{FE} min) of 30.
For example: BC639 (BFY51 equivalent).

Integrated circuits ('chips') and holders

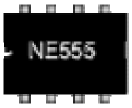

NE555 timer IC, at least 3 (10 if you plan to solder projects). It is not worth ordering other ICs at this stage unless you know they are needed for some of the projects you wish to try.

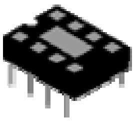

If you are planning to solder circuits on stripboard or PCB you will also need 8-pin, 14-pin and 16-pin DIL sockets (chip holders), at least 10 of each.

Variable Resistors

Presets are cheaper than standard variable resistors but most have pins which are too large for breadboards. For breadboard circuits is probably best to buy standard variable resistors and solder short single core 1/0.6mm wires onto them.

The useful values are: 10k LIN, 100k LIN and 1MΩ LIN, buy 2 of each. A 1MΩ LOG potentiometer is useful too. Knobs are optional because it is easy to turn the spindles by hand. If you buy presets the horizontal style are best, all presets are LIN.

Standard Variable Resistor

Battery Clips

Clip for a 9V PP3 battery; buy 3 (or 10 if you plan to solder projects). Remember to buy a battery too!

Wire

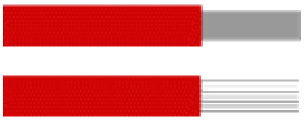

Red and black 7/0.2mm stranded wire, one color of single-core 1/0.6mm wire, 10m (or a reel) of each. If you are planning to build projects on breadboard buy extra colors of the single-core wire, including red and black.

Crocodile Clips

Buy at least one standard crocodile clip to use as a heat sink when soldering. Miniature red and black crocodile clips (buy about 10 of each) are useful for making your own test leads using 7/0.2mm stranded wire.

Switches

Switches are not essential for breadboard circuits because you can make or break links with pieces of wire. The on/off switch from soldered projects can also be omitted if you are willing to unclip the battery instead.

If you wish to buy a few switches the most useful types are push-to-make and miniature SPDT toggle switches, 3 of each.

Stripboard

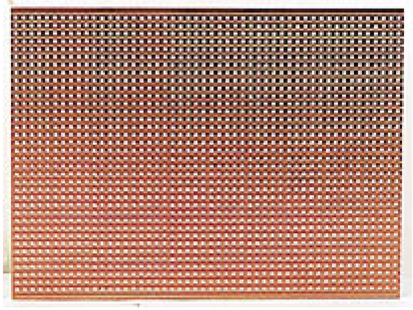

Buy a large sheet (or two) and cut it up as required. You can cut it neatly to size using a junior hacksaw, cutting along the lines of holes is easiest. For quickness you can break it over the edge of a workbench along the lines of holes - take care though because this needs a fairly large force and the edges will be rough. You may need to use a large pair of pliers to nibble away any jagged parts.

Avoid handling stripboard that you are not planning to use immediately because sweat from your hands will corrode the copper tracks and this will make soldering difficult unless you clean the board first.

Breadboard

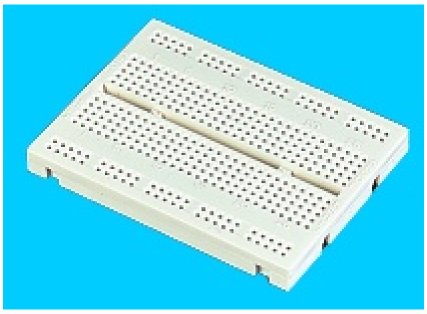

A small breadboard (such as the Protobloc 1 shown in the picture) is suitable for simple circuits with up to two ICs, but if you intend to build more complex circuits such as counters it is best to buy a larger breadboard (such as the Protobloc 2).

Breadboards do not require soldering so the components used on them can be re-used many times. They are ideal for testing your own circuit designs and trying out ideas such as adapting a published project.

Other components to consider

Buy these if you can afford them, or wait until they are needed for a project.

- A light dependent resistor (LDR), ORP12 or NORPS12.
- A thermistor, miniature disc type, NTC about 5kΩ@25°C.
- A piezo transducer with flying leads.
- A buzzer or bleeper.
- A miniature loudspeaker 8Ω.

Miniature piezo transducer

Storage Systems for Components

You can store all your components in a single container, such as a plastic food box, but as you accumulate more items it will become increasingly difficult to find the smaller components! A cheap solution is to organize the parts into small snap-top plastic bags which can be labeled, in fact you may find that some components are supplied like this.

Snap-top plastic bags

Probably the best storage system is a cabinet of plastic drawers. These can be expensive, but you do not need many drawers because there is no need to have a drawer for every single component value. Many parts can be grouped together, such as decades of resistor values. For example you could organize a 15-drawer cabinet like this:

1. Resistors 10Ω+ (third band **black**) only a few, but they tend to be large
2. Resistors 100Ω+ (third band **brown**)
3. Resistors 1kΩ+ (third band **red**)
4. Resistors 10kΩ+ (third band **orange**)
5. Resistors 100kΩ+ (third band **yellow**)
6. Resistors 1MΩ+ (third band **green**) and 10MΩ (third band **blue**)
7. Presets, also variable resistors if they will fit in the drawer
8. Capacitors low values, less than 1µF
9. Capacitors electrolytic 1µF+

Drawer cabinet

10. Diodes and transistors
11. LEDs and lamps (also LED clips and lampholders)
12. ICs (chips) and their holders (DIL sockets)
13. Switches and relays
14. Connectors (crocodile clips, plugs and sockets)
15. Other components (battery clips, piezo transducers, LDR, thermistors)

Frequently Asked Questions

Why is the letter I used to represent current?

The letter I seem to be an odd choice for the English language, but it was chosen in the early days of electricity to represent **intensity of current** which we simply call current today. The unit of current, the ampere, is named after the French scientist André-Marie Ampère in recognition of his work on the relationship between electric current and magnetism. Ampère referred to electric current as "**l'intensité du courant électrique**", so I was a logical choice to represent intensité (intensity). I am grateful to Barry Caruth for suggesting a search of the internet for "Ampère" and "l'intensité du courant électrique" which returns many sites as evidence (most of them French) enabling me to answer this question with confidence.

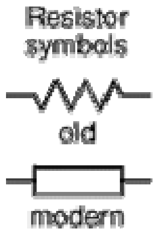

Why do some books use zigzag lines for resistors in circuit diagrams?

The zigzag line is the old symbol for a resistor and you may find it in older books and magazines. Unfortunately a few publications still use it! The correct modern symbol for a resistor is a rectangle.

My project has a resistor labeled 47, does that mean 47kΩ?

No, it means 47Ω which is 1000 times smaller! 47kΩ would be shortened to 47k (or 47K). The ohm (Ω) symbol is often omitted from circuit diagrams and component layouts but the k (meaning kilo = 1000) will always be included if it is needed.

Why do resistors have odd values like 47k and 56k, but not 50k?

There is a good reason for these odd values and it is explained on the Resistors page.

A project on another website lists a 10kW resistor! What does it mean?

It almost certainly means a 10kΩ resistor. This is a common error which occurs when the web page specifies a Greek font. If this font is not available on your computer you see the character in your standard font and it happens to be W which is the symbol for watt, the unit of power! I avoid the problem on this website by using a small image for Ω. In a few projects a low value resistor with a high power rating is required but the power will be something smaller like 5W, never 10kW which is more powerful than an electric heater!

My soldering iron was supplied with a hook, do I really need to buy a stand as well?

For safety you must buy (or make) a stand for your soldering iron. Please don't use the hook because it leaves exposed the very hot element and tip of the iron - it is too easy to accidentally touch them and burn yourself. If you can't afford to buy a stand you could try making your own with a spiral of stiff galvanized iron wire (a coat-hanger?) screwed to a block of wood. Ideally the stand should include a damp sponge for safely wiping the tip of the iron when it needs cleaning.

My teacher says that Christmas tree lights are a series circuit, so when one lamp blew on Christmas Eve why didn't they all go out?

Traditional Christmas tree lights are connected in series and you are correct in thinking that if one lamp blows all the lamps should go out. The problem is that Christmas tree lights are not like ordinary lamps! When they blow they automatically short-circuit (they become like a wire link) so the circuit is still complete and the other lamps remain lit. This makes it easy to see the blown lamp, but do remember to switch off before changing it.

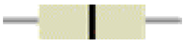

What component has a black stripe in the centre (it looks like a diode)?

A small component about the size of a resistor or signal diode with a single black stripe in the centre is a **zero-ohm resistor**, it is really just a wire link! These components are used on commercial PCBs because they are easier for machines to handle than small pieces of wire. The single black stripe is logical because it means zero in the resistor color code. Ordinary resistors have at least four stripes. Diodes have a single stripe near one end, not in the centre.

What is a "short circuit"?

A "short circuit" is a connection of very low resistance such as a wire (almost 0Ω) which provides a very easy path for current. Think of it as an electrical short-cut. It is normally used to describe a fault or accidental connection rather than a deliberate one.
For example: if the leads from a battery touch one another they create a very low resistance connection across the battery, so we say they have caused a short circuit across the battery. Current will flow through this short circuit rather than through the proper circuit. This stops the circuit working and it may cause a fire because the leads and battery will become hot with a large current flowing.

What does "open circuit" mean?

"Open circuit" means no connection. It is usually used to describe a break in some part of a circuit which could be deliberate (such as a switch in the open or off position) or a fault (such as a broken wire or burnt out component).

How do I choose a relay to use with one of your projects?

The 555 timer chip used in many projects can supply current up to 200mA so it can power most relays directly. However, you must connect a signal diode (a 1N4148 for example) in parallel across the relay coil to protect the 555. Note that this diode is connected 'backwards' so that it will normally **not** conduct.

I want to use a large number of LEDs, do I need a resistor for each one?

No, you can usually connect a few LEDs of the same type in series and just use one resistor. The number of LEDs you can connect in series depends on the circuit's supply voltage. This arrangement has the advantage of reducing the total current required by the circuit. Please see the LEDs page for more details:

I want to build up a stock of components, what should I buy first?

Most people build their first few projects from complete kits, but if you want to try adapting published projects or designing and building your own circuits you will need to have a small stock of components available. There is a page with advice on buying a starter kit of components.

What is a Darlington pair?

A Darlington pair is two transistors connected together so that the current amplified by the first is further amplified by the second transistor, giving a very high gain of 10000 or so.

What does 'sinking a current' mean?

It means current is flowing into the output of a chip. This happens when the output is low (0V) if there is a device connected between the positive supply (+Vs) and the output. It is the opposite of sourcing a current which means current is flowing out of the output. Most chip outputs can both sink and source current.

Are 'time period' and 'time constant' the same thing?

No, they have different meanings although both are time. **Time period** is the duration of a single pulse or the time for one cycle of a repeating electrical signal. **Time constant** is a property of a changing system, such as a capacitor charging and discharging.

What is a PIC?

A PIC is a **P**rogrammable **I**ntegrated **C**ircuit microcontroller, a 'computer-on-a-chip'. They have a processor and memory to run a program responding to inputs and controlling outputs, so they can easily achieve complex functions which would require several conventional ICs. I can strongly recommend the PICAXE system because it is easy to program (and re-program) the PICs with a standard computer - no specialist equipment is required other than a low-cost download lead. The programming software and extensive documentation is available to download free of charge, making the system ideal for education and users at home.

Why does my circuit count 3 or 4 when I press the switch once?

This is likely to happen if a switch is connected directly to the clock input of a counter. Switches contacts tend to rapidly bounce open and closed a few times when the switch is operated. The counter sees this as several clock pulses, not the single pulse you expect. One solution is to make the switch trigger a monostable circuit with a short time constant (0.1s for example) and use this to drive the clock input.

What is a schematic?

It is a circuit diagram.

What does 'SMD' mean?

'SMD' means Surface Mount Device. SMDs are components with small pads instead of leads for their contacts. They are designed for soldering by machine onto specially designed PCBs and are not suitable for educational or hobby circuits constructed on breadboard or stripboard. Do not buy SMD components for your projects!

CHAPTER 24

THE SAILINGS

INTRODUCTION

2400. Introduction

Dead reckoning involves the determination of one's present or future position by projecting the ship's course and distance run from a known position. A closely related problem is that of finding the course and distance from one known point to another known point. For short distances, these problems are easily solved directly on charts, but for long distances, a purely mathematical solution is often a better method. Collectively, these methods are called **The Sailings**.

Navigational computer programs and calculators commonly contain algorithms for computing all of the problems of the sailings. For those situations when a calculator is not available, this chapter also discusses sailing solutions by Table 4, the Traverse Tables.

2401. Rhumb Lines And Great Circles

The principal advantage of a rhumb line is that it maintains constant true direction. A ship following the rhumb line between two places does not change true course. A rhumb line makes the same angle with all meridians it crosses and appears as a straight line on a Mercator chart. For any other case, the difference between the rhumb line and the great circle connecting two points increases (1) as the latitude increases, (2) as the difference of latitude between the two points decreases, and (3) as the difference of longitude increases.

A great circle is the intersection of the surface of a sphere and a plane passing through the center of the sphere. It is the largest circle that can be drawn on the surface of the sphere, and is the shortest distance along the surface between any two points. Any two points are connected by only one great circle unless the points are **antipodal** (180° apart on the earth), and then an infinite number of great circles passes through them. Every great circle bisects every other great circle. Thus, except for the equator, every great circle lies exactly half in the Northern Hemisphere and half in the Southern Hemisphere. Any two points 180° apart on a great circle have the same latitude numerically, but contrary names, and are 180° apart in longitude. The point of greatest latitude is called the **vertex**. For each great circle, there is a vertex in each hemisphere, 180° apart in longitude. At these points the great circle is tangent to a parallel of latitude, and its direction is due east-west. On each side of these vertices the direction changes progressively until the intersection with the equator is reached, 90° in longitude away, where the great circle crosses the equator at an angle equal to the latitude of the vertex.

On a Mercator chart a great circle appears as a sine curve extending equal distances each side of the equator. The rhumb line connecting any two points of the great circle on the same side of the equator is a chord of the curve. Along any intersecting meridian the great circle crosses at a higher latitude than the rhumb line. If the two points are on opposite sides of the equator, the direction of curvature of the great circle relative to the rhumb line changes at the equator. The rhumb line and great circle may intersect each other, and if the points are equal distances on each side of the equator, the intersection takes place at the equator.

Great circle sailing takes advantage of the shorter distance along the great circle between two points, rather than the longer rhumb line. The arc of the great circle between the points is called the **great circle track**. If it could be followed exactly, the destination would be dead ahead throughout the voyage (assuming course and heading were the same). The rhumb line appears the more direct route on a Mercator chart because of chart distortion. The great circle crosses meridians at higher latitudes, where the distance between them is less. This is why the great circle route is shorter than the rhumb line.

The decision as to whether or not to use great-circle sailing depends upon the conditions. The saving in distance should be worth the additional effort, and of course the great circle route cannot cross land, nor should it carry the vessel into dangerous waters. **Composite sailing** (see section 2402 and section 2411) may save time and distance over the rhumb line track without leading the vessel into danger.

Since great circles other than a meridian or the equator are curved lines whose true direction changes continually, the navigator does not attempt to follow it exactly. Rather, he selects a number of points along the great circle, constructs rhumb lines between the points, and follows these rhumb lines from point to point.

2402. Kinds Of Sailings

There are seven types of sailings:

1. **Plane sailing** solves problems involving a single course and distance, difference of latitude, and de-

parture, in which the earth is regarded as a plane surface. This method, therefore, provides solution for latitude of the point of arrival, but not for longitude. To calculate the longitude, the spherical sailings are necessary. Do not use this method for distances of more than a few hundred miles.
2. **Traverse sailing** combines the plane sailing solutions when there are two or more courses and determines the equivalent course and distance made good by a vessel steaming along a series of rhumb lines.
3. **Parallel sailing** is the interconversion of departure and difference of longitude when a vessel is proceeding due east or due west.
4. **Middle- (or mid-) latitude sailing** uses the mean latitude for converting departure to difference of longitude when the course is not due east or due west.
5. **Mercator sailing** provides a mathematical solution of the plot as made on a Mercator chart. It is similar to plane sailing, but uses meridional difference and difference of longitude in place of difference of latitude and departure.
6. **Great circle sailing** involves the solution of courses, distances, and points along a great circle between two points.
7. **Composite sailing** is a modification of great-circle sailing to limit the maximum latitude, generally to avoid ice or severe weather near the poles.

2403. Terms And Definitions

In solutions of the sailings, the following quantities are used:

1. **Latitude (L)**. The latitude of the point of departure is designated L_l; that of the destination, L_2; middle (mid) or mean latitude, L_m; latitude of the vertex of a great circle, L_v; and latitude of any point on a great circle, L_x.
2. **Mean latitude (L_m)**. Half the arithmetical sum of the latitudes of two places on the same side of the equator.
3. **Middle or mid latitude (L_m)**. The latitude at which the arc length of the parallel separating the meridians passing through two specific points is exactly equal to the departure in proceeding from one point to the other. The mean latitude is used when there is no practicable means of determining the middle latitude.
4. **Difference of latitude (l or DLat.)**.
5. **Meridional parts (M)**. The meridional parts of the point of departure are designated M_l, and of the point of arrival or the destination, M_2.
6. **Meridional difference (m)**.
7. **Longitude (λ)**. The longitude of the point of departure is designated λ_1; that of the point of arrival or the destination, λ_2; of the vertex of a great circle, l_v; and of any point on a great circle, λ_x
8. **Difference of longitude (DLo)**.
9. **Departure (p or Dep.)**.
10. **Course** or **course angle (Cn or C)**.
11. **Distance (D or Dist.)**.

GREAT CIRCLE SAILING

2404. Great Circle Sailing By Chart

Navigators can most easily solve great-circle sailing problems graphically. DMAHTC publishes several gnomonic projections covering the principal navigable waters of the world. On these **great circle charts**, any straight line is a great circle. The chart, however, is not conformal; therefore, the navigator cannot directly measure directions and distances as on a Mercator chart.

The usual method of using a gnomonic chart is to plot the route and pick points along the track every 5° of longitude using the latitude and longitude scales in the immediate vicinity of each point. These points are then transferred to a Mercator chart and connected by rhumb lines. The course and distance for each leg is measured on the Mercator chart. See Chapter 25 for a discussion of this process.

2405. Great Circle Sailing By Sight Reduction Tables

Any method of solving a celestial spherical triangle can be used for solving great circle sailing problems. The point of departure replaces the assumed position of the observer, the destination replaces the geographical position of the body, difference of longitude replaces meridian angle or local hour angle, initial course angle replaces azimuth angle, and great circle distance replaces zenith distance (90° - altitude). See Figure 2405. Therefore, any table of azimuths (if the entering values are meridian angle, declination, and latitude) can be used for determining initial great-circle course. Tables which solve for altitude, such as *Pub. No. 229*, can be used for determining great circle distance. The required distance is 90° - altitude.

In inspection tables such as *Pub. No. 229*, the given combination of L_1, L_2, and DLo may not be tabulated. In this case reverse the name of L_2 and use 180° - DLo for entering the table. The required course angle is then 180° minus the tabulated azimuth, and distance is 90° plus the altitude. If neither combination can be found, solution cannot be made by that method. By interchanging L_1 and L_2, one can find the supplement of the final course angle.

Solution by table often provides a rapid approximate check, but accurate results usually require triple interpolation. Except for *Pub. No. 229*, inspection tables do not

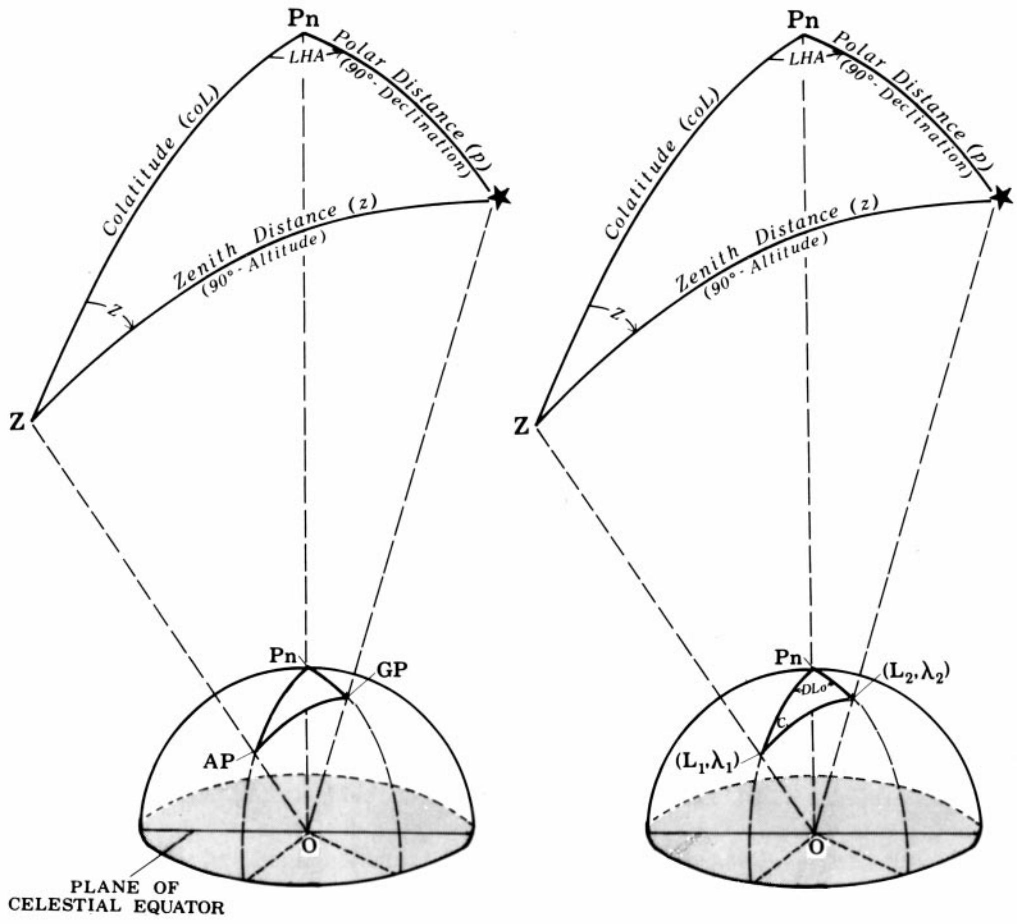

Figure 2405. Adapting the astronomical triangle to the navigational triangle of great circle sailing.

provide a solution for points along the great circle. *Pub. No. 229* provides solutions for these points only if interpolation is not required.

2406. Great Circle Sailing By *Pub. No. 229*

By entering *Pub. No. 229* with the latitude of the point of departure as latitude, latitude of destination as declination, and difference of longitude as LHA, the tabular altitude and azimuth angle may be extracted and converted to great-circle distance and course. As in sight reduction, the tables are entered according to whether the name of the latitude of the point of departure is the same as or contrary to the name of the latitude of the destination (declination). If the values correspond to those of a celestial body above the celestial horizon, 90° minus the arc of the tabular altitude becomes the distance; the tabular azimuth angle becomes the initial great-circle course angle. If the respondents correspond to those of a celestial body below the celestial horizon, the arc of the tabular altitude plus 90° becomes the distance; the supplement of the tabular azimuth angle becomes the initial great-circle course angle.

When the Contrary/Same (CS) Line is crossed in either direction, the altitude becomes negative; the body lies below the celestial horizon. For example: If the tables are entered with the LHA (DLo) at the bottom of a right-hand page and declination (L_2) such that the respondents lie above the CS Line, the CS Line has been crossed. Then the distance is 90° plus the tabular altitude; the initial course angle is the supplement of the tabular azimuth angle. Similarly, if the tables are entered with the LHA (DLo) at the top of a right-hand page and the respondents are found below the CS Line, the distance is 90° plus the tabular altitude; the initial course angle is the supplement of the tabular azimuth angle. If the tables are entered with the LHA (DLo) at the bottom of a right-hand page and the name of L_2 is contrary to L_1, the respondents are found in the column for L_1 on the facing page. In this case, the CS Line has been crossed; the distance is 90° plus the tabular altitude; the initial course angle is the supplement of the tabular azimuth angle.

The tabular azimuth angle, or its supplement, is prefixed N or S for the latitude of the point of departure and suffixed E or W depending upon the destination being east or west of the point of departure.

If all entering arguments are integral degrees, the distance and course angle are obtained directly from the tables without interpolation. If the latitude of the destination is nonintegral, interpolation for the additional minutes of latitude is done as in correcting altitude for any declination increment; if the latitude of departure or difference of longitude is nonintegral, the additional interpolation is done graphically.

Since the latitude of destination becomes the declination entry, and all declinations appear on every page, the great circle solution can always be extracted from the volume which covers the latitude of the point of departure.

Example 1: *Using Pub. No. 229 find the distance and initial great circle course from lat. 32°S, long. 116°E to lat. 30°S, long. 31°E.*

Solution: *Refer to Figure 2405. The point of departure (lat. 32°S, long. 116°E) replaces the AP of the observer; the destination (lat. 30°S, long. 31°E) replaces the GP of the celestial body; the difference of longitude (DLo 85°) replaces local hour angle (LHA) of the body.*

Enter Pub. 229, Volume 3 with lat. 32° (Same Name), LHA 85°, and declination 30°. The respondents correspond to a celestial body above the celestial horizon. Therefore, 90° minus the tabular altitude (90° - 19°12.4' = 70°47.6') becomes the distance; the tabular azimuth angle (S66.0°W) becomes the initial great circle course angle, prefixed S for the latitude of the point of departure and suffixed W due to the destination being west of the point of departure.

Answer:
D = 4248 nautical miles
C = S66.0°W = 246.0°.

Example 2: *Using Pub. No. 229 find the distance and initial great circle course from lat. 38°N, long. 122°W to lat. 24°S, long. 151°E.*

Solution: *Refer to Figure 2405. The point of departure (lat. 38°N, long. 122°W) replaces the AP of the observer; the destination (lat. 24°S, long. 151°E) replaces the GP of the celestial body; the difference of longitude (DLo 87°) replaces local hour angle (LHA) of the body*

Enter Pub. No. 229 Volume 3 with lat. 38° (Contrary Name), LHA 87°, and declination 24°. The respondents correspond to those of a celestial body below the celestial horizon. Therefore, the tabular altitude plus 90° (12°17.0' + 90° = 102°17.0') becomes the distance; the supplement of tabular azimuth angle (180° - 69.0° = 111.0°) becomes the initial great circle course angle, prefixed N for the latitude of the point of departure and suffixed W since the destination is west of the point of departure.

Note that the data is extracted from across the CS Line from the entering argument (LHA 87°), indicating that the corresponding celestial body would be below the celestial horizon.

Answer:
D = 6137 nautical miles
C = N111.0°W = 249°.

2407. Great Circle Sailing By Computation

In Figure 2407, 1 is the point of departure, 2 the destination, P the pole nearer 1, l-X-V-2 the great circle through 1 and 2, V the vertex, and X any point on the great circle. The arcs P1, PX, PV, and P2 are the colatitudes of points 1, X, V, and 2, respectively. If 1 and 2 are on opposite sides of the equator, P2 is $90° + L_2$. The length of arc 1-2 is the great-circle distance between 1 and 2. Arcs 1-2, P1, and P2 form a spherical triangle. The angle at 1 is the initial great-circle course from 1 to 2, that at 2 the supplement of the final great-circle course (or the initial course from 2 to 1), and that at P the DLo between 1 and 2.

Great circle sailing by computation usually involves solving for the initial great circle course; the distance; latitude and longitude, and sometimes the distance, of the vertex; and the latitude and longitude of various points (X) on the great circle. The computation for initial course and the distance involves solution of an oblique spherical triangle, and any method of solv-

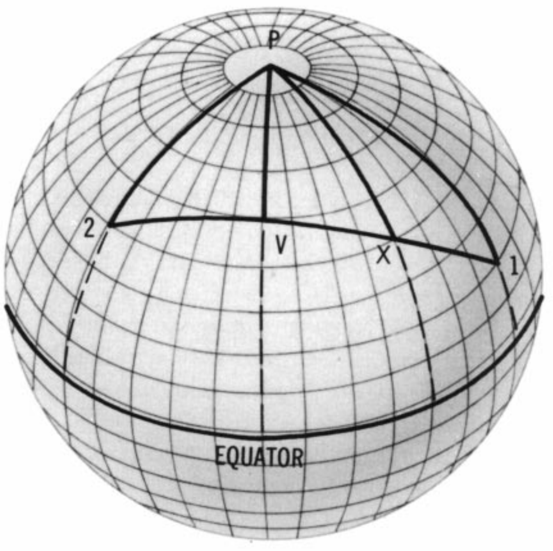

Figure 2407. The navigational triangle and great circle sailing.

ing such a triangle can be used. If 2 is the geographical position (GP) of a celestial body (the point at which the body is in the zenith), this triangle is solved in celestial navigation, except that 90° - D (the altitude) is desired instead of D. The solution for the vertex and any point X usually involves the solution of right spherical triangles.

2408. Points Along The Great Circle

If the latitude of the point of departure and the initial great-circle course angle are integral degrees, points along the great circle are found by entering the tables with the latitude of departure as the latitude argument (always Same Name), the initial great circle course angle as the LHA argument, and 90° minus distance to a point on the great circle as the declination argument. The latitude of the point on the great circle and the difference of longitude between that point and the point of departure are the tabular altitude and azimuth angle, respectively. If, however, the respondents are extracted from across the CS Line, the tabular altitude corresponds to a latitude on the side of the equator opposite from that of the point of departure; the tabular azimuth angle is the supplement of the difference of longitude.

Example 1: *Find a number of points along the great circle from latitude 38°N, longitude 125°W when the initial great circle course angle is N111°W.*

Solution: *Entering the tables with latitude 38° (Same Name), LHA 111°, and with successive declinations of 85°, 80°, 75°, etc., the latitudes and differences in longitude from 125°W are found as tabular altitudes and azimuth angles respectively:*

Answer:

D (NM)	*300*	*600*	*900*	*3600*
D (arc)	*5°*	*10°*	*15°*	*60°*
dec	*85°*	*80°*	*75°*	*30°*
Lat.	*36.1° N*	*33.9° N*	*31.4° N*	*3.6° N*
Dep.	*125° W*	*125° W*	*125° W*	*125° W*
DLo	*5.8°*	*11.3°*	*16.5°*	*54.1°*
Long	*130.8°W*	*136.3°W*	*141.5°W*	*179.1°*

Example 2: *Find a number of points along the great circle track from latitude 38°N, long. 125°W when the initial great circle course angle is N 69° W.*

Solution: *Enter the tables with latitude 38° (Same Name), LHA 69°, and with successive declinations as shown. Find the latitudes and differences of longitude from 125°W as tabular altitudes and azimuth angles, respectively:*

Answer:

D (NM.)	*300*	*600*	*900*	*6600*
D (arc)	*5°*	*10°*	*15°*	*110°*
dec	*85°*	*80°*	*75°*	*20°*
Lat.	*39.6° N*	*40.9° N*	*41.9° N*	*3.1° N*
Dep.	*125° W*	*125° W*	*125° W*	*125° W*
DLo	*6.1°*	*12.4°*	*18.9°*	*118.5°*
Long	*131.1°W*	*137.4°W*	*143.9°W*	*116.5°E*

2409. Finding The Vertex

Using *Pub. No. 229* to find the approximate position of the vertex of a great circle track provides a rapid check on the solution by computation. This approximate solution is also useful for voyage planning purposes.

Using the procedures for finding points along the great circle, inspect the column of data for the latitude of the point of departure and find the maximum value of tabular altitude. This maximum tabular altitude and the tabular azimuth angle correspond to the latitude of the vertex and the difference of longitude of the vertex and the point of departure.

Example 1: *Find the vertex of the great circle track from lat. 38°N, long. 125°W when the initial great circle course angle is N69°W.*

Solution: *Enter Pub. No. 229 with lat. 38° (Same Name), LHA 69°, and inspect the column for lat. 38° to find the maximum tabular altitude. The maximum altitude is 42°38.1' at a distance of 1500 nautical miles (90° - 65° = 25°) from the point of departure. The corresponding tabular azimuth angle is 32.4°. Therefore, the difference of longitude of vertex and point of departure is 32.4°.*

Answer:

Latitude of vertex = 42°38.1'N.
Longitude of vertex = 125° + 32.4° = 157.4°W.

2410. Altering A Great Circle Track To Avoid Obstructions

Land, ice, or severe weather may prevent the use of great circle sailing for some or all of one's route. One of the principal advantages of solution by great circle chart is that the presence of any hazards is immediately apparent. The pilot charts are particularly useful in this regard. Often a relatively short run by rhumb line is sufficient to reach a point from which the great circle track can be followed. Where a choice is possible, the rhumb line selected should conform as nearly as practicable to the direct great circle.

If the great circle route crosses a navigation hazard, change the track. It may be satisfactory to follow a great circle to the vicinity of the hazard, one or more rhumb lines along the edge of the hazard, and another great circle to the destination. Another possible solution is the use of composite sailing; still another is the use of two great circles, one from the point of departure to a point near the maximum latitude of unobstructed water and the second from this point to the destination.

2411. Composite Sailing

When the great circle would carry a vessel to a higher latitude than desired, a modification of great circle sailing called **composite sailing** may be used to good advantage. The composite track consists of a great circle from the point of departure and tangent to the limiting parallel, a course line along the parallel, and a great circle tangent to the limiting parallel and through the destination.

Solution of composite sailing problems is most easily made with a great circle chart. For this solution, draw lines from the point of departure and the destination, tangent to the limiting parallel. Then measure the coordinates of various selected points along the composite track and transfer them to a Mercator chart, as in great circle sailing. Composite sailing problems can also be solved by computation, using the equation:

$$\cos DLo_{vx} = \tan L_x \cot L_v$$

The point of departure and the destination are used successively as point X. Solve the two great circles at each end of the limiting parallel, and use parallel sailing along the limiting parallel. Since both great circles have vertices at the same parallel, computation for C, D, and DLo_{vx} can be made by considering them parts of the same great circle with L_1, L_2, and L_v as given and $DLo = DLo_{v1} + DLo_{v2}$. The total distance is the sum of the great circle and parallel distances.

TRAVERSE TABLES

2412. Using Traverse Tables

Traverse tables can be used in the solution of any of the sailings except great circle and composite. They consist of the tabulation of the solutions of plane right triangles. Because the solutions are for integral values of the course angle and the distance, interpolation for intermediate values may be required. Through appropriate interchanges of the headings of the columns, solutions for other than plane sailing can be made. For the solution of the plane right triangle, any value N in the distance (Dist.) column is the hypotenuse; the value opposite in the difference of latitude (D. Lat.) column is the product of N and the cosine of the acute angle; and the other number opposite in the departure (Dep.) column is the product of N and the sine of the acute angle. Or, the number in the D. Lat. column is the value of the side adjacent, and the number in the Dep. column is the value of the side opposite the acute angle. Hence, if the acute angle is the course angle, the side adjacent in the D. Lat. column is meridional difference m; the side opposite in the Dep. column is DLo. If the acute angle is the midlatitude of the formula $p = DLo \cos Lm$, then DLo is any value N in the Dist. column, and the departure is the value $N \times \cos L_m$ in the D. Lat. column.

The examples below clarify the use of the traverse tables for plane, traverse, parallel, mid latitude, and Mercator sailings.

2413. Plane Sailing

In plane sailing the figure formed by the meridian through the point of departure, the parallel through the point of arrival, and the course line is considered a plane right triangle. This is illustrated in Figure 2413a. P_1 and P_2 are the points of departure and arrival, respectively. The course angle and the three sides are as labeled. From this triangle:

$$\cos C = \frac{l}{D} \qquad \sin C = \frac{p}{D} \qquad \tan C = \frac{p}{l}.$$

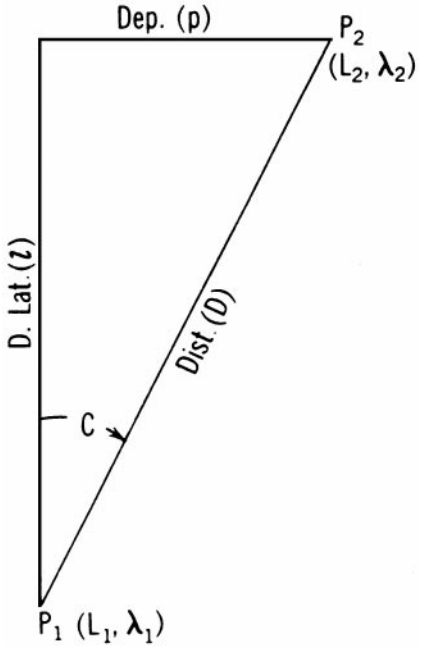

Figure 2413a. The plane sailing triangle.

From the first two of these formulas the following relationships can be derived:

$l = D \cos C \qquad D = l \sec C \qquad p = D \sin C.$

Label l as N or S, and p as E or W, to aid in identification of the quadrant of the course. Solutions by calculations and traverse tables are illustrated in the following examples:

Example 1: *A vessel steams 188.0 miles on course 005°.*

Required: *(1) (a) Difference of latitude and (b) departure by computation. (2) (a) difference of latitude and (b) departure by traverse table.*

Solution:

(1) (a) Difference of latitude by computation:

$$\begin{aligned} \text{diff latitude} &= D \times \cos C \\ &= 188.0 \text{ miles} \times \cos(005°) \\ &= 187.3 \text{ arc min} \\ &= 3° \ 07.3' \text{ N} \end{aligned}$$

(1) (b) Departure by computation:

$$\begin{aligned} \text{departure} &= D \times \sin C \\ &= 188.0 \text{ miles} \times \sin(005°) \\ &= 16.4 \text{ miles} \end{aligned}$$

Answer:
Diff. Lat. = 3° 07.3' N
departure = 16.4 miles

(2) Difference of latitude and departure by traverse table:

Refer to Figure 2413b. Enter the traverse table and find course 005° at the top of the page. Using the column headings at the top of the table, opposite 188 in the Dist. column extract D. Lat. 187.3 and Dep. 16.4.

(a) D. Lat. = 187.3' N.
(b) Dep. = 16.4 mi. E.

Example 2: *A ship has steamed 136.0 miles north and 203.0 miles west.*

Required: *(1) (a) Course and (b) distance by computation. (2) (a) course and (b) distance by traverse table.*

Solution:

(1) (a) Course by computation:

$$C = \arctan \frac{departure}{diff.\ \text{lat.}}$$

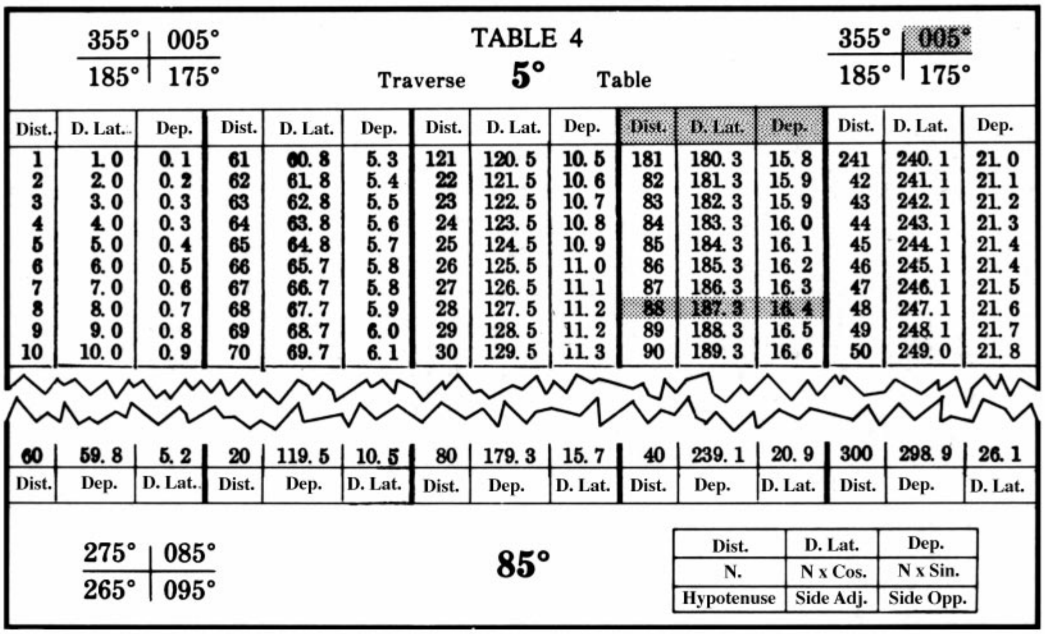

355° | 005°
185° | 175°

TABLE 4
Traverse 5° Table

355° | 005°
185° | 175°

Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.
1	1.0	0.1	61	60.8	5.3	121	120.5	10.5	181	180.3	15.8	241	240.1	21.0
2	2.0	0.2	62	61.8	5.4	22	121.5	10.6	82	181.3	15.9	42	241.1	21.1
3	3.0	0.3	63	62.8	5.5	23	122.5	10.7	83	182.3	15.9	43	242.1	21.2
4	4.0	0.3	64	63.8	5.6	24	123.5	10.8	84	183.3	16.0	44	243.1	21.3
5	5.0	0.4	65	64.8	5.7	25	124.5	10.9	85	184.3	16.1	45	244.1	21.4
6	6.0	0.5	66	65.7	5.8	26	125.5	11.0	86	185.3	16.2	46	245.1	21.4
7	7.0	0.6	67	66.7	5.8	27	126.5	11.1	87	186.3	16.3	47	246.1	21.5
8	8.0	0.7	68	67.7	5.9	28	127.5	11.2	88	187.3	16.4	48	247.1	21.6
9	9.0	0.8	69	68.7	6.0	29	128.5	11.2	89	188.3	16.5	49	248.1	21.7
10	10.0	0.9	70	69.7	6.1	30	129.5	11.3	90	189.3	16.6	50	249.0	21.8
60	59.8	5.2	20	119.5	10.5	80	179.3	15.7	40	239.1	20.9	300	298.9	26.1
Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.

275° | 085°
265° | 095°

85°

Dist.	D. Lat.	Dep.
N.	N x Cos.	N x Sin.
Hypotenuse	Side Adj.	Side Opp.

Figure 2413b. Extract from Table 4.

$$C = arc\,tan\frac{203.0}{136.0}$$

$$C = N\ \ 56°\ 10.8'\ W$$

$$C = 304°(to\ nearest\ degree)$$

Draw the course vectors to determine the correct course. In this case the vessel has gone north 136 miles and west 203 miles. The course, therefore, must have been between 270° and 360°. No solution other than 304° is reasonable.

(1) (b) Distance by computation:

$$\begin{aligned} D &= diff.\ latitude \times sec\ C \\ &= 136\ miles \times sec\ (304°) \\ &= 136\ miles \times 1.8 \\ &= 244.8\ miles \end{aligned}$$

Answer:
C = 304°
D = 244.8 miles

(2) Solution by traverse table:

Refer to Figure 2413c. Enter the table and find 136 and 203 beside each other in the columns labeled D. Lat. and Dep., respectively. This occurs most nearly on the page for course angle 56°. Therefore, the course is 304°. Interpolating for intermediate values, the corresponding number in the Dist. column is 244.3 miles.

Answer:

(a) C = 304°
(b) D = 244.3 mi.

2414. Traverse Sailing

A **traverse** is a series of courses or a track consisting of a number of course lines, such as might result from a sailing vessel beating into the wind. **Traverse sailing** is the finding of a single equivalent course and distance.

Though the problem can be solved graphically on the chart, traverse tables provide a mathematical solution. The distance to the north or south and to the east or west on each course is tabulated, the algebraic sum of difference of latitude and departure is found, and converted to course and distance.

Example: *A ship steams as follows: course 158°, distance 15.5 miles; course 135°, distance 33.7 miles; course 259°, distance 16.1 miles; course 293°, distance 39.0 miles; course 169°, distance 40.4 miles.*

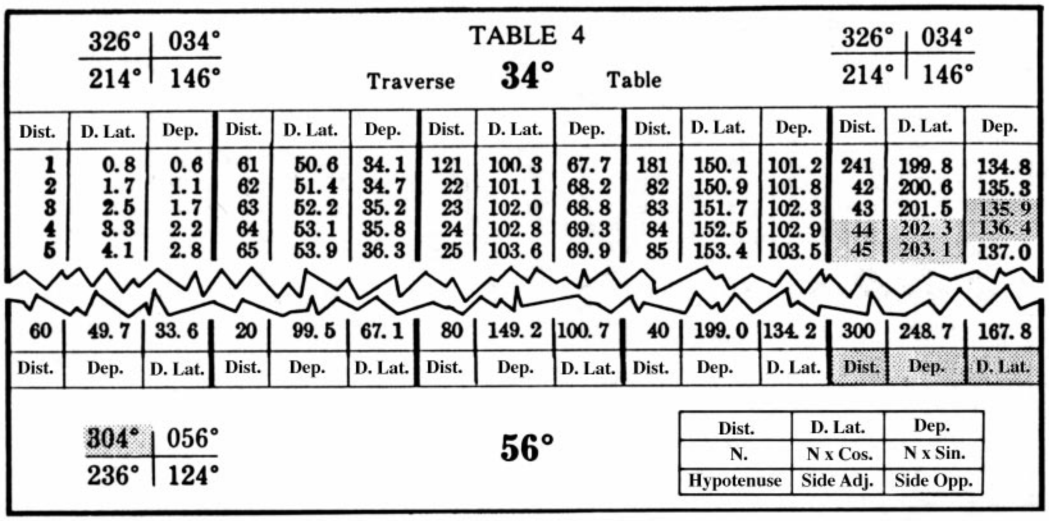

326° | 034°
214° | 146°

TABLE 4
Traverse 34° Table

326° | 034°
214° | 146°

Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.
1	0.8	0.6	61	50.6	34.1	121	100.3	67.7	181	150.1	101.2	241	199.8	134.8
2	1.7	1.1	62	51.4	34.7	22	101.1	68.2	82	150.9	101.8	42	200.6	135.3
3	2.5	1.7	63	52.2	35.2	23	102.0	68.8	83	151.7	102.3	43	201.5	135.9
4	3.3	2.2	64	53.1	35.8	24	102.8	69.3	84	152.5	102.9	44	202.3	136.4
5	4.1	2.8	65	53.9	36.3	25	103.6	69.9	85	153.4	103.5	45	203.1	137.0
60	49.7	33.6	20	99.5	67.1	80	149.2	100.7	40	199.0	134.2	300	248.7	167.8
Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.

304° | 056°
236° | 124°

56°

Dist.	D. Lat.	Dep.
N.	N x Cos.	N x Sin.
Hypotenuse	Side Adj.	Side Opp.

Figure 2413c. Extract from Table 4 .

Required: *Equivalent single (1) course (2) distance.*

Solution: *Solve each leg as a plane sailing and tabulate each solution as follows. For course 158°, extract the values for D. Lat. and Dep. opposite 155 in the Dist. column. Then, divide the values by 10 and round them off to the nearest tenth. Repeat the procedure for each leg of the vessel's journey.*

Course	*Dist.*	*N*	*S*	*E*	*W*
degrees	*mi.*	*mi.*	*mi.*	*mi.*	*mi.*
158	*15.5*		*14.4*	*5.8*	
135	*33.7*		*23.8*	*23.8*	
259	*16.1*		*3.1*		*15.8*
293	*39.0*	*15.2*			*35.9*
169	*40.4*		*39.7*	*7.7*	
Subtotals		*15.2*	*81.0* *-15.2*	*37.3*	*51.7* *-37.3*
N/S Total			*65.8 S*		*14.4 W*

Thus, the latitude difference is S 65.8 miles and the departure is W 14.4 miles. Convert this to a course and distance using the formulas discussed in section 2413.

Answer:

(1) C = 192.3°
(2) D = 67.3 miles.

2415. Parallel Sailing

Parallel sailing consists of the interconversion of departure and difference of longitude. It is the simplest form of spherical sailing. The formulas for these transformations are:

$$\text{DLo} = p \sec L \qquad p = \text{DLo} \cos L$$

Example 1: *The DR latitude of a ship on course 090° is 49°30' N. The ship steams on this course until the longitude changes 3°30'.*

Required: *The departure by (1) computation and (2) traverse table.*

Solution:

(1) Solution by computation:

$DLo = 3° 30'$
$DLo = 210$ arc min
$p = DLo \times \cos L$
$p = 210$ arc minutes $\times \cos (49.5°)$
$p = 136.4$ miles

Answer:

p = 136.4 miles

(2) Solution by traverse table:

Refer to Figure 2415a. Enter the traverse table with latitude as course angle and substitute DLo as the heading of the Dist. column and Dep. as the heading of the D. Lat. column. Since the table is computed for integral degrees of course angle (or latitude), the tabulations in the pages for 49° and 50° must be interpolated for the intermediate value (49°30'). The departure for latitude 49° and DLo 210' is 137.8 miles. The departure for latitude 50° and DLo 210' is 135.0 miles. Interpolating for the intermediate latitude, the departure is 136.4 miles.

Answer:

p = 136.4 miles

Example 2: *The DR latitude of a ship on course 270° is 38°15'S. The ship steams on this course for a distance of 215.5 miles.*

Required: *The change in longitude by (1) computation and (2) traverse table.*

Solution:

(1) Solution by computation

$DLo = 215.5$ arc min $\times \sec (38.25°)$
$DLo = 215.5$ arc min $\times 1.27$
$DLo = 274.4$ minutes of arc (west)
$DLo = 4° 34.4'$ W

Answer:

DLo = 4° 34.4' W

(2) Solution by traverse table

Refer to Figure 2415b. Enter the traverse tables with latitude as course angle and substitute DLo as the heading of the Dist. column and Dep. as the heading of the D. Lat. column. As the table is computed for integral degrees of course angle (or latitude), the tabulations in the pages for 38° and 39° must be interpolated for the minutes of latitude. Corresponding to Dep. 215.5 miles in the former is DLo

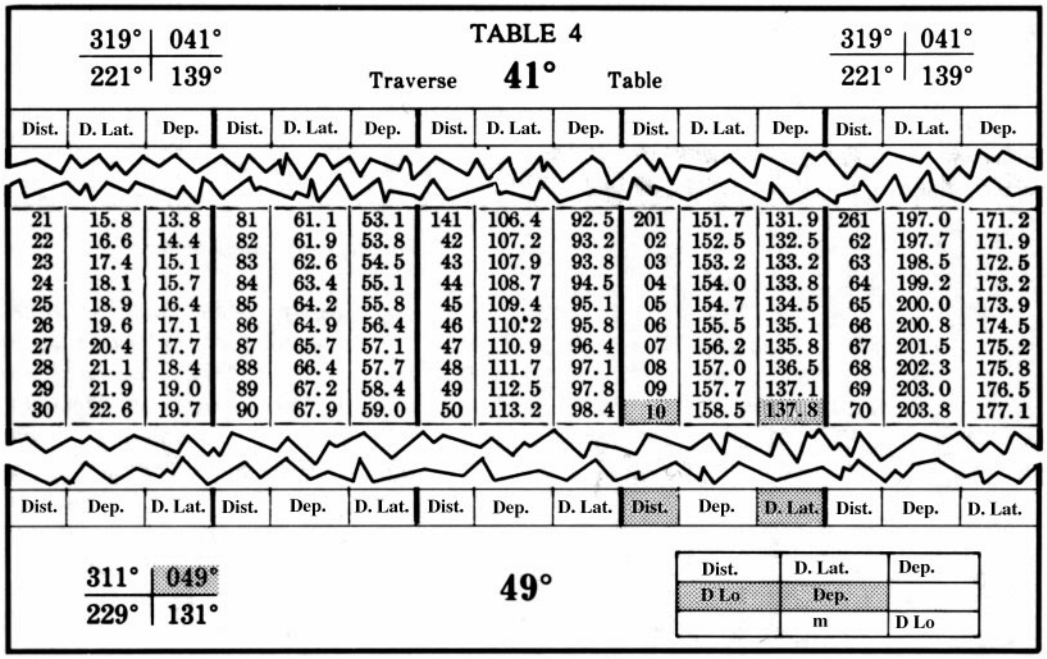

319° | 041°
221° | 139°

TABLE 4
Traverse 41° Table

319° | 041°
221° | 139°

Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.
21	15.8	13.8	81	61.1	53.1	141	106.4	92.5	201	151.7	131.9	261	197.0	171.2
22	16.6	14.4	82	61.9	53.8	42	107.2	93.2	02	152.5	132.5	62	197.7	171.9
23	17.4	15.1	83	62.6	54.5	43	107.9	93.8	03	153.2	133.2	63	198.5	172.5
24	18.1	15.7	84	63.4	55.1	44	108.7	94.5	04	154.0	133.8	64	199.2	173.2
25	18.9	16.4	85	64.2	55.8	45	109.4	95.1	05	154.7	134.5	65	200.0	173.9
26	19.6	17.1	86	64.9	56.4	46	110.2	95.8	06	155.5	135.1	66	200.8	174.5
27	20.4	17.7	87	65.7	57.1	47	110.9	96.4	07	156.2	135.8	67	201.5	175.2
28	21.1	18.4	88	66.4	57.7	48	111.7	97.1	08	157.0	136.5	68	202.3	175.8
29	21.9	19.0	89	67.2	58.4	49	112.5	97.8	09	157.7	137.1	69	203.0	176.5
30	22.6	19.7	90	67.9	59.0	50	113.2	98.4	10	158.5	137.8	70	203.8	177.1
Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.

311° | 049°
229° | 131°

49°

Dist.	D. Lat.	Dep.
D Lo	Dep.	
	m	D Lo

Figure 2415a. Extract fromTable 4.

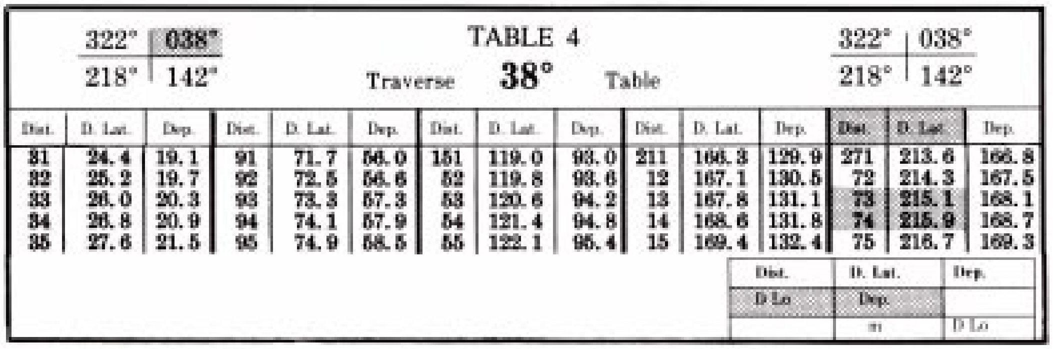

322° | 038°
218° | 142°

TABLE 4
Traverse 38° Table

322° | 038°
218° | 142°

Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.
31	24.4	19.1	91	71.7	56.0	151	119.0	93.0	211	166.3	129.9	271	213.6	166.8
32	25.2	19.7	92	72.5	56.6	52	119.8	93.6	12	167.1	130.5	72	214.3	167.5
33	26.0	20.3	93	73.3	57.3	53	120.6	94.2	13	167.8	131.1	73	215.1	168.1
34	26.8	20.9	94	74.1	57.9	54	121.4	94.8	14	168.6	131.8	74	215.9	168.7
35	27.6	21.5	95	74.9	58.5	55	122.1	95.4	15	169.4	132.4	75	216.7	169.3

Dist.	D. Lat.	Dep.
D Lo	Dep.	
	m	D Lo

Figure 2415b. Extract from Table 4.

273.5', and in the latter DLo 277.3'. Interpolating for minutes of latitude, the DLo is 274.4'W.

Answer:

DLo = 4° 34.4'

2416. Middle-Latitude Sailing

Middle-latitude sailing combines plane sailing and parallel sailing. Plane sailing is used to find difference of latitude and departure when course and distance are known, or vice versa. Parallel sailing is used to interconvert departure and difference of longitude. The mean latitude (L_m) is normally used for want of a practicable means of determining the middle latitude, the latitude at which the arc length of the parallel separating the meridians passing through two specific points is exactly equal to the departure in proceeding from one point to the other. The formulas for these transformations are:

$$DLo = p \sec L_m \qquad p = DLo \cos L_m.$$

The mean latitude (L_m) is half the arithmetical sum of the latitudes of two places on the same side of the equator. It is labeled N or S to indicate its position north or south of the equator. If a course line crosses the equator, solve each course line segment separately.

Example 1: *A vessel steams 1,253 miles on course 070°*

from lat. 15°17.0' N, long. 151°37.0' E.

Required: *Latitude and longitude of the point of arrival by (1) computation and (2) traverse table.*

Solution:

(1) Solution by computation:

$l = D \cos C$; $p = D \sin C$; and $DLo = p \sec L_m$.

D = 1253.0 miles.

C = 070°

l = 428.6' N

p = 1177.4 miles E
L_1 = 15°17.0' N
l = 7°08.6' N
L_2 = 22°25.6' N

L_m = 18°51.3' N

DLo = 1244.2' E
λ_1 = 151°37.0' E
DLo = 20°44.2' E
λ_2 = 172° 21.2' E

Answer:

L_2 = 22° 25.6' N
λ_2 = 172° 21.2' E

(2) Solution by traverse tables:

Refer to Figure 2416a. Enter the traverse table with course 070° and distance 1,253 miles. Because a number as high as 1,253 is not tabulated in the Dist. column, obtain the values for D. Lat. and Dep. for a distance of 125.3 miles and multiply them by 10. Interpolating between the tabular distance arguments yields D. Lat. = 429' and Dep. = 1,178 miles. Converting the D. Lat. value to degrees of latitude yields 7° 09.0'. The point of arrival's latitude, therefore, is 22° 26' N. This results in a mean latitude of 18° 51.5' N.

Reenter the table with the mean latitude as course angle and substitute DLo as the heading of the Dist. column and Dep. as the heading of the D. Lat. column. Since the table is computed for integral degrees of course angle (or latitude), the tabulations in the pages for 18° and 19° must be interpolated for the minutes of L_m. *In the 18° table, interpolate for DLo between the departure values of 117.0 miles and 117.9 miles. This results in a DLo value of 123.9. In the 19° table, interpolate for DLo between the departure values of 117.2 and 118.2. This yields a DLo value of 124.6.*

Having obtained the DLo values corresponding to mean latitudes of 18° and 19°, interpolate for the actual value of the mean latitude: 18° 51.5' N. This yields the value of DLo: 124.5. Multiply this final value by ten to obtain DLo = 1245 minutes = 20° 45' E.

Add the changes in latitude and longitude to the original position's latitude and longitude to obtain the final position.

Answer:

L_2 = 22° 26' N
λ_2 = 172° 22.0' E

Example 2: *A vessel at lat. 8°48.9'S, long. 89°53.3'W is to proceed to lat. 17°06.9'S, long. 104°51.6'W.*

Required: *Course and distance by (1) computation and (2) traverse table.*

Solution:

(1) Solution by computation:

$$p = DLo \cos L_m;\ \tan C = \frac{p}{l};\ \text{and}\ D = l \sec C$$

DLo = 14° 58.3'
DLo = 898.3'
L_m = 12° 57.9' S
p = 893.8 arc min × cos (12° 57.9')
p = 875.4 arc min

l = 17.1° - 8.8°
l = 8.3°
l = 498 arc min

$$C = \arctan\frac{875.4\,\text{arc min}}{498\,\text{arc min}}$$

C = S 60.4° W
C = 240.4°

D = 498 arc min × sec (60.4°)
D = 1008.2 miles

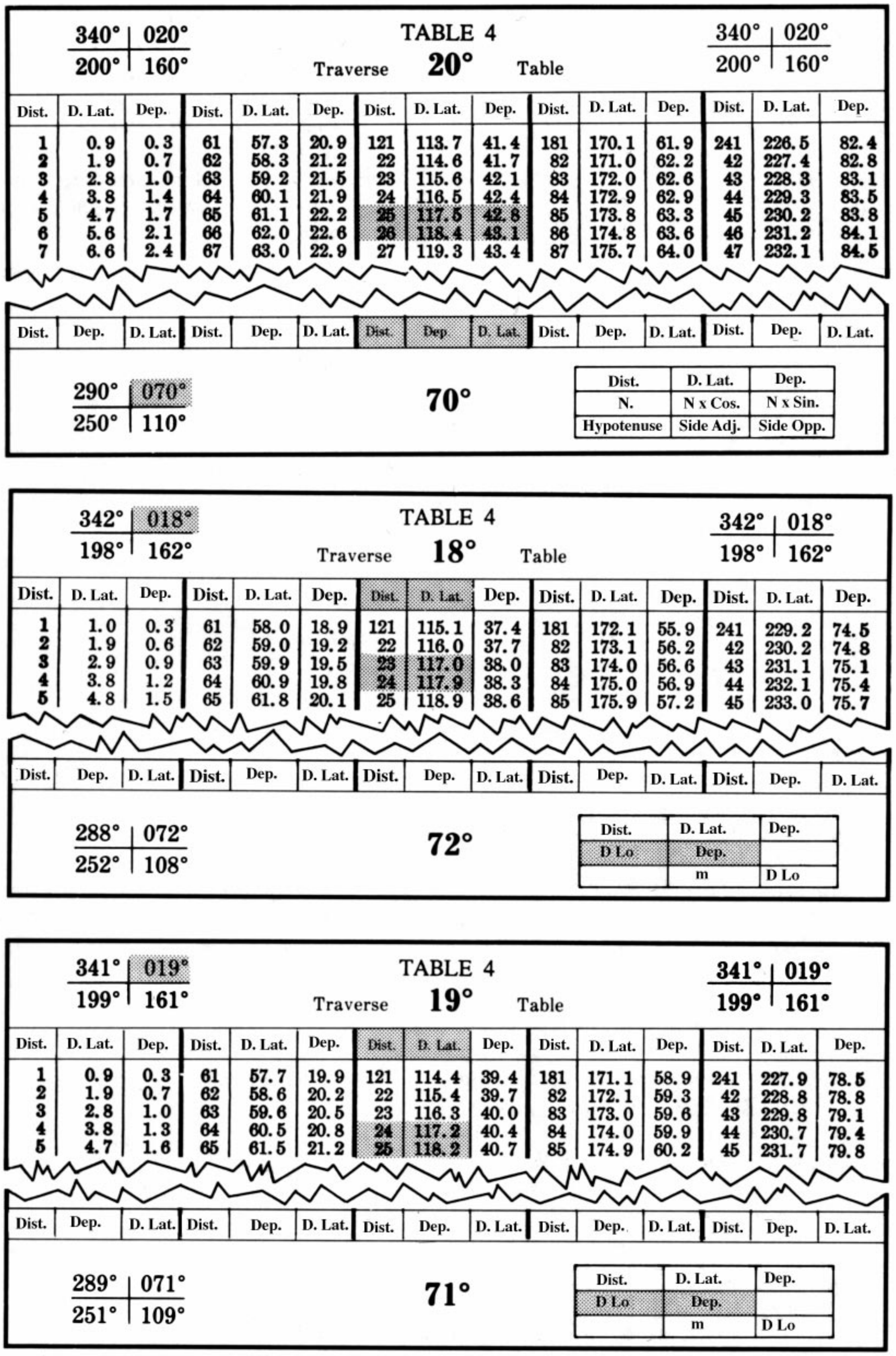

340° | 020°
200° | 160°

TABLE 4
Traverse **20°** Table

340° | 020°
200° | 160°

Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.
1	0.9	0.3	61	57.3	20.9	121	113.7	41.4	181	170.1	61.9	241	226.5	82.4
2	1.9	0.7	62	58.3	21.2	22	114.6	41.7	82	171.0	62.2	42	227.4	82.8
3	2.8	1.0	63	59.2	21.5	23	115.6	42.1	83	172.0	62.6	43	228.3	83.1
4	3.8	1.4	64	60.1	21.9	24	116.5	42.4	84	172.9	62.9	44	229.3	83.5
5	4.7	1.7	65	61.1	22.2	25	117.5	42.8	85	173.8	63.3	45	230.2	83.8
6	5.6	2.1	66	62.0	22.6	26	118.4	43.1	86	174.8	63.6	46	231.2	84.1
7	6.6	2.4	67	63.0	22.9	27	119.3	43.4	87	175.7	64.0	47	232.1	84.5
Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.

290° | 070°
250° | 110°

70°

Dist.	D. Lat.	Dep.
N.	N x Cos.	N x Sin.
Hypotenuse	Side Adj.	Side Opp.

342° | 018°
198° | 162°

TABLE 4
Traverse **18°** Table

342° | 018°
198° | 162°

Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.
1	1.0	0.3	61	58.0	18.9	121	115.1	37.4	181	172.1	55.9	241	229.2	74.5
2	1.9	0.6	62	59.0	19.2	22	116.0	37.7	82	173.1	56.2	42	230.2	74.8
3	2.9	0.9	63	59.9	19.5	23	117.0	38.0	83	174.0	56.6	43	231.1	75.1
4	3.8	1.2	64	60.9	19.8	24	117.9	38.3	84	175.0	56.9	44	232.1	75.4
5	4.8	1.5	65	61.8	20.1	25	118.9	38.6	85	175.9	57.2	45	233.0	75.7
Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.

288° | 072°
252° | 108°

72°

Dist.	D. Lat.	Dep.
D Lo	Dep.	
	m	D Lo

341° | 019°
199° | 161°

TABLE 4
Traverse **19°** Table

341° | 019°
199° | 161°

Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.
1	0.9	0.3	61	57.7	19.9	121	114.4	39.4	181	171.1	58.9	241	227.9	78.5
2	1.9	0.7	62	58.6	20.2	22	115.4	39.7	82	172.1	59.3	42	228.8	78.8
3	2.8	1.0	63	59.6	20.5	23	116.3	40.0	83	173.0	59.6	43	229.8	79.1
4	3.8	1.3	64	60.5	20.8	24	117.2	40.4	84	174.0	59.9	44	230.7	79.4
5	4.7	1.6	65	61.5	21.2	25	118.2	40.7	85	174.9	60.2	45	231.7	79.8
Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.

289° | 071°
251° | 109°

71°

Dist.	D. Lat.	Dep.
D Lo	Dep.	
	m	D Lo

Figure 2416a. Extracts from the Table 4.

Answer:

C = 240.4°
D = 1008.2 miles

The labels (N, S, E, W) of l, p, and C are determined by noting the direction of motion or the relative positions of the two places.

(2) Solution by traverse tables:

Refer to Figure 2416b. Enter the traverse table with the mean latitude as course angle and substitute DLo as the heading of the Dist. column and Dep. as the heading of the D. Lat. column. Since the table is computed for integral values of course angle (or latitude), it is usually necessary to extract the value of departure for values just less and just greater than the L_m *and then interpolate for the minutes of Lm. In this case where* L_m *is almost 13°, enter the table with* L_m *13° and DLo 898.3' to find Dep. 875 miles. The departure is found for DLo 89.9', and then multiplied by 10.*

Reenter the table to find the numbers 875 and 498 beside each other in the columns labeled Dep. and D. Lat., respectively. Because these high numbers are not tabulated, divide them by 10, and find 87.5 and 49.8. This occurs most nearly on the page for course angle 60° (fig. 2414c). Interpolating for intermediate values, the corresponding number in the Dist. column is about 100.5. Multiplying this by 10, the distance is about 1005 miles.

Answer:

C = 240°
D = 1005 miles.

The labels (N, S, E, W) of l, p, DLo, and C are determined by noting the direction of motion or the relative positions of the two places.

2417. Mercator Sailing

Mercator sailing problems can be solved graphically on a Mercator chart. For mathematical solution, the formulas of Mercator sailing are:

$$\tan C = \frac{\mathrm{DLo}}{m} \qquad \mathrm{DLo} = m \tan C.$$

After solving for course angle by Mercator sailing, solve for distance using the plane sailing formula:

$$D = l \sec C$$

Example 1: *A ship at lat. 32°14.7'N, long. 66°28.9'W is to head for a point near Chesapeake Light, lat. 36°58.7'N, long. 75°42.2'W.*

Required: *Course and distance by (1) computation and (2) traverse table.*

Solution:
(1) Solution by computation:

$$\tan C = \frac{DLo}{m}, \text{ and } D = l \sec C.$$

First calculate the meridional difference by entering Table 6 and interpolating for the meridional parts for the original and final latitudes. The meridional difference is the difference between these two val-

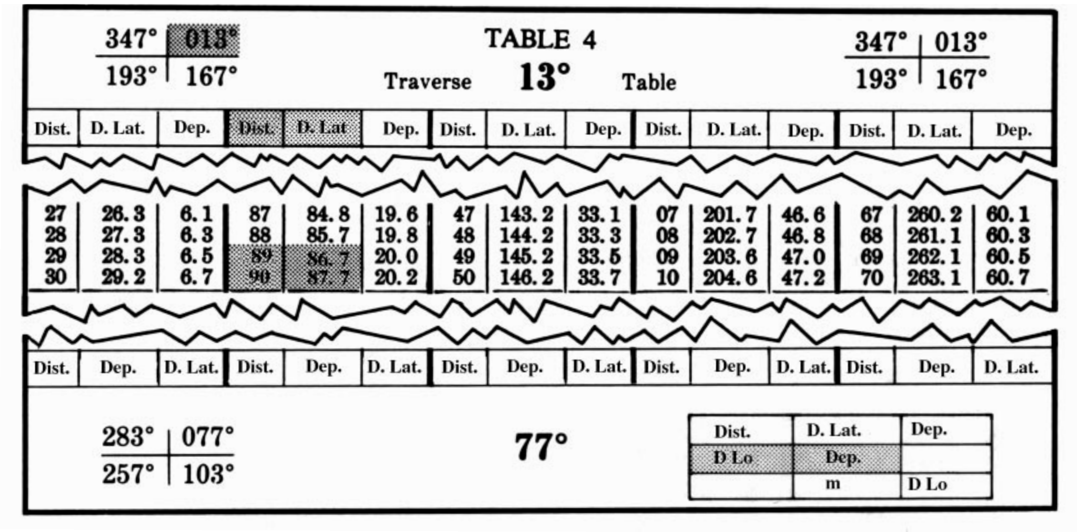

347° | 013°
193° | 167°

TABLE 4
Traverse 13° Table

347° | 013°
193° | 167°

Dist.	D. Lat.	Dep.	Dist.	D. Lat	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.
27	26.3	6.1	87	84.8	19.6	47	143.2	33.1	07	201.7	46.6	67	260.2	60.1
28	27.3	6.3	88	85.7	19.8	48	144.2	33.3	08	202.7	46.8	68	261.1	60.3
29	28.3	6.5	89	86.7	20.0	49	145.2	33.5	09	203.6	47.0	69	262.1	60.5
30	29.2	6.7	90	87.7	20.2	50	146.2	33.7	10	204.6	47.2	70	263.1	60.7
Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.

283° | 077°
257° | 103°

77°

Dist.	D. Lat.	Dep.
D Lo	Dep.	
	m	D Lo

Figure 2416b. Extract from Table 4.

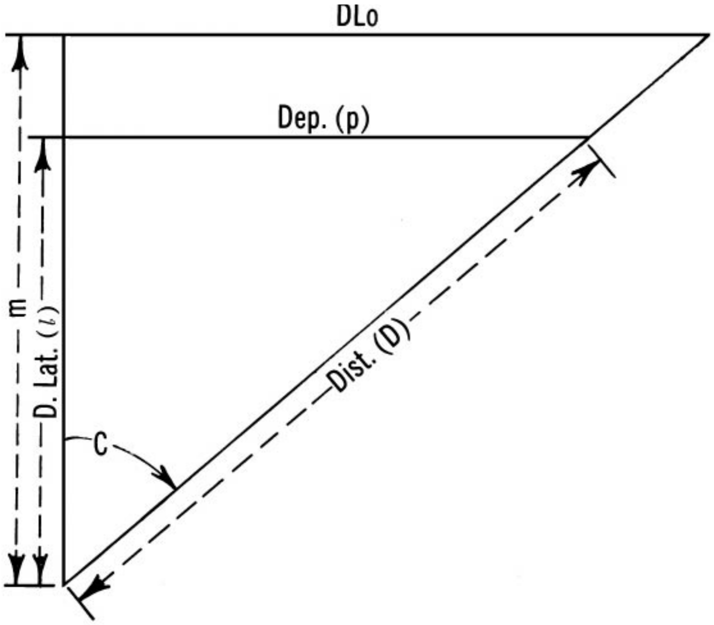

Figure 2417a. Mercator and plane sailing relationships.

ues. Having calculated the meridional difference, simply solve for course and distance from the equations above.

M_2 (36° 58.7' N) = 2377.5
M_1 (32° 14.7' N) = 2032.9
m = 344.6

λ_2 = 075° 42.2' W
λ_1 = 066° 28.9' W
DLo = 9° 13.3' W
DLo = 553.3' W

C = arctan (553.3÷344.6')
C = N 58.1° W
C = 301.9°

L_2 = 36° 58.7' N
L_1 = 32° 14.7' N
l = 4° 44.0' N
l = 284.0'

D = 284.0 arc min × sec (58.1°)
D = 537.4 miles

Answer:

C = 301.9°
D = 537.4 miles
(2) Solution by traverse table:

Refer to Figure 2417b. Substitute m as the heading of the D. Lat. column and DLo as the heading of the Dep. column. Inspect the table for the numbers 343.7 and 553.3 in the columns relabeled m and DLo, respectively.

Because a number as high as 343.7 is not tabulated in the m column, it is necessary to divide m and DLo by 10. Then inspect to find 34.4 and 55.3 abreast in the m and DLo columns, respectively. This occurs most nearly on the page for course angle 58° or course 302°.

Reenter the table with course 302° to find Dist. for D. Lat. 284.0'. This distance is 536 miles.

Answer:

C = 302°
D = 536 miles

Example 2: *A ship at lat. 75°31.7' N, long. 79°08.7'W, in Baffin Bay, steams 263.5 miles on course 155°.*

Required: *Latitude and longitude of point of arrival by (1) computation and (2) traverse table.*

Solution:

(1) Solution by computation:

$l = D \cos C$; and $DLo = m \tan C$

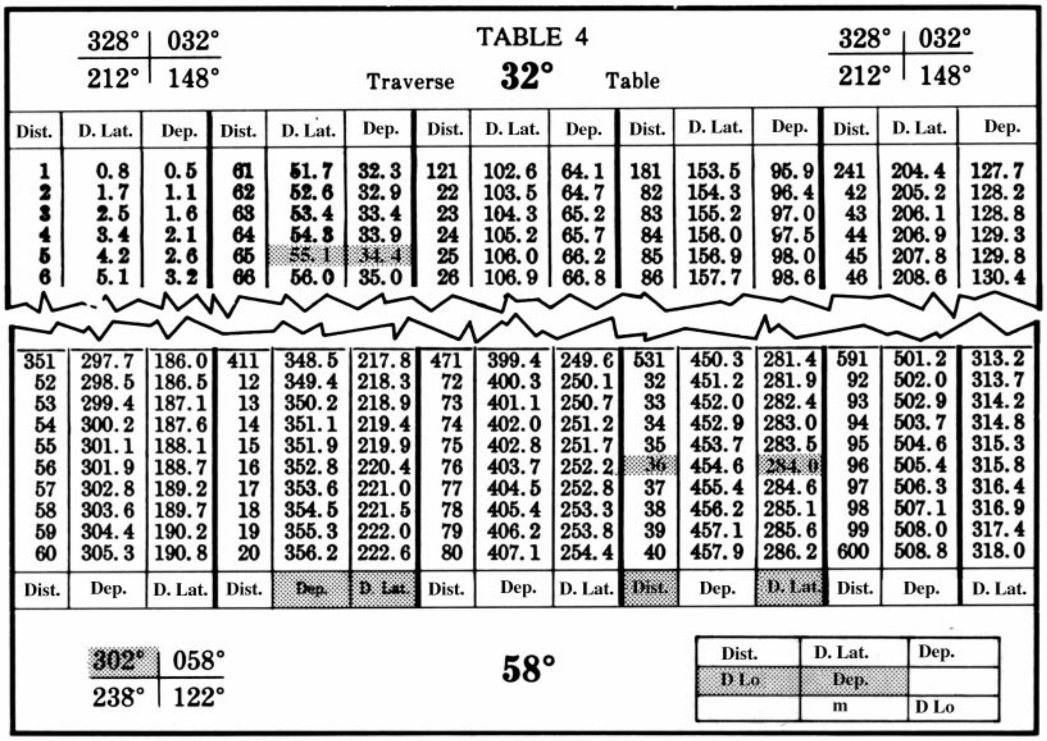

328° | 032°
212° | 148°

TABLE 4
Traverse 32° Table

328° | 032°
212° | 148°

Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.
1	0.8	0.5	61	51.7	32.3	121	102.6	64.1	181	153.5	95.9	241	204.4	127.7
2	1.7	1.1	62	52.6	32.9	22	103.5	64.7	82	154.3	96.4	42	205.2	128.2
3	2.5	1.6	63	53.4	33.4	23	104.3	65.2	83	155.2	97.0	43	206.1	128.8
4	3.4	2.1	64	54.3	33.9	24	105.2	65.7	84	156.0	97.5	44	206.9	129.3
5	4.2	2.6	65	55.1	34.4	25	106.0	66.2	85	156.9	98.0	45	207.8	129.8
6	5.1	3.2	66	56.0	35.0	26	106.9	66.8	86	157.7	98.6	46	208.6	130.4

351	297.7	186.0	411	348.5	217.8	471	399.4	249.6	531	450.3	281.4	591	501.2	313.2
52	298.5	186.5	12	349.4	218.3	72	400.3	250.1	32	451.2	281.9	92	502.0	313.7
53	299.4	187.1	13	350.2	218.9	73	401.1	250.7	33	452.0	282.4	93	502.9	314.2
54	300.2	187.6	14	351.1	219.4	74	402.0	251.2	34	452.9	283.0	94	503.7	314.8
55	301.1	188.1	15	351.9	219.9	75	402.8	251.7	35	453.7	283.5	95	504.6	315.3
56	301.9	188.7	16	352.8	220.4	76	403.7	252.2	36	454.6	284.0	96	505.4	315.8
57	302.8	189.2	17	353.6	221.0	77	404.5	252.8	37	455.4	284.6	97	506.3	316.4
58	303.6	189.7	18	354.5	221.5	78	405.4	253.3	38	456.2	285.1	98	507.1	316.9
59	304.4	190.2	19	355.3	222.0	79	406.2	253.8	39	457.1	285.6	99	508.0	317.4
60	305.3	190.8	20	356.2	222.6	80	407.1	254.4	40	457.9	286.2	600	508.8	318.0
Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.

302° | 058°
238° | 122°

58°

Dist.	D. Lat.	Dep.
D Lo	Dep.	
	m	D Lo

Figure 2417b. Extract from Table 4 composed of parts of left and right hand pages for course angle 58°.

D = 263.5 mi.
C = 155°
l = 238.8 ' S
l = 3° 58.8 ' S

L_1 = 75°31.7' N
l = 3° 58.8' S
L_2 = 71°32.9' N

M_1 = 7072.4
M_2 = 6226.1
m = 846.3

DLo = 394.6' E
DLo = 6°34.6' E

λ_1 = 79°08.7' W
DLo = 6°34.6' E
l_2 = 072°34.1' W

The labels (N, S, E, W) of l, DLo, and C are determined by noting the direction of motion or the relative positions of the two places.

Answer:

L_2 = 71° 32.9'

λ_2 = 072° 34.1'

(2) Solution by traverse table:

Refer to Figure 2417c . Enter the traverse table with course 155° and Dist. 263.5 miles to find D. Lat. 238.8'. The latitude of the point of arrival is found by subtracting the D. Lat. from the latitude of the point of departure. Determine the meridional difference by Table Table 4 (m = 846.3).

Reenter the table with course 155° to find the DLo corresponding to m = 846.3. Substitute meridional difference m as the heading of the D. Lat. column and DLo as the heading of the Dep. column. Because a number as high as 846.3 is not tabulated in the m column, divide m by 10 and then inspect the m column for a value of 84.6. Interpolating as necessary, the latter value is opposite DLo 39.4'. The DLo is 394' (39.4' × 10). The longitude of the point of arrival is found by applying the DLo to the longitude of the point of departure.

Answer:

L_2 = 71°32.9' N.
λ_2 = 72°34.7' W.

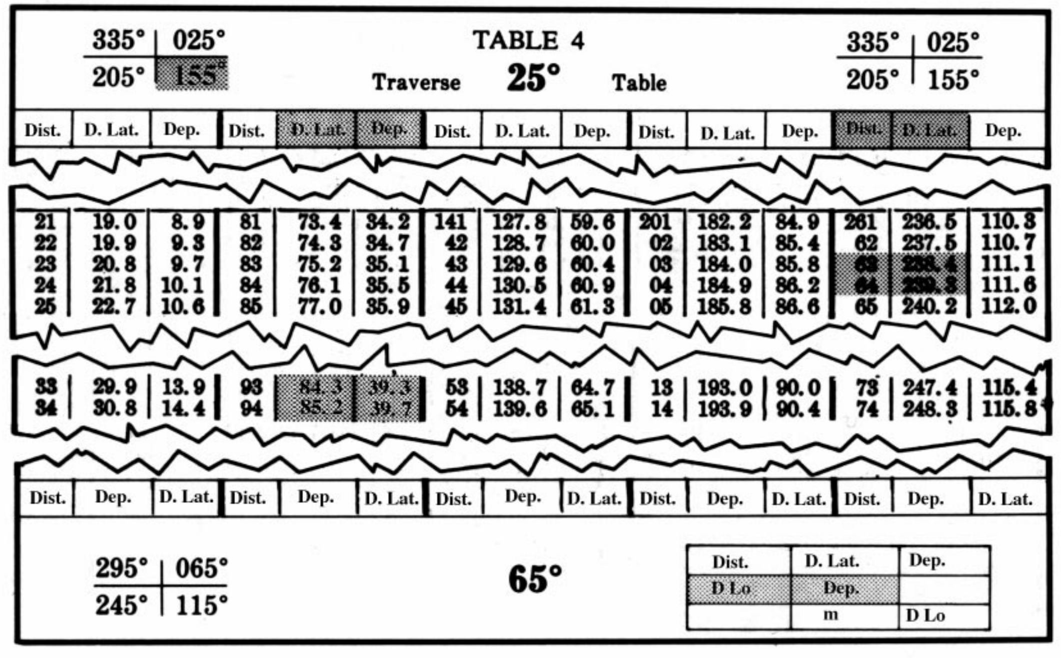

335° | 025°
205° | 155°

TABLE 4
Traverse 25° Table

335° | 025°
205° | 155°

Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.	Dist.	D. Lat.	Dep.
21	19.0	8.9	81	73.4	34.2	141	127.8	59.6	201	182.2	84.9	261	236.5	110.8
22	19.9	9.3	82	74.3	34.7	42	128.7	60.0	02	183.1	85.4	62	237.5	110.7
23	20.8	9.7	83	75.2	35.1	43	129.6	60.4	03	184.0	85.8	63	238.4	111.1
24	21.8	10.1	84	76.1	35.5	44	130.5	60.9	04	184.9	86.2	64	239.3	111.6
25	22.7	10.6	85	77.0	35.9	45	131.4	61.3	05	185.8	86.6	65	240.2	112.0
33	29.9	13.9	93	84.3	39.3	53	138.7	64.7	13	193.0	90.0	73	247.4	115.4
34	30.8	14.4	94	85.2	39.7	54	139.6	65.1	14	193.9	90.4	74	248.3	115.8
Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.	Dist.	Dep.	D. Lat.

295° | 065°
245° | 115°

65°

Dist.	D. Lat.	Dep.
D Lo	Dep.	
	m	D Lo

Figure 2417c. Extract from Table 4.

2418. Additional Problems

Example: *A vessel steams 117.3 miles on course 214°.*

Required: *(1) Difference of latitude, (2) departure, by plane sailing.*

Answers: *(1) l 97.2'S, (2) p 65.6 mi. W.*

Example: *A steamer is bound for a port 173.3 miles south and 98.6 miles east of the vessel's position*

Required: *(1) Course, (2) distance, by plane sailing.*

Answers: *(1) C 150.4°; (2) D 199.4 mi. by computation, 199.3 mi. by traverse table.*

Example: *A ship steams as follows: course 359°, distance 28.8 miles; course 006°, distance 16.4 miles; course 266°, distance 4.9 miles; course 144°, distance 3.1 miles; course 333°, distance 35.8 miles; course 280°, distance 19.3 miles.*

Required: *(1) Course, (2) distance, by traverse sailing.*

Answers: *(1) C 334.4°, (2) D 86.1 mi.*

Example: *The 1530 DR position of a ship is lat. 44°36.3'N, long. 31°18.3'W. The ship is on course 270°, speed 17 knots.*

Required: *The 2000 DR position, by parallel sailing.*

Answer: *2000 DR: L 44°36.3'N, λ 33°05.7'W.*

Example: *A ship at lat. 33°53.3'S, long. 18°23.1'E, leaving Cape Town, heads for a destination near Ambrose Light, lat. 40°27.1'N, long. 73°49.4'W.*

Required: *(1) Course and (2) distance, by Mercator sailing.*

Answers: *(1) C 310.9°; (2) D 6,811.5 mi. by computation, 6,812.8 mi. by traverse table.*

Example: *A ship at lat. 15°03.7'N, long. 151°26.8'E steams 57.4 miles on course 035°.*

Required: *(1) Latitude and (2) longitude of the point of arrival, by Mercator sailing.*

Answers: *(1) L 15°50.7'N; (2) λ 152°00.7'E.*

Teach Yourself
Electricity and Electronics

Fourth Edition

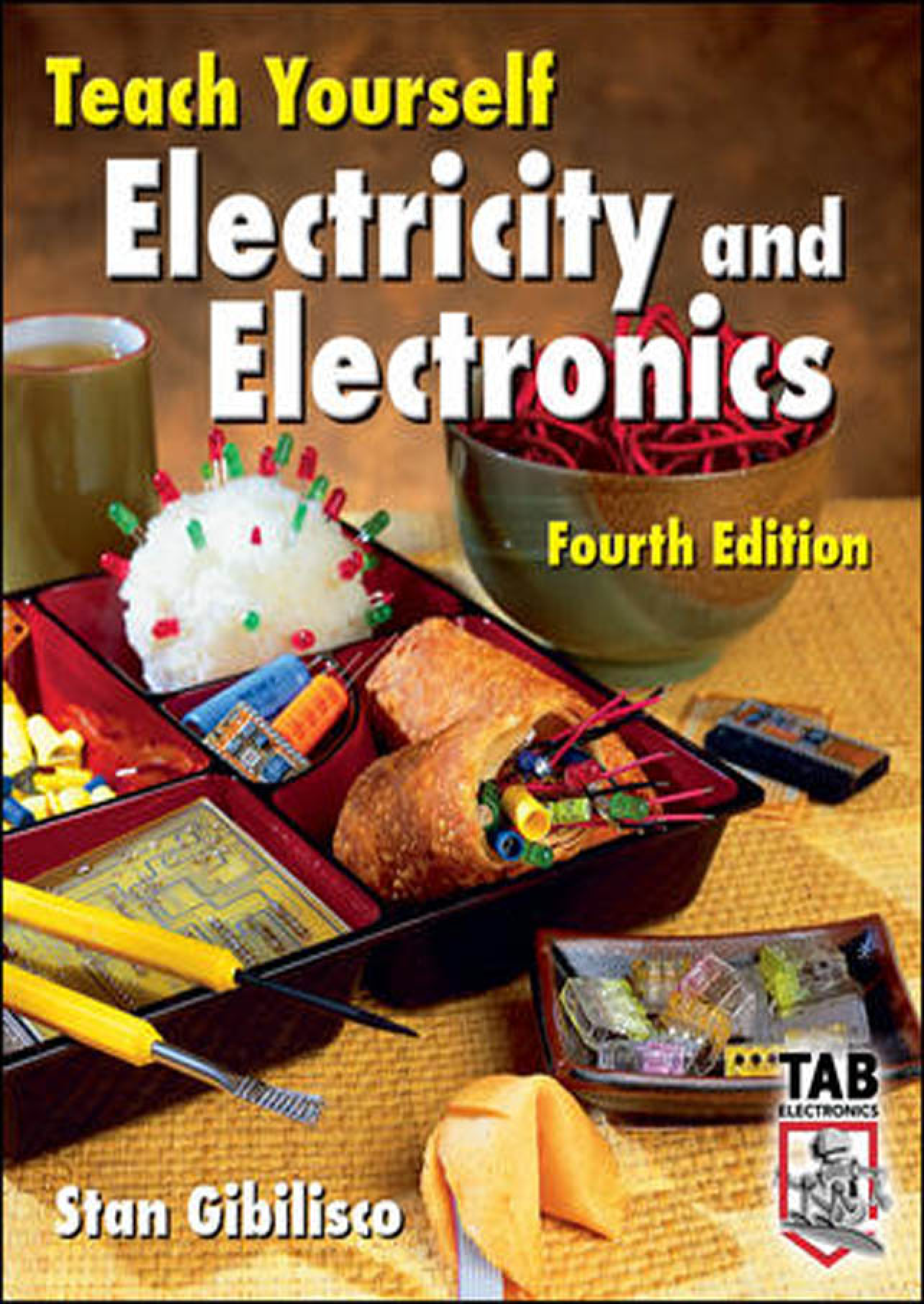

TAB ELECTRONICS

Stan Gibilisco

Teach Yourself Electricity and Electronics

This page intentionally left blank

Teach Yourself Electricity and Electronics

Fourth Edition

Stan Gibilisco

McGraw-Hill
New York Chicago San Francisco Lisbon London Madrid
Mexico City Milan New Delhi San Juan Seoul
Singapore Sydney Toronto

The McGraw-Hill Companies

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. All rights reserved. Manufactured in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

0-07-148709-3

The material in this eBook also appears in the print version of this title: 0-07-145933-2.

All trademarks are trademarks of their respective owners. Rather than put a trademark symbol after every occurrence of a trademarked name, we use names in an editorial fashion only, and to the benefit of the trademark owner, with no intention of infringement of the trademark. Where such designations appear in this book, they have been printed with initial caps.

McGraw-Hill eBooks are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. For more information, please contact George Hoare, Special Sales, at george_hoare@mcgraw-hill.com or (212) 904-4069.

TERMS OF USE

This is a copyrighted work and The McGraw-Hill Companies, Inc. ("McGraw-Hill") and its licensors reserve all rights in and to the work. Use of this work is subject to these terms. Except as permitted under the Copyright Act of 1976 and the right to store and retrieve one copy of the work, you may not decompile, disassemble, reverse engineer, reproduce, modify, create derivative works based upon, transmit, distribute, disseminate, sell, publish or sublicense the work or any part of it without McGraw-Hill's prior consent. You may use the work for your own noncommercial and personal use; any other use of the work is strictly prohibited. Your right to use the work may be terminated if you fail to comply with these terms.

THE WORK IS PROVIDED "AS IS." McGRAW-HILL AND ITS LICENSORS MAKE NO GUARANTEES OR WARRANTIES AS TO THE ACCURACY, ADEQUACY OR COMPLETENESS OF OR RESULTS TO BE OBTAINED FROM USING THE WORK, INCLUDING ANY INFORMATION THAT CAN BE ACCESSED THROUGH THE WORK VIA HYPERLINK OR OTHERWISE, AND EXPRESSLY DISCLAIM ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. McGraw-Hill and its licensors do not warrant or guarantee that the functions contained in the work will meet your requirements or that its operation will be uninterrupted or error free. Neither McGraw-Hill nor its licensors shall be liable to you or anyone else for any inaccuracy, error or omission, regardless of cause, in the work or for any damages resulting therefrom. McGraw-Hill has no responsibility for the content of any information accessed through the work. Under no circumstances shall McGraw-Hill and/or its licensors be liable for any indirect, incidental, special, punitive, consequential or similar damages that result from the use of or inability to use the work, even if any of them has been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause arises in contract, tort or otherwise.

DOI: 10.1036/0071459332

Professional

Want to learn more?

We hope you enjoy this McGraw-Hill eBook! If you'd like more information about this book, its author, or related books and websites, please click here.

To Tony, Samuel, Tim, Roland, Jack, and Sherri

This page intentionally left blank

For more information about this title, click here

Contents

Preface xvii

Part 1 Direct Current

1 Basic Physical Concepts 3

Atoms *3*
Protons, Neutrons, and Atomic Numbers *3*
Isotopes and Atomic Weights *4*
Electrons *4*
Ions *6*
Compounds *6*
Molecules *7*
Conductors *8*
Insulators *8*
Resistors *9*
Semiconductors *9*
Current *10*
Static Electricity *11*
Electromotive Force *12*
Nonelectrical Energy *13*
Quiz *14*

2 Electrical Units 17

The Volt *17*
Current Flow *18*
The Ampere *19*
Resistance and the Ohm *20*
Conductance and the Siemens *22*
Power and the Watt *23*

A Word about Notation *24*
Energy and the Watt-Hour *25*
Other Energy Units *26*
Alternating Current and the Hertz *27*
Rectification and Pulsating Direct Current *28*
Safety Considerations in Electrical Work *30*
Magnetism *30*
Magnetic Units *32*
Quiz *32*

3 Measuring Devices 36
Electromagnetic Deflection *36*
Electrostatic Deflection *38*
Thermal Heating *39*
Ammeters *39*
Voltmeters *41*
Ohmmeters *43*
Multimeters *44*
FET Voltmeters *44*
Wattmeters *45*
Watt-Hour Meters *46*
Digital Readout Meters *46*
Frequency Counters *47*
Other Meter Types *47*
Quiz *51*

4 Direct-Current Circuit Basics 55
Schematic Symbols *55*
Schematic and Wiring Diagrams *56*
Voltage/Current/Resistance Circuits *57*
Ohm's Law *58*
Current Calculations *59*
Voltage Calculations *60*
Resistance Calculations *60*
Power Calculations *61*
Resistances in Series *62*
Resistances in Parallel *63*
Division of Power *64*
Resistances in Series-Parallel *64*
Quiz *65*

5 Direct-Current Circuit Analysis 69
Current through Series Resistances *69*
Voltages across Series Resistances *70*
Voltage across Parallel Resistances *72*
Currents through Parallel Resistances *72*

Power Distribution in Series Circuits 74
Power Distribution in Parallel Circuits 74
Kirchhoff's First Law 75
Kirchhoff's Second Law 77
Voltage Divider Networks 78
Quiz 80

6 Resistors 85
Purpose of the Resistor 85
Fixed Resistors 88
The Potentiometer 90
The Decibel 93
Resistor Specifications 94
Quiz 98

7 Cells and Batteries 102
Electrochemical Energy 102
Grocery Store Cells and Batteries 105
Miniature Cells and Batteries 106
Lead-Acid Batteries 107
Nickel-Based Cells and Batteries 108
Photovoltaic Cells and Batteries 109
Fuel Cells 110
Quiz 111

8 Magnetism 115
The Geomagnetic Field 115
Causes and Effects 116
Magnetic Field Strength 120
Electromagnets 120
Magnetic Properties of Materials 122
Practical Magnetism 123
Quiz 128

Test: Part 1 132

Part 2 Alternating Current

9 Alternating-Current Basics 143
Definition of Alternating Current 143
Period and Frequency 143
The Sine Wave 144
Square Waves 145
Sawtooth Waves 146
Complex and Irregular Waveforms 147

Frequency Spectrum *148*
Fractions of a Cycle *150*
Expressions of Amplitude *151*
The Generator *154*
Why Alternating and Not Direct? *155*
Quiz *156*

10 Inductance 160
The Property of Inductance *160*
Practical Inductors *161*
The Unit of Inductance *162*
Inductors in Series *162*
Inductors in Parallel *163*
Interaction among Inductors *164*
Air-Core Coils *166*
Ferromagnetic Cores *166*
Inductors at RF *169*
Unwanted Inductances *171*
Quiz *171*

11 Capacitance 175
The Property of Capacitance *175*
Practical Capacitors *176*
The Unit of Capacitance *177*
Capacitors in Series *177*
Capacitors in Parallel *178*
Fixed Capacitors *179*
Variable Capacitors *182*
Capacitor Specifications *184*
Interelectrode Capacitance *184*
Quiz *185*

12 Phase 188
Instantaneous Values *188*
Instantaneous Rate of Change *189*
Circles and Vectors *190*
Expressions of Phase Difference *192*
Vector Diagrams of Phase Difference *195*
Quiz *196*

13 Inductive Reactance 200
Coils and Direct Current *200*
Coils and Alternating Current *201*
Reactance and Frequency *202*
Points in the *RL* Plane *203*
Vectors in the *RL* Plane *205*

Current Lags Voltage *205*
How Much Lag? *208*
Quiz *210*

14 Capacitive Reactance 214
Capacitors and Direct Current *214*
Capacitors and Alternating Current *215*
Capacitive Reactance and Frequency *216*
Points in the *RC* Plane *218*
Vectors in the *RC* Plane *219*
Current Leads Voltage *220*
How Much Lead? *222*
Quiz *225*

15 Impedance and Admittance 229
Imaginary Numbers *229*
Complex Numbers *229*
The *RX* Plane *233*
Characteristic Impedance *236*
Conductance *238*
Susceptance *238*
Admittance *240*
The *GB* plane *240*
Quiz *242*

16 *RLC* and *GLC* Circuit Analysis 245
Complex Impedances in Series *245*
Series *RLC* Circuits *248*
Complex Admittances in Parallel *250*
Parallel *GLC* Circuits *253*
Putting It All Together *256*
Reducing Complicated *RLC* Circuits *257*
Ohm's Law for AC Circuits *259*
Quiz *261*

17 Power and Resonance in Alternating-Current Circuits 265
Forms of Power *265*
True Power, VA Power, and Reactive Power *268*
Power Transmission *273*
Resonance *276*
Resonant Devices *280*
Quiz *282*

18 Transformers and Impedance Matching 286
Principle of the Transformer *286*
Geometries *289*

Power Transformers *292*
Isolation and Impedance Matching *294*
Radio-Frequency Transformers *297*
Quiz *299*

Test: Part 2 303

Part 3 Basic Electronics

19 Introduction to Semiconductors 315
The Semiconductor Revolution *315*
Semiconductor Materials *316*
Doping and Charge Carriers *317*
The P-N Junction *318*
Quiz *321*

20 How Diodes Are Used 325
Rectification *325*
Detection *326*
Frequency Multiplication *326*
Signal Mixing *327*
Switching *328*
Voltage Regulation *328*
Amplitude Limiting *329*
Frequency Control *330*
Oscillation and Amplification *330*
Energy Emission *331*
Photosensitive Diodes *332*
Quiz *333*

21 Power Supplies 337
Power Transformers *337*
Rectifier Diodes *338*
Half-Wave Circuit *338*
Full-Wave Center-Tap Circuit *340*
Full-Wave Bridge Circuit *340*
Voltage-Doubler Circuit *341*
Filtering *342*
Voltage Regulation *344*
Protection of Equipment *345*
Quiz *348*

22 The Bipolar Transistor 352
NPN versus PNP *352*
Biasing *353*

Biasing for Amplification *355*
Gain versus Frequency *357*
Common Emitter Circuit *358*
Common Base Circuit *359*
Common Collector Circuit *359*
Quiz *361*

23 The Field Effect Transistor 365
Principle of the JFET *365*
Amplification *369*
The MOSFET *370*
Common Source Circuit *373*
Common Gate Circuit *374*
Common Drain Circuit *375*
Quiz *376*

24 Amplifiers and Oscillators 379
The Decibel *379*
Basic Bipolar Transistor Amplifier *381*
Basic JFET Amplifier *382*
Amplifier Classes *383*
Efficiency in Power Amplifiers *386*
Drive and Overdrive *388*
Audio Amplification *389*
Radio-Frequency Amplification *391*
How Oscillators Work *393*
Common Oscillator Circuits *394*
Oscillator Stability *399*
Audio Oscillators *400*
Quiz *402*

25 Wireless Transmitters and Receivers 407
Oscillation and Amplification *407*
Modulation *407*
Pulse Modulation *414*
Analog-to-Digital Conversion *415*
Image Transmission *416*
The Electromagnetic Field *419*
Wave Propagation *421*
Transmission Media *423*
Two Basic Receiver Designs *424*
Predetector Stages *427*
Detectors *429*
Audio Stages *433*
Television Reception *433*

Specialized Wireless Modes *434*
Quiz *437*

26 Digital Basics 441
Numbering Systems *441*
Logic *443*
Digital Circuits *444*
Binary Digital Communications *448*
The RGB Color Model *453*
Quiz *454*

Test: Part 3 458

Part 4 Specialized Devices and Systems

27 Antennas 471
Radiation Resistance *471*
Half-Wave Antennas *473*
Quarter-Wave Verticals *474*
Loops *476*
Ground Systems *477*
Gain and Directivity *478*
Phased Arrays *480*
Parasitic Arrays *482*
Antennas for Ultrahigh and Microwave Frequencies *483*
Safety *486*
Quiz *486*

28 Integrated Circuits 491
Advantages of IC Technology *491*
Limitations of IC Technology *492*
Linear ICs *492*
Digital ICs *496*
Component Density *498*
IC Memory *499*
Quiz *499*

29 Electron Tubes 504
Tube Forms *504*
Electrodes in a Tube *505*
Circuit Configurations *508*
Cathode-Ray Tubes *509*
Camera Tubes *511*
Tubes for Use above 300 MHz *513*
Quiz *514*

30 Transducers, Sensors, Location, and Navigation 517
Wave Transducers *517*
Displacement Transducers *519*
Detection and Measurement *523*
Location Systems *527*
Navigational Methods *531*
Quiz *534*

31 Acoustics, Audio, and High Fidelity 538
Acoustics *538*
Loudness and Phase *540*
Technical Considerations *541*
Components *541*
Specialized Systems *546*
Recorded Media *547*
Electromagnetic Interference *549*
Quiz *550*

32 Personal and Hobby Wireless 554
Cellular Communications *554*
Satellites and Networks *556*
Amateur and Shortwave Radio *559*
Security and Privacy *561*
Quiz *566*

33 A Computer and Internet Primer 569
The Central Processing Unit *569*
Units of Digital Data *570*
The Hard Drive *571*
External Storage *573*
Memory *574*
The Display *575*
The Printer *577*
The Scanner *578*
The Modem *580*
The Internet *581*
Quiz *584*

34 Monitoring, Robotics, and Artificial Intelligence 587
Keeping Watch *587*
Robot Generations and Laws *591*
Robot Arms *592*
Robot Hearing and Vision *597*
Robot Navigation *600*

Telepresence *603*
The Mind of the Machine *605*
Quiz *606*

Test: Part 4 610

Final Exam 621

Appendix A Answers to Quiz, Test, and Exam Questions 645

Appendix B Schematic Symbols 653

Suggested Additional Reading 671

Index 673

Preface

This book is for people who want to learn the fundamentals of electricity, electronics, and related fields without taking a formal course. The book can also serve as a classroom text. This edition contains new material on transducers, sensors, antennas, monitoring, security, and navigation. Material from previous editions has been updated where appropriate.

As you take this course, you'll encounter hundreds of quiz, test, and exam questions that can help you measure your progress. They are written like the questions found in standardized tests used by educational institutions.

There is a short multiple-choice quiz at the end of every chapter. The quizzes are "open-book." You may refer to the chapter texts when taking them. When you have finished a chapter, take the quiz, write down your answers, and then give your list of answers to a friend. Have the friend tell you your score, but not which questions you got wrong. Because you're allowed to look at the text when taking the quizzes, some of the questions are rather difficult.

At the end of each section, there is a multiple-choice test. These tests are easier than chapter-ending quizzes. Don't look back at the text when taking the tests. A satisfactory score is at least three-quarters of the answers correct.

You will find a final exam at the end of this course. As with the section-ending tests, the questions are not as difficult as those in the chapter-ending quizzes. Don't refer back to the text while taking the final exam. A satisfactory score is at least three-quarters of the answers correct.

The answers to all of the multiple-choice quiz, test, and exam questions are listed in an appendix at the back of this book.

You don't need a mathematical or scientific background for this course. Middle-school algebra, geometry, and physics will suffice. There's no calculus here! I recommend that you complete one chapter a week. That way, in a few months, you'll finish the course. You can then use this book, with its comprehensive index, as a permanent reference.

Suggestions for future editions are welcome.

Stan Gibilisco

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

This page intentionally left blank

1

PART

Direct Current

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

This page intentionally left blank

1

CHAPTER

Basic Physical Concepts

IT IS IMPORTANT TO UNDERSTAND SOME SIMPLE, GENERAL PHYSICS PRINCIPLES IN ORDER TO HAVE A full grasp of electricity and electronics. It is not necessary to know high-level mathematics. In science, you can talk about *qualitative* things or *quantitative* things, the "what" versus the "how much." For now, we are concerned only about the "what." The "how much" will come later.

Atoms

All matter is made up of countless tiny particles whizzing around. These particles are extremely dense; matter is mostly empty space. Matter seems continuous because the particles are so small, and they move incredibly fast.

Each chemical *element* has its own unique type of particle, known as its *atom.* Atoms of different elements are always different. The slightest change in an atom can make a tremendous difference in its behavior. You can live by breathing pure oxygen, but you can't live off of pure nitrogen. Oxygen will cause metal to corrode, but nitrogen will not. Wood will burn furiously in an atmosphere of pure oxygen, but will not even ignite in pure nitrogen. Yet both are gases at room temperature and pressure; both are colorless, both are odorless, and both are just about of equal weight. These substances are so different because oxygen has eight *protons*, while nitrogen has only seven. There are many other examples in nature where a tiny change in atomic structure makes a major difference in the way a substance behaves.

Protons, Neutrons, and Atomic Numbers

The part of an atom that gives an element its identity is the *nucleus.* It is made up of two kinds of particles, the *proton* and the *neutron.* These are extremely dense. A teaspoonful of either of these particles, packed tightly together, would weigh tons. Protons and neutrons have just about the same mass, but the proton has an electric charge while the neutron does not.

The simplest element, hydrogen, has a nucleus made up of only one proton; there are usually no neutrons. This is the most common element in the universe. Sometimes a nucleus of hydrogen

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

has a neutron or two along with the proton, but this does not occur very often. These "mutant" forms of hydrogen do, nonetheless, play significant roles in atomic physics.

The second most abundant element is helium. Usually, this atom has a nucleus with two protons and two neutrons. Hydrogen is changed into helium inside the sun, and in the process, energy is given off. This makes the sun shine. The process, called *fusion*, is also responsible for the terrific explosive force of a hydrogen bomb.

Every proton in the universe is just like every other. Neutrons are all alike, too. The number of protons in an element's nucleus, the *atomic number*, gives that element its identity. The element with three protons is lithium, a light metal that reacts easily with gases such as oxygen or chlorine. The element with four protons is beryllium, also a metal. In general, as the number of protons in an element's nucleus increases, the number of neutrons also increases. Elements with high atomic numbers, like lead, are therefore much denser than elements with low atomic numbers, like carbon. Perhaps you've compared a lead sinker with a piece of coal of similar size, and noticed this difference.

Isotopes and Atomic Weights

For a given element, such as oxygen, the number of neutrons can vary. But no matter what the number of neutrons, the element keeps its identity, based on the atomic number. Differing numbers of neutrons result in various *isotopes* for a given element.

Each element has one particular isotope that is most often found in nature. But all elements have numerous isotopes. Changing the number of neutrons in an element's nucleus results in a difference in the weight, and also a difference in the density, of the element. Thus, hydrogen containing a neutron or two in the nucleus, along with the proton, is called *heavy hydrogen.*

The *atomic weight* of an element is approximately equal to the sum of the number of protons and the number of neutrons in the nucleus. Common carbon has an atomic weight of about 12, and is called carbon 12 or C12. But sometimes it has an atomic weight of about 14, and is known as carbon 14 or C14.

Electrons

Surrounding the nucleus of an atom are particles having opposite electric charge from the protons. These are the *electrons.* Physicists arbitrarily call the electrons' charge *negative*, and the protons' charge *positive.* An electron has exactly the same charge quantity as a proton, but with opposite polarity. The charge on a single electron or proton is the smallest possible electric charge. All charges, no matter how great, are multiples of this unit charge.

One of the earliest ideas about the atom pictured the electrons embedded in the nucleus, like raisins in a cake. Later, the electrons were seen as orbiting the nucleus, making the atom like a miniature solar system with the electrons as the planets (Fig. 1-1). Still later, this view was modified further. Today, the electrons are seen as so fast-moving, with patterns so complex, that it is not even possible to pinpoint them at any given instant of time. All that can be done is to say that an electron will just as likely be inside a certain sphere as outside. These spheres are known as electron *shells.* Their centers correspond to the position of the atomic nucleus. The farther away from the nucleus the *shell*, the more energy the electron has (Fig. 1-2).

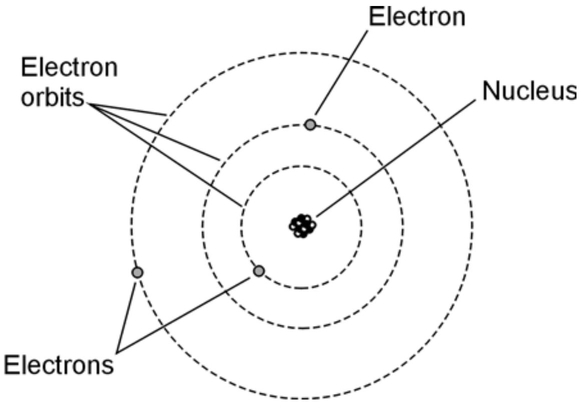
Electron
Electron orbits
Nucleus
Electrons

1-1 An early model of the atom, developed around the year 1900, resembled a miniature solar system. The electrons were held in their orbits around the nucleus by electrostatic attraction.

Electrons can move rather easily from one atom to another in some materials. In other substances, it is difficult to get electrons to move. But in any case, it is far easier to move electrons than it is to move protons. Electricity almost always results, in some way, from the motion of electrons in a material. Electrons are much lighter than protons or neutrons. In fact, compared to the nucleus of an atom, the electrons weigh practically nothing.

Generally, the number of electrons in an atom is the same as the number of protons. The negative charges therefore exactly cancel out the positive ones, and the atom is electrically neutral. But

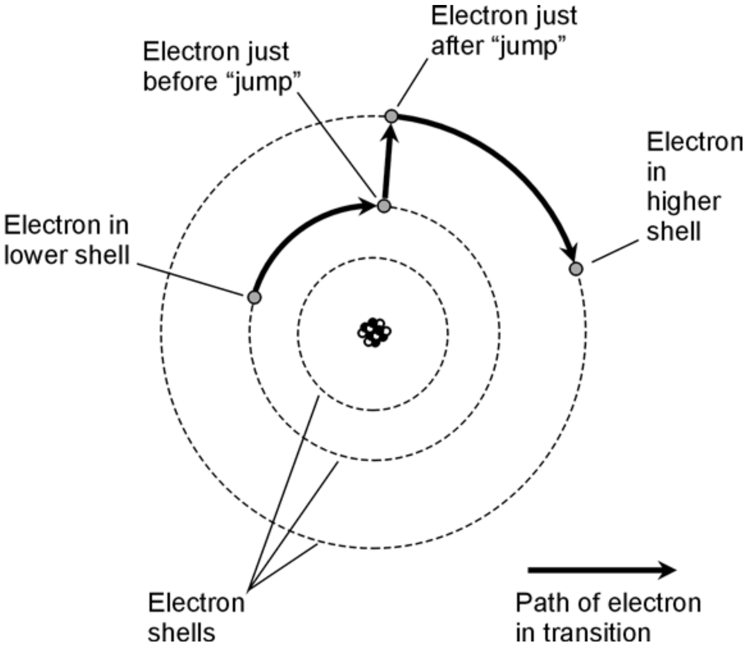
Electron just after "jump"
Electron just before "jump"
Electron in higher shell
Electron in lower shell
Electron shells
Path of electron in transition

1-2 Electrons move around the nucleus of an atom at defined levels, called *shells*, which correspond to discrete energy states. This is a simplified illustration of an electron gaining energy within an atom.

under some conditions, there can be an excess or shortage of electrons. High levels of radiant energy, extreme heat, or the presence of an electric field (discussed later) can "knock" or "throw" electrons loose from atoms, upsetting the balance.

Ions

If an atom has more or less electrons than protons, that atom acquires an electrical charge. A shortage of electrons results in positive charge; an excess of electrons gives a negative charge. The element's identity remains the same, no matter how great the excess or shortage of electrons. In the extreme case, all the electrons might be removed from an atom, leaving only the nucleus. However, it would still represent the same element as it would if it had all its electrons. A charged atom is called an *ion.* When a substance contains many ions, the material is said to be *ionized.*

A good example of an ionized substance is the atmosphere of the earth at high altitudes. The ultraviolet radiation from the sun, as well as high-speed subatomic particles from space, result in the gases' atoms being stripped of electrons. The ionized gases tend to be found in layers at certain altitudes. These layers are responsible for long-distance radio communications at some frequencies.

Ionized materials generally conduct electricity well, even if the substance is normally not a good conductor. Ionized air makes it possible for a lightning stroke to take place, for example. The ionization, caused by a powerful electric field, occurs along a jagged, narrow channel. After the lightning flash, the nuclei of the atoms quickly attract stray electrons back, and the air becomes electrically neutral again.

An element might be both an ion and an isotope different from the usual isotope. For example, an atom of carbon might have eight neutrons rather than the usual six, thus being the isotope C14, and it might have been stripped of an electron, giving it a positive unit electric charge and making it an ion.

Compounds

Different elements can join together to share electrons. When this happens, the result is a chemical *compound.* One of the most common compounds is water, the result of two hydrogen atoms joining with an atom of oxygen. There are literally thousands of different chemical compounds that occur in nature.

A compound is different than a simple mixture of elements. If hydrogen and oxygen are mixed, the result is a colorless, odorless gas, just like either element is a gas separately. A spark, however, will cause the molecules to join together; this will liberate energy in the form of light and heat. Under the right conditions, there will be a violent explosion, because the two elements join eagerly. Water is chemically illustrated in Fig. 1-3.

Compounds often, but not always, appear greatly different from any of the elements that make them up. At room temperature and pressure, both hydrogen and oxygen are gases. But water under the same conditions is a liquid. If it gets a few tens of degrees colder, water turns solid at standard pressure. If it gets hot enough, water becomes a gas, odorless and colorless, just like hydrogen or oxygen.

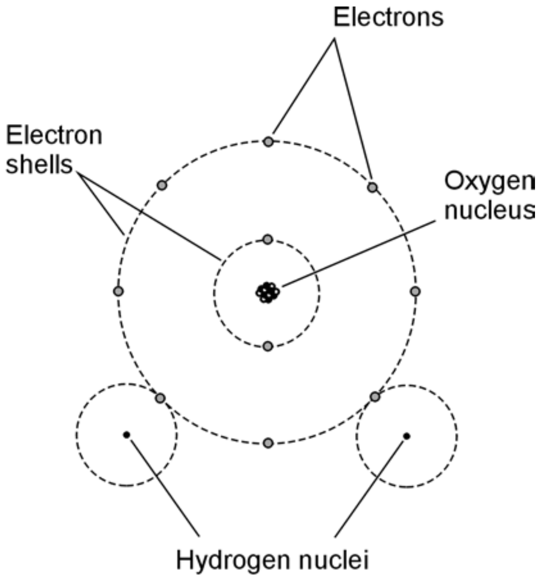

1-3 A simplified diagram of a water molecule. Note the shared electrons.

Another common example of a compound is rust. This forms when iron joins with oxygen. While iron is a dull gray solid and oxygen is a gas, rust is a maroon-red or brownish powder, completely unlike either of the elements from which it is formed.

Molecules

When atoms of elements join together to form a compound, the resulting particles are *molecules.* Figure 1-3 is an example of a molecule of water, consisting of three atoms put together.

The natural form of an element is also known as its molecule. Oxygen tends to occur in pairs most of the time in the earth's atmosphere. Thus, an oxygen molecule is sometimes denoted by the symbol O_2. The "O" represents oxygen, and the subscript 2 indicates that there are two atoms per molecule. The water molecule is symbolized H_2O, because there are two atoms of hydrogen and one atom of oxygen in each molecule.

Sometimes oxygen atoms exist all by themselves; then we denote the molecule simply as O. Sometimes there are three atoms of oxygen grouped together. This is the gas called *ozone*, which has received much attention lately in environmental news. It is written O_3.

All matter, whether solid, liquid, or gas, is made of molecules. These particles are always moving. The speed with which they move depends on the temperature. The hotter the temperature, the more rapidly the molecules move around. In a solid, the molecules are interlocked in a sort of rigid pattern, although they vibrate continuously (Fig. 1-4A). In a liquid, they slither and slide around (Fig. 1-4B). In a gas, they rush all over the place, bumping into each other and into solids and liquids adjacent to the gas (Fig. 1-4C).

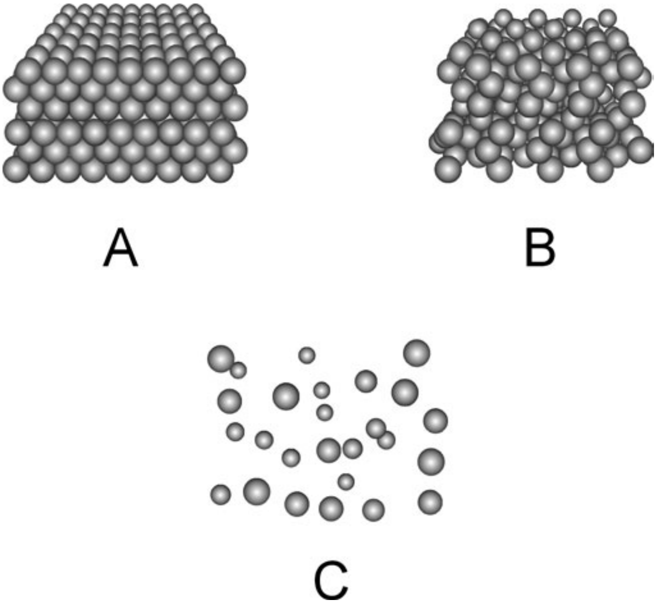

1-4 Simplified renditions of molecular arrangements in a solid (A), a liquid (B), and a gas (C).

Conductors

In some materials, electrons move easily from atom to atom. In others, the electrons move with difficulty. And in some materials, it is almost impossible to get them to move. An electrical *conductor* is a substance in which the electrons are mobile.

The best conductor at room temperature is pure elemental silver. Copper and aluminum are also excellent electrical conductors. Iron, steel, and various other metals are fair to good conductors of electricity. In most electrical circuits and systems, copper or aluminum wire is used. (Silver is impractical because of its high cost.)

Some liquids are good electrical conductors. Mercury is one example. Salt water is a fair conductor. Gases or mixtures of gases, such as air, are generally poor conductors of electricity. This is because the atoms or molecules are usually too far apart to allow a free exchange of electrons. But if a gas becomes ionized, it can be a fair conductor of electricity.

Electrons in a conductor do not move in a steady stream, like molecules of water through a garden hose. Instead, they are passed from one atom to another right next to it (Fig. 1-5). This happens to countless atoms all the time. As a result, literally trillions of electrons pass a given point each second in a typical electrical circuit.

Insulators

An *insulator* prevents electrical currents from flowing, except occasionally in tiny amounts. Most gases are good electrical insulators. Glass, dry wood, paper, and plastics are other examples. Pure water is a

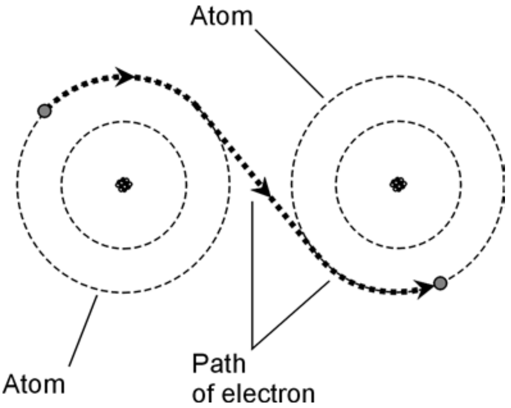

1-5 In an electrical conductor, certain electrons can pass easily from atom to atom.

good electrical insulator, although it conducts some current with even the slightest impurity. Metal oxides can be good insulators, even though the metal in pure form is a good conductor.

Electrical insulators can be forced to carry current. Ionization can take place; when electrons are stripped away from their atoms, they move more or less freely. Sometimes an insulating material gets charred, or melts down, or gets perforated by a spark. Then its insulating properties are lost, and some electrons flow. An insulating material is sometimes called a *dielectric.* This term arises from the fact that it keeps electrical charges apart, preventing the flow of electrons that would equalize a charge difference between two places. Excellent insulating materials can be used to advantage in certain electrical components such as capacitors, where it is important that electrons not flow.

Porcelain or glass can be used in electrical systems to keep short circuits from occurring. These devices, called insulators, come in various shapes and sizes for different applications. You can see them on high-voltage utility poles and towers. They hold the wire up without running the risk of a short circuit with the tower or a slow discharge through a wet wooden pole.

Resistors

Some substances, such as carbon, conduct electricity fairly well but not really well. The conductivity can be changed by adding impurities like clay to a carbon paste, or by winding a thin wire into a coil. Electrical components made in this way are called *resistors.* They are important in electronic circuits because they allow for the control of current flow. The better a resistor conducts, the lower its *resistance;* the worse it conducts, the higher the resistance.

Electrical resistance is measured in units called *ohms.* The higher the value in ohms, the greater the resistance, and the more difficult it becomes for current to flow. For wires, the resistance is sometimes specified in terms of ohms per unit length (foot, meter, kilometer, or mile). In an electrical system, it is usually desirable to have as low a resistance, or ohmic value, as possible. This is because resistance converts electrical energy into heat.

Semiconductors

In a *semiconductor*, electrons flow, but not as well as they do in a conductor. Some semiconductors carry electrons almost as well as good electrical conductors like copper or aluminum; others are almost as bad as insulating materials.

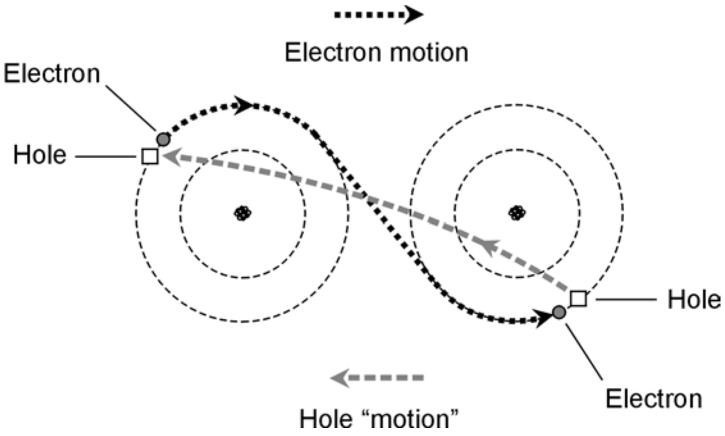

1-6 In a semiconducting material, holes travel in a direction opposite to the direction in which the electrons travel.

Semiconductors are not the same as resistors. In a semiconductor, the material is treated so that it has very special properties.

Semiconductors include certain substances such as silicon, selenium, or gallium, that have been "doped" by the addition of impurities such as indium or antimony. Have you heard of such things as *gallium arsenide*, *metal oxides*, or *silicon rectifiers?* Electrical conduction in these materials is always a result of the motion of electrons. But this can be a quite peculiar movement, and sometimes engineers speak of the movement of *holes* rather than electrons. A hole is a shortage of an electron—you might think of it as a positive ion—and it moves along in a direction opposite to the flow of electrons (Fig. 1-6).

When most of the *charge carriers* are electrons, the semiconductor is called *N-type*, because electrons are negatively charged. When most of the charge carriers are holes, the semiconductor material is known as *P-type* because holes have a positive electric charge. But P-type material does pass some electrons, and N-type material carries some holes. In a semiconductor, the more abundant type of charge carrier is called the *majority carrier.* The less abundant kind is known as the *minority carrier.* Semiconductors are used in *diodes*, *transistors*, and *integrated circuits.* These substances are what make it possible for you to have a computer or a television receiver in a package small enough to hold in your hand.

Current

Whenever there is movement of charge carriers in a substance, there is an electric *current.* Current is measured in terms of the number of electrons or holes passing a single point in 1 second.

A great many charge carriers go past any given point in 1 second, even if the current is small. In a household electric circuit, a 100-watt light bulb draws a current of about *six quintillion* (6 followed by 18 zeros) charge carriers per second. Even the smallest bulb carries *quadrillions* (numbers followed by 15 zeros) of charge carriers every second. It is impractical to speak of a current in terms of charge carriers per second, so it is measured in *coulombs per second* instead. A coulomb is equal to approximately 6,240,000,000,000,000,000 electrons or holes. A current of 1 coulomb per second is called an *ampere*, and this is the standard unit of electric current. A 100-watt bulb in your desk lamp draws about 1 ampere of current.

When a current flows through a resistance—and this is always the case because even the best conductors have resistance—heat is generated. Sometimes light and other forms of energy are emitted as well. A light bulb is deliberately designed so that the resistance causes visible light to be generated.

Electric current flows at high speed through any conductor, resistor, or semiconductor. Nevertheless, it is considerably less than the speed of light.

Static Electricity

Charge carriers, particularly electrons, can build up, or become deficient, on things without flowing anywhere. You've experienced this when walking on a carpeted floor during the winter, or in a place where the humidity was low. An excess or shortage of electrons is created on and in your body. You acquire a *charge* of *static electricity.* It's called "static" because it doesn't go anywhere. You don't feel this until you touch some metallic object that is connected to earth ground or to some large fixture; but then there is a *discharge*, accompanied by a spark.

If you were to become much more charged, your hair would stand on end, because every hair would repel every other. Like charges are caused either by an excess or a deficiency of electrons; they repel. The spark might jump an inch, 2 inches, or even 6 inches. Then it would more than startle you; you could get hurt. This doesn't happen with ordinary carpet and shoes, fortunately. But a device called a *Van de Graaff generator*, found in physics labs, can cause a spark this large (Fig. 1-7). Be careful when using this device for physics experiments!

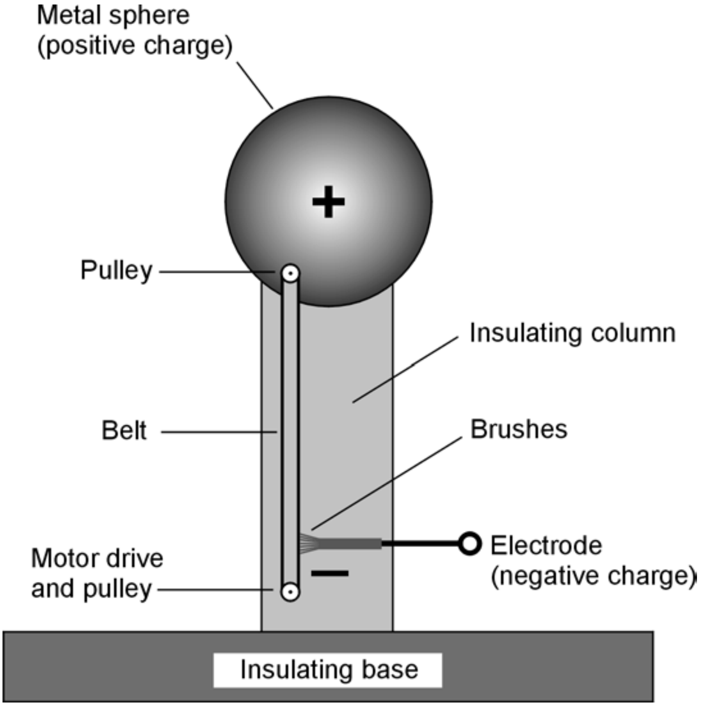

1-7 Simplified illustration of a Van de Graaff generator. This machine can create a charge buildup large enough to produce a spark several centimeters long.

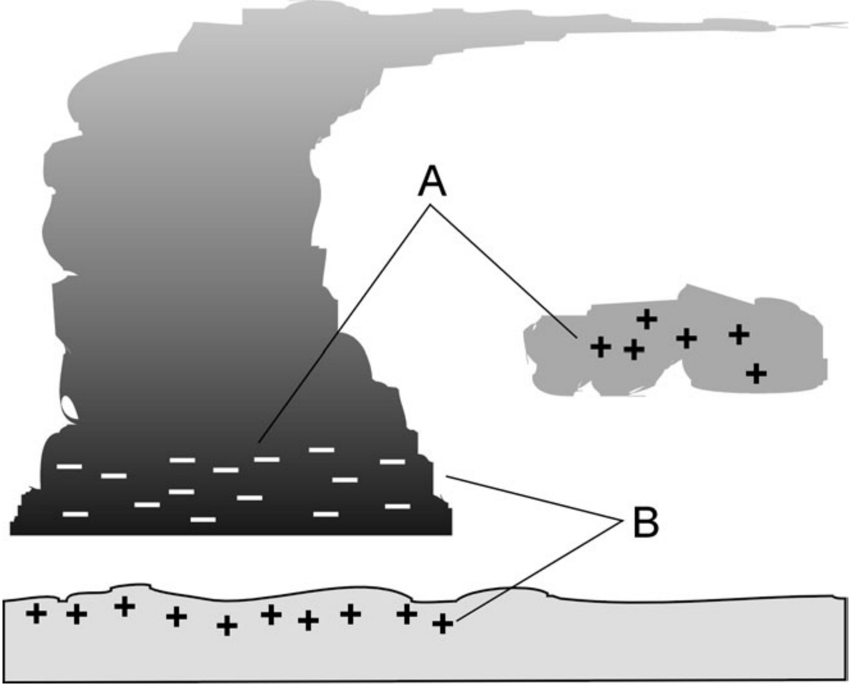

1-8 Electrostatic charges can build up between clouds in a thunderstorm (A), or between a cloud and the surface of the earth (B).

In the extreme, lightning occurs between clouds, and between clouds and ground in the earth's atmosphere. This spark, called a *stroke*, is a magnified version of the spark you get after shuffling around on a carpet. Until the stroke occurs, there is a static charge in the clouds, between different clouds or parts of a cloud, and the ground. In Fig. 1-8, cloud-to-cloud (A) and cloud-to-ground (B) static buildups are shown. In the case at B, the positive charge in the earth follows along beneath the storm cloud. The current in a lightning stroke is usually several tens of thousands, or hundreds of thousands, of amperes. But it takes place only for a fraction of a second. Still, many coulombs of charge are displaced in a single bolt of lightning.

Electromotive Force

Current can only flow if it gets a "push." This can be caused by a buildup of static electric charges, as in the case of a lightning stroke. When the charge builds up, with positive polarity (shortage of electrons) in one place and negative polarity (excess of electrons) in another place, a powerful *electromotive force* (EMF) exists. This force is measured in units called *volts.*

Ordinary household electricity has an effective voltage of between 110 and 130; usually it is about 117. A car battery has an EMF of 12 to 14 volts. The static charge that you acquire when walking on a carpet with hard-soled shoes is often several thousand volts. Before a discharge of lightning, millions of volts exist. An EMF of 1 volt, across a resistance of 1 ohm, will cause a current of 1 ampere to flow. This is a classic relationship in electricity, and is stated generally as *Ohm's Law.* If the EMF is doubled, the current is doubled. If the resistance is doubled, the current is cut in half. This important law of electrical circuit behavior is covered in detail later in this book.

It is possible to have an EMF without having any current. This is the case just before a lightning stroke occurs, and before you touch a metal object after walking on a carpet. It is also true between the two wires of an electric lamp when the switch is turned off. It is true of a dry cell when there is nothing connected to it. There is no current, but a current is possible given a conductive path between the two points. Voltage, or EMF, is sometimes called *potential* or *potential difference* for this reason.

Even a huge EMF does not necessarily drive much current through a conductor or resistance. A good example is your body after walking around on the carpet. Although the voltage seems deadly in terms of numbers (thousands), there are not many coulombs of static-electric charge that can accumulate on an object the size of your body. Therefore, in relative terms, not that many electrons flow through your finger when you touch a radiator. This is why you don't get a severe shock.

If there are plenty of coulombs available, a small voltage, such as 117 volts (or even less) can cause a lethal current. This is why it is dangerous to repair an electrical device with the power on. The power plant will pump an unlimited number of coulombs of charge through your body if you are not careful.

Nonelectrical Energy

In electricity and electronics, there are phenomena that involve other forms of energy besides electrical energy. Visible light is an example. A light bulb converts electricity into radiant energy that you can see. This was one of the major motivations for people like Thomas Edison to work with electricity. Visible light can also be converted into electric current or voltage. A *photovoltaic cell* does this.

Light bulbs always give off some heat, as well as visible light. Incandescent lamps actually give off more energy as heat than as light. You are certainly acquainted with electric heaters, designed for the purpose of changing electricity into heat energy. This heat is a form of radiant energy called *infrared* (IR). It is similar to visible light, except that the waves are longer and you can't see them.

Electricity can be converted into other radiant-energy forms, such as *radio waves*, *ultraviolet* (UV), and *X rays*. This is done by specialized devices such as radio transmitters, sunlamps, and electron tubes. Fast-moving protons, neutrons, electrons, and atomic nuclei are an important form of energy. The energy from these particles is sometimes sufficient to split atoms apart. This effect makes it possible to build an atomic reactor whose energy can be used to generate electricity.

When a conductor moves in a magnetic field, electric current flows in that conductor. In this way, mechanical energy is converted into electricity. This is how an *electric generator* works. Generators can also work backward. Then you have a *motor* that changes electricity into useful mechanical energy.

A magnetic field contains energy of a unique kind. The science of *magnetism* is closely related to electricity. Magnetic phenomena are of great significance in electronics. The oldest and most universal source of magnetism is the *geomagnetic field* surrounding the earth, caused by alignment of iron atoms in the core of the planet.

A changing magnetic field creates a fluctuating electric field, and a fluctuating electric field produces a changing magnetic field. This phenomenon, called *electromagnetism*, makes it possible to send wireless signals over long distances. The electric and magnetic fields keep producing one another over and over again through space.

Chemical energy is converted into electricity in *dry cells*, *wet cells*, and *batteries.* Your car battery is an excellent example. The acid reacts with the metal electrodes to generate an electromotive force. When the two poles of the batteries are connected, current results. The chemical reaction continues, keeping the current going for a while. But the battery can only store a certain amount of chemical energy. Then it "runs out of juice," and the supply of chemical energy must be restored by *charging.* Some cells and batteries, such as lead-acid car batteries, can be recharged by driving current through them, and others, such as most flashlight and transistor-radio batteries, cannot.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct answers out of these 20 questions. The answers are listed in the back of this book.

1. The atomic number of an element is determined by
    (a) the number of neutrons.
    (b) the number of protons.
    (c) the number of neutrons plus the number of protons.
    (d) the number of electrons.

2. The atomic weight of an element is approximately determined by
    (a) the number of neutrons.
    (b) the number of protons.
    (c) the number of neutrons plus the number of protons.
    (d) the number of electrons.

3. Suppose there is an atom of oxygen, containing eight protons and eight neutrons in the nucleus, and two neutrons are added to the nucleus. What is the resulting atomic weight?
    (a) 8
    (b) 10
    (c) 16
    (d) 18

4. An ion
    (a) is electrically neutral.
    (b) has positive electric charge.
    (c) has negative electric charge.
    (d) can have either a positive or negative charge.

5. An isotope
    (a) is electrically neutral.
    (b) has positive electric charge.
    (c) has negative electric charge.
    (d) can have either a positive or negative charge.

6. A molecule
   (a) can consist of a single atom of an element.
   (b) always contains two or more elements.
   (c) always has two or more atoms.
   (d) is always electrically charged.
7. In a compound,
   (a) there can be a single atom of an element.
   (b) there must always be two or more elements.
   (c) the atoms are mixed in with each other but not joined.
   (d) there is always a shortage of electrons.
8. An electrical insulator can be made a conductor
   (a) by heating it.
   (b) by cooling it.
   (c) by ionizing it.
   (d) by oxidizing it.
9. Of the following substances, the worst conductor is
   (a) air.
   (b) copper.
   (c) iron.
   (d) salt water.
10. Of the following substances, the best conductor is
    (a) air.
    (b) copper.
    (c) iron.
    (d) salt water.
11. Movement of holes in a semiconductor
    (a) is like a flow of electrons in the same direction.
    (b) is possible only if the current is high enough.
    (c) results in a certain amount of electric current.
    (d) causes the material to stop conducting.
12. If a material has low resistance, then
    (a) it is a good conductor.
    (b) it is a poor conductor.
    (c) the current flows mainly in the form of holes.
    (d) current can flow only in one direction.
13. A coulomb
    (a) represents a current of 1 ampere.
    (b) flows through a 100-watt light bulb.
    (c) is equivalent to 1 ampere per second.
    (d) is an extremely large number of charge carriers.

14. A stroke of lightning
    - (a) is caused by a movement of holes in an insulator.
    - (b) has a very low current.
    - (c) is a discharge of static electricity.
    - (d) builds up between clouds.

15. The volt is the standard unit of
    - (a) current.
    - (b) charge.
    - (c) electromotive force.
    - (d) resistance.

16. If an EMF of 1 volt is placed across a resistance of 2 ohms, then the current is
    - (a) half an ampere.
    - (b) 1 ampere.
    - (c) 2 amperes.
    - (d) impossible to determine.

17. A backward-working electric motor, in which mechanical rotation is converted to electricity, is best described as
    - (a) an inefficient, energy-wasting device.
    - (b) a motor with the voltage connected the wrong way.
    - (c) an electric generator.
    - (d) a magnetic field.

18. In a battery, chemical energy can sometimes be replenished by
    - (a) connecting it to a light bulb.
    - (b) charging it.
    - (c) discharging it.
    - (d) no means known; when a battery is dead, you must throw it away.

19. A fluctuating magnetic field
    - (a) produces an electric current in an insulator.
    - (b) magnetizes the earth.
    - (c) produces a fluctuating electric field.
    - (d) results from a steady electric current.

20. Visible light is converted into electricity
    - (a) in a dry cell.
    - (b) in a wet cell.
    - (c) in an incandescent bulb.
    - (d) in a photovoltaic cell.

2
CHAPTER
Electrical Units

THIS CHAPTER EXPLAINS, IN MORE DETAIL, STANDARD UNITS THAT DEFINE THE BEHAVIOR OF DIRECT-current (dc) circuits. Many of these rules also apply to utility alternating-current (ac) circuits.

The Volt

In Chap. 1, you learned a little about the volt, the standard unit of electromotive force (EMF) or *potential difference.*

An accumulation of electrostatic charge, such as an excess or shortage of electrons, is always associated with a voltage. There are other situations in which voltages exist. Voltage can be generated at a power plant, produced in an electrochemical reaction, or caused by light rays striking a semiconductor chip. It can be produced when an object is moved in a magnetic field, or is placed in a fluctuating magnetic field.

A potential difference between two points produces an *electric field*, represented by *electric lines of flux* (Fig. 2-1). There is a pole that is relatively positive, with fewer electrons, and one that is relatively negative, with more electrons. The positive pole does not necessarily have a deficiency of electrons compared with neutral objects, and the negative pole does not always have a surplus of electrons relative to neutral objects. But the negative pole always has more electrons than the positive pole.

The abbreviation for volt (or volts) is V. Sometimes, smaller units are used. The *millivolt* (mV) is equal to a thousandth (0.001) of a volt. The *microvolt* (μV) is equal to a millionth (0.000001) of a volt. It is sometimes necessary to use units larger than the volt. One *kilovolt* (kV) is one thousand volts (1000 V). One *megavolt* (MV) is 1 million volts (1,000,000 V) or one thousand kilovolts (1000 kV).

In a dry cell, the voltage is usually between 1.2 and 1.7 V; in a car battery, it is 12 to 14 V. In household utility wiring, it is a low-frequency alternating current of about 117 V for electric lights and most appliances, and 234 V for a washing machine, dryer, oven, or stove. In television sets, transformers convert 117 V to around 450 V for the operation of the picture tube. In some broadcast transmitters, the voltage can be several kilovolts.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

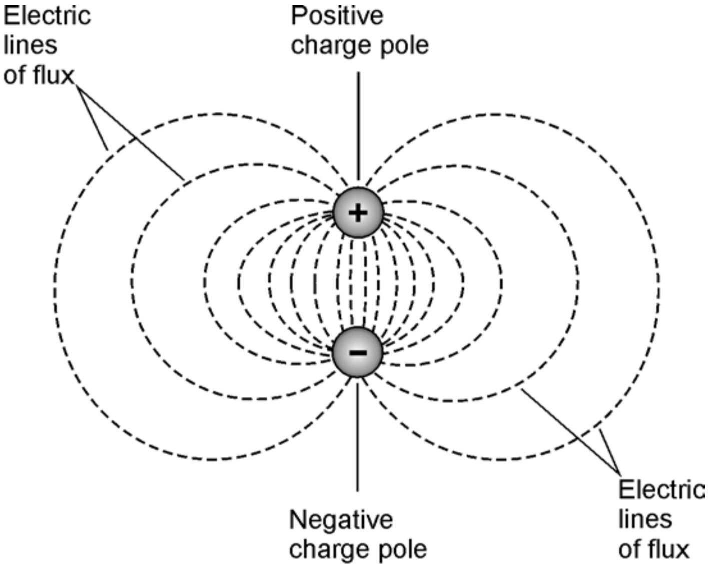

2-1 Electric lines of flux always exist near poles of electric charge.

The largest voltages on our planet occur between clouds, or between clouds and the ground, in thundershowers. This potential difference can build up to several tens of megavolts. The existence of a voltage always means that *charge carriers*, which are electrons in a conventional circuit, flow between two points if a conductive path is provided. Voltage represents the driving force that impels charge carriers to move. If all other factors are held constant, high voltages produce a faster flow of charge carriers, and therefore larger currents, than low voltages. But that's an oversimplification in most real-life scenarios, where other factors are hardly ever constant!

Current Flow

If a conducting or semiconducting path is provided between two poles having a potential difference, charge carriers flow in an attempt to equalize the charge between the poles. This flow of *current* continues as long as the path is provided, and as long as there is a charge difference between the poles.

Sometimes the charge difference is equalized after a short while. This is the case, for example, when you touch a radiator after shuffling around on the carpet while wearing hard-soled shoes. It is also true in a lightning stroke. In these instances, the charge is equalized in a fraction of a second. In other cases, the charge takes longer to be used up. This happens if you short-circuit a dry cell. Within a few minutes, the cell "runs out of juice" if you put a wire between the positive and negative terminals. If you put a bulb across the cell, say with a flashlight, it takes an hour or two for the charge difference to drop to zero.

In household electric circuits, the charge difference is never equalized, unless there's a power failure. Of course, if you short-circuit an outlet (don't!), the fuse or breaker will blow or trip, and the charge difference will immediately drop to zero. But if you put a 100-watt bulb at the outlet, the charge difference will be maintained as the current flows. The power plant can keep a potential difference across a lot of light bulbs indefinitely.

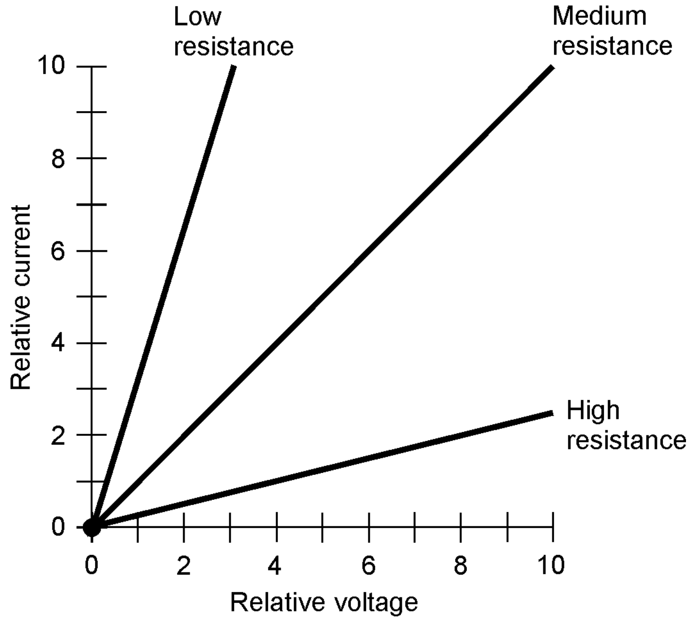

2-2 Relative current as a function of relative voltage for low, medium, and high resistances.

Have you heard that it is current, not voltage, that kills? This is a literal truth, but it plays on semantics. It's like saying "It's the heat, not the fire, that burns you." Naturally! But there can only be a deadly current if there is enough voltage to drive it through your body. You don't have to worry when handling flashlight cells, but you'd better be extremely careful around household utility circuits. A voltage of 1.2 to 1.7 V can't normally pump a dangerous current through you, but a voltage of 117 V almost always can.

In an electric circuit that always conducts equally well, the current is directly proportional to the applied voltage. If you double the voltage, you double the current. If the voltage is cut in half, the current is cut in half too. Figure 2-2 shows this relationship as a graph in general terms. It assumes that the power supply can provide the necessary number of charge carriers.

The Ampere

Current is a measure of the rate at which charge carriers flow. The standard unit is the *ampere.* This represents one coulomb (6,240,000,000,000,000,000) of charge carriers flowing every second past a given point.

An ampere is a comparatively large amount of current. The abbreviation is A. Often, current is specified in terms of *milliamperes*, abbreviated mA, where 1 mA = 0.001 A, or a thousandth of an ampere. You will also sometimes hear of *microamperes* (μA), where 1 μA = 0.000001 A or 0.001 mA, which is a millionth of an ampere. It is increasingly common to hear about *nanoamperes* (nA), where 1 nA = 0.001 μA = 0.000000001 A, which is a thousandth of a millionth of an ampere.

A current of a few milliamperes will give you a startling shock. About 50 mA will jolt you severely, and 100 mA can cause death if it flows through your chest cavity. An ordinary 100-watt light bulb draws about 1 A of current in a household utility circuit. An electric iron draws approximately

10 A; an entire household normally uses between 10 and 50 A, depending on the size of the house and the kinds of appliances it has, and also on the time of day, week, or year.

The amount of current that flows in an electrical circuit depends on the voltage, and also on the *resistance.* There are some circuits in which extremely large currents, say 1000 A, can flow. This will happen through a metal bar placed directly at the output of a massive electric generator. The resistance is extremely low in this case, and the generator is capable of driving huge numbers of charge carriers through the bar every second. In some semiconductor electronic devices, such as microcomputers, a few nanoamperes will suffice for many complicated processes. Some electronic clocks draw so little current that their batteries last as long as they would if left on the shelf without being put to any use.

Resistance and the Ohm

Resistance is a measure of the opposition that a circuit offers to the flow of electric current. You can compare it to the diameter of a hose. In fact, for metal wire, this is an excellent analogy: small-diameter wire has high resistance (a lot of opposition to current), and large-diameter wire has low resistance (not much opposition to current). The type of metal makes a difference too. For example, steel wire has higher resistance for a given diameter than copper wire.

The standard unit of resistance is the *ohm.* This is sometimes symbolized by the uppercase Greek letter omega (Ω). You'll sometimes hear about *kilohms* (symbolized k or kΩ), where 1 kΩ = 1000 Ω, or about *megohms* (symbolized M or MΩ), where 1 MΩ = 1000 kΩ = 1,000,000 Ω.

Electric wire is sometimes rated for *resistivity.* The standard unit for this purpose is the *ohm per foot* (ohm/ft or Ω/ft) or the *ohm per meter* (ohm/m or Ω/m). You might also come across the unit *ohm per kilometer* (ohm/km or Ω/km). Table 2-1 shows the resistivity for various common sizes of solid copper wire at room temperature, as a function of the wire size as defined by a scheme known as the *American Wire Gauge* (AWG).

Table 2-1. Approximate resistivity at room temperature for solid copper wire as a function of the wire size in American Wire Gauge (AWG).

Wire size, AWG #	Resistivity, ohms/km
2	0.52
4	0.83
6	1.3
8	2.7
10	3.3
12	5.3
14	8.4
16	13
18	21
20	34
22	54
24	86
26	140
28	220
30	350

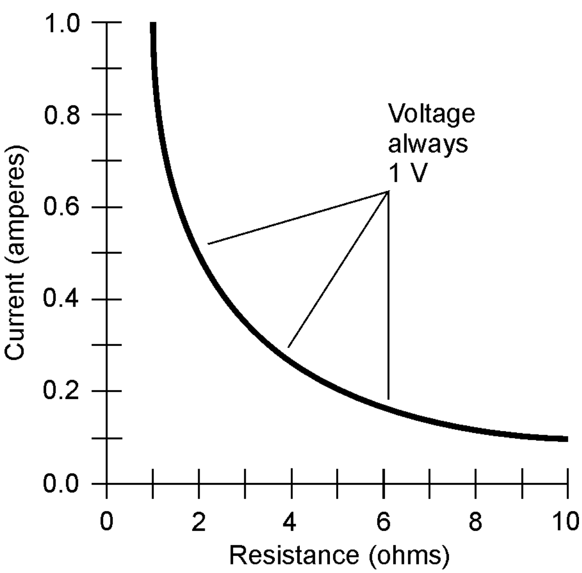

2-3 Current as a function of resistance through an electric device for a constant voltage of 1 V.

When 1 V is placed across 1 Ω of resistance, assuming that the power supply can deliver an unlimited number of charge carriers, there is a current of 1 A. If the resistance is doubled to 2 Ω, the current decreases to 0.5 A. If the resistance is cut by a factor of 5 to 0.2 Ω, the current increases by the same factor, to 5 A. The current flow, for a constant voltage, is said to be *inversely proportional* to the resistance. Figure 2-3 is a graph that shows various currents, through various resistances, given a constant voltage of 1 V across the whole resistance.

Resistance has another property. If there is a current flowing through a resistive material, there is always a potential difference across the resistive component (called a *resistor*). This is shown in Fig. 2-4. In general, this voltage is directly proportional to the current through the resistor. This behavior of resistors is useful in the design of electronic circuits, as you will learn later in this book.

Electrical circuits always have some resistance. There is no such thing as a perfect conductor. When some metals are chilled to temperatures near *absolute zero*, they lose practically all of their resistance, but they never become absolutely perfect, resistance-free conductors. This phenomenon, about which you might have heard, is called *superconductivity.*

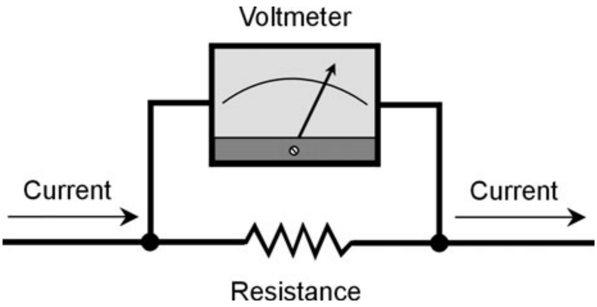

2-4 Whenever current passes through a component having resistance, a voltage exists across that component.

Just as there is no such thing as a perfectly resistance-free substance, there isn't a truly infinite resistance, either. Even air conducts to some extent, although the effect is usually so small that it can be ignored. In some electronic applications, materials are selected on the basis of how "nearly infinite" their resistance is.

In electronics, the resistance of a component often varies, depending on the conditions under which it is operated. A transistor, for example, might have high resistance some of the time, and low resistance at other times. High/low resistance variations can be made to take place thousands, millions, or billions of times each second. In this way, oscillators, amplifiers, and digital devices function in radio receivers and transmitters, telephone networks, digital computers, and satellite links (to name just a few applications).

Conductance and the Siemens

Electricians and electrical engineers sometimes talk about the *conductance* of a material, rather than about its resistance. The standard unit of conductance is the *siemens*, abbreviated S. When a component has a conductance of 1 S, its resistance is 1 Ω. If the resistance is doubled, the conductance is cut in half, and vice versa. Therefore, conductance is the reciprocal of resistance.

If you know the resistance of a component or circuit in ohms, you can get the conductance in siemens: divide 1 by the resistance. If you know the conductance in siemens, you can get the resistance: divide 1 by the conductance. Resistance, as a variable quantity, is denoted by an italicized, uppercase letter R. Conductance, as a variable quantity, is denoted as an italicized, uppercase letter G. If we express R in ohms and G in siemens, then the following two equations describe their relationship:

$$G = 1/R$$
$$R = 1/G$$

Units of conductance much smaller than the siemens are often used. A resistance of 1 kΩ is equal to 1 *millisiemens* (1 mS). If the resistance is 1 MΩ, the conductance is one *microsiemens* (1 μS). You'll sometimes hear about *kilosiemens* (kS) or *megasiemens* (MS), representing resistances of 0.001 Ω and 0.000001 Ω (a thousandth of an ohm and a millionth of an ohm, respectively). Short lengths of heavy wire have conductance values in the range of kilosiemens. Heavy metal rods can have conductances in the megasiemens range.

Suppose a component has a resistance of 50 Ω. Then its conductance, in siemens, is 1/50 S, which is equal to 0.02 S. We can call this 20 mS. Or imagine a piece of wire with a conductance of 20 S. Its resistance is 1/20 Ω, which is equal to 0.05 Ω. You will not often hear the term *milliohm.* But you could say that this wire has a resistance of 50 mΩ, and you would be technically right.

Determining *conductivity* is tricky. If wire has a resistivity of 10 Ω/km, you can't say that it has a conductivity of 1/10, or 0.1, S/km. It is true that a kilometer of such wire has a conductance of 0.1 S, but 2 km of the wire has a resistance of 20 Ω (because there is twice as much wire). That is not twice the conductance, but half. If you say that the conductivity of the wire is 0.1 S/km, then you might be tempted to say that 2 km of the wire has 0.2 S of conductance. That would be a mistake! Conductance decreases with increasing wire length.

Figure 2-5 illustrates the resistance and conductance values for various lengths of wire having a resistivity of 10 Ω/km.

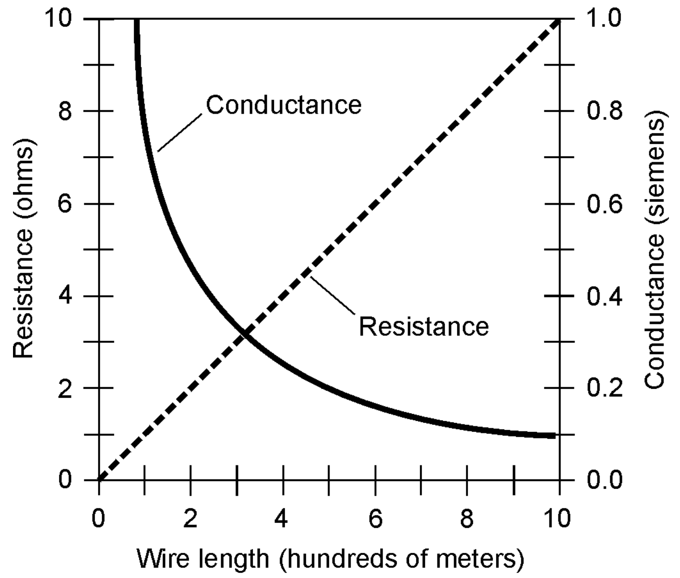

2-5 Resistance and conductance for various lengths of wire having a resistivity of 10 ohms per kilometer.

Power and the Watt

Whenever current flows through a resistance, heat results. The heat can be measured in *watts* (symbolized W) and represents electrical *power.* (As a variable quantity in equations, power is denoted by the uppercase italic letter *P.*) Power can be manifested in many forms, such as mechanical motion, radio waves, visible light, or noise. But heat is always present, in addition to any other form of power, in an electrical or electronic device. This is because no equipment is 100 percent efficient. Some power always goes to waste, and this waste is almost all in the form of heat.

Look again at Fig. 2-4. There is a certain voltage across the resistor, not specifically indicated. There's also a current flowing through the resistance, and it is not quantified in the diagram, either. Suppose we call the voltage E and the current I, in volts (V) and amperes (A), respectively. Then the power in watts dissipated by the resistance, call it P, is the product of the voltage in volts and the current in amperes:

$$P = EI$$

If the voltage E across the resistance is caused by two flashlight cells in series, giving 3 V, and if the current I through the resistance (a light bulb, perhaps) is 0.1 A, then $E = 3$ V and $I = 0.1$ A, and we can calculate the power P in watts as follows:

$$P = EI = 3 \times 0.1 = 0.3 \text{ W}$$

Suppose the voltage is 117 V, and the current is 855 mA. To calculate the power, we must convert the current into amperes: 855 mA = 855/1000 A = 0.855 A. Then:

$$P = EI = 117 \times 0.855 = 100 \text{ W}$$

Table 2-2. Prefix multipliers from 0.000000000001 (trillionths, or units of 10^{-12}) to 1,000,000,000,000 (trillions, or units of 10^{12}).

Prefix	Symbol	Multiplier
pico-	p	0.000000000001 (or 10^{-12})
nano-	n	0.000000001 (or 10^{-9})
micro-	µ	0.000001 (or 10^{-6})
milli-	m	0.001 (or 10^{-3})
kilo-	k	1000 (or 10^{3})
mega-	M	1,000,000 (or 10^{6})
giga-	G	1,000,000,000 (or 10^{9})
tera-	T	1,000,000,000,000 (or 10^{12})

You will often hear about *milliwatts* (mW), *microwatts* (µW), *kilowatts* (kW), and *megawatts* (MW). By now, you should be able to tell from the prefixes what these units represent. Otherwise, you can refer to Table 2-2. This table lists the most commonly used *prefix multipliers* in electricity and electronics.

Sometimes you need to use the power equation to find currents or voltages. Then you should use $I = P/E$ to find current, or $E = P/I$ to find voltage. Always remember to convert, if necessary, to the standard units of volts, amperes, and watts before performing the calculations.

A Word about Notation

Have you noticed some strange things about the notation yet? If you're observant, you have! Why, you might ask, are italics sometimes used, and sometimes not used? Something should be said early in this course about notation, because it can get confusing with all the different symbols and abbreviations. Sometimes, symbols and abbreviations appear in italics, and sometimes they do not. You'll see subscripts often, and sometimes even they are italicized! Here are some rules that apply to notation in electricity and electronics:

- Symbols for specific units, such as volts, amperes, and ohms, are not italicized.
- Symbols for objects or components, such as resistors, batteries, and meters, are not italicized.
- Quantifying prefixes, such as "kilo-" or "micro-," are not italicized.
- Labeled points in drawings might or might not be italicized; it doesn't matter as long as a diagram is consistent with itself.
- Symbols for mathematical constants and variables, such as time, are italicized.
- Symbols for electrical quantities, such as voltage, current, resistance, and power, are italicized.
- Symbols and abbreviations for modifiers might or might not be italicized; it doesn't matter as long as a document is consistent with itself.
- Numeric subscripts are not italicized.
- For nonnumeric subscripts, the same rules apply as for general symbols.

Some examples are R (not italicized) for resistor, *R* (italicized) for resistance, *P* (italicized) for power, W (not italicized) for watts, V (not italicized) for volts, *E* or *V* (italicized) for voltage, A (not italicized) for amperes, *I* (italicized) for current, *f* (italicized) for frequency, and *t* (italicized) for time.

Once in a while you will see the same symbol italicized in one place and not in another—in the same circuit diagram or discussion! We might, for example, talk about "resistor number 3" (symbolized R_3), and then later in the same paragraph talk about its value as "resistance number 3" (Symbolized R_3). Still later we might talk about "the *n*th resistor in a series connection" (R_n) and then "the *n*th resistance in a series combination of resistances" (R_n).

These differences in notation, while subtle (and, some people will say, picayune) are followed in this book, and they are pretty much agreed upon by convention. They are important because they tell the reader exactly what a symbol stands for in a diagram, discussion, or mathematical equation. "Resistor" and "resistance" are vastly different things—as different from each other as a garden hose (the object) and the extent to which it impedes the flow of water (the phenomenon). With this in mind, let us proceed!

Energy and the Watt-Hour

Have you heard the terms "power" and "energy" used interchangeably, as if they mean the same thing? They don't! *Energy* is power dissipated over a length of time. *Power* is the rate at which energy is expended. Physicists measure energy in units called *joules.* One joule (1 J) is the equivalent of a *watt-second*, which is the equivalent of 1 watt of power dissipated for 1 second of time (1 W · s or Ws). In electricity, you'll more often encounter the *watt-hour* (symbolized W · h or Wh) or the *kilowatt-hour* (symbolized kW · h or kWh). As their names imply, a watt-hour is the equivalent of 1 W dissipated for 1 h, and 1 kWh is the equivalent of 1 kW of power dissipated for 1 h.

A watt-hour of energy can be dissipated in an infinite number of different ways. A 60-W bulb consumes 60 Wh in 1 h, the equivalent of a watt-hour per minute (1 Wh/min). A 100-W bulb consumes 1 Wh in 1/100 h, or 36 s. Besides these differences, the rate of power dissipation in real-life circuits often changes with time. This can make the determination of consumed energy complicated, indeed.

Figure 2-6 illustrates two hypothetical devices that consume 1 Wh of energy. Device A uses its power at a constant rate of 60 W, so it consumes 1 Wh in 1 min. The power consumption rate of

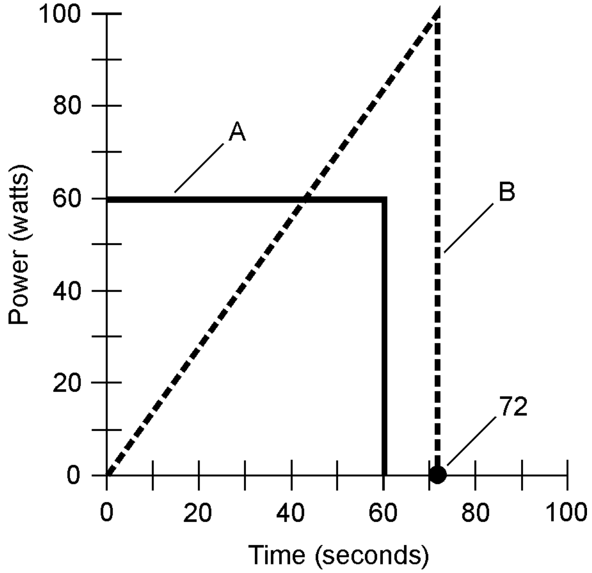

2-6 Two devices that burn 1 Wh of energy. Device A dissipates a constant amount of power. Device B dissipates a variable amount of power.

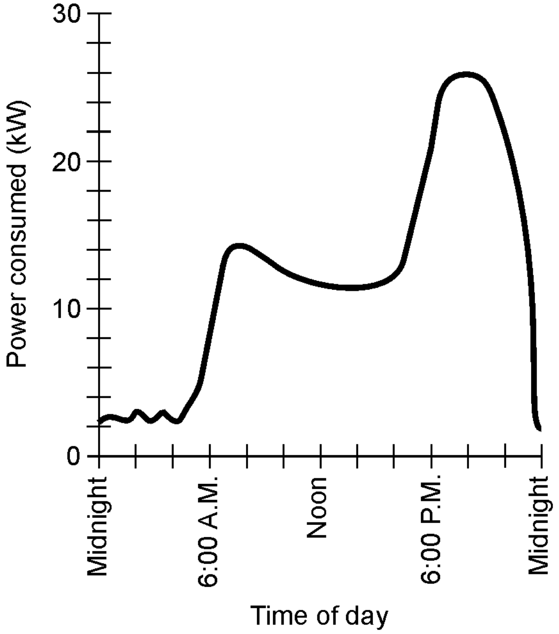

2-7 A graph showing the amount of power consumed by a hypothetical household, as a function of the time of day.

device B varies, starting at zero and ending up at quite a lot more than 60 W. How do you know that this second device really consumes 1 Wh of energy? You must determine the area under the curve in the graph. In this case, figuring out this area is easy, because the enclosed object is a triangle. The area of a triangle is equal to half the product of the base length and the height. Device B is powered up for 72 s, or 1.2 min; this is 1.2/60 = 0.02 h. Then the area under the curve is $1/2 \times 100 \times 0.02 = 1$ Wh.

When calculating energy values, you must always remember the units you're using. In this case the unit is the watt-hour, so you must multiply watts by hours. If you multiply watts by minutes, or watts by seconds, you'll get the wrong kind of units in your answer.

Often, the curves in graphs like these are complicated. Consider the graph of power consumption in your home, versus time, for a day. It might look like the curve in Fig. 2-7. Finding the area under this curve is not easy. But there is another way to determine the total energy burned by your household over a period of time. That is by means of a meter that measures electrical energy in kilowatt-hours. Every month, without fail, the power company sends its representative to read your electric meter. This person takes down the number of kilowatt-hours displayed, subtracts the number from the reading taken the previous month, and a few days later you get a bill. This meter automatically keeps track of total consumed energy, without anybody having to go through high-level mathematical calculations to find the areas under irregular curves such as the graph of Fig. 2-7.

Other Energy Units

The joule, while standard among scientists, is not the only energy unit in existence! Another unit is the *erg*, equivalent to one ten-millionth (0.0000001) of a joule. The erg is used in lab experiments involving small amounts of expended energy.

Table 2-3. Conversion factors between joules and various other energy units.

Unit	To convert energy in this unit to energy in joules, multiply by	To convert energy in joules to energy in this unit, multiply by
British thermal units (Btu)	1055	0.000948
Electron volts (eV)	1.6×10^{-19}	6.2×10^{18}
Ergs	0.0000001 (or 10^{-7})	10,000,000 (or 10^{7})
Foot-pounds (ft-lb)	1.356	0.738
Watt-hours (Wh)	3600	0.000278
Kilowatt-hours (kWh)	3,600,000 (or 3.6×10^{6})	0.000000278 (or 2.78×10^{-7})

Most folks have heard of the *British thermal unit* (Btu), equivalent to 1055 joules. This is the energy unit commonly used to define the cooling or heating capacity of air-conditioning equipment. To cool your room from 85 to 78°F needs a certain amount of energy, perhaps best specified in Btu. If you are getting an air conditioner or furnace installed in your home, an expert will come look at your situation, and determine the size of air-conditioning/heating unit that best suits your needs. That person will likely tell you how powerful the unit should be in terms of its ability to heat or cool in Btu per hour (Btu/h).

Physicists also use, in addition to the joule, a unit of energy called the *electron volt* (eV). This is a tiny unit of energy, equal to just 0.00000000000000000016 joule (there are 18 zeroes after the decimal point and before the 1). The physicists write 1.6×10^{-19} to represent this. It is the energy gained by a single electron in an electric field of 1 V. Machines called *particle accelerators* (or *atom smashers*) are rated by millions of electron volts (MeV), billions of electron volts (GeV), or trillions of electron volts (TeV) of energy capacity.

Another energy unit, employed to denote work, is the *foot-pound* (ft-lb). This is the work needed to raise a weight of one pound by a distance of one foot, not including any friction. It's equal to 1.356 joules.

All of these units, and conversion factors, are given in Table 2-3. Kilowatt-hours and watt-hours are also included in this table. In electricity and electronics, you need to be concerned only with the watt-hour and the kilowatt-hour for most purposes.

Alternating Current and the Hertz

This chapter, and this whole first section, is mostly concerned with *direct current* (dc). That's electric current that always flows in the same direction and that does not change in intensity (at least not too rapidly) with time. But household utility current is not of this kind. It reverses direction periodically, exactly once every 1/120 second. It goes through a complete cycle every 1/60 second. Every repetition is identical to every other. This is *alternating current* (ac).

Figure 2-8 shows the characteristic wave of ac, as a graph of voltage versus time. Notice that the maximum positive and negative voltages are not 117 V, as you've heard about household electricity, but close to 165 V. There is a reason for this difference. The *effective voltage* for an ac wave is never the same as the *instantaneous maximum*, or *peak*, voltage. In fact, for the common waveform shown in Fig. 2-8, the effective value is 0.707 times the peak value. Conversely, the peak value is 1.414 times the effective value.

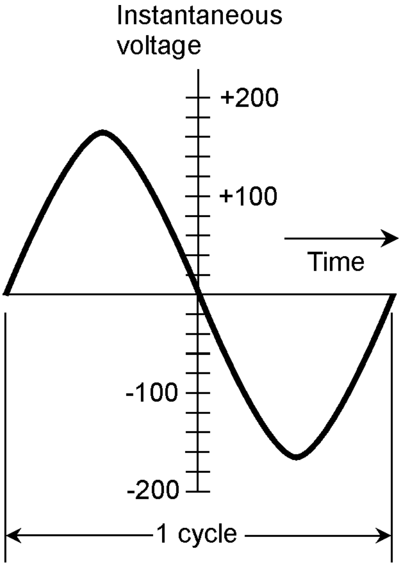

2-8 One cycle of utility alternating current (ac). The instantaneous voltage is the voltage at any particular instant in time. The peak voltages are approximately plus and minus 165 V.

Because the whole cycle repeats itself every 1/60 second, the *frequency* of the utility ac wave is said to be 60 *hertz*, abbreviated 60 Hz. The German word *hertz* literally translates to "cycles per second." In the United States, this is the standard frequency for ac. In some places it is 50 Hz.

In wireless communications, higher frequencies are common, and you'll hear about *kilohertz* (kHz), *megahertz* (MHz), and *gigahertz* (GHz). The relationships among these units are as follows:

$$1 \text{ kHz} = 1000 \text{ Hz}$$
$$1 \text{ MHz} = 1000 \text{ kHz} = 1{,}000{,}000 \text{ Hz}$$
$$1 \text{ GHz} = 1000 \text{ MHz} = 1{,}000{,}000 \text{ kHz} = 1{,}000{,}000{,}000 \text{ Hz}$$

Usually, but not always, the waveshapes are of the type shown in Fig. 2-8. This waveform is known as a *sine wave* or a *sinusoidal* waveform.

Rectification and Pulsating Direct Current

Batteries and other sources of direct current (dc) produce constant voltage. This can be represented by a straight, horizontal line on a graph of voltage versus time (Fig. 2-9). For pure dc, the peak and effective values are identical. But sometimes the value of dc voltage fluctuates rapidly with time. This happens, for example, if the waveform in Fig. 2-8 is passed through a *rectifier* circuit.

Rectification is a process in which ac is changed to dc. The most common method of doing this uses a device called the *diode.* Right now, you need not be concerned with how the rectifier circuit is put together. The point is that part of the ac wave is either cut off, or turned around upside down, so the output is *pulsating dc.* Figure 2-10 illustrates two different waveforms of pulsating dc. In the waveform at A, the negative (bottom) part has been cut off. At B, the negative portion of the wave has been inverted and made positive. The situation at A is known as *half-wave rectification*, because it involves only half the waveform. At B, the ac has been subjected to *full-wave rectification*, because

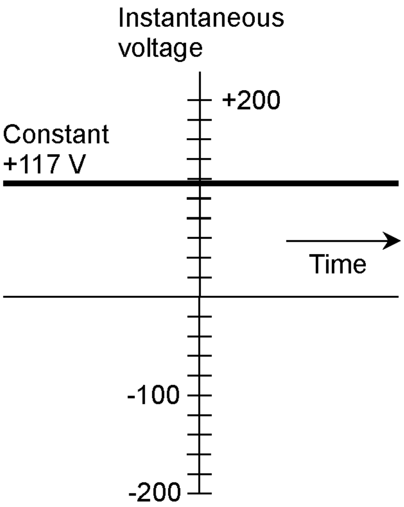

2-9 A representation of pure direct current (dc).

all of the original current still flows, even though the alternating nature has been changed so that the current never reverses.

The effective value, compared with the peak value, for pulsating dc depends on whether half-wave or full-wave rectification is applied to an ac wave. In Fig. 2-10A and B, effective voltage is shown as dashed lines, and the *instantaneous voltage* is shown as solid curves. The instantaneous voltage changes all the time, from instant to instant. (That's how it gets this name!) The peak voltage is

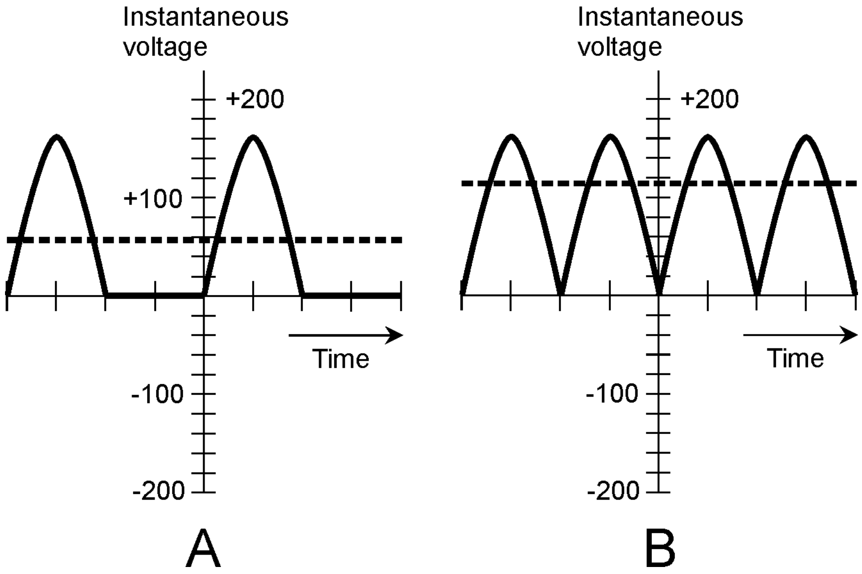

2-10 At A, half-wave rectification of common utility ac. At B, full-wave rectification of common utility ac. Effective voltages are shown by the dashed lines.

the maximum instantaneous voltage. Instantaneous voltage is never any greater than the peak voltage for any wave.

In Fig. 2-10B, the effective voltage is 0.707 times the peak voltage, just as is the case with ordinary ac. The direction of current flow, for many kinds of devices, doesn't make any difference. But in Fig. 2-10A, half of the wave has been lost. This cuts the effective value in half, so that it's only 0.354 times the peak value.

In household ac that appears in wall outlets for conventional appliances in the United States, the peak voltage is about 165 V; the effective value is 117 V. If full-wave rectification is used, the effective value is still 117 V. If half-wave rectification is used, the effective voltage is about 58.5 V.

Safety Considerations in Electrical Work

For our purposes, one rule applies concerning safety around electrical apparatus:

> *If you have any doubt about whether or not something is safe, leave it alone. Let a professional electrician work on it.*

Household voltage, normally about 117 V (but sometimes twice that for large appliances such as electric ranges and laundry machines), is more than sufficient to kill you if it appears across your chest cavity. Certain devices, such as automotive spark coils, can produce lethal currents even from the low voltage (12 to 14 V) in a car battery.

Consult the American Red Cross or your electrician concerning what kinds of circuits, procedures, and devices are safe and which aren't.

Magnetism

Electric currents and magnetic fields are closely related. Whenever an electric current flows—that is, when charge carriers move—a magnetic field accompanies the current. In a straight wire that carries electrical current, *magnetic lines of flux* surround the wire in circles, with the wire at the center, as shown in Fig. 2-11. (The lines of flux aren't physical objects; this is just a convenient way to rep-

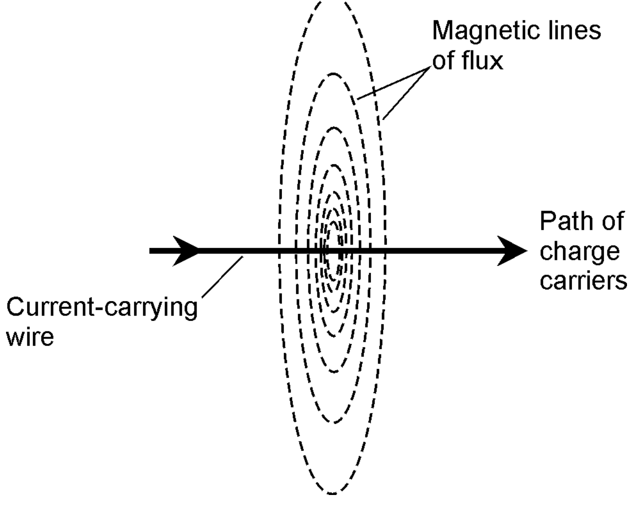

2-11 Magnetic flux lines around a straight, current-carrying wire. The arrows indicate current flow.

resent the magnetic field.) You'll sometimes hear or read about a certain number of flux lines per unit cross-sectional area, such as 100 lines per square centimeter. This is a relative way of talking about the intensity of the magnetic field.

Magnetic fields are produced when the atoms of certain materials align themselves. Iron is the most common metal that has this property. The atoms of iron in the core of the earth have become aligned to some extent; this is a complex interaction caused by the rotation of our planet and its motion with respect to the magnetic field of the sun. The magnetic field surrounding the earth is responsible for various effects, such as the concentration of charged particles that you see as the *aurora borealis* just after a solar eruption.

When a wire is coiled up, the resulting magnetic flux takes a shape similar to the flux field surrounding the earth, or the flux field around a bar magnet. Two well-defined *magnetic poles* develop, as shown in Fig. 2-12.

The intensity of a magnetic field can be greatly increased by placing a special core inside of a coil. The core should be of iron or some other material that can be readily magnetized. Such substances are called *ferromagnetic.* A core of this kind cannot actually increase the total quantity of magnetism in and around a coil, but it will cause the lines of flux to be much closer together inside the material. This is the principle by which an electromagnet works. It also makes possible the operation of electrical transformers for utility current.

Magnetic lines of flux are said to emerge from the magnetic north pole, and to run inward toward the magnetic south pole.

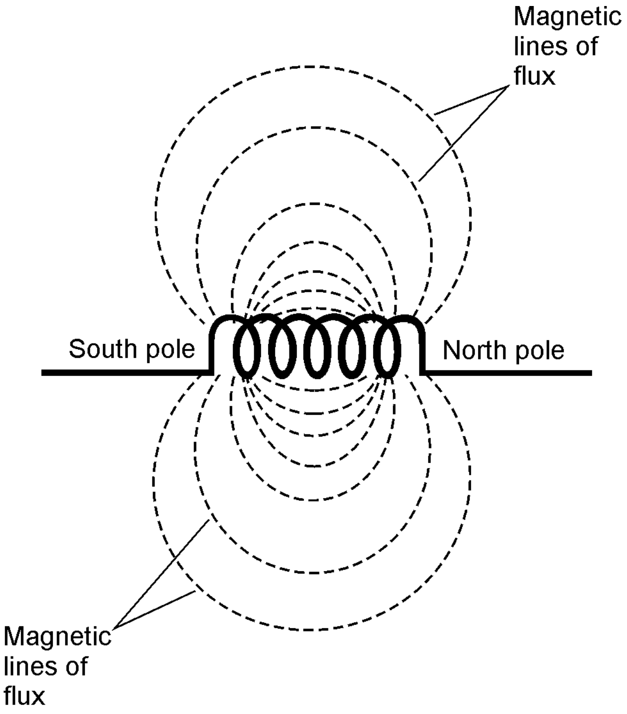

2-12 Magnetic flux lines around a current-carrying coil of wire. The flux lines converge at the magnetic poles.

Magnetic Units

The overall magnitude of a magnetic field is measured in units called *webers*, abbreviated Wb. One weber is mathematically equivalent to one volt-second (1 V · s). For weaker magnetic fields, a smaller unit, called the *maxwell* (Mx), is used. One maxwell is equal to 0.00000001 (one hundred-millionth) of a weber, or 0.01 microvolt-second (0.01 μV · s).

The *flux density* of a magnetic field is given in terms of webers or maxwells per square meter or per square centimeter. A flux density of one weber per square meter (1 Wb/m^2) is called one *tesla* (1 T). One *gauss* (1 G) is equal to 0.0001 T, or one maxwell per square centimeter (1 Mx/cm^2).

In general, as the electric current through a wire increases, so does the flux density near the wire. A coiled wire produces a greater flux density for a given current than a single, straight wire. And the more turns in the coil, the stronger the magnetic field will be.

Sometimes, magnetic field strength is specified in terms of *ampere-turns* (At). This is actually a unit of *magnetomotive force.* A one-turn wire loop, carrying 1 A of current, produces a field of 1 At. Doubling the number of turns, or the current, doubles the number of ampere-turns. Therefore, if you have 10 A flowing in a 10-turn coil, the magnetomotive force is 10 × 10, or 100 At. Or, if you have 100 mA flowing in a 100-turn coil, the magnetomotive force is 0.1 × 100, or 10 At. (Remember that 100 mA = 0.1 A.)

A less common unit of magnetomotive force is the *gilbert* (Gb). This unit is the equivalent of 0.796 At. Conversely, 1 At = 1.26 Gb.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct answers. The answers are listed in the back of this book.

1. A positive electric pole
   (a) has a deficiency of electrons.
   (b) has fewer electrons than the negative pole.
   (c) has an excess of electrons.
   (d) has more electrons than the negative pole.
2. An EMF of 1 V
   (a) cannot drive much current through a circuit.
   (b) represents a low resistance.
   (c) can sometimes produce a large current.
   (d) drops to zero in a short time.
3. A potentially lethal electric current is on the order of
   (a) 0.01 mA.
   (b) 0.1 mA.
   (c) 1 mA.
   (d) 0.1 A.

4. A current of 25 A is most likely drawn by
   (a) a flashlight bulb.
   (b) a typical household.
   (c) a utility power plant.
   (d) a small radio set.
5. A piece of wire has a conductance of 20 S. Its resistance is
   (a) 20 Ω.
   (b) 0.5 Ω.
   (c) 0.05 Ω.
   (d) 0.02 Ω.
6. A resistor has a value of 300 Ω. Its conductance is
   (a) 3.33 mS.
   (b) 33.3 mS.
   (c) 333 μS.
   (d) 0.333 S.
7. A span of wire 1 km long has a conductance of 0.6 S. What is the conductance of a span of this same wire that is 3 km long?
   (a) 1.8 S
   (b) 0.6 S
   (c) 0.2 S
   (d) More information is necessary to determine this.
8. Approximately how much current can a 2-kW generator reliably deliver at 117 V?
   (a) 17 mA
   (b) 234 mA
   (c) 17 A
   (d) 234 A
9. A circuit breaker is rated for 15 A at 117 V. Approximately how much power does this represent?
   (a) 1.76 kW
   (b) 1760 kW
   (c) 7.8 kW
   (d) 0.0078 kW
10. You are told that an air conditioner has cooled a room by 500 Btu over a certain period of time. What is this amount of energy in kWh?
    (a) 147 kWh
    (b) 14.7 kWh
    (c) 1.47 kWh
    (d) 0.147 kWh

11. Of the following energy units, the one most often used to define electrical energy is
    (a) the Btu.
    (b) the erg.
    (c) the foot-pound.
    (d) the kilowatt-hour.
12. The frequency of common household ac in the United States is
    (a) 60 Hz.
    (b) 120 Hz.
    (c) 50 Hz.
    (d) 100 Hz.
13. Half-wave rectification means that
    (a) half of the ac wave is inverted.
    (b) half of the ac wave is cut off.
    (c) the whole ac wave is inverted.
    (d) the effective voltage is half the peak voltage.
14. In the output of a half-wave rectifier,
    (a) half of the ac input wave is inverted.
    (b) the effective voltage is less than that of the ac input wave.
    (c) the effective voltage is the same as that of the ac input wave.
    (d) the effective voltage is more than that of the ac input wave.
15. In the output of a full-wave rectifier,
    (a) half of the ac input wave is inverted.
    (b) the effective voltage is less than that of the ac input wave.
    (c) the effective voltage is the same as that of the ac input wave.
    (d) the effective voltage is more than that of the ac input wave.
16. A low voltage, such as 12 V,
    (a) is never dangerous.
    (b) is always dangerous.
    (c) is dangerous if it is ac, but not if it is dc.
    (d) can be dangerous under certain conditions.
17. Which of the following units can represent magnetomotive force?
    (a) The volt-turn
    (b) The ampere-turn
    (c) The gauss
    (d) The gauss-turn

18. Which of the following units can represent magnetic flux density?
    (a) The volt-turn
    (b) The ampere-turn
    (c) The gauss
    (d) The gauss-turn
19. A ferromagnetic material
    (a) concentrates magnetic flux lines within itself.
    (b) increases the total magnetomotive force around a current-carrying wire.
    (c) causes an increase in the current in a wire.
    (d) increases the number of ampere-turns in a wire.
20. A coil has 500 turns and carries 75 mA of current. The magnetomotive force is
    (a) 37,500 At.
    (b) 375 At.
    (c) 37.5 At.
    (d) 3.75 At.

3
CHAPTER

Measuring Devices

NOW THAT YOU'RE FAMILIAR WITH THE PRIMARY UNITS COMMON IN ELECTRICITY AND ELECTRONICS, let's look at the instruments that are employed to measure these quantities. Many measuring devices work because electric and magnetic fields produce forces proportional to the intensity of the field. Such meters work by means of *electromagnetic deflection* or *electrostatic deflection.* Sometimes, electric current is measured by the extent of heat it produces in a resistance. Such meters work by *thermal heating* principles. Some meters have small motors whose speed depends on the measured quantity. The rotation rate, or the number of rotations in a given time, can be measured or counted. Still other kinds of meters tally up electronic pulses, sometimes in thousands, millions, or billions. These are *electronic counters.*

Electromagnetic Deflection

Early experimenters with electricity and magnetism noticed that an electric current produces a magnetic field. When a magnetic compass is placed near a wire carrying a direct electric current, the compass doesn't point toward magnetic north. The needle is displaced. The extent of the displacement depends on how close the compass is brought to the wire, and also on how much current the wire is carrying.

When this effect was first observed, scientists tried different arrangements to see how much the compass needle could be displaced, and how small a current could be detected. An attempt was made to obtain the greatest possible current-detecting sensitivity. Wrapping the wire in a coil around the compass resulted in a device that could indicate a tiny electric current (Fig. 3-1). This effect is known as *galvanism*, and the meter so devised was called a *galvanometer.* Once this device was made, the scientists saw that the extent of the needle displacement increased with increasing current. Then, the only challenge was to calibrate the galvanometer somehow, and to find a standard so a universal meter could be engineered.

You can make your own galvanometer. Buy a cheap compass, about 2 feet of insulated bell wire, and a 6-volt lantern battery. Set it up as shown in Fig. 3-1. Wrap the wire around the compass four or five times, and align the compass so that the needle points along the wire turns while the wire is disconnected from the battery. Connect one end of the wire to the negative (−) terminal of the bat-

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

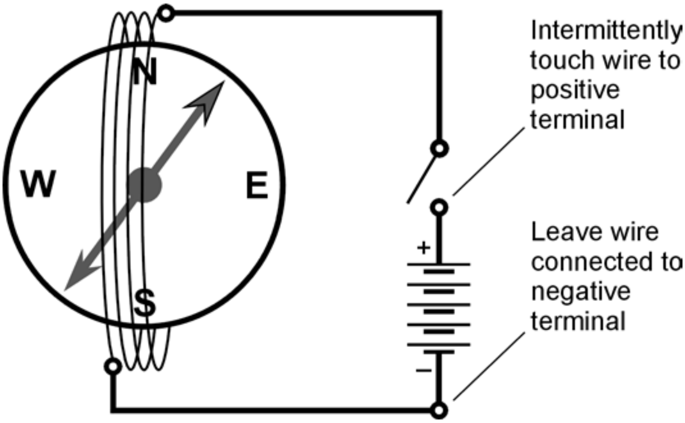
N
W
E
S
+
–
Intermittently touch wire to positive terminal
Leave wire connected to negative terminal

3-1 A simple galvanometer. The compass must lie flat.

tery. Touch the other end to the positive (+) terminal for a second or two, and watch the compass needle. Don't leave the wire connected to the battery for any length of time unless you want to drain the battery in a hurry.

You can buy a *resistor* and a *potentiometer* at a place like RadioShack, and set up an experiment that shows how galvanometers measure current. For a 6-V lantern battery, the fixed resistor should have a value of at least 330 Ω and should be rated for at least ¼ W. The potentiometer should have a maximum value of 10 kΩ. Connect the resistor and potentiometer in series between one end of the bell wire and one terminal of the battery, as shown in Fig. 3-2. The center contact of the potentiometer should be short-circuited to one of the end contacts, and the resulting two terminals used in the circuit.

When you adjust the potentiometer, the compass needle should deflect more or less, depending on the current through the wire. Early experimenters calibrated their meters by referring to the degrees scale around the perimeter of the compass.

3-2 A circuit for demonstrating how a galvanometer indicates relative current.

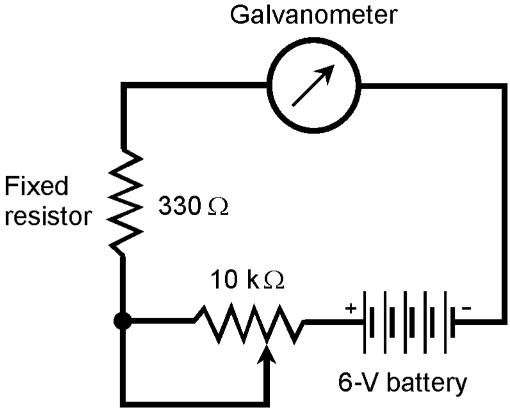
Galvanometer
Fixed resistor
330 Ω
10 kΩ
+
-
6-V battery

Electrostatic Deflection

Electric fields produce forces, just as magnetic fields do. You have noticed this when your hair feels like it's standing on end in very dry or cold weather. You've heard that people's hair really does stand straight out just before a lightning bolt hits nearby. (This is no myth!)

The most common device for demonstrating electrostatic forces is the *electroscope.* It consists of two foil leaves, attached to a conducting rod, and placed in a sealed container so that air currents cannot move the foil leaves (Fig. 3-3). When a charged object is brought near, or touched to, the contact at the top of the rod, the leaves stand apart from each other. This is because the two leaves become charged with like electric poles—either an excess or a deficiency of electrons—and like poles always repel. The extent to which the leaves stand apart depends on the amount of electric charge. It is difficult to measure this deflection and correlate it with charge quantity; electroscopes do not make very good meters. But variations on this theme can be employed, so that electrostatic forces can operate against tension springs or magnets, and in this way, *electrostatic meters* can be made.

An electrostatic meter can quantify alternating (or ac) electric charges as well as direct (or dc) charges. This gives electrostatic meters an advantage over electromagnetic meters such as the galvanometers. If you connect a source of ac to the coil of the galvanometer device in Fig. 3-1 or Fig. 3-2, the compass needle will not give a clear deflection; current in one direction pulls the meter needle one way, and current in the other direction pushes the needle the opposite way. But if a source of ac is connected to an electrostatic meter, the plates repel whether the charge is positive or negative at any given instant in time.

Most electroscopes aren't sensitive enough to show much deflection with ordinary 117-V utility ac. Don't try connecting 117 V to an electroscope anyway. It can present an electrocution hazard if you bring it out to points where you can easily come into physical contact with it.

An electrostatic meter has another property that is sometimes an advantage in electrical or electronic work. This is the fact that the device does not draw any current, except a tiny initial current needed to put a charge on the plates. Sometimes, an engineer or experimenter doesn't want a measuring device to draw current, because this affects the behavior of the circuit under test. Galvanometers, by contrast, always need some current to produce an indication.

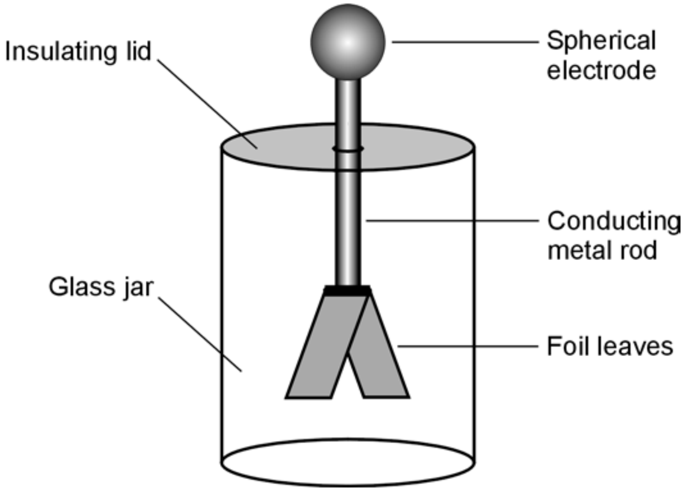

3-3 An electroscope can detect the presence of an electrostatic charge.

If you have access to a laboratory electroscope, try charging it up with a glass rod that has been rubbed against a cloth. When the rod is pulled away from the electroscope, the foil leaves remain standing apart. The charge just sits there! If the electroscope drew any current, the leaves would fall back together again, just as the galvanometer compass needle returns to magnetic north the instant you take the wire from the battery.

Thermal Heating

Another phenomenon, sometimes useful in the measurement of electric currents, is the fact that whenever current flows through a conductor having any resistance, that conductor is heated. All conductors have some resistance; none are perfect. The extent of this heating is proportional to the amount of current being carried by the wire.

By choosing the right metal or alloy, and by making the wire a certain length and diameter, and by employing a sensitive thermometer, and by putting the entire assembly inside a thermally insulating package, a *hot-wire meter* can be made. The hot-wire meter can measure ac as well as dc, because the current-heating phenomenon does not depend on the direction of current flow.

A variation of the hot-wire principle can be used to advantage by placing two different metals into contact with each other. If the right metals are chosen, the junction heats up when a current flows through it. This is called the *thermocouple principle.* As with the hot-wire meter, a thermometer can be used to measure the extent of the heating. But there is also another effect. A thermocouple, when it gets warm, generates dc. This dc can be measured with a galvanometer. This method is useful when it is necessary to have a fast meter *response time.*

The hot-wire and thermocouple effects are sometimes used to measure ac at high frequencies, in the range of hundreds of kilohertz up to tens of gigahertz.

Ammeters

A magnetic compass doesn't make a very convenient meter. It has to be lying flat, and the coil has to be aligned with the compass needle when there is no current. But of course, electrical and electronic devices aren't all oriented so as to be aligned with the north geomagnetic pole! But the external magnetic field doesn't have to come from the earth. It can be provided by a permanent magnet near or inside the meter. This supplies a stronger magnetic force than does the earth's magnetic field, and therefore makes it possible to make a meter that can detect much weaker currents. Such a meter can be turned in any direction, and its operation is not affected. The coil can be attached directly to the meter pointer, and suspended by means of a spring in the field of the magnet. This type of metering scheme, called the *D'Arsonval movement*, has been around since the earliest days of electricity, but it is still used in some metering devices today. The assembly is shown in Fig. 3-4. This is the basic principle of the *ammeter.*

A variation of the D'Arsonval movement can be obtained by attaching the meter needle to a permanent magnet, and winding the coil in a fixed form around the magnet. Current in the coil produces a magnetic field, and this in turn generates a force if the coil and magnet are aligned correctly with respect to each other. This works all right, but the mass of the permanent magnet causes a slower needle response. This type of meter is also more prone to *overshoot* than the true D'Arsonval movement; the inertia of the magnet's mass, once overcome by the magnetic force, causes the needle to fly past the actual point for the current reading, and then to wag back and forth a couple of times before coming to rest in the right place.

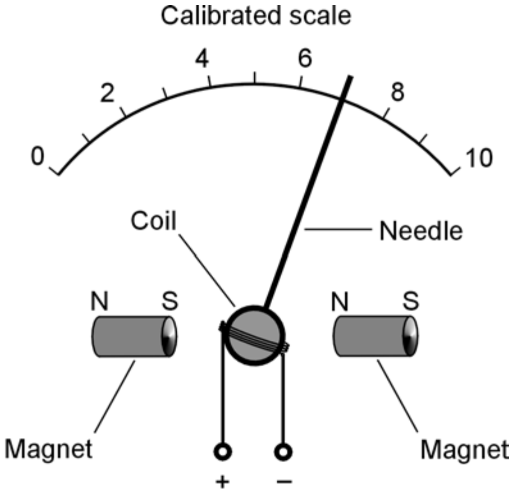

3-4 A functional drawing of a D'Arsonval meter movement (spring bearing not shown).

It is possible to use an *electromagnet* in place of the permanent magnet in the meter assembly. This electromagnet can be operated by the same current that flows in the coil attached to the meter needle. This gets rid of the need for a massive, permanent magnet inside the meter. It also eliminates the possibility that the meter sensitivity will change in case the strength of the permanent magnet deteriorates (such as might be caused by heat, or by severe mechanical vibration). The electromagnet can be either in series with, or in parallel with, the meter movement coil.

The sensitivity of the D'Arsonval-type meter, and of similar designs, depends on several factors. First is the strength of the permanent magnet (if the meter uses a permanent magnet). Second is the number of turns in the coil. The stronger the magnet, and the larger the number of turns in the coil, the less current is needed in order to produce a given magnetic force. If the meter is of the electromagnet type, the combined number of coil turns affects the sensitivity. Remember that the strength of a magnetomotive force is given in terms of ampere-turns. For a given current (number of amperes), the force increases in direct proportion to the number of coil turns. The more force in a meter, the greater the needle deflection for a given amount of current, and the smaller the current necessary to cause a certain amount of needle movement. The most sensitive ammeters can detect currents of just a microampere or two. The amount of current for *full-scale deflection* (the needle goes all the way up without banging against the stop pin) can be as little as about 50 μA in commonly available meters.

Sometimes, it is desirable to have an ammeter that will allow for a wide range of current measurements. The full-scale deflection of a meter assembly cannot easily be changed, because that would mean changing the number of coil turns and/or the strength of the magnet. But all ammeters have a certain amount of *internal resistance.* If a resistor, having the same internal resistance as the meter, is connected in parallel with the meter, the resistor will draw half the current. Then it will take twice the current through the assembly to deflect the meter to full scale, as compared with the meter alone. By choosing a resistor of just the right value, the full-scale deflection of an ammeter can be increased by a large factor, such as 10, or 100, or 1000. This resistor must be capable of carrying the current without burning up. It might have to draw practically all of the current flowing through the assembly, leaving the meter to carry only 1/10, or 1/100, or 1/1000 of the current. This is called a *shunt resistance* or *meter shunt* (Fig. 3-5). Meter shunts are used when it is necessary to measure very large currents, such as hundreds of amperes. They also allow microammeters or milliammeters to be used in a versatile *multimeter*, with many current ranges.

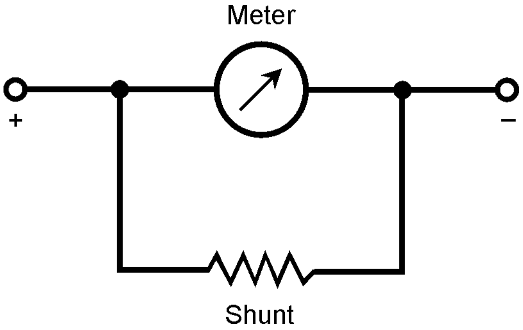

3-5 A resistor, called a *meter shunt*, can be connected across a current-detecting meter to reduce the sensitivity.

Voltmeters

Current, as we have seen, consists of a flow of charge carriers. Voltage, or electromotive force (EMF), or potential difference, is the "pressure" that makes current possible. Given a circuit whose resistance is constant, the current that flows in the circuit is directly proportional to the voltage placed across it. Early electrical experimenters recognized that an ammeter could be used to measure voltage, because an ammeter is a form of constant-resistance circuit. If you connect an ammeter directly across a source of voltage such as a battery, the meter needle deflects. In fact, a milliammeter needle will probably be "pinned" if you do this with it, and a microammeter might well be wrecked by the force of the needle striking the pin at the top of the scale. For this reason, you should never connect milliammeters or microammeters directly across voltage sources. An ammeter, perhaps with a range of 0 to 10 A, might not deflect to full scale if it is placed across a battery, but it's still a bad idea to do this, because it will rapidly drain the battery. Some batteries, such as automotive lead-acid cells, can explode under these conditions.

Ammeters have low internal resistance. They are designed that way deliberately. They are meant to be connected in series with other parts of a circuit, not right across a power supply. But if you place a large resistor in series with an ammeter, and then connect the ammeter across a battery or other type of power supply, you no longer have a short circuit. The ammeter will give an indication that is directly proportional to the voltage of the supply. The smaller the full-scale reading of the ammeter, the larger the resistance that is needed to get a meaningful indication on the meter. Using a microammeter and a very large value of resistance in series, a *voltmeter* can be devised that will draw only a little current from the source.

A voltmeter can be made to have various ranges for the full-scale reading, by switching different values of resistance in series with the microammeter (Fig. 3-6). The internal resistance of the meter is large because the values of the resistors are large. The greater the supply voltage, the larger the internal resistance of the meter, because the necessary series resistance increases as the voltage increases.

A voltmeter should have high internal resistance, and the higher the better! The reason for this is that you don't want the meter to draw much current from the power source. This current should go, as much as possible, toward operating whatever circuit is hooked up to the power supply, and not into getting a reading of the voltage. Also, you might not want, or need, to have the voltmeter constantly connected in the circuit; you might need the voltmeter for testing many different circuits. You don't want the behavior of a circuit to be affected the instant you connect the voltmeter to the supply. The less current a voltmeter draws, the less it affects the behavior of anything that is working from the power supply.

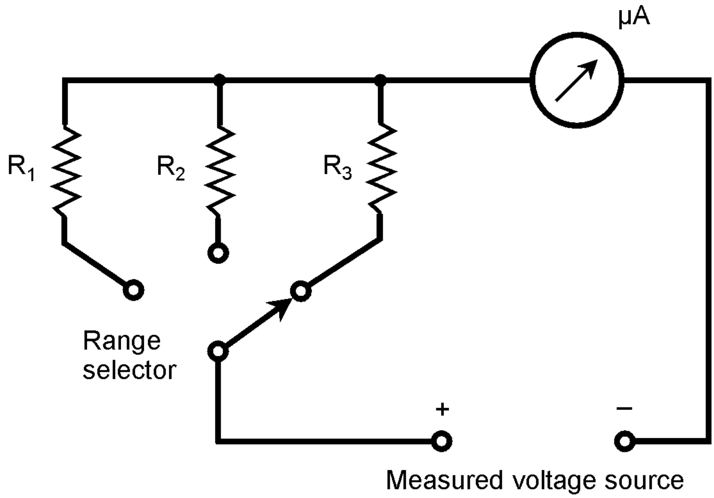
µA
R1
R2
R3
Range selector
+
−
Measured voltage source

3-6 A simple circuit using a microammeter (µA) to measure dc voltage.

A completely different type of voltmeter uses the effect of electrostatic deflection, rather than electromagnetic deflection. Remember that electric fields produce forces, just as do magnetic fields. Therefore, a pair of plates attract or repel each other if they are charged. The *electrostatic voltmeter* takes advantage of the attractive force between two plates having opposite electric charge, or having a large potential difference. Figure 3-7 is a simplified drawing of the mechanics of an electrostatic voltmeter. It draws almost no current from the power supply. The only thing between the plates is air, and air is a nearly perfect insulator. The electrostatic meter can indicate ac voltage as well as dc voltage. The construction tends to be fragile, however, and mechanical vibration can influence the reading.

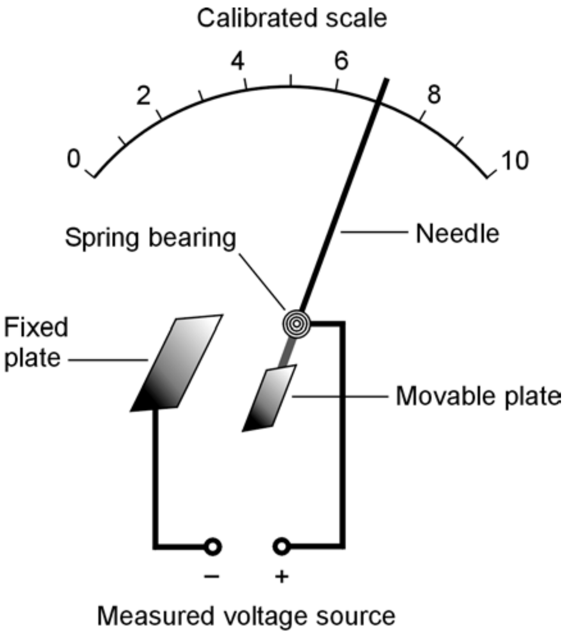
Calibrated scale
0
2
4
6
8
10
Spring bearing
Needle
Fixed plate
Movable plate
−
+
Measured voltage source

3-7 A functional drawing of an electrostatic voltmeter movement.

Ohmmeters

If all other factors are held constant, the current through a circuit depends on the resistance. This provides us with a means for measuring resistance. An *ohmmeter* can be constructed by placing a milliammeter or microammeter in series with a set of fixed, switchable resistances and a battery that provides a known, constant voltage (Fig. 3-8). By selecting the resistances appropriately, the meter gives indications in ohms over any desired range. The zero point on the milliammeter or microammeter is assigned the value of *infinity ohms*, meaning a perfect insulator. The full-scale value is set at a certain minimum, such as 1 Ω, 100 Ω, 1 kΩ, or 10 kΩ.

An ohmmeter must be calibrated at the factory where it is made, or in an electronics lab. A slight error in the values of the series resistors can cause gigantic errors in measured resistance. Therefore, precise *tolerances* are needed for these resistors. That means their values must actually be what the manufacturer claims they are, to within a fraction of 1 percent if possible. It is also necessary that the battery provide exactly the right voltage.

The scale of an ohmmeter is *nonlinear.* That means the graduations are not of the same width everywhere on the meter scale. The graduations tend to be squashed together toward the infinity end of the scale. Because of this, it is difficult to interpolate for high values of resistance unless the appropriate meter range is selected.

Engineers and technicians usually connect an ohmmeter in a circuit with the meter set for the highest resistance range first. Then they switch the range down until the meter needle is in a part of the scale that is easy to read. Finally, the reading is taken, and is multiplied (or divided) by the appropriate amount as indicated on the range switch. Figure 3-9 shows an ohmmeter reading. The meter itself indicates approximately 4.7, but the range switch says 1 kΩ. This indicates a resistance of about 4.7 kΩ, or 4700 Ω.

Ohmmeters give inaccurate readings if there is a voltage between the points where the meter is connected. This is because such a voltage either adds to, or subtracts from, the ohmmeter's own battery voltage. Sometimes, in this type of situation, an ohmmeter might tell you that a circuit has

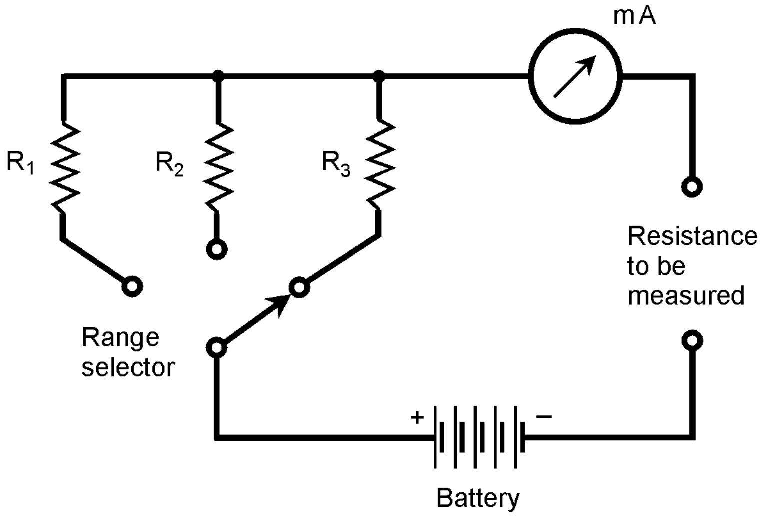

3-8 A circuit using a milliammeter (mA) to measure dc resistance.

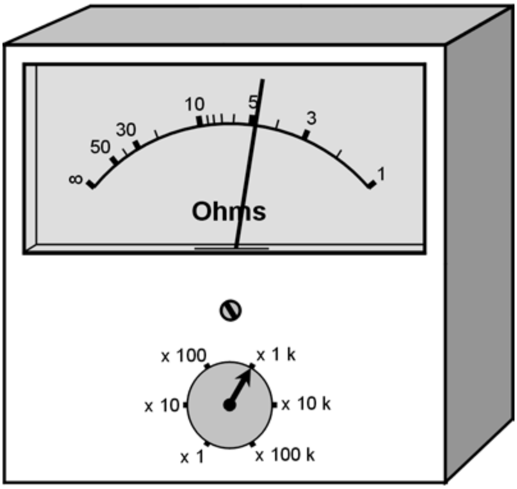

3-9 An example of an ohmmeter reading. This device shows about 4.7 × 1 kΩ = 4.7 kΩ = 4700 Ω.

"more than infinity" ohms! The needle will hit the pin at the left end of the scale. Therefore, when using an ohmmeter to measure resistance, you must always be sure that there is no voltage between the points under test. The best way to do this is to switch off the equipment in question.

Multimeters

In the electronics lab, a common piece of test equipment is the *multimeter*, in which different kinds of meters are combined into a single unit. The *volt-ohm-milliammeter* (VOM) is the most often used. As its name implies, it combines voltage, resistance, and current measuring capabilities. You should not have trouble envisioning how a single milliammeter can be used for measuring voltage, current, and resistance. The preceding discussions for measurements of these quantities have all included methods in which a current meter can be used to measure the intended quantity.

Commercially available multimeters have certain limits in the values they can measure. The maximum voltage is around 1000 V. The measurement of larger voltages requires special probes and heavily insulated wires, as well as other safety precautions. The maximum current that a common VOM can measure is about 1 A. The maximum measurable resistance is on the order of several megohms or tens of megohms. The lower limit of resistance indication is around 0.1 to 1 Ω.

FET Voltmeters

A good voltmeter disturbs the circuit under test as little as possible, and this requires that the meter have high internal resistance. Besides the electrostatic-type voltmeter, there is another way to get high internal resistance. This is to sample a tiny current, far too small for any meter to directly indicate, and then amplify this current so a conventional milliammeter or microammeter can display it. When a minuscule current is drawn from a circuit, the equivalent resistance is always extremely high.

The most effective way to accomplish voltage amplification, while making sure that the current drawn is exceedingly small, is to use a *field-effect transistor*, or FET. (Don't worry about how such amplifiers work right now; you'll learn all about that later in this book.) A voltmeter that uses a FET voltage amplifier to minimize the current drain is known as a *FET voltmeter* (FETVM). It has extremely high input resistance, along with good sensitivity and amplification.

Wattmeters

The measurement of electrical power requires that voltage and current both be measured simultaneously. Remember that in a dc circuit, the power (P) in watts is the product of the voltage (E) in volts and the current (I) in amperes. That is, $P = EI$. In fact, watts are sometimes called *volt-amperes* in dc circuits.

Do you think you can connect a voltmeter in parallel with a circuit, thereby getting a reading of the voltage across it, and also hook up an ammeter in series to get a reading of the current through the circuit, and then multiply volts times amperes to get watts consumed by the circuit? Well, you can. For most dc circuits, this is an excellent way to measure power, as shown in Fig. 3-10.

Sometimes, it's simpler yet. In many cases, the voltage from the power supply is constant and predictable. Utility power is a good example. The effective voltage is always very close to 117 V. Although it's ac, and not dc, power in most utility circuits can be measured in the same way as power is measured in dc circuits: by means of an ammeter connected in series with the circuit, and calibrated so that the multiplication (times 117) has already been done. Then, rather than 1 A, the meter will show a reading of 117 W, because $P = EI = 117 \times 1 = 117$ W. If the meter reading is 300 W, the current is $I = P/E = 300/117 = 2.56$ A. An electric iron might consume 1000 W, or a current of 1000/117 = 8.55 A. A large heating unit might gobble up 2000 W, requiring a current of 2000/117 = 17.1 A. You should not be surprised if this blows a fuse or trips a circuit breaker, because these devices are often rated for 15 A.

Specialized wattmeters are necessary for the measurement of radio-frequency (RF) power, or for peak audio power in a high-fidelity amplifier, or for certain other specialized applications. But almost all of these meters, whatever the associated circuitry, use simple ammeters, milliammeters, or microammeters as their indicating devices.

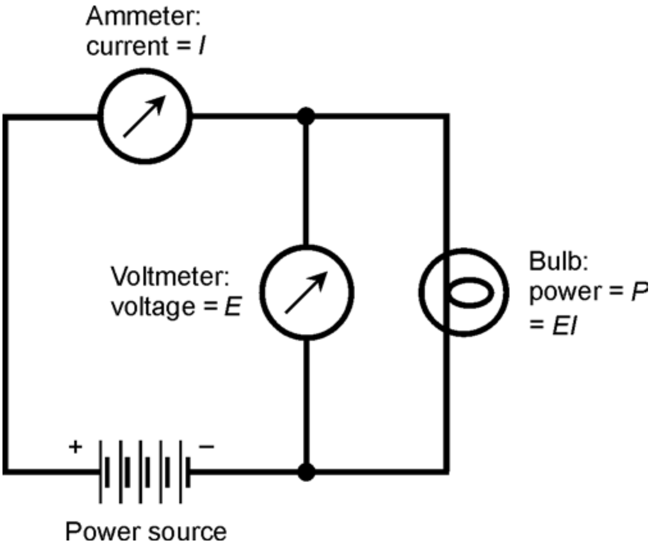

3-10 In a dc circuit, power can be measured with a voltmeter and an ammeter, connected as shown here.

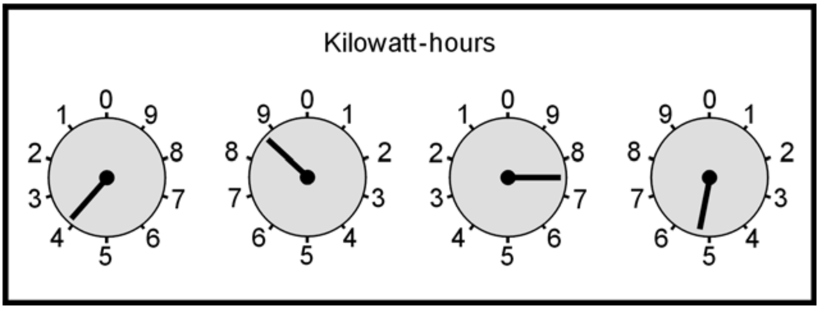

3-11 A utility meter with four rotary analog dials. In this example, the reading is a little more than 3875 kWh.

Watt-Hour Meters

Electrical energy, as you now know, is measured in watt-hours or kilowatt-hours (kWh). Not surprisingly, a metering device that indicates energy in these units is called a *watt-hour meter* or a *kilowatt-hour meter.*

The most often used means of measuring electrical energy is by using a small electric motor, the speed of which depends on the current, and thereby on the power at a constant voltage. The number of turns of the motor shaft, in a given length of time, is directly proportional to the number of kilowatt-hours consumed. The motor is placed at the point where the utility wires enter the building. This is usually at a point where the voltage is 234 V. At this point the circuit is split into some circuits with 234 V (for heavy-duty appliances such as the oven, washer, and dryer) and general household circuits at 117 V (for smaller appliances such as lamps, clock radios, and television sets).

If you've observed a kilowatt-hour meter, you have seen a disk spinning, sometimes fast, other times slowly. Its speed depends on the power being used at any given time. The total number of turns of this little disk, every month, determines the size of the bill you will get, as a function also, of course, of the cost per kilowatt-hour.

Kilowatt-hour meters count the number of disk turns by means of geared rotary drums or pointers. The drum-type meter gives a direct digital readout. The pointer type has several scales calibrated from 0 to 9 in circles, some going clockwise and others going counterclockwise. Reading a pointer-type utility meter is a little tricky, because you must think in whatever direction (clockwise or counterclockwise) the scale goes. An example of a pointer-type utility meter is illustrated in Fig. 3-11. Read from left to right. For each meter scale, take down the number that the pointer has most recently passed. Write down the rest as you go. The meter shown in the figure reads a little more than 3875 kWh.

Digital Readout Meters

Increasingly, metering devices are being designed so that they provide a direct readout. The number on the meter is the indication. It's that simple. Such a meter is called a *digital meter.*

The main advantage of a digital meter is the fact that it's easy for anybody to read, and there is no chance for interpolation errors. This is ideal for utility meters, clocks, and some kinds of ammeters, voltmeters, and wattmeters. It works well when the value of the quantity does not change often or fast.

There are some situations in which a digital meter is a disadvantage. One good example is the signal-strength indicator in a radio receiver. This meter bounces up and down as signals fade, or as you tune the radio, or sometimes even as the signal modulates. A digital meter will show nothing but a constantly changing, meaningless set of numerals. Digital meters require a certain length of time to lock in to the current, voltage, power, or other quantity being measured. If this quantity never settles at any one value for a long enough time, the meter can never lock in.

Meters with a scale and pointer are known as *analog meters.* Their main advantages are that they allow interpolation, they give the operator a sense of the quantity relative to other possible values, and they follow along when a quantity changes. Some engineers and technicians prefer analog metering, even in situations where digital meters would work just as well.

One potential hang-up with digital meters is being certain of where the decimal point goes. If you're off by one decimal place, the error will be by a factor of 10. Also, you need to be sure you know what the units are. For example, a frequency indicator might be reading out in megahertz, and you might forget and think it is giving you a reading in kilohertz. That's a mistake by a factor of 1000! Of course, this latter type of error can happen with analog meters, too.

Frequency Counters

The measurement of energy used by your home is an application to which digital metering is well suited. A digital kilowatt-hour meter is easier to read than the pointer-type meter. When measuring frequencies of radio signals, digital metering is not only more convenient, but far more accurate.

A *frequency counter* measures the frequency of an ac wave by actually counting pulses, in a manner similar to the way the utility meter counts the number of turns of a motor. But the frequency counter works electronically, without any moving parts. It can keep track of thousands, millions, or billions of pulses per second, and it shows the rate on a digital display that is as easy to read as a digital watch.

The accuracy of the frequency counter is a function of the *lock-in time.* Lock-in is usually done in 0.1 second, 1 second, or 10 seconds. Increasing the lock-in time by a factor of 10 will cause the accuracy to increase by one additional digit. Modern frequency counters are good to six, seven, or eight digits; sophisticated lab devices can show frequency to nine or ten digits.

Other Meter Types

Here are a few of the less common types of meters that you will occasionally encounter in electrical and electronics applications.

VU and Decibel Meters

In high-fidelity equipment, especially the more sophisticated amplifiers ("amps"), *loudness meters* are sometimes used. These are calibrated in *decibels*, a unit that you will often have to use, and interpret, in reference to electronic signal levels. A decibel is an increase or decrease in sound or signal level that you can just barely detect, if you are expecting the change.

Audio loudness is given in *volume units* (VU), and the meter that indicates it is called a *VU meter.* The typical VU meter has a zero marker with a red line to the right and a black line to the left, and is calibrated in decibels (dB) below the zero marker and volume units above it (Fig. 3-12). The meter might also be calibrated in *watts rms*, an expression for audio power. As music is played

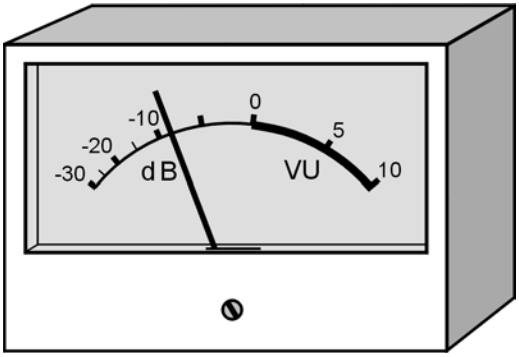

3-12 A VU (volume-unit) meter. The heavy portion of the scale (to the right of 0) is usually red, indicating the risk of audio distortion.

through the system, or as a voice comes over it, the VU meter needle kicks up. The amplifier volume should be kept down so that the meter doesn't go past the zero mark and into the red range. If the meter does kick up into the red scale, it means that distortion is taking place within the amplifier circuit.

Sound level in general can be measured by means of a *sound-level meter*, calibrated in decibels (dB) and connected to the output of a precision amplifier with a microphone of known sensitivity (Fig. 3-13). Have you read that a vacuum cleaner will produce "80 dB" of sound, and a large truck going by will subject your ears to "90 dB"? These figures are determined by a sound-level meter, and are defined with respect to the *threshold of hearing*, which is the faintest sound that a person with good ears can hear.

Light Meters

The intensity of visible light is measured by means of a *light meter* or *illumination meter.* It is tempting to suppose that it's easy to make this kind of meter by connecting a milliammeter to a solar (photovoltaic) cell. As things work out, this is a good way to construct an inexpensive light meter (Fig. 3-14). More sophisticated devices use dc amplifiers, similar to the type found in a FETVM, to enhance sensitivity and to allow for several different ranges of readings.

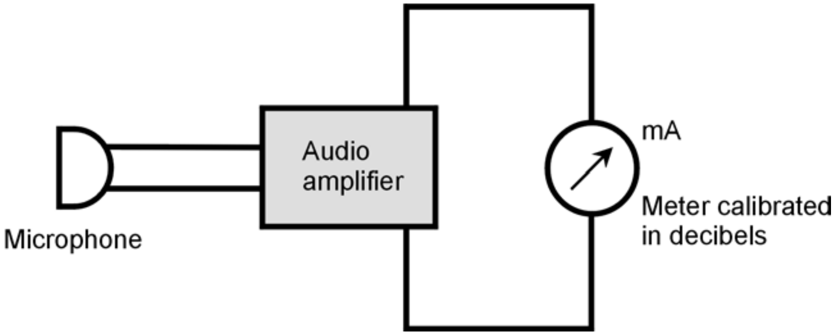

3-13 A meter for measuring sound levels. The output of the audio amplifier is rectified to produce dc that the meter can detect.

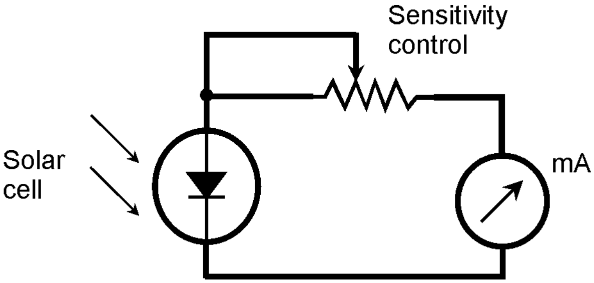

3-14 A simple light meter. A microammeter can be substituted for the milliammeter if greater sensitivity is required.

One problem with this design is that solar cells are not sensitive to light at exactly the same wavelengths as human eyes. This can be overcome by placing a colored filter in front of the solar cell, so that the solar cell becomes sensitive to the same wavelengths, in the same proportions, as human eyes. Another problem is calibrating the meter. This must usually be done at the factory, in standard illumination units such as *lumens* or *candela.*

With appropriate modification, meters such as the one in Fig. 3-14 can be used to measure *infrared* (IR) or *ultraviolet* (UV) intensity. Various specialized photovoltaic cells have peak sensitivity at nonvisible wavelengths, including IR and UV.

Pen Recorders

A meter movement can be equipped with a marking device to keep a graphic record of the level of some quantity with respect to time. Such a device is called a *pen recorder.* The paper, with a calibrated scale, is taped to a rotating drum. The drum, driven by a clock motor, turns at a slow rate, such as one revolution per hour or one revolution in 24 hours. A simplified drawing of a pen recorder is shown in Fig. 3-15.

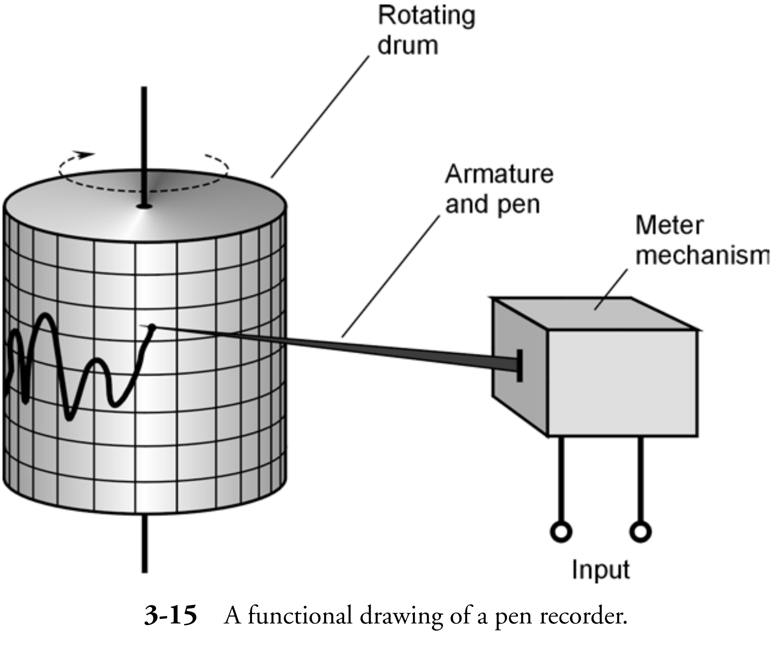

3-15 A functional drawing of a pen recorder.

A device of this kind, along with a wattmeter, can be employed to get a reading of the power consumed by your household at various times during the day. In this way you can find out when you use the most power, and at what particular times you might be using too much.

Oscilloscopes

Another graphic metering device is the *oscilloscope.* This measures and records quantities that vary rapidly, at rates of hundreds, thousands, or millions of times per second. It creates a "graph" by throwing a beam of electrons at a phosphor screen. A *cathode-ray tube*, similar to the kind in a television set, is employed. Some oscilloscopes have electronic conversion circuits that allow for the use of a solid-state *liquid crystal display* (LCD).

Oscilloscopes are useful for observing and analyzing the shapes of signal waveforms, and also for measuring peak signal levels (rather than just the effective levels). An oscilloscope can also be used to approximately measure the frequency of a waveform. The horizontal scale of an oscilloscope shows time, and the vertical scale shows the instantaneous signal voltage. An oscilloscope can indirectly measure power or current, by using a known value of resistance across the input terminals.

Technicians and engineers develop a sense of what a signal waveform should look like, and then they can often tell, by observing the oscilloscope display, whether or not the circuit under test is behaving the way it should. This is a subjective measurement, because it is qualitative as well as quantitative.

Bar-Graph Meters

A cheap, simple kind of meter can be made using a string of light-emitting diodes (LEDs) or an LCD along with a digital scale to indicate approximate levels of current, voltage, or power. This type of meter, like a digital meter, has no moving parts to break. To some extent, it offers the relative-reading feeling you get with an analog meter. Figure 3-16 is an example of a bar-graph meter that is used to show the power output, in kilowatts, for a radio transmitter. This meter can follow along quite well with rapid fluctuations in the reading. In this example, the meter indicates about 0.8 kW, or 800 W.

The chief drawback of the bar-graph meter is that it isn't very accurate. For this reason it is not generally used in laboratory testing. In addition, the LED or LCD devices sometimes flicker when the level is between two values given by the bars. This creates an illusion of circuit instability. With bright LEDs, it can also be quite distracting.

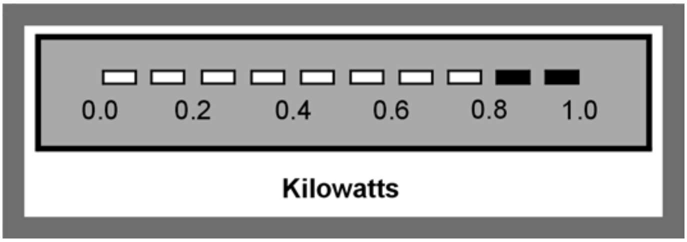

3-16 A bar-graph meter. In this case, the indication is about 80 percent of full-scale, representing 0.8 kW, or 800 W.

Quiz

Refer to the text in this chapter if necessary. A good score is 18 out of 20 correct. Answers are in the back of the book.

1. The attraction or repulsion between two electrically charged objects is called
   (a) electromagnetic deflection.
   (b) electrostatic force.
   (c) magnetic force.
   (d) electroscopic force.
2. The change in the direction of a compass needle, when a current-carrying wire is brought near, is called
   (a) electromagnetic deflection.
   (b) electrostatic force.
   (c) magnetic force.
   (d) electroscopic force.
3. Suppose a certain current in a galvanometer causes the compass needle to deflect by 20 degrees, and then this current is doubled while the polarity stays the same. The angle of the needle deflection will
   (a) decrease.
   (b) stay the same.
   (c) increase.
   (d) reverse direction.
4. One important advantage of an electrostatic meter is the fact that
   (a) it measures very small currents.
   (b) it can handle large currents.
   (c) it can detect and indicate ac voltages as well as dc voltages.
   (d) it draws a large current from a power supply.
5. A thermocouple
   (a) gets warm when dc flows through it.
   (b) is a thin, straight, special wire.
   (c) generates dc when exposed to visible light.
   (d) generates ac when heated.
6. An important advantage of an electromagnet-type meter over a permanent-magnet meter is the fact that
   (a) the electromagnet meter costs much less.
   (b) the electromagnet meter need not be aligned with the earth's magnetic field.
   (c) the permanent-magnet meter has a more sluggish coil.
   (d) the electromagnet meter is more rugged.

7. Ammeter shunts are useful because
   (a) they increase meter sensitivity.
   (b) they make a meter more physically rugged.
   (c) they allow for measurement of large currents.
   (d) they prevent overheating of the meter movement.

8. Voltmeters should generally have
   (a) high internal resistance.
   (b) low internal resistance.
   (c) the greatest possible sensitivity.
   (d) the ability to withstand large currents.

9. In order to measure the power-supply voltage that is applied to an electrical circuit, a voltmeter should be placed
   (a) in series with the circuit that works from the supply.
   (b) between the negative pole of the supply and the circuit working from the supply.
   (c) between the positive pole of the supply and the circuit working from the supply.
   (d) in parallel with the circuit that works from the supply.

10. Which of the following will *not* normally cause a *large* error in an ohmmeter reading?
    (a) A small voltage between points under test
    (b) A slight change in switchable internal resistance
    (c) A small change in the resistance to be measured
    (d) A slight error in the range switch position

11. The ohmmeter in Fig. 3-17 shows a reading of approximately
    (a) 34,000 Ω.
    (b) 3.4 kΩ.
    (c) 340 Ω.
    (d) 34 Ω.

12. The main advantage of a FETVM over a conventional voltmeter is the fact that the FETVM
    (a) can measure lower voltages.
    (b) draws less current from the circuit under test.
    (c) can withstand higher voltages safely.
    (d) is sensitive to ac voltage as well as to dc voltage.

13. Which of the following is *not* a function of a fuse?
    (a) To ensure there is enough current available for an appliance to work right
    (b) To make it impossible to use appliances that are too large for a given circuit
    (c) To limit the amount of power that a device can draw from the electrical circuit
    (d) To make sure the current drawn by an appliance cannot exceed a certain limit

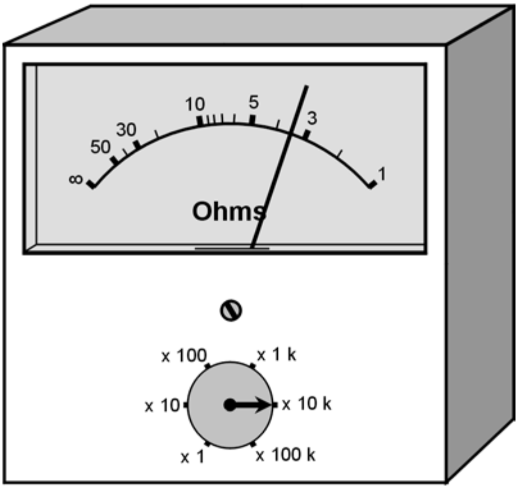

3-17 Illustration for Quiz Question 11.

14. A utility meter's motor speed depends directly on
    (a) the number of ampere-hours being used at the time.
    (b) the number of watt-hours being used at the time.
    (c) the number of watts being used at the time.
    (d) the number of kilowatt-hours being used at the time.
15. A utility meter's readout indicates
    (a) voltage.
    (b) power.
    (c) current.
    (d) energy.
16. A typical frequency counter
    (a) has an analog readout.
    (b) is accurate to six digits or more.
    (c) works by indirectly measuring current.
    (d) works by indirectly measuring voltage.
17. A VU meter is *never* used to get a general indication of
    (a) sound intensity.
    (b) decibels.
    (c) power in an audio amplifier.
    (d) visible light intensity.

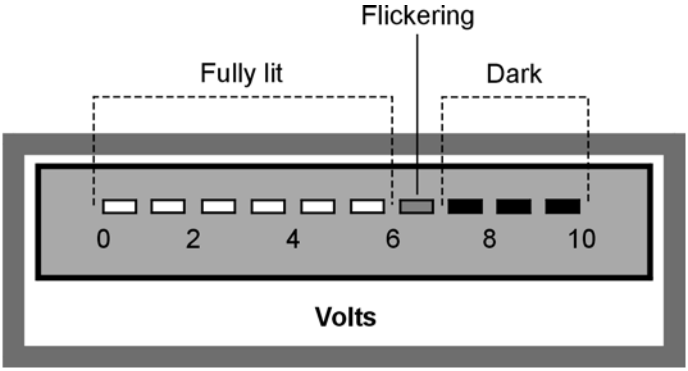

3-18 Illustration for Quiz Question 20.

18. The meter movement in an illumination meter directly measures
    - (a) current.
    - (b) voltage.
    - (c) power.
    - (d) energy.
19. An oscilloscope *cannot* be used to indicate
    - (a) frequency.
    - (b) wave shape.
    - (c) energy.
    - (d) peak signal voltage.
20. What voltage would be expected to produce the reading on the bar-graph meter shown in Fig. 3-18?
    - (a) 6.0 V
    - (b) 6.5 V
    - (c) 7.0 V
    - (d) There is no way to tell because the meter, as shown, is malfunctioning.

4
CHAPTER

Direct-Current Circuit Basics

YOU'VE ALREADY SEEN SOME SIMPLE ELECTRICAL CIRCUIT DIAGRAMS. IN THIS CHAPTER, YOU'LL GET more acquainted with this type of diagram. You'll also learn more about how current, voltage, resistance, and power are related in dc and low-frequency ac circuits.

Schematic Symbols

In this course, the idea is to familiarize you with schematic symbols by getting you to read and use them in action. But right now, why not check out Appendix B, which is a comprehensive table of symbols? Then refer to it frequently in the future, especially when you see a symbol you don't remember or recognize.

The simplest schematic symbol is the one representing a wire or *electrical conductor*: a straight, solid line. Sometimes, dashed lines are used to represent conductors, but usually, dashed lines are drawn to partition diagrams into constituent circuits, or to indicate that certain components interact with each other or operate in step with each other. Conductor lines are almost always drawn either horizontally across or vertically up and down the page. This keeps the diagram neat and easy to read.

When two conductor lines cross, they aren't connected at the crossing point unless a heavy black dot is placed where the two lines meet. The dot should always be clearly visible wherever conductors are to be connected, no matter how many of them meet at the junction. A *resistor* is indicated by a zigzag. A variable resistor, or *potentiometer*, is indicated by a zigzag with an arrow through it, or by a zigzag with an arrow pointing at it. These symbols are shown in Fig. 4-1.

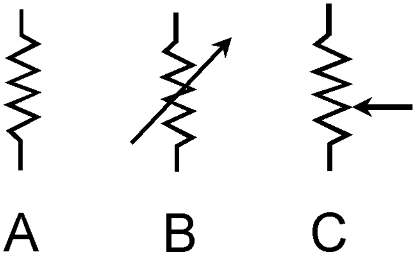

4-1 Schematic symbols for a fixed resistor (A), a two-terminal variable resistor (B), and a three-terminal potentiometer (C).

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

An *electrochemical cell* (such as a common dime-store battery) is shown by two parallel lines, one longer than the other. The longer line represents the plus terminal. A true *battery*, which is a combination of two or more cells in series, is indicated by several parallel lines, alternately long and short. It's not necessary to use more than four lines to represent a battery, although you'll often see 6, 8, 10, or even 12 lines. Symbols for a cell and a battery are shown in Fig. 4-2.

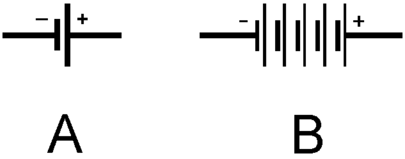

4-2 Schematic symbols for an electrochemical cell (A) and an electrochemical battery (B).

Meters are portrayed as circles. Sometimes the circle has an arrow inside it, and the meter type, such as mA (milliammeter) or V (voltmeter) is written alongside the circle, as shown in Fig. 4-3A. Sometimes the meter type is indicated inside the circle, and there is no arrow (Fig. 4-3B). It doesn't matter which way you draw them, as long as you're consistent throughout a schematic diagram.

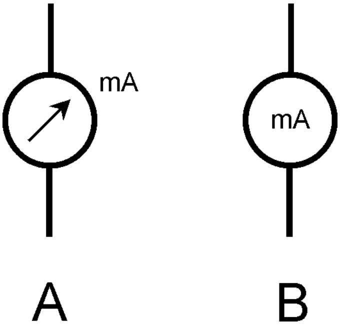

4-3 Meter symbols can have the designator either outside the circle (A) or inside (B). In this case, both symbols represent a milliammeter (mA).

Some other common symbols include the *incandescent lamp*, the *capacitor*, the *air-core coil*, the *iron-core coil*, the *chassis ground*, the *earth ground*, the *ac source*, the set of *terminals*, and the *black box* (general component or device), a rectangle with the designator written inside. These are shown in Fig. 4-4.

Schematic and Wiring Diagrams

Look back through the earlier chapters of this book and observe the electrical diagrams. These are all simple examples of how professionals would draw *schematic diagrams.* In a schematic diagram, the interconnection of the components is shown, but the actual values of the components are not necessarily indicated. You might see a diagram of a two-transistor audio amplifier, for example, with resistors and capacitors and coils and transistors, but without any data concerning the values or ratings of the components. This is a schematic diagram, but not a true *wiring diagram.* It gives the *scheme* for the circuit, but you can't *wire* the circuit and make it work, because there isn't enough information.

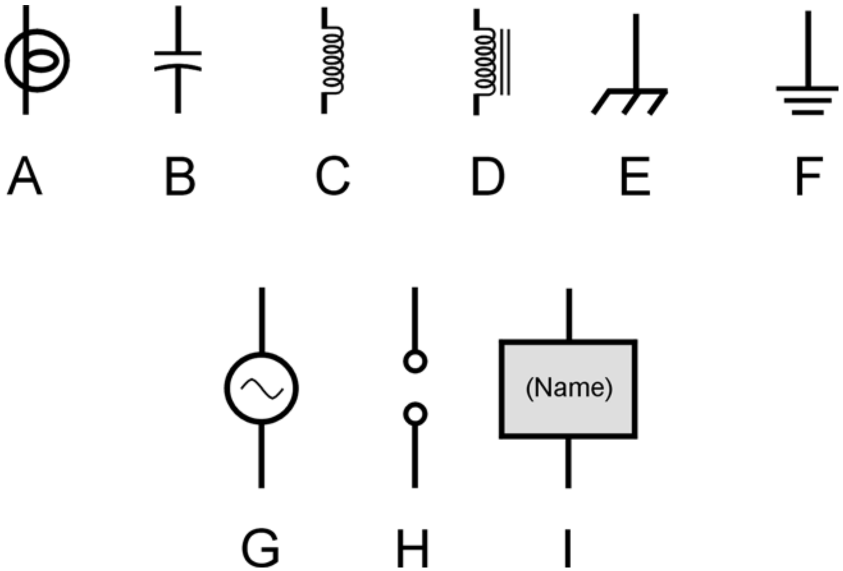

4-4 Schematic symbols for incandescent lamp (A), fixed capacitor (B), fixed inductor with air core (C), fixed inductor with laminated-iron core (D), chassis ground (E), earth ground (F), signal generator or source of alternating current (G), pair of terminals (H), and specialized component or device (I).

Suppose you want to build the circuit. You go to an electronics store to get the parts. What values of resistors should you buy? How about capacitors? What type of transistor will work best? Do you need to wind the coils yourself, or can you get ready-made coils? Are there test points or other special terminals that should be installed for the benefit of the technicians who might have to repair the amplifier? How many watts should the potentiometers be able to handle? All these things are indicated in a wiring diagram. You might have seen this kind of diagram in the back of the instruction manual for a hi-fi amplifier, a stereo tuner, or a television set. Wiring diagrams are especially useful when you want to build, modify, or repair an electronic device.

Voltage/Current/Resistance Circuits

Most dc circuits can be boiled down to three major components: a voltage source, a set of conductors, and a resistance. This is shown in Fig. 4-5. The voltage or EMF source is *E;* the current in the conductor is *I;* the resistance is *R.*

You already know that there is a relationship among these three quantities. If one of them changes, then one or both of the others will change. If you make the resistance smaller, the current will get larger. If you reduce the applied voltage, the current will also decrease. If the current in the circuit increases, the voltage across the resistor will increase. There is a simple arithmetic relationship among these three quantities.

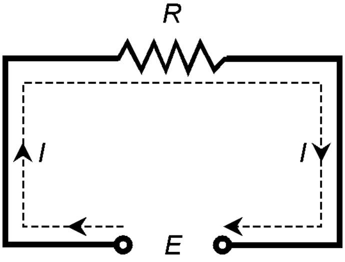

4-5 The basic elements of a dc circuit. The voltage is *E*, the current is *I*, and the resistance is *R*.

Ohm's Law

The interdependence among current, voltage, and resistance in dc circuits is called *Ohm's Law*, named after the scientist who supposedly first quantified it. Three formulas denote this law:

$$E = IR$$
$$I = E/R$$
$$R = E/I$$

You need only remember the first of these formulas in order to derive the others. The easiest way to remember it is to learn the abbreviations *E* for voltage, *I* for current, and *R* for resistance, and then remember that they appear in alphabetical order with the equal sign after the *E*. Sometimes the three symbols are arranged in the so-called Ohm's Law triangle, shown in Fig. 4-6. To find the value of a quantity, cover it up and read the positions of the others.

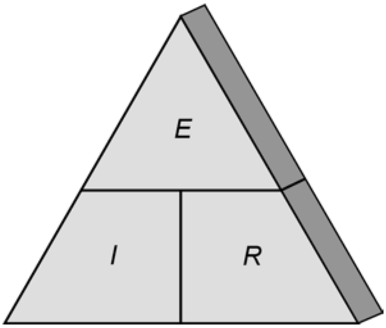

4-6 The Ohm's Law triangle. The voltage is *E*, the current is *I*, and the resistance is *R*. These quantities are expressed in volts, amperes, and ohms, respectively.

Remember that you must use units of volts, amperes, and ohms for the Ohm's Law formulas to yield a meaningful result! If you use, say, volts and microamperes to calculate a resistance, you cannot be sure of the units you'll end up with when you derive the final result. If the initial quantities are given in units other than volts, amperes, and ohms, convert to these units, and then calculate. After that, you can convert the calculated current, voltage, or resistance value to whatever size unit you want. For example, if you get 13,500,000 Ω as a calculated resistance, you might prefer to say that it's 13.5 MΩ.

Current Calculations

The first way to use Ohm's Law is to determine current in dc circuits. In order to find the current, you must know the voltage and the resistance, or be able to deduce them. Refer to the schematic diagram of Fig. 4-7. It consists of a dc voltage source, a voltmeter, some wire, an ammeter, and a calibrated, wide-range potentiometer.

Problem 4-1

Suppose that the dc generator in Fig. 4-7 produces 10 V and the potentiometer is set to a value of 10 Ω. What is the current?

This is solved by the formula $I = E/R$. Plug in the values for E and R; they are both 10, because the units are given in volts and ohms. Then $I = 10/10 = 1.0$ A.

Problem 4-2

Imagine that dc generator in Fig. 4-7 produces 100 V and the potentiometer is set to 10 kΩ. What is the current?

First, convert the resistance to ohms: 10 kΩ = 10,000 Ω. Then plug the values in: $I = 100/10{,}000 = 0.01$ A. You might prefer to express this as 10 mA.

Problem 4-3

Suppose that dc generator in Fig. 4-7 is set to provide 88.5 V, and the potentiometer is set to 477 MΩ. What is the current?

This problem involves numbers that aren't exactly round, and one of them is huge. But you can use a calculator. First, change the resistance value to ohms, so you get 477,000,000 Ω. Then plug into the Ohm's Law formula: $I = E/R = 88.5 / 477{,}000{,}000 = 0.000000186$ A. It is more reasonable to express this as 0.186 μA or 186 nA.

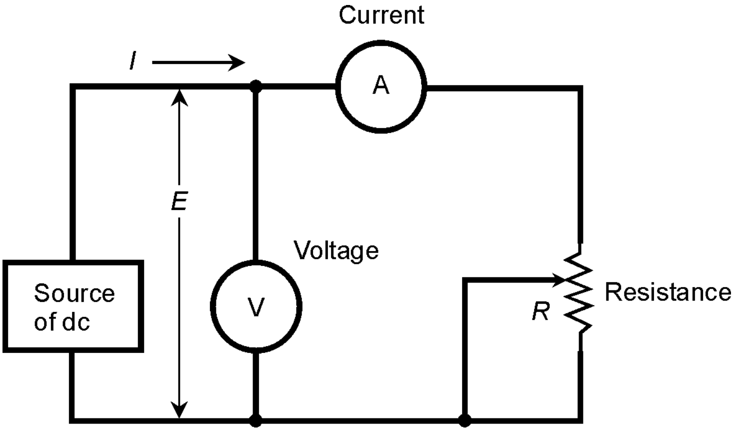

4-7 A circuit for working Ohm's Law problems.

Voltage Calculations

The second application of Ohm's Law is to find unknown dc voltages when the current and the resistance are known. Let's work out some problems of this kind.

Problem 4-4

Suppose the potentiometer in Fig. 4-7 is set to 100 Ω, and the measured current is 10 mA. What is the dc voltage?

Use the formula $E = IR$. First, convert the current to amperes: 10 mA = 0.01 A. Then multiply: $E = 0.01 \times 100 = 1.0$ V. That's a little less than the voltage produced by a flashlight cell.

Problem 4-5

Adjust the potentiometer in Fig. 4-7 to a value of 157 kΩ, and suppose the current reading is 17.0 mA. What is the voltage of the source?

You must convert both the resistance and the current values to their proper units. A resistance of 157 kΩ is 157,000 Ω, and a current of 17.0 mA is 0.0170 A. Then $E = IR = 0.017 \times 157{,}000 = 2669$ V = 2.669 kV. You should round this off to 2.67 kV. This is a dangerously high voltage.

Problem 4-6

Suppose you set the potentiometer in Fig. 4-7 so that the meter reads 1.445 A, and you observe that the potentiometer scale shows 99 Ω. What is the voltage?

These units are both in their proper form. Therefore, you can plug them right in and use your calculator: $E = IR = 1.445 \times 99 = 143.055$ V. This can and should be rounded off—but to what extent? This is a good time to state an important rule that should be followed in all technical calculations.

The Rule of Significant Figures

Competent engineers and scientists go by the *rule of significant figures*, also called the *rule of significant digits.* After completing a calculation, you should always round the answer off to the *least* number of digits given in the input data numbers.

If you follow this rule in Problem 4-6, you must round off the answer to two significant digits, getting 140 V, because the resistance (99 Ω) is only specified to that level of accuracy. If the resistance were given as 99.0 Ω, then you would round off the answer to 143 V. If the resistance were given as 99.00 Ω, then you could state the answer as 143.1 V. However, any further precision in the resistance value would not entitle you to go to any more digits in your answer, unless the current were specified to more than four significant figures.

This rule takes some getting used to if you haven't known about it or practiced it before. But after a while, it will become a habit.

Resistance Calculations

Ohms' Law can be used to find a resistance between two points in a dc circuit when the voltage and the current are known.

Problem 4-7

If the voltmeter in Fig. 4-7 reads 24 V and the ammeter shows 3.0 A, what is the resistance of the potentiometer?

Use the formula $R = E/I$, and plug in the values directly, because they are expressed in volts and amperes: $R = 24/3.0 = 8.0\ \Omega$. Note that you can specify this value to two significant figures, the 8 and the 0, rather than saying simply 8 Ω. This is because you are given both the voltage and the current to two significant figures. If the ammeter reading had been given as 3 A, you would only be entitled to express the answer as 8 Ω, to one significant digit. The digit 0 can be, and often is, just as important in calculations as any of the other digits 1 through 9.

Problem 4-8

What is the value of the resistance in Fig. 4-7 if the current is 18 mA and the voltage is 229 mV?

First, convert these values to amperes and volts. This gives $I = 0.018$ A and $E = 0.229$ V. Then plug into the equation: $R = E/I = 0.229/0.018 = 13\ \Omega$.

Problem 4-9

Suppose the ammeter in Fig. 4-7 reads 52 μA and the voltmeter indicates 2.33 kV. What is the resistance?

Convert to amperes and volts, getting $I = 0.000052$ A and $E = 2330$ V. Then plug into the formula: $R = E/I = 2330/0.000052 = 45{,}000{,}000\ \Omega = 45\ \text{M}\Omega$.

Power Calculations

You can calculate the power P, in watts, in a dc circuit such as that shown in Fig. 4-7, by using the formula $P = EI$. This formula tells us that the power in watts is the product of the voltage in volts and the current in amperes. If you are not given the voltage directly, you can calculate it if you know the current and the resistance.

Recall the Ohm's Law formula for obtaining voltage: $E = IR$. If you know I and R but you don't know E, you can get the power P this way:

$$P = EI = (IR)I = I^2R$$

Suppose you're given only the voltage and the resistance. Remember the Ohm's Law formula for obtaining current: $I = E/R$. Therefore:

$$P = EI = E(E/R) = E^2/R$$

Problem 4-10

Suppose that the voltmeter in Fig. 4-7 reads 12 V and the ammeter shows 50 mA. What is the power dissipated by the potentiometer?

Use the formula $P = EI$. First, convert the current to amperes, getting $I = 0.050$ A. (Note that the last 0 counts as a significant digit.) Then multiply by 12 V, getting $P = EI = 12 \times 0.050 = 0.60$ W.

Problem 4-11

If the resistance in the circuit of Fig. 4-7 is 999 Ω and the voltage source delivers 3 V, what is the power dissipated by the potentiometer?

Use the formula $P = E^2/R = 3 \times 3/999 = 9/999 = 0.009$ W = 9 mW. You are justified in going to only one significant figure here.

Problem 4-12

Suppose the resistance in Fig. 4-7 is 47 kΩ and the current is 680 mA. What is the power dissipated by the potentiometer?

Use the formula $P = I^2R$, after converting to ohms and amperes. Then $P = 0.680 \times 0.680 \times 47{,}000 = 22{,}000$ W = 22 kW. (This is an unrealistic state of affairs: an ordinary potentiometer, such as the type you would use as the volume control in a radio, dissipating 22 kW, several times more than a typical household!)

Problem 4-13

How much voltage would be necessary to drive 680 mA through a resistance of 47 kΩ, as is described in the previous problem?

Use Ohm's Law to find the voltage: $E = IR = 0.680 \times 47{,}000 = 32{,}000$ V = 32 kV. That's the level of voltage you'd expect to find on a major utility power line, or in a high-power tube-type radio broadcast transmitter.

Resistances in Series

When you place resistances in series, their ohmic values add together to get the total resistance. This is easy to imagine, and it's easy to remember!

Problem 4-14

Suppose resistors with the following values are connected in series, as shown in Fig. 4-8: 112 Ω, 470 Ω, and 680 Ω. What is the total resistance of the series combination?

Simply add up the values, getting a total of 112 + 470 + 680 = 1262 Ω. You might round this off to 1260 Ω. It depends on the *tolerances* of the resistors—how precise their actual values are to the ones specified by the manufacturer.

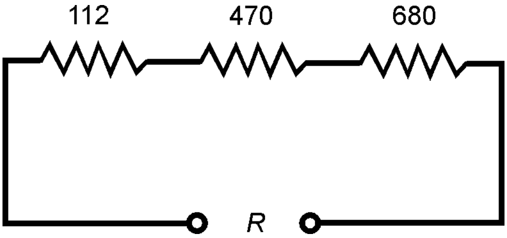

4-8 Three resistors in series. Illustration for Problem 4-14. Resistance values are in ohms.

Resistances in Parallel

When resistances are placed in parallel, they behave differently than they do in series. One way to look at resistances in parallel is to consider them as conductances instead. In parallel, conductances add up directly, just as resistances add up in series. If you change all the ohmic values to siemens, you can add these figures up and convert the final answer back to ohms.

The symbol for conductance is G. This figure, in siemens, is related to the resistance R, in ohms, by these formulas, which you learned in Chap. 2:

$$G = 1/R$$
$$R = 1/G$$

Problem 4-15

Consider five resistors in parallel. Call them R_1 through R_5, and call the total resistance R as shown in Fig. 4-9. Let the resistance values be as follows: $R_1 = 100\ \Omega$, $R_2 = 200\ \Omega$, $R_3 = 300\ \Omega$, $R_4 = 400\ \Omega$, and $R_5 = 500\ \Omega$. What is the total resistance, R, of this parallel combination?

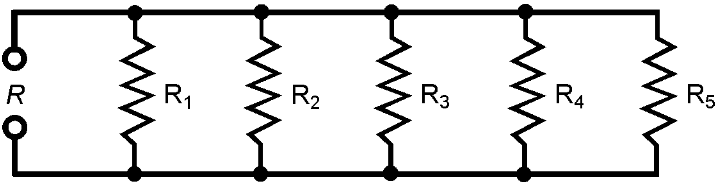

4-9 Five resistors of values R_1 through R_5, connected in parallel, produce a net resistance R. Illustration for Problems 4-15 and 4-16.

Converting the resistances to conductance values, you get: $G_1 = 1/100 = 0.01$ S, $G_2 = 1/200 = 0.005$ S, $G_3 = 1/300 = 0.00333$ S, $G_4 = 1/400 = 0.0025$ S, and $G_5 = 1/500 = 0.002$ S. Adding these gives $G = 0.01 + 0.005 + 0.00333 + 0.0025 + 0.002 = 0.0228$ S. The total resistance is therefore $R = 1/G = 1/0.0228 = 43.8\ \Omega$.

Problem 4-16

Suppose you have five resistors, called R_1 through R_5, connected in parallel as shown in Fig. 4-9. Suppose all the resistances, R_1 through R_5, are 4.70 kΩ. What is the total resistance, R, of this combination?

When you have two or more resistors connected in parallel and their resistances are all the same, the total resistance is equal to the resistance of any one component divided by the number of components. In this example, convert the resistance of any single resistor to 4700 Ω, and then divide this by 5. Thus, you can see that the total resistance is 4700/5 = 940 Ω.

In a situation like this, where you have a bunch of resistors connected together to operate as a single unit, the total resistance is sometimes called the *net resistance.* Take note, too, that R is not italicized when it means *resistor*, but R is italicized when it means *resistance!*

Division of Power

When combinations of resistances are connected to a source of voltage, they draw current. You can figure out how much current they draw by calculating the total resistance of the combination, and then considering the network as a single resistor.

If the resistors in the network all have the same ohmic value, the power from the source is evenly distributed among them, whether they are hooked up in series or in parallel. For example, if there are eight identical resistors in series with a battery, the network consumes a certain amount of power, each resistor bearing $\frac{1}{8}$ of the load. If you rearrange the circuit so that the resistors are in parallel, the circuit will dissipate a certain amount of power (a lot more than when the resistors were in series), but again, each resistor will handle $\frac{1}{8}$ of the total power load.

If the resistances in the network do not all have identical ohmic values, they divide up the power unevenly. Situations like this are discussed in the next chapter.

Resistances in Series-Parallel

Sets of resistors, all having identical ohmic values, can be connected together in parallel sets of series networks, or in series sets of parallel networks. By doing this, the total power-handling capacity of the resistance can be greatly increased over that of a single resistor.

Sometimes, the total resistance of a *series-parallel network* is the same as the value of any one of the resistors. This is always true if the components are identical, and are in a network called an *n-by-n matrix.* That means, when n is a whole number, there are n parallel sets of n resistors in series (Fig. 4-10A), or else there are n series sets of n resistors in parallel (Fig. 4-10B). Either arrangement gives the same practical result.

Engineers and technicians sometimes use series-parallel networks to obtain resistances with large power-handling capacity. A series-parallel array of n by n resistors will have n^2 times that of a single resistor. Thus, a 3 × 3 series-parallel matrix of 2 W resistors can handle up to $3^2 \times 2 = 9 \times$

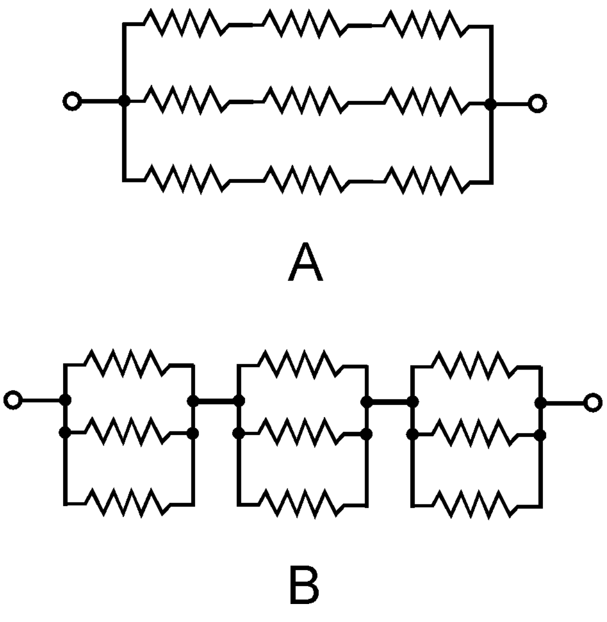

4-10 Series-parallel resistances. At A, sets of series resistors are connected in parallel. At B, sets of parallel resistances are connected in series. These examples show symmetrical *n*-by-*n* matrices with $n = 3$.

2 = 18 W, for example. A 10 × 10 array of 1-W resistors can dissipate up to 100 W. The total power-handling capacity is multiplied by the total number of resistors in the matrix. But this is true only if all the resistors have the same ohmic values, and the same power-dissipation ratings.

It is unwise to build series-parallel arrays from resistors with different ohmic values or power ratings. If the resistors have values and/or ratings that are even a little nonuniform, one of them might be subjected to more current than it can withstand, and it will burn out. Then the current distribution in the network can change so a second component fails, and then a third. It's hard to predict the current and power distribution in an array when its resistor values are all different.

If you need a resistance with a certain power-handling capacity, you must be sure the network can handle at least that much power. If a 50-W rating is required, and a certain combination will handle 75 W, that's fine. But it isn't good enough to build a circuit that will handle only 48 W. Some extra tolerance, say 10 percent over the minimum rating needed, is good, but it's silly to make a 500-W network using far more resistors than necessary, unless that's the only convenient combination given the parts available.

Nonsymmetrical series-parallel networks, made up from identical resistors, can increase the power-handling capability over that of a single resistor. But in these cases, the total resistance is not the same as the value of the single resistors. The overall power-handling capacity is always multiplied by the total number of resistors, whether the network is symmetrical or not, provided all the ohmic values are identical. In engineering work, cases sometimes arise where nonsymmetrical networks fit the need.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct answers. The answers are in the back of the book.

1. Suppose you double the voltage in a simple dc circuit, and cut the resistance in half. The current will
   - (a) become four times as great.
   - (b) become twice as great.
   - (c) stay the same as it was before.
   - (d) become half as great.
2. You can expect to find a wiring diagram
   - (a) on a sticker on the back of a television receiver.
   - (b) in an advertisement for an electric oven.
   - (c) in the service/repair manual for a two-way radio.
   - (d) in the photograph of the front panel of a stereo hi-fi tuner.

For questions 3 through 11, please refer to Fig. 4-7. Remember to take significant figures into account when completing your calculations!

3. Given a dc voltage source delivering 24 V and a resistance of 3.3 kΩ, what is the current?
   - (a) 0.73 A
   - (b) 138 A
   - (c) 138 mA
   - (d) 7.3 mA

4. Suppose the resistance is 472 Ω, and the current is 875 mA. The source voltage must therefore be
   - (a) 413 V.
   - (b) 0.539 V.
   - (c) 1.85 V.
   - (d) none of the above.

5. Suppose the dc voltage is 550 mV and the current is 7.2 mA. Then the resistance is
   - (a) 0.76 Ω.
   - (b) 76 Ω.
   - (c) 0.0040 Ω.
   - (d) none of the above.

6. Given a dc voltage source of 3.5 kV and a resistance of 220 Ω, what is the current?
   - (a) 16 mA
   - (b) 6.3 mA
   - (c) 6.3 A
   - (d) None of the above

7. Suppose the resistance is 473,332 Ω, and the current flowing through it is 4.4 mA. The best expression for the voltage of the source is
   - (a) 2082 V.
   - (b) 110 kV.
   - (c) 2.1 kV.
   - (d) 2.08266 kV.

8. A source delivers 12 V and the current is 777 mA. The best expression for the resistance is
   - (a) 15 Ω.
   - (b) 15.4 Ω.
   - (c) 9.3 Ω.
   - (d) 9.32 Ω.

9. Suppose the voltage is 250 V and the current is 8.0 mA. The power dissipated by the potentiometer is
   - (a) 31 mW.
   - (b) 31 W.
   - (c) 2.0 W.
   - (d) 2.0 mW.

10. Suppose the voltage from the source is 12 V and the potentiometer is set for 470 Ω. The power dissipated in the resistance is approximately
   - (a) 310 mW.
   - (b) 25.5 mW.
   - (c) 39.2 W.
   - (d) 3.26 W.

11. If the current through the potentiometer is 17 mA and its resistance is set to 1.22 kΩ, what is the power dissipated by it?

  (a) 0.24 μW
  (b) 20.7 W
  (c) 20.7 mW
  (d) 350 mW

12. Suppose six resistors are hooked up in series, and each of them has a value of 540 Ω. What is the resistance across the entire combination?

  (a) 90 Ω
  (b) 3.24 kΩ
  (c) 540 Ω
  (c) None of the above

13. If four resistors are connected in series, each with a value of 4.0 kΩ, the total resistance is

  (a) 1 kΩ.
  (b) 4 kΩ.
  (c) 8 kΩ.
  (d) 16 kΩ.

14. Suppose you have three resistors in parallel, each with a value of 0.069 MΩ. Then the total resistance is

  (a) 23 Ω.
  (b) 23 kΩ.
  (c) 204 Ω.
  (d) 0.2 MΩ.

15. Imagine three resistors in parallel, with values of 22 Ω, 27 Ω, and 33 Ω. If a 12-V battery is connected across this combination, as shown in Fig. 4-11, what is the current drawn from the battery?

  (a) 1.4 A
  (b) 15 mA
  (c) 150 mA
  (d) 1.5 A

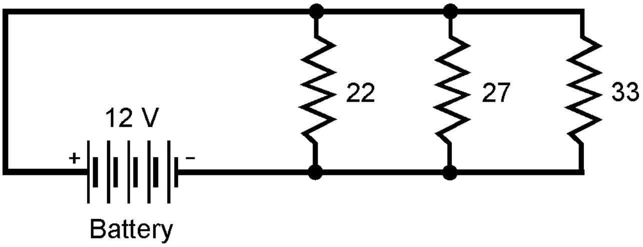

4-11 Illustration for Quiz Question 15. Resistance values are in ohms.

16. Imagine three resistors, with values of 47 Ω, 68 Ω, and 82 Ω, connected in series with a 50-V dc generator, as shown in Fig. 4-12. The total power consumed by this network of resistors is

(a) 250 mW.
(b) 13 mW.
(c) 13 W.
(d) impossible to determine from the data given.

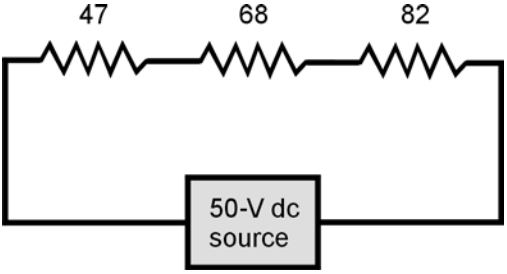

4-12 Illustration for Quiz Question 16. Resistance values are in ohms.

17. Suppose you have an unlimited supply of 1-W, 100-Ω resistors. You need to get a 100-Ω, 10-W resistor. This can be done most cheaply by means of a series-parallel matrix of

(a) 3 × 3 resistors.
(b) 4 × 3 resistors.
(c) 4 × 4 resistors.
(d) 2 × 5 resistors.

18. Suppose you have an unlimited supply of 1-W, 1000-Ω resistors, and you need a 500-Ω resistance rated at 7 W or more. This can be done by assembling

(a) four sets of two resistors in series, and connecting these four sets in parallel.
(b) four sets of two resistors in parallel, and connecting these four sets in series.
(c) a 3 × 3 series-parallel matrix of resistors.
(d) a series-parallel matrix, but something different than those described above.

19. Suppose you have an unlimited supply of 1-W, 1000-Ω resistors, and you need to get a 3000-Ω, 5-W resistance. The best way is to

(a) make a 2 × 2 series-parallel matrix.
(b) connect three of the resistors in parallel.
(c) make a 3 × 3 series-parallel matrix.
(d) do something other than any of the above.

20. Good engineering practice usually requires that a series-parallel resistive network be assembled

(a) from resistors that are all different.
(b) from resistors that are all identical.
(c) from a series combination of resistors in parallel but not from a parallel combination of resistors in series.
(d) from a parallel combination of resistors in series, but not from a series combination of resistors in parallel.

5
CHAPTER

Direct-Current Circuit Analysis

IN THIS CHAPTER, YOU'LL LEARN MORE ABOUT DC CIRCUITS AND HOW THEY BEHAVE UNDER VARIOUS conditions. These principles apply to most ac utility circuits as well.

Current through Series Resistances

Have you ever used those tiny holiday lights that come in strings? If one bulb burns out, the whole set of bulbs goes dark. Then you have to find out which bulb is bad, and replace it to get the lights working again. Each bulb works with something like 10 V; there are about a dozen bulbs in the string. You plug in the whole bunch and the 120-V utility mains drive just the right amount of current through each bulb.

In a series circuit, such as a string of light bulbs (Fig. 5-1), the current at any given point is the same as the current at any other point. The ammeter, A, is shown in the line between two of the bulbs. If it were moved anywhere else along the current path, it would indicate the same current.

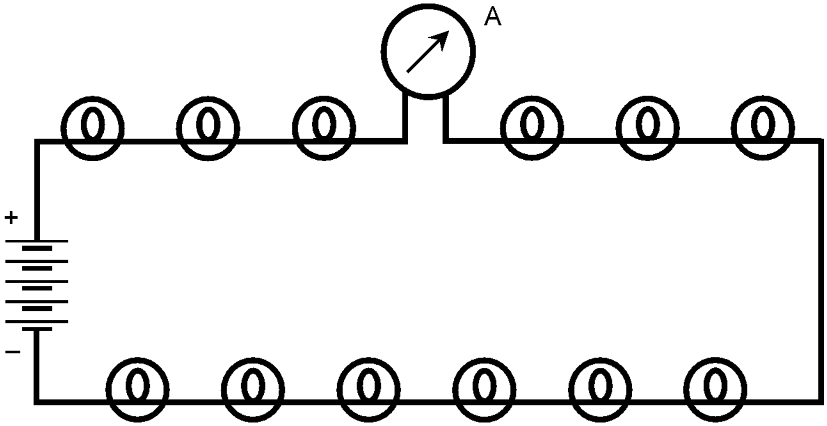

5-1 Light bulbs in series, with an ammeter (A) in the circuit.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

This is true in any series dc circuit, no matter what the components actually are, and regardless of whether or not they all have the same resistance.

If the bulbs in Fig. 5-1 had different resistances, some of them would consume more power than others. In case one of the bulbs in Fig. 5-1 burns out, and its socket is then shorted out instead of filled with a replacement bulb, the current through the whole chain will increase, because the overall resistance of the string will go down. This will force each of the remaining bulbs to carry more current, and pretty soon another bulb would burn out because of the excessive current. If it, too, were replaced with a short circuit, the current would be increased still further. A third bulb would blow out almost right away thereafter.

Voltages across Series Resistances

The bulbs in the string of Fig. 5-1, being all the same, each get the same amount of voltage from the source. If there are a dozen bulbs in a 120-V circuit, each bulb has a potential difference of 10 V across it. This will remain true even if the bulbs are replaced with brighter or dimmer ones, as long as all the bulbs in the string are identical.

Look at the schematic diagram of Fig. 5-2. Each resistor carries the same current. Each resistance R_n has a potential difference E_n across it equal to the product of the current and the resistance of that particular resistor. The voltages E_n are in series, like cells in a battery, so they add together. What if the voltages across all the resistors added up to something more or less than the supply voltage, E? Then there would have to be a "phantom EMF" someplace, adding or taking away voltage. But that's impossible. Voltage cannot come out of nowhere!

Look at this another way. The voltmeter V in Fig. 5-2 shows the voltage E of the battery, because the meter is hooked up across the battery. The voltmeter V also shows the sum of the voltages E_n across the set of resistances, because it's connected across the whole combination. The meter says the same thing whether you think of it as measuring the battery voltage E or as measuring the sum of the voltages E_n across the series combination of resistances. Therefore, E is equal to the sum of the voltages E_n.

How do you find the voltage across any particular resistance R_n in a circuit like the one in Fig. 5-2? Remember Ohm's Law for finding voltage: $E = IR$. Remember, too, that you must use volts, ohms, and amperes when making calculations.

In order to find the current in the circuit, I, you need to know the total resistance and the supply voltage; then $I = E/R$. First find the current in the whole circuit; then find the voltage across any particular resistor.

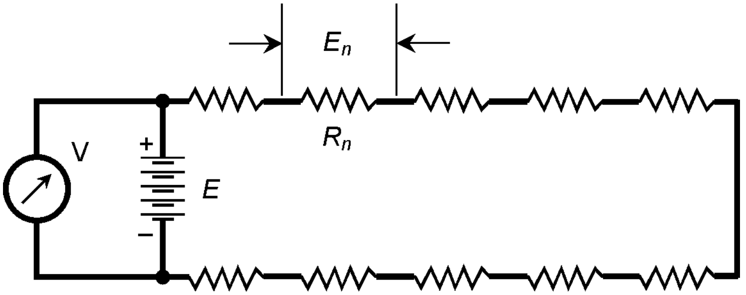

5-2 Analysis of voltages in a series circuit.

Problem 5-1

In Fig. 5-2, there are 10 resistors. Five of them have values of 10 Ω, and the other five have values of 20 Ω. The power source is 15-V dc. What is the voltage across any one of the 10-Ω resistors? Across any one of the 20-Ω resistors?

First, find the total resistance: $R = (10 \times 5) + (20 \times 5) = 50 + 100 = 150$ Ω. Then find the current: $I = E/R = 15/150 = 0.10$ A. This is the current through each of the resistances in the circuit.

- If $R_n = 10$ Ω, then $E_n = IR_n = 0.1 \times 10 = 1.0$ V.
- If $R_n = 20$ Ω, then $E_n = IR_n = 0.1 \times 20 = 2.0$ V.

Let's check to be sure all of these voltages add up to the supply voltage. There are five resistors with 1.0 V across each, for a total of 5.0 V; there are also five resistors with 2.0 V across each, for a total of 10 V. So the sum of the voltages across the resistors is 5.0 + 10 = 15 V.

Problem 5-2

In the circuit of Fig. 5-2, what will happen to the voltages across the resistances if one of the 20-Ω resistances is replaced with a short circuit?

In this case the total resistance becomes $R = (10 \times 5) + (20 \times 4) = 50 + 80 = 130$ Ω. The current is therefore $I = E/R = 15/130 = 0.12$ A. This is the current at any point in the circuit, rounded off to two significant figures.

The voltage E_n across any of the 10-Ω resistances R_n is equal to IR_n, which is $0.12 \times 10 = 1.2$ V. The voltage E_n across any of the 20-Ω resistances R_n is equal to IR_n, which is $0.12 \times 20 = 2.4$ V. Checking the total voltage, add $(5 \times 1.2) + (4 \times 2.4) = 6.0 + 9.6 = 15.6$ V. This rounds off to 16 V when we cut it down to two significant figures.

A "Rounding-Off Bug"

Compare the result for total voltage in Problem 5-2 with the result for total voltage in Problem 5-1. What is going on here? Where does the extra volt come from in the second calculation? Certainly, shorting out one of the resistances cannot cause the battery voltage to change!

This is an example of what can happen when you round off to a certain number of significant figures after calculating the value of some parameter X in a circuit, then change a different parameter Y in the circuit, and finally calculate the value of X again, rounding off to the same number of significant digits as you did the first time. The discrepancy is the result of a "rounding-off bug."

If this bug bothers you (and it should), keep all the digits your calculator will hold while you go through the solution process for Problem 5-2. The current in the circuit, as obtained by means of a calculator that can show 10 digits, should come out as 0.115384615 A. When you find the voltages across all the resistances R_n, accurate to all these extra digits, and then add them up, you'll get a final rounded-off voltage of 15 V.

This example shows why it is a good idea to wait until you get the final answer in a calculation, or set of calculations, involving a particular circuit before you round off to the allowed number of significant digits. Rounding-off bugs of the sort we have just seen can be more than mere annoyances. They are easy to overlook, but they can generate large errors in *iterative processes* involving calculations that are done over and over.

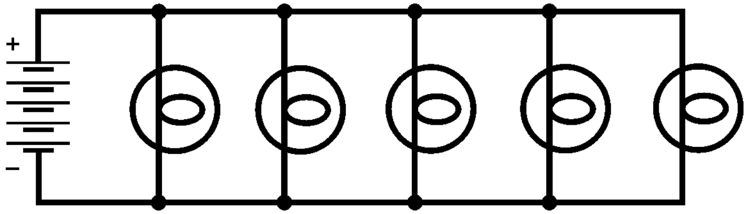

5-3 Light bulbs in parallel.

Voltage across Parallel Resistances

Imagine a set of ornamental light bulbs connected in parallel (Fig. 5-3). This is the method used for outdoor holiday lighting or for bright indoor lighting. It's easier to repair a parallel-wired string of such lights if one bulb should burn out than it is to fix a series-wired string. And in the parallel configuration, the failure of one bulb does not cause total system failure.

In a parallel circuit, the voltage across each component is equal to the supply or battery voltage. The current drawn by each component depends only on the resistance of that particular device. In this sense, the components in a parallel-wired circuit operate independently, as opposed to the series-wired circuit in which they all interact.

If any one branch of a parallel circuit opens up, is disconnected, or is removed, the conditions in the other branches do not change. If new branches are added, assuming the power supply can handle the load, conditions in previously existing branches are not affected.

Currents through Parallel Resistances

Refer to the schematic diagram of Fig. 5-4. The resistances are called R_n. The total parallel resistance in the circuit is R. The battery voltage is E. The current in any particular branch n, containing resistance R_n, is measured by ammeter A and is called I_n. The sum of all the currents I_n is equal to the total current, I, drawn from the battery. The current is divided up in the parallel circuit in a manner similar to the way that voltage is divided up in a series circuit.

Conventional Current

Have you noticed that the direction of current flow in Fig. 5-4 is portrayed as outward from the positive battery terminal? Don't electrons, which are the actual charge carriers in a wire, flow out of the minus terminal of a battery? Yes, that's true; but scientists consider *theoretical current*, more often called *conventional current* (because it is defined by convention), to flow from positive to negative voltage points, rather than from negative to positive.

Problem 5-3

Suppose that the battery in Fig. 5-4 delivers 12 V. Further suppose that there are 12 resistors, each with a value of 120 Ω in the parallel circuit. What is the total current, I, drawn from the battery?

First, find the total resistance. This is easy, because all the resistors have the same value. Just divide $R_n = 120$ by 12 to get $R = 10\ \Omega$. Then the current can be found by Ohm's Law: $I = E/R = 12/10 = 1.2$ A.

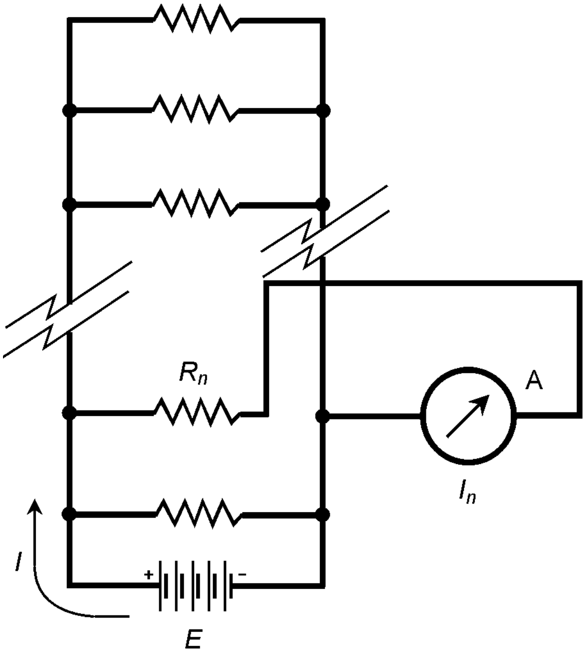

5-4 Analysis of current in a parallel circuit.

Problem 5-4

In the circuit of Fig. 5-4, what does the ammeter say?

This involves finding the current in any given branch. The voltage is 12 V across every branch, and $R_n = 120\ \Omega$. Therefore I_n, the ammeter reading, is found by Ohm's Law: $I_n = E/R_n = 12/120 = 0.10$ A.

Because this is a parallel circuit, all of the branch currents I_n should add up to get the total current, I. There are 12 identical branches, each carrying 0.10 A; therefore the total current is $0.10 \times 12 = 1.2$ A. It checks out.

Problem 5-5

Suppose three resistors are in parallel across a battery that supplies $E = 12$ V. The resistances are $R_1 = 22\ \Omega$, $R_2 = 47\ \Omega$, and $R_3 = 68\ \Omega$. These resistances carry currents I_1, I_2, and I_3, respectively. What is the current, I_3, through R_3?

This problem is solved by means of Ohm's Law as if R_3 is the only resistance in the circuit. There's no need to worry about the parallel combination. The other branches do not affect I_3. Thus $I_3 = E/R_3 = 12/68 = 0.18$ A.

Problem 5-6

What is the total current drawn by the circuit described in Problem 5-5?

There are two ways to go at this. One method involves finding the total resistance, R, of R_1, R_2, and R_3 in parallel, and then calculating I based on R. Another way is to find the currents through R_1, R_2, and R_3 individually, and then add them up.

Using the first method, first change the resistances R_n into conductances G_n. This gives $G_1 = 1/R_1 = 1/22 = 0.04545$ S, $G_2 = 1/R_2 = 1/47 = 0.02128$ S, and $G_3 = 1/R_3 = 1/68 = 0.01471$ S. Adding these gives $G = 0.08144$ S. The resistance is therefore $R = 1/G = 1/0.08144 = 12.279\ \Omega$. Use Ohm's Law to find $I = E/R = 12/12.279 = 0.98$ A. Note that extra digits are used throughout the calculation, rounding off only at the end.

Now let's try the other method. Find $I_1 = E/R_1 = 12/22 = 0.5455$ A, $I_2 = E/R_2 = 12/47 = 0.2553$ A, and $I_3 = E/R_3 = 12/68 = 0.1765$ A. Adding these gives $I = I_1 + I_2 + I_3 = 0.5455 + 0.2553 + 0.1765 = 0.9773$ A, which rounds off to 0.98 A.

Power Distribution in Series Circuits

When calculating the power in a circuit containing resistors in series, all you need to do is find out the current, I, that the circuit is carrying. Then it's easy to calculate the power P_n dissipated by any one of the resistances R_n, based on the formula $P_n = I^2R_n$.

Problem 5-7

Suppose we have a series circuit with a supply of 150 V and three resistances: $R_1 = 330\ \Omega$, $R_2 = 680\ \Omega$, and $R_3 = 910\ \Omega$. What is the power dissipated by R_2?

First, find the current that flows through the circuit. Calculate the total resistance first. Because the resistors are in series, the total is $R = 330 + 680 + 910 = 1920\ \Omega$. The current is $I = 150/1920 = 0.07813$ A. The power dissipated by R_2 is therefore $P_2 = I^2R_2 = 0.07813 \times 0.07813 \times 680 = 4.151$ W. Round this off to three significant digits, because that's all we have in the data, to obtain 4.15 W.

The total wattage dissipated in a series circuit is equal to the sum of the wattages dissipated in each resistance.

Problem 5-8

Calculate the total dissipated power P in the circuit of Problem 5-7 by two different methods.

First, let's figure out the power dissipated by each of the three resistances separately, and then add the figures up. The power P_2 is already known. Let's use all the significant digits we have while we calculate. Thus, as found in Problem 5-7, $P_2 = 4.151$ W. Recall that the current is $I = 0.07813$ A. Then $P_1 = 0.07813 \times 0.07813 \times 330 = 2.014$ W, and $P_3 = 0.07813 \times 0.07813 \times 910 = 5.555$ W. Adding the three power figures gives us $P = P_1 + P_2 + P_3 = 2.014 + 4.151 + 5.555 = 11.720$ W. We should round this off to 11.7 W.

The second method is to find the total series resistance and then calculate the power. The series resistance is $R = 1920\ \Omega$, as found in Problem 5-7. Then $P = I^2R = 0.07813 \times 0.07813 \times 1920 = 11.72$ W. Again, we should round this to 11.7 W.

Power Distribution in Parallel Circuits

When resistances are wired in parallel, they each consume power according to the same formula, $P = I^2R$. But the current is not the same in each resistance. An easier method to find the power P_n dissipated by each of the various resistances R_n is to use the formula $P_n = E^2/R_n$, where E is the voltage of the supply or battery. This voltage is the same across every branch resistance in a parallel circuit.

Problem 5-9

Suppose a dc circuit contains three resistances $R_1 = 22\ \Omega$, $R_2 = 47\ \Omega$, and $R_3 = 68\ \Omega$ across a battery that supplies a voltage of $E = 3.0$ V. Find the power dissipated by each resistance.

Let's find the square of the supply voltage, E^2, first. We'll be needing this figure often: $E^2 = 3.0 \times 3.0 = 9.0$. Then the wattages dissipated by resistances R_1, R_2, and R_3 respectively are $P_1 = 9.0/22 = 0.4091$ W, $P_2 = 9.0/47 = 0.1915$ W, and $P_3 = 9.0/68 = 0.1324$ W. These should be rounded off to $P_1 = 0.41$ W, $P_2 = 0.19$ W, and $P_3 = 0.13$ W. (But let's remember the values to four significant figures for the next problem!)

In a parallel circuit, the total dissipated wattage is equal to the sum of the wattages dissipated by the individual resistances.

Problem 5-10

Find the total consumed power of the resistor circuit in Problem 5-9 using two different methods.

The first method involves adding P_1, P_2, and P_3. Let's use the four-significant-digit values to avoid the possibility of encountering the rounding-off bug. The total power thus calculated is $P = 0.4091 + 0.1915 + 0.1324 = 0.7330$ W. Now that we've finished the calculation, we should round it off to 0.73 W.

The second method involves finding the net resistance R of the parallel combination. You can do this calculation yourself. Determining it to four significant digits, you should get a net resistance of $R = 12.28\ \Omega$. Then $P = E^2/R = 9.0/12.28 = 0.7329$ W. Now that the calculation is done, this can be rounded to 0.73 W.

It's the Law!

In electricity and electronics, dc circuit analysis can be made easier if you are acquainted with certain axioms, or *laws.* Here they are:

- The current in a series circuit is the same at every point along the way.
- The voltage across any resistance in a parallel combination of resistances is the same as the voltage across any other resistance, or across the whole set of resistances.
- The voltages across resistances in a series circuit always add up to the supply voltage.
- The currents through resistances in a parallel circuit always add up to the total current drawn from the supply.
- The total wattage consumed in a series or parallel circuit is always equal to the sum of the wattages dissipated in each of the resistances.

Now, let's get acquainted with two of the most famous laws that govern dc circuits. These rules are broad and sweeping, and they make it possible to analyze complicated series-parallel dc networks.

Kirchhoff's First Law

The physicist Gustav Robert Kirchhoff (1824–1887) was a researcher and experimentalist in a time when little was understood about how electric currents flow. Nevertheless, he used certain common-sense notions to deduce two important properties of dc circuits.

Kirchhoff reasoned that dc ought to behave something like water in a network of pipes, and that the current going into any point ought to be the same as the current going out of that point. This, Kirchhoff thought, must be true for any point in a circuit, no matter how many branches lead into or out of the point.

Two examples of this principles are shown in Fig. 5-5. Examine illustration A. At point X, I, the current going in, equals $I_1 + I_2$, the current going out. At point Y, $I_2 + I_1$, the current going in, equals I, the current going out. Now look at illustration B. In this case, at point Z, the current $I_1 + I_2$ going in is equal to the current $I_3 + I_4 + I_5$ going out. These are examples of *Kirchhoff's First Law.* We can also call it *Kirchhoff's Current Law* or the principle of *conservation of current.*

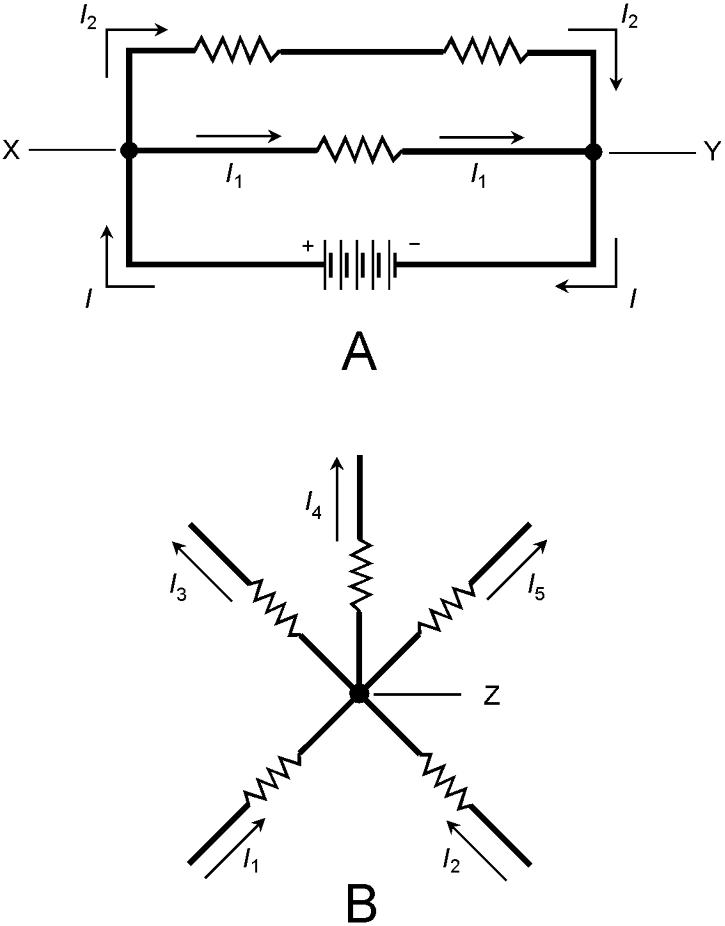

5-5 Kirchhoff's First Law. At A, the current into point X or point Y is the same as the current out of that point. That is, $I = I_1 + I_2$. At B, the current into point Z equals the current flowing out of point Z. That is, $I_1 + I_2 = I_3 + I_4 + I_5$. Illustration for Quiz Questions 13 and 14.

Problem 5-11

Refer to Fig. 5-5A. Suppose all three resistors have values of 100 Ω, and that $I_1 = 2.0$ A and $I_2 = 1.0$ A. What is the battery voltage?

First, find the current I drawn from the battery: $I = I_1 + I_2 = 2.0 + 1.0 = 3.0$ A. Next, find the resistance of the entire network. The two 100-Ω resistances in series give a value of 200 Ω, and this is in parallel with 100 Ω. You can do the calculations and find that the total resistance, R, connected across the battery is 66.67 Ω. Then $E = IR = 66.67 \times 3.0 = 200$ V.

Problem 5-12

In Fig. 5-5B, suppose each of the two resistors below point Z has a value of 100 Ω, and all three resistors above point Z have values of 10.0 Ω. Suppose the current through each 100-Ω resistor is 500 mA. What is the current through any one of the 10.0-Ω resistors, assuming that the current through all three 10.0-Ω resistors is the same? What is the voltage across any one of the three 10.0-Ω resistors?

The total current into point Z is 500 mA + 500 mA = 1.00 A. This is divided equally among the three 10-Ω resistors. Therefore, the current through any one of them is 1.00/3 A = 0.333 A. The voltage across any one of the 10.0-Ω resistors can thus found by Ohm's Law: $E = IR = 0.333 \times 10.0 = 3.33$ V.

Kirchhoff's Second Law

The sum of all the voltages, as you go around a circuit from some fixed point and return there from the opposite direction, and taking polarity into account, is always zero. Does this seem counterintuitive? Let's think about it a little more carefully.

What Kirchhoff was expressing, when he wrote his second law, is the principle that voltage cannot appear out of nowhere, nor can it vanish. All the potential differences must ultimately cancel each other out in any closed dc circuit, no matter how complicated that circuit happens to be. This is *Kirchhoff's Second Law.* We can also call it *Kirchhoff's Voltage Law* or the principle of *conservation of voltage.*

Remember the rule you've already learned about series dc circuits: The sum of the voltages across all the individual resistances adds up to the supply voltage. This statement is true as far as it goes, but it is an oversimplification, because it ignores polarity. The *polarity* of the potential difference across each resistance is *opposite* to the polarity of the potential difference across the battery. So when you add up the potential differences all the way around the circuit, taking polarity into account for every single component, you always get a net voltage of zero.

An example of Kirchhoff's Second Law is shown in Fig. 5-6. The voltage of the battery, E, has polarity opposite to the sum of the potential differences across the resistors, $E_1 + E_2 + E_3 + E_4$. Therefore, $E + E_1 + E_2 + E_3 + E_4 = 0$.

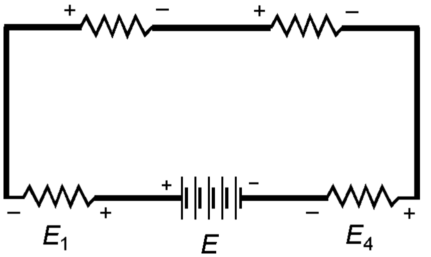

5-6 Kirchhoff's Second Law. The sum of the voltages across the resistances is equal to, but has opposite polarity from, the supply voltage. Therefore, $E + E_1 + E_2 + E_3 + E_4 = 0$. Illustration for Quiz Questions 15 and 16.

Problem 5-13

Refer to the diagram of Fig. 5-6. Suppose the four resistors have values of 50 Ω, 60 Ω, 70Ω, and 80 Ω, and that the current through each of them is 500 mA. What is the battery voltage, E?

Find the voltages E_1, E_2, E_3, and E_4 across each of the resistors. This can be done using Ohm's Law. For E_1, say with the 50-Ω resistor, calculate $E_1 = 0.500 \times 50 = 25$ V. In the same way, you can calculate $E_2 = 30$ V, $E_3 = 35$ V, and $E_4 = 40$ V. The supply voltage is the sum $E_1 + E_2 + E_3 + E_4 = 25 + 30 + 35 + 40 = 130$ V. Kirchhoff's Second Law tells us that the polarities of the voltages across the resistors are in the opposite direction from that of the battery.

Problem 5-14

In the situation shown by Fig. 5-6, suppose the battery provides 20 V. Suppose the resistors labeled with voltages E_1, E_2, E_3, and E_4 have ohmic values in the ratio 1:2:3:4 respectively. What is the voltage E_3?

This problem does not provide any information about current in the circuit, nor does it give you the exact resistances. But you don't need to know these things to solve for E_3. Regardless of what the actual ohmic values are, the ratio E_1:E_2:E_3:E_4 will be the same as long as the resistances are in the ratio 1:2:3:4. We can plug in any ohmic values we want for the values of the resistors, as long as they are in that ratio.

Let R_n be the resistance across which the voltage is E_n, where n can range from 1 to 4. Now that we have given the resistances specific names, suppose $R_1 = 1.0$ Ω, $R_2 = 2.0$ Ω, $R_3 = 3.0$ Ω, and $R_4 = 4.0$ Ω. These are in the proper ratio. The total resistance is $R = R_1 + R_2 + R_3 + R_4 = 1.0 + 2.0 + 3.0 + 4.0 = 10$ Ω. You can calculate the current as $I = E/R = 20/10 = 2.0$ A. Then the voltage E_3, across the resistance R_3, is given by Ohm's Law as $E_3 = IR_3 = 2.0 \times 3.0 = 6.0$ V.

Voltage Divider Networks

Resistances in series produce ratios of voltages, and these ratios can be tailored to meet certain needs by means of *voltage divider networks.*

When a voltage divider network is designed and assembled, the resistance values should be as small as possible without causing too much current drain on the battery or power supply. (In practice, the optimum values depend on the nature of the circuit being designed. This is a matter for engineers, and specific details are beyond the scope of this course.) The reason for choosing the smallest possible resistances is that, when the divider is used with a circuit, you do not want that circuit to upset the operation of the divider. The voltage divider "fixes" the intermediate voltages best when the resistance values are as small as the current-delivering capability of the power supply will allow.

Figure 5-7 illustrates the principle of voltage division. The individual resistances are R_1, R_2, R_3, . . . , R_n. The total resistance is $R = R_1 + R_2 + R_3 + \ldots + R_n$. The supply voltage is E, and the current in the circuit is therefore $I = E/R$. At the various points P_1, P_2, P_3, . . . , P_n, the potential differences relative to the negative battery terminal are E_1, E_2, E_3, . . . , E_n, respectively. The last voltage, E_n, is the same as the battery voltage, E. All the other voltages are less than E, and ascend in succession, so that $E_1 < E_2 < E_3 < \ldots < E_n$. (The mathematical symbol < means "is less than.")

The voltages at the various points increase according to the sum total of the resistances up to each point, in proportion to the total resistance, multiplied by the supply voltage. Thus, the voltage E_1 is equal to ER_1/R. The voltage E_2 is equal to $E(R_1 + R_2)/R$. The voltage E_3 is equal to $E(R_1 + R_2 + R_3)/\mathrm{R}$. This process goes on for each of the voltages at points all the way up to $E_n = E(R_1 + R_2 + R_3 + \ldots + R_n)/R = ER/R = E$.

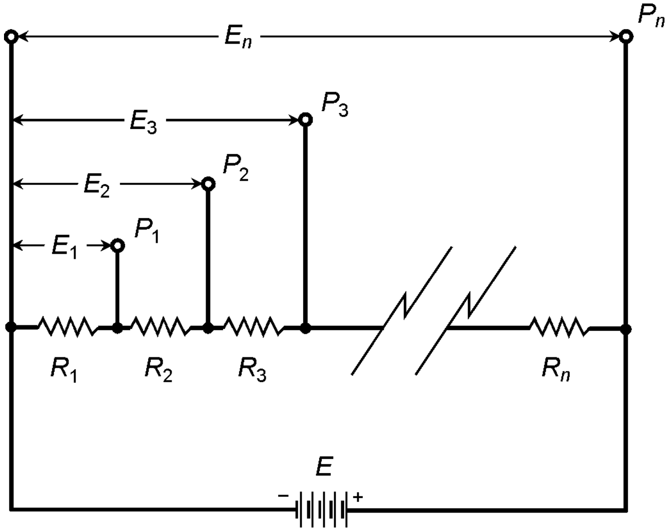

5-7 General arrangement for a voltage-divider circuit. Illustration for Quiz Questions 19 and 20.

Problem 5-15

Suppose you are building an electronic circuit, and the battery supplies 9.0 V. The minus terminal is at common (chassis) ground. You need to provide a circuit point where the dc voltage is +2.5 V. Give an example of a pair of resistors that can be connected in a voltage divider configuration, such that +2.5 V appears at some point.

Examine the schematic diagram of Fig. 5-8. There are infinitely many different combinations of resistances that will work here! Pick some total value, say $R = R_1 + R_2 = 1000\ \Omega$. Keep in mind that the ratio R_1:R will always be the same as the ratio E_1:E. In this case, $E_1 = 2.5$ V, so E_1:E = 2.5/9.0 = 0.28. This means that you want the ratio R_1:R to be equal to 0.28. You have chosen to

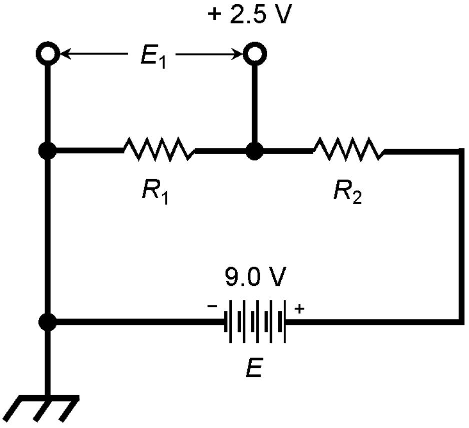

5-8 A voltage divider network in which 2.5 V dc is derived from a 9.0-V dc source.

make R equal to 1000 Ω. This means R_1 must be 280 Ω in order to get the ratio $R_1{:}R = 0.28$. The value of R_2 is the difference between R and R_1. That is $1000 - 280 = 720$ Ω.

In a practical circuit, you would want to choose the smallest possible value for R. This might be less than 1000 Ω, or it might be more, depending on the nature of the circuit and the current-delivering capability of the battery. It's not the actual values of R_1 and R_2 that determine the voltage you get at the intermediate point, but their ratio.

Problem 5-16

What is the current I, in milliamperes, drawn by the entire network of series resistances in the situation described in Problem 5-15 and its solution?

Use Ohm's Law to get $I = E/R = 9.0/1000 = 0.0090$ A = 9.0 mA.

Problem 5-17

Suppose that it is all right for the voltage divider network to draw up to 100 mA of current in the situation shown by Fig. 5-8 and posed by Problem 5-15. You want to design the network to draw this amount of current, because that will offer the best voltage regulation for the circuit to be operated from the network. What values of resistances R_1 and R_2 should you use?

Calculate the total resistance first, using Ohm's Law. Remember to convert 100 mA to amperes! That means you use the figure $I = 0.100$ A in your calculations. Then $R = E/I = 9.0/0.100 = 90$ Ω. The ratio of resistances that you need is $R_1{:}R_2 = 2.5/9.0 = 0.28$. You should use $R_1 = 0.28 \times 90 = 25$ Ω. The value of R_2 is the difference between R and R_1. That is, $R_2 = R - R_1 = 90 - 25 = 65$ Ω.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct answers. The answers are in the back of the book.

1. In a series-connected string of ornament bulbs, if one bulb gets shorted out, which of the following will occur?
   (a) All the other bulbs will go out.
   (b) The current in the string will go up.
   (c) The current in the string will go down.
   (d) The current in the string will stay the same.

2. Imagine that four resistors are connected in series across a 6.0-V battery, and the ohmic values are $R_1 = 10$ Ω, $R_2 = 20$ Ω, $R_3 = 50$ Ω, and $R_4 = 100$ Ω, as shown in Fig. 5-9. What is the voltage across the resistance R_2?
   (a) 0.18 V
   (b) 33 mV
   (c) 5.6 mV
   (d) 0.67 V

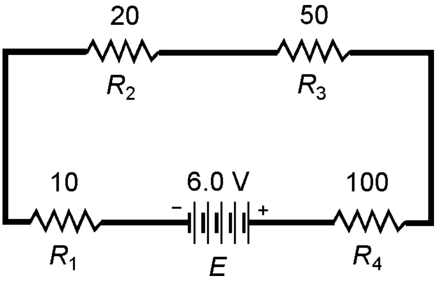

5-9 Illustration for Quiz Questions 2, 3, 8, and 9. Resistance values are in ohms.

3. In the scenario shown by Fig. 5-9, what is the voltage across the combination of R_3 and R_4?
   (a) 0.22 V
   (b) 0.22 mV
   (c) 5.0 V
   (d) 3.3 V

4. Suppose three resistors are connected in parallel across a battery that delivers 15 V, and the ohmic values are $R_1 = 470\ \Omega$, $R_2 = 2.2\ \text{k}\Omega$, and $R_3 = 3.3\ \text{k}\Omega$, as shown in Fig. 5-10. The voltage across the resistance R_2 is
   (a) 4.4 V.
   (b) 5.0 V.
   (c) 15 V.
   (d) not determinable from the data given.

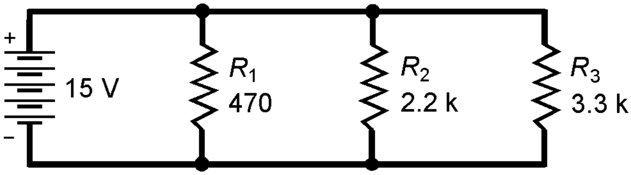

5-10 Illustration for Quiz Questions 4, 5, 6, 7, 10, and 11. Resistances are in ohms, where k indicates multiplication by 1000.

5. In the situation shown by Fig. 5-10, what is the current through R_2?
   (a) 6.8 mA
   (b) 43 mA
   (c) 0.15 A
   (d) 6.8 A

6. In the situation shown by Fig. 5-10, what is the total current drawn from the source?
   (a) 6.8 mA
   (b) 43 mA
   (c) 0.15 A
   (d) 6.8 A

7. In the situation shown by Fig. 5-10, suppose that resistor R_2 opens up. The current through the other two resistors will

(a) increase.
(b) decrease.
(c) drop to zero.
(d) not change.

8. Suppose that four resistors are connected in series with a 6.0-V supply, with values shown in Fig. 5-9. What is the power dissipated by the whole combination?

(a) 0.2 W
(b) 6.5 mW
(c) 200 W
(d) 6.5 W

9. In the situation shown by Fig. 5-9, what is the power dissipated by R_4?

(a) 11 mW
(b) 0.11 W
(c) 0.2 W
(d) 6.5 mW

10. Suppose that three resistors are in parallel as shown in Fig. 5-10. What is the power dissipated by the whole set of resistors?

(a) 5.4 W
(b) 5.4 μW
(c) 650 W
(d) 0.65 W

11. In the situation shown by Fig. 5-10, what is the power dissipated in resistance R_1?

(a) 32 mW
(b) 0.48 W
(c) 2.1 W
(d) 31 W

12. Fill in the blank in the following sentence to make it true: "In a series or parallel dc circuit, the sum of the ____s in each component is equal to the total ____ provided by the power supply."

(a) current
(b) voltage
(c) wattage
(d) resistance

13. Look at Fig. 5-5A. Suppose the resistors each have values of 33 Ω and the battery supplies 24 V. What is the current I_1?

(a) 1.1 A
(b) 0.73 A

(c) 0.36 A
(d) Not determinable from the information given

14. Look at Fig. 5-5B. Let each resistor have a value of 820 Ω. Suppose the top three resistors all lead to identical light bulbs. If $I_1 = 50$ mA and $I_2 = 70$ mA, what is the power dissipated in the resistor carrying current I_4?

(a) 33 W
(b) 40 mW
(c) 1.3 W
(d) It can't be found using the information given.

15. Refer to Fig. 5-6. Suppose the resistances R_1, R_2, R_3, and R_4 are in exactly the ratio 1:2:4:8 from left to right, and the battery supplies 30 V. What is the voltage E_2?

(a) 4.0 V
(b) 8.0 V
(c) 16 V
(d) It is not determinable from the data given.

16. Refer to Fig. 5-6. Suppose the resistances are each 3.3 kΩ, and the battery supplies 12 V. If the plus terminal of a dc voltmeter is placed between resistances R_1 and R_2 (with voltages E_1 and E_2 across them, respectively), and the minus terminal of the voltmeter is placed between resistances R_3 and R_4 (with voltages E_3 and E_4 across them, respectively), what will the meter register?

(a) 0.0 V
(b) 3.0 V
(c) 6.0 V
(d) 12 V

17. In a voltage divider network, the total resistance

(a) should be large to minimize current drain.
(b) should be as small as the power supply will allow.
(c) is not important.
(d) should be such that the current is kept to 100 mA.

18. The maximum voltage output from a voltage divider

(a) is a fraction of the power supply voltage.
(b) depends on the total resistance.
(c) is equal to the supply voltage.
(d) depends on the ratio of resistances.

19. Refer to Fig. 5-7. Suppose the battery voltage E is 18.0 V, and there are four resistances in the network such that $R_1 = 100$ Ω, $R_2 = 22.0$ Ω, $R_3 = 33.0$ Ω, and $R_4 = 47.0$ Ω. What is the voltage E_3 at P_3?

(a) 4.19 V
(b) 13.8 V

(c) 1.61 V
(d) 2.94 V

20. Refer to Fig. 5-7. Suppose the battery voltage is 12 V, and you want to obtain intermediate voltages of 3.0 V, 6.0 V, and 9.0 V. Suppose that a maximum of 200 mA is allowed to be drawn from the battery. What should the resistances, R_1, R_2, R_3, and R_4 be, respectively?

(a) 15 Ω, 30 Ω, 45 Ω, and 60 Ω
(b) 60 Ω, 45 Ω, 30 Ω, and 15 Ω
(c) 15 Ω, 15 Ω, 15 Ω, and 15 Ω
(d) There isn't enough information given here to design the circuit.

6
CHAPTER

Resistors

ALL ELECTRICAL COMPONENTS, DEVICES, AND SYSTEMS HAVE SOME RESISTANCE. IN EVERYDAY PRACTICE, there is no such thing as a perfect electrical conductor. You've seen some examples of circuits containing components that are deliberately designed to oppose the flow of current. These components are *resistors.* In this chapter, you'll learn all about them.

Purpose of the Resistor

Resistors play diverse roles in electrical and electronic equipment. Here are a few of the more common ways they are used.

Voltage Division

You've learned how voltage dividers can be designed using resistors. The resistors dissipate some power in doing this job, but the resulting voltages can provide the proper *biasing* of electronic circuits. This ensures, for example, that an amplifier or oscillator will function in the most efficient, reliable way possible.

Bias

The term *bias* means, in the case of a *bipolar transistor*, a *field-effect transistor*, or a *vacuum tube*, that the control electrode—the *base*, *gate*, or *grid*—is provided with a certain voltage, or made to carry a certain current, relative to the *emitter*, *source*, or *cathode.* Networks of resistors can accomplish this.

A radio transmitting amplifier is biased differently than an oscillator or a low-level receiving amplifier. Sometimes voltage division is required for biasing. Other times it isn't necessary. Figure 6-1 shows a bipolar transistor whose base is biased using a pair of resistors in a voltage divider configuration.

Current Limiting

Resistors interfere with the flow of electrons in a circuit. Sometimes this is essential to prevent damage to a component or circuit. A good example is a receiving amplifier. A resistor can keep the

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

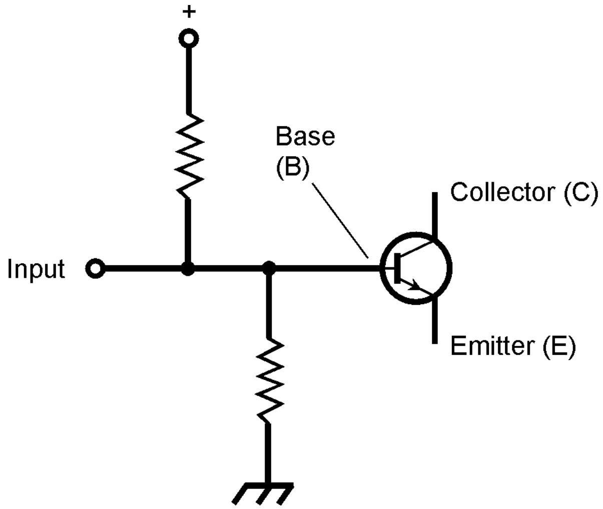

6-1 A pair of resistors can act as a voltage divider to bias the base of a transistor.

transistor from using up a lot of power just getting hot. Without resistors to limit or control the current, the transistor can be overstressed carrying direct current that doesn't contribute to the signal. Figure 6-2 shows a current-limiting resistor between the emitter of a bipolar transistor and electrical ground.

Power Dissipation

The dissipation of power in the form of heat is not always a bad thing. Sometimes a resistor can be used as a dummy component, so a circuit sees the resistor as if it were something more complicated. When testing a radio transmitter, for example, a resistor can be used to take the place of an antenna. This keeps the transmitter from interfering with communications on the airwaves. The transmitter output heats the resistor without radiating any signal. But as far as the transmitter knows, it's connected to a real antenna (Fig. 6-3)—and a perfect one, too, if the resistor has just the right ohmic value!

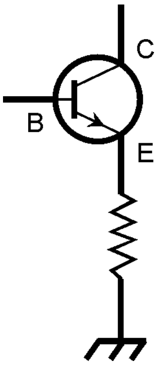

6-2 A resistor can limit the current that passes through the emitter of a transistor.

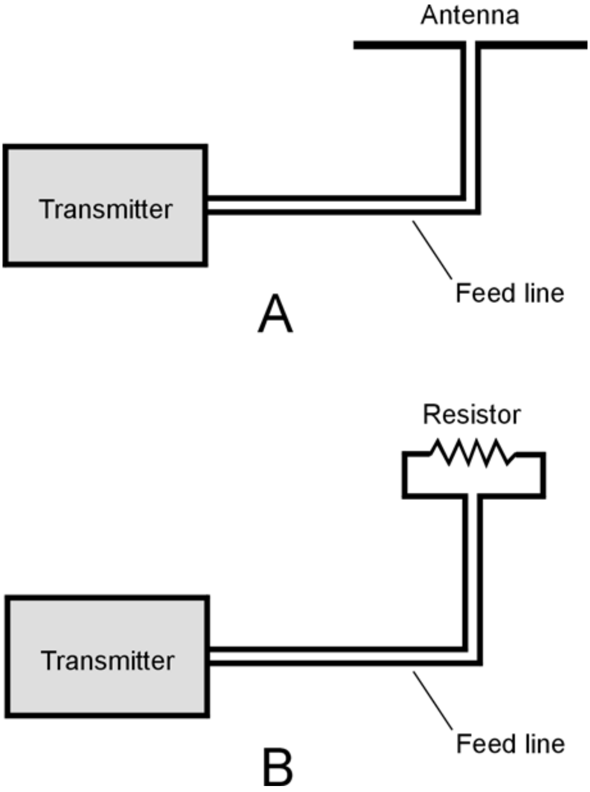

6-3 At A, a radio transmitter is connected to a real antenna. At B, the same transmitter is connected to a resistive dummy antenna.

Another situation in which power dissipation is useful is at the input of a power amplifier, such as the sort used in high-fidelity audio equipment. Sometimes the circuit *driving* the amplifier (supplying its input signal) has too much power. A resistor, or network of resistors, can dissipate this excess so that the amplifier doesn't get too much *drive.* In any type of amplifier, *overdrive* (an excessively strong input signal) can cause distortion, inefficiency, and other problems.

Bleeding Off Charge

In a high-voltage, dc power supply, capacitors are used to smooth out the fluctuations in the output. These capacitors acquire an electric charge, and they store it for a while. In some power supplies, these *filter capacitors* hold the full output voltage of the supply, say something like 750 V, even after the supply has been turned off, and even after it is unplugged from the wall outlet. If you attempt to repair such a power supply, you can be electrocuted by this voltage. *Bleeder resistors*, connected across the filter capacitors, drain their stored charge so that servicing the supply is not dangerous. In Fig. 6-4, the bleeder resistor, R, should have a value high enough so that it doesn't interfere with the operation of the power supply, but low enough so it will discharge the capacitor, C, in a short time after the power supply has been shut down.

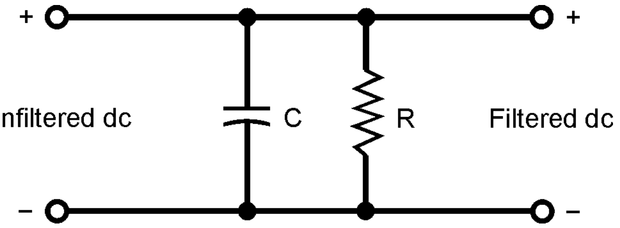

6-4 A bleeder resistor (R) is connected across the filter capacitor (C) in a power supply.

It's a good idea to short out all filter capacitors, using a screwdriver with an insulated handle and wearing heavy, insulated gloves, before working on a dc power supply. Even if the supply has bleeder resistors, they might take a while to get rid of the residual charge. In addition, bleeder resistors can, and sometimes do, fail.

Impedance Matching

A more sophisticated application for resistors is in the *coupling* in a chain of amplifiers, or in the input and output circuits of amplifiers. In order to produce the greatest possible amplification, the *impedances* must agree between the output of a given amplifier and the input of the next. The same is true between a source of signal and the input of an amplifier. Also, this applies between the output of an amplifier and a *load*, whether that load is a speaker, a headset, or whatever.

Impedance is the ac "big brother" of dc resistance. You will learn about impedance in Part 2 of this book.

Fixed Resistors

There are several ways in which *fixed resistors* (units whose resistance does not change, or cannot be adjusted) are manufactured. Here are the most common types.

Carbon-Composition Resistors

The cheapest method of making a resistor is to mix up powdered carbon (a fair electrical conductor) with some nonconductive substance, press the resulting claylike stuff into a cylindrical shape, and insert wire leads in the ends (Fig. 6-5). The resistance of the final product depends on the ratio

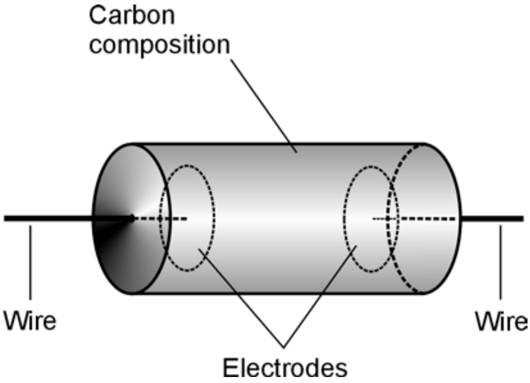

6-5 Construction of a carbon-composition resistor.

of carbon to the nonconducting material, and also on the physical distance between the wire leads. This results in a *carbon-composition resistor.*

Carbon-composition resistors can be manufactured in a wide range of resistance values. This kind of resistor also has the advantage of being *nonreactive*, meaning that it introduces almost pure resistance into the circuit, and not much capacitance or inductance. This makes carbon-composition resistors useful in radio receivers and transmitters.

Carbon-composition resistors dissipate power according to how big, physically, they are. Most of the carbon-composition resistors you see in electronics stores can handle ¼ W or ½ W. There are ⅛-W units available for miniaturized, low-power circuitry, and 1- or 2-W units for circuits where some electrical ruggedness is needed. Occasionally you'll see a carbon-composition resistor with a much higher power rating, but these are rare.

Wirewound Resistors

Another way to get resistance is to use a length of wire that isn't a good conductor. The wire can be wound around a cylindrical form as a coil (Fig. 6-6). The resistance is determined by how well the wire metal conducts, by its diameter or *gauge*, and by its stretched-out length. This type of component is called a *wirewound resistor.*

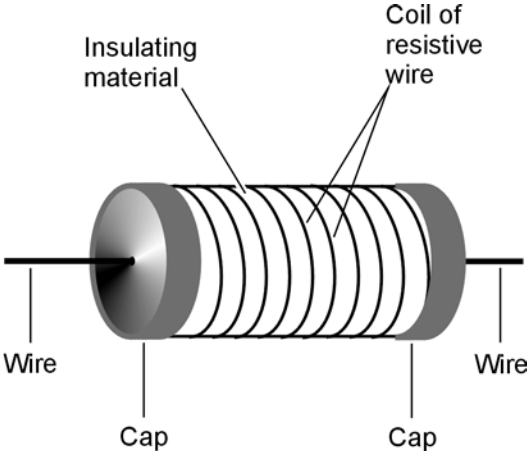

6-6 Construction of a wirewound resistor.

Wirewound resistors can be manufactured to have values within a very close range. They are *precision components.* Also, wirewound resistors can be made to handle large amounts of power. A disadvantage of wirewound resistors, in some applications, is that they act like inductors. This makes them unsuitable for use in most radio-frequency circuits. Wirewound resistors usually have low to moderate values of resistance.

Film-Type Resistors

Carbon, resistive wire, or some mixture of ceramic and metal can be applied to a cylindrical form as a film, or thin layer, in order to obtain a specific resistance. This type of component is called a *carbon-film resistor* or *metal-film resistor.* Superficially, it looks like a carbon-composition resistor, but the construction is different (Fig. 6-7).

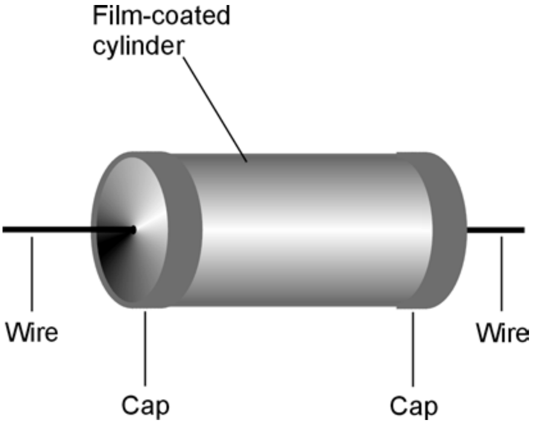

6-7 Construction of a film-type resistor.

The cylindrical form is made of an insulating substance, such as porcelain. The film is deposited on this form by various methods, and the value tailored as desired. Metal-film resistors can be made to have nearly exact values. Film-type resistors usually have low to medium-high resistance.

A major advantage of film-type resistors is that they, like carbon-composition resistors, do not have much inductance or capacitance. A disadvantage, in some applications, is that they can't handle as much power as carbon-composition or wirewound types.

Integrated-Circuit (IC) Resistors

Resistors can be fabricated on a semiconductor wafer known as an *integrated circuit* (IC), also called a *chip.* The thickness, and the types and concentrations of impurities added, control the resistance of the component. Integrated-circuit resistors can handle only a tiny amount of power because of their small size.

The Potentiometer

Figure 6-8 is a simplified drawing of the construction of a *potentiometer*, or variable resistor. A resistive strip, similar to that found on film-type fixed resistors, is bent into a nearly complete circle, and terminals are connected to either end. This forms a fixed resistance. To obtain the variable resistance, a sliding contact is attached to a rotatable shaft and bearing, and is connected to a third terminal. The resistance between the middle terminal and either of the end terminals can vary from zero up to the resistance of the whole strip.

Some potentiometers use a straight strip of resistive material, and the control moves up and down or from side to side. This type of variable resistor, called a *slide potentiometer*, is used in hi-fi audio *graphic equalizers*, as the volume controls in some hi-fi audio amplifiers, and in other applications when a linear scale is preferable to a circular scale. Potentiometers are manufactured to handle low levels of current, at low voltage.

Linear-Taper Potentiometer

One type of potentiometer uses a strip of resistive material whose density is constant all the way around. This results in a *linear taper.* The resistance between the center terminal and either end terminal changes at a steady rate as the control shaft is turned.

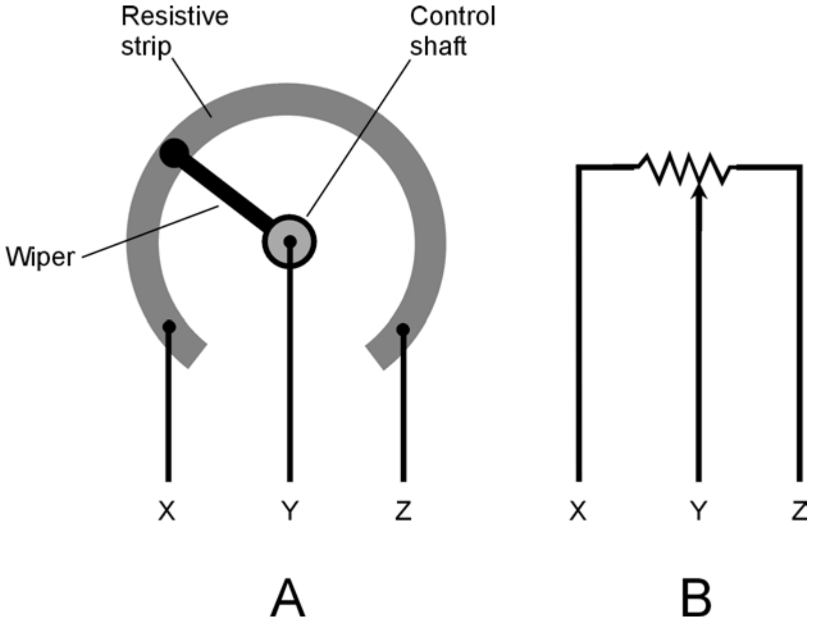

6-8 A simplified functional drawing of a rotary potentiometer (A), and the schematic symbol (B).

Suppose a *linear-taper potentiometer* has a value of zero to 280 Ω. In most units the shaft can be rotated through about 280°, or a little more than three-quarters of a circle. The resistance between the center and one end terminal will increase right along with the number of angular degrees that the shaft is turned. The resistance between the center and the other end terminal will be equal to 280 minus the number of degrees the shaft is turned. The resistance is a *linear function* of the angular shaft position.

Linear-taper potentiometers are commonly used in electronic test instruments and in various consumer electronic devices. Figure 6-9 is a graph of relative resistance versus relative angular shaft displacement for a linear-taper potentiometer.

Audio-Taper Potentiometer

In some applications, linear taper potentiometers don't work well. The volume control of a radio receiver or hi-fi audio amplifier is a good example. Humans perceive sound intensity according to the *logarithm* of the actual sound power. If you use a linear-taper potentiometer as the volume control for a radio or other sound system, the sound volume will vary too slowly in some parts of the control range, and too fast in other parts of the control range.

To compensate for the way in which people perceive sound level, an *audio-taper potentiometer* is used. In this device, the resistance between the center and end terminal increases as a nonlinear function of the angular shaft position. The device is sometimes called a *logarithmic-taper potentiometer* or *log-taper potentiometer* because the nonlinear function is logarithmic. This precisely compensates for the way the human ear-and-brain "machine" responds to sounds of variable intensity. Audio-taper potentiometers are manufactured so that as you turn the shaft, the sound intensity

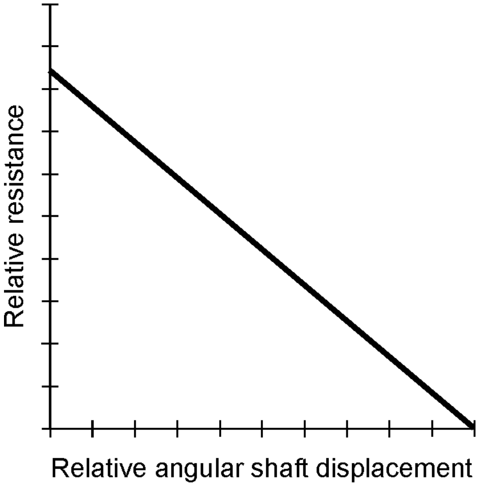

6-9 Resistance as a function of angular displacement for a linear-taper potentiometer.

seems to increase in a smooth, natural way. Figure 6-10 is a graph of relative resistance versus relative angular shaft displacement for an audio-taper potentiometer.

The Rheostat

A variable resistor can be made from a wirewound element, rather than a solid strip of material. This is called a *rheostat.* It can have either a rotary control or a sliding control. This depends on whether the resistive wire is wound around a donut-shaped form (*toroid*) or a cylindrical form (*solenoid*). Rheostats have inductance as well as resistance. They share the advantages and disadvantages of fixed wirewound resistors.

A rheostat is not continuously adjustable, as is a potentiometer. This is because the movable contact slides along from turn to turn of the wire coil. The smallest possible increment is the resistance in one turn of the coil.

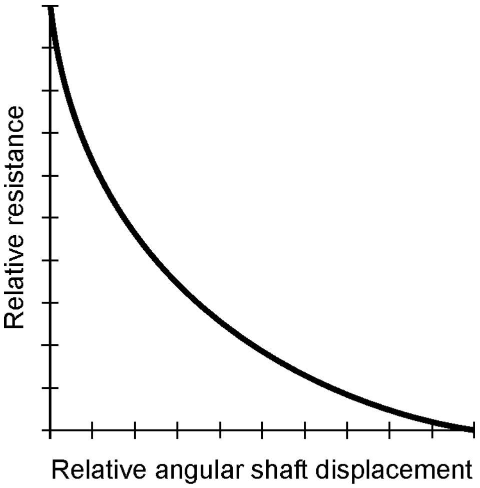

6-10 Resistance as a function of angular displacement for an audio-taper potentiometer.

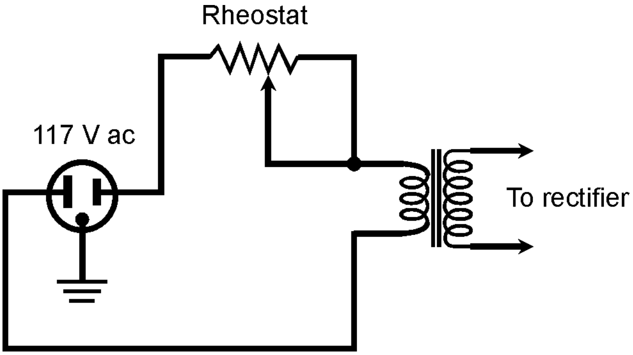

6-11 Connection of a rheostat in a variable-voltage power supply.

Rheostats are used in high-voltage, high-power applications. A good example is in a variable-voltage power supply. This kind of supply uses a transformer that steps up the voltage from the 117-V utility mains, and diodes to change the ac to dc. The rheostat can be placed between the utility outlet and the transformer (Fig. 6-11). This results in a variable voltage at the power-supply output.

The Decibel

As stated in the preceding paragraphs, perceived levels of sound change according to the logarithm of the actual sound power level. The same is true for various other phenomena, too, such as visible-light intensity and radio-frequency signal strength. Specialized units have been defined to take this into account.

The fundamental unit of sound-level change is called the *decibel*, symbolized as dB. A change of +1 dB is the minimum increase in sound level that you can detect if you are expecting it. A change of −1 dB is the minimum detectable decrease in sound volume, when you are anticipating the change. Increases in volume are given positive decibel values, and decreases in volume are given negative decibel values.

If you aren't expecting the level of sound to change, then it takes about +3 dB or −3 dB to make a noticeable difference.

Changes in intensity, when expressed in decibels, are sometimes called *gain* and *loss.* Positive decibel changes represent gain, and negative decibel changes represent loss. The sign (plus or minus) is usually absent when speaking of changes in terms of decibel gain or decibel loss. If you say that a certain system causes 5 dB of loss, you are saying that the gain of that circuit is −5 dB.

Calculating Decibel Values

Decibel values are calculated according to the logarithm of the ratio of change. Suppose a sound produces a power of P watts on your eardrums, and then it changes (either getting louder or softer) to a level of Q watts. The change in decibels is obtained by dividing out the ratio Q/P, taking its base-10 logarithm (symbolized as $\log_{10}$ or simply as log), and then multiplying the result by 10. Mathematically:

$$\text{dB} = 10 \log (Q/P)$$

As an example, suppose a speaker emits 1 W of sound, and then you turn up the volume so that it emits 2 W of sound power. Then $P = 1$ and $Q = 2$, and dB = 10 log (2/1) = 10 log 2 = 10 × 0.3 = 3 dB. This is the minimum detectable level of volume change if you aren't expecting it: doubling of the actual sound power!

If you turn the volume level back down again, then $P/Q = 1/2 = 0.5$, and you can calculate dB = 10 log 0.5 = 10 × −0.3 = −3 dB.

A gain or loss of 10 dB (that is, a change of +10 dB or −10 dB, often shortened to ±10 dB) represents a 10-fold increase or decrease in sound power. A change of ±20 dB represents a 100-fold increase or decrease in sound power. It is not unusual to encounter sounds that vary in intensity over ranges of ±60 dB, which represents a 1,000,000-fold increase or decrease in sound power!

Sound Power in Terms of Decibels

The preceding formula can be worked inside out, so that you can determine the final sound power, given the initial sound power and the decibel change. To do this, you use the *inverse of the logarithmic function*, symbolized as $\log^{-1}$ or antilog. This function, like the logarithmic function, can be performed by any good scientific calculator, or by the calculator program in a personal computer when set to scientific mode.

Suppose the initial sound power is P, and the change in decibels is dB. Let Q be the final sound power. Then:

$$Q = P \text{ antilog } (\text{dB}/10)$$

As an example, suppose the initial power, P, is 10 W, and the perceived volume change is −3 dB. Then the final power, Q, is equal to 10 antilog (−3/10) = 10 × 0.5 = 5 W.

Decibels in the Real World

Sound levels are sometimes specified in decibels relative to the *threshold of hearing*, defined as the faintest possible sound that a person can detect in a quiet room, assuming his or her hearing is normal. This threshold is assigned the value 0 dB. Other sound levels can then be quantified as figures such as 30 dB or 75 dB.

If a certain noise has a loudness of 30 dB, that means it's 30 dB above the threshold of hearing, or 1000 times as loud as the quietest detectable noise. A noise at 60 dB is 1,000,000 (or 10^6) times as powerful as a sound at the threshold of hearing. *Sound-level meters* are used to determine the decibel levels of various noises and acoustic environments.

A typical conversation occurs at a level of about 70 dB. This is 10,000,000 (or 10^7) times the threshold of hearing, in terms of actual sound power. The roar of the crowd at a rock concert might be 90 dB, or 1,000,000,000 (10^9) times the threshold of hearing. A sound at 100 dB, typical of the music at a large rock concert if you are sitting in the front row, is 10,000,000,000 (10^{10}) times as loud, in terms of power, as a sound at the threshold of hearing.

Resistor Specifications

When choosing a resistor for a particular application in an electrical or electronic device, it's important to get a unit that has the correct properties, or *specifications.* Here are some of the most important specifications to watch for.

Ohmic Value

In theory, a resistor can have any ohmic value from the lowest possible (such as a shaft of solid silver) to the highest (dry air). In practice, it is unusual to find resistors with values less than about 0.1 Ω or more than about 100 MΩ.

Resistors are manufactured with ohmic values in power-of-10 multiples of 1.0, 1.2, 1.5, 1.8, 2.2, 2.7, 3.3, 3.9, 4.7, 5.6, 6.8, and 8.2. Thus, you will often see resistors with values of 47 Ω, 180 Ω, 6.8 kΩ, or 18 MΩ, but hardly ever with values such as 384 Ω, 4.54 kΩ, or 7.297 MΩ.

In addition to these standard values, there are others that are used for resistors made with greater precision, or tighter *tolerance.* These are power-of-10 multiples of 1.1, 1.3, 1.6, 2.0, 2.4, 3.0, 3.6, 4.3, 5.1, 6.2, 7.5, and 9.1.

Tolerance

The first set of numbers above represents standard resistance values available in tolerances of plus or minus 10 percent (±10%). This means that the resistance might be as much as 10 percent more or 10 percent less than the indicated amount. In the case of a 470-Ω resistor, for example, the value can be larger or smaller than the rated value by as much as 47 Ω, and still be within tolerance. That's a range of 423 to 517 Ω.

Tolerance is calculated according to the specified value of the resistor, not the actual value. You might measure the value of a 470-Ω resistor and find it to be 427 Ω, and it would be within ±10% of the specified value. But if it measures 420 Ω, it's outside the rated range, and is therefore a reject. The second set, along with the first set, of numbers represents standard resistance values available in tolerances of plus or minus 5 percent (±5%). A 470-Ω, 5 percent resistor will have an actual value of 470 Ω plus or minus 24 Ω, or a range of 446 to 494 Ω.

Some resistors are available in tolerances tighter than ±5%. These precision units are employed in circuits where a little error can make a big difference. In most audio and radio-frequency oscillators and amplifiers, the ±10% or ±5% tolerance is good enough. In many cases, even a ±20% tolerance is satisfactory.

Power Rating

All resistors are given a specification that determines how much power they can safely dissipate. Typical values are ¼ W, ½ W, and 1 W. Units also exist with ratings of ⅛ W or 2 W. These dissipation ratings are for *continuous duty*, meaning they can dissipate this amount of power constantly and indefinitely.

You can figure out how much current a given resistor can handle by using the formula for power (P) in terms of current (I) and resistance (R). That formula, you should recall, is $P = I^2R$. Work this formula backward, plugging in the power rating in watts for P and the resistance in ohms for R, and solve for the current I in amperes. Alternatively, you can find the square root of P/R.

The power rating for a given resistor can, in effect, be increased by using a network of 2 × 2, 3 × 3, 4 × 4, or more units in series-parallel. If you need a 47-Ω, 45-W resistor, but all you have is a bunch of 47-Ω, 1-W resistors, you can make a 7 × 7 network in series-parallel, and this will handle 49 W.

Resistor power dissipation ratings are specified with a margin for error. A good engineer never tries to take advantage of this and use, say, a ¼-W unit in a situation that needs to draw 0.27 W. In fact, good engineers usually include their own safety margin. Allowing 10 percent, a ¼-W resistor should not be called upon to handle more than about 0.225 W.

Temperature Compensation

All resistors change value when the temperature changes dramatically. And because resistors dissipate power, they can get hot just because of the current they carry. Often, this current is so tiny that it doesn't appreciably heat the resistor. But in some cases it does, and the resistance will change. Then a circuit might behave differently than it did when the resistor was still cool.

There are various ways to approach problems of resistors changing value when they get hot. One method is to use specially manufactured resistors that do not appreciably change value when they get hot. Such units are called *temperature-compensated.* But one of these can cost several times as much as an ordinary resistor. Another approach is to use a power rating that is much higher than the actual dissipated power in the resistor. This will keep the resistor from getting very hot. Still another scheme is to use a series-parallel network of identical resistors to increase the power dissipation rating. Alternatively, you can take several resistors, say three of them, each with about three times the intended resistance, and connect them all in parallel. Or you can take several resistors, say four of them, each with about one-fourth the intended resistance, and connect them in series.

It is unwise to combine resistors with different values. This can result in one of them taking most of the load while the others "loaf," and the combination will be no better than the single hot resistor you started with.

How about using two resistors with half (or twice) the value you need, but with *opposite* resistance-versus-temperature characteristics, and connecting them in series or parallel? It is tempting to suppose that if you do this, the component whose resistance decreases with heat (*negative temperature coefficient*) will have a canceling-out effect on the component whose resistance goes up (*positive temperature coefficient*). This can sometimes work, but in practice it's difficult to find a pair of resistances that will do this job just right.

The Color Code for Resistors

Some resistors have *color bands* that indicate their values and tolerances. You'll see three, four, or five bands around carbon-composition resistors and film resistors. Other units are large enough so that the values can be printed on them in ordinary numerals.

On resistors with *axial leads* (wires that come straight out of both ends), the first, second, third, fourth, and fifth bands are arranged as shown in Fig. 6-12A. On resistors with *radial leads* (wires that come off the ends at right angles to the axis of the component body), the colored regions are arranged as shown in Fig. 6-12B. The first two regions represent numbers 0 through 9, and the third region represents a multiplier of 10 to some power. (For the moment, don't worry about the fourth and fifth regions.) Refer to Table 6-1.

Suppose you find a resistor whose first three bands are yellow, violet, and red, in that order. Then the resistance is 4700 Ω. Read yellow = 4, violet = 7, red = ×100. As another example, suppose you find a resistor with bands of blue, gray, orange. Refer to Table 6-1 and determine blue = 6, gray = 8, orange = ×1000. Therefore, the value is 68,000 Ω = 68 kΩ.

The fourth band, if there is one, indicates tolerance. If it's silver, it means the resistor is rated at ±10%. If it's gold, the resistor is rated at ±5%. If there is no fourth band, the resistor is rated at ±20%.

The fifth band, if there is one, indicates the maximum percentage that the resistance can be expected to change after 1000 hours of use. A brown band indicates a maximum change of ±1% of the rated value. A red band indicates ±0.1%. An orange band indicates ±0.01%. A yellow band indicates ±0.001%. If there is no fifth band, it means that the resistor might deviate by more than ±1% of the rated value after 1000 hours of use.

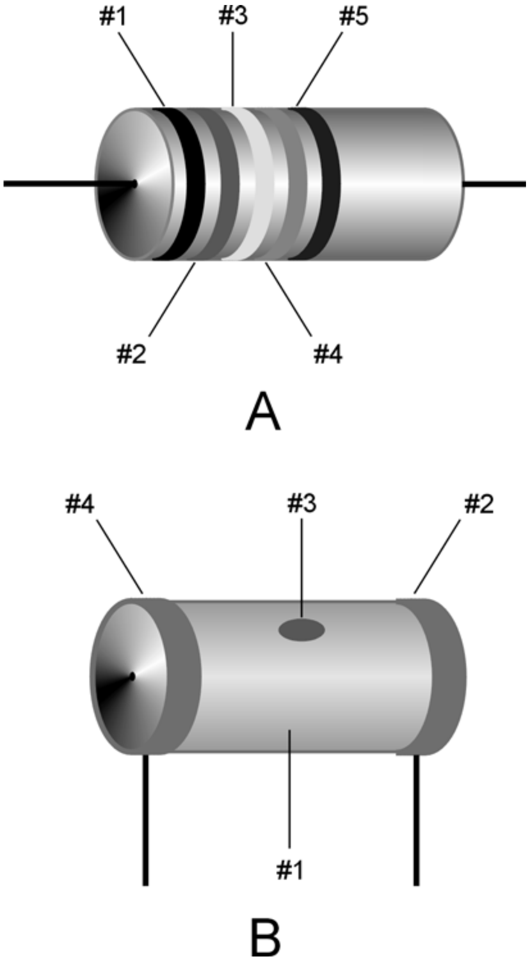

6-12 At A, locations of color-code bands on a resistor with axial leads. At B, locations of color code designators on a resistor with radial leads.

Table 6-1. The color code for the first three bands that appear on fixed resistors. See text for discussion of the fourth and fifth bands.

Color of band	Numeral (first and second bands)	Multiplier (third band)
Black	0	1
Brown	1	10
Red	2	100
Orange	3	1000 (1 k)
Yellow	4	10^4 (10 k)
Green	5	10^5 (100 k)
Blue	6	10^6 (1 M)
Violet	7	10^7 (10 M)
Gray	8	10^8 (100 M)
White	9	10^9 (1000 M or 1 G)

A competent engineer or technician always tests a resistor with an ohmmeter before installing it in a circuit. If the component happens to be labeled wrong, or if it is defective, it's easy to catch this problem while assembling or servicing a circuit. But once the circuit is all together, and it won't work because some resistor is labeled wrong or is bad, it's difficult to troubleshoot.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. Proper biasing in an amplifier circuit
   (a) causes it to oscillate.
   (b) prevents an impedance match.
   (c) can be obtained using a voltage divider network.
   (d) maximizes current flow.
2. A transistor can be protected from needless overheating by
   (a) a current-limiting resistor.
   (b) bleeder resistors.
   (c) maximizing the drive.
   (d) shorting out the power supply when the circuit is off.
3. A bleeder resistor
   (a) is connected across the capacitor in a power supply.
   (b) keeps a transistor from drawing too much current.
   (c) prevents an amplifier from being overdriven.
   (d) optimizes the efficiency of an amplifier.
4. Carbon-composition resistors
   (a) can handle gigantic levels of power.
   (b) have capacitance or inductance along with resistance.
   (c) have essentially no capacitance or inductance.
   (d) work better for ac than for dc.
5. A logical place for a wirewound resistor is
   (a) in a radio-frequency amplifier.
   (b) in a circuit where a noninductive resistor is called for.
   (c) in a low-power radio-frequency circuit.
   (d) in a high-power dc circuit.
6. A metal-film resistor
   (a) is made using a carbon-based paste.
   (b) does not have much inductance.

(c) can dissipate large amounts of power.
(d) has considerable inductance.

7. What type of resistor, or combination of resistors, would you use as the meter-sensitivity control in a test instrument, when continuous adjustment is desired?

(a) A set of switchable, fixed resistors
(b) A linear-taper potentiometer
(c) An audio-taper potentiometer
(d) A wirewound resistor

8. What type of resistor, or combination of resistors, would you use as the volume control in a stereo compact-disc (CD) player?

(a) A set of switchable, fixed resistors
(b) A linear-taper potentiometer
(c) An audio-taper potentiometer
(d) A wirewound resistor

9. If a sound triples in actual power level, approximately what is this, expressed in decibels?

(a) +3 dB
(b) +5 dB
(c) +6 dB
(d) +9 dB

10. Suppose a sound changes in volume by −13 dB. If the original sound power is 1.0 W, what is the final sound power?

(a) 13 W
(b) 77 mW
(c) 50 mW
(d) There is not enough information given here to answer this question.

11. The sound from a portable radio is at a level of 50 dB. How many times the threshold of hearing is this, in terms of actual sound power?

(a) 50
(b) 169
(c) 5000
(d) 100,000

12. An advantage of a rheostat over a potentiometer is the fact that

(a) a rheostat can handle higher frequencies.
(b) a rheostat is more precise.
(c) a rheostat can handle more current.
(d) a rheostat works better with dc.

13. A resistor is specified as having a value of 68 Ω, but is measured with an ohmmeter as 63 Ω. The value is off by which of the following percentages?

(a) 7.4%
(b) 7.9%
(c) 5%
(d) 10%

14. Suppose a resistor is rated at 3.3 kΩ ±5%. This means it can be expected to have a value between

(a) 2970 Ω and 3630 Ω.
(b) 3295 Ω and 3305 Ω.
(c) 3135 Ω and 3465 Ω.
(d) 2.8 kΩ and 3.8 kΩ.

15. A package of resistors is rated at 56 Ω ±10%. You test them with an ohmmeter. Which of the following values indicates a reject?

(a) 50.0 Ω
(b) 53.0 Ω
(c) 59.7 Ω
(d) 61.1 Ω

16. A resistor has a value of 680 Ω, and you expect that it will have to draw 1 mA maximum continuous current in a circuit you're building. What power rating is good for this application, but not needlessly high?

(a) ¼ W
(c) ½ W
(c) 1 W
(d) 2 W

17. Suppose a 1-kΩ resistor will dissipate 1.05 W, and you have a good supply of 1-W resistors of various ohmic values. If there's room for 20 percent resistance error, the cheapest solution is to use

(a) four 1-kΩ, 1-W resistors in series-parallel.
(b) a pair of 2.2-kΩ, 1-W resistors in parallel.
(c) a set of three 3.3-kΩ, 1-W resistors in parallel.
(d) a single 1-kΩ, 1-W resistor, because all manufacturers allow for a 10 percent margin of safety when rating resistors for their power-handling capability.

18. Suppose a carbon-composition resistor has the following colored bands on it: red, red, red, gold. This indicates a resistance of

(a) 22 Ω.
(b) 220 Ω.
(c) 2.2 kΩ.
(d) 22 kΩ.

19. The actual resistance of the component described in the previous question can be expected to vary above or below the specified ohmic value by up to what amount?

    (a) 11 Ω
    (b) 110 Ω
    (c) 22 Ω
    (d) 220 Ω

20. Suppose a carbon-composition resistor has the following colored bands on it: gray, red, yellow. This unit can be expected to have a value within approximately what range?

    (a) 660 kΩ to 980 kΩ
    (b) 740 kΩ to 900 kΩ
    (c) 7.4 kΩ to 9.0 kΩ
    (d) The manufacturer does not make any claim.

7
CHAPTER

Cells and Batteries

IN ELECTRICITY AND ELECTRONICS, A *CELL* IS A UNIT SOURCE OF DC ENERGY. WHEN TWO OR MORE cells are connected in series, the result is known as a *battery.* There are many types of cells and batteries, and new types are constantly being invented.

Electrochemical Energy

Early in the history of electrical science, laboratory physicists found that when metals came into contact with certain chemical solutions, voltages appeared between the pieces of metal. These were the first *electrochemical cells.*

A piece of lead and a piece of lead dioxide immersed in an acid solution (Fig. 7-1) acquire a persistent potential difference. This can be detected by connecting a galvanometer between the pieces of metal. A resistor of about 1000 Ω must be used in series with the galvanometer in experiments of this kind, because connecting the galvanometer directly will cause too much current to flow, possibly damaging the galvanometer and causing the acid to boil.

The chemicals and the metal have an inherent ability to produce a constant exchange of charge carriers. If the galvanometer and resistor are left hooked up between the two pieces of metal for a long time, the current will gradually decrease, and the electrodes will become coated. All the *chemical energy* in the acid will have been turned into electrical energy as current in the wire and galvanometer. In turn, this current will have heated the resistor (another form of kinetic energy), and escaped into the air and into space.

Primary and Secondary Cells

Some electrical cells, once their chemical energy has all been changed to electricity and used up, must be thrown away. These are called *primary cells.* Other kinds of cells, such as the lead-and-acid type, can get their chemical energy back again by means of *recharging.* Such a cell is a *secondary cell.*

Primary cells include the ones you usually put in a flashlight, in a transistor radio, and in various other consumer devices. They use dry electrolyte pastes along with metal electrodes. They go by names such as *dry cell*, *zinc-carbon cell*, or *alkaline cell.* Go into a department store and find a rack of batteries, and you'll see various sizes and types of primary cells, such as *AAA batteries*, *D batteries*,

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

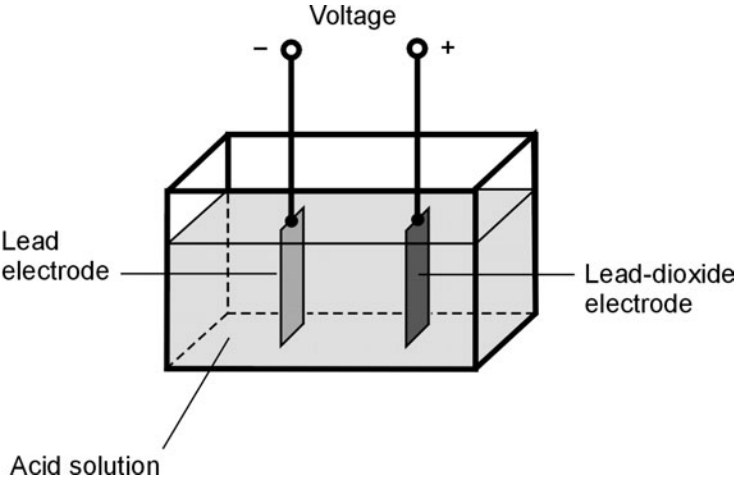

7-1 Simplified drawing of the construction of a lead-acid electrochemical cell.

camera batteries, and *watch batteries.* (These are actually cells, not true batteries.) You'll also see 9-V *transistor batteries* and large 6-V *lantern batteries.*

Secondary cells can also be found in consumer stores. *Nickel-based cells* are common. The most common sizes are AA, C, and D. These cost several times as much as ordinary dry cells, and a charging unit also costs a few dollars. But if you take care of them, these rechargeable cells can be used hundreds of times and will pay for themselves several times over if you use a lot of batteries in everyday life.

The battery in your car is made from secondary cells connected in series. These cells recharge from the alternator or from an outside charging unit. This battery has cells like the one in Fig. 7-1. It is dangerous to short-circuit the terminals of such a battery, because the acid (sulfuric acid) can bubble up and erupt out of the battery casing. Serious skin and eye injuries can result. In fact, it's a bad idea to short-circuit any cell or battery, because it can get extremely hot and cause a fire, or rupture and damage surrounding materials, wiring, and components.

The Weston Standard Cell

Most electrochemical cells produce 1.2 to 1.8 V. Different types vary slightly. A mercury cell has a voltage that is a little less than that of a zinc-carbon or alkaline cell. The voltage of a cell can also be affected by variables in the manufacturing process. Most consumer-type dry cells can be assumed to produce 1.5 V.

There are certain cells whose voltages are predictable and exact. These are called *standard cells.* A good example is the *Weston cell*, which produces 1.018 V at room temperature. It has a solution of cadmium sulfate, a positive electrode made from mercury sulfate, and a negative electrode made from mercury and cadmium. The device is set up in a container, as shown in Fig. 7-2.

Storage Capacity

Recall that the common electrical units of energy are the watt-hour (Wh) and the kilowatt-hour (kWh). Any electrochemical cell or battery has a certain amount of electrical energy that can be obtained from it, and this can be specified in watt-hours or kilowatt-hours. More often, though, it's given in *ampere-hours* (Ah).

A battery with a rating of 2 Ah can provide 2 A for 1 h, or 1 A for 2 h, or 100 mA for 20 h.

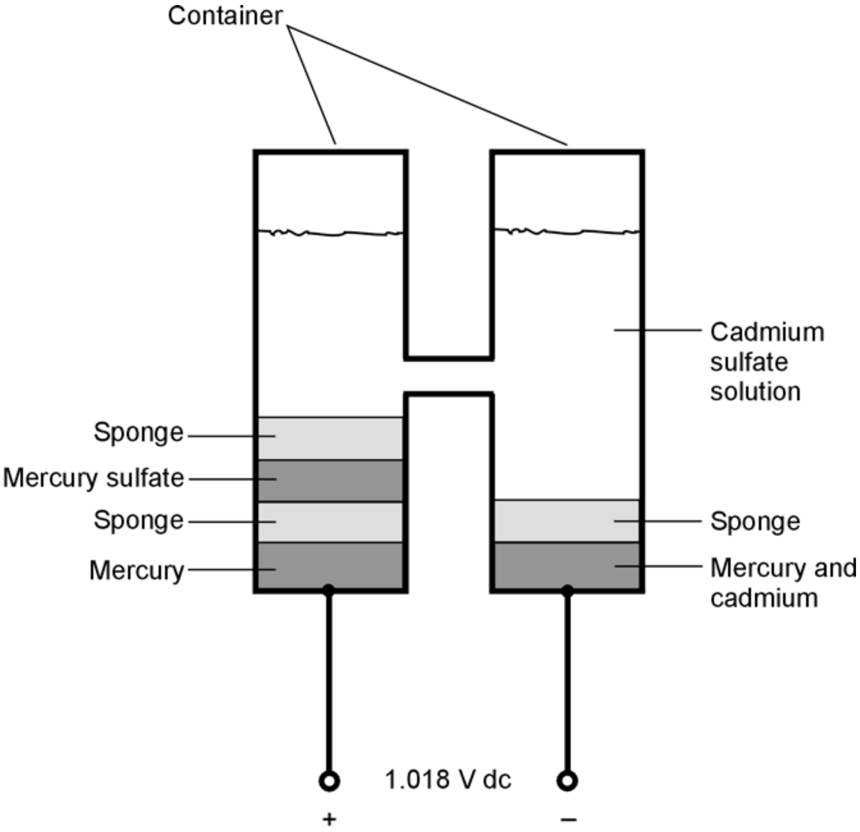

7-2 Simplified drawing of the construction of a Weston standard cell.

There are infinitely many possibilities here, as long as the product of the current in amperes and the use time in hours is equal to 2. The limitations are the *shelf life* at one extreme, and the *maximum deliverable current* at the other. Shelf life is the length of time the battery will last if it is never used; this can be years. The maximum deliverable current is the highest amount of current that the battery can provide before its voltage drops because of its own *internal resistance.*

Small cells have storage capacity of a few milliampere-hours (mAh) up to 100 or 200 mAh. Medium-sized cells can supply 500 mAh to 1 Ah. Large automotive or truck batteries can provide upward of 50 Ah. The energy capacity in watt-hours is the ampere-hour capacity multiplied by the battery voltage.

An *ideal cell* or *ideal battery* (a theoretically perfect cell or battery) delivers a constant current for a while, and then the current starts to drop (Fig. 7-3). Some types of cells and batteries approach this level of perfection, which is represented by a *flat discharge curve.* But many cells and batteries are far from perfect; they deliver current that declines gradually, almost right from the start. When the current that a battery can provide has tailed off to about half of its initial value, the cell or battery is said to be *weak.* At this time, it should be replaced. If it's allowed to run all the way out, until the current actually goes to zero, the cell or battery is *dead.* The area under the curve in Fig. 7-3 is a graphical representation the total capacity of the cell or battery in ampere-hours.

7-3 A flat discharge curve. This is considered ideal.

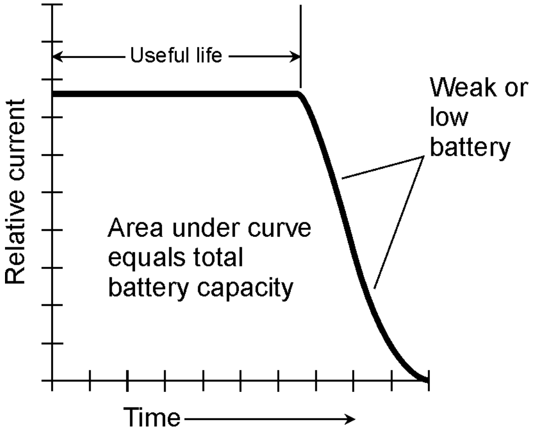

Grocery Store Cells and Batteries

The cells you see in grocery stores, department stores, drugstores, and hardware stores provide 1.5 V, and are available in sizes known as AAA (very small), AA (small), C (medium large), and D (large). Batteries are widely available that deliver 6 or 9 V.

Zinc-Carbon Cells

Figure 7-4 is a translucent drawing of a *zinc-carbon cell.* The zinc forms the case and is the negative electrode. A carbon rod serves as the positive electrode. The electrolyte is a paste of manganese dioxide and carbon. Zinc-carbon cells are inexpensive and are good at moderate temperatures and in applications where the current drain is moderate to high. They are not very good in extreme cold.

Alkaline Cells

The *alkaline cell* has granular zinc as the negative electrode, potassium hydroxide as the electrolyte, and a device called a *polarizer* as the positive electrode. The construction is similar to that of the zinc-carbon cell. An alkaline cell can work at lower temperatures than a zinc-carbon cell. It lasts longer in most electronic devices, and is therefore preferred for use in transistor radios, calculators,

7-4 Simplified drawing of the construction of a zinc-carbon electrochemical cell.

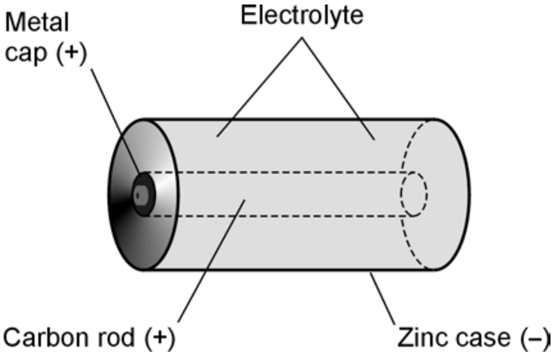

and portable cassette players. Its shelf life is much longer than that of a zinc-carbon cell. As you might expect, it costs more.

Transistor Batteries

A *transistor battery* consists of six tiny zinc-carbon or alkaline cells in series. Each of the six cells supplies 1.5 V. Thus, the battery supplies 9 V. Even though these batteries have more voltage than individual cells, the total energy available from them is less than that from a C cell or D cell. This is because the electrical energy that can be obtained from a cell or battery is directly proportional to the amount of chemical energy stored in it, and this, in turn, is a direct function of the *volume* (physical size) of the cell or the *mass* (quantity of chemical matter) of the cell. Cells of size C or D have more volume and mass than a transistor battery, and therefore contain more stored energy for the same chemical composition.

Transistor batteries are used in low-current electronic devices such as remote-control garage-door openers, television (TV) and hi-fi remote controls, and electronic calculators.

Lantern Batteries

The *lantern battery* has much greater mass than a common dry cell or transistor battery, and consequently it lasts much longer and can deliver more current. Lantern batteries are usually rated at 6 V, and consist of four good-size zinc-carbon or alkaline cells. Two lantern batteries connected in series make a 12-V battery that can power a 5-W citizens band (CB) or ham radio transceiver for a while. They're also good for scanner radio receivers in portable locations, for camping lamps, and for other medium-power needs.

Miniature Cells and Batteries

In recent years, cells and batteries—especially cells—have become available in many different sizes and shapes besides the old cylindrical cells, transistor batteries, and lantern batteries. These are used in wristwatches, small cameras, and various microminiature electronic devices.

Silver-Oxide Cells and Batteries

A *silver-oxide cell* is usually found in a buttonlike shape, and can fit inside a small wristwatch. These types of cells come in various sizes and thicknesses, all with similar appearances. They supply 1.5 V, and offer excellent energy storage for the weight. They also have a nearly flat discharge curve, like the one shown in the graph of Fig. 7-3. Zinc-carbon and alkaline cells and batteries, in contrast, have current output that declines more steadily with time, as shown in Fig. 7-5. This is known as a *declining discharge curve.*

Silver-oxide cells can be stacked to make batteries. Several of these miniature cells, one on top of the other, can provide 6, 9, or even 12 V for a transistor radio or other light-duty electronic device. The resulting battery is about the size of an AAA cylindrical cell.

Mercury Cells and Batteries

A *mercury cell*, also called a *mercuric-oxide cell*, has properties similar to those of silver-oxide cells. They are manufactured in the same general form. The main difference, often not of significance, is a somewhat lower voltage per cell: 1.35 V. If six of these cells are stacked to make a battery, the re-

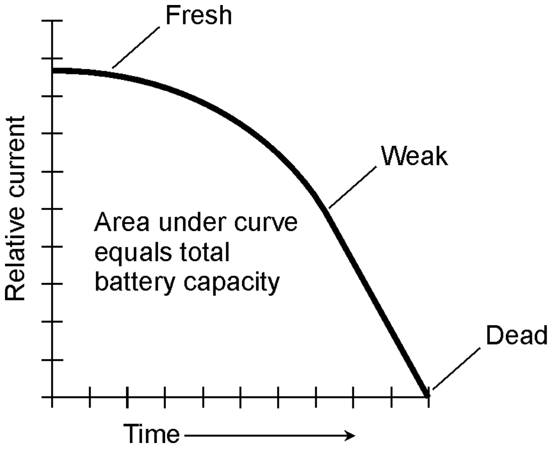

7-5 A declining discharge curve.

sulting voltage will be about 8.1 V rather than 9 V. One additional cell can be added to the stack, yielding about 9.45 V.

There has been a decline in the popularity of mercury cells and batteries in recent years, because of the fact that mercury is toxic to humans and animals, even in trace amounts. When mercury cells and batteries are dead, they must be discarded. Eventually the mercury or mercuric oxide leaks into the soil and groundwater. Mercury pollution has become a significant concern throughout the world.

Lithium Cells and Batteries

Lithium cells gained popularity in the early 1980s. There are several variations in the chemical makeup of these cells; they all contain lithium, a light, highly reactive metal. Lithium cells can be made to supply 1.5 to 3.5 V, depending on the particular chemistry used. These cells, like silver-oxide and mercury cells, can be stacked to make batteries.

The first application of lithium batteries was in memory backup for electronic microcomputers. Lithium cells and batteries have superior shelf life, and they can last for years in very-low-current applications such as memory backup or the powering of a digital liquid crystal display (LCD) watch or clock. These cells also provide high energy capacity per unit volume or mass.

Lead-Acid Batteries

You've seen the basic configuration for a lead-acid cell. This has a solution of sulfuric acid, along with a lead electrode (negative) and a lead-dioxide electrode (positive). These cells are rechargeable.

Automotive batteries are made from sets of lead-acid cells having a free-flowing liquid acid. You cannot tip such a battery on its side, or turn it upside-down, without running the risk of having some of the acid electrolyte spill out. Lead-acid batteries are also available in a construction that uses a semisolid electrolyte. These batteries are sometimes used in consumer electronic devices that require a moderate amount of current. The most common example is an uninterruptible power supply (UPS) that can keep a desktop personal computer running for a few minutes if the utility power fails.

A large lead-acid battery, such as the kind in your car or truck, can store several tens of ampere-hours. The smaller ones, like those in a UPS, have less capacity but more versatility. Their main attributes are that they can be charged and recharged many times, and they are not particularly expensive.

Nickel-Based Cells and Batteries

Nickel-based cells include the *nickel-cadmium* (NICAD or NiCd) type and the *nickel-metal-hydride* (NiMH) type. *Nickel-based batteries* are available in packs of cells. These packs can be plugged into equipment, and sometimes form part of the case for a device such as a portable radio transceiver. All nickel-based cells are rechargeable, and can be put through hundreds or even thousands of *charge/discharge cycles* if they are properly cared for.

Configurations and Applications

Nickel-based cells are found in various sizes and shapes. *Cylindrical cells* look like ordinary dry cells. *Button cells* are those little things you find in cameras, watches, memory backup applications, and other places where miniaturization is important. *Flooded cells* are used in heavy-duty applications, and can have storage capacity in excess of 1000 Ah. *Spacecraft cells* are made in packages that can withstand the rigors of a deep-space environment.

Most orbiting satellites are in darkness half the time and in sunlight half the time. *Solar panels* can be used while the satellite is in sunlight, but during the times that the earth eclipses the sun, batteries are needed to power the electronic equipment on board the satellite. The solar panels can charge a nickel-based battery, in addition to powering the satellite, for the daylight half of each orbit. The nickel-based battery can provide the power during the dark half of each orbit.

Cautions

Never discharge nickel-based cells all the way until they totally die. This can cause the polarity of a cell, or of one or more cells in a battery, to reverse. Once this happens, the cell or battery is ruined.

A phenomenon peculiar to nickel-based cells and batteries is known as *memory* or *memory drain.* If a nickel-based unit is used over and over, and is discharged to the same extent every time, it might begin to die at that point in its discharge cycle. Memory problems can usually be solved. Use the cell or battery almost all the way up, and then fully recharge it. Repeat the process several times.

Nickel-based cells and batteries work best if used with charging units that take several hours to fully replenish the charge. So-called high-rate or quick chargers are available, but these can sometimes force too much current through a cell or battery. It's best if the charger is made especially for the cell or battery type being charged. An electronics dealer, such as the manager at a RadioShack store, should be able to tell you which chargers are best for which cells and batteries.

In recent years, concern has grown about the toxic environmental effects of discarded heavy metals, including cadmium. For this reason, NiMH cells and batteries have replaced NICAD types in many applications. In most practical scenarios, a NICAD battery can be directly replaced with a NiMH battery of the same voltage and current-delivering capacity, and the powered-up device will work satisfactorily.

Some vendors and dealers will call a nickel-based cell or battery a NICAD, even when it is actually a NiMH cell or battery.

Photovoltaic Cells and Batteries

The *photovoltaic* (PV) *cell* is different from any electrochemical cell. It's also known as a *solar cell.* This device converts visible light, infrared (IR), and/or ultraviolet (UV) directly into electric current.

Solar Panels

Several, or many, photovoltaic cells can be combined in series-parallel to make a *solar panel.* An example is shown in Fig. 7-6. Although this shows a 3 × 3 series-parallel array, the matrix does not have to be symmetrical. And it's often very large. It might consist of, say, 50 parallel sets of 20 series-connected cells. The series scheme boosts the voltage to the desired level, and the parallel scheme increases the current-delivering ability of the panel. It's not unusual to see hundreds of solar cells combined in this way to make a large panel.

Construction and Performance

The construction of a photovoltaic cell is shown in Fig. 7-7. The device is a flat semiconductor *P-N junction*, and the assembly is made transparent so that light can fall directly on the *P-type* silicon. The metal ribbing, forming the positive electrode, is interconnected by means of tiny wires. The negative electrode is a metal backing or *substrate*, placed in contact with the *N-type* silicon.

Most solar cells provide about 0.5 V. If there is very low current demand, dim light will result in the full-output voltage from a solar cell. As the current demand increases, brighter light is needed to produce the full-output voltage. There is a maximum limit to the current that can be provided from a solar cell, no matter how bright the light. This limit is increased by connecting solar cells in parallel.

Practical Applications

Solar cells have become cheaper and more efficient in recent years, as researchers have looked to them as an alternative energy source. Solar panels are used in satellites. They can be used in conjunction with rechargeable batteries, such as the lead-acid or nickel-cadmium types, to provide power independent of the commercial utilities.

A completely independent solar/battery power system is called a *stand-alone* system. It uses large solar panels, large-capacity lead-acid batteries, power converters to convert the dc into ac, and a sophisticated charging circuit. These systems are best suited to environments where there is sunshine a high percentage of the time.

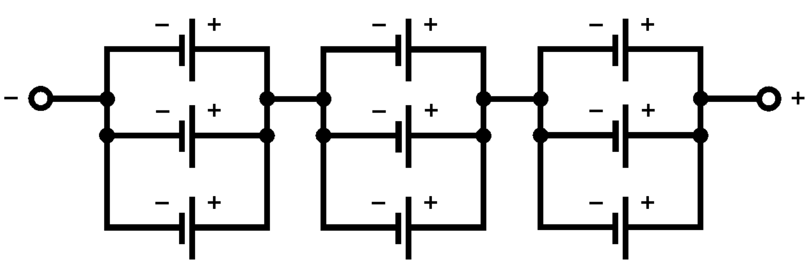

7-6 Connection of cells in series-parallel.

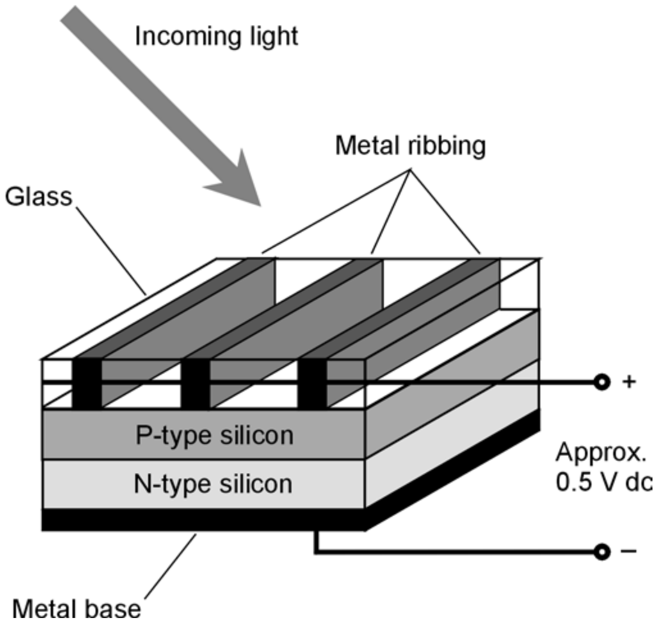

7-7 Construction of a silicon photovoltaic cell.

Solar cells, either alone or supplemented with rechargeable batteries, can be connected into a home electric system in an *interactive* arrangement with the electric utilities. When the solar power system can't provide for the needs of the household all by itself, the utility company can take up the slack. Conversely, when the solar power system supplies more than enough for the needs of the home, the utility company can buy the excess.

Fuel Cells

In the late 1900s, a new type of electrochemical power device emerged that is believed by some scientists and engineers to hold promise as an alternative energy source: the *fuel cell.*

Hydrogen Fuel

The most talked-about fuel cell during the early years of research and development became known as the *hydrogen fuel cell.* As its name implies, it derives electricity from hydrogen. The hydrogen combines with oxygen (that is, it *oxidizes*) to form energy and water. There is no pollution, and there are no toxic by-products. When a hydrogen fuel cell "runs out of juice," all that is needed is a new supply of hydrogen, because its oxygen is derived from the atmosphere.

Instead of combusting, the hydrogen in a fuel cell oxidizes in a more controlled fashion, and at a much lower temperature. There are several schemes for making this happen. The *proton exchange membrane* (PEM) *fuel cell* is one of the most widely used. A PEM hydrogen fuel cell generates approximately 0.7 V of dc. In order to obtain higher voltages, individual cells are connected in series. A series-connected set of fuel cells is technically a battery, but the term used more often is *stack.*

Fuel-cell stacks are available in various sizes. A stack about the size and weight of an airline suitcase filled with books can power a subcompact electric car. Smaller cells, called *micro fuel cells*, can provide dc to run devices that have historically operated from conventional cells and batteries. These include portable radios, lanterns, and notebook computers.

Other Fuels

Hydrogen is not the only chemical that can be used to make a fuel cell. Almost anything that will combine with oxygen to form energy has been considered.

Methanol, a form of alcohol, has the advantage of being easier to transport and store than hydrogen, because it exists as a liquid at room temperature. *Propane* is another chemical that has been used for powering fuel cells. This is the substance that is stored in liquid form in tanks for barbecue grills and some rural home heating systems. *Methane*, also known as natural gas, has been used as well.

Some scientists and engineers object to the use of these fuels because they, especially propane and methane, closely resemble fuels that are already commonplace, and on which society has developed the sort of dependence that purists would like to get away from. In addition, they are derived from so-called *fossil fuel* sources, the supplies of which, however great they might be today, are nevertheless finite.

A Promising Technology

As of this writing (2006), fuel cells have not yet replaced conventional electrochemical cells and batteries. Cost is the main reason. Hydrogen is the most abundant and simplest chemical element in the universe, and it does not produce any toxic by-products. This would at first seem to make it the ideal choice for use in fuel cells. But storage and transport of hydrogen has proven to be difficult and expensive. This is especially true for fuel cells and stacks intended for systems that aren't fixed to permanent pipelines.

An interesting scenario, suggested by one of my physics teachers all the way back in the 1970s, is the piping of hydrogen gas through the lines designed to carry methane. Some modification of existing lines would be required in order to safely handle hydrogen, which escapes through small cracks and openings more easily than methane. But hydrogen, if obtained at reasonable cost and in abundance, could be used to power large fuel-cell stacks in common households and businesses. The dc from such a stack could be converted to utility ac by power inverters similar to those used with PV energy systems. The entire home power system would be about the size of a gas furnace.

Quiz

Refer to the text in this chapter if necessary. A good score is 18 correct. Answers are in the back of the book.

1. The chemical energy in a battery or cell
    (a) is a form of kinetic energy.
    (b) cannot be replenished once it is gone.
    (c) changes to electrical energy when the cell is used.
    (d) is caused by electric current.

2. A cell that cannot be recharged is known as
    (a) a dry cell.
    (b) a wet cell.
    (c) a primary cell.
    (d) secondary cell.
3. A Weston cell is generally used
    (a) as a current reference source.
    (b) as a voltage reference source.
    (c) as a power reference source.
    (d) as a fuel cell.
4. The voltage produced by a battery of multiple cells connected in series is
    (a) less than the voltage produced by a cell of the same composition.
    (b) the same as the voltage produced by a cell of the same composition.
    (c) more than the voltage produced by a cell of the same composition.
    (d) always a whole-number multiple of 1.018 V.
5. A direct short-circuit of a large battery can cause
    (a) an increase in its voltage.
    (b) no harm other than a rapid discharge of its energy.
    (c) the current to drop to zero.
    (d) a physical rupture or explosion.
6. Suppose a cell of 1.5 V delivers 100 mA for 7 hours and 20 minutes, and then it is replaced. How much energy is supplied during this time?
    (a) 0.49 Wh
    (b) 1.1 Wh
    (c) 7.33 Wh
    (d) 733 mWh
7. Suppose a 12-V automotive battery is rated at 36 Ah. If a 100-W, 12-V bulb is connected across this battery, approximately how long will the bulb stay aglow, assuming the battery has been fully charged?
    (a) 4 hours and 20 minutes
    (b) 432 hours
    (c) 3.6 hours
    (d) 21.6 minutes
8. Alkaline cells
    (a) are cheaper than zinc-carbon cells.
    (b) generally work better in radios than zinc-carbon cells.
    (c) have higher voltages than zinc-carbon cells.
    (d) have shorter shelf lives than zinc-carbon cells.

9. The energy in a cell or battery depends mainly on
   (a) its physical size.
   (b) the current drawn from it.
   (c) its voltage.
   (d) all of the above.

10. In which of the following devices would a lantern battery most likely be found?
   (a) A heart pacemaker
   (b) An electronic calculator
   (c) An LCD wall clock
   (d) A two-way portable radio

11. In which of the following devices would a transistor battery be the best power choice?
   (a) A heart pacemaker
   (b) An electronic calculator
   (c) An LCD wall clock
   (d) A two-way portable radio

12. For which of the following applications would you choose a lithium battery?
   (a) A microcomputer memory backup
   (b) A two-way portable radio
   (c) A stand-alone solar-electric system
   (d) A rechargeable lantern

13. Where would you most likely find a lead-acid battery?
   (a) In a portable audio CD player
   (b) In an uninterruptible power supply
   (c) In an LCD wall clock
   (d) In a flashlight

14. A cell or battery that maintains a constant current-delivering capability almost until it dies is said to have
   (a) a large ampere-hour rating.
   (b) excellent energy capacity.
   (c) a flat discharge curve.
   (d) good energy storage capacity per unit volume.

15. Where might you find a nickel-based battery?
   (a) In a satellite
   (b) In a portable cassette player
   (c) In a handheld radio transceiver
   (d) More than one of the above

16. A disadvantage of mercury cells and batteries is the fact that
    - (a) they don't last as long as other types.
    - (b) they have a flat discharge curve.
    - (c) mercury is destructive to the environment.
    - (d) they need to be recharged often.
17. Which kind of battery should never be used until it dies?
    - (a) Silver-oxide
    - (b) Lead-acid
    - (c) Nickel-based
    - (d) Mercury
18. The useful current that is delivered by a solar panel can be increased by
    - (a) connecting capacitors in parallel with the solar cells.
    - (b) connecting resistors in series with the solar cells.
    - (c) connecting two or more groups of solar cells in parallel.
    - (d) connecting resistors in parallel with the solar cells.
19. An interactive solar power system
    - (a) allows a homeowner to sell power to the electric company.
    - (b) lets the batteries recharge at night.
    - (c) powers lights, but not electronic devices.
    - (d) is totally independent from the electric company.
20. An advantage of methanol over hydrogen for use in fuel cells is the fact that
    - (a) methanol is the most abundant element in the universe.
    - (b) methanol is not flammable.
    - (c) methanol is a solid at room temperature.
    - (d) methanol is easier to transport and store.

8
CHAPTER

Magnetism

ELECTRIC AND MAGNETIC PHENOMENA INTERACT. MAGNETISM WAS MENTIONED BRIEFLY NEAR THE end of Chap. 2. Here, we'll look at it more closely.

The Geomagnetic Field

The earth has a core made up largely of iron, heated to the extent that some of it is liquid. As the earth rotates, the iron flows in complex ways. It is thought that this flow is responsible for the magnetic field that surrounds the earth. Some other planets, notably Jupiter, have magnetic fields as well. Even the sun has one.

The Poles and Axis

The *geomagnetic field*, as it is called, has poles, just as a bar magnet does. The *geomagnetic poles* are near, but not at, the *geographic poles.* The *north geomagnetic pole* is located in far northern Canada. The *south geomagnetic pole* is near Antarctica. The *geomagnetic axis* is therefore tilted relative to the axis on which the earth rotates.

The Solar Wind

Charged subatomic particles from the sun, streaming outward through the solar system, distort the geomagnetic *lines of flux* (Fig. 8-1). This stream of particles is called the *solar wind.* That's a good name for it, because the fast-moving particles produce measurable forces on sensitive instruments in space. This force has actually been suggested as a possible means to drive space ships, equipped with *solar sails*, out of the solar system!

At and near the earth's surface, the geomagnetic field is not affected very much by the solar wind, so the geomagnetic field is nearly symmetrical. As the distance from the earth increases, the distortion of the field also increases, particularly on the side of the earth away from the sun.

The Magnetic Compass

The presence of the geomagnetic field was first noticed in ancient times. Some rocks, called *lodestones*, when hung by strings, would always orient themselves a certain way. This was correctly at-

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

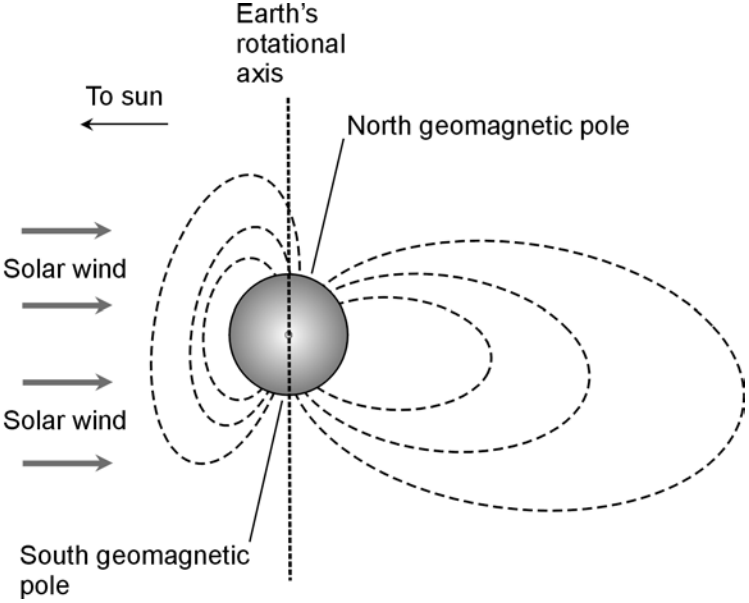

8-1 Geomagnetic flux lines (dashed curves) are distorted by the solar wind, so the geomagnetic field is not symmetrical with respect to the earth.

tributed to the presence of a "force" in the air. This effect was put to use by early seafarers and land explorers. Today, a *magnetic compass* can still be a valuable navigation aid, used by mariners, backpackers, and others who travel far from familiar landmarks.

The geomagnetic field interacts with the magnetic field around a compass needle, and a force is thus exerted on the needle. This force works not only in a horizontal plane (parallel to the earth's surface), but vertically at most latitudes. The vertical component is zero only at the *geomagnetic equator*, a line running around the globe equidistant from both geomagnetic poles.

As the geomagnetic latitude increases, toward either the north or the south geomagnetic pole, the magnetic force pulls up and down on the compass needle more and more. One end of the needle seems to insist on touching the compass face, while the other end tilts up toward the glass. The needle tries to align itself parallel to the geomagnetic *lines of flux.* The vertical angle, in degrees, at which the geomagnetic lines of flux intersect the earth's surface at any given location is called the *geomagnetic inclination.*

Because geomagnetic north is not the same as geographic north in most places on the earth's surface, there is an angular difference between the two. This horizontal angle, in degrees, is called *geomagnetic declination.* It, like inclination, varies with location.

Causes and Effects

Magnets are attracted to some, but not all, metals. Iron, nickel, and alloys containing either or both of these elements are known as *ferromagnetic* materials. They "stick" to magnets. They can also be made into *permanent magnets.* When a magnet is brought near a piece of ferromagnetic material, the atoms in the material become lined up, so that the material is temporarily magnetized. This produces a *magnetic force* between the atoms of the ferromagnetic substance and those in the magnet.

Attraction and Repulsion

If a magnet is brought near another magnet, the force can be repulsive or attractive, depending on the way the magnets are oriented. The force gets stronger as the magnets are brought near each other. Some magnets are so strong that no human being can pull them apart if they get stuck together, and no person can bring them all the way together against their mutual repulsive force. This is especially true of *electromagnets*, discussed later in this chapter.

The tremendous forces produced by electromagnets are of use in industry. A large electromagnet can be used to carry heavy pieces of scrap iron from place to place. Other electromagnets can provide sufficient repulsion to suspend one object above another. This phenomenon is called *magnetic levitation.* It is the basis for low-friction, high-speed commuter trains now in use in some metropolitan areas.

Charge in Motion

Whenever the atoms in a ferromagnetic material are aligned, a *magnetic field* exists. A magnetic field can also be caused by the motion of electric charge carriers, either in a wire or in free space.

The magnetic field around a permanent magnet arises from the same cause as the field around a wire that carries an electric current. The responsible factor in either case is the motion of electrically charged particles. In a wire, electrons move along the conductor, being passed from atom to atom. In a permanent magnet, the movement of orbiting electrons occurs in such a manner that an effective electrical current is produced.

Magnetic fields are also generated by the motion of charged particles through space. The sun is constantly ejecting protons and helium nuclei. These particles carry a positive electric charge. Because of this, and the fact that they are in motion, they are surrounded by tiny magnetic fields. When the particles approach the earth and their magnetic fields interact with the geomagnetic field, the particles are accelerated toward the geomagnetic poles.

When there is a *solar flare*, the sun ejects far more charged particles than normal. When these approach the geomagnetic poles, the result is considerable disruption of the geomagnetic field. This type of event is called a *geomagnetic storm.* It causes changes in the earth's ionosphere, affecting long-distance radio communications at certain frequencies. If the fluctuations are intense enough, even wire communications and electric power transmission can be interfered with. Aurora (northern or southern lights) are frequently observed at night during these events.

Flux Lines

Have you seen the well-known experiment in which iron filings are placed on a horizontal sheet of paper, and then a magnet is placed underneath the paper? The filings arrange themselves in a pattern that shows, roughly, the shape of the magnetic field in the vicinity of the magnet. A bar magnet has a field with a characteristic form (Fig. 8-2). Another popular experiment involves passing a current-carrying wire through a horizontal sheet of paper at a right angle, as shown in Fig. 8-3. The iron filings become grouped along circles centered at the point where the wire passes through the paper.

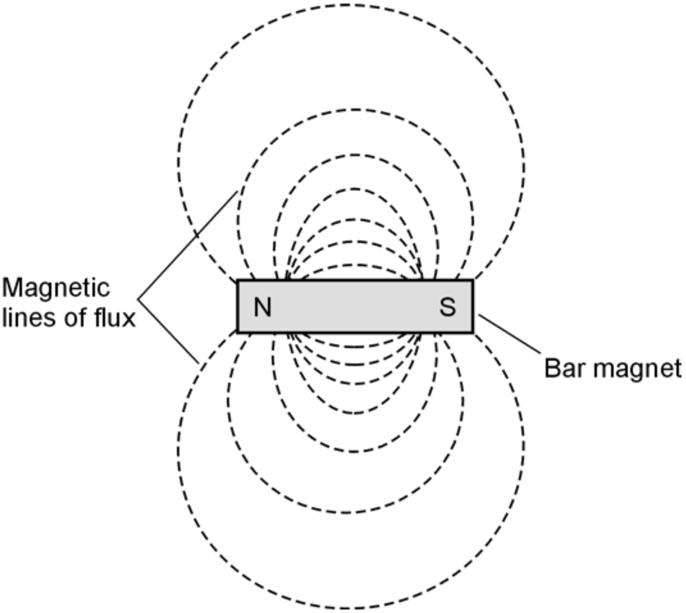

8-2 The pattern of magnetic flux lines (dashed curves) around a bar magnet (rectangle). The N and S represent north and south magnetic poles, respectively.

The intensity of a magnetic field is determined according to the number of flux lines passing through a certain cross section, such as a square centimeter or a square meter. The lines don't exist as real objects, but it is intuitively appealing to imagine them that way. The iron filings on the paper really do bunch themselves into lines (curves, actually) when there is a magnetic field of sufficient strength to make them move. Sometimes lines of flux are called *lines of force.* But technically, this is a misnomer.

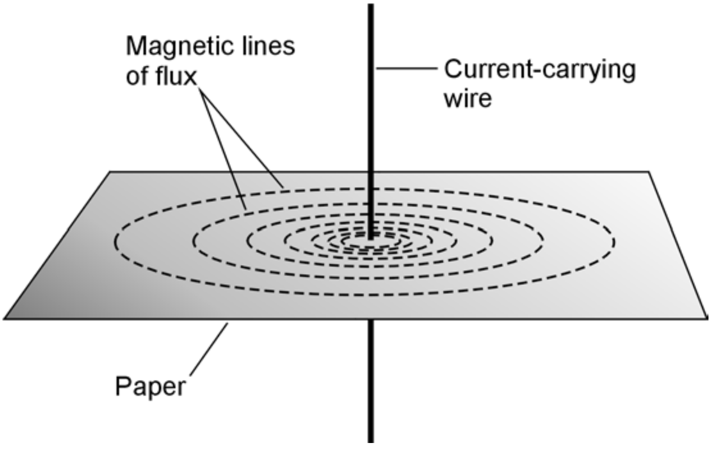

8-3 The pattern of magnetic flux lines (dashed curves) around a straight, current-carrying wire can be seen when the wire passes through a horizontal sheet of paper sprinkled with iron filings.

Poles

A magnetic field has a specific direction, as well as a specific intensity, at any given point in space near a current-carrying wire or a permanent magnet. The flux lines run parallel with the direction of the field. A magnetic field is considered to begin at the north magnetic pole, and to terminate at the south magnetic pole. In the case of a permanent magnet, it is obvious where the magnetic poles are. In the case of a current-carrying wire, the magnetic field goes in endless circles around the wire.

A charged electric particle, such as a proton or electron, hovering all by itself in space, constitutes an *electric monopole.* The *electric lines of flux* around an isolated, charged particle in free space are straight, and they "run off to infinity" (Fig. 8-4). A positive electric charge does not have to be mated with a negative electric charge.

A magnetic field is different. All magnetic flux lines, at least in ordinary real-world situations, are *closed loops.* With permanent magnets, there is a starting point (the north pole) and an ending point (the south pole). Around a straight, current-carrying wire, the loops are closed circles, even though the starting and ending points are not obvious. A pair of magnetic poles is called a *magnetic dipole.*

At first you might think that the magnetic field around a current-carrying wire is caused by a monopole, or that there aren't any poles at all, because the concentric circles don't actually converge anywhere. But you can envision a half plane, with the edge along the line of the wire, as a magnetic dipole. Then the lines of flux go around once in a 360° circle from the "north face" of the half plane to the "south face."

The greatest flux density, or field strength, around a bar magnet is near the poles, where the lines converge. Around a current-carrying wire, the greatest field strength is near the wire.

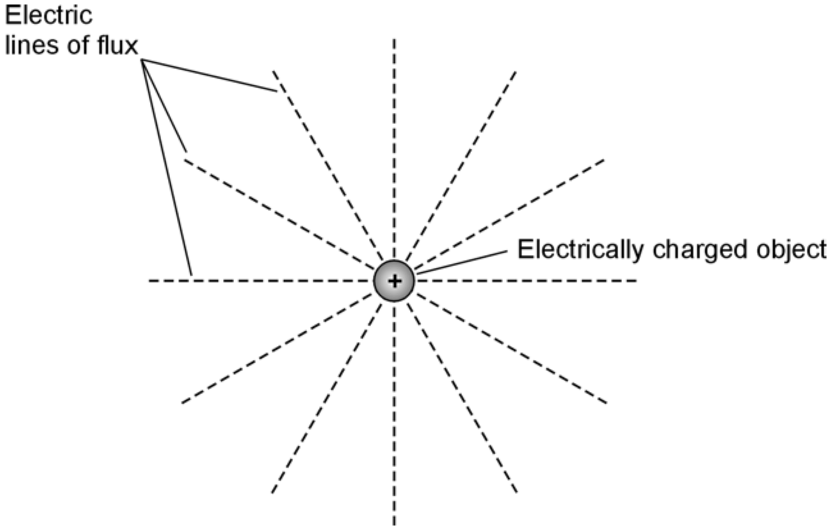

8-4 Electric flux lines (dashed lines) around an electrically charged object. This example shows a positive charge. The pattern of flux lines for a negative charge is identical.

Magnetic Field Strength

The overall magnitude of a magnetic field is measured in units called *webers* (Wb). A smaller unit, the *maxwell* (Mx), is sometimes used if a magnetic field is weak. One weber is equivalent to 100,000,000 (10^8) maxwells. Conversely, 1 Mx = 0.00000001 Wb = 10^{-8} Wb.

The Tesla and the Gauss

If you have access to a permanent magnet or electromagnet, you might see its strength expressed in terms of webers or maxwells. But usually you'll hear units called *teslas* (T) or *gauss* (G). These units are expressions of the concentration, or intensity, of the magnetic field within a certain cross section. The *flux density*, or number of lines per square meter or per square centimeter, is a more useful expression for magnetic effects than the overall quantity of magnetism. A flux density of 1 tesla (1 T) is equal to 1 weber per square meter (1 Wb/m^2). A flux density of 1 gauss (1 G) is equal to 1 maxwell per square centimeter (1 Mx/cm^2). It turns out that the gauss is equal to 0.0001 tesla (10^{-4} T). Conversely, the tesla is equivalent to 10,000 gauss (10^4 G).

The Ampere-Turn and the Gilbert

With electromagnets, another unit is employed: the *ampere-turn* (At). This is technically a unit of *magnetomotive force*, which is the magnetic counterpart of electromotive force. A wire, bent into a circle and carrying 1 A of current, produces 1 At of magnetomotive force. If the wire is bent into a loop having 50 turns, and the current stays the same, the resulting magnetomotive force is 50 At. If the current is then reduced to 1/50 A or 20 mA, the magnetomotive force will go back down to 1 At.

The *gilbert* (Gb) is also used to express magnetomotive force, but it is less common than the ampere-turn. One gilbert (1 Gb) is equal to 0.796 At. Conversely, 1 At = 1.26 Gb.

Electromagnets

Any electric current, or movement of charge carriers, produces a magnetic field. This field can become intense in a tightly coiled wire that has many turns and carries a large current. When a ferromagnetic core is placed inside the coil, the magnetic lines of flux are concentrated in the core, and the field strength in and near the core can become tremendous. This is the principle of an *electromagnet* (Fig. 8-5). Electromagnets are almost always cylindrical in shape. Sometimes the cylinder is long and thin; in other cases it is short and fat. But whatever the ratio of diameter to length for the core, the principle is the same: the magnetic field produced by the current results in magnetization of the core.

Direct-Current Types

You can build a dc electromagnet by taking a large bolt, such as a stove bolt, and wrapping a few dozen or a few hundred turns of wire around it. These items are available in any good hardware store. Be sure the bolt is made of ferromagnetic material. (If a permanent magnet sticks to the bolt, the bolt is ferromagnetic.) Ideally, the bolt should be at least 1 cm (approximately ⅜ in) in diameter and several inches long. You must use insulated wire, preferably made of solid, soft copper. "Bell wire" works well. Be sure all the wire turns go in the same direction. A large 6-V lantern battery can provide plenty of current to work the electromagnet. Never leave the coil connected to the battery for more than a few seconds at a time. And never use a car battery for this experiment! The acid can boil out of this type of battery, because the electromagnet places a heavy load on it.

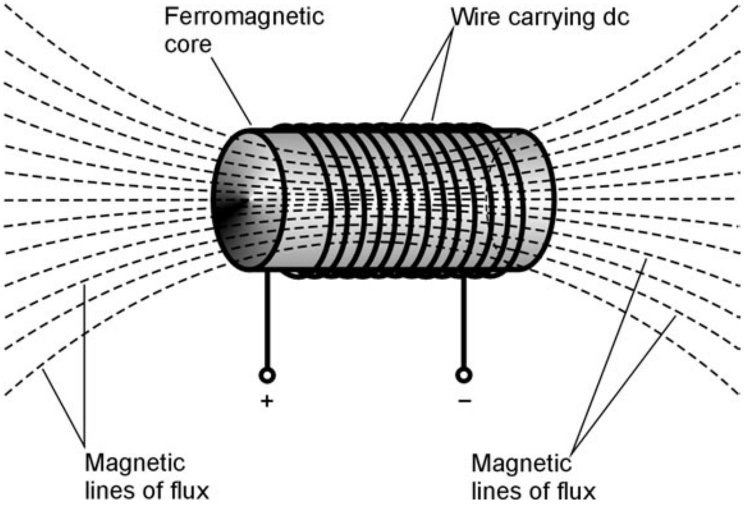

8-5 In an electromagnet, the magnetic flux is concentrated in a ferromagnetic rod surrounded by a current-carrying coil.

Direct-current electromagnets have defined north and south poles, just like permanent magnets. The main difference is that an electromagnet can get much stronger than any permanent magnet. You will see evidence of this if you do the preceding experiment with a large enough bolt and enough turns of wire.

Alternating-Current Types

Do you get the idea that an electromagnet can be made far stronger if, rather than using a lantern battery for the current source, you plug the wires into a wall outlet? In theory, this is true. In prac-

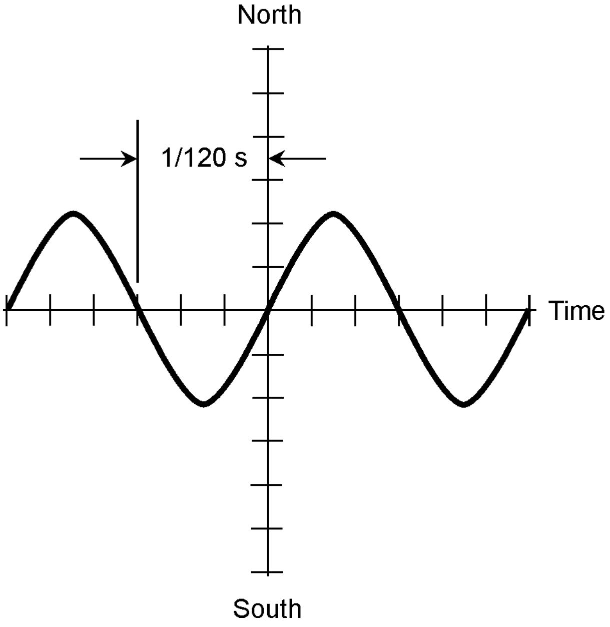

8-6 Polarity change in an ac electromagnet. The polarity changes every 1/120 second for 60-Hz utility current.

tice, you'll blow the fuse or circuit breaker. *Do not try this!* The electrical circuits in some buildings are not adequately protected and it can create a fire hazard. Also, you can get a lethal shock from the utility mains.

Some electromagnets use ac, and these magnets will stick to ferromagnetic objects. But the polarity of the magnetic field reverses every time the direction of the current reverses. With conventional household ac in the United States, there are 120 fluctuations, or 60 complete north-to-south-to-north polarity changes (Fig. 8-6), per second. If a permanent magnet, or a dc electromagnet, is brought near either "pole" of an ac electromagnet, there is no net force because the poles are alike half the time and opposite half the time, producing an equal amount of attractive and repulsive force. But if a piece of iron or steel is brought near a strong ac electromagnet, watch out! The attractive force will be powerful.

Magnetic Properties of Materials

There are four important properties that materials can have with respect to magnetic flux. These properties are *ferromagnetism*, *diamagnetism*, *permeability*, and *retentivity.*

Ferromagnetism

Some substances cause magnetic lines of flux to bunch closer together than they would in the medium of air or a vacuum. This property is called *ferromagnetism*, and materials that exhibit it are called *ferromagnetic.* You've already learned something about this!

Diamagnetism

Another property is known as *diamagnetism*, and materials that exhibit it are called *diamagnetic.* This type of substance decreases the magnetic flux density by causing the magnetic flux lines to diverge. Wax, dry wood, bismuth, and silver are examples. No diamagnetic material reduces the strength of a magnetic field by anywhere near the factor that ferromagnetic substances can increase it. Diamagnetic materials are generally used to keep magnetic objects apart, while minimizing the interaction between them. In recent years, they have also found some application in *magnetic levitation* devices.

Permeability

Permeability is a quantitative indicator of the extent to which a ferromagnetic material concentrates magnetic lines of flux. It is measured on a scale relative to a vacuum, or free space. Free space is assigned permeability 1. If you have a coil of wire with an air core, and a current is forced through the wire, then the flux in the coil core is at a certain density, just about the same as it would be in a vacuum. Therefore, the permeability of pure air is about equal to 1. If you place an iron core in the coil, the flux density increases by a large factor. The permeability of iron can range from 60 (impure) to as much as 8000 (highly refined).

If you use certain ferromagnetic alloys as the core material in electromagnets, you can increase the flux density, and therefore the local strength of the field, by as much as a million times. Such substances thus have permeability as great as 1,000,000 (10^6).

Table 8-1 gives permeability values for some common materials.

Retentivity

When a substance, such as iron, is subjected to a magnetic field as intense as it can handle, say by enclosing it in a wire coil carrying a massive current, there will be some *residual magnetism* left

Table 8-1. Permeability values for some common materials.

Substance	Permeability (approx.)
Air, dry, at sea level	1
Alloys, ferromagnetic	3000–1,000,000
Aluminum	Slightly more than 1
Bismuth	Slightly less than 1
Cobalt	60–70
Iron, powdered and pressed	100–3000
Iron, solid, refined	3000–8000
Iron, solid, unrefined	60–100
Nickel	50–60
Silver	Slightly less than 1
Steel	300–600
Vacuum	1
Wax	Slightly less than 1
Wood, dry	Slightly less than 1

when the current stops flowing in the coil. *Retentivity*, also sometimes called *remanence*, is a measure of how well the substance "memorizes" the magnetism and thereby becomes a permanent magnet.

Retentivity is expressed as a percentage, and is symbolized B_r. If the flux density in the material is x tesla or gauss when it is subjected to the greatest possible magnetomotive force, and then goes down to y tesla or gauss when the current is removed, the retentivity is equal to $100(y/x)\%$.

Suppose that a metal rod can be magnetized to 135 G when it is enclosed by a coil carrying an electric current. Imagine that this is the maximum possible flux density that the rod can be forced to have. (For any substance, there is always such a maximum.) Now suppose that the current is shut off, and 19 G remain in the rod. Then the retentivity, B_r, is calculated as follows:

$$B_r = 100(19/135)\% = (100 \times 0.14)\% = 14\%$$

Some ferromagnetic substances have high retentivity. These materials are excellent for making permanent magnets. Other substances have low retentivity. They work well as electromagnets, but not as permanent magnets.

If a ferromagnetic substance has poor retentivity, it is especially well-suited for use as the core material for an ac electromagnet, because the polarity of the magnetic flux can reverse within the material at a rapid rate. Materials with high retentivity do not work well for ac electromagnets, because they resist the polarity reversal that takes place with ac.

Practical Magnetism

Magnetism has numerous applications in common consumer devices and systems. Here are some of the more common ways in which magnetic phenomena can be put to use.

Permanent Magnets

Permanent magnets are manufactured by using a high-retentivity ferromagnetic material as the core of an electromagnet for an extended period of time. The coil of the electromagnet carries a large direct current, causing intense magnetic flux of constant polarity within the material. (Don't try to do this at home. The high current can heat the coil and overload a battery or power supply, which produces a fire hazard and/or the risk of battery explosion.)

If you want to magnetize a screwdriver a little bit so that it will hold onto screws, just stroke the shaft of the screwdriver with the end of a bar magnet several dozen times. Once you have magnetized a tool in this way, however, it is nearly impossible to demagnetize it.

A Ringer Device

Figure 8-7 is a simplified diagram of a *bell ringer*, also called a *chime.* The main functional component is called a *solenoid*, and it is an electromagnet. The core has a hole going along its axis. The coil has several layers, but the wire is always wound in the same direction, so that the electromagnet is powerful. A movable steel rod runs through the hole in the electromagnet core.

When there is no current flowing in the coil, the steel rod is held down by the force of gravity. When a pulse of current passes through the coil, the rod is pulled forcibly upward so that it strikes the ringer plate. This plate is like one of the plates in a xylophone. The current pulse is short, so the steel rod falls back down again to its resting position, allowing the plate to reverberate.

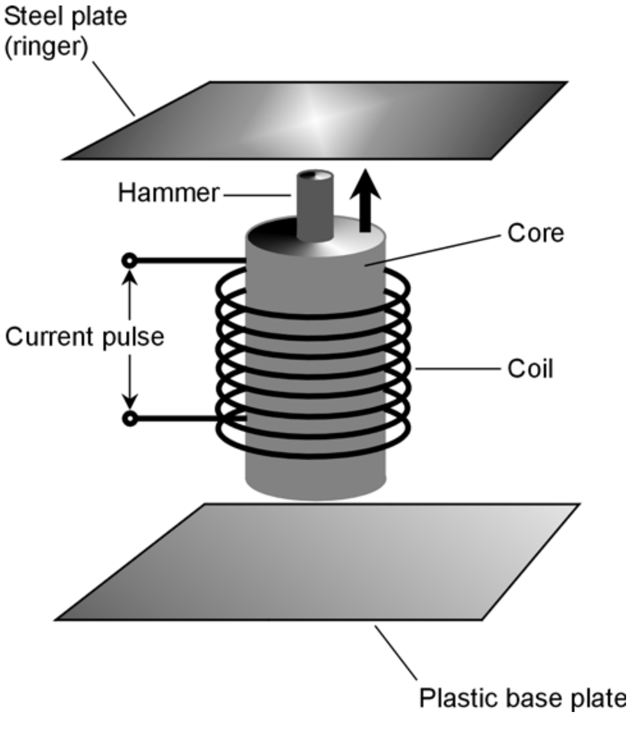

8-7 A solenoid-coil bell ringer.

The Relay

A *relay* makes use of a solenoid to allow remote-control switching of high-current circuits. A diagram of a relay is shown in Fig. 8-8. The movable lever, called the *armature*, is held to one side by a spring when there is no current flowing through the electromagnet. Under these conditions, terminal X is connected to Y, but not to Z. When a sufficient current is applied, the armature is pulled over to the other side. This disconnects terminal X from terminal Y, and connects X to Z.

There are numerous types of relays. Some are meant for use with dc, and others are for ac; a few will work with either dc or ac. A *normally closed relay* completes the circuit when there is no current flowing in its electromagnet coil, and breaks the circuit when current flows through the coil. A *normally open relay* is just the opposite, completing the circuit when current flows through the electromagnet coil, and opening the circuit when current ceases to flow through the coil. *Normal*, in this context, refers to the condition of no current applied to the electromagnet.

The relay shown in Fig. 8-8 can be used as either a normally open or normally closed relay, depending on which contacts are selected. It can also be used to switch a line between two different circuits.

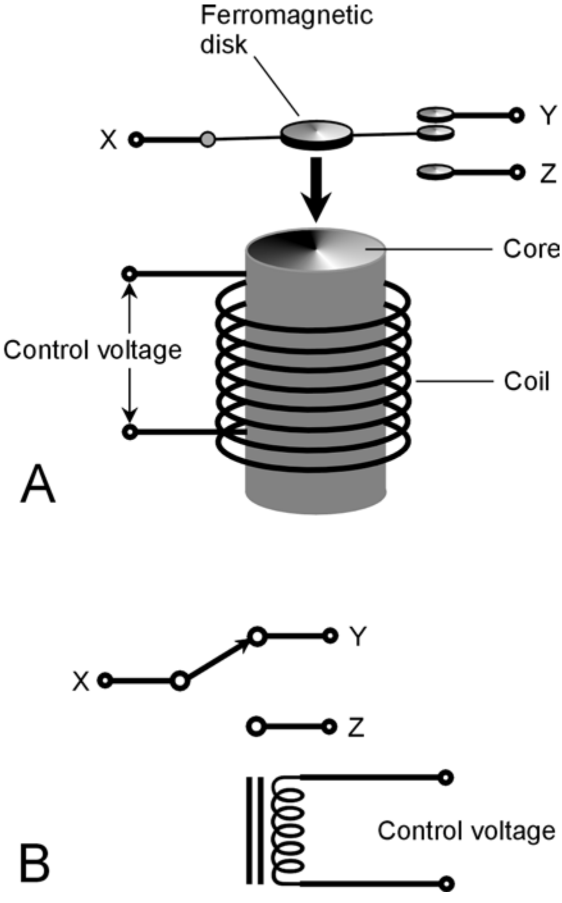

8-8 At A, pictorial diagram of a simple relay. At B, the schematic symbol for the same relay.

Some relays have several sets of contacts. Some relays are meant to remain in one state (either with current or without) for a long time, while others are meant to switch several times per second. The fastest relays can operate several dozen times per second. In recent years, relays have been largely supplanted by switching transistors and diodes, except in applications where extremely high current or high voltage is involved.

The DC Motor

Magnetic forces can be harnessed to do work. One common device that converts direct-current energy into rotating mechanical energy is a *dc motor.* In a dc motor, the source of electricity is connected to a set of coils, producing magnetic fields. The attraction of opposite poles, and the repulsion of like poles, is switched in such a way that a constant *torque*, or rotational force, results. As the current in the coils increases, the torque that the motor can provide also increases.

Figure 8-9 is a simplified, cutaway drawing of a dc motor. One set of coils, called the *armature coil*, rotates along with the motor shaft. The other set of coils, called the *field coil*, is stationary. The current direction is periodically reversed during each rotation by means of the *commutator.* This keeps the rotational force going in the same angular direction, so the motor continues to rotate rather than oscillating back and forth. The shaft is carried along by its own inertia, so that it doesn't come to a stop during those instants when the current is being switched in polarity.

Some dc motors can also be used to generate dc. These motors contain permanent magnets in place of one of the sets of coils. When the shaft is rotated, a pulsating dc flows in the coil.

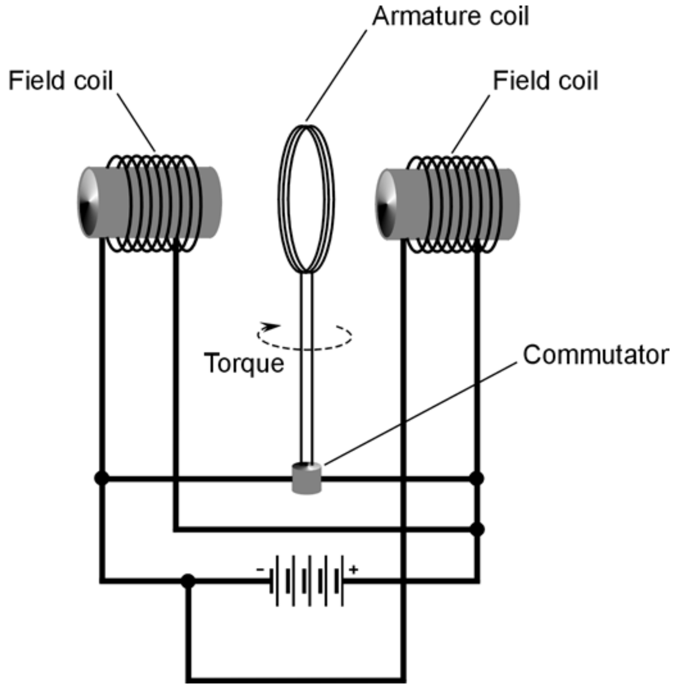

8-9 A functional diagram of a dc motor.

Magnetic Tape

Magnetic tape, also called *recording tape*, consists of millions of ferromagnetic particles attached to a flexible, thin plastic strip. In the *tape recorder*, a fluctuating magnetic field, produced by the *recording head*, polarizes these particles. As the field changes in strength next to the recording head, the tape passes by at a constant speed. This produces regions in which the ferromagnetic particles are polarized in either direction (Fig. 8-10).

When the tape is run at the same speed through the recorder in the playback mode, the magnetic fields around the individual particles cause a fluctuating field that is detected by the *pickup head.* This field has the same pattern of variations as the original field from the recording head.

Magnetic tape is available in various widths and thicknesses. Thicker tapes result in cassettes that don't play as long, but the tape is more resistant to stretching. The speed of the tape determines the fidelity of the recording. Higher speeds are preferred for music and video, and lower speeds for voice and data.

The impulses on a magnetic tape can be distorted or erased by external magnetic fields. Therefore, tapes should be protected from such fields. Keep the tape away from magnets. Extreme heat can also result in loss of data, and can cause permanent physical damage to the tape.

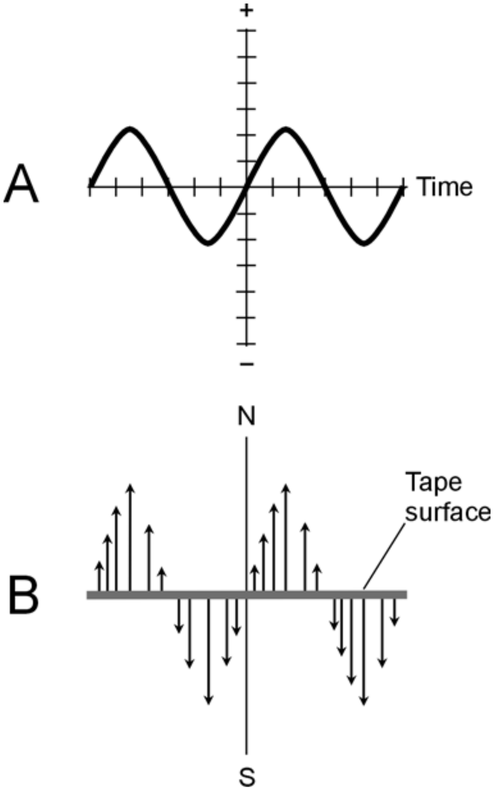

8-10 On recording tape, particles are magnetized in a pattern that follows the input waveform. Graph A shows an example of an audio input waveform. Graph B shows relative polarity and intensity of magnetization for selected particles on the tape surface.

Magnetic Disk

Since the advent of the personal computer, ever-more compact data-storage systems have evolved. One of the most versatile is the *magnetic disk.*

Hard disks, also called *hard drives*, store the most data, and are generally found inside of computer units. *Diskettes* are 8.9 cm (3.5 in) across, and can be inserted and removed from recording/playback machines called *diskette drives.* In recent years, magnetic diskettes have been largely supplanted by nonmagnetic *compact disc recordable* (CD-R) and *compact disc rewritable* (CD-RW) media.

The principle of the magnetic disk, on the microscale, is the same as that of magnetic tape. The information is stored in *binary digital* form; that is, there are only two different ways that the particles are magnetized. This results in almost perfect, error-free storage. On a larger scale, the disk works differently than the tape because of the difference in geometry. On a tape, the information is spread out over a long span, and some bits of data are far away from others as measured along the medium itself. But on a disk, no two bits are ever farther apart than the diameter of the disk. This means that data can be stored to, and retrieved from, a disk much faster than is possible with tape.

The same precautions should be observed when handling and storing magnetic disks as are necessary with magnetic tape.

Bubble Memory

Bubble memory is a sophisticated method of storing data that gets rid of the need for moving parts such as are required in tape machines and disk drives. Data is stored as tiny magnetic fields, in a medium that is made from magnetic film and semiconductor materials.

Bubble memory makes use of all the advantages of magnetic data storage, as well as the favorable aspects of electronic data storage. Advantages of electronic memory include rapid storage and recovery, and high density (a lot of data can be put in a tiny volume of space). Advantages of magnetic memory include *nonvolatility* (it can be stored for a long time without needing a constant current source), high density, and comparatively low cost.

Bubble memory seems to go through phases. Just as it is declared obsolete, someone comes up with a new and improved way to make it work. Check the Internet to find out its current status; enter "bubble memory" or "magnetic bubble memory" into a search engine.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. The geomagnetic field
   (a) makes the earth like a huge horseshoe magnet.
   (b) runs exactly through the geographic poles.
   (c) makes a compass work.
   (d) makes an electromagnet work.
2. Geomagnetic lines of flux
   (a) are horizontal at the geomagnetic equator.
   (b) are vertical at the geomagnetic equator.

   (c) are never horizontal, no matter where you go.
   (d) are perfectly symmetrical around the earth, even far out in space.

3. A material that can be permanently magnetized is generally said to be
   (a) ultramagnetic.
   (b) electromagnetic.
   (c) diamagnetic.
   (d) ferromagnetic.
4. The force between a magnet and a piece of ferromagnetic metal that has not been magnetized
   (a) can be either repulsive or attractive.
   (b) is never repulsive.
   (c) gets smaller as the magnet gets closer to the metal.
   (d) depends on the geomagnetic field.
5. The presence of a magnetic field can always be attributed to
   (a) ferromagnetic materials.
   (b) diamagnetic materials.
   (c) motion of electric charge carriers.
   (d) the north geomagnetic pole.
6. Lines of magnetic flux are said to originate
   (a) in atoms of ferromagnetic materials.
   (b) at a north magnetic pole.
   (c) at points where the lines are straight.
   (d) in electric charge carriers.
7. The magnetic flux around a straight, current-carrying wire
   (a) gets stronger with increasing distance from the wire.
   (b) is strongest near the wire.
   (c) does not vary in strength with distance from the wire.
   (d) consists of straight lines parallel to the wire.
8. The gauss is a unit of
   (a) overall magnetic field strength.
   (b) ampere-turns.
   (c) magnetic flux density.
   (d) magnetic power.
9. A unit of overall magnetic field quantity is the
   (a) maxwell.
   (b) gauss.
   (c) tesla.
   (d) ampere-turn.

10. If a wire coil has 10 turns and carries 500 mA of current, what is the magnetomotive force?
    (a) 5000 At
    (b) 50 At
    (c) 5.0 At
    (d) 0.02 At

11. If a wire coil has 100 turns and carries 1.30 A of current, what is the magnetomotive force?
    (a) 130 Gb
    (b) 76.9 Gb
    (c) 164 Gb
    (d) 61.0 Gb

12. Which of the following can occur during a geomagnetic storm?
    (a) Charged particles stream out from the sun.
    (b) The earth's magnetic field is affected.
    (c) Electrical power transmission is disrupted.
    (d) More than one of the above can occur.

13. An ac electromagnet
    (a) attracts only permanent magnets.
    (b) attracts pure, unmagnetized iron.
    (c) repels all permanent magnets.
    (d) either attracts or repels permanent magnets, depending on the polarity.

14. An advantage of an electromagnet over a permanent magnet is the fact that
    (a) an electromagnet can be switched on and off.
    (b) an electromagnet does not have specific polarity.
    (c) an electromagnet requires no power source.
    (d) permanent magnets must always be cylindrical, but electromagnets can have any shape.

15. A substance with high retentivity
    (a) can make a good ac electromagnet.
    (b) repels both north and south magnetic poles.
    (c) is always a diamagnetic material.
    (d) is well suited to making a permanent magnet.

16. Suppose a relay is connected into a circuit so that a device gets a signal only when the relay coil carries current. The relay is
    (a) an ac relay.
    (b) a dc relay.
    (c) normally closed.
    (d) normally open.

17. A device that repeatedly reverses the polarity of a magnetic field in order to keep a dc motor rotating is known as
    - (a) a solenoid.
    - (b) an armature coil.
    - (c) a commutator.
    - (d) a field coil.

18. A high tape-recorder motor speed is generally used for
    - (a) voice recording and playback.
    - (b) video recording and playback.
    - (c) digital data storage and retrieval.
    - (d) all of the above.

19. An advantage of a magnetic disk, compared with magnetic tape, for data storage and retrieval is that
    - (a) a disk lasts longer.
    - (b) data can be stored and retrieved more quickly with disks than with tapes.
    - (c) disks look better.
    - (d) disks are less susceptible to magnetic fields.

20. A magnetic hard disk is usually part of
    - (a) a computer.
    - (b) a dc motor.
    - (c) a tape recorder.
    - (d) an electromagnet.

Test: Part 1

Do not refer to the text when taking this test. A good score is at least 37 correct. Answers are in the back of the book. It's best to have a friend check your score the first time, so you won't memorize the answers if you want to take the test again.

1. An application in which an analog meter would almost always be preferred over a digital meter is
   - (a) the signal-strength indicator in a radio receiver.
   - (b) a meter that shows power-supply voltage.
   - (c) a utility watt-hour meter.
   - (d) a clock.
   - (e) a device in which a direct numeric display is wanted.

2. Which of the following statements is false?
   - (a) The current in a series dc circuit is divided up among the resistances.
   - (b) In a parallel dc circuit, the voltage is the same across each component.
   - (c) In a series dc circuit, the sum of the voltages across all the components, going once around a complete circle and taking polarity into account, is zero.
   - (d) The net resistance of a parallel set of resistors is less than the value of the smallest resistor.
   - (e) The total wattage consumed in a series circuit is the sum of the wattages consumed by each of the components.

3. The ohm is a unit of
   - (a) electrical charge quantity.
   - (b) the rate at which charge carriers flow.
   - (c) opposition to electrical current.
   - (d) electrical conductance.
   - (e) potential difference.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

4. A wiring diagram differs from a schematic diagram in that
   (a) a wiring diagram is less detailed than a schematic diagram.
   (b) a wiring diagram always shows the component values, but a schematic diagram might not.
   (c) a schematic does not show all the interconnections between the components, but a wiring diagram does.
   (d) a schematic diagram shows pictures of components, while a wiring diagram shows the electronic symbols.
   (e) a schematic diagram shows the electronic symbols, while a wiring diagram shows pictures of the components.

5. In which of the following places would you be most likely to find a wirewound resistor?
   (a) A dc circuit location where a large amount of power must be dissipated
   (b) The input circuit of a radio-frequency amplifier
   (c) The output circuit of a radio-frequency amplifier
   (d) In an antenna system, to limit the transmitter power
   (e) Between ground and the chassis of a power supply

6. The number of protons in the nucleus of an element is known as the
   (a) electron number.
   (b) atomic number.
   (c) valence number.
   (d) charge number.
   (e) proton number.

7. A hot-wire ammeter
   (a) can measure ac as well as dc.
   (b) registers current changes very fast.
   (c) can indicate very low voltages.
   (d) measures electrical energy.
   (e) works only when current flows in one direction.

8. Which of the following units indicates the rate at which energy is expended?
   (a) The volt
   (b) The ampere
   (c) The coulomb
   (d) The ampere-hour
   (e) The watt

9. Which of the following correctly states Ohm's Law?
   (a) Volts equal amperes divided by ohms.
   (b) Ohms equal amperes divided by volts.
   (c) Amperes equal ohms divided by volts.

(d) Amperes equal ohms times volts.
(e) Ohms equal volts divided by amperes.

10. The current flowing into a point in a dc circuit is always equal to the current
    (a) delivered by the power supply.
    (b) through any one of the resistances.
    (c) flowing out of that point.
    (d) at any other point.
    (e) in any single branch of the circuit.

11. A loudness meter in a hi-fi system is generally calibrated in
    (a) volts.
    (b) amperes.
    (c) decibels.
    (d) watt-hours.
    (e) ohms.

12. An electrically charged atom (either positive or negative) is known as
    (a) a molecule.
    (b) an isotope.
    (c) an ion.
    (d) an electron.
    (e) a fundamental particle.

13. Suppose a battery delivers 12.0 V to a bulb, and current flowing through the bulb is 3.00 A. The resistance of the bulb is which of the following?
    (a) 36.0 Ω
    (b) 4.00 Ω
    (c) 0.250 Ω
    (d) 108 Ω
    (e) 0.750 Ω

14. The peak voltage in an ac wave is always
    (a) greater than the average voltage.
    (b) less than the average voltage.
    (c) greater than or equal to the average voltage.
    (d) less than or equal to the average voltage.
    (e) fluctuating.

15. Suppose a resistor is specified a having a value of 680 Ω, and a tolerance of ±5%. You measure the actual resistance with a precision digital ohmmeter. Which of the following meter readings indicates a reject?
    (a) 648 Ω
    (b) 712 Ω

(c) 699 Ω
(d) 636 Ω
(e) 707 Ω

16. A primitive device for indicating the presence of an electric current is
    (a) an electrometer.
    (b) a galvanometer.
    (c) a voltmeter.
    (d) a coulometer.
    (e) a wattmeter.
17. A disadvantage of mercury cells is the fact that they
    (a) can adversely affect the environment when discarded.
    (b) supply dangerously high voltage.
    (c) can reverse polarity unexpectedly.
    (d) must be physically larger than other types of cells that have the same current-delivering capacity.
    (e) must be kept right-side up to keep the mercury from spilling out.
18. Suppose a battery supplies 6.0 V to a bulb rated at 12 W. The bulb draws how much current?
    (a) 2.0 A
    (b) 0.5 A
    (c) 72 A
    (d) 40 mA
    (e) 72 mA
19. Which of the following is not a common use for a resistor or set of resistors?
    (a) Biasing for a transistor
    (b) Voltage division
    (c) Current limiting
    (d) As a dummy antenna
    (e) Helping a capacitor to hold its charge for a long time
20. When an electrical charge exists but there is no flow of current, the charge is said to be
    (a) ionizing.
    (b) atomic.
    (c) molecular.
    (d) electronic.
    (e) static.
21. The sum of the voltages, going around a dc circuit, but not including the power supply, has
    (a) an equal value and the same polarity as the supply.
    (b) a value that depends on the ratio of the resistances.
    (c) a different value from, but the same polarity as, the supply.

(d) an equal value as, but the opposite polarity from, the supply.
(e) a different value from, and the opposite polarity from, the supply.

22. A watt-hour meter measures
(a) voltage.
(b) current.
(c) power.
(d) energy.
(e) charge.

23. Every chemical element has its own unique type of particle, which is known as its
(a) neutron.
(b) electron.
(c) proton.
(d) atom.
(e) isotope.

24. An advantage of a magnetic disk over magnetic tape for data storage is the fact that
(a) data is too closely packed on the tape.
(b) the disk is immune to the effects of magnetic fields.
(c) data storage and retrieval is faster on disk.
(d) disks store computer data in analog form.
(e) tapes cannot be used to store digital data.

25. Suppose a 6-V battery is connected across a series combination of resistors. The resistance values are 1.0 Ω, 2.0 Ω, and 3.0 Ω. What is the current through the 2.0-Ω resistor?
(a) 1.0 A
(b) 3.0 A
(c) 12 A
(d) 24 A
(e) 72 A

26. A sample of material with resistance so high that it can be considered infinite for most practical purposes is known as
(a) a semiconductor.
(b) a paraconductor.
(c) an insulator.
(d) a resistor.
(e) a diamagnetic substance.

27. Primary cells
(a) can be used over and over.
(b) have higher voltage than other types of cells.
(c) all supply exactly 1.500 V.

(d) cannot be recharged.
(e) are made of zinc and carbon.

28. A rheostat
    (a) can be used in high-voltage and/or high-power dc circuits.
    (b) is ideal for tuning a radio receiver.
    (c) is often used as a bleeder resistor.
    (d) is better than a potentiometer for low-power audio.
    (e) offers the advantage of having no inductance.

29. How much dc voltage does a typical dry cell provide?
    (a) 12 V
    (b) 6 V
    (c) 1.5 V
    (d) 117 V
    (e) Any of the above

30. A geomagnetic storm
    (a) causes solar wind.
    (b) causes the earth's magnetic field to disappear.
    (c) can disturb the earth's magnetic field.
    (d) can pollute the earth's atmosphere.
    (e) stabilizes the ac utility grid.

31. An advantage of an alkaline cell over a zinc-carbon cell is the fact that
    (a) the alkaline cell provides more voltage.
    (b) the alkaline cell can be recharged.
    (c) the alkaline cell can deliver useful current at lower temperatures.
    (d) the alkaline cell is far less bulky for the same amount of energy capacity.
    (e) the alkaline cell can produce ac as well as dc.

32. Suppose a battery delivers 12 V across a set of six 4.0-Ω resistors in a series voltage dividing combination. This provides six different voltages, differing by equal increments of which of the following?
    (a) 0.25 V
    (b) 0.33 V
    (c) 1.0 V
    (d) 2.0 V
    (e) 3.0 V

33. A unit of electrical charge quantity is the
    (a) volt.
    (b) ampere.
    (c) watt.

(d) tesla.
(e) coulomb.

34. A unit of conductance is the
(a) volt per meter.
(b) ampere per meter.
(c) anti-ohm.
(d) siemens.
(e) ohm per meter.

35. Suppose a 24-V battery is connected across a set of four resistors in parallel. Each resistor has a value of 32 Ω. What is the total power dissipated by the set of resistors?
(a) 0.19 W
(b) 3.0 W
(c) 0.19 kW
(d) 0.33 W
(e) 72 W

36. The main difference between a lantern battery and a transistor battery is the fact that
(a) a lantern battery has higher voltage than a transistor battery.
(b) a fresh lantern battery has more energy stored in it than a fresh transistor battery.
(c) a lantern battery cannot be used with electronic devices such as transistor radios, but a transistor battery can.
(d) a lantern battery can be recharged, but a transistor battery cannot.
(e) a lantern battery is more compact than a transistor battery.

37. Nickel-based batteries would most likely be found
(a) in disposable flashlights.
(b) in large lanterns.
(c) as car and truck batteries.
(d) in handheld radio transceivers.
(e) in electromagnets.

38. A voltmeter should have
(a) low internal resistance.
(b) electrostatic plates.
(c) a sensitive amplifier.
(d) high internal resistance.
(e) the highest possible full-scale value.

39. The purpose of a bleeder resistor is to
(a) provide bias for a transistor.
(b) serve as a voltage divider.

(c) protect people against the danger of electric shock.
(d) reduce the current in a power supply.
(e) smooth out the ac ripple in a power supply.

40. A dc electromagnet
(a) has constant polarity.
(b) requires an air core.
(c) does not attract or repel a permanent magnet.
(d) has polarity that periodically reverses.
(e) cannot be used to permanently magnetize anything.

41. The rate at which charge carriers flow is measured in
(a) amperes.
(b) coulombs.
(c) volts.
(d) watts.
(e) watt-hours.

42. Suppose a 12-V battery is connected to a set of three resistors in series. The resistance values are 1.0 Ω, 2.0 Ω, and 3.0 Ω. What is the voltage across the 3.0-Ω resistor?
(a) 1.0 V
(b) 2.0 V
(c) 4.0 V
(d) 6.0 V
(e) 12 V

43. Suppose nine 90-Ω resistors are connected in a 3 × 3 series-parallel network. What is the total (net) resistance of the network?
(a) 10 Ω
(b) 30 Ω
(c) 90 Ω
(d) 270 Ω
(e) 810 Ω

44. A device commonly used for remote switching of high-current circuits is
(a) a solenoid.
(b) an electromagnet.
(c) a potentiometer.
(d) a photovoltaic cell.
(e) a relay.

45. Memory in a nickel-based cell or battery
(a) occurs whenever the battery is discharged.
(b) indicates that the cell or battery is dead.

(c) can usually be remedied by repeated discharging and recharging.
(d) can cause an explosion.
(e) causes a reversal in polarity.

46. Suppose a 100-W bulb burns for 100 hours. It has consumed how many units of energy?
(a) 0.10 kWh
(b) 1.00 kWh
(c) 10.0 kWh
(d) 100 kWh
(e) 1000 kWh

47. A material with high permeability
(a) increases magnetic field quantity.
(b) is necessary if a coil is to produce a magnetic field.
(c) always has high retentivity.
(d) concentrates magnetic lines of flux.
(e) reduces flux density.

48. A chemical compound
(a) consists of two or more atoms.
(b) contains an unusual number of neutrons.
(c) is technically the same as an ion.
(d) has a shortage of electrons.
(e) has an excess of electrons.

49. Suppose a 6.00-V battery is connected to a parallel combination of two resistors whose values are 8.00 Ω and 12.0 Ω. What is the power dissipated in the 8-Ω resistor?
(a) 0.300 W
(b) 0.750 W
(c) 1.25 W
(d) 1.80 W
(e) 4.50 W

50. The main problem with bar-graph meters is the fact that
(a) they are not very sensitive.
(b) they are unstable.
(c) they cannot give very precise readings.
(d) you need special training to read them.
(e) they can display only peak values.

2
PART

Alternating Current

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

This page intentionally left blank

9

CHAPTER

Alternating-Current Basics

DIRECT CURRENT CAN BE EXPRESSED IN TERMS OF TWO VARIABLES: DIRECTION (POLARITY) AND intensity (amplitude). Alternating current (ac) is a little more complicated. This chapter will acquaint you with some common forms of ac.

Definition of Alternating Current

You have learned that dc has polarity that stays constant over time. Although the amplitude (the number of amperes, volts, or watts) can fluctuate from moment to moment, the charge carriers always flow in the same direction at any point in the circuit.

In ac, the polarity reverses at regular intervals. The *instantaneous amplitude* (that is, the amplitude at any given instant in time) of ac usually varies because of the repeated reversal of polarity. But there are certain cases where the amplitude remains constant, even though the polarity keeps reversing.

The rate of change of polarity is the variable that makes ac so much different from dc. The behavior of an ac wave depends largely on this rate: the *frequency.*

Period and Frequency

In a *periodic ac wave*, the kind that is discussed in this chapter (and throughout the rest of this book), the function of *instantaneous amplitude versus time* repeats itself over and over, so that the same pattern recurs indefinitely. The length of time between one repetition of the pattern, or one *cycle*, and the next is called the *period* of the wave. This is illustrated in Fig. 9-1 for a simple ac wave. The period of a wave can, in theory, be anywhere from a minuscule fraction of a second to many centuries. Period, when measured in seconds, is denoted by T.

Originally, ac frequency was specified in *cycles per second* (cps). High frequencies were sometimes given in *kilocycles*, *megacycles*, or *gigacycles*, representing thousands, millions, or billions (thousand-millions) of cycles per second. But nowadays, the unit is known as the *hertz* (Hz). Thus, 1 Hz = 1 cps, 10 Hz = 10 cps, and so on. Higher frequencies are given in *kilohertz* (kHz), *megahertz* (MHz), or *gigahertz* (GHz). The relationships are as follows:

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

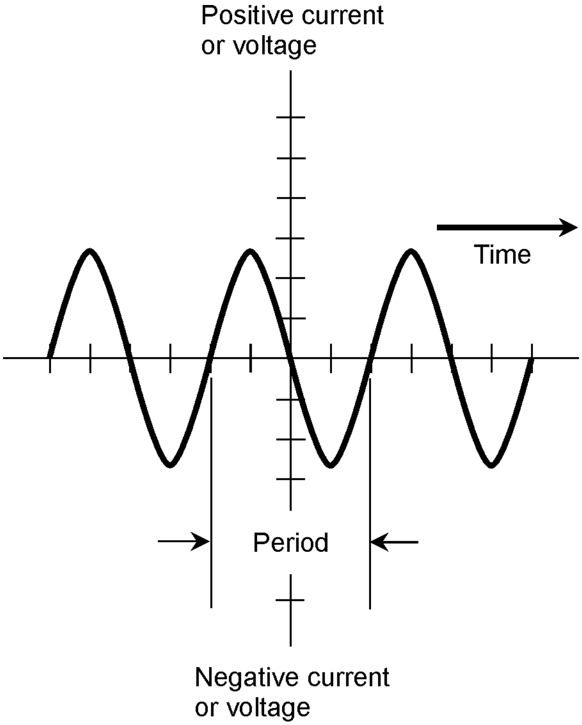

9-1 A sine wave. The period is the length of time it takes for one cycle to be completed.

$$1 \text{ kHz} = 1000 \text{ Hz}$$
$$1 \text{ MHz} = 1000 \text{ kHz} = 1{,}000{,}000 \text{ Hz} = 10^6 \text{ Hz}$$
$$1 \text{ GHz} = 1000 \text{ MHz} = 1{,}000{,}000{,}000 \text{ Hz} = 10^9 \text{ Hz}$$

Sometimes an even bigger unit, the *terahertz* (THz), is used to specify ac frequency. This is a trillion (1,000,000,000,000, or 10^{12}) hertz. Electrical currents generally do not attain such frequencies, although some forms of *electromagnetic radiation* do.

The frequency of an ac wave, denoted f, in hertz is the reciprocal of the period in seconds. Mathematically, these two equations express the relationship:

$$f = 1/T \qquad \text{and} \qquad T = 1/f$$

Some ac waves have only one frequency. These waves are called *pure.* But often, there are components at multiples of the main, or *fundamental*, frequency. There can also be components at odd frequencies. Some ac waves have hundreds, thousands, or even infinitely many different component frequencies.

The Sine Wave

Sometimes, alternating current has a *sine-wave*, or *sinusoidal*, nature. This means that the direction of the current reverses at regular intervals, and that the current-versus-time curve is shaped like the trigonometric *sine function.* The waveform in Fig. 9-1 is a sine wave.

Any ac wave that consists of a single frequency has a perfectly sinusoidal shape. Any perfect sinusoidal ac source has only one component frequency. In practice, a wave might be so close to a sine wave that it looks exactly like the sine function on an oscilloscope, when in reality there are traces of other frequencies present. Imperfections are often too small to see. But pure, single-frequency ac not only looks perfect, but actually is a perfect replication of the trigonometric sine function.

The current at the wall outlets in your house is an almost perfect ac sine wave with a frequency of 60 Hz.

Square Waves

Earlier in this chapter, it was said that there can be an ac wave whose instantaneous amplitude remains constant, even though the polarity reverses. Does this seem counterintuitive? Think some more! A *square wave* is such a wave.

On an oscilloscope, a square wave looks like a pair of parallel, dashed lines, one with positive polarity and the other with negative polarity (Fig. 9-2A). The oscilloscope shows a graph of voltage on the vertical scale and time on the horizontal scale. The transitions between negative and positive for a theoretically perfect square wave would not show up on the oscilloscope, because they would be instantaneous. But in practice, the transitions can often be seen as vertical lines (Fig. 9-2B).

True square waves have equal negative and positive peaks. Thus, the absolute amplitude of the wave is constant. Half of the time it's +*x*, and the other half of the time it's –*x* (where *x* can be expressed in volts, amperes, or watts).

Some squared-off waves are lopsided; the negative and positive amplitudes are not the same. Still others remain at positive polarity longer than they remain at negative polarity (or vice versa). These are examples of *asymmetrical square waves*, more properly called *rectangular waves.*

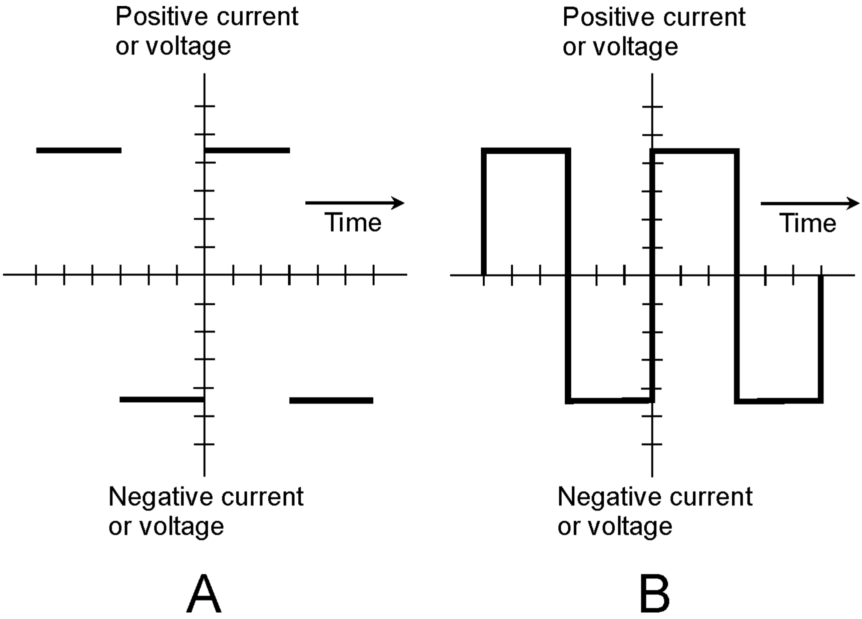

9-2 At A, a perfect square wave; the transitions are instantaneous and therefore do not show up on the graph. At B, the more common rendition of a square wave, showing the transitions as vertical lines.

Sawtooth Waves

Some ac waves rise and/or fall in straight, sloping lines as seen on an oscilloscope screen. The slope of the line indicates how fast the magnitude is changing. Such waves are called *sawtooth waves* because of their appearance. Sawtooth waves are generated by certain electronic test devices. They can also be generated by electronic sound synthesizers.

Fast Rise, Slow Decay

Figure 9-3 shows a sawtooth wave in which the positive-going slope (called the *rise*) is extremely steep, as with a square wave, but the negative-going slope (called the *decay*) is not so steep. The period of the wave is the time between points at identical positions on two successive pulses.

Slow Rise, Fast Decay

Another form of sawtooth wave is just the opposite, with a defined, finite rise and an instantaneous decay. This type of wave is often called a *ramp* because it looks like an incline going upward (Fig. 9-4). This waveshape is useful for scanning in television sets and oscilloscopes. It tells the electron beam to move, or *trace*, at constant speed from left to right across the screen during the rise. Then it retraces, or brings the electron beam back, instantaneously during the decay so the beam can trace across the screen again.

Variable Rise and Decay

Sawtooth waves can have rise and decay slopes in an infinite number of different combinations. One common example is shown in Fig. 9-5. In this case, the rise and the decay are both finite and equal. This is known as a *triangular wave.*

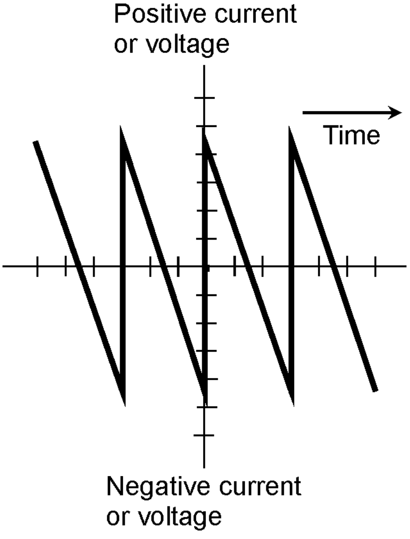

9-3 A sawtooth wave with a fast rise and a slow decay.

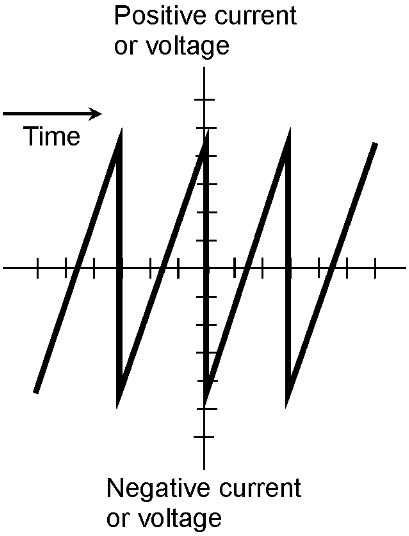
Positive current or voltage
Time
Negative current or voltage

9-4 A sawtooth wave with a slow rise and a fast decay.

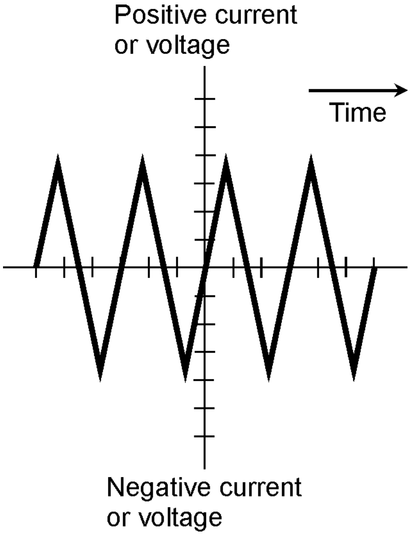
Positive current or voltage
Time
Negative current or voltage

9-5 A triangular wave with rise and decay rates that are the same.

Complex and Irregular Waveforms

As long as a wave has a definite period, and as long as the polarity keeps switching back and forth between positive and negative, it is ac, no matter how complicated the actual shape of the waveform. Figure 9-6 shows an example of a complex ac wave. There is a definable period, and therefore a definable frequency. The period is the time between two points on succeeding wave repetitions.

With some waves, it can be difficult or almost impossible to ascertain the period. This is because the wave has two or more components that are of nearly the same amplitude. When this hap-

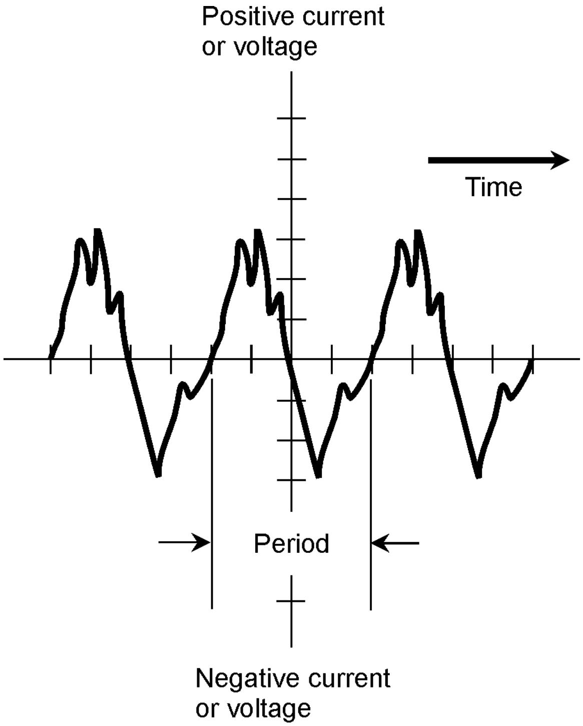

9-6 An irregular waveform.

pens, the *frequency spectrum* of the wave is multifaceted. That means the wave energy is split up more or less equally among multiple frequencies.

Frequency Spectrum

An oscilloscope shows a graph of amplitude as a function of time. Because time is on the horizontal axis and represents the *independent variable* or *domain* of the function, the oscilloscope is said to be a *time-domain* instrument. But suppose you want to see the amplitude of a complex signal as a function of frequency, rather than as a function of time? This can be done with a *spectrum analyzer.* It is a *frequency-domain* instrument. Its horizontal axis shows frequency as the independent variable, ranging from some adjustable minimum frequency (at the extreme left) to some adjustable maximum frequency (at the extreme right).

An ac sine wave, as displayed on a spectrum analyzer, appears as a single *pip*, or vertical line (Fig. 9-7A). This means that all of the energy in the wave is concentrated at one frequency. But many, if not most, ac waves contain *harmonic* energy along with energy at the fundamental frequency. A harmonic frequency is a whole-number multiple of the fundamental frequency. For example, if 60 Hz is the fundamental frequency, then harmonics can exist at 120 Hz, 180 Hz, 240 Hz, and so on. The 120-Hz wave is the *second harmonic;* the 180-Hz wave is the *third harmonic;* the 240-Hz wave is the *fourth harmonic;* and so on.

In general, if a wave has a frequency equal to *n* times the fundamental (where *n* is some whole number), then that wave is called the n*th harmonic.* In Fig. 9-7B, a wave is shown along with several harmonics, as it would look on the display screen of a spectrum analyzer.

Square waves and sawtooth waves contain harmonic energy in addition to energy at the fundamental frequency. Other waves can get more complicated. The exact shape of a wave depends on the amount of energy in the harmonics, and the way in which this energy is distributed among them.

Irregular waves can have any imaginable frequency distribution. Figure 9-8 shows an example. This is a spectral (frequency-domain) display of an *amplitude-modulated* (AM) voice radio signal. Much of the energy is concentrated at the center of the pattern, at the frequency shown by the vertical line. That is the *carrier frequency.* There is also plenty of energy near, but not exactly at, the carrier frequency. That's the part of the signal that contains the voice.

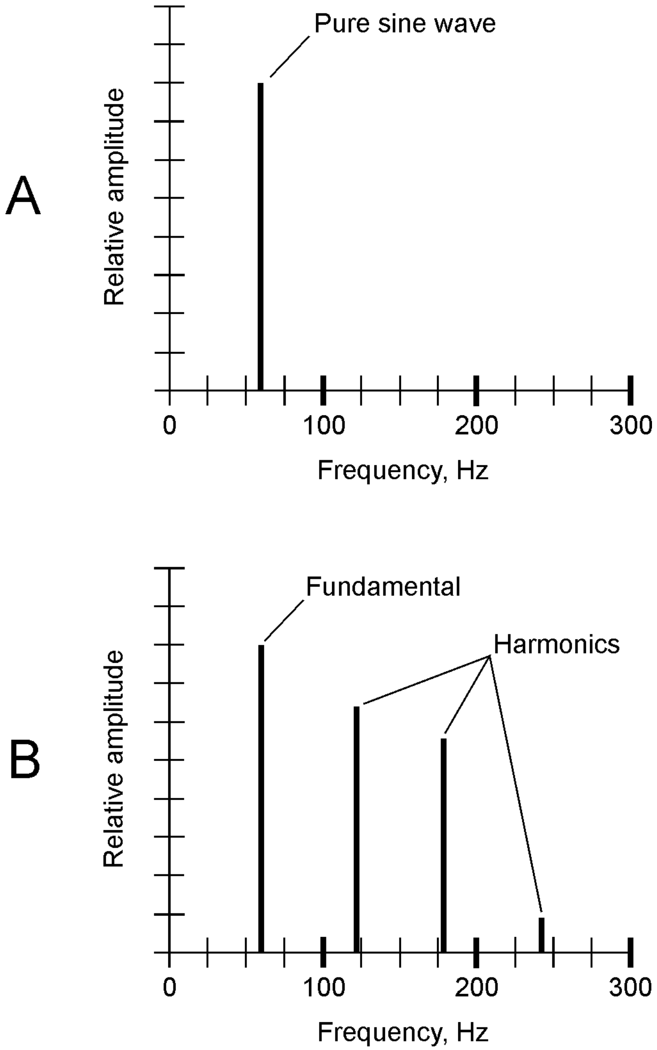

9-7 At A, a spectral diagram of a pure, 60-Hz sine wave. At B, a spectral diagram of a 60-Hz wave with three harmonics.

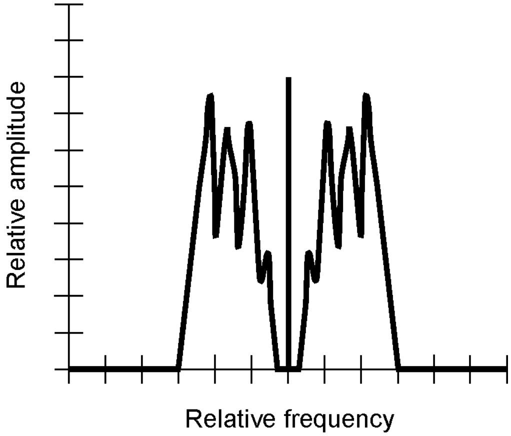

9-8 A spectral diagram of a modulated radio signal.

Fractions of a Cycle

Engineers break the ac cycle down into small parts for analysis and reference. One complete cycle can be compared to a single revolution around a circle.

Degrees

One method of specifying the phase of an ac cycle is to divide it into 360 equal parts, called *degrees* or *degrees of phase*, symbolized by a superscript, lowercase letter *o* (°). The value 0° is assigned to the point in the cycle where the magnitude is zero and positive-going. The same point on the next cycle is given the value 360°. The point one-fourth of the way through the cycle is 90°; the point halfway through the cycle is 180°; the point three-fourths of the way through the cycle is 270°. This is illustrated in Fig. 9-9. Degrees of phase are used mainly by engineers and technicians.

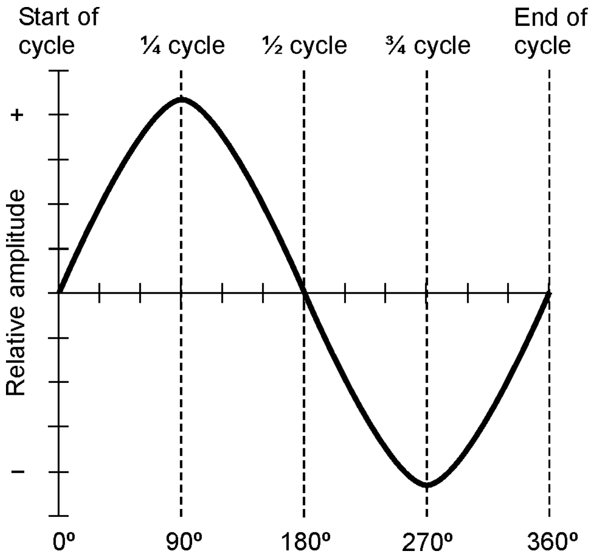

9-9 A cycle is divided into 360 equal parts, called *degrees.*

Radians

The other method of specifying phase is to divide the cycle into 2π equal parts, where π (pi) is a geometric constant equal to the number of diameters of any circle that can be laid end to end around the circumference of that circle. This constant is approximately equal to 3.14159. A *radian* (rad) of phase is thus equal to about 57.3°. Sometimes, the frequency of an ac wave is measured in *radians per second* (rad/s) rather than in hertz. Because there are about 6.28 radians in a complete cycle of 360°, the *angular frequency* of a wave, in radians per second, is equal to about 6.28 times the frequency in hertz. Radians of phase are used mainly by physicists.

Phase Difference

Even if two ac waves have exactly the same frequency, they can have different effects because they are out of sync with each other. This is especially true when ac waves are added together to produce a third, or *composite*, wave.

If two pure ac sine waves have identical frequencies and identical amplitudes but differ in phase by 180° (a half cycle), they cancel each other out, and the composite wave is zero; it ceases to exist! If the two waves are exactly in phase, the composite wave has the same frequency, but twice the amplitude, of either signal alone.

If two pure ac sine waves have the same frequency but different amplitudes, and if they differ in phase by 180°, the composite signal has the same frequency as the originals, and an amplitude equal to the difference between the two. If two such waves are exactly in phase, the composite has the same frequency as the originals, and an amplitude equal to the sum of the two.

If two pure ac sine waves have the same frequency but differ in phase by some odd amount such as 75° or 110°, the resulting signal has the same frequency, but does not have the same waveshape as either of the original signals. The variety of such cases is infinite.

Household electricity from 117-V wall outlets consists of a 60-Hz sine wave with only one phase component. But the energy is transmitted over long distances in three phases, each differing by 120° or one-third of a cycle. This is what is meant by *three-phase ac.* Each of the three ac waves carries one-third of the total power in a utility transmission line.

Expressions of Amplitude

Amplitude is also called *magnitude*, *level*, *strength*, or *intensity.* Depending on the quantity being measured, the amplitude of an ac wave can be specified in amperes (for current), volts (for voltage), or watts (for power). In addition to this, there are several different ways in which amplitude can be expressed.

Instantaneous Amplitude

The *instantaneous amplitude* of an ac wave is the amplitude at some precise moment, or instant, in time. This constantly changes. The manner in which it varies depends on the waveform. Instantaneous amplitudes are represented by individual points on the wave curves.

Peak Amplitude

The *peak* (pk) *amplitude* of an ac wave is the maximum extent, either positive or negative, that the instantaneous amplitude attains. In many situations, the positive and negative peak amplitudes of

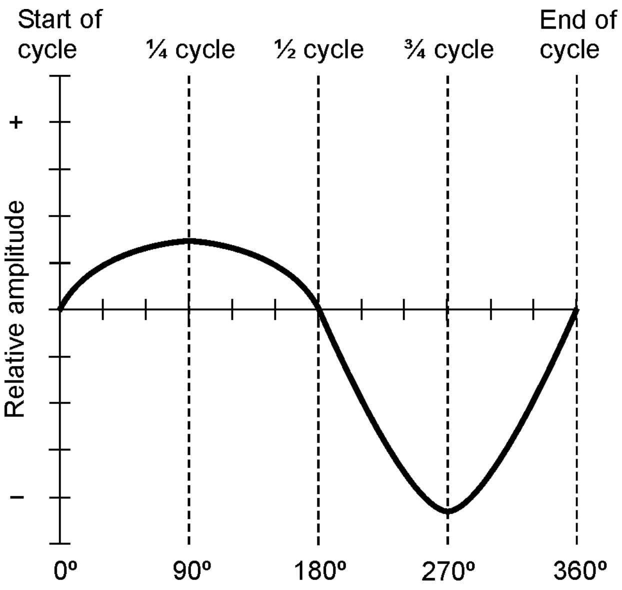

9-10 A wave with unequal positive and negative peak amplitudes.

an ac wave are the same. But sometimes they differ. Figure 9-9 is an example of a wave in which the positive peak amplitude is the same as the negative peak amplitude. Figure 9-10 is an illustration of a wave that has different positive and negative peak amplitudes.

Peak-to-Peak Amplitude

The *peak-to-peak* (pk-pk) *amplitude* of a wave is the net difference between the positive peak amplitude and the negative peak amplitude (Fig. 9-11). The peak-to-peak amplitude is equal to the positive peak amplitude plus the negative peak amplitude. When the positive and negative peak amplitudes of an ac wave are equal, the peak-to-peak amplitude is exactly twice the peak amplitude.

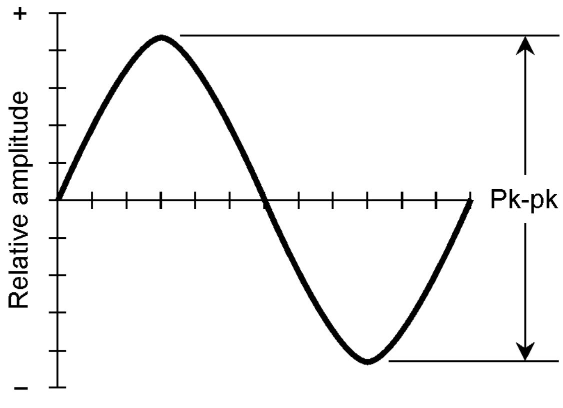

9-11 Peak-to-peak (pk-pk) amplitude of a sine wave.

Root-Mean-Square Amplitude

Often, it is necessary to express the *effective amplitude* of an ac wave. This is the voltage, current, or power that a dc source would have to produce in order to have the same general effect as a given ac wave. When you say a wall outlet provides 117 V, you mean 117 effective volts. This is not the same as the peak or peak-to-peak voltage.

The most common expression for effective ac intensity is called the *root-mean-square* (rms) *amplitude.* The terminology reflects the fact that the ac wave is mathematically operated on by taking the square root of the mean (average) of the square of all its instantaneous amplitudes.

In the case of a perfect ac sine wave, the rms value is equal to 0.707 times the peak value, or 0.354 times the peak-to-peak value. Conversely, the peak value is 1.414 times the rms value, and the peak-to-peak value is 2.828 times the rms value. The rms amplitude is often specified when talking about utility ac, radio-frequency (RF) ac, and audio-frequency (AF) ac.

For a perfect square wave, the rms value is the same as the peak value, and half the peak-to-peak value. For sawtooth and irregular waves, the relationship between the rms value and the peak value depends on the exact shape of the wave. But the rms value is never greater than the peak value for any type of ac wave.

Superimposed DC

Sometimes a wave has components of both ac and dc. The simplest example of an ac/dc combination is illustrated by the connection of a dc voltage source, such as a battery, in series with an ac voltage source, like the utility mains. An example is shown in the schematic diagram of Fig. 9-12. Imagine connecting a 12-V automotive battery in series with the wall outlet. (Do not try this experiment in real life!) When this is done, the ac wave is displaced either positively or negatively by 12 V, depending on the polarity of the battery. This results in a sine wave at the output, but one peak is 24 V (twice the battery voltage) more than the other.

Any ac wave can have dc components along with it. If the dc component exceeds the peak value of the ac wave, then fluctuating, or pulsating, dc will result. This would happen, for example, if a 200-V dc source were connected in series with the output of a common utility ac outlet, which has peak voltages of approximately ±165 V. Pulsating dc would appear, with an average value of 200 V but with instantaneous values much higher and lower. The waveshape in this case is shown in Fig. 9-13.

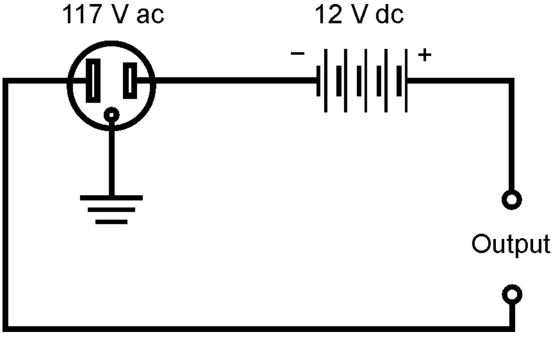

9-12 Connection of a dc source in series with an ac source.

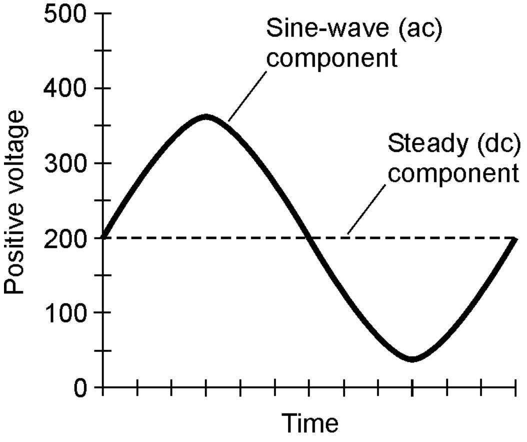

9-13 Waveform resulting from a 117-V ac sine-wave source connected in series with a +200-V dc source.

The Generator

Alternating current can be generated by a rotating coil of wire inside a powerful magnet, as shown in Fig. 9-14. An ac voltage appears between the ends of the wire coil. The ac voltage that a generator can produce depends on the strength of the magnet, the number of turns in the wire coil, and the speed at which the magnet or coil rotates. The ac frequency depends only on the speed of rota-

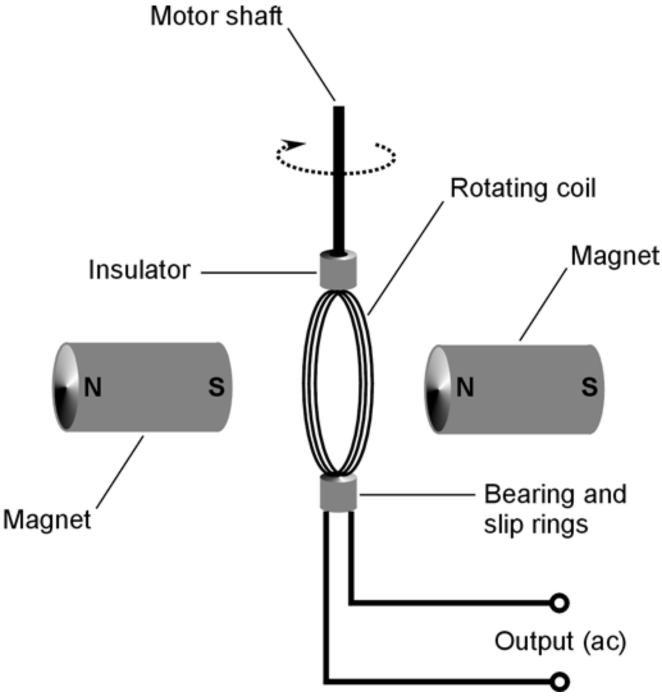
Motor shaft
Rotating coil
Insulator
Magnet
N
S
N
S
Magnet
Bearing and slip rings
Output (ac)

9-14 A functional diagram of an ac generator.

tion. Normally, for utility ac, this speed is 3600 revolutions per minute (rpm), or 60 complete revolutions per second (rps), so the ac output frequency is 60 Hz.

When a load, such as a light bulb or heater, is connected to an ac generator, it becomes more difficult, mechanically, to turn the generator shaft, compared to when there is nothing connected to the output. As the amount of electrical power demanded from a generator increases, so does the mechanical power required to drive it. This is why it is impossible to connect a generator to a stationary bicycle and pedal an entire city into electrification. There's no way to get something for nothing. The electrical power that comes out of a generator can never be more than the mechanical power driving it. In fact, there is always some energy lost, mainly as heat in the generator. Your legs might generate enough power to run a small radio or television set, but nowhere near enough to provide electricity for a household.

The *efficiency* of a generator is the ratio of the electrical power output to the mechanical driving power, both measured in the same units (such as watts or kilowatts), multiplied by 100 to get a percentage. No generator is 100 percent efficient, but a good one can come fairly close.

At power plants, generators are driven by massive turbines. The turbines are turned by various natural sources of energy such as moving water, steam heated by combustion of fossil fuels, or steam taken directly from deep inside the earth. These energy sources can provide tremendous mechanical power, and this is why power plants can produce megawatts of electrical power.

Why Alternating and Not Direct?

Do you wonder why ac is used at all? Isn't it a lot more complicated than dc? Well, ac may be more complicated in theory, but in practice it is a lot simpler to use when it is necessary to provide electricity to a large number of people.

Alternating current lends itself well to being transformed to lower or higher voltages, according to the needs of electrical apparatus. It is not so easy to change dc voltages. Electrochemical cells produce dc directly, but they are impractical for the needs of large populations. Serving millions of consumers requires the immense power of falling or flowing water, the ocean tides, wind, fossil fuels, controlled nuclear reactions, or geothermal heat. All of these energy sources can be used to drive turbines that turn ac generators.

Technology is advancing in the realm of solar-electric energy; someday a significant part of our electricity might come from photovoltaic power plants. These would generate dc. High voltages could be attained by connecting giant arrays of solar panels in series. But there would be a problem transforming this voltage down to manageable levels for consumer use.

Thomas Edison is said to have favored dc over ac for electrical power transmission in the early days, as the electric utilities were first being devised and constructed. His colleagues argued that ac would work better. But perhaps Edison knew something that his contemporaries did not. There is one advantage to dc in utility applications, and it involves the transmission of energy over great distances using wires. Direct currents, at extremely high voltages, are transported more efficiently than alternating currents. The wire has less effective resistance with dc than with ac, and there is less energy lost in the magnetic fields around the wires. Direct-current *high-tension* transmission lines are being considered for future use. Right now, the main problem is expense. Sophisticated power-conversion equipment is needed. If the cost can be brought within reason, Edison will be vindicated.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. Which of the following can vary with ac, but never with dc?
   (a) Power
   (b) Voltage
   (c) Frequency
   (d) Amplitude

2. The length of time between a point in one cycle and the same point in the next cycle of an ac wave is the
   (a) frequency.
   (b) magnitude.
   (c) period.
   (d) polarity.

3. On a spectrum analyzer, an ac signal having only one frequency component looks like
   (a) a single pip.
   (b) a sine wave.
   (c) a square wave.
   (d) a sawtooth wave.

4. The period of an ac wave, in seconds, is
   (a) the same as the frequency in hertz.
   (b) not related to the frequency in any way.
   (c) equal to 1 divided by the frequency in hertz.
   (d) equal to the peak amplitude in volts divided by the frequency in hertz.

5. The sixth harmonic of an ac wave whose period is 1.000 millisecond (1.000 ms) has a frequency of
   (a) 0.006 Hz.
   (b) 167.0 Hz.
   (c) 7.000 kHz.
   (d) 6.000 kHz.

6. A degree of phase represents
   (a) 6.28 cycles.
   (b) 57.3 cycles.
   (c) $^1/_{60}$ of a cycle.
   (d) $^1/_{360}$ of a cycle.

7. Suppose that two ac waves have the same frequency but differ in phase by exactly $^1\!/_{20}$ of a cycle. What is the phase difference between these two waves?
   (a) 18°
   (b) 20°
   (c) 36°
   (d) 5.73°

8. Suppose an ac signal has a frequency of 1770 Hz. What is its angular frequency?
   (a) 1770 rad/s
   (b) 11,120 rad/s
   (c) 282 rad/s
   (d) Impossible to determine from the data given

9. A triangular wave exhibits
   (a) an instantaneous rise and a defined decay.
   (b) a defined rise and an instantaneous decay.
   (c) a defined rise and a defined decay, and the two are equal.
   (d) an instantaneous rise and an instantaneous decay.

10. Three-phase ac
    (a) has sawtooth waves that add together in phase.
    (b) consists of three sine waves in different phases.
    (c) is a sine wave with exactly three harmonics.
    (d) is of interest only to physicists.

11. If two perfect sine waves have the same frequency and the same amplitude, but are in opposite phase, the composite wave
    (a) has twice the amplitude of either input wave alone.
    (b) has half the amplitude of either input wave alone.
    (c) is complex, but has the same frequency as the originals.
    (d) has zero amplitude (that is, it does not exist), because the two input waves cancel each other out.

12. If two perfect sine waves have the same frequency and the same phase, the composite wave
    (a) is a sine wave with an amplitude equal to the difference between the amplitudes of the two input waves.
    (b) is a sine wave with an amplitude equal to the sum of the amplitudes of the two original waves.
    (c) is not a sine wave, but has the same frequency as the two input waves.
    (d) has zero amplitude (that is, it does not exist), because the two input waves cancel each other out.

13. In a 117-V rms utility circuit, the positive peak voltage is approximately
    (a) +82.7 V.
    (b) +165 V.
    (c) +234 V.
    (d) +331 V.

14. In a 117-V rms utility circuit, the peak-to-peak voltage is approximately
    (a) 82.7 V.
    (b) 165 V.
    (c) 234 V.
    (d) 331 V.

15. In a perfect sine wave, the peak-to-peak amplitude is equal to
    (a) half the peak amplitude.
    (b) the peak amplitude.
    (c) 1.414 times the peak amplitude.
    (d) twice the peak amplitude.

16. If a 45-V dc battery is connected in series with the 117-V rms utility mains as shown in Fig. 9-15, the peak voltages will be approximately
    (a) +210 V and −120 V.
    (b) +162 V and −72 V.
    (c) +396 V and −286 V.
    (d) +117 V and −117V.

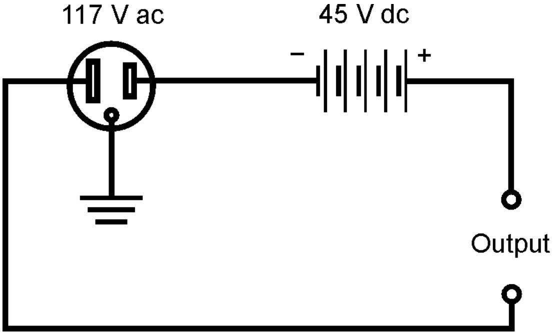

9-15 Illustration for Quiz Question 16.

17. In the situation described in question 16 and illustrated in Fig. 9-15, the peak-to-peak voltage will be approximately
    (a) 117 V.
    (b) 210 V.
    (c) 331 V.
    (d) 396 V.

18. Which one of the following does *not* affect the power output available from a particular ac generator?
    - (a) The strength of the magnet
    - (b) The number of turns in the coil
    - (c) The type of natural energy source used
    - (d) The speed of rotation of the coil or magnet

19. If a 175-V dc source were connected in series with the utility mains from a standard wall outlet, the result would be
    - (a) smooth dc at a constant voltage.
    - (b) pure ac with equal peak voltages.
    - (c) ac with one peak voltage greater than the other.
    - (d) fluctuating dc.

20. An advantage of ac over dc in utility applications is the fact that
    - (a) ac is easier to transform from one voltage to another.
    - (b) ac is transmitted with lower loss in wires.
    - (c) ac can be easily obtained from dc generators.
    - (d) ac can be generated with less-dangerous by-products.

10
CHAPTER

Inductance

IN THIS CHAPTER, YOU'LL LEARN ABOUT ELECTRICAL COMPONENTS THAT OPPOSE THE FLOW OF AC BY temporarily storing energy as magnetic fields. These devices are called *inductors*, and their action is known as *inductance.* Inductors often, but not always, consist of wire coils. Sometimes a length of wire, or a pair of wires, is used as an inductor.

The Property of Inductance

Suppose you have a wire 1 million miles long (about 1.6 million kilometers). Imagine that you make this wire into a huge loop, and connect its ends to the terminals of a battery (Fig. 10-1). An electrical current will flow through the loop of wire, but this is only part of the picture.

If the wire was short, the current would begin to flow immediately, and it would attain a level limited by the resistance in the wire and in the battery. But because the wire is extremely long, it takes a while for the electrons from the negative terminal to work their way around the loop to the positive terminal. It will take a little time for the current to build up to its maximum level.

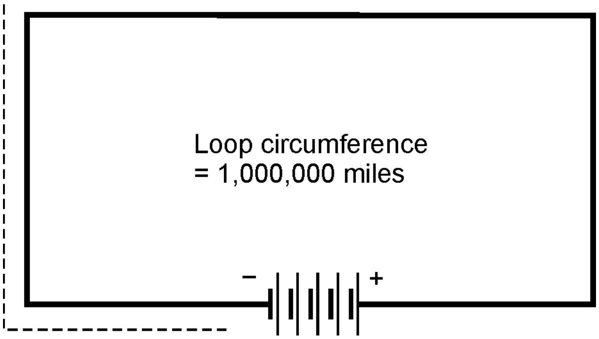

10-1 A huge, imaginary loop of wire can be used to illustrate the principle of inductance.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

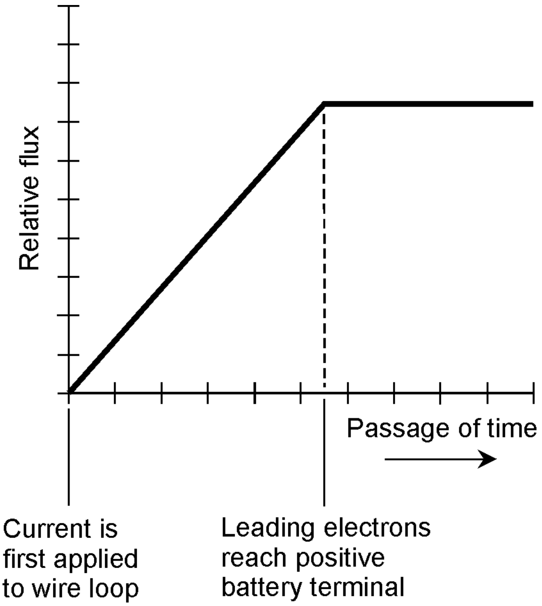

10-2 Relative magnetic flux in and around a huge loop of wire connected to a current source, as a function of time.

The magnetic field produced by the loop will be small during the first few moments when current flows in only part of the loop. The magnetic field will build up as the electrons get around the loop. Once a steady current is flowing around the entire loop, the magnetic field will have reached its maximum quantity and will level off (see Fig. 10-2). A certain amount of energy is stored in this magnetic field. The amount of stored energy depends on the *inductance* of the loop, which is a function of its overall size. Inductance, as a property or as a mathematical variable, is symbolized by an italicized, uppercase letter *L*. The loop constitutes an *inductor*, the symbol for which is an uppercase, nonitalicized letter L.

Practical Inductors

It is impractical to make wire loops 1 million miles in circumference. But lengths of wire can be coiled up. When this is done, the magnetic flux is increased for a given length of wire compared with the flux produced by a single-turn loop.

The magnetic flux density inside a coil is multiplied when a ferromagnetic core is placed within it. The increase in flux density has the effect of increasing the inductance, too, so *L* is many times greater with a ferromagnetic core than with an air core or a nonmagnetic core such as plastic or wood. The current that an inductor can handle depends on the diameter (gauge) of the wire. But the value of *L* is a function of the number of turns in the coil, the diameter of the coil itself, and the overall shape of the coil.

In general, the inductance of a coil is directly proportional to the number of turns of wire. Inductance is directly proportional to the diameter of the coil. The length of a coil, given a certain number of turns and a certain diameter, has an effect as well. If a coil having a certain number of turns and a certain diameter is "stretched out," its inductance decreases. Conversely, if it is "squashed up," its inductance increases.

The Unit of Inductance

When a battery is first connected across an inductor, the current builds up at a rate that depends on the inductance. The greater the inductance, the slower the rate of current buildup for a given battery voltage. The unit of inductance is an expression of the ratio between the rate of current buildup and the voltage across an inductor. An inductance of 1 *henry* (1 H) represents a potential difference of 1 volt (1 V) across an inductor within which the current is changing at the rate of 1 ampere per second (1 A/s).

The henry is a huge unit of inductance. You won't often see an inductor this large, although some power-supply filter chokes have inductances up to several henrys. Usually, inductances are expressed in *millihenrys* (mH), *microhenrys* (μH), or *nanohenrys* (nH). You should know your prefix multipliers by now, but in case you've forgotten:

$$1 \text{ mH} = 0.001 \text{ H} = 10^{-3} \text{ H}$$
$$1 \text{ μH} = 0.001 \text{ mH} = 10^{-6} \text{ H}$$
$$1 \text{ nH} = 0.001 \text{ μH} = 10^{-9} \text{ H}$$

Small coils with few turns of wire produce small inductances, in which the current changes quickly and the induced voltages are small. Large coils with ferromagnetic cores, and having many turns of wire, have high inductances in which the current changes slowly and the induced voltages are large. The current from a battery, building up or dying down through a high-L coil, can give rise to a deadly potential difference between the end terminals of the coil—many times the voltage of the battery itself. This is how spark coils work in internal combustion engines. Be careful around them!

Inductors in Series

When the magnetic fields around inductors do not interact, inductances in series add like resistances in series. The total value is the sum of the individual values. It's important to be sure that you are using the same size units for all the inductors when you add their values. After that, you can convert the result to any inductance unit you want.

Problem 10-1

Suppose three 40.0-μH inductors are connected in series, and there is no interaction, or *mutual inductance*, among them (Fig. 10-3). What is the total inductance?

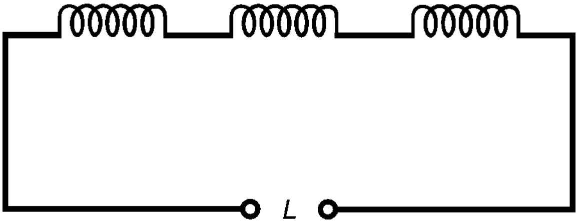

10-3 Inductances in series simply add up, as long as the inductors do not interact.

Add up the values. Call the inductances of the individual components L_1, L_2, and L_3, and the total inductance L. Then $L = L_1 + L_2 + L_3 = 40.0 + 40.0 + 40.0 = 120\ \mu\text{H}$.

Problem 10-2

Imagine three inductors, with no mutual inductance, with values of 20.0 mH, 55.0 μH, and 400 nH. What is the total inductance, in millihenrys, of these components if they are connected in series as shown in Fig. 10-3?

First, convert all the inductances to the same units. Microhenrys are a good choice because that unit makes the calculation process the least messy. Call $L_1 = 20.0\ \text{mH} = 20{,}000\ \mu\text{H}$, $L_2 = 55.0\ \mu\text{H}$, and $L_3 = 400\ \text{nH} = 0.400\ \mu\text{H}$. The total inductance is therefore $L = 20{,}000 + 55.0 + 0.400 = 20{,}055.4\ \mu\text{H}$. This is 20.1 mH after converting and rounding off.

Inductors in Parallel

If there is no mutual inductance among two or more parallel-connected inductors, their values add up like the values of resistors in parallel. Suppose you have inductances L_1, L_2, L_3, . . . , L_n all connected in parallel. Then you can find the reciprocal of the total inductance, $1/L$, using the following formula:

$$1/L = 1/L_1 + 1/L_2 + 1/L_3 + \ldots + 1/L_n$$

The total inductance, L, is found by taking the reciprocal of the number you get for $1/L$. Again, as with inductances in series, it's important to remember that all the units have to agree during the calculation process. Once you have completed the calculation, you can convert the result to any inductance unit.

Problem 10-3

Suppose there are three inductors, each with a value of 40 μH, connected in parallel with no mutual inductance, as shown in Fig. 10-4. What is the net inductance of the combination?

Let's call the inductances $L_1 = 40\ \mu\text{H}$, $L_2 = 40\ \mu\text{H}$, and $L_3 = 40\ \mu\text{H}$. Use the preceding formula to obtain $1/L = 1/40 + 1/40 + 1/40 = 3/40 = 0.075$. Then $L = 1/0.075 = 13.333\ \mu\text{H}$. This should be rounded off to 13 μH, because the original inductances are specified to only two significant digits.

Problem 10-4

Imagine four inductors in parallel, with no mutual inductance and values of $L_1 = 75.0\ \text{mH}$, $L_2 = 40.0\ \text{mH}$, $L_3 = 333\ \mu\text{H}$, and $L_4 = 7.00\ \text{H}$. What is the net inductance of this combination?

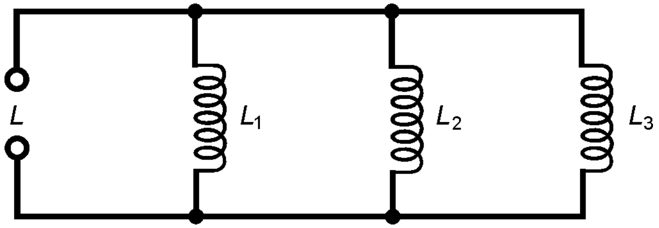

10-4 Inductances in parallel.

You can use henrys, millihenrys, or microhenrys as the standard units in this problem. Suppose you decide to use henrys. Then $L_1 = 0.0750$ H, $L_2 = 0.0400$ H, $L_3 = 0.000333$ H, and $L_4 = 7.00$ H. Use the preceding formula to obtain $1/L = 13.33 + 25.0 + 3003 + 0.143 = 3041.473$. The reciprocal of this is the inductance $L = 0.00032879$ H = 328.79 μH. This should be rounded off to 329 μH. This is only a little less than the value of the 333 μH inductor alone.

If there are several inductors in parallel, and one of them has a value that is much smaller than the values of all the others, then the total inductance is a little smaller than the value of the smallest inductor.

Interaction among Inductors

In real-world circuits, there is almost always some mutual inductance between or among solenoidal coils. The magnetic fields extend significantly outside such coils, and mutual effects are difficult to avoid or eliminate. The same is true between and among lengths of wire, especially at high ac frequencies. Sometimes, mutual inductance has no detrimental effect, but in some situations it is not wanted. Mutual inductance can be minimized by using *shielded* wires and *toroidal* inductors. The most common shielded wire is *coaxial cable.* Toroidal inductors are discussed later in this chapter.

Coefficient of Coupling

The *coefficient of coupling*, symbolized *k*, is an expression of the extent to which two inductors interact. It is specified as a number ranging from 0 (no interaction) to 1 (the maximum possible interaction). Two coils separated by a sheet of solid iron, or by a great distance, have a coefficient of coupling of zero ($k = 0$); two coils wound on the same form, one right over the other, have the maximum possible coefficient of coupling ($k = 1$). Sometimes, the coefficient of coupling is multiplied by 100 and expressed as a percentage from 0 to 100 percent.

Mutual Inductance

The *mutual inductance* between two inductors is symbolized *M*, and is expressed in the same units as inductance: henrys, millihenrys, microhenrys, or nanohenrys. The value of *M* is a function of the values of the inductors, and also of the coefficient of coupling.

In the case of two inductors having values of L_1 and L_2 (both expressed in the same size units), and with a coefficient of coupling equal to *k*, the mutual inductance *M* is found by multiplying the inductance values, taking the square root of the result, and then multiplying by *k.* Mathematically:

$$M = k(L_1 L_2)^{1/2}$$

where the ½ power represents the square root. The value of *M* thus obtained will be in the same size unit as the values of the inductance you input to the equation.

Effects of Mutual Inductance

Mutual inductance can either increase or decrease the net inductance of a pair of series-connected coils, compared with the condition of zero mutual inductance. The magnetic fields around the coils either reinforce each other or oppose each other, depending on the phase relationship of the ac applied to them. If the two ac waves (and thus the magnetic fields they produce) are in phase, the inductance is increased compared with the condition of zero mutual inductance. If the two waves are in opposing phase, the net inductance is decreased relative to the condition of zero mutual inductance.

When two inductors are connected in series and there is *reinforcing* mutual inductance between them, the total inductance L is given by the following formula:

$$L = L_1 + L_2 + 2M$$

where L_1 and L_2 are the inductances, and M is the mutual inductance. All inductances must be expressed in the same size units.

When two inductors are connected in series and the mutual inductance is *opposing*, the total inductance L is given by this formula:

$$L = L_1 + L_2 - 2M$$

where, again, L_1 and L_2 are the values of the individual inductors.

It is possible for mutual inductance to increase the total series inductance of a pair of coils by as much as a factor of 2, if the coupling is total and if the flux reinforces. Conversely, it is possible for the inductances of two coils to completely cancel each other. If two equal-valued inductors are connected in series so their fluxes oppose (or *buck* each other) and $k = 1$, the result is theoretically zero inductance.

Problem 10-5

Suppose two coils, having inductances of 30 μH and 50 μH, are connected in series so that their fields reinforce, as shown in Fig. 10-5. Suppose that the coefficient of coupling is 0.500. What is the total inductance of the combination?

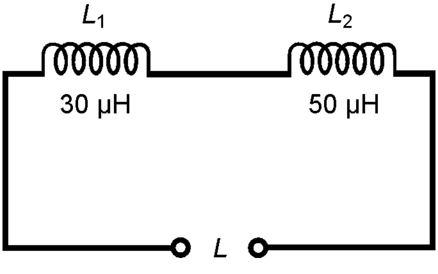

10-5 Illustration for Problem 10-5.

First, calculate M from k. According to the formula for this, given previously, $M = 0.500(50 \times 30)^{1/2} = 19.4$ μH. Then figure the total inductance. It is equal to $L = L_1 + L_2 + 2M = 30 + 50 + 38.8 = 118.8$ μH, rounded to 120 μH because only two significant digits are justified.

Problem 10-6

Imagine two coils with inductances of $L_1 = 835$ μH and $L_2 = 2.44$ mH. Suppose they are connected in series so that their coefficient of coupling is 0.922, acting so that the coils oppose each other, as shown in Fig. 10-6. What is the net inductance of the pair?

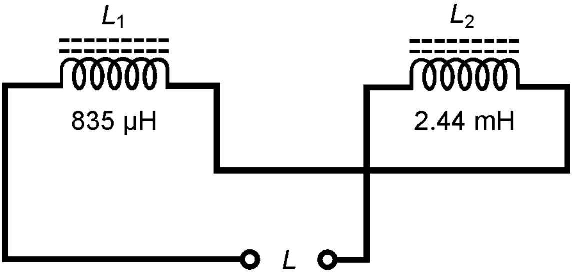

10-6 Illustration for Problem 10-6.

First, calculate M from k. The coil inductances are specified in different units. Let's use microhenrys for our calculations, so $L_2 = 2440$ μH. Then $M = 0.922(835 \times 2440)^{1/2} = 1316$ μH. Then figure the total inductance. It is $L = L_1 + L_2 - 2M = 835 + 2440 - 2632 = 643$ μH.

Air-Core Coils

The simplest inductors (besides plain, straight lengths of wire) are coils. A coil can be wound on a hollow cylinder of plastic or other nonferromagnetic material, forming an *air-core coil.* In practice, the maximum attainable inductance for such coils is about 1 mH.

Air-core coils are used mostly in radio-frequency transmitters, receivers, and antenna networks. In general, the higher the frequency of ac, the less inductance is needed to produce significant effects. Air-core coils can be made to have almost unlimited current-carrying capacity, simply by using heavy-gauge wire and making the radius of the coil large. Air does not dissipate much energy in the form of heat. It's efficient, even though it has low permeability.

Ferromagnetic Cores

Ferromagnetic substances can be crushed into dust and then bound into various shapes, providing core materials that greatly increase the inductance of a coil having a given number of turns. Depending on the mixture used, the increase in flux density can range from a factor of a few times, up through many thousands of times. A small coil can thus be made to have a large inductance. There are two main types of ferromagnetic material in common use as coil cores. These substances are known as *powdered iron* and *ferrite.*

Advantages and Limitations

Powdered-iron cores are common at high and very high radio frequencies. Ferrite is a special form of powdered iron that has exceptionally high permeability, causing a great concentration of magnetic flux lines within the coil. Ferrite is used at audio frequencies, as well as at low, medium, and high radio frequencies. Coils using these materials can be made much smaller, physically, than can air-core coils having the same inductance.

The main trouble with ferromagnetic cores is that, if the coil carries more than a certain amount of current, the core will *saturate.* This means that the ferromagnetic material is holding as much flux as it possibly can. When a core becomes saturated, any further increase in coil current will not produce a corresponding increase in the magnetic flux in the core. The result is that the inductance changes, decreasing with coil currents that are more than the critical value. In extreme cases, ferromagnetic cores can also waste considerable power as heat. This makes a coil *lossy.*

Permeability Tuning

Solenoidal coils can be made to have variable inductance by sliding ferromagnetic cores in and out of them. The frequency of a radio circuit can be adjusted in this way, as you'll learn later in this book.

Because moving the core in and out of a coil changes the effective permeability within the coil, this method of tuning is called *permeability tuning.* The in/out motion can be precisely controlled by attaching the core to a screw shaft, and anchoring a nut at one end of the coil (Fig. 10-7). As the screw shaft is rotated clockwise, the core enters the coil, and the inductance increases. As the screw shaft is rotated counterclockwise, the core moves out of the coil, and the inductance decreases.

Toroids

Inductor coils do not have to be wound on cylindrical forms, or on cylindrical ferromagnetic cores. There's another coil geometry, called the *toroid.* It gets its name from the shape of the ferromagnetic core. The coil is wound over a core having this shape (Fig. 10-8), which resembles a donut or bagel.

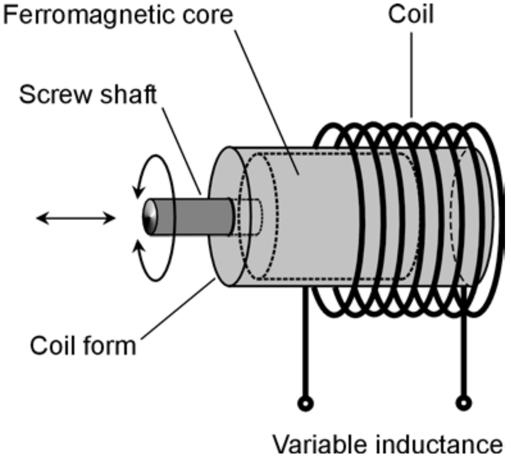

10-7 Permeability tuning can be accomplished by moving a ferromagnetic core in and out of a solenoidal coil.

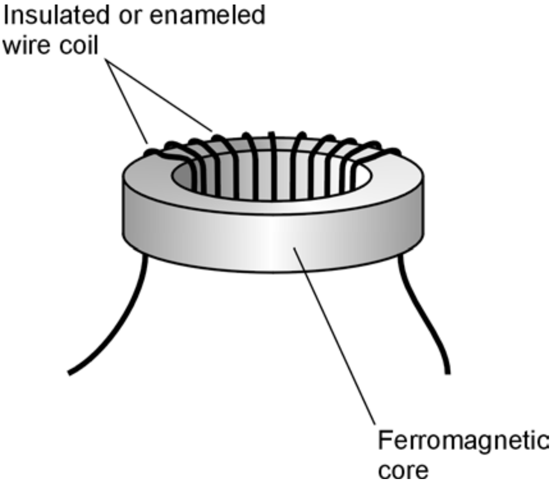

10-8 A toroidal coil is wound on a donut-shaped ferromagnetic core.

There are several advantages to toroidal coils over solenoidal, or cylindrical, ones. First, fewer turns of wire are needed to get a certain inductance with a toroid compared to a solenoid. Second, a toroid can be physically smaller for a given inductance and current-carrying capacity. Third, practically all the flux is contained within the core material. This reduces unwanted mutual inductances with components near the toroid.

Toroidal coils have limitations, too. It is more difficult to permeability-tune a toroidal coil than it is to tune a solenoidal one. Toroidal coils are harder to wind than solenoidal ones. Sometimes, mutual inductance between or among physically separate coils is actually desired; with a toroid, the coils have to be wound on the same form for this to be possible.

Pot Cores

There is another way to confine the magnetic flux in a coil so that unwanted mutual inductance does not occur: wrap ferromagnetic core material around a coil (Fig. 10-9). A wraparound core of this sort is known as a *pot core.*

A typical pot core comes in two halves, inside one of which the coil is wound. Then the parts are assembled and held together by a bolt and nut. The entire assembly looks like a miniature oil tank. The wires come out of the core through small holes or slots.

Pot cores have the same advantages as toroids. The core tends to prevent the magnetic flux from extending outside the physical assembly. Inductance is greatly increased compared to solenoidal windings having a comparable number of turns. In fact, pot cores are even better than toroids if the main objective is to get a large inductance in a small space. The main disadvantage of a pot core is that tuning, or adjustment of the inductance, is all but impossible. The only way to do it is by switching in different numbers of turns, using taps at various points on the coil.

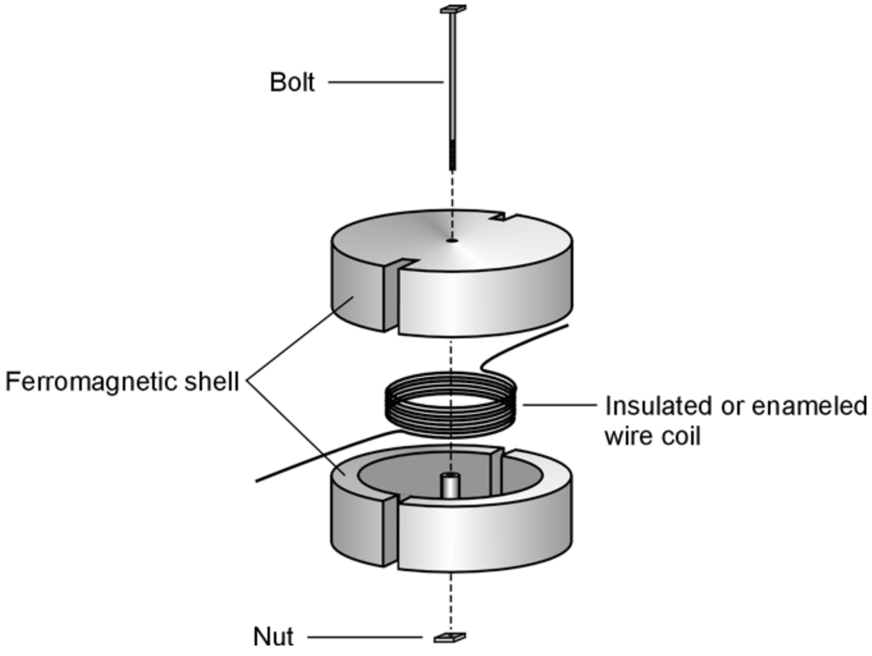

10-9 Exploded view of a pot core. The coil winding is inside the ferromagnetic shell.

Filter Chokes

The largest values of inductance that can be obtained in practice are on the order of several henrys. The primary use of a coil this large is to smooth out the pulsations in direct current that result when ac is *rectified* in a power supply. This type of coil is known as a *filter choke.* You'll learn more about power supplies later in this book.

Inductors at AF

Inductors for audio frequency (AF) applications range in value from a few millihenrys up to about 1 H. They are almost always toroidally wound, or are wound in a pot core, or comprise part of an audio transformer. Ferromagnetic cores are the rule.

Inductors can be used in conjunction with moderately large values of capacitance in order to obtain *AF-tuned circuits.* However, in recent years, audio tuning has been largely taken over by active components, particularly *integrated circuits.*

Inductors at RF

The radio frequency (RF) spectrum ranges from a few kilohertz to well above 100 GHz. At the low end of this range, inductors are similar to those at AF. As the frequency increases, cores having lower permeability are used. Toroids are common up through about 30 MHz. Above that frequency, air-core coils are more often used.

In RF applications, coils are routinely connected in series or in parallel with capacitors to obtain tuned circuits. Other arrangements yield various characteristics of *attenuation versus frequency*, serving to let signals at some frequencies pass through, while rejecting signals at other frequencies. You'll learn more about this in the discussion about *resonance* in Chap. 17.

Transmission-Line Inductors

At frequencies about 100 MHz, another type of inductor becomes practical. This is the type formed by a length of *transmission line.* A transmission line is generally used to get energy from one place to another. In radio communications, transmission lines get energy from a transmitter to an antenna, and from an antenna to a receiver.

Most transmission lines are found in either of two geometries, the *parallel-wire* type or the *coaxial* type. A parallel-wire transmission line consists of two wires running alongside each other with constant spacing (Fig. 10-10). The spacing is maintained by polyethylene rods molded at regular in-

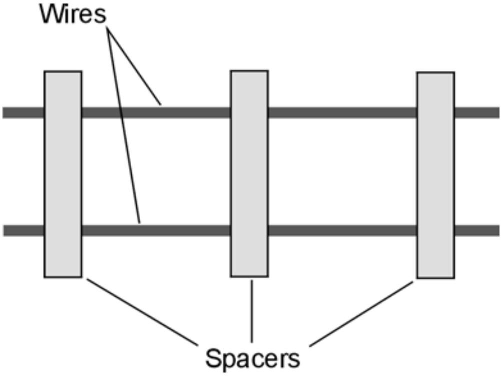

10-10 Parallel-wire transmission line. The spacers are made of sturdy insulating material.

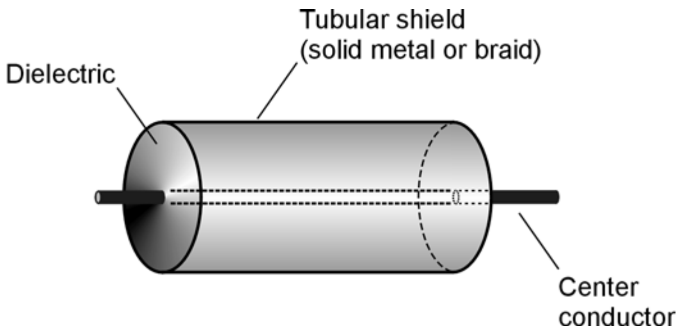

10-11 Coaxial transmission line. The dielectric material keeps the center conductor along the axis of the tubular shield.

tervals to the wires, or by a solid web of polyethylene. The substance separating the wires is called the *dielectric* of the transmission line. A coaxial transmission line has a wire conductor surrounded by a tubular braid or pipe (Fig. 10-11). The wire is kept at the center of this tubular *shield* by means of polyethylene beads, or more often, by solid or foamed polyethylene, all along the length of the line.

Line Inductance

Short lengths of any type of transmission line behave as inductors, as long as the line length is less than 90° ($1/4$ of a wavelength). At 100 MHz, 90° in free space is 75 cm, or a little more than 2 ft. In general, if f is the frequency in megahertz, then $1/4$ wavelength in free space, expressed in centimeters (s_{cm}), is given by this formula:

$$s_{\text{cm}} = 7500/f$$

The length of a quarter-wavelength section of transmission line is shortened from the free-space quarter wavelength by the effects of the dielectric. In practice, $1/4$ wavelength along the line can be anywhere from about 0.66 (or 66 percent) of the free-space length for coaxial lines with solid polyethylene dielectric to about 0.95 (or 95 percent) of the free-space length for parallel-wire line with spacers molded at intervals of several centimeters. The factor by which the wavelength is shortened is called the *velocity factor* of the line.

The shortening of the wavelength in a transmission line, compared with the wavelength in free space, is a result of a slowing down of the speed with which the radio signals move in the line compared with their speed in space (the speed of light). If the velocity factor of a line is given by v, then the preceding formula for the length of a quarter-wave line, in centimeters, becomes:

$$s_{\text{cm}} = 7500v/f$$

Very short lengths of line—a few electrical degrees—produce small values of inductance. As the length approaches $1/4$ wavelength, the inductance increases.

Transmission line inductors behave differently than coils in one important way: the inductance of a coil, particularly an air-core coil, is independent of the frequency. But the inductance of a transmission-line section changes as the frequency changes. At first, the inductance becomes larger as the frequency increases. At a certain limiting frequency, the inductance becomes theoretically infinite. Above that frequency, the line becomes *capacitive* rather than inductive. You'll learn about capacitance in the next chapter.

Unwanted Inductances

Any length of wire has some inductance. As with a transmission line, the inductance of a wire increases as the frequency increases. Wire inductance is more significant at RF than at AF.

In some cases, especially in radio communications equipment, the inductance of, and among, wires can become a major problem. Circuits can oscillate when they should not. A receiver might respond to signals that it's not designed to intercept. A transmitter can send out signals on unauthorized and unintended frequencies. The frequency response of any circuit can be altered, degrading the performance of the equipment. Sometimes the effects of this *stray inductance* are so small that they are not important; this might be the case in a stereo hi-fi set located at a distance from other electronic equipment. But in some situations, stray inductance can cause serious equipment malfunctions.

A good way to minimize stray inductance is to use coaxial cables between and among sensitive circuits or components. The shield of the cable is connected to the *common ground* of the apparatus. In some cases, enclosing individual circuits in metal boxes can prevent stray inductance from causing feedback and other problems.

Quiz

Refer to the text in this chapter if necessary. A good score is 18 correct. Answers are in the back of the book.

1. An inductor works by
   (a) charging a piece of wire.
   (b) storing energy as a magnetic field.
   (c) choking off dc.
   (d) introducing resistance into a circuit.

2. Which of the following does *not* affect the inductance of an air-core coil, if all other factors are held constant?
   (a) The frequency
   (b) The number of turns
   (c) The diameter of the coil
   (d) The length of the coil

3. In a small inductance
   (a) energy is stored and released slowly.
   (b) the current flow is always large.
   (c) the current flow is always small.
   (d) energy is stored and released quickly.

4. A ferromagnetic core is placed in an inductor mainly to
   (a) increase the current carrying capacity.
   (b) increase the inductance.
   (c) limit the current.
   (d) reduce the inductance.

5. Inductors in series, assuming there is no mutual inductance, combine
   (a) like resistors in parallel.
   (b) like resistors in series.
   (c) like batteries in series with opposite polarities.
   (d) in a way unlike any other type of component.

6. Suppose two inductors are connected in series, without mutual inductance. Their values are 33 mH and 55 mH. What is the net inductance of the combination?
   (a) 1.8 H
   (b) 22 mH
   (c) 88 mH
   (d) 21 mH

7. If the same two inductors (33 mH and 55 mH) are connected in parallel without mutual inductance, the combination will have a value of
   (a) 1.8 H.
   (b) 22 mH.
   (c) 88 mH.
   (d) 21 mH.

8. Suppose three inductors are connected in series without mutual inductance. Their values are 4.00 nH, 140 μH, and 5.07 H. For practical purposes, the net inductance will be very close to
   (a) 4.00 nH.
   (b) 140 μH.
   (c) 5.07 H.
   (d) none of the above.

9. Suppose the three inductors mentioned above are connected in parallel without mutual inductance. The net inductance will be close to
   (a) 4.00 nH.
   (b) 140 μH.
   (c) 5.07 H.
   (d) none of the above.

10. Suppose two inductors, each of 100 μH, are connected in series, and the coefficient of coupling is 0.40. The net inductance, if the coil fields reinforce each other, is
   (a) 50.0 μH.
   (b) 120 μH.
   (c) 200 μH.
   (d) 280 μH.

11. If the coil fields oppose in the foregoing series-connected arrangement, assuming the coefficient of coupling does not change, the net inductance is
    (a) 50.0 μH.
    (b) 120 μH.
    (c) 200 μH.
    (d) 280 μH.

12. Suppose two inductors, having values of 44.0 mH and 88.0 mH, are connected in series with a coefficient of coupling equal to 1.0 (the maximum possible mutual inductance). If their fields reinforce, the net inductance is approximately
    (a) 7.55 mH.
    (b) 132 mH.
    (c) 194 mH.
    (d) 256 mH.

13. If the fields in the previous situation oppose, assuming the coefficient of coupling does not change, the net inductance will be approximately
    (a) 7.55 mH.
    (b) 132 mH.
    (c) 194 mH.
    (d) 256 mH.

14. With permeability tuning, moving the core further into a solenoidal coil
    (a) increases the inductance.
    (b) reduces the inductance.
    (c) has no effect on the inductance, but increases the current-carrying capacity of the coil.
    (d) raises the frequency.

15. A significant advantage, in some situations, of a toroidal coil over a solenoid is the fact that
    (a) the toroid is easier to wind.
    (b) the solenoid cannot carry as much current.
    (c) the toroid is easier to tune.
    (d) the magnetic flux in a toroid is practically all within the core.

16. A major feature of a pot core inductor is
    (a) high current capacity.
    (b) large inductance in small volume.
    (c) excellent efficiency at very high frequencies.
    (d) ease of inductance adjustment.

17. As an inductor core material, air
    (a) has excellent efficiency.
    (b) has high permeability.
    (c) allows large inductance to exist in a small volume.
    (d) has permeability that can vary over a wide range.

18. At a frequency of 400 Hz, which is in the AF range, the most likely form for an inductor would be
    (a) air-core.
    (b) solenoidal.
    (c) toroidal.
    (d) transmission-line.

19. At a frequency of 95.7 MHz, which is in the frequency-modulation (FM) broadcast band and is considered part of the very high frequency (VHF) radio spectrum, a good form for an inductor would be
    (a) air-core.
    (b) pot core.
    (c) either (a) or (b).
    (d) neither (a) nor (b).

20. A transmission-line inductor made from coaxial cable having velocity factor of 0.66 and working at 450 MHz, which is in the ultrahigh frequency (UHF) radio spectrum, should, in order to measure less than ¼ electrical wavelength, be cut shorter than
    (a) 16.7 m.
    (b) 11 m.
    (c) 16.7 cm.
    (d) 11 cm.

11
CHAPTER
Capacitance

ELECTRICAL COMPONENTS CAN OPPOSE THE FLOW OF AC IN THREE WAYS, TWO OF WHICH YOU'VE learned about. *Resistance* slows the flow of ac or dc charge carriers (usually electrons) by brute force. *Inductance* impedes the flow of ac charge carriers by temporarily storing the energy as a magnetic field. *Capacitance*, about which you'll learn in this chapter, impedes the flow of ac charge carriers by temporarily storing the energy as an *electric field.*

The Property of Capacitance

Imagine two huge, flat sheets of metal that are excellent electrical conductors. Suppose they are each the size of the state of Nebraska, and are placed one over the other, separated by only 1 foot of space. If these two sheets of metal are connected to the terminals of a battery, as shown in Fig. 11-1, they will become charged electrically, one positively and the other negatively.

If the plates were small, they would both become charged almost instantly, attaining a relative voltage equal to the voltage of the battery. But because the plates are gigantic, it will take a little time for the negative plate to reach full negative potential, and an equal time for the other plate to reach full positive potential. Eventually, the voltage between the two plates will equal the battery voltage,

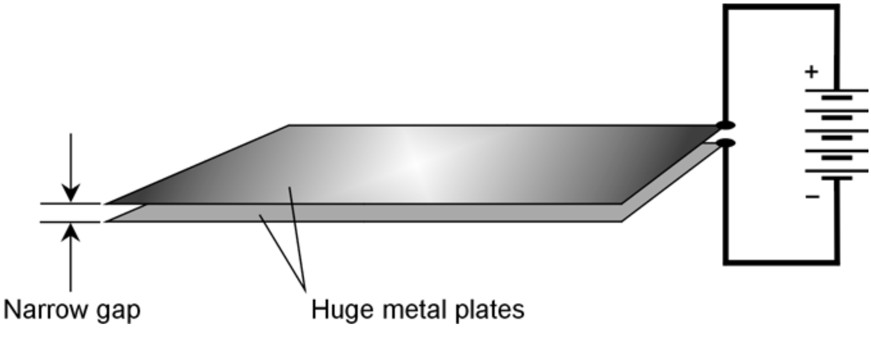

11-1 A hypothetical gigantic capacitor.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

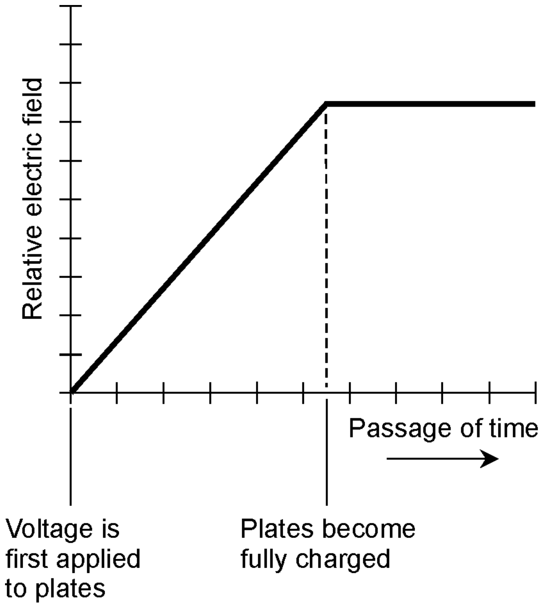

11-2 Relative electric field intensity between metal plates connected to a voltage source, as a function of time.

and an electric field will exist in the space between the plates. This electric field will be small at first, because the plates don't charge up right away. But the charge will increase over a period of time, depending on how large the plates are, and also depending on how far apart they are. Figure 11-2 is a relative graph showing the intensity of the electric field between the plates as a function of time, elapsed from the instant the plates are connected to the battery terminals.

Energy will be stored in this electric field. The ability of the plates, and of the space between them, to store this energy is the property of *capacitance.* As a quantity or variable, capacitance is denoted by the uppercase italic letter *C.*

Practical Capacitors

It's out of the question to make a capacitor of the preceding dimensions. But two sheets, or strips, of foil can be placed one on top of the other, separated by a thin, nonconducting sheet such as paper, and then the whole assembly can be rolled up to get a large effective surface area. When this is done, the electric flux becomes great enough so that the device exhibits significant capacitance. Alternatively, two sets of several plates each can be meshed together with air in between them, and the resulting capacitance is significant at high ac frequencies.

In a capacitor, the electric flux concentration is multiplied when a *dielectric* of a certain type is placed between the plates. This increases the effective surface area of the plates, so that a physically small component can be made to have a large capacitance. The voltage that a capacitor can handle depends on the thickness of the metal sheets or strips, on the spacing between them, and on the type of dielectric used.

In general, capacitance is directly proportional to the surface area of the conducting plates or sheets. Capacitance is *inversely proportional* to the separation between conducting sheets. In other words, the closer the sheets are to each other, the greater the capacitance. The capacitance also depends on the *dielectric constant* of the material between the plates. A vacuum has a dielectric constant of 1; some substances have dielectric constants that multiply the effective capacitance many times.

The Unit of Capacitance

When a battery is connected between the plates of a capacitor, the potential difference between the plates builds up at a rate that depends on the capacitance. The greater the capacitance, the slower the rate of change of voltage in the plates. The unit of capacitance is an expression of the ratio between the current that flows and the rate of voltage change between the plates as the plates become charged. A capacitance of 1 *farad* (1 F) represents a current flow of 1 A while there is a voltage increase of 1 V/s. A capacitance of 1 F also results in 1 V of potential difference for an electric charge of 1 C.

The farad is a huge unit of capacitance. You'll almost never see a capacitor with a value of 1 F. Commonly employed units of capacitance are the *microfarad* (μF) and the *picofarad* (pF). A capacitance of 1 μF represents 0.000001 (10^{-6}) F, and 1 pF is a millionth of a microfarad, or 0.000000000001 (10^{-12}) F.

Physically small components can be made to have fairly large capacitance values. Conversely, some capacitors with small values take up large physical volumes. The physical size of a capacitor, if all other factors are held constant, is proportional to the voltage that it can handle. The higher the rated voltage, the bigger the component.

Capacitors in Series

With capacitors, there is rarely any mutual interaction. This makes capacitors easier to work with than inductors. We don't have to worry about *mutual capacitance* very often, the way we have to be concerned about mutual inductance when working with wire coils.

Capacitors in series add together like resistors or inductors in parallel. Suppose you have several capacitors with values C_1, C_2, C_3, . . . , C_n connected in series. You can find the reciprocal of the total capacitance, $1/C$, using the following formula:

$$1/C = 1/C_1 + 1/C_2 + 1/C_3 + \dots + 1/C_n$$

The net capacitance of the series combination, C, is found by taking the reciprocal of the number you get for $1/C$.

If two or more capacitors are connected in series, and one of them has a value that is tiny compared with the values of all the others, the net capacitance is roughly equal to the smallest capacitance.

Problem 11-1

Suppose two capacitors, with values of $C_1 = 0.10$ μF and $C_2 = 0.050$ μF, are connected in series (Fig. 11-3). What is the net capacitance?

Using the preceding formula, first find the reciprocals of the values. They are $1/C_1 = 10$ and $1/C_2 = 20$. Then $1/C = 10 + 20 = 30$, and $C = 1/30 = 0.033$ μF. Note that we can work with reciprocal capacitances in this calculation only because the values of the components are specified in the same units.

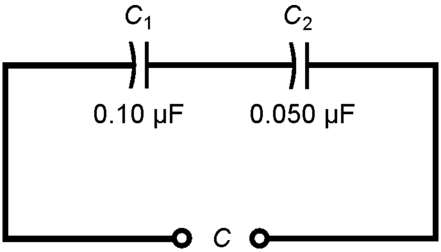

11-3 Capacitors in series. Illustration for Problem 11-1.

Problem 11-2

Suppose two capacitors with values of 0.0010 μF and 100 pF are connected in series. What is the net capacitance?

In this case, you must convert to the same size units before doing any calculations. A value of 100 pF represents 0.000100 μF. Thus, C_1 = 0.0010 μF and C_2 = 0.000100 μF. The reciprocals are $1/C_1 = 1000$ and $1/C_2 = 10{,}000$. Therefore, $1/C = 1000 + 10{,}000 = 11{,}000$, so $C = 1/11{,}000 = 0.000091$ μF. (You might rather say it's 91 pF.)

Problem 11-3

Suppose five capacitors, each of 100 pF, are in series. What is the total capacitance?

If there are n capacitors in series, all of the same value so that $C_1 = C_2 = C_3 = \ldots = C_n$, the net capacitance C is equal to $1/n$ of the capacitance of any of the components alone. Because there are five 100-pF capacitors here, the total is $C = 100/5 = 20.0$ pF.

Capacitors in Parallel

Capacitances in parallel add like resistances in series. The total capacitance is the sum of the individual component values. If two or more capacitors are connected in parallel, and one of the capacitances is far larger than any of the others, the total capacitance can be taken as approximately the value of the biggest one.

Problem 11-4

Suppose three capacitors are in parallel, having values of C_1 = 0.100 μF, C_2 = 0.0100 μF, and C_3 = 0.001000 μF, as shown in Fig. 11-4. What is the total capacitance?

Add them up: $C = 0.100 + 0.0100 + 0.001000 = 0.111000$. Because two of the values are given to only three significant figures, the final answer should be stated as $C = 0.111$ μF.

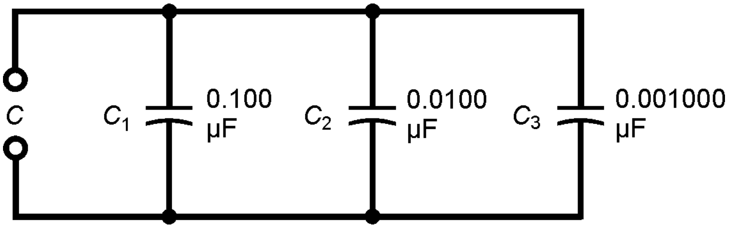

11-4 Capacitors in parallel. Illustration for Problem 11-4.

Problem 11-5

Suppose two capacitors are in parallel, one with a value of 100 μF and one with a value of 100 pF. What is the net capacitance?

In this case, you can say right away that the net capacitance is 100 μF for practical purposes. The 100-pF capacitor has a value that is only one-millionth of the capacitance of the 100-μF component. The smaller capacitance contributes essentially nothing to the net capacitance of this combination.

Fixed Capacitors

A *fixed capacitor* has a value that cannot be adjusted, and that (ideally) does not vary when environmental or circuit conditions change. Here are some of the characteristics, and common types, of fixed capacitors.

Dielectric Materials

Just as certain solids can be placed within a coil to increase the inductance, materials exist that can be sandwiched in between the plates of a capacitor to increase the capacitance. The substance between the plates is called the *dielectric* of the capacitor. Air is an efficient dielectric; it has almost no loss. But it is difficult to get very much capacitance using air as the dielectric. Some kind of solid material is usually employed as the dielectric for most fixed capacitors.

Dielectric materials accommodate electric *fields* well, but they are poor conductors of electric *currents.* In fact, dielectric materials are known as good insulators. Solid dielectrics increase the capacitance for a given surface area and spacing of the plates. Solid dielectrics also allow the plates to be rolled up, squashed, and placed very close together (Fig. 11-5). This geometry acts to maximize the capacitance per unit volume.

Paper Capacitors

In the early days of electronics, capacitors were commonly made by placing paper, soaked with mineral oil, between two strips of foil, rolling the assembly up, attaching wire leads to the two pieces of foil, and enclosing the rolled-up foil and paper in an airtight cylindrical case. *Paper capacitors* can still sometimes be found in older electronic equipment. They have values ranging from about 0.001 μF to 0.1 μF, and can handle low to moderate voltages, usually up to about 1000 V.

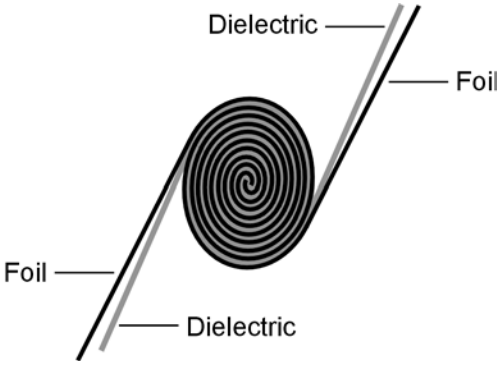

11-5 A cross-sectional drawing of a capacitor consisting of two foil sheets rolled up, and two sheets of dielectric material rolled up between them.

Mica Capacitors

Mica is a naturally occurring, solid, transparent mineral substance that flakes off in thin sheets. It makes an excellent dielectric for capacitors. *Mica capacitors* can be manufactured by alternately stacking metal sheets and layers of mica, or by applying silver ink to sheets of mica. The metal sheets are wired together into two meshed sets, forming the two terminals of the capacitor. This scheme is shown in Fig. 11-6.

Mica capacitors have low loss, and are therefore highly efficient, provided their voltage rating is not exceeded. Voltage ratings can be up to several thousand volts if thick sheets of mica are used. But mica capacitors are large physically in proportion to their capacitance. The main application for mica capacitors is in radio receivers and transmitters. Their capacitances are a little lower than those of paper capacitors, ranging from a few tens of picofarads up to about 0.05 μF.

Ceramic Capacitors

Ceramic materials work well as dielectrics. Sheets of metal are stacked alternately with wafers of ceramic to make these capacitors. The meshing/layering geometry of Fig. 11-6 is used. Ceramic, like mica, has low loss and allows for high efficiency.

For small values of capacitance, only one layer of ceramic is needed, and two metal plates can be glued to the disk-shaped material, one on each side. This type of component is known as a *disk-ceramic* capacitor. Alternatively, a tube or cylinder of ceramic can be employed, and metal ink applied to the inside and outside of the tube. Such units are called *tubular* capacitors. Ceramic capacitors have values ranging from a few picofarads to about 0.5 μF. Their voltage ratings are comparable to those of paper capacitors.

Plastic-Film Capacitors

Plastics make good dielectrics for the manufacture of capacitors. *Polyethylene* and *polystyrene* are commonly used. The method of manufacture is similar to that for paper capacitors. Stacking methods can be used if the plastic is rigid. The geometries can vary, and these capacitors are therefore found in various shapes.

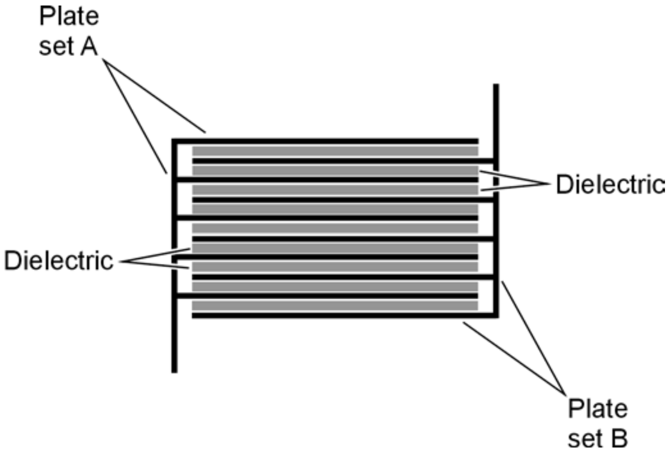

11-6 A cross-sectional drawing of a capacitor consisting of two meshed sets of several metal plates, separated by layers of dielectric material.

Capacitance values for plastic-film units range from about 50 pF to several tens of microfarads. Most often they are in the range of 0.001 μF to 10 μF. Plastic capacitors are employed at AF and RF, and at low to moderate voltages. The efficiency is good, although not as high as that for mica-dielectric or air-dielectric units.

Electrolytic Capacitors

All of the aforementioned types of capacitors provide relatively small values of capacitance. They are also *nonpolarized*, meaning that they can be hooked up in a circuit in either direction. An *electrolytic* capacitor provides greater capacitance than any of the preceding types, but it must be connected in the proper direction in a circuit to work right. An electrolytic capacitor is a *polarized* component.

Electrolytic capacitors are made by rolling up aluminum foil strips, separated by paper saturated with an *electrolyte* liquid. The electrolyte is a conducting solution. When dc flows through the component, the aluminum oxidizes because of the electrolyte. The oxide layer is nonconducting, and forms the dielectric for the capacitor. The layer is extremely thin, and this results in a high capacitance per unit volume. Electrolytic capacitors can have values up to thousands of microfarads, and some can handle thousands of volts. These capacitors are most often seen in AF circuits and in dc power supplies.

Tantalum Capacitors

Another type of electrolytic capacitor uses tantalum rather than aluminum. The tantalum can be foil, as is the aluminum in a conventional electrolytic capacitor. It can also take the form of a porous pellet, the irregular surface of which provides a large area in a small volume. An extremely thin oxide layer forms on the tantalum.

Tantalum capacitors have high reliability and excellent efficiency. They are often used in military applications because they almost never fail. They can be used in AF and digital circuits in place of aluminum electrolytics.

Semiconductor Capacitors

Later in this book, you'll learn about *semiconductors.* These materials have revolutionized electrical and electronic circuit design in the past several decades.

Semiconductor materials can be employed to make capacitors. A semiconductor *diode* conducts current in one direction, and refuses to conduct in the other direction. When a voltage source is connected across a diode so that it does not conduct, the diode acts as a capacitor. The capacitance varies depending on how much of this *reverse voltage* is applied to the diode. The greater the reverse voltage, the smaller the capacitance. This makes the diode act as a *variable capacitor.* Some diodes are especially manufactured to serve this function. Their capacitances fluctuate rapidly along with pulsating dc. They are called *varactor diodes* or simply *varactors.*

Capacitors can be formed in the semiconductor materials of an integrated circuit (also called an *IC* or *chip*) in much the same way. Sometimes, IC diodes are fabricated to serve as varactors. Another way to make a capacitor in an IC is to sandwich an oxide layer into the semiconductor material, between two layers that conduct well. Most ICs look like little boxes with protruding metal prongs (Fig. 11-7). The prongs provide the electrical connections to external circuits and systems.

Semiconductor capacitors usually have small values of capacitance. They are physically tiny, and can handle only low voltages. The advantages are miniaturization and an ability, in the case of the varactor, to change in value at a rapid rate.

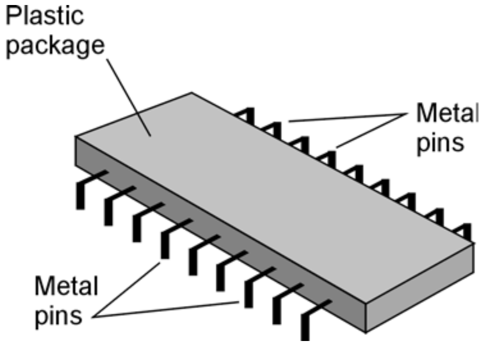

11-7 A typical integrated-circuit package is a tiny plastic box with protruding metal pins.

Variable Capacitors

The capacitance of a component can be varied at will by adjusting the mutual surface area between the plates, or by changing the spacing between the plates. The two most common types of variable capacitors (besides varactors) are the *air variable* and the *trimmer.* You will also sometimes encounter *coaxial capacitors.*

Air Variables

By connecting two sets of metal plates so that they mesh, and by affixing one set to a rotatable shaft, a variable capacitor is made. The rotatable set of plates is called the *rotor*, and the fixed set is called the *stator.* This is the type of component you might have seen in older radio receivers, used to tune the frequency. Such capacitors are still used in transmitter output tuning networks. Figure 11-8 is a functional drawing of an air-variable capacitor.

Air variables have maximum capacitance that depends on the number of plates in each set, and also on the spacing between the plates. Common maximum values are 50 to 500 pF; minimum values are a few picofarads. The voltage-handling capability depends on the spacing between the plates. Some air variables can handle many kilovolts.

Air variables are used primarily in RF applications. They are highly efficient, and are nonpolarized, although the rotor is usually connected to common ground (the chassis or circuit board).

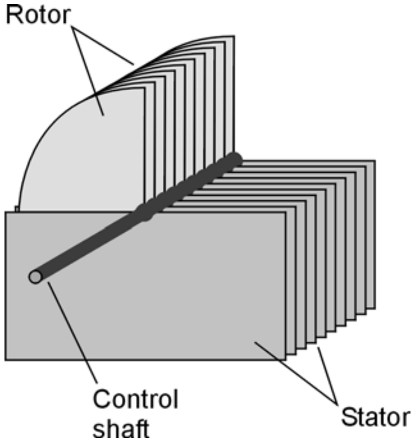

11-8 A simplified drawing of an air-variable capacitor.

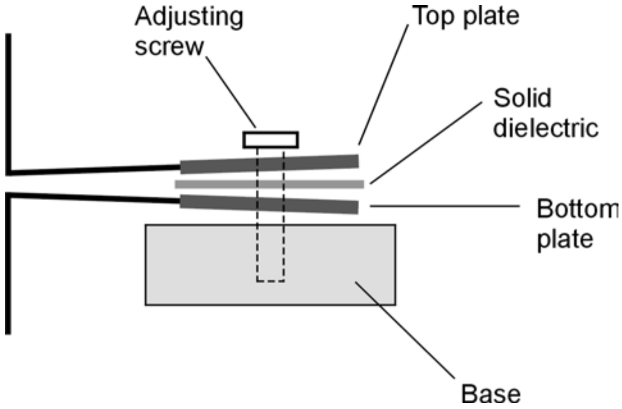

11-9 A cross-sectional drawing of a trimmer capacitor.

Trimmer Capacitors

When it is not necessary to change the value of a capacitor very often, a trimmer can be used. It consists of two plates, mounted on a ceramic base and separated by a sheet of plastic, mica, or some other solid dielectric. The plates are flexible, and can be squashed together more or less by means of a screw (Fig. 11-9). Sometimes two sets of several plates are interleaved to increase the capacitance.

Trimmers can be connected in parallel with an air variable, so that the range of the air variable can be adjusted. Some air-variable capacitors have trimmers built in. Typical maximum values for trimmers range from a few picofarads up to about 200 pF. They handle low to moderate voltages, are highly efficient, and are nonpolarized.

Coaxial Capacitors

You recall from the previous chapter that sections of transmission lines can work as inductors. They can act as capacitors, too. If a section of transmission line is less than $\frac{1}{4}$ wavelength long, and is left open at the far end (rather than shorted out), it behaves as a capacitor. The capacitance increases with length.

The most common transmission-line capacitor uses two telescoping sections of metal tubing. This is called a *coaxial capacitor.* It works because there is a certain effective surface area between the inner and the outer tubing sections. A sleeve of plastic dielectric is placed between the sections of tubing, as shown in Fig. 11-10. This allows the capacitance to be adjusted by sliding the inner section in or out of the outer section.

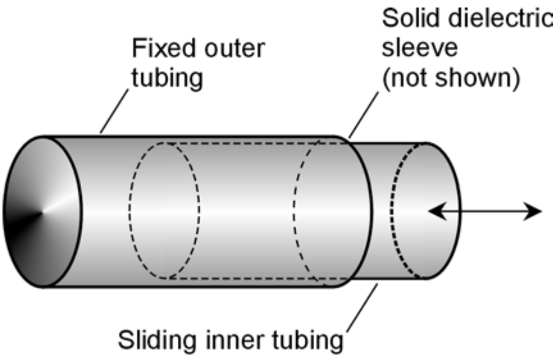

11-10 A simplified drawing of a coaxial variable capacitor.

Coaxial capacitors are used in RF applications, particularly in antenna systems. Their values are generally from a few picofarads up to about 100 pF.

Capacitor Specifications

When you are looking for a capacitor for a particular application, it's important to find a component that has the right specifications for the job. Here are two of the most important specifications to watch for.

Tolerance

Capacitors are rated according to how nearly their values can be expected to match the rated capacitance. The most common tolerance is ±10%; some capacitors are rated at ±5% or even at ±1%.

The lower (or *tighter*) the tolerance number, the more closely you can expect the actual component value to match the rated value. For example, a ±10% capacitor rated at 100 pF can range from 90 to 110 pF. But if the tolerance is ±1%, the manufacturer guarantees that the capacitance will be between 99 and 101 pF.

Problem 11-6

A capacitor is rated at 0.10 μF ±10%. What is its guaranteed range of capacitance?

First, multiply 0.10 by 10 percent to get the plus-or-minus variation. This is $0.10 \times 0.10 = 0.010$ μF. Then add and subtract this from the rated value to get the maximum and minimum possible capacitances. The result is a range of 0.09 to 0.11 μF.

Temperature Coefficient

Some capacitors increase in value as the temperature increases. These components have a *positive temperature coefficient.* Some capacitors decrease in value as the temperature rises; these have a *negative temperature coefficient.* Some capacitors are manufactured so that their values remain constant over a certain temperature range. Within this span of temperatures, such capacitors have *zero temperature coefficient.*

The temperature coefficient is specified in percent per degree Celsius (%/°C). Sometimes, a capacitor with a negative temperature coefficient can be connected in series or parallel with a capacitor having a positive temperature coefficient, and the two opposite effects cancel out over a range of temperatures. In other instances, a capacitor with a positive or negative temperature coefficient can be used to cancel out the effect of temperature on other components in a circuit, such as inductors and resistors.

Interelectrode Capacitance

Any two pieces of conducting material, when they are brought near each other, can act as a capacitor. Often, this *interelectrode capacitance* is so small that it can be neglected. It rarely amounts to more than a few picofarads. In utility circuits and at AF, interelectrode capacitance is not usually significant. But it can cause problems at RF. The chances for trouble increase as the frequency increases. The most common phenomena are *feedback*, and/or a change in the frequency characteristics of a circuit.

Interelectrode capacitance can be minimized by keeping wire leads as short as possible, by using shielded cables, and by enclosing sensitive circuits in metal housings.

Quiz

Refer to the text in this chapter if necessary. A good score is 18 correct. Answers are in the back of the book.

1. Capacitance acts to store electrical energy as
   (a) current.
   (b) voltage.
   (c) a magnetic field.
   (d) an electric field.
2. As capacitor plate area increases, all other things being equal,
   (a) the capacitance increases.
   (b) the capacitance decreases.
   (c) the capacitance does not change.
   (d) the current-handling ability decreases.
3. As the spacing between plates in a capacitor is made smaller, all other things being equal,
   (a) the capacitance increases.
   (b) the capacitance decreases.
   (c) the capacitance does not change.
   (d) the resistance increases.
4. A material with a high dielectric constant
   (a) acts to increase capacitance per unit volume.
   (b) acts to decrease capacitance per unit volume.
   (c) has no effect on capacitance.
   (d) causes a capacitor to become polarized.
5. A capacitance of 100 pF is the same as which of the following?
   (a) 0.01 μF
   (b) 0.001 μF
   (c) 0.0001 μF
   (d) 0.00001 μF
6. A capacitance of 0.033 μF is the same as which of the following?
   (a) 33 pF
   (b) 330 pF
   (c) 3300 pF
   (d) 33,000 pF
7. If five 0.050-μF capacitors are connected in parallel, what is the net capacitance of the combination?
   (a) 0.010 μF
   (b) 0.25 μF

(c) 0.50 μF
(d) 0.025 μF

8. If five 0.050-μF capacitors are connected in series, what is the net capacitance of the combination?
(a) 0.010 μF
(b) 0.25 μF
(c) 0.50 μF
(d) 0.025 μF

9. Suppose that two capacitors are connected in series, and their values are 47 pF and 33 pF. What is the net capacitance of this combination?
(a) 80 pF
(b) 47 pF
(c) 33 pF
(d) 19 pF

10. Suppose that two capacitors are in parallel. Their values are 47.0 pF and 470 μF. What is the net capacitance of this combination?
(a) 47.0 pF
(b) 517 pF
(c) 517 μF
(d) 470 μF

11. Suppose that three capacitors are in parallel. Their values are 0.0200 μF, 0.0500 μF, and 0.10000 μF. What is the net capacitance of this combination?
(a) 0.0125 μF
(b) 0.1700 μF
(c) 0.1000 μF
(d) 0.1250 μF

12. The main advantage of air as a dielectric material for capacitors is the fact that it
(a) has a high dielectric constant.
(b) is not physically dense.
(c) has low loss.
(d) allows for large capacitance in a small volume.

13. Which of the following is *not* a characteristic of mica capacitors?
(a) Excellent efficiency
(b) Small size, even for large values of capacitance
(c) High voltage-handling capacity
(d) Low loss

14. Which of the following capacitance values is most typical of a disk-ceramic capacitor?
    (a) 100 pF
    (b) 33 μF
    (c) 470 μF
    (d) 10,000 μF

15. Which of the following capacitance values is most typical of a paper capacitor?
    (a) 0.001 pF
    (b) 0.01 μF
    (c) 100 μF
    (d) 3300 μF

16. Which of the following capacitance ranges is most typical of an air-variable capacitor?
    (a) 0.01 μF to 1 μF
    (b) 1 μF to 100 μF
    (c) 1 pF to 100 pF
    (d) 0.001 pF to 0.1 pF

17. Which of the following types of capacitors is polarized?
    (a) Paper
    (b) Mica
    (c) Interelectrode
    (d) Electrolytic

18. If a capacitor has a negative temperature coefficient, then
    (a) its capacitance decreases as the temperature rises.
    (b) its capacitance increases as the temperature rises.
    (c) its capacitance does not change with temperature.
    (d) it will not work if the temperature is below freezing.

19. Suppose that a capacitor is rated at 33 pF ±10%. Which of the following actual capacitance values is outside the acceptable range?
    (a) 30 pF
    (b) 37 pF
    (c) 35 pF
    (d) 31 pF

20. Suppose that a capacitor, rated at 330 pF, shows an actual value of 317 pF. By how many percent does its actual capacitance differ from its rated capacitance?
    (a) –0.039%
    (b) –3.9%
    (c) –0.041%
    (d) –4.1%

12
CHAPTER
Phase

IN ALTERNATING CURRENT, EACH 360° CYCLE IS EXACTLY THE SAME AS EVERY OTHER. IN EVERY CYCLE, the waveform of the previous cycle is repeated. In this chapter, you'll learn about the most common type of ac waveform: the *sine wave.*

Instantaneous Values

An ac sine wave has a characteristic shape, as shown in Fig. 12-1. This is the way the graph of the function $y = \sin x$ looks on an (x,y) coordinate plane. (The abbreviation *sin* stands for *sine* in trigonometry.) Suppose that the peak voltage is ±1 V, as shown. Further imagine that the period is 1 s, so the frequency is 1 Hz. Let the wave begin at time $t = 0$. Then each cycle begins every time the value of t is a whole number. At every such instant, the voltage is zero and *positive-going.*

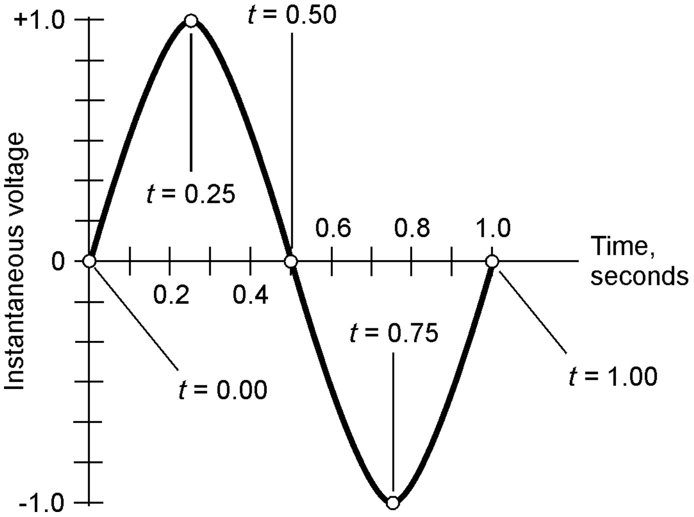

12-1 A sine wave with a period of 1 second. It thus has a frequency of 1 Hz.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

If you freeze time at, say, $t = 446.00$, the voltage is zero. Looking at the diagram, you can see that the voltage will also be zero every so-many-and-a-*half* seconds, so it will be zero at $t = 446.5$. But instead of getting more positive at these instants, the voltage will be *negative-going.*

If you freeze time at so-many-and-a-*quarter* seconds, say $t = 446.25$, the voltage will be +1 V. The wave will be exactly at its positive peak. If you stop time at so-many-and-*three-quarter* seconds, say $t = 446.75$, the voltage will be exactly at its negative peak, −1 V. At intermediate times, say, so-many-and-*three-tenths* seconds, the voltage will have intermediate values.

Instantaneous Rate of Change

Figure 12-1 shows that there are times the voltage is increasing, and times it is decreasing. *Increasing,* in this context, means "getting more positive," and *decreasing* means "getting more negative." The most rapid increase in voltage occurs when $t = 0.0$ and $t = 1.0$. The most rapid decrease takes place when $t = 0.5$.

When $t = 0.25$, and also when $t = 0.75$, the instantaneous voltage neither increases nor decreases. But this condition exists only for a vanishingly small moment, a single point in time.

Suppose n is some whole number. Then the situation at $t = n.25$ is the same as it is for $t = 0.25$; also, for $t = n.75$, things are the same as they are when $t = 0.75$. The single cycle shown in Fig. 12-1 represents every possible condition of the ac sine wave having a frequency of 1 Hz and a peak value of ±1 V. The whole wave recurs, over and over, for as long as the ac continues to flow in the circuit.

Now imagine that you want to observe the *instantaneous rate of change* in the voltage of the wave in Fig. 12-1, as a function of time. A graph of this turns out to be a sine wave, too—but it is displaced to the left of the original wave by $\frac{1}{4}$ of a cycle. If you plot the instantaneous rate of change of a sine wave against time (Fig. 12-2), you get the *derivative* of the waveform. The derivative of a sine wave is a *cosine wave.* This wave has the same shape as the sine wave, but the *phase* is different by $\frac{1}{4}$ of a cycle.

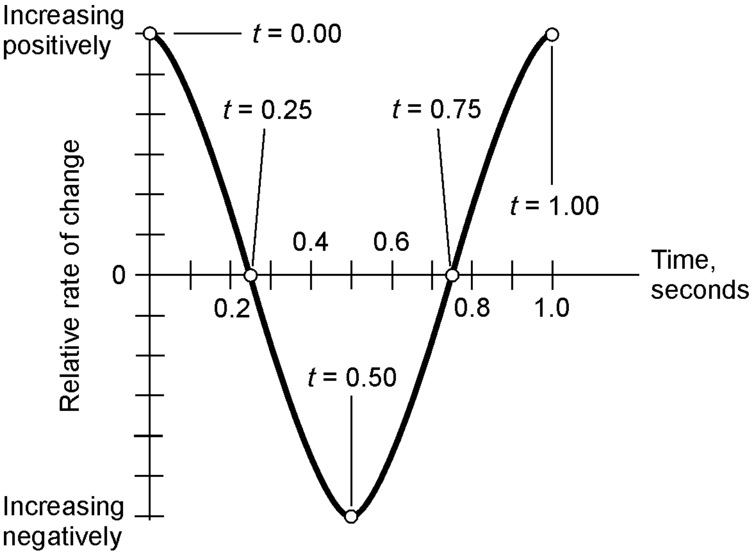

12-2 A sine wave representing the rate of change in the instantaneous voltage of the wave shown in Fig. 12-1.

Circles and Vectors

An ac sine wave represents the most efficient possible way that an electrical quantity can alternate. It has only one frequency component. All the wave energy is concentrated into this smoothly seesawing variation. It is like a pure musical note.

Circular Motion

Suppose that you swing a ball around and around at the end of a string, at a rate of one revolution per second (1 rps). The ball describes a circle in space (Fig. 12-3A). If a friend stands some distance away, with his or her eyes in the plane of the ball's path, your friend sees the ball oscillating back and forth (Fig. 12-3B) with a frequency of 1 Hz. That is one complete cycle per second, because you swing the ball around at 1 rps.

If you graph the position of the ball, as seen by your friend, with respect to time, the result is a sine wave (Fig. 12-4). This wave has the same fundamental shape as all sine waves. Some sine waves are taller than others, and some are stretched out horizontally more than others. But the general waveform is the same in every case. By multiplying or dividing the amplitude and the wavelength of any sine wave, it can be made to fit exactly along the curve of any other sine wave. The standard sine wave is the function $y = \sin x$ in the coordinate plane.

You might whirl the ball around faster or slower than 1 rps. The string might be made longer or shorter. This would alter the height and/or the frequency of the sine wave graphed in Fig. 12-4. But the sine wave can always be reduced to the equivalent of constant, smooth motion in a circular orbit. This is known as the *circular motion model* of a sine wave.

Rotating Vectors

Back in Chapter 9, *degrees of phase* were discussed. If you wondered then why phase is spoken of in terms of angular measure, the reason should be clearer now. A circle has 360°. A sine wave can be represented as circular motion. Points along a sine wave thus correspond to angles, or positions, around a circle.

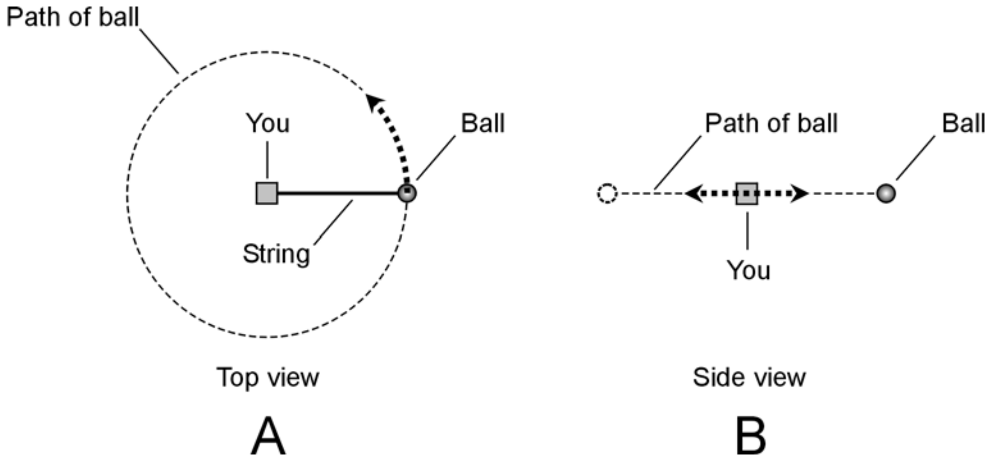

12-3 Swinging ball and string as seen from above (A) and from the side (B).

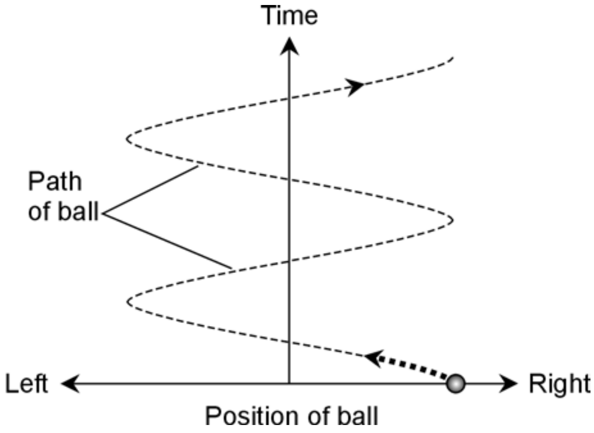

12-4 Position of ball (horizontal axis) as seen from the side, graphed as a function of time (vertical axis).

Figure 12-5 shows the way a rotating *vector* can be used to represent a sine wave. A vector is a quantity with two independent properties, called *magnitude* (or amplitude) and *direction.* At A, the vector points east, and this is assigned the value of 0°, where the wave amplitude is zero and is increasing positively. At B, the vector points north; this is the 90° instant, where the wave has attained its maximum positive amplitude. At C, the vector points west. This is 180°, the instant where the wave has gone back to zero amplitude and is getting more negative. At D, the wave points south. This is 270°, and it represents the maximum negative amplitude. When a full circle (360°) has been completed, the vector once again points east.

The four points in Fig. 12-5 are shown on a sine wave graph in Fig. 12-6. Think of the vector as revolving counterclockwise at a rate that corresponds to one revolution per cycle of the wave. If the wave has a frequency of 1 Hz, the vector goes around at a rate of 1 rps. If the wave has a frequency of

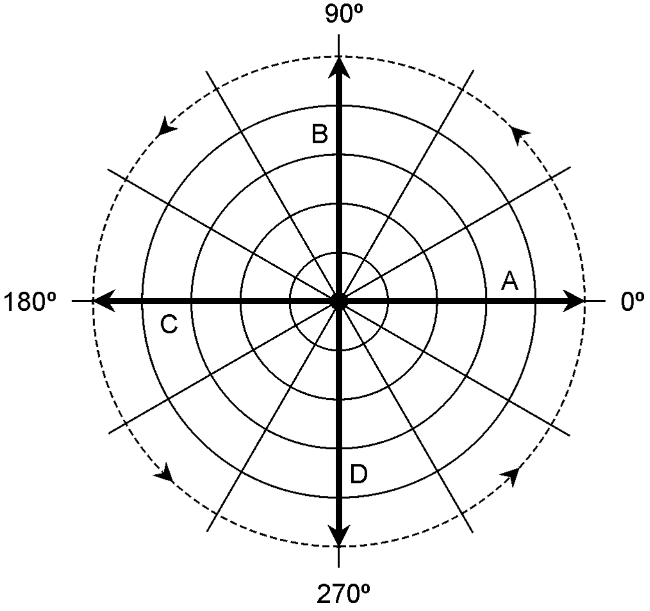

12-5 Rotating-vector representation of a sine wave. At A, at the start of the cycle; at B, one-fourth of the way through the cycle; at C, halfway through the cycle; at D, three-fourths of the way through the cycle.

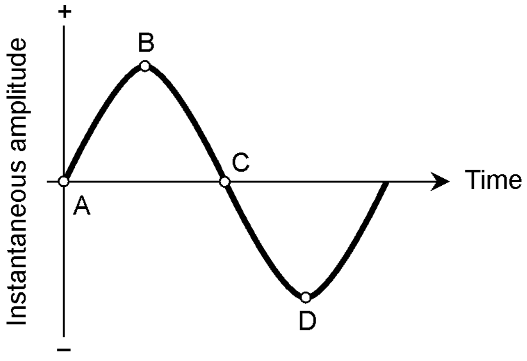

12-6 The four points for the vector model of Fig. 12-5, shown in the standard amplitude-versus-time graphical manner.

100 Hz, the speed of the vector is 100 rps, or a revolution every 0.01 s. If the wave is 1 MHz, then the speed of the vector is 1 million rps (10^6 rps), and it goes once around every 0.000001 s (10^{-6} s).

The peak amplitude of a pure ac sine wave corresponds to the length of its vector. In Fig. 12-5, time is shown by the angle counterclockwise from due east. Amplitude is independent of time. The vector length never changes, but its direction does.

Expressions of Phase Difference

The *phase difference*, also called the *phase angle*, between two waves can have meaning only when those two waves have identical frequencies. If the frequencies differ, even by just a little bit, the relative phase constantly changes, and it's impossible to specify a value for it. In the following discussions of phase angle, let's assume that the two waves always have identical frequencies.

Phase Coincidence

Phase coincidence means that two waves begin at exactly the same moment. They are "lined up." This is shown in Fig. 12-7 for two waves having different amplitudes. The phase difference in this

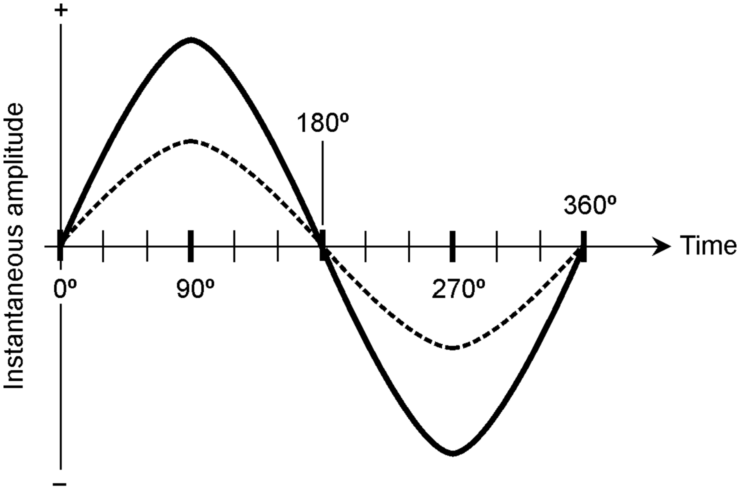

12-7 Two sine waves in phase coincidence.

case is 0°. You could say it's some whole-number multiple of 360°, too—but engineers and technicians rarely speak of any phase angle of less than 0° or more than 360°.

If two sine waves are in phase coincidence, and if neither wave has dc superimposed, then the resultant is a sine wave with positive or negative peak amplitudes equal to the sum of the positive and negative peak amplitudes of the composite waves. The phase of the resultant is the same as that of the composite waves.

Phase Opposition

When two sine waves begin exactly ½ cycle, or 180°, apart, they are said to be in *phase opposition.* This is illustrated by the drawing of Fig. 12-8. In this situation, engineers sometimes say that the waves are *out of phase*, although this expression is a little nebulous because it could be taken to mean some phase difference other than 180°.

If two sine waves have the same amplitudes and are in phase opposition, they cancel each other out. This is because the instantaneous amplitudes of the two waves are equal and opposite at every moment in time.

If two sine waves are in phase opposition, and if neither wave has dc superimposed, then the resultant is a sine wave with positive or negative peak amplitudes equal to the difference between the positive and negative peak amplitudes of the composite waves. The phase of the resultant is the same as the phase of the stronger of the two composite waves.

Any sine wave without superimposed dc has the unique property that, if its phase is shifted by 180°, the resultant wave is the same as turning the original wave upside down. Not all waveforms have this property. Perfect square waves do, but some rectangular and sawtooth waves don't, and irregular waveforms almost never do.

Intermediate Phase Differences

Two sine waves can differ in phase by any amount from 0° (phase coincidence), through 90° (*phase quadrature*, meaning a difference a quarter of a cycle), 180° (phase opposition), 270° (phase quadrature again), to 360° (phase coincidence again).

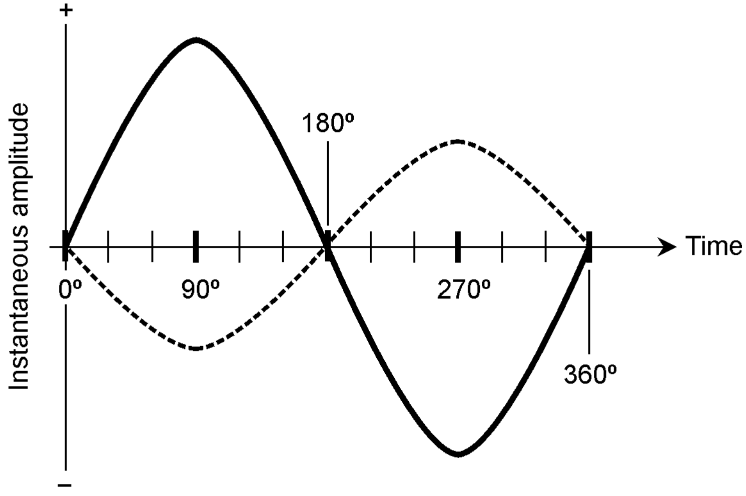

12-8 Two sine waves in phase opposition.

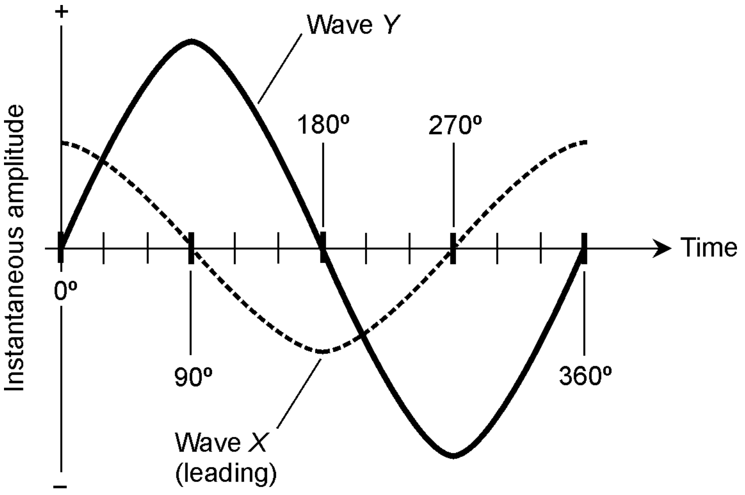
+
Wave *Y*
180°
270°
Instantaneous amplitude
Time
0°
90°
360°
Wave *X* (leading)
–

12-9 Wave *X* leads wave *Y* by 90° of phase (¼ of a cycle).

Leading Phase

Imagine two sine waves, called wave *X* and wave *Y*, with identical frequency. If wave *X* begins a fraction of a cycle *earlier* than wave *Y*, then wave *X* is said to be *leading* wave *Y* in phase. For this to be true, *X* must begin its cycle less than 180° before *Y*. Figure 12-9 shows wave *X* leading wave *Y* by 90°.

Note that if wave *X* (the dashed line in Fig. 12-9) is leading wave *Y* (the solid line), then wave *X* is displaced to the *left* of wave *Y*. In a time-domain graph or display, displacement to the left represents earlier moments in time, and displacement to the right represents later moments in time.

Lagging Phase

Suppose that some sine wave *X* begins its cycle more than 180°, but less than 360°, ahead of wave *Y*. In this situation, it is easier to imagine that wave *X* starts its cycle *later* than wave *Y*, by some value between 0° and 180°. Then wave *X* is not leading, but *lagging*, wave *Y*. Figure 12-10 shows wave *X* lagging wave *Y* by 90°.

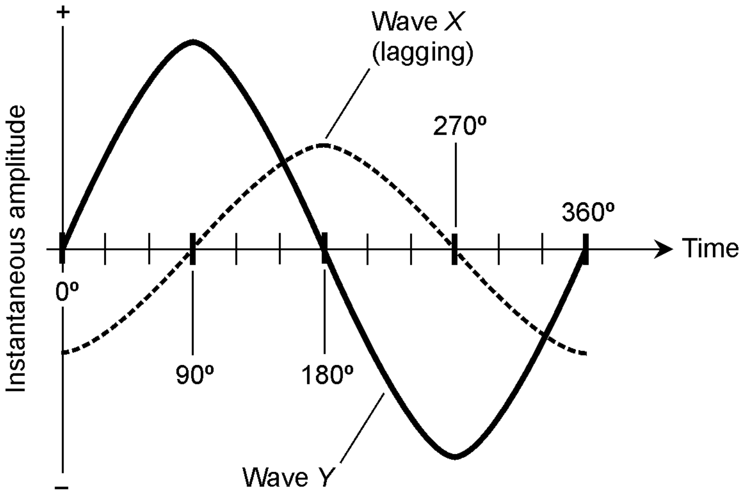
+
Wave *X* (lagging)
270°
360°
Instantaneous amplitude
Time
0°
90°
180°
Wave *Y*
–

12-10 Wave *X* lags wave *Y* by 90° of phase (¼ of a cycle).

Vector Diagrams of Phase Difference

The vector renditions of sine waves, such as are shown in Fig. 12-5, are well suited to showing phase relationships.

If a sine wave *X* leads a sine wave *Y* by some number of degrees, then the two waves can be drawn as vectors, with vector **X** being that number of angular degrees *counterclockwise* from vector **Y.** If a sine wave *X* lags a sine wave *Y* by some number of degrees, then **X** appears to point in a direction that is *clockwise* from **Y** by that number of angular degrees. If two waves are in phase coincidence, then their vectors point in exactly the same direction. If two waves are in phase opposition, then their vectors point in exactly opposite directions.

The drawings of Fig. 12-11 show four phase relationships between two sine waves *X* and *Y.* At A, *X* is in phase with *Y.* At B, *X* leads *Y* by 90°. At C, *X* and *Y* are 180° apart in phase. At D, *X* lags *Y* by 90°. In all of these examples, think of the vectors rotating *counterclockwise* as time passes, but always maintaining the same angle with respect to each other, and always staying at the same lengths. If the frequency in hertz is *f*, then the pair of vectors rotates together, counterclockwise, at an angular speed of *f*, expressed in complete 360° revolutions per second.

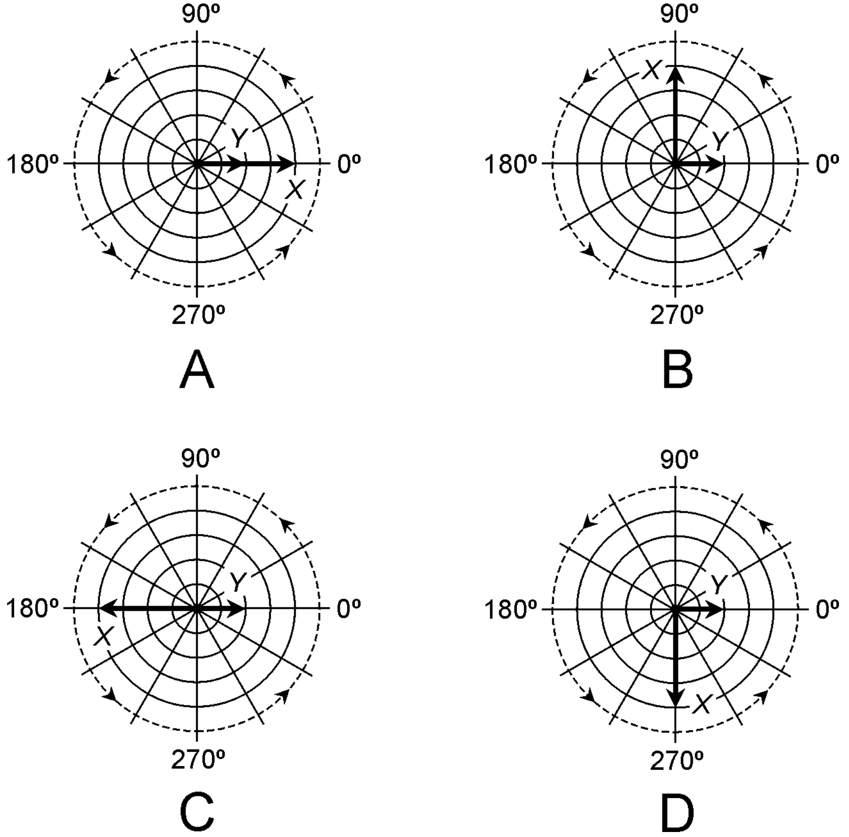

12-11 Vector representations of phase difference. At A, waves *X* and *Y* are in phase. At B, *X* leads *Y* by 90°. At C, *X* and *Y* are 180° out of phase. At D, *X* lags *Y* by 90°. Time is represented by counterclockwise motion of both vectors at a constant angular speed.

Quiz

Refer to the text in this chapter if necessary. A good score is 18 correct. Answers are in the back of the book.

1. Which of the following is *not* a general characteristic of an ac wave?
   (a) The wave shape is identical for each cycle.
   (b) The polarity reverses periodically.
   (c) The electrons always flow in the same direction.
   (d) There is a definite frequency.

2. All sine waves
   (a) have similar general appearance.
   (b) have instantaneous rise and fall times.
   (c) are in the same phase as cosine waves.
   (d) rise instantly, but decay slowly.

3. The derivative of a sine wave
   (a) is shifted in phase by ½ cycle from the sine wave.
   (b) is the rate of change in the instantaneous value.
   (c) has instantaneous rise and decay times.
   (d) rises instantly, but decays slowly.

4. A phase difference of 180° in the circular motion model of a sine wave represents
   (a) ¼ revolution.
   (b) ½ revolution.
   (c) a full revolution.
   (d) two full revolutions.

5. You can add or subtract a certain number of degrees of phase to or from a wave, and end up with exactly the same wave again. This number is
   (a) 90, or any whole-number multiple of it.
   (b) 180, or any whole-number multiple of it.
   (c) 270, or any whole-number multiple of it.
   (d) 360, or any whole-number multiple of it.

6. You can add or subtract a certain number of degrees of phase to or from a sine wave, and end up with an inverted (upside-down) representation of the original. This number is
   (a) 90, or any odd whole-number multiple of it.
   (b) 180, or any odd whole-number multiple of it.
   (c) 270, or any odd whole-number multiple of it.
   (d) 360, or any odd whole-number multiple of it.

7. Suppose a wave has a frequency of 300 kHz. How long does one complete cycle take?
   (a) 1.300 s
   (b) 0.00333 s
   (c) 1/3000 s
   (d) 3.33×10^{-6} s

8. If a wave has a frequency of 440 Hz, how long does it take for 10° of a cycle to occur?
   (a) 0.00273 s
   (b) 0.000273 s
   (c) 0.0000631 s
   (d) 0.00000631 s

9. Suppose two waves are in phase coincidence. One has peak values of ±3 V and the other has peak values of ±5 V. The resultant has voltages of
   (a) ±8 V pk, in phase with the composites.
   (b) ±2 V pk, in phase with the composites.
   (c) ±8 V pk, in phase opposition with respect to the composites.
   (d) ±2 V pk, in phase opposition with respect to the composites.

10. As shown on a graph, shifting the phase of an ac sine wave by 90° is the same thing as
   (a) moving it to the right or left by a full cycle.
   (b) moving it to the right or left by $\frac{1}{4}$ cycle.
   (c) turning it upside down.
   (d) leaving it alone.

11. Two pure sine waves that differ in phase by 180° can be considered to
   (a) be offset by two full cycles.
   (b) be in phase opposition.
   (c) be separated by less than $\frac{1}{4}$ cycle.
   (d) have a frequency of $\frac{1}{2}$ cycle.

12. Suppose two sine waves are in phase opposition. Wave X has a peak amplitude of ±4 V and wave Y has a peak amplitude of ±8 V. The resultant has voltages of
   (a) ±4 V pk, in phase with the composites.
   (b) ±4 V pk, out of phase with the composites.
   (c) ±4 V pk, in phase with wave X.
   (d) ±4 V pk, in phase with wave Y.

13. If wave X leads wave Y by 45°, then
   (a) wave Y is $\frac{1}{4}$ cycle ahead of wave X.
   (b) wave Y is $\frac{1}{4}$ cycle behind wave X.
   (c) wave Y is $\frac{1}{8}$ cycle behind wave X.
   (d) wave Y is 1.16 cycle ahead of wave X.

14. If wave *X* lags wave *Y* by ⅓ cycle, then
    (a) wave *Y* is 120° ahead of wave *X*.
    (b) wave *Y* is 90° ahead of wave *X*.
    (c) wave *Y* is 60° ahead of wave *X*.
    (d) wave *Y* is 30° ahead of wave *X*.
15. Refer to Fig. 12-12. In this example,
    (a) *X* lags *Y* by 45°.
    (b) *X* leads *Y* by 45°.
    (c) *X* lags *Y* by 135°.
    (d) *X* leads *Y* by 135°.
16. Which of the drawings in Fig. 12-13 represents the situation of Fig. 12-12?
    (a) Drawing A
    (b) Drawing B
    (c) Drawing C
    (d) Drawing D
17. In vector diagrams such as those of Fig. 12-13, the length of the vector represents
    (a) the average amplitude of a sine wave.
    (b) the frequency of a sine wave.
    (c) the phase of a sine wave.
    (d) the peak amplitude of a sine wave.
18. In vector diagrams such as those of Fig. 12-13, the angle between two vectors represents
    (a) the average of the peak amplitudes of two sine waves.
    (b) the frequency difference between two sine waves.
    (c) the phase difference between two sine waves.
    (d) the difference between the peak amplitudes of two sine waves.

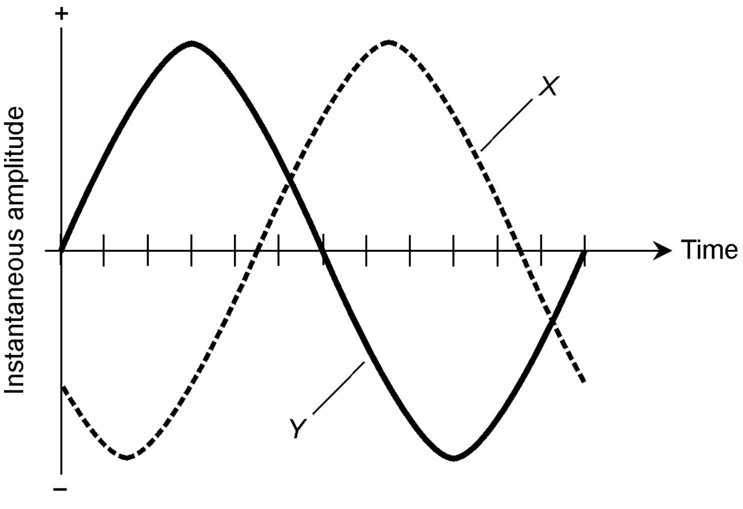

12-12 Illustration for Quiz Question 15.

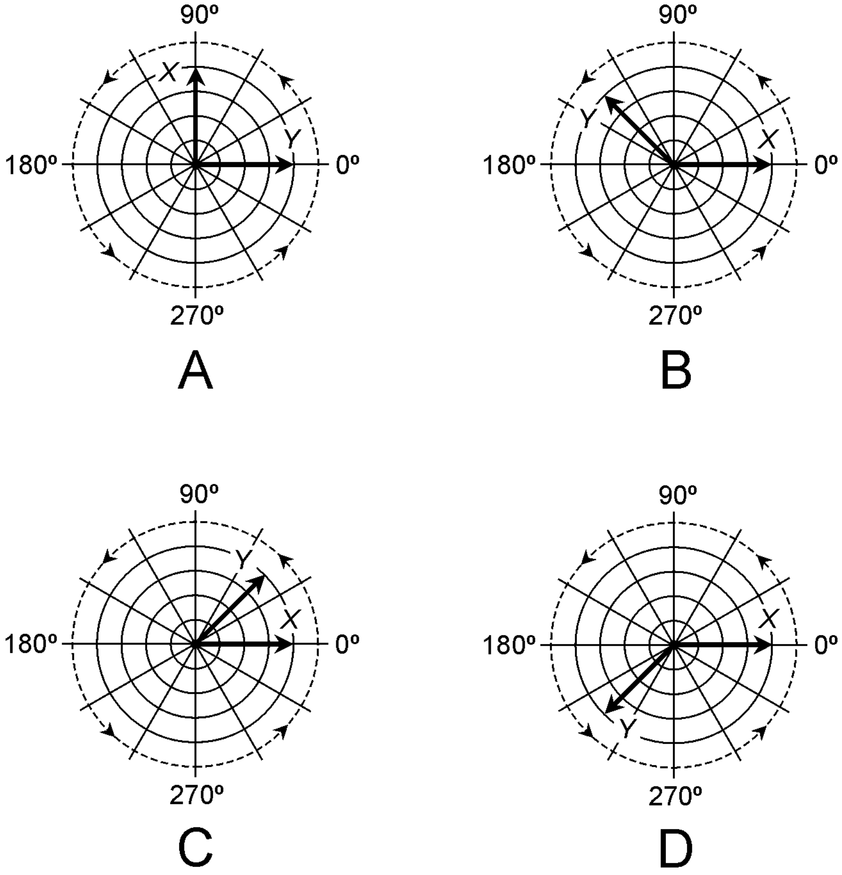

12-13 Illustration for Quiz Questions 16 through 20.

19. In vector diagrams such as those of Fig. 12-13, the distance from the center of the graph represents
    - (a) average amplitude.
    - (b) frequency.
    - (c) phase.
    - (d) peak amplitude.

20. In diagrams like those of Fig. 12-13, the progression of time is sometimes depicted as
    - (a) movement of a vector to the right.
    - (b) movement of a vector to the left.
    - (c) counterclockwise rotation of a vector.
    - (d) clockwise rotation of a vector.

13
CHAPTER

Inductive Reactance

IN DC CIRCUITS, RESISTANCE CAN BE EXPRESSED AS A NUMBER RANGING FROM ZERO (REPRESENTING a perfect conductor) to extremely large values. Physicists call resistance a *scalar* quantity, because it can be expressed on a one-dimensional *scale*, as shown in Fig. 13-1.

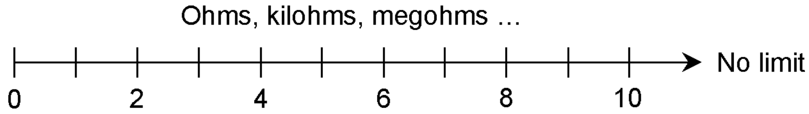

13-1 Resistance can be represented as numerical values (corresponding to ohms) along a half line or ray.

Coils and Direct Current

Suppose you have some wire that conducts electricity very well. If you wind a length of the wire into a coil and connect it to a source of dc (Fig. 13-2), the wire draws a large current. It doesn't matter

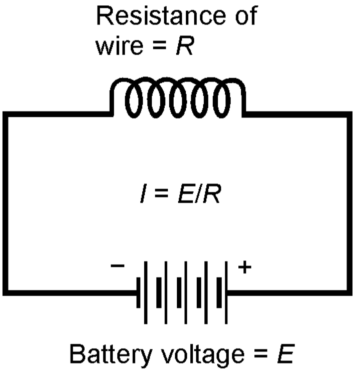

13-2 An inductor connected across a source of dc.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

whether the wire is a single-turn loop, or whether it's lying haphazardly on the floor, or whether it's wrapped around a stick. The current amperes is equal to the applied voltage in volts divided by the wire resistance in ohms. It's that simple.

You can make an electromagnet, as you've already seen, by passing dc through a coil wound around an iron rod. Electromagnets are known for the high current they draw from batteries or power supplies. The coil of an electromagnet heats up as energy is dissipated in the resistance of the wire. If the voltage of the battery or power supply increases, the wire in the coil gets hotter. Ultimately, if the supply can deliver enough current, the wire will melt.

Coils and Alternating Current

Suppose you change the voltage source, connected across the coil, from dc to ac (Fig. 13-3). Imagine that you can vary the frequency of the ac, from a few hertz to hundreds of hertz, then kilohertz, then megahertz.

At first, the current will be high, just as it is with dc. But the coil has a certain amount of inductance, and it takes some time for current to establish itself in the coil. Depending on how many turns there are and on whether the core is air or a ferromagnetic material, you'll reach a point, as the ac frequency increases, when the coil starts to get sluggish. That is, the current won't have time to get established in the coil before the polarity of the ac voltage reverses. At high ac frequencies, the current through the coil will have difficulty following the voltage placed across the coil. This sluggishness in a coil for ac is, in effect, similar to dc resistance. As the frequency is raised, the effect gets more pronounced. Eventually, if you keep increasing the frequency of the ac source, the coil will not even come near establishing a current with each cycle. Then the coil will act like a high resistance.

The opposition that the coil offers to ac is called *inductive reactance.* It, like resistance, is measured in ohms. It can vary, just as resistance does, from near zero (a short piece of wire) to a few ohms (a small coil) to kilohms or megohms (bigger and bigger coils). Like resistance, inductive reactance affects the current in an ac circuit. But, unlike simple resistance, reactance changes with frequency. This effect is not merely a decrease in the current, although in practice this does happen. Inductive reactance produces a change in the way the current flows with respect to the voltage.

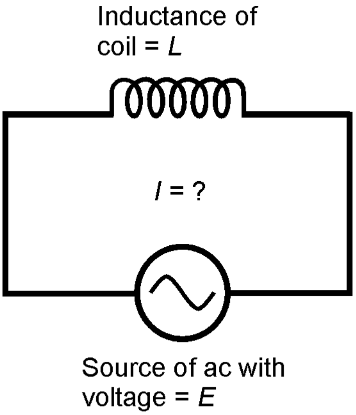

13-3 An inductor connected across a source of ac.

Reactance and Frequency

Inductive reactance is one of two form of reactance. (The other form, called *capacitive reactance*, will be discussed in the next chapter.) Reactance in general is symbolized by the italic uppercase letter X. Inductive reactance is symbolized X_L.

If the frequency of an ac source is given, in hertz, as f, and the inductance of a coil in henrys is given as L, then the inductive reactance in ohms, X_L, is calculated as follows:

$$X_L = 2\pi f L$$

In this formula, the symbol π stands for the mathematical constant *pi*, which is the number of diameters around the circumference of a circle. It is equal to approximately 3.14. We can consider the value of 2π to be equal to 6.28 in most practical situations. Therefore, the preceding formula can be written a little more simply as:

$$X_L = 6.28 f L$$

This same formula applies if the frequency, f, is in kilohertz and the inductance, L, is in millihenrys. And it also applies if f is in megahertz and L is in microhenrys. Just remember that if frequency is in thousands, inductance must be in thousandths, and if frequency is in millions, inductance must be in millionths.

Inductive reactance increases *linearly* with increasing ac frequency. This means that the function of X_L versus f is a straight line when graphed. Inductive reactance also increases linearly with inductance. Therefore, the function of X_L versus L also appears as a straight line on a graph. The value of X_L is *directly proportional* to f, and is also directly proportional to L. These relationships are graphed, in relative form, in Fig. 13-4.

Problem 13-1

Suppose a coil has an inductance of 0.500 H, and the frequency of the ac passing through it is 60.0 Hz. What is the inductive reactance?

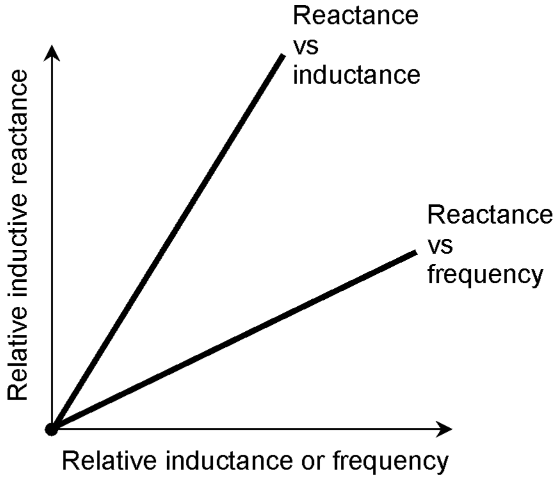

13-4 Inductive reactance is directly proportional to inductance, and is also directly proportional to frequency.

Using the preceding formula, calculate $X_L = 6.28 \times 60.0 \times 0.500 = 188\ \Omega$. This is rounded to three significant figures.

Problem 13-2

What will be the inductive reactance of the preceding coil if the supply is a battery that supplies pure dc?

Because dc has a frequency of zero, $X_L = 6.28 \times 0 \times 0.500 = 0\ \Omega$. That is, there will be no inductive reactance. Inductance doesn't have any practical effect with pure dc.

Problem 13-3

If a coil has an inductive reactance of 100 Ω at a frequency of 5.00 MHz, what is its inductance?

In this case, you need to plug numbers into the formula and solve for the unknown L. Start out with the equation $100 = 6.28 \times 5.00 \times L = 31.4 \times L$. Because the frequency is given in megahertz, the inductance will come out in microhenrys. You can divide both sides of the equation by 31.4, getting $L = 100/31.4 = 3.18\ \mu\text{H}$.

Points in the *RL* Plane

Inductive reactance can be plotted along a half line, just as can resistance. In a circuit containing both resistance and inductance, the characteristics become two-dimensional. You can orient the resistance and reactance half lines perpendicular to each other to make a quarter-plane coordinate system, as shown in Fig. 13-5. Resistance is plotted horizontally, and inductive reactance is plotted vertically upward.

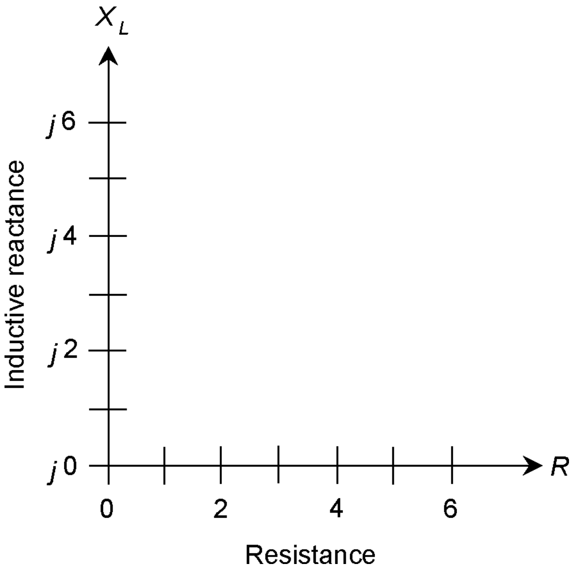

13-5 The quarter plane for inductive reactance (X_L) and resistance (R). This is also known as the *RL* quarter-plane, or simply as the *RL* plane.

In this scheme, resistance-inductance (*RL*) combinations form *complex impedances.* (The term *impedance* comes from the root *impede*, and fully describes how electrical components impede, or inhibit, the flow of ac. You'll learn all about this in Chap. 15.) Each point on the *RL plane* corresponds to one unique complex impedance value. Conversely, each complex impedance value corresponds to one unique point on the *RL* plane.

You might ask, "What's the little j doing in Fig. 13-5?" For reasons that will be made clear in Chap. 15, impedances on the *RL* plane are written in the form $R + jX_L$, where R is the resistance in ohms, and X_L is the inductive reactance in ohms. The little j is called a *j operator* and is a mathematical way of expressing the fact that reactance is denoted at right angles to resistance in complex-impedance graphs.

If you have a pure resistance, say $R = 5\ \Omega$, then the *complex* impedance is $5 + j0$, and is at the point (5,0) on the *RL* plane. If you have a pure inductive reactance, such as $X_L = 3\ \Omega$, then the complex impedance is $0 + j3$, and is at the point $(0,j3)$ on the *RL* plane. These points, and a couple of others, are shown in Fig. 13-6.

In real life, all coils have some resistance, because no wire is a perfect conductor. All resistors have at least a tiny bit of inductive reactance, because they take up some physical space and they have wire leads. So there is really no such thing as a mathematically perfect pure resistance such as $5 + j0$, or a mathematically perfect pure reactance like $0 + j3$. But sometimes you can get extremely close to theoretical ideals in real life.

Often, resistance and inductive reactance are both deliberately placed in a circuit. Then you get impedances values such as $2 + j3$ or $4 + j1.5$. These are shown in Fig. 13-6 as points on the *RL* plane.

Remember that values for X_L are *reactances*, not actual inductances. Because of this, they vary with the frequency in an *RL* circuit. Changing the frequency has the effect of making complex impedance points move around in the *RL* plane. They move vertically, going upward as the ac frequency increases, and downward as the ac frequency decreases. If the ac frequency goes down to zero, the inductive reactance vanishes. Then $X_L = 0$, we have pure dc, and the point is right on the resistance axis.

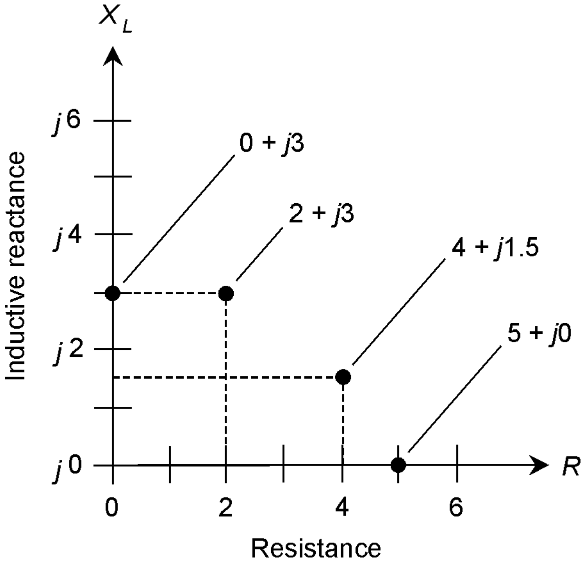

13-6 Four points in the *RL* plane.

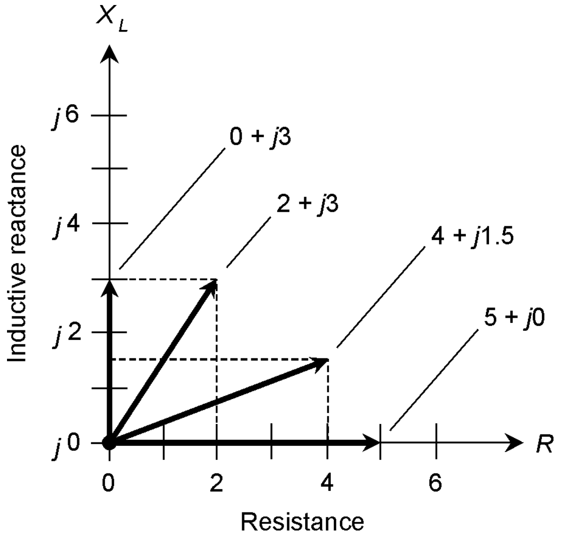

13-7 Four vectors in the *RL* plane, corresponding to the points shown in Fig. 13-6.

Vectors in the *RL* Plane

Engineers sometimes represent points in the *RL* plane as vectors. Recall that a vector is a mathematical quantity that has a defined magnitude (length) and defined direction (orientation). Expressing a point in the *RL* plane as a vector thus gives that point a unique magnitude and a unique direction.

In Fig. 13-6, four different points are shown. Each point is represented by a certain distance to the right of the *origin* (0,*j*0), and a certain distance upward from the origin. The first of these is the resistance, *R*, and the second is the inductive reactance, X_L. Thus, the *RL* combination is a two-dimensional quantity. There is no way to uniquely define *RL* combinations as single numbers, or *scalars*, because there are two different quantities that can vary independently.

Another way to depict these points is to draw lines from the origin out to them. Then you can think of the points as *rays*, each having a certain length, or magnitude, and a certain direction, or angle counterclockwise from the resistance axis. These rays, going out to the points, are *complex impedance vectors* (Fig. 13-7).

Current Lags Voltage

When an ac voltage is placed across an inductor and starts to increase (either positively or negatively) from zero, it takes a fraction of a cycle for the current to follow. Once the voltage starts decreasing from its maximum peak (either positive or negative) in the cycle, it again takes a fraction of a cycle for the current to follow. The instantaneous current can't quite keep up with the instantaneous voltage, as it does in a pure resistance. Thus, in a circuit containing inductive reactance, the current is said to *lag* the voltage in phase.

Pure Inductance

Suppose that you place an ac voltage across a coil, with a frequency high enough so that the inductive reactance, X_L, is much larger than the resistance, *R*. In this situation, the current is $\frac{1}{4}$ of a cycle behind the voltage. That is, the current lags the voltage by 90°, as shown in Fig. 13-8.

At very low frequencies, large inductances are normally needed in order for the current lag to be

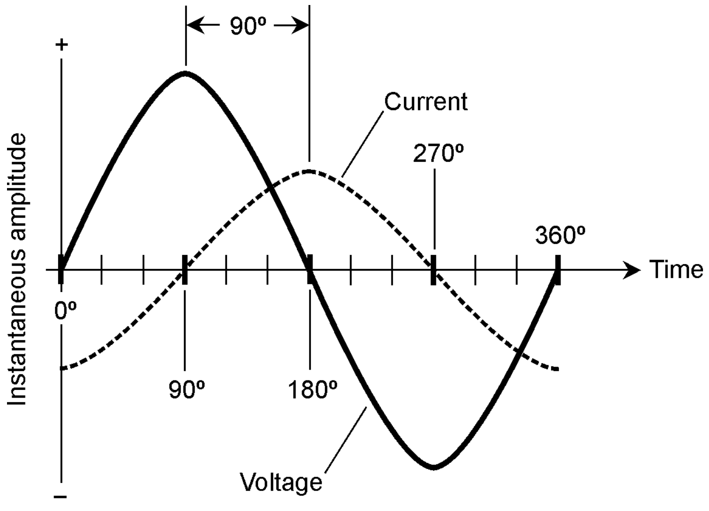

13-8 In a pure inductive reactance, the current lags the voltage by 90°.

a full 90°. This is because any coil has some resistance; no wire is a perfect conductor. If some wire were found that had a mathematically zero resistance, and if a coil of any size were wound from this wire, then the current would lag the voltage by 90° in this inductor, no matter what the ac frequency.

When the value of X_L is very large compared with the value of R in a circuit—that is, when there is an essentially pure inductive reactance—the vector in the RL plane points straight up along the X_L axis. Its angle is 90° from the R axis, which is considered the *zero line* in the RL plane.

Inductance with Resistance

When the resistance in a resistance-inductance (RL) circuit is significant compared with the inductive reactance, the current lags the voltage by something less than 90° (Fig. 13-9). If R is small compared with X_L, the current lag is almost 90°, but as R gets larger relative to X_L, the lag decreases.

The value of R in an RL circuit can increase relative to X_L because resistance is deliberately placed in series with the inductance. It can also happen because the ac frequency gets so low that X_L

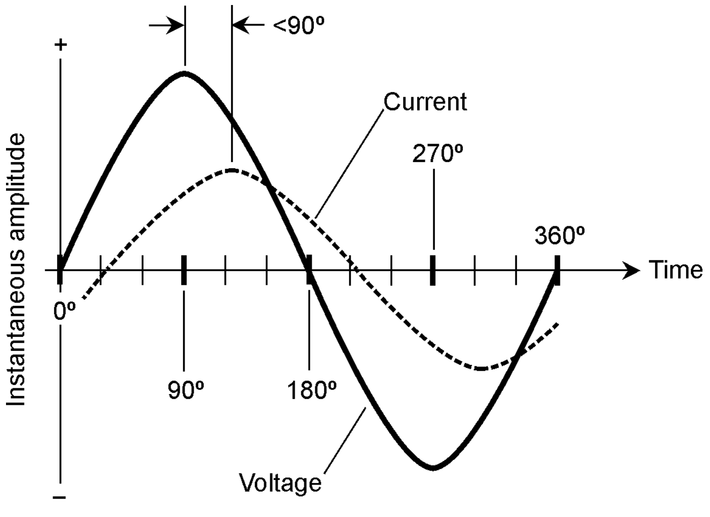

13-9 In a circuit with inductive reactance and resistance, the current lags the voltage by less than 90°.

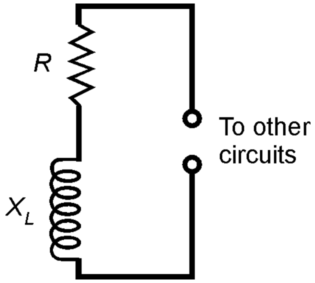

13-10 Schematic representation of a circuit containing resistance and inductive reactance.

decreases until it is comparable to the loss resistance R in the coil winding. In either case, the situation can be schematically represented by an inductance in series with a resistance (Fig. 13-10).

If you know the values of X_L and R, you can find the *angle of lag*, also called the *RL phase angle*, by plotting the point $R + jX_L$ on the RL plane, drawing the vector from the origin out to that point, and then measuring the angle of the vector, counterclockwise from the resistance axis. You can use a protractor to measure this angle, or you can compute its value using trigonometry.

Actually, you don't have to know the actual values of X_L and R in order to find the angle of lag. All you need to know is their ratio. For example, if $X_L = 5\ \Omega$ and $R = 3\ \Omega$, you get the same RL phase angle that you get if $X_L = 50\ \Omega$ and $R = 30\ \Omega$, or if $X_L = 20\ \Omega$ and $R = 12\ \Omega$. The angle of lag is the same for any values of X_L and R in the ratio 5:3.

Pure Resistance

As the resistance in an RL circuit becomes large with respect to the inductive reactance, the angle of lag gets small. The same thing happens if the inductive reactance gets small compared with the resistance. When R is many times greater than X_L, the vector in the RL plane lies almost on the R axis, going east (to the right). The RL phase angle in this case is close to 0°. The current is nearly in phase with the voltage.

In a pure resistance, with no inductance at all, the current is precisely in phase with the voltage (Fig. 13-11). A pure resistance doesn't store and release energy as an inductive circuit does, so there is no sluggishness in it.

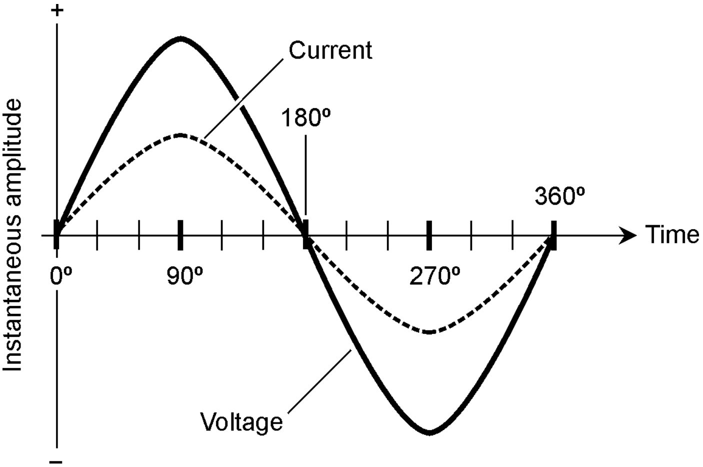

13-11 In a circuit with pure resistance (no reactance), the current is in phase with the voltage.

How Much Lag?

If you know the ratio of the inductive reactance to the resistance (X_L/R) in an *RL* circuit, then you can find the phase angle. Of course, you can also find the phase angle if you know the actual values of X_L and R.

Pictorial Method

It isn't necessary to construct an entire *RL* plane to find phase angles. You can use a ruler that has centimeter (cm) and millimeter (mm) markings, and a protractor. First, draw a line a little more than 10 cm long, going from left to right on a sheet of paper. Use the ruler and a sharp pencil. Then, with the protractor, construct a line off the left end of this first line, going vertically upward. Make this line at least 10 cm long. The horizontal line, or the one going to the right, is the R axis of a coordinate system. The vertical line, or the one going upward, is the X_L axis.

If you know the values of X_L and R, divide them down or multiply them up so they're both between 0 and 100. For example, if $X_L = 680\ \Omega$ and $R = 840\ \Omega$, you can divide them both by 10 to get $X_L = 68$ and $R = 84$. Plot these points lightly by making hash marks on the vertical and horizontal lines you've drawn. The R mark in this example will be 84 mm to the right of the origin, and the X_L mark will be 68 mm up from the origin.

Next, draw a line connecting the two hash marks, as shown in Fig. 13-12. This line will run at a slant, and will form a triangle along with the two axes. Your hash marks, and the origin of the coordinate system, form the three *vertices* of a *right triangle.* The triangle is called *right* because one of its angles is a right angle (90°). Measure the angle between the slanted line and the R axis. Extend one or both of the lines if necessary in order to get a good reading on the protractor. This angle will be between 0 and 90°, and represents the phase angle in the *RL* circuit.

The complex impedance vector, $R + jX_L$, is found by constructing a rectangle using the origin and your two hash marks as three of the four vertices, and drawing new horizontal and vertical lines to complete the figure. The vector is the diagonal of this rectangle, as shown in Fig. 13-13. The

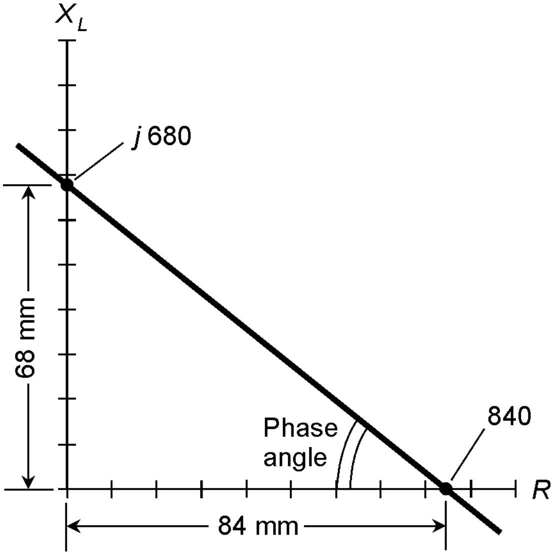

13-12 Pictorial method of finding phase angle in a circuit containing resistance and inductive reactance.

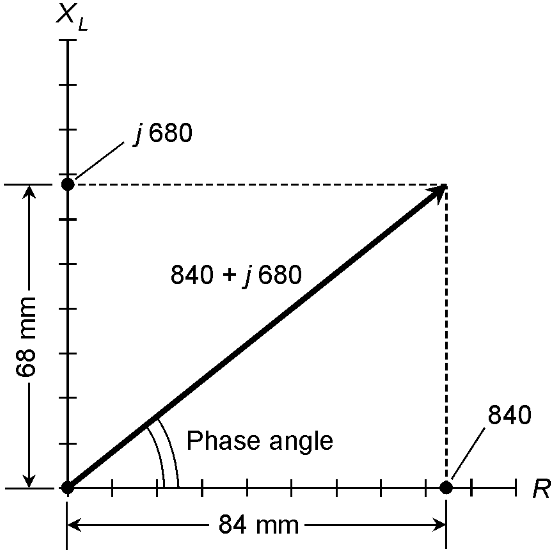

13-13 Another pictorial method of finding phase angle in a circuit containing resistance and inductive reactance. This method shows the actual impedance vector.

phase angle is the angle between this vector and the R axis. It will be the same as the angle of the slanted line in Fig. 13-12.

Trigonometric Method

If you have a good scientific calculator that can find the *arctangent* of a number (also called the *inverse tangent* and symbolized either as arctan or $\tan^{-1}$), you can determine the *RL* phase angle more precisely than the pictorial method allows. Given the values of X_L and R, the *RL* phase angle is the arctangent of their ratio. Phase angle is symbolized by the lowercase Greek letter phi (pronounced "fie" or "fee" and written ϕ). Therefore:

$$\phi = \tan^{-1}(X_L/R) \qquad \text{or} \qquad \phi = \arctan(X_L/R)$$

Problem 13-4

Suppose the inductive reactance in an *RL* circuit is 680 Ω and the resistance is 840 Ω. What is the phase angle?

The ratio X_L/R is 680/840. A calculator will display this quotient as something like 0.8095 and some more digits. Find the arctangent of this number. You should get 38.99 and some more digits. This can be rounded off to 39.0°.

Problem 13-5

Suppose an *RL* circuit operates at a frequency of 1.0 MHz with a resistance of 10 Ω and an inductance of 90 μH. What is the phase angle? What does this tell us about the nature of this *RL* circuit at this frequency?

Find the inductive reactance using the formula $X_L = 6.28fL = 6.28 \times 1.0 \times 90 = 565\ \Omega$. Then find the ratio $X_L/R = 565/10 = 56.5$. The phase angle is equal to arctan 56.5, which, rounded to two significant figures, is 89°. The circuit contains an almost pure inductive reactance, because the phase angle is close to 90°. The resistance contributes little to the behavior of this *RL* circuit at 1.0 MHz.

Problem 13-6

What is the phase angle for the preceding circuit at a frequency of 10 kHz? With that information, what can we say about the behavior of the circuit at 10 kHz?

This requires that X_L be calculated again, for the new frequency. Let's use megahertz, so it goes in the formula with microhenrys. A frequency of 10 kHz is the same as 0.010 MHz. Calculating, we get $X_L = 6.28fL = 6.28 \times 0.010 \times 90 = 5.65\ \Omega$. The ratio X_L/R is $5.65/10 = 0.565$. Therefore, the phase angle is arctan 0.565, which, rounded to two significant figures, is 29°. This is not close to either 0° or 90°. Thus, at 10 kHz, the resistance and the inductive reactance both play significant roles in the behavior of the circuit.

Quiz

Refer to the text in this chapter if necessary. A good score is 18 correct. Answers are in the back of the book.

1. As the number of turns in a coil that carries ac increases without limit, the current in the coil will
   (a) eventually become very large.
   (b) stay the same.
   (c) decrease, approaching zero.
   (d) be stored in the core material.
2. As the number of turns in a coil increases, the reactance at a constant frequency
   (a) increases.
   (b) decreases.
   (c) stays the same.
   (d) is stored in the core material.
3. As the frequency of an ac wave gets lower, the value of X_L for a particular coil of wire
   (a) increases.
   (b) decreases.
   (c) stays the same.
   (d) depends on the voltage.
4. Suppose a coil has an inductance of 100 mH. What is the reactance at a frequency of 1000 Hz?
   (a) 0.628 Ω
   (b) 6.28 Ω
   (c) 62.8 Ω
   (d) 628 Ω

5. Suppose a coil shows an inductive reactance of 200 Ω at 500 Hz. What is its inductance?
   (a) 0.637 H
   (b) 628 H
   (c) 63.7 mH
   (d) 628 mH
6. Imagine a 400-μH inductor with a reactance of 33 Ω. What is the frequency?
   (a) 13 kHz
   (b) 0.013 kHz
   (c) 83 kHz
   (d) 83 MHz
7. Suppose an inductor has $X_L = 555\ \Omega$ at $f = 132$ kHz. What is *L?*
   (a) 670 mH
   (b) 670 μH
   (c) 460 mH
   (d) 460 μH
8. Suppose a coil has $L = 689$ μH at $f = 990$ kHz. What is X_L?
   (a) 682 Ω
   (b) 4.28 Ω
   (c) 4.28 kΩ
   (d) 4.28 MΩ
9. Suppose an inductor has $L = 88$ mH with $X_L = 100\ \Omega$. What is f?
   (a) 55.3 kHz
   (b) 55.3 Hz
   (c) 181 kHz
   (d) 181 Hz
10. Each point in the *RL* plane
    (a) corresponds to a unique resistance.
    (b) corresponds to a unique inductance.
    (c) corresponds to a unique combination of resistance and inductive reactance.
    (d) corresponds to a unique combination of resistance and inductance.

11. If the resistance R and the inductive reactance X_L both are allowed to vary from zero to unlimited values, but are always in the ratio 3:1, the points in the *RL* plane for all the resulting impedances will lie along
    (a) a vector pointing straight up.
    (b) a vector pointing east.
    (c) a circle.
    (d) a ray of indefinite length, pointing outward from the origin.

12. Each specific complex impedance value defined in the form $R + jX_L$
    - (a) corresponds to a specific point in the *RL* plane.
    - (b) corresponds to a specific inductive reactance.
    - (c) corresponds to a specific resistance.
    - (d) All of the above are true.

13. A vector is defined as a mathematical quantity that has
    - (a) magnitude and direction.
    - (b) resistance and inductance.
    - (c) resistance and reactance.
    - (d) inductance and reactance.

14. In an *RL* circuit, as the ratio of inductive reactance to resistance (X_L/R) decreases, the phase angle
    - (a) increases.
    - (b) decreases.
    - (c) stays the same.
    - (d) becomes alternately positive and negative.

15. In a circuit containing inductive reactance but no resistance, the phase angle is
    - (a) constantly increasing.
    - (b) constantly decreasing.
    - (c) equal to 0°.
    - (d) equal to 90°.

16. If the inductive reactance and the resistance in an *RL* circuit are equal (as expressed in ohms), then what is the phase angle?
    - (a) 0°
    - (b) 45°
    - (c) 90°
    - (d) It depends on the actual values of the resistance and the inductive reactance.

17. In Fig. 13-14, the impedance shown is which of the following?
    - (a) 8.0 Ω
    - (b) 90 Ω
    - (c) $90 + j8.0$
    - (d) $8.0 + j90$

18. Note that in the diagram of Fig. 13-14, the R and X_L scale divisions are of different sizes. The phase angle can nevertheless be determined. It is
    - (a) about 50°, from the looks of it.
    - (b) 48°, as measured with a protractor.

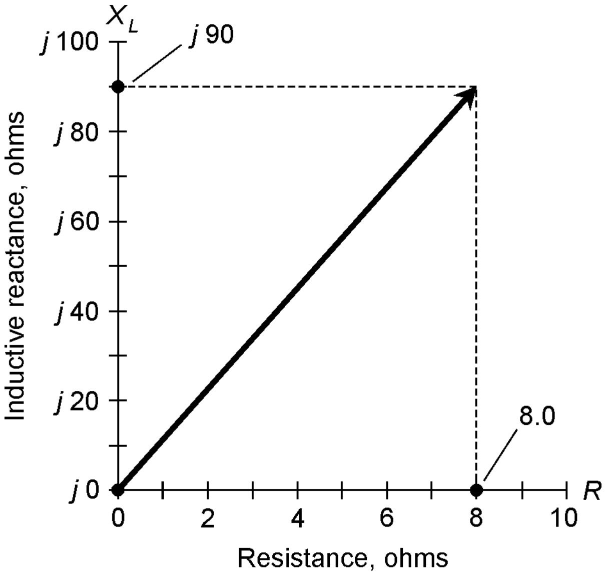

13-14 Illustration for Quiz Questions 17 and 18.

(c) 85°, as calculated using trigonometry.
(d) 6.5°, as calculated using trigonometry.

19. Consider an *RL* circuit that consists of a 100-μH inductor and a 100-Ω resistor. What is the phase angle at a frequency of 200 kHz?

(a) 45.0°
(b) 51.5°
(c) 38.5°
(d) There isn't enough data given to calculate it.

20. Suppose an *RL* circuit has an inductance of 88 mH, and the resistance is 95 Ω. At 800 Hz, what is the phase angle?

(a) 78°
(b) 12°
(c) 43°
(d) 47°

14
CHAPTER
Capacitive Reactance

CAPACITIVE REACTANCE IS THE NATURAL COUNTERPART OF INDUCTIVE REACTANCE. IT, LIKE INDUCTIVE reactance, can be represented as a ray. The capacitive-reactance ray goes in a negative direction and is assigned negative ohmic values. When the capacitive-reactance and inductive-reactance rays are joined at their endpoints (both of which correspond to a reactance of zero), a complete number line is the result, as shown in Fig. 14-1. This line depicts all possible values of reactance.

Capacitors and Direct Current

Suppose you have two big, flat metal plates, both of which are excellent electrical conductors. Imagine that you stack them one on top of the other, with only air in between. If you connect a source of dc across the plates (Fig. 14-2), the plates will become electrically charged, and will reach a potential difference equal to the dc source voltage. It won't matter how big or small the plates are; their mutual voltage will always be the same as that of the source, although, if the plates are huge, it will take awhile for them to become fully charged. Once the plates are fully charged, the current will drop to zero.

If you put some insulating material, such as glass, between the plates, their mutual voltage will not change, although the charging time will increase. If you increase the source voltage, the poten-

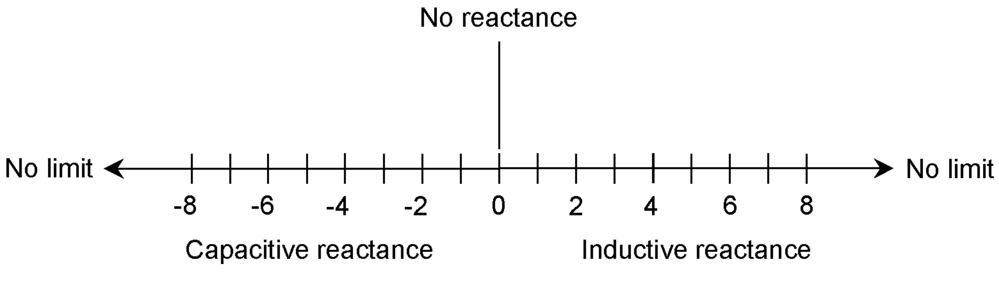

14-1 Inductive and capacitive reactance can be represented as numerical values (corresponding to ohms multiplied by *j*) along a number line.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

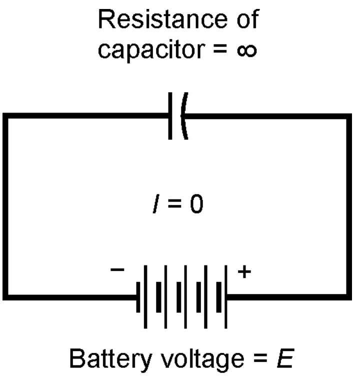

14-2 A capacitor connected across a source of dc.

tial difference between the plates will follow along, more or less rapidly, depending on how large the plates are and on what is between them. If the voltage is increased without limit, arcing will eventually take place. That is, sparks will begin to jump between the plates.

Capacitors and Alternating Current

Now, imagine that the voltage source connected across the plates is changed from dc to ac (Fig. 14-3). Imagine that you can adjust the frequency of this ac from a low value of a few hertz, to hundreds of hertz, to many kilohertz, megahertz, and gigahertz.

At first, the voltage between the plates will follow just about exactly along as the ac source polarity reverses. But the set of plates has a certain amount of capacitance. Perhaps they can charge up fast, if they are small and if the space between them is large, but they can't charge instantaneously. As you increase the frequency of the ac voltage source, there will come a point at which the plates do not get charged up very much before the source polarity reverses. The charge won't have time to get established with each ac cycle. At high ac frequencies, the voltage between the plates will have trouble following the current that is charging and discharging them. Just as the plates begin to get a good charge, the ac current will pass its peak and start to discharge them, pulling electrons out of the negative plate and pumping electrons into the positive plate.

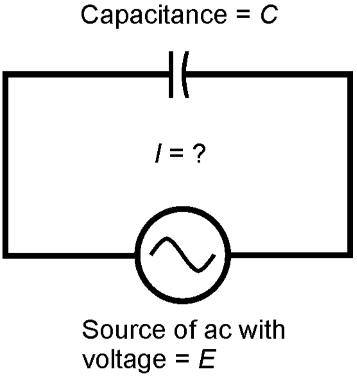

14-3 A capacitor connected across a source of ac.

As the frequency is raised without limit, the set of plates starts to act more and more like a short circuit. When the frequency is low, there is a small charging current, but this quickly drops to zero as the plates become fully charged. As the frequency becomes high, the current flows for more and more of every cycle before dropping off; the charging time remains constant while the *period* of the charging/discharging wave is getting shorter. Eventually, if you keep on increasing the frequency, the period of the wave will be much shorter than the charging/discharging time, and current will flow in and out of the plates in just about the same way as it would flow if the plates were shorted out.

The opposition that the set of plates offers to ac is the *capacitive reactance.* It is measured in ohms, just like inductive reactance, and just like resistance. But it is, by convention, assigned negative values rather than positive ones. Capacitive reactance, denoted X_C, can vary, just as resistance and inductive reactance do, from near zero (when the plates are huge and close together, and/or the frequency is very high) to a few negative ohms, to many negative kilohms or megohms.

Capacitive reactance, like inductive reactance, varies with frequency. But X_C gets larger (negatively) as the frequency goes down. This is the opposite of what happens with inductive reactance, which gets larger (positively) as the frequency goes up.

Often, capacitive reactance is talked about in terms of its *absolute value*, with the minus sign removed. Then we say that the absolute value of X_C increases as the frequency goes down, or that the absolute value of X_C is decreases as the frequency goes up.

Capacitive Reactance and Frequency

In one sense, capacitive reactance behaves like a reflection of inductive reactance. But looked at another way, X_C is an extension of X_L into negative values.

If the frequency of an ac source (in hertz) is given as f, and the capacitance (in farads) is given as C, then the capacitive reactance in ohms, X_C, is calculated as follows:

$$X_C = -1/(2\pi fC)$$

Again, we meet our friend π! And again, for most practical purposes, we can take 2π to be equal to 6.28. Thus, the preceding formula can be expressed like this:

$$X_C = -1/(6.28fC)$$

This same formula applies if the frequency, f, is in megahertz and the capacitance, C, is in microfarads.

Capacitive reactance varies *inversely* with the frequency. This means that the function X_C versus f appears as a curve when graphed, and this curve "blows up" as the frequency gets close to zero. Capacitive reactance also varies inversely with the actual value of capacitance, given a fixed frequency. Therefore, the function of X_C versus C also appears as a curve that blows up as the capacitance approaches zero.

The negative of X_C is *inversely proportional* to frequency, and also to capacitance. Relative graphs of these functions are shown in Fig. 14-4.

Problem 14-1

Suppose a capacitor has a value of 0.00100 μF at a frequency of 1.00 MHz. What is the capacitive reactance?

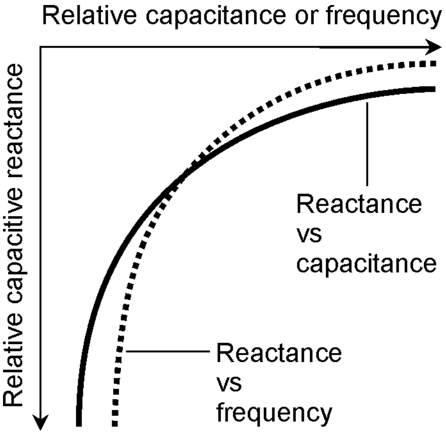

14-4 Capacitive reactance is negatively, and inversely, proportional to capacitance. Capacitive reactance is also negatively, and inversely, proportional to frequency.

Use the formula and plug in the numbers. You can do this directly, because the data is specified in microfarads (millionths) and in megahertz (millions):

$$X_C = -1/(6.28 \times 1.0 \times 0.00100) = -1/(0.00628) = -159\ \Omega$$

This is rounded to three significant figures, because all the data is given to that many digits.

Problem 14-2

What is the capacitive reactance of the preceding capacitor if the frequency decreases to zero (that is, if the voltage source is pure dc)?

In this case, if you plug the numbers into the formula, you get a zero denominator. Mathematicians will tell you that such a quantity is undefined. But we can say that the reactance is negative infinity for all practical purposes.

Problem 14-3

Suppose a capacitor has a reactance of $-100\ \Omega$ at a frequency of 10.0 MHz. What is its capacitance?

In this problem, you need to put the numbers in the formula and solve for the unknown C. Begin with this equation:

$$-100 = -1/(6.28 \times 10.0 \times C)$$

Dividing through by -100, you get:

$$1 = 1/(628 \times 10.0 \times C)$$

Multiply each side of this by C, and you obtain $C = 1/(628 \times 10.0)$. This can be worked out with a calculator. You should find that $C = 0.000159$ to three significant figures. Because the frequency is given in megahertz, the capacitance comes out in microfarads. That means $C = 0.000159\ \mu\text{F}$. You can also say it is 159 pF. (Remember that 1 pF = 0.000001 μF.)

Points in the *RC* Plane

In a circuit containing resistance and capacitive reactance, the characteristics are two-dimensional in a way that is analogous to the situation with the *RL* plane from the previous chapter. The resistance ray and the capacitive-reactance ray can be placed end to end at right angles to make a quarter plane called the *RC plane* (Fig. 14-5). Resistance is plotted horizontally, with increasing values toward the right. Capacitive reactance is plotted downward, with increasingly negative values as you go down.

The combinations of R and X_C in this *RC* plane form impedances. You'll learn about impedance in greater detail in the next chapter. Each point on the *RC* plane corresponds to one and only one impedance. Conversely, each specific impedance coincides with one and only one point on the plane.

Any impedance that consists of a resistance R and a capacitive reactance X_C can be written in the form $R + jX_C$. Remember that X_C is always negative or zero. Because of this, engineers will often write $R - jX_C$ instead.

If an impedance is a pure resistance R with no reactance, then the complex impedance is $R - j0$ (or $R + j0$; it doesn't matter if j is multiplied by 0!). If $R = 3\ \Omega$ with no reactance, you get an impedance of $3 - j0$, which corresponds to the point $(3,j0)$ on the *RC* plane. If you have a pure capacitive reactance, say $X_C = -4\ \Omega$, then the complex impedance is $0 - j4$, and this is at the point $(0,-j4)$ on the *RC* plane. Again, it's important, for completeness, to write the "0" and not just the "$-j4$." The points for $3 - j0$ and $0 - j4$, and two others, are plotted on the *RC* plane in Fig. 14-6.

In practical circuits, all capacitors have some *leakage resistance.* If the frequency goes to zero (pure dc), a tiny current always flows, because no capacitor has a perfect insulator between its plates. In addition to this, all resistors have a little capacitive reactance because they occupy a finite physical space. So there is no such thing as a mathematically perfect resistor, either. The points $3 - j0$ and

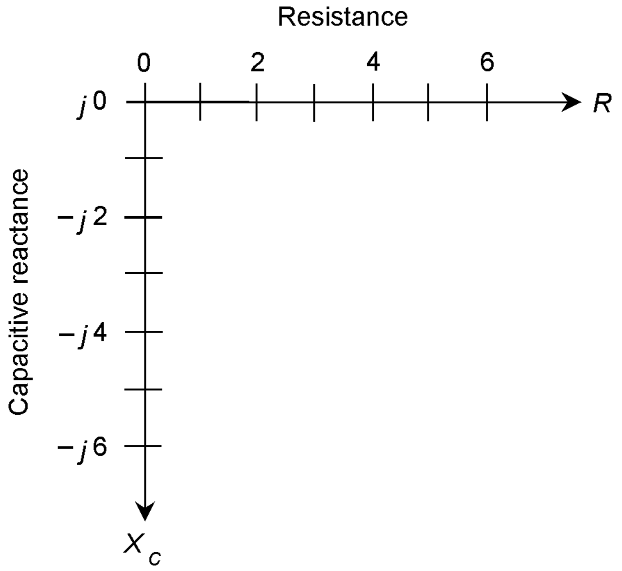

14-5 The quarter plane for capacitive reactance (X_C) and resistance (R). This is also known as the *RC* quarter-plane, or simply as the *RC* plane.

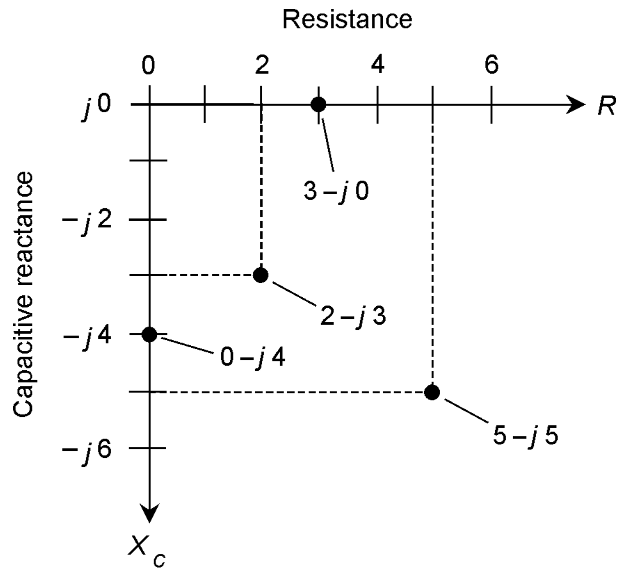

14-6 Four points in the *RC* plane.

0 − *j*4 represent an *ideal resistor* and an *ideal capacitor*, respectively—components that can be worked with in theory, but that you will never see in the real world.

Sometimes, resistance and capacitive reactance are both placed in a circuit deliberately. Then you get impedances such as 2 − *j*3 and 5 − *j*5, both shown in Fig. 14-6.

Remember that the values for X_C are reactances, not the actual capacitances. If you raise or lower the frequency, the value of X_C will change. A higher frequency causes X_C to get smaller negatively (closer to zero). A lower frequency causes X_C to get larger negatively (farther from zero, or lower down on the *RC* plane). If the frequency goes to zero, then the capacitive reactance drops off the bottom of the *RC* plane to negative infinity!

Vectors in the *RC* Plane

Recall from the last chapter that *RL* impedances can be represented as vectors. The same is true for *RC* impedances.

In Fig. 14-6, four different complex impedance points are shown. Each point is represented by a certain distance to the right of the origin (0,*j*0), and a certain displacement downward. The first of these is the resistance, *R*, and the second is the capacitive reactance, X_C. The complex *RC* impedance is a two-dimensional quantity.

Impedance points in the *RC* plane can be rendered as vectors, just as they can in the *RL* plane. Then the points become rays, each with a certain length and direction. The magnitude and direction for a vector, and the coordinates for the point, both uniquely define the same complex impedance. The length of the vector is the distance of the point from the origin, and the direction is the angle measured *clockwise* from the resistance (*R*) line, and specified in *negative degrees.* The equivalent vectors, for the points in Fig. 14-6, are shown in Fig. 14-7.

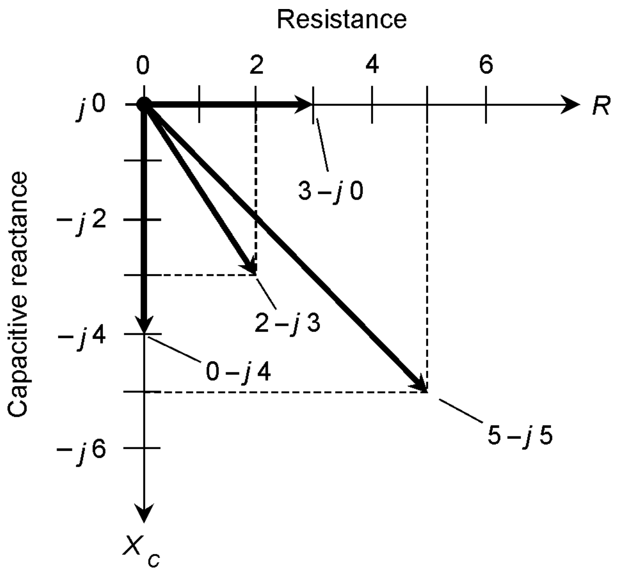

14-7 Four vectors in the *RC* plane, corresponding to the points shown in Fig. 14-6.

Current Leads Voltage

When ac is driven through a capacitor and starts to increase (in either direction), it takes a fraction of a cycle for the voltage between the plates to follow. Once the current starts decreasing from its maximum peak (in either direction) in the cycle, it again takes a fraction of a cycle for the voltage to follow. The instantaneous voltage can't quite keep up with the instantaneous current, as it does in a pure resistance. Thus, in a circuit containing capacitive reactance, the voltage lags the current in phase. Another, and more often used, way of saying this is that the current *leads* the voltage.

Pure Capacitance

Suppose an ac voltage source is connected across a capacitor. Imagine that the frequency is low enough, and/or the capacitance is small enough, so the absolute value of the capacitive reactance, X_C, is extremely large compared with the resistance, R. Then the current leads the voltage by just about 90° (Fig. 14-8).

The situation depicted in Fig. 14-8 represents a pure capacitive reactance. The vector in the *RC* plane in this situation points straight down. Its angle is −90° from the R axis.

Capacitance and Resistance

When the resistance in a resistance-capacitance circuit is significant compared with the absolute value of the capacitive reactance, the current leads the voltage by something less than 90° (Fig. 14-9). If R is small compared with the absolute value of X_C, the difference is almost a quarter of a cycle. As R gets larger, or as the absolute value of X_C becomes smaller, the phase difference decreases. A circuit containing resistance and capacitance is called an *RC circuit.*

The value of R in an *RC* circuit might increase relative to the absolute value of X_C because resistance is deliberately put into a circuit. It can also happen if the frequency becomes so high that the absolute value of the capacitive reactance drops to a value comparable with the loss resistance in

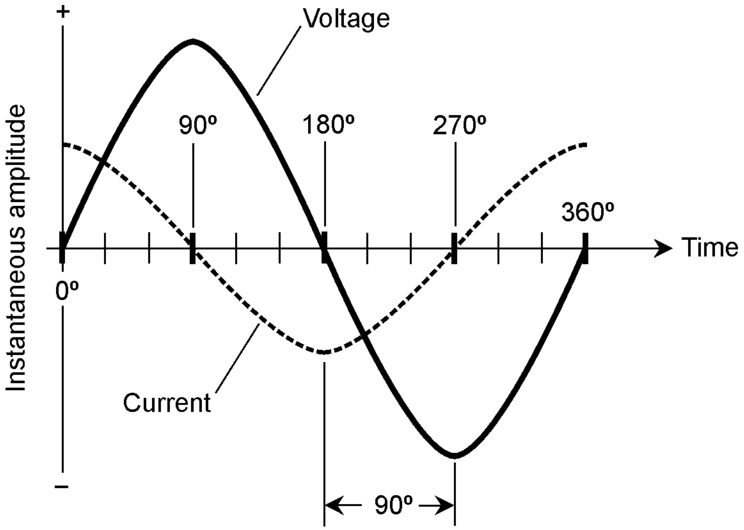

14-8 In a pure capacitive reactance, the current leads the voltage by 90°.

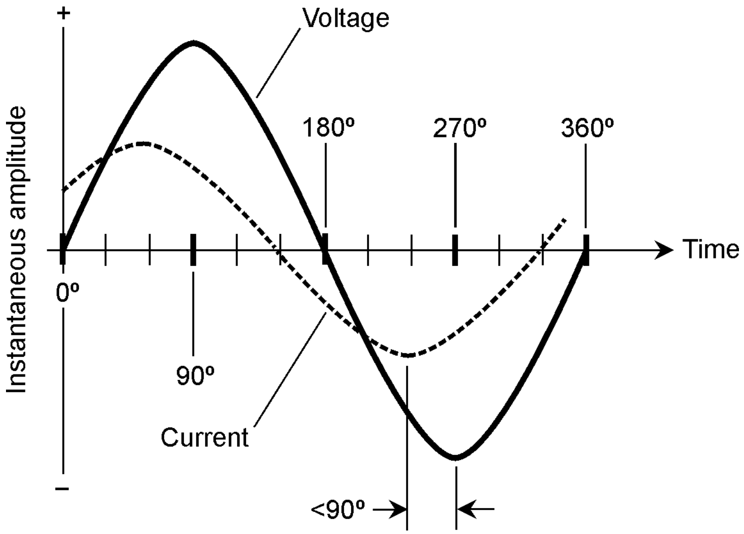

14-9 In a circuit with capacitive reactance and resistance, the current leads the voltage by less than 90°.

the circuit conductors. In either case, the situation can be represented by a resistance, R, in series with a capacitive reactance, X_C (Fig. 14-10).

If you know the values of X_C and R, you can find the *angle of lead*, also called the *RC phase angle*, by plotting the point $R - jX_C$ on the RC plane, drawing the vector from the origin $0 - j0$ out to that point, and then measuring the angle of the vector clockwise from the R axis. You can use a protractor to measure this angle, as you did in the previous chapter for RL phase angles. Or you can use trigonometry to calculate the angle.

As with RL circuits, you need only know the ratio of X_C to R to determine the phase angle. For example, if $X_C = -4\ \Omega$ and $R = 7\ \Omega$, you'll get the same angle as with $X_C = -400\ \Omega$ and $R = 700\ \Omega$, or with $X_C = -16\ \Omega$ and $R = 28\ \Omega$. The phase angle will be the same whenever the ratio of X_C to R is equal to –4:7.

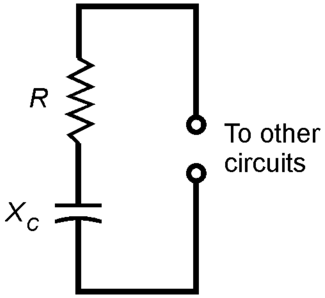

14-10 Schematic representation of a circuit containing resistance and capacitive reactance.

Pure Resistance

As the resistance in an *RC* circuit gets large compared with the absolute value of the capacitive reactance, the angle of lead becomes smaller. The same thing happens if the absolute value of X_C gets small compared with the value of *R*.

When *R* is many times larger than the absolute value of X_C, whatever their actual values, the vector in the *RC* plane points almost along the *R* axis. Then the *RC* phase angle is close to 0°. The voltage comes nearly into phase with the current. The plates of the capacitor do not come anywhere near getting fully charged with each cycle. The capacitor is said to "pass the ac" with very little loss, as if it were shorted out. But it will still have an extremely high X_C for any ac signals at much lower frequencies that might exist across it at the same time. (This property of capacitors can be put to use in electronic circuits. An example is when an engineer wants to let radio-frequency signals get through while blocking signals at audio frequencies.)

Ultimately, if the absolute value of the capacitive reactance gets small enough, the circuit acts as a pure resistance, and the current is in phase with the voltage.

How Much Lead?

If you know the ratio of capacitive reactance to resistance, or X_C/R, in an *RC* circuit, then you can find the phase angle. Of course, you can find this angle if you know the precise values, too.

Pictorial Method

You can use a protractor and a ruler to find phase angles for *RC* circuits, just as you did with *RL* circuits in the previous chapter, as long as the angles aren't too close to 0° or 90°. First, draw a line somewhat longer than 10 cm, going from left to right on the paper. Then, use the protractor to construct a line going somewhat more than 10 cm vertically downward, starting at the left end of the horizontal line. The horizontal line is the *R* axis of an *RC* plane. The line going down is the X_C axis.

If you know the actual values of X_C and *R*, divide or multiply them by a constant, chosen to make both values fall between −100 and 100. For example, if $X_C = -3800\ \Omega$ and $R = 7400\ \Omega$, divide them both by 100, getting −38 and 74. Plot these points on the lines. The X_C point goes 38 mm down from the intersection point between your two axes. The *R* point goes 74 mm to the right of the intersection point. Next, draw a line connecting the two points, as shown in Fig. 14-11. This line will be at a slant and will form a triangle along with the two axes. This is a right triangle, with the right angle at the origin of the *RC* plane. Measure the angle between the slanted line and the *R* axis. Use the protractor for this. Extend the lines, if necessary, using the ruler, to get a good reading

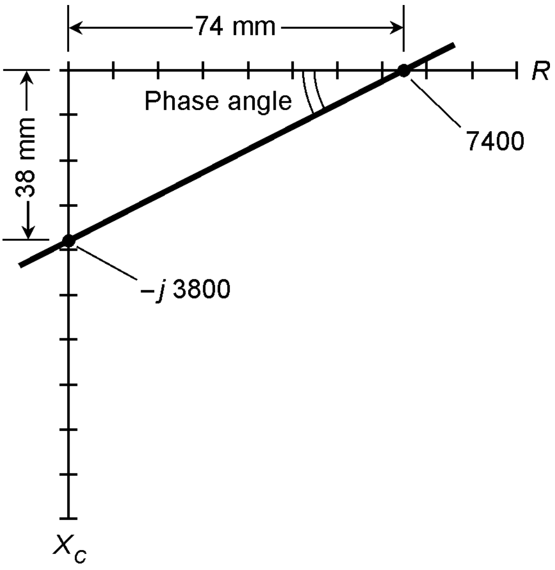

14-11 Pictorial method of finding phase angle in a circuit containing resistance and capacitive reactance.

on the protractor. This angle will be between 0 and 90°. Multiply this reading by −1 to get the *RC* phase angle. That is, if the protractor shows 27°, the *RC* phase angle is −27°.

The actual vector is found by constructing a rectangle using the origin and your two points, making new perpendicular lines to complete the figure. The vector is the diagonal of this rectangle, running out from the origin (Fig. 14-12). The phase angle is the angle between the *R* axis and this vector, multiplied by −1. It will have the same measure as the angle of the slanted line you constructed in Fig. 14-11.

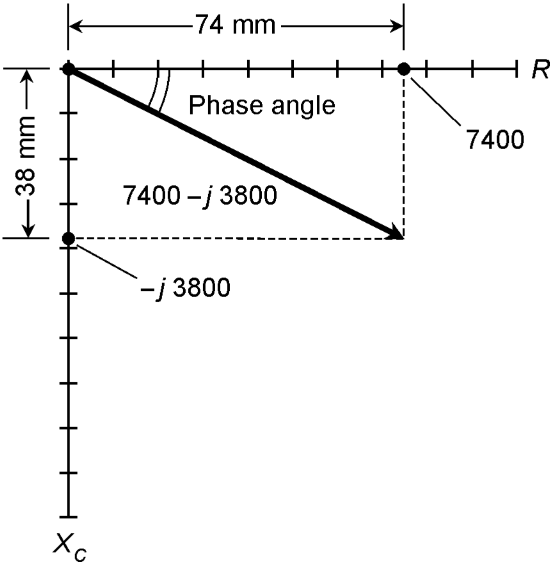

14-12 Another pictorial method of finding phase angle in a circuit containing resistance and capacitive reactance. This method shows the actual impedance vector.

Trigonometric Method

Using trigonometry, you can determine the *RC* phase angle more precisely than the pictorial method allows. Given the values of X_C and R, the *RC* phase angle is the arctangent of their ratio. Phase angle in *RC* circuits is symbolized by the lowercase Greek letter ϕ, just as it is in *RL* circuits. Here are the formulas:

$$\phi = \tan^{-1} (X_C/R) \qquad \text{or} \qquad \phi = \arctan (X_C/R)$$

When doing problems of this kind, remember to use the *capacitive reactance* values for X_C, and not the capacitance values. This means that, if you are given the capacitance, you must use the formula for X_C in terms of capacitance and frequency and then calculate the phase angle. You should get angles that come out negative or zero. This indicates that they're *RC* phase angles rather than *RL* phase angles (which are always positive or zero).

Problem 14-4

Suppose the capacitive reactance in an *RC* circuit is –3800 Ω and the resistance is 7400 Ω. What is the phase angle?

Find the ratio $X_C/R = -3800/7400$. The calculator display should show you something like –0.513513513. Find the arctangent, or $\tan^{-1}$, getting a phase angle of –27.18111109° on the calculator display. Round this off to –27.18°.

Problem 14-5

Suppose an *RC* circuit works at a frequency of 3.50 MHz. It has a resistance of 130 Ω and a capacitance of 150 pF. What is the phase angle?

First, find the capacitive reactance for a capacitor of 150 pF at 3.50 MHz. Convert the capacitance to microfarads, getting $C = 0.000150$ μF. Remember that *micro*farads go with *mega*hertz (millionths go with millions to cancel each other out). Then:

$$\begin{aligned} X_C &= -1/(6.28 \times 3.50 \times 0.000150) \\ &= -1/0.003297 = -303\ \Omega \end{aligned}$$

Now you can find the ratio $X_C/R = -303/130 = -2.33$. The phase angle is equal to the arctangent of –2.33, or –66.8°.

Problem 14-6

What is the phase angle in the preceding circuit if the frequency is raised to 7.10 MHz?

You need to find the new value for X_C, because it will change as a result of the frequency change. Calculating:

$$\begin{aligned} X_C &= -1/(6.28 \times 7.10 \times 0.000150) \\ &= -1/0.006688 = -150\ \Omega \end{aligned}$$

The ratio X_C/R in this case is equal to –150/130, or –1.15. The phase angle is the arctangent of –1.15, which turns out to be –49.0°.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. As the size of the plates in a capacitor increases, all other things being equal,
   (a) the value of X_C increases negatively.
   (b) the value of X_C decreases negatively.
   (c) the value of X_C does not change.
   (d) we cannot say what happens to X_C without more data.

2. If the dielectric material between the plates of a capacitor is changed, all other things being equal,
   (a) the value of X_C increases negatively.
   (b) the value of X_C decreases negatively.
   (c) the value of X_C does not change.
   (d) we cannot say what happens to X_C without more data.

3. As the frequency of a wave gets lower, all other things being equal, the value of X_C for a capacitor
   (a) increases negatively.
   (b) decreases negatively.
   (c) does not change.
   (d) depends on the current.

4. What is the reactance of a 330-pF capacitor at 800 kHz?
   (a) −1.66 Ω
   (b) −0.00166 Ω
   (c) −603 Ω
   (d) −603 kΩ

5. Suppose a capacitor has a reactance of −4.50 Ω at 377 Hz. What is its capacitance?
   (a) 9.39 μF
   (b) 93.9 μF
   (c) 7.42 μF
   (d) 74.2 μF

6. Suppose a 47-μF capacitor has a reactance of −47 Ω. What is the frequency?
   (a) 72 Hz
   (b) 7.2 MHz
   (c) 0.000072 Hz
   (d) 7.2 Hz

7. Suppose a capacitor has $X_C = -8800\ \Omega$ at $f = 830$ kHz. What is C?
   - (a) 2.18 μF
   - (b) 21.8 pF
   - (c) 0.00218 μF
   - (d) 2.18 pF
8. Suppose a capacitor has $C = 166$ pF at $f = 400$ kHz. What is X_C?
   - (a) −2.4 kΩ
   - (b) −2.4 Ω
   - (c) $-2.4 \times 10^{-6}\ \Omega$
   - (d) −2.4 MΩ
9. Suppose a capacitor has $C = 4700$ μF and $X_C = -33\ \Omega$. What is f?
   - (a) 1.0 Hz
   - (b) 10 Hz
   - (c) 1.0 kHz
   - (d) 10 kHz
10. Each point in the *RC* plane
    - (a) corresponds to a unique inductance.
    - (b) corresponds to a unique capacitance.
    - (c) corresponds to a unique combination of resistance and capacitance.
    - (d) corresponds to a unique combination of resistance and reactance.
11. If R increases in an *RC* circuit, but X_C is always zero, the vector in the *RC* plane will
    - (a) rotate clockwise.
    - (b) rotate counterclockwise.
    - (c) always point straight toward the right.
    - (d) always point straight down.
12. If the resistance R increases in an *RC* circuit, but the capacitance and the frequency are nonzero and constant, then the vector in the *RC* plane will
    - (a) get longer and rotate clockwise.
    - (b) get longer and rotate counterclockwise.
    - (c) get shorter and rotate clockwise.
    - (d) get shorter and rotate counterclockwise.
13. Each complex impedance value $R - jX_C$
    - (a) represents a unique combination of resistance and capacitance.
    - (b) represents a unique combination of resistance and reactance.
    - (c) represents a unique combination of resistance and frequency.
    - (d) All of the above are true.

14. In an *RC* circuit, as the ratio X_C/R approaches zero, the phase angle
    (a) approaches −90°.
    (b) approaches 0°.
    (c) stays the same.
    (d) cannot be found.

15. In a purely resistive circuit, the phase angle is
    (a) increasing.
    (b) decreasing.
    (c) 0°.
    (d) −90°.

16. If $X_C/R = -1$, then what is the phase angle?
    (a) 0°
    (b) −45°
    (c) −90°
    (d) Impossible to find because there's not enough data given

17. In Fig. 14-13, the impedance shown is
    (a) $8.02 + j323$.
    (b) $323 + j8.02$.
    (c) $8.02 - j323$.
    (d) $323 - j8.02$.

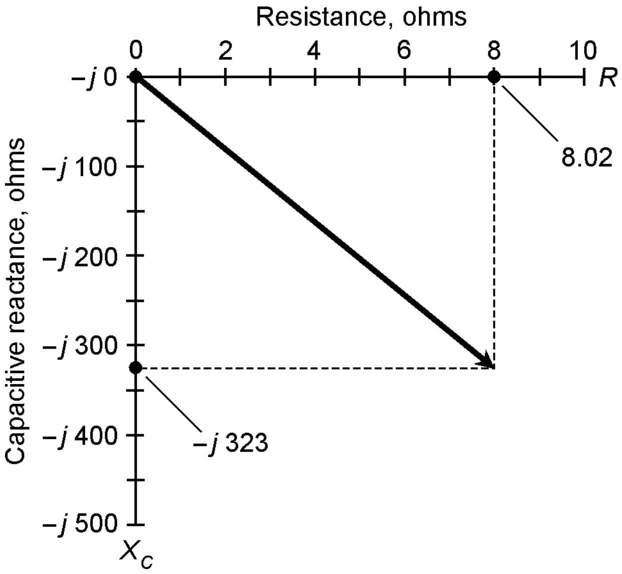

14-13 Illustration for Quiz Questions 17 and 18.

18. In Fig. 14-13, note that the R and X_C scale divisions are not the same size. What is the actual phase angle?

(a) −1.42°
(b) About −60°, from the looks of it
(c) −58.9°
(d) −88.6°

19. Suppose an RC circuit consists of a 150-pF capacitor and a 330-Ω resistor in series. What is the phase angle at a frequency of 1.34 MHz?

(a) −67.4°
(b) −22.6°
(c) −24.4°
(d) −65.6°

20. Suppose an RC circuit has a capacitance of 0.015 μF. The resistance is 52 Ω. What is the phase angle at 90 kHz?

(a) −24°
(b) −0.017°
(c) −66°
(d) None of the above

15
CHAPTER
Impedance and Admittance

IN THIS CHAPTER, A COMPLETE, WORKING DEFINITION OF COMPLEX IMPEDANCE IS DEVELOPED. YOU'LL also get acquainted with *admittance*, the extent to which an ac circuit allows (or admits) current flow, rather than impeding it. As we develop these concepts, let's review, and then expand on, some of the material presented in the previous couple of chapters.

Imaginary Numbers

Have you been wondering what *j* actually means in expressions of impedance? Well, *j* is nothing but a number: the positive square root of −1. There's a negative square root of −1, too, and it is equal to −*j*. When either *j* or −*j* is multiplied by itself, the result is −1. (Pure mathematicians often denote these same numbers as *i* or −*i*.)

The positive square root of −1 is known as the *unit imaginary number*. The *set of imaginary numbers* is composed of real-number multiples of *j* or −*j*. Some examples are *j*4, *j*35.79, −*j*25.76, and −*j*25,000.

The square of an imaginary number is always negative. Some people have trouble grasping this, but when you think long and hard about it, all numbers are abstractions. Imaginary numbers are no more imaginary (and no less real) than so-called real numbers such as 4, 35.79, −25.76, or −25,000.

The unit imaginary number *j* can be multiplied by any real number on a conventional *real number line*. If you do this for all the real numbers on the real number line, you get an *imaginary number line* (Fig. 15-1). The imaginary number line should be oriented at a right angle to the real number line when you want to graphically portray real and imaginary numbers at the same time.

In electronics, real numbers represent resistances. Imaginary numbers represent reactances.

Complex Numbers

When you add a real number and an imaginary number, you get a *complex number*. In this context, the term *complex* does not mean "complicated." A better word would be *composite*. Examples are

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

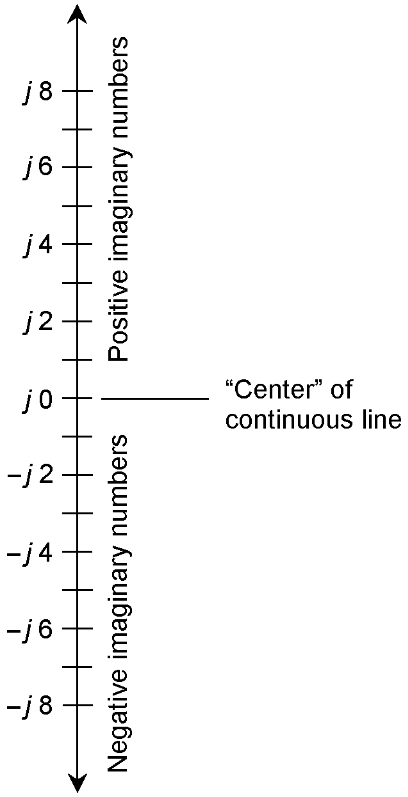
j 8
j 6
j 4
j 2
j 0
−j 2
−j 4
−j 6
−j 8
Positive imaginary numbers
Negative imaginary numbers
"Center" of continuous line

15-1 The imaginary number line.

$4 + j5$, $8 - j7$, $-7 + j13$, and $-6 - j87$. The set of complex numbers needs two dimensions—a *plane*—to be graphically defined.

Adding and Subtracting Complex Numbers

Adding complex numbers is just a matter of adding the real parts and the complex parts separately. For example, the sum of $4 + j7$ and $45 - j83$ works out like this:

$$\begin{aligned}&(4 + 45) + j(7 - 83)\\&= 49 + j(-76)\\&= 49 - j76\end{aligned}$$

Subtracting complex numbers is a little more involved; it's best to convert a difference to a sum. For example, the difference $(4 + j7) - (45 - j83)$ can be found by multiplying the second complex number by -1 and then adding the result:

$$\begin{aligned}&(4 + j7) - (45 - j83)\\&= (4 + j7) + [-1(45 - j83)]\\&= (4 + j7) + (-45 + j83)\\&= -41 + j90\end{aligned}$$

Multiplying Complex Numbers

When you multiply these numbers, you should treat them as sums of number pairs, that is, as *binomials.* It's easier to give the general formula than to work with specifics here. If *a*, *b*, *c*, and *d* are real numbers (positive, negative, or zero), then:

$$\begin{aligned}&(a + jb)\ (c + jd)\\&= ac + jad + jbc + j^2bd\\&= (ac - bd) + j(ad + bc)\end{aligned}$$

Fortunately, you won't encounter complex number multiplication problems very often in electronics. Nevertheless, a working knowledge of how complex numbers multiply can help you get a solid grasp of them.

The Complex Number Plane

A complete *complex number plane* is made by taking the real and imaginary number lines and placing them together, at right angles, so that they intersect at the zero points, 0 and $j0$. This is shown in Fig. 15-2. The result is a *Cartesian coordinate plane*, just like the ones people use to make graphs of everyday things such as stock price versus time.

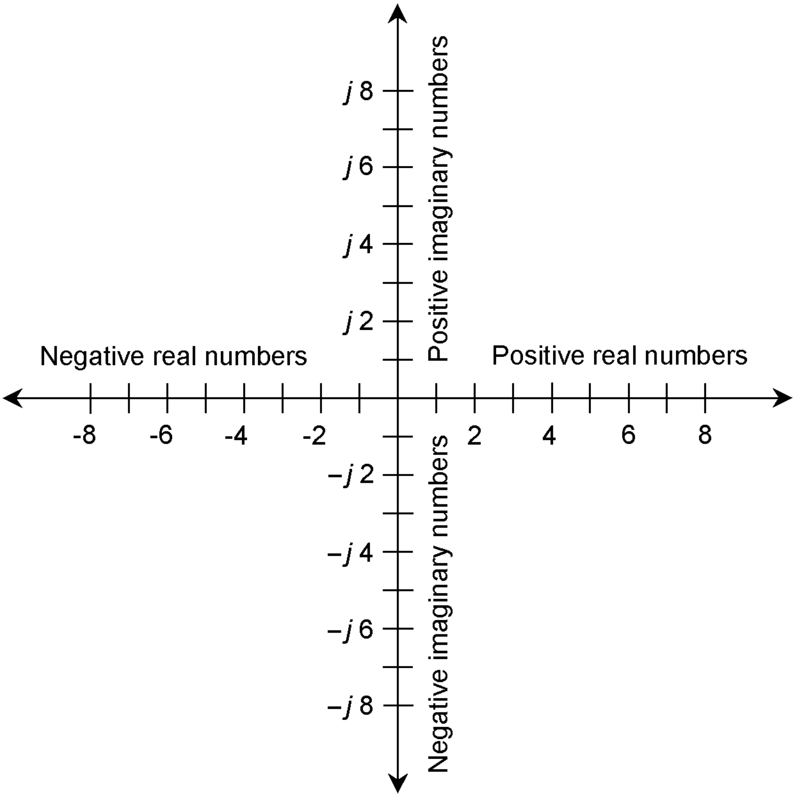
j 8
j 6
j 4
j 2
Positive imaginary numbers
Negative real numbers
Positive real numbers
-8
-6
-4
-2
2
4
6
8
–j 2
–j 4
–j 6
–j 8
Negative imaginary numbers

15-2 The complex number plane.

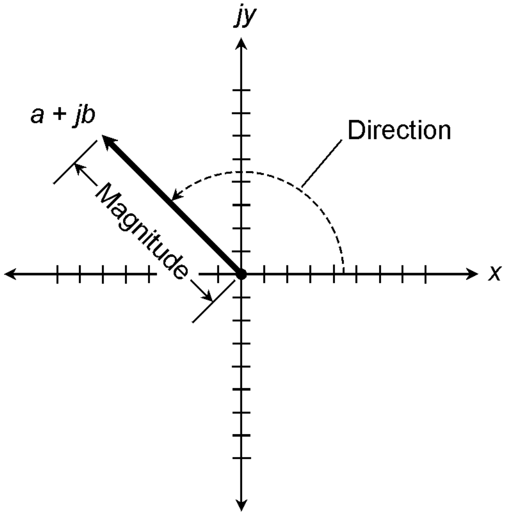

15-3 Magnitude and direction of a vector in the complex number plane.

Complex Number Vectors

Complex numbers can also be represented as vectors. This gives each complex number a unique *magnitude* and a unique *direction.* The magnitude is the distance of the point $a + jb$ from the origin $0 + j0$. The direction is the angle of the vector, expressed counterclockwise from the positive real-number axis. This is shown in Fig. 15-3.

Absolute Value

The *absolute value* of a complex number $a + jb$ is the length, or magnitude, of its vector in the complex plane, measured from the origin (0,0) to the point (a,b).

In the case of a *pure real number* $a + j0$, the absolute value is simply the real number itself, a, if a is positive. If a is negative, then the absolute value of $a + j0$ is equal to $-a$.

In the case of a *pure imaginary number* $0 + jb$, the absolute value is equal to b, if b (a real number) is positive. If b is negative, the absolute value of $0 + jb$ is equal to $-b$.

If the number $a + jb$ is neither pure real or pure imaginary, the absolute value must be found by using a formula. First, square both a and b. Then add them. Finally, take the square root. This is the length, c, of the vector $a + jb$. The situation is illustrated in Fig. 15-4.

Problem 15-1

Find the absolute value of the complex number $-22 - j0$.

This is a pure real number. Actually, it is the same as $-22 + j0$, because $j0 = 0$. Therefore, the absolute value of this complex number is $-(-22) = 22$.

Problem 15-2

Find the absolute value of $0 - j34$.

This is a pure imaginary number. The value of b in this case is -34, because $0 - j34 = 0 + j(-34)$. Therefore, the absolute value is $-(-34) = 34$.

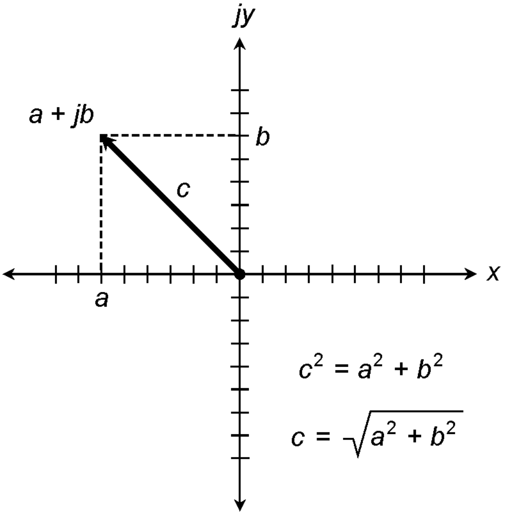

15-4 Calculation of absolute value, or vector length. Here, the vector length is represented by *c*.

Problem 15-3

Find the absolute value of $3 - j4$.

In this number, $a = 3$ and $b = -4$. Squaring both of these, and adding the results, gives us $3^2 + (-4)^2 = 9 + 16 = 25$. The square root of 25 is 5. Therefore, the absolute value of this complex number is 5.

The *RX* Plane

Recall the planes for resistance (R) and inductive reactance (X_L) from Chap. 13. This is the same as the upper-right quadrant of the complex number plane shown in Fig. 15-2. Similarly, the plane for resistance and capacitive reactance (X_C) is the same as the lower-right quadrant of the complex number plane. Resistances are represented by nonnegative real numbers. Reactances, whether they are inductive (positive) or capacitive (negative), correspond to imaginary numbers.

No Negative Resistance

There is no such thing, strictly speaking, as negative resistance. You cannot have anything better than a perfect conductor. In some cases, a supply of direct current, such as a battery, can be treated as a negative resistance; in other cases, you can have a device that acts as if its resistance were negative under certain changing (or dynamic) conditions. But for most practical applications in the *RX plane*, the resistance value is always positive. You can remove the negative axis, along with the upper-left and lower-left quadrants, of the complex number plane, obtaining a half plane, as shown in Fig. 15-5, and still get a complete set of coordinates for depicting complex impedances.

"Negative Inductors" and "Negative Capacitors"

Capacitive reactance, X_C, is effectively an extension of inductive reactance, X_L, into the realm of negatives. Capacitors act like "negative inductors." It's equally true to say that inductors act like "negative capacitors," because the negative of a negative number is a positive number. Reactance can vary from extremely large negative values, through zero, to extremely large positive values.

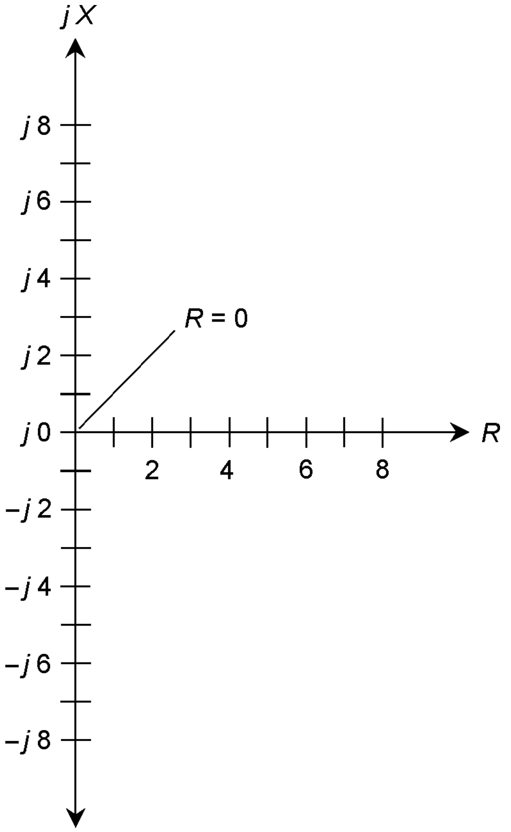

15-5 The complex impedance plane, also called the resistance-reactance (*RX*) plane.

Vector Representation of Impedance

Any impedance $R + jX$ can be represented by a complex number of the form $a + jb$. Just let $R = a$ and $X = b$. Now try to envision how the impedance vector changes as either R or X, or both, are varied. If X remains constant, an increase in R causes the vector to get longer. If R remains constant and X_L gets larger, the vector grows longer. If R stays the same but X_C gets larger negatively, the vector grows longer.

Think of the point $R + jX$ moving around in the *RX* plane, and imagine where the corresponding points on the axes lie. These points can be found by drawing dashed lines from the point $R + jX$ to the R and X axes, so that the dashed lines intersect the axes at right angles. Some examples are shown in Fig. 15-6.

Now think of the points for R and X moving toward the right and left, or up and down, on their axes. Imagine what happens to the point $R + jX$ in various scenarios. This is how impedance changes as the resistance and reactance in a circuit are varied.

Resistance is one-dimensional. Reactance is also one-dimensional. But impedance is two-dimensional. To fully define impedance, you must render it on a two-dimensional coordinate system such as the *RX* plane. The resistance and the reactance can change independently of one another.

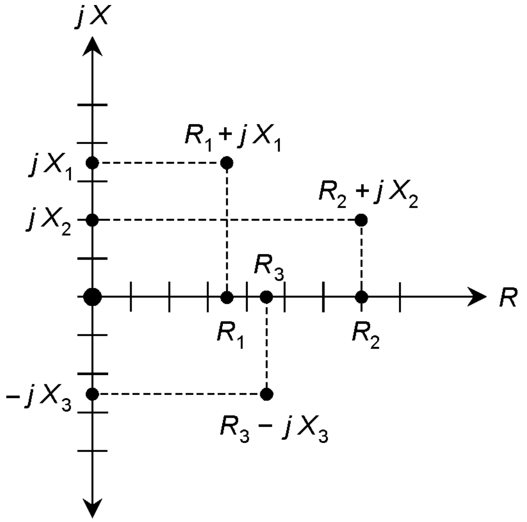

15-6 Some points in the complex impedance plane, and their resistive and reactive components on the axes.

Absolute-Value Impedance

You'll occasionally read or hear that the "impedance" of some device or component is a certain number of ohms. For example, in audio electronics, there are "8-Ω" speakers and "600-Ω" amplifier inputs. How, you ask, can manufacturers quote a single number for a quantity that is two-dimensional and needs two numbers to be completely expressed?

That's a good question, and there are two answers. First, figures like this refer to devices that have *purely resistive impedances*, also known as *nonreactive impedances.* Thus, the 8-Ω speaker really has a complex impedance of $8 + j0$, and the 600-Ω input circuit is designed to operate with a complex impedance at, or near, $600 + j0$. Second, you can talk about the length of the impedance vector (that is, the absolute value of the complex impedance), calling this a certain number of ohms. If you talk about impedance this way, however, you are being ambiguous. There can exist an infinite number of different vectors of any given length in the *RX* plane.

Sometimes, the uppercase italic letter Z is used in place of the word *impedance* in general discussions. This is what engineers mean when they say things like "$Z = 50$ Ω" or "$Z = 300$ Ω nonreactive." In this context, if no specific impedance is given, "$Z = 8$ Ω" can theoretically refer to $8 + j0$, $0 + j8$, $0 - j8$, or any other complex impedance point on a half circle consisting of all points 8 units from $0 + j0$. This is shown in Fig. 15-7.

Problem 15-4

Name seven different complex impedances that can theoretically be meant by the expression "$Z = 10$ Ω."

It's easy name three: $0 + j10$, $10 + j0$, and $0 - j10$. These represent pure inductance, pure resistance, and pure capacitance, respectively.

A right triangle can exist having sides in a ratio of 6:8:10 units. This is true because $6^2 + 8^2 = 10^2$. (Check it and see!) Therefore, you can have $6 + j8$, $6 - j8$, $8 + j6$, and $8 - j6$, all complex impedances whose absolute value is 10.

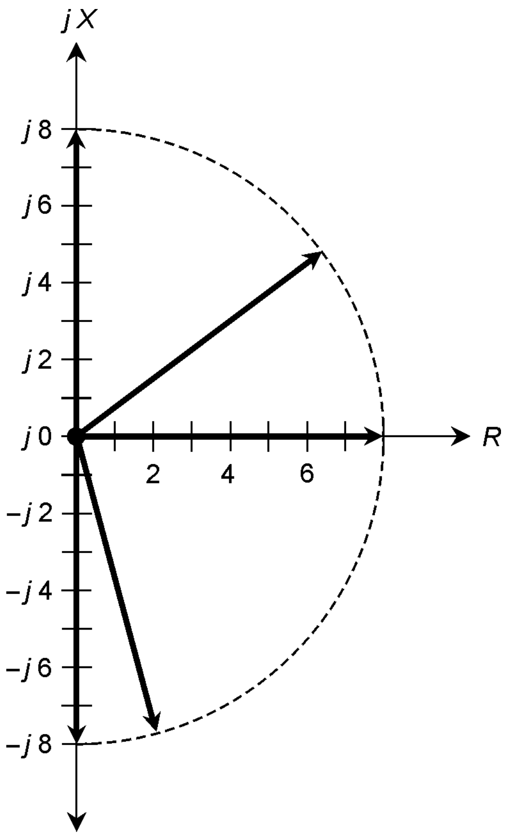

15-7 Vectors representing an absolute-value impedance of 8 Ω.

Characteristic Impedance

There is a rather exotic property of certain electronic components that you'll sometimes hear or read about. It is called *characteristic impedance* or *surge impedance*, and is symbolized Z_o. It is a specification of an important property of *transmission lines.* It can always be expressed as a positive real number, in ohms.

Transmission Lines

When it is necessary to get energy or signals from one place to another, a transmission line is required. These almost always take either of two forms, *coaxial* or *two-wire* (also called *parallel-wire*). Cross-sectional renditions of both types are shown in Fig. 15-8. Examples of transmission lines include the "ribbon" that goes from a television antenna to the receiver, the cable running from a hi-fi amplifier to the speakers, and the set of wires that carries electricity over the countryside.

Factors Affecting Z_o

The Z_o of a parallel-wire transmission line depends on the diameter of the wires, on the spacing between the wires, and on the nature of the insulating material separating the wires. In general, the Z_o

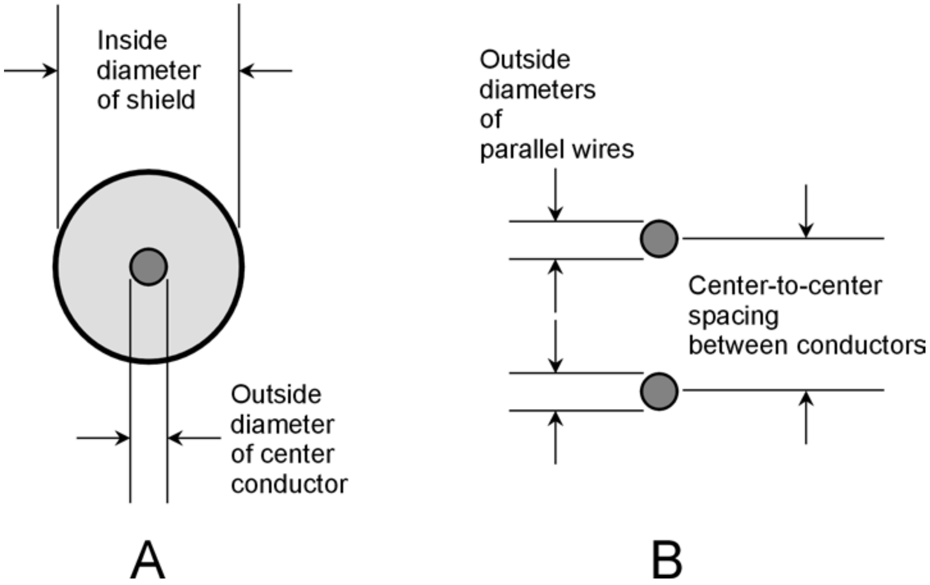
Inside diameter of shield
Outside diameter of center conductor
A
Outside diameters of parallel wires
Center-to-center spacing between conductors
B

15-8 Edge-on views of coaxial transmission line (A) and parallel-wire line (B). In either type of line, Z_o depends on the conductor diameters and spacing, and on the nature of the dielectric material between the conductors. See text for discussion.

increases as the wire diameter gets smaller, and decreases as the wire diameter gets larger, all other things being equal.

In a coaxial line, as the center conductor gets thicker, the Z_o decreases if the shield stays the same size. If the center conductor stays the same size and the *shield* increases in diameter, the Z_o increases.

For either type of line, the Z_o increases as the spacing between wires, or between the center conductor and the shield, gets larger. The Z_o decreases as the spacing is reduced. Solid dielectric materials such as polyethylene reduce the Z_o of a transmission line, compared with air or a vacuum, when placed between the conductors.

An Example of Z_o in Practice

In rigorous terms, the ideal characteristic impedance for a transmission line is determined according to the nature of the *load* with which the line works.

For a system having a purely resistive impedance of a certain number of ohms, the best line Z_o value is that same number of ohms. If the load impedance is much different from the characteristic impedance of the transmission line, excessive power is wasted in heating up the transmission line.

Imagine that you have a so-called 300-Ω frequency-modulation (FM) receiving antenna, such as the folded-dipole type that you can mount indoors. Suppose that you want the best possible reception. Of course, you should choose a good location for the antenna. You should make sure that the transmission line between your radio and the antenna is as short as possible. But you should also be sure that you purchase 300-Ω TV ribbon. It has a value of Z_o that has been optimized for use with antennas whose impedances are close to $300 + j0$.

Impedance matching is the process of making sure that the impedance of a load (such as an antenna) is purely resistive, with an ohmic value equal to the characteristic impedance of the transmission line connected to it. This concept will be discussed in more detail in the next chapter.

Conductance

In an ac circuit, electrical *conductance* works the same way as it does in a dc circuit. Conductance is symbolized by the capital letter G. It was introduced in Chap. 2. The relationship between conductance and resistance is simple: $G = 1/R$. The standard unit of conductance is the *siemens.* The larger the value of conductance, the smaller the resistance, and the more current will flow. Conversely, the smaller the value of G, the greater the value of R, and the less current will flow.

Susceptance

Sometimes, you'll come across the term *susceptance* in reference to ac circuits. Susceptance is symbolized by the capital letter B. It is the reciprocal of reactance. Susceptance can be either capacitive or inductive. These quantities are symbolized as B_C and B_L, respectively. Therefore we have these two relations:

$$B_C = 1/X_C$$
$$B_L = 1/X_L$$

All values of B theoretically contain the j operator, just as do all values of X. But when it comes to finding reciprocals of quantities containing j, things get tricky. The reciprocal of j is equal to its negative! Expressed mathematically, we have these two facts:

$$1/j = -j$$
$$1/(-j) = j$$

As a result of these properties of j, the sign reverses whenever you find a susceptance value in terms of a reactance value. When expressed in terms of j, inductive susceptance is negative imaginary, and capacitive susceptance is positive imaginary—just the opposite situation from inductive reactance and capacitive reactance.

Suppose you have an inductive reactance of 2 Ω. This is expressed in imaginary terms as $j2$. To find the inductive susceptance, you must find $1/(j2)$. Mathematically, this expression can be converted to a real-number multiple of j in the following manner:

$$\begin{aligned} 1/(j2) &= (1/j)(\tfrac{1}{2}) \\ &= (1/j)0.5 \\ &= -j0.5 \end{aligned}$$

Now suppose you have a capacitive reactance of 10 Ω. This is expressed in imaginary terms as $-j10$. To find the capacitive susceptance, you must find $1/(-j10)$. Here's how this can be converted to the straightforward product of j and a real number:

$$\begin{aligned}1/(-j10) &= (1/-j)(^1\!/_{10})\\ &= (1/-j)0.1\\ &= j0.1\end{aligned}$$

When you want to find an imaginary value of susceptance in terms of an imaginary value of reactance, first take the reciprocal of the real-number part of the expression, and then multiply the result by −1.

Problem 15-5

Suppose you have a capacitor of 100 pF at a frequency of 3.00 MHz. What is B_C?

First, find X_C by the formula for capacitive reactance:

$$X_C = -1/(6.28fC)$$

Note that 100 pF = 0.000100 μF. Therefore:

$$\begin{aligned}X_C &= -1/(6.28 \times 3.00 \times 0.000100)\\ &= -1/0.001884 = -531\ \Omega\end{aligned}$$

The imaginary value of X_C is equal to $-j531$. The susceptance, B_C, is equal to $1/X_C$. Thus, $B_C = 1/(-j531) = j0.00188$, rounded to three significant figures.

The general formula for capacitive susceptance in siemens, in terms of frequency in hertz and capacitance in farads, is:

$$B_C = 6.28fC$$

This formula also works for frequencies in megahertz and capacitances in microfarads.

Problem 15-6

Suppose an inductor has $L = 163$ μH at a frequency of 887 kHz. What is B_L?

Note that 887 kHz = 0.887 MHz. You can calculate X_L from the formula for inductive reactance:

$$\begin{aligned}X_L &= 6.28fL\\ &= 6.28 \times 0.887 \times 163\\ &= 908\ \Omega\end{aligned}$$

The imaginary value of X_L is equal to $j908$. The susceptance, $B_L =$ is equal to $1/X_L$. It follows that $B_L = -1/j908 = -j0.00110$.

The general formula for inductive susceptance in siemens, in terms of frequency in hertz and inductance in henrys, is:

$$B_L = -1/(6.28fL)$$

This formula also works for frequencies in kilohertz and inductances in millihenrys, and for frequencies in megahertz and inductances in microhenrys.

Admittance

Real-number conductance and imaginary-number susceptance combine to form *complex admittance*, symbolized by the capital letter Y. This is a complete expression of the extent to which a circuit allows ac to flow.

As the absolute value of complex impedance gets larger, the absolute value of complex admittance becomes smaller, in general. Huge impedances correspond to tiny admittances, and vice versa.

Admittances are written in complex form just like impedances. But you need to keep track of which quantity you're talking about! This will be obvious if you use the symbol, such as $Y = 3 - j0.5$ or $Y = 7 + j3$. When you see Y instead of Z, you know that negative j factors (such as in the quantity $3 - j0.5$) mean there is a net inductance in the circuit, and positive j factors (such as in the quantity $7 + j3$) mean there is net capacitance.

Admittance is the complex composite of conductance and susceptance. Thus, complex admittance values always take the form $Y = G + jB$. When the j factor is negative, a complex admittance may appear in the form $Y = G - jB$.

Do you remember how resistances combine with reactances in series to form complex impedances? In Chaps. 13 and 14, you saw series *RL* and *RC* circuits. Did you wonder why parallel circuits were ignored in those discussions? The reason was the fact that admittance, not impedance, is best for working with parallel ac circuits. Resistance and reactance combine in a messy fashion in parallel circuits. But conductance (G) and susceptance (B) merely add together in parallel circuits, yielding admittance (Y). Parallel circuit analysis is covered in detail in the next chapter.

The *GB* Plane

Admittance can be depicted on a plane similar to the complex impedance (*RX*) plane. Actually, it's a half plane, because there is ordinarily no such thing as negative conductance. (You can't have a component that conducts worse than not at all.) Conductance is plotted along the horizontal, or G, axis on this coordinate half plane, and susceptance is plotted along the B axis. The *GB plane* is shown in Fig. 15-9, with several points plotted.

It's Inside Out

The *GB* plane looks superficially identical to the *RX* plane. But mathematically, the two could not be more different! The *GB* plane is mathematically inside out with respect to the *RX* plane. The center, or origin, of the *GB* plane represents the point at which there is no conduction for dc or for ac. It is the zero-admittance point, rather than the zero-impedance point. In the *RX* plane, the origin represents a perfect short circuit, but in the *GB* plane, the origin corresponds to a perfect open circuit.

As you move out toward the right (east) along the G, or conductance, axis of the *GB* plane, the conductance improves, and the current gets greater. When you move upward (north) along the jB axis from the origin, you have ever-increasing positive (capacitive) susceptance. When you go down (south) along the jB axis from the origin, you encounter increasingly negative (inductive) susceptance.

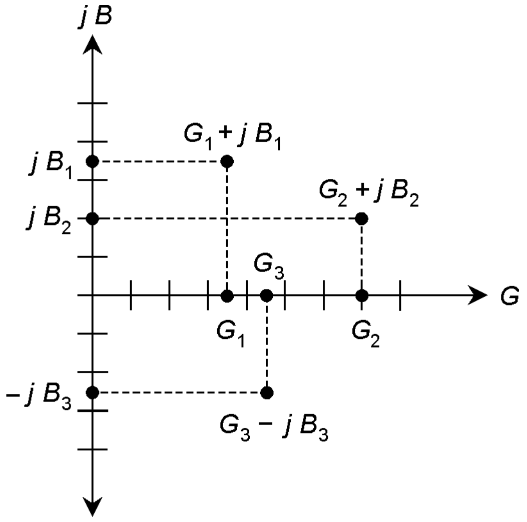

15-9 Some points in the complex admittance plane, and their conductive and susceptive components on the axes.

Vector Representation of Admittance

Complex admittances can be shown as vectors, just as can complex impedances. In Fig. 15-10, the points from Fig. 15-9 are rendered as vectors.

Generally, long vectors in the *GB* plane indicate large currents, and short vectors indicate small currents. Imagine a point moving around on the *GB* plane, and think of the vector getting longer and shorter and changing direction. Vectors pointing generally northeast, or upward and to the right, correspond to conductances and capacitances in parallel. Vectors pointing in a more or less southeasterly direction, or downward and to the right, are conductances and inductances in parallel.

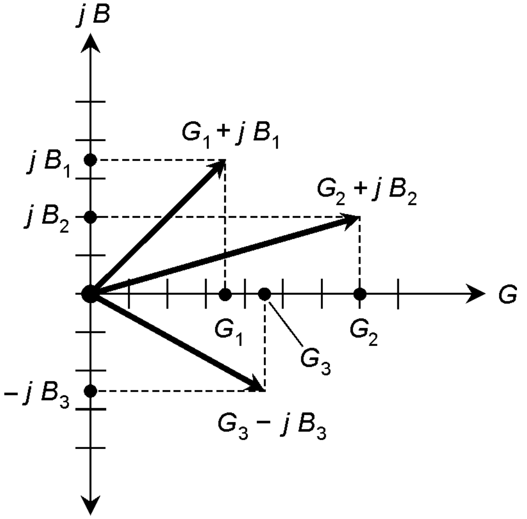

15-10 Vectors representing the points of Fig. 15-9.

Quiz

Refer to the text in this chapter if necessary. A good score is 18 or more correct. Answers are in the back of the book.

1. The square of an imaginary number
   (a) can never be negative.
   (b) can never be positive.
   (c) can be either positive or negative.
   (d) is equal to j.
2. A complex number
   (a) is the same thing as an imaginary number.
   (b) has a real-number part and an imaginary-number part.
   (c) is one-dimensional.
   (d) is a concept reserved for elite mathematicians.
3. What is the sum of $3 + j7$ and $-3 - j7$?
   (a) $0 + j0$
   (b) $6 + j14$
   (c) $-6 - j14$
   (d) $0 - j14$
4. What is $(-5 + j7) - (4 - j5)$?
   (a) $-1 + j2$
   (b) $-9 - j2$
   (c) $-1 - j2$
   (d) $-9 + j12$
5. What is the product $(-4 - j7)(6 - j2)$?
   (a) $24 - j14$
   (b) $-38 - j34$
   (c) $-24 - j14$
   (d) $-24 + j14$
6. What is the magnitude of the vector $18 - j24$?
   (a) 6
   (b) 21
   (c) 30
   (d) 52
7. The complex impedance value $5 + j0$ represents
   (a) a pure resistance.
   (b) a pure inductance.

(c) a pure capacitance.
(d) an inductance combined with a capacitance.

8. The complex impedance value $0 - j22$ represents
(a) a pure resistance.
(b) a pure inductance.
(c) a pure capacitance.
(d) an inductance combined with a resistance.

9. What is the absolute-value impedance of $3.0 - j6.0$?
(a) $Z = 9.0\ \Omega$
(b) $Z = 3.0\ \Omega$
(c) $Z = 45\ \Omega$
(d) $Z = 6.7\ \Omega$

10. What is the absolute-value impedance of $50 - j235$?
(a) $Z = 240\ \Omega$
(b) $Z = 58{,}000\ \Omega$
(c) $Z = 285\ \Omega$
(d) $Z = -185\ \Omega$

11. If the center conductor of a coaxial cable is made to have a smaller diameter, all other things being equal, what will happen to the Z_o of the transmission line?
(a) It will increase.
(b) It will decrease.
(c) It will not change.
(d) There is no way to determine this without knowing the actual dimensions.

12. If a device is said to have an impedance of $Z = 100\ \Omega$, you can reasonably expect that this indicates
(a) $R + jX = 100 + j0$.
(b) $R + jX = 0 + j100$.
(c) $R + jX = 100 + j100$.
(d) the reactance and the resistance add up to $100\ \Omega$.

13. Suppose a capacitor has a value of 0.050 μF at 665 kHz. What is the capacitive susceptance, stated as an imaginary number?
(a) $B_C = j4.79$
(b) $B_C = -j4.79$
(c) $B_C = j0.209$
(d) $B_C = -j0.209$

14. An inductor has a value of 44 mH at 60 Hz. What is the inductive susceptance, stated as an imaginary number?
    (a) $B_L = -j0.060$
    (b) $B_L = j0.060$
    (c) $B_L = -j17$
    (d) $B_L = j17$

15. Susceptance and conductance add to form
    (a) complex impedance.
    (b) complex inductance.
    (c) complex reactance.
    (d) complex admittance.

16. Absolute-value impedance is equal to the square root of which of the following?
    (a) $G^2 + B^2$
    (b) $R^2 + X^2$
    (c) Z_o
    (d) $Y^2 + R^2$

17. Inductive susceptance is defined in
    (a) imaginary ohms.
    (b) imaginary henrys.
    (c) imaginary farads.
    (d) imaginary siemens.

18. Capacitive susceptance values can be defined by
    (a) positive real numbers.
    (b) negative real numbers.
    (c) positive imaginary numbers.
    (d) negative imaginary numbers.

19. Which of the following is false?
    (a) $B_C = 1/X_C$.
    (b) Complex impedance can be depicted as a vector.
    (c) Characteristic impedance is complex.
    (d) $G = 1/R$.

20. In general, as the absolute value of the impedance in a circuit increases,
    (a) the flow of ac increases.
    (b) the flow of ac decreases.
    (c) the reactance decreases.
    (d) the resistance decreases.

16
CHAPTER

RLC and *GLC* Circuit Analysis

WHEN YOU SEE AN AC CIRCUIT THAT CONTAINS COILS AND/OR CAPACITORS, YOU SHOULD ENVISION a complex-number plane, either *RX* (resistance-reactance) or *GB* (conductance-admittance). The *RX* plane applies to series circuit analysis. The *GB* plane applies to parallel circuit analysis.

Complex Impedances in Series

When you see resistors, coils, and capacitors in series, each component has an impedance that can be represented as a vector in the *RX* plane. The vectors for resistors are constant, regardless of the frequency. But the vectors for coils and capacitors vary with frequency.

Pure Reactances

Pure inductive reactances (X_L) and capacitive reactances (X_C) simply add together when coils and capacitors are in series. Thus, $X = X_L + X_C$. In the *RX* plane, their vectors add, but because these vectors point in exactly opposite directions—inductive reactance upward and capacitive reactance downward (Fig. 16-1)—the resultant sum vector inevitably points either straight up or straight down, unless the reactances are equal and opposite, in which case they cancel and the result is the zero vector.

Problem 16-1

Suppose a coil and capacitor are connected in series, with $jX_L = j200$ and $jX_C = -j150$. What is the net reactance?

Just add the values: $jX = jX_L + jX_C = j200 + (-j150) = j(200 - 150) = j50$. This is a pure inductive reactance, because it is positive imaginary.

Problem 16-2

Suppose a coil and capacitor are connected in series, with $jX_L = j30$ and $jX_C = -j110$. What is the net reactance?

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

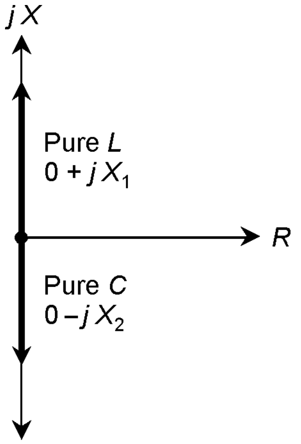

16-1 Pure inductance and pure capacitance are represented by reactance vectors that point straight up and down.

Again, add the values: $jX = j30 + (-j110) = j(30 - 110) = -j80$. This is a pure capacitive reactance, because it is negative imaginary.

Problem 16-3

Suppose a coil of inductance $L = 5.00$ µH and a capacitor of capacitance $C = 200$ pF are connected in series. Suppose the frequency is $f = 4.00$ MHz. What is the net reactance?

First, calculate the reactance of the inductor at 4.00 MHz. Proceed as follows:

$$\begin{aligned} jX_L &= j6.28fL \\ &= j(6.28 \times 4.00 \times 5.00) \\ &= j126 \end{aligned}$$

Next, calculate the reactance of the capacitor at 4.00 MHz. Proceed as follows:

$$\begin{aligned} jX_C &= -j[1/(6.28fC)] \\ &= -j[1/(6.28 \times 4.00 \times 0.000200)] \\ &= -j199 \end{aligned}$$

Finally, add the inductive and capacitive reactances to obtain the net reactance:

$$\begin{aligned} jX &= jX_L + jX_C \\ &= j126 + (-j199) \\ &= -j73 \end{aligned}$$

This is a pure capacitive reactance.

Problem 16-4

What is the net reactance of the aforementioned inductor and capacitor combination at the frequency $f = 10.0$ MHz?

First, calculate the reactance of the inductor at 10.0 MHz. Proceed as follows:

$$\begin{aligned} jX_L &= j6.28fL \\ &= j(6.28 \times 10.0 \times 5.00) \\ &= j314 \end{aligned}$$

Next, calculate the reactance of the capacitor at 10.00 MHz. Proceed as follows:

$$\begin{aligned} jX_C &= -j[1/(6.28fC)] \\ &= -j[1/(6.28 \times 10.0 \times 0.000200)] \\ &= -j79.6 \end{aligned}$$

Finally, add the inductive and capacitive reactances to obtain the net reactance:

$$\begin{aligned} jX &= jX_L + jX_C \\ &= j314 + (-j79.6) \\ &= j234 \end{aligned}$$

This is a pure inductive reactance. For series-connected components, the condition in which the capacitive and inductive reactances cancel is known as *series resonance.* We'll deal with this in more detail in the next chapter.

Adding Impedance Vectors

In the real world, there is resistance, as well as reactance, in an ac series circuit containing a coil and capacitor. This occurs because the coil wire has some resistance (it's never a perfect conductor). It can also be the case because a resistor is deliberately connected into the circuit.

Whenever the resistance in a series circuit is significant, the impedance vectors no longer point straight up and straight down. Instead, they run off toward the northeast (for the inductive part of the circuit) and southeast (for the capacitive part). This is illustrated in Fig. 16-2.

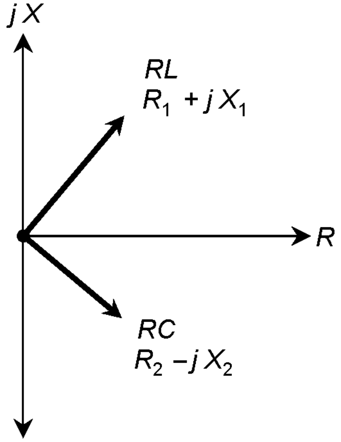

16-2 When resistance is present along with reactance, impedance vectors point at angles; they are neither vertical nor horizontal.

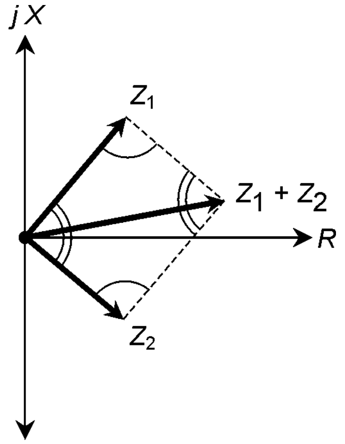

16-3 Parallelogram method of complex-impedance vector addition.

When two impedance vectors don't lie along a single line, you must use *vector addition* to be sure that you get the correct net impedance. In Fig. 16-3, the geometry of vector addition is shown. Construct a *parallelogram*, using the two vectors $Z_1 = R_1 + jX_1$ and $Z_2 = R_2 + jX_2$ as two adjacent sides of the figure. The diagonal of the parallelogram is the vector representing the net complex impedance. (Note that in a parallelogram, pairs of opposite angles have equal measures. These equalities are indicated by single and double arcs in Fig. 16-3.)

Formula for Complex Impedances in Series

Suppose you are given two complex impedances, $Z_1 = R_1 + jX_1$ and $Z_2 = R_2 + jX_2$. The net impedance, Z, of these in series is their vector sum, given by the following formula:

$$\begin{aligned} Z &= (R_1 + jX_1) + (R_2 + jX_2) \\ &= (R_1 + R_2) + j(X_1 + X_2) \end{aligned}$$

Calculating a vector sum using the formula is easier than doing it geometrically with a parallelogram. The arithmetic method is also more exact. The resistance and reactance components add separately. Just remember that if a reactance is capacitive, then it is negative imaginary in this formula.

Series *RLC* Circuits

When an inductance, capacitance, and resistance are connected in series (Fig. 16-4), the resistance R can be imagined as belonging entirely to the coil, when you use the preceding formulas. Then you have two vectors to add, when finding the impedance of the series *RLC* circuit containing three such components:

$$\begin{aligned} Z &= (R + jX_L) + (0 + jX_C) \\ &= R + j(X_L + X_C) \end{aligned}$$

Again, remember that X_C is never positive! So, although the formulas here have addition symbols in them, you're adding a negative number when you add in a capacitive reactance.

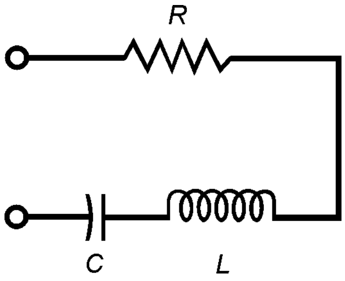

16-4 A series resistance-inductance-capacitance (*RLC*) circuit.

Problem 16-5

Suppose a resistor, a coil, and a capacitor are connected in series with $R = 50\ \Omega$, $X_L = 22\ \Omega$, and $X_C = -33\ \Omega$. What is the net impedance, Z?

Consider the resistor to be part of the coil, obtaining two complex vectors, $50 + j22$ and $0 - j33$. Adding these gives the resistance component of $50 + 0 = 50$, and the reactive component of $j22 - j33 = -j11$. Therefore, $Z = 50 - j11$.

Problem 16-6

Consider a resistor, a coil, and a capacitor that are connected in series with $R = 600\ \Omega$, $X_L = 444\ \Omega$, and $X_C = -444\ \Omega$. What is the net impedance, Z?

Again, imagine the resistor to be part of the inductor. Then the complex impedance vectors are $600 + j444$ and $0 - j444$. Adding these, the resistance component is $600 + 0 = 600$, and the reactive component is $j444 - j444 = j0$. Thus, $Z = 600 + j0$. This is a purely resistive impedance, and you can rightly call it 600 Ω.

Problem 16-7

Suppose a resistor, a coil, and a capacitor are connected in series. The resistor has a value of 330 Ω, the capacitance is 220 pF, and the inductance is 100 μH. The frequency is 7.15 MHz. What is the complex impedance of this series *RLC* circuit at this frequency?

First, calculate the inductive reactance. Remember that $X_L = 6.28fL$ and that megahertz and microhenrys go together in the formula. Multiply to obtain the following:

$$\begin{aligned} jX_L &= j(6.28 \times 7.15 \times 100) \\ &= j4490 \end{aligned}$$

Next, calculate the capacitive reactance using the formula $X_C = -1/(6.28fC)$. Convert 220 pF to microfarads to obtain $C = 0.000220$ μF. Then calculate:

$$\begin{aligned} jX_C &= -j[1/(6.28 \times 7.15 \times 0.000220)] \\ &= -j101 \end{aligned}$$

Now, lump the resistance and the inductive reactance together, so one of the impedance vectors is $330 + j4490$. The other is $0 - j101$. Adding these gives $Z = 330 + j4389$; this rounds off to $Z = 330 + j4390$.

Problem 16-8

Suppose a resistor, a coil, and a capacitor are connected in series. The resistance is 50.0 Ω, the inductance is 10.0 μH, and the capacitance is 1000 pF. The frequency is 1592 kHz. What is the complex impedance of this series *RLC* circuit at this frequency?

First, calculate $X_L = 6.28fL$. Convert the frequency to megahertz; 1592 kHz = 1.592 MHz. Then:

$$\begin{aligned} jX_L &= j(6.28 \times 1.592 \times 10.0) \\ &= j100 \end{aligned}$$

Then calculate $X_C = -1/(6.28fC)$. Let's convert picofarads to microfarads, and use megahertz for the frequency. Therefore:

$$\begin{aligned} jX_C &= -j[1/(6.28 \times 1.592 \times 0.001000)] \\ &= -j100 \end{aligned}$$

Let the resistance and inductive reactance go together as one vector, $50.0 + j100$. Let the capacitive reactance be represented as $0 - j100$. The sum is $Z = 50.0 + j100 - j100 = 50.0 + j0$. This is a pure resistance of 50.0 Ω. You can correctly say that the impedance is 50.0 Ω in this case.

Complex Admittances in Parallel

When you see resistors, coils, and capacitors in parallel, remember that each component, whether it is a resistor, an inductor, or a capacitor, has an admittance that can be represented as a vector in the *GB* plane. The vectors for pure conductances are constant, even as the frequency changes. But the vectors for the coils and capacitors vary with frequency.

Pure Susceptances

Pure inductive susceptances (B_L) and capacitive susceptances (B_C) add together when coils and capacitors are in parallel. Thus, $B = B_L + B_C$. Remember that B_L is never positive, and B_C is never negative. This is just the opposite situation from reactances.

In the *GB* plane, pure jB_L and jB_C vectors add. Because such vectors always point in exactly opposite directions—inductive susceptance down and capacitive susceptance up—the sum, jB, inevitably points either straight down or straight up (Fig. 16-5), unless the susceptances are equal and opposite, in which case they cancel and the result is the zero vector.

Problem 16-9

Suppose a coil and capacitor are connected in parallel, with $jB_L = -j0.05$ and $jB_C = j0.08$. What is the net susceptance?

Just add the values as follows: $jB = jB_L + jB_C = -j0.05 + j0.08 = j0.03$. This is a capacitive susceptance, because it is positive imaginary.

Problem 16-10

Suppose a coil and capacitor are connected in parallel, with $jB_L = -j0.60$ and $jB_C = j0.25$. What is the net susceptance?

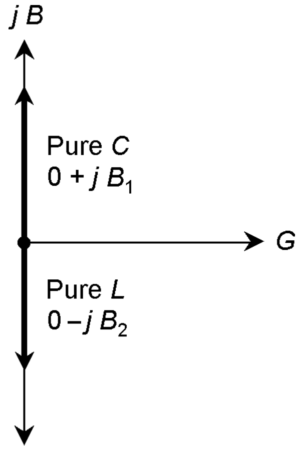

16-5 Pure capacitance and pure inductance are represented by susceptance vectors that point straight up and down.

Again, add the values: $jB = -j0.60 + j0.25 = -j0.35$. This is an inductive susceptance, because it is negative imaginary.

Problem 16-11

Suppose a coil of $L = 6.00$ μH and a capacitor of $C = 150$ pF are connected in parallel. The frequency is $f = 4.00$ MHz. What is the net susceptance?

First calculate the susceptance of the inductor at 4.00 MHz, as follows:

$$\begin{aligned} jB_L &= -j[1/(6.28fL)] \\ &= -j[1/(6.28 \times 4.00 \times 6.00)] \\ &= -j0.00663 \end{aligned}$$

Next, calculate the susceptance of the capacitor (converting its value to microfarads) at 4.00 MHz, as follows:

$$\begin{aligned} jB_C &= j(6.28fC) \\ &= j(6.28 \times 4.00 \times 0.000150) \\ &= j0.00377 \end{aligned}$$

Finally, add the inductive and capacitive susceptances to obtain the net susceptance:

$$\begin{aligned} jB &= jB_L + jB_C \\ &= -j0.00663 + j0.00377 \\ &= -j0.00286 \end{aligned}$$

This is a pure inductive susceptance.

Problem 16-12

What is the net susceptance of the above parallel-connected inductor and capacitor at a frequency of $f = 5.31$ MHz?

First calculate the susceptance of the inductor at 5.31 MHz, as follows:

$$\begin{aligned} jB_L &= -j[1/(6.28fL)] \\ &= -j[1/(6.28 \times 5.31 \times 6.00)] \\ &= -j0.00500 \end{aligned}$$

Next calculate the susceptance of the capacitor (converting its value to microfarads) at 5.31 MHz, as follows:

$$\begin{aligned} jB_C &= \mathrm{j}(6.28fC) \\ &= j(6.28 \times 5.31 \times 0.000150) \\ &= j0.00500 \end{aligned}$$

Finally, add the inductive and capacitive susceptances to obtain the net susceptance:

$$\begin{aligned} jB &= jB_L + jB_C \\ &= -j0.00500 + j0.00500 \\ &= j0 \end{aligned}$$

This means that the circuit has no susceptance at 5.31 MHz. The situation in which there is no susceptance in an *LC* circuit is known as *parallel resonance.* It is discussed in the next chapter.

Adding Admittance Vectors

In real life, there is a small amount of conductance, as well as susceptance, in an ac parallel circuit containing a coil and capacitor. This occurs when the capacitor lets a little bit of current leak through. More often, though, it is the case because a *load* is connected in parallel with the coil and capacitor. This load can be an antenna, the input to an amplifier circuit, a test instrument, a transducer, or some other device.

When the conductance in a parallel circuit containing inductance and capacitance is significant, the admittance vectors do not point straight up and down. Instead, they run off toward the northeast (for the capacitive part of the circuit) and southeast (for the inductive part). This is illustrated in Fig. 16-6.

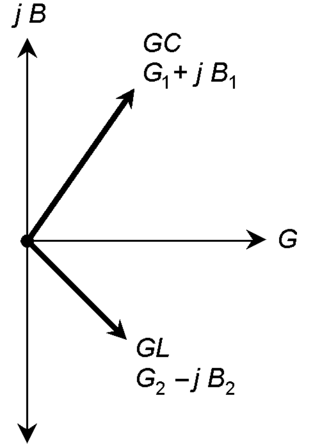

16-6 When conductance is present along with susceptance, admittance vectors point at angles; they are neither vertical nor horizontal.

You've seen how vectors add in the *RX* plane. In the *GB* plane, the principle is the same. The net admittance vector is the sum of the component admittance vectors.

Formula for Complex Admittances in Parallel

Given two admittances, $Y_1 = G_1 + jB_1$ and $Y_2 = G_2 + jB_2$, the net admittance Y of these in parallel is their vector sum, as follows:

$$\begin{aligned} Y &= (G_1 + jB_1) + (G_2 + jB_2) \\ &= (G_1 + G_2) + j(B_1 + B_2) \end{aligned}$$

The conductance and susceptance components add separately. Just remember that if a susceptance is inductive, then it is negative imaginary in this formula.

Parallel *GLC* Circuits

When a coil, capacitor, and resistor are connected in parallel (Fig. 16-7), the resistance should be thought of as a *conductance*, whose value in siemens (symbolized S) is equal to the reciprocal of the value in ohms. Think of the conductance as all belonging to the inductor. Then you have two vectors to add, when finding the admittance of a parallel *GLC* (conductance-inductance-capacitance) circuit:

$$\begin{aligned} Y &= (G + jB_L) + (0 + jB_C) \\ &= G + j(B_L + B_C) \end{aligned}$$

Again, remember that B_L is never positive! So, although the formulas here have addition symbols in them, you're adding a negative number when you add in an inductive susceptance.

Problem 16-13

Suppose a resistor, a coil, and a capacitor are connected in parallel. Suppose the resistor has a conductance $G = 0.10$ S, and the susceptances are $jB_L = -j0.010$ and $jB_C = j0.020$. What is the complex admittance of this combination?

Consider the resistor to be part of the coil. Then there are two complex admittances in parallel: $0.10 - j0.010$ and $0.00 + j0.020$. Adding these gives a conductance component of $0.10 + 0.00 = 0.10$ and a susceptance component of $-j0.010 + j0.020 = j0.010$. Therefore, the complex admittance is $0.10 + j0.010$.

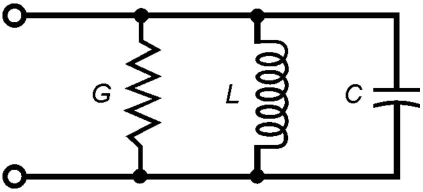

16-7 A parallel conductance-inductance-capacitance (*GLC*) circuit.

Problem 16-14

Suppose a resistor, a coil, and a capacitor are connected in parallel. Suppose the resistor has a conductance $G = 0.0010$ S, and the susceptances are $jB_L = -j0.0022$ and $jB_C = j0.0022$. What is the complex admittance of this combination?

Again, consider the resistor to be part of the coil. Then the complex admittances are $0.0010 - j0.0022$ and $0.0000 + j0.0022$. Adding these, the conductance component is $0.0010 + 0.0000 = 0.0010$, and the susceptance component is $-j0.0022 + j0.0022 = j0$. Thus, the admittance is $0.0010 + j0$. This is a purely conductive admittance.

Problem 16-15

Suppose a resistor, a coil, and a capacitor are connected in parallel. The resistor has a value of 100 Ω, the capacitance is 200 pF, and the inductance is 100 μH. The frequency is 1.00 MHz. What is the net complex admittance?

First, you need to calculate the inductive susceptance. Recall the formula, and plug in the numbers as follows:

$$\begin{aligned} jB_L &= -j[1/(6.28fL)] \\ &= -j[1/(6.28 \times 1.00 \times 100)] \\ &= -j0.00159 \end{aligned}$$

Megahertz and microhenrys go together in the formula. Next, you must calculate the capacitive susceptance. Convert 200 pF to microfarads to go with megahertz in the formula; thus $C = 0.000200$ μF. Then:

$$\begin{aligned} jB_C &= j(6.28fC) \\ &= j(6.28 \times 1.00 \times 0.000200) \\ &= j0.00126 \end{aligned}$$

Finally, consider the conductance, which is $^1/_{100} = 0.0100$ S, and the inductive susceptance as existing together in a single component. That means that one of the parallel-connected admittances is $0.0100 - j0.00159$. The other is $0.0000 + j0.00126$. Adding these gives $0.0100 - j0.00033$.

Problem 16-16

Suppose a resistor, a coil, and a capacitor are in parallel. The resistance is 10.0 Ω, the inductance is 10.0 μH, and the capacitance is 1000 pF. The frequency is 1592 kHz. What is the complex admittance of this circuit at this frequency?

First, calculate the inductive susceptance. Convert the frequency to megahertz; 1592 kHz = 1.592 MHz. Plug in the numbers as follows:

$$\begin{aligned} jB_L &= -j[1/(6.28fL)] \\ &= -j[1/(6.28 \times 1.592 \times 10.0)] \\ &= -j0.0100 \end{aligned}$$

Next, calculate the capacitive susceptance. Convert 1000 pF to microfarads to go with megahertz in the formula; thus $C = 0.001000$ μF. Then:

$$\begin{aligned} jB_C &= j(6.28fC) \\ &= j(6.28 \times 1.592 \times 0.001000) \\ &= j0.0100 \end{aligned}$$

Finally, consider the conductance, which is 1/10.0 = 0.100 S, and the inductive susceptance as existing together in a single component. That means that one of the parallel-connected admittances is $0.100 - j0.0100$. The other is $0.0000 + j0.0100$. Adding these gives $0.100 + j0$.

Converting Complex Admittance to Complex Impedance

The *GB* plane is, as you have seen, similar in appearance to the *RX* plane, although mathematically they are different. Once you've found a complex admittance for a parallel *RLC* circuit, you will usually want to transform this back to a complex impedance.

The transformation from a complex admittance $G + jB$ to a complex impedance $R + jX$ can be carried out using the following two formulas, one for R and the other for X:

$$\begin{aligned} R &= G/(G^2 + B^2) \\ X &= -B/(G^2 + B^2) \end{aligned}$$

If you know the complex admittance, first find the resistance and reactance components individually using the preceding formulas. Then assemble the two components into the complex impedance, $R + jX$.

Problem 16-17

Suppose the complex admittance of a certain parallel circuit is $0.010 - j0.0050$. What is the complex impedance of this same circuit, assuming the frequency does not change?

In this case, $G = 0.010$ S and $B = -0.0050$ S. First find $G^2 + B^2$, as follows:

$$\begin{aligned} G^2 + B^2 &= 0.010^2 + (-0.0050)^2 \\ &= 0.000100 + 0.000025 \\ &= 0.000125 \end{aligned}$$

Now it is easy to calculate R and X, like this:

$$\begin{aligned} R &= G/0.000125 \\ &= 0.010/0.000125 \\ &= 80\ \Omega \end{aligned}$$

$$\begin{aligned} X &= -B/0.000125 \\ &= 0.0050/0.000125 \\ &= 40\ \Omega \end{aligned}$$

The complex impedance is therefore $80 + j40$.

Putting It All Together

When you're confronted with a parallel circuit containing resistance, inductance, and capacitance, and you want to determine the complex impedance of the combination, do these things:

1. Find the conductance $G = 1/R$ for the resistor. (It will be positive or zero.)
2. Find the susceptance B_L of the inductor using the appropriate formula. (It will be negative or zero.)
3. Find the susceptance B_C of the capacitor using the appropriate formula. (It will be positive or zero.)
4. Find the net susceptance $B = B_L + B_C$. (It might be positive, negative, or zero.)
5. Compute R and X in terms of G and B using the appropriate formulas.
6. Assemble the complex impedance $R + jX$.

Problem 16-18

Suppose a resistor of 10.0 Ω, a capacitor of 820 pF, and a coil of 10.0 μH are in parallel. The frequency is 1.00 MHz. What is the complex impedance?

Proceed according to the above steps, as follows:

1. Calculate $G = 1/R = 1/10.0 = 0.100$.
2. Calculate $B_L = -1/(6.28fL) = -1/(6.28 \times 1.00 \times 10.0) = -0.0159$.
3. Calculate $B_C = 6.28fC = 6.28 \times 1.00 \times 0.000820 = 0.00515$. (Remember to first convert the capacitance to microfarads, to go with megahertz.)
4. Calculate $B = B_L + B_C = -0.0159 + 0.00515 = -0.0108$.
5. Define $G^2 + B^2 = 0.100^2 + (-0.0108)^2 = 0.010117$. Then $R = G/0.010117 = 0.100/0.010117 = 9.88\ \Omega$, and $X = -B/0.010117 = 0.0108/0.010117 = 1.07\ \Omega$.
6. The complex impedance is $R + jX = 9.88 + j1.07$.

Problem 16-19

Suppose a resistor of 47.0 Ω, a capacitor of 500 pF, and a coil of 10.0 μH are in parallel. What is their complex impedance at a frequency of 2.252 MHz?

Proceed as before:

1. Calculate $G = 1/R = 1/47.0 = 0.021277$.
2. Calculate $B_L = -1/(6.28fL) = -1/(6.28 \times 2.252 \times 10.0) = -0.00707$.
3. Calculate $B_C = 6.28fC = 6.28 \times 2.252 \times 0.000500 = 0.00707$. (Remember to first convert the capacitance to microfarads, to go with megahertz.)
4. Calculate $B = B_L + B_C = -0.00707 + 0.00707 = 0.00000$.
5. Define $G^2 + B^2 = 0.021277^2 + 0.00000^2 = 0.00045271$. Then $R = G/0.00045271 = 0.021277/0.00045271 = 46.999\ \Omega$, and $X = -B/0.00045271 = 0.00000/0.00045271 = 0.00000$.
6. The complex impedance is $R + jX = 46.9999 + j0.00000$. When we round it off to three significant figures, we get $47.0 + j0.00$. This a pure resistance equal to the value of the resistor in the circuit.

Reducing Complicated *RLC* Circuits

Sometimes you'll see circuits in which there are several resistors, capacitors, and/or coils in series and parallel combinations. Such a circuit can be reduced to an equivalent series or parallel *RLC* circuit that contains one resistance, one capacitance, and one inductance.

Series Combinations

Resistances in series simply add. Inductances in series also add. Capacitances in series combine in a somewhat more complicated way. If you don't remember the formula, here it is:

$$1/C = 1/C_1 + 1/C_2 + \cdots + 1/C_n$$

where C_1, C_2, . . . , and C_n are the individual capacitances, and C is the total capacitance. Once you've found $1/C$, take its reciprocal to obtain C. Figure 16-8A shows an example of a complicated series *RLC* circuit. The equivalent circuit, with one resistance, one capacitance, and one inductance, is shown in Fig. 16-8B.

Parallel Combinations

In parallel, resistances and inductances combine the way capacitances do in series. Capacitances simply add up. An example of a complicated parallel *RLC* circuit is shown in Fig. 16-9A. The equivalent circuit, with one resistance, one capacitance, and one inductance, is shown in Fig. 16-9B.

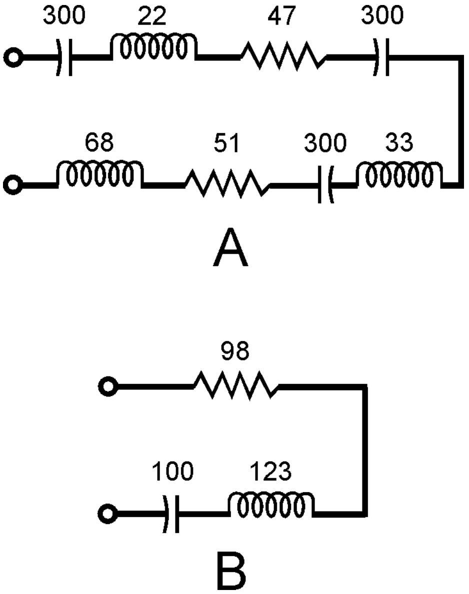

16-8 At A, a complicated series circuit containing multiple resistances and reactances. At B, the same circuit simplified. Resistances are in ohms; inductances are in microhenrys (μH); capacitances are in picofarads (pF).

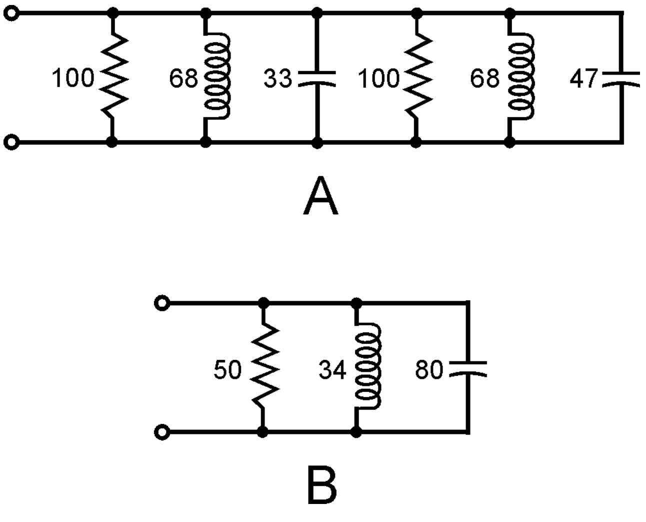

16-9 At A, a complicated parallel circuit containing multiple resistances and reactances. At B, the same circuit simplified. Resistances are in ohms; inductances are in microhenrys (μH); capacitances are in picofarads (pF).

Nightmare Scenarios

Imagine an *RLC* circuit like the one shown in Fig. 16-10. How would you find the complex impedance of this circuit at some particular frequency, such as 8.54 MHz? Don't waste much time worrying about circuits like this. You'll rarely encounter them. But rest assured that, given a frequency, a complex impedance does exist, no matter how complicated an *RLC* circuit happens to be.

An engineer could use a computer to find the theoretical complex impedance of a circuit such as the one in Fig. 16-10 at a specific frequency, or as a function of the frequency. The experimental approach would be to build the circuit, connect a signal generator to it, and then measure *R* and *X* at various frequencies with a device called an *impedance bridge.*

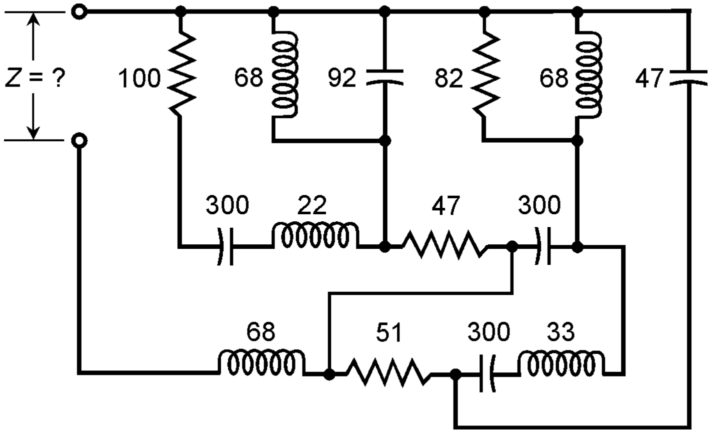

16-10 A series-parallel nightmare circuit containing multiple resistances and reactances. Resistances are in ohms; inductances are in microhenrys (μH); capacitances are in picofarads (pF).

Ohm's Law for AC Circuits

Ohm's Law for a dc circuit is a simple relationship among three variables: the current I (in amperes), the voltage E (in volts), and the resistance R (in ohms). Here are the formulas, in case you don't recall them:

$$E = IR$$
$$I = E/R$$
$$R = E/I$$

In ac circuits containing no reactance, these same formulas apply, as long as you work with root-mean-square (rms) voltages and currents. If you need a refresher concerning the meaning of rms, refer to Chapter 9.

Purely Resistive Impedances

When the impedance Z in an ac circuit contains no reactance, so that all of the current and voltage exist through and across a pure resistance R, Ohm's Law for an ac circuit is expressed as follows:

$$E = IZ$$
$$I = E/Z$$
$$Z = E/I$$

where $Z = R$, and the values I and E are rms current and voltage.

Complex Impedances

When you want to determine the relationship among current, voltage, and resistance in an ac circuit that contains resistance and reactance, things get interesting. Recall the formula for the square of the absolute-value impedance in a series RLC circuit:

$$Z^2 = R^2 + X^2$$

This means that Z is equal to the square root of the quantity $R^2 + X^2$, as follows:

$$Z = (R^2 + X^2)^{1/2}$$

This is the length of the vector $R + jX$ in the complex impedance plane. You learned this in Chap. 15. This formula applies only for series RLC circuits.

The square of the absolute-value impedance for a parallel RLC circuit, in which the resistance is R and the reactance is X, is defined this way:

$$Z^2 = R^2X^2/(R^2 + X^2)$$

This means that the absolute-value impedance, Z, must be calculated using the rather arcane formula:

$$Z = [R^2X^2/(R^2 + X^2)]^{1/2}$$

The $\frac{1}{2}$ power of a quantity represents the positive square root of that quantity.

Problem 16-20

Suppose a series *RX* circuit (shown by the generic block diagram of Fig. 16-11) has a resistance of $R = 50.0\ \Omega$ and a capacitive reactance of $X = -50.0\ \Omega$. Suppose 100-V rms ac is applied to this circuit. What is the current?

First, calculate $Z^2 = R^2 + X^2 = 50.0^2 + (-50.0)^2 = 2500 + 2500 = 5000$. Then Z is the square root of 5000, or 70.7. Therefore, $I = E/Z = 100/70.7 = 1.41$ A rms.

Problem 16-21

What are the rms ac voltages across the resistance and the reactance, respectively, in the circuit described in Problem 16-20?

The Ohm's Law formulas for dc will work here. Because the current is $I = 1.41$ A rms, the voltage drop across the resistance is equal to $E_R = IR = 1.41 \times 50.0 = 70.5$ V rms. The voltage drop across the reactance is the product of the current and the reactance: $E_X = IX = 1.41 \times (-50.0) = -70.5$ V rms. This is an rms ac voltage of equal magnitude to that across the resistance. But the phase is different.

Note that voltages across the resistance and the reactance—a capacitive reactance in this case, because it's negative—don't add up to 100 V rms, which is placed across the whole circuit. This is because, in an *RX* ac circuit, there is always a difference in phase between the voltage across the resistance and the voltage across the reactance. The voltages across the components always add up to the applied voltage *vectorially*, but not always *arithmetically.*

Problem 16-22

Suppose a series *RX* circuit (Fig. 16-11) has $R = 10.0\ \Omega$. and $X = 40.0\ \Omega$. The applied voltage is 100-V rms ac. What is the current?

Calculate $Z^2 = R^2 + X^2 = 100 + 1600 = 1700$. This means that Z is the square root of 1700, or 41.2. Therefore, $I = E/Z = 100/41.2 = 2.43$ A rms.

Problem 16-23

What are the rms ac voltages across the resistance and the reactance, respectively, in the circuit described in Problem 16-22?

Knowing the current, calculate $E_R = IR = 2.43 \times 10.0 = 24.3$ V rms. Also, $E_X = IX = 2.43 \times 40.0 = 97.2$ V rms. If you add $E_R + E_X$ arithmetically, you get $24.3 + 97.2 = 121.5$ V as the total

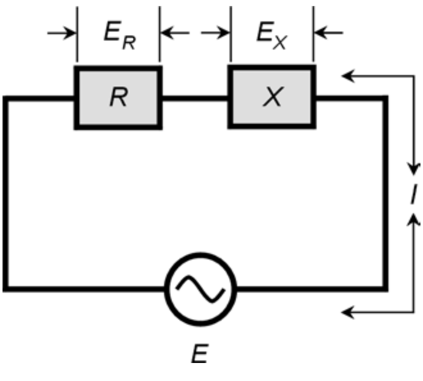

16-11 A series circuit containing resistance and reactance. Illustration for Problems 16-20 through 16-23.

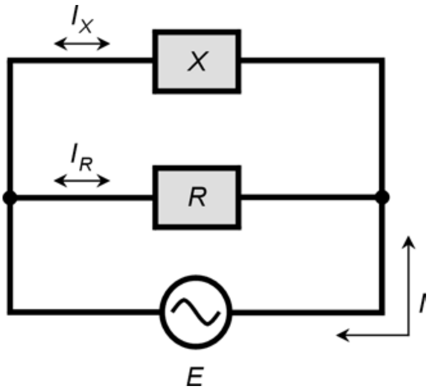

16-12 A parallel circuit containing resistance and reactance. Illustration for Problems 16-24 and 16-25.

across R and X. Again, this differs from the applied voltage! The simple dc rule does not work here, for the same reason it didn't work in the scenario of Problem 16-21.

Problem 16-24

Suppose a parallel RX circuit (shown by the generic block diagram of Fig. 16-12) has $R = 30.0\ \Omega$ and $X = -20.0\ \Omega$. The ac supply voltage is 50.0 V rms. What is the total current drawn from the ac supply?

First, find the square of the absolute-value impedance, remembering the formula for parallel circuits: $Z^2 = R^2X^2/(R^2 + X^2) = 360{,}000/1300 = 277$. The absolute-value impedance Z is the square root of 277, or 16.6. The total current is therefore $I = E/Z = 50/16.6 = 3.01$ A rms.

Problem 16-25

What are the rms currents through the resistance and the reactance, respectively, in the circuit described in Problem 16-24?

The Ohm's Law formulas for dc will work here. For the resistance, $I_R = E/R = 50.0/30.0 = 1.67$ A rms. For the reactance, $I_X = E/X = 50.0/(-20.0) = -2.5$ A rms. Note that these currents don't add up to 3.01 A, the total current. The reason for this is the same as the reason ac voltages don't add arithmetically in ac circuits that contain reactance. The constituent currents, I_R and I_X, differ in phase. Vectorially, they add up to 3.01 A rms, but arithmetically, they don't.

Quiz

Refer to the text in this chapter if necessary. A good score is 18 correct. Answers are in the back of the book.

1. Suppose a coil and capacitor are connected in series. The inductive reactance is 250 Ω, and the capacitive reactance is –300 Ω. What is the complex impedance?
   - (a) $0 + j550$
   - (b) $0 - j50$
   - (c) $250 - j300$
   - (d) $-300 + j250$

2. Suppose a coil of 25.0 μH and capacitor of 100 pF are connected in series. The frequency is 5.00 MHz. What is the complex impedance?
   (a) $0 + j467$
   (b) $25 + j100$
   (c) $0 - j467$
   (d) $25 - j100$
3. When $R = 0$ in a series *RLC* circuit, but the net reactance is not zero, the impedance vector
   (a) always points straight up.
   (b) always points straight down.
   (c) always points straight toward the right.
   (d) None of the above is correct.
4. Suppose a resistor of 150 Ω, a coil with a reactance of 100 Ω, and a capacitor with a reactance of –200 Ω are connected in series. What is the complex impedance?
   (a) $150 + j100$
   (b) $150 - j200$
   (c) $100 - j200$
   (d) $150 - j100$
5. Suppose a resistor of 330 Ω, a coil of 1.00 μH, and a capacitor of 200 pF are in series. What is the complex impedance at 10.0 MHz?
   (a) $330 - j199$
   (b) $300 + j201$
   (c) $300 + j142$
   (d) $330 - j16.8$
6. Suppose a coil has an inductance of 3.00 μH and a resistance of 10.0 Ω in its winding. A capacitor of 100 pF is in series with this coil. What is the complex impedance at 10.0 MHz?
   (a) $10 + j3.00$
   (b) $10 + j29.2$
   (c) $10 - j97$
   (d) $10 + j348$
7. Suppose a coil has a reactance of 4.00 Ω. What is the complex admittance, assuming there is nothing else is in the circuit?
   (a) $0 + j0.25$
   (b) $0 + j4.00$
   (c) $0 - j0.25$
   (d) $0 - j4.00$
8. What will happen to the susceptance of a capacitor if the frequency is doubled and all other factors remain constant?
   (a) It will decrease to half its former value.
   (b) It will not change.

(c) It will double.
(d) It will quadruple.

9. Suppose a coil and capacitor are in parallel, with $jB_L = -j0.05$ and $jB_C = j0.03$. What is the complex admittance, assuming that nothing is in series or parallel with these components?

(a) $0 - j0.02$
(b) $0 - j0.07$
(c) $0 + j0.02$
(d) $-0.05 + j0.03$

10. Imagine a coil, a resistor, and a capacitor connected in parallel. The resistance is 1.0 Ω, the capacitive susceptance is 1.0 S, and the inductive susceptance is −1.0 S. Then, suddenly, the frequency is cut to half its former value. What is the complex admittance at the new frequency?

(a) $1.0 + j0.0$
(b) $1.0 + j1.5$
(c) $1.0 - j1.5$
(d) $1.0 - j2.0$

11. Suppose a coil of 3.50 μH and a capacitor of 47.0 pF are in parallel. The frequency is 9.55 MHz. There is nothing else in series or parallel with these components. What is the complex admittance?

(a) $0 + j0.00282$
(b) $0 - j0.00194$
(c) $0 + j0.00194$
(d) $0 - j0.00758$

12. A vector pointing southeast in the *GB* plane would indicate

(a) pure conductance with zero susceptance.
(b) conductance and inductive susceptance.
(c) conductance and capacitive susceptance.
(d) pure susceptance with zero conductance.

13. Suppose a resistor with conductance 0.0044 S, a capacitor with susceptance 0.035 S, and a coil with susceptance −0.011 S are all connected in parallel. What is the complex admittance?

(a) $0.0044 + j0.024$
(b) $0.035 - j0.011$
(c) $-0.011 + j0.035$
(d) $0.0044 + j0.046$

14. Suppose a resistor of 100 Ω, a coil of 4.50 μH, and a capacitor of 220 pF are in parallel. What is the complex admittance at a frequency of 6.50 MHz?

(a) $100 + j0.00354$
(b) $0.010 + j0.00354$
(c) $100 - j0.0144$
(d) $0.010 + j0.0144$

15. Suppose the complex admittance of a circuit is $0.02 + j0.20$. What is the complex impedance, assuming the frequency does not change?

    (a) $50 + j5.0$
    (b) $0.495 - j4.95$
    (c) $50 - j5.0$
    (d) $0.495 + j4.95$

16. Suppose a resistor of 51.0 Ω, an inductor of 22.0 μH, and a capacitor of 150 pF are in parallel. The frequency is 1.00 MHz. What is the complex impedance?

    (a) $51.0 - j14.9$
    (b) $51.0 + j14.9$
    (c) $46.2 - j14.9$
    (d) $46.2 + j14.9$

17. Suppose a series circuit has 99.0 Ω of resistance and 88.0 Ω of inductive reactance. An ac rms voltage of 117 V is applied to this series network. What is the current?

    (a) 1.18 A
    (b) 1.13 A
    (c) 0.886 A
    (d) 0.846 A

18. What is the voltage across the reactance in the preceding example?

    (a) 78.0 V
    (b) 55.1 V
    (c) 99.4 V
    (d) 74.4 V

19. Suppose a parallel circuit has 10 Ω of resistance and 15 Ω of reactance. An ac rms voltage of 20 V is applied across it. What is the total current?

    (a) 2.00 A
    (b) 2.40 A
    (c) 1.33 A
    (d) 0.800 A

20. What is the current through the resistance in the preceding example?

    (a) 2.00 A
    (b) 2.40 A
    (c) 1.33 A
    (d) 0.800 A

17
CHAPTER

Power and Resonance in Alternating-Current Circuits

ONE OF THE BIGGEST CHALLENGES IN ELECTRICITY AND ELECTRONICS IS OPTIMIZING THE EFFICIENCY with which power is transferred from one place to another, or converted from one form to another. Also important, especially for the radio-frequency (RF) engineer, is the phenomenon of *resonance.* Power and resonance are closely related.

Forms of Power

What is power, exactly? Here is an all-encompassing definition: *Power is the rate at which energy is expended, radiated, or dissipated.* This definition can be applied to mechanical motion, chemical effects, dc and ac electricity, sound waves, radio waves, sound, heat, infrared (IR), visible light, ultraviolet (UV), X rays, gamma rays, and high-speed subatomic particles. In all cases, the energy is converted from one form into another form at a certain rate.

Units of Power

The standard unit of power is the *watt*, abbreviated W. A watt is equivalent to a *joule per second* (J/s). Sometimes power is given as *kilowatts* (kW or thousands of watts), *megawatts* (MW or millions of watts), or *gigawatts* (GW or billions of watts). It is also sometimes expressed as *milliwatts* (mW or thousandths of watts), *microwatts* (µW or millionths of watts), or *nanowatts* (nW or billionths of watts).

Volt-Amperes

In dc circuits, and also in ac circuits having no reactance, power can be defined this way: *Power is the product of the voltage across a circuit or component and the current through that same circuit or component.* Mathematically this is written $P = EI$. If E is in volts and I is in amperes, then P is in *volt-amperes* (VA). This translates into watts when there is no reactance in the circuit (Fig. 17-1). The root-mean-square (rms) values for voltage and current are always used to derive the effective, or average, power.

Volt-amperes, also called *VA power* or *apparent power*, can take various forms. A resistor converts electrical energy into heat energy, at a rate that depends on the value of the resistance and the current through it. A light bulb converts electricity into light and heat. A radio antenna converts high-

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

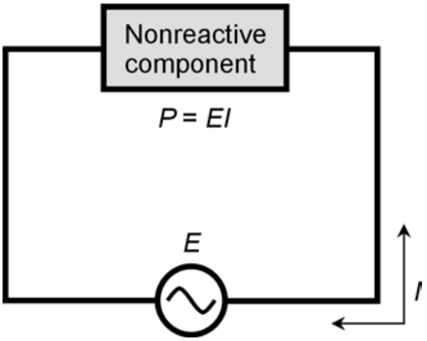

17-1 When there is no reactance in an ac component, the power P is the product of the voltage E across the component and the current I through the component.

frequency ac into radio waves. A speaker converts low-frequency ac into sound waves. The power in these forms is a measure of the intensity of the heat, light, radio waves, or sound waves.

Instantaneous Power

Usually, but not always, engineers think of power based on the rms, or effective, ac value. But for VA power, peak values are sometimes used instead. If the ac is a sine wave, the peak current is 1.414 times the rms current, and the peak voltage is 1.414 times the rms voltage. If the current and the voltage are exactly in phase, the product of their peak values is twice the product of their rms values.

There are instants in time when the VA power in a reactance-free, sine-wave ac circuit is twice the effective power. There are other instants in time when the VA power is zero; at still other moments, the VA power is somewhere between zero and twice the effective power level (Fig. 17-2). This constantly changing power is called *instantaneous power.*

In some situations, such as with a voice-modulated radio signal or a fast-scan television signal, the instantaneous power varies in an extremely complicated fashion. Have you ever seen the *modulation envelope* of such a signal displayed on an oscilloscope?

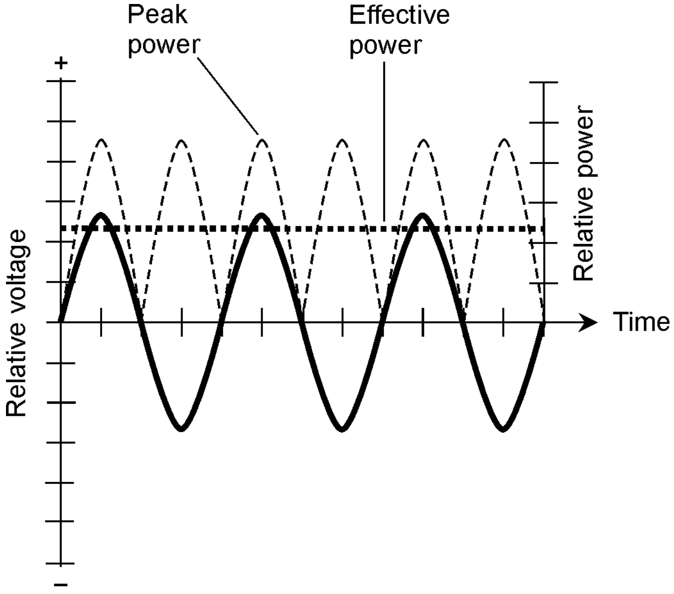

17-2 Peak versus effective power for a sine wave. The left-hand vertical scale shows relative voltage. The right-hand vertical scale shows relative power. The solid curve represents the voltage as a function of time. The light and heavy dashed waves show peak and effective power, respectively, as functions of time.

Imaginary Power

If an ac circuit contains reactance, things get interesting. In a pure resistance, the rate of energy expenditure per unit time (or *true power*) is the same as the VA power (also known as *apparent power*). But when inductance and/or capacitance exists in an ac circuit, the VA power is greater than the power actually manifested as heat, light, radio waves, or whatever. The apparent power is then greater than the true power! The extra power is called *imaginary power*, because it exists in the reactance, and reactance can be, as you have learned, rendered in mathematically imaginary numerical form. Imaginary power is also known as *reactive power.*

Inductors and capacitors store energy and then release it a fraction of a cycle later. This phenomenon, like true power, is expressible as the rate at which energy is changed from one form to another. But rather than existing as a usable form of power, such as heat, light, radio waves, sound waves, or mechanical motion, imaginary power is stored up as a magnetic or electric field, and then released back into the circuit or system. This storage and release of power takes place over and over with each repeating ac cycle.

True Power Does Not Travel

A common and usually harmless misconception about true power is the notion that it can travel. For example, if you connect a radio transmitter to a cable that runs outdoors to an antenna, you might say you're "feeding power" through the cable to the antenna. Everybody says this, even engineers and technicians. But true power always involves a change in form, such as from electrical current and voltage into radio waves. It doesn't go from place to place. It simply happens in a specific place. It's the imaginary power that moves in situations like this, especially in transmission lines between power stations and power users, or between radio transmitters and radio antennas.

In a real-life radio antenna system, some true power is dissipated as heat in the transmitter amplifiers and in the feed line (Fig. 17-3). The useful dissipation of true power occurs when the imaginary power, in the form of electric and magnetic fields, gets to the antenna, where it is changed into *electromagnetic waves.*

You will often hear expressions such as "forward power" and "reflected power," or "power is fed from this amplifier to these speakers." It is all right to talk like this, but it can sometimes lead to

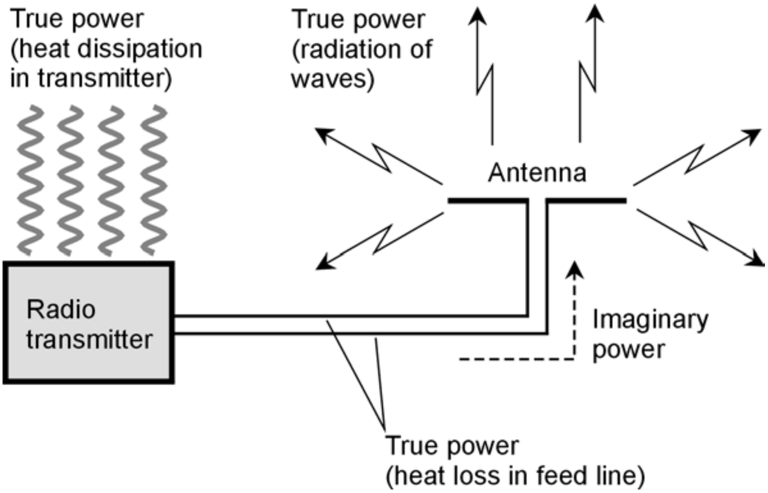

17-3 True power and imaginary power in a radio transmitter and antenna system.

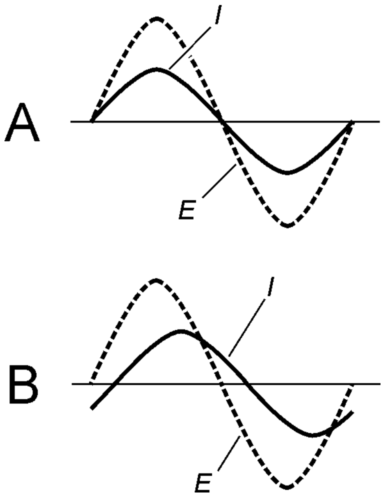

17-4 At A, current (*I*) and voltage (*E*) are in phase in a nonreactive ac circuit. At B, *I* and *E* are not in phase when reactance is present.

wrong conclusions, especially concerning impedance and standing waves. Then, you need to be keenly aware of the distinction among true, imaginary, and apparent power.

Reactance Does Not Consume Power

A pure inductance or a pure capacitance cannot dissipate any power. The only thing that such a component can do is store energy and then give it back to the circuit a fraction of a cycle later. In real life, the dielectrics or wires in coils and capacitors dissipate some power as heat, but ideal components would not do this.

A capacitor, as you have learned, stores energy as an electric field. An inductor stores energy as a magnetic field.

A component that contains reactance causes ac to shift in phase, so that the current is no longer exactly in step with the voltage. In a circuit with inductive reactance, the current lags the voltage by up to 90°, or one-quarter cycle. In a circuit with capacitive reactance, the current leads the voltage by up to 90°.

In a resistance-reactance circuit, true power is dissipated only in the resistive components. The reactive components exaggerate the VA power compared with the true power. Why, you ask, does reactance cause this discrepancy? In a circuit that is purely resistive, the voltage and current march right along in step with each other, and therefore, they combine in the most efficient possible way (Fig. 17-4A). But in a circuit containing reactance, the voltage and current are out of step with each other (Fig. 17-4B) because of their phase difference. Therefore, the actual energy expenditure, or true power, is not as great as the product of the voltage and the current.

True Power, VA Power, and Reactive Power

In an ac circuit or system containing nonzero resistance and nonzero reactance, the relationships among true power P_T, apparent (VA) power P_{VA}, and imaginary (reactive) power P_X are as follows:

$$P_{VA}^2 = P_T^2 + P_X^2$$
$$P_T < P_{VA}$$
$$P_X < P_{VA}$$

If there is no reactance in the circuit or system, then $P_{VA} = P_T$, and $P_X = 0$. Engineers strive to minimize, and if possible eliminate, the reactance in power-transmission systems.

Power Factor

In an ac circuit, the ratio of the true power to the VA power, P_T/P_{VA}, is called the *power factor.* If there is no reactance, the ideal case, then $P_T = P_{VA}$, and the power factor (PF) is equal to 1. If the circuit contains all reactance and no resistance of any significance (that is, zero or infinite resistance), then $P_T = 0$, and therefore $PF = 0$.

When a *load*, or a circuit in which you want power to be dissipated, contains resistance and reactance, then PF is between 0 and 1. That is, $0 < PF < 1$. The power factor can also be expressed as a percentage between 0 and 100, written $PF_{\%}$. Mathematically, we have these formulas for the power factor:

$$PF = P_T/P_{VA}$$
$$PF_{\%} = 100P_T/P_{VA}$$

When a load has some resistance and some reactance, then some of the power is dissipated as true power, and some is rejected by the load as imaginary power. In a sense, this imaginary power is sent back to the power source.

There are two ways to determine the power factor in an ac circuit that contains reactance and resistance. One method is to find the cosine of the phase angle. The other method involves the ratio of the resistance to the absolute-value impedance.

Cosine of Phase Angle

Recall that in a circuit having reactance and resistance, the current and the voltage are not in phase. The phase angle (ϕ) is the extent, expressed in degrees, to which the current and the voltage differ in phase. If there is no reactance, then $\phi = 0°$. If there is a pure reactance, then either $\phi = +90°$ (if the reactance is inductive) or else $\phi = -90°$ (if the reactance is capacitive). The power factor is equal to the cosine of the phase angle:

$$PF = \cos \phi$$

Problem 17-1

Suppose a circuit contains no reactance, but a pure resistance of 600 Ω. What is the power factor?

Without doing any calculations, it is evident that $PF = 1$, because $P_{VA} = P_T$ in a pure resistance. That means $P_T/P_{VA} = 1$. But you can also look at this by noting that the phase angle is 0°, because the current is in phase with the voltage. Using your calculator, you can see that $\cos 0° = 1$. Therefore, $PF = 1 = 100\%$. The vector for this case is shown in Fig. 17-5.

Problem 17-2

Suppose a circuit contains a pure capacitive reactance of −40 Ω, but no resistance. What is the power factor?

Here, the phase angle is −90° (Fig. 17-6). A calculator will tell you that $\cos -90° = 0$. Therefore, $PF = 0$, and $P_T/P_{VA} = 0 = 0\%$. None of the power is true; all of it is reactive.

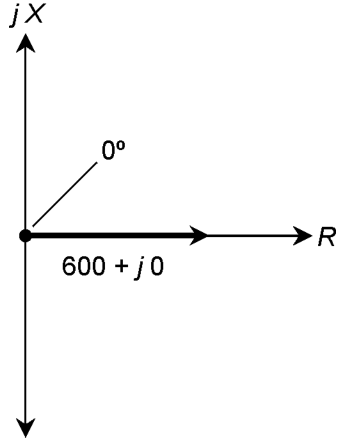

17-5 Vector diagram showing the phase angle for a purely resistive impedance of 600 + $j0$. The R and jX scales are relative.

Problem 17-3

Suppose a circuit contains a resistance of 50 Ω and an inductive reactance of 50 Ω in series. What is the power factor?

The phase angle in this case is 45° (Fig. 17-7). The resistance and reactance vectors have equal lengths and form two sides of a right triangle, with the complex impedance vector forming the hypotenuse. To determine the power factor, you can use a calculator to find cos 45° = 0.707. This means that $P_T/P_{VA} = 0.707 = 70.7\%$.

The Ratio *R/Z*

The second way to calculate the power factor is to find the ratio of the resistance R to the absolute-value impedance Z. In Fig. 17-7, this is visually apparent. A right triangle is formed by the resistance vector R (the base), the reactance vector jX (the height), and the absolute-value impedance Z (the hypotenuse). The cosine of the phase angle is equal to the ratio of the base length to the hypotenuse length; this represents R/Z.

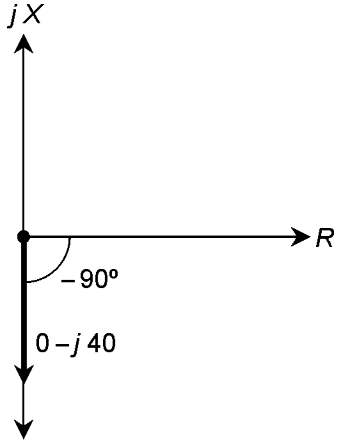

17-6 Vector diagram showing the phase angle for a purely capacitive impedance of 0 − $j40$. The R and jX scales are relative.

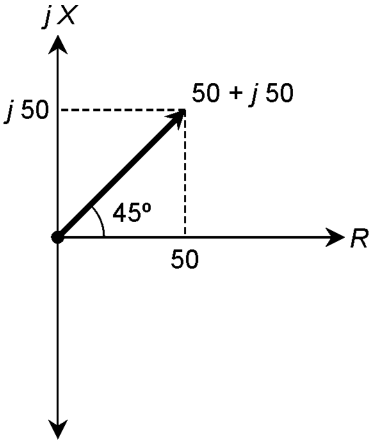

17-7 Vector diagram showing the phase angle for a complex impedance of 50 + *j*50. The *R* and *jX* scales are relative.

Problem 17-4

Suppose a circuit has an absolute-value impedance *Z* of 100 Ω, with a resistance *R* = 80 Ω. What is the power factor?

Simply find the ratio *PF* = *R*/*Z* = 80/100 = 0.8 = 80%. Note that it doesn't matter whether the reactance in this circuit is capacitive or inductive.

Problem 17-5

Suppose a circuit has an absolute-value impedance of 50 Ω, purely resistive. What is the power factor?

Here, *R* = *Z* = 50 Ω. Therefore, *PF* = *R*/*Z* = 50/50 = 1 = 100%.

Problem 17-6

Suppose a circuit has a resistance of 50 Ω and a capacitive reactance of –30 Ω in series. What is the power factor? Use the cosine method.

First, find the phase angle. Remember the formula: $\phi = \arctan (X/R)$, where *X* is the reactance and *R* is the resistance. Therefore, $\phi = \arctan (-30/50) = \arctan (-0.60) = -31°$. The power factor is the cosine of this angle; $PF = \cos (-31°) = 0.86 = 86\%$.

Problem 17-7

Suppose a circuit has a resistance of 30 Ω and an inductive reactance of 40 Ω. What is the power factor? Use the *R*/*Z* method.

Find the absolute-value impedance: $Z^2 = R^2 + X^2 = 30^2 + 40^2 = 900 + 1600 = 2500$. Therefore, $Z = 2500^{1/2} = 50\ \Omega$, so $PF = R/Z = 30/50 = 0.60 = 60\%$. This problem can be represented vectorially by a 30:40:50 right triangle, as shown in Fig. 17-8.

How Much of the Power Is True?

The preceding formulas allow you to figure out, given the resistance, reactance, and VA power, how many watts are true or real power, and how many watts are imaginary or reactive power. This is important in RF equipment, because some RF wattmeters display VA power rather than true power. When there is reactance in a circuit or system, the wattage reading is therefore exaggerated.

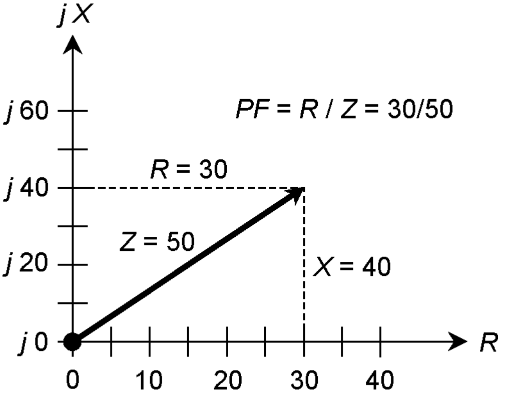

17-8 Illustration for Problem 17-7. (The vertical and horizontal scale increments differ; this is a common practice in graphs, often done for illustration convenience.)

Problem 17-8

Suppose a circuit has 50 Ω of resistance and 30 Ω of inductive reactance in series. A wattmeter shows 100 W, representing the VA power. What is the true power?

First, calculate the power factor. Suppose you use the phase-angle method. Then:

$$\begin{aligned}\phi &= \arctan\,(X/R)\\ &= \arctan\,(30/50) = 31^\circ\end{aligned}$$

The power factor is the cosine of the phase angle. Thus:

$$PF = \cos 31^\circ = 0.86 = 86\%$$

Remember that $PF = P_T/P_{VA}$. This formula can be rearranged to solve for true power:

$$\begin{aligned}P_T &= PF \times P_{VA}\\ &= 0.86 \times 100\\ &= 86\text{ W}\end{aligned}$$

Problem 17-9

Suppose a circuit has a resistance of 1000 Ω in parallel with a capacitance of 1000 pF. The frequency is 100 kHz. If a wattmeter designed to read VA power shows a reading of 88.0 W, what is the true power?

This problem is rather complicated because the components are in parallel. To begin, be sure the units are all in agreement so the formulas will work right. Convert the frequency to megahertz: $f = 100$ kHz $= 0.100$ MHz. Convert capacitance to microfarads: $C = 1000$ pF $= 0.001000$ µF. From the previous chapter, recall the formula for capacitive susceptance, and calculate it for this situation:

$$\begin{aligned}B_C &= 6.28fC\\ &= 6.28 \times 0.100 \times 0.001000\\ &= 0.000628\text{ S}\end{aligned}$$

The conductance of the resistor, G, is the reciprocal of the resistance, R, as follows:

$$\begin{aligned} G &= 1/R \\ &= 1/1000 \\ &= 0.001000 \text{ S} \end{aligned}$$

Now, use the formulas for calculating resistance and reactance in terms of conductance and susceptance in parallel circuits. First, find the resistance:

$$\begin{aligned} R &= G/(G^2 + B^2) \\ &= 0.001000/(0.001000^2 + 0.000628^2) \\ &= 0.001000/0.000001394 \\ &= 717\ \Omega \end{aligned}$$

Then, find the reactance:

$$\begin{aligned} X &= -B/(G^2 + B^2) \\ &= -0.000628/0.000001394 \\ &= -451\ \Omega \end{aligned}$$

Next, calculate the phase angle:

$$\begin{aligned} \phi &= \arctan\ (X/R) \\ &= \arctan\ (-451/717) \\ &= \arctan\ (-0.629) \\ &= -32.2^\circ \end{aligned}$$

The power factor is found from the phase angle as follows:

$$\begin{aligned} PF &= \cos \phi \\ &= \cos\ (-32.2^\circ) \\ &= 0.846 = 84.6\% \end{aligned}$$

The VA power, P_{VA}, is given as 88.0 W. Therefore:

$$\begin{aligned} P_{\text{T}} &= PF \times P_{\text{VA}} \\ &= 0.846 \times 88.0 \\ &= 74.4 \text{ W} \end{aligned}$$

Power Transmission

Consider how electricity gets to your home. Generators produce large voltages and currents at a power plant. The problem: getting the electricity from the plant to the homes, businesses, and other facilities that need it. This process involves the use of long wire transmission lines. Transformers are also required to step the voltages up or down. As another example, consider a radio broadcast or communications station. The transmitter produces high-frequency ac. The problem is getting the power to be radiated by the antenna, located some distance from the transmitter. This involves the use of an RF transmission line. The most common type is coaxial cable. Two-wire line is also sometimes used. At ultrahigh and microwave frequencies, another kind of transmission line, known as a *waveguide*, is often employed.

Loss: The Less, The Better!

The overriding concern in any *power transmission* system is minimizing the loss. Power wastage occurs almost entirely as heat in the transmission line conductors and dielectric, and in objects near the line. Some loss can take the form of unwanted electromagnetic radiation from the line. Loss also occurs in transformers. Power loss in an electrical system is analogous to the loss of usable work produced by friction in a mechanical system. The less of it, the better!

In an ideal power transmission system, all of the power is VA power; that is, it is in the form of ac in the conductors and an alternating voltage between them. It is undesirable to have power in a transmission line or transformer exist in the form of true power, because that translates into either heat loss, or radiation loss, or both. The place for true power dissipation or radiation is in the load, such as electrical appliances or radio antennas.

Power Measurement in a Transmission Line

In an ac transmission line, power is measured by placing an ac voltmeter between the conductors, and an ac ammeter in series with one of the conductors (Fig. 17-9). Then the power P (in watts) is equal to the product of the rms voltage E (in volts) and the rms current I (in amperes). This technique can be used in any transmission line. But this is not necessarily an indication of the true power dissipated by the load at the end of the line.

Recall that any transmission line has a *characteristic impedance*. This value, Z_o, depends on the diameters of the line conductors, the spacing between the conductors, and the type of dielectric material that separates the conductors. If the load is a pure resistance R containing no reactance, and if $R = Z_o$, then the power indicated by the voltmeter/ammeter scheme will be the same as the true power dissipated by the load—provided that the voltmeter and ammeter are placed at the load end of the transmission line.

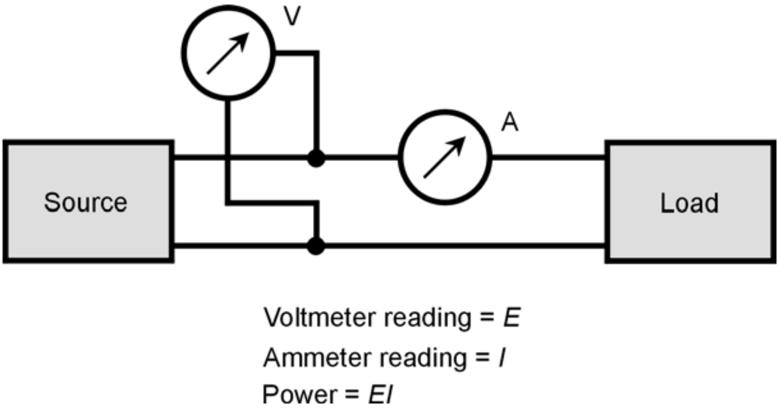

17-9 Power measurement in a transmission line. Ideally, the voltage and the current should be measured at the same physical point on the line.

If the load is a pure resistance but it differs from the characteristic impedance of the line, then the voltmeter and ammeter will not give an indication of the true power. Also, if there is any reactance in the load, the voltmeter/ammeter method will not be accurate, even if the resistive component happens to be the same as the characteristic impedance of the line. The physics of this is rather complicated, and we won't get into the details here. But you should remember that it is optimum for the impedance of a load to be a pure resistance R, such that $R = Z_o$. When this is not the case, an *impedance mismatch* is said to exist.

Small impedance mismatches can often be tolerated in power transmission systems. But this is not always the case. In very high frequency (VHF), ultrahigh frequency (UHF), and microwave radio transmitting systems, even a small impedance mismatch between the load and the line can cause excessive power losses in the line. An impedance mismatch can usually be corrected by means of a *matching transformer* between a transmission line and the load, and/or the deliberate addition of reactance at the load end of the line to cancel out any existing load reactance.

Loss in a Mismatched Line

When a transmission line is terminated in a resistance $R = Z_o$, then the current and the voltage are constant all along the line, provided the line has no loss. The ratio of the voltage to the current, E/I, is equal to R and also equal to Z_o. But this is an idealized case. No line is completely *lossless.*

In a real-world transmission line, the current and voltage gradually decrease as a signal makes its way from the source to the load. But if the load is a pure resistance equal to the characteristic impedance of the line, the current and voltage remain in the same ratio at all points along the line (Fig. 17-10).

Standing Waves

If the load is not perfectly matched to the line, the current and voltage vary in a complicated way along the length of the line. In some places, the current is high; in other places it is low. The max-

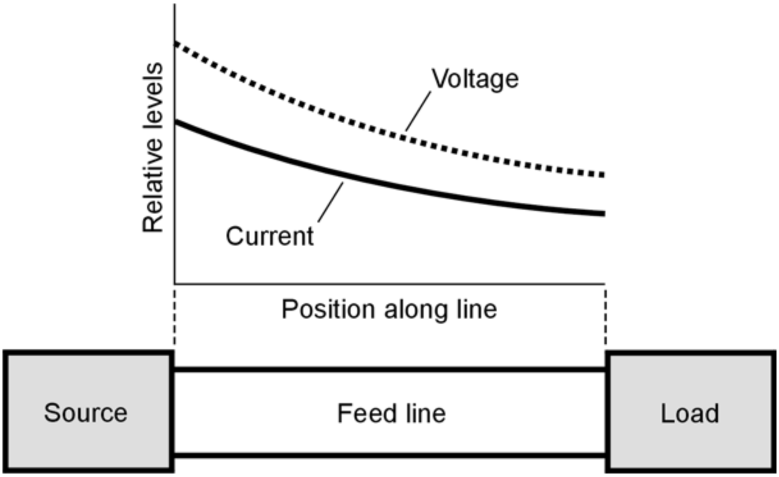

17-10 In a matched line, the ratio of the voltage to the current (E/I) is constant everywhere along the line, although the actual values of E and I decrease with increasing distance from the source.

ima and minima are called *loops* and *nodes*, respectively. At a *current loop*, the voltage is minimum (a *voltage node*), and at a *current node*, the voltage is maximum (a *voltage loop*). The current and voltage loops and nodes along a mismatched transmission line, if graphed as functions of the position on the line, form wavelike patterns that remain fixed over time. They just stand there. For this reason, they are called *standing waves.*

Standing-Wave Loss

At current loops, the loss in line conductors reaches a maximum. At voltage loops, the loss in the dielectric reaches a maximum. At current nodes, the loss in the conductors reaches a minimum. At voltage nodes, the loss in the dielectric reaches a minimum. It is tempting to suppose that everything would average out here, but it doesn't work that way! Overall, in a mismatched line, the line losses are greater than they are in a perfectly matched line. This extra line loss increases as the mismatch gets worse.

Transmission-line mismatch loss, also called *standing-wave loss*, occurs in the form of heat dissipation. It is true power. Any true power that goes into heating up a transmission line is wasted, because it cannot be dissipated in the load.

The greater the mismatch, the more severe the standing-wave loss becomes. The more loss a line has to begin with (that is, when it is perfectly matched), the more loss is caused by a given amount of mismatch. Standing-wave loss also increases as the frequency increases, if all other factors are held constant. This loss is the most significant, and the most harmful, in long lengths of transmission line, especially in RF practice at VHF, UHF, and microwave frequencies.

Line Overheating

A severe mismatch between the load and the transmission line can cause another problem: physical damage to, or destruction of, the line!

A feed line might be able to handle a kilowatt (1 kW) of power when it is perfectly matched. But if a severe mismatch exists and you try to feed 1 kW into the line, the extra current at the current loops can heat the conductors to the point where the dielectric material melts and the line shorts out. It is also possible for the voltage at the voltage loops to cause arcing between the line conductors. This perforates and/or burns the dielectric, ruining the line.

When an RF transmission line must be used with a mismatch, *derating functions* are required to determine how much power the line can safely handle. Manufacturers of prefabricated lines such as coaxial cable can supply you with this information.

Resonance

One of the most important phenomena in ac circuits, especially in RF engineering, is the property of *resonance.* This is a condition that occurs when capacitive and inductive reactance cancel each other out.

Series Resonance

Recall that capacitive reactance, X_C, and inductive reactance, X_L, can be equal in magnitude, although they are always opposite in effect. In any circuit containing an inductance and capacitance, there exists a frequency at which $X_L = -X_C$. This condition constitutes resonance. In a simple *LC* circuit, there is only one such frequency. But in some circuits involving transmission lines or antennas,

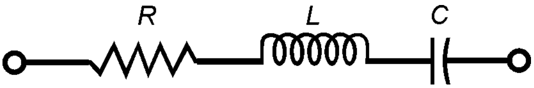

17-11 A series *RLC* circuit.

there can be many such frequencies. The lowest frequency at which resonance occurs is called the *resonant frequency*, symbolized f_o.

Refer to the schematic diagram of Fig. 17-11. You should recognize this as a series *RLC* circuit. At some particular frequency, $X_L = -X_C$. This is inevitable if L and C are finite and nonzero. This frequency is f_o for the circuit. At f_o, the effects of capacitive reactance and inductive reactance cancel out. The result is that the circuit appears as a pure resistance, with a value that is theoretically equal to R.

If $R = 0$, that is, if the resistor is a short circuit, then the circuit is called a *series LC circuit*, and the impedance at resonance will be theoretically $0 + j0$. The circuit will offer no opposition to the flow of alternating current at the frequency f_o. This condition is *series resonance.* In a practical series *LC* circuit, there is always a little bit of loss in the coil and capacitor, so the real part of the complex impedance is not exactly equal to 0 (although it can be extremely small).

Parallel Resonance

Refer to the circuit diagram of Fig. 17-12. This is a *parallel RLC circuit.* Remember that, in this sort of situation, the resistance R should be thought of as a conductance G, with $G = 1/R$. Then the circuit can be called a *parallel GLC circuit.*

At some particular frequency f_o, the inductive susceptance B_L will exactly cancel the capacitive susceptance B_C; that is, $B_L = -B_C$. This is inevitable for some frequency f_o, as long as the circuit contains finite, nonzero inductance and finite, nonzero capacitance. At the frequency f_o, the susceptances cancel each other out, leaving theoretically zero susceptance. The admittance through the circuit is then very nearly equal to the conductance, G, of the resistor.

If the circuit contains no resistor, but only a coil and capacitor, it is called a *parallel LC circuit*, and the admittance at resonance will be theoretically $0 + j0$. That means the circuit will offer great opposition to alternating current at f_o, and the complex impedance will theoretically be infinite! This condition is *parallel resonance.* In a practical parallel *LC* circuit, there is always a little bit of loss in the coil and capacitor, so the real part of the complex impedance is not infinite (although it can be extremely large).

Calculating Resonant Frequency

The formula for calculating resonant frequency f_o, in terms of the inductance L in henrys and the capacitance C in farads, is as follows:

$$f_o = 1/[2\pi(LC)^{1/2}]$$

17-12 A parallel *RLC* circuit.

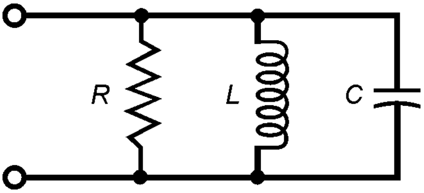

Considering $\pi = 3.14$ to three significant figures, this formula can be simplified to:

$$f_o = 0.159/(LC)^{1/2}$$

The ½ power of a quantity represents the positive square root of that quantity. The preceding formulas are valid for series-resonant and parallel-resonant *RLC* circuits.

The formula will also work if you want to find f_o in megahertz (MHz) when L is given in microhenrys (μH) and C is in microfarads (μF). These values are far more common than hertz, henrys, and farads in electronic circuits. Just remember that millions of hertz go with million*th*s of henrys, and with million*th*s of farads.

The Effects of *R* and *G*

Interestingly, the value of R or G does not affect the resonant frequency in either type of circuit. But these quantities are significant, nevertheless! The presence of nonzero resistance in a *series-resonant circuit*, or nonzero conductance in a *parallel-resonant circuit*, makes the resonant frequency less well-defined. Engineers say that the *resonant frequency response* becomes "more broad" or "less sharp."

In a series circuit, the resonant frequency response becomes more broad as the *resistance* increases. In a parallel circuit, the resonant frequency response becomes more broad as the *conductance* increases. The sharpest possible responses occur when $R = 0$ in a series circuit, and when $G = 0$ (that is, $R = \infty$) in a parallel circuit.

Problem 17-10

Find the resonant frequency of a series circuit with an inductance of 100 μH and a capacitance of 100 pF.

First, convert the capacitance to microfarads: 100 pF = 0.000100 μF. Then find the product $LC = 100 \times 0.000100 = 0.0100$. Take the square root of this, getting 0.100. Finally, divide 0.159 by 0.100, getting $f_o = 1.59$ MHz.

Problem 17-11

Find the resonant frequency of a parallel circuit consisting of a 33-μH coil and a 47-pF capacitor.

Again, convert the capacitance to microfarads: 47 pF = 0.000047 μF. Then find the product $LC = 33 \times 0.000047 = 0.00155$. Take the square root of this, getting 0.0394. Finally, divide 0.159 by 0.0394, getting $f_o = 4.04$ MHz.

Problem 17-12

Suppose you want to design a circuit so that it has $f_o = 9.00$ MHz. You have a 33-pF fixed capacitor available. What size coil will be needed to get the desired resonant frequency?

Use the formula for the resonant frequency, and plug in the values. This will allow you to use simple arithmetic to solve for L. Convert the capacitance to microfarads: 33 pF = 0.000033 μF. Then calculate as follows:

$$\begin{aligned} f_o &= 0.159/(LC)^{1/2} \\ 9.00 &= 0.159/(L \times 0.000033)^{1/2} \\ 9.00^2 &= 0.159^2/(0.000033 \times L) \\ 81.0 &= 0.0253/(0.000033 \times L) \end{aligned}$$

$$\begin{aligned} 81.0 \times 0.000033 \times L &= 0.0253 \\ 0.00267 \times L &= 0.0253 \\ L &= 0.0253/0.00267 \\ &= 9.48\ \mu H \end{aligned}$$

Problem 17-13

Suppose a circuit must be designed to have $f_o = 455$ kHz. A coil of 100 µH is available. What size capacitor is needed?

Convert the frequency to megahertz: 455 kHz = 0.455 MHz. Then the calculation proceeds in the same way as with the preceding problem:

$$\begin{aligned} f_o &= 0.159/(LC)^{1/2} \\ 0.455 &= 0.159/(100 \times C)^{1/2} \\ 0.455^2 &= 0.159^2/(100 \times C) \\ 0.207 &= 0.0253/(100 \times C) \\ 0.207 \times 100 \times C &= 0.0253 \\ 20.7 \times C &= 0.0253 \\ C &= 0.0253/20.7 \\ &= 0.00122\ \mu\text{F} \\ &= 1220\ \text{pF} \end{aligned}$$

In practical circuits, variable inductors and/or variable capacitors are often placed in tuned circuits, so that small errors in the frequency can be compensated for. The most common approach is to design the circuit for a frequency slightly higher than f_o, and to use a *padder capacitor* in parallel with the main capacitor (Fig. 17-13).

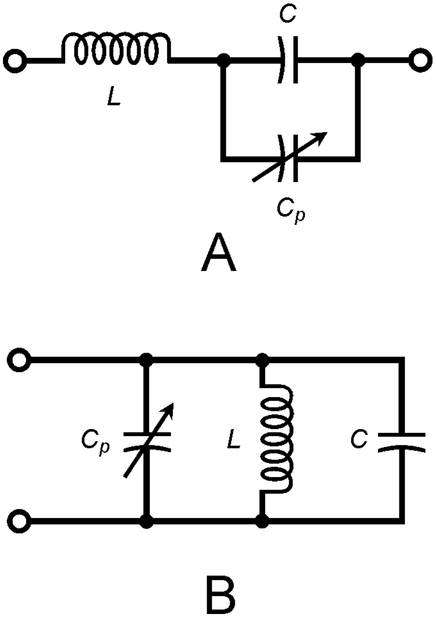

17-13 Padding capacitors (C_p) allow limited adjustment of the resonant frequency in a series *LC* circuit (as shown at A), or in a parallel *LC* circuit (as shown at B).

Resonant Devices

Resonant circuits often consist of coils and capacitors in series or parallel, but there are other kinds of hardware that exhibit resonance. Some of these are as follows.

Piezoelectric Crystals

Pieces of quartz, when cut into thin wafers and subjected to voltages, will vibrate at high frequencies. Because of the physical dimensions of such a *piezoelectric crystal*, these vibrations occur at a precise frequency f_o, and also at whole-number multiples of f_o. These multiples, $2f_o$, $3f_o$, $4f_o$, and so on, are called *harmonic frequencies* or simply *harmonics.* The frequency f_o is called the *fundamental frequency* or simply the *fundamental.* The fundamental, f_o, is defined as the lowest frequency at which resonance occurs. Quartz crystals can be made to act like *LC* circuits in electronic devices. A crystal exhibits an impedance that varies with frequency. The reactance is zero at f_o and the harmonic frequencies.

Cavities

Lengths of metal tubing, cut to specific dimensions, exhibit resonance at very high, ultrahigh, and microwave radio frequencies. They work in much the same way as musical instruments resonate with sound waves. But the waves are electromagnetic, rather than acoustic. Such *cavities*, also called cavity *resonators*, have reasonable physical dimensions at frequencies above about 150 MHz. Below this frequency, a cavity can be made to work, but it is long and unwieldy. Like crystals, cavities resonate at a fundamental frequency f_o, and also at harmonic frequencies.

Sections of Transmission Line

When a transmission line is cut to ¼ wavelength, or to any whole-number multiple of this, it behaves as a resonant circuit. The most common length for a transmission-line resonator is a ¼ wavelength. Such a piece of transmission line is called a *quarter-wave section.*

When a quarter-wave section is short-circuited at the far end, it acts like a parallel-resonant *LC* circuit, and has a high resistive impedance at the resonant frequency f_o. When it is open at the far end, it acts as a series-resonant *LC* circuit, and has a low resistive impedance at f_o. In effect, a quarter-wave section converts an ac short circuit into an ac open circuit and vice versa, at a specific frequency f_o.

The length of a quarter-wave section depends on the desired f_o. It also depends on how fast the electromagnetic energy travels along the line. This speed is specified in terms of a *velocity factor*, abbreviated v. The value of v is given as a fraction of the speed of light. Typical transmission lines have velocity factors ranging from about 0.66 to 0.95 (or 66 percent to 95 percent). This factor is provided by the manufacturers of prefabricated lines such as coaxial cable.

If the frequency in megahertz is f_o and the velocity factor of a line is v, then the length L_{ft} of a quarter-wave section of transmission line, in feet, is given by this formula:

$$L_{ft} = 246v/f_o$$

The length L_m in meters is given by this:

$$L_m = 75.0v/f_o$$

We use L here to stand for “length,” not “inductance”!

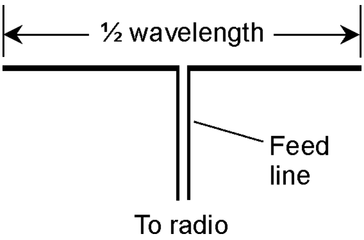

17-14 The half-wave, center-fed dipole is a simple and efficient antenna.

Antennas

Many types of antennas exhibit resonant properties. The simplest type of resonant antenna, and the only kind that will be mentioned here, is the center-fed, half-wavelength *dipole antenna* (Fig. 17-14).

The length L_{ft}, in feet, for a dipole antenna at a frequency of f_o, in megahertz, is given by the following formula:

$$L_{ft} = 468/f_o$$

This takes into account the fact that electromagnetic fields travel along a wire at about 95 percent of the speed of light. A straight, thin wire in free space has a velocity factor of approximately 0.95.

If the length of the half-wave dipole is specified in meters as L_m, then:

$$L_m = 143/f_o$$

A half-wave dipole has a purely resistive impedance of about 73 Ω at its fundamental frequency f_o. But this type of antenna is also resonant at all harmonics of f_o. The dipole is a full wavelength long at $2f_o$; it is $^3\!/_2$ wavelength long at $3f_o$; it is two full wavelengths long at $4f_o$, and so on.

Radiation Resistance

At f_o and all of the odd harmonics, the antenna behaves like a series-resonant *RLC* circuit with a fairly low resistance. At all even harmonics, the antenna acts like a parallel-resonant *RLC* circuit with a high resistance. Does this confuse you? There's no resistor in Fig. 17-14! Where, you ask, does the resistance come from in the half-wave dipole? The answer to this is rather esoteric, and it brings to light an interesting property that all antennas have. It is called *radiation resistance*, and is a crucial factor in the design and construction of all RF antenna systems.

When electromagnetic energy is fed into an antenna, power is radiated into space in the form of radio waves. This is a manifestation of true power, just as the dissipation of power in a pure resistance is a manifestation of true power. Although there is no physical resistor in Fig. 17-14, the radiation of radio waves is like power dissipation in a pure resistance. In fact, if a half-wave dipole antenna were replaced with a 73-Ω nonreactive resistor that could dissipate enough power without burning out, a radio transmitter connected to the opposite end of the line wouldn't know the difference. (But a receiver would!)

Problem 17-14

How many feet long is a quarter-wave section of transmission line at 7.05 MHz, if the velocity factor is 0.800?

Just use the formula:

$$\begin{aligned} L_{\text{ft}} &= 246v/f_{\text{o}} \\ &= (246 \times 0.800)/7.05 \\ &= 197/7.05 \\ &= 27.9 \text{ ft} \end{aligned}$$

Quiz

Refer to the text in this chapter if necessary. A good source is 18 or more correct. Answers are in the back of the book.

1. The power in a pure reactance is
   (a) radiated.
   (b) true.
   (c) imaginary.
   (d) apparent.

2. Which of the following is *not* an example of true power?
   (a) Power in the form of heat, produced by dc flowing through a resistor
   (b) Power in the form of electromagnetic fields, radiated from a radio antenna
   (c) The product of the rms ac through a capacitor and the rms voltage across it
   (d) Power in the form of heat, produced by losses in an RF transmission line

3. Suppose the apparent power in a circuit is 100 W, and the imaginary power is 40 W. What is the true power?
   (a) 92 W
   (b) 100 W
   (c) 140 W
   (d) It is impossible to determine from this information.

4. Power factor is equal to
   (a) apparent power divided by true power.
   (b) imaginary power divided by apparent power.
   (c) imaginary power divided by true power.
   (d) true power divided by apparent power.

5. Suppose a circuit has a resistance of 300 Ω and an inductance of 13.5 μH in series, and is operated at 10.0 MHz. What is the power factor?
   (a) 0.334
   (b) 0.999
   (c) 0.595
   (d) It cannot be determined from the information given.

6. Suppose a series circuit has $Z = 88.4\ \Omega$, with $R = 50.0\ \Omega$. What is the power factor, expressed as a percentage?

   (a) 99.9 percent
   (b) 56.6 percent
   (c) 60.5 percent
   (d) 29.5 percent

7. Suppose a series circuit has $R = 53.5\ \Omega$, with $X = 75.5\ \Omega$. What is the power factor, expressed as a percentage?

   (a) 70.9 percent
   (b) 81.6 percent
   (c) 57.8 percent
   (d) 63.2 percent

8. The phase angle in an ac circuit is equal to

   (a) arctan (Z/R).
   (b) arctan (R/Z).
   (c) arctan (R/X).
   (d) arctan (X/R).

9. Suppose an ac ammeter and an ac voltmeter indicate that there are 220 W of VA power in a circuit that consists of a resistance of 50 Ω in series with a capacitive reactance of −20 Ω. What is the true power?

   (a) 237 W
   (b) 204 W
   (c) 88.0 W
   (d) 81.6 W

10. Suppose an ac ammeter and an ac voltmeter indicate that there are 57 W of VA power in a circuit. The resistance is known to be 50 Ω, and the true power is known to be 40 W. What is the absolute-value impedance?

    (a) 50 Ω
    (b) 57 Ω
    (c) 71 Ω
    (d) It is impossible to determine on the basis of this data.

11. Which of the following should be minimized in an RF transmission line?

    (a) The load impedance
    (b) The load resistance
    (c) The line loss
    (d) The transmitter power

12. Which of the following does *not* increase the loss in a transmission line?
    - (a) Reducing the power output of the source
    - (b) Increasing the degree of mismatch between the line and the load
    - (c) Reducing the diameter of the line conductors
    - (d) Raising the frequency

13. Which of the following is a significant problem that standing waves can cause in an RF transmission line?
    - (a) Line overheating
    - (b) Excessive power loss
    - (c) Inaccuracy in power measurement
    - (d) All of the above

14. Suppose a coil and capacitor are in series. The inductance is 88 mH and the capacitance is 1000 pF. What is the resonant frequency?
    - (a) 17 kHz
    - (b) 540 Hz
    - (c) 17 MHz
    - (d) 540 kHz

15. Suppose a coil and capacitor are in parallel, with $L = 10.0\ \mu\text{H}$ and $C = 10$ pF. What is f_o?
    - (a) 15.9 kHz
    - (b) 5.04 MHz
    - (c) 15.9 MHz
    - (d) 50.4 MHz

16. Suppose you want to build a series-resonant circuit with $f_o = 14.1$ MHz. A coil of 13.5 µH is available. How much capacitance is needed?
    - (a) 0.945 µF
    - (b) 9.45 pF
    - (c) 94.5 pF
    - (d) 945 pF

17. Suppose you want to build a parallel-resonant circuit with $f_o = 21.3$ MHz. A capacitor of 22.0 pF is available. How much inductance is needed?
    - (a) 2.54 mH
    - (b) 254 µH
    - (c) 25.4 µH
    - (d) 2.54 µH

18. A $^1\!/\!_4$-wave section of transmission line is cut for use at 21.1 MHz. The line has a velocity factor of 0.800. What is its physical length in meters?
    - (a) 11.1 m
    - (b) 3.55 m

(c) 8.87 m
(d) 2.84 m

19. What is the fourth harmonic of 800 kHz?
(a) 200 kHz
(b) 400 kHz
(c) 3.20 MHz
(d) 4.00 MHz

20. Suppose you want to build a $^1/_2$-wave dipole antenna designed to have a fundamental resonant frequency of 3.60 MHz. How long should you make it, as measured from end to end in feet?
(a) 130 ft
(b) 1680 ft
(c) 39.7 ft
(d) 515 ft

18
CHAPTER
Transformers and Impedance Matching

TRANSFORMERS ARE USED TO OBTAIN THE OPTIMUM VOLTAGE FOR THE OPERATION OF A CIRCUIT OR system. Transformers can also match impedances between a circuit and a load, or between two different circuits. Transformers can be used to provide dc isolation between electronic circuits while letting ac pass. Another application is to mate balanced and unbalanced circuits, feed systems, and loads.

Principle of the Transformer

When two wires are near each other and one of them carries a fluctuating current, a fluctuating current is induced in the other wire. This effect is known as *electromagnetic induction.* All ac transformers work according to the principle of electromagnetic induction. If the first wire carries sine-wave ac of a certain frequency, then the *induced current* is sine-wave ac of the same frequency in the second wire.

The closer the two wires are to each other, the greater is the induced current, for a given current in the first wire. If the wires are wound into coils and placed along a common axis (Fig. 18-1), the induced current will be greater than if the wires are straight and parallel. Even more *coupling*, or efficiency of induced-current transfer, is obtained if the two coils are wound one atop the other.

Primary and Secondary

The two windings, along with the core on which they are wound, constitute a transformer. The first coil is called the *primary winding*, and the second coil is known as the *secondary winding.* These are often spoken of simply as the *primary* and the *secondary.* The induced current in the secondary creates a voltage between its end terminals. In a *step-down* transformer, the secondary voltage is less than the primary voltage. In a *step-up* transformer, the secondary voltage is greater than the primary voltage. The primary voltage is abbreviated E_{pri}, and the secondary voltage is abbreviated E_{sec}. Unless otherwise stated, effective (rms) voltages are always specified.

The windings of a transformer have inductance, because they are coils. The required inductances of the primary and secondary depend on the frequency of operation, and also on the resistive part of the impedance in the circuit. As the frequency increases, the needed inductance decreases. At high resistive impedances, more inductance is generally needed than at low resistive impedances.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

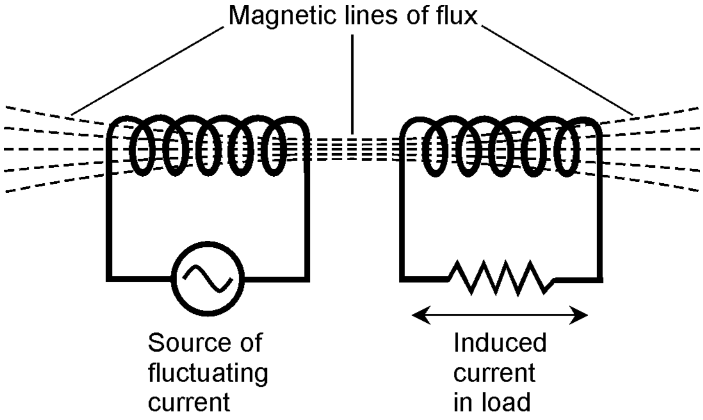

18-1 Magnetic lines of flux between two aligned coils of wire when one of the coils carries fluctuating or alternating current.

Turns Ratio

The ***primary-to-secondary turns ratio*** in a transformer is the ratio of the number of turns in the primary, T_{pri}, to the number of turns in the secondary, T_{sec}. This ratio is written T_{pri}:T_{sec} or T_{pri}/T_{sec}. In a transformer with excellent primary-to-secondary coupling, the following relationship always holds:

$$E_{pri}/E_{sec} = T_{pri}/T_{sec}$$

That is, the primary-to-secondary voltage ratio is always equal to the primary-to-secondary turns ratio (Fig. 18-2).

Problem 18-1

Suppose a transformer has a primary-to-secondary turns ratio of exactly 9:1. The ac voltage at the primary is 117 V rms. Is this a step-up transformer or a step-down transformer? What is the voltage across the secondary?

This is a step-down transformer. Simply plug in the numbers in the preceding equation and solve for E_{sec}, as follows:

$$\begin{aligned} E_{pri}/E_{sec} &= T_{pri}/T_{sec} \\ 117/E_{sec} &= 9.00 \\ 1/E_{sec} &= 9.00/117 \\ E_{sec} &= 117/9.00 \\ &= 13.0 \text{ V rms} \end{aligned}$$

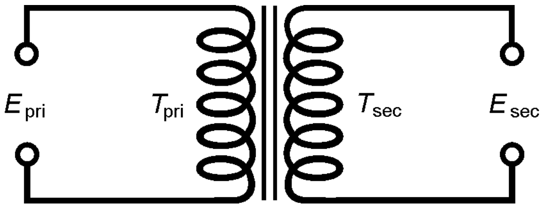

18-2 The primary voltage (E_{pri}) and secondary voltage (E_{sec}) in a transformer depend on the number of turns in the primary winding (T_{pri}) versus the number of turns in the secondary winding (T_{sec}).

Problem 18-2

Consider a transformer with a primary-to-secondary turns ratio of exactly 1:9. The voltage at the primary is 121.4 V rms. Is this a step-up transformer or a step-down transformer? What is the voltage at the secondary?

This is a step-up transformer. Plug in numbers and solve for E_{sec}, as follows:

$$\begin{aligned} 121.4/E_{sec} &= 1/9.000 \\ E_{sec}/121.4 &= 9.000 \\ E_{sec} &= 9.000 \times 121.4 \\ &= 1093 \text{ V rms} \end{aligned}$$

Sometimes the *secondary-to-primary turns ratio* is given, rather than the primary-to-secondary turns ratio. This is written T_{sec}/T_{pri}. In a step-down unit, T_{sec}/T_{pri} is less than 1. In a step-up unit, T_{sec}/T_{pri} is greater than 1. When you hear someone say that such-and-such a transformer has a certain "turns ratio," say 10:1, be sure of which ratio is meant, T_{pri}/T_{sec} or T_{sec}/T_{pri}! If you get it wrong, you'll have the secondary voltage wrong by a factor of the *square* of the turns ratio.

Ferromagnetic Cores

If a ferromagnetic substance such as laminated iron or powdered iron is placed within the pair of coils, the extent of coupling is increased far above that possible with an air core. But this improvement in coupling is obtained at a price. Some energy is invariably lost as heat in the core. Also, ferromagnetic cores limit the maximum frequency at which a transformer will work well.

The schematic symbol for an air-core transformer consists of two inductor symbols back-to-back (Fig. 18-3A). If a laminated iron core is used, two parallel lines are added to the schematic symbol (Fig. 18-3B). If the core is made of powdered iron, the two parallel lines are broken or dashed (Fig. 18-3C).

In transformers for 60-Hz utility ac, and also for low audio-frequency (AF) use, sheets of an alloy called *silicon steel*, glued together in layers, are often employed as transformer cores. The silicon steel is sometimes called *transformer iron*. The reason layering is used, rather than making the core from a single mass of metal, is that the magnetic fields from the coils cause currents to flow in a solid core. These *eddy currents* go in circles, heating up the core and wasting energy that would otherwise be transferred from the primary to the secondary. Eddy currents are choked off by breaking up the core into layers, so that currents cannot flow very well in circles.

A rather esoteric form of loss, called *hysteresis loss*, occurs in all ferromagnetic transformer cores, but especially laminated iron. Hysteresis is the tendency for a core material to be sluggish in accept-

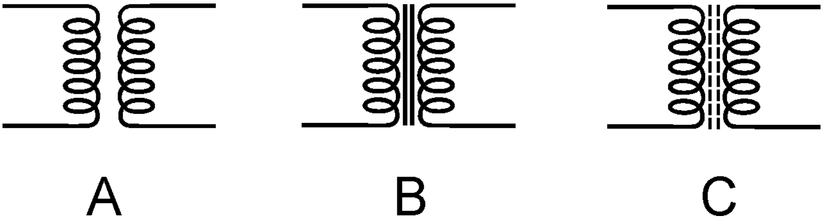

18-3 Schematic symbols for transformers. At A, air core. At B, laminated iron core. At C, ferrite or powdered iron core.

ing a fluctuating magnetic field. Laminated cores exhibit high hysteresis loss above the AF range, and are therefore not good above a few kilohertz.

At frequencies up to several tens of megahertz, *powdered iron* works well for RF transformers. This material has high magnetic permeability and concentrates the flux efficiently. High permeability cores minimize the number of turns needed in the coils, and this minimizes the loss that occurs in the wires.

Geometries

The properties of a transformer depend on the shape of its core, and on the way in which the wires are wound on it. There are several different *geometries* used with transformers.

E Core

A common core for a power transformer is the *E core*, so named because it is shaped like the capital letter E. A bar, placed at the open end of the E, completes the core assembly after the coils have been wound on the E-shaped section (Fig. 18-4A).

The primary and secondary windings can be placed on an E core in either of two ways. The simpler winding method is to put both the primary and the secondary around the middle bar of the E (Fig. 18-4B). This is called the *shell method* of transformer winding. It provides maximum coupling between the windings. However, this scheme results in considerable capacitance between the primary and the secondary. Such *interwinding capacitance* can sometimes be tolerated, but often it cannot. Another disadvantage of the shell geometry is that, when windings are placed one on top of the other, the transformer cannot handle very much voltage. High voltages cause arcing between the windings, which can destroy the insulation on the wires and lead to permanent short circuits.

Another winding method is the *core method.* In this scheme, one winding is placed at the bottom of the E section, and the other winding is placed at the top (Fig. 18-4C). The coupling occurs

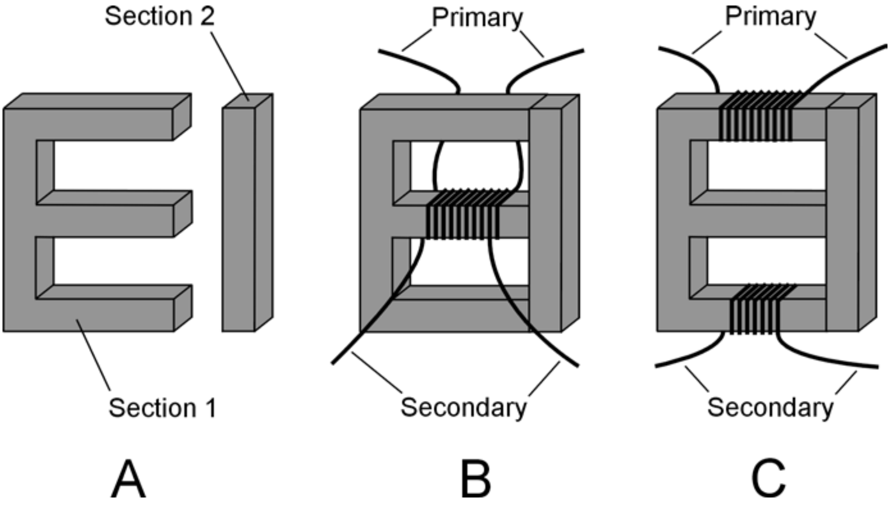

18-4 At A, a utility transformer E core, showing both sections. At B, the shell winding method. At C, the core winding method.

by means of magnetic flux in the core. The interwinding capacitance is lower than it is in a shell-wound transformer because the windings are physically farther apart. Also, a core-wound transformer can handle higher voltages than a shell-wound transformer of the same physical size. Sometimes the center part of the E is left out of the core when the core winding scheme is used.

Shell-wound and core-wound transformers are almost universally employed at 60 Hz. These configurations are also common at AF.

Solenoidal Core

A pair of cylindrical coils, wound around a rod-shaped piece of powdered iron or ferrite, was once a common configuration for RF transformers. Sometimes this type of transformer is still seen, although it is most often used as a *loopstick antenna* in portable radio receivers and in radio direction-finding equipment. The coil windings can be placed one atop the other, or they can be separated (Fig. 18-5) to reduce the capacitance between the primary and secondary.

In a loopstick antenna, the primary serves to pick up the radio signals. The secondary winding provides an optimum impedance match to the first amplifier stage, or *front end*, of the radio receiver. The use of transformers for impedance matching is discussed later in this chapter.

Toroidal Core

The *toroidal core* (or *toroid*) has become common for winding RF transformers. The core is a donut-shaped ring of powdered iron. The coils are wound around the donut. The complete assembly is called a *toroidal transformer.* The primary and secondary can be wound one over the other, or they can be wound over different parts of the core (Fig. 18-6). As with other transformers, when the windings are one on top of the other, there is more interwinding capacitance than when they are separated.

Toroids confine practically all the magnetic flux within the core material. This allows toroidal coils and transformers to be placed near other components without inductive interaction. Also, a toroidal coil or transformer can be mounted directly on a metal chassis, and the operation is not affected (assuming the wire is insulated or enameled).

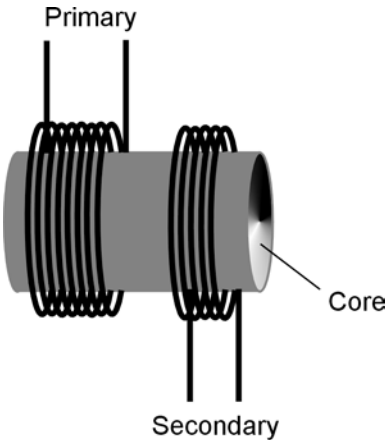

18-5 A solenoidal-core transformer.

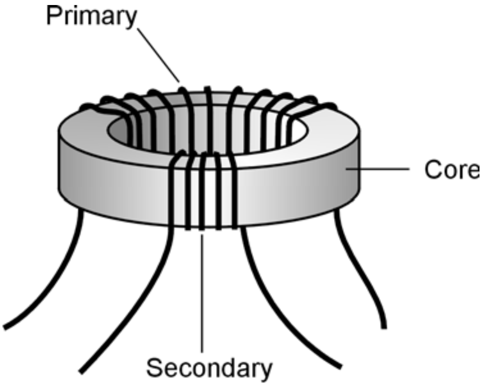

18-6 A toroidal-core transformer.

A toroidal core provides considerably more inductance per turn, for the same kind of ferromagnetic material, than a solenoidal core. It is common to see toroidal coils or transformers that have inductance values as high as 100 mH.

Pot Core

Even more inductance per turn can be obtained with a *pot core.* This is a shell of ferromagnetic material that is wrapped around a loop-shaped coil. The core is manufactured in two halves (Fig. 18-7). You wind the coil inside one of the halves, and then bolt the two together. The final core completely surrounds the loop, and the magnetic flux is confined to the core material.

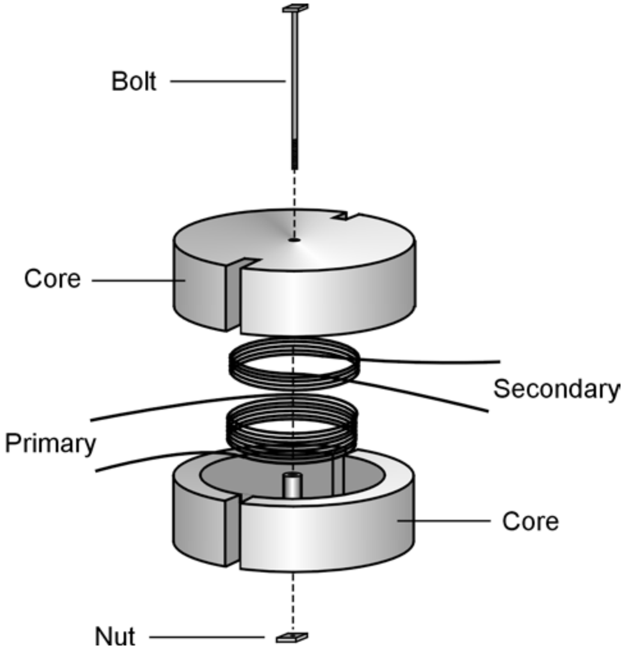

18-7 Exploded view of a pot-core transformer.

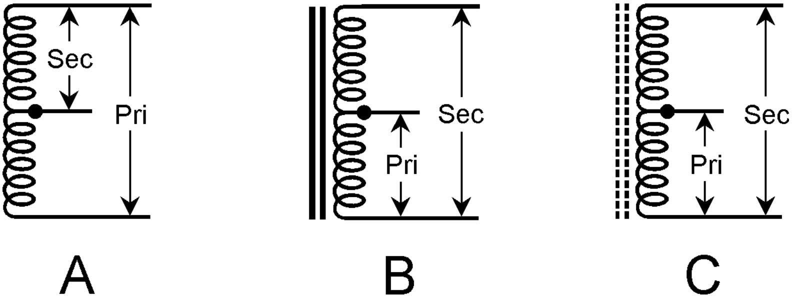

18-8 Schematic symbols for autotransformers. At A, air core, step-down. At B, laminated iron core, step-up. At C, ferrite or powdered iron core, step-up.

Like the toroid, the pot core is self-shielding. There is essentially no coupling to external components. A pot core can be used to wind a single, high-inductance coil. Inductance values of more than 1 H are possible with a reasonable number of wire turns.

In a pot-core transformer, the primary and secondary must be wound next to each other. This is unavoidable because of the geometry. Therefore, the interwinding capacitance of a pot-core transformer is high. Pot cores are useful at AF and the lowest-frequency parts of the RF spectrum. They are rarely employed at high radio frequencies.

Autotransformer

In some situations, there is no need to provide dc isolation between the primary and secondary windings of a transformer. In a case of this sort, an *autotransformer* can be used. It has a single, tapped winding.

Figure 18-8 shows three autotransformer configurations. The unit shown at A has an air core, and is a step-down type. The unit at B has a laminated iron core, and is a step-up type. The unit at C has a powdered iron core, and is a step-up type.

You'll sometimes see autotransformers in radio receivers or transmitters. Autotransformers work well in impedance-matching applications, and also perform well as solenoidal loopstick antennas. Autotransformers are occasionally, but not often, used in AF applications and in 60-Hz utility wiring. In utility circuits, autotransformers can step the voltage down by a large factor, but they aren't used to step voltages up by more than a few percent.

Power Transformers

Any transformer used in the 60-Hz utility line, intended to provide a certain rms ac voltage for the operation of electrical circuits, is a *power transformer.* Power transformers exist in a vast range of physical sizes, from smaller than a tennis ball to as big as a room.

At the Generating Plant

The largest transformers are employed at the places where electricity is generated. Not surprisingly, high-energy power plants have bigger transformers that develop higher voltages than low-energy, local power plants. These transformers must be able to handle high voltages and large currents simultaneously.

When electrical energy must be sent over long distances, extremely high voltages are used. This is because, for a given amount of power ultimately dissipated by the loads, the current is lower when the voltage is higher. Lower current translates into reduced loss in the transmission line.

Recall the formula $P = EI$, where P is the power (in watts), E is the voltage (in volts), and I is the current (in amperes). If you can make the voltage 10 times larger, for a given power level, then the current is reduced to $1/10$ as much. The *ohmic losses* in the wires are proportional to the *square* of the current. Remember that $P = I^2R$, where P is the power (in watts), I is the current (in amperes), and R is the resistance (in ohms). Engineers can't do much about the wire resistance or the power consumed by the loads, but they can adjust the voltage, and thereby the current.

Suppose the voltage in a power transmission line is increased by a factor of 10, and the load at the end of the line draws constant power. This increase in the voltage reduces the current to $1/10$ of its previous value. As a result, the ohmic loss is cut to $(1/10)^2$, or $1/100$, of its previous amount. That's a major improvement in the efficiency of the transmission line, at least in terms of the loss caused by the resistance in the wires—and it is the reason why regional power plants have massive transformers capable of generating hundreds of thousands of volts.

Along the Line

Extreme voltage is good for *high-tension* power transmission, but it's certainly of no use to an average consumer. The wiring in a high-tension system must be done using precautions to prevent arcing (sparking) and short circuits. Personnel must be kept at least several meters away from the wires. Can you imagine trying to use an appliance, say a home computer, by plugging it into a 500,000-V rms electrical outlet?

Medium-voltage power lines branch out from the major lines, and step-down transformers are used at the branch points. These lines fan out to still lower-voltage lines, and step-down transformers are employed at these points, too. Each transformer must have windings heavy enough to withstand the product $P = EI$, the amount of VA power delivered to all the subscribers served by that transformer, at periods of peak demand.

Sometimes, such as during a heat wave, the demand for electricity rises above the normal peak level. This loads down the circuit to the point that the voltage drops several percent. This is called a *brownout*. If consumption rises further still, a dangerous current load is placed on one or more intermediate power transformers. Circuit breakers in the transformers protect them from destruction by opening the circuit. Then there is a temporary *blackout*.

At individual homes and buildings, transformers step the voltage down to either 234 V rms or 117 V rms. Usually, 234-V rms electricity is provided in the form of three sine waves, called *phases*, each separated by 120°, and each appearing at one of the three slots in the outlet (Fig. 18-9A). This voltage is commonly employed with heavy appliances, such as the kitchen oven/stove (if they are electric), heating (if it is electric), and the laundry washer and dryer. A 117-V rms outlet supplies just one phase, appearing between two of the three slots in the outlet. The third opening in the outlet leads to an earth ground (Fig. 18-9B).

In Electronic Devices

The smallest power transformers are found in electronic equipment such as television sets, ham radios, and home computers. Most solid-state devices use low voltages, ranging from about 5 V up to perhaps 50 V. This equipment needs step-down power transformers in its power supplies.

Solid-state equipment usually (but not always) consumes relatively little power, so the transformers are usually not very bulky. The exception is high-powered AF or RF amplifiers, whose tran-

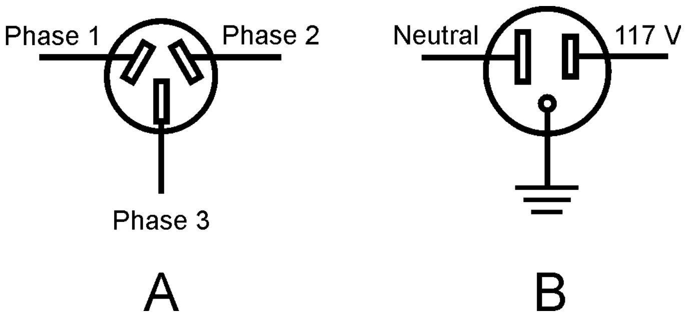

18-9 At A, an outlet for three-phase, 234-V rms utility ac. At B, a conventional single-phase utility outlet for 117-V rms utility ac.

sistors can demand more than 1000 W (1 kW) in some cases. At 12 V, this translates to a current demand of 90 A or more.

Television sets have cathode-ray tubes that need several hundred volts. This is derived by using a step-up transformer in the power supply. Such transformers don't have to supply a lot of current, though, so they are not very big or heavy. Another type of device that needs rather high voltage is a ham-radio amplifier with vacuum tubes. Such an amplifier requires from 2 kV to 5 kV.

Any voltage higher than about 12 V should be treated with respect. *Warning: The voltages in televisions and ham radios can present an electrocution hazard, even after the equipment has been switched off. Do not try to service such equipment unless you are trained to do so!*

At Audio Frequencies

Transformers for use at AF are similar to those employed for 60-Hz electricity. The differences are that the frequency is somewhat higher (up to 20 kHz), and that audio signals exist in a band of frequencies (20 Hz to 20 kHz) rather than at only one frequency.

Most AF transformers are constructed like miniature utility transformers. They have laminated E cores with primary and secondary windings wound around the crossbars, as shown in Fig. 18-4. Audio transformers can be either the step-up or the step-down type. However, rather than being made to produce a specific voltage, AF transformers are designed to match impedances.

Audio circuits, and in fact all electronic circuits that handle sine-wave or complex-wave signals, exhibit impedance at the input and output. The load has a certain impedance; a source has another impedance. Good audio design strives to minimize the reactance in the circuitry, so that the absolute-value impedance Z is close to the resistance R. This means that X must be zero or nearly zero. In the following discussion of impedance-matching transformers, for both AF and RF applications, assume that the reactance is zero, so the impedance is purely resistive with $Z = R + j0$.

Isolation and Impedance Matching

Transformers can provide *isolation* between electronic circuits. While there is *inductive coupling* in a transformer, there is comparatively little *capacitive coupling.* The amount of capacitive coupling can be reduced by using cores that minimize the number of wire turns needed in the windings, and by keeping the windings physically separated from each other (rather than overlapping).

Balanced and Unbalanced Loads and Lines

A *balanced load* is one whose terminals can be reversed without significantly affecting circuit behavior. A plain resistor is a good example. The two-wire antenna input in a television receiver is another example of a balanced load. A *balanced transmission line* is usually a two-wire line, such as old-fashioned *TV ribbon*, also called *twinlead.*

An *unbalanced load* is a load that must be connected a certain way. Switching its leads will result in improper circuit operation. In this sense, an unbalanced load is a little like a polarized component such as a battery or capacitor. Many wireless antennas are of this type. Usually, unbalanced sources and loads have one side connected to ground. The coaxial input of a television receiver is unbalanced; the shield (braid) of the cable is grounded. An *unbalanced transmission line* is usually a coaxial line, such as you find in a cable television system.

Normally, you cannot connect an unbalanced line to a balanced load, or a balanced line to an unbalanced load, and expect good performance. But a transformer can allow for mating between these two types of systems. In Fig. 18-10A, a *balanced-to-unbalanced transformer* is shown. Note that the balanced side is center-tapped, and the tap is grounded. In Fig. 18-10B, an *unbalanced-to-balanced transformer* is illustrated. Again, the balanced side has a grounded center tap.

The turns ratio of a balanced-to-unbalanced transformer (also called a *balun*) or an unbalanced-to-balanced transformer (also known as an *unbal*) can be 1:1, but this need not be the case, and often it is not. If the impedances of the balanced and unbalanced parts of the systems are the same, then a 1:1 turns ratio is ideal. But if the impedances differ, the turns ratio should be such that the impedances are matched. Shortly, we'll see how the turns ratio of a transformer can be manipulated to transform one purely resistive impedance into another.

Transformer Coupling

Transformers are sometimes used between amplifier stages in electronic equipment where a large *amplification factor* is needed. There are other methods of *coupling* from one amplifier stage to another, but transformers offer some advantages, especially in RF receivers and transmitters.

Part of the problem in getting a radio to work is that the amplifiers must operate in a stable manner. If there is too much feedback, a series of amplifiers will oscillate, and this will severely degrade the performance of the radio. Transformers that minimize the capacitance between the amplifier stages, while still transferring the desired signals, can help to prevent this oscillation.

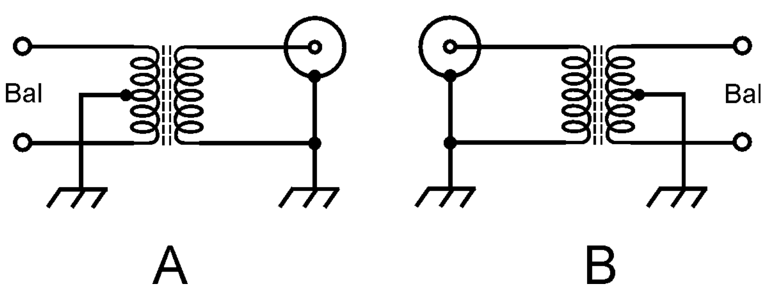

18-10 At A, a balanced-to-unbalanced transformer. At B, an unbalanced-to-balanced transformer.

Impedance Transfer Ratio

In RF and AF systems, transformers are employed to match impedances. Thus, you will sometimes hear or read about an *impedance step-up transformer* or an *impedance step-down transformer.*

The *impedance transfer ratio* of a transformer varies according to the square of the turns ratio, and also according to the square of the voltage-transfer ratio. If the primary (source) and secondary (load) impedances are purely resistive and are denoted Z_{pri} and Z_{sec}, then the following relations hold:

$$Z_{\text{pri}}/Z_{\text{sec}} = (T_{\text{pri}}/T_{\text{sec}})^2$$
$$Z_{\text{pri}}/Z_{\text{sec}} = (E_{\text{pri}}/E_{\text{sec}})^2$$

The inverses of these formulas, in which the turns ratio or voltage-transfer ratio are expressed in terms of the impedance-transfer ratio, are:

$$T_{\text{pri}}/T_{\text{sec}} = (Z_{\text{pri}}/Z_{\text{sec}})^{1/2}$$
$$E_{\text{pri}}/E_{\text{sec}} = (Z_{\text{pri}}/Z_{\text{sec}})^{1/2}$$

Problem 18-3

Consider a situation in which a transformer is needed to match an input impedance of 50.0 Ω, purely resistive, to an output impedance of 300 Ω, also purely resistive. What is the required turns ratio $T_{\text{pri}}/T_{\text{sec}}$?

The required transformer will have a step-up impedance ratio of $Z_{\text{pri}}/Z_{\text{sec}} = 50.0/300 = 1/6.00$. From the preceding formulas:

$$\begin{aligned} T_{\text{pri}}/T_{\text{sec}} &= (Z_{\text{pri}}/Z_{\text{sec}})^{1/2} \\ &= (1/6.00)^{1/2} \\ &= 0.16667^{1/2} \\ &= 0.408 \\ &= 1/2.45 \end{aligned}$$

Problem 18-4

Suppose a transformer has a primary-to-secondary turns ratio of 4.00:1. The load, connected to the transformer output, is a pure resistance of 37.5 Ω. What is the impedance at the primary?

The impedance-transfer ratio is equal to the square of the turns ratio. Therefore:

$$\begin{aligned} Z_{\text{pri}}/Z_{\text{sec}} &= (T_{\text{pri}}/T_{\text{sec}})^2 \\ &= (4.00/1)^2 \\ &= 4.00^2 \\ &= 16.0 \end{aligned}$$

We know that the secondary impedance, Z_{sec} is 37.5 Ω. Thus:

$$\begin{aligned} Z_{\text{pri}} &= 16.0 \times Z_{\text{sec}} \\ &= 16.0 \times 37.5 \\ &= 600\ \Omega \end{aligned}$$

Radio-Frequency Transformers

In radio receivers and transmitters, transformers can be categorized generally by the method of construction used. Some have primary and secondary windings, just like utility and audio units. Others employ transmission-line sections. These are the two most common types of transformer found at radio frequencies.

Wire-Wound Types

In wire-wound RF transformers, powdered-iron cores can be used up to quite high frequencies. Toroidal cores are common, because they are self-shielding (all of the magnetic flux is confined within the core material). The number of turns depends on the frequency, and also on the permeability of the core.

In high-power applications, air-core coils are often preferred. Although air has low permeability, it has negligible hysteresis loss, and will not heat up or fracture as powdered-iron cores sometimes do. The disadvantage of air-core coils is that some of the magnetic flux extends outside of the coil. This affects the performance of the transformer when it must be placed in a cramped space, such as in a transmitter final-amplifier compartment.

A major advantage of coil-type transformers, especially when they are wound on toroidal cores, is that they can be made to work over a wide band of frequencies, such as from 3.5 MHz to 30 MHz. These are called *broadband transformers.*

Transmission-Line Types

As you recall, any transmission line has a characteristic impedance, or Z_o, that depends on the line construction. This property is sometimes used to make impedance transformers out of coaxial or parallel-wire line.

Transmission-line transformers are always made from quarter-wave sections. From the previous chapter, remember the formula for the length of a quarter-wave section:

$$L_{ft} = 246v/f_o$$

where L_{ft} is the length of the section in feet, v is the velocity factor expressed as a fraction, and f_o is the frequency of operation in megahertz. If the length L_m is specified in meters, then:

$$L_m = 75v/f_o$$

Suppose that a quarter-wave section of line, with characteristic impedance Z_o, is terminated in a purely resistive impedance R_{out}. Then the impedance that appears at the input end of the line, R_{in}, is also a pure resistance, and the following relations hold:

$$Z_o^2 = R_{in}R_{out}$$
$$Z_o = (R_{in}R_{out})^{1/2}$$

This is illustrated in Fig. 18-11. The first of the preceding formulas can be rearranged to solve for R_{in} in terms of R_{out}, or vice versa:

$$R_{in} = Z_o^2/R_{out}$$
$$R_{out} = Z_o^2/R_{in}$$

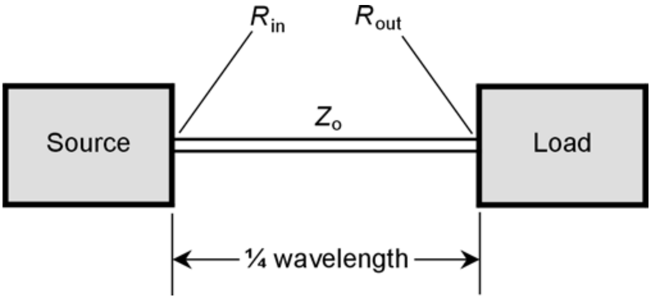

18-11 A quarter-wave matching section of transmission line. The input impedance is R_{in}, the output impedance is R_{out}, and the characteristic impedance of the line is Z_o.

These equations are valid at the frequency f_o for which the line length measures $\frac{1}{4}$ wavelength. Sometimes, the word "wavelength" is replaced by the lowercase Greek letter lambda (λ), so you will occasionally see the length of a quarter-wave section denoted as $(\frac{1}{4})\lambda$ or 0.25λ.

Neglecting line losses, the preceding relations hold at all *odd harmonics* of f_o, that is, at $3f_o$, $5f_o$, $7f_o$, and so on. At other frequencies, a quarter-wave section of line does not act as a transformer. Instead, it behaves in a complex manner that is beyond the scope of this discussion.

Quarter-wave transmission-line transformers are most often used in antenna systems, especially at the higher frequencies, where their dimensions become practical. A quarter-wave matching section should be made using unbalanced line if the load is unbalanced, and balanced line if the load is balanced.

A disadvantage of quarter-wave sections is the fact that they work only at specific frequencies. But this is often offset by the ease with which they are constructed, if radio equipment is to be used at only one frequency, or at odd-harmonic frequencies.

Problem 18-5

Suppose an antenna has a purely resistive impedance of 100 Ω. It is connected to a $\frac{1}{4}$-wave section of 75-Ω coaxial cable. What is the impedance at the input end of the section?

Use the formula from above:

$$\begin{aligned} R_{in} &= Z_o^2/R_{out} \\ &= 75^2/100 \\ &= 5625/100 \\ &= 56\ \Omega \end{aligned}$$

Problem 18-6

Consider an antenna known to have a purely resistive impedance of 600 Ω. You want to match it to the output of a radio transmitter designed to work into a 50.0-Ω pure resistance. What is the characteristic impedance needed for a quarter-wave matching section?

Use this formula:

$$\begin{aligned} Z^2 &= R_{in}R_{out} \\ &= 600 \times 50 \\ &= 30{,}000 \end{aligned}$$

Therefore:

$$
\begin{aligned}
Z_o &= (30{,}000)^{1/2} \\
&= 173\ \Omega
\end{aligned}
$$

It may be difficult to find a commercially manufactured transmission line that has this particular characteristic impedance. Prefabricated lines come in standard Z_o values, and a perfect match might not be obtainable. In that case, the closest obtainable Z_o should be used. In this case, it would probably be 150 Ω. If nothing is available anywhere near the characteristic impedance needed for a quarter-wave matching section, then a coil-type transformer can be used instead.

What about Reactance?

Things are simple when there is no reactance in an ac circuit using transformers. But often, especially in RF antenna systems, pure resistance doesn't occur naturally. It has to be obtained by using inductors and/or capacitors to cancel the reactance out. The presence of reactance in a load makes a perfect match impossible with an impedance-matching transformer alone.

Recall that inductive and capacitive reactances are opposite in effect, and that their magnitudes can vary. If a load presents a complex impedance $R + jX$, it is possible to cancel the reactance X by deliberately introducing an equal and opposite reactance $-X$. This can be, and often is, done by connecting an inductor or capacitor in series with a load that contains reactance as well as resistance. The result is a pure resistance with a value equal to $(R + jX) - jX$, or simply R.

When wireless communications is contemplated over a wide band of frequencies, adjustable impedance-matching and reactance-canceling networks can be placed between the transmitter and the antenna system. Such a device is called a *transmatch* or an *antenna tuner.* These devices not only match the resistive portions of the transmitter and load impedances, but they can tune out reactances in the load. Transmatches are popular among amateur radio operators, who use equipment capable of operation from less than 2 MHz up to the highest known radio frequencies.

Quiz

Refer to the text in this chapter if necessary. A good score is 18 or more correct. Answers are in the back of the book.

1. In a step-up transformer,
   (a) the primary impedance is greater than the secondary impedance.
   (b) the secondary winding is right on top of the primary.
   (c) the primary voltage is less than the secondary voltage.
   (d) All of the above are true.

2. The capacitance between the primary and the secondary windings of a transformer can be minimized by
   (a) placing the windings on opposite sides of a toroidal core.
   (b) winding the secondary right on top of the primary.
   (c) using the highest possible frequency.
   (d) using a center tap on the balanced winding.

3. A transformer steps a voltage down from 117 V to 6.00 V. What is its primary-to-secondary turns ratio?
   - (a) 1:380
   - (b) 380:1
   - (c) 1:19.5
   - (d) 19.5:1

4. A step-up transformer has a primary-to-secondary turns ratio of 1:5.00. If 117 V rms appears at the primary, what is the ac rms voltage across the secondary?
   - (a) 23.4 V rms
   - (b) 585 V rms
   - (c) 117 V rms
   - (d) 2.93 kV rms

5. A transformer has a secondary-to-primary turns ratio of 0.167. This transformer is
   - (a) a step-up unit.
   - (b) a step-down unit.
   - (c) neither a step-up unit nor a step-down unit.
   - (d) a reversible unit.

6. Which of the following statements is false, concerning air cores compared with ferromagnetic cores?
   - (a) Air concentrates the magnetic lines of flux.
   - (b) Air works at higher frequencies than ferromagnetics.
   - (c) Ferromagnetics are lossier than air.
   - (d) A ferromagnetic-core transformer needs fewer turns of wire than an equivalent air-core transformer.

7. Eddy currents cause
   - (a) an increase in efficiency.
   - (b) an increase in coupling between windings.
   - (c) an increase in core loss.
   - (d) an increase in usable frequency range.

8. Suppose a transformer has an ac voltage of 117 V rms across its primary, and 234 V rms appears across its secondary. If this transformer is reversed (that is, connected backward), assuming that this be done without damaging the windings, what will be the voltage at the output?
   - (a) 234 V rms
   - (b) 468 V rms
   - (c) 117 V rms
   - (d) 58.5 V rms

9. The shell method of transformer winding
   - (a) provides maximum coupling.
   - (b) minimizes capacitance between windings.

(c) withstands more voltage than other winding methods.
(d) has windings far apart but along a common axis.

10. Which of these core types is best if you need a winding inductance of 1.5 H?
    (a) Air core
    (b) Ferromagnetic solenoid core
    (c) Ferromagnetic toroid core
    (d) Ferromagnetic pot core

11. An advantage of a toroid core over a solenoid core is the fact that
    (a) the toroid works at higher frequencies.
    (b) the toroid confines the magnetic flux.
    (c) the toroid can work for dc as well as for ac.
    (d) it is easier to wind the turns on a toroid.

12. High voltage is used in long-distance power transmission because
    (a) it is easier to regulate than low voltage.
    (b) the I^2R losses are minimized.
    (c) the electromagnetic fields are strong.
    (d) small transformers can be used.

13. In a household circuit, 234-V rms electricity usually has
    (a) one phase.
    (b) two phases.
    (c) three phases.
    (d) four phases.

14. In a transformer, a center tap often exists in
    (a) the primary winding.
    (b) the secondary winding.
    (c) an unbalanced winding.
    (d) a balanced winding.

15. An autotransformer
    (a) can be adjusted automatically.
    (b) has a center-tapped secondary.
    (c) consists of a single tapped winding.
    (d) is useful only for impedance matching.

16. Suppose a transformer has a primary-to-secondary turns ratio of 2.00:1. The input impedance is 300 Ω, purely resistive. What is the output impedance?
    (a) 75 Ω, purely resistive
    (b) 150 Ω, purely resistive
    (c) 600 Ω, purely resistive
    (d) 1200 Ω, purely resistive

17. Suppose a purely resistive input impedance of 50 Ω must be matched to a purely resistive output impedance of 450 Ω. The primary-to-secondary turns ratio of the transformer must be which of the following?

    (a) 9.00
    (b) 3.00
    (c) 1/3.00
    (d) 1/9.00

18. Suppose a quarter-wave matching section has a characteristic impedance of 75.0 Ω. The input impedance is 50.0 Ω, purely resistive. What is the output impedance?

    (a) 150 Ω, purely resistive
    (b) 125 Ω, purely resistive
    (c) 100 Ω, purely resistive
    (d) 113 Ω, purely resistive

19. Suppose a purely resistive impedance of 75 Ω must be matched to a purely resistive impedance of 300 Ω. A quarter-wave section would need to have

    (a) $Z_o = 188\ \Omega$.
    (b) $Z_o = 150\ \Omega$.
    (c) $Z_o = 225\ \Omega$.
    (d) $Z_o = 375\ \Omega$.

20. If there is reactance in the load to which a transformer is connected, then

    (a) the transformer will be destroyed.
    (b) a perfect impedance match cannot be obtained.
    (c) a center tap must be used in the secondary.
    (d) the turns ratio must be changed to obtain an impedance match.

Test: Part 2

Do not refer to the text when taking this test. A good score is at least 37 correct. Answers are in the back of the book. It's best to have a friend check your score the first time, so you won't memorize the answers if you want to take the test again.

1. Consider a series circuit that has a resistance of 100 Ω and a capacitive reactance of −200 Ω. What is the complex impedance?
   (a) $-200 + j100$
   (b) $100 + j200$
   (c) $200 - j100$
   (d) $200 + j100$
   (e) $100 - j200$

2. Mutual inductance causes the net value of a set of coils to
   (a) cancel out, resulting in zero inductance.
   (b) be greater than what it would be with no mutual coupling.
   (c) be less than what it would be with no mutual coupling.
   (d) double.
   (e) vary, depending on the extent and phase of mutual coupling.

3. Refer to Fig. Test 2-1. Wave *A* is
   (a) leading wave *B* by 90°.
   (b) lagging wave *B* by 90°.
   (c) leading wave *B* by 180°.
   (d) lagging wave *B* by 135°.
   (e) lagging wave *B* by 45°.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

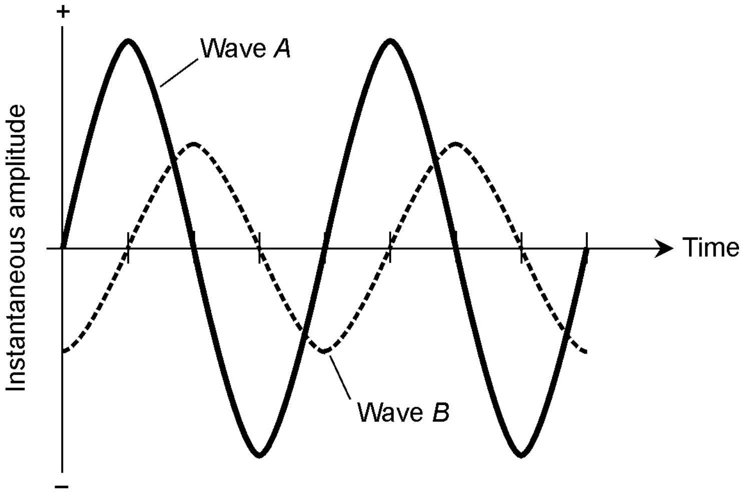

Test 2-1 Illustration for Part 2 Test Question 3.

4. If a pure sine wave with no dc component has a positive peak value of +30.0 V pk, what is its rms voltage?
   (a) 21.2 V rms
   (b) 30.0 V rms
   (c) 42.4 V rms
   (d) 60.0 V rms
   (e) 90.0 V rms

5. Suppose four capacitors are connected in parallel. Their values are 100 pF each. What is the net capacitance?
   (a) 25 pF
   (b) 50 pF
   (c) 100 pF
   (d) 200 pF
   (e) 400 pF

6. Suppose an ac transformer has a primary-to-secondary turns ratio of 8.88/1. The input voltage is 234 V rms. What is the output voltage?
   (a) 2.08 kV rms
   (b) 18.5 kV rms
   (c) 2.97 V rms
   (d) 26.4 V rms
   (e) 20.8 V rms

7. In a series *RL* circuit, as the resistance becomes small compared with the reactance, the angle of lag approaches which of the following?
   (a) 0°
   (b) 45°

(c) 90°
(d) 180°
(e) 360°

8. Suppose an ac transmission line carries 3.50 A rms and 150 V rms. Imagine that the line is perfectly lossless, and that the load impedance is a pure resistance equal to the characteristic impedance of the line. What is the *true power* in this transmission line?
(a) 525 W
(b) 42.9 W
(c) 1.84 W
(d) Nonexistent, because true power is dissipated, not transmitted
(e) Variable, depending on standing-wave effects

9. In a parallel configuration, susceptances
(a) simply add up.
(b) add like capacitances in series.
(c) add like inductances in parallel.
(d) must be changed to reactances before you can work with them.
(e) cancel out.

10. Consider a sine wave that has a frequency of 200 kHz. How many degrees of phase change occur in a microsecond (a millionth of a second)?
(a) 180°
(b) 144°
(c) 120°
(d) 90°
(e) 72°

11. At a frequency of 2.55 MHz, what is the reactance of a 330-pF capacitor?
(a) −5.28 Ω
(b) −0.00528 Ω
(c) −189 Ω
(d) −18.9 kΩ
(e) −0.000189 Ω

12. Suppose a transformer has a step-up turns ratio of 1/3.16. The impedance of the load connected to the secondary is 499 Ω, purely resistive. What is the impedance at the primary?
(a) 50.0 Ω, purely resistive
(b) 158 Ω, purely resistive
(c) 1.58 kΩ, purely resistive
(d) 4.98 kΩ, purely resistive
(e) Impossible to calculate from the data given

13. If a complex impedance is represented by $34 - j23$, what is the absolute-value impedance?
    (a) 34 Ω
    (b) 11 Ω
    (c) −23 Ω
    (d) 41 Ω
    (e) 57 Ω

14. Suppose a coil has an inductance of 750 μH. What is the inductive reactance at 100 kHz?
    (a) 75.0 Ω
    (b) 75.0 kΩ
    (c) 471 Ω
    (d) 47.1 kΩ
    (e) 212 Ω

15. If two sine waves are 180° out of phase, it represents a difference of
    (a) ⅛ of a cycle.
    (b) ¼ of a cycle.
    (c) ½ of a cycle.
    (d) 1 full cycle.
    (e) 2 full cycles.

16. If *R* denotes resistance and *Z* denotes absolute-value impedance, then *R*/*Z* represents the
    (a) true power.
    (b) imaginary power.
    (c) apparent power.
    (d) absolute-value power.
    (e) power factor.

17. Suppose two components are connected in series. One component has a complex impedance of $30 + j50$, and the other component has a complex impedance of $50 - j30$. What is the impedance of the series combination?
    (a) $80 + j80$
    (b) $20 + j20$
    (c) $20 - j20$
    (d) $-20 + j20$
    (e) $80 + j20$

18. Suppose two inductors, having values of 140 μH and 1.50 mH, are connected in series. What is the net inductance?
    (a) 141.5 μH
    (b) 1.64 μH
    (c) 0.1415 mH

(d) 1.64 mH
(e) 0.164 mH

19. Which of the following types of capacitor is polarized?
(a) Mica
(b) Paper
(c) Electrolytic
(d) Air variable
(e) Ceramic

20. A coil with a toroidal, ferromagnetic core
(a) has less inductance than an air-core coil with the same number of turns.
(b) is essentially self-shielding.
(c) works well as a loopstick antenna.
(d) is ideal as a transmission-line transformer.
(e) cannot be used at frequencies below 10 MHz.

21. The efficiency of an electric generator
(a) depends on the mechanical driving power source.
(b) is equal to the electrical output power divided by the mechanical input power.
(c) depends on the nature of the electrical load.
(d) is equal to driving voltage divided by output voltage.
(e) is equal to driving current divided by output current.

22. Admittance is
(a) the reciprocal of reactance.
(b) the reciprocal of resistance.
(c) a measure of the opposition a circuit offers to ac.
(d) a measure of the ease with which a circuit passes ac.
(e) another expression for absolute-value impedance.

23. The absolute-value impedance Z of a parallel RLC circuit, where R is the resistance and X is the net reactance, is found according to which of the following formulas?
(a) $Z = R + X$
(b) $Z^2 = R^2 + X^2$
(c) $Z^2 = R^2X^2/(R^2 + X^2)$
(d) $Z = 1/(R^2 + X^2)$
(e) $Z = R^2X^2/(R + X)$

24. Complex numbers are used to represent impedance because
(a) reactance cannot store power.
(b) reactance isn't a real physical thing.

(c) they provide a way to represent what happens in resistance-reactance circuits.
(d) engineers like to work with sophisticated mathematics.
(e) Forget it! Complex numbers are never used to represent impedance.

25. Which of the following (within reason) has no effect on the value, in farads, of a capacitor?
    (a) The mutual surface area of the plates
    (b) The dielectric constant of the material between the plates
    (c) The spacing between the plates
    (d) The amount of overlap between plates
    (e) The frequency

26. The 0° phase point in an ac sine wave is usually considered to be the point in time at which the instantaneous amplitude is
    (a) zero and negative-going.
    (b) at its negative peak.
    (c) zero and positive-going.
    (d) at its positive peak.
    (e) any value; it doesn't matter.

27. The inductance of a coil can be adjusted in a practical way by
    (a) varying the frequency of the signal applied to the coil.
    (b) varying the number of turns using multiple taps.
    (c) varying the current in the coil.
    (d) varying the wavelength of the signal applied to the coil.
    (e) varying the voltage across the coil.

28. Power factor is defined as the ratio of
    (a) true power to VA power.
    (b) true power to imaginary power.
    (c) imaginary power to VA power.
    (d) imaginary power to true power.
    (e) VA power to true power.

29. Consider a situation in which you want to match a feed line with $Z_o = 50\ \Omega$ to an antenna with a purely resistive impedance of $200\ \Omega$. A quarter-wave matching section should have which of the following?
    (a) $Z_o = 150\ \Omega$
    (b) $Z_o = 250\ \Omega$
    (c) $Z_o = 125\ \Omega$
    (d) $Z_o = 133\ \Omega$
    (e) $Z_o = 100\ \Omega$

30. The vector $40 + j30$ in the *RX* plane represents
    (a) 40 Ω of resistance and 30 μH of inductance.
    (b) 40 μH of inductance and 30 Ω of resistance.
    (c) 40 Ω of resistance and 30 Ω of inductive reactance.
    (d) 40 Ω of inductive reactance and 30 Ω of resistance.
    (e) 40 μH of inductive reactance and 30 Ω of resistance.

31. In a series *RC* circuit where $R = 300\ \Omega$ and $X_C = -30\ \Omega$,
    (a) the current leads the voltage by a few degrees.
    (b) the current leads the voltage by almost 90°.
    (c) the voltage leads the current by a few degrees.
    (d) the voltage leads the current by almost 90°.
    (e) the voltage leads the current by 90°.

32. In a step-down transformer,
    (a) the primary voltage is greater than the secondary voltage.
    (b) the purely resistive impedance across the primary is less than the purely resistive impedance across the secondary.
    (c) the secondary voltage is greater than the primary voltage.
    (d) the output frequency is higher than the input frequency.
    (e) the output frequency is lower than the input frequency.

33. Suppose a capacitor of 470 pF is in parallel with an inductor of 4.44 μH. What is the resonant frequency?
    (a) 3.49 MHz
    (b) 3.49 kHz
    (c) 13.0 MHz
    (d) 13.0 GHz
    (e) It cannot be calculated from the data given.

34. A pure sine wave contains energy at
    (a) only one specific frequency.
    (b) a specific frequency and its even harmonics.
    (c) a specific frequency and its odd harmonics.
    (d) a specific frequency and all its harmonics.
    (e) a specific frequency and its second harmonic only.

35. Inductive susceptance is
    (a) the reciprocal of inductance.
    (b) negative imaginary.
    (c) equivalent to capacitive reactance.
    (d) the reciprocal of capacitive susceptance.
    (e) positive imaginary.

36. The rate of change (derivative) of a pure sine wave is another pure sine wave that has the same frequency as the original wave, and

(a) is in phase with the original wave.
(b) is 180° out of phase with the original wave.
(c) leads the original wave by 45°.
(d) lags the original wave by 90°.
(e) leads the original wave by 90°.

37. True power is equal to

(a) VA power plus imaginary power.
(b) imaginary power minus VA power.
(c) the vector difference between VA and reactive power.
(d) VA power; the two are the same thing.
(e) 0.707 times the VA power.

38. Consider a circuit in which three capacitors are connected in series. Their values are 47 μF, 68 μF, and 100 μF. The total capacitance of this combination is

(a) 215 μF.
(b) between 68 μF and 100 μF.
(c) between 47 μF and 68 μF.
(d) 22 μF.
(e) not determinable from the data given.

39. The reactance of a section of transmission line depends on all of the following factors *except*

(a) the velocity factor of the line.
(b) the length of the section.
(c) the current in the line.
(d) the frequency of the signal in the line.
(e) the wavelength of the signal in the line.

40. When analyzing a parallel *RLC* circuit to find the complex impedance, you should

(a) add the resistance and reactance to get $R + jX$.
(b) find the net conductance and susceptance, convert to resistance and reactance, and then add these to get $R + jX$.
(c) find the net conductance and susceptance, and add these to get $R + jX$.
(d) rearrange the components so they're connected in series, and find the complex impedance of that circuit.
(e) subtract reactance from resistance to get $R - jX$.

41. The illustration in Fig. Test 2-2 shows a vector $R + jX$ representing

(a) $X_C = 60\ \Omega$ and $R = 25\ \Omega$.
(b) $X_L = 60\ \Omega$ and $R = 25\ \Omega$.
(c) $X_L = 60\ \mu\text{H}$ and $R = 25\ \Omega$.

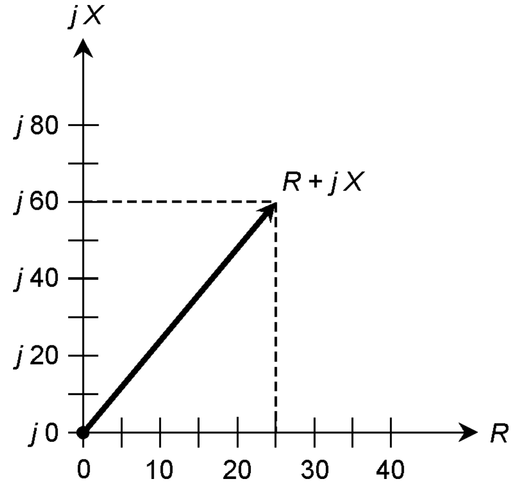

Test 2-2 Illustration for Part 2 Test Question 41.

(d) $C = 60$ µF and $R = 25$ Ω.
(e) $L = 60$ µH and $R = 25$ Ω.

42. Suppose two pure sine waves have no dc components, have the same frequency, and have the same peak-to-peak voltages, but they cancel each other out when combined. What is the phase difference between the waves?

(a) 45°
(b) 90°
(c) 180°
(d) 270°
(e) 360°

43. Suppose a series *RC* circuit has a resistance of 50 Ω and a capacitive reactance of −37 Ω. What is the phase angle?

(a) 37°
(b) 53°
(c) −37°
(d) −53°
(e) It cannot be calculated from the data given.

44. Suppose a 200-Ω resistor is in series with a coil and capacitor, such that $X_L = 200$ Ω and $X_C = -100$ Ω. What is the complex impedance?

(a) $200 - j100$
(b) $200 - j200$
(c) $200 + j100$
(d) $200 + j200$
(e) Impossible to determine from the data given

45. The characteristic impedance of a transmission line
    - (a) is negative imaginary.
    - (b) is positive imaginary.
    - (c) depends on the frequency.
    - (d) depends on the construction of the line.
    - (e) depends on the length of the line.

46. Suppose the period of a pure sine wave is 2×10^{-8}s. What is the frequency?
    - (a) 2×10^8 Hz
    - (b) 20 MHz
    - (c) 50 kHz
    - (d) 50 MHz
    - (e) 500 MHz

47. Suppose a series *RC* circuit has a resistance of 600 Ω and a capacitance of 220 pF. What is the phase angle?
    - (a) −20°
    - (b) 20°
    - (c) −70°
    - (d) 70°
    - (e) Not determinable from the data given

48. A capacitor with a negative temperature coefficient
    - (a) works less well as the temperature increases.
    - (b) works better as the temperature increases.
    - (c) heats up as its value is made larger.
    - (d) cools down as its value is made larger.
    - (e) exhibits increasing capacitance as the temperature drops.

49. Suppose three coils are connected in parallel. Each has an inductance of 300 μH. There is no mutual inductance. What is the net inductance?
    - (a) 100 μH
    - (b) 300 μH
    - (c) 900 μH
    - (d) 17.3 μH
    - (e) 173 μH

50. Suppose a coil has 100 Ω of inductive reactance at 30.0 MHz. What is its inductance?
    - (a) 0.531 μH
    - (b) 18.8 mH
    - (c) 531 μH
    - (d) 18.8 μH
    - (e) It can't be found from the data given.

3
PART
Basic Electronics

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

This page intentionally left blank

19
CHAPTER

Introduction to Semiconductors

SINCE THE 1960S, WHEN THE TRANSISTOR BECAME COMMON IN CONSUMER DEVICES, *SEMICONDUCTORS* have acquired a dominating role in electronics. The term *semiconductor* arises from the ability of these materials to conduct some of the time, but not all the time. The conductivity can be controlled to produce effects such as amplification, rectification, oscillation, signal mixing, and switching.

The Semiconductor Revolution

Decades ago, *vacuum tubes*, also known as *electron tubes*, were the only devices available for use as amplifiers, oscillators, detectors, and other electronic circuits and systems. A typical tube (called a *valve* in England) ranged from the size of your thumb to the size of your fist. They are still used in some power amplifiers, microwave oscillators, and video display units.

Tubes generally require high voltage. Even in modest radio receivers, 100 V to 200 V dc was required when tubes were employed. This mandated bulky power supplies, and created an electrical shock hazard. Nowadays, a transistor of microscopic dimensions can perform the functions of a tube in most situations. The power supply can be a couple of AA cells or a 9-V transistor battery.

Even in high-power applications, transistors are smaller and lighter than tubes. Figure 19-1 is a size comparison drawing between a transistor and a vacuum tube for use in an AF or RF power amplifier.

Integrated circuits (ICs), hardly larger than individual transistors, can do the work of hundreds or even thousands of vacuum tubes. An excellent example of this technology is found in personal computers and the peripheral devices used with them.

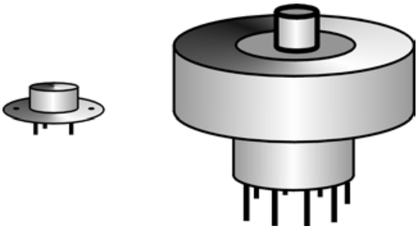

19-1 A power-amplifier transistor (at left) is much smaller than a vacuum tube of comparable power-handling capacity (right).

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

Semiconductor Materials

Various elements, compounds, and mixtures can function as semiconductors. The two most common materials are *silicon* and a compound of gallium and arsenic known as *gallium arsenide* (often abbreviated GaAs). In the early years of semiconductor technology, *germanium* formed the basis for many semiconductors; today it is seen occasionally, but not often. Other substances that work as semiconductors are *selenium*, *cadmium* compounds, *indium* compounds, and the oxides of certain metals.

Silicon

Silicon (chemical symbol Si) is widely used in diodes, transistors, and integrated circuits. Generally, other substances, or *impurities*, must be added to silicon to give it the desired properties. The best quality silicon is obtained by growing crystals in a laboratory. The silicon is then fabricated into *wafers* or *chips.*

Gallium Arsenide

Another common semiconductor is the compound gallium arsenide. Engineers and technicians call this material by its acronym-like chemical symbol, GaAs, pronounced "gas." If you hear about "gasfets" and "gas ICs," you're hearing about gallium-arsenide technology.

GaAs devices require little voltage, and will function at higher frequencies than silicon devices because the charge carriers move faster through the semiconductor material. GaAs devices are relatively immune to the effects of ionizing radiation such as X rays and gamma rays. GaAs is used in light-emitting diodes (LEDs), infrared-emitting diodes (IREDs), laser diodes, visible-light and infrared (IR) detectors, ultra-high-frequency (UHF) amplifying devices, and a variety of integrated circuits.

Selenium

Selenium exhibits conductivity that varies depending on the intensity of visible light or IR radiation that strikes it. All semiconductor materials exhibit this property, known as *photoconductivity*, to some degree; but in selenium the effect is especially pronounced. For this reason, selenium is useful for making *photocells.* Selenium is also used in certain types of *rectifiers.* A rectifier is a component or circuit that converts ac to pulsating dc.

A significant advantage of selenium is the fact that it is electrically rugged. Selenium-based components can withstand brief *transients*, or spikes, of abnormally high voltage, better than components made with most other semiconductor materials.

Germanium

Pure elemental germanium is a poor electrical conductor. It becomes a semiconductor only when *impurities* are added. Germanium was used extensively in the early years of semiconductor technology. Some diodes and transistors still use it.

A germanium diode has a low voltage drop (0.3 V, compared with 0.6 V for silicon and 1 V for selenium) when it conducts, and this makes it useful in some situations. But germanium is easily destroyed by heat. Extreme care must be used when soldering the leads of a germanium component.

Metal Oxides

Certain metal oxides have properties that make them useful in the manufacture of semiconductor devices. When you hear about MOS (pronounced "moss") or CMOS (pronounced "sea moss") technology, you are hearing about *metal-oxide semiconductor* and *complementary metal-oxide semiconductor* devices, respectively.

An advantage of MOS and CMOS devices is the fact that they need almost no power to function. They draw so little current that a battery in a MOS or CMOS device lasts just about as long as it would on the shelf. Another advantage is high speed. This allows operation at high frequencies in RF equipment, and makes it possible to perform many switching operations per second for use in computers.

Certain types of transistors, and many kinds of ICs, make use of this technology. In integrated circuits, MOS and CMOS allow for a large number of discrete diodes and transistors on a single chip. Engineers would say that MOS/CMOS has *high component density.*

The biggest problem with MOS and CMOS technology is the fact that the devices are easily damaged by static electricity. Care must be used when handling components of this type. Technicians working with MOS and CMOS components must literally ground themselves by wearing a metal wrist strap connected to a good earth ground. Otherwise, the electrostatic charges that normally build up on their bodies can destroy MOS and CMOS components when equipment is constructed or serviced.

Doping and Charge Carriers

For a semiconductor material to have the properties necessary in order to function as electronic components, impurities are usually added. The impurities cause the material to conduct currents in certain ways. The addition of an impurity to a semiconductor is called *doping.* Sometimes the impurity is called a *dopant.*

Donor Impurities

When an impurity contains an excess of electrons, the dopant is called a *donor impurity.* Adding such a substance causes conduction mainly by means of electron flow, as in an ordinary metal such as copper or aluminum. The excess electrons are passed from atom to atom when a voltage exists across the material. Elements that serve as donor impurities include antimony, arsenic, bismuth, and phosphorus. A material with a donor impurity is called an *N-type semiconductor*, because electrons have negative (N) charge.

Acceptor Impurities

If an impurity has a deficiency of electrons, the dopant is called an *acceptor impurity.* When a substance such as aluminum, boron, gallium, or indium is added to a semiconductor, the material conducts by means of *hole flow.* A *hole* is a missing electron—or more precisely, a place in an atom where an electron should be, but isn't. A semiconductor with an acceptor impurity is called a *P-type semiconductor*, because holes have, in effect, a positive (P) charge.

Majority and Minority Carriers

Charge carriers in semiconductor materials are either electrons, each of which has a unit negative charge, or holes, each of which has a unit positive charge. In any semiconductor substance, some

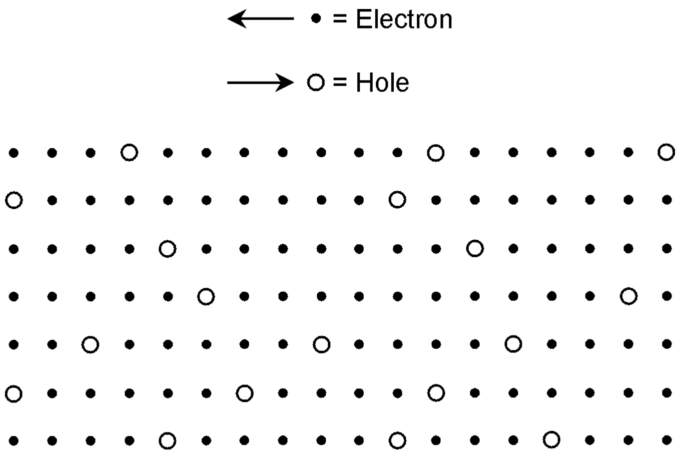

19-2 Pictorial representation of hole flow. Solid black dots represent electrons, moving in one direction. Open circles represent holes, moving in the opposite direction.

of the current takes the form of electrons passed from atom to atom in a negative-to-positive direction, and some of the current occurs as holes that move from atom to atom in a positive-to-negative direction.

Sometimes electrons account for most of the current in a semiconductor. This is the case if the material has donor impurities, that is, if it is of the N type. In other cases, holes account for most of the current. This happens when the material has acceptor impurities, and is thus of the P type. The dominating charge carriers (either electrons or holes) are called the *majority carriers.* The less abundant ones are called the *minority carriers.* The ratio of majority to minority carriers can vary, depending on the way in which the semiconductor material has been manufactured.

Figure 19-2 is a simplified illustration of electron flow versus hole flow in a sample of N-type semiconductor material, where the majority carriers are electrons and the minority carriers are holes. The solid black dots represent electrons. Imagine them moving from right to left in this illustration as they are passed from atom to atom. Small open circles represent holes. Imagine them moving from left to right in the illustration. In this particular example, the positive battery or power-supply terminal (or "source of holes") would be out of the picture toward the left, and the negative battery or power-supply terminal (or "source of electrons") would be out of the picture toward the right.

The P-N Junction

Merely connecting up a piece of semiconducting material, either P or N type, to a source of current can be interesting, and a good subject for science experiments. But when the two types of material are brought together, the boundary between them, called the *P-N junction*, behaves in ways that make semiconductor materials truly useful in electronic components.

The Semiconductor Diode

Figure 19-3 shows the schematic symbol for a *semiconductor diode*, formed by joining a piece of P-type material to a piece of N-type material. The N-type semiconductor is represented by the short, straight line in the symbol, and is called the *cathode.* The P-type semiconductor is represented by the arrow, and is called the *anode.*

19-3 Schematic symbol for a semiconductor diode.

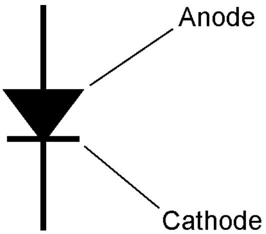

In the diode as shown in Figure 19-3, electrons can move easily in the direction opposite the arrow, and holes can move easily in the direction in which the arrow points. But current cannot, under most conditions, flow the other way. Electrons normally do not move with the arrow, and holes normally do not move against the arrow.

If you connect a battery and a resistor in series with the diode, you'll get a current to flow if the negative terminal of the battery is connected to the cathode and the positive terminal is connected to the anode, as shown in Fig. 19-4A. No current will flow if the battery is reversed, as shown in Fig. 19-4B. (The resistor is included in the circuit to prevent destruction of the diode by excessive current.)

It takes a specific, well-defined minimum applied voltage for conduction to occur through a semiconductor diode. This is called the *forward breakover voltage.* Depending on the type of material, the forward breakover voltage varies from about 0.3 V to 1 V. If the voltage across the junction is not at least as great as the forward breakover voltage, the diode will not conduct, even when it is connected as shown in Fig. 19-4A. This effect, known as the *forward breakover effect* or the *P-N junction threshold effect*, can be of use in circuits designed to limit the positive and/or negative peak voltages that signals can attain. The effect can also be used in a device called a *threshold detector*, in which a signal must be stronger than a certain amplitude in order to pass through.

19-4 Series connection of a battery, a resistor, a current meter, and a diode. At A, forward bias results in a flow of current. At B, reverse bias results in no current.

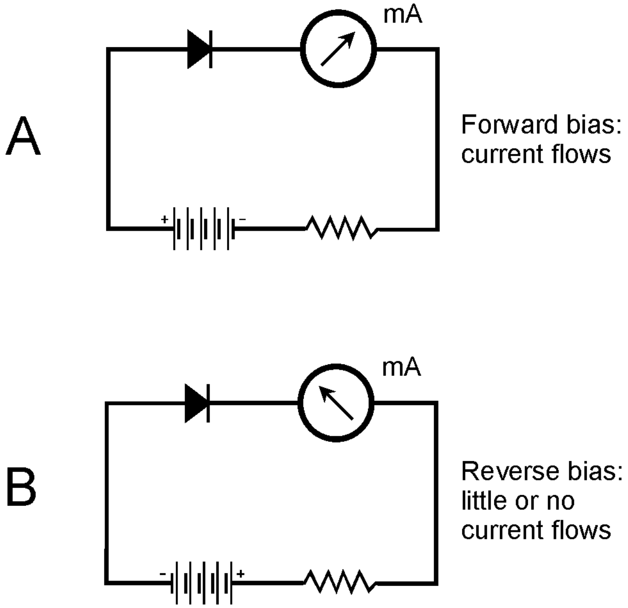

How the Junction Works

When the N-type material is negative with respect to the P type, as in Fig. 19-4A, electrons flow easily from N to P. The N-type semiconductor, which already has an excess of electrons, receives more; the P-type semiconductor, with a shortage of electrons, has some more taken away. The N-type material constantly feeds electrons to the P type in an attempt to create an electron balance, and the battery or power supply keeps robbing electrons from the P-type material. This condition is illustrated in Fig. 19-5A, and is known as *forward bias.* Current can flow through the diode easily under these circumstances.

When the battery or dc power-supply polarity is switched so the N-type material is positive with respect to the P type, the situation is called *reverse bias.* Electrons in the N-type material are pulled toward the positive charge pole, away from the P-N junction. In the P-type material, holes are pulled toward the negative charge pole, also away from the P-N junction. The electrons are the majority carriers in the N-type material, and the holes are the majority carriers in the P-type material. The charge therefore becomes depleted in the vicinity of the P-N junction, and on both sides of it, as shown in Fig. 19-5B. This zone, where majority carriers are deficient, is called the *depletion region.* A shortage of majority carriers in any semiconductor substance means that the substance cannot conduct well. Thus, the depletion region acts like an electrical insulator. This is why a semiconductor diode will not normally conduct when it is reverse-biased. A diode is, in effect, a one-way current gate—usually!

Junction Capacitance

Some P-N junctions can alternate between conduction (in forward bias) and nonconduction (in reverse bias) millions or billions of times per second. Other junctions are slower. The main limiting

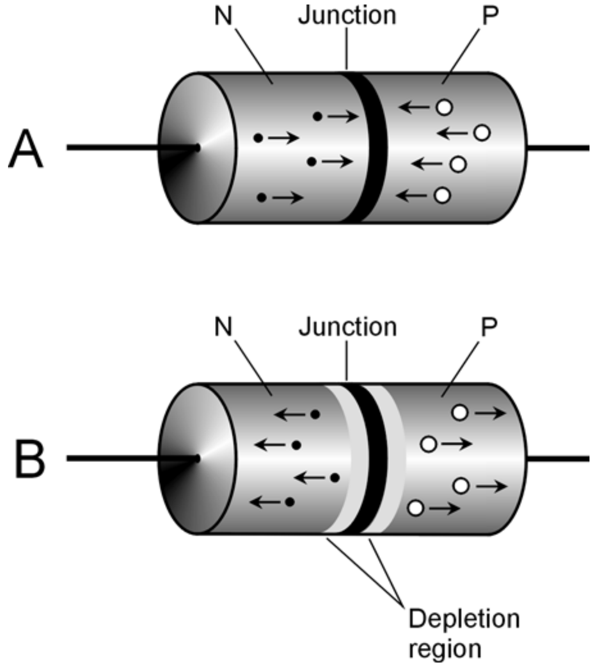

19-5 At A, forward bias of a P-N junction. At B, reverse bias of the same junction. Solid black dots represent electrons. White dots represent holes. Arrows indicate direction of charge-carrier movement.

factor is the capacitance at the P-N junction during conditions of reverse bias. As the *junction capacitance* of a diode increases, maximum frequency at which it can alternate between the conducting state and the nonconducting state decreases.

The junction capacitance of a diode depends on several factors, including the operating voltage, the type of semiconductor material, and the cross-sectional area of the P-N junction. If you examine Fig. 19-5B, you might get the idea that the depletion region, sandwiched between two semiconducting sections, can play a role similar to that of the dielectric in a capacitor. This is true! In fact, a reverse-biased P-N junction actually is a capacitor. Some semiconductor components, called *varactor diodes*, are manufactured with this property specifically in mind.

The junction capacitance of a diode can be varied by changing the reverse-bias voltage, because this voltage affects the width of the depletion region. The greater the reverse voltage, the wider the depletion region gets, and the smaller the capacitance becomes.

Avalanche Effect

Sometimes, a diode conducts when it is reverse-biased. The greater the reverse-bias voltage, the more like an electrical insulator a P-N junction gets—up to a point. But if the reverse bias rises past a specific critical value, the voltage overcomes the ability of the junction to prevent the flow of current, and the junction conducts as if it were forward-biased. This phenomenon is called the *avalanche effect* because conduction occurs in a sudden and massive way, something like a snow avalanche on a mountainside.

The avalanche effect does not damage a P-N junction (unless the voltage is extreme). It's a temporary thing. When the voltage drops back below the critical value, the junction behaves normally again.

Some components are designed to take advantage of the avalanche effect. In other cases, the avalanche effect limits the performance of a circuit. In a device designed for voltage regulation, called a *Zener diode*, you'll hear about the *avalanche voltage* or *Zener voltage* specification. This can range from a couple of volts to well over 100 V. Zener diodes are often used in voltage-regulating circuits.

For *rectifier diodes* in power supplies, you'll hear or read about the *peak inverse voltage* (PIV) or *peak reverse voltage* (PRV) specification. It's important that rectifier diodes have PIV ratings great enough so that the avalanche effect will not occur (or even come close to happening) during any part of the ac cycle.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. The term *semiconductor* arises from
    (a) resistor-like properties of metal oxides.
    (b) variable conductive properties of some materials.
    (c) the fact that electrons conduct better than holes.
    (d) insulating properties of silicon and GaAs.
2. Which of the following is *not* an advantage of semiconductor devices over vacuum tubes?
    (a) Smaller size
    (b) Lower working voltage

   (c) Lighter weight
   (d) Ability to withstand high voltage spikes

3. Of the following substances, which is the most commonly used semiconductor?
   (a) Germanium
   (b) Galena
   (c) Silicon
   (d) Copper

4. GaAs is
   (a) a compound.
   (b) an element.
   (c) a mixture.
   (d) a gas.

5. A disadvantage of MOS devices is the fact that
   (a) the charge carriers move fast.
   (b) the material does not react to ionizing radiation.
   (c) they can be damaged by electrostatic discharges.
   (d) they must always be used at high frequencies.

6. Selenium works especially well in
   (a) photocells.
   (b) high-frequency detectors.
   (c) RF power amplifiers.
   (d) voltage regulators.

7. Of the following, which material allows the lowest forward voltage drop in a diode?
   (a) Selenium
   (b) Silicon
   (c) Copper
   (d) Germanium

8. A CMOS integrated circuit
   (a) can only work at low frequencies.
   (b) requires very little power to function.
   (c) requires considerable power to function.
   (d) can only work at high frequencies.

9. The purpose of doping is to
   (a) make the charge carriers move faster.
   (b) cause holes to flow.

(c) give a semiconductor material specific properties.
(d) protect devices from damage in case of transients.

10. A semiconductor material is made into N type by
    (a) adding an acceptor impurity.
    (b) adding a donor impurity.
    (c) injecting protons.
    (d) taking neutrons away.

11. Which of the following does not result from adding an acceptor impurity?
    (a) The material becomes P type.
    (b) Current flows mainly in the form of holes.
    (c) Most of the carriers have positive electric charge.
    (d) The substance acquires an electron surplus.

12. In a P-type material, electrons are
    (a) the majority carriers.
    (b) the minority carriers.
    (c) positively charged.
    (d) entirely absent.

13. Holes move from
    (a) minus to plus.
    (b) plus to minus.
    (c) P-type to N-type material.
    (d) N-type to P-type material.

14. When a P-N junction does not conduct even though a voltage is applied, the junction is
    (a) reverse-biased at a voltage less than the avalanche voltage.
    (b) overdriven.
    (c) biased past the breaker voltage.
    (d) in a state of avalanche effect.

15. Holes flow the opposite way from electrons because
    (a) charge carriers flow continuously.
    (b) they have opposite electric charge.
    (c) they have the same electric charge.
    (d) Forget it! Holes flow in the same direction as electrons.

16. If an electron is considered to have a charge of −1 unit, then a hole can be considered to have
    (a) a charge of −1 unit.
    (b) no charge.

(c) a charge of +1 unit.
(d) a charge that depends on the semiconductor type.

17. When a P-N junction is forward-biased, conduction will not occur unless
    (a) the applied voltage exceeds the forward breakover voltage.
    (b) the applied voltage is less than the forward breakover voltage.
    (c) the junction capacitance is high enough.
    (d) the depletion region is wide enough.
18. If the reverse bias exceeds the avalanche voltage in a P-N junction,
    (a) the junction will be destroyed.
    (b) the junction will insulate; no current will flow.
    (c) the junction will conduct current.
    (d) the capacitance will become extremely low.
19. Avalanche voltage is routinely exceeded when a P-N junction acts as a
    (a) current rectifier.
    (b) variable resistor.
    (c) variable capacitor.
    (d) voltage regulator.
20. Which of the following does not affect the junction capacitance of a diode?
    (a) the cross-sectional area of the P-N junction
    (b) the width of the depletion region
    (c) the phase of an applied ac signal
    (d) the reverse-bias voltage

20
CHAPTER

How Diodes Are Used

IN THE EARLY YEARS OF ELECTRONICS, NEARLY ALL DIODES WERE VACUUM TUBES. TODAY, MOST ARE made from semiconductors. Contemporary diodes can do almost everything that the old ones could, and also some things that people in the tube era could only dream about.

Rectification

The hallmark of a *rectifier diode* is that it passes current in only one direction. This makes it useful for changing ac to dc. Generally speaking, when the cathode is negative with respect to the anode, current flows; when the cathode is positive relative to the anode, there is no current. The constraints on this behavior are the forward breakover and avalanche voltages, as you learned about in Chap. 19.

Examine the circuit shown at A in Fig. 20-1. Suppose a 60-Hz ac sine wave is applied to the input. During half the cycle, the diode conducts, and during the other half, it doesn't. This cuts off half of every cycle. Depending on which way the diode is hooked up, either the positive half or the negative half of the ac cycle will be removed. Drawing B in Fig. 20-1 shows a graph of the output

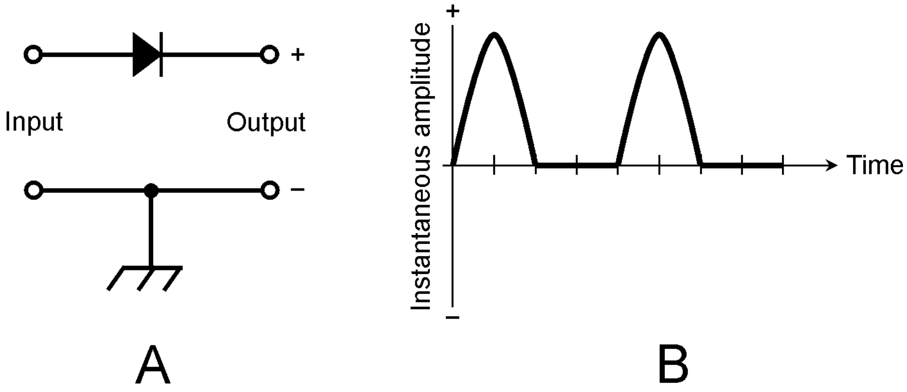

20-1 At A, a half-wave rectifier circuit. At B, the output of the circuit shown at A when an ac sine wave is applied to the input.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

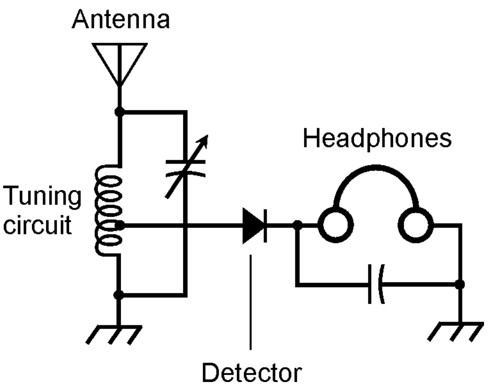

20-2 Schematic diagram of a crystal-set radio receiver.

of the circuit at A. Remember that electrons flow from negative to positive, against the arrow in the diode symbol.

The circuit and wave diagram of Fig. 20-1 show a *half-wave rectifier* circuit. This is the simplest possible rectifier. That's its chief advantage over other, more complicated rectifier circuits. You'll learn about the various types of rectifier diodes and circuits in Chap. 21.

Detection

One of the earliest diodes, existing even before vacuum tubes, was actually a primitive semiconductor device. Known as a *cat whisker*, it consisted of a fine piece of wire in contact with a small piece of the mineral *galena.* This strange-looking contraption had the ability to act as a rectifier for extremely weak RF currents. When the cat whisker was connected in a circuit such as the one shown in Fig. 20-2, the result was a device capable of picking up amplitude-modulated (AM) radio signals and producing audio output that could be heard in the headset.

The galena, sometimes called a "crystal," gave rise to the nickname *crystal set* for this primitive radio receiver. You can still build a crystal set today, using a simple RF diode, a coil, a tuning capacitor, a headset, and a long-wire antenna. Notice that there's no battery! The audio is provided by the received signal alone.

The diode in Fig. 20-2 acts to recover the audio from the radio signal. This process is called *detection*; the circuit is called a *detector* or *demodulator.* If the detector is to be effective, the diode must be of the proper type. It must have low junction capacitance, so that it can work as a rectifier (and not as a capacitor) at radio frequencies. Some modern RF diodes are microscopic versions of the old cat whisker, enclosed in a glass case with axial leads.

Frequency Multiplication

When current passes through a diode, half of the cycle is cut off, as shown in Fig. 20-1B. This occurs no matter what the frequency, from 60-Hz utility current through RF, as long as the diode capacitance is not too great.

The output wave from the diode looks much different than the input wave. This condition is known as *nonlinearity.* Whenever there is nonlinearity of any kind in a circuit—that is, whenever the output waveform is shaped differently from the input waveform—there are harmonics in the output. These are waves at integer multiples of the input frequency. (If you've forgotten what harmonics are, refer to Chap. 9.)

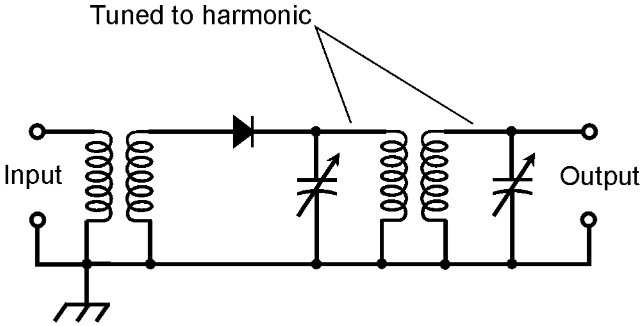

20-3 A frequency-multiplier circuit using a semiconductor diode.

Often, nonlinearity is undesirable. Then engineers strive to make the circuit *linear*, so the output waveform has exactly the same shape as the input waveform. But sometimes harmonics are desired. Then nonlinearity is introduced deliberately to produce *frequency multiplication.* Diodes are ideal for this purpose. A simple frequency-multiplier circuit is shown in Fig. 20-3. The output *LC* circuit is tuned to the desired *n*th harmonic frequency, nf_o, rather than to the input or fundamental frequency, f_o.

For a diode to work as a frequency multiplier, it must be of a type that would also work well as a detector at the same frequencies. This means that the component should act like a rectifier, but not like a capacitor.

Signal Mixing

When two waves having different frequencies are combined in a nonlinear circuit, new waves are produced at frequencies equal to the sum and difference of the frequencies of the input waves. Diodes can provide this nonlinearity.

Suppose there are two signals with frequencies f_1 and f_2. For mathematical convenience, let's assign f_2 to the wave with the higher frequency, and f_1 to the wave with the lower frequency. If these signals are combined in a nonlinear circuit, new waves result. One of them has a frequency of $f_2 + f_1$, and the other has a frequency of $f_2 - f_1$. These sum and difference frequencies are known as *beat frequencies.* The signals themselves are called *mixing products* or *heterodynes* (Fig. 20-4).

Figure 20-4, incidentally, is an illustration of a *frequency domain* display. The amplitude (on the vertical scale or axis) is shown as a function of the frequency (on the horizontal scale or axis). This sort of display is what engineers see when they look at the screen of a lab instrument known as a *spectrum analyzer.* In contrast, an ordinary oscilloscope displays amplitude (on the vertical scale or axis) as a function of time (on the horizontal scale or axis). The oscilloscope provides a *time domain* display.

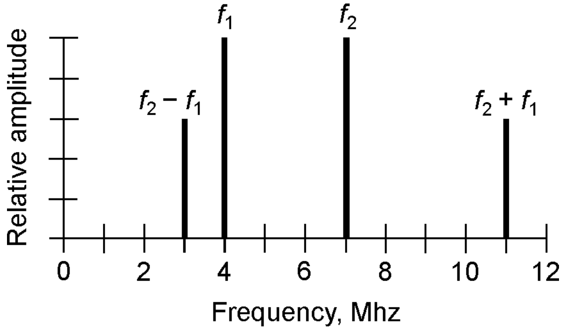

20-4 Spectral (frequency-domain) illustration of signal mixing.

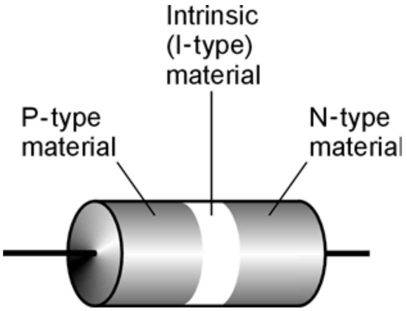

20-5 The PIN diode has a layer of intrinsic (I type) semiconductor material at the P-N junction.

Switching

The ability of diodes to conduct with forward bias, and to insulate with reverse bias, makes them useful for switching in some electronic applications. Diodes can perform switching operations much faster than any mechanical device.

One type of diode, made for use as an RF switch, has a special semiconductor layer sandwiched in between the P-type and N-type material. The material in this layer is called an *intrinsic* (or *I-type*) *semiconductor.* The *intrinsic layer* (or *I layer*) reduces the capacitance of the diode, so that it can work at higher frequencies than an ordinary diode. A diode with an I-type semiconductor layer sandwiched in between the P- and N-type layers is called a *PIN diode* (Fig. 20-5).

Direct-current bias, applied to one or more PIN diodes, allows RF currents to be effectively channeled without using relays and cables. A PIN diode also makes a good RF detector, especially at very high frequencies.

Voltage Regulation

Most diodes have an avalanche breakdown voltage that is much higher than the reverse bias ever gets. The value of the avalanche voltage depends on how a diode is manufactured. *Zener diodes* are specially made so they exhibit well-defined, constant avalanche voltages.

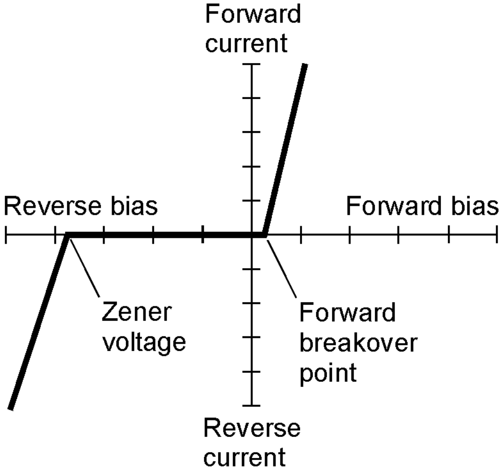

20-6 Current through a Zener diode as a function of the bias voltage.

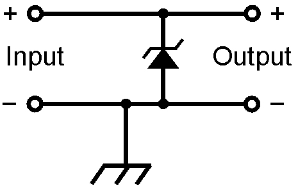

20-7 Connection of a Zener diode for voltage regulation.

Suppose a certain Zener diode has an avalanche voltage, also called the *Zener voltage*, of 50 V. If reverse bias is applied to the P-N junction, the diode acts as an open circuit as long as the bias is less than 50 V. But if the reverse-bias voltage reaches 50 V—even for a brief instant of time—the diode conducts. This effectively prevents the reverse-bias voltage from exceeding 50 V.

The current through a Zener diode, as a function of the voltage, is shown in Fig. 20-6. The Zener voltage is indicated by the abrupt rise in reverse current as the reverse-bias voltage increases. A simple Zener-diode voltage-limiting circuit is shown in Fig. 20-7. Note the polarity of the diode: the cathode is connected to the positive pole, and the anode is connected to the negative pole.

Amplitude Limiting

In Chap. 19, you learned that a diode will not conduct until the forward-bias voltage is at least as great as the forward breakover voltage. There's a corollary to this: a diode will always conduct when the forward-bias voltage reaches or exceeds the forward breakover voltage, when the device is conducting current in the forward direction. In the case of silicon diodes this is approximately 0.6 V. For germanium diodes it is about 0.3 V, and for selenium diodes it is about 1 V.

This phenomenon can be used to advantage when it is necessary to limit the amplitude of a signal, as shown in Fig. 20-8. By connecting two identical diodes back-to-back in parallel with the signal path (A), the maximum peak amplitude is limited, or *clipped*, to the forward breakover voltage of the diodes. The input and output waveforms of a clipped signal are illustrated at B. This scheme is sometimes used in radio receivers to prevent "blasting" when a strong signal comes in.

The downside of the *diode limiter* circuit, such as the one shown in Fig. 20-8, is the fact that it introduces distortion when clipping occurs. This might not be a problem for reception of digital signals, for frequency-modulated signals, or for analog signals that rarely reach the limiting voltage. But for amplitude-modulated signals with peaks that rise past the limiting voltage, it can cause trouble.

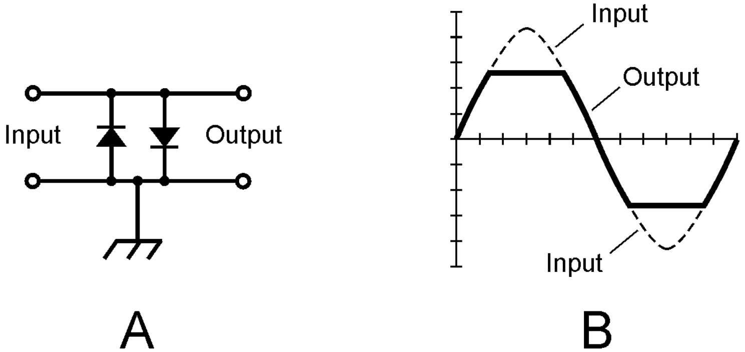

20-8 At A, connection of two diodes to act as an ac limiter. At B, illustration of sine-wave peaks cut off by the action of the diodes in an ac limiter.

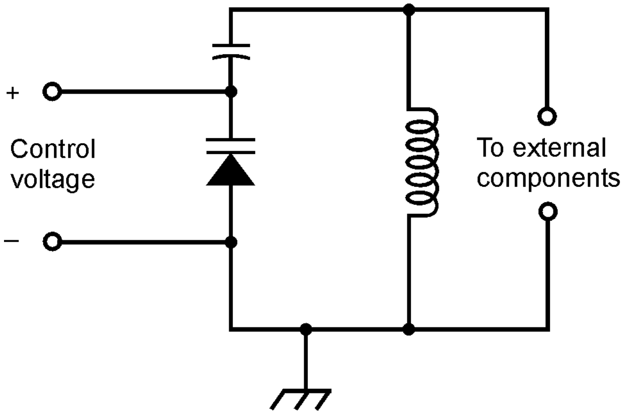

20-9 Connection of a varactor diode in a tuned circuit.

Frequency Control

When a diode is reverse-biased, there is a region at the P-N junction with dielectric (insulating) properties. As you know from Chap. 19, this is called the depletion region, because it has a shortage of majority charge carriers. The width of this zone depends on several things, including the reverse-bias voltage.

As long as the reverse bias is less than the avalanche voltage, varying the bias affects the width of the depletion region. This in turn varies the junction capacitance. This capacitance, which is always small (on the order of picofarads), varies inversely with the square root of the reverse-bias voltage, as long as the reverse bias remains less than the avalanche voltage. Thus, for example, if the reverse-bias voltage is quadrupled, the junction capacitance drops to one-half; if the reverse-bias voltage is decreased by a factor of 9, then the junction capacitance increases by a factor of 3.

Some diodes are manufactured especially for use as variable capacitors. Such a device is known as varactor diode, as you learned in Chap. 19. Varactors are used in a special type of circuit called a *voltage-controlled oscillator* (VCO). Figure 20-9 is a simple example of the *LC* circuit in a VCO, using a coil, a fixed capacitor, and a varactor. This is a parallel-tuned circuit. The fixed capacitor, whose value is large compared with that of the varactor, serves to keep the coil from short-circuiting the control voltage across the varactor. Notice that the symbol for the varactor has two lines on the cathode side.

Oscillation and Amplification

Under certain conditions, diodes can be made to produce microwave RF signals. Three types of diodes that can do this are *Gunn diodes*, *IMPATT diodes*, and *tunnel diodes.*

Gunn Diodes

A Gunn diode can produce up to 1 W of RF power output, but more commonly it works at levels of about 0.1 W. Gunn diodes are usually made from gallium arsenide. A Gunn diode oscillates because of the *Gunn effect*, named after J. Gunn of International Business Machines (IBM), who first observed it in the 1960s. A Gunn diode doesn't work like a rectifier, detector, or mixer. Instead, the oscillation takes place as a result of a quirk called *negative resistance.*

Gunn-diode oscillators are often tuned using varactor diodes. A Gunn-diode oscillator, connected directly to a microwave horn antenna, is known as a *Gunnplexer.* These devices are popular with amateur-radio experimenters at frequencies of 10 GHz and above.

IMPATT Diodes

The acronym *IMPATT* comes from the words *imp*act *a*valanche *t*ransit *t*ime. This, like negative resistance, is a rather esoteric phenomenon. An *IMPATT diode* is a microwave oscillating device like a Gunn diode, except that it uses silicon rather than gallium arsenide.

An IMPATT diode can be used as an amplifier for a microwave transmitter that employs a Gunn-diode oscillator. As an oscillator, an IMPATT diode produces about the same amount of output power, at comparable frequencies, as a Gunn diode.

Tunnel Diodes

Another type of diode that will oscillate at microwave frequencies is the *tunnel diode*, also known as the *Esaki diode.* It produces enough power so it can be used as a local oscillator in a microwave radio receiver, but not much more.

Tunnel diodes work well as amplifiers in microwave receivers, because they generate very little unwanted noise. This is especially true of gallium arsenide devices.

Energy Emission

Some semiconductor diodes emit radiant energy when a current passes through the P-N junction in a forward direction. This phenomenon occurs as electrons fall from higher to lower energy states within atoms.

LEDs and IREDs

Depending on the exact mixture of semiconductors used in manufacture, visible light of almost any color can be produced by diodes when bias is applied to them in the forward direction. Infrared-emitting devices also exist. The most common color for a *light-emitting diode* (LED) is bright red. An *infrared-emitting diode* (IRED) produces energy at wavelengths slightly longer than those of visible red light.

The intensity of the radiant energy from an LED or IRED depends to some extent on the forward current. As the current rises, the brightness increases, but only up to a certain point. If the current continues to rise, no further increase in brilliance takes place. The LED or IRED is then said to be in a state of *saturation.*

Digital Displays

Because LEDs can be made in various different shapes and sizes, they are ideal for use in digital displays. You've seen digital clock radios that use them. They are common in car radios. They make good indicators for "on/off," "a.m./p.m.," "battery low," and other conditions.

In recent years, LED displays have been largely replaced by *liquid crystal displays* (LCDs). The LCD technology has advantages over LED technology, including lower power consumption and better visibility in direct sunlight. However, LCDs require backlighting when the ambient illumination is low.

Communications

Both LEDs and IREDs are useful in communications because their intensity can be modulated to carry information. When the current through the device is sufficient to produce output, but not enough to cause saturation, the LED or IRED output follows along with rapid current changes. Analog and digital signals can be conveyed over light beams in this way. Some modern telephone systems make use of modulated light, transmitted through clear fibers. This is known as *fiber-optic* technology.

Special LEDs and IREDs produce *coherent radiation.* These are called *laser diodes.* The rays from these diodes aren't the intense, parallel beams that most people imagine when they think about lasers. A laser LED or IRED generates a cone-shaped beam of low intensity. But it can be focused into a parallel beam, and the resulting rays have some of the same advantages found in larger lasers, including the ability to travel long distances with little decrease in their intensity.

Photosensitive Diodes

Virtually all P-N junctions exhibit conductivity that varies with exposure to radiant electromagnetic energy such as IR, visible light, and UV. The reason that conventional diodes are not affected by these rays is that they are enclosed in opaque packages. Some *photosensitive diodes* have variable dc resistance that depends on the intensity of the electromagnetic rays. Other types of diodes produce their own dc in the presence of radiant energy.

Silicon Photodiodes

A silicon diode, housed in a transparent case and constructed in such a way that visible light can strike the barrier between the P-type and N-type materials, forms a *silicon photodiode.* A reverse-bias voltage is applied to the device. When radiant energy strikes the junction, current flows. The current is proportional to the intensity of the radiant energy, within certain limits.

Silicon photodiodes are more sensitive at some wavelengths than at others. The greatest sensitivity is in the *near infrared* part of the spectrum, at wavelengths just a little bit longer than the wavelength of visible red light.

When radiant energy of variable intensity strikes the P-N junction of a reverse-biased silicon photodiode, the output current follows the light-intensity variations. This makes silicon photodiodes useful for receiving modulated-light signals of the kind used in fiber-optic communications systems.

The Optoisolator

An LED or IRED and a photodiode can be combined in a single package to get a component called an *optoisolator.* This device, the schematic symbol for which is shown in Fig. 20-10, creates a modulated-light signal and sends it over a small, clear gap to a receptor. An LED or IRED converts an electrical signal to visible light or IR; a photodiode changes the visible light or infrared back into an electrical signal.

When a signal is electrically coupled from one circuit to another, the two stages interact. The input impedance of a given stage, such as an amplifier, can affect the behavior of the circuits that feed power to it. This can lead to various sorts of trouble. Optoisolators overcome this effect, because the coupling is not done electrically. If the input impedance of the second circuit changes, the impedance that the first circuit sees is not affected, because it is simply the impedance of the LED

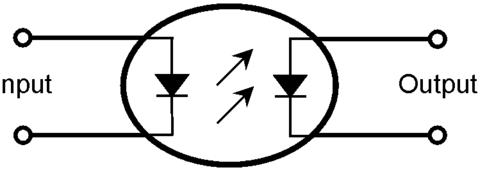

20-10 An optoisolator has an LED or IRED at the input and a photodiode at the output.

or IRED. That is where the "isolator" in "optoisolator" comes from. The circuits can be *electronically coupled*, and yet at the same time remain *electrically isolated.*

Photovoltaic Cells

A silicon diode, with no bias voltage applied, can generate dc all by itself if enough electromagnetic radiation hits its P-N junction. This is known as the *photovoltaic effect.* It is the principle by which solar cells work.

Photovoltaic cells are specially manufactured to have the greatest possible P-N junction surface area. This maximizes the amount of light that strikes the junction. A single silicon photovoltaic cell can produce about 0.6 V of dc electricity. The amount of current that it can deliver, and thus the amount of power it can provide, depends on the surface area of the junction.

Photovoltaic cells can be connected in series-parallel combinations to provide power for solid-state electronic devices such as portable radios. These arrays can also be used to charge batteries, allowing for use of the electronic devices when radiant energy is not available (for example, at night!). A large assembly of solar cells, connected in series-parallel, is called a *solar panel.* The power produced by a solar panel depends on the intensity of the light that strikes it, the sum total of the surface areas of all the cells, and the angle at which the light strikes the cells. Some solar panels can produce several kilowatts of electrical power in direct sunlight that shines in such a way that the sun's rays arrive perpendicular to the surfaces of all the cells.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. When a diode is forward-biased, the anode voltage
   (a) is negative relative to the cathode voltage.
   (b) is positive relative to the cathode voltage.
   (c) is the same as the cathode voltage.
   (d) alternates between positive and negative relative to the cathode voltage.

2. If a diode is connected in series with the secondary winding of an ac transformer, and if the peak voltage across the diode never exceeds the avalanche voltage, then the output of the complete transformer-diode circuit is
   (a) ac with half the frequency of the input.
   (b) ac with the same frequency as the input.
   (c) ac with twice the frequency of the input.
   (d) none of the above.

3. A crystal set
    - (a) can be used to transmit radio signals.
    - (b) requires a battery with long life.
    - (c) requires no battery.
    - (d) is used for rectifying 60-Hz ac.
4. A diode detector
    - (a) is used in power supplies.
    - (b) is employed in some radio receivers.
    - (c) is used to generate microwave RF signals.
    - (d) changes dc into ac.
5. If the output wave in a circuit has the same shape as the input wave, then
    - (a) the circuit is operating in a linear manner.
    - (b) the circuit is operating as a frequency multiplier.
    - (c) the circuit is operating as a mixer.
    - (d) the circuit is operating as a rectifier.
6. Suppose the two input signal frequencies to a mixer circuit are 3.522 MHz and 3.977 MHz. At which of the following frequencies can we expect a signal to exist at the output?
    - (a) 455 kHz
    - (b) 886 kHz
    - (c) 14.00 MHz
    - (d) 1.129 MHz
7. Fill in the blanks to make the following sentence correct: "A spectrum analyzer provides a display of ____ as a function of ____."
    - (a) amplitude/time
    - (b) time/frequency
    - (c) frequency/time
    - (d) amplitude/frequency
8. Zener voltage is a specialized manifestation of
    - (a) forward breakover voltage.
    - (b) peak forward voltage.
    - (c) avalanche voltage.
    - (d) reverse bias.
9. The forward breakover voltage of a silicon diode is
    - (a) about 0.3 V.
    - (b) about 0.6 V.
    - (c) about 1.0 V.
    - (d) dependent on the avalanche voltage.

10. A diode audio limiter circuit
    (a) is useful for voltage regulation.
    (b) always uses Zener diodes.
    (c) rectifies the audio to reduce distortion.
    (d) can cause distortion under some conditions.

11. The capacitance of a varactor varies with the
    (a) forward voltage.
    (b) reverse voltage.
    (c) avalanche voltage.
    (d) forward breakover voltage.

12. The purpose of the I layer in a PIN diode is to
    (a) minimize the junction capacitance.
    (b) optimize the avalanche voltage.
    (c) reduce the forward breakover voltage.
    (d) increase the current through the diode.

13. Which of these diode types can be used as the key element in the oscillator circuit of a microwave radio transmitter?
    (a) A rectifier diode
    (b) A PIN diode
    (c) An IMPATT diode
    (d) None of the above

14. A Gunnplexer is often used as a
    (a) microwave communications device.
    (b) low-frequency RF detector.
    (c) high-voltage rectifier.
    (d) signal mixer or frequency divider.

15. The most likely place you would find an LED would be in
    (a) a rectifier circuit.
    (b) a mixer circuit.
    (c) a digital frequency display.
    (d) an oscillator circuit.

16. Coherent electromagnetic radiation is produced by a
    (a) Gunn diode.
    (b) varactor diode.
    (c) rectifier diode.
    (d) laser diode.

17. Suppose you want a circuit to operate in a stable manner when the load impedance varies. You might consider a coupling method that employs

    (a) a Gunn diode.
    (b) an optoisolator.
    (c) a photovoltaic cell.
    (d) a PIN diode.

18. The electrical power that a solar panel can provide in direct sunlight depends on all of the following factors except

    (a) the ac voltage applied to the panel.
    (b) the total surface area of all the cells in the panel.
    (c) the angle at which the sunlight strikes the cells.
    (d) the intensity of the sunlight that strikes the cells.

19. Emission of energy in an IRED is caused by

    (a) high-frequency radio waves.
    (b) rectification.
    (c) changes in electron energy within atoms.
    (d) none of the above.

20. A photodiode, when not used as a photovoltaic cell, has

    (a) reverse bias.
    (b) no bias.
    (c) forward bias.
    (d) negative resistance.

21
CHAPTER

Power Supplies

A *POWER SUPPLY* CONVERTS UTILITY AC TO DC FOR USE WITH CERTAIN ELECTRICAL AND ELECTRONIC devices. In this chapter, we'll examine the components of a typical power supply.

Power Transformers

Power transformers can be categorized as step-down or step-up. As you remember, the output, or secondary, voltage of a step-down unit is lower than the input, or primary, voltage. The reverse is true for a step-up transformer.

Step-down

Most solid-state electronic devices, such as radios, need only a few volts. The power supplies for such equipment use step-down power transformers. The physical size of the transformer depends on the current. Some devices need only a small current and a low voltage. The transformer in a radio receiver, for example, can be physically small. A ham radio transmitter or hi-fi amplifier needs more current. This means that the secondary winding of the transformer must consist of heavy-gauge wire, and the core must be bulky to contain the magnetic flux.

Step-up

Some circuits need high voltage. The cathode-ray tube (CRT) in a conventional home television set needs several hundred volts. Some ham radio power amplifiers use vacuum tubes working at more than 1 kV dc. The transformers in these appliances are step-up types. They are moderate to large in size, because of the number of turns in the secondary, and also because high voltages can spark, or *arc*, between wire turns if the windings are too tight. If a step-up transformer needs to supply only a small amount of current, it need not be big. But for ham radio transmitters and radio or television broadcast amplifiers, the transformers are large, heavy, and expensive.

Transformer Ratings

Transformers are rated according to output voltage and current. For a given unit, the *volt-ampere* (VA) capacity is often specified. This is the product of the voltage and current.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

A transformer with 12-V output, capable of delivering 10 A, has 12 V × 10 A = 120 VA of capacity. The nature of power-supply filtering, to be discussed later in this chapter, makes it necessary for the power-transformer VA rating to be greater than the wattage consumed by the load.

A high-quality, rugged power transformer, capable of providing the necessary currents and/or voltages, is crucial in any power supply. The transformer is usually the most expensive component to replace.

Rectifier Diodes

Rectifier diodes are available in various sizes, intended for different purposes. Most rectifier diodes are made of silicon, and are known as *silicon rectifiers.* Some are fabricated from selenium, and are called *selenium rectifiers.* Two important features of a power-supply diode are the *average forward current* (I_o) rating and the *peak inverse voltage* (PIV) rating.

Average Forward Current

Electric current produces heat. If the current through a diode is too great, the heat will destroy the P-N junction. When designing a power supply, it is wise to use diodes with an I_o rating of at least 1.5 times the expected average dc forward current. If this current is 4.0 A, for example, the rectifier diodes should be rated at $I_o = 6.0$ A or more.

Note that I_o flows through the *diodes.* The current drawn by the *load* is often different from this. Also, note that I_o is an *average* figure. The *instantaneous* forward current is another thing, and can be 15 or 20 times the I_o, depending on the nature of the filtering circuit.

Some diodes have *heatsinks* to help carry heat away from the P-N junction. A selenium diode can be recognized by the appearance of its heatsink, which looks something like a baseboard radiator built around a steam pipe.

Diodes can be connected in parallel to increase the current rating over that of an individual diode. When this is done, small-value resistors should be placed in series with each diode in the set to equalize the current. Each resistor should have a value such that the voltage drop across it is about 1 V under normal operating conditions.

Peak Inverse Voltage

The PIV rating of a diode is the instantaneous reverse-bias voltage that it can withstand without the avalanche effect taking place. A good power supply has diodes whose PIV ratings are significantly greater than the peak ac input voltage. If the PIV rating is not great enough, the diode or diodes in a supply conduct for part of the reverse cycle. This degrades the efficiency of the supply because the reverse current bucks the forward current.

Diodes can be connected in series to get a higher PIV capacity than a single diode alone. This scheme is sometimes seen in high-voltage supplies, such as those needed for tube-type power amplifiers. High-value resistors, of about 500 Ω for each peak-inverse volt, are placed across each diode in the set to distribute the reverse bias equally among the diodes. In addition, each diode is shunted by (that is, connected in parallel with) a capacitor of 0.005 μF or 0.1 μF.

Half-Wave Circuit

The simplest rectifier circuit, called the *half-wave rectifier* (Fig. 21-1A), has a single diode that chops off half of the ac cycle. The effective (eff) output voltage from a power supply that uses a

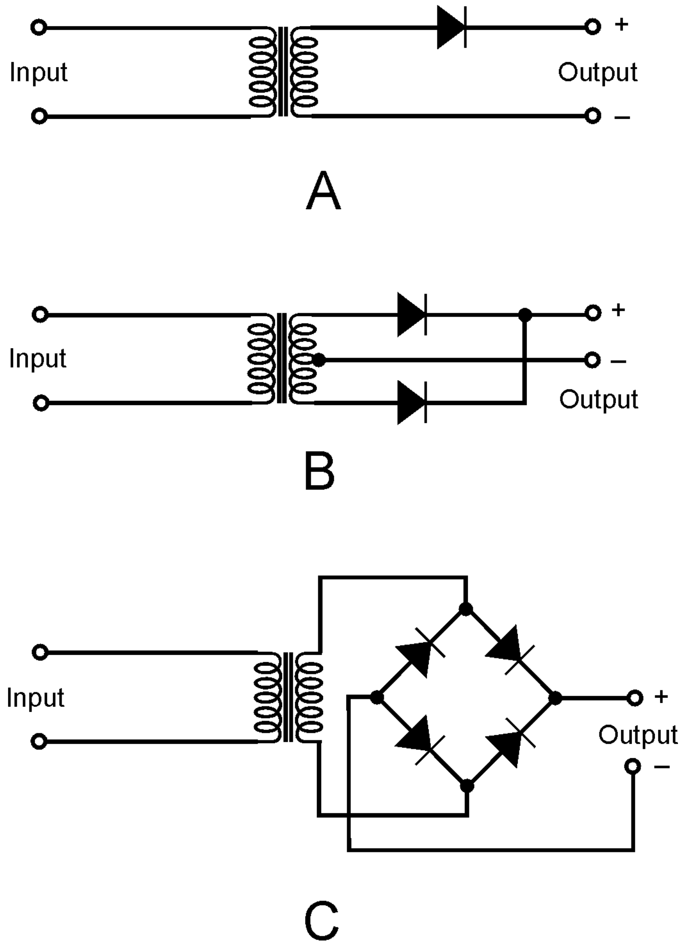

21-1 At A, a half-wave rectifier circuit. At B, a full-wave center-tap rectifier circuit. At C, a full-wave bridge rectifier circuit.

half-wave rectifier is much less than the peak transformer output voltage, as shown in Fig. 21-2A. The peak voltage across the diode in the reverse direction can be as much as 2.8 times the applied rms ac voltage.

Most engineers like to use diodes whose PIV ratings are at least 1.5 times the maximum expected peak reverse voltage. Therefore, in a half-wave rectifier circuit, the diodes should be rated for at least 2.8 × 1.5, or 4.2, times the rms ac voltage that appears across the secondary winding of the power transformer.

Half-wave rectification has shortcomings. First, the output is difficult to filter. Second, the output voltage can drop considerably when the supply is required to deliver high current. Third, half-wave rectification puts a strain on the transformer and diodes because it *pumps* them. The circuit works the diodes hard during half the ac cycle, and lets them loaf during the other half.

Half-wave rectification is usually adequate for use in a power supply that is not required to deliver much current, or when the voltage can vary without affecting the behavior of the equipment connected to it. The main advantage of a half-wave circuit is that it costs less than more sophisticated circuits.

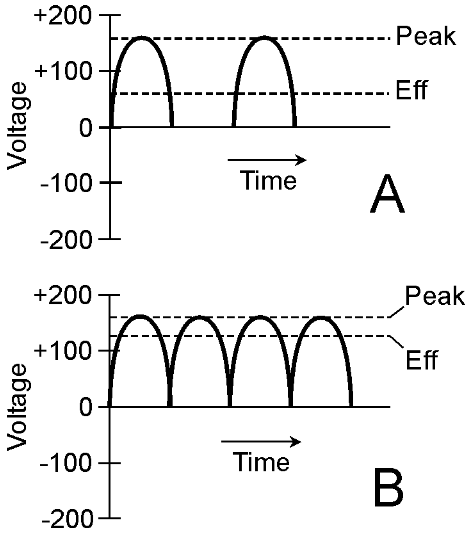

21-2 At A, the output of a half-wave rectifier. At B, the output of a full-wave rectifier. Note the difference in how the effective (eff) voltages compare with the peak voltages.

Full-Wave Center-Tap Circuit

A better scheme for changing ac to dc takes advantage of both halves of the ac cycle. A *full-wave center-tap rectifier* has a transformer with a tapped secondary (Fig. 21-1B). The center tap is connected to *electrical ground*, also called *chassis ground*. This produces voltages and currents at the ends of the winding that are in phase opposition with respect to each other. These two ac waves can be individually half-wave rectified, cutting off one half of the cycle and then the other, over and over.

The effective output voltage from a power supply that uses a full-wave center-tap rectifier is greater, relative to the peak voltage, than is the case with the half-wave rectifier (Fig. 21-2B). The PIV across the diodes can, nevertheless, be as much as 2.8 times the applied rms ac voltage. Therefore, the diodes should have a PIV rating of at least 4.2 times the applied rms ac voltage to ensure that they won't break down.

The output of a full-wave center-tap rectifier is easier to filter than that of a half-wave rectifier because the frequency of the pulsations in the dc (known as the *ripple frequency*) from a full-wave rectifier is twice the ripple frequency of the pulsating dc from a half-wave rectifier, assuming identical ac input frequency in either situation. If you compare Fig. 21-2B with Fig. 21-2A, you will see that the full-wave-rectifier output is closer to pure dc than the half-wave rectifier output. Another advantage of a full-wave center-tap rectifier is the fact that it's gentler with the transformer and diodes than a half-wave rectifier. Yet another asset: When a load is applied to the output of a power supply that uses a full-wave center-tap rectifier circuit, the voltage drops less than is the case with a half-wave supply. But because the transformer is more sophisticated, the full-wave center-tap circuit costs more than a half-wave circuit that delivers the same output voltage at the same rated maximum current.

Full-Wave Bridge Circuit

Another way to get full-wave rectification is the *full-wave bridge rectifier*, often called simply a *bridge*. It is diagrammed in Fig. 21-1C. The output waveform is similar to that of the full-wave center-tap circuit (Fig. 21-2B).

The effective output voltage from a power supply that uses a full-wave bridge rectifier is somewhat less than the peak transformer output voltage, as shown in Fig. 21-2B. The peak voltage across the diodes in the reverse direction is about 1.4 times the applied rms ac voltage. Therefore, each diode needs to have a PIV rating of at least 1.4 × 1.5, or 2.1, times the rms ac voltage that appears at the transformer secondary.

The bridge circuit does not require a center-tapped transformer secondary. It uses the entire secondary winding on both halves of the wave cycle, so it makes even more efficient use of the transformer than the full-wave center-tap circuit. The bridge is also easier on the diodes than half-wave or full-wave center-tap circuits.

Voltage-Doubler Circuit

Diodes and capacitors can be interconnected to deliver a dc output that is approximately twice the positive or negative peak ac input voltage. This is called a *voltage-doubler power supply.* This circuit works well only when the load draws low current. Otherwise, the *voltage regulation* is poor; the voltage drops a lot when the current demand is significant.

The best way to build a high-voltage power supply is to use a step-up transformer, not a voltage-doubling scheme. Nevertheless, a voltage-doubler power supply can be, and sometimes is, used when the cost of the circuit must be minimized and the demands placed on it are expected to be modest.

Figure 21-3 is a simplified diagram of a voltage-doubler power supply. It works on the entire ac cycle, so it is called a *full-wave voltage doubler.* This circuit subjects the diodes to voltage peaks in the reverse direction that are 2.8 times the applied rms ac voltage. Therefore, the diodes should be rated for PIV of at least 4.2 times the rms ac voltage that appears across the transformer secondary. When the current drawn is low, the dc output voltage of this type of power supply is approximately 2.8 times the rms ac input voltage.

Proper operation of a voltage-doubler power supply depends on the ability of the capacitors to hold a charge under maximum load. The capacitors must have large values, as well as be capable of handling high voltages. The capacitors serve two purposes: to boost the voltage and to filter the out-

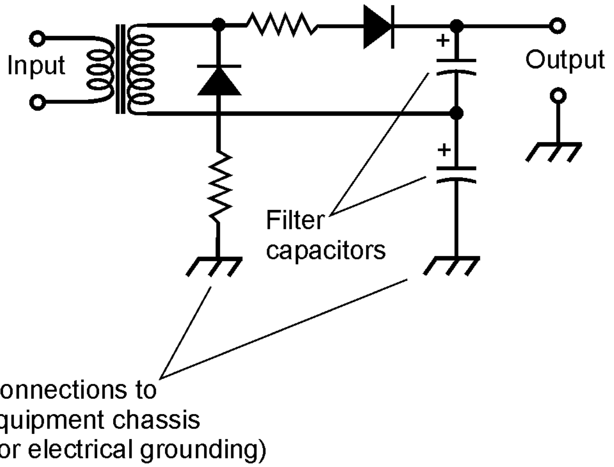

21-3 A full-wave voltage-doubler power supply.

put. The resistors, which have low ohmic values and are connected in series with the diodes, protect the diodes against *surge currents* that occur when the power supply is first switched on.

Filtering

Most dc-powered devices need something better—more pure—than the rough, pulsating dc that comes right out of a rectifier circuit. The pulsations (ripple) in the rectifier output can be eliminated by a *filter.*

Capacitors Alone

The simplest power-supply filter consists of one or more large-value capacitors, connected in parallel with the rectifier output (Fig. 21-4). A good component for this purpose is known as an *electrolytic capacitor.* This type of capacitor is *polarized*, meaning that it must be connected in the correct direction in the circuit. Each capacitor is also rated for a certain maximum voltage. Pay attention to these ratings if you ever work with electrolytic capacitors!

Filter capacitors work by trying to maintain the dc voltage at its peak level, as shown in Fig. 21-5. This is easier to do with the output of a full-wave rectifier (drawing A) than with the output of a half-wave rectifier (drawing B). With a full-wave rectifier receiving a 60-Hz ac electrical input, the ripple frequency is 120 Hz, but with a half-wave rectifier it is 60 Hz. The filter capacitors are thus recharged twice as often with a full-wave rectifier, as compared with a half-wave rectifier. This is why full-wave rectifier circuits produce more pure dc than half-wave rectifier circuits.

Capacitors and Chokes

Another way to smooth out the dc from a rectifier is to place a large-value inductor in series with the output, and a large-value capacitor in parallel. The inductor is called a *filter choke.*

In a filter that uses a capacitor and an inductor, the capacitor can be placed on the rectifier side of the choke. This is a *capacitor-input filter* (Fig. 21-6A). If the filter choke is placed on the rectifier side of the capacitor, the circuit is a *choke-input filter* (Fig. 21-6B). Capacitor-input filtering can be used when a power supply is not required to deliver much current. The output voltage, when the load is light (not much current is drawn), is higher with a capacitor-input filter than with a choke-input filter having identical input. If the supply needs to deliver large or variable amounts of current, a choke-input filter is a better choice, because the output voltage is more stable.

If the output of a power supply must have an absolute minimum of ripple, two or three capacitor/choke pairs can be connected in *cascade* (Fig. 21-7). Each pair constitutes a *section* of the filter. Multisection filters can consist of capacitor-input or choke-input sections, but the two types are never

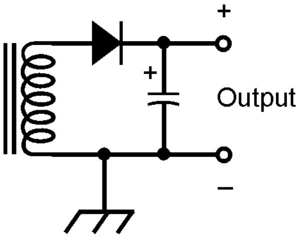

21-4 A large-value capacitor can be used all by itself as a power-supply filter.

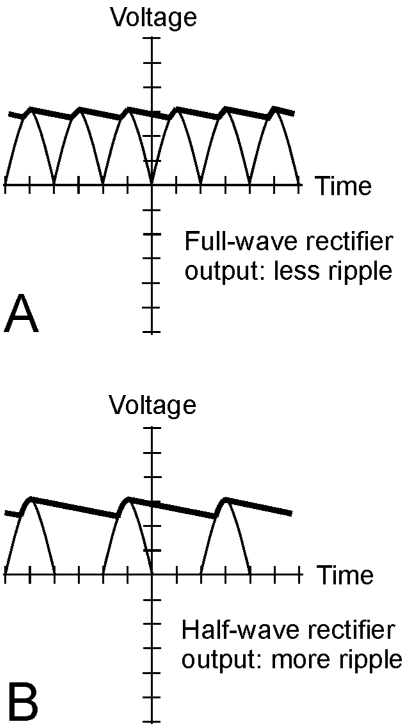

21-5 Filtering of ripple from a full-wave rectifier (A) and from a half-wave rectifier (B).

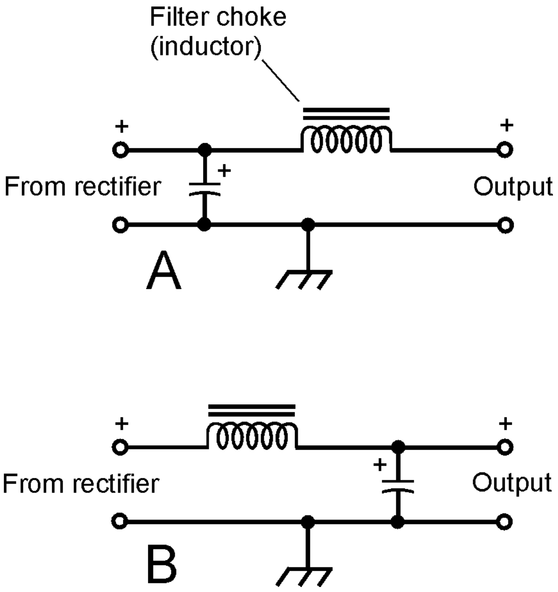
Filter choke
(inductor)
+
+
+
From rectifier
Output
A
+
+
+
From rectifier
Output
B

21-6 At A, a capacitor-input filter. At B, a choke-input filter.

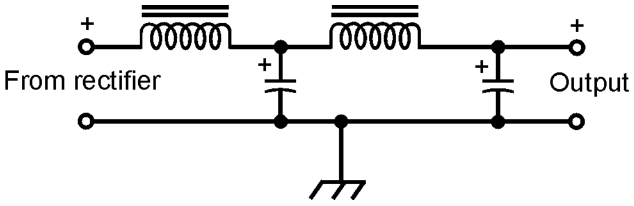

21-7 Two choke-input filter sections in cascade.

mixed. In the example of Fig. 21-7, both capacitor/choke pairs are called *L sections.* If the second capacitor is omitted, the filter becomes a *T section.* If the second capacitor is moved to the input and the second choke is omitted, the filter becomes a *pi section.* These sections are named because their schematic diagrams look something like the uppercase English L, the uppercase English T, and the uppercase Greek Π, respectively.

Voltage Regulation

If a special diode called a *Zener diode* is connected in parallel with the output of a power supply, the diode limits the output voltage. The diode must have an adequate power rating to prevent it from burning out. The limiting voltage depends on the particular Zener diode used. Zener diodes are available for any reasonable power-supply voltage.

Figure 21-8 is a diagram of a full-wave bridge dc power supply including a Zener diode for voltage regulation. Note the direction in which the Zener diode is connected in this application: with the arrow pointing from minus to plus. This is contrary to the polarity used for rectifier diodes. It's important that the polarity be correct with a Zener diode, or it will burn out.

A Zener-diode voltage regulator is inefficient when the supply is used with equipment that draws high current. When a supply must deliver a lot of current, a *power transistor* is used along with the Zener diode to obtain regulation. Figure 21-9 shows such a circuit.

Voltage regulators are available in *integrated-circuit* (IC) form. The *regulator IC,* also called a *regulator chip,* is installed in the power-supply circuit at the output of the filter. In high-voltage power supplies, *electron tubes* are sometimes used as voltage regulators. These are particularly rugged, and can withstand much higher temporary overloads than Zener diodes, transistors, or chips. However, some engineers consider such *regulator tubes* archaic.

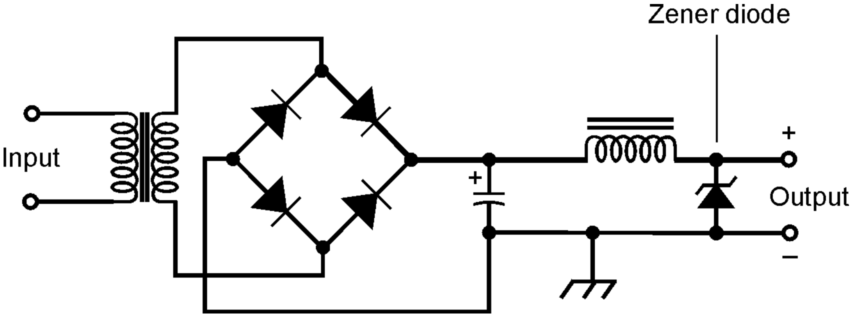

21-8 A power supply with a Zener-diode voltage regulator in the output.

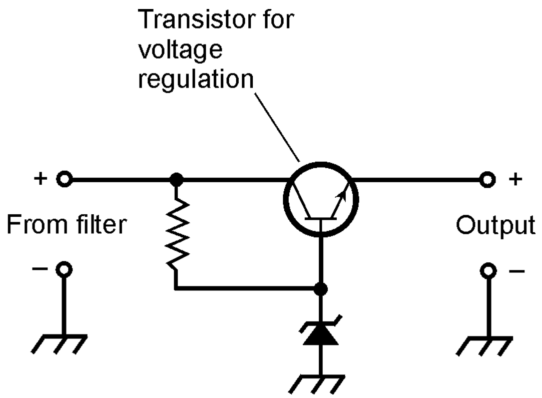

21-9 A voltage-regulator circuit using a Zener diode and an NPN transistor.

Protection of Equipment

The output of a power supply should be free of sudden changes that can damage equipment or components, or interfere with their proper performance. It is also important that voltages not appear on the external surfaces of a power supply, or on the external surfaces of any equipment connected to it.

Grounding

The best electrical ground for a power supply is the third-wire ground provided in up-to-date ac utility circuits. In an ac outlet, this connection appears as a hole shaped like an uppercase letter D turned on its side. The contacts inside this hole should be connected to a wire that ultimately terminates in a metal rod driven into the earth at the point where the electrical wiring enters the building. That constitutes an *earth ground.*

In older buildings, *two-wire ac systems* are common. These can be recognized by the presence of two slots in the utility outlets, but no ground hole. Some of these systems employ reasonable grounding by means of a scheme called *polarization*, where one slot is longer than the other, the longer slot being connected to electrical ground. But this is not as good as a *three-wire ac system*, in which the ground connection is independent of both the outlet slots.

Unfortunately, the presence of a three-wire or polarized outlet system does not always mean that an appliance connected to an outlet is well grounded. If the appliance design is faulty, or if the ground holes at the outlets were not grounded by the people who installed the electrical system, a power supply can deliver unwanted voltages to the external surfaces of appliances and electronic devices. This can present an electrocution hazard, and can also hinder the performance of equipment connected to the supply.

- **Warning: All exposed metal surfaces of power supplies should be connected to the grounded wire of a three-wire electrical cord. The third prong of the plug should never be defeated or cut off. Some means should be found to ensure that the electrical system in the building has been properly installed, so you don't work under the illusion that your system has a good ground when it actually does not. If you are in doubt about this, consult a professional electrician.**

Surge Currents

At the instant a power supply is switched on, a surge of current occurs, even with nothing connected to the supply output. This is because the filter capacitors need an initial charge, so they draw a large current for a short time. The *surge current* is far greater than the normal operating current. An extreme current surge of this sort can destroy the rectifier diodes if they are not sufficiently rated and/or protected. The phenomenon is worst in high-voltage supplies and voltage-multiplier circuits. Diode failure as a result of current surges can be prevented in at least three ways:

- Use diodes with a current rating of many times the normal operating level.
- Connect several diodes in parallel wherever a diode is called for in the circuit. *Current-equalizing resistors* are necessary (Fig. 21-10). The resistors should have small, identical ohmic values. The diodes should all be identical.
- Use an *automatic switching* circuit in the transformer primary. This type of circuit applies a reduced ac voltage to the transformer for a second or two, and then applies the full input voltage.

Transients

The ac that appears at utility outlets is a sine wave with a constant voltage near 117 V rms or 234 V rms. But there are often *voltage spikes*, known as *transients*, that can attain positive or negative peak values of several thousand volts. Transients are caused by sudden changes in the load in a utility circuit. A thundershower can produce transients throughout an entire town. Unless they are suppressed, transients can destroy the diodes in a power supply. Transients can also cause problems with sensitive electronic equipment such as computers or microcomputer-controlled appliances.

The simplest way to get rid of common transients is to place a small capacitor of about 0.01 μF, rated for 600 V or more, between each side of the transformer primary and electrical ground, as shown in Fig. 21-11. A good component for this purpose is a *disk ceramic capacitor* (not an electrolytic capacitor). Disk ceramic capacitors have no polarity issues. They can be connected in either direction to work equally well.

Commercially made *transient suppressors* are available. These devices, often mistakenly called "surge protectors," use sophisticated methods to prevent sudden voltage spikes from reaching levels where they can cause problems. It is a good idea to use transient suppressors with all sensitive electronic devices, including computers, hi-fi stereo systems, and television sets. In the event of a thundershower, the best way to protect such equipment is to physically unplug it from the wall outlets until the event has passed.

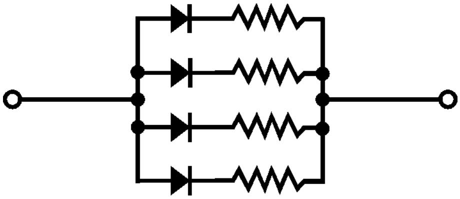

21-10 Diodes in parallel, with current-equalizing resistors in series with each diode.

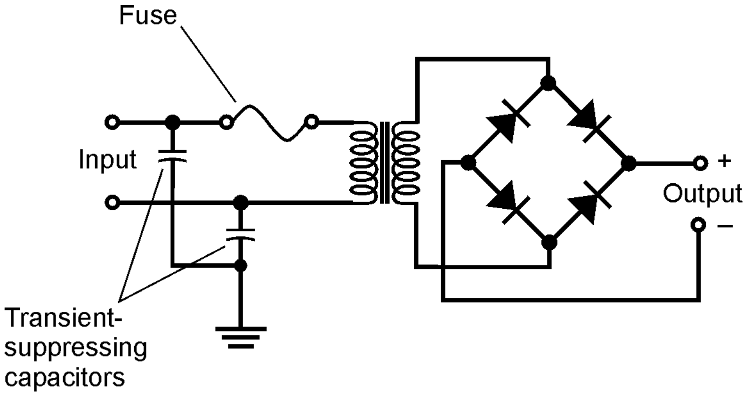

21-11 A full-wave bridge rectifier with transient-suppression capacitors and a fuse in the transformer primary circuit.

Fuses

A *fuse* is a piece of soft wire that melts, breaking a circuit if the current exceeds a certain level. A fuse is placed in series with the transformer primary, as shown in Fig. 21-11. A short circuit or overload anywhere in the power supply, or in equipment connected to it, will burn the fuse out. If a fuse blows out, it must be replaced with another of the same rating. Fuses are rated in amperes (A). Thus, a 5-A fuse will carry up to 5 A before blowing out, and a 20-A fuse will carry up to 20 A.

Fuses are available in two types: the *quick-break fuse* and the *slow-blow fuse.* A quick-break fuse is a straight length of wire or a metal strip. A slow-blow fuse usually has a spring inside along with the wire or strip. It's best to replace blown-out fuses with new ones of the same type. Quick-break fuses in slow-blow situations can burn out needlessly, causing inconvenience. Slow-blow fuses in quick-break environments might not provide adequate protection to the equipment, letting excessive current flow for too long before blowing out.

Circuit Breakers

A *circuit breaker* performs the same function as a fuse, except that a breaker can be reset by turning off the power supply, waiting a moment, and then pressing a button or flipping a switch. Some breakers reset automatically when the equipment has been shut off for a certain length of time. Circuit breakers are rated in amperes, just like fuses.

If a fuse or breaker keeps blowing out or tripping, or if it blows or trips immediately after it has been replaced or reset, then something is wrong with the power supply or with the equipment connected to it. Burned-out diodes, a bad transformer, and shorted filter capacitors in the supply can all cause this trouble. A short circuit in the equipment connected to the supply, or the connection of a device in the wrong direction (polarity), can cause repeated fuse blowing or circuit-breaker tripping.

Never replace a fuse or breaker with a larger-capacity unit to overcome the inconvenience of repeated fuse/breaker blowing/tripping. Find the cause of the trouble, and repair the equipment as needed. The "penny in the fuse box" scheme can endanger equipment and personnel, and it increases the risk of fire in the event of a short circuit.

The Complete System

Figure 21-12 is a block diagram of a complete power supply. Note the sequence in which the portions of the system, called *stages*, are connected. A final note of warning is in order here:

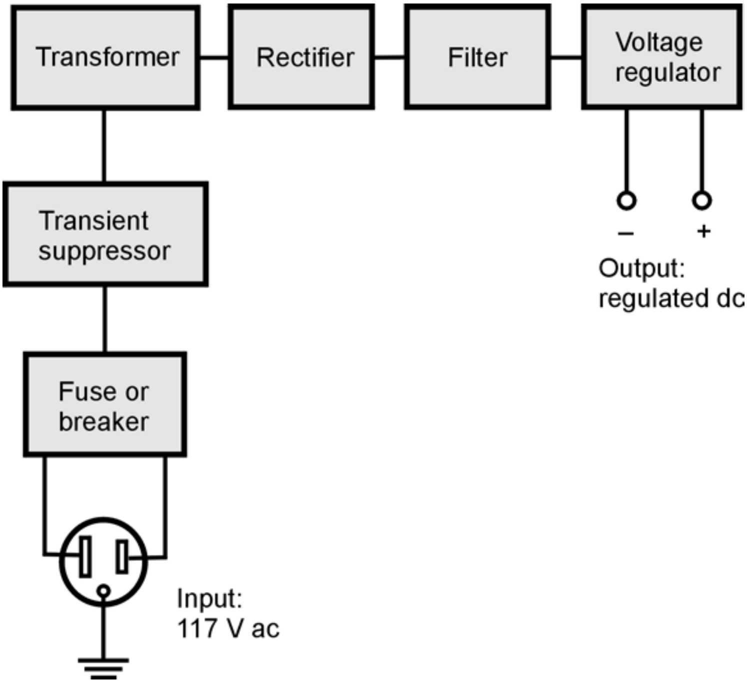
Transformer
Rectifier
Filter
Voltage regulator
Transient suppressor
Fuse or breaker
–
+
Output: regulated dc
Input: 117 V ac

21-12 Block diagram of a complete power supply that can deliver high-quality dc output with ac input.

- **High-voltage power supplies can retain deadly voltages after they have been switched off and unplugged. This is because the filter capacitors retain their charge for some time. If you have any doubt about your ability to safely build or work with a power supply, leave it to a professional.**

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. The output of a rectifier circuit without a filter included is
   - (a) 60 Hz ac.
   - (b) smooth dc.
   - (c) pulsating dc.
   - (d) 120 Hz ac.

2. Which of the following components is not necessarily required in a power supply designed to produce 12-V dc output with 117-V rms ac input?
   - (a) The transformer
   - (b) The filter
   - (c) The rectifier
   - (d) All of the above components are required.

3. Of the following appliances, which would need the biggest transformer?
   (a) A clock radio
   (b) A television broadcast transmitter
   (c) A shortwave radio receiver
   (d) A home television set

4. An advantage of full-wave bridge rectification is the fact that
   (a) it uses the whole transformer secondary for the entire ac input cycle.
   (b) it costs less than other rectifier types.
   (c) it cuts off half of the ac wave cycle.
   (d) it never needs a filter.

5. In a power supply designed to provide high power at low voltage, the best rectifier circuit would probably be the
   (a) half-wave arrangement.
   (b) full-wave, center-tap arrangement.
   (c) quarter-wave arrangement.
   (d) voltage doubler arrangement.

6. The part of a power supply immediately preceding the regulator is
   (a) the transformer.
   (b) the rectifier.
   (c) the filter.
   (d) the ac input.

7. If a half-wave rectifier is used with 165-V pk ac input, the effective dc output voltage is
   (a) considerably less than 165 V.
   (b) slightly less than 165 V.
   (c) exactly 165 V.
   (d) slightly more than 165 V.

8. If a full-wave bridge circuit is used with a transformer whose secondary provides 50 V rms, the peak voltage that occurs across the diodes in the reverse direction is approximately
   (a) 50 V pk.
   (b) 70 V pk.
   (c) 100 V pk.
   (d) 140 V pk.

9. What is the principal disadvantage of a voltage-doubler power supply circuit?
   (a) Excessive current
   (b) Excessive voltage
   (c) Insufficient rectification
   (d) Poor regulation under heavy loads

10. Suppose a transformer secondary provides 10-V rms ac to a voltage-doubler circuit. What is the approximate dc output voltage with no load?
    - (a) 14 V
    - (b) 20 V
    - (c) 28 V
    - (d) 36 V

11. The ripple frequency from a full-wave rectifier is
    - (a) twice that from a half-wave circuit.
    - (b) the same as that from a half-wave circuit.
    - (c) half that from a half-wave circuit.
    - (d) ¼ that from a half-wave circuit.

12. Which of the following would make the best filter for a power supply?
    - (a) A capacitor in series
    - (b) A choke in series
    - (c) A capacitor in series and a choke in parallel
    - (d) A capacitor in parallel and a choke in series

13. If you need exceptionally good ripple filtering for a power supply, which of the following alternatives will yield the best results?
    - (a) Connect several capacitors in parallel.
    - (b) Use a choke-input filter.
    - (c) Connect several chokes in series.
    - (d) Use two capacitor/choke filtering sections in cascade.

14. Voltage regulation can be accomplished by a Zener diode connected in
    - (a) parallel with the filter output, forward-biased.
    - (b) parallel with the filter output, reverse-biased.
    - (c) series with the filter output, forward-biased.
    - (d) series with the filter output, reverse-biased.

15. A current surge takes place when a power supply is first turned on because
    - (a) the transformer core is suddenly magnetized.
    - (b) the diodes suddenly start to conduct.
    - (c) the filter capacitor(s) must be initially charged.
    - (d) arcing takes place in the power switch.

16. Transient suppression is of importance mainly because it minimizes the risk of
    - (a) diode failure.
    - (b) transformer imbalance.
    - (c) filter capacitor overcharging.
    - (d) poor voltage regulation.

17. If a fuse blows, and it is replaced with one having a lower current rating, there is a good chance that
    - (a) the power supply will be severely damaged.
    - (b) the diodes will not rectify.
    - (c) the fuse will blow out right away.
    - (d) transient suppressors won't work.

18. Suppose you see a fuse with nothing but a straight wire inside. You can assume that this fuse
    - (a) is a slow-blow type.
    - (b) is a quick-break type.
    - (c) has a low current rating.
    - (d) has a high current rating.

19. In order to minimize the risk of diode destruction as a result of surge currents that can occur when a power supply is first switched on, which of the following techniques can be useful?
    - (a) Connecting multiple diodes in parallel, with low-value resistors in series with each diode
    - (b) Connecting multiple diodes in parallel, with low-value capacitors in series with each diode
    - (c) Connecting multiple diodes in series, with low-value chokes across each diode
    - (d) Connecting multiple diodes in series, with low-value resistors across each diode

20. To repair a damaged power supply with which you are not completely familiar, you should
    - (a) install bleeder resistors before beginning your work.
    - (b) remove the fuse before beginning your work.
    - (c) leave it alone and have a professional work on it.
    - (d) short out all the diodes before beginning your work.

22
CHAPTER

The Bipolar Transistor

THE WORD *TRANSISTOR* IS A CONTRACTION OF "CURRENT-*TRANS*FERRING RES*ISTOR*." A *BIPOLAR transistor* has two P-N junctions. There are two configurations: a P-type layer sandwiched between two N-type layers, or an N-type layer between two P-type layers.

NPN versus PNP

A simplified drawing of an *NPN bipolar transistor* is shown in Fig. 22-1A, and the schematic symbol is shown in Fig. 22-1B. The P-type, or center, layer is called the *base.* One of the N-type semiconductor layers is the *emitter*, and the other is the *collector.* Sometimes these are labeled *B*, *E*, and *C* in schematic diagrams. A *PNP bipolar transistor* has two P-type layers, one on either side of a thin N-type layer (Fig. 22-2A). The schematic symbol is shown in Fig. 2-22B.

It's easy to tell whether a bipolar transistor in a diagram is NPN or PNP. If the device is NPN, the arrow at the emitter points outward. If the device is PNP, the arrow at the emitter points inward.

Generally, PNP and NPN transistors can perform the same functions. The differences are the polarities of the voltages and the directions of the resulting currents. In most applications, an NPN device can be replaced with a PNP device or vice versa, the power-supply polarity can be reversed, and the circuit will work in the same way—as long as the new device has the appropriate specifications.

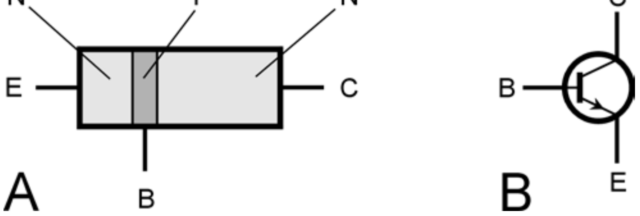

22-1 At A, pictorial diagram of an NPN transistor. At B, the schematic symbol. Electrodes are E = emitter, B = base, and C = collector.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

22-2 At A, pictorial diagram of a PNP transistor. At B, the schematic symbol. Electrodes are E = emitter, B = base, and C = collector.

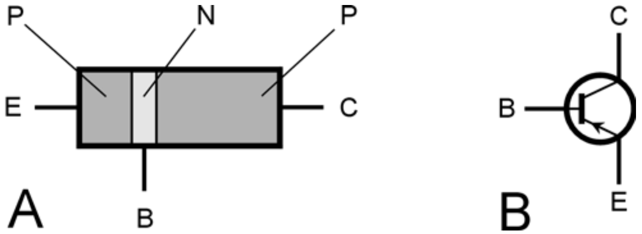

Biasing

Imagine a bipolar transistor as consisting of two diodes in reverse series. You can't normally connect two diodes this way and get a working transistor, but the analogy is good for *modeling* the behavior of bipolar transistors. A dual-diode NPN transistor model is shown in Fig. 22-3A. The base is formed by the connection of the two anodes. The emitter is one of the cathodes, and the collector is the other cathode. Figure 22-3B shows the equivalent real-world NPN transistor circuit.

The NPN Case

The normal method of biasing an NPN transistor is to have the collector voltage positive with respect to the emitter. This is shown by the connection of the battery in Figs. 22-3A and 22-3B. Typical dc voltages for a transistor power supply range between 3 V and about 50 V. A typical voltage is 12 V.

In the model and also in the real-world transistor circuit, the base is labeled "control," because the flow of current through the transistor depends critically on what happens at this electrode.

Zero Bias for NPN

Suppose the base of a transistor is at the same voltage as the emitter. This is known as *zero bias.* When the forward bias is zero, the emitter-base current, often called simply *base current* and de-

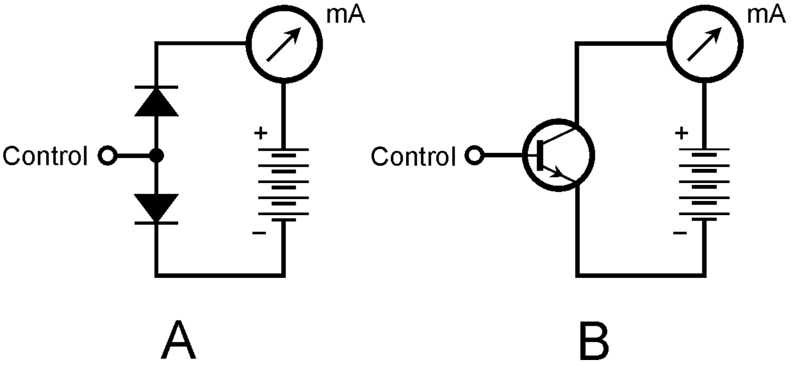

22-3 At A, the dual-diode model of a simple NPN circuit. At B, the actual transistor circuit.

noted I_B, is zero, and the emitter-base (E-B) junction does not conduct. This prevents current from flowing between the emitter and collector, unless a signal is injected at the base to change the situation. Such a signal must, at least momentarily, attain a positive voltage equal to or greater than the forward breakover voltage of the E-B junction.

Reverse Bias for NPN

Now imagine that a second battery is connected between the base and the emitter in the circuit of Fig. 22-3B, with the polarity such that E_B becomes negative with respect to the emitter. The addition of this new battery will cause the E-B junction to be *reverse-biased.* No current flows through the E-B junction in this situation (as long as the new battery voltage is not so great that avalanche breakdown occurs). A signal might be injected at the base to cause a flow of current, but such a signal must attain, at least momentarily, a positive voltage high enough to overcome both the reverse bias and the forward breakover voltage of the junction.

Forward Bias for NPN

Now suppose that E_B is made positive with respect to the emitter, starting at small voltages and gradually increasing. This is *forward bias.* If the forward bias is less than the forward breakover voltage, no current will flow. But as the base voltage E_B reaches the breakover point, the E-B junction will start to conduct.

The base-collector (B-C) junction of a bipolar transistor is normally reverse-biased. It will remain reverse-biased as long as E_B is less than the supply voltage (in this case 12 V). In practical transistor circuits, it is common for E_B to be set at a fraction of the supply voltage. Despite the reverse bias of the B-C junction, a significant emitter-collector current, called *collector current* and denoted I_C, will flow once the E-B junction conducts.

In a real transistor circuit such as the one shown in Fig. 22-3B, the meter reading will jump when the forward breakover voltage of the E-B junction is reached. Then even a small rise in E_B, attended by a rise in I_B, will cause a large increase in I_C, as shown in Fig. 22-4.

If E_B continues to rise, a point will eventually be reached where the I_C versus E_B curve levels off. The transistor is then said to be *saturated* or *in saturation.* It is wide open, conducting as much as it can.

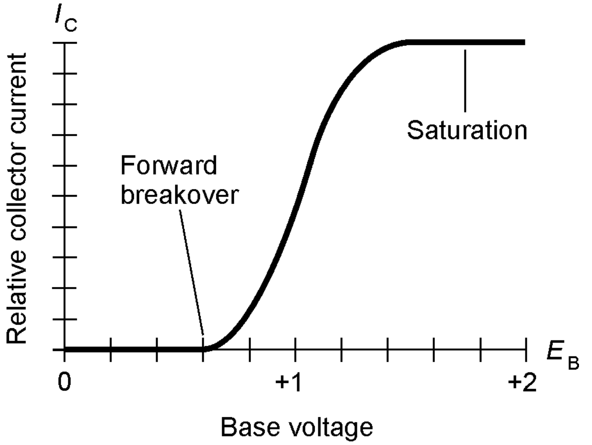

22-4 Relative collector current (I_C) as a function of base voltage (E_B) for a hypothetical NPN silicon transistor.

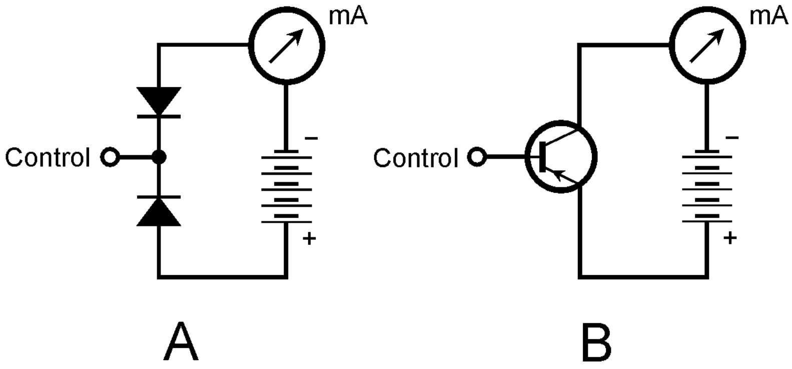

22-5 At A, the dual-diode model of a simple PNP circuit. At B, the actual transistor circuit.

PNP Biasing

For a PNP transistor, the situation is a mirror image of the case for an NPN device. The diodes are reversed, the arrow points inward rather than outward in the transistor symbol, and all the polarities are reversed. The dual-diode PNP model, along with the real-world transistor circuit, are shown in Fig. 22-5. In the preceding discussion, replace every occurrence of the word "positive" with the word "negative." Qualitatively, the same things happen: small changes in E_B cause small changes in I_B, which in turn produce large fluctuations in I_C.

Biasing for Amplification

Because a small change in I_B causes a large variation in I_C when the bias is just right, a transistor can operate as a *current amplifier.* If you look at Fig. 22-6, you'll see that there are some bias values at which a transistor won't provide any current amplification. If the transistor is in saturation, the I_C

22-6 Three different transistor bias points. The most amplification is obtained when the bias is near the middle of the straight-line portion of the curve.

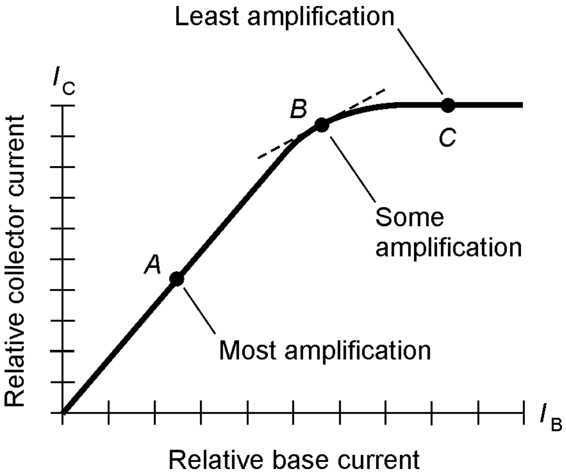

versus I_B curve is horizontal. A small change in I_B, in these portions of the curve, causes little or no change in I_C. But if the transistor is biased near the middle of the straight-line part of the curve in Fig. 22-6, the transistor will work as a current amplifier.

The same situation holds for the curve in Fig. 22-4. At some bias points, a small change in E_B does not produce much, if any, change in I_C; at other points, a small change in E_B produces a dramatic change in I_C. Whenever we want a transistor to amplify a signal, it's important that it be biased in such a way that a small change in the base current or voltage will result in a large change in the collector current.

Static Current Amplification

Current amplification is often called *beta* by engineers. It can range from a factor of just a few times up to hundreds of times. The beta of a transistor can be expressed as the *static forward current transfer ratio*, abbreviated H_{FE}. Mathematically, this is the collector current divided by the base current:

$$H_{FE} = I_C / I_B$$

For example, if a base current, I_B, of 1 mA results in a collector current, I_C, of 35 mA, then $H_{FE} = 35/1 = 35$. If $I_B = 0.5$ mA and $I_C = 35$ mA, then $H_{FE} = 35/0.5 = 70$. The H_{FE} specification for a particular transistor represents the greatest amount of current amplification that can be obtained with it.

Dynamic Current Amplification

A more practical way to define current amplification is as the ratio of the difference in I_C to the difference in I_B that occurs when a small signal is applied to the base of a transistor. Abbreviate the words "the difference in" by *d.* Then, according to this second definition:

$$\text{Current amplification} = dI_C / dI_B$$

Figure 22-6 is a graph of the collector current as a function of the base current (I_C versus I_B) for a hypothetical transistor. Three different points are shown, corresponding to three different bias scenarios. The ratio dI_C/dI_B is different for each of the points in this graph. Geometrically, dI_C/dI_B at a given point is the slope of a line tangent to the curve at that point. The tangent line for point *B* in Fig. 22-6 is a dashed straight line; the tangent lines for points *A* and *C* lie right along the curve and are therefore not shown. The steeper the slope of the line, the greater is dI_C/dI_B. Point *A* provides the highest value of dI_C/dI_B, provided the input signal is not too strong. This value is very close to H_{FE}.

For small-signal amplification, point *A* in Fig. 22-6 represents a good bias level. Engineers would say that it's a good *operating point.* At point *B*, dI_C/dI_B is smaller than at point *A*, so point *B* is not as good for small-signal amplification. At point *C*, dI_C/dI_B is practically zero. The transistor won't amplify much, if at all, if it is biased at this point.

Overdrive

Even when a transistor is biased for the greatest possible current amplification (at or near point *A* in Fig. 22-6), a strong ac input signal can drive it to point *B* or beyond during part of the signal cycle.

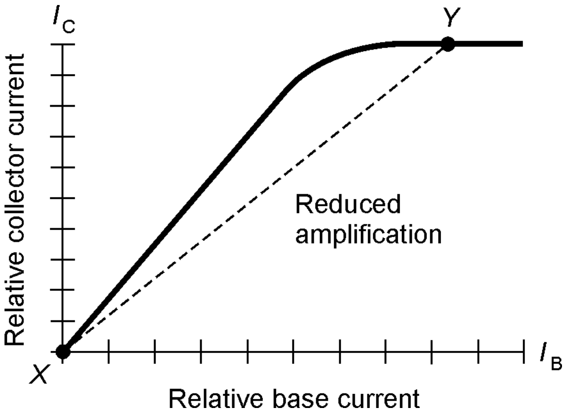

22-7 Excessive input reduces amplification.

Then, dI_C/dI_B is reduced, as shown in Fig. 22-7. Points X and Y in the graph represent the instantaneous current extremes during the signal cycle in this particular case.

When conditions are like those in Fig. 22-7, a transistor amplifier will cause *distortion* in the signal. This means that the output wave will not have the same shape as the input wave. This phenomenon is known as *nonlinearity.* It can sometimes be tolerated, but often it is undesirable. When the input signal to a transistor amplifier is too strong, the condition is called *overdrive*, and the amplifier is said to be *overdriven.*

Overdrive can cause problems other than signal distortion. An overdriven transistor is in or near saturation during part of the input signal cycle. This reduces circuit efficiency, causes excessive collector current, and can overheat the base-collector (B-C) junction. Sometimes overdrive can destroy a transistor.

Gain versus Frequency

Another important specification for a transistor is the range of frequencies over which it can be used as an amplifier. All transistors have an *amplification factor*, or *gain*, that decreases as the signal frequency increases. Some devices work well only up to a few megahertz; others can be used to several gigahertz.

Gain can be expressed in various ways. In the preceding discussion, you learned a little about *current gain*, expressed as a ratio. You will also hear about *voltage gain* or *power gain* in amplifier circuits. These, too, can be expressed as ratios. For example, if the voltage gain of a circuit is 15, then the output signal voltage (rms, peak, or peak-to-peak) is 15 times the input signal voltage. If the power gain of a circuit is 25, then the output signal power is 25 times the input signal power.

Two expressions are commonly used for the gain-versus-frequency behavior of a bipolar transistor. The *gain bandwidth product*, abbreviated f_T, is the frequency at which the gain becomes equal to 1 with the emitter connected to ground. If you try to make an amplifier using a transistor at a frequency higher than its f_T specification, you are bound to fail. The *alpha cutoff frequency* of a transistor is the frequency at which the gain becomes 0.707 times its value when the input signal frequency is 1 kHz. A transistor can have considerable gain at its alpha cutoff frequency. By looking at this

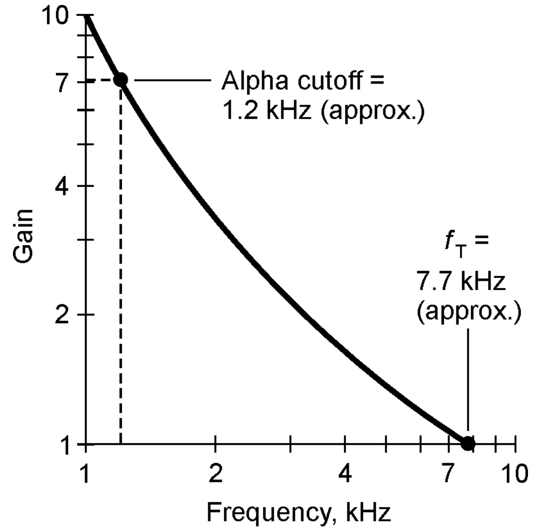

22-8 Alpha cutoff and gain bandwidth product for a hypothetical transistor.

specification for a particular transistor, you can get an idea of how rapidly it loses its ability to amplify as the frequency goes up. Some devices die off faster than others.

Figure 22-8 shows the gain bandwidth product and alpha cutoff frequency for a hypothetical transistor, on a graph of gain versus frequency. Note that the scales of this graph are not linear; that is, the divisions are not evenly spaced. This type of graph is called a *log-log* graph because both scales are *logarithmic* rather than linear.

Common Emitter Circuit

A transistor can be hooked up in three general ways. The emitter can be grounded for signal, the base can be grounded for signal, or the collector can be grounded for signal. An often-used arrangement is the *common emitter circuit.* "Common" means "grounded for the signal." The basic configuration is shown in Fig. 22-9.

A terminal can be at ground potential for a signal, and yet have a significant dc voltage. In the circuit shown, capacitor C_1 appears as a short circuit to the ac signal, so the emitter is at *signal ground.* But resistor R_1 causes the emitter to have a certain positive dc voltage with respect to ground (or a negative voltage, if a PNP transistor is used). The exact dc voltage at the emitter depends on the resistance of R_1, and on the bias. The bias is set by the ratio of the values of resistors R_2 and R_3. The bias can be anything from zero, or ground potential, to +12 V, the supply voltage. Normally it is a couple of volts.

Capacitors C_2 and C_3 block dc to or from the input and output circuitry (whatever that might be) while letting the ac signal pass. Resistor R_4 keeps the output signal from being shorted out through the power supply. A signal enters the common emitter circuit through C_2, where it causes the base current, I_B, to vary. The small fluctuations in I_B cause large changes in the collector current, I_C. This current passes through resistor R_4, causing a fluctuating dc voltage to appear across this resistor. The ac part of this passes unhindered through capacitor C_3 to the output.

The circuit of Fig. 22-9 is the basis for many amplifiers, from audio frequencies through ultrahigh radio frequencies. The common emitter configuration produces the largest gain of any arrangement. The output wave is 180° out of phase with respect to the input wave.

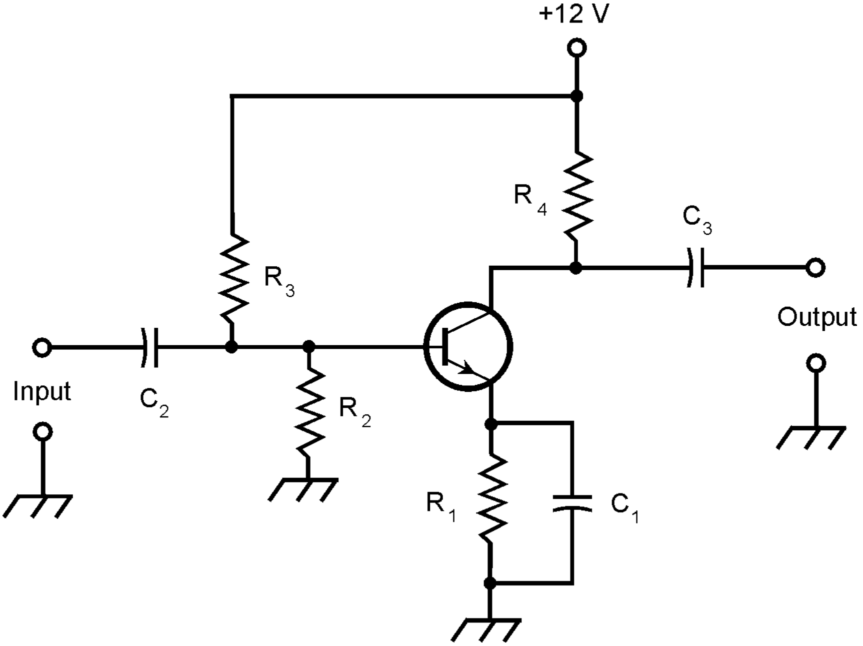

22-9 Common emitter configuration. This diagram shows an NPN transistor circuit.

Common Base Circuit

As its name implies, the *common base circuit* (Fig. 22-10) has the base at signal ground. The dc bias is the same as for the common emitter circuit, but the input signal is applied at the emitter, instead of at the base. This causes fluctuations in the voltage across R_1, causing variations in I_B. The result of these small current fluctuations is a large change in the current through R_4. Therefore, amplification occurs. The output wave is in phase with the input wave.

The signal enters through capacitor C_1. Resistor R_1 keeps the input signal from being shorted to ground. Bias is provided by R_2 and R_3. Capacitor C_2 keeps the base at signal ground. Resistor R_4 keeps the signal from being shorted out through the power supply. The output is taken through C_3.

The common base circuit provides somewhat less gain than a common emitter circuit. However, it is more stable than the common emitter configuration in some applications, especially RF power amplifiers.

Common Collector Circuit

A *common collector circuit* (Fig. 22-11) operates with the collector at signal ground. The input is applied at the base, just as it is with the common emitter circuit. The signal passes through C_2 onto the base of the transistor. Resistors R_2 and R_3 provide the correct bias for the base. Resistor R_4 limits the current through the transistor. Capacitor C_3 keeps the collector at signal ground. A fluctuating direct current flows through R_1, and a fluctuating dc voltage therefore appears across it. The ac

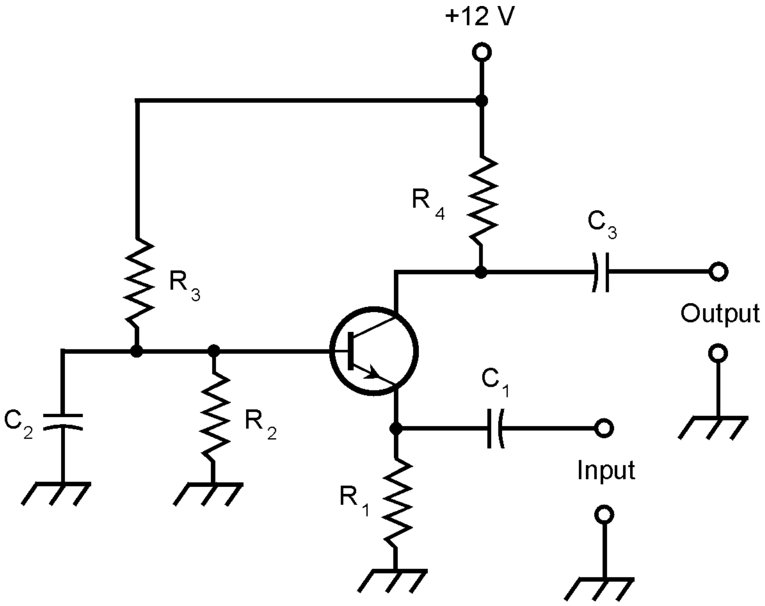
+12 V
R_4
C_3
R_3
Output
C_1
C_2
R_2
Input
R_1

22-10 Common base configuration. This diagram shows an NPN transistor circuit.

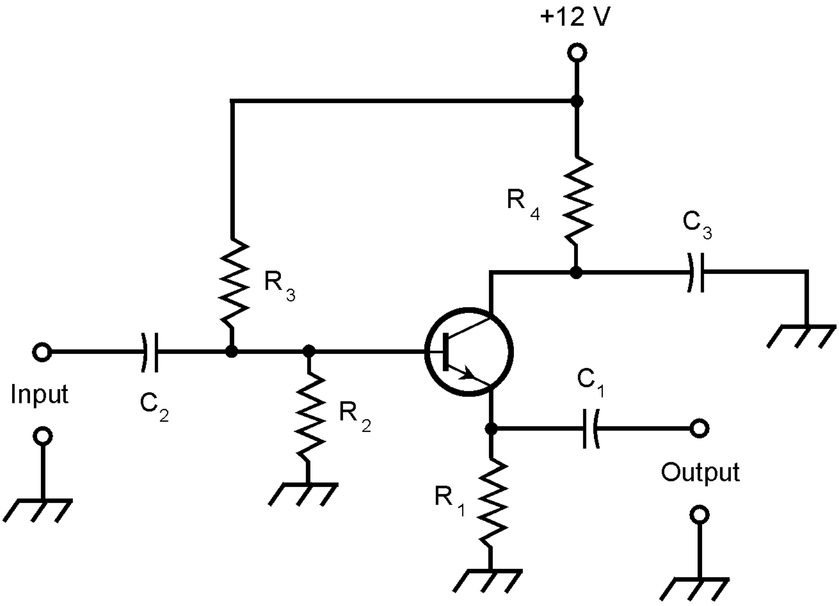
+12 V
R_4
C_3
R_3
Input
C_2
C_1
R_2
Output
R_1

22-11 Common collector configuration, also known as an emitter follower. This diagram shows an NPN transistor circuit.

part of this voltage passes through C_1 to the output. Because the output follows the emitter current, this circuit is sometimes called an *emitter follower circuit.*

The output wave of a common collector circuit is in phase with the input wave. This circuit is unique because its input impedance is high, while its output impedance is low. For this reason, the common collector circuit can be used to match high impedances to low impedances. When well designed, an emitter follower works over a wide range of frequencies, and is a low-cost alternative to a broadband impedance-matching transformer.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. In a PNP circuit, the collector
   (a) has an arrow pointing inward.
   (b) is positive with respect to the emitter.
   (c) is biased at a small fraction of the base bias.
   (d) is negative with respect to the emitter.

2. In many cases, a PNP transistor can be replaced with an NPN device and the circuit will do the same thing, provided that
   (a) the power supply or battery polarity is reversed.
   (b) the collector and emitter leads are interchanged.
   (c) the arrow is pointing inward.
   (d) Forget it! A PNP transistor can never be replaced with an NPN transistor.

3. A bipolar transistor has
   (a) three P-N junctions.
   (b) three semiconductor layers.
   (c) two N-type layers around a P-type layer.
   (d) a low avalanche voltage.

4. In the dual-diode model of an NPN transistor, the emitter corresponds to
   (a) the point where the cathodes are connected together.
   (b) the point where the cathode of one diode is connected to the anode of the other.
   (c) the point where the anodes are connected together.
   (d) either of the diode cathodes.

5. The current through a transistor depends on
   (a) E_C.
   (b) E_B relative to E_C.
   (c) I_B.
   (d) more than one of the above.

6. With no signal input, a bipolar transistor would have the least I_C when
   - (a) the emitter is grounded.
   - (b) the E-B junction is forward-biased.
   - (c) the E-B junction is reverse-biased.
   - (d) the E-B current is high.

7. When a transistor is conducting as much as it can, it is said to be
   - (a) in a state of cutoff.
   - (b) in a state of saturation.
   - (c) in a state of reverse bias.
   - (d) in a state of avalanche breakdown.

8. Refer to the curve shown in Fig. 22-12. Which operating point is best if a large amplification factor is desired with a weak signal input?
   - (a) Point *A*
   - (b) Point *B*
   - (c) Point *C*
   - (d) Point *D*

9. In Fig. 22-12, the forward breakover point for the E-B junction is nearest to
   - (a) no point on this graph.
   - (b) point *B*.
   - (c) point *C*.
   - (d) point *D*.

10. In Fig. 22-12, saturation is nearest to
   - (a) point *A*.
   - (b) point *B*.
   - (c) point *C*.
   - (d) point *D*.

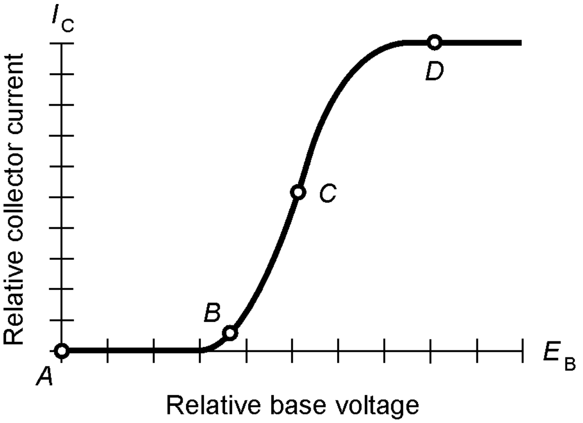

22-12 Illustration for Quiz Questions 8 through 11.

11. In Fig. 22-12, the greatest gain occurs at
    (a) point *A*.
    (b) point *B*.
    (c) point *C*.
    (d) point *D*.

12. In a common emitter circuit, the gain bandwidth product is
    (a) the frequency at which the gain is 1.
    (b) the frequency at which the gain is 0.707 times its value at 1 MHz.
    (c) the frequency at which the gain is greatest.
    (d) the difference between the frequency at which the gain is greatest, and the frequency at which the gain is 1.

13. The bipolar-transistor configuration most often used for matching a high input impedance to a low output impedance puts signal ground at
    (a) the emitter.
    (b) the base.
    (c) the collector.
    (d) any point; it doesn't matter.

14. The output is in phase with the input in
    (a) a common emitter circuit.
    (b) a common base circuit.
    (c) a common collector circuit.
    (d) more than one of the above.

15. The greatest possible amplification is obtained in
    (a) a common emitter circuit.
    (b) a common base circuit.
    (c) a common collector circuit.
    (d) more than one of the above.

16. The input is applied to the collector in
    (a) a common emitter circuit.
    (b) a common base circuit.
    (c) a common collector circuit.
    (d) none of the above.

17. The configuration noted for its stability in RF power amplifiers is the
    (a) common emitter circuit.
    (b) common base circuit.
    (c) common collector circuit.
    (d) emitter follower circuit.

18. In a common base circuit, the output is taken from
    - (a) the emitter.
    - (b) the base.
    - (c) the collector.
    - (d) more than one of the above.

19. Suppose that the input signal to a transistor amplifier results in saturation during part of the cycle. This produces
    - (a) the greatest possible amplification.
    - (b) reduced efficiency.
    - (c) avalanche effect.
    - (d) nonlinear output impedance.

20. Suppose that the gain of a transistor in a common emitter circuit is 100 at a frequency of 1 kHz, and the gain is 70.7 at 335 kHz. The gain drops to 1 at 210 MHz. The alpha cutoff frequency is
    - (a) 1 kHz.
    - (b) 335 kHz.
    - (c) 210 MHz.
    - (d) impossible to define based on this data.

23
CHAPTER

The Field Effect Transistor

THE BIPOLAR TRANSISTOR ISN'T THE ONLY SEMICONDUCTOR TRANSISTOR THAT CAN AMPLIFY. THE other major category of transistor is the *field effect transistor* (FET). There are two main types of FET: the *junction FET* (JFET) and the *metal-oxide FET* (MOSFET).

Principle of the JFET

In a JFET, the current varies because of the effects of an electric field within the device. Charge carriers (electrons or holes) flow from the *source* (S) electrode to the *drain* (D) electrode. This results in a *drain current,* I_D, that is normally the same as the *source current,* I_S. The rate of flow of charge carriers—that is, the current—depends on the voltage at a control electrode called the *gate* (G). Fluctuations in *gate voltage*, E_G, cause changes in the current through the *channel*, which is the path between the source and the drain. The current through the channel is normally equal to I_D. Small fluctuations in E_G can cause large variations in I_D. This fluctuating drain current can produce significant fluctuations in the voltage across an output resistance.

N-Channel versus P-Channel

A simplified drawing of an *N-channel JFET*, and its schematic symbol, are shown in Fig. 23-1. The N-type material forms the channel, or the path for charge carriers. The majority carriers are electrons. The drain is placed at a positive dc voltage with respect to the source.

In an N-channel device, the gate consists of P-type material. Another section of P-type material, called the *substrate*, forms a boundary on the side of the channel opposite the gate. The voltage on the gate produces an electric field that interferes with the flow of charge carriers through the channel. The more negative E_G becomes, the more the electric field chokes off the current through the channel, and the smaller I_D becomes.

A *P-channel JFET* (Fig. 23-2) has a channel of P-type semiconductor. The majority charge carriers are holes. The drain is negative with respect to the source. The more positive E_G gets, the more the electric field chokes off the current through the channel, and the smaller I_D becomes.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

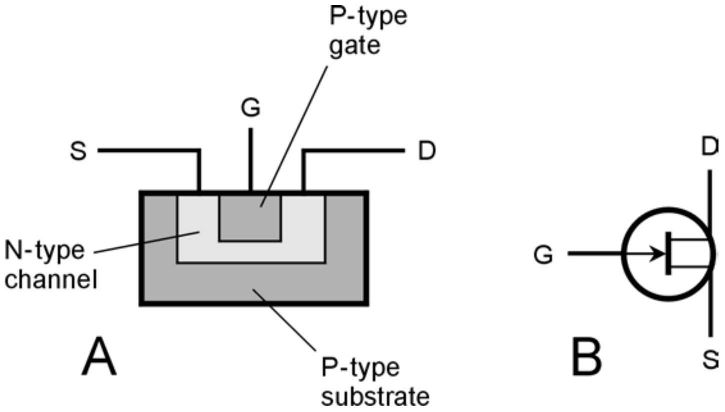

23-1 At A, pictorial diagram of an N-channel JFET. At B, the schematic symbol. Electrodes are S = source, G = gate, and D = drain.

You can recognize the N-channel JFET in schematic diagrams by the arrow pointing inward at the gate, and the P-channel JFET by the arrow pointing outward. Also, you can tell which is which (sometimes arrows are not included in schematic diagrams) by the power-supply polarity. A positive drain indicates an N-channel JFET, and a negative drain indicates a P-channel JFET.

In electronic circuits, N-channel and P-channel devices can do the same kinds of things. The main difference is the polarity. An N-channel device can almost always be replaced with a P-channel JFET, and the power-supply polarity reversed, and the circuit will still work if the new device has the right specifications. Just as there are different kinds of bipolar transistors, there are various types of JFETs, each suited to a particular application. Some JFETs work well as weak-signal amplifiers and oscillators; others are made for power amplification.

Field effect transistors have certain advantages over bipolar transistors. Perhaps the most important is that FETs are available that generate less internal noise than bipolar transistors. This makes them excellent for use as weak-signal amplifiers at very high or ultrahigh frequencies. Field effect transistors have high input impedance, which can also be an advantage in weak-signal amplifiers.

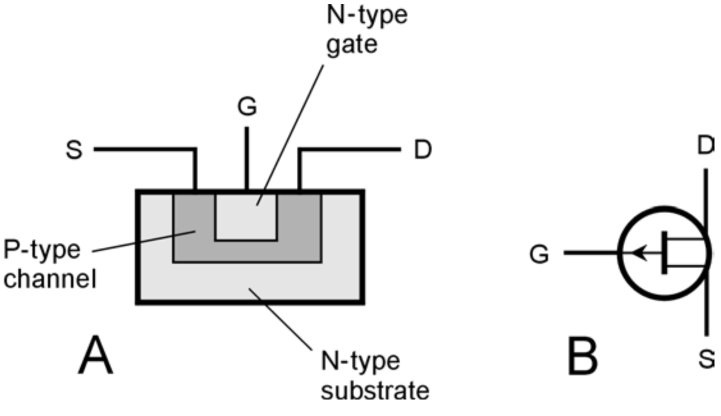

23-2 At A, pictorial diagram of a P-channel JFET. At B, the schematic symbol. Electrodes are S = source, G = gate, and D = drain.

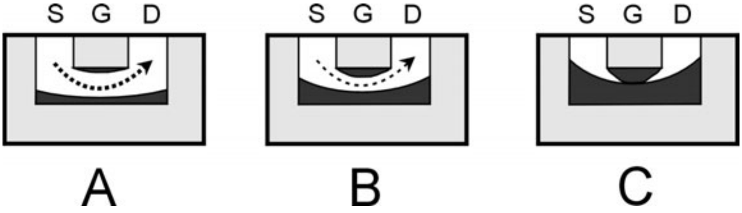

23-3 At A, the depletion region (darkest area) is narrow, the channel (white area) is wide, and many charge carriers (heavy dashed line) flow. At B, the depletion region is wider, the channel is narrower, and fewer charge carriers flow. At C, the depletion region obstructs the channel, and no charge carriers flow.

Depletion and Pinchoff

The JFET works because the voltage at the gate causes an electric field that interferes, more or less, with the flow of charge carriers along the channel. A simplified drawing of the situation for an N-channel device is shown in Fig. 23-3.

As the drain voltage E_D increases, so does the drain current I_D, up to a certain level-off value. This is true as long as the gate voltage E_G is constant, and is not too large negatively. But as E_G becomes increasingly negative (Fig. 23-3A), a *depletion region* (shown as a solid dark area) begins to form in the channel. Charge carriers cannot flow in this region; they must pass through a narrowed channel. The more negative E_G becomes, the wider the depletion region gets, as shown in drawing B. Ultimately, if the gate becomes negative enough, the depletion region completely obstructs the flow of charge carriers. This condition is called *pinchoff*, and is illustrated at C.

JFET Biasing

Two biasing methods for N-channel JFET circuits are shown in Fig. 23-4. In Fig. 23-4A, the gate is grounded through resistor R_2. The source resistor, R_1, limits the current through the JFET. The drain current, I_D, flows through R_3, producing a voltage across this resistor. The ac output signal passes through C_2. In Fig. 23-4B, the gate is connected through potentiometer R_2 to a voltage that is negative with respect to ground. Adjusting this potentiometer results in a variable negative E_G between R_2 and R_3. Resistor R_1 limits the current through the JFET. The drain current, I_D, flows through R_4, producing a voltage across it. The ac output signal passes through C_2.

In both of these circuits, the drain is positive relative to ground. For a P-channel JFET, reverse the polarities in Fig. 23-4. Typical power-supply voltages in JFET circuits are comparable to those for bipolar transistor circuits. The voltage between the source and drain, abbreviated E_D, can range from about 3 V to 150 V dc; most often it is 6 to 12 V dc. The biasing arrangement in Fig. 23-4A is preferred for weak-signal amplifiers, low-level amplifiers, and oscillators. The scheme at B is more often employed in power amplifiers having substantial input signal amplitudes.

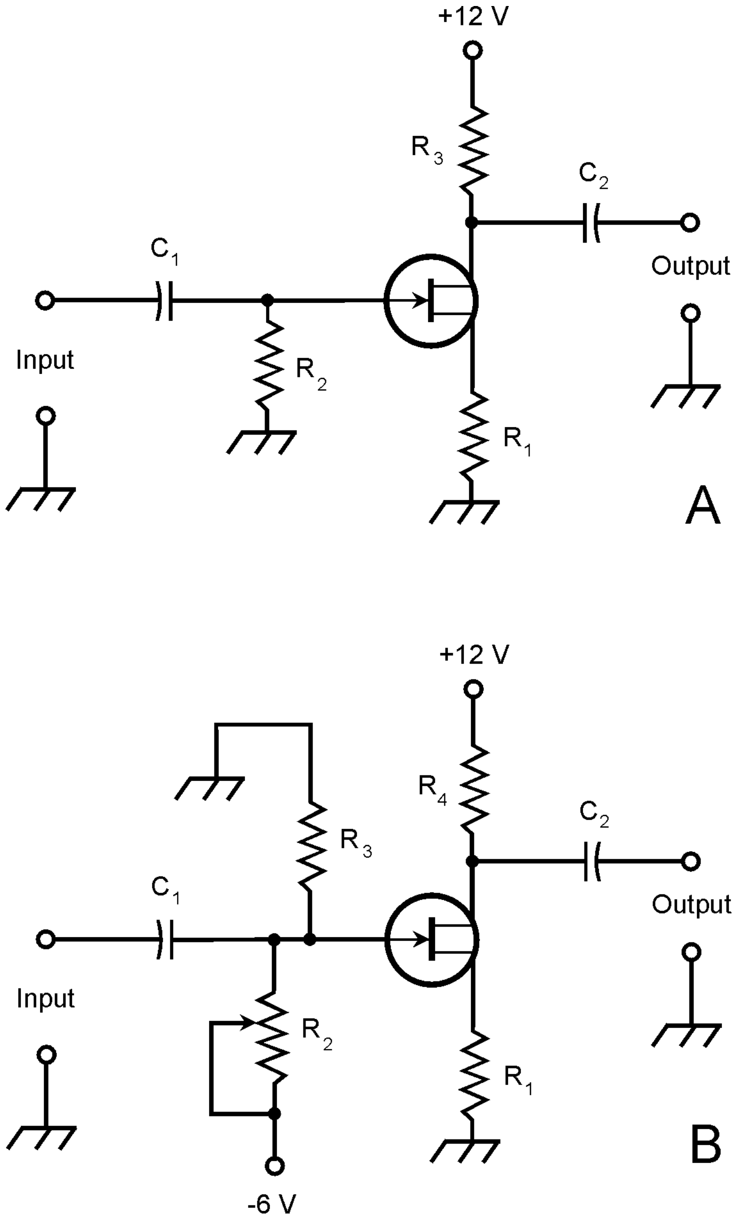
+12 V
R3
C2
C1
Output
Input
R2
R1
A
+12 V
R4
R3
C2
C1
Output
Input
R2
R1
B
-6 V

23-4 Two methods of biasing an N-channel JFET. At A, fixed gate bias; at B, variable gate bias.

Amplification

The graph of Fig. 23-5 shows I_D as a function of E_G for a hypothetical N-channel JFET. The drain voltage, E_D, is assumed to be constant. When E_G is fairly large and negative, the JFET is pinched off, and no current flows through the channel. As E_G gets less negative, the channel opens up, and I_D begins flowing. As E_G gets still less negative, the channel gets wider and I_D increases. As E_G approaches the point where the S-G junction is at forward breakover, the channel conducts as well as it possibly can. If E_G becomes positive enough so the S-G junction conducts, some of the current in the channel leaks out through the gate. This is usually an unwanted phenomenon.

The FET Amplifies Voltage

The best amplification for weak signals is obtained when E_G is such that the slope of the curve in Fig. 23-5 is the greatest. This is shown roughly by the range marked *X.* For power amplification, however, results are often best when the JFET is biased at or beyond pinchoff, in the range marked *Y.*

In either circuit shown in Fig. 23-4, I_D passes through the drain resistor. Small fluctuations in E_G cause large changes in I_D, and these variations in turn produce wide swings in the dc voltage across R_3 (in the circuit at A) or R_4 (in the circuit at B). The ac part of this voltage goes through capacitor C_2, and appears at the output as a signal of much greater ac voltage than that of the input signal at the gate.

Drain Current versus Drain Voltage

Do you suspect that the current I_D, passing through the channel of a JFET, increases in a linear manner with increasing drain voltage E_D? This seems reasonable, but it is not what usually happens. Instead, I_D rises for awhile as E_D increases steadily, and then I_D starts to level off. The current I_D can be plotted graphically as a function of E_D for various values of E_G. When this is done, the result is a *family of characteristic curves* for the JFET. The graph of Fig. 23-6 shows a family of characteristic

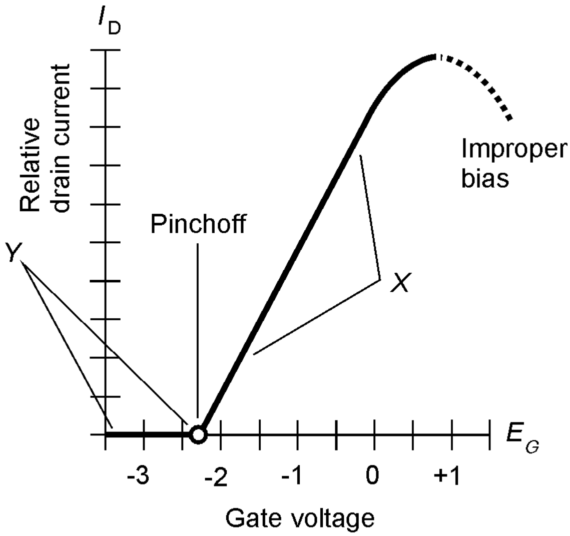

23-5 Relative drain current (I_D) as a function of gate voltage (E_G) for a hypothetical N-channel JFET.

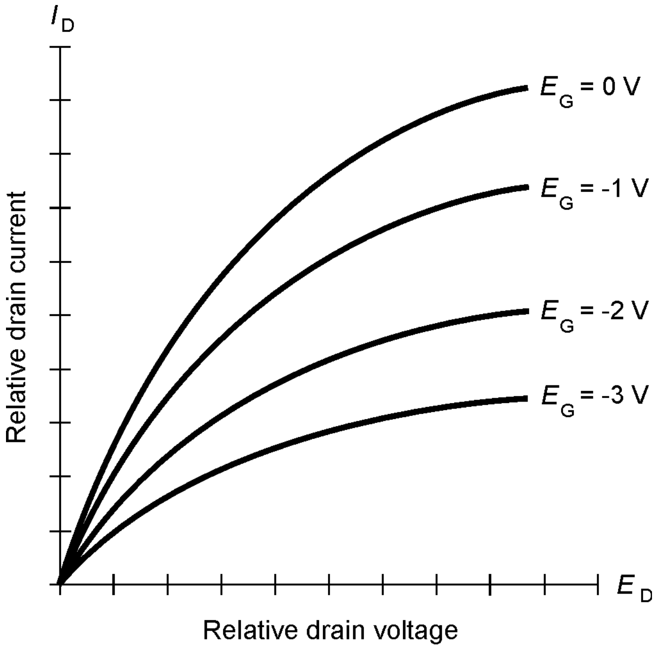

23-6 A family of characteristic curves for a hypothetical N-channel JFET.

curves for a hypothetical N-channel device. The graph of I_D versus E_G, one example of which is shown in Fig. 23-5, is also an important specification that engineers consider when choosing a JFET for a particular application.

Transconductance

Recall the discussion of *dynamic current amplification* from the last chapter. This is a measure of how well a bipolar transistor amplifies a signal. The JFET equivalent of this is called *dynamic mutual conductance* or *transconductance.*

Refer again to Fig. 23-5. Suppose that E_G is a certain value, resulting in a certain current I_D. If the gate voltage changes by a small amount dE_G, then the drain current will change by a certain increment dI_D. The transconductance is the ratio dI_D/dE_G. Geometrically, this translates to the slope of a line tangent to the curve of Fig. 23-5. The value of dI_D/dE_G is not the same at every point along the curve. When the JFET is biased beyond pinchoff, in the region marked *Y*, the slope of the curve is zero. Then there is no fluctuation in I_D when E_G changes by small amounts. There can be a change in I_D when there is a change in E_G only when the channel conducts current. The region where the transconductance, dI_D/dE_G, is the greatest is the region marked *X*, where the slope of the curve is steepest. This region of the curve represents conditions where the most gain can be obtained from the device.

The MOSFET

The acronym *MOSFET* (pronounced "*moss*-fet") stands for *m*etal-*o*xide-*s*emiconductor *f*ield *e*ffect *t*ransistor. A simplified cross-sectional drawing of an N-channel MOSFET, along with the schematic symbol, is shown in Fig. 23-7. The P-channel device is shown in the drawings of Fig. 23-8.

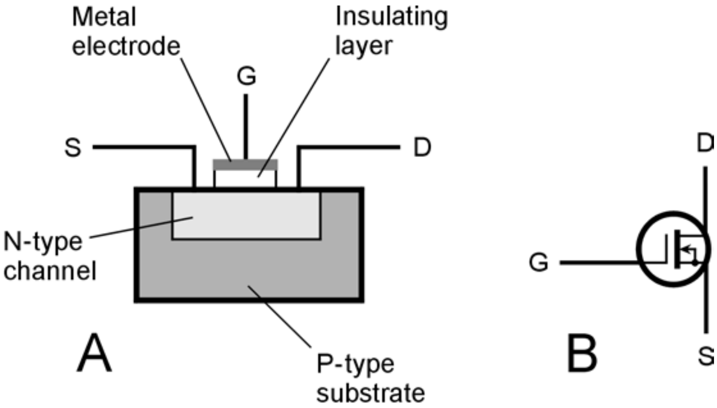
Metal
electrode
Insulating
layer
G
S
D
N-type
channel
P-type
substrate
A
D
G
S
B

23-7 At A, pictorial diagram of an N-channel MOSFET. At B, the schematic symbol. Electrodes are S = source, G = gate, and D = drain.

Extremely High Input Impedance

When the MOSFET was first developed, it was called an *insulated-gate FET* or *IGFET.* That's still a good name for it. The gate electrode is insulated from the channel by a thin layer of dielectric material. As a result, the input impedance is even higher than that of a JFET. The gate-to-source resistance of a typical MOSFET is, as a matter of fact, comparable to that of a typical capacitor! This means that a MOSFET draws essentially no current, and therefore no power, from the signal source. This makes the device ideal for low-level and weak-signal amplifiers. But MOS devices aren't quite perfect. They have an Achilles heel. They are electrically fragile.

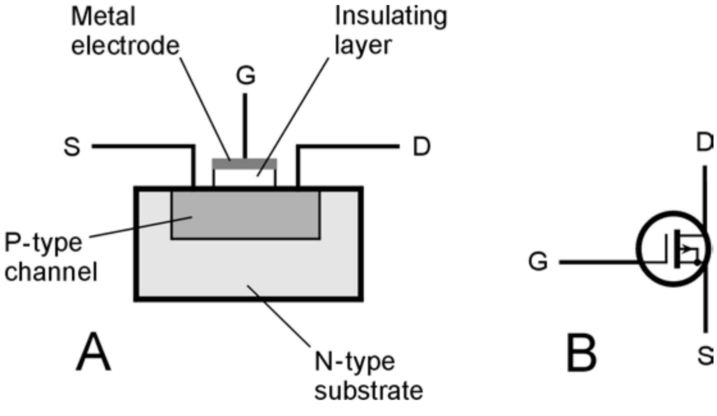
Metal
electrode
Insulating
layer
G
S
D
P-type
channel
N-type
substrate
A
D
G
S
B

23-8 At A, pictorial diagram of a P-channel MOSFET. At B, the schematic symbol. Electrodes are S = source, G = gate, and D = drain.

Beware of Static!

The trouble with MOSFETs is that they can be easily damaged by electrostatic discharges. When building or servicing circuits containing MOS devices, technicians must use special equipment to ensure that their hands don't carry electrostatic charges that might ruin the components. If a discharge occurs through the dielectric of a MOS device, the component is permanently destroyed. Warm and humid climates do not offer total protection against the hazard. This author learned that fact by ruining several MOSFETs while designing circuits in the summertime—in Miami, Florida!

Flexibility in Biasing

In electronic circuits, an N-channel JFET can sometimes be replaced directly with an N-channel MOSFET, and P-channel devices can be similarly interchanged. But the characteristic curves for MOSFETs are not the same as those for JFETs. The main difference is that the S-G junction in a MOSFET is not a P-N junction. Therefore, forward breakover cannot occur. A gate bias voltage, E_G, more positive than +0.6 V can be applied to an N-channel MOSFET, or an E_G more negative than −0.6 V to a P-channel device, without a current leak taking place.

A family of characteristic curves for a hypothetical N-channel MOSFET is shown in the graph of Fig. 23-9.

Depletion Mode versus Enhancement Mode

Normally the channel in a JFET is wide open; as the depletion region gets wider and wider, choking off the channel, the charge carriers are forced to pass through a narrower and narrower path. This is known as the *depletion mode* of operation for a field effect transistor. A MOSFET can also

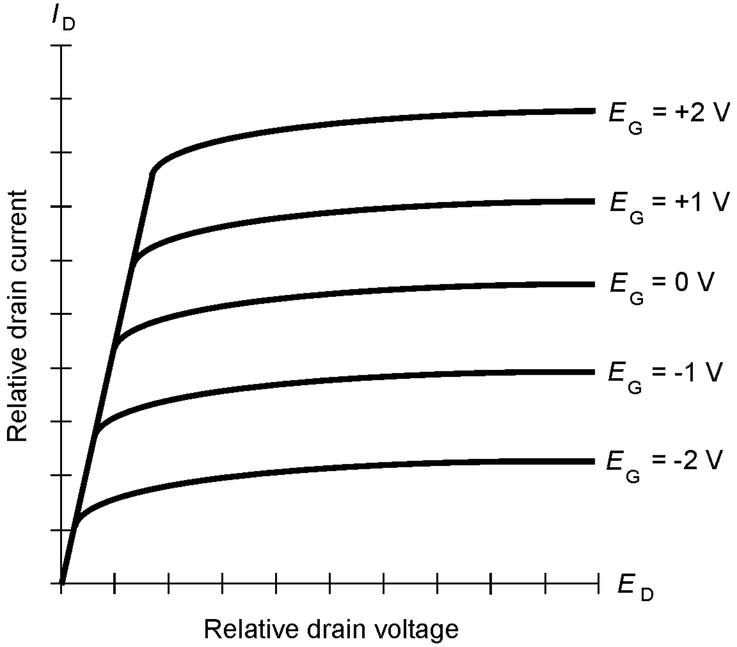

23-9 A family of characteristic curves for a hypothetical N-channel MOSFET.

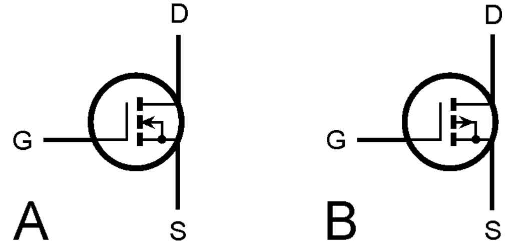

23-10 Schematic symbols for enhancement mode MOSFETs. At A, the N-channel device; at B, the P-channel device.

be made to work in the depletion mode. The drawings and schematic symbols of Figs. 23-7 and 23-8 show depletion-mode MOSFETs.

Metal-oxide semiconductor technology also allows an entirely different means of operation. An *enhancement-mode* MOSFET normally has a pinched-off channel. It is necessary to apply a bias voltage, E_G, to the gate so that a channel will form. If $E_G = 0$ in an enhancement-mode MOSFET, then $I_D = 0$ when there is no signal input. The schematic symbols for N-channel and P-channel enhancement-mode devices are shown in Fig. 23-10. Note that the vertical line is broken. This is how you can recognize an enhancement-mode device in circuit diagrams.

Common Source Circuit

There are three different circuit hookups for FETs, just as there are for bipolar transistors. These three arrangements have the source, the gate, or the drain at signal ground. In the *common source circuit*, the source is placed at signal ground. Signal input is applied to the gate. The general configuration is shown in Fig. 23-11. An N-channel JFET is used here, but the device could be an

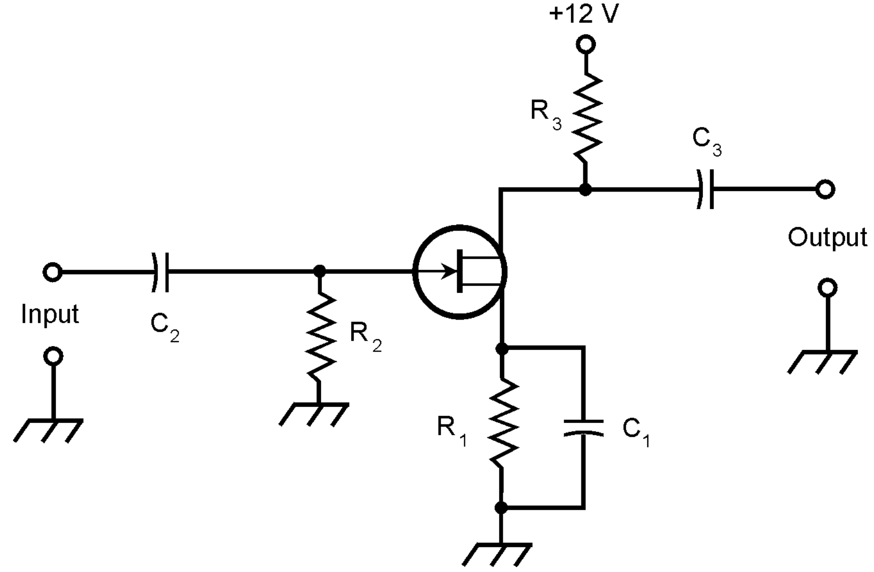

23-11 Common source configuration. This diagram shows an N-channel JFET circuit.

N-channel, depletion-mode MOSFET and the circuit diagram would be the same. For an N-channel enhancement-mode device, an extra resistor would be necessary, running from the gate to the positive power supply terminal. For P-channel devices, the supply would provide a negative, rather than a positive, voltage.

Capacitor C_1 and resistor R_1 place the source at signal ground while elevating the source above ground for dc. The ac signal enters through C_2. Resistor R_2 adjusts the input impedance and provides bias for the gate. The ac signal passes out of the circuit through C_3. Resistor R_3 keeps the output signal from being shorted out through the power supply. The circuit of Fig. 23-11 is the basis for low-level RF amplifiers and oscillators.

The common source arrangement provides the greatest gain of the three FET circuit configurations. The output is 180° out of phase with the input.

Common Gate Circuit

In the *common gate circuit* (Fig. 23-12), the gate is placed at signal ground. The input is applied to the source. The illustration shows an N-channel JFET. For other types of FETs, the same considerations apply as previously described for the common source circuit. Enhancement-mode devices would require a resistor between the gate and the positive supply terminal (or the negative terminal if the MOSFET is P-channel).

The dc bias for the common gate circuit is basically the same as that for the common source arrangement. But the signal follows a different path. The ac input signal enters through C_1. Resistor R_1 keeps the input from being shorted to ground. Gate bias is provided by R_1 and R_2. Capacitor

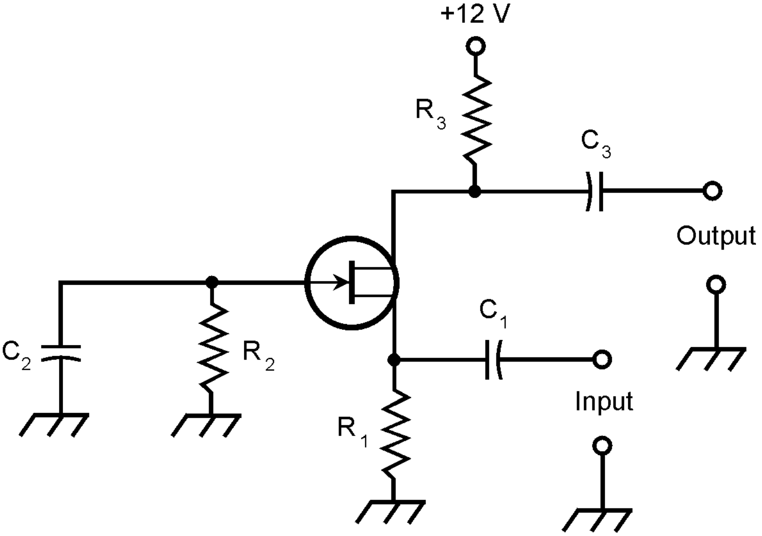

23-12 Common gate configuration. This diagram shows an N-channel JFET circuit.

C_2 places the gate at signal ground. In some common gate circuits, the gate is directly grounded, and R_2 and C_2 are not necessary. The output signal leaves the circuit through C_3. Resistor R_3 keeps the output signal from being shorted through the power supply.

The common gate arrangement produces less gain than its common source counterpart. But a common gate amplifier is not likely to break into unwanted oscillation, making it a good choice for power-amplifier circuits, especially at RF. The output is in phase with the input.

Common Drain Circuit

A *common drain circuit* is shown in Fig. 23-13. In this circuit, the collector is at signal ground. It is sometimes called a *source follower.* The FET is biased in the same way as for the common source and common gate circuits. In the illustration, an N-channel JFET is shown, but any other kind of FET could be used, reversing the polarity for P-channel devices. Enhancement-mode MOSFETs would need a resistor between the gate and the positive supply terminal (or the negative terminal if the MOSFET is P-channel).

The input signal passes through C_2 to the gate. Resistors R_1 and R_2 provide gate bias. Resistor R_3 limits the current. Capacitor C_3 keeps the drain at signal ground. Fluctuating dc (the channel current) flows through R_1 as a result of the input signal; this causes a fluctuating dc voltage to appear across R_1. The output is taken from the source, and its ac component passes through C_1.

The output of the common drain circuit is in phase with the input. This scheme is the FET analog of the bipolar common collector arrangement. The output impedance is rather low, making this circuit a good choice for broadband impedance matching.

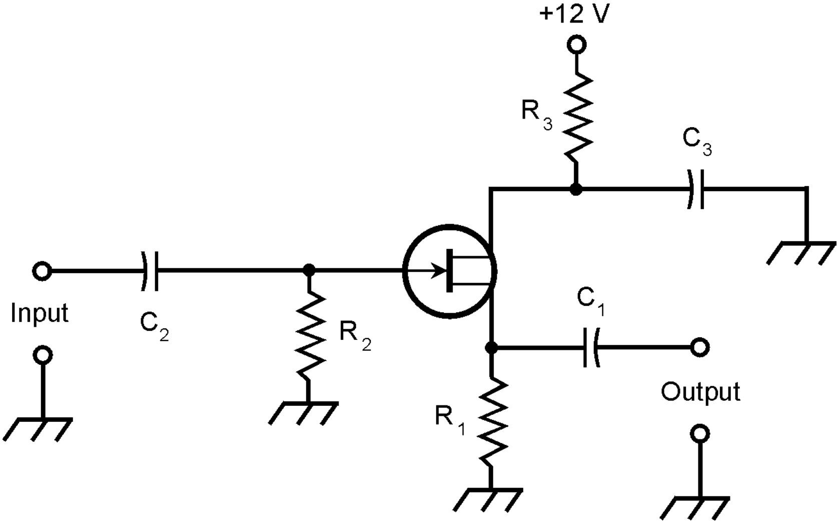

23-13 Common drain configuration, also known as a source follower. This diagram shows an N-channel JFET circuit.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. The current through the channel of a JFET is directly affected by all of the following, except the
   (a) drain voltage.
   (b) transconductance.
   (c) gate voltage.
   (d) gate bias.
2. In an N-channel JFET, pinchoff occurs when the gate bias is
   (a) small and positive.
   (b) zero.
   (c) small and negative.
   (d) large and negative.
3. The current consists mainly of holes when a JFET
   (a) has a P-type channel.
   (b) is forward-biased.
   (c) is zero-biased.
   (d) is reverse-biased.
4. A JFET might work better than a bipolar transistor in
   (a) a high-voltage rectifier.
   (b) a weak-signal RF amplifier.
   (c) a power-supply filter.
   (d) a power transformer.
5. In a P-channel JFET,
   (a) the drain is forward-biased.
   (b) the source-gate junction is forward-biased.
   (c) the drain is negative relative to the source.
   (d) the gate must be at dc ground.
6. A JFET is sometimes biased at or beyond pinchoff in
   (a) a power amplifier.
   (b) a rectifier.
   (c) a filter.
   (d) a weak-signal amplifier.
7. The gate of a JFET exhibits a
   (a) forward bias.
   (b) high impedance.

(c) low reverse resistance.
(d) low avalanche voltage.

8. Which of the following conditions is not normally desirable in a JFET?
   (a) A conducting channel
   (b) Holes as the majority carriers
   (c) A forward-biased P-N junction
   (d) A high input impedance
9. When a JFET is pinched off,
   (a) the value of dI_D/dE_G is very large with no signal input.
   (b) the value of dI_D/dE_G might vary considerably with no signal input.
   (c) the value of dI_D/dE_G is negative with no signal input.
   (d) the value of dI_D/dE_G is zero with no signal input.
10. Transconductance is the ratio of
    (a) a change in drain voltage to a change in source voltage.
    (b) a change in drain current to a change in gate voltage.
    (c) a change in gate current to a change in source voltage.
    (d) a change in drain current to a change in drain voltage.
11. Characteristic curves for JFETs generally show
    (a) drain voltage as a function of source current.
    (b) drain current as a function of gate current.
    (c) drain current as a function of drain voltage.
    (d) drain voltage as a function of gate current.
12. A disadvantage of MOS components is the fact that
    (a) they can be easily damaged by static electricity.
    (b) they need a high input voltage in order to amplify.
    (c) they draw large amounts of current.
    (d) they produce a great deal of heat.
13. The input impedance of a MOSFET is
    (a) lower than that of a JFET.
    (b) lower than that of a bipolar transistor.
    (c) between that of a bipolar transistor and a JFET.
    (d) extremely high.
14. A significant difference between MOSFETs and JFETs is the fact that
    (a) MOSFETs can handle a wider range of gate bias voltages.
    (b) MOSFETs can deliver greater output power.
    (c) MOSFETs are more rugged.
    (d) MOSFETs last longer.

15. The channel in a zero-biased JFET is normally
    (a) pinched off.
    (b) in a state of avalanche breakdown.
    (c) in a conducting state.
    (d) made of P-type semiconductor material.

16. When an enhancement-mode MOSFET is at zero bias,
    (a) the drain current is high with no signal.
    (b) the drain current fluctuates with no signal.
    (c) the drain current is low with no signal.
    (d) the drain current is zero with no signal.

17. An enhancement-mode MOSFET can be recognized in schematic diagrams by the presence of
    (a) an arrow pointing inward.
    (b) a broken vertical line inside the circle.
    (c) an arrow pointing outward.
    (d) a solid vertical line inside the circle.

18. In a source follower, which of the electrodes receives the input signal?
    (a) Any of them (doesn't matter)
    (b) The source
    (c) The gate
    (d) The drain

19. Which of the following circuits produces an output signal wave that is 180° out of phase with the input signal wave?
    (a) The common source circuit
    (b) The common gate circuit
    (c) The common drain circuit
    (d) All of the above

20. Which of the following circuits can produce the greatest signal gain (amplification factor)?
    (a) The common source circuit
    (b) The common gate circuit
    (c) The common drain circuit
    (d) All of the above circuits can amplify to the same extent.

24
CHAPTER

Amplifiers and Oscillators

NOW THAT YOU'VE LEARNED THE BASICS OF BIPOLAR AND FIELD EFFECT TRANSISTORS, YOU'RE READY to learn about amplifiers and oscillators that use these devices. As a prelude to that, let's define the most common unit that engineers use to evaluate changes in signal strength.

The Decibel

The human senses of hearing and vision don't perceive the intensity of sound and light in a linear manner. Instead, our perception of these phenomena is *logarithmic.* Electronic circuits behave in a similar way when subjected to signals that change in amplitude. Because of this, scientists and engineers invented a unit called the *decibel* (symbolized dB), in which amplitude changes are expressed according to the *base-10 logarithm* of the actual change.

More or Less?

In the decibel scheme, increases in amplitude are assigned *positive gain* values, and decreases in amplitude are assigned *negative gain* values. A negative gain figure is sometimes expressed as a *loss* by removing the minus sign.

If the output signal amplitude from a circuit is +6 dB relative to the input signal amplitude, then the output is stronger than the input. If the output is at −14 dB relative to the input, then the output is weaker than the input. In the first case, the circuit has a gain of +6 dB. In the second case, we can say that the circuit has a gain of −14 dB or a loss of 14 dB.

How large a unit is the decibel, anyway? The answer is that it has no absolute size; it is a relative unit. The decibel has meaning only when two or more signals are compared. An amplitude increase or decrease of plus or minus one decibel (±1 dB) is roughly the smallest change a listener can detect if the change is expected. If the change is not expected, the smallest difference a listener can notice is about ±3 dB.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

For Voltage

Consider a circuit with an rms ac input voltage of E_{in} and an rms ac output voltage of E_{out}, specified in the same units as E_{in}. Then the *voltage gain* of the circuit, in decibels, is given by this formula:

$$\text{Gain (dB)} = 20 \log(E_{\text{out}}/E_{\text{in}})$$

The base-10 logarithm of a value x is written $\log x$. Log functions of base 10 and base e (another type of logarithm sometimes used in physics and engineering) can be determined easily using scientific calculators. From now on, when we say "logarithm," we mean the base-10 logarithm unless otherwise specified.

Problem 24-1

Suppose a circuit has an rms ac input of 1.00 V and an rms ac output of 14.0 V. What is the gain in decibels?

First, find the ratio $E_{\text{out}}/E_{\text{in}}$. Because $E_{\text{out}} = 14.0$ V rms and $E_{\text{in}} = 1.00$ V rms, the ratio is 14.0/1.00, or 14.0. Next, find the logarithm of 14.0. This is about 1.146128036. Finally, multiply this number by 20, getting something like 22.92256071. Round this off to three significant figures, because that's all you're entitled to, getting 22.9 dB.

Problem 24-2

Suppose a circuit has an rms ac input voltage of 24.2 V and an rms ac output voltage of 19.9 V. What is the gain in decibels?

Find the ratio $E_{\text{out}}/E_{\text{in}} = 19.9/24.2 = 0.822 \ldots$. (The three dots indicate extra digits introduced by the calculator. You can leave them in until the final roundoff.) Find the logarithm of this: $\log 0.822 \ldots = -0.0849 \ldots$. The gain is 20 times this, or $-1.699 \ldots$ dB, which rounds off to -1.70 dB.

For Current

Current gain or loss is calculated in the same way as voltage gain or loss. If I_{in} is the rms ac input current and I_{out} is the rms ac output current specified in the same units as I_{in}, then:

$$\text{Gain (dB)} = 20 \log(I_{\text{out}}/I_{\text{in}})$$

For Power

The *power gain* of a circuit, in decibels, is calculated according to a slightly different formula. If P_{in} is the input signal power and P_{out} is the output signal power expressed in the same units as P_{in}, then:

$$\text{Gain (dB)} = 10 \log (P_{\text{out}}/P_{\text{in}})$$

The *coefficient* (that is, the factor by which the logarithm is to be multiplied) in the formula for power gain is 10, whereas for voltage and current gain it is 20.

Problem 24-3

Suppose a power amplifier has an input of 5.03 W and an output of 125 W. What is the gain in decibels?

Find the ratio $P_{\text{out}}/P_{\text{in}} = 125/5.03 = 24.85\ldots$. Then find the logarithm: $\log 24.85\ldots = 1.395\ldots$. Finally, multiply by 10 and round off. The gain is thus $10 \times 1.395\ldots = 14.0$ dB.

Problem 24-4

Suppose an *attenuator* (a circuit designed deliberately to produce power loss) provides 10 dB power reduction. The input power is 94 W. What is the output power?

An attenuation of 10 dB represents a gain of −10 dB. We know that $P_{\text{in}} = 94$ W, so the unknown in the power gain formula is P_{out}. We must solve for P_{out} in this formula:

$$-10 = 10 \log (P_{\text{out}}/94)$$

First, divide each side by 10, getting:

$$-1 = \log (P_{\text{out}}/94)$$

To solve this, we must take the *base-10 antilogarithm*, also known as the *antilog*, or the *inverse log*, of each side. The antilog function undoes the log function. The antilog of a value x is written antilog x. It can also be denoted as $\log^{-1} x$ or 10^x. Antilogarithms can be determined with any good scientific calculator. The solution process goes like this:

$$\begin{aligned} \text{antilog}\,(-1) &= \text{antilog}\,[\log (P_{\text{out}}/94)] \\ 0.1 &= P_{\text{out}}/94 \\ 94 \times 0.1 &= P_{\text{out}} \\ P_{\text{out}} &= 9.4 \text{ W} \end{aligned}$$

Decibels and Impedance

When determining the voltage gain (or loss) and the current gain (or loss) for a circuit in decibels, you should expect to get the same figure for both parameters only when the input impedance is identical to the output impedance. If the input and output impedances differ, the voltage gain or loss is generally not the same as the current gain or loss.

Consider how transformers work. A step-up transformer, in theory, has voltage gain, but this alone doesn't make a signal more powerful. A step-down transformer can exhibit theoretical current gain, but again, this alone does not make a signal more powerful. In order to make a signal more *power*ful, a circuit must increase the signal *power*—the *product* of the voltage and the current!

When determining power gain (or loss) for a particular circuit in decibels, the input and output impedances don't matter. In this sense, positive power gain always represents a real-world increase in signal strength. Similarly, negative power gain (or power loss) always represents a true decrease in signal strength.

Basic Bipolar Transistor Amplifier

In the previous chapters, you saw some circuits that use bipolar and field effect transistors. A signal can be applied to some control point, causing a much greater signal to appear at the output. This is the principle by which all amplifiers work.

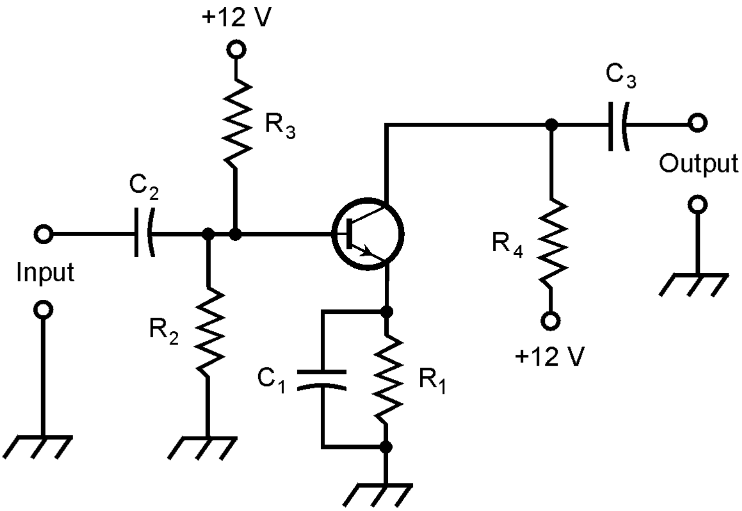

24-1 An amplifier circuit with a bipolar transistor. Component designators are discussed in the text.

In Fig. 24-1, an NPN bipolar transistor is connected as a *common emitter amplifier.* The input signal passes through C_2 to the base. Resistors R_2 and R_3 provide the base bias. Resistor R_1 and capacitor C_1 allow for the emitter to have a dc voltage relative to ground, while keeping it grounded for signals. Resistor R_1 also limits the current through the transistor. The ac output signal goes through capacitor C_3. Resistor R_4 keeps the ac output signal from being short-circuited through the power supply.

In this amplifier, the optimum capacitance values depend on the design frequency of the amplifier, and also on the impedances at the input and output. In general, as the frequency and/or circuit impedance increase, less capacitance is needed. At audio frequencies and low impedances, the capacitors might be as large as 100 µF. At radio frequencies and high impedances, values will be only a fraction of a microfarad, down to picofarads at the highest frequencies and impedances. The optimum resistor values also depend on the application. In the case of a weak-signal amplifier, typical values are 470 Ω for R_1, 4.7 kΩ for R_2, 10 kΩ for R_3, and 4.7 kΩ for R_4.

Basic JFET Amplifier

Figure 24-2 shows an N-channel JFET hooked up as a *common source amplifier.* The input signal passes through C_2 to the gate. Resistor R_2 provides the bias. Resistor R_1 and capacitor C_1 give the source a dc voltage relative to ground, while grounding it for signals. The output signal goes through C_3. Resistor R_3 keeps the output signal from being short-circuited through the power supply.

A JFET has a high input impedance, and therefore the value of C_2 should usually be small. If the device is a MOSFET, the input impedance is higher still, and C_2 will be smaller yet, sometimes 1 pF or less. The resistor values depend on the application. In some instances, R_1 and C_1 are not used, and the source is grounded directly. If R_1 is used, its optimum value will depend on the input

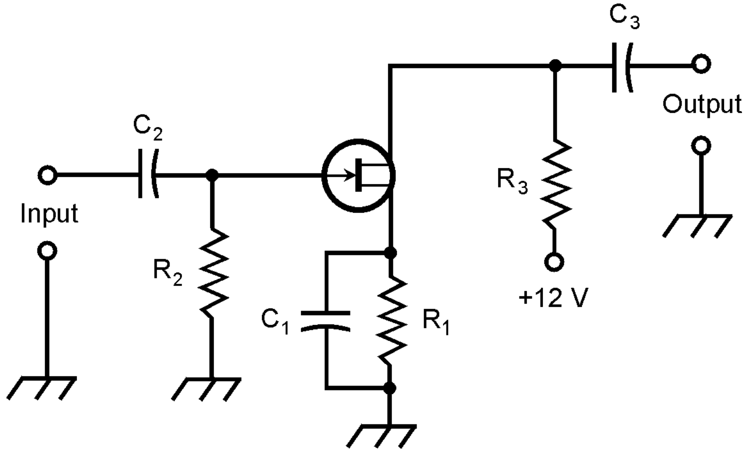

24-2 An amplifier circuit with an FET. Component designators are discussed in the text.

impedance and the bias needed for the FET. For a weak-signal amplifier, typical values are 680 Ω for R_1, 10 kΩ for R_2, and 100 Ω for R_3.

Amplifier Classes

Amplifier circuits can be categorized as *class A*, *class AB*, *class B*, and *class C*. Each class has its own special characteristics, and works best in its own unique set of applications.

The Class A Amplifier

With the previously mentioned component values, the amplifier circuits in Figs. 24-1 and 24-2 operate in class A. This type of amplifier is *linear*, meaning that the output waveform has the same shape as (although a much greater amplitude than) the input waveform.

For class A operation with a bipolar transistor, the bias must be such that, with no signal input, the device is near the middle of the straight-line portion of the I_C versus I_B (collector current versus base current) curve. This is shown for an NPN transistor in Fig. 24-3. For PNP, reverse the polarity signs. With a JFET or MOSFET, the bias must be such that, with no signal input, the device is near the middle of the straight-line part of the I_D versus E_G (drain current versus gate voltage) curve. This is shown in Fig. 24-4 for an N-channel JFET. For P channel, reverse the polarity signs.

In a class A amplifier, it is important that the input signal not be too strong. An excessively strong input signal will drive the device out of the straight-line part of the characteristic curve during part of the cycle. When this occurs, the output waveshape will not be a faithful reproduction of the input waveshape, and the amplifier will become *nonlinear*. Class A amplifiers are supposed to operate in a linear fashion at all times.

The Class AB Amplifier

Class A operation is inefficient because the transistor draws current whether there is a signal input or not. For weak-signal work, efficiency is not too critical; the things that matter are the gain and

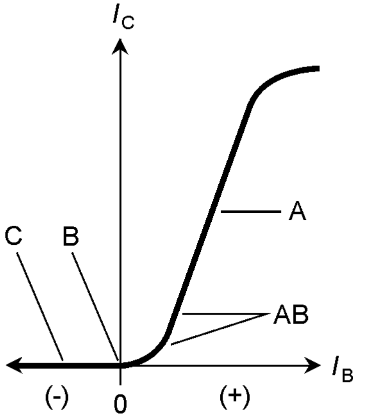

24-3 Various classes of amplifier operation for an NPN bipolar transistor.

the sensitivity. In power amplifiers, efficiency is more important, and gain and sensitivity don't matter as much.

When a bipolar transistor is biased close to cutoff under no-signal conditions, or when an FET is near pinchoff, the input signal always drives the device into the nonlinear part of the operating curve. Typical bias zones for class AB are shown in Figs. 24-3 and 24-4. A small collector or drain current flows when there is no input, but it is less than the no-signal current that flows in a class A amplifier. This is called *class AB operation.* It's more efficient than class A, but the gain and sensitivity are not as high.

There are two modes of class AB amplification. If the bipolar transistor or FET is never driven into cutoff/pinchoff during any part of the signal cycle, the amplifier is working in *class* AB_1. If the device goes into cutoff pinchoff for any part of the cycle (up to almost half), the amplifier is working in *class* AB_2.

In a class AB amplifier, the output signal waveform is not identical with the input signal waveform. But if the signal wave is *modulated*, such as in a voice radio transmitter, the waveform of the *modulating signal* comes out undistorted anyway. Thus, class AB operation is useful in RF power amplifiers.

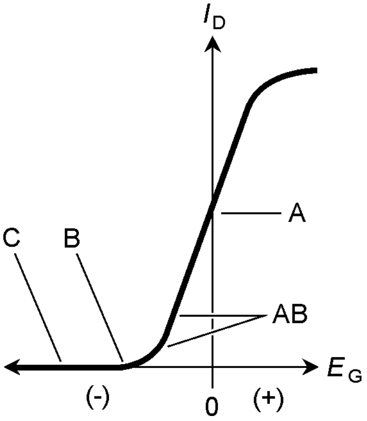

24-4 Various classes of amplifier operation for an N-channel JFET.

The Class B Amplifier

When a bipolar transistor is biased exactly at cutoff, or an FET is biased exactly at pinchoff under zero-input-signal conditions, an amplifier is working in *class B.* These operating points are labeled on the curves in Figs. 24-3 and 24-4. The class B scheme lends itself well to RF power amplification.

In class B operation, there is no collector or drain current when there is no signal. This saves energy, because the circuit does not consume power unless there is a signal going into it. (Class A and class AB amplifiers consume some power even when the input is zero.) When there is an input signal, current flows in the device during exactly half of the cycle. The output signal waveform is greatly different from the input waveshape in a class B amplifier. In fact, it is half-wave rectified.

You'll sometimes hear of class AB or class B "linear amplifiers," especially in ham radio. The term "linear" refers to the fact that the *modulation waveform* is not distorted by such an amplifier, even though the *carrier waveform* is distorted because the transistor is not biased in the straight-line part of the operating curve.

Class AB_2 and class B amplifiers draw power from the input signal source. Engineers say that such amplifiers require a certain amount of *drive* or *driving power* to function. Class A and class AB_1 amplifiers theoretically need no driving power, although there must be an input voltage.

The Class B Push-Pull Amplifier

Sometimes two bipolar transistors or FETs are used in a class B circuit, one for the positive half of the cycle and the other for the negative half. In this way, signal waveform distortion is eliminated. This is called a *class B push-pull amplifier.* This type of circuit, using two NPN bipolar transistors, is illustrated in Fig. 24-5. Resistor R_1 limits the current through the transistors. Capacitor C_1 keeps the input transformer center tap at signal ground, while allowing for some dc base bias. Resistors R_2 and R_3 bias the transistors precisely at their cutoff points. The two transistors must be identical. Not only should their part numbers be the same, but ideally they should be chosen *by experiment* to ensure that their characteristic curves are as closely matched as possible.

Class B push-pull is a popular arrangement for audio-frequency (AF) power amplification. It combines the efficiency of class B with the low distortion of class A. Its main disadvantage is that it needs two center-tapped transformers, one at the input and the other at the output. This makes push-pull amplifiers rather bulky and expensive compared to other types.

The Class C Amplifier

A bipolar transistor or FET can be biased past cutoff or pinchoff, and it will still work as a power amplifier (PA), provided that the drive is sufficient to overcome the bias during part of the cycle. This is known as *class C* operation. Bias points for class C are labeled in Figs. 24-3 and 24-4.

Class C amplifiers are nonlinear, even for amplitude modulation waveforms. Because of this, a class C circuit is useful only for signals that are either full-on or full-off. Such signals include old-fashioned Morse code, and digital schemes in which the frequency or phase (but not the amplitude) of the signal is varied.

A class C amplifier needs a lot of driving power. The gain is low. For example, it might take 300 W of signal drive to get 1 kW of signal power output. However, the efficiency is better than that of class A, AB, or B amplifiers. Let's take a closer look, now, at what *amplifier efficiency* is all about.

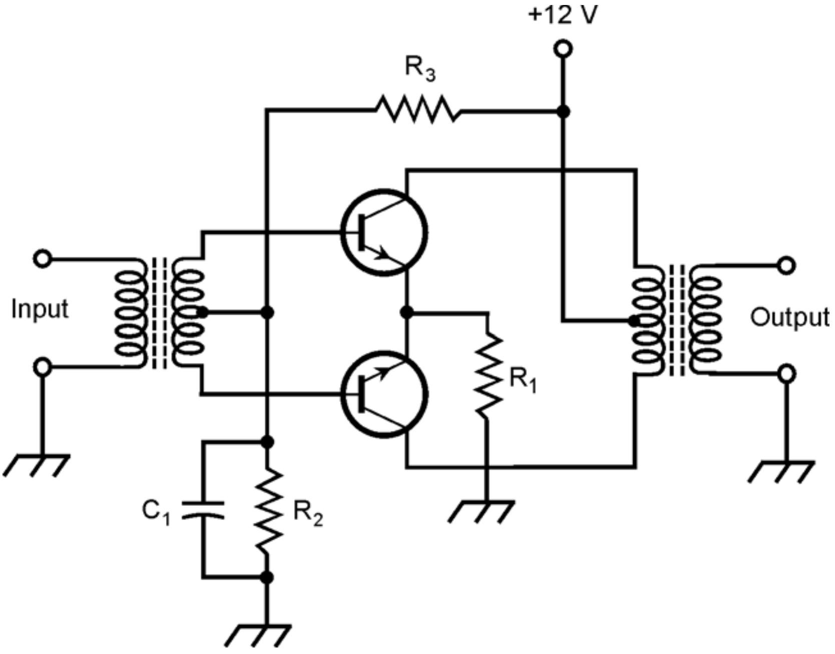

24-5 A class B push-pull amplifier using NPN bipolar transistors. Component designators are discussed in the text.

Efficiency in Power Amplifiers

In power amplification, efficiency is important. It not only provides optimum output power with minimum heat generation and minimum strain on the transistors, but it conserves energy as well. This translates into reduced cost, reduced size and weight, and longer equipment life compared with inefficient power amplifiers.

Determining dc Power Input

Suppose you connect an ammeter or milliammeter in series with the collector or drain of an amplifier and the power supply. While the amplifier is in operation, this meter will have a certain reading. The reading might appear constant, or it might fluctuate with changes in the input signal level. The dc *collector power input* to a bipolar-transistor amplifier circuit is the product of the collector current (I_C) and the collector voltage (E_C). Similarly, for an FET, the *dc drain power input* is the product of the drain current (I_D) and the drain voltage (E_D). These power figures can be further categorized as *average* or *peak* values. This discussion involves only average power.

The dc collector or drain power input can be high even when there is no input signal applied to an amplifier. A class A circuit operates this way. In fact, when a signal is applied to a class A amplifier, the meter reading, and therefore the dc collector or drain power input, does not change compared to the value under conditions of no input signal. In class AB_1 or class AB_2, there is low current (and therefore low dc collector or drain power input) with zero input signal, and a higher current (and therefore a higher dc power input) with an input signal. In class B and class C, there is no current (and therefore zero dc collector or drain power input) when there is no input signal. The current, and therefore the dc power input, increases with increasing signal input. The dc collector or drain power input is usually measured in watts, the product of amperes and volts. It can be indicated in milliwatts for low-power amplifiers, or kilowatts for high-power amplifiers.

Power Output

The *power output* of an amplifier must be measured by means of a specialized ac wattmeter. The design of AF and RF wattmeters is a sophisticated specialty in engineering.

When there is no signal input to an amplifier, there is no signal output, and therefore the power output is zero. This is true no matter what the class of amplification. The greater the signal input, in general, the greater the power output of a power amplifier, up to a certain point.

Power output, like dc input, is measured in watts. For very low power circuits, it can be in milliwatts; for high-power circuits, it is sometimes given in kilowatts.

Definition of Efficiency

The *efficiency* of a power amplifier is the ratio of the ac power output to the dc collector or drain power input.

In a bipolar-transistor amplifier, let P_C be the dc collector power input, and let P_{out} be the ac power output. For an FET amplifier, let P_D be the dc drain power input, and let P_{out} be the ac power output. Then the efficiency, eff, of the bipolar transistor amplifier is given by:

$$\text{eff} = P_{out}/P_C$$

For the FET circuit, the efficiency is:

$$\text{eff} = P_{out}/P_D$$

These are ratios, and they are always between 0 and 1. Efficiency is often expressed as a percentage instead of a ratio, so the preceding formulas are modified as follows:

$$\text{eff}_{\%} = 100\ P_{out}/P_C$$

and

$$\text{eff}_{\%} = 100\ P_{out}/P_D$$

Problem 24-5

Suppose a bipolar-transistor amplifier has a dc collector input of 115 W and an ac power output of 65.0 W. What is the efficiency in percent?

Use the formula for the efficiency of a bipolar transistor amplifier expressed as a percentage: $\text{eff}_{\%} = 100\ P_{out}/P_C = 100 \times 65/115 = 100 \times 0.565 = 56.5\%$.

Problem 24-6

Suppose an FET amplifier is 60 percent efficient. If the power output is 3.5 W, what is the dc drain power input?

Plug in values to the formula for the efficiency of an FET amplifier expressed as a percentage. The resulting equation is solved as follows:

$$
\begin{aligned}
60 &= 100 \times 3.5/P_D \\
60 &= 350/P_D \\
60/350 &= 1/P_D \\
P_D &= 350/60 \\
&= 5.8 \text{ W}
\end{aligned}
$$

Efficiency versus Class

Class A amplifiers have efficiency figures from 25 percent to 40 percent, depending on the nature of the input signal and the type of transistor used. A good class AB_1 amplifier is 35 percent to 45 percent efficient. A class AB_2 amplifier, if well designed and properly operated, can be up to about 50 percent efficient. Class B amplifiers are typically 50 percent to 65 percent efficient. Class C amplifiers can have efficiency levels as high as 75 percent.

Drive and Overdrive

Class A and AB_1 power amplifiers do not, in theory, take any power from the signal source to produce significant output power. This is one of the advantages of these classes of operation. It is only necessary that a certain voltage be present at the control electrode (the base, gate, emitter, or source) for these circuits to produce useful output signal power. Class AB_2 amplifiers need some driving power to produce ac power output. Class B amplifiers require more drive than class AB_2, and class C amplifiers need still more drive.

Whatever kind of PA is used in a given situation, it is important that the driving signal not be too strong. If *overdrive* takes place, distortion occurs in the output signal. An oscilloscope can be used to determine whether or not an amplifier is being overdriven. The scope is connected to the amplifier output terminals, and the waveform of the output signal is examined. The output waveform for a particular class of amplifier always has a characteristic shape. Overdrive is indicated by a form of distortion known as *flat topping.*

In Fig. 24-6A, the output signal waveshape for a properly operating class B amplifier is shown. In Fig. 24-6B, the output of an overdriven class B amplifier is shown. Note that the peaks are blunted or truncated. The result of this can be distortion in the modulation on a radio signal, and

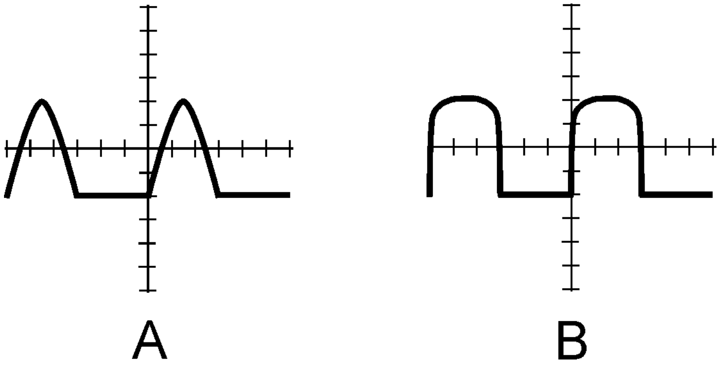

24-6 At A, an oscilloscope display of the signal output waveform from a properly operating class B power amplifier. At B, a display showing distortion in the waveform caused by overdrive.

also an excessive amount of signal output at harmonic frequencies. The efficiency of the circuit can be degraded, as well. The flat tops of the distorted waves don't contribute anything to the strength of the signal at the desired frequency, but they cause a higher-than-normal dc power input, which translates into a lower-than-normal efficiency.

Audio Amplification

The circuits you've seen so far have been general, not application-specific. With capacitors of several microfarads, and when biased for class A, these circuits are representative of audio amplifiers.

Frequency Response

High-fidelity audio amplifiers, of the kind used in music systems, must have more or less constant gain from 20 Hz to 20 kHz. This is a frequency range of 1000:1. Audio amplifiers for voice communications must work from 300 Hz to 3 kHz, a 10:1 span of frequencies. In digital communications, audio amplifiers are designed to work over a narrow range of frequencies, sometimes less than 100 Hz wide.

Hi-fi amplifiers are usually equipped with resistor-capacitor (RC) networks that tailor the frequency response. These are *tone controls*, also called *bass* and *treble* controls. The simplest hi-fi amplifiers use a single knob to control the tone. More sophisticated amplifiers have separate controls, one for bass and the other for treble. The most advanced hi-fi systems make use of *graphic equalizers*, having controls that affect the amplifier gain over several different frequency spans.

Volume Control

Audio amplifier systems usually consist of two or more *stages.* A stage is one bipolar transistor or FET (or a push-pull combination), plus peripheral resistors and capacitors. Stages can be cascaded one after the other to get high gain. In one of the stages in an audio system, a *volume control* is used. This control can be as simple as a potentiometer that allows the gain of a stage to be adjusted without affecting its linearity.

An example of a basic volume control is shown in Fig. 24-7. In this amplifier, the gain through the transistor is constant. The ac output signal passes through C_1 and appears across R_1, a potentiometer. The *wiper* (indicated by the arrow) of the potentiometer picks off more or less of the ac output signal, depending on the position of the control shaft. Capacitor C_2 isolates the potentiometer from the dc bias of the following stage.

A volume control should normally be placed in a stage where the audio power level is low. This allows the use of a low-wattage, low-cost potentiometer.

Transformer Coupling

Transformers can be used to transfer (or *couple*) signals from one stage to the next in a cascaded amplifier system (also known as an *amplifier chain*). An example of *transformer coupling* is shown in Fig. 24-8. Capacitors C_1 and C_2 keep one end of the transformer primary and secondary at signal ground. Resistor R_1 limits the current through the first transistor, Q_1. Resistors R_2 and R_3 provide the proper base bias for transistor Q_2.

The main disadvantage of this scheme is that it costs more than *capacitive coupling.* But transformer coupling can provide an optimum signal transfer between amplifier stages. By selecting a transformer with the correct turns ratio, the output impedance of Q_1 can be perfectly matched to the input impedance of Q_2.

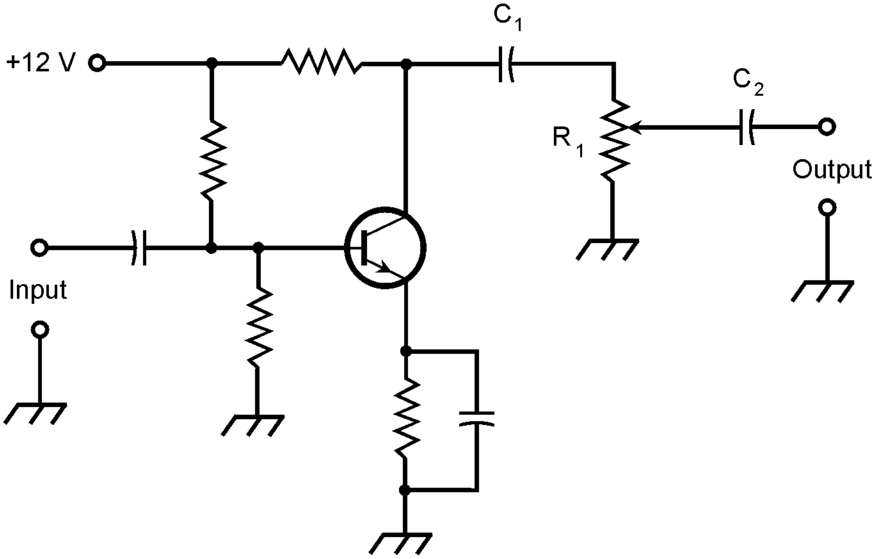

24-7 A basic volume control (potentiometer R_1) can be used to vary the gain in a low-power audio amplifier.

Tuned-Circuit Coupling

In some amplifier systems that employ transformer coupling, capacitors are added across the primary and/or secondary of the transformer. This results in resonance at a frequency determined by the capacitance and the transformer winding inductance. If the set of amplifiers is intended for use at only one frequency (and this is often the case in RF systems), this method of coupling, called

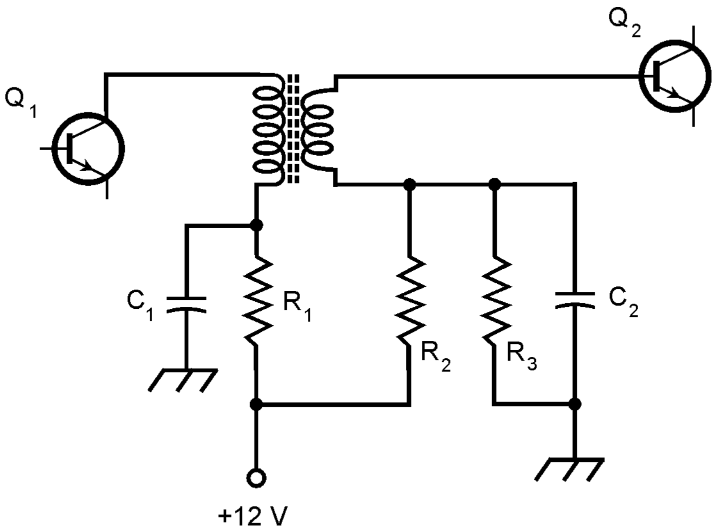

24-8 An example of transformer coupling between amplifier stages. Component designators are discussed in the text.

tuned-circuit coupling, enhances the system efficiency. But care must be taken to be sure that the amplifier chain doesn't oscillate at the resonant frequency of the tuned circuits!

Radio-Frequency Amplification

The *RF spectrum* extends upward in frequency to well over 300 GHz. The exact lower limit is a matter of disagreement in the literature. Some texts put it at 3 kHz, some at 9 kHz, some at 10 kHz, and some at the upper end of the AF range, which is normally considered to be 20 kHz.

Weak-Signal Amplifiers versus Power Amplifiers

Some RF amplifiers are designed for weak-signal work. The *front end*, or first amplifying stage, of a radio receiver requires the most sensitive possible amplifier. Sensitivity is determined by two factors: the *gain*, which has already been discussed here, and the *noise figure*, a measure of how well a circuit can amplify desired signals while generating a minimum of electronic noise.

All bipolar transistors or FETs create some *white noise* because of the movement of the charge carriers among the atoms. In general, JFETs produce less noise than bipolar transistors. Gallium arsenide FETs, also called *GaAsFETs* (pronounced "gasfets"), are the least noisy of all.

The higher the frequency at which a weak-signal amplifier is designed, the more important the noise figure gets. This is because there is less *atmospheric noise* at the higher radio frequencies, as compared with the lower frequencies. At 1.8 MHz, for example, the airwaves contain much atmospheric noise, and it doesn't make any difference if the receiver introduces a little noise itself. But at 1.8 GHz, atmospheric noise is almost nonexistent, and receiver performance depends critically on the amount of internally generated noise.

Weak-signal amplifiers almost always use resonant circuits. This optimizes the amplification at the desired frequency, while helping to cut out noise on unwanted frequencies. A typical tuned GaAsFET weak-signal RF amplifier is diagrammed in Fig. 24-9. It is designed for operation at about 10 MHz.

Broadband PAs

At RF, a PA can be either *broadband* or *tuned.* The main advantage of a broadband PA is ease of operation, because it does not need tuning. The operator need not worry about critical adjustments, nor bother to change them when changing the frequency. However, broadband PAs are slightly less efficient than tuned PAs. Another disadvantage of broadband PAs is the fact that they will amplify any signal in the design frequency range, whether or not this is desired. For example, if some earlier stage in a radio transmitter is oscillating at a frequency different from the intended signal frequency, and if this undesired energy falls within the design frequency range of the broadband PA, it will be amplified. The result will be unintended RF emission from the radio transmitter. Such unwanted signals are called *spurious emissions.*

Figure 24-10 is a schematic diagram of a typical broadband PA. The NPN bipolar transistor is a power transistor. It will reliably provide several watts of continuous RF power output over a range of frequencies from 1.5 MHz through 15 MHz. The transformers are a critical part of this circuit. They must be designed to work efficiently over a 10:1 range of frequencies.

Tuned PAs

A tuned RF power amplifier offers improved efficiency compared with broadband designs. Also, the tuning helps to reduce the chances of spurious signals being amplified and transmitted over the air. Another advantage of tuned PAs is that they can work into a wide range of load impedances. In ad-

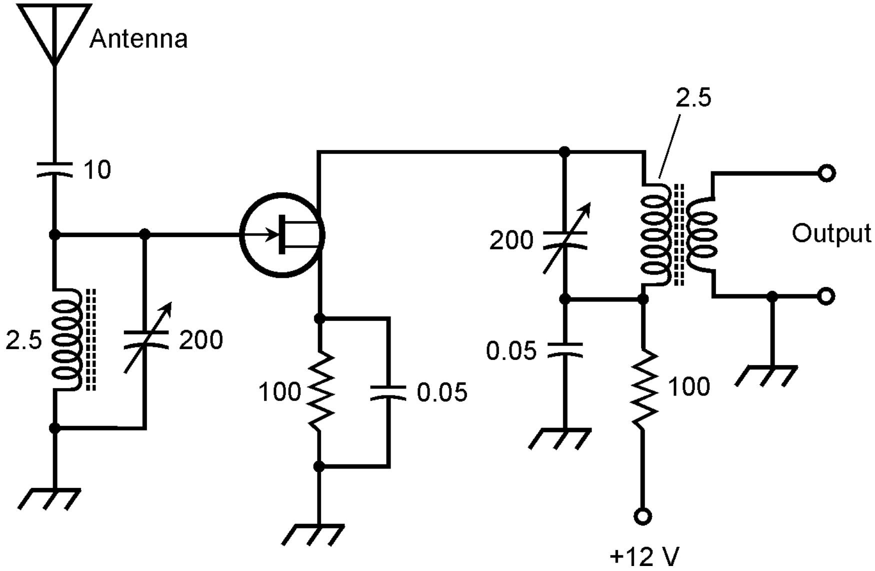
Antenna
10
2.5
200
100
0.05
200
0.05
2.5
Output
100
+12 V

24-9 A tuned RF amplifier for use at about 10 MHz. Resistances are in ohms. Capacitances are in microfarads (μF) if less than 1, and in picofarads (pF) if more than 1. Inductances are in microhenrys (μH).

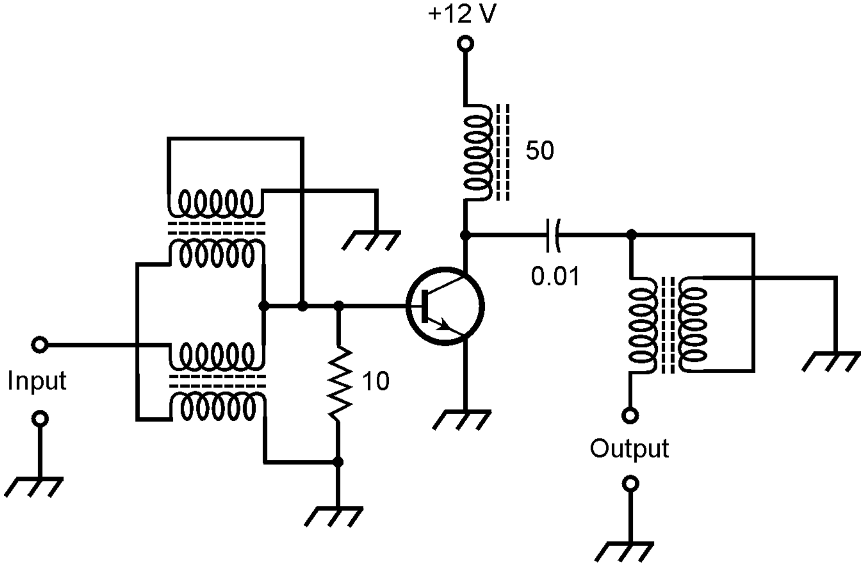
+12 V
50
0.01
Input
10
Output

24-10 A broadband RF power amplifier, capable of producing a few watts output. Resistances are in ohms. Capacitances are in microfarads (μF). Inductances are in microhenrys (μH).

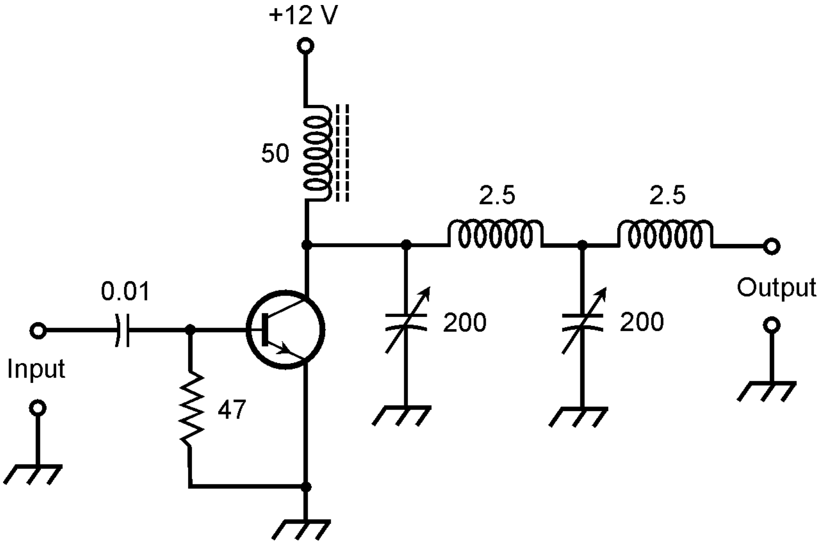

24-11 A tuned RF power amplifier, capable of producing a few watts output. Resistances are in ohms. Capacitances are in microfarads (μF) if less than 1, and in picofarads (pF) if more than 1. Inductances are in microhenrys (μH).

dition to a *tuning control*, or resonant circuit that adjusts the output of the amplifier to the operating frequency, there is a *loading control* that optimizes the signal transfer between the amplifier and the load (usually an antenna).

The main drawback of a tuned PA is that the adjustment takes time, and improper adjustment can result in damage to the transistor. If the tuning and/or loading controls are out of kilter, the efficiency of the amplifier will be extremely low (sometimes practically zero) while the dc collector or drain power input is high. Solid-state devices overheat quickly under these conditions.

A tuned RF PA, providing a few watts' output at 10 MHz or so, is shown in Fig. 24-11. The transistor is the same type as for the broadband amplifier discussed previously. The tuning and loading controls (left-hand and right-hand variable capacitors, respectively) should be adjusted for maximum RF power output as indicated by an RF wattmeter.

How Oscillators Work

Once you know how amplifiers work, it's easy to understand oscillators. All oscillators are amplifiers with positive feedback. In radio communications, oscillators generate the waves, or signals, that are ultimately sent over the air. Audio-frequency oscillators find applications in such devices as music synthesizers, modems, doorbells, sirens, alarms, and electronic toys.

Positive Feedback

Feedback can be in phase or out of phase. For a circuit to oscillate, the feedback must be in phase (positive). Out-of-phase (negative) feedback reduces the gain of an amplifier. In fact, negative feedback is used in some amplifiers to *prevent* oscillation.

The output of a common emitter or common source amplifier is out of phase from the input. If you couple the collector to the base through a capacitor, you won't get oscillation. It is necessary to reverse the phase in the feedback process in order for oscillation to occur. In addition, the amplifier gain must be high, and the coupling from the output to the input must be good. The positive feedback path must be easy for a signal to follow. Most oscillators are common emitter or common source amplifier circuits with positive feedback.

The output of a common base or common gate amplifier is in phase with the input. But these circuits have limited gain, and it's hard to make them oscillate. Common collector and common drain circuits don't have enough gain to make oscillators.

Feedback at a Single Frequency

The frequency of an oscillator is controlled by means of tuned, or resonant, circuits. These are usually inductance-capacitance (*LC*) or resistance-capacitance (*RC*) combinations. The *LC* scheme is common in radio transmitters and receivers; the *RC* method is more often used in audio work. The tuned circuit makes the feedback path easy for a signal to follow at one frequency, but hard to follow at all other frequencies. As a result, oscillation takes place at a predictable and stable frequency, determined by the inductance and capacitance, or by the resistance and capacitance.

Common Oscillator Circuits

Many circuit arrangements can be used to produce oscillation. The following several circuits are all known as *variable-frequency oscillators* (VFOs), because their frequencies are adjustable over a wide range.

The Armstrong Circuit

A common emitter or common source class A amplifier can be made to oscillate by coupling the output back to the input through a transformer that reverses the phase of the fed-back signal. The schematic diagram of Fig. 24-12 shows a common-source amplifier whose drain circuit is coupled to the gate circuit by means of a transformer. The frequency is controlled by a capacitor in series with the secondary winding. The inductance of the secondary, along with the capacitance, forms a resonant circuit that passes energy easily at one frequency while attenuating the energy at other frequencies. This circuit is known as an *Armstrong oscillator.* A bipolar transistor can be used in place of the JFET, as long as the device is biased for class A amplification.

The Hartley Circuit

Another method of obtaining controlled feedback at RF is shown in Fig. 24-13. In this circuit, a PNP bipolar transistor is used. The circuit uses a single coil with a tap on the windings to provide the feedback. A variable capacitor in parallel with the coil determines the oscillating frequency, and allows for frequency adjustment. This circuit is called a *Hartley oscillator.*

In the Hartley circuit, as well as in most other RF oscillator circuits, it is important to use only the minimum amount of feedback necessary to get oscillation. The amount of feedback is controlled by the position of the coil tap. The circuit shown in Fig. 24-13 uses about 25 percent of its amplifier power to produce feedback. The other 75 percent of the power can be used as output.

Oscillators usually produce less than 1 W of RF power output. If more power is needed, the signal can be boosted by one or more stages of amplification.

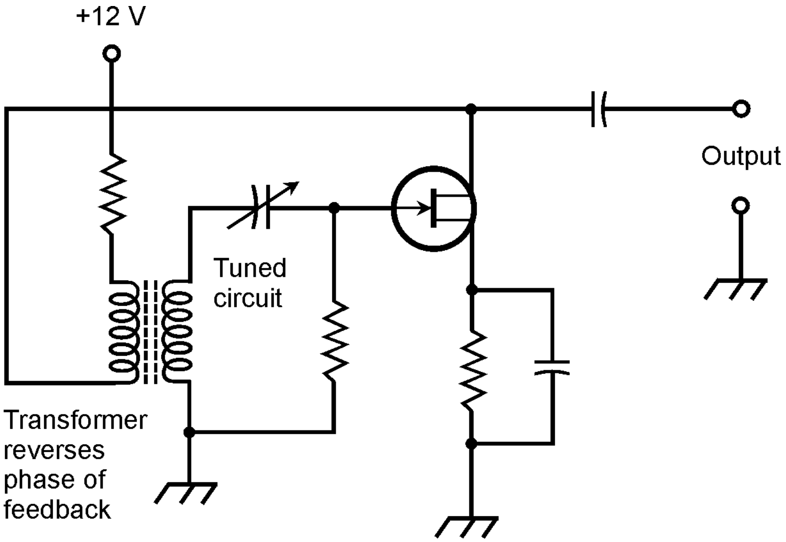

24-12 An Armstrong oscillator using an N-channel JFET. This is a common-source amplifier with positive feedback through a tuned circuit.

The Colpitts Circuit

The capacitance can be tapped, instead of the inductance, in the tuned circuit of an RF oscillator. Such a circuit is called a *Colpitts oscillator*, and a P-channel JFET version is diagrammed in Fig. 24-14. The amount of feedback is controlled by the ratio of the capacitances. A variable inductor provides for frequency adjustment. This is a matter of convenience, because it can be difficult to find a dual

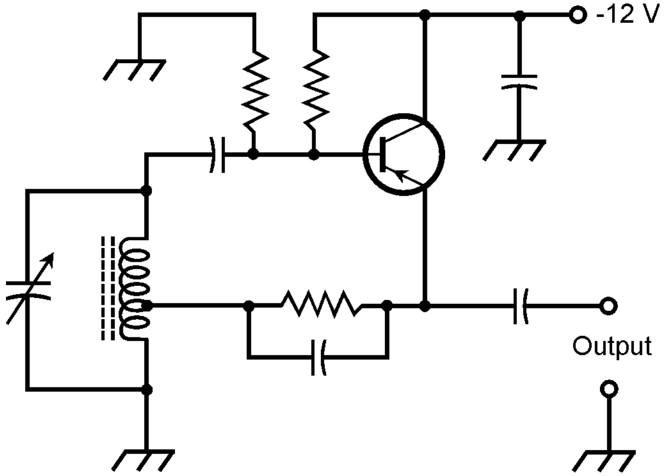

24-13 A Hartley oscillator using a PNP bipolar transistor. The Hartley circuit can be recognized by the tapped inductor in the tuned *LC* circuit.

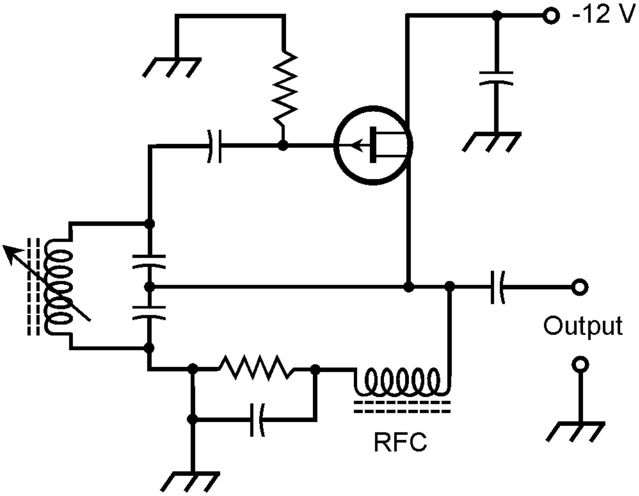

24-14 A Colpitts oscillator using a P-channel JFET. The Colpitts circuit can be recognized by the split capacitance in the tuned *LC* circuit.

variable capacitor that maintains the correct ratio of capacitances throughout its tuning range. Using fixed capacitors eliminates this problem, and it costs less, too!

Unfortunately, finding a good variable inductor for use in a Colpitts oscillator can be just about as hard as obtaining a suitable dual-gang variable capacitor. A *permeability-tuned* coil can be used, but ferromagnetic cores impair the frequency stability of an RF oscillator. A *roller inductor* can be employed, but these are bulky and expensive. An inductor with several switch-selectable taps can be used, but this does not allow for continuous frequency adjustment. Despite these shortcomings, the Colpitts circuit offers exceptional stability and reliability when properly designed, and is preferred by some engineers for this reason.

The Clapp Circuit

A variation of the Colpitts oscillator makes use of series resonance, instead of parallel resonance, in the tuned circuit. Otherwise, the circuit is basically the same as the parallel-tuned Colpitts oscillator. Figure 24-15 is a schematic diagram of a *series-tuned Colpitts oscillator* circuit that uses an NPN bipolar transistor. This circuit is also known as a *Clapp oscillator.* Its frequency won't change much when high-quality components are used. The Clapp oscillator is a reliable circuit. It isn't hard to get it to oscillate and keep it going. Another advantage of the Clapp circuit is that it allows the use of a variable capacitor for frequency control, while accomplishing feedback through a capacitive voltage divider.

Getting the Output

In the Hartley, Colpitts, and Clapp oscillators just described and shown in Figs. 24-13 through 24-15, the output is taken from the emitter or source, not from the collector or drain. There's a reason for this. The output of an oscillator can be taken from the collector or drain, just as is done in a common emitter or common-source amplifier to get maximum gain. But in an oscillator, stabil-

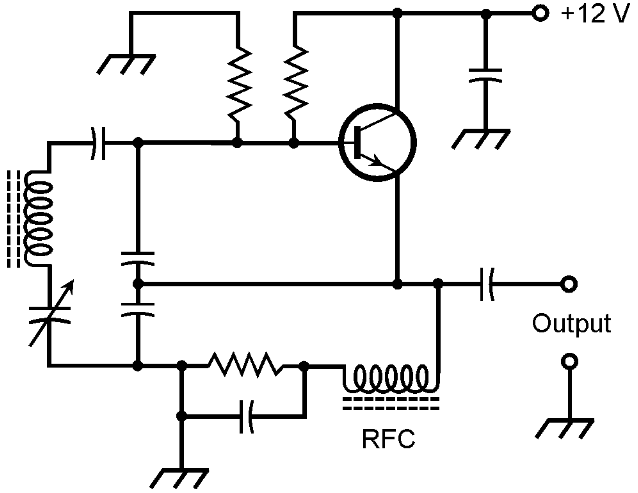

24-15 A series-tuned Colpitts oscillator, also known as a Clapp oscillator. This circuit uses an NPN bipolar transistor.

ity is more important than gain. The stability of an oscillator is better when the output is taken from the emitter or source, as compared with taking it from the collector or drain. Variations in the load impedance are less likely to affect the frequency of oscillation, and a sudden decrease in load impedance is less likely to cause the circuit to stop oscillating.

To prevent the output signal from being short-circuited to ground, an *RF choke* (RFC) is connected in series with the emitter or source in the Colpitts and Clapp oscillator circuits. The choke lets dc pass while blocking ac (just the opposite of a *blocking capacitor*). Typical values for RF chokes range from about 100 μH at high frequencies, such as 15 MHz, to 10 mH at low frequencies, such as 150 kHz.

The Voltage-Controlled Oscillator

The frequency of a VFO can be adjusted by means of a varactor diode in the tuned *LC* circuit. Recall that a varactor, also called a *varicap*, is a semiconductor diode that works as a variable capacitor when it is reverse-biased. The capacitance depends on the reverse-bias voltage. The greater this voltage, the lower the value of the capacitance.

The Hartley and Clapp oscillator circuits lend themselves well to varactor-diode frequency control. The varactor is placed in series or parallel with the tuning capacitor, and is isolated for dc by blocking capacitors. In Chap. 20, we saw an example of how a varactor can be connected in a tuned circuit (Fig. 20-9). The resulting oscillator is called a *voltage-controlled oscillator* (VCO).

Varactors are cheaper than variable capacitors or inductors. They're also less bulky. These are the chief advantages of a VCO over an old-fashioned *LC* tuned VFO.

Diode Oscillators

At ultrahigh frequencies (UHF) and microwave radio frequencies, certain types of diodes can be used as oscillators. You learned about these diodes in Chap. 20.

Crystal-Controlled Oscillators

Quartz crystals can be used in place of tuned *LC* circuits in RF oscillators, as long as it isn't necessary to change the frequency often. Crystal oscillators offer frequency stability far superior to that of *LC* tuned VFOs.

There are several ways that crystals can be connected in bipolar or FET circuits to get oscillation. One common circuit is the *Pierce oscillator.* An N-channel JFET and quartz crystal are connected in a Pierce configuration as shown in the schematic diagram of Fig. 24-16. The crystal frequency can be varied somewhat (by about 0.1 percent, or 1 part in 1000) by means of an inductor or capacitor in parallel with the crystal. But the frequency is determined mainly by the thickness of the quartz wafer, and by the angle at which it is cut from the original quartz sample.

Crystals change in frequency as the temperature changes. But they are far more stable than *LC* circuits, most of the time. Some crystal oscillators are housed in temperature-controlled chambers called *crystal ovens.* In this environment, crystals maintain their frequency so well that they are sometimes used as *frequency standards* against which other oscillators are calibrated.

The Phase-Locked Loop

One type of oscillator that combines the flexibility of a VFO with the stability of a crystal oscillator is known as a *phase-locked loop* (PLL). This makes use of a circuit called a *frequency synthesizer.*

The heart of the PLL is a VCO. The output of this oscillator passes through a *programmable multiplier/divider*, a digital circuit that divides and/or multiplies the VCO frequency by integral (whole-number) values chosen by the operator. As a result, the output frequency can be any rational-number multiple of the crystal frequency. A well-designed PLL circuit can be tuned in small digital increments over a wide range of frequencies.

The output frequency of the multiplier/divider is locked, by means of a *phase comparator*, to the signal from a crystal-controlled *reference oscillator.* As long as the output from the multiplier/divider is exactly on the reference oscillator frequency, the two signals are in phase, and the output of the phase comparator is 0 V dc. If the VCO frequency begins to drift, the output frequency of the multiplier/divider will drift, too (although at a different rate). But even a frequency change of less than 1 Hz causes the phase comparator to produce a dc *error voltage.* This error voltage is either pos-

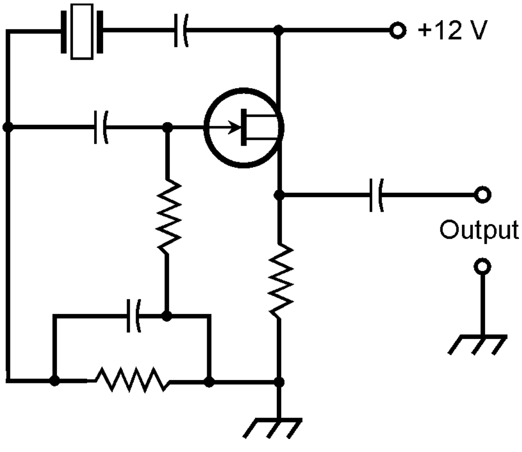

24-16 A Pierce oscillator circuit using an N-channel JFET.

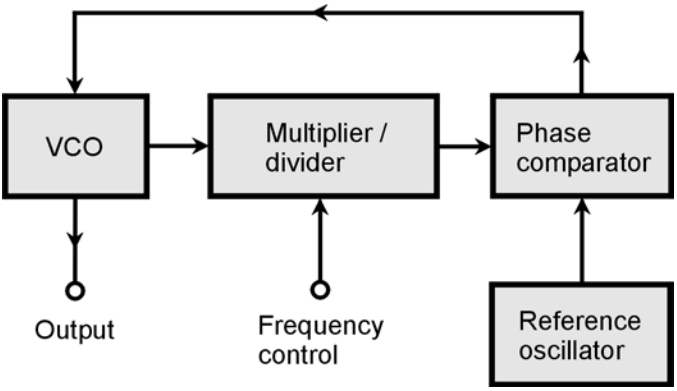

24-17 Block diagram of a phase-locked loop (PLL).

itive or negative, depending on whether the VCO has drifted higher or lower in frequency. The error voltage is applied to a varactor, causing the VCO frequency to change in a direction opposite to that of the drift. This forms a *dc feedback* circuit that maintains the VCO frequency at a precise value. It is a *loop* circuit that *locks* the VCO onto a particular frequency by means of *phase* sensing, hence the term *phase-locked loop* (PLL).

The key to the stability of the PLL lies in the fact that the reference oscillator is crystal-controlled. A block diagram of a PLL circuit is shown in Fig. 24-17. When you hear that a radio receiver, transmitter, or transceiver is synthesized, it usually means that the frequency is determined by a PLL.

The stability of a synthesizer can be enhanced by using an amplified signal from the shortwave time-and-frequency broadcast station WWV at 2.5, 5, 10, or 15 MHz, directly as the reference oscillator. These signals are frequency-exact to a minuscule fraction of 1 Hz, because they are controlled by atomic clocks. Most people don't need precision of this caliber, so you won't see consumer devices like ham radios and shortwave receivers with *primary-standard* PLL frequency synthesis. But it is employed by some corporations and government agencies.

Oscillator Stability

In an oscillator, the term *stability* has two meanings: constancy of frequency (or minimal *frequency drift*), and reliability of performance.

Constancy of Frequency

When designing a VFO of any kind, it's essential that the components maintain constant values, as much as possible, under all anticipated conditions. Some types of capacitors maintain their values better than others as the temperature rises or falls. Among the best are polystyrene capacitors. Silver-mica capacitors also work well when polystyrene units can't be found. Inductors are most temperature-stable when they have air cores. They should be wound, when possible, from stiff wire with strips of plastic to keep the windings in place. Some air-core coils are wound on hollow cylindrical cores, made of ceramic or phenolic material. Ferromagnetic solenoidal or toroidal cores aren't very good for VFO coils, because these materials change their permeability as the temperature varies. This changes the inductance, in turn affecting the oscillator frequency.

The best oscillators, in terms of frequency stability, are crystal-controlled. This includes circuits that oscillate at the fundamental frequency of the quartz crystal, circuits that oscillate at one of the crystal harmonic frequencies, or circuits that oscillate at frequencies derived from the crystal frequency by multiplier/dividers.

Reliability

An oscillator should always start working as soon as power is supplied. It should keep oscillating under all normal conditions. The failure of a single oscillator can cause an entire receiver, transmitter, or transceiver to stop working.

When an oscillator is built and put to use in a radio receiver, transmitter, or audio device, *debugging* is always necessary. This is a trial-and-error process of getting the flaws, or bugs, out of the circuit. Rarely can an engineer build something straight from the drawing board and have it work just right the first time. In fact, if two oscillators are built from the same diagram, with the same component types and values in the same geometric arrangement, one circuit might work fine, and the other might not. This usually happens because of differences in the quality of components that don't show up until the acid test.

Oscillators are designed to work into a certain range of load impedances. It's important that the load impedance not be too low. (You need never be concerned that it might be too high. In general, the higher the load impedance, the better.) If the load impedance is too low, the load will draw significant power from an oscillator. Then, even a well-designed oscillator might become unstable. Oscillators aren't meant to produce powerful signals. High power can be obtained using amplification after the oscillator.

Audio Oscillators

Audio oscillators are used in myriad electronic devices including doorbells, ambulance sirens, electronic games, telephone sets, and toys that play musical tunes. All AF oscillators are, in effect, AF amplifiers with positive feedback.

Audio Waveforms

At AF, oscillators can use *RC* or *LC* combinations to determine frequency. If *LC* circuits are used, the inductances must be large, and ferromagnetic cores are necessary.

At RF, oscillators are usually designed to produce a sine wave output. A pure sine wave represents energy at one and only one frequency. Audio oscillators, by contrast, don't always concentrate all their energy at a single frequency. (A pure AF sine wave, especially if it is continuous and frequency-constant, can be annoying.) The various musical instruments in a band or orchestra all sound different from each other, even when they play the same note (such as middle C). The reason for this is that each instrument has its own unique waveform. A clarinet sounds different than a trumpet, which in turn sounds different than a cello or piano.

Suppose you were to use an oscilloscope to look at the waveforms of musical instruments. This can be done using a high-fidelity microphone, a sensitive, low-distortion audio amplifier, and an oscilloscope. You'd see that each instrument has its own signature. Thus, each instrument's unique sound qualities can be reproduced using AF oscillators whose waveform outputs match those of the instrument. Electronic music synthesizers use audio oscillators to generate the tones you hear.

The Twin T Oscillator

An audio oscillator circuit that is popular for general-purpose use is the *twin T oscillator* (Fig. 24-18). The frequency is determined by the values of the resistors R and capacitors C. The output is a near-perfect sine wave. The small amount of distortion helps to alleviate the irritation produced by an absolutely pure sinusoid. The circuit shown in this example uses two PNP bipolar transistors. They are biased for class A amplification.

The Multivibrator

Another popular AF oscillator circuit makes use of two identical common emitter or common source amplifier circuits, hooked up so that the signal goes around and around between them. This is sometimes called a *multivibrator* circuit, although that is technically a misnomer, the term being more appropriate to various digital signal-generating circuits.

In the example of Fig. 24-19, two N-channel JFETs are connected to form a multivibrator for use at AF. Each stage amplifies the signal in class A, and reverses the phase by 180°. Therefore, the signal goes through a 360° phase shift each time it gets back to any particular point. A 360° phase shift is equivalent to no phase shift at all, so it results in positive feedback.

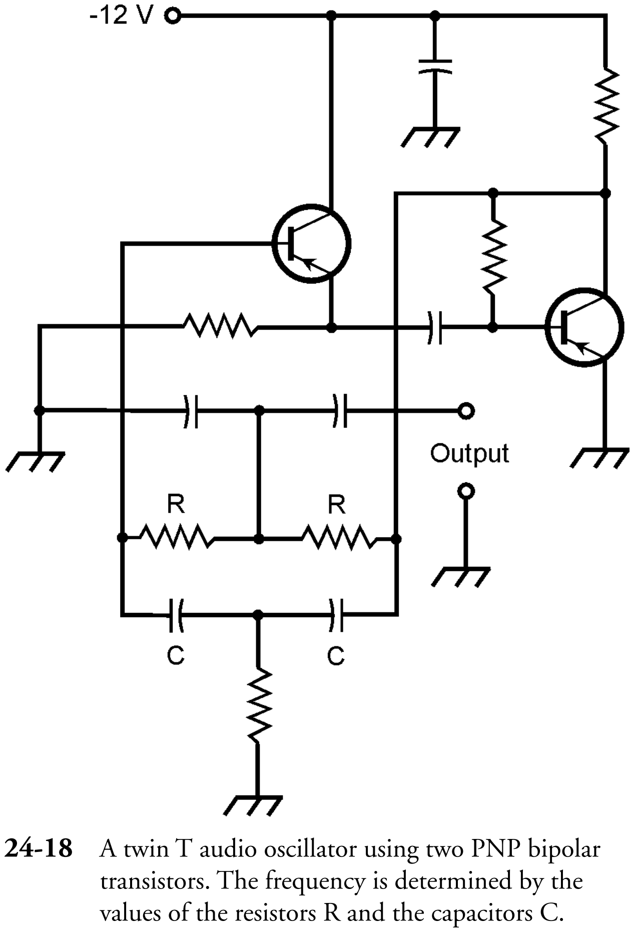

24-18 A twin T audio oscillator using two PNP bipolar transistors. The frequency is determined by the values of the resistors R and the capacitors C.

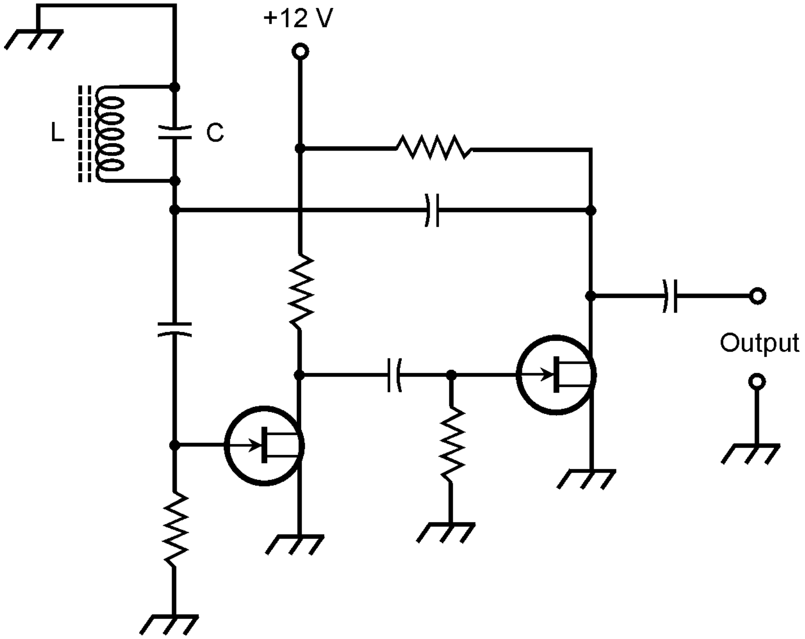

24-19 A multivibrator audio oscillator using two N-channel JFETs. The frequency is determined by the value of the inductor L and the capacitor C.

The frequency of the circuit shown in Fig. 24-19 is set by means of an *LC* circuit. The coil uses a ferromagnetic core, because stability is not of great concern and because such a core is necessary to obtain the large inductance needed for resonance at AF. Toroidal cores or pot cores are excellent in this application. The value of *L* can range from about 10 mH to as much as 1 H. The capacitance is chosen according to the formula for resonant circuits, to obtain an audio tone at the frequency desired.

IC Oscillators

In recent years, solid-state technology has advanced to the point that whole circuits can be etched onto silicon chips. Such devices are called *integrated circuits* (ICs). The *operational amplifier*, also called an *op amp*, is a type of IC that is especially useful as an audio oscillator because it has high gain, and it can easily be connected to produce positive feedback.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. The decibel is a unit of
   (a) relative signal strength.
   (b) voltage.

(c) power.
(d) current.

2. An oscillator at RF requires the use of
   (a) a common drain or common collector circuit.
   (b) a stage with gain.
   (c) a tapped coil.
   (d) a quartz crystal.
3. Suppose a circuit is found to have a gain figure of −15 dB. Which of the following statements is true?
   (a) The output signal is stronger than the input signal.
   (b) The input signal is stronger than the output signal.
   (c) The input signal is 15 times as strong as the output signal.
   (d) The output signal is 15 times as strong as the input signal.
4. In an oscillator circuit, the feedback should be
   (a) as great as possible.
   (b) kept to a minimum.
   (c) just enough to sustain oscillation.
   (d) done through a transformer whose wires can be switched easily.
5. A power gain of 44 dB is equivalent to which of the following output/input power ratios?
   (a) 44:1
   (b) 160:1
   (c) 440:1
   (d) 25,000:1
6. An RF choke
   (a) passes RF signals but blocks dc.
   (b) passes both RF signals and dc.
   (c) passes dc but blocks RF signals.
   (d) blocks both dc and RF signals.
7. The optimum capacitance values in an amplifier circuit depend on
   (a) the power-supply voltage.
   (b) the power-supply polarity.
   (c) the input signal strength.
   (d) the input signal frequency.
8. An oscillator might fail to start for any of the following reasons, except
   (a) low battery voltage.
   (b) low stage gain.

(c) in-phase feedback.
(d) a high-impedance load.

9. In which of the following FET amplifier types does drain current flow for exactly 50 percent of the signal cycle?
(a) Class A
(b) Class AB_1
(c) Class AB_2
(d) Class B

10. The frequency at which a quartz crystal oscillator produces energy is largely dependent on
(a) the load impedance.
(b) the physical thickness of the quartz wafer.
(c) the amount of resistance through the crystal.
(d) the power-supply voltage.

11. Which bipolar amplifier type has some distortion in the signal wave, with collector current during most, but not all, of the cycle?
(a) Class A
(b) Class AB_1
(c) Class AB_2
(d) Class B

12. An RF oscillator usually
(a) produces an output signal with an irregular waveshape.
(b) has most or all of its energy at a single frequency.
(c) produces a sound that depends on its waveform.
(d) employs an RC circuit to determine the output amplitude.

13. A class C amplifier can be made linear by
(a) reducing the bias.
(b) increasing the drive.
(c) using two transistors in push-pull.
(d) no means; a class C amplifier is always nonlinear.

14. A frequency synthesizer has
(a) high power output.
(b) high frequency drift rate.
(c) exceptional stability.
(d) an adjustable waveshape.

15. A graphic equalizer is a form of
(a) bias control for an NPN bipolar transistor.

(b) gain control for an RF oscillator.
(c) tone control that can be used in an audio amplifier.
(d) circuit for adjusting the waveform of an RF oscillator.

16. If the impedance of the load connected to the output of an oscillator is extremely high,
(a) the frequency will drift excessively.
(b) the power output will be reduced.
(c) the oscillator might fail to start.
(d) it is no cause for concern; in fact, it is a good thing.

17. Suppose a certain bipolar-transistor PA is 66 percent efficient. The output power is 33 W. The dc collector power input is
(a) 22 W.
(b) 50 W.
(c) 2.2 W.
(d) impossible to determine without more information.

18. The arrangement in the block diagram of Fig. 24-20 represents
(a) a waveform analyzer.
(b) an audio oscillator.
(c) an RF oscillator.
(d) a sine wave generator.

19. A tuned RF PA must always be
(a) set to work over a wide range of frequencies.
(b) adjusted for maximum power output.
(c) operated at an even harmonic of the input frequency.
(d) operated in class C.

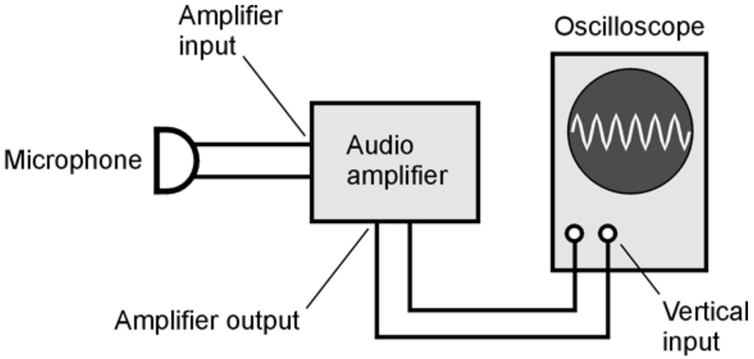

24-20 Illustration for Quiz Question 18.

20. Class B amplification can be used to obtain low distortion for audio applications
    (a) by connecting two amplifiers in cascade, thereby maximizing the gain and generating a pure sine wave output.
    (b) by biasing the bipolar transistor or FET beyond cutoff or pinchoff, thereby ensuring that the output is in phase with the input.
    (c) by connecting two identical bipolar transistors or FETs, biased exactly at cutoff or pinchoff, in a push-pull configuration.
    (d) by biasing the bipolar transistor or FET in the middle of the straight-line portion of the characteristic curve.

25
CHAPTER

Wireless Transmitters and Receivers

IN RADIO OR WIRELESS COMMUNICATIONS, A *TRANSMITTER* CONVERTS DATA INTO *ELECTROMAGNETIC* (EM) *waves* intended for recovery by one or more *receivers.* In this chapter, we'll look at how data is converted to EM waves and transmitted, and then examine how the resulting *EM fields* can be intercepted and received.

Oscillation and Amplification

A radio transmitter employs one or more oscillators to generate an RF signal, and amplifiers to generate the required power output. You just learned how these circuits work. Most transmitters have *mixers* in addition to the oscillating and amplifying stages. Signal mixing is commonly done with diodes, and was discussed in Chap. 20.

Modulation

Modulation is the process of writing data onto an electric current or EM wave. The process can be done by varying the amplitude, the frequency, or the phase of the current or wave. Another method is to transmit a series of pulses, whose duration, amplitude, or spacing is made to vary.

The Carrier

The heart of a wireless signal is a sine wave known as the *carrier.* The lowest carrier frequency used for radio communications is a few kilohertz (kHz). The highest frequency is in the hundreds of gigahertz (GHz). For efficient data transfer, the carrier frequency must be at least 10 times the highest frequency of the modulating signal.

On/Off Keying

The simplest form of modulation is *on/off keying.* This can be done in the oscillator of a radio transmitter to send Morse code, which is a binary digital mode. The duration of a Morse-code *dot* is one *bit* (binary digit). A *dash* is 3 bits long. The space between dots and dashes within a *charac-*

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

ter is 1 bit. The space between characters in a *word* is 3 bits. The space between words is 7 bits. The key-down (full-carrier) condition is called *mark*, and the key-up (no-signal) condition is called *space.*

Morse code is slow. Human operators use speeds ranging from about 5 words per minute (wpm) to 40 or 50 wpm. A few human operators can work at 60 to 70 wpm. These people usually copy the signals in their heads.

Frequency-Shift Keying

Digital data can be sent over wireless links by means of *frequency-shift keying* (FSK). In some FSK systems, the carrier frequency is shifted between mark and space conditions, usually by a few hundred hertz or less. In other systems, a two-tone audio-frequency (AF) sine wave modulates the carrier. This is known as *audio-frequency-shift keying* (AFSK).

The two most common codes used with FSK and AFSK are *Baudot* (pronounced "baw-DOE") and *ASCII* (pronounced "ASK-ee"). The acronym ASCII stands for *American Standard Code for Information Interchange.*

In *radioteletype* (RTTY) FSK and AFSK systems, a *terminal unit* (TU) converts the digital signals into electrical impulses that operate a teleprinter or display characters on a computer screen. The TU also generates the signals necessary to send RTTY as an operator types on a keyboard. A device that sends and receives AFSK is sometimes called a *modem.* This acronym stands for *modulator/demodulator.* A modem is basically the same as a TU. Figure 25-1 is a block diagram of an AFSK transmitter.

The main advantage of FSK or AFSK over on/off keying is the fact that there are fewer errors or misprints, because the space part of the signal is identified as such, rather than existing as a gap or pause in the data. A sudden noise burst in an on/off keyed signal can confuse a receiver into reading the space as a mark signal, but when the space is positively represented by its own signal, this is less likely to happen.

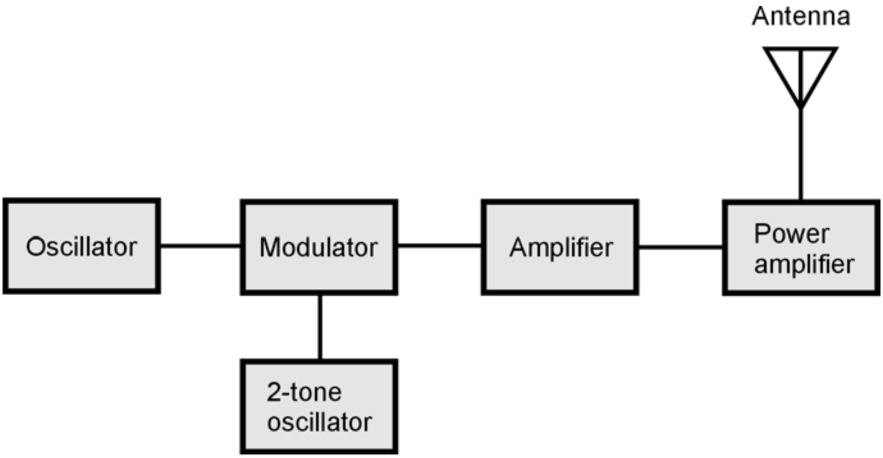

25-1 Simplified block diagram of an AFSK transmitter.

Amplitude Modulation

An AF voice signal has frequencies mostly in the range between 300 Hz and 3 kHz. Some characteristic of an RF carrier can be varied, or *modulated*, by these waveforms, thereby transmitting voice information. Figure 25-2 shows a simple circuit for obtaining *amplitude modulation* (AM). This circuit can be imagined as an RF amplifier for the carrier, with the instantaneous gain dependent on the instantaneous audio input amplitude. Another way to think of this circuit is as a mixer that combines the RF carrier and audio signals to produce sum and difference signals at frequencies just above and below that of the carrier.

The circuit shown in Fig. 25-2 works well, provided the AF input amplitude is not too great. If the AF input is excessive, then distortion occurs, intelligibility is degraded, system efficiency is reduced, and the bandwidth of the signal is increased unnecessarily.

The extent of AM is expressed as a percentage, from 0 percent (an unmodulated carrier) to 100 percent (full modulation). Increasing the modulation past 100 percent causes the same problems as excessive AF input. In an AM signal modulated 100 percent, $\frac{1}{3}$ of the power is used to convey the data, and the other $\frac{2}{3}$ is consumed by the carrier wave.

Figure 25-3 shows a *spectral display* of an AM voice radio signal. The horizontal scale is calibrated in increments of 1 kHz per division. Each vertical division represents 3 dB of change in signal strength. The maximum (reference) amplitude is 0 dB relative to 1 mW (abbreviated as 0 dBm). The data exists in *sidebands* above and below the carrier frequency. These sidebands resemble the sum and difference signals produced by a mixer. In this case the mixing occurs between the AF input signal and the RF carrier. The RF between –3 kHz and the carrier frequency constitutes the *lower*

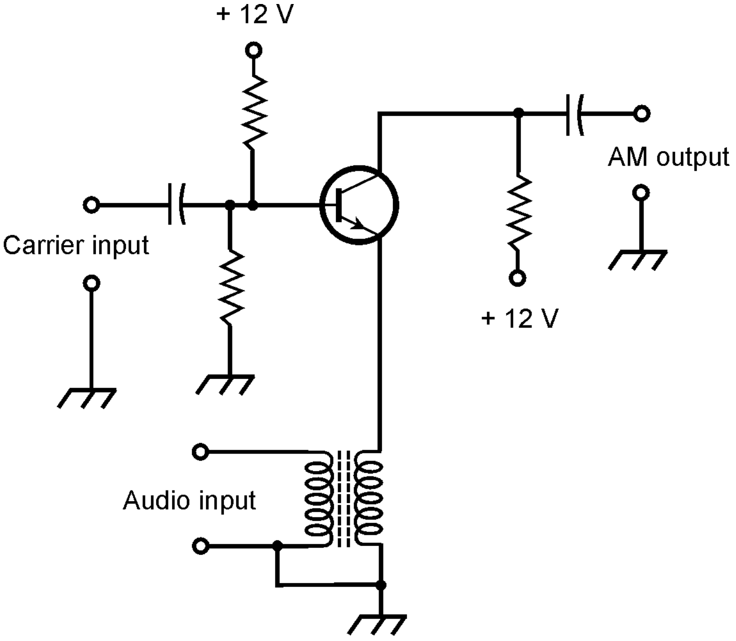

25-2 An amplitude modulator using an NPN bipolar transistor.

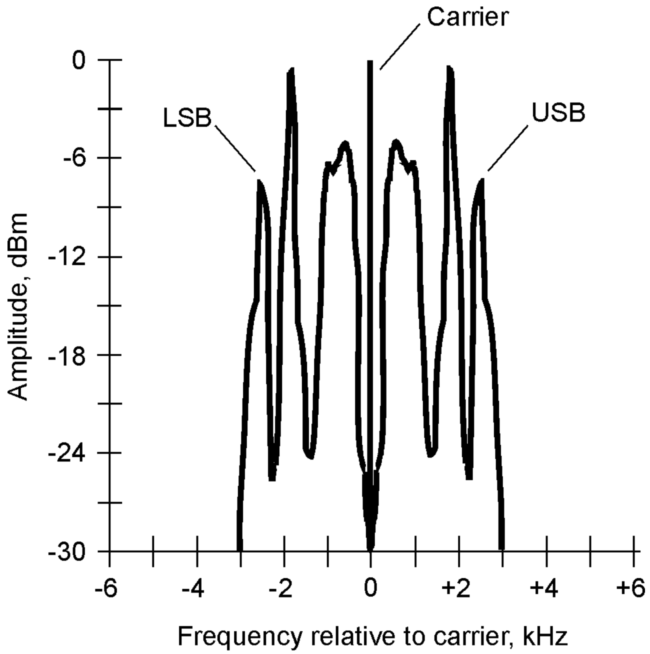

25-3 Spectral display of a typical AM voice communications signal.

sideband (LSB); the RF from the carrier frequency to +3 kHz represents the *upper sideband* (USB). The *bandwidth* is the difference between the maximum and minimum sideband frequencies, in this case 6 kHz.

In an AM signal, the bandwidth is twice the highest audio modulating frequency. In the example of Fig. 25-3, all the voice energy is at or below 3 kHz, so the signal bandwidth is 6 kHz. This is typical of a communications signal. In standard AM broadcasting, the AF energy is spread over a wider bandwidth, nominally 10 kHz.

Single Sideband

In AM, most of the RF signal power is consumed by the carrier alone; the two sidebands are mirror-image duplicates. This is inefficient, and is also unnecessarily redundant! If the carrier and one of the sidebands is eliminated, these shortcomings can be overcome. That makes the signal stronger for a given amount of RF power, or allows the use of lower RF power in a given communications scenario. Another bonus is the fact that the bandwidth is reduced to less than half that of an AM signal modulated with the same data, so more than twice as many signals can fit into a specific range, or *band*, of frequencies.

When the carrier is removed from an AM signal along with one of the sidebands, the remaining RF energy has a spectral display resembling Fig. 25-4. This is *single sideband* (SSB) transmission. Either the LSB or the USB alone can be used, with equally good results.

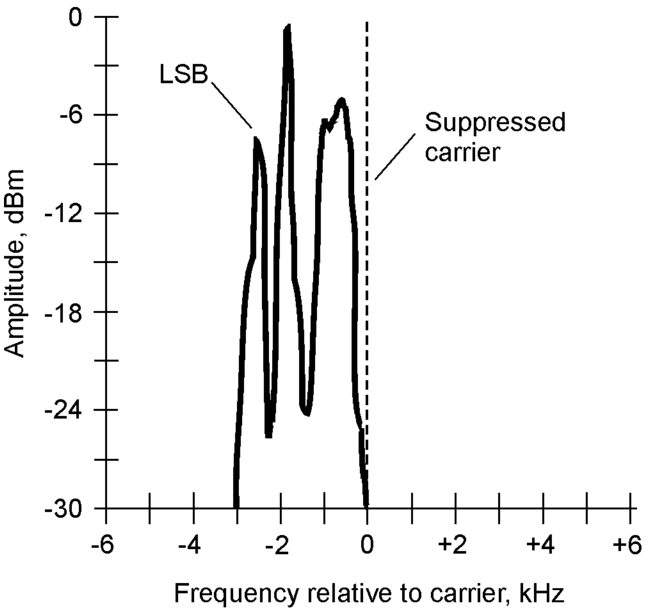

25-4 Spectral display of a typical SSB voice communications signal. In this example, the carrier and the USB energy are eliminated, leaving only the LSB energy.

An SSB signal can be obtained with a *balanced modulator*, which is an amplitude modulator/amplifier using two transistors with the inputs in push-pull and the outputs in parallel (Fig. 25-5). This cancels the carrier wave in the output, leaving only LSB and USB energy. The result is a *double sideband suppressed carrier* (DSBSC) signal, often called simply *double sideband* (DSB). At some stage following the balanced modulator, one of the sidebands is removed from the DSB signal by a *bandpass filter* to obtain an SSB signal.

Figure 25-6 is a block diagram of a simple SSB transmitter. The balanced modulator is placed in a low-power section of the transmitter. The RF amplifiers that follow any type of amplitude modulator, including a balanced modulator, must all be linear to avoid distortion and unnecessary spreading of signal bandwidth ("splatter"). They generally work in class A, except for the PA, which works in class AB or class B.

Frequency and Phase Modulation

In *frequency modulation* (FM), the instantaneous amplitude of a signal remains constant, and the instantaneous frequency is varied instead. A nonlinear PA such as a class C amplifier can be used in an FM transmitter without causing signal distortion, because the amplitude does not fluctuate.

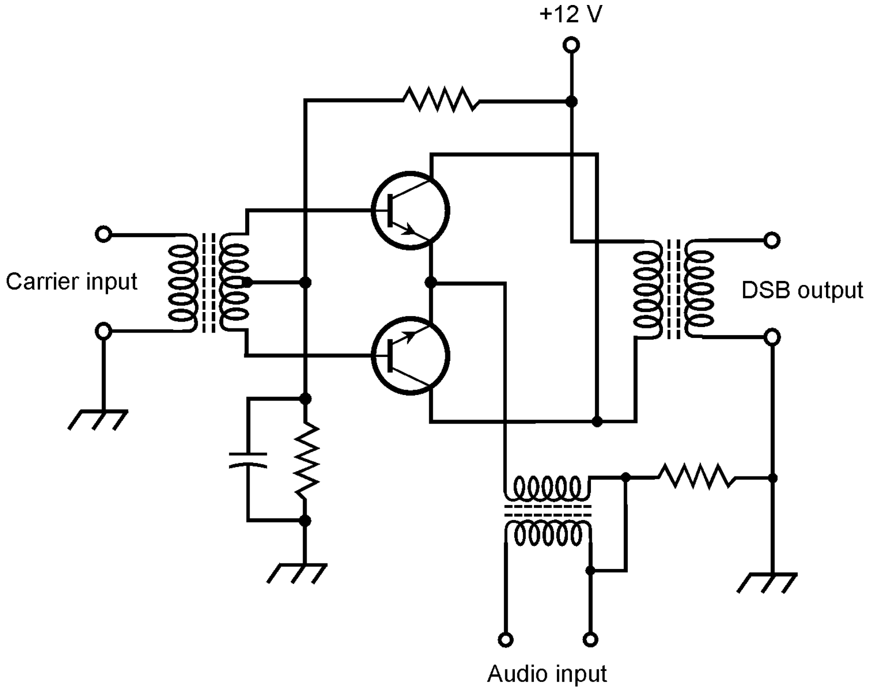
+12 V
Carrier input
DSB output
Audio input

25-5 A balanced modulator using two NPN bipolar transistors. The inputs are in push-pull, but the outputs are in parallel.

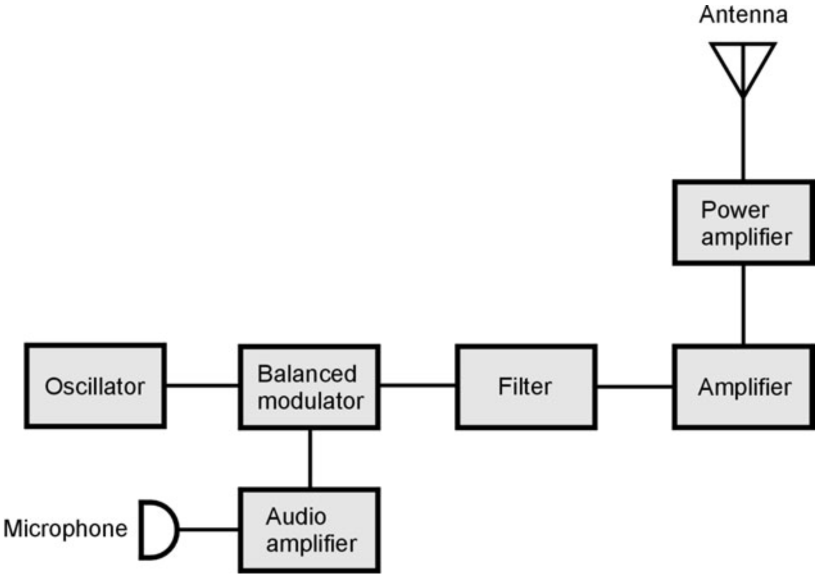
Antenna
Power amplifier
Oscillator
Balanced modulator
Filter
Amplifier
Microphone
Audio amplifier

25-6 Block diagram of a basic SSB transmitter.

Frequency modulation can be obtained by applying the audio signal to a *varactor* in a tuned oscillator. An example of this scheme, known as *reactance modulation*, is shown in Fig. 25-7. The varying voltage across the varactor causes its capacitance to change in accordance with the audio waveform. The changing capacitance results in variation of the resonant frequency of the inductance-capacitance (LC) tuned circuit, causing small fluctuations in the frequency generated by the oscillator.

Another way to get FM is to modulate the phase of the oscillator signal. This causes small variations in the frequency, because any instantaneous phase change shows up as an instantaneous frequency change (and vice versa). When *phase modulation* is used, the audio signal must be processed, adjusting the frequency response of the audio amplifiers. Otherwise the signal will sound unnatural when it is received.

Deviation is the maximum extent to which the instantaneous-carrier frequency differs from the unmodulated-carrier frequency. For most FM voice communications transmitters, the deviation is standardized at ±5 kHz. This is known as *narrowband FM* (NBFM). The bandwidth of an NBFM signal is comparable to that of an AM signal containing the same modulating information. In FM hi-fi music broadcasting, and in some other applications, the deviation is much greater than ±5 kHz. This is called *wideband FM* (WBFM).

The deviation obtainable by means of direct FM is greater, for a given oscillator frequency, than the deviation that can be obtained by means of phase modulation. The deviation of a signal can be increased by a *frequency multiplier*. When an FM signal is passed through a frequency multiplier, the deviation is multiplied along with the carrier frequency.

The deviation in an FM signal should be equal to the highest modulating audio frequency if optimum fidelity is to be obtained. Thus, ±5 kHz is more than enough for voice. For music, a deviation of ±15 kHz or even ±20 kHz is required for good reproduction when the signal is received.

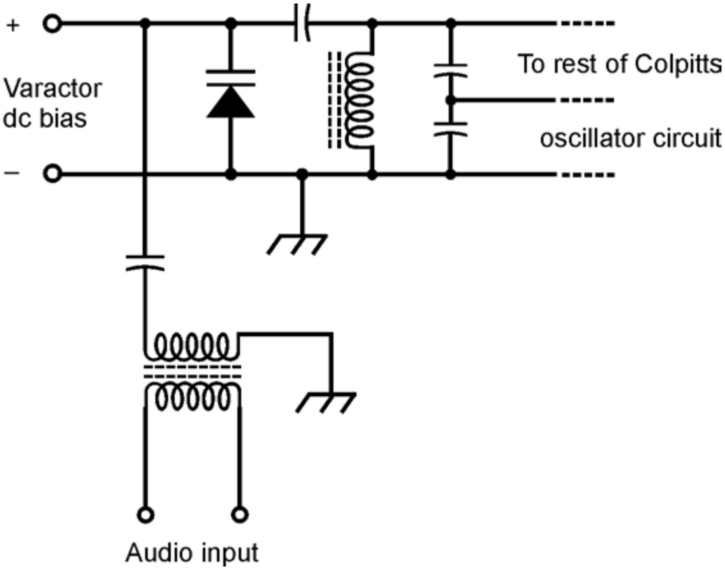

25-7 Generation of FM by reactance modulation of a Colpitts oscillator. Other oscillator types can be similarly modified.

In any FM signal, the ratio of the frequency deviation to the highest modulating audio frequency is called the *modulation index.* Ideally, this figure is between 1:1 and 2:1. If it is less than 1:1, the signal sounds muffled or distorted, and efficiency is sacrificed. Increasing it beyond 2:1 broadens the bandwidth without providing significant improvement in intelligibility or fidelity.

Pulse Modulation

Another method of modulation works by varying some aspect of a constant stream of signal pulses. Several types of *pulse modulation* (PM) are briefly described in the following sections. They are diagrammed in Fig. 25-8 as amplitude-versus-time graphs. The modulating waveform in each case is shown as a dashed curve, and the pulses are shown as vertical gray bars.

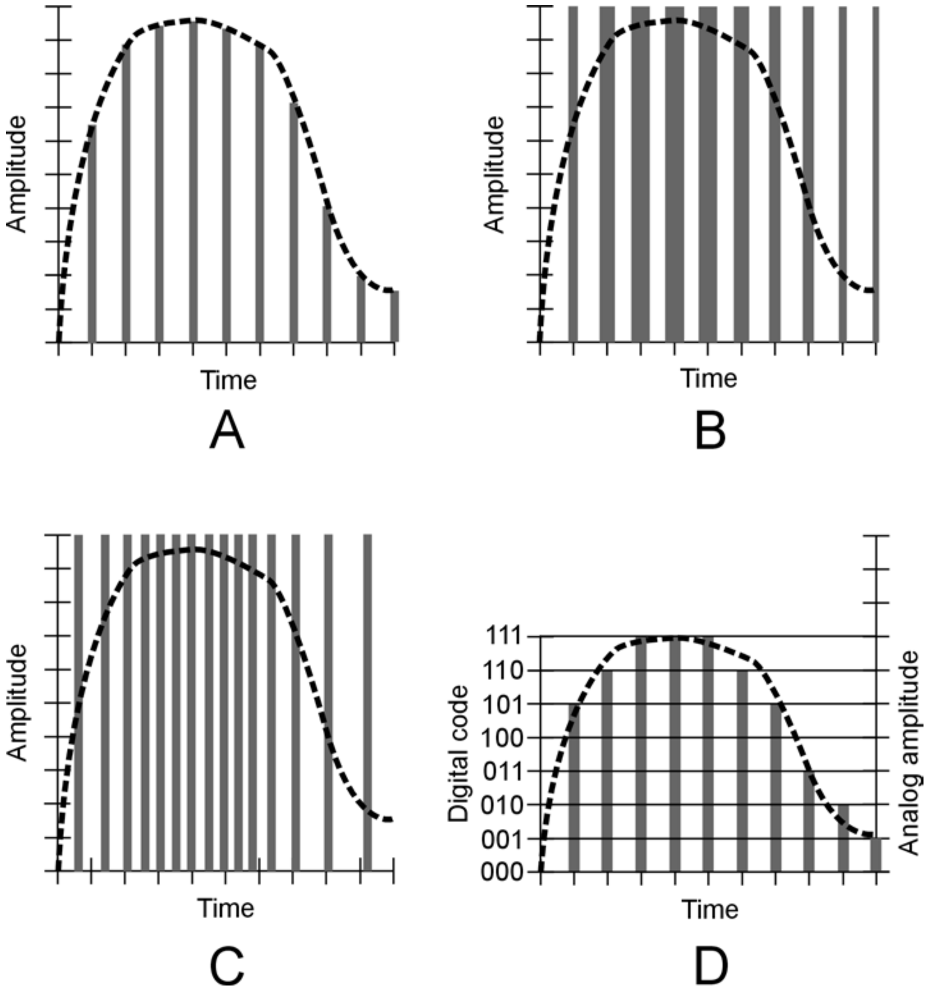

25-8 Time-domain graphs of various modes of pulse modulation. At A, pulse amplitude modulation; at B, pulse width modulation; at C, pulse interval modulation; at D, pulse code modulation.

Pulse Amplitude Modulation

In *pulse amplitude modulation* (PAM), the strength of each individual pulse varies according to the modulating waveform. In this respect, PAM resembles AM. An amplitude-versus-time graph of a hypothetical PAM signal is shown in Fig. 25-8A. Normally, the pulse amplitude increases as the instantaneous modulating-signal level increases (*positive PAM*). But this can be reversed, so higher audio levels cause the pulse amplitude to go down (*negative PAM*). Then the signal pulses are at their strongest when there is no modulation. The transmitter works a little harder if negative PAM is used.

Pulse Width Modulation

Another way to change the transmitter output is to vary the width (duration) of the pulses. This is called *pulse width modulation* (PWM) or *pulse duration modulation* (PDM), and is shown in Fig. 25-8B. Normally, the pulse width increases as the instantaneous modulating-signal level increases (*positive PWM*). But this can be reversed (*negative PWM*). The transmitter must work harder to accomplish negative PWM. Either way, the peak pulse amplitude remains constant.

Pulse Interval Modulation

Even if all the pulses have the same amplitude and the same duration, modulation can still be accomplished by varying how often they occur. In PAM and PWM, the pulses are always sent at the same time interval, known as the *sampling interval.* But in *pulse interval modulation* (PIM), also called *pulse frequency modulation* (PFM), pulses can occur more or less frequently than they do when there is no modulation. A hypothetical PIM signal is shown in Fig. 25-8C. Every pulse has the same amplitude and the same duration, but the time interval between them changes. When there is no modulation, the pulses are evenly spaced with respect to time. An increase in the instantaneous data amplitude might cause pulses to be sent more often, as is the case in Fig. 25-8C (*positive PIM*). Or, an increase in instantaneous data level might slow down the rate at which the pulses are sent (*negative PIM*).

Pulse Code Modulation

In recent years, the transmission of data has been done more and more by *digital* means. In *digital communications*, the modulating data attains only certain defined states, rather than continuously varying. Digital transmission offers better efficiency than analog transmission. With digital modes, the *signal-to-noise* (S/N) *ratio* is better, the bandwidth is narrower, and there are fewer errors. In *pulse-code modulation* (PCM), any of the above aspects—amplitude, duration, or interval—of a pulse sequence (or *pulse train*) can be varied. But rather than having infinitely many possible states, there are finitely many. The number of states is a power of 2, such as 4, 8, or 16. The greater the number of states, the better the fidelity. An example of 8-level PCM is shown in Fig. 25-8D.

Analog-to-Digital Conversion

Pulse code modulation, such as is shown at Fig. 25-8D, is one form of *analog-to-digital* (A/D) *conversion.* A voice signal, or any continuously variable signal, can be *digitized*, or converted into a train of pulses whose amplitudes can achieve only certain defined levels.

Resolution

In A/D conversion, the number of states is always a power of 2, so that it can be represented as a binary-number code. Fidelity improves as the exponent increases. The number of states is called the *sampling resolution*, or simply the *resolution.* A resolution of $2^3 = 8$ (as shown in Fig. 25-8D) is good enough for voice transmission, and is the standard for commercial digital voice circuits. A resolution of $2^4 = 16$ is adequate for high-fidelity (hi-fi) music reproduction.

Sampling Rate

The efficiency with which a signal can be digitized depends on the frequency at which sampling is done. In general, the *sampling rate* must be at least twice the highest data frequency. For an audio signal with components as high as 3 kHz, the minimum sampling rate for effective digitization is 6 kHz; the commercial voice standard is 8 kHz. For hi-fi digital transmission, the standard sampling rate is 44.1 kHz, a little more than twice the frequency of the highest audible sound (approximately 20 kHz).

Image Transmission

Nonmoving images can be sent within the same bandwidth as voice signals. For high-resolution, moving images, the necessary bandwidth is greater.

Facsimile

Nonmoving images (also called *still images*) are commonly transmitted by *facsimile*, also called *fax.* If data is sent slowly enough, any amount of detail can be transmitted within a 3-kHz-wide band, the standard for voice communications. This is why detailed fax images can be sent over a *plain old telephone service* (POTS) line.

In an electromechanical fax machine, a paper document or photo is wrapped around a *drum.* The drum is rotated at a slow, controlled rate. A spot of light scans from left to right; the drum moves the document so a single line is scanned with each pass of the light spot. This continues, line by line, until the complete *frame* (image) has been scanned. The reflected light is picked up by a *photodetector.* Dark parts of the image reflect less light than bright parts, so the current through the photodetector varies. This current modulates a carrier in one of the modes described earlier, such as AM, FM, or SSB. Typically, black is sent as a 1.5-kHz audio sine wave, and white as a 2.3-kHz audio sine wave. Gray shades produce audio sine waves having frequencies between these extremes.

At the receiver, the scanning rate and pattern can be duplicated, and a *cathode-ray tube* (CRT), *liquid crystal display* (LCD), or printer can be used to reproduce the image in *grayscale* (shades of gray ranging from black to white, without color).

Slow-Scan Television

One way to think of *slow-scan television* (SSTV) is to imagine "fast fax." An SSTV signal, like a fax signal, is sent within a band of frequencies as narrow as that of a human voice. And, like fax, SSTV transmission is of still pictures, not moving ones. The big difference between SSTV and fax is that SSTV images are sent in much less time. The time required to send a complete frame (image or scene) is 8 seconds, rather than several minutes. This speed bonus comes with a tradeoff: lower *resolution*, meaning less image detail.

Some SSTV signals are received on CRT displays. A computer can be programmed so that its monitor will act as an SSTV receiver. Converters are also available that allow SSTV signals to be viewed on a consumer-type TV set.

An SSTV frame has 120 lines. The black and white frequencies are the same as for fax transmission; the darkest parts of the picture are sent at 1.5 kHz and the brightest at 2.3 kHz. *Synchronization* (*sync*) *pulses*, that keep the receiving apparatus in step with the transmitter, are sent at 1.2 kHz. A *vertical sync pulse* tells the receiver that it's time to begin a new frame; it lasts for 30 milliseconds (ms). A *horizontal sync pulse* tells the receiver that it's time to start a new line in a frame; its duration is 5 ms. These pulses prevent *rolling* (haphazard vertical image motion) or *tearing* (lack of horizontal synchronization).

Fast-Scan Television

Conventional television is also known as *fast-scan TV* (FSTV). The frames are transmitted at the rate of 30 per second. There are 525 lines per frame. The quick frame time, and the increased resolution, of FSTV make it necessary to use a much wider frequency band than is the case with fax or SSTV. A typical video FSTV signal takes up 6 MHz of spectrum space, or 2000 times the bandwidth of a fax or SSTV signal.

Fast-scan TV is almost always sent using conventional AM. Wideband FM can also be used. With AM, one of the sidebands can be filtered out, leaving just the carrier and the other sideband. This mode is called *vestigial sideband* (VSB) transmission. It cuts the bandwidth of an FSTV signal down to about 3 MHz.

Because of the large amount of spectrum space needed to send FSTV, this mode isn't practical at frequencies below about 30 MHz. All commercial FSTV transmission is done above 50 MHz, with the great majority of channels having frequencies far higher than this. Channels 2 through 13 on your TV receiver are sometimes called the *very high frequency* (VHF) *channels;* the higher channels are called the *ultrahigh frequency* (UHF) *channels.*

Figure 25-9 is a time-domain graph of the waveform of a single line in an FSTV video signal. This represents $^1/_{525}$ of a complete frame. The highest instantaneous signal amplitude corresponds to the blackest shade, and the lowest amplitude to the lightest shade. Thus, the FSTV signal is sent negatively. The reason that FSTV signals are sent this way is that *retracing* (moving from the end of one line to the beginning of the next) must be synchronized between the transmitter and receiver. This is guaranteed by a defined, strong *blanking pulse.* This pulse tells the receiver when to retrace; it also shuts off the beam while the receiver display is retracing. Have you noticed that weak TV signals have poor *contrast*? (You have, if you're old enough to remember "rabbit ears"!) Weakened blanking pulses result in incomplete retrace blanking. But this is better than having the TV receiver completely lose track of when it should retrace.

Color FSTV works by sending three separate monochromatic signals, corresponding to the primary colors red, blue, and green. The signals are literally black-and-red, black-and-blue, and black-and-green. These are recombined at the receiver and displayed on the screen as a fine, interwoven matrix of red, blue, and green dots. When viewed from a distance, the dots are too small to be individually discernible. Various combinations of red, blue, and green intensities result in reproduction of all possible hues and saturations of color.

High-Definition Television

The term *high-definition television* (HDTV) refers to any of several similar methods for getting more detail into a TV picture, and for obtaining better audio quality, compared with standard FSTV.

A standard FSTV picture has 525 lines per frame, but HDTV systems have between 787 and 1125 lines per frame. The image is scanned about 60 times per second. High-definition TV is often

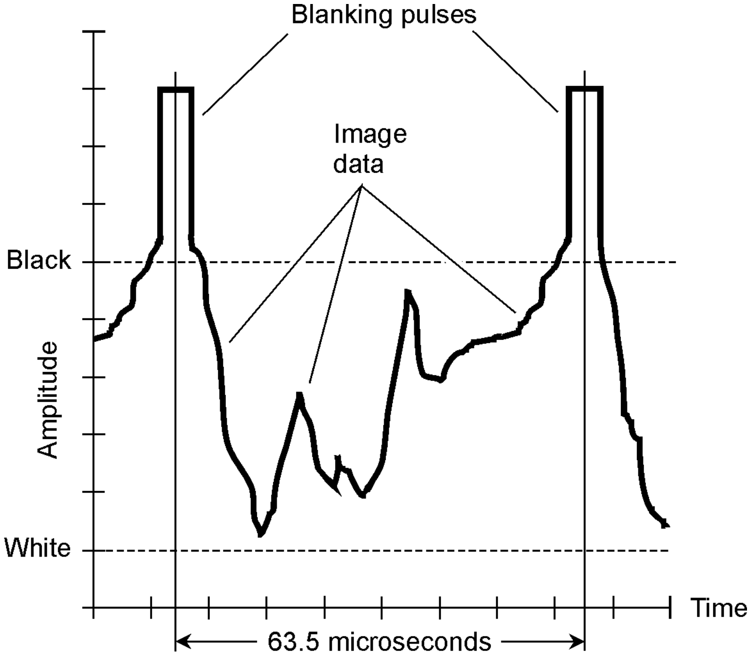

25-9 Time-domain graph of a single line in an FSTV video frame.

sent in a digital mode; this offers another advantage over conventional FSTV. Digital signals propagate better, are easier to deal with when they are weak, and can be processed in ways that analog signals cannot.

Some HDTV systems use *interlacing* in which two *rasters* are meshed together. This effectively doubles the image resolution without doubling the cost of the hardware. But it can cause annoying *jitter* in fast-moving or fast-changing images.

Digital Satellite TV

Until the early 1990s, a satellite television installation required a dish antenna several feet in diameter. A few such systems are still in use. The antennas are expensive, they attract attention (sometimes unwanted), and they are subject to damage from ice storms, heavy snows, and high winds. Digitization has changed this situation. In any communications system, digitization allows the use of smaller receiving antennas, smaller transmitting antennas, and/or lower transmitter power levels. Engineers have managed to get the diameter of the receiving dish down to about 2 ft.

A pioneer in digital TV was RCA (Radio Corporation of America), which developed the *Digital Satellite System* (DSS). The analog signal is changed into digital pulses at the transmitting station via A/D conversion. The digital signal is amplified and sent up to a geostationary satellite. The satellite has a *transponder* that receives the signal, converts it to a different frequency, and retransmits it back toward the earth. The return signal is picked up by a portable dish. A *tuner* selects the channel. *Digital signal processing* (DSP) can be used to improve the quality of reception under marginal conditions. The digital signal is changed back into analog form, suitable for viewing on a conventional FSTV set, by means of *digital-to-analog* (D/A) *conversion.*

The Electromagnetic Field

In a radio or television transmitting antenna, electrons are moving back and forth at an extreme speed. Their velocity is constantly changing as they speed up in one direction, slow down, reverse direction, speed up again, and so on. Any change of velocity (that is, of speed and/or direction) constitutes *acceleration.*

How It Happens

When electrons move, a magnetic (M) field is produced. When electrons accelerate, a changing magnetic field is produced. An alternating M field gives rise to an alternating electric (E) field, and this generates another alternating M field. This process repeats over and over, endlessly, and the effect *propagates* (travels) through space at the speed of light. The E and M fields expand alternately outward from the source in spherical wavefronts. At any given point in space, the E flux is perpendicular to the M flux. The direction of wave travel is perpendicular to both the E and M flux lines. This is an *electromagnetic* (EM) *field.*

An EM field can have any conceivable frequency, ranging from many years per cycle to quadrillions of cycles per second. The sun has a magnetic field that oscillates with a 22-year cycle. Radio waves oscillate at thousands, millions, or billions of cycles per second. Infrared (IR), visible light, ultraviolet (UV), X rays, and gamma rays are EM fields that alternate at many trillions (million millions) of cycles per second.

Frequency versus Wavelength

All EM fields have two important properties: the *frequency* and the *wavelength.* These are inversely related. You've already learned about frequency. Wavelength, for an EM field, is a rather sophisticated concept. It is measured between any two adjacent points on the wave at which the E and M fields have exactly the same amplitudes, and occur in exactly the same relative directions. The following equations relate the frequency and the wavelength of an EM field in free space (the air or a vacuum).

Let f_{MHz} be the frequency of an EM wave in megahertz, and L_{ft} be the wavelength in feet. Then the two are related as follows:

$$L_{\mathrm{ft}} = 984 / f_{\mathrm{MHz}}$$

If the wavelength is given as L_{m} in meters, then

$$L_{\mathrm{m}} = 300 / f_{\mathrm{MHz}}$$

The inverses of these formulas, for finding the frequency if the wavelength is known, are

$$f_{\mathrm{MHz}} = 984 / L_{\mathrm{ft}}$$
$$f_{\mathrm{MHz}} = 300 / L_{\mathrm{m}}$$

Velocity Factor

In media other than free space, the speed at which EM fields propagate is slower than the speed of light. As a result, wavelength is shortened by a factor known as the *velocity factor*, symbolized v. The value of v can be anything between 0 (representing zero speed of propagation) and 1 (representing the speed of propagation in free space, which is approximately 186,000 mi/s or 300,000 km/s). The velocity factor can also be expressed as a percentage $v_{\%}$. In that case, the smallest possible value is 0 percent, and the largest is 100 percent. The velocity factor in practical situations is rarely less than about 0.60, or 60 percent.

Velocity factor is important in the design of RF transmission lines and antenna systems, when sections of cable, wire, or metal tubing must be cut to specific lengths measured in wavelengths or fractions of a wavelength. Taking the velocity factor v, expressed as a ratio, into account, the preceding four formulas become:

$$L_{\mathrm{ft}} = 984v / f_{\mathrm{MHz}}$$
$$L_{\mathrm{m}} = 300v / f_{\mathrm{MHz}}$$
$$f_{\mathrm{MHz}} = 984v / L_{\mathrm{ft}}$$
$$f_{\mathrm{MHz}} = 300v / L_{\mathrm{m}}$$

The Electromagnetic Spectrum

The entire range of EM wavelengths is called the *electromagnetic* (EM) *spectrum.* Scientists use logarithmic scales to depict the EM spectrum, as shown in Fig. 25-10. The *RF spectrum*, which includes radio, television, and microwaves, is blown up in this illustration, and is labeled for frequency.

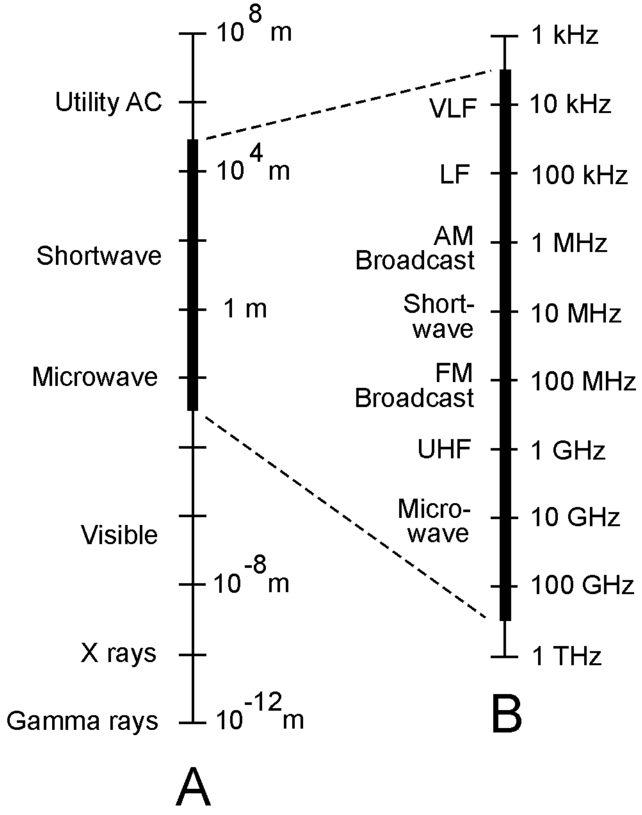

10^8 m
Utility AC
10^4 m
Shortwave
1 m
Microwave
Visible
10^-8 m
X rays
Gamma rays
10^-12 m
A
1 kHz
VLF
10 kHz
LF
100 kHz
AM Broadcast
1 MHz
Short-wave
10 MHz
FM Broadcast
100 MHz
UHF
1 GHz
Micro-wave
10 GHz
100 GHz
1 THz
B

25-10 At A, the EM spectrum from 10^8 m to 10^{-12} m, with each vertical division representing two orders of magnitude (an increase or decrease of the wavelength by a factor of 100). At B, the RF spectrum, with each vertical division representing one order of magnitude (an increase or decrease of the frequency by a factor of 10).

Wave Propagation

Radio-wave propagation has been a fascinating science ever since Marconi and Tesla discovered, around the year 1900, that EM fields can travel over long distances without any supporting infrastructure whatsoever. Let's look at a few of the factors that affect wireless communications at radio frequencies.

Polarization

The orientation of the E field *lines of flux* is defined as the *polarization* of an EM wave. If the E field flux lines are parallel to the earth's surface, you have *horizontal polarization.* If the E field flux lines are perpendicular to the surface, you have *vertical polarization.* Polarization can also be slanted, of course.

In some situations, the E flux lines rotate as the wave travels through space. This is *circular polarization* if the E field intensity remains constant. If the E field intensity is more intense in some planes than in others, the polarization is *elliptical.* Rotating polarization can be *clockwise* or *counterclockwise*, viewed as the wavefronts approach. The rotational direction is called the *sense* of polarization.

The Line-of-Sight Wave

Electromagnetic waves follow straight lines unless something makes them bend. *Line-of-sight* propagation can take place even when the receiving antenna cannot be seen from the transmitting antenna. To some extent, radio waves penetrate nonconducting objects such as trees and frame houses. The line-of-sight wave consists of two components: the *direct wave* and the *reflected wave.*

The direct wave: The longest wavelengths are least affected by obstructions. At very low, low, and medium frequencies, direct waves can *diffract* around things. As the frequency rises, especially above about 3 MHz, obstructions have a greater and greater blocking effect.

The reflected wave: Electromagnetic waves reflect from the earth's surface and from conducting objects like wires and steel beams. The reflected wave always travels farther than the direct wave. The two waves are usually not in phase at the receiving antenna. If they're equally strong but 180° out of phase, a *dead spot* occurs. This phenomenon is most noticeable at the highest frequencies. At VHF and UHF, an improvement in reception can sometimes result from moving the transmitting or receiving antenna just a few inches. In mobile operation, when the transmitter and/or receiver are moving, dead spots produce rapid, repeated interruptions in the received signal. This is called *picket fencing.*

The Surface Wave

At frequencies below about 10 MHz, the earth's surface conducts ac quite well. Because of this, vertically polarized EM waves follow the surface for hundreds or even thousands of miles, with the earth actually helping to transmit the signals. The lower the frequency, the lower the *ground loss*, and the farther the waves can travel by *surface-wave propagation.* Horizontally polarized waves do not travel well in this mode, because horizontal E field flux is shorted out by the earth. Above about 10 MHz, the earth becomes lossy, and surface-wave propagation is not useful for more than a few miles.

Sky-Wave EM Propagation

Ionization in the upper atmosphere, caused by solar radiation, can return EM waves to the earth at certain frequencies. The ionization takes place at three or four distinct layers.

The lowest ionized region is called the *D layer.* It exists at an altitude of about 30 mi (50 km), and is ordinarily present only on the daylight side of the planet. This layer absorbs radio waves at some frequencies, impeding long-distance ionospheric propagation.

The *E layer*, about 50 mi (80 km) above the surface, also exists mainly during the day, although nighttime ionization is sometimes observed. The E layer can provide medium-range radio communication at certain frequencies.

The uppermost layers are called the *F_1 layer* and the *F_2 layer.* The F_1 layer, normally present only on the daylight side of the earth, forms at about 125 mi (200 km) altitude; the F_2 layer exists at about 180 mi (300 km) over most, or all, of the earth. Sometimes the distinction between the F_1 and F_2 layers is ignored, and they are spoken of together as the *F layer.* Communication by means of F-layer propagation can usually be accomplished between any two points on the earth at some frequencies between 5 MHz and 30 MHz.

Tropospheric Propagation

At frequencies above about 30 MHz, the lower atmosphere bends radio waves toward the surface. *Tropospheric banding* occurs because the *index of refraction* of air, with respect to EM waves, decreases with altitude. The effect is similar to the way sound waves sometimes travel long distances over the surface of a calm lake in the early morning or early evening, letting you hear a conversation more than a mile away. Tropospheric propagation makes it possible to communicate for hundreds of miles when the ionosphere will not return waves to the earth.

Another type of tropospheric propagation is called *ducting.* It takes place when EM waves are trapped in a layer of cool, dense air sandwiched between two layers of warmer air. Like bending, ducting occurs almost entirely at frequencies above 30 MHz.

Still another tropospheric-propagation mode is known as *troposcatter.* This takes place because air molecules, dust grains, and water droplets scatter some of the EM field. This effect is commonly seen at VHF and UHF.

Tropospheric propagation in general, without mention of the specific mode, is sometimes called *tropo.*

Auroral Propagation

In the presence of unusual solar activity, the *aurora* (northern lights or southern lights) can return radio waves to the earth. This is known as *auroral propagation.* The aurora occur at altitudes of about 40 to 250 mi (65 to 400 km). Theoretically, auroral propagation is possible, when the aurora are active, between any two points on the surface from which the same part of the aurora lie on a line of sight. Auroral propagation seldom occurs when one end of the circuit is at a latitude less than 35° north or south of the equator.

Auroral propagation is characterized by rapid and deep fading. This almost always renders analog voice and video signals unintelligible. Digital modes are most effective for communication via auroral propagation, but the carrier is often spread out over several hundred hertz as a result of phase modulation induced by auroral motion. This severely limits the maximum data transfer rate. Auroral propagation is often accompanied by deterioration in ionospheric propagation.

Meteor Scatter

Meteors produce ionized trails that persist for approximately 0.5 s up to several seconds, depending on the size of a particular meteor, its speed, and the angle at which it enters the atmosphere. This is not enough time for the transmission of much data, but during a *meteor shower*, multiple trails can result in almost continuous ionization for a period of hours. Such ionized regions reflect radio waves at certain frequencies. This is *meteor scatter propagation.* It can take place at frequencies considerably above 30 MHz, and occurs over distances ranging from just beyond the horizon up to about 1500 mi (2400 km), depending on the altitude of the ionized trail and the relative positions of the trail, the transmitting station, and the receiving station.

Moonbounce

The moon, like the earth, reflects EM fields. This makes it possible to communicate by means of *earth-moon-earth* (EME), also called *moonbounce.* High-powered transmitters, sophisticated antenna systems, and sensitive receivers are needed for EME. Some moonbounce communication is done by radio amateurs at frequencies from 50 MHz to over 2 GHz.

Transmission Media

Data can be transmitted over various *media.* The most common are *cable*, *radio* (also called *wireless*), *satellite links* (a specialized form of wireless), and *fiber optics.* Cable, radio/TV, and satellite communications use the RF spectrum. Fiber optics uses IR or visible light energy.

Cable

The earliest cables were simple wires that carried dc. Nowadays, data transmission cables more often carry ac at radio frequencies. One advantage of using RF is the fact that the signals can be amplified at intervals on a long span. This greatly increases the distances over which data can be sent by cable. Another advantage of using RF is the fact that numerous signals can be carried over a single cable, with each signal on a different frequency.

Cables can consist of pairs of wires, somewhat akin to lamp cords. But more often coaxial cable, of the type described and illustrated at the end of Chap. 10, is used. This has a center conductor surrounded by a cylindrical shield. The shield is grounded, and the center conductor carries the signals. The shield keeps signals confined to the cable, and also keeps external EM fields from interfering with the signals.

Radio

All radio and TV signals are electromagnetic waves. The radio or TV transmitter output is coupled into an *antenna system* located at some distance from the transmitter. The energy follows a *transmission line*, also called a *feed line*, from the transmitter output to the antenna itself.

Most radio antenna transmission lines are coaxial cables. There are other types, used in special applications. At microwaves, hollow tubes called *waveguides* are used to transfer the energy from a transmitter to the antenna. A waveguide is more efficient than coaxial cable at the shortest radio wavelengths. Radio amateurs sometimes use *parallel-wire* transmission lines, resembling the ribbon cable popular for use with consumer TV receiving antennas. In a parallel-wire line, the RF currents in the two conductors are always 180° out of phase, so that their EM fields cancel each other. This keeps the transmission line from radiating, guiding the EM field along toward the antenna. The energy is radiated when it reaches the antenna.

The RF bands are generally categorized from *very low frequency* (VLF) through *extremely high frequency* (EHF), according to the breakdown in Table 25-1. As noted in Chap. 24, the exact lower limit of the VLF range is a matter of disagreement in the literature. In Table 25-1, it is defined as 3 kHz, which is consistent with defining the frequency boundaries between RF bands by order of magnitude.

Table 25-1. Bands in the RF spectrum.

Frequency designation	Frequency range	Wavelength range
Very Low (VLF)	3 kHz–30 kHz	100 km–10 km
Low (LF)	30 kHz–300 kHz	10 km–1 km
Medium (MF)	300 kHz–3 MHz	1 km–100 m
High (HF)	3 MHz–30 MHz	100 m–10 m
Very High (VHF)	30 MHz–300 MHz	10 m–1 m
Ultra High (UHF)	300 MHz–3 GHz	1 m–100 mm
Super High (SHF)	3 GHz–30 GHz	100 mm–10 mm
Extremely High (EHF)	30 GHz–300 GHz	10 mm–1 mm

Satellite Systems

At very high frequencies (VHF) and above, many communications circuits use satellites in *geostationary orbits* around the earth. If a satellite is directly over the equator at an altitude of 22,300 mi (36,000 km) and orbits from west to east, it follows the earth's rotation, thereby staying in the same spot in the sky as seen from the surface, and is thus a *geostationary satellite.*

A single geostationary satellite is on a line of sight with about 40 percent of the earth's surface. Three such satellites, placed at 120° ($^1/_3$ circle) intervals around the planet, allow coverage of all populated regions. A *dish antenna* can be aimed at a geostationary satellite, and once the antenna is in place, it need not be turned or adjusted.

Another form of satellite system uses multiple satellites in low orbits that take them over the earth's poles. These satellites are in continuous, rapid motion with respect to the surface. But if there are enough of them, they can act like repeaters in a cell phone network, and maintain reliable communications between any two points on the surface at all times. Directional antennas are not necessary in these systems, which are called *low earth orbit* (LEO) satellite networks.

Fiber Optics

Beams of IR or visible light can be modulated, just as can RF carriers. The frequency of an IR or visible light beam is higher than the frequency of any RF signal, allowing modulation by data at rates higher than anything possible with radio.

Fiber-optic technology offers several advantages over wire cables (which are sometimes called *copper* because the conductors are usually made of that metallic element). A fiber-optic cable is cheap. It is light in weight. It is immune to interference from outside EM fields. A fiber-optic cable does not corrode as metallic wires do. Fiber-optic cables are inexpensive to maintain and easy to repair. An optical fiber can carry far more signals than a cable, because the frequency bands are far wider in terms of megahertz or gigahertz.

The whole RF spectrum, from VLF through EHF, can (at least in theory) be imprinted on a single beam of light and sent through an optical fiber no thicker than a strand of hair!

Two Basic Receiver Designs

A wireless or *radio receiver* converts EM waves into the original messages sent by a distant transmitter. Let's begin our study of receivers by defining a few of the most important criteria for operation, and then we'll look at two common designs.

Specifications

The *specifications* of a receiver indicate how well it can do the functions it is designed to perform.

Sensitivity: The most common way to express receiver sensitivity is to state the number of microvolts that must exist at the antenna terminals to produce a certain *signal-to-noise* (S/N) *ratio* or *signal-plus-noise-to-noise* (S+N/N) *ratio* in decibels (dB). Sensitivity is related to the gain of the *front end* (the amplifier or amplifiers connected to the antenna), but the amount of noise this stage generates is more significant, because subsequent stages amplify the front-end noise output as well as the signal output.

Selectivity: The *passband*, or bandwidth that the receiver can hear, is established by a wideband *preselector* in the early RF amplification stages, and is honed to precision by narrowband filters in later amplifier stages. The preselector makes the receiver optimally sensitive within a range of approximately plus-or-minus 10 percent (±10%) of the desired signal frequency. The narrowband filter responds only to the frequency or channel of a specific signal to be received; signals in nearby channels are rejected.

Dynamic range: The signals at a receiver input vary over several orders of magnitude (multiples or powers of 10) in terms of absolute voltage. Dynamic range is the ability of a receiver to maintain a fairly constant output, and yet to maintain its rated sensitivity, in the presence of signals ranging from very weak to very strong. The dynamic range in a good receiver is in excess of 100 dB.

Noise figure: The less internal noise a receiver produces, in general, the better is the S/N ratio. Excellent S/N ratio in the presence of weak signals is only possible when the *noise figure*, a measure of internally generated receiver noise, is low. This is paramount at VHF, UHF, and microwave frequencies. Gallium-arsenide field effect transistors (GaAsFETs) are well known for the low levels of noise they generate, even at quite high frequencies. Other types of FETs can be used at lower frequencies. *Bipolar transistors* tend to be rather noisy.

Direct-Conversion Receiver

A *direct-conversion receiver* derives its output by mixing incoming signals with the output of a tunable (that is, variable frequency) *local oscillator* (LO). The received signal is fed into a mixer, along with the output of the LO. Figure 25-11 is a block diagram of a direct-conversion receiver.

For the reception of on/off keyed Morse code, also called *radiotelegraphy* or *continuous wave* (CW), the LO, also called a *beat-frequency oscillator* (BFO), is set a few hundred hertz above or below the signal frequency. This can also be done in order to receive FSK signals. The audio output has a frequency equal to the difference between the LO and incoming carrier frequencies. For reception of AM or SSB signals, the LO is set to precisely the same frequency as that of the signal carrier. This condition is known as *zero beat* because the *beat frequency*, or difference frequency, between the LO and the signal carrier is equal to zero.

A direct-conversion receiver provides rather poor *selectivity.* That means it can't separate incoming signals very well when they are close together in frequency. This is because signals on either side of the LO frequency can be heard at the same time. A *selective filter* can theoretically eliminate this. Such a filter must be designed for a fixed frequency if it is to work well. But in a direct-conversion receiver, the RF amplifier works over a wide range of frequencies.

Superheterodyne Receiver

A *superheterodyne receiver*, also called a *superhet*, uses one or more local oscillators and mixers to obtain a constant-frequency signal. A fixed-frequency signal is more easily processed than a signal that

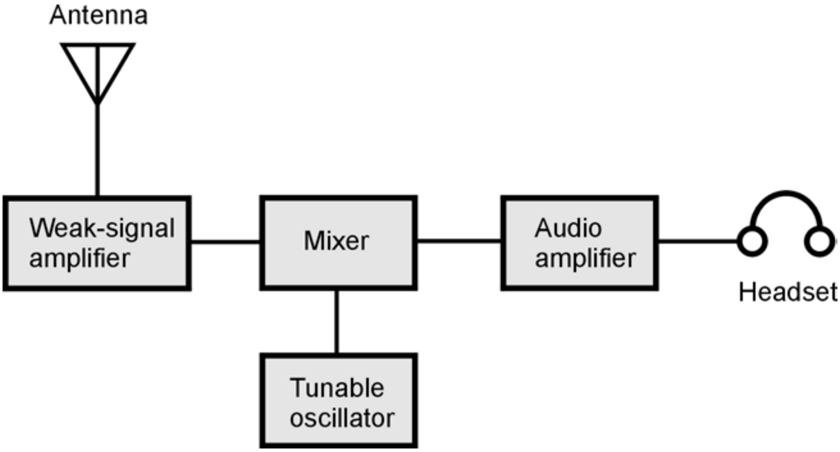

25-11 Block diagram of a direct-conversion receiver.

changes in frequency. The incoming signal is first passed through a tunable, sensitive front end. The output of the front end is mixed with the signal from a tunable, unmodulated LO. Either the sum or the difference signal is amplified. This is the *first intermediate frequency* (IF), which can be filtered to obtain a high degree of selectivity.

If the first IF signal is detected, the radio is a *single-conversion receiver.* Some receivers use a second mixer and second LO, converting the first IF to a lower-frequency *second IF.* This is a *double-conversion receiver.* The IF bandpass filter can be constructed for use on a fixed frequency, allowing superior selectivity and facilitating adjustable bandwidth. The sensitivity is enhanced because fixed IF amplifiers are easy to keep in tune.

A superheterodyne receiver can intercept or generate unwanted signals. False signals external to the receiver are called *images;* internally generated signals are called *birdies.* If the LO frequencies are carefully chosen, images and birdies do not cause problems during ordinary operation of the receiver.

Figure 25-12 is a block diagram of a generic single-conversion superheterodyne receiver. (Individual receiver designs vary somewhat.) Here's what each stage does.

Front end: The front end consists of the first RF amplifier, and often includes *LC* bandpass filters between the amplifier and the antenna. The dynamic range and sensitivity of a receiver are determined by the performance of the front end.

Mixer: A mixer stage converts the variable signal frequency to a constant IF. The output is either the sum or the difference of the signal frequency and the tunable LO frequency.

IF stages: The IF stages are where most of the gain takes place. These stages are also where optimum selectivity is obtained.

Detector: The detector extracts the information from the signal. Common circuits are the *envelope detector* for AM, the *product detector* for SSB, FSK, and CW, and the *ratio detector* for FM.

Audio amplifier: Following the detector, one or two stages of audio amplification are employed to boost the signal to a level suitable for a speaker or headset. Alternatively, the signal can be fed to a printer, facsimile machine, or computer.

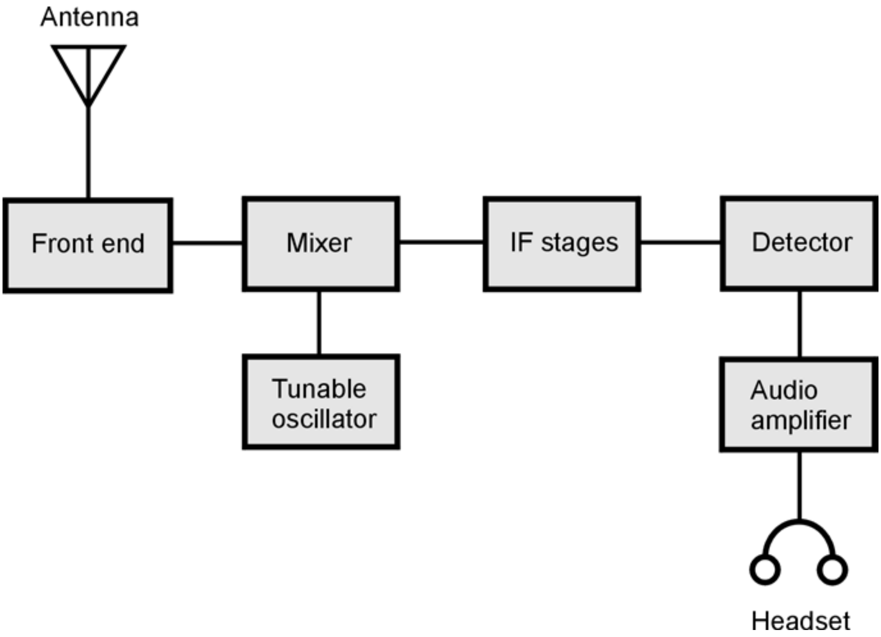

25-12 Block diagram of a single-conversion superheterodyne receiver.

Predetector Stages

In a superhet, the stages preceding the first mixer must be designed so they provide reasonable gain, but produce as little noise as possible. They must also be capable of handling strong signals without *desensitization* (losing gain), also known as *overloading*.

Preamplifier

All preamplifiers operate in class A, and most employ FETs. An FET has a high input impedance that is ideally suited to weak-signal work. Figure 25-13 shows a simple RF preamplifier circuit. Input tuning reduces noise and provides some selectivity. This circuit produces 5 dB to 10 dB gain, depending on the frequency and the choice of FET.

It is important that the preamplifier be linear, and that it remain linear in the presence of strong input signals. Nonlinearity results in unwanted mixing in RF amplifiers. The *mixing products* produce *intermodulation distortion* (IMD), or *intermod*. That can wreak havoc in a receiver, producing numerous false signals. It also degrades the S/N ratio by generating *hash*, the result of complex mixing of many false signals over a wide range of frequencies.

The Front End

At low and medium frequencies, there is considerable atmospheric noise, and the design of a front-end circuit is simple. Above 30 MHz, atmospheric noise diminishes, and the main factor that limits the sensitivity is noise generated within the receiver. For this reason, front-end design becomes increasingly critical as the frequency rises through the VHF, UHF, and microwave spectra.

The front end, like a preamplifier, must be as linear as possible; the greater the degree of nonlinearity, the more susceptible the circuit is to the generation of mixing products. The front end should also have the greatest possible dynamic range.

Preselector

The preselector provides a bandpass response that improves the S/N ratio, and reduces the likelihood of receiver overloading by a strong signal far removed from the operating frequency. The preselector provides *image rejection* in a superheterodyne circuit. Most preselectors have a 3-dB bandwidth that is a few percent of the received frequency.

A preselector can be tuned by means of *tracking* with the tuning dial, but this requires careful design and alignment. Some receivers incorporate preselectors that must be adjusted independently of the receiver tuning.

IF Chain

A high IF (several megahertz) is preferable to a low IF (less than 1 MHz) for image rejection. But a low IF is better for obtaining good selectivity. Double-conversion receivers have a comparatively high first IF and a low second IF to get the best of both worlds.

Intermediate-frequency amplifiers can be cascaded with tuned-transformer coupling. The amplifiers follow the mixer and precede the detector. Double-conversion receivers have two chains of IF amplifiers. The first IF chain follows the first mixer and precedes the second mixer, and the second IF chain follows the second mixer and precedes the detector.

The selectivity of the IF chain in a superheterodyne receiver can be expressed mathematically. The bandwidths are compared for two power-attenuation values, usually 3 dB and 30 dB. This gives an indication of the shape of the bandpass response. The ratio of the 30-dB selectivity to the 3-dB selectivity is called the *shape factor.* A *rectangular response* is desirable in most applications. The smaller the shape factor, the more rectangular the response.

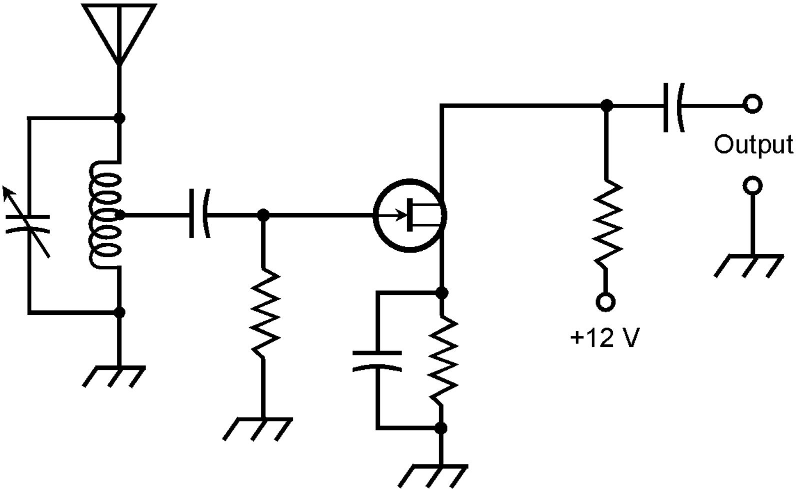

25-13 A tunable FET preamplifier for use in a radio receiver.

Detectors

Detection, also called *demodulation*, is the recovery of information such as audio, images, or printed data from a signal.

Detection of AM

The modulating waveform can be extracted from an AM signal by rectifying the carrier wave. A simplified time-domain view of this is shown in Fig. 25-14A. The rapid pulsations occur at the car-

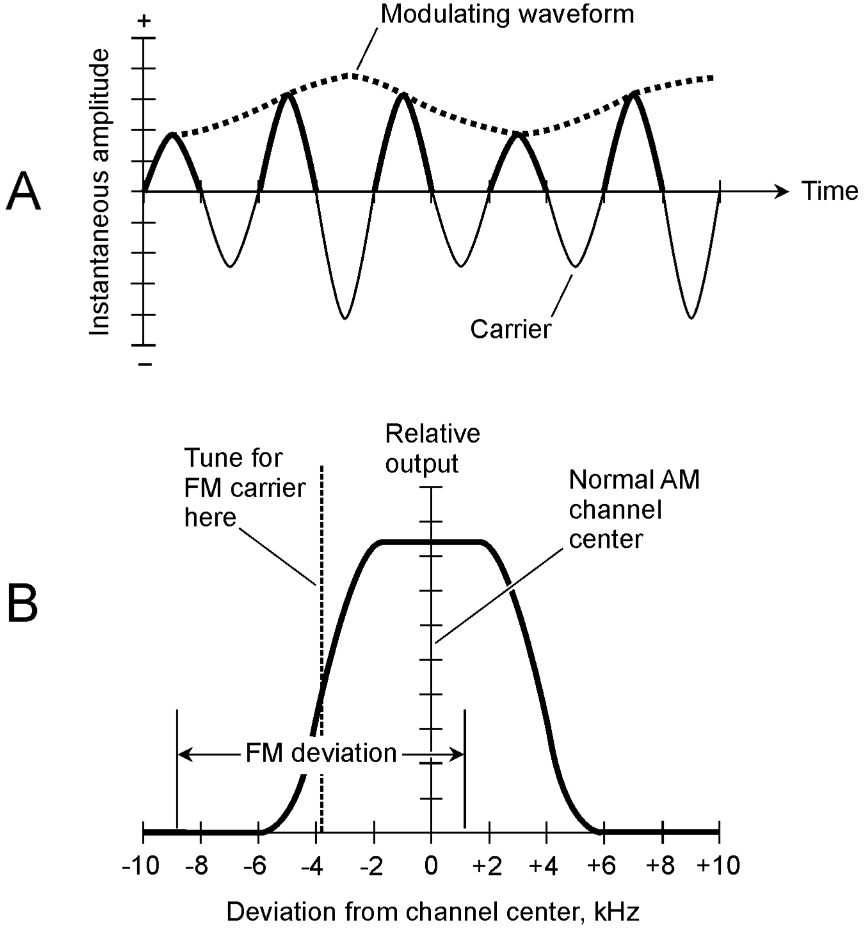
A
+
Modulating waveform
Instantaneous amplitude
Time
Carrier
−
B
Tune for
FM carrier
here
Relative
output
Normal AM
channel
center
FM deviation
-10
-8
-6
-4
-2
0
+2
+4
+6
+8
+10
Deviation from channel center, kHz

25-14 At A, envelope detection of AM, shown in the time domain. At B, slope detection of FM, shown in the frequency domain.

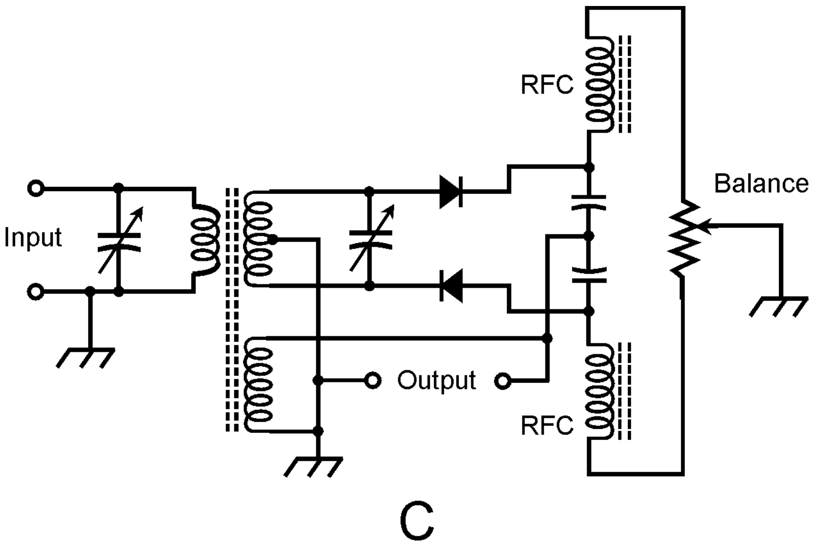

25-14 At C, a ratio detector circuit for demodulating FM signals.

rier frequency; the slower fluctuation is a duplication of the modulating data. The carrier pulsations are smoothed out by passing the output through a capacitor large enough to hold the charge for one carrier current cycle, but not so large that it smoothes out the cycles of the modulating signal. This scheme is known as *envelope detection.*

Detection of CW

For detection of CW signals, it is necessary to inject a signal into the receiver a few hundred hertz from the carrier. The injected signal is produced by a tunable *beat-frequency oscillator* (BFO). The BFO signal and the desired CW signal are mixed, or *heterodyned*, to produce audio output at the difference frequency. The BFO is tuned to a frequency that results in a comfortable listening pitch, usually 500 to 1000 Hz. This is called *heterodyne detection.*

Detection of FSK

The detection of FSK signals can be done using the same method as CW detection. The carrier beats against the BFO in the mixer, producing an audio tone that alternates between two different pitches.

With FSK, the BFO frequency is set a few hundred hertz above or below both of the carrier frequencies—that is, of both the *mark frequency* and the *space frequency.* The *frequency offset*, or difference between the BFO and signal frequencies, determines the audio output frequencies, and must

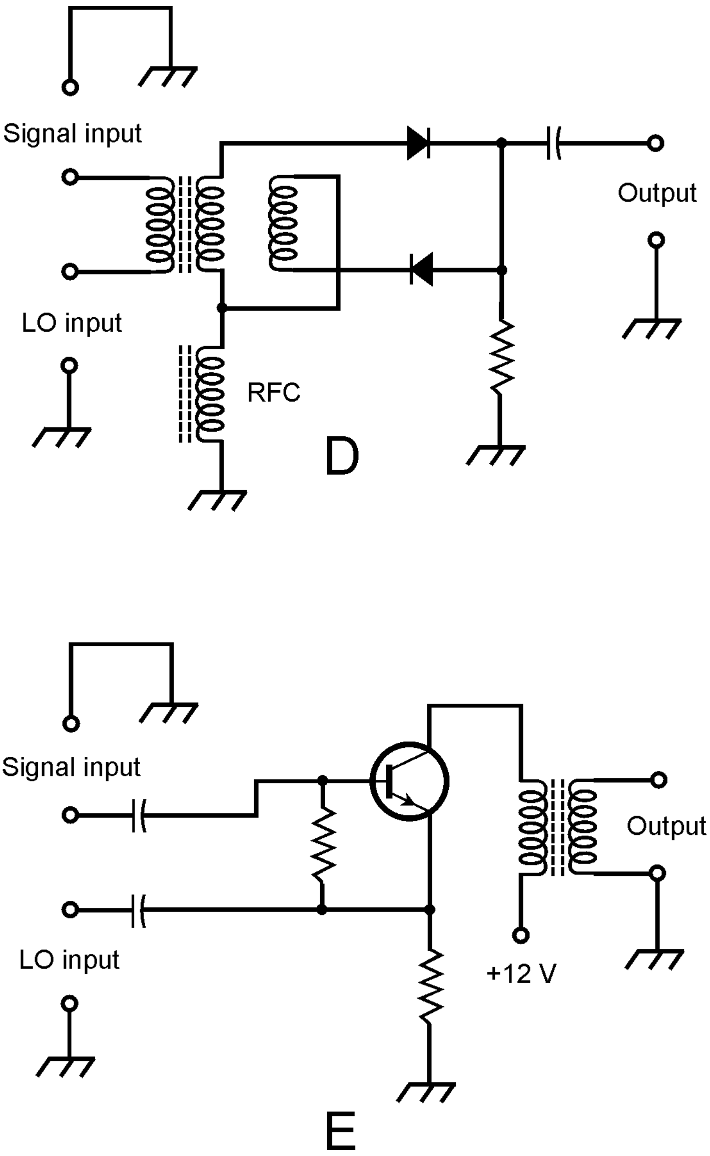
Signal input
LO input
RFC
Output
D
Signal input
LO input
+12 V
Output
E

25-14 At D, a product detector using diodes. At E, a product detector using an NPN bipolar transistor biased as a class B amplifier. Both of these circuits can also be used as signal mixers.

be set so certain standard tone pitches result (such as 2125 Hz and 2295 Hz in the case of 170-Hz shift). Unlike the situation with CW reception, there is little tolerance for BFO adjustment variation or error.

Detection of FM

Frequency-modulated (FM) signals can be detected in various ways. These methods also work for phase modulation.

Slope detection: An AM receiver can detect FM in a crude manner by setting the receiver frequency near, but not on, the FM unmodulated-carrier frequency. An AM receiver has a filter with a passband of a few kilohertz, having a selectivity curve such as that shown in Fig. 25-14B. If the FM unmodulated-carrier frequency is near either edge, or *skirt*, of the filter response, frequency variations in the incoming signal cause it to swing in and out of the receiver passband. This causes the instantaneous receiver output to vary. The relationship between the instantaneous FM deviation and the instantaneous output amplitude is not linear, however, because the skirt of the passband is not a straight line, as is apparent in the figure. The result is an unnatural-sounding received signal.

PLL: If an FM signal is injected into a PLL, the loop produces an error voltage that is a duplicate of the modulating waveform. A *limiter*, which keeps the signal amplitude from varying, can be placed ahead of the PLL so the receiver does not respond to AM. Weak signals tend to abruptly appear and disappear, rather than fading, in an FM receiver that employs limiting.

Discriminator: This type of FM detector produces an output voltage that depends on the instantaneous signal frequency. When the signal is at the center of the passband, the output voltage is zero. If the frequency falls below center, the output voltage becomes positive. If the frequency rises above center, the output becomes negative. The relationship between the instantaneous FM deviation and the instantaneous output amplitude is linear, so the output is a faithful reproduction of the incoming signal data. A discriminator is sensitive to amplitude variations, but this can be overcome by a limiter.

Ratio detector: This type of FM detector is a discriminator with a built-in limiter. The original design was developed by RCA (Radio Corporation of America), and is used in high-fidelity receivers and in the audio portions of TV receivers. A simple ratio detector circuit is shown in Fig. 25-14C. The balance potentiometer should be adjusted for the best received signal quality.

Detection of SSB

For reception of SSB signals, a *product detector* is preferred, although a direct-conversion receiver can also do the job. A product detector also works well for the reception of CW and FSK. The incoming signal combines with the output of an unmodulated LO, producing audio or video. Product detection is done at a single frequency, rather than at a variable frequency as in direct-conversion reception. The single, constant frequency is obtained by mixing the incoming signal with the output of the LO.

Two product-detector circuits, which are also representative of the mixers used in superhet receivers, are shown in Fig. 25-14D and E. At D, diodes are used; there is no amplification. At E, a bipolar transistor is employed; this circuit provides some gain. The essential characteristic of either circuit is the nonlinearity of the semiconductor devices. This generates the sum and difference frequency signals that result in audio or video output.

Audio Stages

Enhanced selectivity can be obtained by tailoring the frequency response in the AF amplifier stages following the detector, in addition to the RF selectivity provided in the IF amplifier stages preceding the detector.

Filtering

A voice signal requires a band of frequencies ranging from about 300 Hz to 3000 Hz. An *audio bandpass filter*, with a passband of 300 Hz to 3000 Hz, can improve the quality of reception with some voice receivers. An ideal voice audio filter has little or no attenuation within the passband range, and high attenuation outside the range, with a near-rectangular response.

A CW or FSK signal requires only a few hundred hertz of bandwidth to be clearly read. Audio CW filters can narrow the response bandwidth to 100 Hz or less, but passbands narrower than about 100 Hz produce *ringing*, degrading the quality of reception. With FSK, the bandwidth of the filter must be at least as large as the difference (shift) between mark and space, but it need not, and should not, be much greater.

An *audio notch filter* is a *band-rejection filter* with a sharp, narrow response. An interfering carrier that produces a tone of constant frequency in the receiver output can be greatly attenuated with this type of filter. Audio notch filters are tunable from at least 300 Hz to 3000 Hz. Some sophisticated AF notch filters can tune themselves automatically. When an interfering carrier appears and remains for a few tenths of a second, the notch is activated and centers itself on the audio frequency of the detected offending signal.

Squelching

A *squelch* silences a receiver when no signal is present, allowing reception of signals when they appear. Most FM communications receivers use squelching systems. The squelch is normally closed, allowing no audio output, when no signal is present. The squelch opens, allowing everything to be heard, if the signal amplitude exceeds the *squelch threshold*, which can be adjusted by the operator.

In some systems, the squelch does not open unless the signal has certain characteristics. This is known as *selective squelching*. The most common way to achieve this is the use of subaudible (below 300 Hz) tone generators, or AF tone-burst generators. The squelch opens only when an incoming signal is modulated by a tone, or sequence of tones, having the proper characteristics. This can prevent unauthorized transmissions from accessing repeaters or being picked up by receivers.

Television Reception

A television (TV) receiver has a tunable front end, an oscillator and mixer, a set of IF amplifiers, a video demodulator, an audio demodulator and amplifier chain, a picture CRT or display with associated peripheral circuitry, and a loudspeaker.

Fast-Scan TV

Figure 25-15 is a block diagram of a receiver for conventional analog FSTV. In the United States, conventional FSTV broadcasts are made on 68 different channels numbered from 2 through 69. Each channel is 6 MHz wide, including video and audio information. Channels 2 through 13 comprise the *VHF TV broadcast channels*. Channels 14 through 69 are the *UHF TV broadcast channels*.

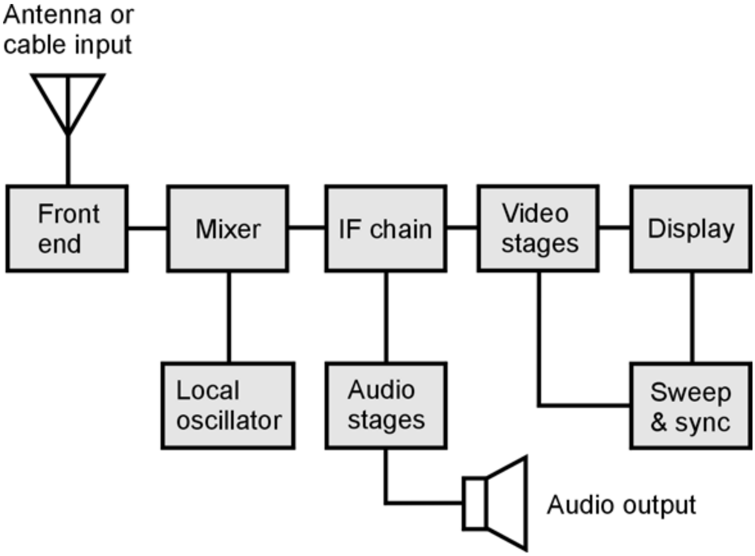

25-15 Block diagram of a conventional FSTV receiver.

In digital cable television, there are more channels. The number of possible channels is virtually unlimited, because signals are not transmitted over the air and do not consume EM spectrum space.

Slow-Scan TV

A *slow-scan television* (SSTV) communications station needs a transceiver with SSB capability, a standard TV set or personal computer, a video camera, and a *scan converter* that translates between the SSTV signal and either FSTV imagery or computer video data. The scan converter consists of two data converters (one for receiving and the other for transmitting), some digital *memory*, a *tone generator*, and a TV detector. Scan converters are commercially available. Computers can be programmed to perform this function. Some amateur radio operators build their own scan converters.

Specialized Wireless Modes

Some less common wireless communications techniques are effective under certain circumstances.

Dual-Diversity Reception

A *dual-diversity receiver* can reduce the fading that occurs in radio reception at high frequencies (approximately 3 to 30 MHz) when signals are propagated by means of the ionosphere. Two receivers are used. Both are tuned to the same signal, but they employ separate antennas, spaced several wavelengths apart. The outputs of the receiver detectors are fed into a common audio amplifier, as shown in Fig. 25-16.

Dual-diversity tuning is critical, and the equipment is expensive. In some installations, three or more antennas and receivers are employed. This provides superior immunity to fading, but it compounds the tuning difficulty and further increases the expense.

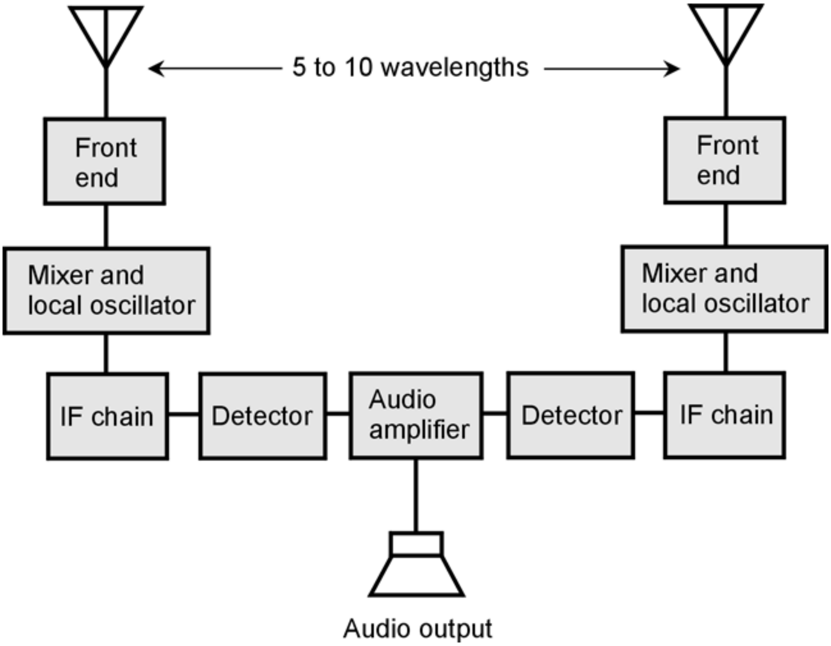

25-16 Block diagram of a dual-diversity radio receiver system.

Synchronized Communications

Digital signals require less bandwidth than analog signals to convey a given amount of information per unit time. *Synchronized communications* refers to a specialized digital mode, in which the transmitter and receiver operate from a common time standard to optimize the amount of data that can be sent in a communications channel or band.

In synchronized digital communications, also called *coherent communications*, the receiver and transmitter operate in lock-step. The receiver evaluates each transmitted binary digit, or *bit*, for a block of time lasting for the specified duration of a single bit. This makes it possible to use a receiving filter having extremely narrow bandwidth. The synchronization requires the use of an external frequency/time standard. The broadcasts of standard time-and-frequency radio stations such as WWV or WWVH can be used for this purpose. Frequency dividers are employed to obtain the necessary synchronizing frequencies. A tone or pulse is generated in the receiver output for a particular bit if, but only if, the average signal voltage exceeds a certain value over the duration of that bit. False signals, such as can be caused by filter ringing, sferics, or ignition noise, are generally ignored, because they rarely produce sufficient average bit voltage.

Experiments with synchronized communications have shown that the improvement in S/N ratio, compared with nonsynchronized systems, is several decibels at low to moderate data speeds. Further improvement can be obtained by the use of DSP.

How DSP Can Improve Reception

In DSP with analog modes such as SSB or SSTV, the signals are first changed into digital form by A/D conversion. Then the digital data is cleaned up so the pulse timing and amplitude adhere strictly to the protocol (standards) for the type of digital data being used. Finally, the digital signal is changed back to the original voice or video by D/A conversion.

Digital signal processing can extend the range of a wireless communications circuit, because it allows reception under worse conditions than would be possible without it. Digital signal processing also improves the quality of marginal signals, so that the receiving equipment or operator makes fewer errors. In circuits that use only digital modes, A/D and D/A conversion are irrelevant, but DSP can still be used to clean up the signal. This improves the accuracy of the system, and also makes it possible to copy data over and over many times (that is, to produce multigeneration duplicates).

The DSP circuit minimizes noise and interference in a received digital signal as shown in Fig. 25-17. A hypothetical signal before DSP is shown at the top; the signal after processing is shown at the bottom. If the incoming signal is above a certain level for an interval of time, the DSP output is *high* (also called *logic 1*). If the level is below the critical point for a time interval, then the output is *low* (also called *logic 0*).

Multiplexing

Signals in a communications channel or band can be intertwined, or *multiplexed*, in various ways. The most common methods are *frequency division multiplexing* (FDM) and *time division multiplexing* (TDM). In FDM, the channel is broken down into subchannels. The carrier frequencies of the signals are spaced so they do not overlap. Each signal is independent of the others. In TDM, signals are broken into segments by time, and then the segments are transferred in a rotating sequence. The receiver must be synchronized with the transmitter by means of a time standard such as WWV. Multiplexing requires an *encoder* that combines or intertwines the signals in the transmitter, and a *decoder* that separates or untangles the signals in the receiver.

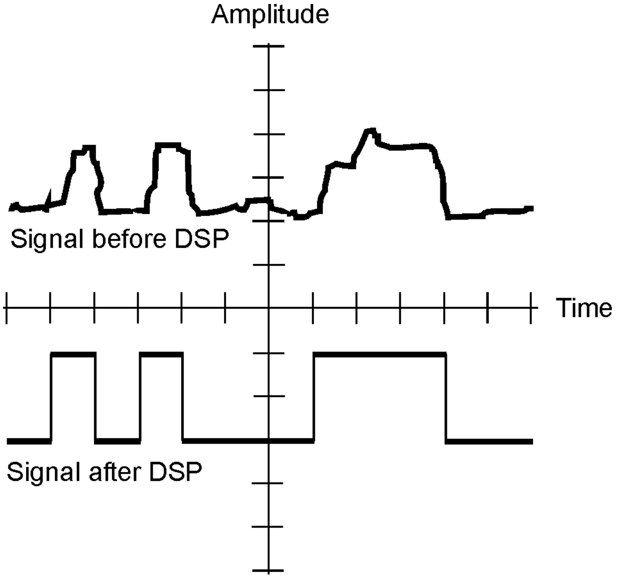

25-17 Digital signal processing can clean up a signal, improving reception.

Spread Spectrum

In *spread-spectrum communications*, the main carrier frequency is rapidly varied independently of signal modulation, and the receiver is programmed to follow. As a result, the probability of *catastrophic interference*, in which one strong interfering signal can obliterate the desired signal, is near zero. It is difficult for unauthorized people to eavesdrop on a spread-spectrum communications link unless they gain access to the *sequencing code*, also known as the *frequency spreading function.* Such a function can be complex, and can be kept secret. If the transmitting and receiving operator do not divulge the function to anyone, and if they do not tell anyone about the existence of their contact, then no one else on the band will know the contact is taking place.

During a spread-spectrum contact between a given transmitter and receiver, the operating frequency can fluctuate over a range of kilohertz, megahertz, or tens of megahertz. As a band becomes occupied with an increasing number of spread-spectrum signals, the overall noise level in the band appears to increase. Therefore, there is a practical limit to the number of spread-spectrum contacts that a band can handle. This limit is roughly the same as it would be if all the signals were constant in frequency, and had their own discrete channels.

A common method of generating spread spectrum is *frequency hopping.* The transmitter has a list of channels that it follows in a certain order. The receiver must be programmed with this same list, in the same order, and must be synchronized with the transmitter. The *dwell time* is the interval at which the frequency changes occur, which is the same as the length of time that the signal remains on any given frequency. The dwell time should be short enough so that a signal will not be noticed, and not cause interference, on any frequency. There are numerous *dwell frequencies*, so the signal energy is diluted to the extent that, if someone tunes to any particular frequency in the sequence, the signal is not noticeable.

Another way to get spread spectrum, called *frequency sweeping*, is to frequency-modulate the main transmitted carrier with a waveform that guides it up and down over the assigned band. This FM is independent of signal intelligence. A receiver can intercept the signal if, but only if, its tuning varies according to the same waveform, over the same band, at the same frequency, and in the same phase as that of the transmitter.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. A radio wave has a frequency of 1.55 MHz. The highest modulating frequency that can be used effectively is about
   (a) 1.55 kHz.
   (b) 15.5 kHz.
   (c) 155 kHz.
   (d) 1.55 MHz.
2. The reflected wave
   (a) arrives in phase with the direct wave.
   (b) arrives out of phase with the direct wave.
   (c) arrives in a variable phase compared with the direct wave.
   (d) is always horizontally polarized.

3. An advantage of FSK over on/off keying is the fact that FSK
   (a) offers better frequency stability.
   (b) can provide faster data rates.
   (c) reduces the number of receiving errors.
   (d) Forget it! On-off keying is just as good as FSK.

4. The highest layer of the ionosphere is
   (a) the D layer.
   (b) the E layer.
   (c) the F layer.
   (d) none of the above.

5. If an AM signal is modulated with audio having frequencies up to 5 kHz, then what is the complete signal bandwidth?
   (a) 10 kHz
   (b) 20 kHz
   (c) 50 kHz
   (d) 500 kHz

6. An LSB, suppressed-carrier signal can be demodulated by
   (a) an envelope detector.
   (b) a diode.
   (c) a ratio detector.
   (d) a product detector.

7. Which of the following modes is used to send image data over telephone lines?
   (a) On/off keying
   (b) Fax
   (c) AM
   (d) Product detection

8. The S/N ratio is a measure of
   (a) the sensitivity of a receiver.
   (b) the selectivity of a receiver.
   (c) the dynamic range of a receiver.
   (d) the efficiency of a transmitter.

9. Suppose an SSB suppressed carrier is at 14.335 MHz, and audio data is contained in a band from 14.335 to 14.338 MHz. What is this mode?
   (a) DSB
   (b) LSB
   (c) USB
   (d) AFSK

10. A receiver that responds to a desired signal, but not to another signal very close by in frequency, has good
    (a) sensitivity.
    (b) noise figure.
    (c) dynamic range.
    (d) selectivity.

11. Fill in the blank in the following sentence to make it true: "The deviation of a narrowband voice FM signal normally extends up to ____ either side of the unmodulated-carrier frequency."
    (a) 3 kHz
    (b) 5 kHz
    (c) 10 kHz
    (d) 3 MHz

12. An FM detector with built-in limiting is
    (a) a ratio detector.
    (b) a discriminator.
    (c) an envelope detector.
    (d) a product detector.

13. In which mode of PM does the peak power level of the pulses vary?
    (a) PAM
    (b) PDM
    (c) PIM
    (d) PFM

14. A continuously variable signal (such as music audio) can be recovered from a signal having only a few discrete levels or states by means of
    (a) a ratio detector.
    (b) a D/A converter.
    (c) a product detector.
    (d) an envelope detector.

15. In which mode are signals intertwined in the time domain at the transmitter, and then separated again at the receiver?
    (a) FDM
    (b) AFSK
    (c) PCM
    (d) None of the above

16. Which of the following modes can be demodulated with an envelope detector?
    (a) AM
    (b) CW

(c) FSK
(d) USB

17. The bandwidth of a fax signal is kept narrow by
    (a) sending the data at a slow rate of speed.
    (b) maximizing the number of digital states.
    (c) optimizing the range of colors sent.
    (d) using pulse modulation.

18. The dynamic range in a superhet is largely influenced by the performance of the
    (a) local oscillator.
    (b) product detector.
    (c) front end.
    (d) selectivity in the IF chain.

19. Frequency sweeping can be used to get a transmitter to produce
    (a) spread-spectrum signals.
    (b) time division multiplexed signals.
    (c) narrowband AM signals.
    (d) double sideband, suppressed carrier signals.

20. Fill in the blank to make the following sentence true: "The reception of ____ can be improved by the use of DSP."
    (a) SSB signals
    (b) SSTV signals
    (c) synchronized communications signals
    (d) any of the above

26
CHAPTER

Digital Basics

A SIGNAL IS *DIGITAL* WHEN IT CAN ATTAIN ONLY SPECIFIC LEVELS. THIS IS IN CONTRAST TO *ANALOG* signals or quantities that vary over a continuous range. A simple analog waveform is shown at Fig. 26-1A; note that the amplitude varies smoothly over time. Figure 26-1B is an example of a digital approximation of the same signal.

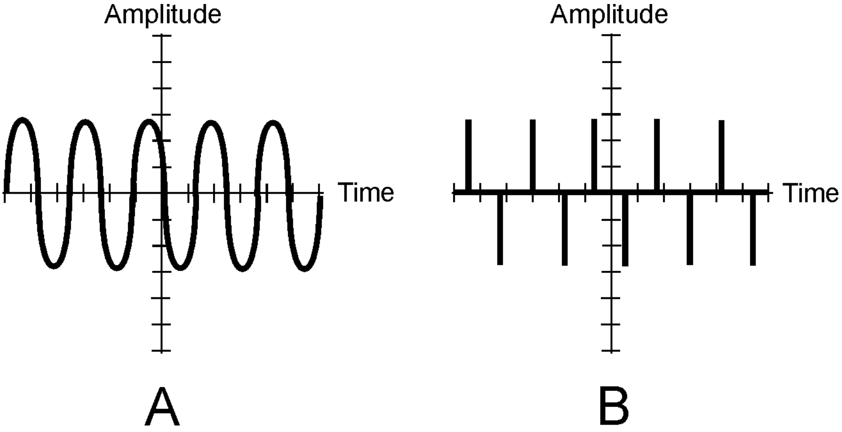

26-1 An analog wave (A), and a digital rendition of this wave (B).

Numbering Systems

People are used to dealing with the *decimal number system*, which has 10 different digits. But machines use schemes that have some power of 2 digits, such as 2 (2^1), 4 (2^2), 8 (2^3), 16 (2^4), 32 (2^5), 64 (2^6), and so on.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

Decimal

The *decimal number system* is also called *base 10* or *radix 10.* The set of possible digits is {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}. The first digit to the left of the radix or "decimal" point is multiplied by 10^0, or 1. The next digit to the left is multiplied by 10^1, or 10. The power of 10 increases as you move farther to the left. The first digit to the right of the decimal point is multiplied by a factor of 10^{-1}, or $1/10$. The next digit to the right is multiplied by 10^{-2}, or $1/100$. This continues as you go farther to the right. Once the process of multiplying each digit is completed, the resulting values are added up. For example:

$$\begin{aligned} & 2 \times 10^3 \\ + & 7 \times 10^2 \\ + & 0 \times 10^1 \\ + & 4 \times 10^0 \\ + & 5 \times 10^{-1} \\ + & 3 \times 10^{-2} \\ + & 8 \times 10^{-3} \\ + & 1 \times 10^{-4} \\ + & 6 \times 10^{-5} \\ = & 2704.53816 \end{aligned}$$

Binary

The *binary number system* is a method of expressing numbers using only the digits 0 and 1. It is sometimes called *base 2* or *radix 2.* The digit immediately to the left of the radix point is the "ones" digit. The next digit to the left is a "twos" digit; after that comes the "fours" digit. Moving farther to the left, the digits represent 8, 16, 32, 64, and so on, doubling every time. To the right of the radix point, the value of each digit is cut in half again and again, that is, $1/2$, $1/4$, $1/8$, $1/16$, $1/32$, $1/64$, and so on.

Consider the decimal number 94. In the binary number system, this number is written as 1011110. It breaks down as follows:

$$\begin{aligned} & 0 \times 2^0 \\ + & 1 \times 2^1 \\ + & 1 \times 2^2 \\ + & 1 \times 2^3 \\ + & 1 \times 2^4 \\ + & 0 \times 2^5 \\ + & 1 \times 2^6 \\ = & 1011110 \end{aligned}$$

When you work with a computer or calculator, you give it a decimal number that is converted into binary form. The computer or calculator does its operations entirely using the digits 0 and 1. When the process is complete, the machine converts the result back into decimal form for display.

In a communications system, binary numbers can represent alphanumeric characters, shades of color, frequencies of sound, and other variable quantities.

Octal

Another scheme, sometimes used in computer programming, is the *octal number system*, so named because it has eight symbols (according to our way of thinking), or 2^3. Every digit is an element of the set {0, 1, 2, 3, 4, 5, 6, 7}. This system is also known as *base 8* or *radix 8.*

Hexadecimal

Another system used in computer work is the *hexadecimal number system.* It has 16 (2^4) symbols. These digits are the usual 0 through 9 plus six more, represented by A through F, the first six letters of the alphabet. The digit set is {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F}. This system is sometimes called *base 16* or *radix 16.*

Logic

Logic refers to the reasoning used by electronic machines. The term is also used in reference to the circuits that make up digital devices and systems.

Boolean Algebra

Boolean algebra is a system of mathematical logic using the numbers 0 and 1 with the operations AND (multiplication), OR (addition), and NOT (negation). Combinations of these operations are NAND (NOT AND) and NOR (NOT OR). This peculiar form of mathematical logic, which gets its name from the nineteenth-century British mathematician George Boole, is used in the design of digital logic circuits.

Symbology

In Boolean algebra, the AND operation, also called *logical conjunction*, is written using an asterisk (*), a multiplication symbol (×), or by running two characters together, for example, X*Y.

The NOT operation, also called *logical inversion*, is denoted by placing a tilde (~) over the quantity, as a minus sign (–) or dash (-) followed by the quantity, as a “lazy inverted L” (¬) followed by the quantity, or as the quantity followed by an accent or “prime sign” (′). An example is –X.

The OR operation, also called *logical disjunction*, is written using a plus sign (+), for example, X + Y.

The foregoing are the symbols used by engineers. Table 26-1A shows the values of these functions, where 0 indicates “falsity” and 1 indicates “truth.”

In mathematics and philosophy courses involving logic, you may see other symbols used for conjunction and disjunction. The AND operation in some texts is denoted by a detached arrow-

Table 26-1A. Boolean operations.

X	Y	–X	X*Y	X+Y
0	0	1	0	0
0	1	1	0	1
1	0	0	0	1
1	1	0	1	1

Table 26-1B. Common theorems in Boolean algebra.

Theorem (logic equation)	What it's called
$X+0 = X$	OR identity
$X*1 = X$	AND identity
$X+1 = 1$	
$X*0 = 0$	
$X+X = X$	
$X*X = X$	
$-(-X) = X$	Double negation
$X+(-X) = 1$	
$X*(-X) = 0$	Contradiction
$X+Y = Y+X$	Commutativity of OR
$X*Y = Y*X$	Commutativity of AND
$X+(X*Y) = X$	
$X*(-Y)+Y = X+Y$	
$X+Y+Z = (X+Y)+Z = X+(Y+Z)$	Associativity of OR
$X*Y*Z = (X*Y)*Z = X*(Y*Z)$	Associativity of AND
$X*(Y+Z) = (X*Y)+(X*Z)$	Distributivity
$-(X+Y) = (-X)*(-Y)$	DeMorgan's Theorem
$-(X*Y) = (-X)+(-Y)$	DeMorgan's Theorem

head pointing up (∧) or by an ampersand (&), and the OR operation is denoted by a detached arrowhead pointing down (∨).

Theorems

Table 26-1B shows several logic equations. Such facts are called *theorems.* Statements on either side of the equals sign in each case are *logically equivalent.*

When two statements are logically equivalent, it means that one is true *if and only if* (*iff*) the other is true. For example, the statement X = Y means that X implies Y, and also that Y implies X. Logical equivalence is sometimes symbolized by a double arrow with one or two shafts (↔ or ⇔).

Boolean theorems are used to analyze and simplify complicated logic functions. This makes it possible to build a circuit to perform a specific digital function, using the smallest possible number of logic switches.

Digital Circuits

All binary digital devices and systems employ high-speed electronic switches that perform Boolean operations. These switches are called *logic gates.* By combining logic gates, sophisticated digital systems can be built up. Even the most advanced computers are, at the basic level, comprised of logic gates.

Positive and Negative Logic

Usually, the binary digit 1 stands for "true" and is represented by a voltage of about +5 V. The binary digit 0 stands for "false" and is represented by about 0 V. This is *positive logic.* There are other logic forms, the most common of which is *negative logic.* In a common form of negative logic, the digit 1 (the logic high state) is represented by about 0 V, and the digit 0 (the logic low state) is represented by about +5 V. The remainder of this chapter deals with positive logic.

Basic Gates

An *inverter* or *NOT gate* has one input and one output. It reverses the state of the input. An *OR gate* can have two or more inputs. If both, or all, of the inputs are 0 (low), then the output is 0. If any of the inputs is 1 (high), then the output is 1. An *AND gate* can have two or more inputs. If both, or all, of the inputs are 1, then the output is 1. Otherwise the output is 0.

Other Gates

Sometimes an inverter and an OR gate are combined. This produces a *NOR gate.* If an inverter and an AND gate are combined, the result is a *NAND gate.*

An *exclusive OR gate*, also called an *XOR gate*, has two inputs and one output. If the two inputs are the same (either both 1 or both 0), then the output is 0. If the two inputs are different, then the output is 1.

The functions of logic gates are summarized in Table 26-2. Their schematic symbols are shown in Fig. 26-2.

Black Boxes

Logic gates can be combined to form circuits with many inputs and outputs. When two or more logic gates are combined, the outputs are always specific *logical functions* of the inputs. A complex combination of logic gates is sometimes called a *black box.*

The functions of a black box can always be determined using Boolean algebra, if the gates inside, and the way they are interconnected, is known. Conversely, if a certain complex logical function is needed for an application, a black box can be designed to perform that function by using Boolean algebra to break the function down into components of NOT, OR, AND, NOR, NAND, and XOR.

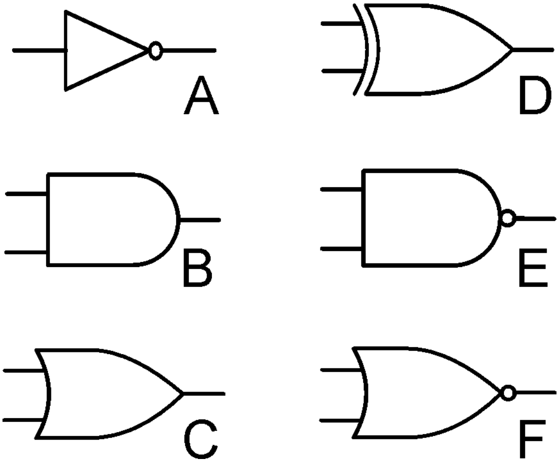

26-2 An inverter or NOT gate (A), an AND gate (B), an OR gate (C), an XOR gate (D), a NAND gate (E), and a NOR gate (F).

Table 26-2. Logic gates and their characteristics.

Gate type	Number of inputs	Remarks
NOT	1	Changes state of input.
OR	2 or more	Output high if any inputs are high. Output low if all inputs are low.
AND	2 or more	Output low if any inputs are low. Output high if all inputs are high.
NOR	2 or more	Output low if any inputs are high. Output high if all inputs are low.
NAND	2 or more	Output high if any inputs are low. Output low if all inputs are high.
XOR	2	Output high if inputs differ. Output low if inputs are the same.

Forms of Binary Data

In communications, binary (two-level) data is less susceptible to noise and other interference than analog or multilevel digital data. There are several forms.

- *Morse code* is the oldest binary means of sending and receiving messages. It is a binary code because it has only two possible states: on (key-down) and off (key-up). It is used mainly by amateur radio operators in their hobby activities. A "human ear/brain machine," scrutinizing a Morse code signal, is an amazingly effective digital communications receiver.
- *Baudot*, also called the *Murray code*, is a five-unit digital code not widely used by today's digital equipment, except in some radioteletype communications.
- *ASCII* (American National Standard Code for Information Interchange) is a seven-unit code for the transmission of text and some programs. Letters, numerals, symbols, and control operations are represented. ASCII is designed for computers. There are 2^7, or 128, possible representations. Both upper- and lowercase letters can be represented, along with numerals and certain symbols.

Flip-flops

A *flip-flop* is also known as a *sequential logic gate.* In a sequential gate, the output state depends on both the inputs and the outputs. A flip-flop has two states, called *set* and *reset.* Usually, the set state is logic 1 (high), and the reset state is logic 0 (low). Here are some common types.

- *R-S flip-flop* inputs are labeled R (reset) and S (set). The outputs are Q and –Q. (Often, rather than –Q, you will see Q′, or perhaps Q with a line over it.) The outputs are always in logically opposite states. The symbol for an R-S flip-flop, also known as an *asynchronous flip-flop*, is shown in Fig. 26-3A. The *truth table* (a specialized form of table denoting logic functions) for an R-S flip-flop is at Table 26-3A.
- *Synchronous flip-flop* states change when triggered by the signal from a *clock*. In *static triggering*, the outputs change state only when the clock signal is either high or low. This type of circuit is sometimes called a *gated flip-flop*. In *positive-edge triggering*, the outputs change state at the instant the clock pulse is positive-going. The term *edge triggering* derives from the fact that the abrupt rise or fall of a pulse looks like the edge of a cliff (Fig. 26-3B). In *negative-edge triggering*, the outputs change state at the instant the clock pulse is negative-going.
- *Master/slave* (*M/S*) *flip-flop* inputs are stored before the outputs are allowed to change state. This device essentially consists of two R-S flip-flops in series. The first flip-flop is called the *master*, and the second is called the *slave*. The master flip-flop functions when the clock output is high, and the slave acts during the next low portion of the clock output. This time delay prevents confusion between the input and output.
- *J-K flip-flop* operation is similar to that of an R-S flip-flop, except that the J-K has a predictable output when the inputs are both 1. Table 26-3B shows the input and output states for this type of flip-flop. The output changes only when a triggering pulse is received. The symbol for a J-K flip-flop is shown in Fig. 26-3C.
- *R-S-T flip-flop* operation is similar to that of an R-S flip-flop, except that a high pulse at the T input causes the circuit to change state.
- The *T flip-flop* has only one input. Each time a high pulse appears at the T input, the output state is reversed.

Clocks

In electronics, the term *clock* refers to a circuit that generates pulses at high speed and at precise intervals. It sets the tempo for the operation of digital devices. In a computer, the clock acts like a metronome for the *microprocessor*. Clock speeds are measured and expressed in hertz (Hz), kilohertz (kHz), megahertz (MHz), or gigahertz (GHz).

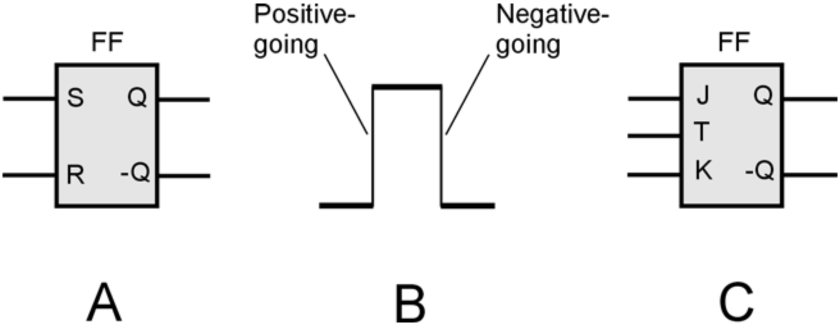

26-3 At A, the symbol for an R-S flip-flop. At B, pulse edges are either negative-going or positive-going. At C, the symbol for a J-K flip-flop.

Table 26-3. Flip-flop states.

A:
R-S Flip-flop

R	S	Q	–Q
0	0	Q	–Q
0	1	1	0
1	0	0	1
1	1	?	?

B:
J-K Flip-flop

J	K	Q	–Q
0	0	Q	–Q
0	1	1	0
1	0	0	1
1	1	–Q	Q

Counters

A *counter* consists of a set of flip-flops or equivalent circuits. Each time a pulse is received, the binary number stored by the counter increases by 1. A *frequency counter* measures the frequency of a wave by tallying the cycles in a given interval of time. The circuit consists of a *gate*, which begins and ends each counting cycle at defined intervals. The accuracy is a function of the length of the *gate time;* the longer the time base, the better the accuracy. The readout is in base-10 digital numerals.

Binary Digital Communications

The use of binary data provides excellent communications accuracy and efficiency. If multilevel signaling is required, then all the levels can be represented by groups of binary digits. A group of 3 binary digits, for example, can represent 2^3, or 8, levels. A group of 4 binary digits can represent 2^4, or 16, levels. The term *binary digit* is commonly contracted to *bit.*

Bits, Bytes, and Baud

A bit is almost always represented by either 0 or 1. A group of 8 bits is a called an *octet*, and in many systems this also corresponds to a unit called a *byte.* Large quantities of data can be expressed either according to powers of 2, or according to powers of 10. This can cause some confusion and gives rise to endless debates over semantics.

One *kilobit* (kb) is equal to 1000 bits. A *megabit* (Mb) is 1000 kilobits, or 1,000,000 bits. A *gigabit* (Gb) is 1000 megabits, or 1,000,000,000 bits. When data is expressed in bits, powers of 10 are used to define large quantities. If you hear about a modem that operates at 56 kbps, it means 56,000 *bits per second* (bps). Bits, kilobits, megabits, and gigabits per second (bps, kbps, Mbps, and Gbps) are commonly used to express data in communications.

Data quantity in storage and memory is usually specified in *kilobytes* (units of 2^{10} = 1,024 bytes), *megabytes* (units of 2^{20} = 1,048,576 bytes), and *gigabytes* (units of 2^{30} = 1,073,741,824 bytes). The abbreviations for these units are KB, MB, and GB, respectively. Note that the uppercase K represents 2^{10} or 1024, while the lowercase k represents 10^3 or 1000. But M and G are always uppercase, no matter whether powers of 2 or 10 are used. (Is all of this confusing to you? Don't be discouraged. It confuses almost everybody.)

Larger data units are being used as memory and storage media continue to grow. The *terabyte* (TB) is 2^{40} bytes, or 1024 GB. The *petabyte* (PB) is 2^{50} bytes, or 1024 TB. The *exabyte* (EB) is 2^{60} bytes, or 1024 PB.

The term *baud* refers to the number of times per second that a signal changes state. The units of bps and baud are not equivalent, even though people often speak of them as if they are. These days, baud (or "baud rate") is seldom used to express data speed.

When computers are linked in a *network*, each computer has a *modem* (modulator/demodulator) connecting it to the communications medium. The slowest modem determines the speed at which the machines communicate. Table 26-4 shows common data speeds and the approximate time periods required to send 1, 10, and 100 pages of double-spaced, typewritten text at each speed.

Forms of Conversion

Any analog (continuously variable) signal can be converted into a string of pulses whose amplitudes have a finite number of states, usually some power of 2. This is analog-to-digital (A/D) conversion, and its reverse is digital-to-analog (D/A) conversion, as you've already learned. The difference between analog and digital signals can be intuitively seen by examining Fig. 26-4. This is essentially a rendition of 8-level pulse code modulation (PCM), as described in Chap. 25 and illustrated in Fig. 25-8D. Imagine the curve being sampled as a train of pulses (A/D), or the pulses being smoothed into the curve (D/A).

Table 26-4. Time needed to send data at various speeds.
Abbreviations: s = second, ms = millisecond (0.001 s), μs = microsecond (0.000001 s).

A:

Speed, kbps	Time for one page	Time for 10 pages	Time for 100 pages
28.8	0.38 s	3.8 s	38 s
38.4	280 ms	2.8 s	28 s
57.6	190 ms	1.9 s	19 s
100	110 ms	1.1 ms	11 s
250	44 ms	440 ms	4.4 s
500	22 ms	220 ms	2.2 s

B:

Speed, Mbps	Time for one page	Time for 10 pages	Time for 100 pages
1.00	11 ms	110 ms	1.1 s
2.50	4.4 ms	44 ms	440 ms
10.0	1.1 ms	11 ms	110 ms
100	110 μs	1.1 ms	11 ms

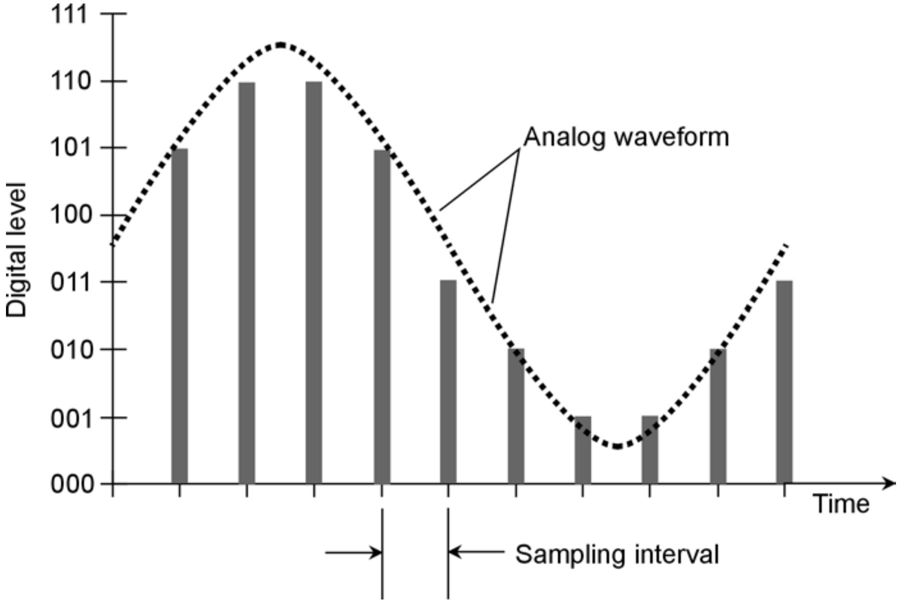

26-4 An analog waveform (dashed curve) and an 8-level digital representation (vertical bars).

Binary digital data can be sent and received one bit at a time along a single line or channel. This is *serial data transmission.* Higher data speeds can be obtained by using multiple lines or a wideband channel, sending independent sequences of bits along each line or subchannel. This is *parallel data transmission.*

Parallel-to-serial (P/S) *conversion* receives bits from multiple lines or channels, and transmits them one at a time along a single line or channel. A *buffer* stores the bits from the parallel lines or channels while they are awaiting transmission along the serial line or channel. *Serial-to-parallel* (S/P) *conversion* receives bits from a serial line or channel, and sends them in batches along several lines or channels. The output of an S/P converter cannot go any faster than the input, but the circuit is useful when it is necessary to interface between a serial-data device and a parallel-data device.

Figure 26-5 illustrates a circuit in which a P/S converter is used at the source (transmitting station), and an S/P converter is used at the destination (receiving station). In this example, the words are 8-bit bytes. However, the words could have 16, 32, 64, or even 128 bits, depending on the communications scheme.

Data Compression

Data compression is a way of maximizing the amount of digital information that can be stored in a given space, or sent in a certain period of time.

Text files can be compressed by replacing often-used words and phrases with symbols such as =, #, &, $, and @, as long as none of these symbols occurs in the uncompressed file. When the data is received, it is uncompressed by substituting the original words and phrases for the symbols.

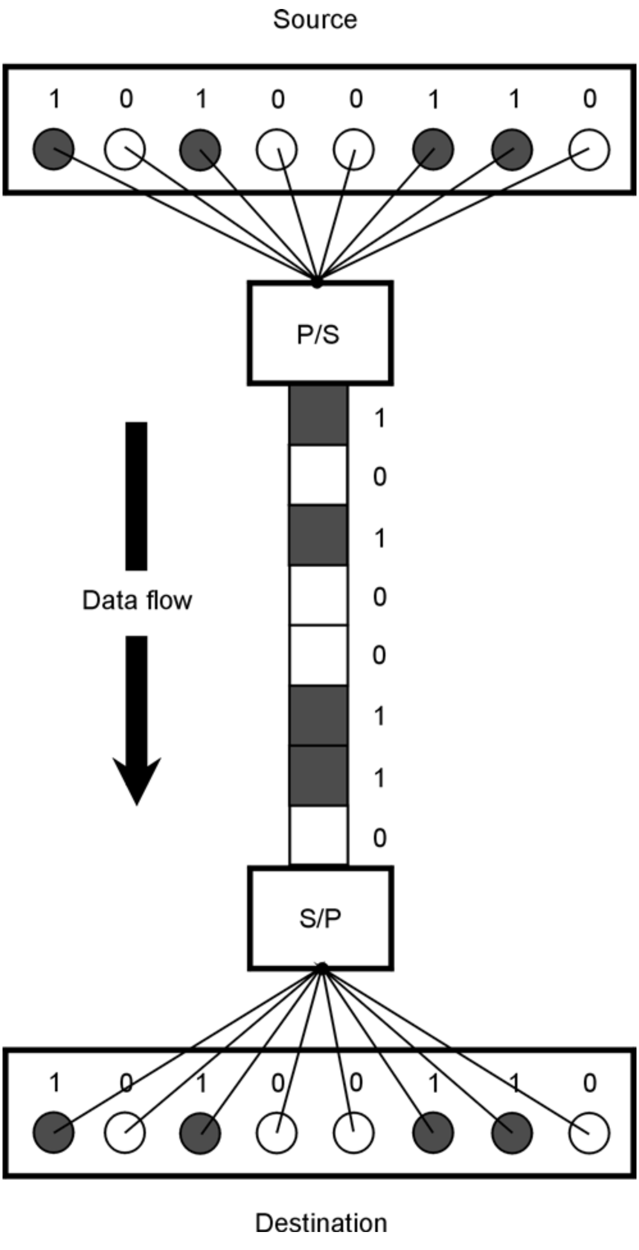

26-5 A communications circuit employing parallel-to-serial (P/S) conversion at the source, and serial-to-parallel (S/P) conversion at the destination.

Digital images can be compressed in either of two ways. In *lossless image compression*, detail is not sacrificed; only the redundant bits are eliminated. In *lossy image compression*, some detail is lost, although the loss is usually not significant.

Packet Wireless

In *packet wireless*, a computer is connected to a radio transceiver using a *terminal node controller* (TNC), which is similar to a modem. An example is shown in Fig. 26-6A. The computer has a tele-

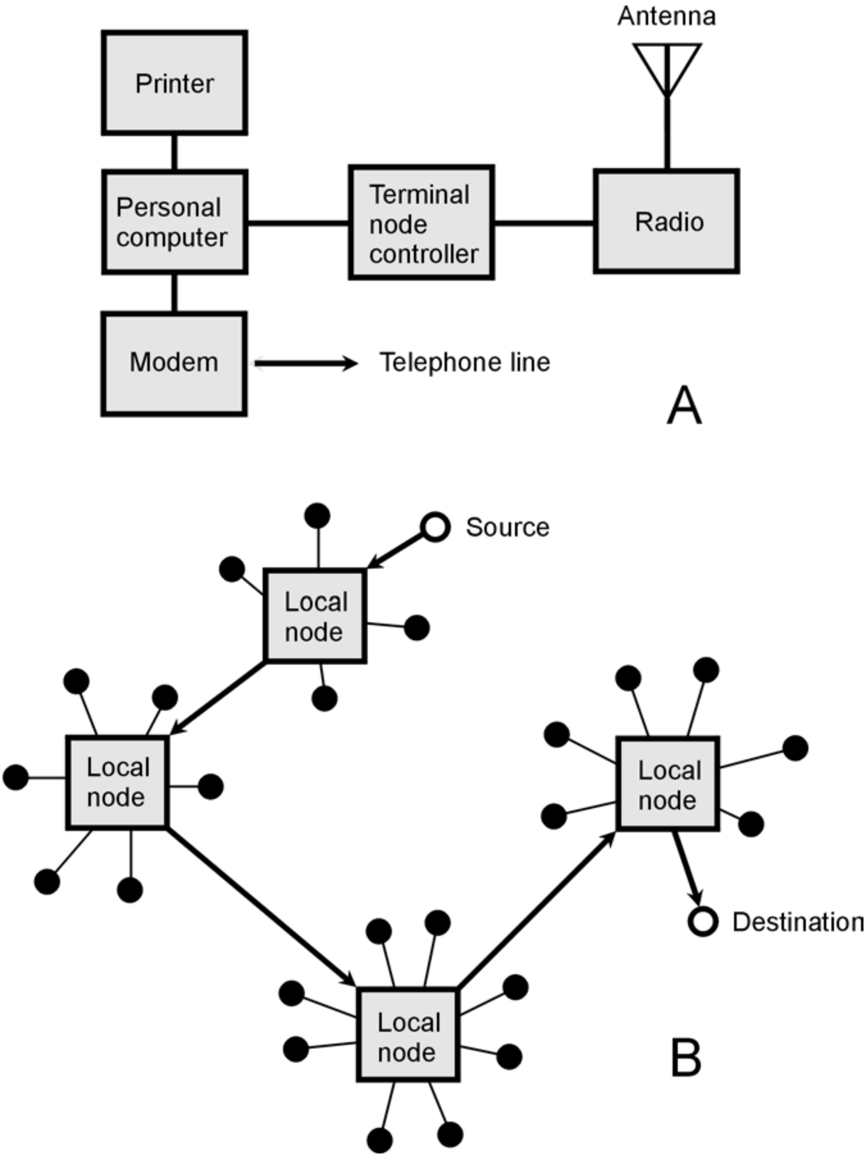
Antenna
Printer
Personal computer
Terminal node controller
Radio
Modem
Telephone line
A
Source
Local node
Local node
Local node
Local node
Destination
B

26-6 At A, a packet-wireless station. At B, passage of a packet through nodes in a wireless communications circuit.

phone modem as well as a TNC, so messages can be sent and received using conventional online services as well as radio.

Figure 26-6B shows how a packet wireless message is routed. Black dots represent *subscribers.* Rectangles represent local nodes, each of which serves subscribers by means of short-range links at very high, ultra high, or microwave radio frequencies. The nodes are interconnected by terrestrial radio links if they are relatively near each other. If the nodes are widely separated, satellite links are used.

The RGB Color Model

All visible colors can be obtained by combining red, green, and blue light. The *red/green/blue* (RGB) *color model* takes advantage of this fact.

Hue, Saturation, and Brightness

Color, as perceived by the human eye, is a function of the wavelength of optical EM energy. When most of the energy is concentrated near a single wavelength, you see an intense *hue.* The vividness of a hue is called the *saturation.* The *brightness* of a color is a function of how much total energy the light contains. In most video displays, there is a control for adjusting the brightness, also called *brilliance.*

3D Color

A color *palette* is obtained by combining pure red, green, and blue in various ratios. Assign each primary color an axis in Cartesian three-dimensional (3D) space as shown in Fig. 26-7. Call the axes R (for red), G (for green), and B (for blue). Color brightness can range from 0 to 255, or binary

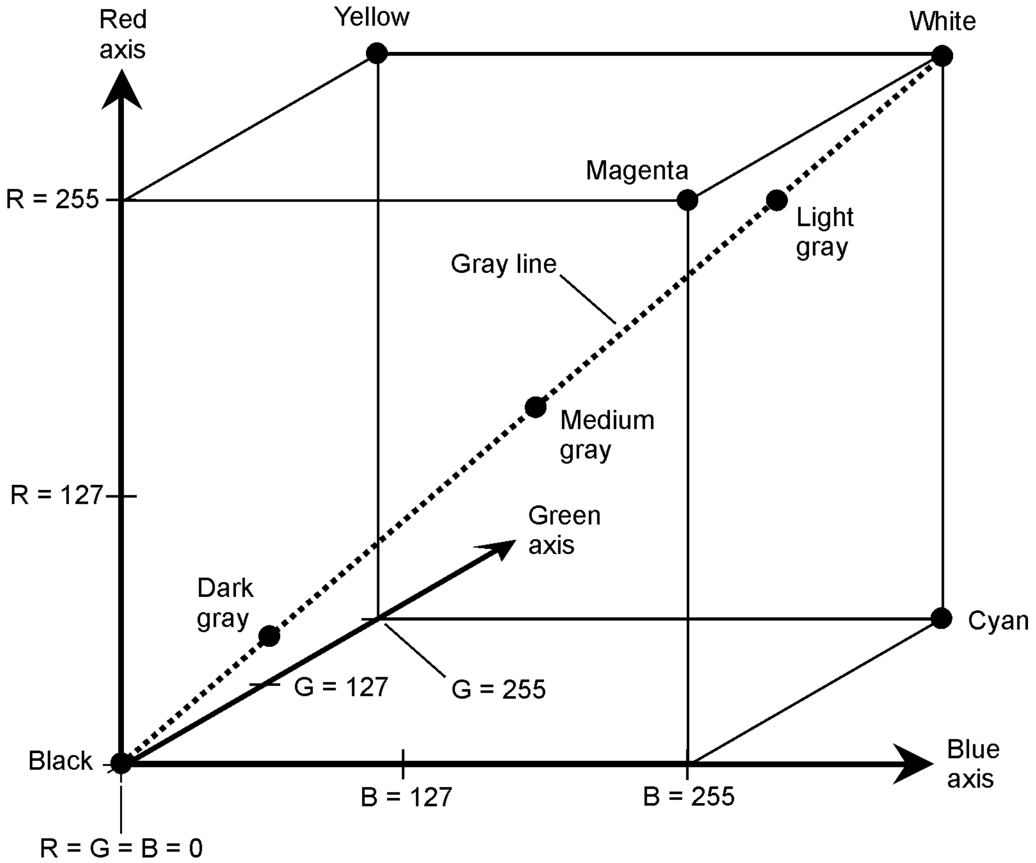
Red axis
Yellow
White
Magenta
R = 255
Light gray
Gray line
Medium gray
R = 127
Green axis
Dark gray
Cyan
G = 127
G = 255
Black
Blue axis
B = 127
B = 255
R = G = B = 0

26-7 The RGB color model, depicted as a cube in Cartesian three-dimensional (3D) space.

00000000 to 11111111. The result is 16,777,216 (256^3) possible colors. Any point within the cube represents a unique color.

Colors in the RGB model can also be represented by six-digit hexadecimal numbers, such as 005CFF. In this scheme, the first two digits represent the red (R) intensity in 256 levels ranging from 00 to FF. The middle two digits represent the green (G) intensity, and the last two digits represent the blue (B) intensity.

Some RGB systems use only 16 levels for each primary color (binary 0000 through 1111). This results in 4096 possible colors.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. Suppose inverters are placed in series with both inputs of an AND gate. Under what conditions is the output of the resulting black box high?
   (a) If and only if both inverter inputs are high
   (b) If and only if both inverter inputs are low
   (c) If and only if one inverter input is high and the other is low
   (d) Under no conditions (the output is always low)

2. Suppose an AND gate is followed by an inverter. Under what conditions is the output of the resulting black box low?
   (a) If and only if both inputs are high
   (b) If and only if both inputs are low
   (c) If and only if one input is high and the other is low
   (d) Under no conditions (the output is always high)

3. What is the binary equivalent of decimal 29?
   (a) 10101
   (b) 11101
   (c) 10111
   (d) 11011

4. In Boolean algebra, addition represents
   (a) the logical NOT operation.
   (b) the logical AND operation.
   (c) the logical OR operation.
   (d) the logical NAND operation.

5. If the output to a logical inverter is low, it means that
   (a) both of the two inputs are high.
   (b) both of the two inputs are low.

(c) the single input is high.
(d) the single input is low.

6. What is the decimal equivalent of binary 1111?
   (a) 15
   (b) 31
   (c) 63
   (d) 127
7. The output of an AND gate is high if and only if
   (a) one input is high and the others are low.
   (b) one input is low and the others are high.
   (c) any of the inputs is low.
   (d) all inputs are high.
8. DeMorgan's Theorem states that, for all logical statements X and Y,
   (a) −(X*Y) is equivalent to X+Y.
   (b) X*Y is equivalent to −(X+Y).
   (c) (−X)+(−Y) is equivalent to X*Y.
   (d) (−X)+(−Y) is equivalent to −(X*Y).
9. If you see a number represented by FF in the documentation for an electronic circuit or system, you can be certain that the number is
   (a) radix 16.
   (b) radix 10.
   (c) radix 8.
   (d) radix 2.
10. Which of the following codes is digital?
    (a) Morse
    (b) Baudot
    (c) ASCII
    (d) All of the above
11. Which of the following voltages could normally represent a 1 in positive logic?
    (a) 0.0 V
    (b) +0.5 V
    (c) +5 V
    (d) +50 V
12. Which of the following voltages might normally represent a 1 in negative logic?
    (a) 0.0 V
    (b) +12 V

(c) +5 V
(d) +50 V

13. Fill in the blank to make the following sentence true: "____ data transmission refers to digital bits sent along a single line, one bit after another."

(a) Serial
(b) Synchronous
(c) Parallel
(d) Analog

14. The inputs of an R-S flip-flop are known as

(a) low and high.
(b) octal and binary.
(c) synchronous and asynchronous.
(d) reset and set.

15. When both inputs of an R-S flip-flop are 0,

(a) the outputs stay as they are.
(b) Q = 0 and –Q = 1.
(c) Q = 1 and –Q = 0.
(d) the resulting outputs can be absurd.

16. The fifth digit from the right in a binary number carries which decimal value?

(a) 64
(b) 32
(c) 24
(d) 16

17. What is the octal equivalent of binary 1010?

(a) 4
(b) 10
(c) 12
(d) There is no way to tell without more information.

18. What is the largest possible radix 10 number that can be represented as a six-digit binary number?

(a) Decimal 256
(b) Decimal 128
(c) Decimal 64
(d) Decimal 63

19. If one of the inputs to a two-input NOR gate is high, what is the output state?

(a) There is not enough information to tell.
(b) Low

(c) High
(d) It depends on the state of the other input.

20. Suppose a logic circuit has four inputs W, X, Y, and Z. How many possible input combinations are there?
(a) 4
(b) 8
(c) 16
(d) 32

Test: Part 3

DO NOT REFER TO THE TEXT WHEN TAKING THIS TEST. A GOOD SCORE IS AT LEAST 37 CORRECT. Answers are in the back of the book. It's best to have a friend check your score the first time, so you won't memorize the answers if you want to take the test again.

1. In a JFET, the control electrode is usually the
   (a) source.
   (b) emitter.
   (c) drain.
   (d) base.
   (e) gate.
2. A diode can be used as a frequency multiplier because of its
   (a) junction capacitance.
   (b) nonlinearity.
   (c) avalanche voltage.
   (d) forward breakover.
   (e) charge carrier concentration.
3. Which of the following is not a common form of data transmission?
   (a) Polarization modulation
   (b) Frequency modulation
   (c) Amplitude modulation
   (d) Phase modulation
   (e) Pulse modulation
4. A very brief, high-voltage spike on an ac power line is called
   (a) a bleeder.
   (b) an arc.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

(c) a transient.
(d) an avalanche.
(e) a clipped peak.

5. Which of the following is not characteristic of an oscillator?
   (a) Negative feedback
   (b) Good output-to-input coupling
   (c) Reasonably high transistor gain
   (d) Alternating current signal output
   (e) Usefulness as a signal generator

6. Which layer of the ionosphere exists at the lowest altitude?
   (a) The F layer
   (b) The E layer
   (c) The D layer
   (d) The C layer
   (e) The B layer

7. The beta of a bipolar transistor is another name for its
   (a) current amplification factor.
   (b) voltage amplification factor.
   (c) power amplification factor.
   (d) maximum amplification frequency.
   (e) optimum amplification frequency.

8. In a schematic diagram, the symbol for a PNP bipolar transistor can be recognized by
   (a) a broken vertical line inside the symbol circle.
   (b) an arrow inside the symbol circle, pointing from the base toward the emitter.
   (c) an arrow inside the symbol circle, pointing outward from the collector.
   (d) a collector biased at a positive voltage with respect to the emitter.
   (e) an arrow inside the symbol circle, pointing from the emitter toward the base.

9. Fill in the blank in the following sentence to make it true: "In an oscillator, ____ is an expression of the extent to which the circuit maintains constant signal frequency output under variable operating conditions, and is also an expression of its overall reliability."
   (a) sensitivity
   (b) drift ratio
   (c) gain
   (d) selectivity
   (e) stability

10. A Zener diode would most likely be found in
    (a) the mixer in a superheterodyne receiver.
    (b) the PLL in a circuit for detecting FM.

(c) the product detector in a receiver for SSB.
(d) the voltage regulator in a power supply.
(e) the AF oscillator in an AFSK transmitter.

11. When the bias in an FET stops the flow of current, the condition is called
    (a) forward breakover.
    (b) cutoff.
    (c) reverse bias.
    (d) pinchoff.
    (e) avalanche.

12. The VA rating of a transformer is an expression of
    (a) the maximum frequency at which it can function.
    (b) the type of core material it has.
    (c) the voltage step-up or step-down ratio.
    (d) the impedance transfer ratio.
    (e) none of the above.

13. In an N-type semiconductor, the minority carriers are
    (a) electrons.
    (b) protons.
    (c) holes.
    (d) neutrons.
    (e) positrons.

14. A disadvantage of a half-wave rectifier is the fact that
    (a) the output voltage is excessive compared to that of a full-wave rectifier.
    (b) the output current is excessive compared to that of a full-wave rectifier.
    (c) the output waveform is harder to filter than is the case with a full-wave rectifier.
    (d) it requires several expensive diodes, whereas a full-wave rectifier requires only a single cheap diode.
    (e) it requires an expensive center-tapped transformer, whereas a full-wave rectifier does not need a transformer at all.

15. A power gain of 30 dB is equivalent to which amplification factor?
    (a) 0.001
    (b) $1/30$
    (c) 30
    (d) 1000
    (e) None of the above

16. Suppose an RF power amplifier has a dc collector input of 60 W, and is 75 percent efficient. What is the RF signal output power?

(a) 80 W
(b) 60 W
(c) 45 W
(d) 40 W
(e) Impossible to determine from this data

17. If a listener is not expecting any change in the intensity of a signal or sound, the smallest change that can be noticed is approximately which of the following?

(a) ±0.3 dB
(b) ±1 dB
(c) ±3 dB
(d) ±10 dB
(e) ±30 dB

18. A common base circuit is commonly employed as

(a) a microwave oscillator.
(b) a low-pass filter.
(c) a noise generator.
(d) a phase-locked loop.
(e) an RF power amplifier.

19. A semiconductor material especially noted for its photoconductivity, and which is often used in making photocells, is

(a) selenium.
(b) lithium salt.
(c) copper chloride.
(d) aluminum oxide.
(e) polyethylene.

20. Which type of PA circuit provides the highest efficiency when used with the appropriate amount of driving power?

(a) Class A
(b) Class AB
(c) Class B
(d) Class C
(e) All of the above circuits are equally efficient.

21. Baudot is a form of

(a) video modulation.
(b) diode.

(c) digital code.
(d) voice modulation.
(e) AM detector.

22. The most stable type of oscillator circuit uses
    (a) a tapped coil.
    (b) a split capacitor.
    (c) negative feedback.
    (d) a common base arrangement.
    (e) a quartz crystal.

23. If the source-gate junction in an FET conducts,
    (a) it is a sign of improper bias.
    (b) the device will work in class C.
    (c) the device will oscillate.
    (d) the device will work in class A.
    (e) the circuit will have good stability.

24. What is the radix of the octal number system?
    (a) 2
    (b) 2^2
    (c) 2^3
    (d) 2^4
    (e) 2^5

25. Signal-to-noise ratio (S/N) is often specified when stating
    (a) the selectivity of a receiver.
    (b) the stability of an oscillator.
    (c) the modulation coefficient of a transmitter.
    (d) the sensitivity of a receiver.
    (e) the polarization of an EM wave.

26. In a reverse-biased semiconductor diode, the junction capacitance depends on
    (a) the width of the depletion region.
    (b) the reverse current.
    (c) the P:N ratio.
    (d) the gate bias.
    (e) the avalanche voltage.

27. Suppose the bandwidth of the channel in a receiver's IF filter is 4.5 kHz at 30-dB power attenuation, and 3.0 kHz at 3-dB power attenuation. What is the shape factor?
    (a) 10:1
    (b) 1.5:1

- (c) 1:1
- (d) 1:10
- (e) More information is necessary to determine this.

28. A simple power supply filter can be built with
    - (a) a capacitor in series with the dc output.
    - (b) an inductor in parallel with the dc output.
    - (c) a rectifier in parallel with the dc output.
    - (d) a resistor in series and an inductor in parallel with the dc output.
    - (e) a capacitor in parallel with the dc output.

29. Which of the following bipolar transistor circuits can, in theory, provide the most amplification?
    - (a) Common emitter
    - (b) Common base
    - (c) Common collector
    - (d) Common gate
    - (e) Common drain

30. The ratio of the difference in I_C to the difference in I_B that occurs when a small signal is applied to the base of a bipolar transistor is called the
    - (a) static current amplification.
    - (b) dynamic current amplification.
    - (c) power factor.
    - (d) efficiency.
    - (e) current attenuation factor.

31. An example of a device that is commonly used as an oscillator at microwave frequencies is
    - (a) a rectifier diode.
    - (b) a weak-signal diode.
    - (c) a Gunn diode.
    - (d) a Zener diode.
    - (e) an avalanche diode.

32. In the operation of a PNP bipolar transistor, which of the following is normal concerning the emitter and collector voltages?
    - (a) The collector is positive relative to the emitter.
    - (b) The collector is at the same voltage as the emitter.
    - (c) The collector is negative relative to the emitter.
    - (d) The collector can be either positive or negative relative to the emitter.
    - (e) The collector must be at ground potential.

33. What can be done to minimize the capacitance of the P-N junction in a semiconductor diode, thereby making the component effective as a high-speed RF switch?
    - (a) The surface area of the P-N junction can be maximized.
    - (b) A layer of intrinsic semiconductor can be placed between the P- and N-type materials.
    - (c) The frequency of the applied signal can be made as high as possible.
    - (d) The diode can be forward-biased with a high voltage.
    - (e) A small-value resistor can be connected in series with each of several diodes, and the series diode-resistor combinations can all be connected in parallel.
34. Which type of modulation consists of one voice sideband, with a suppressed carrier?
    - (a) AM
    - (b) LSB
    - (c) FM
    - (d) RTTY
    - (e) PCM
35. Suppose a series-parallel network of resistors is assembled into an *RF attenuator pad*, which is a circuit deliberately intended to produce a constant signal loss over a wide band of frequencies. If the input RF voltage to this attenuator pad is 500 μV and the output RF voltage is 100 μV, what is the loss in decibels?
    - (a) 14.0 dB
    - (b) 6.99 dB
    - (c) 5.00 dB
    - (d) 1.40 dB
    - (e) 0.699 dB
36. In an AND gate, the output is high
    - (a) if any input is high.
    - (b) only when all inputs are low.
    - (c) if any input is low.
    - (d) only when all inputs are high.
    - (e) only when all inputs have identical logic states.
37. A voltage-controlled oscillator makes use of
    - (a) a varactor diode.
    - (b) a Zener diode.
    - (c) negative feedback.
    - (d) a split capacitance.
    - (e) adjustable gate or base bias.
38. The best electrical ground for a power supply is obtained by
    - (a) connecting a small-value capacitor between the hot wire and a good earth ground, such as an 8-ft copper rod driven into the earth.

(b) ensuring that the ground wire from the power supply is connected to the third wire hole in one of the utility outlets of a properly installed three-wire electrical circuit.
(c) using a center-tapped transformer and a full-wave rectifier circuit, in conjunction with good filtering, voltage regulation, and transient suppression.
(d) connecting a large-value inductor in series with the hot wire, thereby ensuring that the voltage on that wire cannot be shunted to ground.
(e) making sure that all equipment connected to the supply is provided with fuses and/or circuit breakers of the proper ratings.

39. Suppose the output signal voltage from an amplifier is 35 times the input signal voltage. If the input and output impedances are both pure resistances of the same value, what is the gain?
(a) 15 dB
(b) 31 dB
(c) 35 dB
(d) 350 dB
(e) 700 dB

40. In an exclusive OR gate, the output is high
(a) if either input is high.
(b) only when both inputs are low.
(c) if either input is low.
(d) only when both inputs are high.
(e) only when the inputs have opposite logic states.

41. A ratio detector is a circuit for demodulating which of the following?
(a) AM
(b) PM
(c) FM
(d) SSB
(e) AFSK

42. Suppose there is a binary digital black box with two inputs, called X and Y, and an output called Z. If X = 0 and Y = 1, then Z = 0. In all other instances, Z = 1. Which of the following logical expressions represents the contents of the black box?
(a) X + (–Y)
(b) X * (–Y)
(c) X – (*Y)
(d) X = (–Y)
(e) X – (+Y)

43. A method of modulation in which the strength of pulses varies is called
(a) pulse ratio modulation.
(b) pulse position modulation.

(c) pulse frequency modulation.
(d) pulse amplitude modulation.
(e) pulse width modulation.

44. Boolean algebra is
(a) useful for calculating amplifier gain in decibels.
(b) a useful tool in binary digital circuit design.
(c) used to calculate the value of an unknown.
(d) used with negative pulse modulation schemes.
(e) used with positive pulse modulation schemes.

45. A voltage-doubler power supply is best for use in
(a) circuits that need high voltage but do not draw much current.
(b) low-voltage devices.
(c) high-current appliances.
(d) all kinds of electronic equipment.
(e) broadcast transmitter power amplifiers.

46. An optoisolator consists of
(a) two Zener diodes back-to-back.
(b) an LED and a photodiode.
(c) two NPN transistors in series.
(d) an NPN transistor followed by a PNP transistor.
(e) a PNP transistor followed by an NPN transistor.

47. When a semiconductor is reverse-biased with a large enough voltage, it will conduct. This phenomenon is known as
(a) bias effect.
(b) avalanche effect.
(c) forward breakover.
(d) saturation.
(e) conduction effect.

48. Synchronizing pulses in a conventional analog TV signal
(a) keep the brightness constant.
(b) keep the contrast constant.
(c) keep the image from tearing or rolling.
(d) ensure that the colors are right.
(e) keep the image in good focus.

49. In an enhancement-mode MOSFET,
(a) the channel conducts fully with zero gate bias.
(b) the channel conducts partially with zero gate bias.

(c) the channel conducts ac but not dc.
(d) the channel conducts dc but not ac.
(e) the channel does not conduct with zero gate bias.

50. In a step-up power transformer,
    (a) the primary voltage is more than the secondary voltage.
    (b) the secondary voltage is more than the primary voltage.
    (c) the primary and secondary voltages are the same.
    (d) the secondary must be center-tapped.
    (e) the primary must be center-tapped.

This page intentionally left blank

4
PART

Specialized Devices and Systems

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

This page intentionally left blank

27
CHAPTER

Antennas

ANTENNAS CAN BE CATEGORIZED INTO TWO MAJOR CLASSES: RECEIVING TYPES AND TRANSMITTING types. Most transmitting antennas can also function effectively for reception. Some receiving antennas can efficiently transmit EM signals; others cannot.

Radiation Resistance

When RF current flows in an electrical conductor, some EM energy is radiated into space. If a resistor is substituted for the antenna, in combination with a capacitor or inductor to mimic any inherent reactance in the antenna, the transmitter behaves in the same manner as when connected to the actual antenna. For any antenna operating at a specific frequency, there exists a specific resistance, in ohms, for which this can be done. This is known as the *radiation resistance* (R_R) of the antenna at the frequency in question. Radiation resistance is specified in ohms. This phenomenon was introduced in Chap. 17. Now you'll see why it's important!

Determining Factors

Suppose a thin, straight, lossless vertical antenna is placed over perfectly conducting ground. Then R_R is a function of the vertical-antenna conductor height in wavelengths (Fig. 27-1A). Suppose a thin, straight, lossless wire is placed in free space and fed at the center. Then R_R is a function of the overall conductor length in wavelengths (Fig. 27-1B).

Antenna Efficiency

Efficiency is rarely crucial to the performance of a receiving antenna system, but it is important in any transmitting antenna system. Radiation resistance always appears in series with *loss resistance* (R_L). The antenna efficiency, *Eff*, is given by:

$$Eff = R_R/(R_R + R_L)$$

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

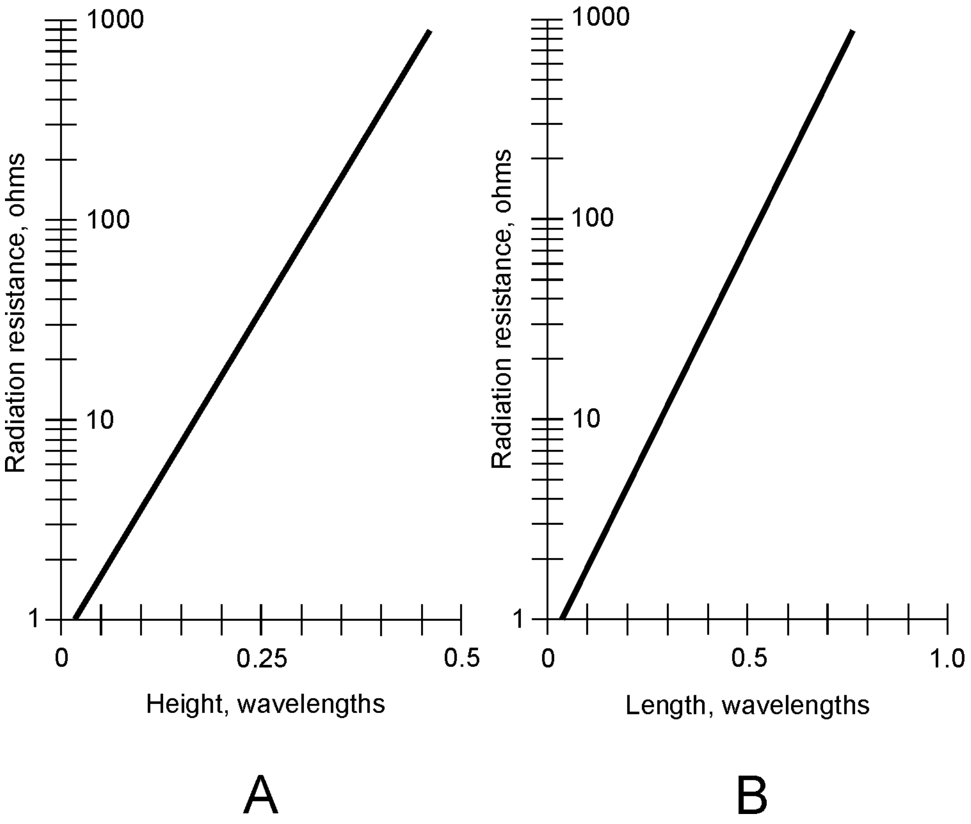

27-1 Approximate values of radiation resistance for vertical antennas over perfectly conducting ground (A) and for center-fed antennas in free space (B).

which is the ratio of the radiation resistance to the total antenna system resistance. As a percentage,

$$Eff_{\%} = 100\ R_{\mathrm{R}}/(R_{\mathrm{R}} + R_{\mathrm{L}})$$

High efficiency in a transmitting antenna is obtained when the radiation resistance is much larger than the loss resistance. Then most of the power goes into useful EM radiation, and relatively little is dissipated as heat in the earth and in objects surrounding the antenna. When the opposite situation exists—radiation resistance comparable to, or smaller than, the loss resistance, a transmitting antenna is inefficient. This is often the case for extremely short radiators, because they always have low radiation resistance. To some extent this can be overcome by careful antenna design and location to minimize loss resistance, but even the most concerted efforts rarely reduce the loss resistance to less than a few ohms.

If an antenna system has a high loss resistance, it can nevertheless work efficiently if its radiation resistance is high enough. When an antenna radiator is just the right length at a given frequency, and if it is constructed of wire or thin tubing, its radiation resistance can be in excess of 1000 Ω. This makes it easy to construct an efficient antenna even when the loss resistance is fairly high.

Half-Wave Antennas

A half wavelength in free space is given by the equation:

$$L_{ft} = 492/f_o$$

where L_{ft} is the linear distance in feet, and f_o is the fundamental frequency, in megahertz, at which the antenna exhibits resonance. A half wavelength in meters, L_m, is given by:

$$L_m = 150/f_o$$

For ordinary wire, the results as obtained above should be multiplied by a velocity factor, v, of 0.95 (95 percent). For tubing or large-diameter wire, v can range down to about 0.90 (90 percent).

Open Dipole

An *open dipole* or *doublet* is a half-wavelength (λ/2) radiator fed at the center (Fig. 27-2A). Each leg of the antenna is a quarter wavelength (λ/4) long. For a straight wire radiator, the length L_{ft}, in feet, at a design frequency f_o, in megahertz, for a center-fed, λ/2 dipole is approximately:

$$L_{ft} = 468/f_o$$

The length in meters is approximately:

$$L_m = 143/f_o$$

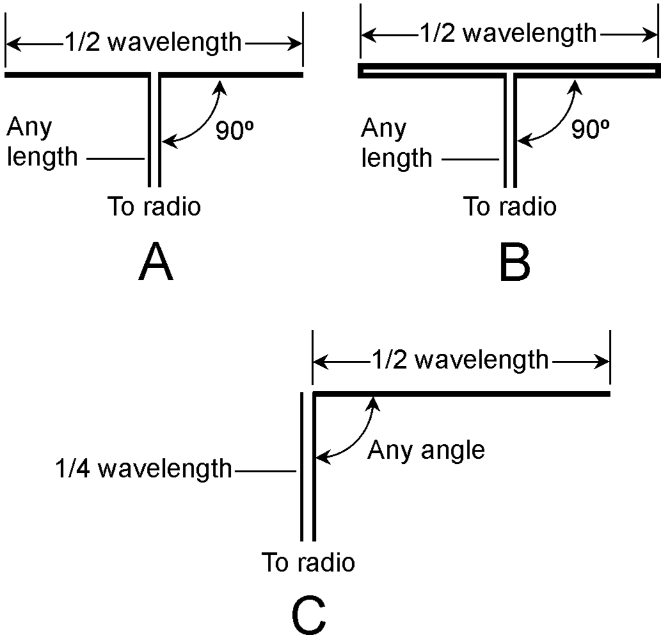
1/2 wavelength
Any length
90°
To radio
A
1/2 wavelength
Any length
90°
To radio
B
1/2 wavelength
1/4 wavelength
Any angle
To radio
C

27-2 Basic half-wave antennas. At A, dipole antenna. At B, folded-dipole antenna. At C, zepp antenna.

These values assume $v = 0.95$. In free space, the impedance at the feed point is a pure resistance of approximately 73 Ω. This represents the radiation resistance alone, in the absence of reactance at the resonant frequency.

Folded Dipole

A *folded dipole antenna* is a λ/2, center-fed antenna constructed of two parallel wires with their ends connected together (Fig. 27-2B). The feed-point impedance of the folded dipole is a pure resistance of approximately 290 Ω. This makes the folded dipole ideal for use with high-impedance, parallel-wire transmission lines in applications where gain and directivity are not especially important.

Half-Wave Vertical

A radiator measuring λ/2 can be stood up, fed at the base (the bottom end) against an earth ground with a transmatch or antenna tuner designed for high impedances, and connected to a radio by a coaxial cable feed line. This type of antenna is an efficient radiator even in the presence of considerable loss resistance, because the radiation resistance is extremely high.

Zepp

A *zeppelin antenna*, also called a *zepp*, is a λ/2 radiator, fed at one end with a λ/4 section of open-wire line (Fig. 27-2C). The impedance at the feed point is an extremely high, pure resistance. At the transmitter end of the line, the impedance is a low, pure resistance. A zeppelin antenna can operate well at all harmonics of the design frequency. If an *antenna tuner*, also called a *transmatch*, is available to tune out reactance, the feed line can be of any length. Feed-line radiation can be minimized by carefully cutting the radiator to λ/2 at the fundamental frequency, and by using the antenna only at this frequency or one of its harmonics. Zepp antennas are rarely used at frequencies above 30 MHz, except when modified to form a *J pole.*

J Pole

A zepp can be oriented vertically, and the feed line placed so it lies in the same line as the radiating element. This antenna, called a J pole, radiates equally well in all horizontal directions. The J pole is used as a low-budget antenna at frequencies from approximately 10 MHz up through 300 MHz. It is, in effect, a λ/2 vertical fed with an *impedance matching section* consisting of a length of transmission line measuring λ/4. It does not require any radials, and this makes it convenient in locations where space is at a premium.

Some radio amateurs hang long J poles, cut for 3.5 MHz or 1.8 MHz, from kites or helium-filled balloons. Such antennas work well, but they are dangerous if they are not properly tethered to prevent them from breaking off and flying away, or if they are placed where they might fall on power lines. They can develop considerable electrostatic charge, even in clear weather. They are deadly if flown in or near thunderstorms.

Quarter-Wave Verticals

A *quarter wavelength* is related to frequency by the equation:

$$L_{\mathrm{ft}} = 246v/f_{\mathrm{o}}$$

where L_{ft} represents λ/4 in feet, f_o represents the frequency in megahertz, and v represents the velocity factor. If the length is expressed in meters as L_m, then the formula is:

$$L_m = 75v/f_o$$

For a typical wire conductor, $v = 0.95$ (95 percent); for metal tubing, v can range down to approximately 0.90 (90 percent).

A λ/4 antenna must be operated against a low-loss RF ground. The feed-point impedance over perfectly conducting ground is approximately 37 Ω. This represents radiation resistance in the absence of reactance, and provides a reasonable impedance match to most coaxial transmission lines.

Ground-Mounted Vertical

The simplest vertical antenna is a λ/4 radiator mounted at ground level. The radiator is fed with a coaxial cable. The center conductor is connected to the base of the radiator, and the shield is connected to a ground system.

Unless an extensive *ground radial* system is installed, a ground-mounted vertical is likely to be inefficient unless the surface happens to be an excellent conductor (saltwater or a salt marsh, for example). Another problem is the fact that vertically polarized antennas receive more human-made noise than horizontal antennas. In addition, the EM fields from ground-mounted transmitting antennas are more likely to cause interference to nearby electronic devices than are the EM fields from antennas installed at a height.

Ground Plane

A *ground-plane antenna* is a vertical radiator, usually measuring λ/4, operated against a system of λ/4 conductors called *radials.* The *feed point,* where the feed line joins the antenna radiator and the hub of the radial system, is elevated. When the feed point is at least λ/4 above the ground, only three or four radials are necessary to obtain a low-loss RF ground system. The radials extend outward from the base of the antenna at an angle between 0° and 45° with respect to the horizon. Figure 27-3A illustrates a typical ground-plane antenna with four horizontal radials.

A ground-plane antenna should be fed with coaxial cable. The feed-point impedance of a ground-plane antenna having a λ/2 radiator is about 37 Ω if the radials are horizontal; the impedance increases as the radials droop, reaching about 50 Ω at a *droop angle* of 45°.

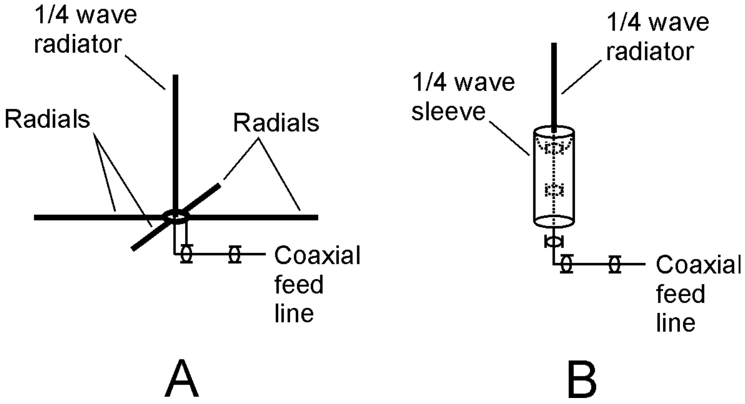

27-3 Basic quarter-wave verticals. At A, ground-plane antenna. At B, coaxial antenna.

Coaxial Antenna

The radials in a ground-plane antenna can extend straight downward, and then can be merged to form a $\lambda/2$ cylinder or sleeve that is concentric with the feed line. In this case, the feed-point impedance is approximately 73 Ω. This is known as a *coaxial antenna* (Fig. 27-3B).

Loops

Any receiving or transmitting antenna, consisting of one or more turns of wire forming a dc short circuit, comprises a *loop antenna.*

Small Loop

A *small loop antenna* has a circumference of less than 0.1λ (for each turn) and is suitable for receiving, but generally not for transmitting. It is the least responsive to signals coming from along its axis, and most responsive in the plane perpendicular to its axis. A capacitor can be connected in series or parallel with the loop to provide a resonant response. An example of such an arrangement is shown in Fig. 27-4.

Small loops are useful for *radio direction finding* (RDF), and also for reducing interference caused by human-made noise or strong local signals. The *null* along the axis is sharp and deep, and can be pointed in the direction of an offending signal or noise source.

Loopstick

For receiving applications at frequencies up to approximately 20 MHz, a *loopstick antenna* is sometimes used. This antenna, a variant of the small loop, consists of a coil wound on a solenoidal (rod-shaped), powdered-iron core. A series or parallel capacitor, in conjunction with the coil, forms a tuned circuit. A loopstick displays directional characteristics similar to those of the small loop antenna shown in Fig. 27-4. The sensitivity is maximum off the sides of the coil, and a sharp null occurs off the ends.

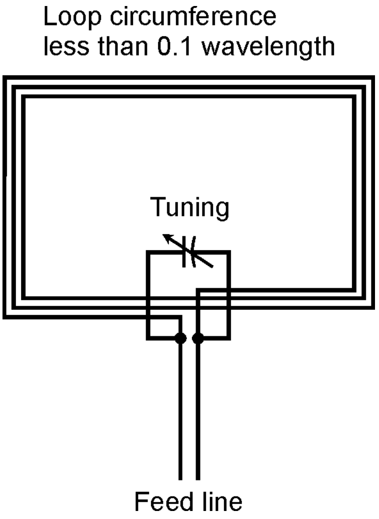

27-4 A small loop antenna with a capacitor for adjusting the resonant frequency.

Large Loop

A *large loop antenna* usually has a circumference of either λ/2 or λ (a full wavelength), is circular or square in shape, and lies entirely in a single plane. It can work well for transmitting or receiving. The λ/2 loop presents a high impedance at the feed point, and maximum radiation/response occurs in the plane of the loop. The λ loop presents an impedance of about 100 Ω at the feed point, and the maximum radiation/response occurs along the axis (that is, perpendicular to the plane containing the loop).

The λ/2 loop exhibits a slight power loss relative to a λ/2 dipole in its favored directions. The λ loop shows a slight gain over a λ/2 dipole in its favored directions. These properties hold for loops up to several percent larger or smaller than exact λ/2 or λ circumferences (as determined for the wavelength in free space). Resonance can be obtained by means of an antenna tuner if the loop is fed with open-wire transmission line.

Sometimes, loop antennas measuring several wavelengths in circumference are strung up horizontally among multiple supports. These are technically large loops, but their gain and directional characteristics are hard to predict. If fed with open-wire line and an antenna tuner, and if placed at least λ/4 above the surface, such an antenna can be exceptionally effective.

Ground Systems

End-fed λ/4 antennas, such as the ground plane, require low-loss RF ground systems in order to perform efficiently. Center-fed λ/2 antennas, such as the dipole, do not. However, good grounding is advisable for any antenna system in order to minimize interference and electrical hazards.

Electrical versus RF Ground

Electrical grounding is important for personal safety. It can help protect equipment from damage if lightning strikes in the vicinity. It also minimizes the risk of *electromagnetic interference* (EMI) to and from radio equipment. In a three-wire electrical system, the ground prong on the plug should never be defeated, because such modification can result in dangerous voltages appearing on exposed metal surfaces.

A good RF ground system can help minimize EMI, even if it is not necessary for efficient antenna operation. Figure 27-5 shows a proper RF ground scheme (A) and an improper one (B). In a good RF ground system, each device is connected to a common *ground bus*, which in turn runs to the earth ground through a single conductor. This conductor should be as short as possible. A poor ground system contains *ground loops* that can act like loop antennas and increase the risk of EMI.

Radials and the Counterpoise

With a surface-mounted vertical antenna, there should be as many radials as possible, and they should be as long as possible. They can lie on the surface or be buried a few inches underground. The greater the number of radials of a given length, the better the antenna will work. Also, the longer the radials for a given number, the better. The radials should all converge toward, and be connected to, a ground rod at the feed point.

A *counterpoise* is a means of obtaining an RF ground or ground plane without a direct earth-ground connection. A grid of wires, a screen, or a metal sheet is placed above the surface and oriented horizontally, to provide *capacitive coupling* to the earth. This minimizes RF ground loss. Ideally, the radius of a counterpoise should be at least λ/4 at the lowest operating frequency.

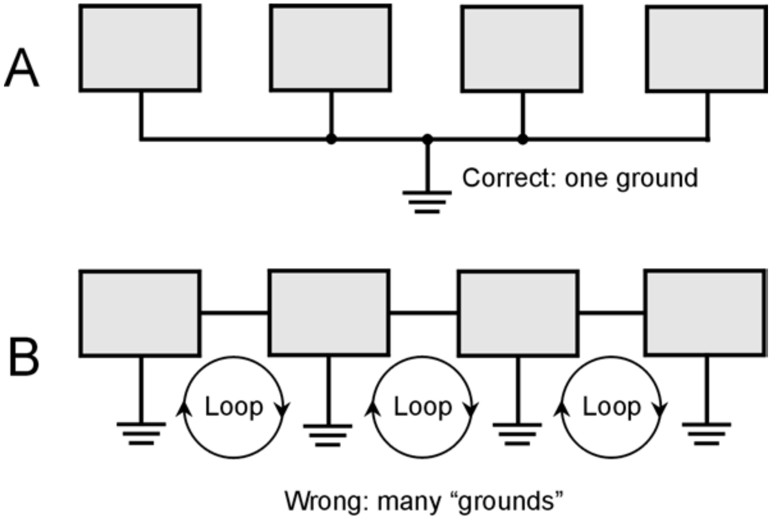

27-5 At A, the correct method for grounding multiple units. At B, an incorrect method creates RF ground loops.

Gain and Directivity

The *power gain* of a transmitting antenna is the ratio of the maximum *effective radiated power* (ERP) to the actual RF power applied at the feed point. Power gain is expressed in decibels (dB).

Suppose the ERP, in watts, for a given antenna is P_{ERP}, and the applied power, also in watts, is P. Then the following equation holds:

$$\text{Power gain (dB)} = 10 \log_{10} (P_{\mathrm{ERP}}/P)$$

Power gain is always measured in the favored direction or directions of an antenna. These are the directions in which the antenna performs the best.

For power gain to be defined, a *reference antenna* must be chosen with a gain that is defined as 0 dB. This reference antenna is usually a $\lambda/2$ dipole in free space. Power-gain figures taken with respect to a dipole (in its favored directions) are expressed in units called dBd. The reference antenna for power-gain measurements can also be an *isotropic antenna*, which theoretically radiates and receives equally well in all directions in three dimensions. In this case, units of power gain are called dBi. For any given antenna, the power gains in dBd and dBi are different by approximately 2.15 dB:

$$\text{Power gain (dBi)} = 2.15 + \text{Power gain (dBd)}$$

Directivity Plots

Antenna radiation and response patterns are represented by plots such as those shown in Fig. 27-6. The location of the antenna is assumed to be at the center (or *origin*) of a *polar coordinate system.* The greater the radiation or reception capability of the antenna in a certain direction, the farther from the center the points on the chart are plotted.

A dipole antenna, oriented horizontally so that its conductor runs in a north-south direction, has a *horizontal plane* (or *H plane*) pattern similar to that in Fig. 27-6A. The *elevation plane* (or *E plane*) pattern depends on the height of the antenna above effective ground at the viewing angle. With the dipole oriented so that its conductor runs perpendicular to the page, and the antenna 1⁄4 wavelength above effective ground, the E plane antenna pattern for a half-wave dipole resembles the graph shown at B.

Forward Gain

Forward gain is expressed in terms of the ERP in the *main lobe* (favored direction) of a *unidirectional* (one-directional) antenna compared with the ERP from a reference antenna, usually a half-wave dipole, in its favored directions. This gain is calculated and defined in dBd at microwave frequencies; large dish antennas can have forward gain upward of 35 dBd. In general, as the wavelength decreases (the frequency gets higher), it becomes easier to obtain high forward gain figures.

Front-to-Back Ratio

The *front-to-back* (f/b) *ratio* of a unidirectional antenna is an expression of the concentration of radiation/response in the main lobe, relative to the direction opposite the center of the main lobe. Figure 27-7 shows a hypothetical directivity plot for a unidirectional antenna pointed north. The outer circle depicts the RF *field strength* in the direction of the center of the main lobe, and represents 0 dB. The next smaller circle represents a field strength 5 dB down with respect to the main lobe. Continuing inward, circles represent 10 dB down, 15 dB down, and 20 dB down. The origin represents 25 dB down, and also shows the location of the antenna. The f/b ratio is found, in this case, by comparing the signal levels between north (azimuth 0°) and south (azimuth 180°).

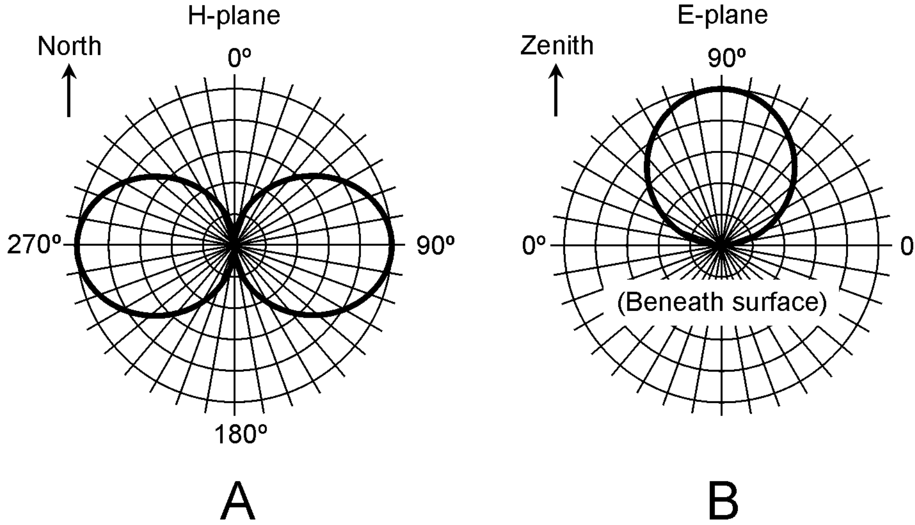

27-6 Directivity plots for a dipole antenna. At A, the H plane (horizontal plane) plot; at B, the E plane (elevation plane) plot.

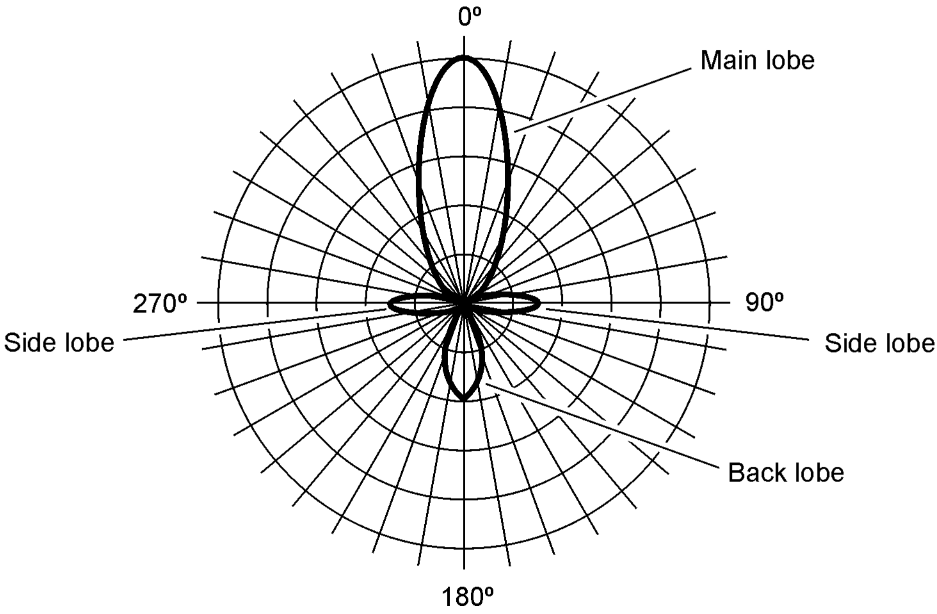

27-7 Directivity plot for a hypothetical antenna. Front-to-back and front-to-side ratios can be determined from such a graph. This plot is in the H plane.

Front-to-Side Ratio

The *front-to-side* (f/s) *ratio* is another expression of the directivity of an antenna system. The term applies to unidirectional antennas, and also to *bidirectional* antennas. The f/s ratio is expressed in decibels (dBd), just as is the f/b ratio. The EM field strength in the favored direction is compared with the field strength at right angles to the favored direction. An example is shown in Fig. 27-7. The f/s ratios are found, in this case, by comparing the signal levels between north and east (right-hand f/s), or between north and west (left-hand f/s). The right-hand and left-hand f/s ratios are usually the same in theory, although they can differ slightly in practice.

Phased Arrays

A *phased array* uses two or more *driven elements* (radiators connected directly to the feed line) to produce gain in some directions at the expense of other directions.

End-Fire Array

A typical *end-fire array* consists of two parallel λ/2 dipoles fed 90° out of phase and spaced λ/4 apart (Fig. 27-8A). This produces a unidirectional radiation pattern. Or, the two elements can be driven in phase and spaced at a separation of λ (Fig. 27-8B). This results in a bidirectional radiation pattern. In the phasing system, the branches of the transmission line must be cut to precisely the correct lengths, and the velocity factor of the line must be known and be taken into account.

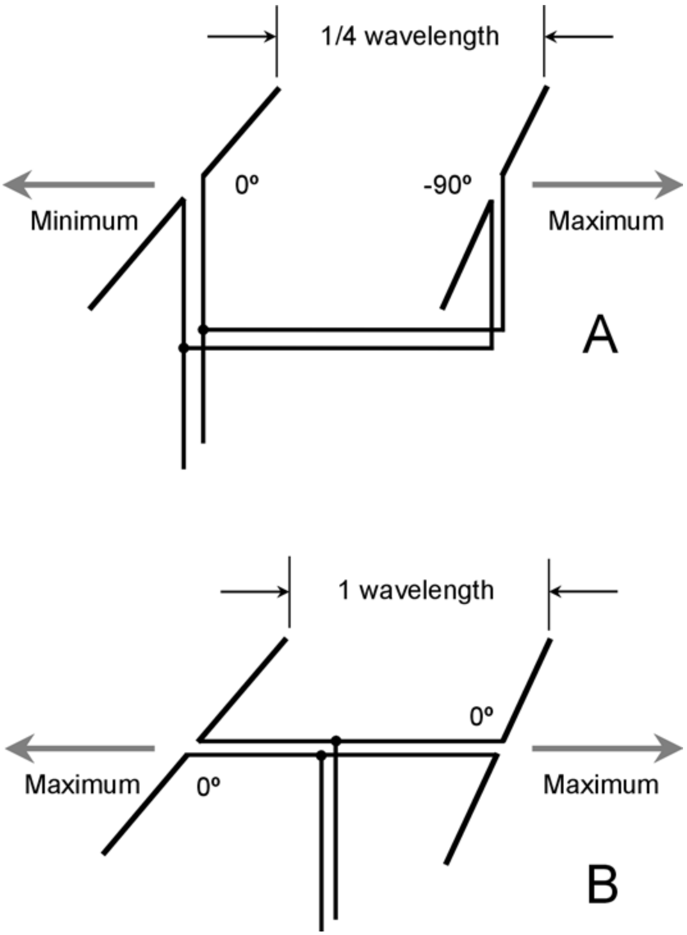

27-8 At A, a unidirectional phased system. At B, a bidirectional phased system.

Longwire

A wire antenna measuring λ or more, and fed at a high-current point or at one end, is a *longwire antenna.* A longwire antenna offers gain over a half-wave dipole. As the wire is made longer, the main lobes get more nearly in line with the antenna, and their amplitudes increase. The gain is a function of the length of the antenna; the longer the wire, the greater the gain. A longwire antenna must be as straight as possible for proper operation.

Broadside Array

Figure 27-9 shows the geometric arrangement of a *broadside array.* The driven elements can each consist of a single radiator, as shown in the figure, or they can consist of more complex antennas with directive properties. In any case, all the elements are identical. If a reflecting screen is placed behind the array of dipoles in Fig. 27-9, the system is known as a *billboard antenna.* The directional properties depend on the number of elements, whether or not the elements have gain themselves, and on the spacing among the elements. In general, the larger the number of elements, the greater the forward gain, and the greater the f/b and f/s ratios.

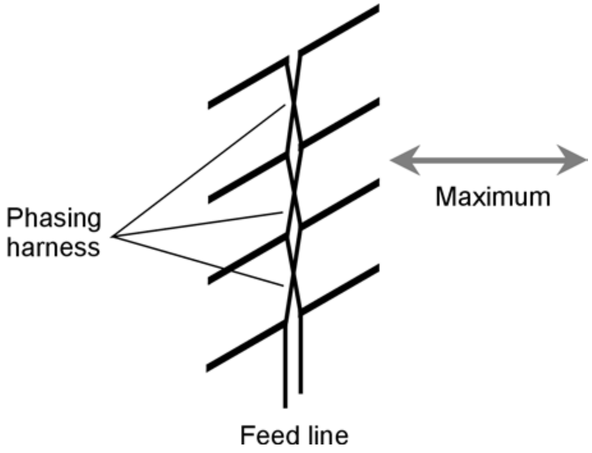

27-9 A broadside array. The elements are all fed in phase.

Parasitic Arrays

Parasitic arrays are used at frequencies ranging from approximately 5 MHz into the microwave range for obtaining directivity and forward gain. Examples include the *Yagi antenna* and the *quad antenna.*

Concept

A *parasitic element* is a conductor that forms an important part of an antenna system, but is not directly connected to the feed line. Parasitic elements operate by means of *EM coupling* to the driven element. When gain is produced in the direction of the parasitic element, the element is a *director.* When gain is produced in the direction opposite the parasitic element, the element is a *reflector.* Directors are a few percent shorter than the driven element; reflectors are a few percent longer.

Yagi

The *Yagi antenna*, sometimes called a "beam," is an array of parallel, straight elements. A two-element Yagi is formed by placing a director or a reflector parallel to, and a specific distance away from, a single $\lambda/2$ driven element. The optimum spacing for a driven-element/director Yagi is 0.1λ to 0.2λ, with the director tuned 5 percent to 10 percent higher than the resonant frequency of the driven element. The optimum spacing for a driven-element/reflector Yagi is 0.15λ to 0.2λ, with the reflector tuned 5 percent to 10 percent lower than the resonant frequency of the driven element. The gain of a well-designed *two-element Yagi* is approximately 5 dBd.

A Yagi with one director and one reflector, along with the driven element, increases the gain and f/b ratio compared with a two-element Yagi. An optimally designed *three-element Yagi* has approximately 7 dBd gain. An example is shown in Fig. 27-10. (This drawing should not be used as an engineering blueprint.)

The gain, f/b ratio, and f/s ratio of a properly designed Yagi antenna all increase as elements are added. This is usually done by placing extra directors in front of a three-element Yagi. Each director is slightly shorter than its predecessor.

Quad

A *quad antenna* operates according to the same principles as the Yagi, except full-wavelength loops are used instead of straight half-wavelength elements.

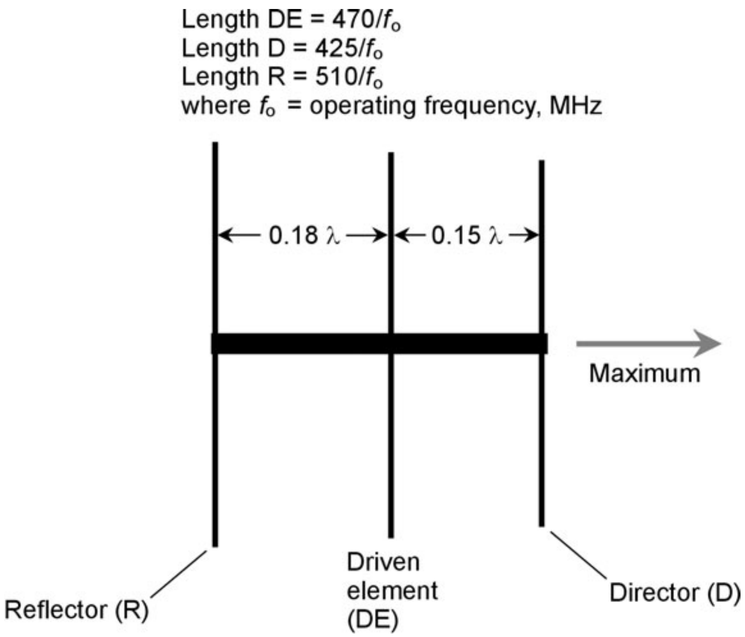

27-10 A three-element Yagi antenna. Dimensions are discussed in the text.

A *two-element quad* can consist of a driven element and a reflector, or it can have a driven element and a director. A *three-element quad* has one driven element, one director, and one reflector. The director has a perimeter of 0.95λ to 0.97λ, the driven element has a perimeter of exactly λ, and the reflector has a perimeter of 1.03λ to 1.05λ. These dimensions, as all antenna element dimensions, are electrical dimensions (taking the velocity factor of wire or tubing into account), and not free-space dimensions.

Additional directors can be added to the basic three-element quad design to form quads having any desired numbers of elements. The gain increases as the number of elements increases. Each succeeding director is slightly shorter than its predecessor. Long quad antennas are practical at frequencies above 100 MHz. At lower frequencies, their construction tends to be unwieldy.

Antennas for Ultrahigh and Microwave Frequencies

At ultrahigh frequencies (UHF) and microwave frequencies, high-gain antennas are reasonable in size because the wavelengths are short.

Waveguides

In Chap. 25, waveguides were briefly mentioned as an option for feeding antenna systems at UHF and microwave frequencies. Here's a little more about them.

A waveguide is a hollow metal pipe having a rectangular or circular cross section. The EM field travels down the pipe, provided that the wavelength is short enough (or the cross-sectional dimensions of the pipe are large enough). In order to efficiently propagate an EM field, a *rectangular waveguide* must have height and width that both measure at least 0.5λ, and preferably more than 0.7λ. A *circular waveguide* should be at least 0.6λ in diameter, and preferably 0.7λ or more.

The characteristic impedance (Z_o) of a waveguide varies with frequency. In this sense, it differs from coaxial or parallel-wire lines, whose Z_o values are generally independent of the frequency.

A properly installed and well-maintained waveguide is an exceptional transmission line because dry air has practically no loss, even at UHF and microwave frequencies. But it is important that the interior of a waveguide be kept free from dirt, dust, insects, spiderwebs, and condensation. Even a small obstruction can seriously degrade the performance and cause the waveguide to become *lossy.*

The main limitation of a waveguide, from a practical standpoint, is inflexibility. You can't run a waveguide from one point to another in a haphazard fashion, as you can do with coaxial cable. Bends or turns in a waveguide present a particular problem, because they must be made gradually. Installing a waveguide for a UHF or microwave antenna is a little like putting in the conduit for a new electrical circuit in a home or business. It's a significant construction project!

Waveguides are impractical for use at frequencies below approximately 300 MHz, because the required cross-sectional dimensions become prohibitively large.

Dish

A *dish antenna* must be correctly shaped and precisely aligned. The most efficient shape, especially at the shortest wavelengths, is a *paraboloidal reflector.* However, a *spherical reflector*, having

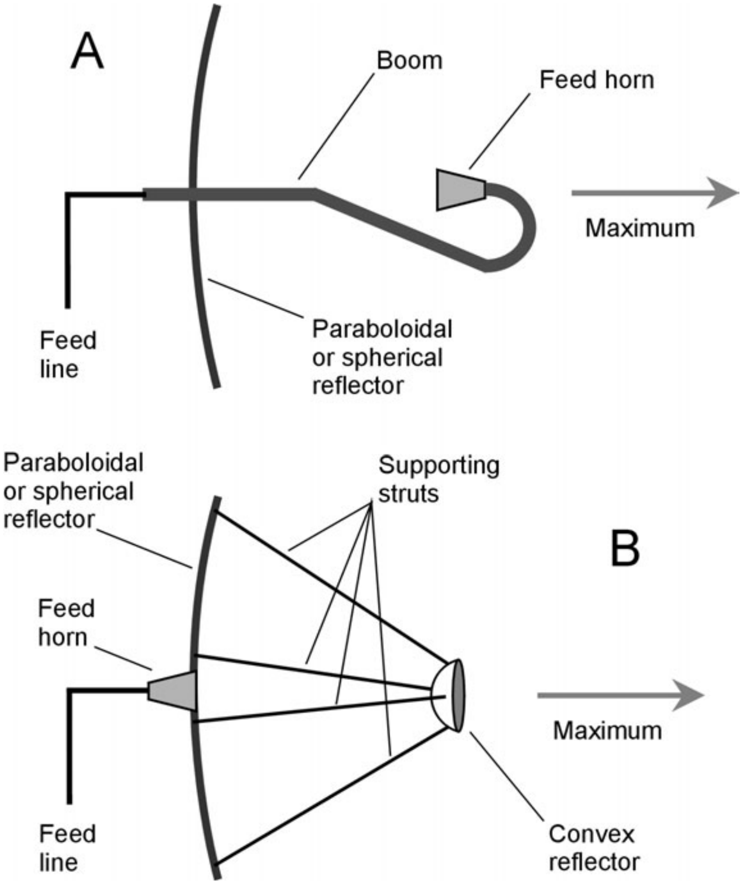
A
Boom
Feed horn
Maximum
Feed line
Paraboloidal or spherical reflector
Paraboloidal or spherical reflector
Supporting struts
B
Feed horn
Maximum
Feed line
Convex reflector

27-11 Dish antennas with conventional feed (A) and Cassegrain feed (B).

the shape of a section of a sphere, can also work well. The feed system consists of a coaxial line or waveguide from the receiver and/or transmitter, and a horn or helical driven element at the focal point of the reflector. *Conventional dish feed* is shown in Fig. 27-11A. *Cassegrain dish feed* is shown in Fig. 27-11B.

The larger the diameter of the reflector in wavelengths, the greater the gain, the f/b ratio, and the f/s ratio, and the narrower the main lobe. A dish antenna must be at least several wavelengths in diameter for proper operation. The reflecting element can be sheet metal, a screen, or a wire mesh. If a screen or mesh is used, the spacing between the wires must be a small fraction of a wavelength.

Helical

A *helical antenna* is a circularly polarized, high-gain, unidirectional antenna. A typical helical antenna is shown in Fig. 27-12. The reflector diameter should be at least 0.8λ at the lowest operating frequency. The radius of the helix should be approximately 0.17λ at the center of the intended operating frequency range. The longitudinal spacing between helix turns should be approximately 0.25λ in the center of the operating frequency range. The overall length of the helix should be at least λ at the lowest operating frequency. A helical antenna can provide about 15 dBd forward gain. Helical antennas are sometimes used in space communications systems.

Corner Reflector

A *corner reflector*, employed with a λ/2 dipole driven element, is illustrated in Fig. 27-13. This provides some gain over a λ/2 dipole by itself. The reflector is made of wire mesh, screen, or sheet metal. The *flare angle* of the reflecting element is approximately 90°. Corner reflectors are widely used in terrestrial communications at UHF and microwave frequencies. Several λ/2 dipoles can be fed in phase and placed along a common axis with a single, elongated reflector, forming a *collinear corner reflector array.*

Horn

The *horn antenna* is shaped like a squared-off trumpet or trombone horn. It provides a unidirectional radiation and response pattern, with the favored direction coincident with the opening of the horn. The feed line is a waveguide that joins the antenna at the narrowest point (throat) of the horn.

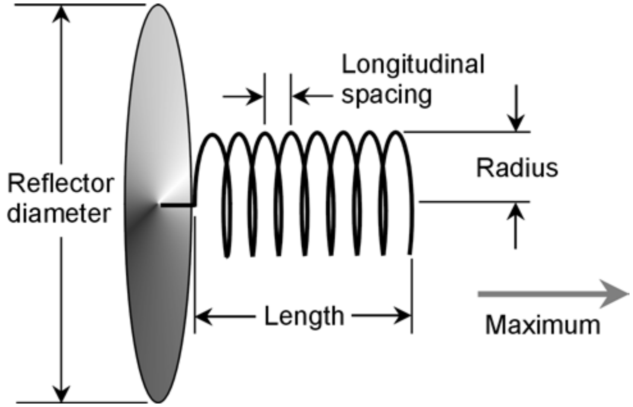

27-12 A helical antenna with a flat reflector.

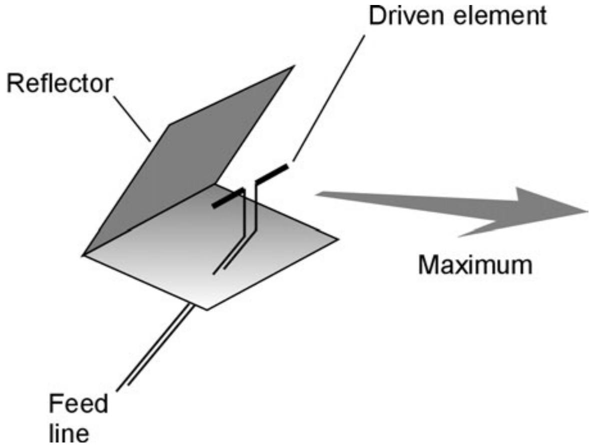

27-13 A corner reflector with a dipole as the driven element.

Horns are sometimes used all by themselves, but they are also used to feed large dish antennas. This optimizes the f/s ratio by minimizing extraneous radiation and response that occurs if a dipole is used as the driven element for the dish.

Safety

Antennas should never be placed in such a way that they can fall or blow down on power lines. Also, it should not be possible for power lines to fall or blow down on an antenna.

Wireless equipment having outdoor antennas should not be used during thundershowers, or when lightning is anywhere in the vicinity. Antenna construction and maintenance should never be undertaken when lightning is visible or thunder can be heard, even if a storm appears to be many miles away. Ideally, antennas should be disconnected from electronic equipment, and connected to a substantial earth ground, at all times when the equipment is not in use.

Tower and antenna climbing is a job for professionals. Under no circumstances should an inexperienced person attempt to climb such a structure.

Indoor transmitting antennas can expose operating personnel to EM field energy. The extent of the hazard, if any, posed by such exposure has not been firmly established. However, there is sufficient concern to warrant checking the latest publications on the topic.

For detailed information concerning antenna safety, consult a professional antenna engineer and/or a comprehensive text on antenna design and construction.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. Imagine a center-fed, straight wire antenna in free space, whose length can be varied at will. At which of the following lengths is the radiation resistance at the feed point the highest?

(a) $\lambda/4$

(b) λ

(c) λ/3
(d) λ/2

2. What is the approximate length, in meters (m), of a center-fed, λ/2 wire antenna at a frequency of 18.1 MHz? Assume the velocity factor is typical for wire.
   (a) 7.90 m
   (b) 8.29 m
   (c) 13.6 m
   (d) 25.9 m

3. What is the approximate height, in feet (ft), of a ground-mounted, λ/4 vertical antenna, made of aluminum tubing, at a frequency of 7.025 MHz? Assume the velocity factor is 92 percent.
   (a) 64.43 ft
   (b) 32.22 ft
   (c) 19.64 ft
   (d) 9.822 ft

4. In order to obtain reasonable efficiency with a λ/4 vertical antenna mounted over poorly conducting soil at a frequency of 10 MHz, it is necessary to
   (a) make the antenna out of large-diameter aluminum or copper tubing.
   (b) suspend the antenna from a kite or balloon.
   (c) feed the antenna with a waveguide.
   (d) minimize the loss resistance by installing numerous ground radials.

5. A driven element in a parasitic array is
   (a) grounded for dc.
   (b) a loop with a circumference of λ/2.
   (c) connected to the feed line.
   (d) a straight length of wire.

6. In a zepp antenna, feed-line radiation can be minimized by
   (a) using the antenna only at frequencies where its length is a whole-number multiple of λ/2.
   (b) using the antenna only at frequencies where its length is a whole-number multiple of λ/4.
   (c) using the antenna only at frequencies where its length is an even-number multiple of λ/10.
   (d) using the antenna only at frequencies where its length is an odd-number multiple of λ/4.

7. Consider a single-turn, closed loop having a circumference of 95 ft. Suppose this loop is circular, lies entirely in a single plane, is fed with open-wire transmission line and an antenna tuner, and is operated at 10 MHz. The maximum radiation and response for this loop can be expected to occur
   (a) in the plane containing the loop.
   (b) along a line perpendicular to the plane containing the loop.

(c) along a line coinciding with the path of the feed line.
(d) equally in all directions in three-dimensional space.

8. Suppose an antenna system has a radiation resistance of 40 Ω at 8.5 MHz. What is the efficiency of this system?
(a) 25 percent
(b) 40 percent
(c) 80 percent
(d) Impossible to determine without more information

9. Suppose an antenna system has a radiation resistance of 40 Ω and a loss resistance of 10 Ω. What is the efficiency of this system?
(a) 25 percent
(b) 40 percent
(c) 80 percent
(d) Impossible to determine without more information

10. The null in a loopstick antenna exists
(a) in the plane perpendicular to the coil axis.
(b) in the line that coincides with the coil axis.
(c) at a 45° angle with respect to the coil axis.
(d) nowhere, because a loopstick is an isotropic antenna.

11. As elements are added to a properly designed Yagi antenna,
(a) the forward gain decreases.
(b) the f/b ratio decreases.
(c) the f/s ratio increases.
(d) None of the above apply.

12. When the radiation resistance in an antenna system represents most of the total system resistance,
(a) the system can be expected to have high loss.
(b) the system can be expected to have high efficiency.
(c) the system can be expected to have a high f/s ratio.
(d) the system can be expected to have a high f/b ratio.

13. What state of affairs is optimal in a waveguide?
(a) A colony of spiders inside
(b) A cross-sectional measurement of less than $\lambda/4$
(c) Operation at frequencies below 3 MHz
(d) None of the above

14. Consider a single-turn, closed loop having a circumference of 65 ft. Suppose this loop is circular, lies entirely in a single plane, is fed with open-wire transmission line and an antenna tuner, and is operated at 7 MHz. The maximum radiation and response for this loop can be expected to occur

(a) in the plane containing the loop.
(b) along a line perpendicular to the plane containing the loop.
(c) along a line coinciding with the path of the feed line.
(d) equally in all directions in three-dimensional space.

15. Suppose the total resistance in an antenna system is 80 Ω, and the loss resistance is 20 Ω. The efficiency of the system is

(a) 80 percent.
(b) 75 percent.
(c) 25 percent.
(d) impossible to determine without more information.

16. In a parasitic array, a director is usually

(a) tuned to a slightly higher frequency than the driven element.
(b) connected to a system of ground radials.
(c) a closed loop that is grounded for RF.
(d) a straight length of wire or tubing.

17. A J pole is a modified form of

(a) ground-plane antenna.
(b) folded dipole antenna.
(c) quad antenna.
(d) zepp antenna.

18. Which of the following antenna types is well suited to use with parallel-wire transmission line?

(a) The ground plane
(b) The ground-mounted vertical
(c) The folded dipole
(d) None of the above

19. Which of the following antenna types is not designed for transmitting at 10 GHz?

(a) A horn antenna
(b) A dish antenna
(c) A zepp antenna
(d) A helical antenna

20. A vertical antenna measuring λ/2 in height does not require a ground system with low resistance because

    (a) its loss resistance is high.
    (b) its radiation resistance is low.
    (c) its main lobe is sharp.
    (d) None of the above apply.

28
CHAPTER

Integrated Circuits

INTEGRATED CIRCUITS (ICs), ALSO CALLED *CHIPS*, HAVE STIMULATED AS MUCH EVOLUTION AS ANY single development in the history of electronic technology. Most ICs look like gray or black plastic boxes with protruding pins. The schematic symbols are triangles or rectangles with lines emerging from them, labeled according to their functions.

Advantages of IC Technology

Integrated circuits have advantages over *discrete components* (individual transistors, diodes, capacitors, and resistors). Here are some of the most important considerations.

Compactness

An obvious asset of IC design is economy of space. They are far more compact than equivalent circuits made from discrete components. A corollary to this is the fact that more complex circuits can be built, and kept down to a reasonable size, using ICs as compared with discrete components.

High Speed

The interconnections among the components within an IC are physically tiny, making high switching speeds possible. Electric currents travel fast, but not instantaneously. The less time charge carriers need to get from component *X* to component *Y*, in general, the more computations are possible within a given span of time, and the less time is needed for complex operations.

Low Power Consumption

Another advantage of ICs is the fact that they use less power than equivalent discrete-component circuits. This is crucial if batteries are to be used for operation. Because ICs use so little current, they produce less heat than their discrete-component equivalents. This translates into better efficiency. It also minimizes the problems that bedevil equipment that gets hot with use, such as frequency drift and generation of internal noise.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

Reliability

Integrated circuits fail less often, per component-hour of use, than systems that use discrete components. This is because all the component interconnections are sealed within the IC case, preventing corrosion or the intrusion of dust. The reduced failure rate translates into less *downtime*, or time during which the equipment is out of service for repairs.

Ease of Maintenance

Integrated-circuit technology lowers maintenance costs. Repair procedures are simplified when failures occur. Many appliances use sockets for ICs, and replacement is simply a matter of finding the faulty IC, unplugging it, and plugging in a new one. Special desoldering equipment is used with appliances having ICs soldered directly to the circuit boards.

Modular Construction

Modern IC appliances use *modular construction.* In this scheme, individual ICs perform defined functions within a circuit board; the circuit board or *card*, in turn, fits into a socket and has a specific purpose. Computers, programmed with customized software, are used by repair technicians to locate the faulty card in an appliance. The whole card can be pulled and replaced, getting the appliance back to the consumer in the shortest possible time. Then a computer can be used to troubleshoot the faulty card, getting the card ready for use in the next appliance that happens to come along with a failure in the same card.

Limitations of IC Technology

No technological advancement ever comes without a downside. Integrated circuits have some limitations that must be considered when designing an electronic system.

Inductors Impractical

While some components are easy to fabricate onto chips, other components defy the IC manufacturing process. Inductances, except for extremely low values, are an example. Devices using ICs must generally be designed to work with discrete inductors, external to the ICs themselves. However, resistance-capacitance (RC) circuits are capable of doing most things that inductance-capacitance (LC) circuits can do, and these can exist inside an IC.

High Power Impossible

The small size and low current consumption of ICs comes with an inherent limitation: high-power amplifiers cannot, in general, be fabricated onto chips. High power necessitates a certain amount of physical mass and volume. High-power devices and systems invariably generate large amounts of heat, which must be efficiently radiated or conducted away.

Linear ICs

A *linear IC* is used to process *analog signals* such as voices, music, and radio transmissions.

The term *linear* arises from the fact that, in general, the amplification factor is constant as the input amplitude varies. That is, the output signal strength is a *linear function* of the input signal strength (Fig. 28-1).

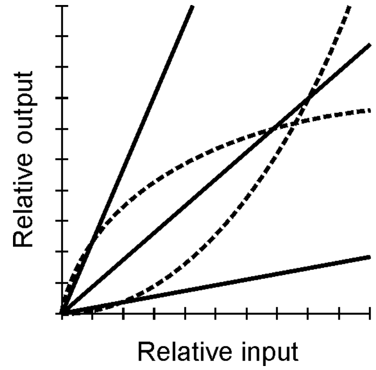

28-1 In a linear IC, the relative output is a linear (straight-line) function of the relative input. The solid lines show examples of linear IC characteristics. The dashed curves show functions not characteristic of properly operating linear ICs.

Operational Amplifier

An *operational amplifier*, or *op amp*, is a specialized linear IC that consists of several bipolar transistors, resistors, diodes, and capacitors, interconnected so that high gain is possible over a wide range of frequencies. An op amp might comprise an entire IC. Or, an IC might consist of two or more op amps. Thus, you'll sometimes hear of *dual op amps* or *quad op amps.* Some ICs have op amps in addition to other circuits.

An op amp has two inputs, one *noninverting* and one *inverting*, and one output. When a signal goes into the noninverting input, the output is in phase with the input; when a signal goes into the inverting input, the output is 180° out of phase with the input. An op amp has two power supply connections, one for the emitters of the transistors (V_{ee}) and one for the collectors (V_{cc}). The usual schematic symbol for an op amp is a triangle. The inputs, output, and power-supply connections are drawn as lines emerging from it (Fig. 28-2).

The gain of an op amp is determined by one or more external resistors. Normally, a resistor is connected between the output and the inverting input. This is called the *closed-loop configuration.* The feedback is negative (out of phase), causing the gain of the op amp to be less than it would be if there were no feedback (the *open-loop configuration*). Figure 28-3 is a schematic diagram of a noninverting closed-loop amplifier.

The reason for providing negative feedback in an op amp circuit is the fact that without it, the gain may be too great. Excessive amplifier gain can cause problems. Open-loop op amps are prone to instability, especially at low frequencies. They also generate a lot of internal noise.

When an RC combination is used in the feedback loop of an op amp, the gain depends on the input-signal frequency. It is possible to get a *lowpass response*, a *highpass response*, a *resonant peak*, or a *resonant notch* using an op amp and various RC feedback arrangements. These four types of responses are shown in Fig. 28-4.

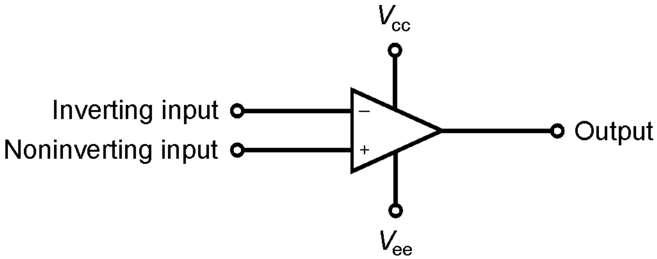

28-2 Schematic symbol for an op amp. Connections are discussed in the text.

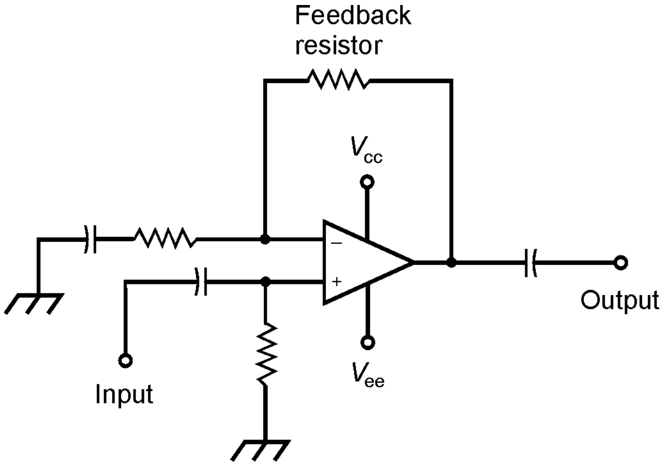
Feedback resistor
V_{cc}
–
+
V_{ee}
Input
Output

28-3 A closed-loop op amp circuit with negative feedback. If the feedback resistor is removed, it becomes an open-loop circuit.

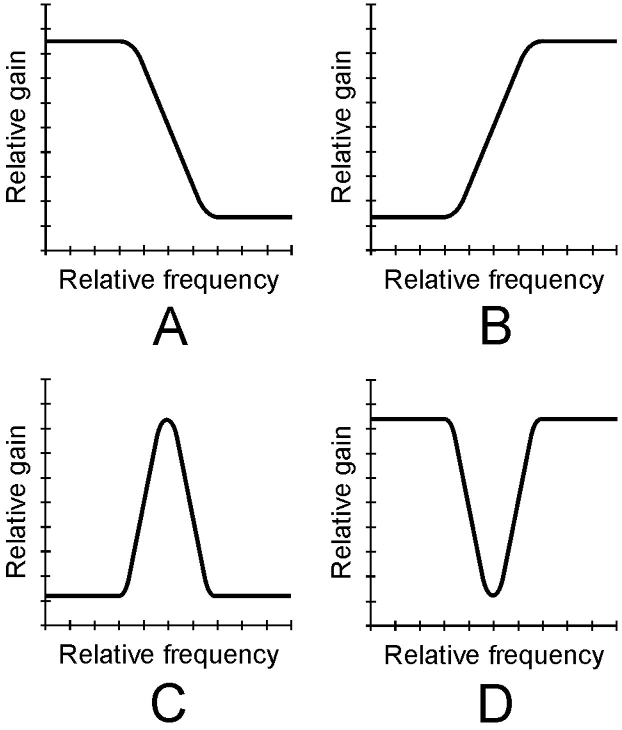

28-4 Gain-versus-frequency response curves.
At A, lowpass; at B, highpass; at C, resonant peak; at D, resonant notch.

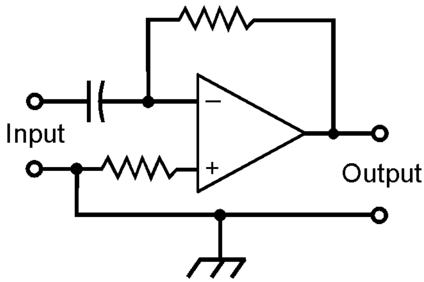
–
+
Input
Output

28-5 A differentiator circuit that uses an op amp.

Op Amp Differentiator

A *differentiator* is an electronic circuit whose instantaneous output amplitude is proportional to the rate at which the input amplitude changes. The circuit mathematically *differentiates* the input signal. Op amps can be used as differentiator circuits. An example is shown in Fig. 28-5.

When the input to a differentiator is a constant dc voltage, the output is zero (no signal). When the input amplitude is increasing, the output is a positive dc voltage. When the input decreases, the output is a negative dc voltage. If the input waveform fluctuates periodically (the usual case), the output varies according to the instantaneous rate of change of the input amplitude. This results in an output signal with the same frequency as that of the input signal, although the waveform is often quite different. A pure sine wave input produces a pure sine wave at the output, but the phase is shifted 90° to the left (that is, $\frac{1}{4}$ cycle earlier in time). Complex input waveforms can produce a wide variety of outputs from a differentiator.

Op Amp Integrator

An *integrator* is an electronic circuit whose instantaneous output amplitude is proportional to the accumulated input signal amplitude as a function of time. The circuit mathematically *integrates* the input signal. The function of an integrator is basically the inverse, or opposite, of the function of a differentiator circuit. Figure 28-6 shows how an op amp can be connected to obtain an integrator.

If an integrator circuit is supplied with an input signal waveform that fluctuates periodically (the usual case), the output voltage varies according to the *integral*, or *antiderivative*, of the input voltage. This results in an output signal with the same frequency as that of the input signal, although the waveform is likely to be different. A pure sine wave input produces a pure sine wave output, but the phase is shifted 90° to the right (that is, $\frac{1}{4}$ cycle later in time). Complex input waveforms can produce many types of output waveforms.

An indefinite rise, either negatively or positively, in output voltage cannot occur in a practical integrator. If the mathematical integral (in pure theory) of an input function is an endlessly increasing output function, the actual output voltage rises to a certain maximum, either positive or negative, and stays there. This maximum is less than or equal to the power supply or battery voltage.

Voltage Regulator

A *voltage regulator IC* acts to control the output voltage of a power supply. This is important with precision electronic equipment. These ICs are available with various voltage and current ratings. Typical voltage regulator ICs have three terminals, and because of this, they are sometimes mistaken for power transistors.

28-6 An integrator circuit that uses an op amp.

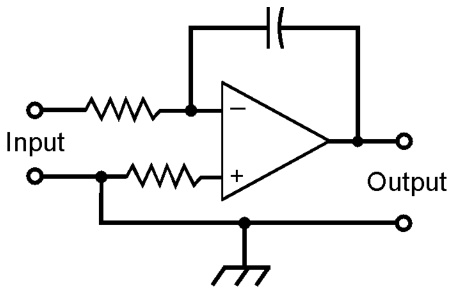

Timer

A *timer IC* is a specialized oscillator that produces a delayed output. The delay is adjustable to suit the needs of a particular device. The delay is generated by counting the number of oscillator pulses; the length of the delay can be adjusted by means of external resistors and capacitors. Timers are commonly used in circuits such as digital frequency counters, where a precise time interval or window must be provided.

Multiplexer

A *multiplexer IC* allows several different signals to be combined in a single channel by means of a process called *multiplexing*. An analog multiplexer can also be used in reverse; then it works as a *demultiplexer*. Thus, you'll sometimes hear or read about a *multiplexer/demultiplexer IC*.

Comparator

A *comparator IC* has two inputs. It compares the voltages at the two inputs, which are called A and B. If the voltage at input A is significantly higher than the voltage at input B, the output is about +5 V. This is logic 1, or high. If the voltage at input A is lower than or equal to the voltage at input B, the output voltage is about +2 V. This is designated as logic 0, or low.

Voltage comparators are available for a variety of applications. Some can switch between low and high states at a rapid rate of speed, while others are slow. Some have low input impedance, and others have high impedance. Some are intended for AF or low-frequency RF use; others are fabricated for video or high-frequency RF applications. Voltage comparators can be used to actuate, or *trigger*, other devices such as relays and electronic switching circuits.

Digital ICs

A *digital IC*, also sometimes called a *digital-logic IC*, operates using two discrete states: high (logic 1) and low (logic 0). Digital logic is discussed in Chap. 26. Digital ICs contain massive arrays of logic gates that perform Boolean operations at high speed.

Transistor-Transistor Logic

In *transistor-transistor logic* (TTL), arrays of bipolar transistors, some with multiple emitters, operate on dc pulses. This technology has several variants, some of which date back to around 1970. A basic TTL gate is illustrated in Fig. 28-7. This gate uses two NPN bipolar transistors, one of which is a *dual-emitter* device. The transistors are always either completely cut off, or else completely saturated. Because of this, TTL is relatively immune to external noise.

Emitter-Coupled Logic

Another bipolar-transistor logic form is known as *emitter-coupled logic* (ECL). In an ECL device, the transistors are not operated at saturation, as they are with TTL. This increases the speed. But noise pulses have a greater effect in ECL, because unsaturated transistors are sensitive to, and can actually amplify, external signals and noise. The schematic of Fig. 28-8 shows a basic ECL gate using four NPN bipolar transistors.

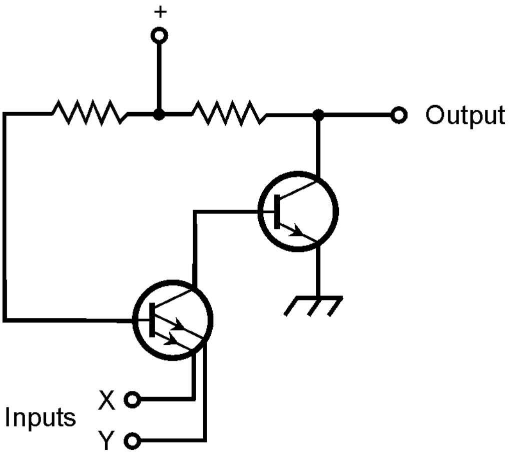

28-7 A transistor-transistor logic (TTL) gate. In this gate, two NPN bipolar transistors are used. Note that one of the transistors has two emitters.

Metal-Oxide-Semiconductor Logic

Digital ICs can also be constructed using metal-oxide-semiconductor (MOS) devices. *N-channel MOS* (NMOS) *logic*, pronounced "EN-moss logic," offers simplicity of design, along with high operating speed. *P-channel MOS logic*, pronounced "PEA-moss logic," is similar to NMOS logic, but the speed is slower. An NMOS or PMOS digital IC is like a circuit that uses only N-channel MOSFETs, or only P-channel MOSFETs, respectively.

Complementary-metal-oxide-semiconductor (CMOS) *logic*, pronounced "SEA-moss logic," employs both N-type and P-type silicon on a single chip. This is analogous to using both N-channel

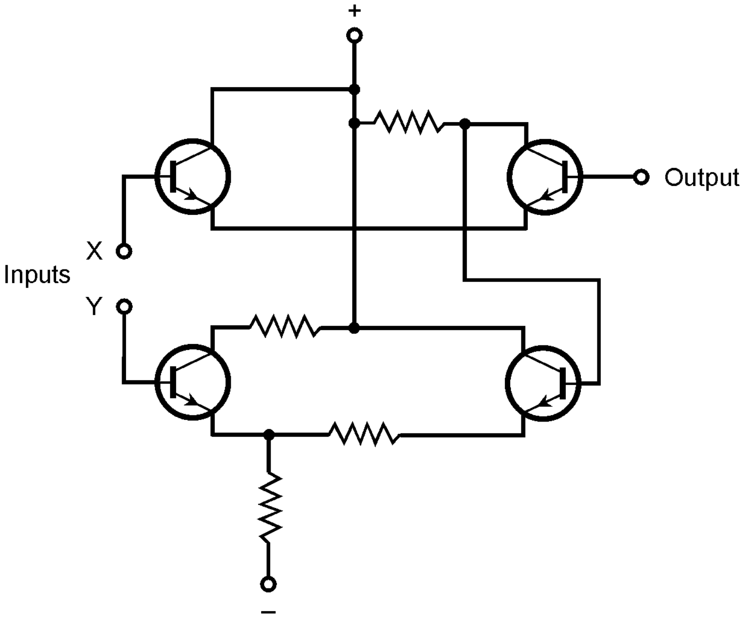

28-8 An emitter-coupled logic (ECL) gate using four NPN bipolar transistors.

and P-channel MOSFETs in a circuit. The advantages of CMOS technology include extremely low current drain, high operating speed, and immunity to noise.

All forms of MOS logic ICs require care in handling to prevent destruction by electrostatic discharges. The precautions are the same as those that are required when handling MOSFETs. All technical personnel who work with the devices should be grounded. This can be achieved by having technicians wear metal wrist straps connected to a good electrical ground, and by ensuring that the relative humidity in the lab is not allowed to get too low. When MOS ICs are stored, the pins should be pushed into special conductive foam that is manufactured for that purpose.

Component Density

The number of elements per chip in an IC is called the *component density.* There has been a steady increase in the number of components that can be fabricated on a single chip. There is an absolute limit on component density, imposed by the atomic structure of the semiconductor material—literally, the size of the atoms! Technology has begun to approach that barrier.

Small Scale

In *small scale integration* (SSI), there are fewer than 10 transistors on a chip. These types of ICs can carry the largest currents, and can be useful in voltage regulation and other moderate-power applications.

Medium Scale

In *medium scale integration* (MSI), there are 10 to 100 transistors per chip. This allows for considerable miniaturization, but it is not a high level of component density, relatively speaking. An advantage of MSI (in a few applications) is that fairly large currents can be carried by the individual gates. Both bipolar and MOS technologies can be adapted to MSI.

Large Scale

In *large scale integration* (LSI), there are 100 to 1000 transistors per semiconductor chip. This is an *order of magnitude* (a factor of 10 times) more dense than MSI. Electronic wristwatches, single-chip calculators, and simple microcomputers are examples of devices using LSI ICs.

Very-Large-Scale Devices

Very-large-scale integration (VLSI) devices have from 1000 to 1,000,000 transistors per chip. This can be up to three orders of magnitude more dense than LSI. Microcomputers and *memory* ICs are made using VLSI.

Ultra-Large-Scale Devices

You might sometimes hear of *ultra-large-scale integration* (ULSI). Devices of this kind have more than 1,000,000 transistors per chip. The principal uses for this technology are in the fields of high-level computing, supercomputing, robotics, and artificial intelligence (AI).

IC Memory

Binary digital data, in the form of high and low levels (logic ones and zeros), can be stored in *memory ICs.* These devices can take various physical forms.

Random Access

A *random access memory* (RAM) chip stores binary data in *arrays.* The data can be *addressed* (selected) from anywhere in the matrix. Data is easily changed and stored back in RAM, in whole or in part. A RAM chip is sometimes called a *read/write memory.*

An example of the use of RAM is a word-processing computer file that you are actively working on. This paragraph, this chapter, and in fact the whole text of this book was written in semiconductor RAM in small sections before being incrementally stored on the computer hard drive, and ultimately on external media.

There are two major categories of RAM: *dynamic RAM* (DRAM) and *static RAM* (SRAM). A DRAM chip contains transistors and capacitors, and data is stored as charges on the capacitors. The charge must be replenished frequently, or it will be lost through discharge. Replenishing is done automatically several hundred times per second. An SRAM chip uses a flip-flop to store the data. This gets rid of the need for constant replenishing of charge, but the tradeoff is that SRAM ICs require more elements than DRAM chips to store a given amount of data.

With any RAM chip, the data in it will vanish when power is removed, unless some provision is made for *memory backup.* The most common means of memory backup is the use of a small cell or battery with a long shelf life. Modern IC memories need so little current to store their data that a *backup battery* lasts as long in the circuit as it would on the shelf. Memory that disappears when power is removed is called *volatile memory.* If memory is retained when power is removed, it is *nonvolatile.*

Read Only

By contrast to RAM chips, the data in a *read-only memory* (ROM) chip can be easily accessed, in whole or in part, but not easily written over. A standard ROM chip is *programmed* at the factory. This permanent programming is known as *firmware.* But there are also ROM chips that you can program and reprogram yourself. The data contents of ROM chips are generally nonvolatile. That means a power failure is no cause for concern.

An *erasable programmable ROM* (EPROM) chip is an IC whose memory is of the read-only type, but that can be reprogrammed by a certain procedure. It is more difficult to rewrite data in an EPROM than in a RAM; the usual process for erasure involves exposure to ultraviolet (UV). An EPROM IC can be recognized by the presence of a transparent window with a removable cover, through which the UV is focused to erase the data. The IC must be taken from the circuit in which it is used, exposed to UV for several minutes, and then reprogrammed.

The data in some EPROM chips can be erased by electrical means. Such an IC is called an *electrically erasable programmable read-only memory* (EEPROM) chip. These do not have to be removed from the circuit for reprogramming.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. Because of the small size of ICs compared with equivalent circuits made from discrete components,
   - (a) more heat is generated.
   - (b) higher power output is possible.
   - (c) higher switching speeds are attainable.
   - (d) fewer calculations need to be done per unit time.
2. Which of the following is not an advantage of ICs over discrete components?
   - (a) Higher component density
   - (b) Ease of maintenance
   - (c) Lower gain
   - (d) Lower current consumption
3. In which of the following devices would you never find an IC as the main component?
   - (a) The final amplifier in a large radio broadcast transmitter
   - (b) The microprocessor in a personal computer
   - (c) A battery-powered calculator
   - (d) A low-power audio amplifier
4. Which, if any, of the following component types is not practical for direct fabrication on a chip?
   - (a) Resistors
   - (b) Capacitors
   - (c) Diodes
   - (d) All three of the above component types are practical for direct fabrication on a chip.
5. An op amp circuit usually employs negative feedback to
   - (a) maximize the gain.
   - (b) control the gain.
   - (c) allow oscillation over a wide band of frequencies.
   - (d) Forget it! Op amp circuits never have negative feedback.
6. Suppose a communications channel carries several signals at once. Which type of IC might be used to single one of the signals out for reception?
   - (a) An op amp
   - (b) A timer
   - (c) A comparator
   - (d) A multiplexer/demultiplexer
7. Which type of IC is commonly used to determine whether two dc voltages are the same or different?
   - (a) An op amp
   - (b) A timer

(c) A comparator
(d) A multiplexer/demultiplexer

8. Which type of digital IC is least susceptible to noise?
(a) Transistor-transistor logic
(b) Base-coupled logic
(c) Emitter-coupled logic
(d) N-channel-coupled logic

9. Which of the following is *not* a characteristic of CMOS chips?
(a) Sensitivity to damage by electrostatic discharge
(b) Low current demand
(c) Ability to work at high speed
(d) Ability to handle extremely high power

10. An absolute limit to the component density that can be achieved in IC technology is determined by
(a) the maximum current levels needed.
(b) the maximum attainable impedance.
(c) the size of the semiconductor atoms.
(d) Forget it! There is no limit to the component density that can be achieved in IC technology.

11. With a ROM chip,
(a) it is easier to read data from the device than to write data into it.
(b) it is more difficult to read data from the device than to write data into it.
(c) it is easy to read data from the device, and just as easy to write data into it.
(d) it is difficult to read data from the device, and equally difficult to write data into it.

12. With a RAM chip,
(a) it is easier to read data from the device than to write data into it.
(b) it is more difficult to read data from the device than to write data into it.
(c) it is easy to read data from the device, and just as easy to write data into it.
(d) it is difficult to read data from the device, and equally difficult to write data into it.

13. Which type of IC must be physically removed from the circuit to have its memory contents changed?
(a) An EEPROM chip
(b) An EPROM chip
(c) An SRAM chip
(d) A DRAM chip

14. With respect to memory, which of the following terms means that the data contents vanish if power is removed from a chip?
    (a) Volatility
    (b) Component density
    (c) Multiplexing
    (d) Corruptibility

15. In terms of the maximum number of transistors on a single chip, how many orders of magnitude larger is VLSI than MSI?
    (a) Two
    (b) Three
    (c) Four
    (d) Forget it! The maximum number of transistors on a VLSI chip is not larger, but smaller than the maximum number of transistors on an MSI chip.

16. Which of the following illustrations in Fig. 28-9 shows the input and output of a differentiator circuit?
    (a) A
    (b) B
    (c) C
    (d) None of them

17. Which of the following illustrations in Fig. 28-9 shows the input and output of an integrator circuit?
    (a) A
    (b) B
    (c) C
    (d) None of them

18. Flip-flops are used to store data in
    (a) timers.
    (b) op amps.
    (c) multiplexers.
    (d) none of the above.

19. Capacitors are used to store data in
    (a) timers.
    (b) op amps.
    (c) multiplexers.
    (d) none of the above.

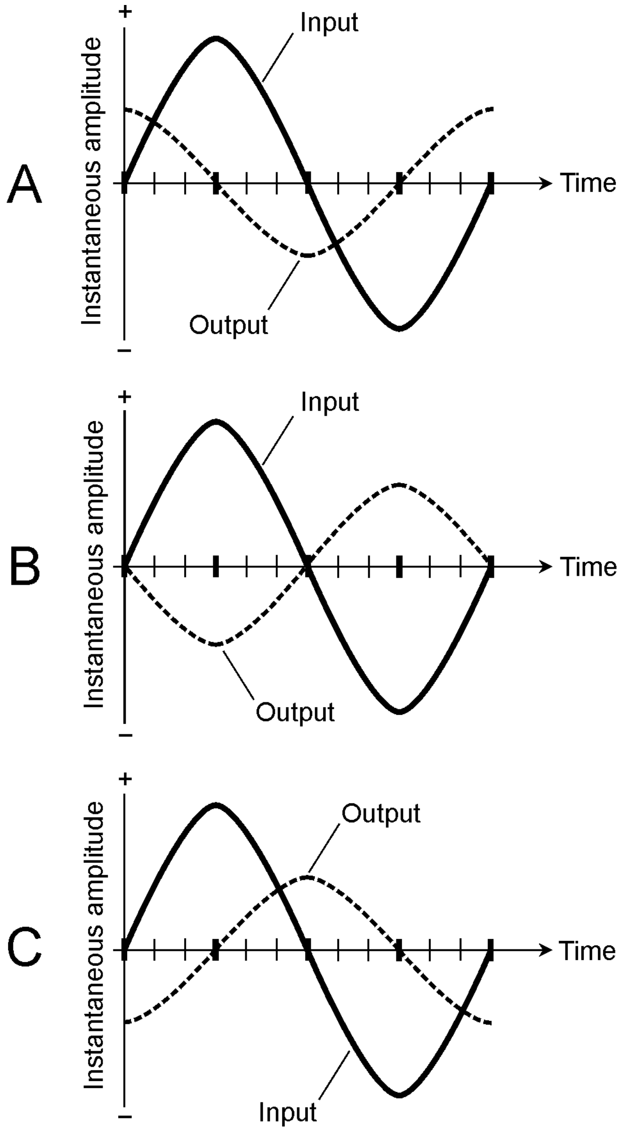

28-9 Illustration for Quiz Questions 16 and 17.

20. The phase of a sine wave signal can be shifted 180° by
    (a) applying the signal to the input of a timer.
    (b) applying the signal to the inverting input of an op amp.
    (c) applying the signal to the input of a flip-flop.
    (d) applying the signal to the input of an SRAM chip.

29
CHAPTER

Electron Tubes

ELECTRON TUBES, ALSO CALLED *TUBES* OR *VALVES* (IN ENGLAND), ARE USED IN SOME ELECTRONIC equipment. In a tube, the *charge carriers* are free electrons that travel through space between electrodes inside the device. This makes tubes fundamentally different from semiconductor devices, in which charge carriers move among atoms in a solid medium.

Tube Forms

There are two basic types of electron tube: the *vacuum tube* and the *gas-filled tube.* As their names imply, vacuum tubes have virtually all the gases removed from their envelopes. Gas-filled tubes contain elemental vapor at low pressure.

Vacuum Tube

Vacuum tubes accelerate electrons to high speeds, resulting in large currents. This current can be focused into a beam and guided in a particular direction. The intensity and/or beam direction can be changed with extreme rapidity, producing effects such as rectification, detection, oscillation, amplification, signal mixing, waveform displays, spectral displays, and video imaging.

Gas-Filled Tube

Gas-filled tubes have a constant voltage drop, no matter what the current. This makes them useful as voltage regulators for high-voltage, high-current power supplies. Gas-filled tubes can withstand conditions that would destroy semiconductor regulating devices. Gas-filled tubes emit infrared (IR), visible light, and/or ultraviolet (UV). This property can be put to use for decorative lighting. A small *neon bulb* can be employed to construct an AF *relaxation oscillator* (Fig. 29-1).

Diode Tube

Even before the year 1900, scientists knew that electrons could carry electric current through a vacuum. They also knew that hot electrodes emit electrons more easily than cool ones. These phenom-

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

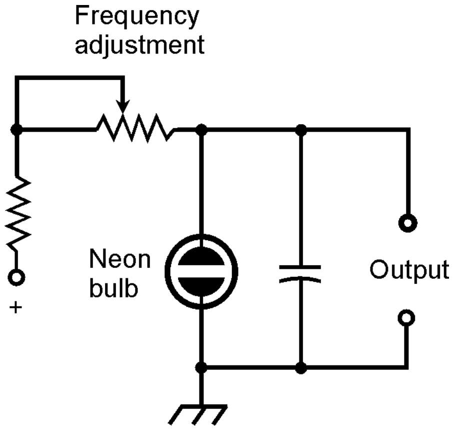

29-1 A neon bulb oscillator, also known as a relaxation oscillator.

ena were put to use in the first electron tubes, known as *diode tubes*, for the purpose of rectification. Diode tubes are rarely used nowadays, although they can still be found in some power supplies that are required to deliver several thousand volts for long periods at a 100 percent *duty cycle* (that is, continuous operation).

Electrodes in a Tube

In a tube, the electron-emitting electrode is the *cathode.* The cathode is usually heated by means of a wire *filament,* similar to the glowing element in an incandescent bulb. The heat drives electrons from the cathode. The cathode of a tube is analogous to the source of an FET, or to the emitter of a bipolar transistor. The electron-collecting electrode is the *anode,* also called the *plate.* The plate is the tube counterpart of the drain of an FET or the collector of a bipolar transistor. In most tubes, intervening *grids* control the flow of electrons from the cathode to the plate. The grids are the counterparts of the gate of an FET or the base of a bipolar transistor.

Directly Heated Cathode

In some tubes, the filament also serves as the cathode. This type of electrode is called a *directly heated cathode.* The negative supply voltage is applied directly to the filament. The filament voltage for most tubes is 6 V or 12 V dc. It is important that dc be used to heat the filament in this type of tube, because ac will tend to modulate the output. The schematic symbol for a diode tube with a directly heated cathode is shown in Fig. 29-2A.

Indirectly Heated Cathode

In many types of tubes, the filament is enclosed within a cylindrical cathode, and the cathode gets hot from the IR radiated by the filament. This is an *indirectly heated cathode.* The filament normally receives 6 V or 12 V ac or dc. In an indirectly heated cathode arrangement, ac does not cause modulation problems, as it can with a directly heated cathode tube. The schematic symbol for a diode tube with an indirectly heated cathode is shown in Fig. 29-2B.

Because the electron emission in a tube depends on the filament or *heater,* tubes need a certain amount of time to warm up before they can operate properly. This time can vary from a few seconds

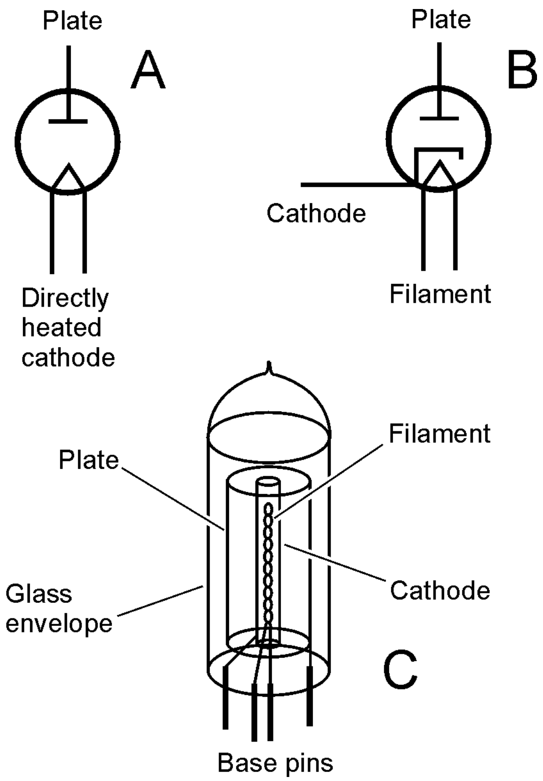

29-2 At A, schematic symbol for diode tube with directly heated cathode. At B, symbol for diode tube with indirectly heated cathode. At C, simplified rendition of the construction of a diode tube.

(for a small tube with a directly heated cathode) to a couple of minutes (for massive power-amplifier tubes with indirectly heated cathodes). The warm-up time for a large tube is about the same as the boot-up time for a personal computer.

Cold Cathode

In a gas-filled tube, the cathode does not have a filament to heat it. Such an electrode is called a *cold cathode.* Various chemical elements are used in gas-filled tubes. In fluorescent devices, neon, argon, and xenon are common. In gas-filled voltage-regulator (VR) tubes, mercury vapor is used. In a mercury-vapor VR tube, the warm-up period is the time needed for the elemental mercury, which is a liquid at room temperature, to vaporize (approximately 2 minutes).

Plate

The plate, or anode, of a tube is a cylinder concentric with the cathode and filament, as shown in Fig. 29-2C. The plate is connected to the positive dc supply voltage. Tubes operate at plate voltages ranging from 50 V to more than 3 kV. These voltages are potentially lethal. Technicians unfamiliar with vacuum tubes should not attempt to service equipment that contains them. The output of a tube-type amplifier circuit is almost always taken from the plate circuit. The plate exhibits high impedance for signal output, similar to that of an FET.

Control Grid

The flow of current can be controlled by means of an electrode between the cathode and the plate. This electrode, the *control grid* (or simply the *grid*), is a wire mesh or screen that lets electrons pass through. The grid impedes the flow of electrons if it is provided with a negative voltage relative to the cathode. The greater the negative *grid bias*, the more the grid obstructs the flow of electrons through the tube.

Triode Tube

A tube with one grid is a *triode.* The schematic symbol is shown at A in Fig. 29-3. In this case the cathode is indirectly heated; the filament is not shown. This omission is standard in schematics showing tubes with indirectly heated cathodes. When the cathode is directly heated, the filament symbol serves as the cathode symbol. The control grid is usually biased with a negative dc voltage ranging from near 0 to as much as half the positive dc plate voltage.

Tetrode Tube

A second grid can be added between the control grid and the plate. This is a spiral of wire or a coarse screen, and is called the *screen grid* or *screen.* This grid normally carries a positive dc voltage at 25 to 35 percent of the plate voltage.

The screen grid reduces the capacitance between the control grid and plate, minimizing the tendency of a tube amplifier to oscillate. The screen grid can also serve as a second control grid, allowing two signals to be injected into a tube. This tube has four elements, and is known as a *tetrode.* Its schematic symbol is shown at B in Fig. 29-3.

Pentode Tube

The electrons in a tetrode can bombard the plate with such force that some of them bounce back, or knock other electrons from the plate. This so-called secondary emission can hinder tube perfor-

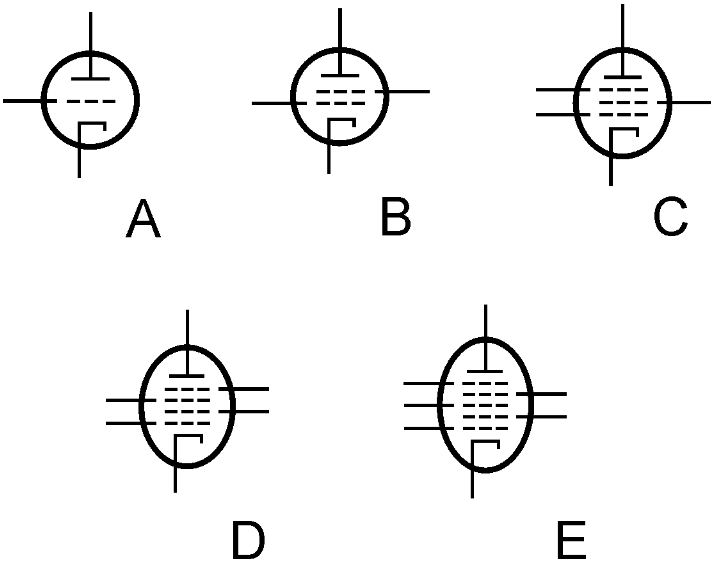

29-3 Schematic symbols for vacuum tubes with grids: triode (A), tetrode (B), pentode (C), hexode (D), and heptode (E).

mance and, at high power levels, cause screen current so high that the electrode is destroyed. This problem can be eliminated by placing another grid, called the *suppressor grid* or *suppressor*, between the screen and the plate. The suppressor repels *secondary electrons* emanating from the plate, preventing most of them from reaching the screen. The suppressor also reduces the capacitance between the control grid and the plate more than a screen grid by itself.

Greater gain and stability are possible with a *pentode*, or tube with five elements, than with a tetrode or triode. The schematic symbol for a pentode is shown at C in Fig. 29-3. The suppressor is often connected directly to the cathode.

Hexode and Heptode Tubes

In some older radio and TV receivers, tubes with four or five grids were sometimes used. These tubes had six and seven elements, respectively, and were called *hexode* and *heptode*. The usual function of such tubes was signal mixing. The schematic symbol for a hexode is shown at D in Fig. 29-3; the symbol for a heptode is at E. You will not encounter hexodes and heptodes in modern electronics, because solid-state components are used for signal mixing.

Interelectrode Capacitance

In a vacuum tube, the cathode, grid(s), and plate exhibit *interelectrode capacitance* that is the primary limiting factor on the frequency range in which the device can produce gain. The interelectrode capacitance in a typical tube is a few picofarads. This is negligible at low frequencies, but at frequencies above approximately 30 MHz, it becomes a significant consideration. Vacuum tubes intended for use as RF amplifiers are designed to minimize this capacitance.

Circuit Configurations

The most common application of vacuum tubes is in amplifiers, especially in radio and television transmitters at power levels of more than 1 kW. Some high-fidelity audio systems also employ vacuum tubes. In recent years, tubes have gained favor with some popular music bands. Some musicians insist that "tube amps" provide richer sound than amplifiers using power transistors. There are two basic vacuum-tube amplifier circuit arrangements: the *grounded-cathode* configuration and the *grounded-grid* configuration.

Grounded Cathode

Figure 29-4 is a simplified schematic diagram of a grounded-cathode circuit using a triode tube. This circuit is the basis for many tube-type RF power amplifiers and audio amplifiers. The input impedance is moderate, and the output impedance is high. Impedance matching between the amplifier and the load can be obtained by tapping a coil in the output circuit, or by using a transformer.

Grounded Grid

Figure 29-5 shows a basic grounded-grid RF amplifier circuit. The input impedance is low, and the output impedance is high. The output impedance is matched by the same means as with the grounded-cathode arrangement. The grounded-grid configuration requires more driving (input) power than the grounded-cathode scheme. A grounded-cathode amplifier might produce 1 kW of

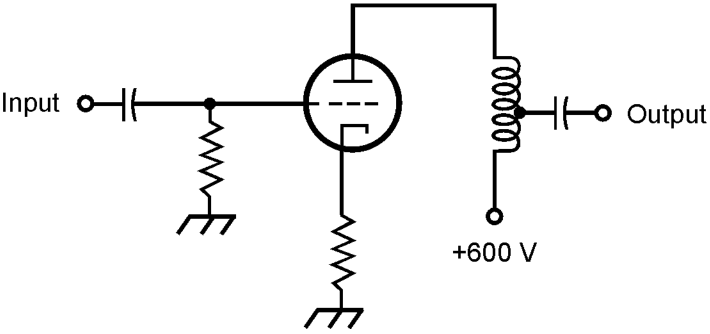

29-4 Simplified schematic diagram of a grounded-cathode RF amplifier circuit using a triode tube.

RF output for 10-W input, but a grounded-grid amplifier needs 50 W to 100 W of drive to produce 1 kW of RF output. A grounded-grid amplifier has a significant advantage, however: it is less likely to break into unwanted oscillation than a grounded-cathode circuit.

Plate Voltage

The plate voltages (+600 V dc) in the circuits of Figs. 29-4 and 29-5 are given as examples. The amplifiers shown could produce 75- to 150-W signal output provided they receive sufficient drive and are properly biased. An amplifier rated at 1-kW output would require a plate voltage of +2 kV dc to +5 kV dc. In high-power radio and TV broadcast transmitters producing in excess of 50-kW RF output, even higher dc plate voltages are used.

Cathode-Ray Tubes

Many TV receivers, and some desktop computer monitors, use *cathode-ray tubes* (CRTs). So do older oscilloscopes, spectrum analyzers, and radar sets.

Electron Beam

In a CRT, a specialized cathode called an *electron gun* emits an electron beam that is focused and accelerated as it passes through positively charged *anodes.* The beam then strikes a glass screen whose inner surface is coated with *phosphor.* The phosphor glows visibly, as seen from the face of the CRT, because of the effect of the high-speed electrons striking it.

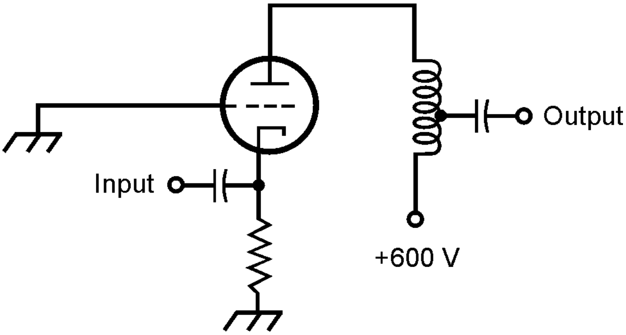

29-5 Simplified schematic diagram of a grounded-grid RF amplifier circuit using a triode tube.

The beam *scanning* pattern is controlled by magnetic or electrostatic fields. One field causes the beam to scan rapidly across the screen in a horizontal direction. Another field moves the beam vertically. When complex waveforms are applied to the electrodes that produce the deflection of the electron beam, a display pattern results. This pattern can be the graph of a signal wave, a fixed image, an animated image, a computer text display, or any other type of visible image.

Electromagnetic CRT

A simplified cross-sectional drawing of an *electromagnetic CRT* is shown in Fig. 29-6. There are two sets of *deflecting coils*, one for the horizontal plane and the other for the vertical plane. (To keep the illustration reasonably clear, only one set of deflecting coils is shown.) The greater the current in the coils, the greater the intensity of the magnetic field, and the more the electron beam is deflected. The electron beam is bent at right angles to the magnetic lines of flux.

In an *oscilloscope*, the horizontal deflecting coils receive a sawtooth waveform. This causes the beam to scan, or *sweep*, at a precise, adjustable speed across the screen from left to right as viewed from in front. After each timed left-to-right sweep, the beam returns, almost instantly, to the left side of the screen for the next sweep. The vertical deflecting coils receive the waveform to be analyzed. This waveform makes the electron beam move up and down. The combination of vertical and horizontal beam motion produces a display of the input waveform as a function of time.

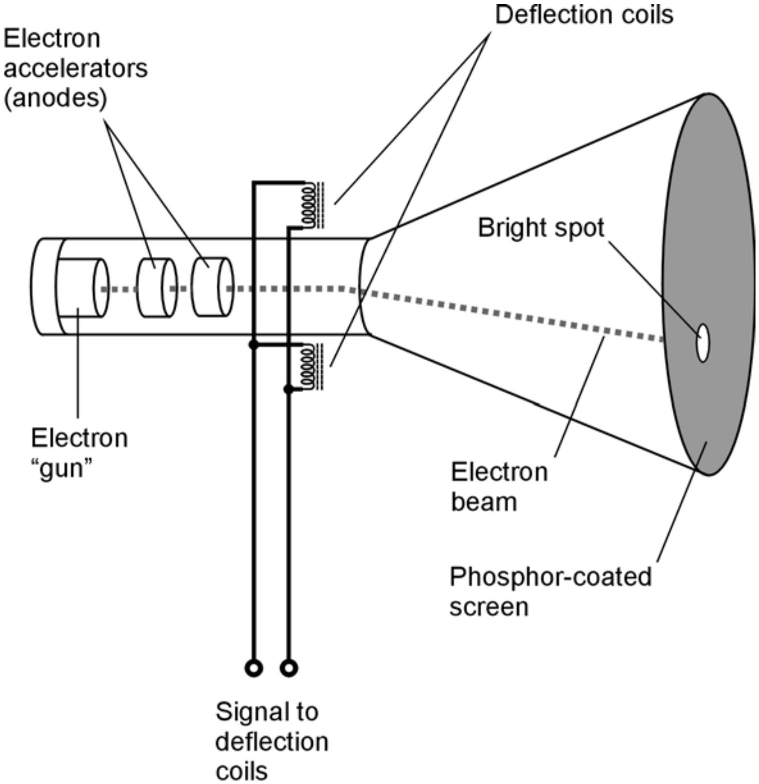

29-6 Simplified cross-sectional rendition of an electromagnetic CRT.

Electrostatic CRT

In an *electrostatic CRT*, charged metal plates, rather than current-carrying coils, are used to deflect the electron beam. When voltages appear on these *deflecting plates*, the beam is bent in the direction of the electric lines of flux. The greater the voltage applied to a deflecting plate, the stronger the electric field, and the greater the extent to which the beam is deflected.

The principal advantage of an electrostatic CRT is the fact that it generates a far less intense magnetic field than an electromagnetic CRT. This so-called extremely low frequency (ELF) energy is a cause for concern, because it might have adverse effects on people who use CRT-equipped devices, such as desktop computers, for extended periods of time. In recent years, with the evolution of *liquid crystal displays* (LCDs) and *plasma displays* as alternatives to the CRT type of display, ELF has become a much less significant concern.

Camera Tubes

Some video cameras use a form of electron tube that converts visible light into varying electric currents. The two most common types of *camera tube* are the *vidicon* and the *image orthicon.*

Vidicon

In the vidicon, a lens focuses the incoming image onto a photoconductive screen. An electron gun generates a beam that sweeps across the screen as a result of the effects of deflecting coils, in a manner similar to the operation of an electromagnetic CRT. The sweep in the vidicon is synchronized with any CRT that displays the image.

As the electron beam scans the photoconductive surface, the screen becomes charged. The rate of discharge in a certain region on the screen depends on the intensity of the visible light striking that region. A simplified cutaway view of a vidicon tube is shown in Fig. 29-7.

The main advantage of the vidicon is its small physical size and mass. A vidicon is sensitive, but its response can be sluggish when the level of illumination is low. This causes images to persist for a short while, resulting in poor portrayal of fast-motion scenes.

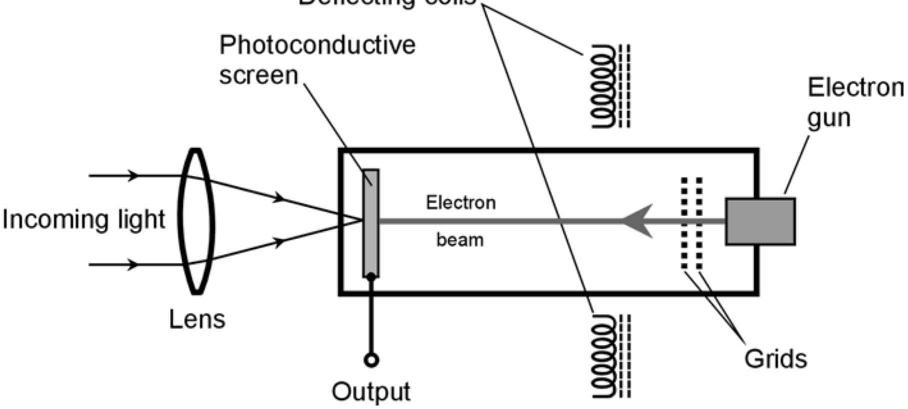

29-7 Functional diagram of a vidicon.

Image Orthicon

Another type of camera tube, also quite sensitive but having a quicker response to image changes, is the image orthicon. It is constructed much like the vidicon, except that there is a target electrode behind the *photocathode* (Fig. 29-8). When a single electron from the photocathode strikes the target electrode, multiple secondary electrons are emitted as a result. The image orthicon thus acts as a video signal amplifier, as well as a camera.

A fine beam of electrons, emitted from the electron gun, scans the target electrode. The secondary electrons cause some of this beam to be reflected back toward the electron gun. Areas of the target electrode with the most secondary electron emission produce the greatest return beam intensity, and regions with the least emission produce the lowest return beam intensity. The greatest return beam intensity corresponds to the brightest parts of the video image. The return beam is modulated as it scans the target electrode and is picked up by a receptor electrode.

One significant disadvantage of the image orthicon is that it produces considerable noise in addition to the signal output. But when a fast response is needed and the illumination ranges from dim to very bright, the image orthicon is the camera tube of choice.

Photomultiplier

A *photomultiplier* is a vacuum tube that generates a variable current depending on the intensity of the light that strikes it. It multiplies its own output, thereby obtaining high sensitivity. Photomultipliers can be used to measure light intensity at low levels.

The photomultiplier consists of a photocathode, which emits electrons in proportion to the intensity of the light striking it. These electrons are focused into a beam, and this beam strikes an electrode called a *dynode.* The dynode emits several secondary electrons for each electron that strikes it. The resulting beam is collected by the anode.

A photomultiplier can have several dynodes, resulting in high gain. The extent to which the sensitivity can be increased by cascading dynodes is limited by the amount of background electron emission or *dark noise* from the photocathode.

Dissector

A *dissector*, also known as an *image dissector,* is a form of photomultiplier in which the light is focused by a lens onto a translucent photocathode. This surface emits electrons in proportion to the

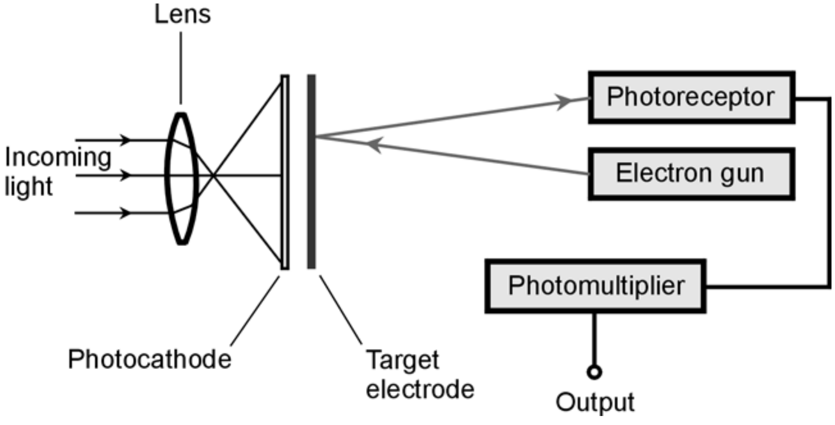

29-8 Functional diagram of an image orthicon.

light intensity. The electrons from the photocathode are directed to a barrier containing a small aperture. The vertical and horizontal deflection plates, supplied with synchronized scanning voltages, move the beam from the photocathode across the aperture. The electron stream passing through the aperture is modulated depending on the light and dark nature of the image.

The *image resolution* of the dissector tube depends on the size of the aperture. The smaller the aperture, the sharper the image, down to a certain limiting point. However, there is a limit to how small the aperture can be, while still allowing enough electrons to pass, and avoiding the generation of *interference patterns.* The image dissector tube produces very little dark noise, and this allows for excellent sensitivity.

Tubes for Use above 300 MHz

Specialized vacuum tubes are required for RF operation at frequencies above 300 MHz. These bands are known as *ultrahigh frequency* (UHF) *band*, which ranges from 300 MHz to 3 GHz, and the *microwave band*, which ranges from 3 GHz up. The *magnetron* and the *Klystron* are examples of tubes that are used to generate and amplify signals at these frequencies.

Magnetron

A magnetron contains a cathode and a surrounding anode. The anode is divided into sections, or *cavities*, by radial barriers. The output is taken from an opening in the anode, and passes into a waveguide that serves as a transmission line for the RF output energy.

The cathode is connected to the negative terminal of a high-voltage source, and the anode is connected to the positive terminal. Therefore, electrons flow radially outward. A magnetic field is applied lengthwise through the cavities. As a result, the electron paths are bent into spirals. The electric field produced by the high voltage, interacting with the longitudinal magnetic field and the effects of the cavities, causes the electrons to bunch up into *clouds.* The swirling movement of the electron clouds causes a fluctuating current in the anode. The frequency depends on the shapes and sizes of the cavities. Small cavities result in the highest oscillation frequencies; larger cavities produce oscillation at relatively lower frequencies.

A magnetron can generate more than 1 kW of RF power at a frequency of 1 GHz. As the frequency increases, the realizable power output decreases. At 10 GHz, a typical magnetron generates about 20 W of RF power output.

Klystron

A Klystron has an electron gun, one or more cavities, and a device that modulates the electron beam. There are several different types. The most common are the *multicavity Klystron* and the *reflex Klystron.*

In a multicavity Klystron, the electron beam is *velocity-modulated* in the first cavity. This causes the density of electrons (the number of particles per unit volume) in the beam to change as the beam moves through subsequent cavities. The electrons tend to bunch up in some regions and spread out in other regions. The intermediate cavities increase the magnitude of the electron beam modulation, resulting in amplification. Output is taken from the last cavity. Peak power levels in some multicavity Klystrons can exceed 1 MW (10^6 W), although the average power is much less.

A reflex Klystron has a single cavity. A *retarding field* causes the electron beam to periodically reverse direction. This produces a phase reversal that allows large amounts of energy to be drawn from the electrons. A typical reflex Klystron can produce signals on the order of a few watts at frequencies of 300 MHz and above.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. A major difference between a triode tube and an N-channel FET is the fact that
   (a) triodes work with lower signal voltages.
   (b) triodes are more compact.
   (c) triodes need higher power-supply voltages.
   (d) triodes don't need filaments.

2. The control grid of a vacuum tube is the electrical counterpart of the
   (a) source of a MOSFET.
   (b) collector of a bipolar transistor.
   (c) anode of a diode.
   (d) gate of an FET.

3. In a tetrode tube, the charge carriers are
   (a) free electrons, which pass among the electrodes.
   (b) holes, which are conducted within the electrodes.
   (c) holes or electrons, depending on whether the device is P type or N type.
   (d) nuclei of whatever elemental gas happens to exist in the tube.

4. Which factor is most significant in limiting the maximum frequency at which a tube can operate?
   (a) The power-supply voltage
   (b) The capacitance among the electrodes
   (c) The physical size of the tube
   (d) The current passing through the filament

5. In a tube with a directly heated cathode,
   (a) the filament is separate from the cathode.
   (b) the grid is connected to the filament.
   (c) the filament serves as the cathode.
   (d) there is no filament.

6. In a tube with a cold cathode,
   (a) the filament is separate from the cathode.
   (b) the grid is connected to the filament.
   (c) the filament serves as the cathode.
   (d) there is no filament.

7. A screen grid enhances tube operation by
   (a) increasing the gain, helping the circuit to oscillate more easily.
   (b) decreasing the plate voltage required to produce oscillation.
   (c) minimizing the risk that a tube amplifier will break into oscillation.
   (d) pulling excess electrons from the plate.

8. A tube with three grids is called a
   (a) triode.
   (b) tetrode.
   (c) pentode.
   (d) hexode.

9. A tube with four grids is called a
   (a) triode.
   (b) tetrode.
   (c) pentode.
   (d) hexode.

10. An advantage of a grounded-grid RF power amplifier over a grounded-cathode RF power amplifier is the fact that the grounded-grid circuit
    (a) has excellent sensitivity.
    (b) exhibits high input impedance.
    (c) produces little or no noise in the input.
    (d) is more stable.

11. A heptode tube has
    (a) one plate.
    (b) two plates.
    (c) three plates.
    (d) four plates.

12. The electron gun in a CRT is another name for its
    (a) cathode.
    (b) anode.
    (c) control grid.
    (d) screen grid.

13. The electron beam in an electrostatic CRT is bent by
    (a) magnetic fields produced by current-carrying coils.
    (b) electric fields produced by charged electrodes.
    (c) a variable voltage on the screen grid.
    (d) visible light striking the electrodes.

14. A grounded-grid RF power amplifier
    (a) requires more driving power, for a given RF power output, than a grounded-cathode RF power amplifier.
    (b) requires less driving power, for a given RF power output, than a grounded-cathode RF power amplifier.
    (c) oscillates at a more stable frequency than a grounded-cathode RF power amplifier.
    (d) oscillates at a less stable frequency than a grounded-cathode RF power amplifier.

15. In a Klystron, the electron-beam density varies as a result of
    (a) amplitude modulation.
    (b) pulse modulation.
    (c) velocity modulation.
    (d) frequency modulation.

16. A vidicon camera tube is noted for its
    (a) poor signal-to-noise ratio.
    (b) large size and heavy weight.
    (c) slow response to image movement in dim light.
    (d) excellent selectivity and electrical ruggedness.

17. The plate in a tetrode tube is normally connected to
    (a) a positive dc power-supply voltage.
    (b) a negative dc power-supply voltage.
    (c) electrical ground.
    (d) RF ground.

18. The screen grid in a tetrode tube is normally connected to
    (a) a positive dc power-supply voltage.
    (b) a negative dc power-supply voltage.
    (c) electrical ground.
    (d) RF ground.

19. Which of the following is most suitable for measuring the intensity of dim light?
    (a) A triode gas-filled tube
    (b) A photomultiplier tube
    (c) An electrostatic CRT
    (d) An electromagnetic CRT

20. In a dissector tube, the aperture size is directly related to the
    (a) operating voltage.
    (b) signal-to-noise ratio.
    (c) response speed.
    (d) image resolution.

30
CHAPTER

Transducers, Sensors, Location, and Navigation

IN THIS CHAPTER, YOU'LL LEARN ABOUT ELECTRONIC DEVICES THAT CONVERT ENERGY FROM ONE form to another, devices that can detect phenomena and measure their intensity, systems that can help you find out where you are (or where some other object is), and devices that facilitate navigation for vessels such as ships, aircraft, and robots.

Wave Transducers

In electronics, *wave transducers* convert ac or dc into acoustic or electromagnetic (EM) waves. They can also convert these waves into ac or dc signals.

Dynamic Transducer for Sound

A *dynamic transducer* is a coil-and-magnet device that translates mechanical vibration into varying electrical current, and can also do the reverse. The most common examples are the *dynamic microphone* and the *dynamic speaker.*

Figure 30-1 shows a functional diagram of a dynamic transducer. A *diaphragm* is attached to a coil that is mounted so it can move back and forth rapidly along its axis. A permanent magnet is placed inside the coil. Sound waves cause the diaphragm to move; this moves the coil, which causes fluctuations in the magnetic field within the coil. The result is ac output from the coil, having the same waveform as the sound waves that strike the diaphragm.

If an audio signal is applied to the coil, it generates a magnetic field that produces forces on the coil. This causes the coil to move, pushing the diaphragm back and forth, creating acoustic waves in the surrounding medium.

Electrostatic Transducer for Sound

An *electrostatic transducer* takes advantage of the forces produced by electric fields. Two metal plates, one flexible and the other rigid, are placed parallel to each other and close together (Fig. 30-2).

In an *electrostatic pickup,* incoming sound waves vibrate the flexible plate. This produces small, rapid changes in the spacing, and therefore the capacitance, between the two plates. A dc voltage is

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

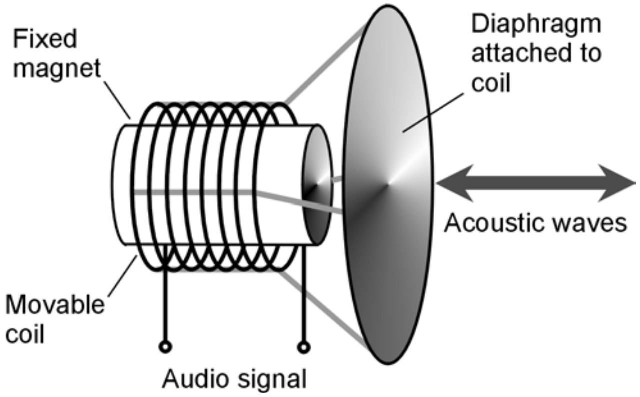

30-1 Functional diagram of a dynamic sound transducer.

applied between the plates. As the interplate capacitance varies, the electric field intensity between them fluctuates. This produces variations in the current through the primary winding of the transformer. Audio signals appear across the secondary.

In an *electrostatic emitter*, fluctuating currents in the transformer produce changes in the voltage between the plates. This results in electrostatic field variations, pulling and pushing the flexible plate in and out. The motion of the flexible plate produces sound waves.

Electrostatic transducers can be used in most applications where dynamic transducers are employed. Advantages of electrostatic transducers include light weight and good sensitivity. The relative absence of magnetic fields can also be an asset in certain situations.

Piezoelectric Transducer for Sound and Ultrasound

Figure 30-3 shows a *piezoelectric transducer.* This device consists of a *crystal* of quartz or ceramic material, sandwiched between two metal plates. When sound waves strike one or both of the plates, the metal vibrates. This vibration is transferred to the crystal. The crystal generates weak electric currents when subjected to this mechanical stress. Therefore, an ac voltage develops between the two metal plates, with a waveform similar to that of the sound.

If an ac signal is applied to the plates, it causes the crystal to vibrate in sync with the current. The metal plates vibrate also, producing an acoustic disturbance.

Piezoelectric transducers can function at higher frequencies than can dynamic or electrostatic transducers. For this reason, they are favored in ultrasonic applications, such as intrusion detectors and alarms.

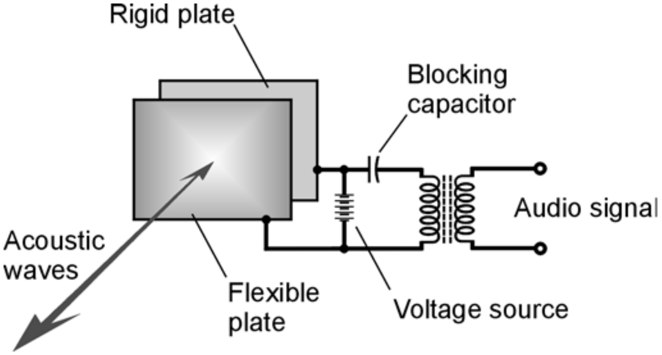

30-2 Functional diagram of an electrostatic sound transducer.

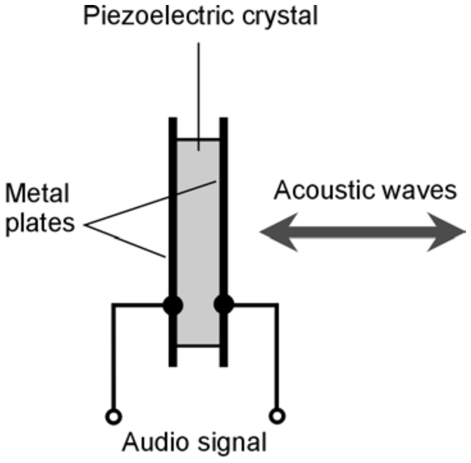

30-3 Functional diagram of a piezoelectric transducer for sound and ultrasound.

Transducers for RF Energy

The term *radio-frequency* (RF) *transducer* is a fancy name for an *antenna.* There are two basic types: the *receiving antenna* and the *transmitting antenna.* You learned about antennas in Chap. 27.

Transducers for IR and Visible Light

Many wireless devices transmit and receive energy at IR wavelengths. Infrared energy has a frequency higher than that of radio waves, but lower than that of visible light. Some wireless devices transmit and receive their signals in the visible range, although these are encountered much less often than IR devices.

The most common IR transmitting transducer is the infrared-emitting diode (IRED). Fluctuating dc is applied to the device, causing it to emit IR rays. The fluctuations in the current constitute modulation, and this produces rapid variations in the intensity of the rays emitted by the semiconductor P-N junction. The modulation contains information, such as which channel your television set should seek, or whether the volume is to be raised or lowered. Infrared energy can be focused by optical lenses and reflected by optical mirrors. This makes it possible to *collimate* IR rays (make them parallel) so they can be transmitted for distances up to several hundred meters.

Infrared receiving transducers resemble photodiodes or photovoltaic cells. The fluctuating IR energy from the transmitter strikes the P-N junction of the receiving diode. If the receiving device is a photodiode, a current is applied to it, and this current varies rapidly in accordance with the signal waveform on the IR beam from the transmitter. If the receiving device is a photovoltaic cell, it produces the fluctuating current all by itself, without the need for an external power supply. In either case, the current fluctuations are weak, and must be amplified before they are delivered to whatever equipment (television set, garage door, oven, security system, etc.) is controlled by the wireless system. Infrared wireless devices work best on a line of sight.

Displacement Transducers

A *displacement transducer* measures a distance or angle traversed, or the distance or angle separating two points. Conversely, it can convert a signal into movement over a certain distance or angle. A device that measures or produces movement in a straight line is a *linear displacement transducer.* If it measures or produces movement through an angle, it is an *angular displacement transducer.*

Pointing and Control Devices

A *joystick* is a control device capable of producing movement, or controlling variable quantities, in two dimensions. The device consists of a movable lever and a ball bearing within a control box. The lever can be moved by hand up and down, and to the right and left. Joysticks are used in computer games, for entering coordinates into a computer, and for the remote control of robots. In some joysticks, the lever can be rotated, allowing control in a third dimension.

A *mouse* is a peripheral commonly used with personal computers. By sliding the mouse around on a flat surface, a cursor or arrow is positioned on the display. Pushbutton switches on the top of the unit actuate the computer to perform whatever function the cursor or arrow shows. These actions are called *clicks.*

A *trackball* resembles an inverted mouse, or a two-dimensional joystick without the lever. Instead of the device being pushed around on a surface, the user moves a ball bearing, causing the display cursor to move vertically and horizontally. Pushbutton switches on a computer keyboard, or on the trackball box itself, actuate the functions.

An *eraser-head pointer* is a rubber button approximately 5 mm in diameter, usually placed in the center of a computer keyboard. The user moves the cursor on the display by pushing against the button. Clicking and double clicking are done with button switches on the keyboard.

A *touch pad* is a sensitive plate that is about the size of a business card. The user places an index finger on the plate and moves the finger around. This results in intuitive movement of the display cursor. Clicking and double clicking are done in the same way as with the trackball and eraser-head pointer.

Electric Motor

An *electric motor* converts electrical energy into angular (and in some cases linear) mechanical energy. Motors can operate from ac or dc, and range in size from tiny devices used in microscopic robots to huge machines that pull passenger trains.

The basics of dc motors were discussed in Chap. 8. In a motor designed to work with ac, there is no commutator. The alternations in the current keep the polarity correct at all times, so the shaft does not lock up. The rotational speed of an ac motor depends on the frequency of the applied ac. With 60-Hz ac, for example, the rotational speed is 60 revolutions per second (60 rps) or 3600 revolutions per minute (3600 rpm).

When a motor is connected to a load, the rotational force required to turn the shaft increases. The greater the required force becomes, the more power is drawn from the source.

Stepper Motor

A *stepper motor* turns in small increments, rather than continuously. The *step angle*, or extent of each turn, varies depending on the particular motor. It can range from less than 1° of arc to a quarter of a circle (90°). A stepper motor turns through its step angle and then stops, even if the current is maintained. When a stepper motor is stopped with a current going through its coils, the shaft resists external rotational force.

Conventional motors run at hundreds, or even thousands, of revolutions per minute (rpm). But a stepper motor usually runs at much lower speeds, almost always less than 180 rpm. A stepper motor has the most turning power when it is running at its slowest speeds, and the least turning power when it runs at its highest speeds.

When a pulsed current is supplied to a stepper motor, the shaft rotates in increments, one step for each pulse. In this way, a precise speed can be maintained. Because of the braking effect, this speed is constant for a wide range of mechanical turning resistances. Stepper motors can be controlled using microcomputers. This type of motor is especially well suited for point-to-point motion. Complicated, intricate tasks can done by computer-controlled robots using stepper motors.

Selsyn

A *selsyn* is an indicating device that shows the direction in which an object is oriented. It consists of a transmitting unit and a receiving (or indicator) unit. As the shaft of the transmitting unit rotates, the shaft of the receiving unit, which is a stepper motor, follows along exactly. A common application is as a direction indicator for a wind vane (Fig. 30-4). When the wind vane rotates, the indicator unit shaft moves through the same number of angular degrees as the transmitting unit shaft. A selsyn for azimuth (compass) bearings has a range of 0° to 360°. A selsyn for elevation bearings has range of 0° to 90°.

A *synchro* is a selsyn used for the control of mechanical devices. Synchros are well suited for robotic *teleoperation*, or remote control. Some synchros are programmable. The operator inputs a number into the generator, and the receiver changes position accordingly. Synchros are commonly used as rotators and direction indicators for directional communications antennas such as the Yagi or dish.

Electric Generator

An *electric generator* is constructed in much the same way as an ac motor, although it functions in the opposite sense. Some generators can also be used as motors; devices of this type are called *motor/generators.*

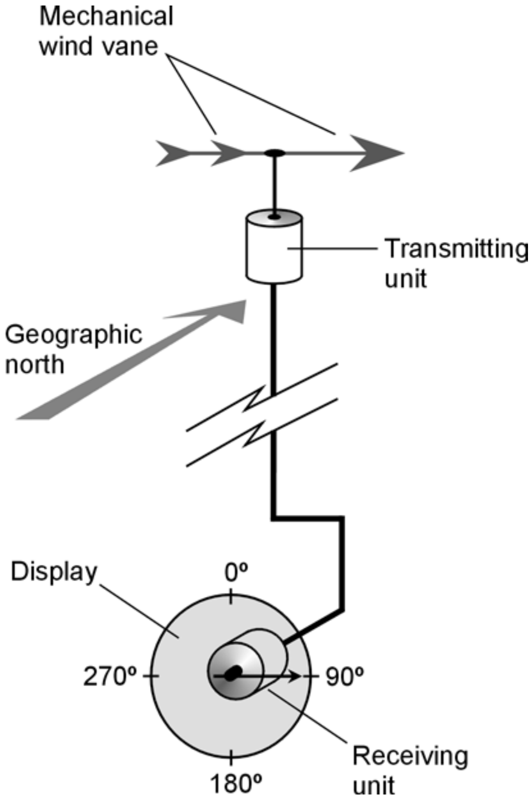

30-4 A selsyn can be used to remotely indicate the direction in which a mechanical device is pointed.

A typical generator produces ac from the mechanical rotation of a coil in a strong magnetic field. Alternatively, a permanent magnet can be rotated within a coil of wire. The rotating shaft can be driven by a gasoline engine, a steam turbine, a water turbine, a windmill, or even by human power. A commutator can be used with a generator to produce pulsating dc output, which can be filtered to obtain pure dc for use with electronic equipment.

Small portable gasoline-powered generators, capable of delivering a few kilowatts, can be purchased in department stores or home-and-garden stores. Larger generators, which usually burn propane or methane ("natural gas"), allow homes or buildings to keep their electrical power in the event of an interruption in the utility. The largest generators are found in power plants, and can produce many kilowatts.

Small generators can be used in synchro systems. These specialized generators allow remote control of robotic devices. A generator can be used to measure the speed at which a vehicle or rolling robot moves. The shaft of the generator is connected to one of the wheels, and the generator output voltage and frequency vary directly with the angular speed of the wheel. This is an example of a *tachometer*, a device familiar to people with automotive experience.

Optical Encoder

In digital radios, frequency adjustment is done in discrete steps. A typical increment is 10 Hz for shortwave radios and 200 kHz for FM broadcast radios. An alternative to mechanical switches or

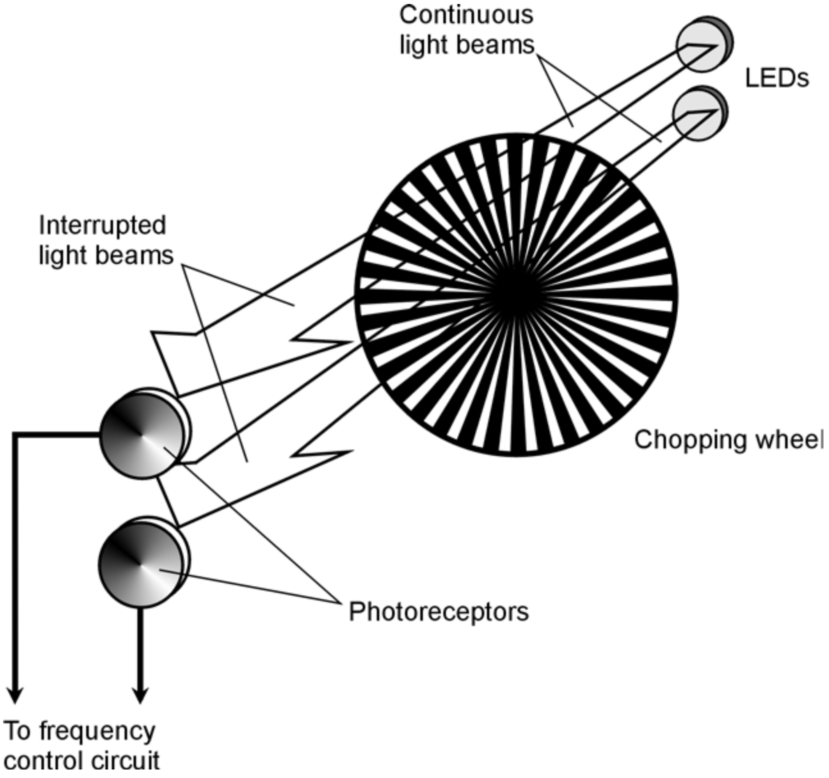

30-5 An optical encoder uses LEDs and photodetectors to sense the direction and extent of shaft rotation.

gear-driven devices, which wear out with time, is the *optical encoder*, also called the *optical shaft encoder.*

An optical encoder consists of two LEDs, two photodetectors, and a device called a *chopping wheel.* The LEDs shine on the photodetectors through the wheel. The wheel has radial bands, alternately transparent and opaque (Fig. 30-5). The wheel is attached to the tuning shaft, which is attached to a large knob. As the tuning knob is rotated, the light beams are interrupted. Each interruption causes the frequency to change by a specified increment. The difference between "frequency up" and "frequency down" (clockwise and counterclockwise shaft rotation, respectively) is determined according to which photodetector senses each sequential beam interruption first.

Detection and Measurement

A *sensor* employs one or more transducers to detect and/or measure phenomena such as temperature, humidity, barometric pressure, texture, proximity, and the presence of certain substances.

Capacitive Pressure Sensor

A *capacitive pressure sensor* is shown in Fig. 30-6. Two metal plates are separated by a layer of dielectric (electrically insulating) foam, forming a capacitor. This component is connected in parallel with an inductor. The resulting inductance/capacitance (*LC*) circuit determines the frequency of an oscillator. If an object strikes the sensor, the plate spacing momentarily decreases. This increases the capacitance, causing a drop in the oscillator frequency. When the object moves away from the transducer, the foam springs back, the plates return to their original spacing, and the oscillator frequency returns to normal.

The output of a capacitive pressure sensor can be converted to digital data using an *analog-to-digital converter* (ADC). This signal can be sent to a microcomputer such as a robot controller. Pressure sensors can be mounted in various places on a mobile robot, such as the front, back, and sides. Then, for example, physical pressure on the sensor in the front of the robot can send a signal to the controller, which tells the machine to move backward.

A capacitive pressure sensor can be fooled by massive conducting or semiconducting objects in its vicinity. If such a mass comes near the transducer, the capacitance may change even if direct contact is not made. This phenomenon is known as *body capacitance.* When the effect must be avoided, an *elastomer* device can be used for pressure sensing.

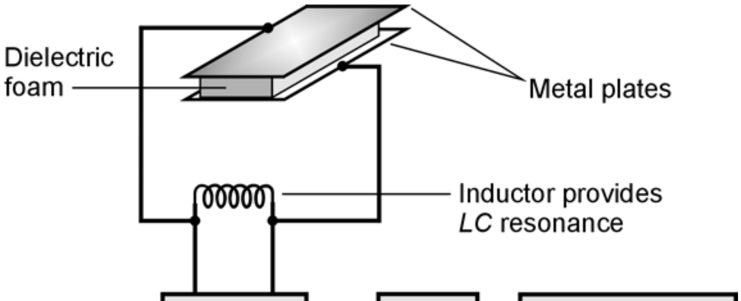

30-6 A capacitive pressure sensor. When force is applied, the spacing between the plates decreases, causing the capacitance to increase and the oscillator frequency to go down.

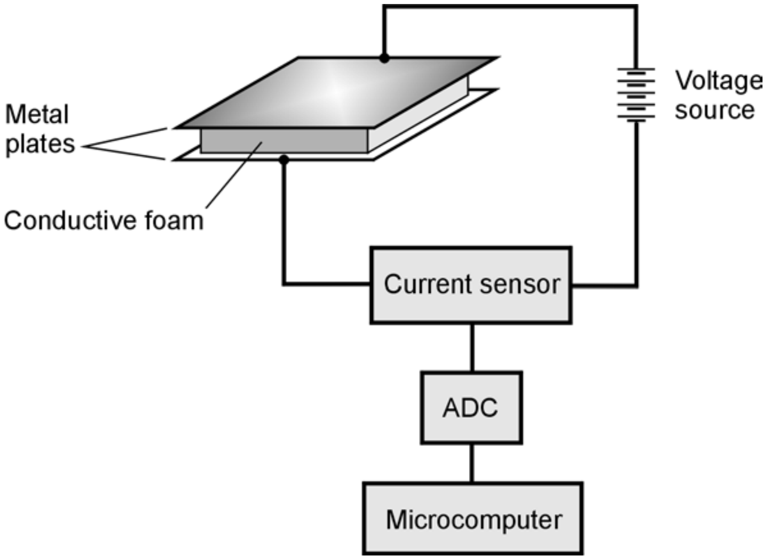

30-7 An elastomer pressure sensor detects applied force without unwanted capacitive effects.

Elastomer

An *elastomer* is a flexible substance resembling rubber or plastic that can be used to detect the presence or absence of mechanical pressure. Figure 30-7 illustrates how an elastomer can be used to detect, and locate, a pressure point. The elastomer conducts electricity fairly well, but not perfectly. It has a foam-like consistency, so that it can be compressed. Conductive plates are attached to the pad.

When pressure appears at some point in the elastomer pad, the material is compressed, and this lowers its electrical resistance. This is detected as an increase in the current between the plates. The greater the pressure becomes, the more the elastomer is compressed, and the greater is the increase in the current. The current-change data can be sent to a microcomputer such as a robot controller.

Back-Pressure Sensor

A motor produces a measurable pressure that depends on the torque being applied. A back-pressure sensor detects and measures the torque that the motor is applying at any given time. The sensor produces a signal, usually a variable voltage, that increases as the torque increases. Figure 30-8 is a functional block diagram of a back-pressure sensor.

Back-pressure sensors are used to limit the force applied by robot grippers, arms, drills, hammers, or other *end effectors.* The *back voltage*, or signal produced by the sensor, reduces the torque applied by the motor. This can prevent damage to objects being handled by the robot. It also helps to ensure the safety of people working around the robot.

Capacitive Proximity Sensor

A *capacitive proximity sensor* uses an RF oscillator, a frequency detector, and a metal plate connected into the oscillator circuit (Fig. 30-9). The oscillator is designed so that a change in the capacitance

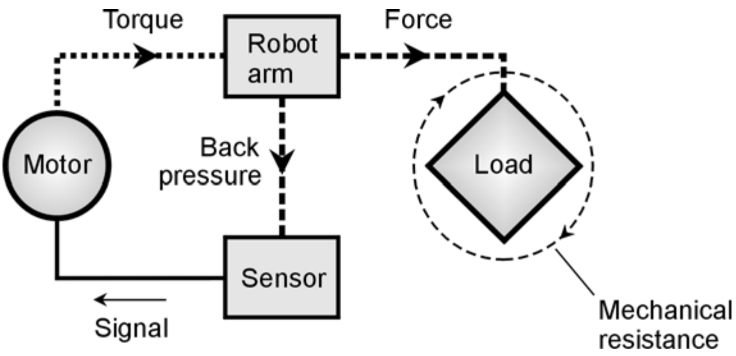
Torque
Robot arm
Force
Motor
Back pressure
Load
Sensor
Signal
Mechanical resistance

30-8 A back-pressure sensor governs the force applied by a robot arm or other mechanical device.

of the plate, with respect to the environment, causes the oscillator frequency to change. This change is sensed by the frequency detector, which sends a signal to a microcomputer or robot controller.

Substances that conduct electricity to some extent, such as metal, saltwater, and living tissue, are sensed more easily by capacitive transducers than are materials that do not conduct, such as dry wood, plastic, glass, or dry fabric.

Photoelectric Proximity Sensor

Reflected light can provide a way for a mobile robot to tell if it is approaching something. A *photoelectric proximity sensor* uses a light-beam generator, a photodetector, a frequency-sensitive amplifier, and a microcomputer (Fig. 30-10).

The light beam reflects from the object and is picked up by the photodetector. The light beam is modulated at a certain frequency, say 1000 Hz. The amplifier responds only to light modulated at that frequency. This prevents false imaging that can otherwise be caused by lamps or sunlight. (Such light sources are unmodulated, and will not actuate a sensor designed to respond only to modulated light.) If the robot is approaching an object, its controller senses that the reflected, modulated beam is getting stronger. The robot can then steer clear of the object.

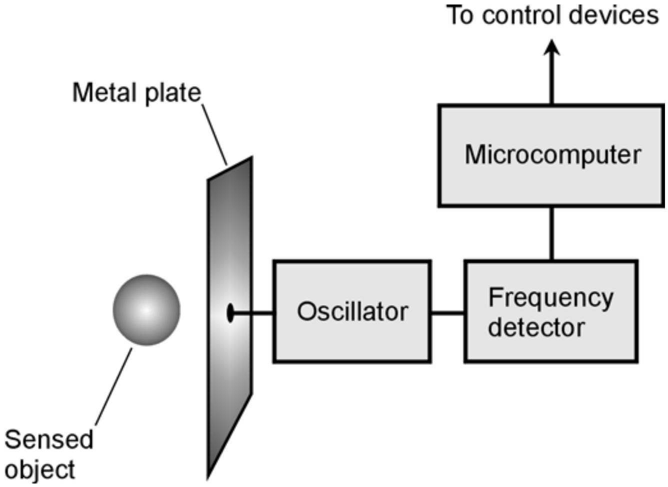
To control devices
Metal plate
Microcomputer
Oscillator
Frequency detector
Sensed object

30-9 A capacitive proximity sensor can detect nearby conducting or semiconducting objects.

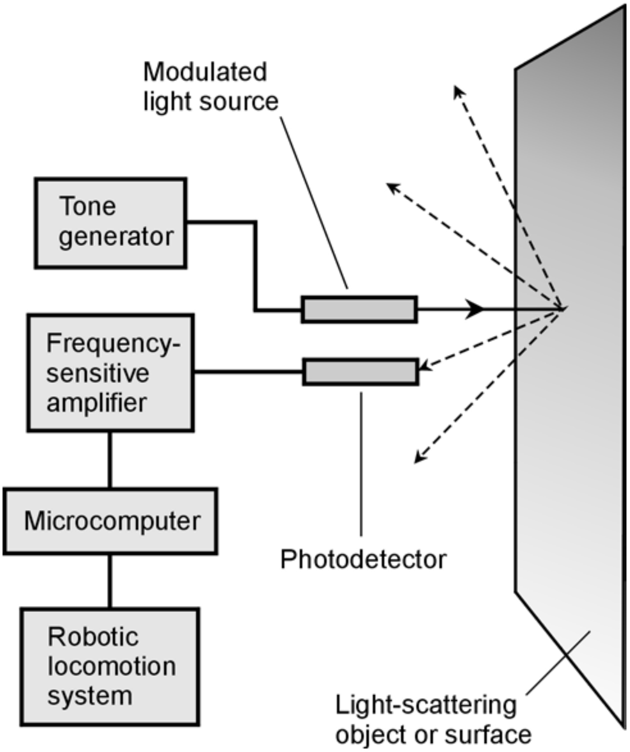
Modulated light source
Tone generator
Frequency-sensitive amplifier
Microcomputer
Robotic locomotion system
Photodetector
Light-scattering object or surface

30-10 A photoelectric proximity sensor. Modulation of the light beam allows the device to distinguish between sensor-generated light and background illumination.

This method of proximity sensing does not work for objects that do not reflect light, or for windows or mirrors approached at a sharp angle. In these scenarios, the light beam is not reflected back toward the photodetector, so the object is invisible.

Texture Sensor

Texture sensing is the ability of a machine to determine whether a surface is shiny or rough (matte). Basic texture sensing involves the use of a laser and several light-sensitive sensors.

Figure 30-11 shows how a laser (L) and sensors (S) can be used to tell the difference between a shiny surface (at A) and a rough or matte surface (at B). The shiny surface, such as the polished hood of a car, reflects light at the incidence angle only. But the matte surface, such as a sheet of paper, scatters light in all directions. The shiny surface reflects the beam back entirely to the sensor in the path of the beam whose reflection angle equals its incidence angle. The matte surface reflects the beam back to all the sensors. A microcomputer can be programmed to tell the difference.

Certain types of surfaces can confuse the texture sensor shown in Fig. 30-11. For example, a pile of glass marbles can be defined as shiny on a small scale but irregular on a large scale. Depending on the diameter of the laser beam, the texture sensor might interpret such a surface as either shiny or matte. The determination can also be affected by motion of the sensor relative to the surface. A surface that is interpreted as shiny when standing still relative to the sensor might be interpreted as matte when moving relative to the sensor.

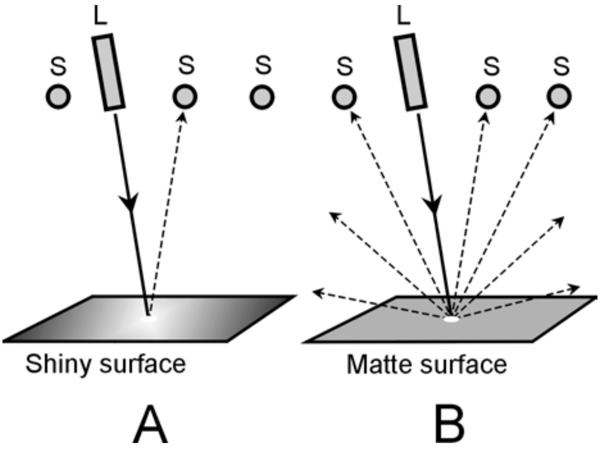

30-11 In texture sensing, lasers (L) and sensors (S) analyze a shiny surface (at A) and a matte surface (at B). Solid lines represent incident light; dashed lines represent reflected light.

Location Systems

Transducers and sensors can operate over long distances. The above-described devices are intended mainly for short-range applications (with the exception of RF antennas). In this section, a few medium-range and long-range applications of transducers and sensors are described. These applications fall into the broad category of *location systems.*

Radar

The term *radar* is derived from the words *radio detection and ranging.* Electromagnetic (EM) waves having certain frequencies reflect from various objects, especially if those objects contain metals or other good electrical conductors. By ascertaining the direction(s) from which radio signals are returned, and by measuring the time it takes for a pulsed beam of EM energy to travel from the transmitter location to a target and back again, it is possible to pinpoint the geographic positions of distant objects. During the Second World War in the 1940s, this property of radio waves was put to use for the purpose of locating aircraft.

In the years following the war, it was discovered that radar can be useful in a variety of applications, such as measurement of automobile speed (by the police), weather forecasting (rain and snow reflect radar signals), and even the mapping of the moon and the planet Venus. Radar is extensively used in aviation, both commercial and military. In recent years, radar has also found uses in robot guidance systems.

A complete radar set consists of a transmitter, a directional antenna with a narrow main lobe and high gain, a receiver, and an indicator or display. The transmitter produces intense pulses of radio microwaves at short intervals. The pulses are propagated outward in a narrow beam from the antenna, and they strike objects at various distances. The reflected signals, or *echoes*, are picked up by the antenna shortly after the pulse is transmitted. The farther away the reflecting object, or target, the longer the time before the echo is received. The transmitting antenna is rotated so that all azimuth bearings (compass directions) can be observed.

A typical circular radar display consists of a CRT or LCD. Figure 30-12 shows the basic display configuration. The observing station is at the center of the display. Azimuth bearings are indicated

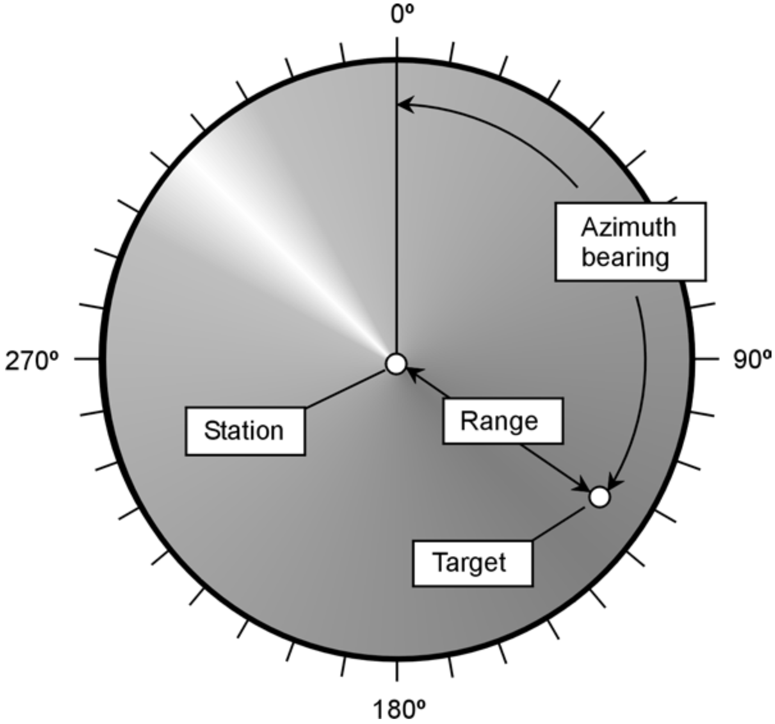

30-12 A radar display. The light radial band shows the azimuth direction in which the microwave beam is currently transmitted and received. (Not all radar displays show this band.)

in degrees clockwise from true north, and are marked around the perimeter of the screen. The distance, or *range*, is indicated by the radial displacement of the echo; the farther away the target, the farther from the display center the echo or blip. The radar display is, therefore, a set of polar coordinates. In the drawing, a target is shown at an azimuth of about 125° (east-southeast). Its range is near the maximum for the display.

The maximum range of a radar system depends on the height of the antenna above the ground, the nature of the terrain in the area, the transmitter output power and antenna gain, the receiver sensitivity, and the weather conditions in the vicinity. Airborne long-range radar can detect echoes from several hundred kilometers (km) away under ideal conditions. A low-power radar system, with the antenna at a low height above the ground, might receive echoes from only 50 to 70 km.

The fact that precipitation reflects radar echoes is a nuisance to aviation personnel, but it is invaluable for weather forecasting and observation. Radar has made it possible to detect and track severe thunderstorms and hurricanes. A *mesocyclone*, which is a severe thunderstorm likely to produce tornadoes, causes a hook-shaped echo on radar. The eye of a hurricane, and the eyewall and rainbands surrounding it, all show up clearly on a radar display.

Some radar sets can detect changes in the frequency of the returned pulse, thereby allowing measurement of wind speeds in hurricanes and tornadoes. This is called *Doppler radar.* This type of radar is also employed to measure the speeds of approaching or receding targets.

Sonar

Sonar is a medium-range method of proximity sensing. The acronym derives from the words *sonic navigation and ranging*. The basic principle is simple: Bounce acoustic waves off of objects, and measure the time it takes for the echoes to return.

An elementary sonar system consists of an ac pulse generator, an acoustic emitter, an acoustic pickup, a receiver, a delay timer, and an indicating device such as a numeric display, CRT, LCD, or pen recorder. The transmitter sends out acoustic waves through the medium, usually water or air. These waves are reflected by objects, and the echoes are picked up by the receiver. The distance to an object is determined on the basis of the echo delay, provided the speed of the acoustic waves in the medium is known.

A simple sonar system is diagrammed in Fig. 30-13A. A *computer map* can be generated on the basis of sounds returned from various directions in two or three dimensions. This can help a mobile robot or vessel navigate in its environment. However, the system can be fooled if the echo delay is equal to or longer than the time interval between pulses, as shown at B. To overcome this, a computer can instruct the generator to send pulses of various frequencies in a defined sequence. The computer keeps track of which echo corresponds to which pulse.

Acoustic waves travel faster in water than in air. The amount of salt in water makes a difference in the propagation speed when sonar is used on boats, for example, in depth finding. The density of water can vary because of temperature differences as well. If the true speed of the acoustic waves is not accurately known, false readings will result. In freshwater, the speed of sound is about 1400 meters per second (m/s), or 4600 feet per second (ft/s). In saltwater, it is about 1500 m/s (4900 ft/s). In air, sound travels at approximately 335 m/s (1100 ft/s).

In the atmosphere, sonar can make use of audible sound waves, but ultrasound is often used instead. Ultrasound has a frequency too high to hear, ranging from about 20 kHz to more than 100 kHz. One advantage of ultrasound is that the signals will not be heard by people working around machines equipped with sonar. Another advantage is the fact that it is less likely to be fooled by people talking, machinery operating, and other noises. At frequencies higher than the range of human hearing, acoustical disturbances do not normally occur as often, or with as much intensity, as they do within the hearing range.

In its most advanced forms, sonar can rival *vision systems* (also called *machine vision*) as a means of mapping the environment for a mobile robot or vessel. Sonar has one significant limitation: all acoustic waves, including sound and ultrasound, require a gaseous or liquid medium in order to propagate. Therefore, sonar is useless in outer space, which is practically a vacuum.

Signal Comparison

A machine or vessel can find its geographical position by comparing the signals from two fixed stations whose positions are known, as shown in Fig. 30-14A. By adding 180° to the bearings of the sources X and Y, the machine or vessel (small square block) obtains its bearings as seen from the sources (round dots). The machine or vessel can determine its direction and speed by taking two readings separated by a certain amount of time.

In the old days, diagrams such as the one in Fig. 30-14A were actually plotted on maps by the captains of oceangoing vessels and aircraft. Nowadays, computers do that work, with more accurate results.

Figure 30-14B is a block diagram of an *acoustic direction finder* such as can be used by a mobile robot. The receiver has a signal-strength indicator and a servo that turns a directional ultrasonic

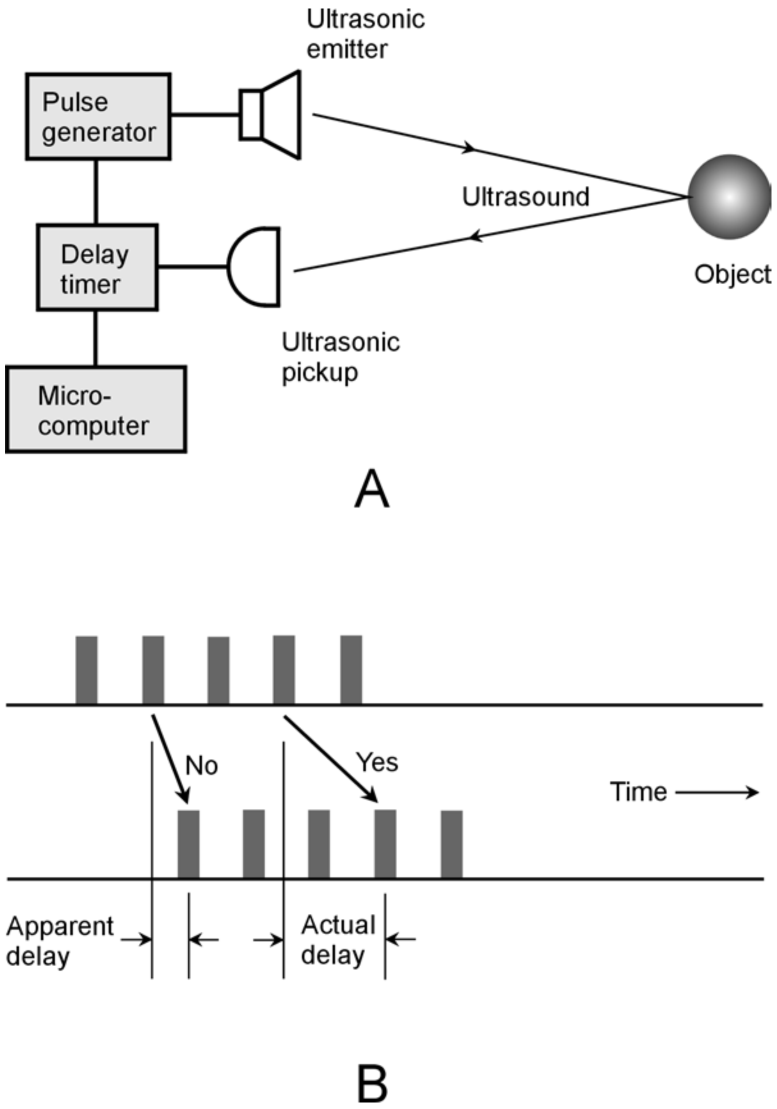
Ultrasonic emitter
Pulse generator
Ultrasound
Delay timer
Object
Ultrasonic pickup
Micro-computer
A
No
Yes
Time ⟶
Apparent delay
Actual delay
B

30-13 At A, a block diagram of medium-range sonar system. At B, a sonar system can be confused by long delays.

transducer. There are two signal sources at different frequencies. When the transducer is rotated so the signal from one source is maximum, a bearing is obtained by comparing the orientation of the transducer with some known standard such as a magnetic compass. The same is done for the other source. A computer determines the precise location of the robot, based on this data.

Radio Direction Finding (RDF)

A radio receiver, equipped with a signal-strength indicator and connected to a rotatable, directional antenna, can be used to determine the direction from which signals are coming. *Radio direction find-*

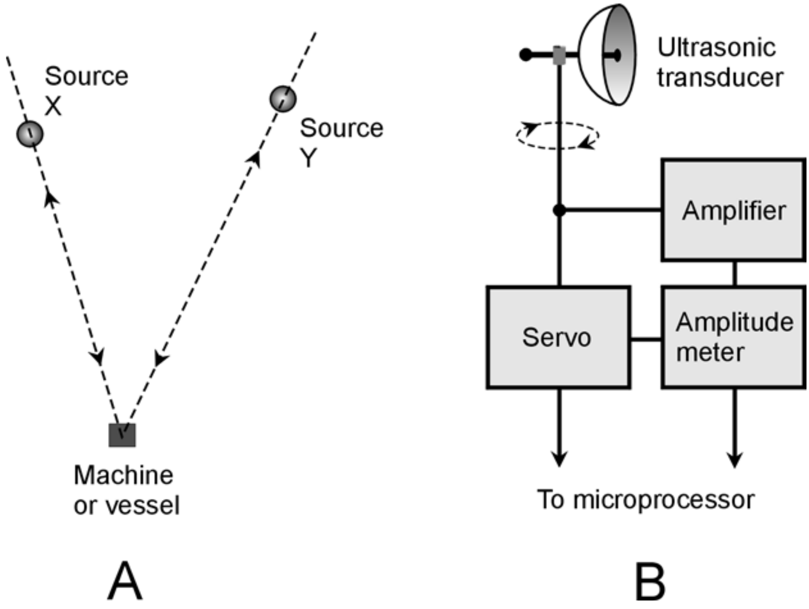

30-14 A simple direction-finding scheme (A) and an ultrasonic direction finder (B).

ing (RDF) equipment aboard a mobile vehicle facilitates determining the location of a transmitter. An RDF receiver can also be used to find one's own position with respect to two or more transmitters operating on different frequencies.

In an RDF receiver for use at frequencies below about 300 MHz, a small loop or loopstick antenna is used. It is shielded against the electric component of radio waves, so it picks up only the magnetic part of the EM field. The loop is rotated until a sharp dip, or null, occurs in the received signal strength, indicating that the axis of the loop lies along a line toward the transmitter. When readings are taken from two or more locations separated by a sufficient distance, the transmitter can be pinpointed by finding the intersection point of the azimuth bearing lines on a map.

At frequencies above approximately 300 MHz, a directional transmitting and receiving antenna, such as a Yagi, quad, dish, or helical type, gives better results than a small loop. When such an antenna is employed for RDF, the azimuth bearing is indicated by a signal peak rather than by a null.

Navigational Methods

Navigation involves the use of location devices over a period of time, thereby deriving a function of position versus time. This technique can be used to determine whether or not a vessel is on course. It can also be used to track the paths of military targets, severe thunderstorms, and hurricanes.

Fluxgate Magnetometer

When conventional position sensors do not function in a particular environment for a mobile robot, a *fluxgate magnetometer* can be used. This system employs sensitive magnetic receptors and a microcomputer to sense the presence of, and detect changes in, an artificially generated magnetic field. Navigation within a room can be done by checking the orientation of magnetic lines of flux generated by electromagnets in the walls, floor, and ceiling of the room. For each point in the room, the magnetic flux lines have a unique direction and intensity. There is a one-to-one correspondence between the magnetic flux intensity/direction and the points within the robot's operating environment. This correspondence can be represented as a two-variable mathematical function of every location in the room. The robot controller is programmed to know this function. This makes it possible for the machine to pinpoint its position with extreme accuracy, in some cases to within a few millimeters.

Epipolar Navigation

Epipolar navigation works by evaluating the way an image changes as viewed from a moving perspective. Suppose that you are piloting an aircraft over the ocean. The only land in sight is a small island. The on-board computer sees an image of the island that constantly changes shape. Figure 30-15 shows three sample sighting positions (A, B, C) and the size/shape of the island as seen by a machine vision system in each case. The computer has the map data, so it knows the true size, shape, and location of the island. The computer compares the shape and size of the image it sees at each point in time, from the vantage point of the aircraft, with the actual shape and size of the island from the map data. From this, the computer can ascertain the altitude of the aircraft, its speed and direction of movement relative to the surface, its latitude, and its longitude.

Log Polar Navigation

In *log polar navigation*, a computer converts an image in *polar coordinates* to an image in *rectangular coordinates.* The polar radius is mapped onto the vertical rectangular axis, and the polar angle is mapped onto the horizontal rectangular axis.

Radial coordinates are unevenly spaced in the polar map, but are uniform in the rectangular map. During the transformation, the logarithm of the radius is taken to maximize the geographical area that the system can observe. As a result of this logarithmic transformation (that's where the "log" comes from in the term), long-distance resolution is sacrificed, but the close-in resolution is enhanced.

The so-called log polar transform distorts the way a scene appears to human observers, but translates images and motions into data that can be efficiently dealt with by a computerized scanning system.

Loran

The term *loran* is an acronym derived from the words *long range navigation.* Loran is one of the oldest electronic navigation schemes, and is still used by some ships and aircraft. The system employs RF pulse transmission at low frequencies from multiple transmitters at specific geographic locations.

In loran, a computer on board a vessel can determine the location of the ship by comparing the time difference in the arrival of the signals from two different transmitters at known locations. Based on the fact that radio waves propagate at the speed of light in free space (approximately 299,792 km/s or 186,282 mi/s), it is possible to determine the distance to each transmitter, and from this, the location of the ship relative to the transmitters.

In recent years, loran has been largely supplanted by the *Global Positioning System* (GPS).

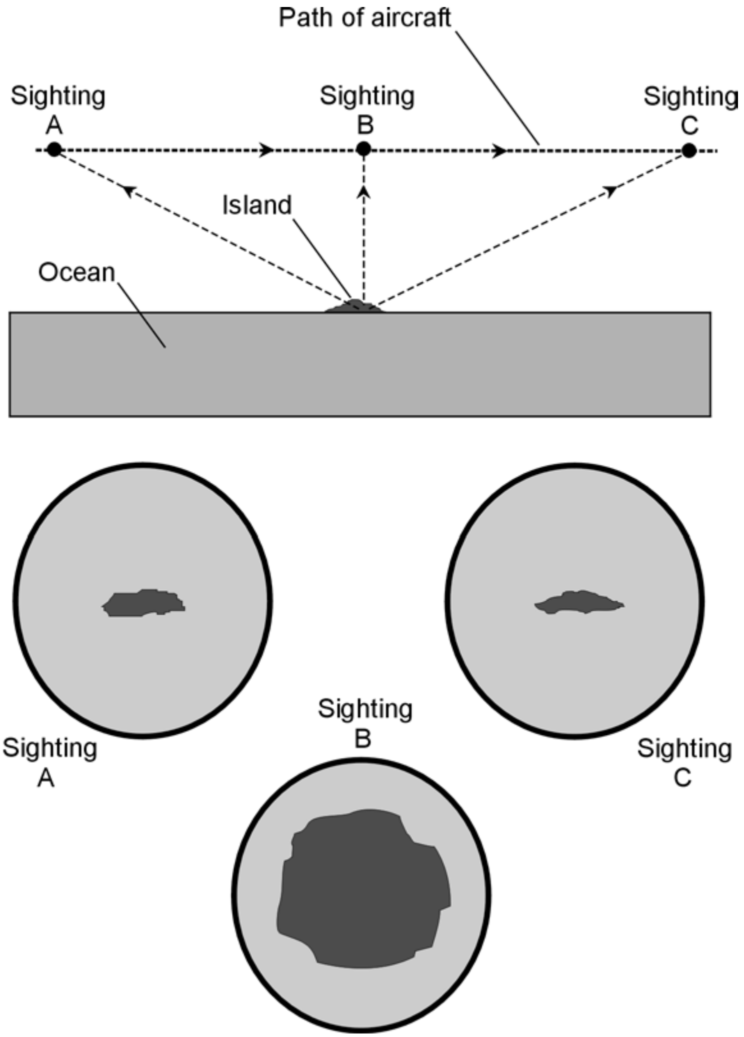

30-15 Epipolar navigation is the optoelectronic counterpart of human spatial perception.

Global Positioning System (GPS)

The Global Positioning System (GPS) is a network of radiolocation/radionavigation apparatus that operates on a worldwide basis. The system employs several satellites, and allows determination of latitude, longitude, and altitude.

All GPS satellites transmit signals in the microwave part of the radio spectrum. The signals are modulated with codes that contain timing information used by the receiving apparatus to make measurements. A GPS receiver determines its location by measuring the distances to several different satellites. This is done by precisely timing the signals as they travel between the satellites and the receiver. The process is similar to triangulation, except that it is done in three dimensions (in space) rather than in two dimensions (on the surface of the earth).

There is some reduction in the propagation speed of EM microwaves in the ionosphere, as compared with the propagation speed in free space. The extent of this reduction depends on the signal frequency. The GPS employs dual-frequency transmission to compensate for this effect. A GPS receiver uses a computer to process the information received from the satellites. From this information, it can give the user an indication of position down to a few meters.

Increasingly, automobiles, trucks, and pleasure boats come with GPS receivers installed as standard equipment. If you are driving in a remote area and you get lost, you can use GPS to locate your position. Using a cell phone, citizens band (CB) radio transceiver, or amateur (ham) radio transceiver, you can call for help and inform authorities of your exact position, which is displayed on a detailed map of the area in which you are located.

Quiz

Refer to the text in this chapter if necessary. A good score is at least 18 correct. Answers are in the back of the book.

1. Which of the following devices would most likely be used for adjusting the frequency setting of a digital radio transmitter or receiver?
   (a) A piezoelectric transducer
   (b) A dynamic transducer
   (c) An optical shaft encoder
   (d) A capacitive proximity sensor

2. Collimation of IR rays can be done by means of
   (a) an ordinary lens.
   (b) a selsyn.
   (c) a log polar transform system.
   (d) any of the above.

3. A computer map
   (a) can help a robot find its way around.
   (b) requires the use of IR beacons.
   (c) is generated using optical encoders.
   (d) requires the use of a low-frequency RDF loop antenna.

4. The distance of a target from a radar station is called
   (a) the resolution.
   (b) the azimuth.
   (c) the range.
   (d) the transform.

5. Which of the following devices is best for use as an ultrasonic pickup?
   (a) A back-pressure transducer
   (b) A piezoelectric transducer

(c) An elastomer pressure transducer
(d) A capacitive pressure transducer

6. What do loran and the GPS have in common?
(a) They both operate using IR energy.
(b) They both employ loop antennas at the transmitting and receiving stations.
(c) They both involve measurement of azimuth angles to determine the location of a transmitter.
(d) They both involve measurement of distances to determine the location of a receiver.

7. A bright spot close to, and directly to the left of, the center of a conventional radar display indicates the presence of a target
(a) right over the observing station.
(b) at close range, west of the observing station.
(c) at close range, south of the observing station.
(d) far from the observing station, azimuth unknown.

8. Which of the following systems enables a vessel to navigate entirely from its own frame of reference, without the need for any electronic devices external to itself?
(a) An epipolar navigation system
(b) The GPS
(c) Loran
(d) Forget it! There is no such thing.

9. Ultrasonic waves travel through a vacuum
(a) at the same speed as they travel in air.
(b) at a slightly higher speed than they travel in air.
(c) at a slightly lower speed than they travel in air.
(d) Forget it! No form of acoustic wave can travel through a vacuum.

10. Which of the following devices or techniques makes use of artificially generated magnetic fields to facilitate location within a work environment?
(a) A dynamic transducer
(b) Epipolar navigation
(c) Capacitive proximity sensing
(d) None of the above

11. Which of the following is not an example of a transducer?
(a) A microphone
(b) A radio antenna
(c) A headset
(d) All of the above are examples of transducers.

12. A foam-like material, having a resistance that varies depending on how much it is compressed, is known as
    - (a) an elastomer.
    - (b) a fluxgate magnetometer.
    - (c) a piezoelectric substance.
    - (d) none of the above.
13. A stepper motor
    - (a) has torque that increases as its speed increases.
    - (b) resists applied torque when stopped with current going through its coils.
    - (c) has turning power that does not depend on the speed.
    - (d) is easily fooled by body capacitance.
14. A permanent magnet is a key component of
    - (a) an elastomer pressure sensor.
    - (b) a capacitive pressure sensor.
    - (c) a dynamic microphone.
    - (d) a piezoelectric microphone.
15. An electric motor is an example of
    - (a) an electromechanical transducer.
    - (b) a pressure sensor.
    - (c) a proximity sensor.
    - (d) a texture sensor.
16. A capacitive proximity sensor should not be expected to detect the presence of
    - (a) a wooden chair.
    - (b) a metal desk.
    - (c) a large dog.
    - (d) a steel post.
17. A well-designed photoelectric proximity sensor
    - (a) is easily fooled by loud noises.
    - (b) cannot detect the presence of objects that reflect light.
    - (c) is sensitive to magnetic fields.
    - (d) is not easily fooled by stray sources of light.
18. An acoustic transducer can translate an ac electrical signal into
    - (a) visible light.
    - (b) sound or ultrasound.
    - (c) a dc electrical signal.
    - (d) mechanical torque.

19. Body capacitance can be a problem with
    (a) electrostatic transducers.
    (b) elastomer pressure sensors.
    (c) capacitive pressure sensors.
    (d) all of the above.

20. Which of the following devices can be used to limit the torque or force applied by an electromechanical device such as a robot?
    (a) A back-pressure sensor
    (b) A capacitive proximity sensor
    (c) A dynamic transducer
    (d) An optical shaft encoder

31
CHAPTER

Acoustics, Audio, and High Fidelity

IN SOUND RECORDING AND REPRODUCTION, ESPECIALLY WITH MUSIC, FIDELITY SUPERSEDES ALL other considerations. In these applications, low distortion and "sound esthetics" are more important than amplifier efficiency or gain. In this chapter, we'll look at basic acoustics principles, and outline some *high-fidelity* (hi-fi) components and design concepts.

Acoustics

Acoustics is the science of sound waves. Sound consists of molecular vibrations at audio frequency (AF), ranging from about 20 Hz to 20 kHz. Young people can hear the full range of AF; older people lose hearing sensitivity at the upper and lower extremes.

Audio Frequencies

In music, the AF range is divided into three broad, vaguely defined parts, called *bass* (pronounced "base"), *midrange*, and *treble.* The bass frequencies start at 20 Hz and extend to 150 or 200 Hz. Midrange begins at this point, and extends up to 2 or 3 kHz. Treble consists of the audio frequencies higher than midrange. As the frequency increases, the wavelength becomes shorter.

In air, sound travels at about 1100 feet per second (ft/s), or 335 meters per second (m/s). The relationship between the frequency f of a sound wave in hertz, and the wavelength λ_{ft} in feet, is as follows:

$$\lambda_{\mathrm{ft}} = 1100/f$$

The relationship between f in hertz and λ_{m} in meters is given by:

$$\lambda_{\mathrm{m}} = 335/f$$

This formula is also valid for frequencies in kilohertz and wavelengths in millimeters.

A sound disturbance in air at 20 Hz has a wavelength of 55 ft (17 m). A sound of 1.0 kHz produces a wave measuring 1.1 ft (34 cm). At 20 kHz, a sound wave in air is only 0.055 ft (17 mm)

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

long. In substances other than air at sea level, such as air at extreme altitudes, freshwater, saltwater, or metals, the preceding formulas do not apply.

Waveforms

The frequency, or *pitch*, of a sound is only one of several variables that acoustic waves can possess. Another important factor is the shape of the wave itself. The simplest acoustic waveform is a sine wave (or *sinusoid*), in which all of the energy is concentrated at a single frequency. Sinusoidal sound waves are rare in nature. A good artificial example is the beat note, or heterodyne, produced by a steady carrier in a communications receiver.

In music, most of the notes are complex waveforms, consisting of energy at a specific fundamental frequency and its harmonics. Examples are sawtooth, square, and triangular waves. The shape of the waveform depends on the distribution of energy among the fundamental and the harmonics. There are infinitely many different shapes that a wave can have at a single frequency such as 1 kHz. As a result, there is infinite variety in the *timbre* that a single musical note can have.

Path Effects

A flute, a clarinet, a guitar, and a piano can each produce a sound at 1 kHz, but the tone quality is different for each instrument. The waveform affects the way a sound is reflected from objects. Acoustics engineers must consider this when designing sound systems and concert halls. The goal is to make sure that all the instruments sound realistic everywhere in the room.

Suppose you have a sound system set up in your living room, and that, for the particular placement of speakers with respect to your ears, sounds propagate well at 1, 3, and 5 kHz, but poorly at 2, 4, and 6 kHz. This affects the way musical instruments sound. It distorts the sounds from some instruments more than the sounds from others. Unless all sounds, at all frequencies, reach your ears in the same proportions that they come from the speakers, you do not hear the music the way it originally came from the instruments.

Figure 31-1 shows a listener, a speaker, and three sound reflectors, also known as *baffles.* The waves reflected by the baffles (X, Y, and Z), along with the direct-path waves (D), add up to some-

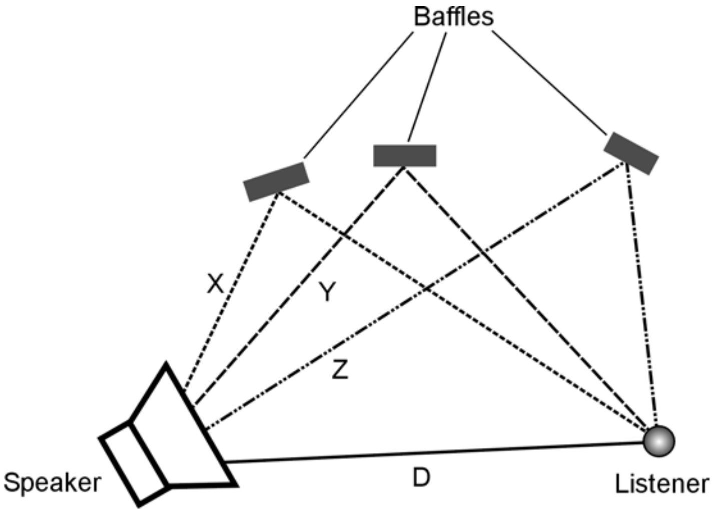

31-1 Sounds from reflected paths (such as X, Y, and Z) combine with direct-path sound (D) to produce what a listener hears.

thing different, at the listener's ears, for each frequency of sound. This phenomenon is impossible to prevent. That is why it is so difficult to design an acoustical room, such as a concert auditorium, to propagate sound well at all frequencies for every listener.

Loudness and Phase

You do not perceive the *loudness* (also called *volume*) of sound in direct proportion to the power contained in the disturbance. Your ears and brain sense sound levels according to the logarithm of the actual intensity. Another variable is the *phase* with which waves arrive at your ears. Phase allows you to perceive the direction from which a sound is coming, and it also affects perceived sound volume.

The Decibel in Acoustics

You have already learned about decibels in terms of signal voltage, current, and power. Decibels are also used in acoustics, and in this application, they are considered in terms of relative power. If you change the volume control on a hi-fi set until you can just barely tell the difference, the increment is one decibel (1 dB). If you use the volume control to halve or double the actual acoustic-wave power coming from a set of speakers, you perceive a change of 3 dB.

For decibels to have meaning in acoustics, there must be a reference level against which everything is measured. Have you read that a vacuum cleaner produces 80 dB of sound? This is determined with respect to the *threshold of hearing*, which is the faintest sound that a person with good hearing can detect in a *quiet room* specially designed to have a minimum of background noise.

Phase in Acoustics

Even if there is only one sound source, acoustic waves reflect from the walls, ceiling, and floor of a room. In Fig. 31-1, imagine the baffles as two walls and the ceiling in a room. As is the case with baffles, the three sound paths X, Y, and Z are likely to have different lengths, so the sound waves reflected from these surfaces will not arrive in the same phase at the listener's ears. The direct path (D), a straight line from the speaker to the listener, is always the shortest path. In this situation, there are at least four different paths by which sound waves can propagate from the speaker to the listener. In some practical scenarios, there are dozens.

Suppose that, at a certain frequency, the acoustic waves for all four paths happen to arrive in exactly the same phase in the listener's ears. Sounds at that frequency will be exaggerated in volume. The same phase coincidence might also occur at harmonics of this frequency. This is undesirable because it causes acoustic peaks, called *antinodes*, distorting the original sound. At certain other frequencies, the waves might mix in phase opposition. This produces acoustic nulls called *nodes* or *dead zones.* If the listener moves a few feet, the volume at any affected frequency will change. As if this isn't bad enough, a new antinode or node might then present itself at another set of frequencies.

One of the biggest challenges in acoustical design is the avoidance of significant antinodes and nodes. In a home hi-fi system, this can be as simple as minimizing the extent to which sound waves reflect from the ceiling, the walls, the floor, and the furniture. Acoustical tile can be used on the ceiling, the walls can be papered or covered with cork tile, the floor can be carpeted, and the furniture can be upholstered with cloth. In large auditoriums and music halls, the problem becomes more complex because of the larger sound propagation distances involved, and also because of the fact that sound waves reflect from the bodies of the people in the audience!

Technical Considerations

Regardless of its size, a good hi-fi sound system must have certain characteristics. Here are two of the most important technical considerations.

Linearity

Linearity is the extent to which the output waveform of an amplifier is a faithful reproduction of the input waveform. In hi-fi equipment, all the amplifiers must be as linear as the state of the art allows.

If you connect a dual-trace oscilloscope (one that lets you observe two waveforms at the same time) to the input and output terminals of a hi-fi audio amplifier with good linearity, the output waveform is a vertically magnified duplicate of the input waveform. When the input signal is applied to the horizontal scope input and the output signal is applied to the vertical scope input, the display is a straight line. In an amplifier with poor linearity, the instantaneous output-versus-input function is not a straight line. The output waveform is not a faithful reproduction of the input, and distortion occurs. In some RF amplifiers this is all right. In hi-fi audio systems, it is not.

Hi-fi amplifiers are designed to work with input signals up to a certain peak (maximum instantaneous) amplitude. If the peak input exceeds this level, the amplifier becomes nonlinear, and distortion is inevitable. In a hi-fi system equipped with VU or distortion meters, excessive input causes the needles to kick up into the red range of the scale during peaks.

Dynamic Range

Dynamic range is a prime consideration in hi-fi recording and reproduction. As the dynamic range increases, the sound quality improves for music or programming having a wide range of volume levels. Dynamic range is expressed in decibels (dB).

At low volume levels, the limiting factor in dynamic range is the *background noise* in the system. In an analog system, most of this noise comes from the audio amplification stages. In tape recording, there is also some *tape hiss.* A scheme called *Dolby* (a trademark of Dolby Laboratories) is used in professional recording studios, and also in high-end consumer tape equipment, to minimize this hiss. Digital recording systems produce less internal noise than analog systems.

At high volume levels, the power-handling capability of an audio amplifier limits the dynamic range. If all other factors are equal, a 100-W audio system can be expected to have greater dynamic range than a 50-W system. The speaker size is also important. As speakers get physically larger, their ability to handle high power improves, resulting in increased dynamic range. This is why serious audio enthusiasts sometimes purchase sound systems with amplifiers and speakers that seem unnecessarily large.

Components

There are myriad ways to set up a hi-fi system. A true *audiophile* ("sound lover") assembles a complex system over a period of time. Here are some basic considerations that can serve as guidelines when choosing system components.

Configurations

The simplest type of home stereo system is contained in a single box, with an AM/FM radio receiver and a *compact disk* (CD) player. The speakers are generally external, but the connecting cables are short. The assets of a so-called compact hi-fi system are small size and low cost.

More sophisticated hi-fi systems have separate boxes containing components such as the following:

- An AM tuner
- An FM tuner
- An amplifier or pair of amplifiers
- A CD player
- A computer and its peripherals

The computer is optional, but it facilitates downloading music files or *streaming audio* from the Internet, creating ("burning") CDs, and composing and editing electronic music. A *satellite radio* receiver, a *tape player*, a *turntable*, or other nonstandard peripheral may also be included. The individual hardware units in this type of system, known as a *component hi-fi system*, are interconnected with shielded cables. A component system costs more than a compact system, but the sound quality is better, you get more audio power, you can do more tasks, and you can tailor the system to your preferences.

Some hi-fi manufacturers build all their equipment cabinets to a standard width so they can be mounted one above the other in a *rack.* A so-called rack-mounted hi-fi system saves floor space and makes the system look professional. The rack can be mounted on wheels so the whole system, except for the external speakers, can be rolled from place to place.

Figure 31-2 is a block diagram of a typical home stereo hi-fi system. The amplifier chassis is grounded to minimize hum and noise, and to minimize susceptibility to interference from external

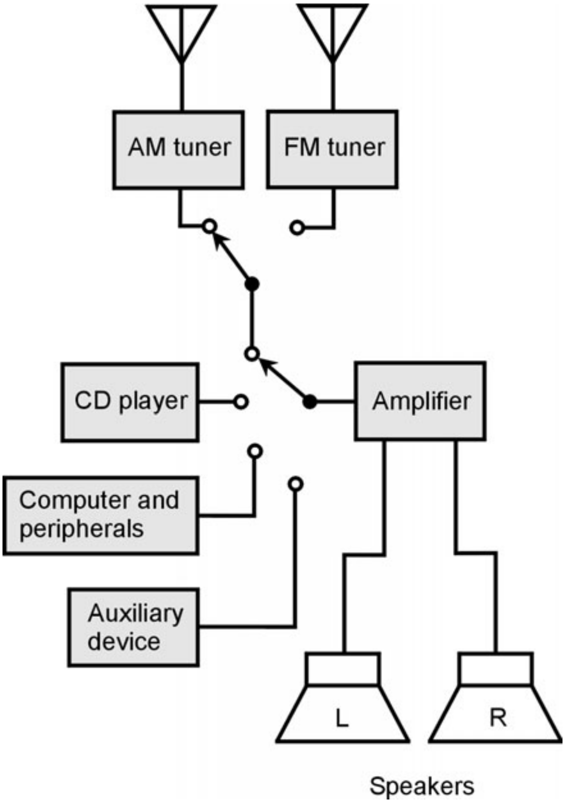

31-2 A basic stereo hi-fi system.

sources. The AM antenna is usually a loopstick built into the cabinet or mounted on the rear panel. The FM antenna can be an indoor type, such as television rabbit ears, or a directional outdoor antenna equipped with lightning protection hardware.

The Tuner

A *tuner* is a radio receiver capable of receiving signals in the standard AM broadcast band (535 to 1605 kHz) and/or the standard FM broadcast band (88 to 108 MHz). Tuners don't have built-in amplifiers. A tuner can provide enough power to drive a headset, but an amplifier is necessary to provide sufficient power to a pair of speakers.

Modern hi-fi tuners employ frequency synthesizers and have digital readouts. Most tuners have several *memory channels.* These are programmable, and allow you to select your favorite stations with a push of a single button, no matter where the stations happen to be in the frequency band. Most tuners also have *seek* and/or *scan* modes that allow the radio to automatically search the band for any station strong enough to be received clearly.

The Amplifier

In hi-fi, an amplifier delivers medium or high audio power to a set of speakers. There is at least one input, but more often there are three or more: one for a CD player, another for a tuner, and still others for auxiliary devices such as a tape player, turntable, or computer. Input requirements are a few milliwatts; the output can range up to hundreds of watts.

Amplifier prices increase with increasing power output. A simplified hi-fi amplifier forms the basis for a public-address system. Massive amplifiers are used by popular music bands. Some such systems employ vacuum tubes, because tubes offer electrical ruggedness and excellent linearity.

Speakers and Headsets

No amplifier can deliver sound that is better than the speakers will allow. Speakers are rated according to the audio power they can handle. It's a good idea to purchase speakers that can tolerate at least twice the audio output power that the amplifier can deliver. This will ensure that speaker distortion will not occur during loud, low-frequency sound bursts. It will also prevent physical damage to the speakers that might otherwise result from accidentally overdriving them.

Good speakers contain two or three individual units within a single cabinet. The *woofer* reproduces bass. The *midrange speaker* handles medium and, sometimes, treble (high) audio frequencies. A *tweeter* is designed especially for enhanced treble reproduction.

Headsets are rated according to how well they reproduce sound. This is a subjective consideration. Equally expensive headsets can, and often do, exhibit huge differences in the quality of the sound that they put out. Not only that, but people disagree about what constitutes good sound.

Balance Control

In hi-fi stereo sound equipment, the *balance control* allows adjustment of the relative volumes of the left and right channels.

In a basic hi-fi system, the balance control consists of a single rotatable knob connected to a pair of potentiometers. When the knob is rotated counterclockwise, the left-channel volume increases and the right-channel volume decreases. When the knob is rotated clockwise, the right-channel volume increases and the left-channel volume decreases. In more sophisticated sound systems, the balance is adjusted by means of two independent volume controls, one for the left channel and the other for the right channel.

Proper balance is important in stereo hi-fi. A balance control can compensate for such factors as variations in speaker placement, relative loudness in the channels, and the acoustical characteristics of the room in which the equipment is installed.

Tone Control

The amplitude versus frequency characteristics of a hi-fi sound system are adjusted by means of a tone control or controls. In its simplest form, a tone control consists of a single knob or slide device. The counterclockwise, lower, or left-hand settings of this control result in strong bass and weak treble audio output. The clockwise, upper, or right-hand settings result in weak bass and strong treble. When the control is set to midposition, the audio response of the amplifier is more or less flat; that is, the bass, midrange, and treble are in roughly the same proportions as in the recorded or received signal.

Figure 31-3A shows how a single-knob tone control can be incorporated into an audio amplifier. The amplifier is designed so that the treble response is exaggerated. The potentiometer attenuates the treble to a variable extent.

A more versatile tone control has two capacitors and two potentiometers, as shown in Fig. 31-3B. One combination is in series, and the other is in parallel. The series-connected resistance-capacitance (*RC*) circuit is connected in parallel with the audio output, and it attenuates the treble to a variable extent. The parallel *RC* circuit is in series with the audio path, and it attenuates the bass to a variable extent. The two potentiometers can be adjusted separately, although there is some interaction.

Audio Mixer

If you simply connect two or more audio sources to the same input of an amplifier, you can't expect good results. Different signal sources (such as a computer, a tuner, and a CD player) are likely to

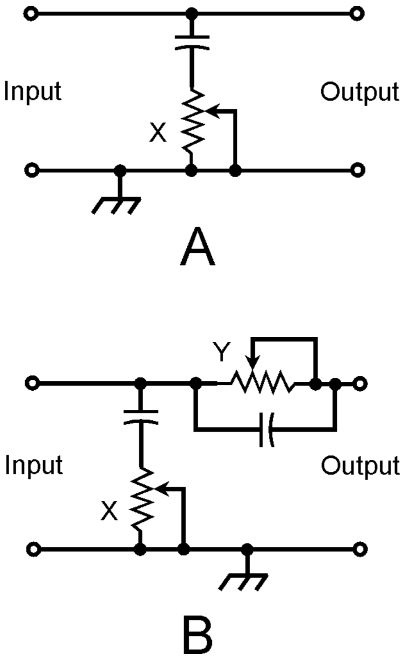

31-3 Methods of tone control. At A, a single potentiometer/capacitor combination (X) provides treble attenuation only. At B, one potentiometer/capacitor combination (X) attenuates the treble, and the other (Y) attenuates the bass.

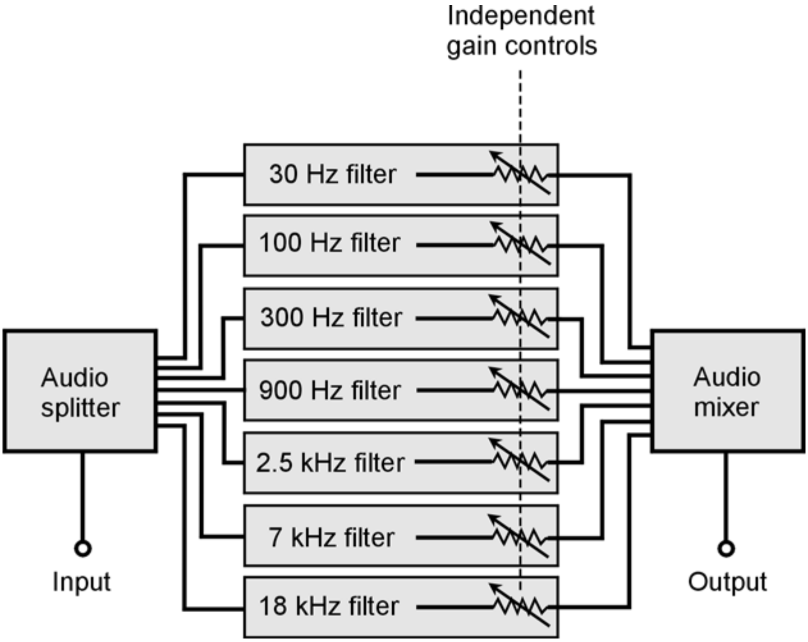

31-4 A graphic equalizer consists of an audio splitter, several bandstop filters, several gain controls, and an audio mixer.

have different impedances. When connected together, the impedances appear in parallel. This can cause impedance mismatches for most or all of the sources, as well as at the amplifier input. The result will be degradation of system efficiency and poor overall performance.

Another problem arises from the fact that the signal amplitudes from various sources almost always differ. A microphone produces minuscule audio-frequency currents, whereas some tuners produce enough power to drive a pair of small loudspeakers. Connecting both of these together will cause the microphone signal to be obliterated by the signal from the tuner. In addition, the tuner output audio might damage the microphone.

An *audio mixer* eliminates all of the problems involved with the connection of multiple devices to a single channel. First, it isolates the inputs from each other, so there is no impedance mismatch or competition among the sources. Second, the gain at each input can be varied independently. This allows adjustment of amplitudes so the signals blend in the desired ratio.

Graphic Equalizer

A *graphic equalizer* is a device for adjusting the relative loudness of audio signals at various frequencies. It allows tailoring of the amplitude versus frequency output of hi-fi sound equipment. Equalizers are used in recording studios and by serious hi-fi stereo enthusiasts. There are several independent gain controls, each one affecting a different part of the audible spectrum. The controls are slide potentiometers with calibrated scales. The slides move up and down, or, in some cases, left to right. When the potentiometers are set so that the slides are all at the same level, the audio output or response is flat, meaning that no particular range is amplified or attenuated with respect to the whole AF spectrum. By moving any one of the controls, the user can adjust the gain within a certain frequency range without affecting the gain outside that range. The positions of the controls on the front panel provide an intuitive graph of the output or response curve.

Figure 31-4 is a block diagram of a hypothetical graphic equalizer with seven gain controls. The input is fed to an *audio splitter* that breaks the signal into seven paths of equal impedance, and prevents interaction among the circuits. The seven signals are fed to *audio attenuators*, also called *bandstop filters*, each filter having its own gain control. (The center frequencies of the attenuators in this example are at 30, 100, 300, and 900 Hz, and 2.5, 7, and 18 kHz. These are not standard frequencies, and are given here only for illustrative purposes.) The slide potentiometers affect the extent to which each filter affects the gain within its frequency range. Finally, the signals pass through an audio mixer, and the composite is sent to the output.

There are several challenges in the design and proper use of graphic equalizers. The filter gain controls must not interact. Judicious choice of filter frequencies and responses is important. The filters must not introduce distortion. The active devices must not generate significant audio noise. Graphic equalizers are not built to handle high power, so they must be placed at low-level points in an audio amplifier chain. In a multichannel circuit such as a stereo sound system, a separate graphic equalizer can be used for each channel.

Specialized Systems

Mobile and portable hi-fi systems operate at low dc voltages. Typical audio power levels are much lower than in home hi-fi systems. Speakers are much smaller also. In portable systems, headsets are often used in place of speakers.

Mobile Systems

Mobile hi-fi systems, designed for cars and trucks, usually have four speakers. The left and right channels each supply a pair of speakers. The left stereo channel drives the left front and left rear speakers; the right stereo channel drives the right front and right rear speakers. The balance control adjusts the ratio of sound volume between the left and right channels for both the front and rear speaker sets. Another control adjusts the ratio of sound volume between the front and rear sets.

A mobile hi-fi system typically has an AM/FM receiver and a CD player. Some older vehicles have systems with cassette tape players. Newer, high-end cars and trucks have satellite radio receivers as well as conventional AM/FM receivers and CD players. One note of caution: compact disks and cassettes are heat-sensitive, so they should not be stored in a car or truck that will be left out in the sun on a warm day.

Portable Systems

Portable hi-fi systems operate from sets of dry cells or rechargeable cells. The most well-known portable hi-fi set is the so-called headphone radio. There are literally hundreds of different designs. Some include only an FM radio; some have AM/FM reception capability. Some have a small box with a cord that runs to the headset; others are entirely contained within the headset. There are portable CD players and even portable satellite radio receivers. The sound quality from these systems can be excellent. The defining factor is the quality of the headset.

Another form of portable hi-fi set, sometimes called a *boom box*, can produce several watts of audio output, and delivers the sound to a pair of speakers built into the box. A typical boom box is about 8 in high by 18 in wide by 6 in deep. It includes an AM/FM radio and a CD player. The system gets its name from the loud bass acoustic energy peaks its speakers can deliver.

Quadraphonic Sound

Quadraphonic sound refers to four-channel audio recording and reproduction. It is also called *quad stereo* or *four-channel stereo.* Each of the four channels is independent of the other three. In a well-designed quadraphonic sound system, the speakers should be level with the listener, equidistant from the listener, and separated by angles of 90° from the listener's point of view. If the listener is facing north, for example, the left front speaker is to the northwest, the right front speaker is to the northeast, the left rear speaker is to the southwest, and the right rear speaker is to the southeast. This provides optimum balance, and also facilitates the greatest possible left-to-right and front-to-rear contrast in the perceived sounds.

Recorded Media

Methods of recording sound, particularly music, have evolved dramatically since the ascent of digital technology. Several types of media are available, but the compact disk (CD) is the most common. Other media include *analog audio tape*, *digital audio tape*, and (of historical interest) *vinyl disk.*

Compact Disk

A compact disk (often spelled "disc" in hi-fi applications), also called a CD, is a plastic disk with a diameter of 4.72 in (12.0 cm), on which data can be recorded in digital form. Any kind of data can be digitized: sound, images, and computer programs and files. This data can also be stored on other digital media of sufficient capacity, such as a computer hard drive or a backup tape drive.

Digital sound, recorded on the surface of a CD, is practically devoid of the hiss and crackle that bedevil recordings on analog media. This is because the information on the disk is binary: a bit (binary digit) is either 1 (high) or 0 (low). The distinction between these two states is more clear-cut than the subtle fluctuations of an analog signal.

When a CD is prepared, the sound is first subjected to *analog-to-digital* (A/D) *conversion.* This changes the continuously variable audio waves into logic bits. These bits are then burned (literally) into the surface of the disk in the form of microscopic pits. The pits are arranged in a spiral track that would, if unwound, measure several kilometers in length. *Digital signal processing* (DSP) minimizes the noise introduced by environmental factors such as microscopic particles on the disk or random electronic impulses in circuit hardware. A *scrambling* process "smears" recordings throughout the disk, rather than burning the pits in a direct linear sequence. This further improves the signal-to-noise (S/N) ratio.

Compact-disk players recover the sound from a disk without any hardware physically touching the surface. A laser beam scans the disk. The beam is scattered by the pits and is reflected from the unpitted plastic. The result is a digitally modulated beam that is picked up by a sensor and converted into electrical currents. These currents proceed to a *descrambling* circuit, a *digital-to-analog* (D/A) *converter*, a DSP circuit, and audio amplifiers. Speakers or headphones convert the audio currents into sound waves.

With a CD player, the track location processes are entirely electronic, and they can all be done quickly. Tracks are assigned numbers that you select by pressing buttons. It is impossible to damage the CD, no matter how much you skip around among the songs. You can move instantly to any desired point within an individual track. You can program the system to play only those tracks you want, ignoring the others.

Analog Audio Tape

Analog audio tape recorders can be classified as either cassette or reel-to-reel. A typical audio cassette plays for 30 minutes on each side; longer-playing cassettes allow recording for as much as 60 minutes per side. The longer tapes are thinner and more subject to stretching than the shorter tapes.

A reel-to-reel tape feed system resembles that of an old-fashioned movie projector. The tape is wound on two flat spools called the *supply reel* and the *take-up reel.* The reels rotate counterclockwise as the tape passes through the recording/playback mechanism. When the take-up reel is full and the supply reel is empty, both reels can be flipped over and interchanged. This allows recording or playback on the other side of the tape. (Actually, the process takes place on the same side of the tape, but on different tracks.) The speed is usually 1.875, 3.75, or 7.5 inches per second (ips).

Figure 31-5 is a simplified functional diagram of an analog audio tape recording/playback mechanism.

In the *record mode*, the tape moves past the *erase head* before anything is recorded on it. If the tape is not blank (that is, if magnetic impulses already exist on it), the erase head removes these before anything else is recorded. This prevents *doubling*, or the simultaneous presence of two programs on the tape. The erase head can be disabled in some tape recorders when doubling is desired. The *recording head* is a small electromagnet. It generates a fluctuating magnetic field whose instantaneous flux density is proportional to the instantaneous level of the audio input signal. This magnetizes the tape in a pattern that duplicates the waveform of the signal. The *playback head* is usually not activated in the record mode. However, the playback head and circuits can be used while recording to create an echo effect.

In the *playback mode*, the erase head and recording head are not activated. The playback head acts as a sensitive magnetic-field detector. As the tape moves past it, the playback head is exposed to a fluctuating magnetic field whose waveform is identical to that produced by the recording head when the audio was originally recorded on the tape. This magnetic field induces weak alternating

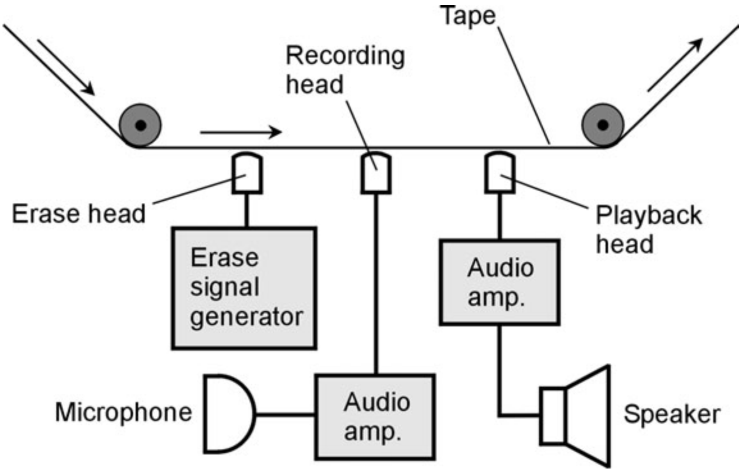

31-5 Simplified functional diagram of the recording and playback hardware in an analog audio tape recorder/player.

currents in the playback head. These currents are amplified and delivered to a speaker, headset, or other output device.

Digital Audio Tape

Digital audio tape (DAT) is magnetic recording tape on which binary digital data can be recorded. In digital audio recording, tape noise is practically eliminated because such noise is analog in nature. Some electronic noise is generated in the analog amplification stages following D/A conversion, but this is minimal compared with the noise generated in older, fully analog systems. The reduced noise in DAT equipment provides more true-to-life reproduction than is possible with analog methods.

With DAT, multigeneration copies can be made with practically no degradation in audio fidelity. The reason for this is the same as the reason a computer can repeatedly read and overwrite data on a magnetic disk. On DAT, the bits are represented by distinct magnetized regions on the tape. While analog signals are fuzzy in the sense that they vary continuously, digital signals are crisp. Imperfections in the recording apparatus, the tape itself, and the pickup head affect digital signals less than they affect analog signals. Digital signal processing can eliminate the minute flaws that creep into a digital signal each time it is recorded and played back.

Vinyl Disk

Vinyl disks were superseded years ago by CDs and Internet downloads, but some audiophiles are still intrigued by vinyl. Some vinyl disks, and the turntables that can play them, have attained considerable value as collectors' items. The main trouble with vinyl is that it can be physically damaged by even the slightest mishandling. Electrostatic effects can produce noise when the humidity is low, as in alpine or far northern regions in the winter.

Vinyl disks require a turntable that spins at speeds of 33 and 45 revolutions per minute (rpm). There are various drive systems. These are called *rim drive*, *belt drive*, and *direct drive.* The best type is a matter of individual taste because there are various factors to consider, such as cost, audio quality, ruggedness, and durability.

Electromagnetic Interference

As hi-fi equipment becomes more sophisticated and complex, the circuits become more susceptible to interference from outside sources, particularly electromagnetic (EM) fields. This problem is known as *electromagnetic interference* (EMI). Sometimes it is called *radio-frequency interference* (RFI).

If a radio transmitter is operated near a hi-fi stereo system, the radio signals can be intercepted by the hi-fi wiring and peripherals, and delivered to the amplifier. Unshielded interconnecting cables act as radio receiving antennas. This problem is exacerbated if any of the connecting cables happen to resonate at the operating frequency of the radio transmitter. In the hi-fi amplifier, the RF currents are rectified, causing changes in the audio gain. Sometimes the signal data can be heard in the speakers or headset.

In most cases when EMI takes place in a hi-fi setup, the fault exists in the stereo system design, not in the radio transmitter. The transmitter system is merely doing its job: generating and radiating electromagnetic signals!

Several steps can be taken when installing a stereo hi-fi system to minimize the likelihood that EMI will occur:

- Connect the stereo amplifier chassis to a good electrical ground.
- Use shielded interconnecting cables as much as possible.
- Use shielded (coaxial) speaker cables.
- Keep all cables as short as possible.

If you have an amateur or citizens band (CB) radio station in your house and it causes EMI to your hi-fi system, perform these additional steps:

- Locate the radio transmitting antenna as far from the hi-fi equipment as possible.
- Use the lowest possible transmitter output power that will ensure reliable communications.

Unfortunately, EMI problems can sometimes prove nigh impossible to eliminate. This can be especially troublesome for amateur and CB radio operators when it damages relations with neighbors. In these cases, old-fashioned diplomacy may work better than engineering-based attempts at resolution.

Quiz

Refer to the text in this chapter if necessary. A good score is 18 correct. Answers are in the back of the book.

1. A young person can hear sounds at frequencies as high as approximately which of the following?
   (a) 20 Hz
   (b) 20 kHz
   (c) 20 MHz
   (d) 20 GHz
2. Electromagnetic interference to a hi-fi amplifier can occur in the presence of
   (a) a nearby radio broadcast station.
   (b) improperly designed receiving antennas.
   (c) excessive utility voltage.
   (d) improper balance between the left and right channels.
3. The midrange audio frequencies
   (a) are exactly halfway between the lowest and highest audible frequencies.
   (b) represent sounds whose volume levels are not too loud or too soft.
   (c) are above the treble range but below the bass range.
   (d) None of the above are true.
4. In the acoustical design of a room intended for a home audio system,
   (a) the use of small speakers can minimize distortion.
   (b) reflection of sound waves from walls should be minimized.
   (c) the walls should all intersect at perfect 90° angles.
   (d) wooden furniture, without upholstery, should be used.

5. A change of +10 dB in an audio signal represents
    - (a) a doubling of acoustic power.
    - (b) a threefold increase in acoustic power.
    - (c) a tenfold increase in acoustic power.
    - (d) no change in acoustic power, but a change in frequency.

6. What is the frequency of an acoustic disturbance whose wavelength is 120 mm in air?
    - (a) 279 Hz
    - (b) 2.79 kHz
    - (c) 35.8 Hz
    - (d) 358 Hz

7. What is the frequency of a sound wave that propagates at a speed of 1100 ft/s?
    - (a) 33.5 Hz
    - (b) 335 Hz
    - (c) 3.35 kHz
    - (d) Forget it! The frequency of a sound wave is independent of the propagation speed.

8. The relative phase of two acoustic waves from the same source at the same time, one wave direct and one wave reflected from a wall, can affect
    - (a) the positions of antinodes and nodes.
    - (b) the perceived frequency.
    - (c) the positions of antinodes and nodes, and the perceived frequency.
    - (d) neither the positions of the antinodes and nodes, nor the perceived frequency.

9. In an acoustic sine wave,
    - (a) the frequency and phase are identical.
    - (b) the sound power is inversely proportional to the frequency.
    - (c) the sound power is directly proportional to the frequency.
    - (d) all of the sound power is concentrated at a single frequency.

10. Vinyl disks are
    - (a) susceptible to physical damage.
    - (b) useful primarily in high-power sound systems.
    - (c) digital media.
    - (d) preferred for off-the-air sound recording.

11. If an amplifier introduces severe distortion in the waveforms of input signals, then that amplifier is
    - (a) not delivering enough power.
    - (b) operating at the wrong frequency.
    - (c) operating in a nonlinear fashion.
    - (d) being underdriven.

12. Suppose a 10-W amplifier is used with speakers designed for a 100-W amplifier. Which of the following statements is true?

   (a) The speakers are capable of handling the amplifier output.
   (b) The amplifier might be damaged by the speakers.
   (c) Electromagnetic interference is more likely to occur than would be the case if the speakers were designed for a 10-W amplifier.
   (d) The speakers are likely to produce excessive distortion.

13. Which of the following frequencies cannot be received by an AM/FM tuner?

   (a) 830 kHz
   (b) 95.7 kHz
   (c) 90.1 MHz
   (d) 107.3 MHz

14. Which of the following statements about woofers is true?

   (a) They are especially useful for reproducing the sounds of barking dogs.
   (b) They are designed to handle short, intense bursts of sound.
   (c) They should not be used with graphic equalizers.
   (d) They are specifically designed to reproduce low-frequency sounds.

15. Suppose you have an amateur radio station and its transmitter causes EMI to your hi-fi system. Which of the following would almost certainly *not* help?

   (a) Build a new transmitter that works on the same frequencies with the same power output as your existing transmitter.
   (b) Reduce the transmitter output power.
   (c) Install shielded speaker wires in the hi-fi system, and be sure the system is well grounded.
   (d) Move the amateur radio transmitting antenna to a location farther away from the hi-fi system.

16. In an analog audio tape recorder/player, the recording head

   (a) converts sound waves to radio signals.
   (b) converts sound waves to fluctuating electric current.
   (c) converts AF current to a fluctuating magnetic field.
   (d) converts dc to AF current.

17. A rack-mounted hi-fi system

   (a) can save floor space.
   (b) is more susceptible to EMI than a compact system.
   (c) is cheaper than a compact system.
   (d) is designed especially for use with vinyl disks and turntables.

18. An audio mixer

   (a) cannot match impedances among interconnected components.
   (b) cannot increase the audio output of an amplifier.

(c) eliminates EMI problems in all but the worst cases.
(d) allows a microphone to be used as a speaker.

19. Which of the following devices or circuits ordinarily employs D/A conversion?
    (a) A portable CD player
    (b) A turntable for use with vinyl disks
    (c) An analog audio tape recording/playback system
    (d) An *RC* tone control

20. Which of the following is an advantage of digital audio over analog audio?
    (a) Digital media can be used to record and play back audio at higher frequencies than can analog media.
    (b) Multigeneration copies of digital audio programs can be made without degradation in fidelity, but this is not true of analog audio programs.
    (c) Digital audio is compatible with vinyl disks and old-fashioned reel-to-reel and cassette tape, but analog audio is not.
    (d) There are no genuine advantages of digital audio over analog audio. In fact, analog audio is superior to digital audio in every respect.

32
CHAPTER

Personal and Hobby Wireless

THE TERM *WIRELESS* AROSE IN THE EARLY 1900S WHEN INVENTORS AND EXPERIMENTERS BEGAN sending and receiving messages using electromagnetic fields. Gradually, the terminology changed, and wireless became known as *radio*, *television*, and *electromagnetic communications.* In the 1980s and 1990s, the term *wireless* emerged again, this time in a consumer context.

Cellular Communications

Wireless telephone sets are used in a specialized communications system called *cellular.* Originally, the cellular communications network was used mainly by traveling business people. Nowadays, many people regard *cell phones* as necessities, and some of them come equipped with extra features such as text messaging, Web browsing, and digital cameras.

How Cellular Systems Work

A cell phone looks like a cross between a cordless telephone receiver and a walkie-talkie, but smaller. The unit contains a radio transmitter and receiver combination called a *transceiver.* Transmission and reception take place on different frequencies, so you can talk and listen at the same time, and easily interrupt the other party if necessary. This capability is known as *full duplex.*

In an ideal cellular network, all the transceivers are always within range of at least one *repeater.* The repeaters pick up the transmissions from the portable units and retransmit the signals to the telephone network and to other portable units. The region of coverage for any repeater (also known as a *base station*) is called a *cell.*

When a cell phone is in motion, say in a car or on a boat, the set goes from cell to cell in the network. This situation is shown in Fig. 32-1. The curved, dashed line is the path of the vehicle. Base stations (dots) transfer access to the cell phone. This is called *handoff.* The hexagons show the limits of the transmission/reception range for each base station. All the base stations are connected to the regional telephone system. This makes it possible for the user of the portable unit to place

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

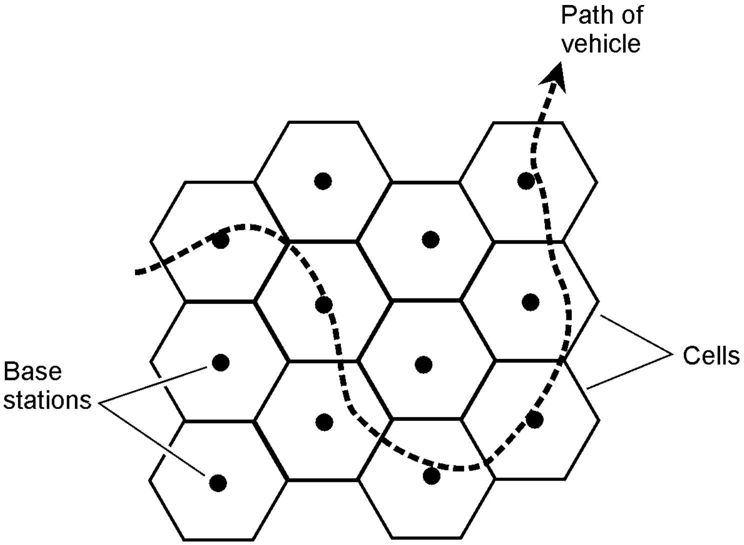

32-1 In an ideal cellular system, a moving vehicle is always within range of at least one base station.

calls to, or receive calls from, anyone else in the system, whether those other subscribers have cell phones or regular phones.

Older cellular systems occasionally suffer from call loss or breakup when signals are handed off from one repeater to another. This problem has been largely overcome by a technology called *code-division multiple access* (CDMA). In CDMA, the repeater coverage zones overlap significantly, but signals do not interfere with each other because every phone set is assigned a unique signal code. Rather than abruptly switching from one base-station zone to the next, the signal goes through a region in which it is sent through more than one base station at a time. This *make-before-break* scheme gets rid of one of the most annoying problems inherent in cellular communications.

In order to use a cellular network, you must purchase or rent a transceiver and pay a monthly fee. The fees vary, depending on the location and the amount of time per month you use the service. When using such a system, it is important to keep in mind that your conversations are not necessarily private. It is easier for unauthorized people to eavesdrop on wireless communications than to intercept wire or cable communications.

Cell Phones and Computers

You can connect a *personal computer* (PC) to a cell phone with a portable *modem* that converts incoming computer data from analog to digital form, and also converts outgoing data from digital to analog form. In this way, you can access the Internet from anywhere within range of a cellular base station. Figure 32-2 is a block diagram of this scheme.

Most commercial aircraft have telephones at each row of seats, complete with jacks into which you can plug a modem. If you plan to access the Internet from an aircraft, you must use the phones provided by the airline, not your own cell phone, because radio transceivers can cause interference to flight instruments. You must also observe the airline's restrictions concerning the operation of electronic equipment while in flight.

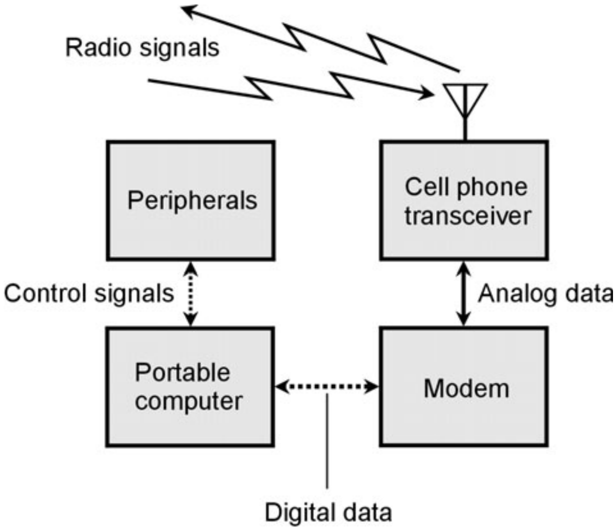

32-2 A cell phone can be equipped with a modem, allowing portable or mobile access to online computer networks.

Satellites and Networks

A satellite system is like a huge cellular network with the base stations (repeaters) located in space rather than on the earth's surface. The zones of coverage are large, and they change in size and shape if the satellite moves relative to the earth's surface.

Geostationary-Orbit Satellites

You learned about geostationary orbits in Chap. 25. Geostationary satellites are used in television (TV) broadcasting, in telephone and data communication, for gathering weather and environmental data, and for radiolocation.

In geostationary satellite networks, earth-based stations can communicate through a single "bird" only when the stations are both on a line of sight with the satellite. If two stations are nearly on opposite sides of the planet, say in Australia and Wisconsin, they must operate through two satellites to obtain a link (Fig. 32-3). In this situation, signals are relayed between the two satellites, as well as between either satellite and its respective earth-based station.

A potential problem with geostationary satellite links is the fact that the signal path is long enough so that perceptible propagation delays occur. This delay, and its observed effect, is known as *latency.* This doesn't cause problems with casual communications or Web browsing, but it slows things down when computers are linked with the intention of combining their processing power.

Low-Earth-Orbit Satellites

The earliest communications satellites orbited only a few hundred miles above the earth. They were *low-earth-orbit* (LEO) *satellites.* Because of their low orbits, LEO satellites took only about 90 minutes to complete one revolution. Communication was spotty, because a satellite was in range of any given ground station for only a few minutes at a time. This is the main reason why geostationary satellites became predominant once rocket technology progressed to the point where the necessary altitude and orbital precision could be obtained.

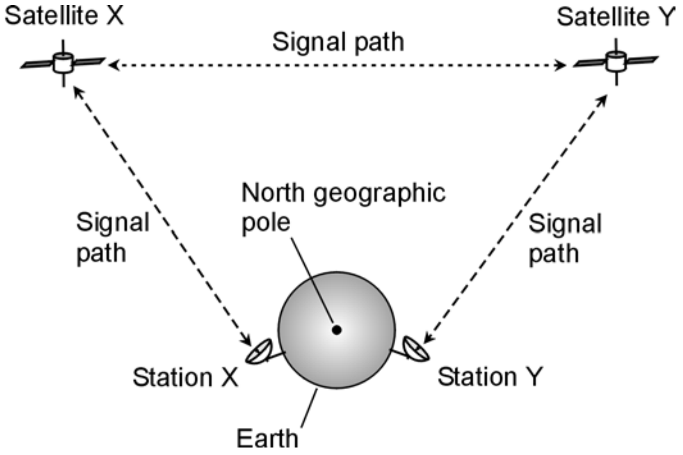

32-3 A communications link that employs two geostationary satellites.

However, geostationary satellites have certain limitations. A geostationary orbit requires constant adjustment, because a tiny change in altitude will cause the satellite to get out of sync with the earth's rotation. Geostationary satellites are expensive to launch and maintain. When communicating through them, there is always a delay because of the path length. It takes high transmitter power, and a sophisticated, precisely aimed antenna, to communicate reliably. These problems with geostationary satellites have brought about a revival of the LEO scheme. Instead of one single satellite, the new concept is to have a large fleet of them.

A good LEO satellite system is launched and maintained in such a way that, for any point on the earth, there is always at least one satellite in direct line-of-sight range. The satellites can relay messages throughout the fleet. Thus, any two points on the surface can always make, and maintain, contact through the satellites. The satellites are placed in *polar orbits* (routes that pass over or near the earth's geographic poles) to optimize the geographical coverage. Even if you're at or near the north geographic pole or the south geographic pole, you can use a LEO satellite system. This is not true of geostationary satellite networks, where the regions immediately around the geographic poles are not seen by the satellites.

A LEO satellite wireless communications link is easier to access and use than a geostationary satellite link. A small, simple antenna will suffice, and it doesn't have to be aimed in any particular direction. The transmitter can reach the network using only a few watts of power. The latency is less than 100 milliseconds (ms), compared with as much as 400 ms for geostationary satellite links.

Medium-Earth-Orbit Satellites

Some satellites revolve in orbits higher than those normally considered low-earth, but at altitudes lower than the geostationary level of 22,300 mi (36,000 km). These intermediate birds are called *medium-earth-orbit* (MEO) *satellites.* A MEO satellite takes several hours to complete each orbit. MEO satellites operate in fleets, in a manner similar to the way LEO satellites are deployed. Because the average MEO altitude is higher than the average LEO altitude, each bird can cover a larger region on the surface at any given time. A fleet of MEO satellites can be smaller than a comparable fleet of LEO satellites, and still provide continuous, worldwide communications.

The orbits of geostationary satellites are essentially perfect circles, and most LEO satellites orbit in near-perfect circles. But MEO satellites often have elongated, or elliptical, orbits. The point of

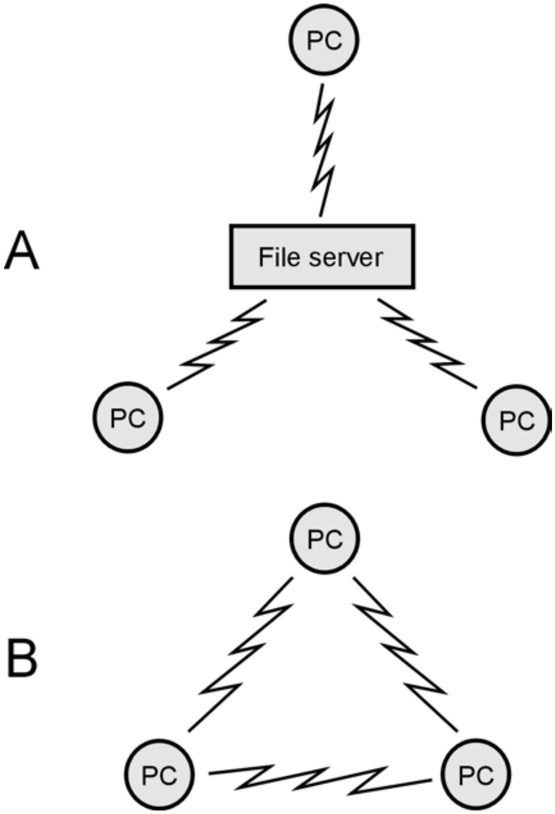

32-4 At A, a client-server wireless LAN. At B, a peer-to-peer wireless LAN.

lowest altitude is called *perigee*; the point of greatest altitude is called *apogee.* The apogee can be, and often is, much greater than the perigee. Such a satellite orbits at a speed that depends on its altitude. The lower the altitude, the faster the satellite moves. A satellite with an elliptical orbit crosses the sky rapidly when it is near perigee, and slowly when it is near apogee; it is easiest to use when its apogee is high above the horizon, because then it stays in the visible sky for a long time.

Every time a MEO satellite completes one orbit, the earth rotates beneath it. The rotation of the earth rarely coincides with the orbital period of the satellite. Therefore, successive apogees for a MEO satellite occur over different points on the earth's surface. This makes the tracking of individual satellites a complicated business, requiring computers programmed with accurate orbital data. For a MEO system to be effective in providing worldwide coverage without localized periodic blackouts, the orbits must be diverse, yet coordinated in a precise and predictable way. In addition, there must be enough satellites so that each point on the earth is always on a line of sight with one or more of the satellites, and preferably, there should be at least one bird in sight near apogee at all times.

Wireless Networks

A *local area network* (LAN) is a group of computers linked together within a building, campus, or other small region. The interconnections in early LANs were made with wire cables, but wireless links are increasingly common. A *wireless LAN* offers flexibility, because the computer users can move around without having to bother with plugging and unplugging cables. This arrangement is ideal when notebook computers (also known as laptops) are used. The way in which a LAN is arranged is called the *LAN topology.* There are two major wireless LAN topologies: the *client-server wireless LAN* and the *peer-to-peer wireless LAN.*

In the client-server topology (Fig. 32-4A), there is one large, powerful, central computer called a *file server*, to which all the smaller personal computers (labeled PC) are linked. The file server has enormous computing power, high speed, and large storage capacity, and can contain all the data for every user. End users do not communicate directly. All the data must pass through the file server.

In a peer-to-peer LAN (Fig. 32-4B), all of the computers in the network are PCs with more or less equal computing power, speed, and storage capacity. Each user maintains his or her own data. Subscribers can, and almost always do, communicate directly without the data having to pass through any intermediary. This offers greater privacy and individuality than the client-server topology, but it is slower when a large number of users need to share data.

Client-server LANs are favored by large institutions. Small businesses and schools, or departments within a larger corporation or university, prefer to use peer-to-peer LANs, mainly because they are cheaper and easier to maintain. In these illustrations, only three PCs are shown in the networks. But any LAN can have as few as two, or as many as several dozen, computers.

Home Internet users sometimes employ a modified version of the arrangement shown in Fig. 32-4A. In place of the file server, a device called a *wireless router* provides a hub through which the computers can communicate. The router is connected to the Internet by a high-speed interface such as a *cable modem*, allowing several computers in a household to have Internet access at the same time.

Amateur and Shortwave Radio

In most countries of the world, people can obtain government-issued licenses to send and receive messages by radio for nonprofessional purposes. In America, this hobby is called *amateur radio* or *ham radio.* If you want only to listen to communications and broadcasting, and not to transmit signals, you do not need a license in the United States (although you do need one in some countries).

Who Uses Amateur Radio?

Anyone can use ham radio, provided they can pass the tests necessary to obtain a license. Amateur radio operators (or hams) can communicate in numerous modes, including voice, Morse code, television, and various forms of text messaging. Text messaging can be done in real time, or by posting messages similar to *electronic mail* (e-mail). Radio amateurs have set up their own radio networks. Some of these networks have Internet gateways. This mode is known as *packet* radio.

Some radio hams chat about anything they can think of (except business matters, which are illegal to discuss via ham radio). Others like to practice emergency communications skills, so they can be of public service during crises such as hurricanes, earthquakes, or floods. Still others like to go out into the wilderness and talk to people far away while sitting out under the stars and using battery power.

Amateur radio operators communicate from cars, boats, aircraft, and bicycles; this is called *mobile operation.* When transceivers are used while walking or hiking, it is known as *portable* or *handheld operation.*

Amateur Equipment and Licensing

A simple ham radio station has a transceiver (transmitter/receiver), a microphone, and an antenna. A small station can fit on a desktop, and is about the size of a home computer or hi-fi stereo system.

Accessories can be added until a ham “rig” is a large installation, comparable to a small commercial broadcast station.

Figure 32-5 shows an example of a fixed, computer-controlled amateur radio station. The computer can be used to control the functions of the transceivers, and also to communicate using digital modes with other hams who own computers. The station can be equipped to interface with the telephone services, also called *landline.* The computer can control the antennas for the station, and can keep a log of all stations that have been contacted.

The best way to learn about ham radio in the United States is to contact the headquarters of the American Radio Relay League (ARRL), 225 Main Street, Newington, CT 06111. They maintain a Web site at www.arrl.org. If you live outside the United States, the ARRL can direct you to an organization in your home country that can help you obtain a license and get on the air.

Shortwave Listening

The high-frequency (HF) portion of the radio spectrum, at frequencies between 3 and 30 MHz, is sometimes called the *shortwave band.* This is a misnomer by contemporary standards; the waves are long compared with ultrahigh frequencies (UHF), microwaves, and IR, which are commonly used in wireless devices. In free space, a frequency of 3 MHz corresponds to a wavelength of 100 m; 30 MHz corresponds to 10 m. In the early years of radio when the shortwave band got its name, the wavelengths between 3 and 30 MHz were short compared with the wavelengths of most broadcast and communications signals, which had wavelengths in the kilometer range.

Anyone can build or obtain a shortwave or general-coverage radio receiver, install a modest outdoor antenna, and listen to signals from all around the world. This hobby is called *shortwave listening* or *SWLing.* In the United States, the proliferation of computers and online communications has, to some extent, overshadowed SWLing, and many young people grow up today ignorant of a

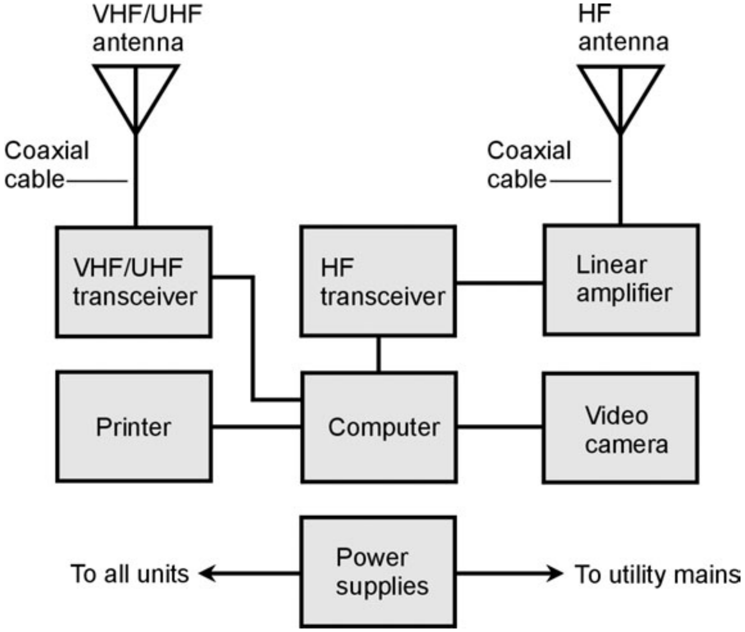

32-5 A computer-controlled amateur radio station.

realm of broadcasting and communications that still predominates in much of the world. But some people are still fascinated by the fact that people can contact each other using wireless devices alone, without the need for any human-made infrastructure other than an antenna at the source and another antenna at the destination. The ionosphere returns shortwave signals to the earth's surface, and allows reliable global broadcasting and communication to take place today, just as it has since the early 1900s.

Various commercially manufactured shortwave receivers are available. Most electronics stores carry one or more models, along with antenna equipment, for a complete installation. Amateur radio conventions, called "hamfests," can be a good source of shortwave receiving equipment at bargain prices. For information about events of this sort in your area, you can contact the American Radio Relay League at www.arrl.org, or pay a visit to your local amateur radio club.

Security and Privacy

People are becoming more and more concerned about the security and privacy of electronic communications, particularly wireless. When a wireless system is compromised, the intrusion is not detected until harm has been done to the system or to its subscribers.

Wireless versus Wired

Wireless eavesdropping differs from conventional *wiretapping* in two fundamental ways. First, eavesdropping is easier to do in wireless systems than in hard-wired systems. Those old-fashioned hard-wired phone sets might not always be convenient, but your privacy is more likely to be maintained than in the case with a system that uses any form of wireless. Second, eavesdropping of a wireless link is nigh impossible to physically detect, but a tap can usually be found in a hard-wired system.

If any portion of a communications link is done by wireless, then an eavesdropping receiver can be positioned within range of the RF transmitting antenna (Fig. 32-6) and the signals intercepted. The existence of a *wireless tap* has no effect on the electronic characteristics of any equipment in the system.

Levels of Security

There are four levels of telecommunications security, ranging from zero (no security) to the most secure connections technology allows.

No Security (*Level 0*): In a communications system with *level 0 security*, anyone can eavesdrop on a connection at any time, provided they are willing to spend the money and time to obtain the necessary equipment. Two examples of level 0 links are amateur radio and citizens band (CB) voice communications.

Wire Equivalent Security (*Level 1*): An end-to-end hard-wired connection requires considerable effort to tap, and sensitive detection apparatus can usually reveal the existence of any wiretap. A communications system with *level 1 security* must have certain characteristics in order to be effective and practical:

- The cost must be affordable.
- The system must be reasonably safe for transactions such as credit-card purchases.
- When network usage is heavy, the degree of privacy afforded to each subscriber should not decrease, relative to the case when network usage is light.

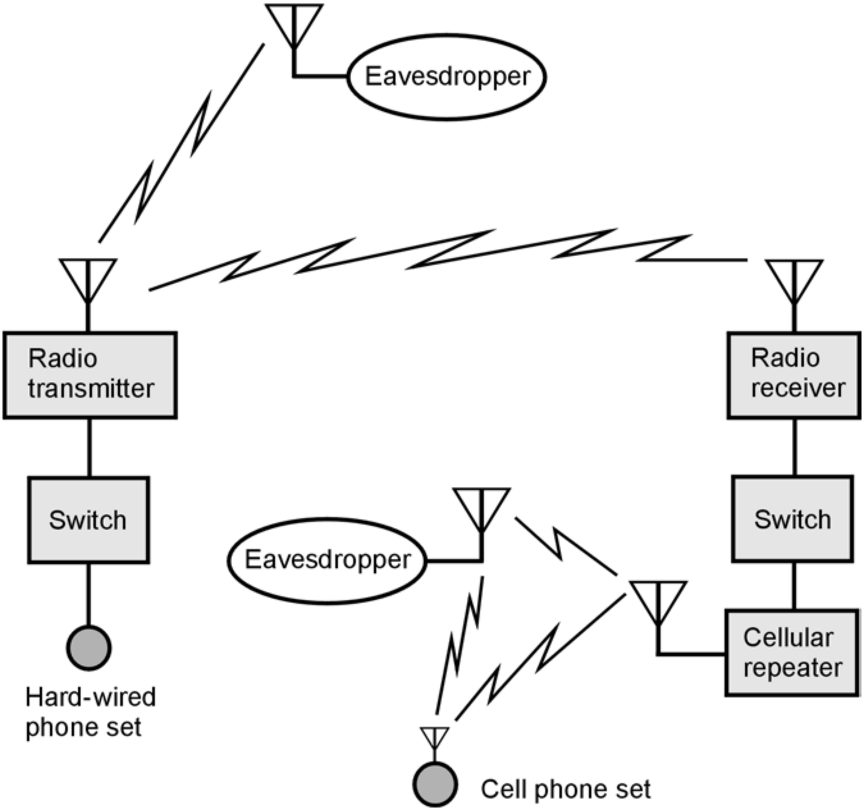

32-6 Eavesdropping on RF links in a telephone system. Heavy, straight lines represent wires or cables; zigzag lines represent RF signals.

- Ciphers, if used, should be unbreakable for at least 12 months, and preferably for 24 months or more.
- Encryption technology, if used, should be updated at least every 12 months, and preferably every six months.

Security for Commercial Transactions (*Level 2*): Some financial and business data demands protection even beyond the wire equivalent level. Many companies and individuals refuse to transfer money by electronic means because they fear criminals will gain access to an account. In a communications system with *level 2 security*, the encryption used in commercial transactions should be such that it would take a potential intruder (also called a *hacker*) at least 10 years, and preferably 20 years or more, to break the cipher. The technology should be updated at least every 10 years, but preferably every 3 to 5 years.

Military Level Security (*Level 3*): Security to *military specifications* (also called *mil spec*) involves the most sophisticated encryption available. Technologically advanced countries, and entities with economic power, have an advantage here. However, as technology gains ever more (and arguably too much) power over human activities, aggressor nations and terrorists might injure powerful nations by seeking out, and striking at, the weak points in communications infrastructures. In a communications system with *level 3 security*, the encryption scheme should be such that engineers believe it would take a hacker at least 20 years, and preferably 40 years or more, to break the cipher. The technology should be updated as often as economics allow.

Extent of Encryption

Security and privacy in wireless networks and communications systems can be achieved by means of *digital encryption.* The idea is to render signals readable only to receivers with the necessary *decryption key.* This makes it difficult for unauthorized people to gain access to the system.

For level 1 security, encryption is required only for the wireless portion(s) of the circuit. The cipher should be changed at regular intervals to keep it fresh. The block diagram of Fig. 32-7A shows *wireless-only encryption* for a hypothetical cellular telephone connection.

For security at levels 2 and 3, *end-to-end encryption* is necessary. The signal is encrypted at all intermediate points, even those for which signals are transmitted by wire or cable. Figure 32-7B shows this scheme in place for the same hypothetical cellular connection as depicted at A.

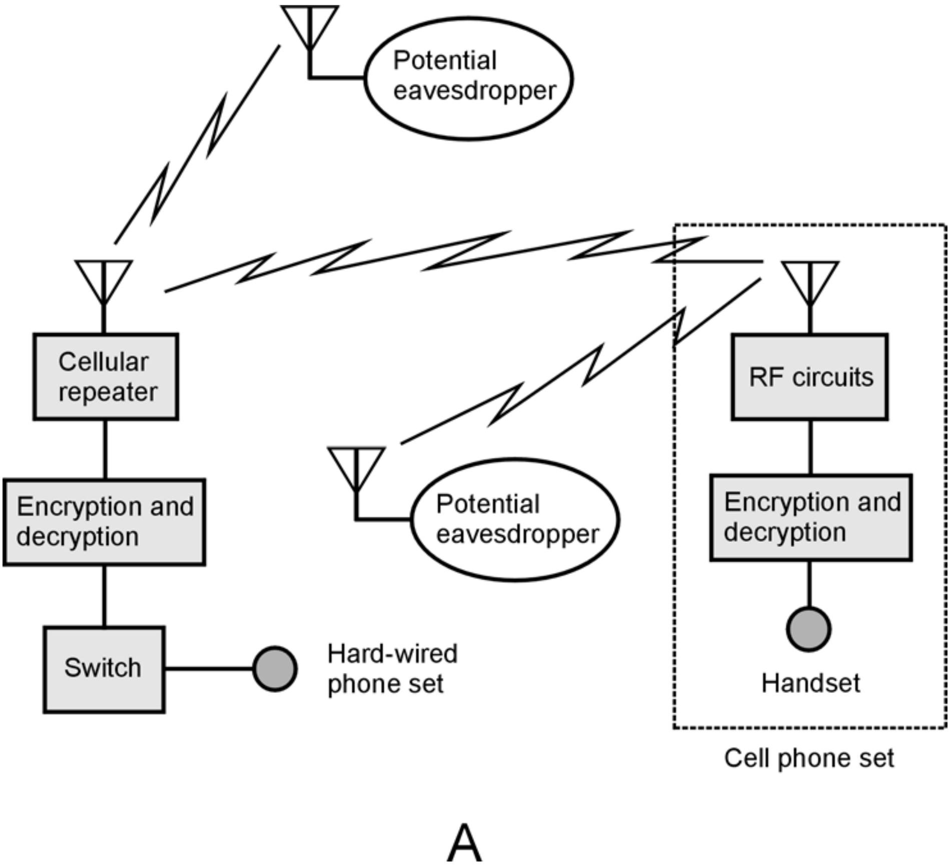
Potential eavesdropper
Cellular repeater
Encryption and decryption
Switch
Hard-wired phone set
Potential eavesdropper
RF circuits
Encryption and decryption
Handset
Cell phone set
A

32-7A Wireless-only encryption. Heavy, straight lines represent wires or cables; zigzag lines represent RF signals.

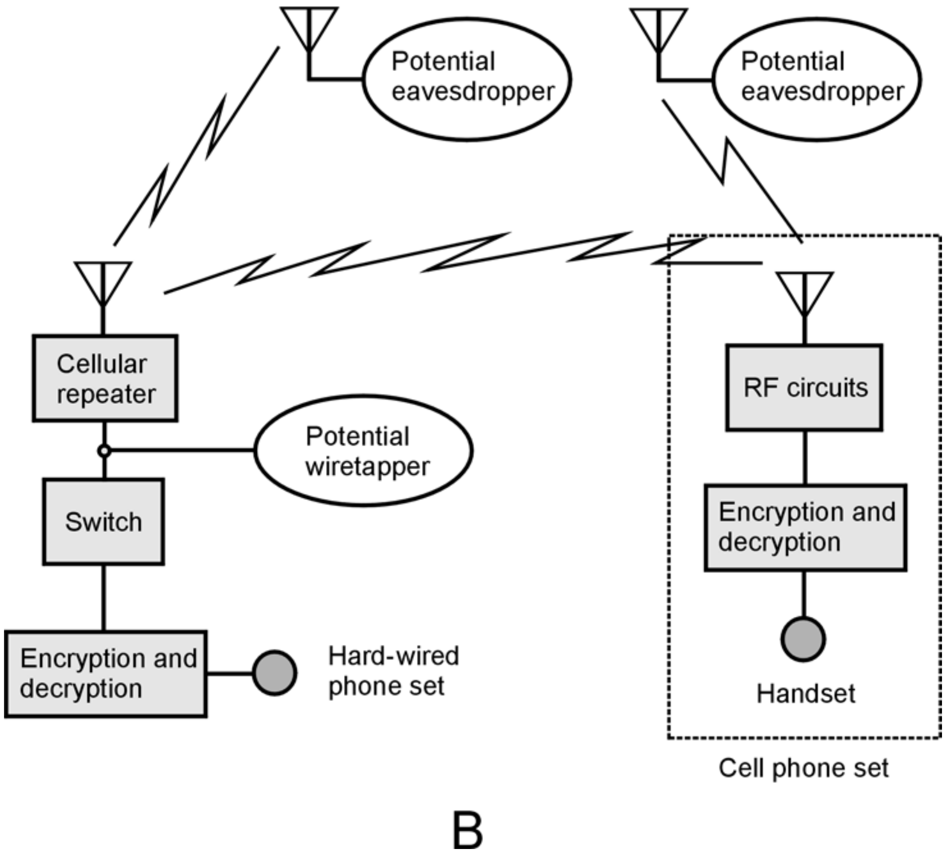
Potential eavesdropper
Potential eavesdropper
Cellular repeater
Potential wiretapper
Switch
Encryption and decryption
Hard-wired phone set
RF circuits
Encryption and decryption
Handset
Cell phone set
B

32-7B End-to-end encryption. Heavy, straight lines represent wires or cables; zigzag lines represent RF signals.

Security with Cordless Phones

Most cordless phones are designed to make it difficult for unauthorized people to pirate a telephone line. Prevention of eavesdropping is a lower priority, except in expensive cordless systems. If there is concern about using a cordless phone in a particular situation, a hard-wired phone set should be used.

If someone knows the frequencies at which a cordless handset and base unit operate, and if that person is determined to eavesdrop on conversations that take place using that system, it is possible to place a *wireless tap* on the line. The conversation can be intercepted at a point near the cordless phone set and its base unit, and then transmitted to a remote site (Fig. 32-8) and recorded there.

Security with Cell Phones

Cellular telephones are, in effect, long-range cordless phones. The wider coverage of cellular repeaters, as compared with cordless base units, increases the risk of eavesdropping and unauthorized use. Some cell phone vendors advertise their systems as “snoop proof.” Some of these claims have more merit than others. The word “proof” (meaning “immune”) should be regarded with skepticism. Digital encryption is the most effective way to maintain privacy and security of cellular communications. Nothing short of this is really of any use against a determined hacker.

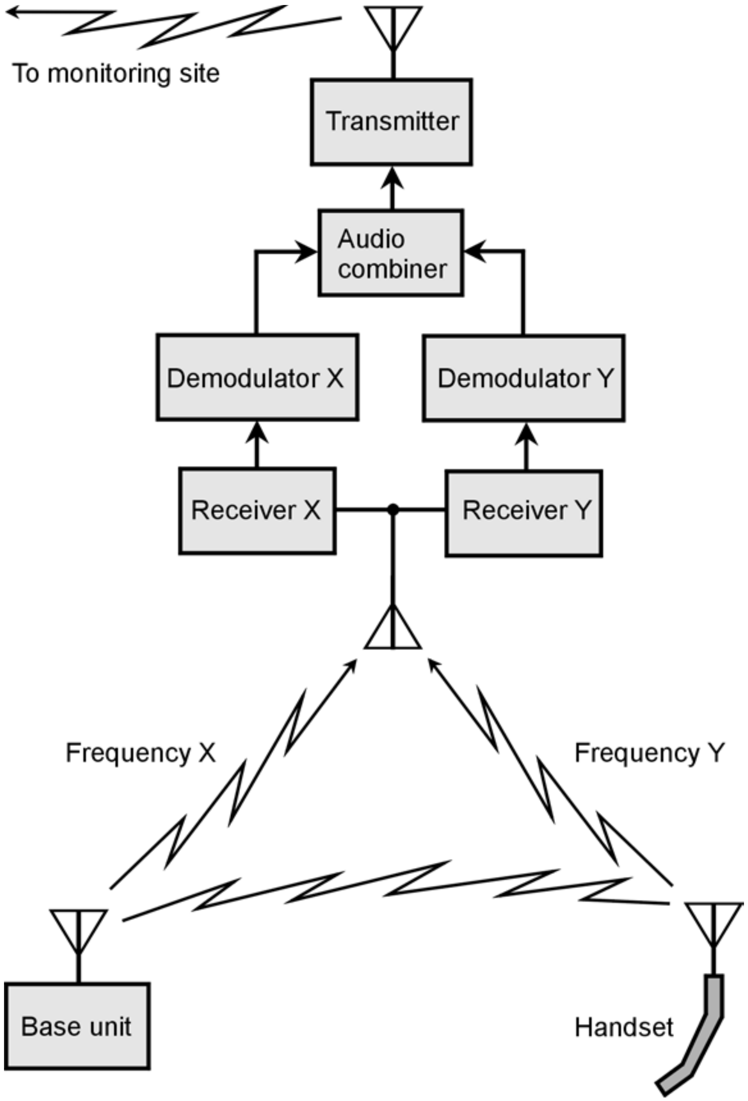
To monitoring site
Transmitter
Audio combiner
Demodulator X
Demodulator Y
Receiver X
Receiver Y
Frequency X
Frequency Y
Base unit
Handset

32-8 Wireless tapping of a cordless telephone.

Access and privacy codes, as well as data, must be encrypted if a cell phone system is to be maximally secure. If an unauthorized person knows the code with which a cell phone set accesses the system (the "name" of the set), rogue cell phone sets can be programmed to fool the system into thinking they belong to the user of the authorized set. This is known as *cell phone cloning.*

In addition to digital encryption of data, *user identification* (user ID) must be employed. The simplest is a *personal identification number* (PIN). More sophisticated systems can employ *voice-pattern recognition*, in which the phone set functions only when the designated user's voice speaks into it. *Hand-print recognition*, *electronic fingerprinting*, or *iris-print recognition* can also be employed. These are examples of *biometric security* measures.

Quiz

Refer to the text in this chapter if necessary. A good score is 18 correct. Answers are in the back of the book.

1. A network that employs one central computer and multiple personal computers is called
   (a) a wireless network.
   (b) a local area network.
   (c) a client-server network.
   (d) a peer-to-peer network.
2. In order for a receiver to render a digitally encrypted signal intelligible, that receiver must
   (a) have a directional antenna.
   (b) have a biometric security system.
   (c) have a decryption key.
   (d) be licensed by the government.
3. Which of the following devices or systems is *not* wireless?
   (a) An SWL station
   (b) An amateur radio station
   (c) A cordless phone set
   (d) A computer with a telephone modem
4. Eavesdropping of a wireless link
   (a) requires digital encryption.
   (b) requires a decryption key.
   (c) is difficult or impossible to detect.
   (d) includes all of the above.
5. In the United States, a license is required for
   (a) using a receiver on amateur radio frequencies.
   (b) transmitting on amateur radio frequencies.
   (c) using cordless phone sets.
   (d) using cell phone sets.
6. In a LEO satellite system, the repeaters are located
   (a) in space.
   (b) at strategic locations in most countries of the world.
   (c) at the north and south geographic poles.
   (d) over the equator.
7. In a geostationary satellite link, the repeater is located
   (a) on a buoy in the ocean.
   (b) at strategic locations in most countries of the world.
   (c) at the north and south geographic poles.
   (d) over the equator.

8. The term *shortwave*, in reference to radio, refers to signals having wavelengths of approximately
   (a) 10 mm to 100 mm.
   (b) 100 mm to 1 m.
   (c) 1 m to 10 m.
   (d) 10 m to 100 m.

9. Which of the following is illegal to do in the United States using a ham radio station?
   (a) Advertise used cars for sale
   (b) Transmit Morse code
   (c) Receive outside the amateur radio bands
   (d) Receive signals without a license

10. A file server is
    (a) a sensitive radio receiver equipped with D/A conversion.
    (b) a radio transmitter equipped with digital signal processing (DSP).
    (c) a central computer in a local area network.
    (d) a repeater in a LEO satellite system.

11. A device consisting of a receiver and transmitter in the same case or enclosure, and commonly used for wireless communications, is called
    (a) a file server.
    (b) a transceiver.
    (c) a transverter.
    (d) an encoder/decoder.

12. End-to-end encryption provides
    (a) maximum range in a wireless LAN.
    (b) a higher degree of security than wireless-only encryption.
    (c) simultaneous transmission and reception in an amateur radio station.
    (d) an easy means for hackers to clone cell phone sets.

13. A LAN in which each computer has more or less equal status, and in which each computer stores its own data, is known as a
    (a) wireless router.
    (b) wide-area LAN.
    (c) LAN topology.
    (d) peer-to-peer LAN.

14. With respect to security, the term *mil spec* refers to
    (a) wireless-only encryption.
    (b) a peer-to-peer network.
    (c) the use of a wireless router.
    (d) the highest obtainable level of security.

15. A radio signal is considered to be in the shortwave band if its frequency is which of the following:
    (a) 5 MHz
    (b) 50 MHz
    (c) 500 MHz
    (d) Any of the above

16. In a cellular network, the region covered by a single base station is called a
    (a) zone.
    (b) cell.
    (c) radius.
    (d) link.

17. An advantage of a conventional hard-wired telephone over a cell phone is the fact that a hard-wired phone set
    (a) affords better privacy.
    (b) offers greater portability.
    (c) is more convenient for use around the house.
    (d) employs digital encryption.

18. An advantage of a cell phone over a hard-wired phone is the fact that a cell phone set
    (a) affords better privacy.
    (b) costs less.
    (c) offers better mobility.
    (d) provides higher data speed.

19. In a two-way communications system with full duplex,
    (a) either party can hear the other at all times.
    (b) a cell phone transmits on two frequencies at the same time.
    (c) end-to-end encryption is necessary.
    (d) a file server or router is necessary.

20. An advantage of LEO satellite systems over geostationary satellite systems is the fact that
    (a) the latency is much greater with LEO satellite systems.
    (b) the equatorial regions are covered by LEO satellite systems, but not by geostationary satellite systems.
    (c) the regions near the geographic poles are covered by LEO satellite systems, but not by geostationary satellite systems.
    (d) All of the above are advantages of LEO satellite systems over geostationary satellite systems.

33
CHAPTER

A Computer and Internet Primer

COMPUTERS ARE USED FOR CALCULATIONS, COMMUNICATIONS, WORD PROCESSING, DATA PROCESSING, drawing, photo processing, music composition and editing, location, navigation, information searches, robot control, and many other purposes. Figure 33-1 is a block diagram showing the major parts of a typical computer system used in a home or small business.

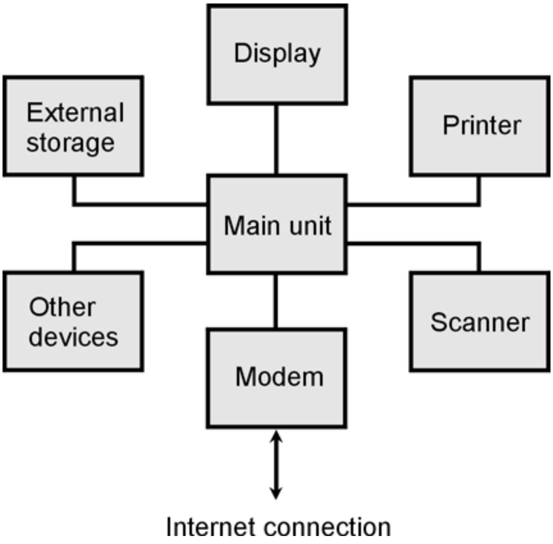

33-1 Components of a basic computer system for personal or small business use.

The Central Processing Unit

The *microprocessor* is the IC, or chip, that forms the core of your computer's "brain." It coordinates all the action and does all the calculations. It is located on the *motherboard*, or main circuit board. This board is sometimes called the *logic board.* The microprocessor, together with various other circuits, comprises the *central processing unit* (CPU). Auxiliary circuits can be fabricated onto the same

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

chip as the microprocessor, but they are often separate. The external chips contain *memory*, *storage*, and *programming* instructions.

You might think of the microprocessor as the computer's conscious mind, which directs the behavior of the machine by deliberate control. The CPU, controlled by the microprocessor, represents the PC's entire mind, conscious and subconscious. All the ancillary circuits, in conjunction with the CPU, make up the nervous system. Peripherals such as printers, disk drives, mice, speech recognition/synthesis apparatus, modems, and displays are the hands, ears, eyes, and mouth of the system. In advanced computer systems there can also be robots, vision systems, various home appliances, surveillance apparatus, medical devices, and other exotic equipment under the control of the CPU.

Units of Digital Data

You learned about bits and bytes in Chap. 26. Let's examine them again in more detail, with emphasis on their relevance to computers. Recall that one bit (1 b) is a single binary digit, and one byte (1 B) is a unit of digital data consisting of a string of eight contiguous bits (8 b) in most systems. One byte constitutes roughly the same amount of data as one character, such as a letter, numeral, punctuation mark, or space.

Memory and Storage Capacity

Computer memory and storage involves files that are huge in terms of bytes. Therefore, *kilobytes* (units of $2^{10} = 1024$ bytes), *megabytes* (units of $2^{20} = 1{,}048{,}576$ bytes), and *gigabytes* (units of $2^{30} = 1{,}073{,}741{,}824$ bytes) are used. The abbreviations for these units are KB, MB, and GB, respectively. Alternatively you might see them abbreviated as K, M, and G.

As computer technology advances, you'll hear more and more about a unit of data called a *terabyte* (TB or T). This is equivalent to 2^{40} bytes, or 1,048,576 MB. Someday we will commonly use the terms *petabyte* (PB or P), which refers to 2^{50} bytes or 1,048,576 GB, and *exabyte* (EB or E), which refers to 2^{60} bytes or 1,048,576 TB.

Here are all these data units listed as a hierarchy:

1 KB = 1024 B
1 MB = 1024 KB
1 GB = 1024 MB
1 TB = 1024 GB
1 PB = 1024 TB
1 EB = 1024 PB

Computer memory is usually specified in megabytes or gigabytes. The same holds true for removable data storage media. The *hard drive* in a computer generally has capacity measured in gigabytes, although a few get into the terabyte range. Some external storage media, used for data *archiving* (saving it for long-term reference), have capacity measured in terabytes and petabytes.

Data Speed

The speed at which computers send digital data to and from each other is almost always expressed in *bits per second* (bps) and large multiples thereof. Multiples of bits per second involve the same prefixes as multiples for bytes, but when talking or writing about bits per second, these prefixes refer to powers of 10 rather than powers of 2.

A *kilobit per second* (1 kbps) is equal to $10^3 = 1000$ bps, a *megabit per second* (1 Mbps) is equal to $10^6 = 1{,}000{,}000$ bps, and a *gigabit per second* (1 Gbps) is equal to $10^9 = 1{,}000{,}000{,}000$ bps. (You won't often hear of data speed units larger than the gigabit per second.) Hierarchically:

1 kbps = 1000 bps
1 Mbps = 1000 kbps
1 Gbps = 1000 Mbps

Note the lowercase "k" in reference to "kilo-" meaning 10^3 or 1000, and the uppercase "K" in reference to "kilo-" meaning 2^{10} or 1024. Note also that for the other prefix multipliers, uppercase letters are always used. These are not typos! The "k versus K" distinction is a notational peculiarity, and is often ignored or overlooked. You will sometimes see "Kbps" rather than "kbps," or "kB" rather than "KB," in technical papers and other documents. As long as you know whether the author is writing about data speed (bits per second) or memory/storage (bytes), it should not be a problem. But if you want to be rigorous, this peculiarity is worth remembering.

The Hard Drive

A hard drive, also known as a *hard disk*, is a common form of mass storage for computer data. The drive consists of several disks, called *platters*, arranged in a stack. They are made of rigid, durable material that is coated with a ferromagnetic substance similar to that used in audio or video tape. The platters are spaced a fraction of a centimeter apart. Each has two sides (top and bottom) and two *read/write heads* (one for the top and one for the bottom). The assembly is enclosed in a sealed cabinet. Figure 33-2A is an edgewise, cutaway view of the platters and heads in a typical hard drive.

Drive Action

When the computer is switched off, the hard-drive mechanism locks the heads in a position away from the platters. This prevents damage to the heads and platters if the computer is moved. When the computer is powered up, the platters spin at several thousand revolutions per minute (rpm). The heads hover a few millionths of a centimeter above and below the platter surfaces.

When you type a command or click on an icon telling the computer to read or write data, the hard-drive mechanism goes through a series of rapid, complex, and precise movements. The head positions itself over the particular spot on the platter where the data is located or is to be written; then the head detects the magnetic fields and translates them into tiny electric currents. All this takes place in a small fraction of a second.

Data Arrangement and Capacity

The data on a hard drive is arranged in concentric, circular *tracks.* There are hundreds or even thousands of tracks per radial centimeter of the platter surface. Each circular track is broken into a number of arcs called *sectors.* A *cylinder* is the set of equal-radius tracks on all the platters in the drive. Tracks and sectors are set up on the hard drive during the initial *formatting* process. There are also data units called *clusters.* These are units consisting of one to several sectors, depending on the arrangement of data on the platters. Figure 33-2B is a face-on view of a single hard-disk platter, showing a track and one of its constituent sectors.

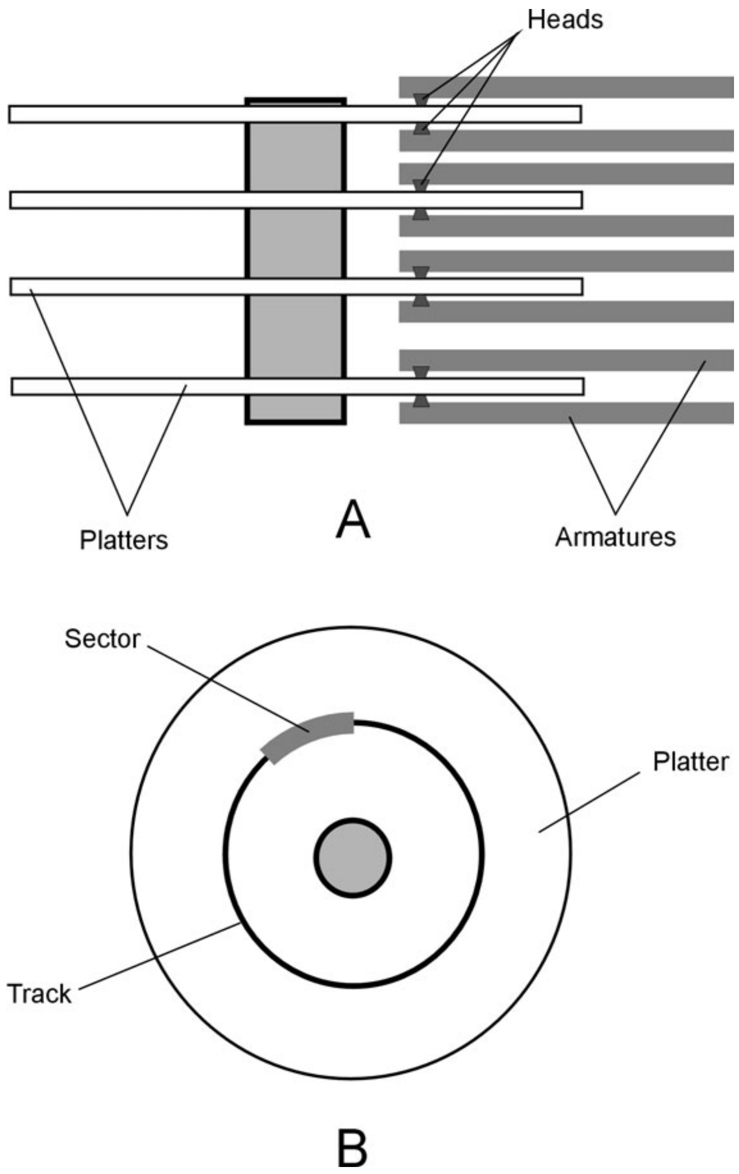

33-2 At A, an edgewise view of a hard drive, showing four platters. At B, a broadside view of a platter, showing one track and one sector.

When you buy a computer, whether it is a desktop, notebook (also called laptop), or portable (also called handheld) unit, it will have a hard drive built in. The drive comes installed and formatted. Most new computers are sold with several commonly used programs installed on the hard drive. Some computer users prefer to buy new computers with only the operating system, by means of which the programs run, installed; this frees up hard-drive space and gives the user control over which programs to install (or not to install).

External Storage

There are several types of *external storage* (besides the hard drive) in which data can be kept in large quantities. Computer experts categorize external storage in two ways: *access time* and *cost per megabyte* (or gigabyte, or terabyte).

Disk Media

Disk media offer the advantage of speed, convenience, and reliability. For personal computers, there are many forms of external disks. Here are three of the most well-known.

An *external hard drive* is exactly what its name suggests. This type of device is exceptionally fast, and has storage capacity similar to the hard drives inside computers. They can be easily connected to any personal computer with a short cable. Most of these devices require a power supply, often found in the form of a "brick" that contains a transformer, rectifier, and filter that converts 120 V ac into the necessary dc for the device.

Compact disk recordable (CD-R) and *compact disk rewritable* (CD-RW) are popular for backing up and archiving computer data. You can buy these disks for various other applications, too, such as storage for digital photos and home videos. They are the same size, physically, as conventional CD-ROMs, which are used for commercial software, databases, and digital publications. The main asset of CD-R and CD-RW is moderately large capacity and long shelf life. Most new computers have built-in drives for these disks.

Diskettes, also called (imprecisely) "floppies," are about 9 cm (3.5 in) in diameter and enclosed in a rigid, square case about 4 mm (0.15 in) thick. Their capacity, individually, is limited. They are all but obsolete. Increasingly, new computers are sold without drives for floppies.

Tape Media

The earliest computers used magnetic tape to store data. This is still done in some systems. You can get a *tape drive* for making an emergency backup of the data on your hard drive, or for archiving data you rarely need to use. Magnetic tape has high storage capacity. There are microcassettes that can hold more than 1 GB of data; standard cassettes can hold many gigabytes. But tapes are extremely slow because, unlike their disk-shaped counterparts, they are a *serial-access* storage medium. This means that the data bits are written in a string, one after another, along the entire length of the tape. The drive might have to mechanically rewind or fast-forward through a football field's length of tape to get to a particular data bit, whereas on a disk medium, the read/write head never has to travel farther than the diameter of the disk to reach a given data bit.

Flash Memory

Flash memory is an all-electronic form of storage that is useful especially in high-level graphics, big-business applications, and scientific work. The capacity is comparable to that of a small hard drive, but there are no moving parts. Because there are no mechanical components, flash memory is faster than any other mass-storage scheme, provided it does not cause a *software conflict* with other programs in the computer. (Software conflicts can cause a computer to slow down or "freeze up.")

Flash memory is available in small modules roughly the size of your index finger, and can be plugged directly into one of the *Universal Serial Bus* (USB) ports provided in all new computers. Some flash memory modules come in the form of *PC cards* (also called *PCMCIA cards*), which are credit-card-sized, removable components.

Memory

In a computer, the term *random access memory* (RAM), often called simply *memory*, refers to ICs that store working data. The amount, and speed, of memory are crucial factors in determining what a computer can and cannot do.

Data Flow

Figure 33-3 shows how data moves in a computer among an internal hard drive, an external hard drive, and a CD-RW storage medium. This process is controlled by the CPU. When you open a file on any medium, the data goes immediately into the memory. The CPU, under direction of the microprocessor, manipulates the data in the memory as you work on the file. Thus, the data in memory changes from moment to moment.

When you hit a key to add a character, or drag the mouse to draw a line that shows up on your display, that character or line goes into memory at the same time. If you hit the backspace key to delete a character, or drag the mouse to erase a line on the screen, it disappears from the memory.

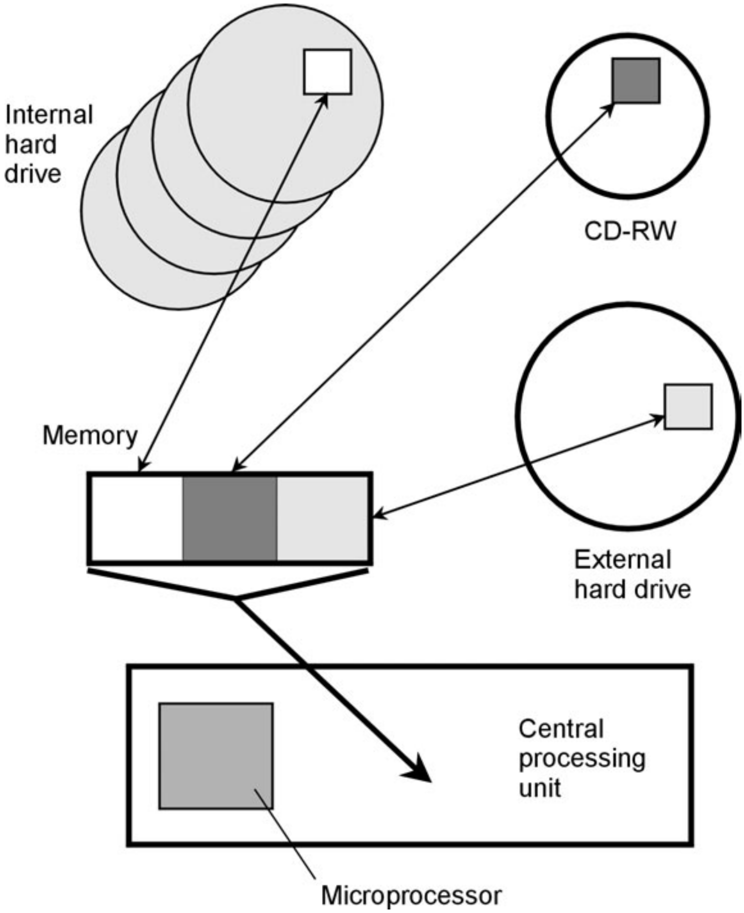

33-3 The flow of data between storage media and memory is controlled by the CPU.

During this time the original file on the storage medium stays as it was before you accessed it. No change is made to that data until you specifically instruct the computer to overwrite it. When you're done working on a file, you tell the microprocessor to close it. Then the data leaves memory and goes back to the medium from which it came, or to some other place, as you direct. If you tell the computer to overwrite the file on the medium from which it came, many programs send the new data (containing the changes you have made) to unused space there; the old data (as it was before you opened the file) stays in its old location. This is a safeguard, in case you decide to undo the changes you made.

All the data passing between the storage media and memory, and between the memory and the CPU, is in *machine language.* This consists of binary digits (bits) 0 and 1. But the data passing between you and the CPU is in plain English (or whatever other language you prefer), or in some high-level programming language, having been translated by the machine into a form you can understand.

Memory Capacity

The maximum number of megabytes or gigabytes of data that can be stored in a computer's memory is known as the *memory capacity.* The main factor that determines memory capacity is the number of transistors that can be fabricated onto a single memory chip. Other factors, such as microprocessor speed, have a practical effect on the usable memory capacity.

A gigantic memory is of little practical value if the microprocessor is slow. Nor is a fast microprocessor worth much if the memory capacity is too small for the applications (programs) you want to run. Most software packages tell you how much memory you need to run the programs they contain. They'll often quote two specifications: a minimum memory requirement and a figure for optimum performance (approximately twice the minimum requirement). If possible, you should equip your computer with enough memory for optimum performance.

When buying a new computer, keep in mind that this year's high-end machine may become next year's ordinary one when it comes to popular software, and in a few years it will no longer be adequate to run many of the programs that will be available. If your computer lacks the memory to run a given application, you can usually add more. But this can only be done up to a certain point. Eventually, your microprocessor will no longer be able to run contemporary software at reasonable speed, no matter how much memory you have.

How much memory do you think your machine will require to run two or three of your favorite applications at the same time? Double or triple it, and you will come close to the amount you are likely to need for the next two or three years, or until you are overcome by the urge to buy a new computer again.

The Display

The visual interface between you and your computer is known as the *display.* In desktop computers, an external display is often called a *monitor.* There are two popular types in use for personal computing today.

A *cathode-ray tube* (CRT) *monitor* resembles a television set without the tuning or volume controls. This type of display is large and heavy, and although some people consider CRTs obsolete, other computer users prefer them, especially the larger ones, which can have a diagonal screen measure of 53 cm (21 in) or more.

A *liquid crystal display* (LCD) is lightweight and thin. This type of display is used in notebook and portable computers. It has become increasingly popular for desktop computers because the technology has improved and has become somewhat less expensive than it used to be. Light weight and a shallow desk profile are decided advantages of this type of display. A good one can provide image crispness and color at least as good as the best CRT display. An LCD also consumes less power.

Other types of displays exist. Of special note is the *plasma display*, which obtains its images from the properties of electric charges passing through rarefied gases. These displays are often found in large department stores, where you can see them hanging from the ceilings, displaying advertisements.

Resolution

The *image resolution*, often called simply the *resolution*, of a computer display is one of its most important specifications. This is the extent to which it can show detail. The better the resolution, the sharper the image.

Resolution can be specified in terms of *dot size* or *dot pitch*. This is the diameter, in millimeters (mm), of the individual elements in the display—the "smallest unbreakable pieces." A good display has a dot pitch that is a small fraction of a millimeter. A typical CRT has a dot pitch of 0.25 or 0.26 mm. The smaller the number, the higher the resolution and, all other factors being equal, the crisper the image in an absolute sense.

Image resolution can be specified in a general sense as a pair of numbers, representing the number of *pixels* (picture elements) the screen shows horizontally and vertically. For a particular screen size, the greater the number of pixels the unit can display, the crisper the image. In personal computers, typical displays have 800 × 600 resolution (800 pixels wide by 600 pixels high) or 1024 × 768 resolution. A few can work up to 1600 × 1200 or even higher.

Screen Size

Screen sizes are given in terms of diagonal measure; a popular size is 43 cm (17 in). This will work quite well at 800 × 600 resolution. For higher resolution, a larger screen is preferable, such as 48 cm (19 in) or 53 cm (21 in). A high-end display is crucial for doing graphics work, when doing serious research on the Internet, in remote-control robotics, and in computer *gaming*. Besides these practical advantages, a sharp display is more pleasant to work with than a marginal one.

Along with memory and hard-drive capacity, the display is one of the most important parts of a computer from a user-friendliness standpoint. On the job, long hours at a computer can get tedious even if the machine is perfect. An inadequate display can give rise to eye strain and headaches, and can also degrade the quality and accuracy of work done by people using the computer.

Interlacing and Refresh Rate

Another important consideration in the choice of a display is whether or not it uses *interlacing*. Interlacing increases the obtainable resolution, but it also results in a lower *refresh rate* (number of times the entire image is renewed). A low refresh rate can cause noticeable flickering in the image, and can be especially disruptive in applications where rapid motion must be displayed, such as high-end computer gaming.

A good refresh rate specification is 70 Hz or more. For applications not involving much motion, 60 or 66 Hz is adequate for some people, but others complain of eye fatigue because they can vaguely sense the flicker. Most people find 56-Hz refresh rates too slow for comfort in any application.

ELF Fields

Extremely low frequency (ELF) fields are electromagnetic (EM) fields that fluctuate or alternate more slowly than conventional radio waves. Such fields are produced by various consumer electronic devices and appliances. An ELF field is nothing like X rays or gamma rays, which cause radiation sickness. Nor is ELF radiation like ultraviolet (UV), which can cause skin cancer over long periods. An ELF field cannot make anything radioactive. In computer systems, the ELF that you've heard about is emitted mainly by electromagnetic CRT monitors. Other parts of a computer system are not responsible for much ELF energy. The LCDs used in many systems today produce essentially none.

As you learned in Chap. 29, the characters and images in a CRT are created as electron beams strike a phosphor coating on the inside of the glass. The electrons constantly change direction as they sweep from left to right, and from top to bottom, on your screen. The sweeping is caused by deflecting coils that steer the beam across the screen. The coils generate magnetic fields that interact with the negatively charged electrons, forcing them to change direction. Because of the positions of the coils, and the shapes of the fields surrounding them, there is more magnetic energy radiated from the sides of an electromagnetic CRT than from the front. If there's any health hazard with ELF, therefore, it is greater for someone sitting off to the side of an electromagnetic CRT monitor, and less for someone watching the screen from directly in front at the same distance.

If you're concerned about the ELF fields produced by CRT monitors, you might consider buying an electrostatic CRT unit that has been designed to minimize ELF fields, or buying a standalone LCD display panel. It's a good idea to arrange your workstation so your eyes are at least 0.5 m (about 18 in) away from the screen of a CRT monitor, and if you are working near other computer users, workstations should be at least 1 m (3 ft) apart.

The Printer

There are several types of *printer* in common use in personal and business computing applications today. The two most common are the *inkjet printer* and the *laser printer.* Less often used are the *thermal printer* and the *dot matrix printer.*

Inkjet Printers

In an inkjet printer, tiny nozzles spray ink onto the paper. Most of these printers are comparatively slow to produce an image, and the ink needs time to dry even after the image comes out, but the quality can be excellent. Inkjet printers are available in single-color and multicolor designs. The best color machines produce images of photo quality. In fact, some are designed especially to print digital photos.

Inkjet printers require periodic replacement of the ink cartridges. These can be quite expensive, and for this reason, inkjet printers are not well suited for high-volume printing. Inkjet printers require paper with low fiber content. This keeps the ink from being carried along by capillary action before it dries, muddling or blurring the printout. Look specifically for "inkjet paper" in computer supply stores and department stores.

Laser Printers

A laser printer works like a photocopy machine. The main difference is that, while a photocopier creates a copy of a real image (the paper original), a laser printer makes a copy of a digital computer image.

When data arrives at the printer from the computer, the encoded image is stored in the printer memory. The memory then sends it along to the laser and other devices. The laser blinks rapidly while it scans a cylindrical *drum.* The drum has special properties that cause it to attract the printing chemical, called *toner*, in some places but not others, creating an image pattern that will ultimately appear on the paper. A sheet of paper is pulled past the drum and also past an *electrostatic charger.* Toner from the drum is attracted to the paper. The image thus goes onto the paper, although it has not yet been permanently fused, or bonded, to the paper. The *fuser*, a hot pair of roller/squeezers, does this job, completing the printing process.

The main asset of laser printers is their ability to produce a large number of copies at high speed. They generally have excellent print and graphics quality for grayscale. Color laser printers are available as well, but they can be quite expensive.

The image resolution of a laser printer ranges from about 300 dots per inch (dpi) for older units to 1200 dpi, 2400 dpi, or even higher resolutions in state-of-the-art machines. As far as the untrained eye can tell, 600 dpi is as good as a photograph. Laser printers can handle graphics and text equally well. If an image can be rendered on a photocopy machine, it can be rendered just as well on a laser printer.

Thermal Printers

A thermal printer uses temperature-sensitive dye and/or paper to create hard-copy text and images. Some thermal printers produce only grayscale images, while others can render full color. Thermal printers are often preferred by traveling executives who use portable computers, because these printers are physically small and light.

A simple grayscale thermal printer employs special paper that darkens when it gets hot. A color thermal printer uses thick, heat-sensitive dyes of the primary pigments: magenta (pinkish red), yellow, and cyan (bluish green). Sometimes black dye is also used, although it can be obtained by combining large, equal amounts of the primary pigments. The print head uses heat to liquefy the dye, so it bleeds onto the paper. This is done for each pigment separately.

Some, if not most, thermal printouts fade after awhile. Have you ever pulled out an old store receipt and found that it was washed-out or blank? Thermal printers can be convenient in a pinch, but you should be aware that some of them have this problem. If you're keeping a receipt for tax purposes or for proof-of-purchase and it has been printed on thermal paper, make a photocopy or a digital scan of the receipt right away. You can recognize the output of a thermal printer because the paper curls up when it's fresh out of the machine.

Dot Matrix Printers

The dot matrix printer is the horse and buggy of the printing family. This type of printer is the least expensive, in terms of both the purchase price and the long-term operating cost. Dot matrix printers produce fair text quality for most manuscripts, reports, term papers, and theses. The mechanical parts are rugged, and maintenance requirements are minimal. Dot matrix printers can render some simple graphic images, but the quality is fair at best, and it can take a long time to print a single image. Dot matrix printers cannot reproduce detailed artwork or photographs with acceptable quality.

The Scanner

A *scanner* is an electromechanical device that converts hard-copy text and graphics into digital form for processing and storage in a computer.

Basic Features

Scanners can save untold hours of tedious manual retyping. Suppose you wrote a book a long time ago, and have lost the digital files (or used an old-fashioned typewriter, and never had any digital files!). The hard copy of that book sits in your basement, awaiting the massive editing that can turn it into a great novel. It needs the power of your computer's word processor. But you can't deal with the prospect of retyping its 1000 pages. A scanner, equipped with *optical character recognition* (OCR), can do away with most of the hard labor involved in getting a hard-copy manuscript onto disk.

A good scanner can be had for a couple of hundred dollars. But the value of optical scanning is hard to measure. For many entrepreneurs, it can make the difference between staying afloat and going bankrupt. It can do the work of one or two full-time typists for a tiny fraction of the long-term cost. Big companies can save money, too. Lawyers and doctors find scanners invaluable for backing up files of all kinds. Aside from storing text, scanners can make digital copies of vital papers, records, and receipts, which can be easily backed up on CD-R or CD-RW media.

A typical scanner can render color images, text, photographs, and everything else needed to make a complete, accurate digital record of any document. Color scanners use three different light beams (red, blue, and green) to get three different images, which are processed and combined in much the same way as a color television camera works. The image resolution of a scanner is measured in dots per inch (dpi), just as is done with printers. The higher the dpi specification, the more detail the scanner can see.

For reliable scanning of text and most images, a resolution of at least 300 dpi is recommended. Virtually all scanners meet this requirement. For images, greater detail translates into more memory consumed. Color increases the amount of memory or storage that an image takes up, if the image resolution remains constant.

Configurations

Scanners come in three basic configurations. The scanner that's best for you will depend on what you want to do with it, and on how much money you're willing to spend for it.

The cheapest type of scanner is a *handheld scanner.* It looks something like a miniature vacuum-cleaner head, or one of the bar-code readers in retail stores. You roll the unit over the paper containing the text and/or graphics you want to scan. Because the unit is not as wide as most pages, you'll have to make two or three passes over the page. Handheld scanners are preferred by people who scan small images, such as snapshots. They are light in weight, and need almost no desk space. One potential problem is that you might try to scan too fast. Some handheld scanners have speed indicators that tell you if you're going too fast. Another potential difficulty is not getting a straight-line scan. Most handheld units have built-in guides (like miniature rolling pins) that minimize this problem.

If you want to scan a book or magazine, a *flatbed scanner* is much easier to use than a handheld scanner. The unit looks something like a photocopier. Using a flatbed scanner is similar to working a small photocopy machine. You lay the page, photo, or sheet down on a clear glass, and the scanning head moves past it, picking up the image. Flatbed scanners consume desk space, which, if you have a couple of printers and a fax machine, might already be at a premium.

A *sheet scanner*, also called a *feedthrough scanner*, resembles a fax machine (and in fact, many of these units can do double duty as fax machines). As its name implies, this type of scanner pulls sheets of paper through, one by one. You can stack several pages, one on top of the other, and the machine will automatically feed and scan them. However, you can't scan bound books or magazines as you can with a flatbed scanner—unless you're willing to rip out individual pages.

Precautions

Even the best scanners make some mistakes when used with OCR. This is especially true if text contains nonstandard symbols. Highly technical material presents the worst problems. Some mathematical symbols are so esoteric that the average person (let alone a machine) is befuddled by them. Ink spots, stray markings, and smudges on a page can cause scanning errors, in much the same way as background noise confuses a speech recognition system.

A human reader can often tell what a printed letter should be, even when it is severely mutilated. But computers lack human intuition. To some extent this can be corrected by a built-in *spell checker.* Some OCR programs have spell checking, but this introduces its own set of problems because it, too, is imperfect.

If a scanner doesn't recognize a character, it will usually print a "tag"—a blank space, underline, or default symbol such as @ or #. Scanned text must always be carefully proofread, and corrections made with word-processing software, after the data has been stored on the hard drive.

The Modem

The term *modem* is a contraction of *modulator/demodulator.* A modem interfaces a computer to a telephone line, *digital subscriber line* (DSL), cable system, fiber-optic network, wireless network, or radio transceiver, allowing you to communicate with other computer users and to "surf the Internet."

Data Speed

Modems work at various speeds, usually measured in kilobits per second (kbps) or megabits per second (Mbps). Sometimes you'll hear about speed units called the *baud* and *kilobaud.* (A kilobaud is 1000 baud.) Baud and bits per second are almost the same units, but they are not identical. People sometimes use the term *baud* when they really mean *bps.*

Modem speeds, particularly in cable, fiber-optic, and wireless networks, keep increasing as computer communications technology advances. Modems are rated according to the highest data speed they can handle. A typical *telephone modem* works at about 56 kbps. Modems for more advanced connections operate much faster.

Slow Modems

A computer works with binary *digital* signals, which are rapidly fluctuating direct currents. For digital data to be conveyed over a telephone or radio circuit, the data must be converted to *analog* form. In a telephone modem or radio-transceiver modem, this is done by changing the digit 1 into an audio tone, and the digit 0 into another tone with a different pitch. The result is an extremely fast back-and-forth alternation between the two tones.

In *modulation*, digital data from the computer is changed into analog data for transmission over the telephone line or radio medium. The modulator is therefore a digital-to-analog converter (D/A converter or DAC). *Demodulation* changes the analog signals from the telephone line or radio medium back to digital signals that a computer can understand. The demodulator is thus an analog-to-digital converter (A/D converter or ADC). The highest practical speed for this type of modem is approximately 56,000 bits per second (bps), or 56 kilobits per second (kbps).

Amateur radio operators use a variety of digital modes at considerably slower speeds than 56 kbps for Internet-like communications. The principal advantage of amateur radio lies in the fact that it can work when all else fails. Such communications are not fast, but no infrastructure is needed. This makes it possible for amateur radio operators to provide emergency communications into and out of areas stricken by natural disasters that destroy the conventional communications infrastructure.

Internal versus External Modems

An *internal modem* is a printed-circuit board, also called a *card.* Virtually all computers, both desktop and notebook, are sold with internal modems that can interface with the telephone line at speeds of up to 56 kbps.

An *external modem* is a self-contained unit. It has a cord that runs to either the computer's *serial data port* (also called the *communications port*) or one of the USB ports, and another cord that runs to the telephone line, cable system, or satellite dish. If you want to use a DSL, cable, or satellite Internet connection, you will usually have to buy or rent an external modem from the service provider.

The Internet

The *Internet* is a worldwide system, or *network*, of computers. It got started in the late 1960s, originally conceived as a network that could survive nuclear war. Back then it was called *ARPAnet*, named after the Advanced Research Project Agency (ARPA) of the United States federal government.

Protocol and Packets

When people began to connect their computers into ARPAnet, the need became clear for a universal set of standards, called a *protocol*, to ensure that all the machines "speak the same language." The modern Internet is such that you can use any computer to take advantage of all the network's resources.

Internet activity consists of computers "talking" to one another. This occurs in machine language. However, the situation is vastly more complicated than when data goes from one place to another within a single computer. In the Internet (often called simply *the Net*), data must often go through several different computers to get from the transmitting or *source* computer to the receiving or *destination* computer. These intermediate computers are called *nodes*, *servers*, *hosts*, or *Internet service providers* (ISPs).

Millions of people are simultaneously using the Net. The most efficient route between a given source and destination can change from moment to moment. The Net is set up in such a way that signals always try to follow the most efficient route. If you are connected to a distant computer, say a machine at the National Hurricane Center, the requests you make of it, and the data it sends back to you, are broken into small units called *packets.* Each packet coming to you has, in effect, your computer's name written on it. But not all packets necessarily travel the same route through the network. Ultimately, all the packets are reassembled into the data you want (such as the infrared satellite image of a hurricane), even though they do not arrive in the same order they were sent.

Figure 33-4 is a simplified drawing of Internet data transfer for a hypothetical file containing five packets transferred during a period of *heavy Net traffic* (meaning that a lot of people are using the Internet at that particular time, so it is operating at near capacity). Nodes are shown as black dots surrounded by circles. In this example, some packets pass through more nodes, and/or over a much greater physical distance, than others. If Net traffic were light, all the packets might follow the same route, or pass through fewer nodes. This is why it takes longer to acquire data on the Net during peak hours of use, as compared with times when there are comparatively few people connected

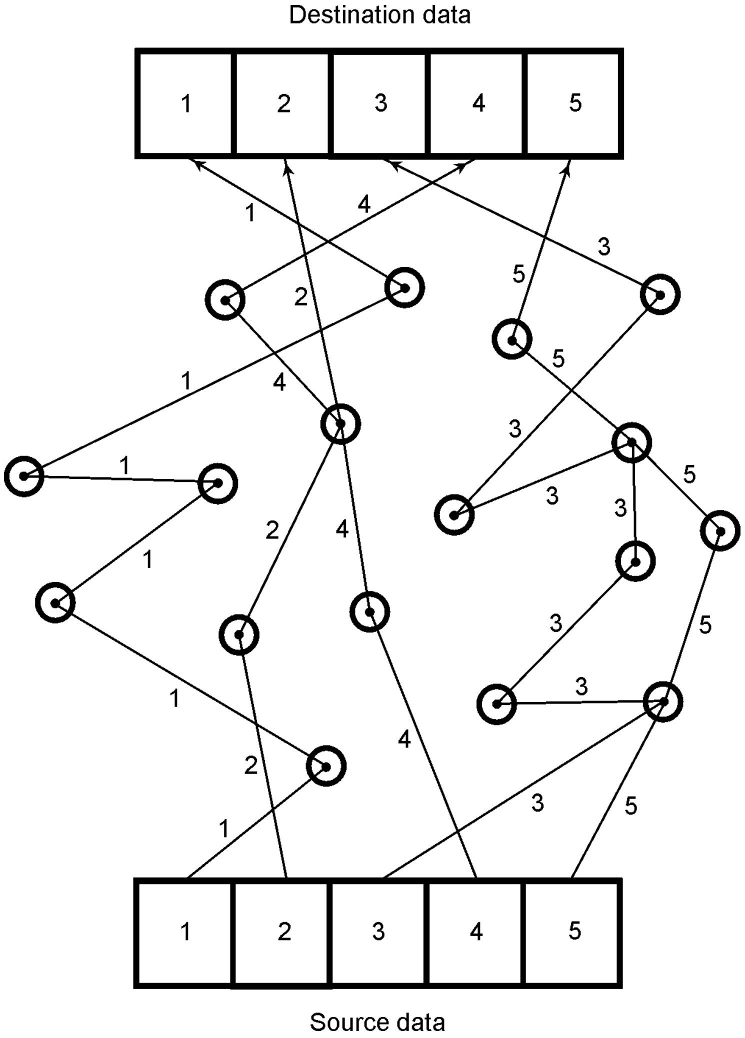

33-4 Internet data flows in packets from the source to the destination.

into it, no matter how fast your connection happens to be. A file cannot be completely reconstructed until all the packets have arrived and the destination computer has ensured that there are no errors.

E-mail and Newsgroups

For many computer users, communication by means of *electronic mail* (*e-mail*) and/or *newsgroups* has practically replaced the postal service. You can leave messages for, and receive them from, friends and relatives scattered throughout the world.

To effectively use e-mail or newsgroups, everyone must have an *Internet address.* These tend to be arcane. An example is sciencewriter@tahoe.com. The first part of the address, before the @ symbol, is the *username.* The word after the @ sign and before the period (or dot) represents the *domain* *name.* The three-letter abbreviation after the dot is the *domain type.* In this case, “com” stands for “commercial.” Tahoe.com is thus a commercial provider. Other common domain types include “net” (network), “org” (organization), “edu” (educational institution), and “gov” (government). In recent years, country abbreviations have been increasingly used at the ends of Internet addresses, such as “us” for United States, “de” for Germany, “uk” for United Kingdom, and “jp” for Japan. Other abbreviations are constantly coming into common usage, such as “info” for “informational site” and “biz” for “business.”

Internet Conversations

You can carry on a “teletype-style” (real-time text) conversation with other computer users using the Internet. When done among users within a single service provider, this is called *chat.* When done among people connected to different service providers, it is called *Internet relay chat* (IRC). Typing messages to, and reading them from, other people in real time is more personal than letter writing, because your addresses get their messages immediately. But it’s less personal than talking on the telephone, because you cannot hear, or make, vocal inflections.

It is possible to digitize voice signals and transfer them via the Internet. This has given rise to hardware and software schemes that claim to provide virtually toll-free long-distance telephone communications. As of this writing, this is similar to cellular telephone communications in terms of reliability and quality of connection. When Net traffic is light, such connections can be good. But when Net traffic is heavy, the quality is marginal. Audio signals, like any other form of Internet data, are broken into packets. All, or nearly all, the packets must be received and reassembled before a good signal can be heard. This takes variable time, depending on the route each packet takes through the Net. If many of the packets arrive disproportionately late, the destination computer can only do its best to reassemble the signal. In the worst case, the signal does not get through at all.

Getting Information

One of the most important features of the Internet is the fact that it can connect you with millions upon millions of sources of information (and misinformation). Data is transferred among computers by means of a protocol that allows the files on the hard drives of distant computers to become available exactly as if the data were stored on your own computer’s hard drive, except the access time is slower. You can also store files on distant computers’ hard drives. When using the Internet for obtaining information, you should be aware of the time at the remote location, and avoid, if possible, accessing files during the peak hours at the remote computer. Peak hours usually correspond to working hours, or approximately 8:00 a.m. to 5:00 p.m. local time, Monday through Friday. You must take time differences into account if you’re not in the same time zone as the remote computer.

The *World Wide Web* (also called *WWW* or *the Web*) is one of the most powerful information servers you will find on the Internet, and in common usage, the terms *Internet* and *Web* have become almost synonymous. The outstanding feature of the Web is *hypertext*, a user-friendly scheme for cross-referencing of documents. In fact, the names of Web sites generally begin with the four letters “http,” which stands for *hypertext transfer protocol.* Certain words, phrases, and images make up *links.* When you select a link in a *Web page* or *Web site* (a document containing text, graphics, and often other types of files), your computer is transferred to another site dealing with the same or a related subject. This site will usually also contain numerous links. Before long, you might find yourself surfing the Web for hours, going from site to site. The word *surfing* derives from the similarity of this activity to television “channel surfing.”

Quiz

Refer to the text in this chapter if necessary. A good score is 18 correct. Answers are in the back of the book.

1. One megabyte is the same amount of data as
   (a) 1024 bytes.
   (b) 1024 KB.
   (c) 1024 GB.
   (d) 1/1024 KB.
2. The Web should be expected to work fastest for a user in New York City at
   (a) 2:00 a.m. local time on a Tuesday.
   (b) 4:00 p.m. local time on a Wednesday.
   (c) 12:30 p.m. local time on a Thursday.
   (d) any time; it doesn't matter.
3. The sharpness of the image on a computer display can be specified in terms of
   (a) refresh rate.
   (b) interlace rate.
   (c) wavelength.
   (d) dot pitch.
4. The term *cylinder* refers to
   (a) a set of tracks in a hard drive.
   (b) a particular type of memory chip.
   (c) a drum-shaped data storage medium.
   (d) the spindle that turns a CD-R or CD-RW.
5. An example of a mass-storage device is a
   (a) hard drive.
   (b) microprocessor.
   (c) modem.
   (d) read-write head.
6. The character string *stangibilisco@rushmore.com* would most likely represent
   (a) a Web site.
   (b) the location of data in a computer's memory.
   (c) an e-mail address.
   (d) a computer's serial number.
7. The megabit per second (Mbps) is a common unit that expresses
   (a) the memory capacity of a chip.
   (b) the storage capacity of an external hard drive.

(c) the image resolution of a printer.
(d) the speed of an Internet connection.

8. A platter is
   (a) part of a hard drive.
   (b) a unit of memory.
   (c) an element of a digital image.
   (d) a semiconductor chip.
9. Protocol ensures that
   (a) a hard drive runs smoothly.
   (b) a display or printer reproduces color accurately.
   (c) a display or printer generates a clear image.
   (d) computers can exchange data.
10. A packet is
    (a) a computer memory module.
    (b) a unit of 2^{10} bytes.
    (c) a piece of data sent over the Net.
    (d) a picture element in a display.
11. The main microprocessor in a computer is located
    (a) on the motherboard.
    (b) in the external hard drive.
    (c) in the memory chip.
    (d) in the power supply.
12. Cross-referencing among Web pages is done with
    (a) digital signal processing.
    (b) an analog-to-digital converter.
    (c) Internet relay chat.
    (d) hypertext links.
13. Web page addresses usually begin with the letters "http," which signifies a form of
    (a) digital signal processing.
    (b) protocol.
    (c) modem configuration.
    (d) color rendition.
14. A telephone modem contains
    (a) an internal CR-R or CD-RW drive.
    (b) a microprocessor.
    (c) an A/D converter.
    (d) an image resolver.

15. An advantage of a flatbed scanner over a feedthrough scanner is the fact that
    (a) a flatbed scanner can be combined with a fax machine, but a feedthrough scanner cannot.
    (b) a flatbed scanner can be used to scan magazines or books intact, but a feedthrough scanner cannot.
    (c) a flatbed scanner can reproduce color drawings and photographs, but a feedthrough scanner cannot.
    (d) a flatbed scanner takes up less desktop space than a feedthrough scanner.

16. Which of the following types of external storage provide the fastest access time, provided no software conflicts occur?
    (a) A magnetic tape drive
    (b) A CD-R or CD-RW drive
    (c) A flash memory module
    (d) An external hard drive

17. Which of the following is a serial-access medium?
    (a) Magnetic tape
    (b) A diskette
    (c) A hard drive
    (d) A CD-R or CD-RW

18. Which of the following character strings represents the proper format for an e-mail address?
    (a) http://www.sciencewriter.net
    (b) www.mcgraw-hill.com
    (c) blackhills.com
    (d) None of the above

19. Which of the following devices is best suited for animated graphics work involving fast motion, such as high-end gaming?
    (a) A laser printer
    (b) An external hard drive or flash memory module
    (c) Hypertext transfer protocol
    (d) A display that does not use interlacing

20. A thermal printer might be an ideal choice for
    (a) a salesperson who is on the road.
    (b) someone working with animated graphics.
    (c) an author who needs to print a huge text document.
    (d) a photographer who needs top-quality color printouts.

34
CHAPTER

Monitoring, Robotics, and Artificial Intelligence

ELECTRONIC AND COMPUTER SYSTEMS FIND APPLICATIONS IN MANY SCIENTIFIC AND TECHNOLOGICAL fields. Increasingly, devices and systems formerly known only to geeks and techies are becoming commonplace in the consumer and hobby market.

Keeping Watch

Most *monitoring systems* consist of radio transmitters and receivers. Some systems employ lasers at IR or visible wavelengths. A few employ wire or cable links. Some have mechanical hardware, such as robots, that are controlled by the signals reaching the receiver. Here are a few specific examples of monitoring systems.

Baby Monitor

A short-range AM or FM radio transmitter and receiver can be used to listen at a distance to the sounds in an infant's room. The transmitter contains a sensitive microphone, a whip antenna, and a power supply. The receiver is battery-powered and portable. It has a short antenna, similar to the antennas on cordless telephone sets. The receiver can pick up signals from the transmitter at distances of up to about 50 m (165 ft). The signals pass easily through the walls, ceilings, and floors in frame houses.

A so-called baby monitor is subject to interference from other units that might be operating nearby on the same channel. Some baby monitors have multiple, selectable channels to help combat this problem. If interference occurs, the channel can be changed. Communications privacy and security are not a concern.

Smoke Detector

Smoke and fire change the characteristics of the atmosphere. Smoke consists of solid particles, and fire burns away oxygen and produces other gases such as carbon dioxide, carbon monoxide, and sulfur dioxide. These changes can be sensed, and alarms set off if the changes exceed certain limits.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

A *smoke detector* senses changes in two characteristics of the air: the *dielectric constant* (an expression of the extent to which pollutants increase the capacitance of the air) and the *ionization potential* (an expression of the extent to which pollutants change the voltage necessary to produce a spark that jumps through the air). Smoke and fire almost always affect one or both of these parameters.

Figure 34-1 is a functional diagram of a smoke detector that operates by sensing changes in the dielectric constant of the air. Two electrically charged plates are positioned a small distance apart. The plates, and the air between them, form a capacitor. A source of dc is connected to the plates. Normally, the plates retain a constant charge, and the current in the circuit is zero. If the dielectric constant of the air increases, the capacitance changes, causing a small, momentary electric current to flow. This current can be detected, and the resulting signal sets off an alarm. The signal can also actuate a robotic system, such as a group of water sprinklers.

Quality Control

Lasers are useful in industrial monitoring and control applications. An example is the *quality control* (QC) checking of bottles for height as they move along an assembly line. A laser/robot combination can find and remove bottles that are not of the correct height. The principle is shown in Fig. 34-2. If a bottle is too short, both laser beams reach the photodetectors. If a bottle is too tall, neither laser beam reaches the photodetectors. In either of these situations, a *robot arm*, equipped with a *gripper*, picks the faulty bottle off the line and discards it. Only when a bottle is within a narrow range of heights (the acceptable range) does the top laser reach its photodetector while the bottom laser is blocked. Then the bottle is allowed to pass.

Of course, proper operation of the QC machine shown in Fig. 34-2 depends on the reliability of the lasers and photodetectors. If, for example, the lower laser burns out and there is no way for the system operator to know about it, the machine will pass all bottles, whether or not they are of the correct height.

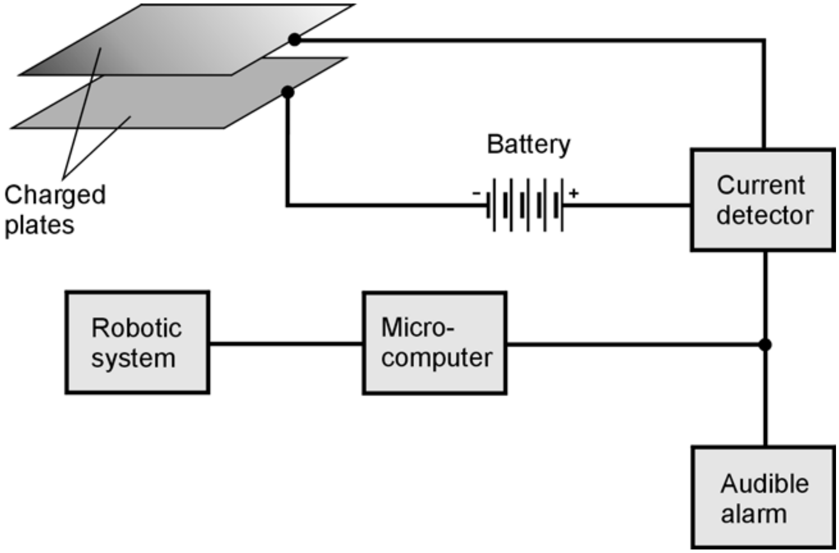

34-1 Simplified functional diagram of a smoke detector.

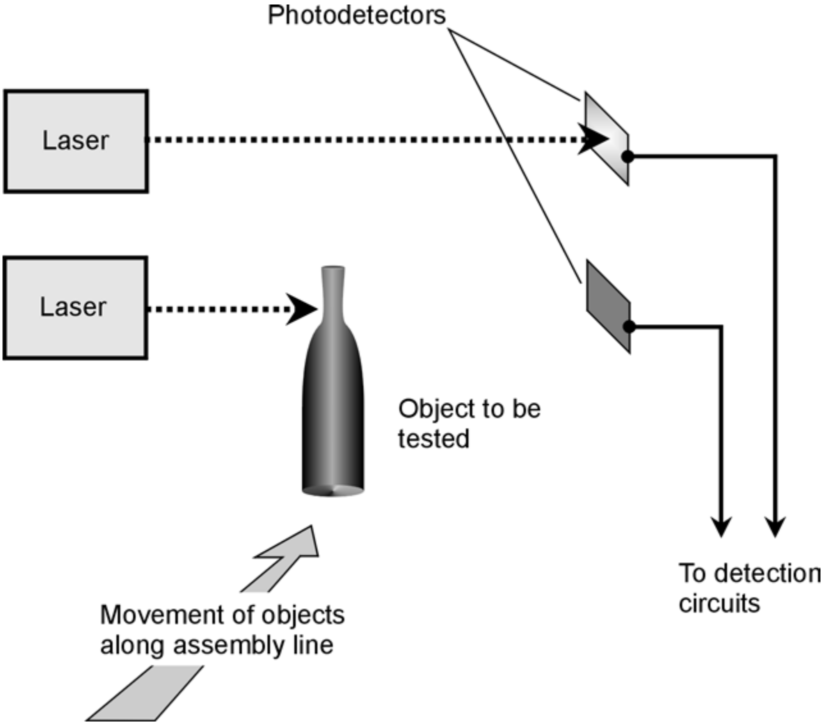

34-2 A method of quality control (QC) for manufactured items.

Tracking People

Suppose that a person is sentenced to house arrest. Compliance can be monitored by having the person carry a conventional beeper (pager). A police or probation officer can page the person at random times; the person must then call the officer within a couple of minutes. The call can be traced, and the location of the telephone verified. This is a simple method of electronically tracking the whereabouts of a person.

A more secure method of ascertaining that a person is at a certain place, at a certain time, is by means of a short-range radio transmitter and receiver. The person wears the transmitting unit. Tamper-proof receiving units are placed at the convict's home, in the car, and at the place of work. The transmitter range is similar to that of a baby monitor. Receiver signals are sent to a central monitoring point. The signals are encoded so the monitoring personnel (or computers) know whether the person is at home, in the car, or at work. Any deviation from normal patterns can be detected.

Radiolocation provides another way to keep track of people. A *transponder* can be carried or worn by the person to be tracked; continuous signals can be sent to the unit asking for a position fix, and the unit can respond through a wireless network such as the cellular telephone system.

Electronic Bug

An *electronic bug* consists of a tiny radio transmitter that can be hidden in a room, placed in a shirt pocket, or planted in a car. The antenna is a length of thin, almost invisible wire. A receiver can be located nearby. The device operates at a low RF power level (on the order of a few milliwatts) to conserve battery energy. If the transmitter is near a wireless repeater that connects into a larger system, eavesdropping can be done anywhere within the coverage of the wireless system. With the advent of low-earth-orbit (LEO) satellite systems, it is theoretically possible to bug a room on the other side of the world.

Wireless electronic bugs can be detected by means of a device called an *RF field-strength meter.* This instrument consists of a microammeter connected to a short whip antenna through a semiconductor diode. The diode rectifies the RF signal, producing dc that shows up as an indication on the meter. When the meter is close to the bug, the current rises. When the bug is within a few centimeters of the meter, the needle may go to full-scale.

The presence of RF fields in your house does not necessarily mean you are being bugged. Many appliances and electronic devices produce RF fields, including computers, radio-operated remote-control units, cell phones, cordless phone sets, and certain medical devices. Even radio receivers and TV sets emit some RF energy.

Electric Eye

The simplest device for detecting an unwanted visitor is an *electric eye.* Narrow beams of IR or visible light are shone across all reasonable points of entry, such as doorways and window openings. A photodetector receives energy from each beam. If, for any reason, the photodetector stops receiving its assigned beam, an alarm is actuated. A person breaking into a property cannot avoid interrupting at least one beam if every possible point of entry has a large enough number of electric eyes spaced at suitable intervals.

IR Motion Detector

A common intrusion alarm device employs an *IR motion detector.* Two or three wide-angle IR pulses are transmitted at regular intervals; these pulses cover most of the room in which the device is installed. A receiving transducer picks up the returned IR energy, normally reflected from the walls, floor, ceiling, and furniture. The intensity of the received pulses is noted by a microprocessor. If anything in the room changes position, there is a change in the intensity of the received energy. The microprocessor detects this change and triggers an alarm (Fig. 34-3). This type of device consumes very little power in regular operation, so batteries can serve as the power source.

Radiant Heat Detector

Certain devices can detect changes in the indoor environment by directly sensing the IR energy (often called *radiant heat*) emanating from objects. Humans, and all warm-blooded animals, emit IR. So does fire. A simple IR sensor, in conjunction with a microprocessor, can detect rapid or large increases in the amount of radiant heat in a room. The time threshold can be set so that gradual or small changes will not trigger the alarm, while significant changes, such as are caused by a person entering the room, will trigger it. The temperature-change threshold can be set so that a small animal will not actuate the alarm, while a full-grown person will. This type of device, like the IR motion detector, can operate from batteries.

The main limitation of radiant-heat detectors is the fact that they can be fooled. False alarms are a risk. The sun, coming out on an overcast day, might suddenly shine directly on the sensor and trigger the alarm. It is also possible that a person clad in a winter jacket, thermal pants, insulated boots, hood, and face mask, entering from a cold outdoor environment, will fail to set off the alarm. For this reason, radiant-heat sensors are used more often as fire-alarm actuators than as intrusion detectors.

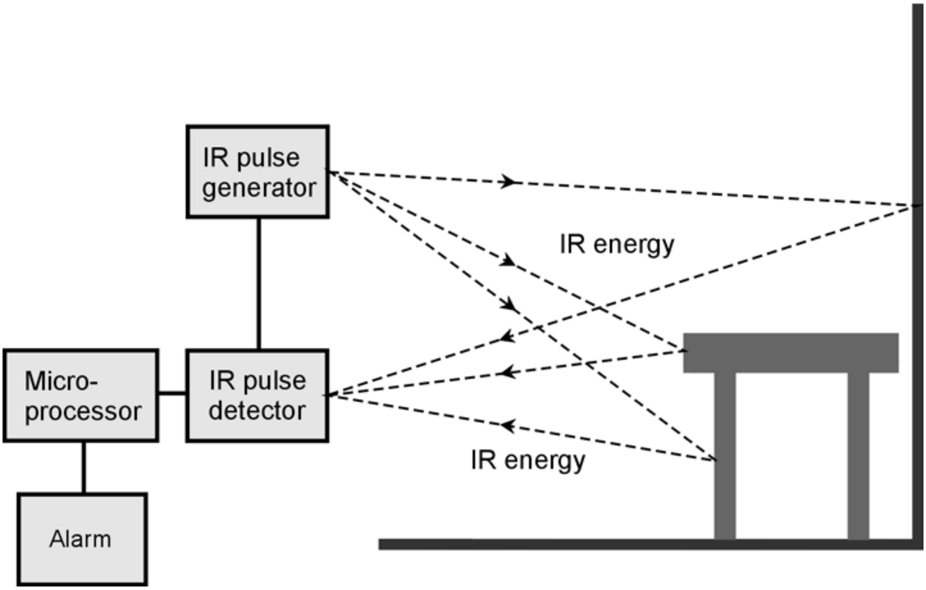

34-3 An IR motion detector.

Ultrasonic Motion Detector

Motion in a room can be detected by sensing the changes in the relative phase of acoustic waves. An *ultrasonic motion detector* employs a set of transducers that emit acoustic waves at frequencies above the range of human hearing (higher than 20 kHz). Another set of transducers picks up the reflected acoustic waves, whose wavelength is on the order of a few millimeters. If anything in the room changes position, the relative phase of the waves, as received by the various acoustic pickups, will change. This data is sent to a microprocessor, which can trigger an alarm and/or notify the police.

Robot Generations and Laws

Some researchers have analyzed the evolution of robots, marking progress according to so-called robot generations. One of the first engineers to make formal mention of robot generations was the Japanese engineer *Eiji Nakano.*

First Generation

According to Nakano, a *first-generation robot* is a simple mechanical arm. Such machines have the ability to make precise motions at high speed, many times, for a long time. They have found widespread industrial application and have been in existence since the middle of the twentieth century. These are the fast-moving systems that install rivets and screws in assembly lines, that solder connections on printed circuits, and that, in general, have taken over tedious, mind-numbing chores that would otherwise have to be done by humans.

First-generation robots can work in groups if their actions are synchronized. The operation of these machines must be constantly watched, because if they get out of alignment and are allowed to keep operating anyway, the result can be a series of bad production units.

Second Generation

A *second-generation robot* has some level of *artificial intelligence* (AI), also called *machine intelligence.* Such a machine is equipped with various sensors that keep it informed about goings-on in the *work environment.* A computer called a *robot controller* processes the data from the sensors and adjusts the operation of the robot accordingly. The earliest second-generation robots came into common use around 1980.

Second-generation robots can stay synchronized with each other, without having to be overseen constantly by a human operator. Periodic checking is needed, however, because things can always go wrong. In fact, as a system becomes more complex, the number of ways in which it can malfunction increases. This is why human beings will never find themselves out of work because of robots. Someone has to make sure the robots keep working properly!

Third Generation

Nakano gave mention to *third-generation robots*, but in the years since the publication of his original paper, some things have changed. Two major avenues are developing for advanced robot technology. These are the *autonomous robot* and the *insect robot.* An autonomous robot is a single machine that works on its own. It contains a controller and can do things largely without supervision, either by an outside computer or by a human being. A good example of this type of third-generation robot is the *personal robot* about which technophiles dream. An insect robot is one of a set of several (or many) identical units that act together to perform a specific task.

Fourth Generation and Beyond

Nakano did not write about anything past the third generation of robots. But we might mention a *fourth-generation robot:* a machine of a sort yet to be deployed. An example is a fleet or population of robots that can reproduce themselves, and perhaps even a system that can evolve to meet changing conditions in its work environment. Past that, we might say that a *fifth-generation robot* is something humans haven't even imagined yet.

Asimov's Three Laws

In one of his early science-fiction stories, the famous author Isaac Asimov first mentioned the word *robotics*, along with three fundamental rules that, in his opinion, all robots ought to obey. These rules were first coined in the 1940s, but *Asimov's three laws of robotics* are still considered valid today:

- First law: A robot must not injure, or allow the injury of, any human being.
- Second law: A robot must obey all orders from humans, except orders that would contradict the first law.
- Third law: A robot must protect itself, except when to do so would contradict the first law or the second law.

Robot Arms

A robot arm, in conjunction with an *end effector* (hand, gripper, or tool), is called a *manipulator.* Some robots, especially industrial robots, are nothing more than sophisticated manipulators. A robot arm can be categorized according to its geometry. Some manipulators resemble human arms. The joints in these machines can be given names like “shoulder,” “elbow,” and “wrist.” Other manipulators are so much different from human arms that these names don't make sense.

Degrees of Freedom

The term *degrees of freedom* refers to the number of different ways in which a robot manipulator can move. Most manipulators move in three dimensions, but often they have more than three degrees of freedom.

You can use your own arm to get an idea of the degrees of freedom that a robot arm might have. Extend your right arm straight out toward the horizon. Extend your index finger so it is pointing. Keep your arm straight, and move it from the shoulder. You can move your shoulder joint in three ways. Up-and-down movement is called *pitch.* Movement to the right and left is called *yaw.* You can rotate your whole arm from the shoulder (albeit to a limited extent); this motion is called *roll.* Your shoulder therefore has three degrees of freedom: pitch, yaw, and roll.

Now move your arm from the elbow only. Holding your shoulder in the same position constantly, you will see that your elbow joint has the equivalent of pitch in your shoulder joint. But that is all. Your elbow, therefore, has one degree of freedom.

Extend your arm toward the horizon again. Now move only your arm below the elbow. Your forearm and wrist can bend up and down, side to side, and it can also twist. Your lower arm has three degrees of freedom.

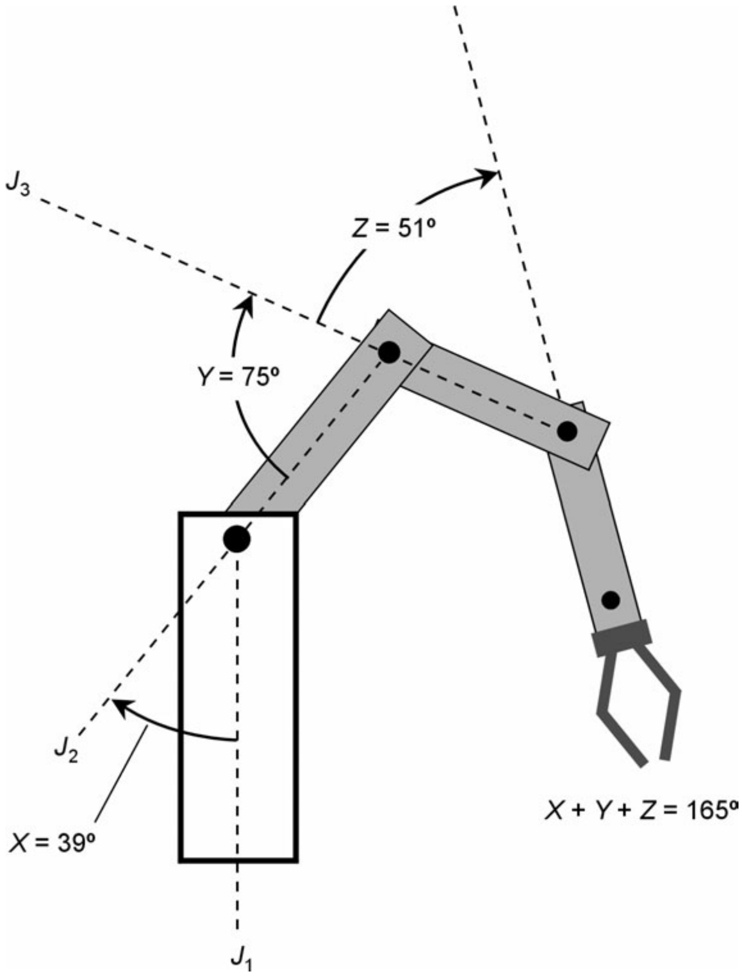

34-4 An example of degrees of rotation in articulated geometry. Angles *X*, *Y*, and *Z* are measured relative to axes J_1, J_2, and J_3, respectively.

In total, your arm has seven degrees of freedom: three in the shoulder, one in the elbow, and three in the arm below the elbow.

It is tempting to suppose that a robot should never need more than three degrees of freedom, because, after all, space has only three dimensions. But the extra possible motions, provided by multiple joints, give a robot arm (and a human arm) versatility that it could not have with only three degrees of freedom.

Degrees of Rotation

The term *degrees of rotation* refers to the extent to which a robot joint, or a set of robot joints, can turn clockwise or counterclockwise with respect to a prescribed linear axis. Some reference point is always used, and the angles are given in degrees with respect to that joint. Rotation in one direction (usually clockwise) is represented by positive angles; rotation in the opposite direction is specified by negative angles. Thus, for example, if angle $X = 58°$, it refers to a rotation of 58° clockwise with respect to the reference axis. If angle $Y = -74°$, it refers to a rotation of 74° counterclockwise.

Figure 34-4 shows a robot arm with three joints. The reference axes are J_1, J_2, and J_3 for rotation angles X, Y, and Z, respectively. The individual angles add together. To move this robot arm to a certain position within its *work envelope* (the region in space that the arm can reach and actually manipulate things), the operator enters data into a computer. This data includes the measures of angles X, Y, and Z. The operator has specified $X = 39°$, $Y = 75°$, and $Z = 51°$.

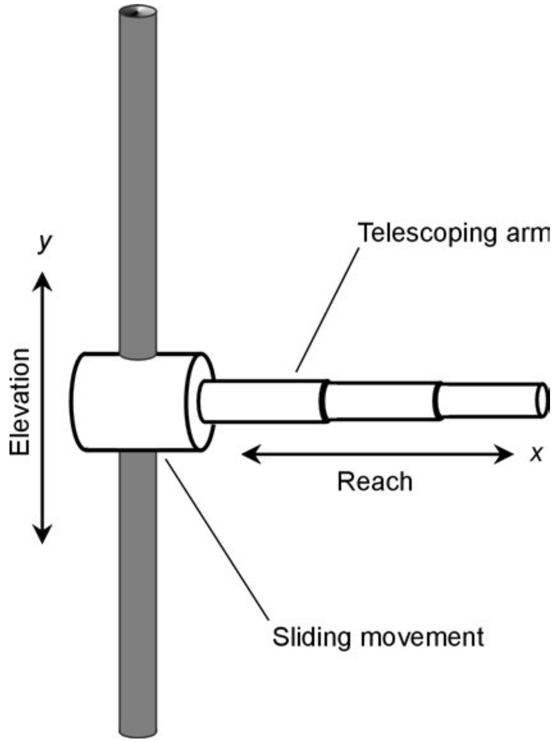

34-5 Cartesian coordinate geometry in two dimensions.

Articulated Geometry

The word *articulated* means "broken into sections by joints." A robot arm with *articulated geometry* bears some resemblance to the arm of a human. The versatility is defined in terms of the number of degrees of freedom. For example, an arm might have three degrees of freedom: base rotation (the equivalent of azimuth), elevation angle, and reach (the equivalent of radius). If you're a mathematician, you might recognize this as a system of *spherical coordinates.* There are several different articulated geometries for any given number of degrees of freedom. Figure 34-4 is a simplified drawing of a robot arm that uses articulated geometry.

Cartesian Coordinate Geometry

Another mode of robot arm movement is known as *Cartesian coordinate geometry* or *rectangular coordinate geometry.* This term comes from the *Cartesian coordinate system* often used for graphing mathematical functions. The axes are always perpendicular to each other. Variables are assigned the letters x and y in a two-dimensional *Cartesian plane*, or x, y, and z in *Cartesian three-space.* The dimensions are called *reach* for the x variable, *elevation* for the y variable, and *depth* for the z variable. Figure 34-5 is a simplified rendition of a robot arm capable of moving in two dimensions using Cartesian coordinate geometry.

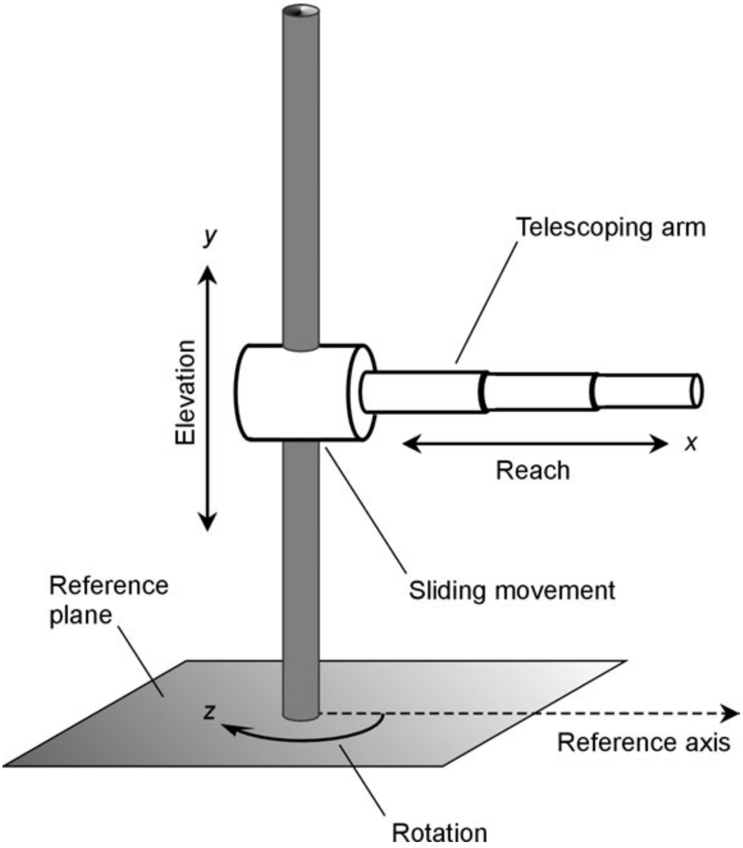

34-6 Cylindrical coordinate geometry in three dimensions.

Cylindrical Coordinate Geometry

A robot arm can be guided by means of a two-dimensional *navigator's polar coordinate system* with an elevation dimension added (Fig. 34-6). This is known as *cylindrical coordinate geometry.* In this system, a *reference plane* is used. An origin point is chosen in this plane. A *reference axis* is defined, running away from the origin in the reference plane. In the reference plane, the position of any point can be specified in terms of reach x, elevation y, and rotation z. The rotation is defined as the angle that the reach arm subtends relative to the reference axis. In this example, it is in the clockwise sense. Note that this is just like the situation for two-dimensional Cartesian coordinate geometry shown in Fig. 34-5, except that the sliding movement is also capable of rotation.

The rotation angle z can range from 0° to 360° clockwise from the reference axis. In some systems, the range is specified as 0° to +180° (up to a half circle clockwise from the reference axis), and 0° to −180° (up to a half circle counterclockwise from the reference axis).

Revolute Geometry

A robot arm capable of moving in three dimensions using *revolute geometry* is shown in Fig. 34-7. The whole arm can rotate through a full circle (360°) at the base point, or *shoulder.* There is also an elevation joint at the base that can move the arm through 90°, from horizontal to vertical. A joint in the middle of the robot arm, at the *elbow*, moves through 180°, from a straight position to dou-

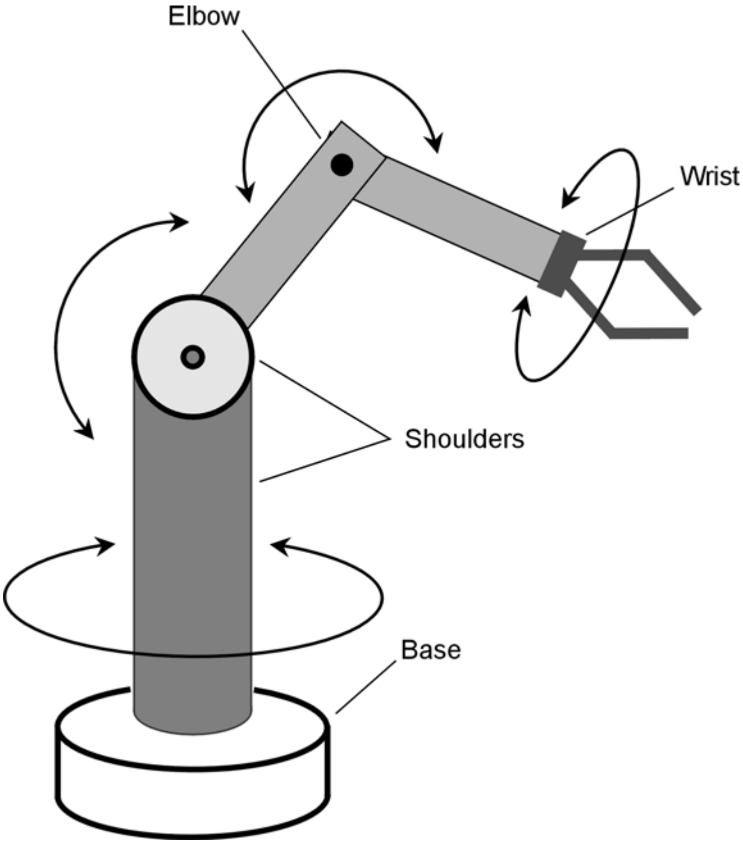

34-7 Revolute geometry in three dimensions.

bled back on itself. There might be, but is not always, a *wrist* joint that can flex like the elbow and/or twist around and around.

A 90° elevation revolute robot arm can reach any point within a half sphere. The radius of the half sphere is the length of the arm when its elbow and wrist (if any) are straightened out. A 180° elevation revolute arm can be designed that will reach any point within a fully defined sphere, with the exception of the small obstructed region around the base.

Robot Hearing and Vision

Machine hearing involves detection of acoustic waves, along with amplification and analysis of the resulting audio signals. Machine vision involves the interception of visible, infrared (IR), or ultraviolet (UV) radiation, and translating this energy into electronic images. Machine hearing and vision can allow robots to locate, and in some cases classify or identify, objects in the environment.

Binaural Hearing

Even with your eyes closed, you can usually tell from which direction a sound is coming. This is because you have *binaural hearing.* Sound arrives at your left ear with a different intensity, and in a different phase, than it arrives at your right ear. Your brain processes this information, allowing you to locate the source of the sound, with certain limitations. If you are confused, you can turn your head until the direction becomes apparent.

Robots can be equipped with binaural hearing. Two acoustic transducers are positioned, one on either side of the robot's head. A microprocessor compares the relative phase and intensity of signals from the two transducers. This lets the robot determine, within certain limitations, the direction from which sound is coming. If the robot is confused, it can turn until the confusion is eliminated and a meaningful bearing is obtained. If the robot can move around and take bearings from more than one position, a more accurate determination of the source location is possible if the source is not too far away.

Visible-Light Vision

A visible-light robotic *vision system* must have a device for receiving incoming images. This is usually a *charge coupled device* (CCD) video camera, similar to the type used in home video cameras. The camera receives an analog video signal. This is processed into digital form by an ADC. The digital signal is clarified by means of DSP. The resulting data goes to the robot controller.

The moving image, received from the camera and processed by the circuitry, contains an enormous amount of information. It's easy to present a robot controller with a detailed and meaningful moving image. But getting the robot controller to know what's happening, and to determine whether or not these events are significant, is another problem altogether.

Optical Sensitivity and Resolution

Optical sensitivity is the ability of a machine vision system to see in dim light or to detect weak impulses at invisible wavelengths. In some environments, high optical sensitivity is necessary. In others, it is not needed and might not be wanted. A robot that works in bright sunlight doesn't need to be able to see well in a dark cave. A robot designed for working in mines, pipes, or caverns must be able to see in dim light, using a system that might be blinded by ordinary daylight.

Optical resolution is the extent to which a machine vision system can differentiate between objects that are close together in the field of vision. The better the optical resolution, the keener the vision. Human eyes have excellent optical resolution, but machines can be designed to have superior resolution.

In general, the better the optical resolution, the more confined the field of vision must be. To understand why this is true, think of a telescope. The higher the magnification, the better its optical resolution will be, up to a certain maximum useful magnification. Increasing the magnification reduces the angle, or field, of vision. Zeroing in on one object or zone is done at the expense of other objects or zones.

Optical sensitivity and resolution are interdependent. If all other factors remain constant, improved sensitivity causes a sacrifice in resolution. Also, the better the optical resolution, the more incident light it requires to function well. In this case, a good analogy is camera film (the old-fashioned kind). The fastest films require more light than slow ones. The corollary to this is the fact that if you want excellent detail in a photograph, you will have to expose the film for a comparatively long period of time.

Invisible and Passive Vision

Robots have an advantage over people when it comes to vision. Machines can see at wavelengths to which humans are blind.

Human eyes are sensitive to EM waves whose length ranges from 390 to 750 nanometers (nm). The nanometer is a thousand-millionth (10^{-9}) of a meter. The longest visible wavelengths look red. As the wavelength gets shorter, the color changes through orange, yellow, green, blue, and indigo. The shortest waves look violet. Infrared (IR) energy is at wavelengths somewhat longer than 750 nm. Ultraviolet (UV) energy is at wavelengths somewhat shorter than 390 nm.

Machines, and most nonhuman living species, can see energy in a range of wavelengths that differs somewhat from the range of wavelengths to which human eyes respond. For example, insects can see UV that humans cannot, but insects are blind to red and orange light that humans can see. (Have you used orange bug lights when camping to keep the flying pests from coming around at night, or those UV devices that attract bugs and then zap them?)

A robot can be designed to see IR and/or UV, as well as (or instead of) visible light, because video cameras can be sensitive to a range of wavelengths much wider than the range humans can see. Robots can be made to see in an environment that is dark and cold, and that radiates too little energy to be detected at any electromagnetic wavelength. In these cases the robot provides its own illumination. This can be a simple lamp, a laser, an IR device, or a UV device. Radar and sonar can also be used.

Binocular Vision

Binocular machine vision is the analog of binocular human vision. It is sometimes called *stereo vision* or *stereoscopic vision.* In humans, binocular vision allows perception of depth. With only one eye—that is, with monocular vision—you can infer depth only to a limited extent, and that perception is entirely dependent on your knowledge of the environment or scene you are observing. Almost everyone has had the experience of being fooled when looking at a scene with one eye covered or blocked. A nearby pole and a distant tower might seem to be adjacent, when in fact they are a city block apart.

Figure 34-8 shows the basic concept of binocular robot vision. High-resolution video cameras, and a sufficiently powerful robot controller, are essential components of such a system.

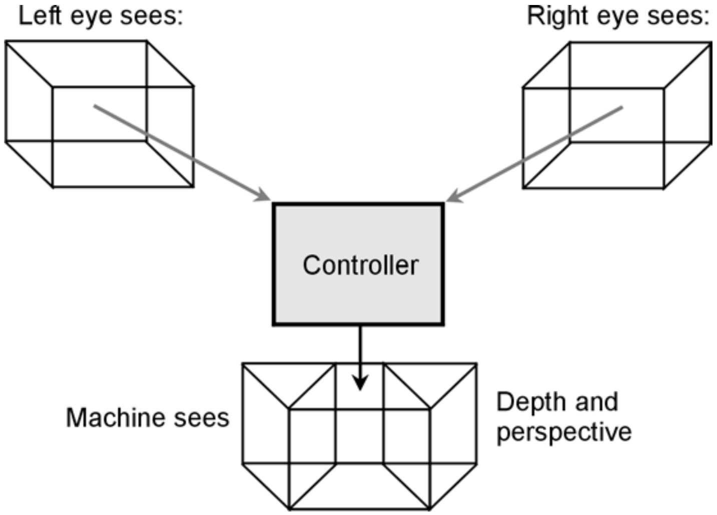

34-8 Binocular machine vision. Two different views of the same object are combined to achieve a sense of depth and perspective.

Color Sensing

Robot vision systems often function only in grayscale, like old-fashioned 1950s television. But color sensing can be added, in a manner similar to the way it is added to television systems. Color sensing can help a robot with AI figure out what an object is. Is that horizontal surface a parking lot, or is it a grassy yard? Sometimes, objects have regions of different colors that have identical brightness as seen by a grayscale system. Such objects, obviously, can be evaluated in more detail with a color-sensing system than with a vision system that sees only shades of gray.

In a typical color-sensing vision system, three grayscale cameras are used. Each camera has a color filter in its lens. One filter passes red light, another passes green light, and another passes blue light. These are the three *primary colors.* All possible hues, levels of brightness, and levels of saturation are made up of these three colors in various ratios. The signals from the three cameras are processed by a microcomputer, and the result is fed to the robot controller.

The Eye-in-Hand System

In order to assist a robot gripper (hand) in finding its way, a camera can be placed in the mechanism. The camera must be equipped for work at close range, from about 1 m down to 1 mm or less. The positioning error must be kept as small as possible. To be sure that the camera gets a good image, lamps are included in the gripper along with the camera (Fig. 34-9). This so-called eye-in-hand system can be used to precisely measure the distance between the gripper and the object it is seeking. It can also make positive identification of the object.

The eye-in-hand system takes advantage of the properties of a *servo.* The robot is equipped with, or has access to, a computer that processes the data from the camera and sends instructions back to the gripper. Most eye-in-hand systems use visible light for guidance and manipulation. Infrared (IR) can be used when it is necessary for the robot gripper to sense differences in temperature.

The Flying Eyeball

In environments hostile to humans, robots find many uses, from manufacturing to exploration. One such device, especially useful underwater, has been called a *flying eyeball.* A cable, containing

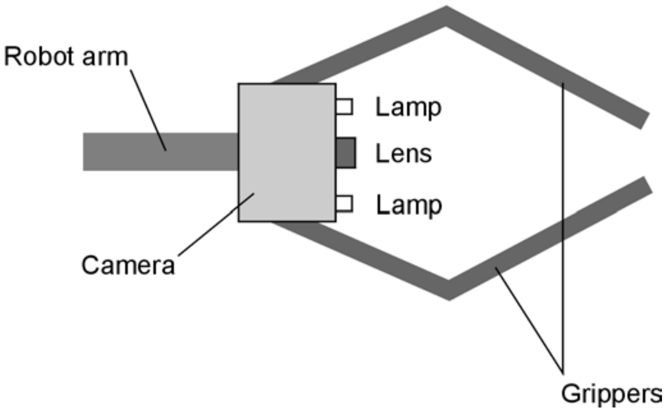

34-9 A robotic eye-in-hand system.

the robot in a special launcher housing, is dropped from a boat. When the launcher gets to the desired depth, it lets out the robot, which is connected to the launcher by a tether. The tether and the drop cable convey data back to the boat.

In some cases, the tether for a flying eyeball can be eliminated, and a wireless link can be used to convey data from the robot to the launcher. The link is usually in the IR or visible red portion of the spectrum. The robot contains a video camera and one or more lamps to illuminate the underwater environment. It also has a set of thrusters (jets or propellers) that let it move around according to control commands sent from the boat. Human operators on board the boat watch the images and guide the robot.

Robot Navigation

Mobile robots must get around in their environment without wasting motion, without running into things, and without tipping over or falling down a flight of stairs. The nature of a *robot navigation system* depends on the size of the work area, the type of robot used, and the sorts of tasks the robot is required to perform. In this section, we'll look at four common methods of robot navigation.

Clinometer

A *clinometer* is a device for measuring the steepness of a sloping surface. Mobile robots use clinometers to avoid inclines that might cause them to tip over or that are too steep for them to ascend while carrying a load.

The floor in a building is almost always horizontal. Thus, its incline is zero. But sometimes there are inclines such as ramps. A good example is the kind of ramp used for wheelchairs, in which a very small elevation change occurs. A rolling robot cannot climb stairs, but it can use a wheelchair ramp, provided the ramp is not so steep that it would upset the robot's balance or cause it to lose its payload.

In a clinometer, a transducer produces an electrical signal whenever the device is tipped from the horizontal. The greater the angle of incline, the greater the electrical output, as shown in the graph of Fig. 34-10A. A clinometer might also show whether an incline goes down or up. A downward slope might cause a negative voltage at the transducer output, and an upward slope a positive voltage, as shown in the graph at Fig. 34-10B.

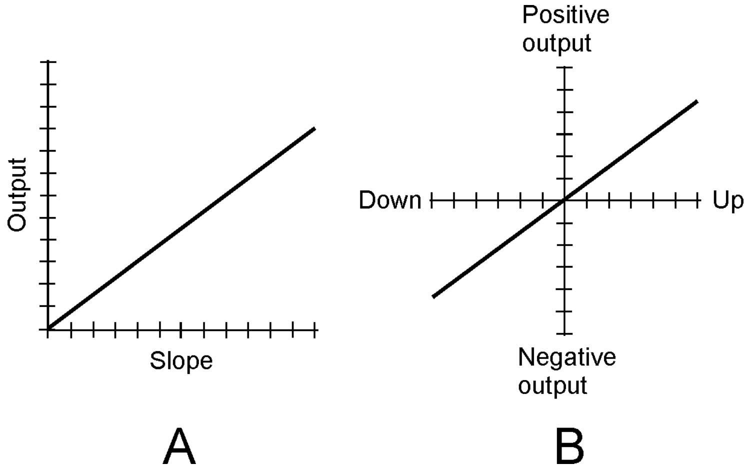

34-10 Response functions for clinometers. At A, the output voltage is a simple function of slope. At B, up-slope causes positive output voltage; downslope causes negative output voltage.

Edge Detection

The term *edge detection* refers to the ability of a robot vision system to locate boundaries. It also refers to the robot's knowledge of what to do with respect to those boundaries. A robot car, bus, or truck can use edge detection to see the edges of a road and use the data to keep itself on the road. But it must stay a certain distance from the right-hand edge of the pavement to avoid crossing into the lane of oncoming traffic (Fig. 34-11). It also must stay off the road shoulder. It must be able to tell the difference between pavement and other surfaces, such as gravel, grass, sand, and snow.

The interior of a home or business contains straight-line edge boundaries of all kinds, and each boundary represents a potential point of reference for a mobile robot's edge detection system. The controller in a personal home robot must be programmed to know the difference between, say, the line where carpet ends and tile begins, and the line where a flight of stairs begins. The vertical line produced by the intersection of two walls would present a different situation than the vertical line produced by the edge of a doorway, even though they might appear identical. Thus, edge detection cannot function very well without a certain amount of AI in the robot controller.

Embedded Path

An *embedded path* is a means of guiding a robot along a specific route. This scheme is commonly used by a mobile robot called an *automated guided vehicle* (AGV). A common embedded path consists of a buried, current-carrying wire. The current in the wire produces a magnetic field that the robot can follow. This method of guidance has been suggested as a way to keep a car on a highway, even if the driver isn't paying attention. The wire needs a constant supply of electricity for this guidance method to work. If this current is interrupted for any reason, the robot will lose its way unless some backup navigation method (or human control) is substituted.

Alternatives to wires, such as colored or reflective paints or tapes, do not need a supply of power, and this gives them an advantage. Tape is easy to remove and put somewhere else; this is difficult to do with paint and practically impossible with wires embedded in concrete. However, tape is obscured by snowfall; and at night, glare from oncoming headlights might be confused for reflections from the tape.

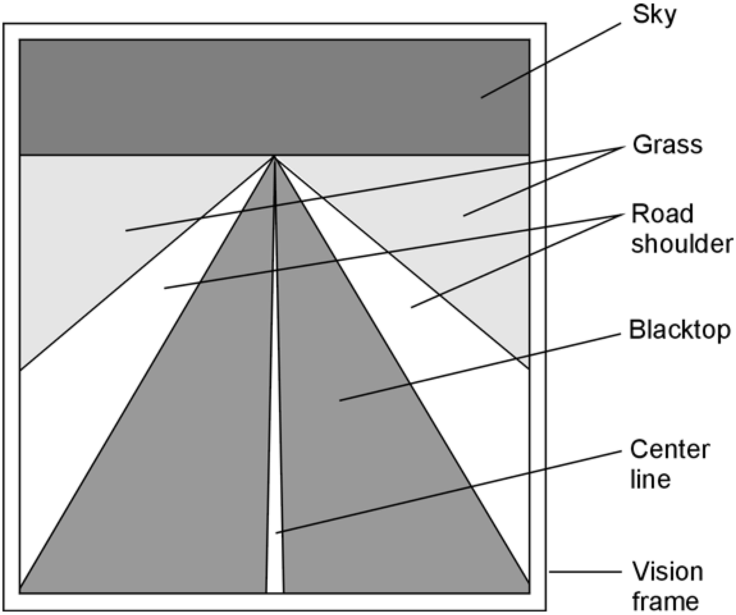

34-11 In edge detection, a robot uses visual boundaries to facilitate navigation. This is what the controller might see in a robot car driving down the center line of a highway.

Range Sensing and Plotting

Range sensing is the measurement of distances to objects in a robot's environment in a single dimension. *Range plotting* is the creation of a graph of the distance (range) to objects, as a function of the direction in two or three dimensions.

In linear or one-dimensional (1D) range sensing, a signal is sent out, and the robot measures the time it takes for the echo to come back. This signal can be sound, in which case the device is sonar. Or it can be a radio wave; this constitutes radar. Laser beams can also be used. Close-range, one-dimensional range sensing is known as *proximity sensing.*

Two-dimensional (2D) range plotting involves mapping the distance to various objects, as a function of direction in a geometric plane. The echo return time for a sonar signal, for example, might be measured every few degrees around a complete circle in the horizontal plane, resulting in a set of range points. A better plot would be obtained if the range were plotted every degree, every tenth of a degree, or even every hundredth of a degree. But no matter how detailed the direction resolution, a 2D range plot is done in only one plane, such as the floor level in a room, or some horizontal plane above the floor. The greater the number of echo samples in a complete circle (that is, the smaller the angle between samples), the more detail can be resolved at a given distance from the robot, and the greater the distance at which a given amount of detail can be resolved.

Three-dimensional (3D) range plotting is done in spherical coordinates: azimuth (compass bearing), elevation (degrees above the horizontal), and range (distance). The distance must be measured for a large number of diverse orientations. In a furnished room, a 3D sonar range plot would show ceiling fixtures, things on the floor, objects on top of a desk, and other details not visible with a 2D plot. The greater the number of echo samples in a complete sphere surrounding the robot, the more detail can be resolved at a given distance, and the greater the range at which a given amount of detail can be resolved.

Telepresence

Telepresence is a refined, advanced form of robot remote control. The robot operator gets a sense of being "on location," even if the remotely controlled machine, or *telechir*, and the operator are miles apart. Control and feedback are done by means of *telemetry* sent over wires, optical fibers, or radio.

What It's Like

Here is an example of a telepresence scenario. The robot is autonomous and has a humanoid form. The control station consists of a suit that you wear or a chair in which you sit with various manipulators and displays. Sensors can give you feelings of pressure, sight, and sound. You wear a helmet with a viewing screen that shows whatever the robot camera sees. When your head turns, the robot head, with its vision system, follows, so you see an image that changes as you turn your head, as if you were in a space suit or diving suit at the location of the robot. Binocular machine vision provides a sense of depth. Binaural machine hearing allows you to perceive sounds. Special vision modes let you see UV or IR; special hearing modes let you hear ultrasound or infrasound.

A telechir can be propelled by means of a *track drive* (similar to those used by military tanks), a *wheel drive*, or *robot legs.* If the propulsion uses legs, you propel the telechir by literally walking around a room in your telepresence suit! In the case of track drive or wheel drive, you sit in a chair and drive the robot like a tank or a car. The telechir, which is a form of android, has two arms, each with grippers resembling human hands. When you want to pick something up, you go through the normal motions with your own hands. Back-pressure sensors and position sensors let you feel (to a limited extent) what's going on. If an object masses 1 kg, it feels as if it masses 1 kg because of resistance provided by the back-pressure sensors. But it will be as if you're wearing thick gloves; you won't be able to feel texture. You might throw a switch, and something that masses 10 kg feels as if it masses only 1 kg. This might be called "strength × 10" mode. If you switch to "strength × 100" mode, a 100-kg object seems to mass only 1 kg.

Functional Description

Figure 34-12 is a simplified diagram of a robot telepresence system using an android and a wireless control link.

At the control end of the system (where the human operator is), electromechanical transducers convert the human operator's movements into electrical signals. The modem converts these signals into analog form. The analog signals modulate the RF produced by the transceiver, which in turn transmits commands by wireless to the telechir end of the system. Radio signals from the telechir arrive at the operator-end transceiver. These signals are converted into analog impulses, and are translated by the modem into electrical signals that power transducers, giving the operator a sense of what is going on.

At the telechir end of the system, the transceiver receives control signals from the human operator. These signals are translated by the modem into impulses that drive electromechanical transducers, propelling or manipulating the telechir. Data obtained by the telechir, such as its visual sense of position, environmental sounds, or the apparent mass of a lifted object, are converted by the modem into electrical signals. These signals modulate the RF energy produced by the transceiver. Signals are thus sent back to the operator end by wireless.

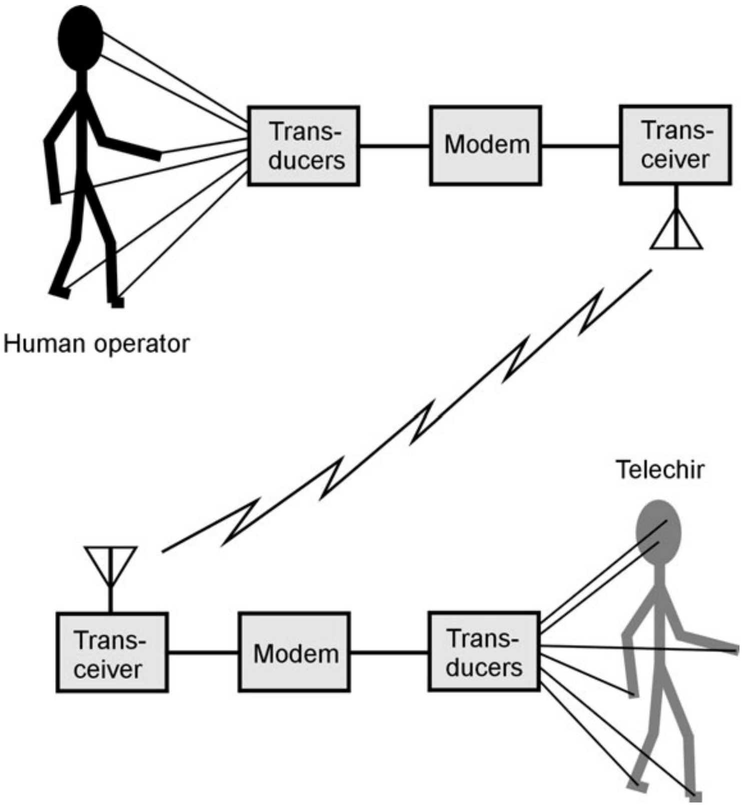

34-12 Telepresence combines remote monitoring and control, giving the operator the feeling of being onsite at a distant location.

Applications

Here are a few potential applications for a telepresence system using a human operator and an android:

- Working in extreme heat or cold
- Working under high pressure, such as on the sea floor
- Working in a vacuum, such as in space
- Working where there is dangerous radiation
- Disarming bombs
- Handling toxic substances
- Serving in high-risk police or military situations

Of course, the robot must be able to survive conditions at its location. Also, it must have some way to recover if it falls or gets knocked over.

Limitations

The technology for telepresence has existed for some time. But certain logistical problems have always bedeviled engineers committed to developing such systems.

The most serious limitation is the fact that telemetry cannot, and never will, travel faster than the speed of light in free space. This is not a problem over short-range links (a few hundred kilometers or less), but it is slow in outer space, and nigh impractical in interplanetary operations. The moon is approximately 1.3 light-seconds away from the earth; that means that any command sent to a telechir from the earth to the moon takes 1.3 s to get there, and any data from the telechir takes another 1.3 s to get back. For planets in the solar system, the delay is on the order of several minutes to several hours. On an interstellar scale, telepresence is out of the question. The nearest stars are at distances of several light-years.

Another problem is the resolution of the robot's vision. A human being with good eyesight can see things in considerable detail. To send images in real-life detail, at reasonable speed, requires a signal with broad bandwidth. There are engineering problems (and cost problems) that go along with this. However, if one is willing to deal with the cost and accommodate the signal bandwidth, robot vision systems can be designed that offer optical resolution superior to human eyesight.

Still another limitation is best put as a question: How will a robot be able to feel something and transmit these impulses to the human brain? For example, an apple feels smooth, a peach feels fuzzy, and an orange feels shiny yet bumpy. How can this sense of texture be realistically transmitted to the human brain?

The Mind of the Machine

A simple electronic calculator doesn't have AI. But a machine that can learn from its mistakes, or that can show reasoning power, does. Between these extremes, there is no precise dividing line. As computers become more powerful, people tend to set higher standards for what they call AI. Things that were once thought of as AI are now ordinary. Things that seem fantastic now will someday be humdrum. There is a tongue-in-cheek axiom: We can call "computer intelligence" true AI only as long as it remains a little bit mysterious.

Robotics and AI

Robotics and AI complement each other. Scientists have dreamed for more than a century about building *smart androids:* robots that look like people, act like people, and can even reason like people. Androids exist, but they aren't very smart. Powerful computers exist, but they lack mobility.

If a machine has the ability to move around under its own power, to lift things, and to move things, it seems reasonable that it should do so with some degree of intelligence if it is to accomplish anything worthwhile. Conversely, if a computer is to manipulate anything, it must be able to cause a machine to do physical work according to a precise program.

Expert Systems

The term *expert systems* refers to a method of reasoning in AI. Sometimes this scheme is called the *rule-based system.* Expert systems are used in the control of smart robots.

The heart of an expert system is a set of facts and rules. In the case of a robotic system, the facts consist of data about the robot's work environment, such as a factory, an office, or a kitchen. The rules are statements of the logical form "If X, then Y," similar to many of the statements in high-level programming languages. An *inference engine* decides which logical rules should be applied in various situations and instructs the robot to carry out certain tasks. But the operation of the system can only be as sophisticated as the data supplied by human programmers.

Expert systems can be used in computers to help people do research, make decisions, and make forecasts. A good example is a program that assists a physician in making a diagnosis. The computer asks questions and arrives at a conclusion based on the answers given by the patient and doctor.

One of the biggest advantages of expert systems is the fact that reprogramming is easy. As the environment changes, the robot can be taught new rules, and supplied with new facts.

How Smart a Machine?

Experts in the field of AI have been disappointed in the past few decades. Computers have been designed that can handle tasks no human could ever contend with, such as navigating a space probe. Machines have been built that can play board games well enough to compete with human masters. Modern machines can understand, as well as synthesize, much of any spoken language. But these abilities, by themselves, don't count for much in the dreams of scientists who hope to create *artificial life.*

The human mind is incredibly complicated. A circuit that would have occupied, and used all the electricity in, a whole city in 1940 can now be housed in a box the size of a vitamin pill and run by a battery. Imagine this degree of miniaturization happening again, then again, and then again. Would that begin to approach the level of sophistication in your body's nervous system?

Is the human brain nothing more than an amazingly complicated digital switching network? Or is there something more to the human mind? No electronic device yet constructed has come anywhere near human intelligence in every respect. Some experts think that a machine will someday be built that is smarter than its creators. Others insist that the very idea is ridiculous.

It is tempting to extrapolate: If technological trends of the past few decades continue indefinitely, will the only limit on machine intelligence be defined by human imagination?

Quiz

Refer to the text in this chapter if necessary. A good score is 18 correct. Answers are in the back of the book.

1. An android takes the form of
    (a) an insect.
    (b) the human body.
    (c) a simple robot arm.
    (d) a stereo vision system.

2. According to Asimov's three laws, under what circumstances is it all right for a robot to injure a human being?
    (a) Under no circumstances
    (b) When the human being specifically requests it
    (c) In case of an accident
    (d) In case the robot controller is infected with a computer virus

3. An RF field strength meter can be used to
    (a) test the performance of a binaural hearing system.
    (b) detect the presence of ionized air.

(c) measure the dielectric constant of the air.
(d) detect the presence of a wireless bugging system.

4. The extent to which a machine vision system can differentiate between two objects that are close together is called the
(a) optical magnification.
(b) optical sensitivity.
(c) optical selectivity.
(d) optical resolution.

5. A robot car or truck can best keep itself traveling down a specific lane of traffic by means of
(a) stereoscopic machine hearing.
(b) epipolar navigation.
(c) edge detection.
(d) proximity sensing.

6. A rule-based system is also known as
(a) a logic gate.
(b) an expert system.
(c) a back-pressure sensor.
(d) a telechir.

7. Suppose you are using a battery-powered, multichannel baby monitor, and you hear one end of a two-way radio conversation on the receiver. You check the baby's room, and it is quiet. How might this problem be resolved?
(a) Put the receiver in a different location.
(b) Switch the monitor to a different channel.
(c) Interchange the transmitting and receiving units.
(d) Use ac power instead of battery power.

8. In robotics, the term *manipulator* refers to
(a) a robot propulsion system.
(b) a robot arm, and the device at its end (such as a gripper).
(c) the system used to remotely control a telechir.
(d) a computer that guides a fleet of mobile robots.

9. A device with an IR sensor can be used to detect the presence of
(a) slow-moving objects.
(b) RF signals.
(c) ionized air.
(d) warm or hot objects.

10. Proximity sensing is most closely akin to
    (a) direction measurement.
    (b) edge detection.
    (c) range plotting.
    (d) binaural machine hearing.

11. A telechir is always used in conjunction with a specialized system of
    (a) track drive.
    (b) wheel drive.
    (c) remote control.
    (d) ionization potential measurement.

12. A limit to the distance over which telepresence is practical is imposed by
    (a) the speed of EM wave propagation.
    (b) the image resolution of the vision system.
    (c) the ability of a robot to determine texture.
    (d) all of the above.

13. The ionization potential of the air can be determined in order to
    (a) detect smoke.
    (b) plot distances and directions.
    (c) measure slope.
    (d) detect boundaries.

14. Two-dimensional range plotting
    (a) takes place along a single geometric line.
    (b) takes place in a single geometric plane.
    (c) is done using spherical coordinates.
    (d) requires an ultrasonic sonar system.

15. Spherical coordinates can uniquely define the position of a point in up to
    (a) one dimension.
    (b) two dimensions.
    (c) three dimensions.
    (d) four dimensions.

16. The total number of ways in which a robot arm can move is known as
    (a) functional orientation.
    (b) degrees of freedom.
    (c) dimensional versatility.
    (d) coordinate geometry.

17. The region in space throughout which a robot arm can accomplish tasks is called its
    (a) coordinate geometry.
    (b) reference axis.
    (c) reference frame.
    (d) work envelope.

18. A robot arm that moves along three independent axes, each of which is straight and perpendicular to the other two, employs
    (a) revolute geometry.
    (b) spherical coordinate geometry.
    (c) Cartesian coordinate geometry.
    (d) cylindrical coordinate geometry.

19. Fill in the blank to make the following sentence true: "A color vision system can use three grayscale cameras, equipped with filters that pass ____ light."
    (a) red, yellow, and blue
    (b) blue, red, and green
    (c) white, black, and gray
    (d) orange, green, and violet

20. A robot typically determines the steepness of a slope by means of
    (a) an epipolar navigation system.
    (b) a clinometer.
    (c) an end effector.
    (d) a proximity sensor.

Test: Part 4

DO NOT REFER TO THE TEXT WHEN TAKING THIS TEST. A GOOD SCORE IS AT LEAST 37 CORRECT. Answers are in the back of the book. It's best to have a friend check your score the first time, so you won't memorize the answers if you want to take the test again.

1. Which of the following tasks can a displacement transducer perform?
   (a) It can detect the peak amplitude of an ac signal and display it on a screen.
   (b) It can convert acoustic energy into radio signals.
   (c) It can detect radiated IR energy and display IR images on a screen.
   (d) It can convert an electrical signal into mechanical rotation through a defined angle.
   (e) It can be used to measure the frequency of a complex wave.

2. Which of the following is *not* an advantage of an IC, compared to a circuit built with discrete components (individual resistors, capacitors, inductors, diodes, and transistors)?
   (a) Compactness
   (b) Reliability
   (c) Ease of maintenance
   (d) Low power consumption
   (e) Unlimited power handling capacity

3. Computer memory is typically measured in
   (a) kilobits, megabits, and gigabits.
   (b) kilobits per second, megabits per second, and gigabits per second.
   (c) kilobytes, megabytes, and gigabytes.
   (d) kilobytes per second, megabytes per second, and gigabytes per second.
   (e) any of the above.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

4. A mode of communications in which either party can hear while talking, thus allowing one party to instantly interrupt the other at any time, is called
   (a) half simplex.
   (b) full simplex.
   (c) half duplex.
   (d) full duplex.
   (e) nothing, because there is no such mode.

5. An FM stereo tuner generally covers a frequency range of
   (a) 535 kHz to 1605 kHz.
   (b) 88 MHz to 108 MHz.
   (c) 3 MHz to 30 MHz.
   (d) 9 kHz to 300 MHz.
   (e) 144 MHz to 148 MHz.

6. Suppose an antenna has a radiation resistance of 35 Ω and a loss resistance of 15 Ω. What is the efficiency of the antenna?
   (a) 20 percent
   (b) 30 percent
   (c) 43 percent
   (d) 70 percent
   (e) Impossible to calculate from this information

7. The peak power output from a multicavity Klystron
   (a) can be much greater than the average power output.
   (b) can be slightly greater than the average power output.
   (c) is about the same as the average power output.
   (d) is slightly less than the average power output.
   (e) is much less than the average power output.

8. To which of the following does the term *platter* apply?
   (a) An individual disk in a hard drive
   (b) An individual disk in CD-R media
   (c) An individual disk in CD-RW media
   (d) An individual disk in flash memory
   (e) An individual disk in RAM

9. Why can signals in the so-called shortwave band sometimes propagate for thousands of kilometers without the need for satellites, repeaters, or any other infrastructure?
   (a) Because EM waves at these frequencies are often returned to the earth by the ionosphere
   (b) Because waves at these frequencies can propagate through the ground or the ocean as electric currents

(c) Because EM waves at these frequencies are reflected back to the earth by the moon
(d) Because EM waves at these frequencies propagate along the boundary between the troposphere and the stratosphere
(e) Forget it! Signals in the shortwave band cannot propagate beyond a line of sight without satellites, repeaters, or other human-made systems.

10. What is the purpose of the filament in a vacuum tube?
(a) It prevents secondary electrons from causing excessive screen grid current.
(b) It causes holes to flow more easily in the collector.
(c) It heats the cathode, thereby increasing the electron emission.
(d) It glows, making the tube easy to locate when servicing is necessary.
(e) Forget it! Vacuum tubes do not have filaments.

11. Figure Test4-1 is a schematic diagram of

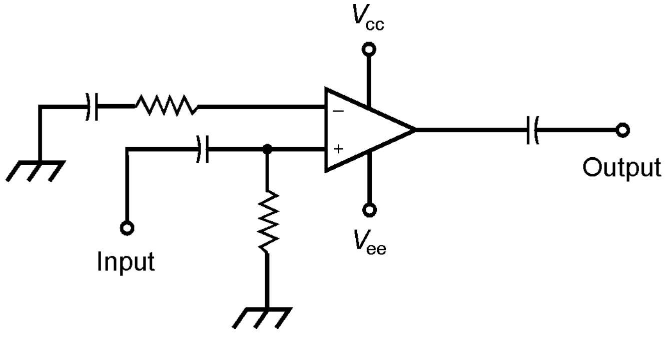

Test 4-1 Illustration for Part 4 Test Questions 11 and 12.

(a) a closed-loop op amp circuit with positive feedback.
(b) a closed-loop op amp circuit with negative feedback.
(c) an open-loop op amp circuit.
(d) an op amp oscillator circuit.
(e) an op amp modulator circuit.

12. Suppose that, in the circuit of Fig. Test4-1, a resistor is connected between the output and the inverting input. What effect will this have?
(a) It will increase the gain.
(b) It will decrease the gain.
(c) It will increase the frequency.
(d) It will decrease the frequency.
(e) It will have no effect.

13. Which of the following is an example of a serial-access storage medium for computer data?
(a) An external hard drive
(b) An internal hard drive

(c) A CD-R
(d) A CD-RW
(e) A magnetic tape

14. Which of the following is an undesirable characteristic of ELF radiation?
(a) It can cause radiation burns like those produced by overexposure to X rays.
(b) It can increase the susceptibility of a CRT to jitter and poor resolution.
(c) It can reduce the efficiency of an antenna system.
(d) It can result in undesirable modulation in a radio transmitter.
(e) None of the above apply.

15. A zepp antenna measuring $\lambda/2$ can be oriented vertically, and the feed line placed so it lies in the same line as the radiating element. This antenna is known as
(a) a vertical dipole.
(b) a Yagi.
(c) a collinear array.
(d) an end-fire array.
(e) a J pole.

16. The GPS can be used to determine the position of a point in
(a) one dimension.
(b) two dimensions.
(c) three dimensions.
(d) four dimensions.
(e) five dimensions.

17. In a component-type hi-fi system, the cables connecting the various devices should
(a) be resonant at the intended operating frequency.
(b) consist of single conductors only.
(c) be connected in parallel.
(d) consist of two parallel wires whenever possible.
(e) be shielded whenever practicable.

18. In a radar display, the azimuth of a target is equal or proportional to
(a) the distance of the blip representing the target from the center of the display.
(b) the clockwise angle between a radial line from the center toward geographic north, and a radial line from the center through the blip representing the target.
(c) the diameter of a circle whose center coincides with the center of the display, and that passes through the blip representing the target.

(d) the distance of the blip representing the target from the outer edge of the display, or from a circle whose diameter is equal to the diameter of the display.
(e) Forget it! A radar set cannot determine the azimuth of a target.

19. Which of the following is an example of wireless technology?
    (a) The Global Positioning System
    (b) A fiber-optic communications system
    (c) Digital-to-analog conversion
    (d) Serial-to-parallel conversion
    (e) All of the above

20. A smoke detector can function by sensing a change in the
    (a) temperature of the air.
    (b) relative humidity of the air.
    (c) barometric pressure of the air.
    (d) ionization potential in the air.
    (e) amount of nitrogen in the air.

21. Computer disk storage is typically measured in
    (a) kilobits, megabits, and gigabits.
    (b) kilobits per second, megabits per second, and gigabits per second.
    (c) kilobytes, megabytes, and gigabytes.
    (d) kilobytes per second, megabytes per second, and gigabytes per second.
    (e) any of the above.

22. Which of the following antennas would most likely be used in space communications?
    (a) A half-wave dipole antenna
    (b) A ground-plane antenna
    (c) A helical antenna
    (d) A small loop or loopstick antenna
    (e) A coaxial antenna

23. The use of a cell phone may be prohibited in a commercial aircraft in flight because
    (a) it can interfere with other people's computers.
    (b) it can interfere with other people's cell phones.
    (c) the interior of the aircraft produces a dangerous RF resonant cavity.
    (d) the area of coverage will not be wide enough to be of any use.
    (e) it can interfere with flight instruments and communications.

24. The circuit shown in Fig. Test4-2 shows two NPN transistors connected in
    (a) reverse series.
    (b) cascade.

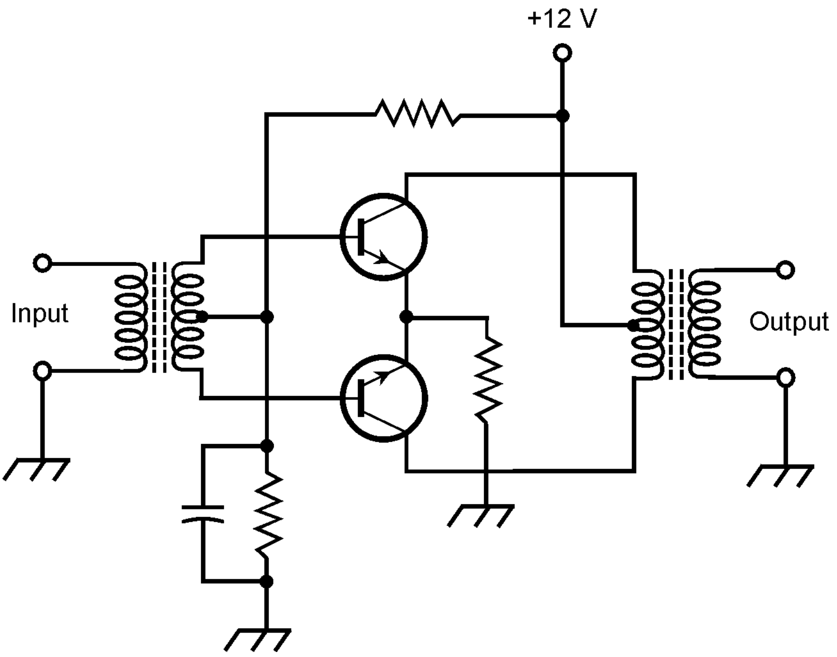

Test 4-2 Illustration for Part 4 Test Questions 24 and 25.

(c) inverse parallel.
(d) push-push.
(e) push-pull.

25. The circuit in Fig. Test4-2 would be a bad choice for use as a medium-power audio amplifier because
   (a) bipolar transistors do not work well in audio applications.
   (b) transformers should not be used in audio amplifiers.
   (c) this type of circuit invariably produces a lot of distortion.
   (d) this type of circuit cannot produce appreciable output power.
   (e) Hold it! The circuit in Fig. Test4-2 would be a good choice for use as a medium-power audio amplifier.

26. In an electrostatic CRT, the electron beam can be diverted up and down by
   (a) synchronization pulses applied to the deflecting coils.
   (b) ELF radiation from the accelerating anodes.
   (c) charged plates that attract or repel moving electrons.
   (d) modulation of the signal on the control grid.
   (e) variable input to the electron gun.

27. A device that translates 500-Hz sound waves into ac by means of the interaction between a permanent magnet and a coil of wire is an example of
    (a) an electromechanical loudspeaker.
    (b) a passive pressure sensor.
    (c) a dynamic transducer.
    (d) an inductive proximity sensor.
    (e) a piezoelectric transducer.

28. Which of the following considerations is most important in a hi-fi audio power amplifier?
    (a) Optimizing the efficiency
    (b) Minimizing the operating voltage
    (c) Maximizing the linearity
    (d) Maximizing the driving power
    (e) Biasing beyond the cutoff point

29. A loopstick antenna, oriented so its ends point toward the eastern and western horizons, exhibits
    (a) nulls in the response to signals coming from the east and west.
    (b) peaks in the response to signals coming from the east and west.
    (c) nulls in the response to signals coming from the north and south.
    (d) an omnidirectional response in three-dimensional space.
    (e) a peak in the response to signals coming from any horizontal direction, and a null in the response to signals coming from the zenith (directly overhead).

30. Fill in the blank to make the following sentence true: "An ____ functions by sensing variations in the relative phase of acoustic waves reflected from objects in the environment."
    (a) IR transducer
    (b) acoustic transponder
    (c) electrostatic loudspeaker
    (d) ultrasonic motion detector
    (e) acoustic pressure sensor

31. Which of the following is an advantage of a large LCD over a large CRT display?
    (a) Lower cost
    (b) Lighter weight
    (c) Greater ELF radiation
    (d) The use of electron beams and deflecting coils
    (e) All of the above

32. An op amp can produce or facilitate
    (a) a low signal-to-noise ratio in an oscillator.
    (b) improved efficiency in an antenna system.

(c) high signal gain over a wide range of frequencies.
(d) a match between a feed line and an antenna.
(e) none of the above.

33. Which of the following is an application of a gas-filled tube?
(a) A high-frequency RF power amplifier
(b) A microwave oscillator
(c) An audio oscillator or amplifier
(d) Decorative lighting
(e) Impedance matching in an antenna system

34. For a communications satellite with an elliptical orbit around the earth, the point at which the altitude is lowest is the
(a) geominimum.
(b) proxima.
(c) perigee.
(d) approach.
(e) optimum.

35. Suppose an antenna has a radiation resistance of 35 Ω and its feed line has a characteristic impedance of 50 Ω. What is the efficiency of the antenna?
(a) 15 percent
(b) 41 percent
(c) 59 percent
(d) 70 percent
(e) Impossible to calculate from this information

36. Which of the following is a characteristic of a stepper motor?
(a) It rotates in defined increments, not continuously.
(b) Its turning power increases as its speed increases.
(c) It cannot rotate faster than approximately 1 rpm.
(d) It is not suitable for use in mechanical devices, but only in sensors.
(e) It operates without an external source of power.

37. With respect to ICs, terms such as MSI, LSI, and VLSI define the
(a) maximum number of transistors on the chip.
(b) maximum frequency at which the chip can operate.
(c) maximum power output the chip can produce.
(d) maximum gain the chip can produce.
(e) maximum number of switching operations the chip can perform per second.

38. In a tetrode vacuum tube amplifier, the screen grid
    (a) reduces the capacitance between the control grid and the plate.
    (b) serves as an auxiliary output in case the plate circuit is overloaded.
    (c) conducts holes away from the control grid.
    (d) prevents undesirable reverse bias between the cathode and the control grid.
    (e) Forget it! Tetrode tubes do not have screen grids.

39. Fill in the blank in the following sentence to make it true: "On the Web, data is transmitted by means of a ____ that allows the files on distant computers to appear as if they are on your own computer."
    (a) protocol
    (b) memory chip
    (c) microprocessor
    (d) CD-R
    (e) tape drive

40. A robot arm with articulated geometry
    (a) has only one degree of freedom.
    (b) has joints, similar to those in a human arm.
    (c) can function in only two dimensions.
    (d) can rotate and move up and down, but cannot bend.
    (e) can perform only coarse movements.

41. The instantaneous output voltage of a differentiator is proportional to
    (a) the peak instantaneous input voltage.
    (b) the rate at which the instantaneous input voltage changes.
    (c) the accumulated input voltage as a function of time.
    (d) the difference between the instantaneous voltages of the two input signals.
    (e) the sum of the instantaneous voltages of the two input signals.

42. What types of devices require particular care in handling, so they are not destroyed by electrostatic discharges that can build up on a technician's body?
    (a) Vacuum tubes
    (b) Ferrite resistors
    (c) Ceramic capacitors
    (d) MOS components
    (e) Rectifier diodes

43. According to Isaac Asimov's rules for the behavior of robots, a robot must not injure, or allow the injury of, any human,
    (a) unless the robot is forced to prevent its own destruction.
    (b) unless the human is committing a crime.

(c) unless the human is trying to turn off the robot's power supply.
(d) unless it is the result of an accident.
(e) under any circumstances.

44. Computer data speed is typically measured in
   (a) kilobits, megabits, and gigabits.
   (b) kilobits per second, megabits per second, and gigabits per second.
   (c) kilobytes, megabytes, and gigabytes.
   (d) kilobytes per second, megabytes per second, and gigabytes per second.
   (e) any of the above.

45. Doppler radar is useful for measuring or estimating
   (a) the frequency of a sine wave that is modulated with a complex signal.
   (b) the wind speed in the funnel cloud of a tornado.
   (c) the distance between a robot gripper and a tool or object to be manipulated.
   (d) the depth of the ocean at a particular location.
   (e) the direction from which a thunderstorm is coming.

46. A device that translates a 30-kHz ac voltage into ultrasonic waves by producing stress on a crystal, thereby causing the crystal to vibrate, is an example of
   (a) an electromechanical loudspeaker.
   (b) a passive pressure sensor.
   (c) a dynamic transducer.
   (d) an inductive proximity sensor.
   (e) a piezoelectric transducer.

47. Fill in the blank in the following sentence to make it true: "A ____ antenna is a vertical radiator, usually measuring $\lambda/4$, elevated above the surface of the earth, and operated against a system of $\lambda/4$ horizontal or slightly drooping radials."
   (a) dipole
   (b) coaxial
   (c) collinear
   (d) ground-plane
   (e) vertizontal

48. Figure Test4-3 shows a robotic system that employs
   (a) Cartesian coordinate geometry.
   (b) spherical coordinate geometry.
   (c) polar coordinate geometry.
   (d) cylindrical coordinate geometry.
   (e) Riemannian coordinate geometry.

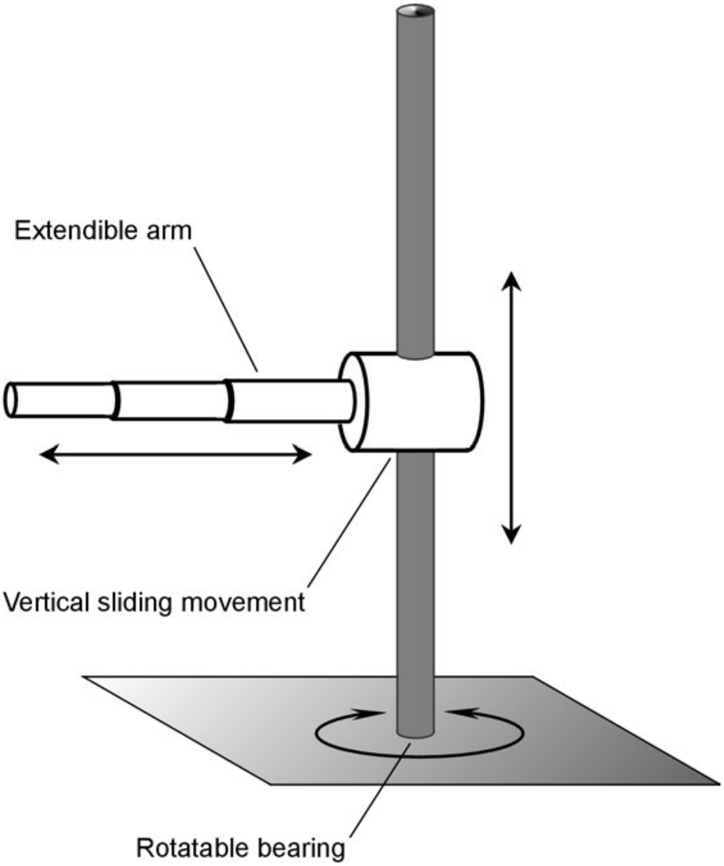
Extendible arm
Vertical sliding movement
Rotatable bearing

Test 4-3 Illustration for Part 4 Test Question 48.

49. A satellite in a LEO system is usually placed in
    (a) an equatorial orbit.
    (b) a geosynchronous orbit.
    (c) a retrograde orbit.
    (d) a polar orbit.
    (e) an orbit midway between the earth and the orbit of the moon.

50. Sound waves in the atmosphere consist of
    (a) rotating magnetic fields.
    (b) alternating electric fields.
    (c) variable electron beams.
    (d) moving molecules.
    (e) rotating electron orbits within atoms.

Final Exam

DO NOT REFER TO THE TEXT WHEN TAKING THIS EXAM. A GOOD SCORE IS AT LEAST 75 CORRECT OUT of 100. Answers are in the back of the book. It's best to have a friend check your score the first time, so you won't memorize the answers if you want to take the exam again.

1. A wire can be coiled around a magnetic compass to make
   (a) an ac wattmeter.
   (b) an ac voltmeter.
   (c) a dc wattmeter.
   (d) a dc ohmmeter.
   (e) a dc galvanometer.

2. The inductive reactance of a fixed air-core coil of wire, assuming the wire has zero dc resistance (it conducts perfectly),
   (a) does not change as the frequency changes.
   (b) increases as the frequency increases.
   (c) decreases as the frequency increases.
   (d) alternately increases and decreases as the frequency increases.
   (e) is zero at all frequencies.

3. *Power* is defined as
   (a) the rate at which current flows in a circuit.
   (b) the product of voltage and resistance in a circuit.
   (c) the rate at which energy is radiated or dissipated.
   (d) the accumulation of energy over time.
   (e) the amount of heat generated in a circuit.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

4. In an amplifier that employs a P-channel JFET, the device can usually be replaced with an N-channel JFET having similar specifications, provided that
   (a) all the resistors are reversed in polarity for the circuit in question.
   (b) the power supply polarity is reversed for the circuit in question.
   (c) the drain, rather than the source, is placed at signal ground.
   (d) the input is supplied to the drain, rather than to the gate.
   (e) the output is taken from the source, rather than from the drain.

5. At the exact moment a 60-Hz ac sine wave is at its positive peak voltage, the instantaneous rate of change in the voltage is
   (a) large and positive.
   (b) small and positive.
   (c) large and negative.
   (d) small and negative.
   (e) zero.

6. In order for a bipolar transistor to conduct under conditions of no signal input, the bias must be
   (a) in the forward direction at the emitter-base (E-B) junction, sufficient to cause forward breakover.
   (b) in the reverse direction at the E-B junction, but not sufficient to cause avalanche effect.
   (c) such that the application of a signal would cause the transistor to go into a state of cutoff.
   (d) such that the application of a signal would cause the transistor to go into a state of saturation.
   (e) such that the application of a signal would cause the transistor to become nonlinear.

7. The risk of electrocution to personnel who service "live" electrical or electronic equipment increases in proportion to the
   (a) maximum amount of power, in watts, that the equipment demands.
   (b) maximum amount of energy, in watt-hours, that the equipment consumes.
   (c) maximum amount of current, in amperes, that the equipment draws.
   (d) maximum potential difference, in volts, that exists in the equipment.
   (e) maximum resistance, in ohms, that exists in the equipment.

8. In a logical AND gate with four inputs, the output is high if and only if
   (a) none of the inputs are high, and all four are low.
   (b) one of the inputs is high, and the other three are low.
   (c) two of the inputs are high, and the other two are low.
   (d) three of the inputs are high, and the other one is low.
   (e) all four of the inputs are high, and none are low.

9. Fill in the blank in the following sentence to make it true: "A ____ receiver derives its audio output by mixing incoming signals with the output of a tunable local oscillator."
   (a) crystal set
   (b) direct conversion

(c) phase-locked
(d) very low frequency
(e) frequency modulation

10. In a power supply filter, the term *L section* refers to a combination of
(a) a capacitor in parallel with the rectifier output, and a choke in series.
(b) a capacitor in parallel with the rectifier output, and another capacitor in series.
(c) a choke in parallel with the rectifier output, and a capacitor in series.
(d) a diode in parallel with the transformer output, and another diode in series.
(e) any of the above.

11. What is the frequency of a pure sine wave that has a period of 50 nanoseconds? (A nanosecond is equal to one thousand-millionth, or 10^{-9}, of a second.)
(a) 2.0 MHz
(b) 20 MHz
(c) 0.20 GHz
(d) 2.0 GHz
(e) 20 GHz

12. Fill in the blank in the following sentence to make it correct: "The peak inverse voltage rating of a rectifier diode is the maximum instantaneous reverse-bias voltage that it can withstand without ____ taking place."
(a) forward breakover
(b) avalanche effect
(c) harmonic generation
(d) oscillation
(e) thermal destruction

13. Suppose a coil and capacitor are connected in series, with $jX_L = j50$ and $jX_C = -j90$, and there is no resistance. What is the complex impedance of this combination?
(a) $0 - j40$
(b) $0 + j40$
(c) $40 + j50$
(d) $50 - j90$
(e) It cannot be determined without more information.

14. In an IC, the term *linear* arises from the fact that
(a) all of the etchings on the semiconductor material are straight lines.
(b) the amplification factor generally decreases in direct proportion to the instantaneous input amplitude.
(c) the amplification factor generally increases in direct proportion to the instantaneous input amplitude.
(d) the amplification factor is generally constant as the instantaneous input amplitude varies.
(e) the device never exhibits distortion under any circumstances.

15. Suppose a dc circuit has 10 kΩ of resistance, and a 12-V battery is connected to it. How much current is drawn from the battery?
    - (a) 1.2 μA
    - (b) 12 μA
    - (c) 1.2 mA
    - (d) 12 mA
    - (e) None of the above

16. The high input impedance of a MOSFET makes this type of device ideal for use in
    - (a) weak-signal amplifiers.
    - (b) high-power oscillators.
    - (c) high-current rectifiers.
    - (d) antenna tuning networks.
    - (e) graphic equalizers.

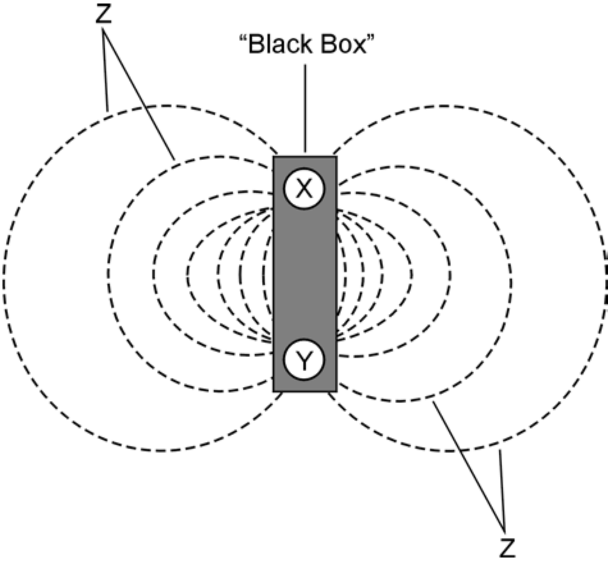

Exam-1 Illustration for Final Exam Question 17.

17. In Fig. Exam-1, the dashed lines labeled (Z) can represent
    - (a) the electric lines of flux in the vicinity of two electrically charged objects (X and Y inside the black box) having opposite polarity.
    - (b) the magnetic lines of flux in the vicinity of two electrically charged objects (X and Y inside the black box) having opposite polarity.
    - (c) the electric lines of flux in the vicinity of two electrically charged objects (X and Y inside the black box) having the same polarity.
    - (d) the magnetic lines of flux in the vicinity of two electrically charged objects (X and Y inside the black box) having the same polarity.
    - (e) any of the above.

18. The drain of a JFET is the analog of the
    - (a) plate of a vacuum tube.
    - (b) emitter of a bipolar transistor.
    - (c) cathode of a diode.
    - (d) positive electrode in a solar cell.
    - (e) substrate of a MOSFET.

19. Imagine a circuit with 21 Ω of inductive reactance and 28 Ω of resistance. What is the absolute-value impedance of this circuit?
    - (a) 7 Ω
    - (b) 24 Ω
    - (c) 35 Ω
    - (d) 49 Ω
    - (e) None of the above

20. Suppose you want to obtain a 12-V battery by connecting 1.5-V size D flashlight cells together. This can be done by
    - (a) connecting 8 cells together in parallel, making sure to always connect the terminals so they are plus-to-plus and minus-to-minus.
    - (b) connecting 8 cells together in series, making sure to always connect the terminals so they are plus-to-minus.
    - (c) either connecting 8 cells together in parallel as described in (a), or connecting 8 cells together in series as described in (b).
    - (d) connecting two sets of 4 cells together in series, making sure to always connect the terminals so they are plus-to-minus, and then connecting these sets together in parallel, making sure to connect the terminals so they are plus-to-plus and minus-to-minus.
    - (e) connecting two sets of 4 cells together in parallel, making sure to always connect the terminals so they are plus-to-plus and minus-to-minus, and then connecting these sets together in series, making sure to connect the terminals so they are plus-to-minus.

21. One of the technical limitations of capacitive proximity sensors is the fact that they
    - (a) are not very sensitive to objects that are poor electrical conductors.
    - (b) are insensitive to objects that reflect light.
    - (c) are insensitive to metallic objects.
    - (d) cannot be used with oscillators.
    - (e) require extreme voltages in order to function properly.

22. A flute sounds different than a violin, even if the two instruments are played at the same pitch, because of a difference in the
    - (a) phase.
    - (b) chamber length.
    - (c) frequency.
    - (d) waveform.
    - (e) bias.

23. In a personal computer, the internal mass-storage medium is nearly always
    - (a) a CD (compact disc) in some form.
    - (b) a diskette drive.
    - (c) a tape drive.
    - (d) read-only memory (ROM).
    - (e) none of the above.

24. A device to which signals can be sent asking for data, and that responds with the requested information by means of radio or a wireless network, is known as a
    - (a) transceiver.
    - (b) transponder.
    - (c) repeater.
    - (d) duplexer.
    - (e) signal reflector.

25. Suppose you want to use a transformer to obtain exactly 30 V rms ac from an electrical outlet that supplies exactly 120 V rms ac. The primary-to-secondary turns ratio of the transformer should be
    - (a) exactly 1:16.
    - (b) exactly 1:4.
    - (c) variable.
    - (d) exactly 4:1.
    - (e) exactly 16:1.

26. Fill in the blank in the following statement to make it true: "Two problems can occur with the ____ type of intrusion alarm. First, the sun, suddenly emerging from clouds or from shadow, might shine on the sensor and trigger the device when there is no threat. Second, someone with malicious intent, wearing winter clothing and entering on a cold day, might fail to set off the alarm in the presence of a real threat."
    - (a) IR radiation detector
    - (b) electromagnetic detector
    - (c) ultrasonic detector
    - (d) ultraviolet detector
    - (e) fluxgate detector

27. Suppose that five resistors, each one having the same ohmic value as all the others, are connected in series, and the entire combination is connected to a 12-V battery. How does the voltage across any one of the resistors compare with the battery voltage, assuming no other loads are connected to the circuit?
    - (a) The voltage across any one of the resistors is 5 times the battery voltage.
    - (b) The voltage across any one of the resistors is the same as the battery voltage.
    - (c) The voltage across any one of the resistors is $1/5$ of the battery voltage.
    - (d) The voltage across any one of the resistors is $1/25$ of the battery voltage.
    - (e) It is impossible to answer this without more information.

28. When the carrier and one of the sidebands are removed from an amplitude-modulated signal, the result is
    - (a) a frequency-modulated signal.
    - (b) a phase-modulated signal.
    - (c) a frequency-shift-keyed signal.
    - (d) a pulse-modulated signal.
    - (e) a single-sideband signal.
29. In log polar navigation, a computer converts a polar coordinate image to
    - (a) a Riemannian coordinate image.
    - (b) a spherical coordinate image.
    - (c) a cylindrical coordinate image.
    - (d) a curvilinear coordinate image.
    - (e) a rectangular coordinate image.
30. The charged particles in the nucleus of an atom are
    - (a) electrons.
    - (b) protons.
    - (c) positrons.
    - (d) neutrons.
    - (e) negatrons.
31. Which of the following functions (a), (b), (c), or (d), if any, is a semiconductor diode never used to perform?
    - (a) Rectification
    - (b) Detection
    - (c) Frequency multiplication
    - (d) Signal mixing
    - (e) A semiconductor diode can perform any of the above functions.
32. The reactance of a fixed capacitor, assuming it has no leakage conductance (that is, it has a theoretically infinite dc resistance),
    - (a) does not change as the frequency changes.
    - (b) increases negatively as the frequency increases.
    - (c) decreases negatively (approaches zero) as the frequency increases.
    - (d) alternately becomes negative and positive as the frequency increases.
    - (e) is theoretically infinite at all frequencies.
33. What is the dc resistance of a component with a dc conductance of 10^{-4} S?
    - (a) 0.0001 Ω
    - (b) 0.01 Ω
    - (c) 100 Ω
    - (d) 10 kΩ
    - (e) None of the above

34. A common-collector transistor circuit is often used
    - (a) to provide high gain and sensitivity over a wide range of frequencies.
    - (b) to match a high impedance to a low impedance.
    - (c) as a high-fidelity audio power amplifier.
    - (d) as an oscillator at microwave frequencies.
    - (e) as the rectifier in a dc power supply.

35. The internal conductance (expressed in siemens) of an ammeter is generally
    - (a) low.
    - (b) directly proportional to the current.
    - (c) inversely proportional to the current.
    - (d) high.
    - (e) any value; it doesn't matter.

36. A half-wave length of wire, fed at one end with a quarter-wave section of parallel-wire transmission line, is known as
    - (a) a zeppelin antenna.
    - (b) a dipole antenna.
    - (c) a Yagi antenna.
    - (d) a random wire antenna.
    - (e) an end-fire antenna.

37. A twin T oscillator is commonly used for generating
    - (a) AF signals.
    - (b) high-frequency RF signals.
    - (c) microwave RF signals.
    - (d) powerful bursts of RF energy.
    - (e) FM signals.

38. Suppose that five resistors, each one having the same ohmic value as all the others, are connected in series, and the entire combination is connected to a 12-V battery. How does the current through any one of the resistors compare with the current drawn from the battery?
    - (a) The current through any one of the resistors is 5 times the current drawn from the battery.
    - (b) The current through any one of the resistors is the same as the current drawn from the battery.
    - (c) The current through any one of the resistors is $1/5$ as great as the current drawn from the battery.
    - (d) The current through any one of the resistors is $1/25$ as great as the current drawn from the battery.
    - (e) It is impossible to answer this without more information.

39. When a nonzero net reactance exists in an ac circuit, the volt-ampere (VA) power, also known as the *apparent power*, is

(a) zero.

(b) equal to the true power.

(c) less than the true power.

(d) greater than the true power.

(e) infinite.

40. When an impurity, also called a *dopant*, is added to a semiconductor material, it can contain either an excess of electrons or a deficiency of electrons. A dopant with an excess of electrons, also known as a *donor impurity*, results in the production of

(a) an NPN transistor.

(b) a PNP transistor.

(c) a P-N junction.

(d) a P type semiconductor.

(e) an N type semiconductor.

41. Suppose an audio amplifier produces 100 W rms output when the input is 5.00 W rms. This represents a power gain of

(a) 1.30 dB.

(b) 2.60 dB.

(c) 13.0 dB.

(d) 20.0 dB.

(e) 26.0 dB.

Exam-2 Illustration for Final Exam Questions 42 and 43.

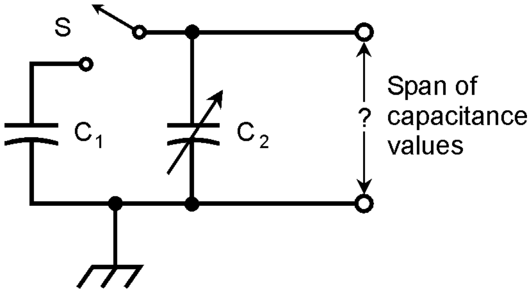

42. In Fig. Exam-2, suppose the value of the fixed capacitor, C_1, is 330 pF, and the range of the variable capacitor, C_2, is 10 to 365 pF. What is the span of capacitance values (minimum to maximum) that can be obtained with this parallel combination, with switch S either open or closed, as desired?

(a) 10 to 365 pF, over a continuous range

(b) 10 to 181 pF, but with a gap in the range

(c) 10 to 181 pF, over a continuous range

(d) 10 to 695 pF, but with a gap in the range
(e) 10 to 695 pF, over a continuous range

43. In Fig. Exam-2, suppose the value of the fixed capacitor, C_1, is 220 pF, and the range of the variable capacitor, C_2, is 5 to 100 pF. What is the span of capacitance values (minimum to maximum) that can be obtained with this parallel combination, with switch S either open or closed, as desired?

(a) 5 to 100 pF, over a continuous range
(b) 5 to 320 pF, but with a gap in the range
(c) 5 to 320 pF, over a continuous range
(d) 5 to 69 pF, but with a gap in the range
(e) 5 to 69 pF, over a continuous range

44. Which of the following units is most often used for specifying data speed in digital communications systems?

(a) Megabytes per second
(b) Megabauds per second
(c) Megabits per second
(d) Megahertz per second
(e) Any of the above units are used equally often when specifying data speed in digital communications systems.

45. Imagine four 100-μH inductors connected in a 2 × 2 series-parallel combination. Suppose there is no mutual inductance among them. What is the net inductance of this matrix?

(a) 25 μH
(b) 50 μH
(c) 100 μH
(d) 200 μH
(e) 400 μH

46. In the output of a dc power supply, voltage regulation can be obtained by

(a) connecting a Zener diode in series, paying attention to the polarity.
(b) connecting a choke in series and an electrolytic capacitor in parallel, paying attention to the polarity.
(c) connecting a capacitor in parallel and a Zener diode in series, paying attention to the polarity.
(d) connecting a Zener diode in parallel, paying attention to the polarity.
(e) connecting a resistor in parallel and another resistor in series.

47. Imagine an inductor and resistor connected in series, such that the inductive reactance and the resistance are both equal to 400 Ω at a frequency of 100 MHz. How are the instantaneous current and the instantaneous voltage related?

(a) They are in phase.
(b) They are 180° out of phase.

(c) The current leads the voltage by 90°.
(d) The current lags the voltage by 90°.
(e) None of the above are true.

48. Suppose you want to use a transformer to match the output of an audio amplifier to a speaker. The amplifier has a purely resistive output impedance of exactly 128 Ω. The speaker has a purely resistive impedance of exactly 8 Ω. The primary-to-secondary turns ratio of the transformer should be

(a) exactly 1:16.
(b) exactly 1:4.
(c) variable.
(d) exactly 4:1.
(e) exactly 16:1.

49. Imagine a string of 20 holiday lights, all connected in parallel and plugged into a 120-V rms ac wall outlet. If one of the bulbs is short-circuited, what will happen?

(a) All of the remaining 19 bulbs will shine a little more brightly than before.
(b) All of the remaining 19 bulbs will shine a little less brightly than before.
(c) The fuse or circuit breaker will blow, and all of the remaining 19 bulbs will go out.
(d) Most of all of the remaining 19 bulbs will burn out, and some may explode.
(e) The remaining 19 bulbs will continue to shine exactly as before—no more or less brightly.

50. Suppose a coil and capacitor are connected in series, with $jX_L = j40$ and $jX_C = -j70$, and the coil has an internal resistance of 10 Ω. Suppose the frequency of operation is 12.5 MHz. This circuit will exhibit resonance at

(a) some frequency below 12.5 MHz.
(b) some frequency above 12.5 MHz.
(c) 12.5 MHz; it is resonant under the conditions stated.
(d) any and all frequencies.
(e) no frequency, because of the internal resistance.

51. Imagine a transmission line consisting of two wires, each having a diameter of exactly 1 mm, and uniformly spaced exactly 10 mm apart with nothing but air as the dielectric between them. If the two wires are moved so they are uniformly spaced exactly 50 mm apart, and the dielectric between them is still nothing but air, what happens to the characteristic impedance?

(a) It does not change.
(b) It increases.
(c) It decreases.
(d) It is impossible to say without knowing what is connected to the line.
(e) It is impossible to say without knowing the frequency.

52. In an FM signal, deviation is

(a) the difference between the highest instantaneous carrier frequency and the lowest instantaneous carrier frequency.
(b) the maximum extent to which the instantaneous carrier frequency differs from the unmodulated-carrier frequency.

(c) the bandwidth of the signal, expressed in hertz, kilohertz, or megahertz.
(d) the rate at which the phase of the signal varies, expressed in hertz per second, kilohertz per millisecond, or megahertz per microsecond.
(e) None of the above is true.

53. A typical angular speed for an old-fashioned vinyl disk in a hi-fi sound system is
(a) 12 rpm.
(b) 24 rpm.
(c) 33 rpm.
(d) 60 rpm.
(e) 90 rpm.

54. An autotransformer can be recognized by the fact that it has
(a) a continuously adjustable turns ratio.
(b) a solenoidal core that can be moved in and out of a coil.
(c) a single, tapped winding.
(d) a special set of terminals for use with an automotive battery.
(e) a toroidal core.

55. One of the purposes of the suppressor grid in a pentode vacuum tube amplifier is to
(a) keep the cathode from emitting too many electrons, which could drive the control grid current to excessive levels.
(b) ensure that the plate is maintained at a constant negative voltage, so electrons will not be attracted to it.
(c) prevent the screen grid from shorting out to the plate, which could cause the power supply to burn out.
(d) repel secondary electrons emitted by the plate, keeping the screen grid current from getting too high.
(e) prevent distortion caused by overdrive in the control grid, screen grid, or cathode circuits.

56. Consider two sine waves having identical frequency. Suppose that one wave leads the other by exactly 1⁄20 of a cycle. What is the phase difference between the two waves?
(a) 9°
(b) 12°
(c) 18°
(d) 20°
(e) 24°

57. In a cellular communications system, a technology in which cells overlap, but signals do not interfere with each other because every phone set is assigned a unique signal code, is known as
(a) code-division multiple access.
(b) diversity reception.
(c) a phase-locked loop.

(d) analog signal mixing.
(e) digital diversity encryption.

58. A steady magnetic field can be produced by
    (a) a straight wire carrying a constant direct current.
    (b) a loop of wire carrying a constant direct current.
    (c) a coil of wire carrying a constant direct current.
    (d) a constant-intensity stream of protons in free space, moving in a straight line.
    (e) any of the above.

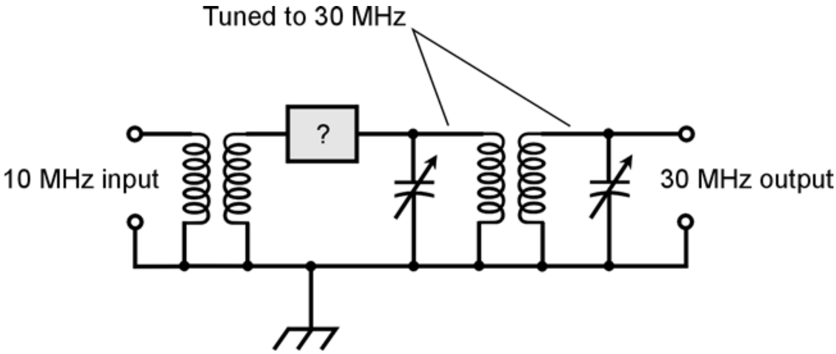

Exam-3 Illustration for Final Exam Question 59.

59. In Fig. Exam-3, what type of component can the box with the query symbol (?) contain in order for the circuit to function as indicated?
    (a) A capacitor
    (b) A resistor
    (c) A battery
    (d) An inductor
    (e) A diode

60. The binary number 1001101 represents which decimal number?
    (a) 44
    (b) 55
    (c) 66
    (d) 77
    (e) 88

61. Suppose a dc circuit has 10 kΩ of resistance, and a 12-V battery is connected to it. How much power is drawn from the battery?
    (a) 1.2 μW
    (b) 12 μW

(c) 1.2 mW
(d) 12 mW
(e) None of the above

62. In which of the following applications would you never find an op amp used as the main active circuit component?

(a) A differentiator
(b) An integrator
(c) A highpass audio filter
(d) A lowpass audio filter
(e) A microwave oscillator

63. Imagine a circuit with 21 Ω of inductive reactance and 28 Ω of resistance. What is the complex impedance of this circuit?

(a) 7 Ω
(b) 24 Ω
(c) 35 Ω
(d) 49 Ω
(e) None of the above

64. In a circuit containing conductance, inductive susceptance, and capacitive susceptance, a condition of resonance exists if and only if

(a) the conductance is zero.
(b) the conductance is infinite.
(c) the inductive susceptance is infinite and the capacitive susceptance is zero.
(d) the inductive and capacitive susceptances are both zero.
(e) the inductive and capacitive susceptances cancel out.

65. Imagine a perfect square wave with no dc component. That means the positive and negative portions have equal amplitudes and equal durations. Suppose you are told that the voltage is 10.0 V pk-pk. What is the average voltage?

(a) 0.00 V
(b) 5.00 V
(c) 7.07 V
(d) 10.0 V
(e) It is impossible to calculate this without more information.

66. Fill in the blank in the following statement to make it true: "It takes a certain minimum ____ voltage, known as the *forward breakover voltage*, for conduction to occur through a P-N junction."

(a) RF
(b) AF
(c) forward bias
(d) negative
(e) positive

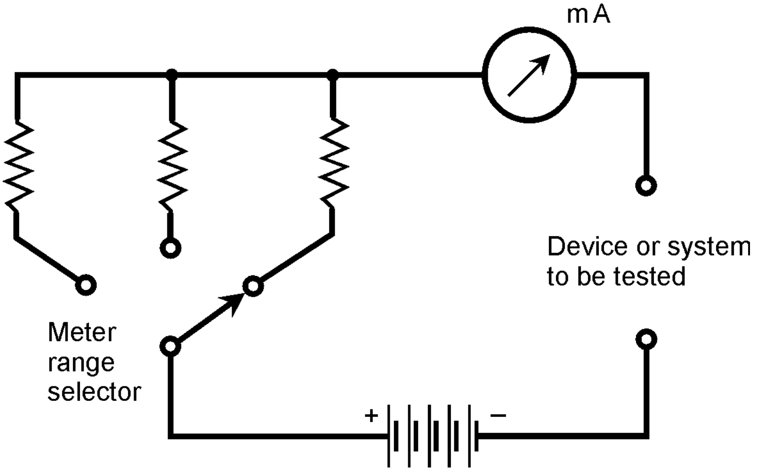

Exam-4 Illustration for Final Exam Question 67.

67. Figure Exam-4 is a generic diagram of a device that can be used to measure
   (a) dc voltage.
   (b) dc resistance.
   (c) dc power.
   (d) ac voltage.
   (e) ac power.

68. Suppose an audio amplifier produces 8.0 V rms output with 80 mV rms input. Suppose the amplifier input impedance is the same as the output load impedance. What is the gain of this amplifier in this situation?
   (a) 20 dB
   (b) 40 dB
   (c) 100 dB
   (d) 200 dB
   (e) We must know the actual input and output impedance values, in ohms, in order to determine the gain in this case.

69. The maximum gain obtainable with an amplifier that employs a given bipolar transistor
   (a) gradually increases as the frequency increases.
   (b) stays the same as the frequency increases.
   (c) gradually decreases as the frequency increases.
   (d) alternately increases and decreases as the frequency increases.
   (e) remains constant up to a certain frequency, and then abruptly drops to zero above that frequency.

70. A phase comparator would most likely be found in
   (a) an AF oscillator.
   (b) an AM detector.

(c) an RF power amplifier.
(d) a PLL frequency synthesizer.
(e) An antenna tuning network.

71. An atom with 7 protons and 5 electrons is an example of
(a) a positive isotope.
(b) a negative isotope.
(c) a positive ion.
(d) a negative ion.
(e) a neutral ion.

72. Electrical energy can be derived from hydrogen on a small scale in a device called
(a) an electrolytic cell.
(b) an alkaline cell.
(c) a fission cell.
(d) a fusion cell.
(e) a fuel cell.

73. The output wave of a common-gate amplifier circuit with a pure sine-wave input
(a) is in phase with the input wave.
(b) lags the input wave by 90° of phase.
(c) leads the input wave by 90° of phase.
(d) is 180° out of phase with the input wave.
(e) is inverted with respect to the input wave.

74. Which of the following capacitor types is polarized?
(a) Electrolytic
(b) Paper
(c) Ceramic
(d) Mica
(e) Air variable

75. The maximum radiation and response from a full-wave loop antenna, assuming there are no nearby conductors or obstructions to distort the pattern, occurs
(a) in the plane of the loop.
(b) perpendicular to the plane of the loop.
(c) at 45° angles to the plane of the loop.
(d) parallel to the horizon, in all directions of the compass.
(e) in all directions in three dimensions; it is a true omnidirectional antenna.

76. Suppose that five resistors, each having a different ohmic value from any of the others, are connected in series, and the entire combination is connected to a 12-V battery. How does the current through any one of the resistors compare with the current drawn from the battery?

(a) The current through any one of the resistors is more than the current drawn from the battery.

(b) The current through any one of the resistors is the same as the current drawn from the battery.

(c) The current through any one of the resistors is less than the current drawn from the battery.

(d) It depends on the relative ohmic values of the resistors.

(e) None of the above statements are true.

77. Suppose a battery-powered circuit has 1.00 Ω of net resistance, and 15.0 A of current flows through it. How much power is demanded from the battery?

(a) 225 W

(b) 150 W

(c) 22.5 W

(d) 66.7 mW

(e) It is impossible to calculate this, based on the information given.

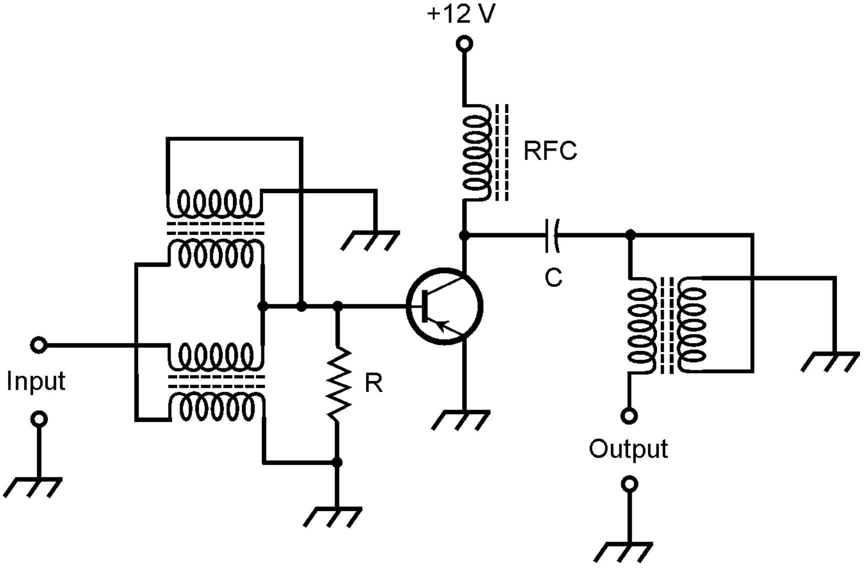

Exam-5 Illustration for Final Exam Question 78.

78. Figure Exam-5 shows a generic circuit that can be used for broadband RF power amplification. What is wrong with the circuit as shown?

(a) The resistor, R, should be replaced with a capacitor.

(b) The capacitor, C, should be replaced with a resistor.

(c) The +12-V power source should be replaced with a −12-V power source.

(d) The choke, RFC, should be replaced with a capacitor.
(e) Nothing is wrong with this circuit.

79. Imagine a string of 20 holiday lights, all connected in parallel and plugged into a 120-V rms ac wall outlet. If one of the bulbs is unscrewed and removed from the circuit, what will happen?
(a) All of the remaining 19 bulbs will shine a little more brightly than before.
(b) All of the remaining 19 bulbs will shine a little less brightly than before.
(c) The fuse or circuit breaker will blow, and all of the remaining 19 bulbs will go out.
(d) Most of all of the remaining 19 bulbs will burn out, and some may explode.
(e) The remaining 19 bulbs will continue to shine exactly as before—no more or less brightly.

80. When a volatile RAM chip is used in a computer,
(a) the contents of the memory can be easily overwritten.
(b) the contents of the memory remain intact even if power is removed.
(c) data in memory cannot be transferred to or from external media.
(d) the contents of memory never last more than a few minutes.
(e) excessive ELF radiation is produced.

81. A nickel-cadmium (NICAD) battery should never be discharged until it is totally dead because
(a) the battery may explode.
(b) the battery will have excessive voltage after it is recharged.
(c) the battery will develop unwanted inductance after it is recharged.
(d) the polarity of one or more of the cells may reverse, ruining the battery.
(e) Forget it! There is no problem with discharging a NICAD battery until it is totally dead.

82. Which of the following is an advantage of a CRT display over an LCD for a computing workstation?
(a) You can get a bigger screen for a lower price.
(b) The CRT emits less ELF energy.
(c) The CRT is less massive.
(d) The CRT takes up less space on the desk.
(e) All of the above are true.

83. The gauss is a unit of
(a) charge carrier flow speed.
(b) magnetic flux density.
(c) electrostatic field strength.
(d) electromagnetic field intensity.
(e) electrical charge quantity.

84. Imagine an inductor and resistor connected in series, such that the inductive reactance is 300 Ω and the resistance is 500 Ω at a particular frequency. If the frequency is decreased, what happens to the relative phase of the instantaneous current and the instantaneous voltage?

(a) Their relative phase does not change.
(b) They become more nearly in phase.
(c) They become more out of phase.
(d) They go alternately into and out of phase.
(e) It is impossible to say anything about this without more information.

85. Imagine two sine waves, one with a frequency of 50.00 Hz and the other with a frequency of 60.00 Hz. The phase difference between these two waves is

(a) not definable.
(b) 10.00°.
(c) 10.00%.
(d) 16.67%.
(e) $5/6$ of a radian.

86. Consider a variable capacitor connected in parallel with a fixed toroidal inductor. Suppose a signal of constant frequency is applied to the combination. If the variable capacitor is adjusted so its value changes from 200 pF to 100 pF, what happens to the reactance of the inductor?

(a) It doubles.
(b) It becomes half as great.
(c) It quadruples.
(d) It becomes $1/4$ as great.
(e) It does not change.

87. When two inductors having the same value are connected in series with a small opposing mutual inductance (say the coefficient of coupling is 5 percent or so), the total reactance of the combination at a specific, constant frequency is

(a) equal to the reactance of either inductor.
(b) slightly greater than the reactance of either inductor.
(c) slightly less than the reactance of either inductor.
(d) slightly greater than twice the reactance of either inductor.
(e) slightly less than twice the reactance of either inductor.

88. The power factor in an ac circuit is defined as

(a) the actual power divided by the maximum power the circuit can handle.
(b) the ratio of the real power to the imaginary power.
(c) the ratio of the apparent power to the true power.
(d) the ratio of the true power to the apparent power.
(e) the ratio of the imaginary power to the apparent power.

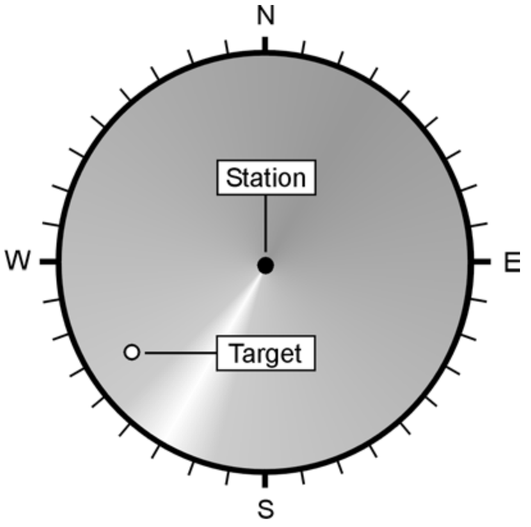

Exam-6 Illustration for Final Exam Question 89.

89. Figure Exam-6 shows a radar display and a single blip, indicating the position of a target. What is the approximate azimuth of this target?

(a) 122°
(b) 212°
(c) 238°
(d) 148°
(e) More information is needed to answer this.

90. Which of the following is an e-mail address in the proper format?

(a) www.sciencewriter.net
(b) members.authorsguild.net/stangib
(c) sciencewriter@tahoe.com
(d) http://google.com
(e) All of the above

91. A robot can be guided along a specific route by means of a current-carrying wire beneath the surface. The robot detects, and follows, the magnetic field produced by the current, thus following the wire. This is called

(a) an epipolar system.
(b) a fluxgate magnetometer.
(c) a servo system.
(d) an electromagnetic router.
(e) an embedded path.

92. Which of the following substances is sometimes used as the semiconductor material in junction field-effect transistors (JFETs)?
   - (a) Gallium arsenide
   - (b) Mica
   - (c) Glass
   - (d) Polystyrene
   - (e) Porcelain

93. Suppose that five resistors, each one having the same ohmic value as all the others, are connected in parallel, and the entire combination is connected to a 12-V battery. How does the current through any one of the resistors compare with the current drawn from the battery?
   - (a) The current through any one of the resistors is 5 times the current drawn from the battery.
   - (b) The current through any one of the resistors is the same as the current drawn from the battery.
   - (c) The current through any one of the resistors is $1/5$ as great as the current drawn from the battery.
   - (d) The current through any one of the resistors is $1/25$ as great as the current drawn from the battery.
   - (e) It is impossible to answer this without more information.

94. Figure Exam-7 illustrates a method of
   - (a) wireless-only encryption.
   - (b) end-to-end encryption.
   - (c) wireless tapping of a cell phone set.
   - (d) wireless tapping of a cordless phone set.
   - (e) dual-diversity shortwave reception.

95. In Fig. Exam-7, the box at the top contains a query symbol (?). What type of device should this be, in order for the system to perform its intended task?
   - (a) A low-frequency receiver
   - (b) A microwave transponder
   - (c) A video and data receiver
   - (d) A wireless transmitter
   - (e) A GPS receiver

96. Some substances cause magnetic lines of flux to bunch closer together than they would be if the magnetic field existed in a vacuum. This property is known as
   - (a) electromagnetism.
   - (b) diamagnetism.
   - (c) flux magnification.
   - (d) flux constriction.
   - (e) ferromagnetism.

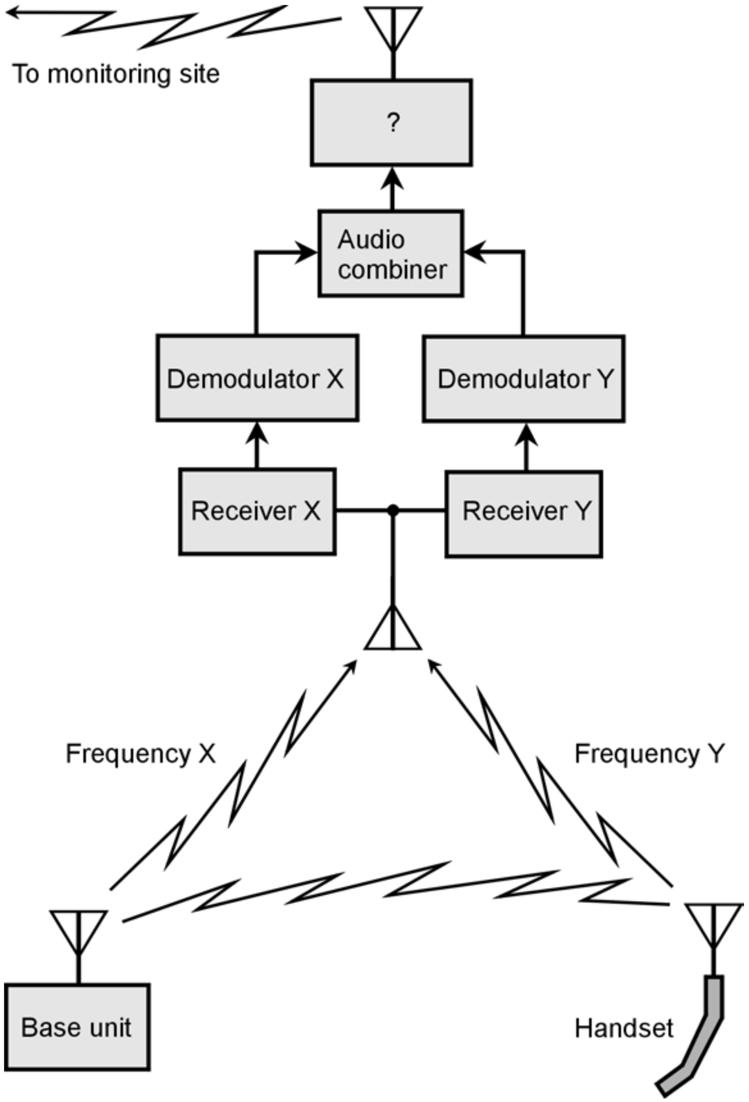
To monitoring site
?
Audio combiner
Demodulator X
Demodulator Y
Receiver X
Receiver Y
Frequency X
Frequency Y
Base unit
Handset

Exam-7 Illustration for Final Exam Questions 94 and 95.

97. In digital audio tape recording and reproduction, the noise can be practically eliminated because
    (a) noise is analog in nature, and can be nulled or filtered out without degrading the digital signals.
    (b) noise is digital in nature, and can be eliminated by means of specialized computer programs.
    (c) music and voices are digital in nature, and can be reproduced multiple times without distortion.
    (d) music and voices are analog in nature, while noise is digital and can therefore be canceled out.
    (e) Forget it! In a digital audio tape recording system, noise is a more serious problem than it is in an analog system.

98. Imagine a capacitor and resistor connected in series, such that the capacitive reactance is equal to –200 Ω and the resistance is equal to 200 Ω at a frequency of 70 MHz. How are the instantaneous current and the instantaneous voltage related?

    (a) They are in phase.
    (b) They are 90° out of phase.
    (c) The current leads the voltage by 45°.
    (d) The current lags the voltage by 45°.
    (e) None of the above are true.

99. The amount of current that a silicon photodiode can deliver in direct sunlight depends on

    (a) the forward breakover voltage.
    (b) the thickness of the substrate.
    (c) the surface area of the P-N junction.
    (d) the applied voltage.
    (e) the reverse bias.

100. Fill in the blank to make the following sentence true: "Electrical current can be expressed in terms of the number of ____ passing a given point per unit time."

    (a) neutrons
    (b) webers or gauss
    (c) charge carriers
    (d) wave cycles
    (e) isotopes

This page intentionally left blank

A
APPENDIX

Answers to Quiz, Test, and Exam Questions

Chapter 1

1. b	2. c	3. d	4. d	5. d
6. a	7. b	8. c	9. a	10. b
11. c	12. a	13. d	14. c	15. c
16. a	17. c	18. b	19. c	20. d

Chapter 2

1. b	2. c	3. d	4. b	5. c
6. a	7. c	8. c	9. a	10. d
11. d	12. a	13. b	14. b	15. c
16. d	17. b	18. c	19. a	20. c

Chapter 3

1. b	2. a	3. c	4. c	5. a
6. d	7. c	8. a	9. d	10. c
11. a	12. b	13. a	14. c	15. d
16. b	17. d	18. a	19. c	20. b

Chapter 4

1. a	2. c	3. d	4. a	5. b
6. d	7. c	8. a	9. c	10. a
11. d	12. b	13. d	14. b	15. a
16. c	17. c	18. a	19. d	20. b

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

Chapter 5

1. b 2. d 3. c 4. c 5. a
6. b 7. d 8. a 9. b 10. d
11. b 12. c 13. b 14. c 15. a
16. c 17. b 18. c 19. b 20. c

Chapter 6

1. c 2. a 3. a 4. c 5. d
6. b 7. b 8. c 9. b 10. c
11. d 12. c 13. a 14. c 15. a
16. a 17. b 18. c 19. b 20. a

Chapter 7

1. c 2. c 3. b 4. c 5. d
6. b 7. a 8. b 9. a 10. d
11. b 12. a 13. b 14. c 15. d
16. c 17. c 18. c 19. a 20. d

Chapter 8

1. c 2. a 3. d 4. b 5. c
6. b 7. b 8. c 9. a 10. c
11. c 12. d 13. b 14. a 15. d
16. d 17. c 18. b 19. b 20. a

Test: Part 1

1. a 2. a 3. c 4. b 5. a
6. b 7. a 8. e 9. e 10. c
11. c 12. c 13. b 14. c 15. d
16. b 17. a 18. a 19. e 20. e
21. d 22. d 23. d 24. c 25. a
26. c 27. d 28. a 29. c 30. c
31. c 32. d 33. e 34. d 35. e
36. b 37. d 38. d 39. c 40. a
41. a 42. d 43. c 44. e 45. c
46. c 47. d 48. a 49. e 50. c

Chapter 9

1. c	2. c	3. a	4. c	5. d
6. d	7. a	8. b	9. c	10. b
11. d	12. b	13. b	14. d	15. d
16. a	17. c	18. c	19. d	20. a

Chapter 10

1. b	2. a	3. d	4. b	5. b
6. c	7. d	8. c	9. a	10. d
11. b	12. d	13. a	14. a	15. d
16. b	17. a	18. c	19. a	20. d

Chapter 11

1. d	2. a	3. a	4. a	5. c
6. d	7. b	8. a	9. d	10. d
11. b	12. c	13. b	14. a	15. b
16. c	17. d	18. a	19. b	20. b

Chapter 12

1. c	2. a	3. b	4. b	5. d
6. b	7. d	8. c	9. a	10. b
11. b	12. d	13. c	14. a	15. c
16. b	17. d	18. c	19. d	20. c

Chapter 13

1. c	2. a	3. b	4. d	5. c
6. a	7. b	8. c	9. d	10. c
11. d	12. d	13. a	14. b	15. d
16. b	17. d	18. c	19. b	20. a

Chapter 14

1. b	2. d	3. a	4. c	5. b
6. a	7. b	8. a	9. a	10. d
11. c	12. b	13. b	14. b	15. c
16. b	17. c	18. d	19. a	20. c

Chapter 15

1. b 2. b 3. a 4. d 5. b
6. c 7. a 8. c 9. d 10. a
11. a 12. a 13. c 14. a 15. d
16. b 17. d 18. c 19. c 20. b

Chapter 16

1. b 2. a 3. d 4. d 5. d
6. b 7. c 8. c 9. a 10. c
11. b 12. b 13. a 14. b 15. b
16. d 17. c 18. a 19. b 20. a

Chapter 17

1. c 2. c 3. a 4. d 5. a
6. b 7. c 8. d 9. b 10. c
11. c 12. a 13. d 14. a 15. c
16. b 17. d 18. d 19. c 20. a

Chapter 18

1. c 2. a 3. d 4. b 5. b
6. a 7. c 8. d 9. a 10. d
11. b 12. b 13. c 14. d 15. c
16. a 17. c 18. d 19. b 20. b

Test: Part 2

1. e 2. e 3. a 4. a 5. e
6. d 7. c 8. d 9. a 10. e
11. c 12. a 13. d 14. c 15. c
16. e 17. e 18. d 19. c 20. b
21. b 22. d 23. c 24. c 25. e
26. c 27. b 28. a 29. e 30. c
31. a 32. a 33. a 34. a 35. b
36. e 37. c 38. d 39. c 40. b
41. b 42. c 43. c 44. c 45. d
46. d 47. e 48. e 49. a 50. a

Chapter 19

1. b 2. d 3. c 4. a 5. c
6. a 7. d 8. b 9. c 10. b
11. d 12. b 13. b 14. a 15. b
16. c 17. a 18. c 19. d 20. c

Chapter 20

1. b 2. d 3. c 4. b 5. a
6. a 7. d 8. c 9. b 10. d
11. b 12. a 13. c 14. a 15. c
16. d 17. b 18. a 19. c 20. a

Chapter 21

1. c 2. d 3. b 4. a 5. b
6. c 7. a 8. b 9. d 10. c
11. a 12. d 13. d 14. b 15. c
16. a 17. c 18. b 19. a 20. c

Chapter 22

1. d 2. a 3. b 4. d 5. d
6. c 7. b 8. c 9. b 10. d
11. c 12. a 13. c 14. d 15. a
16. d 17. b 18. c 19. b 20. b

Chapter 23

1. b 2. d 3. a 4. b 5. c
6. a 7. b 8. c 9. d 10. b
11. c 12. a 13. d 14. a 15. c
16. d 17. b 18. c 19. a 20. a

Chapter 24

1. a 2. b 3. b 4. c 5. d
6. c 7. d 8. c 9. d 10. b
11. c 12. b 13. d 14. c 15. c
16. d 17. b 18. a 19. b 20. c

Chapter 25

1. c 2. c 3. c 4. c 5. a
6. d 7. b 8. a 9. c 10. d
11. b 12. a 13. a 14. b 15. d
16. a 17. a 18. c 19. a 20. d

Chapter 26

1. b 2. a 3. b 4. c 5. c
6. a 7. d 8. d 9. a 10. d
11. c 12. a 13. a 14. d 15. a
16. d 17. c 18. d 19. b 20. c

Test: Part 3

1. e 2. b 3. a 4. c 5. a
6. c 7. a 8. e 9. e 10. d
11. d 12. e 13. c 14. c 15. d
16. c 17. c 18. e 19. a 20. d
21. c 22. e 23. a 24. c 25. d
26. a 27. b 28. e 29. a 30. b
31. c 32. c 33. b 34. b 35. a
36. d 37. a 38. b 39. b 40. e
41. c 42. a 43. d 44. b 45. a
46. b 47. b 48. c 49. e 50. b

Chapter 27

1. b 2. a 3. b 4. d 5. c
6. a 7. b 8. d 9. c 10. b
11. c 12. b 13. d 14. a 15. b
16. a 17. d 18. c 19. c 20. d

Chapter 28

1. c 2. c 3. a 4. d 5. b
6. d 7. c 8. a 9. d 10. c
11. a 12. c 13. b 14. a 15. c
16. a 17. c 18. d 19. d 20. b

Chapter 29

1. c
2. d
3. a
4. b
5. c
6. d
7. c
8. c
9. d
10. d
11. a
12. a
13. b
14. a
15. c
16. c
17. a
18. a
19. b
20. d

Chapter 30

1. c
2. a
3. a
4. c
5. b
6. d
7. b
8. a
9. d
10. d
11. d
12. a
13. b
14. c
15. a
16. a
17. d
18. b
19. c
20. a

Chapter 31

1. b
2. a
3. d
4. b
5. c
6. b
7. d
8. a
9. d
10. a
11. c
12. a
13. b
14. d
15. a
16. c
17. a
18. b
19. a
20. b

Chapter 32

1. c
2. c
3. d
4. c
5. b
6. a
7. d
8. d
9. a
10. c
11. b
12. b
13. d
14. d
15. a
16. b
17. a
18. c
19. a
20. c

Chapter 33

1. b
2. a
3. d
4. a
5. a
6. c
7. d
8. a
9. d
10. c
11. a
12. d
13. b
14. c
15. b
16. c
17. a
18. d
19. d
20. a

Chapter 34

1. b
2. a
3. d
4. d
5. c
6. b
7. b
8. b
9. d
10. c
11. c
12. a
13. a
14. b
15. c
16. b
17. d
18. c
19. b
20. b

Test: Part 4

1. d	2. e	3. c	4. d	5. b
6. d	7. a	8. a	9. a	10. c
11. c	12. b	13. e	14. e	15. e
16. c	17. e	18. b	19. a	20. d
21. c	22. c	23. e	24. e	25. e
26. c	27. c	28. c	29. a	30. d
31. b	32. c	33. d	34. c	35. e
36. a	37. a	38. a	39. a	40. b
41. b	42. d	43. e	44. b	45. b
46. e	47. d	48. d	49. d	50. d

Final Exam

1. e	2. b	3. c	4. b	5. e
6. a	7. d	8. e	9. b	10. a
11. b	12. b	13. a	14. d	15. c
16. a	17. a	18. a	19. c	20. b
21. a	22. d	23. e	24. b	25. d
26. a	27. c	28. e	29. e	30. b
31. e	32. c	33. d	34. b	35. d
36. a	37. a	38. b	39. d	40. e
41. c	42. e	43. b	44. c	45. c
46. d	47. e	48. d	49. c	50. b
51. b	52. b	53. c	54. c	55. d
56. c	57. a	58. e	59. e	60. d
61. e	62. e	63. e	64. e	65. a
66. c	67. b	68. b	69. c	70. d
71. c	72. e	73. a	74. a	75. b
76. b	77. a	78. c	79. e	80. a
81. d	82. a	83. b	84. b	85. a
86. e	87. e	88. d	89. c	90. c
91. e	92. a	93. c	94. d	95. d
96. e	97. a	98. c	99. c	100. c

B
APPENDIX
Schematic Symbols

ammeter	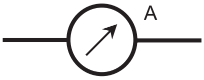
amplifier, general	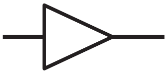
amplifier, inverting	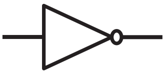

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

amplifier, operational

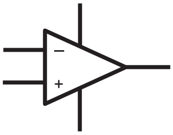

AND gate

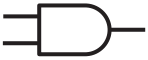

antenna, balanced

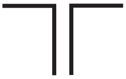

antenna, general

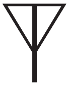

antenna, loop

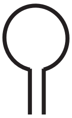

antenna, loop, multiturn

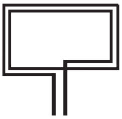

battery, electrochemical

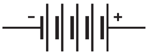

capacitor, feedthrough

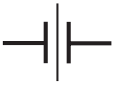

capacitor, fixed

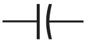

capacitor, variable

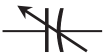

capacitor, variable, split-rotor

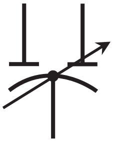

capacitor, variable, split-stator

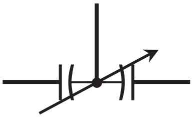

cathode, electron-tube, cold

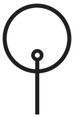

cathode, electron-tube, directly heated

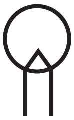

cathode, electron-tube, indirectly heated

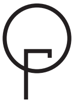

cavity resonator

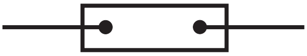

cell, electrochemical

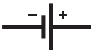

circuit breaker

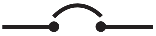

coaxial cable

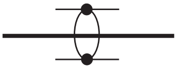

crystal, piezoelectric

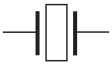

delay line

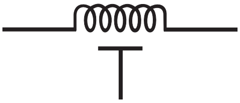

diac

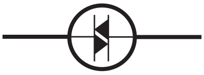

diode, field-effect

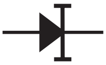

diode, general

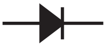

diode, Gunn

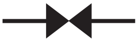

diode, light-emitting	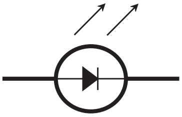
diode, photosensitive	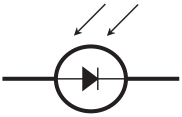
diode, PIN	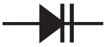
diode, Schottky	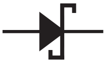
diode, tunnel	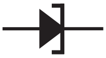
diode, varactor	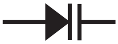
diode, Zener	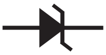
directional coupler	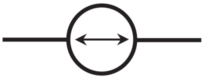
directional wattmeter	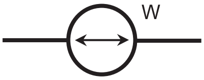
exclusive-OR gate	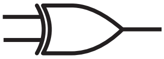
female contact, general	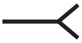
Ferrite bead	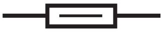
filament, electron-tube	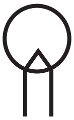
fuse	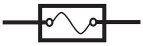

galvanometer	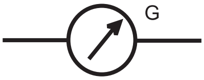
grid, electron-tube	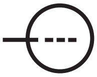
ground, chassis	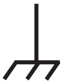
ground, earth	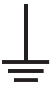
handset	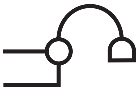
headset, double	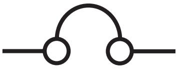
headset, single	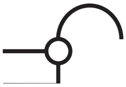
headset, stereo	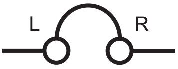
inductor, air core	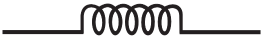
inductor, air core, bifilar	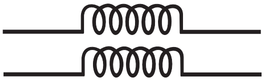
inductor, air core, tapped	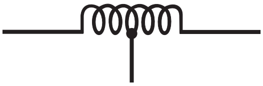
inductor, air core, variable	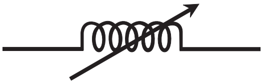
inductor, iron core	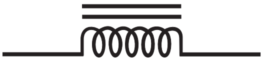

inductor, iron core, bifilar

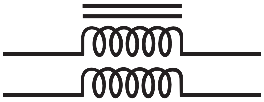

inductor, iron core, tapped

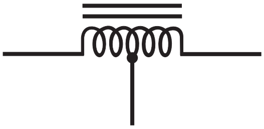

inductor, iron core, variable

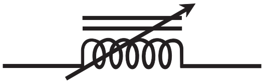

inductor, powdered-iron core

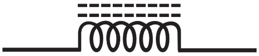

inductor, powdered-iron core, bifilar

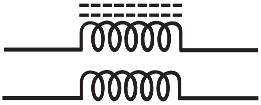

inductor, powdered-iron core, tapped

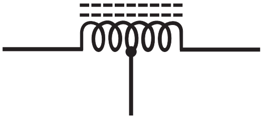

inductor, powdered-iron core, variable

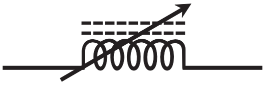

integrated circuit, general

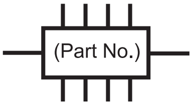

jack, coaxial or phono

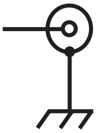

jack, phone, 2-conductor

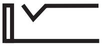

jack, phone, 3-conductor

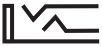

key, telegraph	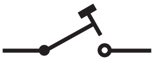
lamp, incandescent	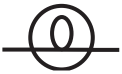
lamp, neon	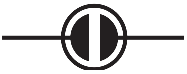
male contact, general	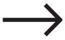
meter, general	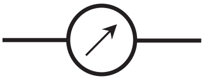
microammeter	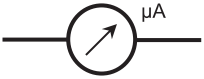
microphone	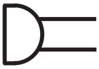
microphone, directional	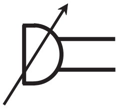
milliammeter	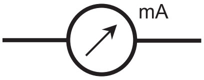
NAND gate	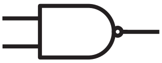
negative voltage connection	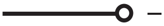
NOR gate	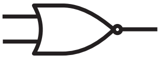
NOT gate	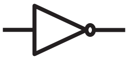
optoisolator	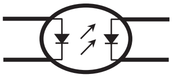

OR gate	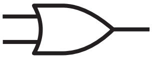
outlet, 2-wire, nonpolarized	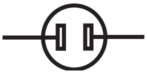
outlet, 2-wire, polarized	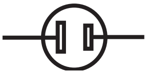
outlet, 3-wire	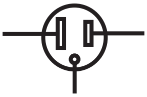
outlet, 234-volt	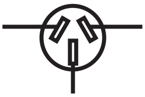
plate, electron-tube	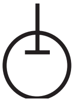
plug, 2-wire, nonpolarized	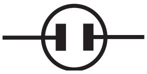
plug, 2-wire, polarized	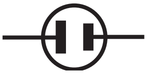
plug, 3-wire	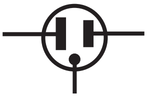
plug, 234-volt	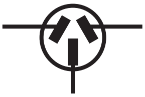
plug, coaxial or phono	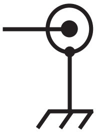

plug, phone, 2-conductor	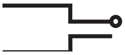
plug, phone, 3-conductor	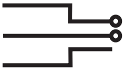
positive voltage connection	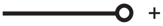
potentiometer	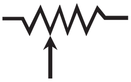
probe, radio-frequency	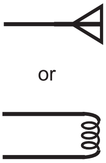
rectifier, gas-filled	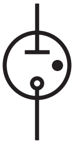
rectifier, high-vacuum	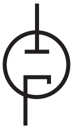
rectifier, semiconductor	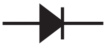
rectifier, silicon-controlled	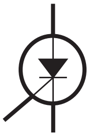
relay, double-pole, double-throw	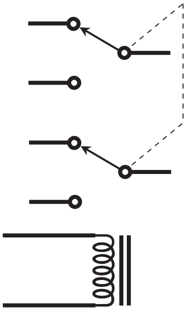

relay, double-pole, single-throw

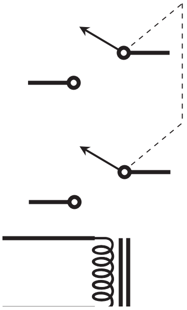

relay, single-pole, double-throw

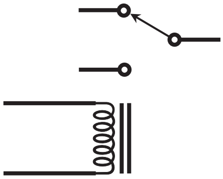

relay, single-pole, single-throw

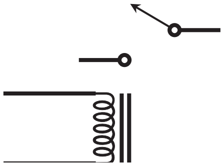

resistor, fixed

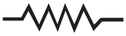

resistor, preset

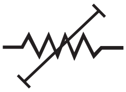

resistor, tapped

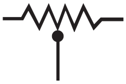

resonator

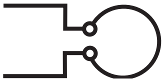

rheostat

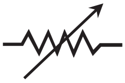

saturable reactor

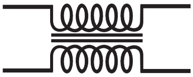

signal generator

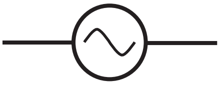

solar battery

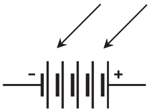

solar cell

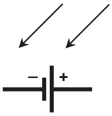

source, constant-current

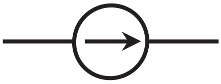

source, constant-voltage

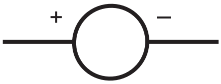

speaker

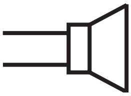

switch, double-pole, double-throw

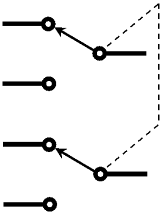

switch, double-pole, rotary

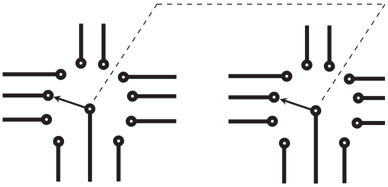

switch, double-pole, single-throw

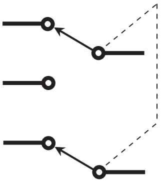

switch, momentary-contact

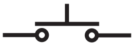

switch, silicon-controlled

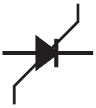

switch, single-pole, double-throw

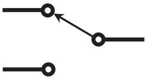

switch, single-pole, rotary

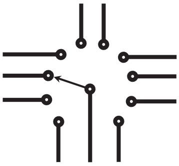

switch, single-pole, single-throw

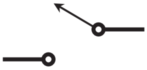

terminals, general, balanced

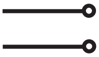

terminals, general, unbalanced

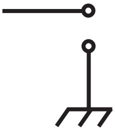

test point

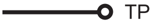

thermocouple

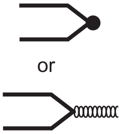

transformer, air core

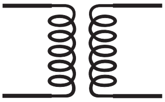

transformer, air core, step-down

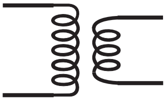

transformer, air core, step-up

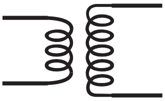

transformer, air core, tapped primary

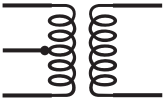

transformer, air core, tapped secondary

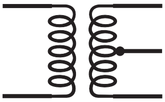

transformer, iron core

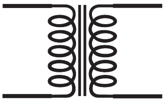

transformer, iron core, step-down

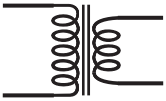

transformer, iron core, step-up

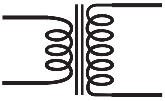

transformer, iron core, tapped primary

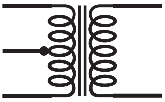

transformer, iron core, tapped secondary

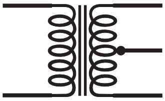

transformer, powdered-iron core

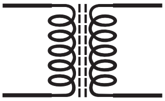

transformer, powdered-iron core, step-down

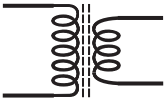

transformer, powdered-iron core, step-up

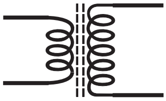

transformer, powdered-iron core, tapped primary

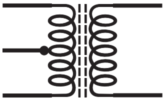

transformer, powdered-iron core, tapped secondary

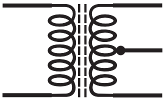

transistor, bipolar, NPN

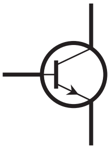

transistor, bipolar, PNP

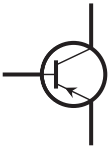

transistor, field-effect, N-channel

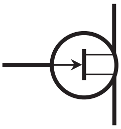

transistor, field-effect, P-channel

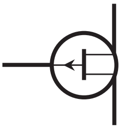

transistor, MOS field-effect, N-channel

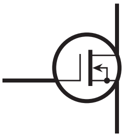

transistor, MOS field-effect, P-channel

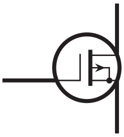

transistor, photosensitive, NPN

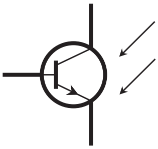

transistor, photosensitive, PNP

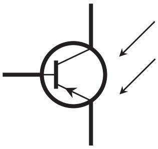

transistor, photosensitive, field-effect, N-channel

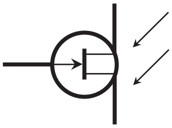

transistor, photosensitive, field-effect, P-channel

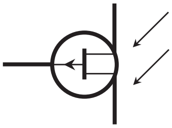

transistor, unijunction

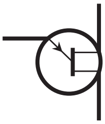

triac

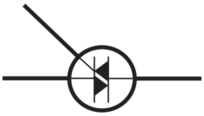

tube, diode

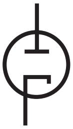

tube, heptode

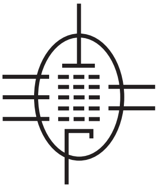

tube, hexode	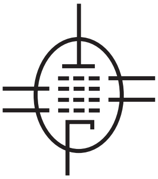
tube, pentode	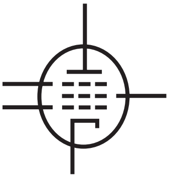
tube, photosensitive	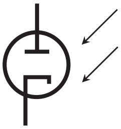
tube, tetrode	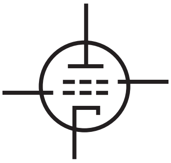
tube, triode	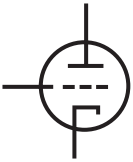
unspecified unit or component	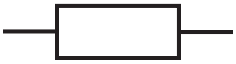
voltmeter	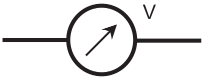
wattmeter	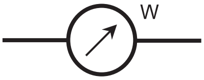
waveguide, circular	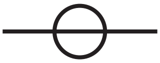
waveguide, flexible	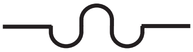

waveguide, rectangular

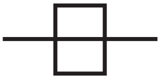

waveguide, twisted

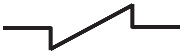

wires, crossing, connected

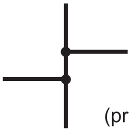

(preferred)

or

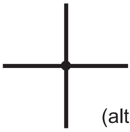

(alternative)

wires, crossing, not connected

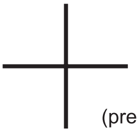

(preferred)

or

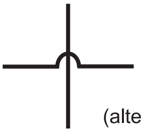

(alternative)

This page intentionally left blank

Suggested Additional Reading

- Brindley, K. *Starting Electronics,* 3d ed. Oxford, England: Newnes, 2004.
- Cutcher, D. *Electronic Circuits for the Evil Genius.* New York: McGraw-Hill, 2005.
- Gibilisco, S. *Electricity Demystified.* New York: McGraw-Hill, 2005.
- Gibilisco, S. *Electronics Demystified.* New York: McGraw-Hill, 2005.
- Horn, D. T. *Basic Electronics Theory with Projects and Experiments,* 4th ed. New York: McGraw-Hill, 1994.
- Long, L. *Personal Computing Demystified.* New York: McGraw-Hill, 2004.
- McComb, G. *Electronics for Dummies.* Hoboken, NJ: John Wiley & Sons, 2005.
- Scherz, P. *Practical Electronics for Inventors.* New York: McGraw-Hill, 2000.
- Slone, G. R. *TAB Electronics Guide to Understanding Electricity and Electronics,* 2d ed. New York: McGraw-Hill, 2000.
- Slone, G. R. *The Audiophile's Project Sourcebook.* New York: McGraw-Hill, 2002.
- Wise, E. *Robotics Demystified.* New York: McGraw-Hill, 2004.

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

This page intentionally left blank

Index

A

ac. *See* alternating current
acceptor impurity, 317–318
acoustic direction finder, 529–531
acoustic transducer, 597
acoustics, 538–540
ADC. *See* analog-to-digital converter
A/D. *See* analog-to-digital
admittance
  complex, 250–253, 255–256
  definition of, 240
  vector representation of, 241, 252–253
AGV. *See* automated guided vehicle
AI. *See* artificial intelligence
air-core coil, 57, 166
alkaline cell, 105–106
alpha cutoff frequency, 357–358
alternating current
  definition of, 27–28, 143
  electromagnet, 121–122
  source, schematic symbol, 57
AM. *See* amplitude modulation
amateur radio, 559–560, 580–581
American Radio Relay League, 560–561
American Red Cross, 30
American Wire Gauge (AWG), 20
ammeter, 39–41
ampere, 10–11, 19–20
ampere-hour, 103–104
ampere-turn, 32, 40, 120
amplification
  audio, 389–391
  factor, 357
  radio-frequency, 391–393
amplifier
  audio, 389–391, 426
  class A, 383
  class AB, 383–384
  class B, 384–386
  class C, 385
  current, 355–356
  distortion in, 388–389
  drive, 385, 388–389
  driving power, 385
  efficiency, 387–388
  flat topping in, 388–389
  hi-fi, 543
  linear, 385
  operational, 402
  overdrive, 388–389
  power, 384–388, 391–393
  push-pull, 385–386
  voltage, 369
  weak-signal, 391
amplitude
  effective, 153
  instantaneous, 151

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

amplitude (*Cont.*)
  limiting, 329
  modulation, 149–150, 409–410, 429–430
  negative peak, 152
  peak, 151–152
  peak-to-peak, 152
  positive peak, 152
  root-mean-square, 153
analog audio tape, 548
analog meter, 47
analog-to-digital
  conversion, 415–416, 449–450, 547
  converter, 523, 547, 580
AND
  gate, 445–446
  operation, 443
android, 605
angle of lag, 207–210
angle of lead, 221–224
angular displacement transducer, 519
angular frequency, 151
anode, 318–319, 505, 509–510
antenna
  billboard, 481
  broadside, 481–482
  coaxial, 475–476
  corner reflector, 485–486
  dipole, 281, 471–472
  directivity plots for, 478–480
  dish, 484–485
  doublet, 471–472
  driven element in, 480
  effective radiated power of, 478
  efficiency of, 471–472
  end-fire, 480–481
  feed point, 475
  folded dipole, 473–474
  forward gain for, 479–480
  front-to-back ratio for, 479–480
  front-to-side ratio for, 480
  ground systems for, 477–478
  ground-mounted vertical, 475
  ground-plane, 475
  half-wave, 473–474
  helical, 485
antenna (*Cont.*)
  horn, 485–486
  isotropic, 478
  longwire, 481
  loop, 476–477
  loopstick, 476
  parasitic, 482–483
  phased, 480–481
  power gain of, 478
  quad, 482–483
  reference, 478
  safety in working with, 486
  tuner, 299, 474
  vertical, 473–476
  Yagi, 482–483
  zepp, 473–474
antinode, acoustic, 540
apogee, 558
apparent power, 255–267
arctangent, 109
armature coil, 126
Armstrong oscillator, 394–395
ARRL. *See* American Radio Relay League
articulated geometry, 593–595
artificial intelligence, 605–606
ASCII, 446
Asimov, Isaac, 592
Asimov's three laws of robotics, 592
asymmetrical square wave, 145
atmospheric noise, 391
atom, 3
atom smasher, 27
atomic number, 3–4
atomic weight, 4
attenuator, 381, 546
audio
  amplifier, 389–391, 426
  attenuator, 546
  band-rejection filter, 433
  bandpass filter, 433
  filtering, 433
  mixer, 544–545
  notch filter, 433
  oscillator, 400–402
  splitter, 546
  tape, 548–549
  waveform, 400

audio-frequency-shift keying, 408
audio-taper potentiometer, 91–92
aurora borealis, 31, 422
auroral propagation, 422
automated guided vehicle, 601
automotive battery, 107
autotransformer, 292
avalanche, 321, 325, 328–330
average forward current, 338
AWG. *See* American Wire Gauge
axial leads, 96
azimuth, 527–528, 602

B

baby monitor, 587
background noise, 541
back-pressure sensor, 524–525, 603
baffle, 539–540
balance control, 543–544
balanced load, 295
balanced modulator, 411–412
balanced transmission line, 295
balun, 295
bandpass filter, 411, 433
band-rejection filter, 433
band-stop filter, 546
bar-graph meter, 50
base
  current, 353–354
  of transistor, 352
bass, 389, 538
battery
  automotive, 107
  defined, 14
  ideal, 104–105
  internal resistance, 104
  lantern, 106
  lead-acid, 107–108
  maximum deliverable current, 104
  mercuric-oxide, 106–107
  mercury, 106–107
  nickel-based, 108
  schematic symbol, 56
  shelf life, 104
  silver-oxide, 106
battery (*Cont.*)
  storage capacity, 103–104
  transistor, 106
baud, 580
Baudot, 446
beat frequency, 327, 425
beat-frequency oscillator, 425
bell ringer, 124
belt drive, 549
bias
  for amplification, 355–357
  in bipolar transistor, 353–357
  definition of, 85, 320, 330
  in field-effect transistor, 367–370
billboard antenna, 481
binary data, 446
binary number system, 442
binaural hearing, 597
binocular vision, 598–599
binomial, 231
biometric security measures, 565
bipolar transistor, 352–364
birdie, 426
bit, 448, 570
bits per second, 448–449, 570–571
black box, 57, 445
blackout, 293
blanking pulse, 417
bleeder resistor, 87–88
body capacitance, 523
boom box, 546–547
Boolean algebra, 443–444
bps. *See* bits per second
bridge
  impedance, 258
  rectifier circuit, 339–341, 347
brightness, in color image, 453
British thermal unit, 27
broadband
  power amplifier, 391–392
  transformer, 297
broadside antenna, 481–482
brownout, 293
bubble memory, 128
buffer, 450
button cell, 108
byte, 448, 570

C

cable
  coaxial, 171, 423
  data transmission, 423
  modem, 559
camera tube, 511–513
candela, 49
capacitance
  body, 523
  definition of, 175–176
  interelectrode, 84, 508
  junction, 320–321, 326
  mutual, 177
  in parallel, 178–179
  in series, 177–178
  units of, 177
capacitive coupling, 389
capacitive pressure sensor, 523
capacitive proximity sensor, 524–525
capacitive reactance
  definition of, 214
  frequency and, 216–217
  RC plane and, 218–219
capacitor
  alternating current and, 215–216
  ceramic, 180
  coaxial, 183–184
  dielectric in, 176–177
  electrolytic, 181, 342
  filter, 87–88
  fixed, 179–182
  integrated-circuit, 181–182
  leakage resistance of, 218
  mica, 180, 399
  padder, 279
  paper, 179
  plastic-film, 180–181
  polyethylene, 180
  polystyrene, 180, 399
  practical, 176–177
  schematic symbol, 57
  semiconductor, 181–182
  silver-mica, 399
  tantalum, 181–182
  temperature coefficient, 184
  tolerance, 184
capacitor (*Cont.*)
  trimmer, 183
  tubular, 180
  variable, 182–184
capacitor-input filter, 342–344
carbon-composition resistor, 88–89
carbon-film resistor, 89–90
carrier
  definition of, 407
  frequency, 149–150, 407
  power consumed by, 409–410
Cartesian coordinate geometry, 594–595
Cartesian coordinate plane, 231
Cartesian three-space, 595
cascade, 342
Cassegrain dish feed, 484–485
cat whisker, 326
catastrophic interference, 437
cathode, 318–319, 505–506
cathode-ray tube, 50, 416, 509–511
cavity resonator, 280
CCD. *See* charge-coupled device
CD. *See* compact disc
CDMA. *See* code-division multiple access
CD-R. *See* compact disc recordable
CD-RW. *See* compact disc rewritable
cell
  alkaline, 105–106
  button, 108
  in communications, 554–556
  cylindrical, 108
  dry, 14
  electrochemical, 56, 102–108
  flooded, 108
  fuel, 110–111
  ideal, 104–105
  internal resistance, 104
  lithium, 107
  maximum deliverable current, 104
  mercuric-oxide, 106–107
  mercury, 106–107
  nickel-based, 103, 108
  phone, 554–556, 564–565
  photovoltaic, 13, 48–49, 109–110
  primary, 102–103
  schematic symbol, 56

cell (*Cont.*)
  secondary, 102–103
  shelf life, 104
  silver-oxide, 106
  spacecraft, 108
  storage capacity, 103–104
  Weston standard, 103–104
  wet, 14
  zinc-carbon, 105
cellular communications, 554–556
central processing unit, 569–570
ceramic capacitor, 180
channel, in field-effect transistor, 365
character, in Morse code, 407–408
characteristic curve, 369–370
characteristic impedance, 236–238, 274–275, 297–299, 484
charge carrier, 10, 18, 317–318
charge-coupled device, 597
chassis ground, 57, 340
chemical energy, 14
chime, 124
choke-input filter, 342–344
chopping wheel, 523
circuit breaker, 347
circular polarization, 421
circular waveguide, 483–484
circular-motion model of sine wave, 190–192
Clapp oscillator, 396–397
class A amplifier, 383
class AB amplifier, 383–384
class B amplifier, 384–386
class C amplifier, 385
clinometer, 600–601
clipping, 329
clock, 447
cloning, cell phone, 565
closed-loop configuration, 493
cluster, in hard drive, 571–572
CMOS. *See* complementary metal-oxide-semiconductor logic
coaxial antenna, 475–476
coaxial cable, 171
coaxial capacitor, 183–184
coaxial transmission line, 170
code-division multiple access, 555
coefficient of coupling, 164
coherent communications, 435
coil
  air-core, 57
  iron-core, 57
cold cathode, 506
collector
  current, 354
  power input, dc, 386–387
  of transistor, 352
color
  code, resistor, 96–98
  sensing, in vision system, 599
Colpitts oscillator, 395–397
commercial-transaction security, 562
common base circuit, 359–360
common collector circuit, 359–361
common drain circuit, 375
common emitter circuit, 358–359, 382
common gate circuit, 374–375
common ground, 171
common source circuit, 373–374, 382–383
communications
  port, 581
  security, 561–565
commutator, 126
compact disc
  in audio recording, 547–548
  recordable, 128, 573
  rewritable, 128, 573
compact hi-fi system, 541–542
comparator, 496
complementary metal-oxide-semiconductor logic, 497–498
complex admittance, 250–253, 255–256
complex impedance, 229–238, 245–248, 255–256, 259
complex number
  absolute value of, 232–233
  addition, 230
  definition of, 229–230
  multiplication, 231
  plane, 231–232
  subtraction, 230
  vector, 232

complex wave, 147
component
  density, 317, 498
  hi-fi system, 542
composite, 151
compound, 6–7
compression, data, 450–451
computer
  central processing unit, 569–570
  diskette, 573
  display, 575–577
  external storage, 573
  flash memory, 573
  gaming, 576
  hard drive, 128, 571–572
  Internet and, 581–583
  logic board, 569
  map, 529
  microprocessor, 447, 569–570
  modem, 580–581
  motherboard, 569
  printer, 577–578
  random-access memory, 574–575
  scanner, 578–580
  tape drive, 573
conductance, 22–23, 238
conductance/inductance/capacitance circuit, 253–255, 277
conductance/susceptance plane, 240–241
conductivity, 22
conductor
  defined, 8–9
  schematic symbol, 55
conservation of current, 76
conservation of voltage, 77
continuous duty, 95
continuous wave, 430
contrast, 417
control grid, 507
conventional current, 72
conventional dish feed, 484–485
cordless phone, 564–565
core method, 289–290
corner reflector, 485–486
coulomb, 10–11
coulombs per second, 10–11
counter, frequency, 47, 448
counterpoise, 477
coupling
  capacitive, 389
  coefficient of, 164
  transformer, 389–391
  tuned-circuit, 390–391
CPU. *See* central processing unit
CRT. *See* cathode-ray tube
crystal-controlled oscillator, 398
crystal set, 326
current, 10–11, 18–19
current
  amplifier, 355–356
  average forward, 338
  calculations, 59
  conservation of, 76
  conventional, 72
  equalizing resistor, 346
  gain, 357, 380
  induced, 286
  lag, 205–210
  limiting, 85–86
  loop, 276
  node, 276
  surge, 342, 346
  theoretical, 72
cutoff, 384–385
CW. *See* continuous wave
cycles per second, 143
cylinder, in hard drive, 571–572
cylindrical cell, 108
cylindrical coordinate geometry, 595–596

D

D/A. *See* digital-to-analog
DAC. *See* digital-to-analog converter
D layer, 421
dark noise, 512
D'Arsonval movement, 39–40
dash, in Morse code, 407–408
data
  compression, 450–451
  speed, 570–571
dc. *See* direct current

dead zone, acoustic, 540
decay, in sawtooth wave, 146
decibel
  for current, 380
  gain, 93–94
  impedance and, 381
  loss, 93–94
  meter, 47–48
  for power, 380–381
  for sound, 93–94
  threshold of hearing and, 94
  for voltage, 380
decimal number system, 442
declining discharge curve, 106–107
decoder, 436
decryption, 563–564
decryption key, 563
deflecting coil, 510
deflecting plate, 511
degrees of freedom, 593–594
degrees of phase, 150
degrees of rotation, 593–594
demodulation
  in modem, 580
  in radio receiver, 326, 426, 429–432
demodulator, 326
demultiplexing, 496
dependent variable, 148
depletion
  mode, 372–373
  region, 320, 330, 367
depth, 595
derating function, 276
derivative of waveform 189
descrambling, 547
desensitization, 427
detection, 326, 429–432
detector, 326, 426, 429–432
diamagnetism, 122
dielectric, 9, 170, 176–177
dielectric constant, 177, 588
differentiator, 494–495
digital audio tape, 549
digital cable television, 434
digital encryption, 563–565
digital integrated circuit, 496–498
digital meter, 46–47
digital satellite television, 418
digital signal processing, 418, 435–436, 547–549
digital-to-analog
  conversion, 418, 449–450, 547
  converter, 547, 580
diode
  amplitude limiting, 329
  Esaki, 331
  IMPATT, 331
  infrared-emitting, 331
  light-emitting, 331
  PIN, 328
  photodiode, 519
  photosensitive, 332–333, 519
  rectifier, 325–326, 338
  semiconductor, 318–319
  signal mixing, 327
  switching, 328
  tube, 504–506
  tunnel, 331
  uses of, 325–336
  varactor, 321
  Zener, 328–329
dipole
  antenna, 281, 471–472
  magnetic, 119
direct current
  feedback, 399
  power input, 386–387
  pulsating, 28–30, 153–154
  superimposed, 153–154
direct drive, 549
direct-conversion receiver, 425, 432
direct-current electromagnet, 120–121
direction
  finder, acoustic, 529–531
  finding, radio, 476, 530–531
  of vector, 191
directivity plot, antenna, 478–480
directly heated cathode, 505–506
director, 482
discharge curve
  declining, 106–107
  flat, 104–105
discriminator, 432

dish antenna, 484–485
diskette, 573
displacement transducer, 519–523
display
  cathode-ray-tube, 50, 416, 509–511, 575
  interlacing, 576
  liquid crystal, 50, 331, 416, 511, 575–576
  plasma, 511, 576
  refresh rate, 576
  screen size, 576
dissector tube, 512–513
distortion, 357, 388–389
Dolby, 541
domain
  of function, 148
  name, in Internet, 582–583
  type, in Internet, 583
donor impurity, 317–318
dopant, 317
doping, 317
Doppler radar, 528
dot
  in Morse code, 407–408
  pitch, in video display, 576
dot-matrix printer, 578
dots per inch, 578–579
double-conversion receiver, 426
double sideband, 411
doublet antenna, 471–472
doubling, 548
drain
  of field-effect transistor, 365
  power input, dc, 386–387
drive, 87, 385, 388–389
driven element, in antenna, 480
driving power, 385
droop angle, of antenna radials, 475
dry cell, 14
DSB. *See* double sideband
DSP. *See* digital signal processing
dual-diversity reception, 434–435
dummy antenna, 87
dwell
  frequency, 437
  time, 437
dynamic
  current amplification, 356
  microphone, 517
  mutual conductance, 370
  range, 425, 541
  speaker, 517
  transducer, 517–519
dynode, 512

E

earth ground, 57
earth-moon-earth, 423
eavesdropping, 561
EB. *See* exabyte
ECL. *See* emitter-coupled logic
E core, 289–290
E layer, 422
echo, 527–528
eddy current, 288
edge
  detection, 601–602
  triggering, 447
Edison, Thomas, 155
EEPROM. *See* electrically erasable programmable read-only memory
effective amplitude, 153
effective power, 266
effective radiated power, 478
effective voltage, 27, 29
efficiency
  of antenna, 471–472
  of power amplifier, 387–388
EHF. *See* extremely high frequency
elastomer pressure sensor, 524
elbow, 596–597
electric eye, 590
electric field, 17–18, 419
electric generator, 13, 521–522
electric lines of flux, 17–18, 119
electric monopole, 119
electric motor, 13, 126, 520
electrical ground, 340, 477
electrically erasable programmable read-only memory, 499

electrochemical
  cell, 56, 102–108
  energy, 102–105
electrolyte, 181
electrolytic capacitor, 181, 342
electromagnet, 31, 40, 117, 120–122
electromagnetic cathode-ray tube, 510
electromagnetic deflection, 36–37
electromagnetic field, 419–420
electromagnetic induction, 286
electromagnetic interference, 477, 549–550
electromagnetic radiation, 144
electromagnetic spectrum, 420
electromagnetic wave polarization, 421
electromagnetic wave propagation, 421–423
electromagnetism, 13
electromotive force, 12–13, 41
electron
  definition of, 4–5
  gun, 509–510
  secondary, 512
  shell, 4–5
  volt, 27
electron tube
  anode in, 504, 509–510
  camera, 511–513
  cathode in, 504
  cathode-ray, 50, 416, 509–511
  control grid in, 507
  diode, 504–506
  dissector, 512–513
  dynode in, 512
  electrodes in, 505–509
  filament in, 505
  forms of, 504–505
  gas-filled, 504–505
  heptode, 507–508
  hexode, 507–508
  image dissector, 512–513
  image orthicon, 512
  interelectrode capacitance in, 508
  Klystron, 513–514
  magnetron, 513
  mercury-vapor, 506
  neon-bulb, 504–505
electron tube (*Cont.*)
  pentode, 507–508
  photomultiplier, 512–513
  plate in, 505–506
  size of, 315
  triode, 507
  vacuum, 504
  vidicon, 511
electronic bug, 589–590
electronic fingerprinting, 565
electronic mail, 582–583
electroscope, 38–39
electrostatic cathode-ray tube, 511
electrostatic charge, 12
electrostatic deflection, 38–39, 42
electrostatic emitter, 518
electrostatic meter, 38, 42
electrostatic pickup, 517–518
electrostatic transducer, 517–518
electrostatic voltmeter, 42
element, 3
elevation
  in robot arm, 595–596
  in spherical coordinates, 602
ELF. *See* extremely low frequency
elliptical orbit, 557
elliptical polarization, 421
embedded path, 601
EME. *See* earth-moon-earth
EMI. *See* electromagnetic interference
emitter
  follower, 360–361
  of transistor, 352
emitter-coupled logic, 496–497
encoder, 436
encryption, 563–565
end effector, 592
end-fire antenna, 480–481
end-to-end encryption, 563–564
energy
  chemical, 14, 102
  definition of, 25
  electrochemical, 102–105
  nonelectrical, 13–14
  units, 25–27
enhancement mode, 373

envelope detector, 429–430
epipolar navigation, 532–533
EPROM. *See* erasable programmable read-only memory
erasable programmable read-only memory, 499
erase head, 548
eraser-head pointer, 520
erg, 26–27
ERP. *See* effective radiated power
error voltage, 398–399
Esaki diode, 331
exabyte, 449, 570
exclusive OR gate, 445–446
expert system, 605–606
external modem, 581
external storage, 573
extremely high frequency, 424
extremely low frequency, 511, 577
eye-in-hand system, 599–600

F

F layer, 422
facsimile, 416
farad, 177
fast-scan television, 417–418, 433–434
fax, 416
feed point, of antenna, 475
feedback
  dc, 399
  negative, 393
  positive, 393–394
  at single frequency, 394
feedthrough scanner, 579
ferromagnetic core, 166–167, 288–289
ferromagnetic material, 31, 116–117, 122
ferromagnetism, 122
FET. *See* field-effect transistor
FET voltmeter, 44–45
fiber optics, 332, 424
field coil, 126
field-effect transistor, 44–45, 365–378
field strength
  magnetic, 120
  radio-frequency, 479–480
field-strength meter, 590
fifth-generation robot, 592
filament, 505
film-type resistor, 89–90
filter
  audio, 433
  bandpass, 411, 433
  band-rejection, 433
  capacitor, 87–88
  capacitor-input, 342–344
  choke, 169
  choke-input, 342–344
  notch, 433
  power-supply, 342–344
  section, 342–344
firmware, 499
first-generation robot, 591
fixed capacitor, 179–182
flare angle, 485
flash memory, 573
flat discharge curve, 104–105
flatbed scanner, 579
flat-topping, 388–389
flip-flop, 446–448
flooded cell, 108
flux density, 32
flux lines, 117–118
fluxgate magnetometer, 531–532
flying eyeball, 599–600
FM. *See* frequency modulation
folded dipole antenna, 471–472
foot-pound, 27
formatting, of hard drive, 571
forward
  bias, 319, 354–355
  breakover, 319, 325, 329
  gain, 479–480
  power, 267
fossil-fuel source, 111
four-channel stereo, 547
fourth-generation robot, 592
frame, 416–418
frequency, 28, 143–144
frequency
  alpha cutoff, 357–358
  angular, 151
  counter, 47, 448

frequency (*Cont.*)
  definition of, 28, 143–144
  domain, 148, 327
  drift, 399–400
  dwell, 437
  fundamental, 280
  harmonic, 280
  hopping, 437
  mark versus space, 430
  modulation, 411–414, 429, 432
  multiplication, 326–327, 413
  offset, 430–432
  spectrum, 148–150
  sweeping, 437
  synthesizer, 398–399
  versus wavelength, 419
frequency-division multiplexing, 436
frequency-shift keying, 408, 430–432
frequency-spreading function, 437
front end, 426–428
front-to-back ratio, 479–480
front-to-side ratio, 480
FSTV. *See* fast-scan television
fuel cell, 110–111
full-scale deflection, 40
full-wave rectification, 28–30, 339–343, 347
full-wave voltage doubler, 341
fundamental frequency, 148–149, 280
fuse
  quick-break, 347
  slow-blow, 347
fusion, 4

G

gain
  in decibels, 379–381
  definition of 93–94, 357
  forward, 479–480
  negative, 379
gain bandwidth product, 357–358
galena, 326
gallium arsenide, 316, 425
galvanism, 36
galvanometer, 36–37
gaming, 576
gas, 7–8
gas-filled tube, 504–505
gate
  of field-effect transistor, 365
  voltage, 365
gated flip-flop, 447
gauss, 32, 120
GB. *See* conductance/susceptance, gigabyte
gbps. *See* gigabit per second
generator
  defined, 154–155
  efficiency of, 155
  electric, 521–522
geographic poles, 115
geomagnetic declination, 116
geomagnetic equator, 116
geomagnetic field, 13, 115–117
geomagnetic inclination, 116
geomagnetic lines of flux, 115
geomagnetic poles, 115
geomagnetic storm, 117
geostationary satellite, 424, 556–557
germanium, 316
gigabit, 448
gigabit per second, 571
gigabyte, 449, 570
gigacycle, 143
gigahertz, 28, 143–144
gigawatt, 265
gilbert, 32, 120
GLC. *See* conductance/inductance/capacitance
Global Positioning System, 533–534
GPS. *See* Global Positioning System
graphic equalizer, 389, 545–546
grayscale, 416
gripper, 588–589
ground
  bus, 477–478
  chassis, 57, 340
  earth, 57, 345
  electrical, 340, 477
  loop, 477–478
  loss, 421
  for power supply, 345
  radial system, 475, 477
  radio-frequency, 477

ground (*Cont.*)
  signal, 358
  systems for antennas, 477–478
grounded-cathode circuit, 508–509
grounded-grid circuit, 508–509
ground-mounted vertical antenna, 475
ground-plane antenna, 475
Gunn diode, 330–331
Gunn effect, 330
Gunnplexer, 331

H

hacker, 562–563
half-wave
  antenna, 473–474
  rectification, 28–29, 325–326, 338–340, 342–343
ham radio, 559–560
handheld scanner, 579
handoff, 554
hand-print recognition, 565
hard drive, 128, 571–572
harmonic, 148–149, 280, 326–327, 539
Hartley oscillator, 394–395
HDTV. *See* high-definition television
headphone radio, 546
headset, 543
heatsink, 338
helical antenna, 485
helium, 4
henry, 161
heptode tube, 507–508
hertz, 27–28, 143–144
heterodyne, 327
hexadecimal number system, 443
hexode tube, 507–508
HF. *See* high frequency
high-definition television, 417–418
high frequency, 424
highpass response, 493–494
high-tension transmission line, 155, 293
hole, 10, 317–318
horizontal polarization, 421
horn antenna, 485–486
hot-wire meter, 39
hue, in color image, 453
hydrogen, 3–4, 110–111
hydrogen fuel cell, 110–111
hypertext, 583
hysteresis loss, 288–289

I

IC. *See* integrated circuit.
ideal battery, 104–105
ideal cell, 104–105
IF. *See* intermediate frequency
illumination meter, 48–49
image
  compression, 451
  dissector, 512–513
  orthicon, 512
  resolution, 513, 576, 578
  in superheterodyne receiver, 426, 428
  transmission, video, 416–418
imaginary number
  definition of, 229
  line, 229–230
  unit, 229
IMD. *See* intermodulation distortion
IMPATT diode, 331
impedance
  absolute value of, 235
  characteristic, 236–238, 274–275, 297–299, 484
  complex, 229–238, 248, 255–256, 259
  matching, 88, 238, 275, 294–299, 474
  measurement of, 258
  mismatch, 275
  nonreactive, 235
  purely resistive, 235, 259
  step-down transformer, 296
  step-up transformer, 296
  surge, 236
  transfer ratio, 296
  vector representation of, 234, 247–248
independent variable, 148
indirectly heated cathode, 505–506
induced current, 286
inductance
  definition of, 160–161
  mutual, 161–162, 164–166, 168
  in parallel, 163–164

inductance (*Cont.*)
  in series, 161–162
  stray, 171
  unit of, 161
  unwanted, 171
inductance/capacitance, 277, 523
inductive reactance
  definition of, 201
  frequency and, 202–203
  RL plane and, 203–205
inductor
  at audio frequencies, 169
  definition of, 161
  practical, 161
  at radio frequencies, 169–170
  roller, 396
  transmission lines as, 169–170
inference engine, 605–606
infrared
  definition of, 13
  intensity measurement, 49
  motion detector, 590–591
  transducer, 519
infrared-emitting diode, 331
inkjet printer, 577
instantaneous amplitude, 143, 151
instantaneous maximum, 27
instantaneous power, 266
instantaneous rate of change, 189
instantaneous voltage, 29, 189
insulated-gate field-effect transistor, 371
insulator, 8–9
integrated circuit
  advantages of, 491–492
  capacitor, 181–182
  comparator, 496
  complementary metal-oxide-semiconductor, 497–498
  demultiplexer, 496
  digital, 496–498
  emitter-coupled logic, 495–496
  limitations of, 492
  linear, 492–496
  memory, 499
  metal-oxide-semiconductor, 497–498
integrated circuit (*Cont.*)
  multiplexer, 496
  operational amplifier, 402, 493–495
  regulator, 344
  resistor, 90
  timer, 496
  transistor-transistor logic, 496–497
  voltage regulator, 495
integrator, 495
interactive photovoltaic system, 110
interelectrode capacitance, 184, 508
interference
  in communications, 437
  pattern, 513
interlacing, 576
intermediate frequency, 426, 428
intermodulation distortion, 427
internal modem, 581
internal resistance, 40, 41, 104
Internet
  address, 582
  cell phones and, 555–556
  domain name, 582–583
  domain type, 583
  electronic mail and, 582–583
  newsgroups, 583–584
  origin of, 581
  packets and, 581
  protocol, 581
  service provider, 581
  World Wide Web and, 583
interwinding capacitance, 289–290
intrinsic semiconductor, 328
inverse tangent, 209
inversion, logical, 444
inverter, logical, 444–446
ion, 6
ionization potential, 588
ionosphere, 421–422
ionospheric propagation, 421–422
interlacing, 418
IRED. *See* infrared-emitting diode
iris-print recognition, 565
iron-core coil, 56
irregular wave, 147–148
isotope, 4

isotropic antenna, 478
iterative process, 71

J
j operator, 204
JFET. *See* field-effect transistor
jitter, 418
J-K flip-flop, 447–448
joule, 25–27
joystick, 520
junction capacitance, 320–321, 326, 330

K
KB. *See* kilobyte
kbps. *See* kilobit per second
keying
  audio-frequency-shift, 408
  frequency-shift, 408
  on/off, 407–408
kilobaud, 580
kilobit, 448
kilobit per second, 571
kilobyte, 449, 570
kilocycle, 143
kilohertz, 28, 143–144
kilohm, 20
kilosiemens, 22
kilovolt, 17
kilowatt, 24, 265
kilowatt-hour, 25–27, 46
kilowatt-hour meter, 46
Kirchhoff's Current Law, 76
Kirchhoff's First Law, 75–77
Kirchhoff's Second Law, 77–78
Kirchhoff's Voltage Law, 77
Klystron, 513–514

L
lag, angle of, 207–210
lagging phase, 194
lamp, incandescent, 57
LAN. *See* local area network
lantern battery, 106
large-scale integration, 498
laser beam, 588–589
laser printer, 577
latency, 556–557
LC. *See* inductance/capacitance
LCD. *See* liquid-crystal display
lead-acid battery, 107–108
leading phase, 194
leakage resistance, 218
LED. *See* light-emitting diode
LEO. *See* low-earth-orbit satellite
LF. *See* low frequency
light meter, 48–49
light-emitting diode, 331
lightning, 12
limiter, 329, 432
linear amplifier, 385
linear displacement transducer, 519
linear function, 492–493
linear integrated circuit, 492–496
linearity, 541
linear-taper potentiometer, 90–91
lines of flux, 117–119
lines of force, 118
liquid, 7–8
liquid-crystal display, 50, 331, 416, 511
lithium battery, 107
lithium cell, 107
LO. *See* local oscillator
load, 237, 252, 295, 397, 400
loading control, 393
local area network, 558–559
local node, 452
local oscillator, 425
location system, 527–531
lock-in time, 47
lodestone, 115–116
logarithmic-taper potentiometer, 91–92
logic
  binary, 443–444
  board, 569
  emitter-coupled, 496–497
  gates, 444–446
  metal-oxide-semiconductor, 497–498

logic (*Cont.*)
  negative, 444–445
  positive, 444–445
  transistor-transistor, 496–497
log-log graph, 358
log-polar navigation, 532–533
log-taper potentiometer, 91–92
longwire antenna, 481
loop, 276
loop antenna, 476–477
loopstick antenna, 476
loran, 532
loss
  in decibels, 93–94, 379–381
  ohmic, 293
  resistance, 471–472
lossless image compression, 451
lossy image compression, 451
loudness, 540
loudness meter, 47–48
low frequency, 424
low-earth-orbit satellite, 424, 556–557
lower sideband, 409–411
lowpass response, 493–494
LSB. *See* lower sideband
LSI. *See* large-scale integration
lumen, 49

M

machine
  hearing, 597
  language, 575
  vision, 529, 597–600
magnet, permanent, 117, 124
magnetic compass, 155–116
magnetic dipole, 119
magnetic disk, 128
magnetic field, 30, 120, 419
magnetic field strength, 120
magnetic force, 117
magnetic flux density, 32
magnetic levitation, 117
magnetic lines of flux, 30–31, 117–118
magnetic pole, 31, 119
magnetic tape, 127
magnetic units, 32
magnetism, 30–32, 115–140
magnetomotive force, 32, 40
magnetron, 513
magnitude of vector, 191
majority carrier, 10, 318
make-before-break, 555
manipulator, 592
mark, 408, 430
master/slave flip-flop, 447
matrix, 64
maximum deliverable current, 104
maxwell, 32, 120
MB. *See* megabyte
mbps. *See* megabit per second
medium-earth-orbit satellite, 557–558
medium frequency, 424
medium-scale integration, 498
megabit, 448
megabit per second, 571
megabyte, 449, 570
megacycle, 143
megahertz, 28, 143–144
megasiemens, 22
megavolt, 17
megawatt, 24, 265
megohm, 20
memory
  backup, 499
  capacity, 575
  channels, in hi-fi tuner, 543
  drain, in nickel-based cell, 108
  flash, 573
  integrated-circuit, 499
  nonvolatile, 499
  random-access, 499, 574–575
  read-only, 499
  volatile, 499
MEO. *See* medium-earth-orbit satellite
mercuric-oxide battery, 106–107
mercuric-oxide cell, 106–107
mercury
  battery, 106–107
  cell, 106–107
mercury-vapor tube, 506
metal-film resistor, 89–90

metal oxide, 317
metal-oxide semiconductor, 370–373
metal-oxide-semiconductor field-effect transistor, 370–373
meteor scatter propagation, 422–423
meteor shower, 422
meter, schematic symbol, 56
meter shunt, 40–41
methane, in fuel cell, 111
methanol, in fuel cell, 111
MF. *See* medium frequency
mica capacitor, 180, 399
micro fuel cell, 111
microammeter, 41
microampere, 19
microfarad, 177
microhenry, 161
microprocessor, 447, 569–570
microsiemens, 22
microvolt, 17
microwatt, 24, 265
microwave band, 513
midrange
  sound, 538
  speaker, 543
military-level security, 562–563
military specifications, 562
milliampere, 19
millihenry, 161
milliohm, 22
millisiemens, 22
millivolt, 17
milliwatt, 24, 265
minority carrier, 10, 318
mixer
  audio, 544–545
  radio-frequency, 426
mixing product, 327
mobile hi-fi system, 546
mobile operation, in amateur radio, 559
modem, 408, 449, 555–556, 559, 580–581
modular construction, 492
modulation
  amplitude, 149–150, 409–410
  definition of, 407
modulation (*Cont.*)
  envelope, 266
  index, 414
  in modem, 580
  phase, 413–414
  pulse, 414–415
  reactance, 413
  waveform, 385
modulator, balanced, 411–412
modulator/demodulator. *See* modem
molecule, 7–8
monitoring system, 587–591
monopole, electric, 119
moonbounce, 423
Morse code, 385, 407–408, 446
MOS. *See* metal-oxide semiconductor
MOSFET. *See* metal-oxide-semiconductor field-effect transistor
motherboard, 569
motion detector
  infrared, 590–591
  ultrasonic, 591
motor, 13, 126, 520–521
motor/generator, 521–522
mouse, 520
M/S flip-flop, 447
MSI. *See* medium-scale integration
multimeter, 40–41, 44
multiplexing, 436, 496
multivibrator oscillator, 401–402
Murray code, 446
music synthesizer, 400
mutual inductance, 161–162, 164–166, 168

N

NAND gate, 445–446
nanoampere, 19
nanohenry, 161
nanowatt, 265
narrowband FM, 413
navigation, robot, 600–602
navigational methods, 531–534
negative feedback, 393
negative peak amplitude, 152
negative resistance, 233, 330

negative temperature coefficient, 96
network
  of computers, 449
  local area, 558–559
  wireless, 558–559
neutron, 3–4
newsgroups, Internet, 583–584
nickel-based battery, 108
nickel-based cell, 103, 108
node
  acoustic, 540
  in communications, 451–452
noise
  atmospheric, 391
  background, 541
  figure, 391, 425
  white, 391
nonelectrical energy, 13–14
nonlinear scale, 43
nonlinearity, 326–327, 357
nonreactive impedance, 235
nonvolatile memory 499
nonvolatility, 128, 499
NOR gate, 445–446
north geomagnetic pole, 115
north magnetic pole, 119
NOT
  gate, 445–446
  operation, 443
notation, 24–25
notch filter, 433
NPN transistor, 352
N-channel field-effect transistor, 365–366
N-type semiconductor, 10, 109–110, 317–318
N-type silicon, 109–110
nucleus, 3
number system
  binary, 442
  decimal, 442
  hexadecimal, 443
  octal, 443

O

OCR. *See* optical character recognition
octal number system, 443
octet, 448
ohm, 9, 20–22
ohm per foot, 20
ohm per kilometer, 20
ohm per meter, 20
ohmic loss, 293
ohmic value, 95
ohmmeter, 43–44
Ohm's Law
  for ac circuits, 259–261
  for dc circuits, 12–13, 58–61
on/off keying, 407–408
op amp, 402, 493–495
open-loop configuration, 493
operating point, 356
operational amplifier, 402, 493–495
optical character recognition, 579–580
optical encoder, 522–523
optical resolution, 598
optical sensitivity, 597–598
optoisolator, 332–333
OR
  gate, 445–446
  operation, 443
oscillator
  Armstrong, 394–395
  audio, 400–402
  beat-frequency, 425
  Clapp, 396–397
  Colpitts, 395–397
  crystal-controlled, 398
  definition of, 393–394
  Hartley, 394–395
  load impedance of, 397, 400
  local, 425
  multivibrator, 401–402
  reference, 398
  reliability of, 400
  stability of, 396–397, 399–400
  twin T, 401
  voltage-controlled, 330, 397
oscilloscope, 50, 327, 510
overdrive, 87, 356–357, 388–389
overloading, 427
overshoot, in meter, 39

oxygen, 7
ozone, 7

P

packet
  radio, 559
  wireless, 451–452
padder capacitor, 279
palette, 453
paper capacitor, 179
paraboloidal reflector, 484
parallel circuit
  power distribution in, 74–75
parallel data transmission, 450
parallel resistances, 63, 72–74
parallel resonance, 252
parallel-to-serial conversion, 450–451
parallel-wire transmission line, 169–170, 423
parasitic antenna, 482–483
parasitic element, 482
particle accelerator, 27
passband, 425
PB. *See* petabyte
PC card, 573
PCMCIA card, 573
P-channel field-effect transistor, 365–366
peak
  amplitude, 151–152
  inverse voltage, 338
  power, 266
peak-to-peak amplitude, 152
pen recorder, 49–50
pentode tube, 507–508
perigee, 557–558
period, 143–144, 146
periodic ac wave, 143
permanent magnet, 117, 124
permeability, 122–123, 166–167
permeability tuning, 167, 396
personal computer. *See* computer
personal identification number, 565
petabyte, 449, 570
phase
  angle, 192–195, 207–210, 221–224, 269–270
  coincidence, 192–193
phase (*Cont.*)
  comparator, 398
  degrees of, 150
  difference, 151, 192–195
  lagging, 194
  leading, 194
  modulation, 413–414
  opposition, 193
  quadrature, 193
  radians of, 151
  of sound wave, 540
phased antenna, 480–481
phase-locked loop, 398–399
phosphor, 509–510
photocathode, 512
photodetector, 416, 588–589
photodiode, 519
photoelectric proximity sensor, 525–526
photomultiplier, 512–513
photosensitive diode, 332–333, 519
photovoltaic cell, 13, 48–49, 109–110, 333, 519
photovoltaic system, interactive, 110
photovoltaic system, stand-alone, 109
picket fencing, 421
pickup head, 127
picofarad, 177
piezoelectric crystal, 280, 518–519
piezoelectric transducer, 518–519
PIN diode, 328
pinchoff, 367, 384–385
pip, 148
pitch
  of dots in video display, 576
  of motion in robot arm, 593
  of sound, 539
pixel, 576
plain old telephone service, 416
plasma display, 511, 576
plastic-film capacitor, 180–181
plate, of electron tube, 505–506
plate voltage, 509
platter, in hard drive, 571
playback
  head, 548
  mode, 548
PM. *See* pulse modulation

P-N junction, 318–321
PNP transistor, 352–353
polar orbit, 557
polarization, wave, 421
polarizer, 105
polyethylene capacitor, 180
polystyrene capacitor, 180, 399
portable hi-fi system, 546–547
portable operation, in amateur radio, 559
position sensor, 603
positive feedback, 393–394
positive peak amplitude, 152
positive temperature coefficient, 96
pot core, 168, 291–292, 401
potential, 13
potential difference, 13, 17, 21, 41
potentiometer
  audio-taper, 91–92
  linear-taper, 90–91
  logarithmic-taper, 91–92
  log-taper, 91–92
  schematic symbol, 55
  slide, 90
POTS. *See* plain old telephone service
power
  amplification, 384–388
  amplifier, 391–393
  apparent, 265–267
  calculations, 61–62
  defined, 23–24, 265
  dissipation, 86–87
  division of, 64
  effective, 266
  effective radiated, 478
  factor, 269–273
  forward, 267
  gain, 357, 380–381
  input, dc, 386–387
  instantaneous, 266
  output, 387
  peak, 266
  rating, of resistor, 95–96
  reactive, 267–273
  reflected, 267
  supply, 337–351
  transformer, 292–294, 337–338
power (*Cont.*)
  transmission, 273–276
  true, 267–275, 281
  units of, 265
  VA, 268–274
prefix multiplier, 24
preamplifier, 427
preselector, 425, 428
pressure sensor
  back, 524–525
  capacitive, 523
  elastomer, 524
primary cell, 102–103
primary colors, 599
primary winding, 286–288
printer
  dot-matrix, 578
  image resolution of, 578
  inkjet, 577
  laser, 577–578
  thermal, 578
product detector, 431–432
programmable multiplier/divider, 398
propagation, wave, 421–423
propane, in fuel cell, 111
proton, 3–4
proximity sensor
  capacitive, 524–525
  photoelectric, 525–526
P-type semiconductor, 10, 109–110, 317–318
P-type silicon, 109–110
pulsating direct current, 28
pulse modulation, 414–415
purely resistive impedance, 235, 259
push-pull amplifier, 385–386

Q

quad antenna, 482–483
quad stereo, 547
quadraphonic sound, 547
quality control, 588–589
quarter-wave vertical antenna, 474–476
quick-break fuse, 347
quiet room, 540

R

rack-mounted hi-fi system, 542
radar, 527–528
radial leads, 96
radial system, ground, 475, 477
radian, 151
radians of phase, 151
radiant-heat detector, 590
radiation resistance, 281, 471–472
radio direction finding, 476, 530–531
radio-frequency
  amplification, 391–393
  bands, 423–424
  field strength, 479–480
  ground, 477
  interference, 549–550
  spectrum, 420
  transformer, 297–299
radiolocation, 589
radioteletype, 408
RAM. *See* random-access memory
random-access memory, 499
range
  plotting, 602
  in radar, 528
  sensing, 602
  in spherical coordinates, 602
raster, 418
ratio detector, 432
RC. *See* resistance/capacitance
RDF. *See* radio direction finding
reach, 595–596
reactance
  capacitive, 214–228, 239
  inductive, 200–213, 239
  modulation, 413
  power consumption and, 268
reactive power, 267–273
read-only memory, 499
read-write
  head, in hard drive, 571
  memory, 499
receiver
  direct-conversion, 425, 432
  double-conversion, 426

receiver (*Cont.*)
  dynamic range, 425
  noise figure, 425
  selectivity, 425
  sensitivity, 425
  single-conversion, 426
  superheterodyne, 425–427
reception
  dual-diversity, 434–435
  wireless, 424–437
record mode, 548
recording head, 127, 548
rectangular coordinate geometry, 595
rectangular response, 428
rectangular wave, 145
rectangular waveguide, 483–484
rectification
  full-wave, 28–30, 339–343
  half-wave, 28–30, 325–326, 338–340, 342–343
rectifier
  diode, 325–326, 338
  full-wave bridge, 339–314
  full-wave center-tap, 339–340
  half-wave, 325–326, 338–340
  selenium, 338
  silicon, 338
red/green/blue, 453–454
reference antenna, 478
refresh rate, 576
reflected power, 267
reflector
  corner, 485–486
  in dish antenna, 484–485
  in helical antenna, 485
  paraboloidal, 484
  in parasitic antenna, 482
  spherical, 484–485
reference oscillator, 398
relay
  normally closed, 125
  normally open, 125
remanence, 122–123
repeater, 554–555
residual magnetism, 122–123

resistance
  calculations, 60–61
  definition of, 9, 20–22
  leakage, 218
  loss, 471–472
  negative, 233, 330
  parallel, 63, 72–74
  radiation, 281, 471–472
  series, 62, 69–71
  series-parallel, 64–65
resistance/capacitance
  circuit, 523
  phase angle, 221–224
  plane, 218–220
resistance/inductance
  phase angle, 207–210
  plane, 203–205
resistance/inductance/capacitance circuit, 248–250, 257–258, 277
resistivity, 20
resistor
  bleeder, 87–88
  carbon-composition, 88–89
  carbon-film, 89–90
  color code, 96–98
  coupling, 88
  current-equalizing, 346
  defined, 9, 21
  film-type, 89–90
  fixed, 88–90
  impedance-matching, 88
  integrated-circuit, 90
  metal-film, 89–90
  ohmic value of, 95
  power rating of, 95–96
  schematic symbol, 55
  temperature-compensated, 96
  tolerance, 95–96
  uses of, 85–88
  wirewound, 89
resolution
  in high-definition television, 418
  image, 513, 576
  in slow-scan television, 471
  optical, 598
  of printer, 578
resolution (*Cont.*)
  sampling, 415–416
  of scanner, 579
resonance
  definition of, 247, 252, 276
  parallel, 252, 277–278
  series, 247, 276–278
resonant devices, 280–282
resonant frequency, 276–278
resonant notch, 493–494
resonant peak, 493–494
response time, in meter, 39
retarding field, 513–514
retentivity, 122–123
retracing, 417
reverse bias, 320, 330, 354–355
revolute geometry, 596–597
RF. *See* radio-frequency
RFI. *See* radio-frequency interference
RGB. *See* red/green/blue
rheostat, 92–93
rim drive, 549
ringer, 124
ripple frequency, 340
rise, in sawtooth wave, 146
robot
  arm, 588–589, 592–597
  generations, 591–592
  hearing, 597
  leg, 603
  navigation, 600–602
  vision system, 597–600
robotics, Asimov's three laws of, 592
roll, of motion in robot arm, 593
roller inductor, 396
ROM. *See* read-only memory
root-mean-square amplitude, 153
rotating vector, 190–192
rotation, 596
rotor, in variable capacitor, 182
"rounding-off bug," 71
R-S flip-flop, 447–448
R-S-T flip-flop, 447–448
RTTY. *See* radioteletype
RX plane, 233–236

S

safety, in electrical work, 30
sampling
  rate, 416
  resolution, 415–416
satellite
  geostationary, 424, 556–557
  low-earth-orbit, 424, 556–557
  medium-earth-orbit, 557–558
  radio, 542
  systems, 424
saturation
  of bipolar transistor, 354–356
  in color image, 453
  of ferromagnetic core, 166
sawtooth wave, 146–147
scan
  converter, 434
  mode, 543
scanner
  basic features, 579
  configurations, 579–580
  optical character recognition and, 579–580
schematic diagram, 55–57
schematic symbols, 55–56
scrambling, 547
secondary cell, 102–103
secondary electron, 512
secondary winding, 286–288
second-generation robot, 592
sector, in hard drive, 571–572
security
  in communications, 561–565
  levels of, 561–563
seek mode, 543
selective squelching, 433
selectivity, of receiver, 425
selenium
  chemical element, 316
  rectifier, 338
selsyn, 521
semiconductor
  capacitor, 181–182
  definition of, 9–10
  diode, 318–319
  intrinsic, 328
semiconductor (*Cont.*)
  materials, 316–317
  N-type, 10, 109–110, 317–318
  P-type, 10, 109–110, 317–318
sensitivity
  optical, 597–598
  receiver, 425
sensor
  back-pressure, 524–525, 603
  capacitive, 523, 524–525
  definition of, 523
  elastomer, 524
  photoelectric, 525–526
  position, 603
  proximity, 524–526
  texture, 526–527
sequencing code, 437
sequential logic gate, 446
serial
  data port, 580
  data transmission, 450
serial-access storage, 573
serial-to-parallel conversion, 450–451
series circuit
  power distribution in, 74
series-parallel resistances, 64–65
series resistances, 62, 69–71
series resonance, 247
series-tuned Colpitts oscillator, 396–397
servo, 599
shape factor, 428
sheet scanner, 579
shelf life, 104
shell, electron, 4–5
shell method, 289–290
SHF. *See* superhigh frequency
shortwave
  band, 560
  listening, 560–561
shoulder, 596
shunt resistance, 40–41
sideband, 409–412
signal
  comparison, 529–531
  mixing, 327
signal-plus-noise-to-noise ratio, 425
signal-to-noise ratio, 425

significant figures, 60
silicon
  chemical element, 316
  rectifier, 338
  steel, 288
silver-mica capacitor, 399
silver-oxide battery, 106
silver-oxide cell, 106
sine function, 144
sine wave, 28, 144–145, 188–199, 400–401, 539
single-conversion receiver, 426
single sideband, 410–412, 432
sky wave, 421–422
slide potentiometer, 90
slope detection, 429, 432
slow-blow fuse, 347
slow-scan television, 416–417
small-scale integration, 498
smoke detector, 587–588
solar cell, 48–49, 333
solar flare, 117
solar panel, 108, 109, 333
solar sail, 115
solar wind, 115–116
solenoid, 92, 124, 168
solenoidal core, 290
solid, 7–8
sonar, 529–530
sound-level meter, 48
sound power, 94
source
  of field-effect transistor, 365
  follower, 375
south geomagnetic pole, 115
south magnetic pole, 119
space, 408, 430
spacecraft cell, 108
spark coil, 161
speaker, 543
spectral display, 409–410
spectrum analyzer, 327
spell checker, 580
spherical coordinates, 602
spherical reflector, 484–485
spread-spectrum communications, 437
square wave, 145
squelch, 433
SSB. *See* single sideband
SSI. *See* small-scale integration
SSTV. *See* slow-scan television
stand-alone photovoltaic system, 109
static current amplification, 356
static electricity, 11–12
static triggering, 447
stator, in variable capacitor, 182
step angle, 520
step-down transformer, 286
stepper motor, 520–521
step-up transformer, 286
stereoscopic vision, 598
storage capacity, 103–104
standing waves, 275–276
stray inductance, 171
streaming audio, 542
stroke, 12
subscriber, 452
substrate, 109–110, 365–366
superheterodyne receiver, 425–427
superhigh frequency, 424
superimposed dc, 153–154
supply reel, 548
surface wave, 421
surge
  current, 342, 346
  impedance, 236
  protector, 346
susceptance
  capacitive, 238–239
  definition of, 238
  inductive, 238–239
switching diode, 328
synchro, 521–522
synchronized communications, 435
synchronous flip-flop, 447–448

T

T flip-flop, 447
tachometer, 522
take-up reel, 548
tantalum capacitor, 181–182

tape
  analog audio, 548
  digital audio, 549
  drive, for computer, 573
  hiss, 541
  magnetic, 127
  player, 542
  recording, 127
TB. *See* terabyte
telechir, 603–605
telemetry, 603–605
teleoperation, 521
telephone modem, 580
telepresence, 603–605
television
  broadcast channels, 433
  digital cable, 434
  digital satellite, 418
  fast-scan, 417–418, 433–434
  high-definition, 417–418
  reception, 433–434
  slow-scan, 416–417
temperature coefficient
  of capacitor, 184
  negative, 96, 184
  positive, 96, 184
  zero, 184
temperature-compensated resistor, 96
temperature compensation, 96
terabyte, 449, 570
terahertz, 144
terminal node controller, 451–452
terminal unit, 408
terminals, schematic symbol, 57
tesla, 32, 120
tetrode tube, 507
texture sensor, 526–527
theoretical current, 72
thermal heating, 39
thermal printer, 578
thermocouple, 39
thermocouple principle, 39
third-generation robot, 592
three-wire ac system, 345
threshold detector, 319
threshold of hearing, 48, 94, 540
timbre, 539
time domain, 148, 194, 327
time-division multiplexing, 436
timer, 496
TNC. *See* terminal node controller
tolerance, 43, 95–96, 184
tone control, 389, 544
toroid, 92, 167–168, 290–291
toroidal core, 290–291, 402
touch pad, 520
track, in hard drive, 571–572
trackball, 520
tracking, in receiver tuning, 428
transceiver, 554–555
transconductance, 370
transducer
  acoustic, 597
  displacement, 519–523
  dynamic, 517–519
  electrostatic, 517–518
  infrared, 519
  piezoelectric, 518–519
  visible-light, 519
  wave, 517–519
transformer
  audio-frequency, 294
  balanced-to-unbalanced, 295
  balun, 295
  broadband, 297
  core method of winding, 289–290
  coupling, 295, 389–391
  E-core, 289–290
  ferromagnetic cores in, 288–289
  for impedance matching, 275, 294–299
  interwinding capacitance of, 289–290
  iron, 288
  pot-core, 291–292
  power, 292–294, 337–338
  primary winding of, 286–288
  principle of, 286–289
  radio-frequency, 297–299
  ratings, 337–338
  secondary winding of, 286–288
  shell method of winding, 289–290
  solenoidal-core, 290
  step-down, 286, 337
  step-up, 286, 337

transformer (*Cont.*)
  toroidal, 290–291
  transmission-line, 297–299
  turns ratio of, 287–288
  unbal, 295
  unbalanced-to-balanced, 295
  volt-ampere capacity of, 337–338
transient
  definition of, 346–347
  suppressor, 346
transistor
  amplifier, basic, bipolar, 381–382
  amplifier, basic, field-effect, 382–383
  base of, 353
  battery, 106
  biasing, 353–357, 367–370
  bipolar, 352–364, 425
  collector of, 352
  drain of, 365
  emitter of, 352
  field-effect, 365–378, 425
  gate of, 365
  NPN, 352
  PNP, 352–353
  power, 344–345
  size of, 315
  source of, 365
transistor-transistor logic, 496–497
transmatch, 299, 474
transmission line
  balanced, 295
  characteristic impedance of, 236–238
  coaxial, 170, 236–237, 423
  dielectric in, 170
  as inductor, 169–170
  mismatched, 275
  parallel-wire, 169–170, 236–237, 423
  power measurement in, 274
  section, 280
  standing waves in, 275–276
  transformer, 297–299
  unbalanced, 295
  velocity factor of, 280
  waveguide, 423, 483–484
transmission, wireless, 407–418
transponder, 589
treble, 389, 538
triangular wave, 146–147
trimmer capacitor, 183
triode tube, 507
tropospheric propagation, 422
true power, 267–275, 281
TTL. *See* transistor-transistor logic
TU. *See* terminal unit
tube. *See* electron tube
tubular capacitor, 180
tuned-circuit coupling, 390–391
tuned power amplifier, 391–393
tuner, 543
tuning control, 393
tunnel diode, 331
turns ratio, 287–288
turntable, 542
TV. *See* television
tweeter, 543
twin T oscillator, 401
twinlead, 295
two-wire ac system, 345

U
UHF. *See* ultrahigh frequency
ULSI. *See* ultra-large-scale integration
ultrahigh frequency, 417, 424, 513
ultra-large-scale integration, 498
ultrasonic motion detector, 591
ultraviolet, 49
unbal, 295
unbalanced load, 295
unbalanced transmission line 295
uninterruptible power supply, 107
Universal Serial Bus, 573
unwanted inductance, 171
upper sideband, 410–411
UPS. *See* uninterruptible power supply
USB. *See* upper sideband, Universal Serial Bus
user identification, 565
username, 582–583

V
VA power, 268–274
vacuum tube, 504
valve. *See* electron tube

Van de Graaff generator, 11
varactor, 321, 413
variable capacitor, 182–184
vector
  admittance, 241
  in complex-number plane, 232
  definition of, 191
  diagram, 195
  direction of, 191, 232
  impedance, 234
  magnitude of, 191, 232
  in RC plane, 219–220
  rendition of sine wave, 195
  in RL plane, 205
  rotating, 190–192
velocity factor, 170, 280, 419–420
vertical antenna, 473–476
vertical polarization, 421
very high frequency, 417, 424
very low frequency, 424
very-large-scale integration, 498
vestigial sideband, 417
vidicon, 511
VHF. *See* very high frequency
vinyl disk, 549
visible-light transducer, 519
vision system, 529, 597–600
VLF. *See* very low frequency
VLSI. *See* very-large-scale integration
voice-pattern recognition, 565
volatile memory, 499
volt, 17–18
voltage
  calculations, 60
  conservation of, 77
  divider network, 78–80
  doubler, 341
  gain, 357, 380
  instantaneous, 29, 189
  loop, 276
  negative-going, 188–189
  node, 276
  peak inverse, 338
  positive-going, 188–189
  regulation, 328–329, 341, 344–345, 495
voltage (*Cont.*)
  spike, 346
  transient, 346–347
  Zener, 319
voltage-controlled oscillator, 330, 397
volt-ampere, 45, 265–266, 268–273, 337–338
voltmeter, 41–42, 44–45
volt-ohm-milliammeter, 44
volume
  control, 389
  of sound, 540
  unit, 47–48
volume unit, 47–48
volume-unit meter, 47–48
VU. *See* volume unit

W

watt, 23–24, 265
watt rms, 47
watt-hour, 25–27, 46
watt-hour meter, 46
wattmeter, 45, 387
wave
  complex, 147–148
  composite, 151
  cosine, 189
  derivative of, 189
  irregular, 147–148
  polarization, 421
  propagation, 421–423
  rectangular, 145
  sawtooth, 146–147, 539
  sine, 28, 144–145, 188–199, 400–401, 539
  sky, 421–422
  square, 145, 539
  surface, 421
  transducer, 517–519
  triangular, 146–147, 539
waveform, 539
waveguide, 423, 483–484
wavelength versus frequency, 419
weak-signal amplifier, 391
weber, 32, 120
Weston standard cell, 103–104

wet cell, 14
white noise, 391
wideband FM, 413
wire-equivalent security, 561–562
wireless
  eavesdropping, 561
  network, 558–559
  router, 559
  tap, 561, 564–565
wireless-only encryption, 563–564
wiretapping, 561
wirewound resistor, 89
wiring diagram, 56–57
woofer, 543
word, in Morse code, 408
work envelope, 594
World Wide Web, 583
wrist, 596–597

X
XOR gate, 445–446

Y
Yagi antenna, 482–483
yaw, of motion in robot arm, 593

Z
Zener diode, 328–329, 344–345
Zener voltage, 329
zepp antenna, 473–474
zero beat, 425
zero bias, 353–354
zero temperature coefficient, 184
zinc-carbon cell, 105

This page intentionally left blank

About the Author

Stan Gibilisco is one of McGraw-Hill's most prolific and popular authors. His clear, reader-friendly writing style makes his books accessible to a wide audience, and his experience as an electronics engineer, researcher, and mathematician makes him an ideal editor for reference books and tutorials. Stan has authored several titles for the McGraw-Hill Demystified library of home-schooling and self-teaching volumes, along with more than 20 other books and dozens of magazine articles. His work has been published in several languages. Booklist named his *McGraw-Hill Encyclopedia of Personal Computing* one of the "Best References of 1996," and named his *Encyclopedia of Electronics* one of the "Best References of the 1980s."

Copyright © 2006, 2002, 1997, 1993 by The McGraw-Hill Companies, Inc. Click here for terms of use.

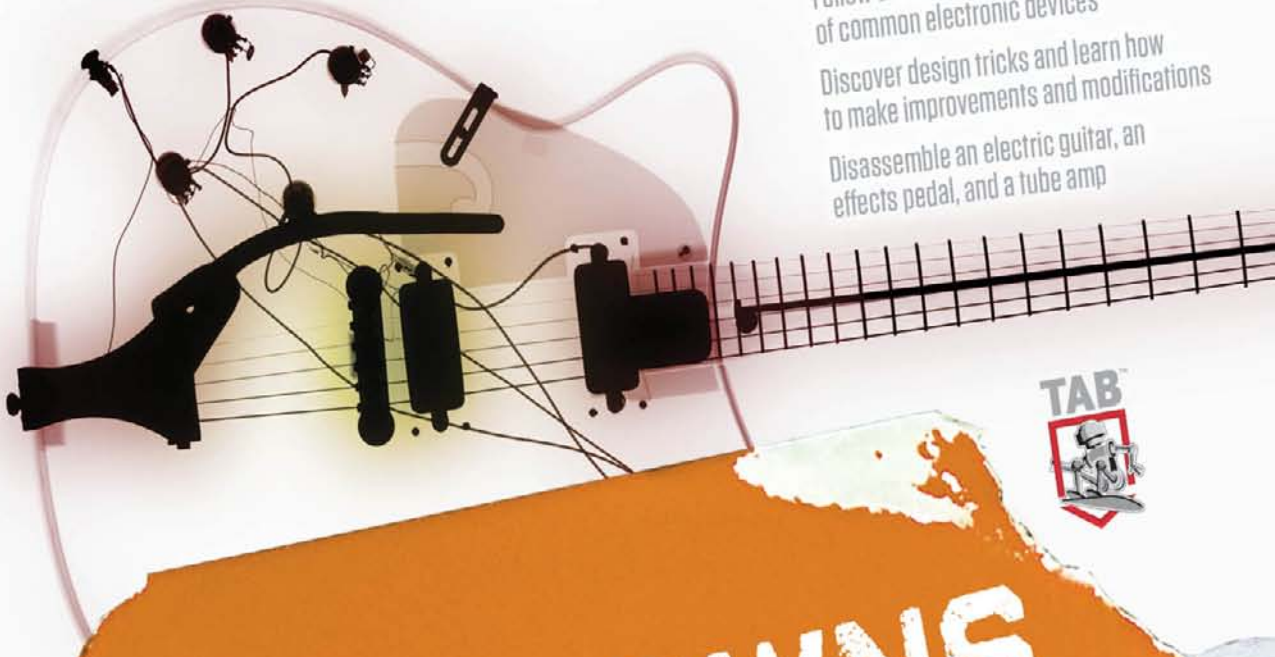

- Follow along with guided teardowns of common electronic devices
- Discover design tricks and learn how to make improvements and modifications
- Disassemble an electric guitar, an effects pedal, and a tube amp

TEARDOWNS

Learn How Electronics Work by Taking Them Apart

Bryan Bergeron

Foreword by Forrest M. Mims III

TEARDO N

About the Author

Bryan Bergeron (Brookline, Mass.) is editor of *Nuts & Volts Magazine* and *Servo Magazine* and the author of several hundred articles, two dozen books, and several patents. He teaches in the HST Division of Harvard Medical School and MIT. His company, Archetype Technologies, Inc., develops intelligent systems for the military.

TEARDO N

Learn How Electronics Work By Taking Them Apart

Bryan Bergeron

New York Chicago San Francisco Lisbon
London Madrid Mexico City Milan New Delhi
San Juan Seoul Singapore Sydney Toronto

McGraw Hill

The McGraw-Hill Companies

Copyright © 2010 by Bryan Bergeron. All rights reserved. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

ISBN: 978-0-07-171335-1

MHID: 0-07-171335-2

The material in this eBook also appears in the print version of this title: ISBN: 978-0-07-171334-4,
MHID: 0-07-171334-4.

All trademarks are trademarks of their respective owners. Rather than put a trademark symbol after every occurrence of a trademarked name, we use names in an editorial fashion only, and to the benefit of the trademark owner, with no intention of infringement of the trademark. Where such designations appear in this book, they have been printed with initial caps.

McGraw-Hill eBooks are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. To contact a representative please e-mail us at bulksales@mcgraw-hill.com.

Information has been obtained by McGraw-Hill from sources believed to be reliable. However, because of the possibility of human or mechanical error by our sources, McGraw-Hill, or others, McGraw-Hill does not guarantee the accuracy, adequacy, or completeness of any information and is not responsible for any errors or omissions or the results obtained from the use of such information.

TERMS OF USE

This is a copyrighted work and The McGraw-Hill Companies, Inc. (“McGrawHill”) and its licensors reserve all rights in and to the work. Use of this work is subject to these terms. Except as permitted under the Copyright Act of 1976 and the right to store and retrieve one copy of the work, you may not decompile, disassemble, reverse engineer, reproduce, modify, create derivative works based upon, transmit, distribute, disseminate, sell, publish or sublicense the work or any part of it without McGraw-Hill’s prior consent. You may use the work for your own noncommercial and personal use; any other use of the work is strictly prohibited. Your right to use the work may be terminated if you fail to comply with these terms.

THE WORK IS PROVIDED “AS IS.” McGRAW-HILL AND ITS LICENSORS MAKE NO GUARANTEES OR WARRANTIES AS TO THE ACCURACY, ADEQUACY OR COMPLETENESS OF OR RESULTS TO BE OBTAINED FROM USING THE WORK, INCLUDING ANY INFORMATION THAT CAN BE ACCESSED THROUGH THE WORK VIA HYPERLINK OR OTHERWISE, AND EXPRESSLY DISCLAIM ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. McGraw-Hill and its licensors do not warrant or guarantee that the functions contained in the work will meet your requirements or that its operation will be uninterrupted or error free. Neither McGraw-Hill nor its licensors shall be liable to you or anyone else for any inaccuracy, error or omission, regardless of cause, in the work or for any damages resulting therefrom. McGraw-Hill has no responsibility for the content of any information accessed through the work. Under no circumstances shall McGraw-Hill and/or its licensors be liable for any indirect, incidental, special, punitive, consequential or similar damages that result from the use of or inability to use the work, even if any of them has been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause arises in contract, tort or otherwise.

For Fred Marshall, my first mentor in the world of science.

This page intentionally left blank

Contents at a Glance

PART I **Around the Home**

1 Dual-sensor Smoke Alarm 3
2 Motion-activated LED Light 19
3 Digital Bathroom Scale 35
4 Surge Protective Devices 53
5 Electronic Pedometer 83
6 Compact Fluorescent Lamp 95
7 Ultrasonic Humidifier 113
8 Digital Hygro Thermometer 133
9 Stereo Power Amplifier 143

PART II **For Tinkerers**

10 Analog Volt-Ohm-Meter 177
11 Laser-guided Sonic Distance Measurer 203

PART III **For Musicians**

12 Electric Guitar 231
13 Effects Pedal 257
14 Vacuum Tube Guitar Amplifier 275

PART IV **Appendixes**

A Component Markings 295
B Resources 301

Index 305

This page intentionally left blank

Contents

Foreword. xix
Acknowledgments xxi
Introduction xxiii

PART I Around the Home

Chapter 1 Dual-sensor Smoke Alarm. 3

- Highlights 4
- Specifications 5
- Operation 5
- Teardown 5
  - Tools and Instruments 7
  - Step by Step 7
- Layout 9
- Components 9
  - Ionization-based Sensor 10
  - Ionization Sensor IC 12
  - Photoelectric Sensor 12
  - Photoelectric Sensor IC 13
  - Piezoelectric Transducer 14
  - C8050 Transistor 14
  - Diodes 15
  - Capacitors 16
  - Switches 16
- How It Works 16
- Mods 18
  - External Alarms 18
  - New Application Areas 18

Chapter 2 Motion-activated LED Light 19

- Highlights 20
- Specifications 20
- Operation 21

Teardown 21
    Tools and Instruments 21
    Step by Step 21
Layout 23
Components 24
    TL0001S (BIS0001) IC 24
    Passive Infrared Sensor 25
    Fresnel Lens 25
    CDS Sensor 26
    7130 Voltage Regulator 27
    NPN Transistor 27
    LEDs 29
    Capacitors 29
    Resistors 29
    Battery 29
How It Works 30
Mods 32
    Narrowed Trigger Zone 32
    External Device Controller 33
    Adjustable Light Sensitivity 33

Chapter 3 Digital Bathroom Scale 35
Highlights 36
Specifications 36
Operation 37
Teardown 37
    Tools and Instruments 37
    Step by Step 37
Layout 40
Components 40
    Microcontroller 41
    EEPROM 42
    Strain Gauges 42
    Piezoelectric Transducer 44
    J3Y Transistor 44
    LCD 45
How It Works 46
Mods 50
    Enhanced Triggering 50
    Create External Load Cells 50

Chapter 4 Surge Protective Devices 53
Surge Suppressors 54
Highlights 54
Specifications 55
    Joule Rating 55

  Response Time 56
  Transient Suppression Voltage 56
  Suppression LG/LN/GN 56
  UL 1449 57
Operation 57
Teardown 57
  Tools and Instruments 59
  Step by Step 59
Layout 61
Components 62
  Thermal Circuit Breaker 62
  MOVs 62
  Thermal Fuses 64
  Rocker Switch Neon Lamp 65
  Transistor 65
  Diodes 65
  Capacitor 65
  Resistors 66
  LED 66
  Circuit Board 66
  Cord 66
How It Works 66
Mods 69
Power Conditioner 69
Highlights 70
Specifications 70
Operation 71
Mini-Teardown 72
  Tools and Instruments 72
  Step by Step 73
Layout 74
Components 75
  MOVs 75
  Thermal Fuse 76
  Thermal Circuit Breaker 76
  Inductors 76
  LEDs 76
  Rocker Switch 77
  Diodes 77
  Capacitors 77
  Relay 77
  Circuit Board 77
How It Works 77
Mods 80
  TMOV Upgrade 80

Add Neutral and Ground MOVs . . . 80
Component Deletions . . . 80
Comparison . . . 80

Chapter 5 Electronic Pedometer . . . 83
Highlights . . . 84
Specifications . . . 84
Operation . . . 84
Teardown . . . 85
Tools and Instruments . . . 85
Step by Step . . . 86
Layout . . . 87
Components . . . 87
Pendulum Sensors . . . 87
Microcontroller . . . 88
Microcontroller Support ICs . . . 89
Crystal Oscillator . . . 89
SMT Potentiometers . . . 90
LCD Panel . . . 90
Silicone-Carbon Elastomeric Buttons . . . 90
How It Works . . . 90
Alternative Technologies . . . 92
Mods . . . 94
Add a Third Axis . . . 94

Chapter 6 Compact Fluorescent Lamp . . . 95
Highlights . . . 96
Specifications . . . 96
Operation . . . 99
Teardown . . . 99
Tools and Instruments . . . 100
Step by Step . . . 101
Layout . . . 102
Components . . . 102
Fluorescent Lamp . . . 103
Transistors and Heat Sinks . . . 103
Toroidal Transformer . . . 104
Inductors . . . 105
Capacitors . . . 105
Resistors . . . 105
Diodes . . . 106
How It Works . . . 107
Mods . . . 110

Chapter 7 Ultrasonic Humidifier . . . 113
Highlights . . . 113
Specifications . . . 113

Operation 114
Teardown 115
  Tools and Instruments 117
  Step by Step 117
Layout 119
Components 121
  Capacitors 121
  Piezoelectric Transducer 121
  Diodes 121
  Bicolor Red/Green LED 121
  Magnetic Reed Switch 122
  SPST Rocker Switch 122
  RF Inductors 122
  Heat Sink 122
  Resistors 123
  Transistors 123
  Fuses 124
  Power Transformer 124
  Brushless DC Motor and Fan 125
How It Works 126
Mods 130
  Closed Loop System 130

Chapter 8 **Digital Hygro Thermometer** **133**
Highlights 133
Specifications 134
Operation 135
Teardown 135
  Tools and Instruments 135
  Step by Step 135
Layout 136
Components 137
  Thermistor 137
  Humidity Sensor 138
  Microcontroller 139
  Crystal Oscillator 139
  LCD Panel 139
How It Works 139
Mods 141
  Adding Calibration Capabilities 141

Chapter 9 **Stereo Power Amplifier** **143**
Highlights 143
Specifications 144
  Frequency Response 145
  SNR 145

THD+N 146
Power Output 147
Power Consumption 148
Auto-On 148
Auto-Input 149
Form Factor 149
Operation 149
Teardown 149
Tools and Instruments 149
Step by Step 151
Layout 153
Components 158
TDA7294 Power ICs 158
Extruded Aluminum Heat Sink 159
5532D and 4558D Operational Amplifiers 162
C2235 NPN Transistor 162
Toroidal Transformer 162
Diodes 162
Capacitors 163
Resistors 163
Stacked Potentiometers 164
Relay 164
Fuse 165
How It Works 165
Mods 172
Disconnect the Switching Circuitry 172
Create a Beefier Power Supply 172

PART II For Tinkerers

Chapter 10 Analog Volt-Ohm-Meter 177
Highlights 177
Specifications 178
Significance 179
Interpretation 179
Operation 180
Teardown 181
Tools and Instruments 181
Step by Step 181
Layout 183
Components 184
Precision Resistors 184
Galvanometer 186
Diodes 186
Rotary Switch 188
Battery 190
MOVs 190

Fuse 190
Electromagnetic Buzzer 190
How It Works 192
DC Voltmeter 192
Battery Tester 194
AC Voltmeter 194
Ammeter 196
Ohmmeter 197
Audible Continuity Tester 199
Mods 200

Chapter 11 Laser-guided Sonic Distance Measurer 203
Highlights 204
Specifications 205
Range 205
Resolution 206
Accuracy 207
Operating Temperature 207
Humidity 207
Aperture Angle 207
Laser Specifications 208
Operation 208
Teardown 209
Tools and Instruments 210
Step by Step 210
Layout 211
Components 213
Ultrasonic Transducer and Horn 213
Thermistor 213
Laser Module 214
LCD Module 214
Microcontroller 215
7130 Positive Fixed Voltage Regulator 215
LM317LM Positive Adjustable Voltage Regulator 216
ST274C Quad CMOS Operational Amplifier 216
Transistors 217
Diodes 217
Crystal 217
Inductors 218
Switches 219
Capacitors 219
Resistors 219
How It Works 220
Operational View 220
Functional View 222
Mods 226

PART III For Musicians

Chapter 12 **Electric Guitar** **231**
Highlights 231
Specifications 232
Operation 234
Teardown 235
    Tools and Instruments 235
    Step by Step 237
Layout 238
Components 239
    EMI Shield 240
    Wiring 240
    Potentiometers 241
    Capacitor 241
    Audio Jack 241
    Switches 242
    Passive Magnetic Pickups 242
Magnetic Pickup Mini-Teardown 243
    Step by Step 244
How It Works 247
Mods 250
    More Pickup Switching Options 250
    Compensated Volume Control 250
    Functional Tone Control 250
    Pickup Variations 251
    A MIDI Interface 253

Chapter 13 **Effects Pedal** **257**
Highlights 258
Specifications 258
Operation 259
Teardown 260
    Tools and Instruments 261
    Step by Step 261
Layout 262
Components 263
    Diodes 263
    Circuit Board and Enclosure 265
    Resistors 265
    Operational Amplifier 265
    Capacitors 268
    Connectors 269
How It Works 269
    Gain 269
    Clipping 270
    Impedance Matching 272

Mods 272
    Change Diode Configuration and Type 273
    Tone Switch 273
    Add True Bypass Switching 274
    Try Different Operational Amplifiers 274

Chapter 14 Vacuum Tube Guitar Amplifier 275
Highlights 276
Specifications 276
Operation 276
Teardown 277
    Tools and Instruments 277
    Step by Step 279
Layout 281
Components 282
    Vacuum Tubes 282
    Transformers 283
    Printed Circuit Boards 283
    Speaker 283
    Fuses 284
    Resistors 284
    Wiring and Cables 284
    LED Power Indicator 284
    Capacitors 284
    Silicon Power Diodes 285
    EMI Shielding 285
How It Works 285
Mods 288
    Convenient Fuse Placement 288
    Improved EMI Shielding 288
    Larger Speaker and Cabinet 289
    Improved Thermal Management 289
    Cleaner, Safer Power 289
    Gain Control 290
    Change the Cathode Capacitor 290
    More Efficient Audio Output Transformer 290
    Permanent Speaker Connection 290
    Higher Quality Tubes 290
    Try an Upgrade Kit 291

PART IV Appendixes

A Component Markings 295
B Resources 301

Index 305

This page intentionally left blank

Foreword

Most electronics books for novices and hobbyists either teach the basics or present a set of do-it-yourself projects. This book breaks the rules by reverting back to what most of us did long before we even read an electronics magazine or book. We took things apart to find out how they worked. And that's precisely what Bryan Bergeron, the editor of *Nuts & Volts* magazine, has so expertly done in this remarkably interesting, nicely illustrated, and fun-to-read book.

While *Teardowns* is appropriately subtitled *How Electronics Work by Taking Them Apart*, the book is much more than this. It begins with an introduction that summarizes the keys to working with basic electronic circuits in only 13 fast moving pages, each of which is packed with practical tips and advice based on the author's extensive experience with hands-on electronics. The book then jumps straight to a series of commercially available electronic products that the author opens, pries into, disassembles, and photographs one step at a time.

The text is written as if Dr. Bergeron, who is a highly experienced electronics practitioner, is speaking directly to the reader with a point-by-point commentary about each teardown, complete with clear explanations of the operation and function of every component. By the time the product is completely disassembled, the reader understands the design tricks, component selection, and packaging choices that enabled the product to reach the market.

The teardowns are divided into three sections. Electronic products commonly found around the home are disassembled first, including a smoke alarm, motion-activated light, digital scale, ultrasonic humidifier, and five others. A section for tinkerers includes an intriguing teardown of an analog volt-ohm meter that effortlessly teaches more about Ohm's law than most college textbooks. A closing section on electronic music takes the reader inside a Fender guitar, an effects pedal, and even a vacuum tube audio amplifier. Along the way, Bergeron suggests various improvements, modifications, and enhancements that can be added to many of the products. He also explores the operation of the various sensors in many of the products he investigates, including those that detect light, pressure, sound, and the vibration of a metal guitar string.

The high quality of the many close-up photographs and circuit diagrams greatly enhances the value of this book. Considerable time was obviously devoted to the photographs, which are crystal clear with excellent lighting and depth of field. Want to know if your solder connections are well done? Just compare yours with the examples of sloppy and good connections in the photos. Want to see the different layout methods manufacturers use to install surface-mount components and conventional parts equipped with pins and wire leads? Just browse through the crisp photos of exposed product circuit boards in the first section.

In the end, the author's top-down approach to explaining electronics provides a unique learning experience and a user-friendly reference for both novices and experienced circuit builders. The book is also a useful teaching tool for electronics instructors. Old timers will wonder why we didn't think of this terrific book idea before the prolific Bryan Bergeron arrived on the scene.

Forrest M. Mims III Geronimo Creek Observatory, Texas www.forrestmims.org

The author of 60 books and a long-time contributor to *Nature, Scientific American, Sky & Telescope, Popular Mechanics, IEEE Spectrum*, and many other magazines and journals, Forrest Mims is the most widely read electronics author in the world. His clients have included the National Geographic Society, the National Science Teachers Association, and NASA's Goddard Space Flight Center. In 1993, he was named a Laureate in the Rolex Awards for Enterprise.

Forrest M. Mims III
Geronimo Creek Observatory, Texas
www.forrestmims.org

Acknowledgments

Every book starts as a spark that sets matter into motion. Roger Stewart, editorial director at McGraw-Hill Professional, provided the spark for this book. Special thanks to Joya Anthony, acquisitions coordinator, for her energetic, day-to-day support throughout the publishing process.

This page intentionally left blank

Introduction

Teardowns: Learn How Electronics Work by Taking Them Apart takes a unique, fun approach to learning about electronics by first guiding you through the disassembly process of popular electronic devices and then discussing the uncovered components and circuitry. That old, broken amplifier or electric guitar in your garage can be transformed from a dust collector into the perfect basis for tutorials on learning the principles of electronics. As such, you can follow the projects discussed in this book as the basis for your own teardowns, even if you are on a restricted budget. Moreover, *Teardowns* is designed as a stand-alone reference, so that you can follow and learn from the photographs and descriptions of teardowns without having access to a shop, tools, or specific devices.

As each device is disassembled, the interaction and function of the components are discussed, providing an immediate working context for otherwise abstract concepts. Moreover, this is much more than a "way things work" book, because the teardown process illustrates the physical nature of device construction and provisions for mechanical stability, heat transfer, radio frequency interference prevention, safety, and design for serviceability—information you won't find on a schematic or specifications sheet. In addition, when appropriate, an improvement or modification is made to the device and the results are discussed. In this way, that old amp rescued from the garage serves as a vehicle for learning and enjoys a second life.

The inspiration for this book stems in part from human nature and in part from my early experiences in electronics. You can probably recall a time when a toy or appliance broke and you followed your natural curiosity and disassembled the device, sometimes discovering a fix for the problem, but always learning about how the device was constructed. In my experience, the texts and booklets on beginning electronics provide the theoretical basis for electronics, and necessarily use over-simplified, easily digested circuits, often without a discussion of practical design considerations.

In contrast with my readings of texts and magazine articles, my most memorable early encounters with electronics were at the workbench, working under the tutelage of an experienced technician. I had informal lessons in diagnostics, around whatever piece of equipment he happened to be repairing. But among the more enjoyable experiences was the teardown. We'd start with a ridiculously inexpensive piece of nonfunctional equipment from a local flea market and either repair it or disassemble it for components and spare parts.

I learned to look for hidden treasures. For example, Tektronix used to include a length of silver solder hidden away in its oscilloscopes for repairs. I also learned what components tended to fail first in different equipment. Through repeated quizzing from my mentor, I learned to differentiate between various types of capacitors, inductors, and solid-state components.

The problem with unguided teardowns is that, without knowledgeable guidance, it's easy for a novice to miss subtle design considerations, such as the mechanical stability of electromagnetic interference (EMI) shielding, component placement that facilitates convection cooling, or the usefulness of knowing the pinouts of a standard transistor package. By providing virtual handholding to virtual teardowns, this book serves as a necessary adjunct to the numerous theory-based books and web sites.

Assumptions

In developing this book, I've made a few assumptions about you, the reader. If you're strictly reading along with the teardowns, I assume that you

- have read at least one introductory book on electronics.
- can recognize basic active and passive components.
- have some experience building basic circuits, either from a kit or a web/magazine article.
- have a natural and insatiable inquisitiveness.

In addition, if you plan to use this book as a field guide of sorts to real teardowns, I assume that, in addition to the above, you

- have access to and know how to use the basic tools of the trade: screwdrivers, cutting and needle-nose pliers, soldering iron, desoldering tools, and a multimeter.
- are aware of basic electrical safety issues.
- have and use eye protection. Odds are, if you get shocked by a charged capacitor or cut yourself with a screwdriver, you'll live and soon forget about it. However, if you happen to launch a piece of wire into your eyes with diagonal cutters, your vision could be permanently degraded—at least until bionic implants are available. Protect your eyes.

Goals

I have done my best to insure that, after reading this book, you will

- have a better understanding of the electronics underlying common electronics devices. The teardowns should give you mental anchor points for an otherwise abstract theory of electronics.

- better understand the connection between a physical device and its corresponding schematic. I'll offer a schematic for each device disassembled.
- be able to tear down and understand electronic devices not covered in the book. I'd like you to be able to tear down any device without fear of personal injury or of permanently making the device inoperable.
- better understand the practical considerations that go into electronic product design from both functional and manufacturing perspectives. It's one thing to design a circuit for a textbook exercise, and another to consider how it might be stressed in a high-temperature environment, when subject to vibration and other physical abuse, and used with an imperfect power source.

Organization

Teardowns is organized in three main parts, based on the types of electronic devices discussed, and two appendices. As you'll see, I don't treat components equally from one teardown to the next, but instead focus on what makes a particular device unique. For example, resistors and Ohm's law are merely mentioned in the teardown of an ultrasonic measuring device, but they take center stage in the teardown of an analog Volt-Ohm-Meter (VOM).

Part I: Around the Home

This section targets electronics you're likely to have in and around your home or that you might find at a garage sale. The components and topics covered in these nine chapters are outlined here.

Chapter 1: Dual-sensor Smoke Alarm

This chapter describes the teardown of a residential smoke alarm that features both photoelectric and ionization sensors. The following key components and topics are covered in this chapter:

- Ceramic disc capacitors
- Decibels
- Electrolytic capacitors
- Epoxy-dipped capacitors
- Infrared LEDs
- Ionization sensors
- Photodiodes
- Photoelectric sensors
- Piezoelectric transducers
- Radiation
- Resistors
- Silicon diodes
- Transistors
- Zener diodes

Chapter 2: Motion-activated LED Light

The focus of this teardown is a passive infrared motion-activated LED light. The following key components and topics are covered in this chapter:

- Ambient light sensors
- Battery amp-hours
- CDS (cadmium sulfide) sensors
- Circuit boards
- Electrolytic capacitors
- Fixed-voltage regulators
- Fresnel lens
- LED current limiting resistors
- Passive infrared (PIR) sensors
- SMT transistors

Chapter 3: Digital Bathroom Scale

The typical digital bathroom scale packs a lot of technology into a small, affordable package. This chapter describes the teardown of a popular digital bathroom scale. The following key topics and components are discussed:

- EEPROMs
- Gauge Factor (GF)
- LCD
- Microcontrollers
- Piezoelectric transducers
- SMT components
- Strain gauges
- Transistors
- Wheatstone bridge
- Zebra connectors

Chapter 4: Surge Protective Devices

This chapter features two teardowns of a common surge suppressor strip and a power conditioner sold to audiophiles and musicians. The following topics and components are covered in this chapter:

- Air gap insulation
- Avalanche diodes
- Bleeder resistors
- Bridge rectifiers
- Ceramic disc capacitors
- Electrical isolation
- Electrolytic capacitors
- Electromagnetic interference (EMI)
- Ferrite inductor
- Filter circuits
- Flame-proof resistors
- Illuminated rocker switches
- Joule rating
- Linear filtering
- Metal oxide varistors (MOVs)
- Metalized polyester film capacitors
- Mylar capacitors
- NEMA plugs and jacks
- Neon bulbs
- Relays
- Response time
- Silicon diodes
- Slow-blow fuses
- Thermal circuit breakers and fuses
- Thermal MOVs (TMOVs)
- Toroidal inductors
- Transformers
- Transient voltage suppression (TVS) diodes
- Tubular ferrite chokes
- UL 1449 rating
- Visible light LED
- Voltage regulators

Chapter 5: Electronic Pedometer

Many dedicated handheld electronic devices, from GPS units to cameras, have been consumed by the smartphone. This chapter describes the teardown of one of the last holdouts, the stand-alone electronic pedometer. The following topics are covered in the teardown:

- Circuit board layout
- Elastomeric buttons
- LCDs
- Mechanical pendulum sensor
- MEMS (micro electro-mechanical machine system) accelerometers
- Microcontrollers
- Piezoelectric pendulum sensors
- Potentiometers
- SMT components
- Tuning fork crystal oscillator
- Zebra connectors

Chapter 6: Compact Fluorescent Lamp

The compact fluorescent (CF) lamp is an evolutionary paradox. Although more energy-efficient than a small length of tungsten wire, the complex device violates most of the laws of basic engineering. The key concepts and components covered in this teardown include the following:

- Capacitive reactance calculations
- Color rendition index (CRI)
- Color temperature
- Diode for alternating current (DIAC)
- Diode peak inverse voltage (PIV) rating
- Electrolytic capacitors
- Electromagnetic interference (EMI)
- Ferrite core inductors
- Flame-proof resistors
- Fusible metal film resistors
- Heat sinks
- High-voltage measurement
- Lumens
- Mercury toxicity
- Power factor
- Silicon diodes
- Silver mica capacitors
- Toroidal transformers
- Total harmonic distortion (THD)
- UV radiation
- Voltage dividers
- Voltage doublers

Chapter 7: Ultrasonic Humidifier

This teardown unveils the hidden complexity of a popular ultrasonic humidifier. Key issues and components discussed include the following:

- Bicolor LEDs
- Bridge rectifiers
- Brushless DC (BLDC) motors
- Capacitive reactance calculations
- Ceramic disc capacitors
- Closed loop control systems
- Diode PIV rating
- Electrolytic capacitors
- Epoxy-dipped capacitors
- Fast-blow fuses
- Ferrite core inductors
- Hall effect sensors
- Heat sinks
- Inductive reactance calculations

- Iron laminate power transformers
- LC resonant frequency
- LED current limiting resistors
- Linear power supply
- Magnetic reed switches
- Metalized polypropylene film capacitors
- Motor controller IC
- Mylar film capacitors
- Open loop control systems
- Open-air inductors
- Piezoelectric transducers
- Potentiometers
- Response time
- RF oscillator
- Rocker switch
- Silicon power diodes
- Switching diodes
- Transistor current gain
- Transistor switches

Chapter 8: Digital Hygro Thermometer

This teardown introduces basic analog sensor circuits. Key component and topics discussed include the following:

- Analog sensors
- Circuit board layout
- Humidity sensors
- LCDs
- Measuring relative humidity
- Microcontrollers
- RC time constants
- SMT components
- Temperature coefficients
- Thermistors
- Tuning fork crystal oscillator

Chapter 9: Stereo Power Amplifier

This teardown features a popular, inexpensive stereo power amplifier with components and circuits you'll likely to find in any modern amplifier. The following key components and topics are discussed:

- ASICs
- Audio taper potentiometers
- Bridge rectifiers
- Bridging amplifiers
- Bypass capacitors
- Center-tapped transformers
- Chokes
- Class AB stereo audio amplifiers
- Decibels
- °C/W ratings
- DPDT switches
- Dual-polarity power supplies
- Frequency response
- Heat sinks
- Impedance matching
- Linear power supplies
- Metalized polyester film capacitors
- Mica insulators
- Mylar film capacitors
- Operational amplifiers
- Precision metal film resistors
- Relays
- Signal diodes
- Signal-to-noise (SNR) ratio
- Silicon power diodes
- Silver mica capacitors
- Slow-blow fuses

- Stacked potentiometers
- Switching diodes
- Tantalum capacitors
- Thermal management
- Toroidal power transformers
- Total harmonic distortion plus noise (THD+N)
- Visible-light LEDs
- Zener diode shunt voltage regulators

Part II: For Tinkerers

This section focuses on two electronic instruments you're likely to have (or want) on your workbench or toolbox: an analog VOM and a laser-guided ultrasonic distance measurer.

Chapter 10: Analog Volt-Ohm-Meter

This chapter reveals the circuitry and components found in a typical VOM. The following topics and components are discussed:

- Battery internal resistance
- D'Arsonval galvanometers
- Decibel measurement
- Diode forward voltage drop
- Diode PIV
- Electromagnetic buzzers
- Fast-acting fuses
- Full-scale accuracy
- Germanium diodes
- Linear potentiometer
- MOVs
- Meter sensitivity
- Nichrome wire resistors
- Precision metal film resistors
- Resolution, precision, and accuracy
- Rotary switch
- Schottky diodes
- Series and shunt resistors
- SMT components
- SMT resistor voltage rating
- Sound pressure calculations

Chapter 11: Laser-guided Sonic Distance Measurer

This chapter explores ultrasonic distance measurers and the physics of ultrasound. Key topics and components covered in this teardown include the following:

- Analog sensors
- Battery power
- Crystal ageing
- Elastomeric switches
- Electrolytic capacitors
- Fixed-voltage regulators
- Frequency response
- Heat sinks
- LC circuit resonant frequency
- LCDs
- Linear taper potentiometers
- Microcontrollers
- Operational amplifiers
- Resonant frequency calculations
- Signal-to-noise (SNR) ratio
- SMT components
- Sound propagation
- Step-up transformers
- Temperature coefficient
- Thermistors
- Ultrasonic transducers
- Variable voltage regulators
- Visible light laser
- Zener diode voltage clamp

Part III: For Musicians

The teardowns in this section, for your inner rock star, include an electric guitar, a tube-type guitar amplifier, and an effects pedal. The components and topics covered in these three chapters are outlined next.

Chapter 12: Electric Guitar

This chapter describes the teardown of a Fender American Deluxe Telecaster, including a detailed teardown of a magnetic pickup. Topics and components in this teardown include the following:

- Active and passive magnetic pickups
- Alnico magnets
- Audio jacks and switches
- Audio taper potentiometers
- Bypass capacitors
- Cable capacitance
- Capacitive reactance calculations
- Ceramic disc capacitors
- EMI shielding
- Frequency response
- Humbucking magnetic pickups
- Identifying magnet polarity
- LC circuit resonant frequency
- Measuring magnetic field strength
- MIDI
- Noise reduction
- Piezoelectric pickups
- RC filters
- Star ground configuration
- Wire insulation

Chapter 13: Effects Pedal

In this chapter, I'll tear down a popular distortion pedal for the electric guitar. The following key components and topics are discussed:

- Amplifier design
- Audio jacks and switches
- Audio taper potentiometers
- Ceramic disc capacitors
- Clipping
- Electrolytic capacitors
- Germanium diodes
- Intentional audio distortion
- Mylar film capacitors
- Operational amplifiers
- Silicon diodes
- Visible light LED

Chapter 14: Vacuum Tube Guitar Amplifier

The focus of this teardown is an all-tube 5W combo amplifier that's a favorite in the modding community because of its affordability and ease of modification. Key topics and components discussed in this teardown include the following:

- Audio output transformers
- Audio taper potentiometers
- Bleeder resistors
- Bridge rectifiers
- Bypass capacitors
- Center-tapped transformers
- Ceramic disc capacitors
- Chokes

- Class A audio amplifiers
- Electrolytic capacitors
- EMI shielding
- Enclosure acoustics
- Flameproof resistors
- Frequency response
- High-voltage power supply design
- High-voltage probe design
- Impedance matching
- Laminate core power transformers
- Linear power supplies
- Metalized polypropylene film capacitors
- Mylar film capacitors
- Silicon power diodes
- Slow-blow fuses
- Speaker design
- Thermal management
- Vacuum tubes

Appendixes

The appendixes include references to help you identify component values and learn more about specific topics introduced in the teardowns. Appendix A provides common component markings used on leaded and SMT capacitors, resistors, and inductors. Appendix B lists online and print resources for more information on components, design principles, and the topics discussed in each teardown.

If You Try This at Home

Sooner or later, you'll want to replicate at least one of the teardowns described in this book, and I encourage you to do so. Teardowns are fun, and there's no better way to learn the practical aspects of electronic device design and construction. However, if you're going to make the most of your educational experience, you'll have to exercise good judgment and take some commonsense precautions to avoid potentially dangerous situations.

Kill the Power

Before you start to disassemble an AC-powered device, verify that it's unplugged. I like to keep the AC plug on my workbench where I can see it, next to the device. Sometimes I'll wrap the plug with masking tape. Otherwise, in the heat of a teardown, you may hurriedly attempt to plug in your soldering iron, only to discover a few minutes later that you have a cold iron and a hot power supply.

Wear Eye Protection

Wear safety glasses. I like the polycarbonate DeWALT Reinforcer Safety Glasses with magnification (www.Dewalt.com) because they're affordable and lightweight, and the built-in magnifier is great for working with surface-mount components.

Cover Your Skin

A long-sleeve shirt and full-length pants are a good idea if you plan to work with a soldering iron. Solder can splash when you extract components from a circuit board. It's also a good idea to keep a pair of leather gloves handy and wear them when there's a risk of being pinched or punctured. Avoid synthetic gloves because a hot soldering iron tip cuts through them like butter.

Don't Inhale

Avoid breathing the fumes from molten solder. The rosin core used with solder is a powerful lung irritant. If you use a soldering iron during a teardown, make sure your work area is ventilated.

Wash Your Hands

Assume the circuit boards you handle are manufactured with lead solder, even if they state otherwise. Wash your hands after a teardown, especially before eating food or playing with children. Ingesting lead can lead to reduced IQ, anemia, attention problems, and seizures, among other problems. Look what lead poisoning did for the Roman Empire.

Beware of High-voltage and High-current Circuits

With the advent of low-voltage wall transformers and low-voltage components, serious electrical shocks from consumer electronics are relatively rare. Even so, I have seen a wedding band vaporized by a 12V, high-current circuit. Fortunately, the wearer didn't lose his finger, but he spent a fortune on reconstructive surgery.

If you tear down an amp or other device with high-voltage vacuum tubes, remember that high voltages can linger in the power circuitry for up to several days after the amplifier has been turned off and unplugged. My motto: When in doubt, short it out. With the amplifier unplugged, take an alligator clip lead, connect one end to the chassis ground, and carefully touch the other end to the positive (+) and negative (–) terminals on power supply filter capacitors. I usually clip a 10Ω, 5W resistor to the lead to limit the initial shorting current.

Don't Panic

You can stay safe by maintaining a sense of self-control. Resist the urge to take out the 5-pound sledge and flatten that circuit board that you've been trying to extract for a half hour. Take a break, come back, and look for the hidden screw (there's always at least one); then proceed thoughtfully. The goal, after all, is to learn and have fun. And electronics is best enjoyed when you're alive, alert, and intact.

Establish a Good Work Environment

For clarity, I photographed the majority of teardowns on a clean white tabletop. However, when working off-camera, I use an indestructible, butcher-block workbench with a hefty vise and power tools that I can use to crack open just about anything. For detailed circuit tracing and analysis, I have an illuminated, ventilated Formica work surface with most of my tools within easy reach (see Figure 1).

You don't need a special work surface for a teardown. Given an understanding spouse, parent, or roommate, a counter top or kitchen table is a fine place for a teardown, as long as you protect the surface from dings and scratches. Wherever you work, do your best to insure that it's well-lit, adequately ventilated, and immune to flying pieces of metal, glass and plastic, and the spray of molten solder.

Use the Appropriate Tools

From a safety perspective, use the tools most appropriate for the job. In other words, use a hex-head driver to remove a hex nut, as opposed to a pair of needle-nose pliers that are likely to slip and either pinch your hand or impale your foot. Not only will you save money on bandages, but your tools will last longer.

The hand tools you'll need for basic teardowns include Phillips and standard screwdrivers, needle-nose pliers, diagonal cutters, crescent wrench, and occasionally a soldering iron and solder wick braid. Eventually, you'll want to increase your arsenal of tools to include a pair of tweezers, a hex driver set, a power screwdriver, drill, and illuminated magnifying glass. A lightweight, desktop vise, such as one from PanaVise (www.panaviseonline.com), is great for working with small devices

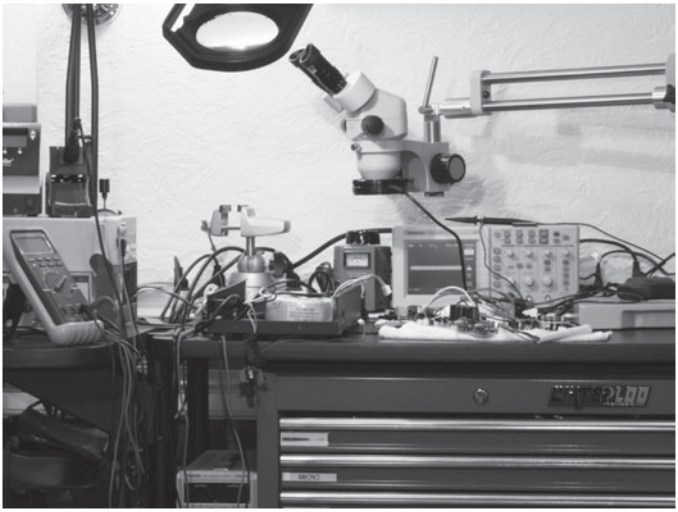

FIGURE 1 The author's work environment

and reworking circuit boards. Once you invest in a base, you can add a variety of heads, from circuit board holders to my favorite, the extra-wide opening head.

In the nice-but-not-necessary category are a variety of power tools. Depending on the task at hand, I choose between a cordless Dremel, a massive Milwaukee 1/2-inch drill, and Proxxon tabletop drill/milling machine. The ultra-precise Proxxon, an eBay find, is my favorite for reworking circuit boards. I also use a basic Weller WD1 soldering station for leaded components and a Weller WRS1002 hot air repair system that I picked up at a flea market for working with surface-mount components.

If you're tearing down a kitchen appliance, you'll want tough tools made for heavy gauge work. Conversely, if you're harvesting a delicate circuit and surface-mount components, you'll appreciate delicate, finely crafted tools. Diagonal cutters and pliers are rated for wire type and gauge. Learn the operating range of your tools. And don't forget the eye protection.

Use Instruments You Can Trust

The only instrument you need for basic teardowns is an inexpensive multimeter. You'll find the resistance function useful for tracing circuits and verifying component values, and the voltage function is indispensable for verifying that high-voltage capacitors are discharged.

If you work with high-voltage power supplies, consider purchasing a variable transformer. I admit that I worked with tube amps and high-voltage supplies for years without a Variac, but that's before the time of $20 audiophile capacitors and $250 power transformers. Besides, applying full voltage to dry electrolytic capacitors can make a real mess. I picked up new a 120V at 3A variable transformer with built-in voltmeter from All Electronics (www.allelectronics.com) for $60, and you can probably do much better on eBay.

If you can afford it, consider purchasing a Variac with built-in ammeter. A less-expensive alternative is to install a Kill A Watt EZ power monitor on the front end of any Variac you can get your hands on. The power monitor, available for about $35 from Amazon, provides a digital readout of voltage, current, and power factor, and you can unplug it and use it to assess the electrical equipment around your house.

Visualizing and assessing the polarity of magnetic fields can be important if you work with electric guitar pickups or electric motors. If you plan to work with magnets and electric motors, consider purchasing a magnet polarity tester and a magnetic field viewing paper. I use a $9 polarity tester from Stewart-MacDonald, a luthier supply shop (www.stewmac.com). The visualizing paper, available from Edmund Scientifics (www.scientificsonline.com) for about $3, is a laminated sheet of plastic containing iron filings suspended in oil. It's cleaner and more efficient than using iron filings.

In the nice-to-have category are an oscilloscope, a signal generator, and an automatic data logging system. The inexpensive Parallax USB Oscilloscope (www.Parallax.com) is a good starter oscilloscope. For more precise work and the ability to work away from a PC, I use a Tektronix TDS2022 color digital oscilloscope, a Fluke 87 DMM, and a Fluke 45 Dual Display Multi-Meter. For automated data logging, I use a USB-1208FS 12-bit data logger from Measurement Computing (www.mccdaq.com). The free software that accompanies the data logger is needlessly crippled in all the wrong places, but the hardware is bulletproof and a real time-saver.

Even if you haven't developed presbyopia, you'll find a magnified worklight handy. I use the three-diopter Electrix Halogen magnifier (www.electrixtask.com). It's bright, it provides a good working distance, and it's good for lighting tasks such as unsoldering leaded components. The downside is that the halogen bulb and anything close to it receives a thermal lashing. For detailed circuit tracing, I use a ProScope HR (www.bodelin.com/proscopehr) with the 100X illuminated head. While affordable, this handheld microscope tethers you to a laptop or desktop PC/Mac, and the working distance—the distance from the objective microscope lens to the subject—is only a few millimeters.

For serious surface mount work and identifying otherwise illegible component markings, I use a Luxo binocular microscope with dimmable LED ring light (www.luxo.com). It's expensive, but the working distance is about 5 inches—plenty of room for fingers, tweezers, and a soldering iron or hot air wand. This is definitely in the eBay, used equipment category.

Jotting down schematics in a notebook as you work is fine, but you'll eventually want to transfer schematics to something more legible. I've had good luck with ExpressSCH (www.expresspcb.com), which is free and easy to use. It's one of a dozen low-cost or free versions of schematic authoring tools on the Web.

Keep It Simple

If you're just starting out, keep it simple. Safety glasses, pliers, wrench, diagonal cutters, screwdrivers, and maybe a multimeter with the ability to measure capacitance and inductance are a reasonable arsenal of tools. You want to focus on the teardown, and not on how to operate a new battery of instruments.

The list of instruments in the preceding section may seem overwhelming at first—and it should. It represents my years of scouring eBay and local flea markets for deals on used but functional equipment. As you progress to designing and building your own circuits, you can gradually add the tools and instruments that suit your particular interests and budget.

Disclaimer

I've done my best to represent the devices and components fairly and accurately in the teardowns. As you'll note, many of the schematics are simplified, and all are based on my analysis. My objective is to highlight specific aspects of a circuit, not to provide an unofficial technical manual for the equipment. If you need a complete, official schematic, contact the manufacturer.

I independently selected and purchased the devices used in the teardowns, usually on Amazon or eBay. In fact, I purchased two and sometimes three of every device discussed in this book. I used one for a destructive, component-level teardown to determine, for example, the turns ratio of a transformer. The second device was for photographing and bench measurements. The occasional third device was for testing and verifying the operation of the second unit.

I don't have a financial interest in any of the manufacturers of the devices discussed in this book. I also don't have anything against any of the manufacturers. When something in a teardown disagrees with marketing material or specifications, I'll point it out. One thing that you'll learn from teardowns is that when the fine print on the packaging states "Specifications subject to change without notice," you're likely to find a few surprises in the teardown.

Web Site

The illustrations of teardowns in this book are necessarily limited to modest-sized grayscale images. To enhance your vicarious enjoyment, I've made additional media freely available at www.mhprofessional.com/computingdownload.

PART I

Around the Home

This page intentionally left blank

Chapter 1

Dual-sensor Smoke Alarm

Smoke alarms, such as the Kidde Pi9000 shown in Figure 1-1, are such commodity items that you probably rarely take notice of the white, saucer-shaped appliances that adorn the ceilings of homes, offices, and classrooms unless they're acting up. And if you've seen and heard one, you've seen and heard them all. Most sport a blinking red LED (light-emitting diode) to indicate the status of the battery or other power source. A test button lets you check out the audible alarm, and an obnoxious chirp indicates that it's time to change the battery. Your home unit likely has a hush button to kill the audible alarm when you accidentally trigger the unit by frying something a little too long on the stove.

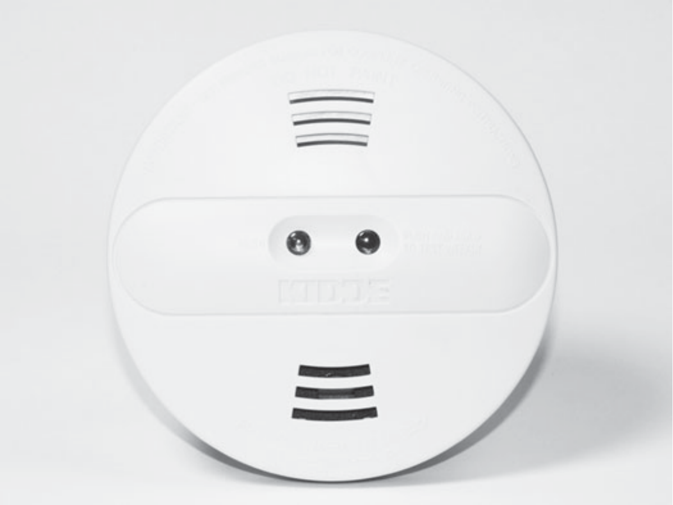

FIGURE 1-1 Kidde Pi9000 smoke alarm

Reflecting this external homogeneity, little variability is found under the hoods of most models. The major differences between the less expensive and advanced commercial and residential units are connectivity and power. Wired or wireless connectivity, a central power source, and triggers for an autodialer for the fire department can double or triple the cost of a system. However, smoke alarms are the same where it matters most—in the sensor technology they use to detect smoke. And smoke inhalation is responsible for more deaths than is direct contact with fire.

The first technology is an *ionization-based sensor*, which is best for detecting the minute particulate by-products of burning liquids and other clean burning fires. The second technology is a *photoelectric sensor*, which responds best to the large particulate by-products from burning cigarettes, bedding, carpet, and other dirty, smoldering fires that produce optically dense smoke. Each technology has its strong points, and no single sensor is best for all situations. Many smoke alarms leverage the advantages of both technologies by offering dual sensors: an ionization sensor for clean fires and a photoelectric sensor for smoldering fires.

In this chapter, I'll tear down a dual-sensor Kidde Pi9000 smoke alarm. The basic sensors and, in many cases, the underlying circuitry and components are similar to units marketed by First Alert, Sentry, and many of the generic brands. In addition to exploring the overall design of this $20 system, I'll walk you through teardowns of the ionizing and photoelectric sensors. Because this is your first teardown, I'll discuss the circuits and components at a relatively high level. I'll dig deeper into the technology in the following teardowns.

Highlights

This is a quick teardown for the basic unit; the sensors are a little more involved and even potentially dangerous. The outer shell of the smoke alarm can be disassembled with a standard screwdriver. Once you're inside, the circuit board simply snaps out. The ionization sensor, the smoke sensor that uses ionization as the basis for its operation, is notable because it contains the synthetic element Americium-241, a potentially lethal alpha particle emitter with a half-life of about 430 years.

During the teardown, note the following:

- **The multiuse circuit board** Kidde apparently uses one board for several of its smoke alarms to cut down on inventory costs.
- **Component count relative to functionality** The Pi9000 is built around two chips that are designed for smoke alarms, which minimizes overall component count.
- **Construction of the two sensors, especially how they facilitate airflow** Particles have to make their way through the plastic outer shell of the alarm housing and then through the narrow slits of the sensors to be detected.

Specifications

Kidde doesn't provide typical technical specifications on its packaging, such as the minimum particulate count that will trip the alarm. The only information that comes close to a technical specification is the loudness of the alarm—85 decibels (dB)—but the number is meaningless because distance isn't specified. Information from the manufacturer of the piezoelectric transducer (which converts electrical pulses to mechanical vibrations—for the alarm sound) suggests that this distance is about 10 feet.

Despite the lack of specifications, the package does provide a feature list and a warning:

- **Dual sensors** An ionization sensor for fast, flaming fires and a photoelectric sensor for slow, smoldering fires
- **Hush button** To silence nuisance alarms
- **Test button and LED** For visual verification of battery and alarm operation
- **Warning chirp** To indicate that the battery needs to be replaced
- **Radiation notice** The product contains 0.9 microcurie (µCi) of Americium-241

Operation

Once the battery is installed, alarm operation is automatic. In normal conditions, the red LED flashes every 30 to 40 seconds to indicate that the unit is operating correctly. When a significant level of smoke is present, the alarm sounds, signaling that it's time to get moving.

Manual operation is limited to pushing and holding the test button and bracing yourself for the blast of audio. If the alarm trips because you're burning something on the stove, you can push the hush button to replace the continuous alarm with a less annoying chirp. However, if you're still smoking up the kitchen after 7 minutes, the alarm logic will override the hush command, and you'll have to bear with a continuous alarm.

Teardown

The teardown of the basic unit is illustrated in Figure 1-2 and discussed in the sections that follow. (I'll discuss the sensors later in the "Components" section.) If you plan to follow along and examine a physical smoke alarm, give yourself about 30 minutes to complete the teardown.

Caution The ionization sensor contains a small amount of Americium, which is extremely dangerous if you inhale or ingest it. If you do plan to replicate this teardown at home, your safest course of action is to skip the teardown of the ionization-based sensor. It's also mandatory that you wash your hands after the teardown.

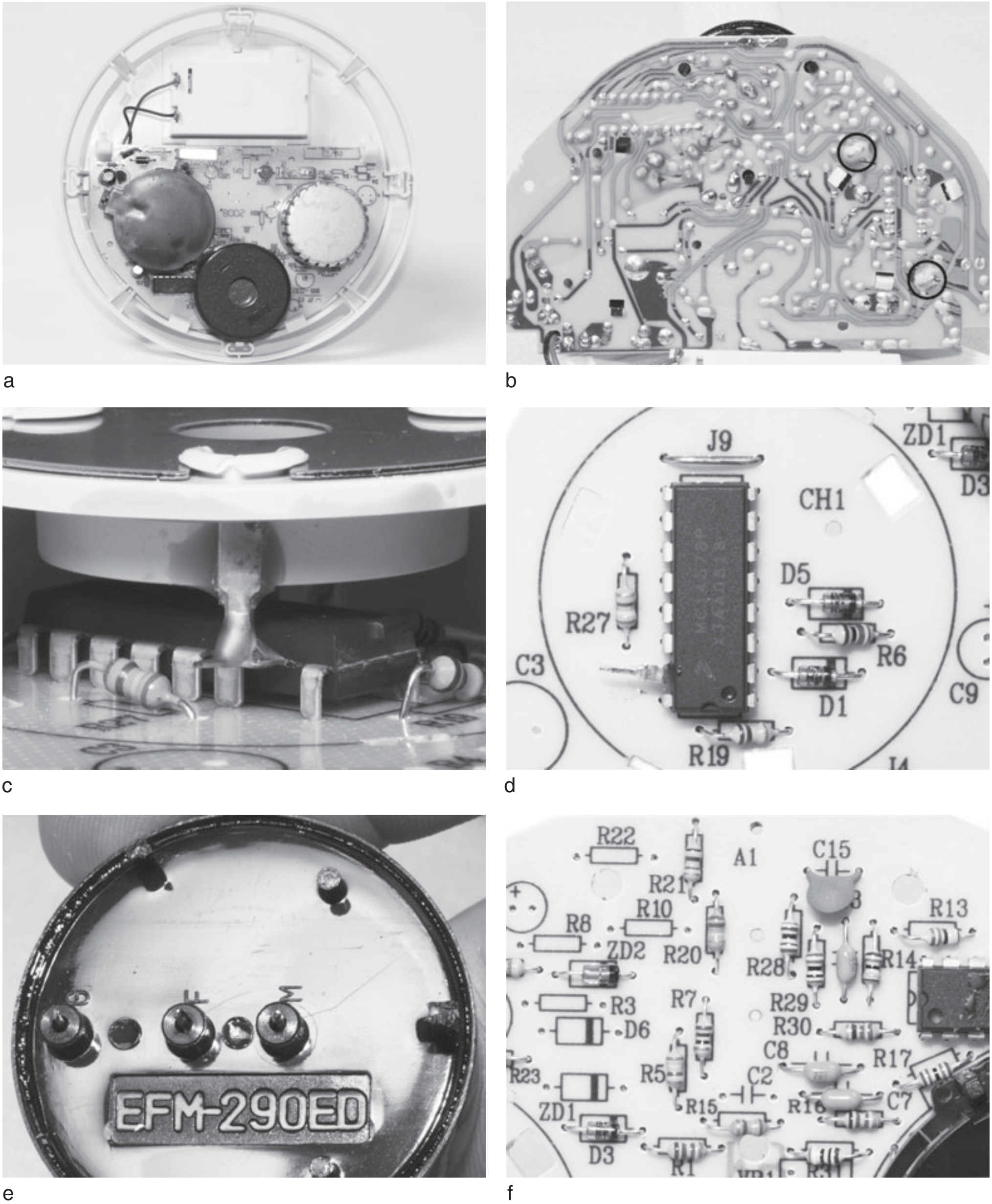
a
b
c
d
e
f
J9
CH1
ZD1
D3
D5
R27
R6
C3
C9
D1
R19
EFM-290ED
R22
A1
C15
R21
R10
R13
R8
ZD2
R20
R28
R14
R3
R7
R29
D6
R30
C8
R17
R23
R5
C2
R15
R16
C7
ZD1
D3
R1

FIGURE 1-2 Teardown sequence

Americium is an alpha particle source. Alpha particles, two protons and two neutrons, are dense and relatively slow and travel perhaps 3 feet in open air. They are easily stopped by a sheet of paper, clothing, and even thick skin. The greatest potential for harm results from inhaling or ingesting the radioactive element. The linings of your lungs and intestines are only a few cells thick, much thinner than a sheet of paper or your hide, and are highly susceptible to radiation damage. So don't try to cut, file, sand, or remove the pencil eraser–sized Americium disc in the ionization-based sensor.

Tools and Instruments

This teardown can be performed with a small standard screwdriver, an old pair of diagonal cutters, a soldering iron with the usual soldering and desoldering accessories, a multimeter, and a magnifier. A Geiger counter or other alpha-radiation detector is required to assess the level of alpha radiation emitted by the ionization source.

Caution If you try this teardown at home, do not crack open the ionization sensor unless you have a means of monitoring the radiation level.

Step by Step

Accessing the single circuit board in the smoke alarm involves little more than removing a plastic cover and unsnapping the circuit board from the plastic base.

Step 1

Crack the case. With the unit face down and the battery compartment empty, use a standard screwdriver blade to disengage the four equally spaced plastic hooks that hold the two halves of the clamshell case together. Remove the front half—the side normally facing down from the ceiling—and set it aside. Orient the bottom half of the alarm so that the 9V battery compartment is away from you, as shown in Figure 1-2a.

Step 2

Identify the sensors. The three prominent cylindrical components, from left to right, are the photoelectric sensor, the piezoelectric transducer, and the steel-encased ionization sensor. I'm tempted to refer to the piezoelectric transducer as the *buzzer* for simplicity, but as you'll see later, it requires an external circuit to create the alarm tone.

Note that each of these components is open to the air. The two sensors feature vertical slits to take in particulate by-products from fire while excluding insects, dust, and, in the case of the photoelectric sensor, light. The cover of the piezoelectric transducer has a large hole in its center to enable the internal piezoelectric disc to interface directly with the air to create ear-piercing sound waves. Locate and identify the 16-pin MC146012ED IC (integrated circuit), just to the left of the piezoelectric transducer and adjacent to the photoelectric sensor.

Step 3 (Optional and Potentially Hazardous)

Remove the outer steel can of the ionization sensor. Remember that this is an optional step that can potentially expose you to alpha radiation.

Caution From a personal liability perspective, I advise against removing this sensor can unless you have supervision from someone knowledgeable about radiation. At a minimum, have a Geiger counter by your side so that you can determine if and when you're being exposed to radiation. Without a counter, there's no way to know if the outside of the sensor is contaminated, for example.

If you're equipped to follow along at this point, put on a pair of latex gloves. The inexpensive surgical variety is fine, but you'll get extra protection by wearing the thicker gloves sold for dishwashing. Next, locate the two metal tabs from the slotted steel ionization sensor cover on the underside of the board, circled in Figure 1-2b. Unsolder the tabs and remove the can to reveal the plastic and steel ionization chamber. Note the MC14578P IC, tucked beneath the chamber, as shown in Figure 1-2c.

The IC is a standard 16-pin DIP (dual inline package) with pin 1 marked by the depression in the top of the package. In Figure 1-2c, pin 1 is in the lower-left corner of the IC, pin 8 is in the upper-right corner, pin 9 is in the upper-left corner, and pin 16 is in the lower-left corner. Don't forget that the pin numbering system is clockwise if you're looking at the underside of the component.

Step 4 (Optional and Potentially Hazardous)

Extract the ionization chamber from the circuit board. Unsolder the connection between the ionization chamber and straightened pin 15 of the MC14578P beneath the chamber, visible in Figure 1-2c. Unsolder the tab from the top plate of the ionization chamber where it connects to the underside of the circuit board, near pin 7 of the IC. Locate the three white plastic tabs on the underside of the board, adjacent to where the two metal tabs were removed from the board. Cut the plastic tabs with a pair of diagonal cutters or push them through the board with the tip of your screwdriver.

Pull the ionization chamber straight up to for a full view of the MC14578P and a few external components, as shown in Figure 1-2d. Put the chamber in a plastic freezer bag and place the sealed bag in a metal drawer or other container. The plastic bag should contain the alpha radiation, but the added protection provided by a metal container can't hurt.

Step 5

Temporarily remove the piezoelectric transducer from the circuit board. From the underside of the circuit board, unsolder the three leads of the transducer. Examine the bottom of the extracted transducer, shown in Figure 1-2e. Note the three terminals, marked *G*, *F*, and *M*. Examine the transducer with an ohmmeter while you have the component out of circuit. You should read several megohms (MΩ) between each terminal. You'll have to replace the transducer to test the logic and sensors, but for now set the transducer aside.

Step 6

Examine the components that were hidden under the piezoelectric transducer, shown in Figure 1-2f. Note the outlines of components that were not installed by the manufacturer. Now is a good time to trace the circuitry of the alarm.

Step 7

Extract the photoelectric sensor. Unsolder the two leads to an IR (infrared) LED at the 10 o'clock position of the sensor and two leads and a tab to an IR photodiode at the 5 o'clock position. You can't see the IR LED or IR photodiode at this point in the teardown because they're hidden by the case of the sensor. Note the C8050 NPN transistor and the silicon power diode external to the sensor at the 11 o'clock position.

Step 8

Replace the piezoelectric transducer. Solder the transducer back in place. Then take a break and wash your hands. Even if you didn't extract the ionization chamber, it's a good idea to wash your hands because of possible contamination with Americium.

Layout

As revealed during the teardown, the 9V battery is located in a separate compartment in the base. I found it easier to work with the circuit by unsoldering the connection to the compartment and using a bench supply for power.

The single-sided circuit board shows stencils for components not included in this product, including a transistor, several diodes, and a few capacitors, probably for the more expensive models with recorded voice alarms. Other than the unorthodox method of soldering the IC pin directly to the ionization sensor and the use of inexpensive spring switches, this is an unremarkable layout.

Components

The two 16-pin chips, the ionization and photoelectric sensors, and the piezoelectric transducer make up the bulk of the smoke alarm circuitry. Let's consider the ionization and photoelectric detection systems separately.

Ionization-based Sensor

The ionization-based sensor is an ionization chamber that requires an MC14578P or equivalent discrete components to perform as a smoke detector. The ionization chamber is constructed of two discs separated by about 1 centimeter of air. The top disc is perforated in the center to allow air and smoke into the chamber, as shown in Figure 1-3. The figure shows the perforated top of the chamber partially removed, revealing a glimpse of the bottom disc.

A pellet of Americium-241 is firmly embedded in the center of the bottom disc. Figure 1-4 shows the view from the top of the sensor. The pellet does not extend through the bottom disc, thereby containing the radiation to the area above the bottom disc. Because the top disc is perforated, ionization leaves the chamber but is stopped by the steel cover that was removed earlier in the teardown.

According to the package back, the pellet contains 0.9μCi of Americium-241. The curie (Ci) is a unit of radioactivity related to the number of decays per second:

$$1\text{Ci} = 3.7 \times 10^{10} \text{ decays/second}$$

$$0.9\mu\text{Ci} = 3.3 \times 10^{6} \text{ decays/second}$$

The number of decays per second, indicative of the number of alpha particles kicked out of the pellet per second, is key to this application, because the alpha particles keep the air within the ionization sensor ionized. And ionized air conducts, unless it's filled with smoke.

If you have access to a radiation detector, you can test the activity of the Americium within the ionization chamber, as shown in Figure 1-5. As you can see, a few centimeters from the source the activity is 1.56 thousand counts per

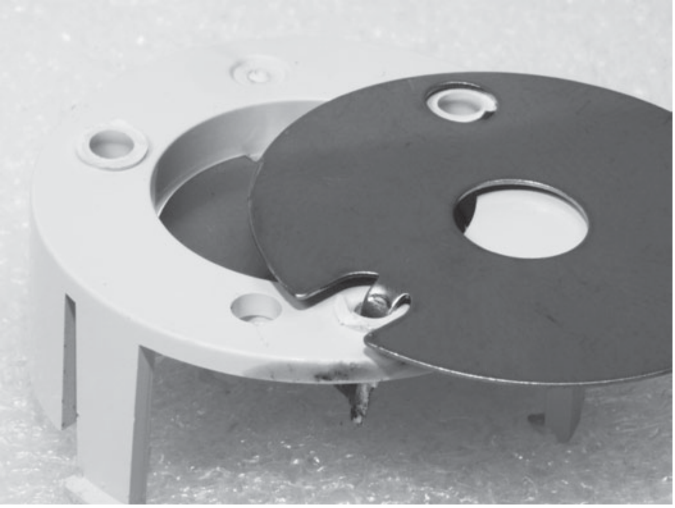

FIGURE 1-3 Ionization chamber with perforated top partially removed

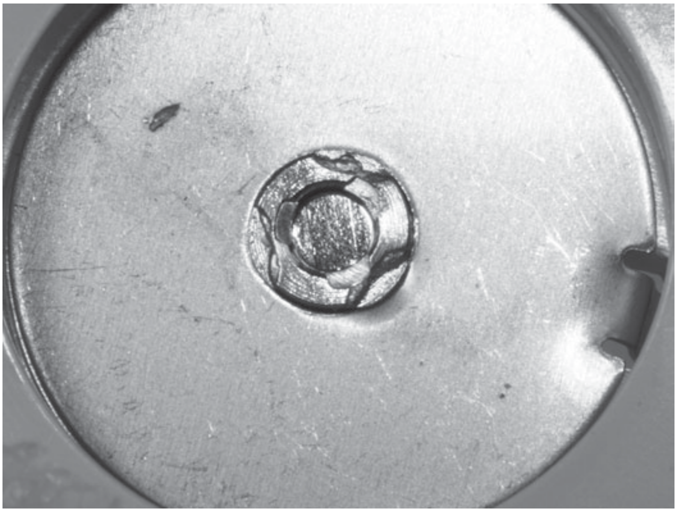

FIGURE 1-4 Americium-241 embedded in the bottom disc of the ionization chamber

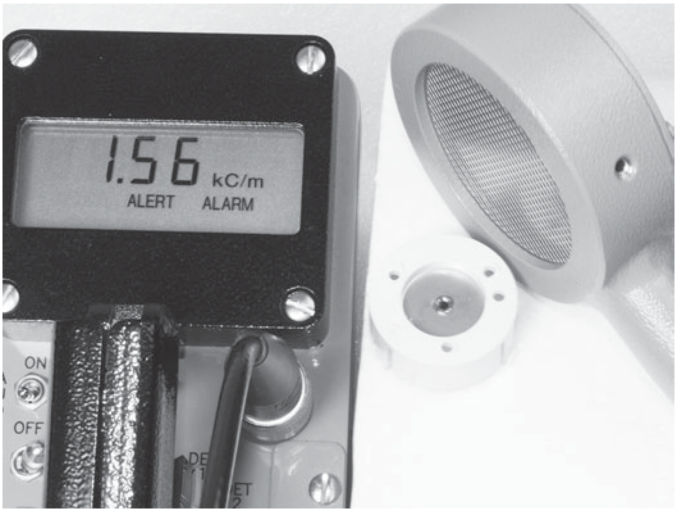

FIGURE 1-5 Radiation meter reading Americium alpha particle activity

minute (kC/m), or about 26 counts per second—enough radiation to cause problems if the Americium is handled carelessly.

Caution If you handle the ionization chamber, make certain to wash your hands immediately afterward.

Ionization Sensor IC

The 16-pin MC14578P is designed specifically to interface with an ionization chamber as part of a smoke alarm. The IC is a CMOS (complementary metal oxide semiconductor) micro-power comparator that requires only one external component—a resistor. The comparator is configured so that minute changes in ionization current result in massive negative or positive swings in the comparator output, depending on the relative change in input of the comparator.

A key parameter of the ionization sensor IC is a standby current of only 10 microamps (μA) and a comparator input current of 1 picoamps (pA). In other words, the chip loafs along consuming only 10μA as long as the ionization current is stable. This ultra-low power consumption contributes to the longevity of 9V batteries in this home smoke alarm.

The chip can also supply 25 milliamps (mA) output current when triggered by a change in ionization current, which is more than enough to signal another device in the smoke alarm. For more detailed information on the chip, see the technical data document available from Freescale Semiconductor, Inc. (www.freescale.com).

Photoelectric Sensor

The photoelectric sensor consists of an IR photodiode and an IR LED mounted in a black, light-tight chamber with slits that are wide enough to allow air and smoke in. As shown in Figure 1-6, the IR LED is oriented at the 1 o'clock position, but the

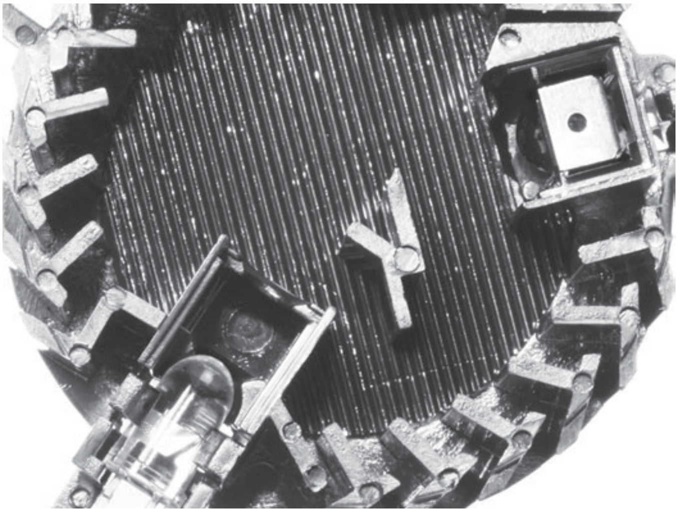

FIGURE 1-6 Photoelectric sensor with cover removed

photodiode is at about the 2 o'clock position. An obstructing wall is positioned near the center of the sensor. Together, the positioning and obstruction wall ensure that, under normal conditions, no direct IR radiation from the LED strikes the photodiode. Light from the LED is absorbed by the corrugated black surface inside the photoelectric sensor.

However, when large particles from a smoldering flame enter the slits of the chamber, they reflect light from the LED, and this radiation is picked up by the photodiode. The more reflected radiation that hits the photodiode, the greater the photodiode current. At a predetermined level, the current trips the alarm and the piezoelectric transducer is activated.

To facilitate testing and experimenting with the sensor, I extracted the LED and IR photodiode from the black plastic case and soldered them in their normal configuration on the circuit board. Figure 1-7 shows the photodiode and IR LED remounted on the circuit board without the light-tight sensor shell.

Photoelectric Sensor IC

The heart of the photoelectric sensor is the photoelectric sensor IC, an MC146012ED, which is marketed by Freescale Semiconductor as a "Low Power CMOS Photoelectric Smoke Detector IC." The chip is designed for one thing: to sense reflected light from large smoke particles and generate an alarm by driving an external piezoelectric transducer. Because the ionization sensor chip does not

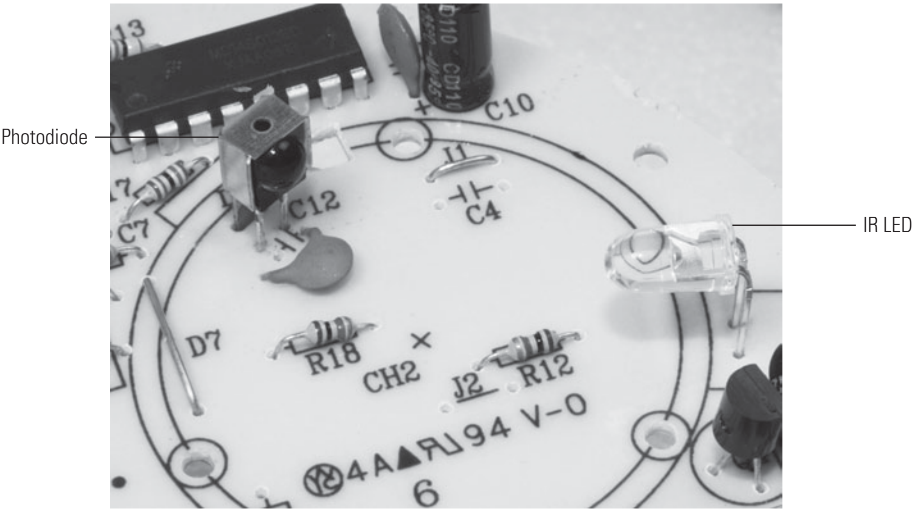

FIGURE 1-7 Photodiode and IR LED free of photoelectric sensor enclosure

have an onboard driver for the piezoelectric transducer, it has to piggyback on the photoelectric detector IC to sound the alarm.

The photoelectric sensor IC provides the logic for the smoke detector, from determining how long to remain silent after being hushed by the user, to when to flash the red LED, to how long to sound while the user is operating the test function. A handful of resistors and capacitors define the sensitivity of the chip. When hibernating, which is most of the time, the chip draws only 6.5μA.

Every 32 seconds, the chip awakens and sends a pulse to the C8050 NPN transistor and reads the photodiode. This gated IR LED pulsing and sensor reading not only minimizes power consumption, but it reduces the effect of noise from other light sources. This cycle continues until the photodiode detects IR light, at which time the chip activates its pulse/read circuitry every 2 seconds.

Piezoelectric Transducer

According to the manufacturer's product sheet, this piezoelectric transducer is capable of emitting a 3.5 kilohertz (kHz), 90dB tone at 12 inches when powered by a fully charged 9V battery.

Caution Because 12 inches is a typical working distance on a bench and 90dB is enough to damage your hearing permanently, it's a good idea to tape over the sound orifice to reduce the output level. Even so, limit exposure to the audio blast, for your sanity if not your hearing.

According to my measurements, the transducer draws only 10mA at full output, which is amazing, given the ear-splitting level. As you discovered during the teardown, the three connections, G, F, and M, connect to dedicated pins on the photoelectric sensor IC. Pin G is the metal support electrode, pin M is the ceramic electrode, and pin F is the feedback electrode. Figure 1-8 shows the pin connections for the paper-thin transducer, partially extracted from the transducer case.

The practical significance of the transducer pin designations is to insure the proper interfacing with the modulation and driver circuitry within the photoelectric sensor IC. As with the two ICs, this is a specialized component specifically designed for use in alarm systems. Recall that this isn't a buzzer, in that no oscillator circuitry resides in the transducer body. However, the transducer is self-resonating at 3.5kHz when properly driven. However, if it's connected to a driver and excited at, say, 2.0 or 6.0kHz, the transducer will likely just sit there.

C8050 Transistor

The IR LED in the photoelectric sensor is pulsed by the C8050 general-purpose silicon NPN transistor that is commonly used as an audio output amplifier in portable radios. This 1 watt (W) transistor, which is packaged in a TO-92 case, is driven by the photoelectric smoke detector IC. Although the IC is probably

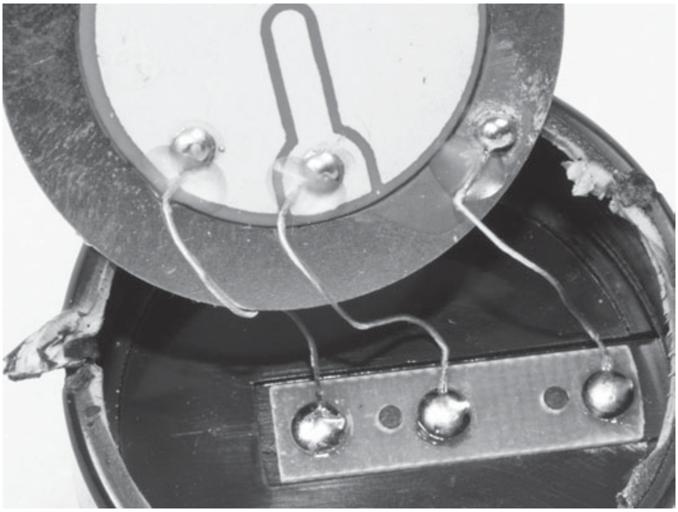

FIGURE 1-8 Piezoelectric transducer construction

capable of driving the IR LED directly, offloading the heavy lifting to the transistor increases the robustness of the circuit and probably extends the functionality of the IC at low battery voltages.

Diodes

This board contains power, switching, and zener diodes. A 1N003 silicon power diode placed across the battery leads where they enter the board insures correct battery polarity. If you accidentally reverse the 9V battery during installation, the diode will shunt the current, preventing the destruction of the two ICs, which are not reverse voltage protected. This isn't the most elegant use of a diode to insure proper battery polarity, in part because it may work only once, ending in the failure of the diode. However, it's efficient. The primary alternative is placing the diode in series with battery supply, but this drops the available voltage by the forward voltage drop of the diode to about 0.7V and wastes energy as heat.

The 1N4148 silicon fast-switching diodes are the most numerous diodes on the board. These glass-cased diodes can handle 200mA of forward current and 100PIV (peak inverse voltage).

The only zener diode on the board is in an unmarked glass package. A zener diode is designed to present a specific reverse breakdown voltage, typically much lower than a typical silicon power diode. The challenge of working with zener diodes is in limiting the reverse current with a series resistor so that the diode doesn't fail.

To determine the breakdown voltage of this zener, I removed the diode from the board and built a simple test circuit with a 4.5M series resistor. I hooked up the circuit to my bench power supply and slowly cranked up the reverse bias voltage across the resistor and zener diode while monitoring the voltage across the zener diode. With 5 to 10VDC (Voltage Direct Current) applied to the series circuit, the voltage across the zener diode varied between 2.5 and 2.8VDC. This test strongly suggests the diode is a 2.5V zener.

If you check the documentation for the ionization sensor chip, you'll see that a zener diode is specified to limit the voltage that can be applied to the negative comparator input pin. This limit is apparently 2.5VDC.

Capacitors

One electrolytic capacitor on the board, rated at 100 microfarad (μf) at 16VDC, is part of the transistor driver circuit. The remaining capacitors support the two ICs, as per the requirements listed on the datasheets. There include three ceramic disc capacitors and three axial lead epoxy-dipped capacitors, about the size of 1/8W resistors. Dipped capacitors are typically used instead of less expensive ceramic disc capacitors when specific capacitance values aren't available and when the circuit demands good temperature and humidity stability.

Switches

The switches in this smoke alarm are notable in that they are constructed of open spring steel that connects with a length of wire on the circuit board when compressed. Given that they'll probably be used a dozen times over the course of five years, someone made the economic decision to go with the cheapest possible switches. I can't think of a cheaper switch mechanism that works most of the time. These switches could be problematic if they become dusty, however.

How It Works

If you've followed the discussion of components to this point, you should have a good idea of the inner workings of the smoke alarm. As shown in the simplified schematic in Figure 1-9, a 9V battery supplies power and a silicon power diode (D1) protects against battery polarity reversal. The MC14578P chip (IC1) together with the ionization chamber forms the ionization sensor. A high-impedance operational amplifier (op amp) configured as a comparator within IC1 monitors current through the chamber.

In normal operation, alpha particles emitted by the Americium-241 ionize the air molecules within the ionization chamber. Electrons are knocked off of the atoms of gas in the chamber, forming equal numbers of positive and negative ions. Because one of the plates is held at a positive charge and the other at a negative charge, a small current flows as positive charged atoms flow toward the negative plate and negatively charged electrons flow toward the positive plate.

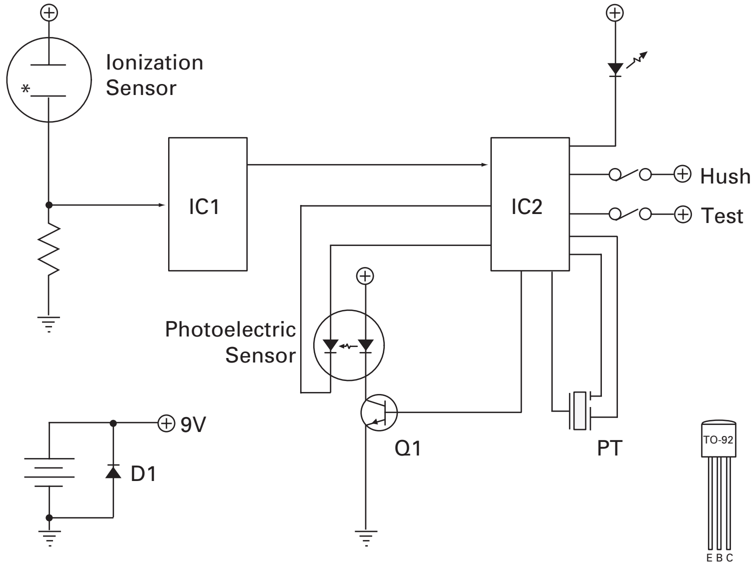

FIGURE 1-9 Simplified schematic

This ionization-induced current is stable until relatively massive smoke particles enter the cloud chamber. Smoke particles each absorb multiple ions, which reduces current flow. If the current flow falls below a predetermined level, an alarm is triggered. A limitation of this basic design is that the sensor current can be adversely affected by changes in humidity and atmospheric pressure, and this could affect the sensitivity of the alarm.

If smoke of sufficient density enters the chamber, the current decreases below a threshold established by the external resistors and capacitors connected to IC1. The chip sends an alarm condition to the MC146012ED photoelectric sensor chip (IC2).

When the air is free of smoke, IC2 periodically triggers the firing of the IR LED through the C8050 NPN transistor (Q1). If light is detected by the IR photodiode or if there is a trigger event from IC1, IC2 sends a 9V square wave to the piezoelectric transducer (PT), resulting in a loud alarm. IC2 also handles the logic of the hush and test buttons as well as the periodic flashing of the red LED. The data sheets for IC1 and IC2, which are readily available on the Web, define the signal levels and components needed for each pin and each function.

Note the outline of the transistor's TO-92 package in the schematic. It's a good idea to become familiar with this and other standard Transistor Outline (TO) packages because you can learn a lot from the package. For example, the standard TO-92 package defines the transistor leads as emitter, base, and collector, from left to right, facing the flat front of the transistor. The package size and composition also reveals a lot about the power rating of the transistor. For example, it's rare to find a transistor in a TO-92 package with a rating greater than 2W.

Mods

The Americium in the ionization sensor is an inexpensive, readily available source of radiation for testing radiation detectors, building a miniature cloud chamber, and performing other experiments using low-level radiation. However, unless you have the experience and equipment for monitoring radiation, it's best to avoid potential harm. Instead, let's focus on the IR photoelectric source and receiver.

External Alarms

If you download the datasheet for the IR photoelectric chip from Freestyle, you'll see that pin 7 is an input/output pin that can be used to activate escape lights, autodialers, and other electronic devices. You can experiment with this output to signal your PC through a serial or USB (Universal Serial Bus) port, or interface directly with a microprocessor.

Note I'm not advocating that you modify the smoke alarms in your home and then put them back into service. What I am suggesting is that you use a smoke alarm as an experimental platform for studying the underlying electronic circuitry. I'm sure that there's a law somewhere that imposes fines and prison time for anyone who tampers with a smoke detector, even in his or her own home. Besides, do you really want to entrust your life to an experimental circuit?

New Application Areas

The test setup shown in Figure 1-7 enables you to work with the LED and photodiode without the constraint of the light-tight container. You can also extend the leads of both components for remote sensing. For example, if you make each component waterproof, can you create a sensor to determine the turbidity of water? Could you use the detector circuit to create a pollen count detector?

Chapter 2

Motion-activated LED Light

Many electronic devices found in the home are interesting gadgets that briefly entertain and then become dust magnets. Other devices fulfill a real need, such as enabling you to navigate to the bathroom in the middle of the night without tripping over a toy or slipper. Of course, I'm referring to the motion-activated lights that turns on automatically when they detect your presence, such as the unit shown in Figure 2-1. The most useful of these lights employ battery power and an LED light source so that they can be placed anywhere they're needed, without regard to the availability of AC power.

Several sensor technologies can be used as a trigger mechanism for automatic lights, ranging from breaking a beam of light, to noise detection, to measuring

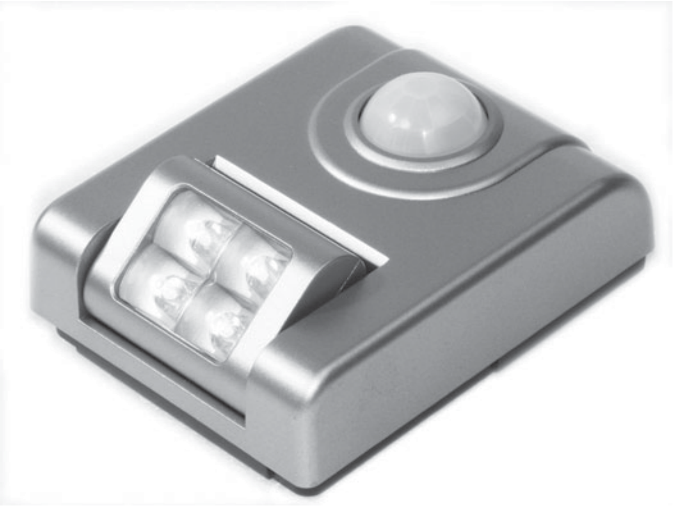

FIGURE 2-1 Fulcrum motion-activated LED light

fluctuations in a radiofrequency (RF) field; to detecting body heat, CO_2 concentration, or other products of respiration; to image recognition. Detecting body heat is a common, inexpensive sensor technology, in part because of the ready availability of passive infrared (PIR) detectors and dedicated support chips that minimize component count, manufacturing costs, and space requirements.

The focus of this teardown is the passive infrared Fulcrum motion-activated LED light shown in Figure 2-1. In addition to walking you through the circuitry, I'll suggest a few mods that will enable you to use the sensing circuitry of the battery-operated unit in additional applications.

Highlights

This teardown involves removing a few screws and extracting two circuit boards. The main board is populated by a variety of surface mount technology (SMT) and leaded components, including ambient light and PIR sensors; a dome-shaped Fresnel lens; a bank of white, high-intensity LEDs; a dedicated PIR integrated circuit; a voltage regulator; and an optoisolator. During the teardown, note the following:

- The quality of the solder joints on the circuit board
- SMT component markings
- Where leaded components are used instead of surface mount components
- Construction of the LED assembly

Specifications

Functionally, this is a relatively simple device. As such, it's not surprising that few specifications are provided by the manufacturer:

- The unit requires 3 AAA cells
- Light duration after triggering is 20 ±4 seconds
- Detection cone is 40 degrees at 10 feet
- The light consists of four white LEDs with an expected lifetime of 50,000 hours

We can deduce from the battery requirement that the device operates on 4.5VDC, assuming the AAA cells are wired in series. It's odd that the specifications include no mention of battery life, but we can measure the average current draw to determine the longevity of the cells under normal use. The detection angle and distance reflect the geometry of the optics and the sensitivity of the PIR sensor. LED lifetime is below average, but much better than equivalent incandescent lights. Overall, these specifications are typical for an inexpensive motion-activated light.

Operation

Once the unit is fitted with batteries and installed on a wall or other surface with double-sided adhesive tape, operation is automatic. If the ambient illumination level is low, the bank of LEDs is energized whenever the sensor detects a change in the level of ambient infrared (IR) radiation. A light sensor in the unit inhibits activation of the LEDs when the area of coverage is illuminated by daylight or an artificial light source. The LEDs remain active for about 20 seconds if no further change in radiation is detected. Otherwise, the LEDs stay on for as long as fluctuations exist in the level of ambient IR radiation.

The PIR sensor technology isn't foolproof, especially when it comes to false triggering. Any source of IR radiation can trigger the unit, from a dog or cat to a nearby heater, heat vents, or other heat source. Reflective surfaces near the sensor can redirect IR radiation from other areas and devices in a room. Fortunately, the worst-case scenario for false triggering is reduced battery life.

Teardown

You should be able to complete the teardown, illustrated in Figure 2-2, in about 10 minutes. If you're following along on your workbench with a battery-powered motion-activated light, you should encounter essentially the same sensor and control configuration. If you're working with an AC-powered unit, you may encounter an additional relay or other high-voltage switching component.

Tools and Instruments

The tools required for this teardown include a small Phillips-head screwdriver, an illuminated magnifier or strong magnifying glass, a multimeter, and forceps. If your circuit board was produced in the same plant as mine, you'll probably have to use the forceps to pull off mounds of excess glue to read component markings. Despite the glue and poor soldering work, with a little effort and a good magnifier, you should have no trouble tracing the circuitry.

Step by Step

There are no surprises in this teardown, other than having to remove gobs of hot melt glue from the main circuit board. Exercise caution when removing the optics so that you don't damage the mounting tabs, and avoid tugging on the fragile 28AWG (American Wire Gauge) wire connecting the battery holder to the main circuit board. Otherwise, the parts involved in the teardown are fairly indestructible.

If you're unfamiliar with the AWG system of classifying wire, now is a good time to get a feel for wire gauge. The smaller the AWG number, the thicker the wire.

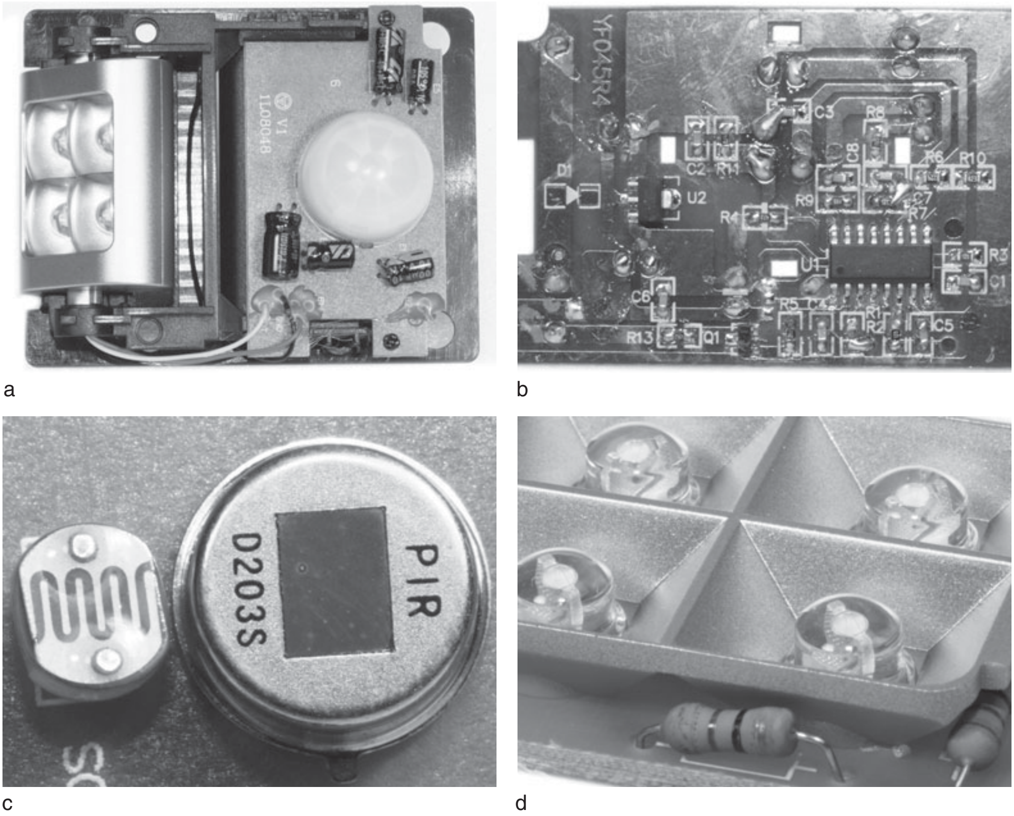

FIGURE 2-2 Teardown sequence

The exact diameter of a particular gauge isn't important, but it's critical that you develop an intuitive grasp for what's an adequate gauge for specific applications. For example, I wouldn't consider using an extension cord made with less than 14AWG copper wire. For standard hookup connections in an amplifier or other relatively high current circuit, I use 22AWG wire. When I'm dealing with robotic sensors and other weight-critical connections, I use 28 or 32AWG, depending on the physical stability of the circuit.

Step 1

With the unit face down, verify that the battery holder is empty and remove the four screws securing the back cover.

Step 2

Tilt the LED bank so that it is perpendicular to the body of the main unit. Pull straight up to separate the back from the top half of the body. Put the empty top half of the unit to the side. Position what was the bottom of the unit face up, as shown in Figure 2-2a. You can see the white Fresnel lens dome, surrounded by five electrolytic capacitors, on the main circuit board.

Also note the globs of glue on the bottom edge of the circuit board, where the wires from the LED bank attach to the board. Because the bank can be tilted, there's added stress on the wires where they attach to the board. The glue helps distribute this stress, thereby minimizing the odds the wires will break.

Step 3

Remove the two screws securing the main circuit board. Pop out the LED bank from the chassis. Flip the main circuit board over, as shown in Figure 2-2b. With the 16-pin IC to the right, you can see two, three-terminal SMT devices, labeled *Q1* and *U2*, and perhaps twenty SMT capacitors and resistors.

I took this photo after spending 15 minutes removing gobs and strands of hot melt glue. The glue not only obscured the components but also huge blobs of solder around some of the components. This is an example of poor SMT work that may have been the result of poor manual soldering technique, a dirty board, or the use of an automated solder machine in which the temperature of the solder was inappropriate.

Step 4

Working on the SMT component side of the board, use an ink pen or miniature screwdriver blade to push out the four plastic tabs securing the white plastic Fresnel lens dome. Turn the board so that the dome faces up. Gently remove the plastic lens from the top of the board. In the newly exposed area, you can see the PIR and CDS (cadmium sulfide) sensors, as shown in Figure 2-2c. Note the reflective window of the PIR sensor and the serpentine tracing on the face of the CDS sensor.

Step 5

Expose the second circuit board that forms the LED bank. Remove the black elastic bands from the stubs on either side of the bank. Pry open the case to reveal the circuit board holding four LEDs, four leaded 0.25W resistors, and the silver reflective assembly, shown in Figure 2-2d.

Layout

As revealed in the teardown, the layout consists of a 4.5VDC power source, a main board with sensors and other active components, and a small board with LEDs and resistors.

Now is a good time to hone your circuit tracing and component identification skills. Start by tracing the power and sensor signal paths with your ohmmeter. Next, insert three AAA cells in the battery pack and put the device through its paces.

What are the current and voltage demands when the device is on standby? What changes when the device is triggered? Any idea of what the three active components on the SMT side of the board are for? Try your best before continuing with this text.

Components

I have to admit that this teardown almost turned into a black box investigation. I spent several hours on the Web trying to find the datasheet on the 16-pin TL0001S integrated circuit. Initially, the best I could do was a spec sheet written in Chinese from the manufacturer in Taiwan.

In a parallel effort, I spent several hours trying to reverse-engineer the chip, working backward from the sensors and external resistors and capacitors. Then, by chance, I found a cross-reference to the BIS0001 IC, and then I was home free with a datasheet written in English. This is a good illustration of the value of looking for cross-references and equivalent components when you're hunting for a spec sheet.

TL0001S (BIS0001) IC

The TL0001S, which is equivalent to the BIS0001, is a special-purpose chip designed expressly to interface with PIR sensors as part of a motion-sensing circuit. In short, this is a plug-and-play chip designed for one thing. I checked out two other makes of PIR motion-sensing devices in my home, and both use the BIS0001 chip. The advantage of this chip over general-purpose chips and discrete components is vastly reduced component count and ease of manufacturing. Instead of mounting two or three op amp chips, a logic chip, three interface transistors for the sensors and output, and a host of bypass and biasing resistors and capacitors, a manufacturer can offer the same functionality with a single, affordable, space-saving chip.

The 16-pin TL0001S is packaged as a small outline integrated circuit (SOIC), as shown in Figure 2-3. Like the DIP (dual inline package) described in Chapter 1, pin 1 is marked with an indentation or other mark on the package top, and the pin numbering is counterclockwise from the top of the chip. This SOIC is about half the size of a 16-pin DIP.

The TL0001S contains five CMOS operational amplifiers, control logic, and dedicated interfaces for a PIR sensor, CDS sensor, and output relay or transistor. You can see in Figure 2-3 that the external component count is minimal. Most of the external components deal with biasing and control. You can modify the sensitivity, time on after motion ceases, and other chip parameters by adjusting resistor and capacitor values.

Two important parameters to consider are that the chip operates from 3 to 5VDC and can sink only up to 10mA. This is more than enough current to drive a single transistor or a solid-state relay, but it's not a typical mechanical relay and certainly not a bank of LEDs. If you look up the spec sheet for the BIS0001 on the Web, you'll find the suggested circuit is surprisingly similar to the circuit used in our motion detector.

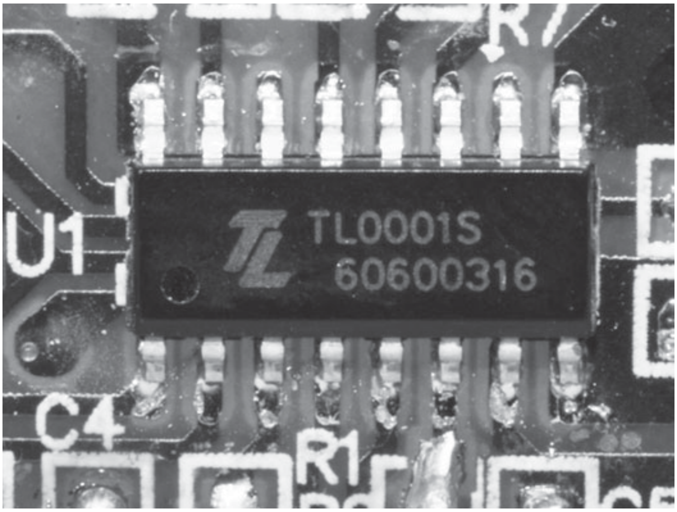

FIGURE 2-3 TL0001S/BIS0001 IC

Passive Infrared Sensor

The PIR sensor used in our motion-sensing light is another example of a special-purpose device that holds a formidable position in the market. The DS203 PIR sensor, by PIR Sensor Co., LTD, is one of the most popular PIR motion-detector sensors available. According to the datasheet, available from Futurlec (www.futurlec.com), behind the single reflective window (see Figure 2-2c) are dual sensor elements that are used to detect variations in radiation in the 5 to 14 micrometer (μm) range. External supporting circuitry for the three-terminal device is minimal. The device requires ground and power and presents its analog output on a single pin.

The PIR sensor has a number of impressive features, such as the use of dual-sensor elements to suppress interference from ambient temperature fluctuations. The sensor isn't perfect, however, especially when used in a high ambient humidity and temperature environment. High humidity is a problem because moisture in the air absorbs IR radiation from the source. Because the sensor responds to fluctuations in temperature, as the temperature of the environment approaches external body temperature, the IR signature from humans blends in with background radiation. In both cases, the range of the sensor suffers.

Fresnel Lens

The dome-shaped Fresnel lens is semi-opaque to visible light, which reduces false triggering from ambient visible light sources, but it is transparent to IR radiation. The annular patterns on the dome's inner surface, visible in Figure 2-4, are

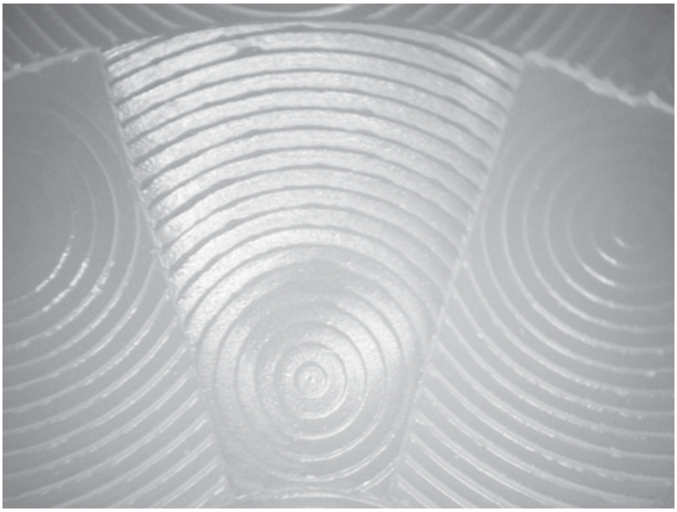

FIGURE 2-4 Fresnel lens close-up showing annular patterns

responsible for focusing the IR light on the window of a PIR sensor mounted at the focal point at the base of the dome.

If you've worked with a plastic Fresnel pocket magnifier, you know the main characteristics of a Fresnel lens compared with a conventional spherical lens. A Fresnel lens is generally thinner and lighter, made of inexpensive plastic, and yet provides significant magnification. The downside of a Fresnel lens is that the boundary between annular sections obscures the magnified image. That's why you don't see many camera lenses or reading glasses based on a Fresnel lens. However, in our application, image quality is irrelevant. We simply want to focus IR radiation within a specific area onto the window of a PIR sensor, and the inexpensive Fresnel lens does this superbly.

CDS Sensor

The CDS sensor, sometimes referred to as a photoresistor, is probably the most common ambient light sensor in use today (see Figures 2-2c and 2-5). The resistance of the CDS sensor in our motion detector changes from more than 1 million ohms (MΩ) in relative darkness to less than 100 ohms (Ω) in full sunlight. From a component perspective, you can think of the sensor as a two-terminal variable resistor. CDS sensors have been supplanted in many applications by more sensitive and responsive phototransistors.

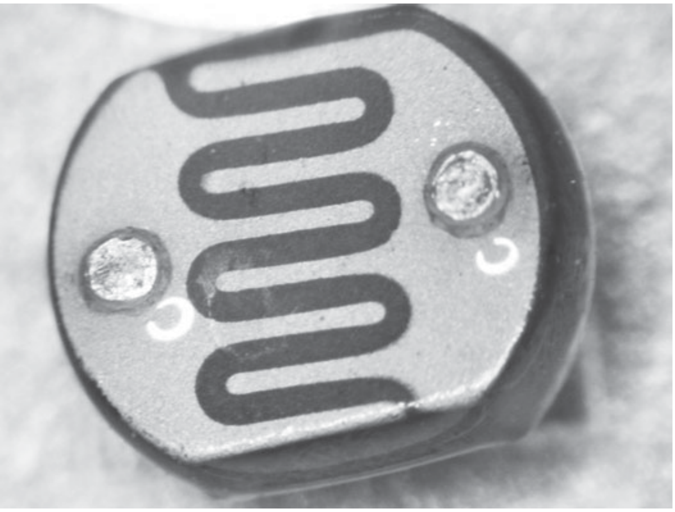

FIGURE 2-5 CDS sensor

7130 Voltage Regulator

The three-terminal 7130 is a 3V at 30mA voltage regulator. It can handle input voltages more than three times that supplied by the three AAA batteries used in our motion detector. More important, this is a *low-dropout regulator*, meaning that the regulator can function with just over 3V input from a nearly depleted battery.

As shown in Figure 2-6, the regulator is packaged in a 3-pin SOT-89 case, with input voltage on the middle pin, which is connected to the tab. In keeping with the SOT-89 standard, the top pin in the figure is connected to ground and the 3.0V output is taken from the bottom pin. If you trace the power from the battery, you'll find one wire to the voltage regulator input and another that goes directly to the bank of LEDs.

NPN Transistor

Near the edge of the main board, between the 16-pin chip and the wires leading to the small circuit board, is Q1, an NPN transistor in an SOT-323 package (see Figure 2-7). The SMT transistor is used to switch on power to the bank of LEDs, as a function of the output of the BIS0001 chip.

As with the voltage regulator, you can use the standard package configuration to identify the transistor leads. With the transistor oriented with the single, center lead to the left, as shown in Figure 2-7, the base lead is on the bottom right, the emitter on the top right, and collector center left. The SOT-323 package is a miniature version of the SOT-23 package, which you'll see plenty of in subsequent teardowns.

If you study Figure 2-7, you'll appreciate my regard for the quality of soldering in this device. The solder pad for the emitter lead of the transistor is nearly touching the trace immediately above it, thanks to the surplus solder on the pad. The

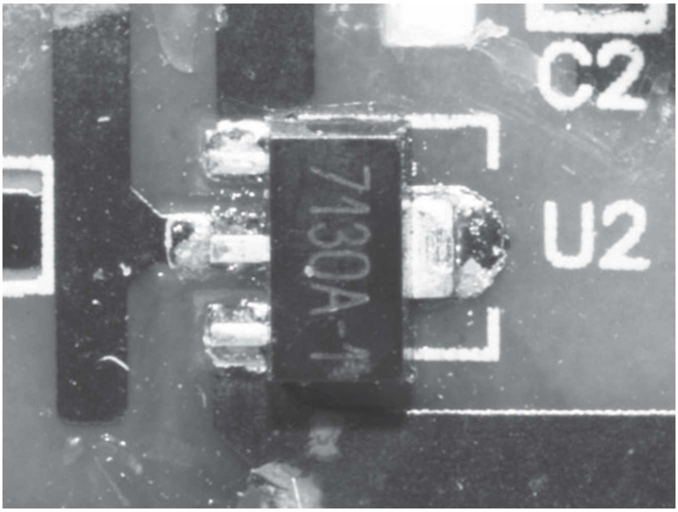

FIGURE 2-6 7130A-1 voltage regulator

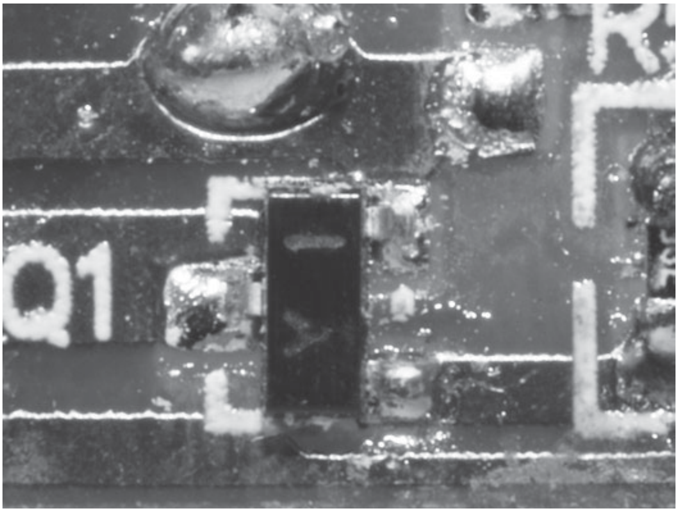

FIGURE 2-7 Q1 SOT-323 package

numerous speckles on the board are remains of large globs of glue that I removed with forceps.

LEDs

White LEDs, once the rage in a world limited to greens and reds, are now available for pennies. The four LEDs used in this motion detector are moderate intensity LEDs of the type typically found in portable reading lights. Each of the four LEDs and its 27Ω series resistor is connected in parallel.

LEDs emit light when forward biased around 2V, with the exact voltage dependent on the diode design. Except for high wattage LEDs, connecting an LED directly to a battery typically results in a very brief flash of light and a burned-out LED. In our motion detector, the 27Ω series resistors limit the current through each LED to about 25mA, which is typical for white-light LEDs used for illumination.

You can calculate the value of the current-limiting resistor for any LED and voltage source with the following equation:

$$R = (VS - VD)/I$$

where *VS* is the source voltage, *VD* is the diode forward voltage drop or operating voltage, and *I* is the operating current. You'll also find several good LED current limiting resistor value calculators on the Web.

Capacitors

The motion detector circuit relies on both leaded electrolytic and SMT ceramic capacitors. The relatively large, leaded electrolytic capacitors are probably used in this application because they're less expensive than their SMT counterparts.

Note If you want to practice reworking an SMT board, the ceramic capacitors are a great target for removal and replacement.

Resistors

The resistors on the main board are standard tolerance, 1/8W, SMT resistors. The 0.25W leaded resistors are used as current-limiting resistors for the LED bank instead of more space-efficient SMT resistors because of the higher wattage requirement.

Battery

The motion detector draws about 1mA when idle and about 138mA with the LEDs energized. Given a typical alkaline AAA is rated at about 1 ampere-hour (Ah), how long will the battery last if the unit is perpetually idle? What about perpetually on? Here's the basic formula for amp-hours:

$$Ah = I \times T$$

where *I* is the current in amps and *T* is the time in hours. For example, a 1Ah battery can theoretically supply 1A for 1 hour or 500mA for 2 hours. In practice, the formula and rating break down at high current and long time intervals. It's safe to say that a 1Ah AAA cell can supply 100mA for 10 hours, but not 100A for 0.01 hour. Similarly, a cell has a finite shelf live. A 1Ah AAA cell won't supply a 1µA circuit for a century.

If you want to verify your Ah calculations, place your ammeter probes across the collector and emitter of Q1. Use a timer and note the starting time and the time when the LEDs are no longer illuminated. If you want to shorten the time of the experiment, use generic carbon zinc cells with a rating of 0.4Ah. On the Web, you can find the rating of standard cells by chemistry, voltage, and manufacturer on Digi-Key (www.digikey.com) or other electronics parts dealers' sites.

How It Works

The sensor technologies that form the basis for most motion-activated lights, cameras, and security systems include active IR, active ultrasound sensing and ranging, and passive IR (PIR) detection. The operation of active IR and active ultrasound sensors is similar to the operation of radar in that a beam of energy is transmitted and the reflection is received and analyzed for changes in timing, amplitude, or phase. If the time between successive reflections decreases, this suggests that the distance between the sensor and an object in the sensor's active area is decreasing. If the amplitude or phase of the reflected signal changes, this suggests that something or someone is interrupting the transmitted or reflected beam.

PIR detection is based on relative changes in the temperature of an area being monitored, based on fluctuations in the level of infrared radiation. Everyone and everything not at absolute zero (–459.67°F or –273.15°C) emits radiation, and some of this radiation is within the infrared spectrum, which spans the range of about 750 nanometers (nm) to 100µm. Different objects and biological systems have different spectral fingerprints, just as objects vary in visible light absorption or perceived color. At normal body temperature, our IR fingerprint has a peak around 10µm. As such, an IR detector that is sensitive to 10µm radiation should be good for detecting the presence and movement of humans.

Each of these methods of motion detection has practical benefits and drawbacks. Passive detection is characterized by relatively low power consumption, for example. A battery-powered PIR-based motion detector can operate for months between battery changes. Moreover, a battery-powered unit is more desirable than an AC-powered monitor in a security system, because AC power can be intentionally disrupted.

Active approaches to motion detection generally have longer range and greater sensitivity, and they are more difficult to evade. For example, because a PIR detector is sensitive to rapid changes in relative temperature, a burglar could avoid detection by moving slowly through an area wearing a heavily insulated jacket and hood. Active IR motion detectors, in contrast, aren't fooled by heat-absorbing insulation. A new object in the path of an active IR pulse results in a new reflection, which can be used to trigger an alarm.

Active ultrasound methods are also impervious to ambient light, and so can be used outside in the daylight. In addition, the range of active ultrasound motion sensors is 30 feet or more—more than twice the range of a typical IR motion detector.

Figure 2-8 shows the simplified schematic of our motion-activated light. Note the package outlines for the 7130 fixed voltage regulator (SOT-89) and the much smaller Q1 (SOT-323). The package outlines are not drawn to scale.

Starting at the upper left of the schematic, the battery supplies 4.5VDC directly to the 7130 3.0VDC fixed voltage regulator and the physically separate bank of LEDs. No on-off switch or fuse lies between the battery and other components.

One reason for the lack of fusing is the relatively low power available from the three AAA cells. Another is that the 7130 voltage regulator is virtually immune to damage from component failure or battery voltage reversal. The fixed voltage regulator can tolerate a short on the output and will protect the components from reverse voltage. A 100µf at 10VDC electrolytic capacitor is used on the output of the voltage regulator to enhance regulation.

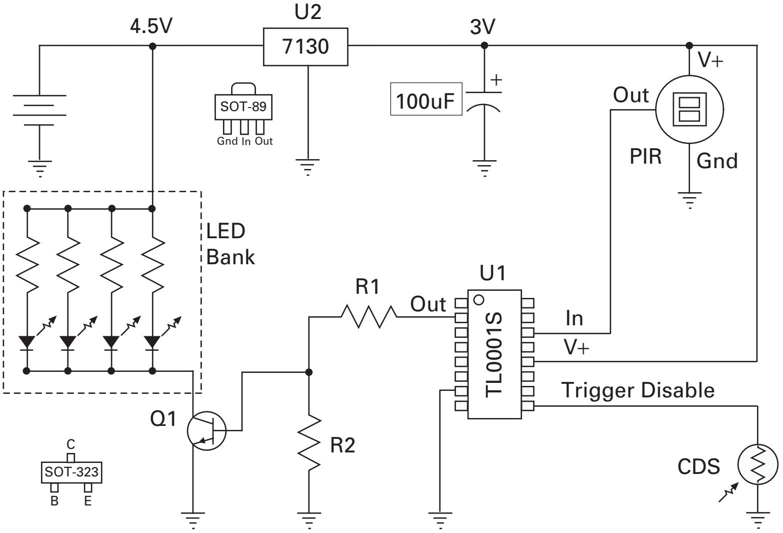

FIGURE 2-8 Simplified schematic of motion-activated light

The circuit lacks a manual on-off switch because the manufacturer can get away without installing one, thereby saving the engineering, manufacturing, and component cost. There's also one less item to fail. A switch isn't absolutely necessary because of the low standby current requirements of the CMOS TL0001S chip and PIR detector, which amounts to only a few microamps. Even when both are triggered, the two components draw less than 100μA.

The other components that could draw down the battery, the bank of LEDs, are similarly insignificant sources of current drain. The switching resistor (Q1) is normally reverse biased, with the base at ground potential, resulting in an essentially open collector-emitter connection.

Although there is no manual on-off switch, the CDS sensor performs an equivalent, automatic function. When the ambient light level is high, the resistance of the CDS is low, and the normally high trigger disable pin of the TL0001S chip is pulled low. This shuts down the chip, making it unresponsive to input from the PIR sensor.

When it's dark, the resistance of the CDS is high, which allows the trigger disable pin of the TL0001S to return to its normally high state. The chip is now responsive to signals from the PIR sensor. When someone walks within the detection range of the PIR sensor, the person causes a fluctuation in the output of the PIR sensor on the order of a few millivolts (mV). This signal is amplified and processed by the logic within the TL0001S chip.

If the fluctuations in PIR output are significant, the normally low output pin of the TL0001S is driven high, and current flows through R1 and R2. The positive voltage drop across R2 forward biases Q1, providing a low resistance ground connection for the four LEDs, which draw a total of 40mA at 4.5VDC.

When the fluctuating heat source moves out of the detection range of the PIR sensor, the sensor output signal will become quiescent in 20 seconds. The TL0001S output pin goes low, as does the base of Q1. Without forward bias, the collector-emitter path returns to a high resistance state, and no current flows through the LEDs. The resistor (R2) from the base of Q1 to ground serves to guarantee the base voltage doesn't float above ground due to noise.

Mods

Because the dedicated chip simplifies the overall circuit, mods are limited but also greatly simplified.

Narrowed Trigger Zone

The quickest mod is narrowing the trigger zone. Let's say you have a hallway with exits to the left and right, and you want to limit triggering to when someone walks to the right. Assuming the 40-degree cone of detection is too broad, you can easily narrow the detection angle to a few degrees by fitting a tube over the Fresnel dome. Take the cardboard tube from a roll of aluminum foil or plastic wrap, cut a 2-inch length with a blade, and paint the inside flat black to absorb stray IR and visible light. Paint the outside silver or the color of your particular motion detector. Mount the tube around the dome with tape or a dab of silicone glue.

You might want to try a few different tube lengths before you glue the tube in place. The longer the tube, the narrower the detection zone. I've tried tubes up to 4 inches long with a successful detection range of about 5 feet.

External Device Controller

You can use the motion detector to trigger anything from a 120VAC light to a speech synthesizer on a mobile robot. Simply take the two wires leading to the LED bank and run them to a solid state or electromechanical relay capable of handling the voltage and current in your circuit. With an isolated output, you can connect anything you like to the output contacts of the relay, without regard for the components and circuitry in the motion detector.

If you opt for an electromechanical relay, make certain the motion detector can meet current demands of the relay coil. Keep the load on Q1 at or below 160mA, which is the load imposed by the LED bank.

You could simply connect the wires intended for the LED bank to the load of the second device. Assuming the device runs off of DC, you can use a battery supply up to about 24VDC, the maximum input to the 7130 voltage regulator. You may have to replace R1 and R2 with a 25KΩ potentiometer, however, because the bias setting will be different with a higher collector-emitter voltage.

Adjustable Light Sensitivity

You can adjust the ambient light sensitivity, time delay, and other basic parameters of the TL0001S by changing the values of the external resistors and capacitors. See the datasheet documentation for details. However, if you don't want to work with SMT components, you can make simple mods within minutes. Want the detector to work in a higher ambient light condition—maybe even daylight? Simply remove the dome and tape over the CDS sensor with black electrical tape. If you want a permanent fix, cut one of the leads to the CDS sensor and replace the dome.

This page intentionally left blank

Chapter 3

Digital Bathroom Scale

Digital bathroom scales, such as the one shown in Figure 3-1, have evolved significantly since they were introduced in the consumer market more than a decade ago. My first all-electronic digital scale had an aluminum chassis about 4 inches thick, weighed 10 pounds, and had a 12VDC external power supply. The cost for this space-age marvel was more than $200. Unfortunately for the average consumer, the market was flooded with cheap knockoff scales that had a digital readout but retained the relatively inaccurate spring-based weighing mechanism.

Today, inexpensive, svelte, feature-laden digital scales with features from body composition to Bluetooth connectivity are the norm. It's actually difficult to find an old-fashioned analog spring steel bathroom scale. After all, why would anyone want a scale with poor repeatability that's based on a mechanism that's destined to rust in a constantly humid environment? Admittedly, balance beam scales—like those used

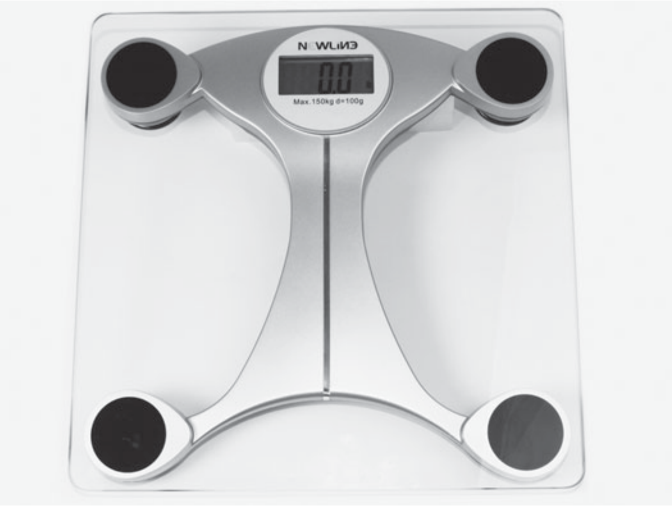

FIGURE 3-1 Newline Well Balance bathroom scale

in doctors' offices—are accurate and can last indefinitely, but they take up valuable space in a home bathroom. It's easy to slide a modern digital scale under a cabinet when it's not in use.

In this chapter, I'll tear down a Newline Well Balance Digital Bathroom Glass Scale (www.newlinescale.com). This scale doesn't talk to you in synthesized speech or indicate your percentage of body fat, but it does measure your weight in pounds or kilograms quickly and with good repeatability. In addition to exploring the overall design of this Chinese-manufactured, $25 scale, we'll review the fundamentals of strain gauge sensors and investigate the scale's unique hibernation and wakeup system.

Highlights

The star of this teardown is the strain gauge, which is the basis for accurate, reproducible measurements. The scale also employs a unique vibration sensing mechanism to activate the scale.

This teardown also features a mystery microcontroller entombed under an unmarked black blob of epoxy. This form of encapsulation is a common method of protecting naked ICs that are wired directly to the circuit board. In effect, the underlying circuit board is part of the IC. Epoxy encapsulation is an economical mounting option for manufacturers that work with multiple custom chips but a pain for curious hobbyists and a speed bump for competitors attempting to reverse-engineer a chip.

During the teardown, note the following:

- The black blob entombing a mystery microcontroller
- The piezoelectric transducer
- The soft, spongy connector between the LCD (liquid crystal display) and the circuit board
- The location, structure, and wiring of the strain gauges

Specifications

Following are the key specifications offered by Newline:

- Strain gauge sensor system
- 330 pound capacity
- Foot tap switch
- 1 inch LCD
- Auto zero resetting/auto power off
- Low power/overload indicator
- Powered by a pair of 3V CR2032 lithium cells

Of note is that there is no mention of accuracy on the Newline web site or in the material accompanying the scale.

Operation

If you check the user feedback on Amazon.com for this scale, you'll see that many people have trouble operating it. This seems paradoxical at first, because this scale is about as simple as it gets. The only setup is selecting pounds or kilograms, using the switch on the underside of the scale. Unlike more advanced scales, there's no age or height data to enter, no memory selector to store the weight for different family members, or any of the other popular features. However, there's no manual on–off switch—the source of the complaints.

According to the manual, to activate the scale, you simply tap it with your toe. In practice, you have to lift the edge of the scale an inch above an uncarpeted floor and then release the scale. If you were sufficiently rough in handling the scale, the LCD display comes to life, indicating that you can step on the scale. I haven't had any problem with the scale, but subjecting the scale to a 1-inch drop onto a ceramic tile floor every day can't be good for the strain gauges.

Teardown

You should be able to complete the teardown of the scale, illustrated in Figure 3-2, in about 15 minutes. The work entails extracting the load cells—the strain gauges and their mechanical mounts—from the four feet and the display and removing the processing unit from the head of the scale.

Tools and Instruments

If you're following along at home with a digital scale, you'll need a good magnifier and forceps in addition to the usual screwdriver, soldering iron, and multimeter. A digital multimeter (DMM) with a high-resolution, accurate ohmmeter function will help when we examine the strain gauges in detail.

Step by Step

Before you dismantle the scale, it's a good idea to test the scale and study the front display, shown in Figure 3-2a, so that you can appreciate the underlying technology. How's the repeatability of your scale? If you stand to one corner of the scale, is the reading different from when you stand with your weight evenly distributed among the four feet? If you have the same model of Newline scale, can you easily activate it?

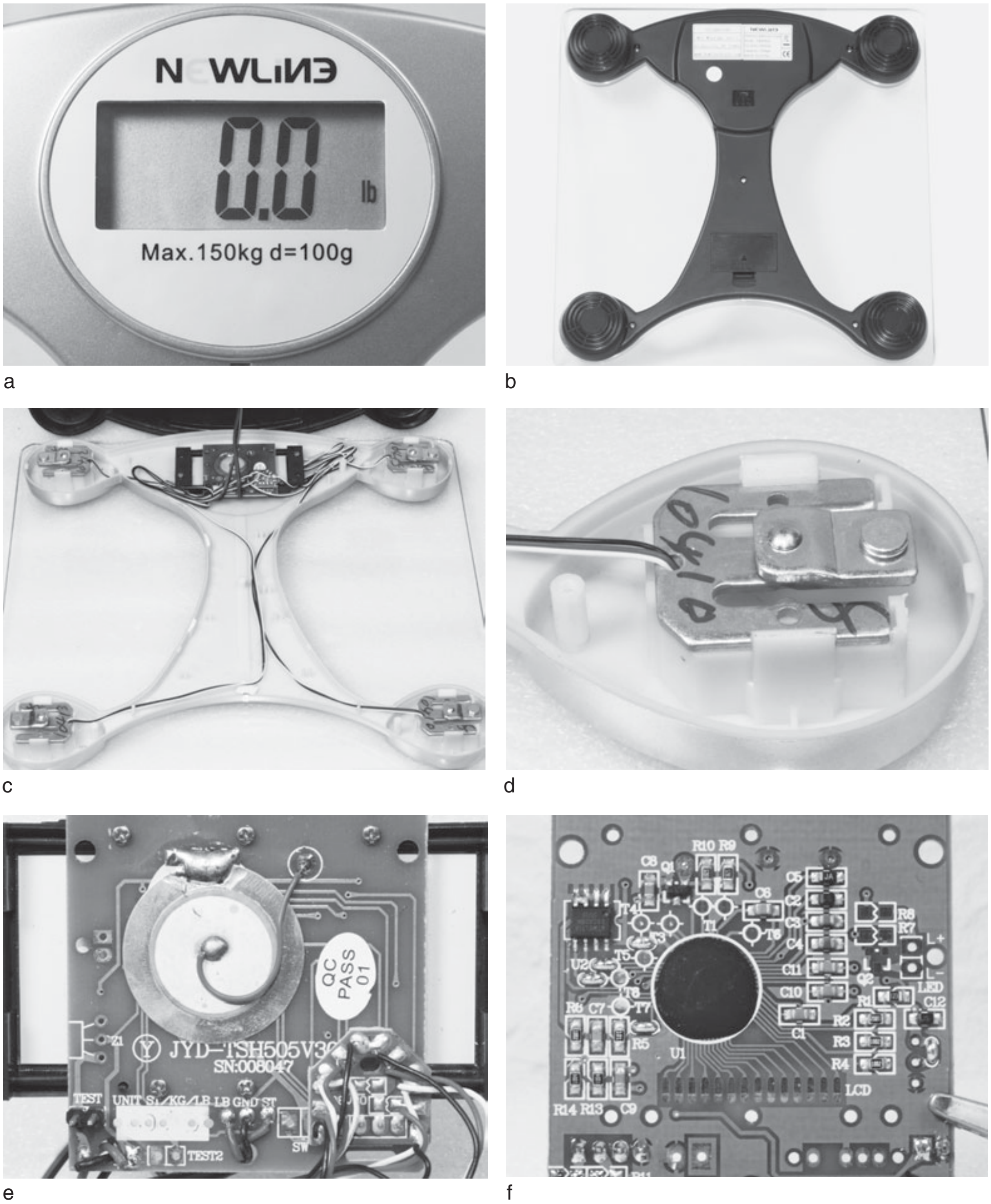
NEWLINE
0.0
lb
Max.150kg d=100g
a
b
c
d
e
JYD-TSH505V3
SN:008047
QC PASS 01
TEST
UNIT
KG/LB
LB
GND
ST
TEST2
SW
f
U1
LCD
LED

FIGURE 3-2 Teardown sequence

Step 1

Remove the two 3V lithium cells. Place the scale face down, as shown in Figure 3-2b, and slide open the battery compartment in the lower half of the scale. Remove the two CR2032 cells and set them aside; you'll need them later.

Step 2

Remove the back of the scale. Remove the five Phillips-head screws from the back—one in the center and one over each of the four arms. Gently pry the black plastic back from the white plastic front, working from the end farthest from the display. Hinge the back away from the remainder of the scale, as shown in Figure 3-2c, taking care not to rip the cables to the pounds/kilograms switch and the battery pack that remain with the black plastic back. Note that the battery pack is wired so that the lithium cells are in parallel, and that the voltage supplied to the processing and display assembly is 3V.

Step 3

Extract the load cells from each of the four feet. Each load cell, such as the cell shown in Figure 3-2d, snaps out of place with a little finger pressure. Glued to the bottom of each load cell is a strain gauge that's connected to the main control unit via a red, white, and black wire.

Step 4

Examine and test each load cell. The load cells in this scale are three-pronged steel devices. A thin foil strain gauge proper is glued to the underside of each cell, and the wires are held in place with a dab of silicone sealant.

Insert the lithium cells in the scale back and tap on the display to activate the scale. Apply pressure to one of the load cells by pushing down on a load cell with the palm of your hand. A weight value should appear on the display.

Now test the load cells. Can you tell if the weight measurement is additive? You have to twist or add pressure to a load within a few seconds of activating the scale or the scale will power down automatically.

Repeat the process with an ammeter in series with the 3V battery pack to determine the current load. You'll have to cut or unsolder one of the wires leading from the battery pack and insert your ammeter in series with the wire. I read 4mA when the display is active and about 5µA when the scale is in standby mode. Remove the lithium cells from the back of the scale when you've finished experimenting.

Step 5

Remove the LCD panel assembly by lifting it straight out of the scale front. Note the main circuit board attached to the display back, as shown in Figure 3-2e. Note the prominent, quarter-sized piezoelectric transducer soldered to the back of the circuit board. The small, six-sided connector board in the lower-right corner of the figure is connected to each of the four load cells. Remove the four-pin connector now.

The wires from the battery pack enter at the lower-left corner of the board. Along the lower edge of the board, between the power and strain gauge connector, is the pounds/kilograms switch.

Step 6

Remove the LCD from the circuit board. Remove the six screws holding the board in place. If you didn't remove the four-pin strain gauge connector in the lower-right corner of the board, you won't be able to get at one of the screws. The LCD panel should fall away from the board. Carefully store the LCD panel and the soft pink connector for use later.

Step 7

Examine the back of the board. Flip the board over and orient it to match Figure 3-2f. To the left of center is a black blob—a 1/2 inch–diameter mound of epoxy that protects a microcontroller and associated wiring. You'll see an SMT (surface mount technology) transistor and 8-pin EEPROM (electrically erasable programmable read-only memory) chip in the upper-left corner of the board. About a dozen SMT resistors and capacitors are distributed to the left and right of the blob. In the lower center of the board are the multiple fingers that mate with the LCD module through the soft connector.

Layout

The teardown revealed a simple and straightforward layout. The four peripheral strain gauge sensors, the weight system switch, and the battery pack are all connected to the central processing and display unit. The double-sided board has numerous markings for test points and markings for components used to drive an LED, which is apparently for some other model of scale.

Components

The real stars of this teardown are the four strain gauges that make this all-electronic scale possible. The flexible coupler between the LCD and the circuit board is worth exploring as well. Although the onboard microcontroller is unmarked, we can deduce the functionality from the product specifications and other hardware on board.

You can use a few SMT resistors and capacitors on the board to practice removing and installing SMT components. The resistors are stamped with values—see Appendix A if you need to brush up on your resistor identification. Without an official schematic, the quickest way to determine capacitor values is to use a capacitance meter with a forceps probe.

Microcontroller

The blob of black epoxy is a microcontroller, because it has analog-to-digital conversion capabilities, on onboard oscillator, timers, and a driver for the LCD. The next best alternative, a microprocessor, doesn't fit the specifications. Microprocessors may be computationally superior to microcontrollers, but they have limited input/output capabilities.

Given the functionality of the scale, we can assume the microcontroller provides the analog-to-digital (A–D) conversion of the strain gauge signals, performs the pounds-kilogram units translation, shuts down the scale after a period of inactivity, powers up the scale when the scale is jostled, and drives the LCD. The peripheral EEPROM, discussed next, suggests that this is an 8-bit microcontroller.

If you want to see what's under the blob, you can use two practical ways to get at the underlying chip: either chemically etch the epoxy or soften the epoxy with an SMT hot air pencil and remove the crumbly mass with a small blade. A third approach that seems to be popular on the Web is to use a rotary tool and rotary file to wear away at the blob until you hit something interesting. In my experience, getting physical with the epoxy results in a small pile of unrecognizable pieces of silicon and epoxy.

Caution If you want to maintain the integrity of the small gold or copper wires between the chip and the circuit board, you can try to dissolve the blob with ATTACK epoxy solvent. (An 8-ounce container is available through online retailers such as Amazon.com for about $10.) The solvent works exactly as advertised, but it's also extremely hazardous. Don't even think of using it indoors, around life forms you value, or without skin protection. The U.S. Postal Service won't even ship it. And I doubt that the board will be useable after your apply the solvent—but it will look good.

The approach I recommend for novices is to simply heat the blob to about 900°F with a hot air pencil. At that temperature, the epoxy behaves like soft clay, and you can easily pick away at it with a hobby knife, scalpel, or other small, sharp instrument. Some of the traces may be removed during the process, but the chip should come out unscathed.

I used the thermal approach to create Figure 3-3, which shows the microcontroller chip sitting on the much larger rectangular pad. To improve the clarity of the image, I superimposed a high-resolution photo of the newly uncovered area onto the original circuit board image shown in Figure 3-2f. Otherwise, the images are untouched.

Visualizing the silicon chip probably doesn't tell you much about the architecture or functionality of the microcontroller. However, a microcontroller designer can tell a lot from viewing the chip, because each section—the mathematical unit, onboard RAM (Random Access Memory), and various registers—has a characteristic appearance. For our purposes, getting a glimpse of the microcontroller is simply part of the fun of a teardown.

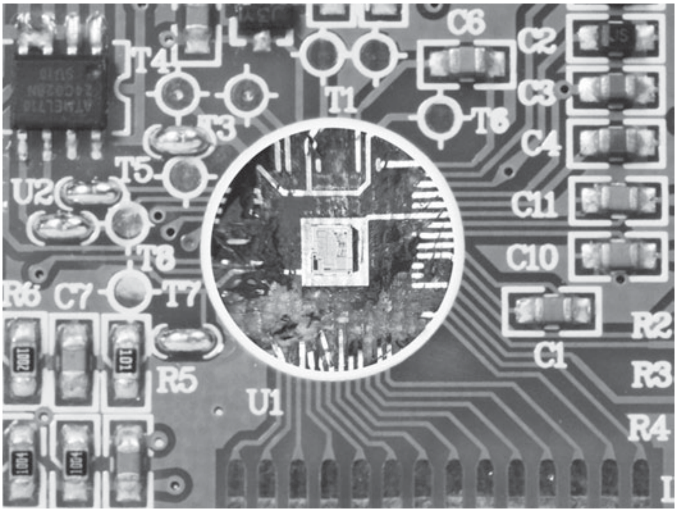

FIGURE 3-3 Circuit board after removal of epoxy blob

EEPROM

The ATMEL 718 24C02BN EEPROM provides 2K of memory, organized with 32 pages of 8 bytes each. This is an 8-bit component, strongly suggesting that the microcontroller is an 8-bit device as well. The microcontroller apparently requires external memory to store the program code required to direct the operation of the scale. There is no evidence that the EEPROM can be erased and reprogrammed by the microcontroller in the field.

The specifications of the chip include a two-wire serial interface and 2.7VDC operating voltage. The two-wire interface agrees with my findings on the board, in that only a few wires connect the microcontroller and EEPROM. The operating voltage is compatible with the supply voltage available in the scale.

It's always a good idea to check the specs against the real-world configuration of a device. In some instances, you may have mischaracterized a device or made incorrect assumptions about the circuit configuration.

Strain Gauges

The weight sensors on this scale are thin foil strain gauges. Each of these gauges is glued to a metal, weight-bearing structure in each of the scale's four feet. Recall that a strain gauge and its supporting, weight-bearing structure are referred to as a *load cell*. Review photos of the load cell in Figure 3-2d, and, if you're following along with your own teardown, examine each load cell carefully. Note how each strain gauge is positioned exactly over the pivot point on the metal support where a metal extension is riveted to the main body.

Both the orientation and placement of each strain gauge on its supporting structure are significant. You can get a better appreciation of the placement of each strain gauge if you remove the white silicone goop from the underside of each load cell with your fingernails. Figure 3-4 shows a strain gauge still glued to the bottom of a weight-bearing structure. The wires were removed for clarity. The three prongs of the load cell open to the right, and the wires normally connected to the sensor, from top to bottom, are white, red, and black.

The metal extension on the support has the effect of subjecting one side of the strain gauge to compression and the other to elongation. This deformation is critical, because strain gauges respond to strain. Recall from basic physics that *strain* is the fractional change in length of a body that is subject to an applied force. Each of the strain gauges in our scale is oriented with the zigzag pattern running from left to right, which is the direction of strain. When you stand on the scale, the strain gauge element on the outside membrane is stretched, the metallic foil pattern is thinned, and resistance increases. Simultaneously, the inside element is compressed, the foil is thickened, and resistance decreases. Because of the zigzag pattern, the relative change in resistance is multiplied relative to what a single conductor would exhibit.

You can confirm these assertions with your DMM. If you check the resistance between the three solder pads of a strain gauge, you'll find that each side has a resistance of about 1KΩ, or 2KΩ when both sides of a strain gauge are measured in series. If you return the load cells to their original housings and measure the resistance between the three leads, you'll find that with no load, the resistance from

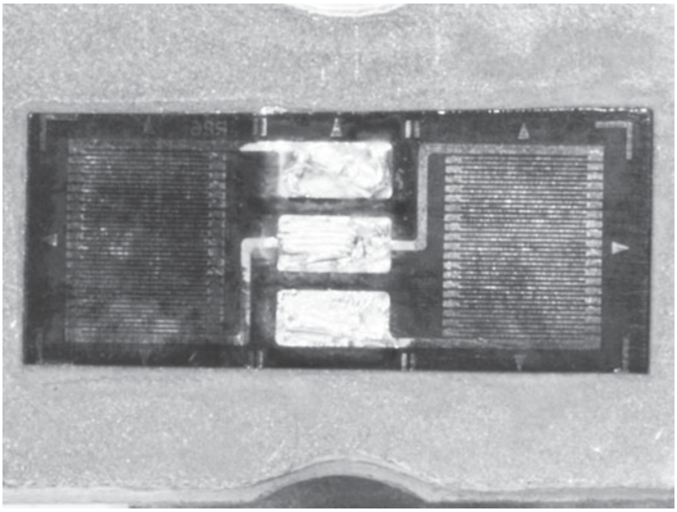

FIGURE 3-4 Foil strain gauge glued to metal pod

the center to either side of a strain gauge is 1KΩ. When you stand on the scale, the resistance of the outer side of each strain gauge should increase by 2 to 4Ω, and the resistance of the inner side should decrease by 2 to 4Ω. Clearly, the position of each strain gauge in its metal pod is such that one side is subject to compression and the other elongation. This finding has major implications, as discussed a bit later.

The metal foil strain gauges used in this scale are popular because they're inexpensive and relatively robust. However, if this were a laboratory scale instead of a bathroom scale, the strain gauges would most likely be made of semiconductors, which are more sensitive, more fragile, and more expensive than metal foil strain gauges. Both types of strain gauges are affected by ambient temperature. However, metal foil strain gauges can be made relatively immune to temperature changes by using alloys that minimally contract or expand with such changes.

Piezoelectric Transducer

The piezoelectric transducer soldered to the back of the main circuit board is a generic component typically used with an oscillator to create a buzzer. In this application, the transducer is used as a sensor. Vibrations caused by dropping the scale are converted to electrical impulses that are used to activate the scale.

Can you think of why the designers decided to add another sensor to the scale when they already had four strain gauges to detect someone stepping on or dropping the scale? It's probably because monitoring the strain gauges requires significant processing power and energy from the battery. It's more efficient to have the microcontroller hibernate until awakened by a pulse from the piezoelectric transducer. An even more efficient, but less attractive, alternative is a mechanical switch, but this solution has limited market appeal.

J3Y Transistor

The output of the high impedance piezoelectric transducer is insufficient to drive the microcontroller directly. The J3Y transistor, an NPN transistor in an SOT-23 case, provides the needed amplification and impedance matching for the microcontroller. The transistor, which is equivalent to the S8050LT, has a collector current rating of 0.5A and a dissipation of 0.3W.

Note In case you're wondering, I don't have an encyclopedic knowledge of transistors. I found the current rating, dissipation, and other specifications of the J3Y transistor the old-fashioned way—by searching the Web. I've had the best luck with Hong Kong supply houses that list transistors and their equivalents. Often the equivalent is easier to locate on yet another supply house list. Make certain you include the full transistor designation on your search. For example, if you search for "8050 transistor," you may find the datasheet for the NPN transistor in a leaded SO-92 case, which has a dissipation of 1W.

Even without a datasheet, identifying the basic transistor parameters is relatively simple. For example, I know the identity of the leads from standard small outline transistor (SOT) packaging. Looking at the component with the single lead on top and two leads on the bottom; the collector is on top, the base lead is on the bottom left, and the emitter is on the bottom right. Following this pinout, the emitter is grounded, the base is connected to the piezoelectric transducer, and the collector is connected to a microcontroller line. The collector is also connected to the 3V battery supply through a resistor.

As discussed in Chapters 1 and 2, given a transistor in an SOT-23 case without a datasheet, I know the emitter, base, and collector leads. It's up to you to decide whether the device is an NPN or PNP transistor. One way to make this determination is to examine the transistor with a multimeter. If you have an NPN transistor, with the positive probe on the base lead, you should read relatively low resistance to the emitter and collector leads. If the resistance is several million ohms, it's a PNP transistor. I prefer an analog ohmmeter to a DMM for checking transistors.

The second method of determining transistor type is to examine the circuit. Given that the emitter is grounded, odds are we're dealing with an NPN transistor. Recall that an NPN transistor is forward biased when the emitter is negative and both base and collector are positive. The forward bias condition for a PNP transistor is the inverse—the emitter is positive and both base and collector are negative. In our circuit, the emitter is grounded and the collector is positive because of the resistor to the 3V supply—clearly conditions for an NPN transistor.

LCD

As we discovered during the teardown, no ribbon cable or soldered connector lies between the circuit board and LCD. Instead, a pink Zebra Elastomeric Connector is friction-fit between the fingers on the circuit board and the etched edge connectors on the LCD. Figure 3-5 shows the conductive fingers on the edge of the LCD panel juxtaposed with the edge of the Zebra connector. The bottom third of the figure shows the LCD panel.

Whoever named the Zebra connector must have had a powerful microscope. The stripes, formed by layer upon layer of insulator alternating with carbon, silver, or gold conductor, are barely visible in Figure 3-5. These layers run perpendicular to the long axis of the connector, so that each conducting layer connects one side of the connector to the other. Think mile-high tiramisu, with each cake layer a conductor and each creamy layer an insulator.

To give you an idea of scale, the spacing between conductive fingers of the LCD in the figure is about 1mm. As you can see, there are about ten stripes or conductor-insulator layers per finger, which makes each stripe on the Zebra connector about 0.1mm thick.

A major limitation of the elastomeric connection is that it provides a relatively high resistance path—on the order of 1KΩ—for low-current applications. This is

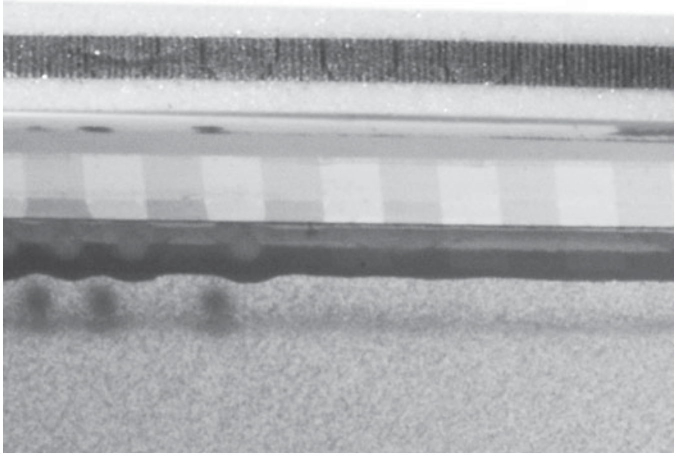

FIGURE 3-5 **Zebra Elastomeric Connector (top) and LCD edge connectors (middle)**

perfect for a low-current LCD, but it wouldn't work for, say, a high-current video board or electric motor connection. For more information on the Zebra connector, try the Fujipoly America web site, at www.fujipoly.com.

How It Works

A simplified schematic of the scale is shown in Figure 3-6. As previously discussed, two 3V lithium cells in parallel supply power. The microcontroller and EEPROM provide programmatic control for the automatic function and LCD display, and a Zebra Elastomeric Connector provides connectivity between the microcontroller and LCD. Note the pinout for the SOT-23 package, which applies to the transistor (Q1).

The trigger circuitry consists of the piezoelectric transducer (PT), NPN transistor (Q1), and pull-up resistor (RP). When the transistor is in the quiescent, nonconducting state, the positive supply voltage is supplied, via the RP, to the microcontroller input line. When the piezoelectric transducer is jolted, it produces an AC signal, the positive component of which forward-biases the transistor. When the transistor conducts, the input line to the microcontroller is pulled down to a few tenths of a volt. This change of state signals the microcontroller to emerge from standby mode and activate the LCD and bridge circuits.

Recall that an NPN transistor is forward-biased when the base and collector are positive relative to the emitter. Also note that, for clarity, the signal from

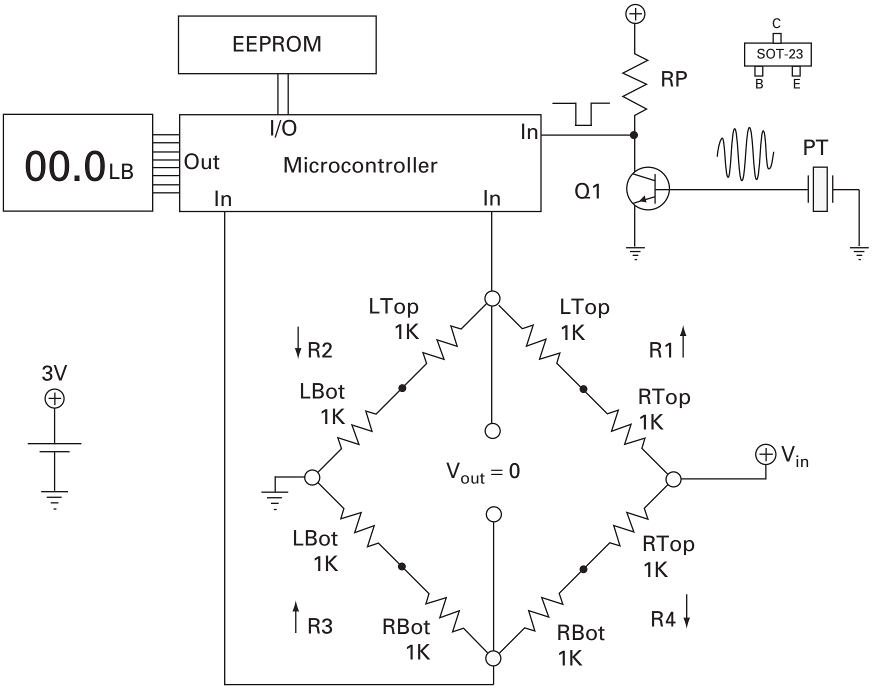

FIGURE 3-6 Simplified schematic of scale

the piezoelectric transducer and output of the trigger circuit are not to scale in the schematic. In addition, only one negative-going pulse is shown on the microcontroller line. In reality, there is a string of pulses, as a function of the severity of the jolt to the piezoelectric transducer. The microcontroller is presumably programmed to ignore activation pulses that closely follow the first pulse.

The most notable feature of the schematic is the resistance bridge formed by the four pairs of strain gauge segments in the load cells. This configuration of resistances, referred to as a *Wheatstone bridge*, produces an output voltage (V_{out}) that is sensitive to minute changes in the relative resistances of each leg of the bridge. This output voltage is sent to the microcontroller, which determines and displays the weight of the person standing on the scale.

The Wheatstone bridge consists of eight separate resistances, the contributions of the two sections of each strain gauge. At rest—that is, without load—the right-top (RTop), right-bottom (RBot), left-top (LTop), and left-bottom (LBot) strain gauges each contribute two 1KΩ resistances to the bridge. The names of each strain gauge correspond to the positions of the load cells in the scale. For example, the RTop strain gauge is located in the right-top foot of the scale, looking at the scale from above, as shown in Figure 3-1.

I labeled the sides of the bridge R1–R4, as shown in Figure 3-6, so that the RTop strain gauge contributed to R1 and R4. The side of the strain gauge that decreases in resistance with weight applied to the load cell contributes to R4, signified by the downward arrow adjacent to R4. The side of the strain gauge that increases in resistance with weight contributes to R1. The remaining three legs of the bridge are similarly configured so that each leg either increases or decreases with weight on the contributing strain gauges.

Now focus on the four sides of the bridge, R1–R4, each 2KΩ in value, with two parallel paths formed by R1–R2 and R3–R4. Hopefully, it's intuitive that the output (V_{out}) from the bridge is 0V, regardless of the input voltage (V_{in}). The same relative voltage drop appears across the voltage dividers R1–R2 and R3–R4.

To illustrate, if V_{in} = 10V, then the voltage drop across R1 = 5V and the voltage drop across R4 = 5V. Using ground as a reference, the junction between R1 and R2—the top corner of the bridge—is at 5V. Similarly, the junction between R3 and R4—the bottom corner of the bridge—is at 5V, relative to ground. Because the top and bottom corners of the bridge are at 5V relative to ground, there is no potential difference between the two corners of the bridge. That is, $V_{out} = 0$.

We can prove this with Ohm's law:

$$V_{out} = V_{in} \times [(R3/(R3 + R4)) - (R2/(R1 + R2))]$$

Assuming all resistances equal 1KΩ, the output voltage is

$$V_{out} = V_{in} \times [(1/2) - (1/2)]$$

$$V_{out} = V_{in} \times 0$$

Regardless of the input voltage (V_{in}), when the resistances in each of the four legs of the bridge are equal, output is 0V. As shown in Figure 3-6, the appropriate LCD display with a 0V input to the microcontroller is 00.0 pound. Of course, R1–R4 aren't exactly 1KΩ, and there will be some output from the bridge with no load. However, the microcontroller uses the initial output of the bridge at startup as 00.0 pound.

Now consider the Wheatstone bridge with a weight of about 173 pounds on the scale—my weight—as shown in Figure 3-7. As you can see, the bridge is unbalanced. Using the previous formula for output

$$V_{out} = V_{in} \times [(R3/(R3 + R4)) - (R2/(R1 + R2))]$$

$$V_{out} = V_{in} \times [(2006/(2006 + 1994)) - (1994/(2006 + 1994))]$$

$$V_{out} = V_{in} \times [(2006/(4000)) - (1994/(4000))]$$

$$V_{out} = V_{in} \times 0.003$$

Assuming an input voltage (V_{in}) of 10V, the output voltage (V_{out}) is 30mV, or about 10mV with an input voltage of 3V. While 10mV may not seem like a significant signal

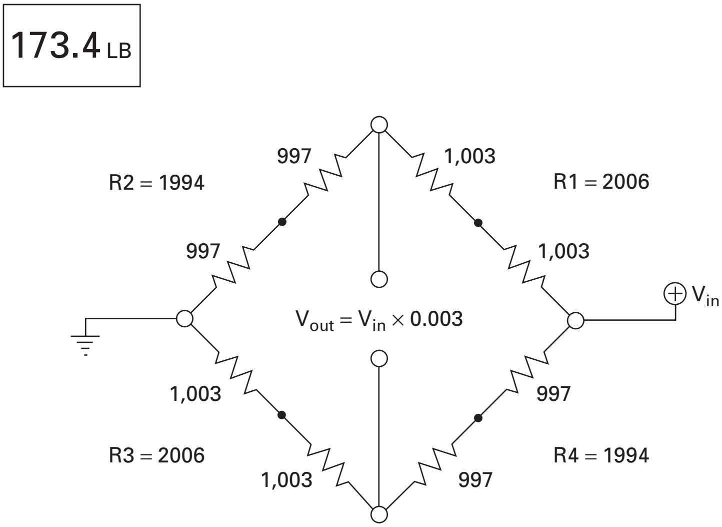

FIGURE 3-7 Wheatstone bridge with 173.4 pounds on the scale

level, it's sufficient for the microcontroller. Strain gauges with greater sensitivity are available for scales with a smaller weight range, such as a kitchen scale with a maximum capacity of a pound or two. In general, strain gauges with higher sensitivity have limited range. Of course, it's possible to purchase strain gauges with high capacity and sensitivity—for a price. If you check the Web for prices of high-capacity, high-sensitivity laboratory scales, you'll find that the dust covers for these scales cost more than our bathroom scale.

The sensitivity of a strain gauge is expressed in terms of Gauge Factor (GF). Mathematically, GF is expressed like so:

$$GF = (\Delta R/R)/(\Delta L/L)$$

Simply put, GF is the ratio of the fractional change in electrical resistance to the fractional change in length of a strain gauge. The fractional change in length, $\Delta L/L$, is the formal definition of strain. The greater the GF, the greater the sensitivity of a strain gauge. As a point of reference, the strain gauges in our scale have a GF of about 2.

If we're interested in measuring the weight of butterfly wings, given enough amplification, we could use the strain gauges in the current load cells for the job. However, given the stiffness of the metal supports, the amplification required would be so great that the signal-to-noise ratio would likely be too low to be useful without significant signal processing. A better approach might be to mount the strain gauges on a more flexible base, or to use a more sensitive strain gauge on a more flexible base.

The best solution depends on the requirements of the scale, including the operating environment. Scale-based strain gauges with a GF of 1000 might be perfect to measure the weight of butterfly wings—assuming the scale is in a lab free of vibration from nearby road traffic, and the scale isn't buffeted by air from a nearby air-conditioning vent.

Although I've focused on sensitivity in this treatment of strain gauges, I'd be amiss not to point out that strain gauges are fundamentally variable resistors. Moreover, the specifications typically associated with fixed and variable resistors apply to strain gauges, including tolerance, repeatability, resistance range, accuracy, linearity, thermal coefficient, power dissipation, operating temperature and humidity, and maximum applied voltage.

Of these variables, the most important from the perspective of an inexpensive bathroom scale is accuracy. The accuracy of a load cell is usually expressed as a percentage of full-scale load, with typical values ranging from 0.03 to about 1 percent of full scale. Assuming our 330-pound capacity scale uses 1 percent load cells, the value displayed on the LCD should be accurate to within ±3.3 pounds. I'm assuming an accuracy of 1 percent, given the price of the scale.

As noted in the discussion of specifications, the manufacturer does not provide an accuracy figure. If the accuracy is in fact 1 percent of full scale, then displaying weight to a tenth of a pound is at best misleading. A display with greater precision doesn't improve the accuracy of the displayed value.

Mods

The easiest mods for this scale are to enhance the triggering circuit and extend measurements beyond the confines of the unit. My favorite source for strain gauges and load cells is Omega Engineering (www.Omega.com). The National Instruments Developer Zone (http://zone.ni.com) is the best source for information on strain gauge design and use.

Enhanced Triggering

Try substituting an outboard microphone and preamp for the piezoelectric transducer to turn on the scale with a hand clap or whistle. Adjust the sensitivity or frequency response of the enhanced trigger so that ordinary room noise doesn't activate the scale, resulting in prematurely depleted batteries.

Create External Load Cells

Although it's a challenge, you can create new load cells by attaching the foil strain gauges to other surfaces. You'll need a steady hand and a sharp blade to peel the strain gauges from their metal mounts. Use superglue (cyanoacrylate) to attach the strain gauges to new platforms that provide both expansion and compression. Try mounts made of thin, relatively flexible metal or plastic to increase the sensitivity of the scale.

You can also use new load cells with different accuracies and sensitivities, as long as the total resistance of each strain gauge is 2KΩ. If you use load cells with a greater equivalent GF than that of the current load cells, the weight displayed on the scale's LCD will be greater than the actual weight. The caveat, of course, is that you'll have to create a lookup table to convert the displayed weight to actual weight.

This page intentionally left blank

Chapter 4

Surge Protective Devices

Surge suppressors and other surge protective devices (SPDs), virtually unheard of before the advent of the PC, are one of the most popular electronics accessories sold. I own at least a half dozen surge suppressors, such as the one shown in Figure 4-1, a pair of power conditioners, and an uninterruptible power supply (UPS).

Why this mix of power technologies? Why spend hundreds of dollars on a power conditioner with fancy meters and lights and-or a UPS capable of an hour of battery backup power instead of a simple surge suppresser? And why not simply use an inexpensive power strip? The reason is that wall socket power can be dirty and fickle: A rare lightning strike can fry your laptop, citywide surges stress devices with linear power supplies, and spike generators in the form of home appliances can be lethal to audiovisual (AV) gear, game consoles, computers, receivers, and other sensitive AC-powered electronics.

FIGURE 4-1 Monster Power AV600 surge suppressor

The most common SPD is the surge suppressor power strip, which provides multiple outlets with active and/or passive electronics that respond to the spikes and, to a lesser degree, surges from the power line. These inexpensive devices, typified by the Monster AV600, offer modest protection against nanosecond-duration events (spikes) and voltage increases that last a second or more (surges).

The typical consumer-grade surge suppressor employs MOVs (metal oxide varistors) to divert and attenuate voltage spikes. MOVs are inexpensive, relatively fast acting, and capable of conducting several thousand amps for a few microseconds, but they also degrade when they absorb energy from a voltage spike. Some commercial-grade surge suppressors combine MOVs with gas tubes, which have a higher energy capacity than MOVs, and avalanche diodes, which are faster than MOVs, to provide a mix of transient response speed, energy handling capacity, and longevity.

Another common SPD, especially in music and professional audio circles, is the power conditioner, which is often marketed as a significant improvement over a basic surge suppressor. Some power conditioners use transformers to provide isolation between equipment and the line and between components, heavy-duty inductors to filter noise from the hot and neutral wires, and a mix of nonlinear components to protect equipment from both surges and spikes. However, many power conditioners are little more than ordinary MOV-based surge suppressors in a fancy enclosure.

In this chapter, we'll tear down two SPDs. The first, an AV600 surge suppressor by Monster Cable, is available on the Web for about $14. The second device is a Furman PL-Plus C power conditioner, available for about $220. The teardowns of these Chinese-built units provide a practical introduction to MOVs, thermal fuses, thermal circuit breakers, component failure mechanisms, and electrical standards.

Surge Suppressors

Surge suppressors are commodity items. Inexpensive surge suppressors from Belkin, Tripp Lite, APC, or other manufacturers typically offer the same suppression features but vary in relatively insignificant features, such as the number and layout of sockets, color, housing size and shape, and length of power cord. I selected the Monster Cable AV600 surge suppressor for this first teardown because it's inexpensive and readily available, and the specifications are representative of similarly priced devices on the market.

Highlights

The AV600 is a typical MOV-based surge suppressor strip constructed of components designed to respond to temperature and heat—properties that aren't easily conveyed by conventional electronic schematics. We'll explore how surges and spikes are handled by separate circuitry and the significance of the "Protection On" LED (light-emitting diode) indicator. During this first teardown, note the following:

- The fiberglass wrap enveloping some of the components
- The slots in the circuit board
- The physical proximity of components
- The size, mass, and number of MOVs

Specifications

According to Monster Cable (www.monstercable.com), the six-outlet unit provides the following:

- Protection against both power surges and voltage spikes
- Suppression from hot to ground, hot to neutral, and ground to neutral
- 555 joule (J) energy rating
- 1 nanosecond (ns) response time
- 15A maximum load capacity at 120VAC, 60Hz
- 330V transient suppression voltage
- Compliance with UL 1449 standards

Let's consider the key specifications in more detail.

Joule Rating

The spike-suppression function of the AV600 works by diverting the spike energy from hot and neutral to ground, and the MOVs that perform this function have a finite energy-handling capacity, which is reflected in the joule rating. The *joule* is a unit of energy, with 1 joule equivalent to 1 watt-second. The greater the surge suppressor's joule rating, the greater its capacity to divert unwanted energy away from sensitive equipment.

Ideally, a surge suppressor can divert an infinite amount of energy from hot and neutral to ground—think lightning strike. However, despite marketing claims, surge suppressors are useless if your house wiring suffers a direct lightning strike. The same goes for a $900 power conditioner or UPSs. Your best bet is to unplug—not simply turn off—your equipment and surge suppressors during a lightning storm.

The energy rating of 555J is the sum of the energy-diverting capabilities of the MOVs used in the AV600. It's difficult to determine the specific energy-diverting capabilities of the AV600 without knowing the number and rating of MOVs installed from hot to ground, hot to neutral, and neutral to ground. And we'll know this after the teardown.

Response Time

Voltage spikes are by definition brief. For a surge suppressor to be useful, it must be fast enough to divert an energy spike before it reaches sensitive equipment. A 1ns response time is excellent for surge suppressors based on MOVs. Even so, 1ns is slow enough that some of the energy of a spike will pass through the suppressor before the MOVs begin to conduct. For example, let's say a high energy spike has a duration of 2ns and that 80 percent of the energy is dispersed in the first nanosecond. Even if the MOV conducts at 100 percent capacity at 1ns, 80 percent of the energy spike gets through the suppressor. In reality, the MOV only begins to conduct at 1ns, meaning that perhaps 90 percent of the spike makes it through the surge suppressor without being attenuated or diverted.

An alternative to the MOV is the gas discharge tube, which lags behind MOVs in response time but has a much greater energy-handling capacity. Avalanche diodes, packaged as transient voltage suppression (TVS) diodes, respond much faster than MOVs but can't handle as much energy.

Transient Suppression Voltage

The TSV is the voltage at which the energy diversion of a MOV kicks in. Given the tolerances of MOVs and other components, the normal variability of the line voltage, and the degradation of MOVs with use, the TSV should be significantly greater than the nominal line voltage. If the TSV is too high, however, the MOVs will be ineffective in protecting equipment. A TSV of 330VAC is a good compromise for an inexpensive surge suppressor.

Suppression LG/LN/GN

Most MOV-based surge suppressors provide energy diversion between hot and ground (LG), hot and neutral (LN), and ground and neutral (GN). The cheapest MOV-based surge suppressor strips often have a single MOV between hot and ground, and therefore can't divert spikes from neutral to ground, clamp voltage spikes between hot and neutral, or divert spikes common to both hot and neutral (common-mode spikes) to ground.

Note

Spike diversion is but one of several methods of protecting equipment from voltage spikes. I'm partial to "heavy iron"—large toroidal transformers that provide complete isolation from the mains, large inductors in the hot and neutral leads that attenuate noise and transients, and nonlinear devices that clamp any residual spikes. But all this comes at a significant price—which is why inexpensive surge suppressors based on 75-cent MOVs are so popular.

UL 1449

The UL 1449 rating is important because it certifies that the AV600 is safe to use. It doesn't certify that the AV600 will protect your equipment from a direct lightning strike or even a minor surge, however. According to the Underwriters Laboratories (www.ul.com), the 1449 rating does not evaluate the effect of the surge suppressor on connected loads, establish a minimum level of attenuation, or assess the adequacy of the voltage protection rating. In short, UL 1449 certification is assurance that the AV600 won't catch fire under normal use.

The UL 1449 standard undergoes periodic updates, and each new edition is more detailed than the last in the assessment of SPDs, including MOVs. UL 1449, third edition, became effective in September 2009. Several UL ratings are relevant to SPDs, depending on the design. For example, UL1414 deals with capacitors connected directly across or in series with the mains. International standards organizations, such as the IEC (International Electrotechnical Commission, www.iec.ch), also establish standards for SPDs.

Operation

I use my AV600s as central power control stations. After all, why flip on a half-dozen on–off switches when one on the surge suppressor will do? Simply hit the illuminated rocker switch and, if both the switch and "Protection On" LED are illuminated, you're good to go.

Before the teardown, I assumed the LED was a ground indicator. After all, a surge suppressor is supposed to divert energy to ground—the green wire, attached to the round conductor on the standard 3-pin NEMA (National Electrical Manufacturers Association) plug. However, in preparation for this teardown, I installed a 2-to-3-prong adapter on the AV600 cord and the "Protection On" LED remained illuminated. I'll discuss the LED circuit in detail later, but you should keep it in mind during the teardown.

Teardown

If you're following along at home with an AV600, you should be able to complete the basic teardown, illustrated in Figure 4-2, in about 5 minutes. If you want to disassemble the main components and remove the MOVs from the circuit board, then give yourself another 20 minutes. Note that this is a destructive teardown. We'll have to destroy the case to get at the components.

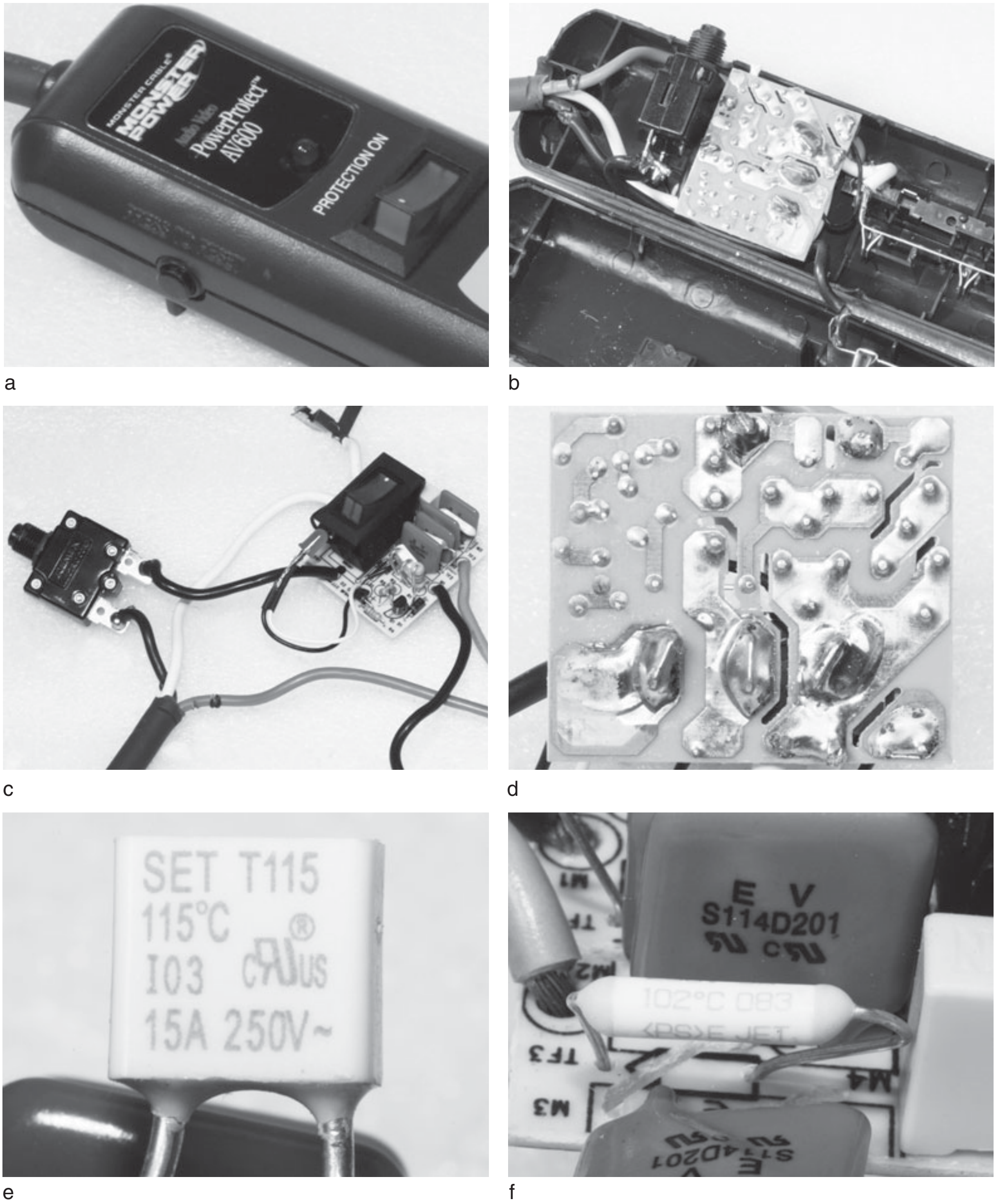

MONSTER CABLE
MONSTER POWER
Audio Video
PowerProtect
AV600
PROTECTION ON
SET T115
115°C
I03
15A 250V~
E V
S114D201
102°C 083
<PS>E JET
M1
M2
M3
M4
TF3

a b c d e f

FIGURE 4-2 Teardown sequence

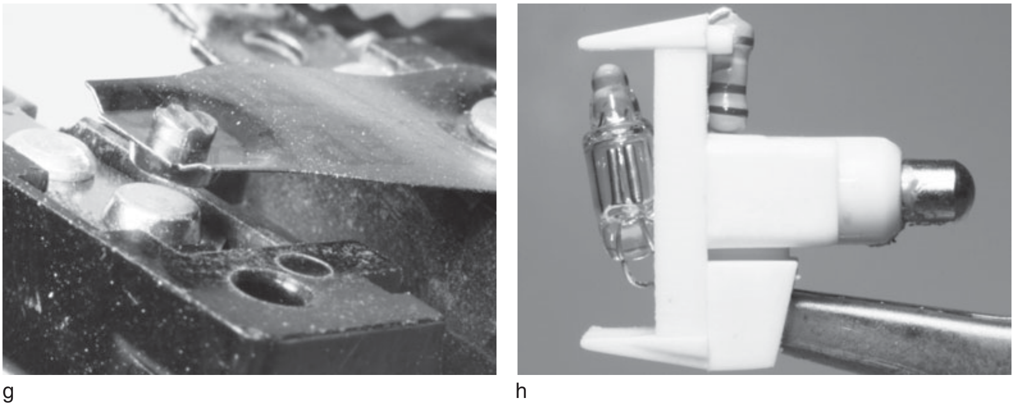

FIGURE 4-2 (*continued*) Teardown sequence

Tools and Instruments

For the basic teardown, you'll need a standard screwdriver with a large blade, a pair of gloves, and eye protection. A large pair of diagonal cutters is useful for extracting the circuit board from the plastic housing. You'll need a soldering iron, pliers, solder wick, or equivalent desoldering tools to extract the components from the circuit board. A drill with a 1/4-inch bit and a pair of heavy-duty lineman's pliers are required for the optional component teardowns.

You'll need a multimeter to test the state of the fuses and other components in the AV600. If you want to experiment with the MOVs, you'll need a high-voltage source, some means of generating voltage spikes, and an oscilloscope. I use a battery-powered photoflash unit to generate spikes. A Variac variable transformer is a great tool for creating surges.

Step by Step

As you progress with this short teardown, you may find it useful to tape down the power cord and other wires as your free components, in preparation for testing.

Caution Before you begin the teardown, verify that the unit is unplugged.

Step 1

Examine the exterior of the unit. There is a 6-foot, three-conductor, 14AWG (American Wire Gauge) power cord attached to one end of the surge suppressor. Note the circuit breaker reset button on the side of the unit, about 1 inch from the cord end of the unit, as shown in Figure 4-2a. Also note the "Protection On" LED, just above the rocker switch.

Step 2

Open the plastic clamshell. With gloved hands, direct the flat blade of a large screwdriver around the unit, breaking the lightly glued seam. Next, insert the blade into the body of the suppressor, around the midline. Pry the halves apart to reveal the circuit board and resettable fuse, as shown in Figure 4-2b. Note the damaged insulation (as shown in the figure) on the ground and hot wires where they enter the unit. Apparently the power cord was not seated properly when the two halves of the clamshell were assembled—an apparent lapse in quality control and a potential fire hazard.

Step 3

Extract the circuit board, LED, and circuit breaker. I removed the board and switch assembly by grasping the plastic body with gloved hands and snapping the body in half. If you're following along at home and want a more controlled approach, use diagonal cutters to nibble away at the plastic body until you free the circuit board. When you're done, you should see the power cord, circuit breaker, circuit board, LED, and the leads to the power jacks, as shown in Figure 4-2c.

To make the photo in Figure 4-2c, I removed a white fiberglass cloth covering the four dark rectangular components (MOVs) to the right of the switch, and a second cloth covering the vertically mounted 1W resistor and the white rectangular component (thermal fuse) near the midline of the board. The separate fiberglass cloths insulate and connect the components thermally and provide protection from fire should a component ignite.

Note how the ground wire from the power cord is soldered to the ground wire of the ground bus, and that the neutral and hot wires are welded to their buses. Note also that the hot (black) wire is the only one that's interrupted by the circuit board. The ground and neutral wires make contact with the board, but they continue on to the receptacle bus. The figure also shows the damaged ground wire insulation.

Step 4

Examine the foil side of the circuit board, as shown in Figure 4-2d. Note the relatively wide circuit tracks, the wide spaces between traces, and the air gaps throughout the middle and right side of the board. Any idea why the manufacturer went through the expense of creating the air gaps? Have you ever seen a circuit board charred by a lightning strike?

Step 5

Identify the components on the board. Unfold the white rectangular thermal fuse adjacent to the 1W resistor, shown in Figure 4-2e. As indicated by the package markings, this is a 115°C thermal fuse with a rating of 15A at 250VAC. Next, examine the double-decker sandwich of four dark rectangular MOVs and two cylindrical thermal fuses, shown in Figure 4-2f. Note the "102°C" marking on the thermal fuse, as well as the markings on the MOVs.

Step 6

Disassemble the thermal circuit breaker (optional). Drill out the five rivets holding the two halves of the component together and use a pair of pliers to remove the sleeve around the push button. You don't have to drill through the casing of the circuit breaker—simply cut through one end of a rivet to release it. Examine the contact mechanism, shown in Figure 4-2g.

Step 7

Disassemble the rocker switch (optional). The rocker switch contains a neon bulb worth studying. If you haven't worked with a neon bulb before, now's the time to harvest one for your parts collection. Simply cut through the soft plastic housing with diagonal cutters to free the bulb and series resistor assembly, as shown in Figure 4-2h.

Layout

As revealed in the teardown, the majority of components are located on a compact circuit board about 1 1/2 × 1 3/4 inches (38mm × 45mm), shown in Figure 4-3. The board contains the illuminated rocker switch, four rectangular MOVs, three

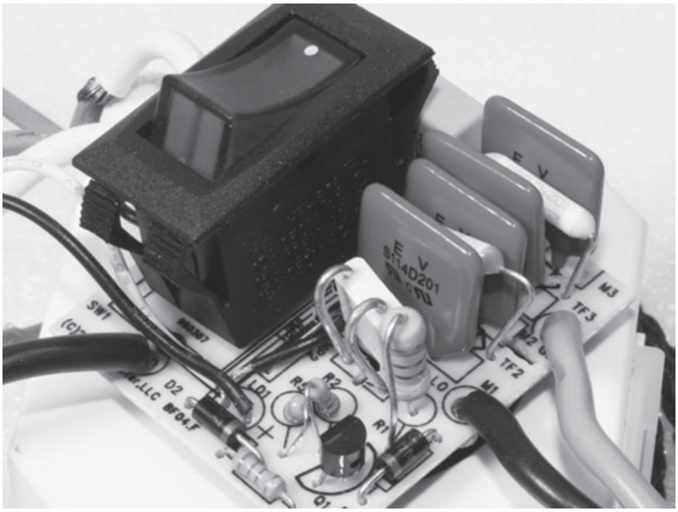

FIGURE 4-3 Circuit board layout

thermal fuses, a transistor, a pair of general-purpose silicon diodes, a capacitor, and a few resistors. The thermal circuit breaker and two of the three power buses—hot and neutral—make connections with the circuit board.

Components

Initially, you might be surprised at the number of components inside such a simple device. After circuit analysis, however, you might be equally impressed at the number of functions these few components can accomplish.

Thermal Circuit Breaker

The largest single component and the only major component not soldered directly to the circuit board is the Joemex (www.joemex.com) thermal circuit breaker, visible in Figure 4-2c. The push-to-reset circuit breaker, rated at 15A and 120VAC, is wired in series with the hot wire, between the power cord and the circuit board.

The circuit breaker consists of a bimetallic band fixed on one end with a contact on the free end, as shown in Figure 4-2g. When the temperature of the strip reaches a predetermined level, whether due to current through the band for a given time or the ambient temperature, the metal band bends, thereby opening the contact. After the band cools for a minute, the plunger on the button can be used to force the contacts together, setting up the circuit breaker for the next over-current event.

The bimetallic arm is a temperature- and time-dependent device. At room temperature, the circuit breaker can carry 15A indefinitely. With a 20A load at room temperature, the circuit breaker will open in about 1 hour. A 30A load will trip the circuit breaker in about 40 seconds at room temperature.

As the ambient temperature rises, the current carrying capacity of the breaker diminishes. For example, at 86°F, the current carrying capacity of the circuit breaker is decreased by 10 percent. At 104°F, a temperature easily achieved in a closed computer closet or AV cabinet, capacity is decreased by about 30 percent.

MOVs

The four EV S114D201 MOVs in the AV600 are rectangular, epoxy-dipped components resembling ceramic disc capacitors. At their normal working voltage of 120VAC, the MOVs are essentially capacitors with a low leakage resistance. I measured the average capacitance of the four MOVs as 800pF and a DC resistance in excess of 10MΩ.

When the voltage across a MOV approaches the transient suppression voltage of 330VAC, the MOV behaves more like a low-resistance shunt, both absorbing and diverting energy. The amount of current that can be carried by a MOV is limited. The higher the current, the higher the heat generated in the body of the MOV, and the greater the likelihood of permanent damage. Recall that the heat produced by a resistance is proportional to the square of current ($P = I^2R$). Unlike resistors, MOVs aren't designed for continuous dissipation of significant energy, but only brief surges.

The 330VAC TSV is based on the specifications of the AV600 unit. Although I was unable to locate the MOVs in a catalog, the component markings suggest a 14mm square body (S114), normal energy rating (D), and a varistor voltage of 200VDC (20×10^1). The varistor voltage, the DC voltage at which 1mA current flows, is one means of identifying the voltage-handling capabilities of a MOV. The varistor voltage is typically, but not necessarily, less than the maximum clamping voltage.

Typical catalog listings for MOVs include the maximum continuous AC and DC voltage, the single pulse transient energy capacity, the varistor voltage, the transient suppression voltage, and capacitance. The transient suppression voltage rating of a MOV is proportional to the thickness of the device, current capacity is proportional to surface area, and energy absorption capacity is proportional to volume (that is, thickness × surface area). The MOVs in the AV600 are 0.5 inch square × 0.06 inch thick (14.6mm sq × 2.1mm).

I set up the experiment illustrated in Figure 4-4 to verify my reading of the component varistor voltage. I used a linear high-voltage power supply from a vacuum tube amplifier to apply a constant DC voltage across a MOV while monitoring the current. Below 200VDC, there was essentially no current flow. However, at 213VDC, the current was 1mA, indicating a varistor voltage of 213VDC. Given typical component tolerance and instrument accuracy, this reading is consistent with a varistor voltage of 200VDC, as per the component marking. Based on varistors listed by DigiKey, a varistor voltage of 200VDC is compatible with the 330VAC transient suppression voltage rating.

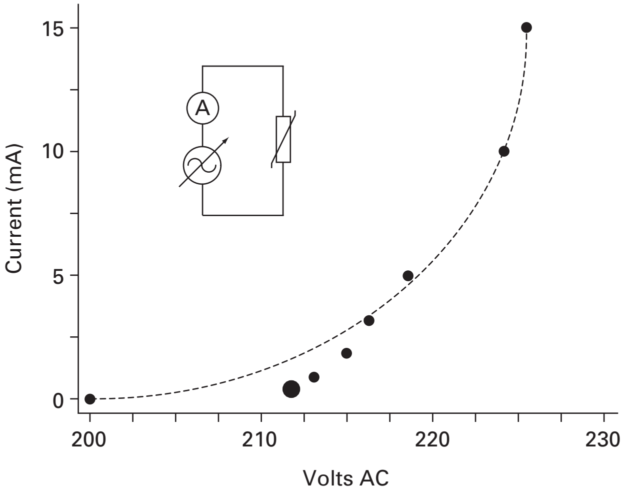

FIGURE 4-4 MOV voltage-current curve experiment results

A limitation of MOVs is that they deteriorate with use. In response to each spike they encounter, a sudden increase in current raises the temperature of the MOV, decreasing its resistance, and, assuming the spike hasn't abated, increasing current. Unless the cycle is checked, it will continue until the MOV fails, typically as a short. In the case of a low-resistance short, the MOV may explode or flame prior to tripping the circuit breaker. This is one reason for the fiberglass wrap; if a MOV fails catastrophically, the heat and shrapnel will be contained within the fiberglass.

If a MOV fails to resistance of, say, 100Ω, the temperature of the component will increase, elevating the temperature of nearby components. Eventually, the MOV will melt, crack, and, ideally, become an open circuit without flaming. Obviously, there's a place for an automatic means of detecting an abnormal rise in MOV temperature and for disconnecting it from the circuit before a damaged MOV becomes a fire hazard.

Because the condition of the MOVs is not apparent through the AV600's opaque case, an indicator of MOV status would be useful. Although it's not practical in an inexpensive surge suppressor, one means of assessing the health of a MOV is to monitor its varistor voltage. It's generally accepted that damage has occurred when the MOV's varistor voltage changes by 10 percent or more from its initial value.

Thermal Fuses

A thermal fuse, also known as a thermal cutoff, permanently opens when its temperature rises above a temperature set point. As with regular current-based fuses, thermal fuses also carry a current rating. However, this figure represents the maximum operating current that a thermal fuse can withstand without derating its temperature set point.

Consider the fuse sandwiched between the two MOVs in Figure 4-2f, for example, which is rated at 5A and 102°C. When the temperature exceeds 216°F (102°C), the fuse permanently opens. If the temperature is only 180°F, the fuse won't blow at a current of 5A or even 10A. The set temperature may decrease from 216°F to 197°F at 10A, but otherwise, within the current carrying capacity of the fuse, it won't open. In my tests, the 102°C/5A thermal fuses in the AV600 can withstand 20A for 1 minute at room temperature before failing. I'm not suggesting you ignore the current rating of a thermal fuse, but just don't confuse the rating with that of a standard fuse. Using a 102°C/5A thermal fuse in a 10A circuit is akin to using a 5A/250VAC slow-blow fuse at 400VAC.

In addition to the thermal fuses sandwiched between MOVs, is a larger, separate thermal fuse, shown in Figure 4-2e, with ratings of 115°C (239°F) and 15A. As with the smaller fuses, the 15A fuse can carry significantly more current than 15A, at the cost of shifting the temperature set point. As you'll see later in the circuit analysis, this thermal fuse provides protection from long-duration voltage surges.

Rocker Switch Neon Lamp

The neon lamp in the rocker switch is worth examining because it represents a class of lamps that is used as voltage regulators and voltage spike diverters. As shown in Figure 4-2h, the neon lamp is constructed of two cylindrical electrodes enclosed in a glass envelope that is filled with neon and a trace of argon under low pressure. When AC or DC voltage is applied to the electrodes, the gas ionizes and gives off an orange-yellow glow. With AC, the area around both electrodes is ionized. With DC voltage, only the area around one electrode is ionized. Do you know which one?

As switch indicators, neon bulbs are useful in line-powered circuits because they're bright, use little power, and require only a series current-limiting resistor to operate at 120VAC. The bulb in the AV600, which is about a third smaller than the ubiquitous NE-2 neon bulb, is connected to a 150KΩ series resistor. This results in an operating current of 120VAC/150KΩ, or 0.8mA.

Unlike an incandescent bulb, the neon bulb exhibits a hysteresis (delayed) effect. The breakdown voltage is about 70VAC/DC, but once the gas in the bulb is ionized, it takes about 10V less voltage to maintain ionization. Another feature of the neon bulb is that the voltage drop across an ionized bulb is constant. Neon bulbs with a breakdown voltage of 200VAC or more are used as spike arrestors in power supplies, power conditioners, and tube-type radio receivers. Neon bulbs are relatively slow to respond to spikes, but their instantaneous shunt current capacity is significant.

Transistor

The single transistor in the AV600 is a KSP44, a high-voltage silicon transistor in a TO-92 package, visible in Figure 4-3. Following the standard package format, looking at the flat face of the transistor, the leads, from left to right, are emitter, base, and collector. The transistor is rated at 500V collector-base and 400V emitter-base voltages, with a collector current of 300mA. The datasheet is available through the Fairchild semiconductor web site (www.fairchildsemi.com).

Diodes

The two diodes in the AV600 are generic silicon power diodes, type 1N4007. They're rated at 1000V PIV (peak inverse voltage) and 1A forward current.

Capacitor

The single, yet prominent, capacitor in the circuit is an MKP X2 metalized polypropylene film capacitor (www.dain.com.tw) rated at 0.1μf and 250VAC. The non-inductive capacitor, constructed with a plastic case sealed with epoxy resin, is sold as an interference-suppression capacitor, for applications that include antenna coupling, bypassing, and electromagnetic interference (EMI) filtering.

Resistors

Of the four metal film resistors in the AV600, the most notable is the 1W, 880Ω, fusible resistor that's thermally isolated with the 15A thermal fuse, as shown in Figure 4-3. The resistor has a flame retardant coating that enables it to fail gracefully, without creating a fire hazard. In normal operation, a band of fiberglass cloth snugly binds the large thermal fuse to the resistor.

LED

The single red LED serves as the "Protection On" indicator light. However, as you'll see in the teardown, what the LED signifies is probably not what you expect.

Circuit Board

The single-sided circuit board is remarkable in the seven cutouts between conductive islands, visible in Figure 4-2d. The cutouts are air-gap insulation between components and leads of the same component that could be compromised by a high-voltage spike. One reason for the gaps is to avoid charring of the circuit board.

Cord

The cord and connector are integral to the AV600. In order for the MOVs to work properly, the unit must be grounded—hence the standard three-pronged NEMA plug. The 6-foot power cord consisting of stranded 14AWG copper wire is sufficient for supporting a 15A load. As a point of reference, 14AWG copper wire has a resistance of about 0.0025Ω/foot at room temperature.

How It Works

The typical single-phase residential power drop-off consists of a hot (black wire), neutral (white wire), and ground (green wire). The hot and neutral lines connect to the power distribution system and the ground is derived locally by connecting a copper wire to a local earth ground, such as a cold water pipe. The flow of current from the hot or neutral to the ground lead—a ground fault—is an abnormal condition. However, the ground lead is a handy place to divert the unwanted energy of a voltage spike, sparing your expensive electronic gadgets.

With this brief background, let's look at the schematic for the AV600, shown in Figure 4-5. Input is on the left and output is on the right, represented by the two NEMA jacks. Note the TO-92 package outline for the PNP transistor (Q1).

The hot lead is interrupted, from left to right in the schematic, by the resettable 15A thermal circuit breaker, by the power switch assembly, and by the 115°C thermal fuse. The symbol for a thermal fuse reflects the typical packaging—a hollow metal tube filled with material from the open end. The overflow of material on the open end tapers to the lead, forming the point of the package. The thermal fuses in the AV600 don't happen to use standard packaging. The other schematic symbol that may be new to you is a rectangle with a diagonal line through its body, which represents a MOV.

The neon bulb (NE) and 150KΩ series resistor, between hot and neutral, illuminate the translucent rocker switch (SW) when the rocker switch is in the on position, assuming the 15A circuit breaker is functional. The 0.1μf capacitor across the rocker switch contacts suppresses the arcing that might occur if a significant load is attached to the surge protector when the switch is moved on or off. Without the capacitor, the arcing across the contacts could damage the equipment plugged in to the AV600 as well as the rocker switch.

Now let's step back and look at the AV600 from a systems' perspective. There are five functional subsystems: a current-limiting system, a spike-diversion system, a spike-clamping system, a power-indicator system, and a surge-protection system. The current-limiting system consists of the 15A thermal circuit breaker. If the load exceeds 15A, the breaker will open, with a response time dependent on the load. A dead short will trip the breaker in a second or two.

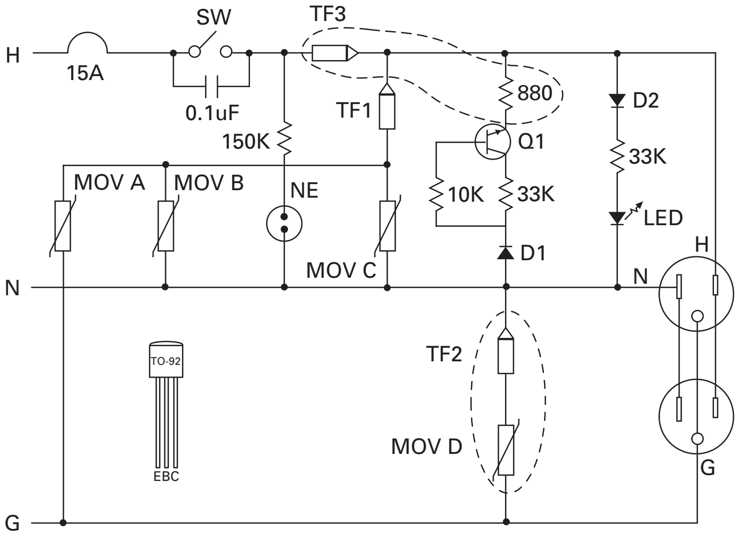

FIGURE 4-5 Schematic diagram of the AV600

The spike-diversion system consists of MOV A between hot and ground and MOV D between hot and neutral. A voltage spike on the hot or neutral lines will be diverted to the ground wire. Note that if the ground wire isn't grounded, the MOVs will be ineffective.

The spike-clamping system consists of MOVs B and C that will clamp a voltage spike between hot and neutral. In compliance with UL 1449, the MOVs are not placed directly across the lines, but are wired in series with thermal fuses. Moreover, the two 102°C thermal fuses (TF1 and TF2) are physically and thermally associated with the MOVs they protect. The dotted ellipse enveloping the TF2 and MOV D signifies one of the thermal links. If MOV A absorbs a spike and fails with a short, its temperature will quickly reach 216°F (102°C) and TF1 will open, removing both MOV A and MOV B from the circuit. This alleviates a potential meltdown of the AV600 but also diminishes the protection provided by the unit.

The power-indicator system consists of the neon bulb within the power switch and the "Protection On" LED between hot and neutral. As long as TF3 is functional, flipping the power switch to the on position should illuminate both neon bulb (NE) and LED. If TF3 is open, the rocker switch will be illuminated, but not the LED. In other words, the state of the LED reflects the status of TF3, independent of the status of the MOVs or other thermal fuses.

The surge-suppression system consists of TG3, a 115°C thermal fuse that is thermally and physically associated with the 880Ω, 1W resistor. The dotted loop enclosing both components at the top of the schematic highlights the thermal connection. If you recall from the teardown, the 115°C thermal fuse and 880Ω resistor are wrapped in an insulating fiberglass cloth.

The 880Ω resistor functions as a thermal indicator of the voltage between hot and neutral. As the voltage increases above 120VAC, the power dissipated by the resistor increases and its temperature rises. The temperature of TF3 tracks the temperature of the resistor. If the sufficiently elevated line voltage persists, the temperature of the resistor and TF3 will reach 239°F (115°C), and TF3 will open. So why the diode (D1) and transistor (Q1)?

Let's examine Figure 4-6 to understand the design of the surge-suppression circuit. The figure shows a plot of voltage versus current and the surface temperature of the 880Ω resistor. The first thing to note is the small line segment labeled "1mA/Volt" in the lower-right quadrant of the graph. It represents the slope of voltage versus current for pure resistance plotted against the same voltage and current scales used in the larger graph.

In comparison, the slope of the voltage versus current for the resistor-transistor circuit combination starts out at about 0.5mA/Volt and then increases to more than 2mA/Volt. As a result, the surface temperature of the 880Ω resistor, and the body temperature of the 115°C thermal fuse, increase rapidly with an applied voltage greater than 120VAC. In other words, with increasing voltage, the relative increase in current increases—and more rapidly than it would with a resistor alone. The purpose of the transistor circuit is to introduce nonlinearity in the flow of current through the 880Ω resistor, thereby increasing the sensitivity of the surge-suppression system to prolonged overvoltages.

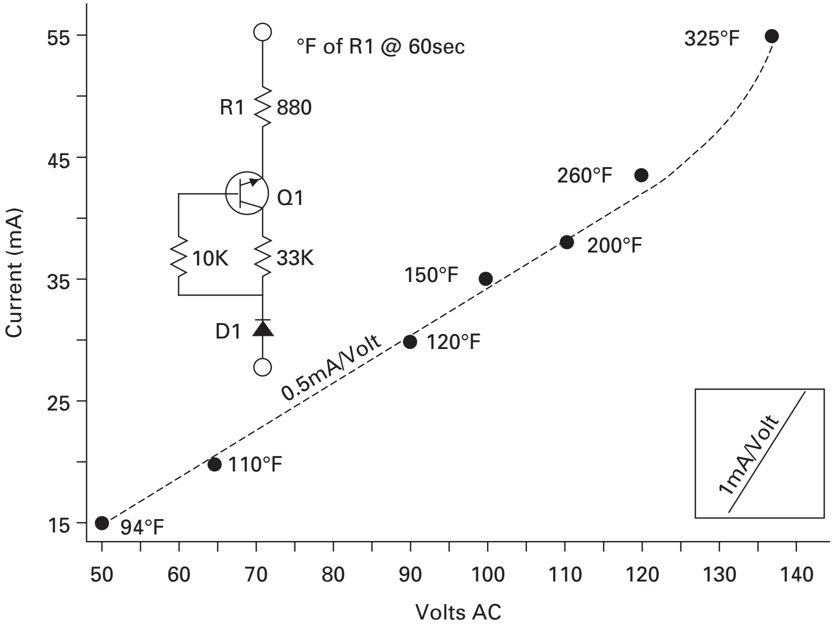

FIGURE 4-6 Surge-suppression circuit and voltage-current plot

The D1 ensures that the transistor conducts only when the hot is negative relative to the neutral wire. The two resistors provide forward bias for the transistor. The values of the 10 and 33KΩ resistors were apparently chosen to position the Q1 in a particular region of its operating curve so that a significant nonlinear change in the voltage-current relationship occurs when the emitter-collector current is about 45mA. See the datasheet for details on the operating characteristics of the KSP44s transistor.

Mods

I don't suggest modding the AV600 for regular service, because of the potential fire hazard. However, as a bench experiment, a worthy goal is implementing a more useful "Protection On" indicator that reflects ground and MOV status. Perhaps you can sell the circuit design to Monster Cable.

Power Conditioner

The subject of this supplemental teardown is the Furman PL-Plus C Power Conditioner, hereafter referred to as simply PL-Plus and shown with its top cover

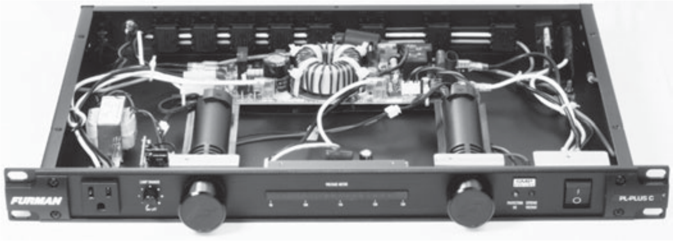

FIGURE 4-7 Furman power conditioner

removed in Figure 4-7. I selected the rackmount Pl-Plus over similarly priced power conditioners from Monster Cable, Tripp Lite, and others for this teardown simply because I own several previous models of Furman conditioners and they've been trouble-free from installation day. Part of the attraction of the PL-Plus is its user interface. Not only does it offer nine outlets, including one in front, but there's an LED voltmeter; pullout, dimmable LED lights; and a BNC socket in the rear for a gooseneck lamp.

Highlights

Given its cost, I assumed the PL-Plus was a super surge suppressor with the added benefits of lights and built-in voltmeter. However, as you'll see in the teardown, I didn't find quite what I expected. Given our previous discussion of surge suppressors, note the following:

- MOV count, capacity, and placement
- Treatment of the ground wire
- Provision for MOV failure

Specifications

Based on the marketing literature, packaging, and markings on my PL-Plus, specifications for the nine-outlet unit include the following:

- Isolated banks of sockets
- 15A maximum load capacity at 120VAC, 60Hz
- 1ns maximum response time
- Protection against both power surges and voltage spikes
- 12W power consumption
- 133VAC RMS spike-clamping voltage and 188V peak at 3000A

- Maximum surge current of 6,500A
- Automatic shutdown with prolonged voltage over 140VAC and under 90VAC
- Spike suppression from hot to neutral
- Noise suppression 10dB at 10kHz, 40dB at 100kHz, 50dB at 500kHz
- Compliance with TÜVRheinland standards

Some of the specifications are unclear. For example, does the 12W power consumption refer to the underlying protection circuitry or simply the drain of the lighting system and LED voltmeter? How does the 3000A figure differ from the 6,500A maximum surge current? How are the banks of isolated sockets segregated, and what is the isolation in dB?

Also of note is what's not mentioned. There is no mention of Joule rating, for example. There is no mention of neutral-to-ground or hot-to-ground transient suppression, only that there is no ground contamination. There's also no mention of UL listing anywhere in the literature or on the body of the unit. Furthermore, a visit to the TÜVRheinland web site (www.tuvdotcom.com) failed to provide me with much insight into the TÜV standard.

In the realm of curious statements, the user manual states that, unlike traditional surge-suppression circuits that sacrifice themselves when exposed to multiple transient spikes, requiring repair, this isn't the case with the PL-Plus. The unit somehow protects itself. Furman also states that MOVs, when used within their design specifications, do not degrade, and that the circuitry within the PL-Plus contains both a 12,000A MOV and a 12,000A TVZ (thermo-fuse varistor). Furthermore, an overvoltage protection circuit protects the MOV from damage caused by overvoltage. Other claims include the shunting of transients only 1.5 to 2.0V above the peak line voltage to an internal capacitor, and a filter circuit that attenuates a typical lightning voltage surge.

Given all of these claims, most of which require additional information for clarification, I was curious about what I'd find inside the PL-Plus. Let's just say I was surprised.

Operation

There isn't much to say on the issue of operation. You plug in your equipment, either in the front or rear sockets, and flip the front on–off rocker switch. If all is well, the LED voltmeter should show the current line voltage and the green "Protection On" LED in the front panel should be illuminated.

If you mistakenly plug the unit into a 240VAC line—hard to do with a NEMA 15 plug—the red "Extreme Voltage" LED will illuminate, the LED voltmeter will dim, and the hot lead to the sockets will be interrupted. The same goes for a brownout—if the line voltage drops below about 90VAC, the hot lead to the nine sockets will be interrupted. It's important to note that the neutral wire is not affected by the over- and under-voltage cutout circuitry.

Mini-Teardown

This is a mini-teardown, perhaps better described as a simple extraction, as illustrated in Figure 4-8. I'm going to focus on the components and circuitry that deal directly with voltage and surge suppression. If you're following along at home, give yourself 5 minutes to pop the top of the unit and another 5 minutes to extract the circuit board.

Tools and Instruments

You'll need a Phillips-head screwdriver with a sharp, small to medium head to get at the cover screws. Use a manual screwdriver—I found the screws easy to strip

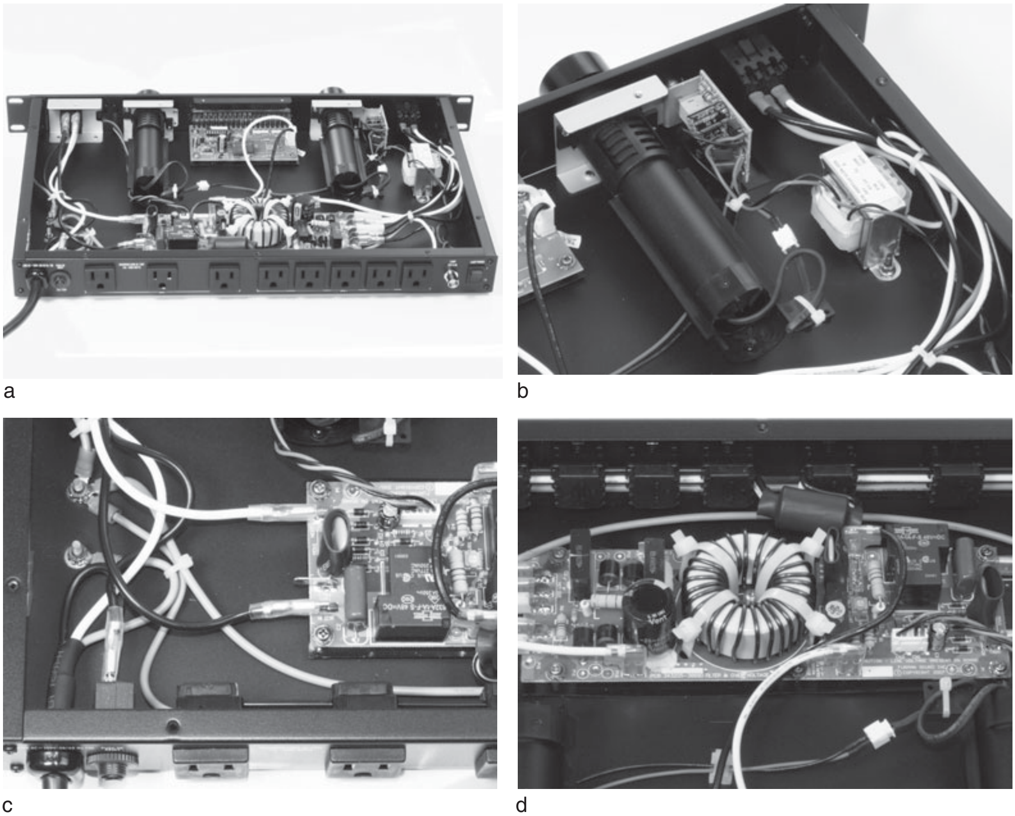

FIGURE 4-8 Mini-teardown sequence

with a power screwdriver. A pair of needle-nose pliers will be useful in breaking the quick-disconnect hardware connections. You'll need a Variac and multimeter to test the voltage cutoff circuitry.

Step by Step

Before you begin the teardown, verify that the unit is unplugged. It's also a good idea to remove any devices plugged into the unit.

Caution Be sure the unit is unplugged before you begin the teardown.

Step 1

Remove the top plate and ground wire. Remove the screws securing the top plate. Next, carefully tilt the plate until you can see the ground cable connection. With a pair of pliers, unplug the ground wire and put the top plate to the side.

Step 2

Identify the main subsystems. Turn the unit so that you're facing the rear sockets, as shown in Figure 4-8a. On the outside of the unit, from left to right, locate the grommet securing the three-conductor, 14AWG power cord to the chassis, the 15A circuit breaker reset button, eight 3-prong NEMA 15 sockets, the 12V external lamp connector, and the oddly placed on–off switch for the external lamp.

Now, focusing on the inside of the chassis, identify the power circuit board, the largest and most prominent subsystem, located adjacent to the output sockets. On the far right is the 12V lamp circuit, including the 12VAC power transformer bolted to the bottom of the chassis and the small dimmer circuit board, shown in Figure 4-8b. Between the cylindrical slide-out LED lamps is the LED voltmeter circuit board, visible in Figure 4-8a. Note the ground wire connections to the chassis, shown in Figure 4-8c, and the lack of connectivity with the power circuit board.

Now, turn the unit so that you're facing the front panel. Note the connections to the power circuit board as well as the wiring of the rear sockets, shown in Figure 4-8d. Notice how the hot and neutral lines attach to the power circuit board on the right and exit on the left.

Hot, neutral, and ground wiring within the unit are 14AWG stranded copper, with the exception of the wire running through the Roto-Tech NEMA sockets in the rear of the unit, which is 14AWG solid copper. Also, note the shrink-wrapped RFI (radio frequency interference) inductor around the line and neutral wires, between two of the Roto-Tech sockets. Note how the insulation of the inductor presses against several components on the power circuit board.

Step 3

Extract the power circuit board. Note the location of the quick disconnect tabs as you remove the wires connected to the circuit board. Next, remove the seven Phillips-head screws securing the board to the bottom of the steel chassis. Flip the board over to examine the trace pattern. The neutral trace runs uninterrupted on one side of the board. The hot trace is along the opposite side of the board, interrupted at the location of the relay and of the toroidal inductor.

Layout

Let's focus on the layout of the 8 × 2 1/2 inch power circuit board, shown in Figures 4-9 and 4-10. As noted during the teardown, the board has connections to the hot and neutral wires. The input section of the board, shown in Figure 4-9, contains two MOVs, a thermal fuse, a relay, a rectangular metalized polyester film capacitor, and circuitry to drive the two status LEDs on the front panel. The output or post-inductor section of the board, shown in Figure 4-10, includes a bridge rectifier circuit, an electrolytic capacitor, a small inductor, and a pair of rectangular metalized polyester film capacitors.

Of note are the two MOVs in Figure 4-9. The MOV on the edge of the board stands alone, soldered across the line and neutral lines. However, the MOV nearest

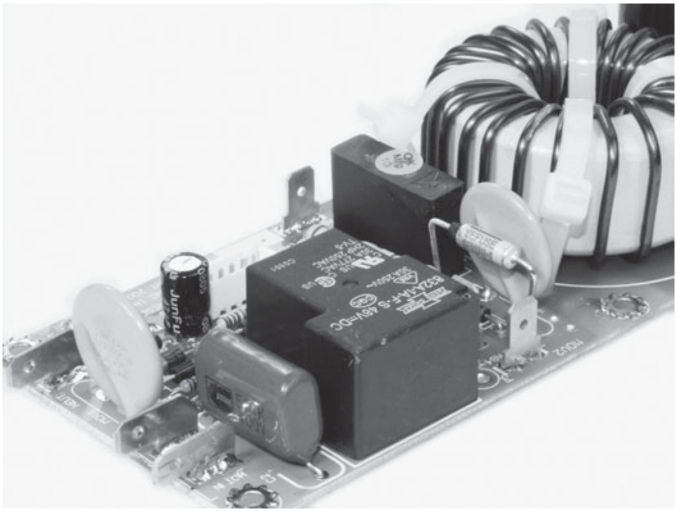

FIGURE 4-9 Input section of the power circuit board

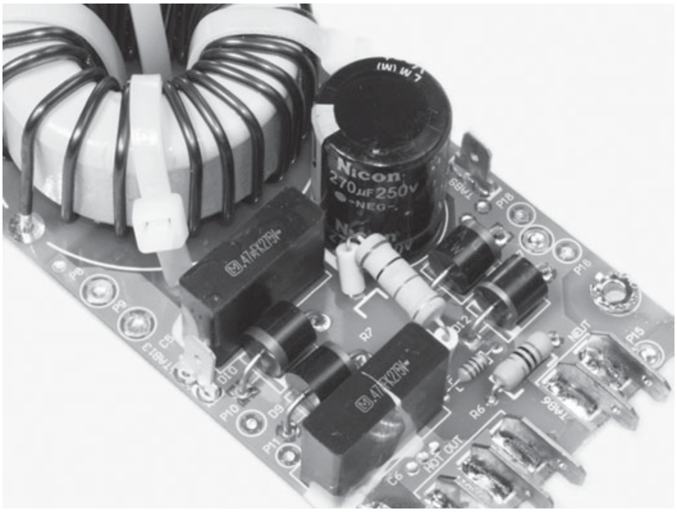

FIGURE 4-10 Output section of power circuit board

the large inductor is wired in series with the thermal fuse. I removed the heat-shrink tubing that physically bonded the thermal fuse and MOV for illustration purposes.

Components

The following discussion of components is limited to those on the power circuit board and mounted on the chassis that are involved in transient and surge suppression.

MOVs

Two Sincera MOVs are on the power circuit board, a 20D391K-N and a 20D201K-N. According to the Sincera documentation (www.worldproducts.com), the 20mm MOVs are standard models, with varistor voltages of 390VDC ($391 = 39 \times 10^1$) and 200VDC ($201 = 20 \times 10^1$). The *K* signifies 10 percent tolerance and the *N* signifies the component is lead-free. The phenolic coating on each MOV is rated at 2000VDC.

Surprisingly, given the specifications of the PL-Plus, high-energy MOVs are not used. In addition, the typical response time for the MOVs is listed as <15ns, as opposed to 1ns response time listed by the PL-Plus manufacturer. The MOV catalog also lists criteria for determining MOV failure—the standard voltage change of

±10 percent from the initially measured varistor voltage, an apparent admission that the MOVs deteriorate with use. More significant is the maximum clamping voltage, which the catalog shows as 650V for the 20D391K-N and 300V for the 20D201K-N. I measured the thickness of the MOVs at 2.5mm for the 200VDC unit and 2.9mm for the 390V unit, which is in agreement with the voltage rating.

Thermal Fuse

The single thermal fuse on the board, which has a trip temperature of 115°C, is physically bonded to the lower voltage MOV with shrink-wrap tubing, as shown in Figure 4-9. This thermal fuse has the traditional shape with one square end and one pointed end.

Thermal Circuit Breaker

The Rong Feng 125VAC/15A thermal circuit breaker is similar to the one discussed earlier. It's one of the few significant components not mounted on the power circuit board.

Inductors

The unit includes three inductors—the large toroidal choke in the center of the board, a small choke on the output section of the board, and a tubular ferrite choke around the hot and neutral leads between several of the rear sockets. The large choke, composed of 26 turns of enameled 12AWG copper wire wrapped on a 2-inch diameter and 1-inch thick core, accounts for the majority of the board's weight of 15 ounces (430g). I measured the inductance of the choke at 0.16 milliHenry (mH). This equates to an inductive reactance of about 1Ω/kHz, using the formula $X_L = 2\pi fL$. At 60Hz, inductive reactance is only 0.001Ω, but at 100kHz, it climbs to a modest 100Ω.

The small 22μH (microHenry) inductor, about the size of the 1/8W film resistor, is on the post-inductor side of the power circuit board. The inductor is part of a series circuit with a resistor and capacitor.

I was unable to determine the inductance of the tubular ferrite choke, which apparently accounts for the manufacturer's claim of isolating the power sockets. Based on ferrite chokes of similar size listed in the Digi-Key (www.digikey.com) catalog, the choke probably introduces only about 50Ω of resistance to RF signals at 1MHz traveling in the hot and neutral lines.

LEDs

Two generic LEDs are located in the front of the unit, one red and one green. The green LED serves as the "Protection On" indicator, and the red the "Extreme Voltage" indicator.

Rocker Switch

The on–off rocker switch is a DPDT (double-pole, double-throw) switch that connects and disrupts both the hot or line and neutral wires. This model has no internal light.

Diodes

Four 6A4 silicon diodes reside on the output side of the power circuit board, shown in Figure 4-10. The 9mm diameter diodes are rated at 400V PIV with a maximum peak forward surge current of 400A. The diodes are part of an unusual bridge circuit on the post-inductor side of the board.

Capacitors

Three 0.47μf at 275VAC metalized polyester film capacitors are packaged in rectangular, flame-retardant plastic cases and noncombustible resin. According to the datasheet, these capacitors are typically used in interference suppressor applications. The large Mylar 0.56μf at 400VDC capacitor, visible in Figure 4-9, is part of the over/undervoltage monitoring circuit.

Relay

The Song Chuan 832A-1A-F-S 48VDC relay is the most prominent component on the input side of the board, as shown in Figure 4-9. The single-pole, normally open relay has silver-tin contacts rated at 250VAC and 30A. According to the Song Chaun relay catalog (www.songchuan.de/en), the 40V coil requires 18mA for activation—that's about 1W to keep the relay energized. Of note are the maximum contact closure time of 15 millisecond (ms) and the 10ms maximum release time. The relay is marketed as a general-purpose power relay for power supplies and home appliances.

Circuit Board

The spacious, sturdy, double-sided circuit board is well laid out, with plenty of room for heat dissipation. My only criticism of the board is related to the placement of the holes for securing the large inductor. While the trace for the hot lead is more than an inch wide in places, it comes to a narrow of about 0.08 inch (2mm) because of a mounting hole drilled through the trace. A nonconductive center bolt that leaves the trace intact is probably a better mounting option.

How It Works

Using the simplified schematic in Figure 4-11 as a guide, let's focus on the circuits related to surge protection. The 12V dimmer circuit for the LED lighting is

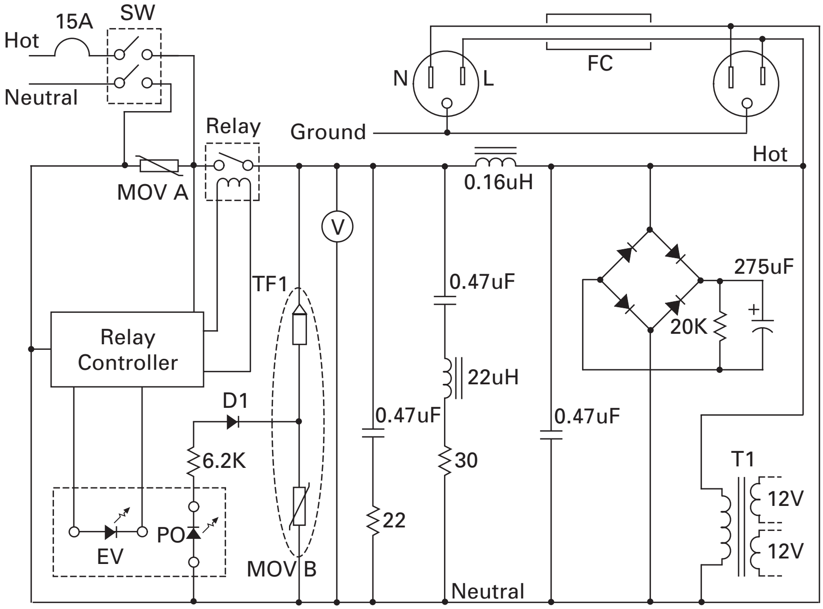

FIGURE 4-11 Simplified schematic

represented by the connection for the primary of transformer T1, between the line and neutral, at the bottom-right corner of the schematic. Similarly, the front-panel LED voltmeter is signified by the encircled *V* between hot and ground.

Power enters from the top-left and power is output through the two NEMA 15 outlets on the top-right. Each jack represents one "bank," as described on the PL-Plus package. The tubular ferrite core (FC) that encircles the line and neutral wires effectively adds inductance to the wire and "isolates" the three final sockets in the chain from the other six sockets. The amount of isolation provided by the ferrite core is not specified.

The hot lead is connected to a 15A circuit breaker prior to the DPDT on–off switch (SW) that interrupts both hot and neutral. After the on–off switch, directly across the hot and neutral lines is MOV A, the 20D391K-N, which has a maximum clamping voltage of 650V. If the MOV shorts, the on–off switch will be subject to short current until the 15A circuit breaker trips. With a maximum clamping voltage of 650V, the MOV is more likely to survive repeated voltage spikes than, say, a MOV rated at 300V, but it will also allow more spike energy to pass unabated.

The relay controller activates the normally open relay when the hot-to-neutral voltage is between about 90 and 140VAC. At voltages in excess of 140VAC, the relay releases, and the red "Extreme Voltage" LED (EV) is illuminated on the front panel. Because of the response time of the relay—a maximum of 10ms to release the contacts—and the relay control circuitry, the relay controller circuit doesn't come into play with subnanosecond voltage spikes.

The brunt of any spike that makes it past the relay is handled by MOV B, which has both a series thermal fuse (TF1) rated at 115°C and a monitoring LED (PO). The silicon power diode (D1) and 6.2KΩ resistor provide current to the green "Protection On" LED (PO) on the front panel. If the MOV overheats and TF1 opens, the LED will not illuminate. In this way, "Protection On" reflects the status of MOV B.

Moving to the right in the schematic is the LED voltmeter (V), which draws only a few milliamperes. The next component across the hot and neutral lines is a 0.47µf capacitor in series with a 22Ω, 1W fusible resistor. Given the capacitive reactance of the capacitor at 60Hz—about 6KΩ—the contribution of the 22Ω resistor is negligible. It's apparently in the circuit as a fuse, in the event the capacitor shorts secondary to damage from a transient. Recall that the MOV A clamps at 650V, which is significantly above the 275VAC rating of the capacitors.

The next circuit, just prior to the large choke in the line wire, is a series circuit composed of a 0.47µf capacitor, a 22µH inductor, and a 30Ω, 1W fusible resistor. Again the resistor apparently serves as a fuse. The inclusion of a small inductor, which, according to my Mouser Electronics catalog, has a maximum current rating of about 100mA, is curious at best. The resonant frequency of the 0.47µf capacitor and 22µH inductor is about 50kHz. That is, 50kHz signals between the hot and neutral wires are shunted, while signals greater or less than 50kHz pass and are not attenuated. This apparent tuning might provide some small improvement in attenuation of 50kHz signals—one of the test frequencies listed in the noise suppression test figures—but at the cost of increasing noise for all other frequencies, relative to a bare capacitor.

As noted earlier, the series inductor, the centerpiece of the power circuit board, provides 0.16µH of inductance, based on my measurements. Depending on the nature of the voltage spike, the inductor can both reduce the amplitude and lengthen the duration of a spike. This may be the rationale for including a 0.47µf capacitor directly across the line and neutral wires after the inductor.

The final circuit element is a diode bridge that feeds a 275µf at 250VDC electrolytic capacitor that leads to a 20KΩ bleeder resistor. This circuit, which I assume represents the manufacturer's linear filtering technology, is in my opinion of little value in protecting sensitive equipment plugged into the PL-Plus from significant voltage spikes.

While the diodes are capable of significant peak forward current from a spike, the current handling capabilities of the capacitor are limited. One of several reasons is that the function of relatively large value electrolytic capacitors degrades with increasing frequency, due to self-inductance. That's why electrolytic capacitors are often bypassed with relatively small value ceramic disc or Mylar capacitors.

As a source of protection from voltage spikes, the PL-Plus is essentially a pair of MOVs across the hot and neutral lines. The large inductor and, to a much lesser extent, linear power supply components probably absorb some of the energy of a transient. The question is how much energy is absorbed. Unlike the AV600, the "Protection On" indicator actually indicates the status of at least one of the MOVs. However, there is no protection from common-mode spikes or spikes from neutral to ground.

Mods

As with the AV600, I'm hesitant to recommend mods to line-powered equipment that's going to be in constant service. However, in my opinion, the PL-Plus is a platform to build on. Following are a few ideas for you to consider.

TMOV Upgrade

The fastest and easiest mod is to upgrade the first, exposed MOV to a fused MOV with a significantly lower clamping voltage rating. I picked up a Littelfuse 20mm TMOV (thermal MOV) with built-in thermal fuse, TMOV20R115E, from Mouser (www.mouser.com) for $1.30. The TMOV is rated at 115VAC continuous, with a clamping voltage of 300V, a 52J energy capacity, and a peak current of 10,000A. Littelfuse also offers a three-lead version, with a lead between the internal MOV and thermal resistor, for externally monitoring the state of the fuse. Both are compliant with UL 1449 2nd Edition.

Add Neutral and Ground MOVs

Install a small circuit board with Littelfuse 20mm TMOVs between ground and hot and ground and neutral. Whether or not "ground contamination" is a real problem for your equipment depends on whether you use equipment with switching power supplies or linear supplies, among other factors.

Component Deletions

Consider replacing the 22μH inductor with a wire jumper. I can see no reason for a filter limited to 50kHz.

Comparison

I don't want to turn this teardown into a product review, but I think it's useful to compare what we've learned about the AV600 and PL-Plus, realizing that they're very different products serving different markets. Based on the teardowns, my overall assessment is that while the PL-Plus is certainly a more robust device, at home with the coolest looking AV equipment on the market, the little plastic AV600 is more sensitive and offers more spike protection per dollar.

The PL-Plus is obviously superior to the AV600 as a power distribution unit and in the areas of over- and undervoltage protection and noise suppression. Overvoltage could be a problem if you have an emergency generator with an improper output setting, or if you generate your own power from solar cells or a wind turbine. In addition, if you have a band that frequently sets up around fluorescent lights, the noise suppression offered by the PL-Plus could be invaluable.

While the PL-Plus clearly has a place next to high-end AV equipment (I now own three Furman power conditioners), the findings of the teardown don't support—and in some cases refute—many of the claims made by Furman. At best, many of the statements in the marketing materials are misleading. For example, the components we discovered during the teardown simply don't support the claim of a spike-clamping voltage of 133VAC RMS. The claim that the circuitry includes both a 12,000A MOV and a 12,000A TVZ was clearly refuted. In addition, the relative effectiveness of the linear filtering circuitry is unsubstantiated.

The specifications of the AV600 are also suspect. For example, I was unable to locate the MOV datasheet to verify the 1ns response time. That the "Protection On" LED functions with the ground wire disconnected and every MOV in the unit blown is problematic.

One issue I have with the PL-Plus as a user is the lack of UL 1449 compliance. Among other requirements, the latest revision requires specific safety measures, such as thermally fused protection of MOVs. It's not clear if the AV600 will pass the latest UL 1449 requirements without further testing or modification, but it's clearly designed to disengage the MOVs quickly if they overheat.

This page intentionally left blank

Chapter 5

Electronic Pedometer

Home entertainment systems have been cited as one reason for the expanding American waistline. That may be, but electronic devices are also helping us trim the fat. A simple but effective electronic aid to fitness that you may own or have access to is the electronic pedometer. In its simplest expression, a pedometer is an instrument that counts the number of steps you take. Advanced units calculate calories burned as they track your pace and time. Try counting your steps as you walk or run and you'll immediately appreciate the utility of a pedometer.

In this chapter, I'll tear down an Omron HJ-112 electronic pedometer, shown in Figure 5-1. This digital meter, which retails for about $40, is widely considered one of the top instruments on the market. In addition to discussing to technologies used in competing products, I'll suggest an experimental mod that can extend the basic functionality of the unit.

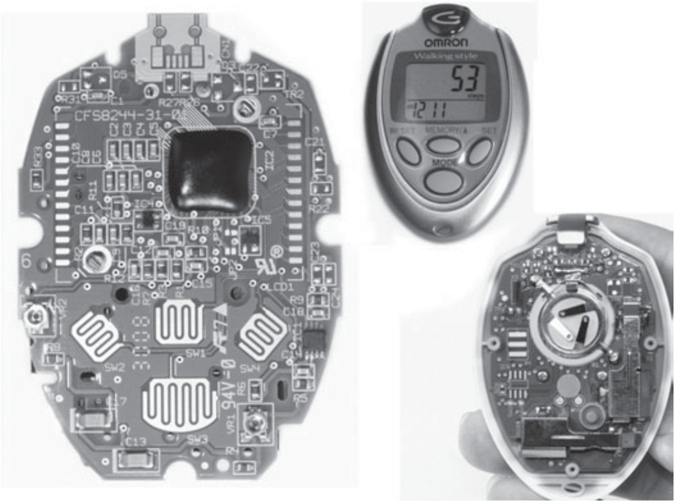

FIGURE 5-1 Omron HJ-112 electronic pedometer

Highlights

I was initially disappointed with this teardown. After reading the marketing copy and rave reviews for the Omron HJ-112, hereafter referred to as the Omron, I expected an all-electronic pedometer, complete with the latest solid-state micro electromechanical systems (MEMS) accelerometers. However, when I first cracked the case and found what appeared to be old-fashioned pendulum switches, I thought of abandoning the teardown. I changed my mind a few minutes later after examining the circuitry and function of the sensors.

This teardown explores an integrated electromechanical system that combines SMT components, a microcontroller that I'll largely ignore, three peripheral chips, and a pair of unique electromechanical sensors. During the teardown, note the following:

- The use of SMT components, including resistors, capacitors, and integrated circuits (ICs)
- The LCD panel, including the mounting hardware
- The tuning fork crystal oscillator
- The construction of the pendulum sensors
- The silicone-carbon switch components

Specifications

The 2.8 × 2.1 × 0.6 inch (HWD) pedometer measures steps with an accuracy of ±5 percent. That is, if you take 1000 steps, the count can be as high as 1050 or as low as 950 and still be within specifications. Based on this step count and elapsed time, the microcontroller calculates and displays calories, distance, and pace. It also displays aerobic steps and minutes, which are the number of steps and elapsed time since 10 minutes of continuous activity. Power is provide by a single CR2032 3V lithium cell, which is rated at six months, assuming 10,000 steps per day. The 1 ounce, Chinese-made Omron has a seven-day memory.

Because the calculation of calories is based on the step count, the calorie counter function is probably no more than ±5 percent accurate. In addition, because age is not used to calculate calories burned, you have to multiply the displayed figure by an age-dependent correction factor from a table in the user manual for an accurate estimate of caloric expenditure. The default equation assumes either a man in his 60s or a woman in her 30s.

Operation

Setup involves entering the time, your weight, and your stride length using the four elastomeric buttons. After that, it's simply a matter of clipping the unit to your belt, backpack, or purse. However, the unit must be perpendicular to the ground during operation. In other words, you can't simply toss the unit into your backpack or briefcase. This is a major limitation of using a dual-axis sensor system.

Teardown

If you're following along with a unit at home, you should be able to complete the basic teardown, illustrated in Figure 5-2, in less than 5 minutes.

Tools and Instruments

You'll need a miniature screwdriver to crack the case and a good magnifier to examine the SMT components. Have a soldering iron handy if you want to examine the two pendulum sensors in detail.

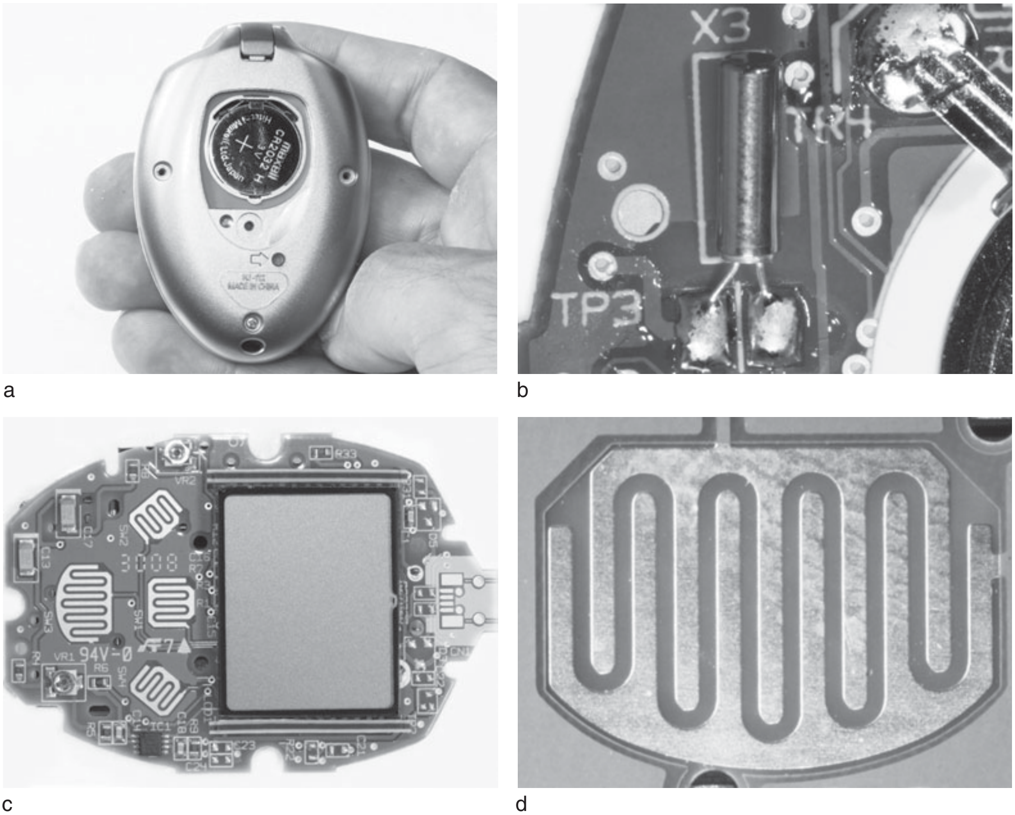
a
b
c
d
X3
TR4
TP3
maxell
CR2032 H
3V
MADE IN CHINA
94V-0
VR1
VR2
SW1
SW2
SW3
SW4
IC1

FIGURE 5-2 Teardown sequence

Step by Step

If you plan to use the pedometer after the teardown, you should probably stop after viewing the back of the circuit board. Once you remove the circuit board completely from the plastic shell, your odds of reseating the LCD panel are slim.

Step 1

Remove the 3V lithium cell. Use a miniature Phillips-head screwdriver to remove the single screw securing the small battery hatch, as shown in Figure 5-2a.

Step 2

Remove the back. Remove the three Phillips-head screws securing the back cover, shown in Figure 5-2a. With one hand, pull up and out on the plastic retention hook at the top of the unit and pull up on the back cover to reveal the back of the circuit board, shown in the lower-right corner of Figure 5-1.

Note the thin rubber gasket that runs in the groove along the periphery of the case. If you plan to use the pedometer after the teardown, do your best to avoid damaging the gasket. Also note the two metal covers of the pendulum switches, immediately to the right and below the lithium cell contacts. Note the silicone reset switch between the two mechanical pendulum switches, and the cylindrical tuning fork crystal oscillator, at the 10 o'clock position relative to the cell contacts. A close-up of the crystal oscillator is shown in Figure 5-2b.

Step 3

Separate the circuit board from the front of the case. Remove the four Phillips-head screws above and below the cell contacts, visible in the lower-right of Figure 5-1. Hold the board in place while you flip the unit over. Gently lift the plastic front plate to reveal the front of the circuit board. Rotate the board so that the LCD panel is on the right, as shown in Figure 5-2c. Note the two SMT potentiometers at the 8 o'clock and 11 o'clock positions relative to the center of the board. Also note the four serpentine pads for silicone-carbon elastomeric switches, which occupy much of the board to the left of the LCD panel. A close-up of a pad is shown in Figure 5-2d.

Now examine how the LCD panel is attached to the board. Referencing Figure 5-2c, note the gray Zebra elastomeric connector that runs along the sides of the LCD panel. As discussed in detail in the teardown of the digital bathroom scale (Chapter 3), the Zebra connector provides hundreds of parallel, electrically insulated connections between the LCD panel and the foil pads on the circuit board directly beneath the panel. One of the many advantages of the Zebra connector over alternative connection technologies is that no solder or special spring mount system is required to connect an LCD to a circuit board. The downside is that reassembling the panel back without a jig to hold everything in place is a challenge.

Step 4

Remove the LCD panel. Secure the circuit board and carefully lift up on the LCD panel. The Zebra connectors may adhere to the circuit board or the LCD panel, or they may simply fall off as you lift the panel. Place the connectors and the panel in a clean, dust-free container. Note the additional components visible, including three ICs and several SMT resistors and capacitors, as shown in Figure 5-1. A microcontroller is hidden under the black epoxy blob.

Layout

As revealed in the teardown, the layout is simple, clean, and compact. Solder joints are well executed on even the smallest SMT components.

The challenge presented by this board is that the components and traces are so small that you need a good magnifier and a multimeter with a set of sharp probes to explore the circuitry. Moreover, many of the traces are hidden under an insulating layer between the battery and circuit board. Fortunately, there are multiple test points on the circuit board to facilitate testing and circuit tracing.

Components

Of the components in this teardown, the most intriguing are the pendulum sensors. The only other consumer devices that come to mind that might include similar switches are car alarms and motion detector alarms for laptops.

Pendulum Sensors

The two pendulum sensors, shown in Figures 5-1, 5-3a, and 5-3b, are aptly named because each sensor consists of a flexible shaft attached to the circuit board at one end and to a broad copper wing at the other end. As shown in Figure 5-3a, the wing-shaped ends of each pendulum are confined to a metal enclosure. If you shake the board in the plane of the pendulum, the copper wing-shaped ends move slightly with no perceptible oscillation. The motion of your torso as you walk should result in the same motion.

Figure 5-3b shows one of the pendulum sensors without its cover. The foam rubber glued to either side of the pendulum arm dampens the mechanical oscillation of the pendulum. Without the foam and the wing-shaped air dampers, each pendulum could move for several seconds and falsely elevate step count.

At first I thought that the copper wings at the end of each pendulum sensor made contact with the grounded enclosure. Then I noticed that the solder connection from the circuit board to the arms of each pendulum isn't a direct connection. As shown in Figure 5-4, a short wire is soldered to what appears to be a length of double-silvered mica. As discussed next, it took a bit of investigation to discover what

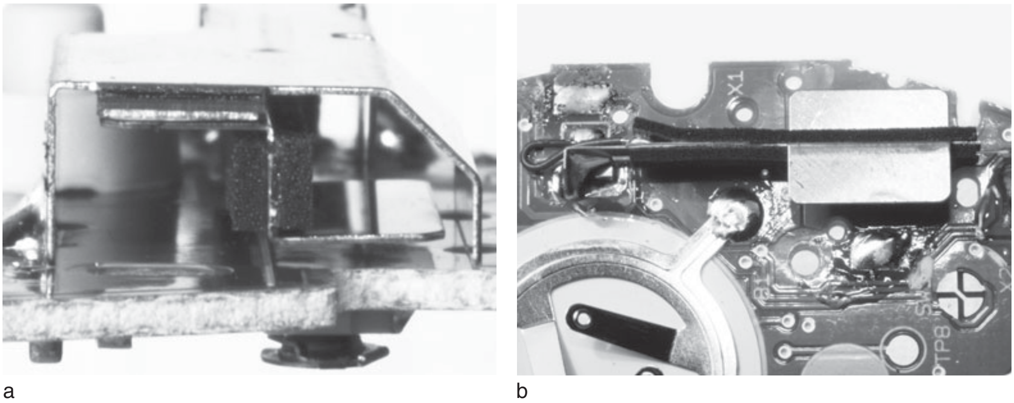

FIGURE 5-3 Pendulum sensor

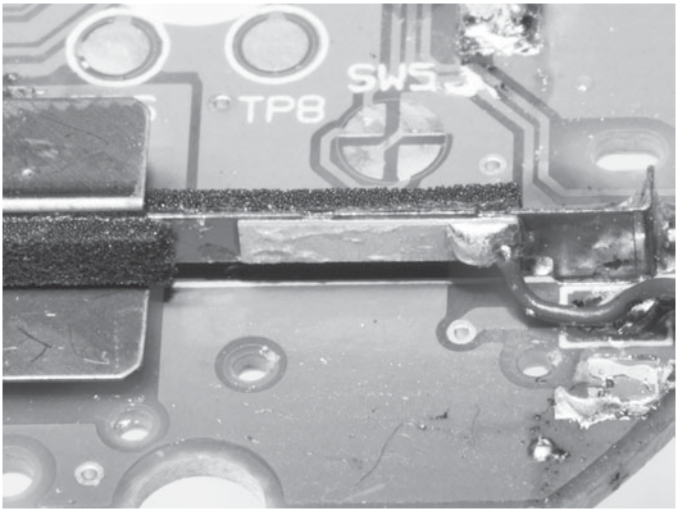

FIGURE 5-4 Pendulum sensor connection

was really going on with the pendulum sensors. If you're working on a pedometer at home, stop reading and try to figure out how the sensors work on your own.

Microcontroller

The microcontroller hidden under the glob of black epoxy, visible in Figure 5-1, handles the LCD display and the momentary switches, and it reads signals from the sensors. Review the discussion of how to remove epoxy blobs in Chapter 3, and you'll see that if you remove the epoxy in this case, you may be able to visualize the raw silicon chip, but that's about it. For our purposes, what's important is that the microcontroller reads the two pendulum sensors, processes the data, and displays the results.

Microcontroller Support ICs

In addition to an external crystal oscillator, the microcontroller relies on three external ICs for normal operation—two CMOS (complementary metal oxide semiconductor) voltage detectors and a dual operational amplifier. The B8K and CUN voltage detectors are packaged as 4-pin SC-82AB devices, as shown in Figure 5-5a. It's difficult to see, but pin 2, in the upper-left corner of the device, is slightly wider than the other three pins. The 8-pin dual operational amplifier is in a standard SOP-8 package, shown in Figure 5-5b.

The voltage detectors, which connect directly to the microcontroller, are triggered when the battery voltage drops to a preset level. The operational amplifiers are used to condition the signals from the pendulum sensors.

Crystal Oscillator

The cylindrical tuning fork crystal oscillator, which operates at 32.768kHz, connects directly to the microcontroller. Tuning fork crystal oscillators are popular in portable devices that need a real-time clock because the oscillator package is small, the operating frequency is stable over a relatively wide temperature range, and the fundamental frequency is low. Clock frequency is a compromise between performance and battery life.

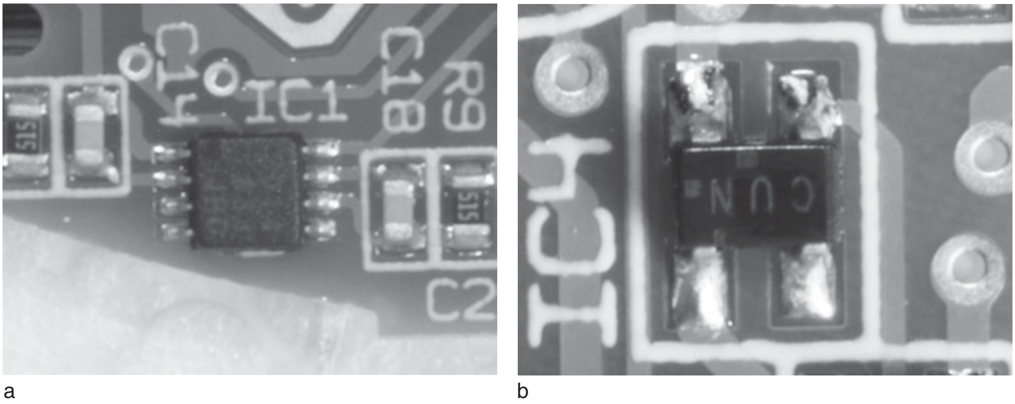

FIGURE 5-5 Microcontroller support ICs

SMT Potentiometers

The two 14KΩ linear potentiometers, shown as VR1 and VR2 in the lower half of the large circuit board shone in Figure 5-1, are used to set the sensitivity of the pendulum sensors. These SMT potentiometers are unlike typical full-sized potentiometers in that the wiper arm is free to rotate continuously. If you rotate the Phillips-head screw clockwise, the resistance from the wiper arm to one of the end terminals cycles from 0 through 14KΩ. The lack of a mechanical stop is common in SMT potentiometers.

LCD Panel

The LCD panel is notable in its use of dual Zebra elastomeric connectors to mate it with the circuit board. The connectors not only facilitate automated assembly of the device, but they protect the panel from vibration and g-forces. Given that the device is likely to be dropped, this protection increases the odds of the panel surviving an impact with the ground.

Silicone-Carbon Elastomeric Buttons

The pedometer relies on four silicone-carbon buttons: a reset button on the back of the circuit board, and buttons for mode, history, set, and reset functions on the front of the pedometer. The bottom of each button is covered with conductive carbon that, when pressed, shorts an underlying circuit board trace, such as the one shown in Figure 5-2d, completing the circuit.

Elastomeric buttons are lightweight and inexpensive, and they provide excellent tactile feedback in a small package. In addition, the design of the silicone button assembly adds to the water-resistance of the pedometer. A related switch design that uses circuit board tracings as part of the circuit is discussed in Chapter 10.

How It Works

The major components and their contributions to the pedometer are shown in the simplified schematic of Figure 5-6. Power is supplied by a single 3V lithium cell that is connected directly to the microcontroller and supporting ICs. Depressing one of the five elastomeric switches, represented by the single switch (S1–5), activates the microcontroller and the LCD. The crystal oscillator (XT) provides the microcontroller with a stable clock signal, which is required for the clock function and to calculate pace.

The horizontal (TH) and vertical (TV) piezoelectric pendulum transducers generate an irregular, alternating signal, as a function of the wearer's motion. IC1 acts as a signal conditioner, producing a clean pulse that's easily processed by the analog-to-digital conversion circuitry within the microcontroller. The 14KΩ SMT

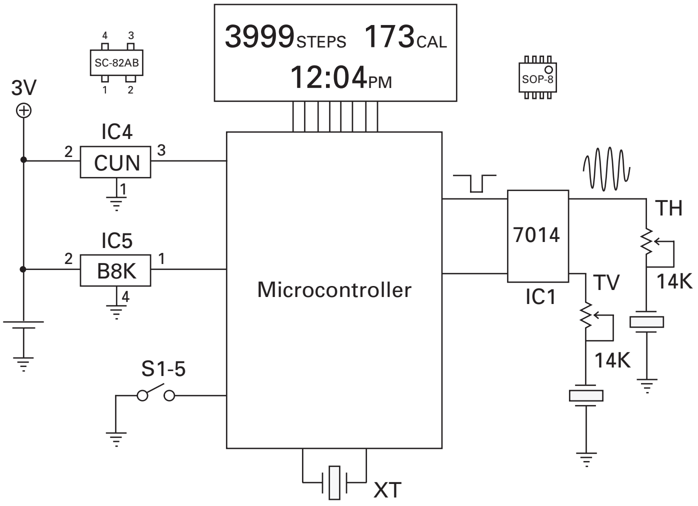

FIGURE 5-6 Simplified schematic of pedometer

potentiometers in series with the piezoelectric transducers are used to control for differences in transducer output. This form of signal conditioning should be familiar to you by now, since I covered the same basic circuit in the teardown of the digital bathroom scale in Chapter 3.

To appreciate what the pulse shaping circuit has to handle, see Figure 5-7, which shows the signal from the horizontal pendulum transducer during one of my steps with the Omron clipped to my belt. As you can see, a single step results in a 500mV signal with duration of about 100ms.

One of the challenges in the design of battery-operated, computerized devices is how to handle a nearly depleted battery at startup and in the middle of an operation. Think of your laptop: it warns you of impending battery death and initiates an automatic shutdown sequence. It also refuses to boot up if the battery is depleted.

As shown in the schematic, the two voltage detectors connect directly to the microcontroller and provide it with information on the state of the battery. IC5, the B8K voltage detector, is triggered at 2.5VDC. IC4, the CUN detector, triggers at about 2.0VDC. Without access to the source code, it's difficult to determine exactly what the microcontroller does with the information provided by the voltage detectors. My assumption is that IC5 triggers a controlled shutdown procedure that protects the data and the state of the microcontroller registers. The state of the CUN voltage detector is probably checked early in the microcontroller startup sequence. If the battery voltage is less than 2.0V, the startup sequence is aborted, thereby avoiding a crash during startup, which could corrupt the registers.

An alternative to the CMOS voltage detectors is to use one or more analog-to-digital converters within the microcontroller to monitor battery voltage periodically. The downside of this solution is the significant overhead of an additional analog channel, in terms of chip complexity, cost, and power requirements. Not only do the

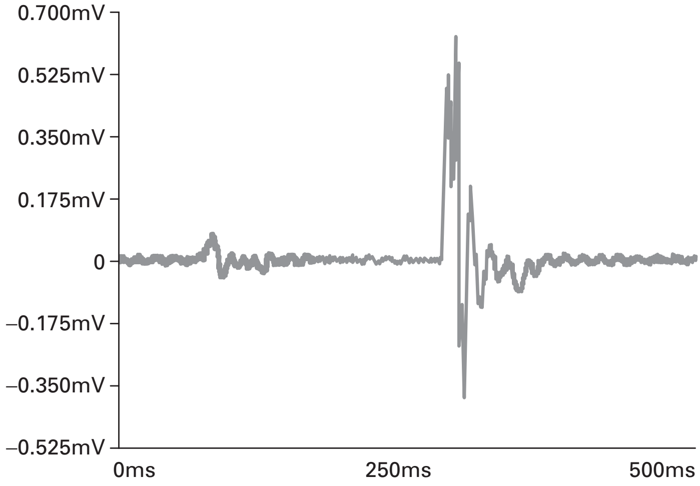

FIGURE 5-7 Raw pendulum sensor output

external voltage detectors provide superior accuracy, but the current requirements are only 0.8µA.

Alternative Technologies

Based on my use, the piezoelectric pendulum transducers are more than sufficient for the pedometer. However, there are alternative sensor technologies that range from cheaper, simpler motion sensors, to more expensive and more advanced transducers that can be integrated with other useful functionality.

Single-axis Pendulum

A simple, inexpensive alternative to the dual-axis piezoelectric pendulum is a single-axis mechanical pendulum, such as the one used in the Korean-made MS-2000 digital counter (see Figure 5-8, left). If you examine the image, you will see the mechanical pendulum switch, including the single wire spring that can be set at one of three notches to adjust for stride length.

Note that in normal use, the pedometer should be worn so that the pendulum arm is horizontal to the ground. When the wearer walks, the weighted pendulum bobs up and down, and in so doing, it makes mechanical contact with the second electrode. Note the black epoxy blob that hides the microprocessor that counts the steps and displays the total count. The unit has a tiny, four-digit LCD display and a single button to zero the count. There is no calculation of calories or an electronic method of specifying weight or stride. The only user-customizable setting is the spring tension.

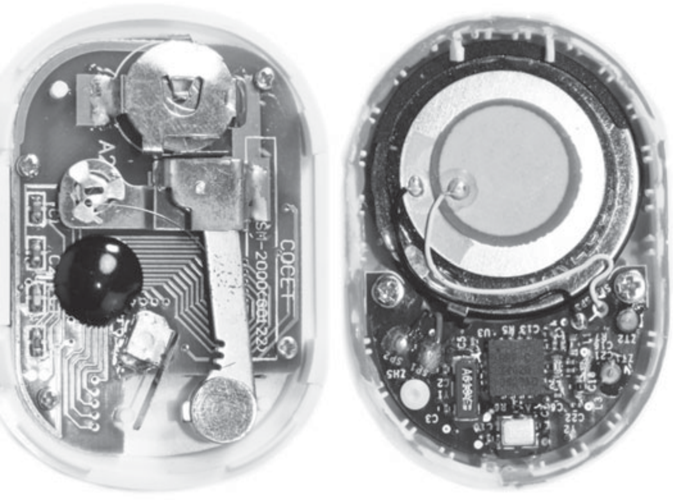

FIGURE 5-8 Mechanical pendulum (left) and piezoelectric step sensor (right)

Piezoelectric Sound Detector

Related to our piezoelectric pendulum sensors is the piezoelectric sound sensor used in the Apple/Nike shoe sensor, shown on the right in Figure 5-8. (Note that the two devices in the image are not to scale. The Apple/Nike sensor is about half the size of the MS-2000 pedometer.) The piezoelectric sensor, which sends signals wirelessly to an iPod with a receiver, triggers on the sound produced by the wearer's foot slapping the ground. Nike sells shoes with built-in holders in the soles. However, I've used the sensor attached to other shoes with good results.

I suspect the sensor would have problems if the wearer runs on loose gravel or soft grass. One of the common complaints of the Nike/Apple sensor is that the case is sealed and that there is no way to change the battery. I used diagonal cutters to nibble away the red and white top to get a view of the sensor and wireless communications circuitry. The sensor should remind you of the bathroom scale teardown in Chapter 3, in which we discovered a piezoelectric sensor used to turn on the scale.

MEMS Accelerometer

If you have an Apple iPhone/iTouch, a Sony/Ericsson w710i phone, or one of several other cell phones, you have at least the potential of owning a pedometer. These devices include two or more multiaxis MEMS accelerometers that can be accessed by software, transforming the device into an expensive but readily available pedometer.

From a functional perspective, the two-axis accelerometer in the Sony/Ericsson phone offers little advantage over our pedometer, other than convenience. The real advantage of MEMS accelerometers is obvious with the Apple iPhone, which sports a three-axis MEMS accelerometer. With three axes covered, the iPhone can respond to steps regardless of its orientation. It can be upside down in a backpack, flat, or at any angle relative to the ground. The limitation of the iPhone, of course, is the initial cost of the phone and the need to buy and install a pedometer app.

MEMS accelerometers vary in mode of operation, but one that I've worked with extensively, the Memsic 2125, uses a heater to warm a bubble of gas that moves in response to gravity and acceleration of the device. Thermopiles near the periphery of the bubble sense identical temperatures when the accelerometer is level and different temperatures when the sensor is tilted. Temperature differences detected by the thermopiles are converted into pulses by the onboard electronics and the angle of tilt is calculated using elementary trigonometry.

If you want to experiment with the Memsic 2125, it's available from www.parallax.com for about the price of our pedometer. And this is the rub. In the current pedometer market, the cost of a three-axis MEMS and peripheral circuitry is prohibitive.

Mods

Given the compact, integrated design, there really aren't many mods that come to mind that could significantly improve upon this pedometer. However, if you're willing to give up portability, the pedometer makes a great platform to explore multiaxis signal detection. If all else fails, the pedometer is a good source for SMT components, and for the piezoelectric pendulum sensors.

Add a Third Axis

A simple bench experiment is to add a third piezoelectric sensor in parallel with one of the existing pendulum sensors. The challenge is presenting the unit with a signal source with amplitude and spectral characteristics compatible with the piezoelectric pendulum sensors. You could try to roll your own third axis sensor with a piezo speaker element, or harvest the sensor from a second Omron pedometer.

Chapter 6

Compact Fluorescent Lamp

One of the casualties of the drive toward carbon neutrality is the tungsten light bulb. LED, halogen, and fluorescent lights are increasingly popular alternatives to the inefficient incandescent bulbs, in part because local governments and utility companies subsidize these green technologies, and in part because they're theoretically cheaper to operate and have a significantly longer lifespan. A rate-limiting step in the move to higher efficiency lighting is the infrastructure of existing screw-in sockets.

The form and function limitations imposed by screw-in sockets have been partially addressed by the compact fluorescent (CF) lamp, such as the one shown partially disassembled in Figure 6-1. To fit and operate in regular sockets and fixtures, the base of CF lamps designed for residential use contains disposable

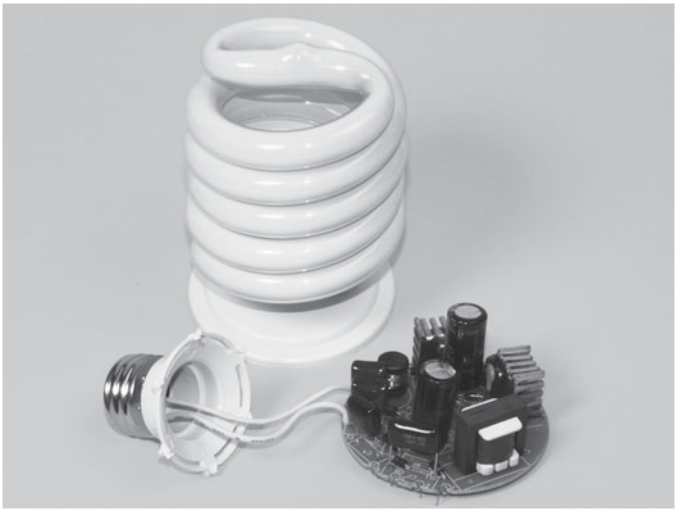

FIGURE 6-1 Compact fluorescent lamp, partially disassembled

ballast circuitry. As such, CF lamps are a bridging technology that will do for now. CF lamps are not as efficient or cost effective as the regular straight fluorescent lamps used in offices and factories. In addition, the production, assembly, and disposal of the electronic components add to the carbon footprint of CF lamps. Disposing of CF lamps also increases the load of mercury, lead, arsenic, and other toxins on the environment.

In this chapter, we'll tear down the 42W TCP SpringLamp shown in Figure 6-1. This CF lamp is available online from vendors such as bulbs.com for less than $20. Although you may not have this particular lamp at home, the layout and component functions differ surprisingly little from one brand to another.

Highlights

If you've never cracked open a CF bulb, you're in for a pleasant surprise. There's quite a bit of technology packed in the base of each bulb. I've opened dozens of different models, and every one had two switching transistors, a toroidal transformer, several inductors, several diodes, a handful of capacitors, and a few resistors. As I noted earlier, it seems paradoxical to create a disposable technology in the name of saving the planet. However, if you're into repurposing hardware, you're the immediate beneficiary of this low-cost source of components.

Caution Now for the disclaimer. It's a good idea to think of the spiral bulb as a snake coiled to strike. If you're careless and unlucky, you could get cut, poisoned, and electrocuted—all within a few seconds. So it's important to keep the bulb intact.

Whether you follow along virtually or with your own CF, during the teardown, note the following:

- **The component types** Do you see any SMT components? If not, why not, given the cramped conditions in the base of the bulb?
- **Provisions for component failure** Are there any fuses or components designed to fail gracefully?
- **The thermal design** Most CF lamps can't operate in closed environments because of heat buildup, and heat kills components. Moreover, many CF lamps can operate only in certain orientations, such as base-up. Are there any heat sinks? Vents? Any active cooling?

Specifications

A typical 100W soft white incandescent bulb is rated at 1500 lumens and 1200 hours. The "soft white" designation means that the color temperature of the lamp is 2700°K. *Color temperature* refers to the color of the light emitted by the

Dangers: Mercury Toxicity and High Voltage

If you apply a little too much pressure to the bulb, it will break. If you're not wearing gloves and the sharp end of the bulb pierces the palm of your hand, it's probably off to the ER for stitches. Leather gloves and eye protection may protect you from glass shards, but they won't protect you from the danger of mercury poisoning. The bulb is filled with a gas mixture composed of mercury vapor, argon, xenon, and neon at a pressure of about 0.003 atmospheres.

To give you an idea of the toxicity of the mercury contained in a CF bulb, consider the recommendation of the US Environmental Protection Agency (EPA). According to the EPA, if you break a CF bulb, you should open a window and leave the room for at least 15 minutes. You should also shut off the central heating or air conditioner to avoid the spread of mercury vapors. When you return, use duct tape to pick up the shards of glass—don't launch the shards into the air with a vacuum cleaner. *Throw away* any clothing that has been in direct contact with the shards. Washing the clothes in machine washer will *contaminate the machine*.

In case you're curious, the effects of mercury toxicity typically aren't immediately apparent after exposure. Symptoms usually appear years after chronic exposure, but a single large exposure can cause immediate problems. According to the National Institutes of Health, you'll likely develop significant neurological deficits—inability to walk, blindness, memory problems, uncontrollable shake, and seizures. Not something you want for yourself or your family.

The other potential health hazard, high voltage, has an immediate effect. If you decide to trace the live circuitry, you're minimally exposing yourself to 120VAC. If you work with a powerful CF lamp, such as the 42W lamp featured in this teardown, you'll be exposed to 600VAC or more. You should have high-voltage equipment and tools on hand if you're working with a live CF bulb.

The bottom line is that if you don't have the protective clothing, tools, and experience, then either enjoy the teardown vicariously or find someone with the skills and tools to work with you.

bulb. The 2700°K is equivalent to the yellowish light produced by a tungsten bulb. 1500 lumens is the amount of light output. Average lifetime is the average service lifetime of a 100W bulb. It's not the average lifetime you can expect from your particular bulb.

In contrast, the specifications of the TCP 42W SpringLamp are as follows:

- **Color temperature:** 5100°K/bright white
- **Average lifetime:** 10,000 hours
- **Lumens:** 2800
- **Base:** Medium screw (E26)
- **Input line voltage:** 120VAC
- **Minimum starting temperature:** –20°F
- **Energy used:** 42W

- **Incandescent equivalent:** 150W
- **Power factor:** >50 percent
- **CRI (color rendering index, aka color temperature):** 84
- **THD (total harmonic distortion):** <150 percent
- **Length:** 7.0 inches
- **Diameter:** 2.8 inches

Although most of these terms should be familiar to you, this list of specifications hints at the complexity of the typical CF relative to the simple tungsten bulb. The SpringLamp features a 5100°K bulb and emits relatively white light. A related specification is the CRI, which relates to how well the light renders colors naturally. A CRI of 100 is perfect. The SpringLamp, like most CF lamps, doesn't produce much red light, so reds are muddied. As a result, the CRI score is only 84—good but not great.

Unlike tungsten lamps, CF lamps employ an initial starting arc, generally assisted with heating elements in the ends of the bulb. The indoor SpringLamp won't start below –20°F—meaning you probably wouldn't want to install it in a walk-in freezer.

The lumens and incandescent equivalent specifications relate the CF to a standard tungsten bulb. For about the cost of running a 42W incandescent bulb, you get 150W of light—or more specifically, 2800 lumens. I qualify the cost because of the horrible power factor figure of >50 percent.

An incandescent bulb, which is simply a resistor that releases about 10 percent of the dissipated energy as light, has a power factor of essentially 100 percent. That is, the real power—in watts—is the same as the apparent power—volts × amps. In contrast, this CF has a power factor of >50 percent, which means that the volt-amp figure can be 50 percent greater than suggested by the 42W specification. The relationship between real power (watts), apparent power (volts × amps), and power factor can be expressed mathematically as follows:

$$\text{Power factor} = (\text{input watts}) / (\text{volts} \times \text{amps})$$

Given our 42W CF and a line voltage of 120VAC, what can we expect for the actual input current? If we were dealing with a resistor or incandescent bulb, each of which has a power factor of 1.0 or 100 percent, the current would be as follows:

$$\text{Input watts} = \text{power factor} \times \text{volts} \times \text{amps}$$

$$42\text{W} = 1 \times 120\text{V} \times \text{amps}$$

$$\text{Amps} = 42\text{W}/120\text{V}$$

$$\text{Amps} = 0.35$$

So, with a purely resistive load, and, by definition, a power factor of 100 percent, a 42W bulb would draw 0.35A. Now, with our CF lamp and a power factor of >50 percent (let's call it 50 percent for simplicity), we have this:

$$\text{Input watts} = \text{power factor} \times \text{volts} \times \text{amps}$$

$$42\text{W} = 0.5 \times 120\text{V} \times \text{amps}$$

$$\text{Amps} = 42\text{W}/(0.5 \times 120\text{V})$$

$$\text{Amps} = 0.70$$

Therefore, the power line has to supply our CF bulb with 0.7A. This shouldn't be a major problem with only one bulb, but with a dozen bulbs in a house, it could be significant. The higher current increases power losses in the distribution systems—recall power lost in wires and components is calculated as $P = I^2R$—and can result in higher charges from the utility company. This charge isn't automatic, because most residential wattmeters measure real, not apparent, power. However, utility companies can easily measure the power factor and adjust your utility bill accordingly.

The reason that the CF has a power factor of only >50 percent is that the input circuit is nonlinear. As you'll see later, the input consists of a pair of diodes that convert 120VAC to DC. This nonlinearity results in the generation of harmonic currents that can wreak havoc on the power distribution system. Better CF lamps have a THD of less than 10 percent. The rating of less than 150 percent for this CF is disappointing. The THD and power factor specifications are interrelated: improving the power factor will improve the THD.

Operation

There is little to say about operating a CF lamp. When you supply 120VAC, the lamp preheats the gases in the bulb for a moment, and then the electronic ballast ignites the lamp with a high-voltage charge. The ballast maintains the proper current flow until the power is removed.

Most CF lamps require full line voltage at all times, so they can't be used with dimmer switches. Similarly, although there are exceptions, CF lamps require good ventilation and risk premature failure and fire if operated in sealed fixtures.

Caution Another hazard when working around CF-generated light is exposure to ultraviolet (UV) radiation. When mercury vapor is ionized, it generates UV light, which excites phosphors in the wall of the spiral tube. These phosphors generate visible light. However, some of the UV light makes it through the tube unimpeded and causes damage to paint pigments, plastics, skin, and eyes. So, it's probably not a good idea to replace your halogen reading lamp with a CF lamp—even if you need to work on your tan.

Teardown

If you're good with your hands and the CF bulb that you're using isn't glued together, this is a 60-second teardown. Even if you have to resort to a hacksaw to open the plastic base, you shouldn't require more than 5 minutes start to finish. The real fun, of course, is figuring out how all those components work together to generate light.

Tools and Instruments

The basic teardown requires safety goggles, a pair of leather or heavy cotton work gloves to protect your hands from a broken fluorescent tube, and either strong fingernails or a standard screwdriver blade to pry apart the plastic housing. If you're working with a high wattage bulb (40W and above), you'll need a multimeter with a high-voltage probe to examine the live circuitry.

Most high-voltage probes are designed for a specific meter or series of meters. For example, I use a Fluke 80K-40HV probe with my Fluke 87 digital multimeter (DMM). The probe is designed so that when it's used with a meter with an input impedance of 10M, it functions as a 1000:1 divider. Figure 6-2 shows the simplified circuit diagram of the HV probe and a DMM with an input impedance of 10MΩ. The input impedance of my probe is about 79M, which is provided primarily by R2. R1 is about 80KΩ. R1 in parallel with the 10M input impedance works out to about 79KΩ, or a ratio of 79MΩ:79KΩ or 1000:1.

It's possible to use the Fluke or other HV probe with a multimeter with a lower input impedance than called for in the design. However, you'll have to multiply the displayed voltage by a correction factor because the displayed voltage will be lower than the actual voltage. See your probe's documentation to determine the correction factor for your multimeter.

The ground lead of the HV probe must be connected to earth ground for the divider to work properly. The closest earth ground is probably the ground pin of a power outlet. Earth ground isn't chassis ground, although a chassis could be

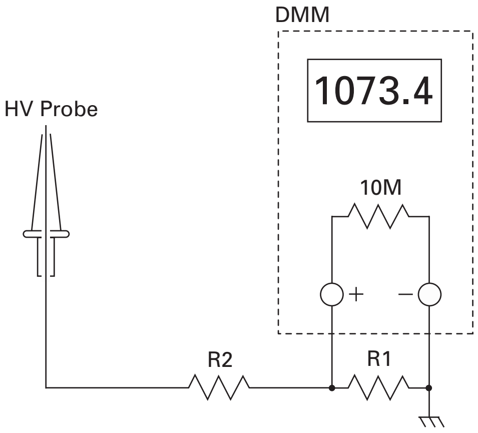

FIGURE 6-2 High-voltage probe circuit

connected to earth ground. If you fail to attach the ground lead of the HV probe to earth ground, you risk melting your DMM and possibly receiving a nasty shock if you serve as the ground return for the voltage you're trying to measure.

Step by Step

The teardown sequence shown in Figure 6-3 is more of a "crack open and look" than a real teardown. Even so, pay attention. If you happen to break a fluorescent tube, treat whatever is left of the tube as a biohazard and seal it in a plastic bag. Check with your public health office for how to best dispose of the hazardous waste.

Step 1

Open the base of the bulb. Grasp the top of the base—not the glass tube—with your gloved hand. Position your hand so that, should the glass bulb break, shards of glass would be blocked from your face and other hand. Now, with your other hand, use your nails, a standard screwdriver, a dime, or a similar thin tool to pry open the enclosure. I used my nails on the base, but you may want to use a screwdriver. Just don't run the screwdriver blade through the electronics and wires in the process of opening the base.

Another option is to use a hacksaw to cut away the plastic skin, right along the seam. However, remember that just below that thin, plastic skin are a dozen or so components. If you opt for the hacksaw approach, secure the lowest part of the base—the part farthest from the bulb—in a vise. And don't hit the glass tube with your hacksaw.

Regardless of the method you use, you should see something like Figure 6-3a when you first crack open the lamp.

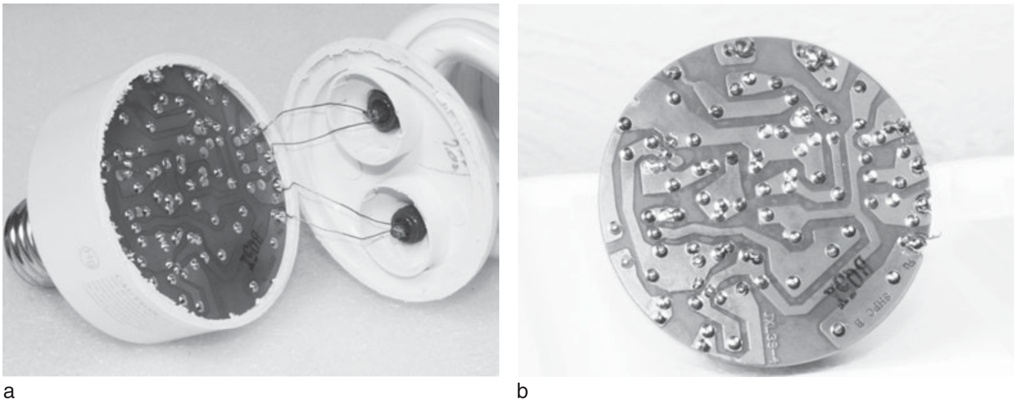

FIGURE 6-3 Teardown sequence

Step 2

Use a pair of diagonal cutters to sever the four wires between the bulb and the circuit board. Put the glass bulb aside for now. Examine the foil side of the circuit board, shown in Figure 6-3b. Note the generous spacing and thickness of the traces, as you'd expect for a high-voltage, high-current circuit.

Layout

The layout of every CF that I've cracked open is remarkably similar, despite differences in power and manufacturer. They all share the same basic layout as the populated circuit board shown in Figure 6-4.

As you can see, the circuit board is a disc, with the two switching transistors and inductors toward the periphery. The electrolytic capacitors are in the center of the disc. The silver mica capacitors, including the starter capacitor used to ignite the bulb, are located on the periphery of the board, along with wires to AC power and the bulb.

Components

Regardless of the make and model of your CF lamp, it probably has a spiral lamp with dual cathode filaments, a pair of switching transistors, a ferrite core toroidal transformer, a pair of inductors, a handful of silicon power diodes, a half-dozen capacitors, and a DIAC (diode for alternating current).

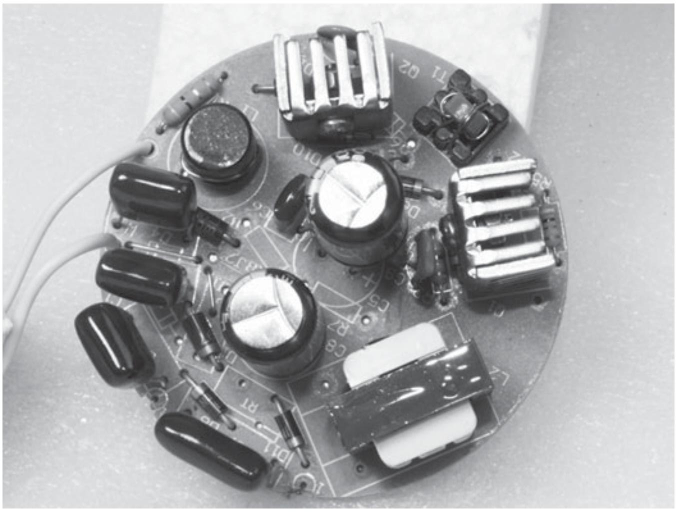

FIGURE 6-4 CF layout

Fluorescent Lamp

The spiral fluorescent lamp is filled with low-pressure mercury vapor that, when ionized by an electrical current, generates UV light. A mixture of phosphors in the tube absorbs this UV radiation and emits lower energy radiation in the visible spectrum.

Most CF bulbs designed for residential use have a rapid start system that requires a combination filament-cathode at each end of the spiral tube. The two 1.5Ω filaments in the SpringLamp CF are used to preheat the bulb for a few tenths of a second, lowering the voltage necessary for ionization of the low-pressure gas mixture.

Transistors and Heat Sinks

This CF features two D13009K NPN switching transistors rated at 400V collector-emitter voltage and 5A collector current. They're packaged in a standard TO-220 case, which has a metal tab extending up from the back of the transistor. As is standard for a TO-220 transistor, the leads are base, collector, emitter, looking from the front of the device. Figure 6-5 shows multifingered heat sinks bolted to the metal tab of each transistor. The heat sinks increase the heat dissipation from each transistor via natural convective cooling. Note the 100mH inductor on the left and a toroidal transformer on the right.

This is a good point to focus on the heat sinks used in this CF and on transistors in general. The goal of using a heat sink is to keep the transistor junction below the maximum rated temperature. But how much of what kind of heat sink is enough?

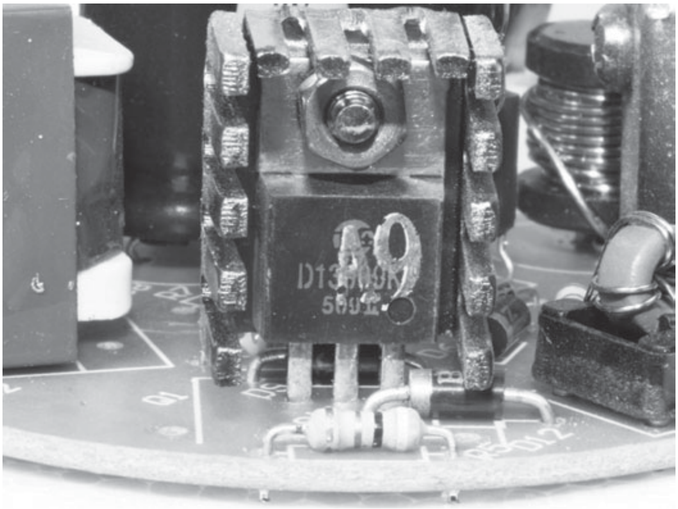

FIGURE 6-5 NPN switching transistor with heat sink

If you search the Web and basic electronics texts, you'll find dozens of formulas that can be used to define the optimum heat sink for any circuit and environment. Most of these formulas consider the temperature and velocity of the ambient air, the amount of heat that must be dissipated, the barometric pressure, and other factors, to determine the heat sink dimensions, composition, number of fins, and required thermal efficiency.

In addition to considering theoretical calculations, practical economic and physical considerations are important in selecting a heat sink. For example, although beryllium copper has a higher thermal conductivity than aluminum, it's generally too expensive to use as heat sink material in CF lamps. Similarly, a massive extruded aluminum heat sink might provide the best protection against thermal overload, but the space limitations imposed by the CF base demand more compact heat sink designs. There's also the issue of availability. If you check parts suppliers for bolt-on heat sinks, you'll find choices for TO-220 devices are limited to aluminum designs rated from 1W to 2.5W. In the end, you go with what you can afford and what fits in the space allotted to the heat sink.

Toroidal Transformer

If you've worked with switching power supplies, the toroidal transformer shown in Figure 6-6 should look familiar. This high-frequency transformer, which, according to my measurements operates at 50kHz, has three windings. Each winding consists of two turns of enameled copper wire. The plastic base provides support and insulation from the underlying circuit board.

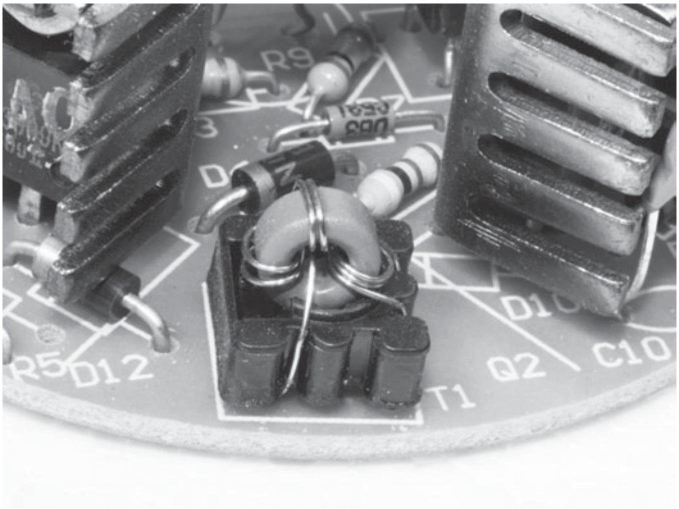

FIGURE 6-6 Toroidal transformer

Note the two transistors on either side of the transformer. The unbanded DIAC is visible in the background.

Inductors

This circuit has two inductors: a 100µH ferrite core inductor used to limit EMI induced into the 120VAC line (see Figure 6-7a), and a much larger 2.1mH iron-laminate core inductor used to limit bulb current once the lamp has been started (see Figure 6-7b).

Capacitors

At the center of the disc are two 47µf electrolytic capacitors rated at 250VDC. The remaining capacitors are silver mica rated between 400 and 1250VDC. Figure 6-8 shows several 0.1µf at 400VDC silver mica capacitors in the foreground and the two radial lead 47µf electrolytic capacitors in the background. Silver mica capacitors are a good choice for this application because of their stability over a wide range of temperatures and availability with high working voltages. Note the diodes adjacent to the capacitor leads in the figure.

Resistors

Several generic 0.25W metal film resistors are in this circuit. Of note, however, is the 0.47Ω fusible metal film resistor in series with the 120VAC mains. This resistor is designed to fail flamelessly in the event of other component failure.

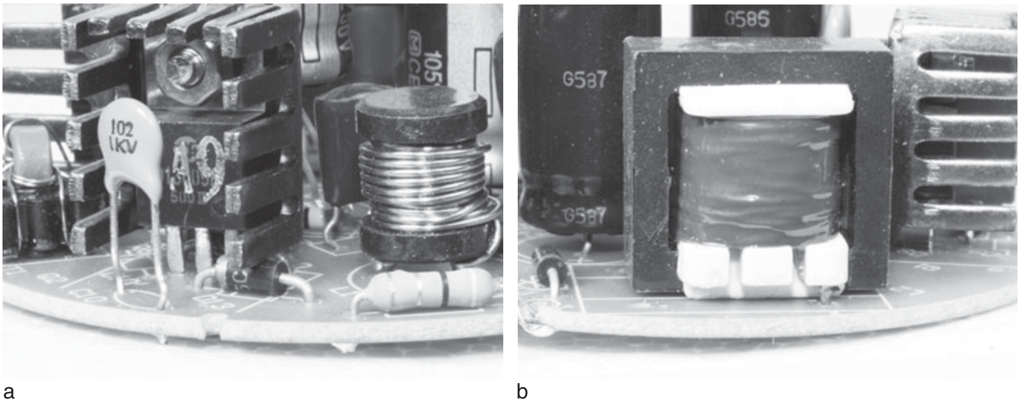

a b

FIGURE 6-7 Ferrite and iron-laminate core inductors

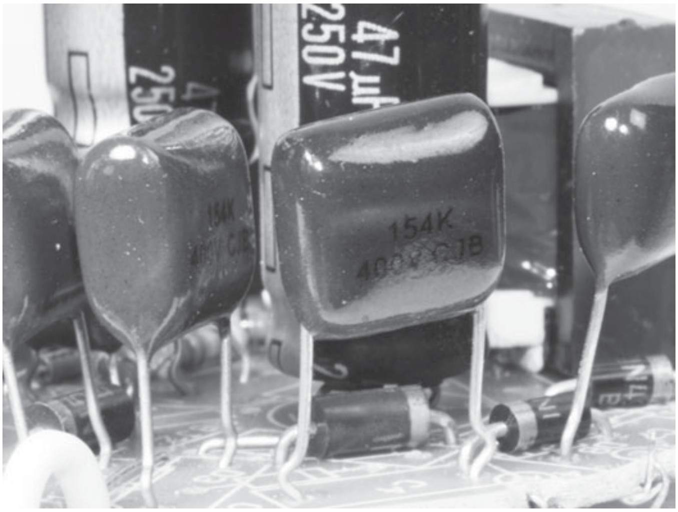

FIGURE 6-8 Silver mica (foreground) and electrolytic capacitors

Note I like to think of fusible resistors as ultra–slow-blow fuses. It might take a minute or more for the resistor to fail—more than enough time for a mosquito that falls across a capacitor's leads to be vaporized, for example. This resistor is designed to save your house from burning down, should your CF lamp fail.

Diodes

As illustrated in Figures 6-4 and 6-8, several diodes are included in this CF circuit: two for the main rectifier, two associated with the lamp, and five associated with the transistors. Two 1N5397 diodes are part of a voltage doubler rectifier circuit. These diodes have a forward current capacity of 1.5A and PIV (peak inverse voltage) rating of 600V.

Each transistor is protected from peak reverse voltages by a pair of BYV26 diodes. These are ultrafast recovery diodes with surge current capacity of 30A and a PIV of 600V. There is a diode from base to emitter and from collector to emitter of each transistor. The diodes essentially prevent the base-emitter or collector-emitter junctions from being reverse-biased by a potentially damaging transient.

A fast recovery 1N4937 diode is used across each of the filament cathodes of the fluorescent bulb. The polarities of the diodes are such that when one diode is forward-biased, the other is reverse-biased. I assume this design is intended to extend the filament lifespan of the bulb. Can you provide another explanation?

A 1N4005 generic silicon power diode rated at 1A and 600V PIV is used in the oscillator circuitry associated with one of the transistors. More intriguing, however, is a DB3C531 DIAC that is connected to the base of one of the transistors. A DIAC is essentially a momentary latching switch formed by the equivalent of two Shockley diodes wired in parallel, cathode to anode. Once a DIAC's firing voltage is reached, it latches on until the voltage drops below a level sufficient to keep the device conducting. With a sinusoidal AC signal, this cyclical firing, latching, and low-current dropout occurs twice per cycle.

How It Works

Figure 6-9 provides a high-level view of CF lamp operation. A line-powered AC-to-DC converter feeds a DC-to-high frequency AC converter, commonly referred to as an *electronic ballast*. The function of the ballast is to generate a high voltage across the bulb to ignite or fire the CF lamp and then reduce the voltage and regulate the current through the lamp for the duration of operation.

The key components in Figure 6-9 are the capacitor in parallel with the fluorescent tube, labeled *CP*, and the inductor and capacitor in series with the lamp, labeled *LS* and *CS,* respectively. While all three components form part of the self-oscillating electronic ballast, the primary role of CP is to build up voltage to fire the fluorescent bulb. The series inductor, *LS*, regulates the flow of current once the bulb starts conducting.

The series capacitor, *CS*, is used to block DC to the bulb and insure that only AC contributes to current flow. Operating a fluorescent bulb with DC causes *cataphoresis*, the migration of mercury vapor and other gases to the end regions of the bulb where they do not to contribute to illumination. The value CS is significantly larger than CP, so CS does not interfere with the buildup of voltage across CP.

Let's plug in some component values and explore the circuit. Assume an operating frequency of 50kHz, a CP of 5.8 nanofarad (nF), a total filament resistance

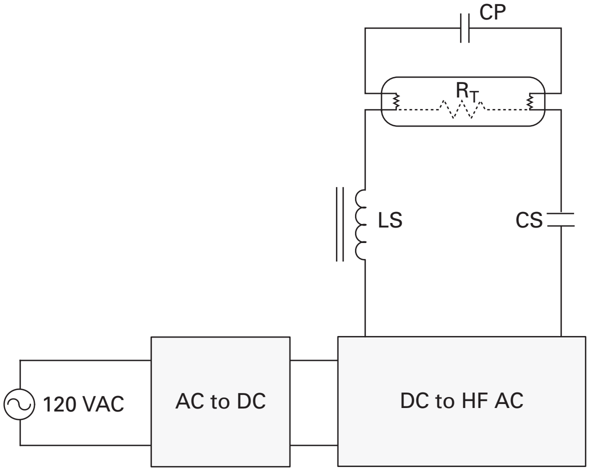

FIGURE 6-9 Block diagram of the CF lamp

of 3Ω, an LS of 2.1mH, and a CS of 0.1µf. When the HF AC is applied, current flows through LS, one filament of the lamp, CP, through the other filament of the lamp, and CS. CP has the following impedance:

$$Xc = 1/(2\pi fc)$$

$$Xc = 1/(2\pi \times 50 \times 10^{3} \times 5.8 \times 10^{-9})$$

$$Xc = 550\Omega$$

Using the same formula, you should be able to determine that the impedance of CS is only 15Ω. As such, the majority of the voltage drop will be across CP.

Prior to firing, the resistance of the bulb, RB, is several megaohms, measured from filament cathode to filament cathode. With each cycle, voltage builds up across CP and the bulb. Simultaneously, the filaments decrease the firing voltage by warming and contributing ions to the mercury vapor inside the bulb. At around 500V, the bulb fires, and the resistance of the bulb, RB, drops to about 100Ω, essentially shunting CP and removing it from the circuit. Next, LS and CS, together with the other components in the electronic ballast, control the frequency and magnitude of current through the bulb.

I shorted the filaments and measured the voltage across CP and the bulb. Voltage across the bulb rose exponentially from a few volts to about 600VAC over a period of 1.5 seconds. This cycle repeated dozens of times before I disconnected the circuit to prevent overheating of the transistors.

For steady state operation, the bulb requires about 220VAC from cathode filament to cathode filament. The reason for this relatively low voltage is the operating frequency. At 50kHz, the gas in the bulb remains virtually fully ionized from cycle to cycle. If the bulb were operated at 60Hz, the gas would deionize between cycles and a higher voltage would be required to reionize the gas. As noted earlier, another benefit of the 50kHz operating frequency is the relatively small LS required to limit current flow and prevent the lamp from self-destructing.

Now let's examine the power supply in detail. As shown in the simplified schematic in Figure 6-10, the power supply includes a simple voltage doubler. The 120VAC line is connected to a 0.47Ω fusible resistor and an EMI filter composed of the 100µH ferrite core inductor and two 0.1µf at 400VDC silver mica capacitors. The two 1N5397 silicon power diodes, D1 and D2, charge the two 47µf electrolytic capacitors that are connected in series.

When the hot lead of the applied AC is positive relative to the neutral lead, D1 conducts, charging C1 approximately to the line voltage. When the hot lead is negative relative to the neutral lead, D2 conducts, charging C2. In this way, the two capacitors in series provide the CF circuitry with 240VDC.

A voltage doubler is a low-cost, low-bulk means of providing the higher voltage needed by more powerful CF lamps. A step-up transformer followed by a bridge rectifier would be more efficient, but it wouldn't fit on the circuit board. Lower powered lamps often employ a standard four-diode bridge rectifier and operate at 120VDC.

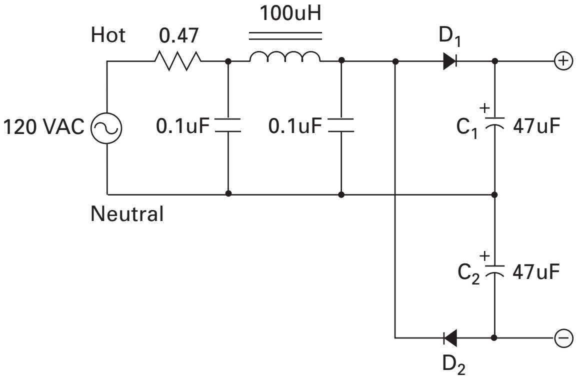
100uH
Hot
0.47
D1
120 VAC
0.1uF
0.1uF
C1
47uF
Neutral
C2
47uF
D2

FIGURE 6-10 Simplified schematic of power supply

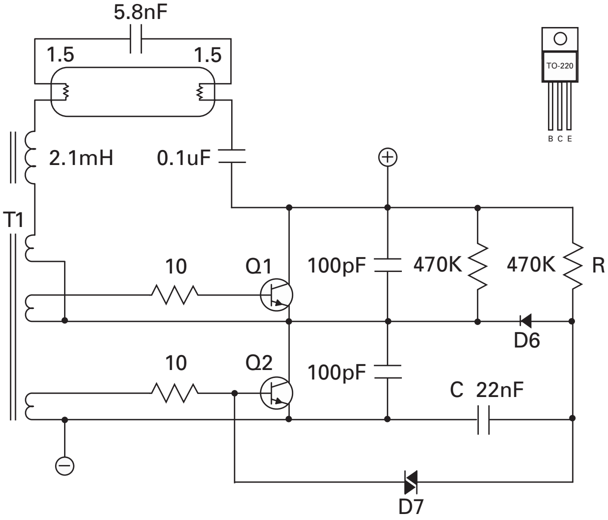
5.8nF
1.5
1.5
TO-220
B C E
2.1mH
0.1uF
T1
10
Q1
100pF
470K
470K
R
D6
10
Q2
100pF
C 22nF
D7

FIGURE 6-11 Simplified schematic of CF lamp

Figure 6-11 shows the simplified schematic of the CF lamp, less the power supply. As you can see, each of the two NPN transistors is connected to a winding of the toroidal transformer T1, and the third winding is connected in series with the bulb circuit. The DIAC is connected to the base of Q2. Note the TO-220 package outline with pinouts for Q1 and Q2.

Despite the relatively low component count, a lot is going on in the circuit that isn't obvious from inspection of the schematic. True analysis requires a multiple-trace oscilloscope and intimate knowledge of the components, especially the transformer, T1, and the transistors, Q1 and Q2. However, at a high level, the manufacturer describes circuit operation as something like the following.

At the moment the power switch is thrown to the on position, the 470KΩ resistor R1, the 22nF capacitor on the emitter of Q2, and the DIAC D7 supply base current to Q2. Q1 is off. Once the bulb fires, the 1N4005 diode D6 inhibits the DIAC. Self-oscillation is maintained by feedback from the base lead of Q2 through the transformer T1. The toroidal core of T1 is formulated such that it easily saturates, thereby removing feedback from the base of Q2, which switches off after a delay defined by the storage time. Meanwhile, T1 switches on Q1, and the circuit oscillation continues.

To get a better handle on this explanation, it helps to know that transistor storage time, or more properly base storage time, is a characteristic of bipolar transistors that limits the maximum speed at which a transistor can be turned off. A charge is stored in the base-collector region of the transistor whenever the transistor is driven to saturation, and this charge takes time to discharge. Q1 and Q2 are both driven to saturation in this circuit.

The 50kHz oscillation frequency is established by the saturation properties of T1 and the transistor storage time—factors also not apparent from the schematic. The saturation properties of T1 are determined by the composition of the toroidal core. The more readily the core saturates, the faster the transistors can switch on and off, within the limits imposed by storage time.

There is a bit of magic in the circuit. The dynamic properties of T1 and storage time don't appear in the schematic, and without these details, it's impossible to specify the changes exactly from startup to steady state. However, I'll offer my impression of what probably happens in the circuit to supplement the condensed version supplied so far.

At the moment the power switch is turned on, the leads of Q1 are at the same potential—the supply voltage—and so the transistor is inactive. In contrast, the emitter of Q2 is at ground, the collector is at positive supply voltage, and the base is forward-biased. In addition, capacitor C charges, and the DIAC fires and latches on. Q2 continues to receive forward-bias from the DIAC and 10Ω resistor in the base lead to ground. Due to T1 saturation, Q1 begins to conduct and has a path to ground through Q2, which is turning off, but not yet off because of storage time. As I see it, it's the slight overlap in on–off time that makes the self-oscillation possible. Otherwise, if Q2 suddenly and completely shuts off, Q1 can't conduct.

Mods

Because a CF bulb is a household appliance, I can't recommend any safe mods if you intend to put the bulb back in regular service. As a bench experiment, you might want to try altering the self-resonant frequency of the CF circuit and record the changes in bulb current, for example. A more appropriate use of

CF lamps, especially those otherwise destined for disposal, is to repurpose the components.

If you search the Web you can easily find examples of a variety of circuits developed with repurposed CF components. As an amateur radio operator (callsign NU1N), one of the more creative ideas I found is to repurpose the components found in a CF lamp to create a radio transmitter. You'll have to rewind T1 and add a crystal to establish the operating frequency. However, other than a DC power supply, everything else can be scavenged from a CF lamp. Coming up with a complete receiver with repurposed components would be more of a challenge.

This page intentionally left blank

Chapter 7

Ultrasonic Humidifier

I live in the US Northeast, where the relative humidity, a measure of the amount of water vapor suspended in the air, routinely dips below 20 percent in the winter. In this environment, electrostatic discharges are an everyday annoyance and a serious threat to sensitive electronic components. The "dry air" also makes it tough to sleep at night. A popular way to avoid this scenario is to use a room humidifier, such as the ultrasonic humidifier shown in Figure 7-1, to raise the relative humidity. A value of around 50 percent is comfortable and suppresses electrostatic buildup.

In this chapter, I'll tear down a Crane ultrasonic humidifier, which is typical of the inexpensive room humidifiers available online for about $30. I'll discuss some ideas for repurposing the components in a mod at the end of the chapter.

Highlights

The humidifier consists of a decorative water container supported by a base filled with electronics. The unit features dual linear power supplies, two transistor-based switching circuits, and a high-voltage radiofrequency (RF) oscillator that drives a submerged piezoelectric transducer at about 1.2MHz. The humidifier also uses a brushless DC motor that we'll examine in detail. During the teardown, note the following:

- The open loop control system
- The safety features, including dual fuses and magnetic reed cutoff switch
- The process of circuit analysis

Specifications

Following are the key technical specifications listed by the manufacturer:

- 32W power requirement
- One gallon capacity with 11-hour runtime

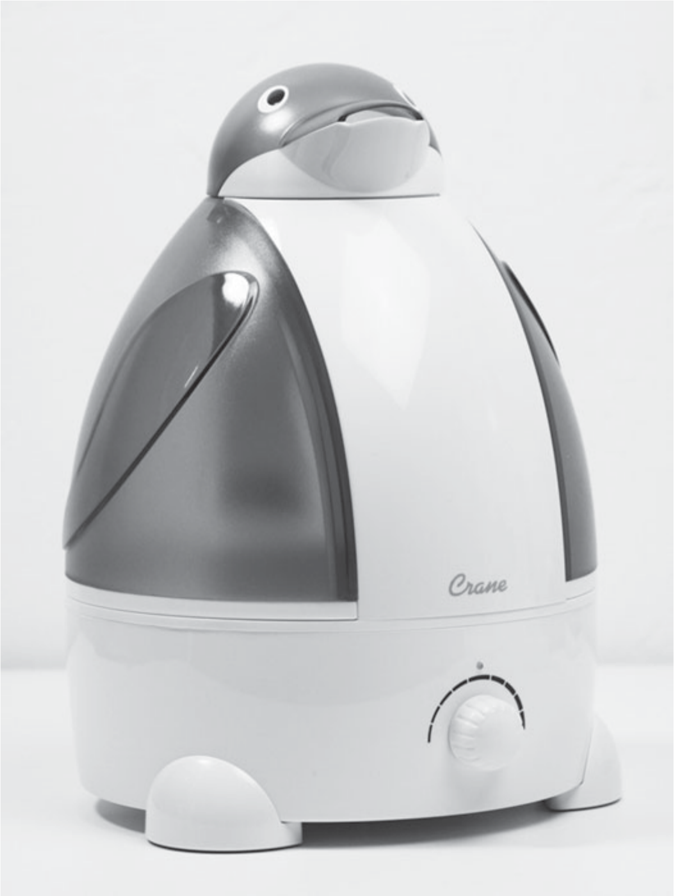

FIGURE 7-1 Crane ultrasonic humidifier

- Auto off function
- Variable humidity control

It's a stretch to claim that the unit has a variable humidity control. The single control knob simply varies the excitation of the ultrasonic transducer and therefore the vapor output. As you'll see during the teardown, the unit has no humidity sensor and no humidity feedback mechanism.

Operation

This is a plug-and-play device. Fill the tank with water, place it on the base, toggle the on–off switch to the on position, and adjust mist output with the knob at the base of the unit. A bicolor LED in the base of the unit glows green as long as there is enough water to operate the ultrasonic transducer. When water level above the transducer falls below a few centimeters, a floating magnet drops, causing a magnetic reed switch to open, which kills the ultrasonic transducer and fan. This also switches the LED from green to red, indicating that it's time to refill the tank.

Teardown

You should be able to complete the basic teardown, illustrated in Figure 7-2, in about 15 minutes if you're working with the same humidifier. Most ultrasonic humidifiers on the market share the basic components of this humidifier—a power supply, fan, output control, and ultrasound transducer. From a control perspective, this is an open loop system, in that the power sent to the ultrasonic element is independent of the ambient relative humidity.

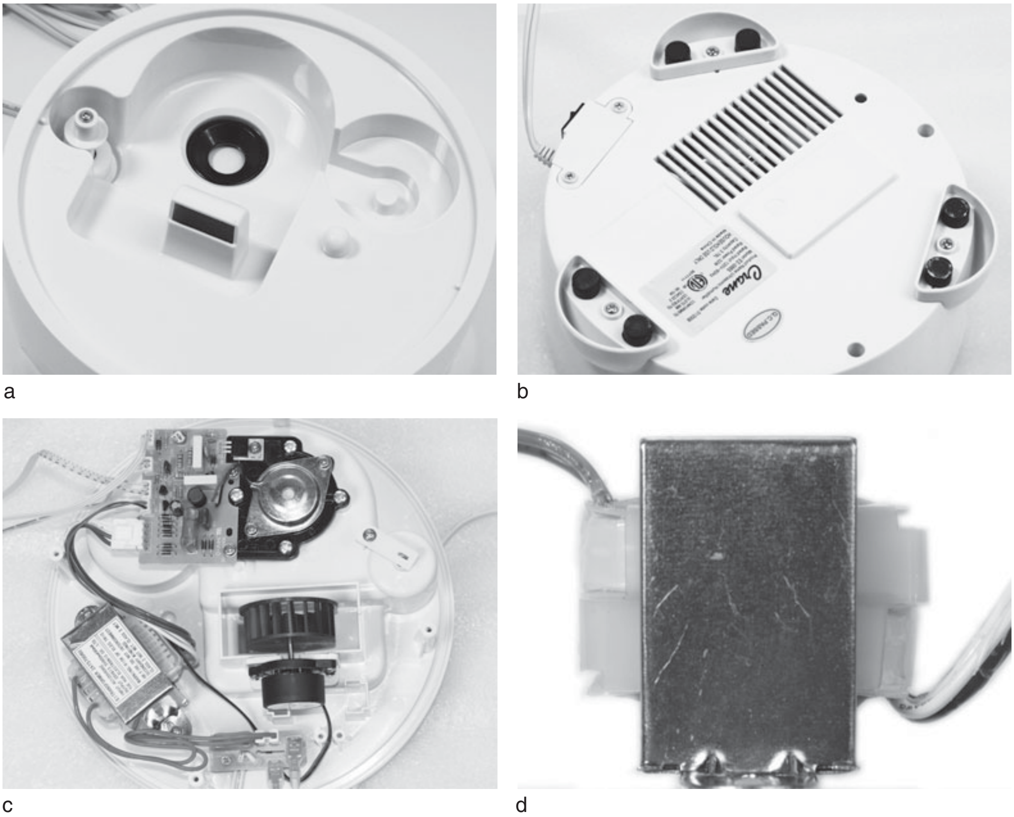

FIGURE 7-2 Teardown sequence

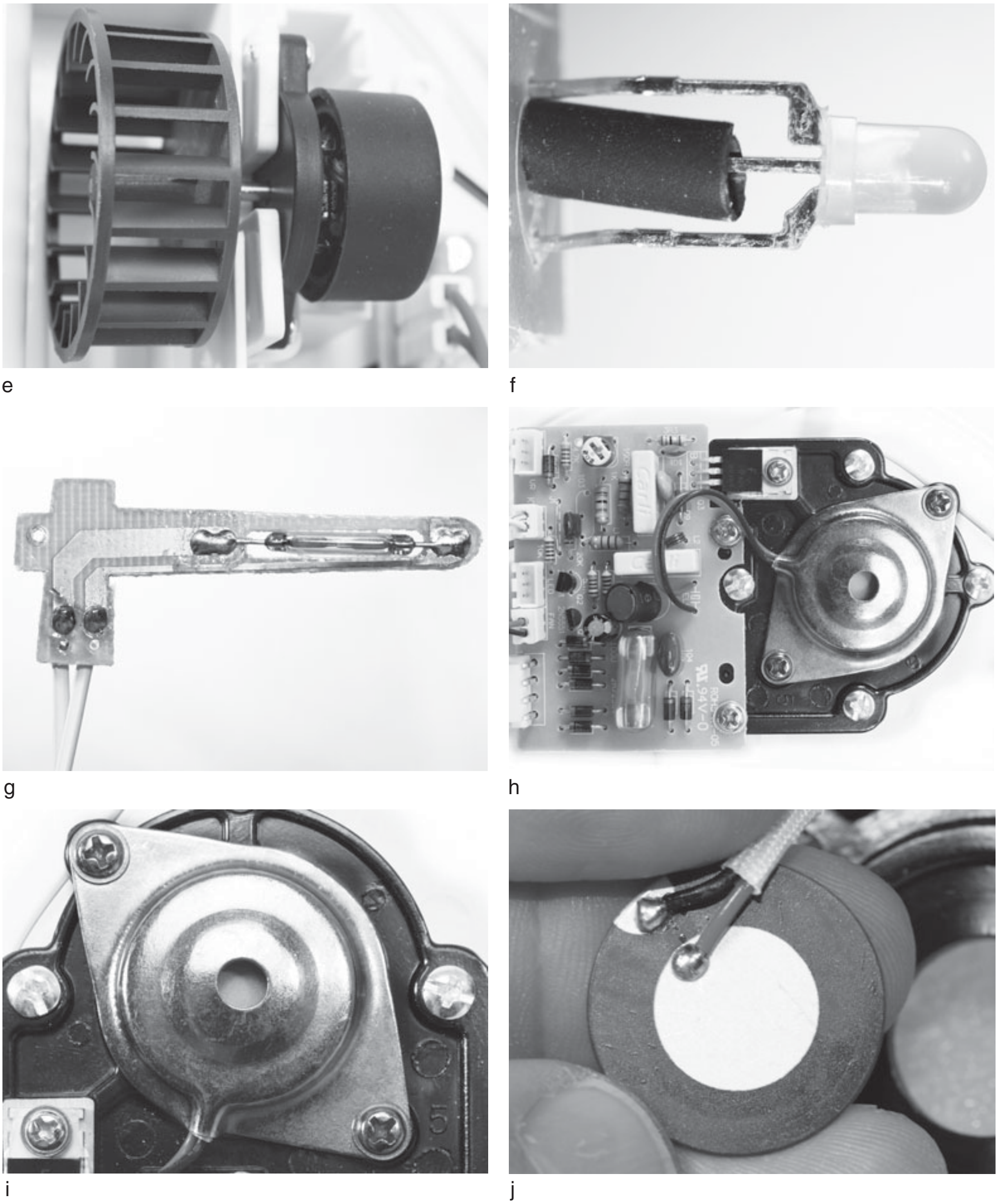

e f g h i j

FIGURE 7-2 (*continued*) Teardown sequence

If you're following along at home with another model of humidifier, your unit may have a relative humidity sensor that is used to control mist output. I'll discuss the implications of a closed system operation later in this chapter.

Caution At this point, it's important to note that, despite its appearance, this electric appliance isn't a toy. Not only are we dealing with the potentially dangerous combination of water and electricity, but the ultrasonic transducer can hurt you. The transducer in a typical home ultrasonic humidifier delivers up to 15W—and that's a lot of power at ultrasound frequencies. As a point of reference, the output of therapeutic ultrasound machines is measured in milliwatts.

Before you think of touching the active transducer in your humidifier, keep in mind that this device is designed to shred water at the molecular level, and remember that your finger is mainly water. Although I've not experienced the jolt from an ultrasonic transducer firsthand, I'm told that touching an active transducer is akin to having your finger slammed in a car door. And that, from personal experience, is something to avoid. I'm not suggesting that you pass up the opportunity to repurpose the components to create an ultrasonic meat tenderizer. Just don't touch the active transducer.

Tools and Instruments

You'll need a Phillips-head screwdriver, crescent wrench, and needle-nose pliers. A multimeter capable of measuring capacitance and inductance modes will come in handy.

Step by Step

If you plan to rebuild your humidifier once you're done with the teardown, sort the screws and other hardware in a cupcake pan or similar compartmentalized container. Use a permanent marker to identify connectors before you disconnect them. And pay particular attention to the integrity of the seal around the ultrasonic transducer—a damaged seal will allow water into the electronics compartment.

Unplug the unit before you begin.

Step 1

Examine the base. Remove the decorative water tank, leaving the base, shown in Figure 7-2a. Thoroughly dry the base with a paper towel. With the output control knob facing you, identify the depressed area with the black-rimmed white disc toward the rear of the base. This disc, about the size of a quarter, is the ultrasonic transducer. Toward the left of the transducer, near the left rim of the base, is the magnetic float assembly. When the water level rises, a ring magnet embedded in a buoyant plastic ring moves with the water level and interacts with a magnetic reed switch on the dry underside of the base. The rectangular fan outlet appears just below the transducer in the figure.

Step 2

Crack the base. Flip the base upside down to reveal the fan input grill and screws securing the feet and base, as shown in Figure 7-2b. Remove the screws and lift the bottom of the base to reveal the electronics.

Arrange the base—actually the underside of the base top, which separates the water from the electronics—as shown in Figure 7-2c. At the top of the figure are the main printed circuit board and ultrasonic transducer. Connectors and cables for sensors and controllers run along the left edge of the circuit board. Near the right rim of the base is the magnetic reed switch used to determine water level. Near the center of the base is the 12VDC brushless squirrel cage fan. At the bottom of the figure is a board for the switched AC input and fuse. In the left lower quadrant of the base is the dual-secondary power transformer.

Step 3

Examine the live circuit. Plug in the unit and measure transformer secondary voltages and examine the logic of the magnetic reed switch. If you have an oscilloscope or frequency meter, measure the operating frequency of the ultrasound generator. However, note that the ultrasonic transducer is designed to operate submerged. If you want to measure voltages and currents with a dry transducer, manipulate the magnetic ring so that power is applied to the ultrasound circuitry in 1- or 2-second bursts—just long enough to get solid measurements. You can also unsolder the transducer to avoid overheating and overstressing the component.

Unplug the unit when you're done.

Step 4

Extract the small circuit board. Remove the small board that holds the 2A at 250V fast-blow fuse and that provides quick-disconnect connections to the power cord and the primary of the power transformer.

Step 5

Extract the dual-secondary power transformer, which provides 28VAC at 800mA and 13.5VAC at 90mA secondary output voltages. Note the construction with the insulated coils for the primary winding on the top half of the transformer and secondary windings on the bottom half, as shown in Figure 7-2d. The transformer is virtually waterproof, thanks to a thick layer of clear sealant on the windings and laminated steel core. As with the small printed circuit board, the transformer is physically isolated from other components in the plastic base of the humidifier.

Step 6

Release the brushless DC fan. The squirrel cage fan blades, shown in Figure 7-2e, are pressure-fit on the motor shaft. Simply pull the cage along the axis of the shaft. Without disconnecting the power cable, remove the two screws securing the motor to the plastic support and set the motor down in the base for now.

Step 7

Extract the bicolor LED and 5KΩ potentiometer from the front of the base. You'll need to use a crescent wrench or nut driver to remove the retaining nut from the potentiometer. Figure 7-2f shows the three-lead, common cathode bicolor LED. Don't disconnect the potentiometer cable from the main circuit board.

Step 8

Extract the magnetic reed switch assembly. Partially remove the screw securing the plastic pressure tab and rotate the tab off of the magnetic reed switch assembly, shown in Figure 7-2g. Pull the assembly straight up, out of its well in the base. You may find it easier to extract the sensor if you temporarily unplug its cable from the main ultrasonic transducer board.

Step 9

Extract the main ultrasonic transducer assembly. Carefully remove the four Phillips-head screws that attach the large aluminum heat sink to the plastic base, as shown in Figure 7-2h. The assembly includes a heat sink about the same size as the main circuit board. Note the clear silicon ring seal between the plastic base and heat sink.

Step 10

Extract the piezoelectric transducer. Remove the two screws securing the silver metal retainer to the black heat sink, shown in Figure 7-2i. Peel the black rubber seal from the piezoelectric disc, taking care not to damage the two wires connecting the transducer to the main printed circuit board. The free transducer is shown in Figure 7-2j.

Layout

The layout of the main circuit board is clean and relatively spacious; you shouldn't have any trouble probing component leads. As a bonus, the component side of the board, shown in Figure 7-3, is marked with component values. As you can see, the board contains traditional leaded components. The connectors along the bottom edge of the board, from left to right, are for the control potentiometer (VR), the magnetic reed switch (K), the bicolor LED (LED), the fan (FAN), and power (C).

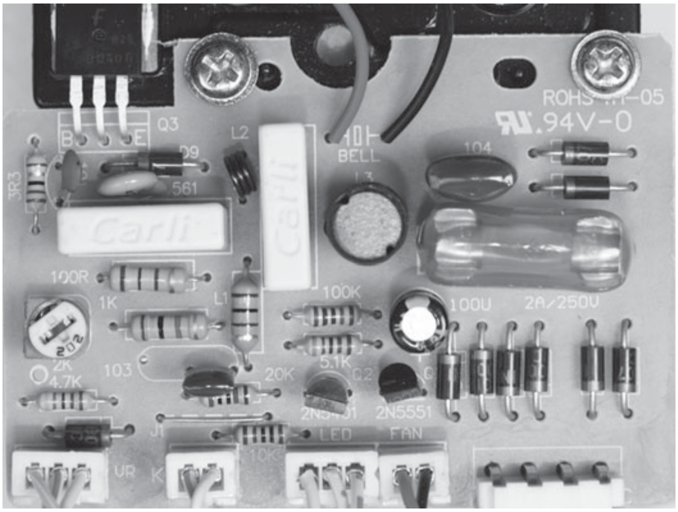

FIGURE 7-3 Main circuit board layout

The right side of the board is populated with components for the two power supplies—eight standard silicon power diodes configured as two bridge rectifiers, one for the 28V and one for the 13.5V transformer outputs. The 100µf at 25VDC electrolytic capacitor (100U) is a filter capacitor in the 13.5V circuit, and the 2A at 250V fast-blow fuse (2A/250V) is in the 28VDC circuit that feeds a high-voltage oscillator. The 0.1µf Mylar film capacitor (104) is used as a bypass capacitor in the power supply.

Moving to the left, immediately adjacent to the fan and LED connectors, are the small switching transistors in TO-92 packages, a 2N551 NPN (negative-positive-negative) silicon transistor for controlling the fan motor and a 2N5401 PNP (positive-negative-positive) silicon transistor for controlling the red element of the bicolor LED.

On the left side of the board, behind the connectors for the magnetic reed switch and control pot, is the drive circuitry for the ultrasonic oscillator. The 2KΩ trim pot (potentiometer) (2KΩ) is wired in series with the external 5KΩ control potentiometer. The oscillator transistor, a BU406 high-voltage NPN silicon transistor in a TO-220 case, is bolted to the large heat sink. You can see a small amount of conductive grease between the transistor's metal tab and the large aluminum heat sink. The transistor's metal tab, which is connected to the collector, is at 28VDC during operation of the humidifier—something to keep in mind if you're probing the live circuit.

Other significant components in the area include an open-air RF inductor composed of three loops of enamel wire (L2), a cylindrical RF inductor (L3) adjacent to the electrolytic filter capacitor, and a 100µH inductor (L1). Adjacent to the open-air RF inductor are two ceramic disc capacitors (101 and 561) and two large, rectangular metalized polypropylene film capacitors.

Components

This simple humidifier provides examples of a variety of components that you'll see in other appliances and in other teardowns in this book. When you discover components during a teardown, make a habit of looking up the datasheets on the Web. Datasheets often feature example circuit diagrams and other information that may not be evident in the teardown.

Capacitors

As discussed earlier, the humidifier employs electrolytic, Mylar film; metalized polypropylene film; and ceramic disc capacitors. The two Mylar film capacitors, marked "2A103J" (0.01µf) and "2A104J" (0.1µf), are perfect for high-humidity environments because they are dipped in a hard epoxy coating for moisture resistance. The small ceramic disc capacitor, marked "101" (100pF), is installed between the base and collector of the oscillator transistor. The large, resin coated disc capacitor, labeled "561K 2KV" (560pF), has a 2KV voltage rating.

Piezoelectric Transducer

The heart of the humidifier is also one of the simplest components, the piezoelectric transducer, which imparts ultrasound energy from the oscillator circuit to the water. There are no markings on the transducer in my humidifier. However, I'd expect the datasheet to include information on the resonant frequency, static capacitance, mist production in cc/minute, voltage input, and maximum drive capacity. The voltage across the disc during operation is more than 120V peak-to-peak (VPP).

Diodes

The ten diodes used in the humidifier are general-purpose 1N4007 silicon diodes rated at 1A and 1000V PIV (peak inverse voltage). Two sets of four diodes are used in bridge rectifier circuits: one is used to protect the high-voltage oscillator, and the other isolates the oscillator circuit from the other circuits on the board.

Bicolor Red/Green LED

The bicolor LED used to signal the status of the humidifier has a common cathode center lead. The LED emits red light when the anode corresponding to the red LED is positive, and it emits green when the other anode is positive.

Magnetic Reed Switch

The slender magnetic reed switch is normally open. In this application, when there is sufficient water in the base, the floating ring magnet keeps the reed switch closed. If the water level drops sufficiently, the ring magnet moves away from the reed switch, allowing it to open.

SPST Rocker Switch

The SPST (single pole, single throw) rocker switch is rated at 6A at 125VAC. If you repurpose this switch for a DC application, remember to derate the current capacity rating by at least 50 percent.

RF Inductors

In addition to L1, the banded 100µH inductor (brown-black-brown-silver), are two unmarked inductors. We can estimate the value of L2, the air-core coil, with the aid of the following formula:

$$L\ (\mu H) = d^2n^2/(18d + 40ln)$$

where d = coil diameter in inches, from wire center to wire center; ln = length of coil in inches; and n = number of turns. My measurements for the RF inductor are as follows: d = 0.16 inch; ln = 0.08 inch; and n = 3. Substituting these measurements into the formula, we have this:

$$L = 0.16^2 3^2/(18 \times 0.16 + 40 \times 0.08)\mu H$$

$$L = 0.23/6.1 = 0.038\mu H$$

Estimating the value of L3 is more problematic because of the unknown permeability of the ferrite core. Ferrite can provide 20–1000 times more inductance per turn than an open-air inductor of the same size. According to my measurements with an LC meter, the inductance of L3 is 32µH.

Heat Sink

The 1.1 ounce (31g), 2.5 × 2.6 inch (65 × 67mm), aluminum heat sink seems more than adequate to cool the oscillator transistor, even with continuous, day-after-day operation. Given the mass of the heat sink and its contact with water, it provides more than sufficient heat capacity for the transistor. The ultrasound transducer is insulated thermally and electrically from the heat sink by a silicone seal.

Resistors

The main circuit board contains a handful of 0.25W metal film resistors; a 0.1W, 2KΩ trim pot; and an external 0.5W, 5KΩ potentiometer. Of note is the similarity of the 100Ω resistor and 100μH inductor. The two components are about the same size and share a color code (brown-black-brown-silver), but the resistor has capped ends and the inductor has conical ends.

Transistors

As noted earlier, three silicon transistors are used in the humidifier, a 2N551 NPN and 2N5401 PNP low-power switching transistor and a hefty BU406 NPN high-voltage transistor. The datasheets for the transistors reveal a current gain (Hfe) of 50–250 for the 2N551 and 2N5401 and 20–40 for the BU406. Let's review the on and off states for NPN and PNP bipolar transistors used as switches, using Figure 7-4 as a guide.

Recall that an NPN bipolar transistor can be configured as a switch between emitter and collector by manipulating the emitter-base current. When the base is positive relative to the emitter, a small current flows from emitter to base and a significantly larger current flows from emitter to collector. The amount of emitter-base current we need to switch a particular load is related to the load and the transistor current gain. With a current gain of 100 and a load of 100mA in the emitter-collector circuit, we'd need an emitter-base current of 100mA/100, or 1mA. Similarly, with a current gain of only 50, we'd need an emitter-base current of 100mA/50, or 2mA, to switch the 100mA load on and off reliably.

The BU406, which is capable of switching 10A at 200V and dissipating 60W, has the bias polarity requirements as the diminutive 2N551. Given a current gain

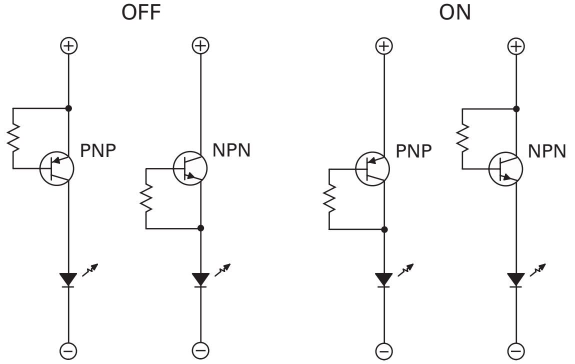

FIGURE 7-4 NPN and PNP transistor switch on and off conditions

of 30 and a load of 10A, we'd need an emitter-base current of 10/30, or 300mA, to reliably switch the 10A load on and off.

A PNP bipolar transistor can also be configured to switch a load in the emitter-collector circuit by manipulating the emitter-base current, as long as the proper bias polarities are observed. When the base is negative relative to the emitter, a small current flows from emitter to base and a significantly larger current flows from emitter to collector. As with the example of the NPN transistor, the transistor's current gain determines the emitter-base current required to switch the expected emitter-collector current reliably. With a current gain of 100 and a load of 20mA, the emitter-base current should be 20mA/100, or 0.20mA. Similarly, with a current gain of 50, the emitter-base current should be 20mA/50, or 0.40mA.

Compared to the diminutive 2N551 and 2N5401 transistors, the BU406 is a powerhouse capable of switching 10A at 200V and dissipating 60W.

Fuses

Both fuses in the humidifier are 2A fast-blow designs. Whereas many appliances feature slow-blow fuses, the fast-blow variety was selected apparently because of the need to kill the power quickly in the event of water entering the electronics area. Even so, the fuses seem overly conservative. At 32W, the humidifier should draw a little more than a quarter amp:

$$I = P/V = 32W/120V = 270mA$$

A 1A fast-blow fuse is probably a better option. One reason for using a 2A fuse would be to allow for power-on surges. However, in my tests, I didn't detect any surges that would blow a 1A fuse.

Note You can usually differentiate fast- and slow-blow fuses by the shape and size of the fuse element. Fast-blow fuses are constructed with a single thin wire from end cap to end cap. Slow-blow fuses, in contrast, are usually made of a flat, wide conductor that narrows slightly toward the middle of the fuse body. When in doubt, read the end caps.

Power Transformer

The step-down power transformer has a laminated steel core with a 120V primary and two secondaries, one rated at 28V at 80mA and one at 13.5V at 90mA.
An individual lamina from the core is shown in Figure 7-5a. The steel core is responsible for efficient coupling of magnetic flux induced by the primary winding to the secondary windings. This particular core design, called an *EI core* because the core is formed by a stack of *E*-shaped steel plates, such as the plate shown in Figure 7-5b, is mated with *I*-shaped steel plates. This common design is inefficient but inexpensive to manufacture.

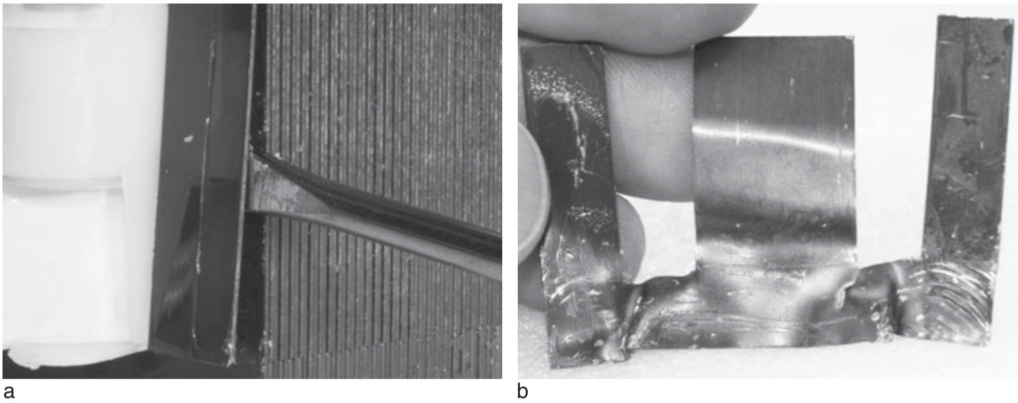

FIGURE 7-5 Iron laminate transformer E plate

Brushless DC Motor and Fan

Let's explore the 12V/80mA DC brushless motor (BLDC) using Figure 7-6 as a guide. Remove the squirrel cage fan with your hands, if you haven't done so yet, and then, with a pair of pliers, twist the shaft counterclockwise, holding the cap steady. The shaft should release from the cap and pass through the center of the coil assembly without resistance.

Figure 7-6a shows the cap containing a DC permanent magnet and the coil assembly. Leads to the 13.5VDC supply are attached to the bottom of the coil assembly. A side view of the coil assembly, shown in Figure 7-6b, reveals the coils of enameled wire and a 4-pin IC flanked by two 3.3µf at 16VDC electrolytic capacitors.

Notice that there are no brushes in the motor. In operation, the coil assembly or stator remains stationary while the DC magnet, which is glued to the inside of the plastic cap, rotates. The only point of contact is the shaft, which isn't strictly necessary as long as the cap can rotate freely around the coil assembly.

Let's take a look at the magnet. It measures about 1 inch (25mm) inside diameter, 1/2 inch (1.6mm) thick, and 3/8 inch (10mm) deep. What isn't clear from visual inspection is the number and orientation of magnetic poles—without instruments. Figure 7-6c shows the four poles of the circular magnet, assessed with the aid of a magnet polarity tester (available from Stewart MacDonald at www.stewmac.com) and magnetic field viewing paper (available from Scientifics at http://scientificsonline.com). Three of the circular magnet's four poles are visible as pie slices in the viewing paper. The polarity tester shows poles alternating between north and south every 90 degrees around the circumference of the magnet. That is, the two north and south poles are on opposite sides of the magnetic band.

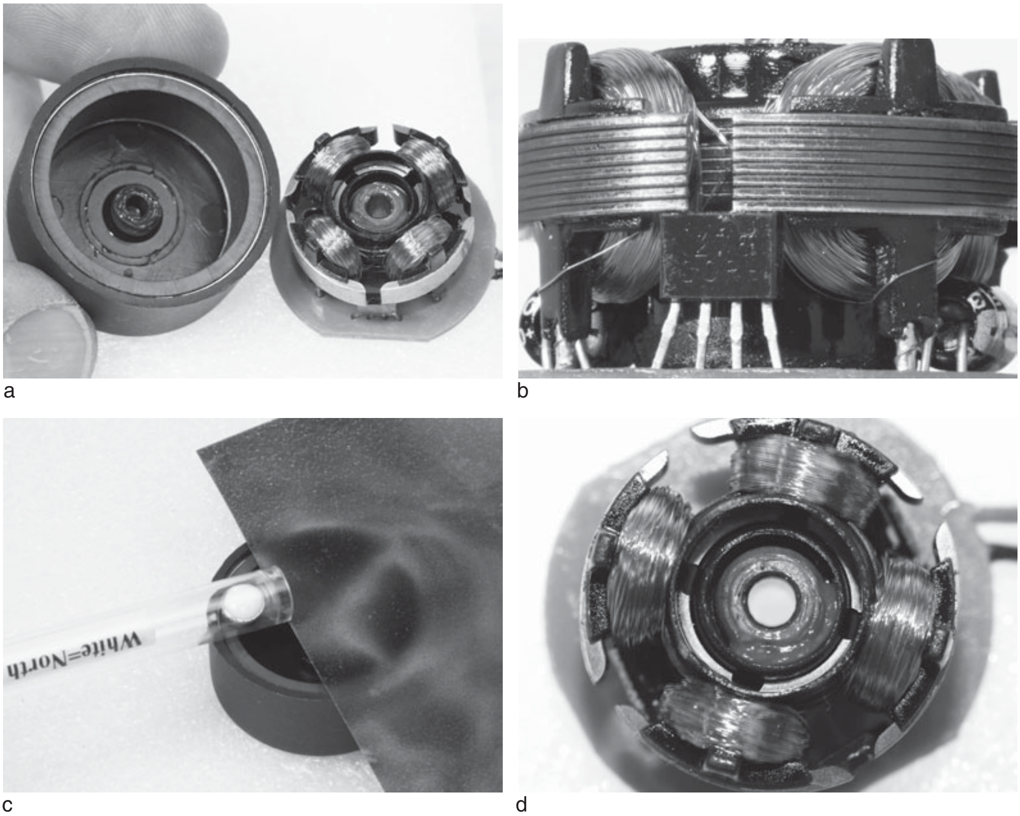

FIGURE 7-6 Brushless DC motor construction

If you look closely at Figure 7-6d, you can see that the silver wings on each of the four ferromagnetic cores is slightly tapered on the counterclockwise-most end. At power-down, the resulting air gaps between the rotating magnet and the stationary coils causes the magnetic rotor to stop so that the north–south pole boundaries are aligned with the spaces between ferro cores. That is, the magnet will stop so that the poles line up over the centers of the coils. As discussed in the next section, this simplifies the startup procedure. The magnetic cap spins counterclockwise relative to the coils shown in the figure.

How It Works

At the heart of a home ultrasonic humidifier is an ultrasonic oscillator that drives a piezoelectric transducer. When excited by high-energy ultrasound from the transducer, small globules of water and organic and inorganic matter suspended in the water are launched into the air. A fan creates an air current that transports this water vapor out into the room or other environment. When the water evaporates, these impurities are distributed around the humidifier in the form of an annoying white dust.

With this background, let's explore the inner workings of our ultrasonic humidifier, starting with the power supply. As shown in Figure 7-7, the power supply is a typical linear design with full-wave bridge rectification in both secondaries. The 13.5VDC supply, which supplies the fan motor, has a 100µf filter capacitor. The 28V supply is unfiltered. There is a 2A fuse in the primary and 28VDC secondary. Of note is that the fuse in the secondary is in the ground lead.

As illustrated in the simplified schematic in Figure 7-8, the power supply provides 13VDC to the fan circuit, composed of Q1 and the brushless DC (BLDC) motor. The power supply also provides 28VDC to Q2, which switches the bicolor

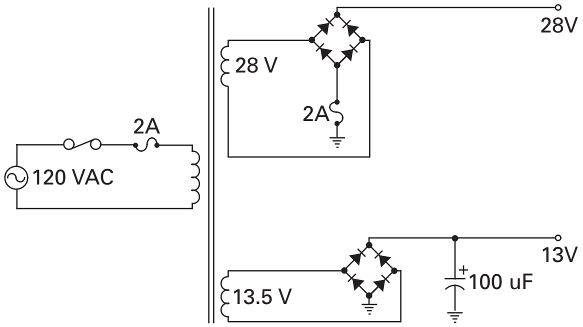
28V
28 V
2A
2A
120 VAC
13V
100 uF
13.5 V

FIGURE 7-7 Schematic of power supply

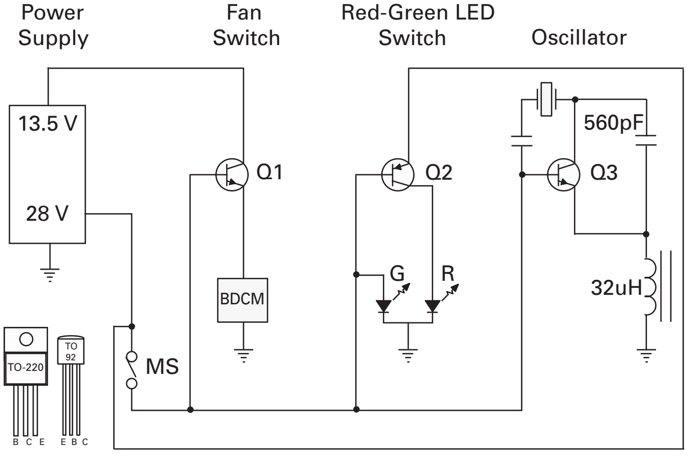
Power Supply
Fan Switch
Red-Green LED Switch
Oscillator
13.5 V
28 V
Q1
Q2
Q3
560pF
BDCM
G
R
32uH
MS
TO-220
TO 92
B C E
E B C

FIGURE 7-8 Simplified schematic of the humidifier

LED, and to Q3, the ultrasonic oscillator. When the magnetic switch (MS) is closed, indicating there is water in the base of the humidifier, the bases of all three transistors are biased positively, turning on Q1 and Q3 and switching off Q2. As a result, the fan and oscillator becomes active, the green LED is illuminated, and the red LED is extinguished. Note that, for clarity, the various bias resistors are not included.

The resonant frequency of the oscillator is established by the series-resonant circuit of the 560pF capacitor and the 32µH inductor. We can calculate the theoretical resonant frequency of the series LC circuit with the following formula:

$$F = 1/(2\pi\ (LC)^{1/2})$$

$$F = 1/(2\pi\ (32 \times 10^{-6} \times 560 \times 10^{-12})^{1/2}) = 1.2\text{MHz}$$

This is in agreement with the actual operating frequency of the oscillator.

Let's look at the ultrasonic oscillator in more detail, using Figure 7-9 as a guide. As discussed earlier, Q3, the BU406 transistor, is forward-biased when the magnetic switch (MS) is closed. The 5KΩ potentiometer establishes the amount of forward bias and therefore the output of the oscillator. The 100µH choke (L1) decouples the RF generated in the oscillator from the 28VDC supply line. At the operating frequency of about 1.2MHz, the choke would present the following impedance:

$$X_L = 2\pi FL$$

$$X_L = 6.2 \times 1.2 \times 10^6 \times 100 \times 10^{-6} = 745\Omega$$

Although the BU406 is a robust transistor, it isn't immune to damage. As noted earlier, I measured 120VPP across the transducer, which is fairly close to the

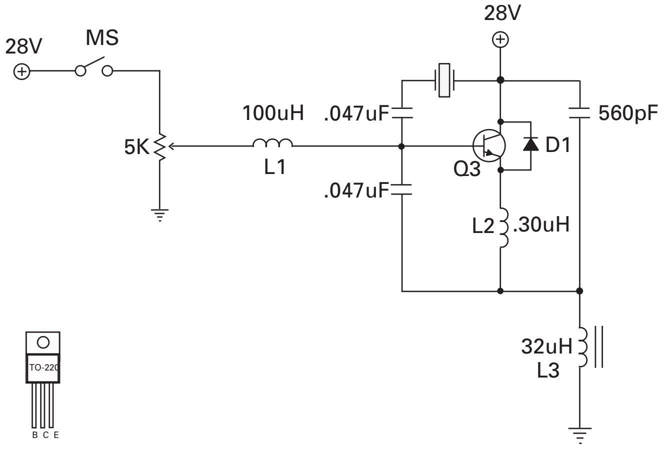

FIGURE 7-9 Simplified schematic of ultrasonic oscillator

transistor's maximum rating of 200V. A diode from the collector-emitter circuit (D1) protects the transistor from reverse overvoltage spikes generated by the crystal transducer. If the collector becomes more negative than the emitter, the diode conducts, shunting the current away from the transistor. The diode has no effect in the amplitude or frequency of the oscillator output.

The role of the 0.30μH open-air inductor (L2) may also be associated with maintaining the integrity of Q1, including inhibiting parasitic oscillations. I am certain it has no bearing on normal operation—the output amplitude and frequency were unaffected when I replaced L2 with a short length of wire.

Now let's turn our attention to the BLDC motor. As noted earlier, a brushless motor is a natural fit for our humidifier because there are no brushes, commutators, slip rings, or other electromechanical components to corrode and malfunction in a humid environment. In addition, there's no need for periodic maintenance, and brushless motors are quieter than their brush counterparts. If you're unfamiliar with the characteristics of a brush motor, think power drill—loud and electromagnetically dirty, but powerful.

Our motor has stationary coils and an outer, rotating plastic cap with a magnetic band glued inside. The motor is fed with 13VDC, with no other electrical input or output. So how does the motor turn? Since the polarity and magnitude of the magnetic field produced by the permanent magnet in our BLDC motor is fixed, the electrical activity necessary for rotation must come from the coils and drive circuitry.

The four-lead controller chip soldered to the BLDC motor circuit board has several potential functions. First, it certainly generates and feeds the coils with an AC signal of the appropriate frequency and amplitude. Second, it may sense the position of the rotating magnet. I couldn't locate the datasheet for the controller chip, so I can't say for certain, but the construction of the motor suggests this may be a sensorless motor design.

The controller chip may detect when magnetic polarity changes as the north–south pole passes immediately in front of it. There are several ways to detect this change in magnetic polarity. One way is to monitor the back electromotive force (EMF) produced by the spinning magnet and stationary coils, just as though the motor were a generator. Another popular approach is to use a Hall effect sensor, which senses magnetic flux density and polarity. It's also possible to go sensorless.

My first impression was that our BLDC motor was sensorless, and that the controller chip functioned solely as an AC generator. To determine whether a Hall effect sensor was buried in the controller chip, I removed the permanent magnet cap and then, with power applied, moved a small magnet across the face of the chip. The magnet triggered a series of trapezoidal AC pulses from the chip—exactly what I'd expect from a chip with a Hall sensor.

However, another factor that suggests sensorless operation is the tapering of the four ferro cores, shown in Figure 7-6d. One of the challenges of working with a BLDC motor in a sensorless configuration is that there is no information on the position of the rotor at startup. Without this information, it's difficult to ensure that the AC signal sent to the coils will result in rotation in the proper direction. One solution to this dilemma is to shape the ferro cores so that they bias the system to rotate in one direction over another. My conclusion is that the BLDC motor uses a combination of a Hall sensor and shaped ferro cores—probably in a redundant configuration, so that one system can take over if the other fails.

Mods

This humidifier is a great source for components that can be repurposed for other projects. I can think of numerous applications for the intact ultrasound module, repackaged so that the transducer is mounted in a probe or in a small container. You could create an ultrasonic cleaner, a mixer for small quantities of fluid, and maybe even a meat tenderizer. Just remember to keep your fingers away from the transducer and don't allow the module to overheat.

Because of safety issues, there isn't much in the way of mods if you plan to use this humidifier unattended in your home on a constant basis. However, this humidifier is a perfect experimental platform for working with closed-system concepts for a science fair project or simply to get a better understanding of the practical application of feedback and control theory.

Closed Loop System

Like most inexpensive humidifiers, our humidifier is an open loop system. You set the rate of vapor production and periodically monitor the status LED for an indication of when it's time to refill the unit. If you happen to forget to fill the reservoir, the unit will shut off automatically. The limitation of this open loop design is that the humidifier output doesn't reflect environmental conditions. Ideally, it would produce more water vapor in dry weather and less in damp, humid weather.

You could form a human-in-the-loop closed loop system by sitting next to the humidifier with a hygrometer in one hand and the humidifier output control in the other. When the humidity drops below your established optimum value, you crank up the output. Conversely, when the humidity exceeds your preset value, you turn down the output. You could automate this closed loop control process by using a humidity sensor, a means of setting the desired humidity, and a difference detector that's linked to the output controller, as shown in Figure 7-10. Keep this in mind when you read the next chapter teardown of a hygro thermometer.

Once you've established a means of detecting when the relative humidity is above or below your set point, you can use a microprocessor to control a relay in series with the magnetic reed switch to activate and deactivate the humidifier at the set points. You'll find that if the high and low humidity control points are too close together, the humidifier will constantly cycle on and off because of the tolerance of the sensor and other components.

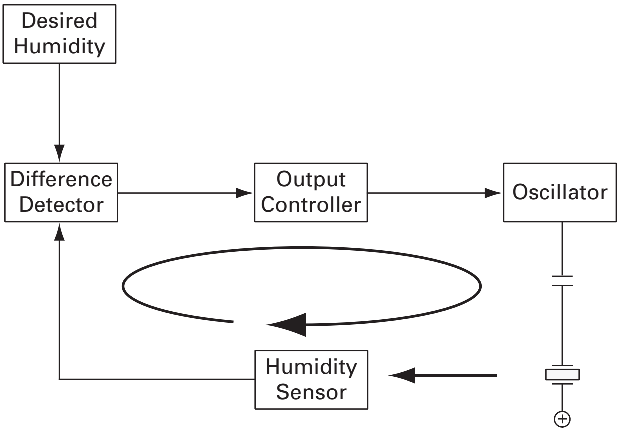

FIGURE 7-10 Closed loop system diagram

The closed loop system would also be affected by temperature and the size and nature of the room. For example, if you put the humidifier in a closet, the humidity will tend to overshoot the set point. If you put the humidifier in a large room, it may take hours for the humidity to change significantly. Open the windows or turn on an air conditioner, and the humidifier may never cycle off or on, depending on set points.

The programming challenge is to control the closed loop so that the unit isn't switching on and off five times a second in response to small deviations in the difference detector, or so that it isn't cycling every 15 minutes, well after the humidity has swung significantly greater or less than the set point. As you might expect, developing a computer algorithm that can maintain an exact relative humidity is more difficult than it sounds.

If you're interested in tackling this challenge, you can find a body of literature and code samples that address closed loop control, including the popular PID (Proportional-Integrative-Derivative) model. The robotics community has also developed practical hardware and software solutions to closed loop control systems that are worth investigating.

This page intentionally left blank

Chapter 8

Digital Hygro Thermometer

If you read the preceding chapter describing the teardown of an ultrasonic humidifier, you know the advantage of having a digital instrument that accurately measures the relative humidity as part of a closed loop control system. Add to this capability a digital thermometer and the instrument becomes an indispensable comfort gauge.

I own several inexpensive digital hygro thermometers and find them indispensable. There's one in my guitar room to help me ensure that my wooden instruments don't turn to splinters or develop mildew, one in the bedroom to monitor the creature comfort levels, and a USB hygro thermometer in a server closet to ensure the environment is suitable for the computers, hard drive arrays, and communications devices. The USB unit is wired to my workstation so that I can monitor the environment remotely.

In this chapter, I'll tear down an Extech Hygro-Thermometer, shown in Figure 8-1. This digital meter is sold under a variety of brands for about $40. I'll discuss some ideas for repurposing the components and a mod at the end of the chapter.

Highlights

The most obvious component of our digital hygro thermometer is the large LCD display. The actual measurement and display hardware, tucked away on a circuit board in the lower fifth of the unit, includes a microcontroller, a crystal oscillator, an analog humidity sensor, and an analog temperature sensor. Unfortunately, the mystery microcontroller is mounted directly on the circuit board and buried under a glob of black epoxy.

During the teardown, note the following:

- The use of SMT components
- The ribbon cable connecting the LCD to the circuit board

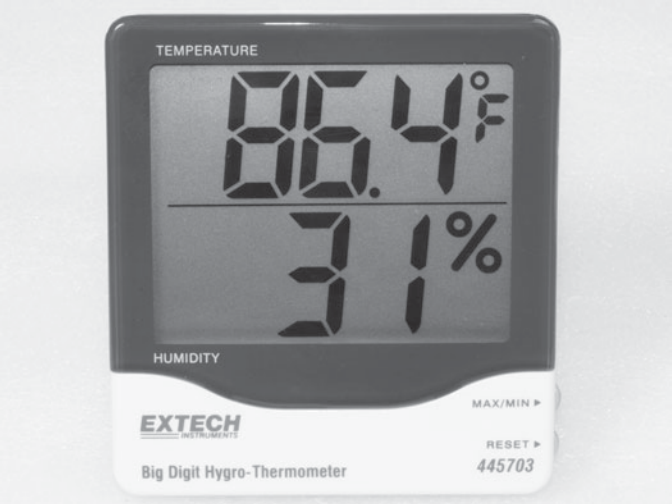

FIGURE 8-1 Digital hygro thermometer

- The crystal oscillator
- The analog temperature and humidity sensors

Specifications

Following are the key specifications of the instrument:

- 4 × 4 × 3/4 inches (HWD)
- Digital display of temperature in degrees Fahrenheit or centigrade (°F or °C) and relative humidity
- Memory function to store and display minimum and maximum temperature and humidity values
- Temperature accuracy is ±1.8°F from 14°F to 122°F (±1°C from –10°C to 50°C)
- Relative humidity accuracy is ±5 percent RH from 25 percent RH to 85 percent RH within the temperature range of 32°F to 122°F (0–50°C).
- Powered by a single 1.5V AAA cell with a six-month life expectancy

These accuracy figures are typical for low-cost digital hygro thermometers. Meteorological digital and analog hygro thermometers capable of measuring relative humidity within ±2 percent of the actual value are available starting at about $200. Of course, laboratory grade instruments with greater accuracy are available.

Operation

Only two buttons and a switch are required to operate this otherwise hands-off meter. A F°/C° switch on the back of the meter determines whether temperature is displayed in Fahrenheit or centigrade, respectively. In the lower-right corner of the front of the meter are the Max/Min and Reset buttons. The Max/Min button recalls the maximum and minimum temperature and relative humidity values since the Reset was pressed. The Reset button erases temperature and humidity values stored in memory.

Teardown

You should be able to complete the basic teardown, illustrated in Figure 8-2, in about 3 minutes. Add another few minutes if you unsolder the sensors.

Tools and Instruments

You'll need a small Phillips-head screwdriver to crack the case and a soldering iron if you want to remove and experiment with the sensors. This is a nice meter for a desk or wall, and you needn't destroy it during the teardown.

Step by Step

Although it's not strictly necessary, you can verify that the AAA cell holder is empty before you begin. You can easily replace the battery once the circuit board is exposed.

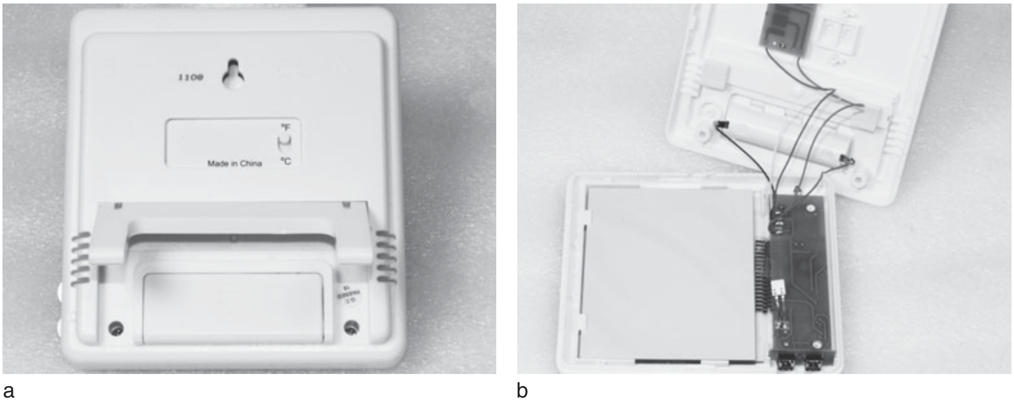

FIGURE 8-2 Teardown sequence

Step 1

Disengage the back. With the meter face down, lift up the plastic easel stand to expose the two Phillips-head screws, as shown in Figure 8-2a. Remove the screws. With your fingernails or a coin, push in the four tabs around the edge of the meter—two on top and two on either side. Gently open the back to reveal the LCD and circuit board, shown in Figure 8-2b. Avoid tugging on the wires connecting the circuit board to the AAA cell holder or the F°/C° switch on the back.

Step 2

Release the circuit board. With a miniature Phillips screwdriver, remove the two retaining screws from either end of the circuit board. Carefully flip the board and LCD to view the front, component side of the board and LCD.

Layout

As shown in Figure 8-2b, the back of the circuit board is sparsely populated. Other than the connections to the F°/C° switch and the AAA cell, the only significant component on the board is the white, rectangular humidity sensor.

The front of the board, shown in Figure 8-3, holds an epoxy dipped thermistor, a cylindrical crystal oscillator, a few SMT resistors and capacitors, and the two momentary contact switches for max/min and reset functions. The microcontroller, hidden under the black epoxy blob, and the adhesive ribbon cable taped to the LCD are also on the front of the circuit board.

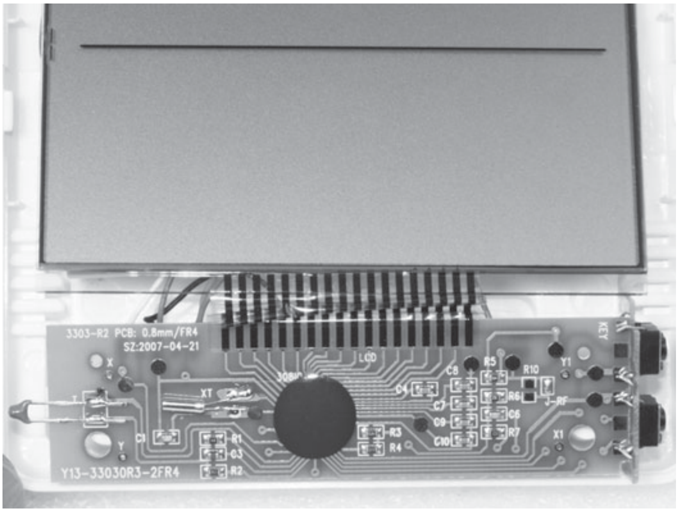

FIGURE 8-3 Front circuit board layout

Components

The microcontroller may be the most advanced component on the circuit board, but the most relevant for this teardown are the temperature and humidity sensors.

Thermistor

The epoxy-dipped thermistor, or temperature-dependent resistor, shown in Figure 8-4, is mounted so that its body is suspended near the air ducts on the plastic case and away from the heat-producing components on the circuit board. This is a negative temperature coefficient (NTC) thermistor, meaning that as temperature increases, the resistance decreases. For example, I measured the resistance of the thermistor in my meter as 8.6KΩ at 80°F and 10KΩ at 60°F.

The ideal thermistor has constant relative changes in resistance for given changes in temperature, regardless of the absolute temperature. In other words, if you plot temperature versus resistance, the relationship is a straight line. In practice, inexpensive thermistors exhibit a nonlinear relationship between temperature and resistance. This nonlinearity can be corrected by placing a resistor with a complementary temperature coefficient in parallel with the thermistor. A second approach is to program a microcontroller either to use a lookup table or execute a corrective formula to linearize the response.

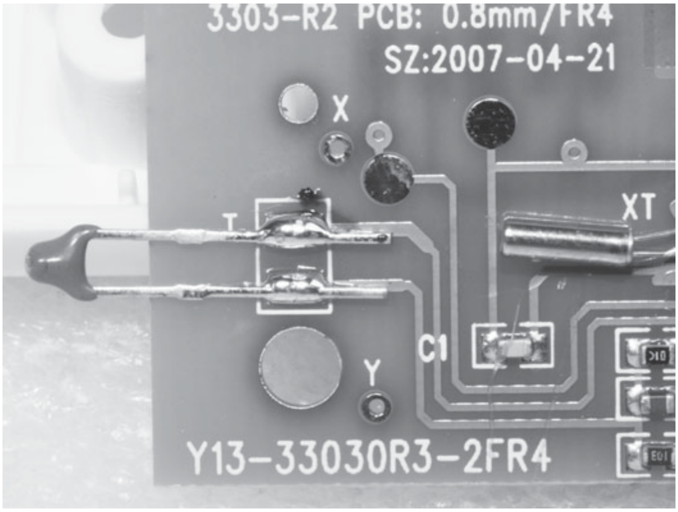

FIGURE 8-4 Thermistor (T) and crystal oscillator (XT)

Humidity Sensor

The relative humidity (RH) sensor used in this meter is a resistive sensor, shown in Figure 8-5, which consists of a moisture-absorbing substrate and conductive traces. With increased moisture in the air, more moisture enters the substrate and the resistance drops. In addition, the moisture in the substrate alters the capacitance between the conductive traces.

The sensor's *impedance*—resistance of the substrate and capacitive reactance of the sensor—is inversely proportional to relative humidity. Because the relationship between impedance and relative humidity is roughly exponential, the raw reading from the sensor must be linearized with a lookup table or formula. The impedance of the RH sensor in my meter measured 400KΩ at 36 percent RH and 40KΩ at 98 percent RH.

The more expensive and more accurate digital hygrometers use capacitive relative humidity sensors that consist of an exposed dielectric (insulating material) between two conductors. Increases in relative humidity increase the dielectric constant of the insulator. Unlike resistive sensors, this response is linear over a wide range of temperatures and relative humidity conditions.

A third type of relative humidity sensor used with digital hygrometers measures the thermal conductivity of the air relative to a sample of dry air. In general, thermal conductivity sensors are used in high-temperature applications where capacitive and resistive sensors would not survive.

A fourth type of relative humidity sensor consists of a quartz crystal with a hygroscopic coating. An increase in relative humidity increases the dimensions of the quartz crystal and causes a downward shift in the resonant frequency. Definitely overkill for a home comfort gauge.

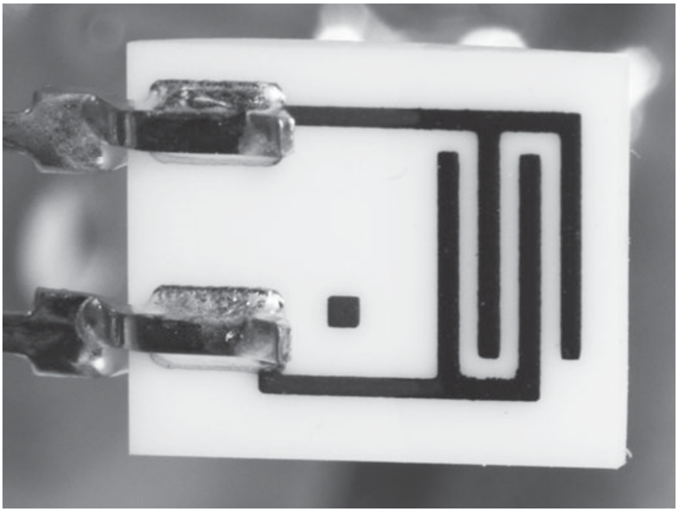

FIGURE 8-5 Humidity sensor

Microcontroller

The microcontroller hidden under the glob of black epoxy handles the LCD display and the momentary switches, and reads signals from the sensors. (See the discussion of how to handle these epoxy blobs in Chapter 3.) For our purposes, what's important is that the microcontroller reads the two sensors, processes the data, and displays the results.

Crystal Oscillator

The cylindrical tuning fork crystal oscillator, which operates at 32.768kHz, can be seen on the circuit board next to the thermistor in Figure 8-4. This is a popular, inexpensive, low-power clock source for microcontrollers and microprocessors, in part because the operating frequency is stable over a relatively wide temperature range.

Many microcontrollers use an onboard crystal oscillator or an external resistance-capacitance (RC) controlled oscillator. However, in applications for which timing is critical, an external, high-accuracy crystal oscillator is often best. As you'll soon see, this is one of those applications.

LCD Panel

The LCD panel, the largest component in this meter, is notable in the edge connector used to mate it with the microcontroller and circuit board. As shown in Figure 8-3, the transparent ribbon cable connects pads on the circuit board to those on the edge of the LCD. The pressure connection is secured by a piece of clear tape over each end of the cable.

Note One of the limitations of LCD displays over competing display technologies, such as LED displays, is operating temperature range. For example, I run with an LCD watch. When the ambient temperature dips below freezing, the digits on the LCD fade to gray. I haven't verified it, but I assume that given the manufacturer lists the accuracy down to 14°F (–10°C), the LCD is capable of operating at that temperature.

Some manufacturers avoid the limitations imposed by the LCD operating environment by adding remote sensing capabilities to their hygro thermometers. This is frequently the design of weather monitors in which the display is kept in a warm room and the sensor is placed outside.

How It Works

The mechanical, analog technologies for measuring temperature and relative humidity—from bimetallic and mercury thermometers, to hygrometers made from strands of horse hair or a sheet of animal intestine—have stiff competition from electronic meters. While mercury thermometers are typically more accurate than similarly priced digital thermometers, there is the issue of environmental toxicity from the mercury. Instruments that rely on the elongation and shortening of horse hair or the tension on a piece of animal intestine stretched over a frame as a function of relative humidity require periodic calibration.

Digital thermometers with resistive sensors rely on the fundamental physical property that the electrical resistance of a conductor varies with temperature. An ordinary metal film resistor, such as a length of copper wire, displays changes in resistance as a function of temperature. A thermistor simply displays this property with a relatively small change in temperature. As noted earlier, the thermistor in this meter is designed to offer decreased resistance as the temperature increases.

You can use several methods to determine the resistance of a thermistor in a digital thermometer. One method is to apply a fixed voltage across the thermistor and measure the current. Another less hardware-intensive approach is to use the thermistor as the resistance component of an RC (resistor-capacitor) circuit and to measure the time required for an applied voltage to drop to a specific value. Recall that the time constant of an RC circuit is the time required for the voltage to drop to 37 percent of its initial value. Mathematically, this is expressed as follows:

$$T = RC$$

where T = time in seconds, R = resistance in ohms, and C = capacitance in farads.

Based on my measurements and assessment of the circuit, our digital thermometer uses the later approach. Every 5 seconds, a series of a dozen pulses, 4V peak-to-peak at 12.5kHz, is applied to the thermistor in series with SMT capacitor C1. The greater the temperature, the lower the resistance of the thermistor and the shorter the time constant. In other words, the greater the temperature, the faster the decay in voltage across C1.

Note the three I/O leads from the microcontroller that interface with the thermistor: R1, R2, and C3 in Figure 8-3. This is reflected in the simplified schematic, shown in Figure 8-6.

For simplicity, assume that the microcontroller reads an input pin as high if the voltage is at or above 0.63V and low for anything below 0.63V. With an accurate clock on board, the microprocessor times how long it takes for the voltage to drop so that the input pin transitions from high to low. This is repeated for each of the pulses, and the average value is used to determine the resistance of the thermistor. Given the resistance, the microcontroller uses a lookup table or a mathematical calculation to convert the thermistor resistance to a temperature value that is then displayed on the LCD.

Note that there are two connections to the C3-thermistor junction. Given that only one connection is required to monitor the voltage across C3, the second lead is probably for auto-calibration or a self-test.

Now let's focus on the hygrometer. You may have noticed that it's difficult to find a digital hygrometer without a thermometer. This is because an accurate calculation

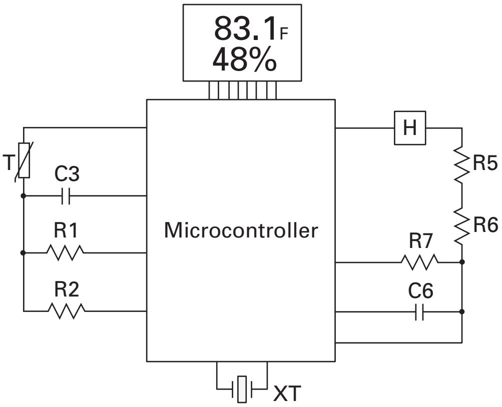

FIGURE 8-6 Simplified schematic

of relative humidity requires a value for the ambient temperature. Because the ambient temperature is used to compute relative humidity and the display circuitry is already in place, it makes sense to add value to the meter by displaying the temperature.

The hygrometer sensor wiring to the microcontroller is similar to that used for the thermistor, as shown in Figure 8-6. One terminal of the humidity sensor is connected directly to an I/O pin of the microcontroller. The other terminal is connected in series with two resistors, R5 and R6, and then to three I/O pins via C6, R7, and a direct connection to the microcontroller. As with the thermistor, every 5 seconds, a series of square waves, in this case about 3VPP (peak-to-peak) at 12.5kHz, is applied to the humidity sensor. The time constant is read by the microcontroller and, along with ambient temperature, is used to calculate and display relative humidity.

Mods

As discussed in Chapter 7, you know that one use for a digital hygro thermometer is to function as part of a closed loop system that controls a heating, cooling, and-or humidification system. The challenge is getting at and interpreting the output of the microcontroller.

Adding Calibration Capabilities

One of the limitations of this meter is that it offers no way of easily calibrating the temperature or relative humidity. You're either stuck with whatever is displayed on the LCD or you can apply correction factors to the values on the display. You can add some adjustability by inserting a subminiature potentiometer in series with the thermistor. If you need to lower the resistance presented by the thermistor, try the potentiometer in parallel with the thermistor. In either case, make certain you use a potentiometer with a low temperature coefficient.

You can try the same approach for the relative humidity sensor. Do your best to avoid handling the sensor, because oils from your fingers can permanently alter the resistance of the sensor.

An obvious issue in calibrating your meter is that you need a standard. Temperature is usually less of a challenge, because accurate analog thermometers are so readily available. Relative humidity is another issue. If you can, borrow a 2 percent hygrometer designed for meteorology to use as a standard. Another option is to bring your meter to a store with high-end hygrometers on display and compare settings after you've given your meter time to stabilize.

Chapter 9

Stereo Power Amplifier

Whether your passion is 7.1 channel surround sound for video games and home theatre or playing vinyl records in stereo, abundant, clean, affordable audio power is a must. And, thanks to specialized, high-performance audio amplifier ICs, audiophile-quality amplifiers are commodity items. All you need to build a state-of-the-art amplifier are a pair of amplifier ICs, a handful of discrete components, a robust DC power supply, and a heat sink large enough to keep the ICs from going into thermal shutdown. The usual amplifier design headaches, from impedance matching and short protection, to providing a standby mode, are handled by the ICs.

In this chapter, I'll tear down the AudioSource AMP 100, shown with its cover removed in Figure 9-1. This Chinese-manufactured stereo amplifier, which retails for about $140, is popular in the audio community because of its simplicity, specifications, and affordability. As a teardown, it's a great introduction to integrated amplifier chips, operational amplifiers, amplifier specifications, and concepts such as bridging. At the end of the chapter, I'll offer a mod that addresses a minor complaint that some have with the amplifier.

Highlights

At first, you might find it intimidating to tear down something as seemingly complex as a stereo amplifier. However, thanks to the simplicity afforded by the power amplifier ICs in the AMP 100, you won't be dealing with dozens of transistors and hundreds of support components. If you can handle a screwdriver, you can handle this teardown. As you'll see, most of the circuit analysis revolves around two circuit boards: the power board containing the power supply and power amplifier ICs and the smaller but more complex logic board.

During the teardown, note the following:

- The simplicity of the power board
- The size, mass, and orientation of the heat sink

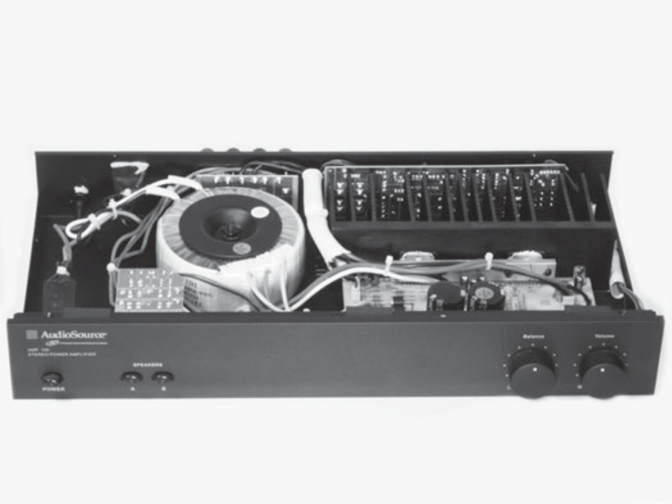

FIGURE 9-1 AudioSource AMP 100 stereo amplifier

- The mass of the toroidal power transformer
- The pair of zener diode voltage regulators

Specifications

The specifications of the AMP 100 reflect those of the two integrated amplifier ICs. Highlights of the specifications, extracted from the user manual and AudioSource web site (www.rodinaudio.com), include the following:

- Frequency response: 20Hz to 20kHz ± 1dB; 5Hz to 65kHz ± 3dB
- Signal-to-noise ratio (SNR): 101dB
- Maximum stereo power output: 60W at 8Ω, 1 percent total harmonic distortion plus noise (THD+N)
- Maximum bridged-mono output: 160W at 8Ω, 1 percent THD+N
- Power consumption: 500W
- A, B, A+B speaker selector switches
- Auto-on (signal sensing)
- Auto input switching
- Front panel controls: power, speakers, volume, and balance
- 3 × 16 1/2 × 9 1/4 inches (HWD)
- 9.6 pounds

Additional specifications, based on the datasheet for the power amplifier chips, include Class AB operation, auto-thermal shutdown, and short-circuit protection.

Together, the specifications detail a tough, compact amplifier. But what do the individual specifications mean? I'll consider the more significant specifications one at a time.

Frequency Response

An often-touted audio amplifier specification is the frequency response or bandwidth, typically at the 3 and 1dB points. The ideal frequency response for a home entertainment amplifier is *flat*—that is, ±0dB through audible audio spectrum. In the context of amplifier bandwidth, a dB, or decibel, is a logarithmic relationship between the power output measured at two frequencies within the bandwidth, defined by this equation:

$$\text{dB} = 10 \log_{10} (\text{P1}/\text{P2})$$

where P1 and P2 are power measures taken within the specified bandwidth. For example, let's say a 4kHz tone is fed to the amplifier input and the output power is measured as 50W. A tone of the same amplitude but frequency of 20Hz is fed to the amplifier input and the output is measured as 45W. Is the amplifier working within spec? Let's apply the formula:

$$\text{dB} = 10 \log_{10} (45/50)$$

$$\text{dB} = -0.45$$

The output is within the ±1dB limit, and therefore in spec. In other words, according to the manufacturer, the power output doesn't fluctuate more than 1dB above or below the nominal power output over the frequency range of 20Hz to 20kHz.

The 1dB point is relevant because it's roughly equal to the just noticeable difference. That is, while most people can't distinguish between 45 and 50W (0.45dB), most can differentiate between 40 and 50W (1dB) and certainly between 50 and 25W (3dB).

The 3dB point is relevant because it represents the half-power point, or where the amplifier output drops to half of the nominal output. In the case of the AMP 100, power drops to half at 5 and 65Hz. For the price of this amplifier, both the 1 and 3dB bandwidth figures are excellent.

SNR

The signal-to-noise ratio is a measure of the amplitude of the music or other desired signal to the amplitude of the background noise. SNR is also measured in dB:

$$\text{SNR (dB)} = 10 \log_{10} (\text{PSignal}/\text{PNoise})$$

where PSignal is the power of the desired signal and PNoise is the amplitude of the noise. From a practical perspective, noise is what you hear when the volume is turned up but no external signal is applied to the amplifier, or when there's a pause between songs. Ideally, SNR is infinite. However, noise is inescapable above absolute zero. For this reason, passing a signal through a resistor introduces noise into the signal and diminishes the SNR.

Considering the price of the AMP 100, an SNR of 101dB is good, if you make a number of assumptions. For example, I picked up an old Marantz integrated amplifier on eBay that was originally an order of magnitude more expensive than the AMP 100. It has an SNR of 90dB, and I'm certain it outperforms the AMP 100. As with power, you can't simply compare two SNR figures without specifying the bandwidth and frequencies at which the signal and noise are measured.

According to the AudioSource web site, the SNR of the AMP 100 is 101dBA, referred to 50W, at 8Ω. The dBA or A-weighted dB is commonly used in measuring the noise of audio equipment, but the reference to 50W at 8Ω is incomplete. A more informative description would include assurance of CEA-490-A R-2008 compliance. This standard of the Consumer Electronics Association (www.ce.org), which defines test conditions and test measurement procedures for determining various performance characteristics of home audio equipment, enables consumers to compare audio amplifiers on an equal basis. A standards organization sometimes cited by manufacturers as a reference for SNR is the Institute of High Fidelity (IHF), now part of the Electronic Industries Alliance (www.eia.org). Unfortunately, most audio amplifier manufacturers simply offer an SNR in dB, with no further elaboration.

THD+N

Total harmonic distortion plus noise is a way of quantifying audio "crud." As you crank up the volume of this or any other home entertainment audio amp, the sound becomes increasingly distorted, noisy, and rich in harmonics. In the context of the AMP 100, harmonics are undesirable. However, this doesn't hold for special-purpose amps, such as guitar amps, which are designed to distort sounds. (See Chapter 14 for more details.)

You can probably get an intuitive handle on THD+N by studying the following formula:

$$\text{THD} + \text{N(bw)} = (\text{sum of harmonic powers} + \text{noise power})/\text{output power}$$

The first thing to note is that THD+N makes sense only when bandwidth (bw) is specified, typically the 1 and 3dB bandwidths. The sum of harmonic powers is the sum of the powers of the harmonics: for example, the power of the second harmonic plus the power of the third harmonic, and so on. Noise power is the power of crud produced by the audio system without an input signal, due to, for example, random movement of electrons within conductors, interference from other devices, and component imperfections. Output power is the power output of the fundamental frequency when the THD measurement is taken.

Common fundamental frequencies for evaluating THD+N are 1 and 4kHz. Increasing the output power typically increases the sum of harmonic powers. For example, if the power is increased at 4kHz, you'd expect the powers at 8, 12, 16, 20kHz, and so on, to increase as well.

THD+N is a ratio measurement expressed as a percentage. Ideally, THD+N is somewhere near 0 percent, but 1 percent with the volume cranked up isn't terrible. One way to reduce THD+N is to reduce output power to well below maximum output power. If a 100W amplifier is loafing along at, say, 50W, then the harmonic power will be down because of the increased dynamic range or headroom available. Background noise isn't as responsive to decreased output power as THD.

Power Output

As the specifications suggest, simply stating that an amplifier has 100W output is incomplete and misleading. The best way to interpret the power output rating of the AMP 100 is that, given an 8Ω load per channel and the limitation of 1 percent THD+N, it will deliver up to 60W per channel, or 120W total. The specification doesn't say 60W is the maximum output power—you may be able to get 75W from the unit, but under different assumptions. For example, if you assume 4Ω speakers and a limitation of 1 percent THD+N, the maximum output power is 75W. Similarly, if you assume 8Ω speakers and up to 10 percent THD+N, you might be able to extract up to 80W from the amplifier. The point is, when you are evaluating amplifier output power and other specifications, make certain you understand the operating assumptions.

A feature of this amplifier that makes it appealing to home theater enthusiasts is that the output is *bridgeable*, meaning both channels of the amplifier can be used to drive a single speaker. Let's say you want to use the AMP 100 as the center (mono) channel of your home theater setup. As with any other stereo amplifier, you could feed the same signal to both left and right input channels of the amplifier, and the left and right speakers would reproduce the same in-phase output.

However, because this is a bridgeable amplifier, you have the option of connecting a single speaker between the left and right output terminals. In the bridge mode, the AMP 100 is configured so that the left and right channels are 180 degrees out of phase. As a result, if each channel swings from 0 to 20V relative to ground, the bridged output will swing 0 to 40V relative to each output terminal. For example, at the instant when the right channel terminal is +20V relative to ground, the left channel terminal is –20V relative to ground, a 40V differential. Otherwise, if the output signals were in phase, which is the case in a regular stereo setup, there would be no voltage difference between the output terminals.

Bridging provides greater output power than simply sending the same signal to each channel of a stereo amplifier. For example, the AMP 100 produces an additional 40W as a bridged-mono output amplifier compared with the stereo out, assuming the same speaker impedance and THD+N. To understand how this is possible, recall the power equation

$$P = V^2/R$$

where P is power in watts, V is the voltage in volts, and R is the resistance in ohms. Let's assume that each channel of a stereo amplifier produces a maximum of 20V across an 8Ω speaker. That's $20^2/8$, or 50W, per channel for a total of 100W. Now consider the bridged output with the same output and speaker impedance. The voltage maximum is now 40V, producing $40^2/8$, or 200W. Although the AMP 100 doesn't provide the theoretical 100W of additional power in bridge configuration, the additional 40W is an increase of just more than 1dB, a noticeable increase in output.

Power Consumption

The AMP 100 is rated at 500W power consumption, and much of this power is destined for the heat sink, as opposed to the speakers. The power consumption, which reflects the efficiency of the amplifier, is largely due to the power requirements of the power ICs, which operate as Class AB amplifiers. Recall that an amplifier configured for Class A operation conducts through the entire input cycle. Using an oscilloscope, you'd see sine wave input and sine wave output. A transistor, IC, or tube amplifier configured for Class B operation, in contrast, conducts for half of the input cycle. On your scope, you'd see sine wave in either the positive or negative half of the sine wave out, but not both.

A pair of Class B amplifiers, configured in a push-pull arrangement, with one assigned to the positive and one to the negative component of the input signal, can faithfully amplify the full sine wave. The advantage of Class B over Class A operation is increased efficiency. The price is increased THD, component count, and, in general, complexity.

Class AB operation is akin to Class B operation with a little Class A operation thrown in to decrease the THD. The Class AB amplifier ICs conduct for a little more than half of each input cycle. The result is significantly improved linearity and reduced THD compared with a Class B configuration, because the crossover from one amplifier to the other is smoother. The price for reduced THD is decreased efficiency.

Auto-On

A feature of this amplifier that you'll either love or hate is the auto-on feature, which is essentially active standby. With the power button depressed, the amplifier is on standby until it detects an audio signal on one of its input connectors. Auto-on is trivial to implement because of a pin dedicated to this function on each power IC.

The advantage of auto-on is that it's one less button to press if you're booting up your home entertainment system, assuming you don't have a power strip with an on–off switch. One disadvantage, in this era of carbon credits and green thinking, is that this is just one more device to add to the dozens of devices consuming energy in your home. Another disadvantage is that you're forced to use this feature.

If the AMP 100 is on and there's a long pause in a music passage, the unit responds by shutting down until triggered by another signal. In this condition, you'll hear the relay click open and closed and the audio will stutter. I haven't experienced this problem, perhaps because I don't watch movies with lots of silence in the soundtracks, but it's a hot topic on the forums.

Auto-Input

Auto-input is like auto-on, in that the AMP 100 automatically senses an audio input signal and then configures itself accordingly. In this case, however, the configuration has to do with selecting the signal source. The front panel is void of a source selector switch. As a result, you have plug in the output cables from two audio devices when you initially configure the system. The amplifier will automatically switch to the active input, line 1 or line 2, when there is signal present. If a signal is present on both input channels, line 1 takes priority.

Form Factor

The specifications relating to physical size and mass are straightforward. Of note is that much of the bulk and mass of the AMP 100 are due to the power transformer and heat sink.

Operation

This amp is plug-and-play. The only significant operating decisions are at setup, when you use the two slider switches in the rear of the unit to select between either standard stereo output or bridged mono output and between normal on–off and auto-detect. The balance and volume controls function as you'd expect.

Teardown

You should be able to complete the teardown, illustrated in Figure 9-2, in about an hour.

Tools and Instruments

You'll need a few Phillips-head screwdrivers to disassemble the housing and extract the circuit boards, and a 10mm nut driver to remove the two potentiometers from the front panel. A multimeter will be useful for tracing the circuits. If you want to trace signals through the circuit boards, a dual-trace oscilloscope is a real time-saver. You might also find a muffin tin useful to store and sort the various screws produced by the teardown.

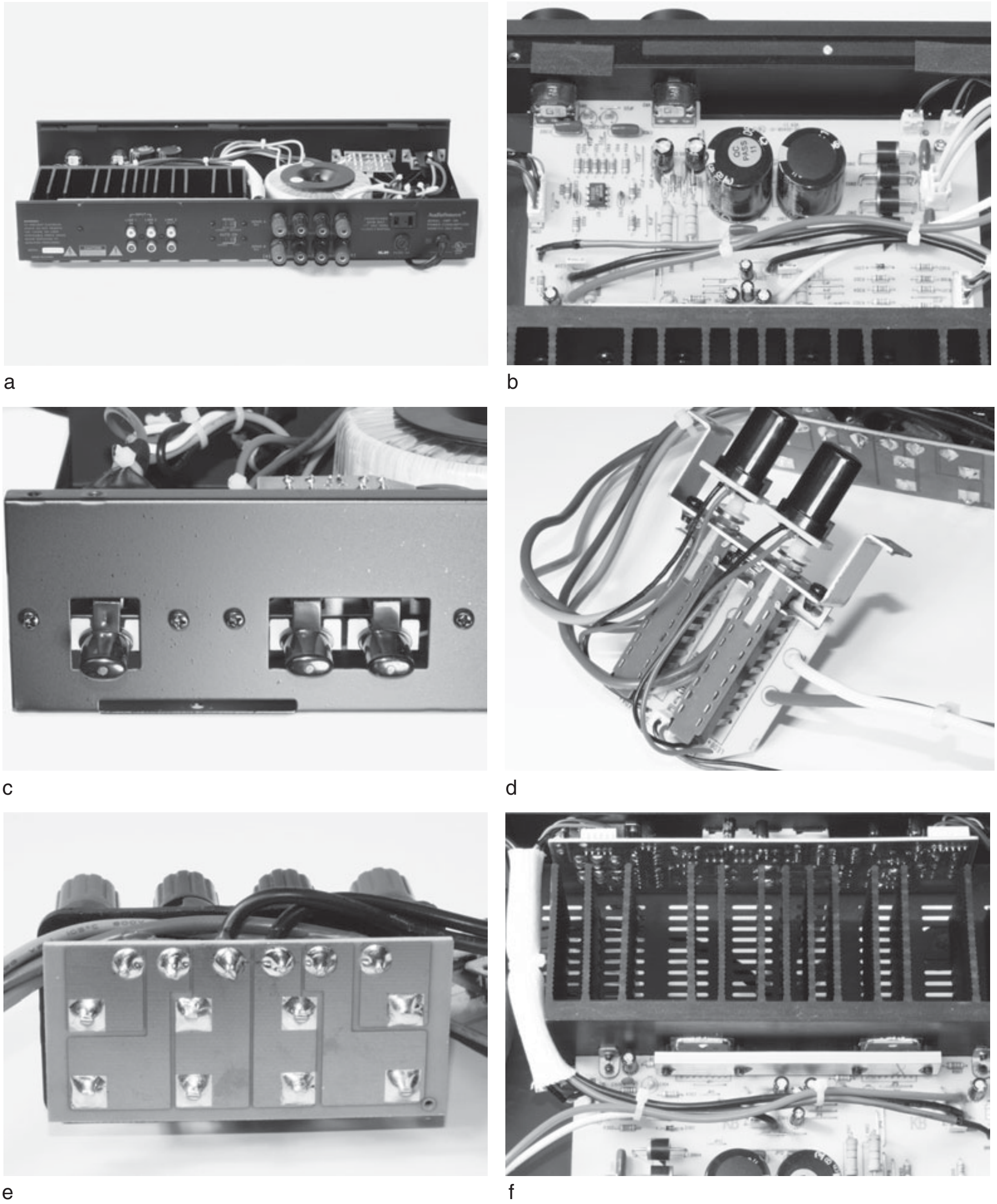

FIGURE 9-2 Teardown sequence

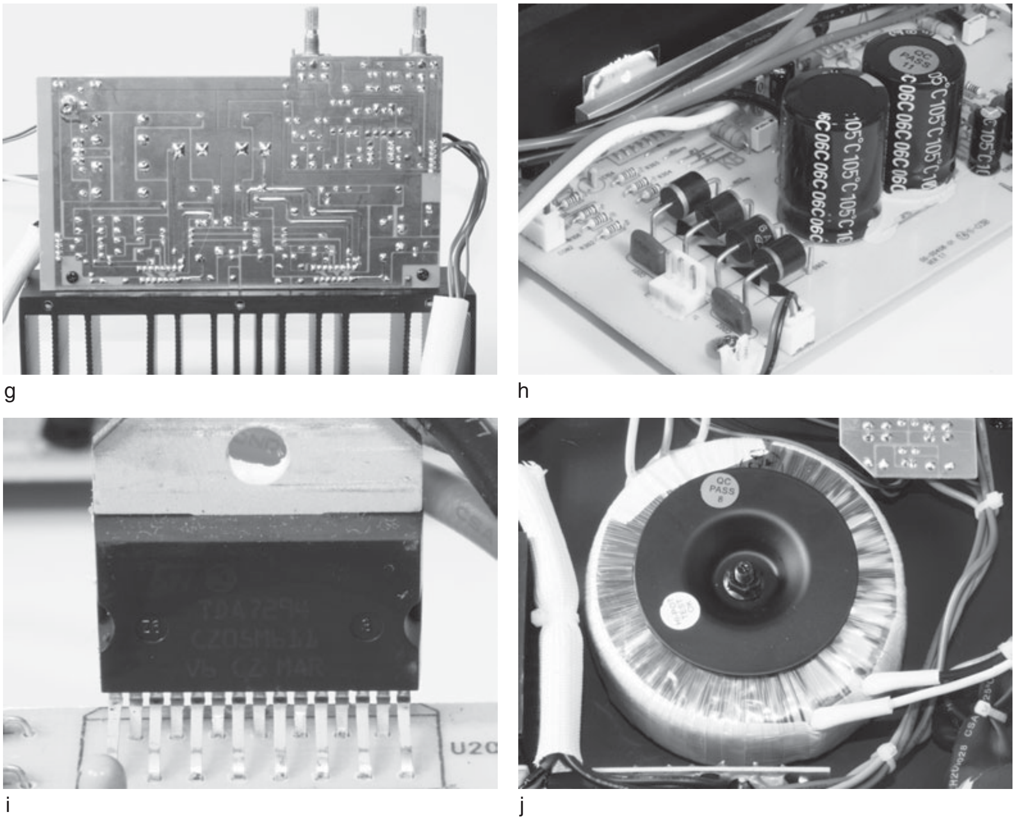

FIGURE 9-2 (*continued*) Teardown sequence

Step by Step

I found it easier to examine the live circuitry by extracting the transformer from the chassis, which involves unsoldering the leads to the fuse and power cord. You'll have to connect another power cord to the transformer, preferably through a Variac or other fused source, to work with the two circuit boards. If you decide to extract the transformer, you'll need a soldering iron, a crescent wrench to remove the toroidal transformer, and another ten minutes for the teardown.

Step 1

Remove the top cover. Remove the machine screws that secure the top cover. You'll find two on the top, near the faceplate; two screws on either side of the unit; and three in the back. Lift the cover up and back and set it aside. Figure 9-2a shows the view from the rear of the unit. Note the relative positions of the input and output jacks and the mode switches on the rear panel, the large heat sink on the left, and the massive toroidal transformer on the right.

Step 2

Remove the faceplate. The AMP 100 has a cosmetic faceplate attached to the chassis. Figure 9-2b shows the faceplate and chassis adjacent to the power board. Note the potentiometer shafts running through the chassis at the upper-left of the figure.

To remove the faceplate, first pull off the volume and balance knobs from the potentiometer shafts using your hands. You shouldn't require tools, but if you do, protect the faceplate from nicks and scratches with a cloth or thin piece of plastic or wood. Next, remove the screw on top and two screws on the bottom of the faceplate. Set the faceplate aside with the cover.

Step 3

Release the front panel components. Remove the 10mm nuts from the shafts of the volume and balance potentiometers. Because the front of the chassis is hidden by the faceplate, you can use a crescent wrench or other means to remove the 10mm nuts, even if you scratch the chassis. However, if you have a hollow shaft nut driver, you should be able to remove the nuts without leaving a mark.

Remove the four screws securing the power and speaker select push buttons to the front of the chassis, as shown in Figure 9-2c. The speaker select push button assembly is shown in Figure 9-2d. The larger wires carry the audio output while the two pairs of smaller wires supply the LEDs.

Step 4

Release the rear panel components. Rotate the unit so that you're facing the rear of the chassis, and remove the screws securing the speaker output lugs on the left. As shown in Figure 9-2e, the left and right channel connections on the speaker output board are electrically isolated from each other. Remove the five machine screws securing the logic board to the right rear of the chassis. This board contains the RCA connectors.

Step 5

Extract the large circuit board and heat sink. Next, remove the screw securing the large power board to the chassis. Note that the heat sink is attached to power amplifier ICs on the power board, as shown in Figure 9-2f. Now, from the bottom of the amplifier, remove the three screws securing the heat sink. Figure 9-2g shows the extracted power board and heat sink assembly from the perspective of the circuit board bottom. The shafts of the volume and balance potentiometers are at the upper-right of the figure.

Step 6

Release the power cables. Disconnect the three-wire transformer secondary cable connector and the two-wire power cable connectors leading to the power board. Figure 9-2h shows the cable connectors on the power board. The two-wire connectors at the bottom of the figure feed the LEDs within the power and speaker selection push buttons. Set aside the main chassis with the toroidal power transformer, power cord, and fuse holder.

Step 7

Release the power ICs from the heat sink. This will allow you to get a good look at the ICs. If you're following along with your amplifier at home, note that the ICs are electrically insulated from the heat sink. If you reassemble the amplifier without ensuring that the insulation is intact, you'll fry the ICs and likely damage the power supply.

Remove the four machine screws securing the two power ICs to the heat sink. The screws run from the heat sink to an aluminum bar that secures the two power ICs to the heat sink. Remove a fifth screw that secures the power board to the heat sink. Note the mica insulators and conductive grease between the heat sink and the metal tabs on the power ICs. Keep the two mica insulators and conductive grease with the heat sink and move the heat sink to a safe place where the mica won't be disturbed. Figure 9-2i shows one of the power ICs unattached to the heat sink.

Step 8

Extract the toroidal transformer. If you want to test the live circuit, you should extract the toroidal transformer from the chassis and supply the primary with 120VAC. Otherwise, you'll be constrained by the bulk of the chassis. First, unsolder the primary winding connections to the line cord and fuse. Next, remove the bolt running through the center of the toroidal transformer, shown in Figure 9-2j, and through the chassis bottom. Remove the transformer and move the chassis to the side. Solder a new power cord to the transformer primary, using electrical tape or shrink-wrap tubing to insulate the solder connections. Don't forget to use a fused source when you plug in the cord.

Layout

The internal layout of the AMP 100 is worth studying. There's no wasted space, and yet the airflow through the heat sink isn't compromised. During the teardown, we encountered seven subcomponents:

- Toroidal transformer, mounted directly to the chassis. The primary is soldered to the fuse and power cord, while the secondary terminates in a three-conductor connector that mates with the power board near the four silicon rectifiers, as shown in Figure 9-2h.
- Power board that contains the power supply, power amplifier ICs, an operational amplifier, and the volume and balance potentiometers. The board is secured by the potentiometer shafts to the front of the chassis, by the aluminum heat sink, and by a metal standoff connected to the chassis bottom.

- Large anodized aluminum heat sink connected to the two power ICs. The heat sink is positioned near the center of the chassis, immediately above the ventilation slots in the base of the chassis and below the ventilation slots in the cover. See Figure 9-2f.
- Logic board with two pairs of RCA audio input jacks for Line 1 and Line 2, one pair of output jacks, mono/stereo and normal/sensing switches, four operational amplifiers, a transistor, and a relay.
- AC power front panel switch assembly. Embedded within each switch is an LED that indicates a depressed condition. The power board supplies power for the LEDs.
- A/B signal selection front panel switch assembly. This switch assembly accepts the right and left channel outputs from the power board and distributes the output to the appropriate terminals on the audio output connector assembly. As with the power button, an LED within each button receives power from the power board. See Figure 9-2d.
- Audio output connector assembly. The output connector, which accepts banana jacks as well as tinned wire, is a little too cramped for my tastes. Fortunately, connecting speakers to the amplifier is usually a one-time affair.

To get the most out of this exercise, disconnect the push button and audio output assemblies from the two circuit boards, as shown in Figure 9-3. This will enable you to examine the layout of the two boards in more detail, with unrestricted access to components while probing for signals.

You can avoid the hassle of connecting the output to speakers by terminating the left and right audio circuits with 8Ω power resistors. If you keep the volume

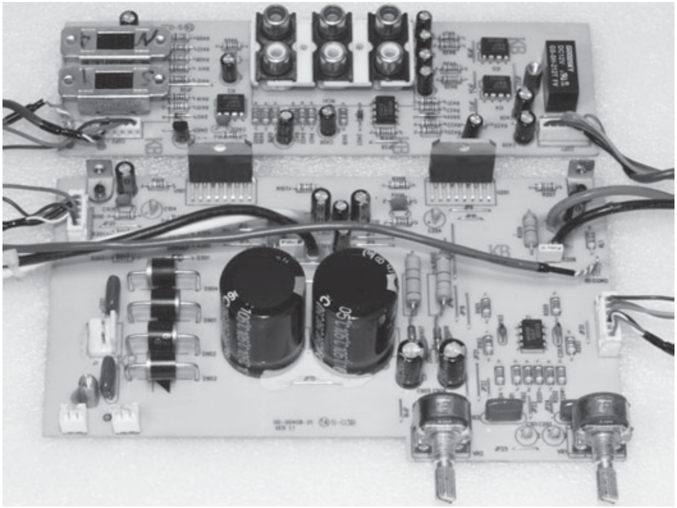

FIGURE 9-3 Circuit analysis setup

potentiometer turned down, you can get away with a 10W resistor on each channel. In addition, because the ICs aren't connected to a heat sink, limit power-on times to a few seconds. Otherwise, there's a risk of thermal shutdown, and you'll have to wait a few minutes for the integrated amplifier ICs to cool before you can return to testing.

Now let's focus on the larger power board and its connections. Orient the board so that the two potentiometers are on the right, as shown in Figure 9-3. In this position, the TDA7294 IC that powers the left channel is on the left and the TDA7294 for the right channel is on the right. The audio output from each power IC exits the board immediately to the right of each chip. Three cables connect the two boards: a six-conductor cable on the right, a five-conductor connector on the left, and a separate bridging cable that extends from the lower-left edge of the logic board to the right edge of the power board, near the right audio output connection.

The five- and six-conductor cables each include a miniature coaxial cable with dual center conductors for signal right and left. The coaxial shield is signal ground. The remaining wires in the six-conductor cable are ±15VDC and power supply ground. The five-conductor cable carries 15VDC and power supply ground.

The bridging cable is worth examining closely. It consists of a coaxial cable containing a twisted pair of wires. The shield is not terminated on either end of the cable, and both inner wires are connected together to solder posts on the circuit boards. Why not use a single wire for the inner conductor? Why use coaxial cable?

The power supply components, less the toroidal transformer, are located on the front half of the power board. Starting on the left edge of the board, we have a three-pin connector for the secondary of the toroidal power transformer, four silicon power diodes, Mylar bypass capacitors, and two 6800µf at 50VDC electrolytic capacitors. Note how the power diodes are mounted above the board for heat dissipation. Immediately behind the balance potentiometer are a pair of zener diodes and fusible resistors, as shown in Figure 9-4. The large filter capacitors feed the power ICs directly with ±36VDC, while the zener diodes provide ±15VDC voltage for the other ICs, a transistor, and a relay.

Caution This is a good point at which to note that 36VDC is enough to cause you serious harm. Moreover, because this is a dual-voltage supply, you could find yourself in contact with 72VDC if you're not careful. If you decide to work with the live circuit, don't become complacent simply because you're dealing with solid-state components.

Behind the volume control is a 5532D dual low-noise operational amplifier, as shown in Figure 9-5. It receives audio input from the adjacent cable and feeds the inputs of the two TDA7294 ICs. Other components on the power board worth noting are the two dozen, five-band, 1/8W metal film resistors; four polarized tantalum capacitors associated with the power IC; and each potentiometer. Most of these components are visible in Figure 9-5. There are no ferrite beads, chokes, or other inductors on the board.

Figure 9-5 also shows the main board with the connection to the heat sink intact. From this perspective, you can see the aluminum compression bar and the thermal grease between the amplifier ICs and the heat sink.

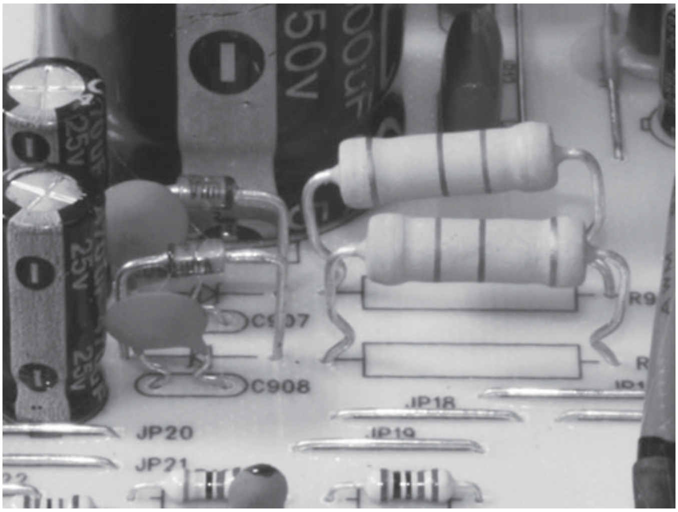

FIGURE 9-4 Zener diodes and fusible resistors on the power board

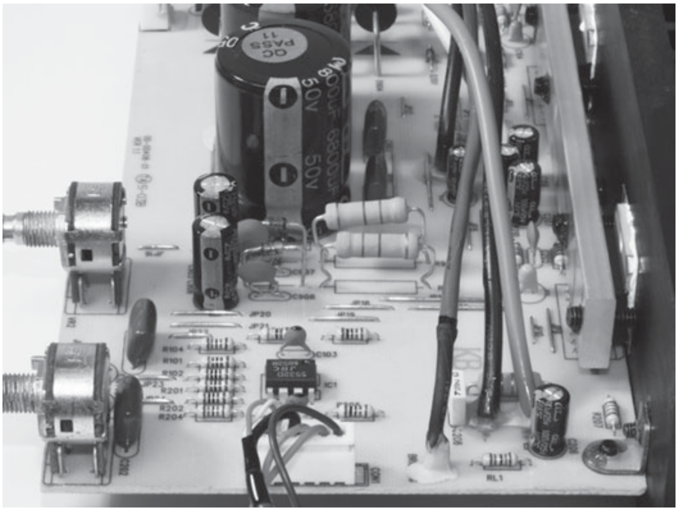

FIGURE 9-5 Right side of power board

One of the features of the single-sided power board is that components are labeled in a logical manner. Components associated with the right channel are numbered in the 200s, such as U201, R205, and C204. Components associated with the left channel are numbered in the 100s, and the labels are applied symmetrically. For example, U101 and U201 are both power ICs and

C104 and C204 are tantalum capacitors of equal value used with U101 and U201, respectively. This labeling makes signal tracing and creating a schematic a breeze.

The logic board is more compact, more complex, and more of a challenge to sort out, compared to the larger power board. Starting on the left of the board, using Figure 9-6 as a reference, you see two DPDT (double-pole, double-throw) slider switches. The top switch is for the switching/normal setting and the bottom switch is for setting the mono/stereo preference. The switches are designed to prevent accidental changes in the settings. There's a slot to use a coin or screwdriver blade to slide each switch. Continuing to the right of the switches is the only transistor on the board, a C2235 NPN transistor in a TO-92 package. To the right of the transistor is a 4558D dual general-purpose operational amplifier in an 8-pin DIP (dual inline package).

Now, moving just to the right of center, in front of the two RCA signal output jacks, is a second 4558D, as shown in Figure 9-7. To the right of the RCA connectors are a pair of 5332D operational amplifiers, identical to the one used on the power board. Finally, there's a 12VDC DPDT relay on the right edge of the board, immediately behind the 6-pin connector. As shown in Figures 9-6 and 9-7, there are a dozen small electrolytic capacitors, a few dozen five-band metal film resistors, and a few signal diodes on the board.

Like the power board, the single-sided logic board is labeled, but other than the ICs, switches, and relay, components are numbered in the 400s. To add to the modest challenge of tracing the circuitry on the logic board, there are jumpers and resistors under the RCA connector block. Don't overlook these as you follow the signal path through the circuitry on the board.

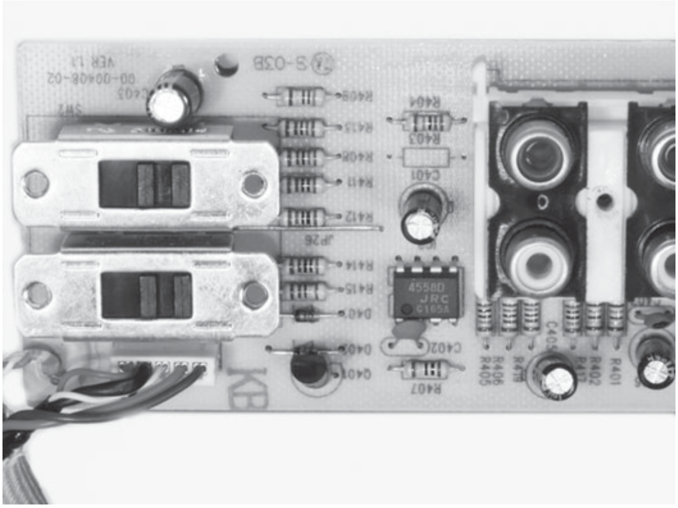

FIGURE 9-6 Left side of logic board

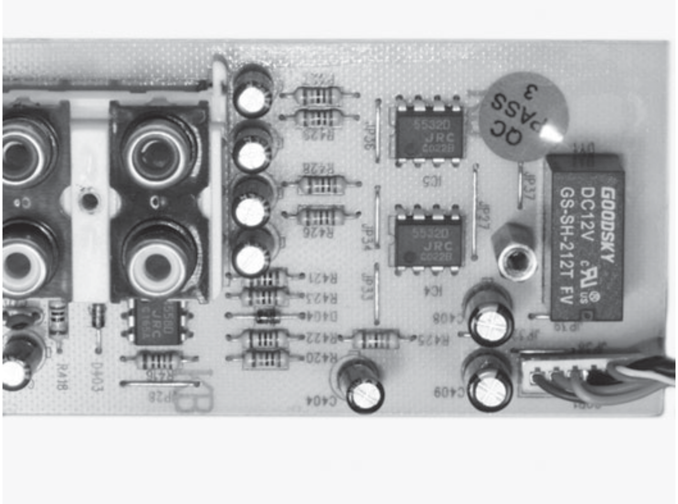

FIGURE 9-7 Right side of logic board

Components

If you're following along with an AMP 100 on your bench, take the time to examine each component carefully as it's discussed. Although I wouldn't unsolder the ICs unless you plan to upgrade the amplifier with higher performance chips, anything else is fair game. Not sure of the value of one of those five-band resistors? Unsolder one end and measure it with an ohmmeter. Want to check out the adequacy of the large electrolytic filter capacitors in the DC supply? Use your oscilloscope to monitor the DC voltage while you drive the amplifier at full output. Just exercise the usual caution around live circuits.

TDA7294 Power ICs

It should be no surprise by now that the two TDA7294 power ICs define the capabilities of the AMP 100. According to the datasheet from STMicroelectronics (www.st.com), each 15-pin IC features a MOS power amplifier that can deliver output current of ±10A when supplied with 80VDC. The IC is surprisingly affordable; at the time of writing, a single unit price was less than $6 from Digi-Key (www.digikey.com). If you own a subwoofer or self-powered speaker, there's a good chance you'll find a TDA7294 or similar IC inside.

The datasheet is invaluable in understanding the features of the TDA7294. For example, it details how the TDA7294 is configured internally to operate with another TDA7294 in bridge mode. What's surprising is the marked similarity between the suggested circuits in the datasheet and the AMP 100 circuitry. By the way, if you're trying to figure out the pin numbering on this oddly packaged chip, the first pin is on the far left if you're facing the front of the package, as shown in Figure 9-2i.

The TDA7294 is theoretically capable of delivering 200W output power, operating as a Class AB amplifier. This level of power is theoretical because it assumes a high-efficiency heat-dissipation system that is practically unachievable. As it is, the heat sink in the AMP 100 occupies a greater volume than that of the two circuit boards combined.

Fortunately, the TDA7294 is internally protected against thermal meltdown. It automatically switches to a muted state when its internal temperature reaches 145°C (293°F) and to standby at 150°C (302°F). In addition, even if you use liquid nitrogen to cool the ICs, the THD without noise exceeds 10 percent when the chip is pushed above 85W output.

Muting and standby functions can be controlled externally as well, independent of the thermal protection circuitry. Muting the audio output of the power chips is useful at startup to avoid switching noise as the AMP 100 selects the input channel and activates the auto-on function. Standby is a power-saving feature that supports the auto-on function as well. Without standby, the ICs and heat sink would be required to dissipate heat at all times.

In addition to mute and standby functions, automatic thermal shutdown, and short-circuit protection, each IC maximizes efficiency by switching on the internal power transistors only during the portion of the signal when more output voltage is needed. Although the IC has separate voltage supply pins for this power transistor circuit and the remainder of the chip, they are tied together in the AMP 100.

The TDA7294 requires only a handful of discrete capacitors and resistors to define its input impedance, gain, standby time, and mute time, and to bypass supply voltage inputs. In addition, the IC doesn't require a highly regulated DC power supply because its supply voltage rejection (SVR) is at least 60dB. SVR, also referred to as Power Supply Rejection Ratio, or PSRR, is a measure of the change in supply voltage divided by the simultaneous change in output voltage, if any. The ideal SVR is infinity, meaning that the supply voltage can swing wildly, while the output voltage remains unaffected. Even so, an SVR of 60dB allows for a simple power supply design.

Caution The metal tab on the IC is connected to the negative voltage supply pin, which is at –36VDC on the AMP 100. Not only does the exposed tab present a shock hazard if you're working with a hot circuit, but if you reattach the heat sink without properly insulating the tab, you'll short-circuit the power supply through the IC.

Extruded Aluminum Heat Sink

As noted in the discussion of the power amplifier ICs, the limiting factor in extracting power from the TDA7294 is removing heat. If you recall from basic physics, heat is transferred by conduction, radiation, and convection, from an object that's at a higher temperature than its surroundings. The AMP 100 relies on all three methods of heat transfer.

The power ICs are physically connected to a large heat sink that is cooled by natural convection. The chassis ventilation is designed with slots immediately below and above the cooling fins, maximizing the flow of heat carrying air, or convection currents, from below to above the unit. Heat from the ICs is conducted through the metal tabs on the IC to the aluminum heat sink. A small amount of IR radiation is emitted from the ICs and heat sink that you could see with an IR camera.

Heat sinks are rated in degrees centigrade/watt (°C/W). The specifications for the power ICs stipulate a heat sink rated at no more than 1.4°C/W, meaning that the temperature of the device attached to the heat sink shouldn't rise more than 1.4°C for each watt dissipated by the heat sink. This temperature differential is relative to the heat sink temperature, not the ambient temperature. For example, given 10W of dissipation and our 1.4°C/W heat sink rating, the ambient temperature may be 23°C (74°F), the heat sink temperature 35°C (95°F), and the device temperature of 35 + 14 = 49°C (120°F).

The AMP 100's heat sink is black anodized aluminum, weighs in at 19 ounces, with dimensions of about 7 3/4 × 2 × 2 1/2 inches. The °C/W rating for our heat sink is a function of the size and number of fins, type of aluminum, surface finish, surface area, and mass of metal. For example, the little corrugations on the fins increase the surface area and therefore the contact with cooling air, both of which decrease the °C/W rating. Even the color of the heat sink can affect the °C/W rating: a heat sink that is anodized flat black is a better thermal radiator than one finished in shiny aluminum.

All else equal, the larger the heat sink, the better the °C/W rating. In addition, the rating can be improved significantly by forced convection cooling, in which air is forced over the fins of the heat sink or chilled water or another coolant is circulated through the heat sink. A solid-state Peltier device, such as those used in battery-powered freezers, can be used to transport heat actively from the ICs to the heat sink or from heat sink to a larger heat sink.

But size and °C/W rating aren't the full story. If you bolt one of these power ICs to a 500-pound block of stainless steel, the average temperature of the block would increase very little, regardless of how hard you drive the TDA7294. However, the temperature of the stainless steel immediately surrounding the IC could rise considerably, even to the point of thermal shutdown.

The issue is thermal resistance, which roughly parallels the electrical resistance of materials. A high thermal resistance, like a high electrical resistance, impedes the flow of heat energy from the source. Silver has the highest thermal conductance of common metals, followed, in order, by copper, gold, and aluminum. If you've ever super-clocked a PC, you know that the better CPU coolers are made of copper for improved conduction over less-expensive aluminum coolers. As a point of reference, the thermal conductance of aluminum is ten times that of stainless steel.

If you removed the heat sink from the ICs during the teardown, you probably noticed that at the point of contact with the tab of the IC, the heat sink surface was smooth, bare aluminum. This maximizes the contact area and minimizes the thermal resistance between the tab of the IC and the heat sink. In addition, you probably encountered the messy thermal grease between the power ICs and the heat sink. The purpose of this thermal grease is to fill the microscopic irregularities in the two metal surfaces with a conductive media instead of thermally insulating air.

Note If you reconnect the heat sink to the ICs, use the smallest amount of conductive grease possible to cover the contact area. More is definitely not better, because the surplus grease diminishes the metal-to-mica contact area and thermal resistance skyrockets.

Because the tab on the TDA7294 is at –36VDC and the heat sink is bolted to the chassis, the IC must be electrically isolated and yet thermally bonded to the heat sink. This represents a challenge, because thermal conductors are generally electrical conductors. Mica is one exception. It's an excellent electrical insulator and a good thermal conductor. Hence the use of the thin mica sheets between the tabs of the IC and the heat sink, as shown in Figure 9-8. The thermal grease on both sides of the mica creates a thermally conducting, electrically insulating mica sandwich of sorts.

Alternatives to mica include polyimide, silicone rubber reinforced with fiberglass, and anodized aluminum. I've used these alternatives on 5W transistors, but I wouldn't dream of using silicone rubber, for example, on a device of more than 10W.

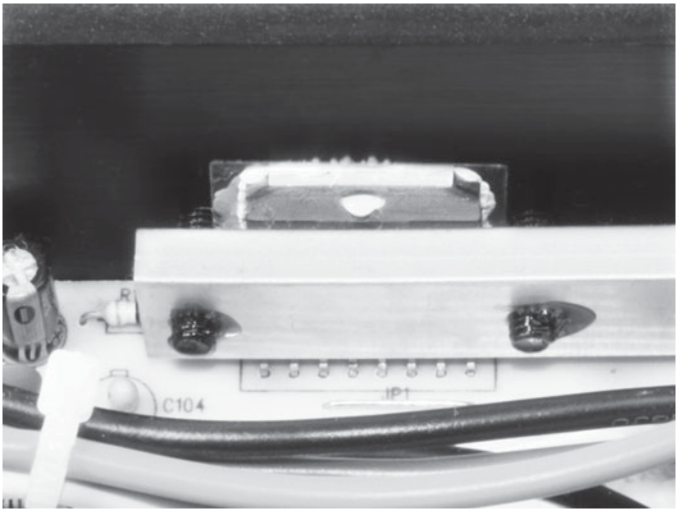

FIGURE 9-8 Mica insulator sheet between heat sink and amplifier IC

5532D and 4558D Operational Amplifiers

As mentioned during the teardown, the two other types of ICs used in the AMP 100 are 5532D and 4558D dual operational amplifiers. The 5532D is excellent for audio work because of its low noise and high bandwidth. The 4558D is a general-purpose operational amplifier, similar to the common 741, but with different pinouts. Both ICs are significantly less expensive than the power ICs. Digi-Key prices range from 15 to 30 cents for these chips in DIP packaging. Each 8-pin DIP chip contains two independent operational amplifiers.

C2235 NPN Transistor

The sole discrete transistor in the AMP 100 is a C2235, located on the logic board. This NPN transistor is rated at 120V maximum collector-emitter breakdown voltage, 800mA maximum collector current, and a current gain (Hfe) of 80. The TO-92 package transistor, capable of dissipating 900mW, is commonly used in audio power amplifier and driver stage amplifier applications.

Toroidal Transformer

The AMP 100's compact form factor is attributable in part to the massive, 62-ounce, low-profile toroidal power transformer that's a mere 1 1/2 inches thick and 4 inches in diameter, shown in Figure 9-1. A conventional, upright laminate power transformer with the same ratings could easily require twice the vertical clearance and nearly double the height of the amplifier chassis. In addition to a compact form factor, properly constructed toroidal transformers offer greater line noise immunity and higher efficiency, compared to conventional EI laminate core power transformers. (See the teardown of the ultrasonic humidifier in Chapter 7 for more information on EI core transformers.)

The toroidal transformer secondary delivers 26.8VAC center tapped, or 53.6VAC across the full secondary, at 3.2A. I measured the inductance of the primary winding at 2.8H and a DC resistance of 3Ω. At 60Hz, the inductive reactance of the primary is 2πfL or $2 \times \pi \times 60 \times 2.8 = 1055\Omega$.

Diodes

The most prominent diodes in the AMP 100 are the four 6A4 silicon power diodes on the power board. These diodes, which are the diameter of AAA batteries, are rated at 6A continuous, 400A peak surge current and 400V PIV (peak inverse voltage). As noted in the teardown, they're also elevated one body diameter above the power board to maximize convection cooling.

There are also four 1N4148 high-speed switching diodes on the switching board and one on the power board. These glass diodes, shown in Figures 9-6 and 9-7, are rated at 450mA and 100V PIV.

The most interesting diodes in the amplifier are the pair of 1N4744 zener diodes on the power board, shown in Figure 9-4. These glass diodes, rated at 15V and 1W, are used as +15VDC and –15VDC shunt regulators in conjunction with the 2W fusible resistors shown in the figure.

Capacitors

Most of the capacitors in the AMP 100 are electrolytic. As shown in Figure 9-3, there are two large 6800µf at 50VDC electrolytic capacitors in the center of the power board. The remaining 21 electrolytic capacitors, with values from 47 to 220µf at 25VDC, are distributed on the power board and logic board. Because the AMP 100 uses a dual supply system, the negative lead of some electrolytic capacitors are at ground potential, while the positive leads of others are grounded.

Two 0.1µf tantalum capacitors are associated with the potentiometers and two with the power ICs. Like electrolytic capacitors, tantalum capacitors are polarized. The half dozen ceramic disc capacitors are primarily low-voltage bypass capacitors. More prominent are the four 0.01µf and two 0.1µf Mylar capacitors used as bypass capacitors in the power supply and with the balance and volume control circuits. Finally, there are two rectangular 100nF metalized polyester film capacitors on the output of the power ICs, immediately adjacent to the audio output cable on the power board, as shown in the lower-right corner of Figure 9-5.

Resistors

A pair of 2W 470Ω fusible resistors is used in the zener diode shunt regulator circuit, shown in Figure 9-5, and a 1W 470Ω fusible is used in each audio output circuit, shown in Figure 9-3. Fusible resistors are designed to fail gracefully—that is, without flame, due to a flame-retardant coating—when subject to current above their specified power level.

The most striking resistors on the two circuits are the 1/8W metal film resistors. Instead of the usual 5 or 10 percent tolerance resistors, the boards are filled with 1 percent tolerance resistors. While the price differential between a 1 percent and a 5 percent resistor is probably a cent or less in large volume purchases, most manufacturers use the cheapest resistors available. The use of 1 percent resistors in the AMP 100 speaks to the quality of the amplifier. The value of these resistors can be a little difficult to decipher at first because of the extra digit band and because the multiplier band is brown, as opposed to the more easily identified gold or silver. Review the resistor color codes in Appendix A if you're not familiar with the five-band system.

Stacked Potentiometers

The four 100KΩ potentiometers in this amplifier are packaged as two stacked potentiometers with two potentiometers per shaft. Each shaft controls both left and right channels of the amplifier. The alternative, separate volume and balance potentiometers for each channel, would demand more front panel space and would likely put off most consumers.

Although they appear identical on the circuit board, there are mechanical and electrical differences between the two stacked potentiometers. A minor mechanical difference is the center notch on the balance potentiometer to indicate the midpoint of rotation. There is no mechanical notch on the volume potentiometer.

Electrically, the potentiometers are considerably different. The left and right balance potentiometers are linear taper, meaning the resistance between the center contact or wiper and one of the terminal contacts is proportional to the distance between the contacts and to the degree of rotation. You can expect the same resistance change per degree of rotation, regardless of the initial position of the wiper.

The volume potentiometer, in contrast, is a dual audio or log taper device, meaning the resistance change is a nonlinear function of the distance between the wiper and a terminal contact. The advantage of log over a linear taper is that the resulting change in volume more closely matches the logarithmic response of the human ear to changes in volume. Although the resistance change, and therefore output volume, is a nonlinear function of the potentiometer rotation, the change in volume is perceived as linear.

If you were to replace the log taper stacked potentiometer with the linear taper stacked potentiometer used for the balance control, the volume control would seem overly sensitive at low volumes and virtually nonresponsive at high volume settings. In fact, you should swap potentiometer positions and experience the effect firsthand.

Relay

The single relay in this amplifier, located on the right edge of the logic board in Figure 9-7, is a Goodsky GS-SH-212T. The DPDT relay has silver alloy contacts rated at 120VAC at 1A and 24VDC at 2A. The contacts close at about 17mA. Because the AMP 100 relies on silent convection cooling, the clicking is noticeable when the relay comes online.

As with manual switches, relay contacts are rated differently for AC and DC operation. In general, the AC specifications must be derated for DC operation. For example, if the Goodsky relay is used at 120VDC, the maximum current is likely around 250mA. (More detailed specifications on the relay are available at www.goodsky.co.uk.)

Fuse

The 4A slow-blow fuse is remarkable mainly in that it and the on–off push button switch are the only components between the power transformer primary and the power mains. There is no EMI filter or surge suppression, which I've come to expect in better electronics.

How It Works

A systems view of the AMP 100 is summarized in the block diagram in Figure 9-9. As shown in the figure, the amplifier consists of a logic board, a power board, a power transformer, a speaker selection switch, a power switch, and an output connector block. The source audio enters the system on the left, through the RCA input jacks on the logic board. The signal is buffered prior to being sent to a pair of drivers and then power amplifier ICs on the power board. While in the logic board, the signal is acted on by the sense circuitry, as a function of the stereo/mono and normal/auto switch settings on the logic board.

The power board contains the power supply, which is connected to the power switch module and the toroidal power transformer. It feeds the driver and power ICs on the power board and the sense circuitry and buffer on the logic board. The power board contains the volume and balance pots, accessible from the front of the AMP 100. The A/B speaker selection switch, accessed from the front of the amplifier, defines the electrical configuration of the output block in the rear of the unit. The output block is shown configured for stereo, with right and left speaker output. However, bridged output, using a single speaker between the left and right speaker output terminals, is also supported.

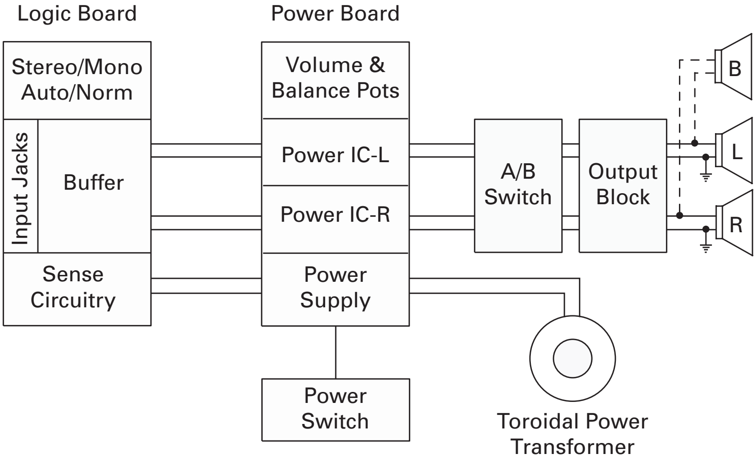
Logic Board
Power Board
Stereo/Mono
Auto/Norm
Input Jacks
Buffer
Sense
Circuitry
Volume &
Balance Pots
Power IC-L
Power IC-R
Power
Supply
Power
Switch
A/B
Switch
Output
Block
B
L
R
Toroidal Power
Transformer

FIGURE 9-9 Block diagram of amplifier

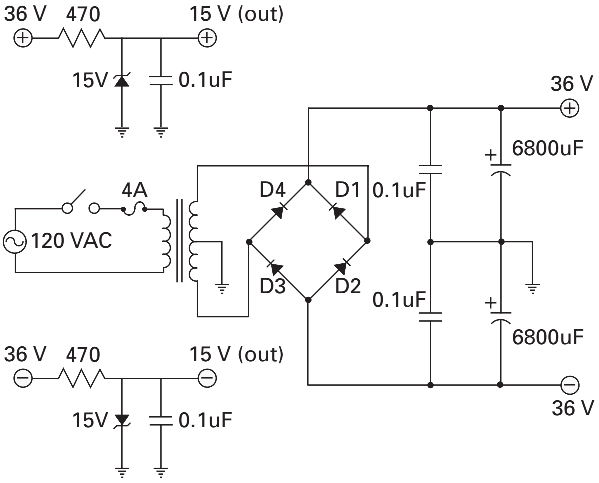

FIGURE 9-10 Schematic of power supply

Now let's examine the power supply schematic, shown in Figure 9-10. The 120VAC primary circuit is a bare-bones power switch and 4A slow-blow fuse. As I noted earlier, there are no inductors or filter capacitors in the primary circuit to minimize susceptibility to, or transmission of, EMI. The main power supply provides ±36VDC to the power ICs and the zener diode shunt regulators, shown at the top-left and bottom-left of the figure.

The secondary of the power transformer, which has a center tap to ground, feeds a bridge rectifier. In a standard bridge configuration, the negative junction of the bridge—in this case, the bottom corner of the bridge—is connected to ground, and the opposite, positive junction is connected to the electrolytic capacitors and other components of the output filter. However, in this power supply, both the negative and positive voltages output from the bridge rectifier are referenced to ground. As a result, instead of a single-ended 72VDC power supply, we have a dual-voltage supply that provides ±36VDC. A 6800µf electrolytic and 0.1µf Mylar capacitor are used on each output to ground. The Mylar capacitors shunt the high-frequency transients to ground, and the electrolytic capacitors smooth out the pulsating DC.

Let's follow the operation of the power supply through one cycle, using Figure 9-10 as a reference. When the top of the secondary winding is positive relative to the bottom of the winding, D1 and D3 conduct. Because the secondary center-tap is grounded, half of the secondary voltage to ground, or 36VDC, is applied to each of the 6800µf capacitors. On the next half cycle, when the top of the secondary winding is negative relative to the bottom of the winding, D2 and D4 conduct. Again 36VDC is applied across each of the 6800µf capacitors. Because this bridge conducts on every half cycle, it is a full-wave rectifier.

The ±36VDC is supplied directly to the TDA7294s. The operational amplifiers, relay, and transistor have much lower working voltages and are supplied by the two

15V zener shunt regulators. Unlike standard silicon power diodes that pop, crack, and fizzle when the peak inverse voltage is exceeded, zener diodes are designed to maintain a fixed voltage drop with reverse current flow. Within the current-carrying capacity of the two zener diodes, the voltage across each reverse-biased diode will be 15VDC, regardless of the applied voltage. In this regard, the zener diodes function as variable resistors. If the voltage across one of the 15V zener increases in magnitude, its resistance drops. If the voltage across the diode decreases, its resistance increases.

This increase and decrease in the resistance of a zener diode with changes in applied voltage can be used to create a voltage regulator when the zener is combined with a series resistor of suitable value—in this case, 470Ω. The role of each series resistor is to drop the 36VDC supply voltage to 15VDC and to control the flow of current through the zener diode. If the resistance of the series resistor is too low, the current capacity of the zener diode will be exceeded and the diode will fail. If the series resistance is too high, the zener diode won't conduct and the output voltage may swing wildly. The optimum value of the series resistor can be determined by Ohm's law:

$$R = V/I$$

where R is the resistance of the series resistor: in this example, 470Ω. V is the voltage across the resistor: in this example 36V – 15V = 21V. The sum of the zener current, I, is the current required by the zener diode to keep it in controlled reverse current breakdown mode, and the current required by the circuit. Given the 21V differential and 470Ω resistor value, total current requirement must be 21V/470Ω, or 45mA. The datasheet for the 1N4744A lists a zener current of 17mA. Assuming the circuit designers used the 17mA figure for zener current, the circuit load must be 45mA – 17mA = 28mA.

The advantages of a shunt regulator include simplicity and low cost. The disadvantages include poor regulation if the input supply voltage is poorly regulated and inefficiency. Using our figures for load and shunt current, the circuit load is 15V × 28mA = 420mW. The power dissipated by the 470Ω resistor is $I^2 \times R$, or $0.045^2 \times 470$ = 950mW. The overhead of delivering 420mW to the load is nearly 1W, which is dissipated as heat by each series resistor.

You might be wondering why use the zener diode at all. After all, if you know the circuit requirement is 28mA, why waste 17mA on a zener diode? If you had a constant load and a fixed supply voltage, you could get away with a series resistor. However, we can't make those assumptions here. Even if the steady-state current requirement of the circuit is 28mA, the startup current requirements are likely significantly greater. For example, during the first second or two of operation, electrolytic capacitors are uncharged and present a low impedance path to the initial voltage surge.

The second issue is variation in supply voltage, both at startup and during normal operation of the amplifier. The output of the nominal ±36VDC supply varies with changes in current drawn by the power ICs, as well as with fluctuations in the line voltage. If series resistors alone supply the ±15VDC, then the ±36VDC voltage fluctuations transfer directly to the ±15VDC output.

Now, for the analysis of the amplifier proper, refer to the simplified schematic in Figure 9-11. Let's follow the Line 1, right-side signal, starting with the RCA connectors on the logic board. A 5532D operational amplifier, configured as a voltage follower or buffer, leads to the relay controlled by the sensing circuitry. The buffer presents high input impedance and low output impedance, and provides unity gain. It isolates the signal source from the amplifier circuit so that each input channel to the AMP 100 essentially looks like the 47µf electrolytic capacitor and 100KΩ shunt resistor. This is an insignificant load for a typical signal source.

From a signal perspective, ideally the input would lead directly to the noninverting input of the 5532D operational amplifier. However, it's possible that the source could contain a DC offset voltage, which would skew the operation of the operational amplifier. The series capacitor ensures that the operational amplifier operates on an AC signal without a DC offset.

Note If you're hungry for more information on operational amplifier function, there's more detail in the teardown of an effects pedal in Chapter 12.

If there is a signal on Line 1, the relay defaults to the position shown in the schematic. The audio signal is sent through the miniature coaxial cable to the six-conductor cable assembly that connects to the right edge of the power board. Otherwise, if there is signal on Line 2 and no signal on the input of Line 1, the switching circuitry connects Line 2 to the power board.

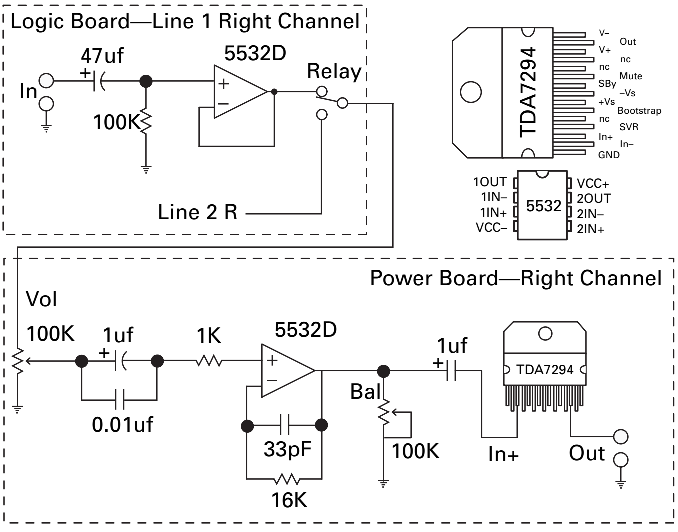
Logic Board—Line 1 Right Channel
47uf
5532D
Relay
In
100K
Line 2 R
TDA7294
V−
Out
V+
nc
nc
Mute
SBy
−Vs
+Vs
Bootstrap
nc
SVR
In+
In−
GND
1OUT
1IN−
1IN+
VCC−
5532
VCC+
2OUT
2IN−
2IN+
Power Board—Right Channel
Vol
100K
1uf
1K
5532D
0.01uf
33pF
16K
Bal
100K
1uf
TDA7294
In+
Out

FIGURE 9-11 Simplified schematic of amplifier circuit

Note that, for clarity, the input circuitry for Line 2, which is identical to that of Line 1, is not shown in Figure 9-11. Similarly, in the following discussion of the right channel of the power board, unless stated otherwise, the circuitry and functionality are mirrored by the left channel.

Following the signal path from the logic board to the power board, the signal is sent to the volume potentiometer. The tap leads to the noninverting input of the only 5532D operational amplifier on the power board. This operational amplifier is configured as a noninverting driver, with the 45µf capacitor in the feedback loop to taper its frequency response.

The 5532D operational amplifier is followed by a balance control potentiometer, which acts a variable attenuator. Rotating the potentiometer shaft clockwise increases the right channel's resistance to ground while simultaneously decreasing the left channel's resistance to ground. The signal is then presented to the noninverting pin of the TDA7294 that handles the right channel.

Now let's analyze the logic board in more detail. This task is less arduous than it could be because the designers did a good job of separating the amplifier functions on the power board from the switching and logic functions on the logic board. Let's start with the stereo-mono switch and function, using the simplified schematic in Figure 9-12 as a guide. For clarity, the logic board circuitry is represented by a single operational amplifier and the stereo-mono switch, *SW*. Similarly, the left channel signal chain on the power board is represented by a volume control, driver operational amplifier, and power IC.

Recall that the function of the stereo-mono switch is to enable the AMP 100 to amplify the signal input to the right channel and either drive both speakers in

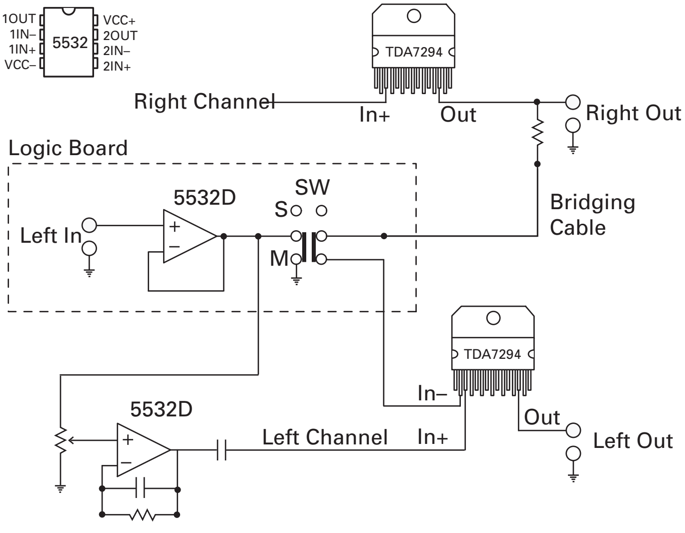

FIGURE 9-12 Simplified schematic of stereo-mono switching circuit

stereo mode or a single speaker in bridged mode. With the stereo-mono switch in the Stereo position, the switch is essentially out of the circuit. The signal input to the right RCA jack is sent to the noninverting (+) input to the TDA7294 dedicated to the right channel, and the signal input to the left RCA jack is sent to the noninverting (+) input of the other power IC. Both signals are subject to the same phase distortions. As such, if two signals are in phase at the input, they are in phase at the output of the amplifier.

When the stereo-mono switch is in the Mono position (M), as shown in Figure 9-12, the signal path is altered significantly. Now the output of the TDA7294 assigned to the right channel is sent to the inverting (–) input to the TDA7294 dedicated to the left channel via the coaxial bridging cable and series resistor. Any signal input to the left channel RCA jack is shunted to ground by the stereo-mono switch. This shunt essentially removes the remainder of the normal left channel signal chain from the circuit. Other than noise generated by the 5532D, no signal is presented to the noninverting (+) input of the left channel TDA7294.

The most important element of Figure 9-12 is the use of the inverting (–) input to the left channel TDA7294. This causes a 180-degree phase shift in the output of the left channel, relative to the output of the right channel. Recall that this phase difference is required for single speaker bridge amplification.

Now let's turn to the simplified schematic shown in Figure 9-13 to examine the circuitry associated with the normal-auto switch and the auto-input select function. As with the preceding schematic, circuit details, such as the components for biasing and operational amplifier feedback, have been omitted for the sake of clarity. The point is to understand the logic applied to signal flow.

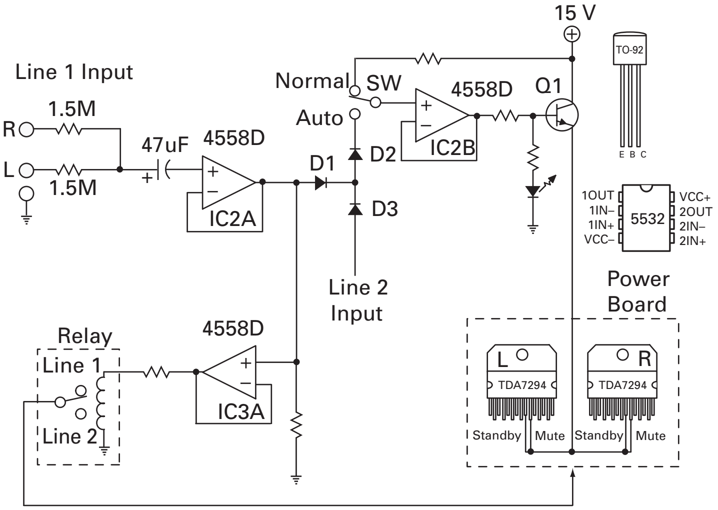
15 V
TO-92
E B C
Line 1 Input
1.5M
R
L
1.5M
47uF
4558D
IC2A
Normal
SW
Auto
4558D
IC2B
Q1
D1
D2
D3
Line 2 Input
1OUT
1IN–
1IN+
VCC–
5532
VCC+
2OUT
2IN–
2IN+
Power Board
Relay
Line 1
Line 2
4558D
IC3A
L
TDA7294
R
TDA7294
Standby
Mute
Standby
Mute

FIGURE 9-13 Simplified schematic of normal-auto and auto-input select circuits

Let's start with the function of the normal-auto switch. Recall that the position of this switch determines whether the front panel on–off switch toggles the AMP 100 between on and off or between on and active standby. With the switch (SW) in the Normal position, the noninverting (+) input of IC2B, a 4558D general-purpose operational amplifier, is tied to the positive 15VDC supply. This drives the output of the operational amplifier high, forward-biasing the C2235 NPN transistor and the LED connected to the base of the transistor and embedded in the front panel power switch.

The forward-biased LED simply indicates that the AMP 100 is on. The forward-biased C2235 transistor has a more significant role. As shown in the schematic, the emitter of the transistor is connected to the mute (*M*) and standby (*SBY*) pins of the two TDA7294 amplifier ICs. When the transistor conducts, the voltage on the pins is raised to about 12VDC, which deactivates the separate mute and standby functions, allowing the pair of TDA7294s to function as amplifiers. When the transistor is in the nonconducting state, these pins are at or near ground potential, which activates the mute and standby functions.

Now let's look at the circuit with the normal-auto switch in the Auto position. In this position, the amplifier is active when a signal is present on one of the input lines. Otherwise, the power ICs are muted and in low-power standby mode. From the preceding discussion, you should know that when the noninverting (+) input to IC2B is high, the LED is illuminated and the TDA7294s function normally. Conversely, when the input is low, the LED is dark and the TDA7294s are silent.

When a signal is present on either the right or left channel of Line 1, a signal is present on the noninverting (+) input to IC2A, a 4558D operational amplifier. The result is a positive output, which propagates through steering diodes D1 and D2 to the noninverting (+) input of IC2B. Similarly, when a signal is present on either channel of Line 2, the positive output propagates through steering diodes D3 and D2, to the noninverting (+) input of IC2B. In both cases, the LED and transistor are forward-biased.

Note that two 1.5M resistors are used to mix the inputs from the right and left channels of Line 1. Although the channel-to-channel resistance of 3.0M is significant, it represents a finite degradation in the separation of the two channels. An equivalent circuit is used for Line 2, which is omitted in the schematic for clarity. Of particular note is the use of switching diodes, D1–D3. Both Line 1 and Line 2 can contribute, independently, to the signal presented to input of IC2B. In addition, this isolation enables the auto-input select circuit to operate properly.

When a signal is present on Line 1, the output of IC2A is high. This signal drives the noninverting (+) input of IC3A, the output of which energizes the relay coil. The relay contacts are configured so that the right and left channels of Line 1 are fed to the power board.

Note that, if not for diode D1, the relay would be activated by a signal on Line 2. However, with D1 in the circuit, when a signal is present on Line 1 and not Line 2, the relay is not energized. The relay contacts route Line 2 signals to the input of the power board.

If you're following along at home with an AMP 100, you'll notice some components not included in Figure 9-13, such as a diode to protect IC3A from the inductive kick from the relay coil. If you're intent on modding the AMP 100, spend a few minutes mapping out the handful of additional components associated with the relay driver circuit. Also note the locations on the logic board where Line 1 and Line 2 are sampled for the switching logic.

Mods

It's hard to improve on a good thing, and this little amplifier is pretty good. However, you can address the most common complaints—erratic clicking of the relay as the auto-sensing circuit drops out during brief periods of low signal amplitude and lack of low frequency response—with the following mods.

Disconnect the Switching Circuitry

In theory, you could increase the SNR figure by disconnecting the sensing circuitry on the logic board. Simply unsolder one lead of each resistor that connects the sensing circuitry to the input connectors, using Figure 9-13 as a guide. Next, remove the resistor in series with the relay coil to avoid the annoying clicks sometimes present at startup. These steps will defeat the auto-input select feature as well as auto-on features, so you'll have to make due with manual on–off an a single input source. But the increased performance may be worth it to you. Fortunately, defeating the switching circuitry is an easy, readily reversible mod.

Create a Beefier Power Supply

If you follow the pro audio community online, you'll probably run across several references to lower than spec power output for the AMP 100 at 20Hz. To my ears, the existing circuitry is fine and I doubt that my speakers, headphones, or ears can handle anything below about 30Hz. Even so, if you want to feel your subwoofer rumble, the generic fix is beef up the power supply with additional filter capacitance.

There are three limits to this upgrade: the cost of high value, high density capacitors; the limits on additional capacitance imposed by the voltage rating of the power amplifier chips; and space. Give the limited height of the AMP 100's enclosure, you can't easily increase capacitance by installing taller capacitors of the same diameter. You might be able to locate higher density, higher capacitance replacement capacitors that fit on the board and in the enclosure, but they'll cost you. A more economical approach is to mount additional capacitors horizontally, adjacent to the two 6800µf capacitors.

If you really want to squeeze every bit of low-end response from the power ICs, consider moving the power transformer and power supply components to a separate enclosure. You'll have to feed the amplifier with ±36VDC and ±15VDC instead of 120VAC. If you opt for an external power supply, avoid the temptation of significantly increasing the capacitance of the filter circuit without modifying the circuit. If you double or quadruple the capacitance, you'll likely kick up the output voltage enough to require new series resistors for the zener shunt regulators. In addition to more capacitance, consider adding a high-current choke, and, if you have luck with locating surplus magnetics, separate transformers for positive and negative 36VDC supplies.

This page intentionally left blank

PART II

For Tinkerers

This page intentionally left blank

Chapter 10

Analog Volt-Ohm-Meter

The analog Volt-Ohm-Meter (VOM), such as the RadioShack VOM in Figure 10-1, has retained a place in the modern test instrument arsenal, despite the onslaught of inexpensive, mega-function digital multimeters (DMMs). For example, I keep both a Fluke 45 DMM and a Triplet 310 VOM on my workbench. The VOM, while not as precise or as accurate as the DMM, is my first choice for working with fluctuating or noisy signals and for rapid troubleshooting.

Not only is the response of my VOM virtually instantaneous, but the inertia of the meter movement serves as a mechanical low-pass filter to smooth fluctuations in the input signal. My five-digit Fluke 45, in comparison, often can't lock-in on a noisy signal. Even when my DMM is presented with a clean signal, it may take 5 or 6 seconds for the analog-to-digital converter and numerical display to settle down. So when accuracy and precision are called for, I turn to my expensive tabletop DMM. However, when I need speed and responsiveness, I turn to my analog VOM.

In this chapter, we'll tear down a RadioShack VOM, model 22-109, shown in Figure 10-1. The construction and feature set is typical of the sub-$20 instruments from companies such as Elenco, Mastech, and Mibuka.

Highlights

Unlike the typical, $10 single-chip DMM, this VOM offers a wealth of components and circuitry to examine. There is a mix of leaded and surface-mount components, including MOVs, Schottky, germanium, and general-purpose silicon diodes; a D'Arsonval galvanometer; a circuit board rotary switch; an electromagnetic buzzer; a linear pot; a fast-acting fuse; and, of course, resistors. This is a great teardown for brushing up on Ohm's law.

During the teardown, note the following:

- The circuit layout of the SMT components
- The marking system for SMT resistor value
- How to determine the tolerance of SMT resistors
- How you would improve the circuit design

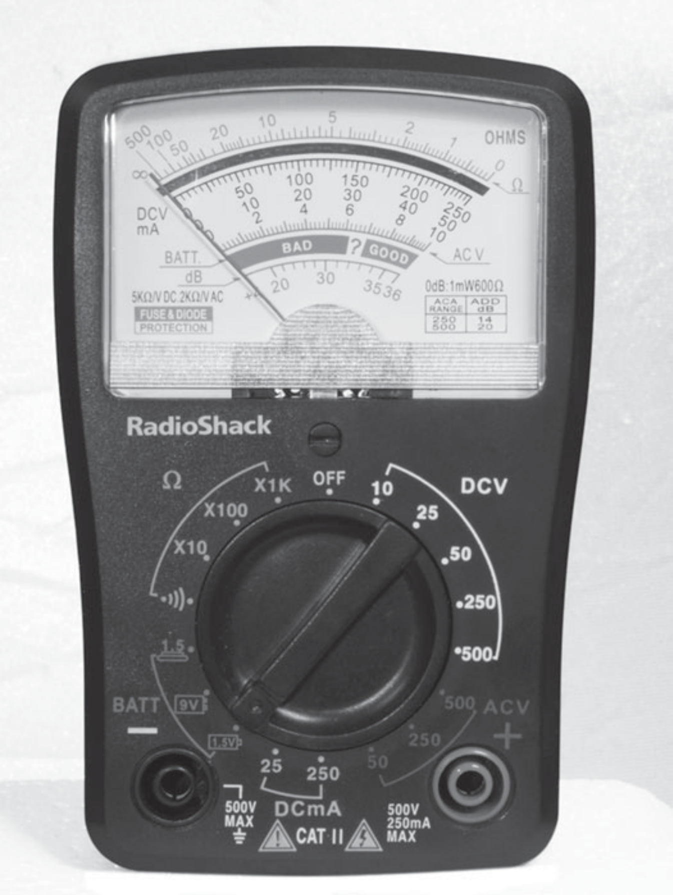

FIGURE 10-1 RadioShack VOM

Specifications

The specifications for the RadioShack 22-109 VOM, hereafter referred to as simply *the VOM*, are typical for inexpensive but useful analog meters. Key specifications include the following:

- 0–500VAC/DC
- DC current from 0 to 250mA
- Resistance from 0 to 500K
- Audible continuity test
- Battery test feature

- Meter face is calibrated to display voltage readings in decibels (dB)
- Accuracy is ±5 percent of full scale
- DC sensitivity of 5000Ω/V

Significance

This is a good time to review the differences between accuracy, precision, and resolution as they relate to test equipment.

Accuracy is the closeness of the measured value to the actual or true value. An instrument is accurate if it indicates the true value. Accuracy can be affected by changes in temperature, drift in internal component values over time, and damage to the internal components from electrical or physical accidents and abuse.

Precision is a measure of dispersion or distribution around a given value, regardless of accuracy. It might help to think of precision as a measure of repeatability. If an instrument reads the gold standard source as 398.1VDC every time, then the meter has good precision—even if it's off the actual value by a couple volts.

Resolution is the granularity of measurement. It's more relevant to digital displays than to analog meters, because your perceptual abilities will limit the resolution with which you can read the position of the needle on the meter face. In the digital world, a measurement of 6.0001VDC has a greater resolution than a measurement of 6.0VDC. Resolution says nothing about accuracy or precision. A cheap DMM might indicate that a 400.00VDC source is 374.04857VDC—nice resolution, but the accuracy falls outside of the 5 percent full-scale accuracy figure of our inexpensive VOM.

The bottom line is don't be fooled by a string of digits after the decimal point. More numbers don't equate to more accuracy or more precision. For this reason, an expensive benchtop DMM that hasn't been calibrated in a while may be less accurate than an inexpensive, but recently calibrated pocket VOM. Unfortunately, however, our inexpensive VOM isn't equipped with a means of calibrating it against a calibration source.

Interpretation

According to my measurements, full-scale deflection of the VOM's galvanometer occurs at 200µA and 145mV. The DC sensitivity of the galvanometer, the reciprocal of the full-scale DC current, is 1/200µA, or 5000Ω/V. This figure agrees with the published specifications.

The greater the sensitivity and the truer the accuracy of a VOM, the better. To illustrate, my pro-grade (read expensive) Triplet 310 pocket VOM is rated at ±3 percent of full-scale accuracy with a DC sensitivity of 20,000Ω/V. The Triplet VOM is clearly a superior instrument—albeit at a significantly greater cost than the RadioShack VOM. However, these figures are unimpressive when compared with the specifications of my Fluke 45 benchtop DMM, which has a resolution of 1µV and accuracy of 0.025 percent. A caveat concerning DMMs is that the least significant digit is subject to rounding and digitization errors.

As with many inexpensive VOMs, our VOM has no AC amperage function. Overall accuracy is ±5 percent of full-scale, meaning a value displayed on the 500VAC scale should fall within ±(0.05 × 500), or ±25, volts of the actual value.

To put the specifications of these two VOMs and the DMM into perspective, let's assume a gold standard 400.00VDC source. Using the 500VDC scale on the RadioShack VOM, the reading should be 400 ± 25VDC. The reading could be as low as 375VDC or as high as 425VDC, and the VOM would still be operating within specification. The Triplet VOM, which has a 0–600VDC scale, should read 400 ± 18VDC because 3 percent of 600VDC is 18VDC. A reading of 375VDC would not be expected from the Triplet. The resolution of the two VOM meter faces is about the same. The Fluke DMM, in comparison, should read 400.00 ± 0.1VDC. The least significant digit might be off by a count or two because of limitations in the A–D converter, but that's a deviation of only a few hundredths of a volt.

Operation

If you're new to analog VOMs, the first thing you'll note is that it has no automatic features. When you plug in the test leads, make certain you plug the positive lead into the red, or positive, jack and the black lead into the black, or negative, jack. Unlike a DMM that will simply display a negative value when you test a negative voltage or current, with an analog VOM, you risk damaging the galvanometer and blowing the relatively expensive fuse when you don't observe proper polarity. In addition, some VOMs have jacks dedicated to specific functions, such as high current. These functions won't work unless the appropriate jacks are used.

Another key manual operation is selecting the operating mode—Resistance, DCV, ACV, Battery Check, DCmA, or Off—and the range within each mode. To maximize accuracy and repeatability, you should select a range that places the needle in the right third of the meter. For example, if you're troubleshooting a 120VAC circuit, you should use the 0–250VAC range instead of the 0–500VAC range. Recall that at an accuracy of ±5 percent of full-scale, a measurement using the 0–500VAC range is ±25VAC, versus ±12.5VAC for the 0–250VAC range. See Figure 10-2 for a close-up of the meter face with scale values. Note the nonlinearity and direction of the resistance scale, with higher resistance values compressed into a few graduations on the left edge of the scale.

You can maximize accuracy and repeatability of all measurements by using the parallax-correcting mirror embedded in the face of the meter. Position yourself so that the needle hides its reflection. If you do this consistently, you'll reduce parallax errors. Because of the space between the needle and the meter face, if you view the meter at an angle, the needle will appear to the left or right of the actual position over the meter face.

You can maximize the accuracy of resistance measurements by zeroing the needle prior to measuring the resistance of a component. With the meter set to resistance and the probes touching, adjust the zero pot so that the meter reads full-scale, or 0Ω.

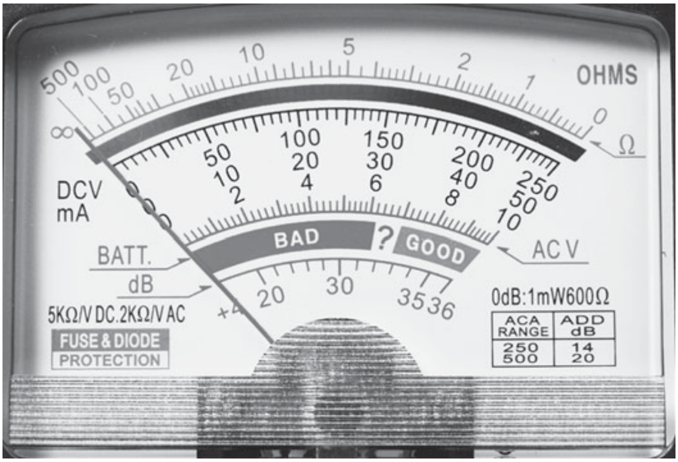
500
100
50
20
10
5
2
1
0
OHMS
∞
Ω
50
10
2
100
20
4
150
30
6
200
40
8
250
50
10
DCV
mA
BATT.
BAD
?
GOOD
AC V
dB
20
30
35
36
5KΩ/V DC.2KΩ/V AC
FUSE & DIODE
PROTECTION
0dB:1mW600Ω
ACA RANGE
ADD dB
250
500
14
20

FIGURE 10-2 Meter face

You can make small changes in the resting position of the needle by adjusting the screw at the front of the meter with the VOM in the off position. This screw mechanically adjusts the galvanometer's spring position, and so the zero position of the needle.

Proper operation can also maximize accuracy. For example, a limitation of this and other galvanometer-based VOMs is susceptibility to magnetic fields generated by unshielded speakers and similar electronic equipment. If you need to take measurements in or around high-powered, unshielded systems, consider using either a shielded VOM or a solid-state DMM.

Teardown

As illustrated in Figure 10-3, this 5-minute teardown involves removing only a handful of screws. You'll have to handle the galvanometer with care once you extract it from the protective VOM enclosure.

Tools and Instruments

You'll need a Phillips-head screwdriver and a multimeter to trace the circuitry. If you decide to remove components for more detailed inspection, you'll need a soldering iron with a thin, clean tip; solder braid; and forceps.

Step by Step

If you want to test the functionality of the meter and practice reading the scale, do this before the teardown. If you're unfamiliar with reading resistance with an

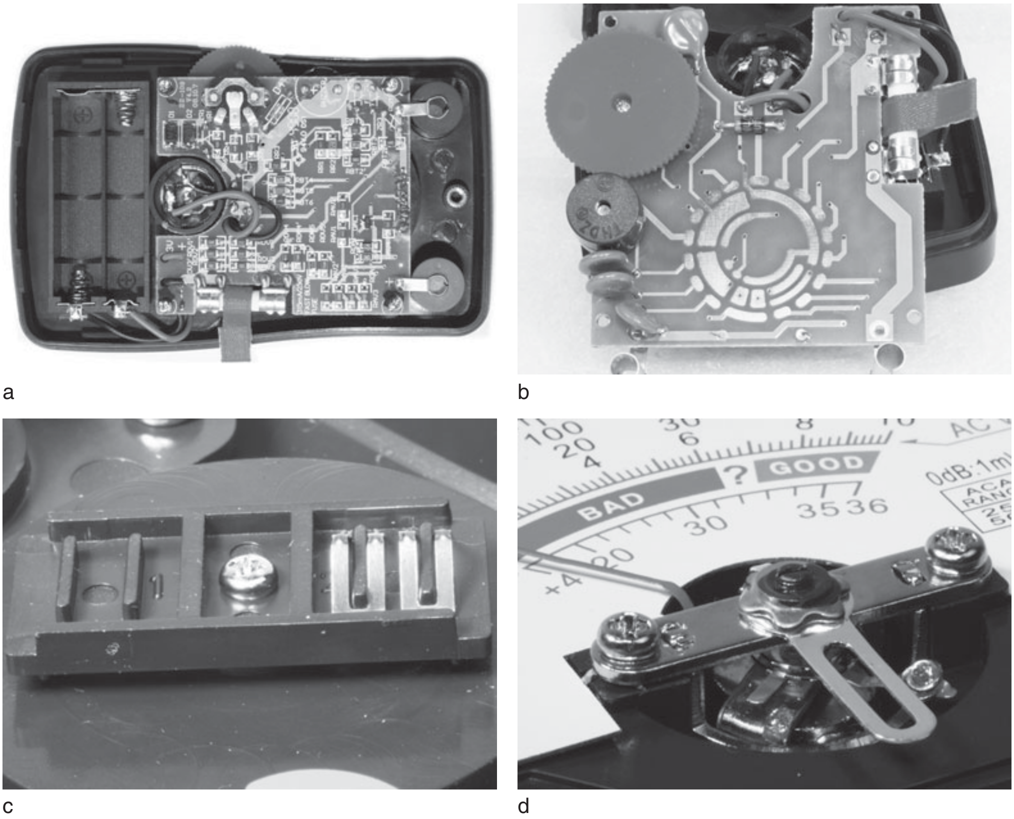

FIGURE 10-3 Teardown sequence

analog ohmmeter, for example, practice with a few resistors of known value. The more you understand the operation of the VOM, the more you'll appreciate the circuitry and components.

Step 1

Remove the back. Place the VOM facedown on a soft cloth or carpet remnant to avoid scratching the clear meter cover.

Step 2

Remove the two AAA cells, as shown in Figure 10-3a. Note the fast-blow fuse, the preponderance of surface-mount components, and the friction-fit positive and negative jack contacts.

Step 3

Release the printed circuit board, shown in Figure 10-3b. Note how the board forms part of the rotary switch. The matching rotary shorting bar is shown in Figure 10-3c.

Step 4

Extract the galvanometer, shown in Figure 10-3d. Pull the battery case to one side and then out, taking care to avoid damaging the needle.

Layout

The components of the VOM are laid out in quadrants. With the back of the VOM open and the empty battery holder oriented away from you, the components associated with DC measurements are primarily in the upper-left area of the board, adjacent to the fast-blow fuse (see Figure 10-4). The resistance components are in the upper-right, adjacent to the ohms zero pot. In the lower-left quadrant are the components associated with the AC measurements, and in the lower-right is the battery-check circuitry.

If you're following along at home with the same model of VOM, it's easy to identify areas devoted to specific meter functions because the components are labeled with circuit keys. For example, the prefix *RAV* is used with resistors in the

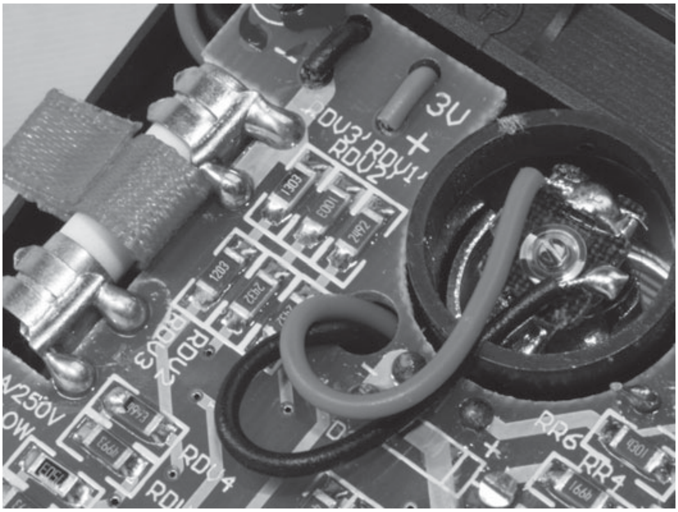

FIGURE 10-4 Component layout, top side, showing SMT resistors, fuse, and galvanometer

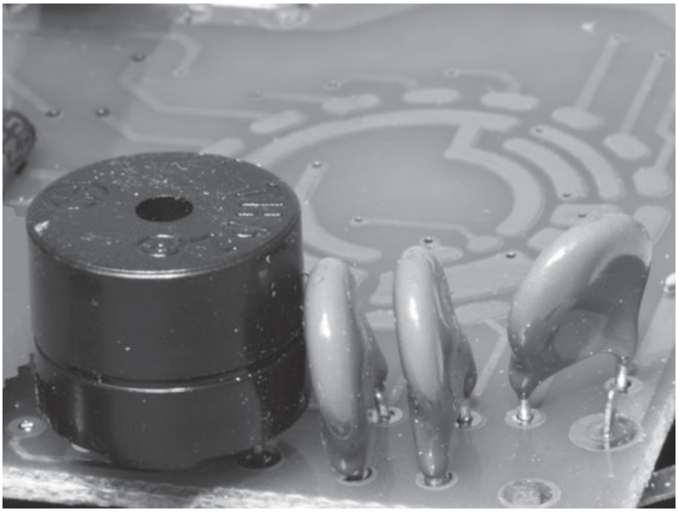

FIGURE 10-5 Component layout, bottom of circuit board

AC voltage section, *RR* for a resistor in the resistance section, *RDV* for a resistor in the DC voltage section, and so on.

As shown in Figure 10-5, the 17-position switch is integral to the bottom of the circuit board. Whereas the top of the board is populated predominantly with surface mount technology (SMT) components, the bottom of the board contains a handful of leaded components, including the large mini-buzzer, and a few leaded MOVs and a diode.

Components

As always, make a habit of looking up component data on the Web. Manufacturer-supplied datasheets often feature example circuit diagrams and other information that may not be evident in the teardown.

Precision Resistors

SMT resistors are the most numerous components in the VOM. The 1/8W resistors are packaged in 1206 cases (3.1mm × 1.5mm), and based on values, are 1 percent metal film. If you're looking for the familiar color bars to identify value and tolerance, you won't find them. Instead, numerical values are stamped on each package. An SMT resistor marked "7501" is "750" + "1" zero, or 7,500Ω. Similarly, a resistor marked "4993" is "499" + "3" zeros, or 499,000Ω.

Although tolerance values aren't marked, they can be easily deduced because resistance tolerance dictates the spacing of values. The lower the tolerance, the greater the spacing between successive values. For example, the progression of values in 5 percent tolerance SMT resistors is 39, 43, 51, and 56KΩ. The progression of 1 percent tolerance resistors includes 39.2 and 49.9KΩ values.

An import feature of SMT resistors is the maximum operating voltage. Common 1206 SMT resistors have a maximum operating voltage of 200VDC. Designers avoid arcing between SMT resistor solder pads and fried components in high-voltage circuits by using several SMT resistors in series.

The linear 15KΩ potentiometer, shown in Figure 10-6, is remarkable in that it's the only variable resistor in the VOM. Better VOMs have several pots that enable the manufacturer and users to calibrate the meter to compensate for variability in initial component values as well as drift in component values over time.

The third type of resistor used in this VOM, shown in Figure 10-7, is a 0.5Ω coil of solid nichrome wire. Nichrome, composed of about 60 percent nickel, with chromium, silicon, and iron for the balance, is commonly used for heater elements and air-wound resistors. Because the resistance per inch is known, creating a low-value resistor is simply a matter of using a specific length of wire where a resistor is needed.

21AWG nichrome wire has a resistance of 0.831Ω/ft. Higher gauge—that is, thinner—wire has greater resistance per foot but lower power-handling capability. The coiled shape does not affect electrical performance from DC to audio frequencies; I measured the inductance of the coil as only 2.4μH. However, a coil is a practical form to store a length of wire within the VOM.

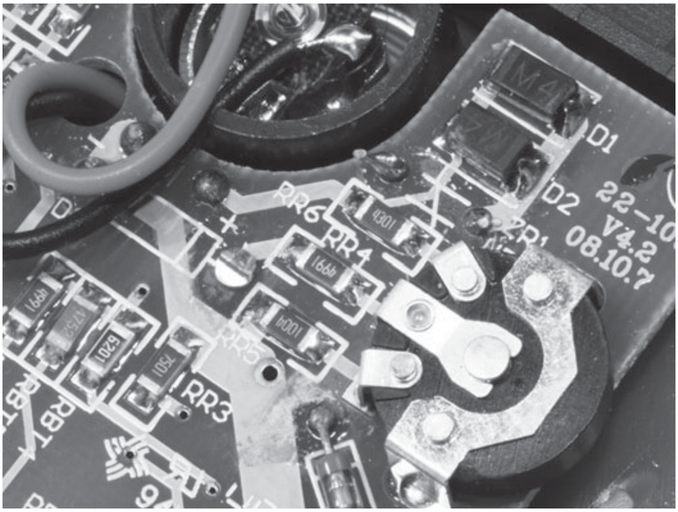

FIGURE 10-6 15KΩ pot and SMT Schottky diodes

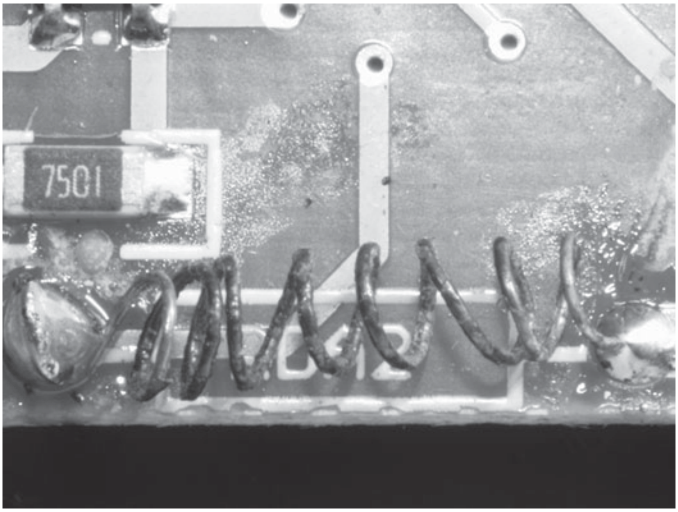

FIGURE 10-7 Nichrome wire resistor

Galvanometer

At the heart of every VOM is a galvanometer. These electromechanical displays rely on a magnetic field acting against spring tension to move a needle by an amount proportional to the current through a coil. The coil and spring of the galvanometer within the VOM are shown in Figure 10-8. This type of galvanometer is referred to as a *D'Arsonval movement*. The downward-pointing metal prongs, which engage with the screw in the meter face, are for adjusting the zero setting of the galvanometer.

As discussed earlier, the characteristics of the galvanometer determine the overall sensitivity of the VOM, and the more sensitive the better. According to my measurements, the resistance of this meter is 725Ω, full-scale voltage is 145mV, and full-scale current is 200μA. I made these measurements with my low-current bench DMM and a voltage divider circuit. If you try to measure the resistance of a galvanometer with an inexpensive VOM, you'll likely fry the galvanometer. The more expensive VOMs and virtually all DMMs use relatively low-current ohmmeter circuits.

Diodes

This VOM employs general-purpose silicon, Schottky, and germanium diodes. A general-purpose 1N4007 silicon diode protects the galvanometer when the VOM is in the off position, and a pair of Schottky SMT diodes protect the galvanometer

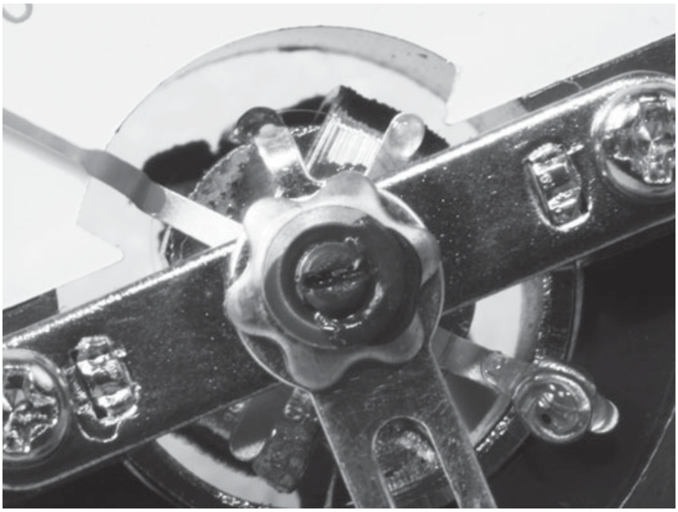

FIGURE 10-8 Galvanometer coil and spring viewed on axis

during measurements. The germanium diode protects the galvanometer against accidental polarity reversal in the battery check mode, and a pair of silicon diodes packaged as a single SMT component rectifies signals for AC voltage measurement.

The 1N4007, one of the few leaded components in the VOM, is rated at 1A and 700V PIV (peak inverse voltage)—the maximum reverse bias voltage that won't result in component failure. This silicon diode, visible at the far left of Figure 10-5, is switched across the galvanometer's positive terminal when the 17-position rotary switch is in the off position. Although the diode protects the galvanometer from damage if you accidentally touch the probes to a live circuit when the VOM is in the off position, it's a poor design guaranteed to sell fuses.

Better—and more expensive—VOMs not only disconnect the input from the leads in the off position, but they also short the leads of the galvanometer together. The short acts as an electromechanical brake that prevents the needle from swinging wildly if you drop the VOM or toss it into a toolkit. Place a clip lead across the galvanometer terminals and see how it dampens the movement of the needle.

The SMT Schottky diodes, shown in the top-right of Figure 10-6, differ from general-purpose silicon diodes in their forward voltage drop and speed of switching. A typical silicon diode, such as the 1N4007, begins conducting when the forward bias reaches about 0.7VDC. Schottky diodes, in comparison, begin conducting with a forward bias as low as 0.2VDC. Moreover, the switching time for a Schottky diode is orders of magnitude shorter than that of a typical silicon diode.

Considering the galvanometer in our VOM requires only 145mV for full-scale deflection, and that even a brief pulse of a significantly higher voltage could easily

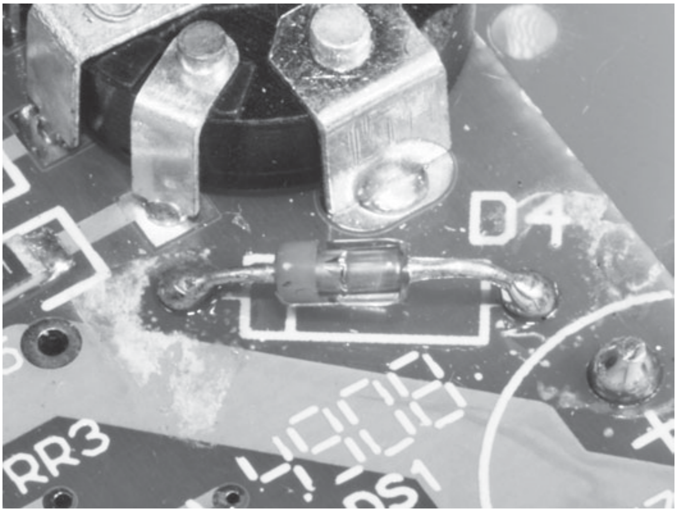

FIGURE 10-9 **Glass-cased germanium diode adjacent to potentiometer**

fry the galvanometer coil, a typical silicon diode would be of little use. A limitation of Schottky diodes is their relatively low PIV ratings. However, because the diodes are wired back-to-back across the galvanometer, the diodes protect each other from reverse voltage spikes.

The glass-cased germanium diode, shown in Figure 10-9, is used in the battery check mode to guard against inevitable reverse-polarity mistakes. As with Schottky diodes, germanium diodes have a lower forward voltage drop than general-purpose silicon diodes—about 0.3VDC. This characteristic is useful, because inserting the germanium diode in series with the voltmeter circuitry will result in a battery check that's only 0.3VDC lower than the actual battery voltage. This is negligible compared to the 0.7VDC drop of a silicon power transistor. The use of a leaded germanium diode over an SMT Schottky diode probably reflects the higher cost of a Schottky diode.

Figure 10-10 shows the dual-diode SMT component (DAC1) used in the AC voltage measurement circuit. The three-terminal device contains two silicon diodes, arranged cathode-to-anode on one end and free cathode and anode at the other. It's not clear why this dual package was used over separate diodes, but component and assembly cost were probably factors.

Rotary Switch

The 17-position rotary switch is formed by a combination of the rotary shorting arm shown in Figure 10-3c and the conductive circuit board pattern shown in Figure 10-11. If you examine the backlit board in Figure 10-11, you can see how the

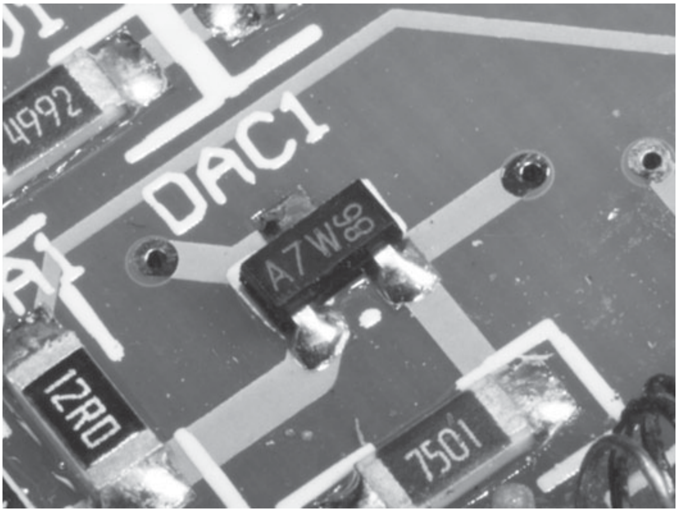

FIGURE 10-10 Dual-diode SMT package, DAC1

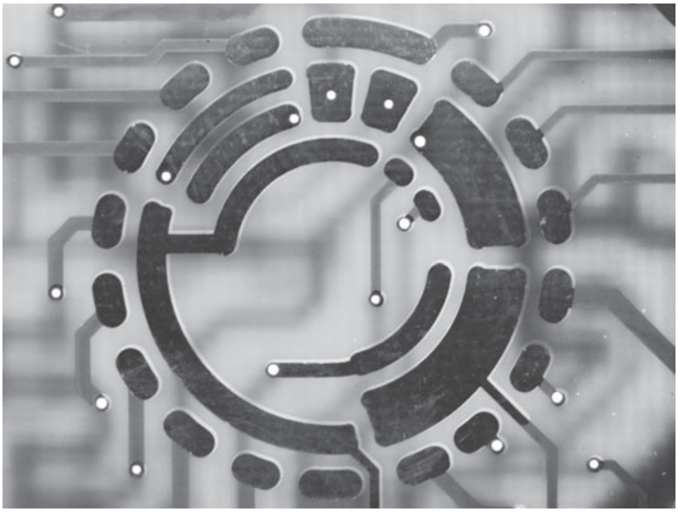

FIGURE 10-11 Rotary switch contacts, back illuminated

copper pads connect through conductive vias to traces on the component side of the board. This form of switch presents a modest obstacle to circuit tracing because the circuit board has to be mounted securely in place for the switch to function. Prepare yourself for removing and reattaching the circuit board a dozen times in the course of tracing the VOM's circuitry.

Battery

The two AAA cells merit discussion because their internal resistance is a component in the resistance mode of the VOM. Lithium, carbon-zinc, alkaline, nickel-metal hydride (NiMH), and nickel-cadmium (NiCd) batteries have different internal resistances that vary with the cell's state of charge. The ohmmeter features a zero adjustment pot to accommodate battery variability.

MOVs

Three unmarked, 0.3-inch diameter MOVs are used in the front ends of the battery check and ohmmeter circuitry to protect the galvanometer from electrical spikes. Recall from preceding chapters that varistors present a high resistance to low voltages and a low resistance once the applied voltage reaches a preset level.

I used a regulated, current-limited power supply to determine that the varistor voltage—the voltage at which current is 1mA—of these MOVs is 20V. At 24V, the current is about 10mA. Based on MOVs of similar diameter and varistor voltage in the Digi-Key catalog (www.digikey.com), the MOVs probably clamp input voltages to less than 30VDC and perhaps 20VAC. (For more information on evaluating MOVs, see Chapter 4 on surge protective devices.)

Fuse

The galvanometer is protected by a 315mA at 250V fast-blow fuse that's designed to fail before the galvanometer's fine copper wiring vaporizes. The speed of the fuse can be attributed to the use of extremely fine wire in the fuse, which unfortunately has a relatively high resistance of 3.9Ω. This resistance is significant at the lower resistance and current ranges of the instrument.

Electromagnetic Buzzer

The miniature electromagnetic buzzer provides audio feedback during continuity tests. The buzzer is a self-contained, high impedance (3.5MΩ) device with a built-in oscillator. According to my measurements, the device draws 20mA at 2V and 30mA at 3V. As shown in Figure 10-12, this buzzer has an outer magnet and an inner electromagnetic element.

The buzzer, visible in Figure 10-5, is the largest component attached to the circuit board. Even though piezoelectric buzzers tend to be more compact and lighter than electromagnetic buzzers, it's difficult to tell the two apart without a teardown. In place of the electromagnetic assembly, piezoelectric buzzers feature a thin disc of piezoelectric material that vibrates when excited by an oscillator built into the base of the buzzer.

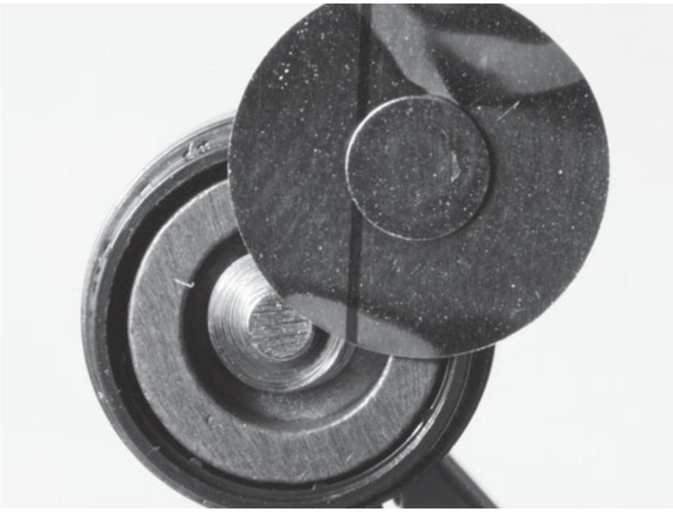

FIGURE 10-12 Electromagnetic buzzer construction

Typical specifications for buzzers include operating voltage and current, operating frequency, and sound pressure level at a specified distance. The sound pressure specification for a typical electromagnetic mini-buzzer is around 75dB at 10cm. Recall that in this context, dB is calculated as follows:

$$\text{dB} = 20 \log_{10} (\text{SPBuzzer}/\text{SPReference})$$

where SPBuzzer is the sound pressure created by the buzzer at a specified distance from the buzzer, and SPReference is the sound pressure reference, equal to the threshold for human hearing. Normally, dB is simply a ratio of two voltages, power, sound pressure, or other measurement. In this instance, dB is an absolute value.

A sound pressure level of 75dB is enough to cause you permanent hearing loss—not to mention headaches—with long-term exposure. Because the sound pressure level is inversely proportional to the distance from the sound source, you can save your hearing by not working directly over the buzzer. Simply moving the mini-buzzer 8 inches away from you would reduce the sound pressure level to a more bearable 69dB.

In case you didn't catch that sleight of hand, recall that doubling the power is equivalent to an increase of 3dB, and that doubling a voltage is equivalent to a 6dB increase. You can think of sound pressure as a voltage—both are pressures of sorts—and so the 6dB relationship applies. And don't forget that dB gain and loss are symmetrical, in that doubling or reducing the sound pressure by 50 percent involves a gain or loss of 6dB.

How It Works

The VOM is a combination of instruments that share a case, selector switch, and function-specific components. Let's examine each instrument function independently.

DC Voltmeter

The ideal DC voltmeter presents a high impedance to the circuit under test—the higher the better. Because a voltmeter is placed across a device or from a circuit element to ground, current flowing through the voltmeter can disrupt the circuit. The sensitivity of the galvanometer determines the current requirement for the voltmeter, which is another way of saying the instrument's input impedance. A 10M resistor draws considerably less current than a 10Ω resistor when placed across a signal source.

The galvanometer in our voltmeter requires 200μA for full deflection. In a 200W audio power amplifier circuit, a measurement that requires up to 200μA may be insignificant. However, in a 200mW preamp, a 200μA drain could shut down the circuit. Some analog VOM designs use a FET front end to minimize the current drawn from the circuit, but let's consider the simple case of a DC voltmeter without active elements.

Figure 10-13 shows the simplified DC voltmeter configuration of the VOM. The circuit consists of the 3.9Ω fuse, one or more series resistors, a pair of Schottky diodes, and the galvanometer. 725Ω is the equivalent resistance of the copper wire coil in the galvanometer. The multiplier resistors R1, R2, and R3, together with the resistances of the fuse and galvanometer, determine the voltage required for full-scale deflection.

As an exercise, let's determine the appropriate multiplier resistance for a 500VDC scale with our 200μA galvanometer. First, notice that the values of the switched multiplier resistors—R1, R2, and R3—are additive. The highest voltage range consists of R1, R2, and R3. The lowest range includes only R3. So keep in mind

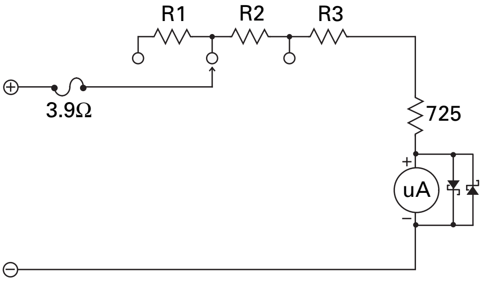

FIGURE 10-13 Simplified schematic of DC voltmeter

that when we solve for total resistance at the highest range, we constrain the possible values of R2 and R3. Let's assume that at the 500VDC range, the 3.9Ω contributed by the fuse and 750KΩ resistance of the galvanometer are insignificant.

Now, to Ohm's law:

$$R1 + R2 + R3 = \text{VFull-Scale/IFull-Scale}$$

$$R1 + R2 + R3 = 500\text{V}/200\mu\text{A}$$

$$R1 + R2 + R3 = 2.5\text{M}\Omega$$

Let's repeat the exercise with a full-scale voltage of 5VDC, and assign this to the middle switch position, which includes R2 and R4, as shown in Figure 10-13. Let's also use the 725Ω galvanometer resistance in our calculation.

$$R2 + R3 = \text{VFull-Scale/IFull-Scale} - 725$$

$$R2 + R3 = 5\text{V}/200\mu\text{A} - 725$$

$$R2 + R3 = 25{,}000 - 725$$

$$R2 + R3 = 24{,}275\Omega$$

Finally, let's assign the third switch position, which inserts R3 in series with the galvanometer. Let's use an arbitrarily low full-scale value of 0.5VDC to illustrate the significance of the resistance of the galvanometer, which is included in the calculation. Repeating the exercise:

$$R3 = \text{VFull-Scale/IFull-Scale} - 725$$

$$R3 = 0.5/200\mu\text{A} - 725$$

$$R3 = 2500 - 725$$

$$R3 = 1775\Omega$$

We can now determine the values of the three multiplier resistors, R1, R2, and R3, for full-scale values of 0.5, 5, and 500 VDC. We've determined the value of the 0.5VDC multiplier, R3:

$$R3 = 1775\Omega$$

We can calculate the value of the 5VDC scale multiplier, R2:

$$R2 + R3 = 24{,}275\Omega$$

$$R2 = 24{,}275 - R3$$

$$R2 = 24{,}275 - 1775$$

$$R2 = 22{,}500\Omega$$

Similarly, for the 500VDC scale multiplier, R1:

$$R1 + R2 + R3 = 2{,}500{,}000$$

$$R1 = 2{,}500{,}000 - R2 - R3$$

$$R1 = 2{,}500{,}000 - 22{,}500$$

$$R1 \approx 2{,}500{,}000\Omega$$

This exercise highlights the significance of resistor tolerances and the relative impact of fixed internal component values, such as the fuse and galvanometer, on the selection of multiplier resistance values.

One note in the overall design is that R1, R2, and R3 could be switched in circuit independently. The advantage of using the current design with several resistors in series with higher voltages is that the overall breakdown voltage is increased. Assuming a breakdown voltage of 200VDC per SMT resistor, three or more SMT resistors should be used in series on the 500VDC range of the VOM.

Battery Tester

The battery tester function of the VOM is essentially the DC voltmeter described previously with multiplier resistors selected for 9V and 1.5V full-scale, together with a dedicated, BAD, (unknown), and GOOD scale on the meter face (see Figure 10-2). In addition, the battery test circuit includes MOVs at each of the three switch positions and a germanium diode in series with the galvanometer. The MOVs absorb overvoltage spikes and the diode protects the galvanometer by insuring proper polarity of the battery under test. A powerful negative swing caused by testing a battery with the wrong lead polarity could permanently damage the galvanometer.

AC Voltmeter

If you compare Figures 10-13 and 10-14, you'll note several differences between the previously discussed DC voltmeter and the AC voltmeter. The most significant difference is a half-wave rectifier in the AC voltmeter used to convert an AC signal to a DC level. In addition, whereas the galvanometer in the DC voltmeter carries all current from the circuit under test, in the AC voltmeter, a proportion of the total current is shunted around the galvanometer.

When you measure a voltage with the AC voltmeter, current flows through the 3.9Ω resistor, through R3, and then to D1 and D2. When the top lead is negative, D1 shunts the current to the negative lead, bypassing the galvanometer. When the top lead is positive, current flows through D2 and the 7.5KΩ resistor in parallel with the galvanometer to the bottom lead.

Why not simply use one diode in series with the galvanometer? Because diodes are imperfect, and they allow a small leakage current when reverse-biased. Without

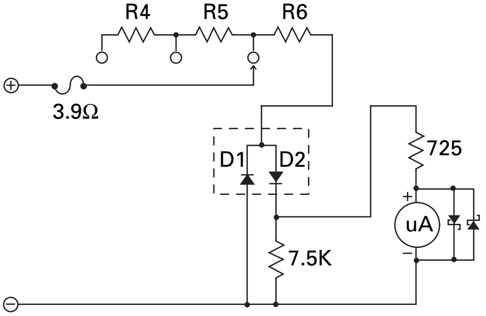

FIGURE 10-14 Simplified schematic of AC voltmeter

D1, some current would flow through the galvanometer when the top lead is negative. In a typical power supply circuit, a few hundred microamperes of leakage current are insignificant. In this circuit, that's enough current to vaporize the galvanometer.

Because the circuit uses half-wave rectification, the DC voltage at the junction of D2 and the 7.5KΩ resistor is a little less than half of the RMS (root mean square) AC applied across D1 and D2. For example, if we applied 120VAC line voltage across D1 and D2, we'd measure a little less than 60VDC across the 7.5KΩ resistor. The actual value of the voltage across the 7.5KΩ resistor, and therefore the galvanometer, isn't that important, as long as the meter face is calibrated to reflect the actual value. For example, if we ignore the 7.5KΩ resistor, we have a basic DC voltmeter with an error of about a half of a volt introduced by the forward bias voltage drop of D2. All we have to do is adjust the values of R4–R6 so that the desired full-scale voltage results in full-scale deflection of the galvanometer.

Are you curious about the purpose of the 7.5KΩ resistor? If you remove the resistor, you'll see that the voltmeter still functions, but the reading is a little off. Let's see by how much. We know full-scale deflection of the galvanometer requires 200μA, which is equivalent to 145mV across the 725Ω device. This voltage would result in a current flow through the 7.5KΩ resistor of the following:

$$I = V/R$$

$$I = 145\text{mV}/7.5\text{K}\Omega$$

$$I = 0.145\text{V}/7500\Omega$$

$$I = 19\mu\text{A}$$

So about 10 percent of the current through D2 passes through the 7.5KΩ resistor, and 90 percent flows through the galvanometer. Why would the designers intentionally decrease the effective sensitivity of the voltmeter? One reason would be to use standard 1 percent resistor values for R4–R6. Recall that we can't use

R1–R3 from the DC voltmeter circuit because of the voltage drop from D2 and the half-wave rectification equivalent of a little less than half (approximately 45 percent) of the RMS AC voltage. Better meters use variable resistors in series with the galvanometer circuit to apply the proper correction. Can you think of another reason for the 7.5KΩ resistor?

Ammeter

The ideal ammeter is a low-resistance device that is placed in series with the circuit under test. Our ammeter circuit, shown in Figure 10-15, illustrates many of the challenges associated with achieving a low-resistance instrument. First, let's ignore the 0.5, 12, and 20Ω shunt resistors and focus on the leads, fuse, and galvanometer. The maximum current we can measure with our bare galvanometer is 200μA. At this level of current, the resistance of the test leads and fuse are insignificant.

Now let's say we want a full-scale current of 250mA. We have to shunt all but 200μA of that current through a low-resistance device in parallel with the galvanometer, as shown in Figure 10-15. But how do we calculate the value of the shunt resistance? The easiest way is to recall that full-scale deflection occurs with 145mV across the galvanometer. So we have this:

RShunt = VFull-Scale/IFull-Scale

RShunt = 145mV/250mA

RShunt = 0.58Ω

I measured the resistance of the nichrome wire shunt used in the 250mA range as 0.5Ω, which is close to the calculated value.

Similarly, for 25mA full-scale, the shunt resistance should be as follows:

RShunt = 145mV/25mA

RShunt = 5.8Ω

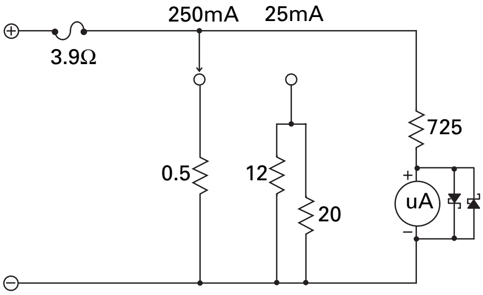

FIGURE 10-15 Simplified schematic of ammeter

Because 5.8Ω isn't a standard 1 percent SMT resistor value, the designers used a 12Ω and 20Ω resistor in parallel for an equivalent resistance of

$$RShunt = (12\Omega \times 20\Omega)/(12\Omega + 20\Omega)$$

$$RShunt = 7.5\Omega$$

Another reason to use multiple resistors in parallel in a current shunt is to increase the dissipation capabilities of the shunt. Two approximately equal 1/8W resistors can dissipate 1/4W when connected in parallel.

Notice that in both circuits, I've ignored the contribution of the 750Ω equivalent resistance of the galvanometer and assumed that most of the current flows through the shunt resistor. Even at 25mA full-scale, 200μA or 0.2mA is insignificant. The difference between the calculated and actual value of the 25mA shunt used in our VOM is significant, however. Again, the rationale for using approximate values is probably a reflection of component cost.

Another significant aspect of the circuit is the resistance of the fuse. On the 250mA scale, the ammeter inserts about 4.5Ω in series with the circuit under test. Assuming a 12V circuit, at 250mA, the voltage drop across the VOM is 4.5Ω × 0.25A = 1.1V. This represents about a 10 percent change in the operating voltage of the circuit, enough to misrepresent the actual current drawn at the full operating voltage. A transistor amplifier circuit might draw less current at a lower operating voltage, but a small DC motor might draw more current, depending on its electrical configuration and mechanical load.

Ohmmeter

The ohmmeter component of our VOM is a little more complicated than either the voltmeter or ammeter, as illustrated by the simplified schematic in Figure 10-16. However, if we ignore the protective elements of the circuit, we have a basic DC ammeter circuit with a built-in current source. The diode D1 protects the meter from reverse voltage resulting from an improperly installed battery—at the cost of a blown fuse. In addition, the MOV protects the meter from inevitable voltage spikes resulting from, for example, checking the in-circuit resistance when there are partially discharged capacitors.

To get a handle on circuit operation, let's make some simplifying assumptions. We can assume that the leads are shorted together, so that R = 0Ω. We can ignore the contributions of the 1MΩ resistor and 3.9Ω fuse. We can assume that the 15KΩ pot is centered and presents a 7.5KΩ resistance in series with the galvanometer. With these assumptions, we have a 3VDC source and a 47Ω resistor in parallel with the galvanometer circuit with a resistance of 4.9KΩ + 7.5KΩ + 4.3KΩ + 725Ω = 12,525Ω. In this configuration, current through the galvanometer is as follows:

$$I = V/R$$

$$I = 3V/12{,}525\Omega$$

$$I = 240mA$$

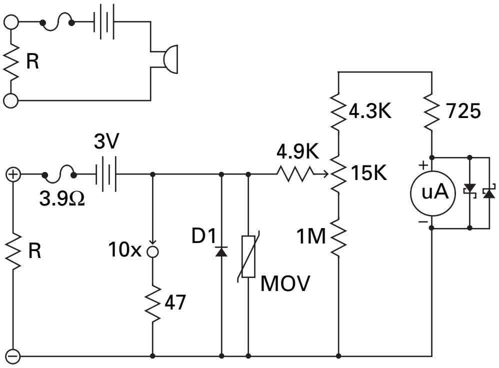

FIGURE 10-16 Simplified schematic of ohmmeter in X10 position (center) and continuity test circuit (top-left)

This is greater than the full-scale value of 200µA, but about right. We can increase the series resistance contributed by the 15KΩ pot so that the total resistance is 3V/200µA = 15KΩ. In actual use, you would adjust the pot to compensate for the contribution of the fuse resistance, the actual battery voltage, as well as the internal resistance of the battery. Recall that the internal resistance of the battery affects current flow according to Ohm's law, and that this resistance depends on the charge status of the battery and the battery chemistry. Lithium batteries have a lower internal series resistance than carbon-zinc batteries, for example.

Now let's see what happens when R = 100Ω. If you look at the meter face, you can see that, on the 10X setting, 100Ω corresponds to about 30 percent deflection of the galvanometer, or a current of about 65µA through the galvanometer circuit. Keeping this in mind, let's return to the schematic in Figure 10-16. Now we have a 3V source, a series resistance of 100Ω, and the 47Ω resistor in parallel with the 15KΩ equivalent resistance of the galvanometer circuit. What's the current through the galvanometer?

We can ignore the 15KΩ resistance and determine the voltage across the 47Ω resistor in series with the 100Ω resistor. Let's start by determining the total current and then calculating the voltage drop across the 47Ω resistor:

$$I = 3V/(100\Omega + 47\Omega)$$

$$I = 20mA$$

Now, let's calculate the voltage across the 47Ω resistor:

$$V = I \times R$$

$$V = 20mA \times 47\Omega$$

$$V = 0.94V$$

Assuming an equivalent series resistance of 15KΩ, current through the galvanometer is

$$I = V/R$$

$$I = 0.94V/15{,}000\Omega$$

$$I = 63\mu A$$

This is fairly close to our original guesstimate, which was based on the meter face markings.

The current through the galvanometer is a function of the ratio of the unknown resistance to the resistance of the 47Ω resistor, and not the absolute value of the unknown resistance. This ratio measurement explains why the scale is compressed at the higher resistance levels. In terms of galvanometric current, a change in the resistance ratio from 1:1 to 2:1 is significant. However, the difference between 10:1 and 11:1 is barely perceptible.

To illustrate, let's use the meter face to estimate galvanometer current. On the X10 setting, testing a 47Ω resistor should place the needle midpoint in the scale, or about 100μA. With a 100Ω resistor, the deflection is about a third, corresponding to 200μA/3, or about 66μA. At 500Ω, deflection is only about 10 percent of full-scale, or 20μA, and so on. Above 500Ω, it's difficult to discern deflection and corresponding galvanometer current.

Keep in mind that the value of the shunt resistor depends on the meter setting. At X100 and X1K, the shunt resistors are 523Ω and 7.5KΩ, respectively. Using a shunt resistor near the value of the unknown resistance results in a galvanometer deflection near mid-scale, away from the compressed, difficult-to-read left edge of the scale. You can use this feature to your advantage if you work with components with a specific range of resistances.

For example, let's say you use your VOM on an assembly line to test guitar pickups, which should have a resistance of 10–12K. If you replace the 47Ω resistor used in the X10 setting with a 20KΩ, 0.1 percent tolerance resistor, you could test coil resistance with the needle in the right third of the scale where small differences in resistance are discernable. You could mark the meter face for correct resistance values, or simply make a note to yourself that components should test within, for example, 1.5 and 2.0 on the existing resistance scale.

Audible Continuity Tester

The audible continuity test mode provides a fast, qualitative test for resistance. As illustrated in the upper-left of Figure 10-16, the unknown resistance simply completes the circuit comprising the 3.9Ω fuse, the battery, and the buzzer. My measurements indicate that the buzzer begins working when it draws about 20mA—which translates to a maximum of 3V/20mA = 150Ω between the test leads. This contrasts with the stated minimum resistance of 300Ω. The continuity tester is great for checking grounds and circuits that don't involve sensitive electronic components.

This VOM is fine for testing diodes, capacitors, resistors, and some fuses, but not for components such as a 200μA galvanometer. The continuity tester essentially places the 3V battery across the component. In our galvanometer, this would result in a brief current of 3V/750Ω = 4000µA. The X10 and X100 settings of the ohmmeter would result in essentially the same current flow. Even the lowest current setting, X1K, would result in a current of 3V/(7500Ω + 725Ω) = 365µA through the galvanometer—more than enough to fry its coil. The ohmmeter function, like the other functions on this VOM, is valuable and useful on and off the workbench. As with any test instrument, you simply have to appreciate the strengths and limitations of the tool.

Mods

You can improve upon the 5 percent full-scale accuracy of the VOM by replacing the 1 percent resistors with 0.1 percent resistors. At about 25 cents per SMT resistor, it should cost about $8 to replace all 32 SMT resistors. In return for your efforts, you might realize a 1 percent improvement in accuracy.

A more practical mod, or perhaps application, of the VOM is to create a high-current meter. If you want to measure currents up to 10A with your VOM, you can add an external shunt resistor and measure the voltage across the shunt using the 10VDC scale. So what's the value of the shunt resistor? Assuming we use the VOMs 10VDC scale, from Ohm's law: R = 10V/10A = 1Ω. The minimum power dissipation required is P = 10V × 10A = 100W.

So if we wanted to use the 10V scale, we'd either need a 1Ω, 100W resistor or some combination of lower wattage, less expensive resistors. In practice, a 150W resistor or resistor network is more practical because of the safety margin. Figure 10-17 shows an external current shunt composed of 20W resistors connected in parallel that I picked up for $5 on eBay.

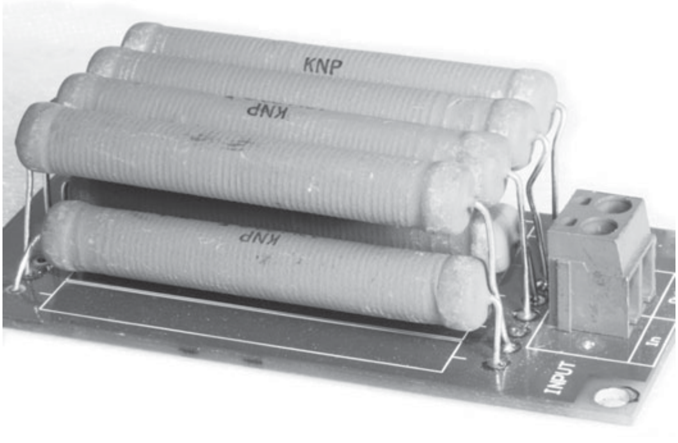

FIGURE 10-17 External current shunt

If you decide to build your own high-current shunt, you should buy a few more resistors than necessary. Because of resistor tolerance, your final shunt could be off by a tenth of an ohm or more. The solution is to use your ohmmeter to select the combination of resistors that will provide a shunt as close as possible to 1Ω.

The power requirements for the shunt illustrate the advantage of a more sensitive VOM. For example, a 20,000Ω/V Triplet 310 has a 3VDC scale. Even though you'd have to multiply the meter reading in volts by 3.33 to get the actual current in amps, you could get by with a 0.33Ω resistor: $R = 3V/10A = 0.33\Omega$. Unfortunately, the power rating would be significant: $P = 3V \times 10A = 30W$. However, with a lower value external shunt, the power requirement is lower as well. This translates to cost savings and, more importantly, to less of an impact on the circuit under test.

This page intentionally left blank

Chapter 11

Laser-guided Sonic Distance Measurer

I own three tape measures—a figure that's probably well below the national average. I have a compact, 6-foot tape measure near my workbench to measure chassis layouts, to get a quick estimate of circuit board sizes, and, in general, to work with small objects. Then there's the 100-foot monster tape, with a built-in crank handle, for measuring outdoor cable runs, spacing between trees, and other odd surveying tasks. Finally, I have a 25-foot tape that I use for measuring carpet, walls, and large interior spaces that the 6-foot tape just can't cover. The problem with the 25-foot and 100-foot tapes is that I generally need someone to hold the other end. Tapes droop unless they're supported by the floor or they simply dislodge because there's no one to ensure that the free end is secured.

Several technological alternatives to the inexpensive but sometimes limited tape measure are available, including handheld ultrasonic and laser distance measuring devices. Ultrasonic distance measurers, from companies such as Zircon (see Figure 11-1), Strait-Line, RadioShack, Mastech, and Black & Decker, are relatively affordable at \$10 to \$80. Sonic measurers, also called ultrasonic tape measures, are based on an ultrasonic transceiver that emits an ultrasonic burst and then times the return. Most of these sonar-like devices have a visible laser for a pointing aid. Sonic measurers for home use operate around 40kHz and generally offer accuracy within a few inches at the maximum range of 50 feet.

Laser distance measuring devices time the round-trip time of laser light instead of an ultrasonic signal. As a result, their maximum range is typically 250 feet when used with a special reflective target, and accuracy is significantly better than with an ultrasonic unit. Prices for nonprofessional laser units from companies including Bosch, Mastech, and Strait-Line range from about \$50 to \$200, however. Too expensive for my tool chest, especially when a good 100-foot tape measure sells for about \$20.

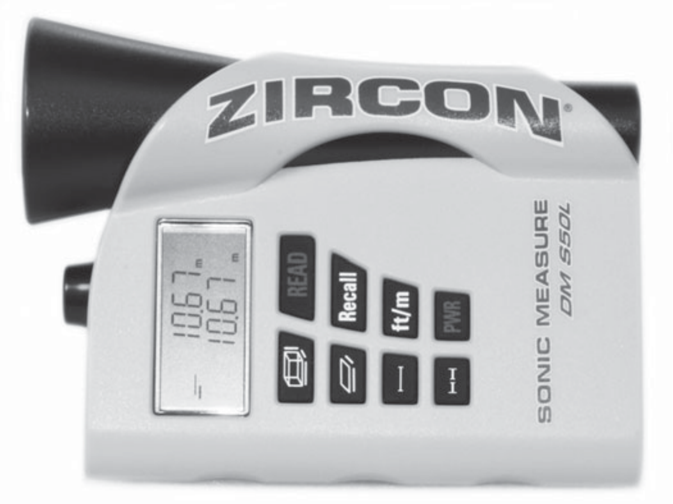

FIGURE 11-1 Zircon sonic measurer

In this chapter, I'll tear down the Chinese-manufactured Zircon Sonic Measure DM S50L, shown in Figure 11-1. You can buy this laser-guided ultrasonic distance measurer, hereafter also referred to as the "sonic measurer," from Amazon for about $40. This teardown provides an excellent means of exploring the physics of sound and an example of interfacing analog sensors and effectors with a microcontroller. At the end of the chapter, I'll suggest a few ways to repurpose some of the circuitry.

Highlights

This is a fun, relatively short teardown. Although the circuit board features a mysterious processor hidden under a blob of black epoxy, we have plenty of discrete components to examine, including an ultrasonic transducer, a thermistor, two voltage regulators, and a visible laser diode assembly.

During the teardown, note the following:

- The layout of the circuit board, which is primarily populated with SMT components
- The bulk of the laser diode module
- The ultrasonic transceiver driver circuitry
- The power regulation and distribution system

Specifications

The specifications of the Zircon DM S50L, extracted from the one-page user guide, the Zircon web site (www.zircon.com), and the back of the unit, include the following:

- **Range:** 2–50 feet
- **Resolution:** 1 inch
- **Accuracy:** ±0.5 percent of distance ±1 digit
- **Operating temperature:** 20°F–120°F
- **Operating humidity:** 0–80 percent (noncondensing)
- **Functions:** Area, volume, addition, memory recall
- **Power requirements**: 9V alkaline battery for 8 hours of continuous use; auto-off after 30 seconds
- **Aperture angle:** ±5 degrees from center
- **Laser:** Class IIIa, <5mW at 630–670nm; compliant with 21 CFR parts 1040 10 and 11
- **Dimensions:** 4 3/4 × 3 1/2 × 1 2/3 inches (HWD)
- **Weight:** 4.7 ounces

These specifications are useful in comparing units from different manufacturers and for insight into the inner workings of the sonic measurer. Let's consider the more significant specifications of the unit in more detail.

Range

The range of the sonic measurer is a function of the properties of ultrasound, the design of the ultrasound transducer, the receiver circuit, and the available processing power. The minimum distance reflects limitations in the processor and transceiver design, which in turn reflect practical cost constraints. The shorter the distance from transceiver to the wall or other object, the quicker the transceiver and processor must respond to the echo. Because the ultrasonic transducer operates in a simplex mode—it can either send or receive, but not both simultaneously—a refractory period follows transmission, during which the transducer cannot receive an echo.

However, there's nothing inherent in the design of a single-transducer transceiver that prevents a minimum range of less than 2 feet. For example, a single-element ultrasonic rangefinder that's popular in the robotics community, the Devantech SRF02 Sensor (www.acroname.com), has a minimum distance of 6 inches. The SRF02 achieves this minimum distance through fast recovery and onboard signal processing. However, the cost of the SRF02 is about $25.

Another approach to minimizing the working distance of an ultrasonic measurer is to use separate transmitter and receiver transducers, separated horizontally by a centimeter or two. For example, the Devantech SRF04 Ranger offers a minimum measurement distance of about 1 inch, in part because the receiver element can be active at all times. However, while a dual-transducer transceiver may be fine for a large robot, there's no space on a handheld unit for two horn assemblies—one for transmit and one for receive.

The maximum range of an ultrasonic measurer is a function of the effective radiated power of the transmitted ultrasound chirps, the sensitivity and selectivity of the receiver, the nature of the target and echo, and the operating frequency. The greater the effective radiated power, the greater the maximum range, everything else held constant. Effective radiated power can be increased by increasing power to the transducer, by using a more efficient transducer, and by employing a horn or acoustic antenna that produces a narrower, more focused acoustic beam. The more focused the transmitted signal, the greater the energy imparted to the target and reflected. A solid, flat, broad surface, perpendicular to the horn, produces the best echoes.

The greater the sensitivity of the receiver and the more focused or selective the horn assembly, the more likely an echo will be detected, assuming perfect target alignment. However, because the target may not be precisely perpendicular to the transmitted signal, a horn that's too narrowly focused may reject the echo. So the directivity of the horn must be low enough to allow the receipt of off-axis signals. The catch is that off-axis reflections from objects other than the intended target can also make it to the receiver if the directivity is too low. Because the transmitter and receiver share the same horn, the directivity of the horn is a compromise between narrowly focusing the transmitted ultrasound signal presenting an aperture wide enough to capture an off-axis echo.

The frequency of the ultrasound signal significantly constrains the maximum distance because of differences in the propagation and of sound through the air as a function of frequency. For example, ultrasonic rangefinders operating at 235kHz have a number of advantages over rangefinders operating at 40kHz, such as a potentially tighter transmitted acoustic beam. However, whereas a sonic measurer based on 40kHz sound can measure up to 50 feet or more, an equivalently configured 235kHz rangefinder, such as the Devantech SRF235 (www.acroname.com), is limited to about 47 inches, because the higher frequency signal is attenuated by the air.

Resolution

If you read the Volt-Ohm-Meter (VOM) teardown in Chapter 10, you know that *resolution* refers to the number of significant digits that can be displayed by the meter. Similarly, resolution in this case is simply the finest level of granularity supported by the sonic measurer—about 1 inch. Resolution is a function of the numeric capabilities and programming of the microcontroller and the number of digits supported by the LCD module.

Accuracy

Also covered in the VOM teardown, accuracy reflects the trueness of the measurement. It's a function of the algorithms used in the microcontroller, rounding errors due to hardware limitations, the accuracy of the microcontroller's timebase, provision for environmental variability, the accuracy of the factory calibration, the nature of the target, and user error. In addition, because components shift value with age, time since calibration is also a factor.

At 50 feet, the claimed accuracy of the sonic measurer is within ±0.5 percent of 50 feet ±1 digit, or ±4 inches. Don't confuse more digits after the decimal point or higher resolution with greater accuracy.

Operating Temperature

All electronic components and equipment are designed to operate within a specific temperature range. At temperature extremes, LCD panels pop or crack, batteries leak or become inoperative, and microprocessors lock up. The operating temperature range is particularly important in the sonic measurer because the speed of sound in air is dependent on the temperature. As you'll see later, the unit includes circuitry specifically intended to monitor the air temperature and adjust the calculations for distance accordingly.

Humidity

Water and electronics generally don't mix well. Most open-air electronics have a limited maximum ambient humidity rating, above which problems with arcing, leakage, and corrosion can occur. Most electronics can't tolerate condensing humidity—the equivalent of dew on components and circuit traces—for obvious reasons. On the other extreme, very low humidity presents a problem because electrostatic discharge occurs more readily.

In addition, as with temperature, the level of humidity affects the propagation of sound in the air. This inexpensive sonic measurer has no provision for monitoring and adjusting the calculation of distance to account for variations in humidity, however.

Aperture Angle

As discussed, the aperture angle of the transmit/receive horn affects the range and accuracy of the measurements. The narrower the aperture angle, the greater the potential maximum range, assuming a perfectly perpendicular target. The larger the aperture angle, the more likely an echo will be recovered, as well as reflections and other noise that could result in erroneous readings.

Laser Specifications

The Code of Federal Regulations (CFR) Title 21, Part 1040, establishes the safety and performance standards for light-emitting products. Ratings range from Class I (no hazard) to Class IV (an acute hazard to the skin and eyes from direct or scattered radiation). This sonic measurer employs a Class IIIa laser, which is considered an acute viewing hazard. Output of the 630–670nm orange-red laser is limited to 5mW. The increasingly popular green lasers, in comparison, operate between 520 and 565nm, and common infrared lasers generally operate at a wavelength of greater than 800nm.

Caution Don't look directly at the beam or its reflection in a mirror. Your retinas can be permanently scarred.

Operation

Operating the sonic measurer is a simple matter of depressing the soft elastomeric PWR (power) button, securing the unit against a wall or other fixed position, and pressing the READ button. The unit first energizes the laser for a couple seconds and then transmits a few rapid pulses of ultrasound or chirps. If all is well, the reading, in either feet or meters, is displayed on the LCD about a second after the chirps have ceased. The distance measurement is made from the piezoelectric transducer at the back of the unit to the wall or other large reflective target.

If the wall or other object within range isn't flat, isn't perpendicular to the sonic measurer, isn't fully covered by the laser pattern (which approximates a ±5-degree cone), or significantly absorbs or disperses the energy of the ultrasonic chirps, the unit will display an error message, "ERR." If the target is within range and is a flat, reflective wall, perpendicular to the unit, the sonic measurer will transmit two chirps and then display the distance. Otherwise, the unit will transmit up to about a dozen chirps before returning either the distance measure or an error message.

In addition to taking basic distance measurements, the unit supports calculations that are accessed through the area, volume, double segment, and Recall buttons. The area and double segment functions each take two consecutive measures, A and B, and compute the surface area (A × B) and total distance (A + B), respectively. The volume function takes the result of three consecutive measures, A, B, and C, and computes the volume (A × B × C). Pushing the Recall button lets you step back through the values used for each calculation and repeat a measurement. For example, after measuring the volume of a room, you can review the three measures and repeat the second measure (B). The unit will recalculate and display the volume.

Pressing the PWR button for a second time turns off the unit. The unit will also shut down automatically after 30 seconds of nonactivity.

Teardown

The teardown, illustrated in Figure 11-2, requires about 5 minutes, including extraction of the ultrasound transducer.

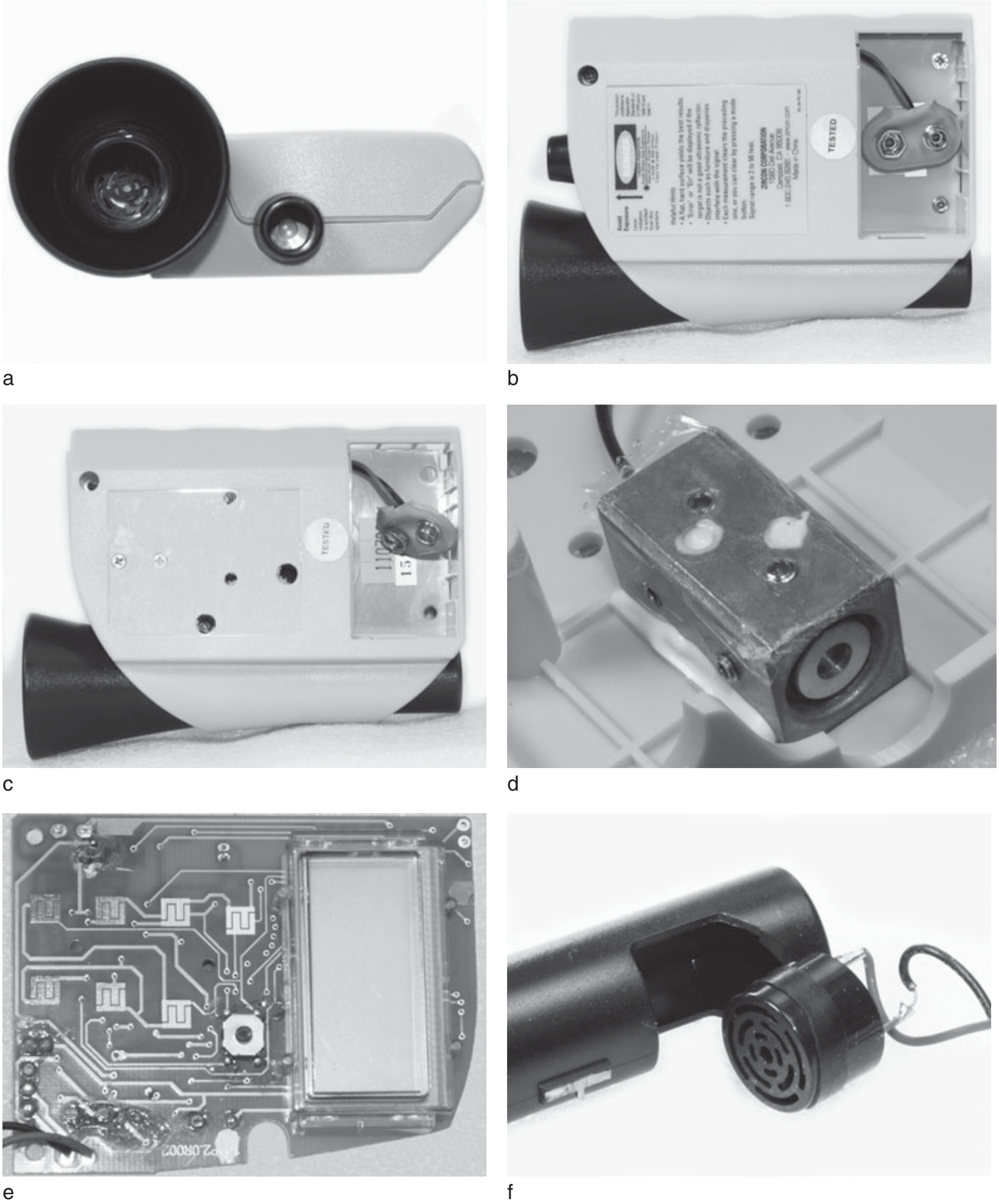

FIGURE 11-2 Teardown sequence

Tools and Instruments

You'll need a sharp knife and a small Phillips-head screwdriver to crack the case. You'll also need an illuminated magnifier and a multimeter with sharp probes to trace the circuitry. The circuitry isn't dense, but the markings on many of the SMT components are impossible to read with the naked eye. You may also need a rough cotton cloth and some rubbing alcohol to remove the glaze over some of the active devices so that you can read their markings. In particular, the coating over the 14-pin IC in my unit was so thick that I couldn't make out a single marking. You may also need a forceps to remove glue between some components and the board.

If your goal is to harvest the SMT components, you'll need a fine-tipped soldering iron and forceps. If you have one, a hot air pen will make the harvest much easier.

Step by Step

If you're following along at home with a sonic measurer, the only thing you have to watch out for is the relatively fragile thermistor bead that's attached to the plastic ultrasound horn. I discovered the hard way that the leads are wispy and easily broken. If you do break a lead, however, don't worry. A bit of solder and tape and it'll be as good as new.

Step 1

Identify the external components. Before you crack the case, identify the ultrasonic transducer in the base of the horn and the emitter of the laser module, which is flush with the top of the unit, as shown in Figure 11-2a. Next, examine the junction of the horn and the case, at the top of the unit. You should see a small, amber bead, about the size of a grain of rice, just inside the seam. This is the epoxy-encapsulated thermistor used to measure the ambient air temperature.

Step 2

Remove the back cover. Remove the battery cover and battery to expose the two Phillips-head screws in the base of the battery compartment. Remove the screws. Next, with the unit oriented as shown in Figure 11-2b, remove the screw in the upper-left corner of the yellow plastic case. If you try to crack the case at this point, you'll encounter stiff resistance.

There are two approaches to dealing with a case that won't open easily. The first is to assume that there's a hidden hinge, clasp, or glue, and apply force until it gives. The second is to search for screws hidden under labels, decals, and rubber feet.

In this instance, the correct approach is to search for hidden screws. To get at the remaining two screws securing the cover, use a sharp knife to peel away the label from the back of the unit, shown in Figure 11-2c. The screws holding the back cover in place are along the top and bottom edges of the newly exposed area. The pair of screws to the left of midline secure the laser module to the inside of the back cover.

Step 3

Free the battery connector and thermistor. As you remove the back cover, note the position of the thin white wires from the circuit board to the thermistor bead identified earlier. Next, push the 9V battery connector through the slot in the plastic back and allow the case to hinge open. Now, carefully remove the thermistor from the horn by tugging on the glue with needle-nose pliers. Avoid the wires and the epoxy bead, which are easily damaged. The plastic back is still tethered by the laser module.

Step 4

Extract the laser module. Free the laser module, shown in Figure 11-2d. The module is secured to the back cover by the two screws uncovered in step 2.

Step 5

Extract the circuit board. Remove the two Phillips-head screws at the base of the circuit board. Don't touch the four small screws that attach the LCD module to the board, near the black epoxy blob. Flip the board over and remove the one-piece set of elastomeric buttons, as in Figure 11-2e.

Step 6 (optional)

Extract the ultrasonic transducer. You'll need an intact horn-transducer assembly to test the circuitry, but if you want a good look at the transducer, simply pop it out of the horn housing, as shown in Figure 11-2f. The transducer is held in place by a friction-fit that's sealed with a black rubber band around the transducer. Push back, toward the lead, to free the cylindrical transducer, which looks like an ordinary piezoelectric buzzer.

Layout

Figure 11-3 shows the layout of the approximately 3 × 2 inch circuit board from the back or component side, oriented with the base of the unit to the left. There are four overlapping areas of the double-sided board that roughly correspond to the functions of power regulation, laser drive, ultrasonic transmission and reception, and numeric processing and display.

In tracing these circuits, I found a short piece of 32AWG wire invaluable. Many of the traces jump from one side of the board to the other, through small vias. When you find a trace that suddenly ends at a plated through-hole or via, insert the wire into the via and flip the board over and continue following the trace marked by the wire.

The power regulation and switching circuitry is in the lower-left corner, where the leads from the 9V battery attach to the board. Key components include a 10µf at 16VDC electrolytic capacitor, a glass SMT diode (D2) that protects against a

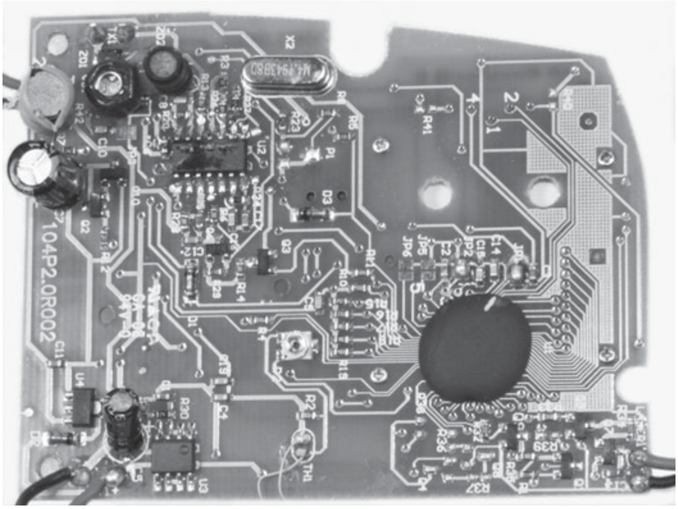

FIGURE 11-3 Circuit board layout, component side

reversed polarity battery, a 3-pin 7130 voltage regulator (U4) that supplies 3.0VDC continuously, and an 8-pin LM317LM voltage regulator (U3) that delivers 3.0VDC when the sonic measurer is powered up.

The laser driver circuitry is in the lower-right corner of the board, where the ultrasonic transducer leads attach to the board. There are two 1F NPN and two 5C PNP transistors in the area. One 5C (Q7) and one 1F (Q5) are driven by the microcontroller to switch the laser diode on and off. If you follow the traces, you'll find that the pair of transistors is supplied with 3.0VDC from the LM317LM voltage regulator. The 1F transistor is driven by the microcontroller, through a 310Ω resistor (R33). The other transistor pair in the area controls the distribution of 9VDC to components, including the LM317LM regulator.

The ultrasonic transceiver circuitry is concentrated in the upper-left quadrant of the board. The most prominent components are the 14-pin 274C quad operational amplifier (U2) that serves as the ultrasound receiver, a pair of zener diodes (ZD1 and ZD2) across the leads to the transducer, a 220μf at 16VDC electrolytic capacitor, the output transformer (T1), an inductor (L1), and a variable capacitor (VC1). The variable capacitor, inductor, and zener diodes in my unit were covered in semitransparent glue that required about 5 minutes of work with a forceps to remove. There are also a pair of 1F NPN transistors, a 6C NPN transistor, a 3F PNP transistor, and a pair of glass diodes (D1 and D3) in the area. The 4.1943MHz crystal (X2), located just above the 274C IC, is connected directly to the microcontroller.

The microcontroller and LCD driver circuitry is roughly confined to the right of the board. Several devices connect directly to the input or output ports of the entombed microcontroller, including the LCD module that is connected directly to the microcontroller, the crystal (X2), and the thermistor (TH1). The thermistor and calibration potentiometer (P1) connect to the microcontroller at the bottom edge of the board, between the LM317LM voltage regulator and microcontroller.

The front of the board is relatively sparse, as shown in Figure 11-2e. In addition to the LCD module is a momentary push button and pads for seven elastomeric buttons. No components are hidden under the LCD module. If you've followed the previous teardowns in this book that feature LCD modules, there's no need for you to remove the module to peek underneath. Besides, we'll need the display to test the other circuitry on the board.

Components

Now that you're oriented to the broad functionality of regions on the board, let's take a look at the individual components. Remember that the best resource for component information is manufacturer datasheets, available from the manufacturer's web site or the catalog of a large parts retailer, such as Digi-Key (www.digikey.com) or Mouser (www.mouser.com).

Ultrasonic Transducer and Horn

The 5-inch plastic horn serves as both transmitter and receiver antenna. The modest taper of the outside diameter, from 20 to 40mm, constrains the transmitted and received aperture angle to ±5 degrees from center. A close-up of the horn and transducer is shown in Figure 11-2f. The piezoelectric element of the transducer is visible through the cutouts in the front of the component. The resistance of the transducer is greater than 10MΩ, which is typical for a piezo transducer.

Thermistor

The 1mm diameter thermistor bead and relatively fragile wire leads, shown in Figure 11-4, provide the basis for adjustments in distance calculations based on ambient temperature. Thermistors are manufactured with metal oxides in a way that enhances their sensitivity to change in ambient temperature. This thermistor contains predominantly iron oxide—you can verify this with a magnet.

The ideal thermistor presents a rapid, significant, linear change in resistance with relatively small changes in ambient temperature. This contrasts with typical metal film resistors, which are formulated to present a constant resistance over a wide range of ambient temperatures.

This thermistor has a negative coefficient, meaning temperature and resistance are inversely correlated. Increases in ambient temperature result in decreased thermistor resistance, and vice versa. At an ambient temperature of 79°F, I measured the resistance as 76K. When I raised the temperature of the thermistor to 95°F by holding the bead between my fingers for 5 seconds, resistance dropped to 52K. I used a Fluke 80T-150U temperature probe and 87 DMM to measure the temperature of the air and of my skin.

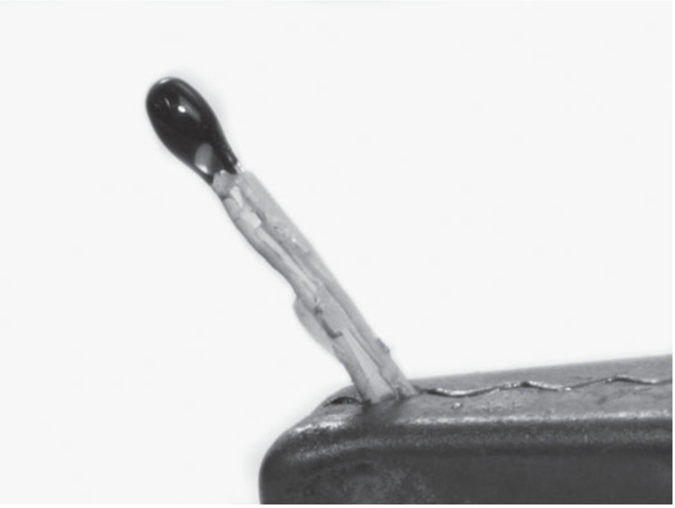

FIGURE 11-4 Thermistor bead

The thermistor is wired in series with a 9KΩ SMT potentiometer (P2) that can be adjusted to calibrate the unit. The two components are connected directly to the microcontroller.

Note See the teardown of the hygro thermometer in Chapter 8 for another example of a thermistor circuit.

Laser Module

The laser is the most entertaining find in this teardown. At 1.2 ounces, the laser module is also the heaviest component in the unit. The cylindrical laser diode and lens assembly are embedded in an aluminum block that's about 1-inch long and a little more than 1/2-inch square, shown in Figure 11-2d.

When the laser module is activated with 2.9VDC, it draws 28mA, or about 93mW ($P = 2.9V \times 0.028mA$). Assuming a typical efficiency of 1 percent, that's less than 1mW output. The majority of power is dissipated as heat, which explains the relatively massive block of aluminum—it serves as a heat sink. Laser diode efficiency and lifespan both suffer at elevated temperatures.

LCD Module

The LCD module in this teardown is essentially identical to the LCDs encountered in teardowns in earlier chapters. For example, as with the bathroom scale and pedometer (Chapters 3 and 5), the LCD panel interfaces with traces on the circuit board through an elastic Zebra connector.

Microcontroller

As noted in previous teardowns that featured epoxy-entombed microcontrollers, including the bathroom scale and pedometer, I consider the hidden slab of silicon a microcontroller instead of a microprocessor because of onboard analog-to-digital (A–D) processing of sensor data.

This microcontroller controls the LCD display, ultrasonic transceiver, and laser module. It also reads and interprets the thermistor value and performs calculations for distance, surface area, and volume. Its internal clock or timebase is dependent on the accuracy and stability of the 4.1943MHz crystal (X2).

7130 Positive Fixed Voltage Regulator

The 7130 fixed voltage regulator (U4) is the 3-pin SOT-89 component just to the left of the electrolytic capacitor, shown in Figure 11-5. The device receives input voltage from the 9VDC battery, via glass diode D2. The tab and middle pin are input, the top pin is ground, and the bottom pin is output. According to the datasheet, the regulator accepts an input of up to 24VDC and provides an output of 3.0VDC at 30mA. Maximum no-load current is 6μA.

The no-load current figure is important in this application because, as you can see in the figure, there is no on–off switch between the battery input and the regulator. I measured the quiescent current flow to the regulator with a fresh 9V battery at 2μA.

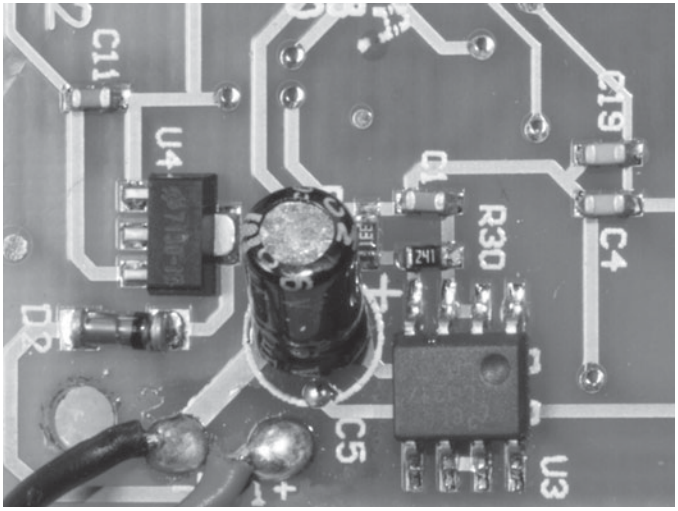

FIGURE 11-5 Voltage regulator section with 3-pin 7130 and 8-pin LM317LM

LM317LM Positive Adjustable Voltage Regulator

The LM317LM voltage regulator (U3) is the SOP-8 component visible in Figure 11-5. Unlike the smaller 7130, which can output only 3.0VDC, this regulator can be configured with external resistors to provide between 1.2 and 28VDC at 100mA. In this application, the output is fixed at 3.0VDC at 100mA.

The LM317LM, which is internally protected against shorts and elevated temperatures, is switched on when the PWR switch is depressed and switched off if there is no activity for 30 seconds. Output is on pins 2, 3, 6, and 7. Input is pin 1, and pins 5 and 8 are for voltage adjustment resistors.

ST274C Quad CMOS Operational Amplifier

The ST24C, which contains four independent CMOS (complementary metal oxide semiconductor) operational amplifiers, is the SO-14 package visible in Figure 11-6. Relevant specifications include low power consumption, owing to the CMOS construction; a supply voltage requirement of from 3 to 16VDC; and a gain bandwidth product (GBP) of 3.5MHz. The GPB is the midband gain of an operational amplifier multiplied by its bandwidth in MHz. In other words, amplifying a 40kHz signal is easily within the capability of the chip.

As noted in the teardown, a semitransparent film obscures the package markings. In this figure, pin 1 is in the upper-right corner of the device. For more information on the ST274C, consult the STMicroelectronics web site (www.st.com).

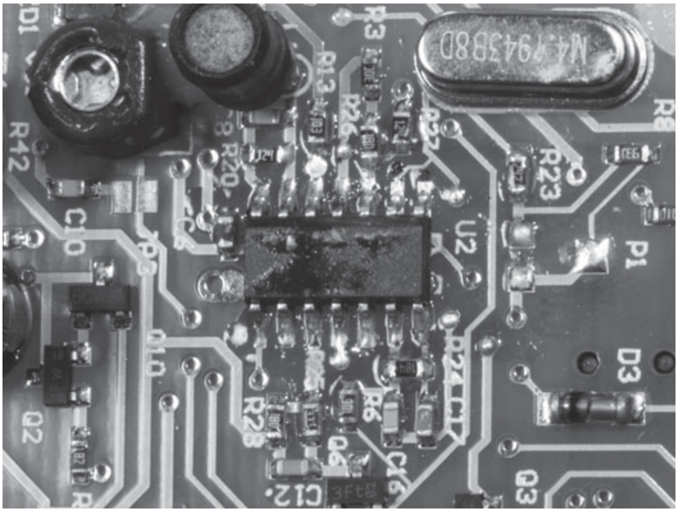

FIGURE 11-6 Transceiver circuitry, including 14-pin ST274C

Transistors

Four models of discrete SMT transistors are on the circuit board, all in SOT-23 packages. This packaging makes circuit tracing a breeze. Looking at the 3-pin transistor package from above, with the single pin up and two pins down, the collector is on top, the base on the lower-left, and emitter on the lower-right.

The 1F (BC847B) and 6C (BC817-40) transistors are general-purpose silicon NPN transistors rated at 50V. The 1F transistor is rated at 100mA collector current and the 6C transistor at 500mA. The 3F (BC857B) and 5C (BC807-40) general-purpose PNP transistors are rated at 45V, with a maximum collector current of 100 and 500mA, respectively. These transistors are employed in NPN-PNP pairs, with the transistor with the lower current rating driving the larger capacity transistor. The transistors listed in parentheses are common equivalents.

Diodes

A few unmarked, apparently general-purpose, glass diodes are also on the board. In addition, two unmarked SMT zener diodes are across the leads of the ultrasound transducer. Recall that zener diodes break down gracefully and exhibit a constant voltage drop, as long as the reverse current doesn't exceed the capacity of the diode. In this circuit, the diodes are used as a voltage clamp to protect the ultrasonic transducer from damage due to overvoltage. Based on my measurements, these are 75V zener diodes configured to constrain the output to the ultrasonic transducer to 150V peak-to-peak.

Crystal

The 4.1943MHz quartz crystal, visible in the top-right corner of Figure 11-6, is central to the operation of the microcontroller. The route from crystal to microcontroller, near test point 2 (TP2), is circuitous, involving traces on both sides of the board, including a trace under the LCD module. The crystal serves as the timebase for the microcontroller. This is a critical function, given that microseconds matter in measuring the speed of sound.

Quartz crystals are rated in terms of frequency tolerance in parts per million (ppm), operating temperature range, load and shunt capacitance, drive level, and ageing in ppm. Load and shunt capacitance and drive level affect component selection during circuit design. Most important from the perspective of sonic measurement accuracy are the frequency tolerance, operating temperature range, and ageing. Frequency tolerance, like resistor or capacitor tolerance, denotes the accuracy of the component at the time of purchase. Crystal ageing, on the other hand, indicates the likely change in resonant frequency while used in a circuit. Heat, vibration, stress from applied voltages, and physical stress from the internal mount all influence how a crystal will age.

Assuming X1 is an inexpensive quartz crystal, it probably has frequency tolerance of ±50ppm, or ±0.005 percent, and ages at ±5ppm, or ±0.0005 percent, per year. According to these specifications, the crystal should oscillate at 4.1943MHz ±210Hz and age ±21Hz per year. These values are overly optimistic, however, because they assume a constant operating temperature, which is an unrealistic assumption with a lightweight, handheld instrument. Precision measurement instruments designed for professional work often place quartz crystals used as a timebase in an insulated chamber that's maintained at a constant, elevated temperature.

Inductors

The two inductors on the board, the transformer (T1) and resonator (L1), are in the ultrasonic transceiver section, as shown in Figure 11-6. Figure 11-7 shows the transformer with its insulation partially removed to reveal the secondary winding composed of 46AWG enameled solid copper wire. The leads to the primary winding of T1 exit the top of the component. The secondary winding is attached to pins on the bottom of the transformer. Both inductors are wound on ferrite cores. You should verify core composition with a magnet. Many cores that look like ferrite are actually plastic.

I measured the inductance of L1 as 28mH at 1kHz and through a destructive teardown found that it's composed of more than 100 turns of 46AWG enameled

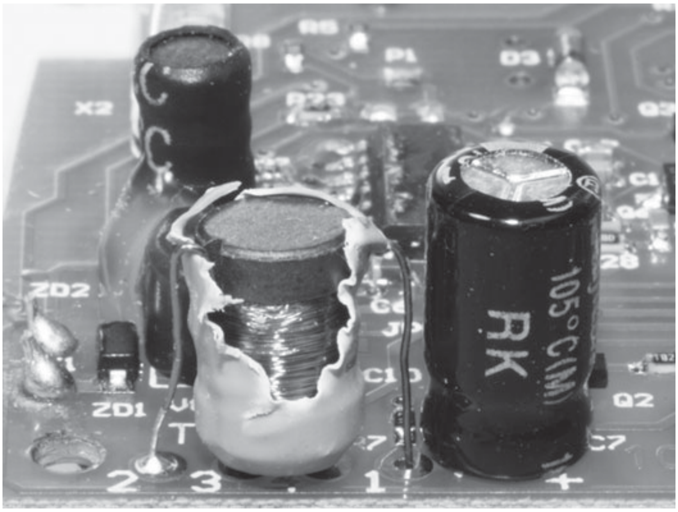

FIGURE 11-7 Transformer with insulation partially removed

copper wire wound tightly over the ferrite core. This inductor is used as part of a filter in the receiver circuit.

The secondary of transformer T1, like L1, is composed of at least 100 turns of fine enameled copper wire that's apparently 46AWG. The primary, which exits at the top of the component, is composed of about 40 turns of 32AWG solid enameled copper wire. The DC resistance of the secondary winding is 80Ω and that of the primary winding is 0.1Ω.

Switches

Seven of the eight switches on the front of the unit are elastomeric. As detailed in the teardown of the electronic pedometer in Chapter 5, depressing a carbon-impregnated elastomeric button shorts underlying traces on the circuit board, completing the circuit. The exception is the READ switch, which is a typical momentary-on SMT switch that is soldered directly over the board tracing of an elastomeric switch. That is, the mechanical switch is an afterthought, a modification made after the board was released.

One reason for using the mechanical switch may be because it's the most often used switch and more prone to failure. This supposes a greater MTBF (mean time before failure) for the particular mechanical switch versus the elastomeric switch. Can you think of another reason for installing the mechanical switch over the existing elastomeric switch?

Capacitors

The most notable capacitor on the board is the ceramic trimmer capacitor in the transceiver circuit, clearly visible in the upper-left corner of Figure 11-6. Trimmer or variable capacitors are considerably more expensive than fixed capacitors of the same voltage rating, but they allow fine-tuning of analog circuits. This capacitor is part of the LC filter in the receiver.

There are also two leaded electrolytic capacitors, and about a dozen garden-variety ceramic SMT capacitors. The 10µf at 16V electrolytic capacitor is used on the output of the LM317LM 3.0V regulator. The second electrolytic, a 220µf at 16V capacitor, is used on the 9V line feeding the switching transistors in the transmitter circuit.

Resistors

The board is populated with standard-tolerance SMT resistors. By now, you should be able to determine the value of resistors, including those with letter markings. If not, check out Appendix A to review SMT resistor markings.

How It Works

Because of the complexity of the physics involved in the operation of the sonic measurer, I've broken down the discussion into operational and functional perspectives.

Operational View

Let's start with an operational view of the sonic measurer, using Figure 11-8 as a guide. As shown in the top half of the figure, a transducer horn transmits a series of 40kHz pulse trains or chirps, separated by 110ms. Each chirp is composed of perhaps a dozen 40kHz pulses. After a little over 25ms, the ultrasonic wave train hits the interface between the air and the target—in this case, a perfectly perpendicular, reflective wall that extends infinitely in all directions.

At the interface, some percentage of the wave train—the echo—is reflected back toward the horn. The part of the echo that is captured by the horn causes the piezoelectric transducer to vibrate, generating an electric pulse. The microcontroller determines the time between events and stores the number. This process is repeated for the second chirp.

Now let's examine the amplitude versus time relationships in the lower half of Figure 11-8. Notice how the recovered pulse or echo is depicted as lower in amplitude and less sharply defined relative to the original pulse train or chirp.

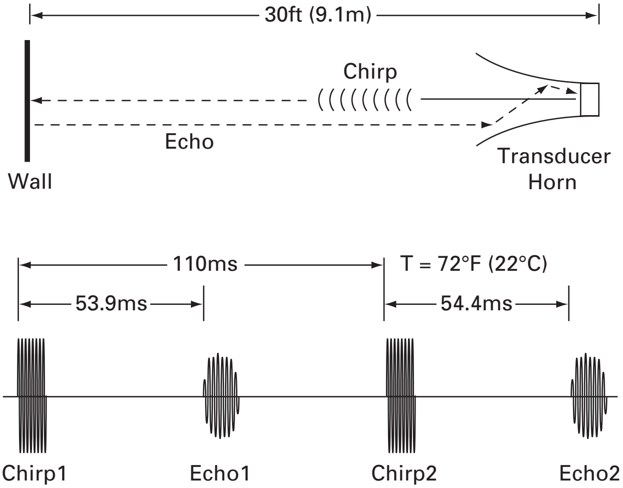

FIGURE 11-8 Operational view of the sonic measurer

This is because the energy of the original chirp is dispersed and attenuated by the air as it leaves the horn and is partially absorbed by the target 30 feet away. The ultrasonic chirp is additionally dispersed and attenuated on the return trip to the horn. If drawn to scale, the echo would not be visible on the amplitude tracing in Figure 11-8.

The interval between the start of each chirp and the start of its corresponding echo is computed by the microcontroller. In this example, the first measurement returned by the sonic measurer is 53.9ms. The interval between the leading edge of the second chirp and that of its echo is longer, at 54.4ms. The microcontroller compares the two interval values and, if they're close enough in value, a distance measure is flashed on the LCD screen. That measure may be the distance computed from the first interval, the second interval, or the average of the two intervals. My vote is for the average.

If the two interval values are too dissimilar, then the microcontroller will either transmit and evaluate another pair of chirps and echoes or generate an error message on the LCD. In my experimentation with the unit, I was unable to get it to repeat the chirp repetition cycle more than six times before it presented an error message. I assume that the similarity score is computed as a percentage of total time. Otherwise, longer intervals would require increased accuracy to get through the system without an error message. You'd need access to the microcontroller source code to verify the algorithms.

Now let's consider the physics of ultrasound in air. The nominal speed of sound in air, 1130 feet per second (ft/s), is an approximation for dry air at room temperature, at sea level, and with a typical CO_2 concentration. A more accurate figure for the velocity of sound in meters per second (m/s) considers the temperature, humidity, and CO_2 concentration, in the form of the following equation:

$$\text{Velocity of Sound} = (n\mathrm{RT/M})^{1/2}\ \mathrm{m/s}$$

where n is the adiabatic constant, a characteristic of the gases in air (nominally 1.4), R is the universal gas constant (8.314J/mol K), T is the absolute temperature in degrees Kelvin, and M is the molecular weight of the gases in kg/mol. According to the equation, the speed of sound in air is proportional to the square root of the absolute temperature and increases slightly with increasing humidity and CO_2 concentration. In practical terms, the speed of sound in air is a function of the density of air, which is dependent on the temperature.

Solving the equation with reasonable speed and accuracy is problematic using integer arithmetic on a typical microcontroller, and I doubt that the chip in our sonic measurer is capable of floating-point math. A more microcontroller-friendly model for the velocity of sound in air, and the one probably used in the sonic measurer, is this:

$$\text{Velocity of Sound} = 331.3 + 0.6\mathrm{C}\ \mathrm{m/s}$$

The 331.3 figure is the speed of sound in m/sec at 0°C, assuming 0 percent humidity. This approximation ignores the contributions of humidity, CO_2, and the frequency of the sound on velocity.

Let's return to Figure 11-8, where the back of the transducer horn is 30 feet from the wall and the ambient temperature is 72°F. According to the second formula, the velocity of sound is 345.5m/s (332.3 + 0.6 × 22) or 1113.5 feet/s. The round-trip transit time for an ultrasonic chirp in this environment is (30 + 30) / 1113.5, or 53.9ms. The first chirp-echo interval agrees with this theoretical figure. The second chirp-echo interval is longer than expected.

Using the same approach, within the operating temperature range of the sonic measurer, 20–120°F, transit time ranges from 55.7ms at 20°F to 50.5ms at 120°F. This 5.2ms difference corresponds to distance of nearly 6 feet at 72°F. Obviously, the sonic measurer has to control for temperature to provide distance measurements within the specified accuracy.

Functional View

Now consider a functional view of the sonic measurer, using the simplified schematic of the systems in Figure 11-9 as a reference. Let's focus first on the power circuitry, which is shown in the upper-left of the schematic.

A 9V alkaline battery supplies power continuously. There is no hard on–off switch. Based on my measurements, the current demanded of the battery varies

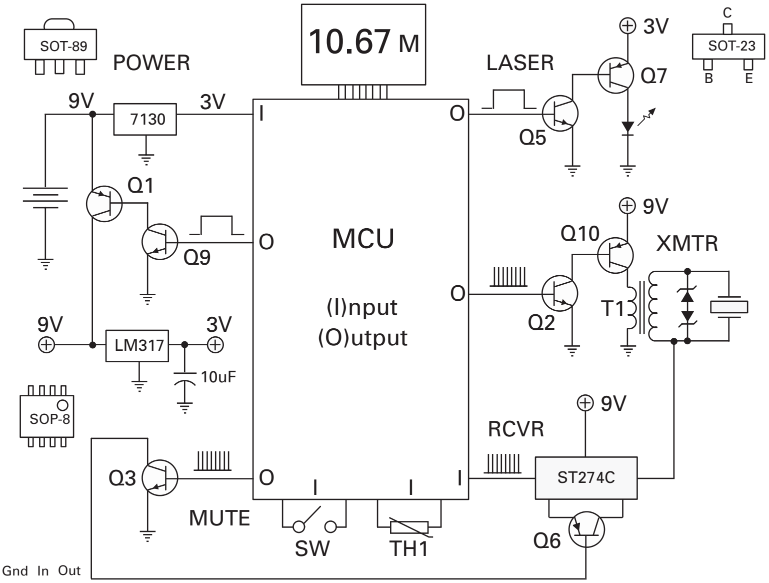
SOT-89
POWER
9V
7130
3V
10.67 M
LASER
3V
C
SOT-23
B
E
Q7
Q5
I
O
Q1
Q9
O
MCU
9V
Q10
XMTR
O
(I)nput
(O)utput
Q2
T1
9V
LM317
3V
10uF
SOP-8
9V
RCVR
Q3
O
I
I
I
ST274C
MUTE
SW
TH1
Q6
Gnd In Out

FIGURE 11-9 Simplified schematic of major systems

from 2µA when the unit is hibernating to 40mA when the laser is on and ultrasonic transmitter is active. Assuming a standard 300mAh (milliamp hours) alkaline battery, the specification of 8-hour battery life with continuous use seems reasonable (300mAh/40mA = 7.5 hours).

Three different supply lines power the circuitry in the sonic measurer. The 7130 voltage regulator provides the microcontroller with 3.0VDC at all times. The switches on the front of the circuit board and the thermistor, which interface directly to the microcontroller, are also continually energized by the output of the 7130 regulator.

The second 3.0VDC line, supplied by the output of the LM317LM, is live only when the 9VDC input to the LM317LM is switched on by the microcontroller through the 1F (Q9) NPN and 5C (Q1) PNP transistor pair in the laser area of the board. When the output of the microcontroller is high, Q9 is forward-biased, which in turn forward-biases Q1, providing a low resistance path from the 9V battery to the input of the LM317LM. In addition to powering the LM317LM, the switched 9VDC also supplies the transceiver circuitry, including the ST274C quad op amp (U2) and associated transistors Q2, Q10, Q3, and Q6.

The basic switching configuration used with Q1 and Q9 is repeated in the laser circuit and twice in the transceiver circuit. In the laser circuit, shown in the upper-right corner of the schematic, the output of the microcontroller goes high, forward-biasing Q5, enabling Q7 and the laser diode to conduct. The voltage across the laser diode is a few tenths of a volt less than the voltage supplied by the LM317LM voltage regulator because of the voltage drop across the junctions of Q7. I didn't crack the laser diode case, but I assume the module contains a series resistor to limit current flow through the diode.

The microcontroller is the source of the 40kHz signal that's eventually transmitted. The transistor pair Q2 and Q10 and step-up transformer T1 are used to increase the amplitude of the signal to achieve a range of at least 50 feet. A signal train of 40kHz pulses is output from the microcontroller, which forward-biases Q2 and Q10, sending pulses from the switched 9VDC source through the primary of transformer T1 to ground. These 9VDC pulses at 40kHz are stepped up by T1 to 150VPP (peak-to-peak). That is, T1 must have a turns ratio of at least 150/9, or 16:1. A pair of 75V zener diodes is placed back-to-back across the transformer secondary to limit voltage across the piezoelectric transducer.

The microcontroller, which is in hibernation mode until the PWR switch (SW) is depressed, employs its own internal power switching circuitry. In addition to monitoring the soft switches, the microcontroller maintains voltage across the thermistor at all times. This minimizes fluctuations in the temperature of the thermistor that would otherwise occur due to self-heating if the thermistor had to be energized before each reading.

As discussed in the teardown of the hygro thermometer in Chapter 8, the process of measuring the resistance of a thermistor involves applying either a continuous or pulsed voltage across the thermistor. The resultant current heats the thermistor, raising its temperature above the ambient temperature. Depending on ambient conditions, this self-heating could be a source of error.

To give you an idea of the significance of the thermistor, consider that I measured a distance with the sonic measurer as 4 feet, 11 inches at ambient temperature of 79°F. I then held the thermistor bead between my fingers, raising the temperature of the bead to 95°F. The instrument read the same distance as 5 feet, 0 inches. That's a difference of 1 inch at only 5 feet. The difference would be proportionally greater at greater distances.

As in most transceivers, the receiver circuitry is considerably more complex than that of the transmitter. Whereas the transmitter circuitry simply has to amplify the pulse train from the microcontroller, the receiver has to distinguish the incoming signals from noise and ignore the transmitted pulses.

As show in simplified form in the lower-right area of Figure 11-9, the analog signal processing for the receive function is performed by the ST274C quad op amp, using the signal produced by the piezoelectric transducer when it receives an ultrasonic impulse. A muting circuit consists of a 3F (Q6) and 1F (Q3) transistor configured in the familiar pairing used with the other systems in the unit.

A more detailed, but still simplified, schematic of the receiver circuit is shown in Figure 11-10. For clarity, I haven't included many of the bypass and feedback capacitors and resistors. I also don't show the zener diodes or transformer secondary across the ultrasonic transducer. Note the two analog data feeds to the microcontroller, from op amps 1 and 3, and that the muting function is controlled by a signal from the microcontroller.

Before continuing with the circuit, let's revisit the needs of our receiver. The first requirement is that it receives echoes from the transmitted chirps or ultrasonic

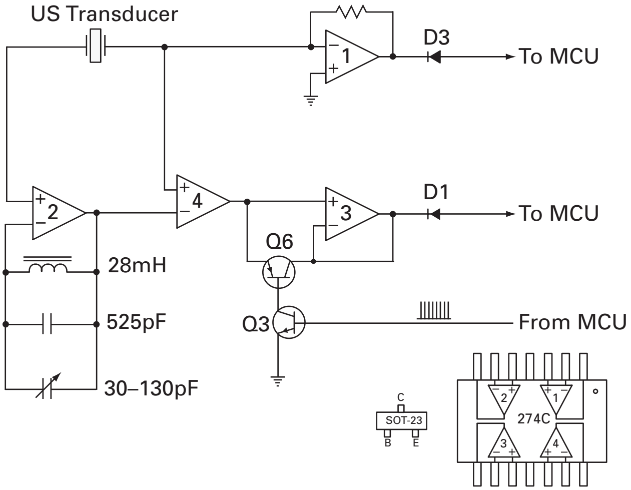

FIGURE 11-10 Simplified schematic of the receiver

pulse trains. The second requirement is that it ignores noise in the environment—that is, the receiver should be selective and provide a good signal-to-noise ratio. If you consider the purpose of the sonic measurer, it's likely to be used in an environment with noise from power tools, ultrasonic rodent repellers, and other noise sources that extend to 40kHz and beyond. The plastic horn provides a degree of directional selectivity, which cuts down on noise. The ultrasonic transducer also has a natural resonant frequency, presumably around 40kHz, that provides attenuation to signals greater and less than 40kHz.

Now, referring back to the schematic in Figure 11-10, let's assume that the piezoelectric transducer is generating an electrical signal secondary to ultrasonic vibrations—that is, there's an AC signal across the transducer.

Op amp 1 is configured as a basic inverting amplifier, meaning that it takes the signal at the inverting (–) input and amplifies it, with the output 180 degrees out of phase with the input. (By the way, if you're new to op amps (operational amplifiers), they're covered in more detail in Chapter 13, a teardown of a guitar effects pedal.) For now, let's stay with a relatively high-level discussion. The output is sent, through diode D3, to the microcontroller (MCU). Because the diode has a minimum forward voltage drop of about 0.7V, only signals greater than that make it to the input of the microcontroller. Noise and weak echoes are blocked by the diode.

In a quiet, indoor setting, a receiver composed of op amp 1 and D3 would probably be sufficient. However, other than the selectivity provided by the transducer and horn assembly, there is no provision for noise rejection by the op amp 1 circuit. Any sound impacting the transducer will result in a signal to the microcontroller.

The remaining three op amps in the 274C chip are used to increase the effective signal-to-noise ratio of the receiver. Op amp 2 is configured as a noninverting amplifier with unity gain at 40kHz—that is, a notch filter. At resonance, the feedback network composed of a parallel resonant LC circuit from the output to the inverting input (–) of op amp 2 is essentially a wire. Above and below 40kHz, the impedance of the feedback increases, thereby increasing the gain of the amplifier circuit.

Recall the resonant frequency of an LC circuit is $1 / (2\pi(LC)^{1/2})$. Applying the formula with a 28mH inductor (L1) in parallel with a 525pF fixed capacitance and a 30–130pF variable capacitance results in a resonant frequency range of about 37 to 40kHz. The variable capacitor enables the manufacturer to tune the circuit to precisely 40kHz. The alternative, using an inductor and fixed capacitor with high tolerance, was probably cost-prohibitive.

Op amp 4 compares the unfiltered signal on one side of the transducer with the output of op amp 2, taken from the other side of the transducer. The input signals to the two op amps are thus 180 degrees out of phase. If the signal from the transducer is 40kHz, then the amplitude of the signals on the inverting (–) and noninverting (+) inputs to op amp 4 should be equal. Assume that at one instant the input to op amp 4 is positive and the input to op amp 2 is negative. The positive input to the noninverting (+) input of op amp 4 results in a positive output. Similarly, the negative input to the inverting (–) input of op amp 4 results in a positive output.

Now let's assume a signal of broadband noise source that extends from 30 to 50kHz. Op amp 1 will amplify the signal and send it to the microcontroller. Op amp 2 will amplify it with a gain of greater than unity. Now op amp 4 is presented with a strong signal on its inverting (–) input, relative to the direct signal from the transducer on its noninverting (+) input. The result is an increase in the amplitude in the output of op amp 4, relative to the output with a 40kHz signal.

Ignoring the contribution of the Q3-Q6 transistor pair for the moment, op amp 3, which is configured as a unity gain buffer, takes the output of op amp 4 and sends it, through diode D1, to the microcontroller. D1 serves the same noise-limiting function as D3. In summary, at this point in our analysis, when a pure 40kHz sound is impacting the transducer, the output from op amps 1 and 3 are about the same in amplitude. When a noise source extends on either side of 40kHz, the output of op amp 3 may be many times that of op amp 1.

My assumption is that the program running on the microcontroller examines the relative amplitudes of both signals to determine whether to return an error message or use the signal to compute distance. As long as the amplitudes of the signals from op amps 1 and 3 are about the same amplitude and great enough to overcome the noise barriers provided by D1 and D3, the signal is considered a valid echo. When the amplitude of the output of op amp 3 is significantly greater than that of op amp 1, the signal is presumably considered noise by the microcontroller.

Finally, let's consider the muting circuit formed by Q6 and Q3. As you can see in the schematic, the collector and emitter of Q6 are across the input ports of op amp 3. As a result, when Q6 conducts, the input ports are shorted and the output of Q3 is zero, regardless of the output of op amp 4. Q3 receives 40kHz pulses from the microcontroller that are identical to those sent to the transmitter. As a result, the output of op amp 3 is muted during transmit.

Given that the microcontroller can be programmed to ignore signals that occur during transmit and that the output of op amp 1 is not muted, my assumption is that the output of op amp 3, if not muted, could damage the input circuitry of the microcontroller. To verify this assumption, we'd have to have access to information on the entombed microcontroller. Can you think of another reason for the muting circuit? Why not simply limit the input to the microcontroller with a diode clamping circuit?

Mods

This sonic measurer is more of a platform for experimentation and repurposing than for mods. The laser module with its built-in ±5-degrees targeting cone pattern is an interesting tool in itself. Given the spread of the pattern, it's easy to estimate the distance of a moving target or of a structure more than 50 feet away. Recall that the ultrasonic measurer requires a stationary target.

Another repurposing project that you might consider is an ultrasonic source detector. Simply remove D3 from the output of op amp 1 and either replace it with an LED or the leads of your multimeter. You might be surprised at the number of significant ultrasound generators in the environment. For example, I discovered that when my bicycle chain needs lubrication, it generates a lot of ultrasonic noise when I turn the pedals with the bike on a stand, even though I can't hear any change in the sound produced by the chain.

I'm sure you can think of ultrasonic diagnostic aids to use with home appliances, car engines, and other mechanical devices. There are also uses for ultrasound as the basis for biological experimentation, from disrupting the flight of bats, to chasing away stray dogs and irritating rodents.

This page intentionally left blank

PART III

For Musicians

This page intentionally left blank

Chapter 12

Electric Guitar

The Fender Telecaster, or "Tele," the first mass-produced, solid-body electric guitar, is the easiest guitar to tear down and mod. And there's good reason—sheer simplicity. As shown in Figure 12-1, the basic design hasn't changed since Fender introduced the guitar in 1949—a solid body, bolt-on neck, one or two magnetic pickups, a volume control, and what passes for a tone control. Models with two pickups have the added complexity of a pickup selector switch, but a stock Tele has no whammy, or tremolo, bar.

In this chapter, we'll tear down a Fender American Deluxe Telecaster, which incorporates two pickups, a pickup selector switch, and volume and tone controls. I'll also walk you through a mini-teardown of a magnetic pickup and discuss several mods, including how to configure the Tele as a MIDI (Musical Instrument Digital Interface) controller.

Highlights

This teardown is deceptively simple. After all, we're working with a slab of wood, six steel strings, tuning machines, a pair of magnetic pickups and potentiometers, a capacitor, a three-way switch, some wire, and an audio output jack. As electronic devices go, it doesn't get much simpler—or does it? As you'll see, there's a hidden elegance in the simplicity of the Tele, and what appears simple at the outset is actually a fairly complex electromechanical device. During the teardown, note the following:

- How shielding is used to minimize EMI
- A star ground configuration that minimizes noise
- Construction and wiring of the pickups—a major contributor to the guitar's tone
- Reliance on traditional components and wiring

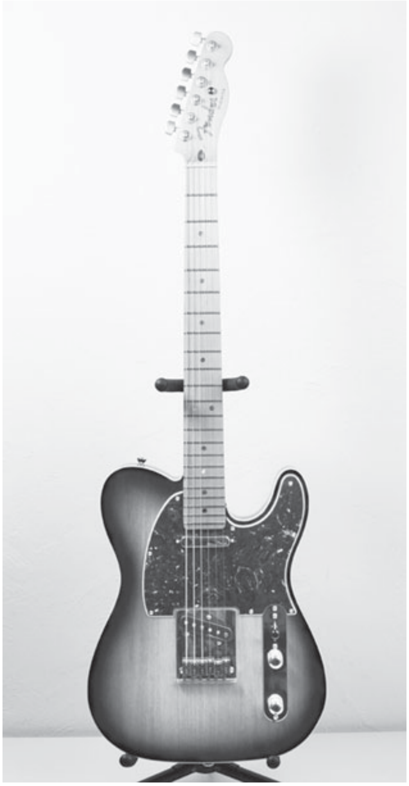

FIGURE 12-1 **Fender American Deluxe Telecaster**

Specifications

Guitar specifications vary markedly from one manufacturer to the next, and even within the same brand. Consider that Fender Teles retail from about $180 to $1800, depending on the quality and origin of the body, neck, and components. You can pick up a knock-off from dozens of vendors for less than $200, and online suppliers offer everything from replacement parts to complete kits. Because they're so easy to mod, you never really know what's inside a pre-owned Tele until you take a look inside. For example, this teardown features a Tele in which I replaced the stock potentiometer and capacitor with a more functional tone control.

If you've shopped recently for an electric guitar, you know that even when specifications are offered, they tend to be qualitative and focused on materials and construction. There's typically nothing quantitative in the way of output level, signal-to-noise (SNR) ratio, or tuning stability. Instead, the marketing brochures list the type of wood—typically swamp ash or alder—and finish, neck configuration and color, and the type of mechanical tuners. There might be a mention of the types of pickups. For example, here's a condensed version of the specifications for the American Deluxe Ash Telecaster, from the Fender site (www.fender.com):

- Ash body
- U-shaped maple neck
- Fingerboard with vintage tint, abalone dot position inlays, and 22 medium jumbo frets
- Stainless steel Telecaster bridge with chrome-plated solid-brass saddles
- Samarium Cobalt Noiseless Telecaster pickups
- S-1 switching

Although everything listed here affects tone and playability, the types of pickups and switching system are the most relevant specifications for our purposes. "S-1 switching" is simply Fender's marketing term for a four-way switch.

Figure 12-2 shows the key electronic elements of a Telecaster. On the far right along the midline, directly under the six steel strings, is the neck magnetic pickup, which is mounted on a springy suspension system in the swirl-patterned plastic pickguard. The strings terminate in a fixed bridge that's bolted directly to the body. Mounted diagonally in the bridge is the bridge magnetic pickup.

Along the bottom, mounted in the control plate, are, from left to right, the tone control, volume control, and pickup selector switch. Not shown in Figure 12-2 is the 1/4-inch audio out jack, which exits the side of the guitar at about the level of the tone control.

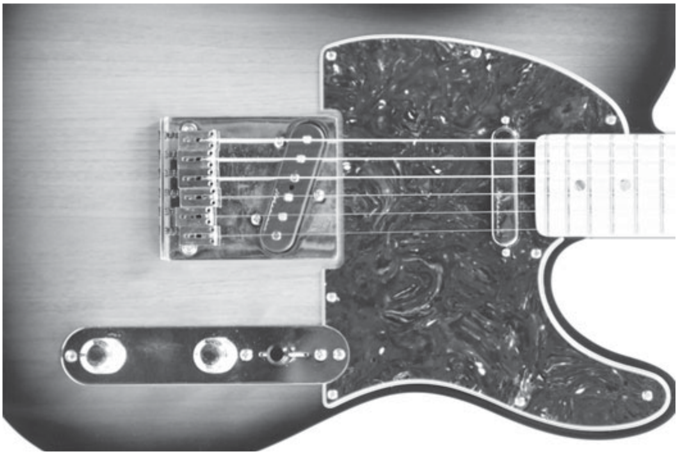

FIGURE 12-2 Electronic elements of the Telecaster

Operation

Mastering the guitar involves patience and years of practice. Designing a guitar that plays correctly is a nontrivial task that requires knowledge of mechanics, electronics, and how to integrate the two. For example, the height of the strings above the pickups significantly affects the spectral content and amplitude of the output signal. Strings that are too close to the magnetic pickups create a distorted overdriven signal, and strings too far from the pickups produce a weak, noisy signal. Similarly, the metal truss rod that runs through the neck must be adjusted so that there is a slight concavity to the fretboard—otherwise, the strings will buzz when struck. But too much concavity adversely affects playability.

In addition, the distance between the two supports for each string—the nut at the head of the guitar and the saddle where each string terminates in the bridge—must be adjusted for correct intonation. Otherwise, the notes played along one part of the neck will be in tune, but notes played at another part of the neck will be out of tune.

Moreover, these electromechanical systems are interdependent. If you modify the concavity of the neck, you affect intonation, the height of the strings above the magnetic pickups, and the guitar tone. As you can see, that simple slab of wood and strings is actually a complex system, and we've only touched on the circuitry.

From an electronics perspective, an electric guitar is one component in a large, interdependent system that minimally includes cables, an effects pedal, an amplifier, and a speaker cabinet. Each element contributes to the tone or sound of the guitar. Without an amp, an electric guitar is useless. Many players wouldn't consider practicing or playing on anything other than a tube-type amp, but others swear by the lightweight, portable, and more affordable solid-state amps with built-in effects.

Back to the operation of the guitar, the player sets everything in motion by strumming one or more steel strings. The vibrating strings interact with the magnetic fields emanating from the neck and bridge pickups, resulting in the generation of electrical signals. Pressing down on a string so that it makes solid contact with a fret on the neck changes the string's frequency of vibration and the frequency of the signal generated by the magnetic pickups. This signal, which may be mixed with signals from other instruments, is routed to the volume and tone control potentiometers, and finally to the audio out jack.

Note the alignment of the six steel strings over the six pole pieces of the bridge pickup, as shown in Figure 12-3. Pole pieces are metal cylinders—typically magnetic—that are embedded vertically in the pickup. The bridge and neck magnetic pickups each produce a distinctive tone, by virtue of their design, adjustment, and relative position. The three-way selector switch enables the player to select the signal from the neck pickup, the bridge pickup, or neck and bridge pickups in parallel. A popular mod, and a standard feature on some Teles, is a four-position switch that allows a fourth option of connecting the neck and bridge magnetic pickups in series.

The volume and tone controls do more or less what their names suggest. The volume control varies the amplitude of the magnetic pickup signals—typically from zero to a few hundred millivolts at an output impedance of a few hundred-thousand

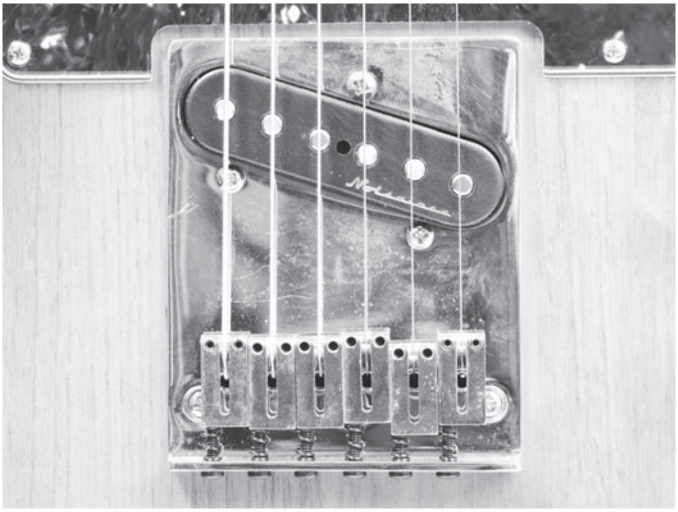

FIGURE 12-3 Bridge close-up

ohms. The stock tone control mainly attenuates the higher frequencies, with the resulting output varying from sparkling to muddy. Because there isn't much need for a muddy tone, it's a common practice to set the tone potentiometer to high and then leave it alone. A popular mod involves replacing the tone potentiometer with a switched bank of capacitors, which results in brighter tone—that is, more treble.

Teardown

You should be able to complete the teardown, illustrated in Figure 12-4, in about an hour. If you're following along at home with your Tele or another electric guitar, remember that electronic components and wiring configuration vary from one guitar to the next, depending on the make, model, year of production, country of manufacture, and, most importantly, the mods performed by previous owner(s).

If you're tearing down a stock Tele, you'll see essentially the same components and wiring described in this teardown. The only significant deviation from a stock guitar is a tone control mod, which I'll discuss in detail later. During the teardown, make note of the wiring and switching arrangement, as these can significantly affect a guitar's tone.

Tools and Instruments

You'll need a Phillips-head screwdriver, wire cutters, and a multimeter. A magnet polarity tester—a small bar magnet or magnetic compass—is useful, but not necessary. Place a towel, carpet remnant, or foam sheet on your workbench to

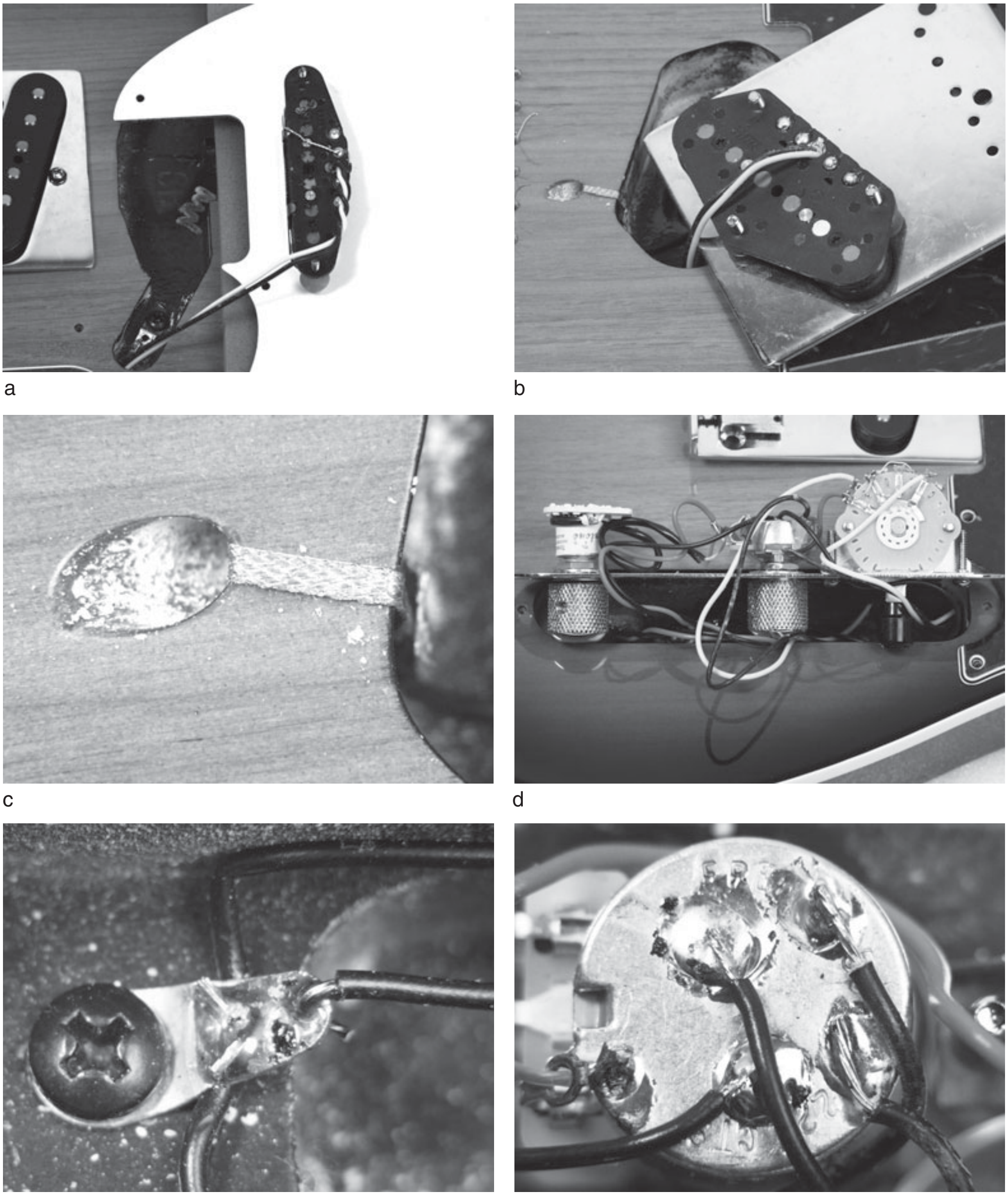

a

b

c

d

e

f

FIGURE 12-4 Teardown sequence

protect your guitar's finish. A few hand towels or small foam sheets would be helpful as well.

Step by Step

If you plan to repair or refurbish a guitar you're tearing down at home, try to minimize the number of times you remove and replace the wood screws. It's not a good idea to remove and reinsert the wood screws more than a few times during the life of a guitar, because the wood eventually strips. However, for your benefit, I'll remove all the wood screws that provide access to the electronics.

Step 1

Remove the guitar strings. Remove the strings as you would if you were preparing to restring the guitar. Start by releasing the tension on the strings by unwinding each of the six tuning machines. Cut the strings in half with diagonal cutters or, if you have them, guitar string snips. Either remove the string remnants from the bridge and the tuners, or trim them so that they're out of the way.

Step 2

Release the pickguard. Remove the eight Phillips-head wood screws retaining the pickguard.

Step 3

Examine neck pickup. Flip over the pickguard onto a cloth, as shown in Figure 12-4a. Note the magnetic neck pickup mounted in the pickguard and the two wires running from the pickup toward the control plate. The white or yellow wire is signal and the black wire is ground. As you'll see later, the polarity, or, more properly, the phasing, of the magnetic pickups matters.

Step 4

Examine the cavity. Note that the cavity is thickly coated with dark gray or black conductive paint. You should also see a metal tab attached with a wood screw to the cavity and a wire leading from the tab toward the controls.

Step 5

Verify the conductivity of the cavity coating. With your multimeter, check the resistance from the metal tab to any point within the cavity. The resistance should vary from a few hundred to perhaps a few thousand ohms, depending on the space between the probes. The cavity is painted with conductive paint to shield the electronics from external magnetic fields.

Step 6

Release the bridge. Remove the three Phillips-head wood screws retaining the bridge. Depending on the bridge design, you may have to push the center saddle—one of the metal blocks securing a string to the bridge—aside to access the middle screw. See Figure 12-3 for a view of the six saddles and their attachment to the bridge.

Step 7

Examine the bridge. Flip over the bridge onto a soft cloth or foam sheet, as shown in Figure 12-4b. Note the white signal wires and the black ground wire. Note also the copper braid running from the lug attached to the center of the cavity to the surface of the guitar body. The exposed braid, shown in more detail in Figure 12-4c, is normally sandwiched between the guitar body and the bridge to ground the bridge. The grounded bridge plate and painted cavity serve as electromagnetic shielding for the bridge pickup and associated wiring. Set the bridge back in place for now.

Step 8

Extract the control plate assembly. Remove the Phillips-head wood screws on either end of the control plate. Extract the assembly and lay it across the cavity, as shown in Figure 12-4d. From left to right are the tone potentiometer, volume potentiometer, and pickup selector switch. The two controls are 250KΩ audio potentiometers, and the switch is a Fender three-position pickup selector switch.

Step 9

Examine the control plate cavity. Push the control plate to the side, taking care not to scratch the guitar finish. You should see the conductive paint throughout the cavity, as well as a metal solder tab attached to the cavity with a wood screw, as shown in Figure 12-4e.

Step 10

Trace the black ground wires from the tabs within each cavity. All ground wires should terminate on the back of the volume potentiometer, shown in Figure 12-4f. In my guitar, four wires are soldered to the back of the pot—one for the bridge cavity ground, one for the neck cavity ground, one for the control plate cavity ground, and one for the audio output jack ground.

Layout

Figure 12-5 shows the basic wiring layout of the volume and tone controls found in most electric guitars. Leads from one or more pickups (the white wire at top right) terminate at a volume potentiometer (right) and tone potentiometer with

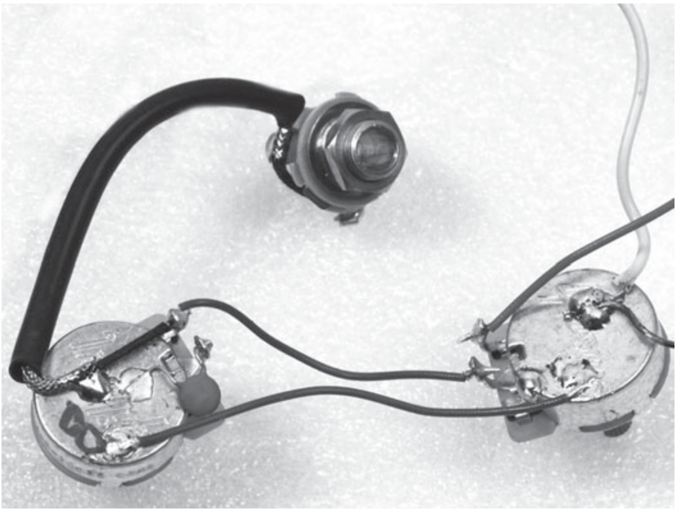

FIGURE 12-5 Basic electric guitar control layout

capacitor (left), and from there, pass to an audio output jack (center). The capacitor and tone potentiometer form a variable low-pass RC (resistor-capacitor) filter. The components shown in Figure 12-5 are from a Gibson Melody Maker guitar, but the same configuration can be found in any number of guitars.

Note that in Figure 12-5, the connection between the tone potentiometer and the output jack is made with coaxial cable. The shield is soldered to the tone potentiometer case, which is in turn wired to the case of the volume potentiometer. A cheaper alternative to coaxial cable is twisted pair, which provides somewhat less noise and EMI protection. This configuration also illustrates that a star ground scheme isn't universally adopted and can be modified. As you can see, three ground connections are soldered to the back of the volume potentiometer, but the ground connection for the audio output jack is made at the tone potentiometer.

At first glance, the use of a potentiometer case for solder connections seems like poor construction technique. After all, a solder lug attached to the control plate would provide a neater, more professional layout. However, this is simply how it's done with legacy guitars from Fender, Gibson, and others. To guarantee authentic, vintage tone, the layout and components used in a stock Tele have remained essentially unchanged for decades.

Components

I don't want to give you the false impression that innovation in the design of electric guitars somehow halted in the 1950s or 1960s. To the contrary, there are numerous examples of electric guitars that push the envelope in electronics and materials, from the lightweight resin-based guitars from Parker (www.parkerguitars.com) and the MIDI-compatible guitars from Breedlove (http://breedlovemusic.com), to the synthesizer guitars from Moog (www.moogmusic.com) and Line 6 (www.line6.com). However, vintage-style guitars remain a significant force in the industry. Fortunately, if you learn the fundamentals of a vintage-style guitar like the Tele, you'll be well positioned to understand the more complex electronic guitar designs.

EMI Shield

Well-built electric guitars are designed to operate in electrically hostile environments—near other instruments, amplifiers, speakers, cables carrying power and audio, fluorescent lights with faulty ballasts, and cell phones. Of these signals, the 60Hz power grid is the most powerful, ubiquitous, and problematic. Without shielding, the wire loops within the magnetic pickups, the wires from the pickups to the controls, and wires to the audio output jack all form excellent antennas for receiving 60Hz hum.

As noted in the teardown, the designers of the Tele attempted to minimize hum pickup by painting the cavities with conductive paint and by grounding the large metal bridge plate. Painting guitar cavities with conductive paint is a common method of reducing noise, hum, and radio interference because it's effective and easy to apply. I use conductive paint from Stewart-MacDonald Instrument Repair Supply (www.stewmac.com), which contains graphite and carbon black. At $60 per pint, it's expensive but easy to work with. If you use it, don't touch the toxic paint while it's wet, and wash your hands after painting.

An alternative to painting is to line each cavity with self-adhesive copper tape. The adhesive backing is conductive, making overlapping seams electrically continuous. Because of the significant expense and time involved in working with copper tape, I prefer paint for cavities and conductive copper tape for the inner surface of the pickguard. Copper tape does look better than black paint in the cavities—a point to consider if you plan to show off your handiwork with a clear pickguard.

Wiring

In addition to using shielding to reduce hum, twisted-pair and coaxial cable can be used to connect the pickups, switches, controls, and output jack; the use of proper grounding techniques can also help. Twisted pairs of wires carrying signals from the pickups can minimize hum pickup, because the 60Hz signals induced in successive twists cancel each other. Coaxial cable minimizes hum pickup by enveloping the inner conductor with a grounded mesh or foil tube.

During the teardown, we saw that the ground wires from each cavity, the audio output jack, and the switch all terminated on the case of the volume pot. The use of a common point in the guitar to which all ground connections are tied is one way to minimize the chances that currents in one ground line will generate a signal in another part of the circuit sharing the same ground. In other words, if there are multiple grounding points, each ground is at a different potential, because there is a finite resistance from one ground point to the next. As such, if one ground path carries a significant current, a voltage drop occurs across the path, and this voltage can interfere with the signals from the magnetic pickups. It's important to note that using a common ground point isn't the only, or necessarily the best, way to insure proper grounding. However, it's tradition, and it usually works.

Another practice that is influenced by tradition is the use of fabric insulation. Some guitar builders insist on using 22AWG tinned, stranded wire insulated with a Celanese (acetate) wrap and waxed, braided cotton outside insulation instead of modern PVC. However, I'm not convinced that any human can hear a difference in tone because of the use of stranded versus solid core wire or of different types of insulation. PVC is less expensive, more stable in environments with varying temperature and humidity, and readily available.

Potentiometers

The Tele controls are 2W, 250KΩ audio taper potentiometers. The audio or log taper potentiometers approximate the logarithmic response of human hearing, so that a given rotation of the volume potentiometer results in equal changes in volume, regardless of the initial potentiometer setting. As discussed in previous teardowns, a linear potentiometer would provide a nonlinear change in volume that is dependent on the initial potentiometer setting.

A 2W potentiometer seems like overkill for a guitar that generates a few milliwatts. However, for many builders, tradition and vintage tone dictate the use of dated hardware. Also, from a practical perspective, the heat absorbed during the process of soldering multiple leads to the case of a metal potentiometer would probably destroy lesser components.

Capacitor

The stock tone control in my Tele uses a 250KΩ audio taper potentiometer and a 0.05μf, 100V ceramic disc capacitor. Ceramic disc capacitors are a good choice for this application because they're inexpensive, they're stable at a range of temperatures, and they fit nicely in the limited space of the control cavity. Mods commonly use different values of ceramic disc capacitors, sometimes in conjunction with one or more film capacitors.

Audio Jack

The 1/4-inch mono audio output jack is standard throughout the guitar industry. The jack is tough enough to withstand abuse from accidentally yanked chords and plugs, is relatively inexpensive, and requires little maintenance.

There are exceptions to the standard output jack. For example, Parker (www.parker.com), Paul Reed Smith (www.paulreedsmith.com), and a few other guitar manufacturers install stereo audio output jacks on their guitars that have both magnetic and piezo pickups. One channel is from the magnetic pickup and the other is from the piezo pickup and built-in preamplifier. In addition, some traveler guitars with built-in audio amplifiers sport 3.5mm jacks for use with lightweight headphones.

Switches

The traditional three- or four-position, spring-action lever switch, shown in Figure 12-6a, is used in virtually all Teles. Some non–Fender Telecaster models feature a modern, sealed lever switch. However, tradition and the pursuit of vintage tone often dictate the traditional, expensive, and somewhat bulky lever switch.

Alternatives to the lever switch include the open three-way toggle switch, shown in Figure 12-6b. This traditional switch, commonly used in Gibson guitars, while more compact than the lever switch used on the Fender Tele, is much larger than a modern, sealed toggle switch. A key advantage of an open toggle switch over a closed switch is that the contacts can be cleaned. Of course, a sealed switch is less likely to require cleaning.

Passive Magnetic Pickups

A guitar's tone is a function of hundreds of variables, from the thickness of the pick and composition of the neck, to the type of paint on the body. However, the magnetic pickups are the heart of the Tele's distinctive sound. They are also the most varied, most valuable, and least understood components on the Tele.

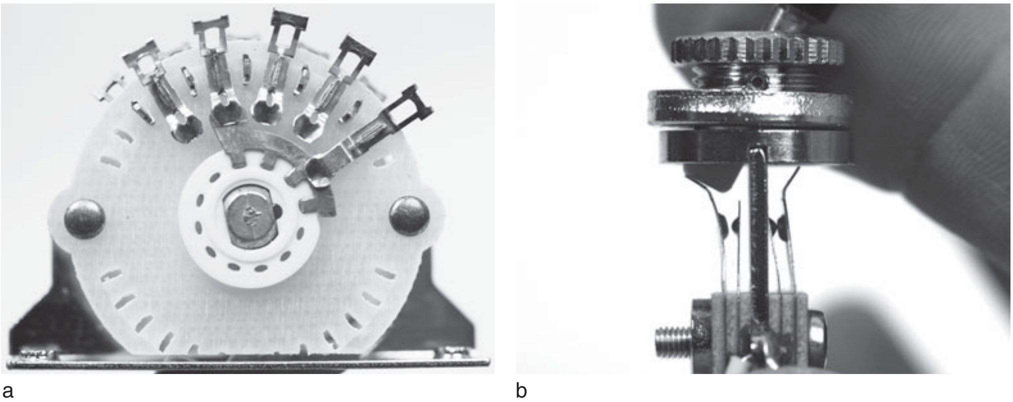

FIGURE 12-6 Four-way lever switch and three-way toggle switch

Passive magnetic pickups are available in hundreds of different models from dozens of manufacturers and an equal number of small artisan shops that cater to the custom market. Most of these pickups share the basic structure of one or more permanent magnets surrounded by a coil of copper. When a steel guitar string enters or vibrates within the magnetic field surrounding the magnet(s), a current is induced in the coil. The signal is routed to the controls, out the mono audio jack, and to the input of the audio amplifier.

Magnetic guitar pickups are based on the same effect used in dynamic microphones, in which a diaphragm moves either a magnet or a coil when sound hits the diaphragm. However, in the case of a guitar pickup, the physical plucking of a string provides the relative motion within the magnetic field.

From an electronics perspective, pickups are commonly classified as active or passive, single coil or double coil (humbucking), high or low impedance, high or low output, and magnetic or piezo. Pickups are further classified by type and configuration of the material(s) used in their construction, whether they are embedded in wax or epoxy, and whether the assembly is open or covered with a grounded shield. Now let's examine a passive magnetic pickup in detail with a mini-teardown.

Magnetic Pickup Mini-Teardown

You can follow the pickup teardown shown in Figure 12-7 at home if you don't mind destroying the pickup. A good pickup can account for up to a third or half of the cost of a guitar, and, other than requiring remagnetization every decade or two, they last forever. I frequently replace a pickup with one of a different design to change the tone of a guitar, but I carefully store the original pickup for future use.

Let's focus on the Tele's smaller, single-coil neck pickup. On inspection, you should see the six round ends of cylindrical pole pieces—one per guitar string. On a Fender pickup, the pole pieces are magnets. Other pickup designs use steel pole pieces to focus the magnetic fields generated by permanent magnets attached to the pole pieces. In either case, the movement of a string above a pole piece disrupts the large number of magnetic lines of force focused through each pole piece.

Let's examine the magnetic field surrounding the pickup. Because we can't sense the field directly, we have to rely on secondary indicators. The simplest approach is to use magnetic field viewing paper, which is a plastic sheet containing iron filings suspended in oil, available from Edmund Scientifics (www.scientificsonline.com). You should see a concentration of magnetic lines of force—that is, greater accumulations of iron filings—immediately above each pole piece.

An alternative to the above qualitative approach is to measure the magnetic field strength with a probe containing a Hall effect chip. The Hall effect is the production of a voltage difference across a conductor, transverse to an electric current in the conductor and a magnetic field perpendicular to the current. If you've used one of those clamp-on current probes, you've worked with a Hall effect probe designed to

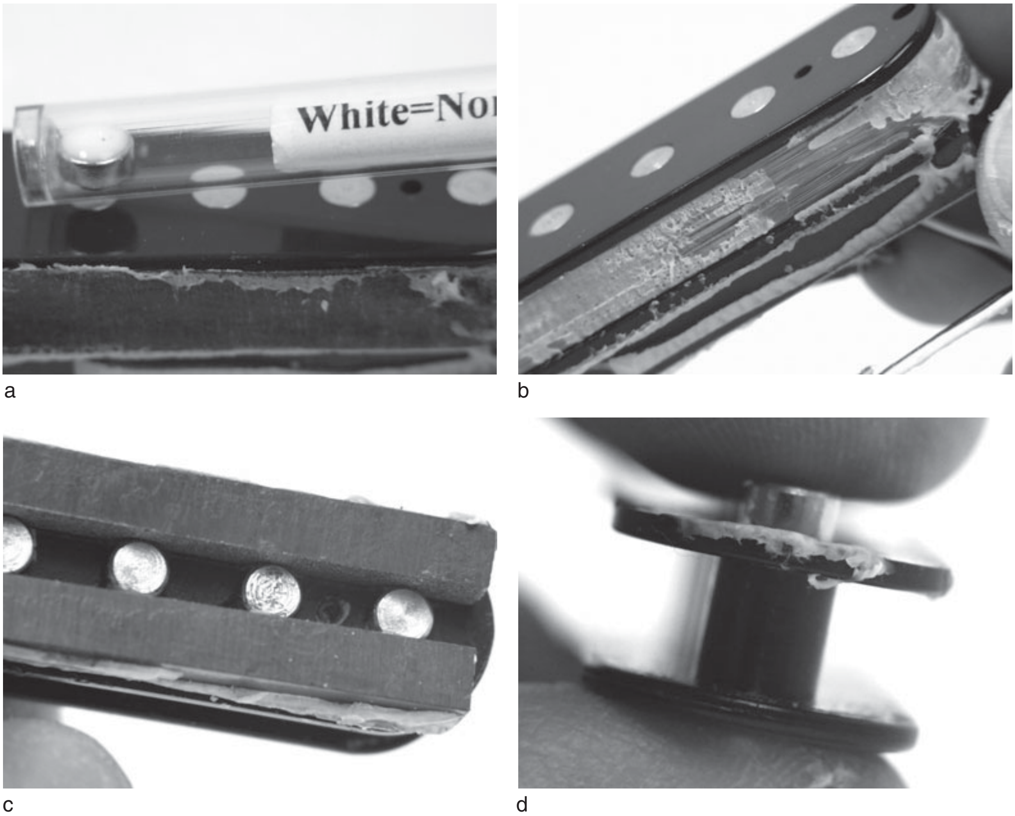

a

b

c

d

FIGURE 12-7 Mini-teardown sequence

measure current. Unfortunately, a precision Hall effect magnetic field probe, such as the B&K PR 26M (www.bkprecision.com), sells for thousands of dollars.

Regardless of how you measure it, what matters is that the stronger the magnetic field, the greater the voltage output from the pickup. Measuring the absolute field strength is more important with older pickups, because magnets degrade with time.

Step by Step

Because of the apparent simplicity of construction, you might get the impression that you're disassembling a simple solenoid or speaker coil. However, subtle differences in pickup construction can result in significant tone differences. For example, as you step through this mini-teardown, notice the windings. Are they regular within a layer? What about between layers?

Step 1

For our purposes, the most relevant magnetic field parameter is polarity, and this is relatively easy to measure. I use a commercially available detector from Stewart-MacDonald—a magnetized steel ball with one pole painted black and one white, contained within a plastic tube. A small keychain compass works almost as well. With your magnetic polarity detector, determine the polarity of each of the pole piece slugs, as shown in Figure 12-7a. If you have a modern Fender Tele, both the bridge and neck pickups should be "south up," where "up" means facing out, toward the steel guitar strings. If you're lucky enough to have a Fender Tele from the early 1950s, the neck pickup should be "north up."

The relative magnetic polarity of each pickup determines the phasing of the pickups, and therefore how the white and black output wires should be connected to the controls for a proper, in-phase signal. If you wire your neck and bridge pickups so that they're out of phase, the signals will partially cancel each other. Each pickup will sound fine by itself, but when the selector switch connects the pickup wires in series or parallel, the resultant output signal will be weak and tinny sounding.

Documenting a pickup's magnetic polarity is useful beyond verifying proper phasing. Polarity also indicates whether the original pickups have been modified—an important consideration when you're verifying the authenticity of a guitar at a garage sale or a guitar shop. For example, if you have been following along closely, you can see in Figure 12-7a that the polarity of the pickup is south up. This is inconsistent for a modern Fender pickup.

There's a good reason for this discrepancy. The pickup in the photo is a single-coil neck pickup from a Gibson Melody Maker that is physically similar to a Fender Tele pickup, but it uses a reverse polarity. I couldn't bring myself to destroy a perfectly good (and expensive) Fender pickup, when I had this relatively inexpensive Gibson pickup to use for a teardown. Besides, even if I had used the stock Fender pickup, it may not have resembled the neck pickup of your Tele if you're following at home.

Pickups used in the Tele differ by model and year of manufacturer. For example, some Tele neck pickups are encased in a metal shield that is connected to ground. Others' designs shield the pickup from 60Hz hum by wrapping the pickup with insulating cloth tape and then wrapping the coil with conductive copper tape, which is then connected to ground. In our teardown, we have a simple, single-coil pickup without a metal shield or copper tape.

Step 2

Now, remove the black paper or cloth tape from the pickup, exposing the 7800 to 8000 windings of fine enameled copper wire, as shown in Figure 12-7b. Tele neck pickups are typically constructed of fragile 43AWG enameled wire, whereas 42AWG wire is used for the bridge pickup. The 43AWG wire is used in the neck pickup because it is shorter than the bridge pickup—the guitar strings are closer together near the neck and wider near the bridge. To get enough windings around the neck pickup, the manufacturer uses thinner wire. The white substance on the surface of the windings in the photo is wax, which minimizes vibration and seals the unit from moisture.

If you have a stock Fender pickup, such as the Vintage Noiseless pickup, the six pole pieces are Alnico (aluminum, nickel, cobalt) magnets. However, if you're following along at home with a non-stock Tele or another guitar, you might encounter one of dozens of possible pickup designs. For example, your pickup may include two rectangular Alnico bar magnets that run on either side of the pickup, as in Figure 12-7c. The magnets make contact with each of the six steel pole pieces. Remove the bar magnets, if present.

Step 3

Remove the wire from the pickup to reveal the plastic bobbin, as shown in Figure 12-7d. If you want to save the wire by unwrapping it, give yourself an hour or two to complete the task. And you're likely to break the fragile 43AWG wire. Alternatively, attack the wire with sharp diagonal cutters and dispose of the short wire clippings. If you're interested in winding your own pickups, 42AWG and 43AWG enameled wire is readily available from several guitar supply houses.

Although the wire used to create magnetic pickups is fragile, it enables pickup designers to pack a lot of inductance into a small space. Consider the Fender Vintage Noiseless pickups, which are single-coil pickups based on Alnico 5 magnets. The 3.7H neck pickup and larger 4H bridge pickup contribute to the Tele's characteristic twang. Unfortunately, the single coil pickup design also works wonderfully as a 60Hz antenna.

Pickup manufacturers address the hum problem by using humbucker pickups, which are designed with two coils of wire and two magnets. A humbucker is constructed so the two coils of wire are out of polarity with each other and each coil is wound around a magnet of opposite polarity. Because the coils are in close proximity to each other, 60Hz currents induced in one coil are induced in the other coil. However, because the coils are wound in opposite polarity, the 60Hz signals are 180 degrees out of phase, canceling the hum.

Consider the contribution of the magnets of opposite polarity. The 60Hz currents are induced by an external electromagnetic field and are not affected by the presence of the magnets. However, because the magnet polarities are reversed, the voltage induced by a vibrating guitar string is doubled. Humbucking pickups not only increase the SNR of a guitar's pickup system, but they create a thicker tone than single-coil pickups. As such, if you replace the Tele's single-coil pickup with a humbucker—a common mod—you'll lose some of the Tele's characteristic biting tone. And that may be a good thing. As with most mods that affect a guitar's tone, it's a question of personal taste.

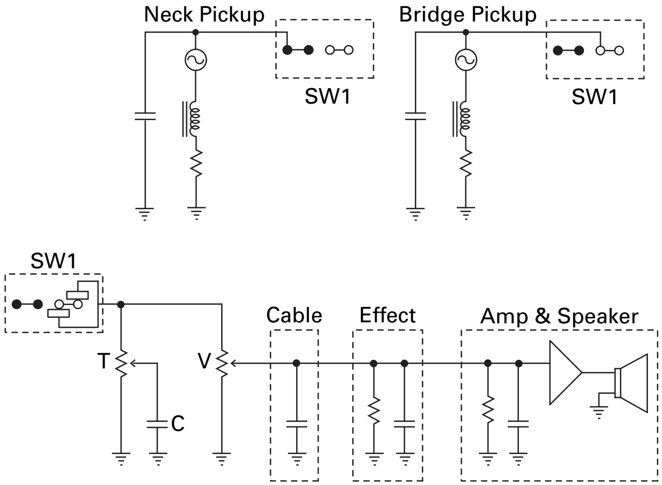

FIGURE 12-8 Simplified schematic of the Telecaster and typical setup

How It Works

An electric guitar is like a CD or DVD player in a home entertainment system, in that the listening experience depends as much on the speakers, amplifiers, mixers, and other devices as it does on the signal source. As such, in considering the operation of our Tele, we have to assume a typical system configuration—at least one cable, perhaps an effects pedal, and an amplifier with a built-in or external speaker system, as in the schematic shown in Figure 12-8. For our purposes, the Tele can be reduced to a magnetic neck pickup, magnetic bridge pickup, pickup switch (SW1), tone potentiometer (T), volume potentiometer (V), and tone capacitor (C).

The electrical equivalent of a pickup is an AC signal source, series resistance, inductance, and parallel capacitance. Recall from our teardown that no resistors or capacitors are associated with a passive magnetic pickup. The resistance is due to the DC resistance of the fine copper wire—typically around 10KΩ for a single coil Tele pickup. The capacitance is the equivalent capacitance between coils of wire. Most references that I've seen suggest that a normal range for coil winding capacitance is 100 to 200pF. As noted earlier, the published inductance values for single coil pickups range from 2 to 5H.

The significance of the equivalent circuit for each pickup is that they appear as classic parallel LC circuits. Recall that at resonance, a parallel LC circuit presents a high impedance and corresponding high voltage across the inductor and capacitor. The lower the value of internal resistance, the sharper the peak of resonance. Conversely, with increased resistance, the resonance peak becomes less pronounced.

If we assume a Fender Vintage Noiseless neck pickup with an inductance of 4H and an equivalent capacitance of 160pF, then the resonant frequency of the neck pickup in free space can be calculated as follows:

$$F = 1 / [2p\ (L\ C)1/2]$$

$$F = 1 / [6.28 \times (4 \times 160 \times 10^{-12})^{1/2}]$$

$$F = 6291\text{Hz}$$

Similarly, the theoretical free space resonant frequency of a 3.7H bridge pickup with an equivalent capacitance of 160pF can be calculated as follows:

$$F = 1 / [6.28 \times (3.7 \times 160 \times 10^{-12})^{1/2}]$$

$$F = 6541\text{Hz}$$

Note If you don't have a calculator handy, a nice LC resonance frequency calculator is available at What Circuits (www.whatcircuits.com). Try plugging in values for the range of equivalent capacitance from, say, 50 to 200pF, and notice the resonant frequency.

Returning to the schematic in Figure 12-8, you'll see that the pickups are not in free space. They can be connected in parallel with each other and with the other systems' components. The capacitance across the inductance of each magnetic pickup includes a contribution from the 0.05µf tone capacitor (C) as well as the equivalent capacitance from the cable, effects pedal, and amplifier input circuitry. The typical cable contributes 200 to 500pF, depending on length and manufacturer. Cable resistance is insignificant, given the low signal levels involved. Effects pedals easily contribute another 100 to 200pF to the system. In addition, pedals that don't completely disconnect when powered off contribute as little as 500KΩ of shunt resistance. The typical amp input also contributes about 1M of shunt resistance and 100pF capacitance. All this has an effect on the tone of the system, namely shunting high frequencies to ground.

The resonant frequencies of the pickups are significant in that they're within the most sensitive band of human hearing. As such, the nonlinearities in frequency response they introduce into the electric guitar system will likely be evident to the listener. However, it's difficult to quantify how the other system elements connected to the guitar can shift the resonant frequency of the system and therefore the tone. You could use SPICE (Simulation Program with Integrated Circuit Emphasis) or a similar tool to simulate the circuit, but you'd be hard pressed to provide the exact electrical parameters of your specific setup—your guitar and pickups, your cables, your effects pedals, and your amp.

For example, it's almost impossible to find capacitance and resistance loading values for effects pedals that don't totally disconnect from the circuit when not in use. And cable manufacturers like to use qualitative superlatives in defining their guitar cables, but few list actual specifications. One exception is Lava Cable (www.lavacable.com), which manufactures cables with capacitance of 25 to 50pF per foot.

To get a feel for the relative contribution of components, let's run a few experiments with the aid of a signal generator and oscilloscope. Since we know the current through a parallel LC circuit will be minimum at resonance, we can apply a constant amplitude audio signal across a pickup and sweep the frequency. A dip in current should indicate resonance.

As shown in the plot in Figure 12-9, the resonance dip in current for the Fender Vintage Noiseless neck pickup is about 3kHz, which is significantly different from the free space resonant frequency calculations. Differences in theoretical and measured values could be due to differences between the actual versus published inductance values and assumed capacitance values. Notice that for illustration purposes, I cleaned up the actual plot, which wasn't as pronounced or clean as that shown in the figure. The change in current at resonance is only on the order of a few milliamps with a 5V sweep generator.

To determine the resonant frequencies of the guitar as a whole, you could use the same method of sweeping a constant amplitude audio signal, but this time with the pickups mounted in the guitar body, with the tone capacitor in the circuit. The audio signal could be input and measured at the audio jack. Similarly, you could measure the resonant frequency of your system as a whole—say, for example, the Tele body, a 30-foot Fender coil cable, an MTX distortion pedal, and a Fender Champion tube amp. You could then substitute components and measure the resonant frequency.

At issue, of course, is the value of these measurements beyond our interest in electronics. We can't assume that our objective measures will reflect the quality of tone produced by the system, and therefore the value to the guitar designer or player. Perhaps you can now appreciate why Fender hasn't changed much in its original Tele design, including the use of antiquated components and wiring techniques. This isn't to say that the guitar shouldn't be modded for fear of

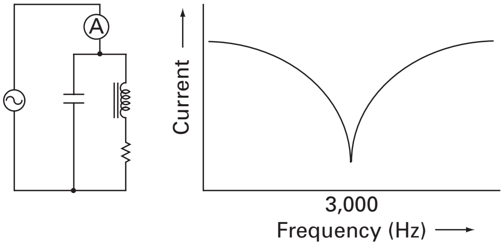

FIGURE 12-9 Idealized dip in current at resonance

introducing some unknown variable into the system and wrecking the tone. On the contrary, the beauty of the Tele is that the design is so simple that most mods can be undone in a few minutes, without any permanent damage.

Mods

Despite the Tele's apparent mechanical simplicity, you might be amazed at the tonal differences that can result from subtle changes in components and wiring. In determining which of the following mods are for you, it's worth repeating that tone is a matter of taste and music style. A stock Tele may be all you ever need—but then perhaps you won't be able to resist manipulating the electronics to create a better tone. If so, here's a list of popular electronic mods for the Tele.

More Pickup Switching Options

One of the most popular and inexpensive mods is adding a serial pickup option by swapping a four-way selector switch for the stock three-way switch. The serial, in-phase tone is distinctive and useful with a variety of styles.

Compensated Volume Control

At low volume settings, the highs in the guitar signal are attenuated disproportionally. A work-around is to set the volume potentiometer to maximum volume and adjust overall volume by adjusting other system components. A popular mod is to use a treble bleeder capacitor—a 0.001μf ceramic disc capacitor—from the high side to the wiper arm of the volume pot. The capacitor provides a low impedance path for high frequencies, regardless of the pot's position.

Functional Tone Control

As noted earlier, a shortcoming of the Tele's design is the suboptimal tone pot. It's a common practice to set the potentiometer to 250KΩ for maximum treble and to forget about it. However, even at maximum, high frequencies are attenuated. For example, at 3000Hz, the total impedance to ground is essentially the value of the pot, given the capacitive reactance is negligible:

$$\begin{aligned}\text{Impedance} &= R + X_C\\ &= 250K\Omega + 1/[2\pi \times F\ C]\\ &= 250K\Omega + 1/[6.28 \times 3000 \times 5 \times 10^{-8}]\\ &= 250K\Omega + 1.1K\end{aligned}$$

The obvious way to maximize the high frequency output of the Tele is to use a higher value tone pot. Substituting a 500KΩ or even a 1MΩ potentiometer for the 250KΩ potentiometer is a common mod.

A commercial mod that does away with the tone potentiometer and 0.05µf tone capacitor, Stellartone, is based on switched capacitors instead of a pot. I've used the Stellartone (www.stellartone.com) on two Teles with good results. However, as with the standard tone pot, I tend to use it wide open, and I use an effects pedal and amp controls to vary the relative contribution of highs to the final tone.

Pickup Variations

You probably don't have enough time or money to try all the various pickups available for the Tele. In this teardown, we touched on the most popular configurations of passive magnetic pickups. In addition to a variety of passive magnetic pickups, there are active magnetic pickups and piezo pickups, both of which may have onboard amplifiers and impedance-matching circuitry that require battery power. Pickup variation mods range from using a different variety of passive pickups to a combination of active magnetic and piezo pickups with a passive magnetic pickup.

Passive Magnetic Pickups

In addition to the stock pickups from the big-name guitar manufacturers, several third-party manufacturers specialize in passive magnetic pickups for specific guitars. Figure 12-10 shows a popular replacement neck pickup for the Tele designed by Kent Armstrong (www.kentarmstrong.com). The pickup, produced in Korea, is more expensive than the equivalent Fender pickup, but it's more affordable than the hand-built pickups available from boutique shops. If you look carefully at the figure, you can see four leads—one pair for each coil in this hum-canceling pickup.

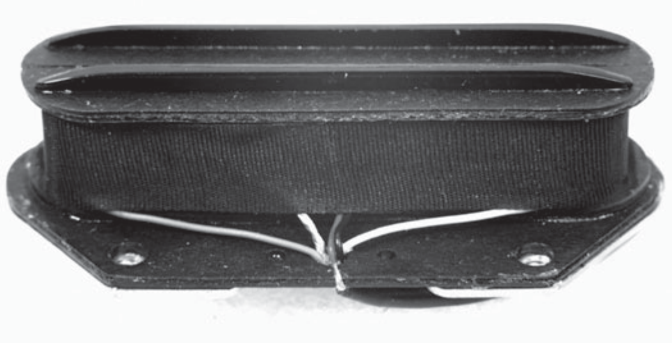

FIGURE 12-10 Kent Armstrong pickup

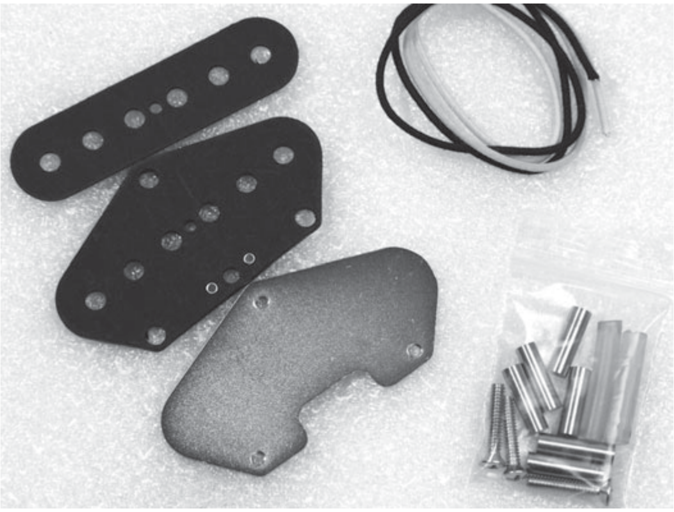

FIGURE 12-11 Single-coil kit

If you can't find or afford the passive magnetic pickup you need, or you simply can't resist working with 43AWG wire, then another option is to wind your own. Figure 12-11 shows a single-coil kit for the Tele that I purchased from Steward-MacDonald (www.stewmac.com).

The kit comes with Alnico 5 magnet pole pieces and the plates to hold the pole pieces in place. You'll need some 43AWG wire and steady hands to complete the pickup without kinking or snapping the wire. You might as well invest in a full roll of wire at the start, given your only option in dealing with a broken wire is to start over. Even if you don't build your own passive magnetic pickup, it's worth visiting the Stewart-MacDonald site to review the specs for common pickups, including the number of turns of wire, magnet polarity, and type of magnet.

Active Magnetic Pickups

Passive magnetic pickups are limited in the signal strength they can achieve. Increase the strength of the magnets and the steel strings are pulled into the pickup, damping the notes. Move the strings closer to the pickup and you have the same problem. One way to address this and other shortcomings of the passive magnetic pickup is to install an amplifier and interfacing circuitry at the point the signal wires exit the magnetic pickup coil. The leader in the active magnetic pickup market is EMG (www.emginc.com).

Active magnetic pickups are considered a "hot" signal source, with a typical output of 1.5VAC. In addition, they are usually shielded internally. Both factors contribute to an SNR that is superior to that of a typical passive magnetic pickup.

Because active pickups have little magnetism compared to passive magnetic pickups, strings can be positioned much closer to the pickup, which can increase playability.

Aside from cost—about double that of a passive magnetic pickup—the major downside of using an active magnetic pickup is the need to supply power, typically 9VDC at 80μA. The issue is typically lack of space for a 9V battery. My suggestion is to use a router to create a cavity in the guitar back and install a battery holder. Steward-MacDonald and WD Music Products (www.wdmusic.com) sell battery holders and router guides. In addition, if you upgrade to active magnetic pickups, you'll have to replace the standard 250KΩ potentiometers with 25KΩ potentiometers.

Piezo Pickups

Piezoelectric, or piezo, pickups are vibration sensors. They're standard issue on most acoustic-electric guitars and increasingly common on hybrid magnetic-piezo electric guitars. Piezo pickups rely on the piezoelectric effect, in which a voltage is produced when a crystal is stressed. Because of their structure, piezo pickups are relatively impervious to magnetic fields, including those that cause hum. The high impedance output of a piezo pickup can be sent directly to an output jack. However, it can't be mixed with the output from a magnetic pickup without an active mixer/preamp.

Properly configured, a piezo pickup can transform a Tele or another electric guitar into a semi-acoustic guitar with shimmering highs. Moreover, with an onboard mixer, the high impedance signal from a piezo pickup can be combined in varying degrees with the signal from an active or passive magnetic pickup, resulting in unique tone combinations somewhere between an acoustic and a traditional Tele.

Examples of piezo pickups compatible with the Tele include the GHOST pickup system from Graph Tech (www.graphtech.com) and the Tune-O-Matic Powerbridge from Fishman (www.fishman.com). Onboard piezo mixers and amplifiers are available from L.R. Baggs (www.lrbaggs.com) and Fishman, among others. These amplifiers and mixers provide isolation and impedance matching for the high-impedance piezo pickup, and typically provide auto-mixing as dictated by a smart 1/4-inch output jack. If the jack senses a stereo plug, magnetic signals are routed to one channel and the piezo signals to another. If it senses a mono plug, the mixer combines the signals from the piezo and magnetic pickups, with the relative contribution of each signal controlled by the position of the mixing pot.

A MIDI Interface

Transforming your Tele into a MIDI controller opens up myriad possibilities, from automatic transcription of music notation to controlling drum machines and synthesizers. You can add a MIDI interface to your Tele in two ways—the easy way and the hard way—and I've done both. The easy way, and the way I recommend if you're new to MIDI, is to purchase an external MIDI pickup, such as the Roland GK-3 (www.roland.com), shown in Figure 12-12.

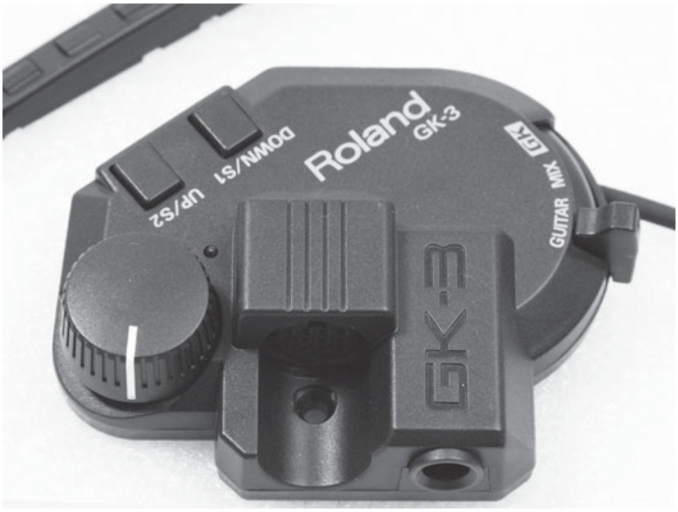

FIGURE 12-12 Roland GK-3 MIDI pickup

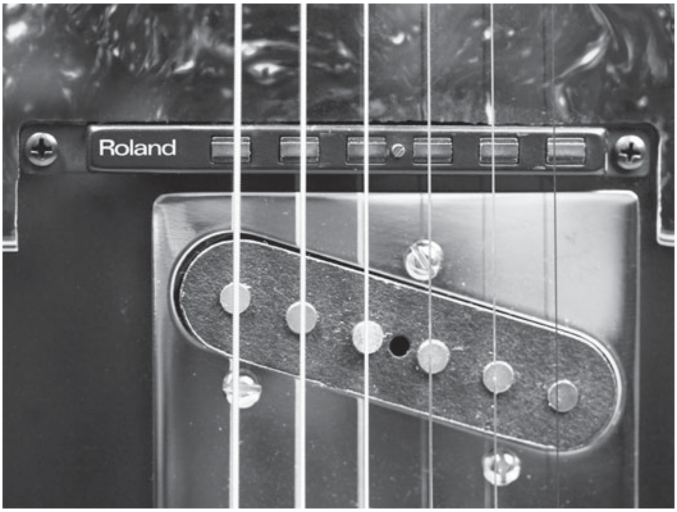

FIGURE 12-13 Pickup installed in reworked pickguard

To install the unit, mount the thin, active magnetic pickup immediately above the bridge plate and attach the GK-3 unit with either temporary adhesive or two wood screws, as shown in Figure 12-13. Notice that the pickup is composed of six separate pickups, as opposed to a single pickup with six pole pieces. Each of the six pickups responds only to the string immediately above it. In this way, the signals are easily parsed and it's possible to determine which string contributes a specific note during play.

Ideally, the active magnetic pickup should be mounted as close to the bridge as possible. The design of Tele's bridge plate is such that you have to mount the pickup just above the plate. However, at this position, the pickguard presses the pickup against the strings. The solution is to cut the pickguard with a jigsaw, coping saw, or band saw to make room for the pickup. Consider purchasing an inexpensive pickguard to cut and use with the MIDI pickup. You can reinstall the original pickguard if you decide to remove the pickup.

The hard—and most attractive—way to attach a MIDI interface to a Tele is to use the embedded version of the Roland pickup. Instead of working with double-sided tape, however, you'll need a drill, soldering iron, and the better part of a day. You'll need to drill three holes in the pickguard and in the body for switches. In addition, you'll have to add a third potentiometer between the tone and volume controls. More significantly, you'll have to bore a 1-inch diameter, 2-inch deep hole from the edge of the guitar to the control area for the 13-pin MIDI connector. And, to add an extra challenge to the effort, the three ribbon cables included in the standard kit aren't long enough for the Tele. You'll either have to extend the cables or find longer versions at an electronics supply house.

I didn't want to wait a week for delivery of longer cables, so I opted for extending the original cables by splicing in a few extra inches of ribbon cable. Fortunately, the operation was successful. With the exception of the thin magnetic pickup mounted just above the bridge, my Tele looks and handles like a standard electric when I'm not using it as a MIDI controller. Next time, however, I'll buy cable extensions and save a few hours of bench time.

This page intentionally left blank

Chapter 13

Effects Pedal

Guitarists who strive for a particular tone for each song or genre they play have two options: they can buy a Fender Stratocaster and a Marshall amp to play one song, a Gibson Les Paul and VOX amp to play another, and so on; or they can fake it with signal processing. Few musicians have the funds or space for the first option, making analog and/or digital signal processing the de facto means of morphing sounds.

Signal processors for guitars are available in a variety of configurations. They can take the form of a chip embedded in a guitar body or amplifier, a rack-mounted unit, or an effects pedal. An effects pedal, such as the one shown in Figure 13-1, is a signal processor in an indestructible case with a rugged on–off button. Effects pedals

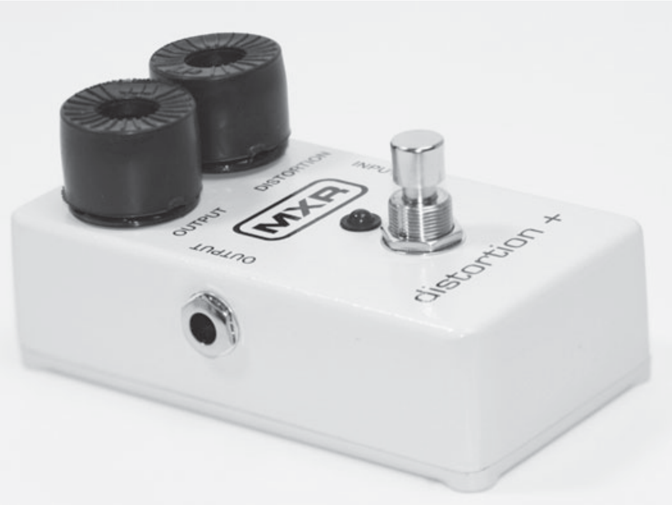

FIGURE 13-1 MXR Distortion+ effects pedal

are designed to interface with the musician's shoes and feet. The on–off buttons are stomped, and guitarists operate rubber-encased knobs with their feet—all without interrupting guitar play.

Effects pedals are available for almost every conceivable sound-processing task, from creating a classic tube sound to adding reverb, chorus, wah, or compression to the signal from a guitar. With a multi-effects pedal or handful of analog or digital single-effects pedals, a skilled guitarist can play authentic-sounding acoustic, hard rock, metal, country, jazz, and blues with a single electric guitar and amp.

In this chapter, I'll tear down a popular analog effects pedal, the Dunlap M-104 MXR Distortion+ pedal, referred to hereafter as simply "the pedal." As its name suggests, this popular signal processor introduces a controllable level of distortion into the guitar signal. The simplicity of the pedal's circuitry lends itself to modding, which I'll cover at the end of this chapter.

Highlights

Give yourself 10 minutes to complete the basic teardown. The pedal consists of an aluminum box that contains a well-laid-out circuit board with about two-dozen components. There's a 741 operational amplifier, a pair of 1/4-inch audio jacks, two audio-taper pots, switching diodes, metal film resistors, an LED, a variety of capacitors, and a foot switch.

During the teardown, note the following:

- **External power requirements** Effects pedals, like most other music electronics, use positive ground instead of conventional negative ground.
- **Layout of the circuit board and overall construction** Effects pedals are built to take abuse.
- **Component types and ratings** Try to date the design of the board, based on components.

Specifications

Effects pedal specifications are often written in qualitative music jargon. For example, Dunlap's materials for this pedal refer to "retro-authentic soft-clipped distortion tones" and "classic fuzz tones." While these terms may be understood by a guitarist, they lack the technical specificity required for an objective comparison of pedals from different vendors.

Many specifications of industry standards are simply understood. For example, virtually all effects pedals are configured so that the audio input jack is on the right side; the audio output jack is on the left; and the controls knobs, LED, and on–off button or pedal are on the top. Figure 13-2 illustrates this standard pedal configuration.

The effects pedals on either side of the Distortion+ pedal in Figure 13-2 are two of my favorites from Roland/Boss, a major manufacturer of single- and

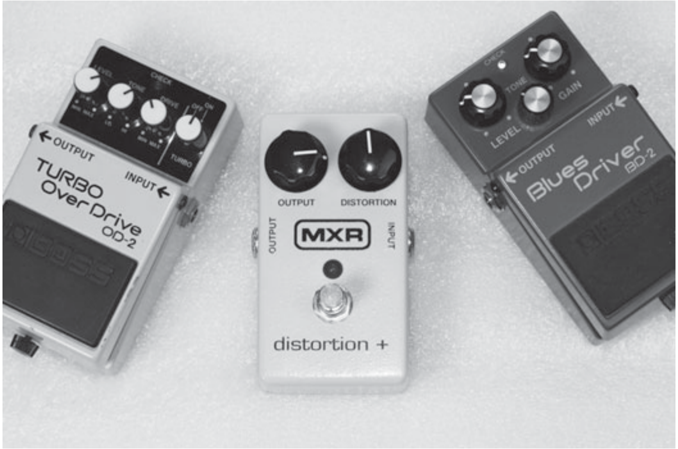

FIGURE 13-2 Standard pedal configurations

multiple-effects pedals. The Roland/Boss OD-2 Turbo Over Drive pedal is similar to the Distortion+ in how it modifies the guitar's signal, but it uses three times as many discrete components. The BD-2 Blues Driver pedal creates a much richer distortion than either of the other pedals, but it uses a high discrete component count.

As you can see from Figure 13-2, an advantage of a common right-input, left-output design is that pedals can be cascaded with short, direct cables. Typical effects pedals, including the three in the figure, present a nominal input impedance of 1MΩ and a nominal output impedance of about 2KΩ, and they require 9VDC (positive ground) at up to 500mA.

Operation

Since this is a stand-alone effects pedal with only two controls and an on–off button, as shown in Figure 13-2, operating the pedal is a no-brainer. Insert a 1/4-inch mono audio plug from your guitar or effects pedal into the input jack, and connect the mono plug to another effects pedal or a power amp into the output jack. Stomp on the large button to activate the circuitry, including the LED power indicator. Turn the DISTORTION knob clockwise to increase the amount of distortion in the signal. You can turn the OUTPUT knob to alter the volume of the guitar signal, without affecting the level of distortion.

Pedals are typically used as part of a sequence of five or six pedals. A typical pedal setup might include, from guitar to amp, a wah-wah pedal, chorus pedal, overdrive pedal, and delay pedal. Because of the interaction among pedals, the same pedals can be arranged to provide a variety of tone.

For example, if you place a reverb pedal before a distortion pedal in the signal chain, you'll get a very different tone compared to using a distortion pedal and then a reverb pedal. And substituting a pedal with a specific effect from, for example, Roland/Boss with one from Ibanez will result in a different tone.

Because of expense and complexity of using multiple, single-function effects pedals, digital, all-in-one, multi-function pedal boards are used by some guitarists. However, dedicated, single-function effects pedals set the standard for guitar signal processing.

Teardown

The teardown is illustrated in Figure 13-3. As you can see, it's a 5- to 10-minute operation. Even if you're following along at home with an effects pedal other than the Distortion+, once you remove the back and expose the circuit board, you're 80 percent there. The remaining 20 percent of effort is in analyzing the circuitry on the board.

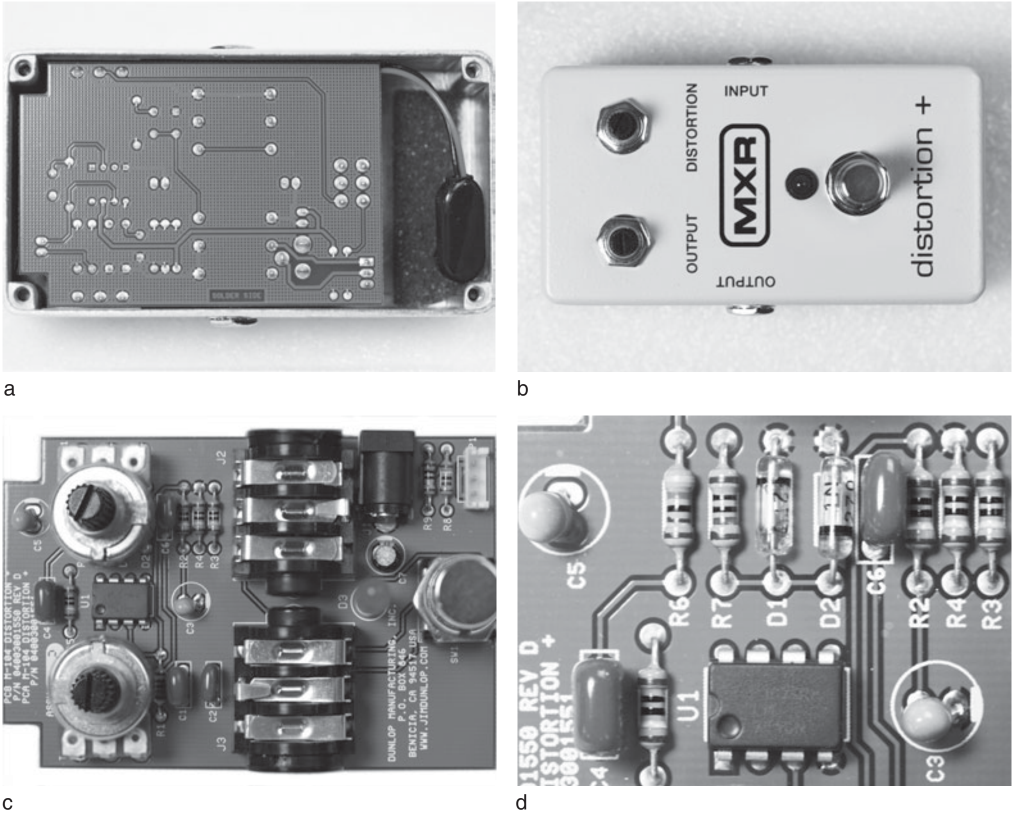

FIGURE 13-3 Teardown sequence

Hundreds of different effects pedals are on the market, and each manufacturer has its own way of achieving a particular tone. For example, the popular Roland/Boss distortion pedals use half a dozen transistors, whereas our pedal achieves similar tones with a single operational amplifier. The distortion pedals from different manufacturers produce similar but different tones. There is no clear-cut better tone—it's a mater of personal taste.

Tools and Instruments

A Phillips-head screwdriver, nut driver set, and crescent wrench are all you need for this teardown. If you intend to make a habit of removing retaining nuts from switches and buttons mounted on front panels, consider picking up a plastic nut removal tool from Steward-MacDonald (www.stewmac.com). Other pedals may have pressure screws in their plastic control knobs, requiring either a small Allen wrench or small flat-blade screwdriver.

A multimeter might help in tracing the schematic, but it's not necessary. A dual-trace oscilloscope would enable you to visualize the effect of the pedal at various distortion settings.

Step by Step

This teardown involves opening the metal box and freeing the controls. However, take your time. It's easy to scratch the paint on an effects pedal when you're removing the retaining nuts on switches and potentiometers.

Step 1

Remove the back cover. Extract the four machine screws with a Phillips-head screwdriver and set aside the back cover, shown in Figure 13-3a. If a 9V battery is installed, remove it from the battery cavity.

Step 2

Remove the external retaining nuts. Carefully remove the retaining nuts from the external jacks, switches, and potentiometers, which are visible in Figure 13-3b. You'll need a 7/16-inch nut driver for the input and output jacks, and a 13mm deep metric driver for the on–off button. Use the plastic nut removal tool to avoid scratching the paint when you remove the retaining nut from the on–off switch. The metal retaining nuts screw into the soft plastic of the input and output jacks—so avoid cross-threading the jacks when you reseat them.

Step 3

Extract the circuit board. Remove the 9V battery connector and lead from the board to avoid damaging the wires.

Step 4

Examine the board. The two potentiometers, input and output jacks, and push button switch dominate the circuit board, as shown in Figure 13-3c. Note the operational amplifier and associated components, shown in Figure 13-3d.

Layout

Thanks to the use of low-density leaded components, the layout of the circuit board is clean and the connections are easy to trace with the aid of a multimeter. As a bonus, component labels are stenciled on the board. Carefully bend the potentiometers out of the way and orient the board so that the on–off button is at the right, as shown in Figure 13-4. From this perspective, the components associated with the power circuitry—the power adapter jack, 100µf at 16V electrolytic capacitor, and 9V battery jack—are at the upper-right. The power-on LED is just to the left of the large button.

The normal signal path is from the right jack to the left jack, which is from top to bottom in this orientation. The jack on the top is a special sensing jack that cuts power to the unit when a plug is removed. The signal processing components, including a 741 operational amplifier, a pair of germanium diodes, 0.25W metal film resistors, and several capacitors, are in the upper-left area of the board. See Figure 13-5 for a close-up of the two germanium diodes and operational amplifier.

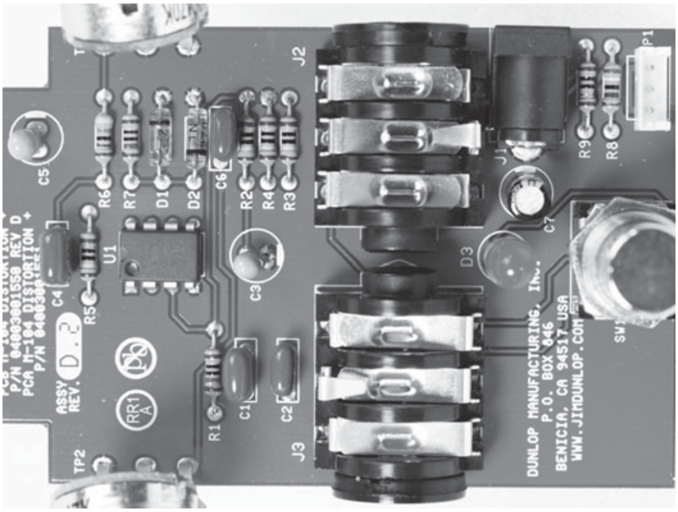
J2
P1
R6
R7
D1
D2
C6
R2
R4
R3
R9
R8
J1
C5
U1
C7
D3
C4
R5
C3
PCA M-104 DISTORTION
ASSY REV. D.2
RR1 A
C1
C2
R1
J3
TP2
SW1
DUNLOP MANUFACTURING, INC.
P.O. BOX 846
BENICIA, CA 94517 USA
WWW.JIMDUNLOP.COM

FIGURE 13-4 Component layout

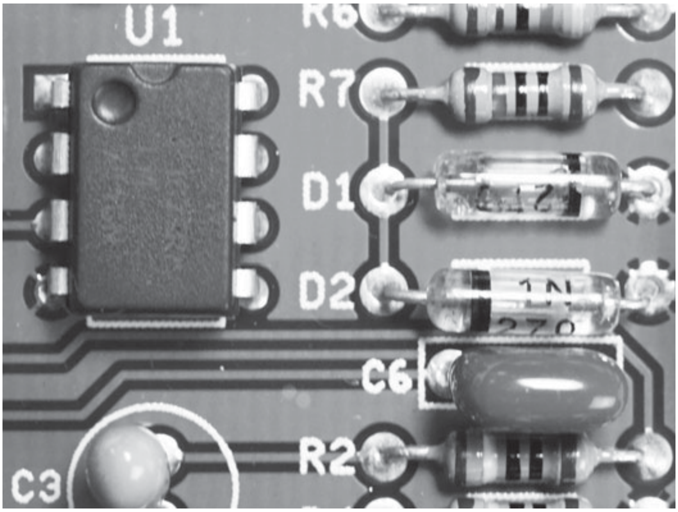

FIGURE 13-5 Germanium diodes and operational amplifier

Components

The components in this effects pedal are hardly representative of the cutting edge of electronics. As with many guitars, you'll find the use of traditional, leaded components common in many effects pedals designed to provide a vintage tone.

Diodes

The centerpiece of the pedal is a pair of 1N270 germanium diodes, labeled *D1* and *D2* in Figure 13-5. Unlike typical silicon power diodes used in power supplies, germanium diodes are relatively fragile, and their uses are limited to low-power switching applications, RF input probes, and AM/FM detectors. The 1N270 can handle an average forward current of only 40mA and a peak inverse voltage of only 100V.

For our purposes, the most significant characteristic of germanium diodes is their relatively low forward conduction voltage. As shown in Figure 13-6, the relationship between forward voltage and current is nonlinear for both germanium and silicon power diodes. With increasing forward voltage, there is an approximately exponential increase in current flow. Figure 13-6 also illustrates that germanium diodes conduct when forward-biased about 0.3V, versus 0.7V for silicon diodes.

Diodes are "soft switches," in that they don't instantly transform from perfect insulator and perfect conductor at a specific forward conduction voltage. Figure 13-6 illustrates that both standard silicon and germanium diodes conduct in varying

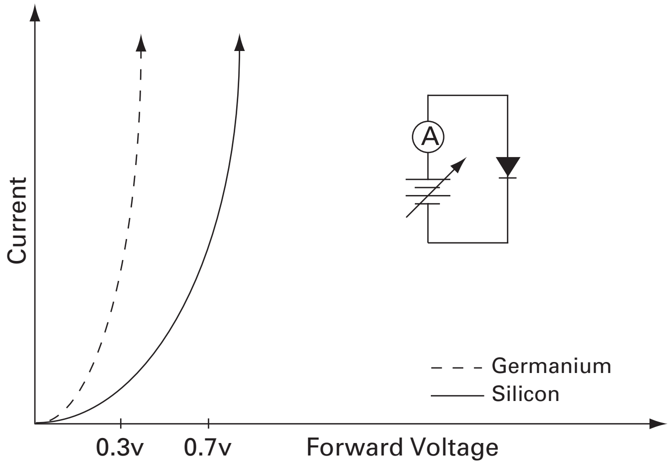

FIGURE 13-6 Silicon versus germanium diode conduction

degrees at 0.3V. Whereas the germanium diode is at nearly full conduction, the silicon diode is barely conducting. Both standard silicon and germanium diodes present high impedance to current when reverse-biased—until their peak inverse voltage is exceeded and an avalanche of current ensues.

The forward-bias conduction voltages of diodes are additive, in that if two germanium diodes are connected in series, cathode-to-anode, the series will conduct when forward-biased at 0.3V + 0.3V = 0.6V. Similarly, a germanium diode in series with a standard silicon diode will conduct when forward-biased at 0.3V + 0.7V = 1.0V.

Note You may be wondering why I keep referencing "standard" silicon diodes. I do this because a special class of silicon diodes called *Schottky* or *hot carrier* diodes are used for high-speed switching. The forward voltage drop is typically between 0.15 and 0.45V and the on–off switching time is often an order of magnitude less than that of standard silicon diodes. However, most effects pedals use either standard silicon or germanium diodes, so we'll focus on those technologies for now.

The fixed forward conduction voltage of a diode or series of diodes can be applied to a variety of applications. One use is to create a constant voltage drop, independent of current. For example, to run a 5.0V circuit from a 6.0V battery, you could insert a silicon power diode in series with the battery, providing the circuit with 6.0V – 0.7V = 5.3V. This may be within the maximum voltage limits for the circuit. Alternatively, you could connect two diodes in series, providing a working voltage of 6.0V – (0.7V × 2) = 4.6V, again, assuming this voltage will work for the circuit.

Another use for the forward-bias voltage drop is to create a clipper or limiter to constrain the amplitude of an AC signal. Within the power-handling capabilities of a diode, when a voltage in excess of the forward conduction voltage is applied to a diode, the device appears as a low-resistance shunt. A series resistor can extend the operating range of a diode clipper or limiter by protecting the diode from excessive current.

A single germanium diode will limit the amplitude of an AC signal applied across it to 0.3V, but only during the part of the cycle when the anode is positive with respect to the cathode. Virtually no current flows when the diode is reverse-biased. If symmetrical clipping is required, two diodes can be used in parallel, arranged anode-to-cathode. When one diode is forward-biased, the other is reverse-biased.

Circuit Board and Enclosure

Effects pedals are known as "stomp boxes" for good reason—they're designed to take abuse. Unlike many other electronics devices, they're intended to be stepped on repeatedly, to withstand an occasional dousing with soda or beer, and to have their power and signal chords yanked out mercilessly. As a result, the circuit board, jacks, and all-metal enclosure of this pedal are substantial. The aluminum enclosure not only provides structural integrity and a hefty heat sink, but also EMI shielding to and from other electronic devices.

The circuit board is also multilayered, meaning a layer of conductive traces are sandwiched between the top and bottom layers. While a multilayered board makes tracing connections between components more challenging, it often includes a large grounded area that shields the circuit from noise.

Resistors

This pedal uses a 47KΩ audio taper pot for the output control, a 470KΩ audio taper pot for the distortion control, and about a dozen metal film fixed resistors. Metal film resistors are appropriate for effects pedals in general because they produce less noise than cheaper carbon resistors. Resistor noise introduced at this point in the signal chain may be amplified several-fold by other pedals and the power amp.

This pedal is a low-power device that could have been produced with SMT resistors and other components. However, given that the size of the enclosure is already as small as practical, and that the input and output jacks need a substantial board for physical support, there is no obvious reason to develop a more compact board.

Operational Amplifier

This pedal is built around the popular LM741CN, a contemporary version of the 741 operational amplifier, or op amp, introduced in the 1960s. This ubiquitous

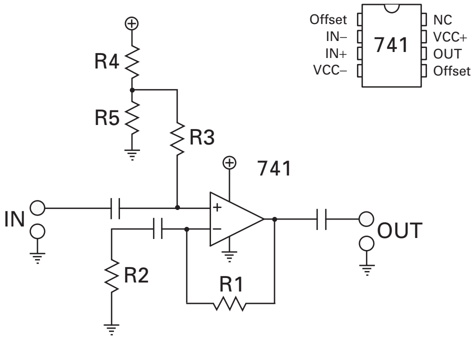

FIGURE 13-7 Noninverting AC operational amplifier circuit

component, containing 20 transistors and about a dozen resistors, has several features that make it ideal for use in an effects pedal. It can be configured as a high-input impedance, low-output impedance amplifier, and gain can be easily adjusted using a few resistors in a feedback network. Moreover, the bandwidth is more than adequate for audio work.

Figure 13-7 show a generic noninverting AC amplifier based on the 741 operational amplifier. The operational amplifier has two inputs, a noninverting input (+) and an inverting input (–). Because the guitar input signal is fed to the noninverting input, the output will be in phase with the input. Conversely, if the input signal is sent to the inverting input (–), the output will be 180 degrees out of phase with the input. Both inputs are critical to manipulating the behavior of the operational amplifier through the external feedback provided by the voltage divider formed by R1 and R2.

My favorite heuristic for working with operational amplifiers, borrowed from *The Art of Electronics* by Horowitz and Hill, is this: "The output will tend to change in a way that forces the external feedback network to minimize the voltage difference between inputs." Armed with this heuristic, you can manipulate the gain of the circuit to suit your needs. You can also use it to understand the operation of any operational amplifier.

Let's start our analysis by defining the voltage gain of the generic AC amp in Figure 13-7 as the ratio of output voltage to input voltage. Note that the input voltage is applied to the noninverting (+) input of the operational amplifier. Mathematically, we have this:

$$\text{Gain} = V_{out}/V_{in}(+)$$

Now assume we apply 1.0V to the noninverting (+) input of the operational amplifier. Based on the heuristic, we know that the output feedback will shift such that 1.0V appears at the inverting input (–) of the operational amplifier. Given the circuit configuration, the only way for 1.0V to appear at the inverting input is for the voltage divider formed by R1 and R2 to deliver 1.0V.

Recall that in working with two resistors in series, the voltage drop across a given resistor is proportional to the ratio of resistance contributed by that resistor to the total resistance. For example, consider R1 and R2 in Figure 13-7. It should be intuitive that when R1 = R2, the AC output voltage of the operational amplifier is equally divided between R1 and R2. Similarly, as the value of R1 approaches zero, the voltage drop across R2 approaches the AC output voltage of the operational amplifier. Mathematically, this can be expressed as follows:

$$\mathrm{VR1} = V_{\mathrm{out}} \times \mathrm{Rl}/(\mathrm{R1} + \mathrm{R2})$$

$$\mathrm{VR2} = V_{\mathrm{out}} \times \mathrm{R2}/(\mathrm{R1} + \mathrm{R2})$$

Following these formulas, given R1 = 25KΩ and R2 = 75KΩ, for a total of 100KΩ, 25 percent of the AC output of the operational amplifier appears across R1 and 75 percent across R2.

Since the voltage drop across R2 is applied to the inverting (–) input of the operational amplifier, we now have this:

$$\mathrm{VR2} = V_{\mathrm{in}}(-) = V_{\mathrm{out}} \times \mathrm{R2}\,/\,(\mathrm{R1} + \mathrm{R2})$$

From our heuristic, we can assume that the values at the inverting (–) and noninverting (+) inputs are equal:

$$V_{\mathrm{in}}(+) = V_{\mathrm{in}}(-)$$

Now we can express the gain as a function of the output voltage:

$$\mathrm{Gain} = V_{\mathrm{out}}/[V_{\mathrm{out}} \times \mathrm{R2}/(\mathrm{R1} + \mathrm{R2})]$$

Simplifying, we have this:

$$\mathrm{Gain} = 1 + \mathrm{R1}/\mathrm{R2}$$

The gain of the generic AC amplifier circuit is defined by the ratio of R1 to R2. Now assume that we need an amplifier that delivers a gain of 10 and that R1 = 1MΩ. We can calculate the value of R2 as follows:

$$\mathrm{Gain} = 10 = 1 + \mathrm{R1}/\mathrm{R2}$$

$$10 = 1 + 1{,}000{,}000/\mathrm{R2}$$

$$9 = 1{,}000{,}000/\mathrm{R2}$$

$$\mathrm{R2} = 110{,}000\Omega$$

Let's work backward to check our work. Given a 1.0V input to the AC amplifier, at a gain of 10, the output voltage should be 10V. And, as we calculated, 10V across the R1–R2 voltage divider will deliver 1/9th of the output voltage, or 1.0V, to the inverting (–) input of the operational amplifier.

Let's consider the power requirements. The 741, like many operational amplifiers, is designed to operate as a dual-supply device, with positive, negative, and ground. However, providing users with both a positive and negative supply voltage with two batteries and a complicated AC power supply is expensive.

An economical workaround, illustrated in Figure 13-7, is to create a reference voltage from a single-sided supply (+V) and to ground the negative voltage supply lead. In this example, R4 and R5, of equal value, create a voltage divider that supplies a bias halfway between ground and +V, or about 4.5VDC with a 9V supply. Notice how capacitors are used to isolate the inputs and outputs of the operational amplifier from the DC supply voltage and ground.

Because of the high input impedance of the 741 operational amplifier, there must be a finite return path to ground for the input current. In our example circuit, R5, which is typically 1M, serves this function.

Capacitors

The use of capacitors in this pedal is typical of audio circuitry. A 100µf at 16V electrolytic capacitor is used to filter the 9VDC supply. Two 1µf at 35V tantalum capacitors and three Mylar film capacitors, from 0.001 to 0.047µf at 50V, are used for signal handling and bypassing. As shown in Figure 13-8, the 1µf tear-shaped tantalum capacitor is polarized. The capacitance value (simply 1 for 1µf), polarity, and voltage rating are stamped plainly on the capacitor body.

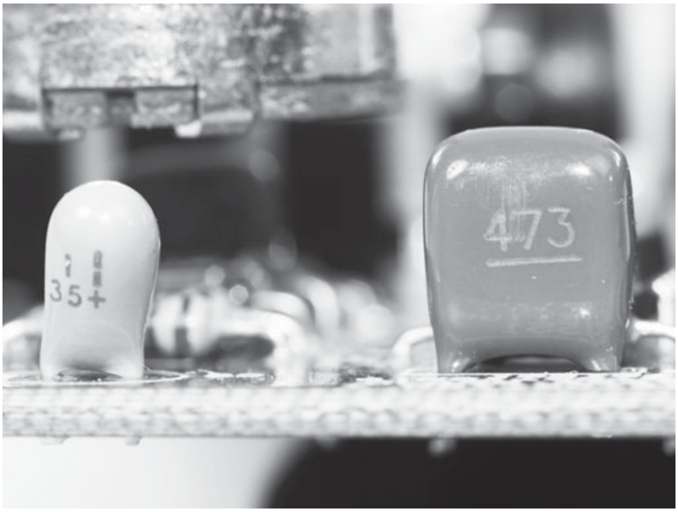

FIGURE 13-8 Tantalum (left) and Mylar capacitors

The Mylar film capacitors, which are dipped in a hard epoxy coating for moisture resistance, are nonpolarized. The code on a Mylar film capacitor indicates capacitance and multiplier. For example, the Mylar capacitor in Figure 13-8, marked *473*, has a capacitance of 47 and a multiplier of 3, or 0.047µf. Similarly, the 0.01µf Mylar capacitor in the pedal is marked *103*, and the 0.001µf capacitor is marked *102*. See Appendix A to review capacitance codes. It's important for you to recognize capacitor values, because much of modding pedals involves replacing capacitors with those of different types and values to achieve different tones.

Connectors

The major connectors in this pedal are for power and audio. The power jack is notable in that it accepts a 2.1mm, 9VDC plug (positive ground) that is standard throughout the pedal industry. If you try to power this or any other effects pedals with a standard negative ground wall wart, you risk frying your pedal.

The mono audio input jack is designed to disconnect power from the circuitry when the plug is removed. If you're debugging a pedal, make sure that you insert a plug in the input jack or you won't be able to activate the signal processing circuit.

This battery-saving feature sounds good in theory, but few musicians take the time to unplug their effects pedals after each practice. Unfortunately, the pedal on–off button is often not a true power on–off switch. For example, with a plug inserted in the input jack of this pedal, the current draw is 0.4mA with the on–off button in the off position. Removing the plug from the input reduces the current draw to zero.

How It Works

This pedal is fundamentally a high-input impedance, symmetrical signal clipper. The audio signal from a guitar is sent to the noninverting input of a 741 operational amplifier configured as an AC amplifier, as shown in the simplified schematic of Figure 13-9. You should recognize most of the major components from the earlier discussion of a generic AC operational amplifier circuit.

Symmetrical signal clipping is the same type of distortion associated with vacuum tubes operated near saturation. By creating symmetrical distortion, this pedal enables a guitar player to achieve the tone characteristic of a powerful tube amp run at full power while using a low-power practice amp.

Gain

As discussed earlier, the gain of the operational amplifier is varied by changing the ratio of R1, a 1MΩ resistor, to R2, the 470KΩ potentiometer. If R2 is set to 470KΩ, the gain of the circuit is as follows:

$$\text{Gain} = 1 + R1/R2 = 1 + 1{,}000{,}000/470{,}000 = 3.1$$

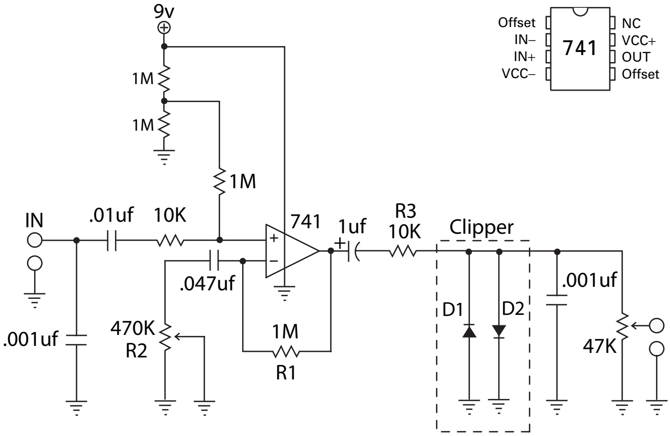

FIGURE 13-9 Simplified schematic of the pedal

With R2 set to 47KΩ, the gain is this:

$$\text{Gain} = 1 + R1/R2 = 1 + 1{,}000{,}000/47{,}000 = 21.3$$

The amplified signal is AC-coupled to a clipper circuit composed of two germanium diodes wired in parallel, anode-to-cathode. The output level is adjusted by varying the tap position of the 47KΩ pot. Because of the clipping effect of the diodes, the theoretical maximum output level of the pedal is ±0.3V, regardless of the gain or level settings.

Clipping

Clipping is performed by two germanium diodes, D1 and D2, protected by a 10KΩ current limiting resistor, R3. The degree of clipping depends on the amplitude of the signal presented to the two diodes, which depends on the setting of the 470KΩ potentiometer. For example, consider Figure 13-10, in which the input signal is approximately ±0.33V. The sinusoidal waveform is minimally clipped at the positive and negative peaks.

Now consider the waveform in Figure 13-11, in which the amplitude of the signal applied across the diodes is approximately ±0.6V. The resulting clipped waveform resembles a square wave, and the sound is harsher than that of the minimally clipped waveform of Figure 13-10.

Figure 13-12 shows an oscilloscope tracing of the pedal input and output signals with the controls set for maximum distortion. I used a BK Precision 4011 function generator to create the 400Hz sine wave input, captured the input and output signal data with a Tektronix TDS2022 100MHz oscilloscope, and plotted the data with a spreadsheet application. Note this figure's resemblance to Figure 13-11.

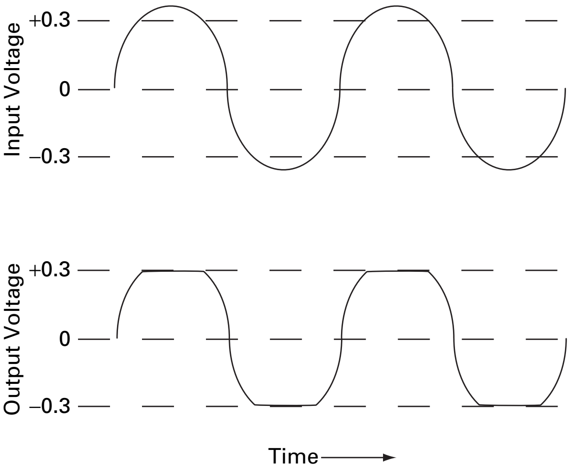

FIGURE 13-10 Minimal clipping

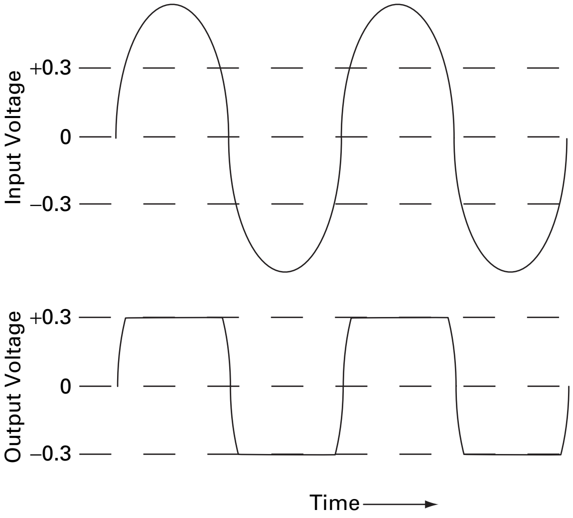

FIGURE 13-11 Pronounced clipping

From examining the figures, you should see that distortion is a function of amplitude of the waveform output from the operational amplifier, relative to the forward conduction voltage of the diode clipper circuit. The lower the resistance setting of the 470KΩ potentiometer, the greater the amplification provided by the operational amplifier, and the greater the distortion. It's also worth noting that the output of the pedal is a little greater than the theoretical limit of ±0.3V.

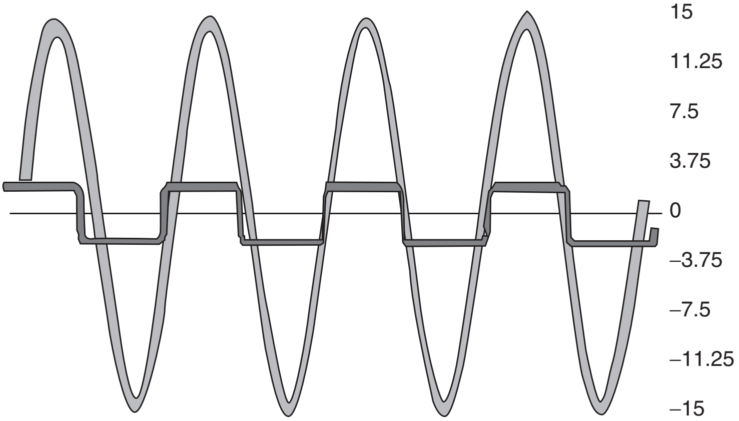

FIGURE 13-12 Oscilloscope trace of clipping

Impedance Matching

The operational amplifier serves as not only a variable gain amplifier to provide varying degrees of clipping, but it also provides a high impedance input for the guitar signal. An electric guitar with standard magnetic pickups is a low-power device designed to work with a high impedance load—on the order of 1MΩ. A lower impedance load will markedly diminish the guitar signal strength and adversely affect tone—notably, the loss of higher frequency signals. Feeding the guitar output directly to a diode clipper would be equivalent to shorting the guitar output through the relatively low-resistance current-limiting resistor.

Mods

There's a healthy economy in pedal modding. For $20, you can purchase a pedal-specific kit with a couple capacitors, diodes, some wire, and a one-page guide of where to install the components. It's common to replace silicon with germanium diodes, a Mylar capacitor with a polypropylene capacitor of higher or lower value, and to replace resistors with wire or resistors of higher value. The value of a kit is information on which capacitors and diodes to replace to achieve a given tone on a specific pedal.

Some generic mods will work with a range of pedals. Following is a list of generic mods that will work for distortion pedals based on a diode clipper.

Change Diode Configuration and Type

Consider different types and configurations of clipping diodes, such as those explained here and shown in Figure 13-13. Each of the clipping circuit configuration circuits depicted here provides a different output tone. The best tone—and therefore configuration—depends on the type of music you want to play and your personal preference. Most guitarists associate the tone produced by germanium diodes with that of vintage effects pedals.

Here's a key to the configurations shown in Figure 13-13:

- **A.** The default version of the pedal, with two germanium diodes, which produces a symmetrical clipping pattern at ±0.3V.
- **B.** A silicon diode paired with a germanium, resulting in an asymmetric clipping at 0.3V on one side of zero and 0.7V on the other.
- **C.** A pair of silicon diodes, resulting in symmetrical clipping at ±0.7V.
- **D.** A silicon diode paired with two germanium diodes in series, resulting in an asymmetrical clipping at 0.7 and 0.6V.
- **E.** A germanium diode paired with two germanium diodes in series, with an asymmetrical clipping at 0.3 and 0.6V.
- **F.** A silicon LED paired with a pair of germanium diodes in series, resulting in an asymmetrical clipping of 1.2 and 0.6V. LEDs conduct when forward-biased at 1.2 to 2.4V, depending on the LED design and composition.

In addition to diodes, you can try transistors and other components with nonlinear junction properties.

Tone Switch

Consider installing a multiposition rotary switch to select one of the diode configurations defined in Figure 13-13. Because of limited space, you'll probably have to mount the switch on the upper side panel of the pedal.

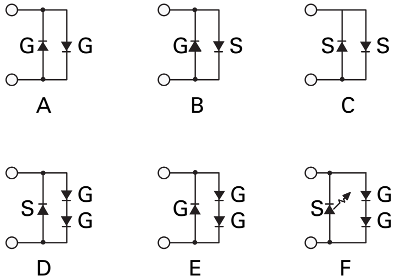

FIGURE 13-13 Clipping circuit configurations

Add True Bypass Switching

The pedal is never completely out of the signal line, even when the on–off button is off. There is a 0.001µf capacitor across the input connector and less than 2MΩ to ground through the bias voltage divider. The on–off button should disconnect these passive elements in the off position to avoid influencing the audio signal.

Try Different Operational Amplifiers

Operational amplifiers of the same type from different manufacturers and op amps of different types have different sound characteristics. To experience the range of possible sounds from your pedal, consider replacing the 741 with an 8-pin DIP (dual inline package) socket so that you can easily change and evaluate different op amps. Check the Web to find modern, high-performance and low-noise operational amplifiers wired to 8-pin DIP carriers compatible with the pin-outs of a 741 operational amplifier.

Chapter 14

Vacuum Tube Guitar Amplifier

In this chapter, I'll tear down the Fender Champion 600, or Champ—a compact 5W combo guitar amplifier that features two vacuum tubes and a 6-inch speaker, as shown in Figure 14-1. The modern Champ is a Chinese reproduction of a small practice amp produced in the United States in the 1950s. It's a favorite of the modding community because of its simple design, affordability, and ready availability of third-party kits and components. Musicians are drawn to the amp because it develops full distortion at only a few watts—a must for bedroom jam sessions when nonmusician parents or roommates don't want to be bothered by the noise.

The Champ retails for about $250, but I've picked up several used amps online for about half that—and spent the difference on various mods. I'll walk you through the circuitry and suggest a few mods that can be applied to most tube-based amps.

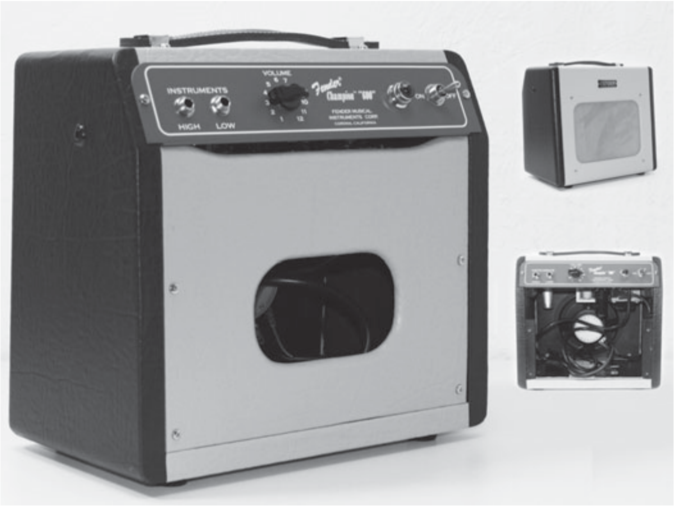

FIGURE 14-1 The Fender Champion 600

Highlights

In this teardown, we'll encounter a high-voltage linear power supply, vacuum tubes, and both power and output transformers. We'll also explore the operating characteristics of Class A amplifiers. You should pay particular attention to the following:

- How shielding is used to minimize noise and EMI
- Provisions for handling heat
- The relatively sparse, open circuit board design
- The use of high-power and high-voltage components
- How the design differs from conventional solid-state devices, in terms of component density, circuit complexity, circuit density, and component count

Specifications

In contrast with the hi-fi industry, Fender and other guitar amplifier suppliers don't provide the typical specifications of their amps, such as frequency response, total harmonic distortion (THD), or noise level. The specifications offered by Fender for the Champ include the following:

- **Dimensions:** 11 × 12 × 7 1/2 inches (HWD)
- **Weight:** 15 pounds
- **Output power:** 5W
- **Speaker:** 4Ω, 6-inch ceramic magnet
- **Preamp tube:** 12AX7A dual triode
- **Power tube:** 6V6GT pentode
- **Inputs:** two

Power input requirements are not specified, but based on my measurements, the amp draws 380mA at 120VAC, or about 40W, regardless of input.

Operation

Operating the Champ involves little more than allowing the tubes to stabilize, plugging in your electric guitar, and turning the volume control clockwise until you like what you hear. Using Figure 14-1 as a guide, on the faceplate, from right to left, is an on–off switch, an LED power indicator, a potentiometer for setting volume and tone, a low-impedance microphone jack, and a high-impedance electric guitar jack.

I usually warm up my amp for about 15 minutes before using it to give the tubes time to reach thermal equilibrium. To give you an idea of the heat involved, consider that after playing for an hour, the temperature of the steel faceplate on my Champ is about 200°F.

Teardown

You should allocate about 2 hours for this teardown, which is illustrated in Figure 14-2.

Caution If you follow along with your own amp, you could be exposed to potentially lethal voltages. Now is a good time to review the safety tips in the Introduction of this book.

Tools and Instruments

For this teardown, you'll need a Phillips-head screwdriver, preferably with insulated handle; a clip lead with medium to large jaws; and a multimeter. Place a towel, carpet remnant, or foam sheet on your workbench to protect the amp's easily torn blonde vinyl.

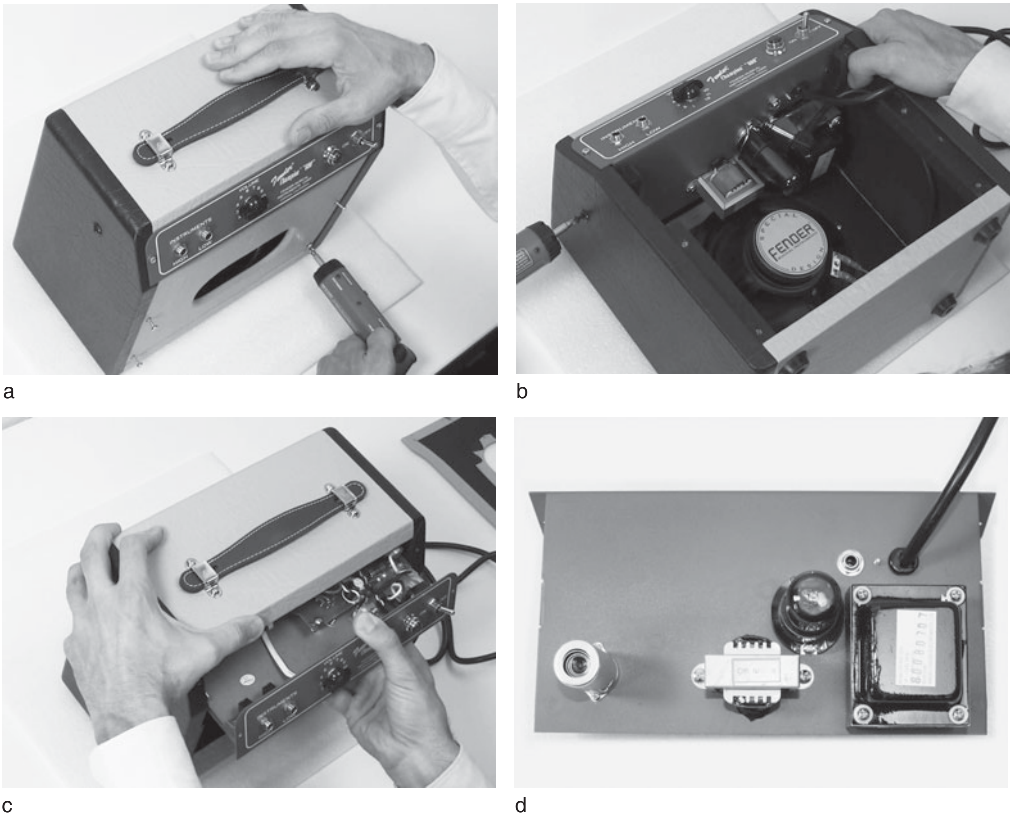

FIGURE 14-2 Teardown sequence

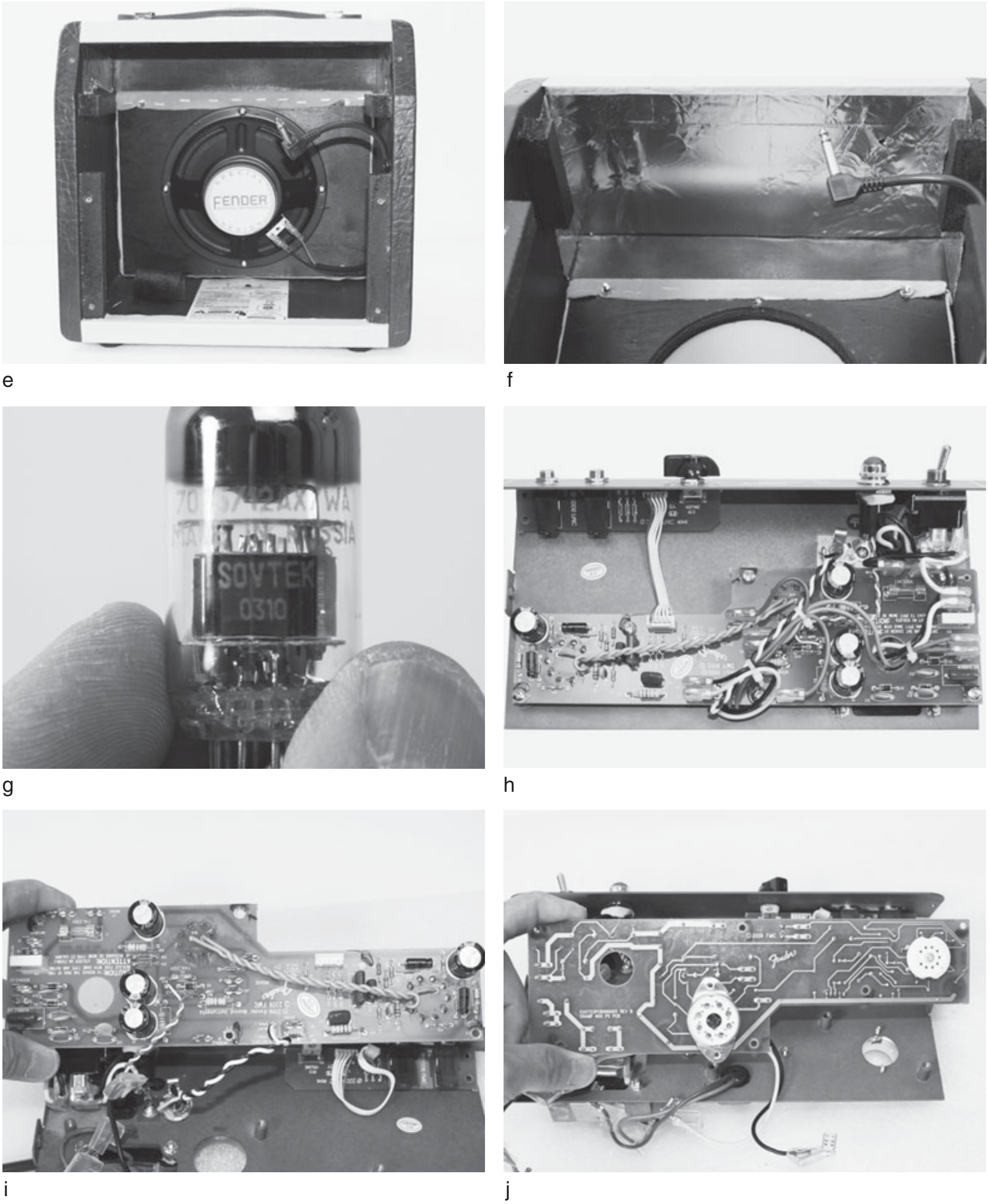

FENDER
SOVTEK
0310

e f g h i j

FIGURE 14-2 (*continued*) Teardown sequence

Step by Step

Before embarking on this teardown, take steps to insure your longevity. If you're not experienced working with high-voltage circuits, unplug the amp and set it aside for a day or two to give the high-voltage filter capacitors time to self-discharge. Because the high-voltage filter capacitors lack bleeder resistors, the high voltage will bleed off over hours instead of minutes.

Step 1

Remove the back plate. As depicted in Figure 14-2a, remove the four Phillips-head wood screws that secure the back plate, which is made of thin pressboard, with the vinyl covering stapled from the inner side.

As shown in Figure 14-1, the rolled steel chassis is housed in the upper 2 inches of the cabinet, with the lower part of the cabinet dedicated to housing the speaker. Steel is used instead of aluminum with many tube-type amps because of the mechanical stability it affords.

Step 2

Remove the chassis. Place the amp speaker-side down and remove the two Phillips-head machine screws on either side of the unit, as shown in Figure 14-2b. Next, remove the two screws on either edge of the faceplate. Unplug the speaker jack from the chassis.

With the amp upright, as shown in Figure 14-2c, carefully slide the chassis out of the cabinet. Avoid jarring the fragile glass envelope of the vacuum tube on the underside of the chassis, and don't touch the circuit board.

Step 3

Examine the chassis layout. Carefully flip it over so that the transformers and tubes are facing up, as shown in Figure 14-2d. Notice the large power transformer on the right and the small audio output transformer near the midline of the chassis. The large, exposed 6V6GT power tube is nestled between the two transformers, the audio output jack, and the power cord entrance. The small 12AX7A dual triode, shielded by a steel cover, is to the left.

Step 4

Examine the cabinet. The wood and vinyl cabinet houses a 6-inch speaker (see Figure 14-2e) and a fragile EMF shield composed of thin aluminum foil. The shield extends about 1 inch below the chassis and makes electrical contact with the chassis at the two tabs on either end. The delicate foil, shown in Figure 14-2f, is inevitably damaged as part of the chassis extraction process.

Step 5

Remove the 6V6GT power tube. Use two fingers of one hand to push down on the retaining spring that secures the base of the tube. With your other hand, grip the base collar—not the glass envelope and not the metal tube pins—and pull straight up. Notice the keyed pin at the center of the base, which will help you align the tube's metal pins with the socket when you replace the tube. Remove and set aside the two screws and lock washers that secure the retaining ring.

Step 6

Remove the 12AX7A tube and shield. Push down on the tube shield and twist a few degrees counterclockwise to disengage the two prongs on the shield base. Allow the spring to push up the cover. Grasp the glass envelope of the tube and pull straight back, perpendicular to the chassis, to extract the tube, shown in Figure 14-2g. Because there is no slotted pin in the center of the tube, you'll have to visually align the pins when you reinsert the tube.

Step 7

Examine the circuit boards. Grasp the large power transformer and flip the chassis to expose the two circuit boards, shown in Figure 14-2h. The main board containing the charged high-voltage electrolytic capacitors is secured by six Phillips-head machine screws, and the smaller board is attached to the faceplate by the volume control and input jack hardware. A ribbon cable connects the two boards.

Step 8

Discharge the high-voltage capacitors. With one hand in your pocket, clip one end of a clip lead to one of the side tabs on the chassis. Clip the other end of the clip lead to the metal shaft of an insulated screwdriver. Now carefully touch the tip of the screwdriver to the anode leads of each silicon power diode. Don't be surprised if you see an arc—it's a sign that you've succeeded in at least partially discharging the capacitors. Verify your work with your multimeter, and repeat the process if the circuit isn't fully discharged.

Step 9

Extract the main circuit board, shown in Figure 14-2i. Unplug the ribbon cable connection to the two boards, unplug the power and transformer leads, and remove the six retaining screws. Although the connection points on the board are labeled, make a note of the transformer and power lead connections.

If you didn't remove the retaining ring of the larger vacuum tube, you won't be able to extract the circuit board because the white ceramic tube sockets are soldered to the main board, as shown in Figure 14-2j. Ceramic provides superior insulation and voltage-handling capabilities compared with plastic tube sockets. You'll have a much easier time of examining the components if you unsolder the audio out and LED indicator twisted pairs that tether the board to the chassis.

Layout

The layout of this amp is utter simplicity—one reason why it's the target of so much mod activity. Position the chassis with the on–off switch in the lower-right corner, as shown in Figure 14-3.

Most of the power supply components are on the right side of the main board, immediately behind the DPDT (double-pole, double-throw) on–off switch and LED power indicator. The four silicon power diodes, each with ceramic disc bypass capacitors, and three 22µf at 450VDC electrolytic capacitors are positioned opposite the power transformer. One of the 22µf capacitors serves as a cathode bypass capacitor for the 6V6GT, and the other two are part of the power supply filter. Two fuses, a 1A slow-blow for the 120VAC mains and a 4A slow-blow for the tube filament supply, are nestled in this area as well.

With the exception of a fourth 22µf at 450VDC electrolytic capacitor immediately adjacent to the 12AX7A tube socket, the right side of the board is devoted to signal amplification. Note that the voltage and power ratings of many components are greater than those in a typical solid-state amp. Components are bulky and leaded, typified by the 2W resistors and orange Mylar signal capacitors rated at 400VDC.

The only notable component on the smaller circuit board is the volume control potentiometer. Instead of a common 2W metal potentiometer, the volume control

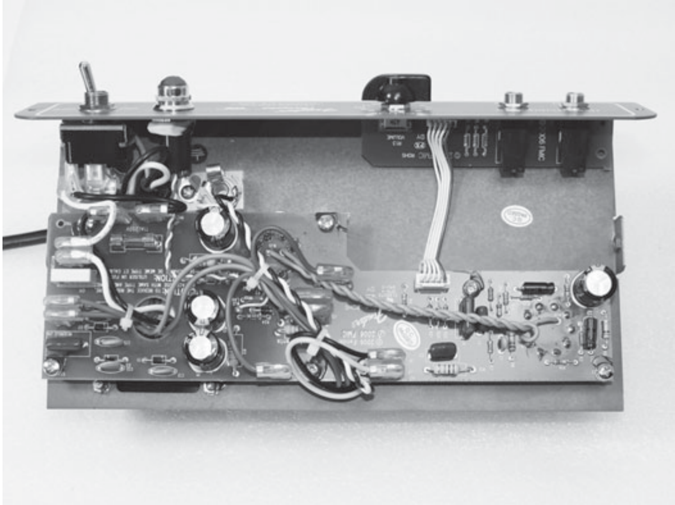

FIGURE 14-3 Layout

is a compact plastic unit. Although it looks out of place with otherwise vintage components, its silky operation imparts a quality feel to the amp.

Components

The major components in this teardown are the vacuum tubes, transformers, and the elements of the high-voltage power supply.

Vacuum Tubes

The Champ's two vacuum tubes are its most distinguishing features. The small 12AX7A dual-triode amplifier has a characteristic stubby evacuated glass envelope and exposed pins. Each triode, which consists of a heater, cathode, control grid, and plate (the heater and cathode count as one element), can produce a voltage gain of up to 100. The 12AX7A excels in low-level audio applications that require high gain.

The maximum plate dissipation of the 12AX7A, the maximum power the plate element can safely dissipate as heat, is a little more than 1W per triode section, or about 2W for the tube as a whole. Exceeding the maximum plate dissipation because of excessive plate current will literally melt the plate. Furthermore, exceeding the maximum plate voltage of 330VDC will result in internal arcing. In either case, the tube and associated circuitry would be destroyed.

The heaters of the 12AX7A are similar to those of other dual tubes in that they can be connected in serial or in parallel, depending on the available voltage. In this application, the two heaters are connected in parallel and supplied by 6.3VAC. Increasing the heater voltage beyond this will result in overheating and premature failure.

The concentric structure of a triode parallels that of a capacitor. As a consequence, significant capacitance exists between the control grid and plate and can result in unwanted oscillation. This unwanted capacitance can be reduced and the stability of the tube improved by inserting additional grids between the control grid and the plate, as in the tetrode (one additional grid) and pentode (two additional grids) tube designs.

The 6V6GT is a single-pentode power tube consisting of a heater, cathode, control grid, screen grid, suppressor grid, and plate. (Again, for categorization purposes, the heater and cathode count as one element.) Although popular as an output tube for audio amplifiers, the 6V6GT is relatively lightweight as power tubes go. Specifications include an amplification factor of about 10, a plate dissipation of 14W, and a maximum plate voltage of 350VDC. The 6V6GT heater requires 6.3VAC.

The majority of tubes sold in the United States are imported from Russia and China. For example, my Champ's 12AX7A was made by a Russian manufacturer and the 6V6GT is from China.

Transformers

By weight, the iron core power and audio output transformers are the most significant components of the Champ. The power transformer has a 120VAC primary and two secondary windings. The high-voltage winding supplies about 280VAC to the capacitor input power supply, and another secondary winding supplies 6.3VAC to the tube filaments. There are two filaments in the 12AX7A and one in the 6V6GT.

The audio output transformer, unshielded and considerably lighter and smaller than the power transformer, enables the 6V6GT to drive the 4Ω speaker. The transformer isolates the high voltage in the output tube circuit from the output speaker and provides impedance matching. In addition, the audio output transformer enables the high-impedance tube circuit to drive the low-impedance speaker.

Printed Circuit Boards

The main circuit board is notable because there are no foil traces on the top side of the board. In addition, as shown in Figure 14-2j, the foil tracks on the underside are wide and generously spaced to accommodate the relatively high currents and voltages in the amp.

The original Champ, like some modern, boutique guitar amplifiers, sports point-to-point wiring. Components are soldered to posts by their leads, and insulated wire is connected to posts to form circuits. Point-to-point wiring was largely abandoned in favor of the circuit board because it was labor-intensive and didn't lend itself to automated fabrication. Although I've used guitar amps from Carr (www.carramps.com) and Gretsch (http://gretsch.com) with point-to-point wiring, I attribute their wonderful tone to quality components, intelligent circuit design, and sturdy enclosure, rather than the antiquated construction techniques.

Speaker

Externally, the 6-inch Fender speaker is unremarkable. The nominal impedance, essentially the average resistance at range of audio frequencies, is also typical at 4Ω. The DC resistance of my speaker is 3.6Ω, which is normal. However, this speaker, like most guitar amp speakers, differs from most hi-fi speakers in frequency response range and intended operating power. Typical guitar speaker frequency response is 75Hz to 5kHz, which is broader than a standard woofer or midrange speaker.

Another distinguishing features of guitar speakers is extended cone and coil travel. In practical terms, this affects optimum operating power. While a typical stereo speaker rated at, say, 100W, can be driven from 100mW to 100W, the Champ's speaker should be driven at 4–5W so that it can contribute to the highly valued distortion.

Fuses

The Champ's power supply has two fuses: a 1A slow-blow on the primary of the power transformer and a 4A slow-blow in the power transformer 6.3VAC secondary. Because the fuses are slow-blow, occasional transients over the stated amperage won't melt either fuse, but overloads of a few seconds or more will.

Unfortunately, if the Champ suddenly dies, the chassis has to be extracted from the speaker enclosure to examine the fuses. Most modern equipment designs incorporate a fuse holder that's accessible from the outside of the device. Curiously, the original Champ sported an external fuse holder where the on–off switch is today.

Resistors

The plastic potentiometer used to control amp volume is an audio taper potentiometer. As discussed in previous teardowns, the resistance between the middle wiper terminal and a terminal is a logarithmic function of the mechanical position of the wiper.

The fixed resistors are 0.25 and 1W metal oxide film. In addition, the 1W resistors have a flame-resistant, nonflammable coating. Flame-resistant resistors are commonly used in high-voltage circuits where there is potential for shorted tubes and other mishaps that could ignite ordinary components.

Wiring and Cables

The transformers are connected to the main board with PVC-coated stranded wire and clip-on connectors. In addition, the main board is connected to the audio output jack via 18-gauge PVC-insulated stranded wire.

The wires carrying audio are tightly twisted, as are wires from the main board to the LED power indicator. Twisting provides some immunity to EMI and unintended coupling between circuit elements.

LED Power Indicator

The LED power indicator, which is incorporated into a vintage-looking incandescent lens hood, is powered by the 6.3VAC secondary of the power transformer. Moreover, the voltage for the LED indicator is taken after the 4A heater fuse. As a result, if the heater fuse blows, the LED power indicator will not be illuminated, even if the high-voltage circuitry is energized. Bottom line: don't trust the LED power indicator.

Capacitors

The most prominent capacitors in this amp are the four 22μf at 450VDC electrolytic capacitors. Three are used to filter the high-voltage supply. The fourth and the two small axial-lead electrolytic capacitors are used to bias the tubes. The dielectric used in electrolytic capacitors dries out with time, and the elevated temperature of a tube circuit accelerates the process.

Silver mica capacitors are used for high-voltage blocking and signal coupling between amplifier sections, and low-cost ceramic disc capacitors are used for bypassing the power diodes. Both silver mica and ceramic disc capacitors are more stable than electrolytic capacitors, and their operating temperature range is compatible with the elevated temperature of tube circuitry.

Silicon Power Diodes

The amp uses five 1N4006 silicon power diodes, conservatively rated at 800V PIV and 1A average output current. Four are used as a full-wave bridge rectifier in the high-voltage power supply, and one rectifies the 6.3VAC heater voltage to drive the LED power indicator.

The four diodes in the high-voltage bridge are each bypassed with 8200pF ceramic disc capacitors. While the 1N4006 has good inrush current properties, it is susceptible to reverse voltage spikes. The bypass capacitors help protect the diodes against reverse spikes from the power transformer.

EMI Shielding

The aluminum foil EMI shield is at best unimpressive. Electrically, the thin aluminum foil shield should reduce EMI. However, as noted earlier, the foil is fragile and easily damaged.

How It Works

Figure 14-4 shows a simplified schematic of the Champ, including labels for the basic tube elements. A detailed schematic is available on the Fender web site (www.Fender.com).

Starting with the power supply, at the lower-right of the schematic, is a simple, unregulated power supply built around a transformer (T1), a diode bridge, and a capacitor input smoothing filter. The power transformer has two secondary windings, one 280VAC for the high-voltage supply, and one 6.3VAC for the three tube heaters—two in the 12AX7A and one in the 6V6GT—and the power indicator LED. For clarity, the LED circuit, consisting of a diode, series resistor, and LED, is not shown.

The 6.3VAC winding is notable in that it's connected to an artificial center tap composed of a 100Ω resistor on either end of the winding to ground. The benefit of this design over a floating 6.3VAC winding is reduction in hum. A better design is a transformer with a real center tap, but center-tapped transformers cost a few cents more than a pair of resistors.

The 280VAC secondary is connected to a bridge rectifier composed of four silicon power diodes. The resulting full-wave rectified DC is smoothed by a bank of three

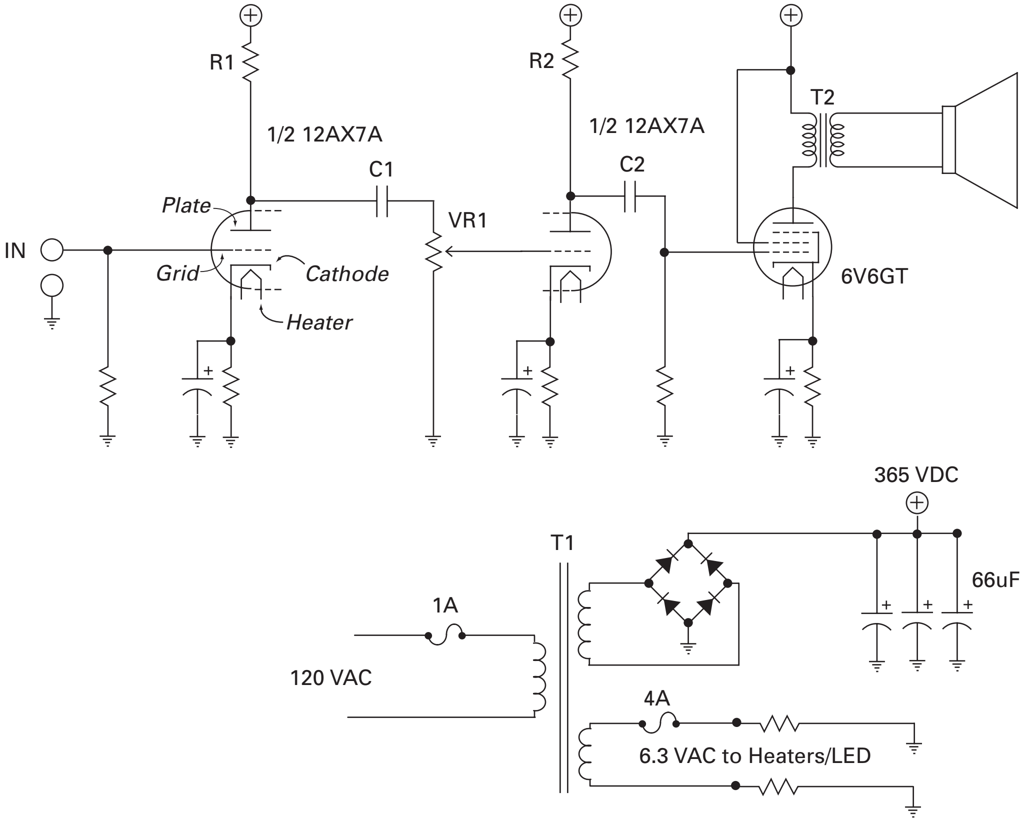

FIGURE 14-4 Simplified schematic

22µf at 450VDC electrolytic capacitors connected in parallel. The capacitor bank, which provides a total of 66µf at 450VDC, boosts the bridge output voltage to about 365VDC, the plate voltage of the 6V6GT output tube. Series resistors (R1 and R1) are used to drop the plate voltage down to about 340VDC for the 12AX7A preamp tube.

The power transformer, which is shielded to reduce hum, is both a step-up and a step-down transformer, supplying both higher and lower voltages than the line voltage applied to the primary winding. Recall that the primary and secondary voltages are related to the turns ratio. Given an input of 120VAC and an output of 280VAC, the primary-to-secondary turns ratio of the power transformer should be 120:280, or 1:2.3, for the high-voltage winding. Similarly, the turns ratio for the primary to 6.3VAC secondary is 120:6.3, or 19:1.

The audio output transformer is a step-down transformer, converting the high voltage, low current of the tube circuit to low voltage, high current for the speaker. While primary and secondary voltages are related by simple turns ratios, primary and secondary impedances are related by turns ratio squared. Given that the typical plate load resistance for a 6V6GT is about 5K and the nominal impedance of Champ's speaker is 4Ω, we can estimate the turns ratio of the audio output transformer. The output-to-speaker impedance ratio is 5000:4, or 1250:1. Taking the square root gives us the turns ratio, or 35:1.

The amplifier circuit proper is built around a two-stage preamplifier using the 12AX7A dual triode. This preamp feeds the 6V6GT power tube, which in turn drives the 4Ω speaker through the audio output transformer. A resistor and capacitor network in the front end of the amp, not shown in the schematic for clarity, provide high- and low-impedance inputs for a guitar and microphone, respectively.

If you're new to tubes, the solid-state equivalent of a triode is an n-channel J-FET (junction field effect transistor). Structurally, the heater element of a 12AX7A triode is surrounded by a cathode, which is in turn surrounded by a porous control grid, which is surrounded by the sheet metal plate. In operation, the heater raises the temperature of the cathode so that it emits electrons. A high positive voltage applied to the plate attracts these electrons. Unimpeded, a plate current flows from cathode to plate. However, a small negative voltage applied to the grid, relative to the cathode, can drastically reduce the flow plate current. In this way, a small varying voltage applied to the grid is amplified in the cathode-plate circuit.

If the control grid is made significantly negative relative to the cathode, electrons streaming from the cathode are repelled and plate current ceases. Conversely, if the grid is made positive relative to the cathode, plate current increases—to a point. A significantly positive grid attracts electrons to itself, resulting in a significant grid current and catastrophic failure of the tube. Establishing the grid voltage relative to the cathode—the tube bias—establishes the operating parameters of the circuit.

Returning to the schematic, the audio signal from the input circuit is applied to the control grid of the first triode. The voltage developed across the plate resistor is coupled through a capacitor (C1) and the volume potentiometer (VR1) to the grid of the second triode. The capacitor blocks the high voltage of the plate circuit while passing the audio signal.

Similarly, the output of the second triode is coupled via C2 to the control grid of the 6V6GT power tube. Variations in the plate current of the 6V6GT are coupled through the audio transformer to the 6-inch speaker. Notice that the first grid after the control grid, the screen grid, is connected to high positive voltage. The grid nearest the plate, the suppressor grid, is connected to the cathode.

A cathode resistor—a resistor in series with the cathode—creates the bias for each tube. The voltage drop across a cathode resistor from plate and grid current makes the cathode positive relative to the control grid. Recall that a negative control grid repels electrons and diminishes plate current. In this way, the amplitude of the negative grid bias affects the amplification factor and the amount of distortion imparted by each tube.

As shown in the schematic, the cathode resistors are bypassed with decoupling capacitors. This decoupling allows for higher amplification because audio-frequency plate current that bypasses the cathode resistor doesn't contribute to a greater voltage drop across the cathode resistor—and a more negative control grid bias voltage.

Without the bypass capacitors, the bias voltage will fluctuate with the input signal. For example, if the input is going positive, the control grid becomes less negative, and plate and therefore cathode current increases. The voltage drop across the cathode resistor makes the control grid more negative, reducing the tube conduction and gain.

The values of the cathode resistors and bypass capacitors are selected so that the plate current is never cut off, regardless of input. This configuration is commonly referred to as a *Class A amplifier*. A characteristic of Class A amplifiers is inefficiency, because the amplifier tube is always conducting, even when there is no input signal. Class A amplifiers are also capable of faithfully reproducing audio with only one output device.

Note the although capacitor coupling is used between tubes, the output of the 6V6GT is coupled to the speaker by the audio output transformer. Transformer coupling from a high-impedance tube to a low-impedance speaker is more efficient than capacitive coupling. Also, recall from our teardown of the solid-state stereo amplifier (Chapter 9) that the low-impedance output of the integrated amplifier chips was connected directly to the speakers.

Mods

Following examples posted on the Web and through my own experimentation, I've made several mods of the Champ to improve the tone, enhance safety, improve usability, and minimize susceptibility to noise. While amplifier specifications such as power and frequency response may be interesting to electronics enthusiasts, guitar players are primarily concerned with tone—a subjective assessment that defies technical definition. Keep this in mind as you consider potential mods for the Champ or another tube guitar amp.

Convenient Fuse Placement

The internal fuses are a pain to check. You can make the mains fuse more convenient to check and change by installing a mains fuse holder and 1A at 120VAC slow-blow in the faceplate, using the space for the low input impedance jack. You'll have to unsolder the unused jack from the small circuit board to make room for the fuse holder. Wire the fuse so that it is on the hot (black) mains lead, between the wall cable and the on–off switch. If the switch shorts, the fuse will blow before the chassis melts. Leave the 4A heater fuse in place.

Improved EMI Shielding

At a minimum, consider reinforcing the area where the chassis tabs make contact with the aluminum foil shield. Use wood screws to attach a 1-inch-wide strip of aluminum—*not foil*—on either side of the inner cabinet. Alternatively, replace the entire shield with copper or aluminum sheeting. Another option is to coat the inner surface of the amp with conductive paint sold through luthier supply stores such as Stewart-MacDonald (www.stewmac.com).

Larger Speaker and Cabinet

The stock speaker, while good at relatively low volumes, bottoms out at maximum volume. It's a simple matter to replace the stock Fender speaker with a 10 or 15W speaker from Celestion, Jensen, or Eminence. If you want to move to an 8-inch or 10-inch speaker, a number of vendors offer cabinet upgrades that accept the Champ's metal chassis.

Improved Thermal Management

The sealed chassis mounted at the top of the cabinet creates an excellent environment for causing early failure of electrolytic capacitors. The simplest thermal management mod is to remove the back plate, at the possible expense of tone. Vacuum tubes are designed to run hot, especially if you want to hear their characteristic distortion.

Another option that doesn't affect tube temperature or performance is to perforate the cabinet top. This mod will definitely decrease resale value, however.

Cleaner, Safer Power

Clean up the power by adding a commercial EMI filter module where the power enters the chassis. An EMI filter is mandatory if you plan to operate the amp around fluorescent lights, cell phones, and computer equipment.

Consider replacing the electrolytic capacitors with a higher ripple current variety. However, avoid the temptation to increase the capacitance as well. If you do, you could increase the voltage to beyond the maximum for the tubes and components. A higher capacitance also places a greater surge load on the diodes, since an uncharged capacitor momentarily looks like a short-circuit to the diode bridge.

Another option is to transform the unregulated, capacitor input supply to a regulated supply by inserting a choke between the bridge rectifier output and capacitor bank. The downside is that a choke-input filter will drop the DC output voltage to about 90 percent of the output from the bridge. A new power transformer with a higher output voltage is a workaround, but an expensive one. Curiously, the original Champ featured a choke-input power supply. I suspect the expensive choke was dropped from the current design as a cost-saving measure.

Finally, add a bleeder resistor across the capacitor bank to bleed off the potentially lethal high voltages after you turn off the amp. A 250K, 1W flame-proof resistor across the high-voltage capacitors should be sufficient.

Gain Control

As discussed earlier, the 12AX7A and 6V6GT operate with fixed bias, by virtue of fixed cathode resistors and bypass capacitors. If you want to adjust the gain of the amp on the fly, replace the fixed cathode resistor of the 6V6GT with a potentiometer. If you make this modification, be certain to stay within the plate dissipation rating of the 6V6GT; otherwise you'll shorten the life of the tube.

The challenge with adding a gain control is that you need a convenient place to mount it. Unless you want to drill a hole in the cabinet, cabinet back, or faceplate, you'll have to remove either the LED power indicator or one of the two input jacks temporarily.

Change the Cathode Capacitor

Changing the value of the 6V6GT cathode bypass capacitor can significantly alter the tone of the amp. Several mod kits increase the value of the bypass capacitor for thicker tone. However, you can experiment with higher and lower values until you find a tone to your liking.

More Efficient Audio Output Transformer

The stock audio output transformer is an inexpensive model from China. A better transformer, either of conventional iron-laminate design or toroidal, could provide higher efficiency. Bigger isn't necessarily better, but all else being equal, a larger audio output transformer will handle more current before saturating.

Permanent Speaker Connection

Minimize the resistance and power loss of the output circuit by wiring the speaker directly to the output of the audio transformer. Consider soldering the leads to the speaker terminals to provide a secure, low-resistance connection.

Higher Quality Tubes

The stock tubes used in many amps don't provide the best possible tone. Although tone preferences vary from person to person, some tubes are know for their superior tone. As noted earlier, the stock Champ comes with relatively inexpensive tubes of Russian and Chinese manufacture. Several companies import tubes in bulk and then test for gain, optimum bias settings, and other parameters, and they resell the tubes under their value-add label. Groove Tubes (www.groovetubes.com), which is owned by Fender, is one such domestic tube reseller. GT tubes are more expensive than their untested counterparts, but you have a better idea of what you're getting.

Try an Upgrade Kit

The easiest way to mod the Champ is to purchase a commercial upgrade kit that addresses a variety of issues with proven mods. I've had success with a combination of a commercial upgrade kit from Mercury Magnetics (www.mercurymagnetics.com) and a new set of tubes from Groove Tubes.

The Mercury Magnetics upgrade includes a power transformer, audio output transformer, choke for the power supply, a handful of components for modifying tube bias levels, and a 15W, 6-inch speaker, shown on the right of Figure 14-5a. The original components appear on the left of the figure. The upgraded amplifier chassis—about 7 pounds heavier than the original—is shown in Figure 14-5b.

The mod kit from Mercury Magnetics is one of dozens available through and reviewed on the Web, and you can find other sources of magnetics for tube amps. For example, Hammond Manufacturing (www.hammondmfg.com) offers an extensive line of transformers and chokes. The benefits of working with a commercial upgrade kit include consistent, known results. You have to determine whether those results are to your liking and worth the expense.

If you make any modifications to the circuitry, use a fused Variac or other brand of variable transformer to increase the AC line voltage gradually while you monitor the input current. I use a Variac with a built-in voltmeter and a digital Kill A Watt EZ (www.p3international.com) to monitor current, but a Variac with a built-in current monitor is best if you can afford one.

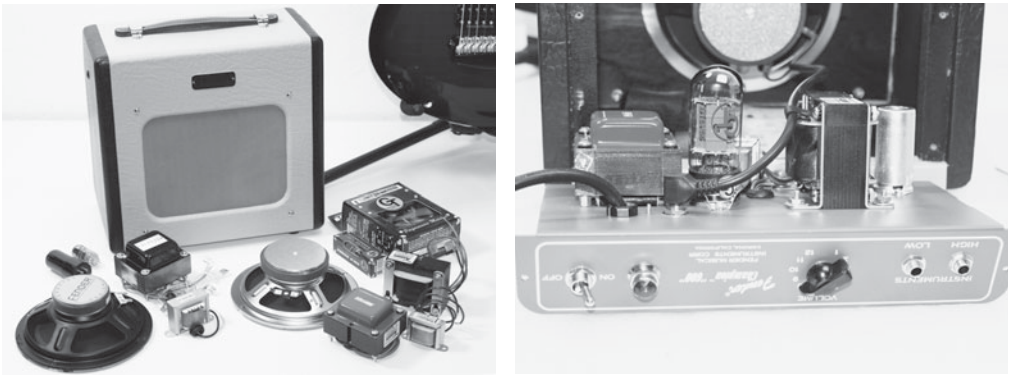

a b

FIGURE 14-5 Mercury Magnetics upgrade

This page intentionally left blank

PART IV

Appendixes

This page intentionally left blank

Appendix A

Component Markings

To fully understand the function of an electronic device, you'll need to identify the value of the components used in the underlying circuitry. Fortunately, most component manufacturers use standard component markings, which are defined in the following tables.

Numeric Notation

Standard numeric notation is listed in Table A-1. Note that alphabetic notation may be used as a decimal point indicator with SMT components.

TABLE A-1 Numeric Notation

Notation	Multiplier	Example	Value
p – pico	$10^{-12} = 0.000000000001$	2.3pF	$2.3 \times 10^{-12}\text{F} = 0.0000000000023\text{F}$
n – nano	$10^{-9} = 0.000000001$	14nH	$14 \times 10^{-9}\text{H} = 0.0000000014\text{H}$
μ – micro	$10^{-6} = 0.000001$	0.01μf	$0.01 \times 10^{-6}\text{F} = 0.0000001\text{F}$
m – milli	$10^{-3} = 0.001$	3mW	$3 \times 10^{-3}\text{W} = 0.003\text{W}$
K – kilo	$10^{3} = 1000$	23KW	$23 \times 10^{3}\text{W} = 23{,}000\text{W}$
		1K2	$1.2 \times 10^{3}\Omega = 1200\Omega$
M – mega	$10^{6} = 1{,}000{,}000$	5MΩ	$5 \times 10^{6}\Omega = 5{,}000{,}000\Omega$
		M47	$0.47 \times 10^{6}\Omega = 470{,}000\Omega$

Leaded Components

With the exception of some capacitors, MOVs, and unmarked glass diodes, most leaded components conform to standard marking systems. The most common standards are reproduced here.

Capacitors

Capacitor marking systems vary by capacitor type. Ceramic disc and most film capacitors are marked in pF using the system in Table A-2. In addition to tolerance markings, film capacitors commonly have a voltage marking. Once you have the system down for a specific vendor or type of capacitor, the most difficult step is converting the value into a range you're familiar with and that's listed in catalogs.

TABLE A-2 Ceramic Disc and Film Capacitor Marking Examples

Marking	Value
101	$10 \times 10^1 = 100pF = 0.1nF = 0.0001\mu f$
102	$10 \times 10^2 = 1000pF = 1nF = 0.001\mu f$
103	$10 \times 10^3 = 1000pF = 10nF = 0.01\mu f$
J	±5%
K	±10%

Electrolytic capacitors, because of their large size relative to other capacitors, have markings for polarity, capacitance, and working voltage. There's generally no guesswork, as the values and units are listed plainly on the component. Where the difficulty arises is deciphering the codes that indicate, for example, whether the capacitor is designed for high temperature, high ripple current, low volume, high stability, and so on. Unfortunately, these codes are vendor-specific. For example, Panasonic Type A, Series EE, is designed for high ripple applications, whereas Type A, Series FM, is a low impedance design.

Resistors

Table A-3 lists the standard resistor and inductor color-code system. Standard value resistors use three bands, the first two for value and the third for the multiplier. A fourth band displays tolerance. Precision resistors may feature a third value band, for a total of five bands.

TABLE A-3 Resistor Color Codes

Color	Value	Multiplier	Tolerance
Black	0	$10^0 = 1$	
Brown	1	$10^1 = 10$	±1%
Red	2	$10^2 = 100$	±2%
Orange	3	$10^3 = 1K$	
Yellow	4	$10^4 = 10K$	
Green	5		
Blue	6		
Violet	7		
Grey	8		
White	9		
<none>			±20%
Silver		$10^{-1} = 0.1$	±10%
Gold		$10^{-2} = 0.01$	±5%

Table A-4 shows examples of color bands applied to standard four-band and precision five-band resistors.

TABLE A-4 Resistor Color-Code Examples

Band 1	Band 2	Band 3	Band 4	Band 5	Value (Ohms)
Black	Red	Red	Silver		1200 ± 10%
Brown	Blue	Orange	Gold		16K ± 5%
Green	Green	Yellow			550K ± 20%
Orange	Red	Red	Gold	Red	32.2 ± 2%
Yellow	Grey	Blue	Brown	Brown	4860 ± 1%

Inductors

Leaded inductors use the same basic color-code scheme for value and multiplier as defined in Table A-3. Table A-5 lists several examples of the color-band marking scheme applied to leaded inductors. Note that the value is in μH.

TABLE A-5 Inductor Color-Code Examples

Band 1	Band 2	Band 3	Band 4	Value (μH)
Black	Red	Red	Silver	1200 ± 10%
Red	Violet	Brown	Gold	270 ± 5%
Green	Green	Black		55 ± 20%
Orange	Red	Red	Gold	3200 ± 5%
Yellow	Grey	Brown	Silver	480 ± 10%

SMT Components

SMT components use characters instead of the color bands or dots to indicate value and operating parameters. However, many of the latest generation of SMT components are too small to mark. You'll have to use a capacitance meter to determine the value of those tiny beige ceramic capacitors, for example.

SMT Capacitors

SMT tantalum capacitors use a standard three-digit and one-letter notation, in which the first two digits indicate the value and the third digit indicates the multiplier, and the letter indicates voltage rating. The anode is marked with a polarity band. Depending on the space available on the face of the capacitor, it may also have a batch ID, manufacturer's logo, and date of manufacture code. See Table A-6 for voltage markings.

TABLE A-6 SMT Tantalum Capacitor Voltage Markings

Code	Voltage	Code	Voltage
X	1.8	C	16
E	2.5	D	20
G	4	E	25
J	6.3	V	35
A	10	T	60

The effect of the multiplier on the value represented by the first two digits of a tantalum capacitor can be determined from this formula:

$$\text{Multiplier} = 10^{n-6}$$

where n is the value of the third digit. See Table A-7 for SMT tantalum capacitor marking examples.

TABLE A-7 SMT Tantalum Capacitor Marking Examples

Marking	Value
475J	$47 \times 105^{-6} = 47 \times 10^{-1} = 4.7$µf at 6.3VDC
476A	$47 \times 106^{-6} = 47 \times 10^{0} = 47$µf at 10VDC
156V	$15 \times 106^{-6} = 15 \times 10^{0} = 15$µf at 35VDC
106G	$10 \times 106^{-6} = 10 \times 10^{0} = 10$µf at 4VDC
227T	$22 \times 107^{-6} = 22 \times 10^{1} = 220$µf at 50VDC

SMT electrolytic capacitors are marked with a straightforward notation that indicates capacitance in µf and working voltage, as in Table A-8.

TABLE A-8 SMT Electrolytic Capacitor Marking Examples

Marking	Value
1/16	1µf at 16VDC
22/25	22µf at 25VDC
100/25	100µf at 25VDC

SMT Resistors

Standard-tolerance SMT resistors are marked with a three-digit code. The first two digits are the value and the third digit is the multiplier. The multiplier is simply 10^n, where n is the value of the third digit. Resistance less than 10Ω is marked with *R* to indicate the position of the decimal point. See Table A-9 for examples.

TABLE A-9 SMT Standard-tolerance Resistor Marking Examples

Marking	Value (Ohms)
475	$47 \times 10^5 = 4.7M$
4R6	4.6
150	$15 \times 10^0 = 15$
151	$15 \times 10^1 = 150$

Precision resistors are marked with a four-digit code, in which the first three digits represent the most significant digits of the component value and the fourth is the multiplier. See Table A-10 for examples. Note that simply because a value is possible doesn't mean that it's available. Manufacturers produce standard resistor values that are applicable to practical engineering design needs. For example, it would be expensive to manufacture 6.2, 6.3, 6.4, and 6.5K ± 5 percent resistors, in part because their values nearly overlap when you consider the tolerance. Standard values in this range and tolerance are 6.2 and 6.8K. Other values are possible, but nonstandard values are more expensive than standard values.

TABLE A-10 SMT Precision-tolerance Resistor Marking Examples

Marking	Value (Ohms)
4995	$499 \times 10^5 = 49.9M$
0R22	0.22
1000	$100 \times 10^0 = 100$
1501	$150 \times 10^1 = 1.5K$

Appendix B

Resources

Practical Theory

Web sites are a great source for component and circuit theory. My favorite first stop on the Web is Wikipedia (www.wikipedia.org). But logging into the Web can be a hassle in the middle of a teardown. If you're just getting into electronics, you owe it to yourself to invest in a modest library of books or ebooks.

Core Resources

If your bookshelf is bare, I suggest starting with *Getting Started in Electronics*, by Forrest M. Mims III (Master Publishing, 2003). Once you've digested the material, invest in a copy of *The Art of Electronics* by Horowitz and Hill (Cambridge University Press, 1989). If your interests are in communications, power supplies, and RF amplifiers, consider adding any year of the *ARRL Handbook for Radio Communications* to your library. All three books are available through Amazon and Barnes & Noble. Go for the second-hand copies of these classics and save yourself some money.

If you're short on cash or simply prefer a local electronic reference, then you have to download *Lessons in Electric Circuits*, by Tony R. Kuphaldt, at www.openbookproject.net/electricCircuits. This excellent, richly illustrated, and regularly updated resource is available for free download as searchable PDF and HTML files.

Topic-specific Sources

If you've exhausted the core resources and want to delve deeper into a topic, try these books and web sites.

Control Theory

- *Feedback Systems: An Introduction for Scientists and Engineers*, by K.J. Astrom and R.M. Murray (Princeton University Press, 2009). www.cds.caltech.edu/~murray/amwiki
- Parallax, Inc., www.Parallax.com Information on microprocessors and control applications using the BASIC Stamp processor.

Distortion and Music

- Electronic Musician tutorials, www.emusician.com/tutorials

Guitar Effects Pedals

- *Guitar Effects Pedals*, by D. Hunter (Backbeat Books, 2004). While mainly dealing with the history and practical application of effects pedals, Hunter's book offers generic schematics for a range of standard effects.

Guitar Magnetic Pickups

- Fender, www.fender.com
- *The Guitar Pickup Handbook*, by D. Hunter (Backbeat Books, 2008). A good book, written more for the musician than electronics enthusiast.
- "The Secrets of Electric Guitar Pickups," by H. Lemme. http://www.buildyourguitar.com/resources/lemme/index.htm A good overview of the complexities of pickups and system interactions.

PIR Sensors

- Parallax, www.Parallax.com Free, downloadable documentation on an inexpensive passive infrared (PIR) sensor based on the BIS0001.

Semiconductors

- National Semiconductor, www.national.com The web site features an interactive program called WEBENCH Power Designer that enables you to design and simulate circuits.

Tube Amplifiers

- Aiken Amplification, www.aikenamps.com
- *Building Valve Amplifiers*, by Morgan Jones (Newnes, 2004).
- *The Guitar Amp Handbook: Understanding Tube Amplifiers and Getting Great Sounds*, by D. Hunter (Backbeat Books, 2005).

Zebra Elastomeric Connectors

- Fujipoly America, www.fujipoly.com

Component Specifications

The best source for information on components is the component's official datasheet, directly from the manufacturer, a datasheet clearinghouse, or through a parts catalog.

General

Here are my favorite online and print catalog references, in order of preference:

- DigiKey, www.digikey.com My first stop for component specifications. Products are linked to an extensive online database of datasheets.
- Mouser Electronics, www.mouser.com My second stop for general components, and my first for hard-to-find and specialized components.
- DatasheetCatalog.com, www.datasheetcatalog.com Free, extensive database of official datasheets of semiconductors by device type and manufacturer.
- Allied Electronics, www.alliedelec.com Excellent source for capacitor and resistor information.
- Jameco Electronics. www.jameco.com Less extensive than DigiKey or Mouser, but easier to navigate, and the online photos make it easy to identify components.
- Parts Express, www.parts-express.com Like Jameco, easy to navigate and loaded with photos of components. Great for specifications on speakers and audio components.

Component-specific Sources

The companies and web sites listed here include sources of components discussed in the teardowns covered in this book. There are hundreds of additional component companies that can be accessed on the Web.

Capacitors

- Carli Electronics Company, www.carli-cap.com.tw
- NTE Electronics, www.nteinc.com
- RFE International, rfeinc.com
- Vishay, www.vishay.com

Diodes

- Diodes Incorporated, www.diodes.com
- National Semiconductor, www.national.com

Electric Guitar Components

- Allparts, www.allparts.com
- Fender, www.fender.com
- Stewart-MacDonald Instrument Repair Supply, www.stewmac.com

Microcontrollers and Supporting Components

- Atmel, www.atmel.com
- Freescale Semiconductor, www.freescale.com

Operational Amplifiers

- National Semiconductor, www.national.com

Resistors

- Bourns, www.bourns.com Fixed, variable, leaded, SMT, and high power resistors.
- Vishay, www.vishay.com

Speakers

- Celestion, www.celestion.com
- Eminence, www.eminence.com

Temperature Sensors

- Diodes Incorporated, www.diodes.com
- National Semiconductor, www.national.com
- Vishay, www.vishay.com

Transistors

- Diodes Incorporated, www.diodes.com
- Vishay, www.vishay.com

Transformers and Inductors

- Hammond Manufacturing, www.hammondmfg.com
- Vishay, www.vishay.com

Vacuum Tubes

- Tube Zone, www.tubezone.net/tubedata.html An extensive source of freely downloadable PDFs of original vacuum tube manufacturers' datasheets.

Index

A

AC voltmeter, 194–196
accelerometers, 93–94
accuracy, 179, 207
activation pulses, 47
Ah calculations, 30
alpha radiation, 7, 8
alphabetic notation, 295
ambient light sensitivity, 33
American Wire Gauge (AWG) system, 21–22
Americium particle emitter, 4, 5–11, 16, 18
ammeter, 196–197
amplifiers
    operational, 216, 265–268, 304
    specifications, 304
    stereo power amplifier, 143–173
    vacuum tube guitar amplifier, 275–291
analog ohmmeter, 45
analog Volt-Ohm-Meter (VOM), 177–201
analog-to-digital converters, 91
aperture angle, 207
audible continuity tester, 199–200
audio jacks, 241–242
audio output transformers, 290
AudioSource AMP 100 amplifier, 143–173
auto-input feature, 149
auto-on feature, 148–149
AV600 surge suppressor, 53, 54–69, 80–81
avalanche diodes, 56
AWG (American Wire Gauge) system, 21–22
axial lead epoxy-dipped capacitors, 16

B

bandwidth, 145
bathroom scale, digital, 35–51
batteries
    analog VOMs, 190, 198
    digital bathroom scale, 50
    lithium, 198
    motion-activated LED light, 20, 27, 29–30, 31
    smoke alarm, 3, 5, 9, 12, 15, 16
    temperature and, 207
battery tester, 194
bipolar transistors, 123, 124
BIS0001S IC, 24–25
BLDC (DC brushless motor), 125–130
books/documents, 301
brushless DC motor/fan, 125–126
buzzers, 44, 190–191, 199. *See also* piezoelectric transducer
bypass switching, 274

C

cables, 284
cadmium sulfide (CDS) sensors, 23, 26–27
calibration, 141–142
calorie counter function, 84
capacitors
    ceramic, 29
    dipped, 16
    effects pedal, 268–269
    electric guitar, 241
    electrolytic, 16, 29, 105, 296, 299
    film, 296
    marking systems, 296
    non-inductive, 65–66
    polypropylene film, 65–66
    SMT, 29, 298–299
    sonic distance measurer, 219
    specifications, 303
    stereo power amplifier, 163
    ultrasonic humidifier, 121
    vacuum tube guitar amplifier, 284–285
cataphoresis, 107
cathode capacitor, 290
cathode resistors, 287–288
CDS (cadmium sulfide) sensors, 23, 26–27, 32, 33
cell phones, 93–94
ceramic capacitors, 29
ceramic disc capacitors, 16, 296
CF (compact fluorescent) lamp, 95–111

circuit boards
  digital hygro thermometer, 136
  double-sided, 77
  effects pedal, 265
  printed, 283
  single-sided, 66
  sonic distance measurer, 211–212, 219, 223
  ultrasonic humidifier, 119–120
  vacuum tube guitar amplifier, 283
circuit breakers, 62, 67, 76
circuit tracing, 23–24
clipping, 270–272
clock frequency, 89
closed loop system, 130–131
CMOS operational amplifiers, 216
CMOS voltage detectors, 89, 91
color rendering index (CRI) score, 98
color temperature, 96–97, 98
compact fluorescent (CF) lamp, 95–111
component identification, 23–24
component markings, 295–300
components
  alphabetic notation, 295
  analog VOMs, 184–191
  compact fluorescent lamp, 102
  digital bathroom scale, 40–46
  digital hygro thermometer, 137–139
  dual-sensor alarm, 9–16
  effects pedals, 263–269
  electric guitar, 239–243
  electronic pedometer, 87–90
  leaded, 296–298
  motion-activated LED light, 24–30
  numeric notation, 295
  SMT, 298–300
  sonic distance measurer, 213–219
  specifications, 303–304
  stereo power amplifier, 158–165
  surge suppressor, 62–66
  ultrasonic humidifier, 121–126
  vacuum tube guitar amplifier, 282–285
component-specific sources, 303–304
connectors
  digital bathroom scale, 45–46, 302
  effects pedal, 269
  electronic pedometer, 86, 90, 302
  identifying, 118
  RCA, 152, 157, 168
  stereo power amplifier, 152, 157, 168
  ultrasonic humidifier, 117–120
  Zebra Elastomeric Connectors, 45–46, 86, 90, 302
control theory, 302
Crane ultrasonic humidifier, 113–131
CRI (color rendering index) score, 98
crystal oscillator, 89, 139
crystals, quartz, 138, 217–218
CUN voltage detector, 89, 91
cyanoacrylate, 51

D

D'Arsonval movement, 186
DC brushless motor (BLDC), 125–130
DC motors, 125–126
DC voltmeter, 192–194
Digi-Keys, 76, 190
digital bathroom scale, 35–51
digital hygro thermometer, 133–142
digital multimeter (DMM), 37, 45, 177, 179–180
diodes
  1N4007, 65
  analog VOMs, 186–188
  avalanche, 56
  CF circuit, 106–107
  effects pedal, 263–265, 273
  germanium, 187–188, 262–265, 270, 273
  Schottky, 186–188, 264
  silicon, 77, 187–188, 263–264, 273, 285
  sonic distance measurer, 217
  specifications, 303
  stereo power amplifier, 162–163
  TVS, 56
  ultrasonic humidifier, 121
dipped capacitors, 16
displays. *See* LCD
distortion, 146–147, 269, 302
DMM (digital multimeter), 37, 45, 177, 179–180
documents/books, 301
dual-sensor smoke alarm, 3–18

E

EEPROM (electrically erasable programmable read-only memory) chip, 40, 41, 42, 46, 47
effects pedal, 257–274, 302
elastomeric buttons, 90
elastomeric connections, 45–46, 86, 90, 302
electric guitar, 231–255, 304
electrically erasable programmable read-only memory (EEPROM) chip, 40, 41, 42, 46, 47
electricity
  joule energy rating, 55
  lightning strikes, 53, 55
  power conditioners, 53, 54, 55, 69–81
  surge protective devices, 53–81
  surge suppressors, 53, 54–69, 80–81
  uninterruptible power supply, 53, 55
  voltage spikes/surges, 54, 55–56, 79
electrolytic capacitors, 16, 29, 31, 296, 299
electromagnetic buzzers, 190–191

electromechanical relay, 33
electronic ballast, 107
electronic devices. *See* components
electronic pedometer, 83–94
EMI shielding, 240, 285, 288–289
energy ratings, 55
epoxy encapsulation, 36, 40, 41, 42
external load cells, 50–51

F

Fairchild Semiconductor, 65
Fender Champion guitar amplifier, 275–291
Fender Telecaster electric guitar, 231–255
ferrite chokes, 76
film capacitors, 296
fluorescent lamp, 95–111
frequency calculator, 248
frequency response, 145
Fresnel lens, 25–26
Fujipoly America, 46
Furman power conditioner, 54, 69–81
Furturlec, 25
fuses
    analog VOMs, 190
    fast-blow vs. slow-blow, 124
    stereo power amplifier, 165
    thermal, 64, 76
    ultrasonic humidifier, 124
    vacuum tube guitar amplifier, 284, 288

G

gain, 269–270
gain control, 290
galvanometer, 181, 186–188, 190, 192–200
Gauge Factor (GF), 49–50, 51
Geiger counter, 7, 8
germanium diodes, 187–188, 262–265, 270, 273
GF (Gauge Factor), 49–50, 51
GN suppression, 56
ground contamination, 80
ground fault, 66
ground lead, 66
guitar magnetic pickups, 243–255, 302
guitars
    amplifiers, 275–291
    effects pedal, 257–274, 302
    electric guitar, 231–255

H

heat sinks
    compact fluorescent lamp, 103–104
    stereo power amplifier, 159–161
    ultrasonic humidifier, 122
high voltage (HT) probes, 100–101
HT (high voltage) probes, 100–101
humidifier, ultrasonic, 113–131
humidity
    ambient, 25, 207
    closed loop system and, 130–131
    electronics and, 207
    measuring, 133–142
    problems caused by, 17, 25
    relative, 113, 130, 134, 138, 140–142
    variable, 114
humidity sensor, 117, 130, 136, 138, 141, 142

I

ICs (integrated circuits)
    BIS0001S, 24–25
    EEPROM, 40, 41, 42, 46, 47
    epoxy encapsulation, 36, 40, 41, 42
    microcontroller support, 89
    SOIC, 24, 25
    TDA7294, 155, 158–159, 169–171
    TL0001S, 24–25, 32, 33
IE core, 124, 125
IEC (International Electrotechnical Commission), 57
impedance, 138
impedance matching, 272
incandescent bulbs, 98
inductors
    color-code system, 296, 298
    core, 105, 108
    leaded, 297
    marking systems, 297–298
    microHenry, 76
    sonic distance measurer, 218–219
    specifications, 304
integrated circuits. *See* ICs
International Electrotechnical Commission (IEC), 57
International standards organizations, 57
ionization chamber, 8, 10, 11, 16
ionization-based sensor, 4, 5, 7, 8, 10–11

J

J3Y transistors, 44–45
Joemex thermal circuit breaker, 62
joule energy rating, 55

K

Kidde Pi9000 smoke alarm, 3–18
KSP44 transistors, 65, 69

L

laser module, 214
laser specifications, 208

laser-guided sonic distance measurer, 203–227
LC resonance frequency calculator, 248
LCD (liquid crystal display)
  digital bathroom scale, 45–46
  digital hygro thermometer, 133, 139
  electronic pedometer, 86–87, 90
  sonic distance measurer, 214
leaded components, 296–298
leaded inductors, 297
LED power indicator, 284
LEDs (light emitting diodes)
  motion-activated, 19–33
  surge suppressors, 57, 66, 76
  ultrasonic humidifier, 121
  white, 29
LG suppression, 56
light emitting diodes. *See* LEDs
light sensitivity, 33
lightning strikes, 53, 55
lights
  compact fluorescent lamp, 95–111
  LED, 19–33
  motion-activated, 19–33
linear potentiometers, 90, 185, 186, 241
liquid crystal display. *See* LCD
lithium cells, 36, 39
LN suppression, 56
load cells, 37, 39, 42–43, 47–51
logic boards, 157–158, 165, 168–169, 172
low-dropout regulators, 27

M

magnetic guitar pickups, 243–255, 302
magnetic switch (MS), 128
memory
  EEPROM, 40, 41, 42, 46, 47
  pedometer, 84
  RAM, 41
MEMS accelerometer, 93–94
mercury thermometers, 139–140
mercury toxicity, 97
mercury vapor, 99, 103
metal film resistors, 66, 105–106
metal oxide varistors (MOVs)
  analog VOMs, 190
  AV600 surge suppressor, 55–56, 62–64
  considerations, 54
  ground, 80
  PL-Plus C power conditioner, 71, 75–76, 79–81
  thermal, 80
  TMOV upgrade, 80
microcontroller support ICs, 89
microcontrollers
  digital bathroom scale, 36, 41–42
  digital hygro thermometer, 137, 139
  electronic pedometer, 88–89
  sonic distance measurer, 215
  specifications, 304
MIDI interface, 253–255
modifications
  analog VOMs, 200–201
  AV600 surge suppressor, 69
  CF lamp, 110–111
  digital bathroom scale, 50–51
  digital hygro thermometer, 141–142
  dual-sensor smoke alarm, 18
  effects pedals, 272–274
  electric guitar, 250–255
  electronic pedometer, 94
  motion-activated LED light, 32–33
  PL-Plus C power conditioner, 80
  sonic distance measurer, 226–227
  stereo power amplifier, 172–173
  ultrasonic humidifier, 130–131
  vacuum tube guitar amplifier, 288–291
mods. *See* modifications
Monster Cable, 54, 55
Monster Power AV600 surge suppressor, 53, 54–69, 80–81
motion-activated LED light, 19–33
Mouser TMOV, 80
MOVs (metal oxide varistors)
  analog VOMs, 190
  AV600 surge suppressor, 55–56, 62–64
  considerations, 54
  ground, 80
  PL-Plus C power conditioner, 71, 75–76, 79–81
  thermal, 80
  TMOV upgrade, 80
MS (magnetic switch), 128
multimeter, 37, 59
multiuse circuit board, 4
music
  effects pedal, 257–274, 302
  electric guitar, 231–255, 304
  guitar magnetic pickups, 243–255, 302
  resources, 302
  vacuum tube guitar amplifier, 275–291

N

National Instruments Zone, 50
negative temperature coefficient (NTC) thermistor, 137
negative-positive-negative. *See* NPN
neon lamp, 65
Newline digital bathroom scale, 35–51
NPN (negative-positive-negative) transistor
  compact fluorescent lamp, 103

  digital bathroom scale, 44–45, 46
  dual-sensor smoke alarm, 9, 14, 17
  motion-activated LED light, 27–29
  stereo power amplifier, 162
  ultrasonic humidifier, 123–124
  vs. PNP transistors, 45
NTC (negative temperature coefficient) thermistor, 137
numeric notation, 295

O

ohmmeter, 9, 37, 45, 158, 197–200, 201
Ohm's law, 48, 167, 177, 198, 200
Omega Engineering, 50
Omron HJ-112 electronic pedometer, 83–94
operating temperature, 207
operational amplifiers, 216, 265–268, 304
operational considerations
  analog VOMs, 180–181
  AV600 surge suppressor, 57
  compact fluorescent lamp, 99
  digital bathroom scale, 37
  digital hygro thermometer, 135
  dual-sensor smoke alarm, 5
  effects pedals, 259–260
  electric guitar, 234–235
  electronic pedometer, 84–85
  Fender Champion guitar amplifier, 276
  motion-activated LED light, 21
  PL-Plus C power conditioner, 71
  sonic distance measurer, 208
  stereo power amplifier, 149
  surge suppressors, 57
  ultrasonic humidifier, 114–115
oscillators, 89

P

passive infrared (PIR) sensors, 20, 21, 25, 26, 32, 302
pedometer, electronic, 83–94
pendulum sensors, 84, 87–88, 92–94
pendulums, single-axis, 92–93
photoelectric sensor, 4, 7, 12–13, 92–94
piezo pickups, 253
piezoelectric sound detector, 93
piezoelectric transducer (PT)
  digital bathroom scale, 44, 46–47, 50
  distance, 5
  smoke alarm, 5, 7, 14, 15
  ultrasonic humidifier, 121, 126–127
PIR (passive infrared) detection, 30–32
PIR (passive infrared) sensors, 20, 21, 25, 26, 32, 302
PL-Plus C power conditioner, 54, 69–81
PNP (positive-negative-positive) transistors
  sonic distance measurer, 212, 217
  ultrasonic humidifier, 123–124
  vs. NPN transistors, 45
polypropylene film capacitor, 65–66
positive-negative-positive. *See* PNP
potentiometers, 90, 164, 185, 186, 241
power conditioners, 53, 54, 55, 69–81
power consumption, 148
power output, 147–148
power strips, 53, 54
power supplies, 53, 55, 127, 155, 166
power transformer, 124–125
precision, 179
precision resistors, 184–186, 296
printed circuit boards, 283
PT. *See* piezoelectric transducer

Q

quartz crystals, 138, 217–218

R

radiation
  alpha, 7, 8
  dual-sensor smoke alarm, 5, 7, 8
  Geiger counter, 7, 8
  PIR. *See PIR entries*
  ultraviolet, 99, 103
radiation levels, 7
RadioShack 22-109 VOM, 177–201
RAM (Random Access Memory), 41
Random Access Memory (RAM), 41
receiver transducers, 205–206
regulators. *See* voltage regulators
relative humidity (RH) sensor, 134, 138
relays, 77, 165
resistance bridge, 47–50
resistors
  cathode, 287–288
  color-code system, 296–297
  effects pedal, 265
  fusible, 106
  marking systems, 296–297
  metal film, 66, 105–106
  precision, 184–186, 296
  SMT. *See* SMT resistors
  sonic distance measurer, 219
  specifications, 304
  stereo power amplifier, 163
  ultrasonic humidifier, 123
  vacuum tube guitar amplifier, 284
  variable, 26, 50, 167, 185, 196
resolution, 179, 206
resonance, 247–249

resonant frequencies, 128, 138, 225, 248–249
resources, 301–304
    books/documents, 301
    specifications, 303–304
    topic-specific sources, 301–302
response time, 56
RF inductors, 122
RH (relative humidity) sensor, 134, 138
rocker switch
    power conditioner, 77
    surge protector, 57, 61, 65, 67–68
    ultrasonic humidifier, 122
rotary switches, 188–189

S

safety issues
    Americium caution, 5, 6–7, 9, 18
    compact fluorescent lamp, 96, 97, 99
    ionization chamber, 8
    ionization sensor, 7, 8
    laser beams, 208
    radiation. *See* radiation
    smoke alarm teardown, 5, 6–9, 18
    stereo power amplifier, 155
    ultrasonic humidifier, 117
    vacuum tube guitar amplifier, 277
scale, digital bathroom, 35–51
schematics
    CF lamp, 108–110
    digital bathroom scale, 40, 46–47
    digital hygro thermometer, 140–141
    electronic pedometer, 90–91
    motion-activated LED light, 31
    power conditioners, 77–79
    smoke alarm, 16, 17
    sonic distance measurer, 222–226
    stereo power amplifier, 166–171
    surge suppressors, 66–67
    vacuum tube guitar amplifier, 285–287
Schottky diodes, 186–188, 264
semiconductors, 44, 302, 303
sensors
    CDS, 23, 26–27, 32, 33
    pendulum, 84, 87–88, 92–94
    photoelectric, 4, 7, 12–13, 92–94
    PIR, 20, 21, 25, 26, 32, 302
    temperature, 133, 304
shunt regulators, 163, 166–167
signal-to-noise ratio (SNR), 49, 145–146, 225, 232
silicon diodes, 77, 187–188, 263–264, 273, 285
silicone-carbon elastomeric buttons, 90
Simulation Program with Integrated Circuit Emphasis (SPICE), 248
single pole, single throw (SPST) rocker switch, 122
small outline integrated circuit (SOIC), 24, 25
small outline transistor. *See* SOT
smoke alarms, 3–18
SMT (surface mount technology), 20, 184
SMT board, 29
SMT capacitors, 29, 298–299
SMT components, 298–300
SMT potentiometers, 90
SMT resistors
    analog VOMs, 184–185
    digital bathroom scale, 40, 41
    motion-activated LED light, 29
    overview, 299–300
SMT Schottky diodes, 186–188, 264
SNR (signal-to-noise ratio), 49, 145–146, 225, 232
SOIC (small outline integrated circuit), 24, 25
Song Chaun relay, 77
sonic distance measurer, 203–227
SOT (small outline transistor) packaging, 27–28, 44–47
SOT-23 case, 45
SOT-89 package, 27, 31, 215
SOT-90 case, 27
SOT-323 package, 27, 28, 31
SPDs (surge protective devices), 53–81
speakers, 283, 289, 290, 304
specifications
    AMP 100 stereo amplifier, 144–149
    AV600 surge suppressor, 55–57
    capacitors, 303
    components, 303–304
    Crane ultrasonic humidifier, 113–114
    digital bathroom scale, 36–37
    digital hygro thermometer, 134
    diodes, 303
    dual-sensor smoke alarm, 5, 7, 8
    effects pedals, 258–259
    electric guitar components, 304
    electronic pedometer, 84
    Fender Champion guitar amplifier, 276
    Fender Telecaster electric guitar, 232–233
    inductors, 304
    laser, 208
    microcontrollers, 304
    motion-activated LED light, 20
    PL-Plus C power conditioner, 70–71
    RadioShack 22-109 VOM, 178–180
    resistors, 304
    resources, 303–304
    speakers, 304
    TCP SpringLamp, 97–98
    transformers, 304
    transistors, 304
    vacuum tubes, 304
    Zircon Sonic Measure DM S50L, 205–208

speech synthesizer, 33
SPICE (Simulation Program with Integrated Circuit Emphasis), 248
spike diversion, 56
spike-clamping voltage, 70, 81
spike-diversion system, 67–68
spikes, 54, 55–56, 79
SPST (single pole, single throw) rocker switch, 122
stacked potentiometers, 164
stereo power amplifier, 143–173
strain gauges, 36, 37, 39, 42–44, 47–51
superglue, 51
surface mount technology. *See* SMT
surge protective devices (SPDs), 53–81
surge suppressors, 53, 54–69, 80–81
surges, 54
switch indicators, 65
switches
    dual-sensor smoke alarm, 16
    electric guitar, 242
    magnetic, 128
    reed, 122
    rocker. *See* rocker switch
    rotary, 188–189
    sonic distance measurer, 219
    SPST, 122
    tone, 273
    ultrasonic humidifier, 122
symmetrical distortion, 269

T

tape measure, sonic distance, 203–227
TCP SpringLamp, 95–111
TDA7294 power ICs, 155, 158–159, 169–171
teardowns
    analog VOMs, 177–201
    compact fluorescent lamp, 95–111
    digital bathroom scale, 35–51
    digital hygro thermometer, 133–142
    dual-sensor smoke alarm, 3–18
    effects pedal, 257–274
    electric guitar, 231–255
    electronic pedometer, 83–94
    motion-activated LED light, 19–33
    resources, 301–304
    sonic distance measurer, 203–227
    stereo power amplifier, 143–173
    surge protective devices, 53–81
    ultrasonic humidifier, 113–131
    vacuum tube guitar amplifier, 275–291
temperature sensors, 133, 304
THD+N (total harmonic distortion plus noise), 146–147
thermal circuit breaker, 62, 67, 76
thermal cutoff, 64
thermal fuses, 64, 76
thermal management, 289
thermal MOV (TMOV), 80
thermistors, 137–139, 213–214
thermo-fuse varistor (TVZ), 71
thermometers
    hygro, 133–142, 223
    mercury, 139–140
TL0001S IC, 24–25, 32, 33
TMOV (thermal MOV), 80
TO (Transistor Outline) packages, 17
tone switches, 273
tools/instruments
    analog VOMs, 181
    compact fluorescent lamp, 100–101
    digital bathroom scale, 37
    digital hygro thermometer, 135
    dual-sensor smoke alarm, 7
    effects pedals, 261–262
    electric guitar, 235–237
    electronic pedometer, 85
    motion-activated LED light, 21
    power conditioners, 72–73
    sonic distance measurer, 210
    stereo power amplifier, 149
    surge suppressors, 59
    ultrasonic humidifier, 117
    vacuum tube guitar amplifier, 277
toroidal transformers, 56, 104–105, 153, 162
total harmonic distortion plus noise (THD+N), 146–147
transducers. *See also* piezoelectric transducer
    pendulum, 90–92
    ultrasonic, 115, 117–119, 213, 217, 225
transformers
    audio output, 290
    power, 124–125
    specifications, 304
    toroidal, 56, 104–105, 153, 162
    vacuum tube guitar amplifier, 283
    variable, 59
transient suppression voltage (TSV), 56, 62–63
transient voltage suppression. *See* TVS
Transistor Outline (TO) packages, 17
transistors
    bipolar, 123, 124
    compact fluorescent lamp, 103–104
    considerations, 44–45
    J3Y, 44–45
    KSP44, 65, 69
    NPN. *See* NPN
    PNP, 45, 123–124
    sonic distance measurer, 217
    SOT packaging, 27–28, 44–47

transistors *(Continued)*
  specifications, 304
  ultrasonic humidifier, 123–124
triggers, enhanced, 50
TSV (transient suppression voltage), 56, 62–63
tube amplifiers, 302
tubes, vacuum, 279, 282, 289, 304
TÜVRheinland web site, 71
TVS (transient voltage suppression), 56
TVS diodes, 56
TVZ (thermo-fuse varistor), 71

U

UL 1449 compliance, 81
UL 1449 rating, 57
ultrasonic humidifier, 113–131
ultrasonic oscillator, 126–127
ultrasonic transducers, 115, 117–119, 213, 217, 225
ultrasound sensing methods, 30, 31
ultraviolet (UV) radiation, 99, 103
Underwriters Laboratories, 57
uninterruptible power supply (UPS), 53, 55
UPS (uninterruptible power supply), 53, 55
UV (ultraviolet) radiation, 99, 103

V

vacuum tube guitar amplifier, 275–291
vacuum tubes, 279, 282, 289, 304
variable resistors, 26, 50, 167, 185, 196
Variac variable transformer, 59
voltage detectors, 89, 91–92
voltage hazards, 97
voltage regulators
  7130 regulator, 27–28, 31, 32, 33, 215
  LM317LM regulator, 212–213, 215, 216, 223
  low-dropout, 27
  motion-activated LED, 27–28, 31, 32
  shunt, 163, 166–167
  sonic distance measurer, 215–216
voltage spikes/surges, 54, 55–56, 79
voltage-current plot, 63, 69
Volt-Ohm-Meter (VOM), 177–201
VOM (Volt-Ohm-Meter), 177–201

W

web sites, resources, 301–302
Wheatstone bridge, 47–50
Wikipedia, 301
wire umper, 80
wiring, 284

Z

Zebra Elastomeric Connectors, 45–46, 86, 90, 302
zener diode, 16, 167
Zircon Sonic Measure DM S50L, 203–227

More than 80 photos

The Basic Soldering Guide

Learn to solder successfully!

Alan Winstanley

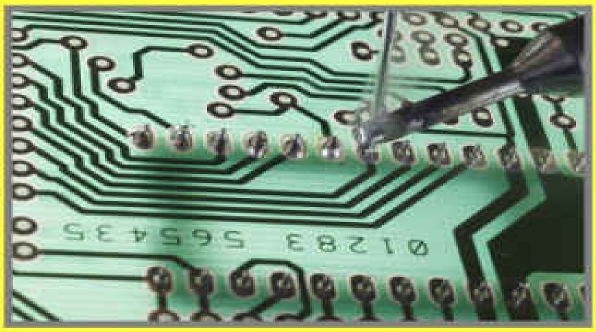

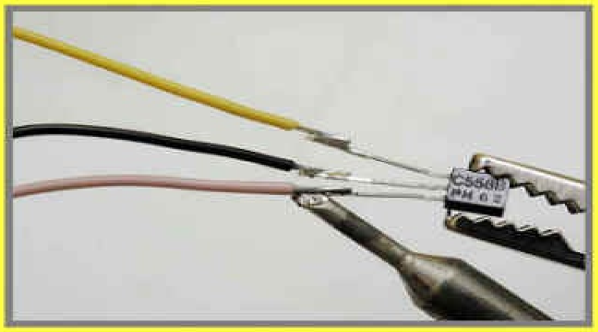

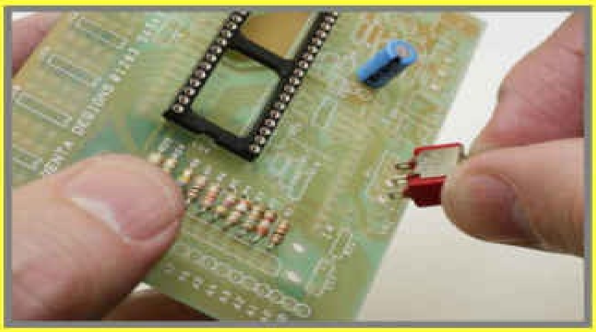

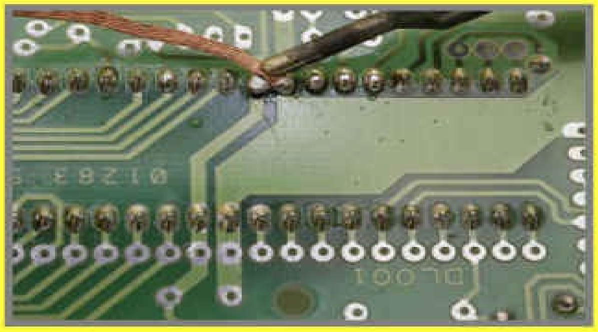

In association with

ANTEX
ANTEX (Electronics) Limited

The Basic Soldering Guide

by Alan Winstanley

EPE Magazine Online Editor

In association with

Antex (Electronics) Ltd.

ANTEX

ANTEX (Electronics) Limited

Copyright Notice and Disclaimer

All text, diagrams and photographs are Copyright © 1996-2013 A. R. Winstanley. All Rights Reserved.

All text, diagrams and photographs are fully protected by copyright and may not be reproduced in any commercial or non-commercial publication or medium, mirrored on any commercial or non-commercial website nor may they be appended to or amended, monetised or used or re-distributed for any commercial reason, without the prior permission of the writer.

Every care has been taken to ensure that the information and guidance given is accurate and reliable, but since conditions of use are beyond the writer's control no legal liability or consequential claims will be accepted for any errors contained herein.

Where mentioned, the U.K. mains voltage supply is 230V a.c. and you should amend ratings for your local conditions.

Preface

In 1996 when the world wide web was still very young, I launched the first and most detailed website ever describing the practical skills of electronic soldering, and my **Basic Soldering Guide** quickly became the #1 web site of its kind in the Google search engine.

Thanks to its in-depth reference text and the unequalled high-quality close-up photography showing soldering step by step, many quickly learned the essential stages needed to make a solder joint successfully. Even novices who had never tried soldering before, gained the skills and confidence needed to acquire this skill.

My Basic Soldering Guide became a key *go-to* online guide for soldering, and I've enjoyed receiving encouraging feedback ever since from the likes of the US Air Force, US Marines, US Coastguard, Honeywell trainers, Atomic Energy authorities, Australian aeronautical suppliers, UK colleges and universities, trainees and many more around the world.

In association with Antex (Electronics) Ltd., the leading UK manufacturer of electronic soldering equipment, I'm delighted to bring you this updated Basic Soldering Guide containing over 80 all-new colour photographs, more background, more detailed information and lots more practical hints and tips.

I've revisited various areas of the guide and refreshed them, taking onboard readers' queries and nearly 17 years of online feedback, experiences and answering reader's questions. With the help of all-new photography, I'm sure you'll master the skills needed to solder electronics successfully using this updated guide.

I welcome feedback and comments, and readers can reach me via my website **www.alanwinstanley.com** or Email me at **alan@epemag.demon.co.uk**

Alan Winstanley

July 2013

Table of Contents

Copyright Notice and Disclaimer
Preface
Introduction
First steps
Soldering iron choices
How to solder
Next steps
Cleanliness and “Tinning” the bit
Getting Ready
Solder and Fluxes
Flux dispensers and Colophony
Temperature flow
Now is the time…
Soldering Step by Step
Interwiring – get hooked up!
Wire joints
Tidying up
Reflow technique
Fatigue & Breakage
Faults & Desoldering techniques
Desolder Braid
Quick Summary Guide
Troubleshooting Guide
Possible Hazards and simple First Aid
Useful resources
Acknowledgements
Photography
Conclusion

Introduction

The first and most important aspect of assembling any electronic project is that of *soldering,* which is a delicate and precise skill that can be mastered with experience. Sometimes called "soft soldering", there's no shortcut to acquiring the necessary expertise, and producing a consistently satisfactory solder joint takes a little practice. However, like riding a bicycle, soldering is an art which once learnt is never forgotten, and the purpose of this new and updated guide is to explain the techniques of soldering and desoldering for beginners, which I hope will set the hobbyist or trainee technician firmly on the road to successful electronic assembly or repairs in the future.

Soldering is the least "aggressive" way of joining non-ferrous metals together, and is used universally in electronics, air conditioning and refrigeration circuits, household plumbing and more besides – applications where the precise joining together of components at fairly moderate temperatures is needed. Further up the scale, **brazing** involves using higher temperatures to melt brazing rods onto larger metal parts, perhaps to repair a metal chair, lawnmower or to fabricate metal components or jewellery into intricate shapes. Lastly, **welding** is a very aggressive way of fabrication using welding rods or wire; steel girders, oil rigs and ships are all welded together, or robotic spot-welding is used for the mass production of, say, washing machines or car bodyshells using sheet steel to make strong rigid assemblies.

Due to the lower temperatures used and the need to make consistently good electrically conductive and mechanically sound joints with precision, soldering is used to connect components together when manufacturing electronic circuits. Small components would quickly be destroyed by brazing or welding, although tiny spot-welding joints do appear in electronics, perhaps to weld a metal tag onto a button battery.

This guide therefore deals with the soldering techniques used in electronics at hobbyist or trainee educational level. It explains what to look for before buying a soldering iron, describes ways of making various solder joints on circuit boards and other electronic components, and also how to *desolder* – removing solder in order to repair a circuit board or replace an electronic component.

You'll also find more details of other aspects of soldering, including an outline of typical solder types and fluxes. In short, everything you need to get started in electronics soldering is here, so let's get started!

First steps

The principle behind soldering sounds quite simple: the idea is to join components together to form an electrical connection, by using a mixture of lead and tin solder or alternatively “lead-free” solder (an alloy of tin and copper), which is melted onto the joint using a *soldering iron*. If you have never picked up a soldering iron before, then this guide will show you everything to help you start soldering with confidence. I also hope that the guidance will help those working in other areas — computer technicians or audio enthusiasts, for example, who may be faced with electronic repairs or modifications using a soldering iron for the first time.

If you’re an electronics hobbyist or trainee, before embarking on any form of ambitious electronic project, it is recommended that you practice your soldering technique on some *brand new* components using *clean* strip board (or protoboard) or a printed circuit board, and select a simple and straightforward constructional design as a starting point. Become acquainted and comfortable with your chosen soldering iron, which likely to become as familiar to you as a favourite pen. Learn how to balance it and handle it with precision. Try soldering an assortment of resistors, capacitors, diodes, transistors and integrated circuits with it, and then try your hand at *desoldering* – removing the solder again to make a repair or modification.

A really good place to start learning is by building a simple electronics kit, such as a Velleman kit that contains a good quality printed circuit board. You’ll learn some of the basic skills of successful soldering this way, and it’ll be a great confidence booster too.

Did you know? In the USA and elsewhere, the letter L in “solder” is silent and they say “soda” or “sodder” – but here in Britain we do pronounce the L and we say “sole-der”.

Soldering iron choices

Search any electronics catalogue or website and you'll see a bewildering array of soldering equipment on sale, including irons, controllers, work stations and desoldering equipment too. A large range of soldering irons is readily obtainable - which one is suitable for you depends on your budget and how serious your interest in electronics is, but there's something for every pocket distributed by a variety of retail, industrial and mail-order outlets.

The Antex range of soldering equipment has been very popular with industry, education and the electronics hobbyist for 60 years; I grew up with an Antex iron and a trusty 15W Antex iron was an everyday part of my hobby electronics all the way through the 1980's. An industrial user or a more dedicated hobbyist – with a bigger budget! – will be interested in a soldering station instead and again Antex offers a range of British-made products for industry or home use.

A very basic mains electric soldering iron can cost from under £5 (US$ 8), but I find that these very cheap irons, as sold on auction websites, are pretty crude and imprecise. They are best suited for simple electrical repairs and DIY rather than precision electronics or printed circuit boards discussed here. They tend to be bulky and uncomfortable for extended use, and they may not have suitable "bits" or tips of various sizes to suit different tasks.

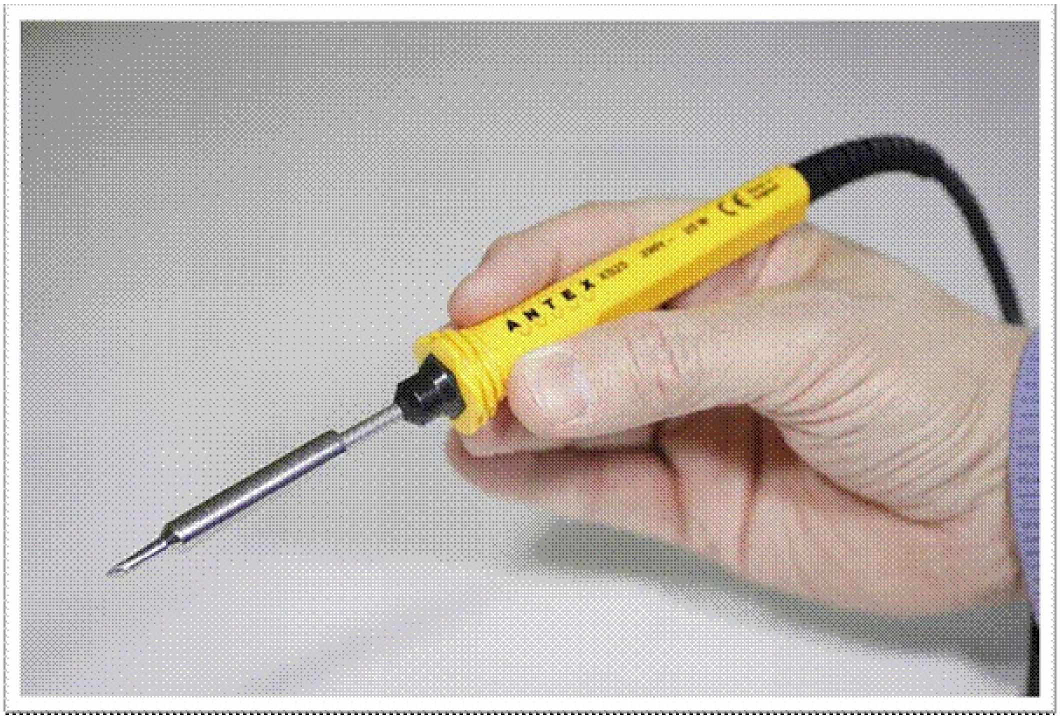

^^ *This classic Antex XS25 "pencil style" 25W mains-powered soldering iron has* *exchangeable tips or "bits" and is great for general hobbyist or educational use. Stands are also available to store them safely in between use.*

A quality pencil-style electric soldering iron such as the Antex XS25 (photo) will be approximately £20 (US$18 tax free) - though it's possible to spend into three figures on a soldering iron "station" if you're really serious about the subject! Don't be tempted to over-spend on an elaborate workstation though, unless you are really very serious about becoming involved in electronics. You will usually obtain perfectly satisfactory results using a fairly modest "pencil" iron model, and you can upgrade to something more sophisticated should your needs change in the future.

ANTEX 660TC

^^ *For enthusiasts or industry, this Antex 660TC soldering station has a separate mains-powered control unit and a matching low-voltage soldering iron rated at 24 Volts, 50 Watts so it's suited to a wider range of tasks than a lower powered one.*

When choosing your soldering iron, certain factors to bear in mind include:

Voltage: for the British market, "mains" electric irons run directly from the mains at 230V a.c. or will be set for other voltages (110V a.c.) depending on the country. However, **low voltage types** (e.g. 12V or 24V) usually form part of a "soldering station" for use with a matching controller made by the same manufacturer. Some low-voltage irons run off batteries (e.g. a car battery or Ni-cads) but these are uncommon.

Wattage: this is an extremely important factor to think about when choosing your iron. Typically, irons for general electronics work may have a power rating of between 15-25 watts or so, which is fine for most electronic assembly tasks, printed circuit boards and inter-wiring. It's important to note that a higher wattage does *not* mean that the iron runs hotter - it simply means that there is more power "in reserve" for coping with larger joints. The maximum electric iron wattage generally available is about 100W, which is OK for DIY electrical repairs but is far too high for general electronics or circuit board use.

A higher wattage iron offers you more flexibility for tackling a wider range of tasks. It has a better "recovery rate" which makes it more "unstoppable" when it comes to heavier-duty work, because it won't be drained of its heat so quickly. So check the power ratings carefully, and anything between 15–40W is fine for general electronics soldering.

Temperature Control - the simplest and cheapest types don't have any form of temperature "regulation". Simply plug them into the mains and switch them on! Thermal regulation is "designed in" (by physics, not electronics!). Sometimes they are described as "thermally balanced" as they have some degree of temperature "matching" – in other words, they warm up as quickly as they lose heat during use, so in a primitive way they maintain roughly a constant temperature. This type of iron is perfectly acceptable for hobby or less demanding professional use. It's essential to use the manufacturer's specified tips (see later) to maintain proper temperature matching, otherwise the iron may not heat up enough – or it may overshoot in temperature.

These unregulated irons form an ideal general-purpose iron for most users, and they cope reasonably well with printed circuit board soldering and general interwiring. However, most of these "miniature" types of iron will be of little use when attempting to solder large joints (e.g. very large terminals or very thick copper wires) because the components being soldered will draw or "sink" heat away from the tip of the iron, cooling it down too much and preventing solder from flowing properly. That's when a higher wattage iron is needed.

A proper **temperature-controlled iron** will be quite a lot more expensive - retailing at say £40 (US$ 60) or more - and will have some form of built-in thermostatic control, to ensure that the temperature of the "bit" (the tip of the iron) is maintained at a fixed level within reasonable limits. This is desirable during frequent use, since it helps to ensure that the temperature will be relatively stable regardless of the workload. Some irons have a bimetallic strip thermostat built into the handle which gives an audible "click" in use, and some may include an adjustable screwdriver control within the handle as well. Others may have electronic controls built in.

^^ *An Antex 660TC soldering station with matching iron and bench stand.*

More expensive still, **soldering stations** cost from £70 (US$ 115) upwards (the iron may be sold separately, so you can pick the type you prefer), and consist of a complete bench-top control unit into which a special *low-voltage* soldering iron is plugged. Some versions might have a built-in digital temperature readout, and a control knob to vary the setting. The temperature could be boosted for soldering larger joints, for example, or for using higher melting-point solders (e.g. lead-free or silver solder). These are designed for the discerning user, or for continuous production line or professional use. A thermocouple will be built into the tip or shaft, which monitors temperature.

The best soldering stations have irons which are well balanced, with comfort-grip handles which remain cool all day, and silicone-based cables which are burn proof. Antex produces a range of irons with silicone cables specially for education use, to help avoid accidents caused by careless use by students.

Anti-static protection: if you need to solder a lot of static-sensitive parts (e.g. CMOS chips or MOSFET transistors), more advanced and expensive soldering iron stations use static-dissipative materials in their construction to avoid static charges accumulating on the iron itself, which could otherwise damage or "zap" some semiconductors. Such irons are listed as "ESD safe" (electro-static discharge proof). The cheapest irons are not classed as ESD-safe but they still perform well enough in most hobby or educational applications provided you take the usual anti-static precautions when handling these components. The iron would need to be well earthed (grounded) in these circumstances, to carry away any static.

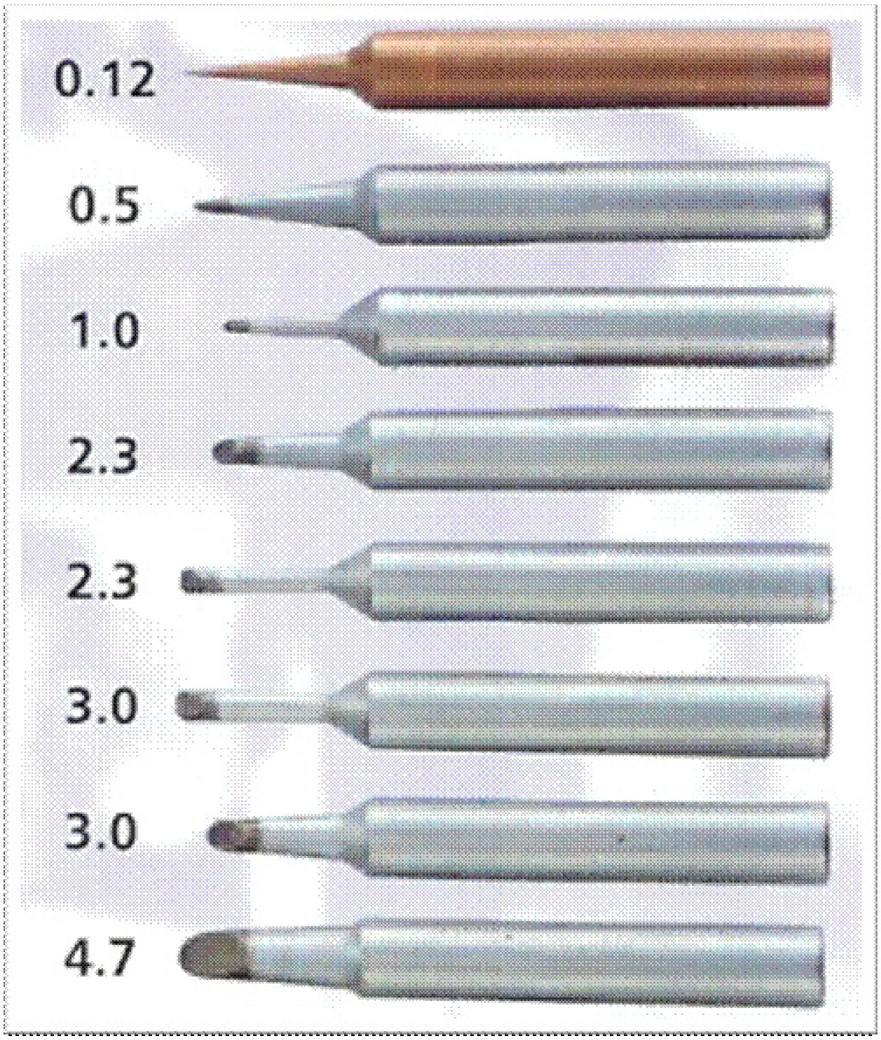

^^ *Just part of the range of spare tips or "bits" produced by Antex for their soldering irons. Use only tips intended for a particular iron, to avoid thermal matching problems.*

Tips or Bits: it's often useful to have a small selection of manufacturer's bits (soldering iron tips) available with different diameters or shapes, which can be changed depending on the type of work in hand. You will probably find that you become accustomed to, and work best with, one particular shape of tip for the majority of your work. Usually, tips are iron-coated or plated to preserve their life and to maintain good tip "hygiene". Be sure only to use tips that are specifically designed for your iron, otherwise thermal problems may arise. I show separately some typical tips, courtesy of Antex.

^^ *Useful accessories from Antex including solder, tips and tip cleaner, desoldering braid and a heatshunt.*

Spare parts: it's always reassuring to know that spares are available in the future if required, so if the element blows, you don't need to replace the entire iron. This is especially the case with expensive irons. Check some websites or mail-order catalogues to see whether spare parts are listed. One drawback is that you may need another soldering iron when exchanging a broken heating element!

^^ *Gascat 40 gas-powered soldering iron kit by Antex, with spare tips and cleaning sponge.*

Gas or electric? So far I've discussed electric soldering irons, but **gas-powered**

soldering irons are sold which use butane propellant rather than mains electricity to operate. They have a flint lighter or (better still) a built-in piezo for ignition, and have a catalytic element which, once warmed up, continues to glow hot when gas passes over them. They tend to be big and bulky compared to electric pencil irons.

Field service engineers use gas-powered irons for working on repairs where there may be no power available, or where a joint is tricky to reach with a normal electric iron, so they are really for occasional "on the spot" use, rather than for mainstream construction or assembly work. I use one when I just need to do a quick repair and can't be bothered getting the electric soldering iron going!

Gas irons can have higher power equivalents than electric ones (eg 125 watts or more) but some gas-powered irons are nothing more than miniaturised blowtorches, which may or may not be useful for occasional heavier duty soldering. In the author's experience they can be difficult to use in confined areas. *Extreme care is needed at all times to ensure hot gas emitted from the iron's exhaust port doesn't damage nearby materials, plastics or wiring.* Gas irons can have useful accessories to convert them into e.g. a hot knife for sealing nylon rope, or a hot-air blower for use with heatshrink tubing. Almost every electronics constructor uses an electric-powered iron though.

A **solder gun** is a pistol-shaped iron, typically running at 100W or more, and is completely unsuitable for soldering modern electronic components: they're too hot, heavy and unwieldy for micro-electronics use, nor are they designed for that. Plumbing or DIY, maybe..!

^^ *A heat resistant soldering iron stand with cleaning sponge. (Antex)*

Soldering irons are best used along with a **heat-resistant bench stand**, where the hot iron can be safely stored in between use (photo). It is extremely important that a hot soldering iron is safely “parked” ready for action, and a bench stand is really a necessity. Soldering stations usually have such a feature, otherwise a separate soldering iron stand is essential, ideally one that’s supplied with a tip-cleaning sponge. You can make your own cleaning sponges using *cellulose* sponge only.

^^ *A benchtop fume extractor fan for hobbyist use. A replaceable carbon filter helps remove particles and air is vented out the back.*

Other equipment worth considering includes the use of **fume extractors**, which are compulsory in the industrial workplace. Solder fumes and flux smoke are not known to be toxic but they can cause irritation. A basic fume extractor (photo) consists of a small bench-top fan which draw fumes and irritating smoke away from the operator’s face and filters out some of the smoke particles, before exhausting the air back out through the fan. The carbon-impregnated foam filters are replaceable. Such devices are quite effective and users soon find them indispensable, even though they can be a bit noisy at close range.

Professional fume extraction systems draw the smoke and fumes directly from the work area via a clip-on tube fitted to the soldering iron, then vent the fumes away through a large filter pump. It is definitely worth considering a small bench top unit for regular hobby or occasional professional use, as having decent ventilation can only be a good thing.

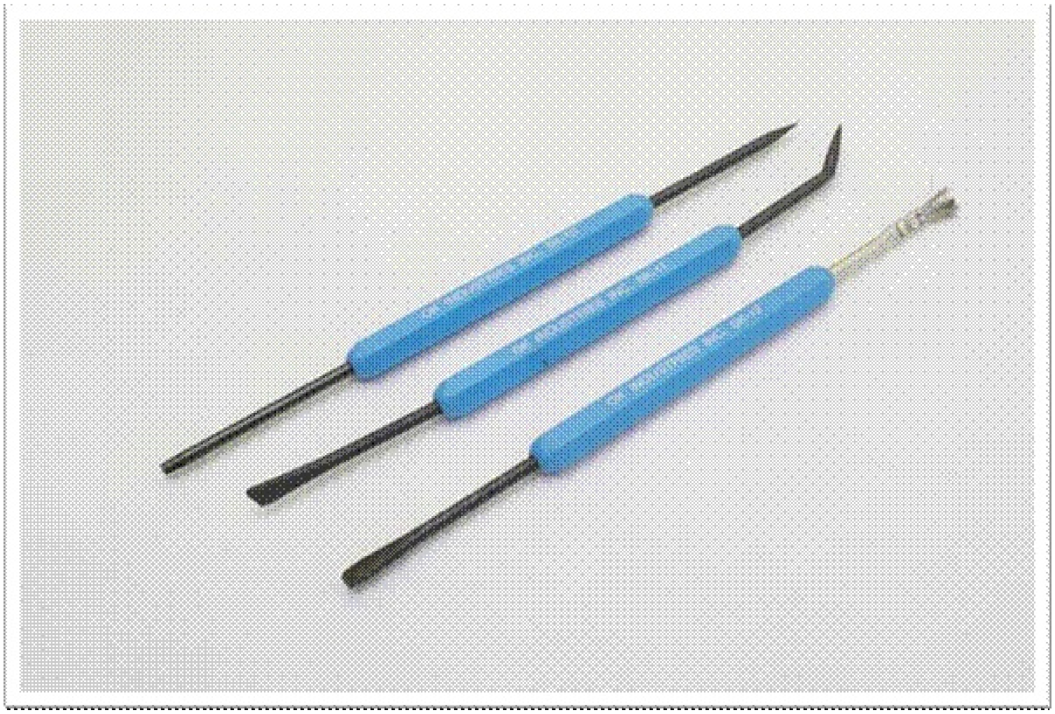

^^ *A soldering accessory toolkit by OK Industries, including probes, scrapers and wire brush*

A variety of specialist **hand tools** are available that assist with soldering, and a good supplier's catalogue will offer a range of small brushes, scrapers and cleaning tools in a handy kit, together with the usual types of wire cutters, pliers and so forth, which are necessary for handling components and tidying up as required. Some specialist service aids (aerosols etc.) are also described later.

Now let's look at how to use soldering irons properly, and later on we will describe the techniques for putting things right when a joint somehow goes wrong — and don't worry, even the experts get it wrong sometimes!

How to solder

This guide will show you step by step how to solder successfully and plenty of photographs are provided to help explain the techniques needed.As you read through the guide I'll explain all the stages in more detail, but let's look at the basics first.

First of all, successful soldering requires that the items being soldered together are held with as little movement as possible. So it's best to *secure the work* as needed, so that your accuracy isn't affected should the workpiece move accidentally.

In the case of a printed circuit board, various holding frames are useful especially when densely populated boards are being soldered: the idea is to insert all the parts on one side (a process called "stuffing the board"), hold them in place with a suitable pad to prevent them falling out again, turn the board over and then snip off the wires with cutters before soldering the joints. The frame saves an awful lot of turning the board over and back again, especially with large boards: all the soldering can be performed in one pass.

^^ The ever-popular "Helping Hands" (left) helps support sundry parts, wires etc. during soldering.
A modeller's vice (right) holds parts firmly. A vacuum base fixes it onto smooth surfaces.

Only the more serious constructor will purchase a holding frame, and hobbyists can retain parts in place by improvising in a variety of ways – the ever popular "Helping

Hands" stands cost a few pounds and is widely sold. They have adjustable crocodile clips to grip parts, and maybe a magnifying glass or soldering iron stand as well. The cast iron base provides stability. Other parts could be held firm in a modeller's small vice, for example.

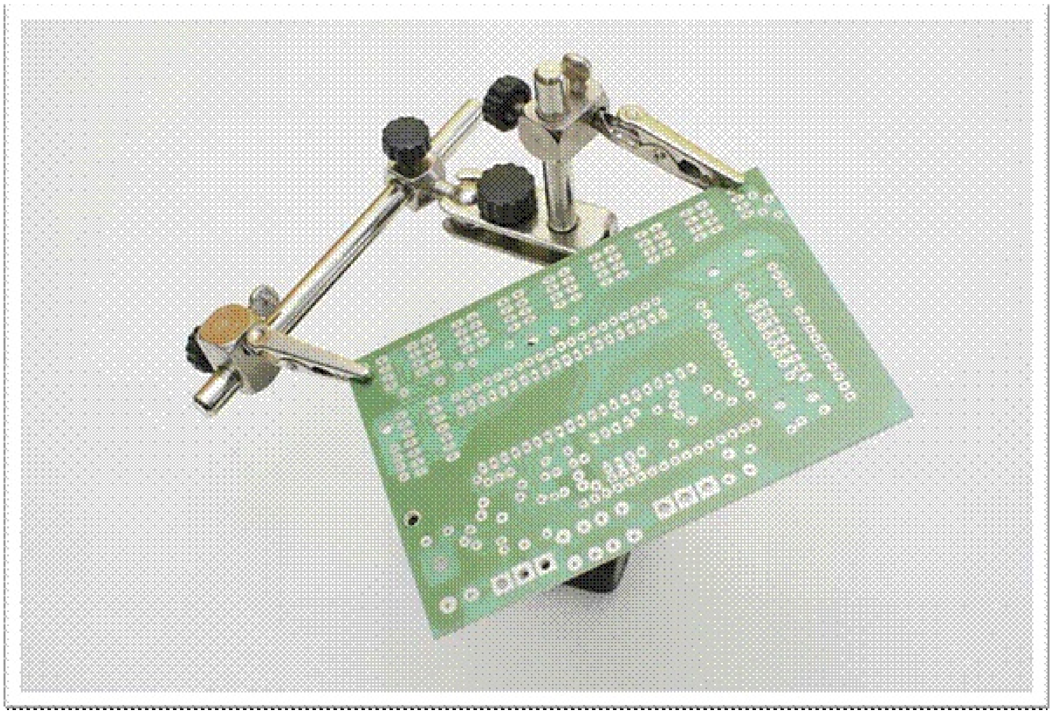

^^ *"Helping Hands" uses crocodile clips to grip parts during soldering. Or just place parts flat on the bench.*

When soldering parts onto an ordinary circuit board, components' wires can simply be bent to the correct pitch (distance apart) to fit through the board, insert the part flush down against the board's surface, splay the wires a little so that the component grips the board under spring tension, and then solder it. This technique is universally used in manual **through-hole soldering,** which is explained in full later.

In the author's view - opinions vary – it's generally better to snip off the surplus wires leads first, to make the joint and neighbouring joints more accessible and also to reduce the mechanical shock transmitted to the p.c.b. copper foil. However, in the case of diodes and transistors the author tends to leave the snipping until after the joint has been made, since the excess wire will help to "sink" heat away from the heat-sensitive semiconductor junction.

A special clip-on heatsink is available which also helps stop excess heat from reaching temperature-sensitive semiconductors like these. I've always managed without one but beginners might find them reassuring. Integrated circuits can either be soldered directly into place if you are confident enough, or better, use a dual-in-line socket to prevent heat damage. The chip can then be swapped out at a later date if needed.

Parts which become hot in operation (e.g. some power resistors), should be raised above the board slightly to allow air to circulate. Some components, especially large electrolytic capacitors, may require a mounting clip to be screwed down to the board first, otherwise the part may eventually break off due to vibration. It’s a good idea to bolt such components firmly into place before soldering their terminals, in order not to strain the soldered joints or the components when fasteners are tightened.

By securing the workpiece as much as possible to prevent movement, you have a much higher chance of producing good-quality reliable solder joints.

Next steps

Let’s get to grips with actual soldering techniques in more detail. The soldering of electronics components utilises **lead/ tin** or **lead-free solder** and the process is compatible with many non-ferrous metals. You can solder copper, lead, brass, gold plate, silver, nickel, tin and tin plate, zinc and more besides but some metals such as nichrome, galvanised or stainless steel require a highly specialist “flux” (see later) to solder them and aren’t discussed here. Some materials such as beryllium, chromium, magnesium and titanium are non-solderable in any case, according to solder manufacturers Multicore.

In electronics we’re mainly concerned with soldering parts or wires onto printed circuit boards or terminals that are usually already “tinned” with solder or plated, ready for soldering with flux-cored solder. The key factors affecting the quality of a solder joint are:

- **Cleanliness** – dirt or impurities drastically hinder good solder coverage.
- **Temperature** – the right level to enable the solder to flow freely!
- **Time** – apply heat for just the right amount of time!
- **Adequate solder coverage** – enough to form a good joint without touching neighbouring areas.

A little effort spent now in soldering the perfect joint may save you — or somebody else —a considerable amount of time in troubleshooting a defective joint in the future. Let’s discuss the basic principles outlined above in more depth.

Cleanliness and "Tinning" the bit

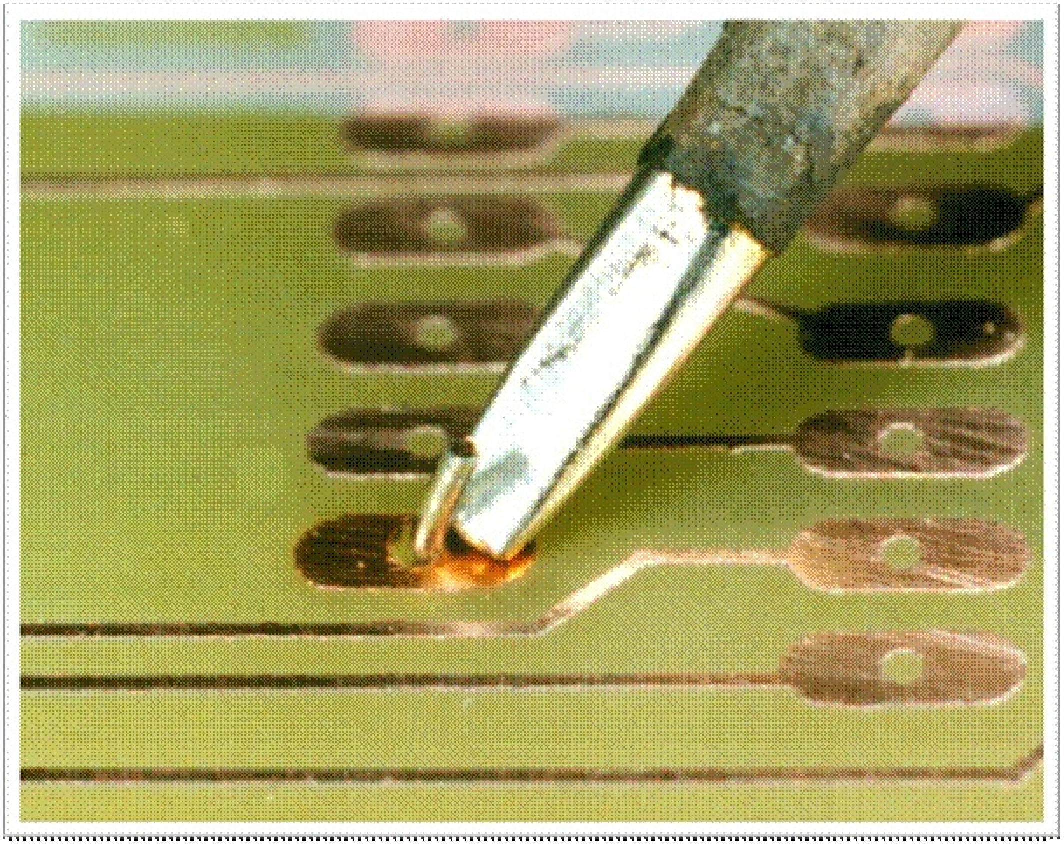

^^ *A clean and shiny soldering tip or "bit" is essential for each and every solder joint you make.*

Firstly, and without exception, all parts - including the soldering iron's tip itself - must be clean and free from contamination. Fact is, solder just will not "take" to dirty parts! Old components or copper board can be impossible to solder because of oxidation that builds up on the surface of the leads. This repels the molten solder and will soon become evident because the solder will "bead" into shiny globules looking like mercury, going everywhere except where you need it. *Dirt and contamination are the enemies of a good quality soldered joint!*

When all the conditions are right for soldering, materials are said to be "wettable" and will accept molten solder, which should flow readily over their surfaces. Hence, it's **really necessary** to ensure that parts are free from grease, oxidation and other contaminants. Note that in any case, some incompatible materials or surface finishes just cannot be soldered using ordinary lead-free solder, no matter how hard you try – e.g. aluminium parts would require special aluminium solder and fluxes (see later) to be used.

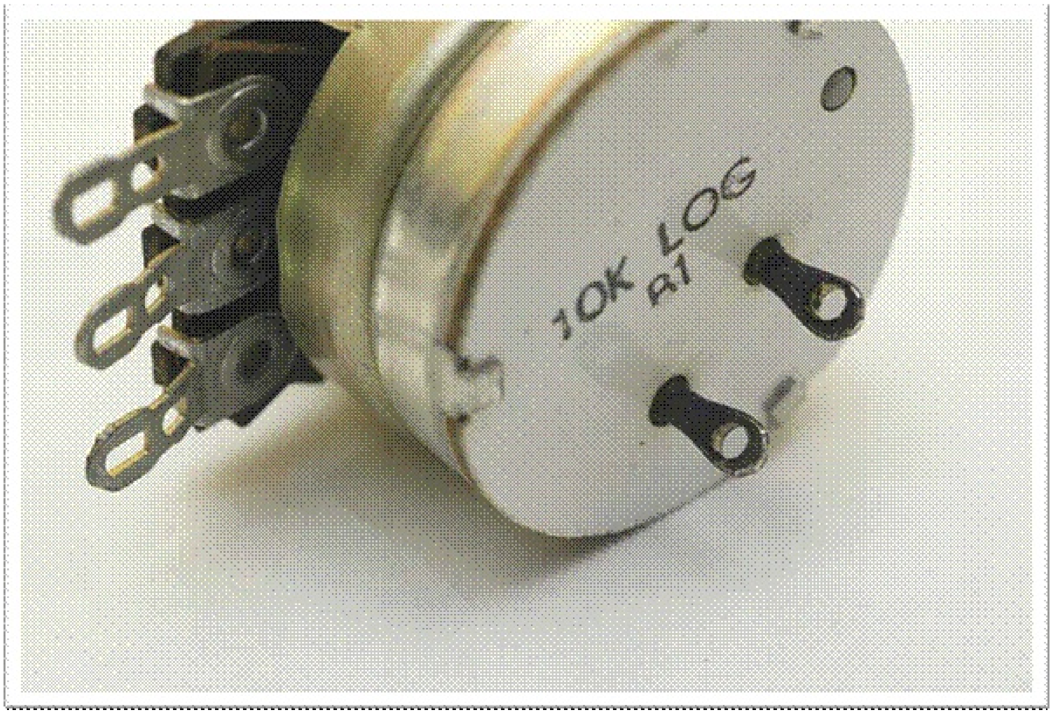

^^ The two tags on this switched potentiometer have blackened with age and oxidation. They must be cleaned before they can be soldered...

In the case of old components that have been stored a long time, where the leads have blackened and oxidised, use a small hand-held file or perhaps fine emery cloth to reveal fresh metal underneath. Stripboard and copper printed circuit board will generally oxidise after a few months, especially where it has tarnished due to fingerprints, so the copper strips could be cleaned using e.g. an abrasive rubber block made for the purpose. Also available is a fibre-glass filament brush (photo), which is used propelling-pencil-like to remove contamination. They're OK for general cleaning but best avoided on fine surfaces (e.g. gold plated switch contacts). These produce tiny particles which can irritate the skin, so avoid accidental contact with any debris.

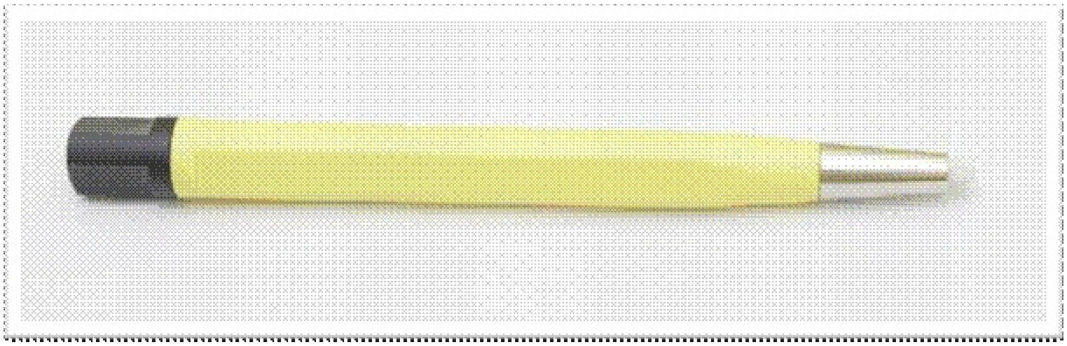

^^ A glass-fibre filament brush like this is useful for cleaning oxidised parts.

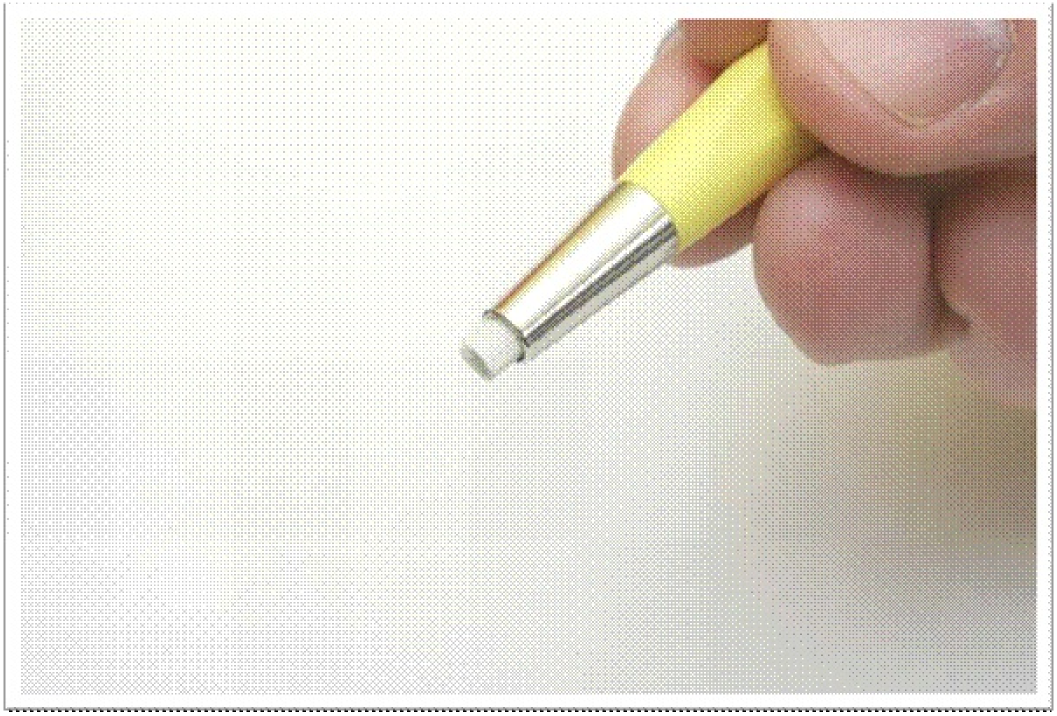

^^ *The glass-fibre brush works like a propelling pencil and produces irritating dust.*

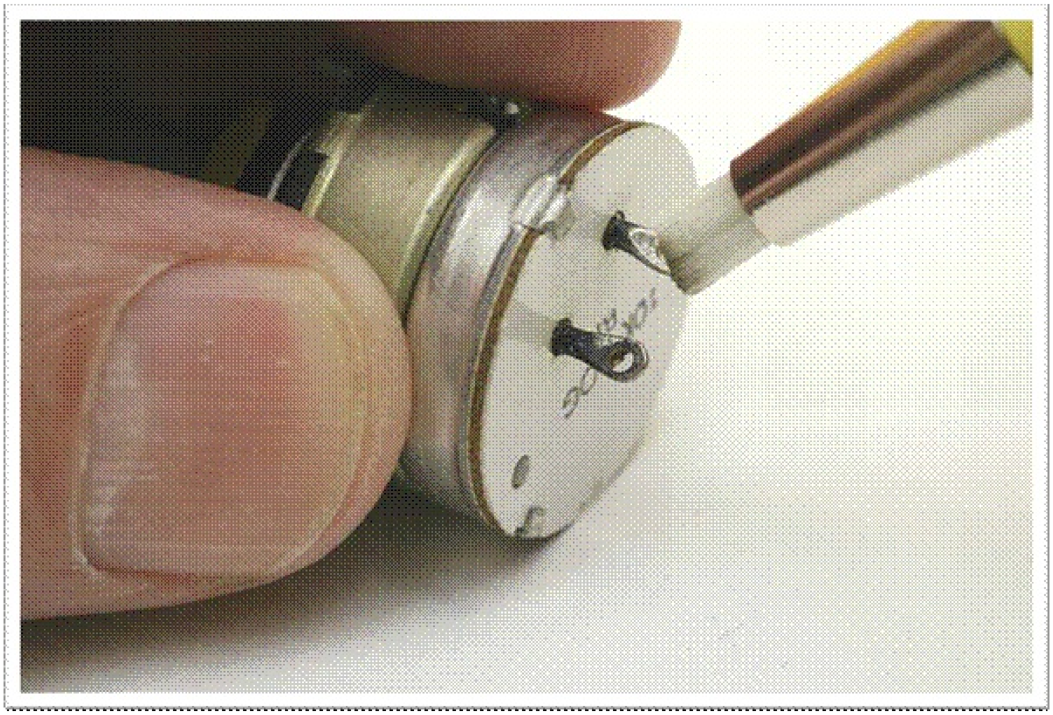

^^ *Cleaning oxidised tags with a glass fibre brush to make them nice and shiny, ready for soldering*

Afterwards, a wipe with a rag soaked in cleaning solvent will remove most grease marks and fingerprints. After preparing the surfaces, avoid touching any parts as far as possible.

Another side effect of trying to solder contaminated or incompatible materials is the tendency to apply more heat and “force the solder to take”. As the materials aren’t wettable the molten solder won’t flow where you want it to flow. This can do more harm than good, because it may be impossible to burn off any contaminants anyway, and the component or the printed circuit board may be overheated and damaged in the process. Semiconductors may be harmed by applying excessive heat for more than a few seconds, and extreme heat applied to printed circuit board tracks can cause irreparable damage, because the tracks will be lifted away from the substrate especially on a delicate or badly designed board. You can avoid trouble by ensuring the surfaces to be soldered are clean and wettable to begin with.

Getting Ready

As already explained, cleanliness in soldering is a major factor in obtaining successful results. A soldering iron stand usually has a sponge that is dampened – it wants to be quite damp but not running wet, so squeeze out any excess.

^^ *Dampen the sponge but don't soak it unduly. Note the hole, which gives an edge to wipe tip with.*

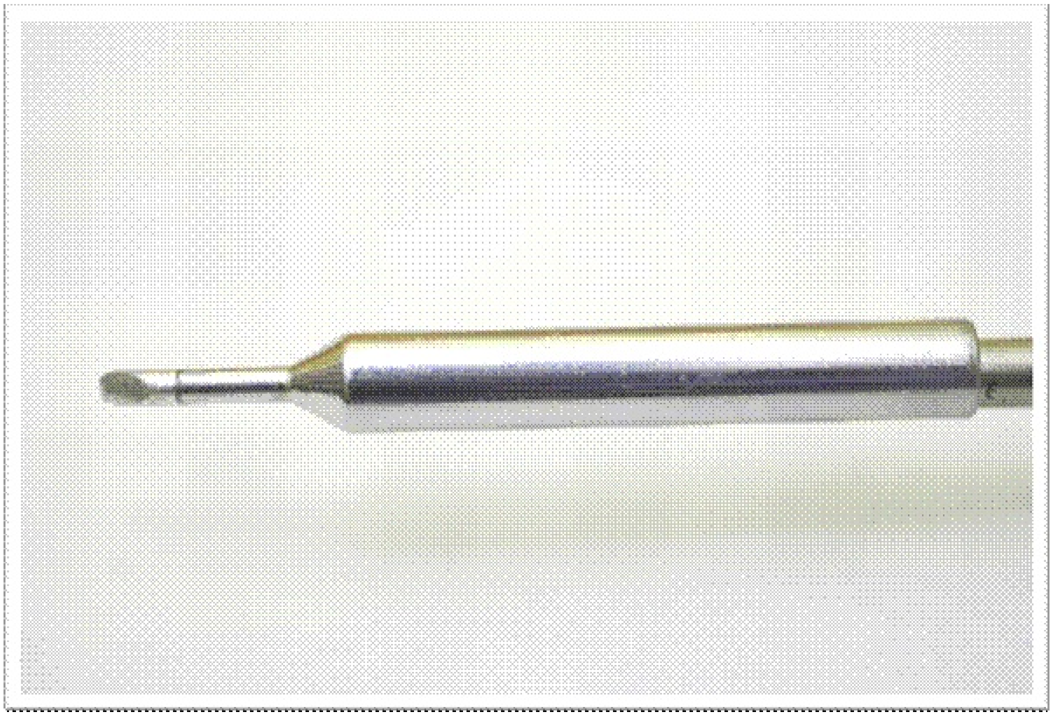

^^ *An Antex soldering iron tip. It simply slides onto the iron's heating element. Only the extreme end is wettable and will accept solder: the tip should be kept nice and shiny.*

Soldering iron bits are typically iron-plated to resist wear, then chrome-plated to prevent molten solder being deposited on them, with the extreme tip being "wettable" to work with molten solder. Before using the iron to make a joint, the hot tip must be "tinned" with a few millimetres of solder*: you should always flood a brand new tip with plenty of solder to tin it immediately, when using it for the first time*.

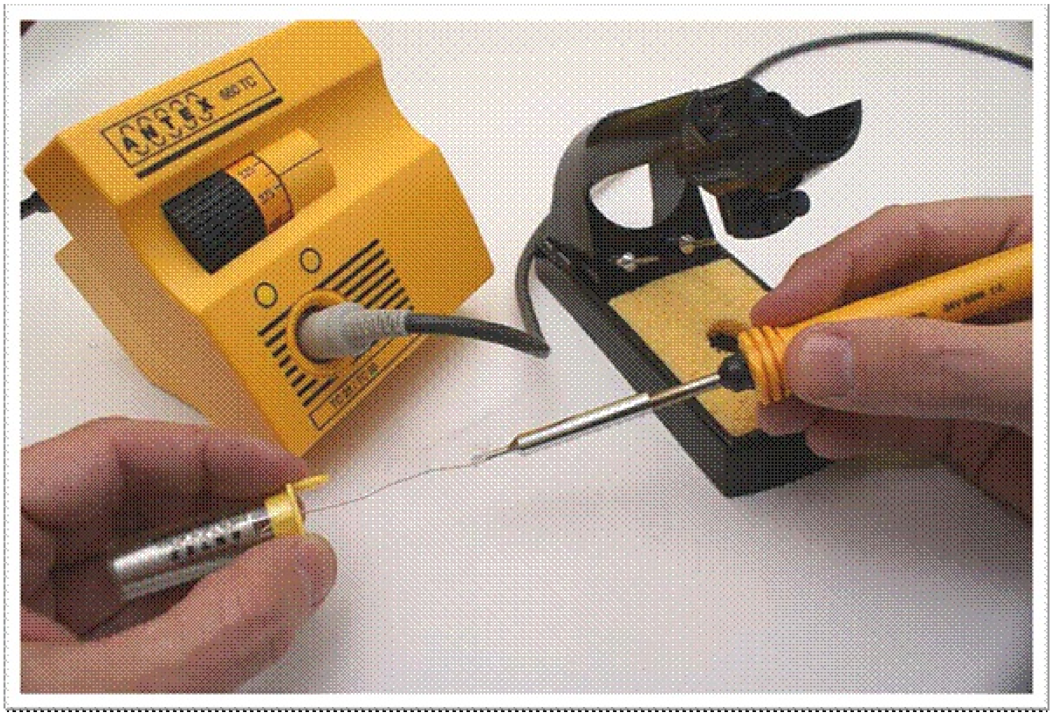

^^ *Tinning the tip ready for use: apply a few millimetres of solder to the hot tip.*

Wipe off excess solder using a damp sponge and it's ready to use. That's why sponges have a hole or well in them – the edge acts as a wiper and the hole catches excess solder.

A useful tip: after tinning the bit, just before using the iron it helps to re-apply a small amount of solder to a clean tip, to improve the thermal contact between the tip and the joint. The molten solder fills the void between the materials and the iron tip, to help transfer heat better so that the solder flows more quickly and easily.

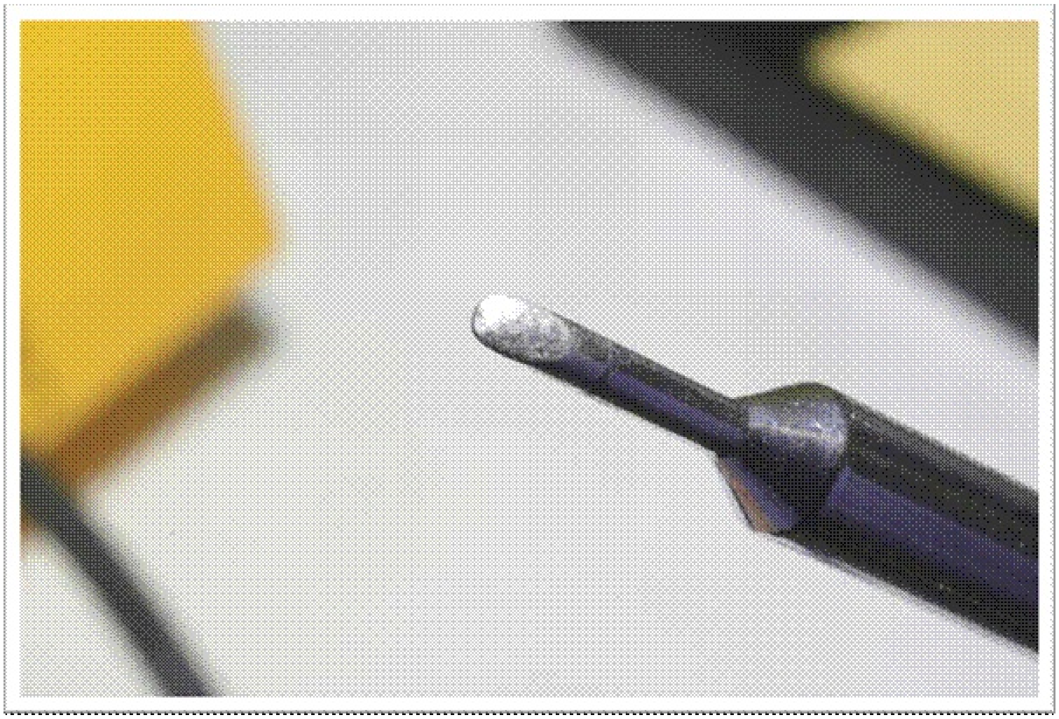

^^ *A tinned tip should look like this, ready to make the next joint.*

It’s sometimes better to tin larger parts as well before making the joint itself, but this is not generally necessary with p.c.b. work. Small tinlets of special paste are available onto which you dab a hot iron - the product cleans and tins the iron ready for use. It’s useful in extreme cases where the bit has some stubborn contamination.

I find special Tip Tinner &Cleaner products to be very useful: gently abrasive in action, they help to clean dirty bits and keep them in good condition. Use them for removing stubborn contaminants, but don’t overdo it.

Multicore
Tip Tinner/Cleaner
TTC1
15g / 0.5oz
Made in Malaysia

^^ *Tip Tinner and Cleaner is a useful aid to maintaining a soldering iron bit ...*

^^ *Just press the hot iron onto the solid paste and scrub it around a little. The tip will be cleaned, tinned, and made ready for use.*

The move to lead-free solders (see next chapter) has had some effect on the life of soldering iron bits, with increased wear and corrosion noted due to the higher temperatures and the fluxes found in tin-based solders. You can therefore expect bits to wear out over time. Once the iron-plating is damaged due to oxidation or erosion, the bit is fast approaching its end of life. Never use an abrasive or file to sand down a tip: the iron-coating will be damaged and the iron's core exposed, so the tip will soon be made useless due to erosion.

Having prepared the soldering iron tip ready for use, in the next chapter solder and fluxes are discussed.

Solder and Fluxes

In recent years there’s been a move towards using more environmentally-friendly materials in electronic products. EU legislation such as *Restriction of the Use of Certain Hazardous Substances in Electrical and Electronic Equipment* (RoHS) aims to reduce toxic heavy metals being sent to landfill. (Look for the RoHS symbol on equipment to indicate compliance.) Due to RoHS compliance, the electronics industry had to change the type of solder it uses in electronic production.

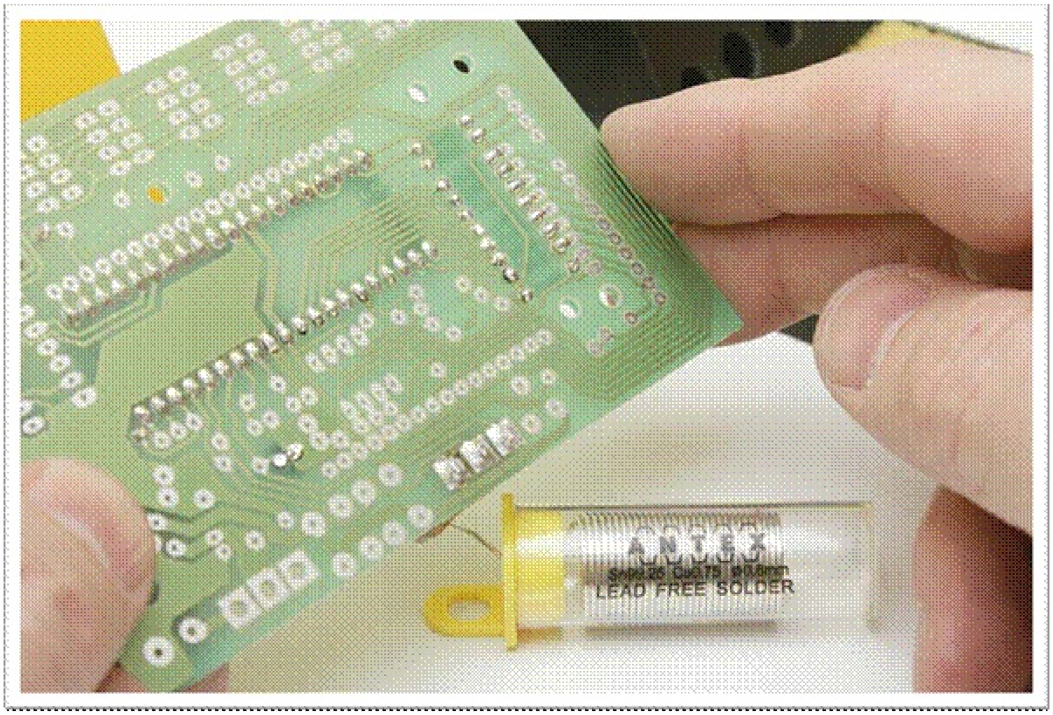

^^ *Electronics solder is supplied in reels or handy dispensers like this.*

Solder comes in various forms including solid bars or pellets for melting in small electric ‘solder pots’ used for treating the ends of wires with solder. Traditional general-purpose electronics-grade solder is in wire form – starting with so-called “60/40” which contains 60% tin (symbol Sn) and 40% lead (symbol Pb) and is sold in handy dispensers or reels. Although tin-lead solder is now banned in industry, there’s nothing to stop the hobbyist from using it but best practice is to use lead-free solder in our work: my advice is to try both, and see which you prefer to work with. “40/60” tin-lead produces lower quality results but is slightly cheaper and perfectly acceptable in hobby circles.

Various diameters of solder wire are marketed. In the UK they’re sold in Standard Width Gauge (SWG) sizes, typically as 18SWG (1.2mm) or 22SWG (0.7mm). The latter is fine for almost all hand-soldering of printed circuit boards or general electronics. For larger solder joints (e.g. larger switch or motor terminals), 18SWG solder would be better as more solder can be dispensed more quickly.

Lead-free solder is universally available and contains typically 99.7% pure tin and 0.3% copper (symbol Cu). It needs a higher melting point which makes it slightly more difficult to work with, but standard soldering irons will cope with it well. Antex lead-free solder (Sn 99.25 / Cu 0.75) is a good compromise at 0.8mm diameter and is sold in small dispensers.

Other solders are produced for specialist work, including aluminium solder (Alu-Sol®) and another solder variant used by professionals is Multicore “Smart” wire which contains a small amount of pure silver (symbol Ag). It produces very clean results and is often associated with SMD (surface mount devices), though some engineers also use it for routine p.c.b. work for producing the best possible finish by hand. As “Smart” wire contains lead it is not RoHS compliant.

An interesting variant is **Eutectic** solder, which is 63/37 Tin/Lead. It goes instantly from solid to liquid when melted and is particularly good for hand-soldering. An almost-equivalent lead-free product would be Stannol Flowtin TC or TSC solder.

The low-down on fluxes

When melting solder with a soldering iron, oxides of metal are produced as a result of the high temperatures involved. Unfortunately, these oxides contaminate the metal surfaces being soldered, which interferes with the flow of molten metal and the production of a good quality solder joint.

All electronics-grade solder wire therefore contains an additive called “flux” which helps the molten solder to flow more easily over the joint. It does this by scrubbing away the oxides which arise naturally during heating, and it will often be seen as a pungent brown fluid bubbling away on the joint, accompanied by some fumes.

Those coming into electronics from other industries should note that flux is already contained within “cored” electronics solder and on no account should any acidic flux be applied separately before using the soldering iron. Plumbers, for example, apply flux paste to copper pipes before soldering them, but electronics-grade solder wire already contains a flux and extra flux is almost never needed. Electronics is no place for acid fluxes!

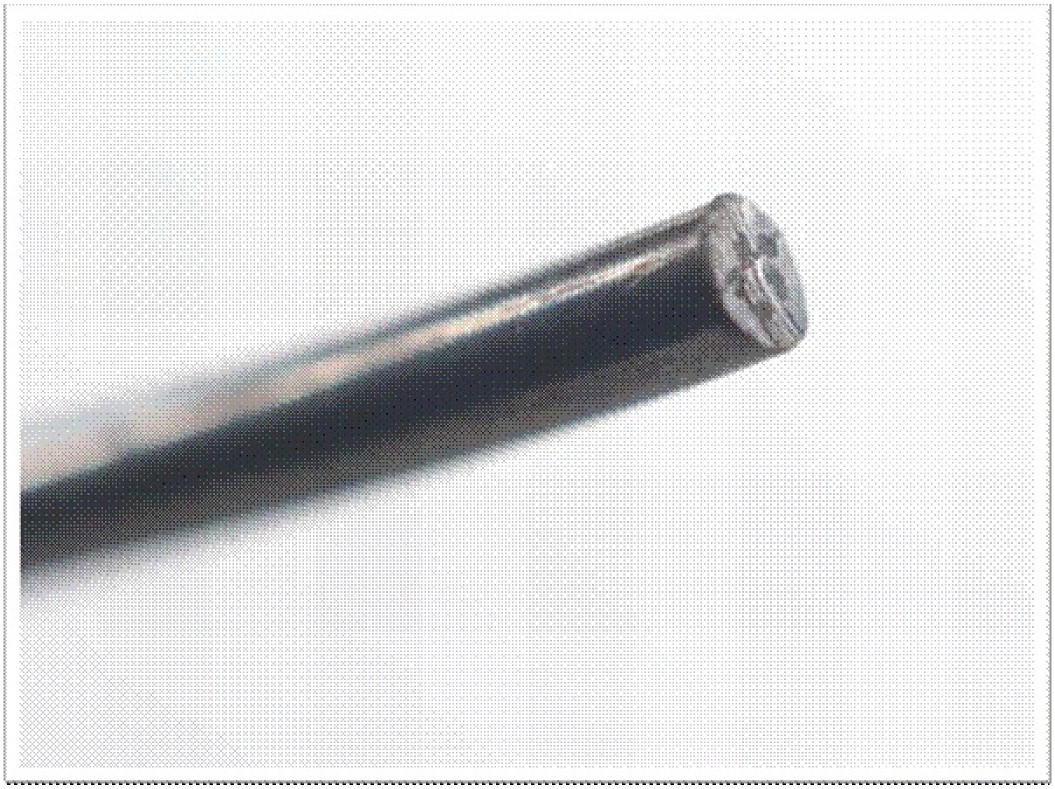

^^ *A close-up of electronics-grade multi-cored solder. Five cores of rosin flux can be seen running through it.*

For almost all electronic hand assembly, solder wire containing “Rosin flux” is used. Cores of flux run through the solder wire like letters running through seaside candy, and they prevent the hot area from being contaminated by oxides, otherwise solder would never flow properly and the result would be an incomplete and unreliable joint.

Flux dispensers and Colophony

Flux dispenser pens are sold that allows special liquid flux to be applied separately onto a work area. These might be handy for difficult or challenging jobs to help solder to flow better: adding more flux this way won't do any harm and may help a solder joint to be made more quickly and reliably. In my hobby electronics, there's hardly ever been a time when I felt the need to apply extra flux but it's useful for some very tricky or demanding jobs.

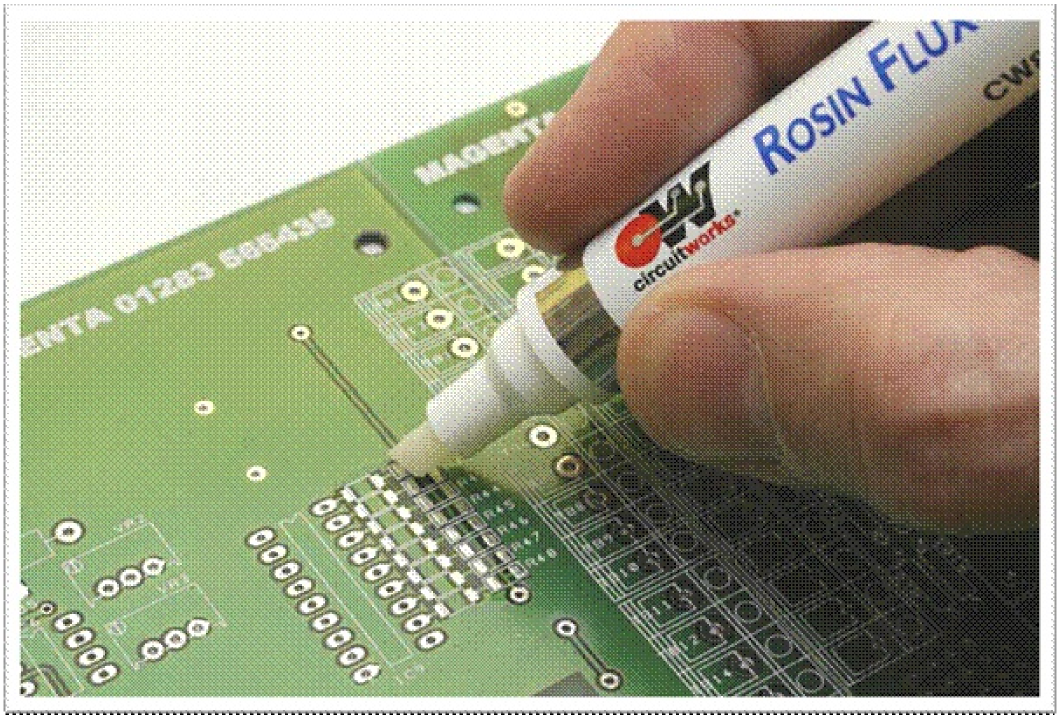

^^ For more demanding work, a flux dispenser pen allows additional flux to be applied

For example I've used specialist Chemtronics flux dispenser pens on tricky, extra-large solder joints involving very thick wires for lead-acid battery connections where I really struggled to make the solder flow properly. You might also use them in micro-electronic surface-mount work. The extra flux can only help, but for the rest of the time rosin-flux core solder wire contains sufficient flux and that's all that you'll need.

^^ *Colophony or rosin flux is available in small tinlets if you want to prepare your own liquid flux.*

One thing that many seasoned electronics enthusiasts will recognise is the distinctive smell of rosin flux: its intense woody pine smell is not unpleasant, and flux fumes themselves are not known to be harmful but solder smoke can be an irritant, especially if you suffer from asthma or other respiratory conditions. (I dealt with solder fume extraction earlier when soldering irons were outlined.)

Rosin flux is also known as **Colophony.** It's an amber resin that's glassy and brittle like sugar candy, distilled from the resins of conifers (mainly pine trees) and it's worth knowing about. For electronics use, Colophony is available in individual tinlets of solid resin (e.g. *Kolophonium* in 20g tins by Donau Elektronik GmbH, sold by Westfalia or Conrad).

As an aside, the resins of various trees are used in incense burners: placing a few fragments on a charcoal burner produces intense delicious fragrances of pine, frankincense or forest but you still wouldn't want to inhale the smoke or fumes directly.

Solids of Colophony can be dissolved to make a semi-liquid flux, for dipping or applying manually prior to soldering. To make your own rosin flux from solids of Colophony, chip off some fragments and crush them into a small tinfoil dish, then apply some isoproponal alcohol (start with 2 parts solvent to 1 part Colophony) and let it dissolve over about 20-30 minutes or more.

^^ *Chip off some solids of colophony and put them in a small tin foil dish*

^^ *Add e.g. x 2 volume isopropanol alcohol and allow the solids to dissolve. Experiment as needed.*

^^ The resulting flux can be very sticky so handle with care and do not spill.

In this form it can be applied directly with a brush or dropper to areas prior to soldering. In an open dish, after a day or two the solvent will mostly have evaporated to leave an extremely sticky resin that should be handled carefully: it's no co-incidence that Colophony is also used in adhesives and varnishes, and different grades of Colophony resins are used to treat violin or double-bass bows to add friction to bow hairs! Try experimenting with different ratios of solvent and making some up and storing it in an old nail varnish bottle with brush, or generally do what works for you.

Unused Colophony crystals can be reformed into their storage tinlet by warming carefully with a hot air gun, but don't overdo it. You can also use flux cleaners or PCB solvent cleaners if you want to remove traces of excess rosin flux after soldering: these are left behind as a brown deposit and are otherwise harmless.

The subject of using your soldering iron to raise the temperature of the materials is discussed next.

Temperature flow

This aspect takes a bit more understanding, but with practice you'll soon understand how temperature flows within a workpiece being soldered and how you have to harness it properly. After ensuring all parts are clean and the soldering iron is ready to go, the next step to successful soldering requires that the **temperature** of **all** the parts is raised to roughly the same level, before solder can be applied.

Imagine the most basic task of soldering a simple resistor onto a printed circuit board: the copper p.c.b. and the resistor lead should both be heated *together* so that the solder will flow readily over the joint. Later I'll show the precise stages step by step.

A beginner will often mistakenly just heat one part of the joint (e.g. a resistor's wire protruding through a printed circuit board) and hope that the resultant "blob" of solder will be enough to tack everything together. That's completely wrong, because the remaining metal in the joint is cold when molten solder is flooded on to it. The joint will be weak, incomplete or unreliable. Flux will not have flowed properly either, so the joint could be contaminated internally.

Another beginner's mistake is to use a soldering iron to carry blobs of molten solder over to the joint, as if to daub solder over it. The secret of success is to control the iron accurately and apply the hot tip onto the workpiece so that it's in contact with *all the parts*. Within a fraction of a second, heat will conduct from the iron and raise the temperature of the entire joint, after which solder can be melted over it. Remove the iron and let the joint cool naturally. It takes some practice with your chosen soldering iron tip to obtain the best heating action, making sure the tip is clean and tinned properly to begin with.

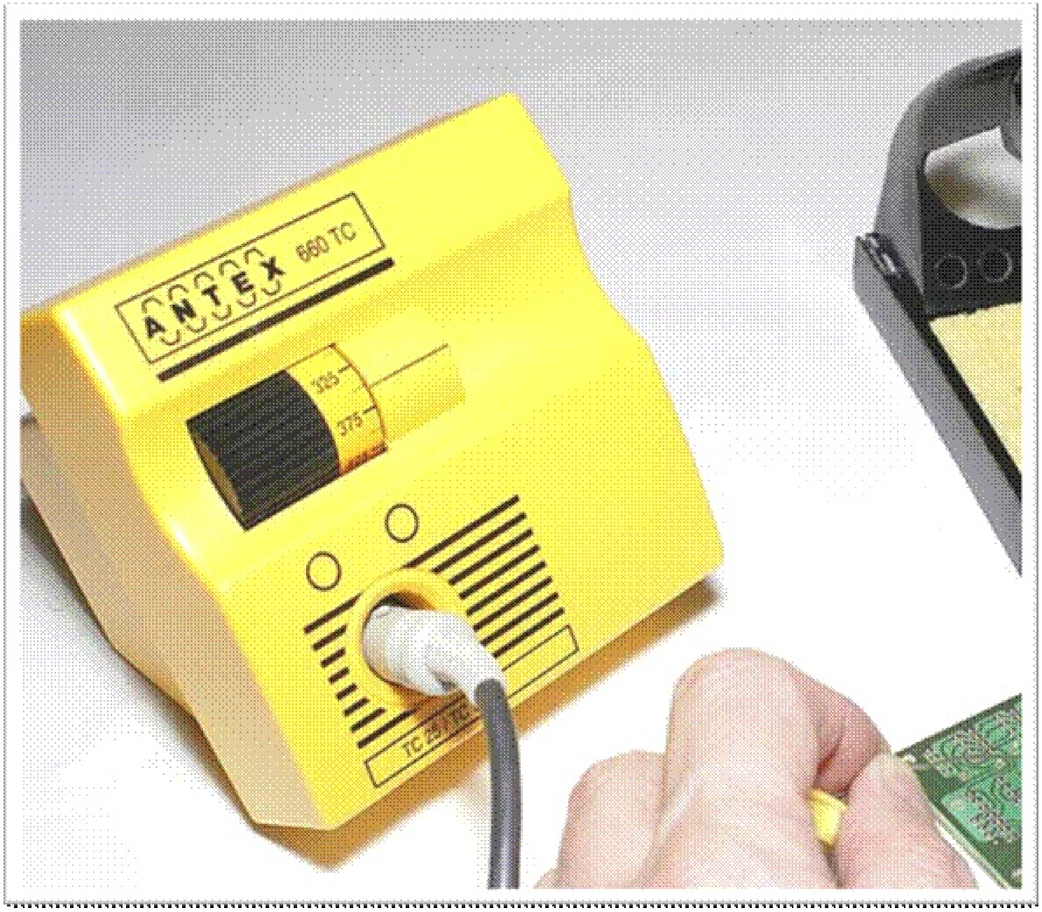

^^ *Set a typical temperature of 330-350 ºC (626-662 ºF) or more, if you have a soldering station.*

The melting point of traditional tin-lead solders is about 190ºC (374ºF) but lead-free tin-based solders require higher temperatures, having melting points of typically 201-227 ºC. As Antex reminds us, the melting point is *not* the temperature of the soldering iron tip: instead you should set a temperature that ensures the solder melts *instantly* onto the tip. Fixed-temperature soldering irons have no adjustability but they'll cope just fine with either type of solder. A soldering station usually has a variable control that gives more control in different circumstances. In practice the iron tip temperature should be set for typically 330-350 ºC (626-662 ºF) or maybe a little more if using lead-free solder.

Now is the time...

The next diagram shows what would happen if you applied a hot soldering iron to an imaginary metal block. In step (1) heat travels out of the iron into the cold metal block, which starts to warm up starting at the edges. Gradually (2), the whole metal block's temperature rises until the middle of the block finally rises in temperature as well. In effect heat is now travelling back "towards" the iron tip, until finally (3) the whole block is at the same temperature as the iron. At this point solder could now be applied.

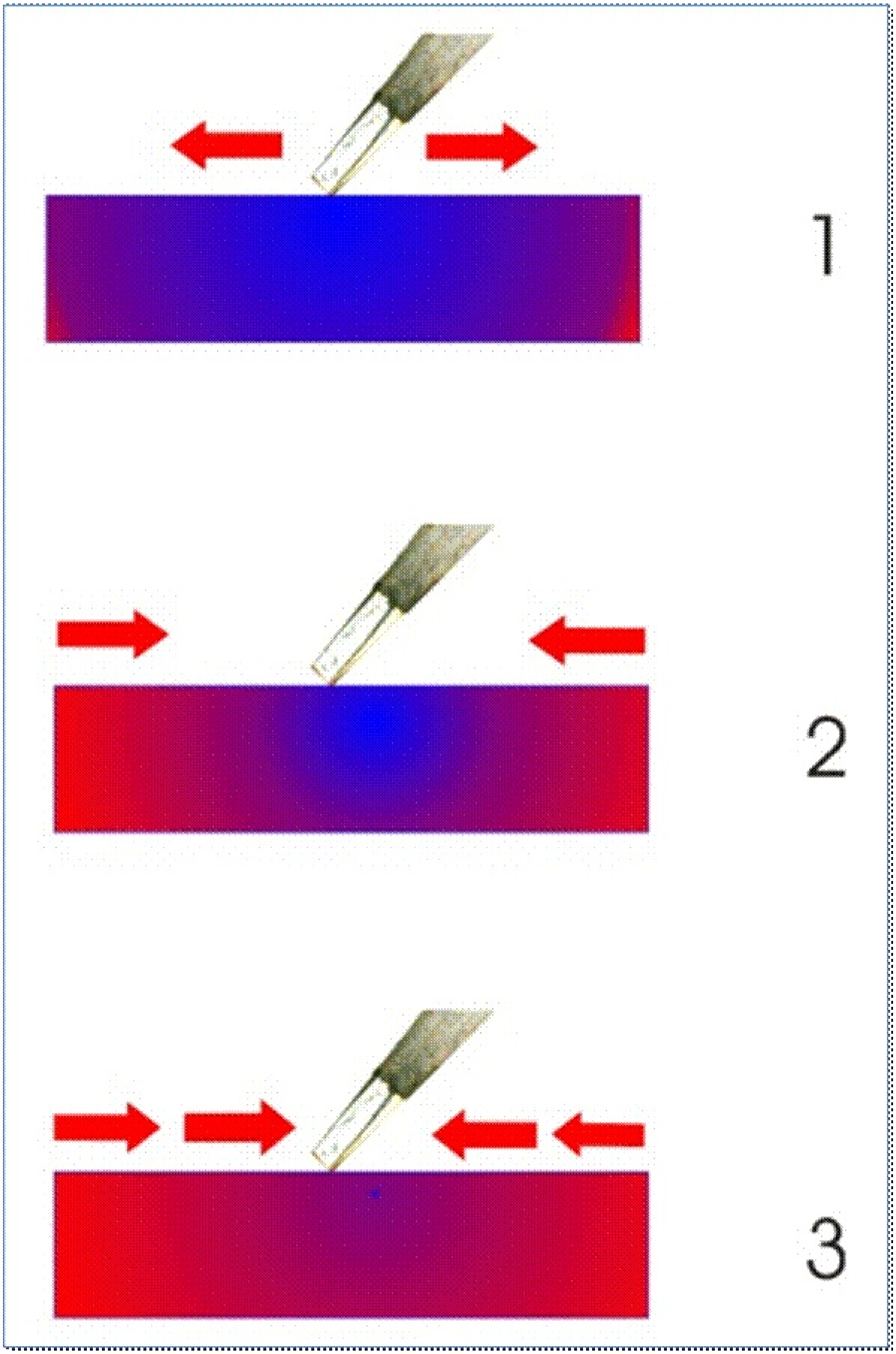

^^ How metal in the joint actually "sinks" heat away from the tip to begin with. Then heat moves back towards the tip until the solder's melting point is finally reached.

You’ll notice this effect as a time delay when soldering any joint. The “working area” of the iron’s tip gets **cooled** to begin with, because the metal in the rest of the joint is sinking heat away from where it’s most wanted. Only after the whole joint has risen in temperature can solder be melted onto it.

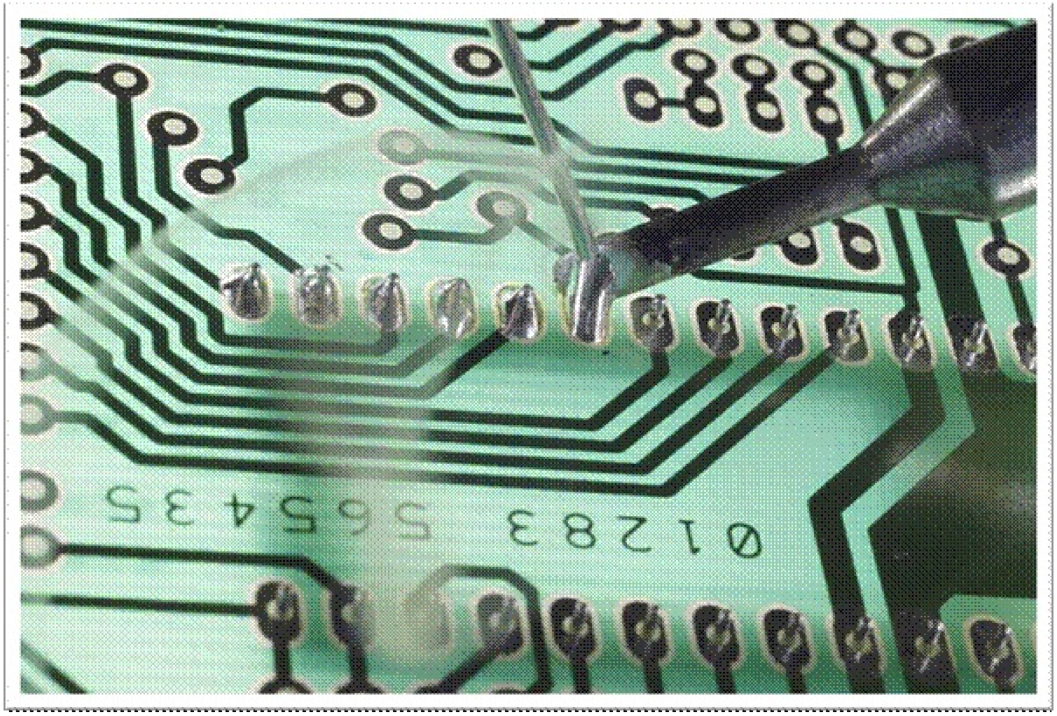

^^ *Soldering a small p.c.b. joint*

With experience, you’ll get a feel for how long it takes before you can apply solder. The more metal that is present in the joint, the longer period that heat must be applied for. A small p.c.b. joint takes well under one second to complete. A large metal terminal could take quite a few seconds or more to heat up. As I explained earlier, higher power (wattage) irons cope better with larger workpieces because they recover more quickly and are more “unstoppable”, making it easier for them to heat larger workpieces without cooling down so much.

If you apply solder too early, it won’t melt properly and the result will be a grey, crystalline joint caused by the solder’s melting point temperature not being reached and the flux not having flowed properly. Semiconductors must be soldered as rapidly as possible as they are heat-sensitive, but they’re a lot more robust than they used to be.

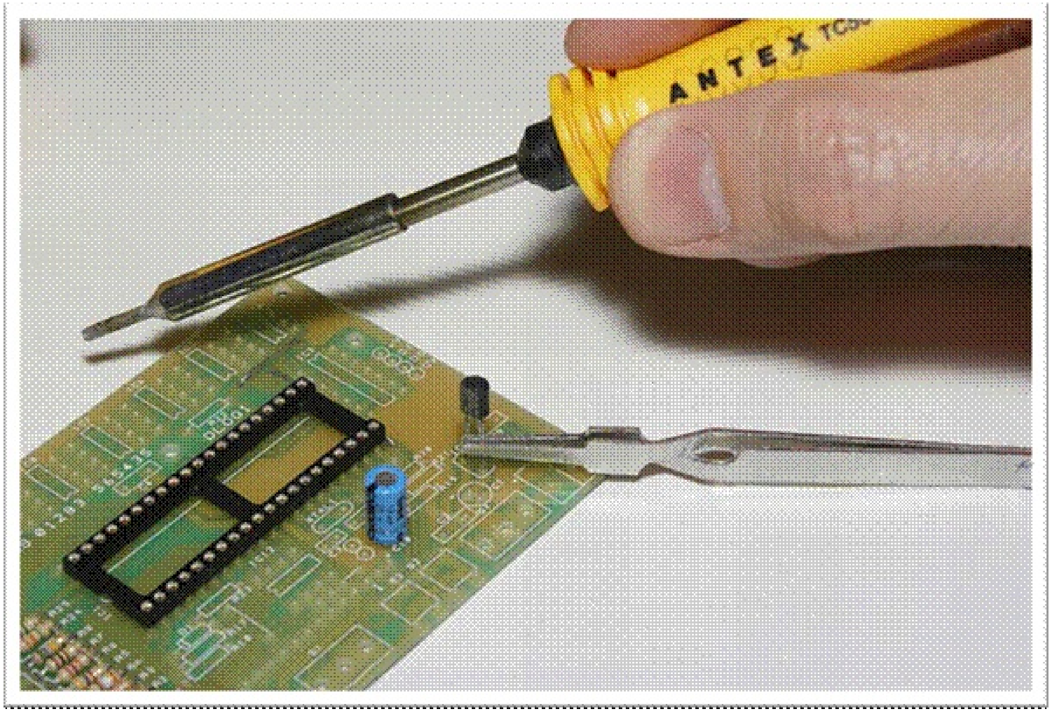

^^ *A clip-on Antex heat shunt fitted to a transistor leg, helps prevent thermal damage due to overheating when soldering it in place. Less essential these days, but beginners find them re-assuring.*

Until they have gained some practice, novices sometimes buy a small clip-on heat-shunt (photo) which resembles a pair of aluminium tweezers. In the example of, say, a transistor, the shunt is attached to one of the leads near to the transistor's body. Any excess heat then diverts up the heat shunt instead of into the transistor junction, thereby avoiding the risk of thermal damage. Applying far too much heat may destroy the part or damage the p.c.b. foil which may lift away from the board.

In due course constructors learns to judge how much solder should be applied to any particular joint. An ideal p.c.b. joint is very slightly concave in shape. If not enough solder is used, the result may be an incomplete joint which may cause an intermittent fault later on. An excess of solder – shaped like a ball bearing - is an unnecessary waste and in extreme cases may cause short circuits, especially on densely-populated boards. There is no need to add more solder “for luck”. Professionally-produced p.c.b.s have a green *solder resist* coating which helps to ensure that solder does not stray onto adjacent pads. As a finishing touch, I usually spray the solder-side of a circuitboard with aerosol spray lacquer afterwards. It keeps the solder joints nice and shiny and helps prevent corrosion.

Some components can create hazards during a soldering operation:

Coin cells and button batteries are commonly used as power, clock or memory backups. If heated excessively they can explode without warning due to the build-up of internal pressure. Spot-welders are used in industry to connect tags to them, but if you need to solder wires to such a cell then it should be done as quickly as possible.

Some **memory back-up capacitors** or **electrolytic capacitors** remain energised for a while even when the circuit is powered down. Molten solder is a perfect electrical conductor and in some cases the component's contacts could be shorted during the soldering (or desoldering) operation. If molten solder shorts it out then the arcing may cause globules of solder to be spattered outwards without warning, potentially risking eyesight damage.

Always ensure that powered components are electrically inert and discharged before soldering. Cells, batteries and battery packs should not be accidentally shorted during the soldering process, to avoid arcing and solder spatter. Note that electrolytic capacitors can also explode after a while if reverse-connected, so observe polarity closely.

Let's now consider the practical stages of soldering components and wires successfully.

Soldering Step by Step

Earlier I explained the individual factors that affect the quality of a solder joint. These are:

- **Cleanliness** – dirt or impurities drastically hinder good solder coverage.
- **Temperature** – the right level to enable the solder to flow freely!
- **Time** – apply heat for just the right amount of time!
- **Adequate solder coverage** – enough to form a good joint without touching neighbouring areas.

These rules apply whether soldering a p.c.b. or performing other tasks such as interwiring (hooking everything together with connecting wire).

We'll now summarise the stages of making a typical solder joint – soldering components onto a printed circuit board (through-hole soldering). Most people insert components into the circuit board and simply splay the wires out to hold them in place under spring tension. I find it best to snip excess wire leads off at this stage, to improve accessibility.

- All parts must be bright, clean and free from dirt and grease.
- Try to secure the work firmly to stop parts moving around.
- "Tin" the soldering iron tip with a small amount of solder. Do this immediately, with new tips being used for the first time.
- Wipe the tip of the hot soldering iron on a damp cellulose sponge to remove excess solder or contamination.
- Many people then add a tiny amount of fresh solder to the cleansed tip just before using it.
- Heat all parts of the joint with the iron typically for under a second or so, until it's heated throughout.
- While heating, then apply sufficient rosin-core solder to form an adequately-covered joint.
- It only takes a second or two at most, to solder the average p.c.b. joint this way.
- Do not move parts until the solder has cooled.
- Remove and return the iron safely to its stand.

This special photo sequence illustrates these stages. It's best to start with the smallest, fiddliest parts first when soldering a blank p.c.b., because that's when you've got the most access on the board. Accessibility will be reduced as more components are added, so we'll start with a simple wire link on a professionally designed p.c.b.

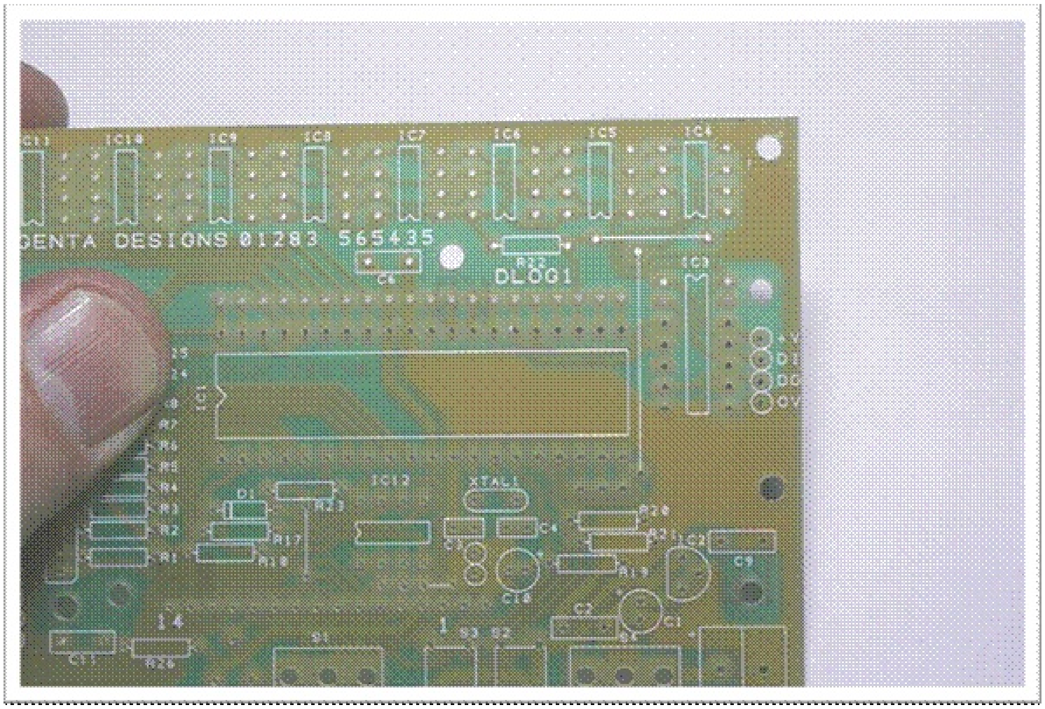

^^ A typical professional blank p.c.b. – silk-screen printing shows what goes where. The underside has been treated with a green solder resist coating, and the solder pads are ready-tinned to help with soldering.

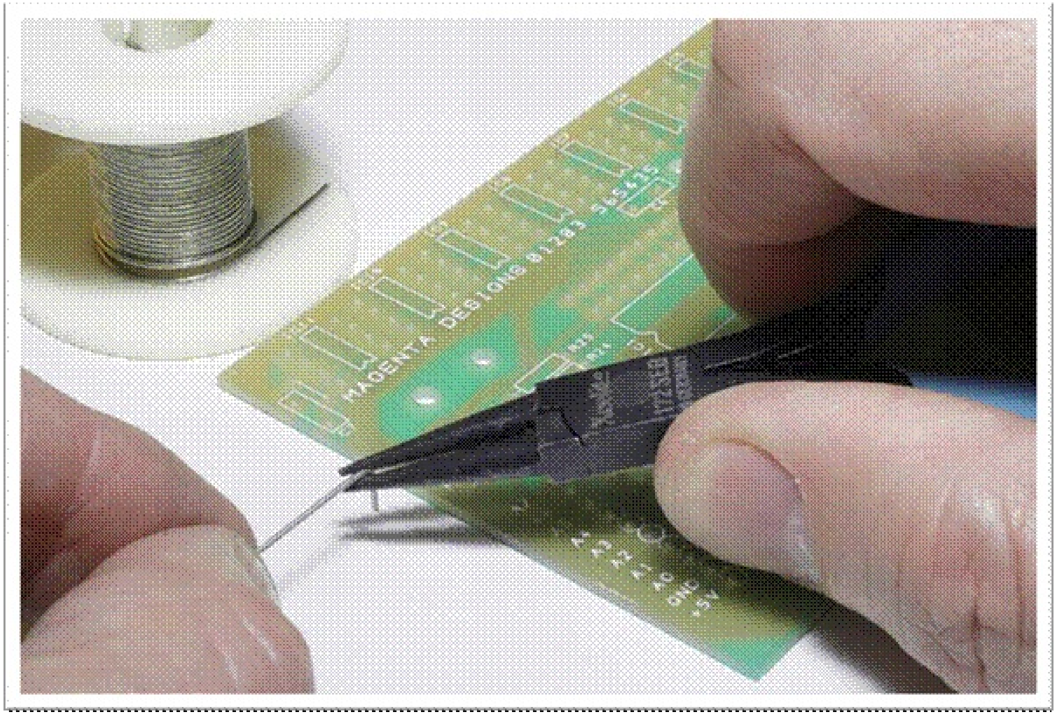

^^ Preparing a wire link for soldering – cut off some tinned copper wire and bend the ends to fit the p.c.b. correctly. Round-nose pliers (shown) are perfect for this, but

ordinary electronics or "radio" pliers will do.

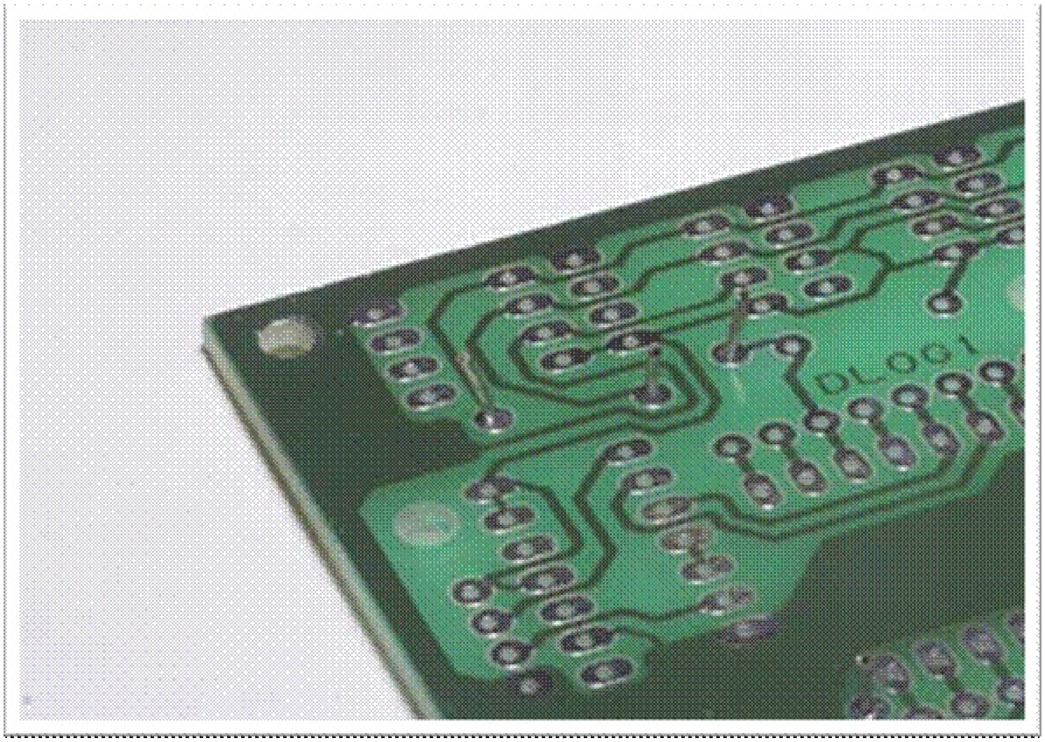

^^ Wires are prodded through the holes in the board, then turn it upside down to view the solder side. You can then "spring" or splay the ends apart slightly, so they are held in place while you solder them.

ANTEX
660 TC

^^ Wipe the hot soldering iron on the damp sponge to clean the tip. Do this periodically when contamination, flux deposits etc. build up on the iron to keep the tip nice and shiny. Tip Tinner & Cleaner helps too.

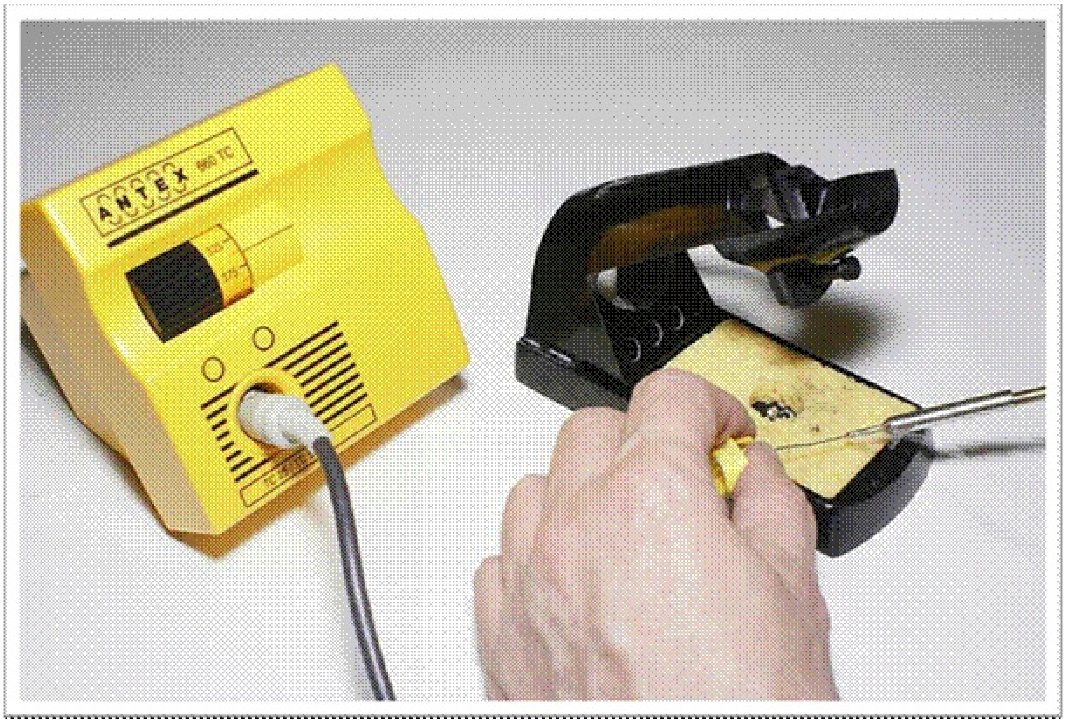

^^ It often helps to dab a tiny amount of solder wire onto the tip, to improve heat transfer.

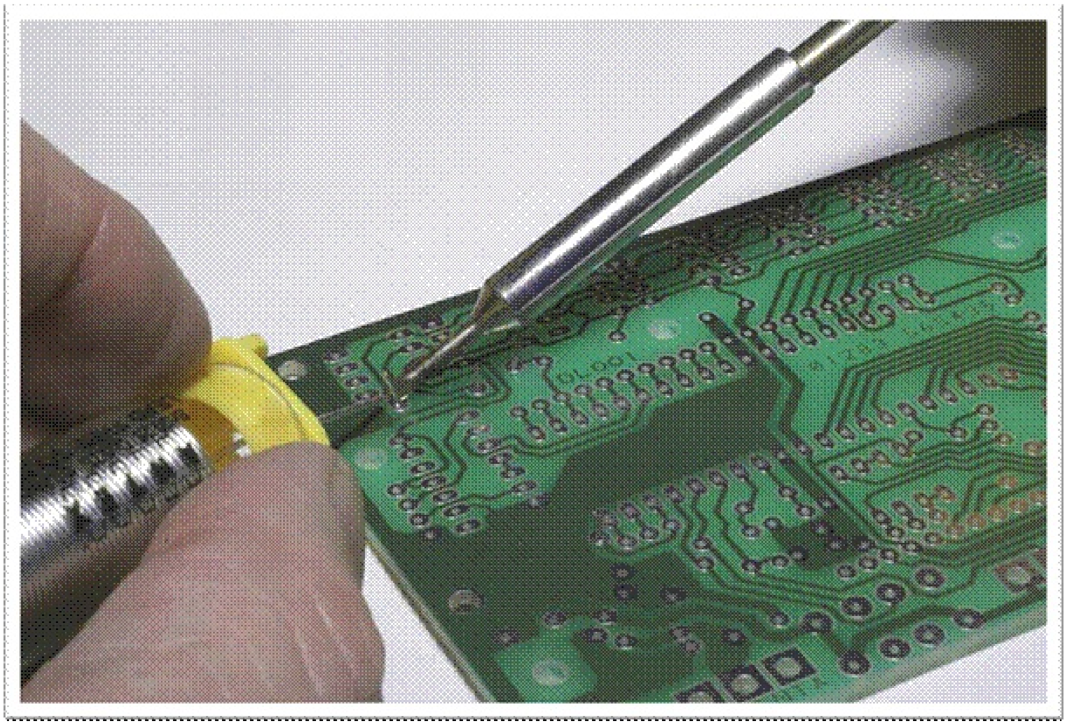

^^ Then apply the soldering iron to heat both the solder "pad" and the wire end at the same time (say <1 second). Apply a few millimetres only, of solder. Then remove the soldering iron immediately and allow the joint to cool down by itself. The green solder-resist coating ensures solder doesn't flow onto neighbouring pads.

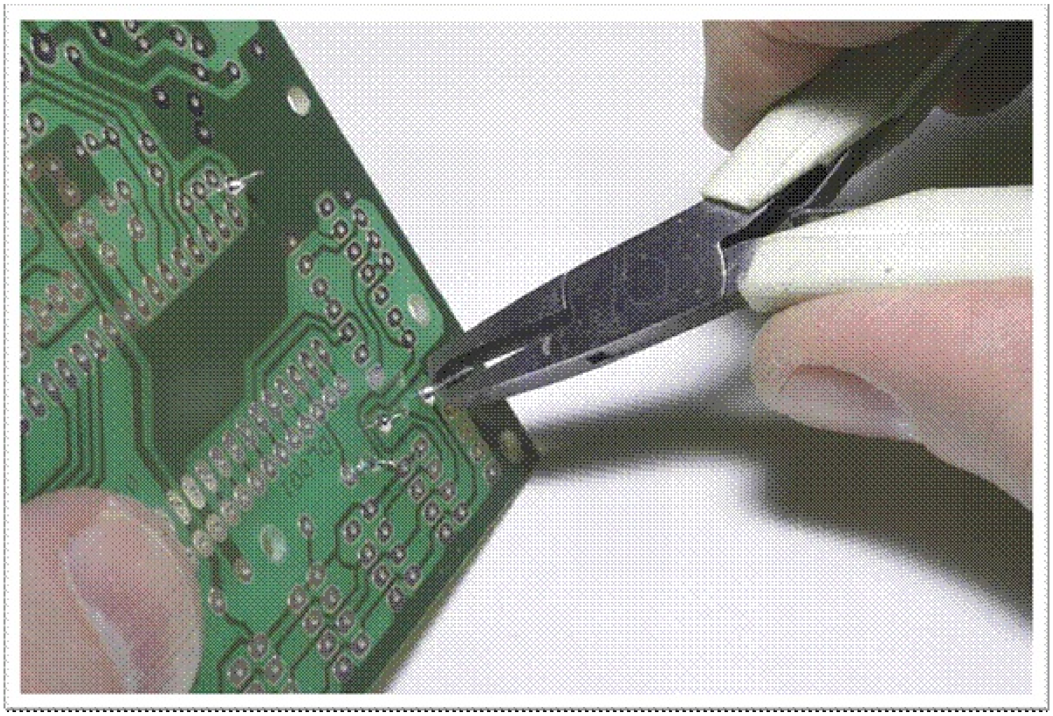

^^ These "end cutters" can then be used to snip off excess wire after soldering....

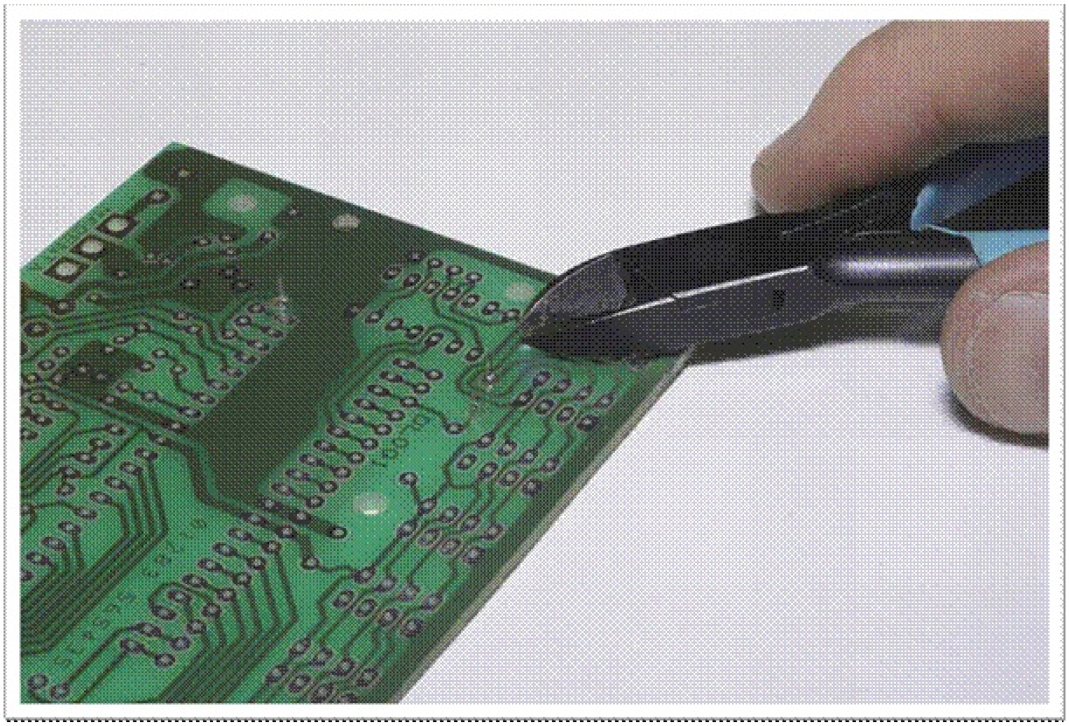

^^ ... but ordinary electronics "side cutters" are fine for trimming or snipping wires.

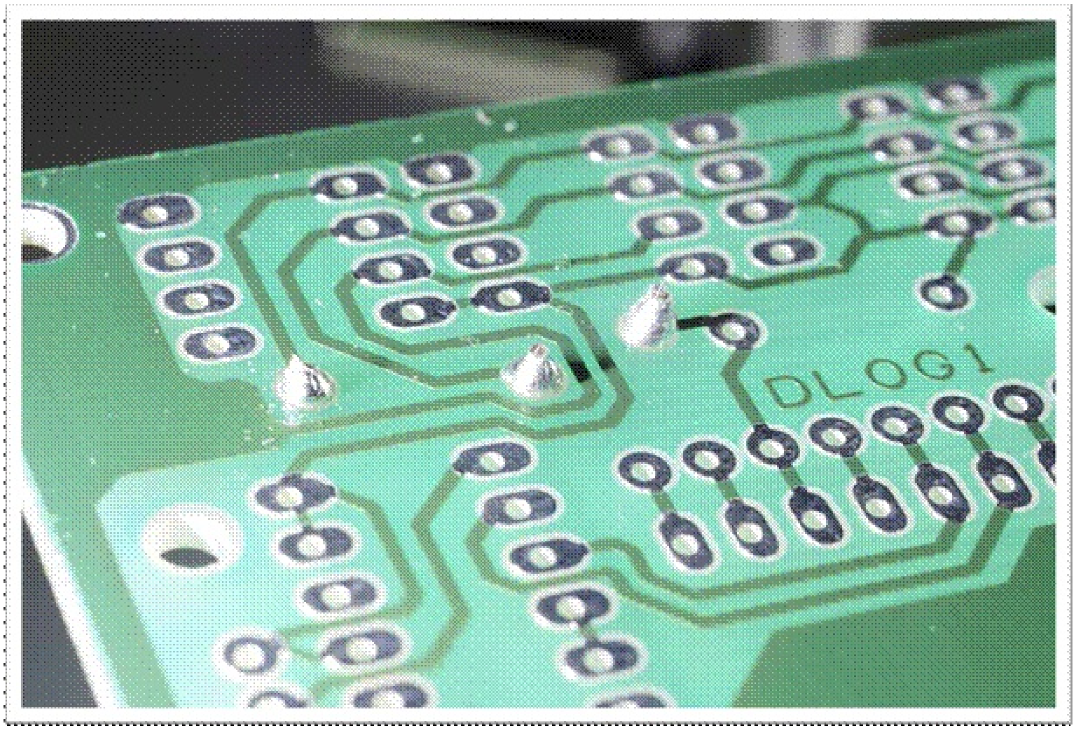

^^ The ideal solder joint should be smooth and slightly concave, quite shiny, not dull or crystalline-looking nor ball-bearing shaped.

Try to get consistent results in your soldering, but don't worry too much about the lack of uniformity – this is soldering by hand after all, not by precision machinery, and some slight variation is OK provided the solder coverage is good and the soldering is generally clean and effective. You'll improve with practice, that's for sure!

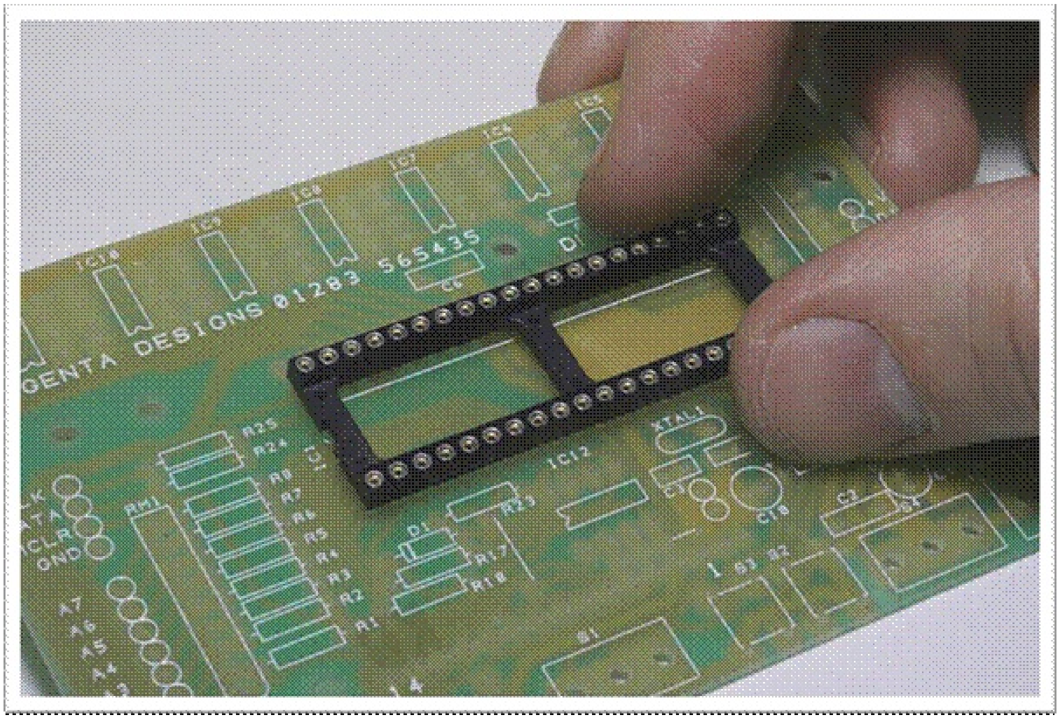

^^ A large i.c. socket can be added next...

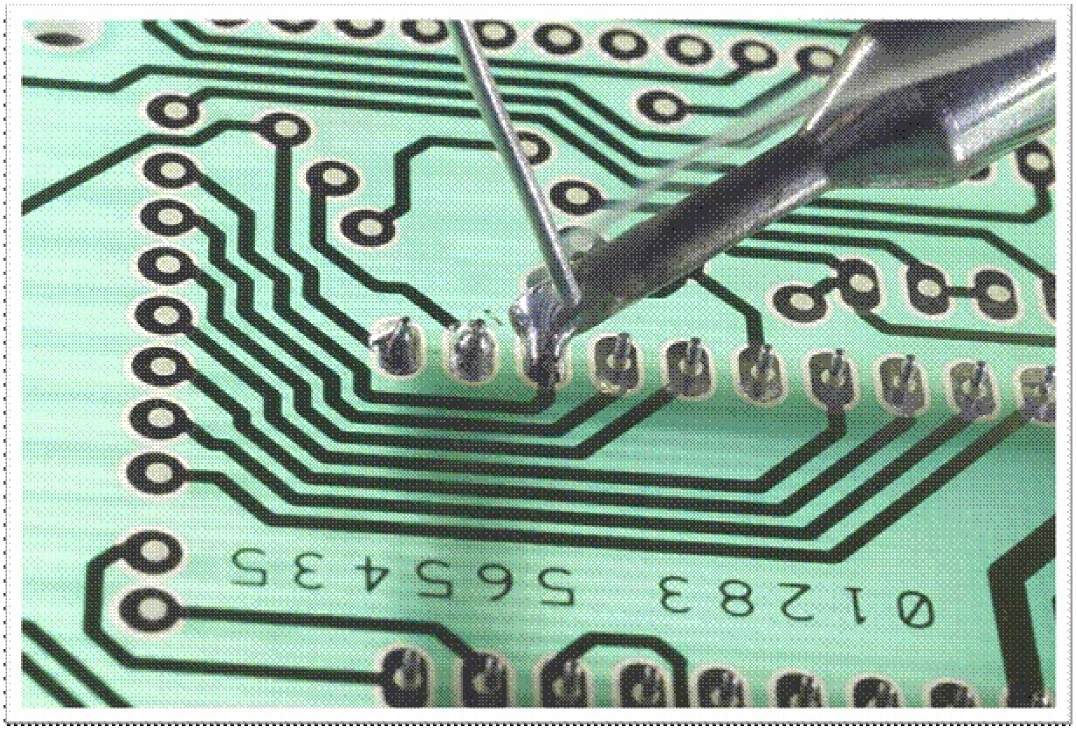
565435 565435 01283

^^ Go along the rows of pins and solder them one at a time. See how simple it is! Heat both the solder pad and pin with the iron, and dab on a little solder wire to ensure full coverage. Ensure the socket is flush against the board.

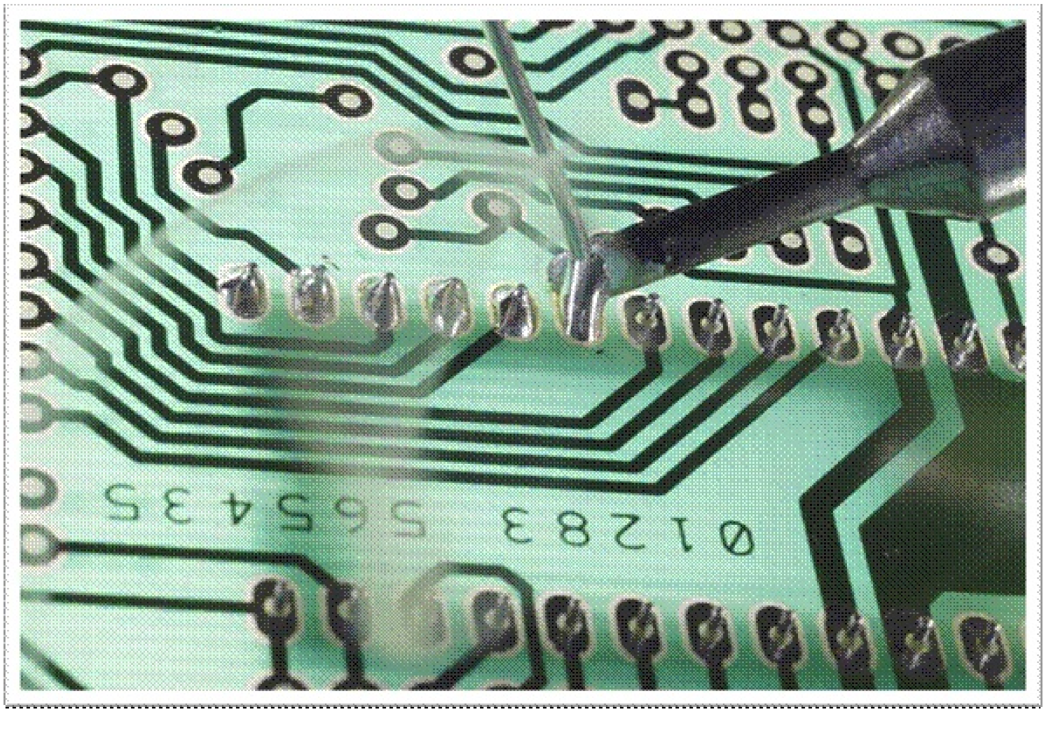
565435 565435 01283

^^ As hardly any metal is involved each pin should take well under a second to solder at most.

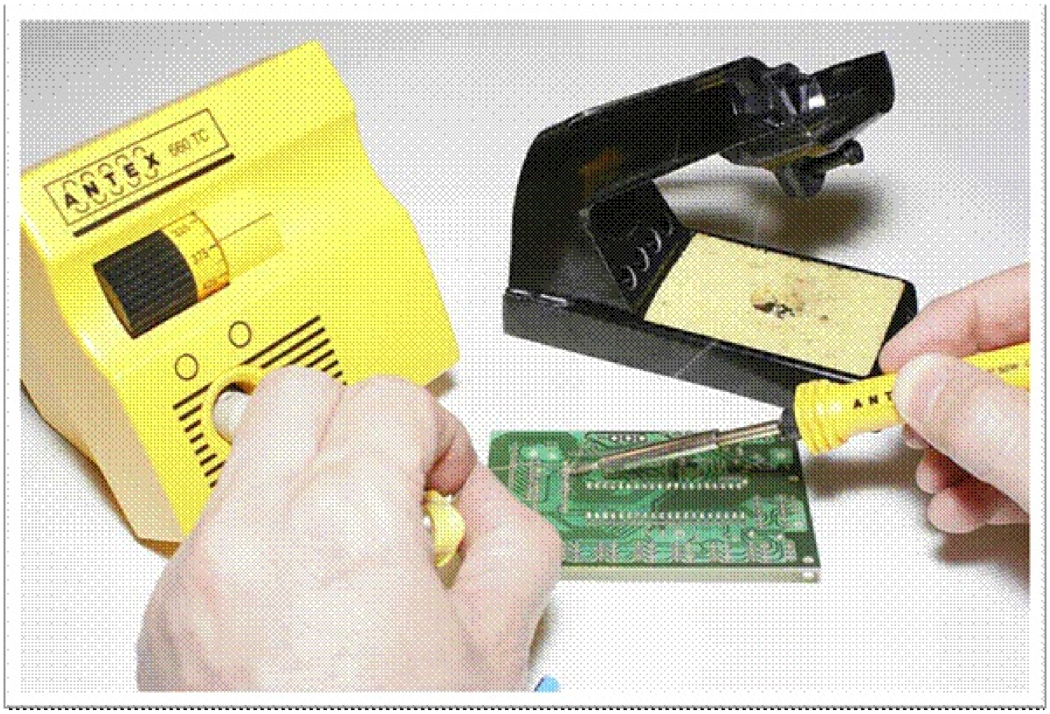

^^ *Continue until the i.c. socket is soldered.*

Once the smallest parts are soldered into place, you can continue to solder the remaining components. It’s easiest to handle the smallest so-called “discrete” parts first while you still have plenty of room on the board. I usually solder resistors and capacitors next. The principle of soldering them is just the same as a simple wire link: insert them from above till they are flush on the board, then splay their wires a little to hold them in place, and preferably snip off at least some of the wire to give you more access.

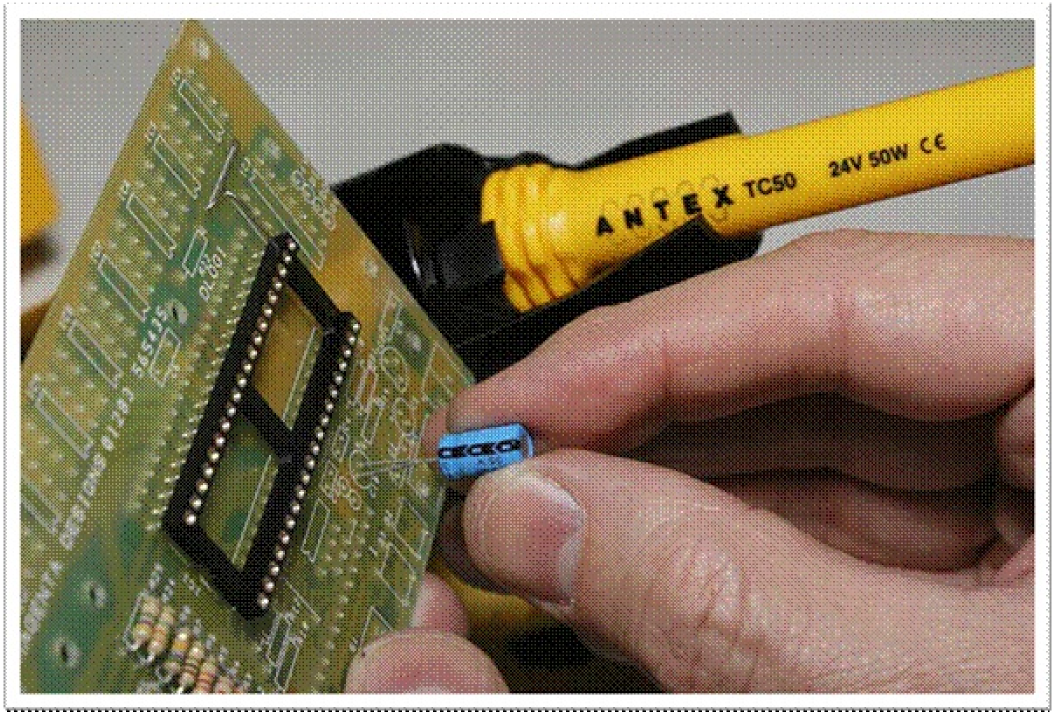

^^ Continue by inserting "discrete" components from above, splay their wires out underneath to secure them in place, then snip and solder the joints exactly as before. Some parts like this blue electrolytic capacitor are polarity sensitive (note the – sign), and must be inserted the right way round. Same is true of every semiconductor.

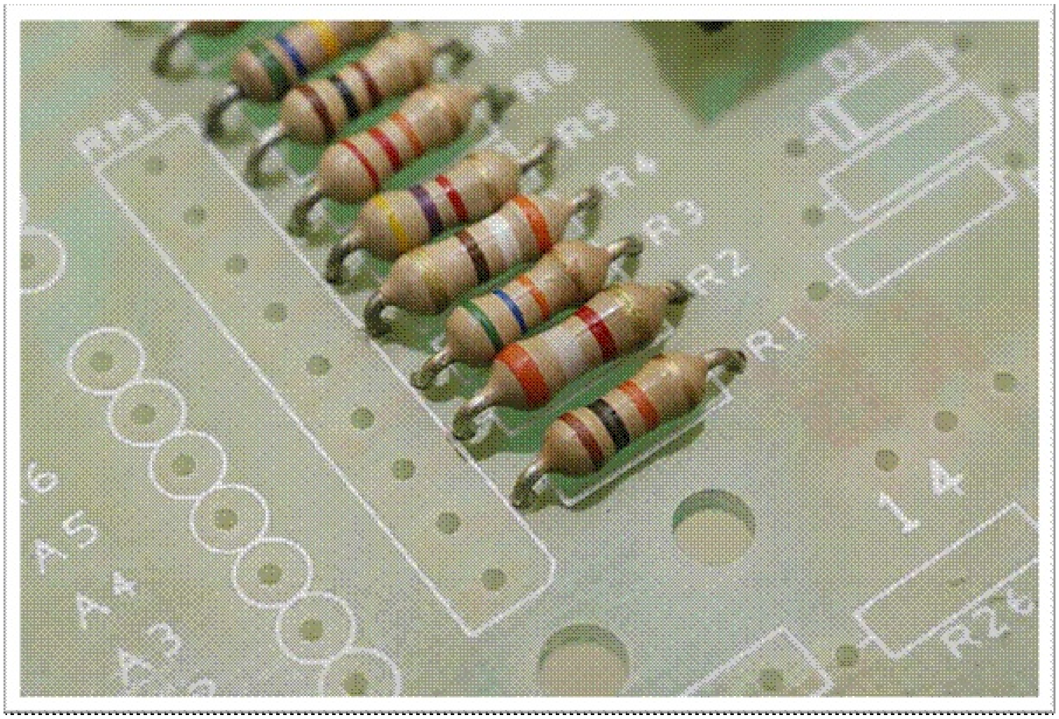

^^ This row of ¼ watt resistors was next. I like to show the colour codes all the *same way round, with gold or silver (tolerance bands) on the right for consistency.*

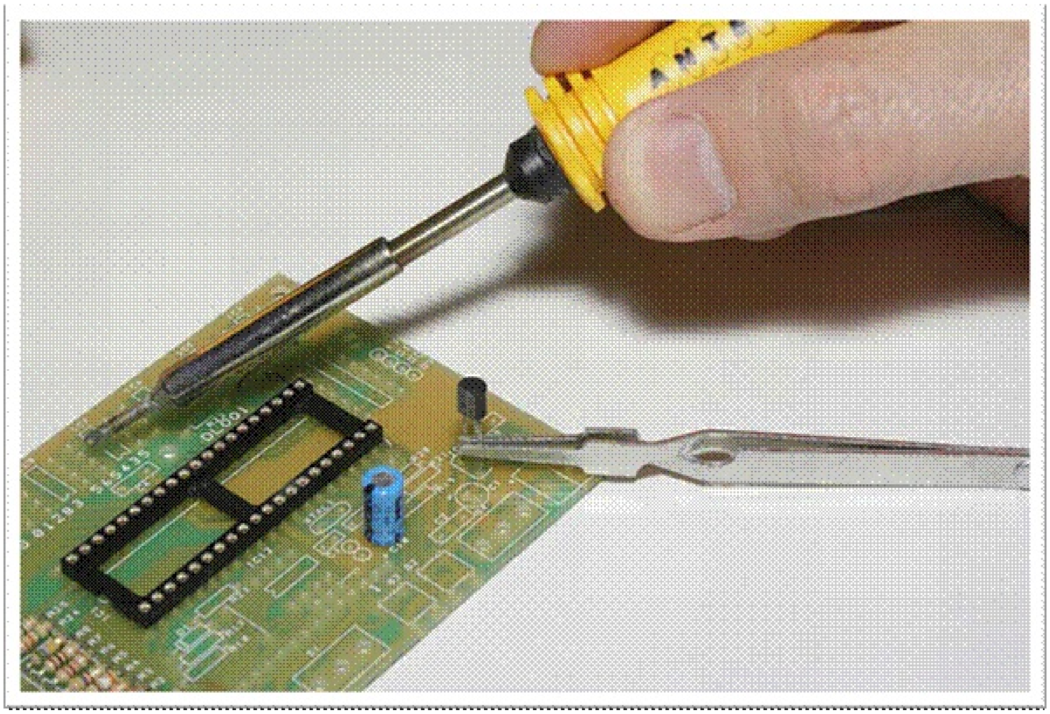

^^ This transistor was next. Solder it quickly to avoid damage, and observe polarity correctly, so it is the right way round on the p.c.b. The heatshunt is optional.

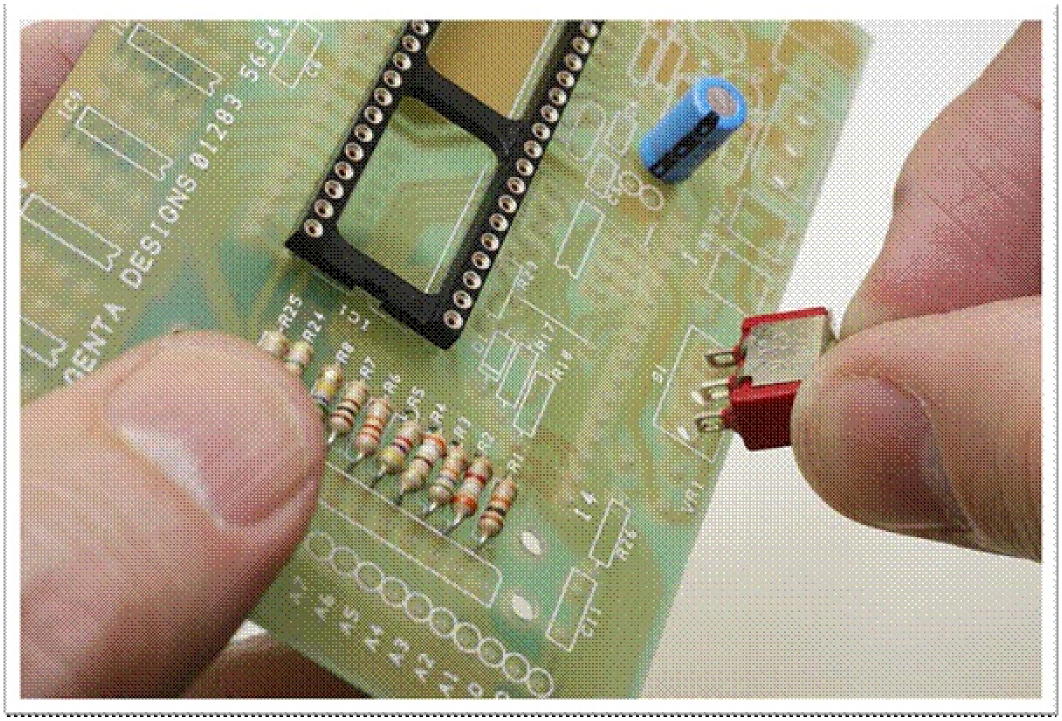
GENTA DESIGNS 01283 5654

^^ *Unusually, this toggle switch fitted directly onto the board as well which saved wiring it up. Neat idea!*

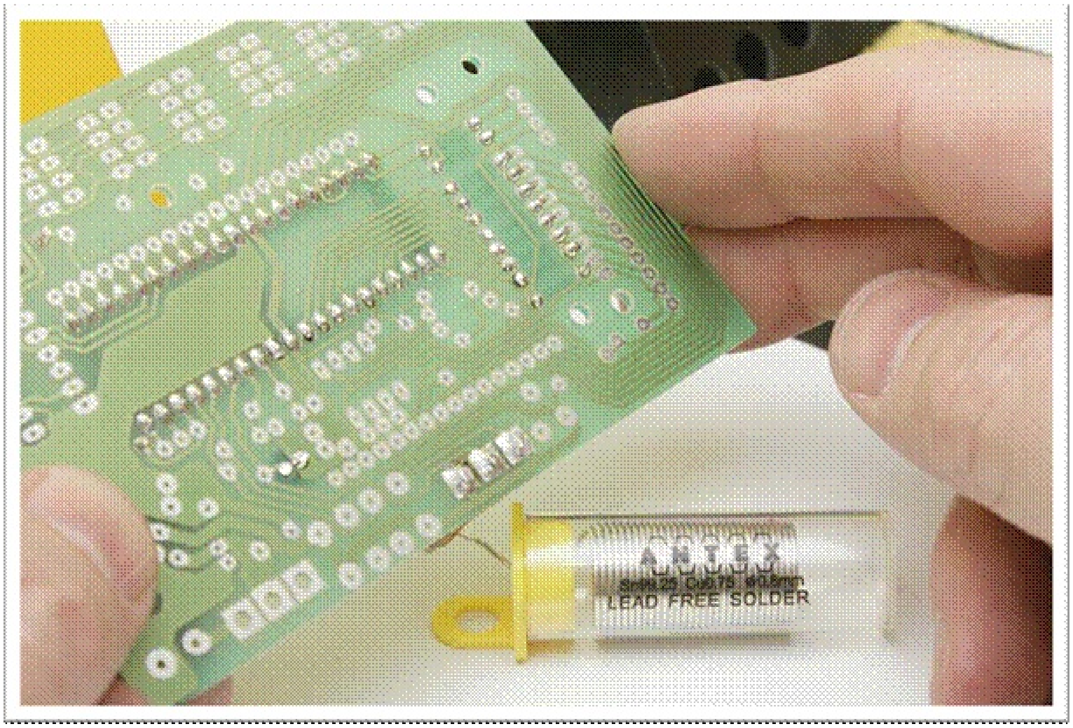
ANTEX
LEAD FREE SOLDER

^^ *As there is comparatively more metal to heat up, it'll take longer to solder the switch terminals, and you'll need more solder as well. Thicker gauge solder is useful at such times. Allow say 2-3 seconds to solder each terminal.*

Don't forget to clean the soldering iron tip on its damp sponge every now and then, to ensure the bit is kept clean and shiny. Later on I'll show you how to correct any problems by desoldering using various techniques. Next, we'll move on from printed circuit board "through-hole" soldering and look at how to handle wires and leads.

Interwiring – get hooked up!

With practice, through-hole soldering of p.c.b.s will become second nature. There's no substitute for tackling some soldering jobs though, particularly trying a simple kit based on a quality p.c.b. which will boost your confidence enormously.

Most electronic devices need connecting up to external components such as battery packs, speakers, l.e.d.s or switches. Usually, multi-stranded connecting wire is used to connect circuit boards and external parts together. Unlike single solid-core wire, *multi-stranded wire* is flexible and vibration-resistant. Hobbyists mainly use 7/0.2mm wire (7 strands, each 0.2mm diameter) for low-voltage hook-ups although much Chinese equipment uses much thinner wires than this. So let's look at some aspects of soldering this type of wire.

In a separate photo sequence I show how a potentiometer (a panel-mounted variable resistor) and a light-emitting diode (l.e.d.) are connected using multi-stranded wire. The same principles of soldering apply to most other components including panel-mounted switches, loudspeakers, buzzers, audio sockets and more.

Components usually have terminals or "tags" to which wires can be soldered. Start by ensuring the component's tags are clean: otherwise solder will not wet properly and the joint will be impossible to solder, so all contamination must be removed. This is especially true of parts that have been in storage a long time. The connections often oxidise or blacken, so clean the solder tags with e.g. an abrasive glass-fibre brush, or a needle file or abrasive paper. Using a glass fibre brush was shown earlier in "Cleanliness and Tinning the Bit".

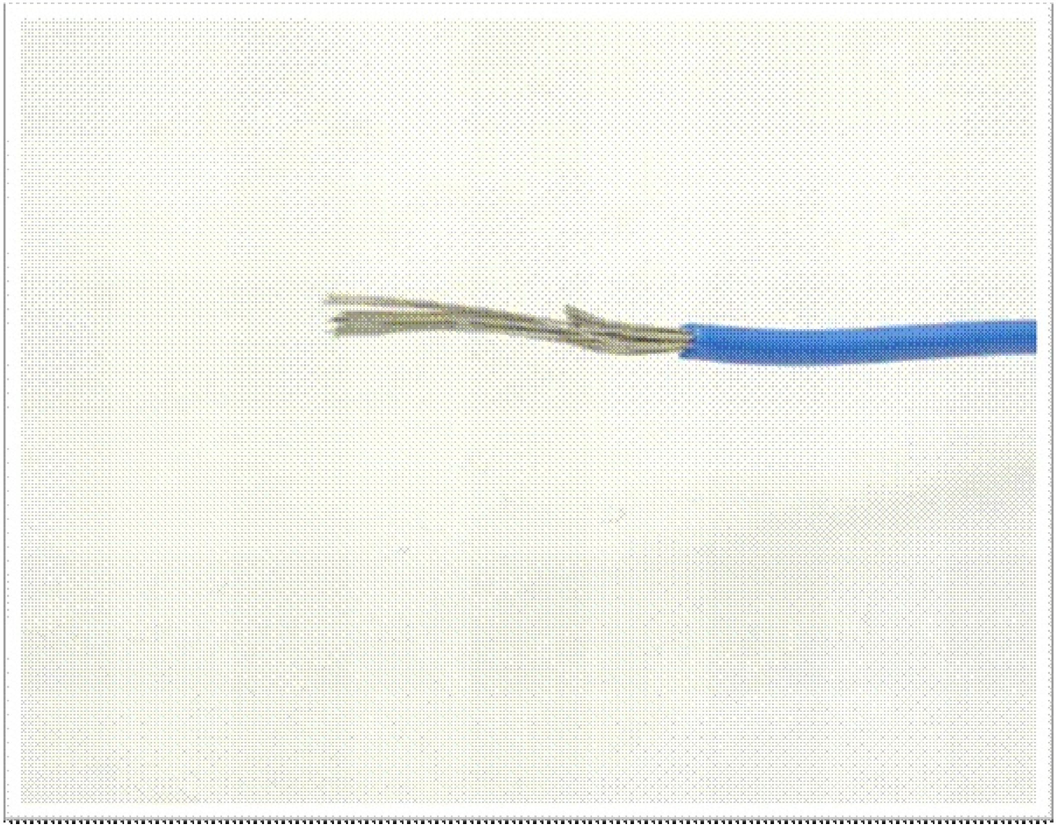

^^ How not to strip insulation off wire: some of the cores have also been cut – avoid doing this!

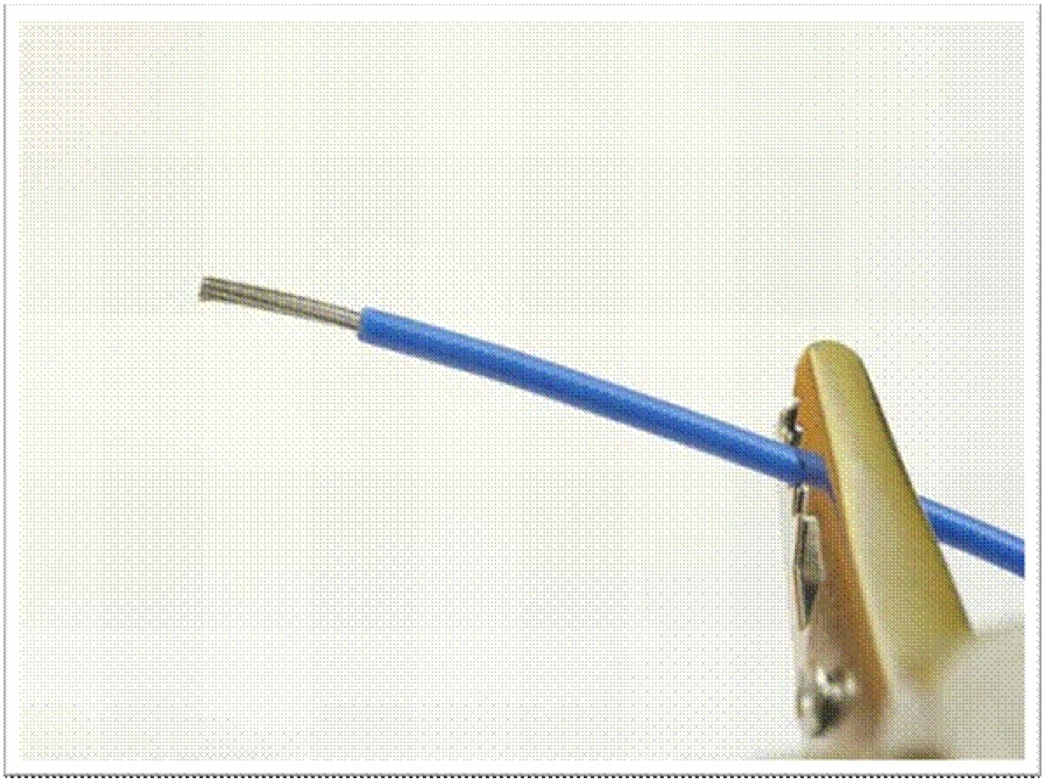

^^ *Gripping a wire end in a "Helping Hand" croc clip can help with soldering.*

After stripping a short length of insulation from the connecting wire, there are two ways to solder it to a component's solder tag. The first way is to "*tin" the stripped wire end to solidify it* - just heat it with the iron and melt a little solder on it, and let it cool. Poke the wire end through the solder tag, apply the hot iron to both parts and solder them together using a few millimetres of solder.

Although the assembly doesn't hold itself together so well during soldering (consider a Helping Hands jig if needed) this is quicker and easier to make and also easy to desolder again, and is perfectly adequate for most joints of this kind. The majority of commercial wire joints seem to be made this way.

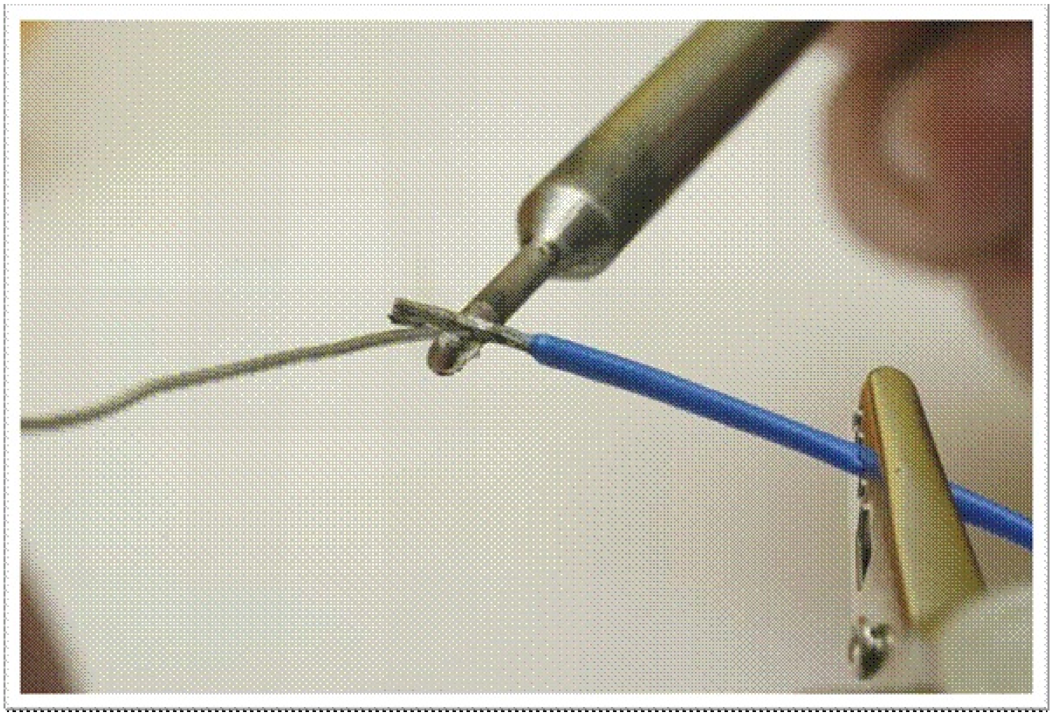

^^ Apply a hot iron and solder, in order to "tin" the wire ends. This makes them into a solid.

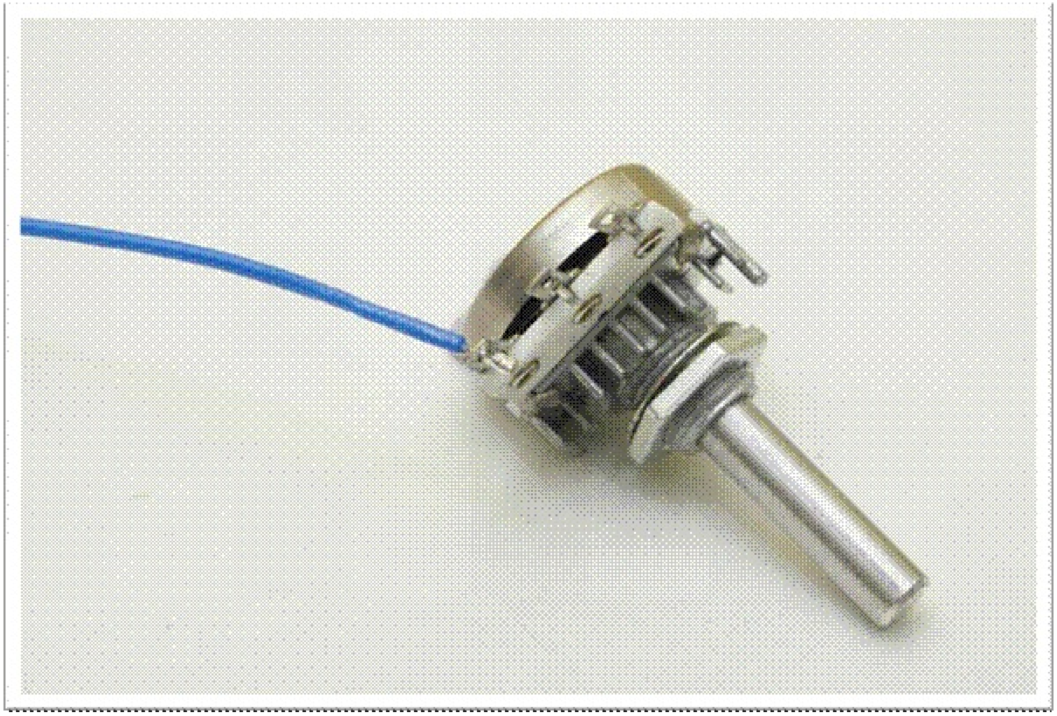

^^ Then feed the tinned wire through the hole in the solder tag. Crop the wire with cutters if needed...

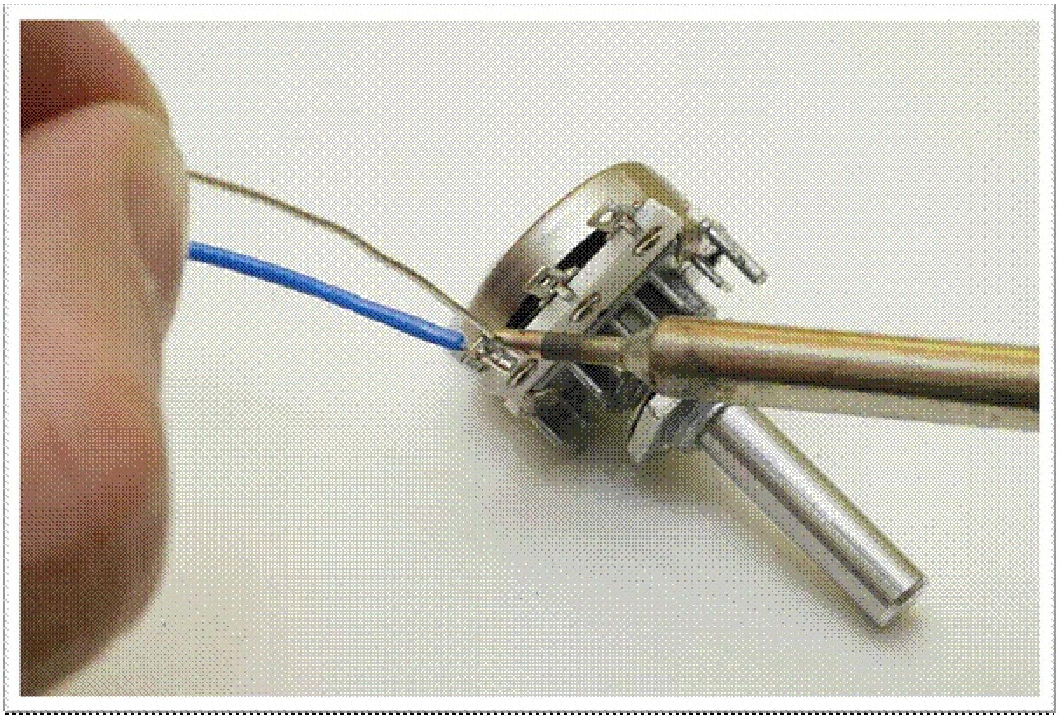

^^ … and apply the iron to heat the joint. As there's quite a lot of metal to heat, allow several seconds to heat it up and then simply solder the wire and solder tag together with a dab of solder wire.

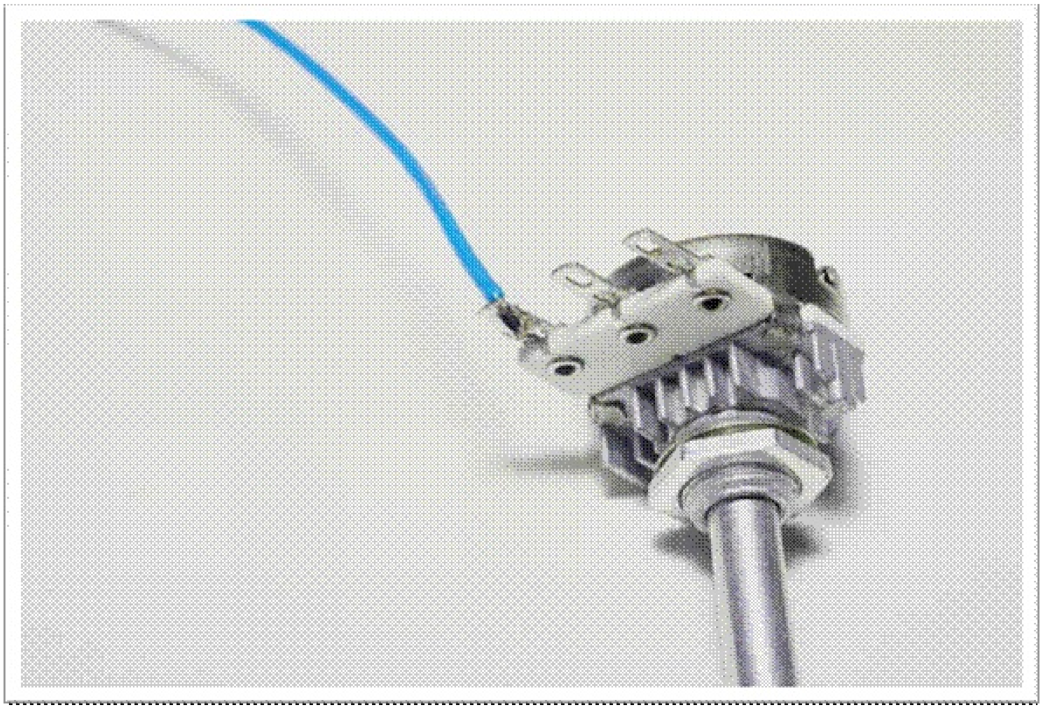

^^ The result is a perfectly satisfactory solder joint.

The second way is to loop the **untinned** wire through the tag a few turns and then solder it. This secures the wire during soldering, but it's messier to desolder if things go wrong (see later):

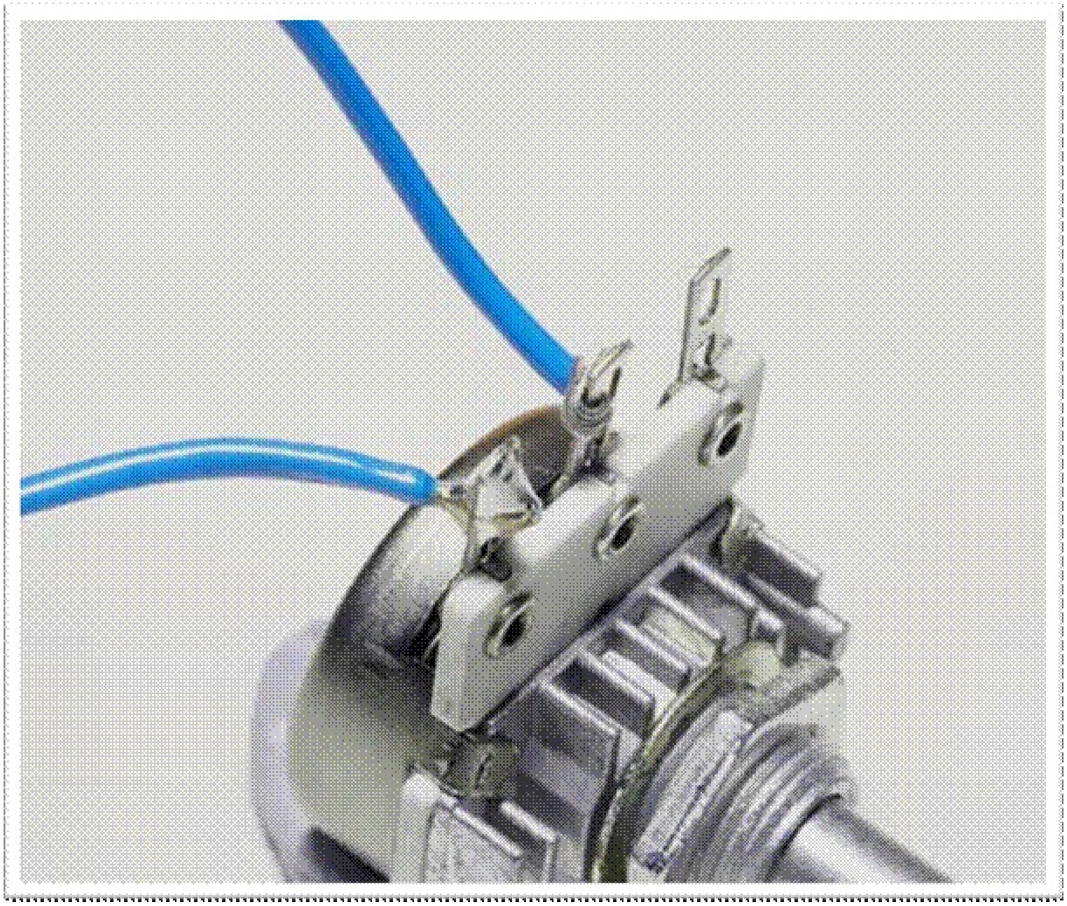

^^ The centre wire (potentiometer wiper terminal) has been wrapped around to secure it before soldering.

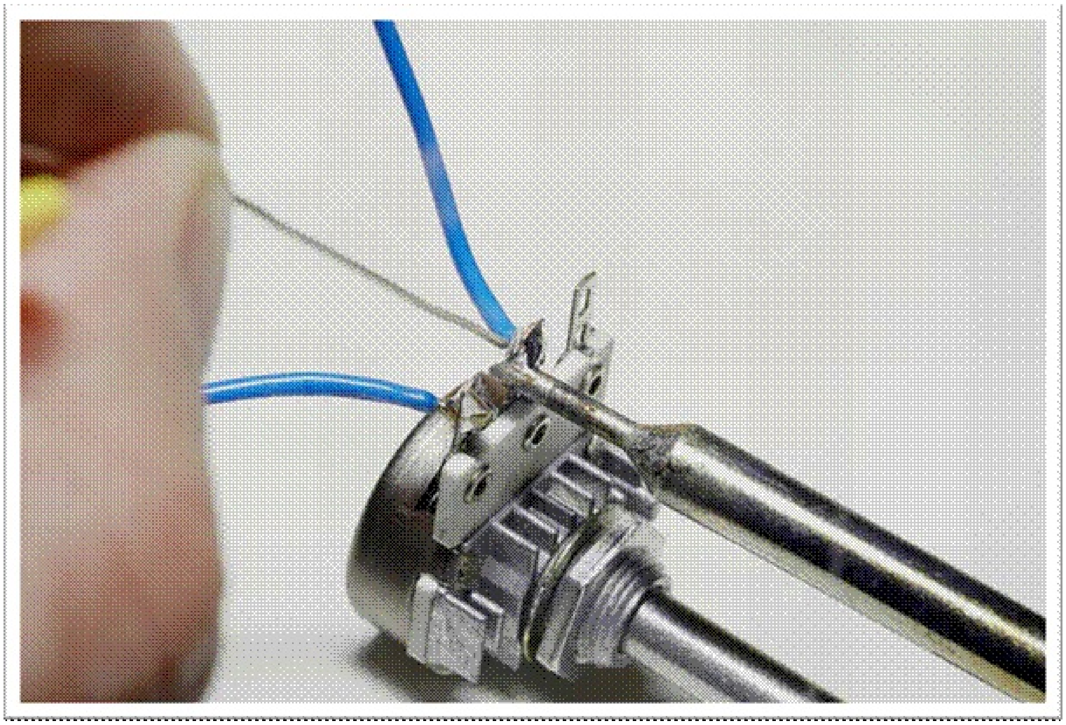

^^ *Solder the joint in just the same way – heat it up and dab on some solder wire.*

Much electronic equipment is connected or inter-wired this way and soldering everything together is a key stage in assembling any electronic project. I'll explain some refinements you can try, later.

Wire joints

Soldering isn't always the best solution to some problems, for reasons that I explain in the next section called Fatigue and Breakage. However you can solder wires together to join them simply and cheaply, and next I show a classic wire joint of this kind, sometimes called a **Western Union** joint.

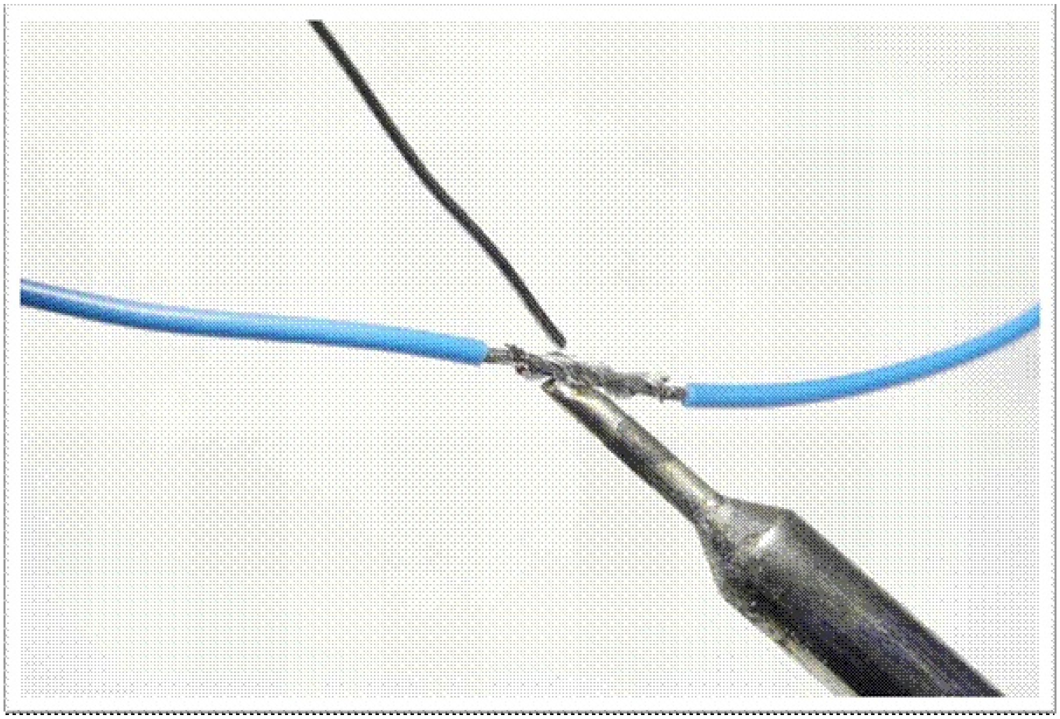

^^ *Soldering a Western Union wire joint.*

Strip the ends from each wire using wire strippers, and try not to cut any of the stranded cores or this will create a weakness. (Stray strands of wire can also be a hazard.) Then twist the wires securely together several times, ideally four or five. Apply a soldering iron to the exposed joint and dab on some solder until it flows fully over the joint, then remove the iron and allow the joint to cool.

The wire's plastic insulation may shrink back a little due to the heat, but try not to overheat it excessively. If you have the resources you can finish off by insulating the joint with some heatshrink tubing (slide it over a wire before you make the solder joint!). Otherwise use ordinary PVC tubing or insulation tape if you have any.

A solder tag can be soldered easily enough, and it's probably best to feed a stripped wire end through the tag and twist it over several times before soldering it, so it's held together when soldering.

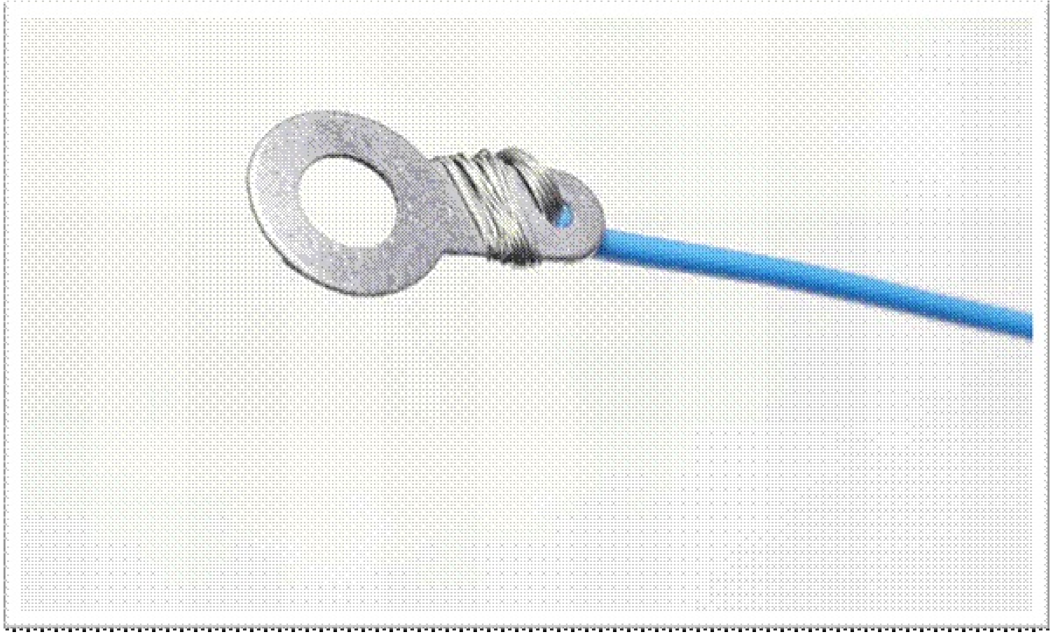

^^ *A solder tag is soldered by wrapping a stripped wire end through the hole and round again, to secure it.*

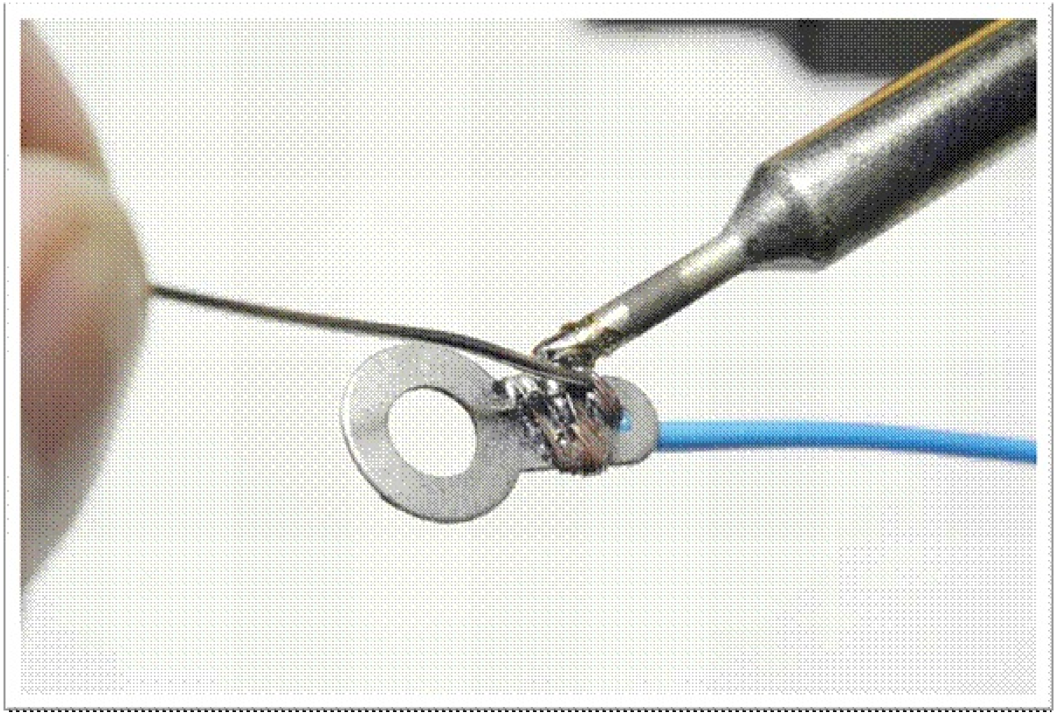

^^ *Apply plenty of heat and solder for a few seconds.*

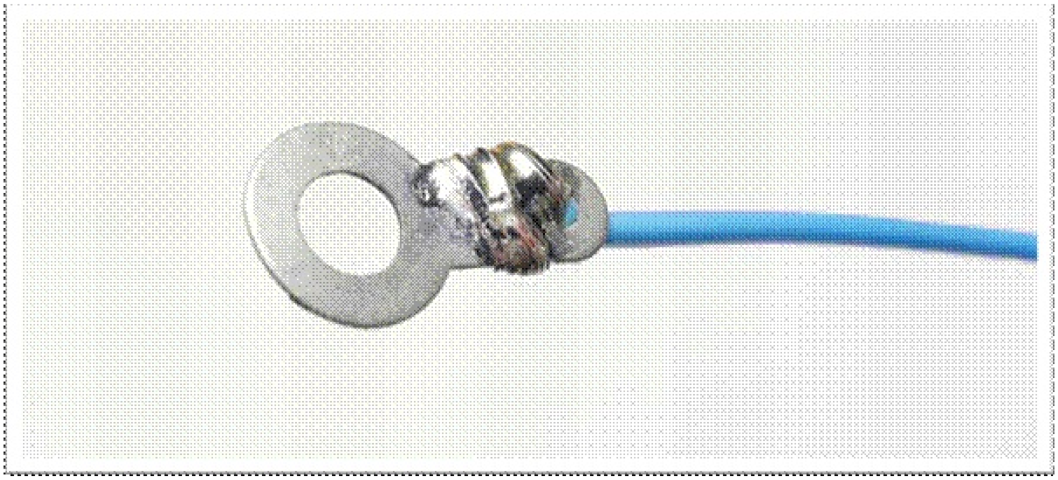

^^ *The completed solder tag*

The next sequence show how the same basic principles of soldering wires are used to connect up a device such as a panel-mounting l.e.d.. A series resistor can be wrapped around and soldered directly to one of the l.e.d. solid wires, and then two multi-stranded connecting leads are soldered to the l.e.d. and resistor. It's best to use heatshrink tubing or PVC sleeve to insulate the solder joints and prevent short circuits afterwards.

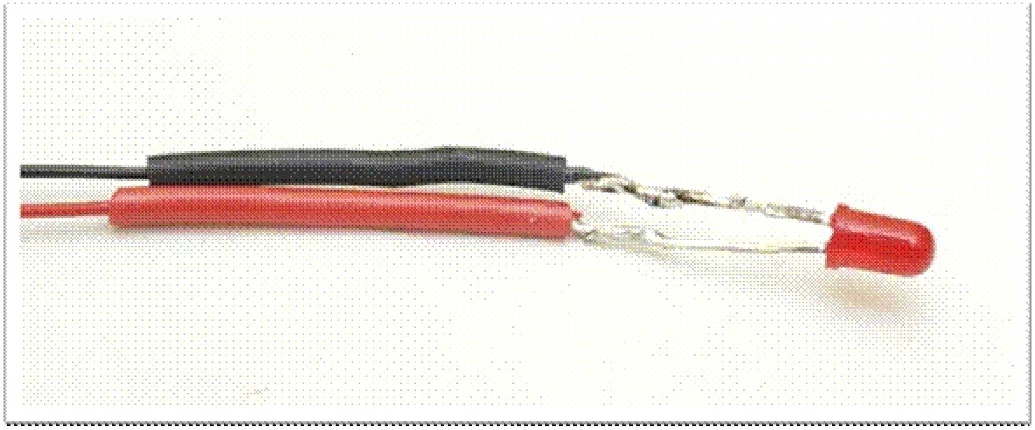

^^ *Multi-stranded wires can be soldered to the solid leads of an l.e.d.: insulate them afterwards. Different coloured wires identify the polarity of anode and cathode.*

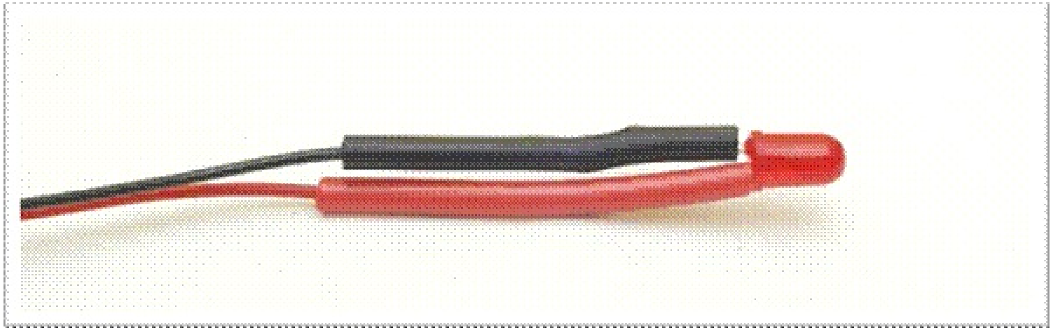

^^ PVC sleeving or heatshrink is needed to prevent short circuits.

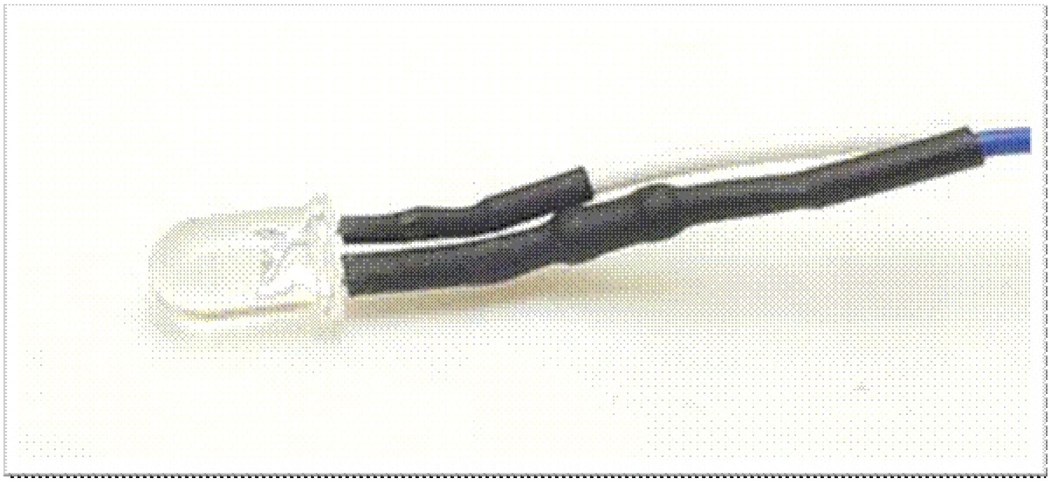

^^ A commercial l.e.d.– the series resistor can be made out in the sleeving.

Wires can also be soldered directly to printed circuit boards by stripping and tinning the ends, provided they fit through the holes in the p.c.b. It's a very common and cheap way of hooking up a board using "flying leads" and you'll see this all the time in consumer electronic equipment. For convenience so-called "solder pins" can be used, onto which flying leads may be soldered from the component-side instead.

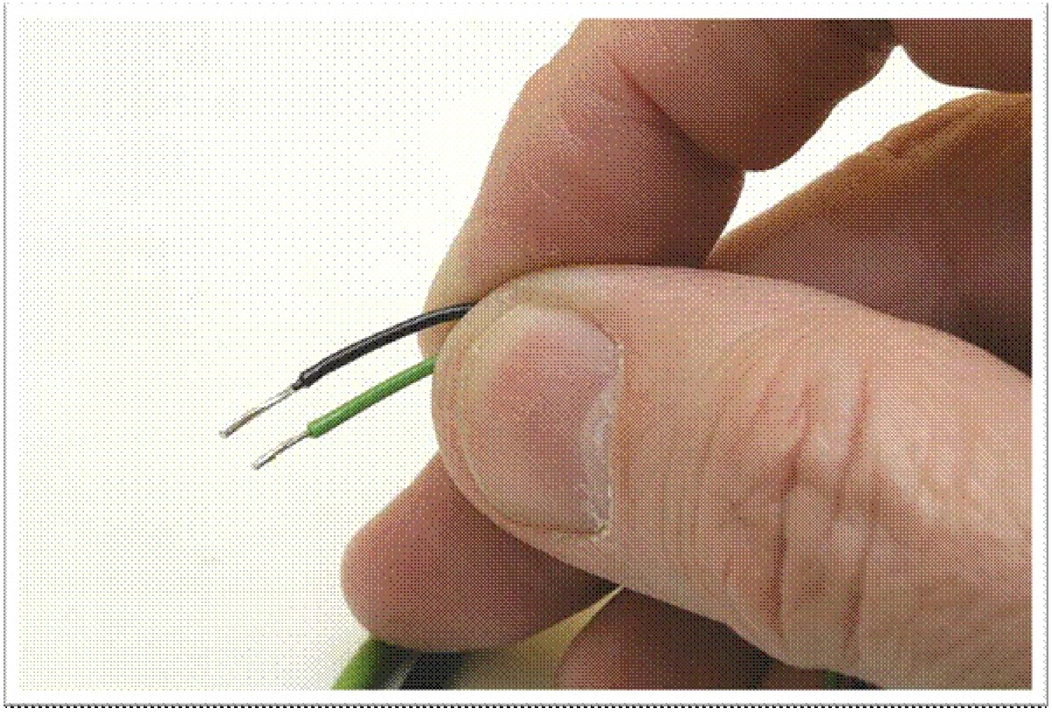

^^ A set of l.e.d.s connecting wires, stripped, tinned and ready to be soldered to the p.c.b.

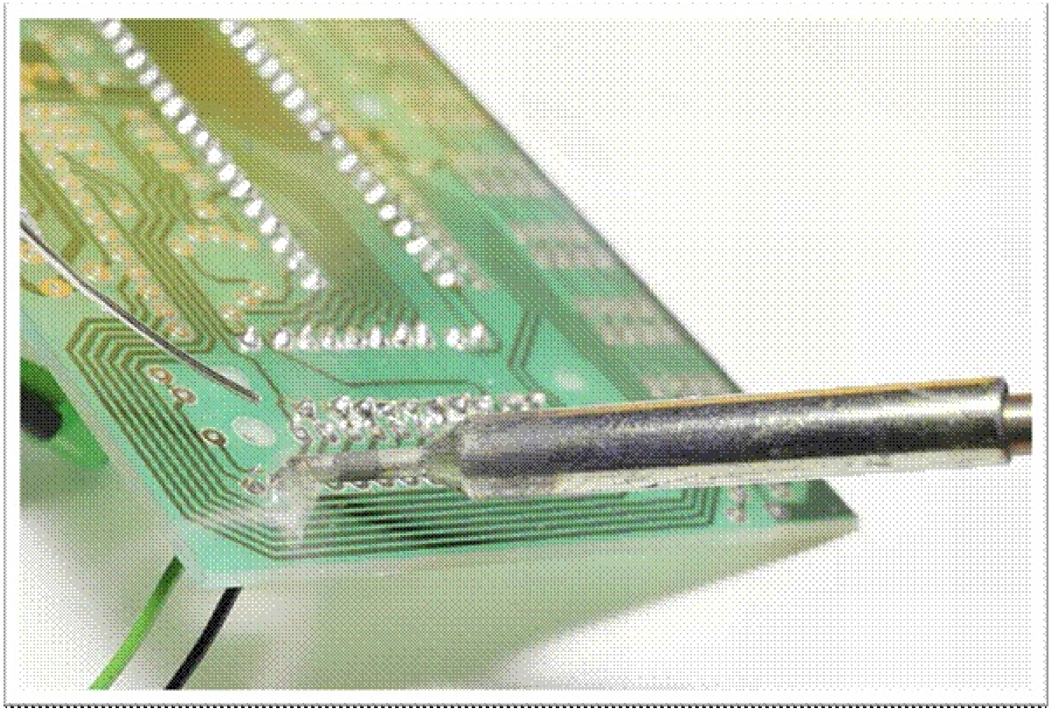

^^ Solder the two "flying leads" to the copper pads in the usual way. Hopefully the holes are big enough for the wires!

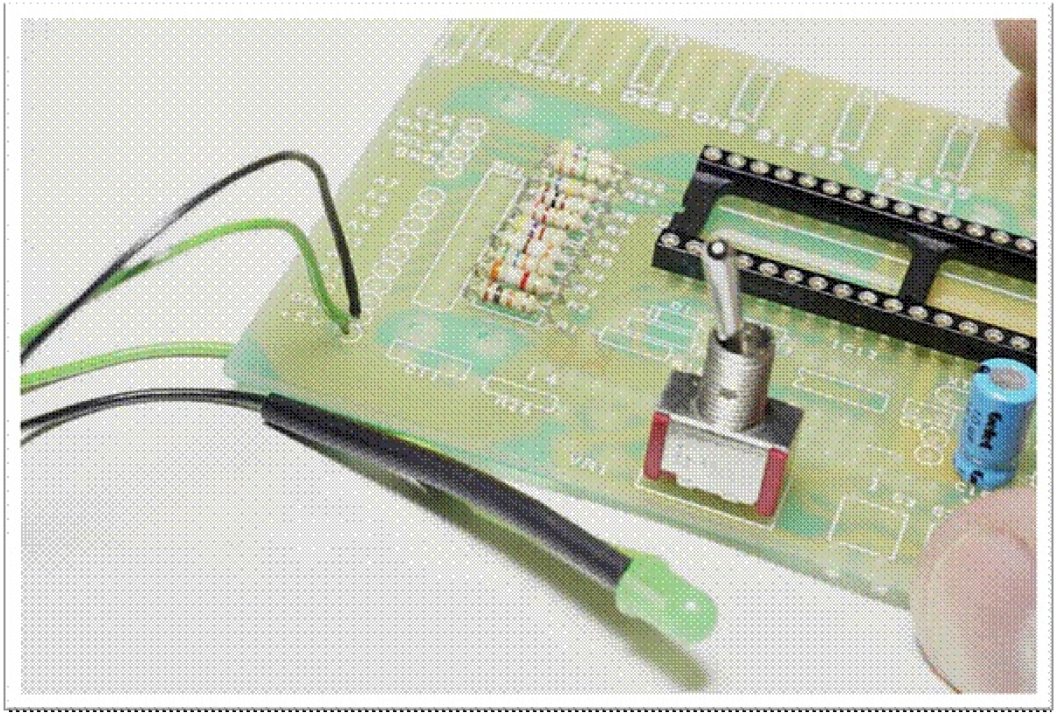

^^ *Connecting an l.e.d. to a circuit board with flying leads.*

Wires can also be attached to the underside (solder-side) of circuit boards, tacking them onto existing solder joints by re-melting them and absorbing the tinned wire end into it ("reflow soldering" – see later). This is a cheap and cheerful, semi-reliable way of rigging wires to a circuit board, and it's used all the time in imported consumer electronics.

Tidying up

I’ll leave it to readers to decide whether they should cut off excess wire before or after soldering. After the soldering is complete, I prefer to tidy up the joint by snipping any excess wire from the joint using a pair of “end cutters” shown earlier. These expensive hand tools have specially angled blades that snip the joint neatly down to the top of the solder joint. Ordinary side cutters will do fine.

It’s worth taking time out to inspect the work closely, looking for any missing solder joints, whiskers of solder or swarf shorting out any solder pads, and all such potential problem areas should be dealt with prior to testing the board. Faultfinding goes beyond the scope of this guide, but it’s true to say that almost always, any problems noticed after powering up the circuit are due to soldering faults or wrong components being used.

Reflow technique

Another technique often used is “reflow” soldering. This is used to “tack” devices or wires together, especially if very small, sensitive or fiddly parts are involved. There might be no room to make a “proper” sturdy joint, or it might just not be necessary to have any mechanical strength in the joint, especially if tiny parts are used.

As an example, imagine a small temperature sensor (I used a transistor) for use in a thermometer project. It could be quite tricky to solder flying leads onto the sensor’s leadouts, so a good approach is to tin both the flying leads and the sensor’s leads, and then simply touch them together and re-melt the solder with the iron. There’s no need to add any more solder, because the solder that’s already there will re-melt and the joint will be made. Sometimes this is called a butt joint.

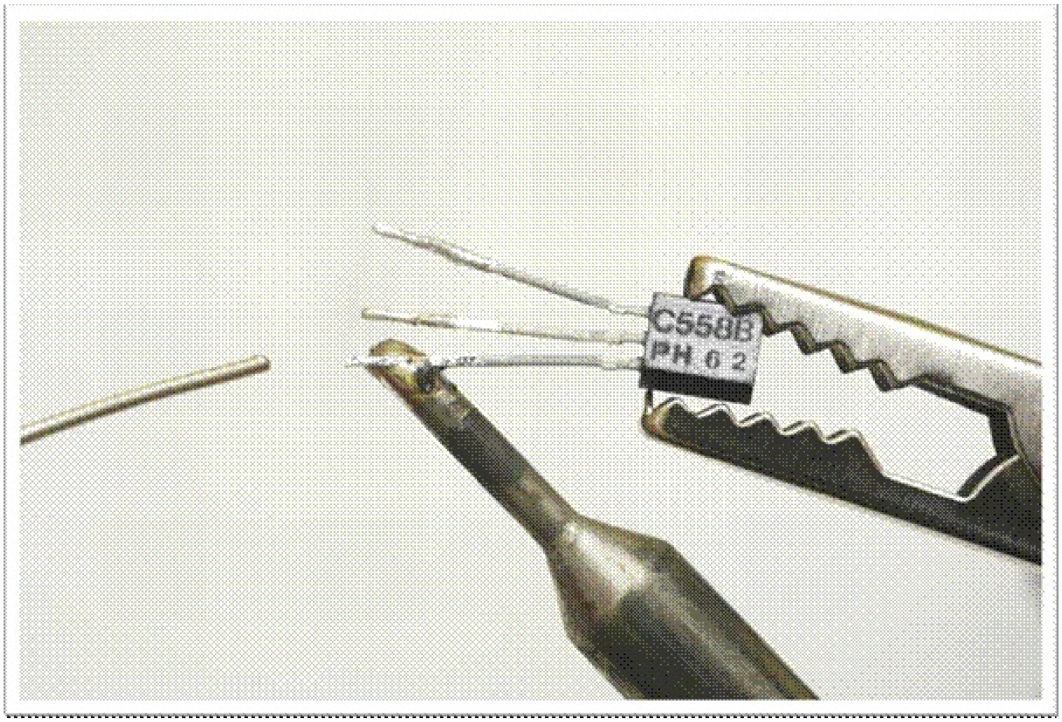

^^ *The leads of this sensor (transistor) have been tinned, ready for flying leads to be tacked onto them with a reflow method. A Helping Hands jig might help!*

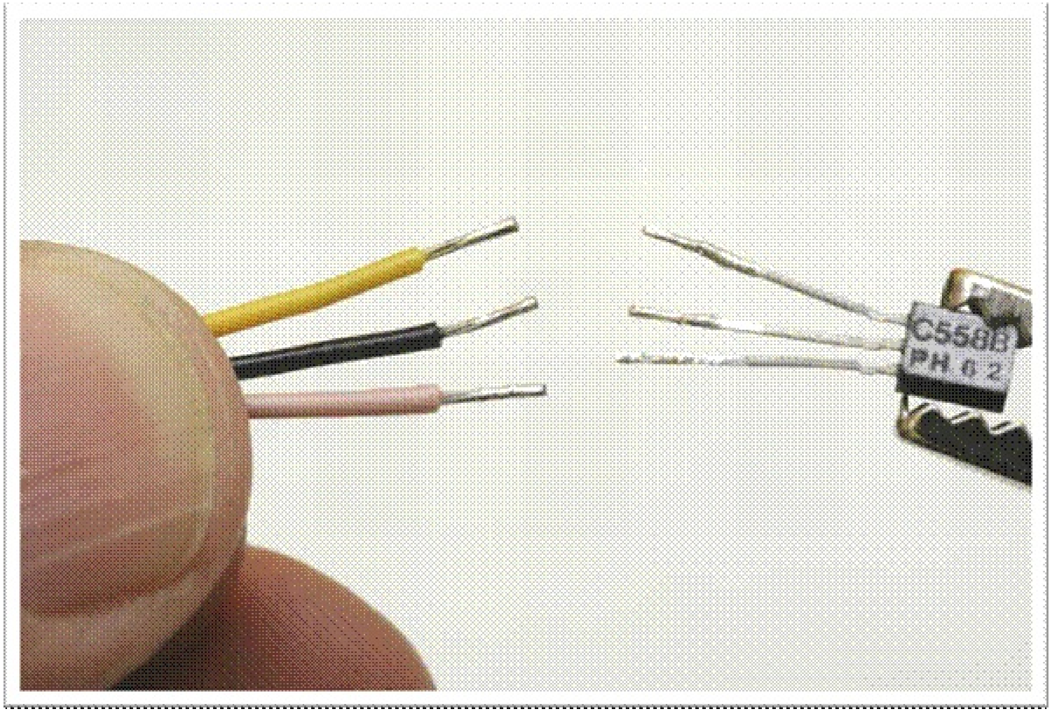

^^ The three flying leads stripped, tinned and ready to be reflow-soldered onto the device.

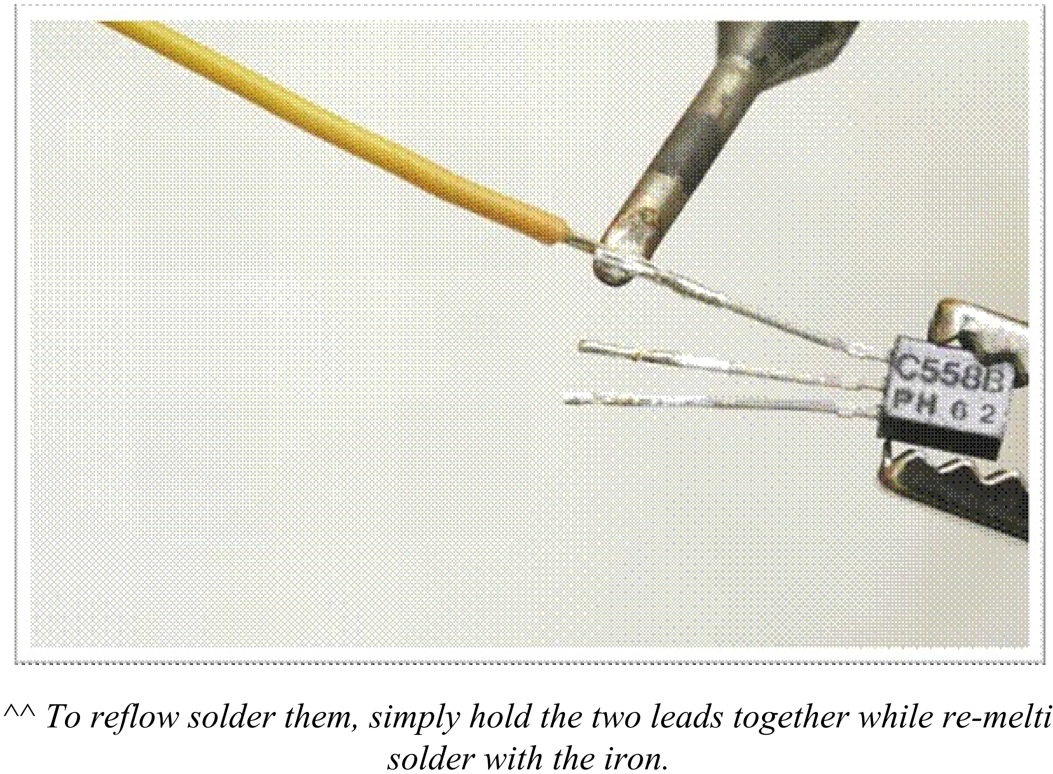

^^ To reflow solder them, simply hold the two leads together while re-melting the solder with the iron.

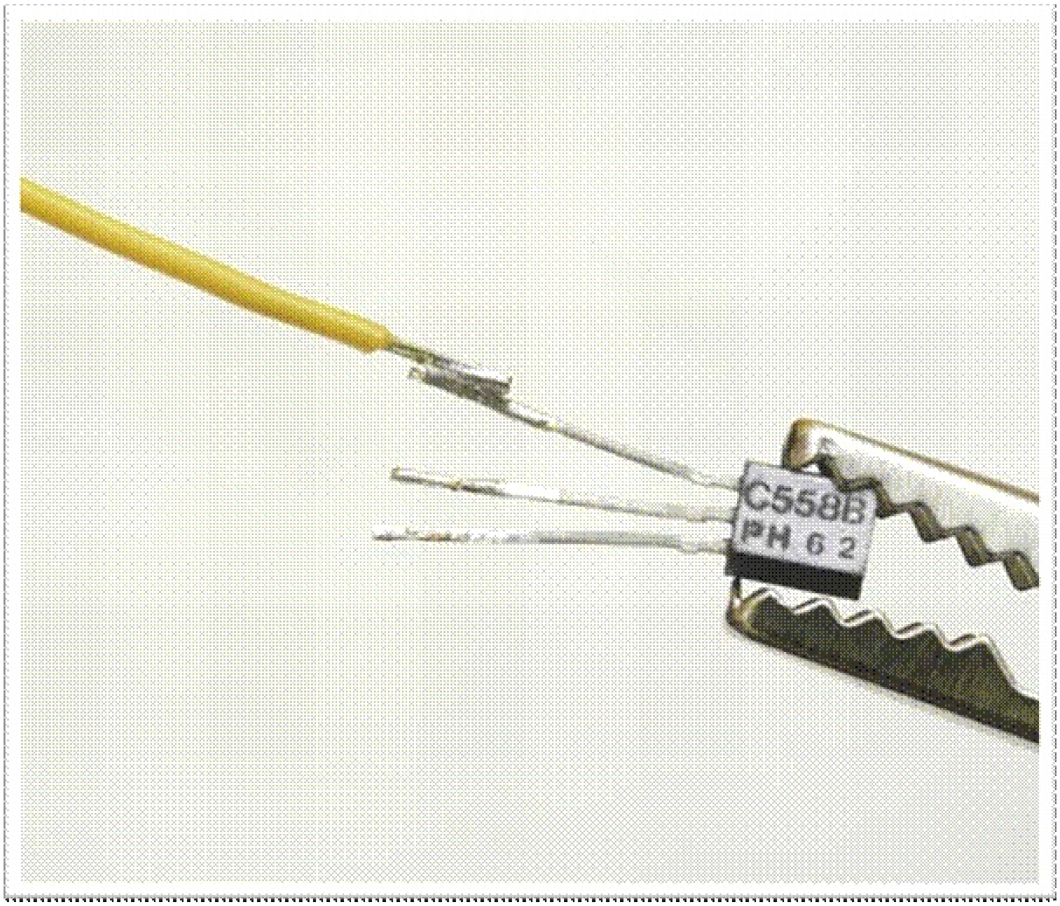

^^ *Remove the iron and let the solder cool. The wire is tacked on.*

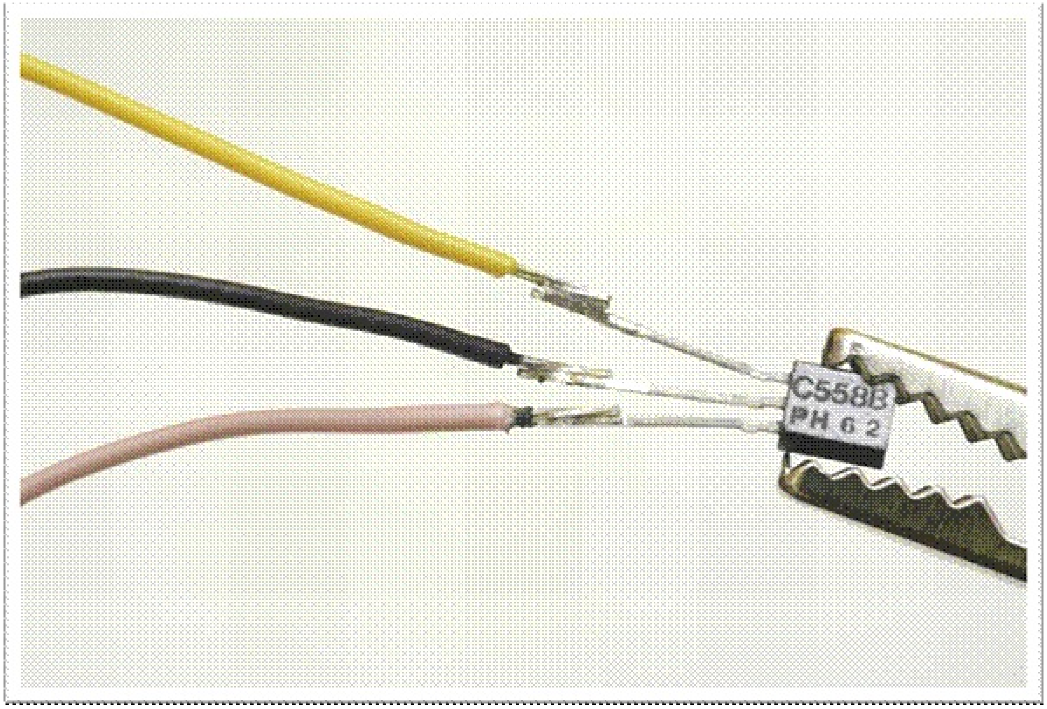

^^ *Repeat for the other leads. They just need insulating with sleeving, then job done!*

Depicted next is a typical "tag strip", an insulated panel with metal solder tags used for making sundry connections. (Entire TVs and radios used to be hand-built with them, in the early-mid 20th Century!) The principles of soldering are exactly the same, but more time is needed when applying the soldering iron because more metal is present, which needs more heat. You'll need more solder for bigger joints like these, so larger diameter solder makes a quicker job of it. Consider adding more flux (see earlier) to see if it helps.

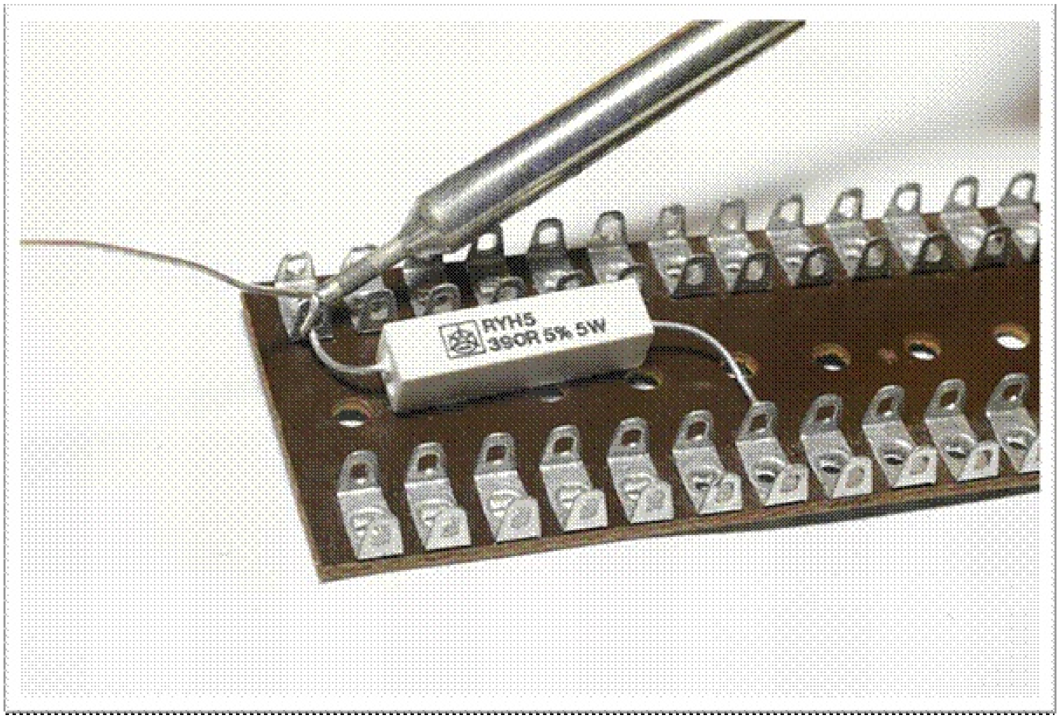

^^ *Wrap wires through the terminals and arrange everything neatly, then solder as normal.*

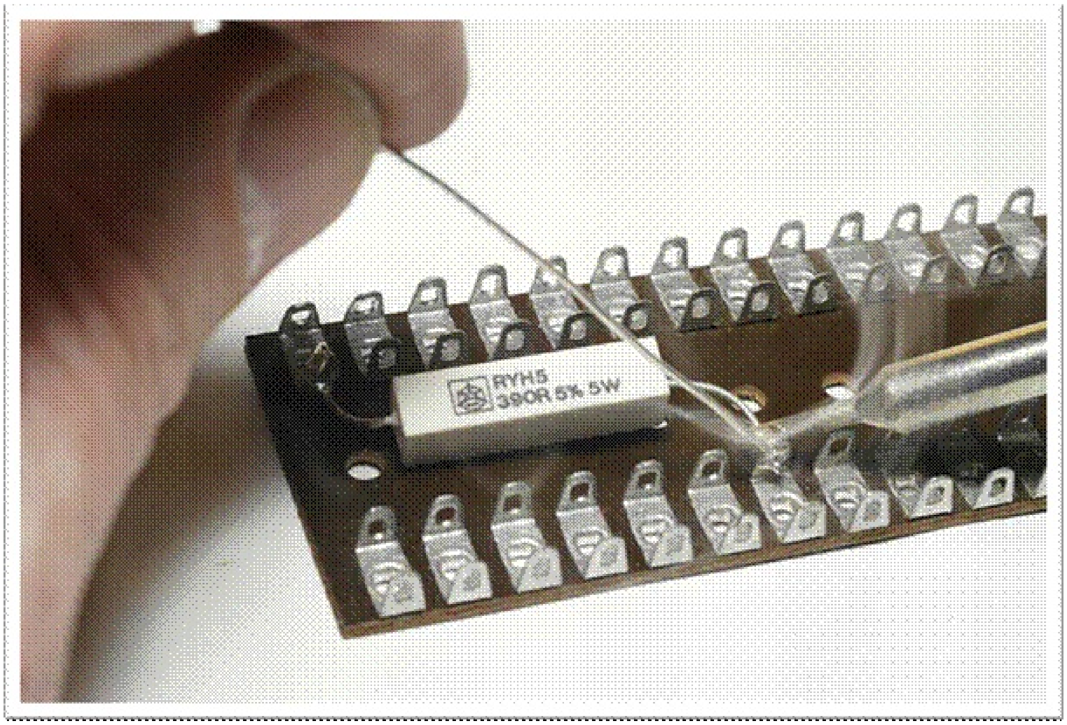

^^ Don't be afraid to apply more heat with larger assemblies like these: they contain more metal than, say, an ordinary p.c.b so allow more time for the solder to flow over everything properly.

Fatigue & Breakage

Earlier I explained how multi-stranded hookup wire is flexible, which also makes it vibration proof – that's why a car's electrics are full of multi-stranded wiring, and test equipment probes use ultra-flexible multi-stranded wire for the same reason. Solid-core wire can be bent and will stay in shape (called "plate wiring") but if it's repeatedly bent or vibrated then it may eventually break somewhere due to fatigue.

The same is true of wires that have been soldered or tinned. No longer is the wire 100% multi-cored and flexible – instead it's been turned into a *single core wire* at the point where it's been soldered. This is potentially a weak spot and could eventually fracture due to fatigue, if subjected to continued vibration (e.g. in a car engine bay or in motorised equipment).

In a lot of equipment problems can be avoided by adding *strain reliefs* of some sort, to stop the wire being flexed where it's been soldered. Heatshrink tubing, or a dab of hot-melt glue, are ways of taking the pressure off the joint and ensuring soldered wires won't snap off due to vibration.

One of the reasons that higher-quality equipment and cars etc. use crimp terminals and connectors is that crimped (as opposed to soldered) connections retain all the flexibility of multi-core stranded wire from end to end, avoiding problems of wires breaking off. A *ferrule* is a very neat way of tidying wire ends and preventing stray strands of wire doing damage. Ferrules can be used to connect wires into screw terminal blocks etc. Simply clamp the terminal block down onto it and the screw will grip the ferrule.

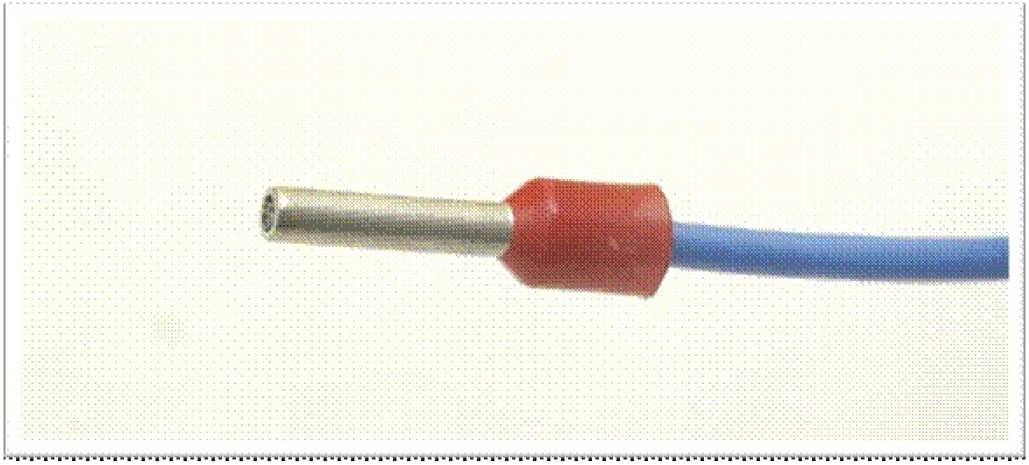

^^ Instead of soldering them, wires can be terminated with ferrules prior to fitting to e.g. screw terminal blocks. This makes them vibration-proof and also avoids any problem of stray wires poking out.

Faults & Desoldering techniques

By putting into practice the guidelines in my *Basic Soldering Guide* practice, there's no reason at all why you should not obtain perfect results and eliminate any potential problems. Hopefully the information gives you plenty of guidance to tackle various soldering projects with confidence. There's no substitute for getting some hands-on experience though, so I'd repeat the advice to try assembling a simple high quality electronic kit or two, such as those produced by Velleman and see how you get on. Powering up your first project successfully is a great thrill as every electronics hobbyist knows.

^^ *Popular desoldering products, a pump and various widths of desolder braid.*

Let's now look at reversing the soldering procedure – what to do if things go wrong, or maybe you have to repair a circuit by replacing a faulty component.

A solder joint which is badly made is likely to be electrically "noisy", unreliable and will probably worsen over time. Expansion and contraction of the joint due to heating and cooling can also throw up intermittent problems later down the line. The joint may look OK but underneath it may have a poor electrical connection, or could work initially and then cause the equipment to fail at a later date! These intermittent problems can be maddening to fix. TV repair technicians have an uncanny ability to go straight to a faulty solder joint because they see the same problem all the time, especially on equipment that has a "reputation".

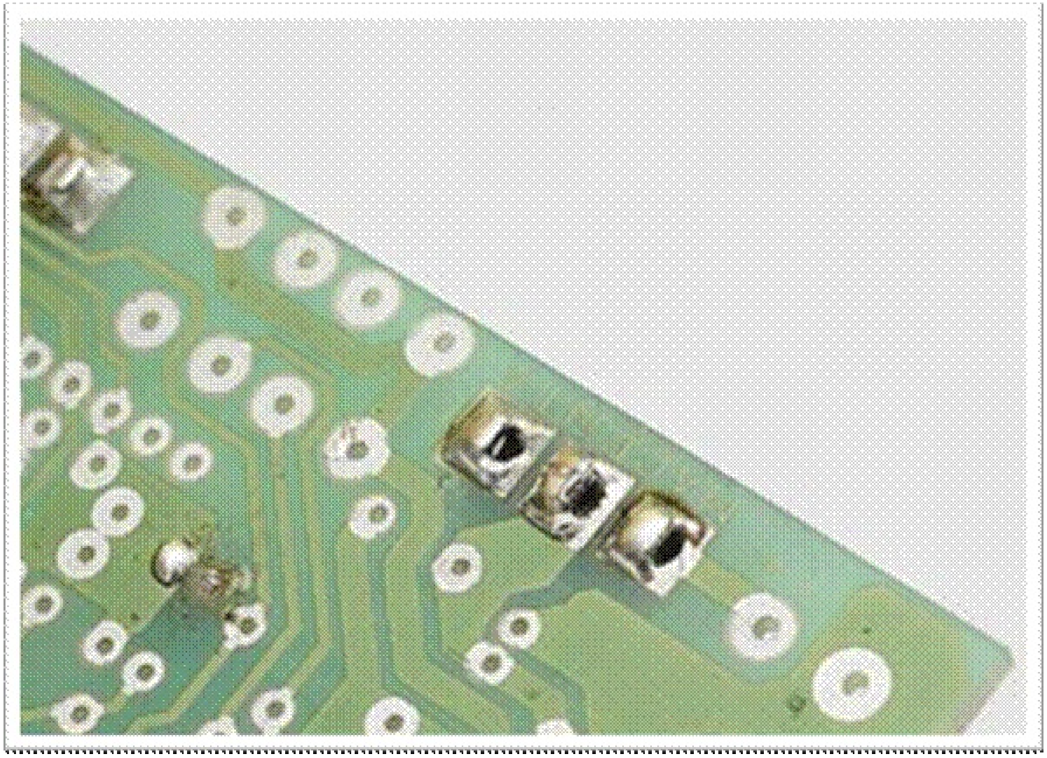

^^ These dry or grey/ gray joints were caused by dirty parts, poor heating and inadequate solder coverage

A solder joint that's poorly formed is called a "dry joint" or "cold joint" or a "grey/ gray" joint. Usually it results from inadequate heating, dirt or grease preventing the solder from melting onto the parts properly, and is often noticeable because the solder tends not to wet the surface properly. Instead it forms beads or globules. Alternatively, if it seems to take a very long time for the solder to wet the joint, that's another sign of contamination and that the joint may be a dry one, or the material is incompatible anyway. A matt, crystalline appearance instead of a shiny joint points to inadequate heating: the solder cooled down far too quickly and didn't flow properly.

Whether you want to replace a faulty component or fix a dry or poor-quality solder joint, it's usually necessary to remove the troublesome solder and then re-solder it afresh. Naturally, there are tools and techniques that make the job easy. It's very bad practice to simply re-melt the joint and then lash out with the board, whiplash style, hoping that the molten solder will be flicked off the board.

The usual way of removing solder from a joint is to use a *desoldering pump*. These work like a small spring-loaded bicycle pump, only in reverse! A plunger is pressed down until it locks into position. It's released by pressing a button which sucks air back through a pointed nozzle, carrying any molten solder with it. It may take one or two attempts to clean up a joint this way, but a small desoldering pump is an invaluable tool especially for p.c.b. work and they are widely available now.

^^ *A hobby desoldering pump primed for use*

Desoldering pumps often have a heatproof P.T.F.E. nozzle which may need replacing occasionally. Each time the button is pressed, a plunger clears the nozzle but sometimes solder particles and swarf will be ejected in the process; when you prime the pump, point the nozzle into a small pot or old aerosol top to catch any debris. Remove the spout and clean out the pump from time to time.

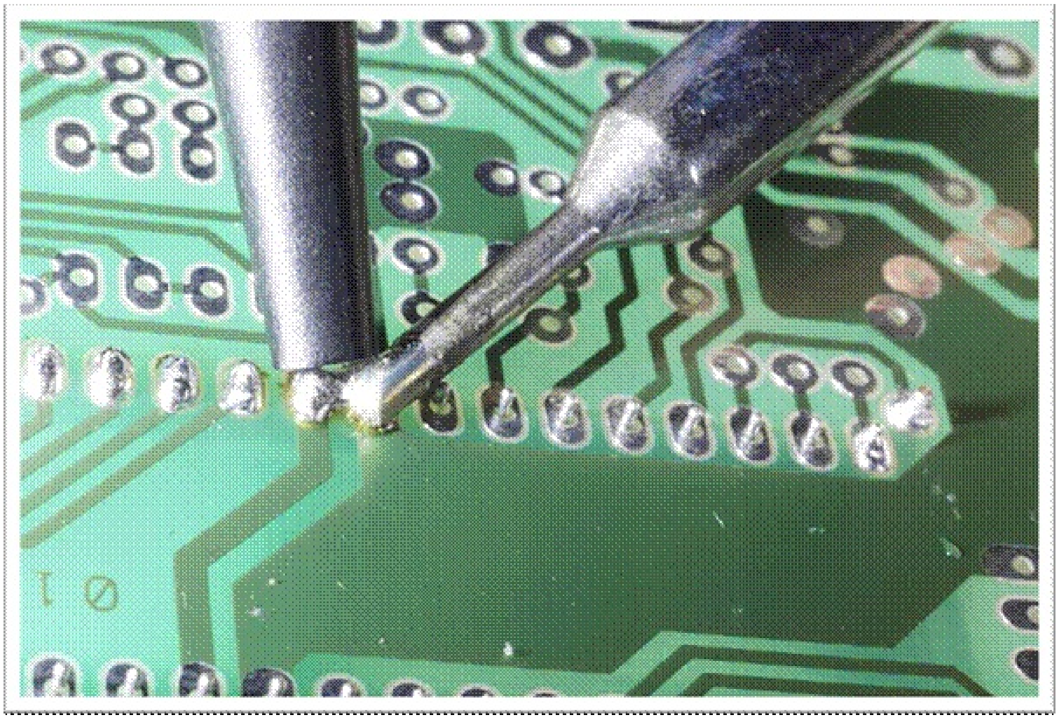

^^ *Suck up molten solder using a desoldering pump*

With very stubborn joints where the last traces of molten solder just can’t be shifted, it sometimes helps to actually *add* more solder and then desolder the whole lot again with a pump.

Desolder Braid

An alternative to a pump is to use *desoldering braid* which arrives in small dispenser reels. It's a flux-impregnated fine copper braid which is applied to the molten joint, and the solder is then drawn up into the wick by capillary action. It's remarkably effective and for certain tasks, it can be more thorough than a pump. I recommend that a small reel is bought (start with 1.5mm width) for the toolbox, to tackle larger or difficult joints which would take several attempts with a pump.

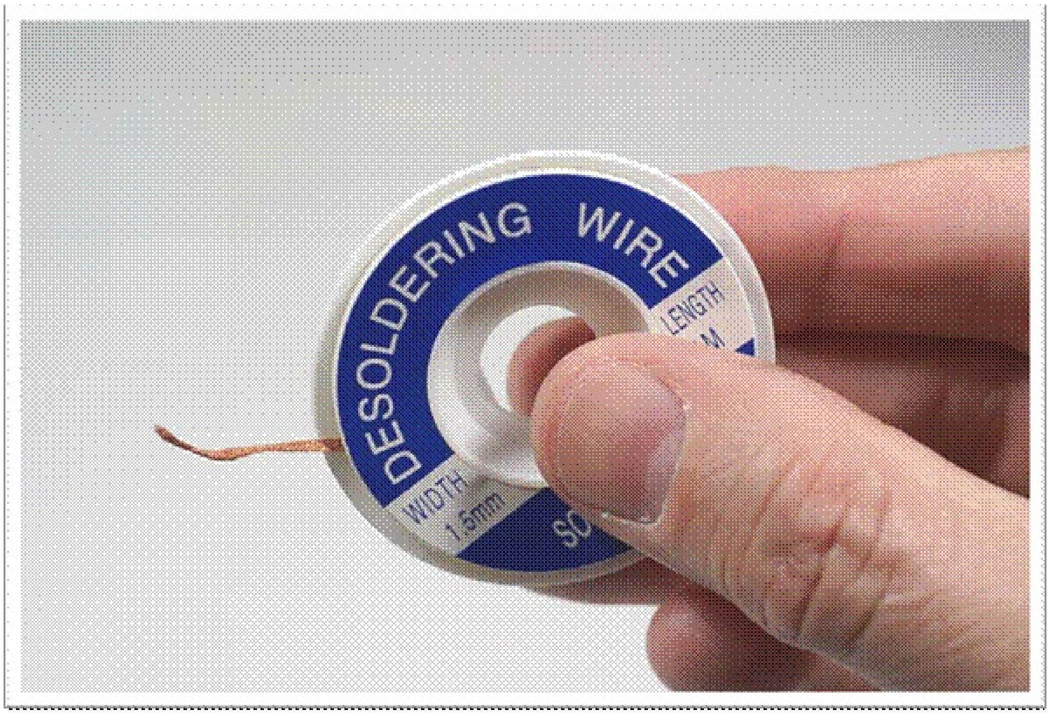

^^ Desolder braid is also handy and can sometimes be more effective than a desolder pump. It comes in various widths to suit the scale of work being tackled.

To use desolder braid, press the end of the braid down onto the joint using the tip of an iron, and let the solder melt underneath: the braid will then absorb the solder. The braid becomes hot so beware of burns. Once the solder's solidified on the braid, cut it off and discard.

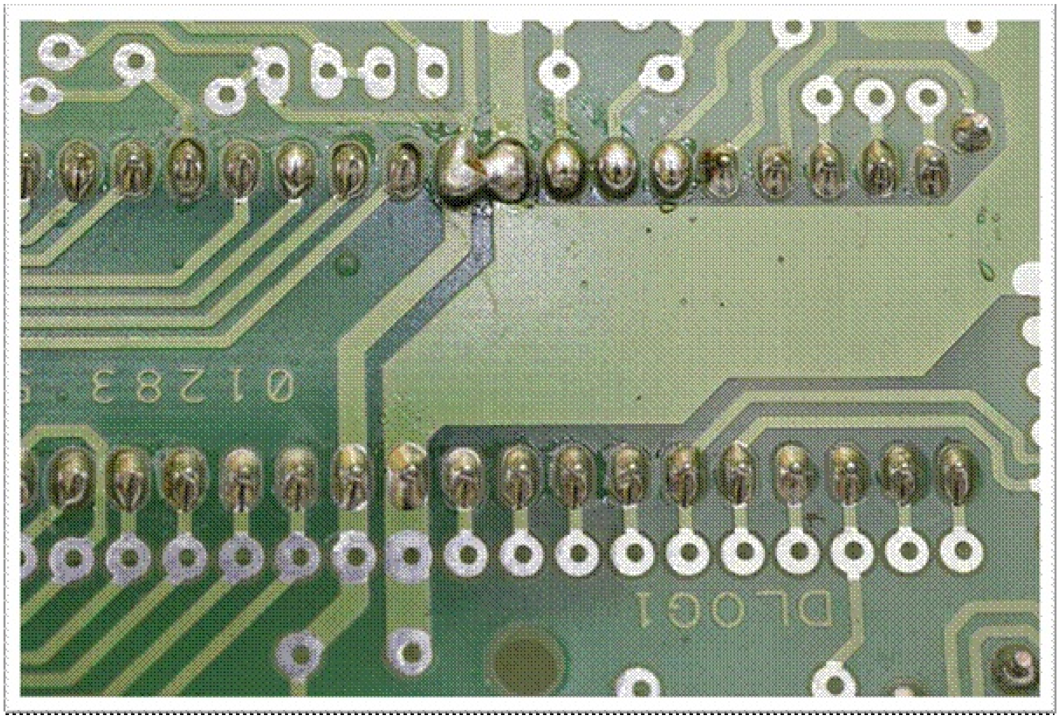
01283 5
DLOG1

^^ Desolder braid can also be used to remove excess solder, e.g. two i.c. pins shorted together.

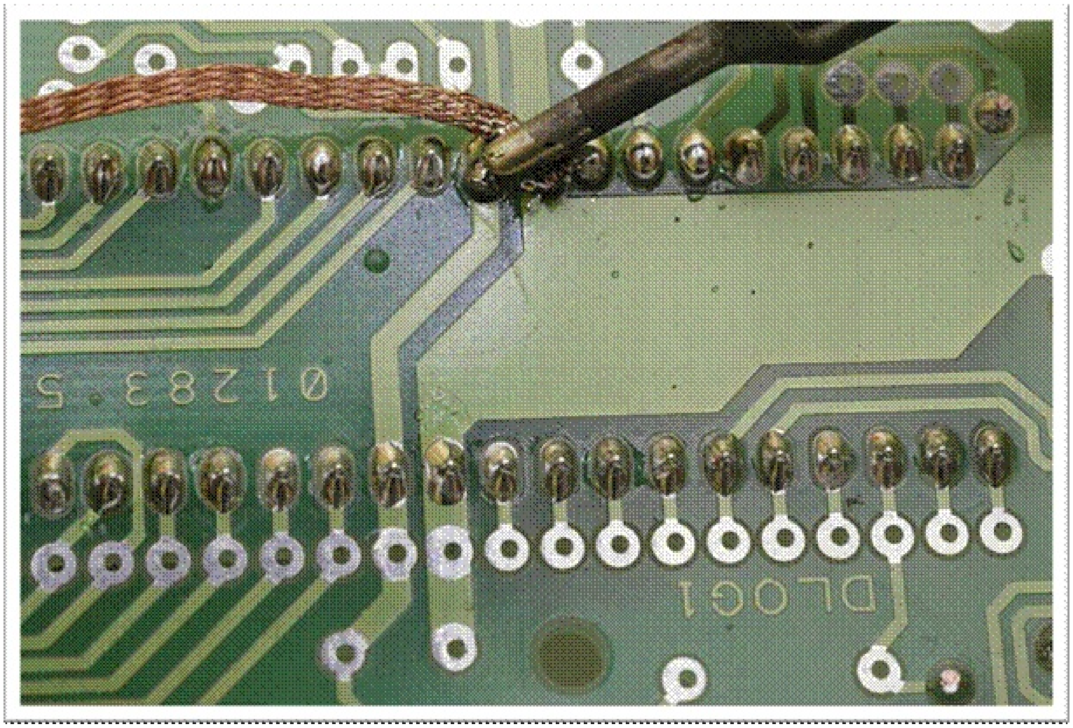
01283 5
DLOG1

^^ Press some desolder braid over the joint, then melt it with an iron to draw molten solder up the braid.

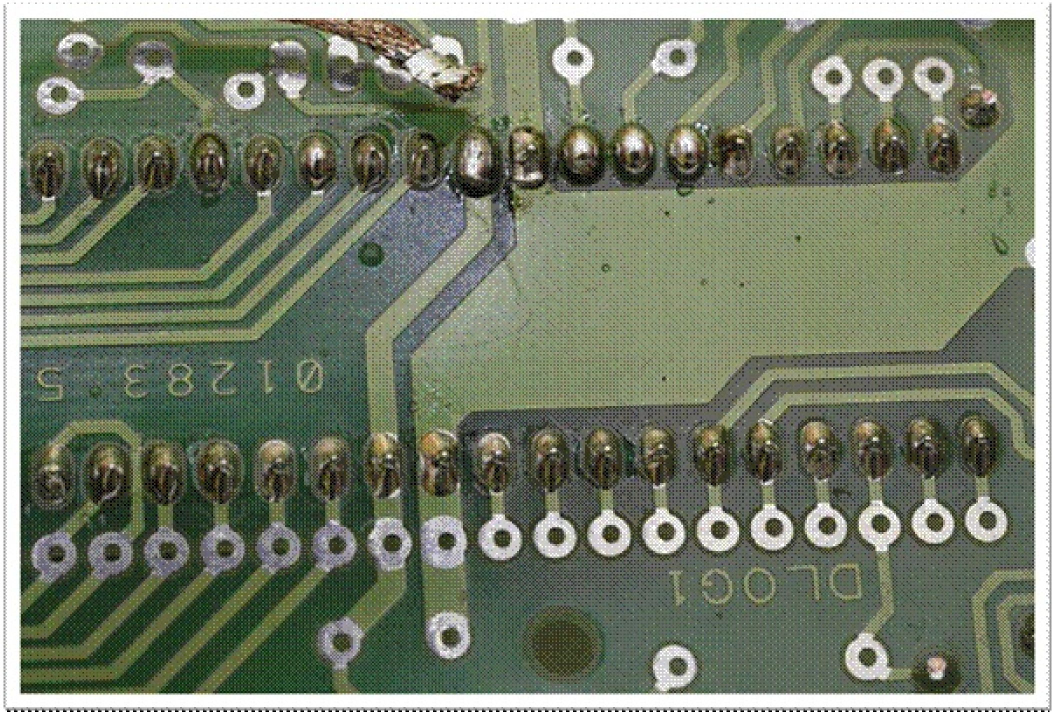

^^ *Remove the braid immediately and don't drag "whiskers" of molten solder around. The excess solder is absorbed by the braid, which is snipped off ready for the next job.*

Be aware that you can damage a printed circuit board accidentally when removing the desolder braid if it's not removed *quickly enough*. The solder will soon solidify, which effectively solders the braid to the printed circuit board! A careless tug may yank copper tracks or pads away from the board, stuck to the braid. You can also drag solder "whiskers" onto neighbouring pads unless the braid is removed cleanly. Why not practice using desolder braid with an old circuit board?

Whichever desoldering method you use, care is needed to ensure that the boards and parts are not damaged by excessive heat. It's not that difficult to apply so much heat when desoldering that the adhesive holding the copper foil tracks on the p.c.b. eventually fails, causing the copper track to lift away – everyone's worst nightmare.

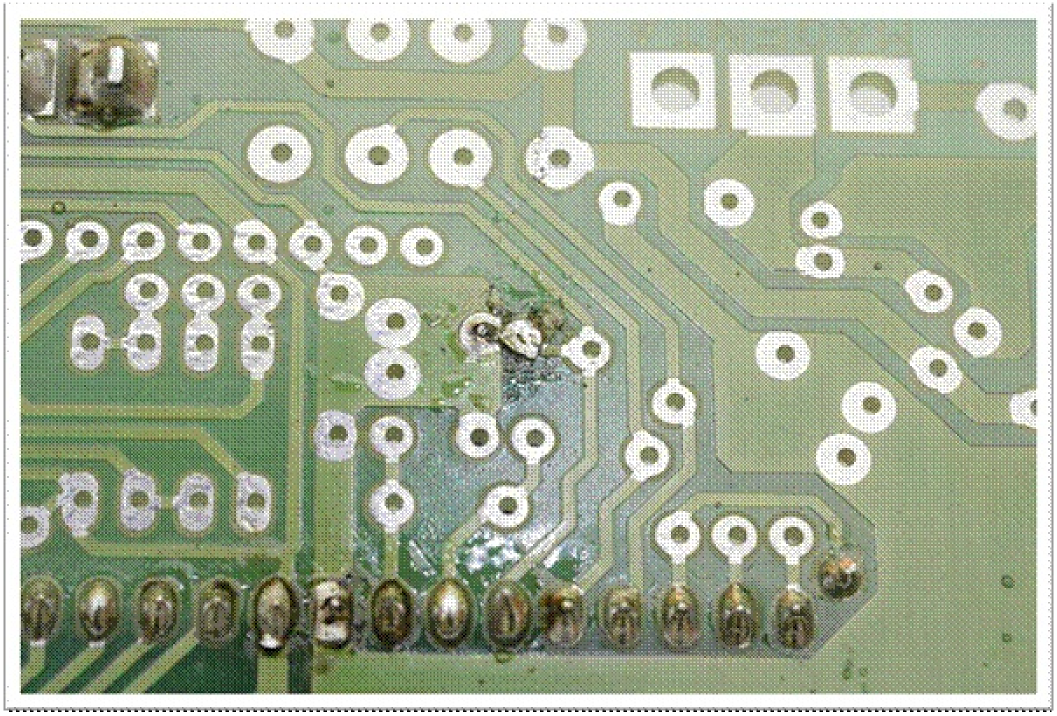

^^ Typical copper track damage (centre) caused by overheating during soldering or desoldering. The track has lifted off, but you can try repairing it by adding extra wiring or SuperGlue it if the track isn't broken.

If this should ever happen, remove the iron immediately and permit the area to cool (a freezer aerosol is valuable at such times). If you're lucky, you can maybe repair the lifted track using a droplet or two of Super Glue, or add "jumper wires" to bypass the damage.

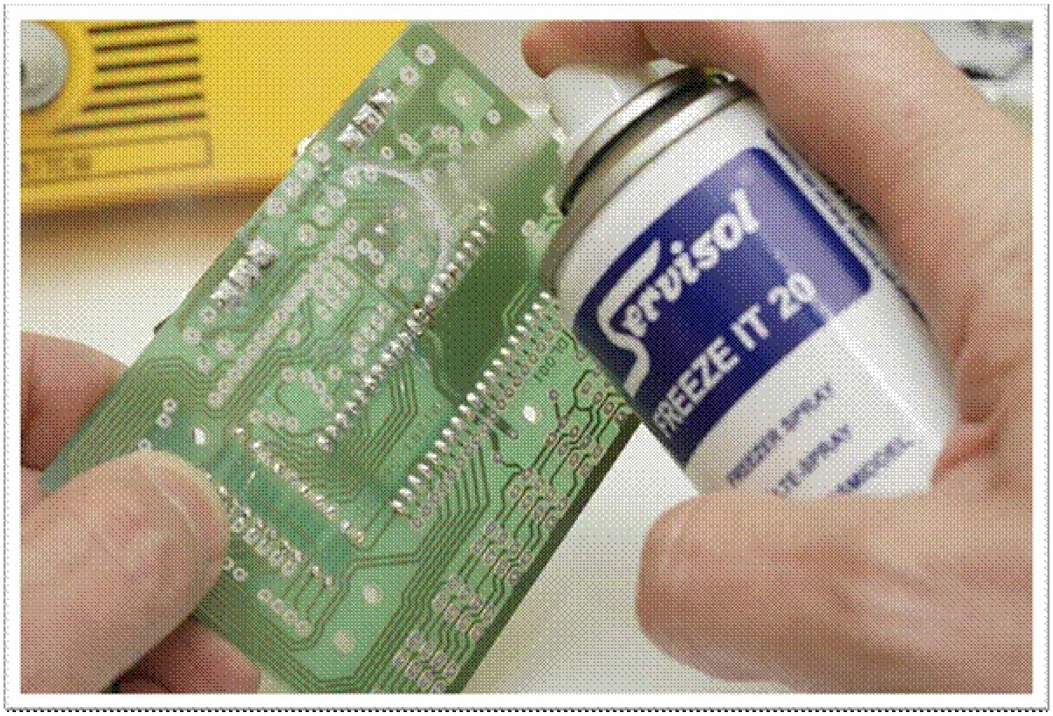

^^ *A Freezer aerosol can give rapid cooling where excess heat has been applied during soldering. Also used in circuit faultfinding to identify overheating parts.*

You now know everything you need to know about making the ideal solder joint, and desoldering it in case you need to make a repair. Just to remind you, a Quick Summary guide follows.

Quick Summary Guide

To round off the Basic Soldering Guide, let’s summarise how to make the perfect solder joint.

- Ensure materials to be soldered are compatible with tin/ lead or lead-free solder.
- All parts must be clean and free from dirt and contaminants.
- Try to secure the workpiece firmly during soldering.
- Brand new soldering iron tips must be flooded with solder immediately, the first time they are used.
- Wipe the tip of the hot soldering iron on a damp cellulose sponge at frequent intervals. Then “tin” the iron tip by applying a small amount of solder.
- Aim to heat all parts of the joint with the iron for under a second or so, to bring them up to the same temperature.
- Continue heating and apply sufficient rosin-core tin/ lead or lead-free solder to form a complete joint.
- It only takes a second at most, to solder the average p.c.b. joint. It should be smooth and shiny, and through-hole joints should be slightly convex in shape.
- Remove the iron and return it safely to its stand.
- Do not move parts until the solder has cooled.
- Tin the soldering iron tip and clean it well, when switching it off, ready for next time.
- Consider using e.g. electronics flux dispenser pens or Colophony (rosin) to help with difficult joints.

Sometimes solder joints don’t go quite to plan, and sooner or later everyone is faced with the need to problem-solve or troubleshoot, so a simple Troubleshooting Guide follows next.

Troubleshooting Guide

This troubleshooting guide may help fix common problems encountered with troublesome solder joints.

SYMPTOMS	LIKELY CAUSES	REMEDY
Solder won't "take" (wet) and won't flow properly over the joint -- molten solder forms beads or "ball bearings" instead of flowing properly.	Grease or contaminants present; Material may not be suitable for soldering with standard lead/tin or lead-free solder, e.g chromium.	Treat contaminated parts with abrasive cleaners etc. as required to expose base metal. Some metals can't be soldered with electronics-grade solder.
Solder doesn't melt or flow very well -- the joint is crystalline or grainy-looking - a grey or dry joint.	Joint has been moved before being allowed to cool naturally, or: Joint was not heated adequately. Too large a joint – too much metal present – and/ or the soldering iron temperature or power rating are too low.	Desolder and remake. Apply heat for a longer period, or use a higher power soldering iron, or check the temperature setting and raise it if possible.
Solder joint forms a "spike" and applying the iron again makes it even worse!	Probably overheated, burning away the flux. The iron, when removed, would cause the solder to stand up in a spike.	It is usually best to desolder and remake the joint freshly again.
The copper foil of my p.c.b. has lifted off the circuit board!	Excessive use of heat has damaged the adhesive. Provided the track hasn't broken, it may be repairable.	You can sometimes repair it with Super Glue, or re-wire the board with jumper wires.
Brown varnish-like deposits are left behind after I finish soldering.	These are the remains of rosin flux and are nothing to worry about.	It can be removed with PCB cleaners or some solvents, if you want to tidy up the board and inspect your work.

Possible Hazards and simple First Aid

It's extremely rare that soldering iron operators receive any burns or other injuries from the use of hot soldering irons. Soldering is perfectly safe provided that common sense precautions are taken during the soldering operation. Here are some of them:

- Components are very hot after soldering, so let them cool before handling them to avoid skin burns.
- Beware of splashes of molten solder caused by careless handling of a hot soldering iron.
- Beware of energised components (capacitors, batteries etc.) being shorted by molten solder and ejecting solder splashes due to arcing.
- Always park a hot iron safely on a stand in between use — never hang it vertically next to the bench.
- Keep a hot soldering iron away from its mains cable (silicone cables reduce the risk of accidental damage).
- Beware of wire offcuts flying off (danger to eyesight) when snipping wires to length before or after soldering.
- Avoid inhalation of solder and flux fumes as this can irritate the respiratory tracts, especially in sensitive cases (e.g. asthma).

Should you receive a more serious skin burn which requires attention, then:

- **Cool the affected area immediately.** Use plenty of cooler running water – but avoid ice cubes etc. as they can cause nerve damage after a time or inhibit the flow of blood to the affected area.
- Remove any **objects** which may be constrictive, before any swelling starts (rings, watches, bracelets).
- Do not prick blisters nor apply ointments, salves or lotions at this stage.
- Local pain relief for small burns can be obtained by spraying *Burneze* aerosol onto unbroken skin.
- Seek medical attention for more serious burns.

Eyesight problems are exceptionally rare, e.g. pieces of wire offcuts or solder splashes lodging in the eye area, and should be treated by a qualified first-aider or A&E. The best you can do is bathe the affected area with e.g. a first-aid eyewash bottle or fresh water. Then seek professional medical help straight away.

Useful resources

Details of the Antex range of soldering equipment, solder tips and spare parts can be obtained from www.antex.co.uk. UK and international distributors are also listed on the Antex (Electronics) web site.

If you're interested in hobby electronics then why not try **EPE Magazine**, the No. 1 UK magazine for hobby electronics enthusiasts, students, trainees and technicians around the world. You can buy a printed edition from newsstands, download a PDF version (for Windows) or try a tablet/ smartphone version from Pocketmags. More details at www.epemag.com

You can learn more about the writer at his website http://www.alanwinstanley.com/

Some mail order suppliers of soldering equipment and electronic parts include:

- ESR Electronic Components Ltd. (soldering equipment, Velleman kits)
- Bowood Electronics (Antex irons and spares, electronic parts)
- Rapid Electronics (components, tools, equipment)
- Cricklewood Electronics (CCTV, Antex soldering irons and spares, components)
- Maplin Electronics (UK electronics retailer and mail order)
- Farnell Electronic Components (major UK industrial supplier)
- RS (UK electronics industrial supplier)
- Brewsters Ltd (soldering equipment mail order specialists)
- Velleman UK (electronic kits)
- Quasar Electronics (Velleman kit mail order retailers)
- Kemo Kits (Germany, trade only)
- Hobbytronics (UK hobby mail order supplier)
- Multicore Solders (now a Henkel brand)

These links were correct at the time of going to press, July 2013.

Acknowledgements

Antex (Electronics) UK generously provided samples of their ever popular British-made soldering equipment, materials and accessories for use in the Basic Soldering Guide. When starting out as a 1970’s teenage schoolboy electronics hobbyist, I and my trusty Antex soldering iron shared many adventures in electronics together, and it has been a pleasure to use the latest Antex soldering irons throughout this guide.

Also I would like to thank Brian Brooks of Magenta Electronics Ltd. (www.magenta2000.co.uk) for supplying me with a variety of their professionally-designed printed circuit boards and a range of components which were the subjects used in the photographs.

Photography

All photographs were taken by the author using Sony Alpha DSLRs and 50mm macro lens with extension tubes, and a Sony macro flash. High-resolution versions of these and more technical photos are available for commercial or educational royalty-free use. Please contact me with any enquiries.

Conclusion

I really hope the Basic Soldering Guide will give you the confidence to try your hand at electronic soldering. It’s really a lot easier than it sounds, and armed with the guide’s advice and photographs, the next thing to do is invest in a decent-quality soldering iron such as the excellent range manufactured by Antex that will serve you well for years to come.

Soldering an electronics kit together such as the professional designs produced by Velleman is a great way of testing out your new skills. Start with a small, simple kit costing a few pounds to gain confidence and experience, and avoid the temptation to tackle something too complex until you’re ready to extend yourself further.

Feedback is welcomed by email to alan@epemag.demon.co.uk

You can learn more about me at www.alanwinstanley.com or read my column in EPE Magazine.

Good luck with your soldering!

Alan Winstanley

Text and Photographs © Copyright © Alan Winstanley 1997-2013

Play fair and support writers who share their skills and expertise

Preparing a resource like this takes months and is based on four decades of practical experience which I gladly share with readers. It costs a small fortune in time, experience, IT and photographic equipment to put an ebook like this together, and the Royalties that you pay to buy the book puts food directly on my table. You paid for your copy (thank you!) so if you see any unauthorised versions of my work anywhere in circulation, I'd really appreciate being told about it. Thank you! A.R.W.

EPE EVERYDAY PRACTICAL ELECTRONICS
Interested in electronics?
Read EPE - the No. 1 UK hobby electronics magazine.
Ask your newsagent, download or subscribe now!
www.epemag.com
EPE EVERYDAY PRACTICAL ELECTRONICS
DRIVEWAY SENTRY

BSG/ ARW/V1.1/0813

energysage

Don't overpay for solar

Compare offers & save 20% or more

Compare Offers

advertisement

The Poor Man's Soldering Iron

By Rubbish Raider (/member/Rubbish+Raider/) in Technology (/technology/) > Soldering (/technology/soldering/) 31,868 254 30 Featured

BY-NC-SA Download Favorite

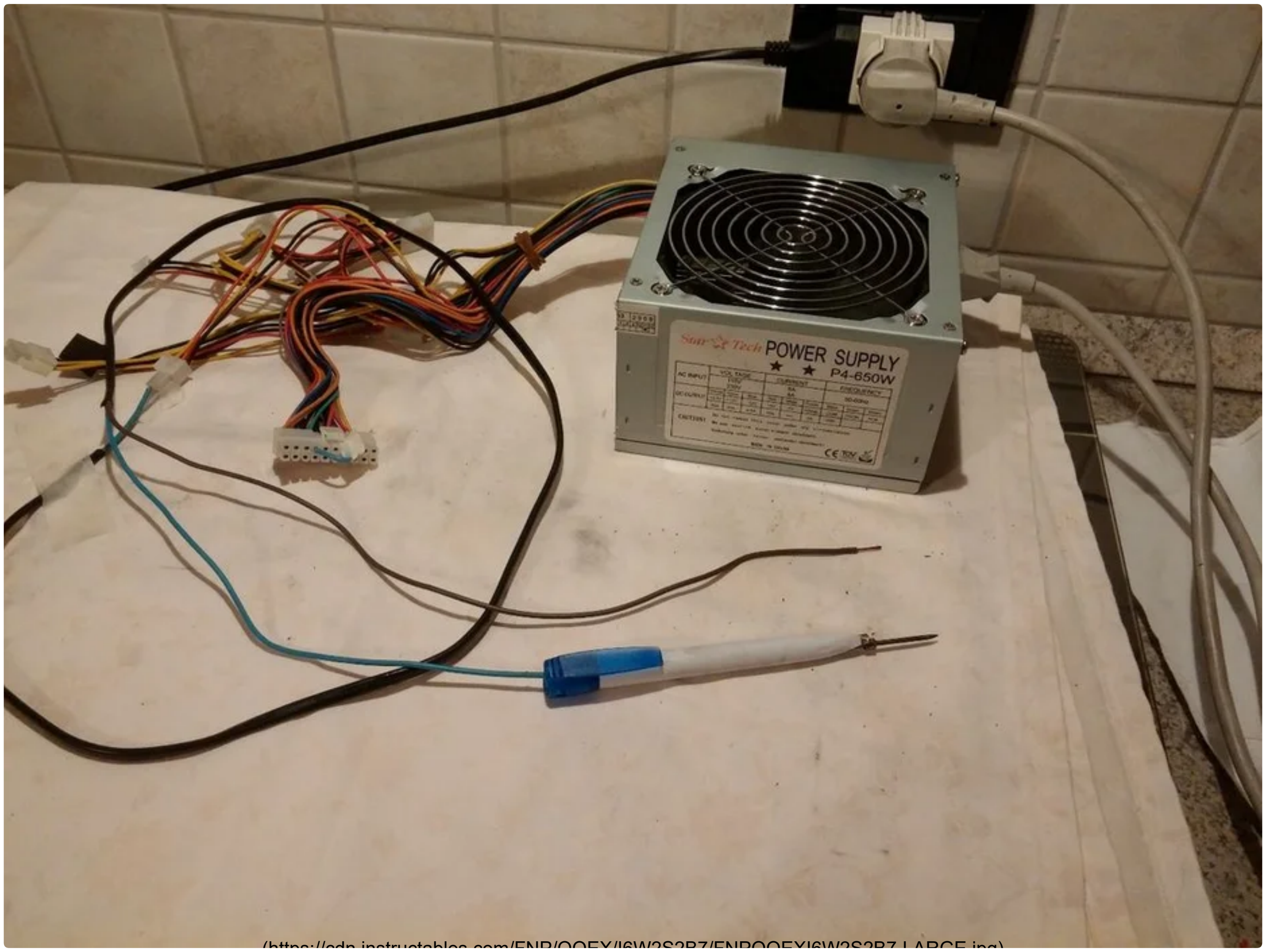

(https://cdn.instructables.com/FNP/OOEX/I6W2S2BZ/FNPOOEXI6W2S2BZ.LARGE.jpg)

Welcome to my first instructable! As said in the title, I will try to teach you how to build a simple soldering iron at little to no expense. "Why should I build a soldering iron instead of buying one?" Apart from the little cost of the project (most people can manage to assemble their soldering iron for free, or almost free) this instructable is for those who like to understand how stuff work, who like to build their tools, and for those (like me) who just broke their soldering iron and needed a quick solution to finish their Arduino project.

The idea came from the "DIY Cold Heat soldering iron" instructable by photozz, for which I thank him very very much. Unfortunately, I didn't have all the materials needed for a Cold Heat style iron, so I decided to modify it, and I ended up with another design.

WARNING!

This project involves dealing with electricity and heat. The fact that you are not going to work with thousands of volts does not mean that you won't hurt yourself.

P.S.: English is not my first language, so don't hesitate to tell me if I say something wrong. And don't hesitate to tell me if you have any idea to improve this project. "If you have an idea and I have an idea and we exchange these ideas, then each of us will have two ideas", as G. B. Shaw said.

Finally... let's start.

Add Tip Ask Question Comment Download

ebay
Here's $5* to Shop Again
*Terms apply. No min. spend.
Shop Now →
advertisement

Step 1: What Do I Need? and What Do I Need to Know?

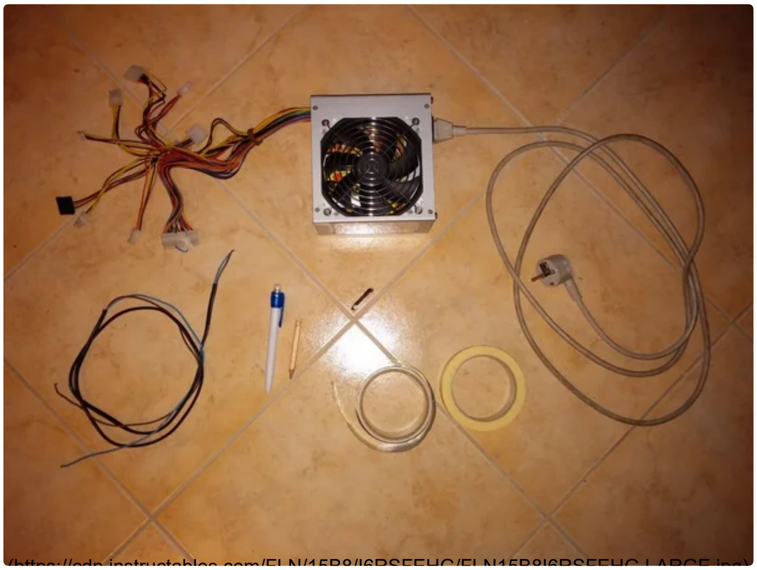
(https://cdn.instructables.com/FLN/15B8/I6RSFFHC/FLN15B8I6RSFFHC.LARGE.jpg)

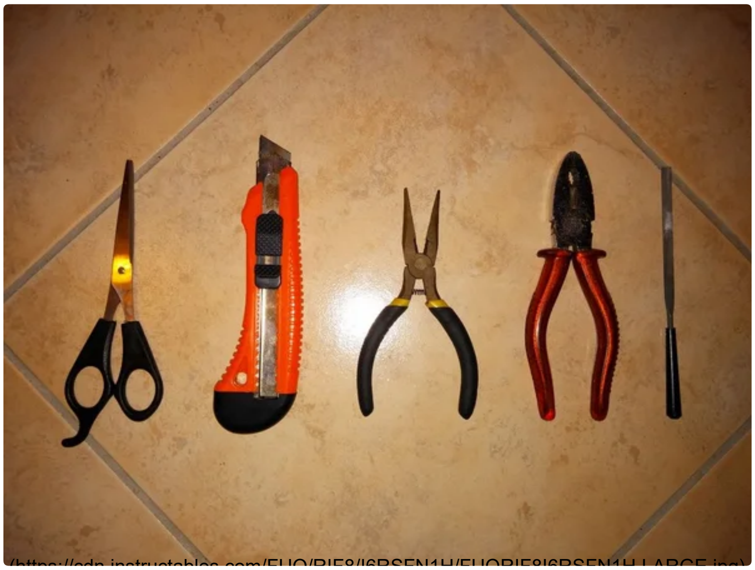
(https://cdn.instructables.com/FUO/BIF8/I6RSFN1H/FUOBIF8I6RSFN1H.LARGE.jpg)

Bill of materials:

- electrical wires, like those used to connect appliances to wall plugs (you'll need two separate pieces of wire)

- an easily bendable piece of metal (can be really anything... it will become clearer later)

- heat resistant adhesive tape (not strictly necessary)

- "normal" adhesive tape (I used paper tape)

- a PSU (Power Supply Unit, there is one in all desktop computers)

- a pencil

- a pen

Handtools:

- a pair of scissors

- an utility knife

- pliers

- a file (or sandpaper, but a file works better in my opinion)

Skills required:

- a bit of creativity

- you should know how to strip electrical wire

This project is quite easy to build, and won't take much time to complete (I spent less than an hour to build it, once I had the materials). However, I understand that some people could find difficulties in gathering some of the materials (ehm... who gave you that weird box with hundreds of wires? NASA?), if they are not used to dumpster dive or don't have a heap of electronic rubbish lying around in their bedroom. Most probably you already have in your home an old pc waiting only for you to dismember it and extract the PSU! You have just thrown it away? Oh... Well, dumpster dive, if you haven't already done it! One man's rubbish... Or ask friends, relatives, classmates (if you have any). Surely one of them has an old computer no longer needed, or knows where you can find what you are searching for. In the unlikely event that, after trying all of these methods, you still are without a PSU, you could ask an electronic retailer. Usually they have many spare parts, and will probably be happy to get rid of a PSU (if asking money, they shouldn't be greedy at all).

Add Tip Ask Question Comment Download

Step 2: Working Principle

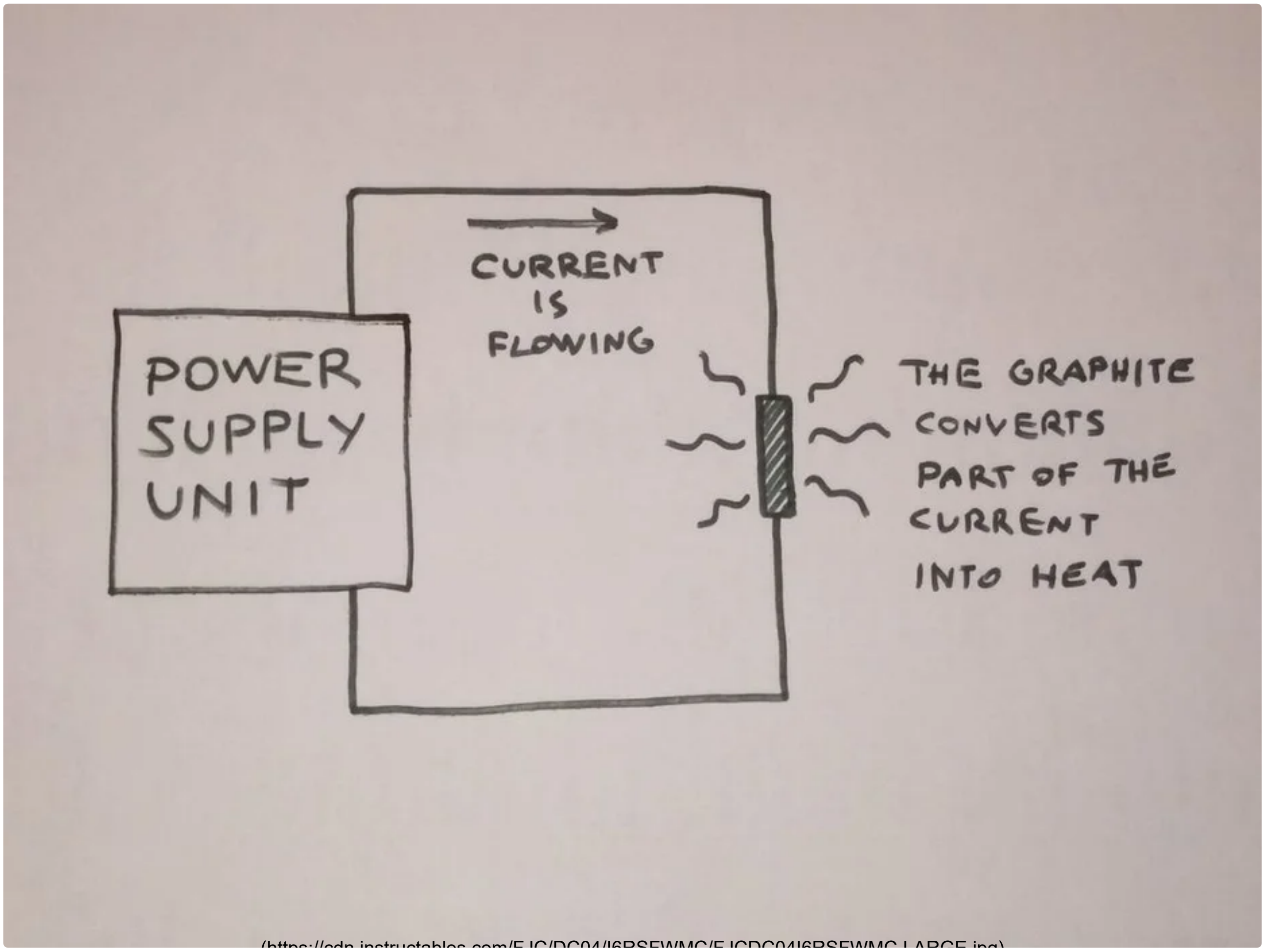

(https://cdn.instructables.com/FJC/DC04/I6RSEWMG/FJCDC04I6RSEWMG.LARGE.jpg)

If you don't care of this part or already know everything, you can just skip to the next step.

Now... a soldering iron generates heat by making an electrical current pass through a resistance. That's it.

Wait, what is a resistance?

When a current passes through any material, the atoms of this material behave like an obstacle to the electron flow. There is what we can consider as a sort of friction. This friction or resistance, like other kinds of friction, generates heat. So part of the current flowing in every circuit is converted into heat, depending on the resistance of the material it's flowing through (generally speaking, metals have a low resistance while materials like glass or plastic have a high resistance).

This principle is widely used in light bulbs, electric ovens and every kind of electric heaters in general.

To make a circuit of this kind, all we need is something that generates current and something that converts part of it into heat.

At this point, some basic formulas could help understanding better:

Current flowing in a circuit = Voltage of the generator(s) / Total resistance of the circuit or $i = V / R$

This formula lets you calculate how much current will flow through a circuit. Let's make an example: we have an AA battery, which outputs 1.5 volts (this is the unit of voltage), and we connect the two poles with a metal wire with a resistance of 1 Ohm (this is the unit of resistance). We don't consider the resistance of the materials of the battery, which is usually very low. The current flowing through this circuit will be approximately 1.5 V / 1 Ohm = 1.5 Amperes (the unit of current). It is important to notice that if we want to increase the current flowing in our circuit, we have to reduce the resistance in that circuit or increase the voltage of the generator.

And here is another formula:

Power dissipated by a material which has a resistance = Voltage applied to this material * Current flowing through it

or $P = V * i$

This formula lets you figure out more or less how much heat will a piece of any material emit (power dissipated is directly proportional to heat generated). Let's calculate how much power is wasting our metal wire connected to the battery:

1.5 V * 1.5 A = 2.25 W (watts, the unit of power)

In order to increase the power wasted and the heat generated, we have to increase the voltage or the current, or both.

Ok, perhaps I'm only frying your brain with not-so-useful math.

As a power supply we are going to use, well, the Power Supply Unit. And as a resistant material (which will be our tip) we are going to use graphite. As outlined by photozz in his instructable, graphite is a very good material to be used as a soldering iron tip, because it has a low resistance, but not too low a resistance, it decomposes at very very high temperatures (Wikipedia says about 3500 °C), it is very easy to find (ehm... pencils) and very easy to clean as well (the solder doesn't stick to graphite). And it seems to return cold in a very short time if compared to a metal tip.

Our circuit will look like what I drew in the picture. We will connect a wire to the PSU and an end of the tip, and another wire to the PSU and the other end of the tip, in order to make a current flow through it.

Step 3: Making a Tip

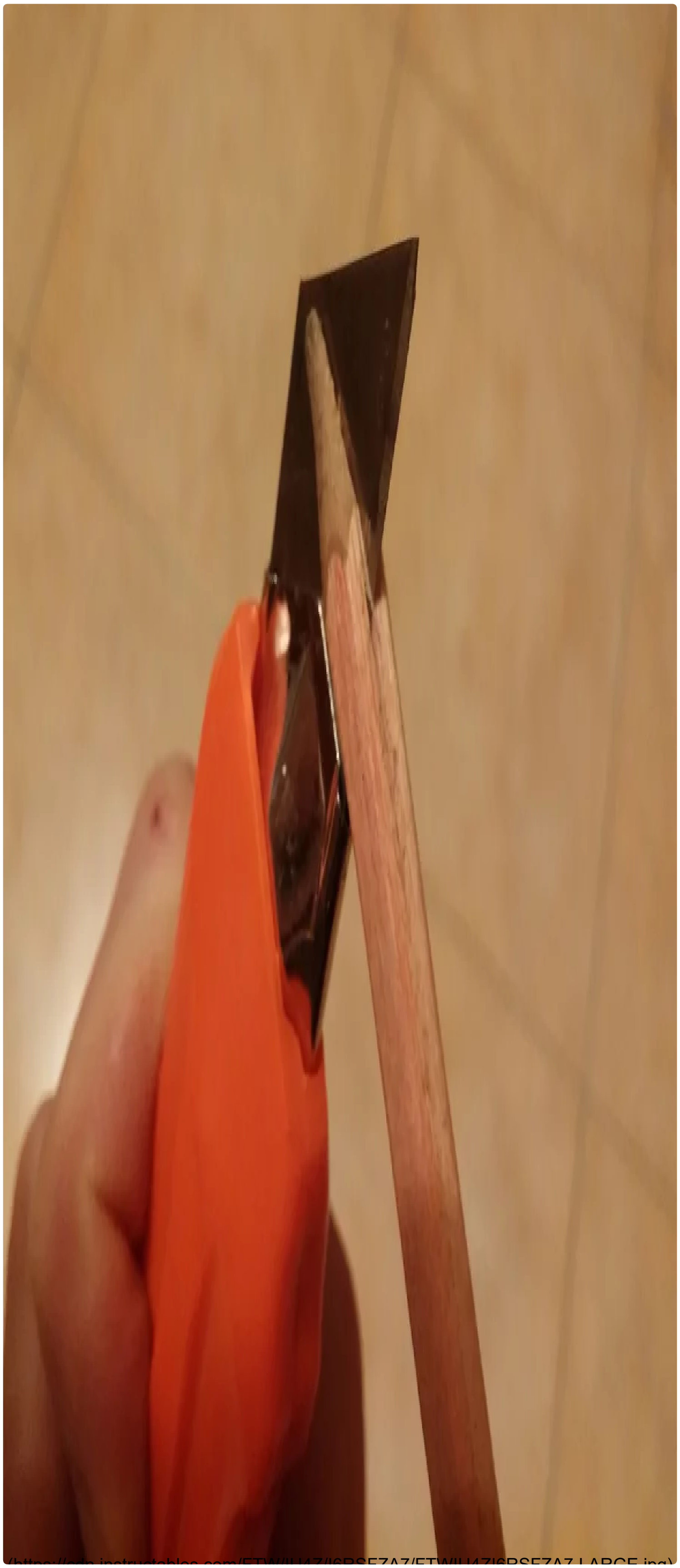
(https://cdn.instructables.com/FTW/IU4Z/I6RSFZAZ/FTWIU4ZI6RSFZAZ.LARGE.jpg)

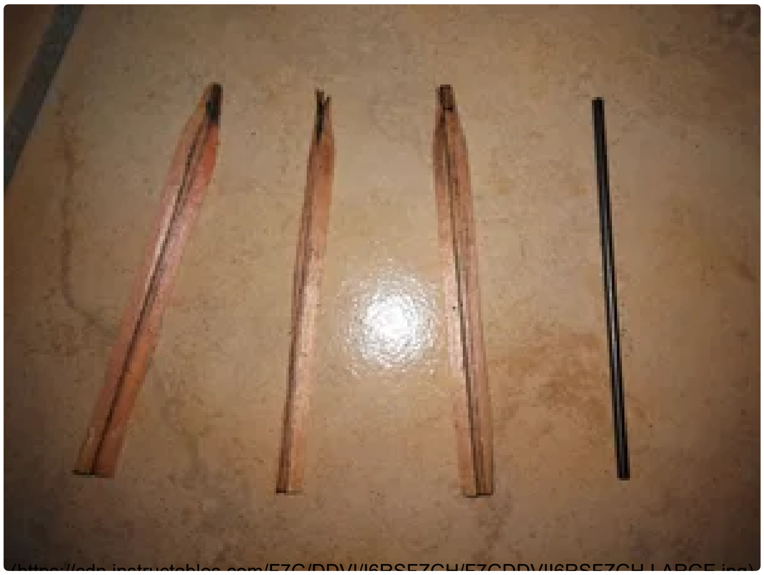
(https://cdn.instructables.com/FZG/DDVI/I6RSFZGH/FZGDDVII6RSFZGH.LARGE.jpg)

(https://cdn.instructables.com/FC9/HBKK/I6W2QB4V/FC9HBKKI6W2QB4V.LARGE.jpg)

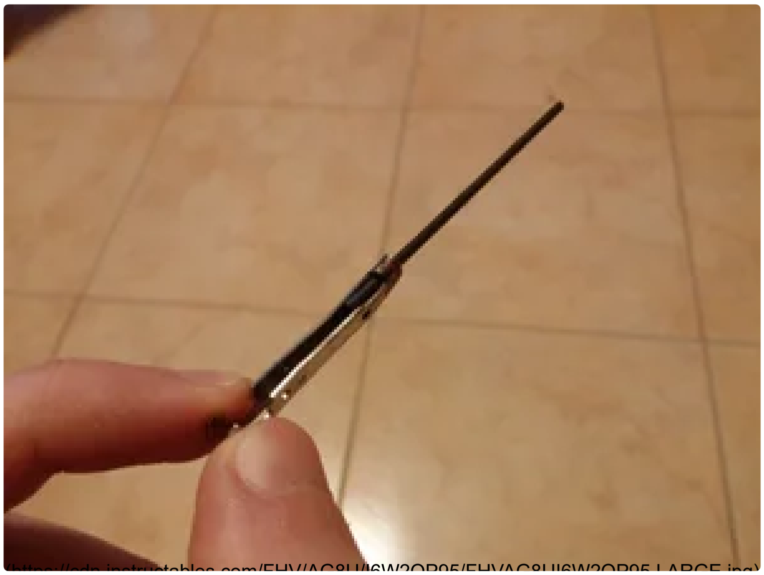
(https://cdn.instructables.com/FHV/AC8U/I6W2QP95/FHVAC8UI6W2QP95.LARGE.jpg)

First, take the pencil.
Second, take the utility knife.
Third, cut the pencil as shown in the first picture. Repeat the procedure on the other sides of the pencil until the graphite comes out of the wood. It is quite important that the graphite

doesn't come out in dozens of small pieces, so be careful.

Now we have the graphite. But how can we hold it?

Here comes in handy the generic piece of metal. It could be a (preferably thick) wire, or a plate, or a strange piece of odd shape (like the one I used... it was the core of an electromagnet, found in one of those cheap wall clocks). The only requirements are that it can be bent to hold the graphite steadily, and has an end that you can wrap in electrical wire. Use your fantasy.

Add Tip Ask Question Comment Download

Step 4: The Tip Holder

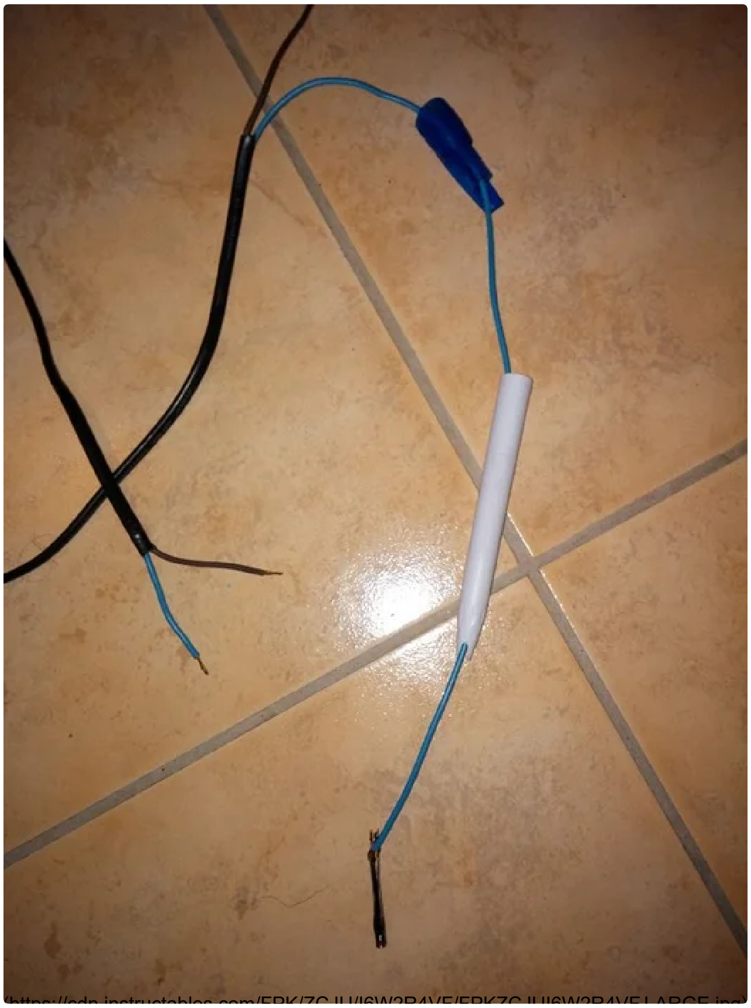

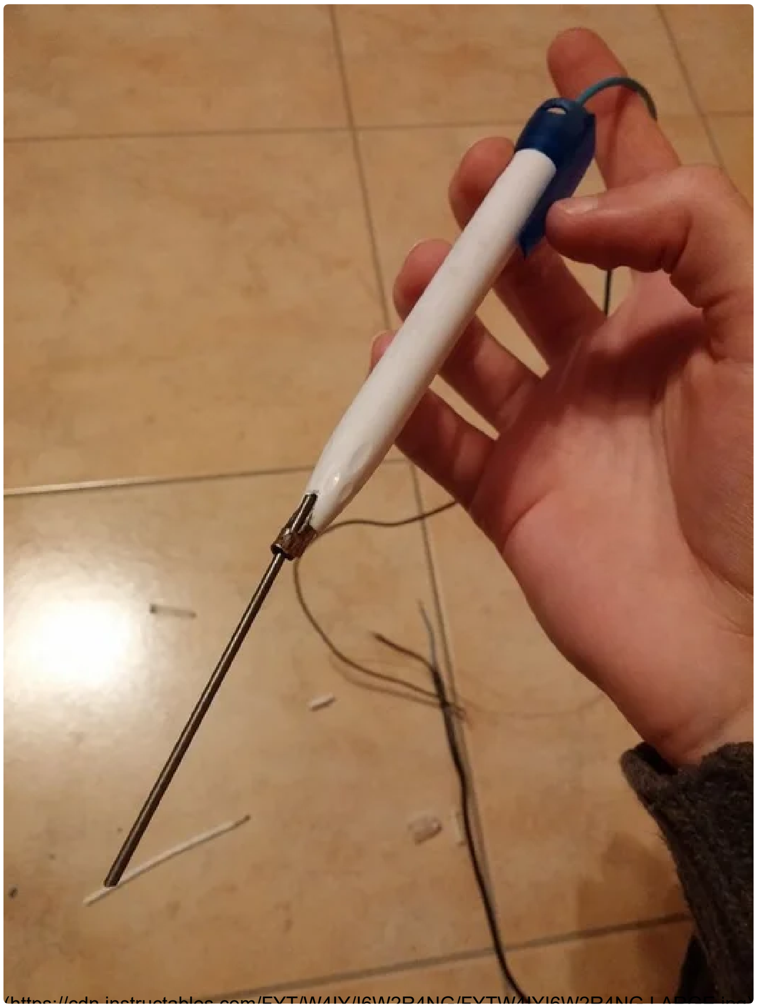

At that moment I didn't have any alternatives, so I used a pen to hold the tip, and discovered that it works quite well (however, some plastics could melt or emit toxic fumes when heated too much. Pay attention to this).
First, remove the internal parts (spring, cartridge, little gear-like components) of the pen. We don't need them.
Second, remove the cap. You will have to disassemble, cut or break it, depending on the pen you are using. Bear in mind that through this hole a wire will pass.
Third, give a careful look at the piece of metal with the graphite. You should put it where the

tip of the pen was. Use the file or cut the pen in whatever shape you need, the important thing is that the piece of metal can be inserted where the tip of the pen was.
Done it? Not yet? Maybe you are going to spend a lot of time thinking about the right way to do it.

When you are done, simply make an electrical wire pass through the pen, wrap it around an end of the piece of metal and make sure they won't disconnect (the best thing would be to solder them, but I assume you don't have a working soldering iron). Now insert the piece of metal into the pen... et voilà! Here is your tip holder. You could wrap the tip in heat resistant tape, but it isn't compulsory to do so.

And yes, my piece of graphite is too long, but we are going to cut it later.

Add Tip Ask Question Comment Download

Step 5: Connecting Everything to the PSU

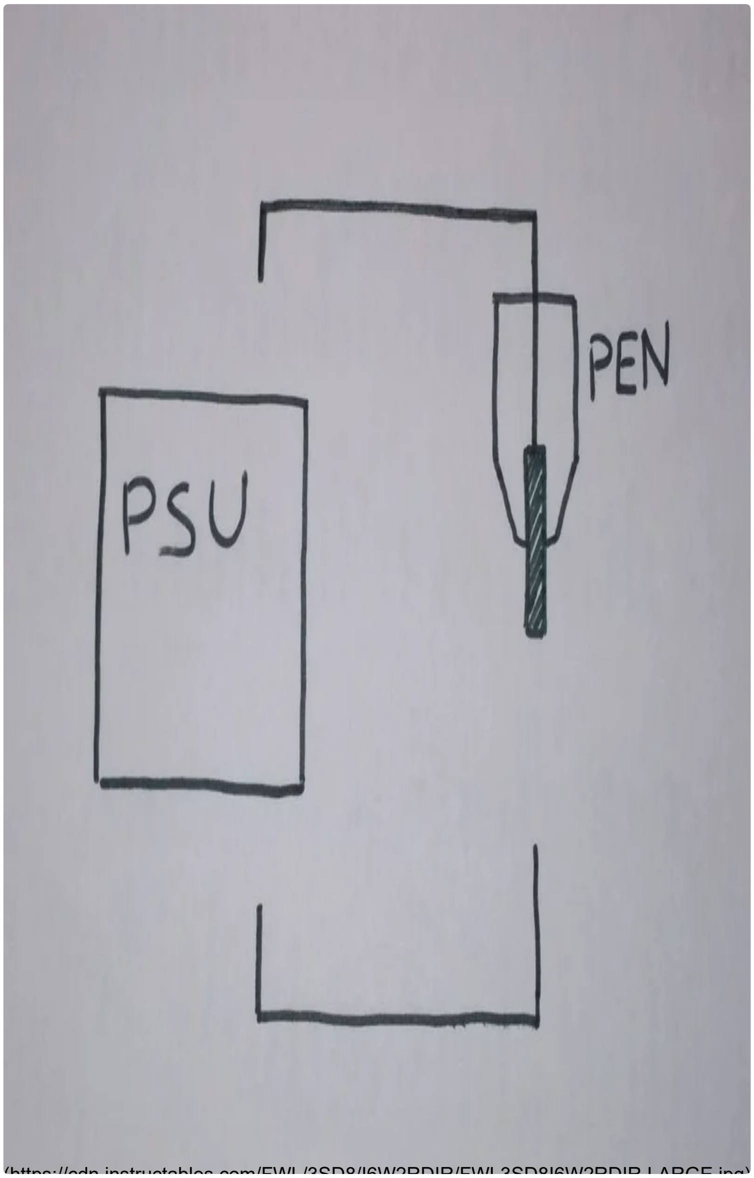

(https://cdn.instructables.com/FWL/3SD8/I6W2BDIP/FWL3SD8I6W2BDIP.LARGE.jpg)

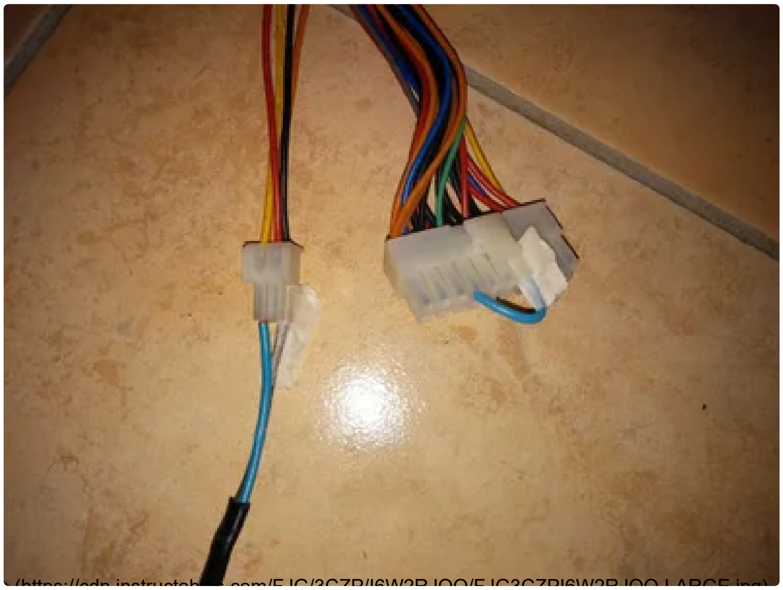

(https://cdn.instructables.com/FJC/3CZP/I6W2BJOQ/FJC3CZPI6W2BJOQ.LARGE.jpg)

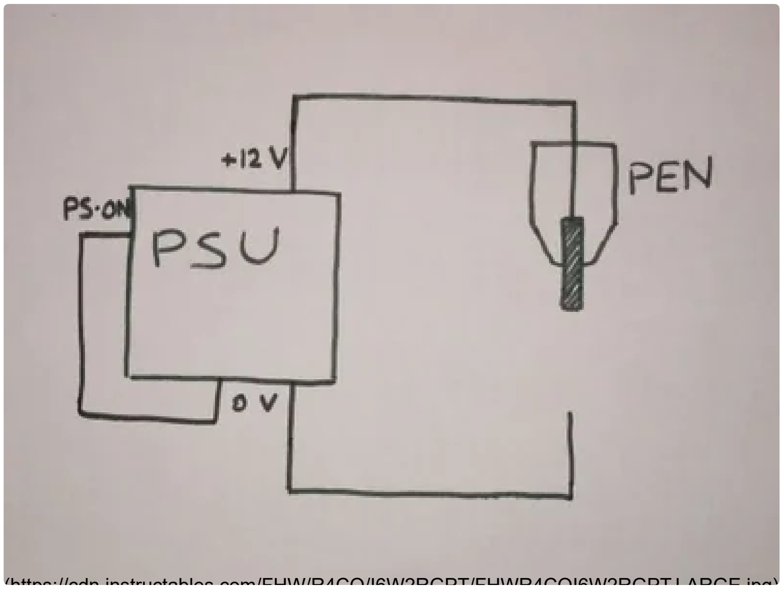

(https://cdn.instructables.com/FHW/R4CQ/I6W2BCPT/FHWR4CQI6W2BCPT.LARGE.jpg)

Tired? The good news are that the difficult part has just finished.

At this point, we should be in the situation shown in the first picture. Now all we need to do is wire the tip to the PSU, which isn't complicated at all, in spite of the hundreds of wires coming out of that box.

On one side of the PSU there should be a legend with all the wire colours explained. Usually, the rail which outputs 12 volts is yellow, 5 volts is red and 3.3 volts is orange, but it isn't always so. There are other colours and rails, but we don't need them. And there is GND, or ground, which is 0 volts. This is important because we will connect a wire to it.

Now look at the amperage (the maximum current they can output). Most PSUs have sufficient amperage to allow the passage of very high currents through your tip, so you can choose whichever rail you want (if, when all is done, your tip doesn't heat, you can simply move to another rail with a higher voltage, remember the first formula? In order to increase the current you have to increase the voltage or decrease the resistance, which means to chop your graphite, because with a shorter piece the resistance will be lower). On my PSU I use a 12 V rail that can output 30 A.

Returning to the PSU wiring, once you have chosen your rail you will have to connect the wire attached to the pen to that rail (I chose 12 V, and so I connected it to a yellow wire), the other wire to ground (leave the other end of this wire free, I will soon explain why), and wrap them in tape, or don't, but wrapping them will keep them more firmly connected (I wrapped only the wire connected to GND, because when I need different voltages I can change rail without unwrapping the other wire).

Certain PSUs won't turn on when plugged in. They need another wire to be connected. Controlling on the legend, you should be able to find a rail named "PS-ON", usually of a weird color like light green. That wire needs to be connected to GND. You can do it with a jumper, or you can put a switch to shut down the PSU without unplugging it. Personally, I prefer to use a jumper and unplug it. And wrap at least one end in tape, in order to keep it firm.

The final situation is shown in the last picture.

Add Tip Ask Question Comment Download

Step 6: Let There Be Light!

POWER SUPPLY
P4-650W

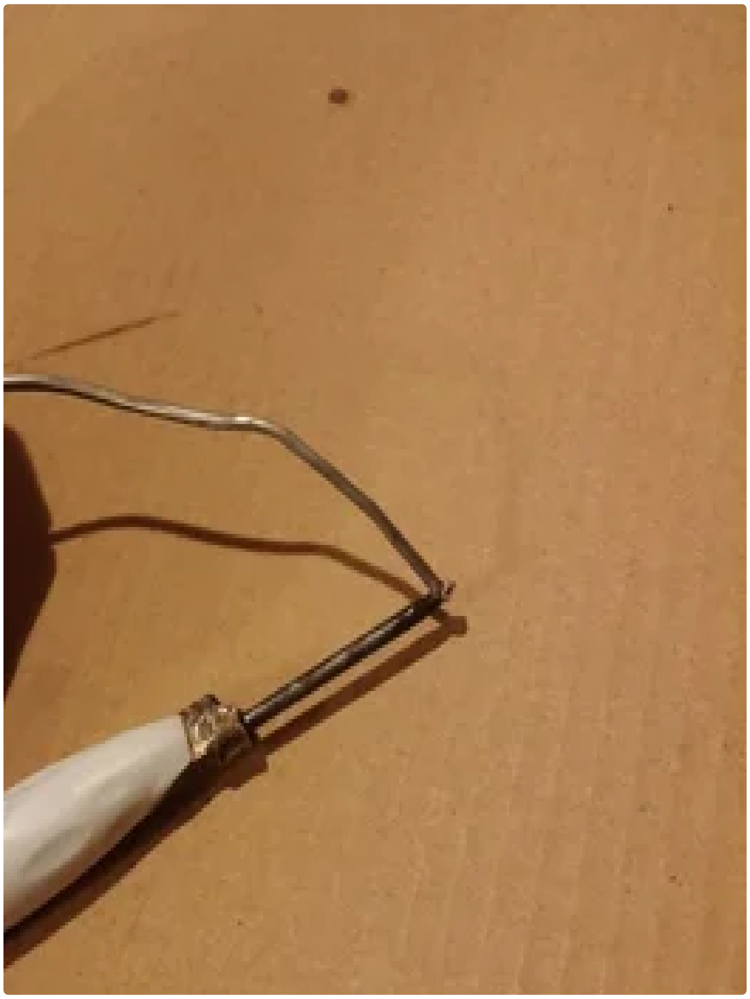

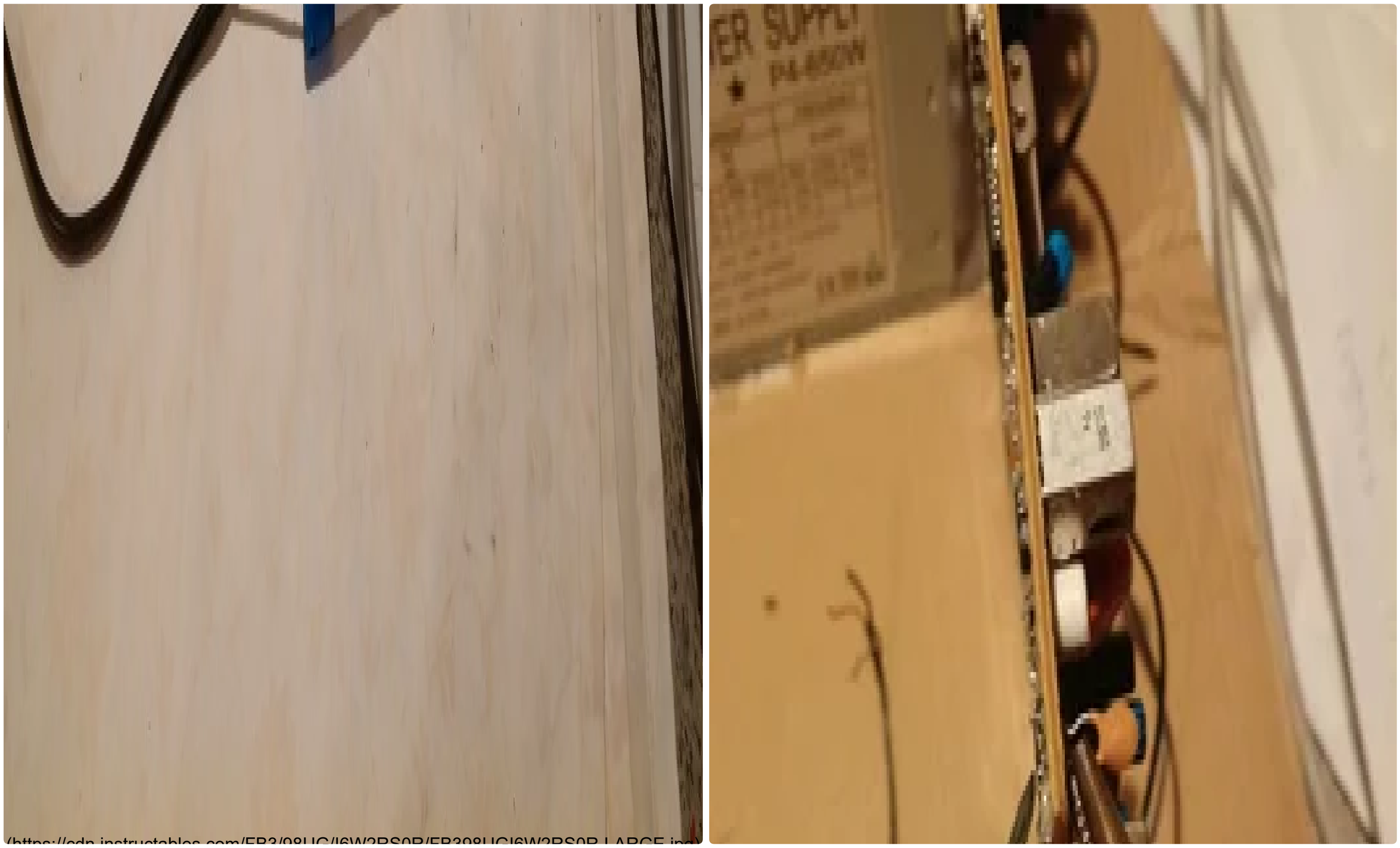

(https://cdn.instructables.com/FB3/98UC/I6W2BS0R/FB398UCI6W2BS0R.LARGE.jpg)

Now your soldering iron is completed!

To heat the tip, simply touch the graphite with the free end of the wire connected to GND (that's why we have left it free). It will heat as long as you keep it in contact with the wire.

You could think this is a drawback, but attach this wire to the solder... and you have a functional iron that stays hot only when the solder is touching the graphite tip (graphite returns cold in seconds). Safety first!

Or you could attach the wire to a screwdriver, or a pair of pliers, and have a rude desoldering iron (pull a pin with the pliers while touching with the graphite the point where it is soldered).

Another big advantage of using graphite is that it can be shaped quite easily: you can make a flat tip, a very sharp tip... the possibilities are endless!

Add Tip Ask Question Comment Download

Step 7: Some Final Caveats

1. Burning in (pun intended!)

The first time you use your soldering iron bear in mind that, together with the graphite, there is also a certain amount of glue. Make it heat gradually, because you will see many strange things happen... the graphite could start to emit fumes (that is the glue evaporating), or bubbles could appear on it (that is the glue boiling), or it could crack (that's due to little bubbles of glue trapped inside, that expand very quickly and break the graphite).

2. Don't make it glow

If you touch your tip with your wire for more than a dozen of seconds (but it depends on the piece of graphite you use, the rail, etc.) it should start to emit light. Never arrive at this point, as the plastic of the pen could emit toxic fumes, your graphite could deform, and the stuff you are working on will be destroyed for sure.

3. There's pencil and pencil

From my personal experience, I have noticed that some kinds of graphite are more conductive than others. High quality pencils like Staedtler (yes, I am so crazy that I dismembered a Staedtler pencil) are way more conductive than low quality ones (perhaps due to a minor amount of glue?). Don't forget this when changing the tip. A new piece of graphite coming from another pencil may heat very quickly, compared to the previous one.

And I warn you against very low quality pencils: they tend to explode as soon as you heat them. Be careful and heat them very very slowly, in order to let the tonnes of glue they put in evaporate gently. But sometimes they will explode anyway.

4. I have seen a big spark and my PSU has stopped working!

One thing you should know is that every PSU has an overcurrent protection, that shuts down the PSU if on a rail is detected a current exceeding the max amperage of that rail. This is only a temporary protection: after some tens of seconds the PSU will return working (however, you could need to unplug and replug it).

If this happens as soon as you touch the graphite with the wire/solder/tool, the resistance of that piece of graphite is too low. Move to a rail with a higher amperage, if possible, or use a longer piece of graphite.

5. I tried to solder a component on the board of my €1500 amplifier and it doesn't work anymore...

Well, if you do such works, it is quite sure you already know, but this kind of soldering iron is not suitable for delicate operations. If you accidentally touch a component on a board with the tip, and another part of the board with the wire, current will try to flow through the board, frying any component that has not been designed to work with such high currents or voltages. You have to be very careful to what you touch, and if you mess around with €1500 amplifiers you should buy a decent soldering station.

My instructable finishes here. I hope I have been clear enough to be understood, if not don't hesitate to tell me.

See you to the next project!

Add Tip Ask Question Comment Download

energysage
Don't overpay for solar
Compare offers & save 20% or more
Compare Offers ❯
advertisement

The Universal Book of Mathematics

From **A**bracadabra to **Z**eno's Paradoxes

David Darling

WILEY
John Wiley & Sons, Inc.

The Universal Book of Mathematics

The Universal Book of Mathematics

From **A**bracadabra to **Z**eno's Paradoxes

David Darling

WILEY
John Wiley & Sons, Inc.

This book is printed on acid-free paper. ♾

Copyright © 2004 by David Darling. All rights reserved

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008.

Limit of Liability/Disclaimer of Warranty: While the publisher and the author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor the author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information about our other products and services, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Darling, David J.
The universal book of mathematics : from abracadabra to Zeno's paradoxes / David Darling.
p. cm.
Includes bibliographical references and index.
ISBN 0-471-27047-4 (cloth: alk. paper)
1. Mathematics–Encyclopedias. I. Title.
QA5 .D27 2004
510'.3–dc22 2003024670
Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Mathematics is not a careful march down a well-cleared highway, but a journey into a strange wilderness, where the explorers often get lost. Rigor should be a signal to the historian that the maps have been made, and the real explorers have gone elsewhere.

–William S. Anglin

But leaving those of the Body, I shall proceed to such Recreation as adorn the Mind; of which those of the Mathematicks are inferior to none.

–William Leybourn (1626–1700)

The last thing one knows when writing a book is what to put first.

–Blaise Pascal (1623–1662)

Contents

Acknowledgments ix

Introduction 1

Mathematics Entries A to Z 3

References 359

Solutions to Puzzles 369

Category Index 373

Acknowledgments

Many people have helped me enormously in assembling this collection of mathematical oddities, delights, whimsies, and profundities. Thanks especially go to Jan Wassenaar (www.2dcurves.com) for drawing many of the plane curves that are featured in the book; Robert Webb (www.software3d.com) for numerous photos of his wonderful, homemade polyhedra; Jos Leys (www.josleys.com) for his mesmerizing fractal artwork; Xah Lee (www.xahlee.org) for a variety of ingenious digital imagery; Sue and Brian Young at Mr. Puzzle Australia (www.mrpuzzle.com.au) and Kate and Dick Jones at Kadon Enterprises (www.gamepuzzles.com) for their advice and photos of puzzles from their product lines and personal collections; Gideon Weisz (www.gideonweisz.com) and Istvan Orosz for stunning recursive and anamorphic art images; my good friend Andrew "Dogs" Barker for stimulating discussions and the solution to one of the problems; William Waite for pictures from his antique math collection; and Peter Cromwell, Lord & Lady Dunsany, Peter Knoppers, John Lienhard, John Mainstone, David Nicholls, Paul and Colin Roberts, Anders Sandberg, John Sullivan, and others for their valuable contributions.

I'm greatly indebted to Stephen Power, senior editor, and to Lisa Burstiner, senior production editor, at John Wiley & Sons, for their encouragement and unfailing attention to detail, and even proffering of alternative, clever solutions to some of the problems in the book. Any errors that remain are entirely my own responsibility. Thanks also to my marvelous agent, Patricia Van der Leun. And last but most of all, thanks to my family for letting me pursue a career that is really a fantasy.

Introduction

You are lost in a maze: How do you find your way out? You want to build a time machine, but is time travel logically possible? How can one infinity be bigger than another? Why can't you drink from a Klein bottle? What is the biggest number in the world to have a proper name, and how can you write it? Who claimed he could see in the fourth dimension? And what does "iteration" mean? And what does "iteration" mean?

Mathematics was never my strong point in school, but because I wanted to become an astronomer, I was told to stick with it. Fortunately, in my last two years before heading off to university, I had a wonderful old-fashioned, eccentric teacher (he actually wore a black gown when teaching), called Mr. Kay (known to one and all as "Danny"), who would suddenly divert from the chalk and blackboard to ask, "But how did the universe come to be *asymmetric*–that's what I want to know," or "These imaginary numbers are *very* interesting; in part, because they are so remarkably real." During lunchbreak, Danny and the senior chemistry teacher, Mr. Erp (whose nickname I need hardly spell out), would always meet in the chemistry prep room for a game of chess. They looked and acted very much like characters from a Wellsian science fiction tale, and I sometimes imagined them musing on formulas for invisibility or doorways to higher dimensions. At any rate, though I was never a shining student, I realize what a profound effect those two deeply imaginative, thoughtful men had on my future career. I did become an astronomer. I did persevere with math to a certain level of competence. But, much more than that, my curiosity was fired by the wonderful and weird possibilities of these subjects: curved space, Möbius bands, parallel universes, patterns in the heart of chaos, alternative realities. These strange possibilities, and a thousand others, make up the stuffing of this book. If you want a comprehensive, academic dictionary of mathematics, look elsewhere. If you want rigor and proof, try the next shelf. Herein you will find only the unusual and the outrageous, the fanciful and the fantastic: a compendium of the mathematics they *didn't* teach you in school.

Entries range from short definitions to lengthy articles on topics of major importance or unusual interest. These are arranged alphabetically according to the first word of the entry name and are extensively cross-referenced. Terms that appear in **bold** type have their own entries. A number of puzzles are included for the reader to try; the answers to these can be found at the back of the book. Also at the back are a comprehensive list of references and a category index. Readers are invited to visit the author's Web site at www.daviddarling.info for the latest news in mathematics and related subjects.

A

abacus

A counting frame that started out, several thousand years ago, as rows of pebbles in the desert sands of the Middle East. The word appears to come from the Hebrew *âbâq* (dust) or the Phoenician *abak* (sand) via the Greek *abax,* which refers to a small tray covered with sand to hold the pebbles steady. The familiar frame-supporting rods or wires, threaded with smoothly running beads, gradually emerged in a variety of places and mathematical forms.

In Europe, there was a strange state of affairs for more than 1,500 years. The Greeks and the Romans, and then the medieval Europeans, calculated on devices with a **place-value system** in which **zero** was represented by an empty line or wire. Yet the written notations didn't have a symbol for zero until it was introduced in Europe in 1202 by **Fibonacci**, via the Arabs and the Hindus.

abacus A special form of the Chinese abacus (c. 1958) consisting of two abaci stacked one on top of the other. *Luis Fernandes*

The Chinese *suan pan* differs from the European abacus in that the board is split into two decks, with two beads on each rod in the upper deck and five beads, representing the digits 0 through 4, on each rod in the bottom. When all five beads on a rod in the lower deck are moved up, they're reset to the original position, and one bead in the top deck is moved down as a carry. When both beads in the upper deck are moved down, they're reset and a bead on the adjacent rod on the left is moved up as a carry. The result of the computation is read off from the beads clustered near the separator beam between the upper and lower decks. In a sense, the abacus works as a 5-2-5-2-5-2 . . . –based number system in which carries and shifts are similar to those in the decimal system. Since each rod represents a digit in a decimal number, the capacity of the abacus is limited only by the number of rods on the abacus. When a user runs out of rods, she simply adds another abacus to the left of the row.

The Japanese *soroban* does away with the dual representations of fives and tens by having only four counters in the lower portion, known as "earth," and only one counter in the upper portion, known as "heaven." The world's largest abacus is in the Science Museum in London and measures 4.7 meters by 2.2 meters.

Abbott, Edwin Abbott (1838–1926)

An English clergyman and author who wrote several theological works and a biography (1885) of Francis Bacon, but is best known for his standard *Shakespearian Grammar* (1870) and the pseudonymously written ***Flatland: A Romance of Many Dimensions*** (by A Square, 1884).[1]

ABC conjecture

A remarkable **conjecture**, first put forward in 1980 by Joseph Oesterle of the University of Paris and David Masser of the Mathematics Institute of the University of Basel in Switzerland, that is now considered one of the most important unsolved problems in **number theory**. If it were proved correct, the proofs of many other famous conjectures and theorems would follow immediately–in some cases in just a few lines. The vastly complex current proof of **Fermat's last theorem**, for example, would reduce to less than a page of mathematical reasoning. The ABC conjecture is disarmingly simple compared to most of the deep questions in number theory and, moreover, turns out to be equivalent to all the main problems that involve **Diophantine equation**s (equations with integer coefficients and integer solutions).

Only a couple of concepts need to be understood to grasp the ABC conjecture. A *square-free number* is an integer that isn't divisible by the square of any number. For example, 15 and 17 are square-free, but 16 (divisible by 4^2) and 18 (divisible by 3^2) are not. The *square-free part* of an integer *n*, denoted sqp(*n*), is the largest square-free number that can be formed by multiplying the prime factors of *n*. For $n = 15$, the prime factors are 5 and 3, and $3 \times 5 = 15$, a square-free number, so that sqp(15) = 15. On the other hand, for $n = 16$, the prime factors are all 2, which means that sqp(16) = 2. In general, if *n* is square-free, the square-free part of *n* is just *n;* otherwise, sqp(*n*) represents what is left over after all the factors that create a square have been eliminated. In other words, sqp(*n*) is the product of the distinct **prime number**s that divide *n*. For example, sqp(9) = sqp(3 × 3) = 3 and sqp(1,400) = sqp(2 × 2 × 2 × 5 × 5 × 7) = 2 × 5 × 7 = 70.

The ABC conjecture deals with pairs of numbers that have no common factors. Suppose *A* and *B* are two such numbers that add to give *C*. For example, if $A = 3$ and $B = 7$, then $C = 3 + 7 = 10$. Now, consider the square-free part of the product $A \times B \times C$: sqp(ABC) = sqp(3 × 7 × 10) = 210. For most values of *A* and *B*, sqp(ABC) > C, as in the prior example. In other words, sqp(ABC)/C > 1. Occasionally, however, this isn't true. For instance, if $A = 1$ and $B = 8$, then $C = 1 + 8 = 9$, sqp(ABC) = sqp(1 × 8 × 9) = sqp(1 × 2 × 2 × 2 × 3 × 3) = 1 × 2 × 3 = 6, and sqp(ABC)/C = $6/9$ = $2/3$. Similarly, if $A = 3$ and $B = 125$, the ratio is $15/64$.

David Masser proved that the ratio sqp(ABC)/C can get arbitrarily small. In other words, given any number greater than zero, no matter how small, it's possible to find integers *A* and *B* for which sqp(ABC)/C is smaller than this number. In contrast, the ABC conjecture says that $[\text{sqp}(ABC)]^n/C$ reaches a minimum value if *n* is any number greater than 1—even a number such as 1.0000000001, which is only barely larger than 1. The tiny change in the expression results in a huge difference in its mathematical behavior. The ABC conjecture in effect translates an infinite number of Diophantine equations (including the equation of Fermat's last theorem) into a single mathematical statement.[144]

Abel, Niels Henrik (1802–1829)

> *The divergent series are the invention of the devil, and it is a shame to base on them any demonstration whatsoever. By using them, one may draw any conclusion he pleases and that is why these series have produced so many fallacies and so many paradoxes.*

A Norwegian mathematician who, independently of his contemporary Évariste **Galois**, pioneered **group** theory and proved that there are no algebraic solutions of the general **quintic** equation. Both Abel and Galois died tragically young—Abel of tuberculosis, Galois in a sword fight.

While a student in Christiania (now Oslo), Abel thought he had discovered how to solve the general quintic algebraically, but soon corrected himself in a famous pamphlet published in 1824. In this early paper, Abel showed the impossibility of solving the general quintic by means of radicals, thus laying to rest a problem that had perplexed mathematicians since the mid-sixteenth century. Abel, chronically poor throughout his life, was granted a small stipend by the Norwegian government that allowed him to go on a mathematical tour of Germany and France. In Berlin he met Leopold Crelle (1780–1856) and in 1826 helped him found the first journal in the world devoted to mathematical research. Its first three volumes contained 22 of Abel's papers, ensuring lasting fame for both Abel and Crelle. Abel revolutionized the important area of elliptic integrals with his theory of **elliptic function**s, contributed to the theory of **infinite series**, and founded the theory of commutative groups, known today as **Abelian group**s. Yet his work was never properly appreciated during his life, and, impoverished and ill, he returned to Norway unable to obtain a teaching position. Two days after his death, a delayed letter was delivered in which Abel was offered a post at the University of Berlin.

Abelian group

A **group** that is **commutative**, that is, in which the result of multiplying one member of the group by another is independent of the order of multiplication. Abelian groups, named after Niels **Abel**, are of central importance in modern mathematics, most notably in **algebraic topology**. Examples of Abelian groups include the **real number**s (with addition), the nonzero real numbers (with multiplication), and all cyclic groups, such as the **integer**s (with addition).

abracadabra

A word famously used by magicians but which started out as a cabalistic or mystical charm for curing various ailments, including toothache and fever. It was first mentioned in a poem called "Praecepta de Medicina" by the Gnostic physician Quintus Severus Sammonicus in the second century A.D. Sammonicus instructed that the letters be written on parchment in the form of a triangle:

A B R A C A D A B R A
A B R A C A D A B R
A B R A C A D A B
A B R A C A D A
A B R A C A D
A B R A C A
A B R A C
A B R A
A B R
A B
A

This was to be folded into the shape of a cross, worn for nine days suspended from the neck, and, before sunrise, cast behind the patient into a stream running eastward. It was also a popular remedy in the Middle Ages. During the Great Plague, around 1665, large numbers of these amulets were worn as safeguards against infection. The origin of the word itself is uncertain. One theory is that it is based on Abrasax, the name of an Egyptian deity.

PUZZLE

A well-known puzzle, proposed by George Polya (1887–1985), asks how many different ways there are to spell abracadabra in this diamond-shaped arrangement of letters:

A
B B
R R R
A A A A
C C C C C
A A A A A A
D D D D D
A A A A
B B B
R R
A

Solutions begin on page 369.

abscissa

The *x*-coordinate, or horizontal distance from the *y*-axis, in a system of **Cartesian coordinates**. Compare with **ordinate**.

absolute

Not limited by exceptions or conditions. The term is used in many different ways in mathematics, physics, philosophy, and everyday speech. Absolute space and absolute time, which, in Newton's universe, form a unique, immutable frame of reference, blend and become deformable in the **space-time** of Einstein. See also **absolute zero**. In some philosophies, the absolute stands behind the reality we see–independent, transcendent, unconditional, and all-encompassing. The American philosopher Josiah Royce (1855–1916) took the absolute to be a spiritual entity whose self-consciousness is imperfectly reflected in the totality of human thought. Mathematics, too, reaches beyond imagination with its absolute **infinity**. See also **absolute value**.

absolute value

The value of a number without regard to its sign. The absolute value, or *modulus,* of a **real number**, *r,* is the distance of the number from zero measured along the real **number line**, and is denoted $|r|$. Being a distance, it can't be negative; so, for example, $|3| = |-3| = 3$. The same idea applies to the absolute value of a **complex number** $a + ib$, except that, in this case, the complex number is represented by a point on an **Argand diagram**. The absolute value, $|a + ib|$, is the length of the line from the origin to the given point, and is equal to $\sqrt{(a^2 + b^2)}$.

absolute zero

The lowest possible temperature of a substance, equal to 0 Kelvin (K), –273.15°C, or –459.67°F. In classical physics, it is the temperature at which all molecular motion ceases. However, in the "real" world of **quantum mechanics** it isn't possible to stop all motion of the particles making up a substance as this would violate the Heisenberg uncertainty principle. So, at 0 K, particles would still vibrate with a certain small but nonzero energy known as the *zero-point energy.* Temperatures within a few billionths of a degree of absolute zero have been achieved in the laboratory. At such low temperatures, substances have been seen to enter a peculiar state, known as the *Bose-Einstein condensate,* in which their quantum wave functions merge and particles lose their individual identities. Although it is possible to approach ever closer to absolute zero, the third law of thermodynamics asserts that it's impossible to ever attain it. In a deep sense, absolute zero lies at the asymptotic limit of low energy just as the speed of light lies, for particles with mass, at the asymptotic limit of high energy. In both cases, energy of motion (kinetic energy) is the key quantity involved. At the high energy end, as the average speed of the particles of a substance approaches the speed of light, the temperature rises without limit, heading for an unreachable ∞ K.

abstract algebra

> *To a mathematician, real life is a special case.*
>
> –Anonymous

Algebra that is not confined to familiar number systems, such as the **real numbers**, but seeks to solve equations that may involve many other kinds of systems. One of its aims, in fact, is to ask: What other number systems are there? The term *abstract* refers to the perspective taken on the subject, which is very different from that of high school algebra. Rather than looking for the solutions to a particular problem, abstract algebra is interested in such questions as: When does a solution exist? If a solution does exist, is it unique? What general properties does a solution possess? Among the structures it deals with are **group**s, **ring**s, and **field**s. Historically, examples of such structures often arose first in some other field of mathematics, were specified rigorously (axiomatically), and were then studied in their own right in abstract algebra.

Abu'l Wafa (A.D. 940–998)

A Persian mathematician and astronomer who was the first to describe geometrical constructions (see **constructible**) possible only with a straightedge and a fixed compass, later dubbed a "rusty compass," that never alters its radius. He pioneered the use of the tangent function (see **trigonometric function**), apparently discovered the secant and cosecant functions, and compiled tables of sines and tangents at 15′ intervals—work done as part of an investigation into the orbit of the Moon.

abundant number

A number that is smaller than the sum of its **aliquot part**s (proper divisors). Twelve is the smallest abundant number; the sum of its aliquot parts is 1 + 2 + 3 + 4 + 6 = 16, followed by 18, 20, 24, and 30. A *weird number* is an abundant number that is not *semiperfect;* in other words, n is weird if the sum of its divisors is greater than n, but n is not equal to the sum of any subset of its divisors. The first few weird numbers are 70, 836, 4,030, 5,830, and 7,192. It isn't known if there are any odd weird numbers. A *deficient number* is one that is greater than the sum of its aliquot parts. The first few deficient numbers are 1, 2, 3, 4, 5, 8, and 9. Any divisor of a deficient (or perfect) number is deficient. A number that is not abundant or deficient is known as a **perfect number**.

Achilles and the Tortoise paradox

See **Zeno's paradoxes**.

Ackermann function

One of the most important **function**s in computer science. Its most outstanding property is that it grows astonishingly fast. In fact, it gives rise to **large numbers** so quickly that these numbers, called *Ackermann numbers,* are written in a special way known as **Knuth's up-arrow notation**. The Ackermann function was discovered and studied by Wilhelm Ackermann (1896–1962) in 1928. Ackermann worked as a high school teacher from 1927 to 1961 but was also a student of the great mathematician David **Hilbert** in Göttingen and, from 1953, served as an honorary professor in the university there. Together with Hilbert he published the first modern textbook on mathematical logic. The function he discovered, and that now bears his name, is the simplest example of a well-defined and total function that is also computable but not primitive recursive (PR). "Well-defined and total" means that the function is internally consistent and doesn't break any of the rules laid down to define it. "Computable" means that it can, in principle, be evaluated for all possible input values of its variables. "Primitive recursive" means that it can be computed using only *for loops*—repeated application of a single operation a predetermined number of times. The recursion, or feedback loop, in the Ackermann function overruns the capacity of any *for loop* because the number of loop repetitions isn't known in advance. Instead, this number is itself part of the computation, and grows as the calculation proceeds. The Ackermann function can only be calculated using a *while loop,* which keeps repeating an action until an associated test returns false. Such loops are essential when the programmer doesn't know at the outset how many times the loop will be traversed. (It's now known that everything computable can be programmed using while loops.)

The Ackermann function can be defined as follows:

$$\begin{aligned} A(0, n) &= n + 1 \text{ for } n = 0 \\ A(m, 0) &= A(m - 1, 1) \text{ for } m = 1 \\ A(m, n) &= A(m - 1, A(m, n - 1)) \text{ for } m, n = 1. \end{aligned}$$

Two positive integers, m and n, are the input and $A(m, n)$ is the output in the form of another positive integer. The function can be programmed easily in just a few lines of code. The problem isn't the complexity of the function but the awesome rate at which it grows. For example, the innocuous-looking $A(4,2)$ already has 19,729 digits! The use of a powerful large-number shorthand system, such as the up-arrow notation, is indispensable as the following examples show:

$$\begin{aligned} A(1, n) &= 2 + (n + 3) - 3 \\ A(2, n) &= 2 \times (n + 3) - 3 \\ A(3, n) &= 2\uparrow(n + 3) - 3 \\ A(4, n) &= 2\uparrow(2\uparrow(2\uparrow(\ldots\uparrow 2))) - 3 \quad (n + 3 \text{ twos}) \\ &= 2\uparrow\uparrow(n + 3) - 3 \\ A(5, n) &= 2\uparrow\uparrow\uparrow(n + 3) - 3, \text{ etc.} \end{aligned}$$

Intuitively, the Ackermann function defines generalizations of multiplication by 2 (iterated additions) and exponentiation with base 2 (iterated multiplications) to iterated exponentiation, iteration of this operation, and so on.[84]

acre

An old unit of **area**, equal to 160 square rods, 4,840 square yards, 43,560 square feet, or 4,046.856 square meters.

acute

From the Latin *acus* for "needle" (which also forms the root for *acid, acupuncture,* and *acumen*). An acute **angle** is less than 90°. An acute **triangle** is one in which all three angles are acute. Compare with **obtuse**.

adjacent

Next to. *Adjacent angles* are next to each other, and thus share one side. *Adjacent sides* of a **polygon** share a vertex.

affine geometry

The study of properties of geometric objects that remain unchanged after parallel projection from one plane to another. During such a projection, first studied by Leonhard **Euler**, each point (x, y) is mapped to a new point $(ax + cy + e, bx + dy + f)$. Circles, angles, and distances are altered by affine transformations and so are of no interest in affine geometry. Affine transformations do, however, preserve collinearity of points: if three points belong to the same straight line, their images (the points that correspond to them) under affine transformations also belong to the same line and, in addition, the middle point remains between the other two points. Similarly, under affine transformations, parallel lines remain parallel; concurrent lines remain concurrent (images of intersecting lines intersect); the ratio of lengths of line segments of a given line remains constant; the ratio of areas of two triangles remains constant; and ellipses, parabolas, and hyperbolas continue to be ellipses, parabolas, and hyperbolas.

age puzzles and tricks

Problems that ask for a person's age or, alternatively, when a person was a certain age, given several roundabout facts. They go back at least 1,500 years to the time of Metrodorus and **Diophantus's riddle**. A number of distinct types of age puzzles sprang up between the sixteenth and early twentieth centuries, in most cases best solved by a little algebra. One form asks: if X is now a years old and Y is now b years old, when will X be c times as old as Y? The single unknown, call it x, can be found from the equation $a + x = c(b + x)$. Another type of problem takes the form: if X is now a times as old as Y and after b years X will be c times as old as Y, how old are X and Y now? In this case the trick is to set up and solve two simultaneous equations: $X = aY$ and $X + b = c(Y + b)$.

PUZZLES

Around 1900, two more variants on the age puzzle became popular. Here is an example of each for the reader to try.

1. Bob is 24. He is twice as old as Alice was when Bob was as old as Alice is now. How old is Alice?
2. The combined ages of Mary and Ann are 44 years. Mary is twice as old as Ann was when Mary was half as old as Ann will be when Ann is three times as old as Mary was when Mary was three times as old as Ann. How old is Ann?"

Solutions begin on page 369.

Various mathematical sleights of hand can seem to conjure up a person's age as if by magic. For example, ask a person to multiply the first number of his or her age by 5, add 3, double this figure, add the second number of his or her age to the figure, and tell you the answer. Deduct 6 from this and you will have their age.

Alternatively, ask the person to pick a number, multiply this by 2, add 5, and multiply by 50. If the person has already had a birthday this year and it's the year 2004, she should add 1,754, otherwise she should add 1,753. Each year after 2004 these numbers need to be increased by 1. Finally, the person should subtract the year they were born. The first digits of the answer are the original number, while the last two digits are the person's age.

Here is one more trick. Take your age, multiply it by 7, then multiply again by 1,443. The result is your age repeated three times. (What you have actually done is multiplied by 10,101; if you multiply by 1,010,101, the repetition is fourfold, and so on.)

Agnesi, Maria Gaetana (1718–1799)

An Italian mathematician and scholar whose name is associated with the curve known as the *Witch of Agnesi.* Born in Milan, Maria was one of 24 children of a professor of mathematics at the University of Bologna. A child prodigy, she could speak seven languages, including Latin, Greek, and Hebrew, by the age of 11 and was solving difficult problems in geometry and ballistics by her early teens. Her father encouraged her studies and her appearance at public debates. However, Maria developed a chronic illness, marked by convulsions and headaches, and, from the age of about 20, withdrew socially and devoted herself to mathematics. Her *Instituzioni analitiche ad uso della gioventu italiana,* published in 1748, became a standard teaching manual, and in 1750, she was appointed to the chair of mathematics and natural philosophy at Bologna. Yet she never fulfilled her early promise in terms of making new

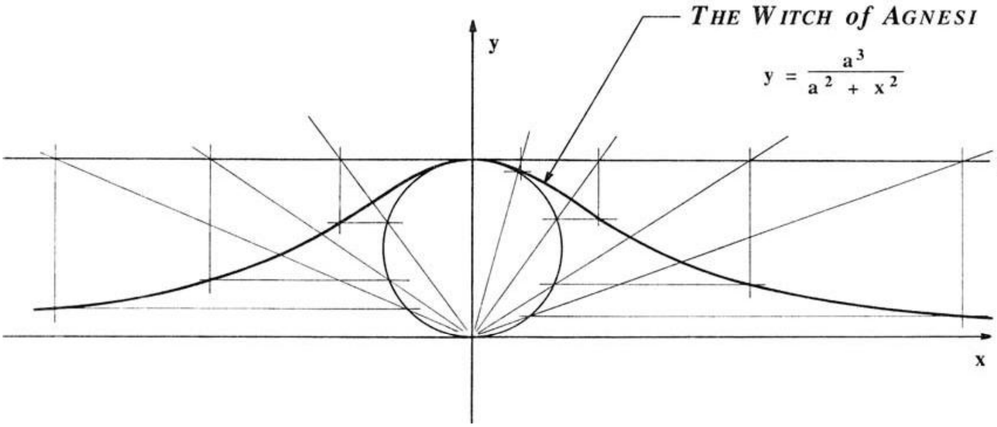

Agnesi, Maria Gaetana The Witch of Agnesi curve. *John H. Lienhard*

breakthroughs. After the death of her father in 1752, she moved into theology and, after serving for some years as the directress of the Hospice Trivulzio for Blue Nuns at Milan, joined the sisterhood herself and ended her days in this austere order.

The famous curve that bears her name had been studied earlier, in 1703, by Pierre de **Fermat** and the Italian mathematician Guido Grandi (1671–1742). Maria wrote about it in her teaching manual and referred to it as the *aversiera,* which simply means "to turn." But in translating this, the British mathematician John Colson (1680–1760), the fifth Lucasian professor of mathematics at Cambridge University, confused *aversiera* with *avversiere* which means "witch," or "wife of the devil." And so the name of the curve came down to us as the Witch of Agnesi. To draw it, start with a circle of diameter a, centered at the point $(0, a/2)$ on the y-axis. Choose a point A on the line $y = a$ and connect it to the origin with a line segment. Call the point where the segment crosses the circle B. Let P be the point where the vertical line through A crosses the horizontal line through B. The Witch is the curve traced by P as A moves along the line $y = a$. By a happy coincidence, it does look a bit like a witch's hat! In Cartesian coordinates, its equation is

$$y = a^3/(x^2 + a^2).$$

Ahmes papyrus

See **Rhind papyrus**.

Ahrens, Wilhelm Ernst Martin Georg (1872–1927)

A great German exponent of recreational mathematics whose *Mathematische Unterhaltungen und Spiele*[6] is one of the most scholarly of all books on the subject.

Alcuin (735–804)

A leading intellectual of his time and the probable compiler of *Propositiones ad Acuendos Juvenes* (Problems to sharpen the young), one of the earliest collections of recreational math problems. According to David **Singmaster** and John Hadley: "The text contains 56 problems, including 9 to 11 major types of problem which appear for the first time, 2 major types which appear in the West for the first time and 3 novel variations of known problems. . . . It has recently been realized that the **river-crossing problems** and the crossing-a-desert problem, which appear here for the first time, are probably the earliest known combinatorial problems."

Alcuin was born into a prominent family near the east coast of England. He was sent to York, where he became a pupil and, eventually, in 778, the headmaster, of Archbishop Ecgberht's School. (Ecgberht was the last person to have known the Venerable Bede.) Alcuin built up a superb library and made the school one of the chief centers of learning in Europe. Its reputation became such that, in 781, Alcuin was invited to become master of Charlemagne's Palace School at Aachen and, effectively, minister of education for Charlemagne's empire. He accepted and traveled to Aachen to a meeting of the leading scholars. Subsequently, he was made head of Charlemagne's Palace School and there developed the Carolingian minuscule, a clear, legible script that became the basis of how letters of the present Roman alphabet are written.

Before leaving Aachen, Alcuin was responsible for the most prized of the Carolingian codices, now called the

Golden Gospels: a series of illuminated masterpieces written largely in gold, on white or purple vellum. The development of Carolingian minuscule had, indirectly, a major impact on the history of mathematics. Because it was a far more easily readable script than the older unspaced capital, it led to many mathematical works being newly copied into this new style in the ninth century. Most of the works of the ancient Greek mathematicians that have survived did so because of this transcription. Alcuin lived in Aachen from 782 to 790 and again from 793 to 796. In 796, he retired from Charlemagne's Palace School and became abbot of the Abbey of St. Martin at Tours, where he and his monks continued to work with the Carolingian minuscule script.

aleph

The first letter of the Hebrew alphabet, ℵ. It was first used in mathematics by Georg **Cantor** to denote the various orders, or sizes, of **infinity**: $\aleph_0$ (aleph-null), $\aleph_1$ (aleph-one), etc. An earlier (and still used) symbol for infinity, ∞, was introduced in 1655 by John **Wallis** in his *Arithmetica infinitorum* but didn't appear in print until the *Ars conjectandi* by Jakob Bernoulli, published posthumously in 1713 by his nephew Nikolaus Bernoulli (see **Bernoulli Family**).

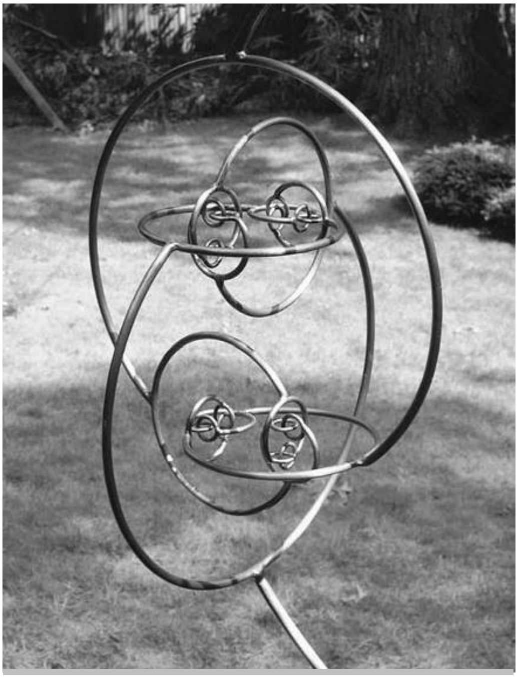

Alexander's horned sphere A sculpture of a five-level Alexander's horned sphere. *Gideon Weisz, www.gideonweisz.com*

Alexander's horned sphere

In **topology,** an example of what is called a "wild" structure; it is named after the Princeton mathematician James Waddell Alexander (1888–1971) who first described it in the early 1920s. The horned sphere is topologically equivalent to the **simply connected** surface of an ordinary hollow **sphere** but bounds a region that is not simply connected. The horns-within-horns consist of a recursive set–a **fractal**–of interlocking pairs of orthogonal rings (rings set at right angles) of decreasing radius. A rubber band around the base of any horn couldn't be removed from the structure even after infinitely many steps. The horned sphere can be embedded in the plane by reducing the interlock angle between ring pairs from 90° to 0°, then weaving the rings together in an over-under pattern. The sculptor Gideon Weisz has modeled a number of approximations to the structure, one of which is shown in the photograph.

algebra

A major branch of mathematics that, at an elementary level, involves applying the rules of arithmetic to numbers, and to letters that stand for unknown numbers, with the main aim of solving equations. Beyond the algebra learned in high school is the much vaster and more profound subject of **abstract algebra**. The word itself comes from the Arabic *al-jebr,* meaning "the reunion of broken parts." It first appeared in the title of a book, *Al-jebr w'al-mugabalah* (The science of reduction and comparison), by the ninth-century Persian scholar **al-Khowarizmi**–probably the greatest mathematician of his age, and as famous among Arabs as Euclid and Aristotle are to the Western world.

algebraic curve

A curve whose equation involves only *algebraic functions.* These are functions that, in their most general form, can be written as a sum of **polynomial**s in x multiplied by powers of y, equal to zero. Among the simplest examples are straight lines and **conic section**s.

algebraic fallacies

Misuse of **algebra** can have some surprising and absurd results. Here, for example, is a famous "proof" that 1 = 2:

$$\begin{gathered}\text{Let } a = b.\\ \text{Then } a^2 = ab\\ a^2 + a^2 = a^2 + ab\\ 2a^2 = a^2 + ab\\ 2a^2 - 2ab = a^2 + ab - 2ab\\ 2a^2 - 2ab = a^2 - ab\\ 2(a^2 - ab) = 1(a^2 - ab).\\ \text{Dividing both sides by } a^2 - ab\\ 2 = 1.\end{gathered}$$

Where's the mistake? The problem lies with the seemingly innocuous final division. Since $a = b$, dividing by $a^2 - ab$ is the same as dividing by **zero**–the great taboo of mathematics.

Another false argument runs as follows:

$$(n + 1)^2 = n^2 + 2n + 1$$
$$(n + 1)^2 - (2n + 1) = n^2.$$

Subtracting $n(2n + 1)$ from both sides and factorizing gives

$$(n + 1)^2 - (n + 1)(2n + 1) = n^2 - n(2n + 1).$$

Adding $\frac{1}{4}(2n + 1)^2$ to both sides yields

$$(n + 1)^2 - (n + 1)(2n + 1) + \tfrac{1}{4}(2n + 1)^2$$
$$= n^2 - n(2n + 1) + \tfrac{1}{4}(2n + 1)^2.$$

This may be written:

$$[(n + 1) - \tfrac{1}{2}(2n + 1)]^2 = [(n - \tfrac{1}{2}(2n + 1)]^2.$$

Taking square roots of both sides,

$$n + 1 - \tfrac{1}{2}(2n + 1) = n - \tfrac{1}{2}(2n + 1).$$

Therefore,

$$n = n + 1.$$

The problem here is that there are *two* square roots for any positive number, one positive and one negative: the square roots of 4 are 2 and –2, which can be written as ±2. So the penultimate step should properly read:

$$\pm(n + 1 - \tfrac{1}{2}(2n + 1)) = \pm(n - \tfrac{1}{2}(2n + 1))$$

algebraic geometry

Originally, the geometry of **complex number** solutions to **polynomial** equations. Modern algebraic geometry is also concerned with algebraic varieties, which are a generalization of the solution sets found in the traditional subject, as well as solutions in fields other than complex numbers, for example *finite fields.*

algebraic number

A **real number** that is a **root** of a **polynomial** equation with integer coefficients. For example, any **rational number** a/b, where a and b are nonzero integers, is an algebraic number of degree one, because it is a root of the linear equation $bx - a = 0$. The **square root of two** is an algebraic number of degree two because it is a root of the quadratic equation $x^2 - 2 = 0$. If a real number is not algebraic, then it is a **transcendental number**. Almost all real numbers are transcendental because, whereas the set of algebraic numbers is countably infinite (see **countable set**), the set of transcendental numbers is uncountably infinite.

algebraic number theory

The branch of **number theory** that is studied without using methods such as **infinite series** and **convergence** taken from **analysis.** It contrasts with **analytical number theory**.

algebraic topology

A branch of **topology** that deals with invariants of a topological space that are algebraic structures, often **group**s.

algorithm

A systematic method for solving a problem. The word comes from the name of the Persian mathematician, **al-Khowarizmi**, and may have been first used by Gottfried **Liebniz** in the late 1600s. It remained little known in Western mathematics, however, until the Russian mathematician Andrei Markov (1903–1987) reintroduced it. The term became especially popular in the areas of math focused on computing and computation.

algorithmic complexity

A measure of **complexity** developed by Gregory **Chaitin** and others, based on Claude **Shannon**'s **information theory** and earlier work by the Russian mathematicians Andrei **Kolmogorov** and Ray Solomonoff. Algorithmic complexity quantifies how complex a system is in terms of the shortest computer program, or set of **algorithm**s, needed to completely describe the system. In other words, it is the smallest model of a given system that is necessary and sufficient to capture the essential patterns of that system. Algorithmic complexity has to do with the mixture of repetition and innovation in a complex system. At one extreme, a highly regular system can be described by a very short program or algorithm. For example, the bit string 01010101010101010101 . . . follows from just three commands: print a zero, print a one, and repeat the last two commands indefinitely. The complexity of such a system is very low. At the other extreme, a totally **random** system has a very high algorithmic complexity since the random patterns can't be condensed into a smaller set of algorithms: the program is effectively as large as the system itself. See also **compressible**.

Alhambra

The former palace and citadel of the Moorish kings of Granada, and perhaps the greatest monument to Islamic mathematical art on Earth. Because the Qur'an consid-

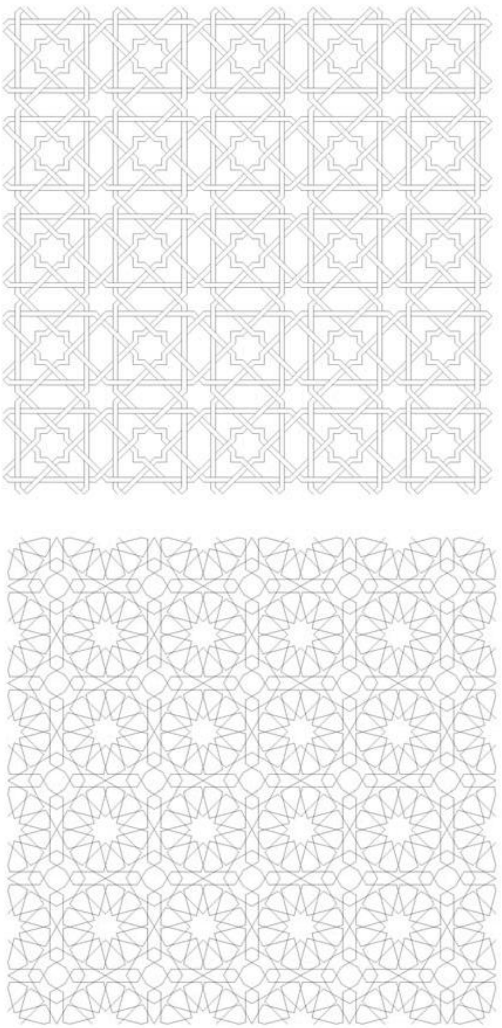

Alhambra Computer-generated tilings based on Islamic tile designs such as those found in the Alhambra. *Xah Lee, www.xahlee.org*

ers the depiction of living beings in religious settings blasphemous, Islamic artists created intricate patterns to symbolize the wonders of creation: the repetitive nature of these complex geometric designs suggests the limitless power of God. The sprawling citadel, looming high above the Andalusian plain, boasts a remarkable array of mosaics with tiles arranged in intricate patterns. The Alhambra **tiling**s are *periodic;* in other words, they consist of some basic unit that is repeated in all directions to fill up the available space. All 17 different groups of isometries–the possible ways of repeatedly tiling the plane–are used at the palace. The designs left a deep impression on Maurits **Escher**, who came here in 1936. Subsequently, Escher's art took on a much more mathematical nature, and over the next six years he produced 43 colored drawings of periodic tilings with a wide variety of symmetry types.

aliquot part

Also known as a *proper divisor,* any divisor of a number that isn't equal to the number itself. For instance, the aliquot parts of 12 are 1, 2, 3, 4, and 6. The word comes from the Latin *ali* ("other") and *quot* ("how many"). An *aliquot sequence* is formed by taking the sum of the aliquot parts of a number, adding them to form a new number, then repeating this process on the next number and so on. For example, starting with 20, we get $1 + 2 + 4 + 5 + 10 = 22$, then $1 + 2 + 11 = 14$, then $1 + 2 + 7 = 10$, then $1 + 2 + 5 = 8$, then $1 + 2 + 4 = 7$, then 1, after which the sequence doesn't change. For some numbers, the result loops back immediately to the original number; in such cases the two numbers are called **amicable numbers.** In other cases, where a sequence repeats a pattern after more than one step, the result is known as an *aliquot cycle* or a *sociable chain.* An example of this is the sequence 12496, 14288, 15472, 14536, 14264, . . . The aliquot parts of 14264 add to give 12496, so that the whole cycle begins again. Do all aliquot sequences end either in 1 or in an aliquot cycle (of which amicable numbers are a special case)? In 1888, the Belgian mathematician Eugène Catalan (1814–1894) conjectured that they do, but this remains an open question.

al-Khowarizmi (c. 780–850)

An Arabic mathematician, born in Baghdad, who is widely considered to be the founder of modern day **algebra**. He believed that any math problem, no matter how difficult, could be solved if broken down into a series of smaller steps. The word ***algorithm*** may have derived from his name.

Allais paradox

A **paradox** that stems from questions asked in 1951 by the French economist Maurice Allais (1911–).[8] Which of these would you choose: (A) an 89% chance of receiving an unknown amount and 11% chance of $1 million; or (B) an 89% chance of an unknown amount (the same amount as in A), a 10% chance of $2.5 million, and a 1% chance of nothing? Would your choice be the same if the unknown amount was $1 million? What if the unknown amount was zero?

Most people don't like risk and so prefer the better chance of winning $1 million in option A. This choice is firm when the unknown amount is $1 million, but seems to waver as the amount falls to nothing. In the latter case, the risk-averse person favors B because there isn't much difference between 10% and 11%, but there's a big difference between $1 million and $2.5 million. Thus the choice between A and B depends on the unknown amount, even though it is the same unknown amount independent of the choice. This flies in the face of the so-called *independence axiom,* that rational choice between two alternatives should depend only on how those two alternatives differ. Yet, if the amounts involved in the problem are reduced to tens of dollars instead of millions of dollars, people's behavior tends to fall back in line with the axioms of rational choice. In this case, people tend to choose option B regardless of the unknown amount. Perhaps when presented with such huge numbers, people begin to calculate qualitatively. For example, if the unknown amount is $1 million the options are essentially (A) a fortune guaranteed or (B) a fortune almost guaranteed with a small chance of a bigger fortune and a tiny chance of nothing. Choice A is then rational. However, if the unknown amount is nothing, the options are (A) a small chance of a fortune ($1 million) and a large chance of nothing, and (B) a small chance of a larger fortune ($2.5 million) and a large chance of nothing. In this case, the choice of B is rational. Thus, the Allais paradox stems from our limited ability to calculate rationally with such unusual quantities.

almost perfect number

A description sometimes applied to the powers of 2 because the **aliquot part**s (proper divisors) of 2^n sum to $2^n - 1$. So a power of 2 is a deficient number (one that is less than the sum of its proper divisors), but only just. It isn't known whether there is an odd number n whose divisors (excluding itself) sum to $n - 1$.

alphamagic square

A form of **magic square**, introduced by Lee **Sallows**,[278–280] in which the number of letters in the word for each number, in whatever language is being used, gives rise to another magic square. In English, for example, the alphamagic square:

5 (five)	22 (twenty-two)	18 (eighteen)
28 (twenty-eight)	15 (fifteen)	2 (two)
12 (twelve)	8 (eight)	25 (twenty-five)

generates the square:

4	9	8
11	7	3
6	5	10

A surprisingly large number of 3 × 3 alphamagic squares exist–in English and in other languages. French allows just one 3 × 3 alphamagic square involving numbers up to 200, but a further 255 squares if the size of the entries is increased to 300. For entries less than 100, none occurs in Danish or in Latin, but there are 6 in Dutch, 13 in Finnish, and an incredible 221 in German. Yet to be determined is whether a 3 × 3 square exists from which a magic square can be derived that, in turn, yields a third magic square–a magic triplet. Also unknown is the number of 4 × 4 and 5 × 5 language-dependent alphamagic squares. Here, for example, is a four-by-four English alphamagic square:

26	37	48	59
49	58	27	36
57	46	39	28
38	29	56	47

alphametic

A type of **cryptarithm** in which a set of words is written down in the form of a long addition sum or some other mathematical problem. The object is to replace the letters of the alphabet with decimal digits to make a valid arithmetic sum. The word *alphametic* was coined in 1955 by James **Hunter**. However, the first modern alphametic, published by Henry **Dudeney** in the July 1924 issue of *Strand Magazine,* was "Send more money," or, setting it out in the form of a long addition:

$$\begin{array}{r} \text{SEND} \\ \underline{\text{MORE}} \\ \text{MONEY} \end{array}$$

and has the (unique) solution:

$$\begin{array}{r} 9567 \\ \underline{1085} \\ 10652 \end{array}$$

PUZZLES

The reader is invited to try to solve the following elegant examples:

1. Earth, air, fire, water: nature. (Herman Nijon)
2. Saturn, Uranus, Neptune, Pluto: planets. (Peter J. Martin)
3. Martin Gardner retires. (H. Everett Moore)

Solutions begin on page 369.

Two rules are obeyed by every alphametic. First, the mapping of letters to numbers is one-to-one; that is, the same letter always stands for the same digit, and the same digit is always represented by the same letter. Second, the digit zero isn't allowed as the left-most digit in any of the numbers being added or in their sum. The best alphametics are reckoned to be those with only one correct answer.

Altekruse puzzle

A symmetrical 12-piece **burr puzzle** for which a patent was granted to William Altekruse in 1890. The Altekruse family is of Austrian-German origin and, curiously, the name means "old cross" in German, which has led some authors to incorrectly assume that it was a pseudonym. William Altekruse came to the United States as a young man in 1844 with his three brothers to escape being drafted into the German army. The Altekruse puzzle has an unusual mechanical action in the first step of disassembly by which two halves move in opposition to each other, unlike the more familiar burr types that have a key piece or pieces. Depending on how it is assembled, this action can take place along one, two, or all three axes independently but not simultaneously.

alternate

A mathematical term with several different meanings: (1) *Alternate angles* are angles on opposite sides and opposite ends of a line that cuts two parallel lines. (2) A well-known theorem called the *alternate segment theorem* involves the segment on the opposite side of a given chord of a circle. (3) An *alternate hypothesis* in statistics is the alternative offered to the **null hypothesis**. (4) To alternate is to cycle backward and forward between two different values, for example, 0, 1, 0, 1, 0, 1, ...

altitude

A perpendicular line segment from one **vertex** of a figure or solid to an edge or face opposite to that vertex. Also the length of such a line segment.

ambiguous figure

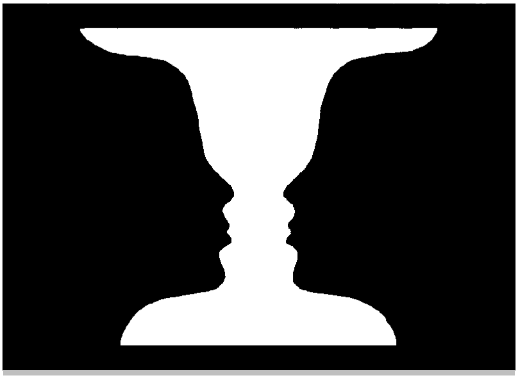

ambiguous figure The Rubin vase illusion: one moment a vase, the next two people face to face.

An **optical illusion** in which the subject or the perspective of a picture or shape may suddenly switch in the mind of the observer to another, equally valid possibility. Often the ambiguity stems from the fact that the figure and ground can be reversed. An example of this is the vase/profile illusion, made famous by the Danish psychologist Edgar John Rubin (1886–1951) in 1915, though earlier versions of the same illusion can be found in many eighteenth-century French prints depicting a variety of vases, usually in a naturalistic setting, and profiles of particular people. The same effect can be created in three dimensions with a suitably shaped solid vase. In some ambiguous figures, the features of a person or of an animal can suddenly be seen as different features of another individual. Classic examples include the old woman–young woman illusion and the duck–rabbit illusion. **Upside-down picture**s involve a special case of dual-purpose features in which the reversal is accomplished not mentally, by suddenly "seeing" the alternative, but physically, by turning the picture 180°. Ambiguity can also occur, particularly in some geometric drawings, when there is confusion as to which are the front and the back faces of a figure, as in the **Necker cube**, the **Thiery figure**, and **Schröder's reversible staircase**.

ambiguous connectivity

See **impossible figure**.

Ames room

The famous distorted room illusion, named after the American ophthalmologist Adelbert Ames Jr. (1880–1955), who first constructed such a room in 1946 based on a concept by the German physicist Hermann Helmholtz in the late nineteenth century. The Ames room looks cubic when seen with one eye through a specially positioned peephole; however, the room's true shape is trapezoidal. The floor, ceiling, some walls, and the far windows are trapezoidal surfaces; the floor appears level but is actually at an incline (one of the far corners being

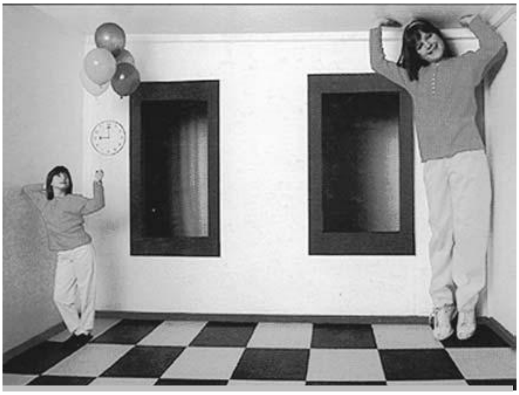

Ames room Misleading geometry makes these identical twins appear totally different in size. *Technische Universitat, Dresden*

much lower than the other); and the walls are slanted outward, though they seem perpendicular to the floor. This shape makes it look as if people or objects grow or shrink as they move from one corner of the room to another. See also **distortion illusion**.[142, 178]

amicable numbers

A pair of numbers, also known as *friendly numbers,* each of whose **aliquot part**s add to give the other number. (An aliquot part is any divisor that doesn't include the number itself.) The smallest amicable numbers are 220 (aliquot parts 1, 2, 4, 5, 10, 11, 20, 22, 44, 55, and 110, with a sum of 284) and 284 (aliquot parts 1, 2, 4, 71, and 142, with a sum of 220). This pair was known to the ancient Greeks, and the Arabs found several more. In 1636, Pierre de **Fermat** rediscovered the amicable pair 17,296 and 18,416; two years later René **Descartes** rediscovered a third pair, 9,363,584 and 9,437,056. In the eighteenth century, Leonhard **Euler** drew up a list of more than 60. Then, in 1866, B. Nicolò Paganini (not the violinist), a 16-year-old Italian, startled the mathematical world by announcing that the numbers 1,184 and 1,210 were amicable. This second-lowest pair of all had been completely overlooked! Today, the tally of known amicable numbers has grown to about 2.5 million. No amicable pair is known in which one of the two numbers is a square. An unusually high proportion of the numbers in amicable pairs ends in either 0 or 5. A *happy amicable pair* is an amicable pair in which both numbers are **happy number**s; an example is 10,572,550 and 10,854,650. See also **Harshad number**.

amplitude

Size or magnitude. The origin of the word is the same Indo-European *ple* root that gives us *plus* and *complement.* The more immediate Latin source is *amplus* for "wide." Today, **amplitude** is used to describe, among other things, the distance a **periodic** function varies from its central value, and the magnitude of a **complex number**.

anagram

The rearrangement of the letters of a word or phrase into another word or phrase, using all the letters only once. The best anagrams are meaningful and relate in some way to the original subject; for example, "stone age" and "stage one." There are also many remarkable examples of long anagrams. " 'That's one small step for a man; one giant leap for mankind.' Neil Armstrong" becomes "An 'Eagle' lands on Earth's Moon, making a first small permanent footprint."

PUZZLES

The reader is invited to untangle the following anagrams that give clues to famous people:

1. A famous German waltz god.
2. Aha! Ions made volts!
3. I'll make a wise phrase.

Solutions begin on page 369.

An *antonymous anagram,* or *antigram,* has a meaning opposite to that of the subject text; for example, "within earshot" and "I won't hear this." *Transposed couplets,* or *pairagrams,* are single word anagrams that, when placed together, create a short meaningful phrase, such as "best bets" and "lovely volley." A rare *transposed triplet,* or *tri-anagram,* is "discounter introduces reductions." See also **pangram**.

anallagmatic curve

A curve that is invariant under inversion (see **inverse**). Examples include the **cardioid**, **Cassinian ovals**, **limaçon of Pascal**, **strophoid**, and **Maclaurin trisectrix**.

analysis

A major branch of mathematics that has to do with *approximating* certain mathematical objects, such as numbers or **function**s, in terms of other objects that are easier to understand or to handle. A simple example of analysis is the calculation of the first few decimal places of **pi** by writing it as the **limit** of an **infinite series**. The origins of analysis go back to the seventeenth century, when people such as Isaac **Newton** began investigating how to approximate locally—in the neighborhood of a point—the behavior of quantities that vary continuously. This led to an intense study of limits, which form the basis of understanding infinite series, differentiation, and integration.

Modern analysis is subdivided into several areas: real analysis (the study of derivatives and integrals of real-valued functions); functional analysis (the study of spaces of functions); **harmonic analysis** (the study of **Fourier series** and their abstractions); **complex analysis** (the study of functions from the complex plane to the complex plane that are complex differentiable); and **nonstandard analysis** (the study of **hyperreal number**s and their functions, which leads to a rigorous treatment of infinitesimals and of infinitely large numbers).

analytical geometry

Also known as *coordinate geometry* or *Cartesian geometry,* the type of geometry that describes points, lines, and shapes in terms of **coordinate**s, and that uses **algebra** to prove things about these objects by considering their coordinates. René **Descartes** laid down the foundations for analytical geometry in 1637 in his *Discourse on the Method of Rightly Conducting the Reason in the Search for Truth in the Sciences,* commonly referred to as *Discourse on Method.* This work provided the basis for **calculus**, which was introduced later by Isaac **Newton** and Gottfried **Leibniz**.

analytical number theory

The branch of **number theory** that uses methods taken from **analysis**, especially **complex analysis**. It contrasts with **algebraic number theory**.

anamorphosis

The process of distorting the perspective of an image to such an extent that its normal appearance can only be restored by the observer completely changing the way he looks at the image. In *catoptric anamorphosis,* a curved mirror, usually of cylindrical or conical shape, is used to restore an anamorphic picture to its undistorted form. In other kinds of anamorphism, the observer has to change her viewing position—for example, by looking at the picture almost along its surface. Some anamorphic art adds deception by concealing the distorted image in an

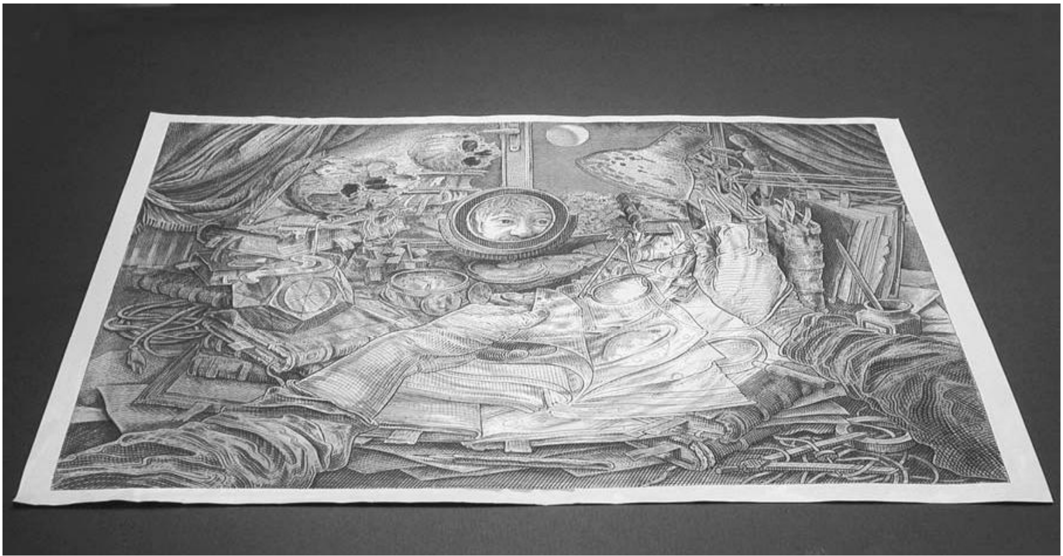

anamorphosis “Self-portrait with Albert” is a clever example of anamorphic art by the Hungarian artist Istvan Orosz. The artist’s hands over his desk and a small round mirror in which the artist’s face is reflected can be seen in the etching. *Istvan Orosz*

(continued)

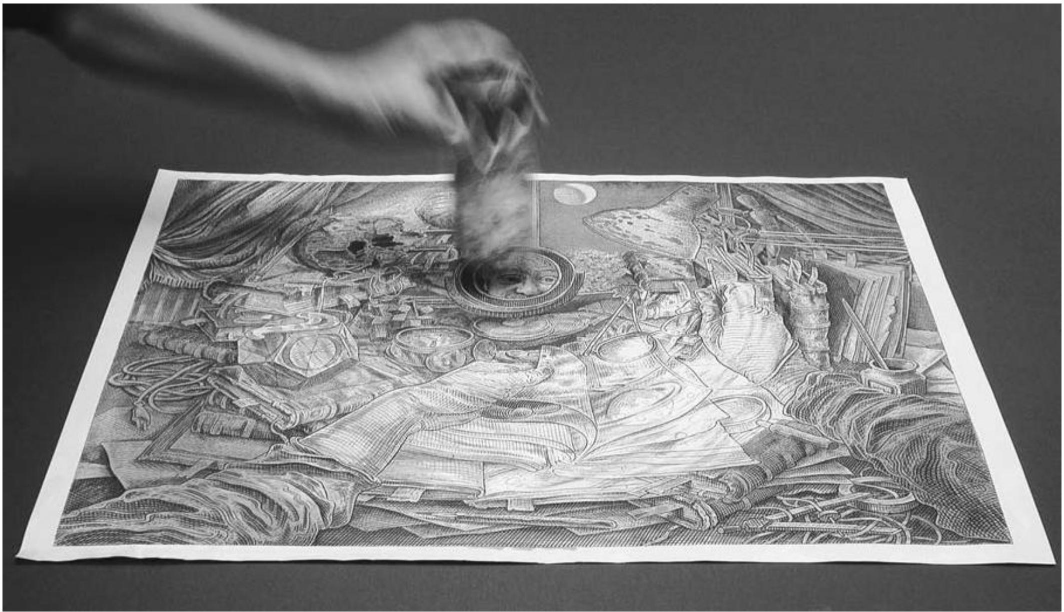

anamorphosis *(continued)* A cylindrical mirror is placed over the circle. *Istvan Orosz*

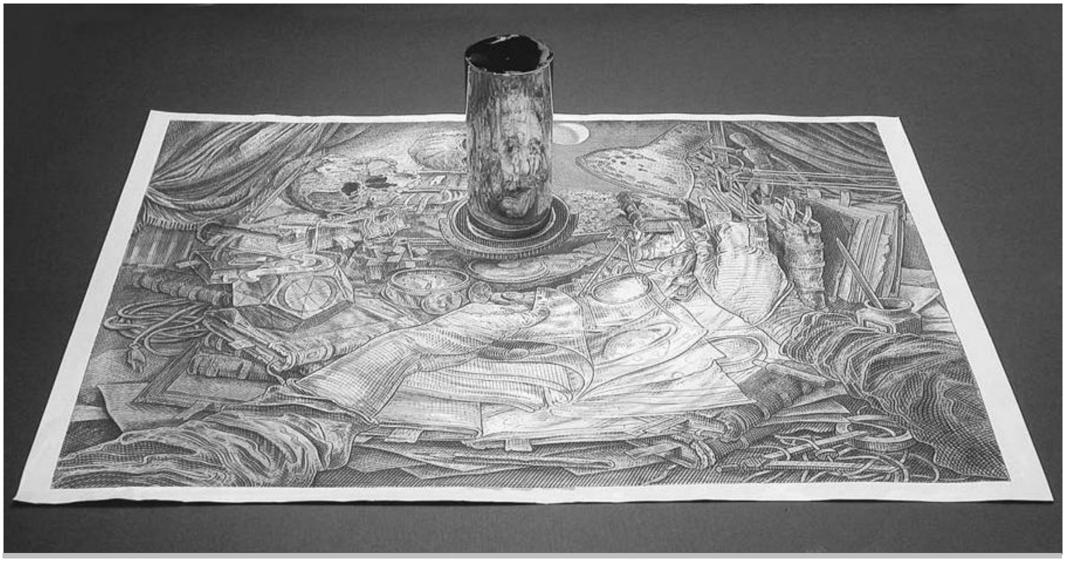

The mirror reveals a previously unsuspected aspect of the picture. The distorting effect of the curved mirror is to undistort a face hidden amid the shapes on the desk: the face of Albert Einstein. Orosz created this etching for an exhibition in Princeton, where the great scientist lived. *Istvan Orosz* otherwise normal looking picture. At one time, artists who had the mathematical knowledge to create anamorphic pictures kept their calculations and grids well-guarded secrets. Now it is relatively easy to create such images by computer.

angle

> *My geometry teacher was sometimes acute, and sometimes obtuse, but he was always right.*
>
> –Anonymous

The opening between two lines or two planes that meet; the word comes from the Latin *angulus* for "sharp bend." Angles are measured in **degree**s. A right angle has 90°, an **acute** angle less than 90°, and an **obtuse** angle has between 90° and 180°. If an angle exceeds the straight angle of 180°, it is said to be convex. *Complementary angles* add to 90°, and *supplementary angles* make a total of 180°.

angle bisection

See **bisecting an angle**.

angle trisection

See **trisecting an angle**.

animals' mathematical ability

Many different species, including rats, parrots, pigeons, raccoons, and chimpanzees, are capable of doing simple calculations. Tests on dogs have shown that they have a basic grasp of cardinality–the number of things on offer. If they're shown a pile of treats and then shown the pile again after it has been concealed and the number of treats changed slightly, they will react differently than if there's been no change. However, not all purported animal math talents stand the test of time. At the turn of the century, a horse named Clever Hans wowed audiences with his counting skills. His trainer would pose a problem, and the horse would tap out the answer. In the end, though, it was found that Hans couldn't really add or subtract but was instead responding to subtle, unintended clues from its trainer, who would visibly relax when the horse reached the correct number.

annulus

The region between the smaller and the larger of two circles that share a common center.

antigravity houses and hills

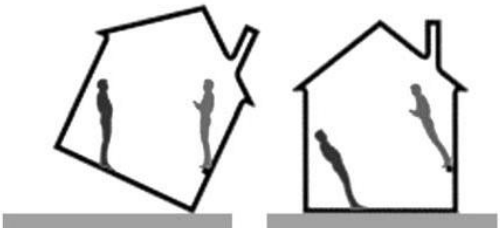

antigravity houses and hills Visitors to a "house of mystery," believing that the floor of the house is horizontal, may be astonished by the apparent gravity-defying effects (right). All these effects are easily understood, however, when it is realized that the entire house tilts at the same angle as a hill on which it is built.

The House of Mystery in the Oregon Vortex, Gold Hill, Oregon, built during the Great Depression in the 1930s, can claim to be the first "antigravity house." It spawned many imitators around the United States and in other parts of the world. Such buildings give rise to some spectacular visual effects, which seem bewildering until the underlying cause is revealed. Of course, the visitor guides are not forthcoming about what is really going on and make fantastic claims about magnetic or gravitational anomalies, UFOs, or other weird and wonderful phenomena. The fact is that all the stunning effects stem from clever construction and concealment that make an incline seem like a horizontal in the mind of the visitor. All antigravity houses are built on hills, with a typical incline of about 25°. But unlike a normal house on the side of a hill, an antigravity house is built so that its walls are perpendicular to the (inclined) ground. In addition, the area around the house is surrounded by a tall fence that prevents the visitor from establishing a true horizontal. Thus compelled to fall back on experience, the visitor assumes that the floor of the house is horizontal and that the walls are vertical with respect to Earth's gravity. All the stunning visual shenanigans follow from this.

In addition to man-made antigravity illusions, there are also a number of remarkable natural locations around the world where gravity seems to be out of kilter. One example is the "Electric Brae," known locally as Croy Brae, in Ayrshire, Scotland. This runs the quarter-mile from the bend overlooking the Croy railway viaduct in the west (86 meters above sea level) to the wooded Craigencroy Glen (92 meters above sea level) to the east. While there is actually a slope of 1 in 86 (a rise of 1 meter for every 86 meters horizontally) upward from the bend at the Glen, the configuration of the land on either side of the road creates the illusion that the slope runs the other way. The author is among countless folk who have parked their cars with the brakes off on this stretch of road and been amazed to see it roll apparently uphill. See also **distortion illusion**.

antimagic square

An $n \times n$ square arrangement of the numbers 1 to n^2 such that the totals of the n rows and n columns and two long diagonals form a sequence of $(2n+2)$ consecutive integers. There are no antimagic squares of size 2×2 and 3×3 but plenty of them for larger sizes. Here is a 4×4 example:

$$\begin{matrix} 1 & 13 & 3 & 12 \\ 15 & 9 & 4 & 10 \\ 7 & 2 & 16 & 8 \\ 14 & 6 & 11 & 5 \end{matrix}$$

See also **magic square**.

antiprism

A **semi-regular polyhedron** constructed from two *n-sided* **polygon**s and $2n$ triangles. An antiprism is like a **prism** in that it contains two copies of any chosen regular polygon, but is unlike a prism in that one of the copies is given a slight twist relative to the other. The polygons are connected by a band of triangles pointing alternately up and down. At each vertex, three triangles and one of the chosen polygons meet. By spacing the two polygons at the proper distance, all the triangles become equilateral. Antiprisms are named square antiprisms, pentagonal antiprisms, and so on. The simplest, the triangular antiprism, is better known as the **octahedron**.

aperiodic tiling

A **tiling** made from the same basic elements or tiles that can cover an arbitrarily large surface without ever exactly repeating itself. For a long time it was thought that whenever tiles could be used to make an aperiodic tiling, those same tiles could also be fitted together in a different way to make a **periodic tiling**. Then, in the 1960s, mathematicians began finding sets of tiles that were uniquely aperiodic. In 1966, Robert Berger produced the first set of 20,426 aperiodic tiles, and soon lowered this number to 104. Over the next few years, other mathematicians reduced the number still further.

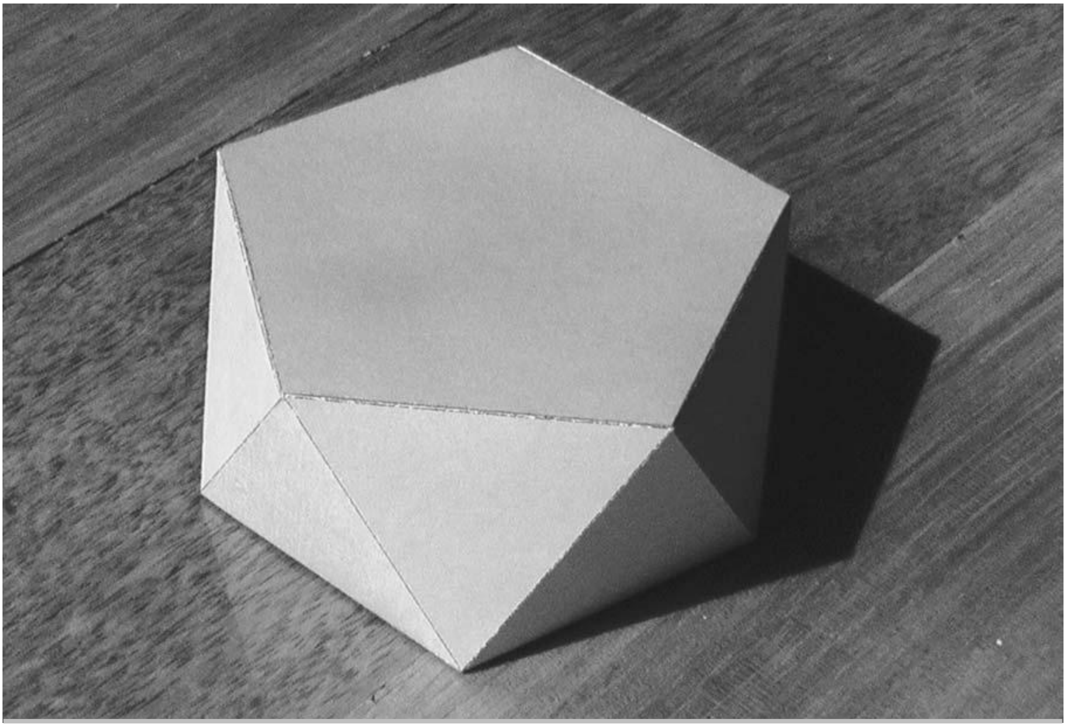

antiprism A pentagonal antiprism. *Robert Webb, www.software3d.com; created using Webb's Stella program*

In 1971, Raphael Robinson found a set of six aperiodic tiles based on notched squares; then, in 1974, Roger **Penrose** found a set of two colored aperiodic tiles (see **Penrose tilings**). The coloring can be dispensed with if the pieces are notched. There is a set of three convex (meaning no notches) aperiodic tiles, but it isn't known if there is a set of two such tiles or even a single tile (see **Einstein problem**). In three dimensions, Robert Ammann found two aperiodic polyhedra, and Ludwig Danzer found four aperiodic tetrahedra.

Apéry's constant

The number defined by the formula $\zeta/(3) = \sum_{n=1}^{\infty} 1/n^3$, where ζ is the **Riemann zeta function**: It has the value 1.202056 . . . and gives the odds (1 in 1.202056 . . .) of any three positive integers, picked at random, having no common divisor. In 1979, the French mathematician Roger Apéry (1916–1994) stunned the mathematical world with a proof that this number is irrational.[11] Whether it is a **transcendental number** remains an open question.

apex

The **vertex** of a cone or a pyramid.

apocalypse number

See **beast number**.

Apollonius of Perga (c. 255–170 B.C.)

A highly influential Greek mathematician (born in a region of what is now Turkey), known as the "Great Geometer," whose eight-part work *On Conics* introduced such terms as *ellipse, parabola,* and *hyperbola.* **Euclid** and others had written earlier about the basic properties of **conic sections** but Apollonius added many new results, particularly related to **normals** and **tangents** to the various conic curves. One of the most famous questions he raised is known as the **Apollonius problem**. He also wrote widely on other subjects including science, medicine, and philosophy. In *On the Burning Mirror* he showed that parallel rays of light are not brought to a focus by a spherical mirror (as had been previously thought), and he discussed the focal properties of a parabolic mirror. A few decades after his death, Emperor Hadrian collected Apollonius's works and ensured their publication throughout his realm.

Apollonius problem

A problem first recorded in *Tangencies,* written around 200 B.C. by **Apollonius of Perga**. Given three objects in the plane, each of which may be a circle *C,* a point *P* (a degenerate circle), or a line *L* (part of a circle with infinite radius), find another circle that is tangent to (just touches) each of the three. There are ten cases: *PPP, PPL, PLL, LLL, PPC, PLC, LLC, LCC, PCC, CCC.* The two easiest involve three points or three straight lines and were first solved by **Euclid**. Solutions to the eight other cases, with the exception of the three-circle problem, appeared in *Tangencies;* however, this work was lost. The most difficult case, to find a tangent circle to any three other circles, was first solved by the French mathematician François Viète (1540–1603) and involves the simultaneous solution of three **quadratic** equations, although, in principle, a solution could be found using just a compass and a straightedge. Any of the eight circles that is a solution to the general three-circle problem is called an *Apollonius circle.* If the three circles are mutually tangent then the eight solutions collapse to just two, which are known as **Soddy circles**. A **fractal** is produced by starting with three mutually tangent circles and creating a fourth—the inner Soddy circle—that is nested between the original three. The process is repeated to yield three more circles nested between sets of three of these, and then repeated again indefinitely. The points that are never inside a circle form a fractal set called the *Apollonian gasket,* which has a fractional dimension of about 1.30568.

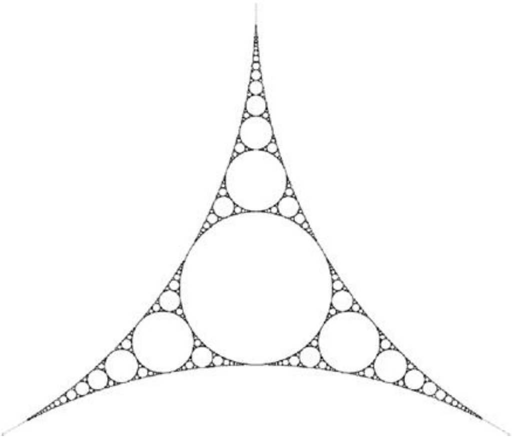

Apollonius problem The Apollonian gasket.

apothem

Also known as a *short radius,* the perpendicular distance from the center of a **regular polygon** to one of its sides. It is the same as the radius of a circle inscribed in the polygon.

apotome

One of Euclid's categories of **irrational numbers**. An apotome has the form $\sqrt{(\sqrt{A} - \sqrt{B})}$. The corresponding number with a "+" sign is called a *binomial* in Euclid's scheme.

applied mathematics

Mathematics for the sake of its use to science or society.

Arabic numeral

A **numeral** written with an Arabic digit alone: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, or in combination: 10, 11, 12, . . . 594, ...

arbelos

A figure bounded by three semicircles *AB, BC,* and *AC,* where *ABC* is a straight line. Archimedes (about 250 B.C.) called it an *arbelos*–the Greek word for a knife of the same shape used by shoemakers to cut and trim leather–and wrote about it in his *Liber assumptorum* (Book of lemmas). Among its properties are that the sum of the two smaller arc lengths is equal to the larger; the area of the arbelos is π/4 times the product of the two smaller diameters (*AB* and *BC*); and the area of the arbelos is equal to the area of a circle whose diameter is the length of a perpendicular segment drawn from the tangent point *B* of the two smaller semicircles to the point *D,* where it meets the larger semicircle. The circles inscribed on each half of *BD* of the arbelos (called *Archimedes's circles*) each have a diameter of $(AB)(BC)/(AC)$. Furthermore, the smallest circumcircle of these two circles has an area equal to that of the arbelos. **Pappus of Alexandria** wrote on the relations of the chain of circles, C_1, C_2, C_3, . . . (called a *Pappus chain* or an *arbelos train*) that are mutually tangent to the two largest semicircles and to each other. The centers of these circles lie on an ellipse and the diameter of the *n*th circle is $(1/n)$ times the base of the perpendicular distance to the base of the semicircle.

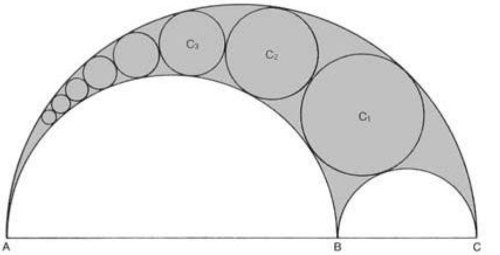

arbelos A Pappus chain of circles, C_1, C_2, C_3, . . . , inside an arbelos (shaded region).

arc

Any part of a curved line or part of the circumference of a circle; the word comes from the Latin *arcus* for a bow, which also gives rise to *arch*. *Arc length* is the distance along part of a curve.

arch

A strong, curved structure, traditionally made from wedge-shaped elements, that may take many different forms and that provides both an opening and a support for overlying material. Two common forms are the semicircular arch, first used by the Romans, and the pointed Gothic arch. The semicircular arch is the weaker of the two because it supports all the weight on the top and tends to flatten at its midpoint. It also requires massive supporting walls since all the stress on the arch acts purely downward. The pointed arch, by contrast, directs stresses both vertically and horizontally, so that the walls can be thinner, though buttressing may be required to prevent the walls from collapsing sideways. See also **Vesica Piscis**.

Archimedean dual

See **Catalan solid**.

Archimedean solid

A convex **semi-regular polyhedron**; a solid made from regular polygonal sides of two or more types that meet in a uniform pattern around each corner. (A regular polyhedron, or **Platonic solid**, has only one type of polygonal side.) There are 13 Archimedean solids (see table "Archimedian Solids"). Although they are named after their discoverer, the first surviving record of them is in the fifth book of the *Mathematical Collection* of **Pappus of Alexandria**. The **dual**s of the Archimedean solids (made by replacing each face with a vertex, and each vertex with a face) are commonly known as **Catalan solid**s. Apart from the Platonic and Archimedean solids, the only other convex uniform polyhedra with regular faces are **prism**s and **antiprism**s. This was shown by Johannes **Kepler**, who also gave the names generally used for the Archimedean solids. See also **Johnson solid**.

Archimedean spiral

A **spiral**, like that of the groove in a phonograph record, in which the distance between adjacent coils, measured radially out from the center, is constant. **Archimedes** was the first to study it and it was the main subject of his treatise *On Spirals*. The Archimedean spiral has a very simple equation in **polar coordinates** (*r*, θ):

$$r = a + b\theta$$

where *a* and *b* can be any real numbers. Changing the parameter *a* turns the spiral, while *b* controls the distance

Archimedean Solids

Name	Number of Vertices	Faces	Edges
Truncated tetrahedron	8 = 4 + 4	12	18
Truncated cube	14 = 8 + 6	24	36
Truncated octahedron	14 = 6 + 8	24	36
Truncated dodecahedron	32 = 20 + 12	60	90
Truncated icosahedron	32 = 12 + 20	60	90
Cuboctahedron	14 = 8 + 6	12	24
Icosidodecahedron	32 = 20 + 12	30	60
Snub dodecahedron	92 = 80 + 12	60	150
Rhombicuboctahedron	26 = 8 + 18	24	48
Great rhombicosidodecahedron	62 = 30 + 20 + 12	120	180
Rhombicosidodecahedron	62 = 20 + 30 + 12	60	120
Great rhombicuboctahedron	26 = 8 + 12 + 6	48	72
Snub cube	38 = 32 + 6	24	60

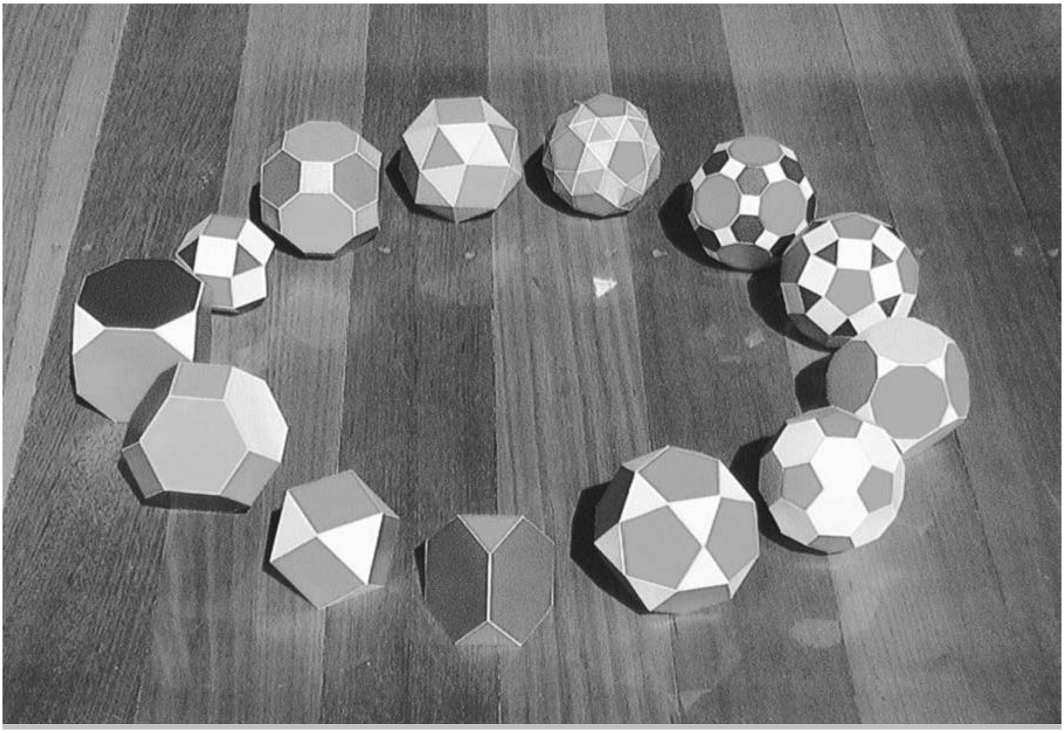

Archimedean solid The complete set of Archimedean solids, starting far left and going clockwise: truncated cube, small rhombicuboctahedron, great rhombicuboctahedron, snub cube, snub dodecahedron, great rhombicosidodecahedron, small rhombicosidodecahedron, truncated dodecahedron, truncated icosahedron (soccer ball), icosidodecahedron, truncated tetrahedron, cuboctahedron, and truncated octahedron. *Robert Webb, www.software3d.com; created using Webb's Stella program* between the arms. The Archimedean spiral is distinguished from the **logarithmic spiral** by the fact that successive arms have a fixed distance (equal to $2\pi b$ if θ is measured in radians), whereas in a logarithmic spiral these distances form a **geometric sequence**. Note that the Archimedean spiral has two possible arms that coil in opposite directions, one for $\theta > 0$ and the other for $\theta < 0$. Many examples of spirals in the man-made world, such as a watch spring or the end of a rolled carpet, are either Archimedean spirals or another curve that is very much like it, the **circle involute**.

Archimedean tessellation

Also known as a *semiregular tessellation,* a **tiling** that uses only regular **polygon**s arranged so that two or more different polygons are around each vertex and each vertex involves the same pattern of polygons. There are eight such tessellations, two involving triangles and squares, two involving triangles and hexagons, and one each involving squares and octagons; triangles and dodecagons; squares, hexagons, and dodecagons; and triangles, squares, and hexagons.

Archimedes of Syracuse (c. 287–212 B.C.)

One of the greatest mathematicians and scientists of all time. He became a popular figure because of his involvement in the defense of Syracuse against the Roman siege in the first and second Punic Wars when his war machines helped keep the Romans at bay. He also devised a scheme to move a full-size ship, complete with crew and cargo, by pulling a single rope, and invented the irrigation device known as the Archimedean screw. According to one of many legends about him, he is said to have discovered the principle of buoyancy while taking a bath and then ran into the street naked shouting “eureka” (“I found it!”).

In his book *The Sand-Reckoner,* he described a positional **number system** and used it to write the equivalent of numbers up to 8×10^{64}–the number of grains of sand he thought it would take to fill the universe. He devised a rule-of-thumb method to do private calculations that closely resembles integral calculus (2,000 years before its “discovery”), but then switched to geometric proof for his results. He demonstrated that the ratio of a circle’s perimeter to its diameter is the same as the ratio of the circle’s area to the square of the radius. Although he didn’t call this ratio “**pi**,” he showed how to work it out to arbitrary accuracy and gave an approximation of it as “exceeding 3 in less than $1/7$ but more than $10/71$.”

Archimedes was the first, and possibly the only, Greek mathematician to introduce mechanical curves (those traced by a moving point) as legitimate objects of study, and he used the **Archimedean spiral** to square the circle. He proved that the area and volume of the **sphere** are in the same ratio to the area and volume of a circumscribed straight cylinder, a result that pleased him so much that he made it his epitaph. Archimedes is probably also the first mathematical physicist on record, and the best before Galileo and Isaac **Newton**. He invented the field of statics, enunciated the law of the lever, the law of equilibrium of fluids, and the law of buoyancy, and was the first to identify the concept of center of gravity. He is also, perhaps erroneously, credited with the invention of a square **dissection** puzzle known as the **loculus of Archimedes**. Many of his original works were lost when the library at Alexandria burned down and they survive only in Latin or Arabic translations. Plutarch wrote of him: “Being perpetually charmed by his familiar siren, that is, by his geometry, he neglected to eat and drink and took no care of his person; that he was often carried by force to the baths, and when there he would trace geometrical figures in the ashes of the fire, and with his finger draws lines upon his body when it was anointed with oil, being in a state of great ecstasy and divinely possessed by his science.”

Archimedes’s cattle problem

A fiendishly hard problem involving very **large numbers** that **Archimedes** presented in a 44-line letter to **Eratosthenes,** the chief librarian at Alexandria. It ran as follows:

> If thou art diligent and wise, O stranger, compute the number of cattle of the Sun, who once upon a time grazed on the fields of the Thrinacian isle of Sicily, divided into four herds of different colors, one milk white, another a glossy black, a third yellow and the last dappled. In each herd were bulls, mighty in number according to these proportions: Understand, stranger, that the white bulls were equal to a half and a third of the black together with the whole of the yellow, while the black were equal to the fourth part of the dappled and a fifth, together with, once more, the whole of the yellow. Observe further that the remaining bulls, the dappled, were equal to a sixth part of the white and a seventh, together with all of the yellow. These were the proportions of the cows: The white were precisely equal to the third part and a fourth of the whole herd of the black; while the black were equal to the fourth part once more of the dappled and with it a fifth part, when all, including the bulls,

> went to pasture together. Now the dappled in four parts were equal in number to a fifth part and a sixth of the yellow herd. Finally the yellow were in number equal to a sixth part and a seventh of the white herd. If thou canst accurately tell, O stranger, the number of cattle of the Sun, giving separately the number of well-fed bulls and again the number of females according to each color, thou wouldst not be called unskilled or ignorant of numbers, but not yet shalt thou be numbered among the wise.
>
> But come, understand also all these conditions regarding the cattle of the Sun. When the white bulls mingled their number with the black, they stood firm, equal in depth and breadth, and the plains of Thrinacia, stretching far in all ways, were filled with their multitude. Again, when the yellow and the dappled bulls were gathered into one herd they stood in such a manner that their number, beginning from one, grew slowly greater till it completed a triangular figure, there being no bulls of other colors in their midst nor none of them lacking. If thou art able, O stranger, to find out all these things and gather them together in your mind, giving all the relations, thou shalt depart crowned with glory and knowing that thou hast been adjudged perfect in this species of wisdom.

The answer to the first part of the problem—the smallest solution for the total number of cattle—turns out to be 50,389,082. But when the extra two constraints in the second part are factored in, the solution is vastly larger. The approximate answer of 7.76×10^{202544} was found in 1880 by A. Amthor, having reduced the problem to a form called a **Pell equation**.[9] His calculations were continued by an ad hoc group called the Hillsboro Mathematical Club, of Hillsboro, Illinois, between 1889 and 1893. The club's three members (Edmund Fish, George Richards, and A. H. Bell) calculated the first 31 digits and the last 12 digits of the smallest total number of cattle to be

77602714064868182695302328332**09 . . .** 719455081800

though the two digits in bold should be 13.[31] In 1931, a correspondent to the *New York Times* wrote: "Since it has been calculated that it would take the work of a thousand men for a thousand years to determine the complete [exact] number [of cattle], it is obvious that the world will never have a complete solution." But *obvious* and *never* are words designed to make fools of prognosticators. Enter the computer. In 1965, with the help of an IBM 7040, H. C. Williams, R. A. German, and C. R. Zarnke reported a complete solution to the cattle problem, though it was 1981 before all 202,545 digits were published, by Harry Nelson, who used a Cray-1 supercomputer to generate the answer, which begins: $7.760271406486818269530232833213 \ldots \times 10^{202544}$.[341]

Archimedes's square

See **loculus of Archimedes**.

area

A measure of surface extension in two-dimensional space. *Area* is the Latin word for a vacant piece of level ground and still carries this common meaning. The French shortened form *are* denotes a square of land with a side length of 10 meters, that is, an area of 100 square meters. A *hectare* is a hundred are.

area codes

North American telephone area codes seem to have been chosen at random. But there was a method to their selection. In the mid-1950s when direct dialing of long-distance calls first became possible, it made sense to assign area codes that took the shortest time to dial to the larger cities. Almost all calls were from rotary dials. Area codes such as 212, 213, 312, and 313 took very little time for the dial to return to its starting position compared, for example, to numbers such as 809, 908, 709. The quickest-to-dial area codes were assigned to the places expected to receive the most direct-dialed calls. New York City got 212, Chicago 312, Los Angeles 213, and Washington, D.C., 202, which is a little longer to dial than 212, but much shorter than others. In order of decreasing size and estimated amount of telephone traffic, the numbers grew larger: San Francisco got 415, Miami 305, and so on. At the other end of the spectrum came places like Hawaii (the last state annexed in 1959) with 808, Puerto Rico with 809, and Newfoundland with 709. The original plan (still in use until about 1993) was that area codes had a certain construction to the numbers: the first digit is 2 through 9, the second digit is 0 or 1, and the third digit is 1 through 9. Three-digit numbers with two zeros are special codes, that is, 700, 800, or 900. Three-digit numbers with two ones are for special local codes such as 411 for local directory assistance, 611 for repairs, and so forth.

Argand diagram

A way of representing **complex number**s as points on a coordinate plane, also known as the *Argand plane* or the *complex plane,* using the *x*-axis as the real axis and the *y*-axis as the imaginary axis. It is named for the French amateur mathematician Jean Robert Argand

(1768–1822) who described it in a paper in 1806.[14] John **Wallis** suggested a similar method 120 years earlier and Casper **Wessel** extensively developed it. But Wessel's paper was published in Danish and wasn't circulated in the languages more common to mathematics at that time. In fact, it wasn't until 1895 that his paper came to the attention of the mathematical community–long after the name "Argand diagram" had stuck.

argument

(1) The input for a **function**. (2) The angle between ZO, where Z is the point representing a complex number on an **Argand diagram** and O is the origin, and the real axis. (3) A mathematical proof, possibly an informal one.

Aristotle's wheel

A **paradox** mentioned in the ancient Greek text *Mechanica,* whose author is unknown but is suspected by some to have been Aristotle. The paradox concerns two concentric circles on a wheel, as shown in the diagram. A **one-to-one** correspondence exists between points on the larger circle and those on the smaller circle. Therefore, the wheel should travel the same distance regardless of whether it is rolled from left to right on the top straight line or on the bottom one. This seems to imply that the two circumferences of the different-sized circles are equal, which is impossible. How can this apparent contradiction be resolved? The key lies in the (false) assumption that a one-to-one correspondence of points means that two curves must have the same length. In fact, the cardinalities of points in a line segment of any length (or even an infinitely long line or an infinitely large n-dimensional Euclidean space) are all the same. See also **infinity**.

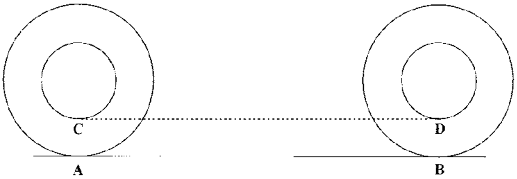

Aristotle's wheel The outer circle turns once when going from A to B, as does the inner circle when going from C to D. Yet AB is the same length as CD. How can this be, since the circles are a different size?

arithmetic

A branch of mathematics concerned with doing calculations with numbers using addition, subtraction, multiplication, and division.

arithmetic mean

The sum of n given numbers divided by n. See also **geometric mean** and **harmonic mean**.

arithmetic sequence

Also known as an *arithmetic progression,* a finite **sequence** of at least three numbers, or an infinite sequence, whose terms differ by a constant, known as the *common difference.* For example, starting with 1 and using a common difference of 4 we can get the finite arithmetic sequence: 1, 5, 9, 13, 17, 21, and also the infinite sequence 1, 5, 9, 13, 17, 21, 25, 29, . . . , $4n + 1$, ... In general, the terms of an arithmetic sequence with the first term a_0 and common difference d, have the form $a_n = dn + a_0$ (n = 1, 2, 3, . . .). Does every increasing sequence of integers have to contain an arithmetic progression? Surprisingly, the answer is no. To construct a counterexample, start with 0. Then for the next term in the sequence, take the *smallest possible integer* that doesn't cause an arithmetic progression to form in the sequence constructed thus far. (There must be such an integer because there are infinitely many integers beyond the last term, and only finitely many possible progressions that the new term could complete.) This gives the nonarithmetic sequence 0, 1, 3, 4, 9, 10, 12, 13, 27, 28, ...

If the terms of an arithmetic sequence are added together the result is an *arithmetic series,* $a_0 + (a_0 + d) + \ldots + (a_0 + (n - 1)d)$, the sum of which is given by:

$$S_n = n/2\ (2a_0 + (n - 1)d) = n/2\ (a_0 + a_n).$$

See also **geometric sequence**.

around the world game

See **Icosian game**.

array

A set of numbers presented in a particular pattern, usually a grid. Matrices (see **matrix**) and **vector**s are examples of arrays.

Arrow paradox

The oldest and best-known **paradox** related to voting. The American economist Kenneth Arrow (1921–) showed that it is impossible to devise a perfect democratic voting system. In his book *Social Choice and Individual Values,*[16] Arrow identified five conditions that are universally regarded as essential for any system in which social decisions are based on individual voting preferences. The Arrow paradox is that these five conditions are logically inconsistent: under certain conditions, at least one of the essential conditions will be violated.

arrowhead

See **dart**.

artificial intelligence (AI)

The subject of "making a machine behave in ways that would be called intelligent if a human were so behaving," according John McCarthy, who coined the term in 1955. How can we tell if a computer has acquired AI at a human level? One way would be to apply the **Turing test**, though not everyone agrees that this test is foolproof (see **Chinese room**). Certainly, AI has not developed at nearly the rate many of its pioneers expected back in the 1950s and 1960s. Meanwhile, progress in fields such as **neural network**s and **fuzzy logic** continues to be made, and most computer scientists have no doubt that it is only a matter of time before computers are outperforming their biological masters in a wide variety of tasks beyond those that call for mere number-crunching ability.

artificial life

A lifelike pattern that may emerge from a **cellular automaton** and appear organic in the way it moves, grows, changes shape, reproduces, aggregates, and dies. Artificial life was pioneered by the computer scientist Chris Langton, and has been researched extensively at the Santa Fe Institute. It is being used to model various **complex system**s such as ecosystems, the economy, societies and cultures, and the immune system. The study of artificial life, though controversial, promises insights into natural processes that lead to the buildup of structure in self-organizing (see **self-organization**) systems.

associative

Three numbers, $x, y,$ and z, are said to be *associative under addition* if

$$x + (y + z) = (x + y) + z,$$

and to be *associative under multiplication* if

$$x \times (y \times z) = (x \times y) \times z.$$

In general, three elements a, b, and c of a set S are associative under the binary operation (an operation that works on two elements at a time) $*$ if

$$a * (b * c) = (a * b) * c.$$

The word incorporates the Greek root *soci,* from which we also get *social,* and may have been first used in the modern mathematical sense by William **Hamilton** around 1850. Compare with **distributive** and **commutative**.

astroid

A **hypocycloid**–the path of a point on a circle rolling inside another circle–for which the radius of the inner circle is four times smaller than that of the larger circle; this ratio results in the astroid having four cusps. The astroid was first studied by the Danish astronomer Ole Römer in 1674, in his search for better shapes for gear teeth, and later by Johann Bernoulli (1691) (see **Bernoulli family**), Gottfried **Leibniz** (1715), and Jean **d'Alembert** (1748). Its modern name comes from the Greek *aster* for "star" and was introduced in a book by Karl Ludwig von Littrow published in Vienna in 1836; before this, the curve had a variety of names, including tetracuspid (still used), cubocycloid, and paracycle. The astroid has the Cartesian equation

$$x^{2/3} + y^{2/3} = r^{2/3}$$

where r is the radius of the fixed outer circle, and $r/4$ is the radius of the rolling circle. Its area is $3\pi r^2/8$, or $3/2$ times that of the rolling circle, and its length is $6r$. The astroid is a sextic curve and also a special form of a **Lamé curve**. It has a remarkable relationship with the quadrifolium (see **rose curve**): the radial, pedal, and orthoptic of the astroid are the quadrifolium, while the cata**caustic** of the quadrifolium is the astroid. The

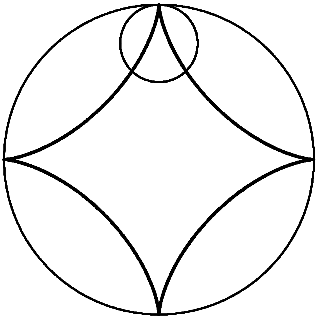

astroid As a small circle rolls around the inside of a larger one with exactly four times its circumference, a point on the rim of the small circle traces out an astroid. *© Jan Wassenaar, www.2dcurves.com* astroid is also the catacaustic of the **deltoid** and the **evolute** of the ellipse.

asymptote

A curve that gets closer and closer to a fixed straight line without ever actually touching it. Imagine facing along the direction of a great wall that is just a meter to your left. Every second, you walk forward a meter and at the same time move sideways slightly so that you halve the distance between you and the wall. The path you follow is an asymptote. The word comes from the Greek roots *a* (not), *sum* (together), and *piptein* (to fall), so that it literally means "not falling together" and was originally used in a wider sense to describe any two curves that don't intersect. **Proclus** writes about both asymptotic lines and symptotic lines (those that do cross). Nowadays, "symptotic" is almost never heard, and "asymptote" is used mainly to denote lines that serve as a limiting barrier for some curve as one of its parameters approaches plus or minus infinity. The "~" symbol is often used to show that one **function** is asymptotic to another. For example, $f(x) \sim g(x)$ indicates that the ratio of the functions $f(x)$ to $g(x)$ approaches 1 as x tends to infinity. Asymptotes are not always parallel to the x- and y-axes, as shown by the graph of $x + 1/x$, which is asymptotic to both the y-axis and the diagonal line $y = x$.

Atiyah, Michael Francis (1929–)

An English mathematician who has contributed to many topics in mathematics, notably dealing with the relationships between geometry and **analysis**. In **topology**, he developed *K-theory.* He proved the *index theorem* on the number of solutions of elliptic differential equations, linking **differential geometry**, topology, and analysis–a theorem that has been usefully applied to quantum theory. Atiyah was influential in initiating work on gauge theories and applications to nonlinear differential equations, and in making links between topology and quantum field theory. Theories of superspace and supergravity, and string theory, were all developed using ideas introduced by him.

Atomium, the

A giant steel monument in Heysel Park, Brussels, Belgium, consisting of 9 spheres that represent the body-centered cubic structure of an iron crystal magnified 150 billion times. Designed by the architect André Waterkeyn and built for the 1958 World's Fair, the 103-meter-high Atomium was originally meant to stand for only 6 months. It may be the world's largest cube. Each of its spheres have a diameter of 18 meters and are connected by escalators. Three of the upper spheres have no vertical support, and so for safety reasons are not open to the public. However, the top sphere offers a panoramic view of Brussels through its windows, and the lower spheres contain various exhibitions.

attractor

A trajectory, or set of points in **phase space**, toward which nearby orbits converge, and which is stable. Specific types of attractor include **fixed-point attractor**, **periodic attractor**, and **chaotic attractor**.

Aubel's theorem

Given a quadrilateral and a square drawn on each side of it, the two lines connecting the centers of the squares on opposite sides are perpendicular and of equal length.

autogram

See **self-enumerating sentence**.

automorphic number

Also known as an *automorph,* a number n whose square ends in n. For instance 5 is automorphic, because $5^2 = 25$, which ends in 5. A number n is called *trimorphic* if n^3 ends in n. For example $49^3 = 117{,}649$, is trimorphic. Not all trimorphic numbers are automorphic. A number n is called *tri-automorphic* if $3n^2$ ends in n; for example 6,667 is tri-automorphic because $3 \times 667^2 = 133{,}346{,}667$ ends in 7.

automorphism

An **isomorphism** from a **set** onto itself. An *automorphism group* of a **group** G is the group formed by the automorphisms of G (bijections from G to itself that preserve the multiplication). Similarly, one can consider the automorphism groups of other structures such as **ring**s and **graph**s, by looking at bijections that preserve their mathematical structure.

Avagadro constant

One of the best known examples of a **large number** in science. It is named after the Italian physicist Amedio Avagadro (1776–1856) and is defined as the number of carbon atoms in 12 grams of pure carbon, or, more generally as the number of atoms of n grams in an element with atomic weight n. It has the value $6.02214199 \times 10^{23}$.

average

A vague term that usually refers to the **arithmetic mean** but can also signify the **median**, **mode**, **geometric mean**, or weighted mean. The word stems from a commercial practice of the shipping age. The root *aver* means to declare, and the shippers of goods would declare the value of their goods. When the goods were sold, a deduction was made from each person's share, based on their declared value, for a portion of the loss or "average."

axiom

A statement that is considered to be true without need of **proof**. The term *axiom* comes from the Greek *axios* meaning "worthy" and was used by many Greek philosophers and mathematicians, including **Aristotle**. Curiously, **Euclid**, whose axioms are best known of all, seems to have favored a more general phrase meaning "common notion."

axiom of choice

An **axiom** in **set theory** that is one of the most controversial axioms in mathematics; it was formulated in 1904 by the German mathematician Ernst Zermelo (1871–1953) and, at first, seems obvious and trivial. Imagine there are many–possibly an unlimited number of–boxes in front of you, each of which has at least one thing in it. The axiom of choice (AC) says simply that you can always choose one item out of each box. More formally, if $\boldsymbol{S}$ is a collection of nonempty sets, then there exists a set that has exactly one element in common with every set S of $\boldsymbol{S}$. Put another way, there exists a function f with the property that, for each set S in the collection, $f(S)$ is a member of S. Bertrand **Russell** summed it up neatly: "To choose one sock from each of infinitely many pairs of socks requires the Axiom of Choice, but for shoes the Axiom is not needed." His point is that the two socks in a pair are identical in appearance, so, to pick one of them, we have to make an arbitrary choice. For shoes, we can use an explicit rule, such as "always choose the left shoe." Russell specifically mentions *infinitely many* pairs, because if the number is finite then AC is superfluous: we can pick one member of each pair using the definition of "nonempty" and then repeat the operation finitely many times using the rules of formal logic.

AC lies at the heart of a number of important mathematical arguments and results. For example, it is equivalent to the *well-ordering principle,* to the statement that for any two **cardinal number**s m and n, then $m < n$ or $m = n$ or $m > n$, and to *Tychonoff's theorem* (the product of any collection of compact spaces in topology is compact). Other results hinge upon it, such as the assertion that every infinite set has a denumerable subset. Yet AC was strongly attacked when it was first suggested, and still makes some mathematicians uneasy. The central issue is what it means to *choose* something from the sets in question and what it means for the choosing function to *exist.* This problem is brought into sharp focus when $\boldsymbol{S}$ happens to be the collection of all nonempty subsets of the **real number**s. No one has ever found a suitable choosing function for this collection, and there are good reasons to suspect that no one ever will. AC just mandates that there *is* such function. Because AC conjures up sets without offering workable procedures, it is said to be *nonconstructive,* as are any theorems whose proofs involve AC. Another reason that some mathematicians aren't greatly enamored with AC is that it implies the existence of some bizarre counterintuitive objects, the most famous and notorious example of which is the **Banach-Tarski paradox**. The main reason for accepting AC, as the majority of mathematicians do (albeit often reluctantly), is that it is useful. However, as a result of work by Kurt **Gödel** and, later, by Paul Cohen, it has been proven to be independent of the remaining axioms of set theory. Thus there are no contradictions in choosing to reject it; among the alternatives are to adopt a contradictory axiom or to use a completely different framework for mathematics, such as **category theory**.

axis

A line with respect to which a curve or figure is drawn, measured, rotated, and so forth. The word comes from the Greek root *aks* for a point of turning or rotation and seems to have first been used in English by Thomas Digges around 1570 in reference to the rotational axis of a right circular cone.

B

Babbage, Charles (1791–1871)

> *On two occasions I have been asked [by members of Parliament], "Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?" I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question.*

An English mathematician who served as Lucasian Professor of Mathematics at Cambridge (1828–1839) and became the most important figure in the prehistory of computers. Babbage noted that astronomical and other mathematical tables of the period were riddled with errors because all the calculations had to be done by hand. This gave him the idea of building a machine that would do the tedious work of computation more accurately, faster, and without ever getting tired.

In 1822, Babbage wrote a letter to one of the top British scientists of the day, Humphrey Davy, in which he talked about the design of an automatic calculator. Shortly after, he was given a grant by the British government to build this device–an elaborate symphony of rods and interlocking gear teeth–which Babbage called the Difference Engine. Construction started but never finished. Despite heroic efforts to construct a working model, the critical tolerances were beyond what engineers could provide in the first half of the eighteenth century (though the resultant gear-making skills gave Britain an edge in precision machinery for several decades and even contributed to the qualitative superiority of the British navy in World War I). The government had spent £17,000, and Babbage contributed a similar amount of his own money, on the project, when Babbage set his sights on something even more ambitious. He grasped

Babbage, Charles The realization of Babbage's dream: the Manchester Mark 1 computer at Manchester University, England, in 1948. This was the first computer that could store both data and programs electronically. *Ferranti Electronics Ltd.*

that the basic mechanisms of the Difference Engine could be generalized to an all-purpose calculating machine, programmable by a punched-card mechanism like that of a Jacquard loom. This vastly more powerful machine was called the Analytical Engine and would have been the world's first true computer. But it never got off the ground. "He was ill-judged enough," wrote the secretary of the Royal Astronomical Society, "to press the consideration of this new machine upon the members of Government, who were already sick of the old one." Prime Minister Robert Peel was less than enthusiastic: "I would like a little previous consideration before I move in a thin house of country gentlemen a large vote for the creation of a wooden man to calculate tables from the formula $x^2 + x + 41$." The government's eventual withdrawal of support for his schemes left Babbage a disappointed and embittered man. However, his ideas survived and proved to be the forerunner of modern computers. Parts of his uncompleted mechanisms are on display in the London Science Museum. In 1991, working from Babbage's original plans, a Difference Engine was completed–and functioned perfectly. Among Babbage's many lesser known accomplishments was his cracking of the Vigenère cipher, a discovery that helped English military campaigns but wasn't published for several years, by which time the credit had gone instead to Friedrich Kasiski, who broke the code some years after Babbage.[329] See also **Byron, Ada**.

Bachet de Méziriac, Claude-Gaspar (1581–1638)

A poet and early mathematician of the French Academy, best known for his 1621 translation of **Diophantus**'s *Arithmetica,* the book that Pierre de **Fermat** was reading when he inscribed the margin with his famous last theorem. Bachet is also remembered as a collector of mathematical puzzles, many of which he published in *Problèmes plaisans et délectables qui font par les nombres* (1612) (Pleasant and delightful problems that involve numbers), including **river-crossing problems**, **measuring and weighing puzzles**, number tricks, and **magic squares**. One of the puzzles is to find the least number of weights that can be used on a scale pan to weigh any integral number of pounds from 1 to 40 inclusive, if the weights can be placed in either of the scale pans. The answer is four: 1, 3, 9, and 27 pounds. On a slightly more serious note, Bachet observed that apparently every positive number can be expressed as a sum of at most four squares; for example, $5 = 2^2 + 1^2$, $6 = 2^2 + 1^2 + 1^2$, $7 = 2^2 + 1^2 + 1^2 + 1^2$, $8 = 2^2 + 2^2$, and $9 = 3^2$. The case of 7 shows that sometimes three squares wouldn't be enough. Bachet said he had checked this for more than 300 numbers but didn't know how to prove it. It wasn't until the late eighteenth century that Joseph **Lagrange** supplied a complete proof.[339]

backgammon

A gambling game for two in which each player seeks to get a set of pieces from one side of the board to the other, while trying to prevent the other player from doing the same. The distance that a piece can be moved at each turn is determined by the throw of dice.

Backgammon has roots stretching back 5,000 years. From Mesopotamia, versions of it spread to Greece and Rome as well as to India and China. The rules of the modern form of the game were largely established in England in 1743 by Edmond Hoyle but benefited from a crucial modification that emerged in American gambling clubs in the 1920s. This final innovation, which added a new level of subtlety, is known as the *doubling cube*.

Backgammon is played with two sets of 15 checkers: one player has black, the other white. The players' checkers move in opposite directions on a board with 24 spaces or *points.* Each player's goal is to be the first to bring all their own checkers "home" (into their own quarter of the board) and then "bear them off" (remove them from the board altogether). The movement of the checkers follows the outcome of a roll of two dice, the numbers on the two dice constituting separate moves.

The actual amount that changes hands at the end of the game can be more than the initial stake. For instance, in certain winning positions called *gammon* and *backgammon,* the stake is doubled or trebled, respectively. The other way the stake can change is by means of the doubling cube. If one of the players thinks that she is in a winning position, she can turn the doubling cube and *announce a double,* which means that the total stake will be doubled. If her opponent *refuses the double,* he immediately loses his (undoubled) stake and the game is finished. If he *accepts the double,* the stakes are doubled and, as a compensation, the doubling cube is handed over to him and he gets the exclusive right to announce the next double. (He is now said to *own the cube.*) If the luck of the game changes so that he later judges that he is now winning, he'll be in a position to announce a *redouble,* which means that the stake is doubled again. If the first player refuses the double, she now loses the doubled stake; if she accepts, the game goes on with a redoubled stake, four times the original value. There's no limit to how many times the stake can be doubled, but the right to announce a double switches from one player to the other every time it's exercised. (Initially either player can double–no one owns the cube.) This aspect of the game adds greatly to the variety of tactical possibilities and problems.

baker's dozen

See **thirteen**.

Bakhshali manuscript

An early mathematical manuscript, written on birch bark and found in the summer of 1881 near the village of Bakhshali in the Yusufzai subdivision of the Peshawar district (now in Pakistan). A large part of the manuscript had been destroyed and only about 70 leaves of birch bark, of which a few were mere scraps, survived to the time of its discovery. Although its date is uncertain, it is most commonly put at about the third or fourth century A.D. and appears to be a commentary on an earlier mathematical work. Among the rules and techniques it sets out for solving problems, mostly in arithmetic and algebra, but also to a lesser extent in geometry and mensuration, is this formula (stated here in modern terms) for calculating the square root of a nonsquare number Q:

$$\sqrt{Q} = \sqrt{(A^2 + b)} = A + b/2A - (b/2A)^2/[2(A + b/2A)]$$

If $Q = 41$ (so that $A = 6$ and $b = 5$) this gives $\sqrt{Q} = 6.403138528$, which compares very favorably with the correct result of 6.403124237.

ball

Mathematicians, unlike the rest of the human race, draw a sharp distinction between a **sphere** and a ball. A sphere (in mathematics) is only a surface, whereas a ball is everything inside, and possibly including, that surface–the filling of the sphere. An *open ball* consists of all the points that are less than a given distance (the radius) away from a given point (the center); a *closed ball* consists of all the points that are less than or equal to the radius.

Mathematical balls can also exist in any number of dimensions. A one-dimensional ball of radius r is just a line segment. It consists of all the points on a line between $-r$ and r, or, in the case of a one-dimensional *unit ball* (a ball with a radius of 1), between -1 and 1. A 1 – d unit ball thus has a length, or "1 – d volume," of 2. A 2 – d unit ball, which is the filling of a unit circle, has an area, or 2 – d volume, of π. The volume of a unit ball in 3 – d is $4/3\pi$. In 4 – d it is $\pi^2/2$. Apparently, as the number of dimensions increases, so does the volume of the unit ball. What does this volume tend to do as the dimension tends to infinity? Intuitively, it might seem that in higher and higher dimensions there's more and more "room" in the unit ball, allowing its volume to become larger and larger. Does the volume become infinite, or does it approach a sufficiently large constant as the dimension increases? The answer is surprising and shows how our intuition is often misleading. Using a technique called *multivariable calculus* the volume of the unit ball in n dimensions, $V(n)$, can be shown to be $\pi^{n/2}/\ \Gamma(n/2 + 1)$, where Γ is the **gamma function** that generalizes the factorial function (i.e., $\Gamma(z + 1) = z!$). For n even, say $n = 2k$, the volume of the unit ball is thus given by $V(n) = \pi^k / k!$.

Ball, Walter William Rouse (1850–1925)

A British mathematician who lectured at Trinity College, Cambridge University, from 1878 to 1905, but is best known as a historian and as the author of the timeless classic *Mathematical Recreations and Essays.*[24] It was first published in 1892 and ran to fourteen editions, the last four with revisions by the great geometer Harold **Coxeter**.

Banach, Stefan (1892–1945)

A great Polish mathematician who founded functional **analysis** and also made important contributions to the understanding of **vector spaces**, **measure theory**, and **set theory**. His name is associated with *Banach space, Banach algebra,* the *Hahn-Banach theorem,* and the remarkable **Banach-Tarski paradox**. Largely self-taught in mathematics, Banach was "discovered" by Hugo **Steinhaus** and when World War II began was president of the Polish Mathematical Society and a full professor at Lvov University. Being on good terms with Soviet mathematicians, he was allowed to hold his chair during the Soviet occupation of Lvov. The German occupation of the city in 1941 resulted in the mass murder of Polish academics. Banach survived, but the only way he could earn a living was by feeding lice with his blood in a German institute where typhoid fever research was conducted. His health declined during the occupation, and Banach died before he could be repatriated from Lvov, which was incorporated into the Soviet Union and returned to Poland after the war. *Théorie des opérations linéaires* (The theory of linear operations) is regarded as his most influential work.

Banach-Tarski paradox

> *There are more things in heaven and earth, Horatio,*
> *than are dreamt of in your philosophy.*
>
> –William Shakespeare

A seemingly bizarre and outrageous claim that it is possible to take a ball, break it into a number of pieces, and then reassemble those pieces to make two identical copies of the ball. The claim can be made even stronger: it is possible to decompose a ball the size of a marble and then reassemble the pieces to make another ball the size of Earth, or, indeed, the size of the known universe!

Before writing off Banach and Tarski as being either very bad mathematicians or very good practical jokers, it's important to understand that this is not a claim about what can actually be done with a real ball, a sharp knife, and some dabs of glue. Nor is there any chance of some entrepreneur being able to slice up a gold ingot and assemble in its place two new ones like the original. The Banach-Tarski paradox tells us nothing new about the physics of the world around us but a great deal about how volume, space, and other familiar-sounding things can assume unfamiliar guises in the strange abstract world of mathematics.

Stefan **Banach** and Alfred **Tarski** announced their startling conclusion in 1924, having built on earlier work by Felix **Hausdorff**, who proved that it's possible to chop up the unit interval (the line segment from 0 to 1) into countably many pieces, slide these bits around, and fit them together to make an interval of length 2. The Banach-Tarski paradox, which mathematicians often refer to as the *Banach-Tarski decomposition* because it's really a proof not a paradox, highlights the fact that among the infinite **set** of points that make up a mathematical **ball**, the concepts of volume and measure can't be defined for all possible subsets. What this boils down to is that quantities that can be measured in any familiar sense are not necessarily preserved when a ball is broken down into subsets and then those subsets reassembled in a different way using just translations (slides) and rotations (turns). These unmeasurable subsets are extremely complex, lacking reasonable boundaries and volume in the ordinary sense, and thus are not attainable in the real world of matter and energy. In any case, the Banach-Tarski paradox doesn't give a prescription for *how* to produce the subsets: it only proves their *existence* and that there must be at least five of them to produce a second copy of the original ball. The fact that the Banach-Tarski paradox depends on the **axiom of choice** (AC), yet is so strongly counterintuitive, has been used by some mathematics to suggest that the AC must be wrong; however, the benefits of adopting the AC are so great that such black sheep of the mathematical family as the paradox are generally tolerated.[324, 340]

Bang's theorem

If all the faces of a **tetrahedron** have the same perimeter, then the faces are all **congruent** triangles.

banker's rounding

For banking or scientific purposes it's often considered correct to round something 0.5 to the nearest *even* number (not always upward). For instance, 5.5 rounds to 6, but 12.5 rounds to 12. This method avoids introducing a bias to a large set of numbers, by rounding up more or less as often as rounding down. Unfortunately, at a lower level, it is often taught to round something 0.5 upward all the time. See also **round-off error**.

Barbaro, Daniele (1513–1570)

A Venetian geometer whose book, *La Practica della Perspectiva* (1568–9), presents the techniques of perspective, illustrated in part with a range of **polyhedra**. Partly based on the methods and writings of the great artist Piero della Francesca (1416–1492), but written in a more readable and humanistic style, it includes the earliest drawing of the *truncated icosidodecahedron* and one of the earliest representations, along with that of the German goldsmith Wenzel Jamnitzer (1508–1585), of a *rhombicosidodecahedron. La Practica* was one of the most respected texts on perspective in the sixteenth century, comparable to Albrecht **Dürer**'s *Painter's Manual.*

barber paradox

See **Russell's paradox**.

Barbier's theorem

See **curve of constant width**.

Barlow, Peter (1776–1862)

A self-educated English mathematician who wrote several important books on the subject but is best known for *New Mathematical Tables* (generally known as *Barlow's Tables*), a compendium of factors, squares, cubes, square roots, reciprocals, and hyperbolic logarithms of all numbers from 1 to 10,000, and his invention of a special telescope lens. "Barlows" are popularly used by amateur astronomers to this day to multiply the power of other lenses. Barlow also worked on the design of bridges and was appointed as royal commissioner for railways, conducting experiments to see if the limitation on gradients and radius of curvature proposed by George Stephenson was correct.

Barnsley's fern

A **fractal** shape, first explored by Michael F. Barnsley at the Georgia Institute of Technology in the 1980s, that has many geometric features in common with a natural fern,

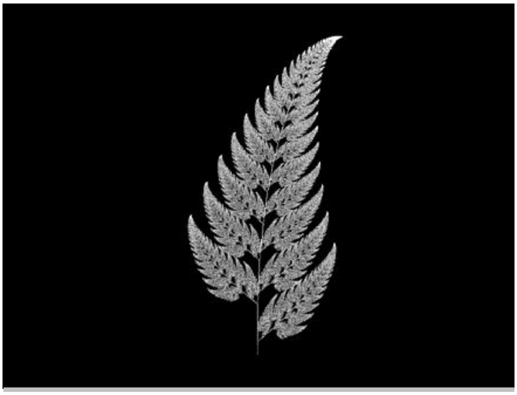

Barnsley's fern *David Nicholls*

most notably the appearance of frondlike forms at different scales. As in the case of real ferns, Barnsley's fern reveals smaller prominences along the edge of each frond that are miniature versions of the overall figure. Along these small prominences are still smaller protuberances, and so on. Barnsley's fern is created by the repetitive application of four relatively simple mathematical rules and is a type of fractal, introduced by Barnsley, known as an *iterated function system* (IFS).[26]

base

(1) The flat plane or straight line upon which a shape or a solid rests. (2) The number upon which a **number system** is based; this is also the number of different characters or figures needed by the number system. The base, or *radix,* of our familiar decimal system is 10. Thus, there are ten symbols, 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9, and a decimal number is written right to left in terms of units, tens, hundreds, and so on. Each move to the left represents a jump by a **power** of 10. The decimal number 375, for example, equals $(3 \times 10^2) + (7 \times 10) + (5 \times 1)$. This can easily be written to another base. Decimal 375_{10} becomes in octal (base 8) $567_8 = (5 \times 8^2) + (6 \times 8) + (7 \times 1)$, or in binary (base 2) 101111001_2.

basin of attraction

The set of all points in **phase space** that are under the influence of an **attractor**, or, more generally, the initial conditions of a system that evolve into the range of behavior allowed by the attractor. If one imagines a **complex system** as a sink, then the attractor can be considered the drain at the bottom, and the basin of attraction is the sink's basin.

basis

In mathematics, usually associated with linear algebra; a minimal set of **vector**s that spans a **vector space**.

Bayes, Thomas (1702–1761)

An English mathematician and theologian, remembered chiefly for the theorem named after him (see **Bayes's theorem**), and the technique of **Bayesian inference** that arises from it. Bayes wrote on probability theory, the logical basis of calculus, and asymptotic series.

Bayesian inference

Statistical inference in which probabilities are interpreted not as frequencies or proportions, but rather as degrees of belief. A prior distribution for a certain random variable is assumed; then this is modified, in the light of experimentation, using **Bayes's theorem**. Pierre **Laplace** applied Bayesian inference to estimate the mass of Saturn and in a variety of other problems.

Bayes's theorem

Also known as *Bayes's rule,* a result in **probability theory**, named after Thomas **Bayes**, who proved a special case of it. It is used in statistical inference to update estimates of the probability that different hypotheses are true, based on observations and a knowledge of how likely those observations are, given each hypothesis. In fact, it is habitually used by scientists in preference to the principle of **induction**. Bayes's theorem says that if an instance X is actually observed, then the probability of a hypothesis H must be multiplied by the following ratio:

$$\frac{\text{probability of observing } X \text{ if } H \text{ is true}}{\text{probability of observing } X}$$

In other words, the probability of a hypothesis H conditional on a given body of data X is equal to the ratio of the unconditional probability of the conjunction of the hypothesis with the data to the unconditional probability of the data alone.

Beale cipher

One of the greatest unsolved puzzles in **cryptography**—or a mere hoax. About a century ago, a fellow by the name of Thomas Beale supposedly buried two wagonloads of pots filled with silver coins in Bedford County, near Roanoke, Virginia. Local rumors claim the treasure was buried near Bedford Lake. Beale wrote three encoded letters telling what was buried, where it was buried, and to whom it belonged. He entrusted these three letters to a friend, went west, and was never heard from again. Several years later, someone examined the letters and was able to crack the code in the second one, which turned out to be based on the text from the Declaration of Independence. A number in the letter indicated which word in the document was to be used. The first letter of that word replaced the number. For example, if the first four words of the document were "We hold these truths," the number 3 in the letter would represent the letter *t.* The second letter translated as follows:

> I have deposited in the county of Bedford about four miles from Bufords in an excavation or vault six feet below the surface of the ground the following articles belonging jointly to the parties whose names are given in number three herewith. The first deposit consisted of ten hundred and fourteen pounds of gold and thirty eight hundred and twelve pounds of silver deposited Nov eighteen nineteen. The second was made Dec eighteen twenty one and consisted of nineteen hundred and seven pounds of gold and twelve hundred and eighty eight of silver, also jewels obtained in St. Louis in exchange to save transportation and valued at thirteen [t]housand dollars. The above is securely

> packed i[n] [i]ron pots with iron cov[e]rs. Th[e] vault is roughly lined with stone and the vessels rest on solid stone and are covered [w]ith others. Paper number one describes th[e] exact locality of the va[u]lt so that no difficulty will be had in finding it.

One of the remaining letters supposedly contains directions on how to find the treasure but, to date, no one has solved the code. One theory is that both the remaining letters are encoded using either the same document in a different way, or another very public document. Or, of course, this could all be an elaborate, but entertaining, wheeze. Those interested may wish to contact the Beale Cypher Association, P.O. Box 975, Beaver Falls, PA 15010.

Beal's conjecture

In 1997, the Texan financier Andrew Beal offered $75,000, later increased to $100,000, to the first person who could prove or provide a counterexample to the following conjecture:

> If $x^m + y^n = z^r$, where x, y, z $m, n,$ and r are all positive integers, and m, n and r are greater than two, then x, y, and z have a common factor (greater than one).

Fermat's last theorem, which was proved in 1994, is a special case of Beal's conjecture. However, no one has yet been able to use this fact to prove or disprove the conjecture, nor has anyone been able to come up with a counterexample as a disproof. It is known that for any set of three exponents m, n, and r, each greater than two, there can be at most finitely many solutions. But is this finite number zero? The prize remains unclaimed.

beast number (666)

Also known as the *Apocalypse number,* the "number of the beast" mentioned in the Bible's book of Revelation, the relevant verse of which (Rev. 13:18) is often quoted as:

> Here is wisdom. Let him that hath understanding count the number of the beast: for it is the number of a man; and his number is six hundred threescore and six.

Leaving aside the thorny issue of what this actually means, the number 666 does have some interesting mathematical properties. Most notably, it is the sum of the first 36 natural numbers (all the numbers on a roulette wheel): $1 + 2 + 3 + \ldots + 36$, which makes it the thirty-sixth **triangular number**. It is also the sum of the squares of the first seven **prime numbers**, $2^2 + 3^2 + 5^2 + 7^2 + 11^2 + 13^2 + 17^2$. Other curious representations of "the beast" include:

$$1^3 + 2^3 + 3^3 + 4^3 + 5^3 + 6^3 + 5^3 + 4^3 + 3^3 + 2^3 + 1^3$$

$$3^6 - 2^6 + 1^6$$

$$6 + 6 + 6 + 6^3 + 6^3 + 6^3.$$

Furthermore, 666 is one member of a **Pythagorean triplet** (216, 630, 666), which can be written in the remarkable form:

$$(6 \times 6 \times 6)^2 + (666 - 6 \times 6)^2 = 666^2.$$

In Roman numerals, 666 represents all the numbers from 500 in descending order, namely D (500) + C (100) + L (50) + X (10) + V (5) + I (1), or DCLXVI. In fact, it's been suggested that the Roman representation of 666 may have something to do with the biblical reference. DCLXVI was often used as a generic way of referring to any unspecified or unknown large number–the Roman equivalent of our "zillion." Thus, the writer of Revelation might simply have been using "666" to mean "big but unspecified."

Beatty sequences

Suppose R is an **irrational number** greater than 1, and let S be the number satisfying the equation $1/R + 1/S = 1$. Let $[x]$ denote the *floor function* of x, that is, the greatest integer less than or equal to x. Then the sequences $[nR]$ and $[nS]$, where n ranges through the set N of positive integers, are the Beatty sequences determined by R. The interesting thing about them is that they partition N; in other words, every positive integer occurs exactly once in one sequence or the other. For example, when R is the **golden ratio** (about 1.618), the two sequences begin with

$$1, 3, 4, 6, 8, 9, 11, 12, 14, 16, 17, 19, 21, \ldots, \text{ and}$$
$$2, 5, 7, 10, 13, 15, 18, 20, 23, 26, 28, 31, 34 \ldots.$$

Beatty sequences are named after the American mathematician Samuel Beatty (1881–1970), who introduced them in 1926 in a problem in the *American Mathematical Monthly.* Beatty was the first person to receive a Ph.D. in mathematics from a Canadian university, and later became the chairman of the mathematics department and chancellor of the University of Toronto.

beauty and mathematics

Many mathematicians and scientists have commented on the beauty they find in the structure and symmetry of the equations that underpin their work, and that beauty is often the first sign of truth. In *A Mathematician's Apology,*[151] G. H. **Hardy** wrote:

> The mathematician's patterns, like the painter's or the poet's must be beautiful; the ideas, like the colors or the words must fit together in a harmonious way. Beauty is the first test: there is no permanent place in this world for ugly mathematics.

The physicist Paul **Dirac** went even further:

> I think that there is a moral to this story, namely that it is more important to have beauty in one's

> equations than to have them fit experiment. If [Erwin] Schrödinger had been more confident of his work, he could have published it some months earlier, and he could have published a more accurate equation. It seems that if one is working from the point of view of getting beauty in one's equations, and if one has really a sound insight, one is on a sure line of progress. If there is not complete agreement between the results of one's work and experiment, one should not allow oneself to be too discouraged, because the discrepancy may well be due to minor features that are not properly taken into account and that will get cleared up with further development of the theory.[176]

The architect Richard Buckminster Fuller also saw beauty as an acid test of truth: "When I'm working on a problem, I never think about beauty. I think only how to solve the problem. But when I have finished, if the solution is not beautiful, I know it is wrong."

Bell, Eric Temple (1883–1960)

A Scottish-born mathematician and writer who, from 1903, spent most of his life in the United States, teaching at the University of Washington from 1912 until 1926, then serving as professor of mathematics at the California Institute of Technology. He did work in **number theory** but is best remembered for his books, including *Algebraic Arithmetic* (1927) and *The Development of Mathematics* (1940), which became classics, and, at a more popular level, *Men of Mathematics* (1937)[32] and *Mathematics, Queen and Servant of Science* (1951). He was also a prolific writer of science fiction under the penname John Taine.

bell curve

The characteristic shape of the graph of a normal (Gaussian) distribution.

Bell number

Named after Eric **Bell,** one of the first to analyze them in depth, the number of ways that *n* distinguishable objects (such as differently colored balls) can be grouped into sets (such as buckets) if no set can be empty. For example, if there are three balls, colored red (R), green (G), and blue (B), they can be grouped in five different ways: (RGB), (RG)(B), (RB)(G), (BG)(R), and (R)(G)(B), so that the third Bell number is 5. The sequence of Bell numbers, 1, 2, 5, 15, 52, 203, 877, 4,140, 21,147, . . . , can be built up in the form of a triangle, as follows. The first row has just the number one. Each successive row begins with the last number of the previous row and continues by adding the number just written down to the number immediately above and to the right of it.

1
1 2
2 3 5
5 7 10 15
15 20 27 37 52
52 . . .

The Bell numbers appear down the left-hand side of the triangle. These normal Bell numbers contrast with *ordered Bell numbers,* which count the number of ways of placing *n* distinguishable objects (balls) into one or more *distinguishable* sets (buckets). The ordered Bell numbers are 1, 3, 13, 75, 541, 4,683, 47,293, 545,835, ... Bell numbers are related to the **Catalan number**s.

Benford's law

If a number is chosen at random from a large table of data or statistics, such as stock quotations, populations of towns in Germany, or half-lives of radioactive atoms, the chance that the first digit is 1 is about 30.1%, that the first digit is 2 is about 17.6%, that it is 3 is 12.4%, . . . , and that it is 9 is 4.5%. These figures fit the rule that the probability that the first digit is d is $\log_{10}(1 + 1/d)$. This rule is called Benford's law after the American physicist Frank Benford, who publicized his findings in 1938.[34] The same discovery had been made 57 years earlier by the astronomer and mathematician Simon Newcomb, who noticed that the front pages of logarithm tables tended to be more dog-eared than pages later on.[232]

Benford tested thousands of different collections of data, including the surface areas of 335 rivers, specific heats and molecular weights of thousands of chemicals, baseball statistics, and the street addresses of the first 342 people listed in the book *American Men of Science.* All these seemingly unrelated sets of numbers followed the same first-digit probability pattern as the worn pages of logarithm tables suggested. In all cases, the number 1 showed up as the first digit about 30% of the time, more often than any other, and seven times more often than the number 9.

It seems extraordinary. Why shouldn't the numbers 1 to 9 take equal turns to be the first digit? Benford's findings have been verified by other researchers. The larger and more varied the sampling of numbers from different data sets, it has been found, the more closely the distribution of numbers approaches what Benford's law predicts. Moreover these probabilities are *scale invariant* and *base invariant.* For example, it doesn't matter whether the numbers are based on the dollar prices of stocks or their prices in yen or euros, nor does it matter if the numbers are in terms of stocks per dollar; provided there are enough numbers in the sample, Benford's law will hold.[164, 263]

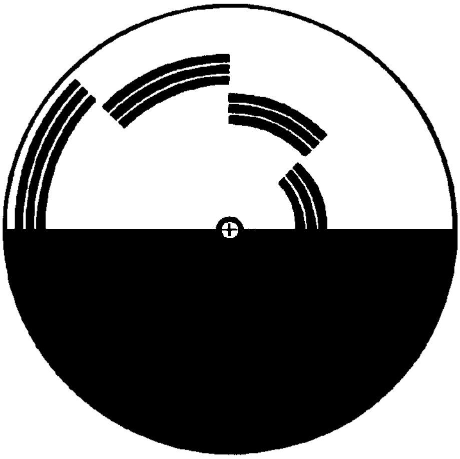

Benham's disk

Benham's disk

A disk, marked with a black and white pattern, which, when spun around, causes people to see colors. Benham's disk, also known as *Benham's wheel* and *Benham's top,* was invented in 1894 by the toy maker C. E. Benham and originally sold through Messrs. Newton and Co. as the Artificial Spectrum Top. It is one of a number of spinning disk color illusions first described by Gustav **Fechner** in 1838. For this reason, the illusory colors are known as *Fechner colors.* From the beginning, it was realized that the root of the illusion probably lay in the variation of retinal response time with wavelength. An online animated version of the disk can be seen at http://www.michaelbach.de/ot/col_benham/index.html.

Bernoulli family

An extraordinary Swiss family from Basle that produced eight outstanding mathematicians within three generations. Together with Isaac **Newton**, Gottfried **Leibniz**, Leonhard **Euler**, and Joseph **Lagrange**, the Bernoulli family dominated mathematics and physics in the seventeenth and eighteenth centuries, making important contributions to differential calculus, geometry, mechanics, ballistics, thermodynamics, hydrodynamics, optics, elasticity, magnetism, astronomy, and probability theory. Unfortunately, the Bernoullis were as conceited and arrogant as they were brilliant, and engaged in bitter rivalries and rows with one another.

The patriarchs of this mathematical dynasty were Jakob I (1654–1705) and his brother Johann I (1667–1748). (The Roman numerals are used to tell fathers, brothers, sons, and cousins apart, as the same Christian names were used repeatedly.) Next came Jakob's son, Nikolaus I, and Johann's three sons, Nikolaus II, Daniel (1700–1772), and Johann II. Finally, came Johann II's mathematical offspring, Johann III and Jakob II.

Jakob I developed a passion for science and mathematics after meeting Robert Boyle during a trip to England in 1676. He largely taught himself in these subjects and went on to lecture in experimental physics at the University of Basle. He also secretly introduced his younger brother to mathematics, against the wishes of his parents who wanted the younger brother to go into commerce. The cooperation between the two brothers soon degenerated, however, into vitriolic argument. Irked by Johann's bragging, Jakob publicly claimed that his younger brother had copied his own results. Later, having been appointed to the chair of mathematics at Basle, Jakob succeeded in blocking his brother's appointment to the same department, forcing Johann to take a teaching job at the University of Groningen instead. Johann proposed the so-called **brachistrochrone problem** and, along with Newton, Leibniz, l'Hospital, and Jakob, managed to solve it–but only after he first came up with a faulty proof and then tried to substitute one of Jakob's in its place! Eventually, Johann was offered a post at Basle as, of all things, the department head of Ancient Greek. But, en route to Basle, Johann learned that Jakob had died of tuberculosis. Upon his arrival he set about lobbying for the vacant position and, in less than two months, got his way. Jakob's most important work, his *Ars Conjectandi* (The art of conjecture), was published posthumously and formed the basis of **probability theory**.

Sadly, Johann I repeated his father's mistake and tried to force the most mathematically talented of his three sons, Daniel, into an unwanted career as a merchant. When the attempt failed, Johann let Daniel study medicine, in order to prevent his son from becoming a competitor. But all three sons followed their father's path, and Daniel, while studying medicine, was tutored in math by his older brother Nikolaus II. In 1720, Daniel went to Venice to work as a physician but won such a big reputation for his work in physics and mathematics that Peter the Great of Russia offered him a chair at the Academy of Science in St. Petersburg. Daniel went, along with Nikolaus II, who was also offered a position at the Academy. However, after just eight months, Nikolaus came down with a fever and died. Upset, Daniel wanted to return to Basle, but Johann I didn't want his son–a potential rival–back home. Instead he sent one of his pupils, none other than the great Leonhard **Euler**, to St. Petersburg to keep Daniel company. The two Swiss mathematicians became good friends, and the six years they spent together in St. Petersburg were the most productive of Daniel's life.

When Daniel finally returned to Basle, quarrels within the family flared up again after he won the prize of the Parisian Academy of Science with a paper, coauthored with his father, on astronomy. Jealous of Daniel's success, Johann threw him out of the family house. And worse was to come. In 1738 Daniel published his magnum opus, *Hydrodynamica.* Johann I read the book, hastily wrote one of his own called *Hydraulica,* back-dated it to 1732, and claimed to be the inventor of fluid dynamics! The plagiarism was soon uncovered, and Johann was ridiculed by his colleagues, but his son never recovered from the blow. See also **St. Petersburg paradox**.

Bernoulli number

A number of the type defined by Jakob **Bernoulli** in connection with evaluating sums of the form $\sum i^k$. The sequence B_0, B_1, B_2, . . . can be generated using the formula

$$x/(e^x - 1) = \sum (B_n x^n)/n!$$

though various different notations are used for them. The first few Bernoulli numbers are: $B_0 = 1$, $B_1 = -1/2$, $B_2 = 1/6$, $B_4 = -1/30$, $B_6 = 1/42$, ... They crop up in many diverse areas of mathematics including the series expansions of $\tan(x)$ and **Fermat's last theorem**.

Berry's paradox

A **paradox**, devised by G. G. Berry of the Bodleian Library at Oxford University in 1906, that involves statements of the form: "The smallest number not nameable in under ten words." At first sight, there doesn't seem anything particularly mysterious about this sentence. After all, there are only so many sentences that have less than ten words, and only a set **S** of these specify unique numbers; so there is clearly some number N that is the smallest integer not in **S**. The trouble is, the Berry sentence itself is a specification for that number *in only nine words!* Berry's paradox shows that the concept of nameability is inherently ambiguous and a dangerous concept to be used without qualification. A similar air of the paradoxical swirls around the notion of **interesting numbers.**[60]

Bertrand's box paradox

A problem, similar to the **Monty Hall problem**, that was published by the French mathematician Joseph Bertrand (1822–1900) in his 1889 text *Calcul des Probabitités.* Suppose there are three desks, each with two drawers. One desk contains a gold medal in each drawer, one contains a silver medal in each drawer, and one contains one of each, but you don't know which desk is which. If you open a drawer and find a gold medal, what are the chances that the other drawer in that desk also contains gold? This comes down, then, to figuring out the probability that you've picked the gold-gold desk instead of the gold-silver desk. Many people quickly jump to the conclusion that there are two possibilities, and since the selection was random, it must be 50-50. But this is wrong. Think of the initial selection as picking from among six drawers:

Before			After		
S	S	G		G	G
S	G	G			G
1	2	3	1	2	3

So, we have it narrowed down to three drawers, with an equal probability of each one being the one that was picked. One of the drawers is in desk 2, so there's a one-third chance that desk 2 was picked. Two of the drawers are in desk 3, so there are two one-third chances (i.e., a two-third chance) that desk 3 was picked.

Bertrand's postulate

Also known as *Betrand's conjecture,* if n is an integer greater than 3, then there is at least one **prime number** between n and $2n - 2$. This postulate (which should now be called a theorem) is named after the French mathematician Joseph Bertrand (1822–1900) who, in 1845, showed it was true for values of n up to 3 million. The Russian Pafnuty Chebyshev (1821–1894) gave the first complete proof in 1850, so that it is sometimes called *Chebyshev's theorem* (although another theorem also goes by this name). In 1932, Paul **Erdös** gave a more elegant proof, using the **binomial coefficient**s, which is the one that appears in most modern textbooks. Bertrand's postulate implies that the nth prime p_n is at most 2^n.

Bessel, Friedrich Wilhelm (1784–1846)

A German astronomer and mathematician who became director of the observatory at Königsberg (see **bridges of Königsberg**). Much of Bessel's work dealt with **perturbation**s (wobbles) in the motion of planets and stars caused by the gravitational influence of other bodies. To help analyze these perturbations he developed certain mathematical **function**s that are now known as *Bessel functions* and are used widely in physics.

beta function

The function

$$B(m, n) = \int_0^1 x^{m-1} (1 - x)^{n-1} \, dx.$$

It can be defined in terms of the **gamma function** by

$$B(m, n) = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}.$$

Many integrals can be reduced to the evaluation of beta functions.

Betti number

An important topological property of a surface, named after the Italian mathematician Enrico Betti (1823–1892). The Betti number is the maximum number of cuts that can be made without dividing the surface into two separate pieces. If the surface has edges, each cut must be a "crosscut," one that goes from a point on an edge to another point on an edge. If the surface is closed, like a sphere, so that it has no edges, each cut must be a "loop cut," a cut in the form of a simple closed curve. The Betti number of a square is 0 because it is impossible to crosscut without leaving two pieces. However, if the square is folded into a tube, its topology changes–it now has two disconnected edges–and its Betti number changes to 1. A **torus**, or donut shape, has a Betti number of 2. See also **chromatic number**.

bicorn

Also known as the *cocked-hat,* any of a collection of **quartic** curves studied by James **Sylvester** in 1864 and by Arthur **Cayley** in 1867. The bicorn has the Cartesian equation

$$y^2(a^2 - x^2) = (x^2 + 2ay - a)^2.$$

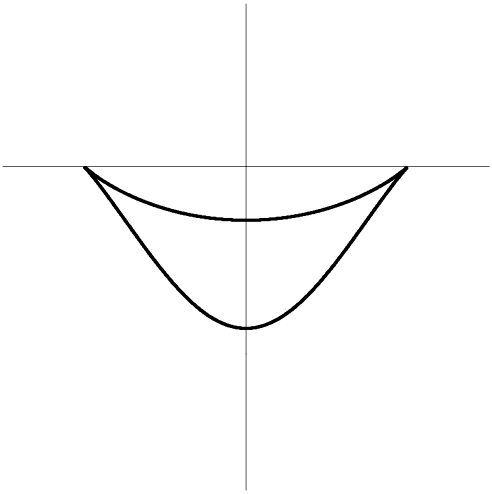

bicorn The bicorn curve *© Jan Wassenaar, www.2dcurves.com*

bicuspid curve

The **quartic** curve given by the equation

$$(x^2 - a^2)(x - a)^2 + (y^2 - a^2)^2 = 0.$$

Bieberbach conjecture

A celebrated conjecture made by the German mathematician Ludwig Bieberbach (1886–1982) in 1916, that was finally proved, after many partial results by others, by Louis de Branges of Purdue University in 1984.[54] Bierberbach is infamous in the history of mathematics because of his outspoken anti-Semitism during the Nazi era. Following the dismissal of Edmund Landau (1877–1938) from the University of Göttingen, Bierberbach wrote: "This should be seen as a prime example of the fact that representatives of overly different races do not mix as students and teachers. . . . The instincts of the Göttingen students felt that Landau was a type who handled things in an un-German manner."

Bieberbach's conjecture (BC) stemmed from the Riemann mapping theorem (RMT), which makes a claim about any region of a plane that is simply connected (in other words, any region, however complicated, that doesn't have any holes). The RMT says there must be some function, or mapping, such that every point in the arbitrary region is associated with one and only one point inside a circle with unit radius. Complex functions are best suited to plane-to-plane mappings and are often easier to work with if they can be represented as a power series. For example, given the complex number z, the function e^z can be expressed as the infinite series $1 + z + z^2/2! + z^3/3! + \ldots$. Bieberbach guessed that there is a link between the conditions imposed on a function by RMT and the numerical coefficients of the terms in a power series that represents the function. The BC says that if a function gives a one-to-one association between points in the unit circle and points in a simply connected region of the plane, the coefficients of the power series that represents the function are never larger than the corresponding power. In other words, given that $f(z) = a_0 + a_1z + a_2z^2 + a_3z^3 + \ldots$, then $|a_n| \leq n|a_1|$ for each n.

bifurcation

The value of a smoothly varying control parameter, or the point in *parameter space,* at which the behavior of a **dynamical system** undergoes a qualitative change. For example, a simple equilibrium, or a **fixed-point attractor**, might give way to a periodic oscillation as the stress on a system increases. Similarly, a **periodic attractor** might become unstable and be replaced by a **chaotic attractor**. To give a real-world example, drops fall individually at equal intervals, from a dripping faucet at low pressure. As the pressure is increased, however, the pattern of dripping abruptly changes so that two drops fall close together, with a longer interval before the next pair fall. In this case, a simple periodic process has given way to a periodic process with twice the period, a process described as "period doubling." If the flow rate of water through the faucet is increased still further, beyond the bifurcation point, often an irregular dripping is found and the behavior can become chaotic. See also **chaos**.

bilateral

Having two sides, or relating to the right and left sides of an object. *Bilateral symmetry* is a symmetrical arrangement, as of an organism or a body part, along a central axis, so that the body is divided into equivalent right and left halves by only one plane. See also **mirror reversal problem**.

bilateral diagram

A device invented by Lewis **Carroll** to represent the different logical states that two objects with two properties can take. Each cell in a four-square array represents one of the four possible object/property states and is covered by a red counter if the state is present and by a gray counter if it is absent.

billion

See **large numbers**.

bimagic square

A **magic square** that becomes a new magic square when each integer is squared. If, in addition to being bimagic, the integers in the square can be cubed and the resulting square is still magic, the square is said to be *trimagic.* To date the smallest bimagic square seems to be of order 8, while the smallest trimagic square is of order 32.

binary

> *There are 10 kinds of people in the world, those who understand binary math, and those who don't.*
>
> –Anonymous

The simplest positional **number system** and the natural one to use when dealing with computers; it employs just two symbols, 0 and 1, which correspond to the possible states of an off-on switch. Each place to the left in a binary number represents the next highest power of two. The binary number 10110_2, for example, means $1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 0 \times 2^0$, or 22_{10} in the familiar decimal notation. Nonintegers can be represented by using negative powers, which are set off from the other digits by means of a *radix point* (called a *decimal point* in base 10). The binary number 11.01_2 thus means $1 \times 2^1 + 1 \times 2^0 + 0 \times 2^{-1} + 1 \times 2^{-2}$ which equals 3.25_{10}. A number that terminates in a decimal doesn't necessarily do so in binary (e.g., $0.3_{10} = 0.0100110011001\ldots_2$), and vice versa. An irrational number, however, is nonperiodic in both systems (e.g., **pi**, $= 3.1415926\ldots_{10} = 11.001001000011111\ldots_2$). Binary arithmetic was first investigated by Gottfried **Leibniz** in 1672, though he didn't publish anything about it until 1701.

binary operation

An operation that involves two operands. For example, addition and subtraction are binary operations.

binomial

An expression containing two terms, joined by + or –. The **binomial theorem** gives the result of raising a binomial expression to a power; the expansion and the series it leads to are called the *binomial expansion* and the *binomial series.* A *binomial distribution* is described by a formula related to the binomial expansion. A binomial equation is a particular kind of equation that contains two terms.

binomial coefficient

A **coefficient** of x in the expansion of $(x + y)^n$. The binomial coefficient $_nC_m$ or $\binom{m}{n}$ gives the number of ways of picking m unordered outcomes from n possibilities, and is also known as a **combination**. It has the value $n!/(n - m)!m!$ The binomial coefficients form the rows of **Pascal's triangle**.

binomial theorem

The result that allows the expansion of a **binomial** expression:

$$(x+y)^n = x^n + a_{n-1}x^{n-1}y + a_{n-2}x^{n-2}y^2 + \ldots + y^n$$

where the coefficients a_i are called **binomial coefficients**.

Birkhoff, George David (1884–1944)

The foremost American mathematician of the early twentieth century and the first prominent dynamicist in the New World. He is known for his work on linear **differential equations** and **difference equations**, and was also deeply interested in and made contributions to the analysis of **dynamical systems**, celestial mechanics, the **four-color map problem**, and function spaces. Although a geometer at heart, he discovered new symbolic solution methods. He saw beyond the theory of oscillations, created a rigorous theory of **ergodic** behavior, and foresaw dynamical models for **chaos**. In addition he wrote on the foundations of relativity and quantum mechanics and, in *Aesthetic Measure* (1933), on art and music.

birthday paradox

The fact–not really a **paradox**–that you need a group of only 23 people for there to be a better than 50/50 chance that two of these people will have the same birthday. This seems surprising because we are used to comparing our particular birthday with others and only rarely finding a perfect match. The probability of any two individuals having the same birthday is just $^1/_{365}$. Even if you were to ask 20 people, the probability of finding someone with your birthday is still less than $^1/_{20}$. But the odds improve dramatically when a group of people *ask each other* about their birthdays because then there are many more opportunities for a matchup. One way to calculate the probability of a birthday match is to count the pairs of people involved. In a room of 23 people, there are $(23 \times 22)/2$, or 253, possible pairs. Each pair has a probability of success of $^1/_{365} = 0.00274$ (0.274%), and thus a probability of failure of $(1 - 0.00274) = 0.99726$ (99.726%). The probability of no match among any of the pairs of people is 0.99726 to the 253rd power, which is 0.499 (49.9%). So the probability of a successful match is $(1 - 0.499)$, or slightly better than even (50/50). With 42 people, the probability of a birthday match climbs to 90%.

birthday surprise

Here is a way to learn someone's birthday by doing a little simple math. Ask a person to take the month number (January = 1, February = 2, and so forth) of their birthday, multiply by 5, add 6, multiply the total by 4, add 9, and multiply the new total by 5 again. Finally, have her add the number of the day on which she was born and give you the total. In your head, subtract 165 and you will have the month and day of her birth. How does this work? If M is the month number and D the day number, then after the seven steps the expression for their calculation is: $5(4(5M + 6) + 9) + D = 100M + D + 165$. Thus, if you subtract the 165, what will remain will be the month in hundreds plus the day.

bisect

To cut in half.

bisecting an angle

Splitting an angle exactly in two. The ancient Greeks knew how to easily do it using only a pair of compasses and a straightedge. Here's how: Put the point of the compass at a point O and draw a circle so that it cuts the two lines coming out from the angle. Call these intersection points A and B. Now put the point of the compass at A and draw an arc that follows within the opening of the angle. Without changing the radius at which the compass is set, move its point to B and draw another arc. Join the point where the two arcs cross, P, to O using the straight edge: angle POB is half of angle AOB. See also **trisecting an angle**.

bishops problem

To find the maximum number of bishops (chess pieces capable of moving any number of spaces along diagonals of their own color) that can be placed on an $n \times n$ chessboard in such a way that no two are attacking each other. The answer is $2n - 2$, which gives the solution 14 for a standard (8×8) chessboard. The numbers of distinct maximal arrangements for $n = 1, 2, \ldots$ bishops are 1, 4, 26, 260, 3,368, ...

bistromathics

The revolutionary new (and totally fictitious) field of mathematics in restaurants, as described by Douglas Adams in his book *Life, the Universe and Everything:*[4]

> Numbers written on restaurant bills within the confines of restaurants do not follow the same mathematical laws as numbers written on any other pieces of paper in any other parts of the Universe. This single statement took the scientific world by storm. . . . So many mathematical conferences got held in such good restaurants that many of the finest minds of a generation died of obesity and heart failure and the science of math was put back by years.

Adams explains that just as Einstein found that space and time are not an absolute but depend on the observer's movement, so numbers are not absolute, but depend on the observer's movement in restaurants:

> The first non-absolute number is the number of people for whom the table is reserved. This will vary during the course of the first three telephone calls to the restaurant, and then bear no apparent relation to the number of people who actually turn up, or to the number of people who subsequently join them after the show/match/party/gig, or to the number of people who leave when they see who else has turned up. The second non-absolute number is the given time of arrival, which is now known to be one of the most bizarre of mathematical concepts, a recipriversexcluson, a number whose existence can only be defined as being anything other than itself. In other words, the given time of arrival is the one moment of time at which it is impossible that any member of the party will arrive. . . . The third and most mysterious piece of nonabsoluteness of all lies in the relationship between the number of items on the bill, the cost of each item, the number of people at the table and what they are each prepared to pay for.

See also **large number**s.

bit

A **binary** digit, either 0 or 1. See also **byte**.

blackjack

Also known as *twenty-one,* the most popular casino game in the world and the only such game with a *fluctuating* probability: the odds of winning change with the makeup of the deck. The cards two to nine have a numerical value equal to the number printed on the card. Tens and all face cards (jack [J], queen [Q], and king [K]) have the value of 10. Aces may be counted as either 11 or one. A dealer plays against one to seven players. Every player and the dealer initially receive two cards each, dealt by the dealer. Each player's hand is played against the dealer's hand only. If a player's hand has a value closer to 21 (without going over) than the dealer's hand, the player wins. The best possible hand is known as a *blackjack* (21 in the first two cards) and consists of an ace and a ten-valued card (10, J, Q, K). The payout for a blackjack is 3-to-2: the player is paid three chips for every two chips bet. When both the player and the dealer have blackjacks, it is a normal *tie* (*push*) situation and the player retains the initial bet. The player has several choices after receiving the first two cards: (1) *Hit* or *draw:* take one or more cards to add-up to a better hand. (2) *Stand:* stop taking more cards. (3) *Double down:* double the initial amount (in cases considered more favorable). (4) *Split pairs:* if the two cards are equal in value they may be played in two separate hands. The dealer must draw until his hand adds up to 17 or more. Both the player and the dealer can go over 21, a situation known as *bust;* the player loses the bet immediately. The dealer plays his hand last, after all the players at the table. This rule creates the so-called *house edge*. John Scarne[282] was the first to calculate the house advantage at blackjack as 5.9%. However, the house edge can be cut to around 1% if the player follows certain rules. The set of rules known as *basic strategy* make blackjack one of the fairest games of any kind, almost as fair as coin tossing.

In 1962 Edward O. Thorp, an IBM computer scientist, published *Beat the Dealer,*[333] which introduced a winning method called *card counting.* This method considered the 10-valued cards and the aces as positive, and the cards 2 to 6 as negative. If the net value of the remaining deck was positive, the player must increase the bet accordingly. The method had visible results when only one deck was used and very few cards remained in the deck. Casinos responded by changing the rules dramatically. The *penetration* was introduced: not all the cards in the deck are played. Shuffling is done unexpectedly. Also, most casinos introduced *multiple-deck blackjack.*

Blanche's dissection

The simplest **dissection** of a square into rectangles of the same areas but different shapes. It is composed of seven pieces; the square is 210 units on a side, and each rectangle has area of $210^2/7 = 6,300$.

Bólyai, János (1802–1860)

A Hungarian mathematician who was one of the founders of **non-Euclidean geometry**, independently coming to almost the same conclusions as Nikolai **Lobachevsky**. He was initially taught by his father, Farkas, also a mathematician, then he studied at the Royal Engineering College in Vienna from 1818 to 1822. Between 1820 and 1823 he prepared his treatise on a complete system of non-Euclidean geometry, commenting, "Out of nothing I have created a strange new universe." It was published in 1832 as an appendix to an essay by his father. Carl **Gauss**, on reading the appendix, wrote to a friend saying, "I regard this young geometer Bólyai as a genius of the first order." It was not until 1848 that Bólyai learned that Lobachevsky had produced a similar piece of work in 1829. Although he never published more than the 24 pages of the appendix, Bólyai left more than 20,000 pages of manuscript of mathematical work when he died. He was an accomplished linguist, speaking nine foreign languages including Chinese and Tibetan.

book-stacking problem

How much of an overhang can be achieved by stacking books on a table before the books overbalance and fall off? Assume each book is one unit long. To balance one book on a table, the center of gravity of the book must be somewhere over the table; to achieve the maximum overhang, the center of gravity should be just over the table's edge. The maximum overhang with one book is

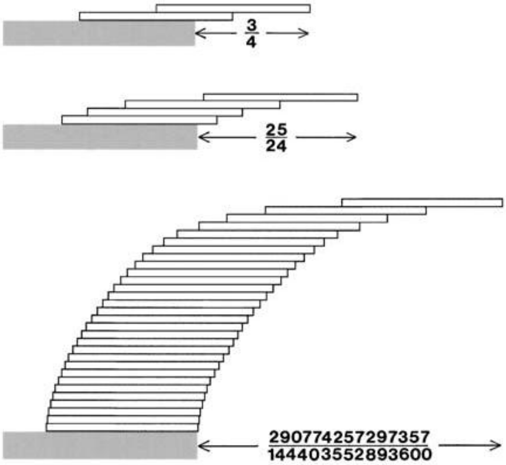

book-stacking problem The solution to the book-stacking problem.

obviously ½ unit. For two books, the center of gravity of the first should be directly over the edge of the second, and the center of gravity of the stack of two books should be directly above the edge of the table. The center of gravity of the stack of two books is at the midpoint of the books' overlap, or (1 + ½)/2, which is ¾ unit from the far end of the top book. It turns out that the overhangs are related to the harmonic numbers H_n, (see **harmonic sequence**), which are defined as $1 + \frac{1}{2} + \frac{1}{3} + \ldots + 1/n$: the maximum overhang possible for *n* books is $H_n/2$. With four books, the overhang $(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4})/2$ exceeds 1, so that no part of the top book is directly over the table. With 31 books, the overhang is 2.0136 book lengths.

Boole (Stott), Alicia (1860–1940)

The third daughter of George **Boole** and an important mathematician in her own right. At the age of 18, she was introduced to a set of wooden cubes devised by her brother-in-law Charles **Hinton** as an aid to visualization of the **fourth dimension**. Despite having had no formal education, she surprised everyone by becoming adept with the cubes and developing an amazing feel for four-dimensional geometry. She introduced the word ***polytope*** to describe a four-dimensional convex solid, and went on to explore the properties of the six regular polytopes and to make 12 beautiful card models of their three-dimensional central cross sections. She sent photographs of these models to the Dutch mathematician Pieter Schoute (1846–1923), who had done similar work and with whom she subsequently published two papers. The models themselves are now housed in the Department of Pure Mathematics and Mathematical Statistics at Cambridge University.

Boole, George (1815–1864)

An English mathematician and philosopher who is regarded as one of the founders of computer science. His great contribution was to approach logic in a new way, reducing it to a simple algebra and thus incorporating logic into mathematics. He pointed out the analogy between algebraic symbols and those that represent logical forms; his algebra of logic became known as **Boolean algebra** and is now used in designing computers and analyzing logic circuits. Although he never studied for a degree, Boole was appointed to the chair of mathematics at Queens College, Cork, Ireland, in 1849. One day in 1864 he walked the two miles in pouring rain from his home to the college and then lectured in wet clothes. A fever followed but whether this alone would have caused his demise is unknown. Certainly his condition wasn't helped by his wife, Mary (a niece of Sir George Everest, after whom the mountain is named), who, following the principle that remedy should resemble cause, put Boole to bed and threw buckets of water over him. He expired shortly after. See also **Boole (Stott), Alicia**.

Boolean

Taking only 0/1, true/false, yes/no values.

Boolean algebra

An **algebra** in which the **binary** operations are chosen to model the **union** and **intersection** operations in **set theory**. For any set *A*, the subsets of *A* form a Boolean algebra under the operations of union, intersection, and complement.

Borel, Emile (1871–1956)

A French mathematician who worked on divergent series, the theory of functions, probability, and **game theory**, and was the first to define games of strategy. He also founded **measure theory**, which applies the theory of sets to the theory of functions, and thus became an originator, with Henri **Lebesgue** and René Louis Baire (1874–1932), of the modern theory of functions of a real variable.

Borges, Jorge Luis (1899–1986)

An Argentinian author, essayist, and poet, many of whose short stories explore paradoxes and other strange avenues of mathematics, logic, philosophy, and **time**. For example, the possibility of branches in time is dealt with in "The Garden of Forking Paths," while the strange notion of the **Universal Library** is the subject of "The Library of Babel." Borges was profoundly influenced by European culture, English literature, and such thinkers as George Berkeley.

Borromean rings

Three rings linked in such a way that although they can't be separated, no two rings are linked; remove any one ring, however, and the other two fall apart. Named after the Italian family of Borromeo whose family crest has borne the rings since the fifteenth century, the design has been used in many places and times as a symbol of strength in unity. A form of the Borromean link known as Odin's triangle or the walknot ("knot of the slain") was used by the Norse folk of Scandinavia in two variants: a set of Borromean triangles and a unicursal curve that makes a trefoil knot. A motif of three interlaced crescent moons, similar to the Borromean rings, can be seen at the Palace of Fontainebleau. Designed by the architect Philibert de l'Orme, it is based on the moon emblem used by Diane de Poitiers (1499–1566), mistress of King Henry II of France. A similar pattern, but with three

Borromean rings A flowerpot on Isola Bella, an island on Lake Maggiore, near Arona in northern Italy, that bears the Borromeo family crest. *Peter Cromwell*

interlaced snakes in place of the crescent moons, occurs at various sites in Wales, including Bangor Cathedral. The Borromean rings are commonly used to symbolize the Christianity Trinity. An early source for this was a thirteenth-century French manuscript, now lost, in which the word *unitas* appears in the center, inside all the circles, and the three syllables of "tri-ni-tas" are distributed in the outer sectors. Borromean rings can also be found on Japanese family emblems, at a Japanese Shinto shrine north of Sakurai in the province of Nara, and in the sculptures of the Australian artist John Robinson. In North America, the design is known as the Ballantine rings after the New Jersey brewing company P. Ballantine and Sons who use it as a trademark with the rings labeled Purity, Body, and Flavor.

The rings first appeared in mathematics in the earliest work on knots by Peter **Tait** in 1876. The pattern of circles can be interlaced by replacing each of the six vertices by a crossing that shows how the circles pass over and under one another. Since there are two choices for each crossing, there are $2^6 = 64$ possible interlaced patterns. However, after taking symmetry into account, these 64 reduce to only 10 geometrically distinct patterns. Two patterns are considered to be the same if one can be obtained from the other by applying one or more of the following operations: rotation by 120°, reflection, and reflection in the plane of the pattern. The last symmetry operation means that the sense of all the crossings is switched. The rings can also be analyzed from the viewpoint of **topology**, which means that the designs are thought of as links made from a flexible and elastic material. If two links can be manipulated and deformed to look like one another (without breaking and joining the rings) then they are topologically equivalent. The 10 geometrically distinct patterns boil down to only five distinct topological types.

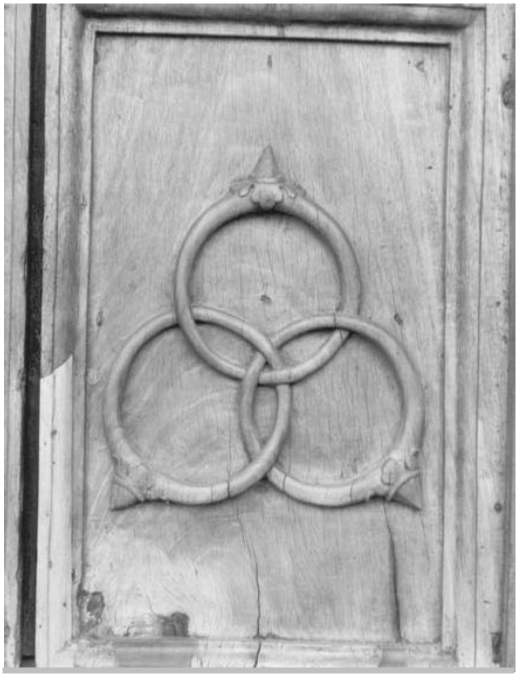

Borromean rings Three alternating rings carved in a panel in a walnut door of the church of San Sigismondo in Cremona, Italy. The emblem is one of several belonging to the Sforza family. *Peter Cromwell*

Borsuk-Ulam theorem

One of the most important and profound statements in **topology**: if there are *n* regions in *n*-dimensional space, then there is some hyperplane that cuts each region exactly in half, measured by volume. All kinds of interesting results follow from this. For example, at any given moment on Earth's surface, there must exist two antipodal points–points on exactly opposite sides of Earth–with the same temperature and barometric pressure! One way to see that this must be true is to think of two opposite points *A* and *B* on the equator. Suppose *A* starts out warmer than *B*. Now move *A* and *B* together around the equator until *A* moves into *B*'s original position, and simultaneously *B* into *A*'s original position. *A* is now cooler than *B,* so somewhere in between they must have been the same temperature. The Borsuk-Ulam theorem implies the **Brouwer fixed-point theorem** and also the **ham sandwich theorem**.

bottle sizes

Wine and champagne come in various standard bottle sizes, as shown in the table "Wine Bottle Sizes." These follow a **geometric sequence**, doubling in size with each step, up to the double-magnum, but thereafter increase in a more complicated way. There are also regional variations (for example, a Nebuchadnezzar may hold from 12 to 15 liters) and differences depending on the type of drink held.

Wine Bottle Sizes

Name of Size	Region	Capacity (liters)	Standard Bottle Equivalents
Baby/split	All	0.1875	0.25
Half-bottle	All	0.375	0.5
Bottle	All	0.75	1
Magnum	All	1.5	2
Double-magnum	All	3	4
Jeroboam	Burgundy, Champagne	5	6.67
Jeroboam	Bordeaux, Cabernet S.	4.5	6
Rehoboam	Burgundy, Champagne	4.5	6
Imperial	Bordeaux, Cabernet S.	6	8
Methuselah	Burgundy, Champagne	6	8
Salmanazar	Burgundy, Champagne	9	12
Balthazar	Burgundy, Champagne	12	16
Nebuchadnezzar	Burgundy, Champagne	15	20

boundary condition

The value of a **function** at the edge of the range of some of its variables. Recognizing the boundary conditions of an unknown function helps in its identification since other unknowns, such as variables in integrations, can then be eliminated.

boundary value problem

An **ordinary differential equation** or a **partial differential equation** given together with **boundary condition**s to ensure a unique solution.

Bourbaki, Nicholas

Not an individual but a collective mathematician. In the 1930s, the Bourbaki group, made up of some of the brightest mathematicians in France, began as a club, holding secret meetings in Strasbourg to update university lectures and texts in the wake of World War I, which had essentially wiped out a generation of young talent. In time, Bourbaki authored encyclopedic accounts of all areas of mathematics, and its influence became widespread.

Its origins can be traced to 1934 and to André Weil and Henri Cartan who were maîtres de conférences (the equivalent of assistant professors) at the University of Strasbourg. One of their duties was to teach differential and integral calculus but they found the standard text on this subject, *Traité d'Analyse* by E. Goursat, wanting. Following a suggestion by Weil to write a new "Treatise on Analysis," a group of about 10 mathematicians began to meet regularly to plan the new work. Quickly, it was decided that the work should be collective, without any acknowledgment of individual contributions, and this became a feature of Bourbaki's output. In the summer of 1935, the pen name Nicholas Bourbaki was chosen, and the initial membership consisted of Weil, Cartan, Claude Chevalley, Jean Delsarte, and Jean Dieudonné–all former students at the École Normale Supérieure in Paris. Over the years, the membership varied; some people in the first group dropped out quickly, others were added, and later there was a regular process of addition and retirement (mandatory by the age of 50).

Bourbaki adopted rules and procedures that to outsiders often seemed eccentric and even bizarre. For example, during meetings to review and revise drafts for the various books the group developed, anyone could express his opinion as loudly as he wanted at any time, so it was not uncommon for several distinguished mathematicians to be on their feet at the same time shouting monologues at the top of their voices. Somehow out of this mayhem emerged work of extreme precision, to the point of pedantry and dryness. Bourbaki would have nothing to do with geometry or any attempt at visualization, and believed that mathematics should distance itself from the sciences. However, despite its tendency to be boring and long-winded, Bourbaki did achieve its goal: to set down in writing what was no longer in doubt in modern mathematics.

brachistochrone problem

A problem with which Johann Bernoulli (see **Bernoulli Family**) challenged his contemporaries in *Acta Eruditorum* in June 1696:

> Following the example set by Pascal, Fermat, etc., I hope to gain the gratitude of the whole scientific community by placing before the finest mathematicians of our time a problem which will test their methods and the strength of their intellect. If someone communicates to me the solution of the proposed problem, I shall publicly declare him worthy of praise. . . . Given two points A and B in a vertical plane, what is the curve traced out by a point acted on only by gravity, which starts at A and reaches B in the shortest time?

Isaac **Newton** reportedly solved the problem between four in the evening and four in the morning after a hard day at the Royal Mint, later commenting: "I do not love to be dunned [pestered] and teased by foreigners about mathematical things. . . ." Other correct solutions came from Gottfried **Leibniz**, the Frenchman Guillaume **de L'Hôpital**, and Johann's brother Jakob. They, like Johann, realized that the solution to the brachistochrone problem, as it was also to the **tautochrone problem**, was a curve known as the **cycloid**.

Brahmagupta (A.D. 598–after 665)

A Hindu astronomer and mathematician who became the head of the observatory at Ujjain–the foremost mathematical center of ancient India. His main work, *Brahmasphutasiddhanta* (The opening of the universe), written in 628, contains some remarkably advanced ideas, including a good understanding of the mathematical role of **zero**, rules for manipulating both positive and **negative number**s, a method for computing square roots, methods of solving linear and some **quadratic** equations, and rules for summing series. His contributions to astronomy were equally ahead of their time. *Brahmagupta's theorem* states that in a cyclic quadrilateral (a four-sided shape whose corners lie on a circle) having perpendicular diagonals, the perpendicular to a side from the point of intersection of the diagonals always bisects the opposite side. **Brahmagupta's formula** for the area of a cyclic quadrilateral with sides of length a, b, c, and d is $\sqrt{(S-a)(S-b)(S-c)(S-d)}$, where $S = (a+b+c+d)/2$. As d goes to zero, this reduces to **Heron's formula**.

braid

A collection of lines or strings that are plaited together and whose ends are attached to two parallel straight lines.

Braid theory was pioneered by the Austrian mathematician Emil Artin (1898–1962) and is related to **knot** theory. It also has other applications: for instance, if we consider the way the roots of a **polynomial** move as one of the polynomial's coefficients changes, this motion can be thought of as a braid.

Brianchon's theorem

Given a **conic section**, if we circumscribe a hexagon about it, then the major diagonals of the hexagon are concurrent.

bridges of Königsberg

A famous routing problem that was analyzed and solved by Leonhard **Euler** in 1736, and that helped spur the development of **graph** theory. The old city of Königsberg, once the capital of East Prussia, is now called Kaliningrad. It falls within a tiny part of Russia known as the Western Russian Enclave, between Poland and Lithuania, which (to the surprise even of many modern Russians) is not connected with the rest of the country! Königsberg lay some four miles from the Baltic Sea on rising ground on both sides of the river Pregel (now the Pregolya), which flowed through the town in two branches before uniting below the Grune Brocke (Green Bridge). Seven bridges (numbered in the diagram) crossed the Pregel and connected various parts of the city (letters A to D), including Kneiphof Island (B), the site of Königsberg University and the grave of its most famous son, the great philosopher Emmanuel Kant (1724–1804).

A question arose among the town's curious citizens: Was it possible to make a journey across all seven bridges *without having to cross any bridge more than once?* No one had been able to do it, but was there a solution? Euler, who was in St. Petersburg, Russia, at the time, heard about this puzzle and looked into it. In 1736, he published a paper called *Solutio problematis ad geometriam situs pertinentis* (The solution of a problem relating to the geometry of position) in which he gave his answer. Euler reasoned that, for such a journey to be possible, each land mass would need to have an even number of bridges connected to it, or, if the journey began at one land mass and ended at another then those two land masses alone could have an odd number of connecting bridges while all the other land masses would have to have an even number of connecting bridges. Since the Königsberg bridges violated this layout, a grand tour that involved only one crossing per bridge was impossible. Euler's paper was important because it solved not just the Königsberg conundrum but the much more general case of any network of points, or *vertices,* that are connected by lines, or *arcs.* What is more, the words *geometry of position* in the title show that Euler realized that he was dealing with a different type of geometry where distance is irrelevant. So this work can be seen as a prelude to the subject of **topology**. See also **Euler path**.

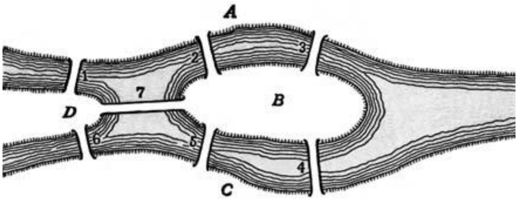

bridges of Königsberg The essential layout of the bridges. *B* represents Kneiphof Island.

KONINGSBERGA

bridges of Königsberg An old map of the city showing the seven bridges.

Briggs, Henry (1561–1630)

An English mathematician who introduced common **logarithms** (to base 10) and was largely responsible for getting scientists to use them. Although a well-regarded mathematician in his own right, holding the Savilian chair of geometry at Oxford, Briggs was most important as a contact and as a public relations man for his field.

Brocard problem

The problem of finding the integer solutions of the equation

$$n! + 1 = m^2.$$

These solutions are called *Brown numbers,* and only three of them are known: (5, 4), (11, 5), and (71, 7). Paul **Erdös** conjectured that there are no others.

broken chessboard

See **polyomino**.

Bronowski, Jacob (1908–1974)

A Polish mathematician who worked first on operation theory and its application to military strategy, but later on the ethics of science. He is remembered for writing and narrating the television series *The Ascent of Man* in 1973.

Brouwer, Luitzen Egbertus Jan (1881–1966)

A Dutch mathematician who opposed the logicist school of Bertrand **Russell** and established the intuitionist school of mathematical thought. He was also one of the founders of **topology**, doing most of his work in this field between 1909 and 1913.

Brouwer fixed-point theorem

An amazing result in **topology** and one of the most useful theorems in mathematics. Suppose there are two sheets of paper, one lying directly on top of the other. Take the top sheet, crumple it up, and put it back on top of the other sheet. Brouwer's theorem says that there must be at least one point on the top sheet that is in exactly the same position relative to the bottom sheet as it was originally. The same idea works in three dimensions. Take a cup of coffee and stir it as much as you like. Brouwer's theorem insists that there must be some point in the coffee that is in exactly the same spot as it was before you started stirring (though it might have moved around in between). Moreover, if you stir again to move that point out of its original position, you can't help but move another point back into its original position! Not surprisingly, the formal definition of Brouwer's theorem makes no mention of sheets of paper or cups of coffee. It states that a continuous function from an n-**ball** into an n-ball (that is, any way of mapping points in one object that is topologically the same as the filling of an n-dimensional sphere to another such object) must have a fixed point. Continuity of the function is essential: for example, if you rip the paper in the previous example then there may not be a fixed point.

Brownian motion

The most common type of continuous **random** motion of a particle, one in which the particle's vibrations have more energy at short length and time scales. It models the motion of a particle in a fluid, fluctuation of stock prices, and many other processes. Brownian motion is named after the Scottish botanist Robert Brown (1773–1858) who first studied it.

Brun's constant

See **twin primes**.

bubbles

Whether alone or in groups joined together, bubbles get their shape by following one simple rule: soap film always tries to form a **minimal surface**. The mathematical study of bubbles and films began in earnest in the 1830s with the experiments of Joseph **Plateau**. A single bubble will always try to form a sphere because this shape, as proposed by **Archimedes** and proved by Hermann Amandus Schwarz (1843–1921) in 1884, is the minimal surface enclosing a single volume. For a long time, mathematicians believed that the minimal surface for enclosing two separate volumes of air is a double bubble separated by a third surface, which meets the other two along a circle at 120° and is flat if the bubbles enclose the same volume and, otherwise, is a spherical surface that bulges a little into the larger of the two. The *double bubble conjecture* was finally confirmed by a team of four mathematicians in 2000.

Another great bubble mystery that has recently been solved is why the bubbles in a glass of Guinness appear to sink rather than rise. Bubbles that rise, like those in a saucepan or those breathed out by a diver, are a familiar sight and easy to explain: gas bubbles are lighter than liquid and experience a buoyancy force that drives them up toward the liquid surface. But many of the bubbles in a glass of Guinness can be seen heading downward. Researchers have found that large bubbles in the center of the glass move upward relatively quickly and drag liquid with them. Since the amount of liquid in the glass stays the same (unless someone drinks it!) the liquid moving upward near the center must eventually move back down near the walls of the glass. This downward-

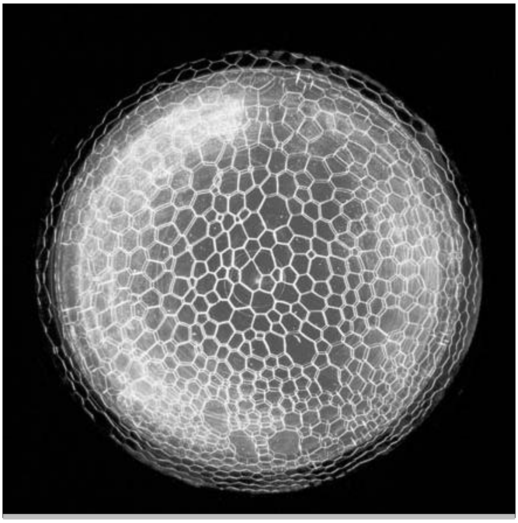

bubbles A honeycomb-like arrangement of tightly packed bubbles.

Australian National University

moving liquid has a dragging effect on the bubbles. Larger bubbles are more buoyant than smaller bubbles, and continue to move upward. Smaller bubbles (less than 0.05 mm in diameter) aren't buoyant enough to resist this drag force, and move downward with the liquid near the sides of the glass. Since Guinness is quite opaque, and these downward-moving small bubbles are close to the side of the glass, it often looks as if almost all the bubbles are moving down. See also **Plateau problem**.

buckyball

Also known as a *fullerene*, a large molecule made of carbon atoms arranged in the form of a convex polyhedral cage. Buckyballs are named after the architect Richard Buckminster Fuller because they look like the geodesic domes that he invented. The first buckyball to be discovered (by accident) was C_{60}, in which 60 carbon atoms are arranged at each of the vertices of a truncated **icosahedron**. This shape, which looks like a soccer ball, has 32 faces, of which 20 are regular hexagons and 12 are regular pentagons. Many

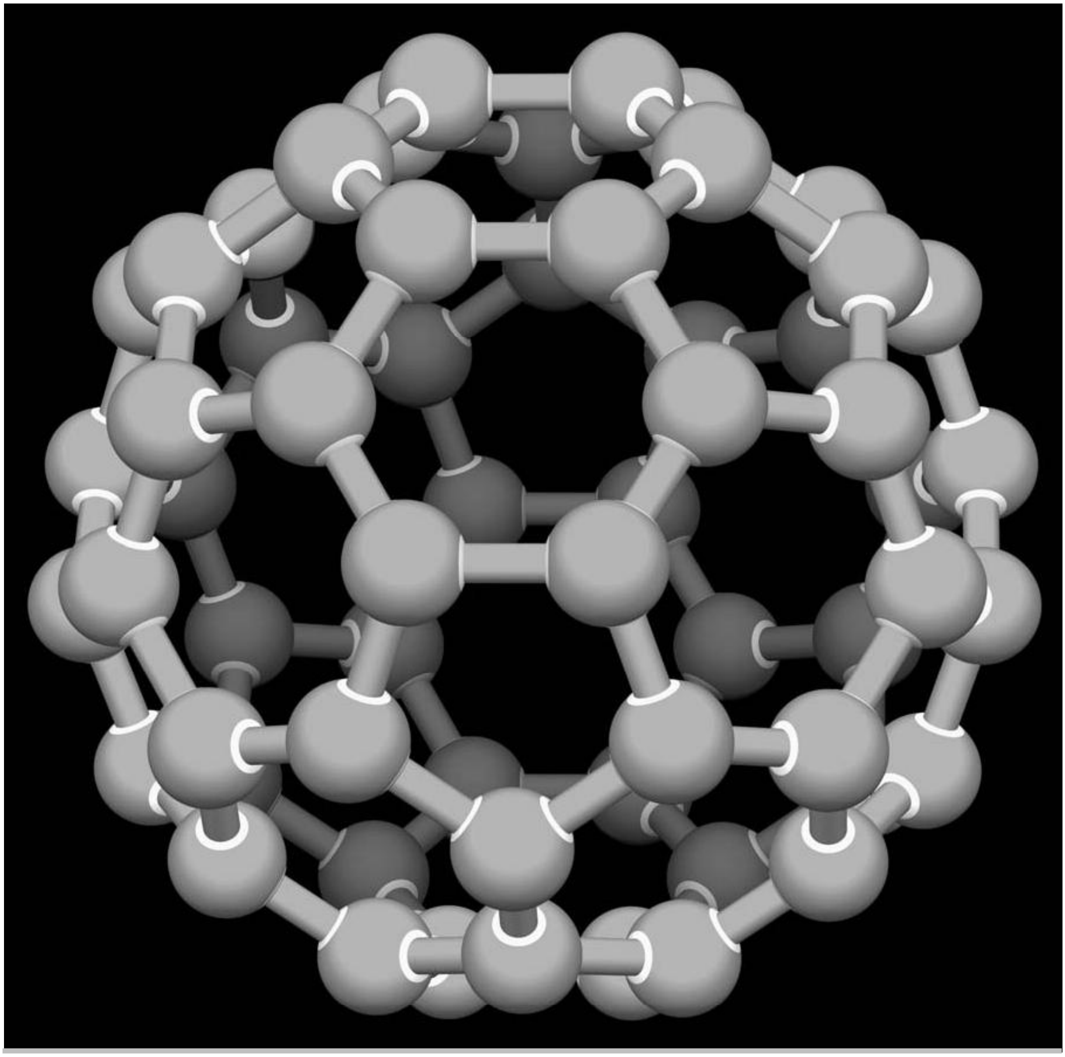

buckyball A molecule of Buckminster fullerene—a buckyball. *Nick Wilson* different types of buckyballs are known. Common ones have 70, 76, and 84 carbon atoms, but all are built up exclusively from hexagonal and pentagonal faces in arrangements that follow **Euler's formula**. This formula ensures that while the number of hexagons can vary from one type of fullerene to another, every fullerene has exactly 12 pentagons. (In fact, buckyballs with heptagonal faces have been seen, but these faces are concave and are regarded as defects.)

Buffon's needle

An early problem in geometrical probability (see **probability theory**) that was investigated experimentally in 1777 by the French naturalist and mathematician Comte Georges Louis de Buffon (1707–1788). It involves dropping a needle repeatedly onto a lined sheet of paper and finding the probability of the needle crossing one of the lines on the page. The result, surprisingly, is directly related to the value of **pi**.

Consider a simple case in which the lines are 1 cm apart and the needle is 1 cm in length. After many drops the probability of the needle lying across a line is found to be very close to $2/\pi$. Why? There are two variables: the distance from the center of the needle to the closest line, d, which can vary between 0 and 0.5 cm, and the angle, θ, at which the needle falls with respect to the lines, which can vary between 0 and 180°. The needle will hit a line if $d \leq \frac{1}{2} \sin\theta$. In a plot of d against $\frac{1}{2} \sin\theta$, the values on or below the curve represent a hit; thus, the probability of a success is the ratio of the area below the curve to the area of entire rectangle. The area below the curve is given by the integral of $\frac{1}{2} \sin\theta$ from 0 to π, which is 1. The area of the rectangle is $\pi/2$. So, the probability of a hit is $1/(\pi/2)$ or $2/\pi$ (about 0.637). Dropping a needle many times onto lined paper gives an interesting (but slow) way to find π. This kind of probabilistic means of performing calculations is the basis of a technique known as the **Monte Carlo method**.

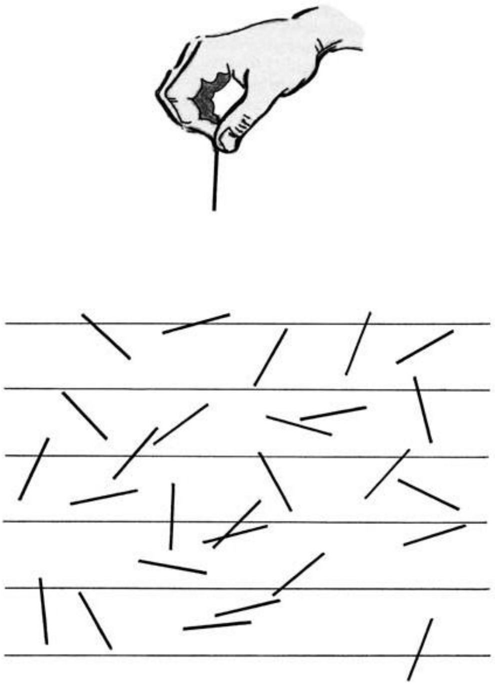

Buffon's needle Needles are dropped randomly onto a lined surface.

bundle

A map between two **topological space**s A and B, where the sets f^{1} (b) for elements b of B (known as fibers), are all **homeomorphic** to a single space. The simplest example is the **Möbius band**, for which A is the Möbius band, B is a circle, and the fibers are homeomorphic to an interval on the real number line.

Burali-Forti paradox

An argument that shows that the collection of **ordinal number**s (numbers that give the position of objects) do not, unlike the **natural number**s, form a **set**. Each ordinal number can be defined as the set of all its predecessors. Thus:

0 is defined as {}, the **empty set**
1 is defined as {0} which can be written as {{}}
2 is defined as {0, 1} which can be written as {{}, {{}}}
3 is defined as {0, 1, 2} which can be written as {{}, {{}}, {{}, {{}}}} . . .
in general, n is defined as $\{0, 1, 2, \ldots n - 1\}$

If the ordinal numbers formed a set, this set would then be an ordinal number greater than any number in the set. This contradicts the assertion that the set contains all ordinal numbers. Although the ordinal numbers don't form a set, they can be regarded as a collection called a class.

Buridan's ass

A paradox of medieval logic concerning the behavior of an ass that is placed equidistantly from two piles of food of equal size and quality. Assuming that the behavior of the ass is entirely rational, it has no reason to prefer one pile to the other. Thus lacking a basis to decide which pile to eat first, it remains in its original position and starves. With one pile it would have lived; having two identical piles it dies. How can this make sense? The paradox is named after the French philosopher Jean Buridan (c. 1295–1356).

burr puzzle

An interlocking wooden puzzle that, when put together, typically looks like three rectangular blocks crossing one another at right-angles. Little is known about its early history, but it was certainly produced both in Asia and Europe in the eighteenth century. It acquired the name *Chinese puzzle,* probably because so many were produced in the Orient from the early 1900s. In 1928, Edwin Wyatt published *Puzzles in Wood,*[355] the first book devoted to the subject, and introduced the term *burr puzzle* because of the likeness of the assembled toy to a seed burr.

Burr puzzles consist of three (the smallest number), six (the most common number), twelve, or other numbers of pieces that are notched in various ways so as to pose a challenge to the would-be assembler. The earliest known reference to the popular six-piece burr appears in a Berlin catalog of 1790, but not until 1917 was a patent taken out on a particular design. In 1977, William **Cutler** proved that 25 possible notchable pieces can be used to make solid six-piece burrs and that they can be put together in 314 ways. (Pieces are considered notchable if they can be made by a sequence of notches that are produced by chiseling out the space between two saw cuts.) Cutler also proved there are 369 general pieces from which solid burrs can be made and that these can be assembled in 119,979 ways. One particular form of burr has six identical pieces, all of which move outward or inward together. Another form, with flat notched pieces, has one piece with an extra notch or an extended notch that allows it to fit in last, either by sliding or twisting, although this isn't initially obvious. This form is sometimes made with equal pieces so that it can only be assembled by force, perhaps after steaming.

butterfly effect

One of the more sensational and loudly touted claims of **chaos** theory: a butterfly beating its wings could, by an intricate chain of causes and effects, give rise to a hurricane. The gist of the argument is that minuscule disturbances can be amplified unpredictably into major phenomena. However, the overwhelming likelihood is that any effect as small as the beating of a butterfly's wing would be quickly dampened out and play no significant part in future events. See also **causality**.

butterfly theorem

Let M be the midpoint of a **chord** PQ of a circle, through which two other chords AB and CD are drawn. If AD intersects PQ at X and CB intersects PQ at Y, then M is also the midpoint of XY. The theorem gets its name from the shape of the resulting figure.

Byron, (Augusta) Ada (1815–1852)

Also known as Lady Lovelace, the daughter of Lord Byron and one of the most picturesque characters in computer prehistory. Her parents separated five weeks after her birth and she was raised by her mother (neé Annabella Milbanke), whom Lord Byron had called his "Princess of Parallelograms" because of her interest in mathematics. She was determined that Ada would become a mathematician and scientist, not a poet like her father. Ada, however, managed to combine both worlds, blending her science with poetical vision and her mathematics with metaphor.

At the age of 17 she was introduced to Mary Somerville, a remarkable woman who translated **Laplace**'s works into English, and whose texts were used at Cambridge (where a

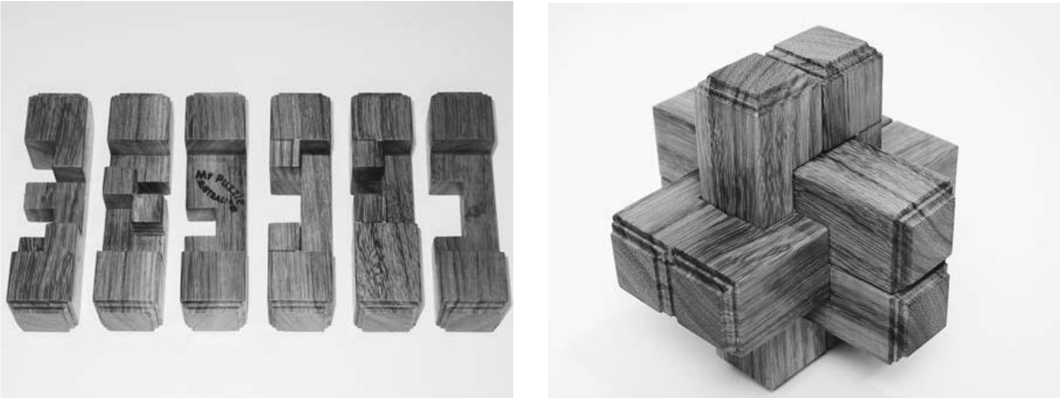

burr puzzle A difficult six-piece burr with 19 false assemblies and a final piece that requires 10 directional moves. *Mr. Puzzle Australia/William Cutler* women's college is now named after her). It was at one of Mary Somerville's dinner parties, in November 1834, that Ada first heard of Charles **Babbage**'s ideas for a new calculating machine, the Analytical Engine, and was immediately intrigued. In 1843, Ada, married and the mother of three children, translated a French article about the Engine and showed it to Babbage. He suggested that she add her own notes, which turned out to be three times the length of the original piece and included prescient comments about how such a machine might be used to compose complex music, produce graphics, and solve scientific problems. A regular correspondence ensued between Ada and Babbage, during which Ada suggested to Babbage a plan for how the engine might calculate **Bernoulli numbers**–a plan now regarded as the first computer program. In recognition of this, a software language developed by the U.S. Department of Defense was named "Ada" in 1979. Like her father, she died at the age of 36, following a series of illnesses.[335]

byte

A string of 8 **bit**s, used to represent a character.

C

caduceus

In mathematics, a pair of curves in space, each of which is a **helix** and which twist in opposite directions around one another. In mythology, the caduceus is the wing-topped staff, wound about by two snakes, carried by Hermes, the Greek messenger of the gods. The snakes became entwined after Hermes threw his staff at them to stop their fighting. A caduceus was carried by Greek officials and became a Roman symbol for truce and neutrality. Since the sixteenth century it has also served as a symbol of medicine. Before modern medicine, people infected by parasitic worms were treated by physicians using a stick and a knife. A slit would be cut in the patient's skin in front of the worm, and as the parasite crawled out of the incision, the worm would be wound around a stick until it was totally removed. The medical treatment of parasitic worm infection by knife and stick is believed to be the inspiration for the original caduceus. It was used as a promotional sign for physicians of that period.

Cage, John (1912–1992)

The American avant-garde composer perhaps best known for the quietest piece of music ever written. His piano composition *4′33″* calls for the player to sit in silence for 273 seconds–this being the number of degrees below zero on the centigrade scale of **absolute zero** at which molecular motion stops. *4′33″* was inspired by Cage's visit to Harvard University's anechoic chamber about which he wrote:

> There is no such thing as empty space or empty time. There is always something to hear or something to see. In fact, try as we might to make a silence, we cannot. For certain engineering purposes, it is desirable to have as silent a situation as possible. Such a room is called an anechoic chamber, its walls made of special materials, a room without echoes. I entered one at Harvard University . . . and heard two sounds, one a high and one a low. When I described them to the engineer in charge, he informed me that the high one was my nervous system and the low one was my blood circulation.

Cage's *4′33″* breaks traditional boundaries by shifting attention from the stage to the audience and even beyond the concert hall. The listener becomes aware of all sorts of sound, from the mundane to the profound, from the expected to the surprising, from the intimate to the cosmic: shifting in seats, riffling programs, breathing, a creaking door, passing traffic, a recaptured memory. Is sitting quietly alone for 273 seconds equivalent to a private performance (and audience) of the piece? Or, in the final analysis, is it all pretentious nonsense? In his essay on "Nothing" Martin Gardner wrote: "I have not heard *4′33″* performed, but friends who have tell me it is Cage's finest composition."

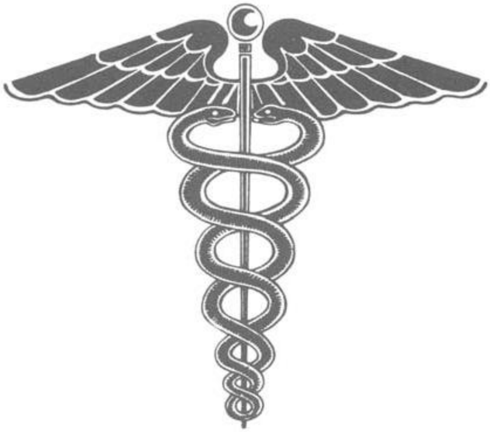

caduceus *Auckland Medical Research Foundation*

Caesar cipher

The simplest and oldest known type of **substitution cipher**, attributed to Julius Caesar, who used it to send government messages. In it, each letter in the alphabet is replaced by another letter using a predefined rule that shifts the alphabet a uniform amount to the right or left. For example, a shift of three units to the right, would turn the "This is secret" into "Wklv lv vhfuhw."

cake-cutting

How can a group of people cut up a cake so that each gets what they consider to be a fair share? In its modern mathematical form, this classic problem of *fair division* dates from World War II, when Hugo **Steinhaus** tackled it using a **game theory** approach.[315] Any number of "players" are allowed. They agree on rules for dividing the cake, and then everyone follows those rules. In the end, each player must get what he or she perceives to be a fair share. In the simplest case, involving just two people, there is an easy, well-known strategy: one cuts, the other chooses. Can this method be extended to three people? Can it be extended so that each person, according to his own judgment, receives the *biggest* piece? Steinhaus was able to prove that a so-called *envy-free division,* where everyone believes they got the best deal exists in every case, for any number of players. However, it was left for others to find actual algorithms that worked for three or more players.

A three-person envy-free method was first devised by John Selfridge and John **Conway**. Suppose the players are called Alice, Bob, and Carol. The method goes like this: (1) Alice cuts the cake into what she thinks are thirds. (2) Bob trims one piece to create a two-way tie for largest, and sets the trimmings aside. (3) Carol picks a piece, then Bob, then Alice. Bob has to take a trimmed piece if Carol does not. Call the person who took the trimmed piece T, and the other (of Bob and Carol) NT. (4) To deal with the trimmings, NT cuts them into what she thinks are thirds. (5) The players pick pieces in this order: T, Alice, NT. The key to the success of the Selfridge-Conway strategy is that for the trimmings, Alice has an "irrevocable advantage" with respect to T, since Alice will never envy T even if T gets all the trimmings. Thus Alice can pick after T and allow the method to end in a finite number of steps.

For four or more cake-cutters, envy-free solutions are very complex and can take arbitrarily long to resolve. However, a general solution of the problem of fair, envy-free division was eventually found in 1992 by the Americans Steven Brams, a political scientist at New York University, and Alan Taylor, a mathematician at Union College in Schenectady, New York.[51, 52] With two players, the first player cuts the cake in half. With three players, the first player cuts the cake into thirds. With four players, Brams and Taylor showed, the first player, say Bob, cuts the cake into five equal-looking pieces. He passes them to Carol, who trims two at most to create a three-way tie for largest in her eyes. She sets the trimmings aside and gives the five pieces to Don, who trims one at most to create a two-way tie for largest in his eyes. Alice, the fourth player, now selects the piece she likes best. Choosing proceeds in the reverse order from cutting, with the proviso that anyone who trimmed one or more pieces must take one of them if any are still available when it's his or her turn to choose. The extra piece to begin with assures that no player gets second-best. If someone takes a piece she likes before it's her turn to choose, an equivalent piece or better always remains on the table. According to a formula Brams and Taylor developed, Bob must cut the cake into at least $2^{(n-2)+1}$ pieces at the start. This amounts to nine pieces for five players, 17 pieces for six, and so on. Bob has to cut all these extra pieces to make sure that, when he finally gets to choose at the end, there will be a piece left that hasn't been either trimmed or chosen by one of the many other players. With 22 players, Bob has to divide the cake into over a million pieces–small crumbs of comfort in the quest for a fairer world.[268]

Calabi-Yau space

A type of mathematical space that enters into **string theory**, where the geometry of the universe is held to consist of at least 10 dimensions–the four familiar dimensions of **space-time** and six compact dimensions of Calabi-Yau space. These extra dimensions are so tightly curled up that they aren't noticed. Although the main application of Calabi-Yau spaces is in theoretical physics, they are also interesting from a purely mathematical standpoint.

calculus

> *The calculus is the greatest aid we have to the application of physical truth in the broadest sense of the word.*
>
> –William Fogg Osgood (1864–1943)

The branch of mathematics that deals with (1) the rate of change of quantities (which can be interpreted as the slopes of curves), known as *differential calculus,* and (2) the length, area, and volume of objects, known as *integral calculus*. Calculus was one of the most important developments in mathematics and also in physics, much of which involves studying how quickly one quantity changes with respect to another. It is no coincidence that one of the founders of calculus was the brilliant English physicist Isaac **Newton**; another was Gottfried **Leibniz**. Although students nowadays learn differential calculus first, integral calculus has older roots.

calculus of variations

Calculus problems, especially **differentiation** and *maximization,* that involve **function**s on a set of functions of a real variable. An example is to find the shape of a cable suspended from both ends.

calendar curiosities

The earliest event in human history for which a definite date is known is a battle between the Lydians (allies of the Greek Spartans) and the Medes (ruled by the Persian king Cyrus) who had been locked in a war for five years. As the two sides faced each other for a crucial daytime confrontation, a solar eclipse occurred. This was taken as a sign of the gods' disapproval and the Lydians and Medes agreed to end the fighting then and there. The dates of solar eclipses can be figured out with great accuracy, and this one is known to have taken place on May 28, 586 B.C.

Much less certain is the birth date of Christ. It was not until A.D. 440 that Christmas was celebrated on December 25. This date was chosen because it coincided with the birth date of Mithras, the Persian sun-god, and was close to the pagan festival of Yule. In A.D. 534, Dionysius Exiguus (also known as Dennis the Little) created the system, still used today, of counting the years from the birth of Christ. Unfortunately, he slipped up in his calculations. No one knows exactly when Jesus was born, but it was probably around 6 B.C. and certainly before the death of Herod the Great in 4 B.C.

As for the future, there's no shortage of predictions about the end of the world. According to the Mayan "long count" linear calendar, it will happen on June 5, 2012. Other calendric curiosities: February 1865 is the only month in recorded history not to have a full moon, and months that begin on a Sunday will always have a Friday the 13th.

Caliban

A pseudonym of Hubert **Phillips**.

Caliban puzzle

A logic puzzle in which one is asked to infer one or more facts from a set of given facts.

cannonball problem

The mathematical analysis of stacks of cannonballs (or of spheres in general) that has its roots in a question posed by Sir Walter Raleigh, explorer, introducer of the potato and tobacco to Britain, and part-time pirate on the high seas. Raleigh asked his mathematical assistant, Thomas Harriot, how he could quickly figure out the number of cannonballs in a square pyramidal stack without having to count them individually. Harriot solved this problem without difficulty. If k is the number of cannonballs along the side of the bottom layer, the number of cannonballs in the pyramid n is equal to $\frac{1}{6}k(1+k)(1+2k)$. For example, if $k = 7$, $n = 420$. A more specific form of the cannonball problem asks what is the smallest number of balls that can first be laid out on the ground as an $n \times n$

cannonball problem Cannonballs stacked in Narbonne, France. *Australia National University* square, then piled into a square pyramid k balls high? In other words, what is the smallest square number that is also a square pyramidal number? This answer is the smallest solution to the **Diophantine equation**

$$\tfrac{1}{6}k(1+k)(1+2k)=n^2$$

and turns out to be $k = 24$, $n = 70$, corresponding to 4,900 cannonballs. The ultimate form of the cannonball problem is to ask if there are any other, larger solutions. In 1875 Edouard **Lucas** conjectured that there weren't, and in 1918 G. N. Watson proved that Lucas was right.[345]

Returning to Elizabethan times, Thomas Harriot's interest in spheres extended far beyond piles of cannonballs. Harriot was an atomist, in the classical Greek sense, and believed that understanding how spheres pack together was crucial to understanding how the basic constituents of nature are arranged. Harriot also carried out numerous experiments in optics and was far ahead of his time in this field. So when, in 1909, Johannes **Kepler** wanted some advice on how to give his own theories on optics a stronger scientific underpinning who better to turn to than the Englishman? Harriot supplied Kepler with important data on the behavior of light rays passing through glass, but he also stimulated the German's interest in the sphere-packing problem. In response, Kepler published a little booklet titled *The Six-Cornered Snowflake* (1611) that would influence the science of crystallography for the next two centuries and that contained what has come to be known as **Kepler's conjecture** about the most efficient way to pack spheres.

canonical form

A form of any given **polyhedron** distorted so that every edge is **tangent** to the unit sphere and the center of gravity of the tangent points is the origin.

Cantor, Georg Ferdinand Ludwig Philipp (1845–1918)

A Russian-born German mathematician who founded **set theory** and introduced the concept of **transfinite numbers**. His shocking and counterintuitive ideas about **infinity** drew widespread criticism before being accepted as a cornerstone of modern mathematical theory.

Cantor was 11 when his family moved from St. Petersburg to Germany. Despite attempts to push him into the more lucrative field of engineering, he eventually won his father's approval to study math at the Polytechnic of Zurich. The following year, 1863, his father died and Cantor switched to the University of Berlin where he studied under some of the greats of the day, including Karl **Weierstrass** and Leopold **Kronecker**. After receiving his doctorate in 1867, he had trouble finding a good job and was forced to accept a position as an unpaid lecturer and later as an assistant professor at the backwater University of Halle. In 1872, he achieved his first breakthrough–and a promotion–by proving that if a **function** is continuous (in other words, its graph is smooth) throughout an interval, it can be represented by a unique trigonometric series. This work, suggested to him by his colleague Heinrich Heine, was crucial because it led Cantor to think about the relations between points, represented by **real numbers**, that make up an unbroken line–the so-called *continuum*. Cantor realized that **irrational numbers** can be represented as infinite sequences of **rational numbers**, so that they can be understood as geometric points on the real-number line, just as rational numbers can. He was now in uncharted territory and at odds with mathematical orthodoxy, which frowned on the idea of actual **infinity**; however, he found like-minded friends in Richard **Dedekind** and, later, Gösta **Mittag-Leffler**.

In 1873 to 1874 Cantor proved that the rational numbers could be paired off, one by one, with the natural numbers and were therefore countable, but that there was no such one-to-one correspondence with the real numbers. He then went on to show, incredibly, that there are exactly the same number of points on a short line as there are on an indefinitely long line, or on a plane, or in any mathematical space of higher dimensions. On this, he wrote to Dedekind: "I see it, but I don't believe it!"

By 1883, Cantor had abandoned his earlier reticence about dealing with irrationals only as sequences of rationals and started to think in terms of a new type of number–the **transfinite numbers**. The sets of natural numbers and of real numbers were, he reasoned, just two elements of a series of different kinds of infinity. This dramatic extension of the number system to allow for legitimate mathematics of the infinite was violently opposed. Henri **Poincaré** said that Cantor's theory of infinite sets would be regarded by future generations as "a disease from which one has recovered." Kronecker went further and did all he could to ridicule Cantor's ideas, suppress publication of his results, and block Cantor's ambition of gaining a position at the prestigious University of Berlin. In the spring of 1884 Cantor suffered the first of several attacks of depression, exacerbated if not induced by the negative reaction of his contemporaries. In between these attacks, he published further results but was increasingly troubled by his failure to prove the **continuum hypothesis**–his belief that the order of infinity of the real numbers came next after that of the natural numbers. Although his later years were spent in and out of a sanatoria, he lived long enough to see his ideas on set theory vindicated and be described by David **Hilbert** as "the finest product of mathematical genius and one of the supreme achievements of purely intellectual human activity."

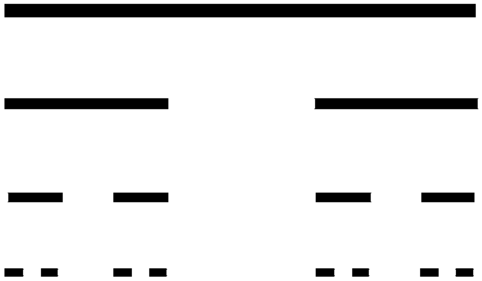

Cantor dust.

Cantor dust

Also known as the *Cantor set,* possibly the first pure **fractal** ever found. It was detected by Georg **Cantor** around 1872. To produce Cantor dust, start with a line segment, divide it in three equal smaller segments, take out the middle one, and repeat this process indefinitely. Although Cantor dust is riddled with infinitely many gaps, it still contains uncountably many points. It has a **fractal dimension** of log 2/log 3, or approximately 0.631. See also **Sierpinski carpet**.

cap

The symbol ∩ used to denote the **union** between two **set**s.

Cardano, Girolamo (1501–1570)

A celebrated Renaissance mathematician, physician, astrologer, and gambler, whose writings on the use of **negative number**s are the earliest known in Europe. As a physician, he gave one of the first clinical descriptions of typhoid fever. The illegitimate child of a mathematically gifted lawyer who was a friend of Leonardo da Vinci, he entered the University of Pavia in 1520 and later studied medicine at Padua. His eccentric and confrontational style earned him few friends, and he had trouble finding work. Eventually, he won a reputation as a physician and his services were highly valued at the courts.

Today, Cardano is remembered mostly for his achievements in algebra. He published the solutions to the **quartic** and **cubic equations** in his book *Ars magna* (1545). The solution to the cubic was communicated to him by Nicoló **Tartaglia** (who later claimed that Cardano had sworn not to reveal it, and became embroiled with Cardano in a decade-long fight), and the quartic was solved by Cardano's student Lodovico Ferrari. Both were acknowledged in the foreword of the book. Cardano was notoriously short of money and kept himself afloat by being an accomplished gambler and chess player. A book by him about games of chance, *Liber de Ludo Aleae* (Book on games of chance), written in the 1560s but published posthumously in 1663, contains the first systematic treatment of **probability theory**, as well as a section on effective cheating methods. Cardano invented several mechanical devices including the combination lock, the *Cardano suspension* (consisting of three concentric circles that allow a supported compass to rotate freely), and the *Cardan shaft,* which allows the transmission of rotary motion at various angles and is used in vehicles to this day. He made several contributions to hydrodynamics and claimed that perpetual motion is impossible, except in celestial bodies. He published two encyclopedias of natural science that contain a wide variety of inventions, facts, and occult superstitions.

Cardano led a beleaguered life. His elder and favorite son was executed in 1560 after he confessed to having poisoned his mercenary, cuckolding wife. Cardano's daughter was allegedly a prostitute who died from syphilis, prompting him to write a treatise about the disease. His younger son was a gambler who stole money from him. And Cardano himself was accused of heresy in 1570 because he computed the horoscope of Jesus Christ. Apparently, his own son contributed to the prosecution. Cardano was arrested and had to spend several months in prison, then was forced to abjure and had to give up his professorship. He moved to Rome, received a lifetime annuity from Pope Gregory XIII, and finished his not-uneventful autobiography. He died on the day he had (supposedly) astrologically predicted. See also **Chinese rings**.

Cardan's rings

See **Chinese rings**.

cardinal number

A number, often called simply a *cardinal,* that is used to count the objects or ideas in a **set** or collection: 0, 1, 2, . . . , 83, and so on. The *cardinality* of a set is just the number of elements the set contains. For finite sets this is always a **natural number**. To compare the sizes of two sets, *X* and *Y,* all that's necessary is to pair off the elements of *X* with those of *Y* and see if there are any left over. This concept is obvious in the case of finite sets but leads to some strange conclusions when dealing with infinite sets (see **infinity**). For example, it is possible to pair off all the natural numbers with all the even numbers, with none left over; thus the set of natural numbers and the set of even numbers have the same cardinality. In fact, an infinite set can be *defined* as any set that has a proper subset of the same cardinality. Every **countable set** that is infinite has a cardinality of **aleph**-null; the set of **real number**s has cardinality aleph-one. See also **ordinal number**.

cardioid

A heart-shaped curve first studied in 1674 by the Danish astronomer Ole Römer who was trying to find the

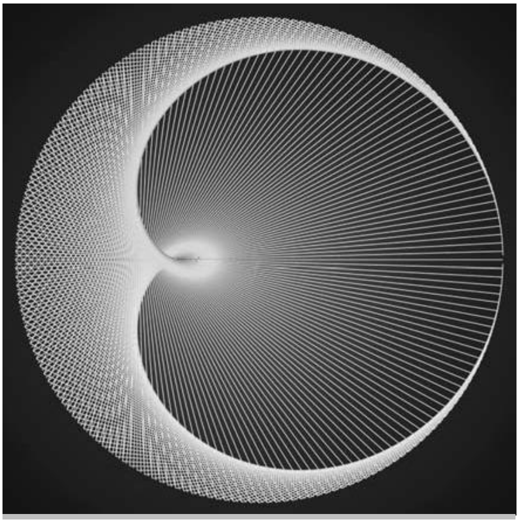

cardioid A cardioid curve spun by thread on a computer loom. *Jos Leys www.josleys.com*

best shape for gear teeth; the curve appears to have been named by Giovanni Salvemini de Castillon (1708–1791). When a circle rolls around another circle of the same size, any point on the moving circle traces out a cardioid. The Greeks used this fact when attempting to describe the motions of the planets. The cardioid is also the envelope of all circles with centers on a fixed circle, passing through one point on the fixed circle. In polar coordinates, it has the equation $r = 2a(1 - \cos\theta)$. It can also be described as an **epicycloid** with one cusp.

cards

The standard deck of 52 cards can be ordered in 52! (see **factorial**), or $8.065817517094 \times 10^{67}$ ways. There are various ways to **shuffle** cards in order to randomize them or to perform tricks with them. Each of the four kings in a deck represents a great leader from history: Charlemagne (hearts), Alexander the Great (clubs), Julius Caesar (diamonds), King David (spades). The king of hearts is the only one without a moustache. See also **blackjack**.

Carmichael number

Also known as an *absolute pseudoprime,* a number n that is a Fermat **pseudoprime** to any **base**, that is, it divides $(a^n - a)$ for any a. Another way of saying this is that a Carmichael number is actually a **composite number** even though **Fermat's little theorem** suggests it is probably a **prime number**. (Fermat's little theorem says that if P is a prime number, then for any number a, $(a^P - a)$ must be divisible by P. Carmichael numbers satisfy this condition to any base despite being composite.) There are only seven Carmichael numbers under 10,000 (they are 561, 1,105, 1,729, 2,465, 2,821, 6,601, and 8,911), and less than a quarter of a million of them under 10^{16}. Nevertheless, in 1994 it was proved that there are infinitely many of them. All Carmichael numbers are the product of at least three distinct primes, for example, $561 = 3 \times 11 \times 17$.

Carroll, Lewis (1832–1898)

The pen name of Charles Lutwidge Dodgson (obtained by anglicizing the Latin translation, "Carolus Lodovicus," of his first two names), an English mathematician, logician, and writer. Carroll was born at the Old Parsonage, Newton-by-Daresbury, Cheshire, his father being the vicar of All Saints Church, Daresbury. There is a commemorative window in the church, and a "Wonderland" weathervane, showing the Mad Hatter, the White Rabbit, and Alice, on the local primary school. Carroll was educated at Rugby School (the student mathematics society there is still called the Dodgson Society in his honor) and then at Christ Church, Oxford, at which college he was to spend the rest of his life, employed mainly as a lecturer.

Carroll's most famous book, *Alice's Adventures in Wonderland* (1865), grew out of a story he told on the

NEHSKIFD
Lewis Carroll's Chess
WORDGAME

Carroll, Lewis Lewis Carroll's Chess Wordgame, a game based on notation in Carroll's diaries with rules devised by Martin Gardner. *Kadon Enterprises, Inc., www.gamepuzzles.com* hot summer afternoon of July 4, 1862, when out rowing with the three young daughters of the Greek scholar H. G. Liddell, dean of Christ Church. Alice, named after Alice Liddell (later Hargreaves, 1852–1934), continued her adventures in *Through the Looking-Glass* (1871). Carroll wrote other books for children, including a long poem, "The Hunting of the Snark" (1876) and published several mathematical works, but was not distinguished academically. He stammered badly, never married, and seemed to find greatest pleasure in the company of little girls, with whom he lost his shyness. He was also an amateur pioneer in photography and an inventor of puzzles, games, ciphers, and mnemonics.[57–59] Carroll was a master of fantasy and his stories have their own logic. Carroll used puns and coined neologisms, including what he called "portmanteau words" like chortle (combining chuckle and snort). He played games with idioms, using such expressions as "beating time" (to music) in a literal sense. He reshaped such animals of fable or rhetoric as the Gryphon, the March Hare (said to have been inspired by a carved hare carrying a satchel located at St. Mary's Church, Beverly, Humberside, where Carroll visited), and the Cheshire Cat, and invented new ones, including the Bandersnatch and the Boojum.

PUZZLES

Here are a few examples of puzzles invented by Carroll:

1. You are given two glasses. One contains 50 tablespoons of milk, the other 50 tablespoons of water. Take one tablespoon of milk and mix it with the water. Now take one tablespoon of the water/milk mixture and mix it with the pure milk to obtain a milk/water mixture. Is there more water in the milk/water mixture or more milk in the water/milk mixture?
2. If you paint the faces of a cube with six different colors, how many ways are there to do this if each face is painted a different color and two colorings of the cube are considered equivalent if you can rotate one to get the other? What if we drop the restriction that the faces be painted different colors?
3. Make a word-ladder from FOUR to FIVE. (Every step in a word ladder differs from the previous step in exactly one letter and each step in the ladder is an English word.)
4. Why is a raven like a writing desk?

Solutions begin on page 369.

CARROLLIAN QUOTES

From *Alice's Adventures in Wonderland:*

- "The different branches of Arithmetic—Ambition, Distraction, Uglification, and Derision."
- "Then you should say what you mean," the March Hare went on.
  "I do," Alice hastily replied; "at least I mean what I say, that's the same thing, you know."
  "Not the same thing a bit!" said the Hatter. "Why, you might just as well say that 'I see what I eat' is the same thing as 'I eat what I see!' "
- "Take some more tea," the March Hare said to Alice, very earnestly.
  "I've had nothing yet," Alice replied in an offended tone, "so I can't take more."
  "You mean you can't take *less,*" said the Hatter. "It's very easy to take *more* than nothing."

From *The Hunting of the Snark:*

- "What I tell you three times is true."

From *Alice through the Looking Glass*

- "Can you do addition?" the White Queen asked.
  "What's one and one and one and one and one and one and one and one and one and one?"
  "I don't know," said Alice. "I lost count."
- "It's very good jam," said the Queen.
  "Well, I don't want any to-day, at any rate."
  "You couldn't have it if you did want it," the Queen said. "The rule is jam tomorrow and jam yesterday but never jam to-day."
  "It must come sometimes to "jam to-day," Alice objected.
  "No it can't," said the Queen. "It's jam every other day; to-day isn't any other day, you know."
  "I don't understand you," said Alice. "It's dreadfully confusing."
- "When I use a word," Humpty Dumpty said, in a rather scornful tone, "it means just what I choose it to mean—neither more nor less."
  "The question is," said Alice, "whether you can make words mean so many different things."
  "The question is," said Humpty Dumpty, "which is to be master—that's all."

Cartesian geometry

See **analytical geometry**.

Cartesian coordinates

An ordered set of **real number**s that defines the position of a point in terms of its projection onto mutually perpendicular number lines. In the plane, each point is defined by two such projections, one onto the *x*-axis and one onto the *y*-axis, and is written as an **ordered pair** of real numbers (x, y). The same system works equally in spaces of three or more dimensions.

Cartesian oval

A curve that actually consists of two ovals, one inside the other. It is the locus of a point whose distances s and t from two fixed points S and T satisfy the equation $s + mt = a$. When c is the distance between S and T then the curve can be expressed in the form:

$$((1 - m^2)(x^2 + y^2) + 2m^2cx + a^2 - m^2c^2)^2 = 4a^2(x^2 + y^2)$$

The curves were first studied by René **Descartes** in 1637 and are sometimes called the *ovals of Descartes;* they were also investigated by Isaac **Newton** in his classification of **cubic curves**. If $m = \pm 1$, then the Cartesian oval is a central conic. If $m = a/c$, then it becomes a **limaçon of Pascal**, in which case the inside oval touches the outside one. Cartesian ovals are **anallagmatic curves**.

Cassinian ovals

Also known as *Cassini's ovals,* a family of curves, each member of which is defined as follows: given two points A and B and a constant c^2, the locus of points P with $PA \times PB = c^2$. The locus has the equation $(x^2 + y^2)^2 - 2a^2(x^2 - y^2)^2 - a^4 + c^4 = 0$, where $a = AB$. Equivalently, Cassinian ovals can be thought of as the set of curves produced when a circular **torus** is sliced at every possible point parallel to its axis. If $c = a$, then the curve is a special case known as the **lemniscate of Bernoulli** (a figure-eight type curve). The ovals are named after the Italian-born astronomer Giovanni Cassini (1625–1712) who first investigated them in 1680 while studying the relative motions of Earth and the Sun. Cassini thought that the Sun traveled round Earth on one of these curves (rather than the **ellipse**, as correctly proposed in Kepler's heliocentric scheme), with Earth at one focus.

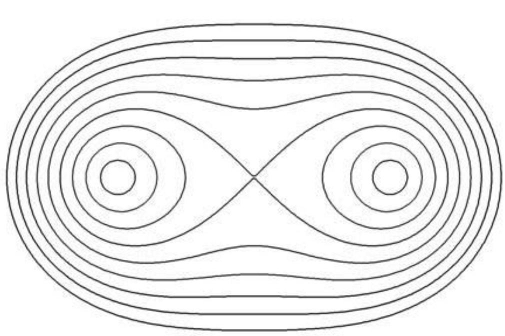

Cassinian ovals The many different ways to slice a doughnut. *Xah Lee, www.xahlee.org*

casting out nines

A method for checking arithmetic that uses the idea of the **digital root** of a number. Let the digital root of a number n be $r(n)$; for example, $r(7{,}586) = 8$. For any two numbers a and b: $r(a + b) = r(r(a) + r(b))$ and $r(a \times b) = r(r(a) \times r(b))$. These rules allow checks on addition and multiplication as the following examples show. Does $7{,}586 + 9{,}492 = 16{,}978$? $r(r(7{,}586) + r(9{,}492)) = r(8 + 6) = 5$; $r(16{,}978) = 4$; so the sum given is incorrect. Does $7{,}586 \times 9{,}492 = 72{,}006{,}312$. $r(r(7{,}586) \times r(9{,}492)) = r(8 \times 6) = r(48) = 3$; $r(72{,}006{,}312) = r(21) = 3$; which suggests that the product given is likely to be correct. The name "casting out nines" comes from the fact that nines need not be included in the calculation of the digital roots, since they have no effect on the final result. This a direct outcome of the fact that we use a **decimal** number system. If we calculated instead in octal (base eight), say, then the process would be one of "casting out sevens." This kind of checking will pick up most errors, but not all. For example, an interchange of two digits will not be detected, nor will replacing a nine by a zero or vice versa. The method appears in the work of ninth-century Arab mathematicians but may have originated earlier with the Greeks and, possibly, the Hindus.

Catalan number

Any number, u_n, from the *Catalan sequence* defined by

$$u_n = (2n)! \,/\, (n + 1)!n!.$$

It begins: 1, 2, 5, 14, 42, 132, 429, 1,430, 4,862, 16,796, 58,786, 208,012, 742,900, ... The values of u_n represent the number of ways a **polygon** with $n + 2$ sides can be cut into n triangles using straight lines joining vertices (see **vertex**). Catalan numbers are named after the Belgian mathematician Eugène Catalan (1814–1894). They also arise in other counting problems, for example in determining how many ways $2n$ beans can be divided into two containers if one container can never have less than the second.

Catalan solid

A **polyhedron** that is a dual of an **Archimedean solid**. (A dual of a polyhedron is obtained by replacing each face with a vertex, and each vertex with a face.) Catalan solids are named after the Belgian mathematician Eugène Catalan (1814–1894) who first described them in 1865. See also **Platonic solid** and **Johnson solid**. (See table, "Catalan solids.")

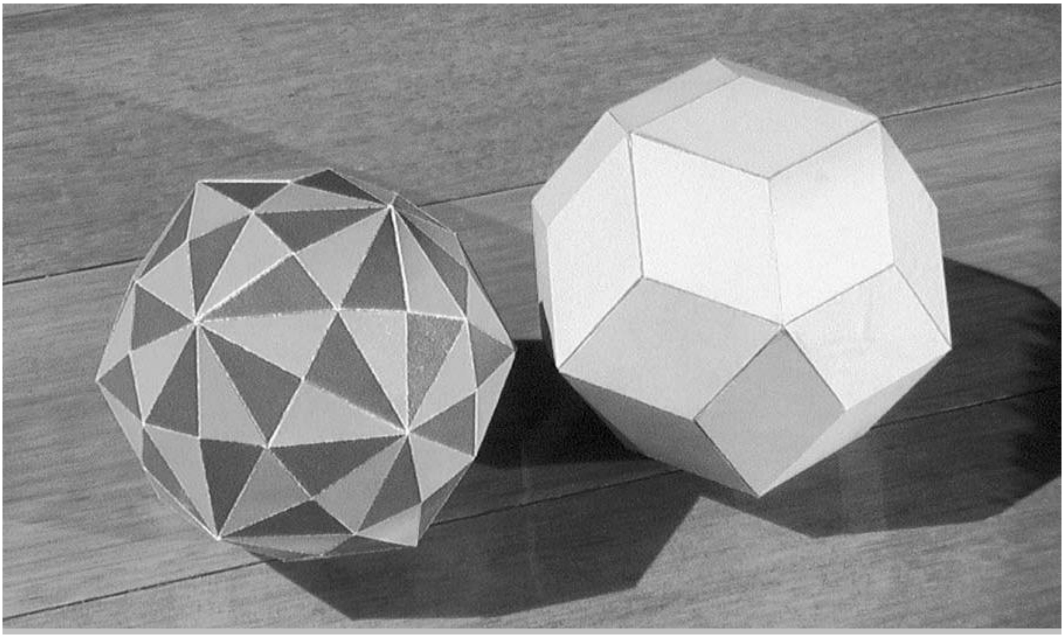

Catalan solid Two of the Catalan solids: the rhombic tricontahedron (right) and the disdyakistriacontahedron (left). *Robert Webb, www.software3d.com; created using Webb's Stella program*

Catalan Solids

Name	Corresponding Archimedean Solid
Triakis tetrahedron	Truncated tetrahedron
Rhombic dodecahedron	Cuboctahedron
Triakis octahedron	Truncated cube
Tetrakis hexahedron	Truncated octahedron
Deltoidal icositetrahedron	Small rhombicuboctahedron
Disdyakis dodecahedron	Great rhombicuboctahedron
Pentagonal icositetrahedron	Snub cube
Rhombic triacontahedron	Icosidodecahedron
Triakis icosahedron	Truncated dodecahedron
Pentakis dodecahedron	Truncated icosahedron
Deltoidal hexecontahedron	Rhombicosidodecahedron
Disdyakis triacontahedron	Great rhombicosidodecahedron
Pentagonal hexecontahedron	Snub dodecahedron

Catalan's conjecture

The hypothesis, put forward by the Belgian mathematician Eugène Catalan (1814–1894) in 1844, that 8 (= 2^3) and 9 (= 3^2) are the only pair of consecutive powers. In other words, the *Catalan equation* for **prime numbers** p and q and positive integers x and y

$$x^p - y^q = 1$$

has only the one solution

$$3^2 - 2^3 = 1.$$

In 1976 R. Tijdeman took the first major step toward showing this is true by proving that for any solution, y^q is less than **e** to the power e to the power e to the power e to the power 730 (a huge number!). Since then, this bound has been reduced many times, and it is now know that the larger of p and q is at most 7.78×10^{16} and the smaller is at least 10^7. On April 18, 2002, the Romanian number theorist Preda Mihailescu sent a manuscript to several mathematicians with a proof of the entire conjecture together with an analysis by Yuri Bilu. It is expected that as soon as this work is completely reviewed by other mathematicians that Catalan's conjecture will have been proved.[266]

Solutions to Catalan's conjecture and **Fermat's last theorem** are special cases of the *Fermat-Catalan equation:*

$$x^p + y^q = z^r,$$

where x, y, and z are positive, **coprime** integers and the exponents are all primes with

$$1/p + 1/q + 1/r \leq 1.$$

The *Fermat-Catalan conjecture* is that there are only finitely many solutions to this system. These solutions include:

$1^p + 2^3 = 3^2$ ($p \geq 2$); $2^5 + 7^2 = 3^4$
$13^2 + 7^3 = 2^9$;
$33^8 + 1{,}549{,}034^2 = 15{,}613^3$
and $43^8 + 96{,}222^3 = 30{,}042{,}907^2$.

Catalan's constant

A constant that crops up regularly in combinatorial problems, especially in the evaluation of certain infinite series and integrals. For example, it is equal to

$\int_0^1 \arctan(x) / x \, dx$, and
$1 - 1/3^2 + 1/5^2 - 1/7^2 + 1/9^2 - \dots$

It is also the solution to the following problem as n becomes arbitrarily large: If you have a $2n \times 2n$ checkerboard and a supply of $2n^2$ dominoes that are just large enough to cover two squares of the checkerboard, how many ways are there to cover the whole board with the dominoes? Catalan's constant has the value 0.915965 . . . ; it is not known if it's an **irrational number**.

catastrophe theory

A theory, developed by the French mathematician René Thom (1923–2003), that attempts to explain the behavior of complex **dynamical system**s by relating it to **topology**. The evolution of such systems consists of steady continuous change interspersed with sudden major jumps, or "catastrophes," when the topology of the set changes. Catastrophe theory has been applied, with varying degrees of success, to phenomena as diverse as earthquakes, stock market crashes, prison riots, and human conflicts, at the personal, group, and societal level. The theory was first developed by Thom in a paper published in 1968 but became well known through his book *Structural Stability and Morphogenesis* (1972).[331] Many mathematicians took up the study of catastrophe theory and it was in tremendous vogue for a while, yet it never achieved the success that its younger cousin **chaos** theory has because it failed to live up to its promise of useful predictions. Late in his career, the surrealist Salvador Dali painted *Topological Abduction of Europe: Homage to René Thom* (1983), an aerial view of a seismically fractured landscape juxtaposed with the equation that strives to explain it.

catch-22

A situation in which a person is frustrated by a paradoxical rule or set of circumstances that preclude any attempt to escape from them. The name comes from the title of a novel by Joseph Heller (1923–1999), based on his personal experiences, about an American airman's attempts to survive the madness of World War II. Heller wrote:

> There was only one catch and that was Catch-22, which specified that concern for one's own safety in the face of dangers that were real and immediate was the process of a rational mind. Orr was crazy and could be grounded. All he had to do was ask; and as soon as he did, he would no longer be crazy and would have to fly more missions. Orr would be crazy to fly more missions and sane if he didn't, but if he was sane he had to fly them. If he flew them he was crazy and didn't have to; but if he didn't want to he was sane and had to.

category theory

The study of abstracted collections of mathematical objects, such as the category of **set**s or the category of **vector space**s, together with abstracted operations sending one object to another, such as the collection of **function**s from one set to another or linear **transformation**s from one vector space to another.

catenary

The shape that a rope or telephone cable makes, under the influence of gravity, when suspended between two points. The word comes from the Latin *catena,* meaning "chain," and was first used by Christiaan **Huygens** while studying the form of suspended chains. Galileo thought the shape would be a **parabola**. In fact, near the **vertex**, a parabola and a catenary do look very similar. When x is slightly greater than three, however, the catenary begins to rapidly outgrow the value of the parabola. The two shapes are related in another way. If a parabola is rolled along a straight line, the **focus** of the parabola moves along a catenary curve. Surprisingly, too, if a bicycle with square (or any polygon-shaped) wheels is ridden along a road made of upturned catenaries the wheels will roll smoothly and the rider will stay at the same height! The St. Louis Arch, which is 192 meters wide at the base and 192 meters tall, follows the form of a catenary, the exact formula for which is displayed inside the arch: $y = 68.8 \cosh(0.01x - 1)$, where cosh is the hyperbolic cosine function.

The general equation of a catenary can be written

$$y = k \cosh(x/k),$$

where k is a constant, or, in terms of the **exponential** function,

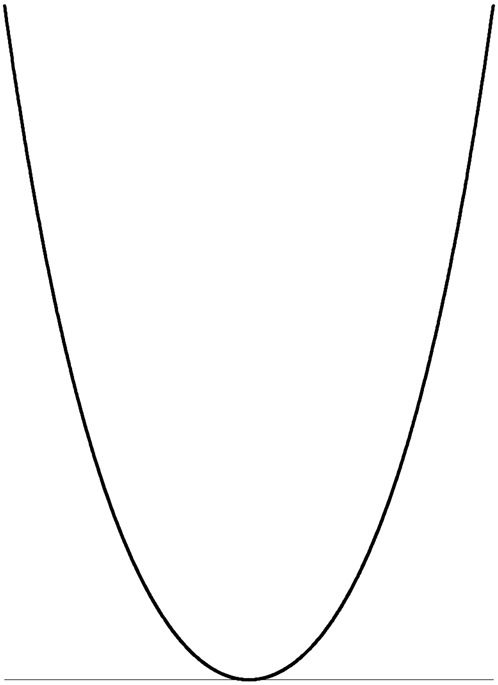

catenary The catenary curve. © *Jan Wassenaar, www.2dcurves.com*

$$y = k(e^{x/k} + e^{-x/k})/2.$$

In the special case where $k = 1$, these reduce to

$$y = \cosh(x) = (e^x + e^{-x})/2.$$

In terms of a polynomial series

$$y = 0.5\ (1 + x + x^2/2! + x^3/3! + x^4/4! + \ldots + 1 - x + x^2/2! - x^3/3! + x^4/4! - \ldots)$$
$$= 1 + x^2/2! + x^4/4! + x^6/6! + \ldots.$$

For small values of x the terms beyond $x^2/2!$ are very small, so that the equation closely approximates that of a parabola, as we have already seen.

catenoid

The **surface of revolution** produced when a **catenary** rotates about its central axis. The catenoid was first described by Leonhard **Euler** in 1740 and is the oldest known **minimal surface** (a shape of least area when bounded by a given closed space). It is the minimal surface connecting two parallel circles of unequal diameter on the same axis; soap film between two circular rings takes this form (see also **bubbles**). The catenoid is the only known minimal surface that is also a surface of revolution, and is one of only four minimal surfaces that have the topological properties of being unbounded, embedded, and non-periodic; the others are the simple plane, the **helicoid**, and *Costa's surface.*

cathetus

A line that is perpendicular to another line. Usually, it refers to one of the lines in a right triangle that is not the **hypotenuse**.

Cauchy, Augustin Louis, Baron (1789–1857)

A French mathematician who founded **complex analysis** by discovering the *Cauchy-Riemann equations* and wrote 789 papers–an output surpassed only by Leonhard **Euler**, George **Cayley**, and Paul **Erdös**. He coined the name for the **determinant** and systematized its study and gave nearly modern definitions of **limit**, **continuity**, and **convergence**.

causality

The relationship between causes and effects. An event or state of affairs A is the cause of an event B if A is the reason that brings about the effect B. For instance, one might say, "my pushing the gas pedal *caused* the car to go faster." An important question in philosophy and other fields is how (and if) causes can bring about effects. In a strict reading, if A causes B, then A must *always* be followed by B. In this sense, for example, smoking doesn't cause cancer. In everyday usage, we therefore often take "A causes B" to mean "A causes an increase in the probability of B." The establishment of cause and effect, even with this relaxed reading, is notoriously difficult. The Scottish philosopher David Hume held that causes and effects are not real, but instead are imagined by our minds to make sense of the observation that A often occurs together with or slightly before B. All we can actually observe are *correlations,* not causations. This is also expressed in the logical fallacy, "correlation implies causation." For instance, the observation that smokers have a dramatically increased lung cancer rate doesn't establish that smoking must be the *cause* of that increased cancer rate: maybe there exists a certain genetic defect which both causes cancer and a yearning for nicotine.[194]

caustic

The envelope of rays of light reflected (or refracted) by a given curve from a given point source of light; a *catacaustic* results from reflection, a *diacaustic* from refraction. Among the caustic curves of a circle are a lima, if the light source is nearby, a **nephroid**, if the source is at infinity, and a **cardioid**, if the source is on the circle.

Cavalieri's principle

If two solids have the same height and the same cross-sectional area at every level, then they have the same volume. This principle is named after the Italian mathematician Bonaventura Cavalieri (1598–1647).

Cayley, Arthur (1821–1895)

A British mathematician who made important contributions to **non-Euclidean geometry** and the algebra of matrices (see **matrix**). The former eventually found its way into the study of the **space-time** continuum, the latter into a formulation of **quantum mechanics** by the German physicist Werner Heisenberg. Cayley was also far ahead of his time in pioneering the idea of abstract **group**s.

Cayley number

See **octonion**.

Cayley's mousetrap

A **permutation** problem invented by Arthur **Cayley**. Write the numbers 1, 2, . . . , n on a set of cards and shuffle the deck. Start counting using the top card. If the card chosen does not equal the count, move it to the bottom of the deck and continue counting forward. If the card chosen *does* equal the count, discard the chosen card and begin counting again at 1. The game is won if all cards are discarded, and lost if the count reaches $n + 1$. The number of ways the cards can be arranged such that at least one card is in the proper place for $n = 1, 2, \ldots$, are 1, 1, 4, 15, 76, 455, ...

Cayley's sextic

A sinusoidal spiral curve described by the Cartesian equation

$$4(x^2 + y^2 - ax)^3 = 27a^2(x^2 + y^2)^2.$$

It was discovered by Colin **Maclaurin** but was first studied in detail by Arthur **Cayley** and named after him by R. C. Archibald in 1900.

ceiling

The largest value that something can take. The *ceiling function* of a number x is the smallest integer that is not smaller than x.

cell

(1) A three-dimensional object that is part of a higher-dimensional object, such as a **polychoron**. A cell is related to higher-dimensional objects in the way that a face, or (two-dimensional) **polygon**, is related to higher-dimensional objects. For example, a cell is to a Four-dimensional polytope, or polychoron, what a face is to a three-dimensional polytope, or **polyhedron**. Often polytopes are classified simply by how many cells they have. For example, the **tesseract** has eight cells, each one of which is a cube. (2) The fundamental spatial unit operated on by the rules of a **cellular automaton** during one generation.

cellular automaton

An array of cells that evolves according to a set of rules based on the states of surrounding cells; for example, a cell might be "on" if its four neighbor cells (east, west, north, and south) are also on. The entire array can self-organize into global patterns that may move around the screen. These patterns can be quite complex even though they emerge from just a few very simple rules governing the connections among the cells. Cellular automata are the simplest models of spatially distributed processes. They were first investigated by John **von Neumann** in about 1952. Von Neumann incorporated a cellular model into his "universal constructor" and also proved that an automaton consisting of cells with four orthogonal neighbors and 29 possible states would be capable of simulating a **Turing machine** for some configuration of about 200,000 cells. The best-known cellular automaton is Life (see **Life, Conway's game of**). Another example is **Langton's ant**. The study of cellular automata and their patterns has led to insights into the way structure is built up in biological and other complex systems, and for this reason forms part of the subject of **artificial life**.

celt

Also known as a *rattleback,* a simple ancient toy that behaves in a very counterintuitive way. When spun one way about its vertical axis, the celt spins for a long time. When spun the other way, however, a wobble quickly sets in that halts the rotation and then, incredibly, reverses it. In his 1986 paper on the subject, the British physicist Hermann Bondi wrote: "Many people, even trained scientists, find it hard to understand that the behaviour of the toy doesn't violate the principle of conservation of angular momentum."[47] The celt's remarkable antics stem from three factors: a curved base that has two different radii–one long radius for the lengthwise curve and one shorter radius for the tighter curve across the width; axes of symmetry that are skewed slightly from the principal axes of inertia; and a different distribution of mass about each of the two horizontal axes of inertia. To understand how the celt can switch direction halfway through its performance, think of the frictional force that acts at the point of contact between the celt and the surface. One component of the friction creates a torque (twisting force) that tends to rotate the celt about its vertical axis. The point of contact is moving all the time and the torque changes. If the inertial and symmetrical axes coincided, the average torque over a single oscillation would be zero. But for the celt, there's a net torque in one direction, which reverses the angular momentum. See also **Tippee Top**.

center of perspective

The point where the lines joining corresponding points of two figures that are in perspective meet.

centillion

See **large number**.

central angle

The angle subtended at the center of a **circle** by an **arc** or a **chord**; in other words, the angle between two radii.

centroid

For a **triangle**, the point of intersection of the **median**s. For any other shape, the point where coordinates are the average of the coordinates of the shape's vertices (see **vertex**). The centroid is the *center of mass* of a figure.

century

A period of 100 years. The original Latin *centuria* means simply "one hundred" and was used to describe any collection of 100 items. In the Roman army, a century was a group of 100 men, each known as a *centurion.* One of the few modern examples of "century" being used other than to denote a period of time is in the game of cricket where a batsman who scores 100 runs in an inning is said to have "made a century."

Ceva, Giovanni (1647–1734)

A Jesuit-trained Italian mathematician who specialized in **geometry**. His greatest discovery, now known as *Ceva's theorem,* can be stated as follows. Given a triangle with vertices (corners), A, B, and C and points D, E, and F on the opposite sides, the lines AD, BE, and CF will intersect at a single point if $BD \times CE \times AF = DC \times EA \times FB$. The term *cevian line* was coined by French geometers around the end of the eighteenth century to honor Ceva. It is defined as any line joining a vertex of a triangle to a point on the opposite side. The **median**, **altitude**, and angle **bisect**or are all examples of cevians. The perpendicular bisector, however, in most cases, is not a cevian because it doesn't usually pass through a vertex.

chained arrow notation

See **Conway's chained arrow notation**.

Chaitin, Gregory (1947–)

An American mathematician and computer scientist at IBM's T. J. Watson Research Center who is the chief architect of a new subject known as *algorithmic information theory,* which has profound consequences for our ideas about **random**ness. In particular, because of the limitations of computers and the programs they run, Chaitin has shown that there is an inherent uncertainty or unknowability in mathematics that is similar to the uncertainty principle in physics. Although there are an infinite number of mathematical facts, they are, for the most part, unrelated to each other and impossible to tie together with unifying theorems. His powerful message is that most of mathematics is true for no particular reason; math is true by accident. See **Chaitin's constant**.

Chaitin's constant

A **real number**, represented by capital omega (Ω) and also known as the *Halting probability,* whose digits are distributed so **random**ly that no rule can be found to predict them. Discovered by Gregory **Chaitin**, Ω is definable but not computable. It has no pattern or structure to it whatsoever, but consists instead of an infinitely long string of zeros and ones in which each digit is as unrelated to its predecessor as one coin toss is to the next. Although called a constant, it is not a constant in the sense that, for example, **pi** is, since its definition depends on the arbitrary choice of computation model and programming language. For each such model or language, Ω is the probability that a randomly produced string will represent a program that, when run, will eventually halt. To derive it, Chaitin considered all the possible programs that a hypothetical computer known as a **Turing machine** could run, and then looked for the probability that a program, chosen at random from among all the possible programs, will halt. He eventually showed that this halting probability turns Turing's question of whether a program ever stops into a real number, somewhere between zero and one. He further showed that, just as there are no computable instructions for deciding in advance whether a computer will halt, there are also no instructions for determining the digits of Ω. Omega is uncomputable and unknowable: we don't know its value for any programming language and we never will. This is extraordinary enough in itself, but Chaitin has found that Ω permeates the whole of mathematics, placing fundamental limits on what we can know.

And Ω is just the beginning. There are more disturbing numbers called Super-Omegas, whose degree of randomness is vastly greater even than that of Ω. If there were an omnipotent computer that could solve the halting problem and evaluate Ω, this mega-brain would have its own unknowable halting probability called Ω′. And if there were a still more godlike machine that could find Ω′, its halting probability would be Ω″. These higher Omegas, it has been recently discovered, are not meaningless abstractions. Ω′, for instance, gives the probability that an infinite computation produces only a finite amount of output. Ω″ is equivalent to the probability that, during an infinite computation, a computer will fail to produce an output–for example, get no result from a computation and move on to the next one–and that it will do this only a finite number of times. Omega and the Omega hierarchy are revealing to mathematicians an unsettling truth: the problems that we can hope ever to solve form a tiny archipelago in a vast ocean of undecidability.[130]

Champernowne's number

The first known **normal number**. It was discovered in 1933 by the English mathematician David G. Champernowne and consists of a decimal fraction in which the decimal integers are written down in increasing order: 0.12345678910111213 . . . . Champernowne's number has been proven to be a normal number in base 10 and also to be an **irrational number**. However, although its digits appear with equal frequency, the sequence of its digits are not unpredictable. An example of a number whose sequence of digits *is* unpredictable is **Chaitin's constant**.

chance

See **probability theory**.

change ringing

> *The art of change ringing is peculiar to the English, and, like most English peculiarities, unintelligible to the rest of the world. To the musical Belgian, for example, it appears that the proper thing to do with a carefully tuned ring of bells is to play a tune upon it. By the English campanologist . . . the proper use of the bells is to work out mathematical permutations and combinations.*
>
> –Dorothy L. Sayers, *The Nine Tailors*

The ringing of a set of bells in a precise relationship to one another to produce a pleasing sound. Bells are numbered 1, 2, 3, 4, 5 . . . from lightest (highest-pitched) to heaviest. After each sequence, or *round,* the order of the bells is changed slightly in a predetermined way. With 5 bells, there are 5 × 4 × 3 × 2 × 1, or 120, possible changes, which take about 4 minutes to ring. With 6, 7, or 8 bells, the number of unique changes is 720, 5,040, and 40,320, respectively. To produce pleasing variations in the sound, bells are made to change places with adjacent bells in the row, for example:

1 2 3 4 5 6 7 8
2 1 4 3 6 5 8 7

These rows are the musical notation of change ringing. No bell moves more than one place in the row at a time, although more than one pair may change in the same row. In order to ring a different row with each pull of the rope, ringers have devised methods for changing pairs in orderly ways. In ringing a method, the bells begin in rounds, ring changes according to the method, and return to rounds without repeating any row along the way. These place changes produce musical patterns, with the sounds of the bells weaving in and out. For example, a "Plain Hunt Minimus" with four bells is rung as shown in the diagram.

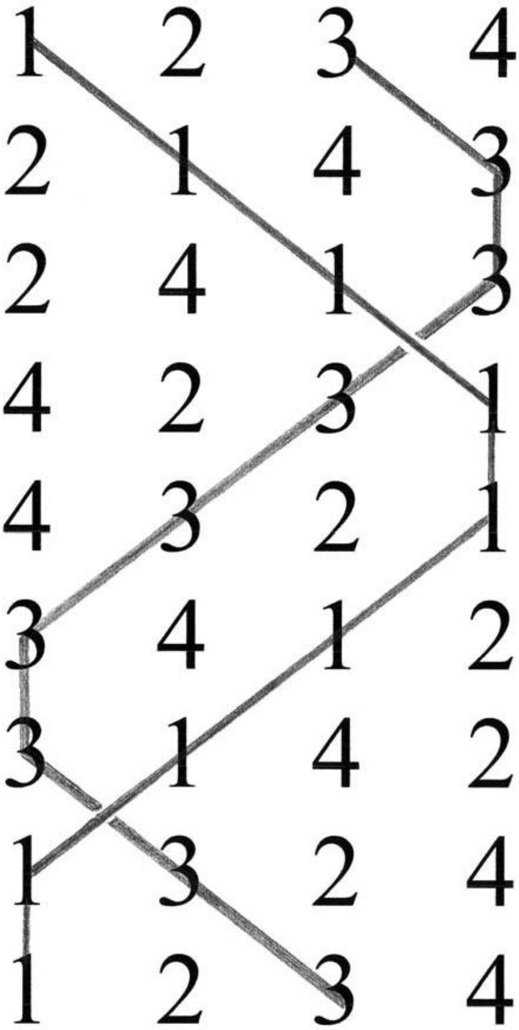

change ringing The "Plain Hunt Minimus" for four bells. The sequences for bells 1 and 3 are shown by lines.

Experienced ringers test and extend their abilities by ringing peals: 5,000 or more changes without breaks or repeating a row. Peals customarily last about three hours. The first peal was rung in England in 1715. Chiming bells (swinging them through a short arc using a rope and a lever) goes well back into the Middle Ages, but it wasn't until the seventeenth century that ringers developed the full wheel, which allowed enough control for orderly ringing. In 1668 Fabian Stedman published *Tintinnalogia* (The art of change ringing), containing all the available information on systematic ringing. The theory of change ringing set forth by Stedman has been refined in later years but remains essentially unchanged today. Bells for change ringing are hung in stout frames that allow the bells to swing through 360°. Each bell is attached to a wooden wheel with a handmade rope running around it and takes about 2 seconds to rotate. The bells are arranged in the frame so their ropes hang in a circle in the ringing chamber below. Into each rope is woven a tuft of brightly colored wool (a *sally*), which marks where the ringer must catch the rope while ringing. Bells are rung from the "mouth up" position. With a pull of the rope, the bell swings through a full circle to the up position again. With the next pull it swings back in the other direction. The plot of Dorothy Sayers's *The Nine Tailors* (1934), considered one of her best works, revolves around the art of change ringing.

chaos

> *We adore chaos because we love to produce order.*
>
> –M. C. Escher

A phenomenon shown by some **dynamical systems**, which consists of a curious, infinitely complex pattern of behavior that lies just beyond the edge of total order. A system is chaotic if it is predictable in principle and yet is unpredictable in practice over long periods because its behavior depends very sensitively on initial conditions. Despite this unpredictability, however, there are certain constants, such as **Feigenbaum's constant**, and certain structures, such as **chaotic attractors**, that are fixed and susceptible to analysis. The weather, the movements of a metal pendulum moving over fixed magnets, and the orbits of closely spaced moons are all examples of chaotic systems. Although the ideas behind modern chaos theory were actively studied at some level throughout most of the twentieth century, the word as a mathematical term dates only from an article in *American Mathematical Monthly* in 1975 called "Period Three Implies Chaos."

In everyday language, chaos has come to mean the exact opposite of order. But the Greek root *khaox* means "empty space" and this meaning still persists in archaic usage where it refers to a canyon or abyss. The evolution of the word to mean disorder seems to come from reference to the time before God created the universe. Empty space was formless and the creation filled the emptiness and established order. Mathematical chaos represents an unexpected third state: a deterministic system subject to simple rules that nevertheless displays infinitely complex behavior.[135]

chaos tiles

See **Penrose tiling**.

chaotic attractor

Also known as a *strange attractor,* a type of **attractor** (i.e., an attracting set of states) in a complex **dynamical system**'s **phase space** that shows **sensitivity** to initial conditions. Because of this property, once the system is on the attractor, nearby states diverge from each other exponentially fast. Consequently, small amounts of noise are amplified. Once sufficiently amplified the noise determines the system's large-scale behavior and the system is then unpredictable. Chaotic attractors themselves are markedly patterned, often having elegant, fixed geometric structures, despite the fact that the trajectories moving within them appear unpredictable. The chaotic attractor's geometric shape is the order underlying the apparent chaos. It functions in much the same way as someone kneading dough. The local separation of trajectories corresponds to stretching the dough and the global attraction property corresponds to folding the stretched dough back onto itself. One result of the stretch-and-fold aspect of chaotic attractors is that they are **fractal**s; that is, some cross section of them reveals similar structure on all scales.

character theory

The study of the traces (sums of the diagonal elements) of the **matrix** representations of a **group**. The information gained is listed in *character tables,* the properties of which give insight into the group's properties.

chess

A game of strategy for two players that probably originated in India, though the earliest documentary references are in Chinese and Persian texts in about A.D. 600. Each player has 16 pieces, either black or white, consisting of eight pawns, two rooks (also known as castles), two knights, two bishops, a queen, and a king. The object is to lay siege to the opposing king in such a way that it cannot escape attack–a position known as *checkmate* (from the Persian phrase *Shah Mat,* meaning "the king is dead"). There are 400 first-move combinations–20 for white × 20 for black (though only 64 of these are regarded as strong), 318,979,564,000 ways of playing the first four moves, and 169,518,829,100,544,000 trillion ways of playing the first

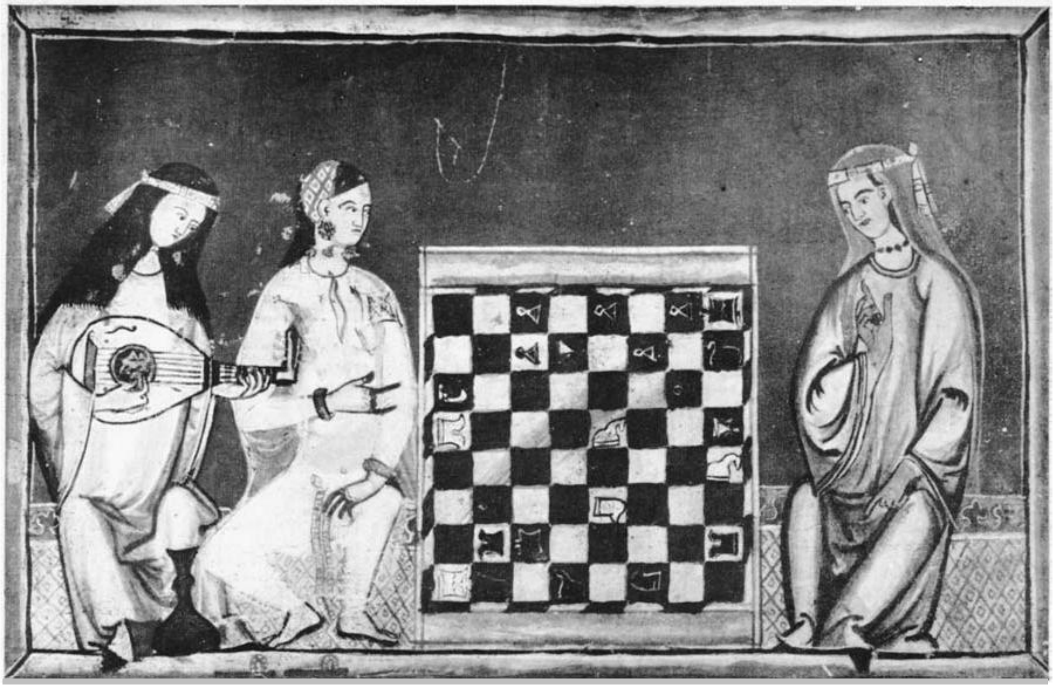

chess An illumination from the *Cantigas de Santa Maria* (a thirteenth-century collection of songs) showing a chess game in progress.

10 moves. The total number of possible board configurations is estimated at 10^{120}; for comparison, that of **Go** is generally put at about 10^{174}.

A standard chessboard is a square plane divided into 64 smaller squares by straight lines at right angles. Originally, it wasn't checkered (that is, made with its rows and columns alternately of dark and light colors), and this feature was introduced merely to help the eye in actual play. In many puzzles based on chess the utility of checkering is questionable, and the board may be generalized to any $n \times n$ size.

One of the first puzzles to use a chessboard was the **wheat and chessboard problem**, posed in 1256 by the Arabic mathematician Ibn Kallikan. Among the earliest problems to involve chess *pieces*, proposed by Guarini di Forli in 1512, asks how two white and two black knights can be interchanged, using normal knight's moves, if they are placed at the corners of a 3 × 3 board. The unusual L-shaped movement of the knight is what makes one of the best known chess puzzles, the **knight's tour**, such a challenge. Other standard puzzles, often called simply the **kings problem**, the **queens puzzle**, the **rooks problem**, the **bishops problem**, and the **knights problem**, ask for the greatest number of each of these pieces that can be placed on an 8 × 8 board or on a generalized $n \times n$ board without attacking each other, and/or the smallest number of each of these pieces that are needed to occupy or attack every square. *Fairy chess* is any variant on the standard game, which may involve a change in the form of the board, the rules of play, or the pieces used. For example, the normal rules of chess can be used but with a cylindrical or **Möbius band** connection of the edges.

Chinese cross

See **burr puzzle**.

Chinese remainder theorem

If there are n numbers, a_1 to a_n, that have no **factor**s in common (i.e., are pairwise relatively prime), then any integer greater than or equal to 0 and less than the product of all the numbers n can be uniquely represented by a series consisting of the remainders of division by the numbers n. For example, if $a_1 = 3$ and $a_2 = 5$, the Chinese remainder theorem (CRT) says that every integer from 0 to 14 will have a unique set of remainders when divided separately by (**modulo**) 3 and 5. Listing out all the possibilities shows that this is true:

0 has a remainder of 0 modulo 3 and a remainder of 0 modulo 5.

1 has a remainder of 1 modulo 3 and a remainder of 1 modulo 5.

2 has a remainder of 2 modulo 3 and a remainder of 2 modulo 5.

3 has a remainder of 0 modulo 3 and a remainder of 3 modulo 5.

4 has a remainder of 1 modulo 3 and a remainder of 4 modulo 5.

5 has a remainder of 2 modulo 3 and a remainder of 0 modulo 5.

6 has a remainder of 0 modulo 3 and a remainder of 1 modulo 5.

7 has a remainder of 1 modulo 3 and a remainder of 2 modulo 5.

8 has a remainder of 2 modulo 3 and a remainder of 3 modulo 5.

9 has a remainder of 0 modulo 3 and a remainder of 4 modulo 5.

10 has a remainder of 1 modulo 3 and a remainder of 0 modulo 5.

11 has a remainder of 2 modulo 3 and a remainder of 1 modulo 5.

12 has a remainder of 0 modulo 3 and a remainder of 2 modulo 5.

13 has a remainder of 1 modulo 3 and a remainder of 3 modulo 5.

14 has a remainder of 2 modulo 3 and a remainder of 4 modulo 5.

CRT enables problems such as the following to be solved: Find the two smallest counting numbers that will each have the remainders 2, 3, and 2 when divided by 3, 5, and 7, respectively. It is said that the ancient Chinese used a variant of this theorem to count their soldiers by having them line up in rectangles of 7 by 7, 11 by 11, and so forth. After counting only the remainders, they solved the associated system of equations for the smallest positive solution.

Chinese rings

One of the oldest known **mechanical puzzle**s, the object of which is to remove all *n* rings from a horizontal loop of stiff wire, and/or put them back on the loop. On the first move it is possible to take up to two rings off the left end of the wire. One or both of those can then be slipped through the wire loop (from top to bottom). If both are removed then the fourth ring can be slipped over the end. If just one of the first two is removed, then the next step is to slip the third ring over the end. Subsequently, rings must be put back on to the wire loop in order to remove other rings, and this procedure is repeated over and over again. In general, the minimum number of moves needed is $(2^{n+1} - 2)/3$ if *n* is even and $(2^{n+1} - 1)/3$ if *n* is odd. For example, with seven rings the solution takes 85 moves. Most of the solution is easy, as each move normally involves going forward or back to the previous state. The key to a correct solution is the first step: if *n* is even, you must remove two rings; if *n* is odd, you must remove only one. The solution is similar to that of the **Tower of Hanoi**. In fact, Edouard **Lucas**, who invented the Tower of Hanoi, gave an elegant solution to the Chinese rings that uses **binary** arithmetic.

Stewart Cullin, the noted nineteenth-century ethnologist, relates that the puzzle was invented by the famous Chinese general Chu-ko Liang (A.D. 181–234), in the second century, as a present to his wife so that, in trying to solve it, she would have something to do while he was away at the wars. However, this is anecdotal and its origins remain obscure. The earliest reference to it in Europe may be in about 1500 in the form of Problem 107 of the manuscript *De Viribus Quantitatis* by Luca **Pacioli** in which the description appears: "*Do Cavare et Mettere una Strenghetta Salda in al Quanti Anelli Saldi, Difficil Caso*" (Remove and put a little bar joined in some joined rings,

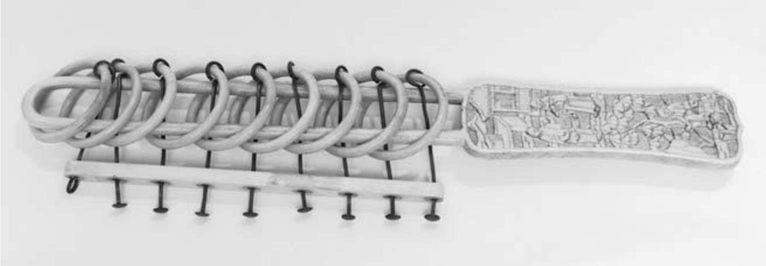

Chinese rings An unusual example of Chinese rings in ivory, dating from the mid-nineteenth century. *Sue & Brian Young/Mr. Puzzle Australia, www.mrpuzzle.com.au* difficult case). It was also mentioned by Girolamo **Cardan** in the 1550 edition of his book *De Subtililate* from which comes the name *Cardan's rings,* and was treated at length in mathematical terms by John **Wallis** in about 1685. By the end of the seventeenth century, it had become popular in many European countries. French peasants used it to lock chests and called it *baguenaudier,* or "time-waster."

Chinese room

An argument first put forward by the American philosopher John Searle (1932–) in 1980 in an attempt to show that the human mind is not a computer and that the **Turing test** is not adequate to prove that a machine can have strong **artificial intelligence** (strong AI)–in other words, can think in a humanlike way.[292] In the Chinese room scenario, a person who understands no Chinese sits in a room into which written Chinese characters are passed. The person uses a complex set of rules, established ahead of time, to manipulate these characters, and pass other characters out of the room. The idea is that a Chinese-speaking interviewer would pass questions written in Chinese into the room, and the corresponding answers would come out of the room in Chinese. Searle maintains that if such a system could indeed pass a Turing test, the person who manipulated the symbols would obviously not understand Chinese any better than he did before entering the room.

Searle proceeds systematically to refute the claims of strong AI by positioning himself as the one who manipulates the Chinese symbols. The first claim is that a system able to pass the Turing test understands the input and output. Searle replies that as the "computer" in the Chinese room, he gains no understanding of Chinese by simply manipulating the symbols according to the formal program (the complex translation rules). The operator in the room need not have any understanding of what the interviewer is asking, or of the replies that he is producing. He may not even know that there is a question-and-answer session going on outside the room.

The second claim of strong AI to which Searle objects is the claim that the system explains human understanding. Searle asserts that since the system is functioning–in this case passing the Turing Test–and yet there is no understanding on the part of the operator, then the system does not understand and therefore could not explain human understanding.

chiral

Having different left-hand and right-hand forms; not mirror symmetric. For example, the *snub cube* (one of the **Archimedean solids**) is chiral, where as the ordinary **cube** is not.

Chladni, Ernst Florens Friedrich (1756–1827)

A German lawyer, musician (he was born in Leipzig in the same year as Mozart and died in the same year as Beethoven), and amateur scientist who founded the science of acoustics. While investigating musical tones, he had the inspired idea of making the sounds visible in a solid material. He spread fine sand over a glass or metal plate and set it into vibration with the bow of a violin by scraping the bow along one edge of the plate. The bow alternately stuck and slipped in rapid succession on the edge of the plate creating waves that moved across the plate and were reflected from the edges. These reflected waves became superimposed on the new waves coming from the bow edge, resulting in symmetrical patterns of nodal lines where the plate wasn't moving. The type of pattern produced on a Chladni plate depends on a variety of factors, including the point or points of support and their location; the point where the bow touches the plate; the frequency of the vibration, which is influenced by the speed the bow; and the shape and other properties of the plate itself.

chord

A straight line that joins two points on a curve. Most commonly, *chord* is used to mean a straight line segment joining, and included between, two points on a **circle**. In this more restricted sense it first appears in English in 1551 in Robert **Recorde**'s *The Pathwaie to Knowledge:* "*Defin.,* If the line goe crosse the circle, and passe beside the centre, then is it called a corde, or a stryngline."

Some surprising results emerge from moving chords. For example, take a chord in a circle C, and slide the chord around the circle so that the midpoint of the chord traces out a smaller concentric circle. Call the area between the two circles $A(C)$. Now do the same thing with a larger circle C' but with the same length chord. Is $A(C')$ larger or smaller than $A(C)$? Surprisingly, they are the same. In other words $A(C)$ doesn't depend on what circle you start with, only the length of the chord. An even more amazing fact is that if you slide a chord of fixed length around *any convex shape C* so that the chord's midpoint traces out another figure D, the area between C and D doesn't depend on what shape you started with.

chromatic number

(1) In **graph theory**, the minimum number of colors needed to color (the vertices of) a **connected graph** so that no two adjacent vertices are colored the same. In the case of simple graphs, this so-called *coloring problem* can be solved by inspection. In general, however, finding the chromatic number of a large graph (and, similarly, an optimal coloring) is an **NP-hard problem**. (2) In **topology**, the maximum number of regions that can be drawn

on a surface in such a way that each region has a border in common with every other region. If each region is given a different color, each color will border on every other color. The chromatic number of a square, tube, or sphere, for example, is 4; in other words, it is impossible to place more than four differently colored regions on one of these figures so that any pair has a common boundary. "Chromatic number" also indicates the least number of colors needed to color any finite map on a given surface. Again, this is 4 in the case of the plane, tube, and sphere, as was proved quite recently in the solution to the **four-color map problem**. The chromatic number, in both senses just described, is 7 for the **torus**, 6 for the **Möbius band**, and 2 for the **Klein bottle**. See also **Betti number**.

chronogram

A phrase or sentence in which certain letters represent, cryptically, a date, epoch, or, in rare cases, a non-date number. For example, the chronogram "My Day Is Closed In Immortality" commemorates the death of Queen Elizabeth the First of England: the capital letters can be rearranged to give MDCIII, or 1603, the year in which she died.

Church, Alonzo (1903–1995)

An American logician and professor at Princeton University who was an early pioneer of theoretical computer science. He is best known for his development, in 1934, of the so-called **lambda calculus**, a model of computation, and his discovery, in 1936, of an "undecidable problem" within it. This result preceded Alan **Turing**'s famous work on the **halting problem**, which also pointed out the existence of a problem unsolvable by mechanical means. Church and Turing then showed that the lambda calculus and the **Turing machine**, which is used in the halting problem, are equivalent in capability. They also demonstrated a variety of alternative "mechanical processes for computation" with equivalent computational abilities. See also **Church-Turing thesis**.

Church-Turing thesis

A logical/mathematical postulate, independently arrived at by Alan **Turing** and Alonzo **Church**, which asserts that as long as a procedure is sufficiently clear-cut and mechanical, there is some **algorithm**ic way of solving it (such as via computation on a **Turing machine**). Thus, there are some processes or problems that are computable according to some set of algorithms, and other processes or problems that are not. A strong form of the Church-Turing thesis claims that all neural and psychological processes can be simulated as computational processes on a computer.

cipher

(1) A cryptographic system (see **cryptography**) in which units of plain text of regular length, usually letters, are arbitrarily transposed (see **transposition cipher**) or substituted (see **substitution cipher**) according to a predetermined code, or a message written or transmitted in such a system. See also **Caesar cipher** and **Beal cipher**. (2) The mathematical symbol (0) for **zero**.

circle

The set of all points in a plane at a given distance, called the *radius,* from a fixed point, called the *center.* A circle is a simple **closed** curve that divides the plane into an interior and exterior. It has a perimeter, called a *circumference,* of length $2\pi r$ and encloses an area of πr^2. In coordinate geometry a circle with center (x_0, y_0) and radius r is the set of all points (x, y) such that:

$$(x - x_0)^2 + (y - y_0)^2 = r^2$$

"Circle" comes from the Latin *circus,* which refers to a large round or rounded oblong enclosure in which the famous Roman chariot races were held.

A line cutting a circle in two places is called a *secant.* The *segment* of a secant bound by the circle is called a *chord,* and the longest chord is that which passes through the center and is known as a *diameter.* The ratio of the circumference to the diameter of a circle is **pi**. The length of a circle between two radii is called an *arc;* the ratio between the length of an arc and the radius defines the *angle* between two radii in *radians.* The area bounded by two radii and an arc is known as a *sector.* A line touching a circle in one place is called a *tangent.* Tangent lines are perpendicular to radii. In **affine geometry** all circles and ellipses become **congruent**, and in **projective geometry** the other **conic section**s join them. A circle is a conic section with eccentricity zero. In **topology** all simple closed curves are **homeomorphic** to circles, and the word *circle* is often applied to them as a result. The three-dimensional analog of the circle is the **sphere,** and the four-dimensional analog is the **hypersphere**.

circle involute

The simplest kind of **spiral** to draw and understand. It is the path that a goat, tethered to a post, would follow if it walked around and around in the same direction, keeping its tether taught until it wound its way to the center. The radial distance between adjacent loops of the spiral is equal to the circumference of the central circle. Except for the innermost loop, the circle involute is hard to distinguish from the **Archimedean spiral**, though the two curves are never identical.

circular cone

A **cone** whose base is a **circle**.

circular helix

See **helix**.

circular prime

A **prime number** that remains prime on any cyclic rotation of its digits. An example (in the decimal system) is 1,193 because 1,931, 9,311, and 3,119 are also prime. Any one-digit prime is circular by default. In base ten, any circular prime with two or more digits can only contain the digits 1, 3, 7, and 9; otherwise when 0, 2, 4, 5, 6, or 8 is rotated into the units place, the result can be divided by 2 or 5. The only circular primes known, listing just the smallest representative from each cycle, are: 2, 3, 5, 7, 11, 13, 17, 37, 79, 113, 197, 199, 337, 1,193, 3,779, 11,939, 19,937, 193,939, 199,933, R_{19}, R_{23}, R_{317}, R_{1031} and possibly R_{49081}. These last five are the known **rep-unit** primes and probable primes. It's generally believed that there are infinitely many rep-unit primes, so there should be infinitely many circular primes. But it's very likely that all circular primes not on the list above are rep-units.

circumcenter

The center of a **circle** that passes through the vertices (see **vertex**) of a given **polygon**, usually a triangle. For a triangle, it is the same as the point of intersection of the perpendicular **bisect**ors of the three sides.

circumcircle

The **circle** that passes through all three vertices (see **vertex**) of a given triangle. It is said to *circumscribe* the triangle.

circumference

The distance around the outside of a **circle**. The word comes the Latin *circus* ("circle") and *ferre* ("to carry"), thus means "to carry around."

cissoid

Given a fixed point A and two curves C and D, the cissoid of the two curves with respect to A is constructed as follows: pick a point P on C, and draw a line l through P and A. This cuts D at Q. Let R be the point on l such that $AP = QR$. The **locus** of R as P moves on C is the cissoid. The name *cissoid,* meaning "ivy-shaped," first appears in the work of Geminus in the first century B.C.

A special case of this curve, now known as the *cissoid of Diocles,* was first explored by **Diocles** in his attempt to solve the classical problem of **duplicating the cube**. Later investigators of the same curve include Pierre de **Fermat**, Christiaan **Huygens**, John **Wallis**, and Isaac **Newton**. The cissoid of Diocles is traced out by the **vertex** of a **parabola** as it rolls, without slipping, on a second parabola of the same size. It has the Cartesian equation

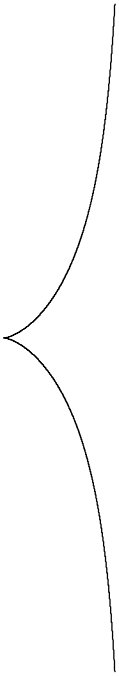

cissoid The cissoid of Diocles. © *Jan Wassenaar, www.2dcurves.com*

$$y^2 = x^3/(2a - x).$$

Interestingly, Diocles investigated the properties of the focal point of a parabola in *On Burning Mirrors* (a similar title appears in the works of Archimedes). The problem, then as now, is to find a mirror surface such that when it is placed facing the Sun, it focuses the maximum amount of heat.

classification

The goal in a branch of mathematics of providing an exhaustive list of some type of mathematical object with no repetitions. For example, the classification of 3-**manifold**s is one of the outstanding problems in **topology**. With the advent of computers, one weak but precise way to state a classification problem is to ask whether there is an **algorithm** to determine whether two given objects are equivalent.

clelia

Also known as a *clelie curve,* the **locus** of a point P that moves on the surface of a **sphere** in such a way that ϕ/θ is constant, where ϕ and θ are the longitude and colatitude (the angular distance from a pole).

Clifford, William Kingdon (1845–1879)

An English mathematician who studied **non-Euclidean geometry** and **topology**. In 1870, he wrote *On the Space Theory of Matter* in which he argued that energy and matter are simply different types of curvature of space–a remarkably advanced idea that would come to fruition in Einstein's general **relativity theory**. Although small of build Clifford was remarkably strong and able to do one-armed chin-ups. His death at an early age was the result of overwork and exhaustion.

clock puzzles

The earliest known clock problem was posed in 1694 by Jacques **Ozanam** in his *Récréations mathématiques et physiques.*

PUZZLES

Here are two clock puzzles invented by Lewis **Carroll**:

1. A clock has hour and minute hands of the same length and no numerals on its face. At what time between 6 and 7 o'clock will the time on the clock appear to be the same as the time read on the reflection of the clock in a mirror?
2. Which has a better chance of giving the right time: a clock that has stopped or one that loses a minute every day?

And here is another from Henry **Dudeney**'s *Amusements in Mathematics* called "The Club Clock":

3. One of the big clocks in the Cogitators' Club was found the other night to have stopped just when the second hand was exactly midway between the other two hands. One of the members proposed to some of his friends that they should tell him the exact time when (if the clock had not stopped) the second hand would next again have been midway between the minute hand and the hour hand. Can you find the correct time that it would happen?

Solutions begin on page 369.

closed

A *closed curve* is one that has no endpoints so that it completely encloses a certain area. A *closed interval,* which corresponds to a *closed set,* is an interval that includes its endpoints.

cochleoid

A **spiral** curve that was first studied by J. Peck in 1700 and Bernoulli in 1726. Its name, meaning "snail-form" (*kochlias* is Greek for "snail"), was coined by Benthan and Falkenburg in 1884. It can be constructed starting from a point O on the y-axis. For all circles through O (tangent to the y-axis), pace a constant distance on the circle. The collection of those points is the cochleoid. In Cartesian coordinates, it is given by the formula

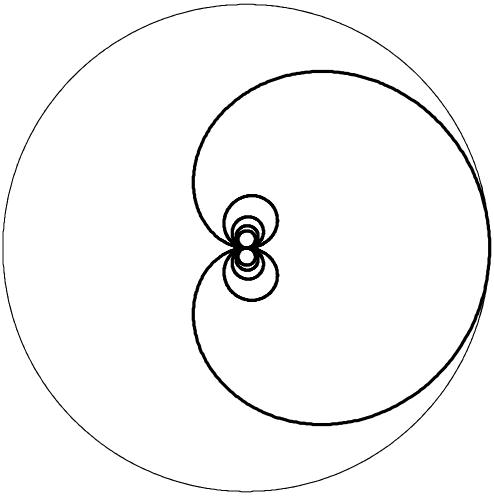

cochleoid A cochleoid inside the circle used to construct it.
© Jan Wassenaar, www.2dcurves.com

$$(x^2 + y^2) \tan^{-1}(y/x) = ay$$

and in polar coordinates by

$$r = a \sin\theta \,/\, \theta.$$

The points of contact of parallel tangents to the cochleoid lie on a **strophoid**.

code

See **cipher**.

codimension

In general, if a mathematical object sits inside or is associated with another object of **dimension** n, then it is said to have codimension k if it has dimension $n - k$.

coding theory

The branch of mathematics concerned with sending data across noisy channels and recovering the message. Whereas **cryptography** is about making messages *hard* to read, coding theory focuses on making messages *easy* to read. The basic problem is that messages, in the form of binary digits or *bits* (strings of 0 or 1) have to be sent along a channel (such as a phone line) in which errors occur randomly, but at a predictable overall rate. To compensate for the errors, more bits have to be sent than are contained in the original message. The easiest way to detect errors in binary data is the **parity** code, which inserts an extra parity bit after every 7 bits from the source message. To correct as well as detect errors, the data has to be retransmitted. A simple way to do this is to repeat each bit a set number of times. The recipient sees which value, 0 or 1, occurs more often and assumes that to be the intended bit. This method can cope with error rates up to 1 error in every 2 bits transmitted but it means that an awful lot of extra bits have to be sent.

In 1948, Claude **Shannon** at Bell Labs began the subject of coding theory by proving the minimum number of extra bits that had to be transmitted to encode messages but without showing ways to find these optimal codes. Two years later, Richard Hamming, also at Bell Labs, gave details of error-correcting codes with information transmission rates more efficient than simple repetition. His first code, in which four data bits were followed by three check bits, allowed not only the detection but the correction of a single error.

While Shannon and Hamming were involved with information transmission in the United States, John Leech devised similar codes while working on **group** theory at Cambridge University. This research also took in the sphere **packing** problem and culminated in the amazing, 24-dimensional *Leech lattice,* the study of which proved crucial to understanding and classifying finite symmetry groups. The value of error-correcting codes for information transmission, both on Earth and from space, was immediately grasped, and a variety of codes were constructed that boosted both economy of transmission and error-correction capacity. Between 1969 and 1973 the NASA Mariner probes used a powerful *Reed-Muller code* capable of correcting 7 errors out of 32 bits transmitted. A less obvious application of error-correcting codes came with the development of the compact disk on which the signal is encoded digitally. To guard against scratches and other damage, two interleaved codes that can correct up to 4,000 consecutive errors are used. By the late 1990s the goal of finding explicit codes that reach the limits predicted by Shannon's original work had been achieved.

codomain

For a given **function** or mapping, a **set** within which the values of the function lie. This is different from the set of values, known as the **range**, that the function actually takes.

coefficient

A number or other factor that multiplies a variable. For example, in the equation $3x - 4ky = 8$, the 3 and $4k$ are coefficients of the variables x and y. The word combines three elements, the Latin *facere* ("to do"), and the prefixes *ex* ("out") and *co* ("with"), to give the overall meaning of joining two things together to bring about a result. The sixteenth-century mathematician Francois Vieta may have coined the word, but it was not commonly used until around the beginning of the eighteenth century.

Coffin, Stewart T.

A leading designer of mechanical puzzles. He is also the author of *The Puzzling World of Polyhedral Dissections,*[64] one of the most significant works produced on this subject.

cohomology

A subject that involves calculating algebraic **invariant**s of **topological space**s that are formally dual to **homology**. The invariants obtained are in general more powerful than those given by homology and usually have more algebraic structure. *Generalized cohomology theories,* both for topological spaces and for purely algebraic structures, have been developed that have some of the formal properties of cohomology but which don't have the same geometric background.

coin paradox

Consider two round coins of equal size. Imagine holding one still and then rolling the other coin around it, making sure that it doesn't slip and that the rims are touching at all times. How many times will the moving coin have rotated after it has completed one revolution of the stationary coin? Most people believe that the answer will be once and are therefore surprised to discover that the truth is in fact *twice.*

coincidence

What an amazing coincidence! Well, not really. Coincidences are bound to happen. In a world where there are a great many potential coincidences each with a small probability of happening, someone, somewhere is going to see one–and be amazed by it. The fact that there are countless numbers of noncoincidences and many people who don't see a significant coincidence in the same period of time is overlooked. Also, we tend to underestimate the probabilities of coincidences in certain situations and are therefore more surprised than we should be when coincidences happen. A classic example of this is the **birthday paradox**.

Obviously some things are extraordinarily unlikely. What are the chances, for example, of a meteorite hitting your car? Next to nothing, but not quite nothing. There are a lot of cars and there are dozens of meteorites that strike Earth every day. Sooner or later, it's bound to happen. In fact, it did happen to Michelle Knapp's Chevy Malibu parked outside her home in Peekskill, New York, on the evening of October 9, 1992. A 12-kilogram space rock smashed through the car's trunk and ended up on the driveway below.

Does coincidence completely explain away all events that might otherwise be put down to precognition? On April 15, 1912, the SS *Titanic* sunk on her maiden voyage, having been holed by an iceberg, and over 1,500 people died. Fourteen years earlier a novel had been published by Morgan Robertson that seemed to foretell the disaster. The book described a ship the same size as *Titanic* that struck an iceberg on its maiden voyage on a misty April night. The name of Robertson's fictional ship was the *Titan.* Mere happenstance or evidence of something deeper? Numerologists often spot matchups that would go unnoticed by the rest of us. Is it so strange that there are almost exactly 500 million inches in the pole-to-pole diameter of Earth? Not if you work in centimeters. And should we make such a fuss over the fact that the speed of light is within 0.1% of 300,000 kilometers per second when we give no attention to the miles-per-second value of 186,282? Yet, surely, there can be no doubt that Shakespeare wrote the Bible. The King James Version was published in 1611, when Shakespeare was 46 years old. Look up Psalm 46. Count 46 words from the beginning of the Psalm. You will find the word "Shake." Count 46 words from the end of the Psalm. You will find the word "Spear." To some, an obvious coded message. See also **thirteen**.

Collatz problem

A problem first posed by the German mathematician Lothar Collatz (1910–1990) in 1937, that is also known variously as the *3n + 1 problem, Kakutani's problem,* the *Syracuse problem, Thwaites' conjecture,* and *Ulam's conjecture.* It runs as follows. Let n be any integer. (1) If n is odd, put n equal to $3n + 1$; otherwise, put n equal to $n/2$. (2) If $n = 1$, stop; otherwise go back to step 1. Does this process always terminate (i.e., end in 1) for any value of n? To date, this question remains unanswered, though the process has been found to stop for all n up to 5.6×10^{13}. British mathematician Bryan Thwaites (1996) has offered a £1,000 reward for a resolution of the problem. However, John **Conway** has shown that Collatz-type problems can be formally undecidable, so it not known if a solution is even possible. The members of sequences produced by the Collatz problem are sometimes known as **hailstone sequences**.[144]

combination

A set of objects selected without reference to the order in which they are arranged. Compare with **permutation**. See also **binomial coefficient**.

combinatorics

The study of the ways of choosing and arranging objects from given collections and the study of other kinds of problems relating to counting the number of ways to do something.

commensurable

Two lines or distances are commensurable if the ratio of their lengths is a **rational number**. If the ratio is an **irrational number**, they are called *incommensurable.*

common fraction

A **fraction** that consists of the **quotient** of two **integer**s.

communication theory

See **information theory**.

commutative

Two numbers, x and z, are said to be *commutative under addition* if

$$x + y = y + x$$

and to be *commutative under multiplication* if

$$x \times y = y \times y.$$

In general, two elements a and b of a **set** S are commutative under the binary operation (an operation that works on two elements at a time) $*$ if

$$a * b = b * b.$$

Compare with **associative** and **distributive**.

complement

That which is needed to complete something. For instance, the complement of a number is what needs to be added to it to make a specified value; the complement of an angle is the angle required to turn it into a right angle. The complement of a **set** is composed of all the elements that are not members of that set.

complete

Describes a **formal system** in which all statements can be proved as being true or false. Most interesting formal systems are not complete, as demonstrated by **Gödel's incompleteness theorem**.

complete graph

A **connected graph** in which exactly one **edge** connects each pair of vertices (see **vertex**). A complete graph with n vertices, denoted K_n, has $n(n - 1)/2$ edges (i.e., the nth **triangular number**), $(n - 1)!$ **Hamilton circuit**s, and a **chromatic number** of n. Every vertex in K_n has **degree** $n - 1$; therefore K_n has an **Euler circuit** if and only if n is odd. In a *weighted complete graph,* each edge has a number called a *weight* attached to it. Each path then has a total weight, which is the sum of the weights of the edges in the path. See also **traveling salesman problem**.

complex adaptive system (CAS)

A nonlinear, interactive, **complex system** with the ability to adapt to a changing environment. CASs evolve by random mutation, **self-organization**, the transformation of their internal models of the environment, and natural selection. Examples include living organisms, the nervous system, the immune system, the economy, corporations, and societies. In a CAS, semiautonomous agents interact according to certain rules of interaction, evolving to maximize some measure like fitness. The agents are diverse in form and capability and they adapt by changing their rules and, hence, behavior, as they gain experience. CASs evolve historically–their experience determines their future trajectory. Their adaptability can either be increased or decreased by the rules shaping their interaction. Moreover, unanticipated, emergent structures can play a determining role in the evolution of such systems, which is why they are highly unpredictable. On the other hand, CASs have the potential of a great deal of creativity that was not programmed into them from the beginning.

complex analysis

The study of **function**s of a complex variable. Often, the most natural proofs for statements in real **analysis** or even **number theory** use techniques from complex analysis. Unlike real functions, which are commonly represented as two-dimensional graphs, complex functions have four-dimensional graphs and may usefully be illustrated by color-coding a three-dimensional graph to suggest four dimensions.

complex number

A **real number** plus a real number times the square root of –1; in other words, a number of the form $z = a + ib$, where a and b are real and $i = \sqrt{-1}$. The term ib is known as an **imaginary number** or the *imaginary part* of the complex number $a + ib$; a is called the real part. The names "complex," "real," and "imaginary," which came about historically, are totally misleading because complex numbers are not particularly complex and imaginary numbers are no less real than real numbers! Another way to represent a complex number is as an **ordered pair** of real numbers (a, b) together with the operations: $(a, b) + (c, d) = (a + c, b + d)$ and $(a, b) \times (c, d) = (ac - bd, bc + ad)$. Alternatively, complex numbers can be shown as points on an **Argand diagram** (a representation of the *complex plane*) in which the horizontal axis is the real number line and the vertical axis represents all possible purely imaginary numbers. Any point that appears on the complex plane off-axis has both real and imaginary parts. On an Argand diagram a complex number can also be shown as a **vector**, or directed line segment (a line of a certain length with an arrow), extending from the origin $(0 + 0i)$ to the number $(a + bi)$. The *absolute value* or *magnitude* of a complex number z, thought of as a point on a plane, is its Euclidean distance from the origin, and is denoted $|z|$; this is always a nonnegative real number. Algebraically, if $z = a + ib$, we can define $|z| = \sqrt{(a^2 + b^2)}$. If the complex number z is written in polar coordinates $z = r\,e^{i\varphi}$, then $|z| = r$.

Complex numbers are a natural extension of real numbers and form what is called an *algebraically closed field*. Because of this, mathematicians sometimes consider the complex numbers to be more "natural" than the real numbers: all **polynomial** equations have solutions among the complex numbers, which is not true for the real numbers. Complex numbers are used in electrical engineering and other branches of physics as a convenient description for periodically varying signals. In an expression $z = re^{i\varphi}$ one may think of r as the amplitude and ϕ as the phase of a sine wave of given frequency. In special and general **relativity theory**, some formulas for the metric on **space-time** become simpler if the time variable is taken to be imaginary.

complex plane

See **Argand diagram**.

complex system

A collection of many simple nonlinear units that operate in parallel and interact locally with each other so as to produce emergent (see **emergence**) behavior.

complexity

A phenomenon that has two distinct and almost opposite meanings. The first, and probably the oldest mathematically, goes back to Andrei **Kolmogorov**'s attempt to give an **algorithm**ic foundation to notions of **random**ness and probability and to Claude **Shannon**'s study of communication channels via his notion of information. In both cases, complexity is synonymous with *disorder* and a lack of structure. The more random a process, the greater its complexity. An ideal gas, for example, with its numerous molecules bouncing around in complete disarray, is complex as far as Kolmogorov and Shannon are concerned. Thus, in this sense, complexity equates to the degree of complication.

The second, and more recent notion of complexity refers instead to how structured, intricate, hierarchical, and sophisticated a natural process is. In particular, it's a property associated with **dynamical system**s in which new, unpredictable behavior arises on scales above the level of the constituent components. The distinction between these two meanings can be revealed by answering a simple question about a system: Is it complex or is it merely complicated? Measures of complexity include **algorithmic complexity**, **fractal dimension**ality; **Lyapunov fractal**s, and **logical depth**.

complexity theory

A part of the theory of computation that has to do with the resources needed to solve a given problem. The most common resources are *time* (how many steps it takes to solve a problem) and *space* (how much memory it takes to solve a problem). Complexity theory differs from **computability theory**, which deals with whether a problem can be solved at all, regardless of the resources required.

composite number

A positive integer that can be factored into smaller positive integers, neither of which is one. If a positive integer is not composite (4, 6, 8, 9, 10, 12, . . .) or one, then it is a **prime number** (2, 3, 5, 7, 11, 13, 17, . . .). As Karl **Gauss** put it in his *Disquisitiones Arithmeticae* (1801): "The problem of distinguishing prime numbers from composite numbers and of resolving the latter into their prime factors is known to be one of the most important and useful in arithmetic." One reason for its importance today is that many secret codes and much of the security of the Internet depends in part on the relative difficulty of factoring large numbers. But more basic to a mathematician is that this problem has always been central to **number theory**. Numbers that, for their size, have a lot of factors are sometimes referred to as *highly composite numbers*. Examples include 12, 24, 36, 48, 60, and 120.

compound polyhedron

An assemblage of two or more polyhedra, usually interpenetrating and having a common center. There are two types: a combination of a solid with its **dual** and an interpenetrating set of several copies of the same **polyhedron**. The simplest example of a compound polyhedron is the compound of two tetrahedra, known as the *stella octangula* and first described by Johannes **Kepler**. This shape is unique in that it falls under both of the above classes, because the **tetrahedron** is the only self-dual **uniform polyhedron**; the edges of the two tetrahedra form the diagonals of the faces of a **cube** in which the stella octangula can be inscribed.

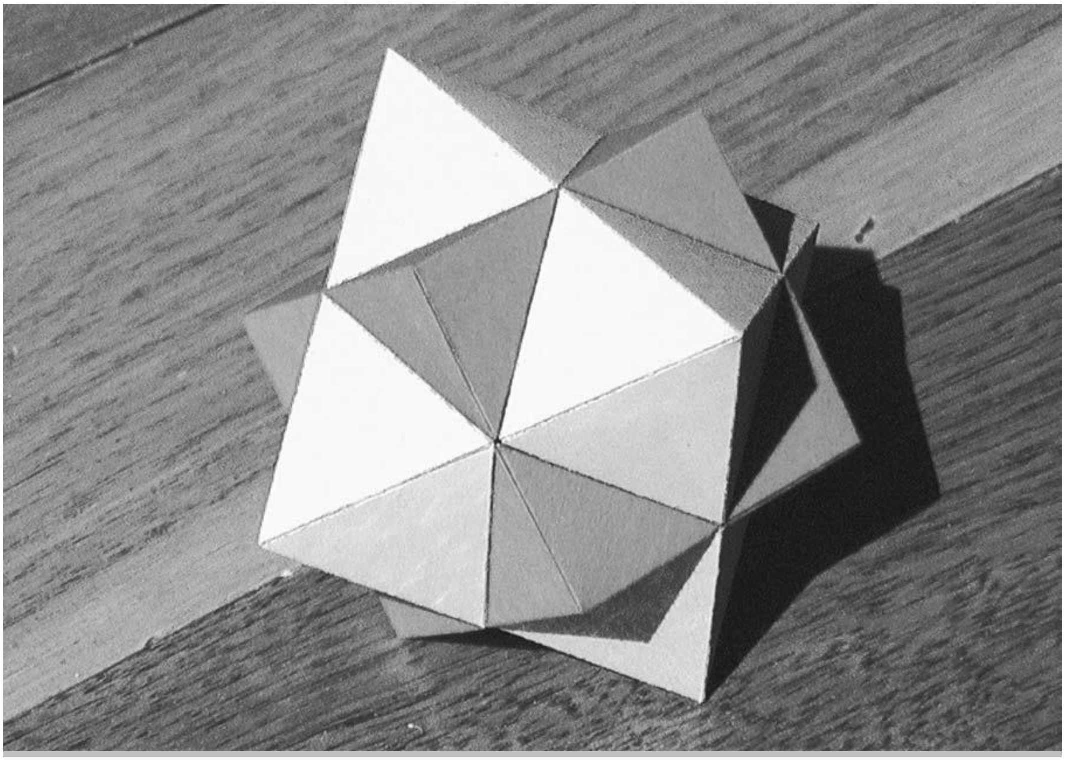

compound polyhedron A compound of duals: the cube and the octahedron. *Robert Webb, www.software3d.com; created using Webb's Stella program*

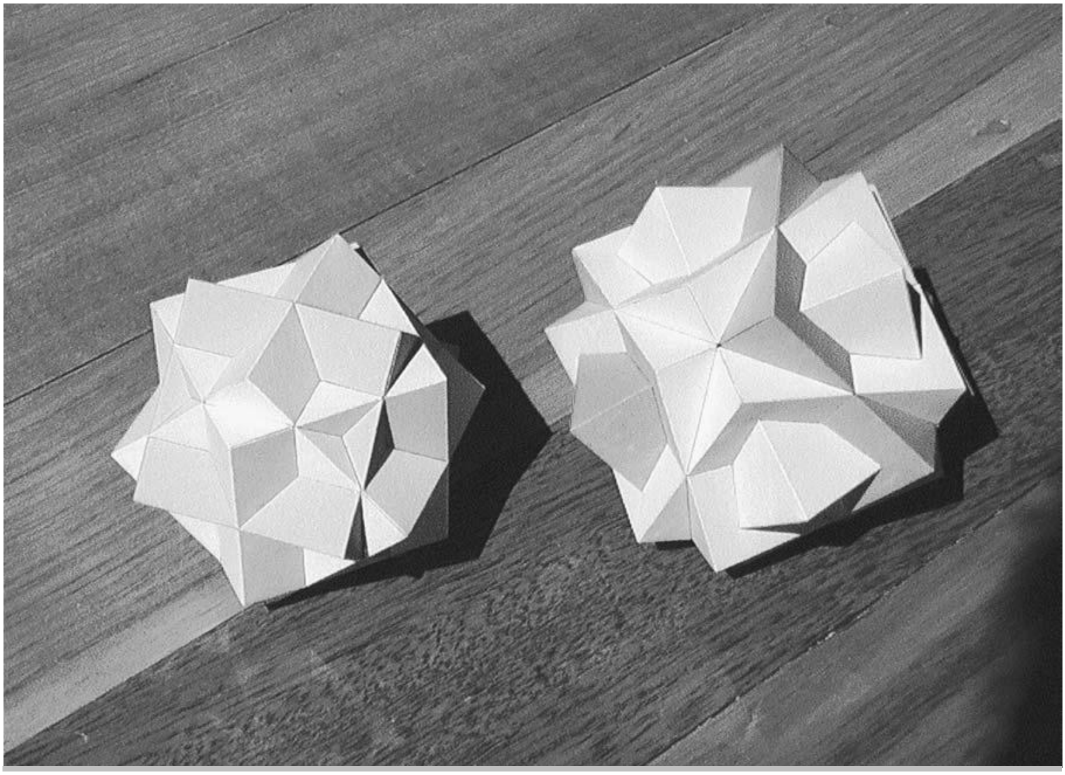

compound polyhedron A compound polyhedron of three cubes (left), such as that used in Escher's picture "Waterfall." The compound of four cubes (right) is also known as Bakos's compound. *Robert Webb, www.software3d.com; created using Webb's Stella program*

Another example of a compound follows from an important Platonic relationship: a cube can be inscribed within a **dodecahedron**. There are five different positions for a cube within a dodecahedron; superimposing all five gives the compound known as the *rhombic triacontahedron*.

compressible

Having a description that is smaller than itself; not **random**; possessing regularity.

computability theory

The part of the theory of computation that deals with problems that are solvable by **algorithms** or–what amounts to the same thing–by **Turing machine**s. Computability theory is concerned with four main questions: What problems can Turing machines solve? What other systems are equivalent to Turing machines? What problems require more powerful machines? What problems can be solved by less powerful machines? Not all problems can be solved computationally. An *undecidable problem* is one that can't be solved by any algorithm, no matter how much time, processing speed, or memory is available. Many examples are known, one of the most famous of which is the **Halting problem**. See also **cellular automaton**.

computable number

A **real number** for which there is an **algorithm** that, given *n,* calculates the *n*th digit. Alan **Turing** was the first to define a computable number and the first to prove that almost all numbers are uncomputable. An example of a number that, even though well-defined, is uncomputable is **Chaitin's constant**.

concave

Curved inward, like the inner surface of a sphere; the word comes from the Latin *concavus* for "hollow." A figure, such as a **polygon** or **polyhedron**, is said to be concave if a line segment joining any points inside the figure goes outside the figure. Similarly, a **set** is concave if it doesn't contain all the line segments connecting any pair of its points.

conchoid

A shell-shaped curve. Given a point A and a curve C, if we pick a point Q on C and draw a line L through A and Q and mark points P and P' on L at some fixed distance in either direction from Q, then the locus of P and P' as Q moves on C is a conchoid. The *conchoid of Nichomedes* is a conchoid in which the given line is a straight line; that is, given a line C and a point A we pick a point Q on C, draw a line L through A and Q, and mark P and P' on L at some fixed distance from Q. The conchoid of Nicomedes is the locus of P and P' as Q moves along C. It has the polar equation $R = a \sec\theta + k$. The *conchoid of de Sluze* is the curve with the Cartesian equation $a(x - a)(x^2 + y^2) = k^2x^2$.

cone

A shape (its name comes from the Greek *konos* for pinecone) that has a circular or elliptical base and a **vertex**, also known as an *apex*, lying outside the plane of the base and that is formed from all the line segments joining points on the edge of the base to the vertex. If the base is a circle, the shape is a *circular cone;* if the line, or *axis*, from the center of the base to the vertex is perpendicular to the base, then it is a *right cone* (an ice-cream cone is a right circular cone); otherwise it's an *oblique cone.* The curved lateral surface of the cone is called a *nappe.* If the cone is extended in both directions from the vertex, the result is a *double cone* or *bicone.* A section through a double cone that has been extended indefinitely in both directions to form a *conic surface* is known as a **conic section**. Another way to think of a cone is as a **surface of revolution** generated by a line that rotates around a fixed point, at a fixed angle from another line (the axis), both lines passing through that fixed point. The volume of a cone, of perpendicular height h and circular base of radius r, is $\frac{1}{3}\pi r^2 h$.

Take a solid cylinder of radius r and height $2r$. Remove the right double cone that passes through the center of the cylinder and extends to meet the circular disks on the cylinder's top and bottom. Interestingly, the volume of the remaining object and the volume of a sphere of radius r are the same.

PUZZLE

The *Cone Puzzle* (no. 202) from Henry **Dudeney**'s *Amusements in Mathematics*[88] runs as follows: "I have a wooden cone. How am I to cut out of it the greatest possible cylinder?"

Solutions begin on page 369.

conformal mapping

A map from the plane to itself that preserves angles. Conformal mapping results in the angle between any two curves being the same as the angle between their images. The Mercator map is a conformal map of Earth's surface.

congruent

In the case of geometric figures, having exactly the same shape and size.

congruum problem

Find a **square** number x^2 such that, when a given number h is added or subtracted, new square numbers are obtained, so that $x^2 + h = a^2$ and $x^2 - h = b^2$. This problem was posed by the mathematicians Théodore and Jean de Palerma in a mathematical tournament organized by Frederick II in Pisa in 1225. The solution is $x = m^2 + n^2$ and $h = 4mn(m^2 - n^2)$, where m and n are integers.

conic section

An important, familiar, and ubiquitous family of curves obtained by slicing a right circular double **cone**, extended indefinitely in both directions, with a plane. Depending on the angle of the slice to the axis of the cone, the resulting curve may be a **circle**, an **ellipse**, a **parabola**, or a **hyperbola**. The circle is a limiting case of the ellipse, when the slice is made at right angles to the axis, while the parabola is the limiting case of both the ellipse and the hyperbola, when the slice is made parallel to the side of the cone. The name *conic sections* comes from the eight-volume work *Conics* (Κωνικα) by **Apollonius**, who also gave us the names *ellipse, parabola,* and *hyperbola.*

Another geometric way to define the conics is as the **locus** of all points in the plane whose distances, r, from a fixed point called the *focus,* and a, from a given straight line called the **directrix**, have a constant ratio. This ratio, r/a, is known as the *eccentricity, e.* The circle has an eccentricity of zero. As the eccentricity increases from near zero, corresponding to a nearly circular ellipse, the ellipse stretches until the right-hand side of it disappears to infinity, e becomes 1, and the ellipse turns into a parabola, with just one open branch. Like the circle, the parabola has only one shape, though it may look different depending on how much it is enlarged or diminished. As the eccentricity increases beyond 1, the "lost" right-hand end of the ellipse reappears from the other side of infinity, so to speak, and turns into the left-hand branch of a hyperbola.

Because a hyperbola is effectively an ellipse split in two by infinity, it comes as no surprise that these curves are related in an inverse way. An ellipse consists of all points whose distances from two foci have a constant sum, while a hyperbola is made from all points whose

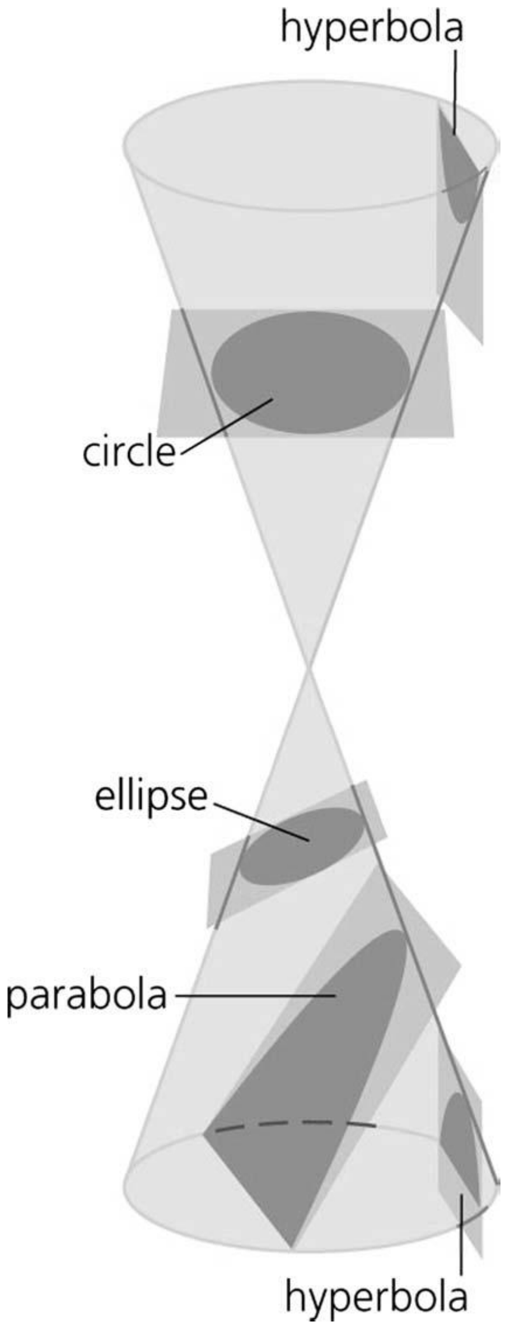

conic section The circle, ellipse, parabola, and hyperbola, obtained by slicing a right double cone in various ways.

distances from two foci have a constant difference. These definitions also apply to the circle and the parabola, if the two foci are considered to coincide in the case of the circle and to be separated by an infinite distance in the case of the parabola.

In terms of algebra, the family of conics represents all the possible **real number** solutions to the general **quadratic** equation $ax^2 + bxy + cy^2 + dx + ey + f = 0$. In other words, the graph of any quadratic with real solutions is always a conic section. The key quantity is the difference $b^2 - 4ac$. If this is less than zero, the graph is an ellipse, a circle, a point, or no curve. If $b^2 - 4ac = 0$, the graph is a parabola, two parallel lines, one line, or no curve; if it is greater than zero, the graph is a hyperbola or two intersecting lines.

conical helix

See **helix**.

conjecture

A mathematical statement that has been put forward as a true statement, but that no one has yet been able to prove or disprove; in mathematics, a conjecture and a hypothesis are essentially the same thing. When a conjecture has been proven to be true, it becomes known as a **theorem**. Famous conjectures include the **Riemann hypothesis**, the **Poincaré conjecture**, the **Goldbach conjecture**, and the **twin primes** conjecture. Just to show how terminology can be used inconsistently, however, the most famous of all conjectures, for centuries before its proof in 1995, was always known as **Fermat's last theorem**!

conjugate

(1) *Conjugate angles* add up to 360°. (2) The *complex conjugate* of a **complex number** $a + bi$ is $a - bi$. (3) *Conjugate lines* of a **conic section** have the property that each contains the pole point of the other, while *conjugate points* of a conic have the property that each lies on the polar line of the other. In general, conjugate indicates that there is a symmetrical relationship between two objects A and B; in other words, there is an operation that will turn A into B and B into A.

connected

A space S is said to be connected if any two points in S can be connected by a curve lying wholly within S. Two spaces can be added by what is called a *connected sum*. Roughly speaking, this involves pulling out a disk from each surface, creating **hole**s, and then sewing the two surfaces together along the boundaries of the holes. In this way, a one-holed **torus** can be added to a two-holed torus to give a three-holed torus; alternatively, a **projective**

plane can be added to a projective plane to give a **Klein bottle**. The operation is **commutative** and **associative** and there is even an **identity** element: for example, adding a sphere to any surface simply returns the same surface. See also **simply connected**.

connected graph

A **graph** in which a path exists between all pairs of vertices (see **vertex**). If the graph is also a **directed graph**, and there exists a path from each vertex to every other vertex, then it is a *strongly connected graph*. If a connected graph is such that exactly one edge connects each pair of vertices, then it is said to be a **complete graph**. See also **Euler path** and **Hamilton path**.

connectionism

A computational approach to modeling the brain that relies on the interconnection of many simple units to produce complex behavior.

connectivity

The amount of interaction in a system, the structure of the weights in a **neural network**, or the relative number of edges in a **graph**.

consistency

An **axiom**atic theory is said to be consistent if it's impossible (within the confines of the theory) to prove simultaneously a statement and its negation. **Godel's incompleteness theorem** states that any (sufficiently powerful) consistent axiomatic theory is incomplete.

constructible

In classical geometry, a figure or length that can be drawn using only an unmarked straightedge and a compass. The Greeks were adept at constructing **polygon**s, but the question of proving which **regular polygons** are constructible and which are not had to wait for the genius of Carl **Gauss**. At the age of only 19, Gauss found that a regular polygon with *n* sides is constructible if and only if *n* is a prime **Fermat number**. The only known such primes are 3, 5, 17, 257, 65,537. It is also possible to construct certain numbers, known as *constructible numbers,* that correspond to line segments, including **rational numbers** and some **irrational numbers**, but no **transcendental numbers**. It turns out that all constructions possible with a compass and straightedge can be done with a compass alone, as long as a line is considered constructed when its two endpoints are located. The reverse is also true, since Jakob **Steiner** showed that all constructions possible with straightedge and compass can be done using only a straightedge, as long as a fixed circle and its center (or two intersecting circles without their centers, or three nonintersecting circles) have been drawn beforehand. Such a construction is known as a *Steiner construction*. The Greeks were unable to achieve certain constructions, such as **squaring the circle**, **duplicating the cube**, and **trisecting an angle**, despite numerous attempts, but it wasn't until hundreds of years later that the problems were proved to be actually impossible under the limitations imposed.

continued fraction

A representation of a **real number** in the form

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ldots}}}$$

which, mercifully for typesetters, can be written in compact notation as

$$x = [a_0; a_1, a_2, a_3, \ldots],$$

where the integers a_i are called *partial quotients*. Although rarely encountered in school and even college math courses, continued fractions (CFs) provide one of the most powerful and revealing forms of numerical expression. Numbers whose decimal expansions look unremarkable turn out, when unfolded as CFs, to have extraordinary symmetries and patterns. CFs also offer a way of constructing rational approximations to **irrational numbers** and of discovering the most irrational numbers.

CFs first appeared in the sixth century in the works of the Indian mathematician Aryabhata, who used them to solve linear equations. They surfaced in Europe in the fifteenth and sixteenth centuries and **Fibonacci** attempted to define them in a general way. The term "continued fraction" first appeared in 1653 in an edition of *Arithmetica Infinitorum* by John **Wallis**. Their properties were also studied by one of Wallis's English contemporaries, William Brouncker, who, along with Wallis, was one of the founders of the Royal Society. At about the same time, in Holland, Christiaan **Huygens** made practical use of CFs in his designs of scientific instruments. Later, in the eighteenth and early nineteenth centuries, Carl **Gauss** and Leonhard **Euler** delved into many of their deeper properties.

CFs can be finite or infinite in length. Finite CFs can be evaluated level by level (starting at the bottom) and will always reduce to a rational fraction; for example, the

CF [1; 3, 2, 4] = 40/31. By contrast, infinitely long CFs produce representations of irrational numbers. Here are the leading terms from a few notable examples of infinite CFs:

$e = [2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, \ldots]$
$\sqrt{2} = [1; 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, \ldots]$
$\sqrt{2} = [1; 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, \ldots]$
$\pi = [3; 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, 1, 14, 2, 1, 1, 2, 2, 2, 2, 1, 84, 2, \ldots]$

Each of these expansions has a simple pattern except that for π (see **pi**), which has no obvious pattern at all. There's also a preference for the quotients to be small numbers.

If an infinite CF is truncated after a finite number of steps, the result is a *rational approximation* to the original irrational. In the case of π, chopping the CF at [3; 7] gives the familiar approximation for π of 22/7 = 3.1428571 . . . . Keeping two more terms leads to [3; 7, 15, 1] = 353/113 = 3.1415929 . . . , which is an even better approximation to the true value of π (3.14159265 . . .). The more terms retained in the CF, the better the rational approximation becomes. In fact, the CF gives yields the best possible rational approximations to a general irrational number. Notice also that if a large number occurs in the expansion of quotients, then truncating the CF after that will produce an especially good rational approximation. Most CF quotients are small numbers (1 or 2), so the appearance in the CF of π of a number as large as 292 so early in the expansion is unusual. It also leads to an extremely good rational approximation to π = [3; 7, 15, 1, 292] = 103,993/33,102.

continuity

A mathematical property that has to do with how smooth or "well-behaved" a **function** or curve is. If two adjacent points on a graph, for example, are not connected or are separated by a jump, this marks a breakdown of continuity. At such a discontinuity it is impossible to obtain a **derivative**, or slope, of the curve. Usually if a curve does misbehave like this, it is only at one or two isolated places; elsewhere the curve is likely to be both continuous and differentiable. However, it is possible to construct a continuous function that has "problem points" everywhere and, therefore, is nowhere differentiable! The first example was found by Karl **Weierstrass** in 1872 and came as a total surprise. It is defined as an infinite series

$$f(x) = \sum_{n=0}^{\infty} B^n \cos(A^n \pi x),$$

where A and B can be any numbers such that B is between 0 and 1, and $A \times B$ is bigger than $1 + (3\pi/2)$.

continuum

Any **set** that can be brought into **one-to-one** correspondence with the set of **real numbers**. Examples include a finite line segment, a square, a circle, and a disk.

continuum hypothesis

In 1874 Georg **Cantor** discovered that there is more than one level of **infinity**. The lowest level is called *countable infinity;* higher levels are known as *uncountable infinities.* The **natural numbers** are an example of a countably infinite set and the **real numbers** are an example of an uncountably infinite set. The continuum hypothesis, put forward by Cantor in 1877, says that the number of real numbers is the *next* level of infinity above countable infinity. It is called the continuum hypothesis (CH) because the real numbers are used to represent a linear continuum. Let c be the cardinality of (i.e., number of points in) a continuum, **aleph**-null (χ_0) be the cardinality of any countably infinite set, and χ_1 be the next level of infinity above χ_0. CH is equivalent to saying that there is no **cardinal number** between χ_0 and c, and that $c = \chi_1$. CH has been, and continues to be, one of the most hotly pursued problems in mathematics.

convergence

A property of some **sequences**. A sequence u_i is said to be convergent if there exists a value u with the property that by choosing a large enough value of i, we can make u_i as close as we wish to u.

convex

Curved outward, like the exterior surface of a sphere; the word comes from the Latin *convexus* for "vaulted." A figure, such as a **polygon** or a **polyhedron**, is said to be convex if every line segment that joins two interior points remains inside the figure. Similarly, a **set** is convex if it contains all the line segments connecting any pair of its points.

Conway, John Horton (1937–)

A British-born (Liverpool) mathematician, who studied and taught at Cambridge University and is now a professor at Princeton University. Conway has been an extraordinarily fertile source of new ideas in mathematics and of mathematical games. His most significant contribution was the discovery of **surreal numbers**, to which he was led after watching the British **Go** champion play at Cambridge. In 1967, he found a cluster of three new sporadic **groups**, now sometimes called *Conway's constellation,* building on an earlier discovery by John Leech of an extremely dense **packing** of unit spheres in a space of 24 dimensions. He has also been active in the field of **knots** and in **coding theory**. Among amateur mathematicians, Conway is best known as the inventor of the games of **Life**, **Sprouts**, and **Phutball**,

Conway, John Horton *Princeton University*

as well as for his detailed analyses of many other games and puzzles, such as the **Soma Cube**.

Conway's chained-arrow notation

One of various methods that have been devised recently for representing extremely **large number**s. Developed by John **Conway**, it is based on **Knuth's up-arrow notation** but is even more powerful. The two systems are related thus:

$$a \to b \to 1 = a \uparrow b$$

$$a \to b \to 2 = a \uparrow\uparrow b$$

$$a \to b \to 3 = a \uparrow\uparrow\uparrow b$$

$$a \to b \to c = a \uparrow\uparrow \ldots \uparrow\uparrow b \text{ (} c \text{ up arrows)}$$

Longer chains are evaluated by the following general rules:

$$a \to \ldots \to b \to c \to 1 = a \to \ldots \to b \to c$$

$$a \to \ldots \to b \to 1 \to d + 1 = a \to \ldots \to b$$

and

$$a \to \ldots \to b \to c + 1 \to d + 1 = a \to \ldots \to b \to (a \to \ldots \to b \to c \to d) \to d$$

It's important to recognize that the Conway arrow isn't an ordinary *dyadic* operator. Where three or more numbers are joined by arrows, the arrows don't act separately but rather the whole chain has to be considered as a unit. The chain might be thought of as a **function** with a variable number of arguments, or as a function whose single argument is an ordered list or **vector**. The **Ackermann function** is equivalent to a three-element chain: $A(m, n) = (2 \to (n + 3) \to (m - 2)) - 3$. It can also be shown that **Graham's number** is bigger than $3 \to 3 \to 64 \to 2$ and smaller than $3 \to 3 \to 65 \to 2$.

coordinate

One of a set of variables that specifies the location of a point in space. If the coordinates are distances measured along perpendicular axes, they are known as **Cartesian coordinates**. See also **polar coordinates**.

coordinate geometry

See **analytical geometry**.

coprime

Two or more numbers are coprime if they have no **factor**s in common other than 1.

cosine

See **trigonometric function**.

countable set

A **set** that is either finite or *countably infinite*. A countably infinite set is one that can be put in **one-to-one** correspondence with the **natural number**s and thus has a **cardinal number** ("size") of **aleph**-null ($\aleph_0$). Examples of countable sets include the set of all people on Earth and the set of all fractions. See also **infinity**.

counterfeit coin problem

Among *n* coins, identical in size, shape, and appearance, one is a counterfeit and has a slightly different weight than the others. Using only a two-pan balance, what is the smallest number of weighings that would guarantee finding the fake coin? The problem of the counterfeit coin (or some other object), especially involving 8, 10, 12, or 13 coins, has cropped up in many guises over the years. Typically, the problem also involves finding whether the counterfeit coin is lighter or heavier than the rest. The answer depends on the specific problem and can involve quite a number of steps.

covariance

The tendency of two **random** variables to move in tandem. This is important in applications such as survey-taking and sociology, as well as in many branches of science, because if two things tend to vary together, there is a good chance they may be causally linked. See also **causality**.

Coxeter, Harold Scott MacDonald (1907–2003)

A British-born, Cambridge-educated mathematician who spent most of his career (from 1936 on) at the University of Toronto and was regarded as the greatest classical geometer of his generation. Always known as "Donald," he is best known for his work on hyperdimensional geometries and regular **polytope**s.

In 1926, at the age of 19, Coxeter discovered a new regular polyhedron, having six hexagonal faces at each vertex. He went on to study the mathematics of kaleidoscopes and, by 1933, had enumerated the *n*-dimensional kaleidoscopes. His algebraic equations expressing how many images of an object may be seen in a kaleidoscope are now known as *Coxeter groups.* His research on icosahedral symmetries played an important role in the discovery by scientists at Rice University, Texas, of the carbon-60 molecule (see **buckyball**), for which they won the 1996 Nobel Prize in Chemistry.

Coxeter was a close friend of the artist M. C. **Escher**, whom he met in 1954, and also of Buckminster Fuller, who used Coxeter's ideas in his architecture. Indeed Coxeter's work was motivated by a strong artistic temperament and a sense of what is beautiful. He had originally intended to be a composer but fascination for symmetry took him toward mathematics and a career about which he said "I am extremely fortunate for being paid for what I would have done anyway."

Several of Coxeter's books are considered classics, including *The Real Projective Plane* (1955), *Introduction to Geometry* (1961),[74] *Regular Polytopes* (1963),[75] *Non-Euclidean Geometry* (1965)[72] and, written jointly with S. L. Greitzer, *Geometry Revisited* (1967). In 1938, he revised and updated Rouse **Ball**'s *Mathematical Recreations and Essays.*[24]

cross

A shape that consists in its most basic form of an upright section and a transverse section. The *Latin cross* has the shape of an irregular **dodecahedron** with a single (vertical) line of symmetry, and can be folded up to make a cube. The *Greek cross* has the shape of a plus sign, has four lines of symmetry, and is used as the emblem of the Red Cross organization. A version of the Greek cross that has flared ends is also known as the *crux immissa* or *cross patée.* A cross of Saint Andrew is an ordinary Greek cross rotated through 45°, and is also called the *crux decussata;* it served as the basis for the multiplication sign. A cross of Saint Anthony takes the form of a capital T. The *Maltese cross* is an irregular dodecahedron whose cross pieces flange out from the center.

crunode

A point where a curve intersects itself so that two branches of the curve have distinct **tangent** lines.

cryptarithm

A number puzzle in which a group of arithmetical operations has some or all of its digits replaced by letters or symbols, and where the original digits must be found. In such a puzzle, each letter or symbol represents a unique digit. The first example appeared in *American Agriculturist* in 1864. Specific types of cryptarithm include the **alphametic**, the **digimetic**, and the **skeletal division**.

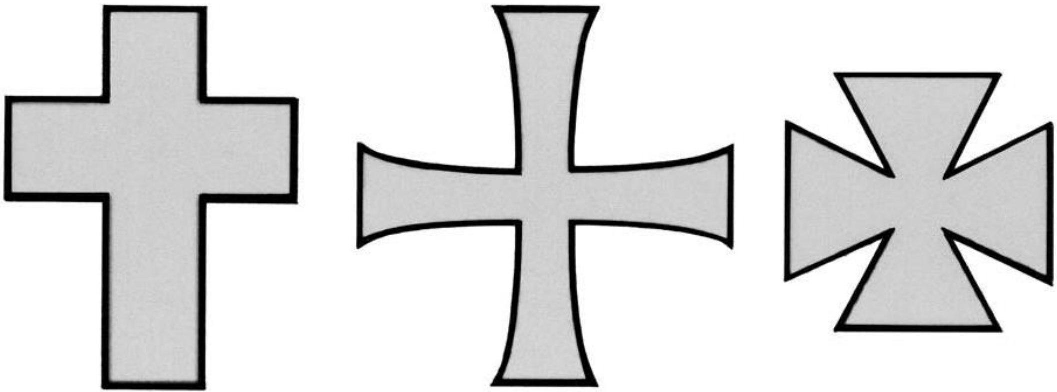

cross From left to right: a Latin cross; a *crux immissa* (a Greek cross with flared ends), also sometimes called a Latin cross; and a Maltese cross.

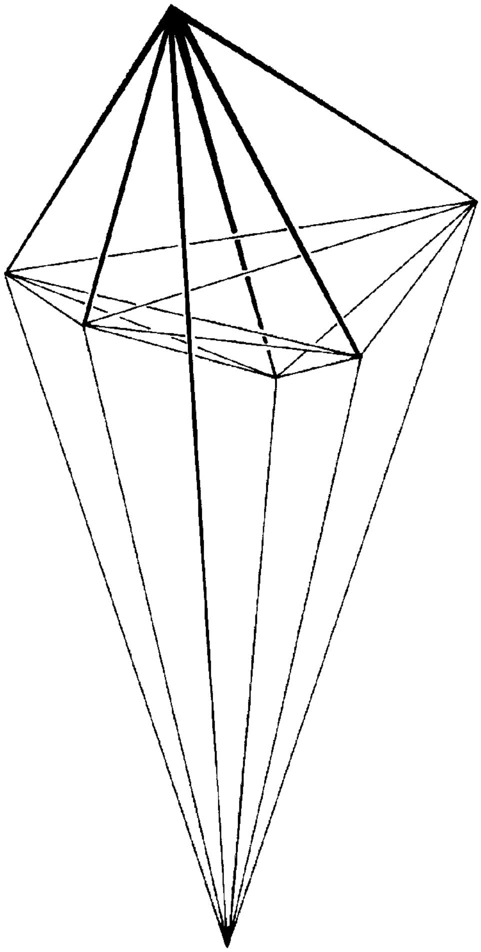

Császár polyhedron

cryptography

The science and mathematics of encoding and decoding information. See also **cipher** and **cryptarithm**.

Császár polyhedron

A **polyhedron**, first described in 1949 by the Hungarian mathematician Ákos Császár,[79] that is a solution to an interesting problem, namely: How many polyhedra exist such that every pair of vertices is joined by an edge? The first clear example is the well known **tetrahedron** (triangular pyramid). Some simple **combinatorics** specify how many vertices, edges, faces, and holes such polyhedra must have. It turns out that, other than the tetrahedron, any such polyhedron must have at least one hole. The first possible polyhedron beyond the tetrahedron has exactly one hole; this is the Császár polyhedron, which is thus topologically equivalent to a **torus** (donut). The Császár polyhedron has 7 vertices, 14 faces, and 21 edges, and is the **dual** of the **Szilassi polyhedron**. It isn't known if there are any other polyhedra in which every pair of vertices is joined by an edge. The next possible figure would have 12 faces, 66 edges, 44 vertices, and 6 holes, but this seems an unlikely configuration–as, indeed, to an even greater extent, does any more complex member of this curious family.

cube

(1) The **Platonic solid** that has a square for every one of its 6 faces; it also has 12 edges and 8 vertices (corners). The 60′ × 30′ × 30′ Double Cube Room of Wilton House (the seat of the Earl of Pembroke), near Salisbury, is considered, together with the Single Cube Room of the same domicile, among the finest surviving rooms in England from the mid-seventeenth century. A favorite with filmmakers, it has provided locations for *Barry Lyndon* by Stanley Kubrick, *The Madness of King George,* and *Sense and Sensibility.* See also **Atomium, the**. (2) To cube something is to raise it to the power of three. The result of cubing is a *cube number:* $1^3 = 1$, $2^3 = 8$, $3^3 = 27$, and so on. To take the *cube root* is the reverse process; thus, 4 cubed (4^3) is 64 and the cube root of 64 ($\sqrt[3]{64}$) is 4. For cube **dissection** problems, see **Hadwiger problem**, **Slothouber-Graatsma puzzle** and **Soma cube**. See also **tesseract** and **Prince Rupert's problem**.

cubic curve

An **algebraic curve** described by a **polynomial** equation of the general form

$$ax^3 + bx^2y + cxy^2 + dy^3 + ex^2 + fxy + gy^2 + hx + iy + j = 0,$$

where $a, b, c, d, e, f, g, h, i,$ and j are constants, such that at least one of $a, b, c,$ and d is nonzero, and x and y are variables. One of Isaac **Newton**'s many accomplishments was the classification of the cubic curves. Newton found 72 different species of curve; later investigators found six more, and it is now known that there are precisely 78 different types of cubic curves. Interesting examples include the **folium** of Decartes and the Witch of **Agnesi**.

cubic equation

A **polynomial** equation of the third degree, the general form of which is

$$ax^3 + bx^2 + cx + d = 0,$$

where $a, b, c,$ and d are constants. There was a great controversy in sixteenth-century Italy between Girolamo **Cardano** and Niccoló **Tartaglia** about who should get credit for solving the cubic. At this time symbolic algebra hadn't been developed, so all the equations were written in words instead of symbols. Early studies of cubics helped legitimize **negative numbers**, give a deeper insight into equations in general, and stimulate work that eventually led to the discovery and acceptance of **complex numbers**. Cardano, in his *Ars Magna,* found negative solutions to equations, but called them "fictitious." He also noted an important fact connecting solutions of a cubic equation to its **coefficients**, namely, that the sum of the solutions is the negation of b, the coefficient of the x^2 term. At one other point, he mentions that the problem of dividing 10 into two parts so that their product is 40 would have to be $5 + v(-15)$ and $5 - v(-15)$. Cardano didn't go further than this observation of what later came to be called complex numbers, but a few years later Rafael Bombelli (1526–1672) gave several examples that involved these strange new mathematical beasts.

cubit

A measure of length used in the ancient world. It is approximately equal to the length of a person's forearm, that is, the part of the arm from the elbow to the fingers. The Romans used a cubit equal to 17.4 modern inches; the Egyptians used one of 20.64 inches.

cuboctahedron

A **polygon** obtained by cutting the corners off a cube or an **octahedron**. It has eight faces that are equilateral triangles and six faces that are squares.

cuboid

Also called a *rectangular prism,* a **hexahedron** of which all of the faces are rectangles and all of the opposite faces are identical. It is not known whether a *perfect cuboid,* whose sides, face diagonals, and space diagonals are all integers, exists. The general suspicion is that it doesn't, although several near misses have been found, including one in which $a = 240$, $b = 117$, $c = 44$, $dab = 267$, $dac = 244$, and $dbc = 125$. If there is a perfect cuboid, it has been shown that the smallest side must be at least $2^{32} = 4{,}294{,}967{,}296$.

Cullen number

A number of the form $(n \times 2^n) + 1$, denoted C_n, and named after the Reverend James Cullen (1867–1933), an Irish Jesuit priest and schoolmaster. Cullen noticed that the first, $C_1 = 3$, was a **prime number**, but with the possible exception of the fifty-third, the next 99 were all composite. Soon afterward, Cunningham discovered that 5,591 divides C_{53}, and noted that all the Cullen numbers are **composite numbers** for n in the range $2 \leq n \leq 200$, with the possible exception of 141. Five decades later Robinson showed that C_{141} is a prime. Currently, the only known Cullen primes are those with n = 1, 141, 4,713, 5,795, 6,611, 18,496, 32,292, 32,469, 59,656, 90,825, 262,419, 361,275, and 481,899. Although the vast majority of Cullen numbers are composite, it has been conjectured that there are infinitely many Cullen primes. Whether n and C_n can simultaneously be prime isn't known. Sometimes, the name "Cullen number" is extended to include the Woodall numbers, $W_n = (n \times 2^n) - 1$. Finally, a few authors have defined a number of the form $(n \times b^n) + 1$, with $n + 2 > b$, to be a *generalized Cullen number.*

Cunningham chain

A sequence of **prime numbers** in which each member is twice the previous one plus one. For example, {2, 5, 11, 23, 47} is the first Cunningham chain of length 5 and {89, 179, 359, 719, 1,439, 2,879} is the first of length 6. In general, a *Cunningham chain of length k of the first kind* is a sequence of k prime numbers, each of which is twice the preceding one plus one. A *Cunningham chain of length k of the second kind* is a sequence of k primes, each of which is twice the preceding one minus one. For example, {2, 3, 5} is a Cunningham chain of length 3 of the second kind and {1,531, 3,061, 6,121, 12,241, 24,481} is a Cunningham chain of length 5 of the second kind. Prime chains of both these forms are said to be complete if they can't be extended by adding either the next larger or the next smaller terms. See also **Sophie Germain prime**.

cup

The symbol ∪, which is used to denote the **union** of two **sets**.

curvature

A measure of the amount by which a curve, a surface, or any other **manifold** deviates from a straight line, a plane, or a hyperplane (the multidimensional equivalent of a plane). For a plane curve, the curvature at a given point has a magnitude equal to one over the radius of an *osculating* circle (a circle that "kisses," or just touches, the curve at the given point) and is a **vector** pointing in the direction of that circle's center. The smaller the radius r of the osculating circle, the greater the magnitude of the curvature ($1/r$) will be. A straight line has zero curvature everywhere; a circle of radius r has a curvature of magnitude $1/r$ everywhere.

For a two-dimensional surface, there are two kinds of curvature: a *Gaussian* (or *scalar*) *curvature* and a *mean curvature.* To compute these at a given point, consider the intersection of the surface with a plane containing a fixed normal vector (an arrow sticking out perpendicularly) at the point. This intersection is a plane and has a curvature; if the plane is varied, this curvature also changes, and there are two extreme values–the maximal and the minimal curvature–which are known as the *main curva-*

tures, $1/R_1$ and $1/R_2$. (By convention, a curvature is taken to be positive if its vector points in the same direction as the surface's chosen normal, otherwise it is negative.) The Gaussian curvature is equal to the product $1/R_1R_2$. It is everywhere positive for a **sphere**, everywhere negative for a **hyperboloid** and **pseudosphere**, and everywhere zero for a plane. It determines whether a surface has elliptic (when it is positive) or hyperbolic (when it is negative) geometry at a point. The integral of the Gaussian curvature over the whole surface is closely related to the surface's **Euler characteristic**. The mean curvature is equal to the sum of the main curvatures, $1/R_1 + 1/R_2$.

A **minimal surface**, like that of a soap film, has a mean curvature of zero. In the case of higher-dimensional manifolds, curvature is defined in terms of a curvature **tensor**, which describes what happens to a vector that is transported around a small loop of the manifold.

curve

A continuous mapping from a one-dimensional space to an *n*-dimensional space. The most familiar mathematical curves are two- and three-dimensional **graph**s. A curve, such as a circle, that lies entirely in a plane is called a *plane curve;* by contrast, a curve that may pass through any region of three-dimensional space is called a *space curve.* See also **space-filling curve**.

curve of constant width

A curve that, when rotated in a square, makes continuous contact with all four sides. It may seem, at first sight, as if there is only one such curve–a circle. But, in fact, there are infinitely many different curves of constant width. The circle is the one with the largest area. The simplest noncircular one, and the one with the smallest area, is the **Reuleaux triangle**. Others can be constructed starting with equilateral (but not necessarily equiangular) stars. Every curve of constant width is **convex**. Moreover, *Barbier's theorem* states that every curve of constant width w has the same perimeter, πw. (The width of a convex figure is defined as the distance between parallel lines–known as *supporting lines*–that bound it.) A curve of constant width can be used in a special drill chuck to cut square holes. A generalization gives solids of constant width. These do not have the same surface area for a given width, but their shadows are curves of constant width with the *same* width.

cusp

In mathematics, a point on a curve where two branches, coming from different directions, meet and have a common **tangent**. If the two branches of the curve approach the tangent from opposite sides the cusp is called a *keratoid* (from the Greek *kera* for "horn") or *first-order cusp.* This is the case, for example, with the curve given by the equation $y^2 = x^2y + x^5$. If the two branches of the curve approach the tangent from the same side the result is a *ramphoid* or *second-order cusp.* "Cusp" derives from the Latin *cuspis* for "sharp." Outside of mathematics, the points of a crescent moon are called cusps and the sharp pointed premolar teeth of children are known as bicuspids.

cute number

A number n such that a square can be cut into n squares of, at most, two different sizes. For example, 4 and 10 are cute numbers.

Cutler, William (Bill)

An Australian puzzle maker and solver who, in 1977, became the first to completely analyze, using a computer, six-piece burrs used to make solid six-piece **burr puzzle**s. Martin Gardner devoted his January 1978 "Mathematical Games" column in *Scientific American* to this and other of Cutler's discoveries. In 2003, Cutler used a computer to enumerate all solutions of the **Loculus of Archimedes**.

cybernetics

The theoretical study of communication and control processes in biological, mechanical, and electronic systems, especially the comparison of these processes in biological and artificial systems. It was pioneered by Norbert **Wiener**.

cyclic number

A number with n digits, which, when multiplied by 1, 2, 3, . . . , n produces the same digits in a different order. For example, 142,857 is a cyclic number: 142,857 × 2 = 285,714; 142,857 × 3 = 428,571; 142,857 × 4 = 571,428; 142,857 × 5 = 714,285; 142,857 × 6 = 857,142, and so on. It has been conjectured, but not yet proven, that an infinite number of cyclic numbers exist.

cyclic polygon

A **polygon** with vertices (see **vertex**) that all lie on the same circle. All triangles are cyclic (but not all of any other kind of polygon) because any set of three points, not lying on a single line, can have a circle drawn through it.

cycloid

The shape defined by a fixed point on a wheel as it rolls; more precisely, it is the **locus** of a point on the rim of a circle rolling along a perfectly straight line. The cycloid was named by Galileo in 1599. It is the solution to both the **tautochrone problem** and the **brachistochrone problem**. In 1634, the French mathematician Gilles de Roberval (1610–1675) showed that the area under a cycloid is three times the area of its generating circle. In 1658, the English architect Christopher Wren showed that the length of a cycloid is four times the diameter of its generating circle. But there was a lot of bickering and a lack of public sharing of information around this time that led to much duplication of effort, particularly over questions related to the cycloid. In fact, the confusion was so bad that the curve was nicknamed the Helen of Geometers, and Jean Montucla referred to it as "*la pomme de discorde*" (the apple of discord).

As well as the ordinary cycloid there is the *curtate cycloid,* which is the path traced out by a point on the inside of a rolling circle, and the *prolate cycloid,* which is followed by a point on the outside of the circle. A prolate cycloid is traced out, for example, by points on the flange of the wheels of a locomotive, which extends below the top of the tracks. This leads to the surprising conclusion that even as the locomotive is moving forward there are always parts of its wheels that are going backward for a moment before moving forward again. See also **epicycloid** and **hypocycloid**.

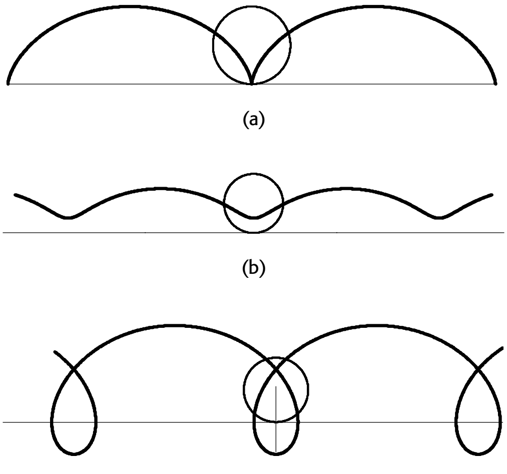

cycloid An ordinary cycloid is traced out by a point on a wheel as it rolls along a flat surface (a). A curtate cycloid is traveled by a point on the wheel that is inside the circumference (b). If the point lies outside the circumference of the wheel, the result is a prolate cycloid (c).

cylinder

A three-dimensional surface described by the Cartesian equation $(x/a)^2 + (y/b)^2 = 1$. If $a = b$ then the surface is a *circular cylinder,* otherwise it is an *elliptic cylinder.* The cylinder is a *degenerate quadric* because at least one of the coordinates (in this case z) doesn't appear in the equation, though by some definitions the cylinder isn't considered to be a quadric at all. In common usage, a *cylinder* is taken to mean a finite section of a right circular cylinder with its ends closed to form two circular surfaces. If the cylinder has a radius r and a length h, then its volume is $V = \pi r^2 h$ and its surface area is $A = 2\pi r^2 + 2\pi rh$. For a given volume, the cylinder with the smallest surface area has $h = 2r$. For a given surface area, the cylinder with the largest volume has $h = 2r$. More unusual types of cylinder include the *imaginary elliptic cylinder:* $(x/a)^2 + (y/b)^2 = -1$, the *hyperbolic cylinder:* $(x/a)^2 - (y/b)^2 = 1$, and the *parabolic cylinder:* $x^2 + 2y = 0$.

D

d'Alembert, Jean Le Rond (1717–1783)

A French mathematician named for the church of St. Jean Baptiste de Rond upon whose steps he was abandoned as a baby, the illegitimate son of a Parisian society hostess. He clarified the concept of a **limit** in calculus, discovered the *Cauchy-Riemann equations* decades before Augustin **Cauchy** or Bernhard **Riemann**, was the first to find and solve the wave equation, and recast **Newton**'s third law in a new and powerful form through what has become known as *d'Alembert's principle.*

Dandelin spheres

If a **cone** is sliced through by a plane, the two spheres that just fit inside the cone, one on each side of the plane and both tangent to it and touching the cone, are known as Dandelin spheres. They are named after the Belgian mathematician and military engineer Germinal Pierre Dandelin (1794–1847) who gave an elegant proof that the two spheres touch the **conic section** at its foci. In 1826, Dandelin showed that the same result applies to the plane sections of a **hyperboloid** of revolution.

dart

Also known as an *arrowhead,* a special kind of quadrilateral that has one **reflex angle**. See also **Penrose tiling**.

de L'Hôpital, Guillaume François Antoine, Marquis de (1661–1704)

A French mathematician who wrote the first textbook on differential calculus, *Analyse des infiniment petits pour l'intelligence des lignes courbes* (1696). This contains the rule, now known as *L'Hôpital's rule,* for finding the limit of a rational function whose numerator and denominator tend to zero at a point. Along with Isaac **Newton**, Gottfried **Leibniz**, and Jacob Bernoulli (see **Bernoulli family**), de L'Hôpital was among the first to solve the **brachistochrone problem**.

de L'Hôpital's cubic

See **Tschirnhaus's cubic**.

de Malves's theorem

Given a **tetrahedron** in which the edges meeting at one **vertex**, *X,* form three right angles (i.e., the tetrahedron is the result of chopping off the corner of a **cuboid**), the square of the face opposite X is equal to the sum of the squares of the other three faces.

de Méré's problem

A question posed in the mid-seventeenth century to Blaise **Pascal** by a French nobleman and inveterate gambler, the Chevalier de Méré, which marked the birth of **probability theory**. One of de Méré's favorite bets was that at least one six would appear during a total of four rolls of a die. From past experience, he knew that this gamble paid off more often than not. Then, for a change, he started betting that he would get a double-six on 24 rolls of two dice. However, he soon realized that his old approach to the game was more profitable. He asked his friend Pascal why. Pascal showed that the probability of getting at least one six in four rolls of a die is $1 - (5/6)^4 \approx 0.5177$, which is slightly higher than the probability of at least one double-six in 24 throws of two dice, $1 - (35/36)^{24} \approx 0.4914$. This problem and others posed by de Méré are thought to have been the original inspiration for a fruitful exchange of letters on probability between Pascal and Pierre de **Fermat**. To tackle these problems, Fermat used combinatorial analysis (finding the number of possible outcomes in ideal games of chance by computing permutation and combination numbers), while Pascal reasoned by recursion (an iterative process that determines the result of the next case by the present case). Their combined work laid the foundations for probability theory as we know it today.

de Moivre, Abraham (1667–1754)

A French-British mathematician who founded analytical trigonometry and stated what has become known as **de Moivre's theorem**. He also worked on probability theory and the normal distribution, and was a good friend of Isaac **Newton**. In 1698 he wrote that the theorem had been known to Newton as early as 1676.

de Moivre's theorem

A theorem, named after Abraham **de Moivre**, that links **complex numbers** and trigonometry. It states that for any real number x and any integer n,

$$(\cos x + i\sin x)^n = \cos(nx) + i\sin(nx).$$

By expanding the left-hand side and then comparing real and imaginary parts, it is possible to derive useful expressions for cos(nx) and sin(nx) in terms of sin(x) and cos(x). Furthermore, the formula can be used to find explicit expressions for the nth **root of unity**: complex numbers z such that $z^n = 1$. It can be derived from (but historically preceded) **Euler's formula** $e^{ix} = \cos x + i \sin x$ and the exponential law $(e^{ix})^n = e^{inx}$.

de Morgan, Augustus (1806–1871)

A British mathematician, born in India, who was an important innovator in the field of mathematical logic. The system he devised to express such notions as the contradictory, the converse, and the transitivity of a relation, as well as the union of two relations, laid some of the groundwork for his friend George **Boole**. De Morgan lost the sight of his right eye shortly after birth, entered Trinity College, Cambridge, at the age of 16, and received his B.A. However, he objected to a theological test required for the M.A. and returned to London to study for the bar. In 1827, he applied for the chair of mathematics in the newly founded University College, London and, despite having no mathematical publications, he was appointed. In 1831, he resigned on principle (after another professor was fired without explanation) but regained his job five years later when his replacement died in an accident. He resigned again in 1861.

His most important published work, *Formal Logic,* included the concept of the *quantification of the predicate,* an idea that solved problems that were impossible under the classic Aristotelian logic. De Morgan coined the phrase "**universe of discourse**," was the first person to define and name mathematical **induction**, and developed a set of rules to determine the **convergence** of a mathematical series. In addition, he devised a decimal coinage system, an almanac of all full moons from 2000 B.C. to A.D. 2000, and a theory on the probability of life events that is still used by insurance companies. De Morgan was also deeply interested in the history of mathematics. In *Arithmetical Books* (1847) he describes the work of over fifteen hundred mathematicians and discusses subjects such as the history of the length of a foot, while in *A Budget of Paradoxes* he gives a marvelous compendium of eccentric mathematics including the poem

> Great fleas have little fleas upon their backs to bite 'em,
> And little fleas have lesser fleas, and so ad infinitum,
> And the great fleas themselves, in turn, have greater fleas to go on,
> While these again have greater still, and greater still, and so on.

The first lines of this poem paraphrase a similar rhyme by Jonathan Swift.

PUZZLE

On one occasion, when asked his age, de Morgan replied: "I was x years old in the year x^2." How old must he have been at the time?

Solutions begin on page 369.

decagon

A **polygon** with 10 sides.

decimal

The commonly used **number system**, also known as *denary,* in which each place has a value 10 times the value of the place at its right. For example, 4,327 in the decimal (base 10) system is shorthand for $(4 \times 10^3) + (3 \times 10^2) + (2 \times 10^1) + (7 \times 10^0)$, where $10^0 = 1$. "Decimal" comes from the Latin *decimus* for "tenth." The verb *decimare,* literally "to take a tenth of," was used to describe a form of punishment applied to mutinous units in the Roman army. The men were lined up and every tenth soldier was killed as a lesson to the rest. From this custom comes our word *decimate,* which we use more loosely–in fact, incorrectly–to indicate near-total destruction. The Latin *decimare* was also used in a less ferocious sense to mean "to tax to the amount of one tenth." However, the usual word describing a one-tenth tax in English is *tithe,* which comes from the Old English *teogotha,* a form of *tenth.*

decimal fraction

A number consisting of an **integer** part, which may be zero, and a decimal part less than unity that follows the decimal marker (which may be a point or a comma). A *finite* or *terminating decimal fraction* has a sequence of decimals with a definite break-off point after which all the places are zeros. Other fractions produce endless sequences of decimals that are *periodic nonterminating.*

Dedekind, (Julius Wilhelm) Richard (1831–1916)

A German mathematician whose most important contribution was the discovery of what became known as the *Dedekind cut.* He realized that every **real number** *r* divides the **rational number**s into two subsets: those greater than *r* and those less than *r.* Dedekind's brilliant idea was to represent the real numbers by such divisions of the rationals. He also provided important support for Georg **Cantor**'s set theory, which was highly controversial at the time.

Dee, John (1527–1609)

A notable English alchemist, mathematician, and astronomer, sometimes referred to as the "last magician" because of his astrological services to Queen Elizabeth I; Dee may also have influenced the writings of Shakespeare. He enrolled at St. Johns College, Cambridge, at age 15, but found the atmosphere there stifling and later went to the Continent to study and lecture. Upon his return to England, Dee cast the horoscope for Queen Mary and later visited Mary's half-sister Elizabeth in jail to determine when Mary would die. Accused of black magic, he was jailed and then released in 1555, three years before Mary's death. When Elizabeth came to the throne she consulted Dee on many matters, including the geography of newly discovered lands, and paid him well. Some of his income he spent on extensive traveling, which may have involved some spying on behalf of his sponsor.

Dee had a large library of books on witchcraft, the occult, and magic, and he wrote 79 manuscripts, only a few of which were published. He married three times and fathered eight children. He also struck up an uneasy partnership with Edward Kelly, a bad-tempered Irishman who claimed to have discovered the alchemical secret of transmuting base metal into gold but had lost his ears for forgery. In 1585 Dee and Kelly went on a four-year trek across the Continent conducting astrological readings for nobility and royalty. But Dee and Kelly had many arguments and eventually parted company. Back in England Dee found his house ransacked and many of his possessions stolen or destroyed. Elizabeth helped pay for the damage and made him warden of Christ's College in Manchester in 1595. However, Elizabeth died in 1603 and her successor James I opposed magic. Dee was forced to retire, his life ending in poverty.

deficient number

See **abundant number**.

degree

(1) The unit of measurement for angles; one degree is 1/360 of a circle. (2) The **exponent** of a variable. For example, the degree of $7x^5$ is 5. See also **degree of freedom**.

degree of freedom

A positive integer that gives the number of pieces of data that are independent.

deletable prime

See **truncatable prime**.

delta curve

A curve that can be turned inside an equilateral **triangle** while continuously making contact with all three sides. There are an infinite number of delta curves, but the simplest are the **circle** and lens-shaped delta-biangle. All the delta curves of height h have the same perimeter $2\pi h/3$. See also **Reuleaux triangle** and **rotor**.

deltahedron

A **polyhedron** whose faces consist of equilateral triangles that are all the same size. Although there are an infinite number of different deltahedra, only eight of them are convex, as O. Rausenberger first showed in 1915. Among this group of eight, faces made of coplanar equilateral triangles sharing an edge (such as the rhombic dodecahedron) aren't allowed. The eight convex deltahedra have 4, 6, 8, 10, 12, 14, 16, and 20 faces.

deltoid

A **hypocycloid** with three cusps, also known as a *tricuspoid* or *Steiner's hypocycloid* after the Swiss mathematician Jakob **Steiner** who investigated the curve in 1856. The deltoid, so-named because it looks like an uppercase Greek delta, Δ, is formed by a point on the circumference of a circle rolling inside another circle with a radius three times as large. While working on a problem in optics in 1745, Leonhard **Euler** was among the first to study its properties. The parametric equations of the cycloid with inner circle of radius r are:

$$x(t) = 2r\cos t + r\cos 2t$$
$$y(t) = 2r\sin t - r\sin 2t$$

The length of the path of the deltoid is $16r/3$, and the area inside the deltoid is $2\pi r^2$. If a tangent is drawn to the deltoid at some point, P, and the points where the tangent crosses the deltoids other two branches are called points A and B, then the length of AB equals $4r$. If the deltoid's tangents are drawn at points A and B, they will

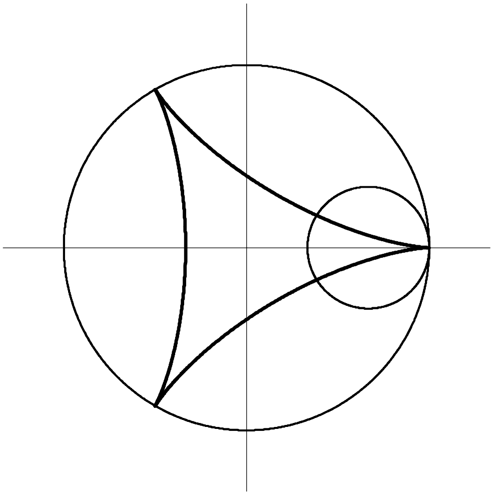

deltoid The deltoid curve. © *Jan Wassenaar, www.2dcurves.com* be perpendicular, and they will intersect at a point inside the deltoid that is the 180° rotation of point P about the center of the fixed circle.

denominator

In a **rational number**, the number below the fraction bar; it indicates into how many parts the whole is divided.

derivative

The result of differentiating a **function**; that is, the infinitesimal change in a function caused by an infinitesimal change in the variable(s) upon which it depends. The derivative gives the rate of change of a function (the slope of its curve) at a particular point. Second and third derivatives give the rate at which the rate of change is changing and the rate at which the rate of rate of change is changing, respectively. In an article in 1996, Hugo Rossi wrote: "In the fall of 1972 President Nixon announced that the rate of increase of inflation was decreasing. This was the first time a sitting president used the third derivative to advance his case for reelection."[269]

Here is a fallacious "proof" that $x = 2x$ based on derivatives. Consider the function $f(x) = x^2$, the derivative of which is $2x$. What is wrong with the following?

$$x^2 = x + x + \ldots + x \text{ (repeated } x \text{ times)}$$

Taking the derivative of both sides gives

$$(x^2)' = 1 + 1 + \ldots + 1 = x.$$

But we have already said that the derivative of x^2 is $2x$. Therefore, $x = 2x$. The error stems from taking the derivative of x different x's. Each of the terms depends not only on x, which was accounted for in taking the derivative, but also on the *number* of terms (which could be fractional) which depends on x, too, and this was not accounted for. Put another way, the derivative measures the rate of change of x^2 as x changes, but as x changes, the number of terms on the right, as well as the terms themselves, increases. For positive x, the correct answer must be larger than x–as indeed it is.

Desargues, Girard (1591–1661)

A French mathematician who is regarded as the chief founder of perspective geometry. His 12-page treatise *La perspective* (1636) consists of a single worked example in which Desargues sets out a method for constructing a perspective image without using any point lying outside the picture field. He considers the representation in the picture of a plane of lines that meet at a point and also of lines that are parallel to each another. In the last paragraph of the work he considers the problem of finding the perspective image of a conic section. Three years later, he wrote his treatise on projective geometry *Brouillon project d'une atteinte aux evenemens des rencontres du cone avec un plan* (Rough draft for an essay on the results of taking plane sections of a cone). The first part of this deals with the properties of sets of straight lines meeting at a point and of ranges of points lying on a straight line. In the second part, the properties of conics are investigated in terms of properties of ranges of points on straight lines and the modern term "point at infinity" appears for the first time. Desargues shows that he has completely grasped the connection between conics and perspective; in fact he treats the fact that any conic can be projected into any other conic as obvious. Given such innovative work it may seem surprising that the subject didn't develop rapidly in the following years. That may be partly due to mathematicians failing to recognize the power of what had been put forward. On the other hand, the algebraic approach to geometry put forward by René **Descartes** at almost exactly the same time (1637) may have diverted attention from Desargues's projective methods.

Descartes, René (1596–1650)

> *If you would be a real seeker after truth, it is necessary that at least once in your life you doubt, as far as possible, all things.*

A hugely influential philosopher and mathematician, born in La Haye (now named Descartes after its most famous son), Indre-et-Loire, France, who is often referred to as the father of modern philosophy and one of the founders of modern mathematics. He studied law at the University of Poitiers but never practiced it, served in the military for a while, and then lived in Holland for 20 years where he did the bulk of his great work. In his *Meditations on First Philosophy,* he tried to establish what can be known as true beyond doubt. His tool was methodological skepticism: the assumption that any idea that can be doubted is false. He gives the example of dreaming: in a dream, one senses things that seem to be real, but that don't actually exist. Thus, the data of the senses can't be fully trusted. Then again, he mused, perhaps there is an "evil genius"–a supremely powerful and devious being who sets out to prevent anyone from knowing the true nature of reality. Given these possibilities, what is it that one can know for certain? Descartes argues that if "I" am being deceived, then surely "I" must exist–the statement famously referred to as *cogito ergo sum* ("I think, therefore I am"), though these words don't actually appear anywhere in the *Meditations.* Descartes concludes that he can be certain that he exists. But in what form? If the senses are unreliable, Descartes reasons, all he can say for sure is that he is a *thinking thing.* He then proceeds to build a system of knowledge, discarding perception as unreliable and instead admitting only deduction as a method. Halfway through the *Meditations* he also claims to prove the existence of a benevolent **God** who has provided him with a working mind and sensory system, and who cannot desire to deceive him, and thus, finally, he establishes the possibility of acquiring knowledge about the world based on deduction *and* perception.

In mathematics, Descartes is important for his discovery of **analytical geometry**. Up to Descartes's time, **geometry**, dealing with lines and shapes, and **algebra**, dealing with numbers, were regarded as completely independent aspects of mathematics. Descartes showed how almost all problems in geometry can be translated into problems in algebra, by regarding them as questions asking for the length of a line segment, and using a coordinate system to describe the problem. Descartes's theory provided the basis for **calculus**, developed by Isaac **Newton** and Gottfried **Leibniz**, and thus for much of modern mathematics. This is particularly amazing when you consider that Descartes intended it merely as an example to support his *Discours de la méthode pour bien conduire sa raison, et chercher la verité dans les sciences* (Discourse on the method to rightly conduct the reason and search for the truth in sciences), better known under the shortened title *Discours de la méthode.*

Descartes died of pneumonia in Stockholm, where he had been invited to serve as tutor to the energetic 19-year-old Queen Christina of Sweden. Accustomed to working in a warm bed till noon, he was shocked into a rapid decline by having to teach philosophy at 5 A.M. in a freezing library. Seventeen years after his death, the Roman Catholic Church placed his works on the Index of Prohibited Books.

Descartes's circle theorem

See **Soddy's formula**.

determinant

A quantity obtained from a square ($n \times n$) array of numbers that can be useful, among other things, in solving systems of **linear** equations (equations in which the unknowns are raised to at most the first power). More generally, a determinant transforms a square **matrix** into a **scalar**–an operation that has many important properties. Two-by-two determinants were considered by Girolamo **Cardano** at the end of the sixteenth century and ones of arbitrary size by Gottfried **Leibniz** about a century later. Determinants are so named because, when applied to systems of linear equations, they "determine" if the systems are singular–that is, have multiple solutions. They also have important geometric applications, because they describe the area of a **parallelogram** and, more generally, the volume of a **parallelepiped**. A three-rowed determinant is defined by:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{aligned} & a_{11} \times a_{22} \times a_{33} + a_{12} \times a_{23} \times a_{31} \\ & + a_{13} \times a_{21} \times a_{32} - a_{13} \times a_{22} \times a_{31} \\ & - a_{12} \times a_{21} \times a_{33} - a_{11} \times a_{23} \times a_{32}\,. \end{aligned}$$

deterministic system

A system in which the later states of the system follow from, or are determined by, the earlier ones. Such a system contrasts with a *stochastic* or *random system* in which future states are not determined from previous ones. An example of a stochastic system would be the sequence of heads or tails of an unbiased coin, or radioactive decay.

If a system is deterministic, this doesn't necessarily imply that later states of the system are predictable from a knowledge of the earlier ones. In this way, **chaos** is similar to a random system. Chaos has been termed "deterministic chaos" since, although it is determined by simple rules, its property of sensitive dependence on initial conditions makes a chaotic system, in practice, largely unpredictable.

devil's curve

Also known as the *devil on two sticks,* a curve with the Cartesian equation

$$y^4 - a^2y^2 = x^4 - b^2x^2$$

and the polar equation

$$r^2\ (\sin^2\Theta - \cos^2\Theta) = a^2\sin^2\Theta - b^2\cos^2\Theta).$$

Early studies of it were carried out in 1750 by the Swiss mathematician Gabriel Cramer (1704–1752), who is

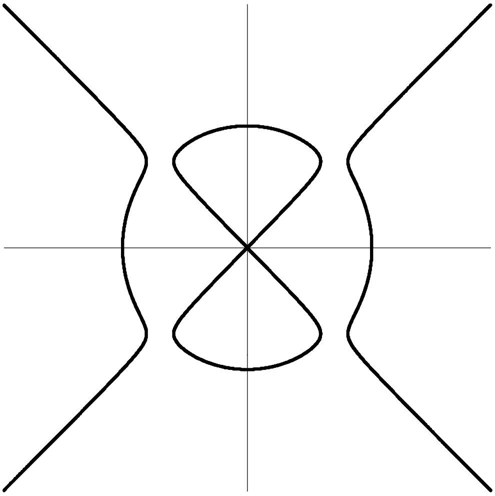

devil's curve *© Jan Wassenaar, www.2dcurves.com* most famous for his work on **determinant**s, and in 1810 by Lacroix. For $a = 25/24$, the curve is called the *electric motor curve.*

Dewdney, Alexander Keewatin (1941–)

A Canadian computer scientist and mathematician at the University of Western Ontario, Canada, best known for his popular books and articles, most notably *The Planiverse: Computer Contact with a Two-dimensional World,* first published in 1984.[81] For several years, Dewdney wrote the "Mathematical Recreations" column for *Scientific American.*

diagonal

A line that joins any two vertices of a **polygon**, if the vertices are not next to each other; or a line that joins two vertices of a **polyhedron** that are not on the same face.

diagonal matrix

A **matrix** that has zero entries along all nondiagonal entries, that is, only the main diagonal may have nonzero values.

diameter

The distance across a **circle** through the center.

dice

Small polyhedra (see **polyhedron**), usually cubes, whose faces are numbered from one to six by patterns of dots, with opposite sides totaling seven. They are thrown, singly or in groups, from the hand or from a cup, onto a flat surface, to provide **random** numbers for gambling and other games. The face of each die that is uppermost when it comes to rest provides the value of the throw. Typical of their use today is the game of craps, in which two dice are thrown together, and bets placed on the total face-up value. Dice probably evolved from knucklebones, which are approximately tetrahedral. Even today, dice are sometimes colloquially referred to as "bones." Ivory, bone, wood, metal, and stone materials have been commonly used to make dice, though the use of plastics is now nearly universal.

Dice found in ancient tombs in the Orient point to an Asiatic origin and dicing is mentioned as an Indian game in the Rig-veda. In its primitive form, knucklebones was essentially a game of skill, played by women and children; gradually, a derivative form evolved for gambling in which four sides of the bones received different values and were counted like dice. Gambling with three, sometimes two, dice was a popular form of amusement in Greece, especially with the upper classes, and was an almost invariable accompaniment to the symposium, or drinking banquet. The Romans were passionate gamblers, and dicing was a favorite form, though it was forbidden except during the festival of Saturnalia (December 17). Throwing dice for money led to many special laws in Rome, one of which decreed that no suit could be brought by a person who allowed gambling in his house, even if he'd been cheated or assaulted! Professional gamblers were common, and some of their loaded dice are preserved in museums.

The Roman historian Tacitus states that the Germans also were passionately fond of dicing–so much so, that, having lost everything, they would even stake their personal liberty. Centuries later, in medieval times, dicing became the favorite pastime of knights, and both dicing schools and guilds of dicers flourished.

Dice are frequently used to randomize allowable moves in board games such as **backgammon**. Loaded dice can be made in many ways to cheat at such games. Weights can be added, or some edges made round while others are sharp, or some faces made slightly off-square, to make some outcomes more likely than would be predicted by pure chance. Dice with non-cubical shapes were once almost exclusively used by fortune-tellers and in other occult practices, but they have become popular lately among players of role-playing and war-games.

difference equation

An equation that describes how something changes in discrete time steps. Numerical solutions to **integral**s are usually realized as difference equations.

differential

A term such as dx used in an expression such as $y dx - x dy$ to denote first-order small changes in the variable. *Differentiation* is the method by which a differential is found.

differential equation

A description of how something continuously changes over time (see **continuity**). Some differential equations have an exact *analytical solution* such that all future states can be known without simulating the time evolution of the system. However, most have a *numerical solution* with only limited accuracy. A differential equation involves the first or higher derivatives of the **function** to be solved for. If the equation only involves first derivatives, it is known as an equation of *order* one, and so on. If only nth powers of the derivatives are involved, the equation is said to have *degree n.* Equations of degree one are called *linear.* Equations in only one variable are called **ordinary differential equation**s to distinguish them from **partial differential equation**s, which have two or more.

differential geometry

The study of geometry using **calculus**; it has many applications in physics, especially in **relativity theory**. The objects studied by differential geometry are known as *Riemannian **manifold**s*. These are geometrical objects, such as surfaces, that locally look like **Euclidean space** and therefore allow the definition of analytical concepts such as tangent **vector**s and tangent space, differentiability (see **differential**), and vector and **tensor** fields. Riemannian manifolds have a **metric**, which opens the door to measurement because it allows distances and angles to be evaluated locally and concepts such as **geodesics**, **curvature**, and torsion to be defined.

differential topology

A branch of **topology** concerned with those properties of **differential geometry** that are preserved by continuous transformation.

differentiation

The method by which the **derivative** of a **function** is found.

digimetic

A **cryptarithm** in which digits are used to represent other digits.

digit

A symbol or numeral that is used to represent an **integer** in a positional **number system**. Examples of digits include the decimal characters 0 through 9, the binary characters 0 or 1, and the hexadecimal digits 0 . . . 9, A . . . F. The word comes from the Latin *digitus* for "finger" or "toe," and retains this meaning, reminding us of the origins of our base 10 number system. The earlier Indo-European root *deik* is related to many other words that hark back to the use of the hands and fingers to "point" out objects, including index, indicate, token, and teach.

digital root

Take a number, n, add its digits, then add the digits of numbers derived from it, and so on, until the remaining number has only one digit. This single digit result is called the digital root of n. For example, in the case of 5381: $5 + 3 + 8 + 1 = 19$; $1 + 9 = 10$; $1 + 0 = 1$; thus, the digital root of 5381 is 1. See also **casting out nines**.

digraph

A **graph** in which each edge has a direction associated with it.

dihedral angle

The angle defined by two given faces meeting at an edge; for example, all the dihedral angles of a cube are 90°. An almost-spherical **polyhedron** (with many faces) has small dihedral angles.

dimension

An extension in some unique direction or sense; the word comes from the Latin *dimetiri* for "measured out." The most common way to think of a dimension is as one of the three spatial dimensions (up-down, left-right, back-forth) in which we live. Mathematicians and science fiction writers alike have long imagined what it would be like in a world with a different number of spatial dimensions. Speculation has particularly focused on **two-dimensional worlds** and, to an even greater extent, on the **fourth dimension**. **Time** is also thought of as a dimension; indeed, in **relativity theory** and as a component of **space-time**, it is treated almost exactly the same as a dimension of space. The universe may have additional spatial dimensions—a total of 10, 11, or 26 are especially favored—according to some theories of the subatomic world (see **string theory** and **Kaluza-Klein theory**), though the additional ones are "curled up" incredibly small and only become important at scales far smaller than those that can be experimentally probed today.

In mathematics, the term *dimension* is used in many different ways. Some of these correspond to the everyday idea of an extension in physical space or to some of the more esoteric meanings in physics. Others are purely abstract and exist only in certain types of theoretical, mathematical space. There are, for example, *Hamel dimensions, Lebesgue covering dimensions,* and **Hilbert space**s. So-called **Hausdorff dimension**s are used to characterize **fractal**s—mathematical objects that have fractional dimensions—by giving a precise meaning to the idea of how well something, such as an extremely "wriggly" curve or surface, fills up the space in which it is embedded.

dinner party problem

See **Ramsey theory**.

Diocles (c. 240–c. 180 B.C.)

A Greek mathematician and contemporary of **Apollonius** who studied the **cissoid** as part of an attempt to duplicate the cube and also was the first to prove the focal property of a parabolic mirror.

Diophantine approximation

The approximation of a **real number** by a **rational number**.

Diophantine equation

An equation that has integer **coefficient**s and for which integer solutions are required. Such equations are named after **Diophantus**. The best known examples are those from **Pythagoras's theorem**, $a^2 = b^2 + c^2$, when a, b, and c are all required to be whole numbers—a **Pythagorean triplet**. Despite their simple appearance Diophantine equations can be fantastically difficult to solve. A notorious example comes from **Fermat's last theorem** (recently solved), $a^n = b^n + c^n$ for $n > 2$. To give a specific example, suppose we want to find integer values for x and y such that

$$x^2 = 1620y^2 + 1.$$

A trial-and-error approach using a computer would quickly find the solution: $y = 4$, $x = 161$. However, just a slight change to the equation to make it

$$x^2 = 1621y^2 + 1$$

would leave the trial-and-error method floundering, even with the resources of the most powerful computers on Earth. The smallest integer solution to this innocent looking formula involves a y-value that is on the order of a thousand trillion trillion trillion trillion trillion trillion! One of the challenges (the tenth one) that David **Hilbert** threw down to twentieth-century mathematicians in his famous list was to find a general method for solving equations of this type. In 1970, however, the Russian mathematician Yu Matijasevic showed that there is no general algorithm for determining whether a particular Diophantine equation is soluble: the problem is undecidable.[217, 218]

Diophantus of Alexandria (A.D. C. 200–C. 284)

A Greek mathematician who developed his own algebraic notation and is sometimes called "the father of algebra." His works were preserved by the Arabs and translated into Latin in the sixteenth century, when they served to inspire momentous new advances. **Diophantine equation**s are named in his honor. It was in the margin of a French translation of Diophantus's work *Aritmetike* from c. 250 that Pierre de **Fermat** scribbled his famous comment that became known as **Fermat's last theorem**.

Diophantus's riddle

One of the oldest known age problems (see **age puzzles and tricks**). It comes from the *Greek Anthology,* a collection of puzzles compiled by Metrodorus in about A.D. 500, and purports to tell how long **Diophantus** lived in the form of a riddle engraved on his tombstone:

> God vouchsafed that he should be a boy for the sixth part of his life; when a twelfth was added, his cheeks acquired a beard; He kindled for him the light of marriage after a seventh, and in the fifth year after his marriage He granted him a son. Alas! late-begotten and miserable child, when he had reached the measure of half his father's life, the chill grave took him. After consoling his grief by this science of numbers for four years, he reached the end of his life.

If d and s are the ages of Diophantus and his son when they died, then the epitaph boils down to these two equations,

$$d = (\tfrac{1}{6} + \tfrac{1}{12} + \tfrac{1}{7})d + 5 + s + 4$$
$$s = \tfrac{1}{2} d,$$

which can be solved simultaneously to give $s = 42$ years and $d = 84$ years.

Dirac, Paul Adrien Maurice (1902–1984)

A British theoretical physicist who played a major role in the development of **quantum mechanics** and predicted the existence of antiparticles. He made his first great breakthrough at Cambridge University in 1928, when he found a wave equation for the electron. This explained aspects of the electron that had previously been observed but not understood, and, incidentally, is the only equation to appear in Westminster Abbey, where it is engraved on Dirac's commemorative plaque. Dirac's electron equation also made the remarkable prediction that there exists a previously unseen type of matter–a particle like the electron, but with the opposite charge. This was startling at the time because only two subatomic particles, the electron and the proton, were known, and there was no suspicion that others might be waiting in the wings. The prediction was fulfilled four years later when the *positron,* as it is now called, was first seen. A central theme of Dirac's work was his belief that **beauty and mathematics** go hand in hand. When a journalist once asked him to explain the concept of mathematical beauty, Dirac asked the journalist "Do you know mathematics?" and when the journalist replied "No," Dirac said, "Then you can't understand the concept of mathematical beauty." A shy, retiring person, Dirac is not as famous as his achievements warrant.

Dirac string trick

Take a cardboard square and tie the four corners to another larger square by loose string. Rotate the small square by 360° about a vertical axis, that is, in a horizontal plane. The strings will become somewhat tangled, and it is not possible to untangle them without rotating the square. Turn the square through another 360°, for a total of 720°. Contrary to all expectations, it is now possible to untangle the string, without further rotation of the square, simply by allowing enough space for the strings to be looped over the top of the square!

Another version of the Dirac string trick has been called the Philippine wineglass trick. A glass of water held in the hand can be rotated continuously through 720° without spilling any water. Surprisingly, these geometrical demonstrations are related to the physical fact that an

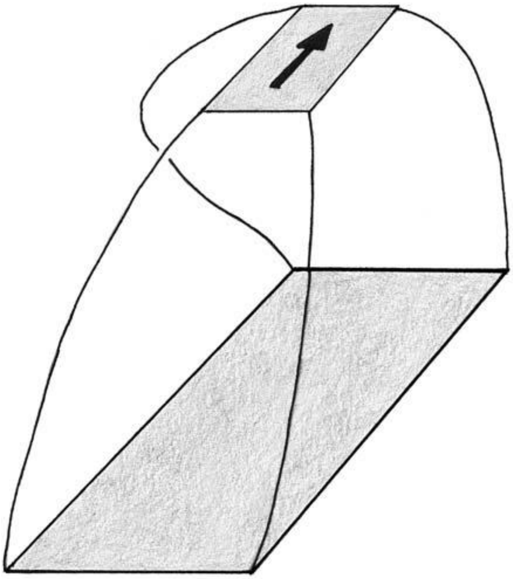

Dirac string trick The equipment needed to simulate an electron's spin property.

electron has spin ½. A particle with spin ½ is something like a ball attached to its surroundings with string. Its amplitude changes under a 360° (2π) rotation and is restored on rotation to 720° (4π).

direct proportion

The relationship two quantities have if the graph of one against the other is a straight line through the origin; so if one doubles then the other doubles, and so forth.

directed graph

Also known as a *digraph,* a **graph** in which each edge is replaced by a directed edge, indicated by an arrow. A directed graph having no multiple edges or loops is called a *simple directed graph.* A **complete graph** in which each edge is bidirected is called a complete directed graph. A directed graph having no symmetric pair of directed edges (i.e., no bidirected edges) is known as an *oriented graph.* A complete oriented graph (i.e., a directed graph in which each pair of vertices is joined by a single edge having a unique direction) is called a *tournament.*

directrix

The line that, together with a point called the **focus**, serves to define a **conic section** as the **locus** of points whose distance from the focus is proportional to the horizontal distance from the directrix. If the ratio $r = 1$, the conic is a **parabola**, if $r < 1$, it is an **ellipse**, and if $r > 1$, it is a **hyperbola**.

Dirichlet, Peter Gustav Lejeune (1805–1859)

A German mathematician who made significant contributions to **number theory**, **analysis**, and mechanics, and who is credited with the modern formal definition of a **function**. He taught at the universities of Breslau (1827) and Berlin (1828–1855) and in 1855 succeeded Carl **Gauss** at the University of Göttingen but died of a heart attack only three years later. Dirichlet continued Gauss's great work on number theory, publishing on **Diophantine equation**s of the form $x^5 + y^5 = kz^5$. His book *Lectures on Number Theory* (published posthumously in 1863) is similar in stature to Gauss's earlier *Disquisitiones* and founded modern **algebraic number theory**. In 1829 he gave the conditions sufficient for a **Fourier series** to converge (though the conditions *necessary* for it to converge are still undiscovered).

Dirichlet's theorem

For any two positive **coprime** integers, a and b, there are infinitely many **prime number**s of the form $an + b$, where $n > 0$. This theorem was first conjectured by Karl **Gauss** and proved by Peter **Dirichlet** in 1835.

discontinuity

Also called a *jump,* a point at which a **function** is not **continuous**.

discrete

Taking only non**continuous** values, for example, **Boolean** or **natural number**s.

discriminant

A quantity that gives valuable information about the solutions of an equation. In the case of the **quadratic** equation $ax^2 + bx + c = 0$, the discriminant is given by $d = b^2 - 4ac$. If $d > 0$, the roots of the equation are two different **real number**s; if $d = 0$, the roots are real and equal; if $d < 0$, the roots are **complex number**s. The concept of discriminant can also be applied in the case of **polynomial**s, **elliptic curve**s, and **metric**s.

disk

Roughly speaking, the "filling" of a **circle**. A flat (two-dimensional) disk of radius r consists of all the points that are at a distance $\leq r$ (*closed disk*) or $< r$ (*open disk*) from a fixed point in the plane. More generally, an n dimensional disk of radius r is the set of all points at a distance $\leq r$ (*closed*) or $< r$ (*open*) from a fixed point in

Euclidean *n*-space. A disk is the two-dimensional analog of a **ball**.

disme

An old word for "tenth." A notation for **decimal** fractions was introduced for the first time in 1585 in a pamphlet called *La Disme* by Simon Stevin of Holland.

dissection

Cutting apart one or more figures and rearranging the pieces to make another figure. Dissection puzzles have been around for thousands of years. The problem of dissecting two equal squares to form one larger square using four pieces dates back to at least the time of Plato (427–347 B.C.). In the tenth century, Arabian mathematicians described several dissections in their commentaries on Euclid's *Elements*. The eighteenth-century Chinese scholar Tai Chen presented an elegant dissection for approximating the value of **pi**. Others worked out dissection proofs of the **Pythagoras's theorem**. In the nineteenth century, dissection puzzles by Sam **Loyd**, Henry **Dudeney**, and others became tremendously popular in magazine and newspaper columns. A classic example is the **Haberdasher's puzzle**. Dissections can get quite elaborate: an eight-piece octahedron becomes a hexagon, a nine-piece five-pointed star becomes a pentagon, and so on. See also **tangrams** and **Loculus of Archimedes**.[201, 202]

dissipative system

A **dynamical system** that contains internal friction that deforms the structure of its **attractor**. Dissipative systems often have internal structure despite being far from **equilibrium**, like a whirlpool that preserves its basic form despite being in the midst of constant change.

distortion illusion

An illusion that distorts an image's shape and/or size. Famous examples include **Poggendorff illusion**, **Zöllner illusion**, **Titchener illusion**, **irradiation illusion**, **Fraser spiral**, **Müller-Lyer illusion**, **Orbison's illusion**, **vertical-horizontal illusion**, and **Ames room**.

distributive

Three numbers *x*, *y*, and *z* are said to be distributive over the operation + if they obey the identity

$$x(y+z) = xy + xz.$$

Compare with **associative** and **commutative**.

diverge

If a **sequence** doesn't converge it is said to diverge (see **convergence**). This can be if it goes to infinity, or if it simply cycles between two or more values without ever staying on one of them. For example, the sequences: 1, 2, 4, 8, 16, 32, . . . and 1, 0, 1, 0, 1, 0, . . . are both divergent.

division

A counterpart to **multiplication** defined so that if

$$a \times b = c$$

where b is nonzero, then

$$a = c/b.$$

In this equation, *a* is the *quotient*, *b* is the *divisor*, and *c* is the *dividend*. A **skeletal division** is a long division in which most or all of the digits are replaced by a symbol (usually asterisks) to form a **cryptarithm**.

dodecagon

A **polygon** with 12 sides.

dodecahedron

A **polyhedron** with 12 faces. A *regular dodecahedron* is made from faces that are identical regular **pentagon**s and is one of the **Platonic solid**s.

dodecahedron A mechanical puzzle in the form of a dodecahedron. *Mr. Puzzle Australia, www.mrpuzzle.com.au*

Dodgson, Charles Lutwidge

See **Carroll, Lewis**.

dollar

The elaborate designs on the various denominations of American dollar bills can be used for some amusing games of the "Can you find . . . ?" variety. On a $1 bill, there is an owl in the upper left-hand corner of the "1" encased in the shield, and a spider hidden in the front upper right-hand corner. There are also at least nine occurrences of **thirteen** things: 13 steps on the pyramid, 13 letters in the Latin above the pyramid, 13 letters in "E Pluribus Unum," 13 stars above the eagle, 13 plumes of feathers on each span of the eagle's wing, 13 bars on the shield, 13 leaves on the olive branch, 13 fruits, and 13 arrows. On the back of a $5 bill the number 172 can be found in the bushes at the base of the Lincoln Memorial, while on the back of the old $10 bill are four cars and eleven light posts. On the new $100 dollar bill the time on the clock tower of Independence Hall reads 4:10.

domain

The set of numbers x for which the **function** $f(x)$ is defined. See also **codomain** and **range**.

domino

A small rectangular tile, marked with spots, that is used to play games. There are many varieties of dominoes and games based on them. Most domino tiles, however, have roughly 2-to-1 proportions and each half of each domino has spots arranged as on a six-sided die; the set generally contains tiles with all possible combinations of two numbers. A tile is identified by the number of spots on each half: for example, "1-6" or "3-3." English/American dominoes include blank sides; Chinese dominoes don't but do duplicate some whole tiles. English/American dominoes can also be bought in larger sets with numbers of spots up to nine or twelve per side. Other than games of strategy, there are many mathematical puzzles that involve dominoes. Some of these puzzles involve tiling variations on the standard 8 × 8 chessboard.

Another common pastime using domino tiles is to stand them on edge in long lines, then topple the first tile, which falls on and topples the second, and so forth, resulting in all of the tiles falling. Arrangements of thousands of tiles have been made that take several minutes to fall. By analogy, phenomena of chains of small events each causing similar events leading to eventual catastrophe are called *domino effects*. The word *domino* was first used to refer to the hooded black cape worn by priests, and later to black masks (of the Lone Ranger type) worn at masquerade balls. The domino is the simplest form of **polyomino**.

PUZZLE

A standard chessboard can easily be tiled by using four dominos in each row. But what if two squares are removed, one each from diagonally opposite corners of the chessboard? Can this reduced board be completely tiled by nonoverlapping dominoes?

Solutions begin on page 369.

domino problem

Is there an algorithm (a set of instructions) that, when given a particular shape as an input, decides if the shape can be used to tile the entire plane? The solution to this unresolved problem is tied up with **Heesch numbers**. The domino problem in turn has a deep connection with the **Einstein problem**.

dozen

See **twelve**.

dragon curve

A classic example of a recursively generated **fractal** shape. Benoit **Mandelbrot** called it the "Harter-Heighway" dragon curve and it formed the subject of one of Martin

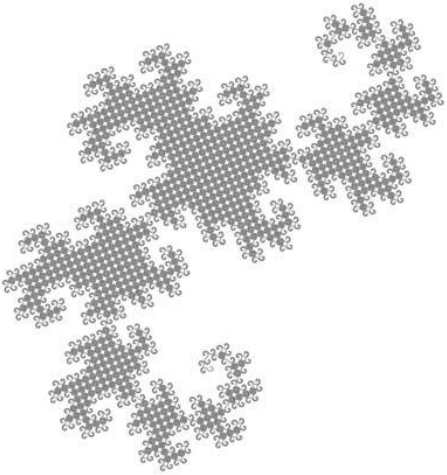

dragon curve *Jos Leys, www.josleys.com*

Gardner's *Mathematical Games* columns in *Scientific American* in 1967.[116] The dragon curve fills out an "island" of positive area with a fractal boundary.

dual

(1) The *dual of a solid* is formed by joining the centers of adjacent faces with straight lines. In the resulting dual solid, each **vertex** of the dual corresponds to a face on the original, each face on the dual to an original vertex, while the edges match, one for one. (2) The *dual of a* ***tessellation*** is obtained by replacing each tile with a point at its center, and each edge between tiles with an edge joining vertices. The dual of a regular tessellation is a regular tessellation; the dual of a semi-regular tessellation is not semi-regular.

Dudeney, Henry Ernest (1857–1930)

An English writer and puzzle-maker who became one of the greatest exponents of recreational mathematics of his time. **Chess** and chess problems captivated him from an early age and he was only 9 when he started contributing puzzles to a local newspaper. His education was limited and he started work as a clerk in the civil service at the age of 13. However, he kept up his interest in math and chess, wrote articles for magazines under the pseudonym "Sphinx," and joined a literary circle that included Arthur Conan Doyle. In 1893 he struck up a correspondence with the American puzzle-maker Sam **Loyd**, the other leading mathematical recreationist of the day, and the two shared many ideas. However, a rift developed after Dudeney accused Loyd of publishing many of Dudeney's puzzles under his own name. One of Dudeney's daughters "recalled her father raging and seething with anger to such an extent that she was very frightened and, thereafter, equated Sam Loyd with the devil." Dudeney was a columnist for the *Strand Magazine* for over 30 years and wrote six books. The first of these, *The Canterbury Puzzles,*[87] published in 1907, purports to include a collection of problems posed by the characters in Chaucer's *The Canterbury Tales*. The answer to the so-called **Haberdasher's puzzle** is Dudeney's best known geometrical discovery. His other books include *Amusements in Mathematics* (1917)[88] and *The World's Best Word Puzzles* (1929). See also **spider-and-fly problem** and **polyomino**.[233]

Dunsany, Lord Edward Plunkett (1878–1957)

An Irish writer who was one of the founders of the fantasy genre of literature. Edward John Moreton Drax Plunkett was born in London to a family whose roots in Ireland predate the Norman invasion. He inherited his father's title in 1899, fought in the Boer War, and returned to the ancestral home, Dunsany Castle, in 1901. Lord Dunsany was a keen marksman and hunter, a fine player of cricket (Dunsany had its own cricket ground near the village), tennis (there is a court beside the Castle), and **chess** (he was an amateur champion and once drew with Grand Master Capablanca. He also wrote chess puzzles for the *Times* over many years and invented his own variant of the game. His first of many books, *The Gods of Pegana,* was published in 1905. In writings that spanned fantasy, drama, poetry, and science fiction, he was an early explorer of such ideas as chess-playing computers (in "The Three Sailors' Gambit" from *The Last Book of Wonder* and, again, in his 1951 novel *The Last Revolution*) and paradoxes in **time travel** (e.g., in "Lost" from *The Fourth Book of Jorkens* and "The King That Was Not" from *Time and the Gods*).

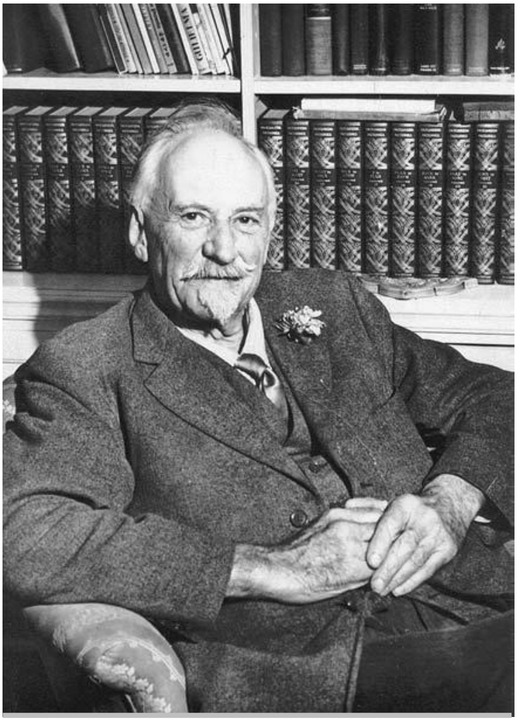

Dunsany, Lord Edward Plunkett *The Dunsany Estate*

Dupin cyclide

The envelope of all **sphere**s touching three given fixed spheres. (Each of the fixed spheres is to be touched in an assigned manner, either externally or internally.) Equivalently, the envelope of all spheres whose centers lie on a given **conic section** and which touch a given sphere. Also equivalently, the inverse of a **torus**.

duplicating the cube

A classical mathematical challenge of antiquity: using only a straightedge and compass, construct a **cube** whose volume is exactly twice that of a given cube. It is often called the "Delian" problem because of a legend that surrounds its origin. The citizens of Athens were being devastated by a plague so that, in 430 B.C., they sought advice from the oracle at Delos on how to rid their community of this pestilence. The oracle replied that the altar of Apollo, which was in the form of a cube, had to be doubled in size. Thoughtless builders merely doubled the edges of the cube, failing to appreciate that this increased the volume of the altar eightfold. The oracle said the gods had been angered; the plague grew worse. Other delegations consulted Plato. When informed of the oracle's admonition, Plato told the citizens "the god has given this oracle, not because he wanted an altar of double the size, but because he wished in setting this task before them, to reproach the Greeks for their neglect of mathematics and their contempt of geometry." Many Greek mathematicians attacked the problem. All failed, because the so-called *Delian constant,* $\sqrt[3]{2}$ (the required ratio of sides of the original cube and that to be constructed), needed for the duplication can't be constructed as prescribed. Cube duplication is possible, however, using a **Neusis construction**. See also **cissoid** of Diocles.

Dürer, Albrecht (1471–1528)

> *Sane judgment abhors nothing so much as a picture perpetrated with no technical knowledge, although with plenty of care and diligence. Now the sole reason why painters of this sort are not aware of their own error is that they have not learnt Geometry, without which no one can either be or become an absolute artist.*
>
> –from *The Art of Measurement,* 1525

A German printmaker who, through applying mathematics to art, brought important ideas to mathematics itself especially in the area of perspective geometry. Dürer was born in Nüremberg, one of 18 children, and showed an early talent for art. After a four-year apprenticeship in painting and woodcutting, he began traveling Europe, especially Italy, in search of new styles and ideas. Back in Nüremberg, he began a serious study of mathematics, absorbing *Elements* by **Euclid** and *De architectura* by the great Roman architect Vitrivius, and studying the work of Leone Alberti (1404–1472) and Luca **Pacioli** on mathematics and art, in particular their work on proportion. His mastery of perspective is clear in woodcuts *Life of a Virgin* (1502–1505). In about 1508, Dürer began to collect material for a major work on mathematics and its applications to the arts. This work was never finished but Dürer did use parts of the material in later published work. One of his most famous engravings *Melancholia,* produced in 1514, contains the first **magic square** to be seen in Europe, cleverly including the date 1514 as two entries in the middle of the bottom row. Also of mathematical interest in *Melancholia* is the **polyhedron** in the picture, the faces of which appear to consist of two equilateral triangles and six somewhat irregular pentagons. In 1825 Dürer published a four-volume treatise, *Underweysung der Messung* (available in English translation as *Painter's Manual*), which dealt with, among other things, the construction of various curves, polygons, and other solid bodies. One of the first books to teach the methods of perspective, it was highly regarded throughout the sixteenth century and presents the earliest known examples of polyhedral **net**s, that is, polyhedra unfolded to lie flat for printing.

Dürer traveled to Italy to learn about perspective and was keen to publish the methods so they weren't kept secret among a few artists. Who he learned from is not known, but Luca Pacioli is a likely possibility. Some of

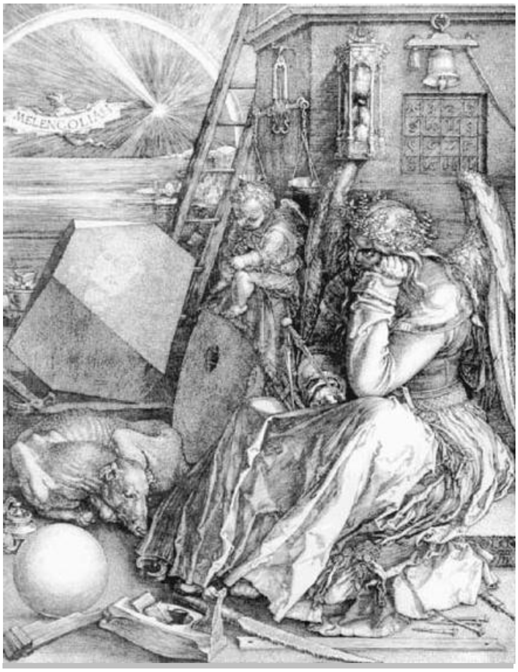

Dürer, Albrecht The famous woodcut *Melancholia,* by Albrecht Dürer, features a magic square and an unusual polyhedron.

the techniques and illustrations also follow closely the work of Piero della Francesca.

Dürer's final work, his *Treatise on Proportion,* was published posthumously and laid the groundwork for descriptive geometry and its rigorous mathematical treatment by Gaspard **Monge**.

Dürer's shell curve

Given a **parabola** and a line that is **tangent** to the parabola, the **glissette** of a point on a line sliding between the parabola and the tangent. It has the equation $(x^2 + xy + ax - b^2) = (b^2 - x^2)(x - y + a)^2$.

dynamical system

A **nonlinear**, interactive system that evolves over time, showing transformations of behavior and an increase in **complexity**. Key to this evolution is the presence and emergence of **attractor**s, most notably **chaotic attractor**s. The changes in the system's organization and behavior are known as **bifurcation**s. Dynamical systems are **deterministic system**s, although they can be influenced by random events. Times series data of dynamical systems can be graphed as phase portraits in **phase space** in order to indicate the qualitative or topological properties of the system and its attractors. For example, various physiological systems, such as the heart, can be conceptualized as dynamical systems. Seeing physiological systems as dynamical systems opens up the possibility of studying various attractor regimes. Moreover, certain diseases can be understood now as "dynamical diseases," meaning that their temporal phasing can be a key to understanding pathological conditions.

dynamics

Pertaining to the change in behavior of a system over time.

E

e

> *Pi goes on and on and on . . .*
> *And e is just as cursed.*
> *I wonder: Which is larger*
> *When their digits are reversed?*
>
> –Martin Gardner

Possibly the most important number in mathematics. Although **pi** is more familiar to the layperson, *e* is far more significant and ubiquitous in the higher reaches of the subject. One way to think of *e* is as the number of dollars you would have in the bank at the end of a year if you invested \$1 at the start of the year and the bank paid an annual interest rate of 100% compounded *continuously*. Compound interest doesn't behave in quite the way intuition suggests. Because more frequent compounding causes the principal to grow faster, it might seem that continuous compounding would make the investor very rich in short order. But the effect tails off. At the end of one year, the \$1 would have grown to a mere \$2.72, rounded to the nearest cent. To a better approximation, *e* is 2.718281828459045 . . . , its decimal expansion stretching out forever, never repeating in any permanent pattern, because *e* is a **transcendental number**. It is the base of natural **logarithms**, which is equivalent to the fact that the area under the curve (the integral of) $y = 1/x$ between $x = 1$ and $x = e$ is exactly equal to one unit. It also features in the **exponential** function $y = e^x$, which is unique in that its value (y) is exactly equal to its growth rate (dy/dx in calculus notation) at every point. As well as showing up in problems involving growth or decay (including compound interest) or in calculus, whenever logarithmic or exponential functions are involved, *e* is at the heart of the statistical **bell curve**; the shape of a hanging cable, known as a **catenary**; the study of the distribution of **prime numbers**; and Stirling's formula for approximating **factorials**.

Like π, *e* pops up as the limit of many **continued fractions** and **infinite series**. Leonhard **Euler**, who was the first to study and to use the symbol *e* (in 1727), found it could be expressed as the curious fraction:

$$1+\cfrac{1}{0+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{4+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{6+\dots}}}}}}}}}}$$

No less remarkable is this infinite series of which *e* is the sum:

$$1 + 1/1! + 1/2! + 1/3! + 1/4! + \dots$$

But of all the places that *e* appears in mathematics none is more extraordinary than **Euler's formula** from which comes the most profound relationship in mathematics: $e^{i\pi} + 1 = 0$, linking *e* and π with **complex numbers**.[215]

Earthshapes

A series of 12 hypothetical Earths as conceived by American airman Joseph Portney in 1968 during a flight to the North Pole onboard a U.S. Air Force KC-135. As the North Pole was reached, Portney looked on the icy terrain below and asked himself, "What if the Earth were . . . ?" The hypothetical Earths, cylindrical, conic, donut-shaped, and so forth, were sketched and captioned by Portney and given to the Litton Guidance & Control Systems graphic arts group to create models. These models were then photographed and became the theme of a Litton publication entitled *Pilots and*

Navigators Calendar for 1969. Each month was introduced with a different one of the 12 hypothetical Earths. The result was an international sensation, attracting awards and heavy fan mail.

eccentricity

See **conic section**.

economical number

A number that has no more digits than there are digits in its prime factorization (including powers). If a number has fewer digits than are in its prime factorization it is known as a *frugal number.* The smallest frugal is 125, which has three digits, but can be written as 5^3, which has only two. The next few frugals are 128 (2^7), 243 (3^5), 256 (2^8), 343 (7^3), 512 (2^9), 625 (5^4), and 729 (3^6). An *equidigital number* is an economical number that has the same number of digits as make up its prime factorization. The smallest equidigitals are 1, 2, 3, 5, 7, and 10 (= 2 × 5). All **prime number**s are equidigital. An *extravagant number* is one that has fewer digits than are in its prime factorization. The smallest extravagant number is 4 (= 2^2), followed by 6, 8, and 9. There are infinitely many of each of these kinds of numbers. Are there also arbitrarily long sequences of consecutive ones? Seven-member strings of consecutive economical numbers start at each of 157; 108,749; 109,997; 121,981; and 143,421. On the other hand, the longest string of consecutive frugal numbers up to 1 million is just two (for example, 4374 and 4375). Even so, it has been proved that if a certain conjecture about prime numbers known as *Dickson's conjecture* is true, then there are arbitrarily long strings of frugals.

Eddington number

"I believe there are 15,747,724,136,275,002,577,605,653, 961,181,555,468,044,717,914,527,116,709,366,231,425, 076,185,631,031,296 protons in the universe and the same number of electrons." So wrote the English astrophysicist Sir Arthur Eddington (1882–1944) in his book *Mathematical Theory of Relativity* (1923). Eddington arrived at this outrageous conclusion after a series of convoluted (and wrong!) calculations in which he first "proved" that the value of the so-called fine-structure constant was exactly $1/136$. This value appears as a factor in his prescription for the number of particles (protons + electrons; neutrons were not discovered until 1930) in the universe: $2 \times 136 \times 2^{256} = 17 \times 2^{260} \cong 3.149544\ldots \times 10^{79}$ (double the number written out in full in the quote above). This is the Eddington number, notable for being the largest *specific* integer (as opposed to an estimate or approximation) ever thought to have a unique and tangible relationship to the physical world. Unfortunately, experimental data gave a slightly lower value for the fine-structure constant, closer to $1/137$. Unfazed, Eddington simply amended his "proof" to show that the value had to be exactly $1/137$, prompting the satirical magazine *Punch* to dub him "Sir Arthur Adding-One." See also **large number**.

edge

A line segment where two faces meet. A **cube**, for example, has 12 edges.

edge coloring theorem

See **Tait's conjecture**.

edge of chaos

The hypothesis that many natural systems tend toward dynamical behavior that borders static patterns and the chaotic regime. See also **chaos**.

egg

Specifically, a chicken's egg and its mathematical equivalent. Eggs are often described as being **oval** in shape, which is effectively tautological since "oval" comes from the Latin *ovus* for "egg." Strictly speaking, an oval is a flat two-dimensional curve, so it is more accurate to say that an egg is shaped like the **surface of revolution** of an oval. In real life, eggs, like ovals, come in a variety of forms all of which can be loosely described as "like an **ellipsoid** but with one end more pointed than the other." Because eggs vary in shape, so too do their mathematical descriptions. Having said this, there are a variety of ways to approximate the shape of a hen's egg by modifying the equation of an ellipsoid, $x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$, so as to

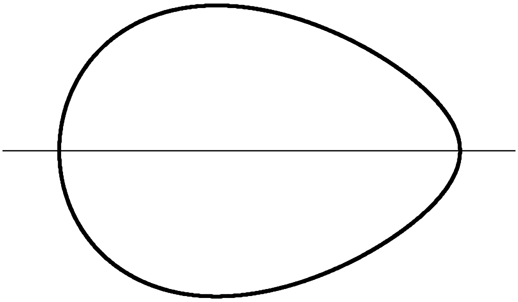

egg A good egg shape is obtained by drawing four circle arcs of different radii. *© Jan Wassenaar, www.2dcurves.com* introduce an asymmetry about the long (say, z-) axis. These involve multiplying z^2/c^2 by a suitable term, so that y becomes larger on the right side of the y-axis and smaller on the left side. For example, $x^2/a^2 + y^2/b^2 + z^2/c^2 (1 - kx) = 1$ gives a good egg. Other useful egg approximations come from surfaces of revolution of **Cartesian ovals**, **Cassinian ovals**, and sections through cones and cylinders. In France, where tennis first became popular, a zero on the scoreboard looked like an egg and was called *l'oeuf*, which is French for "egg." When tennis was introduced in the United States, Americans pronounced it "love."

Why is a hen's, or other bird's, egg shaped as it is? Because it gives strength even though the eggshell is thin enough to allow the young bird to peck its way out when ready. To demonstrate this strength, try balancing a pile of books on four half egg shells. It is even possible for a person's weight to be supported in this way.

Another trick with eggs is to distinguish between a raw egg and a hardboiled one without cracking them open to see which is which. Lay both eggs on their sides on a table, and spin them as you would a spinning top. With a bit of practice, the cooked egg will be made to rise up for a few seconds, while the raw one will remain on its side. The physics of this odd behavior was finally cracked by two mathematicians, Keith Moffat of Cambridge University and Yutaka Shimomura of Keio University, who reported their findings in 2002. They concluded that friction between the egg and the surface produces a gyroscopic effect, which causes some of the kinetic energy of the object to be translated into potential energy, raising its center of gravity (see also **Tippee Top**). As the hardboiled egg spins, its curved surface causes it to touch the tabletop at only one point. The contact point changes and traces out a little circle. If the texture of the tabletop is just right (neither too slippery nor too sticky) the egg will slide a bit as it spins. This sliding slows the spin a bit and causes a wobble. This in turn tilts the egg, lifting one end off the table more than the other, at which point the gyroscopic effect kicks in and swaps some of the kinetic energy of the spinning egg into potential energy and raises its center of gravity in a seemingly paradoxical way. This effect is heightened by the fact that as the end of the egg rises, the egg draws in closer to the axis of spin, causing it to spin more quickly–just as figure-skaters can make themselves pirouette faster by raising their arms above their heads. Why doesn't the effect occur with a raw egg? Because the inside of the egg is runny and it lags behind the shell. This lag serves as a drag, which reduces the spin rate and dissipates the egg's kinetic energy. This in turn reduces the friction between the egg and tabletop, and means that not enough energy is available to be turned into potential energy to raise the egg's center of gravity. As well as solving the mystery of the balancing egg, Moffat also found time to write a limerick to commemorate the event:

> Place a hard-boiled egg on a table,
> And spin it as fast as you're able;
> It will stand on one end
> With vectorial blend
> Of precession and spin that's quite stable.

See also **superegg**.

Egyptian fraction

A unit **fraction**; in other words, a fraction in which the numerator (the number on top) is one. This type of fraction was the only kind used by the ancient Egyptians and appears extensively in the **Rhind papyrus**. Other fractions can be obtained by adding Egyptian fractions together; for example, $5/7 = 1/2 + 1/6 + 1/21$. In 1201 **Fibonacci** proved that every **rational number** can be written as a sum of Egyptian fractions.

eigenvalue

A **complex number**, λ, that satisfies the equation $Ax = \lambda x$, where A is an $n \times n$ **matrix** and x is some **vector**. In this case, x is called an *eigenvector*.

eight

The second smallest cube number (after 1^3): $8 = 2^3 = 2 \times 2 \times 2$. A queen or king in **chess** can move in eight different directions, in the same way that a compass has eight principal points: north, northeast, east, southeast, south, southwest, west, and northwest. In three dimensions, there are eight diagonal ways to move, corresponding to the eight *octants* into which three-dimensional space is divided by three mutually perpendicular planes. Add a **fourth dimension** and movement becomes possible back and forth along four directions at right angles to each other: up and down, left and right, forward and back, and to one other! The Spanish dollar was a gold coin with a value of eight reales, and was sometimes actually cut into eight wedge-shaped pieces–"pieces of eight"–to make change.

eight curve

A curve, also known as the *lemniscate of Gerono*, that has the Cartesian equation

$$x^4 = a^2(x^2 - y^2)$$

and the appearance of a figure eight lying on its side.

Einstein problem

(1) Is there a single shape that will tile a plane aperiodically (see **aperiodic tiling**)? An answer of "no" would imply the existence of a decision method for the **domino problem**. This problem, which is named because of the German translation (*ein* = "one," *stein* = "stone") and was not an invention of the famous scientist, remains unsolved. (2) A logic problem, invented by Albert Einstein, who claimed that 98% of the people in the world couldn't solve it.

PUZZLES

1. There are 5 houses (along a street) in 5 different colors: blue, green, red, white, and yellow.
2. In each house lives a person of a different nationality: Briton, Dane, German, Norwegian, and Swede.
3. These 5 owners drink a certain beverage: beer, coffee, milk, tea, or water; smoke a certain brand of cigar: Blue Master, Dunhill, Pall Mall, Prince, or blend; and keep a certain type of pet: cat, bird, dog, fish, or horse.
4. No owners have the same pet, smoke the same brand of cigar, or drink the same beverage.
5. The Briton lives in a red house. The Swede keeps dogs as pets. The Dane drinks tea. The green house is on the left of the white house (next to it). The green house owner drinks coffee. The person who smokes Pall Mall rears birds. The owner of the yellow house smokes Dunhill. The man living in the house right in the center drinks milk. The Norwegian lives in the first house. The man who smokes blend lives next to the one who keeps cats. The man who keeps horses lives next to the man who smokes Dunhill. The owner who smokes Blue Master drinks beer. The German smokes Prince. The Norwegian lives next to the blue house. The man who smokes blend has a neighbor who drinks water.

The question is: Who keeps the fish?

Solutions begin on page 369.

elementary function

Any real-value algebraic function or transcendental function (trigonometric, hyperbolic, exponential, logarithmic).

eleven

A **palindromic number**, the smallest integer that is not a **Harshad number**, a **prime number** that is a member of a **twin prime** (11 and 13), and the largest integer that is not the sum of two or more distinct primes. There are 11 players on a soccer team and on a cricket team. Strange but true: the youngest pope was 11 years old.

ellipse

A shape that looks like a squashed circle. It is one of the **conic sections** and can be defined as the **locus** of all points in a plane that have the same sum of distances from two given fixed points known as *foci*. If the two foci coincide then the ellipse is a **circle**. The line segment connecting the foci is called the *major axis* of the ellipse; half this is the *semimajor axis, a*. The line passing through the center of the ellipse (the midpoint of the foci) at right angles to the major axis is called the *minor axis*, half of which is the *semiminor axis, b*. An ellipse centered at the origin of an *x-y* coordinate system with its major axis along the *x*-axis is defined by the equation

$$x^2/a^2 + y^2/b^2 = 1.$$

The shape of an ellipse is expressed by a number called the *eccentricity, e*, which is related to a and b by the formula $b^2 = a^2(1 - e^2)$. The eccentricity is a positive number less than 1, or 0 in the case of a circle. The greater the eccentricity, the larger the ratio of a to b, and therefore the more elongated the ellipse. The distance between the foci is $2ae$. The area enclosed by an ellipse is πab. The circumference of an ellipse is $4aE(e)$, where the function E is the complete elliptical integral of the second kind.

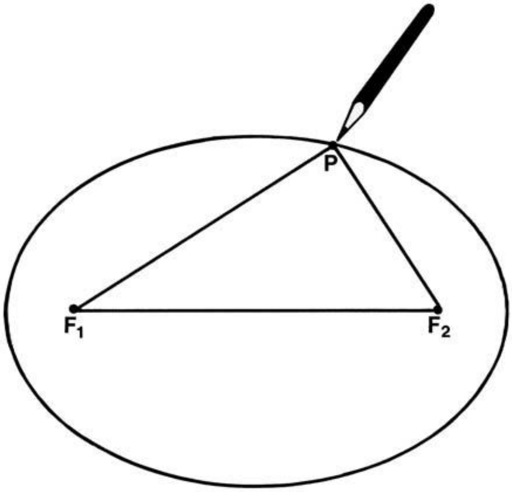

ellipse Nails mark the foci, F_1 and F_2, of an ellipse that is drawn by pencil whose moving tip, P, keeps the string threaded around the nails taut.

ellipsoid

A **quadratic** surface that is the three-dimensional analog of an **ellipse**. The general equation of an ellipsoid in Cartesian coordinates is

$$x^2/a^2 + y^2/b^2 + z^2/c^2 = 1,$$

where a, b, and c are positive real numbers determining the shape. If two of these numbers are equal, the ellipsoid is a **spheroid**; if all three are equal, it is a **sphere**. The intersection of an ellipsoid with a plane is a single point or an ellipse. Ellipsoids can also be defined in higher dimensions.

elliptic curve

The set of solutions to a type of **cubic equation** whose solutions lie on a **torus** (a donut-shaped surface). The particular type of cubic equation whose solutions lead to elliptic curves takes the form

$$y^2 + axy + by = x^3 + cx^2 + dx + e.$$

Elliptic curves, which are said to have a **genus** of 1, have an unusually rich theory and structure, and their study is linked to many other important areas of mathematics and their applications. For example, it was work done on elliptic curves by Andrew **Wiles** that finally led to a proof of **Fermat's last theorem**.

elliptic function

In **complex analysis**, a **function** defined on the complex plane that is **periodic** in two directions. The elliptic functions can be thought of as analogs of the **trigonometric functions** (which have only a single period). Leading eighteenth-century mathematicians, including Leonhard **Euler** and Joseph **Lagrange**, had studied elliptic integrals, such as the integral that gives the arc length of an ellipse; however, these cannot be expressed in terms of the elementary functions (polynomials, exponentials, and trigonometric functions). It was the insight of Karl **Jacobi**, and also of Karl **Gauss** and Niels **Abel**, that the *inverse* functions of elliptic integrals are much easier to study. They turn out to be doubly periodic functions of a complex variable. While a singly periodic function like sine has a number a (specifically $a = 2\pi$) so that $\sin(x + a) = \sin(x)$, a doubly periodic function f has the property that there are two numbers a, b, not rational multiples of each other, so that $f(x + a) = f(x + b) = f(x)$. As Jacobi proved in 1834, the ratio a/b is necessarily an imaginary number.

elliptical geometry

One of the two most important types of **non-Euclidean geometry**: the other is **hyperbolic geometry**. In elliptical geometry, Euclid's **parallel postulate** is broken because no line is parallel to any other line. The original form of elliptical geometry, known as *spherical geometry* or *Riemannian geometry,* was pioneered by Bernhard **Riemann** and Ludwig **Schläfli** and treats lines as **great circle**s on the surface of a sphere. The most familiar example of such circles, which are **geodesic**s (shortest routes) on a spherical surface, are the lines of longitude on Earth. In spherical geometry any two great circles always intersect at exactly two points. Two lines of longitude, for example, meet at the North and South Poles. Working in spherical geometry produces some surprising, nonintuitive results. For instance, it turns out that the shortest flying distance from Florida to the Philippine Islands is a path across Alaska–even though the Philippines are at a more southerly latitude than Florida! The reason is that Florida, Alaska, and the Philippines lie on the same great circle and so are collinear in spherical geometry. Another odd property of spherical geometry is that the sum of the angles of a triangle is greater than 180°. This is always the case on a surface that bulges out or, in mathematical parlance, has positive **curvature**. It was Felix **Klein** who first saw clearly how to rid spherical geometry of its one blemish: the fact that two lines have not one but two common points. He redefined the notion of a point as a **set** of antipodal points. With this definition, any two points determine a unique line so that the traditional form of Euclid's first postulate is restored. Thus modified, spherical geometry became what Klein called elliptical geometry.

embedding

Putting one mathematical object inside another, such as a subgroup within a group or one topological space inside another, while preserving all topological properties.

emergence

The arising of new, unexpected structures, patterns, or processes in a self-organizing system (see **self-organization**). These *emergents* have their own rules, laws, and possibilities, and can be understood as existing on a higher level than that of the components from which they came. The term was first used by the nineteenth-century philosopher G. H. Lewes and came into greater currency in the scientific and philosophical movement known as *emergent evolutionism* in the 1920s and 1930s.

emirp

A **prime number** that becomes a different prime number when its digits are reversed ("emirp" is "prime" spelled backward). The first twenty emirps are 13, 17, 31, 37, 71, 73, 79, 97, 107, 113, 149, 157, 167, 179, 199, 311, 337, 347, 359, and 389. Compare with a *palindromic prime* (see **palindromic number**), which gives the same prime when reversed.

empty set

The **set**, denoted by ∅ or {}, that has no members; also known as the *null set.* This is not the same as **zero**, which is the number of members of ∅. Nor is ∅ the same as **nothing** because a set with nothing in it is still a set, and a set is something. The empty set, for example, is the set of all triangles with four sides, the set of all numbers that are bigger than nine but smaller than eight, and the set of all opening moves in chess that involve a king. Applying the concept of the empty set helps distinguish between the different ways that "nothing" is used in everyday language. In his book *What Is the Name of This Book?* (1978), Raymond **Smullyan** wrote:[304] "Which is better, eternal happiness or a ham sandwich? It would appear that eternal happiness is better, but this is really not so! After all, nothing is better than eternal happiness, and a ham sandwich is certainly better than nothing. Therefore a ham sandwich is better than eternal happiness."

What is wrong with this declaration? The first statement is equivalent to "The set of things that are better than eternal happiness is ∅." The second statement is equivalent to "The set {ham sandwich} is better than the set ∅." The confusion arises because the first is comparing individual things, while the second is comparing sets of things, and ∅ plays a different role in each.

enantiomorph

The mirror image of a given **chiral** polyhedron or other figure.

enormous theorem

The largest theorem in mathematics; it concerns the classification of finite **simple group**s and encapsulates the work of hundreds of mathematicians over many years.

entropy

A measure of a system's degree of **random**ness or disorder.

envelope

A curve or a surface that touches every member of a family of lines, curves, planes, or surfaces.

epicycloid

The path traced out by a point on the circumference of a circle of radius b rolling on the outside of a **circle** of radius a. It is described by the parametric equations:

$$x = (a + b)\cos(t) - b\cos((a/b + 1)t)$$
$$y = (a + b)\sin(t) - b\sin((a/b + 1)t).$$

An epicycloid is like a **cycloid** on the circumference of a circle and is closely related to the **epitrochoid**, **hypocycloid**, and **hypotrochoid**. An epicycloid with one cusp is called a **cardioid**, one with two cusps is called a **nephroid**, and one with five cusps is called a **ranunculoid** (after the buttercup genus *Ranunculus*).

Epimenides paradox

See **liar paradox**.

epitrochoid

A curve traced out by a point that is a distance c from the center of a circle of radius b, where $c < b$, that is rolling around the outside of another circle of radius a. It is described by the parametric equations

$$x = (a + b)\cos(t) - c\cos((a/b - 1)t)$$
$$y = (a + b)\sin(t) - c\sin((a/b + 1)t).$$

Closely related to the epitrochoid are the **epicycloid**, **hypocycloid**, and the **hypotrochoid**. An example of an epitrochoid appears in Albrecht **Dürer**'s work *Instruction in Measurement with Compasses and Straightedge* (1525).

EPORN

An equal product of reversible numbers; defined by the Indian recreational mathematician Shyam Sunder Gupta as a number that can be expressed as the product of two reversible numbers (numbers whose digits are reversed) in two different ways. For example: 4,030 = 130 × 031 = 310 × 013 and 144,648 = 861 × 168 = 492 × 294. The smallest EPORN, 2,520 = 120 × 021 = 210 × 012, is also the least common multiple of all single digit natural numbers in decimal system. The **digital root**, i.e. the ultimate sum of digits, of all EPORNs is always 1, 4, 7, or 9. For example, 2,520 = 2 + 5 + 2 + 0 = 9; 4,030 = 4 + 0 + 3 + 0 = 7; 9,949,716 = 9 + 9 + 4 + 9 + 7 + 1 + 6 = 36 and 3 + 6 = 9.

equichordal point

A point inside a closed **convex** curve in the plane, all the **chord**s through which have the same length.

equilateral

Having sides of equal length, as in the case of an equilateral **polygon**. The equilateral **triangle**, with its three equal angles of 60°, is widely found in historic buildings and structures across Europe. See **Triangular Lodge**.

equilibrium

A term indicating a rest state of a system, for example, when a **dynamical system** is under the sway of a **fixed-point attractor** or **periodic attractor**. The concept originated in ancient Greece when **Archimedes** experimented with levers in balance, literally "equilibrium." The idea was elaborated through the Middle Ages, the Renaissance, and the birth of modern mathematics and physics in the seventeenth and eighteenth centuries. "Equilibrium" has come to mean pretty much the same thing as *stability,* that is, a system that is largely unaffected by internal or external changes since it easily returns to its original condition after being perturbed.

equivalent numbers

Numbers such that the sums of their **aliquot part**s (proper divisors) are the same. For example, 159, 559, and 703 are equivalent numbers because their aliquot parts all sum to 57.

Eratosthenes of Cyrene (c. 276–194 B.C.)

A Greek mathematician, astronomer, and geographer who was born in Cyrene, a Greek colony to the west of Egypt. He studied at Plato's school in Athens and eventually became the chief librarian of the great Library at Alexandria. He wrote works on geography, philosophy, history, astronomy, mathematics, and literary criticism. One of Eratosthenes' contributions to mathematics was his measurement of Earth's circumference, which he calculated to be about 252,000 stadii, or 24,700 miles (about one-tenth the actual value, but still a big improvement on earlier estimates). Eratosthenes is also known in number theory for his **sieve of Eratosthenes**, which finds all **prime number**s less than a given integer *n*.

Eratosthenes's sieve

See **sieve of Eratosthenes**.

Erdös, Paul (1913–1996)

A Hungarian mathematician (his name is pronounced "AIR-dosh"), one of the greatest mathematicians of the twentieth century and, in terms of the number of papers published (more than 1,500), the most prolific in history–beating out even Leonhard **Euler** and inspiring the term "Erdös number." A mathematician has an Erdös number of 1 if he or she has published a paper with Erdös, of 2 if he or she has published with someone who published a paper with Erdös, and so on. Erdös worked almost nonstop, 19 hours a day, 7 days a week. "A mathematician," he quipped, "is a machine for turning coffee into theorems." At age 20, Erdös discovered an elegant proof of a famous theorem in number theory, known as *Chebyshev's theorem,* which says that for each number greater than one, there is always at least one **prime number** between it and its double. **Number theory** remained one of his chief interests, though his work spread across many fields, and he became renowned for posing and solving problems that were often simple to state but notoriously difficult to solve. He did groundbreaking work in a branch of mathematics known as **Ramsey theory** long before it became fashionable in the late 1950s. Bent and slight, often wearing sandals, Erdös had no time for the material side of life. "Property is nuisance," he said. Focused totally on mathematics, Erdös traveled from meeting to meeting, carrying a half-empty suitcase and staying with mathematicians wherever he went. His colleagues took care of him, lent him money, fed him, bought him clothes, and even did his taxes. In return, he showered them with ideas and challenges–with problems to be solved and brilliant ways of attacking them. Ernst Straus, who worked with both Albert Einstein and Erdös, wrote a tribute to Erdös shortly before his own death in 1983. He said of Erdös: "In our century, in which mathematics is so strongly dominated by 'theory doctors,' he has remained the prince of problem solvers and the absolute monarch of problem posers."[170]

ergodic

The property of a **dynamical system** such that all regions of a **phase space** are visited with similar frequency and that all regions will be revisited (within a small proximity) if given enough time.

Escher, Maurits Cornelius (1898–1972)

My work is a game, a very serious game.

A Dutch artist whose graphic explorations of **tiling**, figure-ground ambiguities, **impossible figure**s, and regression has attracted the interest of mathematicians and scientists. His experiences of Moorish art (see **Alhambra**) and his contact with mathematicians, most notably Harold **Coxeter**, led him to explore the way repetitive shapes can be used to tile the plane, and from this to ideas about duality and transformation. Escher's preoccupation with dualities is a constant presence in his work in the form of foreground/background, light and dark, flatness and dimensionality, representation and decoration, frame and scene, large and small, viewpoint and vanishing point, form and negative space, positive and negative, observer and observed, as well as the metaphysical aspects of good and evil. Self-referential images (see **self-referential sentence**) resonate throughout Escher's works–reflections of the artist, hands that draw themselves, the visitor in a picture gallery who looks at a print that contains him. It was for this reason that Douglas **Hofstadter** wove Escher, along with Kurt Gödel and Bach, into the "eternal golden braid" at the heart of his Pulitzer Prize–winning book.[172] Having seen some of Escher's art, Roger **Penrose** was inspired to devise impossible figures, including the **Penrose triangle**, which Escher then incorporated into several of his later works. After the artist's death Penrose regretted that Escher had not lived long enough to take advantage of the discovery of **Penrose tiling**.

escribed circle

A circle that is **tangent** to one side of a triangle and to the extensions of the other sides.

Eternity Puzzle

An enormously difficult jigsaw consisting of 209 pieces, each one different and each made from a unique configuration of equilateral triangles and half-triangles with the same total area as six triangles. The puzzle was to fit them together into an almost-regular 12-sided figure aligned to a triangular grid. The puzzle's inventor, Christopher Monckton, announced a prize of $1 million when the puzzle was released commercially in June 1999, for the first correct solution submitted, assuming there was one, when all the solutions were opened in September 2000. Monckton had run computer searches on much smaller versions of the puzzle, which had convinced him that the sheer size of Eternity would make it intractable. However, the prize was won by two British mathematicians, Alex Selby and Oliver Riordan, with the help of a couple of computers, who sent in a correct tiling on May 15, six weeks ahead of the only other puzzler known to have found a correct solution. Early on, Selby and Riordan made a surprising discovery. As the number of pieces in an Eternity-like puzzle increased, so did the difficulty–but only up to a point. The critical size is about 70 pieces, which would be almost impossible to solve. For larger puzzles, however, the number of possible correct solutions increases. In the case of Eternity itself, with its 209 pieces, there are thought to be at least 10^{95} solutions–far more than the number of subatomic particles in the universe but far, far less than the number of nonsolutions. The puzzle itself is much too large to solve by an exhaustive search but not, as it turns out, by more savvy methods that take into account what shaped regions are easiest to tile and what shaped pieces are easiest to fit. By steadily refining their search algorithm, Selby and Riordan were able to prune out the vast majority of nonsolutions and, with a bit of

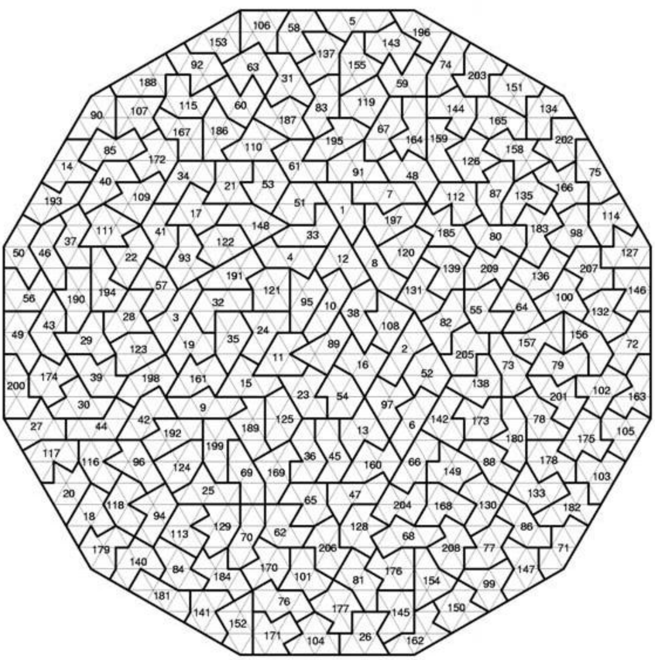
106
58
5
153
137
143
196
92
63
31
155
74
203
188
59
151
90
107
115
60
83
119
187
144
134
167
186
67
165
85
172
110
195
164
159
158
202
126
14
61
91
48
75
40
34
21
53
166
193
109
7
51
112
87
135
1
197
17
114
111
41
148
185
80
183
98
37
33
122
120
50
46
22
93
4
12
8
127
191
139
209
136
207
57
121
131
56
190
194
95
10
100
146
32
38
55
64
132
28
3
108
82
43
24
156
49
29
35
19
89
157
72
123
2
205
11
73
79
16
174
39
198
161
15
52
138
200
23
54
201
102
163
30
9
97
42
125
6
142
173
78
27
44
192
13
180
105
189
175
117
199
116
96
36
45
160
66
88
178
124
69
169
149
103
20
25
65
47
118
94
204
168
130
133
182
18
129
128
70
62
68
86
113
179
206
208
77
71
140
84
184
170
101
81
176
154
147
99
181
141
177
145
150
152
76
171
104
26
162

Eternity Puzzle The solution to the Eternity Puzzle that was awarded a $1 million prize. *Eternity pieces are copyright © 1999 by Christopher Monckton* good fortune, to hit upon a correct solution and claim the prize.

Euclid of Alexandria (c. 330–270 B.C.)

A Greek mathematician who compiled and systematically arranged the geometry and number theory of his day into the famous text *Elements*. This text, used in schools for about 2,000 years, earned him the name "the father of geometry." Even today, the geometries that don't satisfy the fifth of Euclid's "common notions" (now called axioms or postulates) are called non-Euclidean geometries. When, according to the Greek philosopher Proclus, the Egyptian ruler Ptolemy asked if there was a shorter way to the study of geometry than the *Elements,* Euclid told the pharaoh that "there is no royal road to geometry." Little is known of Euclid's life. Proclus wrote (c. A.D. 350) that Euclid lived during the reign of Ptolemy and founded the first school of mathematics in Alexandria–the site of the most impressive library of ancient times with perhaps as many as 700,000 volumes. He wrote books on other subjects such as optics and conic sections, but most of them are now lost. See also **Euclidean geometry**.

Euclidean geometry

Geometry of the type described originally by **Euclid** in his book *Elements* and based on five axioms (see **Euclid's postulates**), one of which is the controversial **parallel postulate**. Various forms of **non-Euclidean geometry** began to emerge in the nineteenth century, with enormous implications for science and philosophy. See also **Euclidean space**.

Euclidean space

Any n-dimensional mathematical space that is a generalization of the familiar two- and three-dimensional spaces described by the axioms of **Euclidean geometry**. The term "n-dimensional Euclidean space" (where n is any positive whole number) is usually abbreviated to "Euclidean n-space", or even just "n-space". Formally, Euclidean n-space is the set $\mathbf{R}^n$ (where $\mathbf{R}$ is the set of **real numbers**) together with the *distance function,* which is obtained by defining the distance between two points $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ to be the square root of $\Sigma(x_i - y_i)^2$, where the sum is over $i = 1, \ldots, n$. This distance function is based on **Pythagoras's theorem** and is called the *Euclidean metric.*

Euclid's postulates

The five postulates, which together with 23 definitions and five "common notions," form the basis of **Euclid**'s great work on geometry, *Elements.* The postulates are:

1. A straight line may be drawn from any one point to any other point.
2. A finite straight line may be produced to any length in a straight line.
3. A circle may be described with any center at any distance from that center.
4. All right angles are equal.
5. If a straight line meets two other lines, so as to make the two interior angles on one side of it together less than two right angles, the other straight lines will meet if produced on that side on which the angles are less than two right angles.

The last postulate is not as obvious as the other four, and Euclid himself was reluctant to use it. Later mathematicians, finding the fifth postulate to be complicated, thought it might be possible to derive it from the other four. However, they only succeeded in replacing it with equivalent statements. The most common of these is the **parallel postulate**.

Eudoxus of Cnidus (c. 408–c. 355 B.C.)

A Greek astronomer, mathematician, and physician whose work on ratios formed the basis for Book V of **Euclid**'s *Elements* and anticipated some aspects of algebra, such as cross multiplying, which is otherwise absent from ancient Greek mathematics. Eudoxus constructed many geometric proofs, found formulas for measuring pyramids, cones, and cylinders, and developed the **method of exhaustion**, a forerunner of integration, later extended by Archimedes. He also studied the kampyle curve, often known as the **kampyle of Eudoxus**, in connection with the classical problem of **duplicating the cube**.

Euler, Leonhard (1707–1783)

A great Swiss mathematician; the second most prolific mathematician in history, after Paul **Erdös**. His greatest contributions were to **number theory**, but Euler also did important work in **calculus**, geometry, algebra, probability, acoustics, optics, mechanics, astronomy, artillery, navigation, and finance. He had a knack for coming up with important results by intuition, he cast calculus and trigonometry in their modern forms, and he showed the importance of the number ***e***. Even the amusing puzzles he invented and, in some cases, solved have opened up new mathematical fields. The **bridges of Königsberg** problem, for example, heralded the beginning of **graph** theory and **topology**, while his **thirty-six officers problem** stimulated important work in **combinatorics**. Euler also worked on **magic squares** and the problem of the **knight's tour**. Having learned some math from his father, a Calvinist preacher, Euler studied at the University of

Basle where he became close friends with members of the **Bernoulli family**. In 1727, he moved to St. Petersburg, to the court of Catherine the Great, becoming professor of physics (1730) and of mathematics (1733). While in Russia, Euler, a devout Christian, met the encyclopedist and philosopher René Diderot, a notorious atheist. When Diderot heard that Euler had a mathematical proof of the existence of **God**, he asked for it and was quoted the equation now often referred to as **Euler's formula**.

Upon losing the use of his right eye, Euler said "Now I will have less distraction." Indeed, the quantity of his output seemed to be inversely proportional to the quality of his sight, because his rate of publication increased after he became almost totally blind in 1766. Euler died moments after calculating the orbit of Uranus on September 18, 1783.

Euler characteristic

An important kind of number, known as a *topological invariant,* that describes a closed surface. In the case of polyhedra, the Euler characteristic is the number of vertices and faces minus the number of edges (see **Euler's formula for polyhedra**).

Euler circuit

A **connected graph** such that starting at a **vertex** *a,* one can traverse every **edge** of the graph once to each of the other vertices and return to vertex *a.* In other words a Euler circuit is a **Euler path** that is a circuit. Thus, using the properties of odd and even degree vertices given in the definition of a Euler path, a Euler circuit exists if and only if every vertex of the graph has an even degree. See also **mazes**.

Euler line

A line that connects the **centroid** and the **circumcenter** of a triangle.

Euler path

A path along a **connected graph** that connects all the vertices (see **vertex**) and that traverses every **edge** of the graph only once. Note that a vertex with an odd degree allows one to travel through it and return by another path at least once, while a vertex with an even degree only allows a number of traversals through, but one cannot end a Euler path at a vertex with even degree. Thus, a connected graph has a Euler path which is a circuit (a **Euler circuit**) if all of its vertices have even degree. A connected graph has a Euler path which is non-circuitous if it has exactly two vertices with odd degree. See also **Hamilton path**.

Euler square

A square array made by combining n objects of two types such that the first and second elements form a **Latin square**. Euler squares are also known as *Graeco-Latin squares, Graeco-Roman squares,* or *Latin-Graeco squares.* For many years, Euler squares were known to exist for $n = 3$, 4, and for every odd n except $n = 3k$. *Euler's Graeco-Roman squares conjecture* maintained that there are no Euler squares of order $n = 4k + 2$ for $k = 1, 2, \ldots$. However, such squares were found to exist in 1959 by Bose and Shrikande, refuting the conjecture.

Euler-Mascheroni constant (γ)

Also known as *Euler's constant* or *Mascheroni's constant,* the limit (as n goes to infinity) of

$$1 + \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{4} + \tfrac{1}{5} + \ldots + 1/n - \log n$$

It is often denoted by a lowercase gamma, γ, and is approximately 0.5772156649 . . . . Even though over one million digits of this number have been calculated, it isn't yet known if it is a rational number (the ratio of two integers a/b). If it is rational, the denominator (b) must have more than 244,663 digits. The constant γ crops up in many places in **number theory**. For example, in 1898, the French mathematician Charles de la Vallée Poussin (who proved the **prime number** theorem) proved the following: Take any positive integer n and divide it by each positive integer m less than n. Calculate the average (mean) fraction by which the quotient n/m falls short of the next integer. The larger n gets, the closer the average gets to gamma.

Euler's conjecture

It always takes n terms to sum to an nth power: two squares, three cubes, four fourth powers, and so on. This hypothesis is now known to be wrong. In 1966, L. J. Lander and T. R. Parkin found the first counterexample: four fifth powers that sum to a fifth power. They showed that $27^5 + 84^5 + 110^5 + 133^5 = 144^5$. In 1988, Noam Elkies of Harvard University found a counterexample for fourth powers: $2{,}682{,}440^4 + 15{,}365{,}639^4 + 187{,}960^4 = 20{,}615{,}673^4$. Subsequently, Roger Frye of Thinking Machines Corporation did a computer search to find the smallest example: $95{,}800^4 + 217{,}519^4 + 414{,}560^4 = 422{,}481^4$.

Euler's constant

See **Euler-Mascheroni constant**.

Euler's formula

For any **real number** x, Euler's formula is

$$e^{ix} = \cos x + i \sin x$$

where ***e*** is a fundamental constant (the base of natural logarithms) and $\boldsymbol{i} = \sqrt{-1}$. If we now put $x = \pi$, we get

$$e^{i\pi} = \cos \pi + i \sin \pi,$$

and since $\cos(\pi) = -1$ and $\sin(\pi) = 0$, this reduces to

$$e^{i\pi} = -1$$

so that

$$e^{i\pi} + 1 = 0.$$

This most extraordinary equation first emerged in Leonhard **Euler**'s *Introductio,* published in 1748. It is remarkable because it links the most important mathematical constants, *e* and $\boldsymbol{\pi}$, the imaginary unit *i,* and the basic numbers used in counting, 0 and 1. In describing the equation to students, the Harvard mathematician Benjamin Peirce said: "Gentlemen, that is surely true, it is absolutely paradoxical; we cannot understand it, and we don't know what it means, but we have proved it, and therefore, we know it must be the truth."

Euler's formula for polyhedra

The earliest known equation in **topology**. If *F* is the number of faces of a **polygon**, *E* the number of edges, and *V* the number of vertices, Euler's formula can be written as

$$F - E + V = 2$$

where F – E + V is known as the **Euler characteristic**. For example, the surface of a cube has six (square) faces, twelve edges, and eight vertices and, sure enough, 6 – 12 + 8 = 2.

even function

A function $f(x)$ such that $f(x) = f(-x)$ for all x.

evolute

The **locus** of the centers of curvature (the envelope) of a plane curve's normals. The original curve is then said to be the **involute** of its evolute. For example, the evolute of an **ellipse** is a **Lamé curve** and the evolute of a **tractrix** is a **catenary**.

excluded middle law

A law in (two-valued) logic which states that there is no third alternative to truth or falsehood. In other words, for any statement *A,* either *A* or not-*A* must be true and the other must be false. This law no longer holds in three-valued logic, in which "undecided" is a valid state, nor does it hold in **fuzzy logic**.

existence

A term that has several different meanings within **mathematics**. In the broadest sense there is the question of what it means for certain concepts, such as **pi**, to exist. Was π, for example, invented or discovered? In other words, does π exist only as an intellectual construct or was it somehow already "out there" waiting for people to find it. If it does exist independently of the human mind, when did its existence start? Does π predate the physical universe? Such ontological questions become even more difficult when applied to more complex or abstract mathematical concepts such as the **Mandelbrot set**, **surreal number**s, or **infinity**. A narrower and more technical type of "existence" in math is implied by an *existence theorem.* Such a theorem is used to prove that a number or other object with particular properties definitely exists, but does not necessarily give a specific example. Finally, there is existence in the sense of particular solutions to problems. If at least one solution can be determined for a given problem, a solution to that problem is said to exist. Something of the flavor of all three types of mathematical existence mentioned here are captured in the following anecdote:

> An engineer, a chemist, and a mathematician are staying in three adjoining cabins at an old motel. First the engineer's coffee-maker catches fire. He smells the smoke, wakes up, unplugs the coffee maker, throws it out the window, and goes back to sleep. Later that night the chemist smells smoke, too. He wakes up and sees that a cigarette butt has set the trash can on fire. He thinks to himself, "How does one put out a fire? One can reduce the temperature of the fuel below the flash point, isolate the burning material from oxygen, or both. This could be accomplished by applying water." So he picks up the trash can, puts it in the shower stall, turns on the water, and, when the fire is out, goes back to sleep. The mathematician has been watching all this out the window. So later, when he finds that his pipe ashes have set the bed sheet on fire, he's not in the least taken aback. "Aha!" he says, "A solution exists!" and goes back to sleep.

exponent

A number that gives the **power** to which a **base** is raised. For example, in 3^2 the base is 3 and the exponent is 2.

exponential

> *Who has not been amazed to learn that the function*
> $y = e^x$, *like a phoenix rising again from its own*
> *ashes, is its own derivative?*
>
> –Francois le Lionnais

Anything that grows at a rate proportional to its size is said to grow exponentially. The simplest form of the exponential function is just $y = e^x$, where ***e*** is about 2.712 . . . The exponential function to **base** *a* can be written as $f(x) = a^x$.

extrapolate

See **interpolate**.

extravagant number

See **economical number**.

F

face

A **polygon** bounding a **polyhedron**. A **cube**, for example, has six square faces. The plane angle formed by adjacent edges of a polygonal angle in space is called a *face angle.*

factor

Also known as a *divisor,* a number or variable that divides evenly into another number or algebraic expression. For example, the factors of 28 are 1, 2, 4, 7, 14, and 28. Although it is true that 28 is also divisible by the negative of each of these, "factors" is usually taken to mean only the positive divisors. *Factorization,* or *factoring,* is the decomposition of an object into a product of factors. For example, the number 15 factorizes into **prime numbers** as 3 × 5; and the **polynomial** $x^2 - 4$ factorizes as $(x - 2)(x + 2)$. The aim of factoring is usually to reduce something to basic building blocks, such as numbers to prime numbers, or polynomials to linear expressions. Factoring integers is covered by the **fundamental theorem of arithmetic** and factoring polynomials by the **fundamental theorem of algebra**. *Integer factorization* for large integers appears to be a difficult problem; there are no known methods for solving it quickly, and, for this reason, it has formed the basis of some public key cryptography algorithms.

factorial

The function, denoted *n!,* that is the product of the positive integers less than or equal to *n*. For example, $1! = 1$; $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$; $10! = 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 3{,}628{,}800$. 0! is defined to be 1, by working the relationship $n! = n \times (n - 1)!$ backward. An interesting equality is $1!\,10!\,22!\,1! = 11!\,0!\,2!\,21!$ in which the same digits are broken up two different ways into factorials. This may be the smallest such example. Factorials are important in **combinatorics** because there are *n*! different ways (permutations) of arranging *n* distinct objects in a sequence. They also turn up in formulas in calculus, for instance in *Taylor's theorem,* because the *n*th derivative of the function x^n is *n*!.

factorion

A natural number that equals of the sums of the **factorials** of its digits in a given **base**. The only known **decimal** factorions are $1 = 1!$, $2 = 2!$, $145 = 1! + 4! + 5!$, and $40{,}585 = 4! + 0! + 5! + 8! + 5!$.

Fadiman, Clifton (1904–1999)

An American essayist, literary critic, and noted intellectual who, among many other works, edited *Fantasia Mathematica*[96] and *The Mathematical Magpie.*[97] He became well known for the encyclopedic knowledge he displayed on the *Information Please* radio programs in the 1930s and '40s.

Fagnano's problem

In a given acute triangle *ABC,* find the inscribed triangle whose perimeter is as small as possible. The answer is the *orthic triangle* of *ABC,* that is, the triangle whose vertices are endpoints of the altitudes from each of the vertices of *ABC.* The problem was proposed and solved using calculus by Giovanni Fagnano (1715–1797) in 1775. Once the answer became known, several purely geometric solutions were also discovered.

fair division

See **cake-cutting**.

Farey sequence

A sequence of numbers named after the English geologist John Farey (1766–1826) who wrote about such sequences in an article called "On a curious property of vulgar fractions" in the *Philosophical Magazine* in 1816. Farey says that he noted the "curious property" while examining the tables of *Complete decimal quotients* produced by Henry Goodwin. To obtain the Farey sequence for a fixed number *n,* consider all **rational numbers** between 0 and 1 which, when expressed in their lowest terms, have a denominator (the number on the bottom of a fraction) not exceeding *n.* Write the sequence in ascending order of magnitude beginning with the smallest. The "curious property" is that each member of the sequence is equal to the rational number whose numerator (the number on top of a fraction) is the sum of the numerators of the fractions on either side, and whose denominator is the sum of the denominators of the fractions on either side. For example, the Farey sequence for $n = 5$ is ($\frac{0}{1}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{1}{1}$), from which it can be seen that $\frac{2}{5} = (1 + 1)/(3 + 2)$, $\frac{1}{3} = (1 + 2)/(4 + 5)$, $\frac{1}{2} = (2 + 3)/(5 + 5)$, $\frac{2}{3} = (3 + 3)/(5 + 4)$, and so forth. Farey wrote: "I am not acquainted whether this curious property of vulgar fractions has been before pointed out?; or whether it may admit of some easy or general demonstration?; which are points on which I should be glad to learn the sentiments of some of your mathematical readers."

One "mathematical reader" was Augustin **Cauchy**, who gave the necessary proof in his *Exercises de mathématique,* published in the same year as Farey's article. Farey was not the first to notice the property. C. Haros, in 1802, wrote a paper on the approximation of decimal fractions by common fractions. He explains how to construct what is in fact the Farey sequence for $n = 99$ and Farey's "curious property" is built into his construction.

Fechner, Gustav Theodor (1801–1887)

A German physicist and psychologist who studied aesthetic aspects of the **golden ratio** and published his findings in *Vorschule der Aesthetik* (Introduction to aesthetics) (1876), arguing that this ratio turns up commonly in human-made rectangular objects and is judged by people to be the most pleasing to the eye (though some later researchers have called his results into question).

Federov, E. S. (1853–1919)

A Russian geologist and crystallographer who helped lay the theoretical foundations for modern crystallography. In his famous two-part paper "Symmetry of Regular Systems of Figures" published in 1891, he proved that there are exactly 17 distinct symmetries in the **wallpaper group**.

feedback

The mutually **reciprocal** effect of one system or subsystem on another. *Negative feedback* is when two subsystems act to dampen the output of the other. For example, the relation of predators and prey can be described by a negative feedback loop since more predators lead to a decline in the population of prey, but when prey decrease too much so does the population of predators since they don't have enough food. *Positive feedback* means that two subsystems are amplifying each other's outputs, e.g., the screech heard in a public address system when the mike is too close to the speaker. The microphone amplifies the sound from the speaker which in turn amplifies the signal from the microphone, and so on. Feedback is a way of talking about the nonlinear interaction among the elements or components in a system and can be modeled by nonlinear differential or difference equations as well as by the activity of cells in a **cellular automaton** array.

Feigenbaum's constant

A universal constant, denoted δ, that governs the behavior of systems that are approaching **chaos**; it was discovered by the American mathematical physicist Mitchell Feigenbaum (1944–) in 1975 and has the value $\delta = 4.6692\ldots$. All one-dimensional chaotic systems have a behavior, as they approach instability, known as period doubling. The Feigenbaum constant gives the rate at which the period of the system doubles.

Fermat, Pierre de (1601–1665)

A French lawyer, magistrate, and gentleman scholar, often called the "Prince of Amateurs," who is best known for the **conjecture**, now proved, known as **Fermat's last theorem**. Although employed as a senior government official, Fermat somehow managed to find time to do an astonishing amount of math, for which he sought little acclaim or acknowledgment. In fact, he published only one important manuscript in his entire lifetime and even then used fake initials. When his fellow French mathematician Gilles Roberval offered to edit and publish some of his works, Fermat replied, "Whatever of my works is judged worthy of publication, I do not want my name to appear there." Most of his results are known through letters to friends, notes in book margins, and challenges to other mathematicians to find proofs for theorems he had devised.

Fermat was one of the founders, with René **Decartes**, of **analytical geometry** and, with Blaise **Pascal**, of **probability theory**. His work on the maxima and minima of curves and tangents to them was seen, by Isaac **Newton**, as a starting point for **calculus**. Yet his greatest love was for **number theory**. In 1640, while studying **perfect numbers**, Fermat wrote to Mersenne that if p is prime, then $2p$ divides $2^p - 2$. Shortly after he expanded this into what is now called **Fermat's little theorem**. As usual, Fermat stated "I would send you a proof, if I did not fear its being too long." His most famous statement of this form accompanied his hasty notes on the "last theorem." See also **Fermat number**.

Fermat number

A number defined by the formula $F_n = 2^{2^n} + 1$ and named after Pierre **Fermat** who conjectured, wrongly, that all such numbers would be **prime**. The first five Fermat numbers, $F_0 = 3$, $F_1 = 5$, $F_2 = 17$, $F_3 = 257$, and $F_4 = 65{,}537$, are prime. However, in 1732, Leonhard **Euler** discovered that 641 divides F_5. It takes only two trial divisions to find this factor because Euler showed that every factor of a Fermat number F_n with n greater than 2 has the form $k \times 2^{n+2} + 1$. In the case of F_5 this is $128k + 1$, so we would try 257 and 641 (129, 385, and 513 are not prime). It is likely that there are only finitely many Fermat primes. Gauss proved that a **regular polygon** of n sides can be inscribed in a circle with Euclidean methods (e.g., by straightedge and compass) if and only if n is a power of two times a product of distinct Fermat primes.

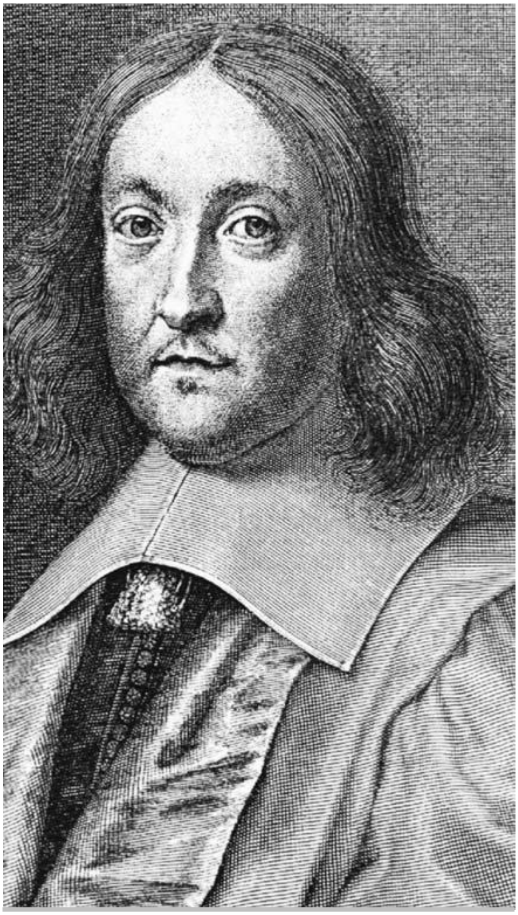

Fermat, Pierre de

Fermat's last theorem

A challenge for many long ages
Had baffled the savants and sages.
Yet at last came the light:
Seems old Fermat was right–
To the margin add 200 pages.

–Paul Chernoff

A conjecture put forward by Pierre de **Fermat** in 1637 in the form of a note scribbled in the margin of his copy of the ancient Greek text *Arithmetica* by **Diophantus**. The note was found after his death, and the original is now lost. However, a copy was included in the appendix to a book published by Fermat's son. Fermat's note read: "It is impossible to write a cube as a sum of two cubes, a fourth power as a sum of fourth powers, and, in general, any power beyond the second as a sum of two similar powers. For this, I have found a truly wonderful proof, but the margin is too small to contain it."

Fermat claimed that the **Diophantine equation** $x^n + y^n = z^n$ has no integer solutions for $n > 2$. It turns out he was right. But the proof had to wait 350 years and involved such advanced techniques, virtually none of which existed in the seventeenth century, that is seems very unlikely that Fermat really had found an elementary proof. Fermat's last theorem–now truly a theorem–was finally proved correct by Andrew **Wiles** in 1994.[353] In order to reach that dizzy height, however, Wiles had to draw on and extend several ideas at the core of modern mathematics. In particular, he tackled the *Shimura-Taniyama-Weil conjecture,* which provides links between the branches of mathematics known as **algebraic geometry** and **complex analysis**. This conjecture dates back to 1955, when it was published in Japanese as a research problem by the late Yutaka Taniyama. Goro Shimura of Princeton and Andre Weil of the Institute for Advanced Study provided key insights in formulating the conjecture, which proposes a special kind of equivalence between the mathematics of objects called elliptic curves and the mathematics of certain motions in space. Interestingly, the Wiles proof of Fermat's last theorem was a byproduct of his deep inroads into proving the Shimura-Taniyama-Weil conjecture. Now, the Wiles effort could help point the way to a general theory of three variable Diophantine equations. Historically, mathematicians have always had to state and solve such problems on a case-by-case basis. An overarching theory would represent a tremendous advance. See also **ABC conjecture**.

Fermat's little theorem

If P is a **prime number** then for any number a, $(a^P - a)$ must be divisible by P. This theorem is useful for testing if a number is *not* prime, though it can't tell if a number is prime. As usual, Pierre de **Fermat** didn't provide a proof (this time saying "I would send you the demonstration, if I did not fear its being too long"). Leonhard **Euler** first published a proof in 1736, but Gottfried **Leibniz** left virtually the same proof in an unpublished manuscript from sometime before 1683.

Fermat's spiral

A parabolic **spiral**.

Fibonacci (c. 1175–1250)

The pen name of Leonardo of Pisa, one of the greatest mathematicians of the Middle Ages. The son of a Pisan merchant who also served as a customs officer in North Africa, he traveled widely in Barbary (Algeria) and was later sent on business trips to Egypt, Syria, Greece, Sicily, and Provence. In 1200 he returned to Pisa and used the knowledge gained on his travels to write *Liber Abaci* (The book of the abacus), published in 1202, which introduced to western Europe the Hindu-Arabic numerals and **decimal** number system that remain in use today. The first chapter of Part 1 begins: "These are the nine figures of the Indians: 9 8 7 6 5 4 3 2 1. With these nine figures, and with this sign *0* which in Arabic is called zephirum, any number can be written, as will be demonstrated."

Fibonacci also showed he was capable of some amazing feats of calculation. For example, he found the positive solution of the **cubic** equation $x^3 + 2x^2 + 10x = 20$ using the Babylonian number system with base 60 (a strange choice, in view of his public advocacy of the decimal system!). He gave the result as 1, 22, 7, 42, 33, 4, 40 which is equivalent to

$$1 + \frac{22}{60} + \frac{7}{60^2} + \frac{42}{60^3} + \frac{33}{60^4} + \frac{4}{60^5} + \frac{40}{60^6}.$$

How on Earth he obtained this, nobody knows; it was 300 years before anyone else could obtain such accurate results. As well as serious mathematics, *Liber Abaci* contains many playful passages and it is for one of these, concerning a problem about counting the offspring of a pair of rabbits, that Fibonacci became best known after Edouard **Lucas** called the sequence of numbers discussed by the rabbit problem the **Fibonacci sequence**.

Fibonacci sequence

The sequence that arises in answer to this problem posed in **Fibonacci**'s great work *Liber Abaci:* "A certain man put a pair of rabbits in a place surrounded on all sides by a wall. How many pairs of rabbits can be produced from that pair in a year if it is supposed that every month each pair begins a new pair which from the second month on becomes productive?"

The number of pairs of rabbits in the *n*th month begins 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, . . . , where each term is the sum of the two terms preceding it. This sequence can be defined recursively as follows: $F(1) = F(2) = 1$, $F(n + 1) = F(n) + F(n - 1)$ for $n > 2$, where $F(n)$ is the *n*th Fibonacci number. Johannes **Kepler** was the first to point out that the growth rate of the Fibonacci numbers, that is, $F(n + 1) / F(n)$, converges to the **golden ratio**, ϕ (phi).

In the nineteenth century Fibonacci numbers were discovered in many natural forms. For example, many types of flower have a Fibonacci number of petals: certain types of daisies tend to have 34 or 55 petals, while sunflowers have 89 or, in some cases, 144. The seeds of sunflowers spiral outward both to the left and the right in a Fibonacci number of spirals. Similarly, the whorls on a pinecone, the numbers of rings on the trunks of palm trees, the patterns of snail shells, and the genealogy of the male bee all follow a sequence of Fibonacci numbers. The arrangement of plant leaves, or phyllotaxis, unfolds to the same pattern because this results in an optimal solution in terms of the spacing of the leaves or the amount of light that can reach them. A familiar spiral form, known as the **logarithmic spiral**, emerges when seeds on a plant grow and space themselves according the Fibonacci sequence. The logarithmic spiral is approximated by the rule: start at the origin of the Cartesian coordinate system, move $F(1)$ units to the right, move $F(2)$ units up, move $F(3)$ units to the left, move $F(4)$ units down, move $F(5)$ units to the right, and so on. By growing in this way, on structures such as sunflowers, pinecones, and pineapples, seeds are able to pack themselves together most efficiently.

Fibonacci numbers have so many interesting mathematical properties that an entire journal, *The Fibonacci Quarterly,* is devoted to them. The sequence of final digits in Fibonacci numbers repeats in cycles of 60. The last two digits repeat in 300, the last three in 1,500, the last four in 15,000, etc. The product of any four consecutive Fibonacci numbers is the area of a **Pythagorean triangle**. The shallow (least steep) diagonals of **Pascal's triangle** sum to Fibonacci numbers. Let *m* and *n* be positive integers, then

$F(n)$ divides $F(mn)$

$\gcd(F(n), F(m)) = F(\gcd(m, n))$, where "gcd" stands for "greatest common divisor."

$(F(n))^2 - F(n + 1)F(n - 1) = (-1)^{n-1}$.

$F(1) + F(3) + F(5) + \ldots + F(2n - 1) = F(2n)$.

For every *n*, there are *n* consecutive composite Fibonacci numbers.

An interesting use of the Fibonacci sequence is for converting miles to kilometers. For instance, if you want to know about how many kilometers 5 miles is, take the Fibonacci number (5) and look at the next one (8) (5 miles is about 8 kilometers). This works because it happens that the conversion factor between miles and kilometers is roughly equal to the Golden Ratio.

The first few Fibonacci numbers that are also **prime numbers** are 3; 5; 13; 89; 233; 1,597; 28,657; 514,229; . . . . It seems likely that there are infinitely many Fibonacci primes, but this has yet to be proven. However, it is relatively easy to show that for $n \geq 4$, $u_n + 1$ is never prime. The Fibonacci sequence is a special case of the **Lucas sequence**.

The *tribonacci series* is made by adding the last two digits: 1, 1, 2, 4, 7, 13, 24, 44, 81, . . . and from this the quadbonacci series, the pentbonacci series, and the

Fibonacci sequence The number of spirals of seeds on a sunflower is always a Fibonacci number—an arrangement that keeps the seeds uniformly packed no matter how large the seed head. *Thomas Stromberg*

hexbonacci series, all the way up to the *n*-bonacci series. Each ratio of successive terms forms a special constant, analogous to ϕ.

field

A number system in which addition, subtraction, multiplication, and division (except by zero) are always defined, and the **associative** and **distributive** laws are valid. For example, the set of **rational number**s is a field, whereas the set of **integer**s is not a field, because the result of dividing one integer by another is not necessarily an integer. The **real number**s also constitute a field, as do the **complex number**s. Compare with **ring**.

Fields Medal

By convention, the most prestigious award for research in mathematics. It is awarded every four years to between two and four mathematicians under the age of 40.

Fifteen Puzzle

A sliding-tile puzzle invented by Sam **Loyd** in the 1870s that became a worldwide obsession, much as **Rubik's cube** did a century later. Fifteen little tiles, numbered 1 to 15, were placed in a four by four frame in serial order except for tiles 14 and 15, which were swapped around; the lower right-hand square was left empty. The object of the puzzle was to get all the tiles in the correct order; the only allowed moves were sliding counters into the empty square. Everyone it seemed was caught up with the craze–playing the game in horse-drawn trams, during their lunch breaks, or when they were supposed to be working. The game even made its way into the solemn halls of the German parliament. "I can still visualize quite clearly the gray-haired people in the Reichstag intent on a square small box in their hands," recalled the geographer and mathematician Sigmund Gunter who was a deputy during the puzzle epidemic. "In Paris the

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Fairylite
15
PUZZLE
The Puzzle of the Age –
CAN YOU SOLVE IT?

Fifteen Puzzle A version of the Fifteen Puzzle produced in England by Fairylite. *Sue & Brian Young/Mr. Puzzle Australia, www.mrpuzzle.com.au*

puzzle flourished in the open air, in the boulevards, and proliferated speedily from the capital all over the provinces," wrote a contemporary French author. "There was hardly one country cottage where this spider hadn't made its nest lying in wait for a victim to flounder in its web." Loyd offered a $1,000 reward for the first correct solution. But, although many claimed it, none were able to reproduce a winning series of moves under close scrutiny. There is a simple reason for this, which is also the reason that Loyd was unable to obtain a U.S. patent for his invention. According to regulations, Loyd had to submit a working model so that a prototype batch could be manufactured from it. Having shown the game to a patent official, he was asked if it were solvable. "No," he replied. "It's mathematically impossible." Upon which the official reasoned there could be no working model and thus no patent!

The puzzle's theory reveals that the more than 20 billion possible starting arrangements of the tiles fall into just two groups: one in which all the tiles can be maneuvered into ascending numerical order (call this group I), and one in which tiles 14 and 15 will be inverted (group II). It's impossible to combine arrangements from these two groups and impossible to turn a group I arrangement into a group II, or vice versa, using the normal rules of the game. Given a random arrangement of tiles, can we know in advance if we have the unsolvable kind? Very easily. Simply count how many instances there are of a tile numbered n appearing after the tile numbered $n + 1$. If there are an even number of such inversions, the puzzle is solvable, otherwise you are wasting your time!

figurate number

A number sequence found by creating consecutive geometrical figures from arrangements of equally spaced points. Here is an example:

```
1   3   5   7
*   *   *   *
    **  *   *
        *** *
            ****
```

The points can be arranged in one, two, three, or more dimensions. There are many different kinds of figurate numbers, such as **polygonal number**s and **tetrahedral number**s.

films and plays involving mathematics

Mathematicians and mathematics rarely make an entrance on the silver screen or the stage, unless as parodies in the form of mad professors and meaningless scrawled equations. A notable exception is *A Beautiful Mind* (2001), directed by Ron Howard and starring Russell Crowe as the brilliant but mentally troubled mathematician John **Nash**. Although a fine love story and a well-crafted film, which won four Oscars, *A Beautiful Mind* is weak on math and inaccurate in many of its details of Nash's life and his battle with schizophrenia. *Rain Man* (1988), also based on a true story, costars Dustin Hoffman as an autistic savant with a photographic memory and a genius for mental arithmetic.

Good Will Hunting (1997), written by and starring Matt Damon and Ben Affleck, and also starring Robin Williams, is about a young man who has led a troubled life but has an amazing talent for mathematics. His abilities are discovered when he comes into conflict with the law, and he soon has to decide if he should pursue his mathematical future and leave his family and friends behind. In Darren Aronofsky's disturbing independent film *Pi* (1998), the main character is a mathematician obsessed with his search for patterns within **pi**'s infinite decimal places. He believes they can be used to predict chaotic behaviors, including that of the stock market. Throughout the film he is pursued by ruthless stock market players and by rabbis trying to find a mathematical way to communicate with God. In the science fiction film *Cube* (1997), six people awake to find themselves trapped in a deadly maze, and one of the characters uses mathematical skills to solve the puzzle and find a way to escape.

Lesser known films with strong mathematical themes include Mario Martone's *Death of a Neapolitan Mathematician;* Peter Greenaway's *Drowning by Numbers;* George Csicsery's *N Is a Number;* and *Moebius,* made by students and faculty at the Universidad del Cine of Buenos Aires. Mathematics has also found its way onto the stage. The musical *Fermat's Last Tango* (2000), a fictionalized account of Andrew **Wiles**'s struggle to prove **Fermat's last theorem**, was performed in New York by the York Theatre company. It followed the Pulitzer prize–winning play *Proof* by David Aubern, about the death of a brilliant mathematician and the repercussions for his daughters and his student.

finite

Limited in extent or scope. In mathematics, a *finite set* is such that the number of elements it contains can be described by a **natural number**. For instance, the set of integers between –18 and 5 is finite, because it has a natural number (17) of elements. The set of all **prime numbers**, on the other hand, is not finite. In physics, "finite" is used to mean both "not infinite" and "nonzero."

finite-state automaton (FSA)

The simplest computing device. Although it is not nearly powerful enough to perform universal computation, it can recognize **regular expression**s. FSAs are defined by a state transition table that specifies how the FSA moves from one state to another when presented with a particular input. FSAs can be drawn as **graph**s.

Fisher, Adrian

A British professional designer and constructor of **maze**s; his company, Adrian Fisher Maze Design, has built a huge variety of mazes in Britain, continental Europe, the United States, and elsewhere. These include the formal hedge maze at Leeds Castle and the largest brick pavement maze in the world at Kentwell Hall in Long Melford. The latter is based on a Tudor rose and has 15 sepals used as locations for a board game in which live players take part in Tudor costume.[101]

Fitchneal

An Irish version of the Viking game **Hnefa-Tafl**; played on a 7 × 7 board (as was the Scottish equivalent, known as *Ard-Ri,* "High King"), it is mentioned in the *Mabinogion* and *Cormac's Glossary* of the ninth century.

five

The length of the hypotenuse of the smallest **Pythagorean triangle** (a right triangle having integral sides). Five is the only **prime number** that is a member of two pairs of **twin primes.** Every integer is the sum of five positive or negative cubes in an infinite number of ways. Five is the smallest degree of a polynomial equation for which there is no general formula for the solutions (see **quintic**).

fixed-point attractor

An **attractor** that is represented by a particular point in **phase space**, sometimes called an *equilibrium point.* As a point it corresponds to a very limited range of possible behaviors of the system. For example, in the case of a pendulum, the fixed-point attractor represents the pendulum when the bob is at rest. This state of rest attracts the system because of gravity and friction.

Flatland: A Romance of Many Dimensions

A satirical novel by Edwin A. **Abbott**,[1] first published in 1884, that portrays a **two-dimensional world**, like the surface of a map, over which its inhabitants move. Flatlanders have no concept of up and down, and appear to each other as mere points or lines. From our three-dimensional perspective we can look down on Flatland and see that its people are "really" a variety of shapes, including straight lines (females), narrow isosceles triangles (soldiers and workmen), equilateral triangles (lower middle-class men), squares and pentagons (professional men, including the pseudonymous author of the tale, A. Square), hexagons and other regular polygons with still more sides (the nobility), and circles (priests). Abbott uses these geometrical distinctions, especially the appearance of Flatland females and the working class, as a commentary on the discrimination against women, the rigid class stratification, and the lack of tolerance for "irregularity" that was prevalent in Victorian Britain.

In a dream, A. Square visits the one-dimensional world of Lineland where he tries, unsuccessfully, to persuade the king that there is such a thing as a second dimension. In turn, the narrator is told of three-dimensional space by a sphere who moves slowly through the plane of Flatland, growing and shrinking as his cross-section changes in size. (If a **hypersphere** were to move through our three-dimensional world, we would see a sphere appear, grow to a maximum size, and then shrink again before disappearing.) Abbott is aware that he cheats a little in his description of what the inhabitants of Flatland actually see. In his preface to the second edition, he gives a lengthy but not-too-convincing reply to the objection, raised by some readers, that a Flatlander, "seeing a Line, sees something that must be *thick* to the eye as well as *long* to the eye (otherwise it would not be visible . . .)." The curious and often-neglected fact is that we are just as unable to imagine what it would truly be like to see in two dimensions as we are to conceive of four dimensions! No matter how hard we try we cannot imagine being able to see a line of *zero* thickness.

flexagon

A flat model constructed from a folded strip of paper, which, when flexed, can be made to reveal a number of hidden faces. Flexagons are amusing toys but they have also caught the interest of mathematicians. They are usually square or rectangular (tetraflexagons) or hexagonal (hexaflexagons). A prefix can be added to the name to indicate the number of faces that the model can display, including the two faces (back and front) that are visible before flexing. For example, a hexaflexagon with a total of six faces is called a hexahexaflexagon. The discovery of the first flexagon, a trihexaflexagon, is credited to the British student Arthur H. Stone who was studying at

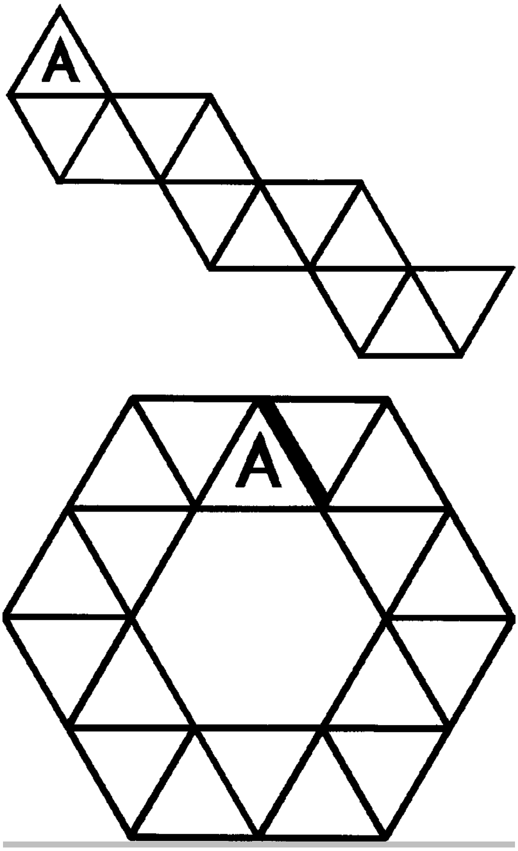

flexagon Two nets for folding into hexaflexagons: the 4-flexagon (top) and the 6-flexagon (bottom). To use the nets, photocopy and enlarge them, and label each side with these numbers:

4-flexagon $\begin{pmatrix}2231 & 2231 & 2231\\ 3144 & 3144 & 3144\end{pmatrix}$,

6-flexagon $\begin{pmatrix}662554 & 662554 & 662554\\ 231431 & 231431 & 231431\end{pmatrix}$.

Start at A using the top row of numbers. Number the other side of the net with the bottom row of numbers so that the top and bottom numbers appear on either side of the first triangle, and so on. *Jill Russell*

Princeton University in 1939. Stone's colleagues Bryant Tuckermann, Richard P. Feynman, and John W. Tukey became interested in the idea. Tuckerman worked out a topological method, called the *Tuckerman traverse,* for revealing all the faces of a flexagon. Tukey and Feynman developed a complete mathematical theory that has not been published. Flexagons were introduced to the general public by Martin **Gardner** writing in *Scientific American.*[110]

floor function

The greatest integer in *x,* that is, the largest integer less than or equal to *x.*

Flower of Life

One of the beautiful arrangements of circles found at the Temple of Osiris at Abydos, Egypt. The pattern also appears in Phoenician art from the ninth century B.C. The circles are placed with six-fold symmetry, forming a mesmerizing pattern of circles and lenses. A related design from the same temple, which recurs in Italian art from the thirteen century, is called the *Seed of Life.*

fly between trains problem

Two trains are approaching each another and a fly is buzzing back and forth between the two trains. Given the (constant) speed of the trains and their initial separation distance, and the (constant) speed of the fly, calculate how far the fly will travel before the trains collide. This problem appears to have been first posed by Charles Ange Laisant (1840–1921) in his *Initiation Mathématique.* There is a long-winded method of getting the answer and a much shorter way. Suppose the trains start out 200 miles apart and are each traveling at 50 miles per hour, and the fly–a speedster of its kind–is moving at 75 miles per hour. The long method involves considering the length of the back-and-forth path that the fly takes and evaluating this as the sum of an infinite series. The quick solution is to notice that the trains will collide in 2 hours and that in this time the fly will travel $2 \times 75 = 150$ miles! When this problem was put to John **von Neumann**, he immediately gave the correct answer. The poser, assuming he had spotted the shortcut, said: "It is very strange, but nearly everyone tries to sum the infinite series." Von Neumann replied: "What do you mean, strange? That's how I did it!"

focal chord

A **chord** of a **conic section** that passes through a **focus**.

focal radius

A line segment from the **focus** of an **ellipse** to a point on the **perimeter** of the ellipse.

focus

A defining point in the construction of a **conic section**. The word comes from the Latin for hearth or fireplace and appears to have been first used in mathematics, in describing an **ellipse**, by Johannes **Kepler**.

foliation

A decoration of a **manifold** in which the manifold is partitioned into sheets of some lower dimension, and the sheets are locally parallel. More technically, the foliated manifold is locally **homeomorphic** to a **vector space** decorated by cosets of a subspace.

folium

A curve, first described by Johannes **Kepler** in 1609, that corresponds to the general equation

$$(x^2 + y^2)(y^2 + x(x + b)) = 4axy^2, \text{ in Cartesian form, or}$$
$$r = -b \cos\theta + 4a \cos\theta \sin^2\theta, \text{ in polar coordinates.}$$

The Latin *folium* means "leaf-shaped." Three types, known as the simple folium, the bifolium (or double folium), and the trifolium, correspond to the cases when $b = 4a$, $b = 0$, and $b = a$, respectively. The *folium of Descartes* is given by the Cartesian equation $x^3 + y^3 = 3axy$ and was first discussed by René **Descartes** in 1638. Although he found the correct shape of the curve in the positive quadrant, he wrongly thought that this leaf shape was repeated in each quadrant like the four petals of a flower. The problem to determine the tangent to the curve was proposed to Gilles de Roberval who, having made the same incorrect assumption, called the curve *fleur de jasmin* after the four-petal jasmine bloom, a name that was later dropped. The folium of Descartes has an asymptote $x + y + a = 0$.

formal system

A mathematical formalism in which statements can be constructed and manipulated with logical rules. Some formal systems, such as **Euclidean geometry**, are built around a few basic **axiom**s and can be expanded with theorems that can be deduced through proofs.

formalism

A mathematical school of thought that was headed by the German mathematician David **Hilbert**. Formalists argue that mathematics must be developed through **axiom**atic systems. Formalists agree with Platonism on the principles of mathematical proof, but Hilbert's followers don't recognize an external world of mathematics. Formalists argue that mathematical objects don't exist until we define them. Humans create the **real number** system, for example, by establishing axioms to describe it. All that mathematics needs are inference rules to progress from one step to the next. Formalists tried to prove that within the framework of established axioms, theorems, and definitions, a mathematical system is consistent and, in the mid-twentieth century, formalism became the predominant philosophical attitude in math textbooks. However, it was undermined by **Gödel's incompleteness theorem** and also the general recognition that results can be usefully applied without having to be proved or derived axiomatically.

Fortune's conjecture

A conjecture about **prime number**s made by the New Zealander social anthropologist Reo Fortune (1903–1979), who had a reputation for unstable behavior bordering on the psychotic. He once attempted to settle an academic dispute with a colleague, Thomas McIlwraith, at the University of Ontario, by challenging him to a duel with any weapon of his choice from the collections of the Royal Ontario Museum. Fortune proposed that if q is the smallest prime greater than $P + 1$, where P is the product of the first n primes, then $q - P$ is prime. For example, if n is 3, then P is $2 \times 3 \times 5 = 30$, $q = 37$, and $q - P$ is the prime 7. These numbers, $q - P$, are now known as *Fortunate numbers*. The conjecture remains unproven but is generally thought to be true. The sequence of Fortunate numbers begins

$$3, 5, 7, 13, 23, 17, 19, 23, 37, 61, 67, \ldots$$

four

The smallest **composite number**, the second smallest **square number**, the first non-Fibonacci number (see **Fibonacci sequence**), the smallest **Smith number**, and the smallest number that can be written as the sum of two **prime number**s. Four is the number of dimensions that make up **space-time** (three of space and one of time). It is the most number of colors needed to color any map so that no two neighboring areas are the same color (see **four-color map problem**). There are four cardinal points on the compass, four Riders of the Apocalypse, and four Gospels.

four-color map problem

A long-standing problem that dates back to 1852 when Francis Guthrie, while trying to color a map of the counties of England noticed that four colors were enough to ensure that no adjacent counties were colored the same. He asked his brother Frederick if it was true that *any* map could be colored using four colors in such a way that adjacent regions (i.e., those sharing a common boundary segment, not just a point) receive different colors. Frederick Guthrie then passed on the conjecture to Augustus **de Morgan**. The first printed reference is due to Arthur **Cayley** in 1878. A year later the false "proof," by the English barrister Alfred Kempe, appeared; *its* incorrectness was pointed out by Percy Heawood 11 years later. Another failed proof is due to Peter **Tait** in 1880, a gap in *his* argument being pointed out by Julius Petersen in 1891. Both false proofs did have some value, though. Kempe discovered what became known as *Kempe chains,* and Tait found an equivalent formulation of the four-color theorem in terms of three-edge coloring.

The next major contribution came from George Birkhoff whose work allowed Philip Franklin in 1922 to prove that the four-color conjecture is true for maps with at most 25 regions. It was also used by other mathematicians to make various forms of progress on the four-color problem. In the 1970s, the German mathematician Heinrich Heesch developed the two main ingredients needed for the ultimate proof–reducibility and discharging. While the concept of reducibility was studied by other researchers as well, it seems that the idea of discharging, crucial for the unavoidability part of the proof, is due to Heesch, and that it was he who conjectured that a suitable development of this method would solve the four-color problem. This was confirmed by Kenneth Appel and Wolfgang Haken of the University of Illinois in 1977, when they published their proof of the four-color theorem.[12] Their controversial proof challenges the basic assumptions of what mathematical proof is. They used more than 1,200 hours of supercomputer time to analyze 1,478 different configurations that in turn can produce every possible map on a plane. Not everyone was happy with the method of the breakthrough, as Appel himself pointed out:

> For almost a century and a half, a Holy Grail of graph theory has been a simple incisive proof of the Four Color Theorem. It has troubled our profession that a problem that can be understood by a school child has yet to be solved in a way that better illuminates the reason that only four colors are needed for planar maps. The feelings of many mathematicians were summed up for me by Herb Wilf's response to being told that it appeared that one could prove the theorem by a long reducibility argument which used computers to test the reducibility of a large number of configurations. He simply said, "God would not allow such a beautiful theorem to have such an ugly proof."

Martin **Gardner** commented, "Whether a simple, elegant proof not requiring a computer will ever be found, is still an open question." It's interesting that such a simple, intuitive puzzle can be so difficult to settle! The four-color theorem is true for maps on a plane or on a sphere. The answer is different for geographic maps on a **torus**: in this case, seven colors are necessary and sufficient.[277]

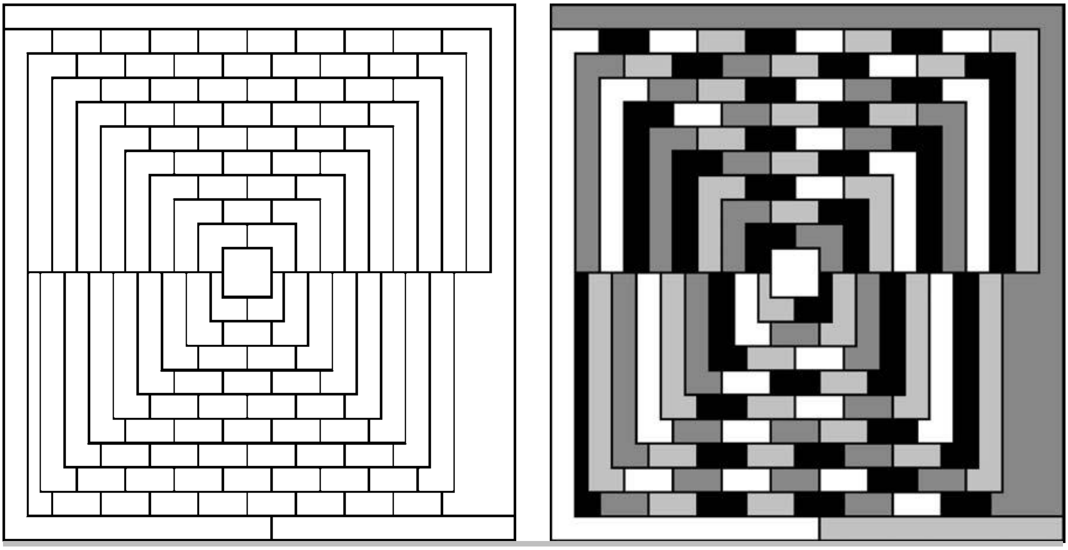

four-color map problem Martin Gardner's spoof counterexample to the four-color theorem (left) and a solution using four colors (right). The different colors are represented here in black, white, and gray.

four coins problem

Given three coins of possibly different sizes, which are arranged so that each is tangent to the other two, find the coin that is tangent to the other three coins. The solution is the inner **Soddy circle**.

four fours problem

Using arithmetic combinations of four 4's, express all the numbers from 1 to 100. For example, 1 = 44/44 and 2 = (4 × 4)/(4 + 4). The problem was first presented in *The Schoolmaster's Assistant: Being a Compendium of Arithmetic Both Practical and Theoretical* (first edition c. 1744), a popular textbook by the English schoolteacher and cleric Thomas Dilworth (d. 1780). Operations and symbols that are allowed include the four arithmetic operations (+, ×, −, /), concatenation (e.g., the use of 44), decimal points (e.g., 4.4), powers (e.g., 4^4), square roots, factorials (e.g., 4!), and overbars for repeating digits (e.g., .4 with an overbar to express $^4/_9$). Ordinary use of parentheses are allowed. One of the trickiest numbers to represent in this way is 73, which calls for something as contorted as

$$\sqrt{(\sqrt{(\sqrt{(4^4)})})} + 4 \,/\, .4'$$

(where .4′ is shorthand for .444 . . .). Of course, the problem can be extended to represent integers greater than 100. The highest value achievable in the four four's puzzle is $10^{8.0723047260281 \times 10153} = 4^{4^{4^4}}$.

four knights puzzle

On a 3 × 3 chessboard are two white knights at the top left-hand and top right-hand squares and two black knights at the bottom left-hand and bottom right-hand squares. The problem is to exchange the black knights with the white knights in the minimum possible number of moves. One move is a normal knight's move on any vacant cell of the board, which renders the center square inaccessible.

Fourier, (Jean Baptiste) Joseph, Baron (1768–1830)

A French mathematician known chiefly for his contribution to the mathematical analysis of heat flow. Although he trained for the priesthood, Fourier didn't take his vows but instead turned toward mathematics. He first studied and later taught mathematics at the newly created Ecole Normale. In 1798 he joined Napoleon's army in its invasion of Egypt as scientific advisor, to help establish educational facilities there and to carry out archaeological explorations. After his return to France in 1801 he was appointed prefect of the department of Isere. Fourier became famous for his *Theorie analytique de la Chaleur* (1822), a mathematical treatment of how heat conducts in solid bodies. He established the **partial differential equation** governing heat flow and solved it by using an infinite series of trigonometric functions, now known as **Fourier series**. Though these series had been used before, Fourier investigated them in much greater detail and prepared the way for later work on trigonometric series and the theory of functions of a real variable. Fourier's belief that his health would be improved by wrapping himself up in blankets proved fatal: thus encumbered he tripped down the stairs of his house and died.

Fourier series

Named after Joseph **Fourier**, the expansion of a **periodic** function as an infinite sum of sines and cosines of various frequencies and amplitudes. This is similar to the approximation of an irrational number by a sum of a series of rational numbers (or a decimal expansion). Human ears effectively produce Fourier series automatically from complex sounds. Tiny hairs, known as *cilia,* vibrate at different specific frequencies. When a wave enters the ear, the cilia vibrate if the wave function contains any component of the corresponding frequency. This enables the hearer to distinguish sounds of various pitches. Fourier series are used a great deal in science and engineering to find solutions to **partial differential equations**, such as those in problems involving heat flow. They can also be used to construct some pathological functions such as ones that are continuous but nowhere differentiable. The study and computation of Fourier series is known as **harmonic analysis**.

fourth dimension

"Do you think that there are things which you cannot understand, and yet which are; that some people see things that others cannot?" said Dr. Van Helsing in Bram Stoker's *Dracula.* Instead of vampires, he may just as easily have been talking about the fourth dimension–an extension at right-angles to the three familiar directions of up-down, forward-backward, and side-to-side. In physics, especially **relativity theory**, **time** is often regarded as the fourth dimension of the **space-time** continuum in which we live. But what meaning can be attached to a fourth *spatial* dimension? The mathematics of the fourth dimension (4-d) can be approached through a simple extension of either the algebra or the geometry of one, two, and three dimensions.

Algebraically, each point in a multidimensional space can be represented by a unique sequence of **real numbers**. One-dimensional space is just the **number line** of real numbers. Two-dimensional space, the plane, corresponds to the set of all **ordered pairs** (x, y) of real numbers, and three-dimensional space to the set of all ordered triplets (x, y, z). By extrapolation, four-dimensional space corresponds to the set of all ordered quadruplets (x, y, z, w). Linked to this concept is that of **quaternions**, which can also be viewed as points in the fourth dimension.

Geometric facts about the fourth dimension are just as easy to state. The fourth dimension can be thought of as a direction perpendicular to every direction in three-dimensional space; in other words, it stretches out along an axis, say the w-axis, that is mutually perpendicular to the familiar x-, y-, and z-axes. Analogous to the cube is a hypercube or **tesseract**, and to the sphere is a 4-d **hypersphere**. Just as there are five regular polygons, known as the **Platonic solids**, so there are six four-dimensional regular **polytopes**. They are: the 4-simplex (constructed from five tetrahedra, with three tetrahedra meeting at an edge); the tesseract (made from eight cubes, meeting three per edge); the 16-cell (made from 16 tetrahedra, meeting four per edge); the 24-cell (made from 24 octahedra, meeting three per edge); the 120-cell (made from 120 dodecahedra, meeting three per edge); and the monstrous 600-cell (made from 600 tetrahedra, meeting five per edge).

Geometers have no difficulty in analyzing, describing, and cataloging the properties of all sorts of 4-d figures. The problem starts when we try to *visualize* the fourth dimension. This is a bit like trying to form a mental picture of a color different from any of those in the known rainbow from red to violet, or a "lost chord," different from any that has ever been played. The best that most of us can hope for is to understand by analogy. For example, just as a sketch of a cube is a 2-d perspective of a real cube, so a real cube can be thought of as a perspective of a tesseract. At a movie, a 2-d picture represents a 3-d world, whereas if you were to watch the action live, in three-dimensions, this would be like a screen projection in four dimensions.

Many books have been written and schemes devised to nudge our imaginations into thinking four-dimensionally. One of the oldest and best is Edwin **Abbott**'s ***Flatland***[1] written more than a century ago, around the time that mathematical discussion of **higher dimensions** was becoming popular. H. G. Wells also dabbled in the fourth dimension, most notably in *The Time Machine* (1895), but also in *The Invisible Man* (1897), in which the central character drinks a potion "involving four dimensions," and in "The Plattner Story" (1896), in which the hero of the tale, Gottfried Plattner, is hurled into a four-spatial dimension by a school chemistry experiment that goes wrong and comes back with all his internal organs switched around from right to left.[285] The most extraordinary and protracted attack on the problem, however, came from Charles **Hinton**, who believed that, through appropriate mental practice involving a complicated set of colored blocks, a higher reality would reveal itself, "bring[ing] forward a complete system of four-dimensional thought [in] mechanics, science, and art."

Victorian-age spiritualists and mystics also latched on to the idea of the fourth dimension as a home for the spirits of the departed. This would explain, they argued, how ghosts could pass through walls, disappear and reappear at will, and see what was invisible to mere three-dimensional mortals. Some distinguished scientists lent their weight to these spiritualist claims, often after being duped by clever conjuring tricks. One such unfortunate was the astronomer Karl Friedrich Zöllner who wrote about the four-dimensional spirit world in his *Transcendental Physics* (1881) after attending séances by Henry Slade, the fraudulent American medium.

Art, too, became enraptured with the fourth dimension in the early twentieth century. When the Cubist painter and theorist Albert Gleizes said, "Beyond the three dimensions of Euclid we have added another, the fourth dimension, which is to say, the figuration of space, the measure of the infinite," he united math and art and brought together two major characteristics of the fourth dimension in early Modern Art theory–the geometric orientation as a higher spatial dimension and the metaphorical association with infinity.[159] See also **Klein bottle**.[25, 156, 163, 212, 213, 231, 254, 267, 271, 273, 340]

Fox and Geese

An English board game that dates back to the Middle Ages and is unusual in that the two sides are unequal, thus making this an example of a **Tafl game**. The lone fox attempts to capture 13 (or, in later versions, 17) geese, while the geese try to hem the fox in so that it can't move. The geese start out by filling up all the points on one side of the cross-shaped grid on the board. The fox–the one counter of a different color–begins on any vacant point remaining. The fox moves first. Each side, in its turn, may move one counter. Both fox and geese can move along a line, forwards, backwards, or sideways, to the next contiguous point. The fox may move along a line or jump over a goose to an empty point, capturing the goose and removing it from the board. Two or more geese may be captured by the fox in one turn, provided that he is able to jump to an empty point after each one. The fox wins if he depletes the gaggle of geese to a number that makes it impossible for them to trap him. The geese can't jump over the fox or capture the fox but instead must try to mob him and trap him in a corner. The geese win if they make it impossible for the fox to move. A modification of this game spread with the British to India, where during the Great Mutiny the game became known as "Officers and Sepoys." In this variant, two officers in a fort attempt to hold off 24 sepoys, who must storm the fort.

fractal

A geometric shape that can be subdivided at any scale into parts that are, at least approximately, reduced-size copies of the whole. The name "fractal," from the Latin *fractus* meaning a broken surface, was coined by Benoit **Mandelbrot** in 1975. The key property of fractals is **self-similarity**, which means that zooming in or zooming out of a fractal produces no overall change in appearance.

One of several technical definitions of a fractal is "a set of points whose topological dimension is less than its **Hausdorff dimension**." The topological dimension is an object's ordinary dimensionality–one in the case of a curve, two in the case of a surface, and so forth–and is always a whole number. The Hausdorff dimension, on the other hand, measures how much space an object fills, and can take non-integer values if the object is very complex and twisty.

Some fractals show a strong regularity and rigid self-similarity and are produced by the repeated application of a set of rules that may be quite simple. Among the best known of these "iterated function" systems are the **Koch snowflake**, the **Peano curve**s, the **Sierpinski carpet**, and the **Sierpinski gasket**. Other fractals, defined by a recurrence relation at each point in space, are among the most complex, beautiful, and beguiling mathematical structures known. They include the well-known **Mandelbrot set** and **Lyapunov fractal**s. Finally, there are random fractals generated by stochastic rather than deterministic processes, for example fractal landscapes. Random fractals have the greatest practical use, and can be used to describe many highly irregular real-world objects, including clouds, mountains, coastlines, and trees. See also **fractal dimension**.

fractal dimension

A non-integer measure of the irregularity or **complexity** of a system; it is an extension of the notion of dimension found in **Euclidean geometry**. Knowing the fractal dimension helps one determine the degree of irregularity and pinpoint the number of variables that are key to determining the dynamics of the system.

fraction

A number that represents a part, or several equal parts, of a whole; examples include one-half, two-thirds, and three-fifths. The word comes from the Latin *frangere,* meaning "to break." A *simple, common,* or *vulgar fraction* is of the form a/b, where a may be any integer and b may be any integer greater than 0. If $a < b$, the fraction is said to be *proper* ("bottom heavy"); otherwise it is *improper* ("top heavy"). A **decimal fraction** has a denominator (number on the bottom) of 10, 100, 1000, and so forth. See also **continued fraction**.

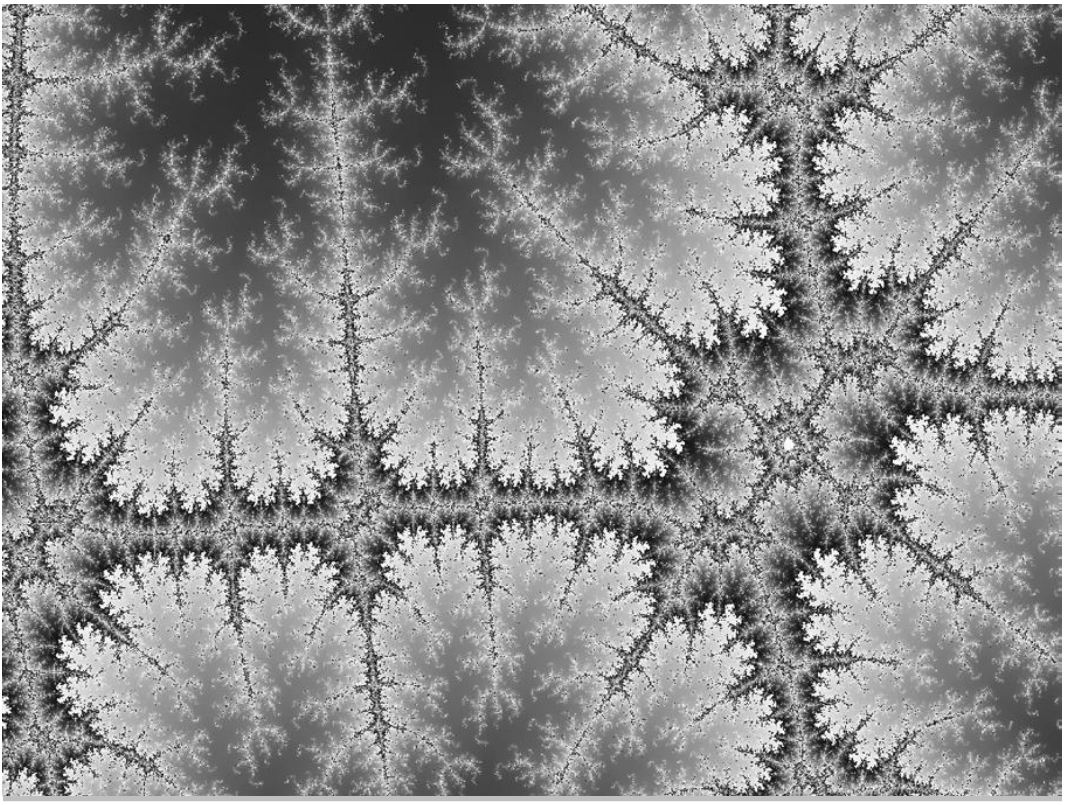

fractal A deep zoom of part of the Mandelbrot set. *Christopher Rowley*

Fraser spiral
A **distortion illusion** in which overlapping black arc segments appear to form a spiral but are in reality a series of overlapping concentric circles. This is easily demonstrated by following one of the curves with your finger. The illusion is named after the British psychologist James Fraser (1863–1936) who first published it in 1908.[104]

Fredholm, Erik Ivar (1866–1927)
A Swedish mathematician who founded the modern theory of **integral equation**s. This became a major research topic in the first quarter of the twentieth century and underpinned important theoretical developments in physics; David **Hilbert**, in particular, extended Fredholm's work to arrive at the concept of **Hilbert space**. Fredholm also devoted time to actuarial science and made a particularly important contribution by proposing an elegant formula to determine the surrender value of a life insurance policy. He earned his Ph.D. from the University of Uppsala but then spent the rest of his academic career at the University of Stockholm.

Freemish crate
See **impossible figure**.

Freeth's nephroid
See **nephroid**.

Frege, Friedrich Ludwig Gottlob (1848–1925)
A German mathematician and philosopher who virtually founded the modern discipline of mathematical logic. In *Die Grundlagen der Arithmetik* (The foundations of arithmetic, 1884), he used **set theory** to define the **cardinal number** of a given class as the class of all classes that are

similar (i.e. can be placed in a **one-to-one** correspondence) to the given class. In *Grundgesetze der Arithmetik* (The basic laws of arithmetic, 2 vols., 1893 and 1903), Frege began attempting to build up mathematics from arithmetic and symbolic logic on a rigorous and contradiction-free basis. When the second volume was in the process of being printed, Bernard **Russell** pointed out a paradox in Frege's work. The paradox, which became known as **Russell's paradox**, stems from the question: "Is the class of all classes that are not members of itself a member of itself or not?" The question leads to a contradiction and cannot be resolved. Frege was thus forced to admit that the foundation of his reasoning was worthless. As he stated at the end of his work, "A scientist can hardly encounter anything more undesirable than to have the foundation collapse just as the work is finished. I was put in this position by a letter from Mr. Bertrand Russell when the work was almost through the press."

Frénicle de Bessy, Bernard (1602–1675)

An eminent French amateur mathematician who extensively researched **magic squares**; his *Des quassez ou tables magiques,* published posthumously in 1693, first identified all 880 magic squares of the fourth order. Frénicle also corresponded with **Descartes**, **Fermat**, **Huygens**, and **Mersenne**, mostly on **number theory**, the work for which he is best known.

frequency

The number of times a value occurs in some time interval.

friendly number

See **amicable numbers**.

Frogs and Toads

A puzzle in which three counters or pegs representing frogs are placed on three successive positions on the left of a string of seven squares, and three different tokens representing toads are placed on the three rightmost squares. Frogs only move to the right, toads only to the left. Every move is either a slide to the adjacent square or a jump over one position, which is allowed only if the latter is occupied by a member of the other species. No two animals are ever allowed on the same square. The goal is to move the toads into the three leftmost positions and the frogs into three rightmost positions in the fewest possible moves. Many different versions of this puzzle have appeared over the centuries and it may be Arabic in origin. The number of pieces on each side may vary, as may the number of empty starting places in the middle; other names for the puzzle have included *Sheep and Goats* and *Sphinxes and Pyramids.*

frugal number

See **economical number**.

frustum

Part of a solid cut off between two parallel planes; in particular, for a **cone** or a **pyramid**, a frustum is determined by the plane of the base and a plane parallel to the base. *Frustum* is Latin for "a piece broken off."

function

> *Old mathematicians never die; they just lose some of their functions.*
>
> –Anonymous

A way of expressing the dependence of one quantity on another quantity or quantities. Traditionally, functions were specified as explicit rules or formulas that converted some input value (or values) into an output value. If f is the name of the function and x is a name for an input value, then $f(x)$ denotes the output value corresponding to x under the rule f. An input value is also called an *argument* of the function, and an output value is called a *value* of the function. The **graph** of the function f is the collection of all pairs $(x, f(x))$, where x is an argument of f. For example, the circumference C of a circle depends on its diameter d according to the formula $C = \pi d$; therefore, one can say that the circumference is a function of the diameter, and the functional relationship is given by $C(d) = \pi d$. Equally well, the diameter can be considered a function of the circumference, with the relationship given by $d(C) = d/\pi$. In modern mathematics, the insistence on specifying an explicit effective rule has been abandoned; all that is required is that a function f associate with every element of some **set** X a unique element of some set Y. This makes it possible to prove the existence of a function without necessarily being able to calculate its values explicitly. Also, it enables general properties of functions to be proved independently of their form. The set X of all admissible arguments is called the *domain* of f; the set Y of all admissible values is called the *codomain* of f. We write $f: X \to Y$.

fundamental group

A group of a **topological space** X that is constructed by looking at how closed paths in X can be combined to get new paths. Under a suitable way of identifying paths (known as **homotopy**) one can get a group structure on the set which gives an algebraic invariant of the space X.

fundamental theorem of algebra

The result that any **polynomial** with real or complex coefficients has a root in the complex plane.

fundamental theorem of arithmetic

Every positive **integer** greater than 1 is a **prime number** or can be expressed as a unique product of primes and powers of primes.

fuzzy logic

A departure from classical two-valued logic in which something is either true or false, to allow a continuous range of truth values. Fuzzy logic was introduced by Lotfi Zadeh of the University of California at Berkeley in the 1960s as a means to model the uncertainty of natural language.

G

Gabriel's horn

The **surface of revolution** of

$$y = 1/x$$

for x greater than 1. Surprisingly this has a finite volume of **pi** cubic units but an *infinitely large* surface area! Gabriel's horn is also known as *Torricelli's trumpet* because it was investigated by the Italian Evangelista Torricelli (1608–1647). As a young man Torricelli studied in Galileo's home at Arcetri, near Florence, and then, upon Galileo's death, succeeded his teacher as mathematician and philosopher for their good friend and patron, the grand duke of Tuscany. Torricelli was amazed by the strange property of his mathematical trumpet and tried various ways to avoid the conclusion that a finite volume could be enclosed by a vessel with an infinite surface area. Unfortunately, he lived before **calculus** came along to explain the apparent paradox in terms of **infinitesimals**.

Galois, Évariste (1811–1832)

A French mathematician who led a short, dramatic life and is often credited with founding modern **group** theory, though the Italian Paolo Ruffini (1765–1822) came up with many of the ideas first. Galois's work wasn't widely acknowledged by his contemporaries, partly because he didn't present his material very well and partly because he held unpopular political views. In fact, he was a republican revolutionary who was twice imprisoned because of his activities. During his second incarceration he fell in love with the daughter of the prison physician, Stephanie-Felice du Motel, and after being released, was killed in a duel over her with Perscheux d'Herbinville. His death started republican riots and rallies which lasted for several days. See also **Galois theory**.

Galois theory

The study of certain **group**s, known as *Galois groups,* that can be associated with **polynomial** equations. Whether

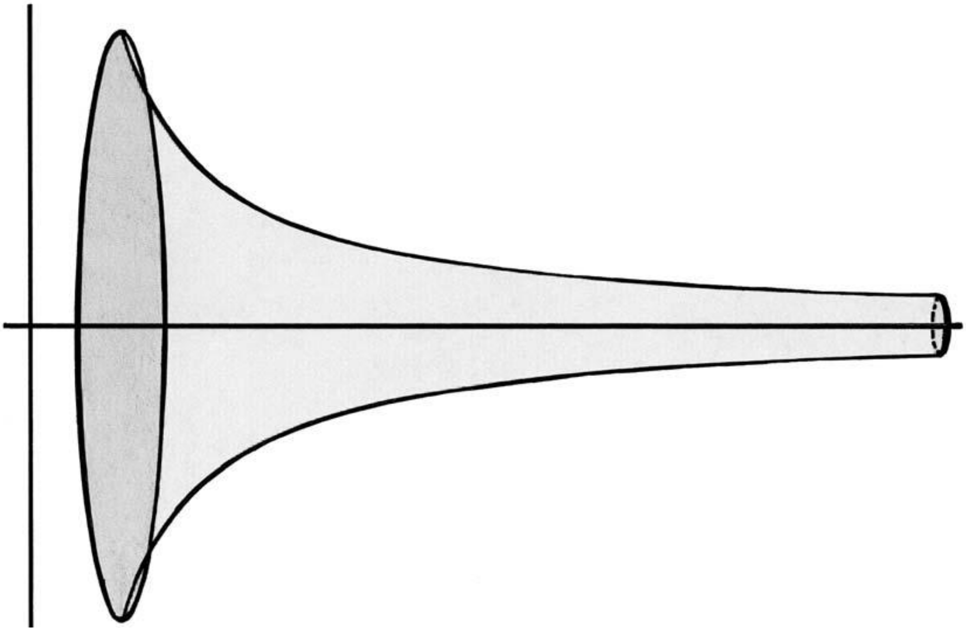

Gabriel's horn The horn for x values between 1 and 10.

or not the solutions to an equation can be written down using rational functions and square roots, cube roots, and so forth depends on certain group-theoretic properties of *Galois groups*.

game

A conflict, with formal rules and a finite number of choices of what to do at each stage, between two or more players. The study of games belongs to a branch of mathematics and logic known as **game theory**. If a game is simple enough, it can be solved for all possible outcomes. This is the case, for example, with **tic-tac-toe** and **Nim**. By harnessing the power of computers to check through vast numbers of moves, even more complicated games are succumbing to a complete analysis. In the 1990s, **nine men's morris** was shown, by searching through billions of possible endgames, to be a certain draw if both players work to an optimal strategy. Checkers may be the next to be fully determined: its roughly 500 million trillion possible positions may soon be within reach of the most powerful supercomputers. See also **blackjack**, **chess**, **Frogs and Toads**, **Ovid's game**, **TacTix**, and **Wythoff's game**.

game theory

A mathematical formalism used to study human **games**, economics, military conflicts, and biology. The goal of game theory is to find the optimal **strategy** for one player to use when his opponent also plays optimally. A strategy may incorporate randomness, in which case it is referred to as a **mixed strategy**.

Early ideas of game theory can be found in writings throughout history as diverse as the Bible and works by René **Descartes**, Sun Tzu (author of the 2,400-year-old *The Art of War*), and Charles Darwin. The basis of modern game theory is an outgrowth of several books that deal with related subjects such as economics and probability. These include Augustin Cournot's *Researches into the Mathematical Principles of the Theory of Wealth* (1838), which gives an intuitive explanation of what would eventually be formalized by John **Nash** as *Nash equilibrium;* Francis Edgeworth's *Mathematical Psychics,* which explored the notion of competitive equilibria in a two-type (or two-person) economy; and Emile **Borel**'s *Algebre et calcul des probabilites* (1927), which gave the first insight into so-called mixed strategies.[49] Game theory finally came of age through the efforts of two European immigrants to the United States working at the Institute of Advanced Studies in Princeton. Around 1940, the idea of the **utility function** was taken up by John **von Neumann**, who had been forced to flee his native Hungary when the Nazis invaded, and the economist Oskar Morgenstern (1902–1976), who had left Austria because he loathed the National Socialists. In Princeton the two immigrants worked together on what they initially thought would be a short paper on the theory of games, but that kept growing until it finally appeared in 1944 as an opus of 600 pages with the title *Theory of Games and Economic Behavior.*[230]

GLOSSARY OF GAME THEORY

categorical game	A game in which a tie is impossible.
finite game	A game in which each player has a finite number of moves and a finite number of choices at each move.
futile game	A game that allows a tie when played properly by both players.
impartial game	A game in which the possible moves are the same for each player in any position.
mixed strategy	A collection of moves together with a corresponding set of weights which are followed probabilistically in the playing of a game.
partisan game	A game for which each player has a different set of moves in any position.
payoff matrix	An $m \times n$ matrix that gives the possible outcome of a two-person zero-sum game when player A has m possible moves and player B has n moves.
strategy	A set of moves that a player plans to follow while playing a game.
zero-sum game	A game in which players make payments only to each other. One player's loss is the other player's gain, so the total amount of "money" available is constant.

gamma

See **Euler-Mascheroni constant**.

gamma function

A generalization of the **factorial** function to the real line and to the complex plane. It is defined by:

$$\Gamma(n+1) = \int_0^\infty x^n e^{-x}\,dx$$

If n is an integer, then $\Gamma(n+1) = n!$ See also **beta function**.

Gardner, Martin (1914–)

An American recreational mathematician best known for his "Mathematical Games" column, which ran in *Scien-*

tific American for 25 years. Through this column he introduced many subjects, including **flexagon**s, **polyomino**s, Piet Hein's **Soma Cube**, and John Conway's Game of **Life**, to a wider audience. He is also an accomplished amateur magician and an active member of the skeptical movement associated with James Randi. Gardner is the author of more than 60 books, including various collections of his *Scientific American* columns, *The Ambidextrous Universe,* and *The Annotated Alice.*[108–131]

gauge theory

A force field in nature, or an analogous **vector** field in mathematics with an enormous amount of **symmetry** that expresses the redundancy or ambiguity of many parameters. The simplest example is the *electromagnetic field.*

Gauss, Carl Friedrich (1777–1855)

A German mathematician, often called the "Prince of Mathematics," whose stature and range of interests rivaled those of **Aristotle** and Isaac **Newton**. Some inkling of what was in store came when, as a 3-year-old, he corrected a mistake in one of his father's lengthy payroll calculations. In school, at age 10, when his teacher gave the class the task of adding all the integers from 1 to 100, Gauss immediately wrote down the correct answer, 5050, on his slate. He had spotted that the numbers can be paired off as (100 + 1), (99 + 2), (98 + 3), . . . , (51 + 50) so that the problem reduces to multiplying 101 by 50. At age 19, Gauss found a way to construct a heptadecagon (a **regular polygon** with 17 sides) using only a straightedge and compass–a feat that had eluded the Greeks. Then Gauss entered the mathematical stratosphere of his time by proving what is now called the *fundamental theorem of algebra,* namely, that every **polynomial** has at last one root that is a **complex number**; in fact, he gave four different proofs, the first of which appeared in his dissertation. In 1801, he proved the *fundamental theorem of arithmetic* (that every natural number can be represented as the product of **prime number**s in only one way); published a brilliant tour de force on the properties of integers in his *Disquisitiones Arithmeticae,* which systematized the study of **number theory**; and showed that every number is the sum of at most three **triangular number**s. In the same year, he also developed the method of least squares fitting and, though he didn't publish it, used it to calculate the orbit of the asteroid Ceres, that had recently been discovered by Piazzi, from only three observations. Gauss published his monumental treatise on celestial mechanics *Theoria Motus* in 1806. He became interested in the compass through surveying, and developed the magnetometer, an instrument with which, together with Wilhelm Weber, he measured the intensity of magnetic forces. With Weber, he also built the first successful telegraph.

Unfortunately for mathematics, Gauss reworked and improved papers incessantly, and, in keeping with his motto "*pauca sed matura*" (few but ripe), he published only a fraction of his work. Many of his results were subsequently repeated by and attributed to others, since his terse diary remained unpublished for years after his death. Only 19 pages long, this diary later confirmed his priority on many breakthroughs, including work on an alternative to the **parallel postulate**, which really makes him the earliest pioneer of **non-Euclidean geometry** despite the fact that János **Bólyai** and Nikolai **Lobachevsky** are normally given this accolade. Gauss did, however, publish his seminal treatment on **differential geometry** in *Disquisitiones circa superticies curvas,* and Gaussian **curvature** is named for him. Gauss wanted a heptadecagon placed on his gravestone, but the carver refused, saying it would be indistinguishable from a circle. The heptadecagon appears in the shape of a pedestal with a statue erected in his honor in his hometown of Braunschweig.

Gaussian

Normally distributed (with a bell-shaped curve) and having a **mean** at the center of the curve with tail widths proportional to the **standard deviation** of the data about the mean.

Gelfond's theorem

Also known as the *Gelfond-Schneider theorem:* a^b is a **transcendental number** if (1) a is an **algebraic number** and not equal to either 0 or 1, and (2) b is algebraic and also an **irrational number**. Gelfond's theorem enables the seventh of David **Hilbert**'s famous problems to be solved.

general relativity

See **relativity theory**.

general topology

See **point-set topology**.

genetic algorithm

A type of evolving computer program, developed by the computer scientist John Holland, whose strategy of arriving at solutions is based on principles taken from genetics. Basically, the genetic **algorithm** utilizes the mixing of genetic information in sexual reproduction, random mutations, and natural selection at arriving at solutions.

genus

In **topology**, roughly speaking, the number of **hole**s in a surface. Spheres, bowling balls (the finger holes aren't true holes because they don't go all the way through), and wine glasses have a genus of 0 and can be represented by **quadratic** equations. Bagels, inner tubes, and teacups have a genus of 1 and can be described by **cubic** equations. Humans are more difficult to specify. However, you will certainly increase your genus by one if you have your ear pierced! Different definitions of genus apply to other types of mathematical objects such as a **curve**, a **knot**, or a **set**.

geoboard

A device commonly used in elementary schools to aid in the teaching of basic geometric concepts. A simple geoboard can be made from a square piece of wood and 25 nails arranged in an evenly spaced grid of 5 vertical lines and 5 horizontal lines. These nails represent the **lattice** points in the plane. Figures are made on the geoboard by stretching rubber bands from one nail to another until the desired shape is formed.

geodesic

A path on a given **surface** that is as straight as possible; in other words, a path that doesn't deviate either to the left or to the right, and only bends when forced to do so by the **curvature** (if any) of the surface. If the surface is an ordinary **plane**, the geodesics are straight lines; on a sphere the geodesics are **great circle**s.

geometric magic square

A square array of $n \times n$ cells each occupied by a distinct geometrical figure (or piece or tile), such that the n pieces contained in every row, column, and diagonal can be fitted together to produce (i.e., tile or pack) a constant shape known as the target. The figures may be of any dimension, but are normally planar (topological disks). In tessellating the target, which may be of any shape, planar pieces are allowed to be flipped. Pieces of three or more dimensions are considered distinct from their mirror images. Geometric **magic squares** using one-dimensional entries have been known for centuries; they are the traditional magic squares in which straight lines pave a constant length, as usually represented by numbers adding to a constant total. The properties of generalized geometric magic squares were first investigated by Lee **Sallows**.

geometric mean

The geometric mean of n numbers is the nth root of the product of the numbers.

geometric sequence

Also known as a *geometric progression,* a finite **sequence** of at least three numbers, or an infinite sequence, whose terms differ by a constant multiple, known as the *common ratio.* For example, starting with 3 and using a common ratio of 2 leads to the finite geometric sequence: 3, 6, 12, 24, 48, and also to the infinite sequence 3, 6, 12, 24, 48, . . . , (3×2^n). . . . In general, the terms of a geometric sequence have the form $a_n = ar^n$ ($n = 0, 1, 2, \ldots$) for fixed numbers a and r. If the terms of a geometric sequence are added together the result is a *geometric series.* If it is a finite series, then we add its terms to get the series sum, $S_n = a + ar + ar^2 + \ldots + ar^n = (a - ar^{n+1})/(1 - r)$. In the case of an **infinite series**, if $|r| < 1$, the sum is $a/(1 - r)$. If $|r| \geq 1$, however, the series diverges and thus has no sum. See also **arithmetic sequence**.

geometry

The study of the properties of shapes and of spaces. See also **Euclidean geometry** and **non-Euclidean geometry**.

geometry puzzles

One of the attractions of puzzles involving shapes, especially **dissection** problems, is that they appeal to the eye and very often don't call for much ability in solving equations and the like. Anyone can try to assemble the pieces of a jigsaw, whether it be of a picture or of a geometric shape, so a mathematical game such as **tangrams** or the **Soma cube** is within everyone's reach. On the other hand, some geometric puzzles call for a basic knowledge of more abstract fields such as algebra and calculus. They may also exploit our sometimes faulty intuition about how different quantities vary in one, two, and three dimensions and about how much information is needed to solve a problem.

As an example of faulty intuition, imagine that Earth, taken to be a perfect sphere with a radius r of 6,378 km, is completely covered by a thin membrane. Now suppose that 1 square meter is added to the area of this membrane to form a larger sphere. By how much does the radius and the volume of this membrane increase? This can be worked out from the formulae for the volume of a sphere ($V = (4/3)\pi r^3$) and the area of a sphere ($A = 4\pi r^2$), respectively. It turns out that if the area of the cover is increased by 1 square meter, then the volume it contains is increased by about 3.25 million cubic meters. This seems like a huge amount. However, the new cover wouldn't be very high above the surface of the planet–only about 6 nanometers! As an example of a problem that is both counterintuitive *and* seems to lack sufficient data for its solution see **hole-through-a-sphere problem**.

Gergonne point

In a triangle, the point at which the lines from the vertices (see **vertex**) to the points of contact of the opposite sides with the **inscribed circle** meet.

Germain, Sophie (1776–1831)

A French mathematician who made notable contributions to **number theory** and to mathematical physics despite a lack of formal training and the social prejudices of her day. She taught herself, against her parents' wishes, often at night, during the Reign of Terror following the French Revolution, with books from her father's library. When deprived of heat and light, she would wrap herself in quilts and use candles. Finally her parents acquiesced to her "incurable" passion for mathematics and let her study. Through Joseph **Lagrange**, to whom she had originally submitted work under a pseudonym, she gained access to a circle of distinguished mathematicians, including Carl **Gauss**. Among her most important work was an analysis of Ernst **Chladni**'s studies of vibrating surfaces, and her proof that if $x, y,$ and z are integers and if $x^5 + y^5 = z^5$ then at least one of $x, y,$ or z must be divisible by 5 (a result now known as *Germain's theorem*); this was an important early step towards proving **Fermat's last theorem**.

Get Off the Earth

A famous **vanishment puzzle** by Sam **Loyd**. The picture is made from a rectangular background topped with a circular card, representing the world, that can be rotated. Parts of a number of Chinese men are on each piece. With the world orientated so that the large arrow on it points to the N.E. point on the background, 13 Chinamen can be counted. But when the earth is turned slightly, so that the arrow points N.W., there are only 12 characters. Where did the thirteenth Chinaman go? The cleverness of the puzzle is that there are many bits of Chinamen–arms, legs, bodies, heads, and swords–and each has tiny slivers missing. When the earth is turned, these pieces get slightly rearranged. In particular, each of the 12 Chinamen gains a sliver of a Chinaman from his neighbor.

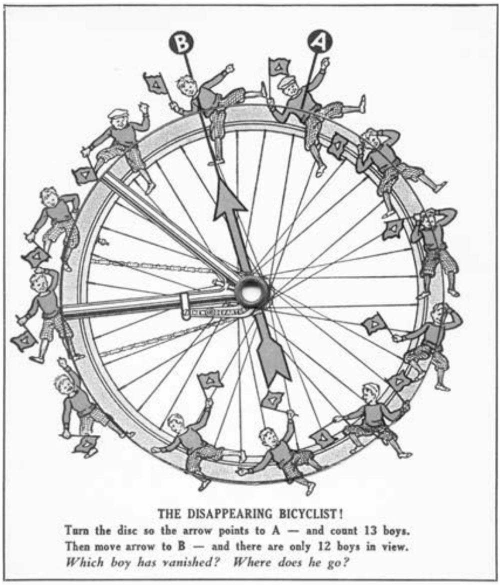
B
A
THE DISAPPEARING BICYCLIST!
Turn the disc so the arrow points to A — and count 13 boys.
Then move arrow to B — and there are only 12 boys in view.
Which boy has vanished? Where does he go?

Get Off the Earth An unusual variant of Loyd's Get Off the Earth puzzle called "The Disappearing Bicyclist." *From the collection of William Waite*

Gettier problem

A thought experiment in philosophy that throws into question the long-held supposition that to know something is equivalent to holding a belief about something that is both true and for which there is justification. Consider a case in which a lecturer has two students in her class called Mr. Havenot and Mr. Havegot. Mr. Havenot claims to own a Ferrari, drives one around, and has papers that state that the car is his. But in fact he does not actually own the car. Mr. Havegot, on the other hand, who shows no sign of Ferrari ownership, secretly has one of these rare cars. On the basis of the evidence, the teacher concludes that one of her students owns a Ferrari–and is correct in this belief. However, there is something wrong. Despite the combination of truth, justification, and belief, it seems that there is no real knowledge. The first examples of such problems were published in 1963 by the American philosopher Edmund Gettier (1927–).

Giant's Causeway

A natural structure that occurs on the coast in County Antrim, Ireland; it is one of the few places in the world where volcanic basalt has cooled in a columnar formation. The columns approximately form a **hexagon**al tessellation (see **tiling**) and tend to break off to produce a pavement with this pattern. The full length of the columns can't be seen, but it is estimated that they may be 20 feet (about 6 meters) high before merging into the underlying irregular basaltic mass. About 99% of the columns are believed to be hexagonal and only one triangular column is known. Though many of the hexagons are fairly regular, some have a side twice as long as their smallest side. Side lengths vary from 8 to 18 inches (20 to 46 cm) and the pillars break up into sections 6 to 36 inches (15 to 90 cm) long, with a concavo-convex junction rather than a plane junction. Other examples of such formations occur at Kirkjubaejarklaustri, in Iceland, and the Devil's Postpile, in California.

Gilbreath's conjecture

A strange hypothesis concerning **prime number**s that was first suggested in 1958 by the American mathematician and amateur magician Norman L. Gilbreath following

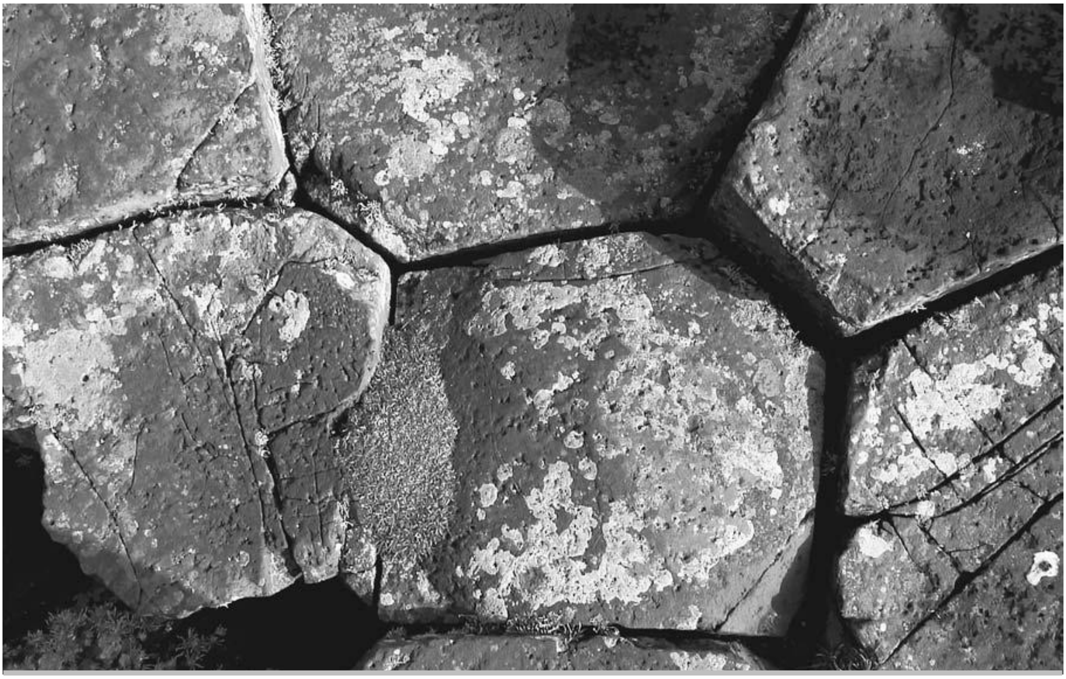

Giant's causeway Hexagonal paving on the Giant's Causeway. *Martin Melaugh, University of Ulster*

some doodlings on a napkin. Gilbreath started by writing down the first few primes:

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, \ldots$$

Under these he put their differences:

$$1, 2, 2, 4, 2, 4, 2, 4, 6, 2, \ldots$$

Under these he put the unsigned difference of the differences:

$$1, 0, 2, 2, 2, 2, 2, 2, 4, \ldots$$

And he continued this process of finding iterated differences:

$$\begin{aligned}&1, 2, 0, 0, 0, 0, 0, 2, \ldots\\ &1, 2, 0, 0, 0, 0, 2, \ldots\\ &1, 2, 0, 0, 0, 2, \ldots\\ &1, 2, 0, 0, 2, \ldots\\ &1, 2, 0, 2, \ldots\\ &1, 2, 2, \ldots\\ &1, 0, \ldots\\ &1, \ldots\end{aligned}$$

Gilbreath's conjecture is that, after the initial two rows, the numbers in the first column are all one. No exception has been found to date, despite searches out to several hundred billion rows, and the conjecture is generally assumed to be true. However, it may have nothing to do with primes as such. The English mathematician Hallard Croft has suggested the conjecture may apply to *any* sequence that begins with 2 and is followed by odd numbers that increase at a "reasonable" rate and with gaps of "reasonable" size. If this is the case, Gilbreath's conjecture may not be as mysterious as it first seems, though it may be very difficult to prove.

glissette

If there are two fixed curves, and a curve S of fixed shape and length that slides with its ends on the fixed curves, then the **locus** of a point moving with S is called a glissette. An example is the locus of the midpoint of a line segment sliding with its ends on two perpendicular lines; this locus is a circle.

gnomon magic square

A 3 × 3 array in which the elements in each 2 × 2 corner have the same sum. See also **magic square**.

Go

A two-player board game that originated in China in about 2000 B.C. It is often compared and contrasted with **chess**. Go is played on a board marked with 19 × 19 lines. Round, lens-shaped pieces called stones are placed, one per move, on the intersections of this grid, of which there are 361. As in chess, the pieces are colored black and white, but in Go black plays first. The board starts blank and pieces once played are not thereafter moved except to be taken off as prisoners. Pieces are captured singly or en masse by being surrounded so that they are not connected to any adjacent open intersection. The player with highest score at the end of the game, or following resignation or time expiry, wins. Go is considered to be a deeply strategic game, unlike chess, which is largely tactical. There are 32,940 opening moves, after symmetry is taken into account, 992 of which are deemed strong. Estimates of the number of possible board configurations vary but are typically on the order of 10^{174}.

God

> *I'm still an atheist, thank God.*
>
> –Luis Buñuel (Spanish film director, 1900–1983)

Mathematicians, logicians, and scientists have long debated the nature, existence, and dice-playing ability of a Higher Power. The pre-Renaissance French philosopher Jean Buridan (c. 1295–1358) used a version of the *liar paradox* to "prove" the existence of God. He wrote these two sentences:

God exists.

None of the sentences in this pair is true.

The only consistent way to have these two sentences be either true or false is for "God exists" to be true. (However, there is nothing to say that such consistency is necessary.) Blaise **Pascal** gave a more persuasive argument, not for the existence of God but for why we should believe in that existence: "If I believe in God and life after death and you do not, and if there is no God, we both lose when we die. However, if there is a God, you still lose and I gain everything." Pierre **Laplace**, on the other hand, replying to **Napoleon Bonaparte**, who had asked why his celestial mechanics made no mention of God, said: "Sir, I have no need of this hypothesis." The German mathematician Leopold **Kronecker** thought that "God made the Integers, all the rest is the work of man." In *The City of God,* however, Saint Augustine seems to imply that the integers were independent of God. He wrote: "**Six** is a number perfect in itself, and not because God created the world in six days; rather the contrary is true. God created the world in six days because this number is perfect, and it would remain perfect, even if the work of the six days did not exist." Augustine's statement can be taken to suggest that six would be a perfect number not only if the universe didn't exist, but even if God didn't exist. As to God's mathematical specialty, Plato said, "God ever geometrizes" while Charles Jacobi insisted that "God ever arithmetizes." James Jeans thought, "The Great Architect of the Universe now begins to appear as a pure mathematician," and Einstein ("God does not play dice") was sure He wasn't a probabilist.

Gödel, Kurt (1906–1978)

An Austrian-American mathematician and logician who, in 1931, proved that within a **formal system** questions exist that are neither provable nor disprovable on the basis of the **axiom**s that define the system. This is known as Gödel's undecidability theorem. He also showed that in a sufficiently rich formal system in which decidability of all questions is required, there will be contradictory statements. This is called **Gödel's incompleteness theorem**. In establishing these theorems Gödel showed that

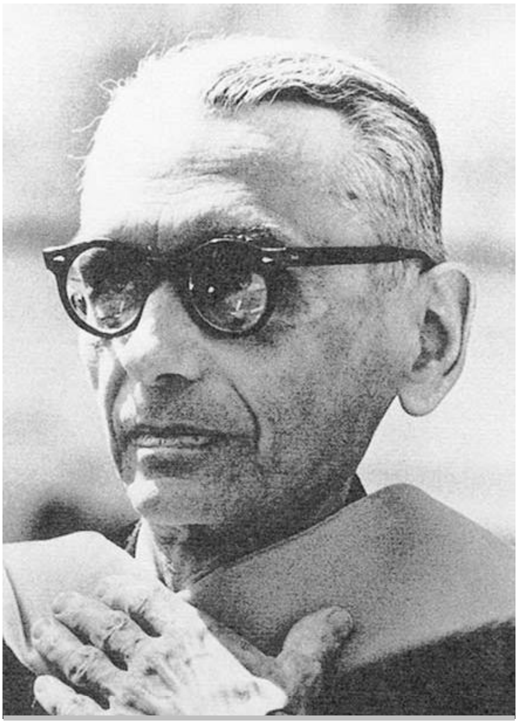

Gödel, Kurt

there are problems that can't be solved by any set of rules or procedures; instead, for these problems one must always extend the set of axioms. This disproved a common belief at the time that the different branches of mathematics could be integrated and placed on a single logical foundation. Gödel was a close friend of Albert Einstein at Princeton and contributed to his general **relativity theory** and cosmology. The so-called *Gödel universe* is a rotating model of the universe in which it is theoretically possible to travel into the past (see **time travel**).

Gödel's incompleteness theorem

In a nutshell: All consistent axiomatic systems contain undecidable propositions. What does this mean? An axiomatic system consists of some undefined terms, a number of **axiom**s that refer to those terms and partially describe their properties, and a rule or rules for deriving new propositions from already existing propositions. Axiomatic systems are powerful because they reduce large bodies of math to a simple description. Also, because they're very abstract, they allow all, and only, the results that follow from things having the formal properties specified by the axioms to be derived. An axiomatic system is *consistent* if, given the axioms and the derivation rules, it doesn't lead to any contradictions. One of the first modern axiomatic systems was a formalization of simple arithmetic (adding and multiplying whole numbers) by Giuseppe **Peano** and now known as *Peano arithmetic.* Kurt **Gödel** showed that every syntactically correct proposition in Peano arithmetic can be represented by a unique integer, called its *Gödel number.* The trick is to replace each symbol in the proposition, *including numerals,* by a different string of digits. If we represent "1" by 01, "2" by 02, "+" by 10, and "=" by 11, then the Gödel number of "1 + 1 = 2" is 0110011102. This allowed Gödel to write down, unambiguously, propositions about propositions. In particular, he was able to write down self-referential (see **self-referential sentence**) propositions–ones that include their own Gödel number. Gödel was then able to prove that, *either* the system of Peano arithmetic is inconsistent, *or* there are true propositions that can't be reached from the axioms by applying the derivation rules. The system is thus *incomplete,* and the truth of those propositions is *undecidable* (within that system). Such undecidable propositions are known as *Gödel propositions* or *Gödel sentences.* Nobody knows what the Gödel sentences for Peano arithmetic are, though people have their suspicions about the **Goldbach conjecture** (every even number is the sum of two prime numbers).

The results of an axiomatic system pertain to more than just Peano arithmetic, they apply to all kinds of things that satisfy the axioms. There are an immense number of other axiomatic systems, which either include Peano numbers among their basic entities or can be constructed from them. It follows that these systems, too, contain undecidable propositions, and are incomplete.

A common misconception is that Gödel's theorem imposes some profound limitation on knowledge, science, and mathematics. In the case of science, this ignores that Gödel's theorem applies to *deduction from axioms,* which is only one source of knowledge and not even a very common mode of reasoning in science. More generally, Gödel's incompleteness result doesn't touch directly on the most important sense of completeness and incompleteness, namely, descriptive completeness and incompleteness–the sense in which an axiom system describes a given field. In particular, the result represents no threat to the notion of truth.

Goldbach conjecture

One of the oldest and easiest-to-understand hypotheses in mathematics that remains unproven. In its original form, now known as the *weak Goldbach conjecture,* it was put forward by the Prussian amateur mathematician and historian Christian Goldbach (1690–1764) in a letter dated June 7, 1742, to Leonhard **Euler**. In this guise it says that every whole number greater than five is the sum of three **prime number**s. Euler restated this, in an equivalent form, as what is now called the *strong Goldbach conjecture* or, simply, the Goldbach conjecture: every even number greater than two is the sum of two primes. Thus, 4 = 2 + 2, 6 = 3 + 3, 8 = 3 + 5, 10 = 3 + 7, . . . , 100 = 53 + 47, ... In fact René **Descartes** knew about the two-prime version of Goldbach's conjecture before either Goldbach or Euler did. So, is it misnamed? Paul **Erdös** said, "It is better that the conjecture be named after Goldbach because, mathematically speaking, Descartes was infinitely rich and Goldbach was very poor." In any event, there is a much more important question, namely, is the conjecture true? The general assumption is that it is, but no one knows for sure. The most significant step toward a proof came in 1966 when the Chinese mathematician Chen Jing-Run showed that every sufficiently large even integer is the sum of a prime and a number that has at most two prime factors. Using powerful computers, the Goldbach conjecture has been checked out to about 400 trillion. But there is no great optimism among mathematicians that a final breakthrough is on the horizon. Even a reward of $1 million dollars for a proof offered by the publishing house Faber & Faber in 2000, to help publicize the novel *Uncle Petros and Goldbach's Conjecture* by the Greek mathematician and author Apostolos Doxiadis, went unclaimed.[85]

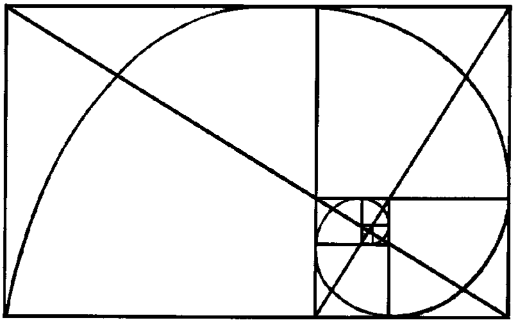

golden ratio A golden rectangle and a logarithmic spiral emerge from a whirling pattern of squares built up from two small squares of equal size at the spiral's center.

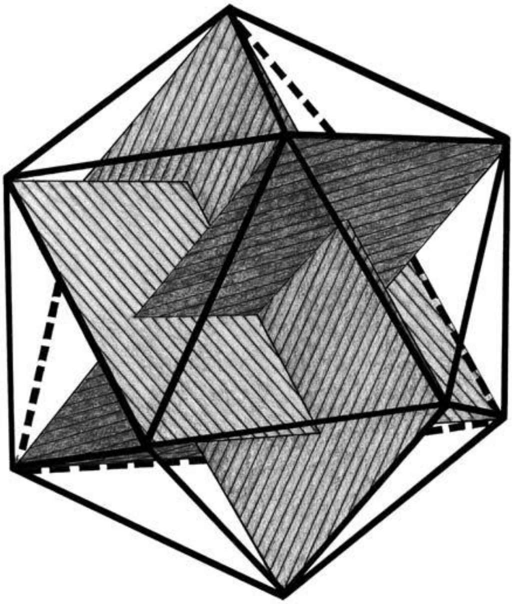

golden ratio The corners of an icosahedron meet the corners of three orthogonal golden rectangles.

golden ratio (phi, ϕ)

A remarkable number that, like **pi** and ***e***, pops up all over the place in mathematics but, in some ways, has a more "human" connection, in that it seems to be linked to aesthetics. Its name, which is also given as the *golden mean*, the *golden section*, the *golden number*, and the *divine proportion*, reflects this sense of a harmonious or pleasing ideal. The golden ratio is an **irrational number** of the type known as an **algebraic number** (in contrast with π and e, which are transcendental) and is represented by the Greek letter ϕ (phi). It can be defined in various ways. For example, it is the only number equal to its own reciprocal plus 1, that is, $\phi = (1/\phi) + 1$, so that $\phi^2 = \phi + 1$. From this comes the quadratic equation $\phi^2 - \phi - 1 = 0$ of which the golden ratio is the positive solution, $(1 + \sqrt{5}) / 2 \approx 1.6180339887\ldots$. The golden ratio is also approximated by the ratio of successive terms in the **Fibonacci sequence**; in fact, $F(n + 1) / F(n)$ gets closer and closer to ϕ as n tends to infinity. Because $1/(1 - \phi) = \phi$, the **continued fraction** representation of ϕ is

$$\phi = 1 + 1/(1 + 1/(1 + 1/(1 + 1/(1 + 1/\ldots$$
$$= [1;1,1,1,1,\ldots].$$

Two quantities are said to be in the golden ratio, if the ratio of the larger one, a, to the smaller one, b, is the same as the ratio of the smaller one to their difference, that is, $a/b = b/(a - b)$. The so-called *golden rectangle* is one whose sides a and b stand in the golden ratio. It is famously said to have great aesthetic appeal and is closely approximated by the dimensions of the front of the Parthenon in Rome. Leonardo da Vinci's masterpiece the *Mona Lisa* is said to have a face that is framed by a golden rectangle; what is certain is that Leonardo was a close personal friend of Luca **Pacioli**, who published a three-volume treatise on the golden ratio, *Divina Proportione*, in 1509. The Swiss-French architect and painter Le Corbusier designed an entire proportional system called the "Modulor," that was based on the golden ratio. The Modulor was supposed to provide a standardized system that would automatically confer harmonious proportions to everything, from door handles to high-rise buildings. Another artist who deliberately used the golden ratio is the surrealist Salvador Dali. The ratio of the dimensions of Dali's *Sacrament of the Last Supper* is equal to the golden ratio. Dali also incorporated in the painting a huge dodecahedron (a twelve-faced Platonic solid in which each side is a pentagon) engulfing the supper table. The dodecahedron, which according to Plato is the solid "which the god used for embroidering the constellations on the whole heaven," is intimately related to the golden ratio–both the surface area and the volume of a dodecahedron of unit edge length are simple functions of the golden ratio. In fact, ϕ turns up frequently in figures that have pentagonal symmetry. For instance the ratio of a regular pentagon's side and diagonal is equal to ϕ, and the vertices of a regular **icosahedron** are located on three orthogonal golden rectangles. The golden ratio is also related to **Penrose tiling** and to the **plastic number**.[205]

Golomb, Solomon W. (1932–)

A mathematician and electrical engineer at the University of Southern California who is best known for his seminal studies of **polyomino**s. His article "Checker Boards and Polyominos," published in the *American Mathematical Monthly* in 1954 when Golomb was a 22-year-old graduate student at Harvard, defined a polyomino as a simply connected set of squares (i.e., a set of squares joined along their edges). Golomb also originated the idea of **graceful graph**s.[137]

golygon

A series of straight-line segments that have lengths of one, two, three, and so on, up to some finite number of units, with the property that every segment connects at a right angle to the segment that is one unit larger, except the longest segment, which meets the shortest segment at a right angle. The name "golygon" was invented by Lee **Sallows**. Golygons have inspired some interesting puzzles as well as some intriguing research problems.

googol

A named coined in 1938 by Milton Sirotta, the 9-year-old nephew of the mathematician Edward **Kasner**, when the child was asked by his uncle to come up with a name for a very **large number**; at the same time, *googolplex* was suggested for a still larger number. Kasner defined these numbers as follows:

1 googol = 10^{100} (i.e., 1 followed by 100 zeros);
1 googolplex = $10^{\text{googol}} = 10^{10^{100}}$ (i.e., 1 followed by a googol number of zeros)

A googol is very roughly the number of years thought to be needed for all the black holes in the universe to evaporate by a process known as Hawking radiation. It is much larger than the number of protons and neutrons in the known universe (about 10^{80}), but much smaller than the number of protons and neutrons needed to pack every cubic centimeter of the known universe (about 10^{128}). The googolplex is the largest number with a proper name of which many people have heard. It is dwarfed, however, by such esoterica as **Graham's number**.

Gordian knot

The earliest reference to a **string puzzle**. In Greek mythology, a Phrygian peasant called Gordius, the father of Minos (see **maze**), became king because he was first to arrive in town after an oracle commanded the Phrygians to select as ruler the first person to drive into the public square in a wagon. In gratitude, Gordius dedicated his wagon to Zeus and placed it in the temple grove, tying the wagon pole to the yoke with a rope of bark. The knot was so intricately entwined that no one could undo it. A saying developed that whoever succeeded in untying the knot would become ruler of all Asia. Many tried, but all failed. According to legend, even Alexander the Great was unable to untie the Gordian knot, so he drew his sword and cut it through with a stroke. The expression "to cut the Gordian knot" is used to refer to a situation in which a difficult problem is solved by a quick and decisive action.

graceful graph

A **graph** of points and connecting lines that can be numbered in a certain way. Say the graph has p points and e lines ("e" for edges) connecting them. Each of the points is assigned an integer; the lowest integer (by convention) is taken to be 0, and no two integers may be alike. Each of the lines is labeled with the difference between the two integers of the points that it connects. Then, if the numbers corresponding with the lines run from 0 through e, the graph is said to be graceful. Graceful graphs were originally defined and developed by Solomon **Golomb**.

gradient

A **vector** of partial **derivative**s of a **function** that operates on vectors. Intuitively, the gradient represents the slope of a high-dimensional surface.

Graham, Ronald L. (1936–)

An American mathematician and leading combinatorialist after whom **Graham's number** is named. Graham is also one of the country's best jugglers and former president of the International Juggler's Association. In his youth, he and two friends were professional trampolinists who performed with a circus as the Bouncing Baers. His office ceiling is covered with a large net that he can lower and attach to his waist so that when he practices juggling with six or seven balls, any that are dropped will roll back to him. Graham is a professor in the department of computer science and engineering at the University of California at San Diego.

Graham's number

A stupendously **large number** that found its way in to the *Guinness Book of Records* as the biggest number ever obtained as part of a mathematical proof; it is named after its discoverer, Ronald **Graham**. Graham's number is the upper bound solution to a very exotic problem in **Ramsey theory**, namely: What is the smallest dimension n of a **hypercube** such that if the lines joining all pairs of corners are two-colored, a planar complete graph K_4 of one color will be forced? This is exactly equivalent to a problem that can be stated in plain language: Take any number of people, list every possible committee that can be formed from them, and consider every possible pair of committees. How many people must be in the original group so that no matter how the assignments are made, there will be four committees in which all the pairs fall in the same group, and all the people belong to an even number of committees. Graham's number is the *greatest* value that the answer could take. It is so large that it can only be written using special big-number notation, such as **Knuth's up-arrow notation**. Even then, it must be built up in stages. First, construct the number $G_1 = 3\uparrow\uparrow \ldots \uparrow\uparrow 3$, where there are $3\uparrow\uparrow\uparrow\uparrow 3$ up-arrows. This, in itself, is a number far, far beyond anyone's ability to even remotely comprehend. Next, construct $G_2 = 3\uparrow\uparrow \ldots \uparrow\uparrow 3$, where there are G_1 up-arrows; then construct $G_3 = 3\uparrow\uparrow \ldots \uparrow\uparrow 3$, where there are G_2 up-arrows; then continue this pattern until the number G_{62} has been made. Graham's number $G = 3\uparrow\uparrow \ldots \uparrow\uparrow 3$, where there are G_{62} up-arrows. While the unthinkably large upper boundary to the problem described earlier is given by Graham's number, nobody, including Graham himself, believes the solution is nearly so large. In fact, it is thought that the actual answer is probably 6![115]

grandfather paradox

One of the most powerful and commonly used arguments against **time travel**. It points out that if you were able to travel into the past you could (if you were so inclined) kill your grandfather when he was very young and thus render your own birth impossible. A simpler version is that you could kill a younger version of yourself so that you would not be alive in the future to travel back in time. The grandfather paradox shows how one form of time travel could violate **causality** by eliminating the cause of a phenomenon that has already taken place in the present.

The most bizarre adaptation of the grandfather paradox is found in Robert Heinlein's classic short story "All You Zombies." A baby girl is mysteriously left at an orphanage in Cleveland in 1945. "Jane" grows up lonely and dejected, not knowing who her parents are, until one day in 1963 she is strangely attracted to a drifter. She falls in love with him. But just when things are finally looking up for Jane, a series of disasters strike. First, she becomes pregnant by the drifter, who then disappears. Second, during the complicated delivery, doctors find that Jane has both sets of sex organs, and to save her life, they are forced to surgically convert "her" to a "him." Finally, a mysterious stranger kidnaps her baby from the delivery room. Reeling from these disasters, rejected by society, scorned by fate, "he" becomes a drunkard and drifter. Not only has Jane lost her parents and her lover, but he has lost his only child as well. Years later, in 1970, he stumbles into a lonely bar, called Pop's Place, and spills out his pathetic story to an elderly bartender. The bartender offers the drifter the chance to avenge the stranger who left her pregnant and abandoned, on the condition that he (Jane) join the "time travelers corps." Both of them enter a time machine, and the bartender drops off the drifter in 1963. The drifter is strangely attracted to a young orphan woman, who subsequently becomes pregnant. The bartender then goes forward nine months, kidnaps the baby girl from the hospital, and drops off the baby in an orphanage back in 1945. Then the bartender drops off the thoroughly confused drifter in 1985, to enlist in the time travelers corps. The drifter eventually gets his life together, becomes a respected and elderly member of the time travelers corps, and then disguises himself as a bartender and has his most difficult mission: a date with destiny, meeting a certain drifter at Pop's Place in 1970. The question is: who is Jane's mother, father, grandfather, grandmother, son, daughter, granddaughter, and grandson? The girl, the drifter, and the bartender, of course, are all the same person. As an exercise (on the road to insanity) try drawing Jane's family tree. You will find that not only is she her own mother and father, she is an entire family tree unto herself!

graph

> *I'll do algebra, I'll do trig, and I'll even do statistics,*
> *but graphing is where I draw the line!*
>
> –Anonymous

(1) In common usage, a plot of x values (the domain) against y values (the codomain) for a given **function**, $y = f(x)$. Such a graph is also known as a *function graph* or the *graph of a function.* (2) In strict mathematical usage, any set of dots, known as *nodes* or *vertices,* in which at least some pairs are joined by lines known as *edges* or *arcs.* What follows applies only to this second definition.

Often the lines on a graph are used to represent relationships between objects (represented by dots). Depending on the application, edges may or may not have a direction, as indicated by an arrow (see **directed graph**); edges joining a node to itself may or may not be allowed, and nodes and/or edges may be assigned weights. A *path* is a series of nodes such that each node is adjacent to both the preceding and succeeding node. A path is considered *simple* if none of the nodes in the path is repeated. The *length* of a path is the number of edges that the path uses, counting multiple edges multiple times. If it's possible to establish a path from any node to any other node of a graph, the graph is said to be a **connected graph**. A *circuit* or *cycle* is a path that begins and ends with the same node and has a length of at least two. A *tree* is a connected acyclic graph, that is, a graph without any circuits. A **complete graph** is one in which every node is adjacent to every other node. An **Euler path** in a graph is a path that uses each edge precisely once. If such a path exists, the graph is said to be *traversable.* An **Euler circuit** is a path that traverses each edge precisely once. A **Hamilton path**

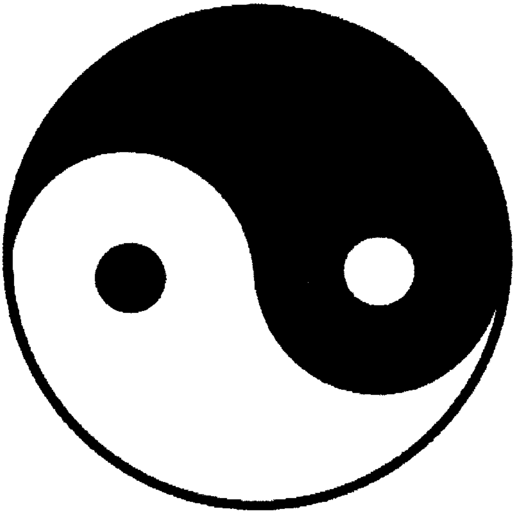

Great Monad

in a graph is a path that visits each node once and only once; a **Hamilton circuit** is a **circuit** that visits each node once and only once. Well known problems whose solution involves graphs and graph theory include the **four-color problem** and the **traveling salesman problem**.

graph theory

The study of **graph**s, either for their own sake, or as models of such diverse things as **group**s (in pure mathematics) or computer networks.

great circle

A circle that goes all the way around a **sphere** and is centered at the center of the sphere. The shortest route between two points on a sphere, such as Earth, is along the great circle that connects these points. A great circle is a **geodesic**.

Great Monad

Also known as *T'ai-Chi,* an important and ubiquitous symbol in traditional Chinese philosophy and cosmology. It represents the underlying harmony of the universe when its opposites or dualities–male and female (*yang* and *yin*), hard and soft, sun and moon, and so forth–are in balance. It occurs everywhere in Chinese art: in books, on walls, porcelain, tablets, and stitched into brocade.

greatest common divisor

The largest integer that divides each of a sequence of integers exactly. Also called the *greatest common factor.*

greatest lower bound

The largest **real number** that is smaller than each of the numbers in a **set** of real numbers.

Green, George (1793–1841)

An English mathematician who published work in the fields of hydrodynamics, electricity, and magnetism, but is best known for his theorem (see **Green's theorem**), which is the basis of **potential theory**. Green took over a bakery and adjoining windmill after the death of his father, but studied mathematics in his spare time. In 1828, he wrote his most important paper, "An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism," which, though generally overlooked at the time, is now regarded as the beginning of mathematical physics in England.

Green's theorem

A connection between path integrals over a well-connected region in the plane and the area of the region bounded in the plane. Green's theorem is a form of the *fundamental theorem of calculus,* and is used today in almost all computer codes that solve **partial differential equation**s.

Grelling's paradox

An equivalent, from the world of words and grammar, of **Russell's paradox**. Grelling's paradox involves dividing all adjectives into two sets: self-applicable and not self-applicable. Words like "English," "written," and "short" are self-applicable, while "Russian," "spoken," and "long" are not self-applicable. Now, define the adjective *heterological* to mean "not self-applicable." To which set of adjectives does "heterological" belong? This strange quandry was devised by the logician and philosopher Kurt Grelling (1886–1941/2), who was persecuted by the Nazis; it is not certain whether he died with his wife in the Auschwitz concentration camp in 1942, or whether he was killed in 1941 in the Pyrenees while trying to escape into Spain.

gross

A group of 144 items. The word comes from the Latin *grossus,* for "thick" or "large," via the Old French *gross douzaine* or "large dozen" (12 dozen), though this grouping may have started out in Germany. "Grocer" has the same origins as "gross" because a grocer is someone who deals in large quantities of food. A *great gross,* or a dozen gross, is 1728. See also **twelve**.

group

> *Wherever groups disclosed themselves, or could be introduced, simplicity crystallized out of comparative chaos.*
>
> –Eric Temple Bell

An abstract and crucially important way of representing **symmetry** and one of the most fundamental concepts in modern **algebra**. Groups were brought into mathematics in the early nineteenth century by the radical young French student Evariste **Galois** as a tool to help solve one of the outstanding problems of his day: to find a formula for solving **polynomial** equations of order five–**quintic**s–and higher. Galois showed, in notes scribbled down the night before he died in a duel, that no such formula exists. The reason for this is that the possible symmetries, or permutations, of the **root**s of fifth-degree polynomial equations are more complex than are the symmetries that can be represented by arithmetical formulas. This fact emerged from the development of the idea of a *permutation group* by Galois and, independently at about the same time, by Niels **Abel**. Half a century later, another Norwegian, Sophus **Lie**, showed how important groups are to the whole of mathematics. The theory of what became known as **Lie group**s links the discrete structure of permutations with the continuous variation of **differential equation**s. Not surprisingly, because group theory forms a common underpinning to algebra and to geometric features such as rotation, reflection, and symmetry, it crops up routinely in modern physics, from the classification of elementary particles to crystallography.

A group is a **set** whose elements are defined by a single operation. The group is called *additive* if the symbol for the operation is "+" and is called *multiplicative* if the symbol is "·" for multiplication. But any other symbol can be substituted for these. There is always a unique element (1, for multiplicative, and 0, for additive, groups) that leaves elements unchanged under the defined operation, like $a + 0 = a$. Also, for every element a there exists a unique inverse b such that, for example, in the case of the additive symbol, $a + b = 0$ and $b + a = 0$. Most often, however, the inverse is denoted as a^{-1}. Lastly, the group operation must be **associative** as in $a \cdot (b \cdot c) = (a \cdot b) \cdot c$. A group is *commutative* or *Abelian* if its operation is symmetric, as in $a + b = b + a$.

Groups come in two types: finite and infinite. The symmetry group of the roots of a polynomial equation is a finite group, because there is only a limited number of permutations possible among the roots of a given polynomial. In contrast, the Lie groups that represent symmetries of solutions of differential equations are infinite because they represent continuous transformations, and continuity carries the potential of an infinite number of changes. Finite groups can be built up from combinations of smaller groups by a process analogous to multiplication. In the same way that a whole number can be written as a product of **prime number**s, a finite group can be expressed as a combination of certain factors known as *simple groups.* Most simple groups belong to one of three families: the *cyclic groups,* the *alternating groups,* or the *groups of Lie type.* Cyclic groups consist of cyclic permutations of a prime number of objects. Alternating groups consist of even permutations–those formed by interchanging the positions of two objects an even number of times. Sixteen subfamilies make up the simple groups of Lie type, each associated with a family of infinite Lie groups. (Confusingly, a Lie group is not a group of Lie type, since the former is infinite and the latter is finite!) Altogether, there are 18 specific families of finite simple groups. There are also 26 simple groups, known as *sporadic groups,* that are highly irregular and fall outside these families. Five sporadic groups were found in the nineteenth century by Emile Mathieu. Then came a hiatus until the 1960s, when suddenly a rush of new sporadics came to light. The most remarkable of these is the so-called **monster group**, which appears to be intimately related to the structure of the universe at the subatomic level.

Grundy's game

See **Nim**.

Guy, Richard Kenneth (1916–)

A British-born mathematician who is professor emeritus of mathematics at the University of Calgary, Canada, and an expert in **combinatorics** and in **number theory**. Guy is the author of more than 250 papers and 10 books, including (as a coauthor) the game theory classic *Winning Ways.* He has been an editor of the "Problems" section of the *American Mathematical Monthly* since 1971.

H

Haberdasher's puzzle

The greatest mathematical discovery of Henry **Dudeney**, it was first published in the *Weekly Dispatch in* 1902 and then as problem no. 26 in his *The Canterbury Puzzles* (1907).[87] One must decide how to cut an equilateral triangle into four pieces that can be rearranged to make a square. The accompanying diagram shows the solution, which Dudeney describes as follows:

> Bisect AB in D and BC in E; produce the line AE to F making EF equal to EB; bisect AF in G and describe arc AHF; produce EB to H, and EH is the length of the side of the required square; from E with distance EH, describe the arc HJ, and make JK equal to BE; now from the points D and K drop perpendiculars on EJ at L and M.

A remarkable feature of the solution is that each of the pieces can be hinged at one vertex, forming a chain that can be folded into the square or the original triangle. Two of the hinges bisect sides of the triangle, while the third hinge and the corner of the large piece on the base cut the base in the approximate ratio 0.982:2:1.018. Dudeney showed just such a model of the solution, made of polished mahogany with brass hinges, at a meeting of the Royal Society on May 17, 1905.

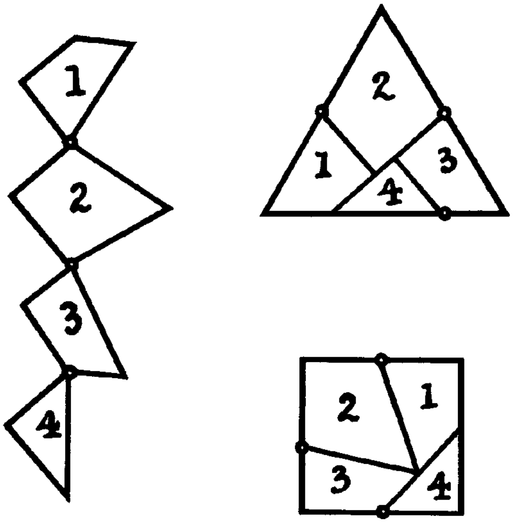

Haberdasher's puzzle The puzzle and its solution as illustrated by Henry Dudeney.

Hadwiger problem

In *d* dimensions, define $L(d)$ to be the largest integer *n* for which a cube cannot be cut into *n* cubes (not necessarily different). The Hadwiger problem is to find $L(d)$. Definite solutions are only known in two and three dimensions: $L(2) = 5$ and $L(3) = 47$. However, it is known that $L(4) \leq 853$ and $L(5) \leq 1{,}890$, and it is considered likely that $L(d)$ is odd for all values of *d.* See also **dissection**.

hailstone sequence

A sequence of numbers produced by the rules of the **Collatz problem**; in other words, a sequence formed in the following way: Start with any positive integer *n.* (1) If *n* is even, divide it by 2; if *n* is odd, multiply it by 3 and add 1. (2) If the result is not 1, repeat step (1) with the new number. For $n = 5$, this produces the sequence 5, 16, 8, 4, 2, 1, 4, 2, 1, ... For $n = 11$, the resulting sequence is 11, 34, 17, 52, 26, 13, 40, 20, 10, 5, 16, 8, 4, 2, 1, 4, 2, 1, ... The name "hailstone" comes from the fact that the numbers in these sequences rise and fall like hailstones in a cloud before finally falling to Earth. It seems from experiment that such a sequence will always eventually end in the repeating cycle 4, 2, 1, 4, 2, 1, . . . , but some values for *n* generate many values before the repeating cycle begins. An unsolved mystery is whether all such sequences eventually hit 1 (and then 4, 2, 1, 4, 2, 1, . . .) or whether there are some sequences that never settle down to a repeating cycle.

hairy ball theorem

If a sphere is covered with hair or fur, like a tennis ball, the hair cannot be brushed so that it lies flat at every point. In mathematical terms: any continuous tangent vector field on the sphere must have a point where the vector is zero. This theorem also means that somewhere on Earth's surface there has to be a point where the horizontal wind speed is zero, even if it's windy everywhere else. Does the same apply to a **torus**? Is there a hairy donut theorem? No! The number of "problem points," where the hair would stick up on a surface, is related to a quantity called the **Euler characteristic** of that surface. Basically, every point on a surface has an index that describes how many times the vector field rotates in a neighborhood of the problem point. The sum of the indices of all the vector fields is the Euler characteristic. Since the torus has Euler number 0, it is possible to have a covering of hair–a vector field–on it that lies flat at every point.

half-line

A **ray**.

half-plane

The part of a **plane** that lies on one side of a given line.

halting problem

Given a program and inputs for it, decide whether it will run forever or will eventually stop. This is not the same thing as actually running a given program and seeing what happens. The halting problem asks whether there is any *general prescription* for deciding how long to run an arbitrary program so that its halting or non-halting will be revealed. In a celebrated 1936 paper,[337] Alan **Turing** proved that the halting problem is undecidable: there's no way to construct an **algorithm** that is *always* able to determine whether another algorithm halts or not. From this it follows that there can't be an algorithm that decides whether a given statement about **natural numbers** is true or not. The undecidability of the halting problem provides an alternative proof of **Gödel's incompleteness theorem**. This is because if there were a complete and consistent axiomatization of all true statements about natural numbers, then we would be able to create a set of rules that decides whether such a statement is true or not. Another amazing consequence of the undecidability of the halting problem is *Rice's theorem,* which states that the truth of *any* nontrivial statement about the **function** that is defined by an algorithm is undecidable. So, for example, the decision problem "will this algorithm halt for the empty string" is already undecidable. Note that this theorem holds for the *function defined by the algorithm* and not the algorithm itself. It is, for example, quite possible to decide if an algorithm will halt within 100 steps, but this isn't a statement about the function that is defined by the algorithm. Many problems can be shown to be undecidable by reducing them to the halting problem. However, Gregory **Chaitin** has given an undecidable problem in algorithmic information theory that doesn't depend on the halting problem.

While Turing's proof shows that there can be no general method or algorithm to determine whether algorithms halt, individual instances of that problem may very well be susceptible to attack. Given a specific algorithm, one can often show that it must halt, and in fact computer scientists often do just that as part of a correctness proof. But every such proof requires new arguments: there is no *mechanical, general way* to determine whether algorithms halt. And there's another caveat. The undecidability of the halting problem relies on the fact that computers are assumed to have a memory of potentially infinite size. If the memory and external storage of a machine is limited, as it is for any real computer, then the halting problem for programs running on that machine can be solved with a general algorithm (albeit an extremely inefficient one).

ham sandwich theorem

Given a sandwich in which bread, ham, and cheese (three finite volumes) are mixed up, in any way at all, there is always a flat slice of a knife (a plane) that bisects each of the ham, bread, and cheese. In other words, however messed up the sandwich–even if it's been in a blender–you can always slice through it in such a way that the two halves have exactly equal amounts, by volume, of the three ingredients. This theorem generalizes to higher-dimensional ham sandwiches, when it essentially becomes the **Borsuk-Ulam theorem**: in *n*-dimensional space in which there are *n* globs of positive volume, there is always a hyperplane that cuts all the globs exactly in half.

Hamilton, William Rowan (1805–1865)

An Irish mathematician who, among other things, invented **quaternions** and a new theory of dynamics. Having excelled in Greek and mathematical physics at Trinity College, Cambridge, Hamilton was appointed Astronomer Royal of Ireland; in this position he served from 1827 to his death and, during all that time, lived in Dunsink Observatory, Dunsink Lane, to the northwest of Dublin. However, he quickly lost interest in staying up at nights to make observations–he hired three of his sisters to help run the place–and preferred instead to write poetry (badly). He was friends with Samuel Coleridge, who introduced him to the philosophy of Kant, which had a great influence on him, and with William Wordsworth, who advised him against writing any more poems.

Hamilton did early work on **caustic** curves and was led from this to his discovery of the *law of least action,* which enabled many physical problems to be expressed more elegantly. One of his greatest triumphs was his treatment of **complex numbers** as pairs of real numbers, an approach that finally exorcised long-standing suspicions about the reality of **imaginary numbers**, and helped clear the way for other algebras. From this he was led to consider ordered quartets of numbers, which he called quaternions. The idea for quaternions came to Hamilton suddenly on October 16, 1843, while he was standing on Brougham ("Broom") Bridge, where Broombridge Street crosses the Royal Canal, Dublin. A commemorative plaque under the bridge, on the towpath, was unveiled by the Taoiseach (head of the Irish parliament), Eamon De Valera, on November 13, 1958. Of his invention, Hamilton wrote:

> The quaternion was born, as a curious offspring of a quaternion of parents, say of geometry, algebra, metaphysics, and poetry. . . . I have never been able to give a clearer statement of their nature and their aim than I have done in two lines of a sonnet addressed to Sir John Herschel:
>
> "And how the One of Time, of Space the Three
> Might in the Chain of Symbols girdled be."

Hamilton's interest in complex numbers was stimulated by his friend and compatriot John Graves, who pointed Hamilton in the direction of John Warren's *A Treatise on the Geometrical Representation of the Square Root of Negative Quantities.* This book explained the concept of the complex plane, which Hamilton turned from geometry into algebra. One of Hamilton's last inventions was a curiosity called the *icosian calculus,* which was another outcome of his friendship with Graves. After a visit to the latter's house, Hamilton wrote: "Conceive me shut up and revelling for a fortnight in John Graves' Paradise of Books! of which he has really an astonishingly extensive collection, especially in the curious and mathematical kinds. Such new works from the Continent he has picked up! and such rare old ones too!" Graves posed some puzzles to Hamilton, and either Graves or his books got Hamilton to thinking about regular polyhedra. When Hamilton returned to Dublin he thought about the symmetry group of the **icosahedron**, and used it to invent an algebra he called the "icosians" and also a game called the **Icosian game**. The only complete example of this game, inscribed to Graves, is now in the keeping of the Royal Irish Academy, of which Hamilton was the president from 1837 to 1847. (In early 1996, a second example of the Icosian game came to light but only included the board.)

In some ways, Hamilton was too far ahead of his time. The operator now referred to as the *Hamiltonian* and the so-called *Hamilton-Jacobi equation* that relates waves and particles only became important when **quantum mechanics** came along, and Felix **Klein** introduced Wernher Schrödinger, the father of wave mechanics, to Hamilton's work.

Hamilton's personal life was not always happy. He fell deeply in love with a woman named Catherine Disney, who was forced by her parents to marry a wealthy man 15 years older than her. Hamilton remained hopelessly in love with her the rest of his life, though he eventually married someone else. He became an alcoholic, then foreswore drink, then relapsed. Many years after their early romance, Catherine began a secret correspondence with Hamilton. Her husband became suspicious, and she attempted suicide by taking laudanum. Five years later, she became seriously ill. Hamilton visited her and gave her a copy of his *Lectures on Quaternions.* They kissed at last, and she died two weeks later. He carried her picture with him ever afterward and talked about her to anyone who would listen.[149]

Hamilton circuit

A **Hamilton path** that starts and ends at the same vertex. See also **traveling salesman problem**.

Hamilton path

Named after William **Hamilton**, a path that traverses every vertex of a **connected graph** once and only once. The problem of the **knight's tour** is equivalent to finding a Hamilton path (or, in the case of a reentrant tour, a **Hamilton circuit**) that corresponds to the legal moves of the knight. Compare with **Euler path**.

Hankel matrix

A **matrix** in which all the elements are the same along any **diagonal** that slopes from northeast to southwest.

happy number

If you iterate the process of summing the squares of the decimal digits of a number and if this process terminates in 1, then the original number is called a happy number. For example 7 → (7^2) 49 → ($4^2 + 9^2$) 97 → ($9^2 + 7^2$) 130 → ($1^2 + 3^2$) 10 → 1. See also **amicable number**.

Hardy, Godfrey Harold (1877–1947)

One of the most prominent English mathematicians of the twentieth century; his legendary collaboration with John Littlewood lasted 35 years and produced nearly 100 papers. Hardy was a precocious child, whose tricks included factorizing hymn numbers during sermons. In 1919, he became Savilian Professor of Geometry at Oxford but returned to Cambridge in 1931 as professor of pure mathematics. His work was mainly in **analysis** and **number theory**.

Hardy had only one other passion in his life–the game of cricket. His daily routine would begin with reading *The Times* and studying the cricket scores over breakfast. Then he would do mathematical research from 9 o'clock till 1 o'clock. After a light lunch, he would walk down to the university cricket ground to watch a game. In the late afternoon he would walk slowly back to his rooms at the college, and take dinner followed by a glass of wine. Hardy was known for his eccentricities. He couldn't stand having his photo taken and only five snapshots of him are known to exist. He also hated mirrors and his first action on entering any hotel room was to cover any mirror with a towel.

Hardy's book *A Mathematician's Apology* (1940)[151] is one of the most vivid descriptions of how a mathematician thinks and the pleasure of mathematics. But the book is more, as C. P. Snow writes:

> *A Mathematician's Apology* is . . . a book of haunting sadness. Yes, it is witty and sharp with intellectual high spirits: yes, the crystalline clarity and candor are still there: yes, it is the testament of a creative artist. But it is also, in an understated stoical fashion, a passionate lament for creative powers that used to be and that will never come again. I know nothing like it in the language: partly because most people with the literary gift to express such a lament don't come to feel it: it is very rare for a writer to realise, with the finality of truth, that he is absolutely finished.

See also **Ramanujan**.

harmonic analysis

The method of expressing **periodic** functions as sums of sines and cosines.

harmonic division

The division of a line segment by two points such that it is divided externally and internally in the same ratio.

harmonic mean

The harmonic mean of two numbers a and b is $2ab/(a+b)$.

harmonic sequence

The sequence: 1, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, $\frac{1}{5}$ . . . . Added together, these become the terms of the *harmonic series:* $1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \ldots$. This series diverges (has no finite sum), though very slowly–a result first proved by the French philosopher and theologian, Nichole d' Oresme (c. 1325–1382). In fact, it still diverges if you take away every other term, and even if you take away nine out of every ten terms. However, if you take the sum of reciprocals of all natural numbers that do not contain the number nine (when written in decimal expansion) the series converges! To show this, group the terms based on the number of digits in their denominator. There are eight terms in $(\frac{1}{1} + \ldots + \frac{1}{8})$, each of which is no larger than 1. Consider the next group $(\frac{1}{10} + \ldots + \frac{1}{88})$. The number of terms is *at most* the number of ways to choose two ordered digits out of the digits 0 . . . 8, and each such term is clearly no larger than $\frac{1}{10}$. So this group's sum is no larger than $9^2/10$. Similarly, the sum of the terms in $(\frac{1}{100} + \ldots + \frac{1}{999})$ is at most $9^3/10^2$, etc. So the entire sum is no larger than

$$9 \times 1 + 9 \times (\tfrac{9}{10}) + 9 \times (9^2/10^2) + \ldots + 9 \times (9^n/10^n) + \ldots$$

This is a geometric series that converges. Thus by the comparison test, the original sum (which is smaller term-by-term) must converge.

Harshad number

A number that is divisible by the sum of its own digits; also known as a Niven number. For example, 1,729 is a Harshad number because 1 + 7 + 2 + 9 = 19 and 1,729 = 19 × 91. A *Harshad amicable pair* is an amicable pair (m, n) such that both m and n are Harshad numbers (see **amicable numbers**). For example, 2,620 and 2,924 are a Harshad amicable pair because 2,620 is divisible by 2 + 6 + 2 + 0 = 10 and 2,924 is divisible by 2 + 9 + 2 + 4 = 17 (2,924/17 = 172). There are 192 Harshad amicable pairs in the first 5,000 amicable pairs.

hat problem

A team of three contestants, Alice, Bob, and Cedric, enter a room and a hat is placed on each one's head so that he or she can't see it. The color of each hat is based on a coin toss–blue (B) for heads, red (R) for tails. After all the contestants enter the room, they look at the colors of one anothers' hats and, based on this information, they guess the color of their own hat. Each can guess red or blue, or, if she can't make up her mind, she can pass. No communication is allowed during the competition, but the players are allowed to agree on a strategy before play begins. The team wins if at least one of them guesses correctly, and none of them guesses incorrectly. What is the team's best strategy? At first sight, it may seem as if no effective strategy is possible beyond each contestant guessing his or her own hat color. In fact, this is the very worst approach since, to succeed, it requires that everyone guess correctly and the probability of this is only $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$. A far better plan is for the contestants to agree that two of them will pass while the third takes a stab at the color of his own hat. Then the odds improve to one in two. Beyond this it's hard to see any way that the probability of success could be increased. Yet there is an even better strategy. The key is to realize that there are only two cases (RRR and BBB) where everyone's hat is the same color but *six* cases where two hats are the same color and the other hat is a different color (RRB, RBR, BRR, BBR, BRB, and RBB). This suggests the following strategy for members of the team: *if you see two hats of opposite colors, pass. If you see two hats of one color, guess that your hat is the other color.* If everyone's hat is the same color, all players on the team will guess wrong and the team will lose. But the chance of this happening is only $\frac{2}{8}$ (= $\frac{1}{4}$). In every other possible case, the odd person out will guess correctly and their teammates will pass, so the team will win. This strategy wins $\frac{6}{8}$ (= $\frac{3}{4}$) of the time, and can't be improved upon. Since half of each player's guesses will be wrong, it's impossible to do better than a strategy in which each player in turn guesses correctly alone three times out of four, and the fourth time all guess wrong.

What if there are more players in the team? Say the number of players is *n*. By the reasoning described previously, it's clear that the team can't hope to win more than $n/(n + 1)$ of the time. Yet it isn't obvious that it can do this well. Having more people, it seems, might make it harder for them to synchronize their wrong guesses. However, it turns out that, if the number of people in the team is *one less than a power of 2,* this best-possible value can be achieved. For example, with a team of 7, the team can win $^7/_8$ of the time, and with a team of 15, they can win $^{15}/_{16}$ of the time. The strategy involved is complicated but is closely linked to *Hamming codes* (see **coding theory**), which are a method of encoding and transmitting information so that even if a small number of errors occur during transmission, the original information can be entirely recovered. For teams of other sizes, such as 9, 10, or 13, mathematicians have yet to find an optimal strategy or establish what proportion of the time the team can be expected to win.

Hausdorff, Felix (1868–1942)

A German mathematician who is considered to be one of the founders of modern **topology** and who also did significant work in **set theory** and functional analysis. Among several concepts named after him is the **Hausdorff dimension**, which gives a way of assigning a fractional dimension to a curve or shape. Hausdorff also published philosophical and literary works under the pseudonym "Paul Mongré." He studied at Leipzig and taught mathematics there until 1910, when he became professor of mathematics at Bonn. When the Nazis came to power, Hausdorff, a Jew, felt that as a respected university professor he would be safe from persecution. However, his abstract mathematics was denounced as useless and "un-German" and he lost his position in 1935. He sent his daughter to Britain but stayed with his wife in Germany. In 1942, when he could no longer avoid being sent to a concentration camp, he committed suicide together with his wife and sister-in-law.

Hausdorff dimension

A way to accurately measure the **dimension** of complicated **set**s such as **fractal**s. The Hausdorff dimension, named after Felix **Hausdorff**, coincides with the more familiar notion of dimension in the case of well-behaved sets. For example, a straight line or an ordinary curve, such as a circle, has a Hausdorff dimension of 1; any **countable set** has a Hausdorff dimension of 0; and an *n*-dimensional **Euclidean space** has a Hausdorff dimension of *n*. But a Hausdorff dimension is not always a natural number. Think about a line that twists in such a complicated way that it starts to fill up the plane. Its Hausdorff dimension increases beyond 1 and takes on values that get closer and closer to 2. The same idea of ascribing a fractional dimension applies to a plane that contorts more and more in the third dimension: its Hausdorff dimension gets closer and closer to 3. As a specific example, the fractal known as the **Sierpinski carpet** has a Hausdorff dimension of just over 1.89.

Heesch number

The maximum number of times that a closed plane figure–a tile–can be completely surrounded by copies of itself. The Heesch number of a triangle, quadrilateral, regular hexagon, or any other single shape that can completely tile the plane (see **tiling**), is infinity. *Heesch's problem* is to find the largest possible *finite* Heesch number, or, more generally, what values other than zero and infinity can Heesch numbers take. In considering this problem, it's helpful to define the Heesch number more precisely. In a tiling, the *first corona* of a tile is the set of all tiles that have a common boundary point with the tile, including the original tile itself. The *second corona* is the set of tiles that share a point with anything in the first corona; and so on. The Heesch number is the maximum value of coronas (*k*) that can surround a shape. For a long time the record holder for the largest finite value of *k* was a shape found by the American computer scientist

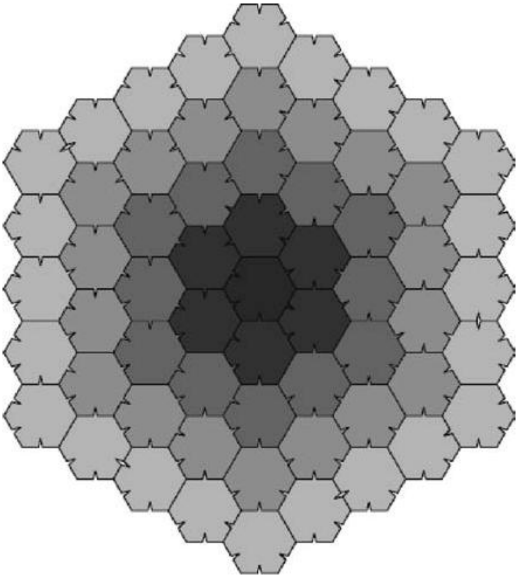

Heesch number A tiling in which copies of the same shape are used out to the fourth surrounding layer, or corona. *David Eppstein*

Robert Ammann, which consists of a regular hexagon with small bumps on two sides and matching notches on three sides. This was thought to have a Heesch number of three; however, in 2000, Alex Day argued that the Ammann hexagon actually has a Heesch number of four, though it isn't clear whether the difference has to do with a definition of tiling. In any event, it has since been shown by Casey Mann, of the University of Arkansas that there exists an infinite family of tiles (consisting of indented and outdented pentahex) with Heesch number five (or six by Day's reckoning)–the largest finite value currently known. Are there any polygons that have higher Heesch numbers? The answer is unknown but Mann thinks that more rounded **polyomino**s than the long skinny ones he's been using may have a better chance of giving unbounded Heesch numbers.

The Heesch number question is connected to two other famous unsolved tiling problems: the **domino problem** and the **Einstein problem**. Aperiodic tiling seems to act as a barrier to the existence of tiling algorithms, so it isn't expected that both of these problems have the same answer. On the other hand, if there's a maximum finite Heesch number k, then it seems that this could be used as the basis of an algorithm to test whether a shape tiles: simply attempt to fill out a tiling to the $(k + 1)$st corona; if successful, the shape must tile the plane, and if not, the shape will not tile. Similar questions can be asked about Heesch numbers for tilings in higher dimensions.

Hein, Piet (1905-1996)

An extraordinarily creative Danish mathematician, scientist, inventor, and poet who often wrote under the Old Norse pseudonym Kumbel, meaning "tombstone." A direct descendant of the Dutch naval hero of the sixteenth century who had the same name, Piet Hein was born in Copenhagen and studied at the Institute for Theoretical Physics of the University of Copenhagen (later the Niels Bohr Institute), the Technical University of Denmark, and the Royal Swedish Academy of Fine Art. He was later awarded an honorary doctorate by Yale University. A good friend of Albert Einstein, he is famed for his many mathematical games, including **Hex**, **Tangloids**, Polytaire, **TacTix**, and the **Soma cube**. These games were featured in numerous columns of Martin **Gardner**'s "Mathematical Recreations" column in *Scientific American* and often achieved worldwide attention in this way. As an artist and constructor, Hein gave form, in the 1950s and 1960s, to elegant pieces of furniture that helped "Scandinavian design" attract international recognition. These pieces, including a dining-room table created in cooperation with the Swedish designer Bruno Mathsson, were based on the **superellipse** curve–a shape that Hein also brought to bear in applications as varied as city planning (it's the basis for Sergel's Square in the center of Stockholm) and toy making (see **superegg**). Hein was a prolific and excellent writer of light verse, producing thousands of short, aphoristic poems known as Grooks. For him there was no unbridgeable gap between the subjectivity of fine art and the objective world of science. "Art," he said, "is a solution to problems which cannot be formulated clearly before they have been solved." His philosophy of life was summed up by his aphorism "co-existence or no existence."

helicoid

The second oldest known **minimal surface**; it was discovered by Jean-Baptiste Meusnier in 1776, thirty years after the **catenoid**. It is the only minimal surface, apart from the simple plane, that is also a **ruled surface**. The helicoid is the surface swept out by a line that always intersects a fixed axis at right angles and that rotates uniformly as its point of intersection moves uniformly along the axis. This line intersects any cylinder concentric with the axis in a **helix**. The helicoid has a wide variety of shapes and is a familiar sight in everyday life, taking the form of many things from spiraling parking ramps to screw threads.

helix

A curve in three dimensions, the tangent to which makes a constant angle with a fixed line. A *circular helix* is formed by winding a line around a **cylinder** so the radius is always the same. A *conical helix* is formed by winding a line around a **cone**, so that, consequently, its radius constantly changes. Springs often take the form of various kinds of helices. In nature, the DNA molecule is in the shape of a double helix.

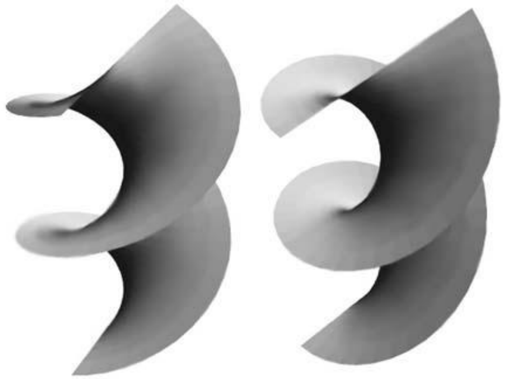

helicoid *Richard Palais*

Henon, Michele

An astronomer at the Nice Observatory in southern France. For a number of years, particularly during the 1960s, he studied the dynamics of stars moving within galaxies, using computers as a way to understand the stability of their motions. His work was very much in the spirit of Henri **Poincaré**'s approach to the classical **three-body problem**: What important geometric structures govern their behavior? The main property of these systems is that the energy of their motion is constant to a very good approximation. Consequently, their chaotic dynamics are not described by **simple attractor**s, but by objects that are markedly more difficult to analyze and visualize, existing on energy "surfaces" in three and higher dimensions. During the 1970s, Henon discovered a very simple iterated mapping that showed a **chaotic attractor**, now called *Henon's attractor,* that allowed him to make a direct connection between deterministic data and **fractal**s. The Henon attractor is self-similar (see **self-similarity**): if you zoom in on the attractor in its state space you find more and more layers, much like filo dough or a croissant.

Henstock integration

See **integration**.

heptagon

A **polygon** with 7 sides.

Hermann grid illusion

An illusion first described by the German physiologist Ludimar Hermann (1838–1914) in 1870. While reading a book on sound by the Irish physicist John Tyndall, Hermann saw gray spots in the intersections of spaces among the figures that Tyndall had arranged in a matrix. Despite the fact that the same intensity of light is reflected all the way along the white spaces in the Hermann Grid, the intersections appear gray. To explain this, consider two regions of the retina. One region views an intersection of a white horizontal and vertical band, while the other views a white band between two intersections (the region going away from the intersection). Although the two regions themselves receive the same amount of light, the situation in their neighboring regions is different. At the intersection, light comes in from all four sides, but the white band that lies between the two intersections is surrounded by two dark sides. This leads to an effect called *lateral inhibition,* which causes a bright surround to an area appear darker and, conversely, a dark surround to an area appear lighter. A similar but more powerful illusion, known as the *Lingelbach illusion* or the *Scintillating grid illusion,* was discovered in 1994 by Elke Lingelbach, the wife of a German mathematics professor, and has not

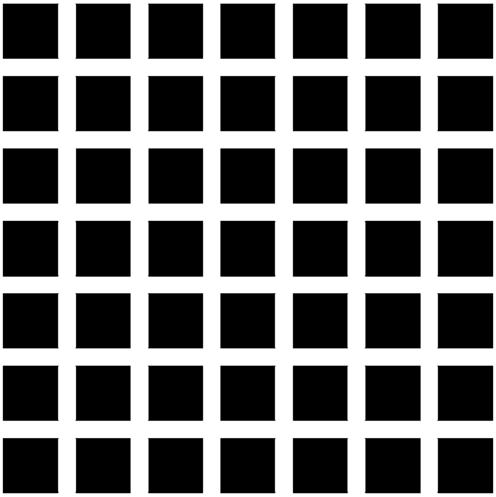

Hermann grid illusion The original illusion, in which the viewer sees gray spots at the intersections.

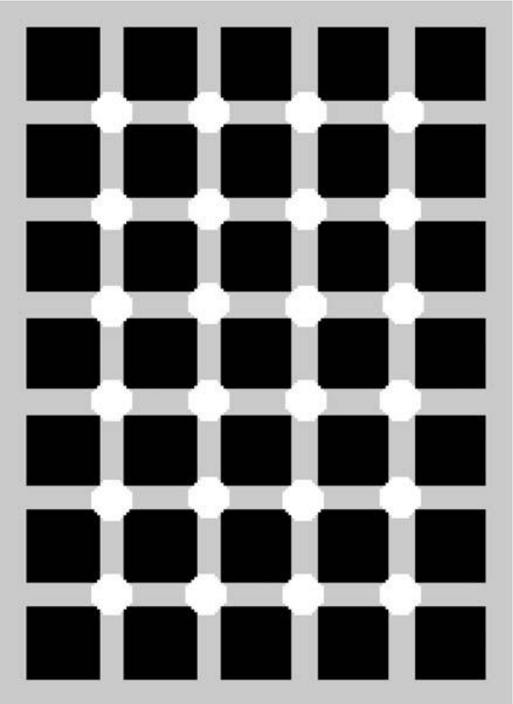

Hermann grid illusion The more striking and recently discovered "scintillating" version of the illusion.

yet been fully explained.[286] Curiously, the effect of the scintillation is lessened by tilting the head through 45°!

Hermite, Charles (1822–1901)

A French mathematician whose work in the theory of functions includes the application of **elliptic function**s to provide the first solution to the general equation of the fifth degree, the **quintic** equation. He also showed that ***e*** is a **transcendental number**, studied a class of **differential equation**s now known as *Hermite polynomials,* which later proved to be of importance in some applications of quantum mechanics, and discovered the properties of *Hermitian matrices.*

Heron of Alexandria (c. A.D. 60)

Also called *Hero,* a Greek geometer and inventor whose writings have helped preserve a knowledge of the mathematics and engineering of Babylonia, ancient Egypt, and the Greco-Roman world. His most important geometric work, *Metrica,* was lost until a fragment was discovered in 1894, followed by a complete copy in 1896. It is a compendium, in three books, of geometric rules and formulas, the best known of which is Proposition 1.8, now known as **Heron's formula**. He invented many devices operated by water, steam, or compressed air, including a fountain, a fire engine, siphons, and an engine in which the recoil of steam revolves a ball or a wheel.

Heron's formula

An important formula in plane geometry that allows the area of any triangle to be calculated without knowing the altitude (perpendicular height) of any of its sides. Let a, b, and c be the side lengths of a triangle and A its area. Heron's formula states that

$$A^2 = s(s - a)(s - b)(s - c),$$

where $s = (a + b + c)/2$. The origin of this formula is historically obscure. A medieval Arab source, for example, ascribes it to **Archimedes**. However, the first definite reference we have to it is by **Heron of Alexandria**. His proof is extremely convoluted, and it seems clear that it must have been determined by an entirely different thought process, and then dressed up in the usual synthetic form that the classical Greeks preferred for their presentations. Heron's formula contains **Pythagoras's theorem** as a degenerate case. A *Heronian triangle* is one with integer sides and integer area.

Herring illusion

A **distortion illusion** first published in 1861 by the German psychologist Ewald Herring (1834–1918), and now named after him. As in the case of the **Zöllner illusion** and others, it shows how geometrical relationships can seem to be distorted by their background (a lined background can make circles, squares, and triangles seem distorted, too). The straight horizontal lines in the illusion appear to bow out in the center. This can be explained if the brain interprets the radiating lines in terms of depth, making the central spot in the Herring diagram, and thus also the heavy black lines near the center of the diagram, appear to be farther away than the edges. Because the heavy black lines are the same thickness at the center as at the edges but are assumed to be farther away, the brain thinks they must be more widely spaced at the center.

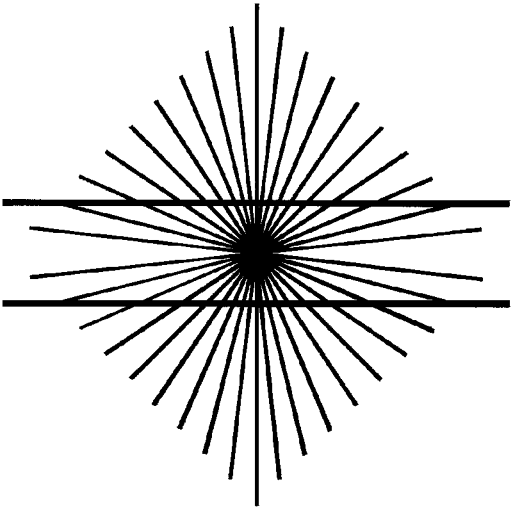

Herring illusion

heuristic argument

An educated guess: something that helps in finding the solution to a problem but is otherwise unjustified or incapable of justification.

Hex

A board game played by two players on a hexagonal grid, usually in the shape of an 11 × 11 rhombus. It was invented by Piet **Hein** in 1942 and independently by John **Nash** in 1948. Hein said that the game occurred to him while contemplating the **four-color problem** and it soon became popular in Denmark under the name *Polygon.* Nash's version was played by math students at Princeton and a number of other American campuses. Players use differently colored pieces–say, red and blue. They take alternate turns placing a piece of their color inside a hexagon, filling in that hexagon with their color. Red's goal is

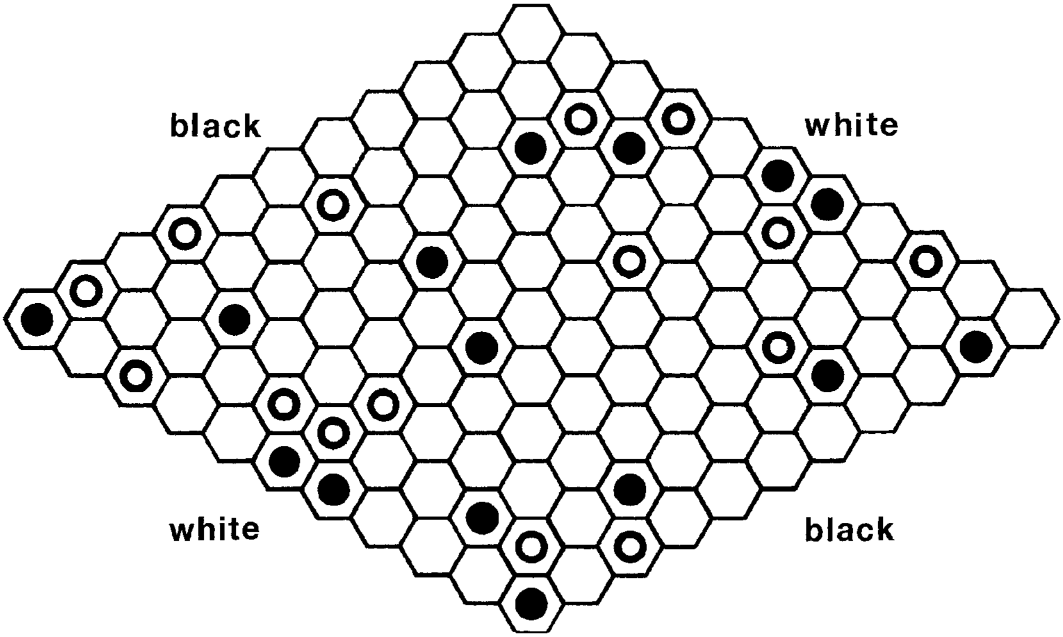

Hex A position in a game of Hex.

to form a red path connecting the top and bottom sides of the parallelogram; Blue's goal is to form a path connecting the left and right sides. The game can never end in a tie, a fact found by Nash. The only way to prevent your opponent from forming a connecting path is to form a path yourself. When the sides of the grid are equal, the game favors the first player and the first player has a winning strategy. There are two ways to make the game fairer. One is to make the second player's sides closer together, playing on a parallelogram rather than a rhombus; however, this has been proven to result in a win for the second player, so it theoretically doesn't improve matters. A better way is to allow the second player to choose his color after the first player makes the first move or to make the first three moves, which encourages the first player to intentionally even out the game.

hexa-

The Greek prefix meaning "six." A *hexagon* is a **six**-sided **polygon**. A *hexahedron* is a six-sided **polyhedron**, otherwise known as a **cube** if it is regular. *Hexadecimal* is the number system with **base** 16 (i.e., six more than the decimal system) and is used mostly in computing (because four binary digits can represent 16 different numbers). *Hexagonal numbers* are **figurate number**s (numbers that can be represented by a regular geometric arrangement of equally spaced points) of the form $n(2n - 1)$; the first few are 1, 6, 15, 28, 45, ... For *hexaflexagon* see **flexagon**. For *hexomino* see **polyomino**. See also **Giant's Causeway**.

higher dimensions

Dimensions beyond the familiar three spatial dimensions (up-down, left-right, back-forth) of which we are aware in every-day life. Intense speculation, both scientific and fictional, has naturally been directed toward the possibility of a **fourth dimension**. One way to think of points in four-dimensional space is as ordered sets of four numbers. Clearly, this algebraic representation can be extended to many arbitrary dimensions: n-dimensional space is defined as the set of the set of points $(a_1, a_2, \ldots, a_n)$ where a_1 to a_n can take any **real number** value. There has been much conjecture that the universe in which we live contains many more than three spatial dimensions. This speculation began with the **Kaluza-Klein theory** but is now firmly embedded in modern **string theory**.[25, 55, 81]

Hilbert, David (1862–1943)

> *One can measure the importance of a scientific work by the number of earlier publications rendered superfluous by it.*

A great German mathematician who was one of the colossi in the field in the twentieth century. His most important discovery was of what is now called **Hilbert space**. He was also a master of mathematical organization. During the early phase of his career, Hilbert reorganized **number theory**, crystallizing his conclusions in the classic book *Der Zahlbericht* (The theory of algebraic number fields, 1897). He then moved into geometry and performed a similar service by setting forth the first rigorous set of geometrical axioms in his *Grundlagen der Geometrie* (Foundations of geometry, 1899). He invented a simple **space-filling curve** now known as the *Hilbert curve* and also proved **Waring's conjecture**. At the Paris International Congress of 1900, Hilbert proposed 23 outstanding problems in mathematics to whose solutions he believed twentieth-century mathematicians should devote themselves. These problems have come to be known as *Hilbert's problems,* and a number still remain unsolved today. Hilbert's mathematical philosophy is partly re-vealed by a couple of remarks, one of which he made after learning that a student in his class had dropped the subject in order to become a poet. "Good," he said. "He did not have enough imagination to become a mathematician." Whether he really believed the second is open to question: "Mathematics is a game played according to certain simple rules with meaningless marks on paper."

Hilbert space

A space of infinite dimensions, named after David **Hilbert**, in which distance is preserved by making the sum of squares of coordinates a convergent sequence; it is of crucial importance in the mathematical formulation of **quantum mechanics**. See also **Fredholm, Erik Ivar**.

Hinton, Charles Howard (1853–1907)

An English-born mathematician best known for his writings and inventions aimed at helping to visualize the **fourth dimension**; he may also have coined the name ***tesseract*** for the four-dimensional analogue of a cube. Hinton matriculated at Oxford and continued to study there, earning a B.A. (1877) and an M.A. (1886), while he also taught, first at Cheltenham Ladies' School and then, from 1880 to 1886, at Uppingham School. At this time, another teacher at Uppingham was Howard Candler, who was a friend of Edwin **Abbott** and thus provides a possible link between these two explorers of other dimensions. In the early 1880s, Hinton published a series of pamphlets starting with "What Is the Fourth Dimension?" and "A Plane World" (a contemporary of Abbott's ***Flatland: A Romance of Many Dimensions***), which were reprinted in the two-volume *Scientific Romances* (1884). Hinton's descriptions owed much to the mathematical models of William **Clifford**, whose theories about four-dimensional spaces were then in vogue. But Hinton went much further in his attempts to break free of three-dimensional thought. He devised an elaborate set of small colored cubes to represent the various cross sections of a tesseract and then memorized the cubes and their many possible orientations in order to gain a window on the fourth dimension.

At the time he was teaching in England, Hinton married Mary Everest Boole, the eldest daughter of George **Boole**, the founder of mathematical logic. Regrettably, he also married a Maud Wheldon and was tried at the Old Bailey in London for bigamy. After serving a day in prison for the offence, he fled with his (first) family to Japan, where he taught for some years, before taking up a post at Princeton University. There, in 1897, he designed a species of baseball gun which, with the help of gunpowder charges, would shoot out balls at speeds of 40 to 70 miles per hour. It was used by the Princeton team for several seasons before being abandoned by the players in fear of their lives.

After a brief spell at the University of Minnesota, Hinton joined the Naval Observatory in Washington, D.C. At the same time, he more rigorously developed his ideas on the fourth dimension and presented his results before the Washington Philosophical Society in 1902. Hinton asked: What would prove the existence of a real fourth spatial dimension? He offered three possibilities, two of which involved a specific molecular structure and a particular case of electrical induction, and have since been explained by science in more mundane ways. However, Hinton's other case, pertaining to right- and left-handedness remains open because there are instances of right- and left-handedness in nature, such as the spin of elementary particles, to which his example could be applied. In any event, Hinton's final assessment that we can only regard a four-dimensional space as possible if three-dimensional mechanics fails to explain known physical phenomena still rings true today.[166, 167, 272] See also **Boole (Stott), Alicia**.

Hippias of Elis (c. 5th century B.C.)

An itinerant Greek philosopher who contributed significantly to mathematics by discovering the quadratrix, a special curve he may have used to trisect an angle (see **quadratrix of Hippias**). Hippias is one of the first mathematicians about whom a good deal is known. He came from a state in the northwest corner of Peloponnesia that was the home of the Olympic games. According to Plato, Hippias boasted, during one of his visits to the Olympics, that everything he wore–his clothing, sandals, ring, and oil flask–he'd made himself. Later, in Athens, Hippias became one of the first to teach for money, a practice forbidden by the Pythagoreans and scorned by Plato. He and other paid teachers became known as "sophists," which was a derogatory term at the time but has since come to mean "wise man."

hippopede

A **quartic** curve described by the equation

$$(x^2 + y^2)^2 + 4b(b - a)(x^2 + y^2) - 4b^2x^2 = 0,$$

where a and b are positive constants. Hippopede means "foot of a horse." It is often known as the *hippopede of Proclus,* after **Proclus** who was the first to study it (together with **Eudoxus**, who used it in his theory of how the planets move), and also the *horse fetter* and the *curve of Booth* because of work done on it by J. Booth (1810–1878). Any hippopede is the intersection of a **torus** (donut) with one of its **tangent** planes–that is, a plane parallel to its axis of rotational symmetry. The curve takes any of a variety of forms depending on where the donut is sliced. It may be a simple oval; an indented oval or *elliptical lemniscate of Booth* ($0 < b < a$); two isolated circles; a figure-eight curve or **lemniscate of Bernoulli** (the only hippopede that is also a **Cassinian oval**); or a *hyperbolic lemniscate of Booth* ($0 < a < b$).

Hi-Q

See **peg solitaire**.

Hnefa-Tafl

The Viking equivalent of chess and a particular form of a **Tafl game**. It effectively models the kind of internal conflict familiar to the Vikings and recounted in *Njal's Saga,* the best known of the Icelandic sagas. The king or chieftain sits in his hall, surrounded by his thanes. His enemies gather in secret, and in numbers sufficient to overwhelm the king's standing forces in a lightning raid. They gather around the king's hall and set it alight, forcing the defenders to fight in the open or burn in the hall. If the king can, by some desperate stratagem, break free and escape the trap, he can rally his people and strike back at his enemies. If not, he dies. Hnefa-Tafl reflects this mode of contest. The board has 19 × 19 grid lines (though sometimes has as few as 7 × 7 grid lines), and the pieces are placed at the points of the intersections (curiously analogous to the oriental game of **Go**, which also uses 19 × 19 lines and where play occurs at the intersections). The opposing forces are unequal in size, and have different objectives: the attackers attempt to trap the king in his hall, while the defenders try to open an escape route for him.

Hoffmann, Louis "Professor" (1839–1919)

The pseudonym of Angelo John Lewis, an English barrister who was the leading writer on magic, cards, and "parlor amusements" at the turn of the twentieth century. His *Puzzles Old and New*[171] (1893) is a major source of information about mathematical recreations.

Hofstadter, Douglas R. (1945–)

A physicist and philosopher best known for his 1980 Pulitzer Prize–winning book *Gödel, Escher, Bach: An Eternal Golden Braid.*[172] He is currently a professor of cognitive science and computer science at Indiana University, Bloomington, and has particular interests in themes of the mind, consciousness, self-reference, translation, and mathematical games. He is the son of the Nobel Prize–winning physicist Robert Hofstadter.

Hofstadter's law

It always takes longer than you think, even when you take Hofstadter's law into account.

Hogben, Lancelot Thomas (1895–1975)

An English zoologist and geneticist famed for his best-selling *Mathematics for the Million* (1933)[173] of which Albert Einstein said, "It makes alive the contents of the elements of mathematics" and H. G. Wells said, "A great book, a book of first class importance." Hogben was born in Southsea, Hampshire, and studied at Cambridge and London. Imprisoned in 1916 as a conscientious objector during World War I, he was released only when his health went into serious decline. He held various academic posts in Britain, Canada, and South Africa, becoming professor of social biology at London University in 1930. During World War II he was put in charge of the medical statistics records for the British Army. After the war he became professor of medical statistics at the University of Birmingham, where he remained until his retirement in 1961. Hogben first began to apply mathematical principles to the study of genetics in the 1930s, focusing on the study of generations of the fruit fly *Drosophila* and how it related to research on heredity in humans. In addition to *Mathematics for the Million,* he authored half a dozen other books, including the popular *Science for the Citizen.* Though trained as a scientist, Hogben was passionately interested in linguistics. In *The Loom of Language,* which he edited, he set out the principles of his own invented language, "Interglossa," based on Greek and Latin roots but with a syntax resembling that of Chinese.

hole

A topological structure (see **topology**) that prevents any object in which it occurs from being continuously shrunk to a point. A **sphere** has no holes; a **torus** and a teacup each have one hole. See also **genus**.

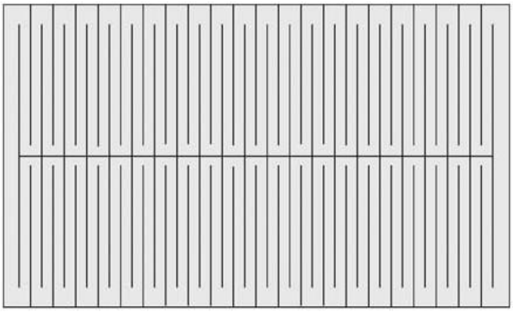

hole-in-a-postcard problem Photocopy the diagram, cut along the lines, and step through the postcard!

hole-in-a-postcard problem

With a pair of scissors, make cuts in a regular-sized postcard to create a hole large enough for a person to step through it. The solution is shown in the diagram. The number of lines determines the size of the resulting aperture. With enough cuts you could literally drive a horse and cart through the card!

hole-through-a-sphere problem

A number of **geometry puzzles** hinge on a surprising fact about spheres that have had holes bored through them. Imagine you have a round bead that is 1 inch (about 2.5 cm) in diameter and that you drill a hole exactly through the middle of this so that the remaining part of the sphere is only half an inch thick. Now imagine that an enormously large drill has been used to bore a hole though Earth so large that the part of Earth that is left behind is only half an inch thick. Amazingly, the residual volumes of these two holey spheres, the drilled bead and the drilled Earth, are exactly the same! It just happens that even though Earth is vastly larger than the bead, the drill has to take out proportionately more in order to make the thickness of the hole the same, so that the volume left doesn't depend separately on the initial size of the sphere or of the hole, but only on their *relation,* which is forced by requiring the hole to be half an inch long. This fact enables the following poem-problem to have a solution even though its seems as if not enough information has been provided:

Old Boniface he took his cheer,
Then he bored a hole through a solid sphere,
Clear through the center, straight and strong,
And the hole was just six inches long.
Now tell me, when the end was gained,
What volume in the sphere remained?
Sounds like I haven't told enough,
But I have, and the answer isn't tough.

Having already learned the secret that the volume that remains of a drilled sphere doesn't depend on the initial size of the sphere, we can cheat and give a kind of meta-argument that is much shorter than the geometric proof. The volume left behind of any sphere with a 6-inch-long hole through it must be the same as the volume left behind of a 6-inch-diameter sphere with a hole of 0 diameter drilled through it. This is equal to $\frac{4}{3}(\pi 6^3)$, or approximately 905 cubic inches.

hole-through-the-earth problem

> *I wonder if I shall fall right* through *the earth!*
> —Alice in *Alice in Wonderland*

Imagine there is a hole going from one point on Earth's surface, all the way through the center of Earth, to the exact opposite (antipodal) point on the other side. What would happen if you dropped something into this hole? The Greek historian Plutarch considered the problem some 2,000 years ago. In 1624, van Etten argued that a millstone dropped down such a hole at 1 mile per minute would take more than $2\frac{1}{2}$ days to reach the center, where "it would hang in the air." The first correct answer was given by Galileo is his *Dialogue Concerning the Two Chief World Systems* (1632). Galileo realized a dropped object would accelerate until it reached the center of Earth, travel through to the other side, then oscillate back and forth.

holism

The idea that the whole is greater than the sum of the parts. Holism is credible on the basis of **emergence** alone, since **reductionism** and bottom-up descriptions of nature often fail to predict complex higher-level patterns.

homeomorphic

In **topology**, two objects are said to be homeomorphic if they can be smoothly deformed into each other. See also **homeomorphism**.

homeomorphism

A **one-to-one** continuous transformation that preserves **open** and **closed** sets.

homology

A way of attaching **Abelian group**s, or more elaborate algebraic objects, to a **topological space** so as to obtain algebraic invariants. In a sense, it detects the presence of "holes" of various dimensions in the space. The methods developed to handle this led to what is now called *homological algebra,* a subject in which homological invariants are calculated for many purely algebraic structures.

homomorphism

A **function** that preserves the operators associated with the specified structure.

homotopy

A continuous transformation from one path in a **topological space** to another, or more generally, of one **function** to another (see **continuity**). Paths connected by a homotopy are called *homotopic* and are said to be in the same *homotopy class*. Properties left unchanged by such homotopies are known as *homotopy invariants*. Homotopy classes of paths can be composed to form the **fundamental group**, or *first homotopy group*. Other maps can be used to form *higher homotopy groups*.

Hordern, L. Edward (d. 2000)

An English puzzlist and leading authority and writer on **sliding-piece puzzles**.[175]

hundred

The smallest three-digit number in the decimal system and the smallest square of a two-digit number (10). A hundred today means 100 but, over the years and in different places, it has stood for different values including 112, 120, 124, and 132. The remnants of these old measures still persist in the *hundredweight* of some countries representing 112 or 120 pounds. A hundred is also a measure of land area, frequently used in colonial America and in England to signify a division of a county or a shire having its own court. A strange custom is invoked if a member of the English Parliament's lower house, the House of Commons, wishes to resign his seat (something that is technically illegal). That member accepts stewardship of the "Chiltern Hundreds," an area of chalk hills near Oxford and Buckingham, which effects his release from Parliament. The "Hundred Years' War" between the English and the French actually lasted 116 years.

hundred fowls problem

A Chinese puzzle found in the sixth-century work of the mathematician Zhang Qiujian. Similar problems involving two constraints and three unknowns are found in early European and Arabic mathematics from about the eighth century A.D. on.

PUZZLE

If a rooster is worth five coins, a hen three coins, and three chicks together are worth one coin, how many roosters, hens, and chicks totaling 100 can be bought for 100 coins? It turns out that there are three different solutions. These can be found the long way, by trial and error, or by using algebra. Call the number of roosters R, the number of hens H, and the number of chicks C. The problem gives two constraints. First, the total number of fowl must be 100, so $R + H + C = 100$. Second, the total cost of the fowl must be 100. The cost of roosters is $5R$, the cost of hens is $3H$, and the cost of chicks is $(1/3)C$, so $5R + 3H + (1/3)C = 100$. These two equations can be used to get rid of one of the unknowns; then it's a question of guess and check.

Solutions begin on page 369.

Hunter, James Alston Hope

An American mathematician and puzzlist who has written numerous articles and several books on recreational mathematics (two in partnership with Joseph **Madachy**), and was the author of a syndicated puzzle column read throughout the United States and Canada. In 1955, he coined the name "**alphametic**" and is probably the most prolific producer of **cryptarithm**s.

Huygens, Christiaan (1629–1695)

A Dutch scientist and mathematician who solved the **tautochrone problem**, proposed a new wave theory of light, designed a new pendulum clock, discovered Saturn's largest moon (Titan), and sketched the first feature on the surface of another planet (Syrtis Major on Mars). In his final years Huygens composed one of the earliest discussions of extraterrestrial life, published after his death as the *Cosmotheoros* (1698).

Hypatia of Alexandria (C. A.D. 370–417)

The first woman known to have made a significant contribution to mathematics. Although there is no evidence that Hypatia did any original research, she assisted her father, Theon of Alexandria, in writing his 11-part commentary on Ptolemy's great work on astronomy and mathematics, the *Almagest*. It's thought that she also helped in producing a new version of **Euclid**'s *Elements*, which formed the basis for all later editions of Euclid. Hypatia became head of the Platonist school at Alexandria in about A.D. 400 and, as a pagan, represented a menace to some Christian sects who felt threatened by her learning and depth of scientific knowledge. In the end, although the exact circumstances are unclear, she was murdered by a mob. The event served as a trigger for the departure of many scholars and the beginning of the decline of Alexandria as a major academic center.

hyperbola

One of the **conic section** family of curves, which also includes the **circle**, the **ellipse**, and the **parabola**; it is obtained if a double cone is cut by a plane inclined to the axis of the cone such that it meets both branches of the cone. Of the four conic curves, the hyperbola is the one least encountered in everyday life. A rare chance to see the complete shape is when a lamp with a cylindrical or conical shade throws shadows on a nearby wall. Part of a hyperbola is produced by the liquid that climbs by capillary action between two microscope slides held vertically and almost touching.

A hyperbola is the path followed by a smaller object that is traveling fast enough to escape completely from the gravitational pull of a larger object. Some comets, for example, have hyperbolic or "open" orbits so that, after one swing around the Sun, they head off into interstellar space never to return. It can be difficult to tell, in some cases, whether a comet's orbit is hyperbolic or is highly elliptical and, therefore, closed. In fact, one way to think of a hyperbola is as a kind of ellipse that is split in half by infinity. Not surprisingly, the hyperbola and the ellipse share many inverse relationships. For example, whereas the eclipse is the **locus** of all points whose distances from two fixed points, called *foci,* have a constant sum, the hyperbola is the locus of all points whose distances, r_1 and r_2, from two fixed points, F_1 and F_2, is a constant difference, $r_2 - r_1 = k$. If a is the distance from the origin to either of the x intercepts of the hyperbola, then $k = 2a$. Also, let the distance between the foci, $F_2 - F_1 = 2c$. Then the *eccentricity,* a measure of the flatness of the hyperbola, is given by $e = c/a$. For all hyperbolas, $e > 1$; the larger the value of e, the more the hyperbola resembles two parallel lines. Just as the circle (for which $e = 0$) is the limiting case of the ellipse (for which $0 < e < 1$), so the parabola ($e = 1$) is the limiting case of both the ellipse and the hyperbola.

A hyperbola has two **asymptotes**: the never-quite-attainable limits of the curve's branches as they run away to infinity. The *transverse axis* of the hyperbola is the line on which both foci lie and that also intersects both vertices (turning points); the *conjugate axis* goes through the center and is perpendicular to the transverse axis.

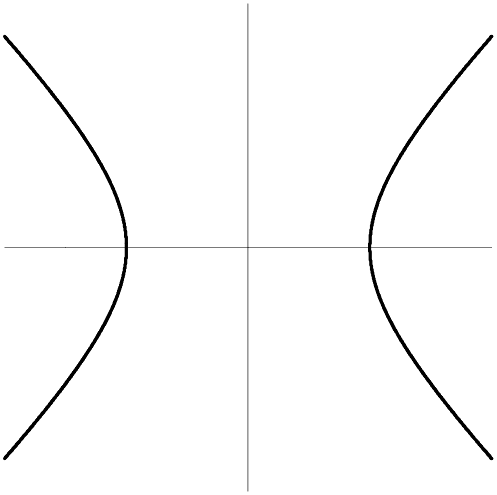

hyperbola © *Jan Wassenaar, www.2dcurves.com*

A *rectangular hyperbola* has an eccentricity of $\sqrt{2}$, asymptotes that are mutually perpendicular, and the property that when stretched along one or both of its asymptotes remains unchanged. This special case of the hyperbola was first studied by **Menaechmus**. **Euclid** and Aristaeus wrote about the general hyperbola but only studied one branch of it, while **Apollonius** was the first to study the two branches of the curve of the hyperbola and is generally thought to have given it its present name.

The **pedal curve** of a hyperbola with one focus as the pedal point is a circle. The pedal of a rectangular hyperbola with its center as pedal point is a **lemniscate of Bernoulli**. The **evolute** of a hyperbola is a **Lamé curve**. If the center of a rectangular hyperbola is taken as the center of inversion, the rectangular hyperbola inverts to a lemniscate. If the vertex is used as the center of inversion, the rectangular hyperbola inverts to a right **strophoid**. If the focus of a hyperbola is taken as the center of inversion, the hyperbola inverts to a limaçon (see **limaçon of Pascal**). In this last case if the asymptotes of the hyperbola make an angle of $\pi/3$ with the axis that cuts the hyperbola, then it inverts to the Maclaurin **trisectrix**. See also **hyperboloid**.

hyperbolic geometry

One of the two main types of **non-Euclidean geometry** and the first to be discovered. It is concerned with saddle-surfaces, which have negative **curvature** and on which the **geodesic**s are **hyperbolas**. In hyperbolic geometry, contrary to the **parallel postulate**, there exists a line m and a point p not on m such that at least two distinct lines parallel to m pass through p. As a result, the sum of the angles of a triangle is less than 180° and, for a right triangle, the square of the hypotenuse is greater than the sum of the squares of the other two sides. See also **elliptical geometry**.

hyperbolic spiral

The curve whose equation in **polar coordinates** is $r\theta = a$.

hyperboloid

A **quadratic** surface of which there are two basic forms: a *hyperboloid of one sheet,* generated by spinning a **hyperbola** around its conjugate axis, and a *hyperboloid of two sheets* produced by rotating a hyperbola about its transverse axis. The hyperboloid of one sheet, first described by **Archimedes**, has some particularly remarkable properties. In 1669 Christopher Wren, the architect who de-

hyperboloid The McDonnell Planetarium in St. Louis is an example of a hyperboloid. *Courtesy of the St. Louis Science Center*

signed St. Paul's Cathedral in London, showed that this kind if hyperboloid is what mathematicians now call a **ruled surface**–a surface composed of infinitely many straight lines. This fact enables a close approximation to a hyperboloid to be made in the form of a string model. Two circular disks, of the same size, are held parallel, one exactly above the other, by a framework. Strings are then run through holes near the circumference of one circle to corresponding holes in the other circle that are a fixed distance farther around the circumference. Each string is perfectly straight but the surface that emerges takes the curved form of a hyperboloid. For the same reason, a cube spun rapidly on one of its corners will appear to describe a hyperbolic curve when viewed side-on.

Prominent examples of hyperboloids can be seen in the form of cooling towers at power stations and, most strikingly, in the shape of the McDonnell Planetarium in St. Louis, Missouri. The designer of this building, Gyo Obata, chose the design because the hyperbolic paths of some comets suggest "the drama and excitement of space exploration."

hypercube

A higher dimensional analog of a cube. A four-dimensional hypercube is known as a **tesseract**.

hyperellipse

See **superellipse**.

hyperfactorial

A number such as 108, which is equal to $3^3 \times 2^2 \times 1^1$. In general, the *n*th hyperfactorial $H(n)$ is given by

$$H(n) = n^n\ (n-1)^{n-1} \ldots 3^3\ 2^2\ 1^1.$$

The first eight hyperfactorials are 1, 4, 108, 27,648, 86,400,000, 4,031,078,400,000, 3,319,766,398,771,200,000, and 55,696,437,941,726,556,979,200,000. See also **large numbers** and **superfactorials**.

hypergeometric function

The sum of the *hypergeometric series:*

$$F(a;\, b;\, c;\, x) = 1 + \frac{ab}{1.c}\,x + \frac{a(a+1)b(b+1)}{1.2.c(c+1)}\,x^2 + \ldots$$

Many common functions can be written as hypergeometric functions.

hyperreal number

Any of a colossal set of numbers, also known as *nonstandard reals,* that includes not only all the **real number**s but also certain classes of infinitely large (see **infinity**) and **infinitesimal** numbers as well. Hyperreals emerged in the 1960s from the work of Abraham **Robinson** who showed how infinitely large and infinitesimal numbers can be rigorously defined and developed in what is called **nonstandard analysis**. Because hyperreals represent an extension of the real numbers, R, they are usually denoted by *R.

Hyperreals include all the reals (in the technical sense that they form an ordered field containing the reals as a subfield) and they also contain infinitely many other numbers that are either infinitely large (numbers whose absolute value is greater than any positive real number) or infinitely small (numbers whose absolute value is less than any positive real number). No infinitely large number exists in the real number system and the only real infinitesimal is zero. But in the hyperreal system, it turns out that each real number is surrounded by a cloud of hyperreals that are infinitely close to it; the cloud around zero consists of the infinitesimals themselves. Conversely, every (finite) hyperreal number x is infinitely close to exactly one real number, which is called its *standard part,* st(x). In other words, there exists one and only one real number st(x) such that x – st(x) is infinitesimal.

hypersphere

A four-dimensional analog of a **sphere**; also known as a *4-sphere.* Just as the shadow cast by a sphere is a circle, the shadow cast by a hypersphere is a sphere, and just as the intersection of a sphere with a plane is a circle, the intersection of a hypersphere with a hyperplane is a sphere. These analogies are reflected in the underlying mathematics.

$x^2 + y^2 = r^2$ is the Cartesian equation of a circle of radius r;

$x^2 + y^2 + z^2 = r^2$ is the corresponding equation of a sphere;

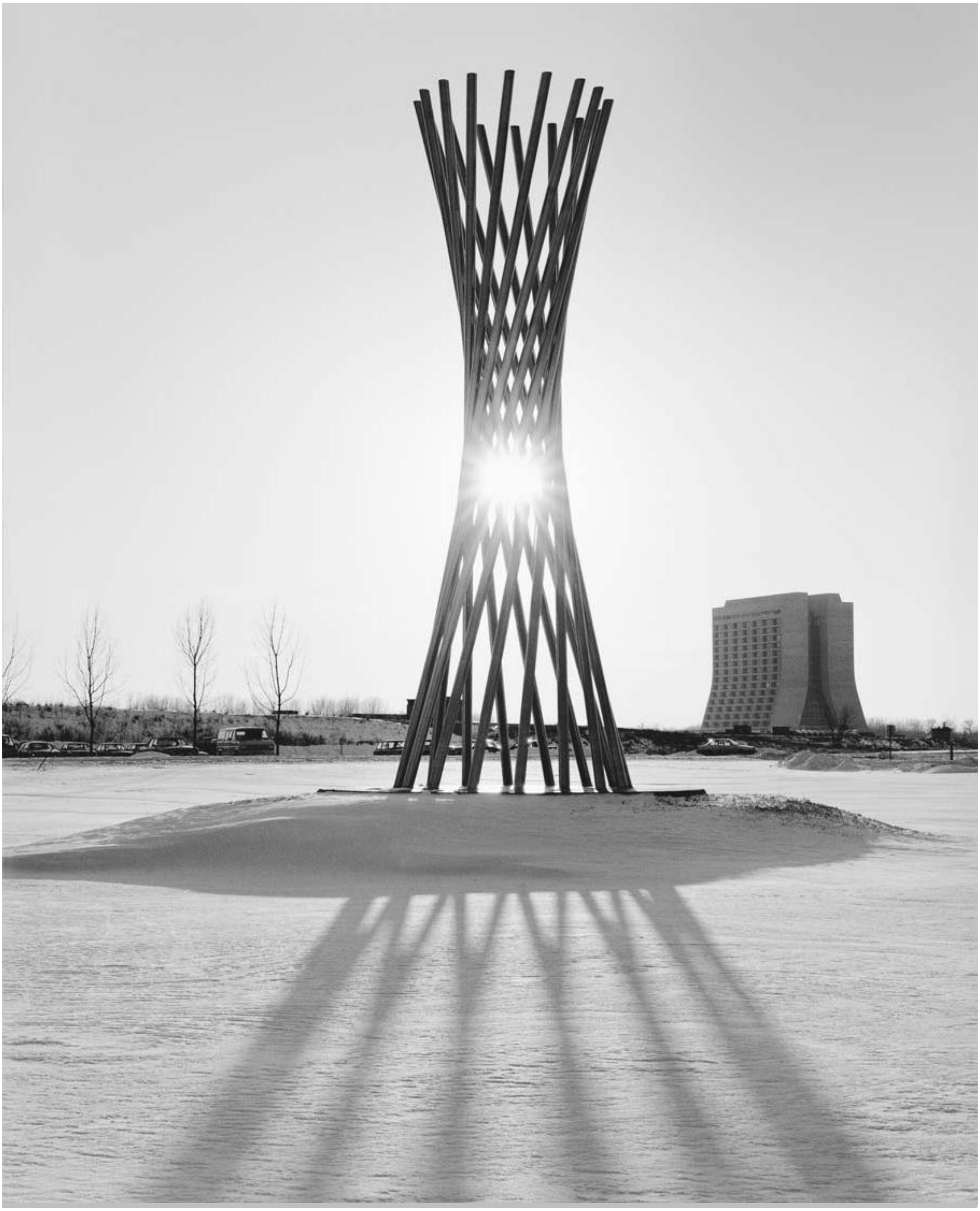

hyperboloid A hyperboloidal sculpture at the Fermi National Accelerator Laboratory. *FNAL*

$x^2 + y^2 + z^2 + w^2 = r^2$ is the equation of a hypersphere, where w is measured along a **fourth dimension** at right angles to the x-, y-, and z-axes.

The hypersphere has a *hypervolume* (analogous to the volume of a sphere) of $\pi^2 r^4/2$, and a *surface volume* (analogous to the sphere's surface area) of $2\pi^2 r^3$. A solid angle of a hypersphere is measured in *hypersteradians,* of which the hypersphere contains a total of $2\pi^2$. The apparent pattern of 2π radians in a circle and 4π steradians in a sphere does not continue with 8π hypersteradians because the *n-volume, n-area,* and number of *n-radians* of an *n-sphere* are all related to the **gamma function** and the way it can cancel out powers of π halfway between integers. In general, the term "hypersphere" may be used to refer to any n-sphere.

hypocycloid

A curve formed by the path of a point attached to a circle of radius b that rolls around the inside of a larger circle of radius a. The parametric equations for a hypocycloid can be written as:

$$x = (a - b)\cos(t) + b\cos((a/b - 1)t)$$
$$y = (a - b)\sin(t) - b\sin((a/b - 1)t)$$

The type of hypocloid depends on where the point whose path is being traced is located on the rolling circle. If it lies on the circumference of the circle, the curve generated is an ordinary hypocycloid. If it lies elsewhere the result is a **hypotrochoid**. A hypocycloid has a closed form–that is, the moving point eventually retraces its steps–when the ratio of the rolling circle and the larger circle, fixed, is equal to a rational number. When this ratio is in its simplest form, the numerator is the number of revolutions covered inside the fixed circle before the curve closes. In the same family of curves as the hypocycloid (and hypotrochoid) are the **epicycloid** and **epitrochoid**.

hypoellipse

See **superellipse**.

hypotenuse

The longest side of a **right triangle**–the one opposite the right angle. The word comes from the Greek roots *hypo* meaning "under" (which also appears, for example, in hypodermic, "under the skin") and *tein* or *ten,* for stretch. Thus hypotenuse is the line segment "stretched under" the right angle.

hypothesis

See **conjecture**.

hypotrochoid

A curve formed by the path of a point attached to a point c, which is not on the circumference, of a circle of radius b that rolls around the inside of a larger circle of radius a. The parametric equations for a hypotrochoid can be written as:

$$x = (a - b)\cos(t) + c\cos((a/b - 1)t)$$
$$y = (a - b)\sin(t) - c\sin((a/b - 1)t)$$

The hypotrochoid is a generalized **hypocycloid** and comes in two varieties: a *prolate hypocycloid* if the starting point is outside the circumference of the rolling circle and a *curtate hypocycloid* if the starting point is inside the rolling circle. In the same family of curves as the hypotrochoid (and hypocycloid) are the **epicycloid** and **epitrochoid**.

I

i

The square root of minus one; *i* is also known as the *imaginary unit number,* $\sqrt{-1}$. The product of *i* and a real number is known as an *imaginary number.* For details, see **complex number**.

I-Ching

A system of Chinese divination based on 64 hexagrams, each consisting of six horizontal lines each of which may be solid (representing the male principle, or yang) or have a break in the middle (representing the female principle, or yin). The first three lines of the hexagram, from the bottom up, constitute the lower trigram and symbolize the inner world. The fourth, fifth, and sixth lines constitute the upper trigram and symbolize the outer world. One of the first Europeans to see the hexagram structure when it was brought out of China in the early 1700s was Gottfried **Leibniz**, the first mathematician to work with base two, or **binary**, arithmetic.

icosahedron

A **polyhedron** with 20 faces. A *regular icosahedron* has faces that are all equilateral **triangle**s, and is one of the five **Platonic solid**s. The length from **vertex** to opposing vertex of a regular icosahedron is $5^{1/4} \times \phi^{1/2} \times d$ where ϕ is the **golden ratio** and d is the length of the side of one of the triangular faces. Chopping off each vertex (corner) of a regular icosahedron reveals the 12 pentagonal and 20 hexagonal faces of the *truncated icosahedron,* which is one of the 13 **Archimedean solid**s (shapes made from truncating Platonic solids in certain ways). See also **buckyball**.

Icosian game

A game devised by William **Hamilton** and first described by him in 1857 at a meeting of the British Association for the Advancement of Science in Dublin. The object of the game is to find a way around the edges of a **dodecahedron** so that every vertex (corner) is visited once and only once. A path such as this became known as a **Hamilton circuit**, though the task of finding a circuit that passes just once through every vertex of a shape seems to have arisen first in connection with Leonhard **Euler**'s study of the **knight's tour**. Two years before Hamilton introduced his game, Thomas **Kirkman** posed the problem explicitly in a paper that he submitted to the Royal Society: Given a graph of a **polyhedron**, does there exist a cycle passing through every vertex?

The Icosian game stemmed from Hamilton's invention of a curious kind of algebra that he called *icosians,* based on the symmetry properties of the **icosahedron**. Hamilton connected the mathematics of his icosians with the problem of traveling along the edges of a dodecahedron, hitting each vertex just once, and coming back to the starting point. His friend and fellow Irishman John Graves (1806–1870) suggested turning the problem into a commercial game and put Hamilton in contact with the London company of John Jacques and Sons, toy-makers and manufacturer of high quality chess sets. Jacques bought the rights to the game for £25 and marketed two versions of it, under the name Around the World. One version, for the parlor, was played on a flat board; another, for the "traveler," consisted of an actual dodecahedron. In both cases, nails at each vertex stood for a major city of the world and the player wrapped a piece of string around these nails as they went. The game was a

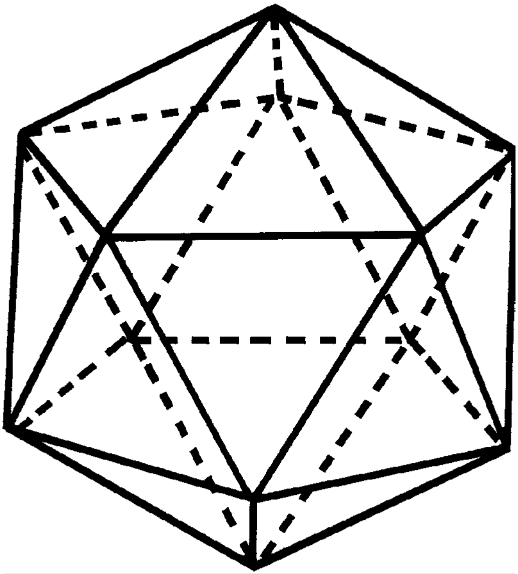

icosahedron The most multifaceted of the Platonic solids.

complete sales flop, mainly because it was too easy, even for children–but not for Hamilton himself who always used the icosian calculus to figure out his moves, instead of just trying different paths like everyone else! See also **traveling salesman problem**.

idempotent

The element x in some algebraic structures is called idempotent if $x \times x = x$.

identity

A statement that two expressions are always equal, for any values of their variables.

imaginary number

A number whose square is negative. Every imaginary number can be written in the form ib where b is a **real number** and i the *imaginary unit* with the property that $i^2 = -1$. Imaginary numbers are **complex numbers** in which the real part is zero. The term "imaginary" is unfortunate because it suggests something that has less reality than a "real" number, which isn't the case.

impossibilities in mathematics

> *Alice laughed: "There's no use trying," she said; "one can't believe impossible things." "I daresay you haven't had much practice," said the Queen. "When I was younger, I always did it for half an hour a day. Why, sometimes I've believed as many as six impossible things before breakfast."*
>
> –Lewis Carroll, *Alice in Wonderland*

Mathematicians are used to believing things that most people would consider impossible or, at least, too outrageous to contemplate, such as the **Banach-Tarski paradox**. However, there are some genuine impossibilities, even in mathematics, including **trisecting an angle**, **duplicating the cube**, and **squaring the circle** using only a straightedge and compass; finding the center of a given circle with a straightedge alone; deriving Euclid's **parallel postulate** from the other four; and representing the **square root of 2** as a rational fraction a/b. Less well known is this little gem from Gustave Flaubert (1821–1880), who sounds as if he had seen too much of this type of problem in school:

> Since you are now studying geometry and trigonometry, I will give you a problem. A ship sails the ocean. It left Boston with a cargo of wool. It grosses 200 tons. It is bound for Le Havre. The mainmast is broken, the cabin boy is on deck, there are 12 passengers aboard, the wind is blowing East-North-East, the clock points to a quarter past three in the afternoon. It is the month of May. How old is the captain?

impossible figure

An image in two dimensions of an object that, because of spatial inconsistencies, is impossible to realize fully in three dimensions. Among the best known impossible figures are the **Penrose triangle**, the **Penrose stairway**, the **Impossible trident**, and the Freemish crate. Notable pioneers of this peculiar form of representation have been Ocar **Reutersvärd**, Roger **Penrose** (and his father), and M. C. **Escher**. See also **optical illusion**.

impossible tribar

See **Penrose triangle**.

impossible trident

One of the most notorious **impossible figures**. It was first seen by many when it appeared on the cover of the March 1965 issue of *Mad* magazine. The two halves of the figure seem perfectly reasonable in themselves. When the top part is covered, the bottom part is taken to be three separate cylinders or tubes. With the bottom part hidden, the foreground figure is interpreted as being built of flat faces making two rectangular prongs. The trouble is that these two aspects of the figure are totally incompatible. Somewhere in the middle, the foreground and background swap places and give rise to an irreconcilable paradox. Over the years, countless adaptations of the trident have appeared with names such as the devil's fork, the three-stick clevis, the blivit, the impossible columnade, the trichotometric indicator support, and, most extravagantly, the triple encabulator tuned manifold. Swedish artist Oscar **Reutersvärd**'s mastery of such figures has led him to draw thousands of variations on the theme. When the figure is drawn long, it is easy to perceive locally as a three-dimensional object and to overlook its inherent inconsistency, because the contradictory clues are too well separated. When the figure is of medium length, the figure is easily interpreted as a three-dimensional object, and its impossibility is quickly perceived. If the prongs are very short, the two different interpretations vie for acceptance within the same local area; thus there is no consistent interpretation and the illusion breaks down. Some early writers commented that the impossible trident couldn't be built in any form in three dimensions. However, this has been shown to be false. In 1985, the Japanese artist Shigeo Fukuda made a three-dimensional model of the trident in the form of classical columns in which the illusion works–from one critical angle.

The origins of the figure are uncertain. It turns out that *Mad* magazine bought the illustration rights from a con-

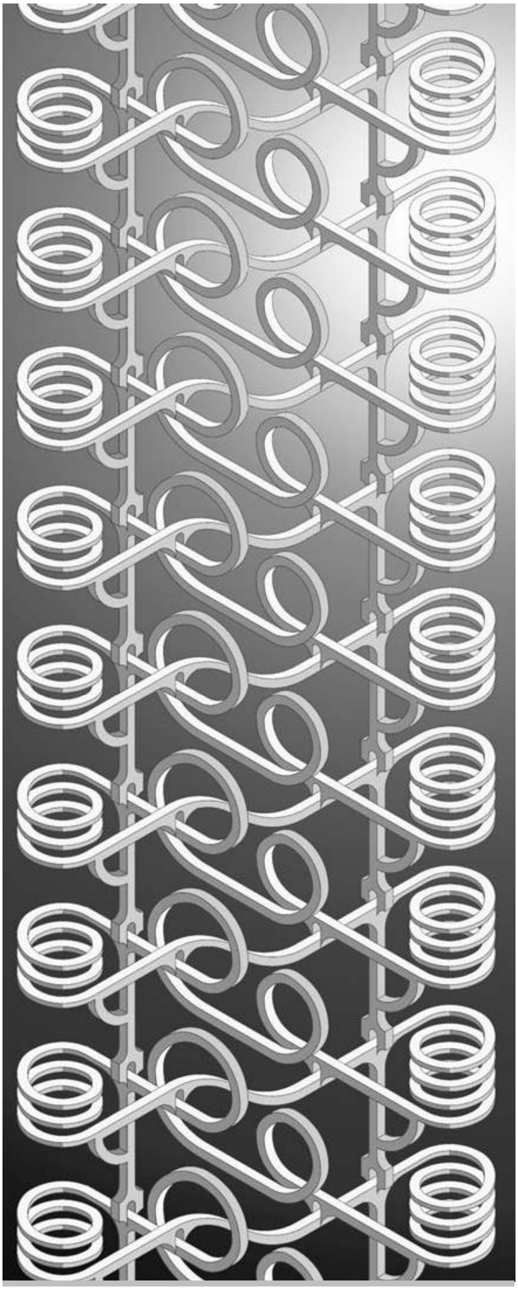

impossible figure "The Rollercoaster," a rendering in two dimensions of a structure impossible to build in three dimensions. *Jos Leys, www.josleys.com*

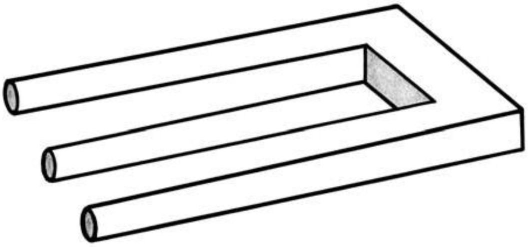

impossible trident

tributor who claimed that it was original; however, the magazine's management soon found out to their embarrassment that the figure had been previously published. It began to surface in several popular engineering, aviation, and science-fiction periodicals in May and June of 1964. D. H. Schuster published an article that same year in the *American Journal of Psychology,* which first brought the figure to the attention of the psychological community.

incircle

The **inscribed circle** of a triangle.

incomputable number

A **real number** with an infinite decimal (or binary) expansion that cannot be enumerated by any **universal computer**.

induction

A method of reasoning by which one infers a generalization from a series of instances. Say there is a hypothesis H that contains the variable n, which is a whole number. To prove by induction that H is true for every value of n is a two-step process: (1) prove that H is true for $n = 1$; (2) prove that H being true for $n = k$ implies that H is true for $n = k + 1$. This is sufficient because (1) and (2) together imply that H is true for $n = 2$, which, from (2), then implies H is true for $n = 3$, which implies H is true for $n = 4$, and so on. H is called an *inductive hypothesis.* Some philosophers don't accept this kind of proof, because it may take infinitely many steps to prove something; however, most mathematicians are happy to use it.

inequality

The statement that one quantity is less than or greater than another.

infinite dimensions

In mathematics, the concept of an infinite-dimensional space considered literally. It is a **vector space** with an infinite basis or a space with infinitely many coordinates.

infinite series

An infinite sum of the form

$$a_1 + a_2 + a_3 + \ldots = \sum_{k=1}^{\infty} a_k.$$

Such series crop up in many areas of modern mathematics. Their development began in the seventeenth century and was continued by Leonhard **Euler**, who, in the process, solved many important problems.

infinitesimal

A number that is greater than zero yet smaller than any positive **real number**. In a sense, infinitesimals are to small what **infinity** is to large. They were first introduced by Isaac **Newton** and Gottfried **Leibniz** in their early versions of **calculus**; however, the lack of a rigorous definition for them stood in the way of calculus being fully accepted. As Bertrand **Russell** later put it: "Calculus required continuity, and continuity was supposed to require the infinitely little; but nobody could discover what the infinitely little might be." In the 1800s, calculus was put on a firmer footing by Augustin **Cauchy**, Karl **Weierstrass**, and others, who clarified and redefined the notion of a **limit** without reference to infinitesimals. When a function $f(x)$ can be made as close as desired to L by taking x close enough to a, then L is the limit of $f(x)$ as x approaches a. This is the classical or *epsilon-delta* formulation of calculus, named for the common use of δ for $|x - a|$ and ε for $|f(x) - L|$. For a long time, this was the only rigorous foundation for calculus, and it is still the one taught in most calculus classes. But in 1960, Abraham **Robinson** discovered **nonstandard analysis**, which provides a rigorous formulation of infinitesimals, confers on them a new significance, and brings them closer to the vision of Newton and Leibniz.

infinity

> *Mystery has its own mysteries, and there are gods above gods. We have ours, they have theirs. That is what's known as infinity.*
>
> –Jean Cocteau (1889–1963),
> French author and filmmaker

A concept that has always fascinated philosophers and theologians, linked as it is to the notions of unending distance or space, eternity, and God, but that was avoided or met with open hostility throughout most of the history of mathematics. Only within the past century or so have mathematicians dealt with it head on and accepted infinity as a number–albeit the strangest one we know.

An early glimpse of the perils of the infinite came to Zeno of Elae through his paradoxes, the best known of which pits Achilles in a race against a tortoise (see **Zeno's paradoxes**). Confident of victory, Achilles gives the tortoise a head start. But then how can he ever overtake the sluggish reptile? First he must catch up to the point where it began, by which time the tortoise will have moved on. When he makes up the new distance that separated them, he finds his adversary has advanced again. And so it goes on, indefinitely. No matter how many times Achilles reaches the point where his competitor was, the tortoise has progressed a bit farther. So perplexed was Zeno by this problem that he decided not only was it best to avoid thinking about the infinite but also that motion was impossible!

A similar shock lay in store for **Pythagoras** and his followers who were convinced that everything in the universe could ultimately be understood in terms of whole numbers (even common fractions being just one whole number divided by another). The **square root of 2**–the length of the diagonal of a right-angled triangle whose shorter sides are both one unit long–refused to fit into this neat cosmic scheme. It was an **irrational number**, inexpressible as the ratio of two integers. Put another way, its decimal expansion goes on forever without ever settling into a recurring pattern.

These two examples highlight the basic problem in coming to grips with infinity. Our imaginations can cope with something that hasn't yet reached an end: we can always picture taking another step, adding one more to a total, or visualizing another term in a long series. But infinity, taken as a whole, boggles the mind. For mathematicians this was a particularly serious problem because mathematics deals with precise quantities and meticulously well-defined concepts. How could they work with things that clearly existed and went on indefinitely–a number like $\sqrt{2}$ or a curve that approached a line ever more closely–while avoiding a confrontation with infinity itself? **Aristotle** provided the key by arguing that there were two kinds of infinity. *Actual infinity,* or *completed infinity,* which he believed could not exist, is endlessness fully realized at some point in time. *Potential infinity,* which Aristotle insisted was manifest in nature–for example, in the unending cycle of the seasons or the indefinite divisibility of a piece of gold–is infinitude spread over unlimited time. This fundamental distinction persisted in mathematics for more than 2,000 years. In 1831, no less a figure than Karl **Gauss** expressed his "horror of the actual infinitude," saying: "I protest against the use of infinite magnitude as something completed, which is never permissible in mathematics. Infinity is merely a way of speaking, the true meaning being a limit which certain ratios approach indefinitely close, while others are permitted to increase without restriction."

By confining their attention to potential infinity, mathematicians were able to address and develop crucial concepts such as those of **infinite series**, **limit**, and **infinitesimal**s, and so arrive at **calculus**, without having to grant that infinity itself was a mathematical object. Yet as early as the Middle Ages certain paradoxes and puzzles arose, which suggested that actual infinity was not an issue to be easily dismissed. These puzzles stem from the principle that it is possible to pair off, or put in *one-to-one correspondence,* all the members of one collection of objects with all those of another of equal size. Applied to indefinitely large collections, however, this principle seemed to flout a commonsense idea first expressed by **Euclid**: the whole is always greater than any of its parts. For instance, it appeared possible to pair off *all* the positive integers with *only those that are even:* 1 with 2, 2 with 4, 3 with 6, and so on, despite the fact that positive integers also include odd numbers. Galileo, in considering such a problem, was the first to show a more enlightened attitude toward the infinite when he proposed that "infinity should obey a different arithmetic than finite numbers." Much later, David **Hilbert** offered a striking illustration of how weird the arithmetic of the endless can get.

Imagine, said Hilbert, a hotel with an infinite number of rooms. In the usual kind of hotel, with finite accommodation, no more guests can be squeezed in once all the rooms are full. But "Hilbert's Grand Hotel" is dramatically different. If the guest occupying room 1 moves to room 2, the occupant of room 2 moves to room 3, and so on, all the way down the line, a newcomer can be placed in room 1. In fact, space can be made for an infinite number of new clients by moving the occupants of rooms 1, 2, 3, etc, to rooms 2, 4, 6, etc, thus freeing up all the odd-numbered rooms. Even if an infinite number of coaches were to arrive each carrying an infinite number of passengers, no one would have to be turned away: first the odd-numbered rooms would be emptied as above, then the first coach's load would be put in rooms $3n$ for $n = 1, 2, 3, \ldots$, the second coach's load in rooms $5n$ for $n = 1, 2, \ldots$, and so on; in general, the people aboard coach number i would empty into rooms p_n where p is the $(i + 1)$th prime number.

Such is the looking-glass world that opens up once the reality of **set**s of numbers with infinitely many elements is accepted. That was a crucial issue facing mathematicians in the late nineteenth century: Were they prepared to embrace actual infinity as a number? Most were still aligned with Aristotle and Gauss in opposing the idea. But a few, including Richard **Dedekind** and, above all, Georg **Cantor**, realized that the time had come to put the concept of infinite sets on a firm logical foundation.

Cantor accepted that the well-known pairing-off principle, used to determine if two finite sets are equal, is just as applicable to infinite sets. It followed that there really are just as many even positive integers as there are positive integers altogether. This was no paradox, he realized, but the *defining property* of infinite sets: the whole is no bigger than some of its parts. He went on to show that the set of all positive integers, 1, 2, 3, . . . , contains precisely as many members–that is, has the same **cardinal number** or cardinality–as the set of all rational numbers (numbers that can be written in the form p/q, where p and q are integers). He called this infinite cardinal number **aleph-**null, $\aleph_0$ ("aleph" being the first letter of the Hebrew alphabet). He then demonstrated, using what has become known as *Cantor's theorem,* that there is a hierarchy of infinities of which $\aleph_0$ is the smallest. Essentially, he proved that the cardinal number of all the subsets–the different ways of arranging the elements–of a set of size $\aleph_0$ is a bigger form of infinity, which he called $\aleph_1$. Similarly, the cardinality of the set of subsets of $\aleph_1$ is a still bigger infinity, known as $\aleph_2$. And so on, indefinitely, leading to an infinite number of different infinities.

Cantor believed that $\aleph_1$ was identical with the total number of mathematical points on a line, which, astonishingly, he found was the same as the number of points on a plane or in any higher n-dimensional space. This infinity of spatial points, known as the *power of the continuum, c,* is the set of all **real number**s (all rational numbers plus all irrational numbers). Cantor's **continuum hypothesis** asserts that $c = \aleph_1$, which is equivalent to saying that there is no infinite set with a cardinality between that of the integers and the reals. Yet, despite much effort, Cantor was never able to prove or disprove his continuum hypothesis. We now know why–and it strikes to the very foundations of mathematics.

In the 1930s, Kurt **Gödel** showed that it is impossible to *disprove* the continuum hypothesis from the standard **axiom**s of set theory. Three decades later, Paul Cohen showed that it *cannot be proven* from those same axioms either. Such a situation had been in the cards ever since the emergence of **Gödel's incompleteness theorem**. But the independence of the continuum hypothesis was still unsettling because it was the first concrete example of an important question that could not be proven either way from the universally accepted system of axioms on which most of mathematics is built.

Currently, the preference among mathematicians is to regard the continuum hypothesis as being false, simply because of the usefulness of the results that can be derived this way. As for the nature of the various types of infinities and the very existence of infinite sets, these depend crucially on what number theory is being used. Different axioms and rules lead to different answers to the question *what lies beyond all the integers?* This can make it difficult or even meaningless to compare the various types of infinities that arise and to determine their relative size, although within any given number system the infinities can usually be put into a clear order. Certain extended number systems, such as the **surreal numbers**, incorporate both the ordinary (finite) numbers and a diversity of infinite numbers. However, whatever number system is chosen, there will inevitably be inaccessible infinities–infinities that are larger than any of those the system is capable of producing.[198, 214, 275]

inflection

A point of inflection of a plane curve is a point where the curve has a stationary **tangent**, at which the tangent is changing from rotating in one direction to rotating in the opposite direction.

information theory

A mathematical theory of information born in 1948 with the publication of Claude **Shannon**'s landmark paper, "A Mathematical Theory of Communication." Its main goal is to discover the laws governing systems designed to communicate or manipulate information, and to set up quantitative measures of information and of the capacity of various systems to transmit, store, and otherwise process information. Among the problems it treats are finding the best methods of using various communication systems and the best methods for separating the wanted information, or signal, from the noise. Another of its concerns is setting upper bounds on what it is possible to achieve with a given information-carrying medium (often called an information channel). The theory overlaps heavily with communication theory but is more oriented toward the fundamental limitations on the processing and communication of information and less oriented toward the detailed operation of the devices used.

injection

A **one-to-one** mapping.

inscribed angle

The angle formed by two **chords** of a curve that meet at the same point on the curve.

inscribed circle

A circle inside a triangle or other **polygon** whose edges are **tangents** to the circle. Such a polygon is said to be *circumscribed* to the circle. The (unique) circle inscribed to a triangle is called the *incircle.*

instability

The condition of a system when it is easily disturbed by internal or external forces or events, in contrast to a *stable system,* which will return to its previous condition when disturbed. A pencil resting vertically on its eraser or a coin standing on edge are examples of systems that have the property of instability since they easily fall over at the slightest breeze or movement of the surface they are resting on. An unstable system is one whose **attractors** can change; thus, instability is a characteristic of a system that is far from equilibrium or at **bifurcation**.

integer

Any positive or negative whole number or zero: . . ., –3, –2, –1, 0, 1, 2, 3, ... *Integer* is Latin for "whole" or "intact." The **set** of all integers is denoted by **Z**, which stands for *zahlen* (German for "numbers"). The integers are an extension of the **natural numbers** to include **negative numbers** and so make possible the solution of all equations of the form $a + x = b$, where a and b are natural numbers. Integers can be added and subtracted, multiplied, and compared. Like the natural numbers, the integers form a countably infinite set. However, the integers don't form a **field** since, for instance, there is no integer x such that $2x = 1$; the smallest field containing the integers is that of the **rational numbers**. An important property of the integers is *division with remainder:* given two integers a and b with $b \neq 0$, it is always possible to find integers q and r such that $a = bq + r$, and such that $0 \leq r < |b|$. q is called the *quotient* and r is called the *remainder* resulting from division of a by b. The numbers q and r are uniquely determined by a and b. From this follows the *fundamental theorem of arithmetic,* which states that integers can be written as products of prime numbers in an essentially unique way.

integral

The area, or a generalization of area, under any section of a graph that is described by a **function**; in other words, the continuous cumulative sum of a function (see **continuity**). Not all functions have an exact formula that allows an integral to be found. In such cases, *numerical integration* has to be used, in which the area is found using approximate numerical techniques. Integrals, together with **derivatives**, are the fundamental objects of **calculus**.

integral equation

An equation that involves an **integral** of the **function** which is to be solved for.

integration

An operation that corresponds to the informal idea of finding the area under the graph of a **function**. The first theory of integration was developed by **Archimedes** with his method of *quadratures,* but this could be applied only in circumstances where there was a high degree of geometric symmetry. In the seventeenth century, Isaac **Newton** and Gottfried **Leibniz** independently discovered the idea that integration was a sort of opposite of **differentiation** (which they had just invented); this allowed mathematicians to calculate a broad class of integrals for the first time. However, unlike Archimedes's method, which was based on **Euclidean geometry**, Newton's and Leibniz's integral calculus lacked a secure foundation.

In the nineteenth century, Augustin **Cauchy** finally developed a rigorous theory of **limit**s, and Bernhard **Riemann** followed this up by formalizing what is now called the **Riemann integral**. To define this integral, one fills the area under the graph with smaller and smaller rectangles and takes the limit of the sums of the areas of the rectangles at each stage. Unfortunately, some functions don't have well-defined limits to these sums, so they have no Riemann integral. Henri **Lesbesgue** invented another method of integration to solve this problem. He first presented his ideas in *Intégrale, Longueur, Aire* (Integral, length, area) in 1902. Instead of using the areas of rectangles, a method that puts the focus on the **domain** of the function, Lebesgue turned to the **codomain** of the function for his fundamental unit of area. Lesbesgue's idea was to build first the integral for what he called simple functions—functions that take only finitely many values. Then he defined it for more complicated functions as the upper bound of all the integrals of simple functions smaller than the function in question. *Lesbesgue integration* has the beautiful property that every function with a Riemann integral also has a Lebesgue integral, and the two integrals agree. But there are many functions with a Lebesgue integral that don't have a Riemann integral. As part of the development of Lebesgue integration, Lebesgue introduced the concept of *Lebesgue measure,* which measures lengths rather than areas. Lebesgue's technique for turning a measure into an integral generalizes easily to many other situations, leading to the modern field of **measure theory**.

The Lebesgue integral was deficient in one respect. The Riemann integral had been generalized to the *improper Riemann integral* to measure functions whose domain of definition was not a **closed** interval. The Lebesgue integral integrated many of these functions (always reproducing the same answer when it did), but not all of them. The *Henstock integral* is an even more general notion of integral (based on Riemann's theory rather than Lebesgue's) that subsumes both Lebesgue integration and improper Riemann integration. However, the Henstock integral depends on specific features of the real number line and so doesn't generalize as well as the Lebesgue integral does.

interesting numbers

Clearly some numbers are of greater interest (at least to mathematicians) than are others. The number **pi**, for instance, is far more interesting than 1.283—or virtually any other number for that matter. Confining our attention to integers, can there be such a thing as an uninteresting number? It is easy to show that the answer must be "no." Suppose there were a set U of uninteresting integers. Then it must contain a least member, u. But the property of being the smallest uninteresting integer makes u interesting! As soon as u is removed from U, there is a new smallest uninteresting integer, which must then also be excluded. And so the argument could be continued until U was empty. Given that all integers are interesting, can they be ranked from least to most interesting? Again, no. To be ranked as "least interesting" is an extremely interesting property, and thus leads to another logical contradiction!

When Srinivasa **Ramanujan**, the great Indian mathematician, was ill with tuberculosis in a London hospital, his colleague G. H. **Hardy** went to visit him. Hardy opened the conversation with: "I came here in taxicab number 1729. That number seems dull to me which I hope isn't a bad omen." Ramanujan replied, without hesitation: "Nonsense, the number isn't dull at all. It's quite interesting. It's the smallest number that can be expressed as the sum of two cubes in two different ways." ($1729 = 1^3 + 12^3$ and $9^3 + 10^3$.)

International Date Line

Is it possible to assign a time to every longitude on Earth, so that each longitude has a *different* time but the times at nearby longitudes are always close? The answer is no, which is mathematically equivalent to saying that there's no way to continuously map points on a **real number** line onto a circle. This explains why an International Date Line is needed. It allows most regions on Earth to have times similar to their neighbors, though, by convention, time changes are (usually) made in chunks of one hour between adjacent time zones. Then it takes care of the inevitable discontinuity by having it happen all at once, as a jump by one whole day, on a longitude that passes mostly through open water in the Pacific. The fact that there doesn't exist any continuous **one-to-one** function from the circle onto the line follows from the **Borsuk-Ulam theorem** in one dimension.

interpolate

To estimate the value of a point that lies between two known values of a **function**. This is often done by approximating a line or a smooth curve between the values, which is the literal meaning of the word. *Inter* is the

Latin prefix for "between," *polire* translates as "to adorn or polish," so together they mean "to smooth between." *Extrapolate* was created as an extension of interpolate to suggest the smoothing of a line *outside* the known points. This operation is often done in statistics when patterns are studied over time to predict future events.

intersection

A place where two or more things meet or overlap. Two lines intersect at a point, two planes can intersect in a line, and so forth. The intersection of two or more **sets**, represented by the symbol ∩, is the set of elements that all the sets have in common; in other words, all the elements contained in every one of the sets.

invariant

(1) Something that stays the same when a particular **transformation** is carried out. (2) A value that is unchanged when a particular **function** is applied. (3) In **topology**, a number, **polynomial**, or other quantity associated with a topological object such as a **knot** or 3-**manifold**, which depends only on the underlying object and not on its specific description or presentation.

invariant theory

The study of quantities that are associated with **polynomial** equations and that are left **invariant** under transformations of the variables. For example, the discriminant $b^2 - 4ac$ is an invariant of the **quadratic** form $ax^2 + bxy + cy^2$.

inverse

(1) The inverse of a number, or a **reciprocal**, is 1 divided by the number; for example, the inverse of 8 is ⅛ and the inverse of ⅗ is 5⁄3. (2) The inverse of a **function** or a **transformation** is the function or transformation that reverses the effect of the function or transformation. For example, the inverse of addition is subtraction, and of clockwise rotation is anticlockwise rotation. (3) The inverse of an element of a **set**, or a number, with respect to a particular operation, is what has to be combined with the element or number in order to obtain that operation's *identity element.*

involute

Attach a string to a point on a curve. Extend the string so that it is **tangent** to the curve at the point of attachment. Then wind the string up, always keeping it taut. The **locus** of points traced out by the end of the string is called the involute of the original curve, and the original curve is called the **evolute** of its involute. Although a curve has a unique evolute, it has infinitely many involutes corresponding to different choices of initial point. An involute can also be thought of as any curve **orthogonal** to all the tangents to a given curve. See also **circle involute**.

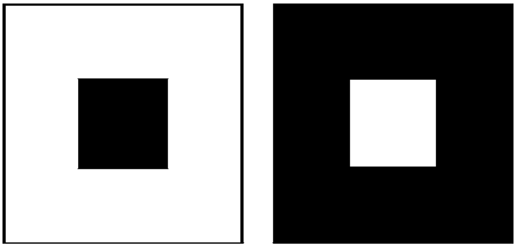

irradiation illusion Despite appearances, the inner squares are the same size.

irradiation illusion

A **distortion illusion** discovered by Hermann von Helmholtz in the nineteenth century. Despite the fact that the two figures are identical in size, the white hole looks bigger than the black one.

irrational number

A **real number** that can't be written as one whole number divided by another; in other words, a real number that isn't a **rational number**. The decimal expansion of an irrational number doesn't come to an end or repeat itself (in equal length blocks), though it may have a pattern such as 0.101001000100001000001 . . . . The vast majority of real numbers are irrational, so that if you were to pick a single point on the real **number line** at random the chances are overwhelmingly high that it would be irrational. Put another way, whereas the set of all rationals is countable, the irrationals form an **uncountable set** and therefore represent a larger kind of **infinity**. Indeed, as the Harvard logician Willard Van Orman Quine pointed out: "The irrationals exist in such variety . . . that no notation whatever is capable of providing a separate name for each of them."

There are two types of irrational number: **algebraic numbers**, such as the **square root of 2**, which are the roots of algebraic equations, and the **transcendental numbers**, such as **pi** and **e**, which aren't. In some cases it isn't known if a number is irrational or not; undecided cases include 2^e, π^e, $\pi^{\sqrt{2}}$, and the **Euler-Mascheroni constant**, γ. An irrational number raised to a rational power can be rational; for instance, $\sqrt{2}$ to the power 2 is 2. Also, an irrational number to an irrational power can be rational. What kind of number is $\sqrt{2}^{\sqrt{2}}$? The answer is irrational. This follows from Gelfond's theorem, which says that if *a* and *b* are roots of **polynomial**s, and *a* is not 0 or 1 and *b* is irrational, then a^b must be irrational (in fact, transcendental).

irreptile

See **rep-tile**.

Ishango bone

A bone tool handle discovered around 1960 in the African area of Ishango, near Lake Edward. It has been dated to about 9000 B.C. and was at first thought to have been a **tally** stick. At one end of the bone is a piece of quartz for writing, and the bone has a series of notches carved in groups on three rows running the length of the bone. The markings on two of these rows each add to 60. The first row is consistent with a number system based on 10, since the notches are grouped as 20 + 1, 20 – 1, 10 + 1, and 10 – 1, while the second row contains the **prime numbers** between 10 and 20! A third seems to show a method for multiplying by 2 that was used in later times by the Egyptians. Additional markings suggest that the bone was also used as a lunar phase counter. The Ishango Bone is kept at the Royal Institute for Natural Sciences of Belgium in Brussels. See also **Lebombo bone**.[45, 157]

isochrone

A set of points with the property that a given process or trajectory will take the same length of time to complete starting from any of the points. The curve formed by such a set of points is called an *isochronous curve*. See also **tautochrone problem**.

isogonal conjugate

Isogonal lines of a triangle are cevians (see **Ceva, Giovanni**) that are symmetric with respect to the angle bisector. Two points are isogonal conjugates if the corresponding lines to the vertices are isogonal (see **vertex**).

isometry

A symmetry operation, which may involve **translation**, **rotation**, and **reflection**, that preserves the distance of any two points. Each isometry of a **wallpaper group**, for example, can be represented by a 3 × 3 matrix.

Ishango bone The marks on this 11,000-year-old bone speak of a surprisingly sophisticated mathematical knowledge. *Museum of Natural Sciences (Brussels)*

isomorphism

In geometry, a **transformation** that doesn't alter the side lengths and the angle sizes of the figure involved. Examples of such transformations include reflections, rotations, translations, or transformations by a glide. In **set theory**, an isomorphism is a **one-to-one** correspondence between the elements of two sets such that the result of an operation on elements of one set corresponds to the result of the analogous operation on their images in the other set.

isoperimetric inequality

For any closed, three-dimensional body with volume V and surface area A, the following inequality always holds:

$$36\pi V^2 \leq A^3$$

isoperimetric problem

Among all shapes of a given perimeter, which encloses the greatest area? This ancient conundrum is alluded to in the tale of Queen Dido in Virgil's *Aeneid*. Threatened by her brother, who had already murdered her father, the queen was obliged to hastily gather her valuables and flee her native city of Tyria in ancient Phoenicia. In due course, her ship landed in North Africa, where she made the following offer to a local chieftain: in return for her fortune he would give as much land as she could isolate with the skin of an ox. This was readily agreed to, whereupon the crafty queen sliced the skin into very fine threads, which she tied together and made into a giant semicircle. Combined with the natural boundary provided by the sea, this enclosed such a large area that a city–Carthage–was eventually built upon it. Two millennia later, Karl **Weierstrass** was the first to prove rigorously, using analysis and calculus, that the solution to the isoperimetric problem is a circle (something the Greeks suspected but were not able to prove geometrically). When the same question is asked in one dimension higher it becomes the **isovolume problem**.

isosceles

Having two sides of the same length, as in the case of an isosceles **triangle**. An isosceles **trapezoid** in the United States is the equivalent of a trapezium in Britain. *Isosceles* comes from the Greek *iso* ("same or equal") and *skelos* ("legs").

isotomic conjugate

Two points on the side of a triangle are *isotomic* if they are equidistant from the midpoint of that side. Two points inside a triangle are isotomic conjugates if the corresponding cevians (see **Ceva, Giovanni**) through these points meet the opposite sides in isotomic points.

isovolume problem

What surface encloses the maximum volume per unit surface? The answer is a **sphere**, for which

$$\text{volume/surface area} = (\tfrac{4}{3}\pi r^3)/(4\pi r^2) = \tfrac{1}{3}r,$$

where r is the sphere's radius. Proof of the solution to the isovolume problem came as recently as 1882 from Hermann Schwarz. See also **bubbles** and **isoperimetric problem**.

iterate

Do something repeatedly. Do something repeatedly. . . .

iteration

> *If at first you don't succeed, try, try, again. Then quit. There's no use being a damn fool about it.*
>
> –W. C. Fields

A **feedback** process that repeats n number of times. Iteration refers to the act of performing the calculation of a certain **function** and then picking the result, or output, as the starting value, or input, for the next calculation of the same function. The operation repeats on and on–even infinitely, despite Fields's comment!

J

Jacobi, Karl Gustav Jacob (1804–1851)

A German mathematician who did important work on **elliptic functions**, **partial differential equations**, and mechanics. Although he was preceded in many of his discoveries about elliptic functions by Carl **Gauss** (who didn't publish) and Niels **Abel**, Jacobi is nevertheless considered one of the founders of the subject. His name is probably best known from the *Jacobian,* an $n \times n$ **determinant** formed from a set of n functions in n unknowns. He wasn't the first to use it–the "Jacobian" appears in an 1815 paper of Agustin **Cauchy**–but Jacobi did write a long memoir about it in 1841, and proved that the Jacobian of n functions vanishes if and only if the functions are related (Cauchy had proved the "if" part). He also did important work on partial differential equations and their application to physics. Along with William **Hamilton**, he developed an approach to mechanics based on generalized coordinates. In this method, the total energy of a mechanical system is represented as a function of generalized coordinates and corresponding generalized momenta; for example, in a double pendulum the two generalized coordinates could be two angles. *Hamilton-Jacobi theory* is the technique of solving the system by transforming coordinates so that the transformed coordinates and momenta are constants.

Jacobi was appointed to a position at the University of Königsberg in 1826. He gained a reputation as a gifted teacher and is credited with introducing the seminar method (giving lectures on his own ongoing research) into the university. After a collapse from overwork in 1843, Jacobi was allowed to stay in Berlin with a generous allowance from the king of Prussia. Five years later, revolution swept Europe and Jacobi was persuaded to run for parliament. This proved a disaster; not only did he loose the election, but his foray into politics annoyed his royal patron, who cut off his pension. Jacobi, with a large family to support, was faced with destitution. Only his reputation as the greatest German mathematician besides Gauss saved him; faced with the prospect of losing Jacobi to the University of Vienna, the king was prevailed upon to restore the pension. Jacobi was a notoriously hard worker (indeed, he had several breakdowns due to overwork), but his death in 1851 was the result of smallpox.

Johnson solid

Any convex **polyhedron** with regular faces that is not a **Platonic solid**, an **Archimedean solid**, or a **prism** (or an antiprism). There are 92 Johnson solids, which are named after Norman W. Johnson who was the first to catalog them in 1966. They include equilateral **deltahedra** and dipyramids (two pyramids placed symmetrically base to base) and any irregular convex solid that can be made by sticking together triangles, squares, pentagons, and so on, in a way that happens to close. Some of the simpler Johnson solids are assemblages of pyramids, prisms, and antiprisms; for example, a gyrobifastigium is two triangular prisms glued together with a twist. Others are fragments of Archimedean solids; for example, a pentagonal rotunda is half an icosidodecahedron. When stretched out with a prism this makes the marvelously named elongated pentagonal orthocupolarotunda.

Johnson's theorem

If three congruent circles all intersect in a single point, then the other three points of intersection will lie on another circle of the same radius. This simple little theorem was discovered by Roger Johnson in 1916.

Jordan, (Marie-Ennemond) Camille (1838–1922)

A French mathematician who made important contributions to **group** theory. He was the first to draw attention to the work of Evariste **Galois**, which had until then been almost entirely ignored. He built on Galois's study by grasping the intimate connection between groups of permutations and the solvability of **polynomial** equations. Jordan also introduced the idea of an infinite group. He passed on his interest in group theory to two of his most outstanding pupils, Felix **Klein** and Sophus **Lie**, both of whom went on to develop the subject in new and important ways.

Jordan curve

A simple, **closed curve**.

Jordan matrix

A **matrix** whose **diagonal** elements are all equal (and nonzero) and whose elements above the principal diagonal are equal to 1, but all other elements are 0.

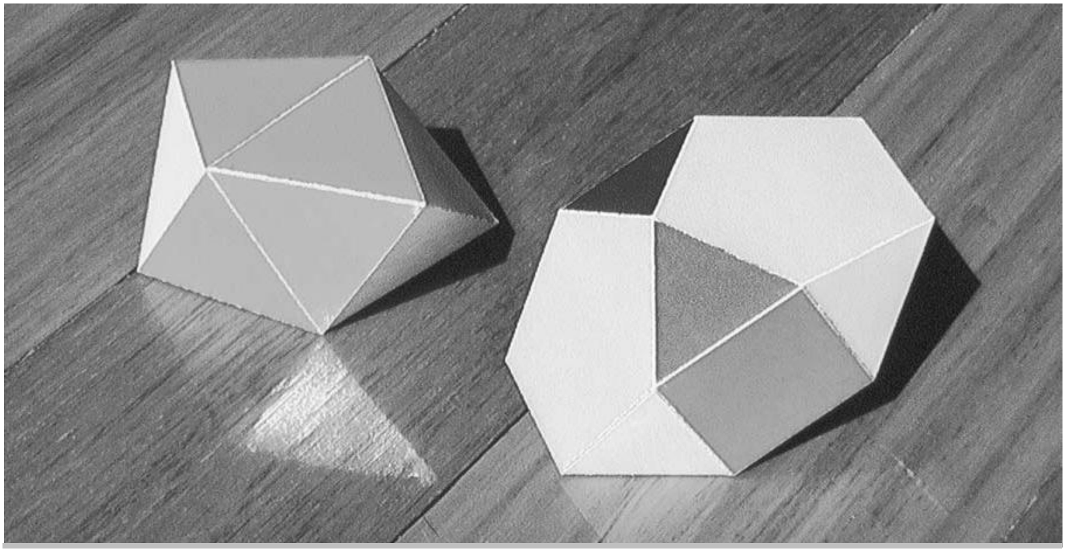

Johnson solid Two of the 92 Johnson solids: the bilunabirotunda (J91, right) and the snub disphenoid (J84, left). *Robert Webb, www.software3d.com; created using Webb's Stella program*

Journal of Recreational Mathematics

The only international journal devoted to the lighter side of mathematics. It was started in 1968 and is currently edited by Charles Ashbacher.

Julia set

Any of any infinite number of **fractal** sets of points on an **Argand diagram** (the complex plane) defined by a simple rule. Given two **complex number**s z and c, and the **recursion** $z_{n+1} = z_n^2 + c$, the Julia set for any given value of c, consists of all values of z for which z, when iterated in the equation above, does not "blow up" or tend to infinity. Julia sets are closely related to the **Mandelbrot set**, which is, in a way, an index of all Julia sets. For any point on the complex plane (which represents a value of c) a corresponding Julia set can be drawn. We can imagine a movie of a point moving about the complex plane with its corresponding Julia set. When the point lies inside the Mandelbrot set the corresponding Julia set is topologically unified or connected. As the point crosses the boundary of the Mandelbrot set, the Julia set explodes into a cloud of disconnected points called *Fatou dust.* If c is on the boundary of the Mandelbrot set, but not a waist point (where two large regions of the Mandelbrot set are connected by a narrow bridge), the Julia set of c looks like the Mandelbrot set in sufficiently small neighborhoods of c.

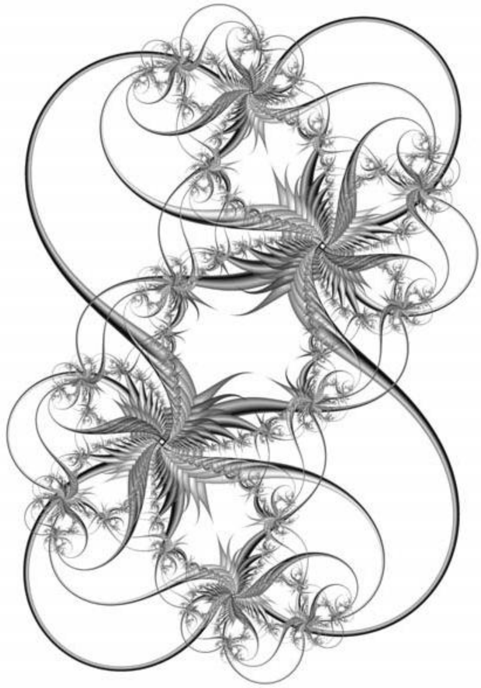

Julia set A fractal image based on a Julia set. *Jos Leys, www.josleys.com*

Julia sets are named after the French mathematician Gaston Julia (1893–1978), whose most famous work, *Memoire sur l'iteration des fonctions rationnelles,* provides the theory for Julia sets before computers were available. In 1918, at the age of 25 Julia was severely wounded in World War I and lost his nose. He wrote his greatest treatise in a hospital between the painful operations necessitated by his wounds.

jump discontinuity

A discontinuity in a **function** where the left- and right-hand limits exist but are not equal to each other.

K

Kaczynski, Theodore John (1942–)

Known as the Unabomber before his true identity was discovered, a Harvard graduate with a genius IQ who abandoned a promising career as a mathematician for a life of social isolation and intermittent terrorist attacks. Kaczynski eluded the FBI for 17 years, during which time he orchestrated 16 explosions that killed 3 people and injured 23 others. He avoided execution following his arrest in 1996 through a plea bargain and was given instead 4 life sentences plus 30 years in prison. During his mathematical career in the late 1960s, Kaczynski published a doctoral dissertation and several papers in academic journals and served as an assistant math professor at the University of California, Berkeley (1967–1969).

Kakeya needle problem

A famous problem named after the Japanese mathematician Sôichi Kakeya, who first posed it in 1917. It asks: What is the smallest-area plane figure inside which a unit straight-line segment can be rotated through 180°? For some years, the answer was thought to be a **deltoid**. However, in 1928, the Russian mathematician Abram Besicovitch shocked the mathematical world by showing that the problem *had no answer* or, to be more precise, that there was no minimum area.[39] In 1917 Besicovitch had been working on a problem in **Riemann integration**, and had reduced it to the question of existence of planar **set**s of **measure** 0, which contain a line segment in each direction. He then constructed such a set, and published his construction in a Russian journal in 1920. Due to the civil war and the blockade, there was hardly any communication between Russia and the rest of the world at the time, so Besicovitch hadn't heard about the challenge posed by Kakeya. Several years later, after he left Russia and learned about the needle problem, Besicovitch modified his original construction and was able to give the startling answer that the area in question may be made arbitrarily small.

Kaluza-Klein theory

A model that seeks to unite classical gravity and electromagnetism by resorting to **higher dimensions**. In 1919 the German mathematician Theodor Kaluza (1885–1954) pointed out that if general **relativity theory** is extended to a five-dimensional **space-time**, the equations can be separated out into ordinary four-dimensional gravitation plus an extra set (which is equivalent to Maxwell's equations for the electromagnetic field) plus an extra field known as the dilaton). Thus, electromagnetism is explained as a manifestation of curvature in a **fourth dimension** of physical space, in the same way that gravitation is explained in Einstein's theory as a manifestation of curvature in the first three. In 1926 the Swedish physicist Oskar Klein (1894–1977) proposed that the reason the extra spatial dimension goes unseen is that it is compact–curled up like a ball with a fantastically small radius. In the 1980s and 1990s, Kaluza-Klein theory experienced a big revival and can now be seen as a precursor of **string theory**.

Kampyle of Eudoxus

An hourglass-shaped curve that was first studied by **Eudoxus** in an attempt to solve the classical problem of **duplicating the cube**. It is described by the Cartesian equation

$$a^2x^4 = b^4(x^2 + y^2).$$

It is also the **radial curve** of the **catenary**.

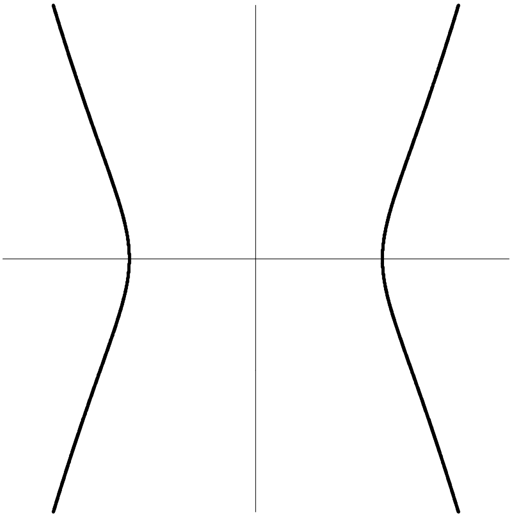

Kampyle of Eudoxus *© Jan Wassenaar, www.2dcurves.com*

kappa curve

Also known as *Gutschoven's curve,* named after Gérard van Gutschoven (1615–1668), who first studied it around 1662. The kappa curve, which resembles the Greek letter κ, was also investigated by Isaac **Newton** and, some years later, by Johann Bernoulli (see **Bernoulli family**). It is given by the Cartesian equation

$$y^2(x^2 + y^2) = a^2x^2.$$

Kaprekar constant

Take any four-digit number whose digits are not all identical. Rearrange the digits to make the largest and smallest 4-digit numbers possible. Subtract the smaller number from the larger. Use the resulting number and repeat the process. For example, starting with 4,731: 7,431 – 1,347 = 6,084; 8,640 – 468 = 8,172; 8,721 – 1,278 = 7,443; 7,443 – 3,447 = 3,996; 9,963 – 3,699 = 6,264; 6,642 – 2,466 = 4,176; 7,641 – 1,467 = 6,174. After this, the result is always 6174. Remarkably, *every* four-digit number whose digits are not all the same will eventually reach 6,174, in at most seven steps, and then stay there. This is called the Kaprekar constant for four-digit numbers, after the Indian mathematician Dattathreya Ramachandra Kaprekar who made the discovery in 1949. The Kaprekar constant for three-digit numbers is 495, which is arrived at for any three-digit number in no more than six iterations. The same process, or algorithm, can be applied to numbers of *n*-digits, where *n* is any whole number. Depending on the value of *n,* the algorithm will result in a nonzero constant, zero (the degenerate case), or a cycle.

Kaprekar number

Take a positive whole number n that has d number of digits. Take the square of n and separate the result into two pieces: a right-hand piece that has d digits and a left-hand piece that has either d or $d - 1$ digits. Add these two pieces together. If the result is n, then n is a Kaprekar number. Examples are 9 ($9^2 = 81$, $8 + 1 = 9$), 45 ($45^2 = 2{,}025$; $20 + 25 = 45$), and 297 ($297^2 = 88{,}209$; $88 + 209 = 297$). The first 10 Kaprekar numbers according to this definition are 1; 9; 45; 55; 99; 297; 703; 999; 2,223; and 2,728. Kaprekar numbers can also be defined by higher powers. For example, $45^3 = 91{,}125$; and $9 + 11 + 25 = 45$. The first 10 numbers with this property are: 1; 8; 10; 45; 297; 2,322; 2,728; 4,445; 4,544; and 4,949. For fourth powers, the sequence begins 1; 7; 45; 55; 67; 100; 433; 4,950; 5,050; 38,212. Notice that 45 is a Kaprekar number for second, third, and fourth powers ($45^4 = 4{,}100{,}625$; and $4 + 10 + 06 + 25 = 45$)–the only number in all three Kaprekar sequences, up to at least 400,000. See also **unique number**.

Kasner, Edward (1876–1955)

An American mathematician at Columbia University best remembered for introducing the words ***googol*** and *googolplex* into the popular mathematical lexicon. He is also well known as the coauthor, with James **Newman**, of *Mathematics and the Imagination,* first published in 1940.[186] In a later edition (1967), he spoke about the term *mathescope,* which was coined by the science reporter Wilson Davis after listening to one of Kasner's public lectures. In Kasner's words: "It is not a physical instrument; it is a purely intellectual instrument, the ever-increasing insight which mathematics gives into the fairyland which lies beyond intuition and beyond imagination." His main field of research was **differential geometry**, which he studied in its applications to mechanics, cartography, and stereographic projections, though he also wrote papers on circle packing and on the horn angle and studied an extension of right triangles to the complex plane.

Kepler, Johannes (1571–1630)

A German astronomer and mathematician best remembered for his three laws of planetary motion. He also did important work in optics, discovered two new regular polyhedra (1619), gave the first mathematical treatment of close **packing** of equal spheres (in 1611, leading to an explanation of the shape of the cells of a honeycomb), gave the first proof of how **logarithm**s work (1624), and devised a method of finding the volumes of **solids of revolution** that (with hindsight) can be seen as contributing to the development of **calculus** (1615–1616). Although his work eventually led to our modern view of the solar system, he himself held beliefs about the arrangement of the planets that smack of Pythagorean mysticism, though ultimately they stemmed from his belief that mathematical harmony was a reflection of God's perfection. In his first cosmological model (*Mysterium cosmographicum,* 1596) he suggested that if a sphere were drawn to touch the inside of the path of Saturn, and a cube were inscribed in the sphere, then the sphere inscribed in that cube would be the sphere circumscribing the path of Jupiter. Then if a regular tetrahedron were drawn in the sphere inscribing the path of Jupiter, the insphere of the tetrahedron would be the sphere circumscribing the path of Mars, and so inward, putting the regular dodecahedron between Mars and Earth, the regular icosahedron between Earth and Venus, and the regular octahedron between Venus and Mercury. Thus the number of (then-known) planets is explained perfectly in terms of the five convex regular solids–the **Platonic solid**s.

Kepler's second work on cosmology (*Harmonices mundi* [Harmony of the world], book V, 1619) offers a more elaborate mathematical model but still with the polyhedra in place. The mathematics in this work includes the first systematic treatment of **tessellation**s, a proof that there are only 13 convex uniform polyhedra (the **Archimedean solid**s) and the first account of two non-convex regular polyhedra (see **Kepler-Poinsot solid**s).

Kepler's mathematical work was full time and continued even during the wedding to his second wife in 1613. The dedicatory letter to a book he wrote shortly after explains that at the wedding celebration he noticed that the volumes of wine barrels were estimated by means of a rod slipped in diagonally through the bunghole, which prompted him to wonder how that could work. The result was a study of the volumes of solids of revolution, *Nova Stereometria Doliorum* (New stereometry of wine barrels, 1615).

Kepler-Poinsot solids

The four regular nonconvex polyhedra that exist in addition to the five regular convex polyhedra known as the **Platonic solid**s. As with the Platonic solids, the Kepler-Poinsot solids have identical regular **polygon**s for all their faces, and the same number of faces meet at each vertex. What is new is that we allow for a notion of "going around twice," which results in faces that intersect each other. In the *great stellated dodecahedron* and the *small stellated dodecahedron,* the faces are pentagrams (five-pointed stars). The center of each pentagram is hidden inside the polyhedron. These two polyhedra were described by Johannes **Kepler** in 1619, and he deserves credit for first understanding them mathematically, though a sixteenth-century drawing by the Nuremberg goldsmith Wentzel Jamnitzer (1508–1585) is very similar to the former and a fifteenth-century mosaic attributed to the Florentine artist Paolo Uccello (1397–1475) illustrates the latter. The *great icosahedron* and *great dodecahedron* were described by Louis **Poinsot** in 1809, though Jamnitzer made a picture of the great dodecahedron in 1568. In these two, the faces (20 triangles and 12 pentagons, respectively) that meet at each vertex "go around twice" and intersect each other, in a manner that is

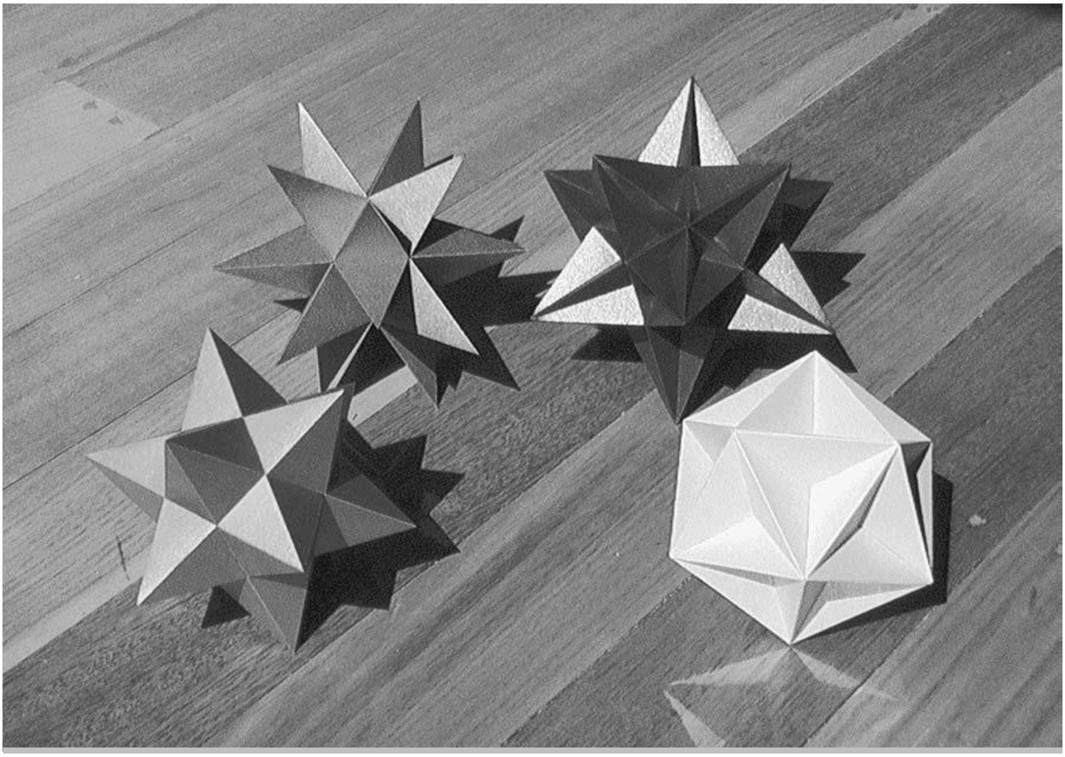

Kepler-Poinsot solids From left to right: the small stellated dodecahedron, great stellated dodecahedron, great icosahedron, and great dodecahedron. *Robert Webb, www.software3d.com; created using Webb's Stella program* the three-dimensional analog to what happens in two dimensions with a **pentagram**. Together, the Platonic solids and the Kepler-Poinsot polyhedra form the set of nine *regular* polyhedra. Augustin **Cauchy** first proved that no other polyhedra can exist with identical regular faces and identical regular vertices.

Kepler's conjecture

No **packing** of spheres of the same radius in three dimensions has a density greater than the face-centered (hexagonal) cubic packing. This claim was first published by Johannes **Kepler** in his monograph *The Six-Cornered Snowflake* (1611), a treatise inspired by his correspondence with Thomas Harriot (see **cannonball problem**). In his slender essay, Kepler asserted that face-centered cubic packing–the kind greengrocers use to stack oranges–is "the tightest possible, so that in no other arrangement could more pellets be stuffed into the same container." The question of whether Kepler's conjecture is right or not has became known, not surprisingly, as *Kepler's problem*. In the nineteenth century, Carl **Gauss** proved that face-centered cubic packing is the densest arrangement in which the centers of the spheres form a regular lattice, but he left open the question of whether an irregular stacking of spheres might be still denser. In 1953, László Tóth reduced the Kepler conjecture to an enormous calculation that involved specific cases, and later suggested that computers might be helpful for solving the problem. This was the approach taken by Thomas Hales, a mathematician at the University of Michigan at Ann Arbor, and which led him, in 1998, to claim that he had proved Kepler was right all along. Hales's proof of Kepler's conjecture remains controversial simply because of the length of the computer calculations involved and the difficulty of verifying them. The casebook on this mystery remains open. See also **monster group**.

Khayyam, Omar (1044–1123)

An outstanding Persian mathematician and astronomer, whose full name was Abu al-Fath Omar ben Ibrahim al-Khayyam, or "tent-maker"–possibly his father's profession. His work on algebra was known throughout Europe in the Middle Ages. He made several important contributions: he discovered a geometrical method to solve **cubic** equations by intersecting a **parabola** with a circle, discussed what would become known as **Pascal's triangle**, and asked if a ratio could be regarded as a number. He is best known, however, as a poet. In 1859, Edward Fitzgerald (1809–1883) translated Khayyam's *Rubaiyat*–a popular collection 600 short four-line poems. Rosehips from Khayyam's tomb were germinated at Kew Gardens, London, and planted on Fitzgerald's tomb in St. Michael's Churchyard, Boulge, Suffolk, in 1893; the original plant has died, but its descendents continue to bloom.

Khintchine's constant

One of the most remarkable, yet poorly understood, constants in mathematics, which captures, in a fascinating way, the behavior of *almost all* **real number**s. Pick a real number at random and write it down as a **continued fraction**. Almost certainly, the **geometric mean** of the terms in this continued fraction will be Khintchine's constant, which has the value 2.685452 . . . . It is important to say "almost certainly" because there are some real numbers including all **rational number**s, roots of third-order **polynomial**s, and certain other classes of number, whose continued fractions would give a different result. These exceptions, however, form a tiny minority of all real numbers.

kings problem

The problem of determining how many nonattacking kings, $K(n)$, can be placed on an $n \times n$ chessboard. For $n = 8$, the solution is 16. The general solution is $K(n) = \frac{1}{4}n^2$ if n is even and $K(n) = \frac{1}{4}(n + 1)^2$ if n is odd. The minimum number of kings needed to attack or occupy all squares on an 8×8 chessboard is nine.

kinship puzzles

Like **age puzzles and tricks**, problems to do with how family members are related go back many centuries. Some of these can be fiendishly convoluted, especially if incestual pairings are allowed (in the puzzle!). Sketching a genealogical tree is sometimes helpful. The following all involve legitimate ties.

PUZZLES

1. Brothers and sisters have I none, but that man's father is my father's son. Who is that man? (This is one of the oldest problems of the kinship variety.)
2. What is the simplest way in which two people can be an uncle to each other? (From Dudeney's *A Puzzle-Mine.*)
3. A certain family party consisted of 1 grandfather, 1 grandmother, 2 fathers, 2 mothers, 4 children, 3 grandchildren, 1 brother, 2 sisters, 2 sons, 2 daughters, 1 father-in-law, 1 mother-in-law, and 1 daughter-in-law. Twenty-three people, you will say. No; there were only seven people present. Can you show how this might be? (From *Amusements in Mathematics.*)

Solutions begin on page 369.

Kirkman, Thomas Penyngton (1806–1895)

An English rector and mathematician who did important work in **combinatorics**. His **schoolgirls problem**, which he posed in 1850, led to a more general study of certain ways of combining objects. He also explored the possibility of finding a route around the edges of a given **polyhedron** that passes through each vertex (corner) of the shape once and only once. His rather complex and ultimately unprovable ideas were later picked up and refined successfully by William **Hamilton**, but not be-fore Kirkman had managed to supply a general proof of the fact that "if a polyhedron has an odd number of vertices and each face has an even number of edges, then there is no circuit which passes through all the vertices." This result introduces the concept of a *bipartite graph*–a **graph** that can be divided into two separate sets of vertices such that every edge in one set joins a vertex.

kite

A **quadrilateral**, with two pairs of congruent adjacent sides, that is named after one of the traditional forms of toy kite. The toy itself probably draws its name from the bird commonly called a kite, or kyte, in England. The old English form of the word, cyte, may in turn derive from an early German name for an owl. A nonconvex kite is often called a *dart,* a term used by Roger **Penrose** in his proof on a nonperiodic tiling of the plane (see also **Penrose tiling**). **Proclus** referred to this shape as a "four-sided triangle" and spoke of it as a geometric paradox. John **Conway** pointed out that there is no special name for a quadrilateral that has two pairs of equal sides that, unlike in the case of a kite, are not parallel sides. He proposed the name *strombus* for such a figure, from the Greek for a spinning top.

Klein, Christian Felix (1849–1925)

A German mathematician noted for his work on **non-Euclidean geometry**, the connections between geometry and **group** theory, and the theory of functions. His Erlangen Programm (1872) for unifying the diverse forms of geometry through the study of equivalence in transformation groups was profoundly influential, especially in the United States, for over 50 years. In his *Lectures on the Icosahedron and the Solution of Equations of the Fifth Degree* (1884, tr. 1888) he showed how the rotation groups of regular solids could be applied to the solution of difficult algebraic problems. Klein was a professor of mathematics successively at the University of Erlangen, the Technical Institute in Munich, and the universities of Leipzig and Göttingen, and was a prolific writer and lecturer on the theory, history, and teaching of mathematics. See also **Klein bottle**.

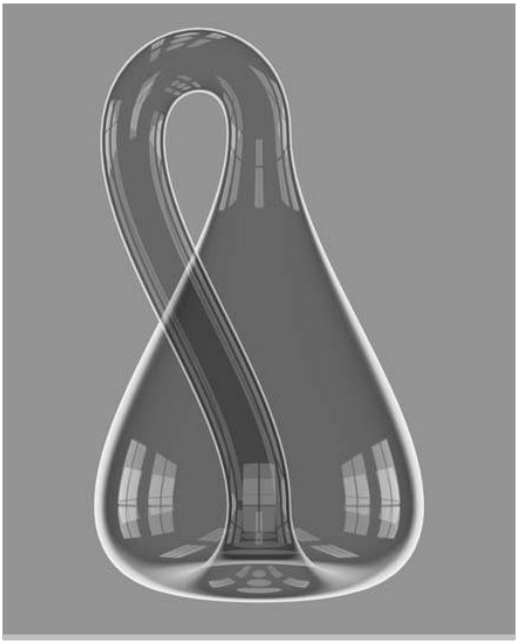

Klein bottle The famous container, with zero volume, rendered in the improbable form of a soap bubble. *© John M. Sullivan, University of Illinois and TU Berlin*

Klein bottle

A mathematician named Klein,
Thought the Möbius band was divine,
He said, "If you glue,
The edges of two,
You'll get a weird bottle like mine."

–Leo Moser

Take a rectangle and join one pair of opposite sides to make a cylinder. Now join the other pair with a half-twist. The result is a Klein bottle. Sound easy? It is if you have access to a **fourth dimension** because that's what is needed to carry out the second step and to allow the surface to pass through itself without a hole. A true Klein bottle is a four-dimensional object. It was discovered in 1882 by Felix **Klein** when he imagined, as in the limerick, joining two **Möbius band**s together to create a single-sided bottle with no boundary.

An ordinary (three-dimensional) bottle has a crease or fold around the opening where the inside and outside of the bottle meet. A sphere doesn't have this crease or fold, but it has no opening. A Klein bottle has an opening but no crease: like a Möbius band, it is a continuous one-sided structure. Because it has no crease or fold, there's no verifiable definition of where its inside and outside begin. Therefore, the volume of a Klein bottle is considered to be zero, and the bottle has no real contents–except itself! As the joke goes: "In topological hell the beer is packed in Klein bottles." Take a coin, slide it across the surface of a Klein bottle until it returns to its starting point, and the coin, as if by magic, will be flipped over. This is because, unlike a sphere or a regular bottle, a Klein bottle is nonorientable.

Although a Klein bottle can't be "embedded" (that is, fully realized) in three dimensions, it can be "immersed" in three dimensions. Immersion is what happens when a higher dimensional object cuts through a lower dimensional one, producing a cross section. When a sphere is immersed in a plane, for example, it produces a circle. The computer-generated soap bubble structure shown here could be created by stretching the neck of a bottle through its side and joining its end to a hole in the base. Except at the side-connection (the nexus), this properly shows the shape of a four-dimensional Klein bottle. Just as the picture here is of a three-dimensional Klein bottle immersion, so an immersion in real life is like a drawing of the true four-dimensional bottle.

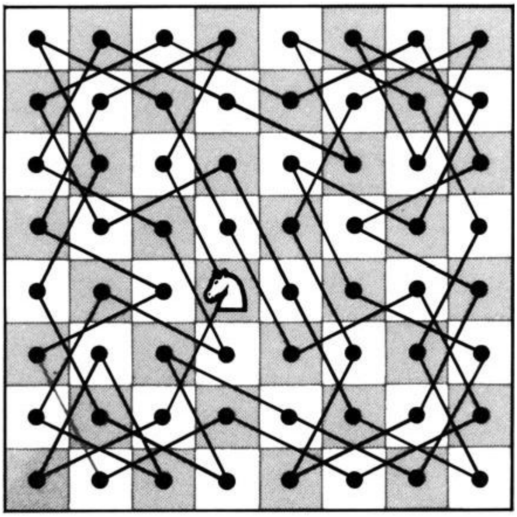

knight's tour After visiting every square of the chessboard exactly once, the knight returns to its starting point.

knights problem

To find the maximum number of knights that can be placed on an $n \times n$ chessboard in such a way that no two pieces are attacking each other. For a standard 8×8 chessboard, the answer is 32 (all on white squares or all on black). In the general, the solution is $\frac{1}{2} n^2$ if n is even and $\frac{1}{2} (n^2 + 1)$ if n is odd, giving the sequence 1, 4, 5, 8, 13, 18, ...

knight's tour

A classic **chess** puzzle: to find a sequence of moves by which a knight can visit each square of a chessboard exactly once. If the final position is a knight's move away from the first position, the tour is said to be *closed* or *reentrant.* The earliest recorded solution for a standard 8×8 chessboard was given by Abraham **de Moivre**; the earliest known reentrant solution came from the French mathematician Adrien-Marie Legendre (1752–1833). Not to be outdone Leonhard **Euler** found a reentrant tour that visits two halves of the board in turn. The problem can be generalized to an $n \times n$ board, with some surprising results; for example, a reentrant tour is not possible on a 4×4 board.

A knight's tour is called a **magic tour** if the resulting arrangement of numbers forms a **magic square**, and a *semimagic tour* if the resulting arrangement of numbers is a **semi-magic square**. It has long been known that magic knight's tours aren't possible on $n \times n$ boards if n is odd. It was also known that such tours are possible for all boards of size $4k \times 4k$ for $k > 2$. However, while a number of semi-magic knight's tours were known on the usual 8×8 chessboard, it was *not* known if any fully magic tours existed on the 8×8 board. This longstanding open problem was finally settled in the negative by an exhaustive computer enumeration of all possibilities. The software for the computation was written by J. C. Meyrignac, and a Web site was established by Guenter Stertenbrink to distribute and collect results for all possible tours. After 61.40 CPU-days, corresponding to 138.25 days of computation at 1 GHz, the project was completed on August 5, 2003. As well as netting a total of 140 distinct semi-magic knight's tours, the computation showed for the first time that no 8×8 magic knight's tour is possible.

More magical and mysterious tours can be conducted on boards on the surfaces of cubes, cylinders, toruses, and multidimensional shapes, such as hypercubes. See also **Hamilton path**.

knot

> *Never cut what you can untie.*
>
> –Joseph Joubert

A **closed** curve in three dimensions. The two simplest nontrivial knots are the trefoil knot, whose picture has three crossings, and the figure-eight knot, whose picture has four. To date, more than 1.7 million nonequivalent knots with pictures of 16 or fewer crossings have been identified. The mathematical theory of knots was born out of attempts to model the atom. Near the end of the nineteenth century, William Thomson (Lord Kelvin) suggested that different atoms were actually different knots tied in the ether that was believed to permeate all of space. Physicists and mathematicians set to work making a table of distinct knots, believing they were making a table of the elements. A pioneer in this effort, alongside Thomson, was Peter **Tait**. By the time the theory of the ether vanished into thin air, *knot theory* was firmly tied into mainstream mathematics. It blossomed with the development of **topology** and eventually led to important applications in DNA research and molecular biology. Today it is one of the most active areas of mathematical research. See also **tie knot**, **loop**, and **braid**.

Knuth's up-arrow notation

A notation for **large number**s developed by the American mathematician Donald Knuth in 1976. A single up-arrow ($\uparrow$) is the same as exponentiation:

$$m \uparrow n = m \times m \times \ldots \times m \ (n \text{ terms}) = m^n.$$

Two up-arrows together represent a **power tower**:

$$m \uparrow\uparrow n = m^{m^{\cdot^{\cdot^{m}}}}\Big\}n.$$

(a tower of height n), which is the same as the operation known as *hyper4* or *tetration*. This can very rapidly generate huge numbers. For example:

$2 \uparrow\uparrow 2 = 2 \uparrow 2 = 4$
$2 \uparrow\uparrow 3 = 2 \uparrow 2 \uparrow 2 = 2 \uparrow 4 = 16$
$2 \uparrow\uparrow 4 = 2 \uparrow 2 \uparrow 2 \uparrow 2 = 2 \uparrow 16 = 65{,}536$
$3 \uparrow\uparrow 2 = 3 \uparrow 3 = 27$
$3 \uparrow\uparrow 3 = 3 \uparrow 3 \uparrow 3 = 3 \uparrow 27 = 7{,}625{,}597{,}484{,}987$
$3 \uparrow\uparrow 4 = 3 \uparrow 3 \uparrow 3 \uparrow 3 = 3 \uparrow 3 \uparrow 27 = 3^{7625597484987}$

Three up-arrows together represent a still more vastly powerful operator, equivalent to *hyper5* or *pentation*, or a power tower of power towers:

$$m \uparrow\uparrow\uparrow n = m \uparrow\uparrow m \uparrow\uparrow \ldots \uparrow\uparrow m \ (n \text{ terms})$$

For example:

$2 \uparrow\uparrow\uparrow 2 = 2 \uparrow\uparrow 2 = 4$
$2 \uparrow\uparrow\uparrow 3 = 2 \uparrow\uparrow 2 \uparrow\uparrow 2 = 2 \uparrow\uparrow 4 = 65{,}536$
$2 \uparrow\uparrow\uparrow 4 = 2 \uparrow\uparrow 2 \uparrow\uparrow 2 \uparrow\uparrow 2 = 2 \uparrow\uparrow 65{,}536 = 2 \uparrow 2 \uparrow \ldots \uparrow 2$ (65,536 terms)
$3 \uparrow\uparrow\uparrow 2 = 3 \uparrow\uparrow 3 = 7{,}625{,}597{,}484{,}987$
$3 \uparrow\uparrow\uparrow 3 = 3 \uparrow\uparrow 3 \uparrow\uparrow 3 = 3 \uparrow\uparrow 7{,}625{,}597{,}484{,}987 = 3 \uparrow 3 \uparrow \ldots \uparrow 3$ (a power tower 7,625,597,484,987 layers high)
$3 \uparrow\uparrow\uparrow 4 = 3 \uparrow\uparrow 3 \uparrow\uparrow 3 \uparrow\uparrow 3 = 3 \uparrow\uparrow 3 \uparrow\uparrow 7{,}625{,}597{,}484{,}987 = 3 \uparrow\uparrow 3 \uparrow \ldots \uparrow 3$ (a tower $3 \uparrow\uparrow 7{,}625{,}597{,}484{,}987$ layers high)

Similarly,

$$m \uparrow\uparrow\uparrow\uparrow n = m \uparrow\uparrow\uparrow m \uparrow\uparrow\uparrow \ldots \uparrow\uparrow\uparrow m \ (n \text{ terms})$$

so that, for example:

$2 \uparrow\uparrow\uparrow\uparrow 2 = 2 \uparrow\uparrow\uparrow 2 = 4$
$2 \uparrow\uparrow\uparrow\uparrow 3 = 2 \uparrow\uparrow\uparrow 2 \uparrow\uparrow\uparrow 2 = 2 \uparrow\uparrow\uparrow 4 = 2 \uparrow 2 \uparrow \ldots \uparrow 2$ (65,536 terms)
$2 \uparrow\uparrow\uparrow\uparrow 4 = 2 \uparrow\uparrow\uparrow 2 \uparrow\uparrow\uparrow 2 \uparrow\uparrow\uparrow 2 = 2 \uparrow\uparrow\uparrow 2 \uparrow 2 \uparrow \ldots \uparrow 2$ (65,536 terms)
$3 \uparrow\uparrow\uparrow\uparrow 2 = 3 \uparrow\uparrow\uparrow 3 = 3 \uparrow 3 \uparrow \ldots \uparrow 3$ (7,625,597,484,987 terms)
$3 \uparrow\uparrow\uparrow\uparrow 3 = 3 \uparrow\uparrow\uparrow 3 \uparrow\uparrow\uparrow 3 = 3 \uparrow\uparrow\uparrow 3 \uparrow 3 \uparrow \ldots \uparrow 3$ (7,625,597,484,987 terms)
$= 3 \uparrow\uparrow 3 \uparrow 3 \uparrow \ldots \uparrow 3$ (3 $\uparrow 3 \uparrow \ldots \uparrow 3$ (7,625,597,484,987 terms)

Even up-arrow notation becomes cumbersome, however, when one is faced with staggeringly large numbers such as **Graham's number**. For such cases, more extensive systems such as **Conway's chained arrow notation** or *Steinhaus-Moser notation* are better suited. See also the **Ackermann function**, to which up-arrow notation is closely related.

Koch snowflake

One of the most symmetric and easy to understand **fractals**; it is named after the Swedish mathematician Helge von Koch (1870–1924), who first described it in 1906. To make the snowflake, start with a straight line and split it into three equal parts. Replace the middle part with two lines, both of the same length as the first three, creating an

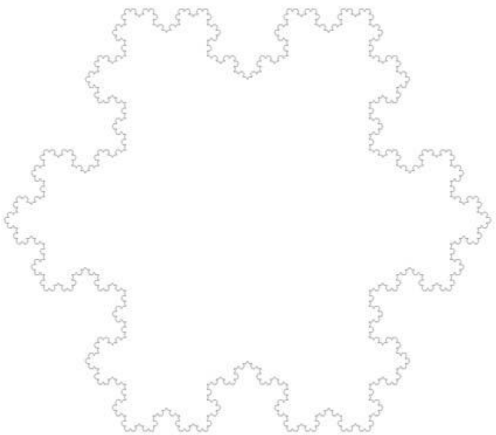

Koch snowflake This fractal shape emerges from continually sprouting equilateral triangles. *Xah Lee, www.xahlee.org* equilateral triangle missing the bottom line. The shape now consists of four straight lines with the same length. For each of these lines, repeat this process, then continue the transformation indefinitely. The Koch snowflake has infinite length because each iteration increases the length of a line segment one-third, and the iterations go on forever.

The same kind of process can be applied to a tetrahedron. Take a regular tetrahedron (all side lengths the same) and glue to each of its faces a smaller regular tetrahedron. (Each smaller tetrahedron is scaled down by a factor of ½ from the larger one, and placed on each face in an inverted fashion, so that it divides the face into four equilateral triangles and covers the center one.) Then iterate this process. Intuition suggests that the end product might be a very strange-looking jagged object. But in fact, in the limit, as the number of iterations tends to infinity, the result is a perfect cube! The cube has side length $t/\sqrt{2}$, where t is the length of one of the edges of the regular tetrahedron you started with. Variations on the flat Koch snowflake include the *exterior snowflake,* the *Koch antisnowflake,* and the *flowsnake* curves.

Kolmogorov, Andrei Nikolaievich (1903–1987)

A Russian mathematician and physicist who advanced the foundations of **probability theory**, the **algorithm**ic theory of **random**ness, and made crucial contributions to the foundations of statistical mechanics, **stochastic** processes, **information theory**, fluid mechanics, and nonlinear dynamics. All of these areas, and their interrelationships, underlie **complex system**s as they are studied today. His work on reformulating probability started with a 1933 paper in which he built up probability theory in a rigorous way from fundamental axioms, similar to Euclid's treatment of geometry. Kolmogorov went on to study the motion of the planets and turbulent fluid flows, publishing two papers in 1941 on turbulence that even today are of fundamental importance.

In 1954 he developed his work on dynamical systems in relation to planetary motion, thus demonstrating the vital role of probability theory in physics and reopening the study of apparent randomness in deterministic systems, along the lines originally conceived by Henri **Poincaré**. In 1965 he introduced the algorithmic theory of randomness via a measure of complexity, now referred to as *Kolmogorov complexity.* According to Kolmogorov, the complexity of an object is the length of the shortest computer program that can reproduce the object. Random objects, in his view, were their own shortest description, whereas periodic sequences have low Kolmogorov complexity, given by the length of the smallest repeating "template" sequence they contain. Kolmogorov's notion of complexity is a measure of randomness, one that is closely related to Claude **Shannon**'s entropy rate of an information source.

Königsberg bridge problem

See **bridges of Königsberg**.

Kronecker, Leopold (1823–1891)

A German mathematician and pioneer in the field of **algebraic number**s who formulated the relationship between the theory of numbers, the theory of equations, and **elliptic function**s. He acquired a passion for **number theory** from Ernst Kummer, his instructor at the Liegnitz Gymnasium. Kronecker, who made a fortune in business before returning to his academic studies, claimed that mathematical argumentation should be based only on integers and finite procedures. He was one of Georg **Cantor**'s sternest critics and refused to accept the validity of **Weierstrass's nondifferentiable function.**

L

labyrinth

See **maze**.

Lagrange, Joseph Louis (1736–1813)

An Italian-born French mathematician who made important contributions to **number theory** and to classical and celestial mechanics. By his mid-twenties he was recognized as one of the greatest living mathematicians because of his papers on wave propagation and the maxima and minima of curves. His prodigious output included his textbook *Mécanique Analytique* (Analytical mechanics, 1788), the basis for all later work in this field. His remarkable discoveries include the *Lagrangian,* a differential operator characterizing a system's physical state, and the *Lagrangian points,* points in space where a small body in the gravitational fields of two large ones remains relatively stable. Under Napoleon, Lagrange was made both a senator and a count; he is buried in the Panthéon.

lambda calculus

A model of computation that is capable of **universal computation**. The Lisp programming language was inspired by lambda calculus.

Lamé curve

Any of a family of curves related to the **ellipse** and that was first recognized and studied in 1818 by the French physicist and mathematician Gabriel Lamé (1795–1870). The formula for the Lamé curve family is a generalization of the equation for an ellipse ($|x/a|^2 + |y/b|^2 = 1$), namely:

$$|x/a|^n + |y/b|^n = 1,$$

where n is any real number. When $n = 0$, the curve reduces to a pair of crossed lines. As n increases, the curve changes from a curved star shape to a rectangle, with diagonals a and b, when $n = 1$. The special case when $n = 2/3$ corresponds to the **astroid**. Between $n = 1$ and $n = 2$ the curve turns from a curved rectangle to an ellipse (or a circle when both a and b are 1). For values of n greater than 2, Lamé curves are known as **superellipses**.

lamination

A decoration of a **manifold** in which some subset is partitioned into sheets of some lower dimension, and the sheets are locally parallel. It may or may not be possible to fill the gaps in a lamination to make a **foliation**.

Langley's adventitious angles

A seemingly simple problem first posed in 1922 by E. M. Langley[195] in connection with an isosceles triangle. In its original form, it is stated as follows: ABC is an isosceles triangle. B = C = 80°. CF at 30° to AC cuts AB in F. BE at 20° to AB cuts AC in E. Prove angle BEF = 30°. (No mention is made of D. Perhaps it is at the intersection of BE and CF.) A number of solutions appeared shortly after, including this one given by J. W. Mercer: Draw BG at 20° to BC, cutting CA in G. Then angle GBF = 60° and angles BGC and BCG are 80°. So BC = BG. Also angle BCF = angle BFC = 50°, so BF = BG and triangle BFG is equilateral. But angle GBE = 40° = angle BEG, so BG = GE = GF. And angle FGE = 40°, hence GEF = 70° and BEF = 30°.

Langton's ant

A type of **cellular automaton**, or a simple form of **artificial life**, named after its designer, Christopher Langton. The ant lives on an infinitely large chessboard, each square of which can be either black or white. Two pieces of information are associated with this digital insect: the

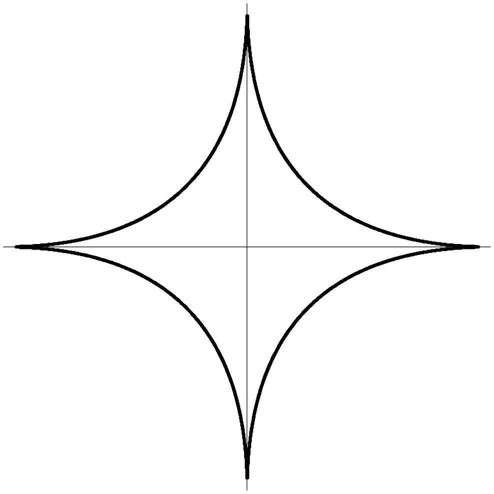

Lamé curve One example from the family of Lamé curves. © *Jan Wassenaar, www.2dcurves.com*

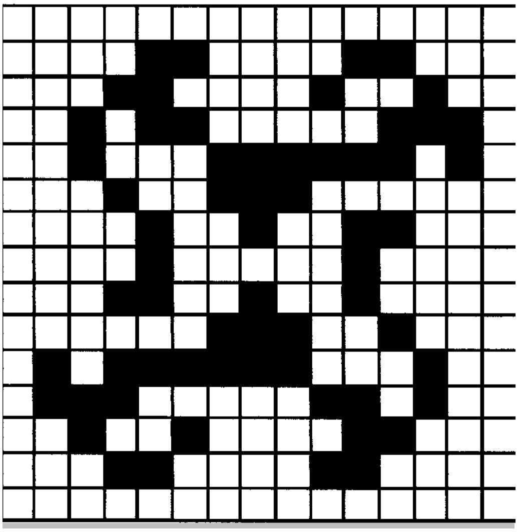

Langton's ant After 368 steps, a symmetrical pattern emerges.

direction that it's facing, and the state of the square that it currently occupies. Three simple rules govern the ant's behavior: (1) If it's on a black square, it turns left. (2) If it's on a white square, it turns right. (3) As it moves to the next square, the one it was on reverses color. Interest in Langton's ant stems from the fact that despite being a completely determined system governed by extremely simple rules, the patterns it produces are fantastically rich and complex. For the first 10,000 moves or so, the ant meanders around, building and then unbuilding structures with little pattern to them. Then, near the end of this chaotic phase, the ant begins to construct a diagonal highway off toward one edge of the board. In fact, this pattern stems from a sequence of 104 moves that, once started, will go on forever. In the language of **chaos** theory, the pattern is a *stable attractor* for the system. Remarkably, no matter what the initial arrangement of squares–even if the white and black squares are set up randomly–the ant will end up building a highway. The ant can be allowed to wrap around the edges of a finite board, thus allowing it to intersect its own path, and it will *still* end up building the highway. Are there any initial states that don't lead to the diagonal-road-building loop? No exceptions have been found from experiments–but proving it is another matter. Most mathematicians believe there is no general analytical method of predicting the position of the ant, or of any such chaotic system, after any given number of moves. Its behavior can't be reduced to the rules that govern it. In this sense, Langton's ant is a simple demonstration of the undecidability of the **halting problem**. The British mathematician Ian **Stewart** and biologist Jack Cohen, in their book *Figments of Reality*,[321] go a step further and use Langton's ant as an analog of an essential stage in the evolution of complex systems such as life: a stage in which the existence of chaotic behavior contains the potential for the spontaneous emergence of unpredictable forms of order.

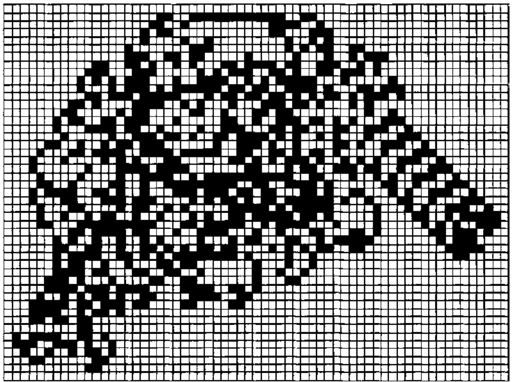

Langton's ant After 10,647 steps the "highway" is under construction.

Laplace, Pierre Simon (1749–1827)

A French mathematician and astronomer who was heavily involved with the development of celestial mechanics. He made an early impact by solving a complex problem of mutual gravitation that had eluded both Leonhard **Euler** and Joseph **Lagrange**. Laplace was among the most influential scientists of his time and was called the Newton of France for his study of and contributions to the understanding of the stability of the solar system. Laplace generalized the laws of mechanics for their application to the motion and properties of the heavenly bodies. He is also famous for his great treatises *Mécanique céleste* (Celestial mechanics, 1799–1825) and *Théorie analytique des probabilités* (Analytical theory of probabilities, 1812) which were advanced in large part by the mathematical techniques that Laplace developed early in his life.

large number

> *Bigger than the biggest thing ever and then some.*
> *Much bigger than that in fact, really amazingly*
> *immense, a totally stunning size, real "wow, that's*
> *big!" time. . . . Gigantic multiplied by colossal multiplied by staggeringly huge is the sort of concept*
> *we're trying to get across here.*
>
> –Douglas Adams, *The Restaurant at the End of the Universe*

Making, naming, and representing very large numbers is itself a big problem. A simple way to start is by adding zeros: 10; 100; 1,000; 10,000; . . . , 1,000,000; . . . . But this quickly gets tedious and exponentiation becomes a more attractive option: 10^1, 10^2, 10^3, . . . , 10^6, ... Naming the various powers of 10 follows a regular pattern of prefixes. In the United States, 10^3 is one thousand, 10^6 is one million, 10^9 is one billion, 10^{12} is one trillion, 10^{15} is one quadrillion, and so on. The "-illion" root kicks in at the sixth power of 10 prefixed by "m" for "mono"; then, every jump of three powers (factor of a thousand) comes the next prefix. Put another way, the U.S. name for 10^{3n} uses the Latinized prefix for $n - 1$. One centillion, which is the largest number with a single-word name in English, is 10^{303}. Elsewhere in the world, "billion," "trillion," and so on, can mean things other than they do in the U.S. system. A British billion, for example, is one million million, or 10^{12}, while the now largely obsolete term *milliard* was used for a thousand million. However, the American forms have become fairly standard internationally and are used without qualification in this book. It's also worth noting that, while "quadrillion," "quintillion," and so forth, are perfectly valid terms, "one thousand trillion," "one million trillion," and so forth, are generally preferred.

Exponentiating quite small numbers seems at first to be a pretty economical way of making and writing large numbers: 10^{30}, for example, is a highly effective shorthand for 1,000,000,000,000,000,000,000,000,000,000. But this method runs out of steam as the numbers get bigger and bigger. Take, for instance, the **googol** and the googolplex. One googol is the unofficial name for 10^{100}, or 1 followed by 100 zeros. This innocuous-looking number is larger than the number of atoms in the universe. What happens, therefore, if we want to represent the number that is 1 followed by *a googol number of zeros?* One way would be to write "1 followed by a googol number of zeroes"! But this is cheating because it couldn't be generalized without first giving a proper name to a fantastically large number of numbers. A better solution is to exponentiate using large numbers. Thus, 1 googolplex = $10^{\text{googol}} = 10^{10100}$. This is the beginning of a **power tower**.

Are numbers as large as the googol, not to mention the googolplex, of any practical importance? Science has given us such colossi as the **Avogadro constant** (the number of molecules in a sample whose weight in grams equals its molecular weight) = 6.023×10^{23}, the **Eddington number** (astrophysicist Arthur Eddington's best estimate of the number of protons in the universe) = 1.575×10^{79}, and the Supermassive Black Hole Evaporation Time = 10^{100} years (or thereabouts)–which brings us to the level of the googol. But there's nothing known or that can be reasonably conjectured in the "real" world of physics that goes much beyond this.

Science fiction can carry us a bit further. In *The Hitchhiker's Guide to the Galaxy*[5] by Douglas Adams appears one of the largest numbers ever used in a work of fiction: 2^{260199}. These are the odds quoted against the characters Arthur Dent and Ford Prefect being rescued by a passing spaceship just after having been thrown out of an airlock. As it happens, they *are* rescued by a spaceship powered by the "infinite improbability drive." By contrast, some special numbers in mathematics make even the googolplex look tiny. **Skewes' number**, $10^{10^{10^{34}}}$, was long held up as an example of a googolplex-dwarfing number that nevertheless served a bonafide purpose in mathematics. However, even this seemingly immense integer is made to look ridiculously small by the likes of more recently described numbers, such as **Graham's number**, the *Mega*, and the *Moser*, which are so utterly vast that it takes a page or two simply to describe the various special notations used to represent them.

Just as writing out a number in full, or in place-value notation, becomes unwieldy with numbers as big as a googol, so exponentiation, in turn, endangers the world's forests if it tries to take on seriously large numbers. A more effective shorthand is *tetration–tetra* (from the Greek meaning "four") because it is the fourth dyadic operation in the series: addition, multiplication, exponentiation, tetration. Dyadic means that two numbers, or arguments, are involved in the operation. Multiplication is repeated addition (e.g., $2 \times 3 = 2 + 2 + 2$), exponentiation is repeated multiplication (e.g., $2^3 = 2 \times 2 \times 2$), and tetration is repeated exponentiation. For example, 2 tetrated to 3, represented as ${}^{3}2$, is $2^{2^2} = 2^4 = 16$; 2 tetrated to 4, or ${}^{4}2$, is $2^{2^{2^2}} = 2^{16} = 65{,}536$; and 2 tetrated to 5, or ${}^{5}2$, is $2^{2^{2^{2^2}}} = 2^{65{,}536}$ = something too big to write out in full. Tetration goes by various other names including *superpower, superdegree,* and, the one used most commonly in mathematical circles and also here, *hyper4*.

Just as the exponentiation of two numbers, a and b, is represented as a^b and defined as $a \times a \times \ldots \times a$ (b terms), the hyper4 of a and b is represented as $a^{(4)}b$ and defined as $a^{a^{a^{\cdots^{a}}}}$ (a power tower with b levels). Alternatively, the hyper4 operator can be represented in **Knuth's up-arrow notation** as $a \uparrow\uparrow b$. Continuing this pattern:

hyper5 of a and $b = a^{(5)}b = a^{(4)}a^{(4)} \ldots a^{(4)} = a \uparrow\uparrow\uparrow b$
hyper6 of a and $b = a^{(6)}b = a^{(5)}a^{(5)} \ldots a^{(5)} = a \uparrow\uparrow\uparrow\uparrow b$
hyper7 of a and $b = a^{(7)}b = a^{(6)}a^{(6)} \ldots a^{(6)} = a \uparrow\uparrow\uparrow\uparrow\uparrow b$

To get some idea of the potency of this kind of representation, consider the sequence

1

10

10,000,000,000

10,000,000,000,000,000,000,000,000,000,000,000,000,000,
000,000,000,000,000,000,000,000,000,000,000,000,000,
000,000,000,000,000,000,000,000,000 (100 zeroes)

. . .

Notice how big even the fourth term is? The seemingly innocuous-looking number 5 ↑↑↑↑↑ 5 (or $5^{(7)}5$) is so huge that it would be around the 100,000,000,000,000,000th (one hundred thousand trillionth) term of this sequence!

Since the dyadic operators discussed previously form a pattern, they can be telescoped into a single *triadic* operator that has three arguments. This can be defined as:

$$\mathrm{hy}(a, n, b) = \begin{cases} 1 + b \text{ for } n = 0 \\ a + b \text{ for } n = 1 \\ a \times b \text{ for } n = 2 \\ a \uparrow b \text{ for } n = 3 \\ a \uparrow \mathrm{hy}(a, 4, b-1) \text{ for } n = 4 \\ \mathrm{hy}(a, n-1, \mathrm{hy}(a, n, b-1)) \text{ for } n > 4 \\ a \text{ for } n > 1, b = 1 \end{cases}$$

Beyond hyper are other triadic operators capable of generating large numbers even faster. The **Ackermann function** and the *Steinhaus-Moser notation* are both equivalent to a triadic operator that is somewhat more powerful than hy(*a*, *n*, *b*). Similarly, **Conway's chained-arrow notation** marks an evolution of Knuth's symbolism. These various techniques and notations can produce immense *finite* numbers. But beyond any of these lie the many different kinds of **infinity**.

lateral inhibition illusion

See **Hermann grid illusion**.

Latin square

An $n \times n$ square grid, or matrix, whose entries consist of n symbols such that each symbol appears exactly once in each row and each column. The following are some examples:

$$\begin{array}{cc} 1 & 2 \\ 2 & 1 \end{array} \quad \begin{array}{ccc} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{array} \quad \begin{array}{cccc} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \\ 4 & 1 & 2 & 3 \end{array} \quad \begin{array}{cccc} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \\ 3 & 4 & 1 & 2 \\ 4 & 3 & 2 & 1 \end{array} \quad \begin{array}{ccccc} \mathrm{M} & \mathrm{A} & \mathrm{G} & \mathrm{I} & \mathrm{C} \\ \mathrm{G} & \mathrm{I} & \mathrm{C} & \mathrm{M} & \mathrm{A} \\ \mathrm{C} & \mathrm{M} & \mathrm{A} & \mathrm{G} & \mathrm{I} \\ \mathrm{A} & \mathrm{G} & \mathrm{I} & \mathrm{C} & \mathrm{M} \\ \mathrm{I} & \mathrm{C} & \mathrm{M} & \mathrm{A} & \mathrm{G} \end{array}$$

Latin squares have a long history, stretching back at least as far as medieval Islam (c. 1200), when they were used on amulets. Abu l'Abbas al Buni wrote about them and constructed, for example, 4 × 4 Latin squares using letters from a name of God. In his famous etching *Melancholia,* the fifteenth-century artist Albrecht **Dürer** portrays a 4 × 4 **magic square**, a relative of Latin squares, in the background. Other early references to them concern the problem of placing the 16 face cards of an ordinary deck of **cards** in the form of a square so that no row, column, or diagonal contains more than one card of each suit and each rank. Leonhard **Euler** began the systematic treatment of Latin squares in 1779 and posed a problem connected with them, known as the **thirty-six officers problem**, that wasn't solved until the beginning of the twentieth century. Arthur **Cayley** continued work on Latin squares, and in the 1930s the concept arose again in the guise of multiplication tables when the theory of quasigroups and loops began to be developed as a generalization of the **group** concept. Latin squares played an important role in the foundations of finite geometries, a subject that was also in development at that time. Also in the 1930s, a large application area for Latin squares was opened by R. A. Fisher who used them and other combinatorial structures in the design of statistical experiments.

lattice

A periodic arrangement of points such as the vertices (see **vertex**) of a **tiling** of space by cubes or the positions of atoms in a crystal. More technically, a discrete Abelian subgroup (see **Abelian group**) of an n-dimensional **vector space** that is not contained in an $(n-1)$-dimensional vector space. Lattices play a central role in the theory of **Lie group**s, in **number theory**, in error-correcting codes, and many other areas of mathematics. See also **geoboard**.

lattice path

A sequence of points in a **lattice** such that each point differs from its predecessor by a finite list of allowed steps. **Random** lattice paths are an interesting model for the random motion of a particle and lattice paths are also important in enumerative **combinatorics**.

lattice point

A point with **integer** coordinates.

latus rectum

A **chord** of an **ellipse** that passes through a **focus** and is perpendicular to the major axis of the ellipse. Its plural is latera recta.

league

An archaic unit of traveling distance. The precise value varied, but was usually around 3 miles (4.8 kilometers).

least common multiple

The smallest **integer** that is an exact multiple of every number in a set of integers.

least upper bound

The smallest number that is larger than every member of a set of numbers.

Lebesgue, Henri Leon (1875–1941)

A French mathematician who introduced the modern definition of an **integral**. Lebesgue graduated from the École Normale Supériere and, from 1921, taught at the College de France. He and Emile **Borel** founded the modern theory of **function**s of a real variable, Lebesgue's great contribution being his new general definition of an integral (1902), which became known as the *Lebesgue integral* (see **integration**). This led to important advances in calculus, curve rectification, and trigonometric series, and, in Borel's hands, marked the start of **measure theory**. Although the Lebesgue integral was an example of the power of generalization, Lebesgue himself wasn't a fan of generalization and spent the rest of his life working on very specific problems, mostly in **analysis**.

Lebombo bone

One of the oldest mathematical artifacts known, a small piece of the fibula of a baboon, found near Border Cave in the Lebombo Mountains between South Africa and Swaziland. Discovered in the 1970s during excavations of Border Cave and dated about 35,000 B.C., the Lebombo bone is marked with 29 clearly defined notches. This suggests it may have been used as a lunar phase counter, in which case African women may have been the first mathematicians, because keeping track of menstrual cycles requires a lunar calendar. Certainly, the Lebombo bone resembles calendar sticks still used by Bushmen in Namibia. See also **Ishango bone**.

left-right reversal

See **mirror reversal problem**.

Leibniz, Gottfried Wilhelm (1646–1716)

A German philosopher, mathematician, and statesman who developed differential and integral **calculus** independently of Isaac **Newton**. He also invented a calculating machine, is considered a pioneer in mathematical logic, and proposed the metaphysical theory that we live in "the best of all possible worlds." In Leibniz's philosophical view, the universe is composed of countless conscious centers of spiritual force or energy known as *monads*. Leibniz talks about the "compossible" elements of any possible world–elements that allow a logically consistent structure. Though one of the finest minds of his age, Leibniz was not immune to blunders: he thought it just as easy to throw 12 with a pair of dice as to throw 11.

Leibniz harmonic triangle

A triangle of fractions that is related to the more famous **Pascal triangle** in a very simple way. Each row of the Leibniz harmonic triangle starts with the **reciprocal** of the row number (or the row number plus one depending on whether one starts counting from 1 or 0). Every entry is the sum of the two numbers just below it. The entries can thus be computed sequentially left to right and top to bottom using subtraction instead of addition.

1/1
1/2 1/2
1/3 1/6 1/3
1/4 1/12 1/12 1/4
1/5 1/20 1/30 1/20 1/5

lemma

A short auxiliary proposition used in the proof of a larger **theorem**.

lemniscate of Bernoulli

A curve "shaped like a figure 8, or a knot, or the bow of a ribbon" in the words of Jacob Bernoulli (see **Bernoulli family**) in an article published in 1694. Bernoulli named the curve "lemniscate" after the Greek *lemniskus* for a pendant ribbon (the type fastened to a victor's garland). It has the Cartesian equation

$$(x^2 + y^2)^2 = a^2(x^2 - y^2).$$

At the time he wrote his article, Bernoulli wasn't aware that the curve he was talking about was a special case of a **Cassinian oval**, which had been described by Cassini in 1680. The general properties of the lemniscate were established by Giovanni Fagnano (1715–1797) in 1750; Leonhard **Euler**'s investigations of the length of arc of the curve (1751) led to later work on **elliptic function**s. There is a relationship between the lemniscate and the rectangular **hyperbola**. If a **tangent** is drawn to the hyperbola and the perpendicular to the tangent is drawn through the origin, the point where the perpendicular meets the tangent is on the lemniscate. See also **hippopede.**

Leurechon, Jean (c. 1591–1670)

A French Jesuit priest and mathematician who wrote *Recreations mathématiques* (1624) under the pseudonym Henrik van Etten. Much of the mathematical content centers around Claude **Bachet**'s problems and may have been copied from it or some common source. The book also gives the earliest known description of the operation of an ear trumpet.

liar paradox

"This statement is false." What do we make of this statement (call it *S*)? If *S* is true, then *S* is false. On the other hand, if *S* is false, then it is true to say *S* is false; but, because the liar sentence is saying precisely that (namely that it is false), *S* is true. So *S* is true if and only if it is false. Since *S* is one or the other, it is both! Debate about sentences like *S* has been going on among philosophers and logicians for more than 2,000 years without any clear resolution.

The roots of the liar paradox stretch back to the philosopher Epimenides in the sixth century B.C. He said: "All Cretans are liars. . . . One of their own poets has said so." Another version of this can be found in the Bible, Titus 1:12–13, "Even one of their own prophets has said, 'Cretans are always liars, evil brutes, lazy gluttons.' This testimony is true." The poet's (or prophet's) statement is sometimes wrongly considered to be paradoxical because he himself is a Cretan. But actually there's no paradox here. A "liar," in everyday language, is someone who *on occasion* knowingly gives false answers. This isn't problematic: the poet, while lying occasionally, this time spoke the truth. However, most formulations of logic define a "liar" as an entity that always produces the negation of the true answer, that is, someone who does nothing but lie. Thus, the poet's statement can't be true: if it were, then he himself would be a liar who just spoke the truth, but liars don't do that. However, no contradiction arises if the poet's statement is taken to be false: the negation of "All Cretans are liars" is "Some Cretans aren't liars," in other words: some Cretans sometimes speak the truth. This doesn't contradict the fact that our Cretan poet just lied. Therefore, the statement "All Cretans are liars," if uttered by a Cretan, is necessarily false, but not paradoxical. Even the statement "I am a liar" is not paradoxical; depending on the definition of "liar" it may be true or false. However, the statement "I am lying now," first attributed to Eubulides of Miletus in the fourth century B.C., definitely is paradoxical. It is exactly equivalent to the sentence, we started with: "This statement is false."

Various elaborations of the basic Eubulides liar paradox have appeared over the ages. In the fourteenth century, the French philosopher Jean Buridan applied it in his argument for the existence of **God**. In 1913, the English mathematician Philip Jourdain (1879–1921) offered a version that is sometimes referred to as "Jourdain's card paradox." On one side of a card is written:

> The sentence on the other side of this card is true.

On the other side is written:

> The sentence on the other side of this card is false.

Yet another popular version of the liar paradox, guaranteed to perplex, is given by the following three sentences written on a card:

1. This sentence contains five words.
2. This sentence contains eight words.
3. Exactly one sentence on this card is true.

Lie, Marius Sophus (1842–1899)

A Norwegian mathematician who, along with his close friend Felix **Klein**, introduced **group** theory into geometry and used it to classify geometries. Lie discovered the *contact transformation*, which maps curves into surfaces (1870), and **Lie group**s, which use continuous or infinitesimal transformations. He used these groups to classify **partial differential equation**s, making the traditional methods of solution all reduce to a single principle. Lie groups also provided a basis for the growth of modern **topology**. See also **Lie algebra**.

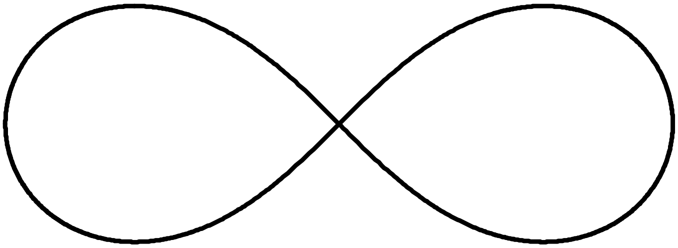

lemniscate of Bernoulli © Jan Wassenaar, www.2dcurves.com

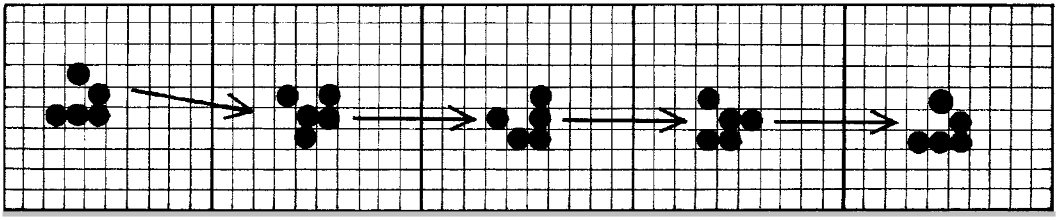

Life, Conway's game of A "glider" advances one cell down and one cell to the right.

Lie algebra

An **algebra**, named after M. Sophus **Lie**, in which multiplication satisfies properties similar to the *bracket operation* on matrices given by $[A, B] = AB - BA$, where the operation on the right-hand side involves ordinary multiplication and subtraction of matrices. The operation is not **associative**.

Lie group

A **group**, named after M. Sophus **Lie**, that is also a **manifold**. Groups of a real **matrix**, such as occur in quantum field theory, give naturally occurring examples of Lie groups. The tangent space at the identity element of a Lie group forms a **Lie algebra** in a natural way.

Life, Conway's game of

The best known example of a **cellular automaton**; it was invented by John **Conway** and first brought to public attention in Martin Gardner's *Scientific American* column in October 1970.[120] Conway's goal in creating Life was to devise a universal **Turing machine**–a sort of infinitely programmable computer. John **von Neumann** had described such a system in the 1950s, but it had very complex rules; Conway wanted to find one that was much simpler to describe and to operate. Life is played on a grid of squares on which each cell is either alive (occupied) or dead (empty). The game starts from an arbitrary initial configuration of live cells, and then progresses through generations as the life and death rules are applied. These rules are very simple: (1) A live cell survives to the next generation if it has two or three neighbors. (2) A live cell dies if it has four or more neighbors (overcrowding) or if it has only one neighbor or none (isolation). (3) A dead cell becomes a live cell in the next generation if it has exactly three neighbors (birth). The rules of Life were developed over a two-year-period during tea and coffee breaks by Conway and a group of graduate students and colleagues. Because **Go** boards and counters were used at this stage, instead of computers, it was important to have a death rule so that populations didn't tend to explode and quickly race off the board. On the other hand, to enable sufficiently interesting behavior so the game had a chance of being a universal system, it was equally important to have a birth rule that prevented populations from dying out. The rules eventually chosen provided a balance between birth and death so that the system tended to be fairly stable yet interesting enough to study. An early sign of success was the discovery of patterns, known as "gliders," that kept their shape while drifting across the plane. This was a hopeful step toward proving universality because it showed that the system had a way to transmit information from one place to another. Conway and his group went on to build nearly all the necessary configurations for arbitrary computations: AND gates, OR gates, and so on, just like the components of an ordinary computer. What they needed next was a way of producing gliders at will–a "glider gun." At this point, Conway sent a letter describing Life and the early findings to Martin Gardner, offering a prize of $50 for a configuration whose population tended toward infinity. The resulting *Scientific American* column sparked the public's imagination and very quickly a glider gun was discovered by a group at the Massachusetts Institute of Technology led by R. W. Gosper. Within two weeks of the discovery of the glider gun, both Conway's group and the group at MIT had shown that the system was indeed universal.

limaçon of Pascal

A snail-shaped curve (*limaçon* is French for "snail") named by the French mathematician Gilles Roberval after Etienne Pascal, the father of Blaise **Pascal**. It had been discovered earlier, however; Albrecht **Dürer** gave a method for drawing it as early as 1525 in his *Underweysung der Messung*. The limaçon of Pascal is a special case of an **epitrochoid** in which the rolling circle and the rolled circle have the same radius, and is also the cata**caustic** and the **pedal curve** of the **circle**. It has the **quartic** Cartesian equation

$$(x^2 + y^2 - 2rx)^2 = k^2(x^2 + y^2),$$

where r is the radius of the rolling circle or the rolled circle and k is a constant.

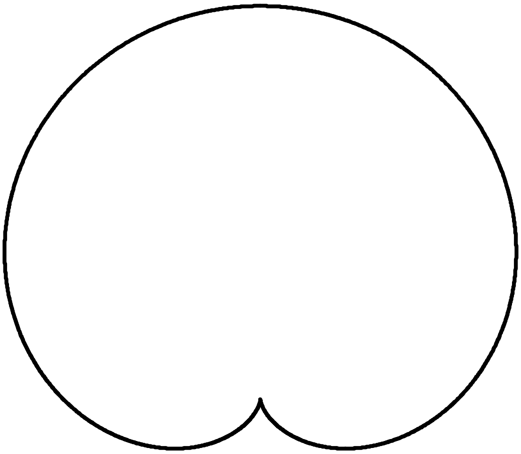

limaçon of Pascal *© Jan Wassenaar, www.2dcurves.com*

Sometimes the term *ordinary limaçon* is used to describe the curve when the value of k is greater than 0 and less than 1. When $k = 0$ the curve is a circle and when $k = 1$ the curve is a **cardioid**, so the ordinary limaçon is a transitional form between these two. The ordinary limaçon is also the **inverse** of the **ellipse**. For values of $k > 1$, a loop or noose appears in the curve. The inverse of a limaçon with a noose is a **hyperbola**. In fact, the constant k is the same as the eccentricity for a **conic section**. When $k = 2$, the limaçon is also called a **trisectrix**.

limerick

Why is this a limerick?:

$$((12 + 144 + 20) + 3\sqrt{4}/7) + 5 \times 11 = 9^2 + 0$$

Because, as its inventor, Jon Saxton, has pointed out:

A dozen, a gross, and a score,
Plus three times the square root of four,
Divided by seven,
Plus five times eleven,
Is nine squared and not a bit more.

limit

The target value that terms in a **sequence** of numbers get closer and closer to. This limit is not necessarily ever reached but can be approached arbitrarily close if the sequence is taken far enough.

limping triangle

A right **triangle** with two shorter sides (i.e., those other than the hypotenuse) that differ in length by one unit. An example is the 20-21-29 triangle ($20^2 + 21^2 = 29^2$).

line

The shortest distance between any two points in **Euclidean space**. A line is implicitly a *straight line;* the alternative is a **curve**. Mathematically, a line may be determined by the presence of any two points in an n-dimensional space (where n is two or more). A *line segment* is a piece of a line with definite endpoints.

linear

Having only a multiplicative factor. If $f(x)$ is a *linear function*, then $f(a + b) = f(a) + f(b)$ and $cf(x) = f(cx)$ must both be true for all values of a, b, c, and x.

linear algebra

The study of **vector**s and **vector space**s. Linear algebra has a central place in modern mathematics, is used widely in both **abstract algebra** and functional **analysis**, and finds a concrete representation in the form of **analytical geometry**. Linear algebra began as the study of vectors in two- and three-dimensions but has now been extended to a generalized n-space.

linear group

A **group** of matrices under **matrix** multiplication.

linear programming

The problem, and associated area of mathematics, of maximizing or minimizing a linear **function** on a **convex** set, especially a **polytope**. Equivalently, maximizing a linear expression in some number of variables subject to linear equalities and inequalities.

linear system

Any system whose change of values of its variables can be represented as a series of points suggesting a straight **line** on a coordinate; hence, *linear* for "line." More generally, a linear system is one in which small changes result in small effects, and large changes in large effects. In a linear system, the components are isolated and noninteractive. Real linear systems are rare in nature since living organisms and their components are not isolated and do interact. Compare with **nonlinear system**.

Liouville number

A **transcendental number** that can be approximated very closely by a **rational number**. Normally, proving that any given number is transcendental is difficult. However, the French mathematician Joseph Liouville (1809–1882) showed the existence of a large (in fact, infinitely large) class of transcendentals whose nature is easy to ascertain. An example of a Liouville number is 0.1010010000001000000000000000000000001 . . . in which the successive groups of zeros are of length 1!, 2!, 3!, 4!, and so on.

Lissajous figure

A figure or graph of the type most often seen on an oscilloscope. The simplest Lissajous figures are circles or ellipses, but they can also take the form of **lemniscates** and other, more complex shapes. They are named for the French scientist Jules Antoine Lissajous (1822–1880), who experimented with them in the 1850s, and are also known as *Bowditch curves,* because they had been written about earlier by the American astronomer and mathematician Nathanial Bowditch (1773–1838).

lituus

A **spiral** curve that was discovered by the English mathematician Roger Cotes (1682–1716) and named by the Scottish mathematician Colin Maclaurin (1698–1746) in 1722. The Latin *lituus* refers to a staff shaped like a bishop's crosier. The lituus is the locus of a point that moves in such a manner that the area of a circular sector remains constant; it has the polar equation

$$r^2 = a^2/\theta.$$

See **polar coordinates**.

Lobachevsky, Nikolai Ivanovich (1793–1856)

A Russian mathematician who was one of the pioneers of **non-Euclidean geometry**. He developed, independently of János **Bólyai**, the self-consistent system of **hyperbolic geometry** in which Euclid's **parallel postulate** is replaced by one allowing more than one parallel through the fixed point. Lobachevsky first announced his system in 1826 and subsequently wrote several expositions of it, including *Geometrical Researches on the Theory of Parallels* (originally published in 1840 in German). Lobachevsky studied and taught at the University of Kazan and eventually became rector of this institution in 1826. However, for some reason, despite serving his country and university well, he fell from favor and in 1846 was relieved by the government of his posts as professor and rector.

localized solution

A solution of a **differential equation**, or a similar mathematical object, that is confined to a small region even though it has the freedom to spread out.

loculus of Archimedes

A **dissection** game, similar to **tangrams**, which consists of 14 polygonal shapes that fit together to make a square. These pieces can be rearranged to make pictures of people, animals, and objects, or reassembled into their original form. There are many references to the game in ancient literature, including a description by the Roman poet and statesman Magnus Ausonius (A.D. 310–395). Only two fragmentary manuscripts, one an Arabic translation and the other a Greek manuscript dating from the tenth century discovered in Constantinople in 1899, connect the puzzle to **Archimedes** by calling it *loculus Archimedius* ("Archimedes's box"). More generally, but for unclear reasons, it is known as the *ostomachion* (Greek for "stomach"), or, in Latin texts, as the *syntemachion.* In 2003, William **Cutler** used a computer program to enumerate all 536 distinct ways (barring rotations and reflections) in which the pieces can be arranged into a square.

locus

The set of all points (usually forming a curve or surface) that satisfy some condition. For example, the locus of points in the plane equidistant from a given point is a circle. The Latin word *locus* simply means "place." (The Greek equivalent is *topos,* which crops up in "topology".)

logarithm

The logarithm of a number or variable x to base b, $\log_b x$, is the **exponent** of b needed to give x. The bases most commonly used in mathematics are ***e*** and 10. A logarithm to base e, written as $\log x$ or $\ln x$, is known as a *natural logarithm.* A logarithm to base 10 is written as $\log_{10}x$ and is known as a *common logarithm.*

logarithmic spiral

A type of **spiral**, also known as an *equiangular spiral,* that is very common in the natural world. Wonderful examples are found in the shells of some mollusks, such as that of the nautilus, and in spider webs. The angle any **tangent** to the curve makes with a tangent to a circle at the same radius, known as the *pitch angle,* is constant and results in a logarithmic spiral being *self-similar:* in other words, any part of it looks like any other part (though possibly rotated). Hawks approach their prey in the form of a logarithmic spiral and their sharpest view is at an angle to their flight direction that is the same as the spiral's pitch. On an altogether different scale, the arms of spiral galaxies are roughly logarithmic spirals. Our own galaxy, the Milky Way, is believed to have four major arms, each of which is a logarithmic spiral with pitch of about 12°. Approximate logarithmic spirals with a pitch of about 17° can be generated using the **Fibonacci sequence** or the **golden ratio**.

In **polar coordinates** (r, θ) the equation of the logarithmic spiral is

$$r = a\, b^{\theta},$$

with positive real numbers a and b. Changing a rotates the spiral while b controls how tightly and in which direction it is wrapped. It can be distinguished from the

Archimedean spiral by the fact that the distance between the arms of a logarithmic spiral increase in a **geometric sequence** while in an Archimedean spiral this distance is constant. Starting at a point *P* and moving inward along the spiral, one has to circle the origin infinitely often before reaching it; yet, the total distance covered is finite. This was first realized by the Italian physicist Evangelista Torricelli (1608–1647) even before calculus had been invented. The total distance covered is $r/\sin\theta$ where θ is the pitch angle and r is the straight-line distance from *P* to the origin.

The logarithmic spiral was first described by René **Descartes** and later studied in depth by Jakob Bernoulli (see **Bernoulli family**), who called it *Spiralis mirabilis* (the wonderful spiral) and wanted one engraved on his tombstone. He did get a spiral, but unfortunately it was a rather crudely cut Archimedean type.

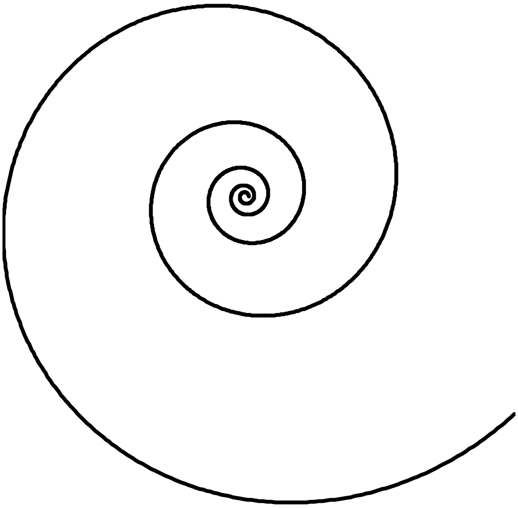

logarithmic spiral *© Jan Wassenaar, www.2dcurves.com*

logic

The branch of **mathematics** concerned with how one statement can imply others, or how sets of statements can

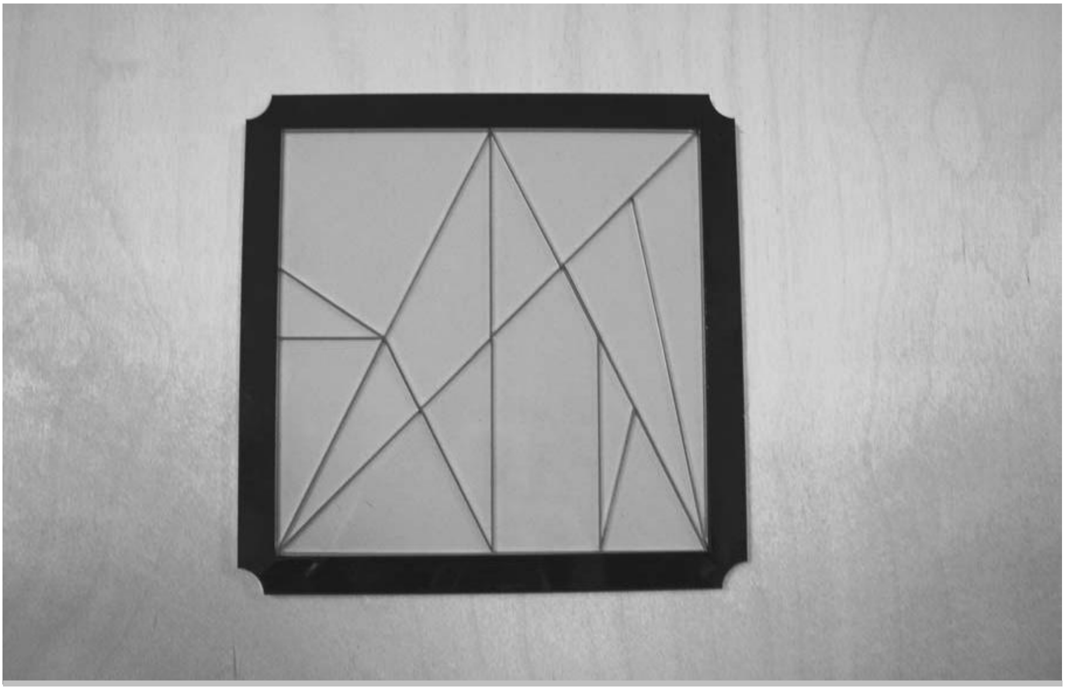

loculus of Archimedes The loculus of Archimedes is one of the most ancient dissection puzzles. *Kadon Enterprises, Inc., www.gamepuzzles.com* be connected by chains of implication. Such relationships can be written down using special symbols, and different sets of rules can be studied to see how they give rise to different sorts of structures.

logical depth

A measure of the **complexity** of a system, developed in 1988 by the American computer scientist and mathematician Charles Bennett. It contrasts with another such measure, **algorithmic complexity**, but, like it, is a measure of the algorithms needed to generate the data from a system.

loop

(1) A **knot** or hitch that holds its form. (2) A degenerate edge of a **graph** that joins a vertex to itself. (3) A sequence of instructions that is repeated either a specified number of times or until a particular condition prevails. Loops lie at the heart of most computer programs. The most common type is the *iterative loop,* signified by keywords such as *for, while, do,* and *repeat,* or their equivalent, in which a given set of instructions is repeated a specified number of times. The *recursive loop* is a more powerful construct that carries out a given set of instructions, typically including **recursive** calls with modified parameters back to the instruction set itself, until a terminating condition is met. Recursive algorithms solve problems by reducing them to smaller and smaller subproblems until a solution is found, reusing the same set of instructions as often as needed.

Lorenz, Edward Norton (1917–)

A research meteorologist at the Massachusetts Institute of Technology who, in the early 1960s, using a simple system of equations to model convection in the atmosphere, ran headlong into the phenomenon of "sensitivity to initial conditions." In the process he sketched the outlines of one of the first recognized **chaotic attractor**s. In Lorenz's meteorological computer modeling, he discovered the underlying mechanism of deterministic chaos: simply formulated systems with only a few variables can display highly complicated behavior that is unpredictable. Using his digital computer, culling through reams of printed numbers and simple strip chart plots of the variables, he saw that slight differences in one variable had profound effects on the outcome of the whole system. This was one of the first clear demonstrations of sensitive dependence on initial conditions. Equally important, Lorenz showed that this occurred in a simple, but physically relevant model. He also appreciated that in real weather situations, this sensitivity could mean the development of a front or pressure-system where there never would have been one in previous models. In his famous 1963 paper, Lorenz picturesquely explained that a butterfly flapping its wings in Beijing could affect the weather thousands of miles away some days later. This sensitivity is now called the **butterfly effect**.

Lorenz system

A system of three **differential equation**s, named after its discoverer Edward **Lorenz**, that was the first concrete example of **chaos** and a **chaotic attractor**.

Lovelace, Lady

See **Byron, Ada**.

loxodrome

A path on Earth's surface that is followed when a compass is kept pointing in the same direction. It is a straight line on a Mercator projection of the globe precisely because such a projection is designed to have the property that all paths along Earth's surface that preserve the same directional bearing appear as straight lines. The loxodrome isn't the shortest distance between two points on a sphere. The shortest distance is an arc of a **great circle**. But, in the past, it was hard for a ship's navigator to follow a great circle because this required constant changes of compass heading. The solution was to follow a loxodrome (from

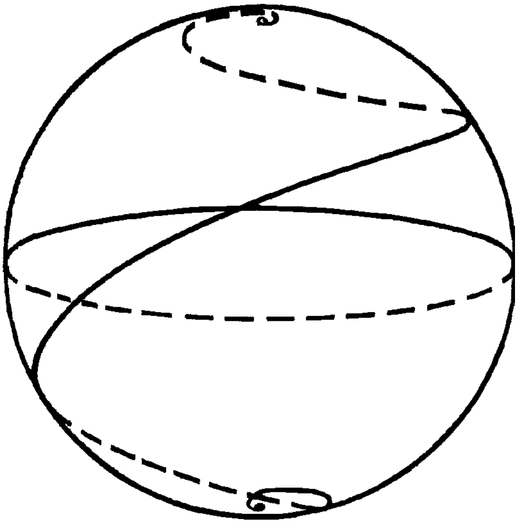

loxodrome Following a constant-compass heading between Earth's poles produces the winding path shown.

the Greek *loxos* for "slanted" and *drome* for "course"), also known as a *rhumb line,* by navigating along a constant direction. In middle latitudes, at least, this didn't lengthen the journey unduly. If a loxodrome is continued indefinitely around a sphere it will produce a spherical **spiral**, or a logarithmic spiral on a polar projection.

Loyd, Sam (1841–1911)

A great American inventor of puzzles among whose best-known creations are the Hoop-Snake Puzzle, **Get Off the Earth**, the **Pony Puzzle**, and, most famous of all, the **Fifteen Puzzle**. By age 17 he was already hailed as one of the world's leading writers of chess problems and had also just invented his deceptively simple-looking trick mules puzzle. The object of this is to cut apart three pieces that show two mules and two jockeys and then reassemble the pieces so that the jockeys are riding the mules. Loyd sold the puzzle to the showman Phineas T. Barnum (of Barnum & Bailey Circus fame) for some $10,000. This was to become his forte: devising puzzles that looked so simple to solve that people felt compelled to try them, only to find, hours later, that they were still trying to figure them out. Loyd became a full-time professional puzzlemaker only in the 1890s. He corresponded with his English counterpart Henry **Dudeney** and worked alongside his son Sam Loyd Jr. After the death of his father, Sam Loyd Jr. continued publishing puzzles that were mainly compilations of his father's work and in 1914 issued a now out-of-print and much sought-after mammoth collection of his father's puzzles called *Cyclopedia of 5,000 Puzzles, Tricks, and Conundrums.*[111, 207, 208]

lozenge

A **rhombus** with a 60° angle.

L-system

A method of constructing a **fractal** that is also a model for plant growth. L-systems use an **axiom** as a starting string and iteratively apply a set of parallel string substitution rules to yield one long string that can be used as instructions for drawing the fractal. Many fractals, including **Cantor dust**, the **Koch snowflake**, and the **Peano curve**, can be expressed as an L-system.

Lucas, (François) Edouard (Anatole) (1842–1891)

A French mathematician well known for his study of the **Fibonacci sequence** and the related **Lucas sequences** named after him. He devised methods of testing for **prime numbers**–work that was later refined by D. H. Lehmer to yield the Lucas-Lehmer test for checking **Mersenne numbers** to see if they are prime. Lucas was also interested in recreational mathematics, the **Tower of Hanoi** being his best known puzzle game. He worked at the Paris Observatory and later became a professor of mathematics in Paris.

Lucas sequences

Generalizations of the **Fibonacci sequence** first investigated by Edouard **Lucas**. One kind can be defined as follows: $L(0) = 0$, $L(1) = 1$, $L(n + 2) = PL(n + 1) + QL(n)$, where the normal Fibonacci sequence is the special case of $P = Q = 1$. Another kind of Lucas sequence begins with $L(0) = 2$, $L(1) = P$. Such sequences are used in **number theory** and in testing for **prime numbers**.

lucky number

A number in a sequence, first identified and named around 1955 by Stanislaw **Ulam**, that evades a particular type of number "sieve" (similar to the famous **sieve of Eratosthenes**), which works as follows. Start with a list of integers, including 1, and cross out every second number: 2, 4, 6, 8, ... The second surviving integer is 3. Cross out every third number not yet eliminated. This removes 5, 11, 17, 23, ... The third surviving number from the left is 7; cross out every seventh integer not yet eliminated: 19, 39, ... Repeat this process indefinitely and the numbers that survive are the "lucky" ones: 1, 3, 7, 9, 13, 15, 21, 25, 31, 33, 37, 43, 49, 51, 63, 67 . . . .

Amazingly, though produced by a sieve based solely on a number's position in an ordered list, the luckies have many properties in common with **prime numbers**. For example, there are 25 primes less than 100 and 23 luckies less than 100. In fact, primes and luckies crop up about equally often between any two given integers. Also, the gaps between successive primes and the gaps between successive luckies widen at roughly the same rate as the numbers increase, and the number of twin primes (primes that differ by 2) is close to the number of twin luckies. The luckies even have their own equivalent of the famous (still unsolved) **Goldbach conjecture**, which states that every even number greater than 2 is the sum of two primes. In the case of luckies, it's conjectured that every even number is the sum of two luckies; no exception has yet been found. Another unresolved problem is whether there are an infinite number of lucky primes. See also **Ulam spiral**.

Ludolph's number

Also known as *Ludolphine,* a name by which the number **pi** was known in Germany for many years following its evaluation to 35 digits by Ludolph von Ceulen (1540–1610).

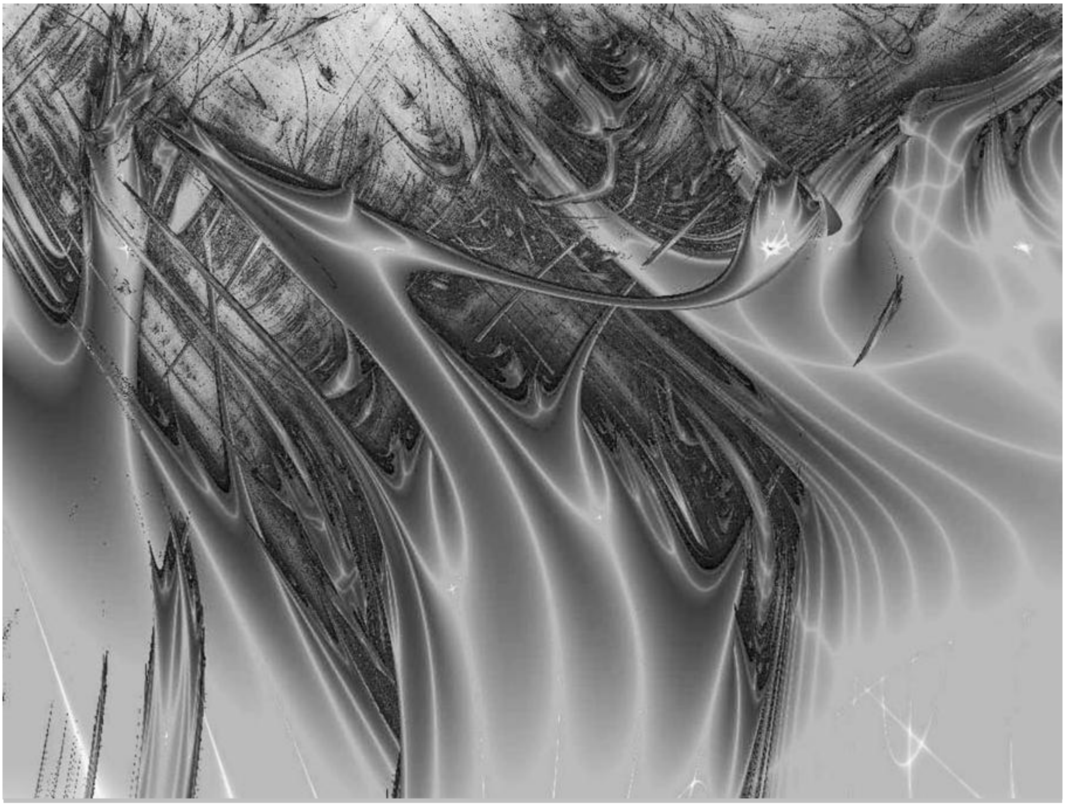

Lyapunov fractal A computer-generated image using Lyapunov exponents. *Radek Novak/Infojet*

lune

The portion of a **sphere** between two great semicircles (see **great circle**) having common endpoints (including the semicircles).

Lyapunov fractal

A particularly photogenic type of **fractal** that is popular with computer artists. It also represents a simple biological model of population growth in which the degree of the growth of the population can periodically alter between two values *a* and *b*.

M

Maclaurin, Colin (1698–1746)

A Scottish mathematician who developed and extended Isaac **Newton**'s work on **calculus** and gravitation, and did notable work on higher plane curves (see **Maclaurin trisectrix**). In his *Treatise of Fluxions* (1742), he gave the first systematic formulation of Newton's methods and set out a method for expanding functions about the origin in terms of series now known as *Maclaurin series*. Maclaurin also invented several devices, made astronomical observations, wrote on the structure of bees' honeycombs, and improved maps of the Scottish isles.

Maclaurin trisectrix

A curve first studied by Colin **Maclaurin** in 1742 with a view to solving one of the great geometric problems of antiquity: **trisecting an angle**. The Maclaurin trisectrix results from the Cartesian equation

$$y^2(a+x) = x^2(3a-x).$$

It is an **anallagmatic curve** that intersects itself at the origin.

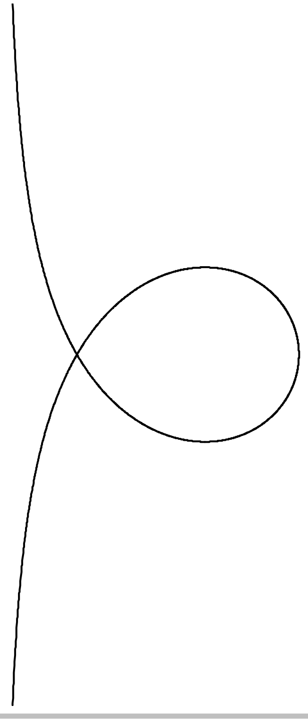

Maclaurin trisectrix *Jan Wassenaar, www.2dcurves.com*

MacMahon, Percy Alexander (1854–1929)

A British mathematician, physicist, and naval officer, born into a military family, whose leanings were evident early on when as a young child he showed a fascination with the way artillery was stacked. MacMahon later did work on missile trajectories, taking resistance into account, and on symmetric functions in the field of **combinatorics**, building on the results of James **Sylvester** and Arthur **Cayley**. His studies in symmetry led him to investigate partitions and to become a world authority on **Latin squares**. He wrote a two-volume treatise *Combinatory Analysis* (1915–1916), which became a classic, and a book on mathematical recreations called *New Mathematical Pastimes* (1921).[210] The latter shows another of the topics that intrigued MacMahon: the construction of patterns that can be repeated to fill the plane. However, much to his regret, as he wrote in the preface, "It has not been found possible to produce the book in colour." See also **MacMahon squares** and **thirty colored cubes puzzle**.

MacMahon squares

The concept of color-matching tiles based on all the permutations of colors on their edges dates from 1926, when Percy **MacMahon** invented and introduced three-color squares and four-color triangles as mathematical pastimes. MacMahon divided squares and triangles into triangles to give each edge of a piece its own color, in all possible combinations. Each set contains 24 different tiles, and MacMahon discovered that they could form a single figure with all adjacent edges matching and just one color all around the outside border. The most extensive research into these sets, over three decades, was done by the American engineer Wade Philpott (1918–1985), of Lima, Ohio, who identified all the possible symmetrical shapes that MacMahon squares and triangles could solve with both matching edge colors and uniform border color, and who calculated all the numbers of solutions for the MacMahon squares' 4 × 6 rectangle. See also **thirty colored cubes puzzle**.

Madachy, Joseph S.

An American mathematician, founder of *Recreational Mathematics Magazine,* editor of the ***Journal of Recreational Mathematics*** for nearly 30 years (now editor emeritus), the author of articles and books on the subject, and a well-known puzzlist.

magic cube

Similar to a **magic square** but having three dimensions instead of two. It contains the integers from 1 to n^3 and has $3n^2 + 4$ lines that sum correctly. All rows, columns, pillars, and the four triagonals (three-dimensional diagonals) must sum to the constant $\frac{1}{2}n(n^3 + 1)$. A cross section through a magic cube is a magic square.

magic square

An $n \times n$ square of the distinct whole numbers $1, 2, \ldots, n^2$, such that the sum of the numbers along any row, column, or main diagonal is the same. This sum is known as the *magic constant* and is equal to $\frac{1}{2}n(n^2 + 1)$. There is only one 3×3 magic square (not counting reflections and rotations), which was known to the Chinese as long ago as 650 B.C. as *lo-shu* and is bound up with a variety of myths. Associations between magic squares and the supernatural are also evident in early Indian and Arabian mathematics. The 3×3 square can be written as:

$$\begin{array}{ccc} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{array}$$

Each row, column, and main diagonal sums to 15. If the rows are read as three-digit numbers, forward and backward, and then squared, we find the interesting relation

$$816^2 + 357^2 + 492^2 = 618^2 + 753^2 + 294^2.$$

The reader may wish to see if the same rule holds for the columns and main diagonals.

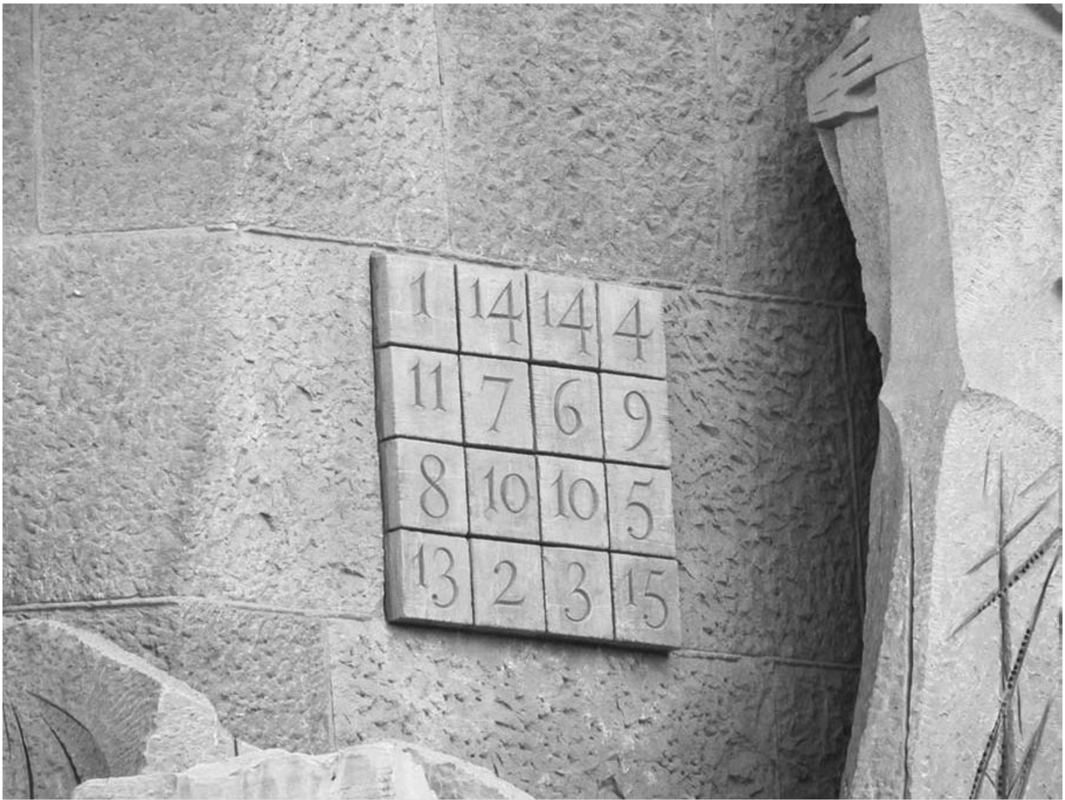

magic square A 4 × 4 magic square carved on the side of the Sagrada Familia cathedral in Barcelona.

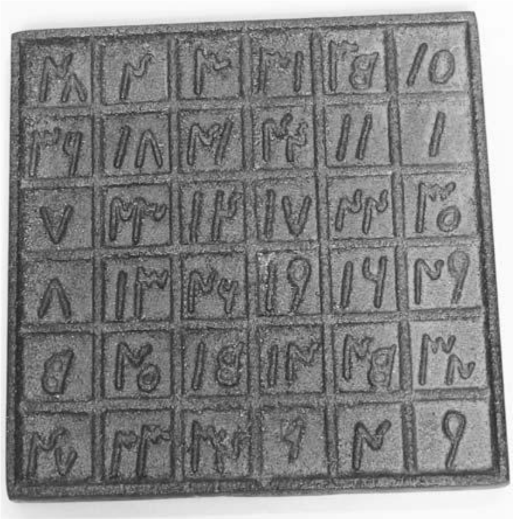

magic square A replica of a cast iron 6 × 6 magic square (vertical, horizontal, and main diagonal lines add to 111) from the Chinese Yuan dynasty (1271–1368), used as a sacred object to ward off evil spirits. *Sue & Brian Young/Mr. Puzzle Australia, www.mrpuzzle.com.au*

In the early sixteenth century Cornelius Agrippa constructed squares for $n = 3, 4, 5, 6, 7, 8$, and 9, which he associated with the seven "planets" then known (including the Sun and the Moon). Albrecht **Dürer**'s famous engraving of *Melancholia* (1514) includes a picture of an order-4 (4 × 4) magic square. There are 880 distinct squares of order-4 and 275,305,224 squares of order-5, but the number of larger squares is unknown. A square that fails to be magic only because one or both of the main diagonal sums don't equal the magic constant is called a *semi-magic square*. If *all* diagonals (including those obtained by wrapping around) of a magic square sum to the magic constant, the square is said to be a *pandiagonal square* (also known as a *panmagic* or *diabolical square*). Pandiagonal squares exist for all orders except 6, 10, 14, . . . , $2(2i + 1)$. There are 48 pandiagonal 4 × 4 squares. If replacing each number n_i by its square n_i^2 produces another magic square, the square is said to be a *bimagic* or *doubly magic square*. If a square is magic for n_i, n_i^2, and n_i^3, it is known as a *trebly magic square*.

A little trial and improvement is all it takes to construct the 3 × 3 magic square, but for building 4 × 4 squares and larger, a systematic method, or **algorithm**, is important. Interestingly, different algorithms are needed depending on whether the square is of an even order or an odd order. Odd order squares are the easier variety to make and there are several standard techniques, including the Siamese (sometimes called de la Loubere's or the Staircase), the Lozenge, and de Meziriac's methods. Here is yet another approach, known as the Pyramid or extended diagonals method: (1) Draw a pyramid of same size squares as the magic square's squares, on each side of the magic square; the number of squares on the pyramid's base should be two less than the number of squares on the side of the magic square. (2) Sequentially place the numbers 1 to n^2 of the $n \times n$ magic square in the diagonals. (3) Relocate any number not in the $n \times n$ square to the opposite hole inside the square.

An *antimagic square* is an $n \times n$ array of integers from 1 to n^2 in which each row, column, and main diagonal gives a different sum such that these sums form a sequence of consecutive integers. A 4 × 4 antimagic square is a square arrangement of the numbers 1 to 16 so that the totals of the four rows, four columns, and two main diagonals form a sequence of 10 consecutive integers, for example:

1	12	3	12
15	9	4	10
7	2	16	8
14	6	11	5

The principle of magic squares can be extrapolated from two dimensions to any number of **higher dimensions**, including **magic cube**s and *magic* ***tesseracts***, whose cross sections consist of magic cubes, and so forth.[10, 160, 162] See also **Latin square**s and **magic tour**.

magic tour

A **tour** by a **chess** piece on an $n \times n$ chessboard, whose squares are numbered from 1 to n^2 along the path of the piece, such that the resulting arrangement of numbers is a **magic square**. The tour is a *semi-magic tour* if the resulting arrangement of numbers is a **semi-magic square**. Magic **knight's tour**s aren't possible on $n \times n$ boards if n is odd. They are possible for all boards of size $4k \times 4k$ for $k > 3$, but are believed to be impossible for $n = 8$.

main diagonal

In the $n \times n$ **matrix** $[a_{ij}]$, the elements $a_{11}, a_{22}, \ldots, a_{nn}$. See also **diagonal**.

major axis

The longest **chord** of an **ellipse**.

Malfatti circles

In 1803 the Italian mathematician Giovanni Malfatti (1731–1807) posed the following problem: Given a triangle, find three nonoverlapping circles inside it such that the sum of their areas is maximal. Malfatti and many other mathematicians thought that the solution is given by the three circles, each of which is **tangent** to the other two and also to two sides of the triangle. Malfatti computed the radii of these circles, and they are now known as Malfatti's circles. Later it became clear that Malfatti's conjecture isn't true. In particular, Goldberg proved, in 1969, that the Malfatti circles *never* give a solution of the Malfatti problem! In other words, for any triangle there are three nonintersecting circles inside it, whose areas are bigger than the area of the Malfatti circles. So far as is known, the Malfatti problem hasn't been solved yet in the general case although it seems reasonable to suppose that the solution is given by what is called the *greedy algorithm:* First inscribe a circle in the given triangle; then inscribe a circle in the smallest angle of the triangle that is tangent to the first circle. The third circle is inscribed either in the same angle or in the middle angle of the triangle, depending on which of them has the bigger area.

Mancala

What is today almost universally called Mancala is, correctly speaking, the game of *Wari.* Mancala refers to a type of counting game of which Wari is a specific example, dating back to ancient times, known in Egypt and throughout Africa and Asia, and probably brought to Europe by sailors returning from those lands. Wari is played with a board and 48 markers or playing pieces, which are usually small colored stones or shells. The board consists of a piece of wood with two rows of six hollowed-out circular spaces, called *cups,* on each side, and one larger oval space at each end called a *reservoir.* The game is set up by placing four markers in each of the twelve cups. The first player picks up the markers in one of the cups on his side, then distributes the markers by placing one (and only one) in the other cups, going counterclockwise around the board. Markers are not placed in the reservoirs, which are for holding captured pieces only. Should there be enough markers that a complete circle of the board is made, the cup just emptied is skipped over. If the last marker put down goes into a cup on the opponent's side, and ends there with a total of two or three markers in that cup, then the markers in that cup are captured, and go into the reservoir at the player's right. Also, if the next cup clockwise from the captured cup has only two or three markers in it, that cup is likewise captured and the markers taken. This continues so long as the cup is clockwise to the last captured cup, has only two or three markers in it, and is an opponent's cup. However, a player can't capture all of an opponent's remaining pieces in a move, leaving him nothing to play. In capturing cups, if there's only one opponent cup left with markers in it, it can't be captured. Also, you cannot leave your opponent with all empty cups if you have a move that would put some pieces in one or more of his cups. When one side of the board is empty, play is ended. This happens when a player has to move his last markers onto his opponent's side. At the end of play each player takes any markers left on his side of the board and adds them to his reservoir. The winner is the person with the most markers in his reservoir.

Mandelbrot, Benoît B. (1924–)

A Polish-born French mathematician, largely responsible for the present interest in **fractal** geometry. A native of Warsaw, he spent most of his early life in France. Mandelbrot was born into a family with a strong academic tradition: his mother was a doctor and his uncle Szolem Mandelbrot was a famous Parisian mathematician. His family left Poland for Paris in the 1930s to escape Hitler's regime. There, Mandelbrot was introduced to mathematics by his two uncles. Educated in France, he developed the mathematics of Gaston **Julia**, and began the (now common) graphing of equations on a computer. Mandelbrot originated what became known as fractal geometry and the object known as the **Mandelbrot set** is named after him. His work on fractals as a mathematician at IBM earned him an emeritus fellowship at the Thomas J. Watson Research Laboratories. In addition to his study of fractals in mathematics, he showed that fractals can be found in many places in nature, leading to entire new fields of exploration in **chaos theory**. He joined the faculty of Yale in 1987.

Mandelbrot set

The best known **fractal** and one of the most complex and beautiful mathematical objects known. It was discovered by Benoît **Mandelbrot** in 1980 and named after him by Adrien Douady and J. Hubbard in 1982. The set is produced by the incredibly simple **iteration** formula

$$z_{n+1} = z_n^2 + c,$$

where z and c are complex numbers and $z_0 = 0$. This can be written without complex numbers as

$$x_{n+1} = x_n^2 - y_n^2 + a, \text{ and}$$
$$y_{n+1} = 2x_n y_n + b,$$

where $z = (x, y)$ and $c = (a, b)$. The Mandelbrot set consists of all the points on the **Argand diagram** for which the function $z^2 + c$ doesn't diverge under iteration. A computer is essential for carrying out the necessary calculations and for producing pictures of this remarkable structure. For the purposes of computation, the Argand diagram is broken down into pixels (picture elements), and the coordinates of each supply the constant c in $z^2 + c$. For each pixel (value of c) the function is iterated. If the function either rapidly diverges (blows up) or rapidly

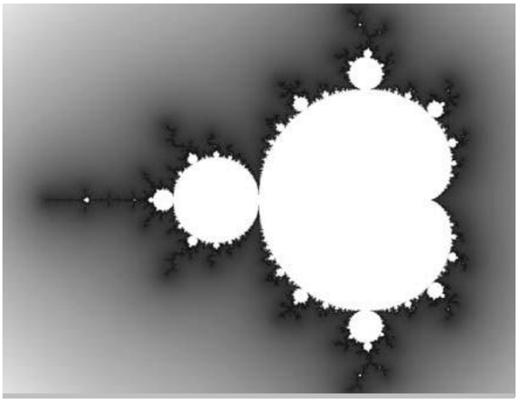

Mandelbrot set A general view of the Mandelbrot set.

converges (collapses), the pixel is left black. If the function is more indecisive about which way it is heading, it is allowed to iterate longer. In some cases the iterations could go on for a very long time before it became clear that the function would ultimately diverge, so a limit is established, known as the *depth,* beyond which iterations are stopped. If the depth is reached without divergence, the corresponding pixel is left black as though it were in the set. At locations where divergence becomes clear prior to hitting the limit, the pixel is displayed according to a scale that represents how many iterations are needed to show divergence. The whole Mandelbrot set lies within a circle of radius 2.5 centered at the origin of the Argand diagram. Although finite in area, the set has a boundary that is infinitely long and has a **Hausdorff dimension** of 2.

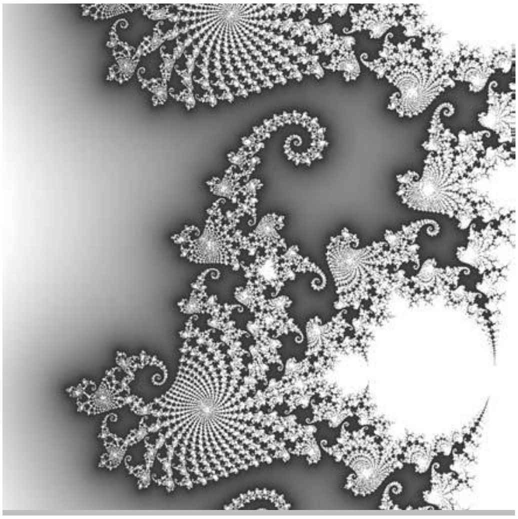

Mandelbrot set A zoom view of the Mandelbrot set reveals some of the infinite variety of remarkable patterns that inhabit it.

The overall appearance of the Mandelbrot set is that of a series of disks. These disks have irregular borders and decrease in size heading out along the negative real axis; moreover, the ratio of the diameter of one disk to the next approaches a constant. More complex shapes branch out from the disks. One region of the Mandelbrot set containing spiral shapes is known as *Seahorse Valley* because it resembles a seahorse's tail. A computer can be used, like a microscope, to zoom in on different parts of the set. This reveals that, although the shape is infinitely complex, it also displays **self-similarity** with regions that look like the outline of the entire set. The Mandelbrot set also reveals symmetry on different levels. It is identically symmetrical about the real axis, and almost symmetrical at smaller scales. This kind of "near-but-not-quite" symmetry is one of the most unexpected properties to find in an object generated from such a simple formula and process. The Mandelbrot set was created by Mandelbrot as an index to the **Julia sets**. Each point in the Argand diagram corresponds to a different Julia set, and those points within the Mandelbrot set correspond precisely to the connected Julia sets.

manifold

A mathematical object that, in geometrical terms, is nearly "flat" on a small scale (though on a larger scale it may bend and twist into exotic and intricate forms). More precisely, a manifold is a **topological space** that looks locally like ordinary **Euclidean space**. Every manifold has a *dimension,* which is the number of coordinates needed to specify it in the local coordinate system. A circle, although curved through two dimensions, is an example of a one-dimensional manifold, or one-manifold. A close-up view reveals that any small segment of the circle is practically indistinguishable from a straight line. Similarly, a sphere's two-dimensional surface, even though it curves through three dimensions, is an example of a two-manifold. Seen locally, the surface, like that of a small portion of Earth, appears flat. A manifold that is smooth enough to have locally well-defined directions is said to be *differentiable.* If it has enough structure to enable lengths and angles to be measured, then it is called a *Riemannian manifold.* Differentiable manifolds are used in mathematics to describe geometrical objects, and are also the most natural and general settings in which to study differentiability. In physics, differentiable manifolds serve as the **phase space** in classical mechanics, while four-dimensional pseudo-Riemannian manifolds are used to model **space-time** in general relativity.

mantissa

The positive fractional part of the representation of a **logarithm**. For example, in the expression log 3,300 = 3.5185 . . . , 0.5185 is the mantissa.

many worlds hypothesis

> *Do you not think it a matter worthy of lamentation*
> *that when there is such a vast multitude of them*
> *[worlds], we have not yet conquered one?*
> –Alexander the Great

An interpretation of **quantum mechanics**, first proposed by the American physicist Hugh Everett III in 1957, according to which, whenever numerous viable possibilities exist, the world splits into many worlds, one world for each different possibility (in this context, the term *worlds* refers to what most people call "universes"). In each of these worlds, everything starts out identical, except for the one initial difference; but from this point on, they develop independently. No communication is possible between the separate universes, so the people living in them (and splitting along with them) would have no idea what was really going on. Thus, according to this view, the world branches endlessly. What is "the present" to us, lies in the pasts of an uncountably huge number of different futures. Everything that *can* happen does happen, somewhere. Until the many worlds interpretation, the generally accepted interpretation of quantum mechanics was (and perhaps still is) the *Copenhagen interpretation.* The Copenhagen interpretation makes a distinction between the observer and the observed; when no one is watching, a system evolves deterministically according to a wave equation, but when someone *is* watching, the wave function of the system "collapses" to the observed state, which is why the act of observing changes the system. The Copenhagen interpretation gives the observer special status, not accorded to any other object in quantum theory, and cannot explain the observer itself, while the many worlds hypothesis models the entire observer-observee system.

map

(1) A synonym for **function**, in which context it is also known as a *mapping.* More generally, the correspondence of elements in one **set** to elements in the same set or another set. (2) A representation, usually on a plane surface, of geographical regions. See **four-color map problem**.

Markov chain

A sequence of **random** variables in which the future variable is determined by the present variable but is independent of the way in which the present state arose from its predecessors. In other words, a Markov chain describes a chance process in which the future state can be predicted from its present state as accurately as if its entire earlier history was known. Markov chains are named after the Russian mathematician Andrei Andrevich Markov (1856–1922) who first studied them in a literary context, applying the idea to an analysis of vowels and consonants in a text by Pushkin. But his work launched the theory of **stochastic** processes and has since been applied in quantum theory, particle physics, and genetics.

Martingale system

A simple, popular, and ultimately disastrous gambling system that, on the face of it, seems like a dream come true. In the short run the player has a good chance of making a few dollars using this method. But, in the long run, two things conspire to defeat him–the table betting limits and the player's available funds. The Martingale system calls for an initial bet of, say, $2. If the player loses, he doubles his bet to $4. Another loss puts the net loss at $6 and requires a doubling of the bet to $8 to recoup the losses and show a profit. Assume the player loses five hands in a row. The sixth bet requires $64. Let's say this wins–the gambler has now won $128 but has lost $124–a net win of $4. Of course, any win, however small, if repeated over and over, could produce a fortune. The trouble is that many losing streaks run longer than 6 or 8 or 10 in a row. The Martingale quickly runs into the table limits. For example, if the player is at a $2 **blackjack** table with a $500 upper limit, he has to retire after 9 losses in a row and is down over $1,000. It would take another 500 winning hands to make up this loss! Basically, one losing streak will put the Martingale gambler in a hole he is unlikely ever to climb back out of.

Mascheroni construction

A construction done using a moveable compass alone, named after the Italian geometer Lorenzo Mascheroni (1750–1800), who, in his *Geometria del compasso* (1797), astonished the mathematical world by showing how every compass-and-straightedge construction can be done in this minimalist way. (Since straight lines can't be drawn with just a compass, it's assumed that two points, obtained by arc intersections, define a straight line.) It is now known that Georg Mohr (1640–1697) proved the same results earlier in his obscure *Euclides danicus* (1672). Mascheroni, or Mohr-Mascheroni, constructions are today primarily of interest to puzzle enthusiasts who try to improve on the older solutions by finding ones with fewer steps.

matchstick puzzle

A form of **mechanical puzzle** that involves rearranging a pattern of ordinary matches, according to a given instruction, to make a new shape or solve some other mathematical problem.

mathematical lifespan

"The mathematical life of a mathematician is short. Work rarely improves after the age of twenty-five or thirty. If little has been accomplished by then, little will ever be accomplished." Thus wrote Alfred Adler in an article titled "Mathematics and Creativity" in the *New Yorker* magazine (1972), echoing a common belief that mathematicians tend to do their best work before the age of 30, physicists before the age of 40, and biologists before the age of 50 (though there are exceptions!). The mathematical physicist Freeman Dyson put it more succinctly: "Young men should prove theorems, old men should write books." On the other hand, there are compensations for early burnout, as G. H. **Hardy** pointed out (in *A Mathematician's Apology*): "Archimedes will be remembered when Aeschylus is forgotten, because languages die and mathematical ideas do not. 'Immortality' may be a silly word, but probably a mathematician has the best chance of whatever it may mean."

mathematics

> *Pure mathematics consists entirely of assertions to the effect that if such and such a proposition is true of anything, then such and such another proposition is true of that thing. It is essential not to discuss whether the first proposition is really true, and not to mention what the anything is of which it is supposed to be true. . . . Thus mathematics may be defined as the subject in which we never know what we are talking about, nor whether what we are saying is true.*
>
> –Bertrand Russell

The science of patterns, real or imagined; *mathematics* comes from the Greek *mathema* for "knowledge" or "that which is learned." Its roots lie in the practical need to carry out commercial calculations, to measure land, and to forecast astronomical events. These activities correspond roughly to the mathematics of structure, space, and change. The investigation of structure begins with numbers–initially the **natural number**s and **integer**s. The rules governing arithmetical operations are dealt with in elementary **algebra**, while the deeper properties of whole numbers are the province of **number theory**. The study of methods to solve equations leads to **abstract algebra**, which deals with structures that generalize the properties of familiar numbers. The physically important concept of **vector**, generalized to **vector space**s and studied in **linear algebra**, embraces both structure and space. The mathematics of space stems from **geometry**, first the **Euclidean geometry** and **trigonometry** of the everyday world, and later the various forms of **non-Euclidean geometry**. The modern fields of **differential geometry** and algebraic geometry generalize geometry in different ways. Differential geometry builds upon the concepts of coordinate system, smoothness, and direction, while algebraic geometry treats geometrical objects as sets of solutions to **polynomial** equations. **Group** theory places the concept of symmetry on an abstract footing and provides a bridge between space and structure. **Topology** links space and change through its emphasis on **continuity**. Analyzing and describing change in the physical world is a perennial theme of the natural sciences, and **calculus** was developed as a tool for doing this. The central concept used to describe a changing variable is that of a **function**. Many problems lead to relations between a quantity and its rate of change, and the methods to solve these are studied in the field of differential equations. The numbers used to represent continuous quantities are **real number**s, and the study of their properties and the properties of real-valued functions is known as real **analysis**. For various reasons, it's convenient to generalize to **complex number**s, which are dealt with in **complex analysis**. Functional analysis focuses attention on (typically infinite-dimensional) spaces of functions, laying the groundwork for **quantum mechanics** among many other things. In order to probe the foundations of mathematics, the fields of **set** theory, mathematical **logic**, and model theory were developed.

Since the time of the ancient Greeks, thought has been given to the ultimate nature of mathematics. What is its role and status in reality? Most crucially, is it invented or discovered? Leopold **Kronecker** was on the side of invention: "God made the integers; all the rest is the work of Man." Charles **Hermite**, by contrast, was clearly a Platonist: "There exists . . . an entire world, which is the totality of mathematical truths, to which we have access only with our mind, just as a world of physical reality exists, the one like the other independent of ourselves, both of divine creation." The German physicist Heinrich Hertz went even further: "One cannot escape the feeling that these mathematical formulas have an independent existence and an intelligence of their own, that they are wiser than we are, wiser even than their discoverers, that we get more out of them than was originally put into them." G. H. **Hardy** summed up what many mathematicians today have a tendency to believe:[151] "that mathematical reality lies outside us, that our function is to discover or observe it, and that the theorems which we prove, and which we describe grandiloquently as our 'creations,' are simply the notes of our observations." In another passage, he made this point more precisely: "317 is a prime, not because we think so, or because our minds are shaped in one way rather than another, but because it is so, because mathematical reality is built that way." We are struck by how well mathematics describes the behavior of the world in which we live. The universe, in fact, appears to have a deep mathematical infrastructure. Martin **Gardner** makes the bold claim that: "Mathematics is not only real, but it is the only reality. [The] . . . entire universe is made of matter. . . . And matter is made of particles. . . . Now what are the particles made out of? They're not made out of anything. The only thing you can say about the reality of an electron is to cite its mathematical properties. So there's a sense in which matter has completely dissolved and what is left is just a mathematical structure."

matrix

A square or rectangular array of numbers, usually written enclosed in a large pair of parentheses. Matrices, which are added and multiplied using a special set of rules, are extremely useful for representing quantities, particularly in some branches of physics. A matrix can be thought of as a **linear** operator on **vector**s. Matrix-vector multiplication can be used to carry out geometric **transformation**s such as scaling, rotation, reflection, and translation.

Matrix comes from the same Latin root that gives us *mother,* and was used to refer to the womb and to pregnant animals. It became generalized to mean any situation or substance that contributes to the origin of something. The first mathematical use of the word *matrix* was around 1850 by James **Sylvester** who saw a matrix as a way of obtaining **determinant**s, but didn't fully appreciate its potential. Within a year of his first mention of the term, he introduced the idea to Arthur **Cayley** who was the first to publish the **inverse** of a matrix and to treat matrices as purely abstract mathematical forms. The use of mathematical arrays to solve problems predates the application of the name by about 2,000 years. Around 200 B.C. in the Chinese text *Juizhang Suanshu* (Nine Chapters on the Mathematical Arts) the author solves a system of three equations in three unknowns by placing the **coefficient**s on a counting board and solving by a process that today would be called *Gaussian elimination.*

maximum

The largest of a set of values.

maze

A network of winding and interconnected passageways that a traveler must negotiate in order to reach some goal. The terms *maze* and *labyrinth* are often used interchangeably, though sometimes a distinction is made based on the layout. A labyrinth is then defined as a construction that leads from a starting point to a goal by a *single* path, with no branches or dead ends. No matter how long and twisting the route, it is predetermined by the builder: a labyrinth, according to this definition, is *unicursal.* A maze, by contrast, is *multicursal* and calls for the traveler to make a series of decisions that affect how quickly the goal is reached. In this book, *maze* and *labyrinth* mean the same thing, and refer to unicursal and multicursal mazes.

The most famous of legendary mazes was the lair (or prison) of the Minotaur at Knossos on Crete. According to Greek mythology, King Minos of Crete had his chief engineer, Daedalus, build the labyrinth in order to keep the half-human half-bovine offspring of his wife Pasiphae and a bull out of the public eye. King Aegeus of Athens was forced to pay a periodic tribute to Minos (the Athenians having earlier murdered Minos's son) in the form of seven young men and seven maidens. These unfortunates were forced to enter the maze below Minos's palace where they would get hopelessly lost and eventually be eaten by the monster. King Aegeus' son, Theseus, decided to put a stop to this and offered to take the place

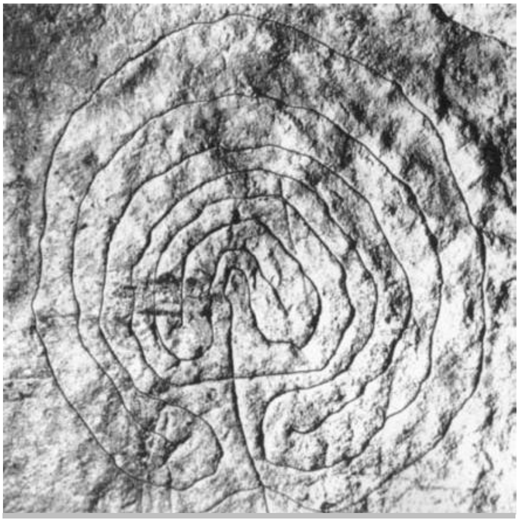

maze Possibly the world's oldest surviving labyrinth: a seven-ring labyrinth rock carving inside the Tomba del Labirinto, a Neolithic tomb at Luzzanas, Sardinia, dating somewhere between 2500 and 2000 B.C. *National University of Singapore*

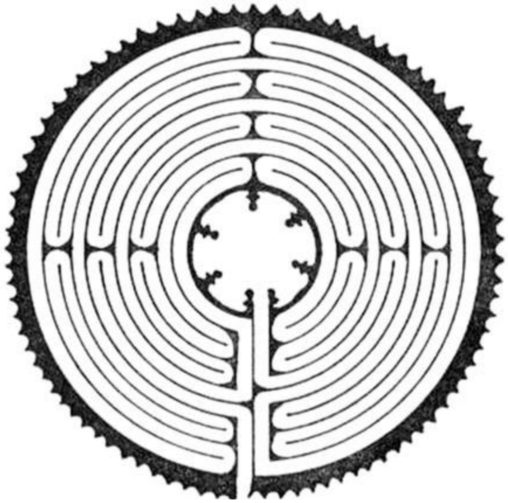

maze A plan of the maze in Chartres Cathedral as drawn by Henry Dudeney.

of one of the sacrificial virgins. He fell in love with Ariadne, one of Minos's daughters, who gave him a *clew,* or ball of yarn, to unravel as he entered the labyrinth so that he would be able to find his way back out. (Originally, *clue* and *clew* were alternative spellings of the same word. The modern sense of *clue* as a guide to solving a problem comes from the legend of the Minotaur, while *clew* retains its ancient meaning.) In true heroic style, Theseus slew the Minotaur but then spoiled the fairytale ending by abandoning Ariadne on the voyage home. Did this labyrinth really exist? In Roman times some writers suggested that a set of winding caves at Gortyan, in southern Crete, might have formed the basis for the tale. Although this complex of natural passages sounds similar to the maze in the myth, the story definitely places the labyrinth at the ancient Cretan capital. Archaeologists have found no evidence of a labyrinth structure at Knossos, but it has been suggested that the palace was so complicated, with its many levels, stairs, and rooms, that it may itself have inspired the story.

Minoan coins from about 300 B.C. bear a round, winding design, thought to be a representation of the labyrinth. A very similar geometric pattern recurs across many different cultures and times–scratched into caves in Cornwall (possibly by visiting Phoenician seafarers), on Roman coins, and in pictures drawn by native American Indians. Almost all these designs, including the one on the Minoan coins, are unicursal. Quite why Theseus would have needed a clew to navigate a maze that had only one way in and out is anybody's guess. But the underlying reason for the early unicursal design isn't hard to find. These mazes were intended not as intellectual puzzles but as symbolic representations of destiny as a matter of fate, beyond personal control. The world's oldest known maze-like designs, dating back about 3,500 years, have been found carved on rocks in northwest Spain and around the shores of the Mediterranean, and according to legend, were walked by fishermen before setting sail to ensure favorable winds and a good catch. In medieval times, mazes started to appear in churches, as art painted on walls or inlaid as "pavement mazes" on the floor. Some of the larger floor versions were traversed by people on their knees as a form of repentance or traversed as a substitute for an actual pilgrimage to the holy city, earning them the name "*Chemin de Jerusalem,*" or Road of Jerusalem. The oldest-known church labyrinth, at the Basilica of Reparatus at Orleansville, Algeria, dates from the fourth century A.D. and measures about 8 feet in diameter. One of the largest, built in 1288, formed part of the floor in the nave of Amiens Cathedral in France and spanned about 42 feet, but was destroyed in 1825. A splendid surviving example, however, is in Chartres Cathedral near Paris. Laid down in about 1200, this is an 11-circuit design (11 concentric windings) divided into quadrants, of the type often found in Gothic cathedrals. The only cathedral maze in Britain is at Ely, 16 miles north of Cambridge; when the Cathedral was restored around 1870, the architect, Gilbert Scott, installed a pavement maze of his own design under the west tower.

Mazes built primarily as puzzles or for recreation represent a different type of structure. For one thing, they are multicursal, offering those who enter a series of choices about which way to proceed. In his *Natural History,* the Roman historian Pliny comments that the classical type of labyrinth is quite distinct from "the mazes formed in the fields for the entertainment of children," suggesting that these diversions may have a long history. But it was in England that they really came of age. Church mazes never caught on in England, but turf mazes became tremendously popular. Ranging from 25 feet to over 80 feet across, they were constructed in or just outside villages across the countryside and were given names such as "Mizmaze," "Troy Town," "Shepherd's Race," and "Julian's Bower." A Welsh history book *Drych y Prif Oesoedd* (Mirror of the first age) published in 1740 notes the curious custom shepherds had of cutting the turf in the form of a labyrinth, which would seem to account for the origin of "Shepherd's Race." "Troy Town" probably refers to a legend that the city of Troy had seven exterior walls arranged as a maze to frustrate an attacking force.

From the turf maze it was no great leap to perhaps the most famous form of full-size maze–the topiary or hedge maze. While the use of hedges in gardens dates back to Roman times, the earliest references to a topiary maze appears in thirteenth-century Belgium. By the sixteenth century the hedge maze had spread to England, as a landscape painting by Tintoretto attests. In the later part of the seventeenth century, Louis XIV had a labyrinth built as a part of the gardens at Versailles, which included 39 groups of hydraulic statuary representing the fables of Aesop. The most famous surviving historic hedge maze is that on the grounds of Hampton Court Palace in England, designed by George London and Henry Wise for William of Orange, planted between 1689 and 1694, and occupying about a third of an acre. The finest turf maze in England may be that at Bridge End Gardens in Saffron Walden, Essex, which was replanted with yew in 1838–1840, abandoned and lost by 1949, then restored in 1983. A second maze, an imitation of that at Hampton Court, is on the Common.

In August 1997, Adrian **Fisher** opened the "World's Largest Maze" at Millets Farm Centre, Frilford, Oxfordshire, England, cut from about six acres of grain. However, his 1995 effort in Shippensburg, Pennsylvania, was in a 30-acre field and the 1997 maze in Reignac sur Indre covers 37 acres.[101]

When faced with a maze, what is the best way of reaching the goal, whether this is a point in the middle of the maze or an exit that forms a second opening to the structure? A unicursal maze calls for no brainwork, only footwork, since it consists of a single winding passage with no offshoots. Multicursal mazes are a different story. The easiest solution is to place a hand on one wall at the outset and follow that wall, come what may. Each blind alley will be traversed one time in and out until the whole maze is completed, or the goal is found. This simple method isn't the most efficient and it will fail altogether if the goal lies within an island in the maze–that is, a section detached from any of the exterior walls. A classic general method of "threading a maze" is: (1) never traverse a path more than twice; (2) when arriving at a new branch point or node, select either path; (3) when arriving at an old node or at a dead end by a new path, return by the same path; and (4) when arriving at an old node by an old path, select a new path, if possible; otherwise, an old path. An

maze A handheld maze, the object of which is to get the silver ball from the starting rectangle on the left to the finishing rectangle on the right. Crafted from Tasmanian blackheart sassafrass by Kym Anderson. *Mr. Puzzle Australia, www.mrpuzzle.com.au* explorer who follows these rules, using marks on the ground, as a record, can be assured of visiting every part of the maze.[48, 219, 240, 294]

See also **Rosamund's bower**.

mean

> *When she told me I was average, she was just being mean.*
>
> –Anonymous

A precisely calculated "typical" value of a group of numbers. There are several kinds of means. The *arithmetic mean,* which is usually what is meant when people talk about an **average**, is the sum of a set of values divided by the number of values. For example, the arithmetic mean of 3, 4, 7, and 10 is $(3 + 4 + 7 + 10)/4 = 6$. The *geometric mean* of n values is the nth root of the product of the values. For example, the geometric mean of 3, 8, and 10 is $(3 \times 8 \times 10)^{1/3}$ or the cube root of 240. The *harmonic mean* is the reciprocal of, or one over, the mean of the reciprocals of the values. For example, the harmonic mean of 3, 8, and 10 is $1/[(\frac{1}{3} + \frac{1}{8} + \frac{1}{10})/3]$. A different kind of average is *median,* which is generally the center term when a group of numbers are ordered by size. If there are an even number of values then the arithmetic mean of the center two is the median. The words *mean* and *median* both come from the Indo-European root *medhyo* meaning "middle."

measure

A way of gauging how big something is in terms of length, volume, or some other quality. One of the strangest facts in mathematics is that some objects exist that can't be measured. In the language of **set**s, the basic rules (somewhat simplified) of mathematical measures are as follows: (1) the measure of any set is a **real number**; (2) the **empty set** has measure zero; (3) if A and B are two sets with no elements in common then the measure of $A \cup B$ (the union of A and B) is equal to the measure of A plus the measure of B. The second of these rules can be very useful, for example, when integrating a **function**, since it allows us to ignore any points where the function jumps around, provided that such points are isolated. A slightly jittery function is one thing; a nonmeasurable set is a very different animal. Imagine a three-dimensional shape so fantastically intricate, so jagged and crinkled, that it is impossible to measure its volume and this gives some idea of the concept of nonmeasurability. From it flow such bizarre conclusions as the **Banach-Tarski paradox**.

measure theory

The part of **mathematics** that investigates the conditions under which **integration** can be carried out. It focuses mainly on the various ways in which the size, or **measure**, of a **set** can be estimated.

measuring and weighing puzzles

Problems that involve measuring a given quantity of liquid by pouring from one vessel into others of known capacity go back to medieval times. One of the earliest to appear in print was given by Niccoló **Tartaglia** and asks for 24 ounces of balsam to be divided into three equal portions using vessels that hold 5, 11, and 13 ounces, respectively. A similar problem was posed by a fellow traveler to the young Siméon **Poisson** while on a journey. Poisson's family had tried to steer him into careers ranging from a surgeon to a lawyer, the last on the theory that he was fit for nothing better. He seemed inept at everything he did. However, he saw the solution to the measuring problem immediately and realized his true calling. Thereafter, he threw himself into mathematics and became one of the greatest mathematicians of the nineteenth century.

The classic weighing problem was proposed by Claude-Gaspar **Bachet** and entails finding the least number of weights needed to weigh any integral number of pounds from 1 to 40 pounds inclusive, when no weights are allowed in either of the two pans. The answer is 1, 3, 9, and 27 pounds. Tartaglia had previously stated the same puzzle with the condition that the weights may only be placed in one pan, in which case the solution is 1, 2, 4, 8, 16, and 32 pounds.

PUZZLE

The following is a measuring problem from Henry **Dudeney**'s *Canterbury Puzzles,* which Dudeney claimed was the most popular of the whole collection and which the reader may like to try:

> Here be a cask of fine London ale, and in my hands do I hold two measures–one of five pints, and the other of three pints. Pray show how it is possible for me to put a true pint into each of the measures.

Solutions begin on page 369.

mechanical puzzle

A puzzle, involving several objects or a single object composed of one or more movable parts, whose solution requires moving from an initial state to a predefined final state. Mechanical puzzles were first classified by Louis **Hoffmann** in *Puzzles Old and New*[171] (1893). A modified form of his scheme is shown in the table "Mechanical Puzzles" on the following page.

medial triangle

The **triangle** whose vertices are the midpoints of the sides of a given triangle (see **vertex**).

Mechanical Puzzles

Type	Subtype	Examples
Assembly	2-dimensional assembly	**Tangrams**, **T-puzzle**, jigsaw
	3-dimensional assembly (noninterlocking)	**Soma cube**
	Matchstick puzzles	
	Miscellaneous	**Puzzle rings**
Disassembly	Trick or secret opening	**Puzzle jug**
	Miscellaneous	Trick locks, keys, etc.
Interlocking solid	**Burr puzzle**	
	3-dimensional jigsaws	
	Miscellaneous	Cubes, other objects
Disentanglement and entanglement	Wire puzzles	**Chinese rings**
	String puzzles	Cat's cradle
	Miscellaneous	
Sequential movement	**Peg solitaire** (peg removal)	
	Other counter (peg rearrangement)	
	Sliding-piece puzzles	**Fifteen Puzzle**
	Miscellaneous	**Tower of Hanoi**
Puzzle vessels	**Puzzle jug**	
	Miscellaneous	Bottom-fill teapots, pitchers
Vanishment puzzle		**Get Off the Earth**
Folding	**Origami**	
	Flexagon	
Impossible figure		**Penrose stairway** **Penrose triangle** **Impossible trident**

median

1. The line from a **vertex** of a **triangle** to the midpoint of the opposite side. 2. See **mean**.

Menaechmus (c. 380–c. 320 B.C.)

A Greek mathematician, thought to have been a pupil of **Eudoxus**, who is famed for his discovery of the **conic sections** and for being the first to show that **ellipse**s, **parabola**s, and **hyperbola**s are formed by cutting a cone in a plane that is not parallel to the base. Menaechmus made his discoveries on conic sections while attempting, unsuccessfully, to solve the problem of **duplicating the cube**. It has also been suggested that he served as a tutor to Alexander the Great.

Menger sponge

A famous **fractal** solid that is the three-dimensional equivalent of the **Sierpinski carpet** (which, in turn, is the

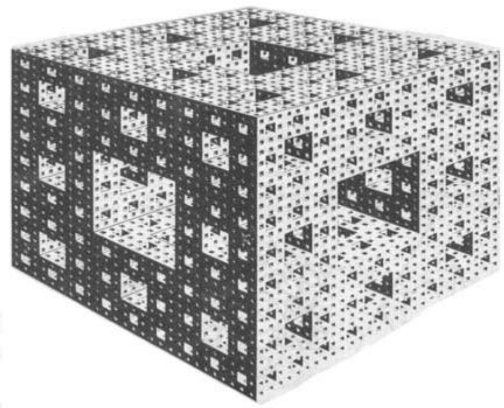

Menger sponge

two-dimensional equivalent of **Cantor dust**). To make a Menger sponge, take a cube, divide it into 27 (= 3 × 3 × 3) smaller cubes of the same size and remove the cube in the center and the six cubes that share faces with it. What's left are the eight small corner cubes and twelve small edge cubes holding them together. Now, imagine repeating this process on each of the remaining 20 cubes. Repeat it again. And again . . . ad infinitum. The Menger sponge was invented in 1926 by the Austrian mathematician Karl Menger (1902–1985).

Mersenne, Marin (1588–1648)

A French monk, philosopher, and mathematician best remembered for his work to find a formula to generate **prime number**s based on what are now known as **Mersenne number**s. However, in addition to being a mathematician, he wrote about music theory and other subjects; edited works of Euclid, Archimedes, and other Greek mathematicians; but most importantly, corresponded extensively with mathematicians and other scientists in many countries. At a time before scientific journals existed, Mersenne was at the heart of a network for information exchange.

Mersenne number

A number of the form $2^n - 1$ (one less than a power of 2), where n is a positive integer. Mersenne numbers are named after Marin **Mersenne** who wrote about them in his *Cogita Physico-Mathematica* (Physical Mathematics Knowledge 1644) and wrongly conjectured that they were prime for $n = 2, 3, 5, 7, 13, 17, 19, 31, 67$, and 257, and composite for $n < 257$. See also **Mersenne prime**.

Mersenne prime

A **prime number** of the form $2^p - 1$, where p is prime. A prime exponent is necessary for a **Mersenne number** to be prime but is not sufficient; for example, $2^{11} - 1 = 2{,}047 = 23 \times 89$. In fact, after an early clustering of Mersenne primes for fairly small values of p, further occurrences become increasingly rare. At the time of writing there are 40 known Mersenne primes, corresponding to values for p of 2; 3; 5; 7; 13; 17; 19; 31; 61; 89; 107; 127; 521; 607; 1,279; 2,203; 2,281; 3,217; 4,253; 4,423; 9,689; 9,941; 11,213; 19,937; 21,701; 23,209; 44,497; 86,243; 110,503; 132,049; 216,091; 756,839; 859,433; 1,257,787; 1,398,269; 2,976,221; 3,021,377; 6,972,593; 13,466,917; and 20,996,011. However, it isn't known if the current largest Mersenne prime is the fortieth in order of size because not all lower exponents have been checked. Mersenne primes rank among the largest of all known primes because they have a particularly simple test for primality, called the *Lucas-Lehmer test*.

The search for Mersenne primes has been going on for centuries. They are named after Marin **Mersenne** who, in 1644, helped the search gain wide recognition by writing to many mathematicians of his conjecture about which small exponents yield primes. Around the time that Mersenne's conjecture was finally settled, in 1947, digital computers gave a new impetus to the hunt for Mersenne primes. As time went on, larger and larger computers found many more Mersennes and, for a while the search belonged exclusively to those with the fastest computers. This changed in 1995 when the American computer scientist George Woltman began the *Great Internet Mersenne Prime Search* (*GIMPS*) by providing a database of what exponents had been checked, an efficient program based on the *Lucas-Lehmer test* that could check these numbers, and a way of reserving exponents to minimize the duplication of effort. Today GIMPS pools the combined efforts of dozens of experts and thousands of amateurs. This coordination has yielded several important results, including the discovery of the Mersennes $M_{3021377}$, $M_{2976221}$, and $M_{20996011}$ and the proof that M_{756839}, M_{859433}, and $M_{3021377}$ are the thirty-second, thirty-third, and thirty-fourth Mersennes.

meter

The basic unit of length adopted under the System International d'Unites (SI units). Over the years the definition of the meter has changed several times. Throughout all these definition changes the length of the meter hasn't changed, but the precision by which it is measured was improved. In 1793, the meter was defined to be 1/10,000,000 of the distance from the pole to the equator. During the nineteenth century, the definition was in terms of the length of standard bars of platinum kept under controlled conditions. In 1983, the current definition was adopted of the distance traveled by light in a vacuum in 1/299,792,458 of a second.

method of exhaustion

Finding an **area** by approximating it by the areas of a sequence of **polygon**s; for example, filling up the interior of a circle by inscribing polygons with more and more sides.

metric

Any **function** $d(x, y)$ that describes the *distance* between two points. Distance is formally defined as a single number with the following properties: (1) $d(x,y) = 0$ if and only if $x = y$; (2) $d(x,y) = d(y, x)$; (3) $d(x,y) + d(y, z) \geq d(x, z)$ (the triangle inequality).

metric space

A **set** that has a **metric**; in other words, a kind of space in which the concept of distance has meaning. Compare with **topological space**.

metrizable

For a **topological space**, the property that there exists a **metric** compatible with the topology. To say that a topological space is metrizable is to treat it as a **metric space**, but without distinguishing any specific or preferred distance function.

Michell, John (1724–1793)

An English natural philosopher and clergyman, educated at Queen's College, Cambridge, who discovered that the force between magnetic poles varies as $1/r^2$. In 1767, Michell became rector of Thornhill, now a suburb of Dewsbury, Yorkshire. Sometime in the early 1770s, he played music with the great astronomer William Herschel and gave Herschel his first telescope. In 1784, Michell deduced the existence of what are now called black holes from Newton's corpuscular theory of light, and suggested that some stars might have dark companions. He also devised and built the torsion balance for determining the universal gravitational constant *G,* but died before having the chance to use it. The balance was passed to the Cambridge physicist F. J. H. Wollaston, and from him to Henry Cavendish, who used it at his house in London and is often, mistakenly, identified as its inventor.

midpoint

The point M is the midpoint of line segment AB if $AM = MB$. That is, M is halfway between A and B.

mile

A measure of distance, the name of which is an abbreviation of the Latin *mille passes* or "one thousand paces." Since the paces (one step with each foot) of the Roman Army were supposed to be two steps, each 2.5 feet long, 1,000 paces is very close to the length now called a statute mile (5,280 feet). A *nautical mile* was developed to be a distance equal to 1 minute of arc ($1/60$ of a degree) distance along a great circle and is equal to 6,076 feet.

million

A thousand thousand, 1,000,000, or 10^6. The word comes from the Latin *mille* for "thousand" (which is also the root for ***mile*** and *millennium*) and the suffix *ion* that implies "large" or "great;" thus a million is literally a "great thousand." Although *million* seems to have come into use as early as the middle of the fourteenth century, most mathematicians would use the phrase "thousand thousands" to avoid confusion and it was not until the 1700s that "million" caught on. It appears in the King James Version of the Bible (Genesis 24:60) and in Shakespeare (*Hamlet* act II scene II)–"for the play, I remember, pleased not the million."

minimal prime

A **prime number** that is a substring of another prime when written in **base** 10. A string *a* is a substring of another string *b,* if *a* can be obtained from *b* by deleting zero or more of the characters in *b.* For example, 392 is a substring of 639,802. The minimal primes are:

2; 3; 5; 7; 11; 19; 41; 61; 89; 409; 449; 499; 881; 991; 6,469; 6,949; 9,001; 9,049; 9,649; 9,949; 60,649; 666,649; 946,669; 60,000,049; 66,000,049; 66,600,049

minimal surface

A surface that, bounded by a given closed curve or curves, has the smallest possible area. A minimal surface has a mean **curvature** of zero. Finding and classifying minimal surfaces and proving that certain surfaces are minimal have been major mathematical problems for over 200 years. If the closed curve is planar then the solution is trivial; for example, the minimal surface bounded by a circle is just a disk. But the problem becomes much more difficult if the bounding curve is nonplanar–in other words, is allowed to move up and down in the third dimension. The first nontrivial examples of minimal surfaces, the **catenoid** and the **helicoid**, were discovered by the French geometer and engineer Jean Meusnier (1754–1793) in 1776, but there was then a gap of almost 60 years before the German Heinrich Scherk found some more. In 1873 the Belgian physicist Joseph **Plateau** carried out experiments that led him to conjecture that soap **bubble**s and soap films always form minimal surfaces. Proving mathematically this was true became known as the **Plateau problem**. Most minimal surfaces are extremely hard to construct and visualize, in part because the majority of them are self-intersecting. However, the development of high-performance computer graphics has provided mathematicians with a powerful tool and the last couple of decades have seen a huge increase in the number of such surfaces that have been defined and investigated.

minimax theorem

A theorem that says there is always a rational solution to a precisely defined conflict between two people whose interests are completely opposite. It is rational in that both parties can convince themselves that they can't expect to do any better, given the nature of the conflict.

minimum

The smallest of a set of values.

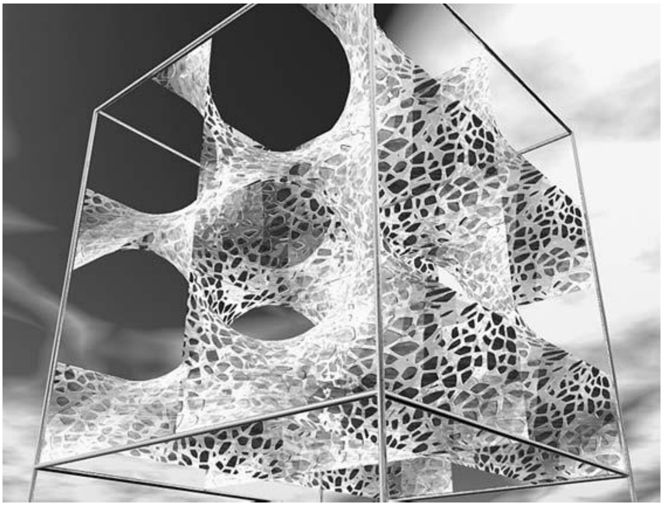

minimal surface A Scherk Surface—a type of minimal surface—portrayed as a membrane. *Anders Sandberg*

Minkowski, Hermann (1864–1909)

A German mathematician, born in Lithuania, who played an important part in the early development of **relativity theory**. Minkowski was the first to realize that the work of Hendrik Lorentz and Albert **Einstein** could be best understood if space and time, formerly thought to be separate entities, were treated as part of a four-dimensional **space-time** with a **non-Euclidean geometry**. The concept of the space-time continuum, which provided a framework for all later mathematical work in relativity, appeared in Minkowski's book *Raum und Zeit* (Space and Time 1907). From 1896 to 1902, Minkowski taught at the Zurich Federal Institute of Technology when Einstein was a student. In fact, Einstein attended several of the courses he gave but didn't create a good impression at the time. Minkowski described him as a "lazy dog" who "never bothered about mathematics at all." In 1902, Minkowski accepted a chair at the University of Göttingen, where he stayed for the rest of his life. His main interest was in pure mathematics, including number theory and geometry, and it was through his understanding of the more abstract side of mathematics and geometry in more than three dimensions that he developed the idea of four-dimensional space-time.

Minkowski space

A finite-dimensional **vector space**, especially a four-dimensional one, together with an indefinite inner product with one positive or timelike direction and many negative or spacelike directions. In particular, Minkowski space is ordinary **space-time** in the special **relativity theory**.

minor axis

The smallest **chord** of an **ellipse**.

minute

(1) One sixtieth of an hour or of a **degree** of arc or angle. (2) One sixtieth of an hour. As a measure of both time and angle, it equals 60 **seconds**.

mirror reversal problem

Why does a mirror reverse right and left, but not up and down? This question crops up perennially in the letter and query columns of magazines and newspapers. It was the inspiration for Lewis **Carroll**'s *Alice through the Looking Glass*. Alice Raikes (not to be confused with Alice Liddell, after whom the fictional Alice was modeled) was another of Carroll's young friends. On one occasion, in 1868, Carroll put an orange in her right hand and then asked her to stand in front of a mirror and say which hand showed the reflection of the orange. She said the left hand and Carroll asked her to explain. She finally replied "If I was on the other side of the glass, wouldn't the orange still be in my right hand?" Carroll said this was the best answer he'd had and later said it gave him the idea for his book.

Others have struggled harder but not always convincingly to explain the phenomenon, appealing variously to gravity, the psychology of perception, and philosophy. Why *does* a mirror reverse right and left, but not up and down? A frequently given answer is that a mirror doesn't reverse right and left. It reverses *front and back*. This is certainly true: the looking glass you is facing in the opposite

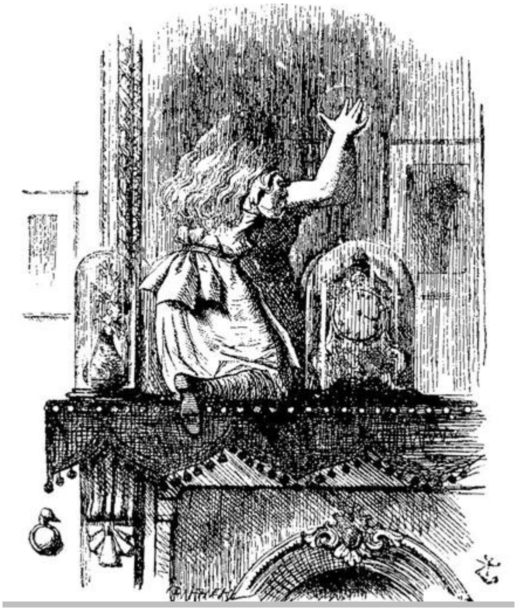

mirror reversal problem A looking-glass world beckons the curious Alice in Lewis Carroll's book *Alice through the Looking Glass.*

direction to the "real" you. But this short, crisp explanation doesn't completely dispel the mystery. If you imagine that the mirror is not there and that instead you are looking at a flesh-and-blood twin of yourself, that twin is differently handed. If you have a watch on your left wrist, the person you are facing has his/her watch on the right wrist. The mirror *has* done a left-right swap, surely! At any rate, something has happened to left and right that hasn't happened to up and down. To be more convinced of this, hold this book up to the mirror and try to read it. If no left-right swap has happened, why is the reflected writing so hard to read? First, remember that you are only looking at an image! The mirror hasn't (Carrollian fantasies aside) created some*thing* of opposite handedness. Secondly, appreciate how the writing appears in the mirror's frame of reference. This is easy to do by looking at the writing from the other side of the page (i.e., back to front, thus undoing the back to front reversal caused by the reflection). From the mirror's point of view the writing looks perfectly normal.

missing dollar problem

A version of this problem first appeared in R. M. Abraham's *Diversions and Pastimes* in 1933.[2] See also **nine nooms paradox**.

PUZZLE

Three people dine at a restaurant and receive a total bill for $30. They agree to split the amount equally and pay $10 each. The waiter hands the bill and the $30 to the manager, who realizes there has been a mistake and the correct charge should be only $25. He gives the waiter five $1 bills to return to the customers, with the restaurant's apologies. However, the waiter is dishonest. He pockets $2, and gives back only $3 to the customers. So, each of the three customers has paid $9 and the waiter has stolen $2 making a total of $29. But the original bill was for $30. Where has the missing dollar gone?

Solutions begin on page 369.

Mittag-Leffler, (Magnus) Gösta (1846–1927)

A Swedish mathematician who, in 1882, founded the international journal *Acta Mathematica,* and was its chief editor for 45 years. He studied in Paris under Charles **Hermite** and in Berlin under Karl **Weierstrass**, and made significant contributions to **analysis**. His best known work concerned the analytic representation of a one-valued **function**s and culminated in the *Mittag-Leffler theorem.* Since he took a special interest in Georg **Cantor**'s discoveries, much of Cantor's work was published in *Acta Mathematica.* Inscribed on the mantlepiece of Mittag-Leffler's home–now a research institute–in Djursholm is the epitaph: "Number is the beginning and end of thought. Thought gave birth to number but reaches not beyond. ML 1903"

mixed strategy

In **game theory**, a **strategy** that uses **random**ness by employing different actions in identical circumstances with different probabilities.

Möbius band

A simple and wonderfully entertaining two-dimensional object, also known as the Möbius strip, that has only one surface and one edge. It is named after the German mathematician and theoretical astronomer August Ferdinand Möbius (1790–1868), who discovered it in September 1858, although his compatriot and fellow mathematician Johann Benedict Listing (1808–1882) independently devised the same object in July 1858. Making a Möbius band is simple: take an ordinary sheet of typing paper, cut an 11″ × 1″ rectangle, bring the two long ends together, twist one of the ends 180°, and tape the two ends together. To prove that the band is single-sided, take a pen and start drawing a line around the band's circumference. When drawing the line, never take the pen off the paper; just

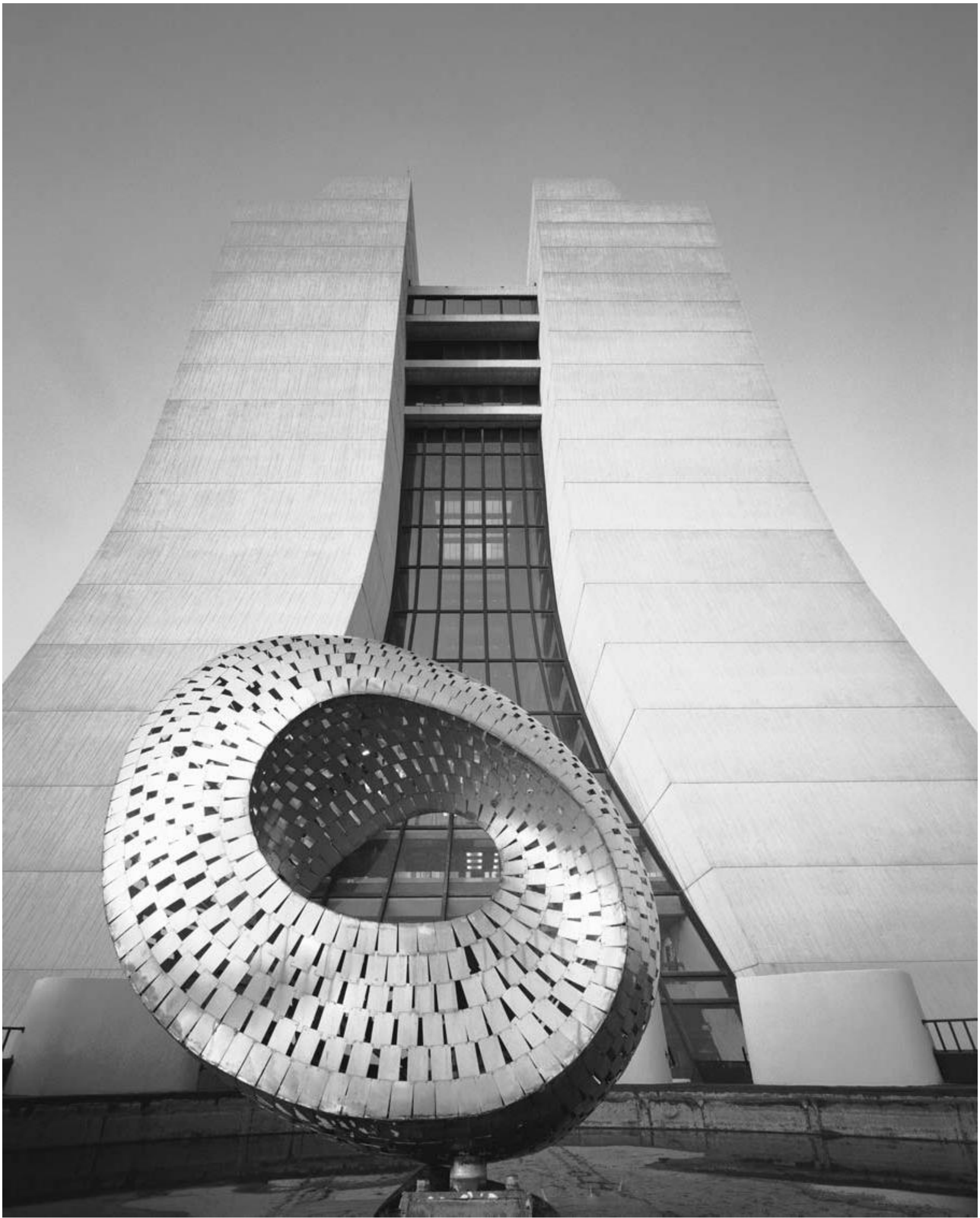

Möbius band A sculpture of a Möbius band outside the Fermi National Accelerator Laboratory. *FNAL* keep drawing the line until the starting point is reached. Once you are finished, look at both sides: there should be a line on both sides, thus proving that it is all the same side because you never took the pen off the paper.

The Möbius band has a lot of curious properties. If you cut down the middle of the band, instead of getting two separate strips, it becomes one long strip with two half-twists in it. If you cut this one down the middle, you get two strips wound around each other. Alternatively, if you cut along the band, about a third of the way in from the edge, you will get two strips; one is a thinner Möbius band, the other is a long strip with two half-twists in it. Other interesting combinations of strips can be obtained by making Möbius bands with two or more flips in them instead of one. Cutting a Möbius band, giving it extra twists, and reconnecting the ends produces unexpected figures called *paradromic rings*.

The Möbius band has provided inspiration both for sculptures and for graphical art. M. C. **Escher** was especially fond of it and based many of his lithographs on it. It is also a recurrent feature in science fiction stories, such as Arthur C. Clarke's *The Wall of Darkness*. A common fictional theme is that our universe might be some kind of generalized Möbius band. There have been technical applications; giant Möbius bands have been used as conveyor belts (to make them last longer, since "each side" gets the same amount of wear) and as continuous-loop recording tapes (to double the playing time).

A closely related strange geometrical object is the **Klein bottle**, which can be produced by gluing two Möbius bands together along their edges; however, this can't be done in ordinary three-dimensional Euclidean space without creating self-intersections.[116]

mode

The most frequently occurring value in a sequence of numbers.

model of computation

An idealized version of a computing device that usually has some simplifications such as infinite memory. A **Turing machine** and the **lambda calculus** are models of computation.

model theory

The study of mathematical structures that satisfy a particular set of **axiom**s, especially in the field of **logic**.

modulo

The integers a and b are said to be congruent modulo m if $a - b$ is divisible by m.

Moiré pattern

A radiating curved pattern created when two repetitive patterns overlap and interfere with one another. A Moiré pattern is seen, for example, when someone on TV wears a herringbone jacket. *Moiré* is the French word for "silk", and *silk moiré*, introduced from China to France in 1754, is the fabric that shows the familiar shifting patterns.

Moivre, Abraham de

See **de Moivre, Abraham**.

monad

A central concept in the Pythagorean worldview (see **Pythagoras of Samos**), in which it is regarded as the first thing that came into existence. Following the monad came, in order: the *dyad*, numbers, lines, two-dimensional entities, three-dimensional entities, bodies, the four elements (earth, air, fire, and water), and the rest of the world. The monad plays a similarly fundamental role in the metaphysics of Leibnitz as an indivisible, impenetrable unit of mental experience. It also has several different technical meanings in modern mathematics. For example, in **nonstandard analysis**, a monad consists of all those numbers infinitesimally closer to a given number. The word comes from the Latin *monas* (single) and Greek *monos* (unit).

Monge, Gaspard (1746–1818)

A French mathematician and physicist who put descriptive geometry, introduced by Albrecht **Dürer**, on a firm mathematical footing. He became professor of mathematics at Mézières (1768) and of hydraulics at the Lycée in Paris (1780), and published his groundbreaking treatise on the application of geometry to the arts of construction in 1795.

monkeys and typewriters

Six monkeys pounding away on typewriters would, by pure chance, if given enough time, be bound to write out all the works held in the British Library (or in all other libraries, for that matter). This idea was first suggested by the biologist Julian Huxley (1887–1975); it was discussed by the physicist James Jeans (1877–1946) in his *Mysterious Universe* (1930); and it has been restated in various forms over the years, in terms of chimpanzees, Shakespeare's sonnets, and the like. It was also the subject of Russell Maloney's short story "Inflexible Logic," first published in the *New Yorker* magazine (1940) and reprinted in Clifton **Fadiman**'s *Fantasia Mathematica*,[96] which tells the tragic tale of what happens when the fantastically improbable comes true. See also **Universal Library**.

monochromatic triangle

A triangle whose vertices are all colored the same (see **vertex**).

monomial

An algebraic expression that consists of just one term.

monotonic

The property of a **function** that is always strictly increasing or strictly decreasing, but never both.

Monster curve

See **Peano curve**.

Monster group

The largest, most fascinating, and most mysterious of the so-called sporadic **group**s; it was constructed by Robert Griess at Princeton in 1982, having been predicted to exist by him and Bernd Fischer in 1973, and was named the Monster by John **Conway**. Think of the Monster group as a preposterous snowflake with more than 1,050 symmetries that exists in a space of 196,883 dimensions. It contains the following number of elements:

$$2^{46} \times 3^{20} \times 5^{9} \times 7^{6} \times 11^{2} \times 13^{3} \times 17 \times 19 \times 23 \times 29 \times 31 \times 41 \times 47 \times 59 \times 71$$
$$= 808{,}017{,}424{,}794{,}512{,}875{,}886{,}459{,}904{,}961{,}710{,}757{,}005{,}754{,}368{,}000{,}000{,}000$$
$$\approx 8 \times 10^{53}$$ (more than the number of quarks in the Sun).

Despite these impressive credentials, however, it is still classified as a *simple group,* meaning that it doesn't have any normal subgroups other than the identity element and itself. All 26 simple groups have now been classified and the Monster is far and away the biggest. At first, it seemed that the Monster was just a curiosity–a Guinness Book record of pure math. Its only "useful" application seemed to be to give the best way for packing spheres in 24 dimensions! In ordinary three-dimensional space (and also four and five dimensions), the grocer's way of stacking oranges in a hexagonal lattice is thought to be the tightest possible (see **Kepler's conjecture**). But as the number of dimensions increases, the optimal packing method changes. A 24-dimensional grocer would get the most efficient arrangement of his 24-dimensional oranges by using the same symmetry as that of the Monster. This is unlikely to be immediately useful. Much more interesting, however, is the connection that has been found between the symmetry of the Monster and one of the most promising unifying theories in physics–**string theory**–which has been revealed by the **Monstrous Moonshine conjecture**.[70]

Monstrous Moonshine conjecture

An outrageous idea that stemmed from an observation made by John McKay of Concordia University in 1978. McKay was leafing through a table of abstruse mathematical data, giving possible values for coefficients of the *j*-function of certain elliptic curves, when he noticed the number 196,884 in the expression $j(q) = q - 1 + 196884q + 21493760q^2 + \ldots$. In a moment of inspiration, he recognized this number as being one more than the number of dimensions in which the **Monster group** can be most simply represented. Looking into this "coincidence" more closely, he found that it was no coincidence at all. In fact, *all* the coefficients of the *j*-function were simple combinations of the degrees of possible representations of the Monster. This pointed to some deep connection between two seemingly unrelated areas of mathematics. On the one hand were the coefficients of what is called an elliptic modular function–exactly the kind of function that would play a key role in the proof of **Fermat's last theorem**. On the other was the number of dimensions, and combinations of degrees, of a crystal lattice whose symmetry rotations and reflections formed the Monster. Subsequently, McKay and a few other mathematicians, including John **Conway** and Simon Norton, drew out the link between elliptic modular functions and the Monster in a proposition christened, because of its fantastic nature, the Monstrous Moonshine. In 1998, this conjecture was proved by Richard Borcherds (a former student of Conway's) at the University of California at Berkeley. Astonishingly, Borcherds's proof reveals a deep relationship between elliptic curves, the Monster Group, and **string theory**–the most promising theory on offer to unify our understanding of nature at the subatomic level. Borcherds showed that the Monster is the group of symmetries of 26-dimensional strings expressed in a form known as *vertex algebra.* Some people believe the connection may run even deeper and that Monstrous Moonshine may hold clues to the very existence of the reality in which we live.[69]

Monte Carlo method

A method of estimating the true value of a quantity by carrying out a lot of **random** samples. For example, suppose we want to know the probability of getting a double six when we roll two dice. We could roll a pair of dice a thousand times, and count how many times, *n,* a double six came up; the estimated probability would then be *n*/1,000. A famous example of using the Monte Carlo method is to calculate **pi**. Get a computer to generate two random numbers x and y, each in the range –1 to 1, so that the point (x, y) lies somewhere randomly inside a square of side 2 units. Do this thousands of times, and count up what proportion of the points also lie inside the circle that inscribes

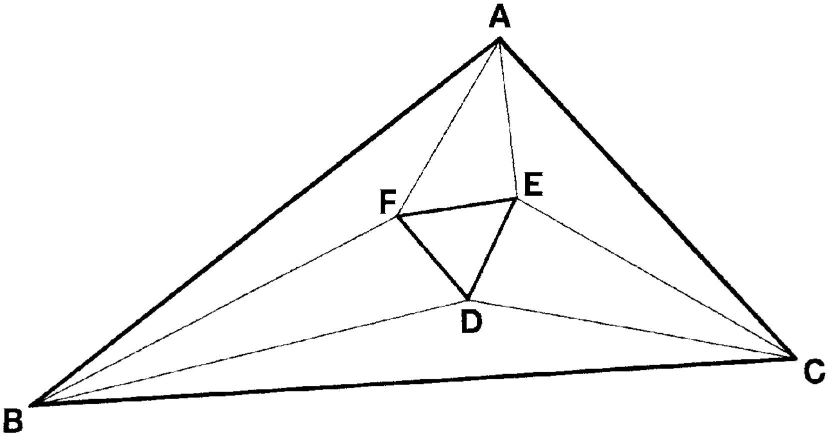

Morley's miracle The trisectors of the angles of triangle *ABC* meet the corners of a inner equilateral triangle *FED* known as Morley's triangle.

the square (you can tell whether a point does or not by working out whether $x^2 + y^2 < 1$). The proportion inside the circle is an approximation to $\pi/4$ (because the circle has area π but the square has area 4); millions of points are needed to obtain a good estimate.

This method was developed by researchers working on the Manhattan Project during World War II. To answer some of their scientific questions, they would repeatedly sample from their best estimates of the partial results, then apply the math they knew to the interactions and study the range of results. This process, which they named after the famous Monaco casino town of Monte Carlo, was created by John **von Neumann** and Stanislaw **Ulam**. The term and a description of the method seems not to have been published until some time after the war.

Montucla, Jean Etienne (1725–1799)

A French writer, mathematician, and scientist, who wrote several important early works on the history of mathematics. His *Histoire des mathématiques* (1758) was published in two volumes, the first of which covers the subject from ancient times to 1700, while the second is entirely devoted to seventeenth-century mathematics. It is considered the first attempt at a history of mathematical ideas and problems, in contrast to earlier works that were mostly lists of names, titles, and dates. Montucla had intended to produce a third volume covering the first half of the eighteenth century but the amount of new developments that had appeared during this time, and the difficulties of putting recent work into its historical context, led him to abandon this aim. A few years later he published another text for which he is famed–a new, greatly expanded and improved edition of Jacques **Ozanam**'s *Récréations mathématiques et physiques* (1778). Montucla's edition was particularly influential in popularizing geometric **dissection** problems. Charles Hutton translated it into English in 1803 and Riddle's edition was published in 1844, called *Recreations in science and natural philosophy*.

Monty Hall problem

A puzzle in probability that was inspired by the American game show *Let's Make a Deal*, hosted by Monty Hall. In its original form it goes like this: at the end of the show, you, the player, are shown three doors. Behind one of them is a new car, behind the other two are goats. Monty knows where the car is, but you don't. You choose a door. Before that door is opened however, Monty opens one of the two other doors with a goat behind it. He then gives you the option of switching to the other closed door. Should you switch or stick? At first glance, it seems as if it shouldn't make any difference. But the answer is surprising.

Suppose you stick. Your original choice made when all three doors were equally likely gives you a probability of winning the car of $\frac{1}{3}$. Now suppose you switch. In other words, you choose a door, wait for Monty to expose a goat, then switch to the other remaining door. This means that you win if the door you chose to begin with had a goat behind it. The odds that your initial choice had a goat is two thirds, so you are twice as likely to win the car if you switch! This can be hard to grasp. To make it easier, suppose there are 100 doors to choose from, but still only one car. You pick a door, Monty opens 98 that have goats behind them, then he gives you the option of switching to the other remaining closed door. Should you? Of course–it's almost certain that the car is behind the other door, and very unlikely that it's behind your original choice.

In a generalization of the original problem there are *n* doors. In the first step, you choose a door. Monty then opens some other door that's a loser. If you want, you may then switch your choice to another door. Monty will then open an as yet unopened losing door, different from your current preference. Then you may switch again, and so on. This carries on until there are only two unopened doors left: your current choice and another one. How many times should you switch, and when, if at all? The answer is: stick all the way through with your first choice but then switch at the very end.

In another variation of the problem, consider that in the actual game show there were two contestants. Both of them were allowed to pick a door but not the same one. Monty then eliminated a player with a goat behind his door (if both players had a goat, one was eliminated randomly, without letting either player know about it), opened the loser's door, and then offered the remaining player a chance to switch. Should the remaining player switch? The answer is *no*. The reason: a switcher in this game will lose if and only if either of two initial choices of the two contestants was correct. How likely is that? Two-thirds. A sticker will win in those ⅔ of the cases. So stickers will win twice as often as switchers.

Morley's miracle

A remarkable theorem, discovered in 1899, by Frank Morley, then professor of Mathematics at Haverford College. Take any triangle. Mark the three points that are the intersections of adjacent angle trisectors. Then, no matter what triangle you start with, these three points will form an equilateral triangle. That such a simple and elegant result was not known to the ancient Greeks may be because it is quite hard to prove.

One of the interesting auxiliary results of some of the proofs is that the side of the equilateral triangle is equal to $8r \sin(A/3) \sin(B/3) \sin(C/3)$, where A, B, and C are the angles of the larger triangle, and r is the radius of the circumcircle. A surprise awaits anyone who takes the intersections of the exterior, as well as the interior, angle trisectors. In addition to the interior equilateral triangle, four exterior equilateral triangles appear, three of which have sides that are extensions of a central triangle.

Moscow papyrus

See **Rhind papyrus**.

mousetrap

See **Cayley's mousetrap**.

moving sofa problem

In Douglas Adams's book *Dirk Gently's Holistic Detective Agency,* the character Richard MacDuff says at one point, "It would be really useful to know before you buy a piece of furniture whether it's actually going to fit up the stairs or around the corner." Mathematicians call this the moving sofa problem and it has been tackled in various forms over the past few decades. One version of it, formulated by Leo Moser in 1966, asks: What is the largest sofa (in terms of area) that can be moved around a right-angled corner in a hallway of unit width? The sofa can be any shape and doesn't even have to resemble a piece of furniture! The question simply asks for the biggest, unbendable area that can be maneuvered around the corner. Several different approaches suggest that the answer is about 2.21 square units. Variations on the problem involve negotiating pianos and other items around different types of bends and passageways.[133]

Mrs. Perkins's quilt

A square **dissection** problem first posed by Henry **Dudeney** in his *Amusements in Mathematics* (1917):[88]

PUZZLE

It will be seen that in this case the square patchwork quilt is built up of 169 pieces. The puzzle is to find the smallest possible number of square portions of which the quilt could be composed and show how they might be joined together. Or, to put it the reverse way, divide the quilt into as few square portions as possible by merely cutting the stitches.

Solutions begin on page 369.

Dudeney's problem can be generalized to the dissection of a square of side n into a number S_n of smaller squares. Unlike a perfect **squaring the square** problem, the smaller squares needn't be all different sizes. In addition, only prime dissections are considered so that patterns that can be dissected on lower order squares aren't allowed. The smallest number of relatively prime dissections of an $n \times n$ quilt for $n = 1, 2, \ldots,$ are 1, 4, 6, 7, 8, 9, 9, 10, 10, 11, 11, 11, 11, 12, ...[66]

Müller-Lyer illusion

A **distortion illusion** in which the orientation of arrowheads makes one line segment look longer than another.

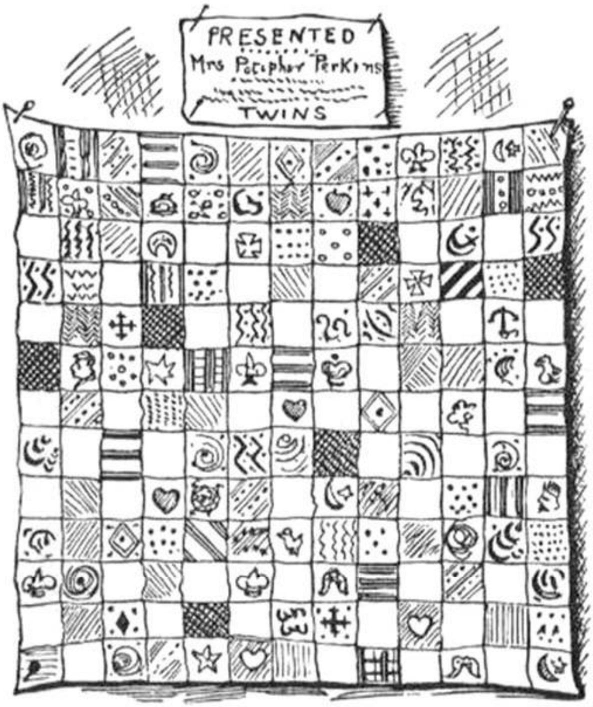

Mrs. Perkins's quilt Dudeney's drawing of one of his most famous puzzles.

multigrade

A set of equations in which the sums of powers of two different sets of numbers are the same for several different exponents. The simplest example is:

$$1 + 6 + 8 = 2 + 4 + 9$$
$$1^2 + 6^2 + 8^2 = 2^2 + 4^2 + 9^2$$

Another multigrade is:

$$1 + 8 + 10 + 17 = 36 = 2 + 5 + 13 + 16$$
$$1^2 + 8^2 + 10^2 + 17^2 = 454 = 2^2 + 5^2 + 13^2 + 16^2$$
$$1^3 + 8^3 + 10^3 + 17^3 = 6426 = 2^3 + 5^3 + 13^3 + 16^3$$

Remarkably, if any integer is added to all the terms of a multigrade it will still hold. Adding 1 to the example above, gives the multigrade (2, 9, 11, 18); (3, 6, 14, 17) ($n = 1, 2, 3$). Some high-order multigrades include: (1, 50, 57, 15, 22, 71); (2, 45, 61, 11, 27, 70); (5, 37, 66, 6, 35, 67) ($n = 1, 2, 3, 4, 5$), and (1, 9, 25, 51, 75, 79, 107, 129, 131, 157, 159, 173); (3, 15, 19, 43, 89, 93, 97, 137, 139, 141, 167, 171) ($n = 1, 3, 5, 7, 9, 11, 13$).

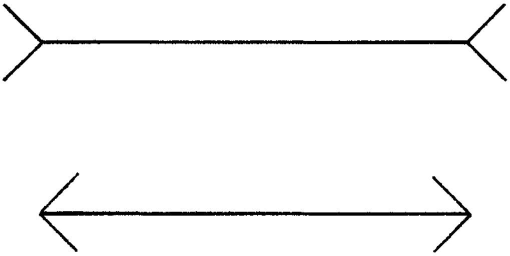

Müller–Lyer illusion Which of the horizontal lines is longer?

multiple

The integer b is a multiple of the integer a if there is an integer d such that $b = da$.

multiplication

A **binary operation** that is the equivalent of repeated addition and the **inverse** of **division**.

Mydorge, Claude (1585–1647)

A French mathematician who trained as a lawyer but was wealthy enough that he didn't have to work for a living. He was interested in mathematical puzzles and his book *Examen du livre des récréations mathématiques* (Study of the book of recreational mathematics, 1630) formed the basis for later works such as that by Denis Henrion (1659). Mydorge edited *Récréations mathématique* and left an unpublished manuscript of over 1,000 geometric problems and their solutions. He was also interested in optics and made a large number of instruments for his close friend René **Descartes**; the two shared a strong interest in explaining vision and the instruments and lenses were designed to help test their theories.

myriad

A term which today is normally synonymous with "very large number." Its origins go back to the Greek word *murious,* meaning "uncountable". The plural of this, *murioi,* evolved into the Latin *myriad,* which the Romans used to represent ten thousand. *Myriapod* is a general name for any many-legged anthropod, such as a millipede or a centipede

N

Nagel point

A point in a triangle where the lines from the vertices (see **vertex**) to the points of contact of the opposite sides with the excircles to those sides meet. (An excircle touches one side of a triangle and also touches the lines extended from the other two sides.)

nano-

Prefix for billionth (10^{-9}), from the Greek *nanos,* meaning "dwarf."

Napier, John (1550–1617)

A Scottish mathematician and theological writer who invented **logarithms** and wrote *Mirifici logarithmorum canonis descriptio* (The description of the wonderful canon of logarithms, 1614), which contains the first logarithmic table and the first use of the word *logarithm.* He also introduced the decimal point in writing numbers. His *Rabdologiae* (1617) describes various shortcuts for carrying out arithmetical calculations. One method of multiplication uses a system of numbered rods called *Napier's rods,* or *Napier's bones*–a major improvement on the ancient system of counters then in use. In 1619, after Napier's death, his *Mirifici logarithmorum canonis constructio,* which gave the method of construction of his logarithms, was published by his son Robert and edited by Henry **Briggs**.

Napoleon Bonaparte (1769–1821)

> *The advancement and perfection of mathematics are intimately connected with the prosperity of the state.*

Emperor of France and a very good amateur mathematician, having excelled in this subject as a student at school and at military college. Even after becoming first consul he was proud of his membership in the Institute de France (the nation's leading scientific society), and was close friends with several mathematicians and scientists, including Joseph **Fourier**, Gaspard **Monge**, Pierre Simon **Laplace**, Chaptal, and Berthollet. Indeed, in his grand expedition to Egypt in 1798 Napoleon brought along (in addition to 35,000 troops) over 150 experts in various fields, among them Monge, Fourier, and Berthollet, not to mention a complete *encyclopedie vivante* with libraries and instruments. One result of the expedition was that Fourier served for a time as the governor of lower Egypt. Likewise Laplace (who interviewed the young Napoleon for admission to the artillery) received titles and high office as a result of his friendship with Bonaparte. However, Laplace was relieved of his duties as the minister of the interior after only six weeks, and Napoleon later commented that Laplace had "sought subtleties everywhere, had only doubtful ideas, and carried the spirit of the infinitely small into administration." The most famous exchange between these two men occurred after Laplace had given Napoleon a copy of his great work, *Mecanique Celeste.* Napoleon looked it over, and remarked that in this massive volume about the universe there was not a single mention of **God**. Laplace replied "Sire, I had no need of that hypothesis." Regarding the idea that Napoleon might have discovered what is now called *Napoleon's theorem* (if equilateral triangles are constructed on the sides of any triangle (all outward or all inward), the centers of these equilateral triangles themselves form an equilateral triangle), Harold **Coxeter** and Samuel Greitzer have said that "The possibility of [Napoleon] knowing enough geometry for this feat is as questionable as the possibility of his knowing enough English to compose the famous palindrome, ABLE WAS I ERE I SAW ELBA."

nappe

Either of the two parts into which a **cone** is divided by the **vertex**. "Nappe" is the French for "tablecloth," which in turn comes from the Latin *mappa* (napkin).

narcissistic number

Also known as an *Armstrong number* or a *plus perfect number,* an *n*-digit number equal to the sum of its digits raised to the *n*th power. For instance, 371 is narcissistic because $3^3 + 7^3 + 1^3 = 371$, and 9474 is narcissistic because $9^4 + 4^4 + 7^4 + 4^4 = 9474$. The smallest narcissistic number of more than one digit is $153 = 1^3 + 5^3 + 3^3$. The largest narcissistic number (in base 10) is 115,132,219,018,763,992,565,095,597,973,971,522,401, which is the sum of the thirty-ninth powers of its digits. The reason there are no larger numbers is related to the fact that, as the number of digits increases, more and more nines are required to get a sum that has *n* digits. For example, $10^{70} - 1$ is a number consisting of 70 nines in a row, and the sum of the seventieth powers of its digits is $70 \times 9^{70} \cong 4.386051 \times 10^{68}$, which is only 69 digits long. So there is no way any 70-digit number can be equal to the sum of the seventieth powers of its digits. The reason we see the last number occur at 39 digits is because, as the limit is approached, the number of big digits like eights and nines has to increase to make sure the sum will be big enough, but this means that there are a lot fewer combinations of digits to choose from.

Nash, John Forbes Jr. (1928–)

An American mathematician not very accurately portrayed in the Oscar-winning film *A Beautiful Mind* (2001), loosely based on the biography of the same name by Sylvia Nasar (1998).[229] Nash, who worked in **game theory** and **differential geometry**, shared the 1994 Nobel prize for economics with two other game theorists, Reinhard Selten and John Harsanyi. After a promising start to his mathematical career, Nash began to suffer from schizophrenia around the age of 30 and battled with the illness for the next quarter of a century. His Ph.D. dissertation, entitled "Non-cooperative Games," contained the definition and properties of what would later be called **Nash equilibrium** and the basis of the work that, 44 years later, would make him a Nobelist. Between 1966 and 1996, Nash published nothing. However, as his mental health slowly began to improve in the mid-1990s, his ability to tackle mathematical problems returned, and he also became interested in computer programming.

Nash equilibrium

In **game theory**, a pair of strategies (see **strategy**) for a game such that neither player can improve his outcome by changing his strategy. A Nash equilibrium sometimes takes the form of a **saddle** structure. In other cases, when a strategy is at a Nash equilibrium with itself, the strategy resembles an evolutionary stable strategy.

natural logarithm

Also called a *Naperian logarith,* a **logarithm** to base **e**. For example, $\log_e 10$ (also written as 1*n* 10) is approximately 2.30258.

natural number

A number used for counting: 1, 2, 3, ... The debate about whether zero should also be included as a natural number has been going on for hundreds of years, and there's no general agreement even today. To avoid confusion, 0, 1, 2, 3, . . . , are often referred to as *nonnegative integers* or *whole numbers,* while 1, 2, 3, . . . , are called *positive integers.* Adding or multiplying natural numbers always produces other natural numbers. However, subtracting them can produce zero or negative integers, while dividing them produces **rational number**s. An important property of the natural numbers is that they are *well-ordered,* in other words, every set of natural numbers has a smallest element. The deeper properties of the natural numbers, such as the distribution of **prime numbers**, are studied in **number theory**. Natural numbers can be used for two purposes: to describe the position of an element in an ordered sequence, which is generalized by the concept of **ordinal number**, and to specify the size of a finite set, which is generalized by the concept of **cardinal number**. In the finite world, these two concepts coincide; however, they differ when it comes to infinite sets (see **infinity**).

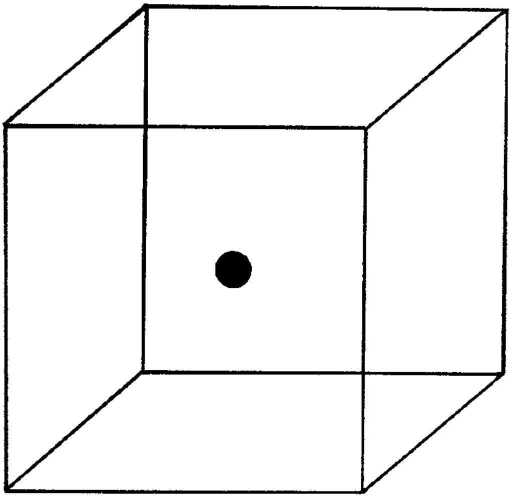

Necker cube Is the circle on the front face of the cube or the back?

Necker cube

A classic example of an **ambiguous figure**. In 1832, the Swiss crystallographer Louis Necker noticed, while examining crystals, that three-dimensional objects can fluctuate in appearance. He published pictures of an unusual **cube** that appeared to assume different orientations as one looked at it. The effect works because the drawing of the cube (an orthographic projection) carefully eliminates all depth cues. In attempting to fit the expected model of a cube to the picture, our brain must resolve the ambiguity as to which corner of the cube is closer.

negative base

The use of a negative **base** to represent numbers gives rise to some intriguing possibilities. Consider "negadecimal," for example, in which the base is *minus* 10 instead of the familiar positive 10. In this system, the number 365 is equivalent to the decimal number 5 + (6 × −10) + (3 × −10 × −10), = 245, while 35 in negadecimal is equivalent to 5 + (3 × −10), = −25, in ordinary decimal. This points to an interesting fact: the negadecimal equivalent of any positive or negative decimal number is always positive and therefore doesn't need to be accompanied by a sign. The Polish UMC-1, of which a few dozen were built in the late 1950s and earlier 1960s, is the only computer ever to use "nega**binary**" (base 2 arithmetic).

negative number

Long denied legitimacy in mathematics, negative numbers are nowhere to be found in the writings of the Babylonians, Greeks, or other ancient cultures. On the contrary, because Greek mathematics was grounded in geometry, and the concept of a negative distance is meaningless, negative numbers seemed to make no sense. They surface for the first time in bookkeeping records in seventh-century India and in a chapter of a work by the Hindu astronomer **Brahmagupta**. Their earliest documented use in Europe is in 1545 in the *Ars magna* of Girolamo **Cardano**. By the early seventeenth century, Renaissance mathematicians were explicitly using negative numbers but also meeting with heavy opposition. René **Descartes** called negative roots "false roots," and Blaise **Pascal** was convinced that numbers "less than zero" couldn't exist. Gottfried **Leibniz** admitted that they could lead to some absurd conclusions, but defended them as useful aids in calculation. By the eighteenth century, negative numbers had become an indispensable part of algebra.

Neile's parabola

Also known as the semi-cubical **parabola**, a curve discovered by the English mathematician William Neile (1637–1670) in 1657; it is the first **algebraic curve** to have its arc length calculated. (Before this, only the arc lengths of transcendental curves such as the **cycloid** and the **logarithmic spiral** had been calculated.) Neile's parabola is described in **Cartesian coordinate**s by the formula

$$y^3 = ax^2.$$

Christiaan **Huygens** showed that this curve satisfies the requirement requested by Gottfried **Leibniz** in 1687, namely, the curve along which a particle may descend under gravity so that it moves equal vertical distances in equal times. Neile's parabola is the **evolute** of a parabola.

nephroid

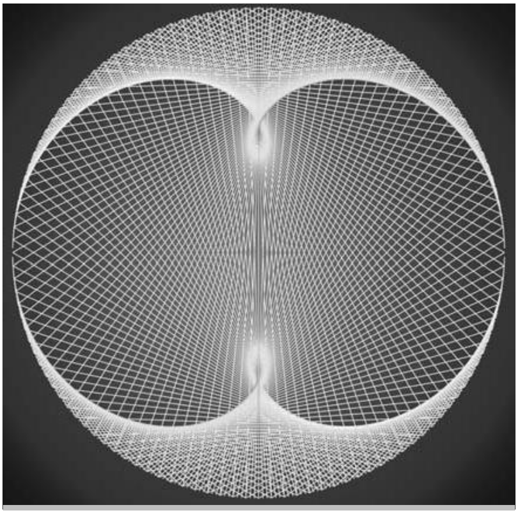

nephroid A nephroid curve spun by thread on a computer loom. *Jos Leys, www.josleys.com*

A type of curve often seen on the surface of a cup of coffee in the sunshine–a crescent of light formed by sunlight reflecting off the inside of the cup onto the surface of the drink. More generally, it is the shape made by parallel rays of light reflecting from the inside of any semicircle. In mathematical terms, this means that the nephroid is the cata**caustic** of a **circle** when the light source is at infinity, a fact first demonstrated by Christiaan **Huygens** in 1678 and published by him in his *Traité de la Lumière* (Treatise on light, 1690). A physical explanation wasn't forthcoming, however, until 1838 when George Airy gave a proof in terms of the wave theory of light. The name *nephroid* (from the Latin for "kidney-shaped") was introduced in 1878 by the English mathematician Richard Proctor in his book *The Geometry of Cycloids*. Prior to that it was known as a two-cusped **epicycloid**. Specifically, the nephroid is the epicycloid formed by a circle of radius *a* rolling around the outside on a fixed circle of radius 2*a*. It has a length of 24*a*, an area of $12\pi^2$, and is given by the parametric equations:

$$x = a(3\cos(t) - 3\cos(3t))$$
$$y = a(3\sin(t) - \sin(3t)).$$

The nephroid is the **involute** of **Cayley's sextic** and is also the **envelope** of circles with their centers on a given circle, touching a given diameter of that circle. The nephroid has been described as the perfect shape for a multiseat dining table. *Freeth's nephroid,* not to be mistaken for the ordinary nephroid just described, is named after the English mathematician T. J. Freeth (1819–1904) who first wrote about it in a paper published by the London Mathematical Society in 1879. Freeth's nephroid is the **strophoid** of a circle and has the polar equation $r = a(1 + 2\sin(\theta/2))$. Freeth's nephroid is also the name of a group of mathematicians, mostly from Royal Holloway College, London, who gather weekly in a pub called the Beehive and compete in games of trivial pursuit.

net

A drawing of a **polyhedron** unfolded along its **edge**s, so as to lie flat in a plane. The earliest known examples of nets to represent polyhedra are by Albrecht **Dürer**.

neural network

An electronic automaton, similar in some ways to a **cellular automaton**, that offers a highly simplified model of a brain. As such, a neural network is a device for machine learning that is based on associative theories of human cognition. Using various algorithms and weightings of different connections between "neurons," neural networks are set up to learn how to recognize a pattern in applications such as voice recognition, visual pattern recognition, robotic control, symbol manipulation, and decision making. Generally, they consist of three layers: input neurons, output neurons, and a layer in between where information from input to output is processed. Initially the network is loaded with a random program, then the output is measured against a desired output which prompts an adjustment in the weights assigned to the connections in response to the discrepancy between the actual and desired output. This is repeated many times so that the network effectively learns as a child does: in a sense, the net discovers its own rules. Changing the rules of interaction between the "neurons" in the net can lead to interesting emergent behavior, so that neural networks have become another tool for investigating **emergence** and **self-organization**.

Neusis construction

A geometric construction that breaks the strict rules of classical Greek straightedge-and-compass construction (see **constructible**) by allowing a marked ruler to be slid into different positions. Neusis construction makes possible **duplicating the cube** and **trisecting an angle**. John **Conway** and Richard **Guy** have also shown how Neusis constructions, based on angle trisection, can be used to draw regular polygons with 7, 9, and 13 sides.

Newcomb's paradox

One of the most simply stated but astonishing of the so-called prediction paradoxes that bear on the problem of free will. It was devised in 1960 by William Newcomb, a theoretical physicist at the Lawrence Livermore Laboratory, while contemplating the **prisoner's dilemma**. A superior being, with super-predictive powers that have never been known to fail, has put $1,000 in box *A* and either nothing or $1 million in box *B*. The being presents you with a choice: (1) open box *B* only, or (2) open both box *A* and *B*. The being has put money in box *B* only if it predicted you will choose option (1). The being put nothing in box *B* if it predicted you will do anything other than choose option (1) (including choosing option (2), flipping a coin, etc.). The question is, what should you do to maximize your winnings? You might argue that since your choice now can't alter the contents of the boxes you may as well open them both and take whatever's there. This seems reasonable until you bear in mind that the being has never been known to have made an incorrect prediction. In other words, in some peculiar way, your mental state is highly correlated with the contents of the box: your choice is linked to the probability that there's money in box B. These arguments and many others have been put forward in favor of either choice. The fact is there is no known "right" answer, despite the concerted attentions of many philosophers and mathematicians over several decades.[123]

Newton, Isaac (1642–1727)

One of the great intellectual giants in human history; an English mathematician, physicist, and sometime head of the Royal Mint, Newton was one of the chief architects of **calculus** (though not in its modern form), discovered the **binomial theorem**, established the principles of universal gravitation, and saw through numerous other developments, any one of which would have brought a lesser person fame. Interestingly, his most productive period was 1665–1666, his so-called "miraculous year," when Cambridge University was closed because of the plague and Newton had to work at home. He was born in the same year that Galileo died. After publishing his *Principia* (1687), the most important and influential scientific book ever written, his interests drifted toward theology, politics, and, for a while, alchemy. His last two decades were largely spent in acrimonious debate with Gottfried **Leibniz**, over priority in the discovery of calculus, and with the astronomer John Flamsteed.

Newton's method

An iterative method for finding the zeros of a **function**.

n-gon

A **polygon** with *n* sides.

Nim

A **game**, of which there are many different versions, that involves two players alternately removing at least one item from one of two or more piles or rows. The person who picks up the last item wins. In one form of the game, five rows of matches are laid out in such a way that there is one match in the first row, two matches in the second, and so on, down to five matches in the bottom row. Players take turns to remove any nonzero number of matches from any one row. The game may have originated in China. The name "Nim" was coined by Charles Bouton, an associate professor of mathematics at Harvard at the turn of the twentieth century, who took it from an archaic English word meaning to steal or to take away. In 1901 he published a full analysis of Nim and proof of a winning strategy.[50] The first Nim-playing computer, the Nimatron, a 1-ton behemoth, was built in 1940 by the Westinghouse Electrical Corporation and was exhibited at the New York World's Fair. It played 100,000 games against spectators and attendants, and won an impressive 90% of the time; most of its losses came at the hands of attendants who were instructed to reassure incredulous onlookers that the machine could be beaten! In 1951 a Nim-playing robot, the Nimrod, was shown at the Festival of Britain, and later at the Berlin trade fair. It was so popular that spectators entirely ignored a bar at the other end of the room where free drinks were being offered. Eventually the local police had to be called in to control the crowds.[110]

nine

A number long considered to have strange, mystic properties. A phrase in a book written during the Dark Ages gave rise to the superstition that cats have nine lives. English author and satirist William Baldwin wrote in his *Beware the Cat,* "It is permitted for a witch to take her cat's body nine times." There were nine Muses, nine rivers of Hades, and nine heads on the Hydra. It took nine days for Vulcan to fall from the heavens. The phrase "nine days' wonder" comes from the proverb "a wonder lasts nine days and then the puppy's eyes are open." A cat-o'-nine-tails is a whip, usually made of nine knotted lines or cords fastened to a handle that produces scars like the scratches of a cat. Being on "cloud nine" may have its origin in Dante's ninth heaven of Paradise, whose inhabitants are blissful because they are closest to God.

The term "the whole 9 yards" came from World War II fighter pilots in the Pacific. When arming their planes on the ground, the .50-caliber machine gun ammo belts measured exactly 27 feet, before being loaded into the fuselage. If the pilots fired all their ammo at a target, it got "the whole 9 yards." Less certain–though there is no shortage of theories–is the source of the expression "dressed to the nines."

Nine is the largest single-digit number and the one that occurs least frequently in most situations; an exception is the tendency of businesses to set prices that end with one or more nines. Because nine is one less than the base of our number system, it is easy to see if a number is divisible by 9 by adding the digits (and repeating on the result if necessary: the result should be nine). This process is sometimes called **casting out nines**. Similar processes can be developed for divisibility by 99, 999, etc. or any number that divides one of these numbers. Nine has many other interesting properties. For example, write down a number containing as many digits as you like, add these digits together, and deduct the sum from the first number. The sum of the digits of this new number will always be a multiple of nine.

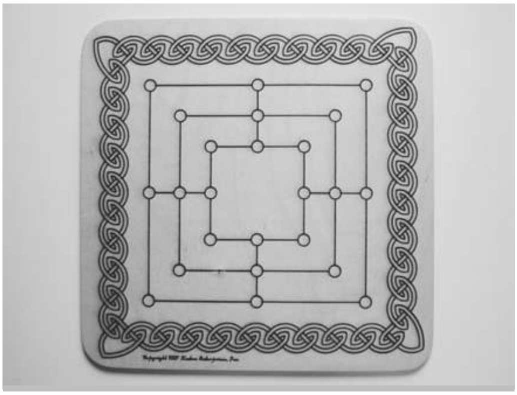

nine men's morris A modern version of the playing board.
Kadon Enterprises, Inc., www.puzzlegames.com

nine holes

See **three men's morris**.

nine men's morris

> *The nine-men's morris is filled up with mud;*
> *And the quaint mazes in the wanton green,*
> *For lack of tread, are indistinguishable.*
> –*Midsummer Night's Dream* (Act 2, scene 1),
> Shakespeare

One of the oldest of board games, known by different names and played with variations of rules in different places and periods. In France it is Marelle, in Austria it is Muhle, and in England it was known as Peg Meryll, Meg Marrylegs, and other names, all referring to a "mill" because that is the name of a run of three counters in the game. Versions of it have been found etched into the roof of the Temple of Kurna in Egypt (dated to about 1400 B.C.), cut into the oak planks that form the deck of the great Viking ship discovered at Gokstad in 1880, and carved in the choir stalls of several English cathedrals. A typical board layout is shown in the accompanying figure; diagonal lines may or may not be included. Joseph Strutt in *The Sports and Pastimes of the People of England* (1801) described the rules in this way:

> Two persons, having each of them nine pieces, or men, lay them down alternately, one by one upon the spots; and the business of either party is to prevent his antagonist from placing three of his pieces so as to form a row of three, without the intervention of an opponent piece. If a row be formed, he that made it is at liberty to take up one of his competitor's pieces from any part he thinks most to his advantage; excepting he has made a row, which must not be touched if he have another piece upon the board that is not a component part of that row. When all the pieces are laid down, they are played backwards and forwards, in any direction that the lines run, but only can move from one spot to another (next to it) at one time. He that takes off all his antagonist's pieces is the conqueror.

In 1996, the German mathematician Ralph Gasser used a computer to prove that nine men's morris is a guaranteed draw if both players make optimal moves from the outset. He programmed the computer to figure out and tabulate 10 billion positions that were known to be a win for one side or the other, then worked forward 18 moves from the beginning of the game until his opening analysis met his endgame analysis. As a result he showed that every potentially winning position could be countered by the opponent in the early stages of the game. See also **three men's morris**.[132]

nine rooms paradox

A puzzle that was first published in *Current Literature* vol. 2, April 1889. It takes the form of a poem and is similar to that of the **missing dollar**.

PUZZLE

Ten weary, footsore travellers,
All in a woeful plight,
Sought shelter at a wayside inn
One dark and stormy night.

'Nine rooms, no more,' the landlord said
'Have I to offer you.
To each of eight a single bed,
But the ninth must serve for two.'

A din arose. The troubled host
Could only scratch his head,
For of those tired men not two
Would occupy one bed.

The puzzled host was soon at ease—
He was a clever man—
And so to please his guests devised
This most ingenious plan.

In a room marked A two men were placed,
The third was lodged in B,
The fourth to C was then assigned,
The fifth retired to D.

In E the sixth he tucked away,
In F the seventh man.
The eighth and ninth in G and H,
And then to A he ran,

Wherein the host, as I have said,
Had laid two travellers by;
Then taking one—the tenth and last—
He logged him safe in I.

Nine single rooms—a room for each—
Were made to serve for ten;
And this it is that puzzles me
And many wiser men.

How has the host managed to bamboozle his patrons?

Solutions begin on page 369.

nine-point circle

Draw a triangle, any triangle (although it may be best to start with an acute triangle). Mark the midpoints of each side. Drop an altitude from each vertex to the opposite side, and mark the points where the altitudes intersect the opposite side. (If the triangle is obtuse, an altitude will be outside the triangle, so extend the opposite side until it intersects.) Notice that the altitudes intersect at a common point. Mark the midpoint between each vertex and this common point. No matter what triangle you start

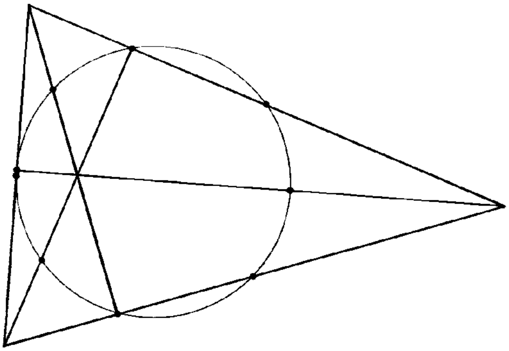

nine-point circle

with, these nine points all lie on a perfect circle! This result was known to Leonhard **Euler** in 1765, but was rediscovered by the German mathematician Karl Feuerbach (1800–1834) in 1822.

node

See **crunode**.

Noether, Emmy (Amalie) (1882–1935)

A German mathematician, one of the most talented of the early twentieth century. The crucial result now known as *Noether's theorem,* which is important in other symmetries in natural systems, is of great importance in physics. She received her doctorate in 1907 and rapidly built an international reputation, but the University of Göttingen refused to let her teach, and her colleague David **Hilbert** had to advertise her courses in the university's catalog under his own name. A long controversy ensued, with her opponents asking what the country's soldiers would think when they returned home and were expected to learn at the feet of a woman. Allowing her on the faculty would also mean letting her vote in the academic senate. Said Hilbert, "I do not see that the sex of the candidate is against her admission as a privatdozent. After all, the university senate is not a bathhouse." She was finally admitted to the faculty in 1919. A Jew, Noether was forced to flee Nazi Germany in 1933 and joined the faculty at Bryn Mawr in the United States.

non-Abelian

Non**commutative** or order-dependent. For example, the group of manipulations of **Rubik's cube** is non-Abelian because the state of the cube depends greatly on the order of the moves performed on it. See also **Abelian group**.

nonagon

A **polygon** with nine sides. A *nonagonal number* is a number of the form $n(7n - 5)/2$.

nonconvex uniform polyhedron

A **uniform polyhedron** of a type obtained by relaxing the conditions used to produce the **Archimedean solid**s (which have regular convex faces and identical convex vertices) to allow both nonconvex faces and vertex types, as in the case of the **Kepler-Poinsot solid**s. The condition that every vertex must be identical, but the faces need not be, gives rise to 53 nonconvex uniform polyhedra. An example is the *great truncated dodecahedron,* obtained by truncating the corners of the great dodecahedron at a depth which gives regular decagons.

non-Euclidean geometry

Any **geometry** in which Euclid's **parallel postulate** doesn't hold. (One way to state the parallel postulate is: given a straight line and a point A not on that line, there is only one exactly straight line through A that never intersects the original line.) The two most important types of non-Euclidean geometry are **hyperbolic geometry** and **elliptical geometry**. The different models of non-Euclidean geometry can have positive or negative **curvature**. The sign of curvature of a surface is indicated by drawing a straight line on the surface and then drawing another straight line perpendicular to it: both these lines are **geodesic**s. If the two lines curve in the same direction, the surface has a positive curvature; if they curve in opposite directions, the surface has negative curvature. Elliptical (and spherical) geometry has positive curvature whereas hyperbolic geometry has negative curvature.

The discovery of non-Euclidean geometry had immense consequences. For more than 2,000 years, people had thought that **Euclidean geometry** was the only geometric system possible. Non-Euclidean geometry showed that there are other conceivable descriptions of space–a realization that transforms mathematics into an altogether more abstract science. Thereafter, it was clear that in mathematics, one could start out with *any* set of self-consistent

Types of Geometry

	Euclidean	**Elliptical**	**Hyperbolic**
Curvature	Zero	Positive	Negative
Given a line m and a point P not on m, the number of lines passing through P and parallel to m	1	0	Many
Sum of interior angles of a triangle	180°	> 180°	< 180°
Square of hypotenuse of a right triangle with sides a and b	$a^2 + b^2$	$< a^2 + b^2$	$> a^2 + b^2$
Circumference of a circle with diameter 1	π	$< \pi$	$> \pi$

postulates and follow through their ramifications. The discovery of non-Euclidean geometry has been compared with Copernicus's theory and Einstein's **relativity theory**, the analogy being that each freed people from long-held models of thought. In fact, Einstein said about non-Euclidean geometry: "To this interpretation of geometry, I attach great importance, for should I have not been acquainted with it, I never would have been able to develop the theory of relativity."

It's important to realize that both Euclidean and non-Euclidean geometry are consistent in that the assumptions on which they rest don't involve any contradictions. In response to a question as to which geometry is true, Henri **Poincaré** said: "One geometry cannot be more true than the other; it can only be more convenient." Which geometry is valid in the physical space in which we live? On a small scale, and for all practical purposes on Earth, Euclidean geometry works just fine. But on larger scales this is no longer true. Einstein's general theory of relativity uses non-Euclidean geometry as a description of **space-time**. According to this idea, space-time has a positive curvature near gravitating matter and the geometry is non-Euclidean. When a body revolves around another body, it appears to move in a curved path due to some force exerted by the central body, but it is actually moving along a **geodesic**, without any force acting on it. Whether all of space-time contains enough matter to give itself an overall positive curvature is one of the many unanswered question in physics today, but it is generally accepted that the geometry of space-time is more non-Euclidean than Euclidean. It is proposed that if space-time does happen to have an overall positive curvature, then the universe will stop expanding after a fixed amount of time and start to shrink resulting in a "big crunch," as opposed to the "big bang" that resulted in its creation. At the moment, astronomical observations seem to favor a universe that is "open" and has a hyperbolic geometry. Another consequence of non-Euclidean geometry is the possibility of the existence of a **fourth dimension**. Just as the surface of the sphere curves in the direction of the third dimension, i.e., perpendicular to its surface, it is believed that space-time curves in the direction of the fourth dimension. Non-Euclidean geometry has applications in other areas of mathematics, including the theory of elliptic curves, which was important in the proof of **Fermat's last theorem**.[72, 140] (See table, "Types of Geometry.")

nonlinear system

Any system in which the data points coming from the measurement of the values of its variables can be represented as a curvilinear pattern on a coordinate plane; hence, "nonlinear" for "not-a-line." More generally, a system in which small changes can result in large effects, and large changes in small effects. Thus, sensitive dependence on initial conditions (see **butterfly effect**) in chaotic systems illustrates the extreme nonlinearity of these systems. In a nonlinear system the components are interactive, interdependent, and exhibit feedback effects.

nonstandard analysis

In a broad sense, the study of the infinitely small; more specifically, the study of **hyperreal numbers**, their functions and properties. Nonstandard analysis, which was pioneered by Abraham **Robinson** in the 1960s, puts the concept of **infinitesimals** on a firm mathematical footing and is, for many mathematicians, more intuitive than real **analysis**.

normal

A line that is perpendicular to a given line or plane.

normal number

A number in which digit sequences of the same length occur with the same frequency. A constant is considered normal to base 10 if any single digit in its decimal expansion appears one-tenth of the time, any two-digit combination one-hundredth of the time, any three-digit combination one-thousandth of the time, and so on. In the case of **pi**, the digit 7 is expected to appear 1 million times among the first 10 million digits of its decimal expansion. It actually occurs 1,000,207 times–very close to the expected value. Each of the other digits also turns up with approximately the same frequency, showing no significant departure from predictions. A number is said to be *absolutely normal* if its digits are normal not only to base 10 but also to every integer base greater than or equal to 2. In base 2, for example, the digits 1 and 0 would appear equally often. Émile **Borel** introduced the concept of normal numbers in 1909 as a way to characterize the resemblance between the digits of a mathematical constant such as π and a sequence of **random numbers**. He quickly established that there are lots of normal numbers, though finding a specific example of one proved to be a major challenge. The first to be found was **Champernowne's number**, which is normal to base 10. Analogous normal numbers can be created for other bases. To date, no specific "naturally occurring" **real number** has been proved to be absolutely normal, even though it is known that almost all real numbers are absolutely normal! However, in 2001, Greg Martin of the University of Toronto found some examples of the opposite extreme–real numbers that are normal to no base whatsoever. To start with, he noted that every

rational number is absolutely abnormal. For example, the fraction 1/7 can be written in decimal form as 0.1428571428571 . . . . The digits 142857 just repeat themselves. Indeed, an expansion of a rational number to any base b or b^k eventually repeats. Martin then focused on constructing a specific *irrational* absolutely abnormal number. He nominated the following candidate, expressed in decimal form, for the honor:

$$\alpha = 0.6562499999956991\underline{999999\ldots9999998}528404201690728\ldots$$

The middle portion (underlined) of the given fragment of α consists of 23,747,291,559 nines. Martin's formulation of this number and proof of its absolute abnormality involved so-called **Liouville number**s.[216]

nothing

The absence of anything; nonexistence. Nothing is not the same as the **empty set**, which exists as the set that mathematically denotes nothing, nor is it the same as **zero**, which exists as the number that denotes how many members the empty set contains. In physics, nothing is not a **vacuum**, because a vacuum not only contains energy but exists in space and time; nor is it a singularity, which contains a great deal of concentrated matter and energy. So, can there be nothing? No. "To be" implies existence of some sort: the one thing we can be absolutely sure has never existed, or will exist, is nothing.

noughts and crosses

See **tic-tac-toe**.

NP-hard problem

A mathematical problem for which, even in theory, no shortcut or smart **algorithm** is possible that would lead to a simple or rapid solution. Instead, the only way to find an optimal solution is a computationally intensive, exhaustive analysis in which all possible outcomes are tested. Examples of NP-hard problems include the **traveling salesman problem** and the popular game **Tetris**. NP stands for "non-deterministic polynomial-time."

nucleation

A process in a physical system, or a mathematical model such as a **cellular automaton** or a statistical model, whereby a bubble or other structure appears spontaneously at a random or unpredictable spot.

null hypothesis

The **hypothesis** that is being tested in a hypothesis-testing situation.

null set

See **empty set**.

number

An abstract measure of quantity. The most familiar numbers are the **natural number**s, 0, 1, 2, . . . , used for counting. If negative numbers are included, the result is the **integer**s. Ratios of integers are called **rational number**s, which can be expressed as terminating or repeating decimals. If all infinite and nonrepeating decimal expansions are thrown in as well, the scope of numbers extends to all **real numbers**, which can be extended to the **complex numbers** in order to include all possible solutions to algebraic equations. More recent developments are the **hyperreal number**s and the **surreal number**s, which extend the real numbers by adding infinitesimal and infinitely large numbers. For measuring the size of infinite **set**s, the natural numbers have been generalized to the **ordinal number**s and to the **cardinal number**s. See also **numeral** and **number system**.

number line

A way of representing numbers by thinking of them as the positions of points on a line.

number system

A way of counting using a particular **base**. The familiar **decimal** number system is a base 10 system, because it uses 10 different digits, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, and combinations of these. Other number systems, used for special purposes or by certain cultures, are the **binary** (base 2), trinary (base 3), hexadecimal (base 16), **vigesimal** (base 20), and **sexagesimal** (base 60). In a *positional number system* of base b, b basic symbols (or digits) corresponding to the first b **natural number**s, including zero, are used. To generate the rest of the numbers, the position of the symbol in the figure is used. The symbol in the last position has its own value, and as it moves to the left its value is multiplied by b. In this way, with only finitely many different symbols, every number can be expressed. This is unlike systems, such as that of **Roman numeral**s, which use different symbols for different orders of magnitude.

number theory

The study of the whole numbers and their properties and relationships. Often, a problem in number theory can be restated in terms of finding the solution, or showing that there is no solution, to a **Diophantine equation**. A Diophantine equation is one where the coefficients are whole numbers and where the solution is also constrained to be a whole number. Sometimes, what appears to be a simple Diophantine equation can lead to an **elliptic curve**.[30]

numeral

A symbol, or combination of symbols, that describes a number. See **Arabic numeral** and **Roman numeral**.

numerator

The number above the fraction bar that indicates the number of parts of the whole there are in a **rational number**.

numerical analysis

The study of methods for approximating the solutions of various classes of mathematical problems including *error analysis*.

O

obelus

The name of the symbol "÷", used as a sign for division. It comes from the Greek *obelos* meaning a pointed stick (a spit) used for cooking. This root word also gave rise to *obelisk* for a pointed stone pillar. The "÷" symbol was originally used as an editing mark in early manuscripts, sometimes only as a line without the two dots, to point out material that the editor thought needed cutting. It was also used occasionally as a symbol for subtraction. As a division symbol it was first employed by the Swiss mathematician Johann Rahn (1622–1676) in his *Teutsche Algebra* in 1659. By a misunderstanding of a credit to John Pell about other material in the book, many English writers started using the symbol and calling it "Pell's notation." Although it appears regularly in literature produced in Britain and the United States, it is virtually unknown in the rest of the world.

oblate spheroid

An **ellipsoid** produced by rotating an **ellipse** through 360° about its **minor axis**.

oblique

Slanted or not perpendicular. *Oblique coordinates* are measured on a plane coordinate system whose axes are not perpendicular. An *oblique angle* is any angle except a **right angle** and an *oblique triangle* is any triangle that doesn't contain a right angle. The first reference to "oblique angled triangles" may have been by Thomas Blundevil in his *Exercises* in 1594.

oblong

An alternate name for any **rectangle** that is not a square. The word comes from the Latin *ob* ("excessive") and *longus* ("long").

oblong number

Any positive integer that is not a **perfect square**.

obtuse

An *obtuse angle* is an **angle** greater than 90° but less than 180°. An *obtuse triangle* is a **triangle** that contains an obtuse angle. The word comes from the Latin *ob* ("against") and *tundere* ("to beat") and thus refers to things that are blunt, dull, or rounded.

octa-

The Greek prefix meaning "eight."

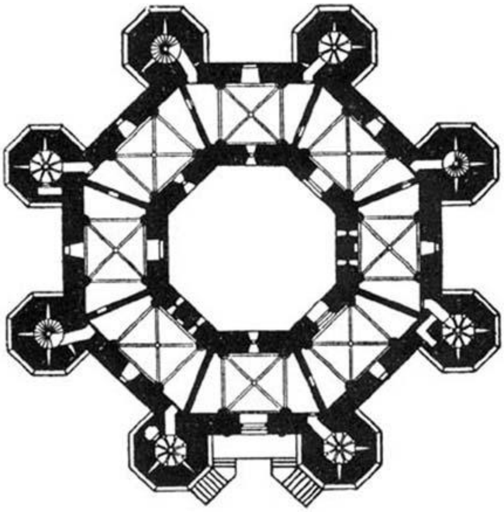

octa- An aerial view of Castel del Monte in Italy's Apulien region.

octagon

An eight-sided **polygon**. One of the most remarkable examples of octagonal design in architecture is the Castel del Monte, in southern Italy, which consists of a central octagonal core containing an inner octagonal courtyard that is surrounded by eight tall, perfectly octagonal towers. Another famous octagon is Oxford University's Radcliffe Observatory tower, built from 1772 to 1794, vaguely based on the Tower of the Winds in Athens, and considered architecturally the finest observatory in Europe.[139]

octahedron

A **polyhedron** with eight faces, each of which is an equilateral triangle. A regular octahedron, whose sides are all equilateral triangles, is one of the **Platonic solid**s; it looks like two square pyramids with their square bases stuck together. See also **octa-**.

octant

Any one of the eight portions of space determined by the three coordinate planes.

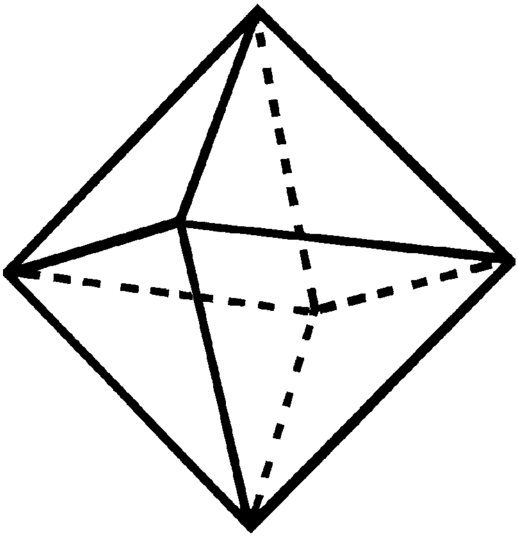

octahedron One of the Platonic solids.

octonion

Also known as a *Cayley number.* Octonions are a non**associative** generalization of the **quaternion**s and the **complex number**s involving numbers with one real **coefficient** and seven imaginary coefficients.

odd

Not divisible by two. An *odd function* is a **function** $f(x)$ with the property that $f(-x) = -f(x)$ for any value of x; an example is $\sin(x)$.

officer problem

See **thirty-six officers problem**.

omega

See **Chaitin's constant**.

one

The first positive **integer** and the first odd number; it is also known as *unity.* From the time of Euclid to the late 1500s, one wasn't generally considered to be a number but instead was thought of as the unit of which bonafide numbers were composed. The Old English (c. 550–c. 1100) *ane* served both for counting and as the indefinite article. Toward the end of the Old English period and the beginning of Middle English (c. 1100–c. 1500), *ane* developed two pronunciations, the first being used for 1 and the other for the indefinite article *an, a.* The existence of different words for the number one and the indefinite article seems to be unique to the English language. "One" can be traced back to the Latin *unus* and the Greek *oine* but probably came into English from the German *eine.*

153

A number with some very curious properties. It is the smallest number that can be expressed as the sum of the cubes of its digits: $153 = 1^3 + 5^3 + 3^3$. It is equal to the sum of the factorials of 1 to 5: $153 = 1! + 2! + 3! + 4! + 5!$ The sum of the digits of 153 is a perfect square: $1 + 5 + 3 = 9 = 3^2$. The sum of the **aliquot part**s of 153 is also a perfect square: $1 + 3 + 9 + 17 + 51 = 81 = 9^2$. On adding the number 153 to its reverse (351), the result is 504 whose square is the smallest square that can be expressed as the product of two different numbers that are the reverse of one another: $153 + 351 = 504$; $504^2 = 288 \times 882$. It can be expressed as the sum of all integers from 1 to 17. In other words, 153 is the seventeenth triangular number; its reverse is also a triangular number. In addition, 153 and its reverse are **Harshad number**s. It can be expressed as the product of two numbers formed from its own digits: $153 = 3 \times 51$. A reference to 153 occurs in the New Testament: the net that Simon Peter drew from the Sea of Tiberias held 153 fishes.

one-to-one

A **function** or map that for every possible output has only one input that yields that particular output; if $f(a) = f(b)$, then $a = b$.

open

An *open interval* is a piece of a straight line that doesn't contain its endpoints; a *half-open interval* contains one endpoint. An *open set* is one in which every point in the set has a neighborhood lying in the set.

operator

Something that acts on a **function** to give another function.

optical illusion

A picture or figure that deceives or confuses the eye and/or brain. Categorization of optical illusions is difficult because several underlying mechanisms may contribute to an effect, or the cause of the illusions may not be completely understood. For the purposes of this book illusions are grouped as shown in the table "Some Types and Examples of Optical Illusions." See individual entries for details. See also **anamorphosis**.

Some Types and Examples of Optical Illusions

Type	Examples
Distortion illusion	Fraser spiral Müller-Lyer illusion Orbison's illusion Poggendorff illusion Titchener illusion Zöllner illusion
Impossible figures and objects	Freemish crate Penrose triangle Penrose stairway Tribar illusion
Ambiguous figure	Ames room Necker cube Schröder's reversible staircase Thiery figure
Lateral inhibition illusion	Hermann grid illusion
Antigravity houses and hills	

Orbison's illusion

A **distortion illusion** in which a background of radiating lines appears to distort the shape of a superimposed figure, such as a square or circle.

orbit

(1) The path, in the form of a **conic section**, that an object takes when under the gravitational influence of another body. (2) More generally, the trajectory of a **differential equation**.

order

A word with many different meanings in mathematics. Among these are: (1) the sequence in which a set of objects or numbers is placed; (2) the number of elements in a set; (3) the number of times a shape can be fitted back onto its own outline during a complete turn (*order of symmetry*); (4) the highest order **power** in a one-variable **polynomial** (*order of a polynomial*, also known as its *degree*); (5) the type of curve described by such a polynomial (*order of a curve*).

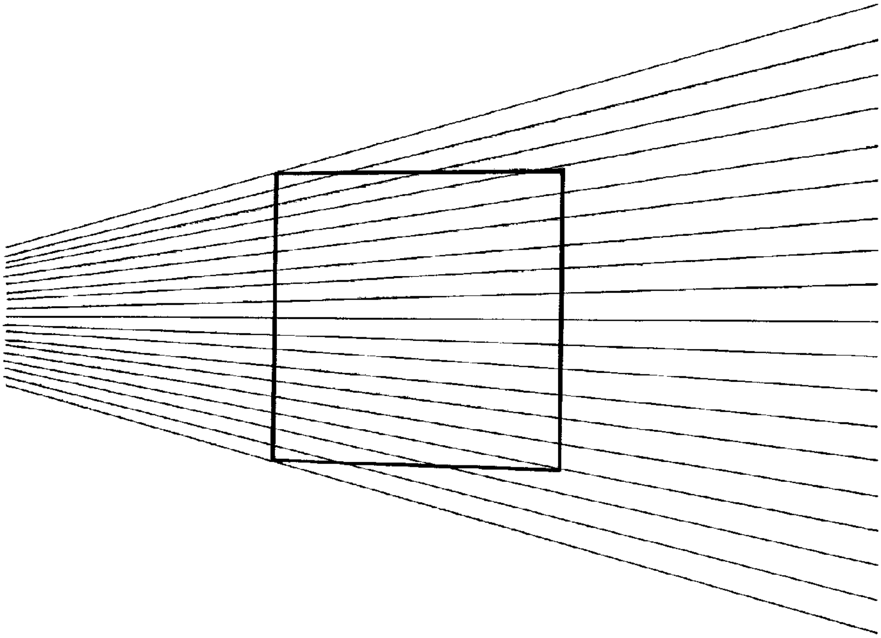

Orbison's illusion Is the square really a square?

ordered pair

A collection of two objects such that one can be distinguished as the *first element* and the other as the *second element.* An ordered pair with first element a and second element b is usually written as (a, b). Two such ordered pairs (a_1, b_1) and (a_2, b_2) are equal if and only if $a_1 = a_2$ and $b_1 = b_2$. *Ordered triples* and *ordered n-tuples* (ordered lists of n terms) are defined in the same way. An ordered triple (a, b, c) can be defined as $(a, (b, c))$, that is, as two nested pairs.

ordinal number

A number used to give the position in an ordered sequence: first, second, third, fourth, ...Ordinal numbers are distinct from **cardinal numbers** (one, two, three, four, . . .), which describe the *size* of a collection. The mathematician Georg **Cantor** showed in 1897 how to extend the concept of ordinals beyond the **natural numbers** to the infinite and how to do arithmetic with the resulting *transfinite ordinals* (see **infinity**).

ordinary differential equation

Any equation relating a **function** of one variable to its **derivatives**. Compare with **partial differential equation**.

ordinate

The y-coordinate, or vertical distance from the x-axis, in a system of **Cartesian coordinates**. Compare with **abscissa**.

origami

The Japanese art of the paper-folding, without cutting and joining, to form definite objects.

origin

The point (0, 0) on the coordinate axes.

orthic triangle

The triangle whose vertices (see **vertex**) are the feet of the **altitudes** of a given triangle.

orthocenter

The point of intersection of the **altitudes** of a **triangle.**

orthogonal

At right angles to; independent of. *Orthogonal curves* are two families of curves with the property that each member of one family meets members of the other family at right angles.

osculating

To share the same **tangent** and **curvature** and at a given point, as in the case of two *osculating curves.* For example the curve $y = x^3$ and the x-axis osculate at the origin.

Oughtred, William (1574–1660)

An English clergyman and mathematician who invented the **slide rule** and who, in the first English edition of his *Clavis Mathematicae* in 1647, first used the name **"pi"** for the number 3.141 . . . . He wrote it $\pi.\delta$, where π stood for the English word *periphery* (what we would call circumference), the dot was his symbol for division, and δ stood for the English word *diameter.* The use of π to represent words starting with the letter *p,* like *periphery,* was not uncommon. Before $\pi = 3.14\ldots$ caught on, π was variously used to indicate a point, a polygon, a positive number, a power, a proportion, the number of primes in a series, and a factorial (which is a product). Oughtred used the same notation in all later English and Latin editions of his book, but not in the earlier first Latin edition. He was a prodigious inventor of mathematical symbols, though most of them have not survived. He did, however, introduce a couple of other symbols we still use: × for multiplication (as distinct from the letter x, which he also used for this purpose) and ± for "plus or minus." Oughtred served as rector of Albury from 1608 or 1610 until his death. There he tutored many young mathematicians of the time, including John **Wallis**, Seth Ward, Charles Scarburgh, and Christopher Wren. Moreover, all English mathematicians for the next century, including Isaac **Newton**, learned algebra from *Clavis Mathematicae* (first published in 1631).

oval

A curve that looks like a squashed **circle** but, in contrast with the **ellipse**, doesn't have a precise mathematical definition. The word *oval* comes from the Latin *ovus* for "egg." Unlike ellipses, ovals sometimes have only a single axis of reflection symmetry (instead of two).

Ovid's game

A board game, described by the Roman poet Ovid in Book III of his *Art of Love,* that was popular in ancient Greece, Rome, and China. Each player has three counters, which they place alternately on to nine points laid out in a 3 × 3 grid, with the object of getting three in a line and so winning. This may sound like **tic-tac-toe**, and there is no doubt that both Ovid's game and **nine men's morris** are antecedents of the familiar noughts and crosses. But in Ovid's game play continues after the six counters are down, if no one has yet won, by moving a single counter on each turn, though not diagonally, to any adjacent square. As in the case of tic-tac-toe, two experts (i.e., "rational" players) will always draw. Because the first player is ensured a win by covering the center square, this move is usually not allowed. Ovid advised women to master the game in order to gain the attention of men!

Ozanam, Jacques (1640–1717)

A French mathematician, scientist, and writer best remembered for his book on mathematical and scientific puzzles *Récréations Mathématiques et Physiques* (four volumes, 1694), which later went through 10 editions. Based on earlier works by Claude **Bachet**, Claude **Mydorge**, Jean **Leurechon**, and Daniel Schwenter (1585–1636), it was later revised and enlarged by Jean **Montucla**, then translated into English by Charles Hutton (1803, 1814). Edward Riddle edited a new edition, which was published in 1844, removing some old material and adding new material so that "[T]he work might continue to be to the present generation a useful manual of scientific recreation, as its predecessors have been to the generation which has passed." Ozanam's original edition contained an early example of a problem about orthogonal **Latin squares**: "Arrange the 16 court cards so that each row and each column contains one of each suit and one of each value."

P

Pacioli, Luca (1445–1517)

An Italian mathematician and Franciscan monk who wrote several influential books. His encyclopedic *Summa de Arithmetica, Geometria, Proportioni et Proportionalita* (Summary of arithmetic, geometry, proportion and proportionality, 1494) summarized what was known about contemporary arithmetic, algebra, geometry, and trigonometry and gave a basis for the major progress in mathematics which subsequently took place in Europe. *Divina proportione* (1509), with drawings by none other than Leonardo da Vinci (surely no mathematical text was more impressively illustrated!), deals with the **golden ratio**, a subject that Pacioli treats from an architectural standpoint in a second volume. At his death, he left a major book unpublished, *De Viribus Amanuensis,* on recreational problems, geometrical problems, and proverbs. It makes frequent reference to Leonardo, who assisted him with the project: many of the problems in this treatise are also in Leonardo's notebooks. Again it is a work for which Pacioli claimed no originality, describing it as a compendium.

packing

A way to place objects of the same kind so that they touch in some specified way, often inside a container with specified properties. The objects to be packed may be polyhedra, polygons, spheres, ellipsoids, hyperspheres, or any other type of shape, and the number of dimensions involved may range upward from two. The fraction of a space filled by a given collection of objects is called the *packing density.* The densest packing of circles in the plane is the hexagonal lattice of the bee's honeycomb, which has a packing density of 0.9069 . . . . In 1611, Johannes **Kepler** proposed that hexagonal, or face-centered cubic, packing is also the densest possible way to arrange spheres in three dimensions—an assertion known as **Kepler's conjecture**. Currently, the worst known **convex** packer in two-dimensions is the smoothed octagon, with a packing density of about 0.902. Stanislaw **Ulam** conjectured that the sphere was the worst packing object in three-dimensional space.

Padovan sequence

See **plastic number**.

palindrome

A word, verse, sentence, or passage that reads the same forward or backward; the term comes from the Greek *palindromos* for "running back again." Well-known examples include: "Madam, I'm Adam;" "A man, a plan, a canal—Panama!;" and "Able was I ere I saw Elba." A slightly longer one, devised by Peter Hilton, a codebreaker on the British team that cracked the German Enigma, is "Doc, note. I dissent. A fast never prevents a fatness. I diet on cod." Credit for inventing the palindrome is often given to Sotades the Obscene of Maronea (third century B.C.). Though only eleven lines of his work have survived, he is thought to have recast the entire Iliad in palindromic verse. Sotades also wrote lines (now sometimes called Sotadic verses) that, when read backward, had the opposite meaning. His acid tongue eventually landed him in jail by order of Ptolemy II, though worse was to follow. Sotades escaped but was captured by Ptolemy's admiral Patroclus, who sealed him in a leaden chest and tossed him into the sea. A musical palindrome is formed by Haydn's Symphony No. 47 in G, sometimes referred to as *The Palindrome,* because in both the minuet and the trio the orchestra plays the music twice forward and then twice backward to arrive at the beginning. See also **palindromic number**.

palindromic number

A number such as 1,234,321 that reads the same forward and backward; more generally, a symmetrical number written in some **base** a as $a_1\ a_2\ a_3 \ldots | \ldots a_3\ a_2\ a_1$. In the familiar base 10 system, there are nine two-digit palindromic numbers: 11, 22, 33, 44, 55, 66, 77, 88, 99; there are 90 palindromics with three digits: 101, 111, 121, 131, 141, . . . , 959, 969, 979, 989, 999; and there are 90 palindromics with four digits: 1,001; 1,111; 1,221; 1,331; 1,441; . . . , 9,559; 9,669; 9,779; 9,889; 9,999, giving a total of 199 palindromic numbers below 10^4. Below 10^5 there are 1,099 palindromics and for other exponents of 10^n there are 1,999; 10,999; 19,999; 109,999; 199,999; 1,099,999; . . . . It is conjectured, but has not been proven, that there are an infinite number of palindromic **prime number**s. With the exception of 11, palindromic primes must have an odd number of digits. A normally quick way to produce a palindromic number is to pick a positive integer of two or more digits, reverse the digits, and add to the original, then repeat this process with the new number, and so on. For example, 3,462 gives the sequence 3,462; 6,105; 11,121; 23,232. Does the series formed by adding a number to its reverse always end in a palindrome? It used to be thought so. However, this conjecture has been proven false for bases 2, 4, 8, and other powers of 2, and seems to be false for base 10 as well. Among the first 100,000 numbers, 5,996 numbers are known that have not produced palindromic numbers by the add-and-reverse method in calculations carried out to date. The first few of these are 196; 887; 1,675; 7,436; 13,783; 52,514; . . . . A proof that these numbers never produce palindromes, however, has yet to be found. The largest known palindromic prime, containing 30,913 digits, was found by David Broadhurst in 2003.

pandiagonal magic square

A **magic square** in which all the broken diagonals as well as the main diagonals add up to the magic constant.

pandigital number

An integer that contains each of the digits from zero to nine exactly once and whose leading digit is nonzero. The smallest pandigital numbers are 1,023,456,789; 1,023,456,798; 1,023,456,879; 1,023,456,897; and 1,023,456,978. Like all pandigitals these are divisible by nine. The first few "zeroless" pandigitals are 123,456,789; 123,456,798; 123,456,879; 123,456,897; 123,456,978; and 123,456,987; and the first few zeroless pandigital primes are 1,123,465,789; 1,123,465,879; 1,123,468,597; 1,123,469,587; and 1,123,478,659. The sum of the first 32,423 (a **palindromic number**) consecutive primes is 5,897,230,146, which is pandigital. No other palindromic number shares this property. Examples of palindromic numbers that are the product of pandigital numbers are 2,970,408,257,528,040,792 (= 1,023,687,954 × 2,901,673,548) and 5,550,518,471,748,150,555 (= 1,023,746,895 × 5,421,768,309), both found in 2001. See also **pandigital product**.

pandigital product

A product in which the digits of the multiplicand, multiplier, and product, taken together, form a **pandigital number**.

pangram

A phrase or sentence that contains every letter of the alphabet at least once. The best known example in English and a familiar typing test is "The quick brown fox jumps over the lazy dog." One of the shortest known that still makes some kind of sense is "The five boxing wizards jump quickly."

Pappus of Alexandria (c. A.D. 300)

The last of the great Greek geometers whose eight-volume *Mathematical Collection* summarized the bulk of mathematics known at that time. In this compendium, Pappus added a considerable number of his own explanations and amplifications of the earlier work of **Euclid**, **Archimedes**, **Apollonius**, and others.

parabola

One of the **conic section**s and one of the most studied curves in the history of mathematics. A parabola is the outline of the figure obtained if a right circular **cone** is cut by a plane that is exactly parallel to the cone's side. Just as the circle is a limiting case of the **ellipse** when the two foci coincide, the parabola is a limiting case of the ellipse when one of the foci is moved to infinity. As the French mathematician Henri Fabre eloquently put it, the parabola is an ellipse that "seeks in vain for its second, lost center." Johannes **Kepler** drew this further connection between the parabola and other conic sections: "Because of its intermediate nature the parabola occupies a middle position [between the ellipse and the **hyperbola**]. As it is produced it does not spread out its arms like the hyperbola but contracts them and brings them nearer to parallel, always encompassing more, yet always striving for less–whereas the hyperbola, the more it encompasses, the more it tries to obtain."

A parabola is the **locus** of all points in a plane that are equidistant from a given line, known as the *directric,* and

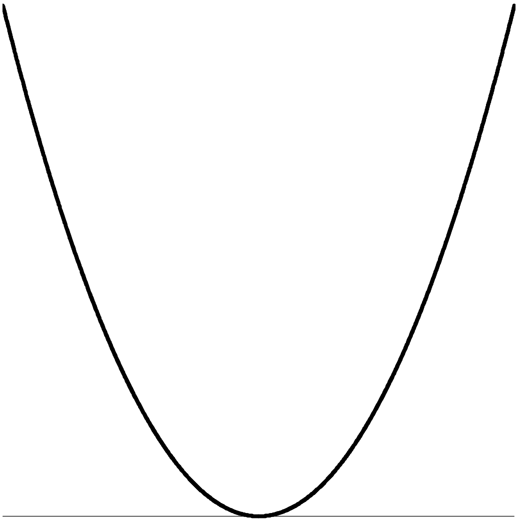

parabola *Jan Wassenaar, www.2dcurves.com* a given point not on the line, known as the focus. The Cartesian equation of a parabola that opens upward and has its vertex (turning point) at the origin is $y = 4ax^2$, where a is the distance from the vertex to the focus, and the quantity $4a$ is known as the *latus rectum.* More generally, any **quadratic** equation of the form $y = ax^2 + bx + c$, where a is not zero, graphs a parabola. The simplest form of this, when $a = 1$, and both b and c are zero, is $y = x^2$. Like the circle, but unlike the ellipse and the hyperbola, the parabola has only one distinct shape. In other words, any parabola can be superimposed exactly on any other parabola simply by rotating, translating (sliding), and/or enlarging or shrinking it.

Euclid dealt with the parabola in his *Conic Sections,* and, although this treatise was lost, it provided a foundation for the first four books by **Apollonius** of the same name. Galileo (1564–1642) discovered that a cannonball, or any other projectile launched at an angle to ground, follows a parabolic path: a result that immediately grabbed the attention of not only scientists but of monarchs and military leaders. René **Descartes**, in writing *La Géométrie* (1637), chose the parabola to illustrate his innovative **analytical geometry**. In 1992, Rudolph Marcus of the California Institute of Technology won the Nobel Prize in Chemistry for his work showing that parabolic reaction surfaces can be used to calculate how fast electrons travel in molecules. His most famous theoretical result, an inverted rate-energy parabola, predicts electron transfer will slow down at very high-reaction free energies.

paraboloid

The surface of revolution of the **parabola**. It is a **quadratic** surface described by the equation $z = a(x^2 + y^2)$.

paradox

> *Please accept my resignation. I don't want to belong to any club that will accept me as a member.*
>
> –Groucho Marx (1895–1977)

A statement that seems to lead to a logical self-contradiction, or to a situation that contradicts common intuition. The word *paradox* comes from the Greek *para* ("beyond") and *doxa* ("opinion" or "belief"). The identification of a paradox based on seemingly simple and reasonable concepts has often led to significant advances in science, philosophy, and mathematics. See **Allais paradox, Arrow paradox, Banach-Tarski paradox, Berry's Paradox, birthday paradox, Burali-Forti paradox, coin paradox, grandfather paradox, Grelling's paradox, liar paradox, Newcomb's paradox, nine rooms paradox, Parrondo's paradox, raven paradox, Russell's paradox, St. Petersburg paradox, Siegel's paradox, unexpected hanging**, and **Zeno's paradoxes**.

parallel

Said of two or more things, such as lines or planes, that are equally distant from one another at all points.

parallel postulate

The fifth and most controversial of **Euclid's postulates** set forth in the Greek geometer's great work, *Elements.* To later mathematicians, the parallel postulate seemed less obvious than the other four and many attempts were made to derive it from them, but without success. In 1823, János **Bólyai** and Nikolai **Lobachevsky** independently realized that entirely self-consistent types of **non-Euclidean geometry** could be created in which the parallel postulate doesn't hold. Carl **Gauss** had made the same discovery earlier but kept the fact secret.

parallelepiped

A **polyhedron** with six faces bounded by three pairs of parallel planes, so that all its faces are **parallelograms**. It is also a **prism** with a parallelogram for a base. A *rectangular parallelepiped* has the shape of a shoebox. The word comes from the Greek *parallelepipedon* for the same shape, which in turn comes from the roots *para* ("beside"), *allel* ("other"), *epi* ("on"), and *pedon* ("ground"). *Parallelepipedon* may have been first used in English by Billingsley in his 1570 translation of Euclid. It seems to have then given way to *parallelepiped* in the last quarter of the nineteenth century.

parallelogram

A **quadrilateral** (four-sided figure) whose opposite sides are parallel, and whose opposite angles, therefore, are equal. The diagonals of a parallelogram bisect each other. A parallelogram of base b and height h has an area

$$bh = ab \sin A = ab\sin B.$$

The height of a parallelogram is $h = a \sin A = a \sin B$. The sides a, b, c, and d and diagonals p and q of a parallelogram satisfy the equality

$$p^2 + q^2 = a^2 + b^2 + c^2 + d^2.$$

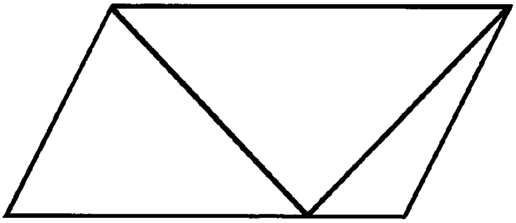

parallelogram The two lines drawn inside this parallelogram have the same length.

Special cases of a parallelogram are a **rhombus**, which has sides of equal length; a **rectangle**, which has two sets of parallel sides that are perpendicular to each other; and a **square**, which meets the conditions of both a rectangle and a rhombus.

In the *parallelogram illusion* shown, despite appearances to the contrary, the interior diagonal lines are of equal length.

parameter

An independent variable; one of the inputs for a **function**. *Parametric equations* are equations in which the variables of interest are given in terms of another variable.

parity

(1) The sum of the digits in a **binary** representation of a number. (2) An indication of whether two numbers are both even or both odd (same parity), or are such that one is even and one is odd (opposite parity). This concept, of same or opposite parity, can be applied to any things that can have two different states–the sense in which a knot is tied, the color of a chessboard square, and so on.

Parrondo's paradox

Two losing gambling games can be set up so that when they are played one after the other, they become winning. This paradox is named after the Spanish physicist Juan Parrondo who, in the late 1990s, discovered how to construct such a scenario. The simplest way is to use three biased coins. Imagine you are standing on stair zero, in the middle of a long staircase with 1001 stairs numbered from –500 to 500. You win if you can get to the top of the staircase, and the way you move depends on the outcome of flipping one of two coins. Heads you move up a stair, tails you move down a stair. In game 1, you use coin *A,* which is slightly biased and comes up heads 49.5% of the time and tails 50.5%. Obviously, these are losing odds. In game 2, you use two coins, *B* and *C.* Coin *B* comes up heads only 9.5% of the time, tails 90.5%. Coin *C* comes up heads 74.5% of the time, tails 25.5%. In game 2 if the number of the stair you are on at the time is a multiple of 3 (that is, . . . , –9, –6, –3, 0, 3, 6, 9, 12, . . .), then you flip coin *B;* otherwise you flip coin *C.* Game 2, it turns out, is also a losing game and would eventually take you to the bottom of the stairs. What Parrondo found, however, is that if you play these two games in succession in random order, keeping your place on the staircase as you switch between games, you will steadily rise to the top of the staircase! This kind of process has been called a *Brownian motion.*[153]

partial differential equation

An equation that involves **derivative**s with respect to more than one variable. Many of the equations used to model the physics of the real world are partial differential equations.

partition number

A number that gives the number of ways of placing *n* indistinguishable balls into *n* indistinguishable urns. For example:

1: (*)

2: (**) (*)(*)

3: (***) (**)(*) (*)(*)(*)

5: (****) (***)(*) (**)(**) (**)(*)(*) (*)(*)(*)(*)

7: (*****) (****)(*) (***)(**) (***)(*)(*) (**)(**)(*)
(**)(*)(*)(*)(*)(*)(*)(*)(*)

11: (******) (*****)(*) (****)(**) (****)(*)(*)
(***)(***) (***)(**)(*)(***)(*)(*)(*)
(**)(**)(**) (**)(**)(*)(*) (**)(*)(*)(*)(*)
(*)(*)(*)(*)(*)(*)

The sequence runs: 1, 2, 3, 5, 7, 11, 15, 22, 30, 42, 56, 77, 101, 135, 176, 231, 297, 385, ... If the urns are distinguishable, the number of ways is 2^n. If the balls are distinguishable, the number of ways is given by the *n*th **Bell number**.

Pascal, Blaise (1623–1662)

A French mathematician, philosopher, and pioneer of **probability theory** whose short life was rich with mathematical invention but whose name, ironically, is most familiar from its association with an array of numbers known as **Pascal's triangle** (which he didn't discover, though he did important work on it). Educated by his father (after whom the **limaçon of Pascal** is named), Pascal showed his intellectual prowess early on by proving one of the most important theorems in **projective geometry** at the age of 16. Three years later he devised the world's second mechanical calculating machine (the first was made by Wilhelm Schickard in 1623) to help with his father's business; he sold about 50 of these "Pascalines," several of which survive. In 1654, he and Pierre de **Fermat**, in an exchange of correspondence, laid the foundation for **probability theory**. They considered the dice problem, already studied by Girolamo **Cardano**, and the problem of points also considered by Cardano and, around the same time, by Luca **Pacioli** and Niccoló **Tartaglia**. The dice problem asks how many times one can expect to throw a pair of dice before getting a double six; the problem of points asks how to divide the stakes if a game of dice is incomplete. Pascal and Fermat solved the problem of points for a two-player game but didn't develop powerful enough methods to solve it for three or more players. In the same year, Pascal was almost killed in an incident in which the horses pulling his carriage bolted and the carriage was left hanging over a bridge above the river Seine. Though rescued unharmed, he shortly after converted to the rigorous Jansenist sect of the Catholic Church. His philosophical work *Pensées,* written between 1656 and 1658 contains his famous argument, often called Pascal's wager, for belief in **God**. Having suffered poor health for most of his adult life, he died in great pain of cancer at the age of 39.

Pascal's mystic hexagon

If a **hexagon** *ADBFCE* (not necessarily convex) is inscribed in a **conic section** (in particular a circle), then the points of intersections of opposite sides (*AD* with *FC, DB* with *CE* and *BF* with *EA*) are collinear. This line is called the *Pascal line* of the hexagon. A special case is when the conic degenerates into two lines; the theorem still holds but is then usually called *Pappus's theorem.*

Pascal's triangle

A triangular pattern of numbers in which each number is equal to the sum of the two numbers immediately above it:

$$\begin{array}{c}
1\\
1\quad 1\\
1\quad 2\quad 1\\
1\quad 3\quad 3\quad 1\\
1\quad 4\quad 6\quad 4\quad 1\\
1\quad 5\quad 10\quad 10\quad 5\quad 1\\
1\quad 6\quad 15\quad 20\quad 15\quad 6\quad 1\\
1\quad 7\quad 21\quad 35\quad 35\quad 21\quad 7\quad 1\\
1\quad 8\quad 28\quad 56\quad 70\quad 56\quad 28\quad 8\quad 1\\
\vdots
\end{array}$$

Although named after Blaise **Pascal**, who studied it, this arithmetic triangle has been recognized since the twelfth century and has a variety of other names. In Italy it is called *Tartaglia's triangle* and in many parts of Asia it is referred to as *Yang Hui's triangle.* Yang Hui was a minor Chinese official who wrote two books, dated 1261 and 1275, that use decimal fractions (long before they appeared in the West) and contain one of the earliest accounts of the triangle. At about the same time, Omar **Khayyam** also wrote about it. The Chinese triangle appears again in 1303 on the front of Chu Shi-Chieh's *Ssu Yuan Yü Chien* (Precious mirror of the four elements), a book in which Chu says the triangle was known in China more than two centuries before his time.

The numbers in Pascal's triangle give the number of ways of picking *r* unordered outcomes from *n* possibilities. This is equivalent to saying that the numbers in each row are the **binomial coefficient**s in the expansion of $(x+y)^n$:

$$\begin{aligned}
(x+y)^0 &= 1\\
(x+y)^1 &= 1x + 1y\\
(x+y)^2 &= 1x^2 + 2xy + 1y^2\\
(x+y)^3 &= 1x^3 + 3x^2y + 3xy^2 + 1y^3\\
(x+y)^4 &= 1x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + 1y^4
\end{aligned}$$

and so on. In addition, the shallow diagonals of the triangle sum to give the **Fibonacci sequence**.

path

See **trajectory**.

pathological curve

A curve often specifically devised to show the falseness of certain intuitive concepts. In particular, the image of **continuity** as a smooth curve in our mind's eye severely misrepresents the situation and is the result of a bias stemming from an overexposure to the much smaller class of differentiable functions (see **differentiation**). A chief lesson of pathological curves is that continuity is a weaker notion than differentiability. Many pathological curves are **fractal**s, such as **Cantor dust**, including **space-filling curve**s, such as the **Peano curve**. The earliest known example is the graph of **Weierstrass' nondifferentiable function**.

payoff

In **game theory**, the amount that a player wins, given the player's and his opponent's actions.

Peano, Guiseppe (1858–1932)

An Italian pioneer of mathematical **logic** and the axiomatization of mathematics. In 1889, he published his first version of a system of mathematical logic in *Arithmetics Principia,* which included his famous **axiom**s of natural numbers, now known as *Peano's axioms.* Two years later, he established a journal, the *Rivista,* in which he proposed the symbolizing of all mathematical propositions into his system. The project, which became known as the *Formulario,* was his chief focus for the next fifteen years.

Peano curve

The first known example of a **space-filling curve**. Discovered by Guiseppe **Peano** in 1890, its effect was like that of an earthquake on the traditional structure of mathematics. Commenting in 1965 on the impact of the curve in Peano's day, N. Ya Vilenkin said: "Everything has come unstrung! It's difficult to put into words the effect that

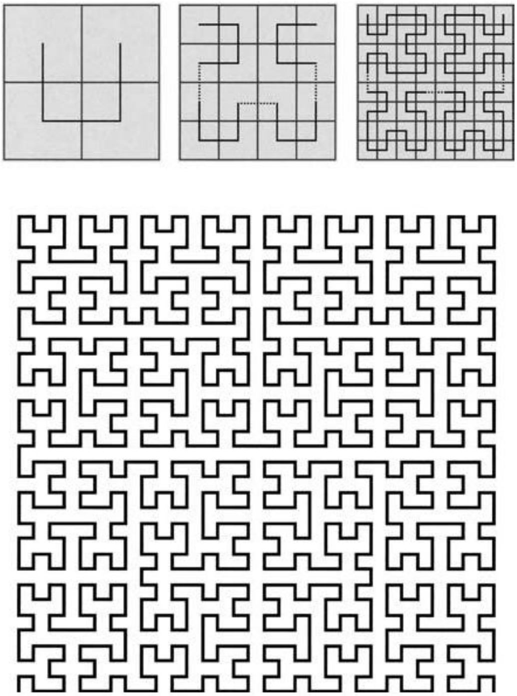

Peano curve A simple iterative process produces a remarkable space-filling shape.

Peano's result had on the mathematical world. It seemed that everything was in ruins, that all the basic mathematical concepts had lost their meaning." Today, the Peano curve is recognized as just one of an infinite class of familiar objects known as **fractal**s. But at the end of the nineteenth century it was an extravagant, completely counterintuitive thing; indeed, it was something that had been believed impossible. Writing of Peano's result in *Grundzüge der Mengenlehre* (Basic features of set theory) in 1914, Felix **Hausdorff** said: "This is one of the most remarkable facts of set theory." Originally, the Peano curve was derived purely analytically, without any kind of drawing or attempt at visualization. But the first few steps in drawing it, as shown in the diagrams, are easy enough, even though the finished product is unattainable in this way–and totally unimaginable. To fill the unit square, as the Peano curve does, without leaving any holes, the curve has to be both continuous and self-intersecting.

pearls of Sluze

Curves that are generated by the Cartesian equation

$$y^n = k(a - x)^p x^m,$$

where *n, m,* and *p* are integers. They were first studied by the French mathematician René de Sluze (1622–1685) and named the pearls of Sluze by Blaise **Pascal**.

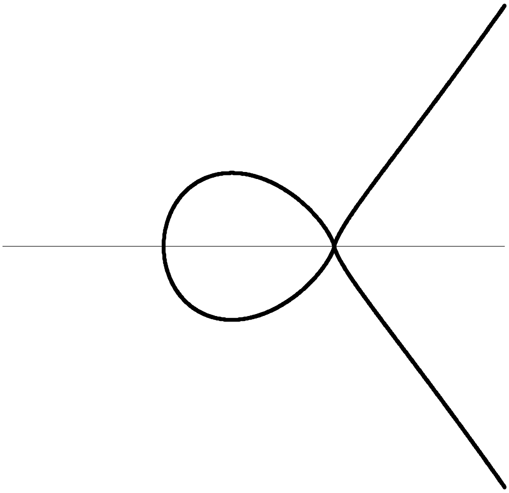

pearls of Sluze *Jan Wassenaar, www.2dcurves.com*

pedal curve

The locus of the feet of the perpendiculars from a given point to the **tangent**s to the given figure. The "given point" is known as the *pedal point.*

pedal triangle

The pedal triangle of a point *P* with respect to a triangle *ABC* is the triangle whose vertices (see **vertex**) are the feet of the perpendiculars dropped from *P* to the sides of triangle *ABC.*

peg solitaire

Also known as Hi-Q, a game that, in its commonest form, is played with 32 pegs or marbles on a rectangular grid, the middle position of which starts out empty. A peg may jump horizontally or vertically, but not diagonally, over a peg in an adjacent hole into a vacant hole immediately beyond. The peg that was jumped over is then removed. The object is to be left with a single peg in the center position.

The quickest solution of peg solitaire was found by Ernest Bergholt in 1912 and was proved to be minimal by John Beasley in 1964. Suppose the rows and columns of the board are each labeled 1 to 7 so that, for example, a peg on the fourth row and the third column is in position 43. The quickest solution is: (1) 46→44. (2) 65→45.

5 LEFT - Good
4 LEFT - Better
3 LEFT - Really Clever
2 LEFT - A Pat on the back
1 LEFT - Take a deep bow
PERFECT SCORE
1 LEFT in centre hole
RULES

peg solitaire A version of peg solitaire made of tin with marbles by the British firm MAR Toys in the 1950s. *Sue & Brian Young/Mr. Puzzle Australia, www.mrpuzzle.com.au*

(3) 57→55. (4) 54→56. (5) 52→54. (6) 73→53. (7) 43→63. (8) 75→73→53. (9) 35→55. (10) 15→35. (11) 23→43→63→65→45→25. (12) 37→57→55→53. (13) 31→33. (14) 34→32. (15) 51→31→33. (16) 13→15→35. (17) 36→34→32→52→54→34. (18) 24→44.[28]

Pell equation

An equation of the form $y^2 = ax^2 + 1$, where a is any positive whole number except a square number. The name comes from the English mathematician John Pell (1611–1685); however, he was wrongly credited. In writing about some of the work done on this type of equation, Leonhard **Euler** gave priority to Pell whereas, in fact, Pell had done no more than copy it in his papers from some of Pierre de **Fermat**'s letters. Fermat had been the first to state that an equation of this form always has an unlimited number of integer solutions. For example, the equation $y^2 = 92x^2 + 1$, has the solutions $x = 0, y = 1$; $x = 120, y = 1{,}151$; $x = 276{,}240, y = 2{,}649{,}601$; and so on. Each successive solution is about 2,300 times the previous solution. In fact, the solutions are every eighth partial fraction (where x is the numerator and y is the denominator) of the **continued fraction** for $\sqrt{92}$. A Pell equation was used in finding the solution to **Archimedes's cattle problem**.

Pell numbers

Numbers, named after John Pell (1611–1685), that are similar to the **Fibonacci sequence** and are generated by the formula $A_n = 2\,A_{n-1} + A_{n-2}$. The sequence runs: 1; 2; 5; 12; 29; 70; 169; 408; 985; 2,378; 5,741; 13,860; 33,461; 80,782; 195,025; . . . . The ratio of successive terms approaches 1 plus the **square root of 2**.

Penrose, Roger (1931–)

An English mathematical physicist famous for his important contributions to cosmology and the physics of black holes, for his controversial views on the nature of human consciousness and its relationship to **quantum mechanics**, and for his work in the field of recreational mathematics. The **Penrose tiling** and the **Penrose triangle** are named after him (but not the **Penrose stairway**, which is named after his father). In his book *The Emperor's New Mind* he argues that there must be errors in the known laws of physics, notably in quantum mechanics, and that true **artificial intelligence** is impossible.[242] The latter claim is based on his assertion that humans can do things outside the power of formal logic systems, such as knowing the truth of unprovable statements, or solving the **halting problem** (claims that were originally made by the philosopher John Lucas of Merton College, Oxford). These are controversial views, with which most of the mathematical and computer science communities disagree.

Penrose stairway

An **impossible figure** named after the British geneticist Lionel Penrose (1898–1972), father of Roger **Penrose**. It served as an inspiration for the staircase in M. C. **Escher**'s famous print "Ascending and Descending." Although the Penrose stairway cannot be realized in

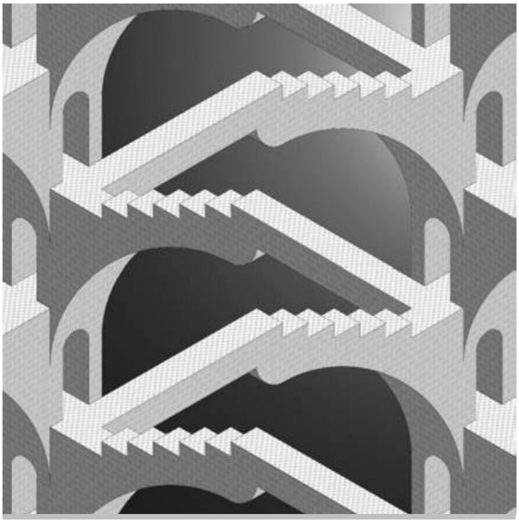

Penrose stairway Steps that ascend and descend simultaneously. *Jos Leys, www.josleys.com*

three dimensions, this impossibility is not immediately perceived and, in fact, the paradox is not even apparent to many people at a quick glance. Although Escher and the Penroses made the Stairway famous, it was, unbeknownst to them, independently discovered and refined years before by the Swedish artist Oscar **Reutersvärd**. In the 1960s the Stanford psychologist Roger Shepard created an auditory analogue of the Stairway. See also **Penrose triangle**.

Penrose tiling

A kind of **aperiodic tiling** discovered by Roger **Penrose**. In 1973 he announced a tiling made from a set of six tiles and then, by slicing and re-gluing, was able to reduce the number of tiles to just two. The most elegant of Penrose tilings use two rhombi (see **rhombus**), a thick one called a "kite" and a thin one called a "dart," which are fitted together so that no two tiles are aligned to form a single parallelogram (otherwise, a single rhombus could be used to make a periodic tiling). All angles are multiples of $\pi/5$ radians (36°, or one-tenth of a circle). Each tile has four sides with a length of one unit. One tile has four corners with the angles 72°, 72°, 108°, and 108° (2, 2, 3, and 3 multiples of 36°); the other has angles of 36°, 36°, 144°, and 144° (1, 1, 4, and 4 multiples of 36°). On each tile one of the vertices (corners) is colored black and two of the sides are marked with arrows. The only rules for assembling the tiles to ensure an aperiodic tiling are that two adjacent vertices must be of the same color, and two adjacent edges must have arrows pointing in the same direction or have no arrows at all. These rules ensure that, taken over a large enough area of the plane, the pattern of tiles doesn't repeat. In a correct Penrose tiling, the ratio of kites to darts is always the same and is equal to the **golden ratio**. Although Penrose tilings started out as nothing more than an interesting mathematical diver-

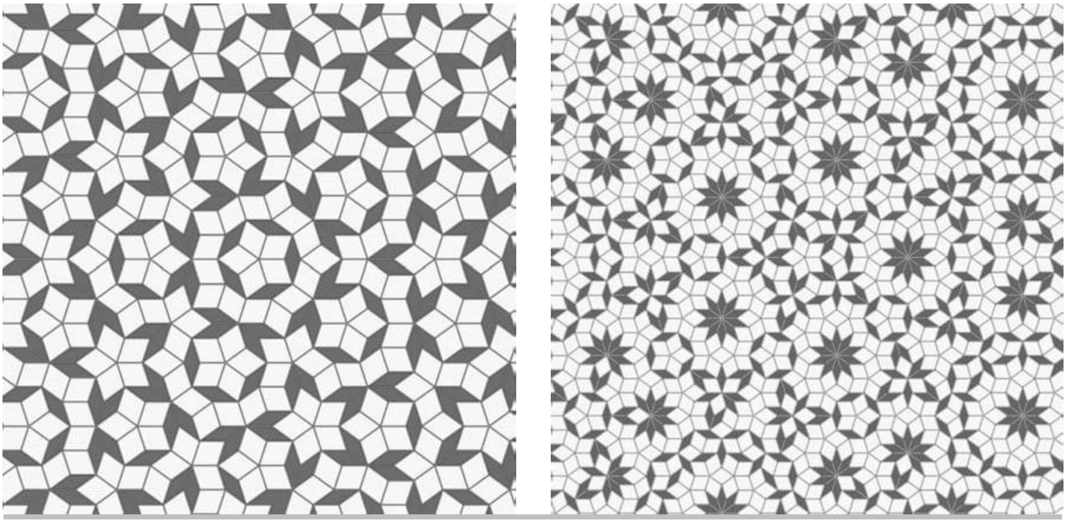

Penrose tiling Two of an infinite variety of nonperiodic patterns possible using Penrose tiles. *Jos Leys, www.josleys.com*

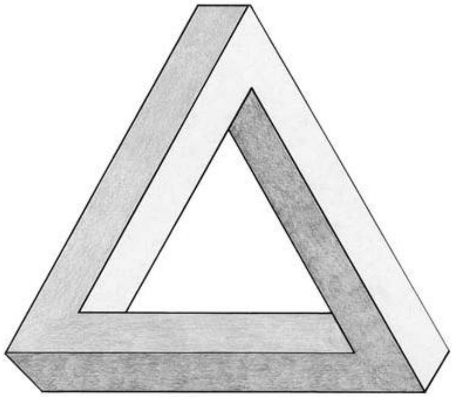

Penrose triangle

sion, they have turned out to mimic the arrangements of atoms in some newly discovered materials, known as **quasicrystal**s.[127]

Penrose triangle

The most famous and one of the simplest **impossible figure**s. Its roots go back to 1934 when Oscar **Reutersvärd** made the first recognizable impossible triangle out of a strange two-dimensional representation of cubes; this artwork appeared on a Swedish postage stamp issued in 1982. In 1954, Roger **Penrose**, after attending a lecture by the artist M. C. **Escher**, rediscovered the impossible triangle and drew it in its most familiar form, which he published in a 1958 article in the *British Journal of Psychology,* coauthored with his father Lionel.[241] Penrose was unfamiliar with the work of Reutersvärd, Giovanni Piranesi, and others who had previously created impossible figures. Penrose's impossible triangle, unlike Reutersvärd's earlier version, was drawn in perspective, which added an additional size paradox to the object. In 1961, Escher, inspired by Penrose's version of the impossible triangle (he was sent a copy of the article by the Penroses), incorporated it into his famous lithograph *Waterfall.*

penta-

The Greek prefix meaning "five." A *pentagon* is a five-sided **polygon** and a *pentahedron,* such as the Great Pyramid of Gisa, is a five-sided **polyhedron**. A *pentagram,* or *pentacle,* is a five-pointed star formed by the pairwise extension of the sides of a regular polygon so that each side meets not with the next side but with the one after that. The Pythagoreans (see **Pythagoras of Samos**) used the pentagram as a secret identification emblem; later it became a trademark of alchemists and, perhaps because of its repeating properties, a sign of the occult. According to legend, a pentagram was used by Doctor Faustus to exorcise Mephistopheles. *Pentagonal numbers* are **figurate number**s (numbers that can be represented by a regular geometric arrangement of equally spaced points) of the form $n(3n - 1)/2$; the first few are 1, 5, 12, 22, ...

pentomino

A five-square **polyomino**.

Perelman, Yakov Isidorovitch (1882–1942)

A Russian scientist and exponent of recreational mathematics whose stature in his own country was, and remains, similar to that of Martin **Gardner** in the United States. Between 1913 and his death, Perelman authored a dozen books and scores of articles covering many different aspects of popular mathematics, physics, and astronomy. His books have been republished so often in Russia that Perelman is universally known among both amateur and professional mathematicians across several generations. He was also a leading proponent of the ideas of the spaceflight visionary Konstantin Tsiolkovsky.

perfect cube

An integer of the form n^3 where n is an integer.

perfect cuboid problem

See **cuboid**.

perfect number

A whole number that is equal to the sum of all its factors except itself. For example, 6 is a perfect number because its factors, 1, 2, and 3 add to give 6. The next smallest is 28 (the sum of $1 + 2 + 4 + 7 + 14$). Augustine (354–430) argued that God could have created the world in an instant but chose to do it in a perfect number of days, 6. Early Jewish commentators saw the perfection of the universe in the Moon's period of 28 days. The next in line are 496, 8,128, and 33,550,336. As René **Descartes** pointed out: "Perfect numbers like perfect men are very rare." All end in six or eight, though what seems to be an alternating pattern of sixes and eights for the first few perfect numbers doesn't continue. All are of the form $2^{n-1}(2^n - 1)$, where $2^n - 1$ is a **Mersenne prime**, so that the search for perfect numbers is the search for Mersenne primes. The largest one found is $2^{3021376}(2^{3021377} - 1)$. It isn't known if there are infinitely many perfect numbers or if there are any odd perfect numbers. A *pseudoperfect number* or *semi-perfect number* is a number equal to the sum of some of its divisors, for example, $12 = 2 + 4 + 6$, $20 = 1 + 4 + 5 + 10$. An *irreducible semi-perfect number* is a semi-perfect number, none of whose factors is semi-perfect, for example, 104. A *quasi-perfect number* would be a number n whose divisors (excluding itself) sum to $n + 1$, but it isn't known if such a number exists. A *multiply perfect number* is a number n whose divisors sum to a multiple of n. An example is 120, whose divisors (including itself) sum to $360 = 3 \times 120$. If the divisors sum to $3n$, n is called multiply perfect of order 3, or *tri-perfect.* Ordinary perfect numbers are multiply perfect of order 2. Multiply perfect numbers are known of order up to 8. See also **abundant number**.

perfect power
An integer of the form m^n where m and n are integers and $n > 1$.

perfect square
A number that is the product of two equal whole numbers; for example, $1 = 1 \times 1$, $4 = 2 \times 2$, $9 = 3 \times 3$, $16 = 4 \times 4$.

Perigal, Henry (1801–1898)
An English amateur mathematician best known for his elegant **dissection** proof of **Pythagoras's theorem**, a diagram of which is carved on his gravestone. Perigal also discovered a number of other interesting geometrical dissections and, though employed modestly for much of his life as a stockbroker's clerk, was well known in British scientific society.

perimeter
The distance around a two-dimensional shape.

period, of a decimal expansion
The length of the smallest block of repeating digits in the decimal expansion of a **rational number** that does not terminate. For example:

$1/3 = 0.333333333333\ldots$
repeating block = 3, period = 1
$5/7 = 0.71428571428571\ldots$
repeating block = 714285, period = 6
$89/26 = 3.4230769230769\ldots$
repeating block = 230769, period = 6

periodic
Refers to motion or to an entity that goes through a finite number of regions, returns to a previous state, and repeats the same fixed pattern forever.

periodic attractor
Also called a *limit cycle attractor,* an **attractor** that consists of a periodic movement back and forth between two or more values. The periodic attractor represents more possibilities for system behavior than does the **fixed-point attractor**. An example of a period two attractor is the oscillating movement of a metronome or, in psychiatry, a bipolar disorder that causes a person's mood to shift back and forth from elation to depression.

periodic tiling
A **tiling** in which a region can be outlined that tiles the plane by translation, that is, by shifting the position of the region without rotating or reflecting it. M. C. **Escher** is famous for his many pictures of periodic tilings with shapes that resemble living things. An infinity of shapes—for instance, the regular hexagon—tile only periodically, though all these fall into 17 distinct **wallpaper groups**. An infinity of other shapes tile both periodically and aperiodically. But it was only quite recently that the first **aperiodic tilings** were discovered.

periphery
A curved **perimeter**.

permutable prime
Also known as an *absolute prime,* a **prime number** with at least two distinct digits, which remains prime on every rearrangement (permutation) of the digits. For example, 337 is a permutable prime because 337, 373, and 733 are all prime. Most likely, in base 10, the only permutable primes are 13, 17, 37, 79, 113, 199, 337, and their permutations. Permutable primes can't have any of the digits 2, 4, 6, 8 or 5, nor can they have all four of the digits 1, 3, 7, and 9 simultaneously.

permutation
A particular ordering of a collection of objects. For example, if an athlete has won three medals, a bronze one (B), a silver one (S), and a gold one (G), there are six ways they can be permuted or lined up: BSG, BGS, SBG, SGB, GBS, and GSB. If six people want to sit on the same park bench, there are 720 ways in which they can organize themselves. In general, n things can be permuted in $n \times (n - 1) \times (n - 2) \times \ldots \times 2 \times 1 = n!$ ways (where "!" is the symbol for **factorial**). How about if there are n distinct objects but we want to permute them in groups of k (where $k \leq n$): how many ways can that be done? The first member of the group can be picked in n ways because there are n objects to pick from. The second member can be filled in $(n - 1)$ ways since one of the n elements has already been taken. The third member can be filled in $(n - 2)$ ways since 2 elements have already been used, and so on. This pattern continues until k things have been chosen. This means that the last member can be filled in $(n - k + 1)$ ways. Therefore a total of $n(n - 1)(n - 2)\ldots(n - k + 1)$ different permutations of k

objects, taken from a pool of *n* objects, exist. If we denote this number by P(*n, k*), we can write $P(n, k) = n! / (n - k)!$

perpendicular

At right angles. Two lines, planes, etc., are said to be perpendicular if they are 90° apart.

Perrin sequence

The sequence of integers defined by the recurrence relation $P(n) = P(n - 2) + P(n - 3)$, with the initial conditions $P(0) = 3$, $P(1) = 0$, $P(2) = 2$. Although the sequence is named after R. Perrin who studied it in 1899, it had been explored earlier, in 1876, by Édouard **Lucas**. As in the case of the similar *Padovan sequence,* the ratio of consecutive Perrin numbers tends toward the **plastic number**. More importantly, it appears that *n* divides $P(n)$ exactly if and only if *n* is a **prime number**. For example, 19 is prime, $P(19) = 209$ and $^{209}/_{19} = 11$; on the other hand, 18 is composite, $P(18) = 158$ and $^{158}/_{18} = 8.777$, which is not a whole number. Lucas conjectured that this is true for all values of *n* so that the Perrin sequence can be used as a test for non-primality: any number *n* that does not divide $P(n)$ is composite. Whether this conjecture is correct remains an open question. No one has ever found a composite *n* that divides $A(n)$, nor has anyone been able to prove that such numbers, known as *Perrin pseudoprimes,* don't exist. In 1991 Steven Arno of the Supercomputing Research Center in Bowie, Maryland, proved that Perrin pseudoprimes must have at least 15 digits.[15] The conjecture that no Perrin pseudoprimes exist is important, because the remainder on dividing $P(n)$ by *n* can be calculated very rapidly.

perturbation

A slight nudge that, temporarily or permanently, displaces an object or system out of **equilibrium**.

peta-

Prefix for 10^{15}, from the Greek *pentakis,* meaning "five times."

phase space

The mathematical space of all possibilities in a given situation. A motion is then described by a *path, trajectory,* or *orbit* in this space. This is not the usual kind of path laid out on the ground, but a series of locations in phase space, describing motion or change over a period of time. The terms do, however, recall the origins of qualitative dynamics in Henri **Poincaré**'s study of planetary motion. The *dimension* of the phase space is the number of initial conditions needed to uniquely specify a path and is equal to the number of variables in the dynamical system. The temporal behavior of the system is viewed as the succession of states in the system's state space. In the case of a simple pendulum, for example, the instantaneous configuration is given by just two numbers–the position of the pendulum bob and its velocity–which completely describe the system's state. For more complex systems, such as a chain of *n* pendulums coupled together, the state of the system is much larger. It requires, in this case, 2*n* numbers to specify the state of the entire system. This collection of all possible configurations is the phase space.

phase transition

In physics, a change from one state of matter to another. In **dynamical system**s theory, a change from one mode of behavior to another.

phi

See **golden ratio**.

Phillips, Hubert (1891–1964)

An English compiler of crosswords and word puzzles who wrote under the pseudonyms "Dogberry" in the *News Chronicle* magazine and "Caliban" in the *New Statesman.* He was also a prolific writer of epigrams, parodies, and satirical verse, and appeared on many radio quizzes. Phillips earned a first class degree in history at Oxford, served in the army throughout World War I, taught economics at Bristol University, and was active in the British Liberal Party. He was an accomplished player of contract bridge, captaining England in 1937 and 1938. Among his many publications are *Caliban's Problem Book, The Complete Book of Card Games, Brush Up Your Wits, My Best Puzzles in Mathematics,* over a hundred crime-problem stories, and a novel, *Charteris Royal.*[249, 250, 251]

Phutball

Also known as *philosopher's football,* a two-player board **game** first described by Elwyn Berlekamp, John **Conway**, and Richard **Guy**, in their book *Winning Ways for your Mathematical Plays.* Phutball is played on the intersections of a 19 × 15 grid using one white piece and as many black pieces as necessary. The objective is to score goals by using the men (black pieces) to move the football (white piece) onto or over the opponent's goal line, that is, either rows 1 or 0 or rows 19 and 20 (rows 0 and 20 being off the board). At the start of the game the football is placed on the central point, although a handicapping scheme exists where the ball can start nearer to the stronger player's goal line. Players alternate making moves, which consist either of adding a man to any vacant point on the board or of moving the ball. There is no difference between men played by the two opponents. The football is moved by a series of jumps over adjacent men. One jump is from the football's current point to the first vacant point in a straight line orthogonally or diagonally over one or more men. The jumped men are then removed from the board (before any subsequent jump occurs). Jumping is optional and can be repeated for as long as there are men available to be jumped and the player desires to do so. In contrast to checkers, multiple men in a row are jumped and removed as a group. If the football ends the move on or over the opponent's goal line then a goal has been scored. If the football passes through a goal line, but ends up elsewhere due to further jumps, the game continues. The game is sufficiently complex that, although theoretically one of the players has a winning strategy, this strategy is not known.

pi (π)

The ratio of the circumference to the diameter of a **circle**. Pi, sometimes known as *Archimedes' constant,* is one of the most important and ubiquitous numbers in mathematics, popping up in all kinds of seemingly unrelated areas. Its approximate value is 3.1419 but, being an **irrational number**, it can't be written as a terminating or recurring decimal. The fact that π can't be expressed as the ratio of two integers was proved by Johann Lambert in 1761. Then, in 1881, Ferdinand Lindemann showed that it is a **transcendental number**, meaning that there is no polynomial with integer or rational coefficients of which π is a root. Because of this, it's impossible to write π in terms of *any* finite number of integers or fractions or their roots. From the outset, this dooms any attempt at **squaring the circle**, that is, constructing a square whose area is equal to the area of a given circle, using straightedge and compass alone.

The association of the Greek letter π with the number 3.141 . . . started with William **Oughtred** in 1647 who used π.δ as a stand-in for "periphery-diameter" in his *Clavis mathematicae.* The Welsh mathematician William Jones was the first to use π as a symbol on its own in his 1706 *Synopsis palmariorum matheseos.* But it fell to the great Leonhard **Euler** to popularize the notation 30 years later.

Calculating π with greater and greater precision became a popular pursuit. Around 1600, Ludolph van Ceulen computed the first 35 decimals and was so proud of his accomplishment that he had the number inscribed on his tombstone. In 1789, the Slovene mathematician Jurij Vega extended known π to 140 places (of which only 137 were correct), while, in 1873, William Shanks pushed back the frontier to 707 places–a record that stood until 1949 when the first electronic computers appeared on the scene. In September 2002, Yasumasa Kanada and his colleagues at Tokyo University used 400 hours of supercomputer time to calculate π to 1.24 trillion places, beating their previous best, set in 1999, of 206 billion places. Although such an exercise may seem pointless, it serves as a benchmark for new high-speed computers and algorithms, and also to test the long-standing, but still unproven assertion that the distribution of digits in π is completely random.

For the record, the first 100 digits of π are: 3.1415926535 8979323846 2643383279 5028841971 6939937510 5820974944 5923078164 0628620899 8628034825 3421170679. The wonderfully pointless pastime of "piphilology" is concerned with devising mnemonics for remembering some of these: the number of letters in each word of the mnemonic gives the corresponding digit. A famous example is Isaac Asimov's "How I want a drink, alcoholic of course, after the heavy lectures involving quantum mechanics!" For the more ambitious, there is Michael Keith's 1995 poem that supplies the first 42 digits:

Poe, E. Near A Raven
Midnights so dreary, tired and weary,
Silently pondering volumes extolling all by-now obsolete lore.
During my rather long nap, the weirdest tap!
An ominous vibrating sound disturbing my chamber's antedoor.
"This," I whispered quietly, "I ignore."

There are some reasonably good approximations to π in the form of ordinary (rational) fractions. The best known is $^{22}/_{7}$, but this is only accurate to two decimal places. A fraction with a larger denominator offers a better chance of getting a more refined estimate. There is also $^{333}/_{106}$, which is good to 5 places. But an outstanding approximation is $^{355}/_{113}$ which is accurate to 6 places; in fact, there is no better approximation among all fractions with denominators less than 30,000. See also **continued fraction**.

In 1897 the General Assembly of the state of Indiana tried to pass legislation to the effect that the exact value of π is 3.2. The bill was referred, for some bizarre reason, to the House Committee on Canals. But then the true motive for the attempted change to the law became clear. By chance a professor of mathematics happened to be present during a debate and heard an ex-teacher saying "The case is perfectly simple. If we pass this bill, which establishes a new and correct value for π, the author offers to our state without cost the use of his discovery and its free publication in our school textbooks, while everyone else must pay him a royalty." Fortunately, the professor was able to teach the senators a little math and the bill was stopped in its tracks.

Any formulas to do with circles or spheres, not surprisingly, have π in them. For example, the circumference of a circle of radius r is $2\pi r$, the area of a circle is πr^2, the volume of a sphere is $(4/3)\pi r^3$, and the surface area of a sphere is $4\pi r^2$. But π also has a strange habit of appearing in the most unexpected places. In the eighteenth century, the French naturalist Comte de Buffon showed that you could estimate π by experiment using the unlikely apparatus of a repeatedly dropped needle and a shove-halfpenny board (see **Buffon's needle**). Two of the most important, but (as far as is known) unrelated, equations in modern physics, Heisenberg's uncertainty principle ($\Delta x \ \Delta p = h/4\pi$) in **quantum mechanics** and Einstein's field equation ($R_{ik} - \frac{1}{2}\, g_{ik}\, R + \Lambda\, g_{ik} = 8\pi G/c^4\ T_{ik}$) in general **relativity theory** include π. The ever-present constant also emerges as a result of various remarkable **infinite series**. These include:

$1/1^2 + 1/2^2 + 1/3^2 + 1/4^2 + \ldots = \pi^2/6$
(found by Leonhard **Euler**)
$1/1 - 1/3 + 1/5 - 1/7 + 1/9 - \ldots = \pi/4$
(found by Gottfried **Leibniz**)
$2/1 \times 2/3 \times 4/3 \times 4/5 \times 6/5 \times 6/7 \times 8/7 \times 8/9 \times \ldots = \pi/2$
(found by John **Wallis**)

Among many other formulas that give rise to or contain π are

the integral

$$\int_{-\infty}^{\infty} e^{-x^2}\, \mathrm{d}x = \sqrt{\pi},$$

Stirling's formula, $n! \sim \sqrt{2\ \pi\ n}\ (n/e)^n$ and, the most intriguing equation in mathematics, $e^{i\pi} + 1 = 0$.

In **number theory**, the probability that two randomly chosen integers have no common divisors (in other words, that they're *relatively prime*) is $6/\pi^2$ or 1 in 1.644934 . . . , and the average number of ways to write a positive integer as the sum of two **perfect square**s is $\pi/4$. Both of these facts are astonishing because it is hard to see how a basically geometric constant to do with circles has any bearing at all on how various types of numbers are distributed. Deep truths are buried here!

In 1995, David Bailey, Peter Borwein, and Simon Plouffe of the University of Quebec at Montreal discovered a new formula for π as an infinite series:[21]

$$\pi = \sum_{k=0}^{8} \frac{1}{16^k} \left[\frac{4}{8k+1} - \frac{2}{8k+4} - \frac{1}{8k+5} - \frac{1}{8k+6} \right]$$

What is so remarkable about this is that it enables the calculation of *isolated digits* of π–say, the trillionth digit–without computing and keeping track of all the preceding digits. How such a formula could possibly arise constitutes a mystery in itself. The only catch is that the formula works for base 2 (**binary**) and 16 (**hexadecimal**) but not base 10. So, it's possible to use the formula to determine, for example, that the five-trillionth binary digit of π is 0, but there's no way to convert the result into its decimal equivalent without knowing all the binary digits that come before the one of interest. The new formula allows the calculation of the *n*th base 2 or base 16 digit of π in a time that is essentially linear in n, with memory requirements that grow logarithmically (very slowly) in n. One possible use of the Bailey-Borwein-Plouffe formula is to help shed light on whether the distribution of π's digits are truly random, as most mathematicians suppose.[29, 42, 68, 114] See also **pyramid**.

Pick's theorem

First published in 1899, a theorem that was brought to broad attention as recently as 1969 through Hugo **Steinhaus**'s popular book *Mathematical Snapshots*.[316] Pick's theorem gives an elegant formula for the area of *lattice polygons*–**polygon**s that have vertices located at the integral nodes of a square grid or lattice that are spaced a unit distance from their immediate neighbors. Pick's theorem says that the area of such a polygon can be found simply by counting the lattice points on the interior and boundary of the polygon. The area is given by

$$i + (b/2) - 1,$$

where i is the number of interior lattice points and b is the number of boundary lattice points. The Austrian mathematician Georg Pick (1859–1942) after whom the result is named, was born in Vienna and perished during World War II in the Theresienstadt concentration camp. Over the past few decades, beginning with a paper by J. E. Reeve in 1957, various generalizations of Pick's theorem have been made to more general polygons, to higher-dimensional polyhedra, and to lattices other than square lattices. Most recently, mathematicians have become interested in the theorem because it provides a link between traditional Euclidean geometry and the modern subject of digital (discrete) geometry.

pico-

Prefix meaning a trillionth (10^{-12}), from the Italian *piccolo,* meaning "small."

pitch drop experiment

The world's longest running experiment. It started in 1927 when Thomas Parnell, the first professor of physics at the University of Queensland in Brisbane, Australia, heated a sample of pitch (a derivative of tar) and poured it into a glass funnel with a sealed stem. Three years were allowed for the pitch to settle, then, in 1930, the sealed stem was cut. From that date on the pitch has slowly dripped out of the funnel–so slowly that, up to the present time, only eight drops have fallen! The experiment

pitch drop experiment Professor John Mainstone alongside the world's longest-running experiment. *John Mainstone, University of Queensland*

stands in a display cabinet in the foyer of the Department of Physics at the University of Queensland demonstrating for all to see the fact that pitch, though it feels like a solid and is brittle enough to smash with a hammer, is really a fluid of very high viscosity (about 100 billion times as viscous as water).[91]

pizza

Order a pizza, pick an arbitrary point in it, and cut the pizza into eight slices by slicing at 45° angles through the chosen point. Color the alternate pieces red, with ketchup, and yellow, with mustard. Measure the area of the red slices and the yellow slices. Surprisingly, it will be found that these areas are the same! This fact holds true for any multiple of four slices cut by using equal angles through a fixed arbitrary point in the pizza. See also **tautology**.

place-value system

A **number system** in which the value of a number symbol depends not only on the symbol itself but also on the position where it occurs.

Planck time

In **quantum mechanics**, the shortest meaningful period of **time**; any two events that are separated by less than this amount of time can be considered simultaneous. It has the value 5.390×10^{-44} second. Related to this is the *Planck length* of 6.160×10^{-35} meters, which is the distance that light can travel in the Planck time.

plane

A flat surface such that a straight **line** joining any two points on it will also lie entirely on the surface.

plane partition

A stack of unit cubes in a rectangular box or in the positive **octant** in space such that to the left, behind, and below every cube lies either another cube or a wall. A plane partition in a box is equivalent to a **lozenge** tiling of a hexagon in the plane.

plastic number

A little-known number that has much in common with the **golden ratio** in that it is closely linked to architecture and to aesthetics. The concept of the plastic number was first described by the Dutchman Hans van der Laan (1904–1991) in 1928, shortly after he had abandoned his

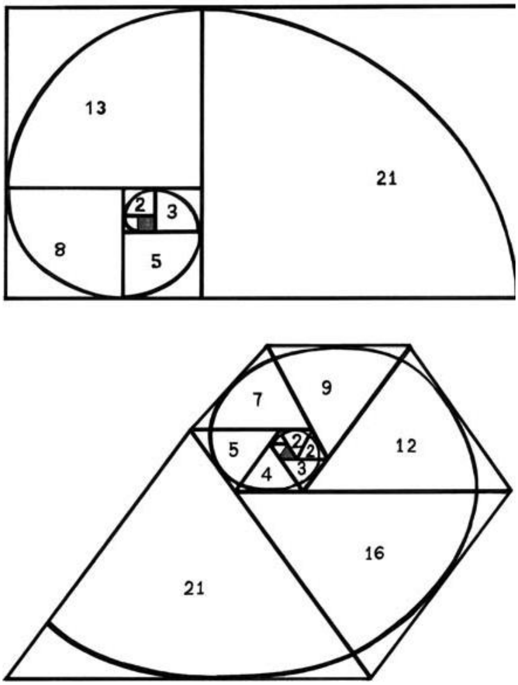

plastic number Spiraling systems illustrate the Fibonacci numbers (top) and the Padovan sequence.

architectural studies and become a novice monk, and has subsequently been explored by the English architect Richard Padovan (1935–). It is derived from a **cubic** equation, rather than a **quadratic** in the case of the golden ratio, and is intimately linked to two ratios, approximately 3:4 and 1:7, which van der Laan considered fundamental in the relationship between human perception and shape and form. These ratios, he believed, express the lower and upper limits of our normal ability to perceive differences of size among three-dimensional objects. The lower limit is that at which things differ just enough to be of distinct types of size. The upper limit is that beyond which they differ too much to relate to each other; they then belong to different orders of size. According to van der Laan, these limits are precisely definable. The mutual proportion of three-dimensional things first becomes perceptible when the largest dimension of one thing equals the sum of the two smaller dimensions of the other. This initial proportion determines, in turn, the limit beyond which things cease to have any perceptible mutual relation.

In mathematical terms, the plastic number is the unique **real number** solution to the equation $x^3 - x - 1 = 0$ and has the approximate value 1.324718 . . . . Just as the golden ratio is approximated better and better by successive terms of the **Fibonacci sequence**, $F(n + 1) = F(n) + F(n - 1)$, where $F(0) = F(1) = 1$, so the plastic number arises as the limit of the ratio of successive numbers in the sequence $P(n + 1) = P(n - 1) + P(n - 2)$ where $P(0) = P(1) = P(2) = 1$. This sequence has been called the *Padovan sequence,* and its members the *Padovan numbers.* The Padovan sequence increases much more slowly than does the Fibonacci sequence. Some numbers, such as 3, 5, and 21, are common to both sequences; however, it is not known if there are finitely many or infinitely many such pairs. Some Padovan numbers, such as 9, 16, and 49, are perfect squares; the square roots of these (3, 4, and 7) are also Padovan numbers, but it is not known if this is a coincidence or a general rule. Another way to generate the Padovan numbers is to mimic the use of squares for Fibonacci numbers, but with cuboid structures–boxes with rectangular faces. A kind of three-dimensional spiral of boxes emerges. Start with a cube of side 1 and place another adjacent to it: the result is a $1 \times 1 \times 2$ cuboid. On the 1×2 face, add another $1 \times 1 \times 2$ box to produce a $1 \times 2 \times 2$ cuboid. Then on a 2×2 face, add a $2 \times 2 \times 2$ cube to form a $2 \times 2 \times 3$ cuboid overall. To a 2×3 face, add a $2 \times 2 \times 3$ to get a $2 \times 3 \times 4$ box overall, and so on. Continue this process, always adding cuboids in the sequence east, south, down, west, north, and up. At each stage the new cuboid formed will have three consecutive Padovan numbers as its sides. Moreover, if successive square faces of the added cuboids are connected by straight lines, the result is a spiral that lies in a plane. The Padovan sequence is very similar to the **Perrin sequence**.[319]

Plateau, Joseph Antoine Ferdinand (1801–1883)

A Belgian physicist who did pioneering experimental work on soap bubbles and soap films, which stimulated the mathematical study of bubbles as **minimal surface**s. In 1829 Plateau carried out an optical experiment that involved looking at the Sun for 25 seconds: this damaged his eyes, and he eventually became blind.

Plateau curves

A set of curves, named after Joseph **Plateau**, that are described by the parametric equations:

$$x = a \sin(m + n)t \,/\, \sin(m - n)t$$
$$y = 2a \sin(mt) \sin(nt) \,/\, \sin(m - n)t$$

If $m = 2n$ the Plateau curves become a circle with a center at (1, 0) and a radius of 2.

Plateau problem

The general problem of determining the shape of the **minimal surface** constrained by a given boundary. It is named after Joseph **Plateau** who noticed that a handful of simple patterns seemed to completely describe the geometry of how soap **bubble**s fit together. Plateau claimed that soap bubble surfaces always make contact in one of two ways: either three surfaces meet at 120° angles along a curve; or six surfaces meet at a vertex, forming angles of about 109°. For instance, in a cluster of bubbles, two intersecting bubbles (of possibly different sizes) will have a common dividing wall (the third surface) that meets the outer surfaces of the bubbles in 120° angles. On the other hand, the edges of the six soap-film faces that emerge within a tetrahedral wire frame, when dipped in a soapy solution, form angles of roughly 109° at a central vertex. Until the American mathematician Jean Taylor came along in the mid-1970s, Plateau's patterns were just a set of empirical rules. However, as a follow-up to her doctoral thesis, Taylor was able to prove that Plateau's rules were a necessary consequence of the energy-minimizing principle–no other yet unobserved configurations were possible–thus settling a question that had been open for more than a century. The forces acting along the surface of a soap bubble all have the same magnitude in all directions. In crystals this is not the case (magnitudes of surface forces differ in different directions, though they may exhibit a grain, analogous to that in a piece of wood), but they still require the least energy to enclose a given volume. Minimal surfaces that model these conditions, like a cube with its corners chopped off or the bottom half of a cone

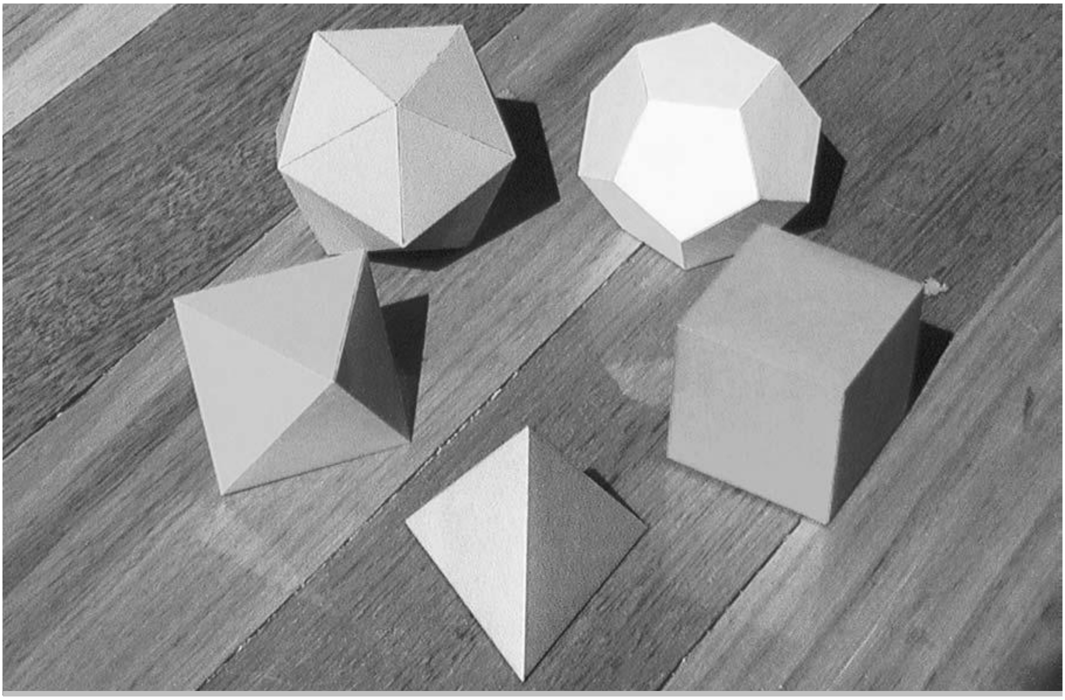

Platonic solid Clockwise from the extreme right: the cube, tetrahedron, octahedron, icosahedron, and dodecahedron. *Robert Webb, www.software3d.com; created using Webb's Stella program*

mounted on a cylinder, are known as *Wullf Shapes*, and provide fertile ground for mathematical study today.[328]

Platonic solid

Any of the five regular **polyhedron**s–solids with regular **polygon** faces and the same number of faces meeting at each corner–that are possible in three dimensions. They are the **tetrahedron** (a pyramid with triangular faces), the **octahedron** (an eight-sided figure with triangular faces), the **dodecahedron** (a 12-sided figure with pentagonal faces), the **icosahedron** (a 20-sided figure with triangular faces), and the **cube** (see table, "The Platonic Solids"). They are named after **Plato** who described them in one of his books, though it was **Euclid** who proved that there are only five regular polyhedra. A regular solid with hexagonal faces cannot exist because if it did, the sum of the angles of any three hexagonal corners that meet would already equal 360°, so such an object would be

The Platonic Solids

	Number of			
Name	**Faces**	**Edges**	**Vertices**	**Schläfi Symbol**
Tetrahedron	4	6	4	{3,3}
Cube	6	12	8	{4,3}
Octagon	8	12	6	{3,4}
Dodecahedron	12	30	20	{5,3}
Icosahedron	20	30	12	{3,5}

planar. See also **Archimedean solid**, **Catalan solid**, and **Johnson solid**.

Platonism

The belief that mathematical objects exist independent of physical models. It is, at the very least, a useful pretense in mathematics, especially in **geometry**.

pleated surface

A surface in **Euclidean space** or hyperbolic space (see **hyperbolic geometry**) that resembles a **polyhedron** in the sense that it has flat faces that meet along edges. Unlike a polyhedron, a pleated surface has no corners, but it may have infinitely many edges that form a **lamination**.

Poggendorff illusion

A **distortion illusion** in which the two ends of a straight line passing behind a rectangle appear offset when, in fact, they are aligned. It was discovered in 1860 by the physicist J. C. Poggendorff, editor of *Annalen der Physik und Chemie* (Annals of physics and chemistry), after receiving a letter from the astronomer Friedrich Zöllner. In this letter, Zöllner described an illusion (see **Zöllner illusion**) he had noticed on a fabric design in which parallel lines intersected by a pattern of short diagonal lines appear to diverge.[134]

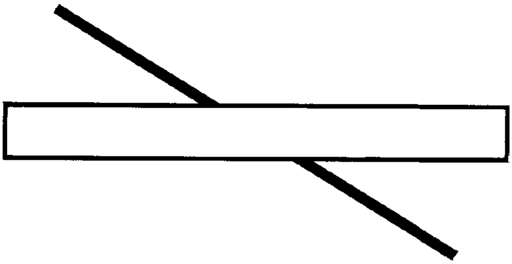

Poggendorff illusion

Poincaré, (Jules) Henri (1854–1912)

A French theoretical physicist and mathematician who almost discovered **relativity theory** before Einstein and who is most famous in mathematics because of his hypothesis known as the **Poincaré conjecture**. Poincaré, who taught for most of his life at the University of Paris, was sometimes referred to as "the last universalist," because of his huge published output on a wide variety of mathematics and mathematical physics. Strangely, he was also clumsy, absent-minded, and inept at simple arithmetic. In order to translate problems in **topology** into questions in **algebra**, he devised (or discovered) ***homotopy** groups*–quantities that capture the essence of multidimensional spaces in algebraic terms, and have the power to reveal similarities between them.

Poincaré conjecture

A proposition in **topology** put forward by Henri **Poincaré** in 1904. Poincaré was led to make his conjecture during his pioneering work in topology, the mathematical study of the properties of objects that stay unchanged when the objects are stretched or bent. In loose terms, the conjecture is that every three-dimensional object that has a set of spherelike properties (i.e., is topologically equivalent to a sphere) can be stretched or squeezed until it is a three-dimensional sphere (a 3-sphere) without tearing (i.e., making a hole in) it. Strictly speaking, the conjecture says that every **closed**, **simply connected** three-**manifold** is **homeomorphic** to the 3-sphere.

Poincaré proved the two-dimensional case and he guessed that the principle would hold in three dimensions. Determining if the Poincaré conjecture is correct has been widely judged the most important outstanding problem in topology–so important that, in 2000, the Clay Mathematics Institute in Boston named it as one of seven Millennium Prize Problems and offered a $1 million prize for its solution. Since the 1960s, mathematicians have shown by various means that the generalized conjecture is true for all dimensions higher than three–the four-dimensional case finally falling in 1982. But none of these strategies work in three dimensions. On April 7, 2002, came reports that the Poincaré conjecture might have been proved by Martin Dunwoody of Southampton University, but within a few days a fatal flaw was found in his proof. Then, in April 2003, what appears to be a genuine breakthrough emerged during a series of lectures delivered at the Massachusetts Institute of Technology by the Russian mathematician Grigori Perelman of the Steklov Institute of Mathematics (part of the Russian Academy of Sciences in St. Petersburg). His lectures, entitled "Ricci Flow and Geometrization of Three-Manifolds," constituted Perelman's first public discussion of important results contained in two earlier preprints. Mathematicians will now scrutinize the validity of Perelman's work (which does not actually mention the Poincaré conjecture by name). In any event, the Clay Institute requires a two-year cooling-off period before the prize can be awarded.

Poincaré disk

The region inside (but not including) a bounding circle, where straight lines are defined to be either diameters of the bounding circle, or arcs of circles that are perpendicular to the bounding circle. The Poincaré disk is a model

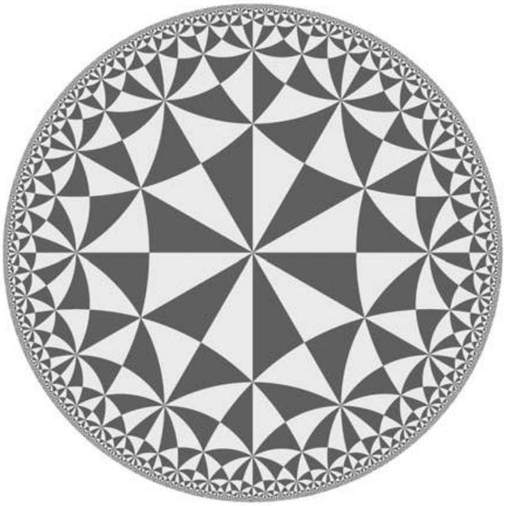

Poincaré disk A pattern on a hyperbolic surface. *Jos Leys, www.josleys.com*

of **hyperbolic geometry**. The great flexibility it allows in specifying the angles of a triangle leads to an infinite variety of ways in which the disk can be tiled.

Poinsot, Louis (1777–1859)

A French mathematician who invented *geometrical mechanics* while investigating how a system of forces acting on a rigid body can be resolved into a single force and a couple (a pair of equal and oppositely directed forces, as when you try to unscrew a bottle cap). Together with Gaspard **Monge**, he helped geometry regain its leading role in mathematical research in France in the eighteenth century. He wrote an important work on polyhedra in 1809, discovering four new regular polyhedra, which are now known as the **Kepler-Poinsot solids**. Two of these had been found by Johannes **Kepler** in 1619 but Poinsot was unaware of this; the two additional ones that Poinsot discovered were the *great dodecahedron* and the *great icosahedron*. In 1810 Augustin **Cauchy** proved that, with this definition of regular, the enumeration of regular polyhedra is complete (although a mistake was discovered in Poinsot's, and hence Cauchy's, definition in 1990 when an internal inconsistency became apparent). Poinsot also worked in number theory, studying **Diophantine equations** with a view to expressing numbers as the difference of two squares and primitive roots.

point

A dimensionless geometric object having no properties other than location or place. More generally, an element in a geometrically described **set**.

point-set topology

Also known as *general topology*, a branch of **topology** concerned with how to put a structure on a **set** in such a way as to generalize the idea of **continuity** for maps from the real numbers to itself. A topology on a set X is a certain set of *open subsets* of the set X which satisfy various axioms. The set X together with this topology is called a **topological space**.

Poisson, Siméon Denis (1781–1840)

A French mathematician whose main interest lay in the application of mathematics to physics, especially in electrostatics and magnetism. He developed a two-fluid theory of electricity and provided theoretical support for the experimental results of others, notably Charles de Coulomb. Poisson also made important contributions to mechanics, especially the theory of elasticity; optics; calculus, especially definite integrals; differential geometry; and probability theory. In all, he wrote more than 300 papers on mathematics, physics, and astronomy, and his *Traité de Mécanique* (1811) was long a standard work.

polar coordinates

A **coordinate** system in which distances are measured from a fixed reference point (the pole) and angles from a fixed reference line. A *polar equation* is one that uses polar coordinates.

pole

(1) The **origin** of a system of **polar coordinates**. (2) In **complex analysis**, a function of a certain simple type of **singularity**. (3) One of the two points, where the axis of rotation of a rotating body, such as Earth, passes through the surface. (4) An old unit of length, also called a *rod*, equal to 5.5 yards.

Pólya's conjecture

A hypothesis put forward by the Hungarian mathematician George Pólya (1887–1985) in 1919. A positive integer is said to be of *even type* if it factorizes into an even number of **prime numbers**; otherwise it is said to be of *odd type*. For example, 4, $= 2 \times 2$, is of even type, whereas 18, $= 2 \times 3 \times 3$, is of odd type. Let $O(n)$ be the number of odd type and $E(n)$ be the number of even type integers in the first n integers. Pólya's conjecture says that $O(n) \geq E(n)$ for all $n \geq 2$. After the conjecture had been checked for all values of n up to 1 million, many mathematicians assumed it was probably true. However, in 1942 A. E. Ingham came up with an ingenious method to show how a counterexample could be constructed, even though there wasn't enough computing power around at the time to do the necessary calculations.[177] Twenty years later, R. S. Lehman ran Ingham's method on a computer to find a counterexample to Pólya's conjecture at $n = 906180359$.[199]

polychoron

An unofficial name (from the Greek *poly* meaning "many" and *choros* meaning "room" or "space") for a four-dimensional **polytope**; it was first advocated by George Olshevsky.

polycube

A **polyhedron** formed by joining unit cubes by their faces. Examples of puzzles that involve polycubes are the **Soma cube** and **Rubik's cube**.

polygon

A plane closed figure whose sides are straight lines. The term *polygon* (from the Greek *poly* for "many" and *gwnos* for "angle") sometimes also refers to the *interior* of the polygon (the open area that this path encloses) or to the **union** of both. A polygon is *simple* if it is described by a single, nonintersecting boundary; otherwise it is said to be *complex.* A simple polygon is called *convex* if it has no internal angles greater than 180°; otherwise it is called *concave.* A polygon is called *regular* if all its sides are of equal length and all its angles are equal (see table, "Regular Polygons"). Any polygon, regular or irregular, has as many angles as it has sides. See **constructible** (for constructible polygons).

Regular Polygons

Name	Sides	Angle (= 180°–360°/sides)
Equilateral **triangle**	3	60°
Square	4	90°
Regular **penta**gon	5	108°
Regular **hexa**gon	6	120°
Regular heptagon	7	128.57° (approx.)
Regular **octa**gon	8	135°
Regular nonagon	9	140°
Regular **decagon**	10	144°
Regular hectagon	100	176.4°
Regular megagon	10^6	179.99964°
Regular **googol**gon	10^{100}	180° (approx.)

polygonal number

The number of equally spaced dots needed to draw a **polygon**. Polygonal numbers, which are a type of **figurate number**, include square numbers, triangular numbers, and hexagonal numbers.

polyhedron

A three-dimensional object whose faces are all **polygon**s and whose edges are shared by exactly two polygons. *Polyhedron* comes from the Greek *poly* for "many" and -*hedron* meaning "base," "seat," or "face." Every polyhedron in three-dimensional space consists of (two-dimensional) faces, (one-dimensional) edges, and (zero-dimensional) vertices. Sometimes the term *polyhedron* is used to apply to figures in more than three dimensions; however, analogs of polyhedra in the **fourth dimension** or higher are also referred to as **polytope**s. Polyhedrons, like polygons, may be *convex* or *nonconvex.* If a line that connects any two points on the surface of a polyhedron is completely inside or on the polyhedron, the figure is convex. Otherwise, it is nonconvex or *concave.* A polygon is *regular* if all of its faces are exactly the same size and shape and if the same number of faces meet at each vertex. There are only five regular convex polyhedra–the **Platonic solid**s. Another four regular nonconvex polyhedra are called **Kepler-Poinsot solid**s. However, the term *regular polyhedra* is sometimes used to describe only the Platonic solids. A convex polyhedron is said to be *semi-regular* if its faces have a similar arrangement of nonintersecting regular plane convex polygons of two or more different types about each vertex. These solids, of which there are 13 different kinds, are commonly called the **Archimedean solids**. A *dual* of a polyhedron is another polyhedron in which faces and vertices occupy complementary locations. The duals of the Archimedean solids are known as the **Catalan solid**s. A **quasi-regular polyhedron** is the solid region interior to two dual regular polyhedra; only two exist: the cuboctahedron and the icoidodecahedron. There are also infinite families of **prism**s and **antiprism**s. In total there are 92 convex polyhedra with regular polygonal faces (and not necessarily equivalent vertices); these are the **Johnson solids**.

The oldest known examples of man-made polyhedra were found on the islands of northeastern Scotland and date back to Neolithic times, between 2000 and 3000 B.C. These stone figures are about 2 inches in diameter and many are carved into rounded forms of regular polyhedra. Examples including cubical, tetrahedral, octahedral, and dodecahedral forms, one which is the dual of the pentagonal prism, are on display in the Museum of Scotland and in Oxford's Ashmolean Museum.

polyiamond

A shape made from identical equilateral **triangle**s that have been joined at their edges.

polynomial

An expression in which whole-numbered powers of a variable multiplied by numerical coefficients are added together. The powers of the variable must be positive integers, or zero. An example of a polynomial is $3x^3 + 7x - 2$. A polynomial *equation* has a polynomial expression, or zero, on each side of the equal sign: for example, $4a^2 - 5.6a + 1.7 = 0$ or $5x^2 - 1 = 8x + 2x^4$. The type of polynomial expression or equation is determined by the highest power present of the variable. A **quadratic**, for example, has nothing higher than a squared term. **Cubic**s, **quartic**s, and **quintic**s have maximum powers of three, four, and five, respectively. Mathematicians who have done pioneering work on each of these higher types of polynomial equations tended, for some reason, to have had colorful and star-crossed lives. Niccolo **Tartaglia**, who first solved the cubic, failed miserably as a mathematician for the rest of his life, largely because he spent it trying to discredit Girolamo **Cardano**. Tartaglia told Cardano his method of solution and swore him to secrecy but Cardano went ahead and published the solution anyway. Cardono himself lived a long unhappy life and his only son was executed for murder. Lodovico Ferrara, Cardano's student, who solved the general quartic, was poisoned, probably by his sister, over an inheritance dispute. Finally, Evariste **Galois**, who showed that the general quintic was unsolvable, died in a duel at the age of 20.

polyomino

A two-dimensional shape made by connecting *n* squares of the same size along their edges. The polyomino is a generalization of the **domino**, which can be thought of as a polyomino with $n = 2$. Several puzzles involving polyominoes made from three or more squares stuck together became popular around the beginning of the twentieth century. The best-known of these is the Broken Chessboard, presented as problem no. 74 in Henry **Dudeney**'s book *The Canterbury Puzzles* (1907).[87] Dudeney gives an amusing introduction to the problem, first quoting from Hayward's *Life of William the Conqueror* (1613) about an incident in which Prince Henry,

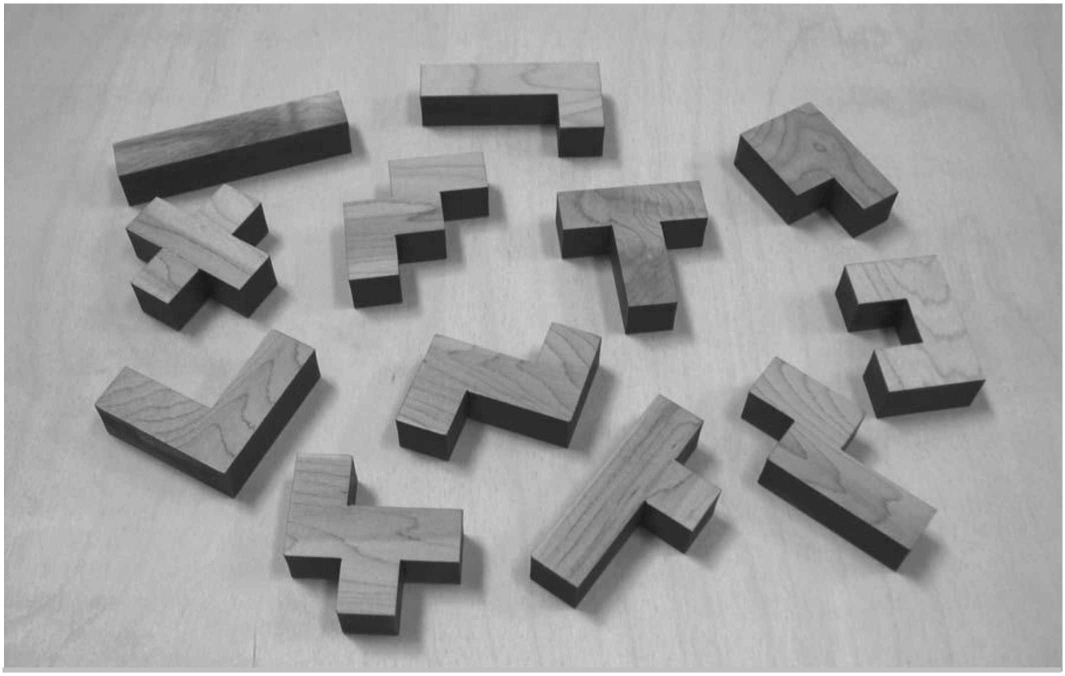

polyomino A complete set of 12 solid pentominos, also known as planar pentacubes. *Kadon Enterprises, Inc., www.gamepuzzles.com* one of William's sons, smashes a chessboard over the head of his brother, Prince Robert, and then adding his own mathematical appendix. By a curious quirk, Dudeney reveals, the board breaks into 13 pieces, 12 of which are the different possible ways in which five squares can be arranged and one square 2 × 2 piece. The puzzle is to reconstitute the board from these fragments. This is the earliest example of a mathematical recreation that involves pentominoes, a name that wasn't coined until 1953 when Solomon **Golomb** first used it in a talk to the Harvard Math Club. Golomb not only invented the nomenclature for polyominoes but did much of the pioneering research on them; his work was brought to a wide audience and popularized by Martin **Gardner** in his *Scientific American* column, beginning in 1957.[114, 136, 137]

Golomb was particularly interested in the pentomino, for a reason that becomes clear (see table, "Types of Polyomino"). The familiar domino comes in just one configuration as, obviously, does the trivial monomino. A triomino can have two different shapes–three squares in a line or in an L shape. The tetromino (or tetramino) has five distinct arrangements and is used in the popular video game **Tetris**. Thus for $n \leq 4$ the number of distinct pieces is ≤ 5, which restricts the variety of combinations possible. On the other hand, for $n > 5$, the number of different pieces is large, making analysis of problems difficult and games based on such polyominoes difficult and unwieldy. However, pentominoes, for which $n = 5$, come in 12 unique configurations, which is just about an ideal balance between tractability and combinatorial richness, hence the high level of interest in this particular type of polyomino. The problem of tiling an 8 × 8 chessboard with a square hole in the center using pentominoes was first solved in 1935 and found, by computer in 1958, to have exactly 65 solutions. Another standard pentomino puzzle is to arrange the set of 12 possible shapes into rectangles without holes: 3 × 20, 4 × 15, 5 × 12, and 6 × 10. Pentominoes are prominently featured in a subplot of the novel *Imperial Earth* by Arthur C. Clarke.

Hexominoes and heptominoes come in 35 and 108 unique arrangements, respectively. One of the 108 heptomino configurations, however, has a hole–a region that is not tiled with squares but that is unconnected to the exterior of the polyomino–and may or may not be counted as a valid piece depending on the rules of a particular game. All polyominoes made of seven or more squares may contain holes. There is no known algorithm or formula for calculating how many distinct polyominoes of each order there are.

Related to polyominoes are **polyiamond**s (formed from equilateral triangles) and polyhexes (formed from regular hexagons). The three-dimensional analog of polyominoes uses cubes instead of squares; an example is the **Soma cube**.

Types of Polyomino

Name	Number of Unit Squares	Number of Distinct Configurations
Monomino	1	1
Domino	2	1
Triomino	3	2
Tetromino	4	5
Pentomino	5	12
Hexomino	6	35
Heptomino	7	108

polytope

A higher-dimensional analogue of a **polygon** or **polyhedron**. The number of possible regular polytopes depends on the number of dimensions. In two dimensions there

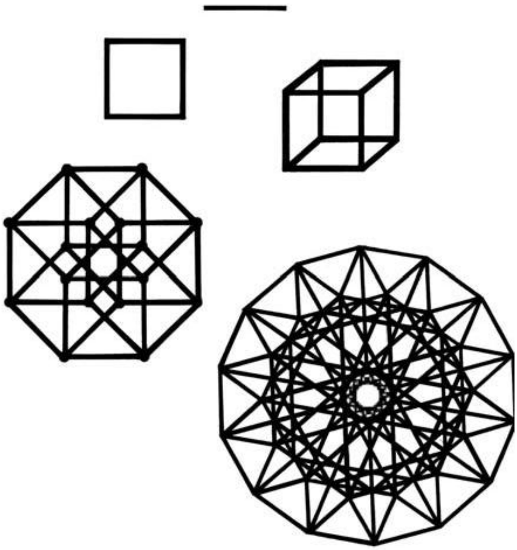

polytope In descending order: a straight line, a square, a cube, a tesseract (four-dimensional cube), and a regular, multidimensional polytope.

are infinitely many possible regular polygons; in three dimensions there are five possible regular polyhedra; in four dimensions there are six possible regular polytopes; and in each number of dimensions higher than four, there are just three possible regular polytopes, analogous to the three-dimensional tetrahedron, cube, and octahedron. A four-dimensional polytope is also sometimes called a *polyhedroid* or a *polychoron*. Just as a polygon has vertices and edges, and a polyhedron has vertices, edges, and faces, a four-dimensional polytope has vertices, edges, faces, and cells, where a cell is a three-dimensional figure.[75] See also **Boole (Stott), Alicia**.

Poncelet, Jean Victor (1788–1867)

A French mathematician who substantially advanced **projective geometry**. With Brianchon, he proved *Feuerbach's theorem* on the **nine-point circle** in 1820–1821, and also suggested the theorem proved by Jakob **Steiner** and now called the *Poncelot-Steiner theorem* that Euclidean constructions (see **constructible**) can be done with a straightedge alone. As a soldier in Napoleon's army, he was captured and imprisoned in Russia. While in prison from 1813–1814, he organized and wrote down his discoveries, and the result was published as *Traité des Propriétés Projectives des Figures* (Treatise on the properties of the projections of shapes, 1822). To serve as an introduction to this work, he also wrote *Applications D'analyse et de Géométrie* (Applications of analysis and geometry, 2 vols., 1862–1864).

Poncelet's theorem

Given an **ellipse**, and a smaller ellipse entirely inside it, start at a point on the outer ellipse, and, moving clockwise, follow a line that is **tangent** to the inner ellipse until you hit the outer ellipse again. Repeat this over and over again. It may be that this path will never hit the same points on the outer ellipse twice. However, if it *does* close up in a certain number of steps, then something amazing is true: *all* such paths, starting at *any point* on the outer ellipse, close up in the same number of steps. This fact is Poncelet's theorem, also known as *Poncelet's closure theorem,* and is named after the Jean **Poncelet**.

Pony Puzzle

One of Sam **Loyd**'s best known and commercially successful puzzles. It consists of just six pieces of the silhouette of a horse that have to be assembled in the most sensible way. Loyd got the idea for the puzzle from the governor of Philadelphia, Andrew Curtin, when the two were on a steamer returning from Europe and talking about the White Horse of Uffington–a famous chalk figure carved on a hillside in Berkshire, England.

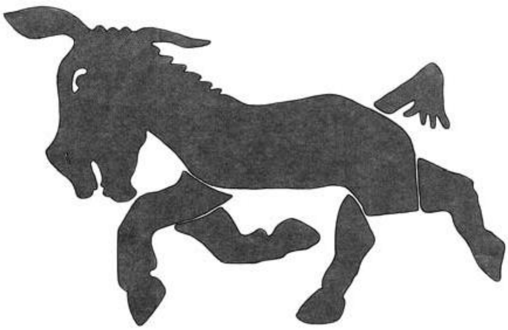

Pony Puzzle One of Sam Loyd's best known brain-teasers.

postage stamp problems

Mathematical puzzles that involve postage stamps have been around almost as long as postage stamps themselves, the first of which, the Penny Black, was issued by Great Britain on May 6, 1840. Some such puzzles ask what postage amounts can or can't be made with stamps of certain values. Others are based on the ways that a block of stamps can be folded or torn along the perforations.

PUZZLE

One of this type appears as problem no. 285 in Henry **Dudeney**'s *Amusements in Mathematics* (1917).[88] It starts by saying you have just bought 12 stamps in a rectangular block of three rows with four stamps in each. (In Dudeney's diagram, they are labeled 1, 2, 3, 4 across the top row, and so on.) He goes on: "[A] friend asks you to oblige him with four stamps, all joined together–no stamp hanging on by a mere corner. In how many different ways is it possible for you to tear off those four stamps? You see, you can give him 1, 2, 3, 4, or 2, 3, 6, 7, . . . , and so on. Can you count the number of different ways in which those four stamps might be delivered?" This can be thought of as a problem involving tetrominoes, which are a type of **polyomino**.

Solutions begin on page 369.

The postage stamp problem, also known as the *Frobenius problem,* is a long-standing challenge in **number theory** and in **computer science**. Suppose a country issues n different denominations of stamps but allows no more than m stamps to be put on a single letter. The postage stamp problem is to write and implement an algorithm (a stepwise set of rules) that, for any given values of m and

n, computes the greatest consecutive range of postage values, from one on up, and all possible sets of denominations that realize that range. For example, for $n = 4$ and $m = 5$, the stamps with values (1, 4, 12, 21) allow the postage values 1 through 71. If the values of the stamps are constant and not part of the input, algorithms can be, and have been, devised that give a short-cut solution. However, in the general case where the number of stamp values *is* part of the input, the postage stamp problem has been shown to be an **NP-hard problem**, and thus not susceptible to an efficient algorithmic approach.

potato paradox

Fred brings home 100 pounds of potatoes, which (being purely mathematical potatoes) consist of 99 percent water. He then leaves them outside overnight so that they consist of 98 percent water. What is their new weight? The surprising answer is 50 pounds!

potential theory

The study of *harmonic functions.* These functions satisfy **Laplace**'s equation, a certain type of **partial differential equation** that commonly arises in physics in problems to do with gravity and electromagnetism.

power

A word that is almost never used in its correct, original sense any more. Strictly speaking, if we write $8 = 2^3$, then 2 is the **base**, 3 is the **exponent**, and 8 is the power. But almost everyone, including most mathematicians, would say that 3 is the power, and that "power" and "exponent" mean the same thing. The misuse has probably come about from a misunderstanding of statements such "eight is the third power of two."

power law

A type of mathematical pattern in which the frequency of an occurrence of a given size is inversely **proportional** to some **power** (or exponent) of its size. For example, in the case of avalanches or earthquakes, large ones are fairly rare, smaller ones are much more frequent, and in between are cascades of different sizes and frequencies. Power laws define the distribution of catastrophic events in self-organized critical systems (see **self-organization**).

power series

An infinite sum of the form

$$a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots,$$

where a is a number of any type and x is the variable. Power series are commonly used to define **function**s. For example, the **sine** function can be written

$$\sin x = x - x^3/3! + x^5/5! - x^7/7! + \ldots,$$

where "!" stands for **factorial**. Although the series has infinitely many terms, these get small so quickly that only the first few terms make much of a contribution.

power set

The **set** of all **subsets** of a given set, including the **empty set** and the original set. For example, if the original set is {*a*, *b*, c} then the power set is {{∅, {*a*}, {*b*}, {*c*}, {*a*, *b*}, {*a*, *c*}, {*b*, *c*}, {*a*, *b*, c}}.

power tower

A way of representing very large numbers in terms of stacks of **exponent**s. For example:

$$10^{3^{3^{3^{3}}}} = 10^{3^{3^{27}}} = 10^{3^{2541865828329}} = \text{enormous}$$

In general, and especially if the number at the top of the tower is fairly large, adding another exponent to the bottom of a tower will make the value of the tower much larger than will increasing the size of the bottom exponent. This leads to the (somewhat counterintuitive) result that to know which of two towers is the larger, you can look at how many exponents are in the tower and know right away which is larger. For example:

$$1.1^{1.1^{1.1^{1,000}}} \text{ is much larger than } 1{,}000^{1{,}000^{1{,}000}}$$

In the case of an infinite power tower of the form

$$x^{x^{x^{\cdots}}},$$

the maximum value that x can take and still cause the tower to converge to a finite value is $e^{(1/e)} = 1.444667\ldots.$ The minimum value of x that will produce convergence is $1/e^e = 0.065988\ldots.$

powerful number

Also known as a *squarefull number,* a positive whole number n such that for every **prime number** p dividing n, p^2 also divides n. Every powerful number can be written as a^2b^3, where a and b are positive integers. The first few powerful numbers are 1, 4, 8, 9, 16, 25, 27, 32, 36, 49, 64, 72, 81, 100, and 108. Pairs of consecutive powerful numbers exist, such as (8,9), (288,289), and (675,676). However, no three consecutive powerful numbers are known and, in 1978, Paul **Erdös** conjectured that none exist.

practical number

A number n such that every positive integer less than n is either a divisor or a sum of distinct divisors of n. The first few practical numbers are 1, 2, 4, 6, 8, 12, 16, 18, 20, and 24. All **perfect number**s are practical.

prime number

> *[U]pon looking at prime numbers one has the feeling of being in the presence of one of the inexplicable secrets of creation.*
>
> –Don Zagier

An **integer** greater than 1 that is divisible only by 1 and itself. Prime numbers have fascinated mathematicians for centuries, in large part because of how they are distributed. At first sight, their occurrence looks random; yet, on closer inspection, they reveal a subtle order or pattern, that seems to hold deep truths about the nature of mathematics and of the world in which we live. The German-born American mathematician Don Zagier (1951–), in his inaugural lecture at Bonn University, put it this way:

> There are two facts about the distribution of prime numbers which I hope to convince you. . . . The first is that despite their simple definition and role as the building blocks of the natural numbers, the prime numbers . . . grow like weeds among the natural numbers, seeming to obey no other law than that of chance, and nobody can predict where the next one will sprout. The second fact is even more astonishing, for it states just the opposite: that the prime numbers exhibit stunning regularity, that there are laws governing their behavior, and that they obey these laws with almost military precision.

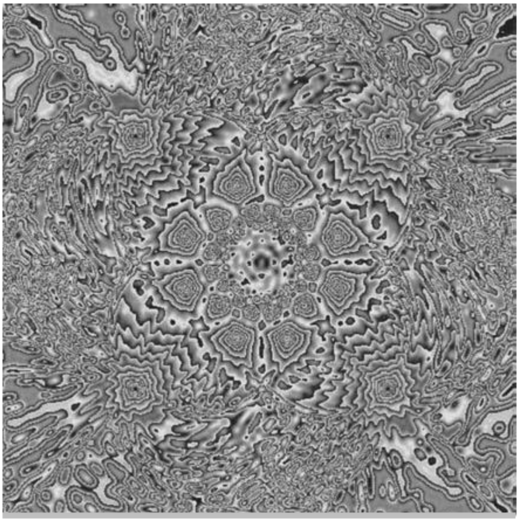

prime number A remarkable pattern generated by the distribution of the first few billion primes. *Jean-Francois Colonna/Ecole Polytechnique*

The prime numbers are 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, ... The *fundamental theorem of arithmetic* declares that the primes are the building blocks of the positive integers: every positive integer is a product of prime numbers in one and only one way, except for the order of the factors. This is the key to their importance: the prime factors of an integer determine its properties. The ancient Greeks proved (c. 300 B.C.) that there are infinitely many primes and that they are irregularly spaced; in fact, there can be arbitrarily large gaps between successive primes. On the other hand, in the nineteenth century it was shown that the number of primes less than or equal to *n* approaches *n*/log *n,* as *n* gets very large (a result known as the *prime number theorem*), so that a rough estimate for the *n*th prime is *n* log *n.* In his *Disquisitiones Arithmeticae* (1801), Carl **Gauss** wrote: "The problem of distinguishing prime numbers from composite numbers and of resolving the latter into their prime factors is known to be one of the most important and useful in arithmetic. It has engaged the industry and wisdom of ancient and modern geometers to such an extent that it would be superfluous to discuss the problem at length. . . . Further, the dignity of the science itself seems to require that every possible means be explored for the solution of a problem so elegant and so celebrated." The earliest known primality test is the **sieve of Eratosthenes**, which dates from around 240 B.C. However, high-speed computers and fast algorithms are needed to identify large primes. New record-breaking primes tend to be of the variety known as **Mersenne prime**s, since these are the easiest to find. About 6,000 prime numbers are known of which the largest is $2^{20996011} - 1$.

Much remains unknown about the primes. As Martin **Gardner** said:[113] "No branch of **number theory** is more saturated with mystery. . . . Some problems concerning primes are so simple that a child can understand them and yet so deep and far from solved that many mathematicians now suspect they have no solution. Perhaps they are 'undecideable.' Perhaps number theory, like quantum mechanics, has its own uncertainty principle that makes it necessary, in certain areas, to abandon exactness for probabilistic formulations." One of the greatest unsolved problems in mathematics is the **Riemann hypothesis** concerning the distribution of prime numbers. See also **Goldbach's conjecture**, **twin primes**, **Ulam spiral**, and **Ishango bone**.

primitive root

A primitive root for a **prime number** *p* is one whose powers generate all the nonzero integers **modulo** (or mod) *p*. For example, 3 is a primitive root modulo 7 since $3 = 3^1$, $2 = 3^2 \bmod 7$, $6 = 3^3 \bmod 7$, $4 = 3^4 \bmod 7$, $5 = 3^5 \bmod 7$, $1 = 3^6 \bmod 7$.

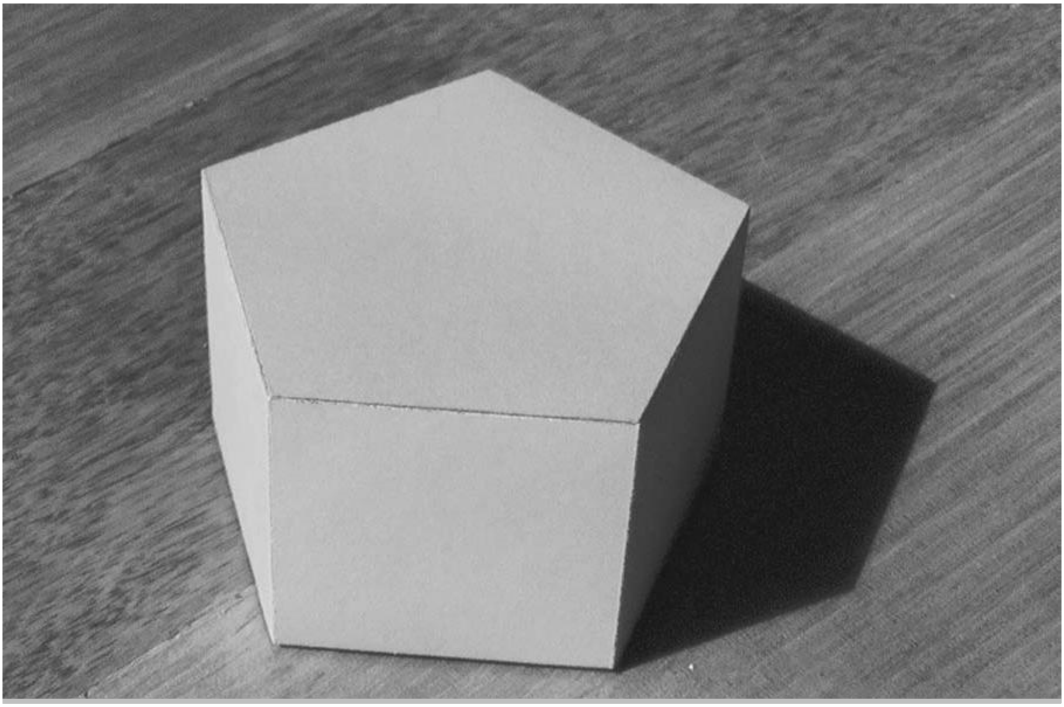

prism A pentagonal prism. *Robert Webb, www.software3d.com; created using Webb's Stella program*

primitive root of unity

The **complex number** z such that $z^n = 1$ but z^k is not equal to 1 for any positive integer k less than n.

primorial

Also known as a *prime factorial,* the product of all **prime numbers** that are less than or equal to a given prime p; it is denoted $p\#$. For example $3\# = 2 \times 3 = 6$, $5\# = 2 \times 3 \times 5 = 30$, and $13\# = 2 \times 3 \times 5 \times 7 \times 11 \times 13 = 30{,}030$.

Prince Rupert's problem

The problem of pushing a **cube** through a hole in another cube of equal or less size; it is named after Prince Rupert (1619–1682), a nephew of England's King Charles I, who won a wager that a hole could be made in one of two equal cubes large enough for the other cube to slide through. The mathematics of cubes passing through cubes was considered by John **Wallis**. Later, in 1816, a solution was published posthumously by the Dutch mathematician Pieter Nieuwland (1764–1794) to the question: What is the largest cube that can be passed through a unit cube (a cube of which each side is one unit long)? Nieuwland answered this by finding the largest square that fits inside a unit cube. When viewed from directly above one apex, a unit cube has the outline of a regular hexagon of side $\sqrt{3}/\sqrt{2}$. The largest square that will go into a cube has a face that can be inscribed within this hexagon; the length of its edge is $\sqrt{6} - \sqrt{2} = 1.03527618$.

prism

A **semi-regular polyhedron** constructed from two **congruent** *n-sided* **polygons** and *n* **parallelograms**. The word comes from the Greek *prizma,* which relates to cutting or sawing. A *prismoid* resembles a prism but has bases that are **similar** rather than congruent, and sides that are **trapezoids** rather than parallelograms. An example of a prismoid is the **frustum** of a **pyramid**. A *prismatoid* is a polyhedron with all its vertices lying in two parallel planes.

prisoner's dilemma

A problem in **game theory** first described by the Canadian-born Princeton mathematician Albert Tucker (1905–1995) in 1950, while addressing an audience of psychologists at Stanford University, where he was a visiting professor. It runs along these lines: Al and Bob have been arrested for holding up the Anyapolis State Bank and have been put in separate cells. Each cares a lot more about his personal freedom than he does about his accomplice's welfare. A clever prosecutor makes the following offer to each. "You may choose to confess or remain silent. If you confess and your accomplice remains silent, I'll drop all charges against you and use your testimony to ensure that your accomplice does serious time. Likewise, if your accomplice confesses while you remain silent, he'll go free while you do the time. If you both confess I get two convictions, but I'll see to it that you both get early parole. If you both remain silent, I'll have to settle for token sentences on firearms possession charges. If you wish to confess, you must leave a note with the jailer before I come back tomorrow morning." The dilemma faced by the prisoners is that, whatever the other does, each is better off confessing than remaining silent. But the outcome when both confess is worse for each than the outcome if both stay silent!

Tucker's paradox was based on puzzles with a similar structure that had been devised in 1950 by Merrill Flood and Melvin Dresher as part of the Rand Corporation's investigations into game theory (which Rand pursued because of possible applications to global nuclear strategy). Flood and Dresher hadn't published much about their work, but the prisoner's dilemma attracted an enormous amount of attention in subjects as diverse philosophy, biology, sociology, political science, and economics, as well as game theory itself. A common view is that the puzzle illustrates a conflict between individual and group rationality. A group whose members pursue rational self-interest may all end up worse off than a group whose members act contrary to rational self-interest. More generally, if the payoffs aren't assumed to represent self-interest, a group whose members rationally pursue any goals may all meet less success than if they hadn't rationally pursued their goals individually.

probability

A measure of how likely it is that some event will occur, given as a number between 0 (impossible) and 1 (certain). Usually, probability is expressed as a ratio: the number of experimental results that would produce the event divided by the number of experimental results considered possible. For example, the probability of drawing the five of hearts from an ordinary deck of cards is one in fifty-two (1:52).

probability theory

The branch of **mathematics** that deals with the possible outcomes of events and their relative likelihoods. While mathematicians agree on how to calculate the probability of certain events and how to use those calculations in certain ways, there's plenty of disagreement as to what the numbers actually mean. Probability divides into two main concepts: *aleatory probability,* which represents the likelihood of future events whose occurrence is governed by some random physical phenomenon like tossing dice or spinning a wheel; and *epistemic probability,* which represents our uncertainty of belief about past events that either did or didn't occur, or uncertainty about the causes of future events. An example of the latter is when we say that it's "probable" that a certain suspect committed a crime based on the available evidence. It is an open question whether aleatory probability is reducible to epistemic probability based on our inability to precisely predict every force that might affect the roll of a die, or whether such uncertainties exist in the nature of reality itself, particularly at the level of **quantum mechanics**. One of the earliest mathematical studies on probability was written by Girolamo **Cardano**. Among other important contributions to the development of the subject were those by Blaise **Pascal**, Pierre de **Fermat**, Jakob Bernouilli (see **Bernouilli family**), Joseph **Lagrange**, Pierre **Laplace**, Carl **Gauss**, Siméon **Poisson**, Abraham **de Moivre**, Pafnuty Chebyshev, Andrei Markov (see **Markov chain**), and Andrei **Kolmogorov**.

PUZZLES

The following are two problems in probability. Both are easier to solve if the various possible outcomes are written down in the form of a table.

1. You meet a stranger on the street, and ask how many children he has. He truthfully says two. You ask "Is the older one a girl?" He truthfully says yes. What is the probability that both children are girls? What would the probability be if your second question had been "Is at least one of them a girl?", with the other conditions unchanged?
2. You are in a game of Russian roulette, but this time the gun (a six-shooter revolver) has three bullets in a row in three of the chambers. The barrel is spun only once. Each player then points the gun at his head and pulls the trigger. If he is still alive, the gun is passed to the other player who then points it at his own head and pulls the trigger. The game stops when one player dies. Would you stand a better chance of surviving if you shoot first or second, or does it make any difference?

Solutions begin on page 369.

See also **Buffon's needle**, **birthday problem**, **Monty Hall problem**, and **St. Petersburg paradox**.

Proclus Diadochus (c. A.D. 410–485)

The last major Greek philosopher; his *Commentary on Euclid* is the main original source we have on the early history of Greek geometry. He also wrote *Hypotyposis,* which gives a detailed account of the Earth-centered astronomical theories of Hipparchus and Ptolemy.

product

The result of one or more multiplications.

projectile

An object, such as a baseball, spear, or cannonball, that is thrown, fired, or otherwise propelled but that can't propel itself. For centuries, philosophers and mathematicians debated what path a projectile follows under gravity. Galilei Galileo was the first to establish that (in the absence of air resistance) this path is a **parabola**.

projective geometry

The branch of **geometry** that deals with properties of geometric figures that remain unchanged under projection. A mathematical theory of perspective grew out of the studies of Renaissance architects and painters who asked themselves how to best represent a three-dimensional object on a two-dimensional surface. The Greeks had done some early work on perspective, and the great geometer **Pappus of Alexandria** is credited with the first theorem in projective geometry. However, the subject reached mathematical maturity through the efforts first of Girard **Desargues**, and then, much later through the work of Jean **Poncelet** and by Karl von Staudt (1798–1867).

The basic elements of projective geometry are points, lines, and planes. These elements retain their characteristics under projection; for example, the projection of a line is another line, and the point of intersection of two lines is projected into another point that is the intersection of the projections of the two original lines. However, lengths and ratios of lengths are not invariant under projection, nor are angles or the shapes of figures. The concept of parallelism doesn't appear at all in projective geometry; any pair of distinct lines intersects in a point, and if these lines are parallel in the sense of **Euclidean geometry**, then their point of intersection is at infinity. The plane that includes the ideal line, or line at infinity, consisting of all such ideal points, is called the **projective plane**. Two properties that are invariant under projection are the order of three or more points on a line and the harmonic relationship, or cross-ratio, among four points, A, B, C, D, that is, $AC/BC : AD/BD$. The most remarkable concept in projective geometry is that of duality. In the plane, the terms *point* and *line* are dual and can be interchanged in any valid statement to yield another valid statement; in space, the terms *plane, line,* and *point* are interchanged with *point, line,* and *plane,* respectively, to yield dual statements. Entire theorems also occur in dual pairs, so that one can be instantly transformed into the other. For example, *Pascal's theorem* (given a hexagon inscribed in a **conic section**, the three pairs of the continuations of opposite sides meet on a straight line) is the dual of *Brianchon's theorem* (given a hexagon circumscribed on a conic section, the lines joining opposite diagonals meet in a single point). In fact, *all* the propositions in projective geometry occur in dual pairs.

projective plane

The surface you would get if you glued the edge of a disk to the edge of a **Möbius band**. This sounds easy to do, since a disk and a Möbius band each have one edge. But the process becomes hopelessly tangled and is, in fact, impossible. The projective plane needs a **fourth dimension**, in addition to the three we live in (up-down, left-right, and back-forth), to be fully realized. The idea of the projective plane arose from the study of perspective by mathematicians and painters in the Renaissance. In trying to represent parallel lines in space on the two-dimensional surface of a painting, it was found useful to introduce the notion of a *line at infinity* on which parallel lines met. The study of the geometry that adds this extra line of ideal points to the ordinary familiar plane came to be known as **projective geometry**, because of its use in studying projections of figures onto different lines. This idea was even more important in three dimensions, since projections are used for representing three-dimensional figures on planes. An interesting property of the projective plane is that any "straight" line on it, followed far enough, comes back to the starting point. (The old arcade version of the game of Asteroids was played on a virtual projective plane: the screen was a disk, and when an asteroid went off one edge of the screen it emerged on the opposite side.) The projective plane is also *nonorientable,* as a result of which any two-dimensional object pushed along a path back to its starting point would be reversed.

prolate

(1) Rounded like an **egg**. (2) Having a polar diameter greater than the equatorial diameter. See also **spheroid**.

pronic number

Also known as a *rectangular* or *oblong number,* a number that is the product of two consecutive integers: 2 (1 × 2), 6 (2 × 3), 12 (3 × 4), 20 (4 × 5), ... The pronic numbers are twice the **triangular numbers**, and represent the lengths that produce the musical intervals: octave (1:2), fifth (2:3), fourth (3:4), major third (4:5) . . . . Pronic seems to be a misspelling of *promic,* from the Greek *promekes,* for "rectangular" or "oblong"; however, the "n" form goes back at least as far as Leonhard **Euler** who used it in series one, volume fifteen of his *Opera.*

proof

A sequence of statements in which each subsequent statement is derivable from one of the previous statements or from an **axiom** of a **formal system**. The final statement of a proof is usually the **theorem** that one has set out to prove.

proper divisor

See **aliquot part**.

proportional

A variable a is said to be *directly proportional* to b if a/b is a constant. The relationship is written $a \propto b$, which implies that $a = kb$, where k is a constant. If a is *inversely proportional* to b, this is written $a \propto 1/b$.

pseudoprime

A number that passes the test of **Fermat's little theorem** (FLT) for **prime numbers** but actually isn't a prime. FLT says that if p is prime and a is **coprime** to p, then $a^{p-1} - 1$ is divisible by p. If a number x is not prime, a is coprime to x, and x divides $a^{x-1} - 1$, then x is called a *pseudoprime to base a.* A number x that is a pseudoprime for all values of a that are coprime to x is called a **Carmichael number**. The smallest pseudoprime in base 2 is 341. This isn't prime because 341 = 11 × 31; however, it satisfies FLT: $2^{340} - 1$ is divisible by 341.

pseudosphere

A saddle-shaped surface that is produced by rotating a **tractrix** about its **asymptote**. The name *pseudosphere,* which means "false sphere," is misleading because it suggests something that is spherelike; however, a pseudosphere is almost exactly the opposite of a sphere. Whereas a sphere has a constant positive **curvature** (equal to *+l/r,* where *r* is the radius) at every point on its surface, a pseudosphere has a constant negative curvature (equal to *–l/r*) everywhere. As a result, a sphere has a closed surface and a finite area, while a pseudosphere has an open surface and an infinite area. In fact, although both the two-dimensional plane and a pseudosphere are infinite, the pseudosphere manages to have more room! One way to think of this is that a pseudosphere is more intensely infinite than the plane. Another result of the pseudosphere's negative curvature is that the angles of a triangle drawn on its surface add up to less than 180°. The geometry on the surfaces of both the sphere and the pseudosphere is a two-dimensional **non-Euclidean geometry**–spherical (or elliptical) geometry in the case of the sphere and **hyperbolic geometry** in the case of the pseudosphere. Astronomers currently suspect that the universe we live in may have a hyperbolic geometry and thus have properties analogous to those of a pseudosphere.

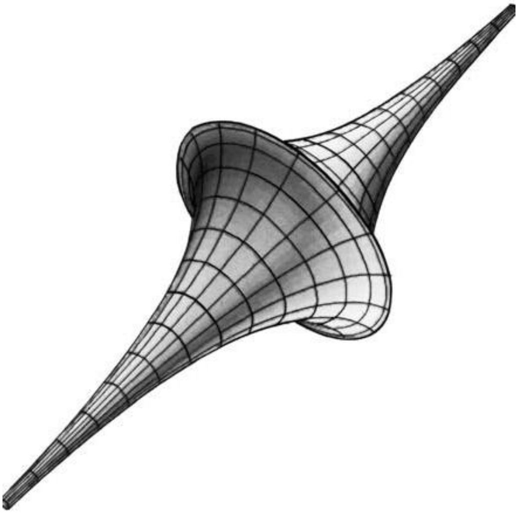

pseudosphere

Ptolemy's theorem

The sum of the products of the two pairs of opposite sides of a convex cyclic **quadrilateral** (see **cyclic polygon**) is equal to the product of the lengths of the diagonals. The theorem is named after the mathematician, astronomer, and geographer Ptolemy of Alexandria.

pure mathematics

Mathematics for the sake of its internal beauty or logical strength. Compare with **applied mathematics**.

pursuit curve

The path an object takes when chasing after another object in the most effective way. Pursuit curves can arise in a variety of situations, for example when a lion is chasing after a gazelle or a heat-seeking missile is homing in on a moving target. Suppose that four ants are at the corners of a square. They start to crawl clockwise at a constant rate, each moving toward its neighbor. At any instant, they mark the corners of a square. As the ants get closer to

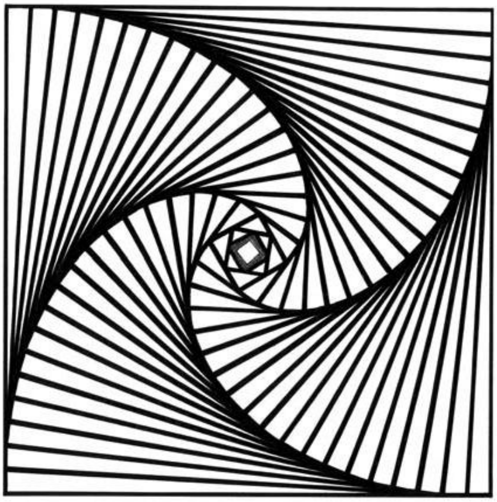

pursuit curve A set of superimposed snapshots shows the lines of sight at regular intervals of four bugs chasing one another, all moving at the same speed after starting at the corners of a square. *John Sharp*

the original square's center, the new square they define rotates and diminishes in size. In reaching the center, each ant travels on a **logarithmic spiral** with a length equal to the side of the original square. Superimposed snapshots of the ants' progress give rise to an intriguing pattern.

puzzle jug

A practical joke in the form of a drinking vessel that, unless the trick is known, spills all over the user when he tries to take a sip. As the jug is tilted, liquid pours out from the pierced decoration surrounding the pot. The puzzle is to learn the secret of how to successfully drink without getting wet. Puzzle jugs are the oldest known **mechanical puzzle**s. Several Phoenician examples are in New York's Metropolitan Museum of Art, and in ninth-century Turkey they were especially popular. Puzzle jugs have been manufactured in Germany, Holland, France, and several other European countries since the end of the thirteenth century. The Exeter puzzle jug, on display in the Royal Albert Museum in Exeter, Devon, was probably made in the Saintonge region of western France, around 1300, and is one of the finest examples of medieval pottery imported to England. Puzzle jugs came back into vogue in the eighteenth and nineteenth centuries and are still manufactured by ceramicists in France, Germany, and Britain. And the secret? That is really quite open, and can be handled by anyone prepared to get to the bottom of it.

puzzle rings

A **mechanical puzzle** consisting of a group of interlocking rings that fit together to form an intricate design; when apart, the rings remain interconnected and pose a challenging puzzle. The puzzle itself has ancient origins in Egypt, though many stories have been told of its use throughout the Middle East. According to one, a person would give the ring to their lover telling them of the magical qualities of the ring. If the ring was ever worn while being amorous to another, it would tell the giver of the ring about the unfaithfulness of the ring wearer. The wearer, wishing to avoid the watchfulness of the ring, might remove it from their finger so that it wouldn't witness an unfaithful act. But once removed, the ring would fall into its separate pieces and so the giver would know their lover had been cheating.

pyramid

A **polyhedron** whose base is a **polygon** and whose other faces are triangles that meet at a common **vertex**, sometimes called the *apex*. A right pyramid has its apex directly above the center of its base. A square-based pyramid, like the pyramids in Egypt, has a square base and four triangular sides. A triangular pyramid, or **tetrahedron**, is made from four triangular sides; if regular, it is one of the **Platonic solid**s. The volume of a pyramid is $\frac{1}{3}A_bh$, where A_b is the area of the base and h is the perpendicular height of the apex above the base; the surface area is ps, where p is the base perimeter and s is the slant height.

Pyramidology, which (in its original form) claims links between biblical prophesy and the construction of the Egyptian pyramids, stemmed from the publication in 1859 of *The Great Pyramid: Why Was It Built? And Who Built It?* by

pyramid The Great Pyramid of Cheops at Giza.

John Taylor, a British amateur mathematician and astronomer who was editor of the *London Observer.* Taylor became convinced that the Great Pyramid of Cheops, in its architectural proportions, embodied various remarkable and deeply meaningful geometric and mathematical properties. Chief among these, Taylor noted, was that the ratio of the perimeter of the pyramid's base to twice its height closely approximates the universal constant **pi**. The said ratio of the Great Pyramid gives a better value for pi than any found in written Egyptian records. The **Rhind papyrus** contains several problems that involve a multistep method for finding the area of a circle from its diameter. This method implies a value of π equal to $^{256}/_{81}$, or 3.1605, which is less than 1% larger than the true value of 3.14159 . . . . The values Taylor used for the Great Pyramid's base and height gave him a value of π correct to two decimal places. Further analysis of the pyramid's dimensions and various manipulations of numbers led him to conclude that the builders of the edifice had used a unit that he called "pyramid inch" (equal to about 1.01 times a standard inch). Twenty-five pyramid inches made a "pyramid cubit," and 10 million pyramid cubits, Taylor pointed out, approximate the length of Earth's polar radius. These and a series of similar calculations provided what Taylor considered to be adequate evidence that the Great Pyramid had been built as a model of Earth. Taylor's fantastic claims may never have become popular had not Charles Piazzi Smyth, the Astronomer Royal of Scotland taken up the cause of pyramidology. He popularized it in Britain, the rest of Europe, and the United States, through a number of books including *Our Inheritance in the Great Pyramid* (1864) and *Life and Work at the Great Pyramid* (1867). (In the churchyard in Sharow, North Yorkshire, Smyth has a pyramidal grave marker.) Others who espoused pyramid numerology have included Helena Blavatsky, the theosophist; Charles Taze Russell, founder of the Watchtower Bible and Tract Society; and Edgar Cayce, the American psychic; not to mention a variety of more recent pseudohistorians who have linked the Great Pyramid to everything from lost civilizations to UFOs. Yet Taylor himself was well aware of other, less dramatic explanations for the pyramid's dimensions, including the possibility that the pyramid had been constructed so that the area of one of its faces would equal the square of its height. The mathematical sophistication required to achieve this is not great and would have resulted in a ratio of the perimeter of the base to twice the height of 3.145–about as close as the approximation of pi used by Smyth in his studies. So, the ratio could have occurred as a completely coincidental byproduct of a design that had nothing to do with the ratio of the circumference of a circle to its diameter. The pyramid is a rich source of the kind of data that Taylor and Smyth worked with, and it would be surprising if they'd been unable to come up with some interesting number combinations given the extent of their efforts.

pyramidal number

The number of dots that may be arranged in a **pyramid** with a regular **polygon** as a base.

Pythagoras of Samos (c. 580–500 B.C.)

A Greek philosopher and mathematician, a native of the Aegean island of Samos, and the founder of a secretive pseudoreligious community in Croton, southern Italy. Pythagoras left no writings and virtually nothing is known about him as an individual, so it is almost impossible to disentangle the beliefs and discoveries of the "Pythagoreans" from those of their leader. To the Pythagoreans, "everything is number" and every number was supposed to be a quantity that could be expressed as the ratio of two integers. We now call such a number a **rational number**. The Pythagoreans used music as an example of the perfection and harmony of numbers that can be expressed as ratios. They showed that pitch could be represented as a simple ratio that came from the length of equally tight strings that could be plucked. Perhaps the most famous of the Pythagoreans' mathematical results is **Pythagoras's theorem**. Then the sky fell in on the Pythagoreans' worldview. Using their very own theorems they showed that not all numbers are rational. Their discovery that the **square root of 2** (the length of the hypotenuse of a triangle with sides 1 and 1) can't be expressed as ratio of two whole numbers was to have been kept a closely guarded a secret, but was later revealed by one of the cult's members.

Pythagoras's lute

The **kite**-shaped figure that forms the enclosing shape for a progression of diminishing **penta**gons and pentagrams, linking the vertices together. The resulting diagram is replete with lines in the **golden ratio**.

Pythagoras's theorem

The square of the length of the **hypotenuse** of a right **triangle** is the sum of the squares of the lengths of the two sides. This is usually expressed as $a^2 + b^2 = c^2$. See also **Pythagorean triplet**.

Pythagorean square puzzle

A deceptively hard assembly puzzle in which a small square piece must be combined with four pieces forming a larger square to make an even larger square.

Pythagorean triangle

A right triangle whose sides are integers. A *primitive Pythagorean triangle* is one whose sides are **coprime**.

Pythagorean triplet

Also called a *Pythagorean triple,* a set of three whole numbers that satisfies **Pythagoras's theorem**, that is, the squares of two of the numbers add up to the square of the third number. Examples include (3, 4, 5), (5, 12, 13), and (7, 24, 25). These are called *primitive triplets* because they have no common divisors. If the members of a primitive triplet are multiplied by the same integer, the result is a new (but not primitive) triplet. In any primitive Pythagorean triplet, one, and only one, of the three numbers must be even (but can't equal 2); the other two numbers are **coprime**. There are infinitely many such triplets, and they are easy to generate using a classic formula, known since ancient times. If the numbers in the triplet are a, b, and c, then: $a = n^2 - m^2$, $b = 2mn$, $c = m^2 + n^2$, where m and n are two integers and m is less than n. Because the **square root of 2** is irrational, there can't be any Pythagorean triplets (a, a, c). However, there are an infinite number of triplets (a, $a + 1$, c), the first three of which (apart from the trivial (0, 1, 1)) are (3, 4, 5), (20, 21, 29), and (119, 120, 169).

There are also an infinite number of Pythagorean quartets (a, b, c, d) such that $a^2 + b^2 + c^2 = d^2$. This is simply the three-dimensional form of Pythagoras's theorem and can be interpreted as the fact that the point in three-dimensions with Cartesian coordinates (a, b, c) lies an integer distance d from the origin. A formula that generates Pythagorean quartets is: $a = m^2$, $b = 2mn$, $c = 2n^2$, $d = (m^2 + 2n^2) = a + c$. Also note that $b^2 = 2ac$. When $m = 1$ and $n = 1$, we get the quartet (1, 2, 2, 3)–the simplest example.

Although there are an infinite number of Pythagorean triplets, **Fermat's last theorem,** which is now know to be true, ensures that there are no triplets for higher powers. See also **Euler's conjecture** and **multigrade**.

PUZZLE

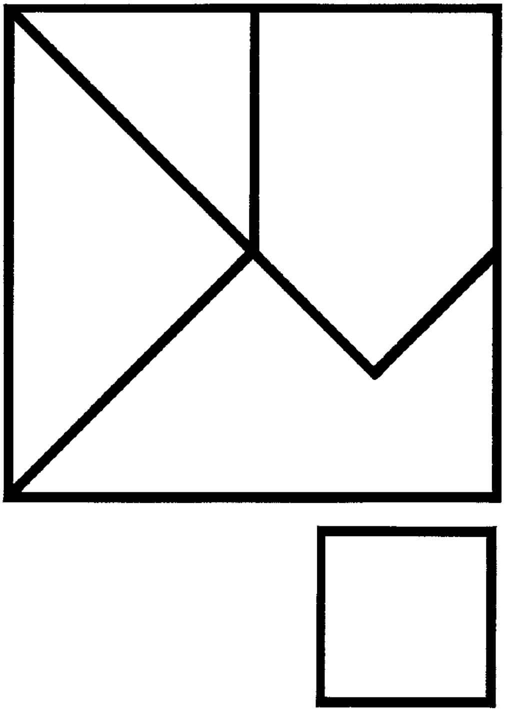

Rearrange the pieces in the Pythagorean square puzzle, including the small square, to make a single larger square.

Solutions begin on page 369.

Q

QED
Abbreviation for *quod erat demonstrandum* ("that which was to be shown"), used to denote the end of a proof.

quadrangle
A plane figure consisting of four points, each of which is joined to two other points by a line segment. A quadrangle may be **concave** or **convex** depending on whether the line segments do or don't intersect. A convex quadrangle is a **quadrilateral**. The word is from the Latin *quadrangulum* for "four-cornered" and is also used to describe a rectangular area surrounded on all four sides by buildings, or to such buildings themselves.

quadrant
Any one of the four portions of the plane into which the plane is divided by the Cartesian coordinate axes.

quadratic
An expression or an equation that contains the variable squared, but not raised to any higher power. For instance a *quadratic equation* in x contains x^2 but not x^3. Similarly a *quadratic expression,* or a *quadratic form,* contains its variable(s) squared but not raised to any higher power. If there is more than one variable (say, x and y), quadratic can mean that they are multiplied together in pairs (xy) but not in threes (such as x^2y). The graph of a quadratic equation is known as a *quadratic curve;* the curve of the general quadratic equation $y = ax^2 + bx + c$ is a **parabola**.

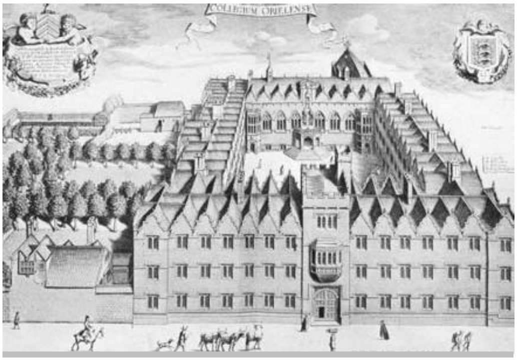

quadrangle The sixteenth-century quadrangle of Oxford University's Oriel College. *Oriel College, Oxford*

quadratrix of Hippias
The first curve in recorded history that was not part of a line or a circle, and the first curve known that is not **constructible** in the classical sense; in other words, it can't be drawn using a straightedge and a compass alone, but instead has to be plotted point by point. The quadratrix can be thought of as the intersection of two lines moving with constant velocity: the first line rotates (e.g., counterclockwise) while the second line moves along (say, in the direction of the positive y-axis). It has the Cartesian equation $y = x\ cot(\pi x/2a)$. The quadratrix was discovered by **Hippias of Elis** in about 430 B.C. and was used by him in his work on **trisecting an angle** and **squaring the circle**. In fact, its name refers to its use in turning curvilinear space into a rectangular area.

quadrature
The determination of the area of a geometric figure.

quadric
A surface in three dimensions that is described by equations containing the squares of x, y, and z, but no higher powers of them. Examples of such surfaces include the **sphere**, **ellipsoid**, **cone**, and **cylinder**.

quadrifolium
See **rose curve**.

quadrilateral
A **polygon** that has four sides and four vertices (corners). Quadrilaterals, and polygons in general, may be **convex** or **concave**. A convex quadrilateral may be further classified as a **trapezoid** or a British **trapezium** (one pair of opposite sides are parallel), a trapezium (no sides parallel); an isosceles trapezoid (United States) or an isosceles trapezium (United Kingdom) (two of the opposite sides parallel, the two other sides equal, and the two ends of each parallel side of equal angles); a **parallelogram** (opposite sides are parallel); a **kite** (two

adjacent sides of equal length, the other two sides of equal length); a **rhombus** (four sides of equal length); a **rectangle** (each angle is a right angle); or a **square** (four sides of equal length, each angle a right angle). *Quadrangular prisms* and *quadrangular pyramids* are ones whose bases are quadrilateral.

quantifier

In symbolic logic, the *universal quantifier* ∀ indicates "for every" or "for all." For example, ∀ *x A, p*(*x*) means for all *x* belonging to *A,* the **proposition** *p*(*x*) is valid. The *existential quantifier* ∃ indicates "there exists." So, for instance, ∃ *x A, p*(*x*) means there exists at least one *x,* belonging in *A,* for which the proposition *p*(*x*) is valid.

quantum field theory

The study of force fields such as the electromagnetic field in the context of **quantum mechanics**, and often special **relativity theory**. The mainstay of modern high-energy physics.

quantum mechanics

The science and mathematics that describe the behavior of nature at the atomic and subatomic level. At the heart of quantum mechanics are two basic concepts: (1) that every small bit of matter or energy can behave as if it were either a particle or a wave; and (2) that certain combinations of properties such as position and velocity, and energy and time, can't be known with arbitrary precision. The latter idea is encapsulated in *Heisenberg's uncertainty principle.* See also **many worlds hypothesis**.

quartic

A **polynomial** or polynomial equation that contains the fourth power of the variable, but no higher power. Many famous curves are described by such equations, including the **bicorn**, **Cartesian oval**, **conchoid**, **deltoid**, **devil's curve**, **folium**, **kampyle of Eudoxus**, and **limacon of Pascal**.

quartile

The first quartile of a **sequence** of numbers is the number such that one quarter of the numbers in the sequence are less than this number.

quasicrystal

A strange type of solid whose atomic structure is very regular but never quite repeats. Quasicrystalline structures don't have a simple unit cell that can be repeated periodically in all directions to fill space, although they do have local patterns that repeat almost periodically. They also have local rotational symmetries, such as those of a **pentagon**, that can't exist in ordinary crystals. Prior to the discovery of quasicrystals, it was thought that five-fold crystal symmetry was impossible, because there are no space-filling periodic tilings of this kind. The best known examples of quasicrystals resemble **Penrose tiling**s, which use repeated copies of two different rhombi to cover an infinite plane in intricate, interlocking patterns. In fact, some quasicrystals can be sliced in such a way that the atoms on the surface follow the exact pattern of the Penrose tiling.[293]

quasiperiodic

Refers to a form of motion that is regular but never exactly repeating. Quasiperiodic motion is always composed of multiple but simpler **periodic** motions. In the general case for motion that is the sum of simpler periodic motions, if there exists a length of time that evenly divides the frequencies of the underlying motions, then the composite motion will also be periodic; however, if no such length of time exists, then the motion will be quasiperiodic.

quasiregular polyhedron

A **polyhedron** that consists of two sets of regular **polygon**s, *m*-sided and *n*-sided respectively, and is constructed so that each polygon in one set is surrounded by members of the other set. There are three convex quasiregular solids: the *cuboctahedron* ($m = 3$, $n = 4$), the *icosidodecahedron* ($m = 3$, $n = 5$), and the **octahedron** ($m = n = 3$). In each case four faces meet at each vertex in the cyclic order (m, n, m, n). Because of this, these polyhedra have some special properties, one of which is that their edges form a system of **great circle**s. The edges of the octahedron form three squares; the edges of the cuboctahedron form four hexagons, and the edges of the icosidodecahedron form six decagons. Among the nonconvex polyhedra are two examples of type (m, n, m, n): the *dodecadodecahedron* ($m = 5$, $n = 5/2$) and the *great icosidodecahedron* ($m = 3$, $n = 5/2$), which can be made by truncating the **Kepler-Poinsot polyhedra** at their edge midpoints. There are also three nonconvex examples of type (m, n, m, n, m, n): the *small triambic icosidodecahedron* ($m = 3$, $n = 5/2$), the *triambic dodecadodecahedron* ($m = 5/3$, $n = 5$), and the *great triambic icosidodecahedron* ($m = 3$, $n = 5$). Finally, there is a group of nine *hemihedra,* in which some faces pass through the polyhedron's center. These hemifaces each cut a sphere into two hemispheres.

quaternion

An ordered **set** of four numbers. Quaternions, first introduced by William **Hamilton**, can also be written in the

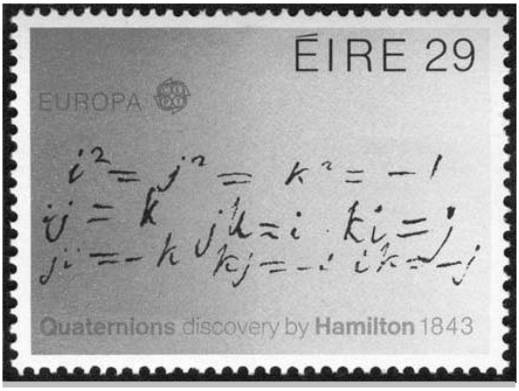
ÉIRE 29
EUROPA
$i^2 = j^2 = k^2 = -1$
$ij = k \quad jk = i \quad ki = j$
$ji = -k \quad kj = -i \quad ik = -j$
Quaternions discovery by Hamilton 1843

quaternion An Irish stamp showing the quaternion equations in Hamilton's own hand.

form $a + bi + cj + dk$, where a, b, c, and d are real numbers and i, j, and k are imaginary numbers, which is similar to that of **complex numbers**. Whereas complex numbers can be represented by points of a two-dimensional plane, quaternions can be viewed as points in the **fourth dimension**. For a while, quaternions were very influential: they were taught in many mathematics departments in the United States in the late 1800s, and were a mandatory topic of study at Dublin, where Hamilton ran the observatory. But then they were driven out by the **vector** notation of William Gibbs and Oliver Heaviside. Had quaternions come along later, when theoretical physicists were trying to understand patterns among subatomic particles, they may have found a place in modern science; after all, the unit quaternions form the group SU(2), which is perfect for studying spin-½ particles. But the way things turned out, quaternions had fallen from favor by the twentieth century and Wolfgang Pauli used 2 × 2 complex matrices instead to describe the generators of SU(2).

queens puzzle

A famous **chess** problem that asks in how many ways eight queens can be placed on a chessboard so that no two attack each other. The generalized problem, to find how many ways n queens can be placed on an $n \times n$ board so that no two attack each other, was first posed by Franz Nauck in 1850. In 1874 Günther and Glaisher described methods for solving this problem based on **determinants**. The number of distinct solutions, not counting rotations and reflections, for board sizes ranging from 1 × 1 to 10 × 10 is 1, 0, 0, 1, 2, 1, 6, 12, 46, and 92, respectively. The 6 × 6 puzzle, for which there is a solitary unique solution, was sold for one penny in Victorian London in the form of a wooden board with 36 holes into which pins were placed.

quine

> *Before we put the motion "that the motion be now put," should we not first put the motion "that the motion 'that the motion be now put' be now put?"*
>
> –Chairman of the meeting of the Society of Logicians

A term named by Douglas **Hofstadter** after the Harvard logician Willard van Orman Quine. It can be used either as a noun or a verb. (1) Quine (noun). A computer program that produces an exact copy of itself (or, alternatively, that prints its own listing.) This means that when the program is run, it must duplicate (or print out) precisely those instructions that the programmer wrote as part of the program, including the instructions that do the copying (or printing) and the data used in the copying (or printing.) A respectable quine–one that doesn't cheat–is not allowed to do anything as underhand or trivial as seeking the source file on the disk, opening it, and copying (or printing) its contents. Although writing a quine is not always easy, and in fact may seem impossible, it can always be done in any programming language that is Turing complete (see **Turing machine**), which includes every programming language actually in use. (2) Quine (verb). To write a sentence fragment a first time, and then to write it a second time, but with quotation marks around it. For example, if we quine "say," we get "say 'say' "). Thus, if we quine "quine", we get "quine 'quine,' " so that the sentence "quine 'quine' " is a quine. In this linguistic analogy, the verb "to quine," plays the role of the code, and "quine" in quotation marks plays the role of the data.

quintic

A **polynomial** or polynomial equation that contains the fifth power of the variable, but no higher power. Niels **Abel** and Evariste **Galois** independently proved that although there exist formulas for the general solution of **quadratic**, **cubic**, and **quartic** equations, no such formula exists for quintic equations.

quipu

A recording device generally associated with the Incas, who ruled Peru before the Spanish conquest. Quipu consisted of a number of color-coded cords; knots of various kinds were tied on these cords to represent a variety of information. An important use of quipu was to record numbers for use in trade, keeping accounts, and calendars, but the knotted strings might also have served as mnemonics for important historical events, astronomical data, and mythology (the Incas had no written language). Some evidence supports the idea that the knots and cords followed in a decimal system. Similar devices were used by several other Indian tribes and also described in Chinese and Persian documents from the fifth and sixth centuries B.C., and are still used by shepherds in the Andes for keeping accounts of their herds.

quotient

The number of times that one number can be divided exactly into another.

R

radian

A unit of angular measurement such that there are 2π radians in a complete circle. One radian = $180/\pi$ degrees. One radian is approximately 57.3°.

radical

The symbol that indicates a **root**, $^n\sqrt{}$. It seems to have been first used in 1525 by Christoff Rudolff (1499–1545) in his *Die Coss*.

radical axis

The **locus** of points of equal power with respect to two **circles**. The *radical center* of three circles is the common point of intersection of the radical axes of each pair of circles.

radius

The distance from the center of a **circle** to its circumference, or from the center of a regular **polygon** to any one of its vertices. The *radius of curvature, r,* at any point of a curve is $r = 1/\kappa$, where κ is the **curvature**.

radix

See **base**.

railroad problems

See **shunting puzzles**.

Ramanujan, Srinivasa Aaiyangar (1887–1920)

An extraordinary, largely self-taught, Indian mathematician who, in the most unorthodox way, made significant contributions to **number theory**, including the subject of **elliptic functions**, **continued fractions**, and **infinite series**. During much of his early work, he was unaware that he was rediscovering results that had taken other mathematicians centuries to achieve. But even in cases where he arrived at conclusions already known, he'd often travel an original route, and, in many cases, almost purely by intuition. Ramanujan was employed in a lowly clerk's position in Madras, when, in 1913, he wrote letters to three eminent mathematicians in England describing some of his results. Two of the three letters were returned unopened. However, G. H. **Hardy** recognized Ramanujan's abilities and arranged for him to come to Cambridge. Because of his lack of formal training, Ramanujan sometimes failed to distinguish between formal proof and apparent truth based on intuition or numerical evidence. His extraordinary innate familiarity with numbers was revealed by an incident recalled by Hardy:[150] "I remember once going to see him when he was lying ill at Putney. I had ridden in taxi cab number 1729 and remarked that the number seemed to me rather a dull one, and that I hoped it was not an unfavorable omen. 'No,' he replied, 'it is a very interesting number; it is the smallest number expressible as the sum of two cubes in two different ways [$1729 = 1^3 + 12^3 = 9^3 + 10^3$].' "

Unfortunately, Ramanujan's health deteriorated rapidly in England, perhaps due to the unfamiliar climate and food, and to the isolation which Ramanujan felt as the sole Indian and a devout Hindu in a culture which was alien to him. Ramanujan was sent home to recuperate in 1919, but tragically died the following year at the age of only 32. Although he published some of his results in journals, much of his work and conclusions have only

SRINIVASA RAMANUJAN
1887–1920
न.पै.
15 nP
INDIA POSTAGE
INDIA SECURITY PRESS

Ramanujan, Srinivasa Aaiyangar The enigmatic mathematician on an Indian commemorative stamp.

come to light more recently, from a scrutiny of his disorganized but fascinating notebooks.[184]

Ramsey theory

A branch of mathematics that asks questions such as: Can order always be found in what appears to be disorder? If so, how much can be found and how big a chunk of disorder is needed to find a particular amount of order in it? Ramsey theory is named after the English mathematician Frank P. Ramsey (1904–1930) who started the field in 1928 while wrestling with a problem in logic. (Frank's one-year-younger brother, Arthur, served as Archbishop of Canterbury from 1961 to 1974.) His life was cut short at the age of 26, following a bout of jaundice. Ramsey suspected that if a system was big enough, even if it seemed to be disorderly to an arbitrary degree, it was bound to contain pockets of order from which information about the system could be gleaned.

random

Without cause; not **compressible**; obeying the statistics of a fair coin toss.

random number

A number generated by a process that is fundamentally nondeterministic and unpredictable. Computer-generated "random numbers," which are calculated through a deterministic process, cannot, by definition, be **random**. Given knowledge of the **algorithm** used to create the numbers and its internal state, it's possible to predict all of the numbers returned by subsequent calls to the algorithm. For this reason the numbers produced by computer-based "random number generators" are often referred to as *pseudorandom numbers*. In the case of genuinely random numbers, knowledge of one number or an arbitrarily long sequence of numbers offers no clue of the next number to be generated. Humans are among the worst random number gen-

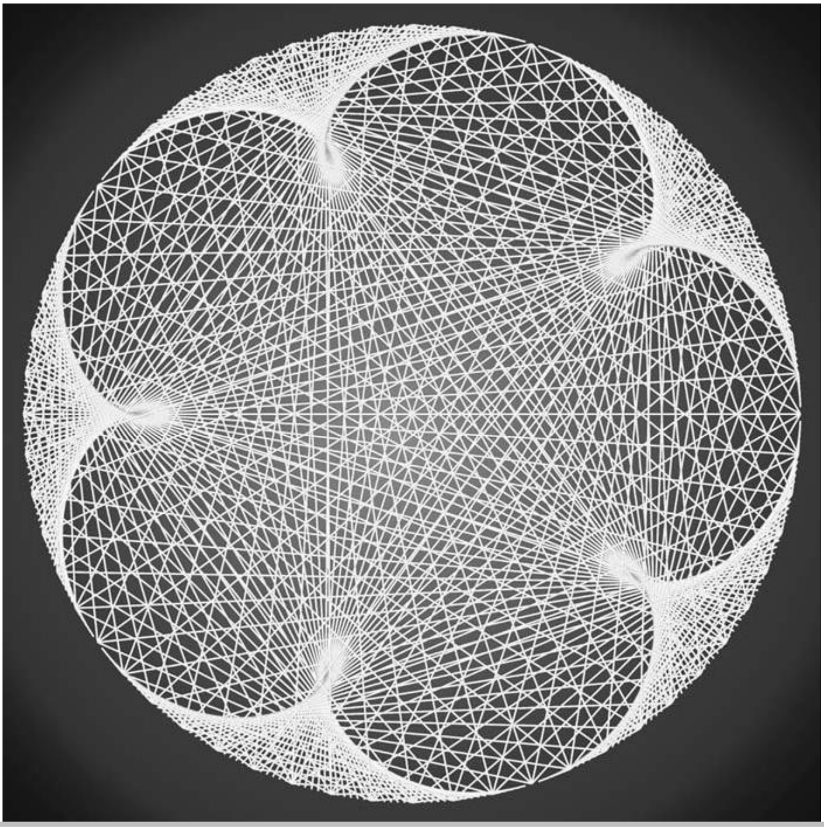

ranunculoid A ranunculoid curve spun by thread on a computer loom. *Jos Leys, www.josleys.com* erators. Ask someone to pick a number "at random" between 1 and 20, and the number they're most likely to choose is 17. *Psychologically random numbers,* such as 17, are usually odd and don't end in 5, so that they frequently tend to be **prime numbers**. See also **Chaitin's constant**.

random walk

A process in which the position of a particle changes by discrete steps of fixed length, and the direction of each step is chosen randomly. Random walks have interesting mathematical properties that vary greatly depending on the number of dimensions in which the walk takes place and whether it is confined to a **lattice**. For a random walk in one dimension there are only two directions to choose from. Imagine a drunken person wandering on the **number line** who starts at 0, and then moves left or right (+/−1) with probability ½. The probability that the walker will eventually return to his starting point is 1; in other words, it is certain to happen. The same is true for a random walk in the plane, moving on the integer lattice points, with probability ¼ in each of the coordinate directions: the probability of ending up back at the starting point is 1. However, the situation changes in three dimensions. Suppose a drunken fly moves randomly from one point to another in a three-dimensional lattice with a probability of 1 in 6 of arriving at any of the six adjacent lattice points on each move. No matter how long the fly roams, it has only a 0.34054 . . . probability of ever getting back to where it started. Probabilists say that random walks on the line and plane are *recurrent,* whereas random walks in three dimensions or more are *transient.* Effectively, this is because there is so much more "space" in three or more dimensions. The numbers giving the probability of eventually returning to the starting point are known as *random walk constants.* The random thermal perturbations in a liquid are responsible for a random walk phenomenon known as *Brownian motion,* and the collisions of molecules in a gas are a random walk responsible for *diffusion.*

range

(1) The set of possible values in which a **function**'s output can be. See also **codomain**. (2) The set of all points on a **line segment**.

rank

(1) Any of the rows of squares running crosswise to the files on a playing board in **chess** or checkers. (2) The rank of a **matrix** is equal to the dimension of the largest submatrix that can be obtained by deleting rows and columns of the parent matrix and that has a nonzero determinant. See also **tensor**.

ranunculoid

An **epicycloid** with five cusps ($n = 5$), named after the buttercup genus *Ranunculus.*

ratio

A **rational number** of the form a/b where a is called the *numerator* and b is called the *denominator.* It may be written with a colon (:), as a fraction, or with the word *to.*

rational number

A number that can be written as an ordinary **fraction**—a ratio, a/b, of two integers, a and b, where b isn't zero—or as a decimal expansion that either stops (like 4.58) or is periodic (like 1.315315 . . .). Other examples include 1, 1.2, 385.66, and ⅓. Rational numbers are countable, which means that, although there are infinitely many of them, they can always be put in a definite order, from smallest to largest, and can thus be counted. They also form what's called a *densely ordered set;* in other words, between any two rationals there always sits another one—in fact infinitely many others. The rational numbers are a subset of the **real numbers**; real numbers that aren't rational are called, rationally enough, **irrational numbers**. Although rationals are dense on the real **number line**, in the sense that any **open** set contains a rational, they're pretty sparse in comparison with the irrationals. One way to think of this is that the **infinity** of rationals (which, strangely enough, is exactly the same size as the infinity of whole numbers) is smaller than the infinity of irrational numbers. Another way to grasp the scarcity versus density issue, is to realize that the rationals can be covered by a set whose "length" is arbitrarily small. In other words, given a string of any positive length, no matter how short, it will still be long enough to cover all the rationals. In mathematical parlance, the rationals are a **measure** *zero* set. The irrationals, by contrast, are a measure one set. This difference in measure means that the rationals and irrationals are quite different even though a rational can always be found between any two irrationals, and an irrational exists between any two rationals.

raven paradox

A **paradox** put forward by the German logician Carl Hempel (1905–1997) in the 1940s to highlight a situation where the logic of **induction** seems to fly in the face of intuition. According to the *principle of induction,* the more that a theory is supported by observation, the greater the probability that the theory is true. Consider, said Hempel, the theory that all ravens are black. After each observation of a black raven, our belief in the theory "all ravens are black" increases. But here's the rub. The statement "all ravens are black" is logically equivalent to the statement "all nonblack things are nonravens." The observation of a white swan is consistent with this statement. A white swan is a nonblack thing, and when we examine it, we observe that it is a nonraven. So by the principle of induction, observing a white swan should increase our belief that all ravens are black!

Many solutions have been offered to this enigma. The American logician Nelson Goodman (1906–1998) suggested imposing restrictions to our reasoning, such as never considering an instance as support for "All P are Q" if it would also support "No P are Q." Others have questioned the *principle of equivalence*. Perhaps seeing a white swan should strengthen our belief in the theory "all nonblack things are nonravens," without increasing our conviction that "all ravens are black." Yet others have argued that our intuition is flawed. Observing a white swan really does increase the probability that all ravens are black! After all, if you were shown all the nonblack things in existence, and you noticed that none was a raven, then you could properly conclude that all ravens were black. The example only seems counterintuitive because the set of nonblack things is vastly larger than the set of ravens. Thus observing one more nonblack thing that is not a raven should make a tiny difference to our degree of belief in the proposition compared to the difference made by observing one more raven that is black.

A way to sidestep the paradox is by using **Bayes's theorem**. According to this the probability of a hypothesis H must be multiplied by the ratio:

$$\frac{\text{probability of observing } X \text{ if } H \text{ is true}}{\text{probability of observing } X}$$

If a swan is picked at random, the probability of it being white is independent of the colors of ravens. The numerator in the above ratio will equal the denominator, the ratio will equal one, and the probability will remain unchanged. Seeing a white swan doesn't affect our belief about whether all ravens are black. If a nonblack thing is chosen at random, and a white swan is shown, then the numerator will be bigger than the denominator by a tiny amount. Seeing the white swan will only slightly increase our belief that all ravens are black. We'd have to see almost every nonblack thing in the universe (and see that they're all nonravens) before our belief in "all ravens are black" would increase appreciably. In both cases, these results are in line with intuition.

ray

A straight path of points that begins at one point and continues in one direction.

real number

Any number that can be represented as a decimal, possibly infinitely long and nonrepeating. Real numbers stand in one-to-one correspondence with the points on a continuous line, known as the *real number line,* that stretches from zero to infinity in both directions. The set of real numbers contains the set of all **rational number**s and the set of all **irrational number**s. The name "real number" is a retronym, coined by René **Descartes** in response to the concept of **imaginary number**s. Number systems that are even more general than the real numbers include the **complex numbers** and, of much more recent discovery, **hyperreal number**s and **surreal number**s.

realm

A term advocated for a three-dimensional version of the two-dimensional **plane**.

reciprocal

One over a given number; for example, the reciprocal of 4 is $\frac{1}{4}$.

Recorde, Robert (c. 1510–1558)

A Welsh physician and mathematician, born in Tenby, Pembrokeshire, and trained at Oxford and Cambridge, who held various positions, including master of the mint in Bristol and later in Ireland, and wrote a number of influential math textbooks. These books formed a complete course and were written in English, rather than the usual Latin or Greek, so that they could be read by anyone. In one of them Recorde introduces the "=" sign for "equals."

rectangle

A **quadrilateral** whose interior angles are all 90°. If all of its sides are the same length, it is a **square**. The smallest square that can be cut into $m \times n$ rectangles, such that all m and n are different integers, is the 11×11 square, and the **tiling** uses five rectangles. The smallest rectangle that can be cut into $m \times n$ rectangles, such that all m and n are different integers, is the 9×13 rectangle; this tiling also uses five rectangles.

rectangular coordinates

See **Cartesian coordinates**.

rectangular hyperbola

See **hyperbola**.

recursion

See **recursion**.

No, seriously. Recursion is a process that wraps back on itself and feeds the output of a process or function back in as the input. Using some sort of *recurrence relation,* an entire class of objects can be built up from a few initial values and a small number of rules. The **Fibonacci sequence**, for example, is defined recursively, as are many **fractal** figures. *Self-recursion,* of which the first two lines of this entry are an example, leads to an endless feedback loop. One, somewhat disturbing, notion of reality is that we live in a *recursive universe* in which nature resembles an infinite nest of Russian dolls. One day, we will simulate the big bang in one of our supercomputers. Inside this artificial universe (rather like an immense *Star Trek* "holodeck" simulation) will evolve new star systems and new life forms. They too will evolve intelligence one day and invent computers. And they too may one day simulate the big bang inside one of their supercomputers. This chain of existence will continue as long as we continue to run our simulation. Our universe will continue to exist as long as our parent universes' continue to run their simulations. And therein lies madness.

recursive function

Strictly speaking, a **function** that is *computable;* however, in the usual sense of the word, a function is said to be recursive if its definition makes reference to itself (see **recursion**). For example, **factorial** can be defined as $x! = x(x - 1)!$ with the base case of 1! equal to 1. See also **self-referential sentence**.

recursively enumerable set

A potentially infinite **set** whose members can be enumerated by a **universal computer**; however, a universal computer may not be able to determine that something is not a member of a recursively enumerable set. The *halting set,* a concept related to the **halting problem**, is recursively enumerable but not recursive.

reductio ad absurdum

"Reduction to the absurd"; the process of demonstrating that an idea is probably false by first assuming its truth, and then showing how that truth leads to conclusions that can't possibly be true. In *A Mathematician's Apology* (1941),[151] G. H. **Hardy** said: "Reductio ad absurdum, which Euclid loved so much, is one of a mathematician's finest weapons. It is a far finer gambit than any chess play: a chess player may offer the sacrifice of a pawn or even a piece, but a mathematician offers the game."

reductionism

The idea that nature can be understood by taking it apart. In other words, knowing the lowest-level details of how things work (at, say, the level of subatomic physics) reveals how higher-level phenomena come about. This is a bottom-up way of looking at the universe, and is the exact opposite of **holism**.

redundancy

The existence of repetitive patterns or structures. In an important sense, redundancy refers to *order* in a **complex system** since order is defined as the existence of structures that maintain themselves over time. In **information theory**, redundancy refers to repetition in patterns of messages in a communication channel. If the message contains these redundancies, they can be compressed further; for example, a message containing a series of 250 ones, could be compressed into a command that effectively says "and then repeat one 250 times," instead of writing out all 250 ones.

reentrant angle

An inward-pointing angle of a **concave** polygon.

reflection

A way of transforming a shape in the same way that a mirror does. The reflection of a shape in a mirror line is an identical shape that has been flipped over. When an object is placed a certain distance in front of a mirror, its image in the mirror appears the same distance away from the edge of the mirror. Likewise, all the points on a shape and all the points on its image are the same distance away from the mirror line.

reflex angle

An angle between 180° and 360°.

reflexible

Having a plane of mirror symmetry. Compare with **chiral**.

regular polygon

A **polygon** in which all the sides are equal and all the angles are equal.

regular polyhedron

A polyhedron in which every face and vertex figure is regular. There are nine regular polyhedra: the five **Platonic solid**s and the four **Kepler-Poinsot solid**s. However, others are sometimes allowed, depending on the definition of polyhedron.

relativity theory

The physical and mathematical theory due to Albert Einstein (1879–1955) that revolutionized our understanding of **space**, **time**, and gravity. In it, space and time are seen as a unified and inseparable whole–the four-dimensional continuum of **space-time**. The curvature of space-time due to the presence of matter becomes the extraordinary new explanation and interpretation of gravity. Einstein's theory of relativity was published by him in two great parts. Special relativity, published in 1905, deals exclusively with *inertial* frames of reference, that is, reference frames that don't accelerate with respect to one another. Its two central premises are that the laws of physics are the same in all reference frames and that the speed of light (in a vacuum) is constant in all reference frames. General relativity, published in 1915, centers on the *equivalence principle,* the idea that acceleration and gravity are equivalent. See also **non-Euclidean space**.

renormalization

A mathematical technique for looking at a physical system at different levels of magnification.

repdigit

A number composed of repetition of a single digit in a given base, generally taken as base 10 unless otherwise specified. For example, the **beast number** 666 is a (base-10) repdigit.

representation theory

A theory that seeks to understand an abstract algebraic system such as a **group** by obtaining it in a more concrete way as a *permutation group* or as a group of matrixes (see **matrix**).

rep-tile

A repetitive **tiling**: a shape with the property that it tiles a larger version of itself, using identical copies of itself. A simple example is a square because four copies of any square tile a larger square. Any triangle also is a rep-tile, because four copies of it tile a larger version of this triangle. Rep-tiles that require *n* tiles to build a larger version of themselves are said to be rep-*n;* thus a square is rep-4. Since any of these larger replicas can be combined to give an even larger, second-generation copy, a rep-*n* tile is also rep-n^2, rep-n^3, and so on. Often tiles have several rep-numbers. If a tile is rep-*n* and rep-*m,* it is also rep-*mn,* since replicas can be built with *n* tiles, then combined, *m* at a time, to give a yet larger version.

The set of rep-tiles is a subset of the set of *irreptiles.* An irreptile is any shape that tiles a larger version of itself using either differently sized or identical copies of itself. The problem to find all irreptiles in the Euclidean plane has been studied but not yet completely solved. A

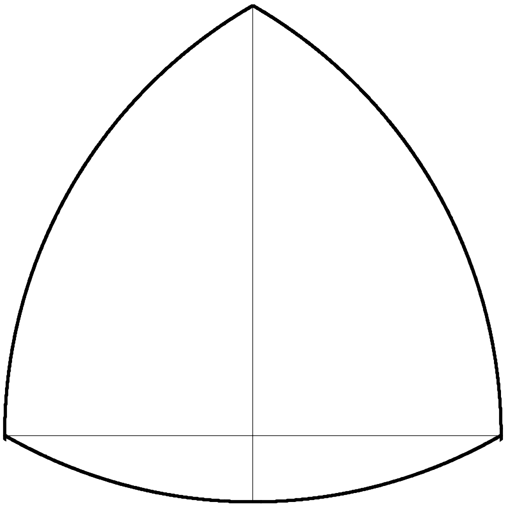

Reuleaux triangle © Jan Wassenaar, www.2dcurves.com

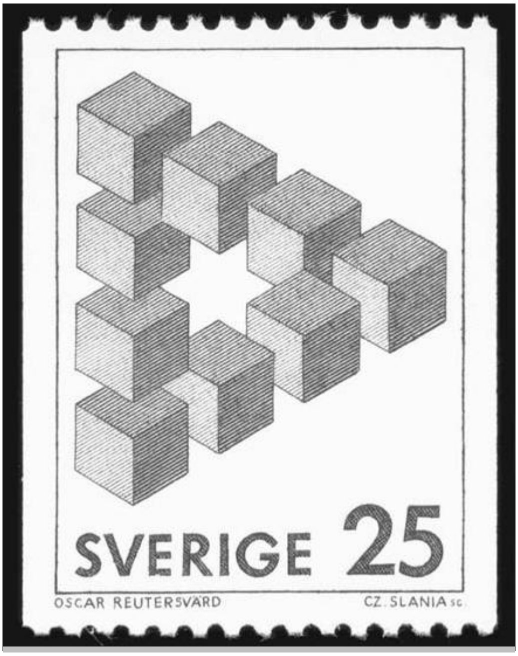

Reutersvärd, Oscar A Swedish stamp depicting one of Reutersvärd's impossible figures.

related set of problems is to find for each irreptile the minimum number of smaller copies needed to tile the original shape; in many cases it is difficult to prove such a minimality. The name "rep-tile" was coined by Simon **Golomb**.

rep-unit

A number whose digits are all units; the rep-unit (repeated unit) with n digits is denoted R_n. For example, $R_1 = 1$, $R_2 = 11$, $R_3 = 111$, and $R_n = (10^n - 1)/9$. R_n divides R_m whenever n divides m. No rep-unit can be a square, but it is not known if one can be a cube. *Rep-unit primes* are rep-units that are **prime numbers**. The only known rep-unit primes are R_2 (11), R_{19}, R_{23}, R_{317}, and R_{1031}, though R_{49081} and R_{86453} are suspected primes.

resultant

A **vector** that is the sum of a given set of vectors.

Reuleaux polytope

A **convex** body in the plane or in higher dimensions that, like the **Reuleaux triangle**, consists of pieces of round spheres, each centered at one of the corners of the convex body.

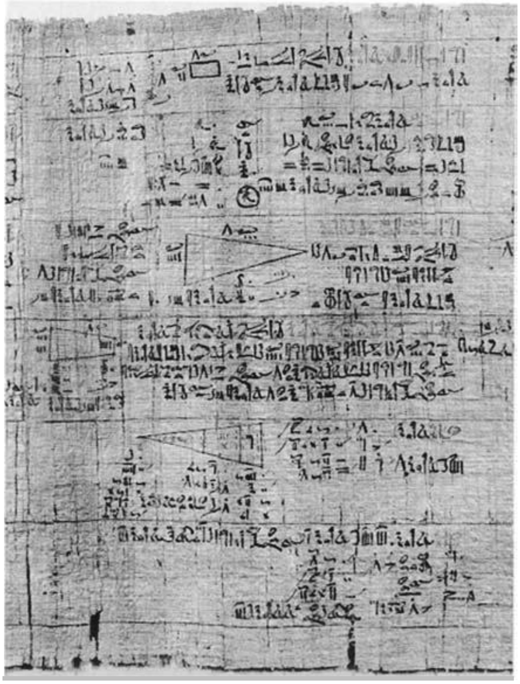

Rhind papyrus *British Museum*

Reuleaux triangle

The simplest noncircular **curve of constant width**; also known as the *Reuleaux wheel*, it is named after the German engineer and mathematician Franz Reuleaux (1829–1905). Although it was known to earlier mathematicians, Reuleaux was the first to show its constant-width properties. To form a Reuleaux triangle, take the three points at the corners of an equilateral triangle and connect each pair of points by a circular arc centered at the remaining point. The ratio of the circumference to the width of the triangle is, remarkably, **pi**. By rotating the centroid of a Reuleaux triangle appropriately, the figure can be made to trace out a square, perfect except for slightly rounded corners. This idea has formed the basis of a drill that will carve out squares, first patented by Harry Watts in 1914. Bits for square, pentagonal, hexagonal, and octagonal holes are still sold by the Watts Brothers Tool Works in Wilmerding, Pennsylvania. The actual drill bit for the square is a Reuleaux triangle made concave in three spots to allow for unobstructed corner-cutting and the discharge of shavings. The Reuleaux triangle may also form the shape of the piston in a rotary, or Wankel, engine, in which gasoline burns in crescent-shaped chambers, turning a rotating piston that drives an axle through its center.

Reutersvärd, Oscar (1915–)

A Swedish artist who pioneered the creation and design of **impossible figures**. His work in this area goes back to one day in 1934 when, as a young student in Stockholm, he started doodling in the margins of a textbook during a long lecture. Reutersvärd's doodle began with an outline of a perfect six-pointed star. Once the star was complete he added cubes around the star, nestled into the spaces between the points. He soon realized that what he'd drawn was paradoxical: something that couldn't be built in the real world. A different version of this figure, independently created by Roger **Penrose**, would later be called the **Penrose triangle**. Thus began a lifelong fascination with such objects, which later included work on the impossible staircase (a design he sketched in 1950 while on a cross-country train ride), which is known as the **Penrose stairway**, and the discovery of the **tribar illusion**. In the early 1980s the Swedish government honored Reutersvärd's achievements by issuing a set of three stamps depicting impossible figures, including a version of his 1934 weird cubes design.

Rhind papyrus

A papyrus scroll, 33 cm high and 565 cm wide, found in a tomb in Thebes, that is the most valuable source of information we have about Egyptian mathematics. The scroll was bought at a market in Luxor in 1858 by a 25-year-old Scotsman, Henry Rhind, who went to Egypt for health reasons and became interested in archeology. After his early death at the age of 30, the scroll found its way to the British Museum in London in 1864 and has remained there ever since. It is often referred to as the *Rhind mathematical papyrus*, or RMP for short. The hieroglyphs on the papyrus were deciphered in 1842, while the Babylonian clay-tablet cuneiform writing was deciphered later in the nineteenth century. The text begins by stating that the scribe "Ahmes" is writing it (in about 1600 B.C., and he is thus the earliest named individual in the history of mathematics) but that he has copied it from "ancient writings," which probably go back to at least 2000 B.C. Although there is some strictly practical mathematics on the papyrus, including calculations needed for surveying, building, and accounting, some of which involve **Egyptian fractions**, many of the problems in the RMP take the form of arithmetic puzzles. One of these is: Seven houses contain seven cats. Each cat kills seven mice. Each mouse had eaten seven ears of grain. Each ear of grain would have produced seven hekats of wheat. What is the total of all of these? This is very similar to the **St. Ives problem**.

There are also four lesser documents preserving Egyptian arithmetic: the Moscow papyrus and the Berlin papyrus (named for the places they are kept), the Kahun papyrus (named for where it was found), and the Leather Roll (named for its composition). The Moscow papyrus is sometimes called the Golenischev papyrus after the Russian V. S. Golenischev, who purchased it in 1893 from two Egyptian brothers who found it in a tomb at Deir el-Bahri. It measures 8 cm high and 540 cm wide and contains 25 problems and their solutions. The most unusual are the tenth, which seems to give the area of the surface of a hemisphere or perhaps a cylinder, and the fourteenth, which gives the formula for the volume of the frustum of a **pyramid**.

rhombus

A **quadrilateral** in which both pairs of opposite sides are parallel and all sides are the same length, that is, an equilateral **parallelogram**. A rhombus is also sometimes called a *rhomb* or a diamond. A rhombus whose acute angles are 45° is called a *lozenge*. The diagonals p and q of a rhombus are perpendicular and satisfy the relationship $p^2 + q^2 = 4a^2$. The area of a rhombus is given by $A = \frac{1}{2}pq$.

Richard's paradox

See **Berry's paradox**.

Riemann, (Georg Friedrich) Bernhard (1826–1866)

A German mathematician who was the first person to provide a thorough treatment of **non-Euclidean geometry** and to see how it might be applied in physics; he thus helped pave the way for the general **relativity theory**. Among several profound aspects of mathematics now named after him are the **Riemann hypothesis** and the related **Riemann zeta function**. His father, a Lutheran pastor, encouraged him to study theology at Göttingen. But even as a child Riemann had shown a tremendous aptitude for mathematics and, in 1847, he persuaded his father to let him go to Berlin to learn mathematics from the likes of Karl **Jacobi**, Peter **Dirichlet**, and Jakob **Steiner**. Two years later, he returned to Göttingen to study for his Ph.D. and begin his climb up the professorial ladder. In 1854, his inaugural lecture, "Concerning the hypotheses which underlie geometry," covered a breathtaking array of topics, including a workable definition of the curvature of space and how it could be measured, the first description of **elliptical geometry**, and, most important of all, the extension of geometry into more than three dimensions with the aid of algebra.

Riemann hypothesis

> *If I were to awaken after having slept for a thousand years, my first question would be: Has the Riemann Hypothesis been proven?*
>
> –David Hilbert

The most important open question in **number theory** and, possibly, in the whole of mathematics. A $1 million prize has been offered by the Clay Mathematics Institute for a proof. The hypothesis was first formulated by Bernhard **Riemann** in 1859, was included in David **Hilbert**'s list of challenging problems for twentieth-century mathematicians, and is widely believed to be true. Yet a proof remains tantalizingly out of reach. What the Riemann hypothesis says is that the nontrivial zeros of the **Riemann zeta function** all have real parts equal to ½. In plain language, the hypothesis asserts that there is an underlying order, akin to musical harmonics, in the way **prime numbers** are distributed. It's known that for any given number n there are approximately $n/\log n$ prime numbers that are less than n. The formula is not exact: sometimes it is a little high and sometimes it is a little low. Riemann looked at these deviations and found that they contain periodicities. His hypothesis quantifies and formalizes this discovery, positing that the zeros of the zeta function can be regarded as the harmonic frequencies in the distribution of primes. If the Riemann hypothesis turns out to be true, what do these harmonics in the "music" of the primes mean? Remarkably, it is been found by the English physicist Michael Berry and his colleagues that there is a deep connection between the harmonics–the Riemann zeros–and the allowable energy states of physical systems that are on the border between the quantum world (see **quantum**

mechanics) and the everyday world of classical physics. The Riemann harmonics, or "magic numbers," behave exactly like the energy levels in quantum systems that classically would be chaotic. This deep connection between number theory and the physics of the real universe, if upheld, is utterly astonishing. If the Riemann hypothesis is proved true, it could open an entirely new window on the nature of reality and the relationship between the abstract world of mathematics and the behavior of matter and energy. On the other hand, if it is disproven, there will be an even deeper mystery to explore: How can the Riemann zeta function so convincingly mimic a quantum system without actually being one?[40]

Riemann integral

The kind of **integral** familiar from **calculus** texts and normally used by scientists and engineers, in which the process of **integration** of function on an interval amounts to finding the area under the curve.

Riemann sphere

A topological sphere consisting of the **complex plane** and the point at infinity; an example of a **Riemann surface**.

Riemann surface

Also known as a *complex curve,* a complex **manifold** with one complex dimension. The Riemann surface is a conformal structure (see **conformal mapping**).

Riemann zeta function $\zeta(s)$

> *We may–paraphrasing the famous sentence of George Orwell–say that "all mathematics is beautiful, yet some is more beautiful than the other." But the most beautiful in all mathematics is the zeta function. There is no doubt about it.*
>
> –Krzysztof Maslanka, Polish cosmologist

One of the most profound and mysterious objects in modern mathematics; from it has sprung the **Riemann hypothesis** and all that this conjecture seems to imply. The Riemann zeta function is closely tied to the distribution of **prime numbers**. It is an extension of the *Euler zeta function,* first studied by Leonhard **Euler**, which is the sum

$$\zeta(s) = 1 + 1/2^s + 1/3^s + 1/4^s + \ldots = \sum_{n=1}^{\infty} 1/n^s.$$

Euler found that this function is linked to the occurrence of prime numbers by the following fundamental relationship:

$$\zeta(s) = 1 + 1/2^s + 1/3^s + 1/4^s + \ldots = 2^n/(2^n - 1)$$
$$\times\, 3^n/(3^n - 1) \times 5^n/(5^n - 1) \times 7^n/(7^n - 1) \times \ldots$$
$$= \prod_p 1/(1 - p^{-s})$$

The Riemann zeta function extends the definition of Euler's zeta function to all **complex numbers**.

Riemannian geometry

See **elliptical geometry**.

right

A *right angle* is an angle of 90°. A *right triangle* is a triangle that contains a right angle.

ring

(1) Another name for an **annulus**. (2) A number system in which addition, subtraction, and multiplication are always defined and the **associative** and **distributive** laws are valid. Compare with **field**.

Rithmomachia

A medieval, chesslike board game for two players that is based on the number theories of Pythagoras and Boethius. Rithmomachia or "battle of numbers" (*rithmo,* "arithmetic, numbers"; *machia* "battle") dates back to about A.D. 1150 although the earliest publication of the rules was by Jean de Boissiere in the sixteenth century. Used as an educational tool (the only game allowed in medieval schools and universities) and as an intellectual exercise, it enjoyed a last wave of popularity during the Renaissance before the early Scientific Revolution led to its disappearance.

The game is played on an 8 × 16 board. Each player starts with either 24 black or white pieces: 8 circles, 8 triangles, 7 squares, and 1 pyramid. Each of these has a *number,* which is how many places it can move. A square can move 4 spaces, a triangle 3, a circle 1, and the pyramid as many spaces as the player chooses. Opponent pieces can be captured in a variety of ways: *siege capture* (surrounding the opponent piece on all four sides; *meeting capture* (attacking a piece with the same type of piece); *assault capture* (achieved if the piece's number times the number of spaces it moved lands it next to an opposing piece that equals the product); *ambuscade* (achieved if two pieces of a player that are on either sides of a piece sum to equal the opponent's piece). There are also a number of different ways to win, including: *de corpore* (players agree on a number of pieces to be captured); *de bonis* (players agree on a number value target); *de lite* (the winner is determined by the sum of the pieces as well as the number of digits on all those pieces); *victoria magna* (if there is a common difference between the pieces a player has captured, or the squares of three consecutive integers are captured; or there is a difference of 2, 4, and 6 in even pieces, or 3, 5, and 7 in odd pieces).

river-crossing problem

A puzzle in which a variety of objects and living things, some of them mutually incompatible, must be conveyed in small groups from one side of a river to another without any loss along the way. The earliest known examples are in *Propositiones ad Acuendos Juvenes* (Propositions for sharpening youths), which is generally attributed to Abbott **Alcuin**. They are: the problem of three jealous husbands (each of whom won't let another man be alone with his wife), the problem of the two adults and two children where the children weigh half as much as the adults (and the boat has a limited weight capacity), and the problem of the wolf, the goat, and the cabbage. In the last case, the difficulty is that only one item can be ferried across at once but, if left unattended, the sheep will eat the cabbage and the wolf will eat the sheep. The solution, which involves a stratagem common to all these types of problems, is to bring back to the starting bank of the river an item that has already been taken across. In this case, the sheep must be taken across first, followed by either the cabbage or the wolf, but then the sheep must be brought back before the next item is taken across to avoid the sheep becoming either a diner or a dinner.

These medieval puzzles were considered and elaborated on by Niccoló **Tartaglia**, Luca **Pacioli**, and Claude-Gaspar **Bachet**, and even more so by later mathematicians such as Edouard **Lucas** and Gaston **Tarry**. Ways of complicating river-crossing problems include adding more people and objects, using a bigger boat, and inserting an island in the river. The reader may care to try the problem of the missionaries and the cannibals.

PUZZLE

Three missionaries and three cannibals must cross a river in a boat that holds a maximum of two people. If the cannibals outnumber the missionaries, on either side of the river, the missionaries are in trouble. Each missionary and each cannibal is capable of rowing the boat. How can all six get across the river safely?

Solutions begin on page 369.

Robinson, Abraham (1918–1974)

A German-born mathematician who founded **nonstandard analysis** and did important work in a wide diversity of fields from aerodynamics to mathematical logic. A modern counterpart to Gottfried **Leibniz** in his range of interests and the significance of his research on **infinitesimal**s, Robinson (he changed his name from Robinsohn) taught at various universities in Israel, England, Canada, and the United States.

Rolle's theorem

Suppose a continuous function (see **continuity**) crosses the *x*-axis at two points *a* and *b* and is differentiable at all points between *a* and *b;* that is, it has a **tangent** at all points on the curve between *a* and *b.* Then there's at least one point between *a* and *b* where the **derivative** is 0, and the tangent is parallel to the *x*-axis.

Roman numerals

A **number system** in which each symbol represents a fixed value regardless of its position; this differs from the place-value system of **Arabic numeral**s. The earliest form of the Roman system was, however, decimal. In this primitive version a series of I's represented any number from 1 to 9, and a new symbol was introduced for each higher power of 10: X for 10, C for 100, and M for 1,000. The symbols V, L, and D, which stand for 5, 50, and 500, are thought to have been introduced by the Etruscans. A common remark is that multiplication and division using Roman numerals is so awkward that it is totally impractical. However, an article by James G. Kennedy in *The American Mathematical Monthly* in 1980 gives algorithms for these operations that are actually more straightforward in the Roman system than in the Arabic. In multiplication the first step is to rewrite the numbers in a simple place-value notation. Seven columns are set up, headed by the symbols M, D, C, L, X, V, and I, and tallies are marked in each column corresponding to the number of times that symbol appears in the multiplicand. For example, if the multiplicand is XIII (13), one tally is marked in the X column and three tallies are marked in the I column. The multiplier is written in the same way. The multiplication itself is done by forming partial products according to two simple rules. In most cases the partial product given by any one tally in the multiplier is simply the set of tallies that represents the multiplicand, shifted to the left an appropriate number of columns. If the multiplier digit is I, the multiplicand is not shifted at all; the multiplicand is shifted one place to the left for V, two places for X, three places for L, and so on. The second rule is applied only when one Etruscan character is multiplied by another. In such cases the tallies representing the multiplicand digit are written twice in the appropriately shifted column and an additional tally is written one column to the right. Once a partial product has been formed for every tally in the multiplier, the tallies in each column are accumulated and replaced by the Roman symbol at the head of the column, giving the final answer. Only a slight change in the method is needed for Roman numerals in "subtractive notation," where 10 is written as IX, and so on. If all this sounds not quite so simple, the method for multiplying Arabic numbers is just as

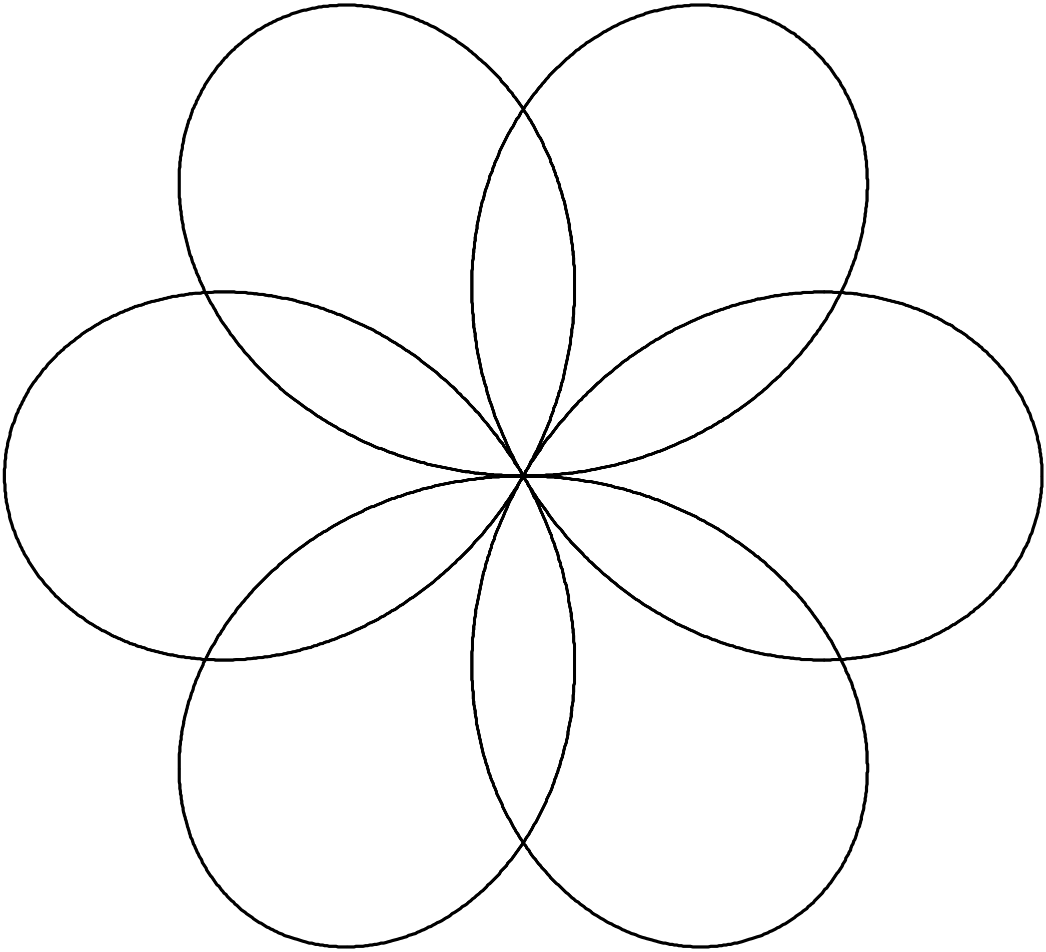

rose curve *© Jan Wassenaar, www.2dcurves.com*

involved if they are written in explicit form. Furthermore, Arabic operations require a multiplication table giving the 100 products of all the possible pairs of Arabic digits. No comparable table is needed with Roman numerals, where all arithmetical operations can be defined in terms of shifting rules, addition, and subtraction.

rooks problem

To find the maximum number of rooks that can be placed on an $n \times n$ **chess**board such that no rook attacks another. Since each rook attacks all squares in the rank and file upon which it rests, this number is *n:* the rooks may be placed along the main diagonal. The total number of ways of placing *n* nonattacking rooks is *n* factorial (*n*!).

root

(1) A number used to build up another number by repeated multiplication. For example, since $2 \times 2 \times 2 = 8$, two is said to be the third root or cube root of eight. (2) A solution of an equation. For example, 3 is a root of the

Rubik's cube *Peter Knoppers, www.buttonius.com*

equation $x^2 = 9$. A root is also called a *zero of a function* because it is a value that will make the function zero ($x = 3$ will make the function $f(x) = x^2 - 9$ zero. The word comes from the Indo-European *werad*, which originally meant the roots of a plant but was later generalized to mean the origins or beginnings of something, whether it was physical or mental.

root of unity

A solution of the equation $x^n = 1$, where n is a positive integer.

rope around the earth puzzle

Imagine a rope that fits snugly all the way around Earth like a ring on a person's finger. Now imagine the rope is made just 1 meter longer and lifted uniformly off the surface until it is once again taught. What will its height be above the surface? This puzzle, or one very like it, appeared in a students' book on Euclid written in 1702 by the English clergyman, mathematician, and natural philosopher William Whiston (1667–1752). The answer in the form just given is remarkable: about 16 cm. It comes simply from the formula for the circumference of a circle. If the extra radius of the rope is r and Earth's radius is R, then

$$2\pi(R + r) = 2\pi R + 100$$
$$\text{so that } r = 100/2\pi \cong 15.9.$$

Rosamund's bower

A legendary **maze**, located in Woodstock Park, Oxfordshire, whose purported site is marked today by a well and fountain. It was supposedly intended to conceal Rosamund Clifford, the mistress of King Henry II (1133–1189), from the queen, Eleanor of Aquitaine. Legend has it that about 1176, Eleanor managed to solve the maze and confronted Rosamund with the choice of a dagger or poison; she drank the poison and Henry never smiled again. Historically, Henry had imprisoned Eleanor for fomenting rebellion by her sons, and Rosamund was his acknowledged mistress. Rosamund probably spent her last days at a nunnery in Godstow, near Oxford. The legend of the bower dates from the fourteenth century and her murder is a later addition. In the nineteenth century, many puzzle collections had a maze called Rosamund's bower.

rose curve

A curve that has the shape of a flower with petals; it was named *rhodonea* ("rose") by the Italian mathematician Guido Grandi in the 1720s. It is given by the polar equation

$$r = a \sin (n\theta).$$

If n is odd the rose has n petals, if n is even the rose has $2n$ petals, and if $n = 2$ the rose becomes the **quadrifolium**. If n is an *irrational* **number**, then there are an infinite number of petals.

rotation

A **transformation** in which a figure turns through a specific angle about a fixed point, called the center of rotation. The center of rotation may be inside or outside the figure that is being transformed. If the figure is turned anticlockwise, the rotation is considered positive, while a negative rotation turns the figure clockwise.

rotor

A **convex** figure that can be rotated inside a **polygon** (or **polyhedron**) while always touching every side (or face). The least area rotor in a square is the **Reuleaux triangle**. The least area rotor in an equilateral triangle is a lens with two 60° arcs of circles and a radius equal to the triangle's altitude. There exist nonspherical rotors for the tetrahedron, octahedron, and cube, but not for the dodecahedron and icosahedron. See also **curve of constant width**.

roulette

(1) The curve traced by a fixed point on a closed **convex** curve as that curve rolls without slipping along a second curve. (2) A gambling game in which players bet on which slot of a rotating disk a small ball will come to rest in. On August 18, 1913, on an unbiased roulette wheel at Monte Carlo, evens came up 26 times in a row. The probability of this occurring is 1 in 136,823,184.

round

In **topology**, the terms ***circle*** and ***sphere*** refer to topological objects and not geometric ones, so that the surface of an egg shape is a sphere. A *round sphere* is, topologically speaking, not a tautology, but a sphere with constant **curvature**; that is, a sphere in the sense of geometry.

rounding

Replacing a number by another number having fewer **significant digits** or, for integer numbers, fewer value-carrying (nonzero digits). For example, 386.804 may be rounded successively to 386.80, 386.8, 387, 390, and 400. Rounding may be carried out in two ways: by *rounding down*, which is equivalent to truncation, and by *rounding up* the last digit to be retained by one unit. See also **banker's rounding**.

round-off error

The error accumulated during a calculation due to rounding intermediate results. See also **banker's rounding**.

Rubik's cube

A 3 × 3 × 3 cube in which the 26 subcubes on the outside are internally hinged in such a way that rotation (by a quarter turn in either direction or a half turn) is possible in any plane of cubes. Each of the six sides is painted a distinct color, and the goal of the puzzle is to return the cube to a state in which each side has a single color after it has been randomized by repeated rotations. Invented in 1974 by the Hungarian Ernö Rubik, patented in 1975, and put on the market in Hungary in 1977, it went on to sell some 100 million copies worldwide over the next decade. Since there are over 43 million trillion different arrangements of the small cubes, only one of which corresponds to the desired goal, to solve Rubik's cube in a time significantly less than the current age of the universe (let alone the world record, which stands at around 20 seconds) calls for some kind of methodical approach. Algorithms exist for solving a cube from an arbitrary initial position, but they are not necessarily optimal (i.e., requiring a minimum number of turns).[270, 297]

Rucker, Rudy (Rudolf von Bitter) (1946–)

An American mathematician best known for his entertaining popular mathematics, science, and science fiction books, including *Infinity and the Mind,*[275] *The Fourth Dimension,*[273] and *Mind Tools.*[274] Rucker has a doctorate in mathematical logic from Rutgers and teaches at San José State University. His great-great-great-grandfather was the famous German philosopher Georg Hegel.

ruled surface

A surface that is built up from an infinite number of perfectly straight lines. A **cylinder**, for example, is a ruled surface of parallel straight lines. A **cone** is a ruled surface of straight lines that meet at the apex of the cone. Also known as *scrolls,* ruled surfaces have been studied for centuries by geometers such as the Jesuits Roger Boscovich and Andre Tacquet as well as by their famous students, including Gaspar **Monge** and Phillippe de Lahire. Examples that stand out because of both their striking shape and their relative ease of construction include **hyperboloid**s, **helicoid**s and **Möbius band**s. Most ruled surfaces, however, are so complicated that, before the computer age, they were almost impossible to construct.

ruler-and-compass construction

See **Mascheroni construction**.

Russell, Bertrand Arthur William (1872–1970)

A British philosopher, mathematician, and logician who rose to prominence with his first major work, *The Principles of Mathematics* (1902), in which he attempted to remove mathematics from the realm of abstract philosophical notions and to give it a precise scientific framework. Russell then collaborated for eight years with the British philosopher and mathematician Alfred North Whitehead (1861–1947) to produce the monumental work *Principia Mathematica* (3 volumes, 1910–1913). This work showed that mathematics can be stated in terms of the concepts of general **logic**, such as class and membership in a class. It became a masterpiece of rational thought. Russell and Whitehead proved that numbers can be defined as classes of a certain type, and in the process they developed logic concepts and a logic notation that established symbolic logic as an important specialization within the field of philosophy. In his next major work, *The Problems of Philosophy* (1912), Russell borrowed from the fields of sociology, psychology, physics, and mathematics to refute the tenets of idealism, the dominant philosophical school of the period, which held that all objects and experiences are the product of the intellect. Russell, a realist, believed that objects perceived by the senses have an inherent reality independent of the mind.

Russell condemned both sides in World War I, and for his uncompromising stand he was fined, imprisoned, and deprived of his teaching post at Cambridge. In prison he wrote *Introduction to Mathematical Philosophy* (1919). After the war he visited the Soviet Union, and in his book *Practice and Theory of Bolshevism* (1920) he expressed his disappointment with the form of socialism practiced there. He felt that the methods used to achieve a Communist system were intolerable and that the results obtained were not worth the price paid. Russell taught at Beijing University during 1921 and 1922, and in the United States from 1938 to 1944, though he was barred from teaching at the College of the City of New York (now City College of the City University of New York) by the state supreme court because of his attacks on religion and his advocacy of sexual freedom. Russell returned to England in 1944 and was reinstated as a fellow of Trinity College. Although he abandoned pacifism to support the Allied cause in World War II, he became an ardent opponent of nuclear weapons. Russell received the 1950 Nobel Prize for Literature and was cited as "the champion of humanity and freedom of thought." He led a movement in the late 1950s advocating unilateral nuclear disarmament by Britain, and at the age of 89 he was imprisoned after an antinuclear demonstration.[276]

Russell's paradox

A **paradox** uncovered by Bertrand **Russell** in 1901 that forced a reformulation of **set theory**. One version of Russell's paradox, known as the barber paradox, considers a town with a male barber who, every day, shaves every man who doesn't shave himself, and no one else. Does the barber shave himself? The scenario as described requires that the barber shave himself if and only if he does not! Russell's paradox, in its original form considers the set of all sets that aren't members of themselves. Most sets, it would seem, aren't members of themselves–for example, the set of elephants is not an elephant–and so could be said to be "run-of-the-mill." However, some "self-swallowing" sets *do* contain themselves as members, such as the set of all sets, or the set of all things except Julius Caesar, and so on. Clearly, every set is either run-of-the-mill or self-swallowing, and no set can be both. But then, asked Russell, what about the set S of all sets that aren't members of themselves? Somehow, S is neither a member of itself nor not a member of itself. Russell discovered this strange situation while studying a foundational work in symbolic logic by Gottlob **Frege**. After he described it, set theory had to be reformulated axiomatically in a way that avoided such problems. Russell himself, together with Alfred North Whitehead (1861–1947), developed a comprehensive system of types in *Principia Mathematica.* Although this system does avoid troublesome paradoxes and allows for the construction of all of mathematics, it never became widely accepted. Instead, the most common version of axiomatic set theory in use today is the *Zermelo-Fraenkel set theory,* which avoids the notion of types and restricts the universe of sets to those that can be built up from given sets using certain axioms. Russell's paradox underlies the proof of **Gödel's incompleteness theorem** as well as Alan **Turing**'s proof of the undecidability of the **halting problem**.

Russian multiplication

Multiplication by repeated doubling, also known as *peasant multiplication*. For example to multiply 17 by 13, double the 17 and halve the 13 rounding down to the next whole number where necessary; then add the doubles that correspond to an odd number in the other column.

So	17×13
doubled and halved	34×6
doubled and halved	68×3
doubled and halved	136×1

adding up the numbers in the first column that correspond to an odd number in the second $(17 + 68 + 136) = 221 = 17 \times 13$.

S

Saccheri, Giovanni Girolamo (1667–1733)

A Jesuit priest, philosopher, and mathematician who did early work on **non-Euclidean geometry**, although he didn't see it as such. His *Euclides ab Omni Naevo Vindicatus* (1733) was actually an attempt to prove Euclid's **parallel postulate** but ended up laying the groundwork for both **hyperbolic geometry** and **elliptical geometry**.

saddle

A type of surface that is neither a peak nor a valley but still has a zero gradient. Saddle points are situated such that moving in one direction takes one uphill, while moving in another direction would be downhill. A *saddle function* is a **function** $f(x, y)$ of two vectors x and y (which typically lie in different **vector space**s) that is **concave** up in x and concave down in y. See also **pseudosphere**.

St. Ives problem

A well-known and simple puzzle in arithmetic set by Mother Goose in the rhyme "As I Was Going to St. Ives":

> As I was going to St. Ives, I met a man with seven wives. Every wife had seven sacks, and every sack had seven cats, every cat had seven kittens. Kittens, cats, sacks, and wives, how many were going to St. Ives?

The various individuals and items make up a **geometric sequence**: 7 wives + 7^2 sacks + 7^3 cats + 7^4 kittens = 7 + 49 + 343 + 2,401. This gives a total of 2,800 (or 2,801 if the narrator also happens to be a wife). It's been said that the real answer to this rhyme is 1, since "I *met* a man with seven wives" means the 2,800 were going in the *opposite* direction!

A similar problem, and solution, is contained in the **Rhind papyrus** by Ahmose, written about 1650 B.C. (some of which is copied from an older document of about 1800 B.C.). Here the geometric series has one more power of 7: 7 houses + 49 cats + 343 mice + 2,401 ears of grain + 16,807 hekats (a measure) of grains, giving a total of 19,607.

St. Petersburg paradox

A strange state of affairs that arises from a game proposed by Nikolaus (I) Bernoulli (see **Bernoulli family**) in 1713. It is named after the fact that a treatise on the paradox was written by Nikolaus' cousin, Daniel, and published (1738) in the *Commentaries of the Imperial Academy of Science of St. Petersburg.* The game goes as follows. You toss a coin. If it shows heads, you win whatever is in the pot and the game is over. If it shows tails, the pot is doubled and you get to toss the coin again. If the coin shows heads on the first toss you win \$2; if it shows tails, you toss again. If the coin now shows heads you win \$4, and so on. After n tosses you get \$$2^n$ if heads appear for the first time. The only catch is you have to pay to play the game. How much should you be willing to pay? Classical decision theory says that you should be willing to pay any amount up to the expected prize, the value of which is obtained by multiplying all the possible prizes by the probability that they are obtained and adding the resulting numbers. The chance of winning \$2 is ½ (heads on the first toss); the chance of winning \$4 is ¼ (tails followed by heads); the chance of winning \$8 is ⅛ (tails followed by tails followed by heads); and so on. Since the expected payoff of each possible consequence is \$1 (\$2 × ½, \$4 × ¼, etc.) and there are an infinite number of them, the total expected payoff is an infinite sum of money. A rational gambler would enter a game if and only if the price of entry was less than the expected value. In the St. Petersburg game, any finite price of entry is smaller than the expected value of the game. Thus, the rational gambler would play no matter how large the entry price was! But there's clearly something wrong with this. Most people would offer between \$5 and \$20 on the grounds that the chance of winning more than \$4 is only 25% and the odds of winning a fortune are very small. And therein lies the paradox: If the expected payoff is infinite, why is no one willing to pay a huge amount to play?

The classical solution to this mystery, provided by Daniel Bernoulli and another Swiss mathematician, Gabriel Cremer, goes beyond probability theory to touch areas of psychology and economics. Bernoulli and Cremer pointed out that a given amount of money isn't always of the same use to its owner. For example, to a millionaire \$1 is nothing, whereas to a beggar it can mean not going hungry. In a similar way, the utility of \$2 million is not twice the utility of \$1 million. Thus, the important quantity in the St. Petersburg game is the *expected utility* of the game (the utility of the prize multiplied by its probability) which is far less than the *expected prize.* This explanation forms the theoretical basis of the insurance business. The existence of a **utility function** means that most people prefer, for example, having $98 in cash to gambling in a lottery where they could win $70 or $130 each with a chance of 50%, even though the lottery has the higher expected prize of $100. The difference of $2 is the premium most of us would be willing to pay for insurance. That many people pay for insurance to avoid any risk, yet at the same time spend money on lottery tickets in order to take a risk of a different kind, is another paradox, which is still waiting to be explained.

salient

At a *salient point,* two branches of a curve meet and stop, and have different **tangent**s. A *salient angle* is an outward-pointing angle of a **polygon**; compare with **reentrant angle**.

salinon

A figure formed from four connected semicircles. The word *salinon* is Greek for "salt cellar," which the figure resembles. In his *Book of Lemmas,* **Archimedes** proved that the salinon has an area equal to the circle having the line segment joining the top and bottom points as its diameter. See also **arbelos**.

Sallows, Lee C. F. (1944–)

A British electronics engineer and puzzle enthusiast at the University of Nijmegen in the Netherlands. Among his many accomplishments in recreational mathematics, he devised the first **self-enumerating sentence** (published in 1982), introduced **alphamagic square**s in 1986, coined the word ***golygon*** in 1990, invented reflexicons (minimal self-enumerating phrases) in 1992, demonstrated the parallelogram theorem for 3 × 3 **magic square**s in 1997, and discovered **geometric magic squares** in 2001.

scalar

A quantity specified by a single number or value (as opposed to a vector, matrix, or array) that contains multiple values. Examples of scalars include mass, volume, and temperature. A *scalar field* is an arrangement of scalar values distributed in a space.

scalene triangle

A **triangle** whose sides are all unequal.

schizophrenic number

An informal name for an **irrational number** that displays such persistent patterns in its decimal expansion, that it

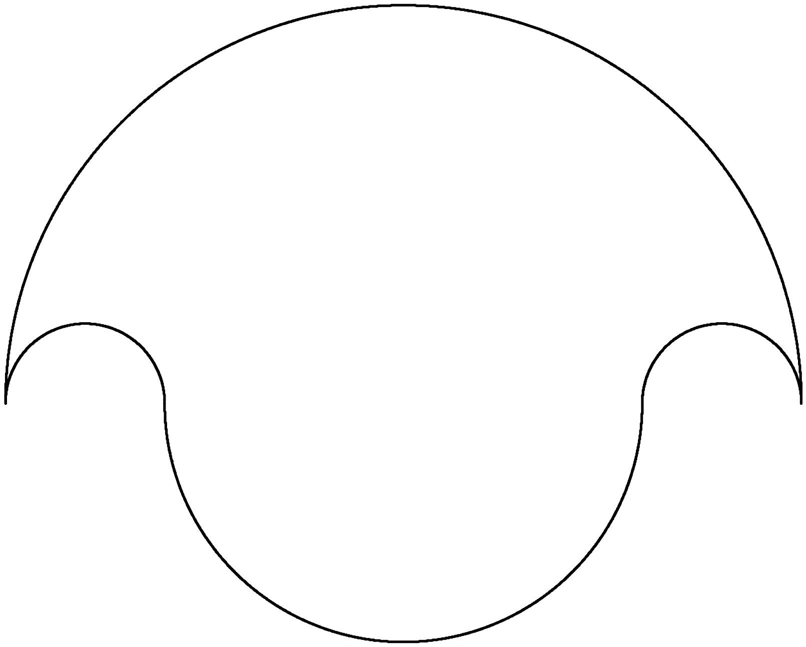

salinon *© Jan Wassenaar, www.2dcurves.com* has the appearance of a **rational number**. A schizophrenic number can be obtained as follows. For any positive integer n let $f(n)$ denote the integer given by the recurrence $f(n) = 10f(n-1) + n$ with the initial value $f(0) = 0$. Thus, $f(1) = 1$, $f(2) = 12$, $f(3) = 123$, and so on. The square roots of $f(n)$ for odd integers n give rise to a curious mixture appearing to be rational for periods, and then disintegrating into irrationality. This is illustrated by the first 500 digits of $\sqrt{f(49)}$:

1111111111111111111111111.1111111111111111111111 0860
555555555555555555555555555555555555555555555 2730541
66666666666666666666666666666666666666666 0296260347
2222222222222222222222222222222222222 0426563940928819
4444444444444444444444444444444 38775551250401171874
9999999999999999999999999999 808249687711486305338541
66666666666666666666666 5987185738621440638655598958
33333333333333333333 0843460407627608206940277099609374
99999999999999 0642227587555983066639430321587456597
222222222 186349201679118083308 1844 . . . .

The repeating strings become progressively shorter and the scrabbled strings become larger until eventually the repeating strings disappear. However, by increasing n we can forestall the disappearance of the repeating strings as long as we like. The repeating digits are always 1, 5, 6, 2, 4, 9, 6, 3, 9, 2, ...

Schläfli, Ludwig (1814–1895)

A German mathematician whose work centered on geometry, arithmetic, and the theory of functions. He made an important contribution to **non-Euclidean geometry** when he proposed that spherical three-dimensional space could be thought of as the surface of a **hypersphere** in Euclidean four-dimensional space. Schläfli started out as a schoolteacher and amateur mathematician. He was also an expert linguist and spoke many languages, including Sanskrit. In 1843 he served as a translator for the great mathematicians Jakob **Steiner**, Karl **Jacobi**, and Peter **Dirichlet** during their visit to Rome and learned a great deal from them. Ten years later he became professor of mathematics at Bern. However, his true importance was only appreciated following the publication of his magnum opus *Theory of Continuous Manifolds* in 1901, several years after his death.

Schläfli symbol

A notation, devised by Ludwig **Schläfli**, which describes the number of edges of each **polygon** meeting at a **vertex** of a regular or semi-regular **tessellation** or solid. For a **Platonic solid**, it is written $\{p, q\}$, where p is the number of sides each face has, and q is the number of faces that touch at each vertex.

schoolgirls problem

A problem in **combinatorics** posed by the Rev. Thomas **Kirkman** in a letter in 1850 following a paper he wrote on the same subject in 1847:[188] A school mistress has fifteen girl pupils and she wishes to take them on a daily walk. The girls are to walk in five rows of three girls each. It is required that no two girls should walk in the same row more than once per week. Can this be done?

In fact, provided n is divisible by 3, we can ask the more general question about n schoolgirls walking for $(n-1)/2$ days so that no girl walks with any other girl in the same triplet more than once. Solutions for $n = 9$, 15, and 27 were given in 1850 and much work was done on the problem thereafter. The general problem of how many triads can be made out of **n** symbols, so that no pair of symbols is comprised more than once among them gave rise to the study of Steiner triple systems. However, Jakob **Steiner** had little to do with them and they should rightfully be named after Kirkman. They are important in the modern theory of **combinatorics**.

Schröder's reversible staircase

A classic example of an **ambiguous figure**, first drawn by Schröder in 1858. Not to be confused with the **Penrose stairway**.

Schubert, Hermann Cäsar Hannibal (1848–1911)

A German mathematician who worked mainly in enumerative geometry–the parts of algebraic geometry that

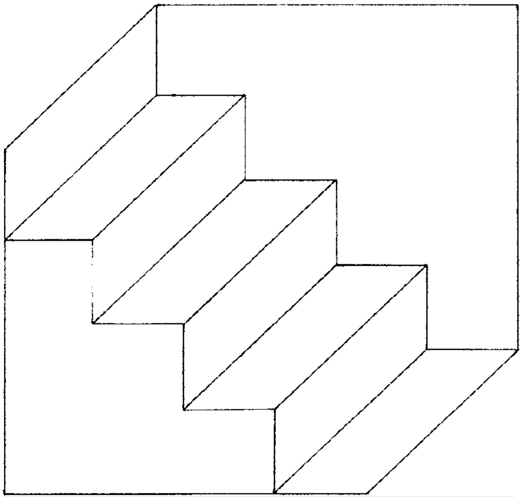

Schröder's reversible staircase

involve a finite number of solutions. He also wrote extensively on recreational math.

Schuh, Frederick (1875–1966)

A Dutch mathematician who wrote many textbooks and a number of books on recreational mathematics, including *The Master Book of Mathematical Recreations* (1943), the English edition of which appeared in 1968 (Dover Publications).[291] Schuh was professor of mathematics at the Technische Hoogeschool at Delft (1907–1909 and 1916–1945) and professor of mathematics at Groningen (1909–1916).

scientific notation

A number of the form $a \times 10^n$, where n is an integer, positive or negative, and a is a **real number** larger than or equal to 1, but less than 10. Scientific notation provides a compact way of writing **large numbers**.

scintillating grid illusion

See **Hermann grid illusion**.

score

A group of 20 items. The word comes from the Old Norse *skor* for a heavy mark used to indicate a string of 20 smaller marks; *skor,* in turn, is descended from the Indo-European *sker,* for "cutting" or "slicing." From about 1400, *score* was also the word for a record or an amount due–the total of the score marks on a tally. It became a common word for the total of a tradesman's or innkeeper's account. So, *to settle the score* originally meant just to pay one's bill. But it acquired the figurative sense of taking revenge on somebody, and that's usually what is meant by the expression now. The more general meaning of score, as a tally, is used daily when the results of sports competitions are reported.

search space

A characterization of every possible solution to a problem instance.

secant

A straight line that meets a curve in two or more points.

second

One-sixtieth of a **minute** in both time and angle. Literally, the second division of the hour or the circle, the minute being the first; from the Latin *secundus.* In the System International d'Unites (SI units) one second is defined as the duration of 9,192,631,770 periods of radiation corresponding to the transition between two hyperfine levels of cesium-133 in a ground state at a temperature of 0°K (Kelvin).

secretary problem

See **sultan's dowry**.

sector

Part of a **circle** bounded by two radii and the included **arc**.

segment

Part of a **circle** bounded by a **chord** and the **arc** subtending the chord.

self-enumerating sentence

Also known as an *autogram,* a **self-referential sentence** whose text consists solely of the enumeration of its letter content. The answer to the question whether such a sentence exists in English was given by Lee **Sallows**, and was first published in *Scientific American* in January 1982:

> Only the fool would take trouble to verify that his sentence was composed of ten a's, three b's, four c's, four d's, forty-six e's, sixteen f's, four g's, thirteen h's, fifteen i's, two k's, nine l's, four m's, twenty-five n's, twenty-four o's, five p's, sixteen r's, forty-one s's, thirty-seven t's, ten u's, eight v's, eight w's, four x's, eleven y's, twenty-seven commas, twenty-three apostrophes, seven hyphens and, last but not least, a single !

This remarkable sentence counts not only its own letters, but also its punctuation marks, although it fails to enumerate three letters of the alphabet (j, q, and z). Sallows went on to devise a "pangram machine"–a computer purposely built to search for sentences of this type. Among its many successes is:

> This pangram lists four a's, one b, one c, two d's, twenty-nine e's, eight f's, three g's, five h's, eleven i's, one j, one k, three l's, two m's, twenty-two n's, fifteen o's, two p's, one q, seven r's, twenty-six s's, nineteen t's, four u's, five v's, nine w's, two x's, four y's, and one z.

self-intersecting

A *self-intersecting polygon* is a **polygon** with edges that cross other edges. A *self-intersecting polyhedron* is a **polyhedron** with faces that cross other faces.

self-organization

A process in a **complex system** whereby new emergent structures, patterns, and properties arise without being externally imposed on the system. Not controlled by a centralized, hierarchical "command and control" center, self-organization is usually distributed throughout a system. It requires a complex, **nonlinear system** under appropriate conditions, variously described as "far-from-equilibrium," critical values of control parameters leading to "bifurcation," or the "edge of chaos." First investigated in the 1960s in physical systems by Ilya Prigogine and his followers, as well as the Synergetics School founded by Hermann Haken, self-organization is now studied mainly through computer simulations (see **cellular automaton**), Boolean networks, and other phenomena of **artificial life**. However, self-organization is now recognized as a crucial way of understanding emergent, collective behavior in a large variety of systems including the economy, the brain and nervous system, the immune system, and ecosystems. The buildup of system order via self-organization is now conceived as a primary tendency of complex systems in contrast to the past emphasis on the degrading of order in association with the principle of entropy (Second Law of Thermodynamics). However, rather than denying entropy, self-organization can be understood as a way that entropy increases in complex, nonlinear systems.

self-organized criticality

A mathematical theory that describes how systems composed of many interacting parts can tune themselves toward dynamical behavior that is critical in the sense that it is neither stable nor unstable but at a region near a **phase transition**. See also **edge of chaos** and **self-organization**.

self-referential sentence

A sentence that refers to itself and nothing else. Here are some examples:

> This statement is short.
>
> This sentence has five words.
>
> The last word of this sentence is "wrong."
>
> "Pentasyllabic" is pentasyllabic.
>
> How long is the answer to this question? Ten letters.

Some self-referential statements take the form of jokes. For example:

> The two rules for success are: Never tell them everything you know.
>
> There are three kinds of people in the world: those who can count and those who can't.

Finally, some take the form of maxims, as in this case by Thomas Macaulay (1800–1859):

> Nothing is so useless as a general maxim.

See also **Hofstadter's law**.

self-similarity

The property an object has when a part of itself looks the same or similar to the whole. Many objects in the real world, such as coastlines, are statistically self-similar: parts of them show the same statistical properties at many scales. Self-similarity is a defining characteristic of **fractal**s.

semigroup

A **set** together with a method of combining elements, such as addition or multiplication, to get new ones, which satisfies only some of the properties required to get a **group**. In particular, a semigroup need not have an **identity** element and elements need not have **inverse**s.

semi-magic square

A square array of *n* numbers such that sum of the *n* numbers in any row or column is a constant (known as the magic sum). See also **magic square**.

semi-regular polyhedron

A **polyhedron** that consists of two or more types of **regular polygon**s, all of whose vertices are identical. This category includes the **Archimedean solid**s, **prism**s and **antiprism**s, and the nonconvex uniform polyhedra (see **nonconvex uniform polyhedron**).

Senet

A popular two-player board game in ancient Egypt, enjoyed by both commoners and nobility, that may be an ancestor of modern **backgammon**. The rules are not known, though about 40 sets have been found in tombs, some in very good condition, together with paintings of games on tomb walls, dating back to the reign of Hesy (c. 2686–2613 B.C.). Senet, or the "game of passing," was played on a rectangular board consisting of three rows of 10 squares called "houses" that represented good or bad fortune. The board could be a grid drawn on a smooth surface or an elaborate box of wood and other precious materials. A perfectly preserved traveling version of Senet was found in Tutankhamen's tomb. The pieces, called *ibau* ("dancers" in Egyptian), varied in number from five to ten per player–five and seven being commonest. Cone-shaped pieces were pitted against reel-shaped pieces. The object was to get one's pieces on the board, then around the board in an S-shaped pattern, and finally off again at the far end. Strategy was mixed with chance (as it is in backgammon), introduced by the throw of four, two-sided sticks (as depicted in the Hesy painting) or, in later times, of knucklebones. Later depictions of the game, in the New Kingdom period, often showed just one player in competition–the opponent being a spirit from the afterlife. This has been interpreted as a change

Senet A modern version of the ancient Egyptian board game. *Fundex Games Ltd.*

in the significance of Senet, from a simple amusement to a symbolic representation of the deceased's journey through the underworld. See also **nine men's morris**.

sensitivity

The tendency of a system, which may be chaotic (see **chaos**) to change dramatically with only small **perturbation**s.

sequence

An ordered list of values that may be finite or infinite in length. Among the different types of sequence are **arithmetic sequence**s, **geometric sequence**s, and the **harmonic sequence**.

series

A sum of all or some of the terms of a **sequence**. A series may or may not *converge* to a particular value as more and more terms are included. A series is said to be *absolutely convergent* if the sum of the absolute values of the terms converges; in this case the series converges no matter how the terms of the sum are arranged. Series that are *conditionally convergent* only converge for some arrangements of the terms and, even then, converge to different values for different arrangements.

serpentine

A curve named and studied by Isaac **Newton** in 1701 and contained in his classification of **cubic curve**s. It had been studied earlier by **de L'Hôpital** and Christiaan **Huygens** in 1692. The curve is given by the Cartesian equation

$$y(x) = abx/(x^2 - a^2).$$

set

A finite or infinite collection of objects known as *elements.* Sets are one of the most basic and important concepts in mathematics. An example of a finite set is the set of whole numbers from 1 to 58; an example of an infinite set is the set of all the **rational number**s. Two sets are equal if, and only if, they contain the same objects. Standard notation uses braces around the list of elements, as in: {red, green, blue}. If A and B are two sets and every x in A is also contained in B, then A is said to be a *subset* of B. Every set has as subsets itself, known as the *improper subset,* and the **empty set**. The *union* of a collection of sets $S = \{S_1, S_2, S_3, \ldots\}$ is the set of all elements contained in at least one of the sets $S_1, S_2, S_3, \ldots$. The *intersection* of a collection of sets $T = \{T_1, T_2, T_3, \ldots\}$ is the set of all elements contained in all of the sets. The union and intersection of sets, say A_1 and A_2, is denoted $A_1 \cup A_2$ and $A_1 \cap A_2$, respectively. The set of all subsets of X is called its *power set* and is denoted 2^X or $P(X)$. See also **set theory**, **Venn diagram**, and **Russell's paradox**.

set of all sets

See **Russell's paradox**.

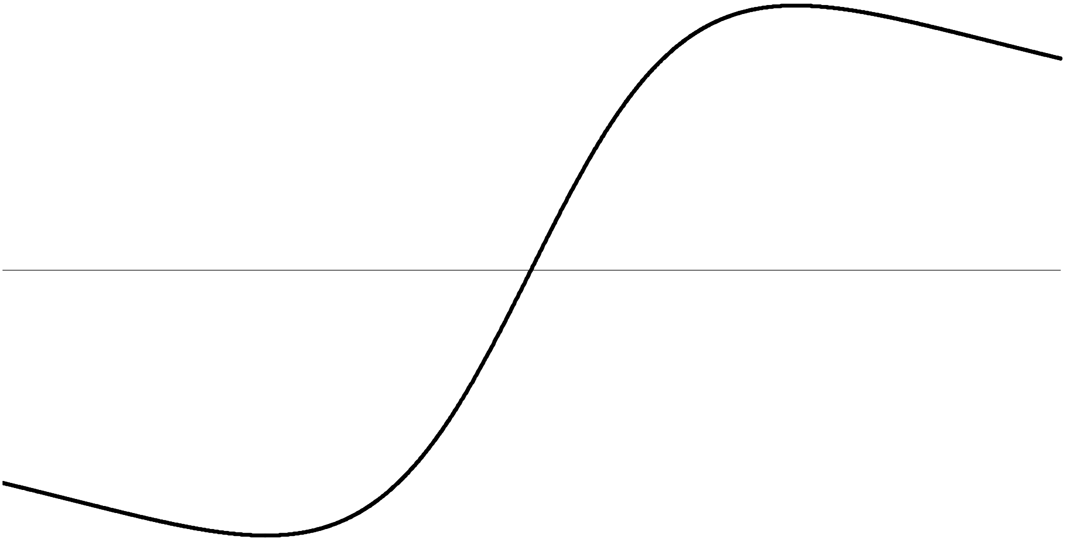

serpentine The serpentine curve. © *Jan Wassenaar, www.2dcurves.com*

set theory

A branch of **mathematics** created by Georg **Cantor** at the end of the nineteenth century. Initially controversial, set theory has come to play a foundational role in modern mathematics, in that it is used to justify assumptions made concerning the existence of mathematical objects (such as numbers or functions) and their properties. Formal versions of set theory also figure centrally in specifying a theoretical ideal of mathematical rigor in proofs. Cantor's basic discovery was that if we define two sets *A* and *B* to have the same number of members (the same *cardinality*), then there is a way of pairing off members of *A* exhaustively with members of *B*. The appearance around the turn of the century of set-theoretical paradoxes, such as **Russell's paradox**, prompted the formulation in 1908 by Ernst Zermelo of an axiomatic theory of sets. The axioms for set theory now most often studied and used are those called the *Zermelo-Fraenkel axioms*, usually together with the **axiom of choice**. The Zermelo-Fraenkel axioms are commonly abbreviated to ZF, or ZFC if the axiom of choice is included. An important feature of ZFC is that every object that it deals with is a set. In particular, every element of a set is itself a set. Other familiar mathematical objects, such as numbers, must be subsequently defined in terms of sets.

seven

A lucky number in the eyes of many people and one that has been given much spiritual significance. The early religious and cultural use of the seven-day week almost certainly stems from the fact that the Moon goes through its four phases in a bit over 28 days, which divides nicely into seven days per phase. There are seven moving objects in the sky visible to the naked eye (the Sun, Moon, Mercury, Venus, Mars, Jupiter, and Saturn), seven seas, seven orders of architecture, seven deadly sins, seven liberal arts and sciences, and seven dwarfs. The seventh son of a seventh son is supposed to be born gifted (Donny Osmond was such a person). In the Bible, there were seven years of famine and seven years of plenty, and seven years were taken to construct King Solomon's Temple. The Pythagoreans were especially intrigued by the number as it is the sum of three and four, which are the number of sides of a triangle and a square–shapes of enormous importance to the sect. These links with Solomon's Temple and the Pythagoreans help explain the importance of seven in freemasonry. Seven is the smallest positive integer whose reciprocal has a pattern of more than one repeating digit: $1/7 = 0.142857142857\ldots$ and is the smallest number for which the digit sequence of $1/n$ is of length $n - 1$ (the longest such a sequence can be). The next such numbers are 17, 19, 23, 29, 47, 59, 61, 97, 109, 113, ... Other curios: the citrus soda 7-UP, created in 1929, was so called because the original containers were 7 ounces and "up" was the direction of the bubbles, and seven is the maximum number of times you can fold any sheet of paper (try it!).

seventeen

The number most often picked in response to the request "Pick a random number from 1 to 20." Seventeen is a *Fermat prime* (a **prime number** of the form $2^{2^n} + 1$, where n is a positive integer), the exponent of a *Mersenne prime* (a prime p for which $2^p - 1$ is prime), and the only prime that is the sum of four consecutive primes (2 + 3 + 5 + 7). Seventeen is also the smallest number for which the sum of the digits of its cube is equal to the number: $17^3 = 4913$, $4 + 9 + 1 + 3 = 17$, and the smallest number that can be written as $a^2 + b^3$ in two different ways: $17 = 3^2 + 2^3 = 4^2 + 1^3$. The pair (8, 9), whose sum is 17, is the only pair of consecutive numbers where one is a square and the other is a cube (a result proved by Leonhard **Euler**.) There are 17 planar crystallographic groups, called **wallpaper groups**. The minimum number of faces on a convex **polyhedron** that has only one stable face is 17. (A stable face is one that the figure can rest on without falling over; most polygons have more than one such face.) Seventeen is also the answer to the following problem: At a party where any two people have previously met each other in one of three other places, what is the least number of people who must be at the party to guarantee that there is at least one group of three people who have met each other before in the same place?

sexagesimal

Of, relating to, or based on the number 60. Sexagesimal refers especially to the **number system** with base 60. The Babylonians began using such a scheme around the beginning of the second millennium B.C. in what was the first example of a **place-value system**. Our degree of 60 minutes, minute of 60 seconds (in both time and angle measure), and hour of 60 minutes hark back to this ancient method of numeration. Why the Babylonians counted using sexagesimal isn't known, but 60 certainly has more factors than any other number of comparable size.

Shannon, Claude Elwood (1916–2001)

An American mathematician and a pioneer of **information theory**. Shannon was the first to realize that any sort of message can be transmitted as a series of 0's and 1's, regardless of whether it consists of words, numbers, pictures, or sound. In his master's thesis, he explained how electrical switches could represent **binary** digits–a 1 when the switch is on and 0 when it is off. He also used **Boolean algebra** to show that complex operations could be carried out automatically on these electrical circuits, thus manipulating the data they were storing. It was in one of his papers, "A Mathematical Theory of Communication" published in 1948, that the word "bit" (short for binary digit) was used for the first time. In fact the framework and terminology for information theory he developed remains standard today. Shannon was driven by curiosity, and in his own words he "just wondered how things were put together." Among his inventions were rocket-powered Frisbees, motorized Pogo sticks, a device that could solve the **Rubik's cube** puzzle, and a juggling machine. (He could ride a unicycle while juggling three balls.) He was involved in pioneering **artificial intelligence** research, which included building the electromechanical mouse called "Theseus" that could navigate a metal **maze** using magnetic signals. Shannon also built a **chess**-playing computer, many years before IBM's Deep Blue, that played well against the world champion of the time, Mikhail Botvinnik (the computer lost only after 42 moves).

shell curve

See **Dürer's shell curve**.

shuffle

How many shuffles does it take to **randomize** a deck of cards–in other words, to mix up the cards about as thoroughly as dropping them all on the table and stirring them around for several minutes (the author's usual method). The answer depends on the kind of shuffle used. The beginner's *overhand shuffle,* for example, is a really bad way to mix cards: about 2,500 such shuffles are need to randomize a deck of 52 cards. A magician's *perfect shuffle,* on the other hand, in which the cards are cut exactly in half and then perfectly interlaced, *never* produces randomization (see below). One of the most effective ways to get a random deck is the *riffle shuffle* in which the deck is cut in half and imperfectly interlaced by dropping cards one by one from either half of the deck with a probability proportional to the current sizes of the deck splits. In 1992 Persi Diaconis (then at Harvard) and David Bayer demonstrated that, starting with a completely ordered deck, it takes *seven* riffle shuffles to produce randomization.[27] Any more than this and there's no significant increase in the randomness; any less and the shuffle is far from random. In fact, not only are five or six riffles not enough to randomize, there are some configurations of cards that are impossible to reach in this number of shuffles! To understand this, suppose the starting order of the cards is marked 1 to 52, top to bottom. After one shuffle, only configurations with two or fewer rising sequences are possible. A rising sequence is a maximal increasing sequential ordering of cards that appear in the deck (with other cards possibly interspersed) as it is run through from top to bottom. For instance, in an eight-card deck, 12345678 is the ordered deck and it has one rising sequence. After one shuffle, 16237845 is a possible configuration, and there are two rising sequences (the underlined numerals form one, the nonunderlined numerals form the other). Clearly the rising sequences are formed when the deck is cut before the cards are interleaved in the shuffle. After two shuffles, there can be at most four rising sequences, since each of the two rising sequences from the first shuffle has a chance of being cut in the second. This pattern continues: the number of rising sequences can at most double during each shuffle. After five shuffles, there are at most 32 rising sequences. But the *reversed* deck, numbered 52 down to 1, has 52 rising sequences. Thus, this is one (of many) arrangements that are unattainable in five riffle shuffles. Interestingly, Diaconis and other researchers have also found that decks can undergo sudden changes in their degree of randomness; after six riffle shuffles, a deck is still visibly ordered, but this order vanishes one shuffle later.

Perfect shuffles do the exact opposite of randomizing: they preserve order at every stage. There are two kinds of perfect shuffles. The *out-shuffle* is one in which the top card stays on top; the *in-shuffle* is one in which the top card moves to the second position of the deck. Amazingly, eight perfect out-shuffles restore the deck to its original order! Magicians use combinations of out and in shuffles to perform a variety of baffling tricks and to control the position of any given card in a deck. How could you make the top card (call it position 0) go to position *n*? Easy: write *n* in **binary** (base 2), read the 0's and 1's from left to right, perform an out-shuffle for a 0 and an in-shuffle for a 1, and, as if by magic, the top card will have materialized at position *n*.

shunting puzzles

Railroad modelers, especially those with limited space available for their layouts, often enjoy setting up track that allows interesting shunting problems to be tried out and solved. The most famous mathematical puzzle of this type, called the *railroad shunting puzzle,* comes in a number of variations, but basically the problem is that there are two trains (*A* and *B* in the diagram) facing each other on a single line with just one short siding, which will only hold one item of rolling stock at a time. In order to enable the two trains to pass each other and to continue their journey, a series of movements using the siding is required. First time around it's quite a brain-teaser, which probably explains why railroad companies all over the world took the more costly but easier way out and built passing sidings!

Siegel's paradox

If a fixed fraction *x* of a given amount of money *P* is lost, and then the same fraction *x* of the remaining amount is gained, the result is less than the original and equal to the final amount if a fraction *x* is first gained, then lost.

Sierpinski, Waclaw Franciszek (1882–1969)

A Polish mathematician who made outstanding contributions to **set theory**, which included research on the **axiom of choice** and the **continuum hypothesis**, **number theory**, and **topology**. Two well-known fractals, the **Sierpinski carpet** and the **Sierpinski gasket**, are named after him.

Sierpinski carpet

A fractal, named after Waclaw **Sierpinski**, that is derived from a square by cutting it into nine equal squares with a 3 × 3 grid, removing the central piece, and then applying the same procedure ad infinitum to the remaining eight squares. It is one of two generalizations of the Cantor set to two dimensions; the other is the **Cantor dust**. The carpet's **Hausdorff dimension** is log 8/log 3 = 1.8928. . . .

Sierpinski gasket

A fractal, also known as the *Sierpinski triangle* or *Sierpinski sieve* after its inventor Waclaw **Sierpinski**. It is produced by the following set of rules: (1) start with any triangle in a plane; (2) shrink the triangle by ½, make three copies, and translate them so that each triangle touches the two other triangles at a corner; (3) repeat step 2 ad infinitum. The gasket can also be made by starting with **Pascal's**

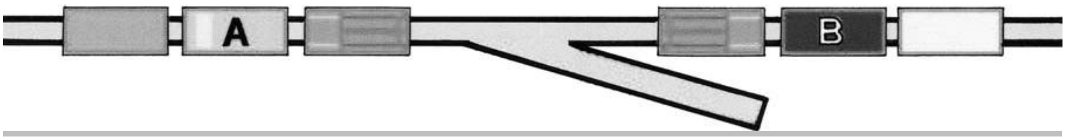

shunting puzzles How can two trains, traveling in opposite directions on a single track, get past each other by using a siding that can only accommodate a single item of rolling stock at a time?

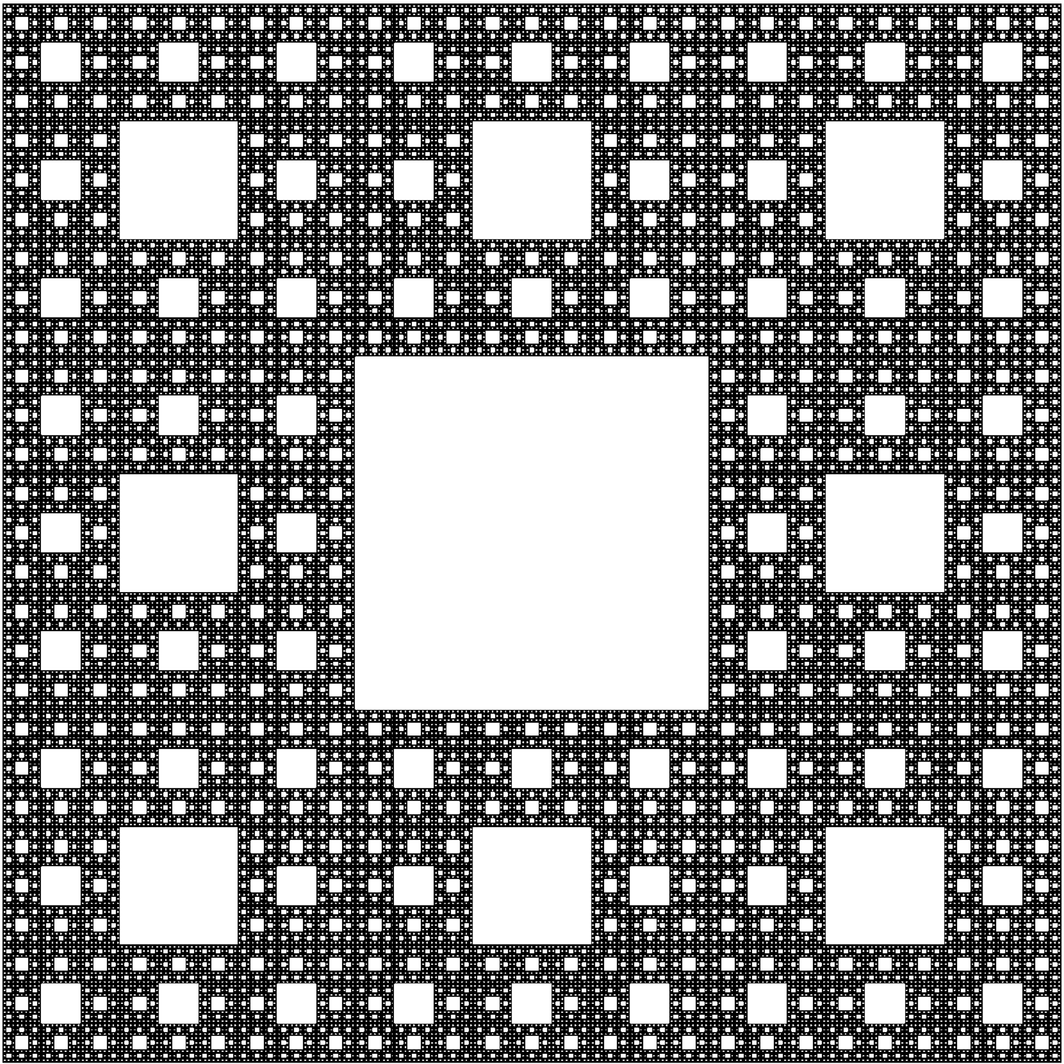

Sierpinski carpet

triangle, then coloring the even numbers white and the odd numbers black. Most curiously, it can be generated by a game of chance. Begin with three points, labeled 1, 2, and 3, and any starting point, *S*. Then select randomly 1, 2, or 3, using a die or some other method. Each random number defines a new point halfway between the latest point and the labeled point that the random number indicates. When the game has gone on long enough the pattern produced is the Sierpinski gasket. The gasket has a **Hausdorff dimension** of log 3/log 2 = 1.585 . . . , which follows from the fact that it is a union of three copies of itself, each scaled by a factor of ½. Adding rounded corners to the defining curve gives a nonintersecting curve that traverses the gasket from one corner to another and which Benoit **Mandelbrot** called the *Sierpinski arrowhead.*

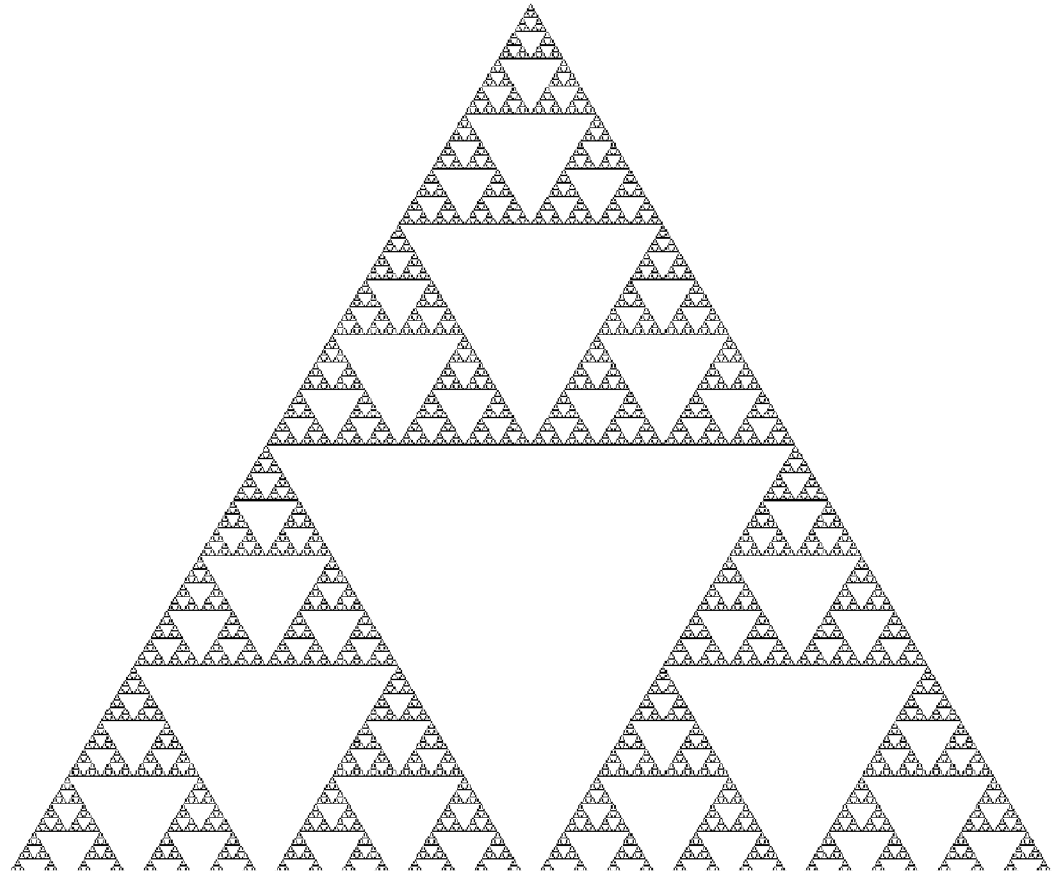

Sierpinski gasket

Sierpinski number

A positive, odd integer *k* such that *k* times $2^n + 1$ is *never* a **prime number** for any value of *n*. In 1960 Waclaw **Sierpinski** showed that there were infinitely many such numbers (though he didn't give a specific example). This is a strange result. Why should it be that while the vast majority of expressions of the form *m* times $2^n + 1$ eventually produce a prime, some don't? For now, mathematicians are focused on a more manageable problem posed by Sierpinski: What is the smallest Sierpinski number? In 1962, John Selfridge discovered the smallest known Sierpinski number, $k = 78{,}557$. The next largest is 271,129. Is there a smaller Sierpinski number? No one yet knows. However, to establish that 78,557 is really the smallest, it would be sufficient to find a prime of the form $k(2^n + 1)$ for every value of *k* less than 78,557. In early 2001, there were only 17 candidate values of *k* left to check: 4,847; 5,359; 10,223; 19,249; 21,181; 22,699; 24,737; 27,653; 28,433; 33,661; 44,131; 46,157; 54,767; 55,459; 65,567; 67,607; and 69,109. In March 2002, Louis Helm of the University of Michigan and David Norris of the University of Illinois started a project called "Seventeen or Bust," the goal of which is to harness the computing power of a worldwide network of hundreds of personal computers to check for primes among the remaining candidates. The team's effort have so far eliminated six candidates–5,359; 44,131; 46,157; 54,767; 65,567; and 69,109. Despite this encouraging start, it may take as long as a decade, with many additional participants, to check the eleven remaining candidates.

sieve of Eratosthenes

The most efficient way to find all of the smallest **prime numbers**. First described by **Eratosthenes of Cyrene**, it involves making a list of all the integers less than or equal to *n* (and greater than one), then striking out the multiples of all primes less than or equal to the square root of *n*. The numbers that are left are the primes. For example, to find all the primes less than or equal to 30, we list the numbers from 2 to 30:

2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30

The first number, 2, is prime, so we keep it and strike out all of its multiples, leaving

2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29

The next number left, 3, is prime, so again we retain it and delete all of its multiples, leaving

2, 3, 5, 7, 11, 13, 17, 19, 23, 25, 29

Now we do the same thing for 5, another prime, to give

2, 3, 5, 7, 11, 13, 17, 19, 23, 29

The next number, 7, is larger than the square root of 30, so all of the numbers left are primes.

significant digits

The digits that define a numerical value. The significant digits of a given number begin with the first nonzero integer digit or, if this number is less than unity, with the first (zero or nonzero) decimal digit. They end with the final (zero or nonzero) decimal digit; the final zero or zeros of an integer may or may not be significant.

similar

Having the same shape but not necessarily the same size. Two triangles are similar if they have equal angles and their corresponding sides, say, a_1, b_1, c_1 and a_2, b_2, c_2, have a common ratio, r: $a_1/a_2 = b_1/b_2 = c_1/c_2$. In general, a *similarity* is a transformation under which the distance between any corresponding pair of points changes by the same factor.

simple group

A **group** that has no nontrivial proper normal **subgroup**s. Simple groups are important because they can be thought of as the blocks out of which other groups can be built. Much activity has been expended in the classification of all finite simple groups.

simplex

The *n*-dimensional generalization of the **triangle** and the **tetrahedron**; in other words, a **polytope** in *n* dimensions with $n + 1$ vertices.

simply connected

The condition of a geometrical object if it consists of one piece and doesn't have any holes or "handles." For example, a line, disk, and sphere are simply connected, but a torus (doughnut) and teapot are not. See also **connected**.

simulation

Experimentation in the space of theories, or a combination of experimentation and theorization. Some numerical simulations are programs that represent a model for how nature works. Usually, the outcome of a simulation is as much a surprise as the outcome of a natural event, due to the richness and uncertainty of computation.

sine

The **trigonometric function** of an angle of a right-angled triangle other than the right angle that is equal to the length of the side adjacent to the angle divided by the length of the hypotenuse. The curve of $y = \sin x$ is called the sine curve, or *sinusoid.*

Singmaster, David

A professor at the school of computing, information systems, and mathematics at South Bank University, London, who is one of the world's leading compilers and historians of mathematical puzzles.

singularity

(1) A point at which the **derivative** does not exist for a given **function** but every neighborhood of which contains points for which the derivative exists. (2) A point in **space-time** at which gravitational forces cause matter to have infinite density and infinitesimal volume, and space and time to become infinitely distorted.

six

The smallest **perfect number**, the number of faces of a **cube**, and the number of sides of a **hexagon**. There are six players on a volleyball team, six kinds of chessmen, and six types of quark (not including antiquarks). A touchdown in American football earns six points and a hit across the boundary rope of a cricket field without bouncing scores six runs. Long ago people indicated a number by pointing to a part of their body; this is echoed in the New Guinea word for six, which is the same as that for "wrist."

six circles theorem

Given a triangle and a circle inside it that touches two of the triangle's sides, draw a second circle touching another two sides and touching the first circle. Draw a third circle touching two sides and the second circle, and so on. This chain ends with the sixth circle, which will touch the first.

sixty

See **sexagesimal**.

skeletal division

A long division in which most or all of the digits are replaced by symbols (usually asterisks) to form a **cryptarithm**.

skew lines

Also known as *crossing lines,* lines that lie in different planes and do not intersect one another.

Skewes' number

A famous **large number**, commonly given as $10^{10^{10^{34}}}$, that was first derived in 1933 by the South African mathematician Samuel Skewes in a proof involving **prime numbers**.[299] G. H. **Hardy** once described Skewes' number as "the largest number which has ever served any definite purpose in mathematics," though it has long since lost that distinction. Skewes' numbers–there are actually two of them–came about from a study of the frequency with which **prime numbers** occur. Gauss's well-known estimate of the number of prime numbers less than or equal to n, pi(n), is the integral from $u = 0$ to $u = n$ of $1/(\log u)$; this integral is called Li(n). In 1914 the English mathematician John Littlewood proved that $\text{pi}(x) - \text{Li}(x)$ assumes both positive and negative values infinitely often. For all values of n up to 10^{22}, which is as far as computations have gone so far, Li(n) has turned out to be an overestimate. But Littlewood's result showed that above some value of n it becomes an underestimate, then at an even higher value of n it becomes an overestimate again, and so on. This is where Skewes' number comes in. Skewes showed that, if the **Riemann hypothesis** is true, the first crossing can't be greater than $e^{e^{e^{79}}}$. This is called the first or *Riemann true Skewes' number.* Converted to base 10, the value can be approximated as $10^{10^{10^{34}}}$, or more accurately as $10^{10^{8.852142 \times 10^{33}}}$ or $10^{10^{8852142197543270606106100452735038.55}}$. In 1987, the Dutch mathematician Herman te Riele[330] reduced dramatically the upper bound of the first crossing to $e^{e^{27/4}}$, or approximately 8.185×10^{370}, while John **Conway** and Richard Guy[68] have made the contradictory claim that the lower bound is 10^{1167}. In any event, Skewes' number is now only of historical interest. Skewes also defined the limit if the Riemann hypothesis is false: $10^{10^{10^{1000}}}$. This is known as the second Skewes' number.

slide rule

A calculating device consisting of two sliding logarithmic scales.

sliding-piece puzzle

A type of *sequential-movement puzzle,* within the larger category of **mechanical puzzles**, that involves sliding one piece at a time into a single vacant opening in order to advance toward the solution–a certain orderly arrangement of the pieces. The best known is Loyd's **Fifteen Puzzle**.

Slocum, Jerry

An American historian of and writer on **mechanical puzzles**.

slope

"Rise over run." For a straight line in the plane, the *slope* is the **tangent** of the angle it forms with the positive x-axis. For a curve, the slope is, by definition, the slope of the *tangent* line. Therefore, if the slope is constant, a line is straight.

Slothouber-Graatsma puzzle

A **packing** puzzle in which six $1 \times 2 \times 2$ blocks and three $1 \times 1 \times 1$ blocks must be fitted together to make a $3 \times 3 \times 3$ cube. There is only one solution. A similar but much more difficult puzzle, named after its inventor, John **Conway**, calls for packing three $1 \times 1 \times 3$ blocks, one $1 \times 2 \times 2$ block, one $2 \times 2 \times 2$ block, and thirteen $1 \times 2 \times 4$ blocks into a $5 \times 5 \times 5$ box.

Smith number

A **composite number**, the sum of whose digits equals the sum of the digits of its prime factors (its **factors** that are **prime numbers**). The name stems from a phone call in 1984 by the mathematician Albert Wilansky to his brother-in-law Smith, during which Wilansky noticed that the phone number, 493-7775, obeyed the condition just mentioned. Specifically:

$$4{,}937{,}775 = 3 \times 5 \times 5 \times 65{,}837$$

$$4 + 9 + 3 + 7 + 7 + 7 + 5 = 3 + 5 + 5 + 6 + 5 + 8 + 3 + 7$$

Trivially, all **prime numbers** have this property, so they are excluded. The first few Smith numbers are: 4, 22, 27, 58, 85, 94, 121, 166, 202, 265, 274, 319, 346, ... In 1987, Wayne McDaniel proved that there are infinitely many Smiths.

smooth

(1) Infinitely differentiable; possessing infinitely many **derivatives**. For example, $\sin(x)$ is a smooth function, while $|x|^3$ is not. More complicated mathematical objects such as **manifolds** are called smooth if they are defined or described by smooth functions. (2) Continuously differentiable (see **continuity**); possessing a continuous **tangent** or derivative.

Smullyan, Raymond (1919–)

An American mathematical logician, puzzle-maker, and magician, who has taught in various colleges but is best known for his books on recreational mathematics. Among these are *What Is the Name of This Book?*, *The Lady or the Tiger*, *The Tao is Silent*, and *This Book Needs No Title: A Budget of Living Paradoxes*.[304–311] Smullyan is an inventive maker of logic paradoxes and a pioneer of chess problems that involve "retrograde analysis," in which the object is to deduce the past history of a game from some given present position.

snow

It is often said that no two snowflakes are alike. While this is hard to prove, individual samples can be captured on a chilled glass microscope slide and preserved with artist's spray fixative. All are six-sided and the more ornate kind, called *dendritic snowflakes*, form when the air temperature is between −12°C and −16°C (10°F and 3°F). Typical snowflakes fall at a rate of a meter or two per second; assuming 1.5 m/s and a cloud base of 3,000 m (roughly the height of nimbostratus clouds) gives a descent time of 20 minutes. One of the great urban legends is that the Inuit have *n* words for "snow," where *n* is a large number. This story may have started in 1911 when anthropologist Franz Boaz casually mentioned that the Inuit–he called them "Eskimos," using the derogatory term of a tribe to the south of them for eaters of raw meat–had four different words for snow. With each succeeding reference in textbooks and the popular press the number grew to as many as 400 words. A problem with trying to pin down exactly how many Inuit words there are for snow and/or ice, or for anything else, is that the various dialects of Inuit are polysynthetic, which means that words can effectively be made up on the spot by concatenating various particles to the root word. For example, the suffix *-tluk*, for "bad," might be added to *kaniktshaq*, for "snow," to give *kaniktshartluk*, "bad snow." This can give rise to any number of snow terms, from *akelrorak* ("newly drifting snow") to *mitailak* ("soft snow over an opening in an ice floe").

snowball prime

Also known as a *right-truncatable prime*, a **prime number** whose digits can be chopped off, one by one, from the right-hand side, yet still leave a prime number. This means that even if you stop writing before you finish the number, you will still have written a prime. The largest snowball prime is 73,939,133 (7, 73, 739, . . . , 73,939,133 are all prime).

snowflake curve

See **Koch snowflake**.

soap film

See **bubbles**.

Soddy circle

A solution to the three-circle form of the **Apollonius problem** in which each of the given circles is tangent to (just touches) the other two.[312] There are two Soddy circles: the *outer Soddy circle*, which surrounds the three given circles, and the *inner Soddy circle*, which is interior to them. The inner Soddy circle is the solution to the **four coins problem**.

Soddy's formula

If four circles *A*, *B*, *C*, and *D*, of radii r_1, r_2, r_3, and r_4, are drawn so that they do not overlap but each touches the other three, and if we let $b_1 = 1/r_1$, etc., then

$$(b_1 + b_2 + b_3 + b_4)^2 = 2(b_1^2 + b_2^2 + b_3^2 + b_4^2).$$

solid

Of or relating to three-dimensional geometric figures or bodies.

solid angle

An angle formed by three or more planes intersecting at a common point. Solid angles are measured in **steradian**s.

solid geometry

The geometry of three-dimensional space.

solidus

The slanted line in a fraction such as *a/b* dividing the **numerator** from the **denominator**.

solitaire

See **peg solitaire**.

solitary number

A number that is not one of a pair of **amicable numbers**. Examples include all **prime number**s, all integer powers or primes, and other numbers such as 9, 16, 18, 52, and 160.

soliton

A solitary wave that can travel for long distances without changing its shape or losing energy. Mathematically, a soliton is a solution to a **partial differential equation** that is localized in some directions but not localized in time, and which does not change its shape. The first soliton to be described was a water wave seen by the engineer and shipbuilder John Scott Russell (1808–1882) in 1834 when he was riding by the Grand Union Canal at Hermiston, Glasgow. He observed that when a canal boat stopped, its bow wave continued onward as a well-defined elevation of the water at constant speed. (Another account says Scott Russell observed it on the Glasgow and Ardrossan Canal when a horse bolted with a light canal boat in tow.) The phenomenon was largely forgotten until the 1960s, when the American physicist Martin Kruskal rediscovered it and called it a soliton wave. On July 12, 1995, a viaduct at Hermiston was renamed the John Scott Russell viaduct, with Kruskal unveiling plaques and attempting to re-create a soliton wave. A classic soliton occurs on the River Severn in England, which starts at the head of the triangular Bristol Channel and narrows rapidly upstream. When the tide comes in, it is greatly compressed and produces the Severn Bore, a tidal wave up to 6 feet (2 m) high which rushes about 20 miles (32 km) up the river to Gloucester at a speed of up to 10 mph (16 km/hr). It is strongest at the spring tides, i.e. at full and new moons, and is now popular with surfers and canoeists.

solution

A value that satisfies the requirements of an equation. See also **lost solutions**.

Soma cube

A mathematical puzzle devised by Piet **Hein** in 1936 during a lecture on quantum mechanics by Werner Heisenberg in which the great German physicist was describing a space sliced into cubes. In a moment of genius, Hein grasped that the result of combining all seven of the irregular shapes that can be made from no more than four unit cubes joined at their faces is a single larger (3 × 3 × 3) cube. The Soma cube was first brought to popular attention by Martin **Gardner** in his "Mathematical Games" column in *Scientific American* in 1958. All 240 possible solutions were first identified by John **Conway** and Mickael Guy in 1961. The pieces can also be used to make a variety of other interesting three-dimensional

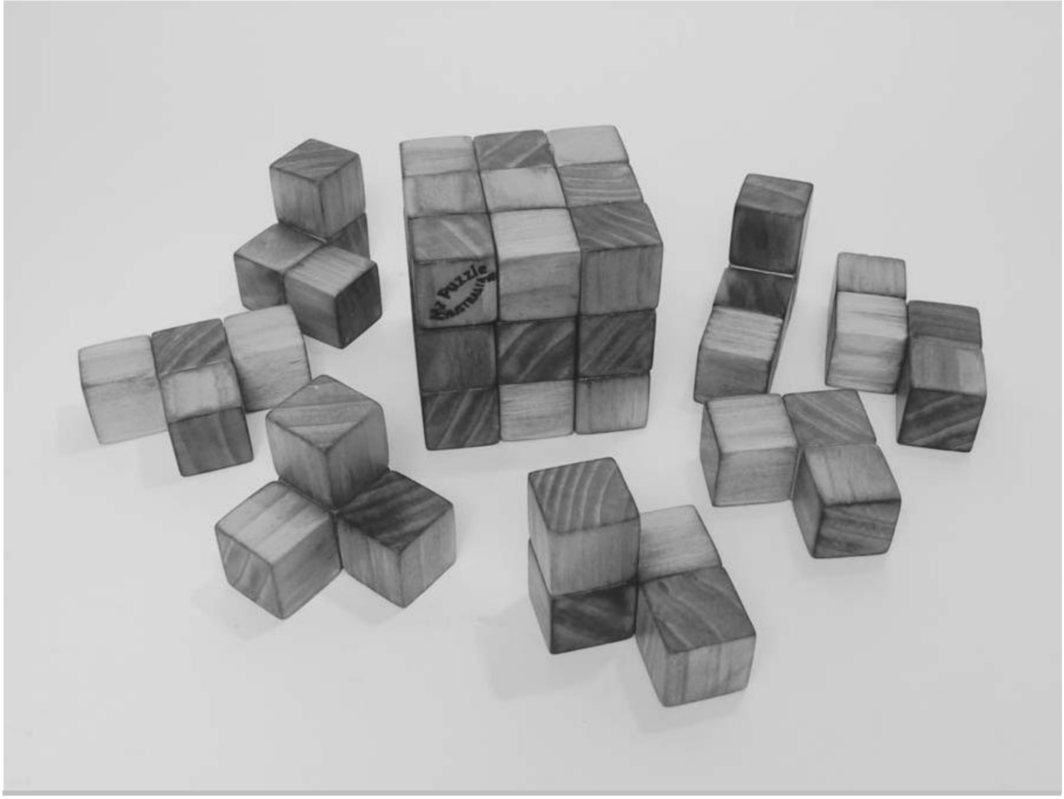

Soma cube How the pieces of the cube fit together. *Mr. Puzzle Australia, www.mrpuzzle.com.au*

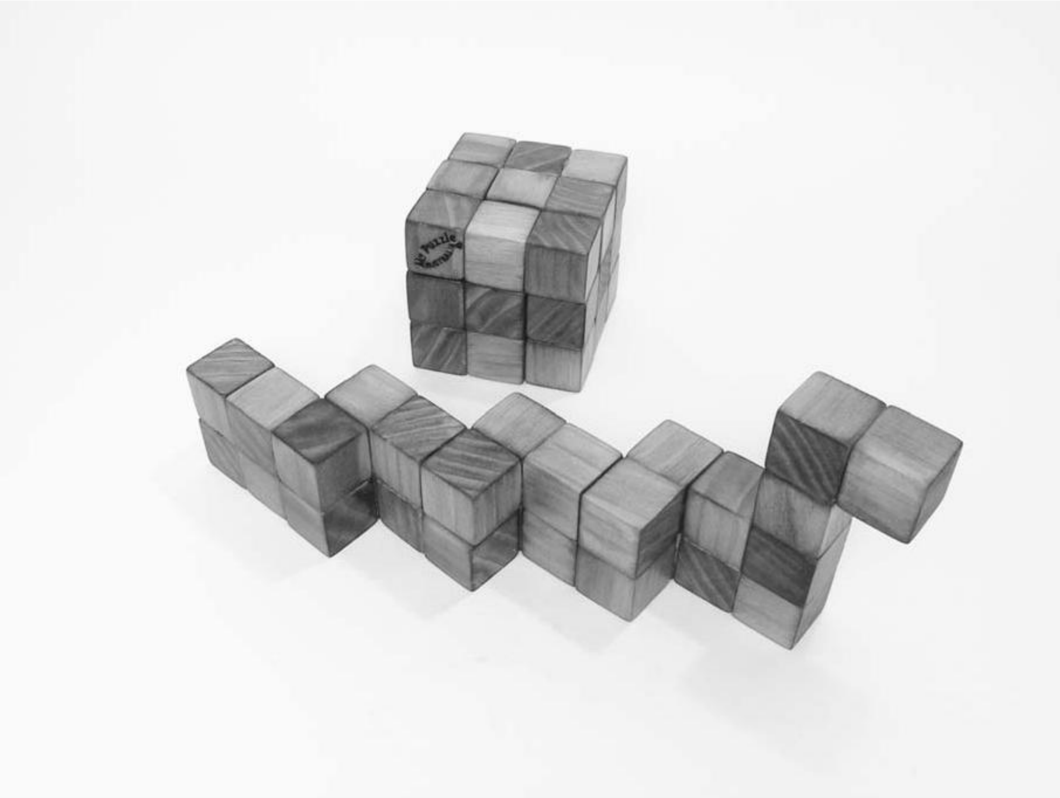

Soma cube One of many animals into which the cube pieces can be arranged. *Mr. Puzzle Australia, www.mrpuzzle.com.au*

shapes so that the Soma cube is often regarded as a three-dimensional analog of **tangrams**. It is possible that the puzzle is named after the fictitious drug "soma" in Aldous Huxley's novel *Brave New World.* See also **polyomino**.[112]

Sophie Germain prime

Any **prime number** p such that $2p + 1$ is also prime; the smallest examples are 2, 3, 5, 11, 23, 29, 41, 53, 83, 89, 113, and 131. Around 1825 Sophie **Germain** proved that the first case of **Fermat's last theorem** (FLT) is true for such primes. Soon after, Adrien-Marie Legendre began to generalize this by showing the first case of FLT also holds for odd primes p such that $kp + 1$ is prime, $k = 4, 8, 10, 14$, and 16. In 1991 Fee and Granville extended this to $k < 100$, where k is not a multiple of three. Many similar results were also shown, but now that FLT has been proven correct, they are of less interest.

soroban

See **abacus**.

space

(1) The three-dimensional theater in which things as we know them can exist or in which events can take place. In the Einsteinian worldview, space and **time** are united inextricably in a **space-time** continuum and there is also the possibility of **higher dimensions**. (See also **fourth dimension**.) (2) In mathematics, there are additionally many other types of space, most of them too abstract to imagine or to describe accurately in a few sentences. Generally, a mathematical space is a **set** of points with additional features. In a *topological space* (see **topology**) every point has a collection of neighborhoods to which it belongs. In an *affine space,* which is a generalization of the familiar concepts of a straight line, a plane, and ordinary three-dimensional space, a defining feature is the ability

to fix a point and a set of coordinate axes through it so that every point in the space can be represented as a "tuple," or ordered set, of coordinates. Other examples of mathematical spaces include *vector spaces, measure spaces,* and *metric spaces.*

space-filling curve

A curve that passes through every point of a finite region (such as a unit square or unit cube) of an n-dimensional space, where $n \geq 2$. A well-known example is the **Peano curve**.

space-time

The inseparable four-dimensional **manifold**, or combination, which **space** and **time** are considered to form in the special and general theories of relativity (see **relativity theory**). A point in space-time is known as an *event*. Each event has four coordinates (x, y, z, t). Just as the x, y, z coordinates of a point depend on the axes being used, so distances and time intervals, (which are invariant in Newtonian physics) may depend (in relativistic physics) on the reference frame of an observer; this can lead to bizarre effects such as *length contraction* and *time dilation*. A *space-time interval* between two events is the invariant quantity analogous to distance in **Euclidean space**. The space-time interval s along a curve is defined by the quantity

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2dt^2,$$

where c is the speed of light. A basic assumption of relativity theory is that coordinate transformations leave intervals invariant. However, note that whereas distances are always positive, intervals may be positive, zero, or negative. Events with a space-time interval of zero are separated by the propagation of a light signal. Events with a positive space-time interval are in each other's future or past, and the value of the interval defines the proper time measured by an observer traveling between them.

special function

A **function**, often named after the person who introduced it, that has a particular use in physics or some branch of mathematics. Examples include *Bessel functions, Lagrange polynomials,* **beta functions**, **gamma functions**, and **hypergeometric functions**.

special relativity

See **relativity theory**.

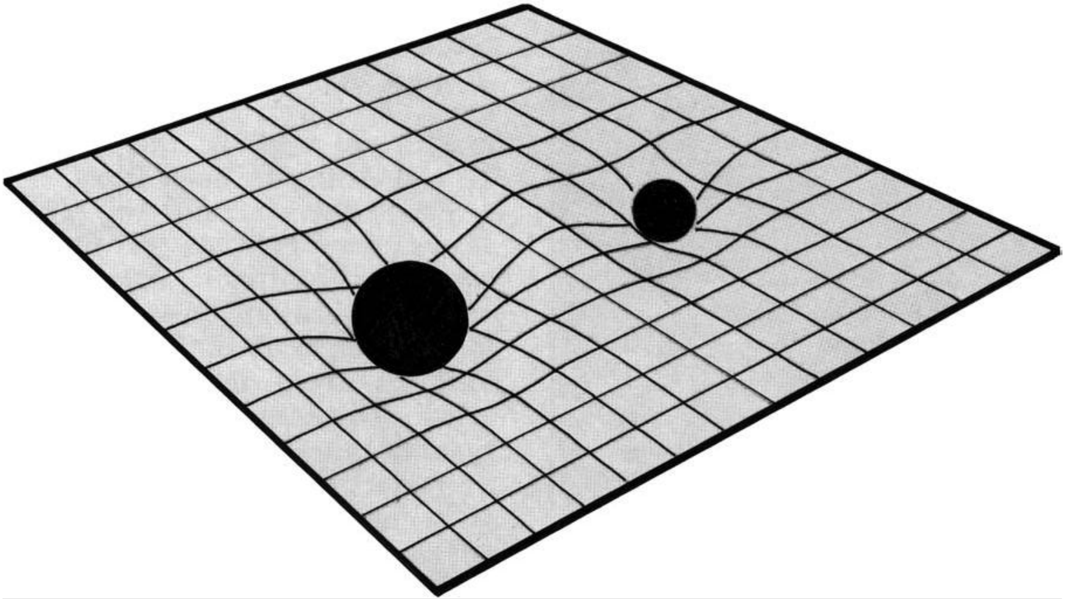

space-time Gravity seen as a curvature of the fabric of space and time.

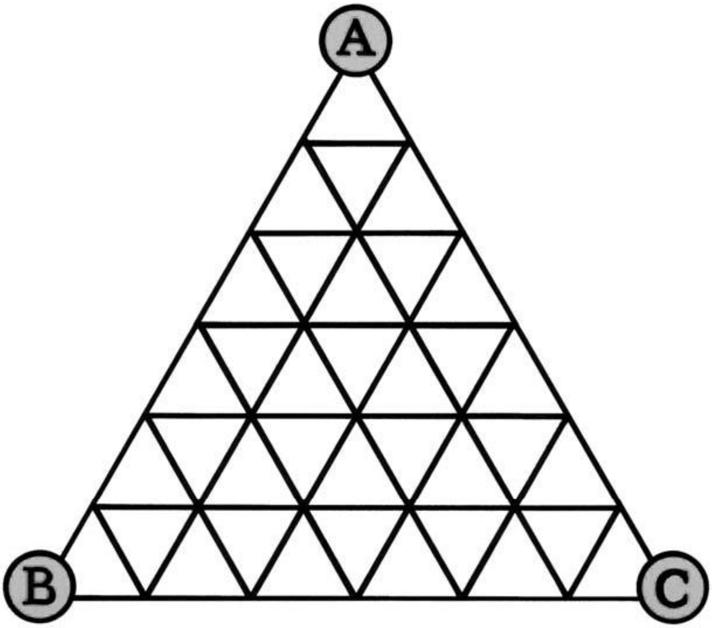

Sperner's lemma A triangular lattice that can be used for demonstrating Sperner's lemma. To understand how, see the accompanying entry.

spectrum

(1) In **quantum mechanics**, the set of allowable energy levels of a particle or system. It is directly related to bright or dark lines in a spectrum of light produced by a prism. (2) In mathematics, the set of **eigenvalue**s of a linear transformation. By historical coincidence, it is equivalent to the notion of a spectrum in quantum mechanics.

Sperner's lemma

Take a triangle *ABC,* labeled counterclockwise, and subdivide it into lots of smaller triangles in any arbitrary way. Then label all the new vertices as follows: (1) vertices along *AB* may be labeled either *A* or *B,* but not *C;* (2) vertices along *BC* may be labeled either *B* or *C,* but not *A;* (3) vertices along *CA* may be labeled either *C* or *A,* but not *B;* (4) vertices inside triangle *ABC* may be labeled *A* or *B* or *C.* Now shade in every small triangle that has three different labels. Use two different shadings to distinguish the triangles that have been labeled counterclockwise (i.e., in the same sense as triangle *ABC*) from the triangles that have been labeled clockwise (i.e., in the sense opposite to that of triangle *ABC*). Then there will be exactly one more counterclockwise triangle than clockwise triangles. In particular, the number of shaded triangles will be odd. This is Sperner's lemma, named after its discoverer, the German mathematician Emanuel Sperner (1905–1980). Sperner's lemma is equivalent to the **Brouwer fixed-point theorem**; a version of it holds in all dimensions.

sphere

Roughly speaking, a ball-shaped object. In everyday usage a sphere is often considered to be solid; mathematicians call this the *interior* of the sphere. In mathematics, a sphere is a **quadric** consisting only of a surface and is therefore hollow. More precisely, a sphere is the set of all points in three-dimensional **Euclidean space** that lie at distance r, the radius, from a fixed point. In **analytical geometry** a sphere with center (x_0, y_0, z_0) and radius r is the set of all points (x, y, z) such that

$$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2.$$

A sphere can also be defined as the **surface of revolution** formed by rotating a circle about its diameter. If the circle is replaced by an **ellipse**, the shape becomes a **spheroid**. The surface area of a sphere is $4\pi r^2$ and its volume is $4\pi r^3/3$. The sphere has the smallest surface area among all surfaces enclosing a given volume and it encloses the largest volume among all closed surfaces with a given surface area. In nature, **bubbles** and water drops tend to form spheres because surface tension always tries to minimize surface area. The circumscribed cylinder for a given sphere has a volume which is 3/2 times the volume of the sphere. If a spherical egg were cut up by an egg-slicer with evenly spaced wires, the bands between the cuts (on the surface of the sphere) would have exactly the same area. Spheres can be generalized to other dimensions. For any natural number n, an n-sphere is the set of points in $(n + 1)$-dimensional **Euclidean space** that lie at distance r from a fixed point of that space. A 2-sphere is therefore an ordinary sphere, while a 1-sphere is a **circle** and a 0-sphere is a pair of points. An n-sphere for which $n = 3$ or more is often called a **hypersphere**.

sphere packing

See **packing**, **Kepler's conjecture**, and **cannonball problem**.

spherical geometry

See **elliptical geometry**.

sphericon

A curious and mathematically delightful three-dimensional object made from a right double-**cone**–two identical, 90° cones joined base to base–and an added twist. To create a sphericon, a right double-cone is sliced along a plane that includes both vertices. The resulting cross section is a square, which enables one of the halves to be rotated through a right angle and the two halves to be glued back together without any overlap. This final twist enables the sphericon to roll in an unusual way. An ordinary cone placed on a flat surface rolls around in circles. A double-cone can roll in a clockwise circle or a counterclockwise one. A sphericon, in contrast, performs a controlled wiggle, with first one conical sector in contact with the flat surface, then the other. Two sphericons placed next to each other can roll on each other's surfaces. Four sphericons arranged in a square block can all roll around one another simultaneously. And eight sphericons can fit on the surface of one sphericon so that any one of the outer solids can roll on the surface of the central one. The sphericon was first found by the En-glishman Colin Roberts in 1969, while he was still in school. In 1999 he brought his discovery to the attention of Ian **Stewart** who subsequently wrote about the new object in his "Mathematical Recreations" column in *Scientific American*.[322]

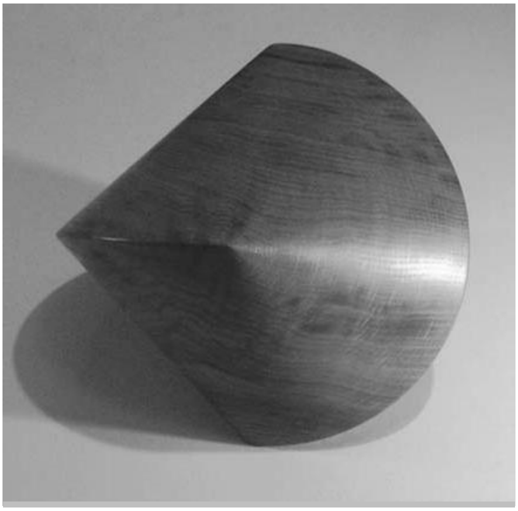

sphericon A model in oak of this remarkable shape. *Paul and Colin Roberts*

spheroid

A surface in three dimensions obtained by rotating an **ellipse** about one of its principal axes. If the ellipse is rotated about its major axis, the surface is called a *prolate spheroid* (similar to the shape of a rugby ball). If the minor axis is chosen, the surface is called an *oblate spheroid* (similar to the shape of Earth). The **sphere** is a special case of the spheroid in which the generating ellipse is a **circle**. A spheroid is a special case of an **ellipsoid** where two of the three major axes are equal.

Sphinx riddle

In Greek mythology, the Sphinx sat outside Thebes and asked this riddle of all travelers who passed by. If the traveler failed to solve the riddle, then the Sphinx killed him/her. And if the traveler answered the riddle correctly, then the Sphinx would destroy herself. The riddle: what goes on four legs in the morning, on two legs at noon, and on three legs in the evening? Oedipus solved the riddle, and the Sphinx destroyed herself. The solution: a man, who crawls on all fours as a baby, walks on two legs as an adult, and walks with a cane in old age. Of course morning, noon, and night are metaphors for the times in a person's life. Such metaphors are common in riddles. There were two Thebes; apparently the Thebes in this myth was the one in Greece, and the Sphinx was different from the one that stands at Giza, in Egypt.

spider-and-fly problem

A puzzle that was originally posed in the *Weekly Dispatch,* an English newspaper, on June 14, 1903, by Henry **Dudeney**, and that appears as one of the problems in *The Canterbury Puzzles* (1907).[87] A simple but elegant exercise in **geodesic**s, it is Dudeney's best-known brain-teaser. In a cuboidal (shoebox-shaped) room measuring 30′ × 12′ × 12′, a spider is in the middle of one 12′ × 12′ wall, 1 foot away from the ceiling. A fly is in the middle of the opposite wall 1 foot away from the floor. If the fly remains stationary, what is the shortest total distance (the geodesic) the spider must crawl along the walls, ceiling, and floor in order to get to the fly? The answer, 40′, can be obtained by flattening out the walls. Note that this distance is shorter than the 42′ the spider would have to travel if it first crawled along the wall to the floor, then across the floor, then up 1′ to get to the fly. A twist to the problem can be obtained by a spider that suspends himself from a strand of cobweb and thus takes a shortcut by not being forced to remain glued to a surface of the room. If the spider attaches a strand of cobweb to the wall at his starting position and lowers himself down to the floor (thus not crawling a single inch), he can then walk across the length of the room (30′) and ascend a single foot, thus reaching his prey after a total crawl of 31′ (although the total distance traveled is of course 42′). If the spider is not proficient with fastening strands to vertical walls, he must first ascend 1′ to the ceiling, from where it can lower himself to the floor, traverse the length of the room, and climb one foot to get to the fly, for a total distance crawled of 32′.

Dudeney and Sam **Loyd** offered several versions of the problem in a rectangular room. In 1926 Dudeney gave a version on a cylindrical glass with the source and the target on opposite sides.

spinor

A mathematical object similar to a **vector**, but which changes sign when rotated through 360°. Spinors were invented by Wolfgang Pauli and Paul **Dirac** (they were named by Paul Ehrenfest) to represent the spin of a subatomic particle. In the early 1930s, Dirac, Piet **Hein**, and others at the Niels Bohr Institute created games such as **Tangloids** to teach and model the calculus of spinors.

spiral

A curve that turns around some central point, getting progressively closer to it or progressively farther from it, depending on which way the curve is followed. Among the best known types are the **Archimedean spiral**, the **logarithmic spiral**, the **circle involute**, and the **lituus**. Like their space-curve cousin the **helix**, all spirals are asymmetric and each come in two forms that are mirror reflections of one another.

spirograph curve

See **roulette**.

Sprague-Grundy theory

A theory of certain classes of games called *impartial games,* discovered independently by Roland Percival Sprague (in 1936) and Patrick Michael Grundy (in 1939)

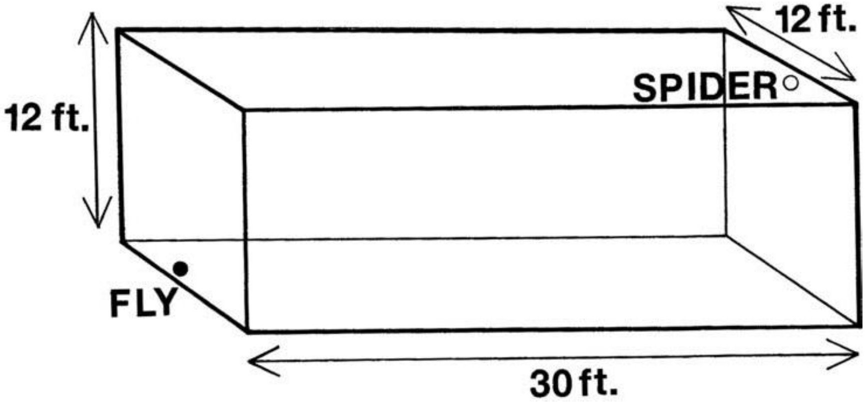

Spider-and-fly problem What is the shortest route for the spider to catch its prey?

and originally applied to **Nim**. In simple terms, they showed that one could take any impartial game and analyze it in terms of Nim heaps that could grow or decrease in size. The theory was developed further by E. R. Berlekamp, John **Conway**, and others, and presented comprehensively in the books *Winning Ways for Your Mathematical Plays* and *On Numbers.* Sprague-Grundy theory has been applied to other combinatorial games, including kayles.

Sprouts

A pencil-and-paper game invented by John **Conway** and Michael S. Paterson at Cambridge University in 1967. Sprouts is played by two players, starting with a few dots (called *spots*) drawn on a sheet of paper. To make a move, a player draws a curve between two spots or a loop from a spot to itself; the curve may not cross any other curve. The player marks a new spot on the curve, dividing it in two. Each spot can have at most three curves connected to it. The player who makes the last move wins. Sprouts has been studied from the perspectives of **graph** theory and of **topology**. It can be proven that a game started with n spots will last at least $2n$ moves and at most $3n - 1$ moves. By enumerating all possible moves, one can show that the first player is guaranteed a win in games involving three, four, or five spots, while the second player can always win a game that starts with one, two, or six spots. Following a 1990 computer analysis of the game at Bell Labs out to 11 spots, David Applegate, Guy Jacobsen, and Daniel Sleator conjectured that the first player has a winning strategy when the number of spots divided by six leaves a remainder of three, four, or five. Sprouts was featured in the plot of the first part of the science fiction novel *Macroscope* by Piers Anthony.

square

(1) A **quadrilateral** with four equal sides that meet at right angles. (2) To square something is to multiply it by itself. To take the *square root* is the reverse process. A *square number* is the square of a whole number (see also **figurate number**). A well-known formula for the *difference of squares* is $a^2 - b^2 = (a + b)(a - b)$. This formula enables some otherwise difficult calculations to be done easily in the head. For example, $43 \times 37 = (40 + 3)(40 - 3) = 40^2 - 3^2 = 1600 - 9 = 1591$.

square free

An integer that is not divisible by a perfect square, n^2, for $n > 1$.

square pyramid problem

See **cannonball problem**.

square root of 2 ($\sqrt{2}$)

The first number shown to be what is now known as an **irrational number** (a number that can't be written in the form a/b, where both a and b are integers). This discovery was made by **Pythagoras** or, at any rate, by the Pythagorean group that he founded. The square root of 2 is the length of the hypotenuse (longest side) of a right **triangle** whose other two sides are each one unit long. A ***reductio ad absurdum*** proof that $\sqrt{2}$ is irrational is straightforward. Suppose that $\sqrt{2}$ is rational, in other words that $\sqrt{2} = a/b$, where a and b are coprime integers (that is, they have no common factors other than 1) and $b > 0$. It follows that $a^2/b^2 = 2$, so that $a^2 = 2b^2$. Since a^2 is even (because it has a factor of 2), a must be even, so that $a = 2c$, say. Therefore, $(2c)^2 = 2b^2$, or $2c^2 = b^2$, so b must also be even. Thus, in a/b, both a and b are even. But we started out by assuming that we'd reduced the fraction to its lowest terms. So there is a contradiction and therefore $\sqrt{2}$ must not be irrational. This type of proof can be generalized to show that any root of any natural number is either a natural number or irrational.

As a **continued fraction**, $\sqrt{2}$ can be written $1 + 1/(2 + 1/(2 + 1/(2 + \ldots)))$, which yields the series of rational approximations: 1/1, 3/2, 7/5, 17/12, 41/29, 99/70, 239/169, ... Multiplying each numerator (number on the top) by its denominator (number on the bottom) gives the series 1; 6; 35; 204; 1,189; 6,930; 40,391; 235,416; . . . which follows the pattern: $A_n = 6A_{n-1} - A_{n-2}$. Squaring each of these numbers gives 1; 36; 1,225; 41,616; 1,413,721; 48,024,900; 1,631,432,881; . . . each of which is also a **triangular number**. The numbers in this sequence are the only numbers that are both square and triangular.

squarefull number

See **powerful number**.

squaring the circle

The problem to find a construction (see **constructible**), using only a straightedge and compass, that would give a square of the same area as a given circle. It is now known that this is impossible because, as the German mathematician Ferdinand von Lindemann showed in 1882, it would amount to finding a **polynomial** expression for **pi**, which can't be done as π is a **transcendental number**. (All the coordinates of all points that can be constructed with ruler and compass are algebraic numbers.) This minor inconvenience, however, hasn't prevented some amateur mathematicians from continuing to claim they have found proofs that the circle can be squared. Some of the most bizarre attempts have involved proposing a different and rational value for π. In 1897, the Indiana

State Legislature came within a hair's breadth of introducing a bill to set the value of π equal to 3.2! In fact, long before π was finally proved to be transcendental, most learned societies had conjectured that a proof was impossible and had stopped considering circle-squaring arguments sent to them.

squaring the square

The problem of how to tile a square with *integral squares* (squares of integral side-length). Of course squaring the square is a trivial task unless additional conditions are set. The most studied restriction is the *perfect squared square:* a square such that each of the smaller squares has a different size. The name was coined in humorous analogy with **squaring the circle** and is first recorded as being studied by R. L. Brooks, C. A. B. Smith, A. H. Stone, and W. T. Tutte at Cambridge University. The first perfect squared square was found by Roland Sprague in 1939. If such a **tiling** is enlarged so that the formerly smallest tile becomes as big as the original square, it becomes clear that the whole plane can be tiled with integral squares, each having a different size. It is still an unsolved problem, however, whether the plane can be tiled with a set of integral square tiles such that each natural number is used exactly once as the size of a tile. A *simple squared square* is one where no subset of the squares forms a rectangle. The smallest simple perfect squared square was discovered by A. J. W. Duijvestin using a computer search. His tiling uses 21 squares, and has been proved to be minimal. Other possible conditions that lead to interesting results are nowhere-neat squared squares and no-touch squared squares. Developments leading to squaring the square can be traced back to 1902 and the first appearance of Henry **Dudeney**'s Lady Isabel's casket, later published as problem #40 in *The Canterbury Puzzles.*[87]

standard deviation

A measure of the spread of a **set** of data. For a **Gaussian** distribution, the standard deviation hints at the width of the tails of the distribution **function**.

Stanhope, Earl

A line of English earls, several of whom were notable mathematicians and polymaths. The third earl, Charles Stanhope (1753–1816), invented the process of stereotyping and, in 1777, devised the first mechanical logical calculator. The fourth earl, Philip Henry Stanhope (1781–1855) was a mathematician who applied his knowledge to mapmaking and **maze** design. From 1818 to 1830 he planted a maze according to a basic plan laid down by the second earl (1714–1786) that was probably the first to incorporate islands that defeat the familiar "hand-on-wall" method. (There are islands in the famous Hampton Court maze, yet the hand-on-wall method solves it.) His designs often had a number of islands, or just isolated lengths of hedge.

Star of David

Also known as a *hexagram,* the six-pointed star obtained by extending the sides of a regular **hexa**gon to the points of intersection.

star of Lakshmi

An eight-pointed star design often used in architecture, particularly as a tiling or other decoration on the floor of a room that has four- or eight-fold symmetry. A notable example is in the **octagonal** central lobby in the Houses of Parliament in London. It was used by Hindus to symbolize *Ashtalakshmi,* the eight forms of wealth.

state space

See **phase space**.

stationary point

A point on the graph of a **function** where the **tangent** to the graph is parallel to the *x*-axis or, equivalently, where the **derivative** of the function is 0. There are four kinds of stationary points: (1) a *local minimum,* where the derivative of the function changes from negative to positive; (2) a *local maximum,* where the derivative changes from positive to negative; (3) a *rising point of inflection,* where the derivative is positive on both sides of the stationary point; and (4) a *falling point of inflection,* where the derivative is negative on both sides of the stationary point.

statistical mechanics

The study of statistical and thermal properties of physical materials and their idealized mathematical models.

statistics

The study of ways that lots of data can be represented using a few numbers and the study of how such numbers can be chosen and used to draw reasonable conclusions about the data. The word *statistics* comes from the Latin *statis* for "political state"; one of the main tasks of the subject involves analyzing facts and figures about governments, resources, and populations. Although a powerful tool, statistics is open to abuse, both intentional and unintentional. Benjamin Disraeli (1804–1881) may have gone a little over the top when he said, "There are lies, damned lies, and statistics," but Scottish author Andrew Lang (1844–1912) could have been describing many a politician when he remarked, "He uses statistics as a drunken man uses lamp posts–for support rather than

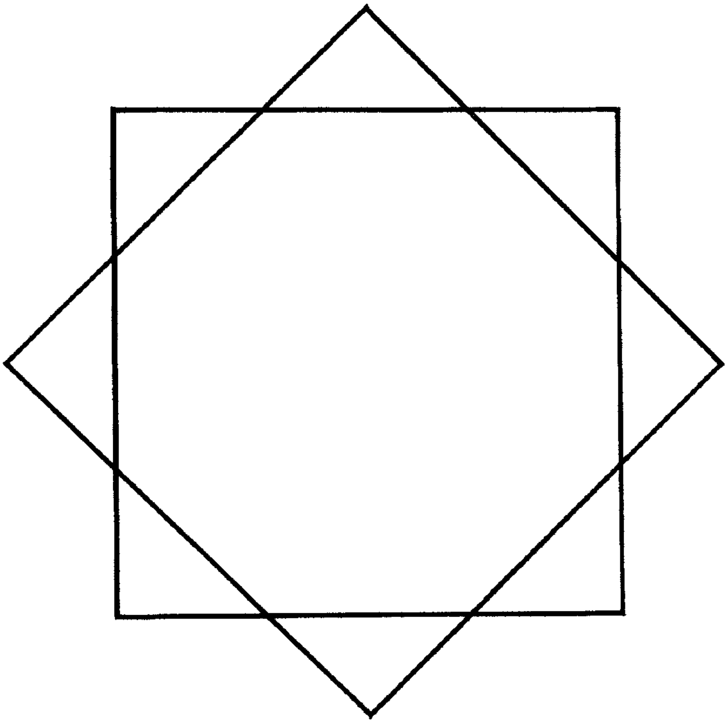

star of Lakshmi

illumination." The branch of statistics most commonly used in recreational mathematics is **probability theory.**

Steiner, Jakob (1796–1863)

A Swiss mathematician, considered by many to be the greatest geometer since **Apollonius** of Perga. Largely self-taught, he became a professor at the University of Berlin, and was a pioneer in the field of **projective geometry**. He had an important influence on his students, including Bernhard **Riemann**.

Steiner-Lehmus theorem

Any triangle that has two equal angle bisectors (each measured from a vertex to the opposite sides) is an **isosceles** triangle. In 1840, a Berlin professor Ludolph Lehmus wondered if this statement is true, given that it is the inverse of the already proven rule: If a triangle is isosceles then two of its internal bisectors are equal. He put the problem to Jakob **Steiner** who was quickly able to show its validity. Shortly after, Lehmus himself found a neater proof and it is has since become a favorite pastime of geometry hobbyists to search for still simpler proofs of the theorem.

Steinhaus, Hugo Dyonizy (1887–1972)

A Polish mathematician who was an influential member of the *Lvov school*, based at the Jan Kazimierz University in Lvov, which also included Stefan **Banach** and which focused on problems in functional **analysis**, real functions, and probability in the 1920s and '30s. Early on, Steinhaus's work revolved around applications of the **Lebesque** measure and integral. In 1923 he published the first rigorous account of the theory of tossing coins based on **measure theory**, and in 1925 was the first to define and discuss the concept of strategy in **game theory**. During World War II, as a Jew he was compelled to hide from persecution by the Nazis, yet continued his mathematical work despite great hardship. In 1944, Steinhaus proposed the problem of dividing a cake into *n* pieces so that it is proportional and envy free (see **cake-cutting**). He is also well known as the author of the widely read *Mathematical Snapshots*.[316]

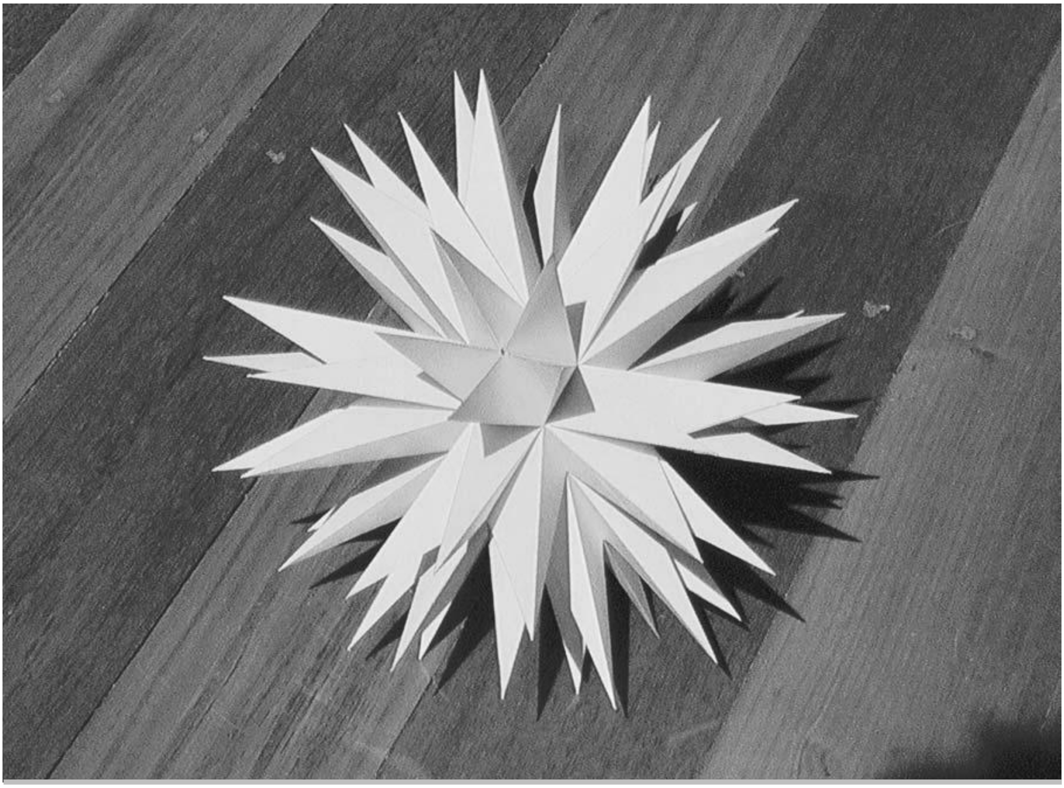

stellation The final stellation of the icosahedron. *Robert Webb, www.software3d.com; created using Webb's Stella program*

stellation

(1) The process of constructing a new **polyhedron** by extending the face planes of a given polyhedron past their edges. (2) The new polyhedron thus obtained. Starting with the **icosahedron**, for example, there are 59 possible stellations, including the *great icosahedron,* which is one of the **Kepler-Poinsot solids**.

steradian

The international (SI) unit in which **solid angle**s are measured. The steradian (sr) is a solid angle with its **vertex** in the center of a **sphere** and which cuts off from the spherical surface an area equal to the square of the radius of the sphere. The full sphere represents a spherical angle of 4π steradians.

Stewart, Ian (1945–)

A British mathematician at Warwick University who has written many books and articles on recreational math and popular science. From 1990 to 2001 he wrote the "Mathematical Recreations" column in *Scientific American.* His books include *Flatterland: Like Flatland Only More So* (a modern version of Edwin Abbott's classic ***Flatland: A Romance of Many Dimensions***); *Does God Play Dice: The New Mathematics of Chaos,* and *Evolving the Alien* (with Jack Cohen).[317, 318, 321, 323]

Stewart toroid

A **polygon** with a hole through it that meets some or all of the criteria specified by Bonnie Madison Stewart in his 1980 book *Adventures among the Toroids.* These criteria are as follows: all faces must be regular; faces that meet mustn't lie in the same plane; the polygon must be quasi-convex; the hole through the polygon must change its **genus**; and, the faces aren't allowed to intersect with themselves or each other. Stewart toroids, pierced relatives of the familiar Platonic solids, combine dazzling complexity with attractive symmetry.

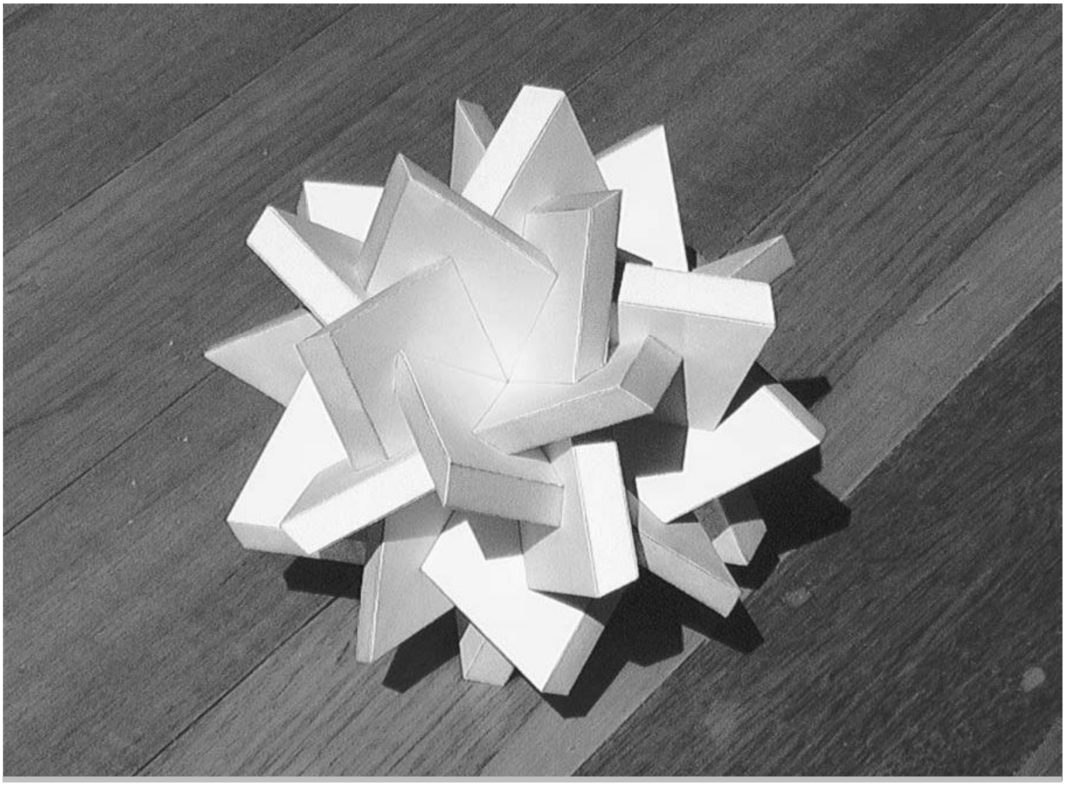

stellation A stellation of the small stellated truncated dodecahedron, a uniform polyhedron. *Robert Webb, www.software3d.com; created using Webb's Stella program*

stochastic
Something that is **random**.

stochastic process
A **dynamical system** with **random** fluctuations at each iteration or that is influenced by random noise. A random variable that, at each stage in time, depends on its previous values and on further random choices. For example, the price of a stock is often modeled as a stochastic process.

Stomachion
See **Loculus of Archimedes**.

straight
Having no deviations. A *straight line* is usually simply called a **line**. A *straight angle,* or *flat angle,* is exactly 180°.

strange attractor
See **chaotic attractor**.

strange loop
A phenomenon in which, whenever movement is made upward or downward through the levels of some hierarchical system, the system unexpectedly arrives back where it started. Douglas **Hofstadter** has used the strange loop as a paradigm in which to interpret paradoxes in logic, such as **Grelling's paradox** and **Russell's paradox**, and has called a system in which a strange loop appears a *tangled hierarchy*.

strategy
In **game theory**, a policy for playing a game. A strategy is a complete recipe for how a player should act in a game under all circumstances. A policy may employ **randomness**, in which case it is referred to as a **mixed strategy**.

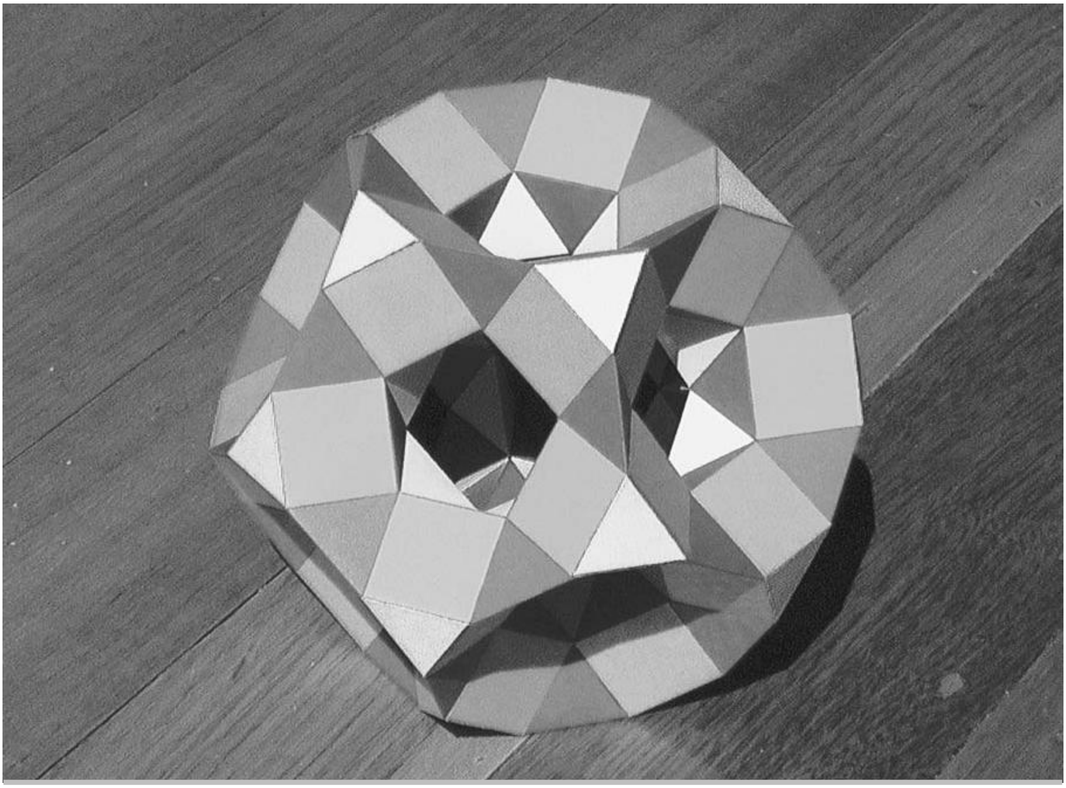

Stewart toroid A drilled truncated dodecahedron. *Robert Webb, www.software3d.com; created using Webb's Stella program*

string

Any sequence of letters, numbers, digits, **bit**s, or symbols.

string puzzle

The origins of puzzles using string, a form of *disentanglement puzzle,* are lost in antiquity but one of the earliest stories involving this type of problem is that of the **Gordian knot**. Today many kinds of string puzzles are available, most of which involve removing a ring from loops of string that are threaded through a solid (typically wooden) body. See also **mechanical puzzle**s and **knot**.

string theory

An important theory in modern physics in which the fundamental particles in nature are thought of as the "musical notes" or excitation modes of vibrating elementary strings. These strings have the shortest meaningful physical length, the so-called *Planck length* (equal to about 10^{-33} cm), and no thickness. Even more bizarre, for the theory to make sense, the universe must have nine space dimensions and one time dimension, for a total of 10 dimensions. This idea of a 10-dimensional universe first appeared in the **Kaluza-Klein theory**. We're familiar with time and three of the space dimensions: the other six together are known as **Calabi-Yau space**s. In string theory, as in the case of a stringed instrument, the string must be stretched under tension in order to vibrate. This tension is fantastically high–equivalent to a weight of about 10^{39} tons. String theories are classified according to whether or not the strings are required to be closed loops, and whether or not the particle spectrum includes *fermions.* In order to include fermions in string theory, there must be a special kind of symmetry called *supersymmetry,* which means that for every *boson* (a particle, of integral spin, that transmits a force) there is a corresponding fermion (a particle, of half-integral spin, that makes up matter). So supersymmetry relates the particles that transmit forces to the particles that make up matter. Supersymmetric partners to currently known particles have not been observed in particle experiments, but theorists believe this is because supersymmetric particles are too massive to be produced by present-day high-energy accelerators. Particle accelerators could be on the verge of finding evidence for high energy supersymmetry in the next decade. Evidence for supersymmetry at high energy would be compelling evidence that string theory was a good mathematical model for nature at the smallest distance scales. In string theory, all of the properties of elementary particles–charge, mass, spin, and so forth–come from the vibration of the string. The easiest of these to understand is mass. The more frantic the vibration, the more energy; and since mass and energy are the same thing, higher mass comes from faster vibration.

strobogrammatic prime

A **prime number** that remains unchanged when rotated through 180°. An example is 619, which looks the same when read upside-down. To be strobogrammatic, a prime cannot contain digits other than 0, 1, and 8, which have a horizontal line of symmetry (ignoring font variations), and 6 and 9, which are vertical reflections of each other. An *invertible prime* is one that yields a different prime when the digits are inverted. Of course, these definitions are not taken seriously by mathematicians!

strophoid

A looping curve, first studied in 1670 by Isaac Barrow (1630–1677), the first Lucasian Professor of Mathematics at Cambridge and the immediate predecessor of Isaac **Newton** in this job. The strophoid, which is a special case of the general **cissoid**, was named by Enrico Montucci in 1846 from the Latin for "twisted belt shape." The elaborate rules for drawing a strophoid are as follows. Let C be a curve, let O be a fixed point called the pole, and let O' be a second fixed point. Let P and P' be points on a line through O meeting C at Q such that $P'Q = QP = QO'$. The locus of P and P' is the strophoid of C with respect to the pole O and fixed point O'. If the curve C is a straight line, the pole P is not on C, and the second point O' is on C, the resulting strophoid is called an *oblique strophoid*. If these same conditions apply except that O' is the point where the perpendicular from O to C cuts C, then the strophoid produced is called a *right strophoid*. On the other hand if C is a circle, O is the center of the circle, and O' a point on its circumference, then the strophoid that results is known as *Freeth's nephroid*. The French mathematician Gilles Roberval (1602–1675) found the strophoid in a different way–as the result of planes cutting a **cone**. When the plane rotates (about the tangent at its vertex) the collection of foci of the obtained conics gives the strophoid.

subgroup

A subset of a **group** that is a group under the same operation.

sublime number

A number such that both the *sum* of its divisors and the *number* of its divisors are **perfect numbers**. The smallest sublime number is 12. There are 6 divisors of 12–1, 2, 3, 4, 6, and 12–the sum of which is 28. Both 6 and 28 are perfect. The second sublime number begins 60,865 . . . , ends . . . 91,264, and has a total of 76 digits! It is not known if there are larger even sublime numbers, nor if there are any odd sublime numbers.

subset

A **set** whose members are members of another set; a set contained within another set.

substitution cipher

A **cipher** that replaces each plaintext (original message) symbol with a ciphertext (coded text) symbol. The receiver decodes using the inverse substitution. A simple example is the **Caesar cipher**.

subtraction

The **binary operation** of finding the difference between two quantities or numbers.

sultan's dowry

A sticky problem in **probability** that first came to light in Martin **Gardner**'s "Mathematical Recreations" column in the February 1960 issue of *Scientific American*. Gardner's original version has become known as the secretary problem. In the exactly equivalent form called the sultan's dowry problem, a sultan has granted a commoner the chance to marry one of his hundred daughters. The commoner will be shown the daughters one at a time and will be told each daughter's dowry. The commoner has only one chance to accept or reject each daughter; he can't go back and choose one that he has previously rejected. The sultan's catch is that the commoner may only marry the daughter with the highest dowry. What is the commoner's best strategy, assuming that he knows nothing in advance about the way the dowries are distributed?

Many mathematicians have tackled this question and numerous papers have been written on the subject. It has even spawned its own area of study within the field of management science. The consensus among those who have worked on the problem is that the commoner's best strategy is to let a certain fraction of the daughters pass and then choose the next one who has a dowry higher than any of the ones seen up to that point. The exact number to skip is determined by the condition that the odds that the highest dowry has already been seen is just greater than the odds that it remains to be seen *and that if it is seen it will be picked.* This amounts to finding the smallest x such that:

$$x/n > x/n \times [1/(x+1) + \ldots + 1/(n-1)]$$

Substituting $n = 100$ leads to the conclusion that the commoner should wait until he has seen 37 of the daughters, then pick the first daughter with a dowry that is bigger than any that have already been revealed. With this strategy, his odds of choosing the daughter with the highest dowry are surprisingly high: about 37%.[224]

supercomputer

A computer that, in its day, is faster than any contemporary conventional computer. Supercomputers are typically used for enormous number-crunching tasks such as weather-forecasting, simulations of fusion experiments or galaxy evolution, design of cars and planes, and code-cracking. At the start of 2004, the world's most powerful computer, by a long way, is the Earth Simulator in Japan, which is capable of a maximum of 40,960 gigaflops, that is, 40.96 trillion "floating point operations" per second. The top 500 supercomputers are listed at the following Web site: http://www.top500.org.

superegg

The **surface of revolution** of a **superellipse** given by the formula $|x/a|^{2.5} + |y/b|^{2.5} = 1$, where $a/b = 4/3$. The superegg was named by Piet **Hein** and singled out by him because of an unusual property: stood on either end it has a peculiar and surprising stability. Supereggs, made of metal, wood, and other materials, were sold as novelties in the 1960s; small, solid-steel ones were marketed as an "executive toy." The world's largest superegg, made of steel and aluminum and weighing 1 ton, was placed outside Kelvin Hall in Glasgow in 1971 to honor Hein's appearance there as a speaker.

superegg

superellipse

A **Lamé curve**, described by the formula $|x/a|^n + |y/b|^n = 1$, for which $n > 2$. Superellipses have a form partway between an **ellipse** and a rounded rectangle (or, if $a = b$, partway between a circle and a rounded square). The Danish poet and architect Piet **Hein** decided that the superellipse with $n = 5/2$ and $a/b = 6/5$ is the most pleasing to the eye. This so-called *Piet Hein ellipse* was quickly adopted as the basic motif for planning an open space at the center of Stockholm and was also incorporated into Scandinavian designs for office tables, desks, beds, and even roundabouts in roads. The **surface of revolution** of a superellipse is a *superellipsoid,* one special form of which has been nicknamed the **superegg**.

superfactorial

A **function** based on the **factorial** that produces **large numbers** very quickly. It is a recent creation that has not yet entered the mathematical mainstream, and has been defined in a couple of different ways. In 1995, in his book *Keys to Infinity,* Clifford Pickover gave the superfactorial as

$$n\$ = \underbrace{n!^{n!^{\cdot^{\cdot^{\cdot n!}}}}}_{n!\text{ terms}}$$

In the same year, Sloane and Plouffe offered an alternative definition:

$$n\$ = \prod_{i}^{n} i!$$

Pickover's superfactorial grows with extraordinary speed. The first two terms are 1 and 4, but the third term 3$, already has more digits than could be written down on this page. Sloan and Plouffe's factorial grows sedately by comparison–the first few values are 1, 1, 2, 12, 288, and 345,690–and is related to **Bell numbers**.

supersymmetry

A theory in physics that postulates a counterintuitive symmetric relationship between fermions, which are particles such as electrons that obey the Pauli exclusion principle and thus cannot occupy the same quantum state, and bosons, which are particles such as photons that can coexist in the same state.

supertetrahedral number

A type of **figurate number** in four dimensions (see **fourth dimension**). Supertetrahedral numbers are obtained by piling up the **tetrahedral number**s 1, 4, 10, 20, 35, etc., as in:

1 = 1
1 + 4 = 5
1 + 4 + 10 = 15
1 + 4 + 10 + 20 = 35
1 + 4 + 10 + 20 + 35 = 70
etc.

supplementary angles

Two angles that add up to 180°.

surd

A **radical** that expresses an **irrational number**. Surds may be *quadratic* (e.g., $\sqrt{2}$), *cubic* (e.g., $\sqrt[3]{2}$), *quartic* (e.g., $\sqrt[4]{2}$), and so on. (The term is sometimes used as a synonym for irrational number.) A *pure surd,* or entire surd, contains no **rational number**; that, is, all its factors or terms are surds (e.g., $\sqrt{2} + \sqrt{3}$). A *mixed surd* contains at least one rational number (e.g., $2 + \sqrt{3}$ or $3\sqrt{2}$).

surface

In mathematics, any object that locally (if you zoom in close enough to it) looks like a piece of a flat **plane**. A **sphere**, a **torus**, a **pseudosphere**, and a **Klein bottle** are examples of different types of surfaces.

surface of revolution

A surface produced by rotating a line or a curve about some axis. For example, a **sphere** is the surface of revolution generated when a circle spins about its diameter.

surreal number

A member of a mind-boggling vast class of numbers that includes all of the **real number**s, all of Georg Cantor's infinite **ordinal number**s (different kinds of infinity), a set of **infinitesimals** (infinitely small numbers) produced from these ordinals, and strange numbers that previously lived outside the known realm of mathematics. Each real number, it turns out, is surrounded by a "cloud" of surreals that lie closer to it than do any other real numbers. One of these surreal clouds occupies the curious space between zero and the smallest real number greater than zero and is made up of the infinitesimals.

Surreal numbers were invented or discovered (depending on your philosophy) by John **Conway** to help with his analysis of certain kinds of games. The idea came to him after watching the British **Go** champion playing in the mathematics department at Cambridge. Conway noticed that endgames in Go tend to break up into a sum of games, and that some positions behave like numbers. He then found that, in the case of infinite games, some positions behaved like a new kind of number–the surreals. The name "surreal" was introduced by Donald **Knuth** in his 1974 book *Surreal Numbers: How Two Ex-Students Turned on to Pure Mathematics and Found Total Happiness.*[191] This novelette is notable as being the only instance where a major mathematical idea has been first presented in a work of fiction. Conway went on to describe the surreal numbers and their use in analyzing games in his 1976 book *On Numbers and Games.*[67] The surreals are similar to the **hyperreal number**s, but they are constructed in a very different way and the class of surreals is larger and contains the hyperreals as a subset.

swastika

An ancient ideogram, signs of which have been found in the Euphrates-Tigris Valley, and in some areas of the Indus Valley, dating back 3,000 years; it became commonly used

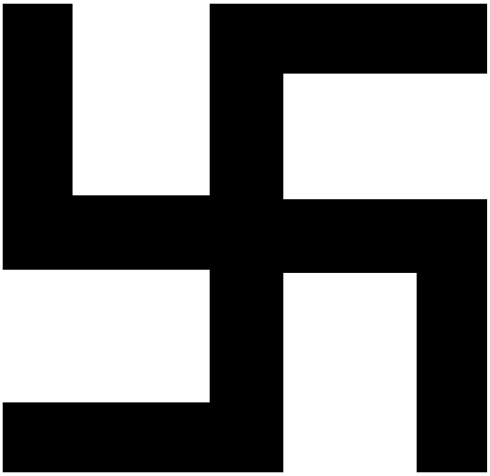

swastika

around 1000 B.C., possibly first in ancient Troy, in the northwest of modern Turkey. The swastika is an irregular icosagon (a 20-sided **polygon**), which in Arabic and Indian culture originally represented good luck. Of course, in more recent times, it was adopted as the symbol of the Nazi Party in Hitler's Germany and thus came to stand for anti-Semitism.

syllogism

An argument composed of three parts–a major premise, a minor premise, and a conclusion. For example: All men are mortal (major premise). Socrates is a man (minor premise). Therefore, Socrates is mortal (conclusion). The syllogism forms the basis of Aristotle's system of logic, which went unchallenged for over 2,000 years. Aristotle believed that by setting out any argument in syllogistic form, it should be possible to avoid fallacies. However, Bertrand **Russell** discovered several formal errors in the doctrine of syllogism.

Sylvester, James Joseph (1814–1897)

An English mathematician and lawyer who, in 1850, first used the term ***matrix*** in mathematics and gave it its present meaning of a rectangular array of numbers from which **determinant**s may be formed. Together with Arthur **Cayley**, he founded the *theory of invariants*. Hotheaded and vociferous in opposition of anti-Semitism, he was thrown out of the University of London for threatening another student with a table knife. Later he studied at Cambridge, emerging as Second Wrangler but without a degree (though in 1871 he earned an MA) because, as a Jew, he refused to accept the articles of the Church of England. After a spell working as an actuary and a barrister, and also taking private pupils (one of whom was Flor-ence Nightingale), he met Cayley, with whom he forged a lifelong friendship and collaboration, and returned to mathematics. He became professor of mathematics at the Royal Military Academy at Woolwich (1855–1870) and at the newly established Johns Hopkins University at Baltimore (1877–1883), founding the *American Journal of Mathematics,* before accepting the Savilian chair at Oxford (1883–1894). Remarkably, especially in the field of mathematics, he produced an extraordinary flood of ideas well into his old age. At 82, he worked out the theory of compound partitions. He also published on the roots of **quintic** equations and on **number theory**. His partnership with Cayley worked perfectly, since Cayley supplied the rigor which the brilliantly creative Sylvester lacked.

Sylvester's problem of collinear points

A problem posed in 1893 by James **Sylvester** who wrote: "Prove that it is not possible to arrange any finite number of real points so that a right line through every two of them shall pass through a third, unless they all lie in the same right line." No correct proof was forthcoming at the time, but the problem was revived by Paul **Erdös** in 1943 and correctly solved by T. Grünwald in 1944.

symmedian

Reflection of a **median** of a triangle about the corresponding angle **bisector**.

symmetric group

The **group** of all **permutation**s of a finite set. The symmetric group of a **set** of size n is denoted S_n and has $n!$ elements.

symmetry

An intrinsic property of a mathematical object that allows it to remain unchanged under certain types of **transformation**, such as rotation, reflection, or more abstract operations. The mathematical study of symmetry is systematized and formalized in the extremely powerful subject known as **group** theory. Symmetries and apparent symmetries in the laws of nature have played a part in the construction of physical theories since the time of Galileo and Newton. The most familiar symmetries are spatial or geometric ones. In a **snow**flake, for example, the presence of a symmetrical pattern can be detected at a glance. One of the most remarkable developments of the past half century has been the emergence of symmetry as a central theme of subatomic physics. This came about through a series of subtle evolutions in the concept of symmetry itself. Many researchers believe that this evolutionary process has not come to an end, and that further meaning of the concept of symmetry, with perhaps new mathematical structures, will develop in the coming years.

symmetry group

The **group** formed by the set of all rigid motions (translations, rotations, reflections, etc.) of Euclidean space that map all points of a subset F into F.

system

Something that can be studied as a whole. Systems may consist of subsystems that are interesting in their own right. Or they may exist in an environment that consists of other similar systems. Systems are generally understood to have an internal state, inputs from an environment, and methods for manipulating the environment or themselves. Since cause and effect can flow in both directions of a system and environment, interesting systems often possess **feedback**, which is self-referential in the strongest case.

Szilassi polyhedron

A **toroid**al heptahedron (seven-sided **polyhedron**) first described in 1977 by the Hungarian mathematician Lajos Szilassi. It has 7 faces, 14 vertices, 21 edges, and 1 hole. The Szilassi polyhedron is the **dual** of the **Császár polyhedron** and, like it, shares with the **tetrahedron** the property that each of its faces touches all the other faces. Whereas a tetrahedron demonstrates that four colors are necessary for a map on a surface topologically equivalent to a sphere, the Szilassi and Császár polyhedra show that seven colors are necessary for a map on a surface topologically equivalent to a **torus**.

T

tachyon

A hypothetical particle that travels faster than the speed of light. Tachyons were first proposed in prerelativistic times by the physicist Arnold Sommerfeld and named in the 1960s by Gerald Feinberg from the Greek *tachys* meaning "swift." By extension of this terminology, particles that travel slower than light are called *tardyons* (or *bradyons* in more modern usage) and particles, such as photons, that travel exactly at the speed of light are called *luxons*. The existence of tachyons is allowed by the mathematics of special **relativity theory**, one of the basic equations of which is

$$E = m/\sqrt{(1 - v^2/c^2)}$$

where E is the mass-energy of a particle, m its rest mass, v its velocity, and c the speed of light. This shows that for tardyons (particles of ordinary matter), E increases as v increases and becomes infinite when $v = c$, thus preventing an initially slower-than-light particle from being accelerated up to the speed of light and beyond. What about a particle for which v is always greater than c? In this case, $v^2/c^2 > 1$, so that the denominator in the equation above is an **imaginary number**–the square root of a negative real number. If m has a real value, E is imaginary, which is hard for physicists to swallow because E is a measurable quantity. If m takes an imaginary value, however (because one imaginary number divided by another is real), then E is real. Tachyons are allowed, therefore, providing (a) they never cross from one side of the light-barrier to the other, and (b) they have an imaginary rest mass (which is physically more acceptable, since the rest mass of an object that never stops isn't directly measurable). Tachyons would slow down if they lost energy, and accelerate if they gained energy. This leads to a problem in the case of charged tachyons because charged particles that move faster than the speed of light in the surrounding medium give off energy in the form of Cherenkov radiation. Charged tachyons would continuously lose energy, even in a vacuum, through Cherenkov emission. This would cause them to gain speed, thus lose energy at an even greater rate, thus accelerate even more, and so on, leading to a runaway reaction and the release of an arbitrarily large amount of energy.

More worrisome, as the physicist Gregory Benford and his colleagues first pointed out in their 1970 paper "The Tachyonic Antitelephone," tachyons seem to lead to a **time travel** paradox because of their ability to send messages into the past. Suppose Alice on Earth and Boole on a planet circling around Sirius can communicate using what has been called a tachyon "antitelephone." They agree in advance that when Boole receives a message from Alice, he will reply immediately. Alice promises to send a message to Boole at noon her time, if and only if she has not received a message from Boole by 10 A.M. The snag is that both messages, being superluminal, travel back in time. If Alice sends her message at noon, Boole's reply could reach her before 10 A.M. "Then," as Benford and colleagues wrote, "the exchange of messages will take place if and only if it does not take place. . . ." Perhaps not surprisingly, despite numerous searches, no tachyon detection has so far been confirmed. The same is true of another hypothetical faster-than-light particle called a dybbuk (Hebrew for a "roving spirit"), which would have imaginary mass, energy, and momentum. Dybbuks, proposed by Raymond Fox of the Israel Institute of Technology, have properties even stranger than those of tachyons yet, interestingly, they avoid the **causality** problem that affects their superluminal cousins.

TacTix

A two-player game of strategy, devised by Piet **Hein**, that is essentially a two-dimensional version of **Nim**. Though the game is nontrivial, the first player can always win, at least in the 5 × 5 matrix version by choosing the center piece and symmetrically mirroring the second player's moves. On a 5 × 5 grid, players alternate taking away as many contiguous pieces as desired from a single row or column.

Tafl game

A type of board game in which the contest is between two forces of unequal number or strength. The earliest form of Tafl (old Norse for "table"), known as **Hnefa-Tafl** ("king's table"), originated in Scandinavia before A.D. 400 and was then exported by the Vikings to Greenland, Iceland (where it is mentioned in the *Grettis Saga* dating back to A.D. 1300), Ireland, England, Wales, and as far east as the Ukraine. Several boards unearthed in both Viking and Anglo-Saxon contexts, including the board found at the Gokstad ship burial, have had

Hnefa-Tafl on one side and **nine men's morris** on the other. Later variants of Tafl include Tabula (the medieval ancestor of **backgammon**, introduced from the French as *Quatre* and thus *Kvatru-Tafl*), **fox and geese** (*Ref-Skak,* "fox chess", *Hala-Tafl* or *Freys-Tafl*), **three men's morris** (*Hrœ-Tafl,* "Quick-Tafl"), and **nine men's morris.**

Tait, Peter Guthrie (1831–1901)

A Scottish scientist and mathematician who carried out the world's first systematic investigation of **knot** theory. Early in his career he formed a friendship with William **Hamilton** and became fascinated in the application of Hamilton's **quaternion**s to problems in physics. In 1857, he also took an interest in Hermann Helmholtz's theories on the behavior of vortex rings, and began experimenting with smoke rings and their interactions. These experiments greatly impressed William Thomson (Lord Kelvin) who saw in them a possible way (wrong, as we now know) to explain atomic structure and the buildup of different elements. This idea, in turn, led Tait, Thomson, and James Maxwell to do seminal work on knot theory, since the basic building blocks in Thomson's vortex atom model were rings knotted in three dimensions. Without any rigorous theory, which would have been well beyond nineteenth-century mathematics, Tait began to classify knots using his geometric intuition. By 1877 he had classified all knots with seven crossings. He then went on to consider the coloring of **graph**s and put forward a hypothesis (see **Tait's conjecture**) that, if true (which it wasn't), would have proved the **four-color map problem**. Among his many other accomplishments, Tait wrote a classic paper on the trajectory of golf balls (1896). This was a subject close to his heart because the third of his four sons was Frederick Gutherie Tait, the leading amateur golfer in 1893 and winner of the Open Golf Championship in 1896 and 1898.

Tait's conjecture

A hypothesis put forward by Peter **Tait** in 1884, which says that every **polyhedron** has a **Hamilton circuit** through its vertices. In other words, it is possible to travel around all the edges of a polyhedron, passing through each vertex (corner) exactly once and arriving back at the starting point. If true, Tait's conjecture would have provided an immediate proof of the **four color map problem**. However, in 1946, the British mathematician William Tutte (1917–2002), whose work at Bletchley Park on cracking the German FISH cipher played an important role in World War II, found a counterexample to the conjecture in the form of a polygon with 25 faces, 69 edges, and 46 vertices.

tally

To count or keep score. In the past this was often done by making marks on a stick; the word comes from the Latin *talea* meaning "one who cuts," which is also the root of *tailor.* The oldest known tally stick is thought to be the **Lebombo bone** dating back about 37,000 years. Until around 1828 British tax records were kept on wooden tally sticks. When the system was finally abandoned, the government was left with a mountain of wood which, in 1834, it decided to dispose of by having a giant bonfire. So successful was that blaze that it also burned down the parliament buildings. What Guy Fawkes had failed to do with dynamite, the Exchequer did with tally sticks!

tangent

(1) A straight line that touches a given curve exactly once, at a given point. (2) In a right triangle, if one of the angles is θ, then the tangent of θ is the ratio of lengths of the side opposite θ to the side next to θ. See also **trigonometric function**.

tangle

A system in which a **strange loop** appears.

tangled graph

A **graph** in 3-dimensional space; equivalently, a graph drawn in the plane so that when edges cross, one edge goes over the other.

Tangloids

A mathematical game for two players devised by Piet **Hein** to model the calculus of **spinor**s. Two flat blocks of wood each pierced with three tiny holes are joined with three parallel strings. Each player holds one of the blocks of wood. The first player holds one block of wood still, while the other player rotates the other block of wood around any axis for 4π radians (two full revolutions). Then the first player tries to untangle the strings without rotating either piece of wood. Only translations (sliding the pieces) are allowed. Afterward, the players reverse roles; whoever can untangle the strings fastest is the winner.

tangrams

A puzzle of Chinese origin, the objective of which is to form given shapes using a set of seven pieces (five triangles of various sizes, one square, and one parallelogram) that come from slicing up a square. The produced shape has to contain all the pieces, which mustn't overlap. Tangrams became popular in England around the middle of the nineteenth century, having been brought back by sailors from Hong Kong. It received a further boost when Lewis **Carroll** used the pieces to create illustrations of the characters in the Alice books. The origin of the name

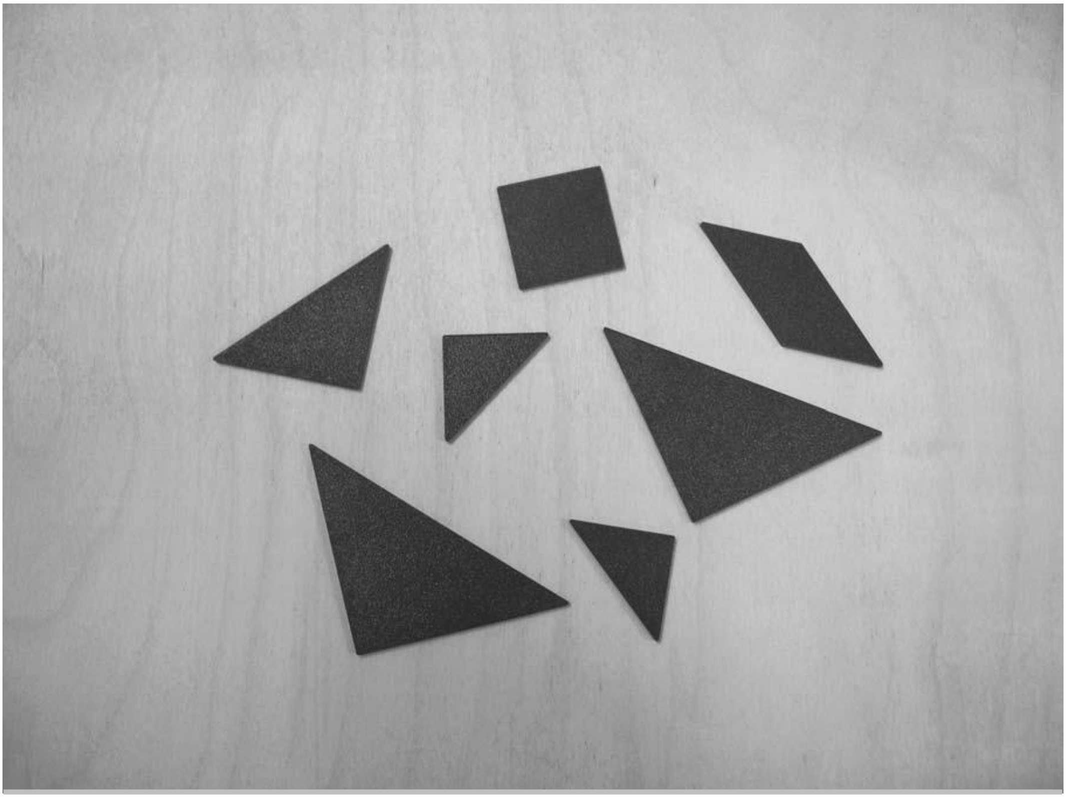

tangrams The seven tangrams. *Kadon Enterprises, Inc., www.gamepuzzles.com*

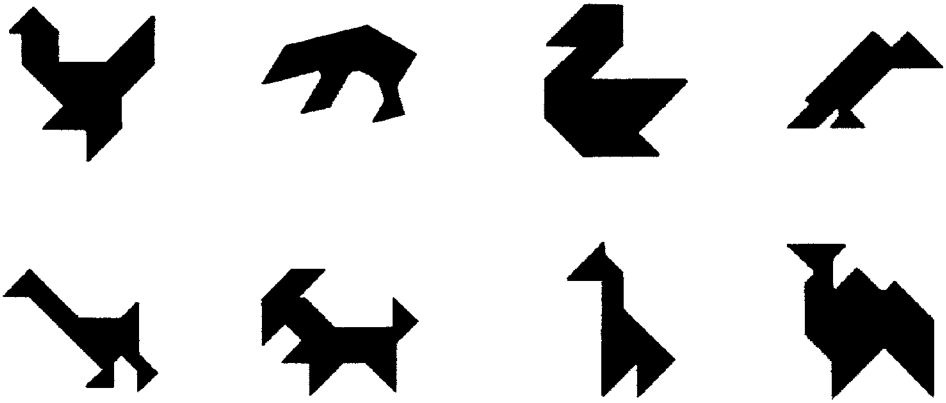

tangrams A tangram menagerie built up from the seven simple shapes.

isn't certain. One theory is that it comes from the Cantonese word for chin. A second is that it stems from a mispronunciation of a Chinese term that the sailors used for the ladies of the night from whom they learned the game! A third suggestion is that it is from the archaic Chinese root for seven, which still persists in the Tanabata festival held on July 7 in Japan. The first definitive appearance of tangrams in the Far East seems to be in Japan in the early eighteenth century; by around 1805 it had become a fad in both China and Europe. The **loculus of Archimedes** is a similar game and it has been suggested, though without evidence, that this was a direct precursor of tangrams, having been transmitted to the East via Arab sources.[93, 265]

Tarry, Gaston (1843–1913)

A French civil servant and amateur mathematician who spent the whole of his working career in Algeria. He published numerous articles on geometry, **number theory**, and **magic square**s from 1882 until his death, and is best known for his contribution to Euler's **thirty-six officers problem**. He also published an algorithm for exploring **maze**s that is named after him.

Tarski, Alfred (1902–1983)

A Polish-American mathematician who, along with Aristotle, Friedrich **Frege**, and Kurt **Gödel** is considered one of the greatest logicians of all time and certainly the most prolific: his collected works, excluding books, runs to 2,500 pages. Tarski made important contributions in many areas of mathematics, including **set theory**, **topology**, algebraic **logic**, and metamathematics. His most important contribution to logic is the *semantic method*–a technique that allows a more exacting study of formal scientific languages.

Tartaglia, Niccoló Fontana (1499–1557)

An Italian mathematician who, along with Girolamo **Cardano**, discovered the algebraic solution of the **cubic**. He was also a well-known inventor of mathematical recreations. He devised many arithmetical problems, and contributed especially to **measuring and weighing puzzles** and to **river-crossing problem**s. Although his real name was Niccoló Fontana, he is always referred to as Tartaglia, "the stammerer."

tautochrone problem

Find the curve down which an object can slide from any point to the bottom (accelerated by gravity and ignoring friction), always in the same length of time. *Tautochrone* comes from the Greek *tauto* for "the same" (which also gives us ***tautology***) and *chronos* for "time." The solution, first found by Christiaan **Huygens** and published in his *Horologium oscillatorium* (1673), is a **cycloid**. Thus, if you were to upturn a cycloid, in the manner of an inverted arch, and then release a marble from any point on it, it would take exactly the same time to reach the bottom, no matter where on the curve you started. Huygens used his discovery to design a more accurate pendulum–one with curved jaws from the point of support that forced the string to follow the right curve no matter how large or small the swing. The cycloid's unique property is mentioned in the following passage from Herman Melville's *Moby Dick:* "[The try-pot] is also a place for profound mathematical meditation. It was in the left-hand try-pot of the *Pequod,* with the soapstone diligently circling round me, that I was first indirectly struck by the remarkable fact, that in geometry all bodies gliding along a cycloid, my soapstone, for example, will descend from any point in precisely the same time." The cycloid is also the curve that answers the **brachistochrone problem**.

tautology

In mathematics, a logical statement in which the conclusion is equivalent to the premise. According to the school of thought known as *logicism,* all of mathematics is derived from **logic** and is thus inherently tautological. Tautology is also the needless, pointless, meaningless, and unwarranted repetition of words and phrases that mean the same thing. Examples are to be found in the previous sentence and the next one. Have a slice of pizza (*pizza* is Italian for "slice") and spin the **roulette** wheel (*roulette* is French for "little wheel").

ten

The **base** of our familiar **number system**, which stems directly from the fact that we have 10 fingers on which to count. Ten is the only **triangular number** that is a sum of consecutive odd squares ($10 = 1^2 + 3^2$) and the only composite integer such that all of its positive integer divisors other than 1 are of the form $x^2 + 1$ ($2 = 1^2 + 1$, $5 = 2^2 + 1$, $10 = 3^2 + 1$). Strange but true: the life span of a taste bud is 10 days.

tensor

A generalization of the concept of a **vector**. A *scalar* is a tensor of rank zero and a vector is a tensor of rank one. There are tensors of rank two, three, and so on, used mainly in manipulating and transforming sets of equations within and between different coordinate systems. A tensor of order n has n^2 components and is often best thought of as an array of values that can be manipulated like a **matrix**. To illustrate a tensor of rank two, imagine a plane surface area with a force acting on it. The total effect depends on two things: the magnitude and direction of the force, and the size of the area and its orientation. In fact this latter property can be represented uniquely by a vector of magnitude proportional to the size of the area, and in a direction normal to the area. So the effect of the force upon the surface depends on two vectors and is equivalent to a tensor of rank two. Tensors were used by Einstein in deriving his law of gravitation in general **relativity theory**.

tera-
Prefix for 10^{15}, from the Greek *pentakis* for "five times."

terminating decimal
A decimal fraction that comes to an end, as in 0.3 or 0.7194. All terminating decimals are **rational numbers**, but not all rational numbers have terminating decimal expansions. For example, ⅓ is rational, but its decimal expansion goes on forever.

ternary
(1) Having the **base** three. (2) Involving three **variables**. From the Latin *ternarius* ("three each").

tessellation
See **tiling**.

tesseract
The four-dimensional analogue of a cube, also known as a 4-space **hypercube** or an 8-cell. The name, possibly coined by Charles **Hinton**, comes from *tesser,* meaning "four," and *aktis,* meaning "ray," thus "four rays." Just as a cube is obtained by "thickening" a square in the third dimension, which can be imagined as stacking infinitely many infinitely thin sheets of paper, a tesseract is a cube thickened in the **fourth dimension**. We can't imagine this because we can't think four dimensionally but it is possible to appreciate that just as perspectives of cubes can be drawn on a 2-d surface, so real cubes can serve as perspectives of tesseracts. A square drawn inside a larger square with the vertices connected by lines is one way to provide a perspective of a cube. Similarly, a hypercube is sometimes portrayed as a small cube within a larger cube with lines drawn from the vertices of the smaller cube to the vertices of the larger cube. This kind of representation is a bit misleading, however, and reveals very little of the nature of a tesseract. It doesn't show, for instance, how a tesseract can be subdivided into smaller 4-d blocks in the same way that a cube can be divided into smaller cubes, or a square into smaller squares. A more useful way to think of a tesseract is as a folding, in the fourth dimension, of a 3-d net of eight cubes, just as a cube is a folding in the third dimension of a 2-d net of six squares. Start with a stack of four cubes, with four more cubes arranged in a cross around the second cube from the top. A tesseract is made by folding (in the fourth dimension) so that the top face of the cube at the top of the stack merges with the bottom face of the bottom cube, and so that the adjacent edges of the cubes in the cross join. (See table, "A Comparison of the Square, Cube, and Tesseract.")

A Comparison of the Square, Cube, and Tesseract

	Vertices	Edges	Squares	Cubes
Square	4	4	1	–
Cube	8	12	6	1
Tesseract	16	32	24	8

A tesseract is bounded by eight hyperplanes, each of which intersects it to form a cube. Two cubes, and so three squares, intersect at each edge. There are three cubes meeting at every vertex, the vertex polyhedron of which is a regular tetrahedron, leading to the **Schläfli symbol** {4,3,3}. The distance between opposite corners of a hypercube is twice the length of a side–much tidier than the corresponding values of $\sqrt{2}$ for a square and $\sqrt{3}$ for a cube.

If a cube is hung from one of its vertices and sliced horizontally through its center, the result is a hexagon. What if the same is done to a tesseract? The slice will yield a 3-dimensional object–but what kind? The answer is an octahedron. By analogy with the slice of the 3-cube, the slice of the 4-cube must cut every "face." The number of faces of a 4-cube is eight and the only regular eight-sided solid is an octahedron.

Tesseracts turn up in both art and literature. Salvador Dali's *Christus Hypercubus* shows Christ being crucified on a tesseract. In Robert Heinlein's short story "And He Built a Crooked House" (1940), a house built as a three-dimensional projection of a tesseract collapses, or folds up, to become a real tesseract–with unusual consequences for the person trapped inside. The tesseract is also mentioned in Madeleine l'Engles children's fantasy *A Wrinkle in Time* as a way of introducing the concept of higher dimensions.[161, 338]

tetragon
The less familiar name for what's normally called a **quadrilateral**.

tetrahedral number
A number that can be made by considering a tetrahedral pattern of beads in three dimensions. For example, if a tri-

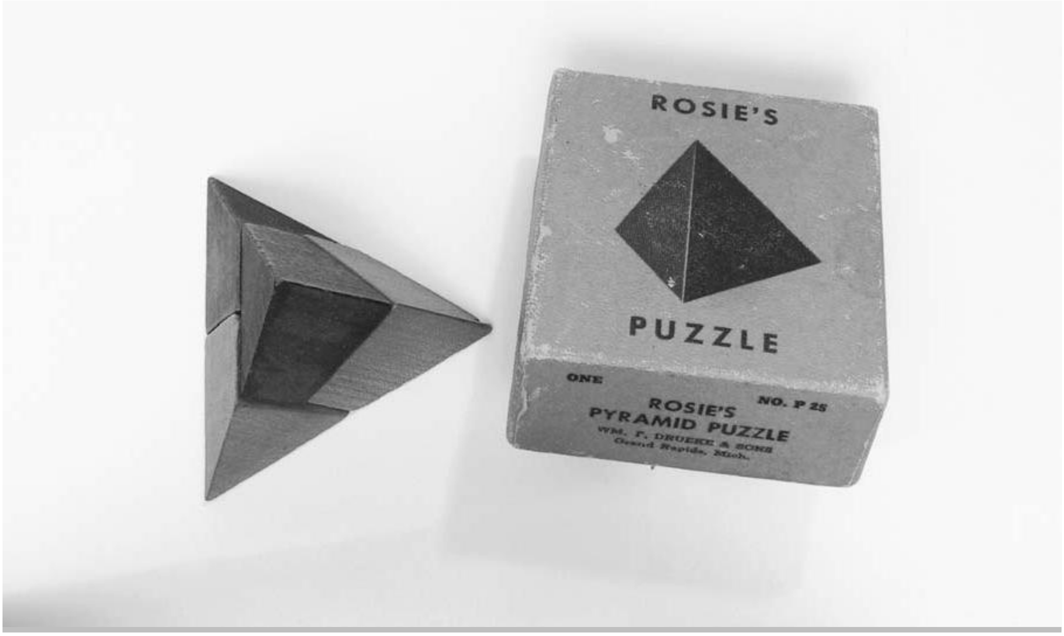

tetrahedron Rosie's pyramid puzzle, a vintage game that calls for four pieces to be assembled into a tetrahedron. *Sue & Brian Young/Mr. Puzzle Australia, www.mrpuzzle.com.au*

angle of beads is made with three beads to a side, and on top of this is placed a triangle with two beads to a side, and on top of that a triangle with one bead to a side, the result is a tetrahedron of beads. In this case the total number of beads is (third triangular number) + (second triangular number) + (first triangular number) = 6 + 3 + 1 = 10. In general the *n*th tetrahedral number is equal to the sum of the first *n* triangular numbers. This is the same as the fourth number from the left in the $(n + 3)$th row of **Pascal's triangle**. We can use the **binomial** formula for numbers in Pascal's triangle to show that the *n*th tetrahedral number is ${}^{n+2}C_3$, or $(n + 2)(n + 1)n/6$. The only numbers that are both tetrahedral and square are 4 ($= 2^2 = T_2$) and 19,600 ($= 140^2 = T_{48}$).

tetrahedron

A four-sided **polyhedron**. A regular tetrahedron, one of the **Platonic solids**, is a regular three-sided pyramid in which the base-edges and side-edges are of equal length. The projection of a regular tetrahedron can be an equilateral triangle or a square. The centers of the faces of a tetrahedron form another tetrahedron.

tetraktys

The sum $1 + 2 + 3 + 4 = 10$ (the fourth **triangular number**), held in reverence by the Pythagoreans (see **Pythagoras of Samos**). The tetraktus, or "holy fourfoldness," was taken to represent the four elements: fire, water, air, and earth.

Tetris

A video and computer game, invented in 1985 by the Russian Alexey Pajitnov, that has become one of the most widely played games of all time. In 2002, computer scientists Erik Demaine, Susan Hohenberger, and David Liben-Nowell of the Massachusetts Institute of Technology (MIT) analyzed the game to determine its computational complexity and found it to be an **NP-hard problem** (one that is immune to simple solution and instead demands exhaustive analysis to work out the best way to be completed). Many people first played Tetris on the Nintendo Gameboy handheld console but it has since become available for virtually every personal computer–based device. The game gives the player the task of creating complete lines from a series of regularly shaped blocks–tetrominos, which are a type of **polyomino**–that advance steadily

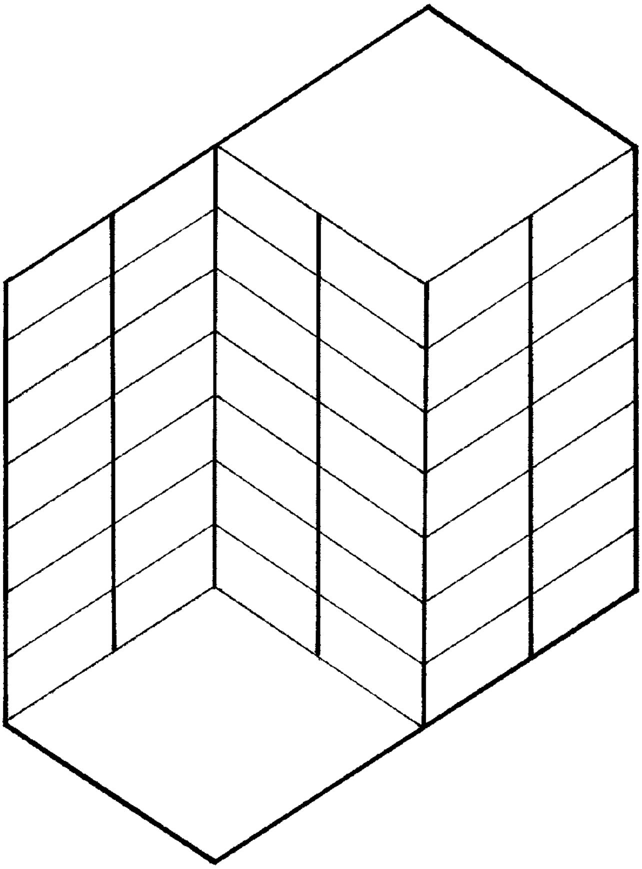

Thiery figure

down a narrow grid. The blocks can be spun to make them fit together better and complete lines. The game gets faster as levels are completed, making it harder to spin and fit blocks together fast enough to form lines. The MIT team found that, subject to certain conditions, Tetris has much in common with some of the knottiest mathematical conundrums, including the **traveling salesman problem**. Because Tetris is NP-hard, there is no easy way to maximize a score for the game, even when the sequence of blocks is known in advance.

tetromino
A four-square **polyomino**.

theorem
A major mathematical proposition that has been proved correct. More precisely, a statement in a **formal system** for which there exists a **proof**. See also **conjecture** and **lemma**.

theory of everything
A unified theory of all fundamental forces and interactions in nature; a grand unified theory that also includes gravity or general relativity (see **relativity theory**).

Thiery figure
A classic **ambiguous figure** devised by the psychologist A. Thiery in the late nineteenth century.

thirteen
The unluckiest of numbers if you happen to be superstitious. This belief has a couple of historical roots. According to biblical tradition, there were 13 people at Christ's Last Supper, and Christ was crucified on a Friday the thirteenth. Further back in time, Alexander the Great decided he wanted to be the thirteenth god alongside the 12 that already stood for each month of the year, so he had a thirteenth statue built on the place of his capital. His death

shortly after gave the number a bad name. Many buildings don't have a floor labeled 13 and many hotels will have a room numbered 12A instead of 13. There is even a name for a morbid fear of 13: triskaidekaphobia.

Fresh disasters involving the number hardly help triskaidekaphobics overcome their affliction. The most notorious of these involved the Apollo 13 Moon mission, which was launched on April 11, 1970 (the sum of 4, 11, and 70 equals 85, the digital sum of which is 13), from Pad 39 (3 × 13) at 13:13 local time, and suffered an explosion on April 13. (The astronauts did, however, make it home safely, which could be considered good luck.) There is always at least one Friday the thirteenth in each year; in some years there are two, and rarely three (e.g., 1998 and 2009). There were 13 original U.S. colonies (hence the 13 stripes on the American flag) and 13 signers of the Declaration of Independence. In Japan, the numbers 4 and 9 are considered unlucky, not because 13 can be represented as sum of these two perfect squares but because of their pronunciation. In Japanese, four is *shi,* which is pronounced the same as the word for death; nine is *ku,* which sounds like the word for torture. And speaking of torture, it was not unusual in times past for bakers to come in for stiff punishment if they short-changed their customers. In ancient Egypt, someone found selling light loaves might end up with his ear nailed to a doorpost, while in medieval Britain the punishment was likely to be a spell in the pillory. This led to the custom of adding a thirteenth loaf to every batch of 12 to be on the safe side, and hence the expression "a baker's dozen."

Mathematically, the reverse of the square of 13 is the same as the square of the reverse of 13: $13^2 = 169$; the reverse of 169 is 961 and the reverse of 13 is 31; $31^2 = 961$. Thirteen is the smallest **prime number** that can be expressed as the sum of the squares of two prime numbers: $13 = 2^2 + 3^2$. Also the sum of all prime numbers up to 13 $(2 + 3 + 5 + 7 + 11 + 13)$ is equal to the thirteenth prime number (41), and this is the largest such number. On the anagrammatical front is this nice equation:

$$\text{ELEVEN} + \text{TWO} = \text{TWELVE} + \text{ONE}$$

See also **dollar**.

thirty colored cubes puzzle

A game devised in 1921 by Percy **MacMahon**, which was marketed under the name "Mayblox." It is played with 30 cubes that have all possible permutations of six different colors on their faces. A number of different games can be played with the blocks. One is to choose a cube at random and then choose seven other cubes to make a $2 \times 2 \times 2$ cube with the same arrangement of colors for its faces as the first chosen cube. Each face of the $2 \times 2 \times 2$ cube has to be a single color and the interior faces have to match in color. After building one such cube, only one other with the same properties can by made from the remaining 22 cubes–the mirror image of the first.

thirty-six officers problem

Arrange 36 officers in a 6 × 6 square so that one officer from each of six regiments appears in each row and one from each of six ranks appears in each column. This problem, first posed by Leonhard **Euler** in 1779, is equivalent to finding two mutually orthogonal **Latin squares** of order six. Euler correctly conjectured that there was no solution; the search for a proof led to important developments in **combinatorics**.

Thompson, D'Arcy Wentworth (1860–1948)

A Scottish naturalist and polymath. Thompson held a professorial chair at St. Andrews and Dundee, Scotland, for the amazing period of 64 years, a record for tenure unlikely ever to be broken. Although he wrote more than 300 scientific articles and books, his reputation is based primarily on his efforts to reduce biological phenomena to mathematics in his magnum opus *On Growth and Form* (1917).[332] In this book, full of marvelous sketches of such things as nautilus shells and honeycombs, Thompson claimed that much about animals and plants could be understood by the laws of physics, as mirrored in the

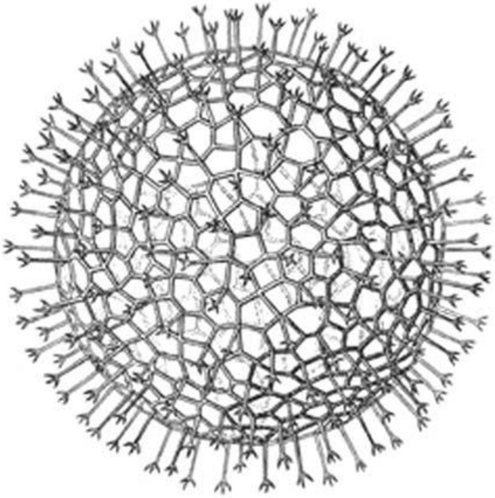

Thompson, D'Arcy Wentworth An illustration from Thompson's *On Growth and Form:* the shell of the radiolarian Aulastrum triceros.

structures and patterns of mathematics. His most novel idea was to show how mathematical functions could be applied to the shape of one organism to continuously transform it into other, physically similar organisms. One memorable example is the squeezing and stretching of a rectangular Cartesian grid that transforms the fish species *Scarus sp.* to the species *Pomacanthus*. Thompson used the same principle to transform skulls of baboons into those of other primates, and to show how corresponding bones like the shoulder blade are related in different species. Doubtless, if he were alive today, he would be heavily into "morphing"–the digital technique that allows a computer to do exactly the same kind of transformation of one object to another, but vastly more efficiently. Thompson acquired a local reputation as a mild eccentric; indeed, older inhabitants of St. Andrews still recall seeing him strolling about town with a parrot on his shoulder.

thousand

A number that, linguistically, comes from an extension of "hundred" and has roots in the Germanic *teue* and *hundt*. *Teue* refers to a thickening or swelling, and *hundt* is the source of our present-day hundred. A thousand, then, literally means a swollen or large hundred. The root *teue* is the basis of such common words today as thigh, thumb, tumor, and tuber. One thousand is the answer to the question: If you were to spell out numbers, how far would you have to go until you would find the letter "A"? While Americans may say "Thanks a million" to express gratitude, Norwegians offer "Thanks a thousand" ("*tusen takk*").

three

> *1 + 1 = 3, for large values of 1.*
>
> –Anonymous

The number of dimensions of space in which we live; three is also the smallest odd **prime number**, the second **triangular number**, and member of the **Fibonacci sequence**. Three is often the number of repetitions in jokes and children's stories (for example, the tale of the *Three Little Pigs* and of the number of chairs, beds, and porridge-eating bears in *Goldilocks*), because it is the minimum number needed to establish a pattern (such as a regular tempo) or to convey the impression of an ongoing sequence or succession. In the Christian tradition, three plays a crucial role: Christ represents one third of the Trinity (the Father, Son, and Holy Spirit), was visited by the three wise men and, 33 years later, when Peter disowned him three times, rose on the third day after the crucifixion, having died at 3 P.M. Other things that come in threes: musketeers, primary colors, wishes, blind mice, bad luck, and London buses. See also **Triangular Lodge** and **tinner's rabbits**.

three-hat problem

See **hat problem**.

three men's morris

An old game played on a 3 × 3 board that is thought to be a direct ancestor of **tic-tac-toe**; it is known by many other names, including nine holes, and is related to six men's morris and **nine men's morris**. The game involves two differently colored sets of four pieces (one set for each player). Players take turns placing pieces on intersection points, and the first person to place three along a line wins the game. The earliest known board for three men's morris was found on the roof of the temple in Kurna, Egypt, dating back almost three and a half thousand years. Its earliest known appearance in literature is in Ovid's *Ars Amatoria*. The Chinese are believed to have played it under the name *Luk tsut K'i* during the time of Confucius (c. 500 B.C.). Boards for three men's morris dating back to the thirteenth century can be found carved into the cloister seats at the cathedrals at Canterbury, Gloucester, Norwich, and Salisbury, and at Westminster Abbey.

three-body problem

The problem of determining the future positions and velocities of three gravitational bodies. The problem was proved unsolvable in the general case by Henri **Poincaré**, which foreshadowed the importance of **chaos**. Although no analytical solutions are possible in the worst case, a numerical solution is sometimes sufficient for many tasks.

Thue-Morse constant

Also known as the *parity constant*, the number defined as follows: Take a string of 1's and 0's and follow it by its complement (the same string with 1's switched to 0's and vice versa) to give a string twice as long. Repeat this process forever (starting with 0 as the initial string) to get the sequence 01101001100101101001011O . . . . Make this a binary fraction, $0.0110100110010110\ldots_2$, and rewrite it in base 10. The resulting **transcendental number**, 0.41245403364 . . . , is the Thue-Morse constant.

tic-tac-toe

Also known as noughts and crosses, and spelled in a variety of ways (such as ticktacktoe), this well-known pastime is not as ancient as popularly believed, though it certainly has roots in older games. The earliest clear description of the rules, but without a name, comes from Charles **Babbage** around 1820. Later on, Babbage started to call the game tit tat to and by slight variants on this, gave the first detailed analysis of it, and designed the first robot to play it! He speculated that "[T]he machine might consist of the figures of two children playing against each other, accompanied by a lamb and a cock. That the child who won the game might clap his hands whilst the cock was crowing, after which, that the child who was beaten might cry and wring his hands whilst the lamb was bleating."

His plan was to exhibit the machine in London to raise money for his more serious projects, including the fabulous Analytical Engine. However, on hearing that similar devices, such as a mechanical composer of Latin verse, had flopped financially he abandoned the scheme.

Two expert players of tic-tac-toe will always draw. In other words, there is no winning strategy against an opponent who knows the game well. Nor is it difficult to become unbeatable. Despite the fact that there are 9 × 8 × 7 × 6 × 5 (= 15,120) possible layouts of noughts and crosses for the first five moves alone, these reduce to just a few basic patterns of play and counterplay that ensure the canny contestant never loses. The game is only interesting when at least one novice is involved. The strongest opening play is in a corner because an unwary opponent will be trapped unless her countermove is in the center. By the same token, a center opening must be countered by a corner response or the first player has an easy win. Of course, a master player will not only never lose but will learn his opponent's weaknesses and exploit them in the most devastating way.

The game also becomes more interesting when played on a larger board and/or in more dimensions. An exception to this is the 3 × 3 × 3 cube which gives an easy win to the first player; in fact, it cannot end in a draw because the first player has 14 plays and it is impossible to make all of these without scoring. The 4 × 4 × 4 cube is much more interesting and is sold commercially as the well-known Score Four game. For the ambitious, tic-tac-toe can be played in four dimensions on a **tesseract** by sectioning it into two-dimensional squares. Older, similar, and mathematically more interesting relatives of tic-tac-toe include **nine men's morris** and **Ovid's game**.

tie knots

How many ways can you tie a tie? For many years there were just three styles of **knot**: the Four-in-Hand, the Windsor, and the Half-Windsor. Then the Pratt was introduced to the world on the cover of the *New York Times* in 1989. Intrigued that only one new knot had been added to the tie-tying repertoire in more than half a century, two researchers from the University of Cambridge's physics department, Thomas Fink and Yong Mao, decided to see how many tie knots were actually possible. To this end, they applied **random walk** theory—a technique useful for describing movement which, although unpredictable in detail, reveals large-scale patterns. Such patterns, the researchers realized, were essential to a successfully accomplished tie-knot. For example, if the end of the tie is moved to the right, its next move can't be to the right again—it has to be either to the left or to the center. This means that each move made in tying a tie can only be followed by one of two alternatives. Fink and Mao found that the simplest possible knot involves just three moves. They went on to discover 85 possible tie-knots, including the four popular knots, six new knots that they consider aesthetically pleasing, and two complicated nine-move knots.[100]

tiling

Also called a *tesselation,* a collection of smaller shapes that precisely covers a larger shape, without any gaps or overlaps. Usually, the shape to be tiled is a flat plane but other shapes and three-dimensional objects can be tiled, too. In a game that involves tiling, certain conditions are applied; for example, all the tiles may have to be identical, or they may all have to be squares but every one of a different size. It's been known for some time that all simple regular tilings in the plane belong to one of the 17 plane symmetry groups known as **wallpaper group**s. All 17 of these patterns are known to exist in the **Alhambra** palace. This doesn't exhaust the apparently simple problem of tiling the plane: adding extra constraints or removing the requirement for regularity leads to a large number of interesting problems. These include *alternating tilings,* for examples of squares or dominoes, such that two tiles have a side or a part of a side in common, or *colored tilings,* in which no two adjacent tiles have the same color. Colored tilings are also called *colored maps.* The most famous problem relating to colored tilings is the **four-color map problem**, which has been solved. Other problems involve *n*-tesselations, in which each tile has an integral area and for each natural number *n* there is exactly one tile with area *n.* See also **Penrose tiling**, **rep-tile**, **rectangle**, **squaring the square**, **dissection**, and **packing**.

time

> *Time is a great teacher. Unfortunately, it kills all its pupils.*
>
> —Hector Berlioz (1803–1869)

One of the most familiar and yet mysterious properties of the universe. The "flow" of time is one of the strongest impressions we have, yet it may simply be

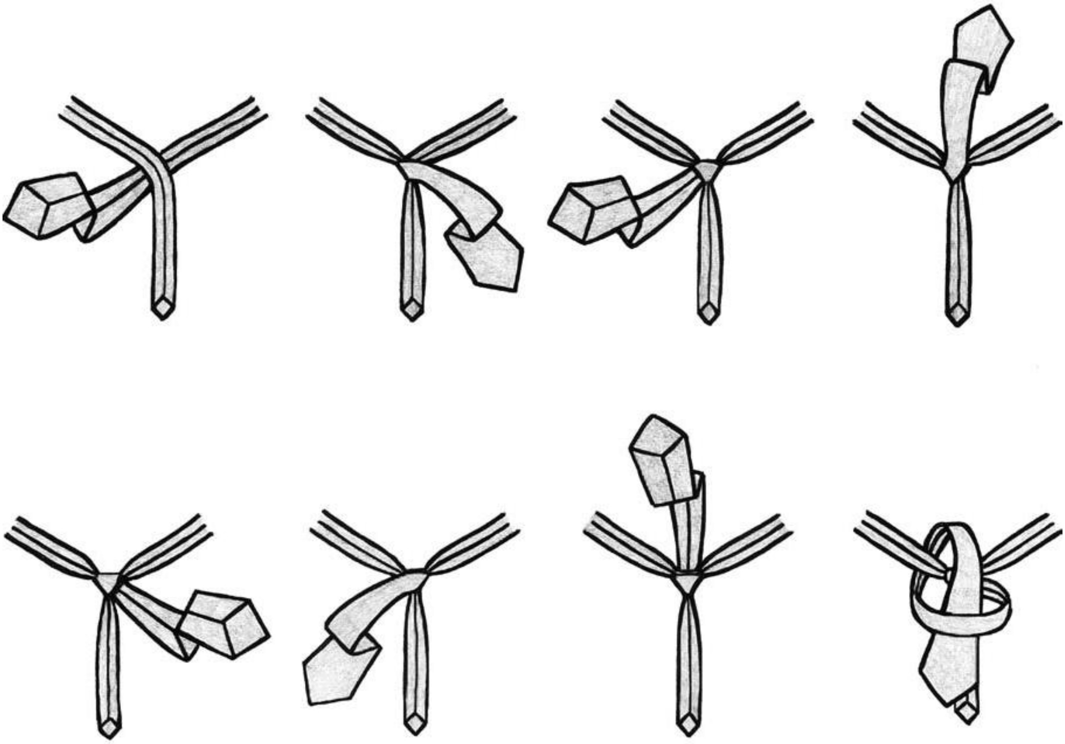

tie knots The “St. Andrew”: one of several new tie knots recently discovered. *Thomas Fink*

an illusion or a product of the conscious mind. The very notion that time somehow moves leads to a logical paradox because, as the Australian philosopher J. J. C. Smart asked: “In what units is the rate of time flow to be measured? Seconds per ____?” John Dunne in his classic book *An Experiment with Time,* argued that the human mind has the ability to rove back and forth along the time line so that precognition is a physical possibility. However his theory involves an infinite regress of time and of the observer that is philosophically hard to swallow. All the same, it may be that the apparent movement from past to present to future has less to do with the universe at large than it has to do with our individual subjective experience. In some way, still to be fathomed, time, consciousness, free will, and the individual are intimately entwined. In physics, by contrast, time is treated no differently (with one important exception, noted below) than space. It is simply another **dimension**–another axis, or extension, of physical reality. Just as the various spatial dimensions prevent everything from happening at a single point, so time prevents everything from happening all at once. As one wag put it, “Time is just one damned thing after another!” In Einstein’s **relativity theory**, time is effectively “spatialized” so that, instead of speaking of an absolute three-dimensional space and a separate one-dimensional time, there is a four-dimensional **space-time** continuum. So closely related are time and space in relativity theory that time can be converted into space and vice versa. In particular, different observers may not agree on the distance or the duration between any two events in space-time, but they will always agree on the *space-time interval.* If the two points events occur at (t, x, y, z) and $(t + dt, x + dx, y + dy, z + dz)$, then the (constant) space-time interval between them is given by

$$s^2 = c^2(t_2^2 - t_1^2) - (x_2^2 - x_1^2) - (y_2^2 - y_1^2) - (z_2^2 - z_1^2).$$

But the time of relativity, like that of classical physics, remains reversible.

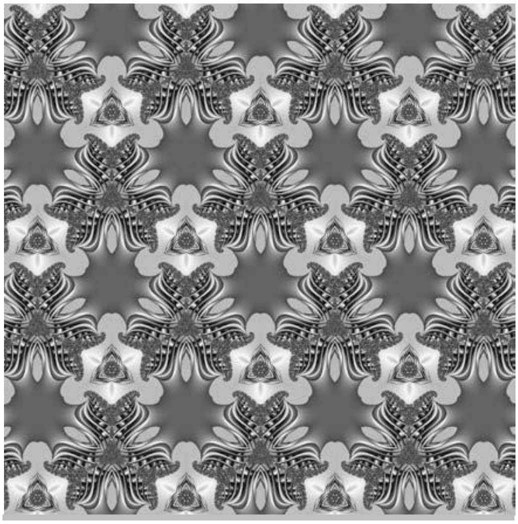

tiling A computer-generated periodic tiling. *Jos Leys, www.josleys.com*

Chrononauts
Official Rules
The Card Game of Time Travel
ID
MISSION
1912
Titanic Avoids Iceberg
Linchpin

time travel The game of Chrononauts, in which players must resolve a series of paradoxes that threaten the time line. *Looney Labs, www.LooneyLabs.com*

time complexity

A **function** that describes the amount of time required for a program to run on a computer to perform a particular task. The function is parameterized by the length of the program's input.

time dilation

See **relativity theory**.

time-reversible

A property of **dynamical system**s that can be run unambiguously both forward and backward in time. The **Lorenz system**, for example, is time-reversible.

time travel

Can we travel through **time**? Of course, we do it all the time! But can we do it at a different rate than normal? Again, the answer is "yes" because of the phenomenon known as *time dilation* in Einstein's **relativity theory**. However, time dilation enables, even in principle, only a limited kind of leap into the future–one from which we cannot return to the present. Genuine time travel is the ability to jump forward or backward through time at a rate other than that of the ordinary progression of events or that enabled by the relativistic time dilation effect.

The possibility of traveling through time poses such a threat to **causality** and opens the door to so many disturbing paradoxes that many scientists feel inclined to dismiss it out of hand. However, it has been a favorite theme of science fiction since the 1880s. In *The Time Machine* (1895), H. G. Wells gives a pleasant preamble about the nature of the **fourth dimension** before whisking his hero 802,000 years into the future. Says the Time Traveller (we never learn his real name), "[A]ny real body must have extension in four directions: it must have Length, Breadth, Thickness, and–Duration. There are really four dimensions, three of which we call the three planes of Space, and a fourth, Time. There is, however, a tendency to draw an unreal distinction between the former three dimensions and the latter, because it happens that our consciousness moves intermittently in one direction along the latter from the beginning to the end of our lives."

Unfortunately for would-be chrononauts (an early version of *The Time Machine* was called *The Chronic Astronauts*), Wells is not specific about how his time traveling device works, though we know that "Parts were of nickel, parts of ivory, parts had certainly been filed or sawn out of rock crystal." In more recent times, physicists, speculating on some of the more esoteric byways of relativity and **quantum mechanics**, have been a little more forthcoming about how time travel might be achieved in practice. These speculations have variously involved wormholes (shortcuts outside of normal space and time), faster-than-light particles known as **tachyon**s, and unusual cosmological models, such as the ***Gödel** universe,* which allow movement to any point in the future or the past. Let us leave aside the practical aspects, however, and focus on the logic of breaking the time barrier.

The various time travel possibilities dealt with in science fiction fall into two broad categories. In the first the timeline, from deepest past to darkest future, is frozen and immutable, like a filmstrip. Any time traveling that takes place is constrained by this preordained structure–effectively, already written into the narrative of the world (the "block universe" of Einsteinian physics)–and is thus prevented from leading to paradoxes. In one variant of this scenario, the *Novikov self-consistency principle* applies. Named after Igor Novikov, an astrophysicist at Copenhagen University, this asserts that any attempt at time travel that would lead to a paradox, such as the **grandfather paradox**, is bound to fail even if the cause of failure is an extremely improbable event. In other words, try as you might to introduce a contradiction into the timeline, like killing yourself or one of your ancestors in the past, circumstances will always conspire to prevent you. An excellent example of this type of universe is found in Robert L. Forward's novel *Timemaster.* Another variant on the fixed timeline concept is that any event that appears to have caused a paradox has, in fact, created a new timeline. The old timeline remains unaltered, and the time traveler becomes part of a new temporal branch line. One difficulty with this arrangement is that it might violate the principle of conservation of mass-energy, unless the mechanics of time travel demand that mass-energy be exchanged in precise balance between past and future at the moment of travel. However, the concept of branching universes and alternative histories is not outrageous in physics where the **many worlds hypothesis** and of Feynmann's sum-over-histories are routinely debated.

The second main type of time travel entertained in science fiction assumes that the timeline is flexible and changeable. This can lead to all sorts of mind-boggling difficulties and contradictions. A way to offset some of these problems is to stipulate that the timeline is very resistant to change. In the extreme case, as writer Larry Niven has argued, it may be a fundamental rule that in any universe where time travel is allowed, no actual time machine is ever invented. The English physicist and mathematician Stephen Hawking put this idea on a more formal footing with his *chronology protection conjecture.* On the other hand if the timeline is presumed to be easily changed, paradoxes threaten to spring up at every turn. One of the most remarkable of these is the *closed causal curve paradox* in which, it seems, something can be gotten for nothing. Samuel Mines summarized the plot of his 1946 short story as follows:

> A scientist builds a time machine, goes 500 years into the future. He finds a statue of himself commemorating the first time traveler. He brings it back to his own time and it is subsequently set up in his honor. You see the catch here? It had to be set up in his own time so that it would be there waiting for him when he went into the future to find it. He had to go into the future to bring it back so it could be set up in his own time. Somewhere a piece of the cycle is missing. When was the statue made?

Closed loops in time can also conjure knowledge out of thin air. A man builds a time machine and travels into the past to give the plans for the device to his younger self who then builds the machine, travels into the past, and so on. Where did the plans originate? A curious thing about time loops is that they have no easily discernible future or past because all the events taking place in them affect one another in a circular way. Time loops also put a question mark over free will. What happens if the younger man, given the time machine plans by his older self, decides not to build the device? *Can* he make that choice given that, in some sense, he has already built it? Perhaps the apparent absence of time travelers and time machines in the real world is a sign that we do not have to worry about such issues–at least, for the present.

tinner's rabbits

A name that has emerged recently to describe a pattern of three rabbits or hares that has been found in many parts of the world, including England and Wales, mainland Europe, China, and Russia. It occurs, for example, on the medieval roof bosses of some churches in Devon and Cornwall and is thought to be connected with the local tin-mining industry. One theory suggests the following link: Tin is alloyed with copper to make bronze, copper came from Cyprus (the words *Cyprus* and *copper* have the same root), Cyprus is the island of the goddess Venus or Aphrodite (she was born there), rabbits are symbols of Venus. Three intertwined fishes are a common Christian symbol, so the three rabbits may also have stood for the Trinity.

Tippee Top

Also known as a *Tippy Top,* a type of top, patented in Britain in 1953, that consists of a peg with a ball-shaped body. If the top is spun quickly on the rounded body, with the peg pointing upward, it flips itself over and spins on its peg. Crucial to this counterintuitive behavior is the shape of the body, which is smooth and spheroidal with no sharp point. After release, the top, like tops of every description, begins to show precession, that is, its axis of rotation moves in a small circle. After a while, the contact point of the top with the table no longer coincides with the rotation axis but instead moves to other points of the top's head. Due to frictional forces and precession, the top seeks a more stable position, which it finds by flipping on to its stalk. See also **celt**.

Titchener illusion

Also known as the *Ebbinghaus size illusion,* a well-known **distortion illusion**. Two circles are surrounded by either six big circles or six small circles. Despite appearances, the two center circles are exactly the same size.

Toeplitz matrix

A **matrix** in which all the elements are the same along any **diagonal** that slopes from northwest to southeast.

topological group

Also called a *continuous group,* a **set** that has both the structure of a **group** and of a **topological space** in such a way that the operations defining the group structure give continuous maps in the topological structure. Many groups of matrices (see **matrix**) give topological groups.

topological dimension

An integer that defines the number of coordinates needed to specify a given point of an object of set X. A single point, therefore, has a topological dimension equal to zero; a curve has dimension one, a surface has dimension two, and so on.

topological space

A type of generalized mathematical space in which the idea of closeness, or limits, is described in terms of relationships between **set**s rather than in terms of distance. Every topological space consists of: (1) a set of points; (2) a class of subsets defined axiomatically as open sets; and (3) the set operations of union and intersection.

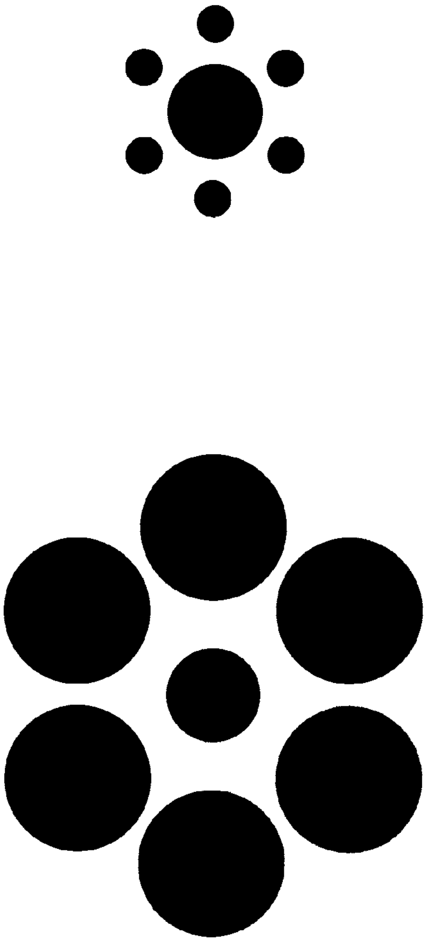

Titchener illusion The center circles are the same size.

topology

The study of those properties of mathematical objects that remain unaffected by smooth deformations, such as stretching and squeezing, but that don't involve tearing. The word comes from the Greek *topos* for "place," and was introduced into English by Solomon Lefschetz in the late 1920s. A topologist has been described as someone who doesn't know the difference between a doughnut and a coffee cup. Substitute "care about" for "know" and this becomes more accurate. Imagine a donut made of soft clay. A potter can easily shape this into a cup with a handle without removing or creating any new holes. Both shapes, in topology, are said to be genus 1–objects with a single hole. A sphere, by contrast, is genus 0 (no holes), while an eyeglass frame, with the lenses removed, is genus 2. For more on topologically intriguing structures, see **Möbius band** and **Klein bottle**.

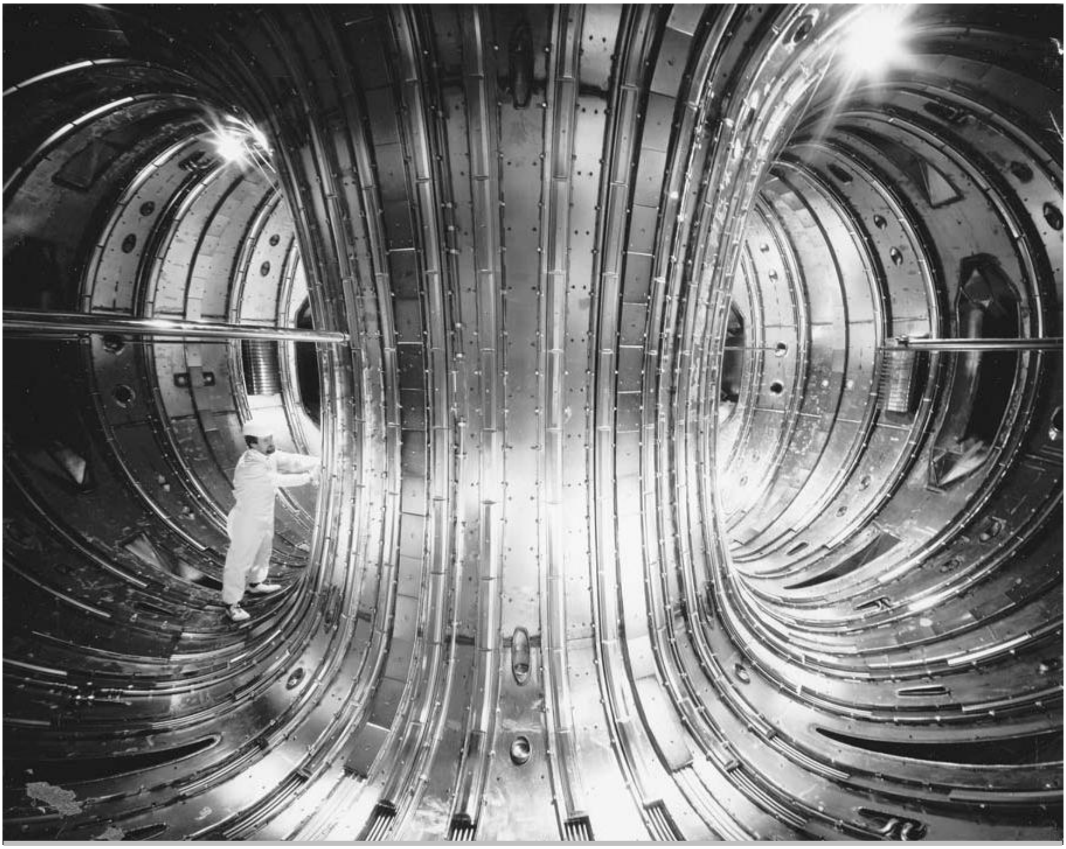

torus Inside the torus of an experimental nuclear fusion reactor. *Joint European Torus*

torus

A doughnut, bagel, or inner-tube shape; the word comes from the Latin for "bulge" and was first used to describe the molding around the base of a column. One way to think of a torus is as a **surface of revolution** obtained by rotating a circle around an axis that lies in the plane of the circle but doesn't intersect the circle. In the general case, where the shape being so rotated is any closed plane curve, the resulting surface is called a *toroid.* Although the usual torus in three-dimensional space is shaped like a doughnut, the concept of the torus is extremely useful in higher dimensional space as well.

tour

A sequence of moves by a **chess** piece on a chessboard in which each square of the board is visited exactly once. See also **knight's tour** and **magic tour**.

Tower of Brahma

A romantic legend manufactured by Edouard **Lucas** as an accompaniment to the popular game he invented, the **Tower of Hanoi**. According to the tale of the Tower of Brahma, in the Indian city of Benares, beneath a dome that marked the center of the world, is to be found a brass plate in which are set three diamond needles, "each a cubit high and as thick as the body of a bee." Brahma placed 64 disks of pure gold on one of these needles at the time of Creation. Each disk is a different size, and each is placed so that it rests on top of another disk of greater size, with the largest resting on the brass plate at the bottom and the smallest at the top. Within the temple are priests whose job it is to transfer all the gold disks from their original needle to one of the others, without ever moving more than one disk at a time. No priest can ever place any disk on top of a smaller one, or anywhere

AND BRAHMA SAID UNTO THE PRIESTS, TRANSFER THESE DISKS FROM THE FIRST PYRAMID, TO THE THIRD, TRANSPOSING ONE DISK AT A TIME ONLY
AND PUTTING IT EITHER ON A VACANT PYRAMID OR ON A LARGER DISK. BY THE TIME YOU HAVE EXECUTED THIS TASK THE END OF THE WORLD WILL BE NEAR.
PYRAMIDS
A PROBLEM FROM THE ANCIENTS
TRADE MARK
A MASTER THINKER CREATION
THIS MODERN VERSION MAY BE SOLVED IN A SHORT TIME..
PYRAMIDS
A GAME
A PUZZLE

Tower of Hanoi A vintage version of the game called Pyramids, manufactured by Knapp Electric, Inc., of New York. *Sue & Brian Young/Mr. Puzzle Australia, www.mrpuzzle.com.au*

else except on one of the needles. When the task is done, and all 64 disks have been successfully transferred to another needle, "tower, temple, and Brahmins alike will crumble into dust, and with a thunder-clap the world will vanish." The prediction (thunder-clap aside) seems fairly safe given that the number of steps required to transfer all the disks is $2^{64} - 1$, which is approximately 1.8447×10^{19}. Assuming one second per move, this would take about five times longer than the current age of the universe! Interestingly, $2^{64} - 1$ is also the answer to the **wheat and chessboard problem**.

Tower of Hanoi

A game invented by Edouard **Lucas** and sold as a toy in 1883. Early versions of it carried the name "Prof. Claus" of the College of "Li-Sou-Stain," but these were quickly discovered to be anagrams for "Prof. Lucas" of the College of "Saint Louis." The game, in its usual form, consists of three pegs on one of which are eight disks, stacked from largest to smallest. The problem is to transfer the tower to either of the vacant pegs in the fewest possible moves, by moving one disk at a time and never placing any disk on top of a smaller one. The minimum number of moves turns out to be $2^n - 1$, where n is the number of disks; this equals 255 in the case of eight disks. The original toy came with a description saying that it was a small version of the great **Tower of Brahma**.[257]

T-puzzle

A surprisingly difficult puzzle, given that there are only four pieces; it dates back to the start of the twentieth century. Photocopy and cut out the four pieces shown in the figure on the following page, and then try to arrange them to make the symmetric capital T. You are allowed to rotate the pieces as you wish and even turn them over, but they mustn't overlap in the final letter. In fact *two* different symmetric capital T letters can be made from the pieces. Also, two other symmetric shapes can be formed from the set, including an isosceles trapezoid. Can you find all of these?

trace

The sum of the terms along the **main diagonal** of a **matrix**.

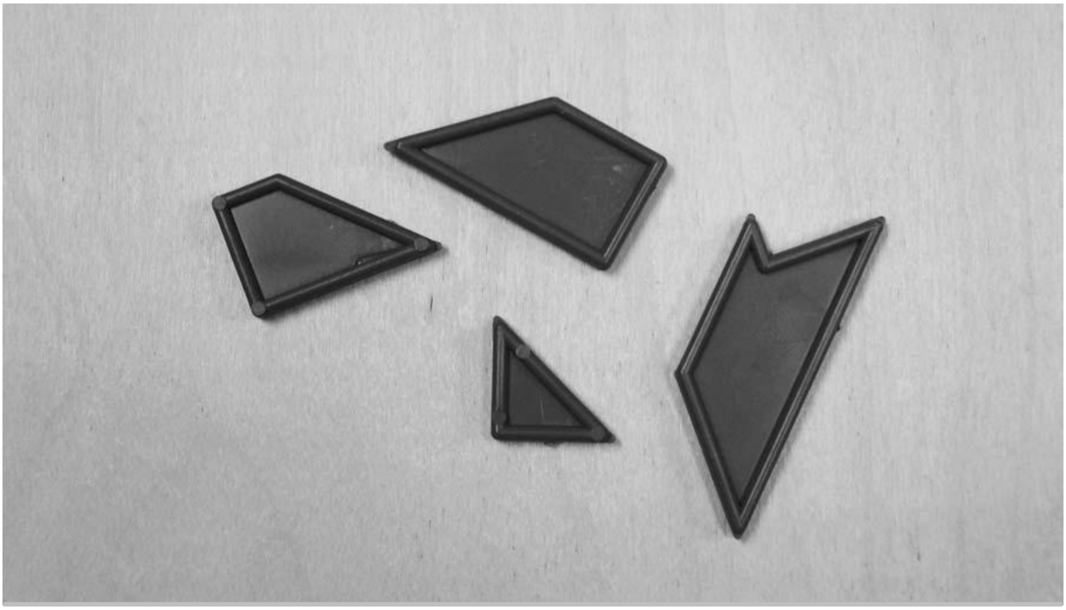

T-puzzle The pieces of the T-puzzle. *Kadon Enterprises, Inc., www.puzzlegames.com*

tractrix

A curve, sometimes called the *trajectory curve* or *equitangential curve,* that is the answer to a question asked by the Frenchman Claude Perrault (1613–1688). Perrault is not a giant in the annals of mathematics; in fact, he trained as a doctor and gained a minor reputation as an architect and an anatomist before dying in unusual style as a result of an infection he caught while dissecting a camel. His greatest claim to fame, aside from his connection with the tractrix, is that he was the brother of the author of "Cinderella" and "Puss-in-Boots."

In 1676, at about the time Gottfried **Leibniz** was doing groundbreaking work on the calculus, Perrault placed his pocket watch on the middle of a table, pulled the end of its chain along the edge of the table, and asked: What is the shape of the curve traced by the watch? The first known solution was given in a letter to a friend in 1693 by Christiaan **Huygens**, who also coined the name "tractrix" from the Latin *tractus* for something that is pulled along. (The corresponding German name is *hundkurve,* or "hound curve," which makes sense if you imagine the path a dog might follow on its leash as its master walks away.) The tractrix can also be found by taking the **involute** of a **catenary**. (Imagine a horizontal bar held at the vertex of the catenary and the point of contact marked as *P*. When the bar is rolled against the catenary without slipping, the path of *P* is a tractrix.) It is described by the parametric equations: $x = 1/\cosh(t)$, $y = t - \tanh(t)$. The **surface of revolution** of the tractrix is the **pseudosphere**, which is the classic model for **hyperbolic geometry** and one possible three-dimensional analog for the shape of the four-dimensional space-time in which we live.

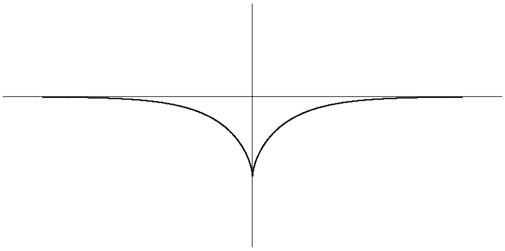

tractrix © *Jan Wassenaar, www.2dcurves.com*

trajectory

(1) The path of a **projectile** or other moving body through space. (2) A curve that intersects all curves of a given family at the same angles; if the intersection is at a right angle, this is an *orthogonal trajectory.* (3) The path through **phase space** taken by a system.

transcendental number

A number that can't be expressed as the root of a **polynomial** equation with integer coefficients. Transcendental numbers are one of the two types of **irrational number**, the other being **algebraic number**s. Their existence was proved in 1844 by the French mathematician Joseph Liouville (1809–1882). Although transcendentals make up the vast majority of real numbers, it is often surprisingly hard, and may even be impossible, to tell whether a certain number is transcendental or algebraic. For example it is known that both **pi** (π) and ***e*** are transcendental and also that at least one of $\pi + e$ and $\pi \times e$ must be transcendental, but it is not known which. It is also known that e^{π} is transcendental. This follows from the *Gelfond-Schneider theorem,* which says that if a and b are algebraic, a is not 0 or 1, and b is not rational, then a^b is transcendental. Using Euler's formula, $e^{i\pi} = -1$, and taking both sides to the power $-i$ gives $(-1)^{-i} = (e^{i\pi})^{-i} = e^{\pi}$. Since the theorem tells us that the left-hand side is transcendental, it follows that the right-hand side is too. (It also follows that $e \times \pi$ and $e + \pi$ are not both algebraic, because if they were, then the equation $x^2 + x(e + \pi) + e\pi = 0$ would have roots e and π, making both numbers algebraic.) But although it is known that e^{π} is transcendental, the status of e^e, π^e, and π^{π} remains uncertain.

transfinite number

Any of the infinite **ordinal number**s first described by Georg **Cantor**.

transformation

In geometry, a change to an object due to a process such as **rotation**, **reflection**, enlargement, or **translation**. In algebra, a transformation is the action of a **function**; in other words, what happens when there is a **one-to-one** mapping between sets of objects.

translation

Any **transformation** that takes the form of a constant offset with no rotation or distortion.

transpose

An operation that flips a **matrix** about the **main diagonal**.

transposition cipher

A **cipher** that encodes a message by reordering the plaintext. The receiver decodes the message using the inverse transposition. A simple kind of transposition cipher writes the message in a rectangle by rows, for example:

Asimplekin
doftranspo
sitionciph
erwritesth
emessagein
toarectang
lebyrowsan
dreadsitou
tbycolumns

and reads it by columns:

Adsee tldts oirmo erbif tweab eymti rsrya cproi serdo lanta cosle ncegt wiuks iseas tmipp tinao nnohh ngnus.

This type of cipher can be made more difficult to crack by permuting the rows and columns. See also **substitution cipher**.

transversal

A line that cuts across parallel lines, intersecting each of them.

trapezoid

A **quadrilateral** with one pair of parallel sides; in Britain this shape is known as a **trapezium**. If the parallel sides are of length a and b, and h is the perpendicular distance between them, then the area of the trapezoid is given by $A = \frac{1}{2}(a + b)h$.

trapezium

The American definition of a trapezium is a **quadrilateral** with no parallel sides. The British definition is equivalent to that of a **trapezoid**.

traveling salesman problem

Given a number n of cities, along with the cost of travel between each pair of them, find the cheapest way of visiting all the cities and returning to the starting point. This is equivalent to finding the **Hamilton circuit** of minimum weight in a weighted **complete graph**. Mathematical problems related to the traveling salesman problem (TSP) were treated in the nineteenth century by William **Hamil-**

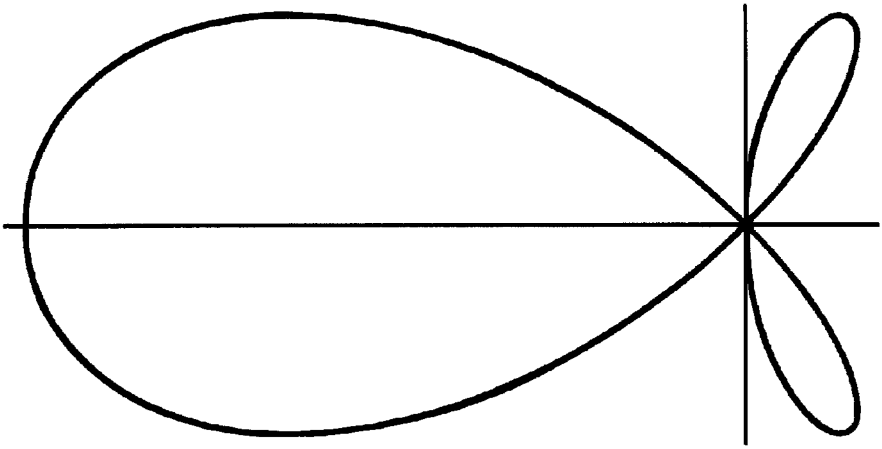

trefoil curve

ton, for example in his **Icosian game**, and by Thomas **Kirkman**. The general form of the TSP appears to be have been first studied by mathematicians in the 1930s, notably by Karl Menger in Vienna and Harvard, and later promoted by Hassler Whitney and Merrill Flood at Princeton. It has become a classic challenge to computer scientists seeking fast **algorithms** to complex problems. An approximate solution to the TSP for 15,112 cities, towns, and villages in Germany was found in 2001 by Princeton researchers using 110 computer processors and the equivalent of more than five years computer time on a 2 GHz machine.

tree

A **graph** with the property that there is a unique path from any **vertex** to any other vertex traveling along the **edges**.

trefoil curve

The plane curve given by the equation

$$x^4 + x^2y^2 + y^4 = x(x^2 - y^2).$$

triangle

A three-sided **polygon**. The sum of the interior angles of a triangle is always 180°, unless the triangle is drawn in a **non-Euclidean geometry**. Triangles can be classified either by their angles, as **acute**, **obtuse**, or **right**; or by their sides, as scalene (all different), **isosceles** (two the same), or **equilateral** (all equal).

Triangular Lodge

One of the few triangular buildings in England; it was built by Sir Thomas Tresham in about 1595 at Rushton, Northamptonshire. Tresham was a Catholic (spending some 15 years in prison because of this) and also a mystic numerologist. The whole design of the Lodge is based on the number **three**, which Tresham saw connected with his own surname and as an expression of his faith in the Christian Trinity. The Lodge's ground plan is a perfect **equilateral** triangle, each side 33 feet long–by tradition, the age of Christ at his death. The building has three floors, each floor has three windows, and each window is a three-fold **trefoil**. There are three gables and three gargoyles on each side. Even the central chimney is three-sided. The inscriptions all have 33 letters. Other buildings with a three-sided equilateral theme in Europe include a triangular castle at Gripsholm in Sweden and part of the Chateau de Chantilly in France, which is based on an equilateral plan of gigantic scale.

triangular number

Any number that can be represented by a triangular array of dots: 1, 3, 6, 10, ... The nth triangular number is $n(n + 1)/2$. Every integer is the sum of at most three triangular numbers. Every triangular number is a **perfect number**. If T is a triangular number, $8T + 1$ is a perfect square and $9T + 1$ is another triangular number. The square of the nth triangular number is equal to the sum of the first n cubes. Certain triangular numbers are also squares, but no triangular number can be a third, fourth, or fifth power, nor can one end in 2, 4, 7, or 9.

triangulation

A **tiling** of some object such as a **manifold** by simplices (see **simplex**).

tribar illusion

See **Penrose triangle**.

tricuspoid

See **deltoid**.

trident of Newton

A curve investigated and named "trident" by Isaac **Newton** as part of his systematic study of **cubic** equations. René **Descartes** also studied it and it is sometimes called the parabola of Descartes, although it isn't a parabola. It has the Cartesian equation

$$xy = cx^3 + dx^2 + ex + f.$$

trifolium

See **rose curve**.

trigonometric curve

Any of the curves produced when a **trigonometric function** is graphed.

trigonometric function

Any of the **functions** *sine* (sin), *cosine* (cos), *tangent* (tan), *secant* (sec), *cosecant* (cosec), or *cotangent* (cot), or their inverses, sin^{-1}, etc, which deal with certain proportions in right **triangles**. For example, the sine of an angle θ, $\sin\theta$, in a right triangle, is equal to the side opposite the angle divided by the hypotenuse (the longest side). Similarly cos is adjacent over hypotenuse and tan is opposite over adjacent. Sec, cosec, and cot are the multiplicative inverses of cos, sin, and tan, respectively: $\sec\theta = 1/\cos\theta$, and so forth. These are not the same as the inverse functions $\cos^{-1}$, $\sin^{-1}$, and $\tan^{-1}$, which are also known as arccos, arcsin, and arctan. Graphs of trigonometric functions produce **trigonometric curves**.

trigonometry

The branch of **mathematics** that deals with the relationships between the sides and the angles of **triangles** and the calculations based on them, particularly the **trigonometric functions**. Sherlock Holmes relies on a little trigonometry to solve a 250-year-old mystery known as the Musgrave Ritual (in a short story of the same name)–an enigmatic series of clues that refers to the shadow of an elm tree when the sun is just visible at the top of a nearby oak to point toward buried treasure. The great detective recalls to Watson his conversation with Reginald Musgrave:

> "Have you any old elms?" . . .
>
> "There used to be a very old one over yonder, but it was struck by lightening ten years ago, and we cut down the stump."
>
> "You can see where it used to be?"
>
> "Oh, yes." . . .
>
> "I suppose it is impossible to find out how high the elm was?"
>
> "I can give it to you at once. It was sixty-four feet. . . . When my old tutor used to give me an exercise in trigonometry, it always took the shape of measuring heights." . . .
>
> I went with Musgrave to his study and whittled myself this peg, to which I tied this long string with a knot at each yard. Then I took two lengths of a fishing-rod, which came to just six feet. . . . The sun was just grazing the top of the oak. I fastened the rod on end, marked out the direction of the shadow. . . . It was nine feet in length. Of course, the calculation was now a simple one. If a rod of six feet threw a shadow of nine, a tree of sixty-four feet would throw one of ninety-six. . . . I measured out the distance . . . and I thrust a peg into the spot.

trillion

In American and general usage, a million million–1,000,000,000,000 or 10^{12}. A European trillion is a million times larger than this, or 10^{18}. Counting one number every second, 24 hours a day, it would take 31,688 years to reach one (American) trillion. The first trillion-dollar lawsuit ($116 trillion) was filed in August 2002 by 600 family members against a company run by Osama bin Laden's family, Saudi Arabian princes, and Sudan. See also **large number**.

trimorphic number

See **automorphic number**.

trinomial

An algebraic expression consisting of three terms.

triomino

Also called a *tromino,* a three-square **polyomino**.

triple

A multiple of **three**. A *triple integral* is one in which the integrand is integrated three times. See also **Pythagorean triplet**.

trisecting an angle

Whereas **bisecting an angle** could hardly be simpler, splitting an angle in three equal parts with compass and straightedge alone is impossible, except in a few special cases such as when the angle happens to be 90°. Trisecting an arbitrary angle can be done if you cheat by using a measuring ruler instead of a plain straightedge, or even if you draw just two marks on the straightedge, but not if you play by the rules and the straightedge is completely blank. The Greeks put a huge effort into the problem but couldn't crack it. In fact, the question of whether trisection could ever be done in the general case remained open until 1837, when it was finally shown to be impossible by Pierre Wantzel, a 23-year-old French mathematician. Why is it impossible? Wantzel showed that the two problems of trisecting an angle and of solving a cubic equation are equivalent. Moreover, he showed that only a very few cubic equations can be solved using the straightedge-and-compass method. He thus deduced that most angles cannot be trisected.

trisector theorem

See **Morley's miracle**.

trisectrix

A general name for curves that can be used in **trisecting an angle**. The name "trisectrix," on its own, is often applied specifically to the **limaçon of Pascal**. Other famous trisectrix curves include the **Maclaurin trisectrix** and the **conchoid** of Nicomedes.

triskaidekaphobia

See **thirteen**.

trochoid

The curve formed by the path of a point on the extension of a radius of a circle as it rolls along a curve or line. It is also the curve formed by the path of a point on a perpendicular to a straight line as the straight line rolls along the convex side of a base curve. By the first definition, the trochoid is derived from the **cycloid**; by the second definition it is derived from the **involute**. See also **roulette**.

truel

A three-cornered gunfight or its logical equivalent. Imagine a truel between Arnie, Bullseye, and Clint, who are standing at the corners of an equilateral triangle. All know that Arnie's chance of hitting a target is 0.3 and Clint's is 0.5, while Bullseye never misses. They have to fire at their choice of target in the order Arnie then Bullseye then Clint until only one man is left. A man who's been hit is out of the fight and can no longer be shot at. What should Arnie's strategy be?

Truels, like this one, have become a significant topic in **game theory** because they're analogs of various real-life situations, from rivalry among animals to competition between television networks. Small changes in the rules can lead to strikingly different, sometimes counterintuitive outcomes. Different firing rules are possible: sequential in fixed order (players fire one at a time in a predetermined, repeating sequence), sequential in random order (the first player to fire and each subsequent player is chosen at random from among the survivors), or simultaneous (all surviving players fire at the same time in every round). In certain truels, a participant is allowed to shoot at the ground rather than try to eliminate an opponent (an optimal strategy if the firing order is fixed and each player has only one bullet and is a perfect shot). If the first shooter misses on purpose, he eliminates himself as a threat, and the other two fight it out, leaving two survivors in the end. Any other course of action would lead to the first shooter's own demise, with only one survivor. Even if the players have an unlimited supply of bullets, the truel may still end with more than one survivor because no player wants to be the first to shoot. Indeed, under the fixed firing order rule, no player has an incentive to eliminate another player. Only in the case of simultaneous firing is there a chance that nobody will survive.

Most of the mathematical research on truels concerns the relationship between a player's marksmanship (probability of hitting a target) and his or her survival probability. It's possible to show, for example, that better marksmanship can hurt in many situations. In a sequential truel in which contendors aren't allowed to shoot in the air, a player maximizes his probability of survival by firing at the opponent against whom he'd less prefer to fight in a duel—regardless of what the other players do. If his shot misses, it makes no difference who the target was. If the shot hits the target, the shooter is better off because his opponent in the next duel is weaker. Thus, the first shooter fires at the opponent who's the better marksman. In general, depending on the marksmanship values, the survival probabilities of the truelists could end up in any order, including one that is the reverse order of shooting skill. Optimal play can be very sensitive to slight changes in the rules, such as the number of rounds of play allowed. On the other hand, some factors are fairly constant: the disadvantage of being the best marksman, the weakness of pacts, the possibility that an endless supply of ammunition may stabilize rather than undermine cooperation, and the deterrent effect of an indefinite number of rounds of play (which can prevent players from trying to get the last shot). Some of these findings are counterintuitive, even paradoxical. An understanding of them might well dampen the desire of aggressive players to score quick but temporary wins, rendering them more cautious. In particular, contemplating the consequences of a long, drawn-out conflict, truelists may come to realize that their own actions, while immediately beneficial, may trigger forces that ultimately lead to their own destruction.

truncate

To slice off a corner of a **polyhedron** around a **vertex**.

truncatable prime

A **prime number** *n* that remains a prime when digits are deleted from it one at a time. For example 410,256,793 is a truncatable prime because each number created by the removal of the digit underlined produces a new prime: 410,256,793; 41,256,793; 4,125,673; 415,673; 45,673; 4,567; 467; 67; 7. It is conjectured that there are infinitely many of these primes. If the digits from a prime can be deleted only from the right, to leave a prime, then *n* is called a *right truncatable prime.* If they can be deleted only from the left, to leave a prime, then *n* is a *left truncatable prime.* The list of primes from which *any* digit can be deleted at each step to leave a prime is very short indeed, because it demands that each digit be a prime and also that no digit occurs twice. Only these numbers satisfy this requirement: 2, 3, 5, 7, 23, 37, 53, and 73.

Tschirnhaus's cubic

A curve with the Cartesian equation $3ay^2 = x(x - a)^2$.

Turing, Alan Mathison

Turing, Alan Mathison (1912–1954)

An English mathematician considered to be one of the fathers of modern digital computing. At an early age, Turing showed signs of the genius and eccentricity that became hallmarks of his adult personality. He taught himself to read in three weeks and made a habit of stopping at street corners to read the serial numbers of traffic lights. Later, he became a near-Olympic-class runner and ran long distances with an alarm clock tied to his waist to time himself.

At Cambridge, Turing studied under G. H. **Hardy** and got involved with problems that David **Hilbert** and Kurt **Gödel** had proposed to do with completeness and decidability in mathematics. In 1936, he introduced the idea of what became known as **Turing machine**s–formal devices capable of solving any conceivable mathematical problem that could be represented by an **algorithm**. However, the Turing machine was only a theoretical possibility at that time and not a working implementation. It would remain for later researchers to solve the various practical difficulties required to make the computer a reality. Turing also showed that there were mathematical problems that a Turing machine could never solve. One of these is the **halting problem**. While his proof was published after that of Alonzo **Church**, Turing's work is more accessible and intuitive. During World War II, Turing was a major player at Bletchley Park, near present-day Milton-Keynes (a town built after the war), in the successful efforts to crack the Nazi Enigma ciphers. While serving at Bletchley Park (1939–1944), he stayed at the Crown Inn, Shenley Brook End, and somewhere near here he buried two silver bars, carefully recording the site with respect to local landmarks. When he returned to recover them, the area had been rebuilt and all his landmarks were gone. Despite several attempts with metal detectors, he never recovered them and no one else is known to have found them. The Crown is now a private house and the area where he buried the bars is a housing estate.

Turing's interest in computing continued after the war, when he worked at the National Physical Laboratory on the development of a stored-program computer (the ACE or Automatic Computing Engine). In 1948 he moved to the University of Manchester, where the first stored-program digital computer ran later that year (see

Babbage, Charles for a photo of this machine). In 1950, in the article "Computing Machinery and Intelligence," Turing tackled the problem of **artificial intelligence**, and proposed an experiment now known as the **Turing test**.

In 1952 his lover helped a compatriot to break into Turing's house and commit larceny. Turing went to the police to report the crime. As a result of the police investigation, he was charged with homosexuality (formerly a crime), offered no defense, and was convicted. Following the well-publicized trial, he was given a choice between incarceration and libido-reducing hormone injections. He chose the latter, which lasted for a year and had side effects including the development of breasts during that period. In 1954 he died of poisoning after eating a cyanide-laced apple. Most (though not his mother) believed that his death was intentional, and the death was ruled a suicide. According to one urban legend the Apple company's logo is symbolic of this event: an apple with two bites (or possibly bytes) out of it and rainbow colors that code for homosexuality. See also **Church-Turing thesis**.[168]

Turing machine

An abstract model of computer execution and storage introduced in 1936 by Alan **Turing** to give a mathematically precise definition of **algorithm**. A Turing machine can be thought of as a black box that carries out a calculation of some kind on an input number. If the calculation reaches a conclusion, or halts, then an output number is returned. Otherwise, the machine theoretically carries on forever (see **halting problem**). There are an infinite number of Turing machines, as there are an infinite number of calculations that can be done with a finite list of rules. A Turing machine that can simulate any other Turing machine is called a *universal Turing machine* or a **universal computer**. The concept of Turing machines is still widely used in theoretical computer science, especially in **complexity theory** and the theory of computation.

Turing test

A proposed way of deciding if a machine has human-level intelligence. First described by Alan **Turing** in 1950, it goes like this: A human judge engages in a natural language conversation with other parties; if the judge can't reliably tell whether the other party is human or machine, then the machine is said to pass the test. It is assumed that both the human and the machine try to appear human. The origin of the test is a party game in which guests try to guess the gender of a person in another room by writing a series of questions on notes and reading the answers sent back. In Turing's original proposal, the human participants had to pretend to be the other gender, and the test was limited to a 5-minute conversation. These features are nowadays not considered to be essential and are generally not included in the specification of the Turing test. Turing proposed the test in order to replace the emotionally charged and, for him, meaningless question "Can machines think?" with a more well-defined one. Turing predicted that machines would eventually be able to pass the test. In fact, he estimated that by the year 2000, machines with 10^9 bits (about 119MB) of memory would be able to fool 30% of human judges during a 5-minute test. He also predicted that people would then no longer consider the phrase "thinking machine" contradictory.

It has been argued that the Turing test can't serve as a valid definition of **artificial intelligence** for at least two reasons: (1) A machine passing the Turing test might be able to simulate human conversational behavior, but this could be much weaker than true intelligence. The machine might just follow some cleverly devised rules. (2) A machine might well be intelligent without being able to chat like a human. Simple conversational programs, such as ELIZA, have fooled people into believing they are talking to another human being; however, such limited successes don't amount to passing the Turing test. Most obviously, the human party in the conversation has no reason to suspect he is talking to anything other than a human, whereas in a real Turing test the questioner is actively trying to determine the nature of the entity he is chatting with. The Loebner Prize is an annual competition to determine the best Turing test competitors. See also **Chinese room**.

twelve

A number heavily used for grouping things (inches, hours, 12-packs), partly because it can be divided evenly in several different ways (by 2, 3, 4, and 6) and partly because there are roughly 12 cycles of the Moon for every one of the Sun. The Latin *duodecim* (two + ten) for 12 forms the root of dodecagon (originally duodecagon), meaning a 12-sided shape, and duodenum, the first part of the intestine that is about 12 inches long. Contracted and modified over the years, *duodecim* became "dozen." Multiples of 12 have also been used by many cultures for various units and measures. A "shock" was 60 or five dozen (a dozen for each finger on one hand), and many cultures had a "great hundred" of 120 or 10 dozen (a dozen for each finger on both hands). The Romans used a fraction system based on 12 and the smallest part, an *uncil,* became our word for "ounce." The French emperor Charlemagne established a monetary system that had a base of 12 and 20, the remnants of which persist. Until 1970, the English pound sterling consisted of 20 shillings, and each shilling contained 12 pence. In 1944, The Duodecimal Society was formed in New York with the purpose of proposing a switch to base 12 for all scientific work. There are 12 signs of the zodiac and there were 12 apostles of Christ. Twelve is the smallest **abundant number**, a **Harshad number**, and a semiperfect number (because $12 = 1 + 2 + 3 + 6$; see **perfect number**). See also **gross**.

twelve-color map problem

If on a plane or sphere each country has at the most one colony, requiring the same color as its parent country, at most 12 different colors are needed to distinguish the political regions on a map. The problem of determining if this is true or not remains open. See also **four-color map problem**.

twenty

Many early cultures, including some in Europe; the Mayans of Central America; and the Ainu, the indigenous people of the Japanese islands, used a **base** of 20 for counting. The base 20 system was retained until about 1970 in the British monetary system, in which there were 20 shillings to the pound. A group of 20 is often called a **score**. Two of the five Platonic solids involve 20: the **icosahedron** has 20 triangular faces and the **dodecahedron** has 20 vertices. Twenty is a **Harshad number**, a semiperfect number (see **perfect number**), and a **practical number**. It is also a tetratrahedral number–the sum of consecutive triangular numbers $(1 + 3 + 6 + 10)$.

twin primes

Pairs of **prime numbers** that differ by two, the first of which are 3 and 5, 5 and 7, 11 and 13, and 17 and 19. The largest example known, as of February 2003, is a pair of 51,090-digit primes discovered by Yves Gallot and Daniel Papp, with the value $33{,}218{,}925 \times 2^{169690} \pm 1$. Other than the first, all twin primes have the form $\{6n - 1, 6n + 1\}$; also, the integers n and $n + 2$ form twin primes if and only if $4[(n - 1)! + 1] = -n \pmod{n(n + 2)}$. In 1919 the Norwegian mathematician Viggo Brun (1885–1978) showed that the sum of the reciprocals of the twin primes converges to a sum now known as *Brun's constant:*

$$(1/3 + 1/5) + (1/5 + 1/7) + (1/11 + 1/13) + \\ (1/17 + 1/19) + \ldots$$

In 1994, by calculating the twin primes up to 10^{14} (and discovering the infamous Pentium bug in the process), Thomas Nicely of Lynchburg College estimated Brun's constant to be 1.902160578. According to the (unsolved) *twin-prime conjecture* there are infinitely many twin primes. The twin-prime conjecture generalizes to prime pairs that differ by any even number n, and generalizes even further to certain finite patterns of numbers separated by specified even differences. For example, the following triplets of primes all fit the pattern k, $k + 2$, and $k + 6$: 5, 7, and 11; 11, 13, and 17; 17, 19, and 23; 41, 43, and 47. It is believed that for any such pattern not outlawed by divisibility considerations there are infinitely many examples. (The pattern k, $k + 2$, and $k + 4$ has only one solution in primes, 3, 5, and 7, because any larger such triplet would contain a number divisible by 3.) Quartets of the form k, $k + 2$, $k + 6$, and $k + 8$ (the smallest example is 5, 7, 11, and 13) are thought to be infinite. For some patterns no example is known, or only one.

twins paradox

See **relativity theory**.

twisted cubic

A curve in three-dimensional space or projective space whose points are given by $(x(t), y(t), z(t))$ for a parameter t and where x, y, z are **polynomials** of at most degree 3.

two

The first even number and the only even **prime number**. The word comes from the Greek *dyo* and the Latin *duo* through the Old English *twa*. Early languages often had both feminine and masculine forms for two and so there are a lot of diverse roots related to "two-ness." Many "two" words use the Greek root *bi* such as biannual, binary, biscuit, and biceps. Others come from the Old English *twa*, such as between, twilight, twist, and twin. From *duo* we get dual, duet, dubious (of two minds), duplex (two layers), and double. The Latin *di* gives us diploma (two papers) and dihedral. The earlier Greek *dyo* produces dyad, composed of two parts. Two is the only positive real number that gives the same result when added to itself as when multiplied by itself. It is conjectured that 2 is the only even integer that cannot be written as the sum of two primes (see **Goldbach conjecture**) and it has recently been proven that 2 is the largest value of n for which the equation $x^n + y^n = z^n$ has nonzero integer solutions (see **Fermat's last theorem**). Two is the base of the **binary** number system.

two-dimensional world

Life in three dimensions is familiar and there is a huge body of literature on the **fourth dimension**. But what would a universe of just two dimensions be like? The first and the most charming book on the subject is Edwin Abbott's ***Flatland: A Romance of Many Dimensions***

(1884).[1] This was followed by Charles **Hinton**'s lengthier *An Episode of Flatland* (1907)[272] in which the 2-d world is not a plane, as in Abbott's yarn, but the rim of a large circular world called Astria. Hinton was the first to explore in some depth what science and technology might be like in two dimensions; in fact, an earlier pamphlet of his called "A Plane World," (reprinted in *Scientific Romances* in 1884) may have helped inspire Abbott's novel. Hinton's speculations were taken much further by Alexander **Dewdney** in *Planiverse* (1984).[81]

U

Ulam, Stanislaw Marcin (1909–1984)

A Polish-born American mathematician and physicist who solved the problem of how to initiate fusion in the hydrogen bomb and also devised the **Monte Carlo method** of solving mathematical problems using statistical sampling. He first came to the United States in 1935 following an invitation from John **von Neumann**. One morning in 1946 an event happened that changed Ulam's life, as colleague Gian-Carlo Rota recalled:

> Ulam, a newly appointed professor at the University of Southern California, awoke to find himself unable to speak. A few hours later, he underwent a dangerous surgical operation after the diagnosis of encephalitis. . . . In time, however, some changes in his personality became obvious to those who knew him . . . [H]is ideas, which he spouted out at odd intervals, were fascinating beyond anything I have witnessed before or since. However, he seemed to studiously avoid going into details . . . . [H]e came to lean on his unimpaired imagination for his ideas, and on . . . others for technical support . . . . A crippling technical weakness coupled with an extraordinarily creative imagination is the drama of Stan Ulam.

Ulam spiral

A remarkable geometric pattern accidentally found among the **prime numbers** by Stanislaw **Ulam**;[314] it is also known as the *Prime Spiral.* During a boring meeting one day in 1963, Ulam drew a square, marked the number 1 at the center, and then wrote the increasing whole numbers as a spiral that wound its way out to the edge of the paper. He then circled all the prime numbers and was immediately struck by how they tended to fall on diagonal lines radiating from the central 1. In Ulam's words the arrangement of primes "appears to exhibit a strongly nonrandom appearance." Ulam rushed home and

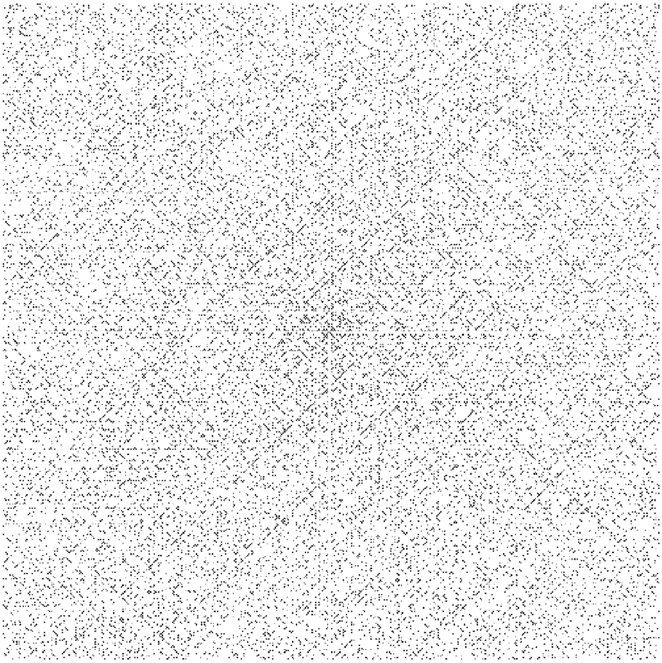

Ulam spiral Patterns amid the primes.

expanded the spiral to cover a much larger portion of the number sequence. The strange pattern persisted. Primes had a tendency to occur in clusters and all clusters tended to make a beautiful image that couldn't be predicted. With the help of computers this pattern can now be explored almost indefinitely and it reveals a wonderfully rich combination of symmetry and surprise–very reminiscent of some **fractal**s.

The Ulam spiral should perhaps be known as the "Clarke spiral" in view of the fact that Arthur C. Clarke described the phenomenon in his novel *The City and the Stars* (1956, ch. 6, p. 54),[63] predating Ulam's discovery by several years. Clarke wrote:

> Jeserac sat motionless within a whirlpool of numbers. The first thousand primes. . . . Jeserac was no mathematician, though sometimes he liked to believe he was. All he could do was to search among the infinite array of primes for special relationships and rules which more talented men might incorporate in general laws. He could find how numbers behaved, but he could not explain why. It was his pleasure to hack his way through the arithmetical jungle, and sometimes he discovered wonders that more skillful explorers had missed. He set up the matrix of all possible integers, and started his computer stringing the primes across its surface as beads might be arranged at the intersections of a mesh.

Ulam's conjecture

See **Collatz problem**.

uncountable set

A **set** of numbers that can't be put in a definite order from smallest to largest and so can't be counted. All uncountable sets are infinite, but not all infinite sets are uncountable. The best known uncountable set is the set of all **real number**s. By contrast the set of all **natural number**s, which represents the "smallest" type of **infinity**, is countable.

undulating number

An integer whose digits, in a given **base**, alternate–that is, one written in the form *ababab* . . . , where *a* and *b* are digits. For example, 434,343 and 101,010,101 are undulating numbers.

unduloid

A member of a family curves that is formed by films or liquid drops suspended between certain boundaries. Examples of unduloids are seen on a spider web when viewed through a microscope. They consist of blobs of viscous liquid that make up the sticky part of the web and are mostly gathered into a lemon shape. The family of unduloids includes shapes ranging from very thin to almost spherical, depending on the diameter of the thread and the volume of liquid in the blob. The shape of the curve is a result of the equality of pressure throughout the blob, which means that the total **curvature** at all points on the surface must be the same. The total curvature is the sum as the curvatures in two planes at right angles, and so varies from one blob to the next. A common property of all unduloids, however, is that they have a constant nonzero main curvature.

unexpected hanging

A remarkable logical **paradox** that appears to have begun circulating by word of mouth in the 1940s, often in the form of a puzzle about a man condemned to be hanged. A judge, with a reputation for reliability, tells a prisoner on Saturday that he will be hung on one of the next seven days but that he will not know which day until he is informed on the morning of the execution. Back in his cell, the prisoner reasons that the judge must be wrong. The hanging cannot be left until Saturday, because the prisoner would certainly know, if this day dawned, that it was his last. But if Saturday is eliminated, the hanging cannot take place on Friday either, because if the prisoner survived Thursday he would know that the hanging was scheduled for the next day. By the same argument, Thursday can be crossed off, then Wednesday, and so forth, all the way back to Sunday. But with every other day ruled out for a possible unexpected hanging, the hangman cannot arrive on Sunday without the prisoner knowing in advance. Thus, the condemned man reasons, the sentence can't be carried out as the judged decreed. But then Wednesday morning comes around and, with it, the hangman–unexpectedly! The judge was right after all and something was awry with the prisoner's seemingly impeccable logic. But what? More than half a century of attack by numerous logicians and mathematicians has failed to produce a resolution that is universally accepted. The paradox seems to stem from the fact that whereas the judge knows beyond doubt that his words are true (the hanging will occur on a day unknown in advance to the prisoner), the prisoner does not have this same degree of certainty. Even if the prisoner is alive on Saturday morning, can he be certain that the hangman will arrive?[61, 260]

uniform polyhedron

A **polyhedron** in which each face is regular and each **vertex** is equivalently arranged. Uniform polyhedra include the **Platonic solid**s, the **Archimedean solid**s, the **prism**s and **antiprism**s, and the nonconvex uniform polyhedra (see **nonconvex uniform polyhedron**).

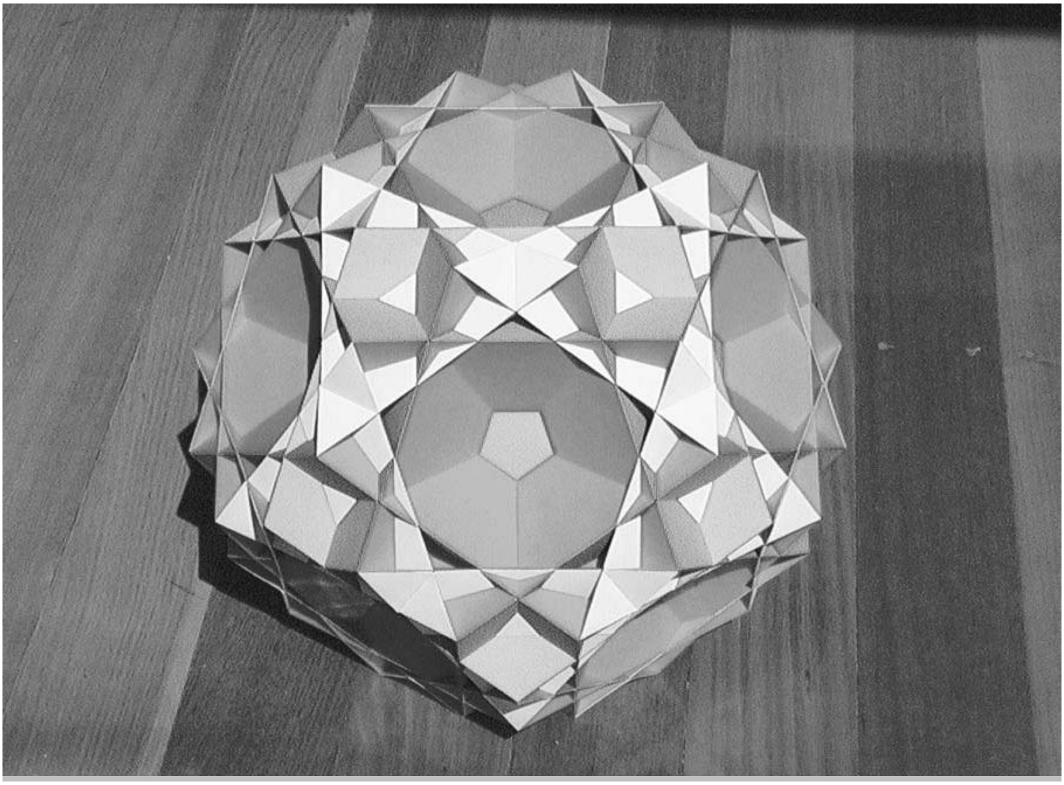

uniform polyhedron The great icosicosidodecahedron, a uniform polyhedron. *Robert Webb, www.software3d.com; created using Webb's Stella program*

unilluminable room problem

Imagine an L-shaped room in which Amy is standing near one corner holding a match. If Bob stands round the corner, he can see light from the match because a light ray can bounce off the two opposite walls. This is true wherever Bob stands in the room: the whole room is illuminated by the one match. Would this be true for a room of *any* shape, or is there at least one room that is *so* complicated that there's somewhere inside it that light from the match never reaches. This problem was first asked by Ernst Strauss in the 1950s. Nobody knew the answer until 1995, when George Tokarsky of the University of Alberta showed that the answer is "yes," there *is* a room that is not completely illuminable. His published floor plan showed a room with 26 sides–the smallest such room currently known. But a mystery remains. The room Tokarsky found contains one particular place where the match can be held which leaves part of the room dark. But if the match is moved slightly, the whole room is lit up again. Is there a room so fiendishly complicated that *wherever* the match is held there are some places that its light can never reach? For the moment we remain in the dark.[320, 334]

unique number

The constant U_n that results if a number A_n consisting of n consecutive digits, in ascending order, is subtracted from the number A_n' obtained by reversing the digits of A_n. For example, a three-digit number 345, if subtracted from its reverse 543, yields a difference of 198. Any other three-digit number subtracted from its reverse gives the same difference. Thus $U_3 = 198$. Similarly for a number with four consecutive digits, the unique number $U_4 = 3{,}087$. The first ten unique numbers are: $U_1 = 0$, $U_2 = 9$, $U_3 = 198$, $U_4 = 3{,}087$, $U_5 = 41{,}976$, $U_6 = 530{,}865$, $U_7 = 6{,}419{,}754$, $U_8 = 75{,}308{,}643$, $U_9 = 864{,}197{,}532$, and $U_{10} = 9{,}753{,}086{,}421$. Unique numbers are related to **Kaprekar numbers**, K_n, by the formula

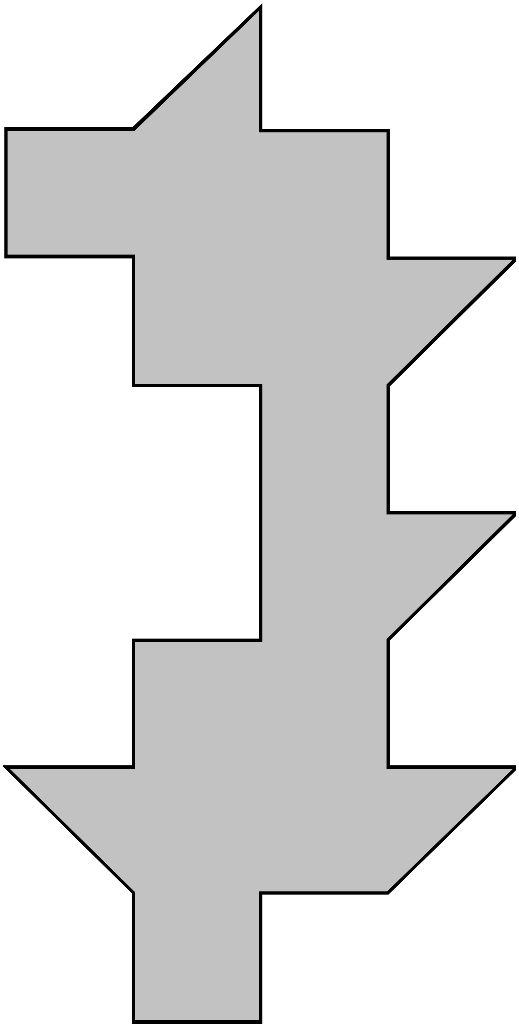

unilluminable room problem No single light source can light up every corner of a room with this shape.

$$U_n + U'_n = K_n + K'_n.$$

For example, when $n = 4$, $K_4 = 6{,}174$, $K_4' = 4{,}716$, $U_4 = 3{,}087$, $U_4' = 7{,}803$, and

$$3{,}087 + 7{,}803 = 10{,}890 = 6{,}174 + 4{,}716.$$

unimodal sequence

A **sequence** that first increases and then decreases.

unimodular matrix

A square **matrix** whose **determinant** is 1.

union

The **set** of all elements that belong to at least one of two or more given sets. It is denoted by the symbol ∪.

unit circle

A circle with radius 1.

unit cube

A cube with edge length 1. A *unit square* has a side length of 1.

unit fraction

A fraction whose **numerator** (number on top) is 1.

universal approximation

Having the ability to approximate any **function** to an arbitrary degree of accuracy. **Neural network**s are universal approximators.

universal computation

Capable of computing anything that can in principle be computed; being equivalent in computing power to a **Turing machine** or the **lambda calculus**.

universal computer

A computer that is capable of **universal computation**, which means that given a description of any other computer or program and some data, it can perfectly emulate this second computer or program. Strictly speaking, home PCs are not universal computers because they have only a finite amount of memory. However, in practice, this is usually ignored.

Universal Library

A library that contains not just one copy of every book that has ever been printed but one copy of every book that it is possible to print. A version of such a fantastic place is described by Jorge Luis **Borges** in his melancholic short story "Library of Babel" from *The Garden of Forking Paths* (1941). It begins: "The universe (which others call the Library) is composed of an indefinite, perhaps infinite number of hexagonal galleries." Each gallery is identical to all the others and contains 800 books identical in format. "[E]ach book contains four hundred ten pages; each page, forty lines, each line, approximately eighty black letters. . . ." There are 25 symbols–22 letters, the comma, the period, and the space. Because the library contains every possible combination of these

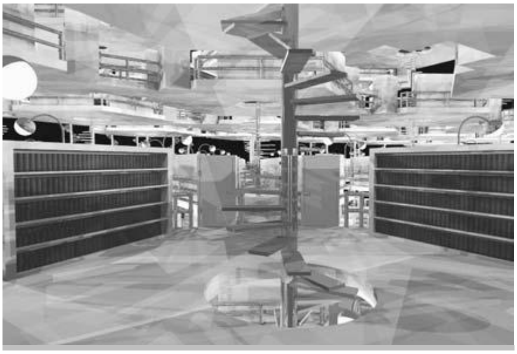

Universal Library Inside the library that goes on forever.
Joseph Formoso

symbols it contains, in addition to vast tracts of gibberish, every truth, falsehood, idea, novel, thought, and description of events, past and future, that are possible. It contains, writes Borges,

> *All*–the detailed history of the future, the autobiographies of the archangels, the faithful catalog of the Library, thousands and thousands of false catalogs, the proof of the falsity of those catalogs, a proof of the falsity of the *true* catalog, the gnostic gospel of Basilides, the commentary upon that gospel, the commentary on the commentary of that gospel, the true story of your death, the translation of every book into every language, the interpolations of every book into all books, the treatise Bede could have written (but did not) on the mythology of the Saxon people, the lost books of Tacitus.

The name "Universal Library" was first used by the German philosopher and science fiction writer Kurt Lasswitz (1848–1910) as the title of a short story published in 1901. He, in turn, borrowed the concept from the German psychologist **Theodor Fechner**. But to get to the root of speculation about listing all possible combinations of words and meanings, we have to go back much further, to Ramon Lully (1235–1315), a Spaniard who was a missionary and a mystic philosopher and who had an idea that later became known as Lully's Great Art. His idea was simply this: if one property of a thing is chosen, say the color of blood, and all the possibilities for that property are listed–blood is green, blood is yellow, etc.–then one of them must be true. One list alone might not be enough to point out the truth. However, other lists could be made that would eliminate some of the possible colors. Done in the right way, the one and only true answer should emerge. Lully even tried to build a device that used a series of concentric rings to bring different combinations of words into alignment. Eventually, the idea came down to Fechner who ruminated on the idea of permuting all combinations of letters to express all possible statements and concepts. There are, however, two barriers to this dream of ultimate truth. The first is that there isn't enough matter or space in the universe to represent all the different ways that book-length sequences of letters can expressed. Second, even if there were, it would take an all-seeing, all-knowing intelligence to sort the rare grain of meaningful wheat from the vast quantities of vapid chaff. See also **monkeys and typewriters**.

universal set

The **set** that contains all elements capable of being accepted to the problem. Also known as the *universe,* as in the **universe of discourse**, it is usually denoted *U.*

universe of discourse

The part of the world under discussion; more precisely, the **set** of all objects presumed or hypothesized to exist for some specific purpose. Objects may be concrete (e.g., a specific carbon atom, Confucius, the Sun) or abstract (e.g., the number 2, the set of all integers, the concept of justice). Objects may be primitive or composite (e.g., a circuit that consists of many subcircuits). Objects may even be fictional (e.g., a unicorn, Sherlock Holmes). The universe of discourse is a familiar concept in **logic**, linguistics, and mathematics.

unknown

A quantity, denoted by a letter, that is to be found by solving one or more equations.

untouchable number

A number that is not the sum of the **aliquot part**s of any other number. The first few untouchable numbers are 2, 5, 52, 88.

up-arrow notation

See **Knuth's up-arrow notation**.

upside-down picture

A picture (or figure) that, when inverted, looks the same or changes into the picture of a different subject. Possibly the most remarkable examples of upside-down art were the cartoons drawn by Gustave Verbeek for the *Sunday New York Herald* in the early 1900s. The first part of the cartoon is read normally; then the newspaper is turned

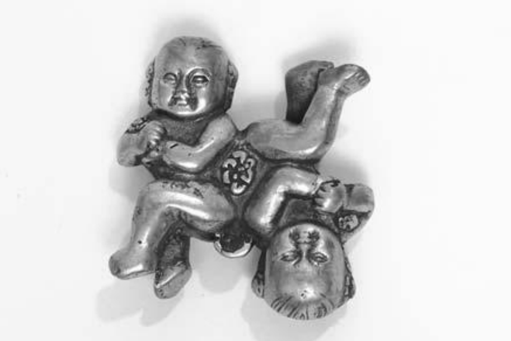

upside-down picture An upside-down figure from China.
Sue & Brian Young/Mr. Puzzle Australia, www.mrpuzzle.com.au

through 180° and the second part read from the same boxes in reverse order. As if by magic, Little Lady Lovekins transforms into Old Man Muffaroo, a giant fish becomes a giant bird, and a pouncing tiger turns into a tiger buried under a pile of stones.

Ussher, James (1581–1656)

An Irish clergyman, born in Fishamble Street, Dublin, who studied at Trinity College, Dublin, and later became a fellow there. He entered the Church and eventually became Archbishop of Armagh. In 1650 he published his famous assertion that the Creation had taken place "upon the entrance of the night preceding" Sunday, October 23, 4004 B.C.

utility function

The lesson of the **St. Petersburg paradox** is that people do not play games as if they are maximizing the expected monetary value they receive. However, certain rationality assumptions about the way people behave, known as the *von Neumann and Morganstern axioms,* imply that people *do* act as though they are maximizing *something.* This "something" is often referred to as a utility function.

V

vampire number

A natural number x that can be factorized as $y \times z$ in such a way that the number of occurrences of a particular digit in the representation of x in a given **base** (say 10) appears the same number of times in the representations in that same base of y and z together. For example, 2,187 is a vampire number since $2{,}187 = 21 \times 87$; similarly 136,948 is a vampire because $136{,}948 = 146 \times 938$. Vampire numbers are a whimsical idea that was introduced by Clifford Pickover in 1995.[253]

van der Pol, Balthazar (1889–1959)

A Dutch electrical engineer who began the modern experimental study of **dynamical systems** in the 1920s and '30s. Van der Pol discovered that electrical circuits employing vacuum tubes display stable oscillations, now called limit cycles, but that when these circuits are driven with a signal whose frequency is near that of the limit cycle, the periodic response shifts its frequency to that of the driving signal. The resulting waveform, however, can be quite complicated and contain a rich structure of harmonics and subharmonics. In 1927, van der Pol and his colleague van der Mark reported that an "irregular noise" was heard at certain driving frequencies between the natural entrainment frequencies. It's now clear that, without realizing it, they had described one of the first experimental instances of **chaos**.

vanishment puzzle

A **mechanical puzzle** in which the total area of a collection of pieces, or the number of items in a picture, appear to change following some manipulation. A well-known puzzle, first seen in 1868, involves an 8 × 8 square that is divided up into two triangles and two trapezoids. The pieces are reassembled into an oblong shape that measures 5 × 13. Where did the extra bit of area come from? The answer is that final shape is not perfectly rectangular but instead has a narrow gap that runs the length of one of the diagonals. Most famous of all vanishment puzzles is **Get off the Earth**.

variable

An **unknown** that has no fixed quantitative value.

vector

A quantity that is specified by a number, indicating size or "magnitude," and a direction; for example, "80 kilometers per hour, heading due south." More generally, a vector is any element of a **vector space** and is also a type of **tensor**. A vector is usually shown in a graph or other diagram as an arrow whose length and direction represent the vector's magnitude and direction, respectively. In n-dimensional space, it is easy to deal with vectors algebraically in the form of n-tuples, which are ordered lists (one-dimensional arrays) of n components.

vector space

Also known as a *linear space,* the most fundamental concept in **linear algebra**. It is a generalization of the set of all geometric **vector**s and is used throughout modern mathematics. Like the concepts of **group**, **ring**, and **field**, that of a vector space is entirely abstract.

Venn diagram

A simple way of representing **set**s and subsets, which makes use of overlapping circles. Venn diagrams are named after the Englishman John Venn (1834–1923), a fellow of Cambridge University. Venn was a cleric in the Anglican Church, an authority on what was then called "moral science," the compiler of a massive index of all Cambridge alumni, and a rather mundane mathematician who worked in logic and probability theory. The diagrams he used for representing **syllogism**s appear to have been first called "Venn diagrams" by Clarence Irving in his book *A Survey of Symbolic Logic* in 1918. However, Venn was lucky to be so immortalized. Both Gottfried **Leibniz** and Leonhard **Euler** used very similar forms of representation many years earlier.

vertex

The point where two sides of a closed figure, or two sides of an angle, meet; otherwise known as a *corner.* A cube, for example, has eight vertices.

vertex figure

The **polygon** that appears if a **polyhedron** is **truncate**d at a **vertex**. The vertex figure of a cube, for example, is an equilateral triangle. To ensure consistency, the truncation may be done at the midpoints of the edges.

vertical-horizontal illusion

Vertical lines looks considerably longer than horizontal ones of the same length. For centuries, it has been known

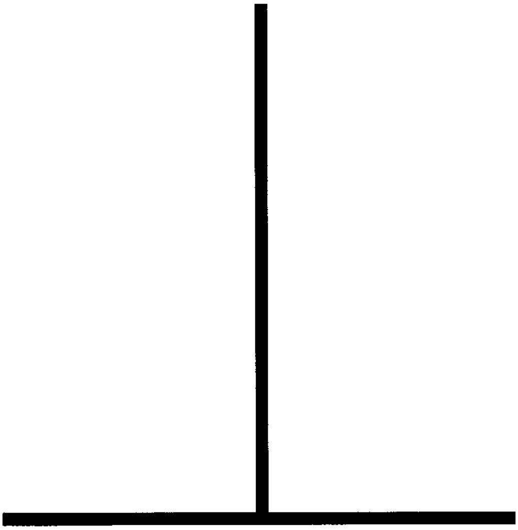

vertical-horizontal illusion

that the height of trees and buildings is perceived to be greater than the horizontal distance between them. This may be because we live in gravity. Rotate the book and you will see that the vertical line in the diagram appears to shrink in relation to what was originally the horizontal line. See also **distortion illusion**.

Vesica Piscis

Literally "fish's bladder," the almond shape formed when two identical circles overlap so that the outside edge of each just touches the center of the other. The Vesica Piscis appears as the Christian fish symbol and also frequently in medieval art and architecture. Most significantly, it provides the template for the pointed Gothic **arch**.

Vickrey auction

A silent auction in which the object being sold goes to the *second-highest bidder.* In a normal silent auction prospective buyers place their bids in sealed envelopes and the highest bid wins. This makes it risky for the seller because if the object is highly valuable and all buyers think they are the only ones who recognize this fact, they may offer much less than they think the object is actually worth. However, a Vickrey auction induces people to bid truthfully. Why? Because when all other bids are fixed and unknown to a given bidder, that bidder's optimal strategy is to bid what she thinks the object is worth. Suppose Alice places a bid for an antique vase. Let V be the amount that Alice thinks the vase is actually worth, and B the bid that she actually makes. Let M be the maximum of all other bids. If M is more than V, then Alice should set her bid B less than or equal to V, so that she does not get the vase for more than she thinks it is worth. If M is less than V, then Alice should set $B = V$, because if she bids any less, she will not get the vase any cheaper, and she may lose it altogether.

vigesimal

Of, relating to, or based on the number 20; the term comes from the Latin *vigesimus* for "twentieth." Mayan arithmetic, which took account of all the toes as well as the fingers, used a vigesimal system. In place of the multiples of 10 used in the decimal system: 1; 10; 100; 1,000; 10,000, . . . , the Mayans dealt in multiples of 20: 1; 20; 400; 8,000; 160,000; . . . .

vinculum

The bar that is placed over repeating decimal fractions to indicate the portion of the pattern that repeats. In the original Latin, *vinculum* referred to a small cord for binding the hands or feet. The symbol was once used in the same way that parentheses and brackets are now used to bind together a group of numbers or symbols. Originally the line was placed under the items to be grouped. What today might be written $7(3x + 4)$ the early users of the vinculum would write $\underline{3x + 4}\ 7$. Sometimes the horizontal fraction bar is called a vinculum as it binds the numerator and denominator into a single value.

Vinogradov's theorem

Every sufficiently large odd number can be expressed as the sum of three **prime numbers**. The theorem was named after the Russian mathematician Ivan Vinogradov (1891–1983) who proved it in 1937. This is a partial solution of the **Goldbach conjecture** and is related to **Waring's conjecture**.

Viviani's curve

The space curve that marks the intersection of the cylinder $(x - a)^2 + y^2 = a^2$ and the sphere $x^2 + y^2 + z^2 = a^2$. It is given by the parametric equations:

$$x = a(1 + \cos t)$$
$$y = a \sin t$$
$$z = 2a \sin(\tfrac{1}{2} t)$$

Viviani's theorem

For a given point inside an equilateral **triangle**, the sum of the perpendicular distances from the point to the sides is equal to the height of the triangle. If the point is outside the triangle, the relationship still holds if one or more of the perpendiculars is treated as a negative value.

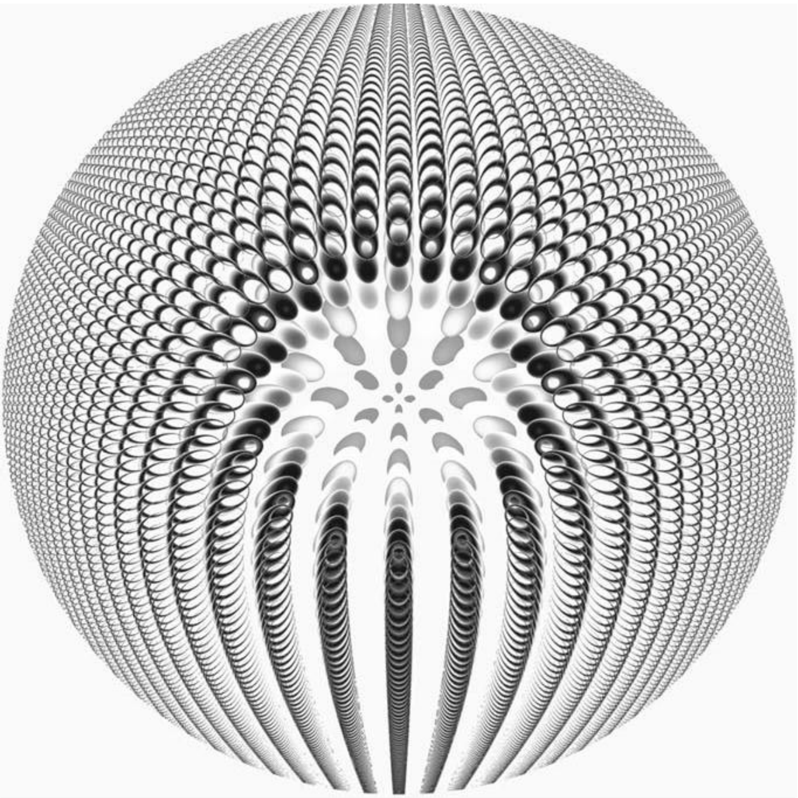

Volvox fractal *Jos Leys, www.josleys.com*

Viviani's theorem generalizes to a regular *n*-sided **polygon**: the sum of the perpendicular distances from an interior point to the *n* sides being *n* times the **apothem** of the figure. The theorem is named for Vincenzo Viviani (1622–1703), a pupil of Galileo and Torricelli, who is also remembered for a reconstruction of a book on the **conic section**s of **Apollonius** and for finding a way of **trisecting an angle** through the use of an equilateral hyperbola.

volume

The measure of space occupied by a solid body.

Volvox fractal

A **fractal** that is similar in appearance to *Volvox*–a unicellular life form that lives in spherical colonies with thousands of members.

von Neumann, John (1903–1957)

> *Young man, in mathematics you don't understand things, you just get used to them.*

A Hungarian-American mathematician who made important contributions to **set theory**, computer science, economics, and **quantum mechanics**. He received a Ph.D. in mathematics from the University of Budapest and later he worked at the Institute for Advanced Study in Princeton. *Theory of Games and Economic Behavior,*[230] which he coauthored with Oskar Morgenstern in 1944, is considered a seminal work in the field of **game theory**. Von Neumann devised the von Neumann architecture used in all modern computers and studied cellular automata (see **cellular automaton**) in order to construct the first examples of self-replicating automata, now known as **von Neumann machine**s. Von Neumann had a mind of great

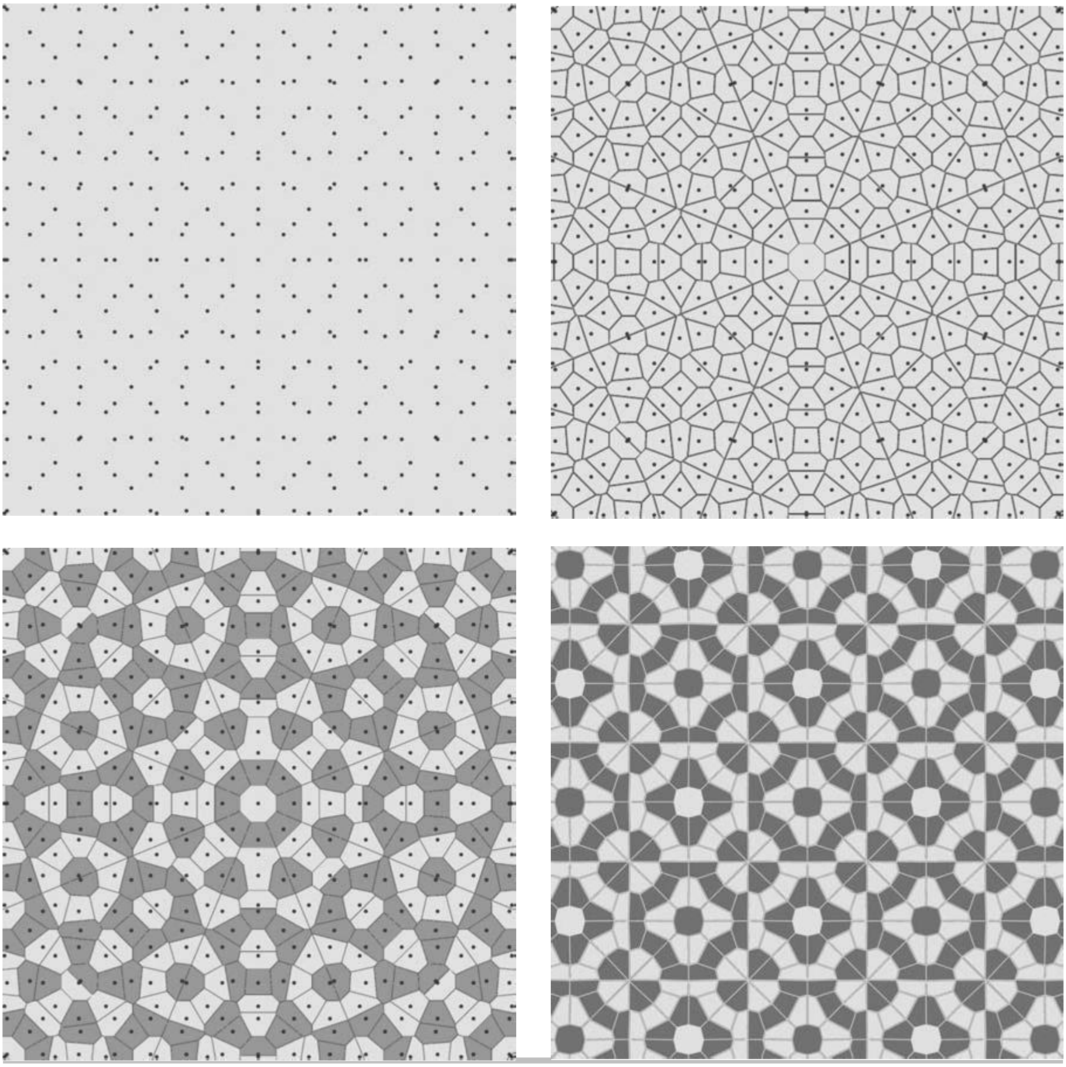

Voronoi diagram Top (left to right): Start from a grid of points. Draw lines that limit areas closest to the points of the grid. Bottom (left to right): Color the areas in some fashion. An infinite variety of patterns is possible. *Jos Leys, www.josleys.com*

ingenuity, nearly total recall of what he'd learned, immense arrogance, and a great love of jokes and humor.[259]

von Neumann machine

(1) A model for a computing machine that uses a single storage structure to hold both the set of instructions on how to perform the computation and the data required or generated by the computation. John **von Neumann** helped to create the model as an example of a general-purpose computing machine. By treating the instructions in the same way as the data, the machine could easily change the instructions. In other words the machine was reprogrammable. (2) A self-replicating machine. In principle, if a machine (e.g., an industrial

robot) could be given enough capability, raw material, and instructions, then that robot could make an exact physical copy of itself. The copy would need to be programmed in order to do anything. If both robots were reprogrammable, then the original robot could be instructed to copy its program to the new robot. Both robots would now have the capability of building copies of themselves. Since such a machine is capable of reproduction, it could arguably qualify as a simple form of life.

Voronoi diagram

Also known as a *Dirichlet tesselation,* a partitioning of space into cells, each of which consists of the points closer to one particular object than to any others. More specifically, in two dimensions a Voronoi diagram consists of breaking up a plane containing *n* points into *n* convex **polygon**s in such a way that each polygon contains exactly one point and every point in a given polygon is closer to its central point than to any other. Voronoi diagrams, their boundaries (known as *medial axes*), and their **dual**s (called *Delaunay triangulations*) have been reinvented, given different names, generalized, studied, and applied many times over in many different fields. Voronoi diagrams tend to be involved in situations where a space should be partitioned into "spheres of influence"; examples include models of crystal and cell growth and protein molecule volume analysis.

vulgar fraction

See **common fraction**.

W

walking in the rain problem

The question is whether you get wetter by walking or running a given distance through rain that is falling at a constant rate. An early appearance of this problem was in Bagley's *Paradox Pie* (1944). A simple answer is that moving faster is better. If the rain falls vertically and the density of water in the air is assumed constant, then, no matter what your speed, you'll sweep out the same volume and will always get the same amount of water hitting your front. However, running rather than walking will reduce the amount of water landing on your head.

Wallis, John (1616–1703)

The most influential English mathematician before Isaac **Newton** and an important contributor to the origins of **calculus**. He was a skilled linguist, was one of the first to proclaim in public Harvey's discovery of the circulation of the blood, and had an extraordinary memory for figures. His *Arithmetica Infinitorum* was described as "the most stimulating mathematical work so far published in England" and introduced the symbol ∞ for infinity (see also **aleph**). It contained the germs of the differential calculus, and it suggested to Newton, who was delighted by it, the **binomial theorem**.

Wallis formula

See **pi**.

wallpaper group

Also known as a *crystallographic group,* a distinct way to tile the plane that repeats indefinitely in two dimensions; that is, a collection of two-dimensional symmetric patterns on a plane surface, containing two nonparallel **translation**s (see **periodic tiling**). There are only 17 kinds of these patterns, known as isometries (see **isometry**), each uniquely identified by its translation and rotation symmetries, as discovered in the late nineteenth century by E. S. **Fedorov** and, independently, by the German A. M. Schoenflies and the Englishman William Barlow. Thirteen of the isometries include some kind of rotational symmetry, while four do not; twelve show rectangular symmetries, while five involve hexagonal symmetries. Every two-dimensional repetitive pattern in wallpaper, textiles, brickwork, or the arrangement of atoms in a plane of a crystal is just a minor variation on one of these 17 patterns.

Wari

See **Mancala**.

Waring's conjecture

A hypothesis given, without proof, by the English mathematician Edward Waring (1734–1798) in his *Meditationes algebraicae* (1770). It states that for every number *k*, there is another number *s* such that every natural number can be represented as the sum of *s* *k*th powers. For example, every natural number can be written as a sum of 4 squares, 9 cubes and so on. Waring's conjecture was first proven in full by David **Hilbert** in 1909.

weak inequality

An **inequality** that permits the equality case. For example, *a* is less than or equal to (≤) *b*.

Weierstrass, Karl Wilhelm Theodor (1815–1897)

A German mathematician who is considered the father of modern **analysis**. Compelled by his father to study law, he instead spent four years at the University of Bonn, fencing, drinking, and reading math. He left under a cloud and

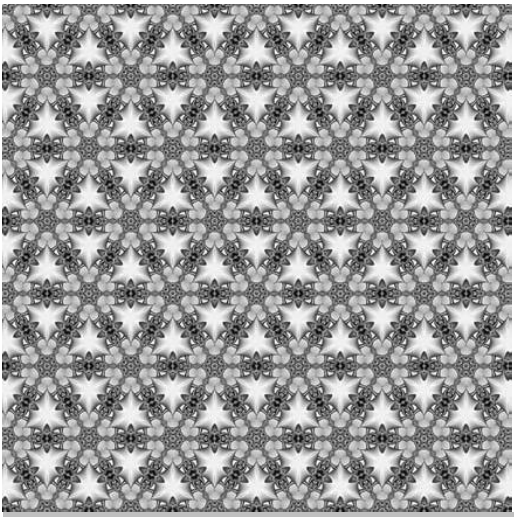

wallpaper group A pattern made from one of the wallpaper groups. *Jos Leys, www.josleys.com* ended up teaching in secondary schools for many years. In 1854 he published a paper, written 14 years earlier when he was fresh out of college, in *Crelle's Journal*, on *Abelian functions* which completed work that Niels **Abel** and Karl **Jacobi** had begun. Its importance was immediately recognized and Weierstrass was appointed a professor at the Royal Polytechnic School and a lecturer at the University of Berlin. He went on to give the first rigorous definitions of **limit**, **derivative**, differentiability, and **convergence**, and investigated under what conditions a power series will converge.

Weierstrass's nondifferentiable function

The earliest known example of a *pathological function*–a **function** that gives rise to a **pathological curve**. It was investigated by Karl **Weierstrass**, but had been first discovered by Bernhard **Riemann**, and is defined as:

$$f(x) = \sum_{k=1}^{\infty} \frac{\sin(\pi k^2 x)}{\pi k^2}$$

The Weierstrass function is everywhere continuous (see **continuity**) but nowhere differentiable; in other words, no **tangent** exists to its curve at any point. Constructed from an infinite sum of trigonometric functions, it is the densely nested oscillating structure that makes the definition of a tangent line impossible.

weighing puzzles

See **measuring and weighing puzzles**.

weird number

See **abundant number**.

Wessel, Caspar (1745–1818)

A Norwegian surveyor whose mathematical fame rests on a single paper, published in 1799, that gave the first geometrical interpretation of **complex number**s. His priority in this discovery, however, went unrecognized for many years. Thus what should really be called a *Wessel diagram* is known instead as an **Argand diagram** after the man whose work on the same subject, published in 1806, first came to the attention of the mathematical world. Wessel's paper, by contrast, wasn't noticed by the mathematical community until 1895 when the Danish mathematician Sophus Juel drew attention to it and, in the same year, Sophus **Lie** republished Wessel's paper. Astonishingly, Wessel's remarkable work was not translated into English until 1999–its bicentenary!

Weyl, Hermann Klaus Hugo (1885–1955)

A German mathematician (known as "Peter" to his close friends) whose work involved symmetry theory, **topology**, and **non-Euclidean geometry**. Weyl studied under David **Hilbert** at Göttingen. Then, as a colleague of Albert Einstein at Zurich 1913, he got involved with **relativity theory** and came to believe (erroneously) that he had found a way to unite gravity and electromagnetism. From 1923 to 1938 he concentrated on **group** theory and made some important contributions to **quantum mechanics**. As the Nazi tide swept over Europe, Weyl came to the United States and spent the rest of his career at the Institute for Advanced Studies at Princeton. He said: "My work always tried to unite the truth with the beautiful, but when I had to choose one or the other, I usually chose the beautiful." See also **beauty and mathematics**.

wff

A well-formed formula.

what color was the bear?

A hunter walks one mile due south, then one mile due east, then one mile north and arrives back at his starting point. He shoots a bear. What color was it? Such a trip is possible if the hunter starts from one of the geographical poles, then circumnavigates the sides of a spherical triangle (see **elliptical geometry**). Since there are no bears in Antarctica, the trip is assumed to have taken place in the Arctic where there are polar bears and thus the answer is "white." Versions of this problem began to appear in the 1940s. A closer examination reveals that there are many more points on the globe, other than an exact pole, from which the hunter could have begun his trek. One example is any point (of which there are an infinite number) on a circle drawn at a distance of slightly more than 1 + $1/2\pi$ mile (about 1.16 miles) from a pole–"slightly more" because of Earth's curvature. But this is not all. The hunter could also satisfy the conditions by starting at points closer to the pole, so that the walk east would carry him exactly twice around the pole, or three times, and so on. Of course, the bear would still be white (except that polar bears don't live that far north!).

wheat and chessboard problem

According to one myth, **chess** was invented by Grand Vizier Sissa Ben Dahir and then given to King Shirham of India. The king was so pleased that he offered his subject a great reward in gold, but the wily vizier said that he would be happy merely to have some wheat: one grain for the first square of the chessboard, two grains for the second square, four for the third, and so on, doubling each time. The king thought this was a very modest request, granted it, and asked for a bag of wheat to be brought in. However the bag was emptied by the twentieth square. The king asked for another bag, but then realized that this entire bag was needed for the next square. In 20 more squares, as many bags would have been exhausted as there

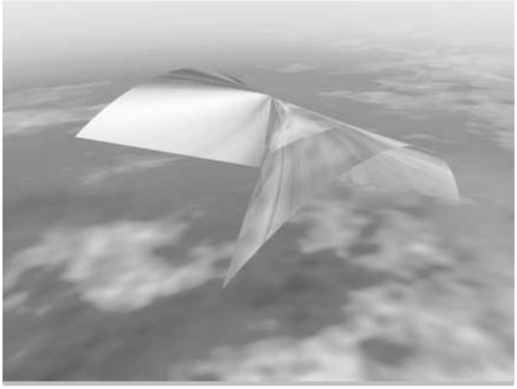

Whitney's umbrella *Takashi Nishimura*

were grains in the first bag! The number of grains on the sixty-fourth square would have been 2^{63}, and the total for the whole board $2^{64} - 1 = 18{,}446{,}744{,}073{,}709{,}551{,}615$. This is more wheat than in the entire whole world; in fact, it would fill a building 40 km long, 40 km wide, and 300 meters tall. See also **Tower of Brahma**.

Whitney's umbrella

A strange-looking, whimsically named geometrical object first studied by Hassler Whitney in the 1940s. It can be pictured as a self-intersecting rectangle in three dimensions. A *pinch point,* also known as a *Whitney singularity* or a *branch point* occurs at the top endpoint of the segment of self-intersection: every neighborhood of the pinch point intersects itself.

Wiener, Norbert (1894–1964)

An American mathematician who established the subject of **cybernetics.** As a precocious youngster, Wiener hopped from subject to subject at college, finally earning a Ph.D. in mathematics from Harvard at 19, before embarking on an even more erratic early career that took him into a variety of activities, including journalism. Having settled upon mathematical research, obtaining a post at the Massachusetts Institute of Technology in 1919, he nevertheless continued to range across fields, from random processes, including ergodic theory (concerned with the onset of chaos in a system), to integral equations, quantum mechanics, and potential theory. Wartime work that involved applying statistical methods to control and communication engineering, led to him extending these studies into control and communication in complex electronic systems and in animals, especially humans–the science of cybernetics.

Wiles, Andrew (1953–)

The English mathematician who, in 1994, finally proved **Fermat's last theorem**. Wiles studied at Oxford (B.A. 1974) and Cambridge (Ph.D. 1977) and has held posts at Cambridge, Oxford, and Princeton. From the mid-1980s his work was focused on proving a proposition known as the *Shimura-Taniyama conjecture,* since from this, it had been shown, Fermat's last theorem would follow. In 1993 he gave a series of lectures at Cambridge University ending on June 23, 1993. At the end of the final lecture he announced he had a proof of Fermat's last theorem. However, when the results were written up for publication, a subtle error was found. Wiles worked hard for about a year, helped in particular by a colleague, R. Taylor, and by September 19, 1994, having almost given up, he decided to have one last try. As he recalled, "suddenly, totally unexpectedly, I had this incredible revelation. It was the most important moment of my working life. . . . [I]t was so indescribably beautiful, it was so simple and so elegant . . . [that] I just stared in disbelief for twenty minutes, then during the day I walked round the department. I'd keep coming back to my desk to see it was still there–it was still there."

In 1994 Wiles was appointed Eugene Higgins Professor of Mathematics at Princeton. His paper that proves Fermat's last theorem is called "Modular elliptic curves and Fermat's Last Theorem" and appeared in the *Annals of Mathematics* in 1995.[296]

Wilson's theorem

Any number p is a **prime number** if, and only if, $(p - 1)! + 1$ is divisible by p. We can easily check this for some small numbers: $(2 - 1)! + 1 = 2$, which is divisible by 2; $(5 - 1)! + 1 = 25$, which is divisible by 5; $(9 - 1)! + 1 = 40{,}321$, which is not divisible by 9. The theorem is named for Sir John Wilson (1741–1793), who came across it (but left no formal proof) while he was a student at Peterhouse College, Cambridge. Wilson went on to become a judge and seems to have done little else in mathematics. The theorem was first published and named after Wilson by Edward Waring (1734–1798) around 1770. However, it is now clear that the result was known to Gottfried **Leibniz** and perhaps, much earlier to Ibn al-Haytham (965–1040). The first known proof was provided by Joseph **Lagrange**.

winding number

The number of times a **closed** curve in the plane passes around a given point in the counterclockwise direction.

wine

The creation of a fine wine may be a complex business, but it is not mathematically intractable according to a Californian winemaker and businessman, Leo McCloskey, who runs a consulting firm. McCloskey's formula takes into account color, fragrance, and flavor and attempts to predict wine quality two to three years in advance. The formula stems from a comparison of key chemical characteristics for a range of wines with what experts made of the final product. Putting this information into a database, McCloskey has claimed, enables the math of wine flavor to be derived. More than 50,000 wines were analyzed for the database but, in each case, only a handful of the 400 to 500 constituent chemicals were considered crucial flags to future color, taste, and bouquet. The proportion of tannins and phenols is particularly crucial in determining a wine's future, thus reducing the number of essential flags for an accurate mathematical model to just 10 or 20. With this system, McCloskey believes, he can predict what a wine will be like before it is bottled, giving winemakers the opportunity to modify the flavor if necessary. However, whether any wine formula can effectively take into account the subjective impressions of the end user remains to be seen. See also **bottle sizes**.

Witch of Agnesi curve

See **Agnesi**.

word puzzles

1. What positive integer, when spelled out, has a Scrabble score equal to that integer?
2. What is the only English word that ends in "mt?"
3. Only four words in the English language end in "-dous." What are they?
4. What is the shortest complete sentence in the English language?
5. Name the only word (not including proper names) in the English language that has two "i" 's together.
6. Name a one-syllable word that becomes a three-syllable word by adding one letter to the end of it.
7. What word begins and ends with "und?"

Solutions begin on page 369.

word trivia

- The hardest word to define briefly is thought to be the word *mamihlapinatopai* from the Fuegian language spoken by the natives of the Andaman Islands. Its simplest definition is "two people looking at each other without speaking hoping that the other will offer to do something which both parties desire but neither are willing to do."
- No word in the English language rhymes with orange, silver, purple, or month.
- The only 15-letter word that can be written without repeating a letter is *uncopyrightable.*
- The combination "ough" can be pronounced in nine different ways. The following sentence contains them all: "A rough-coated, dough-faced, thoughtful ploughman strode through the streets of Scarborough; after falling into a slough, he coughed and hiccoughed."
- *Facetious* and *abstemious* contain all the vowels in the correct order, as does *arsenious,* meaning "containing arsenic."
- The longest word in the English language, according to the Oxford English Dictionary, is *pneumonoultramicroscopicsilicovolcanoconiosis.* The only other word with the same amount of letters is its plural: *pneumonoultramicroscopicsilicovolcanoconioses.* It means an infection of the lungs.
- The longest English word with one vowel is *strengths.*
- The "sixth sick sheik's sixth sheep's sick" is said to be the toughest tongue twister in the English language.
- Richard Millhouse Nixon was the first U.S. president whose name contains all the letters from the word *criminal.*
- In Scotland, a new game was invented. It was called Gentlemen Only Ladies Forbidden and thus the word GOLF entered the vocabulary.
- The longest English words that don't use any of vowels a, e, i, o, or u are "rhythm" and "syzygy."
- The verb *cleave* is one of many English words with two synonyms which are antonyms of each other: adhere and separate.
- There is a seven letter word in the English language that contains ten words without rearranging any of its letters, *therein:* the, there, he, in, rein, her, here, ere, therein, herein.
- The longest English word that consists entirely of consonants is *crwth,* which is from the fourteenth century and means crowd.
- The word *trivia* comes from the Latin *tri-* + *via,* and means "three streets." This is because in ancient times, at an intersection of three streets in Rome, they would have a type of kiosk where ancillary information was listed. You might be interested in it, you might not, hence they were bits of "trivia."

worldline

The path of an object through **space-time**. On a *Minkowski diagram,* in which the three dimensions of space are represented by the horizontal axis and time is represented by the vertical axis, worldlines appear as wiggly curves extending from the past into the future. *My*

Worldline is the title of the autobiography of George Gamow, the Ukrainian-American physicist.

Wundt illusion

A **distortion illusion** devised by the German "father of experimental psychology," and one-time assistant to the physicist Hermann von Helmholtz at Heidelberg, Wilhelm Wundt (1832–1920). In the figure, two horizontal lines are both straight, although they look as if they bow in at the middle. The distortion is induced by crooked lines on the background, as in **Orbison's illusion**. The simplest of all distortion illusions–the **vertical-horizontal illusion**–was also discovered by Wundt.

Wythoff's game

A variation on the game of **Nim** suggested by W. A. Wythoff in 1907. It is played with two heaps of counters in which a player may take any number from either heap or the same number from both. The player who takes the last counter wins.

X

x

The most frequently used symbol for an unknown in an expression or equation.

***x*-axis**

The horizontal axis of a two-dimensional plot in **Cartesian coordinates**.

X-pentomino

A pentomino, a five-square **polyomino**, in the shape of the letter X.

Y

yard

A unit of distance equal to 3 feet (36 inches) or 0.9144 meters. The name originates with the old German word *gazdaz* for a staff or stick, which could be used for measurement. This changed to *gierd* in Old English and eventually to *yard.* The yard-arm of a sailing ship–a tapered spar used to support a square sail–harkens back to the earlier meaning. In France, the equivalent of the yard measure is called a "verge," from the Latin *virga* for a stick or rod.

y-axis

The vertical axis of a two-dimensional plot in **Cartesian coordinates**.

yin-yang symbol

See **Great Monad**.

yocto-/yotta-

The prefix yocto-, for 10^{-24}, derives from the Greek *oktakis* ("eight times"). The prefix yotta-, for 10^{24}, comes from the same source.

Z

Zeller's formula

A formula invented by the German clergyman Christian Zeller (1824–1899) for figuring out the day of the week of a given date, without the need of tables. Let

J be the century-number,
K the last two digits in the year,
e the residue, which remains when J is divided by 4,
m the number of the month,
q the day of the month,
h the number of the day of the week.

Then h is the remainder that results when

$$h = q + 26(m + 1)/10 + K + K/4 - 2e$$

is divided by 7. For the formula to work, January and February have to be taken as months 13 and 14 of the preceding year. For example, Frederick the Great was born on 24 January 1712, so $J = 17$, $e = 1$, $K = 11$ (not 12 because of the special way of numbering the months), $m = 13$, and $q = 24$. Plugging these values into the formula, we get

$$\begin{aligned} &24 + 26(13 + 1)/10 + 11 + 11/4 - 2 \times 1 \\ &= 71 = 70 \times 1 + 1. \end{aligned}$$

The remainder h is 1, so Frederick the Great was born on the first day of the week, Sunday.

Zenodorus (c. 180 B.C.)

The Greek mathematician and philosopher who proposed in *On Isometric Figures* that the circle has the maximum area of all isoperimetric figures in a plane, and that the sphere has the maximum volume of all bodies with equal surface.

Zeno's paradoxes

A series of **paradox**es posed by the philosopher Zeno of Elea (c. 490–c. 425 B.C.). Little is known about Zeno's life. He was born in Elea (now Lucania) in southern Italy and was a friend and student of Parmenides. None of his writings survive but he is known to have written a book, which **Proclus** says contained 40 paradoxes. Four of these, which all concern motion, have had a profound influence on the development of mathematics. They are described in Aristotle's great work *Physics* and are called the Dichotomy, Achilles (and the Tortoise), the Arrow, and the Stadium.

The Dichotomy argues that "there is no motion because that which is moved must arrive at the middle of its course before it arrives at the end." In order to traverse a line segment it's necessary to reach the halfway point, but this requires first reaching the quarter-way point, which first requires reaching the eighth-way point, and so on without end. Hence motion can never begin. This problem isn't alleviated by the well-known infinite sum $1/2 + 1/4 + 1/8 + \ldots = 1$ because Zeno is effectively insisting that the sum be tackled in the reverse direction. What is the first term in such a series?

Zeno's paradox of Achilles is told by Aristotle in this way: "The slower when running will never be overtaken by the quicker; for that which is pursuing must first reach the point from which that which is fleeing started, so that the slower must necessarily always be some distance ahead." Thus, Achilles, however fast he runs, will never catch the plodding Tortoise who started first. And yet, of course, in the real world, faster things do overtake slower ones. So how is the paradox to be solved? The German set theorist Adolf Frankel (1891–1965) is one of many modern mathematicians (Bertrand Russell is another) who have pointed out that 2,000 years of attempted explanations have not cleared away the mysteries of Zeno's paradoxes: "Although they have often been dismissed as logical nonsense, many attempts have also been made to dispose of them by means of mathematical theorems, such as the theory of convergent series or the theory of sets. In the end, however, the difficulties inherent in his arguments have always come back with a vengeance, for the human mind is so constructed that it can look at a continuum in two ways that are not quite reconcilable."

zepto-/zetta-

The prefix zepto-, for 10^{-21}, derives from the Greek *heptakis* ("seven times"). The prefix zetta-, for 10^{21}, comes from the same source.

zero

The **integer**, denoted 0, which, when used as a counting number, indicates that no objects are present. It is the only integer that is neither negative nor positive. Zero is both a **number** and a **numeral**. The number zero is the size of the **empty set** but it is not the empty set itself, nor is it the same thing as **nothing**. The numeral or digit zero is used in positional **number systems**, where the position of a digit signifies its value, with successive positions having higher values, and the digit zero is used to skip a position. The earliest roots of the numeral zero stretch back 5,000 years to the Sumerians in Mesopotamia, who inserted a slanted double wedge between cuneiform characters for numbers, written positionally, to indicate a number's absence. The symbol changed over time as positional notation made its way to India, via the Greeks (in whose own culture zero made a late and only occasional appearance). Our word *zero* derives from the Hindi *sunya* for "void" or "emptiness," through the Arabic *sifr* (which also gives us *cipher*), and the Italian *zevero*. As a number in its own right, aside from its use as a position marker, zero took a much longer time to become established, and even now is not equal in status to other numbers: division by zero is not allowed.

zero divisors

Nonzero elements of a **ring** whose product is 0.

zero of a function

See **root**.

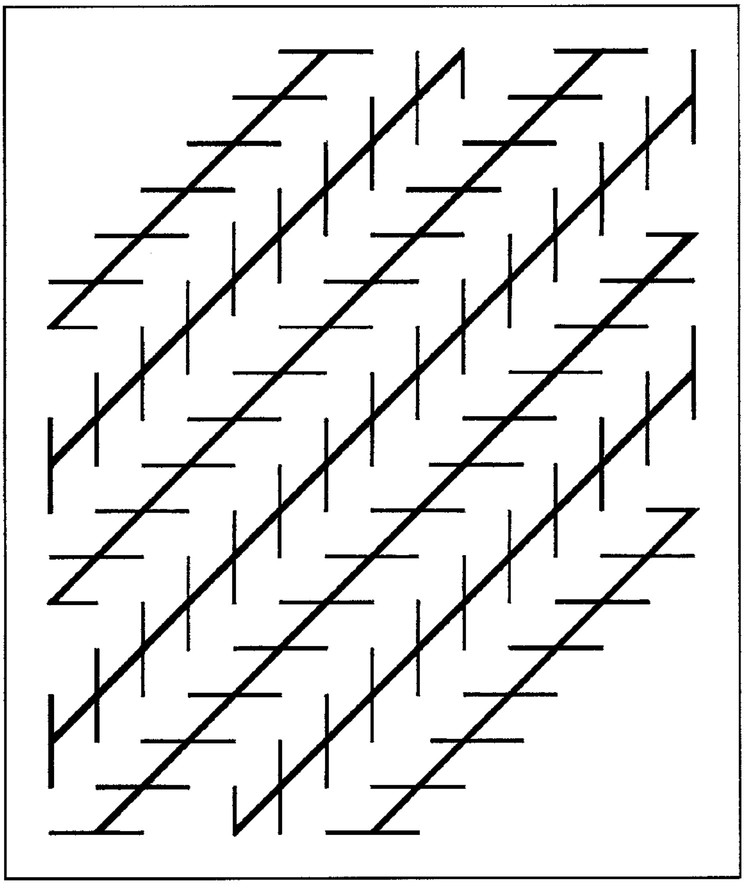

Zöllner illusion

zero-sum game

In **game theory**, a game in which a win for one player results in an equal but opposite loss for the other players.

zeta function

A **function** that has certain properties and is calculated as an infinite sum of negative powers. The most commonly encountered zeta function is the **Riemann zeta function**.

zigzag

A general word for a type of curve that consists of several straight lines joined at points. A zigzag usually goes alternately from side to side.

Zöllner illusion

A line **distortion illusion** first published by the astronomer Johann Zöllner in 1860. The diagonal lines, although parallel, appear not to be. The illusion was one of a series specifically designed to cause errors in optical equipment of that time. They did cause errors, and also great concern among scientists over the validity of all human observations. See also **Poggendorff illusion**.

zombie

A hypothetical being that behaves like us and may share our functional organization and even, perhaps, our neurophysiological makeup, but lacks consciousness or any form of subjective awareness. The concept is used in discussions of **artificial intelligence**.

zone

The portion of a **sphere** between two parallel planes.

zonohedron

A **polyhedron** in which the faces are all **parallelogram**s or parallel-sided. The faces of a zonohedron can be grouped into *zones*–encircling bands of faces which share a common edge direction (and length).

References

1. Abbott, Edwin A. *Flatland: A Romance of Many Dimensions.* London: Seely and Co., 1884. Reprint, Mineola, N.Y.: Dover, 1992.

2. Abraham, R. M. *Diversions and Pastimes.* London: Constable & Co., 1933. Reprinted, 1964.

3. Ackermann, A. S. E. *Scientific Paradoxes and Problems.* London, 1925.

4. Adams, Douglas. *Life, the Universe and Everything.* New York: Harmony Books, 1982.

5. Adams, Douglas. *The Hitchhiker's Guide to the Galaxy.* New York: Ballantine, 1995.

6. Ahrens, W. *Mathematische Unterhaltungen und Spiele.* Leipzig, Germany: Teubner, 1910.

7. Ainley, Stephen. *Mathematical Puzzles.* London: Bell, 1977.

8. Allais, M. "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine." *Econometrica,* 21: 503–546 (1953).

9. Amthor, A., and B. Krumbiegel. "Das Problema bovinum des Archimedes." *Zeitschrift für Mathematik und Physik,* 25: 121–171 (1880).

10. Andrews, William S. *Magic Squares and Cubes,* 2nd ed. Mineola, N.Y.: Dover, 1960.

11. Apéry, R. "Irrationalité de $\zeta(2)$ et $\zeta(3)$." *Astérisque,* 61: 11–13 (1979).

12. Appel, K., and W. Haken. "The Solution of the Four-Color Map Problem." *Scientific American,* 237: 108–121 (1977).

13. ApSimon, Hugh. *Mathematical Byways in Ayling, Beeling, and Ceiling.* New York: Oxford University Press, 1984.

14. Argand, R. *Essai sur une manière de représenter les quantités imaginaires dans les constructions géométriques.* Paris: Albert Blanchard, 1971. Reprint of the 2nd ed., published by G. J. Hoel in 1874. First published in 1806 in Paris.

15. Arno, Steven. "A Note on Perrin Pseudoprimes." *Mathematics of Computation,* 56(193): 371–376 (1991).

16. Arrow, Kenneth. *Social Choice and Individual Values.* New York: John Wiley & Sons, 1951.

17. Ascher, Marcia. *Ethnomathematics: A Multicultural View of Mathematical Ideas.* Pacific Grove, Calif.: Brooks/Cole Publishing Co., 1991.

18. Avedon, Elliot M., and Brian Sutton-Smith. *The Study of Games.* New York: John Wiley & Sons, 1971.

19. Averbach, Bonnie, and Orin Chein. *Mathematics: Problem Solving through Recreational Mathematics.* New York: W. H. Freeman, 1980.

20. Bagley, William A. *Puzzle Pie: A Unique Collection of Scientific Paradoxes, Posers and Oddities.* London: Vawser and Wiles, 1944.

21. Bailey, D. H., P. B. Borwein, and S. Plouffe. "On the Rapid Computation of Various Polylogarithmic Constants." *Mathematics of Computation,* 66(218): 903–913 (1997).

22. Bakst, Aaron. *Mathematical Puzzles and Pastimes.* New York: Van Nostrand, 1954.

23. Bakst, Aaron. *Mathematics: Its Magic and Mystery,* 3rd ed. Princeton, N.J.: Van Nostrand, 1967.

24. Ball, Walter William Rouse. *Mathematical Recreations and Problems.* London: Macmillan, 1892. (Ball, W. W. Rouse and H. S. M. Coxeter. *Mathematical Recreations and Essays,* 13th ed. Mineola, N.Y.: Dover, 1987.)

25. Banchoff, Thomas F. *Beyond the Third Dimension: Geometry, Computer Graphics, and Higher Dimensions.* New York: W. H. Freeman, 1990.

26. Barnsley, Michael. *Fractals Everywhere,* 2nd edition. San Francisco: Morgan Kaufmann, 1993.

27. Bayer, D., and P. Diaconis. "Trailing the Dovetail Shuffle to Its Lair." *Annals of Applied Probability,* 2(2): 294–313 (1992).

28. Beasley, John D. *The Ins and Outs of Peg Solitaire.* New York: Oxford University Press, 1985.

29. Beckmann, P. *A History of Pi,* 3rd ed. New York: Dorset Press, 1989.

30. Beiler, Albert H. *Recreations in the Theory of Numbers.* Mineola, N.Y.: Dover, 1964.

31. Bell, A. H. " 'Cattle Problem.' By Archimedes 251 B.C." *American Mathematical Monthly,* 2: 140 (1895).

32. Bell, E. T. *Men of Mathematics.* New York: Simon & Schuster, 1937.

33. Bell, Robbie, and Michael Cornelius. *Board Games Round the World: A Resource Book for Mathematical Investigations.* New York: Cambridge University Press, 1990.

34. Benford, F. "The Law of Anomalous Numbers."

Proceedings of the American Philosophical Society, 78: 551–572 (1938).

35. Bergholt, E. *Complete Handbook to the Game of Solitaire on the English Board of Thirty-Three Holes.* Routledge: London, 1920.

36. Berlekamp, Elwyn R., John Horton Conway, and Richard K. Guy. *Winning Ways for Your Mathematical Plays,* 2 vols. New York: Academic Press, 1982.

37. Berloquin, Pierre. *Le Jardin du Sphinx.* Paris: Dunod, 1981.

38. Berrondo, Marie. *Mathematical Games.* Englewood Cliffs, N.J.: Prentice Hall, 1983.

39. Besicovitch, A. S. "On Kakeya's Problem and a Similar One." *Mathematische Zeitschrift,* 27: 312–320, (1928).

40. Berry, Michael. "Quantum Physics on the Edge of Chaos." *New Scientist,* 116: 44–47 (1987).

41. Birtwistle, Claude. *The Calculator Puzzle Book.* New York: Bell Publishing Co., 1978.

42. Blatner, David. *The Joy of Pi.* New York: Walker, 1997.

43. Blocksma, Mary. *Reading the Numbers: A Survival Guide to the Measurements, Numbers, and Sizes Encountered in Everyday Life.* New York: Viking Penguin, 1989.

44. Blyth, Will. *Match-Stick Magic, Puzzles, Games and Conjuring Tricks.* London: C. A. Pearson Ltd., 1921.

45. Bogoshi, J., Naidoo, K., and Webb, J. "The Oldest Mathematical Artifact." *Mathematical Gazette,* 71: 458 (1987).

46. Bond, Raymond T. *Famous Stories of Code and Cipher.* New York: Rinehart & Co., 1947.

47. Bondi, Hermann. "The Rigid Body Dynamics of Unidirectional Spin." *Proceedings of the Royal Society,* A405: 265–274, 1986.

48. Bord, Janet. *Mazes and Labyrinths of the World.* London: Latimer, 1976.

49. Borel, Emile. *Algebre et Calcul des Probabilites,* vol. 184. Paris: Comptes Rendus Academie des Sciences, 1927.

50. Bouton, Charles L. "Nim, a Game with a Complete Mathematical Theory." *Annals of Mathematics,* Series 2. 3: 35–39, (1901–1902).

51. Brams, S. J., and A. D. Taylor. "An Envy-Free Cake Division Protocol." *American Mathematical Monthly,* 102: 9–19 (1995).

52. Brams, S. J., and A. D. Taylor. *Fair Division: From Cake-Cutting to Dispute Resolution.* New York: Cambridge University Press, 1996.

53. Brandreth, Gyles. *Numberplay.* New York: Rawson, 1984.

54. Branges, L., de, "A Proof of the Bieberbach Conjecture." *Acta Mathematica,* 154: 137–152 (1985).

55. Burger, Dionys. *Sphereland: A Fantasy about Curved Spaces and an Expanding Universe.* New York: Barnes & Noble, 1983.

56. Cadwell, J. H. *Topics in Recreational Mathematics.* New York: Cambridge University Press, 1966.

57. Carroll, Lewis. *Mathematical Recreations of Lewis Carroll.* 2 vols. Mineola, N.Y.: Dover, 1958.

58. Carroll, Lewis. *Symbolic Logic and the Game of Logic.* 2 vols. Mineola, N.Y.: Dover, 1958.

59. Carroll, Lewis. *Pillow Problems and A Tangles Tale.* New York: Dover, 1958.

60. Chaitin, G. J. "The Berry Paradox." *Complexity,* 1: 26–30 (1995).

61. Chow, T. Y. "The Surprise Examination or Unexpected Hanging Paradox." *American Mathematical Monthly,* 105: 41–51 (1998).

62. Church, Alonzo. *Introduction to Mathematical Logic.* Princeton, N.J.: Princeton University Press, 1956.

63. Clarke, Arthur C. *The City and the Stars.* New York: Harcourt, Brace, 1956.

64. Coffin, Stewart T. *The Puzzling World of Polyhedral Dissections.* New York: Oxford University Press, 1990.

65. Collins, A. Frederick. *Fun with Figures.* New York: D. Appleton & Co., 1928.

66. Conway, J. H. "Mrs. Perkins's Quilt." *Proceedings of the Cambridge Philosophical Society,* 60: 363–368, 1964.

67. Conway, John Horton. *On Numbers and Games.* New York: Academic Press, 1976.

68. Conway, John Horton, and Richard K. Guy. *The Book of Numbers.* New York: Springer-Verlag, 1996.

69. Conway, J. H., and S. P. Norton. "Monstrous Moonshine." *Bulletin London Mathematical Society,* 11: 308–339 (1979).

70. Conway, J. H., and N. J. A. Sloane. "The Monster Group and its 196884-Dimensional Space" and "A Monster Lie Algebra?" Chs. 29–30 in *Sphere Packings Lattices, and Groups,* 2nd ed. New York: Springer-Verlag, pp. 554–571, 1993.

71. Cooper, Necia Grant (ed.). *From Cardinals to Chaos: Reflections on the Life and Legacy of Stanislaw Ulam.* Cambridge: Cambridge University Press, 1989.

72. Coxeter, H. S. M. *Non-Euclidean Geometry.* Toronto: The University of Toronto Press, 1942.

73. Coxeter, H. S. M., M. S. Longuet-Higgins, and J. C. P. Miller. "Uniform Polyhedra." *Philosophical Transactions of the Royal Society of London,* Series, *A* 246: 401–450, 1954.

74. Coxeter, H. S. M. *Introduction to Geometry,* 2nd ed. New York: John Wiley & Sons, 1969.

75. Coxeter, H. S. M. *Regular Polytopes,* 3rd ed. New York: Dover, 1973.

76. Coxeter, H. S. M., P. DuVal, H. T. Flather, and J. F. Petrie. *The Fifty-Nine Icosahedra.* Toronto: University of

Toronto Press, 1938. Reprint, New York: Springer-Verlag, 1982.

77. Coxeter, H. S. M., M. Emmer, R. Penrose, and M. L. Teuber, (eds). *M. C. Escher: Art and Science.* New York: North-Holland, 1986.

78. Cromwell, Peter R. *Polyhedra.* Cambridge: Cambridge University Press, 1997.

79. Császár, Ákos. "A polyhedron without diagonals." *Acta Univ Szegendiensis, Acta Scientia Math,* 13: 140–142, 1949.

80. Cundy, H. Martyn, and A. P. Rollett. *Mathematical Models,* 3rd ed. Oxford: Clarendon Press, 1961; Diss, England: Tarquin Publications, 1981.

81. Dewdney, A. K. *The Planiverse: Computer Contact with a Two-dimensional World.* New York: Simon & Schuster, 1984.

82. Dickson, Leonard E. *History of the Theory of Numbers.* 3 vol. London: Chelsea Pub. Co., 1919–23. Reprint, Providence, R.I.: American Mathematical Society, 1999.

83. Domoryad, A. P. *Mathematical Games and Pastimes.* Elmsford, N.Y.: Pergamon Press, 1963.

84. Dötzel, Gunter. "A Function to End All Functions." *Algorithm: Recreational Programming 2.4,* 16–17 (1991).

85. Doxiadis, Apostolos K. *Uncle Petros and Goldbach's Conjecture.* New York: Bloomsbury, 2000.

86. Dresner, Simon (ed.). *Science World Book of Brain Teasers.* New York: Scholastic Book Services, 1962.

87. Dudeney, H. E. *The Canterbury Puzzles.* London: Nelson, 1907. Reprinted Mineola, N.Y.: Dover, 1958.

88. Dudeney, H. E. *Amusements in Mathematics.* New York: Dover, 1917. Reprinted Mineola, N.Y.: Dover, 1958.

89. Dudeney, Henry E. *The World's Best Word Puzzles.* London: Daily News, 1925.

90. Dudeney, H. E. *536 Puzzles and Curious Problems,* ed. by Martin Gardner. New York: Charles Scribner, 1967.

91. Edgeworth, R., Dalton, B. J., and Parnell, T. "The Pitch Drop Experiment." *European Journal of Physics,* 5: 198–200 (1984).

92. Eiss, Harry Edwin. *Dictionary of Mathematical Games, Puzzles, and Amusements.* Westport, Conn.: Greenwood Press, 1988.

93. Elffers, J. *Tangram—The Ancient Chinese Puzzle.* New York: Harry N. Abrams, 1979.

94. Escher, M. C. *The Graphic Work of M. C. Escher,* New York: Ballantine, 1971.

95. Ewing, John, and Kosniowski, Czes. *Puzzle It Out: Cubes, Groups and Puzzles.* New York: Cambridge University Press, 1982.

96. Fadiman, Clifton. *Fantasia Mathematica: Being a set of stories, together with a group of oddments and diversions, all drawn from the universe of mathematics.* New York: Simon and Schuster, 1958.

97. Fadiman, Clifton. *The Mathematical Magpie: Being more stories, mainly transcendental, plus subsets of essays, rhymes, music, anecdotes, epigrams and other prime oddments and diversions, rational or irrational, all derived from the infinite domain of mathematics.* New York: Simon and Schuster, 1962.

Federico, P. J. "The Melancholy Octahedron." *Mathematics Magazine,* 45: 30–36, 1972.

98. Filipiak, Anthony S. *100 Puzzles: How to Make and Solve Them.* New York: A. S. Barnes, 1942.

99. Filipiak, Anthony S. *Mathematical Puzzles and Other Brain Twisters.* New York: A. S. Barnes, 1964.

100. Fink, Thomas M., and Yong Mao Mao. "Designing Tie Knots Using Random Walks." *Nature,* 398: 31 (1999).

101. Fisher, Adrian. Arian Fisher Maze Design Web site: http://www.mazemaker.com.

102. Fisher, John (ed.). *The Magic of Lewis Carroll.* New York: Simon and Schuster, 1973.

103. Foster, James E. *Mathematics as Diversion.* Astoria, IL: Fulton County Press, 1978.

104. Fraser, James. "A new visual illusion of direction." *British Journal of Psychology,* 2: 307–337 (1908).

105. Frederickson, Greg N. *Dissections: Plane and Fancy.* Cambridge: Cambridge University Press, 1997.

106. Frey, Alexander H., Jr., and David Singmaster. *Handbook of Cubik Math.* Hillside, N.J.: Enslow, 1982.

107. Gamow, G., and M. Stern. *Puzzle-Math.* New York: Viking, 1958.

108. Gardiner, M. *Mathematical Puzzling.* New York: Oxford University Press, 1987.

109. Gardner, Martin. *Mathematics, Magic and Mystery.* Mineola, N.Y.: Dover, 1956.

110. Gardner, M. *Mathematical Puzzles and Diversions.* New York: Simon and Schuster, 1959.

111. Gardner, Martin, ed. *The Mathematical Puzzles of Sam Loyd,* 2 vols. Mineola, N.Y.: Dover, 1959.

112. Gardner, M. *The Second Scientific American Book of Mathematical Puzzles & Diversions.* New York: Simon and Schuster, 1961.

113. Gardner, Martin. "The Remarkable Lore of the Prime Numbers," *Scientific American,* March 1964.

114. Gardner, M. *New Mathematical Diversions from Scientific American.* New York: Simon and Schuster, 1966.

115. Gardner, M. "Mathematical Games." *Scientific American,* 237, 18–28, Nov. 1977.

116. Gardner, Martin. *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleights-of-Mind from Scientific American.* New York: Vintage, 1978.

117. Gardner, Martin. *The Ambidextrous Universe.* New York: Charles Scribner's Sons, 1978.

118. Gardner, Martin. *Mathematical Circus: More Games, Puzzles, Paradoxes, and Other Mathematical*

Entertainments from Scientific American. New York: Alfred A. Knopf, 1979.

119. Gardner, Martin. *Science Fiction Puzzle Tales.* New York: Clarkson N. Potter, 1981.

120. Gardner, Martin. *Wheels, Life, and other Mathematical Amusements.* New York: W. H. Freeman, 1983.

121. Gardner, Martin. *Martin Gardner's Sixth Book of Mathematical Diversions from Scientific American.* Chicago: University of Chicago Press, 1983.

122. Gardner, Martin. *The Magic Numbers of Dr. Matrix.* Buffalo, N.Y.: Prometheus Books, 1985.

123. Gardner, M. *Knotted Doughnuts and Other Mathematical Entertainments.* New York: W. H. Freeman, 1986.

124. Gardner, Martin. *Riddles of the Sphinx And Other Mathematical Puzzle Tales.* Washington, DC: Mathematical Association of America, 1987.

125. Gardner, Martin. *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, 1987.

126. Gardner, Martin. *Hexaflexagons and Other Mathematical Diversions.* Chicago: University of Chicago Press, 1988.

127. Gardner, Martin. *Penrose Tiles and Trapdoor Ciphers . . . and the Return of Dr. Matrix.* New York: W. H. Freeman, 1989.

128. Gardner, Martin. *Mathematical Carnival.* Washington, D.C.: Mathematical Association of America, 1989.

129. Gardner, Martin. *Mathematical Magic Show.* Washington, D.C.: Mathematical Association of America, 1990.

130. Gardner, M. *Fractal Music, Hypercards, and More Mathematical Recreations from Scientific American Magazine.* New York: W. H. Freeman, 1992.

131. Gardner, Martin (introduction and notes). *The Annotated Alice: Alice's Adventures in Wonderland and Through the Looking Glass.* New York: Random House, 1998.

132. Gasser, Ralph. "Solving Nine Men's Morris." *Computational Intelligence,* 12: 24–41 (1996).

133. Gerver, J. L. "On Moving a Sofa Around a Corner." *Geometriae Dedicata,* 42: 267–283 (1992).

134. Gillam, B. "Geometrical Illusions." *Scientific American,* 242: 102–111 (Jan. 1980).

135. Gleick, James. *Chaos: Making a New Science.* New York: Penguin, 1988.

136. Golomb, S. W. "Checker Boards and Polyominoes." *American Mathematical Monthly,* 61: 675–682 (1954).

137. Golomb, Solomon W. *Polyominoes: Puzzles, Patterns, Problems, and Packings.* New York: Charles Scribner's, 1965. Reprint, Princeton, N.J.: Princeton University Press, 1995.

138. Gomme, Alice Bertha. *Traditional Games of England, Scotland and Ireland.* 2 vols. London: Thames and Hudson, 1894–1898.

139. Gotze, Heinz. *Castel del Monte, Geometric Marvel of the Middle Ages.* New York: Prestel, 1998.

140. Greenberg, Marvin J. *Euclidean and non-Euclidean Geometries: Development and History.* San Francisco: W. H. Freeman, 1974.

141. Greenblatt, M. H. *Mathematical Entertainments.* London: George Allen and Unwin, 1968.

142. Gregory, R. L. "Analogue Transactions with Adelbert Ames." *Perception,* 16: 277–282 (1987).

143. Grunbaum, B., and G. C. Shephard. *Tilings and Patterns.* New York: W. H. Freeman, 1987.

144. Guy, Richard K. *Unsolved Problems in Number Theory,* 2nd ed. New York: Springer-Verlag, 1994.

145. Guy, Richard K., and Robert E. Woodrow (eds). *The Lighter Side of Mathematics: Proceedings of the Eugène Strens Memorial Conference on Recreational Mathematics and its History.* Washington, D.C.: Mathematical Association of America, 1994.

146. Haeckel, Ernst. Kunstformen der Natur (Art Forms in Nature). Leipzig, Wien: Bibliographisches Institut 1899–1904. Reprinted Mineola, N.Y.: Dover, 1974.

147. Haldeman-Julius, E. *Problems, Puzzles and Brain-Teasers.* Girard, KS: Appeal to Reason, 1937.

148. Hall, Trevor H. *Old Conjuring Books.* London: Duckworth, 1972.

149. Hankins, Thomas L. *Sir William Rowan Hamilton.* Baltimore: Johns Hopkins University Press, 1980.

150. Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work.* London: Cambridge University Press, 1940.

151. Hardy, G. H. *A Mathematician's Apology.* London: Cambridge University Press, 1940.

152. Hargrave, Catherine Perry. *A History of Playing Cards and a Bibliography of Cards and Gaming.* New York: Houghton Mifflin, 1930. Reprint, New York: Dover, 1966.

153. Harmer, G. P., and D. Abbott. "Losing Strategies Can Win by Parrondo's Paradox." *Nature,* 402: 864 (1999).

154. Hartley, Miles C. *Patterns of Polyhedra.* Ann Arbor, MI: Edwards Brothers, 1957.

155. Heafford, Philip. *The Math Entertainer.* New York: Vintage Press, 1983.

156. Heinlein, Robert. "And He Built a Crooked House," in Isaac Asimov (ed.) *Where Do We Go From Here?* New York: Doubleday, 1971.

157. Heinzelin, J., de, "Ishango." *Scientific American,* 206(6): 105–116 (Jun. 1962).

158. Held, Richard (ed.). *Image, Object and Illusion.* San Francisco: W. H. Freeman, 1974.

159. Henderson, Linda Dalrymple. *The Fourth Dimension and Non-Euclidean Geometry in Modern Art.* Princeton, N.J.: Princeton University Press, 1983.

160. Hendricks, John R. “Magic Tesseracts and N-Dimensional Magic Hypercubes.” *Journal of Recreational Mathematics.* 6(3) (Summer 1973).

161. Hendricks, John R. “Black and White Vertices of a Hypercube.” *Journal of Recreational Mathematics,* 11(4), 1978–1979.

162. Hendricks, John R. “Magic Hypercubes.” Winnipeg, Canada: self-published pamphlet, 1988.

163. Hendricks, John R. “Images from the Fourth Dimension.” Winnipeg, Canada: self-published pamphlet, 1993.

164. Hill, T. P. “The First Digit Phenomenon.” *American Scientist,* 86: 358–363 (1998).

165. Hilton, Peter, and Jean Pedersen. *Build Your Own Polyhedra.* New York: Addison Wesley, 1988.

166. Hinton, Charles Howard. *A New Era of Thought.* London: Swan Sonnenschein, 1888.

167. Hinton, Charles Howard. *The Fourth Dimension.* London: Allen & Unwin, 1904.

168. Hodges, Andrew, and Douglas Hofstadter. *Alan Turing: The Enigma.* New York: Walker & Co., 2000.

169. Hoffman, Paul. *Archimedes’ Revenge: The Joys and Perils of Mathematics.* New York: W. W. Norton, 1988.

170. Hoffman, Paul. *The Man Who Loved Only Numbers: The Story of Paul Erdos and the Search for Mathematical Truth.* New York: Hyperion, 1998.

171. Hoffmann, Professor (Angelo Louis). *Puzzles Old and New.* London: Frederick Warne, 1893. Reprint, L. Edward Hordern, 1993.

172. Hofstadter, Douglas R. *Gödel. Escher, Bach: An Eternal Golden Braid.* New York: Basic Books, 1999.

173. Hogben, Lancelot. *Mathematics for the Million.* New York: W. W. Norton, 1933.

174. Holden, Alan. *Shapes, Spaces and Symmetry.* New York: Columbia University Press, 1971. Reprint, Mineola, N.Y.: Dover, 1991.

175. Hordern, Edward. *Sliding Piece Puzzles.* New York: Oxford University Press, 1986.

176. Hovis, R. Corby, and Helge Kragh. “P. A. M. Dirac and the Beauty of Physics.” *Scientific American,* 268(5): 104–109, May 1993.

177. Ingham, A. E. “On Two Conjectures in the Theory of Numbers.” *American Journal of Mathematics,* 64: 313–319 (1942).

178. Ittelson, W. H., and F. P. Kilpatrick. “Experiments in Perception.” *Scientific American,* 185(2), 50–55, August 1951.

179. Jackson, John. *Rational Amusement for Winter Evenings.* London: Longmans & Co., 1821.

180. Jayne, Caroline Furness. *String Figures and How to Make Them: A Study of Cat’s-Cradle in Many Lands.* New York: Dover, 1975.

181. Jones, John Winter. *Riddles, Charades, and Conundrums.* London, 1821.

182. Jones, Samuel Isaac. *Mathematical Nuts for Lovers of Mathematics.* Nashville, 1932.

183. Kadesch, Robert R. *Math Menagerie.* New York: Harper & Row, 1970.

184. Kanigal, Robert. *The Man Who Knew Infinity: A Life of the Genius Ramanujan.* New York: Scribner, 1991.

185. Kapur, J. N. *The Fascinating World of Mathematics.* 3 vols. New Delhi, India: Mathematical Sciences Trust Society, 1989.

186. Kasner, Edward, and James R. Newman, *Mathematics and the Imagination.* New York: Simon and Schuster, 1940.

187. King, T. *The Best 100 Puzzles Solved and Answered.* Slough, England: W. Foulsham & Co., 1927.

188. Kirkman, T. P. “On a Problem in Combinatorics.” *Cambridge and Dublin Math. J.,* 2: 191–204 (1847).

189. Klarner, David A., ed. *The Mathematical Gardner.* Boston: Prindle, Weber and Schmidt, 1981.

190. Klarner, David A., ed. *Mathematical Recreations: A Collection in Honor of Martin Gardner.* New York: Dover, 1998.

191. Knuth, Donald. *Surreal Numbers: How Two Ex-Students Turned On to Pure Mathematics and Found Total Happiness.* Reading, Mass.: Addison-Wesley, 1974.

192. Körner, T. W. *The Pleasures of Counting.* Cambridge: Cambridge University Press, 1997.

193. Kraitchik, Maurice. *Mathematical Recreations.* New York: Norton, 1942.

194. Krasnikov, S. V. “Causality Violations and Paradoxes.” *Physical Review D,* 55(6): 3427–30 (1997).

195. Langley, E. M. “Problem 644.” *Mathematical Gazette,* 11: 173, 1922.

196. Langman, H. *Play Mathematics.* New York: Hafner, 1962.

197. Lausmann, Raymond F. *Fun With Figures.* New York: McGraw-Hill, 1965.

198. Lavine, S. *Understanding the Infinite.* Cambridge, Mass.: Harvard University Press, 1994.

199. Lehman, R. S. “On Liouville’s Function.” *Mathematics of Computation,* 14: 311–320, 1960.

200. Licks, H. E. *Recreations in Mathematics.* New York: D. Van Nostrand, 1921.

201. Lindgren, Harry. *Geometric Dissections.* New York: Van Nostrand Reinhold, 1964.

202. Lindgren, Harry. *Recreational Problems in Geometrical Dissections and How to Solve Them.* New York: Dover Publications, 1972.

203. Lines, Malcolm. *Think of a Number.* New York: Adam Hilger, 1990.

204. Littlewood, J. E. *A Mathematician’s Miscellany.* London: Methuen and Co., 1953.

205. Livio, Mario. *The Golden Ratio: The Story of Phi, the World’s Most Astonishing Number.* New York: Broadway Books, 2002.

206. Locher J. L., ed. *M. C. Esher: His Life and Complete Graphic Work.* New York: Abradale Press, 1982.

207. Loyd, Sam. *Sam Loyd's Cyclopedia of 5000 Puzzles, Tricks, and Conundrums.* 1914. Reprint, New York: Pinnacle, 1976.

208. Loyd, Sam. *Sam Loyd's Picture Puzzles with Answers.* Brooklyn: S. Loyd, 1924. Reprint, Ontario, Canada: Algrove, 2000.

209. Lucas, Edouard. *Récréations Mathématiques,* 4 vols. Paris: Gauthier-Villars et fils, 1881–1894. Reprinted by Paris: A. Blanchard, 1960–1974.

210. MacMahon, Alexander: *New Mathematical Pastimes.* Cambridge: Cambridge University Press, 1921.

211. Madachy, Joseph S. *Mathematics on Vacation.* New York: Charles Scribner's, 1966. Republished as *Madachy's Mathematical Recreations.* New York: Dover, 1979.

212. Manning, H. P. *Geometry of Four Dimensions.* New York: Dover, 1956.

213. Manning, H. P. *The Fourth Dimension Simply Explained.* New York: Dover, 1960.

214. Maor, E. *To Infinity and Beyond: A Cultural History of the Infinite.* Boston, Mass.: Birkhäuser, 1987.

215. Maor, E. *e: The Story of a Number.* Princeton, N.J.: Princeton University Press, 1994.

216. Martin, G. "Absolutely abnormal numbers." *American Mathematical Monthly,* 108: 746–754, 2001.

217. Matijasevic, Yu. V. "Solution of the Tenth Problem of Hilbert." *Matematikai Lapok,* 21: 83–87 (1970).

218. Matijasevic, Yu. V. *Hilbert's Tenth Problem.* Cambridge, Mass.: MIT Press, 1993.

219. Matthews, W. H. *Mazes & Labyrinths: A General Account of Their History and Development.* London: Longmans, Green and Co., 1922. Reprint, New York: Dover, 1970.

220. Maxwell, E. A. *Fallacies in Mathematics.* New York: Cambridge University Press, 1959.

221. Meeus, J., and P. J. Torbijn. *Polycubes.* Paris: Cedic, 1977.

222. Melzak, Z. A. *Companion to Concrete Mathematics.* 2 vols. New York: John Wiley, 1973, 1976.

223. Mercer, J. W., et al. "Solutions to Langley's Adventitious Angles Problem." *Mathematical Gazette,* 11: 321–323, 1923.

224. Mosteller, F. *Fifty Challenging Problems in Probability with Solutions,* Reading, Mass.: Addison-Wesley, 1965, #47; "Mathematical Plums," edited by Ross Honsberger, pp. 104–110.

225. Mott-Smith, Geoffrey. *Mathematical Puzzles for Beginners and Enthusiasts,* 2nd rev. ed. New York: Dover, 1954.

226. Moyer, Ann E. *The Philosophers' Game: Rithmomachia in Medieval and Renaissance Europe.* Ann Arbor, Mich.: University of Michigan Press, 2001.

227. Murray, H. J. R. *History of Chess.* Oxford: Clarendon Press, 1913.

228. Murray, H. J. R. *A History of Board-Games Other than Chess.* Oxford: Clarendon Press, 1952.

229. Nasar, Sylvia. *A Beautiful Mind: A Biography of John Forbes Nash, Jr.* New York: Simon & Schuster, 1998.

230. Neumann, J. von, and Morgenstern, O. *Theory of Games and Economic Behavior.* New York: Wiley, 1964.

231. Neville, E. H. *The Fourth Dimension.* Cambridge: Cambridge University Press, 1921.

232. Newcomb, S. "Note on the Frequency of the Use of Digits in Natural Numbers." *American Journal of Mathematics,* 4: 39–40, 1881.

233. Newing, Angela. "The Life and Work of H. E. Dudeney." *Mathematical Spectrum,* 21: 37–44 (1988–1989).

234. Northrop, Eugene P. *Riddles in Mathematics. A Book of Paradoxes.* Princeton, N.J.: Van Nostrand, 1944.

235. O'Beirne, T. H. *Puzzles and Paradoxes: Fascinating Excursions in Recreational Mathematics.* Mineola, N.Y.: Dover, 1984.

236. Ogilvy, C. Stanley. *Tomorrow's Math: Unsolved Problems for the Amateur,* 2nd ed. New York: Oxford University Press, 1972.

237. Olivastro, Dominic. *Ancient Puzzles.* New York: Bantam, 1993.

238. Panofsky, Erwin. *The Life and Art of Albrecht Durer,* Princeton, N.J.: Princeton University Press, 1955.

239. Pardon, G. F. *Parlour Pastimes: A Repertoire of Acting Charades, Fire-Side Games, Enigmas, Riddles, Etc.* London, 1868.

240. Pennick, Nigel. *Mazes and Labyrinths.* London: Robert Hale, 1990.

241. Penrose, L. S., and R. Penrose. "Impossible Objects: A Special Type of Illusion," *British Journal of Psychology,* 49: 31, 1958.

242. Penrose Roger. *The Emperor's New Mind.* New York: Oxford University Press, 1989.

243. Péter, Rózsa. *Playing with Infinity: Mathematical Explorations and Excursions.* Mineola, N.Y.: Dover, 1976.

244. Peterson, Ivars. *The Mathematical Tourist: Snapshots of Modern Mathematics.* New York: W. H. Freeman, 1988.

245. Peterson, Ivars. *Islands of Truth: A Mathematical Mystery Cruise.* New York: W. H. Freeman, 1990.

246. Peterson, Ivars. *The Jungles of Randomness: A Mathematical Safari.* New York: Wiley, 1997.

247. Peterson, Ivars. *The Mathematical Tourist: New and Updated Snapshots of Modern Mathematics.* New York: W. H. Freeman, 1998.

248. Peterson, Ivars. *Mathematical Treks: From Surreal Numbers to Magic Circles.* Washington, D.C.: Mathematical Association of America, 2001.

249. Phillips, Hubert. *Journey to Nowhere: A Discursive Autobiography.* London: Macgibbon & Kee, 1960.

250. Phillips, Hubert. *My Best Puzzles in Mathematics.* Mineola, N.Y.: Dover, 1961.

251. Phillips, Hubert. *My Best Puzzles in Logic and Reasoning.* Mineola, N.Y.: Dover, 1961.

252. Picciotto, Henri. *Pentomino Activities.* Mountain View, Calif.: Creative Publications, 1986.

253. Pickover, Clifford. *Keys to Infinity.* New York: W. H. Freeman, 1995.

254. Pickover, Clifford. *Surfing Through Hyperspace.* New York: Oxford University Press, 1999.

255. Pickover, Clifford A. *Wonders of Numbers: Adventures in Mathematics, Mind, and Meaning.* Oxford: Oxford University Press, 2001.

256. Pickover, C. A. *The Zen of Magic Squares, Circles, and Stars: An Exhibition of Surprising Structures across Dimensions.* Princeton, N.J.: Princeton University Press, 2002.

257. Poole, D. G. "The Towers and Triangles of Professor Claus (or, Pascal Knows Hanoi)." *Mathematics Magazine,* 67: 323–344 (1994).

258. Poundstone, William. *The Recursive Universe.* New York: William Morrow, 1984.

259. Poundstone, William. *Prisoner's Dilemma: John Von Neumann, Game Theory and the Puzzle of the Bomb.* New York: Doubleday, 1992.

260. Quine, W. V. O. "On a So-Called Paradox." *Mind,* 62: 65–67 (1953).

261. Rabinowitz, Stanley. *Index to Mathematical Problems 1980–1984.* Westford, Mass.: Mathpro Press, 1992.

262. Rademacher, Hans, and Otto Toeplitz. *The Enjoyment of Mathematics: Selections from Mathematics for the Amateur.* Princeton, N.J.: Princeton University Press, 1957.

263. Raimi, R. A. "The Peculiar Distribution of First Digits." *Scientific American,* 221: 109–119 (Dec. 1969).

264. Ransom, William R. *One Hundred Mathematical Curiosities.* Portland, Maine: Weston Walch, 1955.

265. Read, Ronald C. *Tangrams: 330 Puzzles.* Mineola, N.Y.: Dover, 1965.

266. Ribenboim, P. *Catalan's Conjecture.* Boston, Mass.: Academic Press, 1994.

267. Robbin, Tony. *Fourfield: Computers, Art, and the Fourth Dimension.* Boston: Little, Brown, 1992.

268. Robertson, J., and Webb, W. *Cake-cutting Algorithms.* Natick, Mass.: A. K. Peters, 1998.

269. Rossi, Hugo. "Mathematics Is an Edifice, Not a Toolbox." *Notices of the AMS,* 43(10) (Oct. 1996).

270. Rubik, Erno, et al. *Rubik's Cubic Compendium.* New York: Oxford University Press, 1987.

271. Rucker, Rudolf v.B. *Geometry, Relativity, and the Fourth Dimension.* New York: Dover, 1977.

272. Rucker, R. (ed). *Speculations on the Fourth Dimension: Selected Writings of C. H. Hinton.* New York: Dover, 1980.

273. Rucker, Rudy. *The Fourth Dimension.* Boston: Houghton Mifflin, 1984.

274. Rucker, Rudy. *Mind Tools.* Boston: Houghton Mifflin, 1987.

275. Rucker, Rudy. *Infinity and the Mind.* Princeton, N.J.: Princeton University Press, 1995.

276. Russell, Bertrand. *Bertrand Russell Autobiography.* New York: Routledge, 1992.

277. Saaty, T. L., and P. C. Kainen. *The Four-Color Problem: Assaults and Conquest.* New York: Dover, 1986.

278. Sallows, L. C. F. "Alphamagic squares." *Abacus* 4 (No. 1): 28–45, 1986.

279. Sallows, L. C. F. "Alphamagic squares, part II". *Abacus* 4 (No. 2): 20–29, 43, 1987.

280. Sallows, L. C. F. "Alphamagic squares." In *The Lighter Side of Mathematics: Proceedings of the Eugène Strens Memorial Conference on Recreational Mathematics and Its History,* R. K. Guy and R. E. Woodrow, eds. Washington, D.C.: Mathematical Association of America, 1994.

281. Sanford, Vera. *A Short History of Mathematics.* Boston, Mass.: Houghton Mifflin, 1930.

282. Scarne, John. *Scarne's New Complete Guide to Gambling.* New York: Simon and Schuster, 1986.

283. Schaaf, William Leonard. *A Bibliography of Recreational Mathematics.* 4 vols. Washington, D.C.: National Council of Teachers of Mathematics, 1955–78.

284. Schattschneider, Doris. *M. C. Escher: Visions of Symmetry.* New York: W. H. Freeman, 1990.

285. Scheick, William J. "The Fourth Dimension in Wells's Novels of the 1920s." *Criticism,* 20: 167–190, 1978.

286. Schrauf, M., B. Lingelbach, E. Lingelbach, and E. R. Wist. "The Hermann Grid and the Scintillation Effect." *Perception,* 24, suppl. A: 88–89, 1995.

287. Schreiber, P. "A New Hypothesis on Durer's Enigmatic Polyhedron in His Copper Engraving 'Melencholia I'," *Historia Mathematica,* 26: 369–377, 1999.

288. Schubert, Hermann. *Mathematical Essays and Recreations.* Thomas J. McCormack, trans. Chicago: Open Court, 1899.

289. Schubert, Hermann. *Mathematische Mussestunden.* Walter de Gruyter, 1940. First published 1900? (German).

290. Schubert, Hermann. *Mathematische Mussestunden II.* Berlin: G. J. Goshen'sche, 1909.

291. Schuh, Fred. *The Master Book of Mathematical Recreations.* Mineola, N.Y.: Dover, 1968.

292. Searle, John R. "Minds, Brains and Programs" in *The Brain and Behavioral Sciences,* vol. 3. Cambridge: Cambridge University Press, 1980.

293. Senechal, M. *Quasicrystals and Geometry.* New York: Cambridge University Press, 1995.

294. Shepherd, Walter. *For Amazement Only.* New York: Penguin, 1942. Reprinted as *Mazes and Labyrinths: A Book of Puzzles.* New York: Dover, 1961.

295. Simon, William. *Mathematical Magic.* New York: Charles Scribner's, 1964.

296. Singh, Simon. *Fermat's Enigma: The Epic Quest to Solve the World's Greatest Mathematical Problem.* New York: Walker & Co., 1997.

297. Singmaster, David. *Notes on Rubik's 'Magic Cube.* Hillside, N.J.: Enslow, 1981.

298. Singmaster, David. *Sources in Recreational Mathematics: An Annotated Bibliography,* 7th prelim, ed. Unpublished.

299. Skewes, S. "On the difference π(x) – li(x)." *Journal of the London Mathematical Society,* 8: 277–283 (1933).

300. Slocum, Jerry. *Compendium of Mechanical Puzzles.* 3rd ed. Slocum and Haubrich, 1977.

301. Slocum, Jerry. *The Tangram Book: The Story of the Chinese Puzzle with Over 2000 Puzzles to Solve.* New York: Sterling, 2003.

302. Smith, David Eugene. *Number Stories of Long Ago.* Washington, D.C.: National Council of Teachers of Mathematics, 1919.

303. Smith, David Eugene. *A Source Book in Mathematics.* 2 vols. New York: McGraw-Hill, 1929.

304. Smullyan, Raymond M. *What Is the Name of This Book? The Riddle of Dracula and Other Logical Puzzles.* Englewood Cliffs, N.J.: Prentice Hall, 1978.

305. Smullyan, Raymond M. *The Chess Mysteries of Sherlock Holmes.* New York: Alfred A. Knopf, 1979.

306. Smullyan, Raymond. *This Book Needs No Title: A Budget of Living Paradoxes.* Englewood Cliffs, N.J.: Prentice-Hall, 1980.

307. Smullyan, Raymond M. *The Lady or the Tiger? And Other Logic Puzzles.* New York: Alfred A. Knopf, 1982.

308. Smullyan, Raymond M. *Alice in Puzzleland.* New York: William Morrow, 1982.

309. Smullyan, Raymond M. *To Mock a Mockingbird.* New York: Alfred A. Knopf, 1985.

310. Smullyan, Raymond M. *Forever Undecided: A Puzzle Guide to Godel.* New York: Alfred A. Knopf, 1987.

311. Smullyan, Raymond. *The Tao is Silent.* New York: HarperCollins, 1992.

312. Soddy, F. "The Kiss Precise." *Nature,* 137: 1021, 1936.

313. Sprague, Roland. *Recreation in Mathematics.* Mineola, N.Y.: Dover, 1963.

314. Stein, M. L., S. M. Ulam, and M. B. Wells. "A Visual Display of Some Properties of the Distribution of Primes." *American Mathematical Monthly,* 71: 516–520, 1964.

315. Steinhaus, H. "Sur la division pragmatique." *Ekonometrika* (*Supp.*), 17: 315–319, 1949.

316. Steinhaus, Hugo. *Mathematical Snapshots.* 3rd rev. ed. New York: Dover, 1983.

317. Stewart, Ian. *Does God Play Dice: The New Mathematics of Chaos.* Cambridge, Mass.: Basil Blackwell, 1989.

318. Stewart, Ian. *Game, Set, and Math: Enigmas and Conundrums.* Cambridge, Mass.: Basil Blackwell, 1989.

319. Stewart, Ian. "Tales of a Neglected Number." *Scientific American,* 274: 102–103 (Jun. 1996).

320. Stewart, Ian. "Unilluminable Rooms." *Scientific American,* 275(2), 100–103 (Aug. 1996).

321. Stewart, Ian and Cohen, Jack Cohen. *Figments of Reality.* Cambridge: Cambridge University Press, 1997.

322. Stewart, Ian. "Cone with a Twist." *Scientific American,* 281: 116–117 (Oct. 1999).

323. Stewart, Ian. *Flatterland: Like Flatland Only More So.* New York: Perseus, 2001.

324. Stromberg, K. "The Banach-Tarski Paradox." *American Mathematical Monthly,* 86: 3, 1979.

325. Stubbs, A. Duncan. *Miscellaneous Puzzles.* London: Frederick Warne, 1931.

326. Struik, Dirk Jan, ed. *A Source Book in Mathematics 1200–1800.* Princeton, N.J.: Princeton University Press, 1969.

327. Strutt, Joseph. *The Sports and Pastimes of the People of England.* London: J. White, 1801.

328. Stuwe, M. *Plateau's Problem and the Calculus of Variations.* Princeton, N.J.: Princeton University Press, 1989.

329. Swade, Doron. *The Difference Engine: Charles Babbage and the Quest to Build the First Computer.* New York: Viking, 2000.

330. te Riele, H. J. J. "On the Sign of the Difference π(*x*) – Li(*x*)." *Mathematics of Computation,* 48: 323–328 (1987).

331. Thom, R. *Structural Stability and Morphogenesis: An Outline of a General Theory of Models.* Reading, Mass.: Addison-Wesley, 1972.

332. Thompson, d'Arcy W. *On Growth and Form.* Cambridge: Cambridge University Press, 1917. (Reprint, New York: Dover, 1992.)

333. Thorpe, Edward O. *Beat the Dealer,* rev. ed. New York: Random House, 1966.

334. Tokarsky, George W. "Polygonal Rooms Not Illuminable from Every Point." *American Mathematical Monthly,* 102(10): 867–879, 1995.

335. Toole, Betty A. (ed.). *Ada, the Enchantress of Numbers: A Selection from the Letters of Lord Byron's Daughter and Her Description of the First Computer.* Mill Valley, Calif.: Strawberry Press, 1998.

336. Tormey, Alan and Tormey, Judith Farr. "Renaissance Intarsia: The Art of Geometry." *Scientific American,* 247: 136–143, July 1982.

337. Turing, A. M. "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings*

of the London Mathematical Society, Series 2, 42: 230–265, 1937.

338. Turney, P, "Unfolding the Tesseract." *Journal of Recreational Mathematics,* 17(1): 1–16, 1984–1985.

339. Underwood, Dudley. "The First Recreational Mathematics Book." *Journal of Recreational Mathematics,* 3, 164–169, 1970.

340. Van Cleve, James. "Right, Left, and the Fourth Dimension." *The Philosophical Review,* 96: 33–68 (1987).

341. Vardi, I. "Archimedes' Cattle Problem." *American Mathematical Monthly,* 105: 305–319 (1998).

342. Vardi, Ilan. *Computational Recreations in Mathematica.* Redwood City, Calif.: Addison-Wesley, 1991.

343. Wagon, S. *The Banach-Tarski Paradox.* New York: Cambridge University Press, 1993.

344. Waller, Mary D. *Chladni Figures: A Study in Symmetry.* London: G. Bell & Sons, 1961.

345. Watson, G. N. "The Problem of the Square Pyramid." *Messenger of Mathematics,* 48: 1–22 (1918–1919).

346. Weisstein, Eric W. *The CRC Concise Encyclopedia of Mathematics.* Boca Raton, Fla.: CRC Press, 1998.

347. Wells, David Graham. *Hidden Connections, Double Meanings: A Mathematical Exploration.* Cambridge: Cambridge University Press, 1988.

348. Wells, David. *The Penguin Dictionary of Curious and Interesting Geometry.* London: Penguin Books, 1991.

349. Wells, David. *The Penguin Book of Curious and Interesting Mathematics.* New York: Penguin, 1997.

350. Wells, Kenneth. *Wooden Puzzles and Games.* New York: Sterling Publications, 1983.

351. Weyl, Hermann. *Symmetry.* Princeton, N.J.: Princeton University Press, 1952.

352. White, Alain C. *Sam Loyd and His Chess Problems.* Leeds, England: Whitehead & Miller, 1913.

353. Wiles, A. "Modular Elliptic-Curves and Fermat's Last Theorem." *Annals of Mathematics,* 141: 443–551, 1995.

354. Williams, W. T., and Savage, G. H. *The First Penguin Problems Book.* New York: Penguin Books, 1940.

355. Wyatt, Edwin Mather. *Puzzles in Wood.* Milwaukee, Wis.: Bruce Publishing Company, 1928.

356. Yan, Li, and Du, Shiran. *Chinese Mathematics: A Concise History.* Oxford: Oxford University Press, 1988.

357. Zaslavsky, Claudia. *Africa Counts.* Brooklyn, N.Y.: Lawrence Hill Books, 1973.

Solutions to Puzzles

abracadabra

To solve this problem, redraw the diamond and replace every letter on the outside top half of the diamond with a 1 and every other point in the diamond with a dot. This gives

1
1 1
1 . 1
1 . . 1
1 . . . 1
1 1
. . . . .
. . . .
. . .
. .
.

Next, replace each of the dots, starting at the top, with the sum of the two numbers in the northwest and northeast positions, so that the pyramid now starts

1
1 1
1 2 1
1 3 3 1

After you have worked your way down to the very bottom, the bottom number in the diamond is the answer to Polya's problem: 252.

age puzzles and tricks

1. 18.
2. 16½ (Mary is 27½).

alphametic

1.
67432 (EARTH)
704 (AIR)
8046 (FIRE)
+ 97364 (WATER)
————————
173546 (NATURE)

A = 7, E = 6, F = 8, H = 2, I = 0, R = 4, T = 3, W = 9

2.
127503 (SATURN)
502351 (URANUS)
3947539 (NEPTUNE)
+ 46578 (PLUTO)
————————
4623971 (PLANETS)

A = 2, E = 9, L = 6, N = 3, O = 8, P = 4, R = 0, S = 1, T = 7, U = 5

3.
862903 (MARTIN)
+ 1627342 (GARDNER)
————————
2490245 (RETIRES)

A = 6, D = 7, E = 4, G = 1, I = 0, M = 8, N = 3, R = 2, S = 5, T = 9

anagram

1. Wolfgang Amadeus Mozart.
2. Thomas Alva Edison.
3. William Shakespeare.

Carroll, Lewis

1. There is as much water in the milk/water mixture as milk in the water/milk mixture.
2. There are 30 ways of painting the cube. If the restriction that each face be painted a different color is dropped, there are 2,226 ways of painting the cube.
3. FOUR → FOUL → FOOL → FOOT → FORT → FORE → FIRE → FIVE

Alternatively:

FOUR → POUR → POUT → ROUT → ROUE → ROVE → DOVE → DIVE → FIVE
(Thanks to my editor, Stephen Power, for this one.)

4. Carroll didn't have an answer in mind when he wrote the riddle, though he later came up with: "Because it can produce a few notes, though they are *very* flat; and it is nevar put wrong end in front!" (Note the variant spelling of "never.") Other authors have come up with: "Because Poe wrote on both"

(Sam Loyd); "Because it slopes with a flap" (Cyril Pearson); and "Because both have quills dipped in ink" (David Jodrey).

clock puzzles

1. Approximately 27 minutes and 42 seconds (exactly 360/13 minutes) after 6.
2. The stopped clock, because it will give the right time twice a day, whereas the other is only correct approximately every two years.
3. The positions of the hands shown in the illustration could only indicate that the clock stopped at 44 min. 51 1,143/1,427 sec. after eleven o'clock. The second hand would next be "exactly midway between the other two hands" at 45 min. 52 496/1,427 sec. after eleven o'clock. If we had been dealing with the points on the circle to which the three hands are directed, the answer would be 45 min. 22 106/1,427 sec. after eleven; but the question applied to the hands, and the second hand would not be between the others at that time, but outside them.

cone

The simple rule is that the cone must be cut at one-third of its height.

de Morgan, Augustus

43 (the only number that, when squared, gives a number between the years of de Morgan's birth and death).

domino

No! Suppose it *were* possible to totally cover the modified chessboard with nonoverlapping dominos. In any complete tiling, every domino must cover exactly one white square and one black square. Thus the modified board must have exactly the same number of black and white squares. But the two removed squares, from diagonally opposite corners of a chessboard, must be same color. Since there can't be the same number of white squares and black squares on the modified board it must be impossible to tile the modified board with nonoverlapping dominos.

hundred fowls problem

Using the two equations to eliminate C, gives the equation $7R + 4H = 100$. R must be less than 15 (since $7 \times 15 = 105$). Trial-and-error shows that values of R that allow whole number values of H are 4, 8, and 12, from which it follows that the three possible solutions to the problem are 4 roosters, 18 hens, and 78 chicks; 8 roosters, 11 hens, and 81 chicks; and 12 roosters, 4 hens, and 84 chicks.

kinship puzzles

1. Your son.
2. Sons of two men who married each other's mothers.
3. The party consisted of two girls and a boy, their father and mother, and their father's father and mother.

measuring and weighing puzzles

The innkeeper first filled the 5-pint and 3-pint measures, then turned the tap on the barrel and allowed the rest of its contents to run to waste. He closed the tap and emptied the 3-pint into the barrel; filled the 3-pint from the 5-pint; emptied the 3-pint into the barrel; transferred the 2 pints from the 5-pint to the 3-pint; filled the 5-pint from the barrel, leaving 1 pint now in the barrel; filled the 3-pint from the 5-pint; allowed the company to drink the contents of the 3-pint; filled the 3-pint from the 5-pint, leaving 1 pint now in the 5-pint; drank the contents of the 3-pint; and finally drew off 1 pint from the barrel into the 3-pint. He now had 1 pint of ale in each measure!

missing dollar problem

There is no missing dollar (of course!). Adding \$27 and \$2 (to get \$29) is a bogus operation. They paid \$27, \$2 went to the dishonest waiter, and \$25 went to the restaurant. You have to subtract \$27 minus \$2 to get \$25. There never was a \$29; it has nothing to do with anything.

Mrs. Perkins's quilt

The following diagram shows how the quilt should be made.

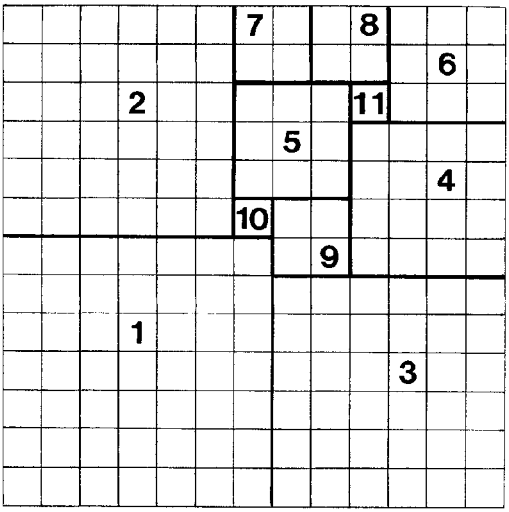

The solution of Mrs. Perkins's quilt.

In *Amusements in Mathematics,* Dudeney writes: "There is, I believe, only one solution to this puzzle. The fewest separate squares must be eleven. The portions must be of the sizes given, the three largest pieces must be arranged as shown, and the remaining group of eight squares may be 'reflected,' but cannot be differently arranged."

nine rooms paradox

One of the two customers who is initially placed in room A, whom we refer to as the "first" customer, is later transferred to room I and treated as if he were also the tenth customer!

postage stamp problems

Dudeney gives the solution to his "The Four Postage Stamps" problem as follows: "[T]he four stamps may be given in the shape 1, 2, 3, 4, in three ways; in the shape 1, 2, 5, 6, in six ways; in the shape 1, 2, 3, 5, or 1, 2, 3, 7, or 1, 5, 6, 7, or 3, 5, 6, 7, in twenty-eight ways; in shape 1, 2, 3, 6, or 2, 5, 6, 7, in fourteen ways; in shape 1, 2, 6, 7, or 2, 3, 5, 6, or 1, 5, 6, 10, or 2, 5, 6, 9, in fourteen ways. Thus there are sixty-five ways in all."

probability theory

1. There are four possibilities:

Oldest Child	Youngest Child
1. Girl	Girl
2. Girl	Boy
3. Boy	Girl
4. Boy	Boy

If your friend says "My oldest child is a girl," he has eliminated cases 3 and 4, and in the remaining cases both are girls ½ of the time. If your friend says "At least one of my children is a girl," he has eliminated case 4 only, and in the remaining cases both are girls ⅓ of the time.

2. There are six possible bullet configurations (B = bullet, E = empty):

B B B E E E → player 1 dies
E B B B E E → player 2 dies
E E B B B E → player 1 dies
E E E B B B → player 2 dies
B E E E B B → player 1 dies
B B E E E B → player 1 dies

One therefore has a ⅔ probability of winning (and a ⅓ probability of dying) by shooting second.

Pythagorean square puzzle

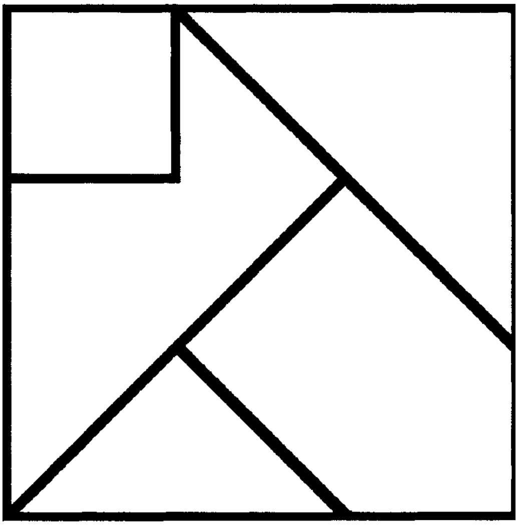

Pythagorean puzzle solution.

river-crossing problem

The missionaries (M) can avoid being eaten by the cannibals (C) if the crossings of the river and the returns are arranged as follows. Crossing #1: 1M + 1C; return #1: 1M; crossing #2: 2C; return #2: 1C; crossing #3: 2M; return #3: 1M + 1C; crossing #4: 2M; return #4: 1C; crossing #5: 2C; return #5: 1C; crossing #6: 2C.

word puzzles

(1) $T_1W_4E_1L_1V_4E_1$.
(2) Dreamt.
(3) Tremendous, horrendous, stupendous, and hazardous.
(4) I am.
(5) skiing
(6) "Are" by an adding an "a" becomes "area."
(7) underground.

Category Index

Algebra
abstract algebra
algebra
algebraic fallacies
algebraic geometry
array
associative
Bieberbach conjecture
binary operation
coefficient
commutative
cubic equation
determinant
diagonal matrix
discriminant
distributive
field
fundamental theorem of algebra
gradient
Hankel matrix
hundred fowls problem
idempotent
invariant
invariant theory
Jordan matrix
Lie algebra
linear algebra
linear system
matrix
monomial
non-Abelian
quadratic
quadric
quartic
quintic
resultant
ring
scalar
solution
tensor
Toeplitz matrix
trace
transpose
trinomial
unimodular matrix
vector
vector space
weak inequality
x
zero divisors

Analytical geometry
abscissa
algebraic curve
analytical geometry
Cartesian geometry. *See* analytical geometry
Cartesian coordinates
coordinate
coordinate geometry. *See* analytical geometry
octant
ordinate
origin
polar coordinates
quadrant
rectangular coordinates. *See* Cartesian coordinates
slope
x-axis
y-axis

See also: **Geometry, general terms and theorems** and **Graphs and graph theory**

Approximations and averages
Banker's rounding
arithmetic mean
average
ceiling
difference equation
extrapolate. *See* interpolate
floor function
geometric mean
interpolate
iterate
iteration
mean
median
mode
Newton's method
rounding numerical analysis
round-off error
significant digits
universal approximation

Arithmetic
arithmetic
binary
casting out nines
cryptarithm
denominator
digimetic
direct proportion
division
fly between trains problem
four fours problem
fundamental theorem of arithmetic
greatest common divisor
greatest lower bound
harmonic mean
least common multiple
multiple
negative base
number system
pandigital product
period, of a decimal expansion
place-value system
product
proportional
quotient
ratio
reciprocal
Russian multiplication
St. Ives problem
skeletal division
subtraction
wheat and chessboard problem

See also: **Number theory**

Biography
Abbott, Edwin Abbott (1838–1926)
Abel, Niels Henrik (1802–1829)
Abu'l Wafa (A.D. 940–998)
Agnesi, Maria Gaetana (1718–1799)
Ahrens, Wilhelm Ernst Martin Georg (1872–1927)
Alcuin (735–804)
al-Khowarizmi (c. 780–850)
Archimedes of Syracuse (c. 287–212 B.C.)
Apollonius of Perga (c. 255–170 B.C.)
Atiyah, Michael Francis (1929–)
Babbage, Charles (1791–1871)

Biography *(Continued)*
Bachet de Méziriac, Claude-Gaspar (1581–1638)
Ball, Walter William Rouse (1850–1925)
Banach, Stefan (1892–1945)
Barbaro, Daniele (1513–1570)
Barlow, Peter (1776–1862)
Bayes, Thomas (1702–1761)
Bell, Eric Temple (1883–1960)
Bernoulli family
Bessel, Friedrich Wilhelm (1784–1846)
Birkhoff, George David (1884–1944)
Bólyai, János (1802–1860)
Boole, George (1815–1864)
Boole (Stott), Alicia (1860–1940)
Borel, Emile (1871–1956)
Borges, Jorge Luis (1899–1986)
Bourbaki, Nicholas
Brahmagupta (A.D. 598–after 665)
Briggs, Henry (1556–1630)
Bronowski, Jacob (1908–1974)
Brouwer, Luitzen Egbertus Jan (1881–1966)
Byron, (Augusta) Ada (1815–1852)
Cage, John (1912–1992)
Caliban
Cantor, Georg Ferdinand Ludwig Philipp (1845–1918)
Cardano, Girolamo (1501–1570)
Carroll, Lewis (1832–1898)
Cauchy, Baron Augustin Louis, (1789–1857)
Cayley, Arthur
Ceva, Giovanni (1647–1734)
Chaitin, Gregory (1947–)
Chladni, Ernst Florens Friedrich (1756–1827)
Church, Alonzo (1903–1995)
Clifford, William Kingdon (1845–1879)
Coffin, Stewart T.
Conway, John Horton (1937–)
Coxeter, Harold Scott MacDonald
Cutler, William ("Bill")
d'Alembert, Jean Le Rond (1717–1783)
de l'Hopital, Guillaume Francis Antoine (1661–1704)
de Moivre, Abraham (1667–1754)
de Morgan, Augustus (1806–1871)
Dedekind, Julius Wilhelm Richard (1831–1916)
Dee, John (1527–1609)
Desargues, Girard (1591–1661)
Descartes, René (1596–1650)
Dewdney, Alexander Keewatin
Diocles (c. 240–c. 180 B.C.)
Diophantus of Alexandria (A.D. c. 200–c. 284)
Dirac, Paul Adrien Maurice (1902–1984)
Dirichlet, Peter Gustav Lejeune (1805–1859)
Dodgson, Charles Lutwidge. *See* Carroll, Lewis
Dudeney, Henry Ernest (1857–1930)
Dunsany, Lord Edward Plunkett (1878–1957)
Dürer, Albrecht (1471–1528)
Eratosthenes of Cyrene (c. 276–194 B.C.)
Erdös, Paul (1913–1996)
Escher, Maurits Cornelius (1898–1972)
Euclid of Alexandria (c. 330–270 B.C.)
Eudoxus of Cnidus (c. 408–c. 355 B.C.)
Euler, Leonhard (1707–1783)
Fadiman, Clifton (1904–1999)
Fechner, Gustav Theodor (1801–1887)
Federo, E. S. (1853–1919)
Fermat, Pierre de (1601–1665)
Fibonacci (c. 1175–1250)
Fisher, Adrian
Fourier, (Jean Baptiste) Joseph, Baron (1768–1830)
Fredholm, Erik Ivar (1866–1927)
Frege, Friedrich Ludwig Gottlob (1848–1925)
Frénicle de Bessy, Bernard (1602–1675)
Galois, Évariste (1811–1832)
Gardner, Martin (1914–)
Gauss, Carl Friedrich (1777–1855)
Germain, Sophie (1776–1831)
Gödel, Kurt (1906–1978)
Golomb, Solomon W. (1932–)
Graham, Ronald L. (1936–)
Green, George (1793–1841)
Guy, Richard Kenneth (1916–)
Hamilton, William Rowan (1805–1865)
Hardy, Godfrey Harold (1877–1947)
Hausdorff, Felix (1868–1942)
Hein, Piet (1905–1996)
Henon, Michele
Hermite, Charles (1822–1901)
Heron of Alexandria (c. A.D. 60)
Hilbert, David (1862–1943)
Hinton, Charles Howard (1853–1907)
Hippias of Elis (c. fifth century B.C.)
Hoffmann, Louis "Professor" (1839–1919)
Hofstadter, Douglas R. (1945–)
Hogben, Lancelot Thomas (1895–1975)
Hordern, L. Edward (d. 2000)
Hunter, James Alston Hope
Huygens, Christiaan (1629–1695)
Hypatia of Alexandria (c. A.D. 370–417)
Jacobi, Karl Gustav Jacob (1804–1851)
Jordan, (Marie-Ennemond) Camille (1838–1922)
Kaczynski, Theodore John (1942–)
Kasner, Edward (1876–1955)
Kepler, Johannes (1571–1630)
Khayyam, Omar (1044–1123)
Kirkman, Thomas Penyngton (1806–1895)
Klein, Christian Felix (1849–1925)
Kolmogorov, Andrei Nikolaievich (1903–1987)
Kronecker, Leopold (1823–1891)
Lagrange, Joseph Louis (1736–1813)
Laplace, Pierre Simon (1749–1827)
Lebesgue, Henri Leon (1875–1941)
Leibniz, Gottfried Wilhelm (1646–1716)
Leurechon, Jean (c. 1591–1670)
Lie, Marius Sophus (1842–1899)
Lobachevsky, Nikolai Ivanovich (1793–1856)
Lorenz, Edward Norton (1917–)
Lovelace, Lady. *See* Byron, Ada
Loyd, Sam (1841–1911)
Lucas, (François) Edouard (Anatole) (1842–1891)
Maclaurin, Colin (1698–1746)
MacMahon, Percy Alexander (1854–1929)
Madachy, Joseph S.
Mandelbrot, Benoit B. (1924–)
Menaechmus (c. 380–c. 320 B.C.)
Mersenne, Marin (1588–1648)
Michell, John (1724–1793)
Minkowski, Hermann (1864–1909)
Mittag-Leffler, (Magnus) Gösta (1846–1927)
Moivre, Abraham de. *See* de Moivre, Abraham
Monge, Gaspard (1746–1818)
Montucla, Jean Etienne (1725–1799)
Mydorge, Claude (1585–1647)
Napier, John (1550–1617)
Napoleon Bonaparte (1769–1821)
Nash, John Forbes, Jr. (1928–)
Newton, Isaac (1642–1727)
Noether, Emmy (Amalie) (1882–1935)
Oughtred, William (1574–1660)
Ozanam, Jacques (1640–1717)
Pacioli, Luca (1445–1517)
Pappus of Alexandria (lived c. A.D. 300)
Pascal, Blaise (1623–1662)
Peano, Guiseppe (1858–1932)
Penrose, Roger (1931–)
Perelman, Yakov Isidorovitch (1882–1942)
Perigal, Henry (1801–1898)
Phillips, Hubert (1891–1964)
Plateau, Joseph Antoine Ferdinand (1801–1883)
Poincaré, (Jules) Henri (1854–1912)
Poinsot, Louis (1777–1859)
Poisson, Siméon Denis (1781–1840)
Poncelot, Jean Victor (1788–1867)
Proclus Diadochus (A.D. c. 410–485)
Pythagoras of Samos (c. 580–500 B.C.)
Ramanujan, Srinivasa Aaiyangar (1887–1920)

Recorde, Robert (c. 1510–1558)
Reutersvärd, Oscar (1915–)
Riemann, (Georg Friedrich) Bernhard (1826–1866)
Robinson, Abraham (1918–1974)
Rucker, Rudy (1946–)
Russell, Bertrand Arthur William (1872–1970)
Saccheri, Giovanni Girolamo (1667–1733)
Sallows, Lee C. F. (1944–)
Schläfli, Ludwig (1814–1895)
Schubert, Hermann (1848–1911)
Schuh, Frederick (1875–1966)
Shannon, Claude Elwood (1916–2001)
Sierpinski, Waclaw Franciszek (1882–1969)
Singmaster, David
Slocum, Jerry
Smullyan, Raymond (1919–)
Stanhope, Earl
Steiner, Jakob (1796–1863)
Steinhaus, Hugo (1887–1972)
Stewart, Ian (1945–)
Sylvester, James Joseph (1814–1897)
Tait, Peter Guthrie (1831–1901)
Tarry, Gaston (1843–1913)
Tarski, Alfred (1902–1983)
Tartaglia, Niccoló Fontana (1499–1557)
Thompson, D'Arcy Wentworth (1860–1948)
Turing, Alan Mathison (1912–1954)
Ulam, Stanislaw Marcin (1909–1984)
Ussher, James (1581–1656)
van der Pol, Balthazar (1889–1959)
von Neumann, John (1903–1957)
Wallis, John (1616–1703)
Weierstrass, Karl Wilhelm Theodor (1815–1897)
Wessel, Caspar (1745–1818)
Weyl, Hermann Klaus Hugo (1885–1955)
Wiener, Norbert (1894–1964)
Wiles, Andrew (1953–)

Board games and chess problems
backgammon
bishops problem
chess
Fitchneal
four knights puzzle
Fox and Geese
Frogs and Toads
Go
Hex
Hi-Q. *See* peg solitaire
Hnefa-Tafl
kings problem
knights problem
knight's tour
Mancala
nine holes. *See* three men's morris
nine men's morris
Ovid's game
peg solitaire
Phutball
queens problem
rooks problem
Rithmomachia
Senet
solitaire. *See* peg solitaire
Tafl game
three men's morris
tour
Wari. *See* Mancala

See also: **Games and game theory**

Calculus and analysis
analysis
boundary condition
boundary value problem
brachistochrone problem
calculus
calculus of variations
codimension
continuity
convergence
derivative
differential
differential equation
differentiation
discontinuity
diverge
Green's theorem
harmonic analysis
Henstock integration. *See* integration
infinitesimal
inflection
integral
integral equation
integration
least upper bound
limit
localized solution
maximum
measure
measure theory
method of exhaustion
minimum
ordinary differential equation
partial differential equation
Riemann integral
Rolle's theorem
smooth
stationary point

See also: **Complex numbers** and **Series and sequences**

Calendars, dates, ages, and clocks
age puzzles and tricks
birthday surprise
calendar curiosities
chronogram
clock puzzle
Diophantus's riddle
International Date Line
kinship puzzles
Zeller's formula

See also: **Time**

Cards
blackjack
cards
Martingale system
shuffle

See also: **Games and game theory**

Chaos, complexity, and dynamical systems
attractor
basin of attraction
bifurcation
butterfly effect
catastrophe theory
chaos
chaotic attractor
complex adaptive system
complex system
complexity
complexity theory
deterministic system
dissipative system
dynamical system
edge of chaos
emergence
equilibrium
ergodic
Feigenbaum's constant
fixed-point attractor
instability
logical depth
nonlinear system
nucleation
periodic attractor
quasiperiodic
redundancy
self-organization
self-organized criticality
sensitivity
strange attractor. *See* chaotic attractor
time-reversible

See also: **Fractals**

Codes and ciphers
Beal cipher
Caesar cipher
cipher
code. *See* cipher
coding theory
cryptography

Codes and ciphers *(Continued)*
substitution cipher
transposition cipher

Combinatorics
binomial
binomial coefficient
binomial theorem
Cayley's mousetrap
change ringing
combination
combinatorics
dinner party problem. *See* Ramsey theory
mousetrap. *See* Cayley's mousetrap
officer problem. *See* thirty-six officers problem
Pascal's triangle
permutation
Pick's theorem
plane partition
Ramsey theory
schoolgirls problem
thirty-six officers problem

Complex numbers
Argand diagram
complex analysis
complex number
complex plane. *See* Argand diagram
de Moivre's theorem
eigenvalue
Euler's formula
i
imaginary number
primitive root of unity
root of unity

Computing, artificial intelligence, and cybernetics
abacus
algorithm
algorithmic complexity
artificial intelligence
artificial life
bit
byte
cellular automaton
Chinese room
Church-Turing thesis
communication theory. *See* information theory
compressible
computability theory
computable number
connectionism
cybernetics
feedback
finite-state automaton
genetic algorithm
halting problem
incomputable number
information theory
lambda calculus
Langton's ant
Life, Conway's game of
model of computation
neural network
NP-hard problem
recursion
simulation
slide rule
soroban. *See* abacus
supercomputer
time complexity
Turing machine
Turing test
universal computer
von Neumann machine
zombie

Differential geometry
curvature
differential geometry
elliptical geometry
geodesic
hairy ball theorem
hyperbolic geometry
metric
Minkowski space
non-Euclidean geometry
Poincaré disk
Riemannian geometry. *See* elliptical geometry
spherical geometry. *See* elliptical geometry

See also: **Topology**

Dimensions, higher and lower
Flatland: A Romance of Many Dimensions
fourth dimension
higher dimensions
hypercube
hypersphere
polychoron
polytope
realm
Reuleaux polytope
simplex
tesseract
two-dimensional worlds

See also: **Solids and surfaces**

Dissection
Archimedes's square. *See* loculus of Archimedes
Blanche's dissection
cake-cutting
dissection
loculus of Archimedes
fair division. *See* cake-cutting
Haberdasher's puzzle
Hadwiger problem
Mrs. Perkins's quilt
Pythagorean square puzzle
stomachion. *See* loculus of Archimedes
T-puzzle

Fractals and pathological curves
Barnsley's fern
Cantor dust
dragon curve
fractal
fractal dimension
Hausdorff dimension
Mandelbrot set
Monster curve. *See* Peano curve
Peano curve
Julia set
Koch snowflake
Lorenz system
L-system
Lyapunov fractal
Menger sponge
pathological curve
self-similarity
Sierpinski carpet
Sierpinski gasket
snowflake curve. *See* Koch snowflake
space-filling curve
Volvox fractal
Weierstrass's nondifferentiable function

Functions
absolute value
Ackermann function
beta function
codomain
cosine. *See* trigonometric function
domain
elementary function
elliptic function
even function
function
gamma function
hypergeometric function
exponent
exponential
factorial
homomorphism
hyperfactorial
injection
jump discontinuity
linear programming
logarithm
mantissa
natural logarithm
one-to-one
operator
parameter
periodic
polynomial

range
recursive function
special function
trigonometric function
zero of a function. *See* root
zeta function

Games
game
game theory
Grundy's game. *See* Nim
minimax theorem
mixed strategy
Nash equilibrium
Nim
noughts and crosses. *See* tic-tac-toe
payoff
prisoner's dilemma
Sprague-Grundy theory
Sprouts
strategy
TacTix
Tangloids
Tetris
thirty colored cubes puzzle
tic-tac-toe
truel
utility function
Wythoff's game
zero-sum game

See also: **Board games and chess problems** and **Cards**

Geometry, general terms and theorems
acute
adjacent
affine geometry
alternate
altitude
angle
apex
apothem
arc
asymptote
Aubel's theorem
Bang's theorem
bisect
Brianchon's theorem
butterfly theorem
cathetus
Cavalieri's principle
center of perspective
central angle
centroid
chord
circumcenter
circumference
commensurable
congruent
constructible
curve
de Malves's theorem
Descarte's circle theorem. *See* Soddy formula
diagonal
diameter
equichordal point
equilateral
Euclidean geometry
Euler line
Fagnano's problem
focal chord
focal radius
focus
geometry
Gergonne point
great circle
half-line
half-plane
harmonic division
Heron's formula
hypotenuse
inscribed angle
isogonal conjugate
isoperimetric inequality
isosceles
isotomic conjugate
isometry
latus rectum
line
locus
main diagonal
major axis
Mascheroni construction
midpoint
minor axis
Morley's miracle
Nagel point
Neusis construction
nine-point circle
normal
oblique
obtuse
orthocenter
orthogonal
osculating
parallel
Pascal's mystic hexagon
perimeter
periphary
perpendicular
plane
point
Poncelot's theorem
projective geometry
prolate
Ptolemy's theorem
Pythagoras's theorem
quadrature
radical axis
ray
radius
reentrant angle
reflection
reflex angle
right
rotation
ruler-and-compass construction. *See* Mascheroni construction
salient
secant
sector
segment
self-intersecting
similar
six circles theorem
skew lines
Soddy's formula
Sperner's lemma
Steiner-Lehmus theorem
straight
supplementary angles
symmedian
tangent
transformation
translation
transversal
triangulation
trigonometry
trisector theorem. *See* Morley's miracle
vertex
Viviani's theorem
volume

See also: **Coordinate geometry; Geometry, problems; Plane curves; Solids and surfaces;** and **Space curves**

Geometry, problems
angle bisection. *See* bisecting an angle
angle trisection. *See* trisecting an angle
Apollonius problem
bisecting an angle
duplicating the cube
four coins problem
geometry puzzles
hole-in-a-postcard problem
hole-through-a-sphere problem
isovolume problem
Johnson's theorem
Kakeya needle problem
Langley's adventitious angles
Prince Rupert's problem
rope around the Earth puzzle
spider-and-fly problem
squaring the circle
squaring the square
Sylvester's problem of collinear points
tautochrone problem

Geometry, problems *(Continued)*
trisecting an angle
unilluminable room problem

Graphs and graph theory
Around the World game. *See* Icosian game
axis
bridges of Königsberg
complete graph
connected graph
digraph
directed graph
Euler circuit
Euler path
four-color map problem
graceful graph
graph
graph theory
Hamilton circuit
Hamilton path
Icosian game
Königsberg bridge problem. *See* bridges of Königsberg
Tait's conjecture
tangled graph
traveling salesman problem
tree
twelve-color map problem

Groups
Abelian group
automorphism
character theory
enormous theorem
fundamental group
Galois theory
group
linear group
Monster group
Monstrous Moonshine conjecture
representation theory
semigroup
simple group
subgroup
symmetric group
symmetry group
topological group
wallpaper group

History of mathematics
Ahmes papyrus. *See* Rhind papyrus
apotome
Bakhshali manuscript
disme
Egyptian fraction
Ishango bone
Lebombo bone
Ludolph's number
Moscow papyrus. *See* Rhind papyrus
quipu
Rhind papyrus
Roman numerals
tally
tetraktys
Tinner's rabbits

See also: **Biographies**

Illusions and impossible figures
ambiguous connectivity. *See* impossible figures
ambiguous figure
Ames room
anamorphosis
antigravity houses and hills
Benham's disk
distortion illusion
Fraser spiral
Freemish crate. *See* impossible figure
Get off the Earth
Hermann grid illusion
Herring illusion
impossible figure
impossible tribar. *See* Penrose triangle
impossible trident
irradiation illusion
lateral inhibition illusion. *See* Hermann grid illusion
Moiré pattern
Müller-Lyer illusion
Necker cube
optical illusion
Orbison's illusion
Penrose stairway
Penrose triangle
Poggendorff illusion
Schröder's reversible staircase
scintillating grid illusion. *See* Hermann grid illusion
Thiery figure
Titchener illusion
tribar illusion. *See* Penrose triangle
upside-down picture
vertical-horizontal illusion
Wundt illusion
Zöllner illusion

Infinity
aleph
continuum hypothesis
infinite dimensions
infinity
transfinite number

Large numbers
chained arrow notation. *See* Conway's chained arrow notation
Conway's chained-arrow notation
Knuth's up-arrow notation
large numbers
power tower
scientific notation
up-arrow notation. *See* Knuth's up-arrow notation

See also: **Numbers, special**

Logic
bilateral diagram
Boolean
Boolean algebra
Caliban puzzle
excluded middle law
existence
fuzzy logic
Gettier problem
logic
QED
quantifier
quine
strange loop
syllogism
tangle
Venn diagram
wff

See also: **Mathematics, foundations**

Magic squares
alphamagic square
antimagic square
bimagic square
magic cube
magic square
magic tour
Euler square
gnomon magic square
Latin square
pandiagonal magic square
semi-magic square

Mathematics, foundations
applied mathematics
axiom
axiom of choice
barber paradox. *See* Russell paradox
category theory
classification
complete
conjecture
consistency
Euclid's postulates
formalism
formal system
Gödel's incompleteness theorem
heuristic argument
hypothesis. *See* conjecture
induction
isomorphism
lemma
mathematics
nonstandard analysis

parallel postulate
Platonism
proof
pure mathematics
reductio ad absurdum
Russell's paradox
set of all sets. *See* Russell's paradox
theorem
universe of discourse

See also: **Logic**

Mathematics, miscellaneous
animals' mathematical ability
area codes
beauty and mathematics
bistromathics
Fields Medal
films and plays involving mathematics
Great Monad
Hofstadter's law
I-Ching
impossibilities in mathematics
Journal of Recreational Mathematics
mathematical lifespan
origami
pizza
ying-yang symbol. *See* Great Monad

Mazes
labyrinth. *See* maze
maze
Rosamund's bower

Measuring and units
acre
area
bottle sizes
counterfeit coin problem
cubit
league
measuring and weighing puzzles
meter
mile
minute
nano-
peta-
pico
radian
second
steradian
tera-
weighing puzzles. *See* measuring and weighing puzzles
yard

Mechanical puzzles
Altekruse puzzle
burr puzzle
Cardan's rings. *See* Chinese rings
Chinese cross. *See* burr puzzle
Chinese rings
Fifteen Puzzle
matchstick puzzles
mechanical puzzles
Pony Puzzle
puzzle jug
puzzle rings
puzzle vessels
Rubik's cube
sliding-piece puzzles
Soma cube
tangrams
Tower of Brahma
Tower of Hanoi
vanishment puzzles

Mirrors and symmetry
bilateral
chiral
enantiomorph
left-right reversal. *See* mirror reversal problem
mirror reversal problem
reflexible

Number theory
ABC conjecture
algebraic number theory
aliquot part
analytical number theory
Archimedes's cattle problem
Beal's conjecture
Brocard problem
Catalan's conjecture
Chinese remainder theorem
Collatz problem
congruum problem
digit
digital root
Diophantine approximation
Diophantine equation
elliptic curve
Euler's conjecture
Fermat number
Fermat's last theorem
Gelfond's theorem
modulo
multigrade
number line
number theory
parity
Pell equation
proper divisor. *See* aliquot part
Riemann hypothesis
Riemann zeta function
Ulam's conjecture. *See* Collatz problem
Waring's conjecture

Numbers, special
Apéry's constant
apocalypse number. *See* beast number
Avagadro constant
baker's dozen. *See* twelve
beast number (666)
billion. *See* large numbers
Catalan's constant
centillion. *See* large numbers
century
Chaitin's constant (Ω)
Champernowne's number
dozen. *See* twelve
e
Eddington number
eight
eleven
Euler-Mascheroni constant
Euler's constant. *See* Euler-Mascheroni constant
five
four
gamma. *See* Euler-Mascheroni constant
golden ratio (phi, ϕ)
googol
Graham's number
gross
hexa-
hundred
Kaprekar constant
Khintchine's constant
Leibniz harmonic triangle
million
myriad
nine
octa-
Omega. *See* Chaitin's constant
one
153
penta-
phi. *See* golden ratio
pi (π)
plastic number
score
seven
seventeen
sexagesimal
six
sixty. *See* sexagesimal
Skewes' number
square root of 2
ten
thirteen
thousand
three
Thue-Morse constant
trillion
triskaidekaphobia. *See* thirteen
twelve
twenty
two
vigesimal
Wallis formula. *See* pi
zero

Numbers, types
abundant number
algebraic number
almost perfect number
amicable numbers
Arabic numeral
automorphic number
Bell number
Bernouilli number
Betti number
cardinal number
Carmichael number
Catalan number
Cayley number. *See* octonion
chromatic number
common fraction
composite number
continued fraction
Cullen number
cute number
cyclic number
decimal
decimal fraction
deficient number. *See* abundant number
economical number
EPORN
equivalent numbers
extravagant number. *See* economical number
factorion
Fibonacci sequence
figurate number
fraction
friendly number. *See* amicable number
frugal number. *See* economical number
happy number
Harshad number
hyperreal number
integer
interesting numbers
irrational number
Kaprekar number
Liouville number
lucky number
Mersenne number
narcissistic number
natural number
negative numbers
normal number
number
numeral
numerator
oblong number
octonion
ordered pair
ordinal number
palindromic number
pandigital number
partition number
Pell numbers
perfect cube
perfect number
perfect power
perfect square
polygonal number
powerful number
practical number
pronic number
pyramidal number
quaternion
rational number
real number
repdigit
rep-unit
schizophrenic number
Sierpinski number
Smith number
solitary number
square free
squarefull number. *See* powerful number
sublime number
supertetrahedral number
surd
surreal number
terminating decimal
tetrahedral number
transcendental number
triangular number
trimorphic number. *See* automorphic number
undulating number
unique number
unit fraction
untouchable number
vampire number
vulgar fraction. *See* common fraction
weird number. *See* abundant number
winding number

See also: **Complex numbers** and **Infinity**

Packing
cannonball problem
Kepler's conjecture
moving sofa problem
packing
Slothouber-Graatsma puzzle
square pyramid problem. *See* cannonball problem
sphere packing. *See* packing, Kepler's conjecture, and cannonball problem

Paradoxes
Achilles and the Tortoise paradox. *See* Zeno's paradoxes
Allais paradox
Aristotle's wheel
Arrow paradox
Banach-Tarski paradox
Berry's paradox
Bertrand's box paradox
Burali-Forti paradox
Buridan's ass
catch-22
coin paradox
Epimenides paradox. *See* liar paradox
Grelling's paradox
liar paradox
Newcomb's paradox
nine rooms paradox
paradox
Parrondo's paradox
potato paradox
raven paradox
Richard's paradox. *See* Berry's paradox
St. Petersburg paradox
Siegel's paradox
unexpected hanging
wheel paradoxes
Zeno's paradoxes

See also: **Illusions and impossible figures** and **Mathematics, foundations**

Places and buildings
Alhambra
Atomium, the
Giant's Causeway
Triangular Lodge

Plane curves
anallagmatic curve
annulus
arbelos
Archimedean spiral
astroid
bicorn
bicuspid curve
Barbier's theorem. *See* curve of constant width
caduceus
cardioid
Cartesian oval
Cassinian ovals
catenary
caustic
Cayley's sextic
circle
circle involute
circumcircle
cissoid
cochleoid
conchoid
conic section
crunode
cubic curve

curve of constant width
cusp
cycloid
de l'Hopital's cubic. *See* Tschirnhaus's cubic
delta curve
deltoid
devil's curve
directrix
Dürer's shell curve
eccentricity. *See* conic section
egg
eight curve
ellipse
epicycloid
epitrochoid
escribed circle
evolute
Flower of Life
folium
Freeth's nephroid. *See* nephroid
glissette
hippopede
hyperbola
hyperbolic spiral
hyperellipse. *See* superellipse
hypocycloid
hypoellipse. *See* superellipse
hypotrochoid
incircle
inscribed circle
involute
isochrone
Jordan curve
Kampyle of Eudoxus
kappa curve
Lamé curve
lemniscate
limaçon of Pascal
Lissajous figure
lituus
logarithmic spiral
Maclaurin trisectrix
Malfatti circles
Neile's parabola
nephroid
node. *See* crunode
oval
parabola
pearls of Sluze
pedal curve
Plateau curves
pursuit curve
quadratrix of Hippias
quadrifolium. *See* rose curve
ranunculoid
rectangular hyperbola. *See* hyberola
Reuleaux triangle
rose curve
rotor
roulette
salinon
serpentine
shell curve. *See* Dürer's shell curve
Soddy circle
spirograph curve. *See* roulette
strophoid
superellipse
tractrix
trefoil curve
tricuspoid. *See* deltoid
trident of Newton
trifolium. *See* rose curve
trisectrix
trochoid
Tschirnhaus's cubic
twisted cubic
unduloid
unit circle
Vesica Piscis
witch of Agnesi curve. *See* Agnesia, Maria Gaetana

See also: **Tiling**

Polygons
cross
cyclic polygon
decagon
dodecagon
golygon
heptagon
kite
limping triangle
lozenge
medial triangle
monochromatic triangle
n-gon
nonagon
oblong
octagon
orthic triangle
parallelepiped
parallelogram
pedal triangle
polygon
Pythagoras's lute
Pythagorean triangle
Pythagorean triplet
quadrangle
quadrilateral
rectangle
rhombus
scalene triangle
square
Star of David
Star of Lakshmi
swastika
trapezium
trapezoid
triangle
vertex figure

See also: **Polyominos**

Polyominos
broken chessboard. *See* polyomino
domino
pentomino. *See* polyomino
polyiamond
polyomino
flexagon
tetromino. *See* polyomino
triomino
X-pentomino

Prime numbers
Bertrand's postulate
Brun's constant. *See* twin primes
circular prime
coprime
Cunningham chain
deletable prime. *See* truncatable prime
Dirichlet's theorem
emirp
Eratosthenes's sieve. *See* sieve of Eratosthenes
factor
Fermat's little theorem
Fortune's conjecture
Gilbreath's conjecture
Goldbach conjecture
Mersenne prime
minimal prime
permutable prime
Pólya's conjecture
prime number
primitive root
primorial
pseudoprime
sieve of Eratosthenes
snowball prime
Sophie Germain prime
strobogrammatic prime
truncatable prime
twin primes
Ulam spiral
Wilson's theorem

Probability and statistics
Bayesian inference
Bayes's theorem
bell curve
Benford's law
birthday paradox
Buffon's needle
chance. *See* probability
coincidence
covariance
de Méré's problem

Probability and statistics *(Continued)*
dice
Gaussian
hat problem
Markov chain
monkeys and typewriters
Monte Carlo method
Monty Hall problem
null hypothesis
power law
probability
probability theory
random
random number
random walk
secretary problem. *See* sultan's dowry
standard deviation
statistics
stochastic
sultan's dowry
three hat problem. *See* hat problem
Vickrey auction

Puzzles, miscellaneous
dollar
missing dollar problem
postage stamp problems
railroad problems. *See* shunting puzzles
river-crossing problems
shunting puzzles
string puzzles
what color was the bear?

Science and philosophy
absolute
absolute zero
book-stacking problem
Brownian motion
bubbles
buckyball
celt
dynamics
entropy
gauge field
general relativity. *See* relativity theory
God
hole-through-the-Earth problem
holism
Kaluza-Klein theory
many worlds hypothesis
monad
nothing
perturbation
phase transition
pitch drop experiment
quantum field theory
quantum mechanics
reductionism
relativity theory
renormalization
soap film. *See* bubbles
soliton
special relativity. *See* relativity theory
spectrum
spinor
statistical mechanics
stochastic process
string theory
supersymmetry
theory of everything
three-body problem
Tippee top
walking in the rain problem
wine

See also: **Chaos, complexity, and dynamical systems** and **Time**

Series and sequences
arithmetic sequence
Beatty sequences
Farey sequence
Fourier series
geometric sequence
hailstone sequence
harmonic sequence
infinite series
Lucas sequences
Padovan sequence. *See* plastic number
Perrin sequence
power series
quartile
sequence
series
unimodal sequence

Sets and set theory
cap
continuum
countable set
cup
empty set
finite
homeomorphism
model theory
null set. *See* empty set
recursively enumerable set
set
set theory
subset
uncountable set
union

Solids and surfaces
antiprism
arch
Archimedean dual. *See* Catalan solid
Archimedean solid
ball
Catalan solid
catenoid
circular cone
compound polyhedron
cone
Császár polyhedron
cube
cuboctahedron
cuboid
cylinder
Dandelin spheres
deltahedron
dihedral angle
dodecahedron
dual
Dupin cyclide
edge
edge coloring theorem. *See* Tait's conjecture
ellipsoid
envelope
Euler's formula for polyhedra
face
frustum
Gabriel's horn
geoboard
hyperboloid
icosahedron
Johnson solid
Kepler-Poinsot solids
minimal surface
nonconvex uniform polyhedron
octahedron
paraboloid
Plateau problem
Platonic solid
lattice
lattice path
lattice point
lune
nappe
net
oblate spheroid
polycube
polyhedron
prism
pseudosphere
pyramid
quasi–regular polyhedron
regular polygon
regular polyhedron
ruled surface
saddle
Schläfli symbol
semi-regular polyhedron
solid
solid angle
sollid geometry
sphere
sphericon
spheroid
stellation
Stewart toroid
superegg

surface
surface of revolution
Szilassi polyhedron
tetrahedron
torus
truncate
uniform polyhedron
unit cube
zone
zonohedron

See also: **Dimensions** and **Space curves**

Space curves
circular helix. *See* helix
clelia
helix
conical helix. *See* helix
helicoid
loxodrome
Viviani's curve

Terminology, general
amplitude
argument
base
basis
canonical form
cell
closed
complement
concave
conjugate
connectivity
convex
degree
degree of freedom
discrete
frequency
identity
inequality
intersection
inverse
linear
loop
map
monotonic
obelus
odd
open
orbit
order
path. *See* trajectory
phase space
pole
power
radical
radix. *See* base
root
search space
solidus
space
state space. *See* phase space
string
symmetry
system
ternary
trajectory
triple
vinculum

Tiling
aperiodic tiling
Archimedean tessellation
arrowhead. *See* dart
chaos tiles. *See* Penrose tiles
dart
domino problem
Einstein problem
Eternity puzzle
Heesch number
irreptile. *See* rep-tile
MacMahon squares
Penrose tiling
periodic tiling
quasicrystal
rep-tile
tessellation. *See* tiling
tiling
Voronoi diagram

Time
causality
grandfather paradox
Planck time
space-time
tachyon
time
time dilation. *See* relativity theory
time travel
twins paradox. *See* relativity theory
worldline

Topology
Alexander's horned sphere
algebraic topology
Borromean rings
Borsuk-Ulam theorem
braid
Brouwer fixed-point theorem
bundle
Calabi-Yau space
cohomology
conformal mapping
connected
differential topology
dimension
Dirac string trick
disk
Earthshapes
embedding
Euclidean space
Euler characteristic
foliation
general topology. *See* point-set topology
genus
Gordian knot
ham sandwich theorem
Hilbert space
hole
homeomorphic
homology
homotopy
Klein bottle
knot
lamination
manifold
metric space
metrizable
Möbius band
pleated surface
Poincaré conjecture
point-set topology
projective plane
Riemann sphere
Riemann surface
round
simply-connected
tie knots
topological dimension
topological space
topology
Whitney's umbrella

Words and word puzzles
abracadabra
alphametic
anagram
autogram. *See* self-enumerating sentence
limerick
palindrome
pangram
self-enumerating sentence
self-referential sentence
snow
Sphinx riddle
tautology
Universal Library
word puzzles
word trivia

MORGAN

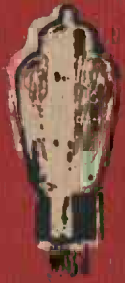

THE BOYS' SECOND BOOK OF
RADIO AND ELECTRONICS

THE BOYS' SECOND BOOK of RADIO and ELECTRONICS

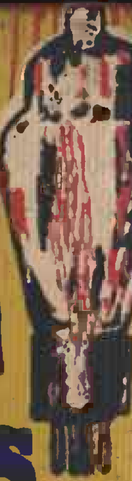

Alfred
Morgan

THE BOYS' SECOND BOOK OF RADIO AND ELECTRONICS

OTHER BOOKS BY ALFRED MORGAN

The Boys' First Book of Radio and Electronics
First Chemistry Book for Boys and Girls
A Pet Book for Boys and Girls
The Boys' Book of Engines, Motors and Turbines
Things A Boy Can Do with Electricity
An Aquarium Book for Boys and Girls
A First Electrical Book for Boys

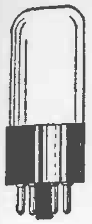

THE BOYS' SECOND BOOK OF

Radio

AND

Electronics

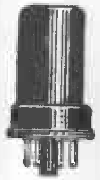

by ALFRED MORGAN

WITH DIAGRAMS BY THE AUTHOR

CHARLES SCRIBNER'S SONS   NEW YORK

COPYRIGHT © 1957 BY ALFRED MORGAN

H-12.63[H]

All rights reserved. No part of this book may be reproduced in any form without the permission of Charles Scribner's Sons.

First printing	May 1957
Second printing	January 1958
Third printing	September 1958
Fourth printing	November 1959
Fifth printing	December 1960
Sixth printing	June 1960
Seventh printing	May 1963
Eighth printing	December 1963

PRINTED IN THE UNITED STATES OF AMERICA
LIBRARY OF CONGRESS CATALOG CARD NUMBER 57-6078

Science Beckons to Youth

Brig. General David Sarnoff
Chairman of the Board
RADIO CORPORATION OF AMERICA

In the guided missiles program and in all fields affecting our national security, research and engineering are of surpassing importance. There is no substitute for brains, or for practical training in a technical age.

Our safety and our industrial strength rest upon our success in expanding the Nation's reservoir of physicists and scientists, trained engineers and technicians. Our economy and national security alike will suffer seriously unless we solve this problem promptly and vigorously.

Science and technology are the very hallmarks of American civilization.

In the not too distant past, science was associated with men of age and experience. Science was moving slowly at that time and its discoveries developed leisurely. The pressure of time was not a vital factor in the birth or use of an invention. Even when a great discovery was made, a skeptical public was slow to accept it because public interest in science was limited and faith in its possibilities was not what it is today.

In contrast with that past, youth today is largely in the forefront of science. Young men and women hold strategic places in its advancement. Young experimenters now enjoy the vast facilities of university and industrial laboratories and a free exchange of knowledge with others.

These young people have noted professors and their libraries from which to draw knowledge from other men's experience. As a result, they may easily learn what has been done before and need not waste their time and energy going over ground already covered by other scientific experimenters.

Science is much like a football game; teamwork is the key to success, but some player trained to take advantage of opportunity will pick up the ball and score a touchdown. That has happened time and time again in science and industry. Science seems to have an affinity for youth. Genius knows no age.

Contents

FOREWORD vii

1 An Adventure in Science 3

2 Building Your First Radio Receiver 15
GERMANIUM SEMI-CONDUCTORS

3 How to Build a More Selective Crystal Receiver 26

4 Taking the Mystery out of Tuning 36

5 A Simple Outdoor Receiving Antenna 48
STATIC AND STRAYS

6 Components of Electronic Circuits 52
CAPACITORS AND CAPACITANCE
RESISTORS AND CONTROLS

7 Transducers—Components of an Electronic Circuit 70
EARPHONES, LOUDSPEAKERS, MICROPHONES AND PICKUPS

8 Transistors—Versatile Midgets Which Can Do the Work of Vacuum Tubes 89

9 Transistor Apparatus Which You Can Build 99
A ONE-STAGE AUDIO AMPLIFIER
A COMBINATION CRYSTAL DIODE AND TRANSISTOR RADIO RECEIVER
A TRANSISTOR TIMER SWITCH
A TRANSISTOR CODE PRACTICE OSCILLATOR
LEARNING TO TELEGRAPH

10 About Radio Tubes 125

11 Introduction to Amplifiers 135

12 The Photocells and Phototubes Which Are Popularly Called Electric Eyes 143
HOW TO BUILD A PHOTOTUBE RELAY

13 Prospecting for Uranium 171
RADIOACTIVITY
GEIGERSCOPES
GEIGER COUNTERS
SCINTILLOMETERS
HOW TO MAKE A GEIGER COUNTER

14 How to Build a Two-stage Audio Amplifier 199

15 How Phonograph Records Are Made 214
BUILD YOUR OWN RECORD PLAYER
YOU CAN PLAY RECORDS WITH THE AMPLIFIER IN A RADIO RECEIVER
HOW TO BUILD THE HUSH-A-TUNE
REPAIRS AND ADJUSTMENTS TO LOUDSPEAKERS

16 Electron Tubes as Oscillators and Some of Their Uses 239
CAPACITY RELAYS
YOU CAN BUILD A CAPACITY RELAY
FUN WITH A CAPACITY RELAY

17 How to Build and Operate a Miniature Home Broadcasting Station 254

Where to Obtain Materials 269

Index 271

THE BOYS' SECOND BOOK OF RADIO AND ELECTRONICS

CHAPTER ONE

An Adventure in Science

We live in a complex technological world. Progress in science and engineering is not only desirable but has become an essential part of our civilization. We must look to this progress for the maintenance of our high standard of living. In it is much of our hope for peace and survival. More and more we are dependent on people who are well trained in scientific and technical skills.

Are you looking for an interesting hobby—a hobby which may lead to a more intelligent understanding of the world in which we live and which may also lead to a scientific career? Are you undecided about your future? Do you still have to make a decision whether you should study law, medicine, engineering, or go into business? You may find an answer in this book. It deals with an intriguing, constantly growing and probably unlimited branch of science and engineering.

America needs twice as many scientists and engineers as we are turning out. By 1960, the U.S.A. will be short at least 100,000 scientists and engineers. Opportunities for employment and advancement in engineering and in the sciences have never been so great or numerous. The best way to discover whether or not you have natural ability for one of these professions is to adopt some branch of engineering or science as a hobby. Electronics is recommened as an enjoyable, engrossing and instructive scientific spare time activity.

An interest in electronics will bring adventures in a scientific wonderland. Radiotelegraphy and radiotelephony, television, hi-fi phonographs, sound motion pictures, radar, telephoto transmission of pictures, guided missiles, "electric eyes" and "snooperscopes" for seeing in the dark are only a few of the marvelous things found in the realm of electronics. Apparatus with which you can communicate with other "amateurs" in far away Hawaii or on the other side of the world in Europe is not difficult to construct. Furthermore, a sustained interest in electronics as a hobby will provide the foundation for a professional career as an engineer or scientist. The principles of electronics are basically electrical principles but physics and chemistry are also involved. When this nation's industries are listed according to their size, electronics is No. 9 on the list.

Electronics is a science and an art. A science, you may know, is systematized knowledge obtained by study and practice. What is an art? Well, an art is the practical application of knowledge and skill.

Electronics includes so much of the science of physics and of electrical engineering that it is difficult to state a

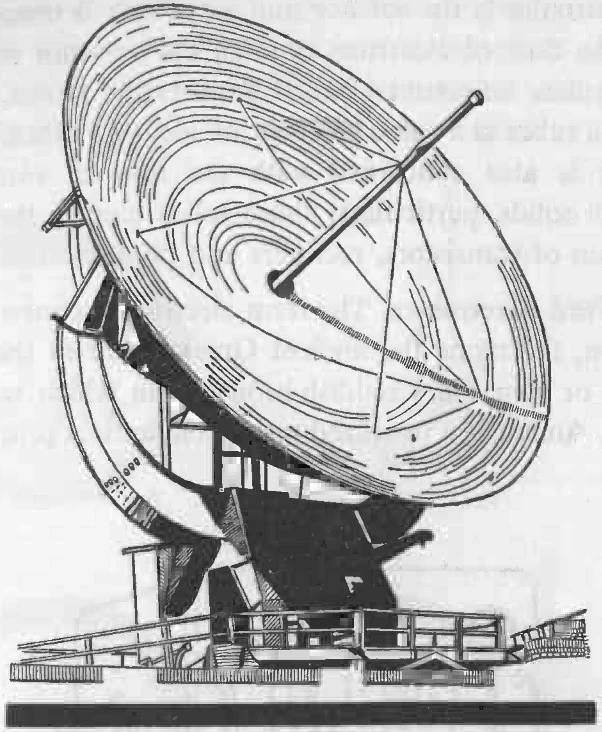

RADIO WAVES FROM THE STARS

This is the 50-foot radio mirror and 21 centimeter antenna at the Naval Research Laboratory in Washington, D. C. where it is used as a *radio telescope*. Radio radiation reaching the earth from outer space is adding to our knowledge of the physical universe of stars and galaxies. Very large antennas and sensitive electronic recording apparatus are required in the operation of a radio telescope.

Until 1931 all our knowledge of the physical universe of stars and galaxies came from studies made with optical telescopes. Atmospheric absorption limits research with optical telescopes to about 5 octaves of the electromagnetic spectrum. On the other hand, radio telescopes can record radiation for the entire range of wave lengths between 1 centimeter and 30 meters. This is roughly 12 octaves of the electromagnetic spectrum.

brief yet comprehensive definition of what is included in the term ELECTRONICS. A short definition, ample for the purpose of this book, is:

Electronics is the science and art which is concerned with the flow of electrons through the vacuum or gas-filled tubes sometimes called ELECTRON TUBES. The vacuum tubes in a radio receiver are electron tubes. Electronics is also concerned with the flow of electrons through solids, particularly those solids used in the construction of transistors, rectifiers and photoelectric cells.

The Word Electronics. The term electronics comes from *electron,* the name the ancient Greeks gave to the pale yellow or sometimes reddish-brown resin which we call amber. Amber is a fossilized resin from extinct pine trees

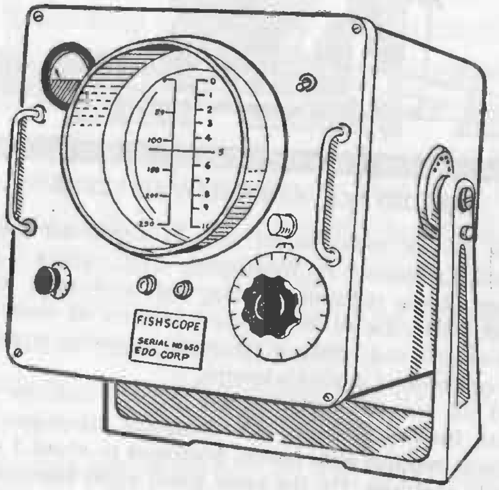

THE EDO FISHSCOPE,
AN ELECTRONIC APPARATUS WHICH FINDS FISH WITH SONAR

The Edo Fishscope employs the principles used in under-water sonar detection gear for the Navy. It gives a clear, accurate picture on an easily readable cathode-ray tube of schooled or even individual fish swimming just below the surface to as deep as 250 fathoms. The Fishscope also serves as a very accurate depth sounder.

which existed many ages ago. Pieces of the resin were worn by ancient people as jewels and decorations. Amber beads which were buried in prehistoric times have been dug up from tombs at Mycenae and Etruria. Amber's

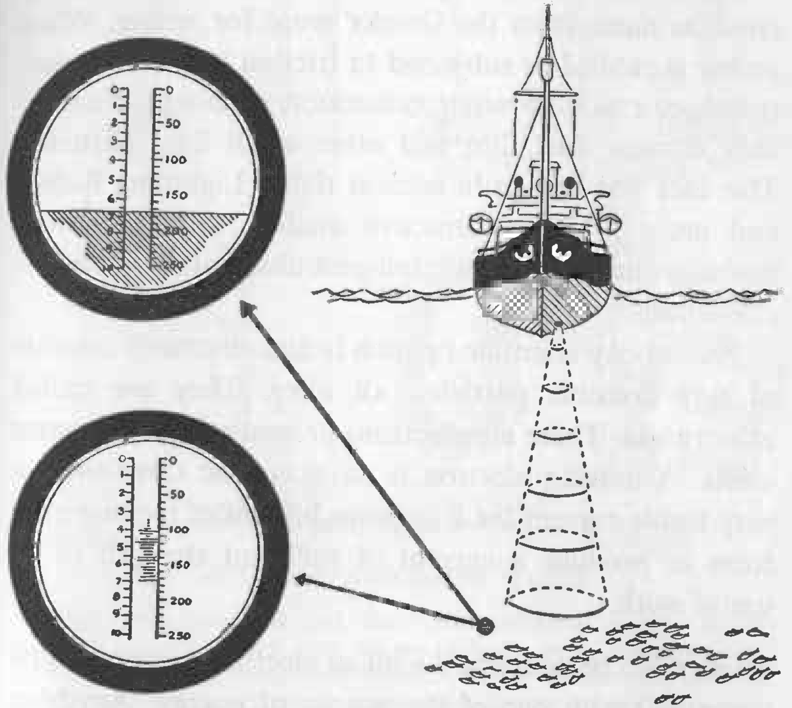

HOW THE EDO FISHSCOPE OPERATES

Sonar signals are sent forth from a transducer located in the bottom of the ship. If the signals strike fish nearby, the signals are reflected back to the transducer and produce "pips" on the screen of a large cathode ray tube in the Fishscope. The right-hand sketch shows how signals from the transducer spread out as they travel toward "bottom." The solid curved lines are the outgoing sonar signals. The dotted curved lines are the sonar signals reflected from the fish. The upper left-hand sketch shows how a school of fish appears on the 0- to 250-fathom scale on the tube. By flipping a switch to the 10-fathom scale, the view is magnified 25 times. The exact depth of the fish is determined by centering the enlarged view on the scope and reading the depth indicated on the scale (*see lower left*).

yellow luster must have reminded the Greeks of native gold, for they applied the same name to amber and gold. Amber and gold were both *electron* or "children of the sun" to the Greeks.

There is a reason why the science of electronics derived its name from the Greeks' word for amber. When amber is rubbed or subjected to friction it becomes electrified, or CHARGED WITH ELECTRICITY, so that it noticeably attracts dust, lint and other small light particles. This fact was known in ancient times. Lightning flashes and amber's magic attractive quality, therefore, were probably mankind's first intelligent observation of electrical action.

Present-day scientific opinion is that electricity consists of tiny invisible particles, all alike. They are called ELECTRONS. There are electrons present wherever matter exists. A moving electron is an ELECTRIC CURRENT—a very feeble current for it requires billions of moving electrons to produce a current of sufficient strength to do useful work.

Energy. To be concerned with an electric current is to be concerned with one of the FORMS of energy. Anything which can be converted into work and which can arise from work is ENERGY. Energy can also be the CAPACITY for doing work. Energy takes different forms. The forms of energy of greatest interest in electronics are electricity, heat, light, mechanical energy, electromagnetic energy, radiant energy and sometimes chemical energy.

All of these different forms of energy can be converted into any of the other forms. All electronic equipment con-

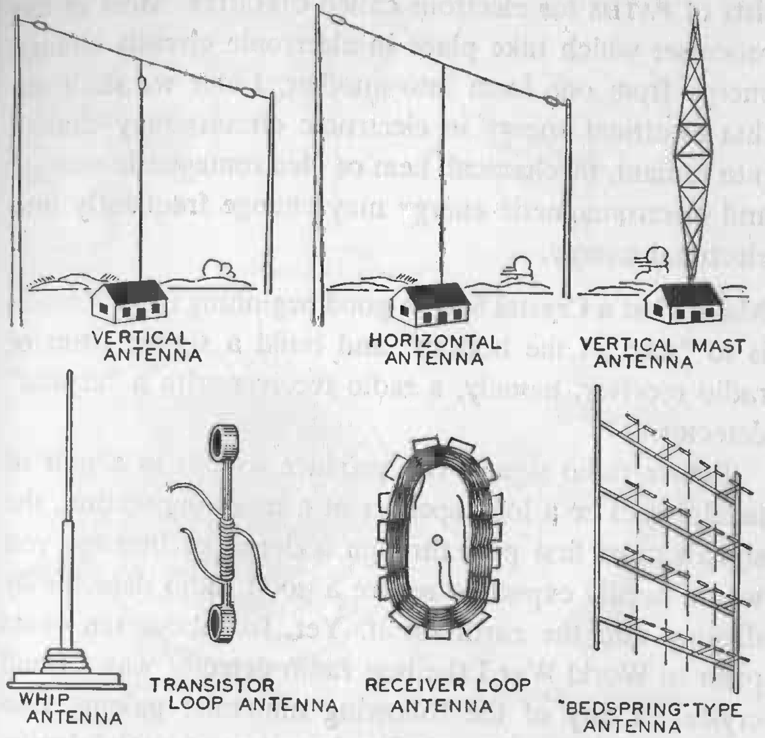

ANTENNAS AND RADIO WAVES

Radar, loran, television and radio communications employ electromagnetic waves to carry their signals through space. Electromagnetic waves are commonly called RADIO WAVES. A wire or conductor, or a system of wires or conductors, known as an ANTENNA, is necessary to radiate radio waves into space and to receive them from space.

Antennas are made in many forms and sizes. Portable broadcast receivers are usually equipped with a flat coil of wire called a LOOP ANTENNA. Many home broadcast receivers are provided with a loop antenna inside the cabinet. When not equipped with a built-in antenna, an external wire strung to a mast or tree is used. Automobile radio receivers and transmitters are connected to a WHIP ANTENNA. This is a metal tube about 6 feet long.

Television waves are of much higher frequency than the waves used for broadcasting. For that reason television transmitters and receivers use antennas which are different from those used for broadcasting waves. Radar antennas are of different shapes still. The illustrations show seven types of antennas. The BEDSPRING type is one form of radar antenna.

sists of PATHS for electrons called CIRCUITS. Most of the processes which take place in electronic circuits change energy from one form into another. Later we shall see that electrical energy in electronic circuits may change into radiant, mechanical, heat or electromagnetic energy, and electromagnetic energy may change frequently into electrical energy.

Make First a Crystal Set. A good beginning in electronics is to "start at the bottom" and build a simple form of radio receiver, namely, a radio receiver with a "crystal" detector.

Before radio signals can produce sounds in a pair of headphones or a loud speaker at a receiving station, the signals must first pass through a detector. Perhaps you would hardly expect to secure a good radio detector by digging into the earth for it. Yet, for about ten years prior to World War I the best radio detector was a small crystal of any of the following minerals: galena, iron pyrites, chalcopyrites, zincite, bornite and molybdenite.

For more than a generation uncounted thousands of interested youngsters and adults have constructed radio receivers with mineral or "crystal" detectors. The fascination of this device does not seem to wane.

Some amateur experimenters may prefer to put together pieces of equipment without knowing the "whys" of what they are doing. On the other hand, young experimenters with greater curiosity and more analytical minds wish to know the "whys" as well as the "hows" of radio. Here is an explanation of some of the fundamental processes that take place in a radio receiver.

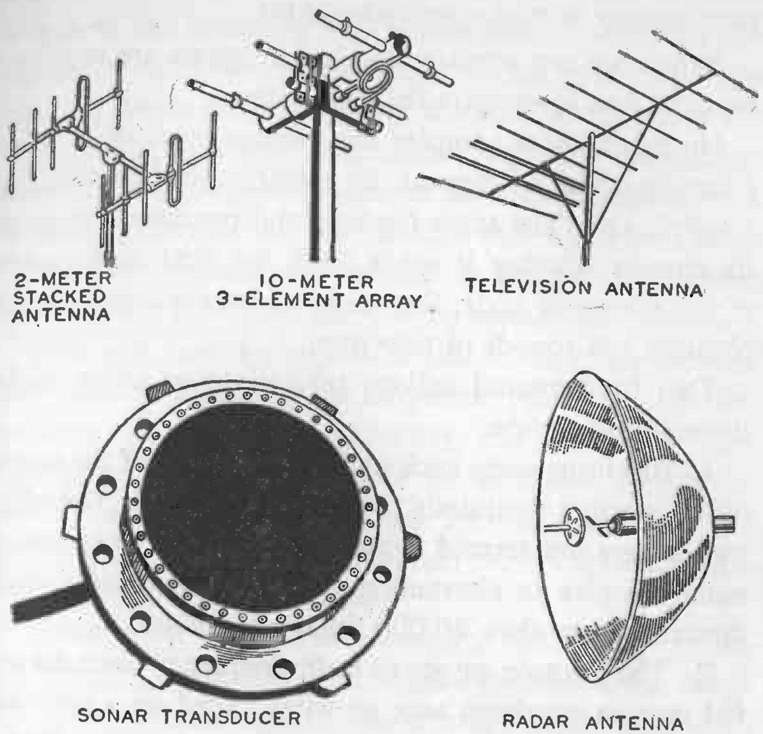

A SUITABLE ANTENNA IS IMPORTANT

Best results are obtained only when radio apparatus is connected to a suitable antenna. An antenna that is efficient at one frequency may be entirely unsatisfactory at another frequency. The antenna most suitable for receiving waves of a certain frequency is not always the best antenna for transmitting waves of the same frequency. The higher the frequency, the shorter or smaller the antenna required. At very low frequencies the physical dimensions of the antenna become so great that it is not always practical to employ the vertical type. These are some of the reasons that antennas are of many forms.

The 2-meter and 10-meter antennas illustrated above are types used frequently by amateur or "ham" operators. The television antenna is one of the many kinds for television receivers. The sonar transducer is a form of antenna for propagating and receiving sonar waves in water. The radar antenna at the lower right is the reflector, or mirror, type.

HOW RADIO SIGNALS ARE RECEIVED

Before we can explain how radio signals are received we must first investigate the transmitter.

No matter how complex the various parts of a radio transmitter make it appear, its performance is essentially a simple one. The same fundamental processes occur in its circuits whether it sends forth the dots and dashes of the telegraph code, the music of broadcasting or the pictures and sounds of television.

Two fundamental actions take place at every radio transmitting station.

1. Electrons surge back and forth in some of the transmitter circuits thousands of times per second. The electron surges are termed RADIO-FREQUENCY CURRENTS, a name we give to alternating currents that change their direction more than 20,000 times per second.

2. The electron surges or radio-frequency currents are fed into an overhead wire or wires called an ANTENNA. The surges send radio waves out through space. These radio waves are the CARRIERS of the radio signals.

When a transmitter is in operation there are several changes of energy from one form into another. The principal changes are: electrical energy into magnetic energy, magnetic energy into electrical energy and electrical energy into electromagnetic energy, the latter being a form of radiant energy related to light, heat and x-rays.

When the waves produced at the transmitting station reach an antenna connected to a receiver they set up electron surges in that antenna. The surges in the receiving antenna are of the same frequency as the electron

surges in the transmitting antenna which produced the waves. In other words, they occur at the same rate. The electron surges in the receiving antenna, like those in the transmitting antenna, are radio-frequency alternating currents. An alternating current reverses its direction at regular intervals. At one instant the electrons surge through the antenna wire toward the ground. At the next instant, they flow from the ground toward the antenna.

You will hear no signals if you connect an antenna and ground directly to an earphone. Antenna currents, because of their high frequency, are unable to flow through the electromagnets in an earphone. Moreover, when radio frequencies produce sounds, the sounds are ultrasonic—that is, their vibrations are so rapid they cannot be heard by human ears.

However, if you connect a crystal detector to an antenna and ground and connect an earphone to the detector, you will hear signals. The crystal detector makes the signals audible. The crystal does this because it functions as a one-way valve or RECTIFIER. Electron surges in the antenna can pass through it more easily in one direction than in the other. Thus the crystal detector can change or rectify an alternating current into a direct current. A single radio-frequency electron surge does not last long enough to have a useful effect on an earphone. But the surges which produce signals come in groups at an AUDIO frequency, which is to say at less than 20,000 per second, and therefore audible to human ears. Due to the rectifying action of the crystal, the portions of the surges in each group which pass through the crystal ALL FLOW IN

THE SAME DIRECTION. They are pulsating direct currents (currents flowing in one direction only) and can PRODUCE SOUNDS IN AN EARPHONE.

Detection or Demodulation. This process of sorting out the audio-frequency groups of pulses generated in an antenna by radio-frequency carrier waves is called DETECTION OR DEMODULATION. There could be no practical radio communication without detection or demodulation.

First Detectors. In the pioneering days of wireless telegraphy many types of detectors were devised and used but they all lacked dependability and great sensitivity. The use of crystal detectors began in 1906 when General H. H. Dunwoody of the U.S. Army discovered that a small piece of Carborundum would act as a sensitive detector. Carborundum is a crystalline abrasive used in the manufacture of grinding wheels. It is made by heating coke, sand, sawdust and salt in an electric furnace. Its chemical name is silicon carbide. A year later (1907) Greenleaf W. Pickard, an American engineer, developed the first crystal detector of the type which is still in use today.

Dr. Lee DeForest discovered the most practical detector in 1906 when he added a wire grid to the Fleming two-element vacuum tube valve. DeForest thus created not only a stable sensitive detector but also an amplifier which would increase the strength of signals.

Building Your First Radio Receiver

GERMANIUM SEMI-CONDUCTORS

Instructions for assembling a simple radio receiver with a crystal detector, using the circuit diagram shown on page 16, have been printed frequently in pamphlets and magazines. The design of the receiver is based on the use of a crystal diode and the standard antenna coil which is manufactured for superheterodyne broadcast receivers. All parts are factory made except the small wood base. The receiver can be assembled and made ready in an hour or two.

The following parts and materials are required for building the Simple Crystal Receiver:

1	Shellacked wood base
1	1N34 germanium crystal diode
1	Antenna coil with frequency range of 540-1750 kilocycles when used with a 365 mmfd. variable capacitor

1	365 mmfd. midget variable capacitor
1	0.001 mfd. mica capacitor
1	2,000-ohm radio headset
6	No. 2 Fahnestock connector clips
6	½-in. No. 5 round head wood screws

The 1N34 crystal diode in the list above is used as a detector. A germanium diode of the right type is the modern form of crystal detector. It is sensitive, and unlike the "cat-whisker" detector which was its predecessor in the early days of radio, a germanium diode needs no adjustment and does not get out of order unless mistreated.

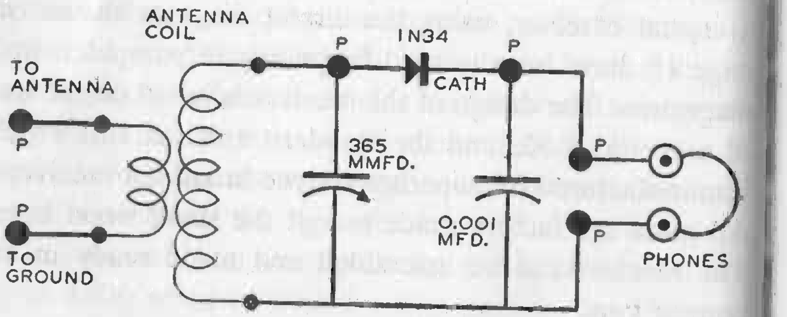

DIAGRAM OF CIRCUIT OF SIMPLE CRYSTAL RECEIVER

This receiver is assembled from standard parts. Nothing is homemade except a 5½ in. x 3¼ in. x ¾ in. wood base. The drawing is a SCHEMATIC diagram. Conventional symbols indicate the various parts or components of the circuit. The symbols used in the circuit diagrams in this book are shown on another page. Memorize them so that you can read any of the schematic diagrams. Schematic diagrams are not as easy at first for the beginner to understand as pictorial diagrams. A pictorial diagram of the same circuit follows this one. Schematic diagrams are more accurate and convey more information than pictorial diagrams. Both kinds appear in this book.

Germanium is of great importance in modern electronics. Since it will be mentioned again in later chapters let's find out what germanium is and what diodes are.

Germanium. Germanium is one of the rare and least known metals of the earth. This grayish-white crystalline substance of great hardness and brittleness was first discovered in 1886 by Clemens Winkler, Professor of Chemistry at the Freiberg School of Mines. He named his discovery germanium in honor of his fatherland. Germanium remained an insignificant metal, merely a laboratory curiosity, for more than half a century. But during the last decade it has become one of the most valuable metallic substances in existence. In 1942 the National Defense Research Council sought metal with certain special properties for use in the field of electronics. Germanium was found to have those properties. Its utility in electronics soon caused it to become five times

PICTORIAL DIAGRAM OF CIRCUIT OF SIMPLE CRYSTAL RECEIVER

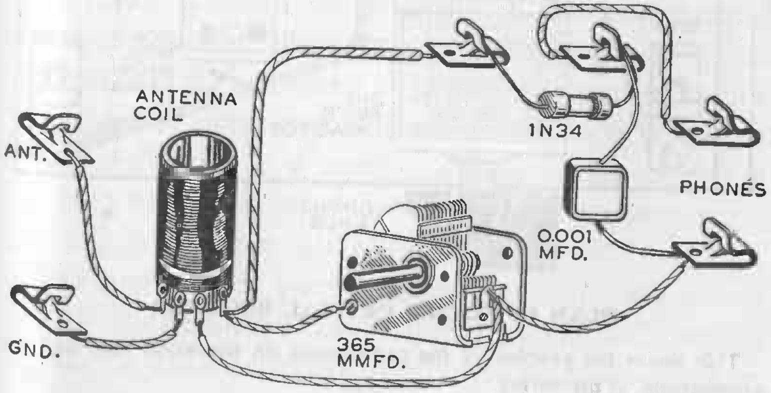

as valuable as gold. But as more sources of germanium were found and better methods of refining it were developed, the cost was reduced. Pure germanium is worth about $300.00 per pound today. Gold is worth $510.40 a pound.

The useful electrical properties which give germanium its importance in electronics are due to its behavior as a SEMI-CONDUCTOR.

Semi-conductors. The materials through which electrical currents flow easily are called CONDUCTORS. Most of the metals are fairly good conductors, copper and silver being the best. Materials through which electric currents do not flow are called NON-CONDUCTORS. Materials such as glass, porcelain, rubber, Bakelite, etc., used to insulate current-carrying wires are non-conductors.

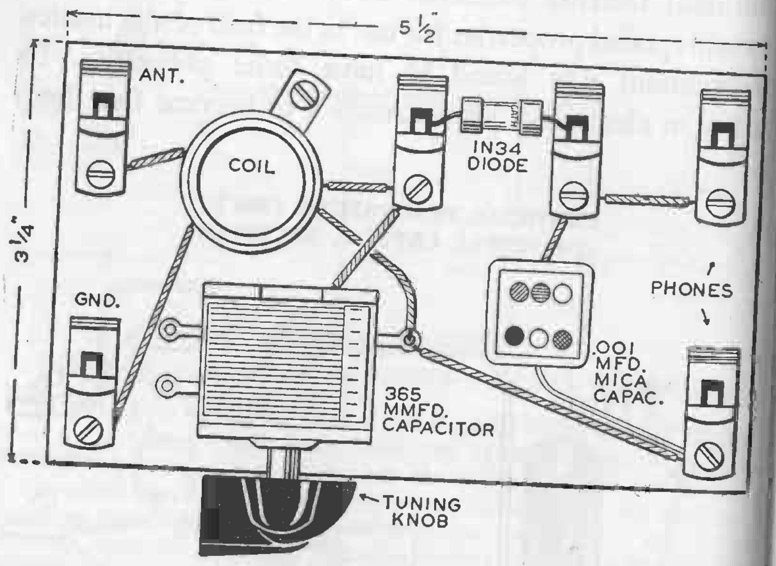

PLAN OF SIMPLE CRYSTAL RECEIVER

This shows the position of the components on the wood base and the arrangement of the wiring.

In addition to conductors and non-conductors there are materials called SEMI-CONDUCTORS. A semi-conductor is neither a good conductor like silver or copper, nor a good non-conductor like glass. But under certain conditions germanium acts like copper, and a current will flow through it. Under other conditions it acts like glass and is practically a non-conductor.

One of the useful properties of a semi-conductor is its ability to permit the flow of electricity in one direction and to resist its flow in the other direction. It acts as a sort of automatic electrical valve, opening to electrons passing in one direction and closing to those which try to move the opposite way. Germanium, selenium and silicon, when combined with certain impurities, are the most useful semi-conductors. There are other materials which possess similar properties but they are not used as extensively as the three mentioned.

The first use for semi-conductors in electronics was as RECTIFIERS—devices which change alternating current into direct current. The galena or silicon crystals with an adjustable "cat's whisker," employed as detectors in the early days of radio, represent one of the earliest uses of a semi-conductor as a rectifier *in electronics.* Selenium has long been used as a semi-conductor in power rectifiers and in "self-generating" photocells.

The first use for germanium was in the construction of tiny electronic rectifiers called DIODES, and used to replace the diode vacuum tubes used as detectors in radar, radio and television receivers. A variety of germanium crystal diodes and also silicon crystal diodes are produced by several manufacturers. They range from 65 cents to $11.00 in retail price. They are small objects, averaging about ½ inch long and ¼ inch in diameter. The construction of all varieties is quite similar and in general consists of a tiny crystal of germanium or silicon to which is welded a small platinum "cat-whisker" wire. The crystal and whisker are enclosed in a small ceramic tube or hermetically sealed in moisture-proof glass cartridges. The germanium or silicon crystal is the cathode and the contact point wire is the anode of the device. A germanium diode is not only much smaller than a vacuum tube diode, which performs the same service, but there are no parts which wear or burn out soon in normal use. A germanium diode has a life about ten times as long as the average life of a vacuum tube diode. A further advantage of the germanium diode is that, unlike a vacuum tube diode, it requires no filament or heater current for its operation.

No. 1N34 Diode. The Sylvania No. 1N34 general purpose germanium diode is an excellent detector. It retails for about 80 cents and can be used for several purposes in electronic circuits.

Although germanium is a costly metal, the amount used in each diode is so small that it does not greatly increase the cost of a diode. The small speck of mirror-surfaced germanium which is used in each diode is only 0.02 of an inch thick and is not worth more than two cents.

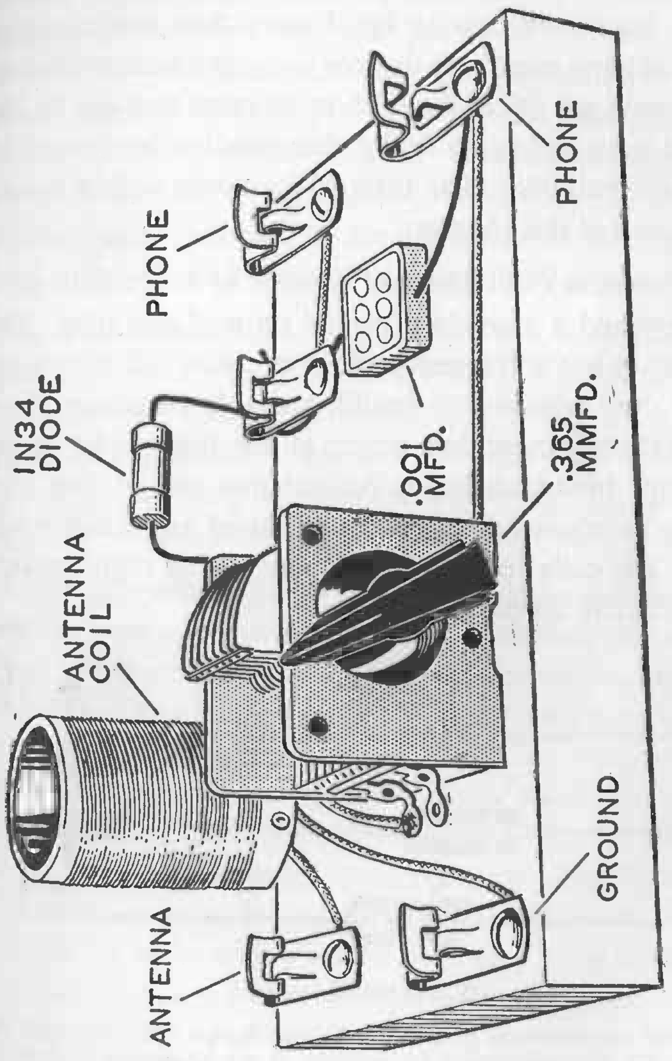

PERSPECTIVE SKETCH OF THE COMPLETED CRYSTAL RECEIVER

Several small paper-covered books have been published by engineers who helped develop the germanium diode. These books tell how to build more than one hundred electrical and electronic devices using germanium diodes. The books are priced from 25 to 50 cents and can be obtained from any radio dealer that handles the diodes, or from the publisher. The titles of the books will be found at the end of this chapter.

The Antenna Coil. The coil consists of two windings, a primary and a secondary wound on a plastic tube. The secondary has a frequency range of 540-1750 kilocycles when used with a 365 mmfd. variable capacitor. This range slightly more than covers all the frequencies in the standard broadcast band. An antenna coil of this type can be purchased at small cost at almost any radio repair shop. The coils are kept in stock for use as replacements for defective coils.

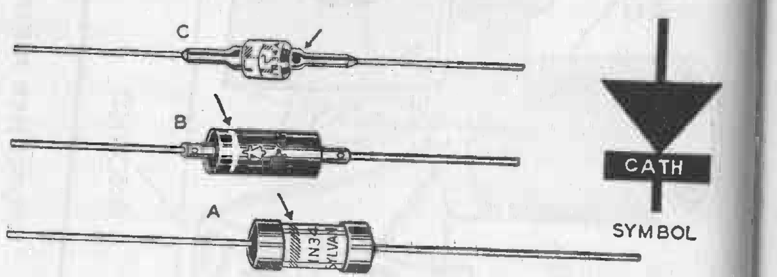

GERMANIUM DIODES

Several manufacturers produce germanium diodes. The 1N34 (A) and 1N34A (C) illustrated above are products of the Electronics Division of Sylvania Electric Products, Inc. The arrows in the illustration point to the markings which indicate the cathode end of the diode. On each Sylvania germanium diode the cathode terminal is indicated by a green band and the letters CATH.

Radio Headset. A radio headset consists of two earphones, or telephone receivers, mounted on a headband. The telephone receivers made for radio use are more sensitive to very small currents than the common telephone receiver. Two types of telephone receivers are used in radio. Do not use "crystal" type headphones with a crystal detector. The headset connected to the Simple Crystal Receiver should be the electromagnetic type and have a resistance rating of at least 2000 ohms.

Mica Capacitor. This component of the circuit is not always required. Some headsets have enough of the quality called "distributed capacitance" to make the mica capacitor unnecessary. In some instances the use of the mica capacitor may not improve the strength of the signals heard in the headset but will improve their tone.

Variable Capacitor. The common 365 single gang mmfd. variable capacitors usually have a ¼ inch diameter shaft. When purchasing the capacitor, obtain also a molded knob which can be attached to the shaft with a setscrew. Some variable capacitors are so made that they can be mounted on a base or against a panel. A capacitor for base mounting has threaded holes in the bottom of the frame. The threaded holes are in the front of the frame in capacitors for panel mounting. If the capacitor which you use is of the latter type it will be necessary to mount it on the back of a small ¼-inch plywood panel attached to the front of the wood base.

ASSEMBLING THE SIMPLE CRYSTAL RECEIVER

The base should be a piece of well seasoned wood which has been shellacked or varnished on all surfaces. Mount the components as shown in the plan and perspective sketch. There are four terminals on the antenna coil, and connection is made to them by soldering. Use rosin-core solder NOT acid-core solder. The short coil is the primary. The long coil is the secondary.

Use both the schematic and pictorial circuit diagram when wiring the receiver, and check each connection carefully to make sure it is correct.

Operating the Receiver. Connect a high-resistance (1000 ohms or more) earphone or a radio headset to the PHONE terminals. Connect an antenna to the ANTENNA terminal and a ground to the GROUND terminal. The antenna should be a wire 50 to 100 feet long and be insulated from its supports. Antennas and grounds are described in more detail in the chapter called "Taking the Mystery Out of Tuning."

Under ordinary conditions the range of the receiver will not be over 25 to 30 miles when listening to a broadcast station. The range of the receiver and the strength of its signals will depend to a great extent upon the height and length of the antenna. If the receiver is located in a metropolitan area where several broadcasting stations are nearby, it may "pick up" more than one program at a time. If two stations are heard at the same time, adjusting the variable capacitor will increase the strength of the signals which you prefer to listen to and weaken those which are causing interference.

All in all, the "simple crystal receiver" is not an entirely satisfactory design. There is no provision for adjusting

or tuning the antenna circuit. It has been described here because of its simplicity and because there are no home-made parts required except the wood base.

If you are willing to wind your own coils, build the receiver described in the next chapter. It is much more sensitive and selective. With it, you can perform experiments which will demonstrate the principles of tuning radio circuits.

BOOKS ABOUT TRANSISTORS

Electronic Shortcuts for Hobbyists, Sylvania Electric Products, Inc., 1740 Broadway, New York 19, N. Y. 25 cents.

Forty Uses for Germanium Diodes, Sylvania Electric Products, Inc., 1740 Broadway, New York 19, N. Y. 25 cents.

Transistor Manual, General Electric Co., Electronics Park, Syracuse, N. Y. 50 cents.

Transistor Applications, Raytheon Mfg. Co., Newton 58, Mass. 50 cents.

Popular Electronics, Ziff-Davis Publishing Co., 366 Madison Avenue, New York 17, N. Y. 35 cents. A magazine published monthly for electronic experimenters.

CHAPTER THREE

How to Build
A More Selective Crystal Receiver

The crystal receiver described in the last chapter is popular because all the parts can be purchased ready-made. It is used in many high schools to demonstrate elementary principles of a radio receiver.

Unfortunately, the advantage of being able to assemble quickly a working receiver sacrifices sensitivity and selectivity. You can build a much better receiver if you wind a special coil to be used in place of a standard antenna coil which, after all, was not originally designed for use in a crystal receiver.

The following parts and materials are required:

1	Shellacked wood base, 10 in. x 4 in. x ¾ in. (1)
1	Shellacked pine coil form 10 in. x 1½ in. x ¾ in. (2)

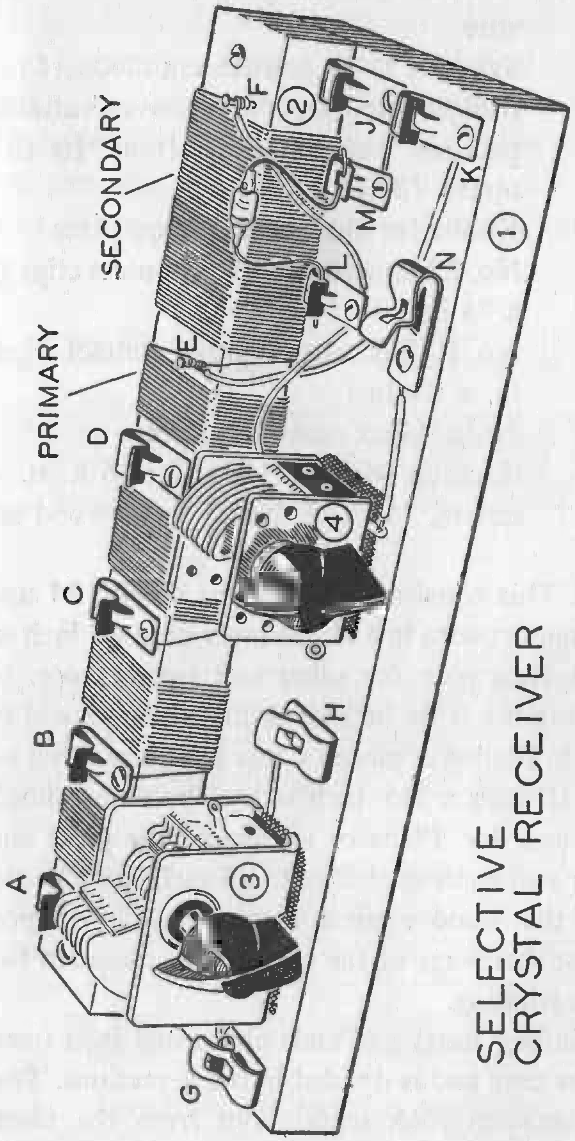

A MORE SELECTIVE CRYSTAL RECEIVER

100	feet No. 24 B.S. gauge enamelled magnet wire
1	Sylvania 1N34 germanium diode (5)
2	Midget one-gang TRF receiver variable capacitors with capacity from 10 to 365 mmfd. (3 and 4)
2	Knobs for the variable capacitors
9	No. 2 Fahnestock spring contact clips (1 in. x ⅜ in.)
2	No. 10 Fahnestock spring contact clips (¾ in. x 5⁄16 in.)
2	¾ in. brass escutcheon pins
	Hook-up wire, 2-1¼-in. No. 6 R. H. wood screws, 11-½-in. No. 5 R. H. wood screws

The Coil. This consists of 245 turns of No. 24 B.S. enamelled magnet wire in a single layer on a 10 inch x 1½ inch x ¾ inch pine (or other soft wood) core. Much common lumber is ¾ inch thick and there should be no difficulty in finding a piece of this thickness from which to saw a 10 inch x 1½ inch strip. The wood should be seasoned and dry. Plane or sandpaper the piece smooth and apply two coats of shellac to all surfaces. The shellac will seal the wood against moisture, help to prevent swelling or shrinkage of the core and consequent loosening of the winding.

The winding starts and ends about one inch from the ends of the core and is divided into five sections. There is a space between each section but from the electrical standpoint the winding is continuous. There is no break in the wire between sections. The spaces between the sections and at the ends of the core provide the room for "taps." Four of the taps are No. 2 Fahnestock connector clips so that a wire can be easily and quickly connected to any one of them and disconnected with equal facility. Two of the taps are brass escutcheon pins. Escutcheon pins are round headed brass nails obtainable at a hardware store and used for fastening small name

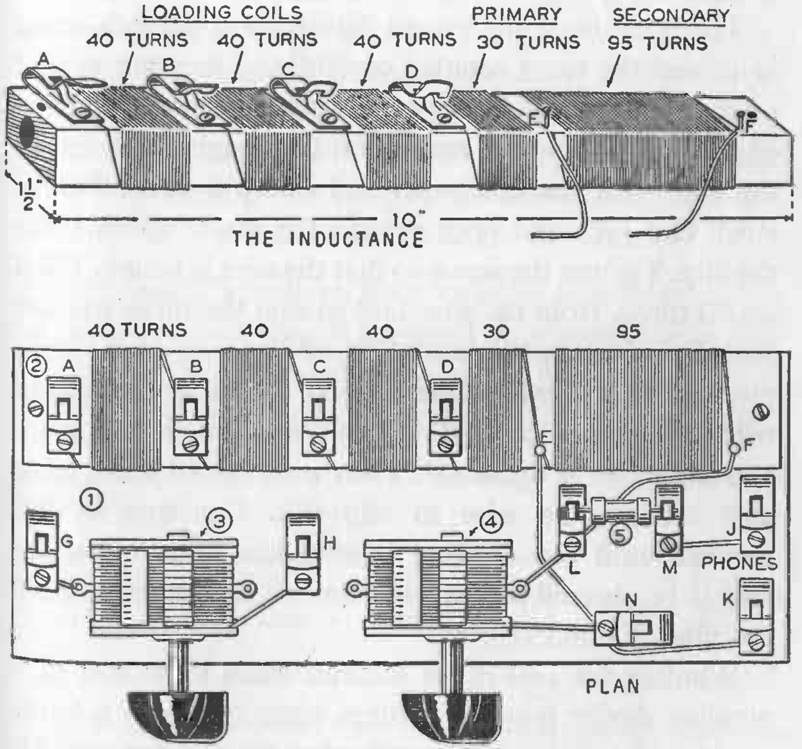

PLAN OF THE CRYSTAL RECEIVER AND DETAILS OF THE TUNING INDUCTANCE

Note the letters identifying each contact clip and the two pins. These letters also appear in the circuit diagram.

plates. The connections made to the brass pins are permanently soldered. The Fahnestock clips are fastened to the wood core with ½-inch No. 5 round head wood screws. The escutcheon pins are driven into the core.

The first three sections of the winding each contain 40 turns of wire and comprise the ANTENNA LOADING COIL. The fourth section consists of 30 turns and is the PRIMARY. The last section contains 95 turns and is the SECONDARY.

The wire must be wound tightly in a smooth, even layer and the turns counted carefully as they are wound on so that there will be the correct number in each section of the coil. Scrape the enamel off the "beginning" end of the wire with fine sandpaper and clamp it under Fahnestock clip (A) and once around the screw which holds the clip. Tighten the screw so that the wire is secure. Wind on 40 turns. Hold the wire taut so that the turns will not loosen, and clean the insulation off the wire at the right place so that it will make a good electrical connection with Fahnestock clip (B) when placed under the latter and the screw is tightened. Then wind on 40 more turns and connect the wire to clip (C). Continue in this manner until the winding is complete. The insulation should be cleaned off the wire where it is wrapped around the pins (E) and (F).

Winding the coil is not difficult when some sort of a winding device is used. Perhaps some one with a lathe-equipped home workshop will wind the coil for you. Or perhaps you can do this yourself on a lathe in the shop at your school. If a metal-turning lathe is used, one end of the wood core is placed in a four-jaw chuck on the

headstock of the lathe. The other end is supported on the center in the tailstock. The lathe should be turned slowly by hand by pulling on the belt. If a wood-turning lathe is used, the wood core is mounted between the head and tail centers as if it were a piece of wood to be turned. Turn the lathe by hand by pulling on the belt. Do not use power. Those who are not fortunate enough to have a lathe available should build the winding machine described at the end of this chapter. It will not require much time to make and the machine will make winding the coil easy.

When the winder is in use it should be held firmly to the top of work bench or table with a C clamp. Coil forms of various lengths can be wound with this same winder by moving the uprights so that the distance between them is 1/16 inch longer than the coil form.

The Variable Capacitors. Both variable capacitors should be the type which can be mounted on a base and be provided with two threaded holes in the bottom of the capacitor frame for that purpose. If the threaded holes are

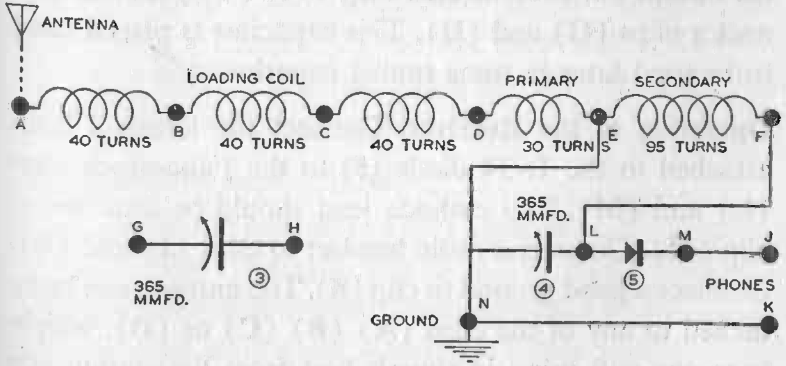

CIRCUIT DIAGRAM OF THE CRYSTAL RECEIVER

in the front of the frame, the capacitor is designed for panel mounting. In that case it will be necessary to attach a small plywood panel (6 in. x 3 in. x ¼ in.) to the front edge of the wood base and mount the capacitors on the back of the panel.

Fixed Capacitor. You may notice that an 0.001 mfd. fixed capacitor to be connected across the headset is not included in the list of parts. Such a capacitor is usually shown in circuit diagrams of crystal receivers. If a standard radio headset is connected to the receiver the 0.001 mfd. capacitor is not necessary. The purpose of the capacitor is to provide a path for radio-frequency currents around the headset. There is sufficient capacitance in the windings and connecting cord of a standard radio headset to provide such a path.

Assembling. Assemble the parts on the base according to the plan in one of the illustrations. To connect the parts, consult the wiring diagram. Notice that the variable capacitor (**3**) to the left on the base is not included in the circuit but is connected only to two Fahnestock connector clips (**G**) and (**H**). This capacitor is placed there to be used later in some tuning experiments.

Operation of the Receiver. Connect the terminal leads attached to the 1N34 diode (**5**) to the Fahnestock clips (**L**) and (**M**). The cathode lead should be attached to clip (**M**). Connect a radio headset to clips (**J**) and (**K**). Connect a good ground to clip (**K**). The antenna can be attached to any of the clips (**A**) (**B**) (**C**) or (**D**), whichever one will bring in signals best from the station you wish to listen to. For instructions regarding the antenna

and ground connection consult the chapter devoted to those subjects. Adjusting the variable capacitor (**4**) will tune the secondary winding and enable you to tune out unwanted signals.

With this receiver you should be able to hear more stations and get louder signals without interference than is possible with the simple receiver described in the previous chapter. The primary of the antenna coil used with that receiver does not contain enough turns to tune in signals efficiently over the entire broadcast band when used with a crystal detector. The coils (**A**) (**B**) (**C**) and (**D**) provided in the more selective receiver are loading coils. They enable you to tune or adjust the antenna circuit to suit the length of the waves which produce the signals you wish to listen to.

Perhaps you will understand how the loading coil improves the receiver after you have read the next chapter.

HOW TO BUILD A WINDING MACHINE FOR WINDING THE TUNING INDUCTANCE

The materials required are several pieces of wood, screws, ½-inch dowel rod, a piece of wire from a coat hanger and a clamp. A handsaw, screwdriver, ½-inch auger bit, brace, hand drill and a No. 28 twist drill will be needed for making the parts.

Parts Required for the Winder

1	Wood base 8 in. x 3¾ in. x ¾ in.
2	Wood uprights 4½ in. x 2¼ in. x ¾ in.
2	Wood battens 3½ in. x 1¼ in. x ¾ in.
1	Piece of steel wire from a coat hanger
8	No. 6 wood screws 1¼ in. long

Cut the wooden parts to the dimensions shown in the illustration. Give them one or two coats of shellac or varnish. Bore a ½-inch diameter hole through both uprights on their center line and 3½ inches from the lower end.

Bore two holes in each upright with a No. 28 twist drill. The holes should be ⅜ inch from the end, and at the end opposite the ½-inch hole. Slip a No. 6 wood screw through each of the small holes in the uprights and fasten each upright to a batten. Fasten each batten to the base, using two screws. The battens should be located so that the distance between the uprights is 1/16 inch longer than the wood core of the inductance. The ½-inch holes in the uprights are bearings and must be in line so that a piece of ½-inch diameter dowel will pass through both holes and turn without binding. To align the holes, first fasten one upright and batten to the base. Then slip a long piece of ½-inch dowel through the bearing holes in both uprights, and fasten the second batten and upright to the base at the proper distance. Assembling the winder with a long dowel in the bearing holes should bring the holes in alignment.

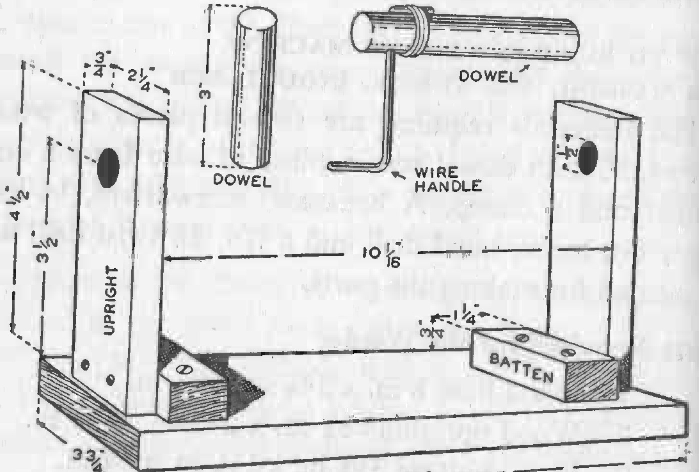

PARTS FOR THE COIL WINDER

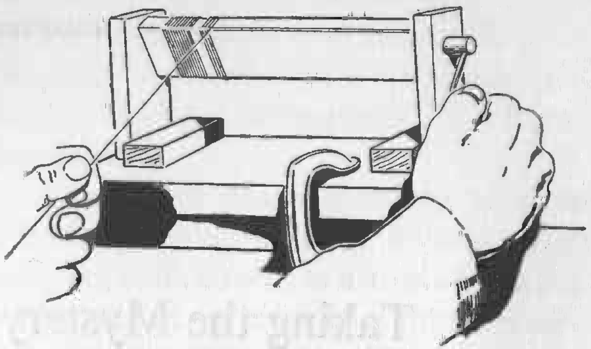

WINDING THE TUNING INDUCTANCE

Two pieces of ½-inch dowel 3 inches long are used as a shaft for the wood core which is to be wound. Bore a ½-inch diameter hole in each end of the core along its axis. Make the holes 1½ inches deep. A piece of steel wire (from a coat hanger) is slipped through a hole near the end of one of the 3-inch dowels and bent to form a crank. This wire handle is used to turn the core when winding it. The two pieces of dowel used as a shaft should fit tightly into the holes in the end of the core. It may be necessary to wrap them with a layer or two of paper to make a snug fit. The dowel shafts must turn freely in the bearing holes in the uprights. If they are too tight in the bearing holes, free them by sandpapering the dowels to make them slightly smaller.

CHAPTER FOUR

Taking the Mystery Out of Tuning

Hundreds of radio stations are transmitting at the same time. Many of them are in the same neighborhood. They transmit at many different frequencies. For example, broadcasting stations transmit over a "band" ranging from 550 kilocycles to 1700 kilocycles. This means that the alternating currents produced at broadcast transmitters and fed into the antenna alternate from 550,000 to 1,700,000 times per second. Each transmitting station sends at a frequency which has been assigned to it by the Federal Communications Commission. Each transmitter is equipped with automatic controlling apparatus which holds it within narrow limits to the frequency assigned to it.

Every receiving antenna is constantly intercepting waves of many different frequencies, but with a selective receiver signals from only one transmitter can be heard at a time. A process called TUNING makes this possible.

Tuning a receiver adjusts it so that it selects signals of one frequency only and rejects the signals of all other frequencies. Tuning is accomplished by adjustable coils and capacitors.

A radio engineer thinks of tuning in mathematical terms. By using analogies instead of technical terms and by comparing radio circuits to a rope clothes-line, tuning can be freed of some of its mystery for the novice.

EXPERIMENTS IN "TUNING" A ROPE

Try these experiments. Hang a light rope or clothesline between two supports 30 to 40 feet apart. The rope

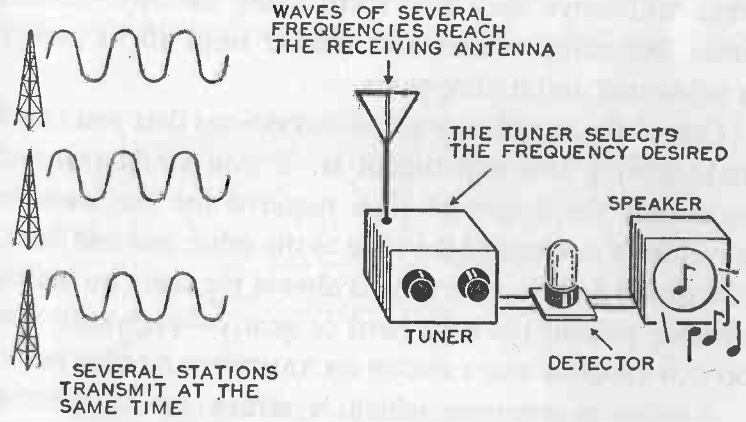
WAVES OF SEVERAL FREQUENCIES REACH THE RECEIVING ANTENNA
THE TUNER SELECTS THE FREQUENCY DESIRED
SPEAKER
TUNER
DETECTOR
SEVERAL STATIONS TRANSMIT AT THE SAME TIME

TUNING

An agreement among the nations of the world requires all radio transmitters to be licensed. Under this agreement, each transmitter is licensed for operation only on the frequency (or frequencies) assigned to it. Consequently it is necessary to adjust or TUNE a transmitter so that it will produce waves only of the frequency or frequencies assigned. Tuning a receiver strengthens the desired signals and bars out signals from other stations which are unwanted at the time.

should not be stretched taut or allowed to hang very slack.

The rope can be made to behave in a mechanical manner somewhat as a radio circuit behaves in an electrical manner, and thereby illustrate some of the principles of tuning.

With your forefinger or a stick, strike the rope a sharp, quick blow near one end. If the rope is properly hung and you use the right amount of force, a visible wave in the rope will move along its length. Some of the energy in your arm has been transferred to the rope. The energy travels along the rope to the opposite end in the form of a wave. When it reaches the end of the rope the wave reverses its motion and travels back toward you. The wave will move back and forth along the rope several times, becoming weaker and weaker until all its energy is exhausted and it disappears.

One of the most important observations that you could make during this experiment is: if you could measure accurately the length of time required for the wave to travel from one end of the rope to the other end and back, you would find that the time is always the same no matter whether you hit the rope hard or gently—PROVIDED YOU DO NOT CHANGE THE LENGTH OR TAUTNESS OF THE ROPE.

A wire or antenna which is struck by an electromagnetic (radio) wave from a transmitter can be likened to the rope struck by your hand. Energy runs along the wire or antenna, not in a visible wave which can be seen, but as a moving group of invisible electrons called a pulse of electric current. The electrons travel along the wire or antenna to one end and then reverse their direction.

They travel back and forth until the energy that moves them is exhausted. Since the electrons reverse their motion they are an ALTERNATING electric current.

Now try another experiment. Tighten the rope slightly so that there is less slack. Tap it again once near one end. A wave will travel from one end of the rope to the other and back as it did before but this time it will travel MORE RAPIDLY. Shortening the rope will produce the same effect as tightening it. The wave will move from one end to the other and back in less time. Lengthening or loosening the rope will INCREASE the time required for the wave to travel back and forth. By this experiment you have demonstrated a method of "tuning" a rope.

A wire or an antenna can be ELECTRICALLY shortened, lengthened, tightened or loosened just as a rope can be MECHANICALLY shortened, lengthened, tightened or loos-

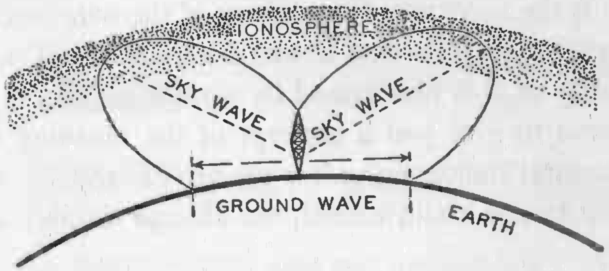

RADIO WAVES—GROUND WAVES AND SKY WAVES

Normally a vertical antenna is more efficient for transmitting radio waves than a horizontal antenna. When a radio wave leaves a vertical antenna, if it were visible, its pattern would resemble a huge doughnut lying on the ground with the antenna in the hole at the center. Part of the wave moves away from the antenna but remains in contact with the earth and is called the GROUND WAVE. The rest of the wave moves upward and outward to form the SKY WAVE. It is the surface part of the ground wave that reaches most receivers tuned to daytime broadcast programs.

ened. In other words, it can be TUNED. If a wire or circuit is either shortened or electrically "tightened," alternating electric currents will travel back and forth in it in less time than before. From a radio or electrical standpoint, the currents will alternate at HIGHER FREQUENCY. Lengthening or electrically "loosening" a wire or circuit will produce the opposite effect. This will increase the time required for an alternating current to travel back and forth therein. The currents will alternate at a LOWER FREQUENCY.

NATURAL FREQUENCY

The number of little waves that run back and forth along a rope in a second are called the NATURAL FREQUENCY of the rope. The number of times that an alternating electric current will surge back and forth along a wire or through a circuit NATURALLY, that is, without being forced is the NATURAL FREQUENCY of the wire or circuit. This explanation is not a literal definition of natural frequency as it is understood by a radio engineer, but it will serve to give you a concept of the meaning of the term natural frequency as it is used in electronics. Whenever we tune a radio circuit, we change its natural frequency.

Changing the physical length of a wire or circuit, that is its length in feet and inches, changes its natural frequency. There are other ways of changing the natural frequency of a wire or circuit besides changing its physical length. It may be ELECTRICALLY tightened or shortened, or lengthened or loosened by changing its INDUCTANCE and CAPACITANCE. A wire wound in the form of a coil

has a different "tension" and a different "electrical length" from the same piece of wire when it is stretched out straight.

ANOTHER EXPERIMENT WITH THE ROPE

Let us consider the rope again. Suppose, for the moment, that your finger is an electromagnetic wave carrying energy from a transmitting station and that the rope is an antenna connected to a radio receiver. You are the transmitter. Your frequency is three. You tap the rope three times per second. At the first tap the energy induced in the rope surges outward toward the far end. It reaches

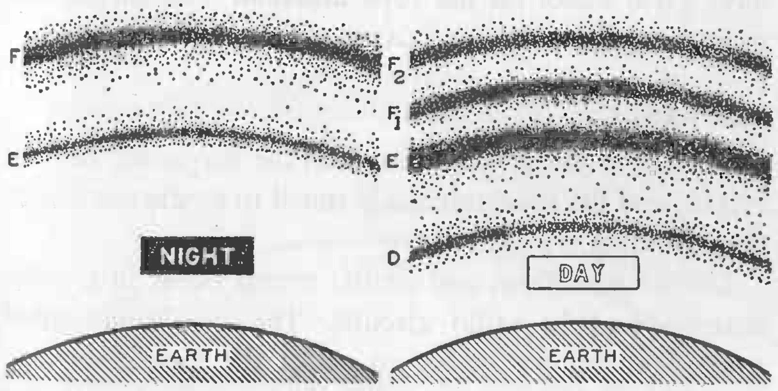

RADIO WAVES AND THE IONOSPHERE

The rarefied atmosphere approximately 40 to 350 miles above the earth contains many more positive and negative ions than the lower atmosphere. This ionized layer is called the IONOSPHERE. It has a marked effect upon the transmission and reception of radio signals. The ions in the ionosphere are produced by ultra violet and particle radiation from the sun. The ionosphere is thicker and closer to the earth in the daytime than at night. Consequently sunshine limits radio transmission. Radio transmission and reception are better at night than in daytime. Much long distance radio transmission is accomplished by receiving sky waves reflected from the ionosphere. The reflected waves come back to earth closer to the transmitter in daytime than at night (*See next illustration*).

the far end and is returning when it meets the second outward traveling surge induced by the second tap of your finger. One surge cancels another when they meet coming from opposite directions. Your waves (taps) do not have much effect on the rope antenna after the first surge because the antenna is NOT IN TUNE with the frequency of the surges.

But wait a moment. Suppose that you lengthen, shorten, loosen or tighten the rope. In other words, tune the rope so that each new outgoing surge set up by the three taps per second of your finger falls in step with the one already travelling back and forth on the rope. Your waves will have great effect on the rope antenna. The surges will move to and fro, or OSCILLATE, without interference.

RESONANCE

A radio engineer would say that the surges are now IN PHASE, and the rope antenna is tuned to RESONANCE with the incoming waves.

Similar conditions and similar events occur in a radio antenna and in radio circuits. The condition called resonance enables tuning to select the signals of one station from the others which are transmitting at the same time. In order for incoming waves to have the most effect in producing alternating currents in a receiving antenna or in radio circuits, the antenna and circuits must be tuned until they are in resonance wih the incoming waves. They must be tuned to suit the frequency of the waves coming from the station to which it is desired to listen.

Every receiving antenna includes a small coil which

couples it to the detector and other receiver parts. A coil is also used to couple transmitting apparatus to its antenna. These antenna coupling coils increase the electrical length of the antenna. A capacitor placed in series with the coil shortens the electrical length of the antenna. These two devices, a coil and a capacitor, are a simple means of increasing or decreasing the ELECTRICAL LENGTH of an antenna or radio circuit or bringing it into resonance with the frequency of the alternating currents which surge in it. In actual practice that is the way most tuning is done. A coil and a capacitor connected to a tuning knob tune the antenna and other circuits.

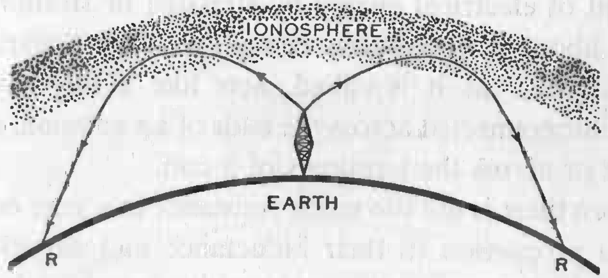

RADIO WAVES—REFLECTION OF THE SKY WAVES BY THE IONOSPHERE

The ionosphere acts as a conductor and absorbs part of the energy in the sky waves radiated from an antenna. It also acts as a radio mirror and bends the sky wave back to the earth as illustrated above. Receivers located at or near either of the points **R** in the illustration will receive signals transmitted from the antenna at point **T**. This explains why low-powered amateur stations can sometimes be heard at night halfway around the world. Their sky waves are bent by the ionosphere and returned to earth in a distant land. It also partly explains the fading of signals. The ionosphere is constantly shifting and changing. Occasionally sudden disturbances in the ionosphere cause absorption of all sky-wave radiations. At other times the reflected sky waves alternately strike and skip over areas.

EXPLANATION OF INDUCTANCE AND CAPACITANCE

One of the first things to be learned about electricity is that in order for a direct current to flow in a circuit, the circuit must be complete—that is, the wire must form a loop so as to provide a return path. How then can an alternating current flow back and forth in a straight wire without a return path, as it does in an antenna?

Here is the explanation. A single wire stretched out straight or in the form of an open coil has two remarkable properties. One of these is its INDUCTANCE. The inductance of a wire or a coil is the property that allows it to store up energy in the form of an electromagnetic field. The other is its CAPACITY or ability to store up a certain amount of electrical energy by stressing or straining the space about it or substances nearby. This "distributed capacitance," as it is called, acts like a tiny invisible capacitor connected across the ends of an antenna, across a wire or across the terminals of a coil.

When there is not too much resistance in a wire or open coil in proportion to their inductance and capacitance, radio-frequency currents can surge back and forth in the wire or coil because of the distributed capacitance present. Low-frequency or direct currents can flow in an open circuit for the merest infinitesimal fraction of a second only while the open coil or wire is storing up energy in electromagnetic or electrostatic form.

The antenna is not the only part of a radio transmitter or receiver that must be tuned. Other circuits in both transmitters and receivers must be adjusted until they are in resonance. The transmitter circuits are tuned to be in

resonance with the frequency at which IT IS DESIRED TO TRANSMIT. The receiving circuits are tuned to resonance with the frequency of the transmitter whose signals IT IS DESIRED TO RECEIVE.

When an antenna or radio circuit is tuned to resonance with any certain frequency, for example 550 kilocycles, the maximum amount of 550 kilocycle current will flow in the circuit. Very little, if any, current of other frequencies will flow.

Tuning therefore, is ADJUSTING CIRCUITS TO RESONANCE at a desired frequency so that the circuits give the maximum response to currents of that frequency.

Tuning can be done by varying the inductance of a

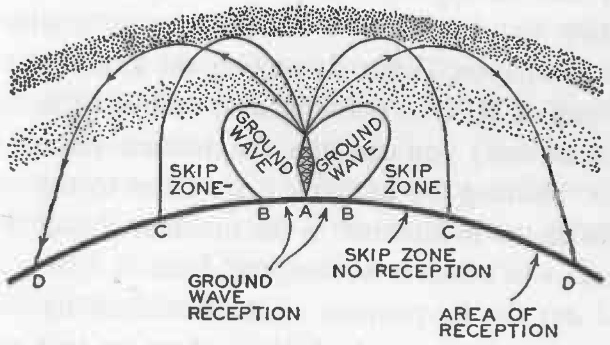

RADIO WAVES—SKIP ZONES

Signals from a transmitter can always be heard over the considerable area (**A** to **B**) surrounding the transmitter where the ground waves reach. Surrounding this area is a SKIP ZONE where there are no ground waves and where the transmitter's signals cannot be heard. Beyond the skip zone there are distant points where the signals can be heard. The illustration helps to explain this. The skip zone is the area beyond the limits of the ground waves. The area beyond the skip zone where signals can be heard again is the area where the sky waves reflected from the ionosphere return to earth (**D**).

coil. Usually it is easier to do it with a capacitor and most tuning is done with a variable capacitor.

The majority of modern receivers employ the superheterodyne circuit invented by E. H. Armstrong. In this circuit, there are tuned radio-frequency amplifiers. They consist of electron tubes, coils and variable capacitors. There is more than one tuned circuit in a superheterodyne receiver. For convenience in tuning, so that there is only one knob to turn, the rotors of two or more identical variable capacitors are mounted on one shaft. Such capacitors are called ganged capacitors.

EXPERIMENTS IN TUNING A CRYSTAL RECEIVER

Now that tuning has been explained it will be more interesting to experiment with the crystal receiver described in Chapter Three. By moving the antenna connection to clip **A**, **B**, **C** or **D** (whichever will bring in the station the loudest) you can tune the antenna circuit of the receiver. Shifting the antenna connection to the various clips varies the inductance in the antenna circuit and the frequencies to which it will respond best.

Find out the frequencies of the broadcasting stations which you can tune in, by looking them up in a printed radio program. Listen to each station until its name is announced. Then jot down its call letters and frequency. You will find that by tuning the antenna circuit you can bring in more frequencies than is possible with the simple receiver with an untuned antenna circuit described in Chapter Two.

The variable capacitor marked (**3**) to the left in the

plan was purposely not made a permanent part of the receiver circuit so that it could be used for experiment. The terminals of capacitor (**3**) are connected to the Fahnestock clips **G** and **H**.

Let us suppose that you have tuned in a station which you wish to listen to and get the best signal when the antenna is connected to clip **B**. Perhaps you will get an even better signal if you connect capacitor (**3**) across the primary and that portion of the loading coil which is in the antenna circuit. To do this connect clip **B** to clip **C** with a short piece of wire and connect clip **H** to escutcheon pin **E** with another piece of wire. Adjusting capacitor (**3**) may improve the signals. Perhaps you will get even better signals if you connect the antenna to clip **C** and disconnect clip **G** from clip **B** and connect **G** to **C**.

If the antenna you are using is a very long one, it may be necessary to connect the antenna to clip **G** and connect clip **H** to clip **A** in order to receive the higher frequencies in the broadcast band.

Remember that in order to tune you do not merely shift the capacitor connections. You must also adjust both capacitors by turning the knobs to the position that produces the loudest signals.

CHAPTER FIVE

A Simple Outdoor Receiving Antenna . . . STATIC AND STRAYS

The simplest outdoor antenna for receiving signals from the broadcasting stations which are in the standard frequency range is a single horizontal wire having a lead-in wire connected to one end. A wire 50 to 100 feet long, properly insulated and as high and clear of surrounding objects as possible, with a height of 30 to 50 feet is desirable. It is important to provide rigid supports for both ends of the antenna. Tree branches are not desirable supports because they sway in a wind. When a tree trunk must be used as an antenna support, the antenna wire should hang slack so that the tree can move without straining the wire.

It is important to keep an antenna wire as far away as possible from chimneys, metal roofs, gutters, drain pipes, telephone and power wires, other antennas and tree branches.

The antenna and lead-in should be one continuous piece of wire. If they are not, all splices should be soldered. Use stranded copper antenna wire, obtainable at any radio store. Use a 3-inch glass insulator at each end of the antenna.

The lead-in wire should not approach the side of a building closer than six inches except at the point where it enters. An antenna lead-in strip will insulate the lead-in and bring it into a building without permitting much leakage of energy. The lead-in strip is a thin copper strip covered with waterproof lacquered webbing. It is thin enough so that it may be laid on the sill and the window sash will close upon it.

An antenna on an apartment building or on a city house brings problems not encountered if you live in a detached house in the country or suburbs. Telephone wires over the rooftops, other antennas close by, steel framework in buildings, and man-made static are some of the difficulties to be met in the city. Preferably the antenna

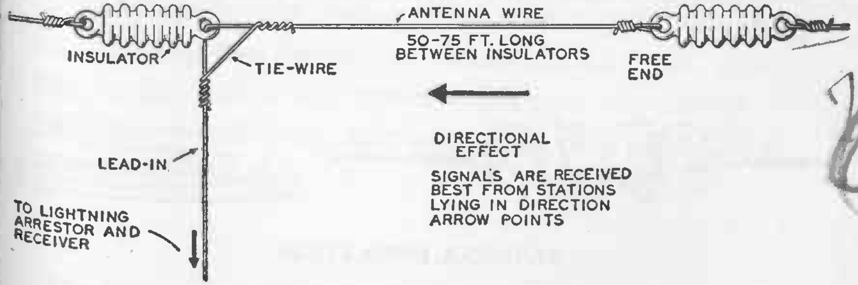
ANTENNA WIRE
50-75 FT. LONG BETWEEN INSULATORS
INSULATOR
TIE-WIRE
FREE END
DIRECTIONAL EFFECT
SIGNALS ARE RECEIVED BEST FROM STATIONS LYING IN DIRECTION ARROW POINTS
LEAD-IN
TO LIGHTNING ARRESTOR AND RECEIVER

AN ANTENNA FOR YOUR HOMEMADE RECEIVER

A single hard drawn No. 14 B.S. copper wire or stranded copper wire 50 to 75 feet long insulated at both ends from its supports.

wire erected on an apartment house or city dwelling should be supported from two wood poles at least ten feet above the roof. The wire, if possible, should be erected at right angles to neighboring wires or antennas.

A horizontal antenna wire whose length is greater than its height above ground will receive signals slightly better in one direction than in others. It will receive best from stations lying in the direction OPPOSITE to which the free end points. The free end is the one opposite to that the lead-in is attached to.

If you can arrange your antenna to point in any desired direction, erect it so that its position will be an advantage in bringing in signals from the station you prefer to listen to.

The single, horizontal wire antenna which has just been discussed is recommended only as the SIMPLEST form of antenna for use with a broadcast band receiver used solely for the reception of comparatively local programs. Receiving antennas of this type were in wide use in the early days of broadcasting. Today they are not considered desirable unless used in a location free of man-made static.

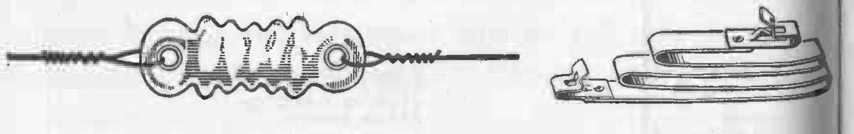

ANTENNA INSULATORS

At the left is a corrugated glass insulator suitable for supporting a 50- to 75-foot receiving antenna. The right-hand sketch shows an insulated lead-in window strip which makes it easy to lead in the antenna over a window sill and under the sash without interfering with the opening or closing of the window.

Static. There are two kinds of objectionable noises called STATIC produced in radio receivers. One is due to nature's electrical disturbances in the atmosphere and the other is "man-made."

Lightning discharges are the cause of the loud crashes and rattling sounds sometimes heard coming from a loudspeaker or headset. Snowflakes often carry minute electrical charges and produce noises in a receiver when they strike the antenna.

Another name for atmospheric electricity which disturbs a radio receiver is STRAYS. Natural static does not affect a frequency-modulated or FM receiver but it cannot be tuned out of an AM receiver. All of the receivers described in this book are the AM type. An FM receiver cannot be assembled by the young novice at radio building.

Man-made static is usually due to induction and sparking by various electrical devices such as elevators, vacuum cleaners, oil burners, sewing machines, electric refrigerators, etc. Apartment houses are fairly alive with these radio disturbers. When a receiver is disturbed the source of the trouble is usually not far away. Nowadays many electrical appliances are equipped with interference prevention devices at the factory. When not equipped, a line filter connected to the power wires near the source of interference will often eliminate man-made static from electrical appliances.

CHAPTER SIX

Components of Electronic Circuits . . . CAPACITORS AND CAPACITANCE • RESISTORS AND CONTROLS

It would be difficult to find an electronic circuit in which there is not at least one capacitor. It is important for the experimenter with electronics to know about capacitors and their quality of CAPACITANCE.

Capacitors were formerly called condensers. But thousands of young men who were trained to handle radio and electronics equipment by the U. S. Navy in World War II learned to use the better and now correct term CAPACITOR. In engineering, the term "condenser" now refers only to certain hydraulic, chemical and engineering devices.

What is a Capacitor? A capacitor is not a complicated device. There are few useful contrivances as simple as a capacitor. It is not much more elaborate than a resistor.

A capacitor consists of two conducting metal plates separated by a layer of insulating material. The conducting plates are called the PLATES and the insulating separator is called the DIELECTRIC. If the plates are separated by air, air is the dielectric. If they are separated by oil, oil is the dielectric. The plates are usually made of sheet aluminum or aluminum foil. The common dielectrics used in capacitors are air, mica, paper, oil and a thin layer of non-conducting aluminum oxide.

The behavior of capacitors when connected to a source of electric current makes them a necessary part of almost every electronic circuit. For example, capacitors will store energy. In addition, when included in a circuit, an efficient capacitor will IN EFFECT allow an alternating current to pass but block a steady direct current. See further explanation on page 69.

How a Capacitor Performs. Let's see what happens when a capacitor is connected to a direct current voltage source, as, for example, to the terminals of a battery. Since the dielectric between the plates is a non-conductor, a capacitor might be expected to behave as a break in the circuit. When connected to a source of steady direct current, it does act as an open circuit except for a very short time

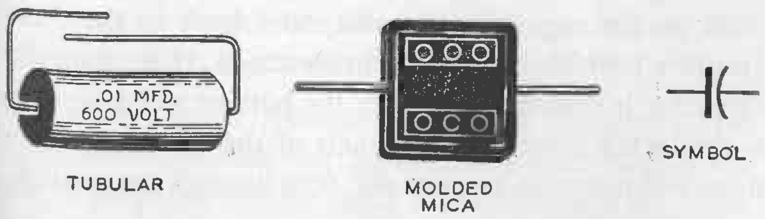

SMALL FIXED CAPACITORS AND THEIR SYMBOLS

The capacitor at the left is a paper by-pass capacitor. That at the right is a molded mica capacitor.

while it is CHARGING. But it does other things as well.

As soon as a capacitor is connected to a source of direct current voltage, electrons start to flow from the plate connected to the positive terminal of the battery. The electrons move through the battery and up to the plate connected to the negative terminal. They collect there. When the capacitor is first connected to the battery, the electrons rush along quickly. Then they slow down. When the charging continues to the point where the voltage across the capacitor terminals equals the voltage across the battery terminals, the electrons no longer move. At this stage the capacitor is fully charged insofar as the voltage of that particular battery is able to charge it. When in this condition, one plate has a positive charge, which is to say it is deficient in electrons. The other plate is negatively charged, that is, has a surplus of electrons. While the electrons were in motion during the charging process, there was a current flow. When the electrons are not moving, there is no current flow.

If the charged capacitor is disconnected from the battery it will remain charged unless the electrons leak across the dielectric or are released by an outside conducting path. Given the opportunity, the excess of electrons on the negative plate will move back to the plate (positive) which is deficient in electrons. If the charged capacitor is disconnected from the battery and a resistor is connected across the terminals of the capacitor, electrons will move—a current will flow through the resistor.

The Meaning of Capacitance. The CAPACITANCE of a capacitor refers to the amount of charge—that is, the

number of electrons required to build up a certain potential difference. Increasing the area of the plates, or decreasing the distance between the plates increases the capacitance. The unit of measurement for capacitance is the FARAD (a shortened form of Faraday in honor of Michael Faraday).

The plates of a capacitor with a capacitance of one farad would be raised to a potential difference of one volt if charged for one second with electrons flowing at the rate of one ampere. When current flows at a rate of one ampere 6 quintillion, 281 quadrillion electrons flow past a given point in the circuit in one second. That is a lot of electrons and you can see what a staggering quantity is involved if you write 6,281 and place fifteen zeros after it like this: 6,281,000,000,000,000,000. A capacitor having a capacitance of ONE FARAD is extremely large and there is little use for so much capacitance in a single unit. The capacitors which are most widely used are much smaller. A more practical unit than the farad for measuring capacitance is the MICROFARAD, this being one millionth of a farad or .000001 farad. The abbreviation for microfarad is mfd. In electronic circuits some of the capacitors are so small that their capacitance is measured in micro-microfarads or millionths of a microfarad. The abbreviation for micro-microfarad is mmfd.

The Effects of the Dielectric. Different dielectrics in a capacitor create different capacitances even though the area of the plates and the space between them is the same. If a capacitor with air as the dielectric has a certain capacity, for example, .001 mfd., the same plates, spaced the same distance apart but with mica as the dielectric will have from 2½ to nearly 7 times the capacitance. In other words, the capacitance will be from .0025 to .0065 mfd. depending upon the variety of mica. Paper substituted for air will increase capacitance from 2 to about 2½ times. The capacitance of .001 mfd. is increased to .002 to .0026 mfd. by paper.

Some dielectrics are stronger than others in respect to the amount of voltage they will withstand without breaking down or puncturing. A mica dielectric .005 inch thick will withstand several times the voltage that would puncture holes in a paper dielectric. Oil substituted for air between the capacitor plates will not only increase the capacitance from 2.2 to 4.7 times but will withstand a much higher voltage before breaking down.

When an electric current flows into a capacitor, a certain amount of energy is lost in the dielectric. Some dielectrics produce greater losses than others.

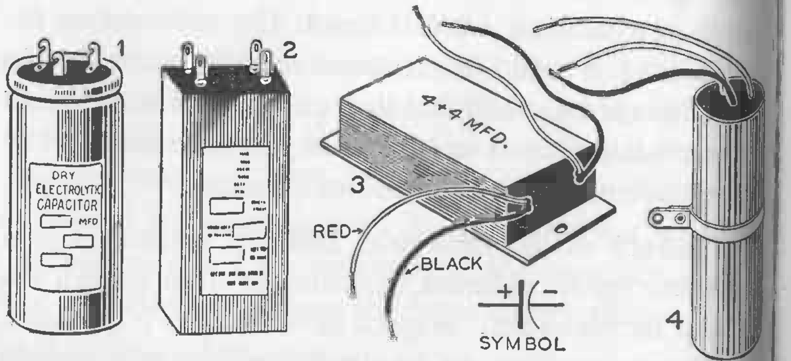

ELECTROLYTIC CAPACITORS

These are fixed capacitors of large capacitance.

For all of these reasons and also because of the requirements of limits in size, weight and cost, capacitors are manufactured in great variety. Those which we will utilize later in carrying out the plans in this book are described here in detail.

Capacitors are an essential part of most electronic circuits. They are also used in automobile ignition equipment, telephone, telegraph and railway signaling apparatus and other electrical devices.

There are two general classes of capacitors used in radio and electronics. They are called FIXED and VARIABLE, accordingly, as the capacity may be varied or not.

FIXED CAPACITORS

Of fixed capacitors there is no end of varieties and sizes. The catalogs of supply houses dealing in electronic components list hundreds of fixed capacitors. Some have mica dielectrics, some have paper, some have oil. They may be enclosed in molded Bakelite, waxed cardboard tubes or sealed in metal cans.

Fixed capacitors may be used to block a direct current from flowing in a circuit and let an alternating current pass. Their purpose in a circuit gives them various names such as BLOCKING CAPACITOR, BY-PASS CAPACITOR, etc.

Capacitors of small capacitance will pass radio-frequency alternating currents but block audio-frequency currents. Capacitors of greater capacitance pass radio-frequency and audio-frequency alternating currents and stop only direct current. By controlling the capacitance of a capacitor, it is possible to permit radio-frequency currents to pass and stop audio-frequency currents that try to flow through the same path. Capacitors used for such purposes are termed BY-PASS capacitors. They are used to by-pass radio-frequency currents in the detector circuits of radio receivers and in amplifiers for radio-frequencies; also to by-pass audio-frequency currents in both detector circuits and audio amplifiers.

Generally speaking, the capacitances used for by-passing radio-frequencies in broadcast receivers usually range from 0.1 to .006 mfd. (depending upon frequency to be passed). Those for by-passing audio frequencies generally range from 0.25 to 1 mfd.

Mica Fixed Condensers. The size and form of small mica fixed condensers varies but the principle of two conducting surfaces separated by a thin layer of mica never changes. Mica fixed condensers are usually enclosed in moulded Bakelite. Two tinned wire terminals are provided. Mica fixed condensers withstand higher voltages and have less energy loss in their operation than paper condensers. They are usually of smaller capacitance than paper condensers. The capacitance may be indicated by color code markings, or numerals showing the actual value may be stamped on the capacitor. If the numerals are a decimal, for example, .002, the capacitance is indicated in microfarads. If it is a whole number, for example, 2000, the figure represents micro-microfarads.

Tubular Paper By-Pass Capacitors. These are cylindrical shaped paper dielectric capacitors with two tinned wire terminals. They are plainly marked to show the capacitance and safe working voltage.

Ceramic Capacitors. The dielectric in this type is a ceramic, that is, special baked clay. The plates are a film of metallic silver baked on the surface of the ceramic. Ceramic capacitors are often used for the same purposes as mica capacitors. They are made only in small capacitances.

Electrolytic Capacitors. This form of capacitor is commonly used in the power supply of electronic circuits where large capacitance is required. An electrolytic capacitor can provide large capacitance in small space. It can also be manufactured at much lower cost than a paper capacitor of equal capacitance.

Electrolytic capacitors may be either wet or "dry." Both types are actually wet inside but the so-called dry type contains no free liquid, whereas the wet type does. The wet capacitor consists of sheets of aluminum foil immersed in a solution of borax in water. The borax solution is an electrolyte. All electrolytes are conductors of electricity. The borax solution serves as one of the conducting surfaces in the capacitor. Chemical action forms a thin film of aluminum oxide on the surface of the aluminum foil immersed in the borax solution. Aluminum oxide is an insulator and the film serves as the dielectric between the aluminum foil and the solution. Wet type electrolytic capacitors are usually enclosed in an aluminum can.

The dry type of electrolytic capacitor also consists of aluminum foil and a solution of borax. However, the electrolyte is contained in wet gauze or paper or in a thick jelly so that there is no free liquid. There is nothing in the dry type to spill.

Electrolytic capacitors are manufactured in many sizes, ranging in capacitance from 1 to 1000 microfarads. The capacitance, polarity and working voltage is usually indicated. The terminals are always marked with a plus and minus sign or the polarity is indicated by red and green or by red and black terminal wires. Red indicates positive; so does a plus sign. The electrolytic capacitors shown in circuit diagrams are always marked to show the location of the positive and negative terminals. The circuit will not function unless the capacitors are connected as indicated. If an electrolytic capacitor is accidentally reversed in a circuit it forms a short circuit and the capacitor usually is ruined. Electrolytic capacitors are manufactured in single, double and triple units—that is, one, two or three capacitors within one case. Double and triple units usually have a common negative terminal.

The working voltage marked on an electrolytic capacitor is important. A capacitor marked 25-v D.C. should be used only in a circuit where the voltage is never above 25 volts. It is well to allow more margin of safety than this. A capacitor which is to be used on 117-v A.C. circuit should be one marked for 150 or 250 working volts.

Although electrolytic capacitors furnish large capacity in small space and at low cost they are inefficient and have a high power loss compared to paper, mica, air and oil dielectric capacitors. The electrolytic film between the aluminum foil and the electrolyte deteriorates with age. Consequently a breakdown in the electrolytic capacitors in commercial radio receivers is one of the most frequent causes of receiver failure. Because of deterioration with

age or while in storage, electrolytic capacitors seldom are used in army and navy equipment.

VARIABLE CAPACITORS

Variable capacitors are used for tuning. They are sometimes called tuning capacitors. The knob on your radio receiver which you turn to tune in different broadcasts is attached to a variable capacitor. The most common type is constructed so that the capacitance may be gradually and continuously increased or decreased by turning a

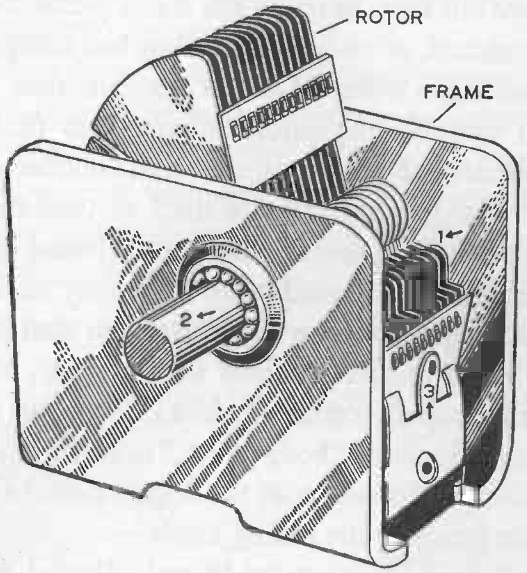

VARIABLE TUNING CAPACITOR

This is the 365 mmfd. variable tuning capacitor used in building the receivers described in this book. It is designed for mounting on a base and has two threaded holes for 6-32 machine screws on the under side. Connection to the frame and rotor is made by a wire placed under the head of one of the mounting screws. The numerals identify: 1, the stator, 2, the shaft and 3, the stator terminal.

knob. It consists of a number of movable, semi-circular aluminum plates which rotate between another group of nonmoving plates so as to interleave. The group of movable plates is called the ROTOR. The group of nonmoving plates is called the STATOR.

The air between the plates is the dielectric. When the movable plates (rotor) are turned so as to mesh with the nonmoving plates (stator) the capacitance increases. The capacitance is greatest when the plates are completely meshed. When the rotor is turned so that the movable plates move out from between the stator plates the capacitance is reduced. A variable capacitor has comparatively small capacitance when the plates are separated.

Several variable capacitors mounted so that all the movable plates are on a common shaft is called a "gang" capacitor. Gang capacitors are used to tune more than one circuit simultaneously. They will be found in modern superheterodyne and tuned radio-frequency receivers.

Variable capacitors are constructed so that the metal frame is electrically connected to the rotor. When the capacitor is properly connected in a circuit, this construction helps to eliminate "body effect," or disturbance of the circuit due to the presence of the fingers close to the rotor shaft while grasping the tuning knob.

Variable capacitors are made and adjusted by experts at the factory. Bending the plates may throw them out of alignment or cause them to short circuit. When handling or working with a variable capacitor in the construction of a project, keep the capacitor closed—that is, plates fully meshed.

Trimmer and Padder Capacitors. The most common variety of this type of capacitor consists of two small metal plates mounted on a ceramic base. A thin sheet of mica between the plates acts as the dielectric. The capacitance can be changed by using a screwdriver to turn a screw which moves the two plates closer together or farther apart. Trimmers and padders are "adjustable" capacitors rather than variable condensers.

Trimmers are usually mounted on the side of variable capacitors for "trimming" or adjusting the capacitors. Padders are commonly used in the oscillator circuit of superheterodyne receivers.

Modern radio receivers employ from two up to ten or even more circuits in order to obtain selectivity. These circuits are useless unless all of them are working at their proper frequencies simultaneously. In order to tune several circuits simultaneously, tuning capacitors with several individual capacitor sections on one shaft are employed.

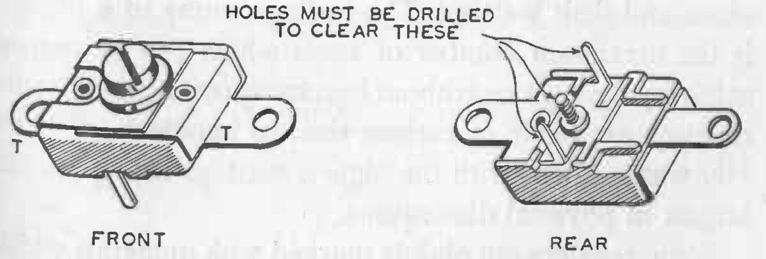

A TRIMMER CAPACITOR

The terminals are marked **T** in the illustration. When the capacitor is mounted on a base or panel, holes must be drilled to provide clearance for the lugs and adjusting screw.

Trimmer capacitors mounted on the variable capacitor and also, in some instances, on the transformers and other coils make it possible to "align" the coils and capacitors so that several circuits can be tuned simultaneously by turning a single tuning control knob.

RESISTORS

Radio men commonly use the term RESISTOR to indicate the small cartridge-shaped fixed resistances which are included in the circuits of radio receivers and amplifiers. A resistor is a device intentionally placed in an electric circuit to provide resistance in the circuit. The term resistor applies also to the various forms of voltage dividers, rheostats, potentiometers, volume controls and ballasts used in electronic circuits. Resistors are utilized in electronic circuits to:

1. Reduce or control the amount of current flowing.
2. Cause differences of potential at certain points.
3. Reduce the voltage applied to a device.

Resistors are rated according to their resistance in ohms and their wattage. The wattage rating of a resistor is the maximum number of watts which can be carried safely by the resistor without becoming overheated. Radio resistors are made in various sizes to handle from ¼ to 100 watts. Those with the highest wattage rating are the largest in physical dimensions.

Some resistors are plainly marked with numerals which show their resistance measured in ohms. Those which are not so marked employ a color code to indicate their resistance. The most common method is to use colored bands.

When the resistor is turned so that the bands are at the left hand end, the first two bands (reading from left to right) indicate the first two figures. The third band is the multiplier, used to multiply the first two figures. If there is no fourth band, the resistance may vary plus or minus one-fifth (20%). A gold fourth stripe indicates a variation of plus or minus one-twentieth (5%) and a silver fourth stripe indicates a variation of plus or minus one-tenth (10%).

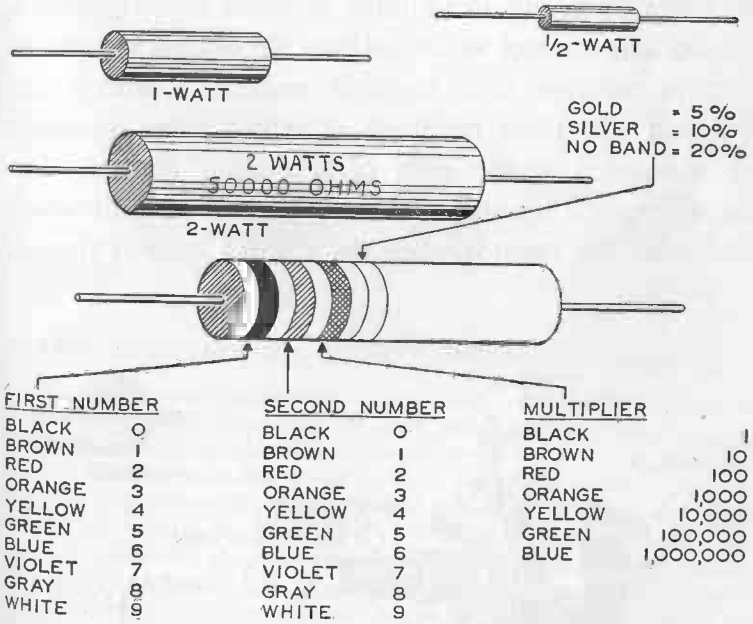

FIRST NUMBER		SECOND NUMBER		MULTIPLIER	
BLACK	0	BLACK	0	BLACK	1
BROWN	1	BROWN	1	BROWN	10
RED	2	RED	2	RED	100
ORANGE	3	ORANGE	3	ORANGE	1,000
YELLOW	4	YELLOW	4	YELLOW	10,000
GREEN	5	GREEN	5	GREEN	100,000
BLUE	6	BLUE	6	BLUE	1,000,000
VIOLET	7	VIOLET	7		
GRAY	8	GRAY	8		
WHITE	9	WHITE	9		

RESISTOR COLOR CODE

To interpret the color bands hold the resistor so that the end bearing the bands is at your left. Each band is code for a numeral. Follow the instructions in the text to determine the resistance. You can identify the capacity of a resistor by its size. The resistors illustrated above are full size.

For example, if a resistor has a blue first band, a gray second band, an orange third band and a silver fourth band, the first number is 6, the second is 8, the multiplier is 1000 and the variation may be plus or minus 10%. Therefore the resistor is a 68,000 ohm unit whose actual resistance in ohms may be somewhere between 61,200 and 74,800 ohms.

CONTROLS

Many of the adjustments required in electronic circuits are made by means of variable resistors called controls. The volume control which adjusts the sound volume of a radio receiver is a variable resistor. Controls are equipped with three terminals. The two outer terminals are connected to the ends of a circular resistor. The center terminal is connected to a movable contact which slides over the resistor when the control knob is turned.

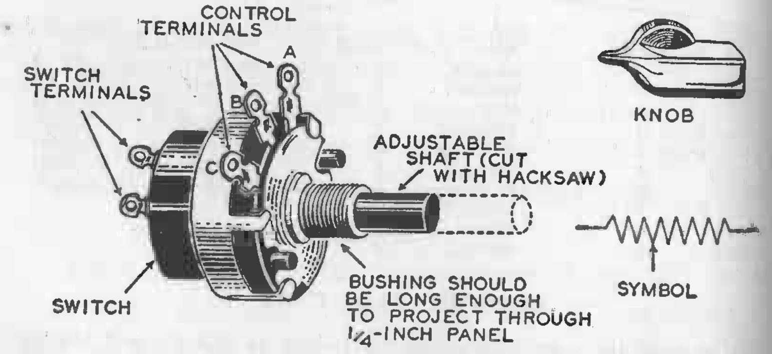
CONTROL TERMINALS
A
B
C
SWITCH TERMINALS
KNOB
ADJUSTABLE SHAFT (CUT WITH HACKSAW)
SWITCH
BUSHING SHOULD BE LONG ENOUGH TO PROJECT THROUGH 1/4-INCH PANEL
SYMBOL

CONTROLS

Controls are resistors whose resistance can be varied by turning a knob. Controls are used in some of the apparatus described in this book.

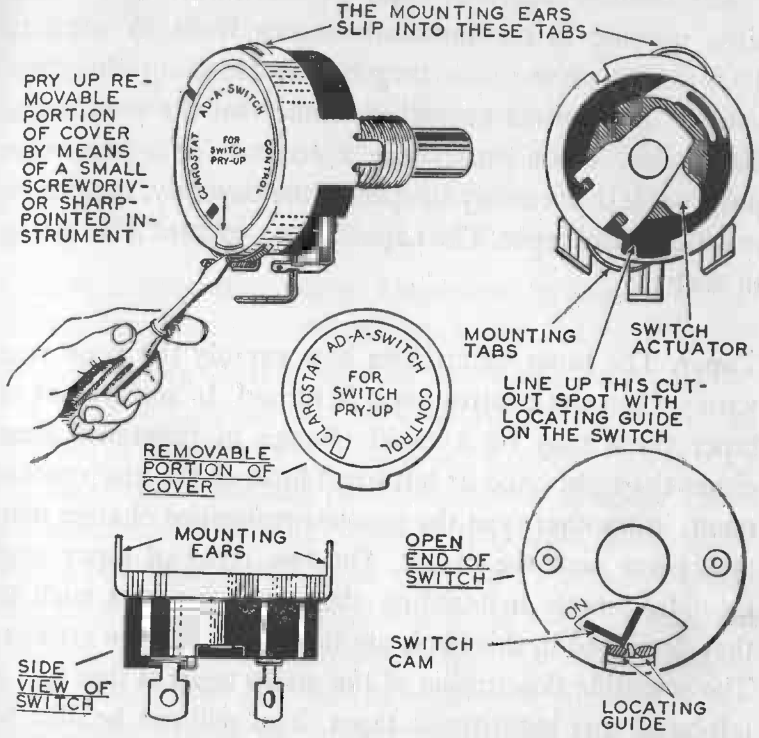
THE MOUNTING EARS SLIP INTO THESE TABS
PRY UP RE-MOVABLE PORTION OF COVER BY MEANS OF A SMALL SCREWDRIV-OR SHARP-POINTED IN-STRUMENT
CLAROSTAT AD-A-SWITCH CONTROL
FOR SWITCH PRY-UP
MOUNTING TABS
SWITCH ACTUATOR
LINE UP THIS CUT-OUT SPOT WITH LOCATING GUIDE ON THE SWITCH
REMOVABLE PORTION OF COVER
MOUNTING EARS
OPEN END OF SWITCH
ON
SWITCH CAM
SIDE VIEW OF SWITCH
LOCATING GUIDE

HOW TO ATTACH AN "ON-OFF" SWITCH TO A CONTROL

Pry off the removable portion of the cover on the control. Hold the control in your left hand with the shaft pointing toward you. Use your right hand to turn the shaft in a clockwise direction as far as it will go. There are two projecting metal tabs or "ears" on the metal rim of the switch—there is also an opening in the switch cover. Inside the switch and visible through the opening is a "V"-shaped cam. Moving the cam operates the switch. Move the cam until it is in the center of the opening. Line up the projecting ears on the switch housing with the metal guides in the control case. Then press the switch firmly against the control and while holding it in that position bend the ears so that the switch cannot slip away from the control.

The shaft on the control will probably prove to be longer than necessary and it should be cut off. If the shaft is the notched type, it can be shortened by holding it in a pair of pliers and breaking it at the notch which will reduce it to the correct length. If the shaft is not notched, cut off the surplus length with a fine tooth hacksaw.

Electronic-equipment catalogs list controls of many sizes ranging in maximum resistance from ½ ohm to 10,000,000 ohms (ten megohms). Hook-up diagrams usually indicate the correct resistance for the controls in the circuit. When purchasing a control for a particular purpose, it is necessary to specify the capacity, maximum resistance and taper. The capacity of a control is its rating in watts.

Taper. The taper determines how rapidly the resistance varies when the control knob is turned. In some types of taper there may be a rapid change in resistance near either the right hand or left hand limit of the knob movement; in another type the greatest resistance change may take place near the center. The two types of taper used most frequently in building electronic appartus such as that described in this book are the AUDIO and the LINEAR. The scientific description of the audio taper is that it is a left-hand true logarithmic taper. You will not be able to understand that definition unless you are an engineer or a mathematician.

The linear taper is not really a tapered control even though it is commonly referred to as such. It has a uniform resistance change from either end. A "linear taper" is used wherever a control should be such that the change in resistance is proportional to the degree of rotation over the entire range of adjustment.

A control can be connected in a circuit so as to regulate either current flow or voltage. When used to regulate current it is a RHEOSTAT and only two terminals are used. When used to regulate voltage the control is a POTENTIOMETER and all three terminals are utilized. Potentiometers usually have a linear taper.

It is sometimes desirable to attach an on-off switch to a control so that the same knob operates both the control and the switch. SPST, SPDT and DPST switches for this purpose are manufactured. One of the illustrations and its caption explains how to attach a switch to a control.

Further Explanation about Capacitor Action. It is customary for electronic technicians to use the word "pass" as if alternating current actually passed THROUGH a capacitor. The word pass may mislead students who are interested in the physics of a capacitor's action. Alternating current does not actually pass THROUGH a capacitor unless the capacitor "leaks." A capacitor may leak because of a defect, low resistance dielectric or other reasons and permit an alternating current to pass through. Without a long explanation of capacitor action, it is more accurate to say briefly that an efficient condenser, in good condition, blocks a steady direct current but permits an alternating current to surge back and forth in a circuit without in reality passing THROUGH the capacitor.

CHAPTER SEVEN

Transducers—Components of an Electronic Circuit: EARPHONES, LOUDSPEAKERS, MICROPHONES AND PICKUPS

There is a word in our language which once belonged only to physicists and the dictionary but now is coming into more common usage. It is the term TRANSDUCER and it means CONVERTER OF ENERGY. You hold two transducers in your hand when you make a telephone call. The microphone into which you direct your voice converts sound energy into electrical energy. The receiver into which you listen converts electrical energy back into sound energy.

There are many kinds of transducers. Some of them, such as the bolometer, glass electrodes and electrical strain gages, are used only in laboratories. A bolometer is an electrical instrument for measuring minute quantities of radiant heat by changes in the resistance of a blackened platinum strip exposed to the radiation. The development of transducers in recent years and their

combination with the electronic amplifier and recording instruments brought a revolution into the broad field of scientific measurements. Practically any physical or chemical property of matter can be made to give an electrical signal by means of transducers. They are invaluable in industrial research and development laboratories.

All of us have at least a nodding acquaintance with a few transducers. Two of these, the telephone receiver and telephone transmitter, have been mentioned. Others in this class are the loudspeaker, microphone, phonograph pickup and photocell. This chapter describes some of them.

TELEPHONE RECEIVERS AND HEADSETS FOR RADIO AND ELECTRONIC CIRCUITS

In the early days of radio, there were no loudspeakers. Loudspeakers had not been developed because they required more energy than a radio receiver could furnish without amplification and there were no amplifiers. A telephone receiver requires very little energy in order to produce sounds in comparison to the energy required for a loudspeaker. Telephone receivers made the signals audible. They were the transducer used to convert small amounts of electrical energy into audible sounds.

Telephone receivers still are used in aircraft, military and communications service, by experimenters and by "ham" (amateur) radio operators. The hams call a telephone receiver a "can."

A telephone receiver is one of the most sensitive of electrical instruments. A few millionth parts of the energy required to light a small flashlight lamp will produce sounds in a telephone receiver. A good radio headset is so sensitive that it will produce audible sounds in response to changes in the almost infinitesimal amount of current generated when the cord tips are touched to the moist tongue and one tip is tapped against the other.

The telephone receivers or ear-phones used in radio and electronics are the small light-weight variety known as "watch case" or headphones. A complete headset consists of a pair of receivers, a connecting cord and a head band which holds the receivers snugly against the ears.

Electromagnetic Telephone Receivers. The most common type of earphone used in headsets is an electromagnetic device which consists of a case and cap enclosing a permanent magnet, a sheet-iron diaphragm and two small electromagnets wound with very fine wire. The parts are assembled so that the space between diaphragm and the iron cores in the electromagnets is small—usually about .002 inch.

It has been explained that there is always a magnetic force or flux at the poles of an electromagnet and that there is also a magnetic force or flux through the core of an electromagnet when an electric current flows through the winding. It has also been shown that a magnetic flux exerts an attractive or pulling force upon iron.

The magnetic flux from the poles of the permanent magnet in a telephone receiver passes through the cores of the electromagnets and exerts a pull upon the iron diaphragm. The pull of the permanent magnet places the

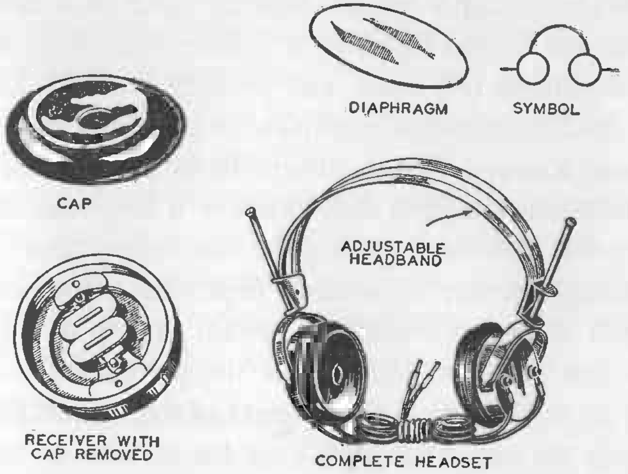

ELECTROMAGNETIC TELEPHONE HEADSET FOR RADIO USE

Telephone receivers make radio signals audible by changing the energy of the rectified high-frequency currents delivered by the detector into sound waves. For radio, it is desirable to have the telephone receivers held firmly to the listener's ears and leave both hands free. A complete radio headset consists of two receivers equipped with a headband so that a receiver is firmly held to each ear. When used for radio work, telephone receivers are wound with much finer wire than is used in the Bell telephone receiver. This makes it possible to wind a large number of turns in small space so that the receiver is more sensitive to weak currents. Incidentally, it also increases the resistance. The Bell telephone receiver has a resistance of approximately 75 ohms but the resistance of a radio headset is 2,000 to 3,000 ohms.

A telephone receiver is one of the most sensitive of electrical instruments. An amazingly small amount of energy flowing through a sensitive telephone receiver will produce sounds audible to human ears. There are two kinds of telephone receivers used in radio, the crystal and the electromagnetic types. Electromagnetic receivers are the "best all around" for amateur experimenters. An electromagnetic telephone receiver consists of a pair of small electromagnets, a permanent magnet and a thin sheet-iron diaphragm. The electromagnets are mounted close to the diaphragm but do not touch it. The permanent magnet exerts a constant pull on the diaphragm. The diaphragm acts like the head of a drum. Only a very small disturbance is required to make it respond. When current flows through electromagnets, the magnetic pull on the diaphragm varies. It becomes stronger or weaker, depending upon whether the current in the windings is aiding or is bucking the pull of the permanent magnet. Each quick change in the pull produces a sound wave.

diaphragm under a tension like the head of a drum. If the tension on the diaphragm is increased or decreased the diaphragm will move and produce a sound. Only a very small disturbance is required to make the diaphragm produce a sound. When current flows through the coils on the electromagnets they produce a magnetic flux of their own. The force pulling the diaphragm then varies, becoming stronger or weaker, depending upon whether the pull of the flux created by current passing through the coils aids or bucks the pull of the permanent magnet. Very small changes in the strength of the current flowing through the electromagnets alter the tension of the diaphragm and cause it to move. Its motions force sound waves out through the hole in the cap.

The strength of the magnetic flux produced when a weak current flows through the electromagnets is greater if the coils consist of many turns of wire than when the same weak current flows through electromagnets having few turns of wire.

The electromagnets in an earphone are wound with hundreds of turns of wire in order to make the phone sensitive to weak currents. The wire must be very small in diameter in order to wind a large number of turns on a small electromagnet. Fine wire has much greater resistance than larger wire. Earphones used in radio headsets usually have a resistance of 1000 to 1500 ohms each. Although radio headsets are rated according to their resistance in ohms, resistance itself is not always the quality which is most desired and it is specified only to give an indication of the turns on the electromagnets. The number of turns of wire wound on the bobbins is more

important than the resistance except in special instances. Headsets having a resistance of 8,000 to 10,000 ohms are manufactured for use in circuits where the high resistance is a desirable quality.

A radio headset requires care to preserve its sensitive condition. Do not drop it and do not unscrew the caps and lift up the diaphragms to see inside. A diaphragm is easily bent and the receiver thereby thrown out of adjustment. Keep the caps screwed tightly in place.

Crystal Telephone Receivers. The fact that certain crystals will be distorted mechanically if a voltage is impressed upon them is used in the construction of "crystal phones." No coils or magnets are used. The crystal is mounted between two electrodes. One electrode is connected to a circular diaphragm. When the electrodes are connected to a source of varying potential the crystal vibrates and transmits its motion to the diaphragm and sounds are produced. Crystal phones respond to a wide range of frequencies (60 to 10,000 cycles) and produce sounds of greater fidelity than electromagnetic receivers.

LOUDSPEAKERS

More energy is required to operate a loudspeaker than is required to produce audible sounds in a sensitive telephone receiver. A pair of headphones connected directly to the detector of a radio receiver will produce sounds which are plainly audible to the person wearing the headset. A loudspeaker connected to the detector under the same conditions would not produce audible sounds. The currents from a detector must be amplified before they will operate a speaker satisfactorily. A loudspeaker connected to the right amplifier will produce sounds which can be heard by a group of people. When used to reproduce voice and music, the sounds from the speaker are a more faithful reproduction of the original than the sounds from a telephone receiver.

The type of speaker used most widely is the dynamic moving-coil PM speaker. PM is an abbreviation for permanent magnet. The mechanism of a PM dynamic speaker is a small coil of wire called the voice coil, a cone-shaped paper diaphragm and a permanent magnet assembled on a metal frame.

The voice coil is a single layer of fine wire wound on a small paper tube and attached to the small end of the paper diaphragm. The diaphragm is mounted on the

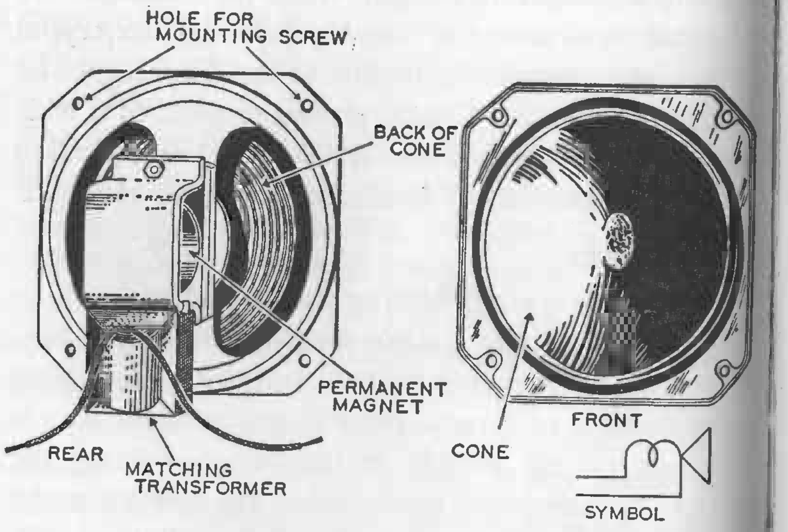
HOLE FOR
MOUNTING SCREW
BACK OF
CONE
PERMANENT
MAGNET
REAR
MATCHING
TRANSFORMER
CONE
FRONT
SYMBOL

FRONT AND BACK OF A SMALL PM LOUDSPEAKER WITH TRANSFORMER

All speakers do not come equipped with a matching transformer.

speaker frame so that the coil attached to its apex "floats" in the narrow space between the poles of the permanent magnet. The coil is close to the magnet poles but does not touch them. The coil is supported so that it is free to move back and forth along its axis but not sideways or up and down. Its movements are like the to and fro motion of a piston in an engine cylinder.

When an electric current is passed through the voice coil it produces a magnetic field. The reaction of the coil's magnetic field produces forces which move the coil. A current flowing through the coil in one direction will cause the coil to be pulled farther into the field of the permanent magnet. A current in the opposite direction

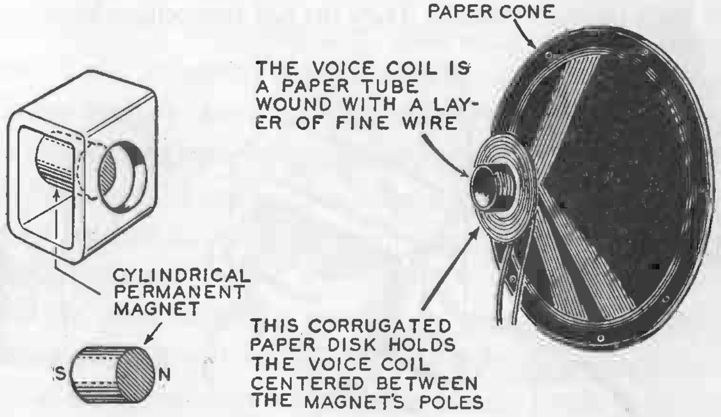
PAPER CONE
THE VOICE COIL IS
A PAPER TUBE
WOUND WITH A LAY-
ER OF FINE WIRE
CYLINDRICAL
PERMANENT
MAGNET
S
N
THIS CORRUGATED
PAPER DISK HOLDS
THE VOICE COIL
CENTERED BETWEEN
THE MAGNET'S POLES

THE MECHANISM OF A PM SPEAKER

The permanent magnet in a PM speaker is usually a small cylinder of Alnico or similar alloy steel. It is mounted in an iron frame with a circular opening in the front. The magnetic path is arranged so that one end of the cylinder is one pole and the edge of the opening in the frame is the other pole. The voice coil moves back and forth in the magnetic field between the poles.

will cause the coil to be pushed away from the permanent magnet. A fluctuating direct current or an alternating current causes the coil to move to and fro. Since the coil is attached to the diaphragm, the diaphragm moves when the coil moves. Movements of the diaphragm push the air and produce sound waves. If the fluctuations in the direct current or alternations of the alternating current occur between approximately 40 to 10,000 per second, the movements of the ordinary PM speaker is able to follow them and produce sounds of corresponding frequencies.

A speaker is rated in size according to the diameter of its cone-shaped diaphragm. A 6-inch speaker has a 6-inch diaphragm; a 12-inch speaker has a 12-inch diaphragm, etc. Speakers smaller than 8 inches are noticeably partial to high-pitched sounds. They do not reproduce base tones

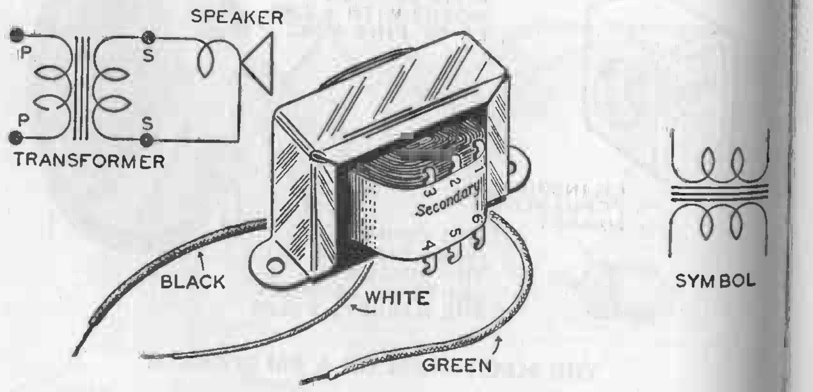

OUTPUT TRANSFORMER AND CIRCUIT CONNECTING A SPEAKER TO AN AMPLIFIER

The black, white and green wires are primary or input terminals.

as well as higher tones. No speaker is equally efficient over the full range of pitch audible to human ears. A 12-inch speaker gives more faithful reproduction than the smaller sizes, but it too favors some tones more than others and does not reproduce faithfully some of the higher pitched sounds. When high fidelity is desirable, three speakers of different size are often used in order to cover the audio-frequency sound range more efficiently.

Speakers are rated in capacity according to the number of watts they will carry. The 12-inch speakers built for home hi-fidelity phonograph installations have a rating of 10 to 25 watts. Some of the 12-inch speakers designed for public address systems have a rating of 35 to 50 watts.

The impedance of the voice coil in a speaker must be taken into consideration before it is connected to an amplifier. The impedance of standard speakers ranges from 3 to 50 ohms. Those used in ordinary radio receivers usually have an impedance of 3 to 4 ohms. Speakers designed for public address systems have an impedance of 6 to 8 ohms. The impedance of speakers manufactured for intercom systems is the highest, usually 45 to 50 ohms.

If the voice coil of a speaker is connected directly to an amplifier, it will not operate satisfactorily. The reason is that the impedance of the plate circuit in the amplifier output tube is much greater than the impedance of the voice coil in a PM speaker, and for efficient operation the impedance of the speaker should match the impedance of the tube. The problem is solved by the use of a small closed core transformer called a "matching" transformer.

The primary of a matching transformer is wound so that its impedance nearly matches the impedance of the plate circuit in the amplifier. It may have an impedance of 14,000 ohms. The secondary of the transformer is wound so that it has about the same impedance as the speaker voice coil. When the primary of the transformer is connected to the output of the amplifier and the secondary is connected to the voice coil, the speaker is "matched" to the amplifier.

Some matching transformers are built with taps on both the primary and secondary so that different impedance values are available. Matching transformers are made in several sizes having a capacity of 3.5 to 35 watts. The proper procedure is to select a transformer which will match most nearly the impedance of the signal source. A difference of 10% is usually of no great importance. If a close match cannot be secured, it is best to select an impedance value which will present a higher rather than a lower impedance to the output tube.

Cabinets for Loudspeakers. The common loudspeaker with a paper "cone" or diaphragm radiates sound waves directly into the air. Normally sound waves will be radiated from both sides of the cone. Those from the back of the cone combine with those from the front of the cone. Under these conditions the waves of low-pitched sounds tend to combine so as to mutually cancel each other. Therefore a loudspeaker is usually placed in an enclosure open to the front of the cone but closed at the back so that no sound waves can be radiated from the back of the diaphragm.

MICROPHONES

A microphone is a device which picks up sound vibrations and produces fluctuating currents or potentials. The ordinary telephone transmitter is a microphone. Usually there is a diaphragm in a microphone and the diaphragm vibrates when struck by sound waves. The vibrations occur in step with those of the sound waves. The movements of the diaphragms actuate an electrical device which may be a carbon button, a small metal ribbon or a piezo-electric crystal.

The Carbon Microphone. A single-button carbon microphone consists of a metal diaphragm which, when it vibrates, exerts mechanical pressure against a group of loosely packed carbon granules contained in a small cup (called the button). The carbon granules lie between two electrodes. One electrode is rigidly fixed but the other is attached to the diaphragm and moves when the diaphragm moves. The electrodes make electrical contact with the granules but there is a certain amount of resistance to the flow of current from one electrode and through the granules to the other electrode. This resistance changes if the granules are disturbed. Vibrations of the diaphragm disturb the granules. If a carbon microphone is made part of an electrical circuit, the resistance of that circuit will fluctuate when sounds strike the microphone.

Single-button carbon microphones are more sensitive than two-button microphones but their tone quality is not as high.

A carbon microphone itself does not generate any cur-

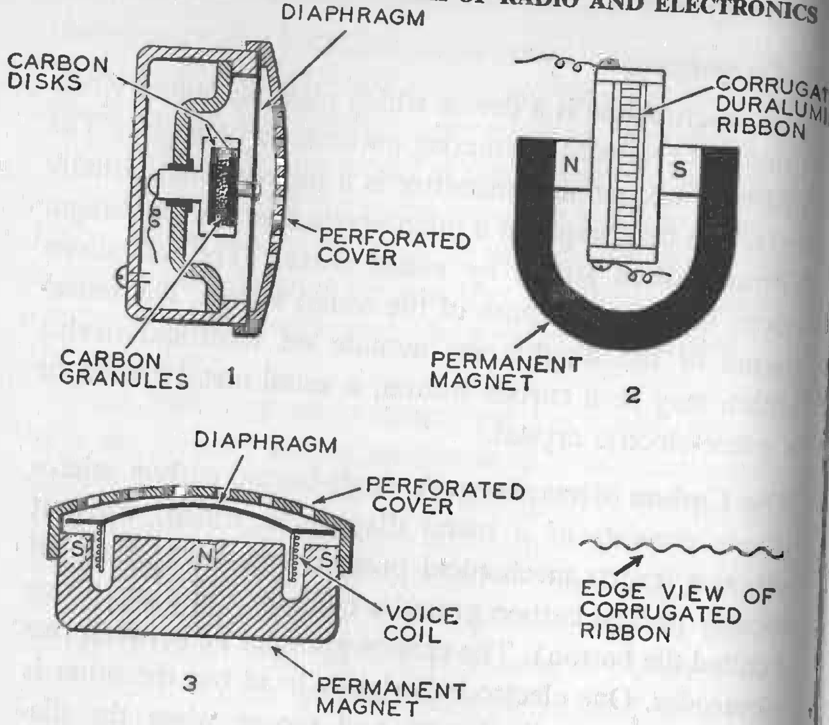
DIAPHRAGM
CARBON DISKS
PERFORATED COVER
CARBON GRANULES
1
CORRUGATE
DURALUMIN
RIBBON
N
S
PERMANENT MAGNET
2
DIAPHRAGM
PERFORATED COVER
S
N
S
VOICE COIL
EDGE VIEW OF CORRUGATED RIBBON
3
PERMANENT MAGNET

MICROPHONES

Several types of microphones are used in radio broadcasting. Here are three. (1) A single-button carbon microphone is shown in cross section at the upper left. The parts marked carbon disks are the electrodes. (2) A velocity or ribbon microphone consists of a thin corrugated Duralumin ribbon suspended between the poles of a permanent magnet. When the ribbon vibrates under the impact of sound waves, it is moving in a magnetic field and an alternating voltage is induced in the ribbon. The ribbon is connected to a small transformer and a high-gain pre-amplifier. The velocity microphone acquired its name from the fact that the ribbon vibrates in response to the velocity of the air particles moving in a sound wave rather than to the sound pressure. (3) The electrodynamic microphone reverses the principle of the speaker. A small coil of wire attached to a diaphragm is suspended in a magnetic field. When the diaphragm vibrates under the impact of sound waves an alternating current is generated in the moving coil. This current is fed into a small transformer, the secondary of which is connected to the input of an amplifier.

rent or voltage. If placed in a circuit with suitable direct current supply it will cause the current to fluctuate in step with the variations in its own resistance.

Crystal Microphones. No separate source of current or voltage is necessary for a crystal microphone. It generates its own voltage. A crystal microphone is practically a crystal earphone in reverse. The diaphragm of a crystal microphone is mechanically connected to a piezo-electric crystal. Sound waves which strike the diaphragm cause it to vibrate. The vibrations twist the crystal. Since it is a piezo-electric crystal, its disturbance generates a small alternating current voltage.

Electrodynamic Microphone. The electrodynamic or moving microphone reverses the principle of the loudspeaker. A small coil of fine wire attached to a diaphragm is suspended between the poles of a magnet. When the diaphragm vibrates under the impact of sound waves the

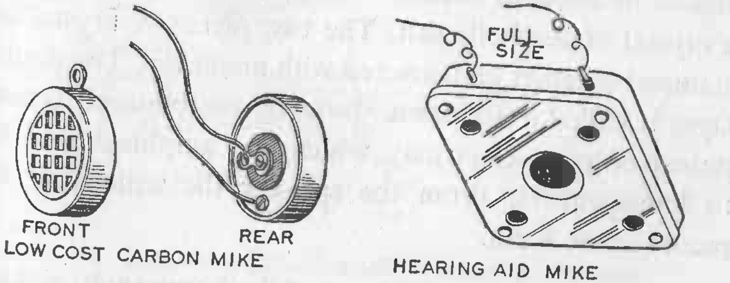
FULL SIZE
FRONT
REAR
LOW COST CARBON MIKE
HEARING AID MIKE

SMALL CARBON MICROPHONES
USEFUL TO THE YOUNG AMATEUR EXPERIMENTER

Microphones such as these can often be purchased at small cost from mail order firms which advertise in radio magazines.

movements of the coil in the magnetic field generates an alternating voltage. An alternating current results if the microphone is connected in a closed circuit.

A PM loudspeaker can be used as an electrodynamic microphone. In fact PM loudspeakers function in intercom sets both as microphones and speakers.

Crystal Pickups. When certain crystalline materials are pressed or twisted, they generate electricity. The electromotive force produced in this manner is called PIEZO-electricity. The word *piezo* was derived from Greek and means *press* or *squeeze*. Crystals which will produce piezo-electricity are called piezo-electric crystals. Piezo-electric crystals have another remarkable and useful property. If placed between two metal plates, the crystals will change in width, in length or in thickness if a voltage is applied to the plates. Piezo-electric crystals of Rochelle salt and of quartz are in common use in electronics. The crystal pickups used to play phonograph records produce piezo-electricity. Movement of the needle as it follows the groove in a record twists two rectangular plates cut from a crystal of Rochelle salt. The two pieces of crystal are clamped together and covered with metal foil. The double layer is called a BIMORPH. Twisting the bimorph creates pulses of piezo-electricity which are amplified and fed to loudspeakers. From the speaker the sounds of the record come forth.

Variable Reluctance Pickups. All phonograph pickups do not utilize piezo-electricity. Some of the best reproduction from high-fidelity phonographs is furnished by variable reluctance pickups. These are magnetic devices.

Movement of the phonograph needle varies the amount of magnetism which passes through two small coils in the pickup. The variations in the magnetism generates a feeble alternating voltage which is greatly amplified before entering the loudspeaker. The voltage generated by a variable reluctance pickup is so small that it requires amplification by a pre-amplifier before it passes on to the ordinary phonograph amplifier.

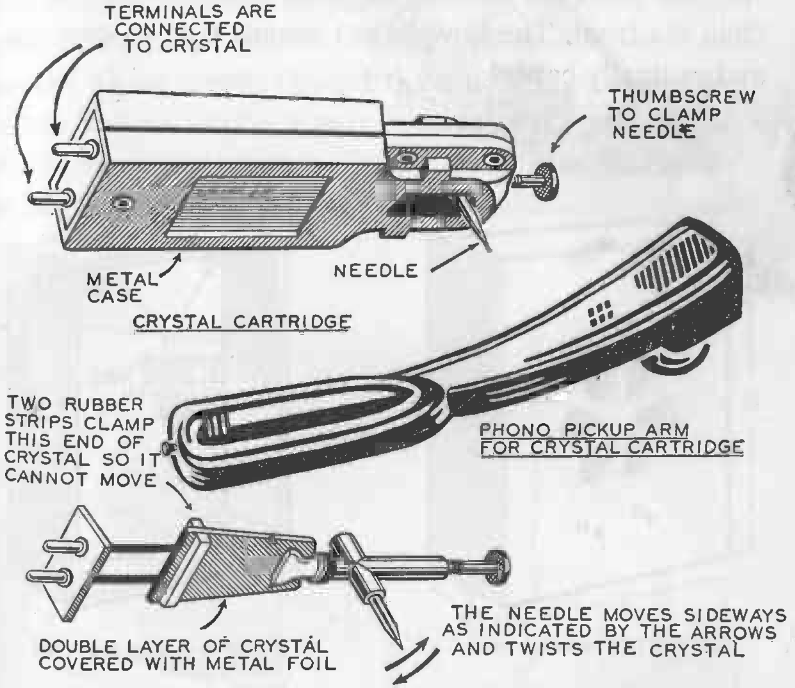
TERMINALS ARE CONNECTED TO CRYSTAL
THUMBSCREW TO CLAMP NEEDLE
METAL CASE
NEEDLE
CRYSTAL CARTRIDGE
TWO RUBBER STRIPS CLAMP THIS END OF CRYSTAL SO IT CANNOT MOVE
PHONO PICKUP ARM FOR CRYSTAL CARTRIDGE
DOUBLE LAYER OF CRYSTAL COVERED WITH METAL FOIL
THE NEEDLE MOVES SIDEWAYS AS INDICATED BY THE ARROWS AND TWISTS THE CRYSTAL

THE TRANSDUCER ON A RECORD PLAYER

Crystal pickups are in common use on record players. Movement of the needle or stylus as it follows the groove in a record twists two plates cut from a crystal of Rochelle salt. The twisting creates voltage pulses which are amplified and then led to a loudspeaker. The two plates of crystal are clamped together. The double layer is called a BIMORPH.

Baffles and Enclosures for Loudspeakers. Loud speakers should be mounted in a baffle or enclosure for two reasons:

1. To protect from dust and mechanical injury.
2. To improve the sound.

Sound waves are radiated from both the back and front of the cone-shaped paper diaphragm. When a speaker is not in a baffle or enclosure of some type, the sound waves radiated from the back of the cone combine with those from the front. The low-pitched sounds tend to combine and mutually cancel.

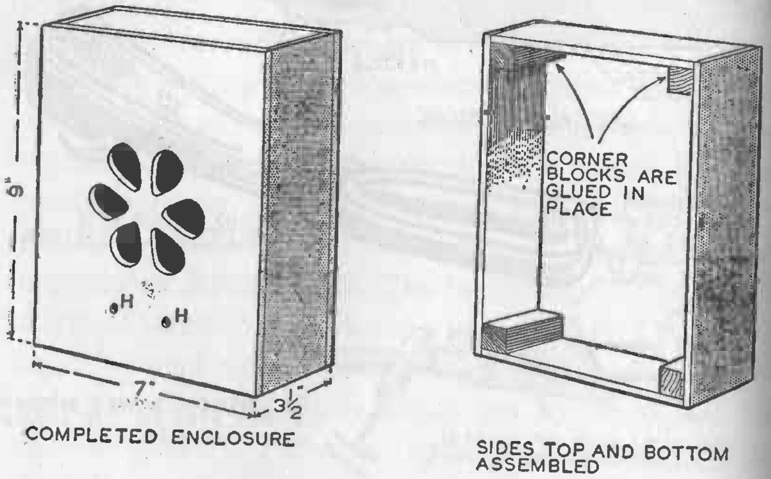

COMPLETED ENCLOSURE

SIDES TOP AND BOTTOM ASSEMBLED

ENCLOSURE FOR SMALL SPEAKER

The enclosure illustrated above is suitable for a 3-, 4- or 5-inch speaker. The six petal-shaped openings in the front can be cut by hand with a fretsaw or with a power jigsaw. The speaker should be mounted on the back of the front panel. The location of the mounting holes depends upon the size of the speaker. The two holes marked HH in the illustration are for binding posts.

The simplest form of baffle is a large sheet of Celotex. A hole of the same diameter as the diameter of cone on the speaker is cut in the center of the sheet. The speaker is mounted on the back directly over the hole.

Two of the illustrations show a small cabinet suitable for a 3-, 4- or 5-inch speaker. This type of enclosure will not produce "hi-fi" tone quality but will protect the speaker and is easily constructed.

Base Reflex Cabinet. This is a common type of enclosure designed to secure the best tonal quality from a loudspeaker. Good results depend upon a cabinet of correct volume and an auxiliary port opening of correct area to suit the size of the speaker. One of the illustrations shows the correct dimensions for a 6-inch speaker. The cabinet

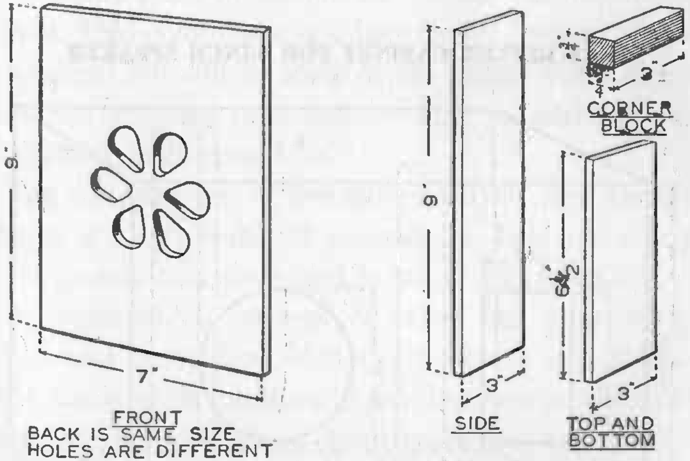

PARTS FOR THE SPEAKER CABINET

The parts for the cabinet, with the exception of the four corner blocks, are cut from ¼- to ⅜-inch plywood.

should be built of ½-inch plywood. The dimensions can be changed slightly to suit particular needs provided the area of the port in square inches and the enclosed volume of the cabinet measured in cubic inches are not changed. The inside surface of both sides and the inside surface of the back should be covered with Tufflex sound insulating material ⅝ inch thick. The Tufflex should be attached to the wood with rubber cement. The cement used for setting rubber floor tiles is satisfactory for this purpose. The speaker should be mounted on the back of the front panel directly over a circular opening of the same diameter as the speaker cone. A base reflex cabinet can be incorporated into a larger cabinet enclosing other portions of a "hi-fi" sound system.

BASE REFLEX CABINET FOR 6-INCH SPEAKER

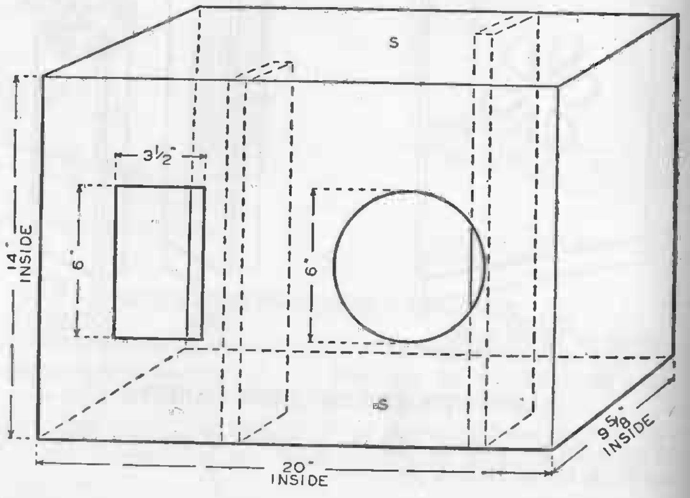

CHAPTER EIGHT

Transistors—Versatile Midgets Which Can Do the Work of Vacuum Tubes

After many years of unique service the vacuum tube has a rival. This competitor is a tiny device, without vacuum or filament but will do some of the things which the conventional vacuum tube can do and, in addition, some things that tubes cannot do.

The essential part of this comparatively new electronic wonder is a tiny wafer of germanium. It is a first cousin of the germanium diode and is called a TRANSISTOR.

A considerable amount of effort has gone into the design and production of this new device and it has become the most revolutionary development in the field of electronics in many years. Transistors are semi-conductor devices capable of doing most of the things that vacuum tubes can do, and as previously stated, they perform some jobs which heretofore have been difficult if not impossible for vacuum tubes alone to handle. Transistors may be used as detectors, amplifiers, oscillators, mixers, modulators, attenuators and relays.

The Advantages of Transistors. Transistors are much smaller, lighter, longer-lived, more rugged and more efficient than vacuum tubes. They require no heater current or "warm up" time. They are ready for instant operation and consume only infinitesimal power. Two coins, one of copper and one of silver, moistened on the tongue will provide enough current to operate a transistor.

Transistors were invented by Dr. John Bardeen and Dr. Walter Brattan, members of a research group in the Bell Telephone Laboratories directed by Dr. William Shockley.

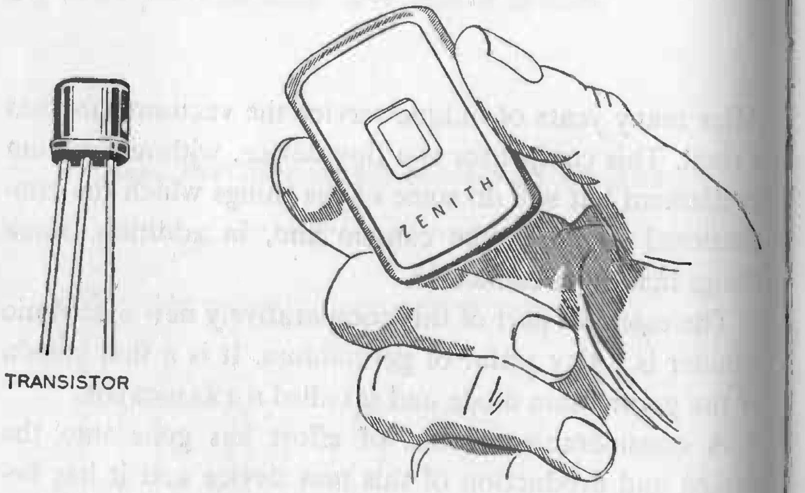

TRANSISTORS REDUCE THE SIZE OF ELECTRONIC EQUIPMENT

Transistors have made possible miniature electronic apparatus operated by tiny, low-voltage batteries. The Zenith "50-X" Hearing Aid, smaller than a package of cigarettes, is an example of such miniaturization. Four transistors supplied with current from a small low-cost battery do the work formerly performed by much larger vacuum tubes requiring expensive batteries.

The small size and low power requirement of transistors are changing the design of much electronic equipment by reducing its size. The first large use for transistors was in hearing aids where they replace the vacuum tubes used as amplifier tubes. Battery replacement is the major cost of operating a hearing aid equipped with amplifier tubes. Transistor-equipped hearing aids are not only much smaller in size but have slashed operating costs from $50 to $75 a year to $5 a year or less.

Transistors have made it possible to build small superheterodyne radio receivers the size of a cigarette package.

Transistor Types. Transistors will probably undergo changes as they are further developed. In their present form, they are triodes. They have three electrodes known as the EMITTER, the COLLECTOR and the BASE. Two types are manufactured.

One, called the POINT-CONTACT transistor, is of little use to the average amateur electronics experimenter. It consists of a very small piece of semiconductor material such as germanium with two properly spaced, pointed, small diameter wires (electrodes) making contact with the surface of the germanium. In appearance the point-contact transistor resembles a crystal diode with an additional electrode. Its mode of operation is entirely different. Point-contact transistors have wide use in switching circuits and in oscillator circuits at frequencies which are normally not possible with the type of transistor known as a JUNCTION transistor.

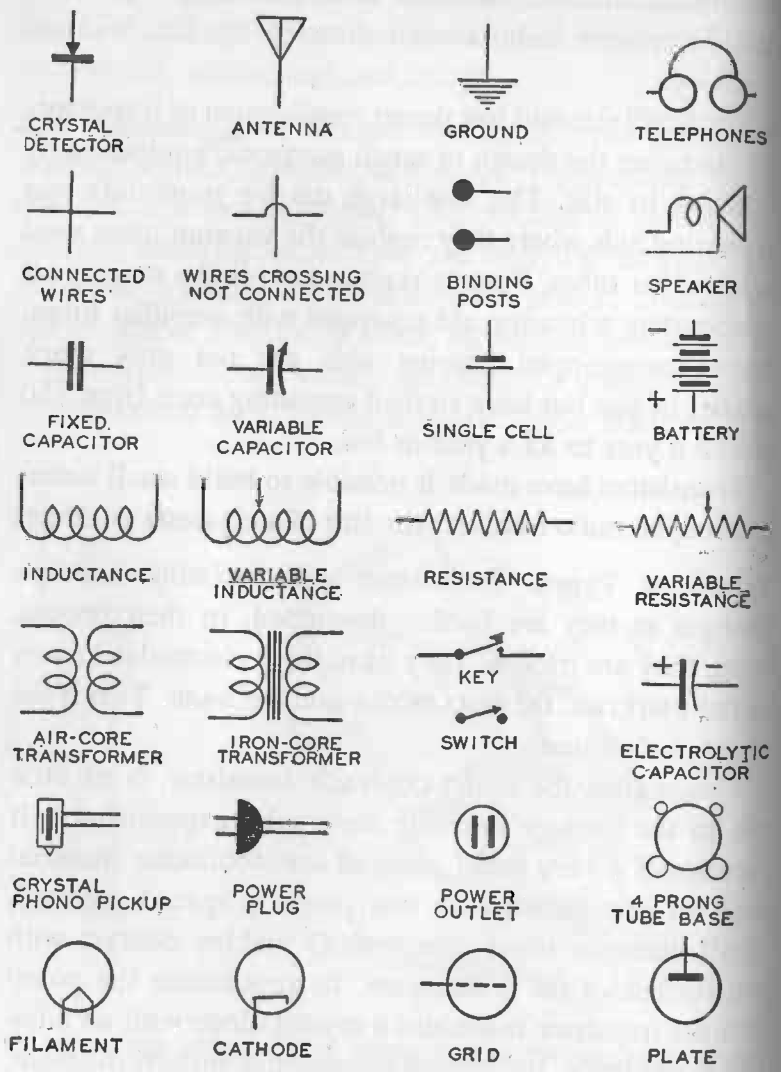
CRYSTAL DETECTOR
ANTENNA
GROUND
TELEPHONES
CONNECTED WIRES
WIRES CROSSING NOT CONNECTED
BINDING POSTS
SPEAKER
FIXED CAPACITOR
VARIABLE CAPACITOR
SINGLE CELL
BATTERY
INDUCTANCE
VARIABLE INDUCTANCE
RESISTANCE
VARIABLE RESISTANCE
AIR-CORE TRANSFORMER
IRON-CORE TRANSFORMER
KEY
SWITCH
ELECTROLYTIC CAPACITOR
CRYSTAL PHONO PICKUP
POWER PLUG
POWER OUTLET
4 PRONG TUBE BASE
FILAMENT
CATHODE
GRID
PLATE

THE SYMBOLS USED IN DIAGRAMS OF RADIO CIRCUITS

Symbols are a sort of radio shorthand used to illustrate schematic circuit diagrams in radio books and magazines.

The junction transistor costs less than the point-contact type. It is of greater interest to the amateur experimenter in electronics and is the type used in building the transistor equipment described in this book.

There are several varieties of junction type transistors. The CK722 transistor manufactured by Raytheon Manufacturing Company of Waltham, Massachusetts, is the "p-n-p" type. It consists of a tiny wafer of germanium protected with about three layers of plastic so that it is hermetically sealed in its plastic jacket. The three wire leads which emerge from the bottom of the jacket are connected to the EMITTER, BASE and COLLECTOR. Current enters the transistor through the emitter lead and

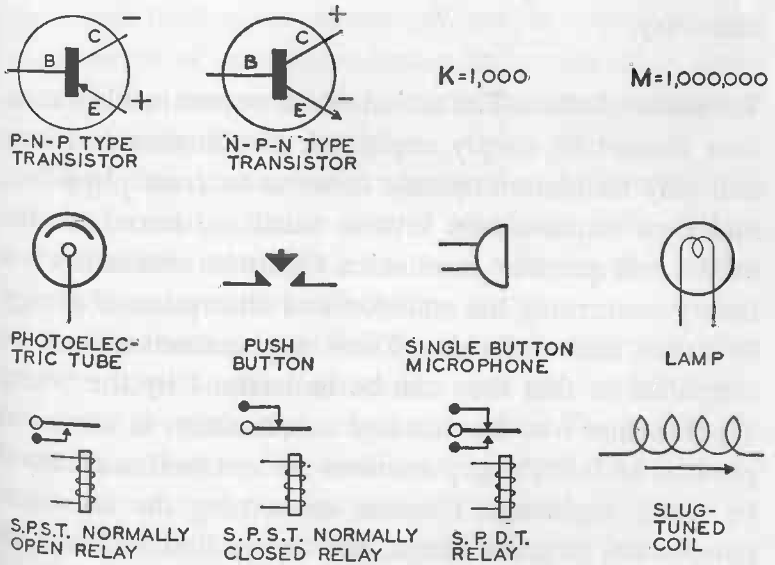
C
B
E
P-N-P TYPE TRANSISTOR
N-P-N TYPE TRANSISTOR
K=1,000
M=1,000,000
PHOTOELECTRIC TUBE
PUSH BUTTON
SINGLE BUTTON MICROPHONE
LAMP
S.P.S.T. NORMALLY OPEN RELAY
S.P.S.T. NORMALLY CLOSED RELAY
S.P.D.T. RELAY
SLUG-TUNED COIL

THE SYMBOLS USED IN DIAGRAMS OF RADIO CIRCUITS, CONTINUED leaves through the collector lead. The emitter is a speck of indium on one side of the germanium wafer. The collector is also a tiny piece of indium but it is located on the opposite side of the wafer from the emitter. The wafer itself is the base. The wire lead on the side marked with the red dot is connected to the collector. The center wire lead is connected to the base. The remaining wire is connected to the emitter. In case you do not know what indium is, it may be interesting to know that indium is one of the metallic elements. It is white and somewhat soft like tin. It leaves a mark when drawn across paper. Indium was first discovered in 1863 by Ferdinand Reich, a professor of physics at the famous School of Mines at Freiberg, where germanium was also discovered. Reich's assistant, Hieronymus Theodor Richter, shared in the discovery.

Transistor Action. The action which occurs inside a tranistor cannot be simply explained. Explanations of how and why transistors operate come to us from physicists, and their explanations involve much advanced mathematics and quantum mechanics. Quantum mechanics is a theory concerning the emission and absorption of energy by atoms and molecules. These explanations cannot be simplified so that they can be understood by the young experimenter who has not had an education in advanced physics. In fact the explanations are not well understood by many engineers. Theories concerning the transistor complicate, to some extent, the theory that we have all been taught that an electric current is a flow of electrons. It is not necessary to know anything about transistor

action in order to build the transistor-equipped apparatus described in the following pages. For these many reasons, no attempt will be made here to explain transistor action except by making a slight comparison between it and the familiar vacuum tube.

Transistor Characteristics. The emitter, base and collector of a transistor compare somewhat to the vacuum tube's cathode, grid and plate. The fundamental differences between transistor and vacuum tube action are:

1. In the vacuum tube electrons are liberated into an evacuated space or VACUUM by a heated cathode. The movement of the electrons toward a positively charged plate is an electric current.

2. In a transistor, electrons or holes (spaces where there are inadequate electrons) are injected into the SOLID BODY of a semiconductor by means of a small battery. The movement of the electrons through the SOLID material is an electric current.

When these statements are summed up in fewer words, the fact most evident is that in a vacuum tube electric current FLOWS THROUGH A VACUUM, whereas in a transistor the current FLOWS THROUGH A SOLID.

Another fundamental difference between vacuum tubes and transistors is in their mechanism of control. Mechanism of control is an engineer's term. To explain the term "mechanism of control" we might call the PRESSURE ON THE TRIGGER the "mechanism of control" in a rifle. Signal VOLTAGE and a minute grid charging current controls the cathode to anode electron flow in a radio receiving or amplifier tube and is the mechanism of control.

Signal ENERGY or CURRENT controls the electron flow in a transistor and is the mechanism of control. The signal source employed to operate a tube need deliver VOLTAGE only, but the signal source employed to operate a transistor must deliver CURRENT.

When a circuit is to be designed to couple a tube or a transistor to microphones, telephone receivers, speakers, etc., the radio engineer bases his plans and calculations on:

1. A vacuum tube is a voltage-actuated high-impedance device.
2. A transistor is a current-operated low-impedance device.

The Care Required in Handling Transistors. Although transistors are fairly rugged they can be damaged if improperly handled. The following precautions should be observed.

1. Do not drop or subject a transistor to severe mechanical shock.
2. Do not operate a transistor at higher voltages or greater current values than those designated for it.
3. When connecting a transistor, observe and comply with the polarity designation. Passing a current through a transistor in the wrong direction may damage it.
4. Do not connect a transistor in a circuit until all other connections have been made.
5. Whenever possible, avoid cutting transistor leads short.

Transistors are easily damaged by too much heat and special precautions must be taken when transistor leads

are soldered. Use rosin-core solder only, never acid-core. Use high quailty solder (60% tin and 40% lead). This quality is available in Ersin, Kester, Alpha and some other brands of solder. Scrape the connections before soldering so that they can be soldered quickly and without much heat. Use a small or miniature soldering iron. Grasp each transistor lead in turn as close to the free end as possible with a pair of long nose pliers. The pliers will conduct much of the heat away from the lead and prevent it from reaching the transistor.

There is no attempt at miniaturization in the design of the transistor-operated apparatus in this book. It is important to become familiar with the design, use and limitations of transistors before attempting to build MINIATURE receivers, amplifiers, etc.

Transistor Kits. If you wish to build miniaturized apparatus, you can buy complete kits of parts for building transistor-operated pocket radio receivers, amplifiers, etc. Miniature and sub-miniature controls, transformers, switches and coils are also available.

Transistor Batteries. Small 1½-volt flashlight cells called "penlite" cells give satisfactory service in many transistor circuits. They can be purchased at any "dime store" or hardware store. There are two sizes. Several manufacturers make them but the list numbers of all the different brands would be out of place here. The Burgess No. 7, called a "slim penlite" cell, or its equivalent of other brands is 1¾ inches long and $\frac{13}{32}$ inches in diameter. The Eveready No. 915, called a "penlite" cell, or its equivalent of other brands is $1\frac{31}{32}$ inches long and $\frac{35}{64}$

inches in diameter. An Eveready cell of the same size, known as No. 1015, is made for hearing aids. The hearing aid cells cost approximately the same price as the penlite cells but will last about six times as long as the No. 915 in transistor service.

Special tubular-shaped transistor batteries are also obtainable. Each tube contains fifteen 1.4 volt alkaline dry cells. Power requirement from 1.4 to 21 volts can be met simply by slicing off the necessary number of cells.

CHAPTER NINE

Transistor Apparatus Which You Can Build . . .

A ONE-STAGE AUDIO AMPLIFIER • A COMBINATION CRYSTAL DIODE AND TRANSISTOR RADIO RECEIVER • A TRANSISTOR TIMER SWITCH • HOW TO BUILD A TRANSISTOR CODE PRACTICE OSCILLATOR • LEARNING TO TELEGRAPH.

Although transistors have been in use less than a decade, they have found so many applications they are being manufactured at the rate of 10,000,000 per year. In a few years transistor production is expected to total 30,000,000 per year. They truly are the basis for the electronics of the future. A transistor is free of many of the vacuum tube's limitations such as bulkiness, short life, fragility and high power consumption. A transistor saves space, weight, heat, power lasts 150 times as long as a vacuum tube and uses 1/1000 of the electric current.

Building radio and electronic apparatus which uses a transistor in place of a radio tube is easy for the amateur experimenter. An important advantage from the stand-

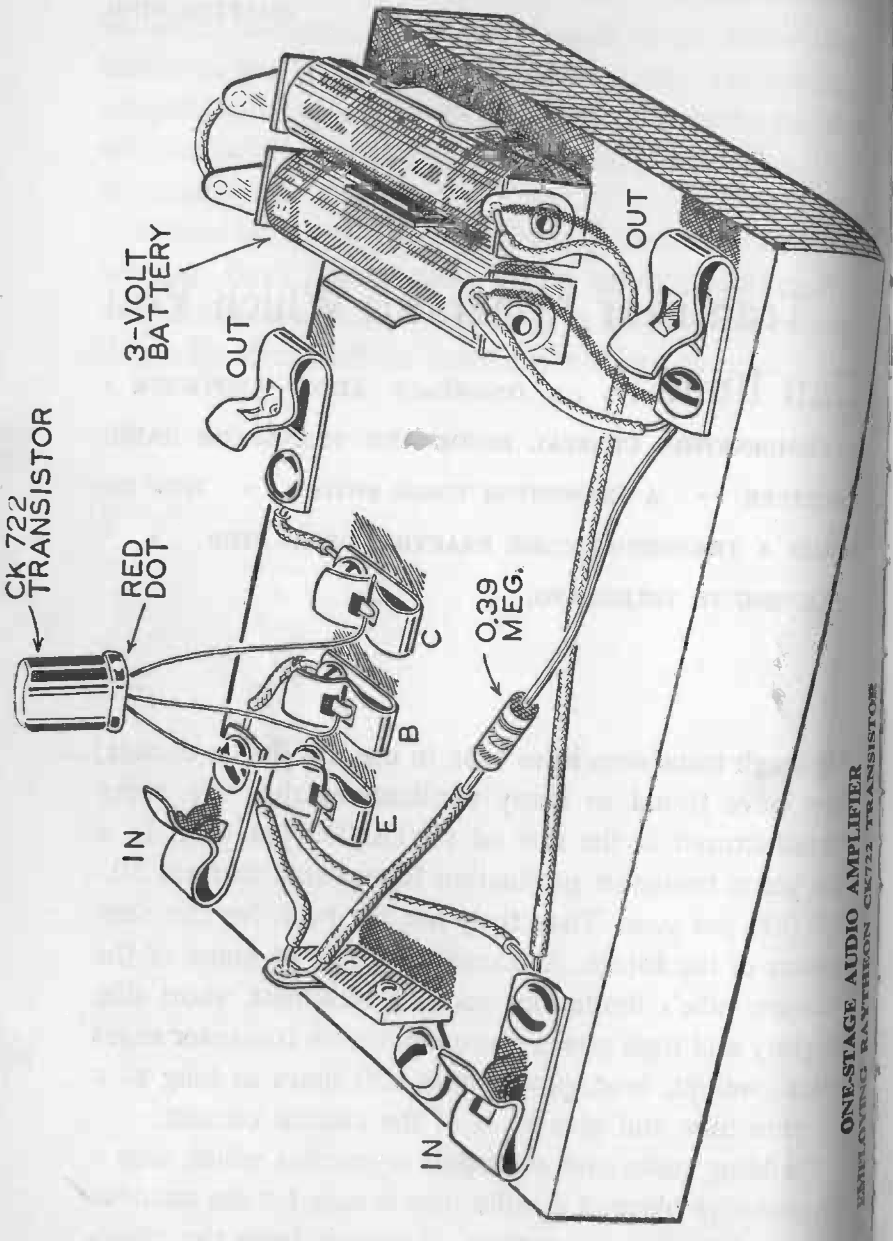
CK 722 TRANSISTOR
RED DOT
3-VOLT BATTERY
OUT
OUT
0.39 MEG.
C
B
E
IN
IN

ONE-STAGE AUDIO AMPLIFIER
EMPLOYING RAYTHEON CK722 TRANSISTOR

point of the young experimenter is: transistor circuits are not connected to the 120-volt power line. No one in your household can object to your hobby on those grounds. Transistor circuits are operated by batteries. The batteries cost only a few pennies and last a long time.

There is no attempt at miniaturization in the design of the instruments described in this chapter. To make or assemble miniature and sub-miniature apparatus requires more mechanical skill than to make or assemble apparatus of larger size. There are many first-class kits of parts on the market for building miniature amplifiers, radio

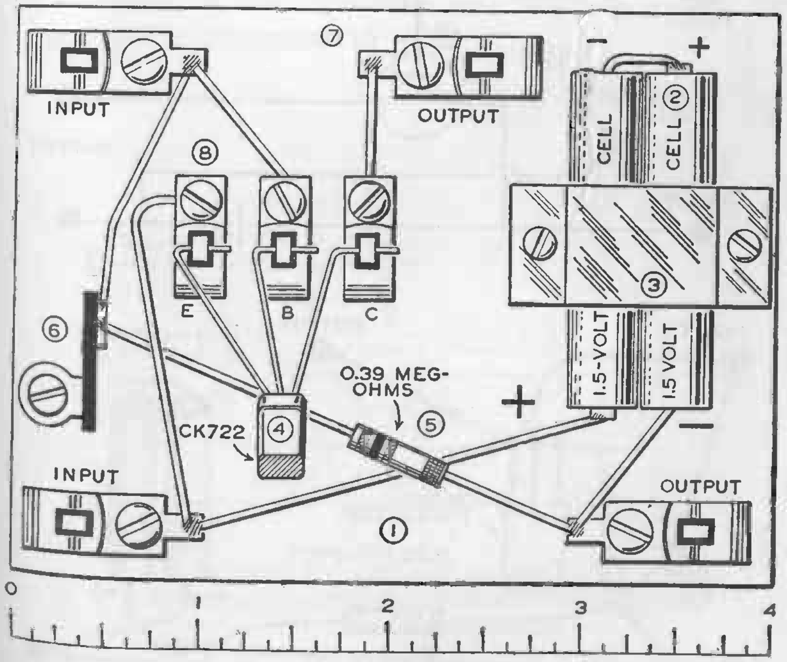
INPUT
⑦
OUTPUT
CELL
CELL ②
⑧
③
E
B
C
⑥
0.39 MEG-OHMS
⑤
+
1.5-VOLT
1.5 VOLT
−
+
−
CK722
④
INPUT
OUTPUT
①
0
1
2
3
4

PLAN OF THE ONE-STAGE TRANSISTOR AUDIO AMPLIFIER receivers, oscillators, timers, etc. If you wish to build miniature equipment, it is advisable to buy the kits and follow the instructions which accompany them.

ONE-STAGE TRANSISTOR AUDIO AMPLIFIER

One of the simplest transistor circuits is the one-stage audio amplifier shown on the following page. In addition to simplicity this amplifier has the advantage of low cost.

SCHEMATIC AND PICTORIAL DIAGRAMS
OF ONE-STAGE AUDIO AMPLIFIER
EMPLOYING RAYTHEON CK722 TRANSISTOR

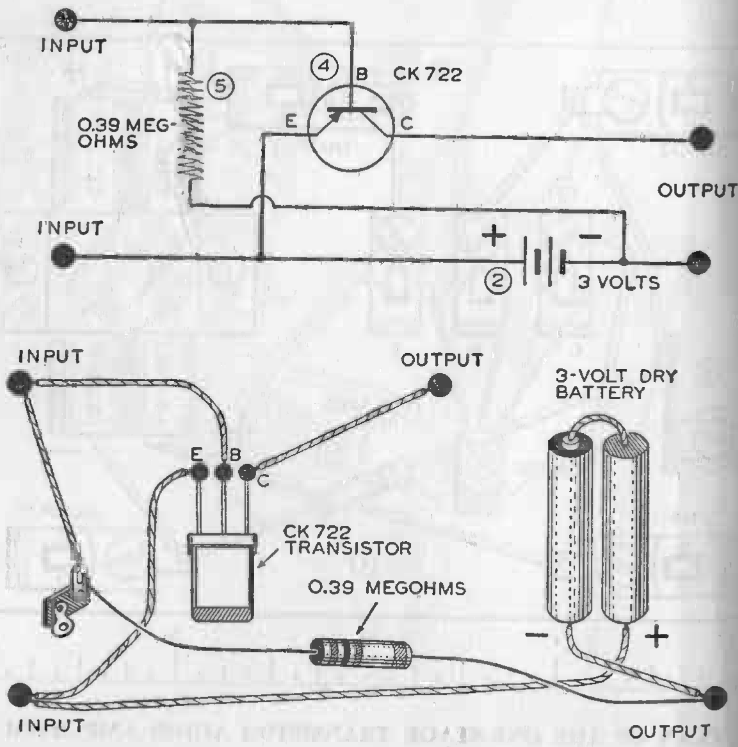
INPUT
④ B CK722
⑤
0.39 MEG-OHMS
E
C
OUTPUT
INPUT
+
−
② 3 VOLTS
INPUT
OUTPUT
3-VOLT DRY BATTERY
E B C
CK722 TRANSISTOR
0.39 MEGOHMS
−
+
INPUT
OUTPUT

The components can be purchased for $1.75 or less. When connected to either of the crystal detector receivers described in Chapter Three, this amplifier will greatly increase the strength of the signals.

The amplifier is so designed that the transistor is held in three Fahnestock clips. This eliminates soldering and possible damage to the transistor by the heat of the soldering iron. The transistor can be removed quickly from the clips for use in another piece of equipment if you wish. The first circuit diagram of the amplifier specifies a a Raytheon CK722 transistor. A General Electric type 2N107 transistor will serve equally well. A circuit diagram for the G.E. 2N107 is also illustrated. Notice that the values of the capacitors and resistors in the two circuits are different.

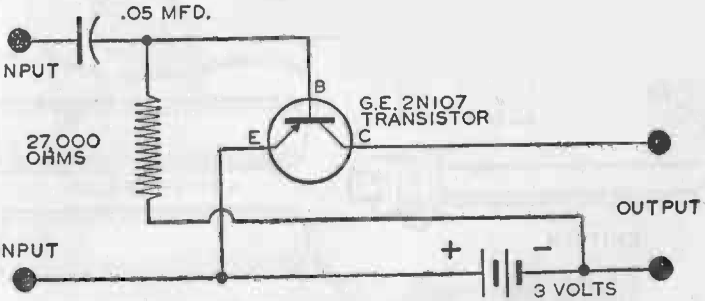
.05 MFD.
INPUT
B
G.E. 2N107 TRANSISTOR
E
C
27,000 OHMS
OUTPUT
INPUT
+
−
3 VOLTS

SCHEMATIC CIRCUIT DIAGRAM OF
ONE-STAGE AUDIO AMPLIFIER EMPLOYING
G. E. 2N107 TRANSISTOR

The diagram shows an .05 mfd. ceramic capacitor which does not appear in the previous circuit diagram.

Here is a list of the parts and materials needed to construct the amplifier:

1	Wood base 4 in. x 3 in. x ¾ in. (1)
1	Raytheon CK722 P-N-P transistor (4)
4	No. 2 Fahnestock Spring Contact Clips (7)
3	No. 10 Fahnestock Spring Contact Clips (8)
2	Burgess 1½-volt No. 7 dry cells (2)
1	Transistor battery holder for two Burgess No. 7 dry cells or 1 strip of thin metal sheet 2⁷⁄₁₆ in. x 1 in. if battery holder is to be home-made (3)
1	½-watt 390,000-ohm resistor (5)
7	⅜-in. No. 6 round head wood screws
3	⅜-in. No. 4 round head wood screws
1	Single contact tie-point (6)
	Hook-up wire

Two size AA flashlight cells can be used instead of Burgess No. 7 cells. If a factory-made transistor battery holder is used, worn out cells can be removed without disturbing any soldered connections. If you wish to make a battery holder, clamp the cells to the wood base with a strip of thin sheet metal as shown in one of the illustrations. The Burgess cells are designed especially for transistors and will last longer than flashlight cells.

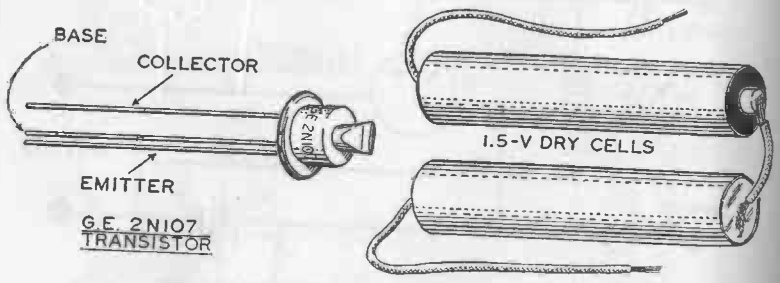

G. E. 2N107 TRANSISTOR

This is a PNP type transistor for experimental and amateur use. The right-hand sketch shows how two penlite cells are connected in series by soldering. If a G.E. 2N107 transistor cannot be obtained easily, a G.E. 2N191 or a G.E. 2N323 should give the same results.

Assembling. Give all surfaces of the wood base a coat of white shellac. When the shellac is dry assemble the parts on the base as shown in the plan.

Although the plan drawing shows wires attached to the spring clips by soldering, it is easy to wire the amplifier without any soldered connections except those to the battery or battery holder. Remove the insulation from both

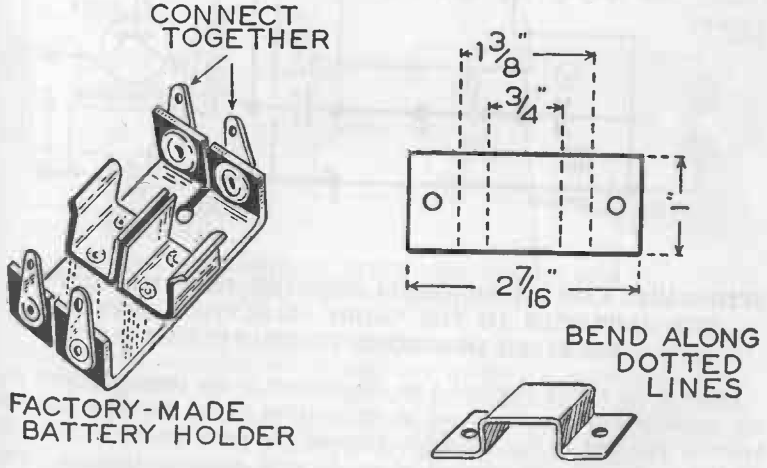

BATTERY HOLDERS FOR TWO 1½-v No. 7 CELLS

The right-hand sketch shows how to make a battery holder for clamping a two-cell battery to a wooden base. The holder is bent from a strip of sheet metal.

ends of each wire, and clamp the bare ends between the wood base and the spring clips. The single-contact tie-point can be omitted. Instead of soldering one terminal of the resistor and two wires to the tie-point, clamp them under a washer held firmly against the wood base by a round head screw which passes through the hole in the washer and into the base. When a factory-made battery holder is used, the cells should be placed in the holder so that they are in series and the terminal voltage of the battery is 3 volts. The positive end of the left-hand cell should be at the front of the holder and the positive end of the

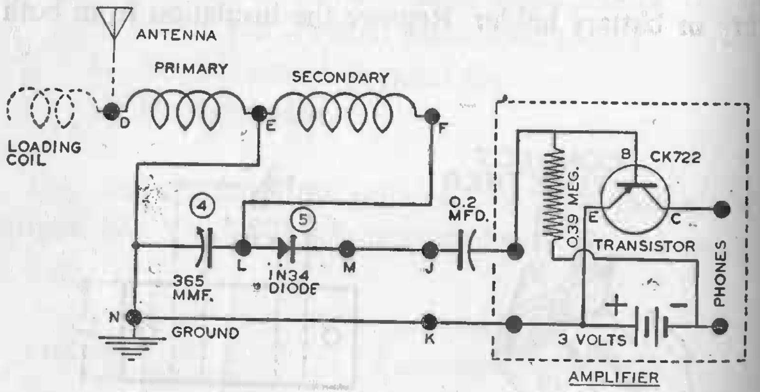

SCHEMATIC CIRCUIT DIAGRAM SHOWING HOW TO CONNECT THE AMPLIFIER TO THE "MORE SELECTIVE" CRYSTAL RECEIVER DESCRIBED IN CHAPTER III

Connect the PHONE terminal J on the receiver to the INPUT terminal on the amplifier which is connected to the EMITTER of the transistor and the POSITIVE terminal of the amplifier. Connect the other PHONE terminal J on the receiver to one terminal of an .02 mfd. ceramic capacitor. Connect the other terminal of the capacitor to the INPUT terminal on the amplifier which connects with the resistor and the base of the transistor. If a ceramic capacitor is not easily obtainable, a paper by-pass capacitor of the same capacitance can be used. The ceramic capacitor is preferable. Connect headphones to the PHONES or OUTPUT terminals of the amplifier.

right-hand cell should be at the back of the holder. Observe the indicated polarity carefully or possible damage to the transistor may occur.

Consult the plan and both circuit diagrams when wiring the amplifier. The center wire on a Raytheon CK722 transistor is the BASE terminal. The wire nearest the red dot on one side of the transistor is the COLLECTOR terminal. The remaining wire is the EMITTER terminal. The transistor terminals must be connected to the Fahnestock clips as shown in the diagrams. The letters **E, B** and **C** in the illustrations are abreviations for EMITTER, BASE and COLLECTOR *respectively.* Observe the transistor connections carefully. Possible damage to the resistor may result if it is not connected in the circuit properly.

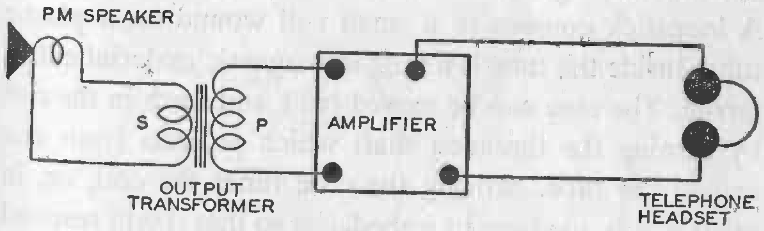

AN EXPERIMENT WITH THE ONE-STAGE TRANSISTOR AMPLIFIER

The gain or amplifying power of the amplifier may be demonstrated by hooking up the amplifier with a small PM speaker, a transformer and a pair of headphones. Connect the voice coil of the speaker to the secondary of a small output or matching transformer. Connect the primary of the transformer to the input terminals of the amplifier. Connect a 2,000-ohm headset to the output terminals of the amplifier. The speaker will act as a self-powered microphone. The wires connecting the headphones to the speaker should be long enough so that the speaker and amplifier are in one room and the phones can be carried to an adjoining room. Words spoken within a few feet of the speaker can be heard in the earphones in the adjoining room.

YOU CAN BUILD THIS COMBINATION CRYSTAL DIODE AND TRANSISTOR RADIO RECEIVER

This receiver is capable of providing good signals from broadcast stations within a radius of 50 miles using a 50-foot antenna and a first-class ground connection. In some locations, signals from stations 500 miles away can be picked up.

The antenna circuit includes a loopstick which can be used to select the station and a miniaturized 365 mmfd. variable capacitor to tune the station selected to a peak. A 1N34 diode is used as the detector.

There are several loopsticks on the market. They are intended as replacements for loop antennas in superheterodyne and other receivers provided with radio-frequency amplification. When used with a crystal detector and antenna a ground must be connected to the loopstick. A loopstick consists of a small coil wound on a plastic tube. Inside the tube is a plug of magnetic material called ferrite. The core can be moved back and forth in the coil by turning the threaded shaft which projects from one end of the tube. Moving the core tunes the coil, or, in other words, changes its impedance so that it will respond to different frequencies. A loopstick called a "tenna-loop" was used to build the receiver which is illustrated here. There are several other brands of equally good loopsticks, such as the "ferri-loopstick," "vari-loop" and Miller No. 6300 which can be used instead. Each loopstick comes equipped with a metal bracket. At one end of each loopstick is a metal ferrule. The loopstick is fastened to the bracket by pushing the ferrule into the large hole in the bracket. The bracket provided with a loopstick is usually

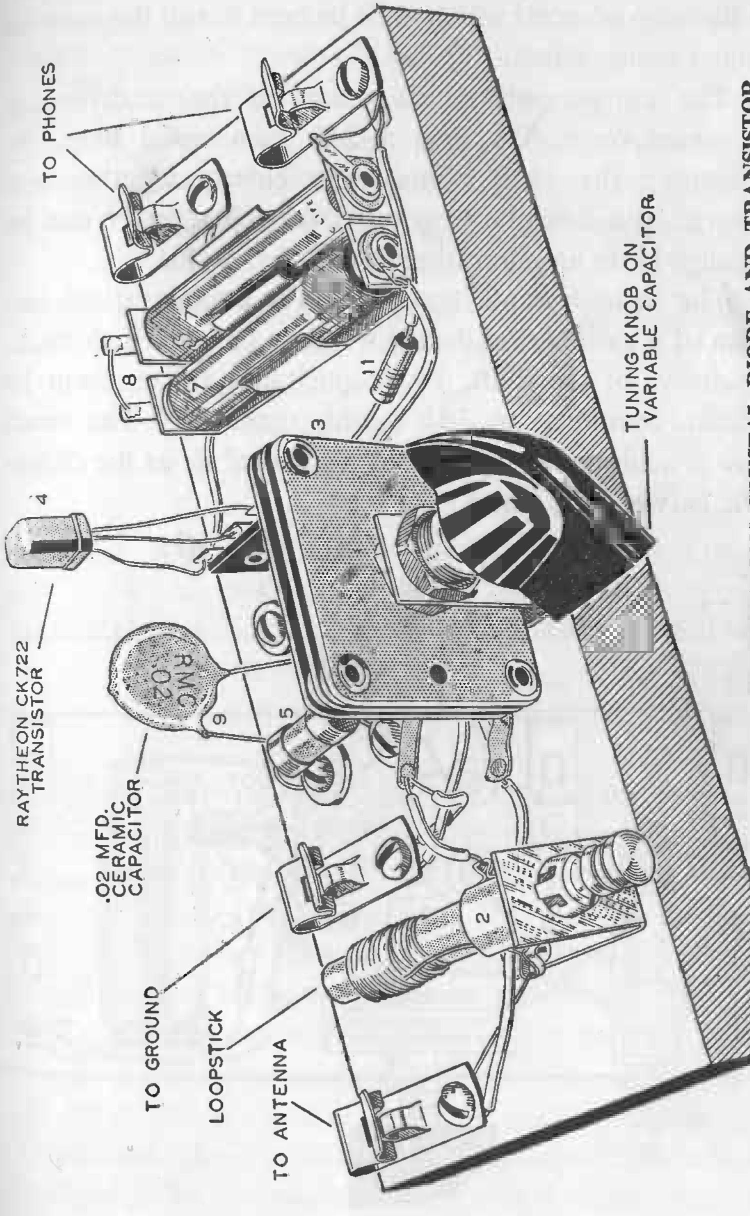
TO PHONES
TUNING KNOB ON VARIABLE CAPACITOR
RAYTHEON CK722 TRANSISTOR
.02 MFD. CERAMIC CAPACITOR
RMC .02
TO GROUND
LOOPSTICK
TO ANTENNA
1
2
3
4
5
6
7
8
11

COMBINATION CRYSTAL DIODE AND TRANSISTOR RADIO RECEIVER WHICH YOU CAN BUILD a flat strip of metal which must be bent to suit the receiver under construction.

The components which comprise the receiver are mounted on a 6⅜ inch x 2¾ inch wood base. By crowding the components more closely together and providing a box-like wooden cover the receiver can be changed into an interesting pocket size radio.

The variable capacitor specified is a miniaturized version of a variable condenser which is only 3/16 inch thick, exclusive of the shaft. The capacitance ranges from 10 mmfd. minimum to 365 mmfd. maximum. The small size is achieved by using mica instead of air as the dielectric between the plates.

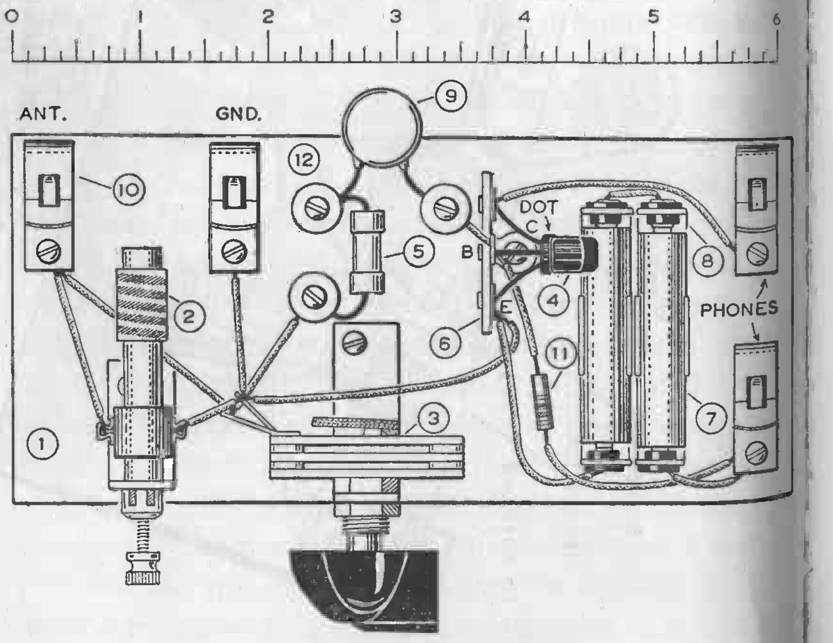

PLAN OF THE RECEIVER

PARTS NEEDED TO BUILD THE RECEIVER

1	Wood base 6⅜ in. x 2¾ in. x ¾ in. **(1)**
1	Loopstick antenna and bracket with small knob to fit threaded shaft on loopstick **(2)**
1	Miniaturized 365 mmfd. variable capacitor knob and bracket **(3)**
1	Raytheon CK722 transistor **(4)**
1	Sylvania IN34 diode crystal **(5)**
1	3-contact tie-point **(6)**
2	No. 7 Burgess dry cells **(7)**
1	Battery holder for cells **(8)**
1	.02 mfd. ceramic capacitor **(9)**
4	No. 2 Fahnestock spring contact clips **(10)**
1	½-watt, 390,000-ohm resistor **(11)**
3	No. 4, ⅜ in., round head wood screws with washers to fit **(12)**
9	No. 6, ⅜ in. round head wood screws

Assembling and Wiring. Mount the parts on the wood base as shown in the plan. The loopstick comes with a short length of bare or enamelled wire attached to one of its terminals. Discard the wire. The terminal to which the wire is attached should be considered the ANT terminal and connected to the stator of the variable capacitor and the ANTENNA Fahnestock clip. The other terminal of the loopstick is the GND terminal and should be connected to the rotor of the variable capacitor and to the GROUND Fahnestock clip. Connections to the loopstick should be soldered. The terminals of the crystal diode and the ceramic capacitor are clamped under brass washers held firmly against the base by a No. 4 round head brass screw through each washer and into the base.

Note that the terminals of the transistor are soldered to a 3-contact tie-point. If you use this method of construction, grasp each transistor terminal in the jaws of a pair of long nose pliers while soldering. Use a very small

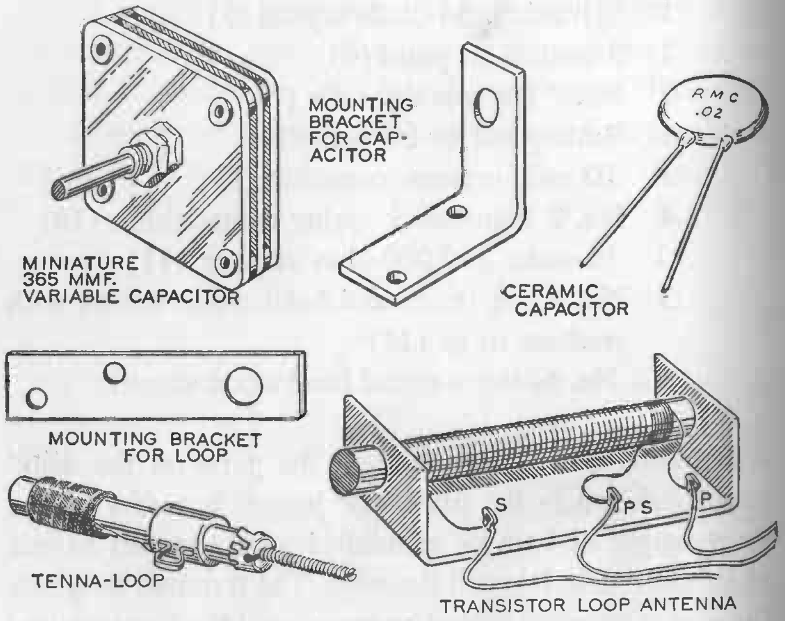
MINIATURE
365 MMF.
VARIABLE CAPACITOR
MOUNTING
BRACKET
FOR CAP-
ACITOR
R M C
.02
CERAMIC
CAPACITOR
MOUNTING BRACKET
FOR LOOP
S
PS
P
TENNA-LOOP
TRANSISTOR LOOP ANTENNA

PARTS FOR TRANSISTOR CIRCUITS

The miniature variable capacitor uses thin sheets of plastic as the dielectric between plates. It is not as efficient as an air dielectric capacitor. Ceramic capacitors are much smaller than paper capacitors of equal capacitance. The transistor loop antenna at the lower right has a non-adjustable ferrite core. It is provided with two windings. **S** is one of the secondary terminals. **P** is a primary terminal. **PS** is a common primary and secondary terminal. The loopstick or tenna-loop at the lower left has an adjustable ferrite core. The winding is tuned by turning the projecting screw.

soldering iron. Transistors are easily damaged by heat and the purpose of the pliers is to draw the heat from the wire terminal and absorb it so that it cannot reach the transistor. If you do not wish to take the risk involved in soldering the transistor terminals you can clamp them under brass washers held to the base by round head screws.

Operation. Connect the ANT clip to a 50-foot antenna and connect the GND clip to a good ground. Connect a 2000-3000-ohm radio headset to the PHONE clips. Place the two dry cells in their holder. Observe the indicated polarity, or damage to the transistor may occur.

Tune the loopstick by turning the threaded shaft which moves the ferrite plug until the desired station is heard. Then tune the variable capacitor until the signal is as loud as it is possible to make it.

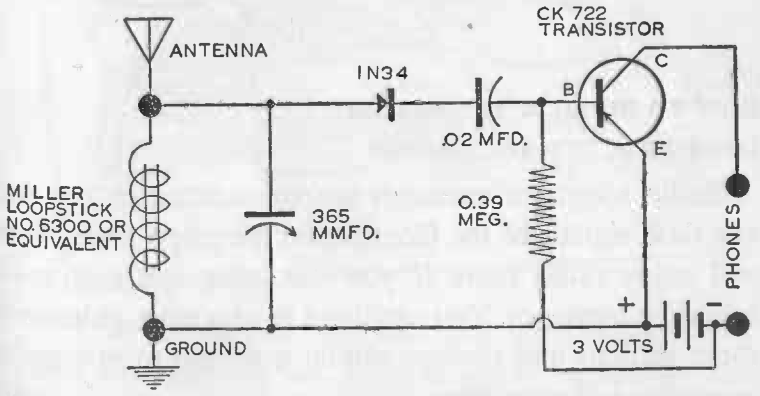
CK 722
TRANSISTOR
ANTENNA
IN34
C
B
E
.02 MFD.
MILLER
LOOPSTICK
NO. 6300 OR
EQUIVALENT
365
MMFD.
0.39
MEG.
PHONES
+
-
GROUND
3 VOLTS

SCHEMATIC CIRCUIT DIAGRAM OF THE COMBINATION CRYSTAL DIODE AND TRANSISTOR RADIO RECEIVER

Trouble Shooting. If no signal is heard check the wiring and connections carefully, the polarity of the transistor and the polarity of the dry cells.

PICTORIAL CIRCUIT DIAGRAM OF THE COMBINATION CRYSTAL DIODE AND TRANSISTOR RADIO RECEIVER

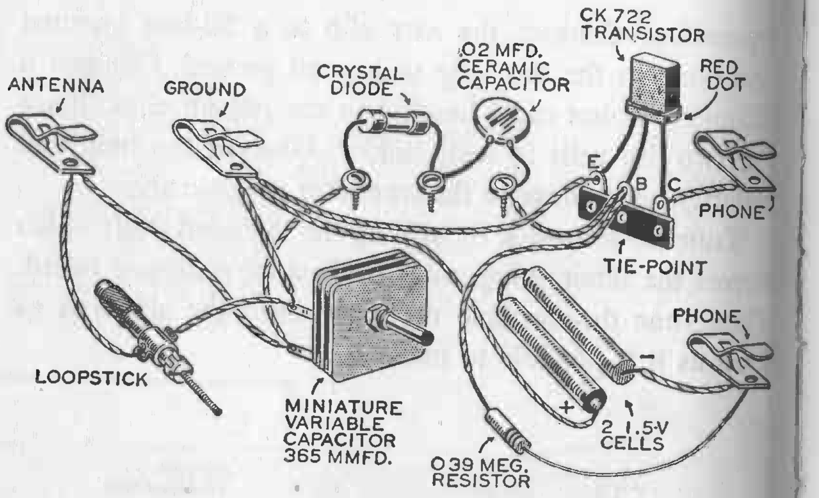
CK 722 TRANSISTOR
.02 MFD. CERAMIC CAPACITOR
CRYSTAL DIODE
RED DOT
ANTENNA
GROUND
E
B
C
PHONE
TIE-POINT
PHONE
LOOPSTICK
+
−
MINIATURE VARIABLE CAPACITOR 365 MMFD.
0.39 MEG. RESISTOR
2 1.5-V CELLS

HOW TO BUILD A TRANSISTOR CODE PRACTICE OSCILLATOR

Radio telegraph messages are transmitted in the dot and dash signals of the Continental telegraph code. You will enjoy radio more if you can telegraph and read telegraph messages. You can listen to amateurs, ships and shore stations and perhaps obtain a license to operate a transmitter of your own.

Anyone who wishes to obtain an amateur license must pass an examination in reading the telegraph code. No fee is required for the license. It is necessary only to pass a simple examination in elementary radio theory, in the Federal Radio Regulations and in sending and receiving the Continental telegraph code.

When you can send and receive 65 letters, numerals and puctuation marks per minute accurately you can pass the government code test for an amateur license. During the examination, the code test is given first. If the applicant passes this, he may proceed with the examination covering the radio laws and simple technical problems.

The first step in learning to telegraph is to memorize

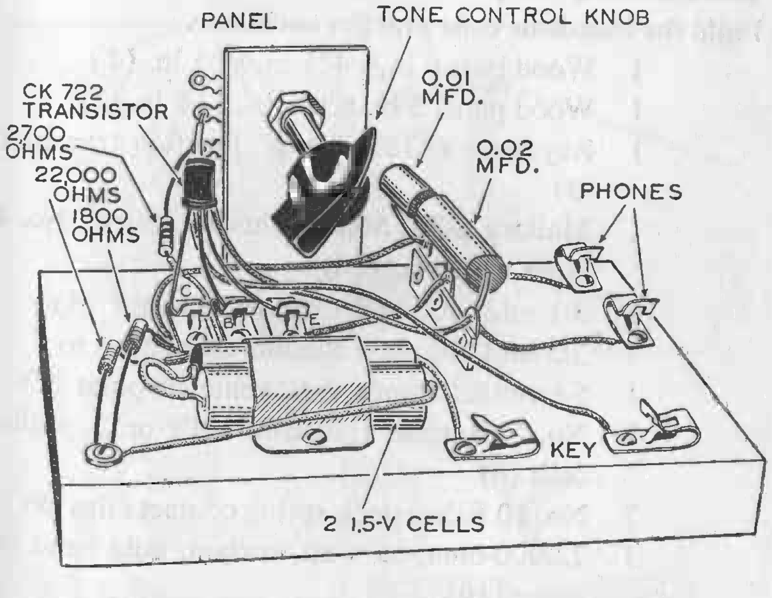
PANEL
TONE CONTROL KNOB
CK 722 TRANSISTOR
0.01 MFD.
2700 OHMS
0.02 MFD.
22,000 OHMS
1800 OHMS
PHONES
C
B
E
KEY
2 1.5-V CELLS

A TRANSISTOR CODE-PRACTICE OSCILLATOR

When the 50,000-ohm tone control is properly adjusted the circuit oscillates at an audio-frequency and sounds are produced by the headphones connected to the circuit. Adjusting the control varies the frequency and an exact imitation of the sound of radiotelegraph signals can be obtained.

the code. The second is to practice sending with a telegraph key. The simplest way to do this is to connect a telegraph key in series with a small high-pitched buzzer and a battery so that you can make DIT and DAH sounds on the buzzer by pressing the key. A code practice oscillator with a key and headphones is better than a key and buzzer. The DIT and DAH sounds are exactly like those heard in a radio receiver.

A compact, low-cost oscillator can be built with a transistor and a few other components.

The following components and materials are required to build the transistor code practice oscillator.

1	Wood base 6 in. x 4¾ in. x ¾ in. **(1)**
1	Wood panel 3 in. x 1½ in. x ⅛ in. **(2)**
1	Raytheon CK722 "p-n-p" junction transistor **(3)**
1	**Mallory U-35, 50,000-ohm midgetrol, No. 4 taper with knob (4)**
1	.01 mfd. 400 volt tubular capacitor **(5)**
1	.02 mfd. 400 volt tubular capacitor **(6)**
1	5-terminal standard Bakelite tie-point **(7)**
2	No. 7 Burgess transistor cells or 2 penlite cells **(8)**
7	No. 10 Fahnestock spring contact clips **(9)**
1	22,000-ohm, ½ watt, carbon, gold band resistor **(10)**
1	2,700-ohm, ½ watt, carbon, gold band resistor **(11)**
1	1,800, ½ watt, carbon, gold band resistor **(12)**

11	No. 4, ⅜-in. round head wood screws
1	Strip sheet metal 3 in. x 1 in. or a battery holder for 2 No. 7 cells

Assembling. Shellac or varnish the wood base and the small panel upon which the volume control is to be mounted. Fasten the Bakelite tie-points and the Fahnestock contact clips to the base in the locations indicated in the plan. Place a solder lug under the head of each

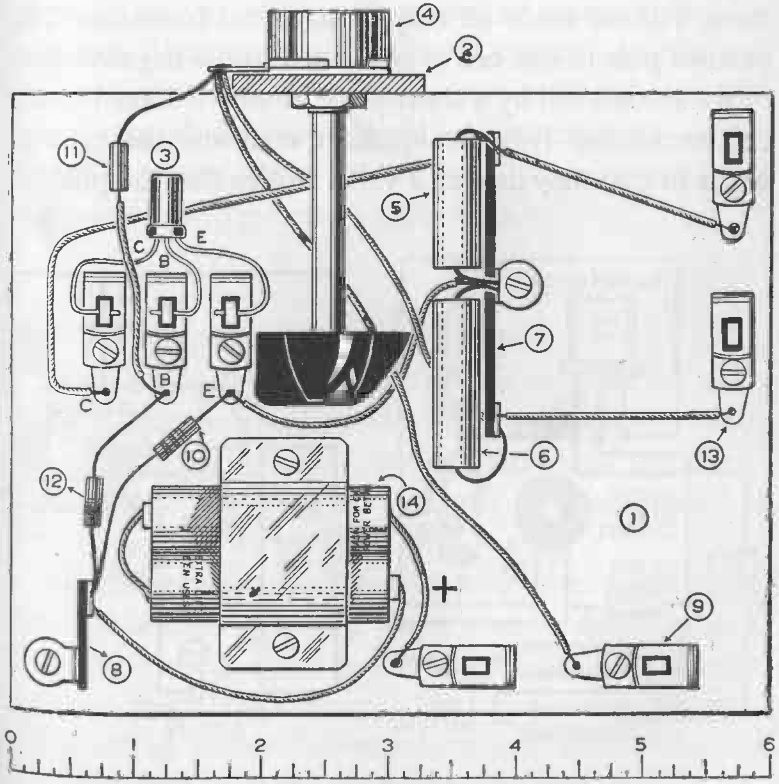

PLAN OF THE CODE PRACTICE OSCILLATOR

screw used to fasten a contact clip. The lugs make it easy to solder a connection to the clips.

Cut the strip of sheet metal 3 inches long and 1 inch wide from a tin can and use the strip to make a battery clamp as shown in one of the illustrations.

Wiring. Use pushback wiring for the connections and solder all wire terminals with rosin-core solder except the terminals of the transistor. Consult the schematic wiring diagram and also the wiring in the plan. Install the dry cells and observe the polarities indicated in the illustrations. Failure to do so may damage the transistor. The positive pole of one cell is connected to the negative pole of the second cell by a short piece of wire soldered to the cell terminals. This arrangement connects the cells in series so that they deliver 3 volts. Notice that the positive pole of the 3-volt battery is connected to the EMITTER of the transistor through the 22,000-ohm resistor. The COLLECTOR terminal on the transistor is marked by a RED dot. The wire terminal in the center is the BASE terminal. The wire terminal on the side of the transistor opposite to that marked by the red dot is the EMITTER terminal. The transistor is connected in the circuit by fastening the ends of the terminal wires in the three Fahnestock clips marked **E**, **B** and **C** in the diagrams. One terminal wire on the 2,700-ohm resistor which is connected to Fahnestock clip (**B**) and a terminal of the volume control is close to Fahnestock clip **C** but must not be allowed to touch the clip and form a short circuit. If you pull a piece of insulation 1¼ inches long off a piece of pushback wire and slip it over the wire terminal of the resistor, it will form an insulating sleeve which will prevent a short circuit.

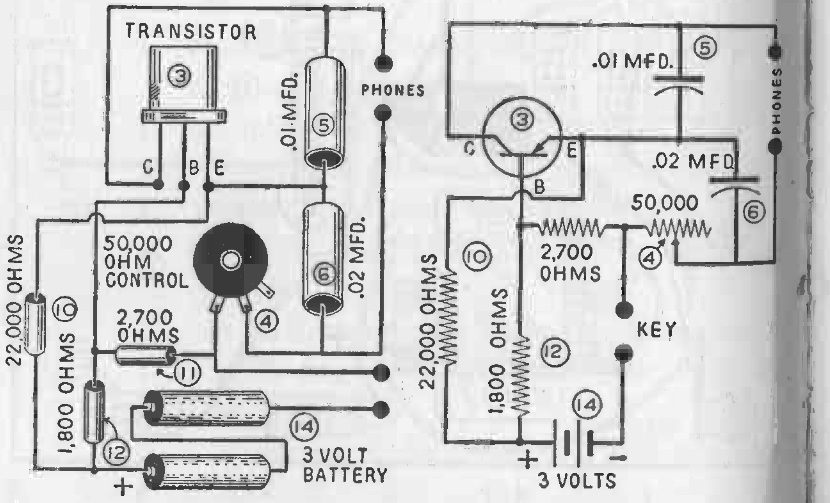

PICTORIAL AND SCHEMATIC CIRCUITS OF THE CODE PRACTICE OSCILLATOR

LEARNING TO TELEGRAPH

When memorizing the code do not think of the characters as dots and dashes but rather as sounds in which the sound DIT represents a dot and the sound DAH represents a dash. For example, the letter A is memorized as DIT DAH and not as "dot dash."

The DIT sound is made in the headphones connected to the code practice oscillator by pressing the key down and letting it come up again just as soon as possible. The DAH sound is made by pressing the key down and holding it down for an instant before releasing it so that the DAH sound is three times as long as the DIT sound.

Proper adjustment of the telegraph key is important.

The key contacts should be adjusted for the beginner so that the space between is about 0.01 inch. If the key lever has more up and down movement than is provided by a space of 0.01 inch between the contacts, the signals sent with it will be "choppy."

The key should be mounted on a table far enough back from the edge so that there is room for the elbow to rest on the table. Learn in the beginning to handle the key properly. Let the fingers rest lightly against the key knob. The thumb should be against the left side of the knob (unless you are a southpaw). The first finger should be

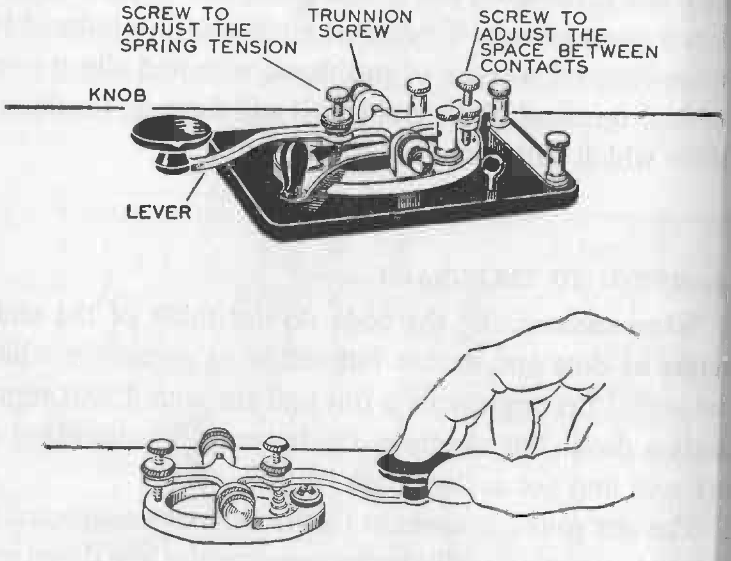
SCREW TO ADJUST THE SPRING TENSION
TRUNNION SCREW
SCREW TO ADJUST THE SPACE BETWEEN CONTACTS
KNOB
LEVER

TELEGRAPH KEYS

Telegraph keys similar to those illustrated above are available from dealers in "ham" radio supplies. The drawing at the bottom shows a method of holding the key knob used by many professional telegraph operators. The forefinger rests on top of the key knob. The thumb and second finger grasp the edges of the knob and help the spring to raise the level. The thumb and both fingers are used to press the knob down.

THE CONTINENTAL CODE

A ●▬
B ▬●●●
C ▬●▬●
D ▬●●
E ●
F ●●▬●
G ▬▬●
H ●●●●
I ●●
J ●▬▬▬
K ▬●▬
L ●▬●●
M ▬▬
N ▬●
O ▬▬▬
P ●▬▬●
Q ▬▬●▬
R ●▬●
S ●●●
T ▬
U ●●▬
V ●●●▬
W ●▬▬
X ▬●●▬
Y ▬●▬▬
Z ▬▬●●
1 ●▬▬▬▬
2 ●●▬▬▬
3 ●●●▬▬
4 ●●●●▬
5 ●●●●●
6 ▬●●●●
7 ▬▬●●●
8 ▬▬▬●●
9 ▬▬▬▬●
0 ▬▬▬▬▬
Call ▬●▬●▬
Invitation to Transmit ▬●▬
Period ●▬●▬●▬
Interrogation ●●▬▬●●
Break ▬●●●▬
Wait ●▬●●●
End of message ●▬●▬●
End of transmission ●●●▬●▬
Received ●▬●
Go ahead ▬●▬
Comma ▬▬●●▬▬

THE CONTINENTAL TELEGRAPH ALPHABET USED IN RADIOTELEGRAPHY

partly on top and against the edge of the knob. The second finger should be against the right edge of the knob. Allow the third and fourth fingers to curve naturally under the palm without tension or rigidity. Do not operate the key with the finger or wrist muscles. Use the forearm muscles. The fingers are merely the medium through which the stroke of the forearm is transmitted to the key and the wrist is merely a sort of hinge through which the motion of the forearm is transmitted to the key. Do not "tap" the key, as newspaper reporters and story writers say. A telegraph operator never "taps." An operator's thumb and fingers do not quit contact with the key knob until he is through sending. Do not think of speed at first during your practice. Take time to send slow, even clean-cut signals. Speed will come with sufficient practice.

MEMORIZING GROUPS

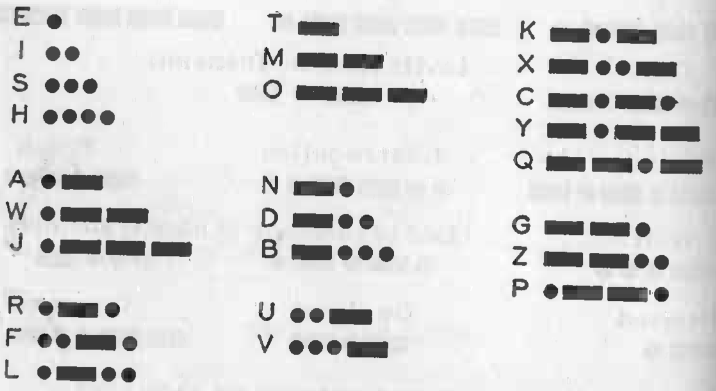

There are two systems of memorizing the telegraph alphabet. One is to study the first four or five letters, then the next four or five, and so on. The other is to study the characters after they have been arranged in easily identifiable "memory groups" as above.

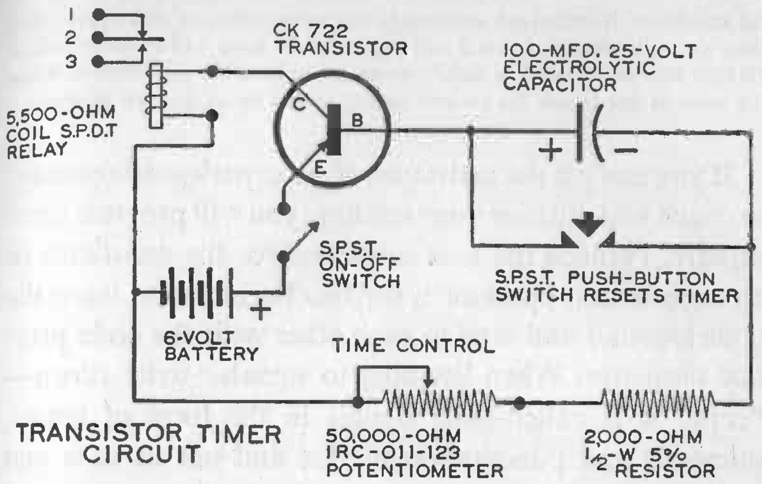
1
2
3
CK 722 TRANSISTOR
100-MFD. 25-VOLT ELECTROLYTIC CAPACITOR
5,500-OHM COIL S.P.D.T RELAY
C
B
E
+
−
S.P.S.T. ON-OFF SWITCH
S.P.S.T. PUSH-BUTTON SWITCH RESETS TIMER
6-VOLT BATTERY
TIME CONTROL
50,000-OHM IRC-011-123 POTENTIOMETER
2,000-OHM ½-W 5% RESISTOR

TRANSISTOR TIMER CIRCUIT

AN OPPORTUNITY TO DO YOUR OWN ENGINEERING

This is a schematic circuit for a timer switch which employs a transistor in its operation. The timer will open or close a switch after a definite period of time has elapsed. It can be used to turn off a photo-printing lamp after the lamp has burned a required number of seconds. A pictorial circuit, plan, list of components and detailed instructions have been omitted so that the young electronics experimenter can make his own layout from the information which can be gained by studying the schematic circuit.

With proper adjustment of the potentiometer and the spring on the relay armature, the relay can be made to operate at the end of any time period ranging from 2 to 40 seconds. A paper dial should be mounted under the knob and pointer on the potentiometer shaft and marked to indicate the time periods. Use a watch with a sweep second hand to test different settings of the potentiometer. Mark the dial to indicate seconds.

If the relay is connected to a lamp used for photo-printing, the lamp will remain lighted for periods of 2 to 40 seconds in accordance with the time indicated on the scale by the pointer. Use the relay terminals marked 1 and 2 in the circuit diagram as part of the lamp circuit. When the ON-OFF switch is closed and the push-button reset is pressed, the 100-mfd. electrolytic condenser is discharged and the relay armature is drawn toward the relay coil. This movement of the armature closes contacts 1 and 2 and lights the photo-printing lamp. As soon as the condenser is discharged by pressing the push button, the 6-volt battery begins to recharge it again. The time required to recharge the condenser depends upon the amount of resistance in the circuit, in other words, the adjustment of the potentiometer. When the condenser is recharged sufficiently the relay armature will move away from the coil, contacts 1 and 2 will open and the lamp will be extinguished. To light the lamp again it is merely necessary to press the push-button. When the timer is not in use the ON-OFF switch should be in the OFF position.

If you can get the assistance of an experienced operator to coach and criticize your sending, you will progress more rapidly. Perhaps the best substitute for the assistance of an experienced operator is for two beginners to learn the code together and send to each other with the code practice oscillator. When listening to signals, write down—"copy" it is called—the signals in the form of letters, numerals and punctuation marks, and not as dots and dashes.

There is not enough space in this book to give more than brief instruction for learning to send and receive radiotelegraph signals.

A copy of the 126-page manual called *The Radio Operator* used in training operators in the army will be found helpful and interesting to any one who wishes to learn to telegraph. Write to the Superintendent of Documents, U. S. Government Printing Office, Washington, D. C. and ask if the book is in print and its price. Ask for TM-454 *The Radio Operator*. The book is not always available. The price is usually nominal (25 cents).

The American Radio Relay League, Hartford, Connecticut, publishes an excellent instruction manual called *Learning the Radio Telegraph Code*. The price is 25 cents.

CHAPTER TEN

About Radio Tubes

The operation of all the apparatus described in the following chapters centers around some type of ELECTRON TUBE, popularly known as a radio tube. An electron tube consists of two or more metal electrodes sealed within a totally or partially evacuated metal or glass envelope. Since radio tubes are the "heart" of most electronic apparatus, it is useful to have some knowledge of how they work and of the circuits which employ them.

Classification of Tubes. There are more than one thousand different types of vacuum tubes regularly manufactured in the United States. This wide variety may be classified in a number of ways. A simple method is to group them according to their use—that is, whether they were designed to be DETECTORS, AMPLIFIERS, OSCILLATORS or RECTIFIERS. Another method is to divide them into groups having such family names as DIODE, TRIODE, TETRODE and PENTODE according to the number of electrodes they employ, DI, TRI, TETR and PENT are from Greek words meaning respectively *two, three, four* and *five*. A DIODE has two electrodes, a TRIODE has three electrodes and so on, the name of the tube thus revealing how many electrodes are enclosed in its glass or metal envelope.

Electron Emission. Interesting things happen inside a radio tube when it operates. These cannot be seen but they can be visualized by the mind when we know certain facts about electrons. A radio tube is essentially a device for moving electrons and utilizing their movement. The operation of all, except a few specialized forms of electron tubes, depends upon the phenomenon of ELECTRON EMISSION from a heated electrode.

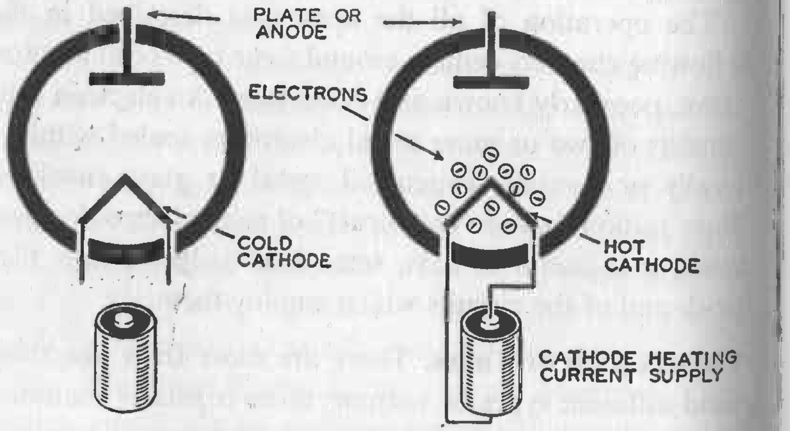

RADIO TUBES ARE ABLE TO FUNCTION BECAUSE OF ELECTRON EMISSION FROM A HEATED FILAMENT

A source of free electrons is essential for the operation of every radio vacuum tube. The electrons are supplied by using heat to drive them out of a filament or a sleeve surrounding a filament. The filament, or sleeve, which supplies the electrons is called the CATHODE. The cathode is heated by an electric current. When the cathode is cold no electrons are emitted.

To get a mental picture of electron emission, consider for a moment the tungsten filament of an incandescent lamp. You probably know that the tungsten is made up of atoms of tungsten clustered together in molecules and that scattered about in the spaces between the molecules are many "free" electrons. The free electrons are those which are not part of any of the atoms. At a temperature of absolute zero there is no motion of the atoms, molecules and free electrons. At higher temperatures there is random motion of these elementary particles. As the temperature increases, the velocity or speed of the motion increases. When an electric current is sent through a tungsten filament the filament is heated. At a dull red heat some of the electrons move with sufficient speed to break through the surface of the hot filament. The filament of an incandescent electric light, when heated to incandescence by the passage of an electric current, emits not only light but also electrons. Ordinarily the electrons do not travel far away from the filament. If undisturbed, they usually curve around and re-enter the filament.

All radio tubes contain a filament or heater similar to the filament of an incandescent lamp. When an electric current passes through the filament and sufficient heat is developed, electron emission occurs.

The Cathode. In some electron tubes, the electrons may be emitted directly from the filament, in which case it forms the CATHODE of the tube. In others, the filament may merely supply heat to a chemically treated sleeve which emits the electrons. In that case, the filament is called the HEATER and the sleeve is the CATHODE.

The Diode. The first radio tubes were diodes. They were called Fleming valves. John A. Fleming, a British scientist, found the first practical application for the diode. He patented its use as a detector of wireless telegraph signals and it was thus employed for several years by the Marconi Company.

A diode contains two electrodes, a cathode in the form of a filament and an anode in the form of a metal cylinder or plate. The anode of a radio tube is commonly called the PLATE in radio language, The cathode is an EMITTER of electrons and the plate is a COLLECTOR of electrons. A diode is an electrical valve in which electrons can flow only from the cathode to the anode. This action is explained by the illustration on page 129 and its caption.

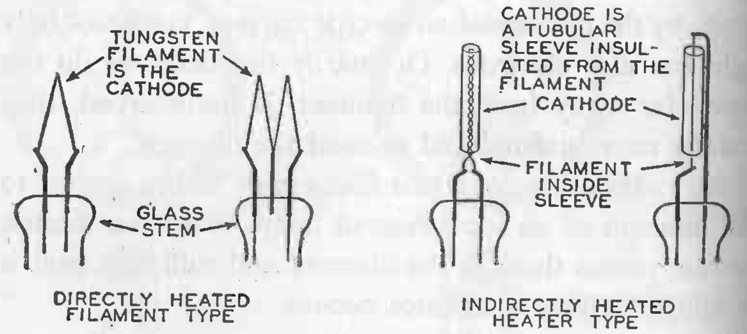
TUNGSTEN FILAMENT IS THE CATHODE
GLASS STEM
CATHODE IS A TUBULAR SLEEVE INSULATED FROM THE FILAMENT
CATHODE
FILAMENT INSIDE SLEEVE
DIRECTLY HEATED FILAMENT TYPE
INDIRECTLY HEATED HEATER TYPE

TWO KINDS OF RADIO-TUBE CATHODES

In the directly heated type (usually employed in battery-operated receivers) the cathode is a tungsten filament. The electrons are emitted directly from the hot tungsten, which is heated by an electric current passing through it.

In the indirectly heated type, the tungsten filament is enclosed in a chemically treated sleeve. The sleeve emits electrons when heated by the filament. The directly heated cathode requires direct current. An indirectly heated cathode may be operated on alternating current.

Since electrons can flow in only one direction through a diode, its basic use is as a rectifier. A rectifier changes alternating current into direct current.

The unidirectional flow of current through a diode is used in principle when a diode is employed as a DETECTOR. Diodes in which the filament is replaced by a combination of filament and cathode and which have been evacuated until only a few molecules of gas remain in the envelope, are used as detectors in some modern radios. Frequency modulation receivers use a diode detector.

Other diodes containing a small amount of mercury vapor to increase their current carrying capacity are used as rectifier tubes to change alternating current into direct

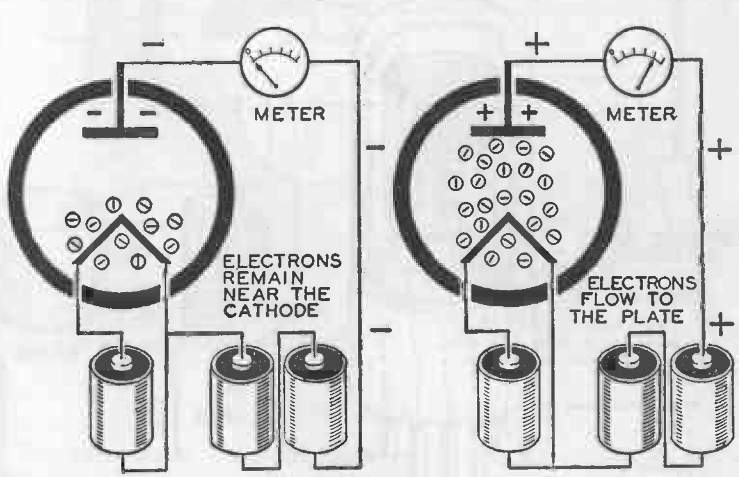
METER
ELECTRONS REMAIN NEAR THE CATHODE
METER
ELECTRONS FLOW TO THE PLATE

THE FUNDAMENTAL ACTION WHICH OCCURS IN A RADIO TUBE

If the plate is neutral or negative in respect to the cathode, the electrons emitted from the cathode remain close to the cathode. Some electrons reenter it. None reaches the plate and no current flows in plate circuit. (See meter in illustration.) If the plate is positive, electrons are drawn to it and pass across the vacuum space between the cathode and plate. Under this condition a current flows in the plate circuit. (See meter.)

current. Direct current is required in the plate circuits of all radio receivers and amplifiers, and in AC-operated receivers is secured by rectifying the AC.

The Triode. The greatest advance in the development of radio tubes was made in 1906 when Lee DeForest added a third electrode to the diode and it became a triode. The triode is similar in construction to the diode, except that a grid of fine wire or screen with openings through which electrons may pass is added between the cathode and plate. The addition of the grid gives to the tube its most

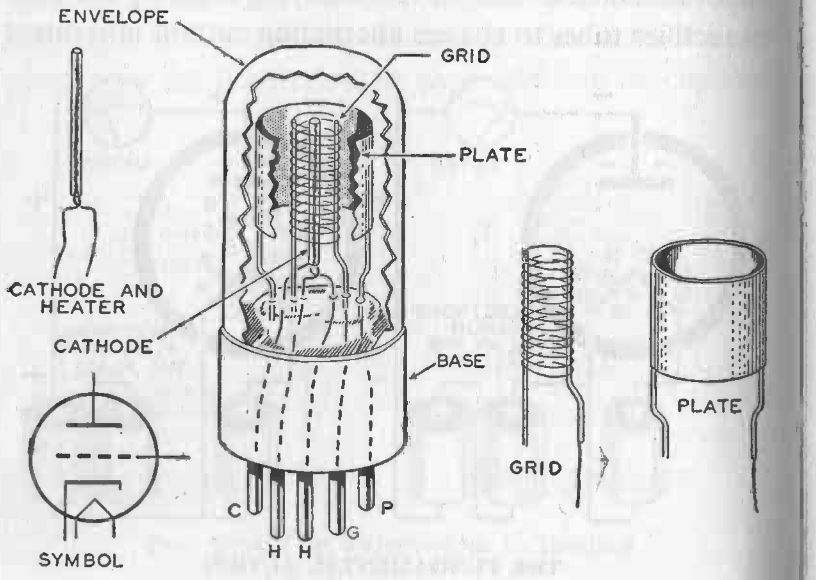
ENVELOPE
GRID
PLATE
CATHODE AND HEATER
CATHODE
BASE
PLATE
GRID
C
H
H
G
P
SYMBOL

THE CONSTRUCTION OF A TYPICAL TRIODE

Part of the glass envelope and part of the plate are shown cut away in this illustration to reveal the arrangement of the electrodes. Electrical connections to the filament, cathode, grid and plate are made through wires which connect to the base pins. The cathode sleeve is insulated from the filament. Notice that the grid is much closer to the cathode than to the plate.

useful functions—the ability to amplify. The grid is called the CONTROL GRID. It makes it possible to control relatively large electron current across the space between the cathode and anode by very small variations in voltage applied to the grid. This can be accomplished without the expenditure of appreciable power in the control circuit. A triode may be used as a sensitive detector, a relay, an amplifier or an oscillator.

When a triode is placed in operation, the plate is con-

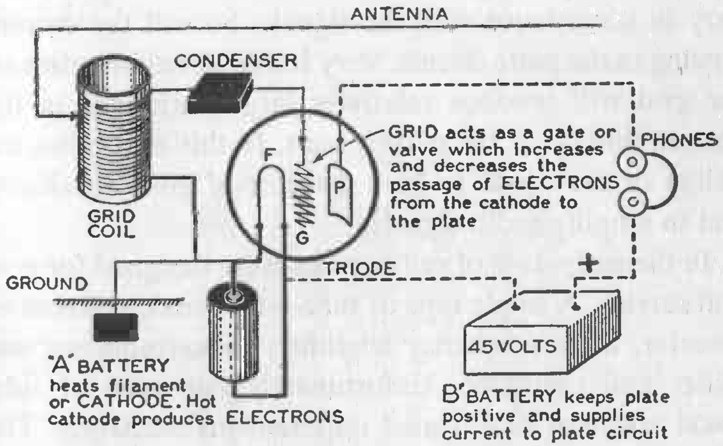
ANTENNA
CONDENSER
GRID acts as a gate or valve which increases and decreases the passage of ELECTRONS from the cathode to the plate
PHONES
F
P
G
GRID COIL
TRIODE
+
GROUND
45 VOLTS
"A" BATTERY heats filament or CATHODE. Hot cathode produces ELECTRONS
"B" BATTERY keeps plate positive and supplies current to plate circuit

THE TRIODE AS A DETECTOR

The three circuits which are necessary in order for a triode to play its part as a detector in a radio receiver are illustrated. The tube represented is the filament cathode type. The grid circuit is indicated by light lines, the filament circuit is indicated by heavy lines and the plate circuit is indicated by broken lines. The high-frequency currents generated in a receiving antenna by incoming signals are alternating currents and as a result the charge upon the grid is alternately positive and negative and varies in degree in accordance with the signals. The number of electrons passing through the grid from the cathode to the plate and likewise the volume of current flowing in the plate circuit varies in accordance with the charge on the grid.

nected to a high positive voltage, usually designated by the abbreviation B+. The grid is connected so that it is negative in respect to the cathode. Since electrons are negative, they tend to be repelled by the negative grid and few of them get through the grid to the plate.

If the triode is used as a detector or an amplifier, the signals which reach the grid are alternately positive and negative. If the grid becomes less negative, more electrons will get through to the plate. If it becomes more negative, fewer electrons will get through to the plate. The number of electrons passing from the cathode to the plate will vary in accordance with the signals. So will the current flowing in the plate circuit. Very feeble signal impulses on the grid will produce relatively large variations in the current flowing in the plate circuit. In this action lies the ability of the triode to be a detector of great sensitivity and to amplify radio signals.

In the early days of radio, tubes were designed for general service. A single type of tube—the triode—served as detector, audio-frequency amplifier, radio-frequency amplifier and oscillator. Unfortunately one type of tube could not meet such varied requirements efficiently. The splendid performance of modern radio apparatus is largely due to the fact that hundreds of different tubes are available and that each one has been developed so as to be most efficient for a certain purpose.

Tetrodes and Pentodes. There is capacitance between the cathode, grid and plate of a triode just as there is between the plates of a small capacitor. For many years when triodes were the only tubes available, the capacitance between the plate and the grid prevented the development of satisfactory radio-frequency amplifiers. The problem was finally solved by adding a SCREEN GRID or fourth electrode between the grid and plate, thus changing the triode into a TETRODE. This reduced the capacitance between the control grid and the plate to a very low value.

With tetrode tubes as radio-frequency amplifiers it became possible to obtain much more amplification than with triodes. But before long it was discovered that a tetrode was not the final answer to the problem of radio-frequency amplification.

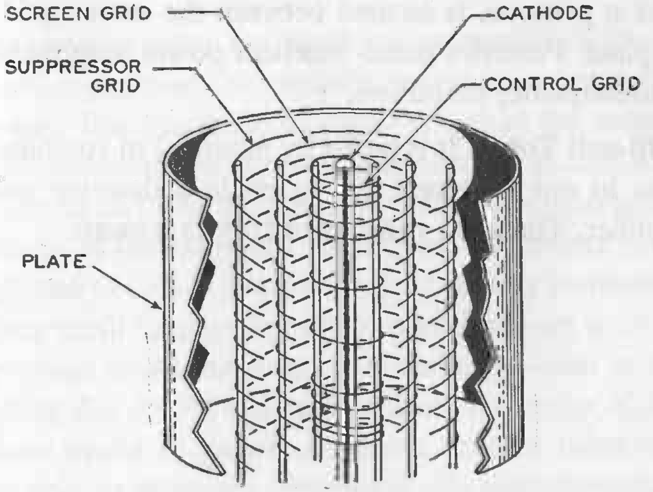

CONSTRUCTION OF A PENTODE

This is an example of the advance in tube knowledge and design since the first triode was created.

Part of the plate is shown cut away to reveal the three grids and cathode of a pentode designed to reduce distortion in radio-frequency amplifiers. Known both as a super-control amplifier tube and as a variable-mu tube, it also reduces "cross modulation." Cross modulation is the effect produced in a radio receiver by an interfering station "riding through" on the carrier wave of the station to which the receiver is tuned.

In all radio tubes the electrons emitted from the cathode and attracted to the plate may, if they strike the plate with sufficient velocity, dislodge other electrons from the plate. These dislodged electrons are called secondary electrons and their release is called secondary emission.

In a tetrode the screen grid offers an attraction to the secondary electrons and some of them are drawn to it with the effect that the plate current is lowered. It was soon found that the troublesome effect of secondary emission in lowering the plate current may be eliminated by adding a fifth electrode known as a SUPPRESSOR GRID. The suppressor grid, which added to a tetrode makes the latter a pentode, is located between the screen grid and the plate. Pentodes make excellent power amplifiers and radio-frequency amplifiers.

Multi-unit Tubes. It is perfectly practical to combine two tubes in one envelope, for example a detector and an amplifier. These are called MULTI-UNIT TUBES.

CHAPTER ELEVEN

Introduction to Amplifiers

More vacuum tubes are employed as amplifiers than for any other purpose. To amplify means to enlarge or to increase. The electronic amplifier is one of the most important inventions of the twentieth century. It is an essential part of all radio, television and radar transmitters and receivers, of the telephone system,* of the modern phonograph and of almost unnumbered laboratory instruments.

Very small changes in voltage or current—so small that they would be useless without an amplifier—can be magnified by this device and made useful. The range of radio stations would be greatly limited if vacuum tubes could serve only as detectors. We would still wear headphones when listening to radio programs. Television, radar, long-distance telephony and talking motion pictures would be dreams and not realities.

Broadcasting and television transmitters put many thousands of watts of power into their signals but the energy picked up from them by your receiver is very small. It is usually only a few millionths of a watt and too weak to do more than produce faint sound in a sensitive telephone receiver. It is too feeble to operate a loud speaker or produce a television picture unless it is fed first into an amplifier. The common electronic amplifier employs vacuum tubes or transistors so arranged that a very small power in one part of the circuit will control a much larger power in an adjoining circuit. Many electronic devices require more than one tube or transistor to secure sufficient amplification. When the output of one tube or transistor is fed to another tube or transistor the arrangement is called a CASCADE AMPLIFIER. When one amplifier tube or transistor is used to do the job, it is called a single-stage amplifier. When two tubes or two transistors are used it is a two-stage amplifier. One, two and three-stage amplifiers are common. Some amplifiers consist of five or six stages.

---

* The amplifier tubes used in telephone circuits are called REPEATERS.

A vacuum tube in itself is not an amplifier, nor is a transistor alone an amplifier. A single-stage amplifier utilizing a vacuum tube consists of the tube, a power source and three adjoining circuits. The three circuits may include resistors, transformers, capacitors, etc. (A single-stage transistor amplifier is illustrated on page 100.)

Amplifiers can be classified according to the frequency ranges over which they operate. There is no "universal" amplifier which will amplify all frequencies efficiently. The three general types are called:

1. Audio amplifiers
2. Video amplifiers
3. Radio-frequency amplifiers

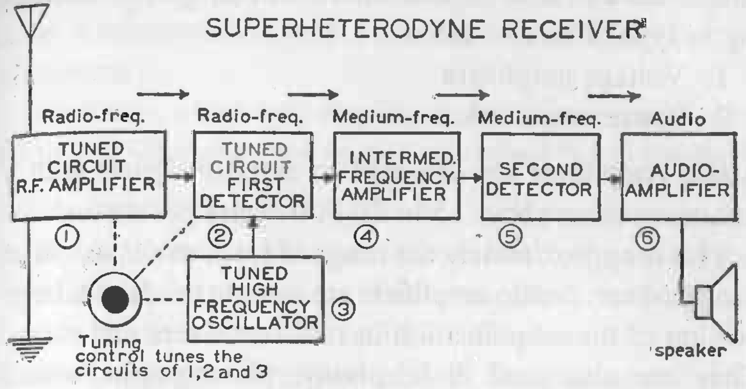
SUPERHETERODYNE RECEIVER

Radio-freq.
Radio-freq.
Medium-freq.
Medium-freq.
Audio
TUNED CIRCUIT R.F. AMPLIFIER
TUNED CIRCUIT FIRST DETECTOR
INTERMED. FREQUENCY AMPLIFIER
SECOND DETECTOR
AUDIO-AMPLIFIER
①
②
④
⑤
⑥
TUNED HIGH FREQUENCY OSCILLATOR
③
Tuning control tunes the circuits of 1, 2 and 3
speaker

WHAT IS A SUPERHETERODYNE RECEIVER?

One of the problems in designing a sensitive broadcast receiver is that it must include a radio-frequency amplifier which will operate over a wide range of frequencies. Such an amplifier cannot be as efficient as one designed for a single frequency. The higher radio-frequencies cannot be amplified as efficiently as the lower frequencies. Most home broadcast receivers overcome this difficulty by employing the circuit invented by Prof. E. H. Armstrong which he named the SUPERHETERODYNE.

The illustration above is a block diagram of a simple superheterodyne receiver. A schematic circuit diagram of a superheterodyne would be very confusing to one who is not familiar with radio circuits. Radio engineers often use block diagrams to represent intricate apparatus. A block diagram is somewhat like an architect's plan for a building. The architect's plan does not show all the individual beams, boards, pipes, bricks and other details. A radio engineer's block diagram does not show all the details of a circuit.

In the superheterodyne circuit, the frequency of an incoming signal is changed to a lower frequency which can be amplified efficiently. The radio frequency signal from the antenna is mixed in the first detector (2) with radio-frequency currents generated by a tuned high-frequency oscillator (3). The combining of the two radio-frequencies results in an entirely new signal called the I.F. signal. The tuning and mixing is done in such a manner that the I.F. signal always has a carrier-frequency of exactly 456 Kc.

After the carrier-frequency of the desired incoming signal has been changed to 456 Kc by the mixer, first detector and oscillator, it is fed into the intermediate frequency (I.F.) amplifier (4). Since the I.F. amplifier works at only one frequency (456 Kc) it is many times more efficient than an amplifier which works at many different frequencies. The greatly amplified I.F. signal then passes through an ordinary detector (5) and audio amplifier (6). The modern superheterodyne receiver is simple to tune and so sensitive that it will pick up extremely faint signals.

Amplifiers can also be classified into two groups according to type of service as:

1. Voltage amplifiers
2. Power amplifiers.

Audio Amplifiers are designed to amplify signals with a frequency from about 15 to 20,000 cycles per second.

This is approximately the range of frequencies AUDIBLE to a good ear. Audio amplifiers are used to produce a large portion of the amplification in radio receivers and sonar. They are also used in telephony, phonographs, sound motion pictures and intercom equipment. The audio amplifiers in the common radio set do not usually have a frequency range of 15 to 20,000 cycles. They commonly cover a narrower frequency band (30 to 11,000 cycles per second).

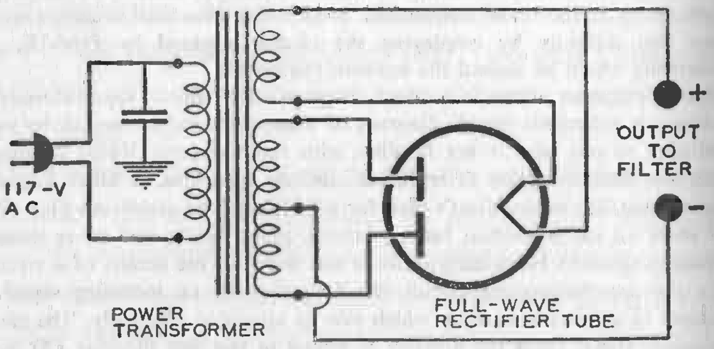

ELECTRON TUBES AS RECTIFIERS

Electron tubes are used for rectifying or changing 117-volt alternating current into direct current. The direct current used as the plate supply in radio receivers, amplifiers, transmitters, etc., is generally secured in this manner. A half-wave rectifier rectifies only one-half of each alternating-current cycle and consequently produces a pulsating direct current. A full-wave tube, connected to a "center-tapped" transformer, rectifies both halves of each alternating cycle. This produces a pulsating direct current also but the current does not have such prominent ripples as the current from a half-wave rectifier and so is easier to smooth out with a filter. (The circuit of a full-wave rectifier is illustrated.)

Amplifiers which will amplify uniformly frequencies of 15 to 20,000 cycles per second are rated "high-fidelity."

Video Amplifiers are designed to amplify uniformly the audio range and, in addition, a wide band of frequencies inaudible to human ears. Video amplifiers amplify all frequencies from about 30 to 6,000,000 cycles per second.

Radio-frequency Amplifiers. Well designed audio-frequency amplifiers amplify all the frequencies in the audio range as equally as possible. Unlike the audio and video

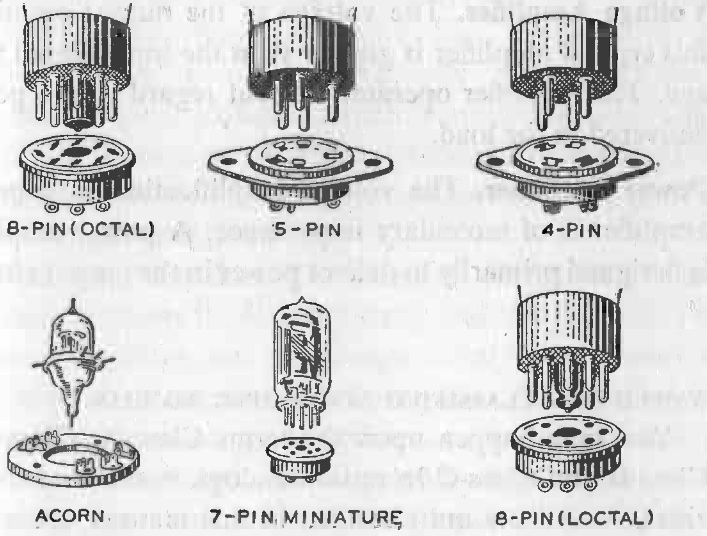

THE BASES OF RECEIVING TUBES AND THEIR CORRESPONDING SOCKETS

The electrodes in a tube are connected to terminal pins which project from the base of the tube. There are six different types of bases commonly used for receiving tubes.

amplifiers, radio-frequency amplifiers need amplify only a very narrow band of radio frequencies. This narrow band may be any part of the wide range of frequencies from 30,000 to several billion cycles per second. Radio-frequency amplifiers use resonant, or tuned, circuits whereas audio-frequency amplifiers use untuned circuits.

AMPLIFIERS CLASSIFIED ACCORDING TO USE

When amplifiers are classified according to use, they can be divided into two groups known respectively as POWER amplifiers and VOLTAGE amplifiers. You will see these two types mentioned from time to time in radio literature.

Voltage Amplifier. The voltage of the output circuit of this type of amplifier is greater than the input signal voltage. The amplifier operates without regard to the power delivered to the load.

Power Amplifier. The voltage amplification in a power amplifier is of secondary importance. A power amplifier is designed primarily to deliver power in the output circuit.

AMPLIFIERS CLASSIFIED ACCORDING TO BIAS

You may happen upon the terms Class-A, Class-AB, Class-B and Class-C in radio catalogs, books and magazines. Amplifiers are classified in this manner according to the "bias voltage applied to their grids." An explanation of grid bias and its effects involves the use of mathematical curves and diagrams which are clear only to those who have studied sufficient mathematics. If you are interested in the meaning and effects of grid bias upon the action of an amplifier tube you may be able to find an explanation in a radio book at the nearest public library.

Class-A amplifiers are used in the output stages of radio receivers and low-power audio systems. They are also used as radio-frequency amplifiers in radio receivers. Class-A amplifiers produce little distortion (an especially desirable quality in an audio amplifier). They are also characterized by low power output and low efficiency.

Class-B amplifiers are used where it is desired to develop a relatively large output in the load circuit. The grid of a class-B amplifier is biased so that plate current flows only during half of the AC signal voltage that is applied to the grid. Class-B amplifiers are used in audio amplifiers. Distortion-free amplification is desired in an audio amplifier. To obtain this quality, two class-B tubes are arranged in a push-pull circuit. Some of the tubes used for this purpose consist of two triodes in the same glass or metal envelope. Then only one socket and what appears to be a single tube is necessary.

Class-AB Amplifiers. Sometimes it is desirable to compromise between the low efficiency and low distortion of class-A amplifiers and the comparatively high distortion and high efficiency of a class-B amplifier. This can be accomplished by biasing an amplifier so that its operation is halfway between that of a class-A amplifier and a class-B amplifier. Such an amplifier is called a class-AB amplifier.

Class-C amplifiers are by far the most efficient of the four classes for radio-frequency amplification.

THE MEANING OF COUPLING

Joining one circuit to another so that energy can be transferred from one circuit to the other is called COUPLING. There are several methods of coupling.

There are two common methods of coupling a circuit to either the INPUT or OUTPUT circuit of an amplifier. One method employs a transformer and the other method uses a resistor and a capacitor. They are called respectively:

1. Transformer coupling
2. Resistance-capacitance coupling

There are two types of transformers used for transformer coupling. They are:

1. Radio-frequency transformers
2. Audio-frequency transformers

Radio-frequency transformers are usually wound on a tube and may have either an air-core or a core made of a powdered iron compound called FERRITE. The ferrite is pressed into a short, straight rod and forms the type of core known as an OPEN-CORE.

Audio-frequency transformers have closed cores made of thin sheets of transformer steel. For further description of audio-frequency transformers refer to the paragraphs on transformers in the chapter on Components.

CHAPTER TWELVE

The Photocells and Phototubes Which Are Popularly Called Electric Eyes . . . HOW TO BUILD A PHOTOTUBE RELAY

There are several varieties of transducers which generate a voltage when light shines on them. They change the energy of light into electrical energy. These devices have several names. They are called "light-sensitive cells," "photocells," "photronic cells," and "electric eyes." Photocell and photoelectric cell are the terms we will use.

Uses for Photocells. Photoelectric cells are far more sensitive than the human eye to small changes in light intensity and color. They have important uses. Here are a few:

1. to measure illumination
2. as light meters to indicate proper exposure in photography
3. to measure the intensity of photographic negatives for automatic photo-printing machines
4. to count vehicles or people passing a given point

5. to open a door when a person approaches it
6. to signal the presence of an intruder
7. to count packages or manufactured articles passing on a conveyor belt
8. to determine the amount of ingredients in a solution
9. to keep color-printing presses in "register" or adjustment
10. to send photos, graphs, drawings, etc. by wire or radio
11. as an essential part of sound motion pictures
12. to level elevators at a floor
13. to turn street lights on when darkness comes and off when the sun comes up
14. to measure high temperatures
15. to signal the boiler room when smoke issues from the chimney of a power plant

Selenium. The first photocell was made with selenium. Selenium is an element that lies between an insulator and a conductor. Many years ago small rods of selenium were used occasionally as resistors in electrical circuits. For a long time it was not known that selenium has a unique quality. Then a discovery was made on a far away Pacific island. There was an ocean cable station on the island and the resistors used at the station were selenium. One of the cable operators noticed that when the sun came up in the morning he had to readjust his apparatus. He discovered that the adjustment was necessary because a change in resistance occured when the sun shone on the selenium resistors. Thus it was found that selenium has the unique quality of changing its resistance to electric current when exposed to light.

Selenium belongs to the same family of chemical elements as sulfur. It was discovered in 1817 by the great Swedish chemist, Jons Jakob Berzelius. Several chemical elements exist in more than one form. Selenium is one of these. It can be a brown shapeless mass or take the form of gray crystals. It is the gray crystalline selenium which has the property of being a conductor of electricity when exposed to light and almost an insulator in darkness.

First Photocell. A selenium photocell consists of two electrodes, separated by a thin layer of selenium. Alexander Graham Bell, the ingenious inventor of the telephone,

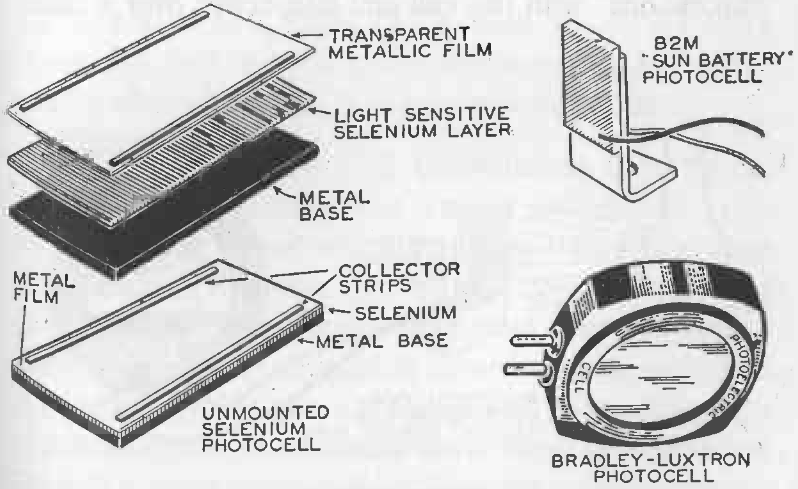
TRANSPARENT METALLIC FILM
LIGHT SENSITIVE SELENIUM LAYER
METAL BASE
B2M "SUN BATTERY" PHOTOCELL
METAL FILM
COLLECTOR STRIPS
SELENIUM
METAL BASE
UNMOUNTED SELENIUM PHOTOCELL
PHOTOELECTRIC CELL
BRADLEY-LUXTRON PHOTOCELL

SELENIUM PHOTOELECTRIC CELL CONSTRUCTION

A selenium photoelectric cell is built in layers as shown at the left. The B2M "Sun Battery" is a low-priced photoelectric cell suitable for amateur experimenters.

was the first to make a USEFUL photocell. After much study of the remarkable behavior of selenium and much laborious research, Bell made a photocell in 1877. He used it to TELEPHONE OVER A BEAM OF LIGHT. The first successful selenium cells made by Bell for this purpose consisted of two copper or brass plates separated by a thin insulating layer of mica. A large number of holes were bored through both plates and mica. Conical brass studs were fixed in the holes in one plate and spaced so that they entered the holes in the other plate but did not touch the plate itself. The annular spaces surrounding the studs were filled with melted brown selenium and heated over a gas flame until the selenium again melted and turned into the slate-colored crystalline form.

Bell and an assistant named Taintner constructed a "photophone" with this cell and telephoned over a beam

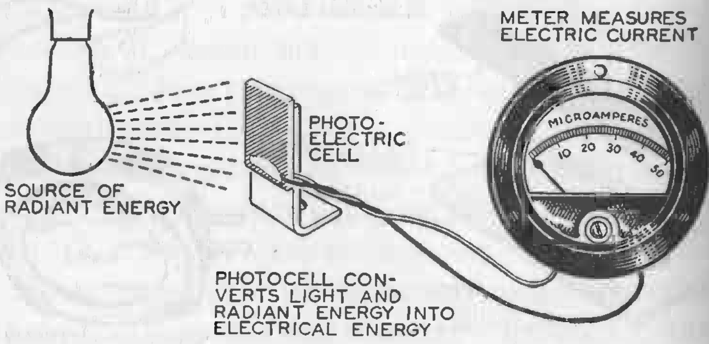
METER MEASURES ELECTRIC CURRENT
MICROAMPERES
10 20 30 40 50
PHOTO-ELECTRIC CELL
SOURCE OF RADIANT ENERGY
PHOTOCELL CONVERTS LIGHT AND RADIANT ENERGY INTO ELECTRICAL ENERGY

LIGHT METERS

Light meters, used in photography for measuring the intensity of light, consist of a selenium photoelectric cell connected to a microammeter (measures millionths of an ampere). The meter is calibrated to indicate the correct exposure under existing light conditions.

of light. In their first experiments, the rays of the sun furnished the source of light. The light was reflected from a mirror attached to the small end of a megaphone. Words spoken into the megaphone caused the mirror to vibrate and vary the light beam reflected from the mirror. When one of the Bell's selenium cells was connected to a battery and a telephone receiver and placed where it would intercept the light beam, the variations in light were reconverted into sound waves which repeated the words spoken into the megaphone.

The Direct Conversion of Sunlight to Electricity. For many decades the selenium cell was apparently only a current-conducting device whose resistance changed when exposed to light. It was not known that a selenium cell could also be a photoelectric cell and in addition to changing its resistance it could also generate a voltage when exposed to light. It is the photoelectric voltage-generating quality of selenium that is now most widely used.

Selenium Photoelectric Cell Construction. The typical selenium photocell consists of a metal plate or disk (usually iron), one side of which is covered with a thin layer of selenium. A very thin film of gold or silver is spattered on the selenium. A current-conducting electrode or collector strips rest on the gold or silver film. The metal film is so extremely thin that it is transparent and permits light to pass and reach the selenium layer. When light shines on the cell a small current flows from the selenium to the metal base. Under proper conditions the flow of current varies almost directly as the amount of light which reaches the cell. Needless to say, the action of a selenium cell is electronic. Several theories explaining how light thus produces a flow of electrons have been proposed.

All photoelectric transducers are not constructed like a selenium cell. One different and widely used type is called a PHOTOTUBE.

Descripition of a Phototube. At first glance a phototube may have the general appearance of a radio tube intended for use in a radio receiver. It consists principally of two electrodes sealed in a glass envelope and has a black plastic base like the base of a radio tube.* The electrode in the

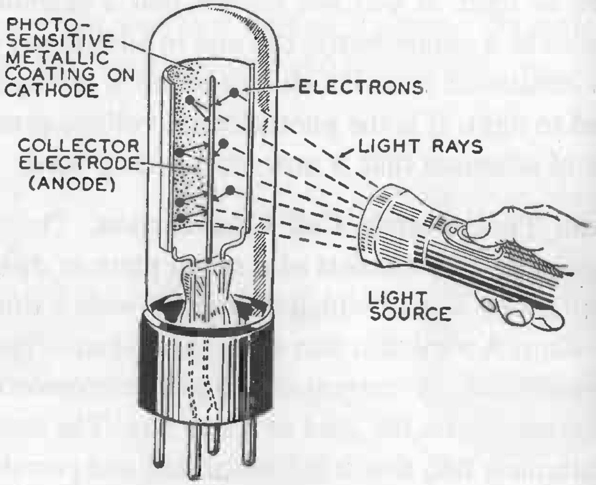
PHOTO-SENSITIVE METALLIC COATING ON CATHODE
ELECTRONS
COLLECTOR ELECTRODE (ANODE)
LIGHT RAYS
LIGHT SOURCE

A PHOTOTUBE HAS A SMALL ANODE AND A LARGE LIGHT-SENSITIVE CATHODE

When light strikes the photosensitive metallic coating on the cathode of the tube, electrons are thrown off from the metal. The electrons (represented by black dots in the sketch) move toward the anode and current flows in the tube circuit. This current can be amplified and made to operate a relay.

---

* This description does not fit some phototubes, namely, those of the cartridge and end types.

shape of a half cylinder, easily visible through the glass envelope, is the cathode. The anode is a short straight rod or wire. When the phototube is in operation the anode is connected to the positive terminal of a voltage source. The cathode emits electrons when its sensitized surface is exposed to light or other radiant energy.

The cathode is the only part of the phototube which is sensitive to light. It is the PRODUCER of electrons. The anode is a COLLECTOR of electrons. The cathode is a strip of thin, silver-plated, sheet copper bent into semicylindrical form. When the phototube is one which is sensitized with cesium, the silver on the concave surface of the cathode is covered with a thin layer of cesium oxide and

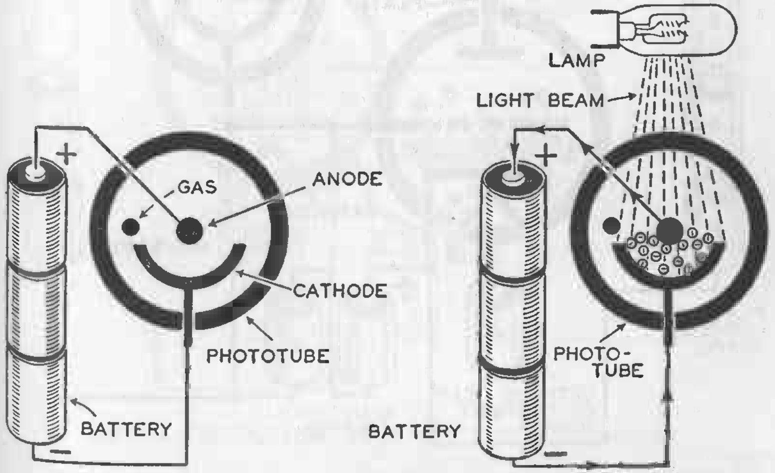

NO ELECTRONS ARE PRODUCED AT THE CATHODE UNLESS LIGHT STRIKES ITS SENSITIVE SURFACE

The large black dot is the conventional symbol for gas in a tube. The small circles enclosing a negative sign in the right-hand sketch represent electrons. When a photoelectric tube is placed in operation the anode is kept at a positive potential so that it will attract electrons.

the oxide is coated with a thin layer of metallic cesium. Cesium is a costly, rare, soft, silvery metal discovered nearly one hundred years ago by Professord Gustav R. Kirchoff and Professor Robert W. Bunsen.* Cesium oxidizes easily in air and decomposes water when immersed in the latter. It must be stored in a vacuum or in a chemically non-active gas.

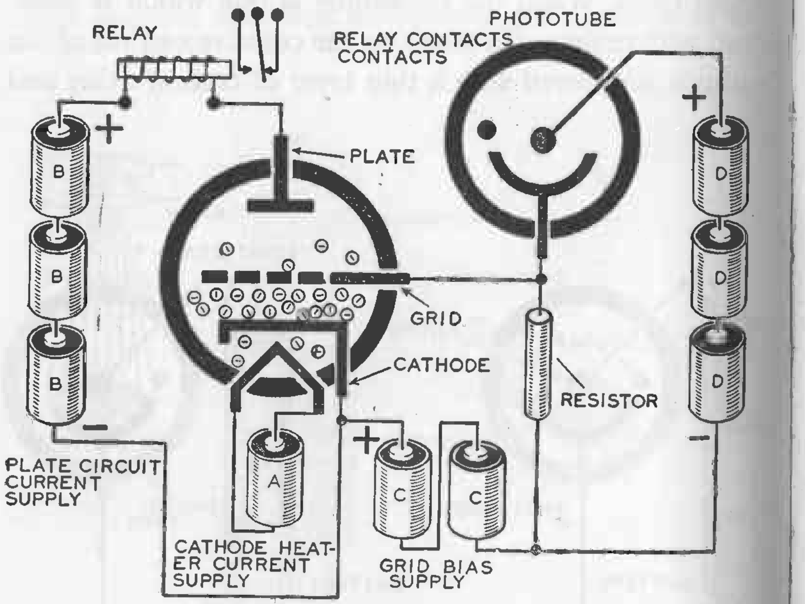
RELAY
RELAY CONTACTS
CONTACTS
PHOTOTUBE
PLATE
GRID
CATHODE
RESISTOR
PLATE CIRCUIT CURRENT SUPPLY
CATHODE HEATER CURRENT SUPPLY
GRID BIAS SUPPLY
A
B
B
B
C
C
D
D
D
+
+
+
−
−

HOW A PHOTOTUBE RELAY OPERATES

When a phototube is put into operation the cathode is connected to a battery (DDD) or other source of positive potential. The circuit also includes an amplifier tube and a sensitive relay. The electric current produced in a phototube is too weak to operate any but the most sensitive relays. Therefore it is fed into an amplifier tube. The amplified current will operate a less sensitive but more rugged type of relay.

---

* Both of these scientists made many important discoveries. Look for their names in the encyclopedia and read about them in "Discovery of the Elements" by Mary Elvira Weeks or in other reference books at the library.

Little practical use has been found for cesium except in phototubes. However, cesium is not the only metal used for sensitizing the cathode in phototubes. Calcium, barium, strontium, etc., and their oxides, are also used for this purpose. Metals such as tantalum, thorium and

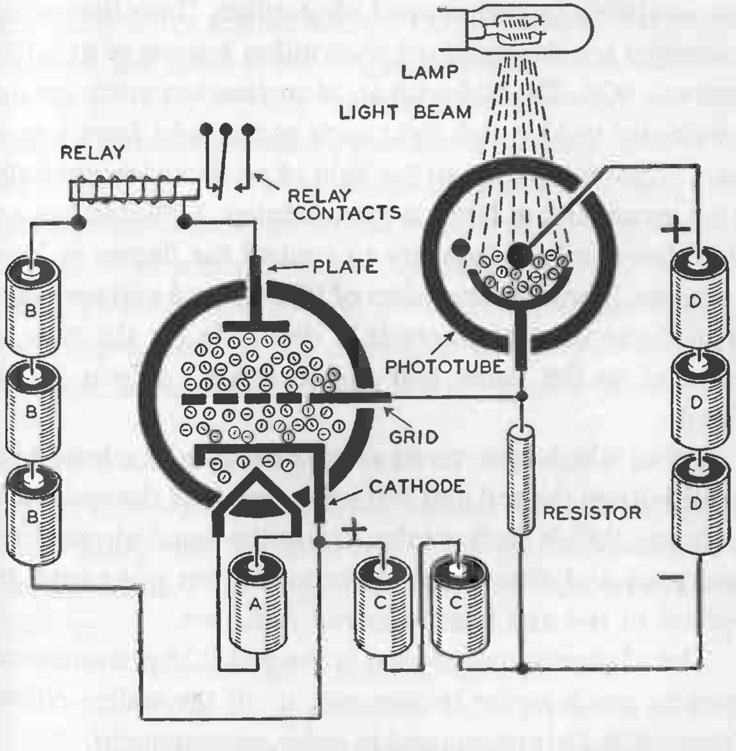
LAMP
LIGHT BEAM
RELAY
RELAY CONTACTS
PLATE
PHOTOTUBE
GRID
CATHODE
RESISTOR
A
B
B
B
C
C
D
D
D
+
+
−

HOW A PHOTOTUBE RELAY OPERATES, continued

Compare this illustration carefully with the preceding one where no light strikes the phototube and consequently few electrons from the amplifier tube's cathode are able to get through the grid and reach the plate. In the illustration above, light is shown striking the cathode of the phototube. Electrons are released. The current flow in the phototube makes the grid of the amplifier tube more positive and electrons from the amplifier cathode reach the plate. Current flows in the plate circuit (**B, B, B** are the plate supply) and the relay armature is drawn toward the coil.

titanium are used on the cathodes of special purpose phototubes which respond to ultra violet light but are not sensitive to light visible to the human eye.

Different metals used as the sensitive surface of the cathode give phototubes different characteristics. Four types of surfaces known respectively as s1, s2, s3 and s4 are available in commercial phototubes. Two interesting examples are the standard phototubes known as RCA 929 and RCA 926. The 929 with an s4 surface has much greater sensitivity to blue rich light such as the light from a mercury vapor lamp than to the light of an incandescent bulb. (An incandescent lamp is blue deficient.) Phototubes are used frequently in industry to control the flames in large furnaces. The characteristics of 929 with s4 surface adapt it to flame control where it is desirable for the tube to respond to the flame and not to heated objects in the flame.

Tubes which have an s1 or s2 photo-surface have high sensitivity in the red and infra-red region of the spectrum. The RCA 930 is such a tube. It has the same physical appearance and dimensions as RCA 929 but gives high response to red and near infra-red radiation.

The s3 photo-surface used in the RCA 926 phototube responds, much as the human eye, to all the visible colors. Hence 926 s3 is often used in color measurement.

What happens? When a phototube is put into operation the anode is connected to the positive terminal of a battery or other source of positive potential. Then if light or other radiant energy strikes the photo-surface of the cathode, electrons are knocked out of the surface, and since electrons are negative they are drawn to the positive

anode. They fly across the space between the cathode and the anode and continue to do so as long as radiant energy strikes the cathode and the anode remains positive. This movement of electrons is an electric current. The number of electrons emitted by the cathode depends upon the wave length of the radiant energy and the amount of light that falls on the cesium surface of the cathode. An intense light of the same wave length produces more electrons and a stronger electric current than a weak light. The electric current produced, even by an intense light is too weak (it is only a few millionths of an ampere) to operate any but the most sensitive relay—a type of relay too delicate to be practical for most applications of the "electric eye." Therefore the current produced in a phototube by light rays is fed into an amplifier. The amplified current will operate less sensitive but more rugged types of relay.

HOW TO BUILD AND USE A PHOTOTUBE RELAY

The components and parts required to build the Phototube Relay are listed below. The letters enclosed in parenthesis are the key to identity the parts in the illustrations.

1	Wood base 7 in. x 4¾ in. x ¾ in.
1	Piece of plywood 2⅝ in. x 1⅛ in. x ¼ in. or a metal bracket to support volume control
1	2500-ohm SPDT sensitive plate circuit relay **(E)**
1	2-watt, 10,000 or 20,000-ohm wire wound linear taper volume control **(B)**
1	SPST switch which can be attached to the volume control **(A)**
1	Knob for volume control **(C)**

1	CE23C gas Cetron phototube
1	117L7/M7-GT radio tube
1	Octal wafer socket for 117L7 tube **(K)**
1	Steatite 4-prong wafer socket for phototube **(L)**
1	16 mfd. 150 DCWV electrolytic capacitor **(J)**
1	10 megohm, ½-watt resistor
1	25,000-ohm, ½-watt resistor
2	single-terminal Bakelite tie-points **(H)**
1	2-terminal Bakelite tie-point **(H)**
1	6-ft. lamp cord and attachment plug **(G)**
1	Screw eye with ¼ in. eye
3	Binding posts or 3 Fahnestock connectors **(D)**
6	½-in. No. 5 round head wood screws
4	1¼-in. No. 5 round head wood screws
4	¾-in. spacers with hole to pass No. 5 wood screw
	shellac, hook-up wire, rosin-core solder

The Phototube. Several types of phototubes were tested for this book and it was found that the CE23C gas-filled Cetron and RCA 868 gave excellent results in a fairly simple circuit. The CE23C proved to be more sensitive than the RCA868. Both are 4-prong tubes which will fit into the same socket and either one may be substituted for the other without changing the circuit.

The Plate Relay. A relay is an electrically operated device for the opening and closing of a circuit, or, in other words, an electrically operated switch. A variety of relays are used in electronic circuits and several types will be found listed in the catalogs of firms dealing in radio, television and industrial electronic supplies. Each type is in-

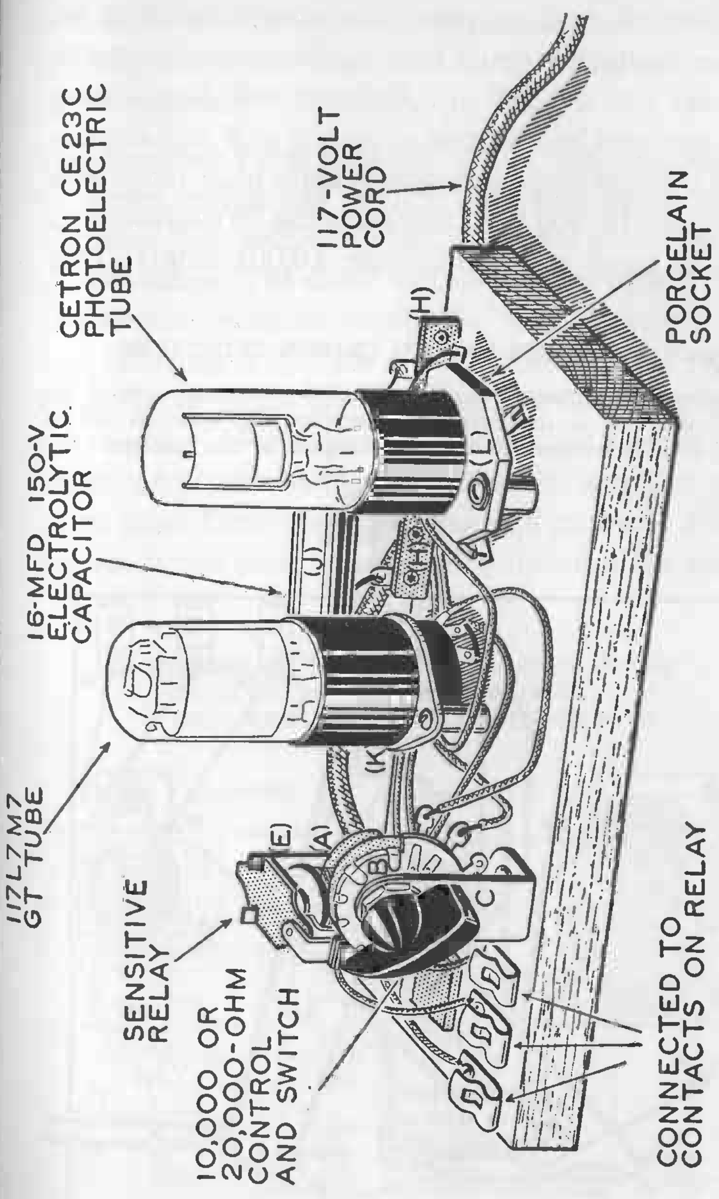
CETRON CE23C PHOTOELECTRIC TUBE
117-VOLT POWER CORD
16-MFD 150-V ELECTROLYTIC CAPACITOR
PORCELAIN SOCKET
117L7 M7 GT TUBE
SENSITIVE RELAY
10,000 OR 20,000-OHM CONTROL AND SWITCH
CONNECTED TO CONTACTS ON RELAY
(H)
(L)
(J)
(H)
(K)
(E)
(A)
(B)
C

PHOTO-RELAY WITH CETRON CE23C TUBE

The letters enclosed in parentheses are the keys for identifying the parts with those in the plan.

tended for a special purpose. The relay selected to use with the phototube should be a medium-cost PLATE CIRCUIT relay with a 2,500 or 5,000-ohm coil. Such a relay can be operated by a current of only 8 or 9 milliamperes (0.008 to 0.009 amps.). Super-sensitive plate relays with a resistance of 5,000 to 10,000 ohms can be operated by a current as small as 1 milliampere (0.001 amp.). They

PLAN OF THE RELAY WITH CETRON CE23C TUBE

The letters in parentheses will aid in identifying the same parts in the circuit diagram and in the preceding illustration. The terminals can be identified with the terminals in the circuit diagram by the numerals.

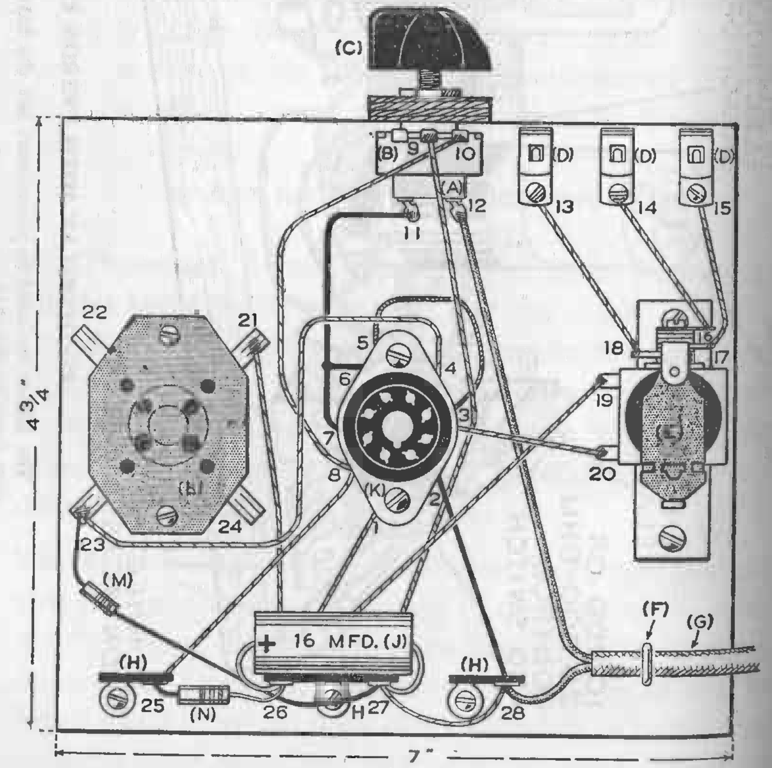
(C)
(B)
9
10
(A)
11
12
(D)
13
(D)
14
(D)
15
22
21
5
4
6
3
7
8
(K)
2
18
16
17
19
(E)
20
(L)
24
23
(M)
(F)
(G)
16 MFD. (J)
(H)
(H)
25
(N)
26
H27
28
4 3/4"
7"

cost more than the type of relay recommended in the list of components and their great sensitivity is not required for this project.

The Amplifier Tube. A 117L7/M7-GT tube is a combined half-wave rectifier and amplifier. A glance at the diagram will show it to be a diode and a pentode within the same glass envelope. The diode section together with a 16 mfd. electrolytic condenser rectify the 117-v alternating current, changing it into the direct current voltage required for the anode of the phototube and for the plate circuit of the amplifier. When electrons are emitted from the cathode of the phototube they are fed to the amplifier section of the tube. Their effect upon the grid produces a flow of current in the plate circuit which operates the relay. A

SCHEMATIC CIRCUIT DIAGRAM OF RELAY WITH CETRON CE23C PHOTOTUBE AND 117L7 AMPLIFIER AND RECTIFIER TUBE

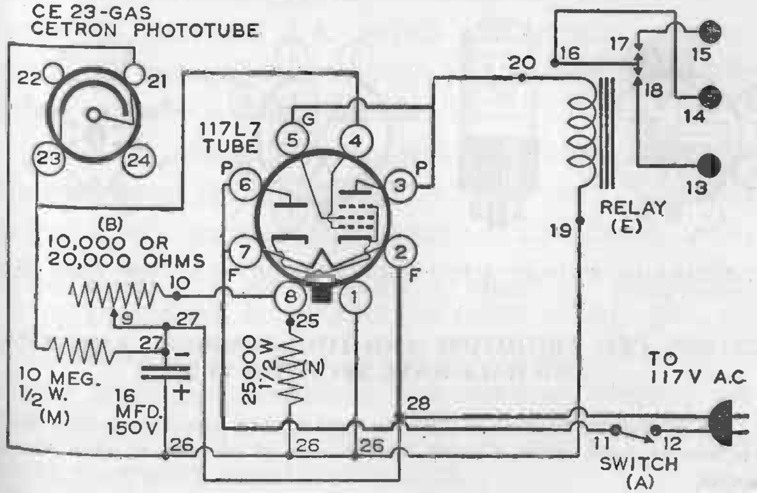
CE 23-GAS
CETRON PHOTOTUBE
22
21
23
24
117L7
TUBE
G
P
5
4
6
3
P
7
2
F
F
8
1
(B)
10,000 OR
20,000 OHMS
10
9
27
27
25
10 MEG.
1/2 W.
(M)
16
MFD.
150 V
25000
1/2 W
(N)
26
26
26
28
20
16
17
18
15
14
13
RELAY
(E)
19
TO
117V A.C.
11
12
SWITCH
(A)

117L7 tube operates directly from the 117-v AC power supply without any form of resistor in series.

Socket. It is not necessary to use a steatite socket for the phototube. Steatite is a low-loss porcelain-like insulating material used in the manufacture of sockets for transmitting tubes. A steatite socket is recommended because it is preferable to a Bakelite or other plastic socket but a Bakelite or plastic socket will be satisfactory for a "homemade electric eye."

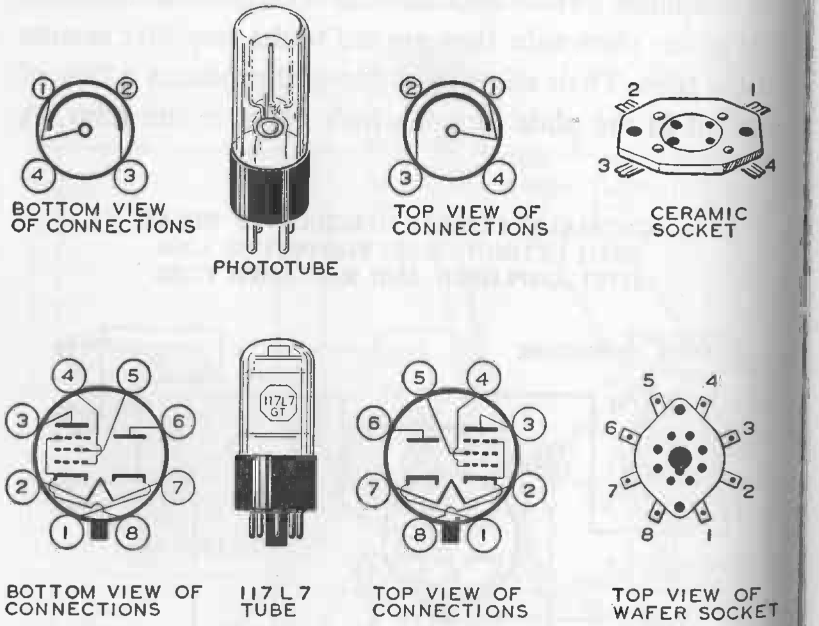

CETRON CE23 PHOTOTUBE AND 117L7 COMBINED AMPLIFIER AND HALF-WAVE RECTIFIER TUBE

The phototube requires a four-pin socket. The amplifier-rectifier tube requires an octal socket. Consult the top view of connections when wiring socket.

Control. The 10,000- to 20,000-ohm, 2-watt, linear taper, wire-wound control specified in the list of components is a type of potentiometer carried in stock by television repairmen for replacement of defective television controls and should not be difficult to find.

Assembling. Smooth the wood base by sandpapering and shellacking all its surfaces. When the shellac has dried mount the sockets, relay, Bakelite tie-points and connectors in the locations shown on the plan. Each socket is raised above the base by two ¾-inch spacers so that the socket terminals do not touch the wood.

The control and switch is mounted on a metal bracket or on a 2⅝ in. x 1⅛ in. plywood strip attached to the front edge of the base at its center.

Begin wiring by tying the 117-v power cord (**G**) to the screw eye **F**. Connect one power wire to a Bakelite tie-point (**H**) 28 near the capacitor. Connect the other power wire to one terminal (**12**) of the switch (**A**). Next connect socket (**K**) terminal 7 to switch (**A**) terminal 11 and socket (**K**) terminal 2 to tie-point (**H**) 28. The heater circuit of the tube is now complete.

Solder all connections with rosin-core solder. Observe the polarity of the electrolytic condenser and connect it as indicated. Consult both the plan and the schematic circuit diagram each time that you put a wire in place. The wiring should not be difficult. In order to help you, all components are marked in the illustrations with an identifying letter in parenthesis. The numerals identify the terminals. After the assembly and wiring is completed, check your work with the diagrams.

Adjustment and Operation. Set the control knob at about the halfway point between its limits of movement. Place the 117L7 tube in its socket but leave the phototube out of its socket temporarily. Allow about one-half minute for warm-up, then place the phototube in its socket. Darken the room or cover the phototube with a small cardboard tube or box. Make a hole about one inch in diameter in the box or tube and position it so that the hole is opposite the center of the concave side of the phototube cathode. Move the control knob back and forth very slowly until you find a point where the relay armature is drawn toward the coil but moves away from the coil if the knob is moved slightly in the opposite direction. Let the control rest at the point where the armature moves away from the coil.

When the phototube relay is used for some practical purpose such as for an annunciator, it should be enclosed in a small cabinet. It can be adjusted to greater sensitivity when all non-essential light is excluded. One of the illustrations shows the details of a 7 inch x 4¾ inch x 4¼ inch cabinet which can be fastened to the relay base with two small hinges. The light beam enters the cabinet through a 1 inch diameter round hole located exactly opposite the center of the concave side of the cathode. Three ⁵⁄₁₆ in. holes are bored in the top of the cabinet directly over the 117 L7/M7-GT tube to let some of the warm air from the tube escape.

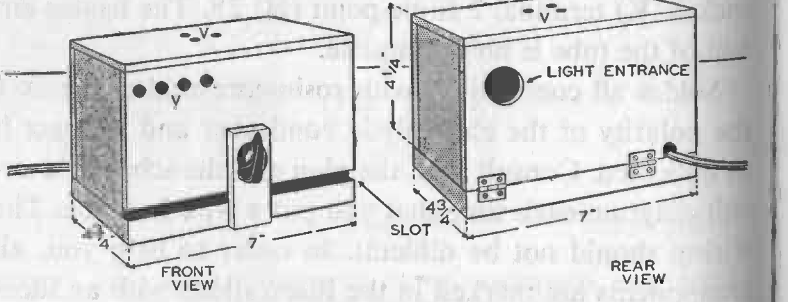

ENCLOSURE FOR PHOTOTUBE RELAY

The ⁵⁄₁₆ in. diameter holes in the top and front of the enclosure (marked V) are for ventilation. The long slot along the lower edge of the front side of the enclosure is also for ventilation.

The Light Source. A match or a flashlight will test and demonstrate the action of the photo-relay. But if the relay is put into practical use, a source of steady light is required. Commercial light sources for photo-relays are fairly expensive. You can build your own at relatively low cost.

Two plans for light sources are illustrated and described. The one with cabinet is a more permanent arrangement than the other.

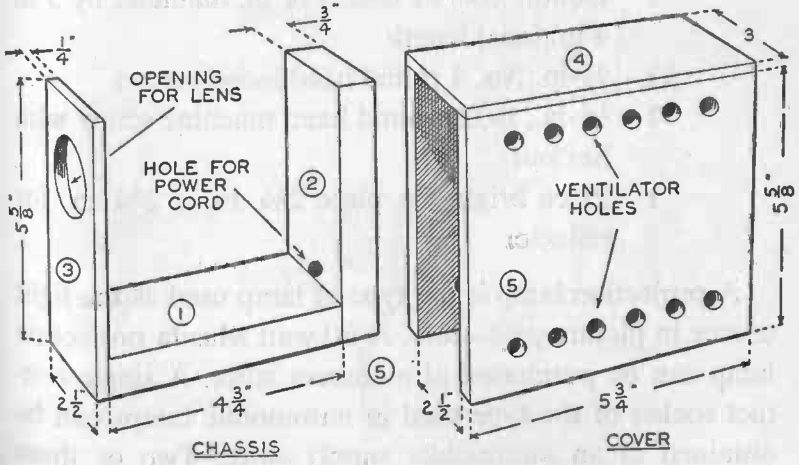

ENCLOSURE FOR LIGHT SOURCE

Materials and Parts Required for Building Light Source

1	Piece of pine 4¾ in. x 2½ in. x ¾ in. (1)
1	Piece of pine 5⅝ in. x 2½ in. x ¾ in. (2)
1	Piece of plywood 5⅝ in. x 2½ in. x ¼ in. (3)
1	Piece of plywood 5¾ in. x 3 in. x ¼ in. (4)
2	Pieces of plywood 5¾ in. x 5⅝ in. x ¼ in. (5)
1	Piece plywood 2⅜ in. x 2⅜ in. x ¼ in. with 1½ in. dia. hole in center (6)
1	115-125-volt, 50-watt, G. E. projection lamp with single contact bayonet candelabra base
1	100-ohm, 20-watt fixed wire wound power resistor with mounting brackets
1	Single-contact, candelabra-base lamp socket
1	Strip sheet metal 4¾ in. x ½ in. x approximately 1⁄32 in.
1	6 ft. length of lamp cord with attachment plug
1	Double convex lens, 1¾ in. diameter by 3 to 4 in. focal length
18	¾-in. No. 4 round head wood screws
1	½-in., 6-32, round head machine screw with hex nut
1	Piece bright tin plate 2¼ in. x 2¾ in. for reflector

A projection lamp is the type of lamp used as the light source in picture projectors. A 50-watt Mazda projection lamp can be purchased at a camera store. A single contact socket of the type used in automobile lamps can be obtained at an automobile supply store. Two or three dealers in surplus and chipped-edge lens advertise in the

mechanical and handicraft magazines. They will supply a chipped-edge lens, approximately 1¾ inches in diameter and with 3 to 4-inch focal length at small cost. A 4-inch

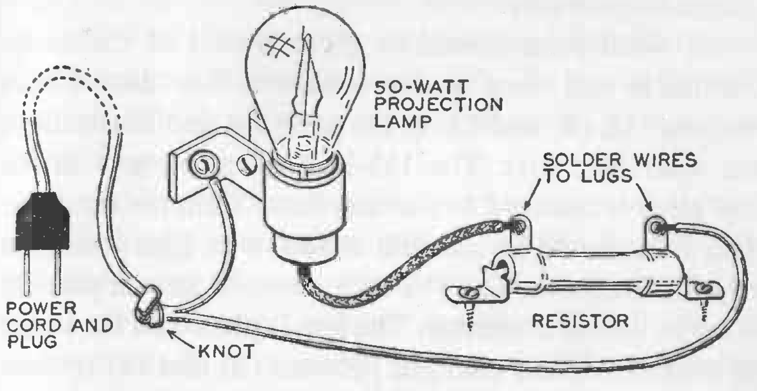
50-WATT PROJECTION LAMP
SOLDER WIRES TO LUGS
POWER CORD AND PLUG
RESISTOR
KNOT

CONNECT LAMP IN SERIES WITH RESISTOR

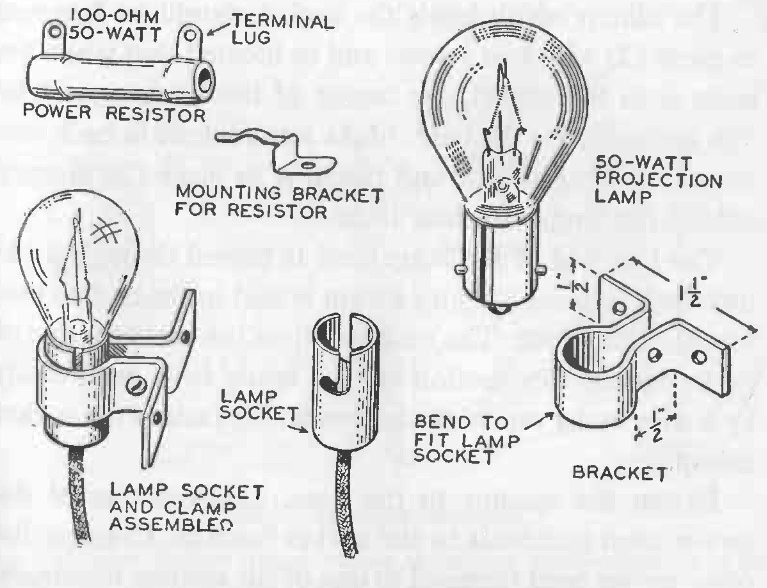
100-OHM 50-WATT
TERMINAL LUG
POWER RESISTOR
MOUNTING BRACKET FOR RESISTOR
50-WATT PROJECTION LAMP
½"
1½"
1"
½"
LAMP SOCKET
BEND TO FIT LAMP SOCKET
BRACKET
LAMP SOCKET AND CLAMP ASSEMBLED

PROJECTION LAMP. SOCKET, RESISTOR AND HOMEMADE BRACKET FOR LIGHT SOURCE focal length is preferable. The bracket for holding the light socket to one end of the cabinet is home-made from the metal strip specified in the parts list. Aluminum is good material for this part. It can be bent into the required shape easily.

All wood parts should be given a coat of shellac or varnish to seal them against moisture. The three pieces marked (**1**), (**2**) and (**3**) in the parts list and illustrations are assembled first. The 1½-inch diameter hole in the end piece is centered 1½ inches down from the top edge. This hole should be cut with a fretsaw or jigsaw and the edge filed smooth. The 1½-inch diameter hole in part (**6**) is cut in the same manner. The lens is placed on the inside of the cabinet and clamped between (**3**) and (**6**) by four screws.

The clamp which holds the socket should be fastened to piece (**2**) with two screws and so located that when the lamp is in the socket, the center of the filament will be 3⅜ inches above the base. Make a small hole in each corner of the tin reflector and fasten it to piece (**2**) directly behind the lamp with four tacks.

The free end of the lamp cord is passed through a ¼-inch hole in piece (**2**) and a knot is tied in the end so that it cannot slip back. The metal shell on the socket is one of its terminals. Connection can be made to it most easily by a wire under one of the screws which fastens the socket clamp.

Fasten the resistor to the base. Connect one of the power cord terminals to the socket bracket. Connect the other power cord terminal to one of the resistor terminals.

Connect the wire which comes out of the bottom of the socket to the other terminal of the resistor. This will connect the lamp and resistor in series.

The rows of ¼ inch holes near the top and bottom edges of the two side pieces (**5**) are necessary for ventilation. The 50-watt lamp and resistor generate considerable heat which must be dispersed. The top of the cabinet should be fastened to the two sides with glue and six or eight small brads.

One of the light sources illustrated in this book consists of a flashlight, transformer and power cord fastened to a wooden frame. The arrangement can be enclosed in a small cabinet if appearance is important. But if you wish to get a steady light source ready as quickly as possible, build it without a cabinet and save time.

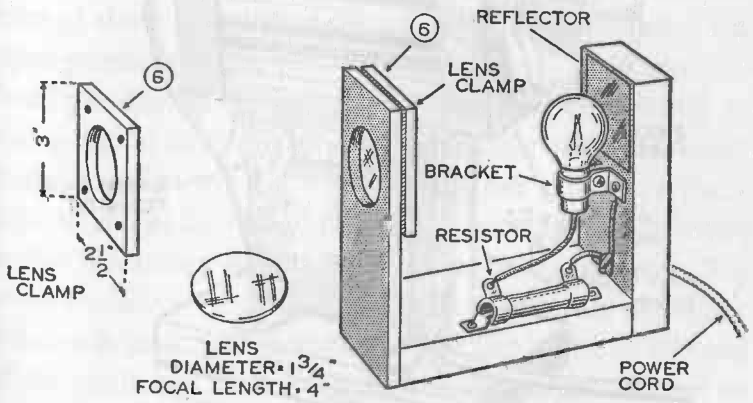

LIGHT SOURCE ASSEMBLY

This shows the light sources without its cover so that the assembly of the lens, socket and resistor can be seen.

The following parts and materials are required:

1	6 in. x 3 in. x ¾ in. wood base **(1)**
1	4¾ in. x 1½ in. x ¾ in. wood block **(2)**
2	Pieces ½ in. dowel rod 4½ in. long **(3)**
1	Piece of pine 1 in. x 1 in. x 4⅜ in. **(4)**
1	2-cell flashlight **(5)**
1	Small filament transformer with 6.3-volt secondary **(6)**
1	**6.2 volt miniature lamp with base to fit socket in flashlight (7)**
1	6 ft. power cord and plug **(8)**
1	Piece of thin brass or copper 2 in. x ¾ **in. (9)**
1	Small screweye **(10)**

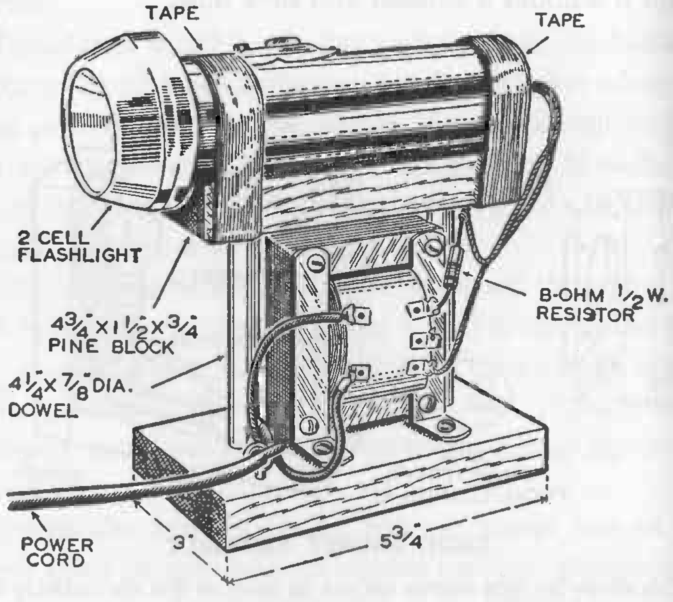

LIGHT SOURCE MADE FROM A FLASHLIGHT
FOR OPERATION ON 117-V A.C.

1	2 ft. length of connecting wire
1	3 ft. length of friction tape
1	¾-in. No. 6 round head brass wood screw

The flashlight should be a 2-cell spotlight or focussed type which throws a narrow concentrated beam. Flashlight batteries do not last long enough to use continuously so they are removed and the 117-v AC takes their place as a power source. The voltage is stepped down by a filament transformer. A filament transformer is used in many radio receivers and amplifiers to reduce 117-v AC to the proper value for the tube heaters. A small filament transformer which will deliver 6.3 volts and about 1 ampere is a standard size and carried in stock by most dealers in radio and electronic parts.

When a round head brass screw is inserted in one end of the 4⅜ inch x 1 inch x 1 inch pine block (4) and the strip of sheet metal (9) fastened to the opposite end, the arrangement will be just the right size to slide into the flashlight case in place of the batteries. A wire should be connected to the brass screw and another wire connected to the metal strip. The wires are led out through a small hole drilled in the end of the flashlight case and connected to the secondary of a 6.3-volt filament transformer. Remove the lamp supplied with the flashlight and substitute a 6.2-volt lamp with a similar base. The screw in the end of the block will make contact with the contact button on the bottom of the lamp base and the metal strip at the opposite end will make contact with the spiral spring in the end of the flashlight case. When the primary of the transformer is connected to the 117-v AC power supply, the switch on the flashlight case will still function and will turn the lamp on and off.

The flashlight is fastened to the framework of the light source by two bands of friction tape and can be removed at any time and restored to normal use by unwinding the tape and installing batteries and suitable lamp. The transformer is mounted on the wood base below the flashlight. The power cord connected to the primary of the transformer is tied to a small screweye in the base so that there will not be a strain on the connections if the cord is pulled.

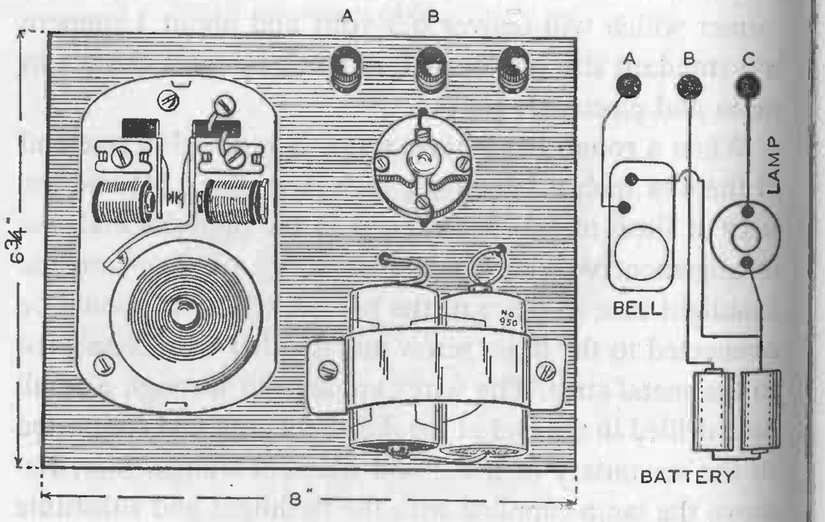

LAMP AND BELL FOR DEMONSTRATING THE PHOTO-RELAY

A bell, flashlight lamp and two size D flashlight cells mounted on an 8 in. x 6¾ in. wood base. The dry cells are connected in series and secured to the base by a clamp made from a sheet metal strip 5¼ in. x 1 in. The 2.5-V flashlight lamp screws into a miniature flat base porcelain receptacle. If binding post terminals B and C are connected to the relay, the lamp will light when the relay operates. If binding post terminals A and B are connected to the relay, the bell will ring when the relay operates.

The Phototube Relay which has just been described can be used as an annunciator. An annunciator is an electrically controlled signal or indicator which "makes an announcement." The small lights which flash on a telephone switchboard are annunciators. When one of the lamps is lighted, it is an announcement that someone is calling.

No one needs to touch a phototube annunciator to set it into operation. A light source and a phototube relay are

ELECTRONIC "BLABBER STOPPER"

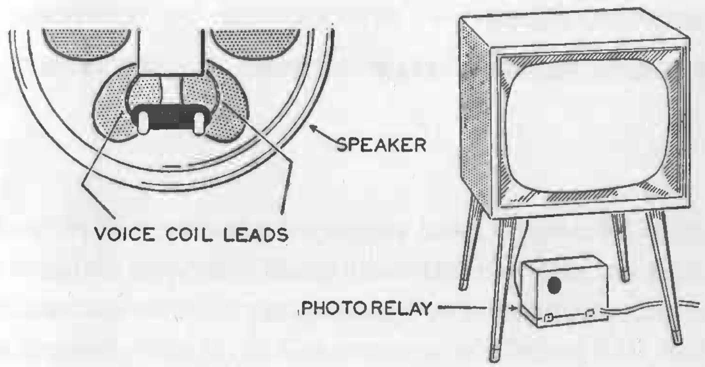

You do not have to leave the comfort of your favorite chair to turn down the volume when you are tired of hearing too much blabber about toothpaste, razor blades, beer, shaving cream, etc. Connect a photoelectric relay so that it will open one of the voice-coil leads on the receiver speaker. Short-circuiting the voice coil may work equally well. To short the voice-coil connect terminal (13) on the photo-relay to one of the voice-coil terminals and terminal (14) to the other voice-coil terminal. To open the voice coil circuit cut one of the voice-coil leads near its terminal and connect the ends of the lead to terminals (14) and (15) on the photo-relay. The voice-coil leads and terminals are located on the back of the speaker as illustrated above.

Adjust the photo-relay so that it will close when you direct the beam of a flashlight at it or turn on a table lamp alongside your chair. When the light is extinguished the program will resume.

placed near a doorway and arranged to announce when anyone passes through the doorway. Anyone who enters or leaves will interrupt the light beam and operate the relay. The relay can be connected to a bell, buzzer, chimes or lamp.

CHAPTER THIRTEEN

Prospecting for Uranium

RADIOACTIVITY • GEIGERSCOPES • GEIGER COUNTERS
SCINTILLOMETERS • HOW TO MAKE A GEIGER COUNTER

Thousands of amateur prospectors have discovered valuable uranium deposits. Many have thereby become rich. Part time and vacation prospecting has sometimes brought large rewards. The U. S. Government is offering $10,000 plus up to $35,000 more in cash bonuses for the discovery of new rich uranium deposits. In the two years just prior to the time when this was written, the U. S. Government paid about four million dollars in bonuses for uranium discoveries. This sum does not include payment for the ore or for land leases. Strikes have been reported throughout the United States from Oregon to Maine and California to Florida.

Prospecting for uranium does not require a great deal of technical "know-how." Uranium reveals itself by its radioactivity. It is not a particularly rare element. It is more abundant than gold and silver. It is found in virtually every kind of rock and natural water but concentrated deposits of uranium are rare.

The search for uranium is not solely to obtain raw material for the manufacture of atomic weapons. There are other uses for this metal. It is a source of energy—a replacement for coal and oil. There are more than 1,000 industrial uses of uranium products. Of perhaps greatest importance are the medical applications of the radioactive uranium products called ISOTOPES. Their healing power may already have saved more lives than were destroyed by the atom bombs at Hiroshima and Nagasaki.

Radioactivity. Some elements continuously undergo a process of atomic disintegration in which energy is liberated. This process is called RADIOACTIVITY and while it is taking place one or more radiations known as particles and rays are spontaneously emitted. The three principal radiations have been named after the first three letters of the Greek alphabet: alpha-particles, beta-particles and gamma-rays. Radium, uranium and thorium and their products are radioactive substances. Their radioactivity cannot be seen, heard or felt directly but there are ways in which it can be revealed. There are several laboratory instruments for detecting radioactivity. Only three are practical to use for prospecting in the field. These are:

1. The Geigerscope
2. The Geiger-Muller Counter
3. The Scintillometer

The Geigerscope (also called a Scintilloscope) is an optical device. The Geiger counter and the scintillometer

are electronic instruments. A description of the Geigerscope and the manner in which it operates makes it easier to understand an explanation of the scintillometer.

What is the Geigerscope? At first glance, a Geigerscope resembles the magnifying glass which a watchmaker uses, called a loupe. The body of the Geigerscope is made of black plastic and is funnel-shaped. At the small end is a magnifying lens and a PHOSPHOR screen.

The word *phosphor* comes from a Greek word meaning *light-carrier*. The name of the element phophorus is from

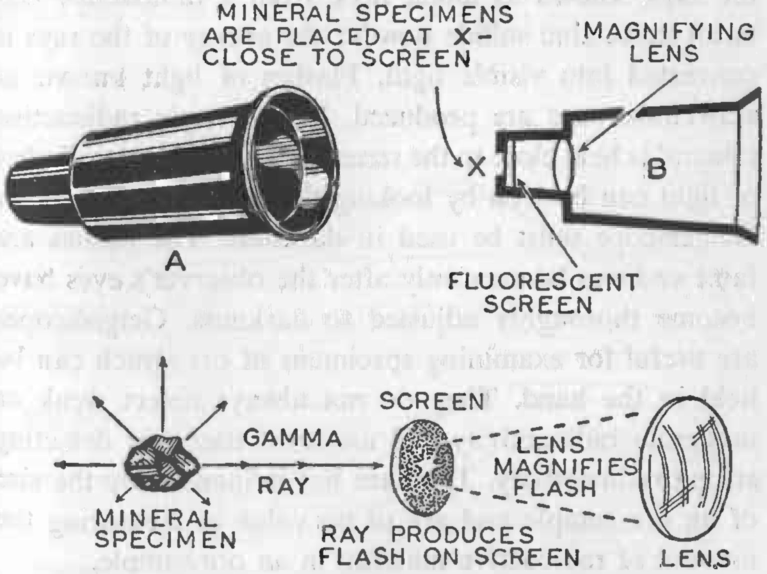

THE GEIGERSCOPE

The screen of a Geigerscope is coated with zinc sulfide. When alpha rays from a radioactive material strike the screen, flashes of light known as scintillations are produced. The scintillations can be seen by looking through the eyepiece. (B is a section of the Geigerscope shown at A.)

the same word. Phophorus emits light when exposed to natural weathering. Many photoluminescent phosphors are employed in television tubes, radar tubes, in paints and inks, in fluorescent lighting and in instruments for detecting various radiations. Photoluminescent phosphors produce light at low temperatures. They absorb energy and then re-emit this energy in the form of visible light. The hands and numerals on a "radium" watch dial which glow in the dark are coated with a photoluminescent phosphor.

The screen of a Geigerscope also is coated with a photoluminescent phosphor, namely zinc sulfide powder. When the rays, known as alpha rays, from a radioactive material strike zinc sulfide powder the energy of the rays is converted into visible light. Flashes of light known as SCINTILLATIONS are produced. If a strongly radioactive mineral is held close to the screen of a Geigerscope, flashes of light can be seen by looking through the eyepiece. A Geigerscope must be used in darkness. The flashes are faint and can be seen only after the observer's eyes have become thoroughly adjusted to darkness. Geigerscopes are useful for examining specimens of ore which can be held in the hand. They do not always detect weak or moderate radioactivity and are most useful in detecting strong radioactivity. They are not influenced by the size of an ore sample and are of no value in measuring the amount of radioactive material in an ore sample.

The Geiger Counter. The Geiger-Muller counter, popularly called Geiger counter, was first devised in 1928 as a laboratory instrument for measurements of radioactivity. When the wide search for uranium ores began during World War II, the laboratory counter was developed into a rugged portable device which could be used by prospectors.

A portable Geiger counter resembles a small battery-operated radio in general appearance. It consists essentially of a Geiger-Muller tube (often spoken of as a G-M tube), a source of high voltage (700-1200 volts), an electronic amplifier and one or more transducers. The transducer provided with inexpensive counters is a telephone receiver or a set of earphones. The more expensive counters are provided with a meter and a neon tube in addition to the earphones. The equipment varies with model and price.

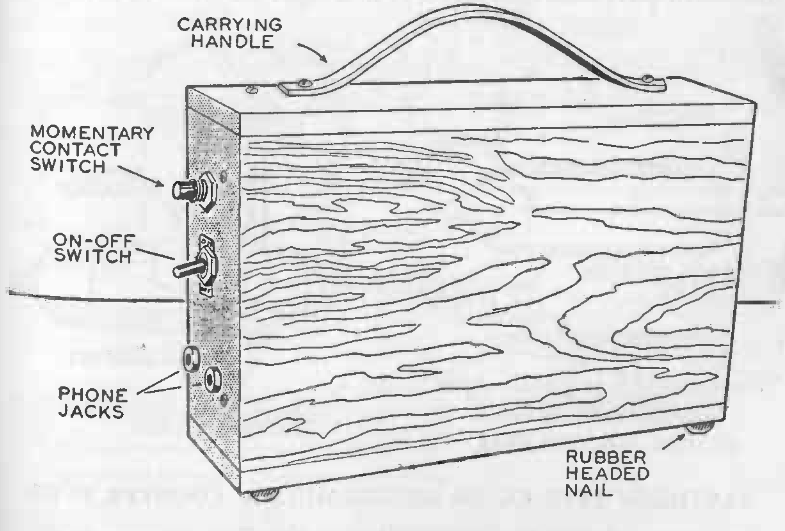

"GAMMA SNIFFER" OR HOMEMADE GEIGER COUNTER FOR DETECTING URANIUM ORES

The Geiger Counter Tube. The "heart" of a Geiger counter is the G-M tube. This may be a few inches long or several feet in length and be made of either glass or metal. In some counters the G-M tube is enclosed within the counter case. In others it is apart from the case but connected thereto by a cable. But all G-M counter tubes, regardless of size or form, operate in the same manner. They all consist of a metal or glass envelope and two electrodes, a cathode and an anode. The envelope is filled with one or more gases. Helium, argon, chlorine and Krypton are some of the gases used in G-M tubes.

When a G-M tube is in operation a high voltage positive charge (500 to 1000 volts depending upon the tube) is applied to the anode. When thus charged, the tube is sensitive to radioactivity. Ores containing such radioactive elements as uranium and radium are radioactive and emit particles and rays. If a Geiger counter is brought close to a radioactive ore a few rays (gamma rays) pass through the wall of the G-M tube and collide with molecules of the gas enclosed in the tube. The collisions knock electrons out of the molecules. The negative electrons are attracted to the positively charged anode and produce a negative electrical pulse each time that an electron collects on the anode. The anode is connected to the grid of an amplifier tube and each time an electron reaches the anode it is indicated by the transducer connected to the amplifier. If the transducer is an earphone it clicks each time an electron reaches the anode. If the transducer is a meter or a neon light it gives a visual signal. The number of clicks in the earphones or flashes in the neon tube which occur in a minute is called the "count" and is a measure of the radioactivity to which the tube is exposed.

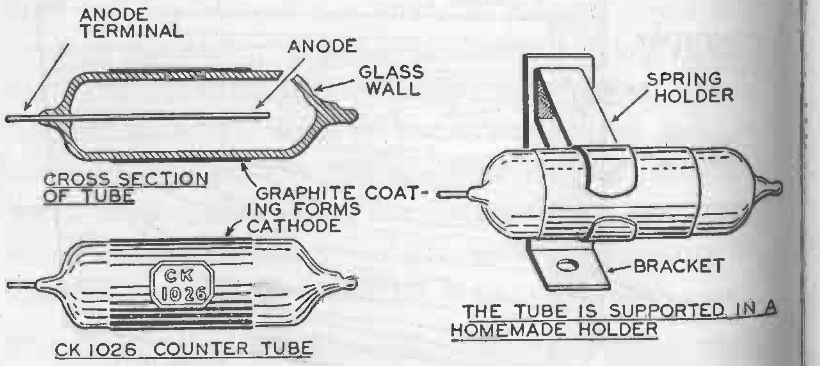
ANODE TERMINAL
ANODE
GLASS WALL
CROSS SECTION OF TUBE
GRAPHITE COATING FORMS CATHODE
CK 1026
CK 1026 COUNTER TUBE
SPRING HOLDER
BRACKET
THE TUBE IS SUPPORTED IN A HOMEMADE HOLDER

RAYTHEON TYPE CK1026 GEIGER-MULLER COUNTER TUBE

This is the sensitive, low-priced G-M tube used in the "Gamma Sniffer." The illustration shows also the homemade holder used to support the tube.

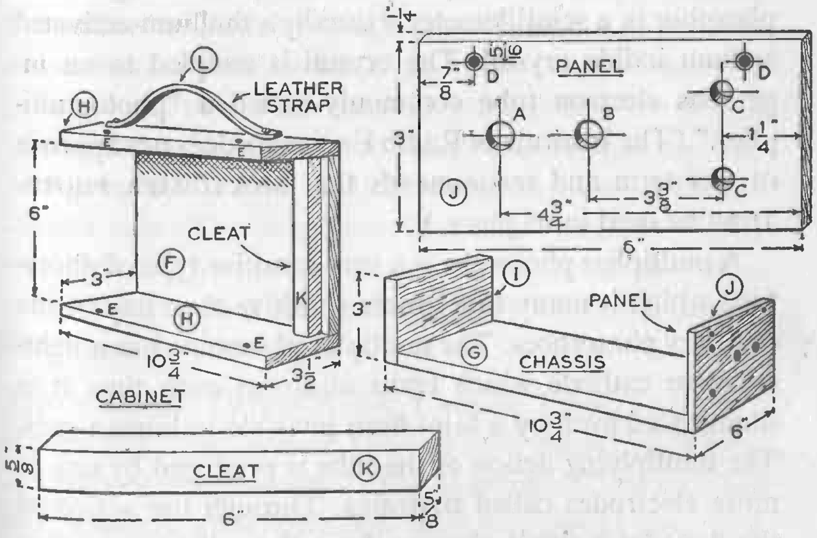
L
LEATHER STRAP
H
E
6"
3"
CLEAT
F
K
3"
H
E
10 3/4
3 1/2
CABINET
1/4"
7/8
9/16
D
PANEL
A
B
3"
C
D
1 1/4
C
J
4 3/4
3 3/8
6"
I
J
PANEL
3"
G
CHASSIS
10 3/4
6"
5/8
CLEAT
K
6"
5/8

DETAILS OF THE WOOD CABINET FOR THE GEIGER COUNTER

When the transducer is a properly calibrated meter the position of the pointer indicates the count.

The Scintillometer. This instrument, (also called a scintillation detector and a scintillation counter) like the Geiger counter was used originally only in laboratories. Portable models for uranium and thorium prospecting are now manufactured. A scintillometer is more than 200 times as sensitive to gamma rays as the most sensitive Geiger counter. It can be used for uranium prospecting from an airplane or a moving vehicle and for the detection of radioactive minerals located from a few feet to hundreds of feet from the instrument.

A scintillometer can be described as an "electrified Geigerscope." As in a Geigerscope the rays from a radioactive material strike a phosphor and produce light. The phosphor in a scintillometer is usually a thallium-activated sodium iodide crystal. The crystal is coupled to an ingenious electron tube commonly called a "photomultiplier." (The Institute of Radio Engineers does not approve of this term and recommends that MULTIPLIER PHOTOTUBE be used in its place.)

A multiplier phototube is a very sensitive type of phototube which is many times more sensitive than any of the ordinary phototubes. The multipler phototube has a light-sensitive cathode which emits electrons each time it is illuminated even by a faint flash invisible to human eyes. The multiplying action of the tube is produced by one or more electrodes called DYNODES. Through the action of the dynodes a single electron from the cathode can start an avalanche of electrons flowing and produce a considerable amount of plate current. Amplification of the order of 100,000 times is possible with this mechanism. A meter properly calibrated and connected to the scintillometer will indicate the percentage of uranium in samples of uranium ore.

HOW TO BUILD THE "GAMMA SNIFFER," OR GEIGER COUNTER

The components of an efficient scintillometer cost more than two hundred dollars at retail and consequently are not within the means of the average amateur experimenter. On the other hand, the components of a simple Geiger counter cost less than fifteen dollars. The Geiger counter, nicknamed "The Gamma Sniffer," and described on the following pages, can be built for slightly less than fifteen dollars by anyone who can make a simple cabinet and solder the circuit wiring. The "Gamma Sniffer" is a simple form of counter but is extremely sensitive and is a practical instrument for uranium prospecting. The model which was built in order to make illustrations for this book employs a type CK1026 Raytheon Geiger-Muller counter tube. This is a low-priced tube (listed by Allied at $3.35) but proved sensitive enough to detect the radioactive fallout which reached New York City from atomic bomb tests in Utah.

Type CK1026 Geiger-Muller Counter Tube. There are several types and sizes of G-M counter tubes. Each type has its own characteristics. The Raytheon Type CK1026 is a low-cost glass envelope tube for use in detecting gamma radiation. It is commonly used in inexpensive prospector's units and demonstration units. The envelope is a sealed glass tube about 2 inches long and ¾ of an inch in diameter. The cathode is a band of colloidal graphite about one inch wide applied to the outside surface of the tube. Connection is made to the graphite band by a spring clamp. The anode is a straight wire which runs through the center of the tube and terminates at one end in a contact pin. A small Fahnestock spring contact clip is the best means of making connection to the pin which forms the anode terminal.

If not broken or mistreated, a CK1026 has a relatively unlimited life. It is not sensitive to light and therefore does not have to be enclosed in a light-tight box or shield.

The Chassis and Cabinet. Commercial Geiger counters are enclosed in a steel or aluminum case. A metal cabinet can be factory-produced in quantities at less cost than a well-finished wood cabinet. Another reason why metal cabinets are used is that a metal enclosure of sufficient thickness will shield the G-M tube from alpha particles and to a great extent from beta particles. The counter is thus made sensitive mainly to gamma rays.

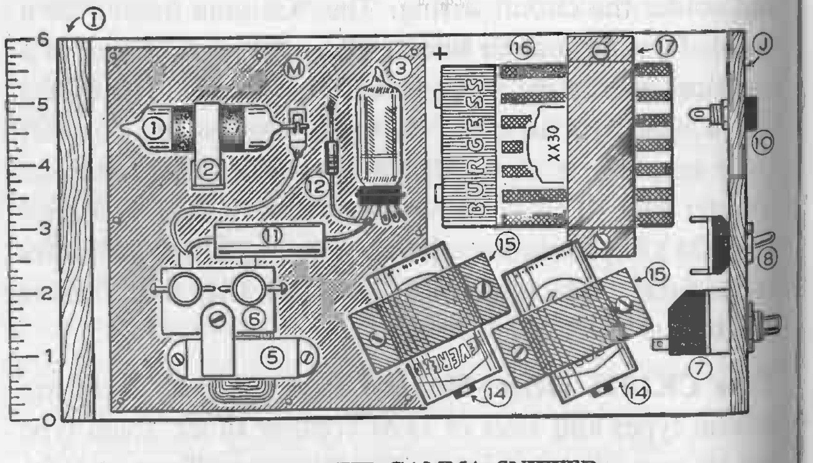

PLAN OF THE GAMMA SNIFFER

This illustration shows the arrangement of the parts on the chassis. The scale at the left can be used to locate their exact position. The letters and numerals correspond to those in the list of parts and can be used for identification.

Some readers of this book may not have the skill or tools required to make a metal cabinet but a wood cabinet is not difficult to make. Wood will permit the G-M tube to respond to all three forms of radioactivity. There is no objection to this.

Some commercial Geiger counters are very compact and the manufacturer features their small size. In designing the Geiger counter for this book, no attempt was made to reduce it to the smallest possible size. It is easier to assemble and wire than it would be if smaller.

The components of the counter are assembled on a wood chassis. When the cabinet is attached to the chassis all components are completely enclosed except those mounted on the panel. Plywood is used to make the

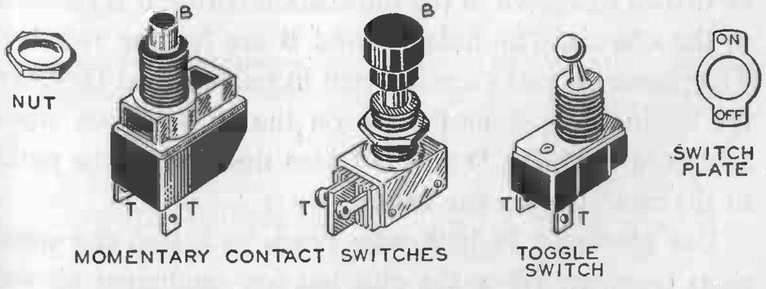

SWITCHES OF THE TYPES USED IN BUILDING THE SNIFFER chassis and cabinet. Six pieces of plywood and one piece of pine cut to the following dimensions are required:

MATERIALS REQUIRED FOR BUILDING THE WOODEN CABINET AND CHASSIS

1	Piece of plywood 10¾ in. x 6 in. x ½ in (F)
1	Piece of plywood 10 in. x 6 in. x ½ in. (G)
2	Pieces of plywood 10¾ in. x 3½ in. x ½ in. (H)
1	Piece of plywood 6 in. x 3 in. x ½ in. (I)
1	Piece of plywood 6 in. x 3 in. x ¼ in. (J)
1	Piece of pine 6 in. x ⅝ in. x ⅝ in. (K)
4	Rubber head tacks
1	Piece leather strap 9 in. x ½ in. (L)
4	1-inch No. 6 flat head wood screws
2	⅜-inch No. 5 round head wood screws
1	Piece sheet metal 5 in. x 5¾ in. (M)
	⅞-in. brads and glue

The pieces marked **I, G** and **J** are assembled to make the chassis. The other four pieces form the cabinet. The 6 inch x 3 inch x ¼ inch piece (**J**), used as a panel, should be drilled as shown in the illustration before it is fastened to the chassis. The holes **A** and **B** are for the switches. The phone tip jacks are inserted in holes **C** and **C**. When the cabinet is put into place on the chassis, two wood screws in the holes **D** and **D** fasten the edge of the panel to the cleat (**K**) on the cabinet.

Use glue and ⅞-inch wire brads to fasten the wood parts together. When the glue has set, sandpaper all surfaces smooth and finish with two coats of shellac. Some

sort of carrying handle should be attached to the top of the cabinet (**H**). A piece of leather strap (**L**) makes a good handle. The holes **E, E, E, E** are for the screws which hold the cabinet to the chassis.

Cut a 5 inch x 5¾ inch rectangle from a piece of thin tinned or galvanized sheet iron. Use the point of an 8 d nail to punch a small hole through the plate at each corner and midway along each edge. Use ⅜-inch wire nails to fasten the plate to the chassis at the end opposite from the panel. (*See plan*). When the chassis has been assembled and the shellac or varnish finish is dry attach a rubber head tack at each corner on the underside.

COMPONENTS OF THE GEIGER-MULLER COUNTER CIRCUIT

All of these components, with the exception of the spark gap and tube holder are standard parts listed in many radio supply catalogs.

1	Type CK1026 Raytheon Geiger-Muller Counter Tube (1)
1	Homemade clamp for counter tube (2)

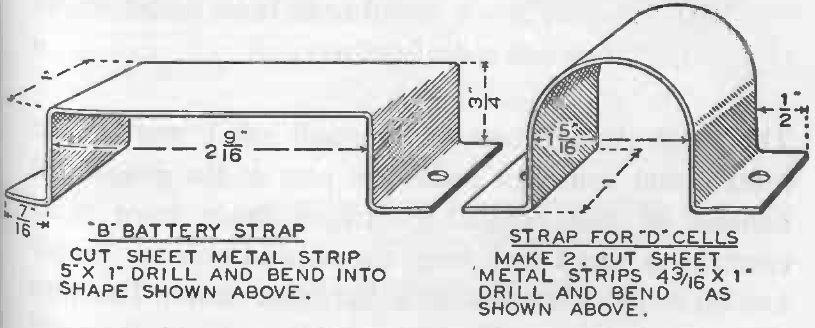
3"/4
2 9"/16
1 5"/16
1"/2
7"/16
"B" BATTERY STRAP
CUT SHEET METAL STRIP 5" X 1" DRILL AND BEND INTO SHAPE SHOWN ABOVE.
STRAP FOR "D" CELLS
MAKE 2. CUT SHEET METAL STRIPS 4 3/16" X 1" DRILL AND BEND AS SHOWN ABOVE.

HOMEMADE CLAMPS FOR SECURING THE BATTERIES TO THE CHASSIS

1	Type 3s4 power pentode miniature radio tube (3)
1	Miniature bottom-mounting Bakelite tube socket for 3s4 tube (4)
1	Output transformer of the type used with 50L6 tubes in small radio receivers (5)
1	Homemade spark gap (6)
1	Momentary contact switch (7)
1	SPST toggle switch (8)
1	ON-OFF switch plate (9)
2	Phone tip jacks (10)
1	.05 mfd. 600 VDC by-pass capacitor (11)
1	10-megohm, ½ watt resistor (12)
1	Small Fahnestock Spring contact clip (13)
2	Size D 1.5 volt flashlight cells (14)
2	Homemade clamps for cells (15)
1	Type xx30 45-volt Burgess B battery with Snapon terminals (16) The equivalent Eveready battery is No. 455 and the equivalent RCA battery is No. vs055
1	Homemade clamp for battery (17)
10	⅜-inch No. 4 round head brass wood screws
1	2000-ohm radio headset

The Spark Gap. There is no small spark gap on the market and you must make this part of the counter. It consists of two pointed electrodes made from brass escutcheon pins 1 inch long. Escutcheon pins are round-headed nails obtainable at a hardware store. The pins are pointed at one end but the point is too blunt. It should be filed until it is as sharp as a needle. Each electrode is

supported in an Eby binding post. Both binding posts are mounted on a hard rubber, Micarta, polystyrene or Bakelite base. Any thin plastic which will insulate high voltage

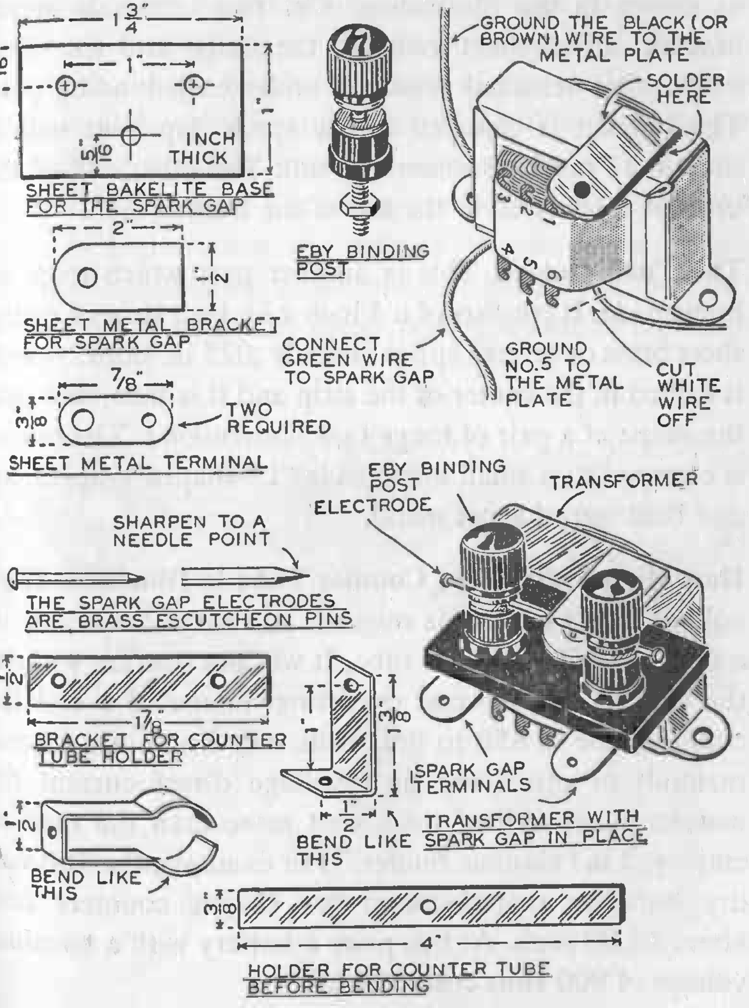
1¾"
1"
3/16"
5/16"
1"
INCH THICK
SHEET BAKELITE BASE FOR THE SPARK GAP
2"
½"
SHEET METAL BRACKET FOR SPARK GAP
⅞"
⅜"
TWO REQUIRED
SHEET METAL TERMINAL
EBY BINDING POST
GROUND THE BLACK (OR BROWN) WIRE TO THE METAL PLATE
SOLDER HERE
CONNECT GREEN WIRE TO SPARK GAP
GROUND NO.5 TO THE METAL PLATE
CUT WHITE WIRE OFF
EBY BINDING POST
ELECTRODE
TRANSFORMER
SHARPEN TO A NEEDLE POINT
THE SPARK GAP ELECTRODES ARE BRASS ESCUTCHEON PINS
½"
1⅞"
BRACKET FOR COUNTER TUBE HOLDER
⅜"
1"
SPARK GAP TERMINALS
½"
BEND LIKE THIS
TRANSFORMER WITH SPARK GAP IN PLACE
½"
BEND LIKE THIS
⅜"
4"
HOLDER FOR COUNTER TUBE BEFORE BENDING

DETAILS OF THE SPARK GAP, BRACKET AND HOLDER FOR G-M TUBE can be used. Bakelite and other plastics can be cut with a fine-tooth hacksaw. A piece 1 inch x 1¾ is required. It does not need to be more than 1/16 inch thick. Drill three holes in the plastic with a No. 28 drill. Locate the holes as shown in the illustration. Cut two terminals and a bracket out of sheet metal of the shape and size illustrated. One terminal is placed under each binding post. The bracket is clamped to the spark gap base with a short 6-32 machine screw and nut. The other end of the bracket is soldered to the top of the transformer.

The Tube Holder. This is another part which must be homemade. It consists of a 4 inch x ⅜ inch strip of spring sheet brass or bronze approximately .025 in. thick. A hole is drilled in the center of the strip and it is then bent into the shape of a pair of tongs (*see illustration*). The holder is clamped to a small homemade "L"-shaped bracket cut and bent out of sheet metal.

How High-Voltage for Counter Tube is Obtained. High voltage direct current is required to charge the anode of a Geiger-Muller counter tube. It will not operate without the charge. The normal operating-voltage of a CK1026 counter tube is 850 to 950 volts D.C. There are several methods of producing high voltage direct current for counter tubes. All of them cost more than the method employed in "Gamma Sniffer." For example, the 300-volt dry batteries manufactured for Geiger counters cost about $8.00 each. At this price a battery with a terminal voltage of 900 volts costs $24.00.

The "Gamma Sniffer" employs a small transformer, a spark gap, two size "D" Flashlight cells and a momentary contact switch. This same method is used in low-priced factory-built Geiger counters. Here is how it operates:

The small transformer is an OUTPUT transformer ordinarily used to couple the plate circuit of an audio amplifier tube to the voice coil of a loud speaker. The primary winding on an output transformer of this type contains a large number of turns of wire and the secondary contains comparatively few turns. When it is used to couple an amplifier tube to a loudspeaker it is a STEP-DOWN transformer. The voltage across the primary terminals is higher than the voltage across the secondary terminals. However, if the primary winding is used as a secondary winding and the secondary is used as a primary

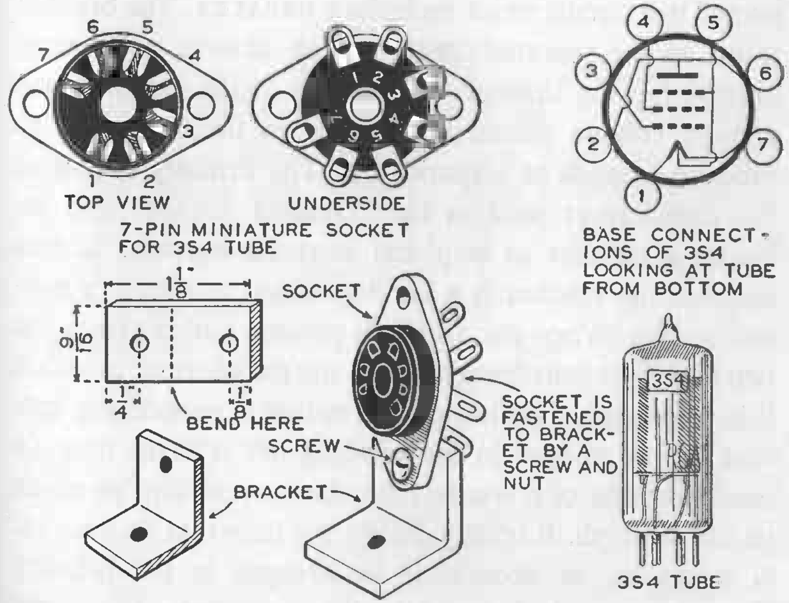
6
5
7
4
3
1
2
TOP VIEW
UNDERSIDE
7-PIN MINIATURE SOCKET FOR 3S4 TUBE
1 2 3 4 5 6 7
BASE CONNECTIONS OF 3S4 LOOKING AT TUBE FROM BOTTOM
1⅛"
9/16"
1/4"
1/8"
BEND HERE
SOCKET
SCREW
BRACKET
SOCKET IS FASTENED TO BRACKET BY A SCREW AND NUT
3S4
3S4 TUBE

THE 3S4 AMPLIFIER TUBE, ITS SOCKET AND BRACKET FOR MOUNTING THE SOCKET ON CHASSIS the transformer becomes a STEP-UP transformer and INCREASES voltage. It will raise the voltage of two flashlight cells (3 volts) to several hundred volts. An output transformer was not designed to produce high voltage for a Geiger counter tube but it serves this purpose quite well and is used because there are no small step-up transformers readily obtainable.

The transformer operates in the same manner as the ignition coil or spark coil in an automobile. When a direct current flows steadily through the primary winding there is no voltage induced in the secondary. It is only while a current BEGINS to flow and CEASES to flow through the primary that a voltage is induced in the secondary. The primary winding of an automobile ignition coil is connected in a circuit which includes a BREAKER. The breaker is an engine-operated switch which STARTS and STOPS current flowing through the primary at the instant when enough voltage should be induced in the secondary to produce a spark at a spark plug. The primary circuit of the transformer used in the "Gamma Sniffer" also includes a breaker to stop and start the current. In this instance the breaker is a hand-operated momentary contact switch. When the switch is pressed current from the two flashlight cells flows through the transformer to which it is connected. The instant the switch is pressed the current begins to flow in the winding but it takes time (a small fraction of a second) for the current flow to reach its full strength. It is only during the time that the current is increasing or decreasing in strength in the primary that a voltage is induced in the secondary. When the current reaches its full strength the voltage in the secondary disappears until the switch is released and the current in the primary winding is cut off. The current is not cut off instantly. It dies away gradually (in a very small fraction of a second) and while it is ceasing to flow a voltage is again induced in the secondary.

Electromagnetism is the agent which causes the current flow to increase gradually when the circuit is closed and to cease gradually when the circuit is opened. It is also electromagnetism which induces a voltage in the secondary. When current flows through the primary winding it builds up a magnetic field in the iron core of the transformer. In other words, current flowing through the primary winding magnetizes the iron core of the trans-

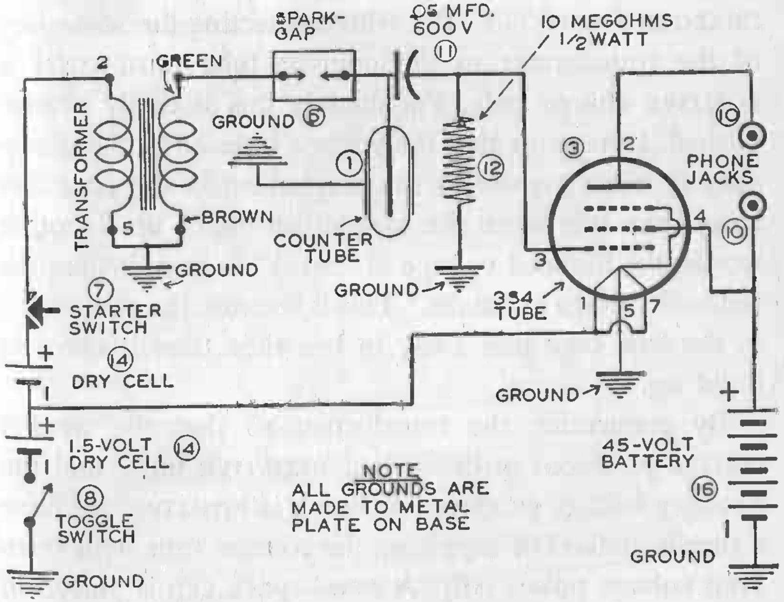
SPARK-GAP
.05 MFD. 600 V
⑪
10 MEGOHMS 1/2 WATT
2
GREEN
TRANSFORMER
GROUND ⑤
①
⑫
③
2
⑩
PHONE JACKS
4
⑩
5
BROWN
COUNTER TUBE
3
GROUND
⑦
STARTER SWITCH
GROUND
3S4 TUBE
1
5
7
+
⑭
DRY CELL
GROUND
+
1.5-VOLT DRY CELL ⑭
NOTE
ALL GROUNDS ARE MADE TO METAL PLATE ON BASE
45-VOLT BATTERY
⑯
⑧
TOGGLE SWITCH
GROUND
GROUND

SCHEMATIC CIRCUIT DIAGRAM OF THE "GAMMA SNIFFER" former. The secondary winding is also wound around the iron core. When current begins to flow through the primary, a short space of time (fraction of a second) is required for the magnetism to reach its full strength. When current ceases to flow through the primary a fraction of a second is required for the magnetism to die away and disappear. It is during the short periods when the magnetism is either building up or dying away, in other words fluctuating, that a voltage is induced in the secondary.

When the magnetism is building up it produces a voltage in the secondary which is of opposite polarity to that produced when the magnetism is dying away. This imposes a small problem when the transformer is used to charge the anode of a counter tube. The anode MUST BE CHARGED POSITIVELY. The wire connecting the secondary of the transformer to the counter tube must carry a POSITIVE charge only. Fortunately this is easily accomplished. It happens that the voltage induced in the secondary is much greater as the magnetism in the core dies away than it is when the magnetism builds up. In other words, the induced voltage at "break" is greater than the induced voltage at "make." This is because the magnetism in the iron core dies away in less time than it takes to build up.

By connecting the transformer so that the weaker voltage produced at break is a NEGATIVE pulse and the stronger voltage produced at break is POSITIVE, we have a simple method of supplying the counter tube with POSITIVE voltage pulses only. A small spark gap is placed in the wire which connects the anode of the counter tube to

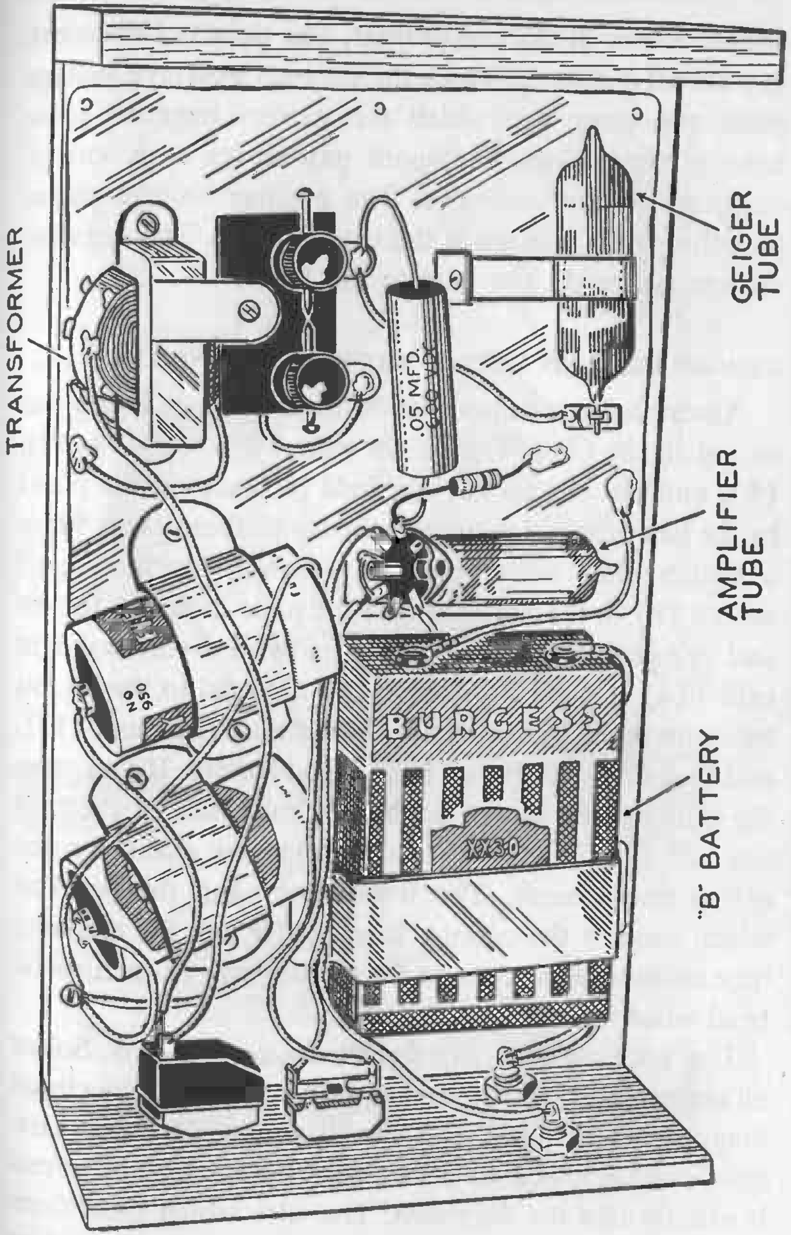

THE ASSEMBLED AND WIRED CHASSIS OF THE GEIGER COUNTER the secondary of the transformer. The points of the spark gap are set a distance which the stronger POSITIVE voltage pulse can jump but which the weaker negative pulse cannot jump. Thus, the spark gap serves as a sort of electrical selector or valve. The positive voltage pulses pass through it and reach the counter tube. The negative voltage pulses do not pass to the tube.

ASSEMBLING AND WIRING THE "GAMMA-SNIFFER"

Assemble the components on the chassis and panel as shown in the plan. The starter switch (7), toggle switch (**8**), and pin jacks (**10**) are held in place on the panel by the hex nuts which are supplied with these parts. Place a switch plate under the nut which holds the toggle switch (**8**) to the panel so that the plate indicates the ON and OFF position of the switch. The two 1.5-volt flashlight cells (**14**) and the B battery (**16**) are held to the chassis by straps made from strips of sheet metal, (**15**) and (**17**), and fastened with round head wood screws. If you trace the 6-inch scale shown at the left hand side of the plan you can use the tracing to determine the exact position of the components. The transformer and the brackets which support the counter tube holder and the amplifier tube socket are fastened to the chassis with ⅜-inch round head wood screws.

Use hook-up wire to connect the components. Solder all connections with rosin-core solder. Refer to the circuit diagrams each time that a wire is put in place and carefully double check to insure that the completed circuit is exactly like the diagrams. The wire which leads from one side of the spark gap to the counter tube should be

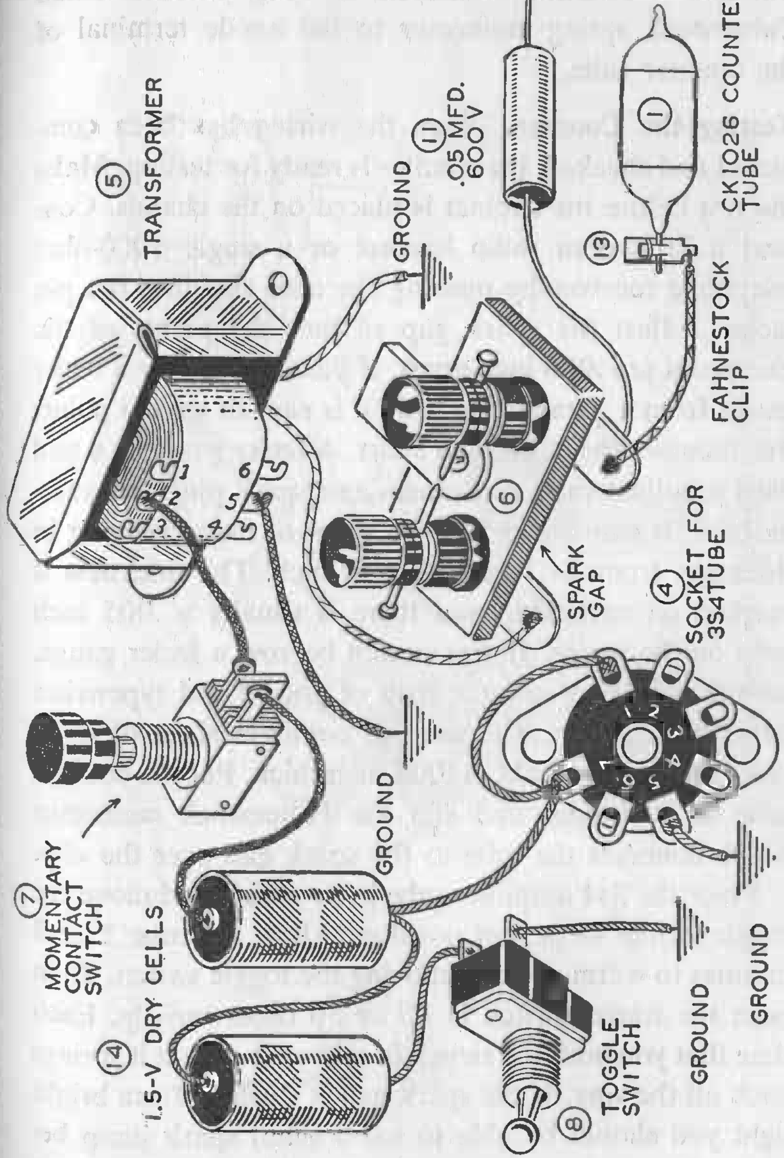
⑤ TRANSFORMER
GROUND
⑪ .05 MFD. 600 V
① CK1026 COUNTER TUBE
⑬ FAHNESTOCK CLIP
1
2
3
4
5
6
⑥
SPARK GAP
④ SOCKET FOR 3S4TUBE
1
2
3
4
5
6
7
GROUND
⑦ MOMENTARY CONTACT SWITCH
GROUND
⑭ 1.5-V DRY CELLS
⑧ TOGGLE SWITCH
GROUND
GROUND

PICTORIAL DIAGRAM OF PART OF THE GEIGER COUNTER CIRCUIT a flexible wire and about 3 inches long. Solder a small Fahnestock spring connector to the anode terminal of the counter tube.

Testing the Counter. When the wiring has been completed and checked, the counter is ready for testing. Make the test before the cabinet is placed on the chassis. Connect a 2000-ohm radio headset or a single 1000-ohm telephone receiver by pushing the cord tips into the pin jacks. Adjust the spark gap so that the points of the electrodes are .005 inch apart. If you can borrow a feeler gauge from a garage mechanic it is easy to set the points five-thousandths of an inch apart. A feeler gauge is a tool used to adjust valve clearances and spark plugs on automobiles. It consists of 8 or 10 strips of metal ranging in thickness from .003 inch to .040 inch. The thickness is marked on each strip and there is usually a .005 inch strip on the gauge. If you cannot borrow a feeler gauge, set the points by using a strip of good bond typewriter paper as a gauge. The average common sheet of bond typewriter paper is about .005 inch thick. Put the counter tube in its holder and slip the Fahnestock connector which connects the tube to the spark gap over the clip.

Place the 3s4 amplifier tube in its socket and move the toggle switch to the ON position. Allow the tube 2 or 3 minutes to warm up after moving the toggle switch. Then push the starter switch in 40 or 50 times rapidly. Each time that you push the switch in release it so that it springs back all the way. If the spark gap is shielded from bright light you should be able to see a small spark jump between the points each time that the button on the switch is released. Bring a sample of radioactive mineral or the radium dial of a watch close to the counter tube and listen

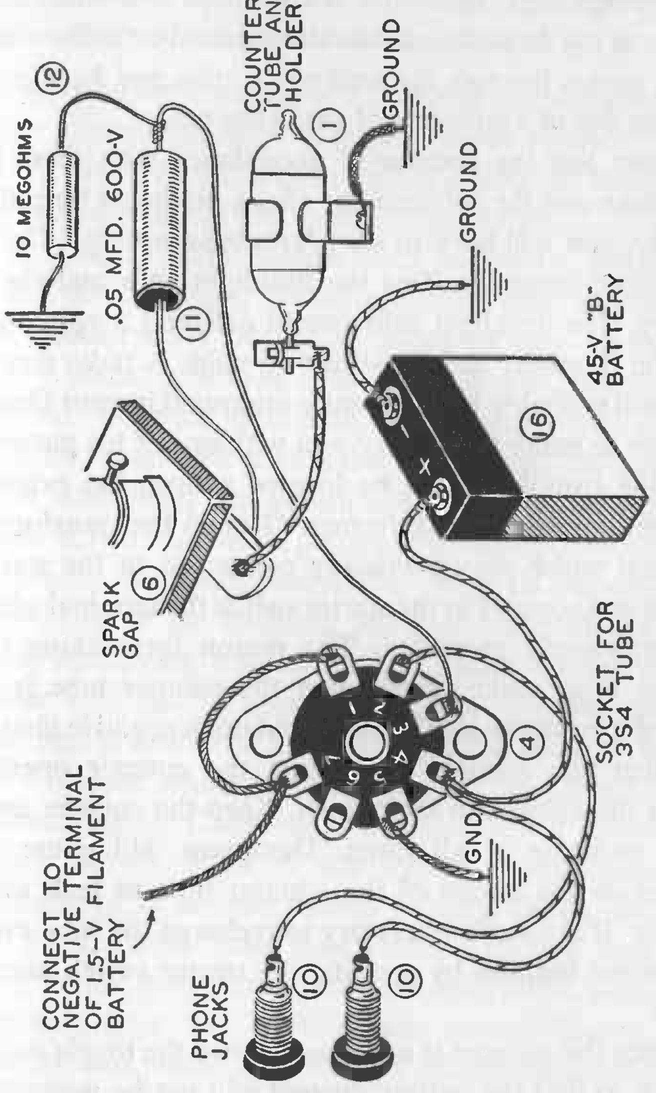

PICTORIAL DIAGRAM OF PART OF THE GEIGER COUNTER CIRCUIT in the headphones. You should hear a loud rapid clicking in the earphones. Each one of the clicks is produced by a gamma ray from the radioactive mineral or radium dial which passes through the wall of the tube and knocks an electron out of a molecule of gas in the tube.

If you test the counter in accordance with these instructions and do not hear the clicks produced by radioactivity, you will have to start "trouble-shooting." Check the wiring carefully. Test the flashlight cells and the B battery. The flashlight cells should deliver 1.5 volts each and the B battery should deliver 45 volts. A radio service man will probably be sufficiently interested in your Geiger counter to make the test for you with one of his meters.

If the trouble cannot be located reverse the primary connections on the transformer. Ground the transformer terminal which was previously connected to the starter switch and connect to the starter switch the terminal which was previously grounded. The reason for making this change is to make certain that the counter tube is receiving a POSITIVE small charge and not a negative charge.

When you succeed in making the counter operate, fasten the cabinet to the chassis. Keep the counter away from moisture at all times. Dampness will cause the charge on the anode of the counter tube to leak away rapidly. It is usually necessary to recharge the tube every five or ten minutes by pressing the starter switch several times.

When the counter is not in use, move the toggle switch to OFF so that the battery current will not be wasted. In order for the counter to operate it is necessary for the 1.5-volt cells to be fresh.

The clicks which occur a few seconds apart when there

is apparently no radioactive material nearby is the "background" count and is due to cosmic rays and general radioactivity in the earth and atmosphere.

Prospecting for Uranium will probably be most successful if the mineral is sought in general areas where

(1) the geologic conditions are similar to those existing where uranium has been found before;

(2) where other metals such as cobalt, zinc, lead, nickel, silver, copper, vanadium and bismuth exist.

Uranium never occurs in nature in pure form. It is always found combined with other substances to form a mineral. There are more than 100 minerals which contain uranium combined with other substances. In some of these uranium is the principal constituent. In others it forms only a very small percentage. Anyone who intends to search for uranium-bearing minerals should have knowledge of the common uranium minerals and their chief identifying characteristics. This information, together with descriptions of the most common thorium minerals, is contained in a small book called *Prospecting for Uranium* which is published by the United States Atomic Energy Commission and the United States Geological Survey. It is for sale by the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D.C. and costs only 45 cents. If you send for a copy do not send postage stamps. Send coins, a check or a money order. The United States Atomic Energy Commission also publishes a booklet called *Prospecting with a Counter*. This also is for sale by the Superintendent of Documents. The price is 30 cents. Both books are worth having if you are interested in radioactivity, uranium or counters.

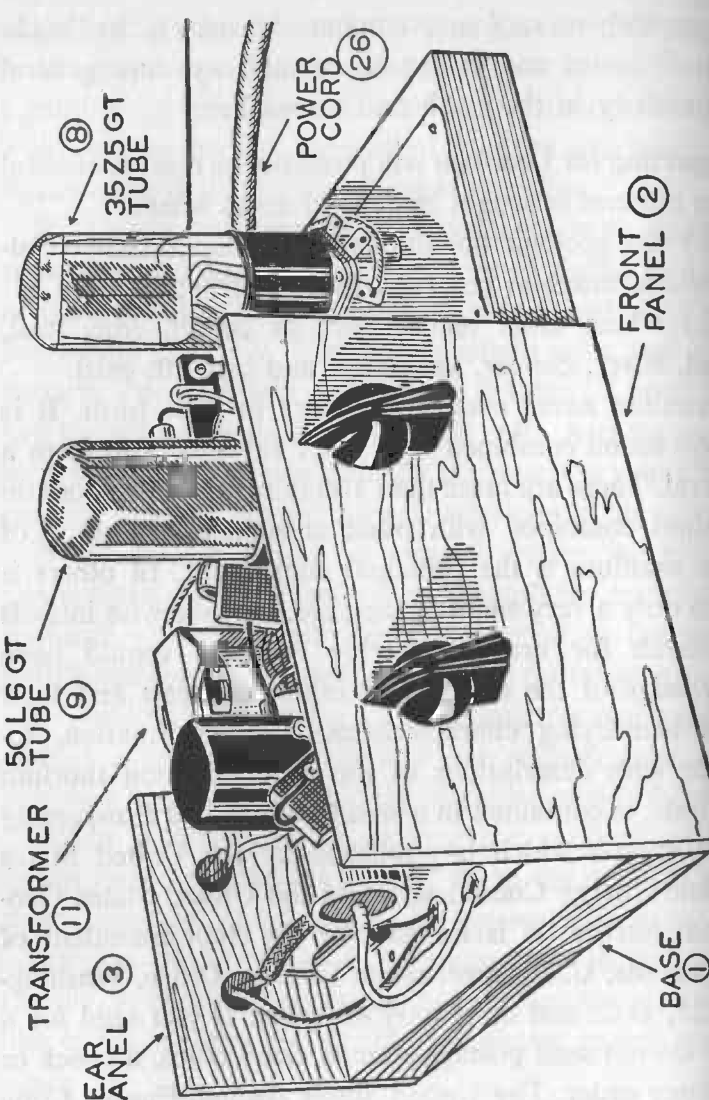

FRONT VIEW OF TWO-STAGE AUDIO-AMPLIFIER

CHAPTER FOURTEEN

How to Build a Two-stage Audio Amplifier

This simple, low cost amplifier is not “hi-fi.” Nevertheless it will provide good musical reproduction when connected to a record player with a crystal pickup. An amplifier which could be rated “hi-fi” costs more and is not so easy to build.

Resistance coupling is used between the amplifier stages and the pickup. An “on-off” switch and tone and volume controls are provided. The amplifier tubes are a 12SQ7 and a 50L6. Plate current is supplied by a 35Z5 half-wave rectifier. All three tubes fit standard octal sockets. 12SQ7 is a metal-envelope, multi-unit tube which contains two diodes and a triode. The triode portion only is used in this amplifier. The triode in a 12SQ7 has characteristics which adapt it for use in resistance coupled amplifiers. The 50L6 is a beam power pentode. In other words, it is a five-element screen grid tube designed for use in the output stage of low-powered amplifiers. 50L6 is frequently used in the output stage of broadcast receivers.

The amplifier is assembled on a wood base or chassis and can be of the open type or enclosed in a cabinet. The wood parts for the open and for the cabinet type are listed separately.

The components and materials required to build the two-stage audio amplifier are:

3 Octal wafer sockets (**6**)
6 Metal spacers ¾ in. long for supporting sockets (**7**)
1 35Z5 radio tube (**8**)
1 50L6 radio tube (**9**)
1 12SQ7 radio tube (**10**)
1 50L6 output transformer with tapped secondary (**11**)
1 1-megohm volume control with SPST switch and knob (**12**)
1 ½-megohm volume control and knob (**13**)
1 Dual section 50-30 mfd., 150-volt electrolytic capacitor with common negative terminal (**14**)
1 .02 mfd., 600-volt tubular capacitor (**15**)
1 .004 mfd., 600-volt tubular capacitor (**16**)
2 .01 mfd., mica capacitors (**17**)
1 150-ohm, 10-watt wire wound resistor (**18**)
1 22,000-ohm, 1-watt carbon resistor (**19**)
2 0.47-megohm, ½-watt carbon resistors (**20**)
1 .5-megohm, ½-watt carbon resistor (**22**)

2 3-terminal Bakelite tie-points (**23**)
2 Phono pin plugs and jacks (**24**)
1 117-volt plug-in receptacle (**25**). This part is optional.
1 6-ft. length of lamp cord with attachment plug (**26**)

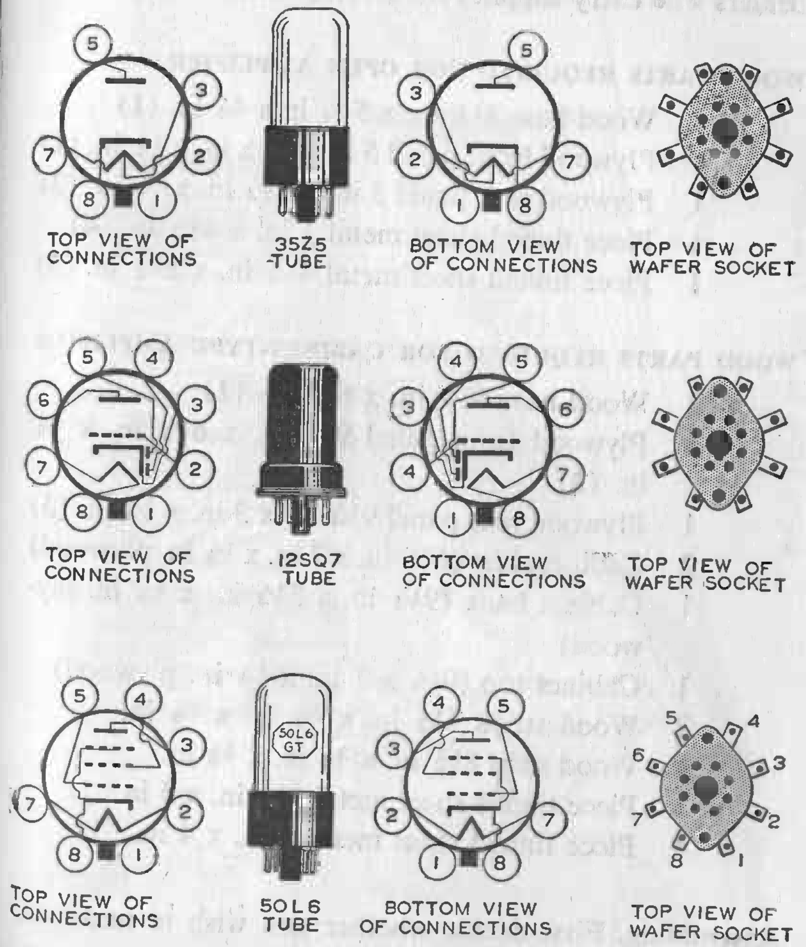
TOP VIEW OF CONNECTIONS
35Z5 -TUBE
BOTTOM VIEW OF CONNECTIONS
TOP VIEW OF WAFER SOCKET
TOP VIEW OF CONNECTIONS
12SQ7 TUBE
BOTTOM VIEW OF CONNECTIONS
TOP VIEW OF WAFER SOCKET
TOP VIEW OF CONNECTIONS
50L6 GT
50L6 TUBE
BOTTOM VIEW OF CONNECTIONS
TOP VIEW OF WAFER SOCKET
1
2
3
4
5
6
7
8

THE TUBES FOR THE AMPLIFIER

Sandpaper, shellac, varnish, push-back wire, shielded phono wire, rosin-core solder, miscellaneous screws and wire brads.

All of the components in the above list are standard electronic parts which can be purchased from radio dealers who carry supplies for service men and amateurs.

WOOD PARTS REQUIRED FOR OPEN AMPLIFIER

1	Wood base 8⅛ in. x 5¾ in x ¾ in. (1)
1	Plywood front panel 5 in. x 3½ in. x ¼ in. (2)
1	Plywood rear panel 3 in. x 2½ in. x ¼ in. (3)
1	Piece tinned sheet metal 7 in. x 4½ in. (4)
1	Piece tinned sheet metal 4½ in. x 2½ in. (5)

WOOD PARTS REQUIRED FOR CABINET-TYPE AMPLIFIER

1	Wood base 9¼ in. x 6½ in. (1)
1	Plywood front panel 9¼ in. x 6½ in. x ¼ in. (2)
1	Plywood rear panel 9¼ in. x 3 in. x ¼ in. (3)
2	Cabinet sides (6½ in. x 7 in. x ¼ in. plywood)
1	Cabinet back (9¼ in. x 3½ in. x ¼ in. plywood)
1	Cabinet top (9¼ x 7 in. x ¼ in. plywood)
2	Wood strips 5½ in. x ¾ in. x ⅜ in.
1	Wood strip 8½ in. x ¾ in. x ⅜ in.
1	Piece tinned sheet metal 8½ in. x 5 in.
1	Piece tinned sheet metal 5 in. x 4 in.

Assembling. First decide whether you wish to make an amplifier which is enclosed in its own cabinet or an am-

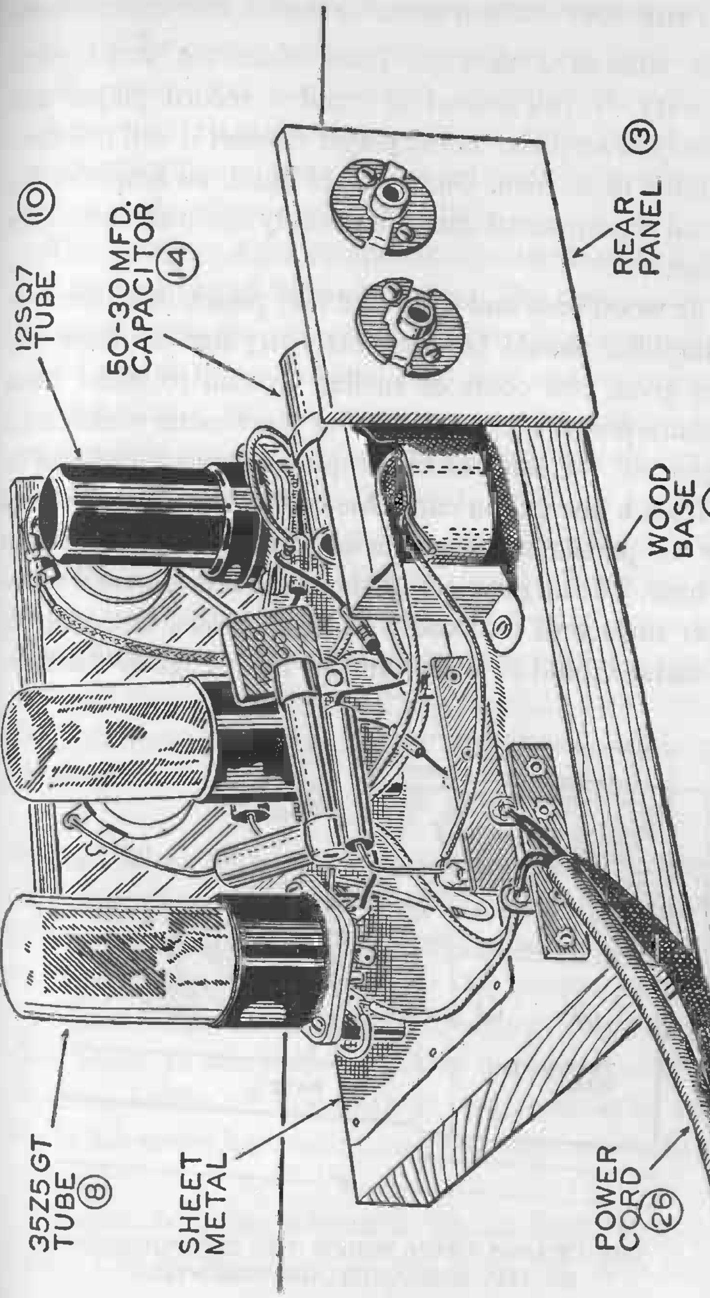

REAR VIEW OF THE AMPLIFIER

plifier without a cabinet. Then make the wood parts necessary. If you intend to build a record player and enclose the amplifier in the player cabinet it will not need a cabinet of its own. On the other hand, an amplifier for general experimental use is preferably enclosed in its own cabinet.

The wood base and front and rear panels of either type of amplifier should be completely dry and all their surfaces given two coats of shellac varnish to make them moisture-proof. The two pieces of sheet metal which form a "ground" for some of the amplifier components can be cut from a one-gallon can. One of the illustrations shows how the panels and sheet metal pieces are assembled on the base. The larger piece of sheet metal is fastened to the upper surface of the base with small brads. One edge of the metal should be flush with the front edge of the base.

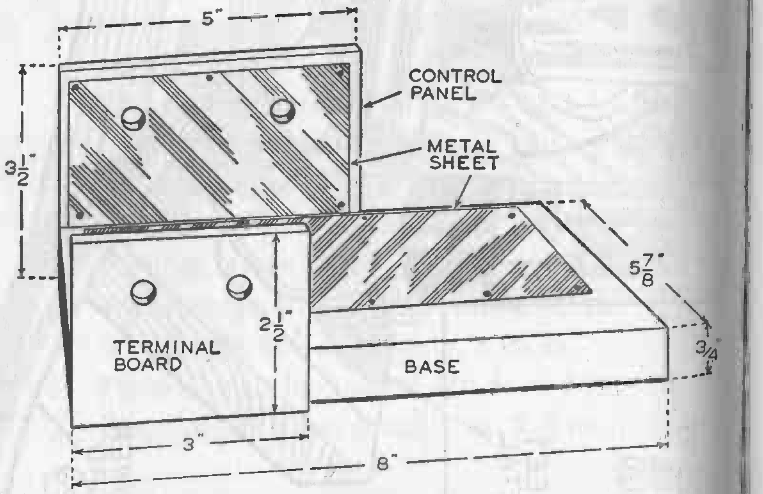

THE CHASSIS UPON WHICH THE COMPONENTS OF THE AMPLIFIER ARE MOUNTED

The smaller metal piece is fastened to the back of the front panel (the panel which supports the tone and volume controls). The two pieces of sheet metal should be electrically connected to each other by soldering at several spots where the edges adjoin.

The location of the components on the base and panels can be determined from the plans. The sockets **(6)** are mounted with spacers so that the socket terminals are raised above the metal plate (**4**) and cannot come into contact with it. If the sockets are located exactly in the positions shown in the plans, the wiring will be easier to follow. The two Bakelite tie-points **(23)** are anchorages for the terminals of the power cord and several of the resistors and capacitors. The electrolytic capacitor **(14)** is fastened to the base by a round head wood screw which passes through the metal mounting strap which is supplied with the capacitor. The SPST switch **(12)** is mounted on the volume control **(12)** before the latter is put in place on the panel.

Wiring. When all the components which are fastened to the base and panel are in place, the amplifier is ready for wiring. Use push-back wire for making the connections, except the wire which connects the phono jack to the volume control. This should be a piece of shielded phono wire. Make all connecting wires as direct and as short as possible. Solder all connections with rosin-core solder. When the solder has cooled, test each connection by pulling on it.

Consult both the schematic circuit diagram and the plans. The wiring shown in the plans has been divided between two illustrations to make it clearer. One illustration, in which the wires are heavy black lines, shows the power wiring—in other words, the power supply to the heaters and the 35z5 rectifier. The second plan shows the audio wiring.

Each resistor and each capacitor must be identified before it can be connected in the circuit. The capacity or resistance of some of these components may be printed on them. Others must be identified from the colors of their dots or bands. Make certain to connect the 50 mfd. section and the 30 mfd. section on the electrolytic capacitor exactly as shown in the illustrations.

The 50L6 output transformer (**11**) has six metal terminals connected to the secondary winding. They are numbered from 1 to 6. The No. 1 terminal should be connected to one of the terminals on the OUTPUT phono jack (**24**).

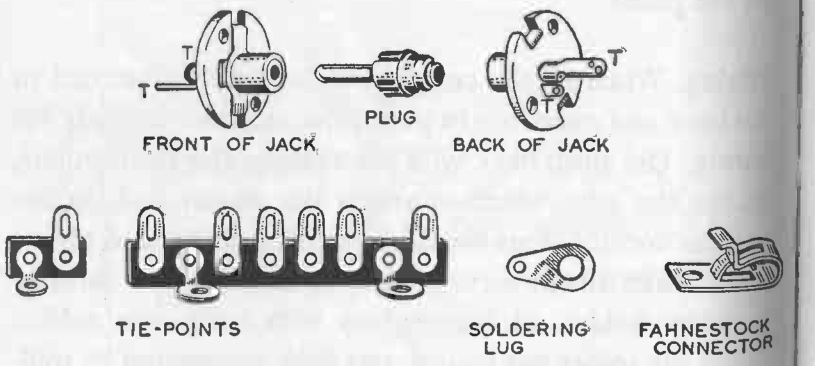
T
T
FRONT OF JACK
PLUG
T
T
BACK OF JACK
TIE-POINTS
SOLDERING LUG
FAHNESTOCK CONNECTOR

PHONOGRAPH JACK AND PLUG, TIE-POINTS, ETC., USED IN BUILDING THE AMPLIFIER

One of the tie-points illustrated above has six terminals. Only three are necessary on the tie-points used for the amplifier.

The other OUTPUT phono jack terminal should be connected to one of the other secondary terminals on the output transformer. The proper terminal is selected by testing when the amplifier is connected to a speaker and in operation. The terminal which gives the best tone is a "match" for the speaker and is the one which should be used.

There are three terminal wires connected to the primary winding of the 50L6 output transformer. Only two are used in this amplifier. Usually one wire is dark brown, one is green and one is white. The white wire is a center

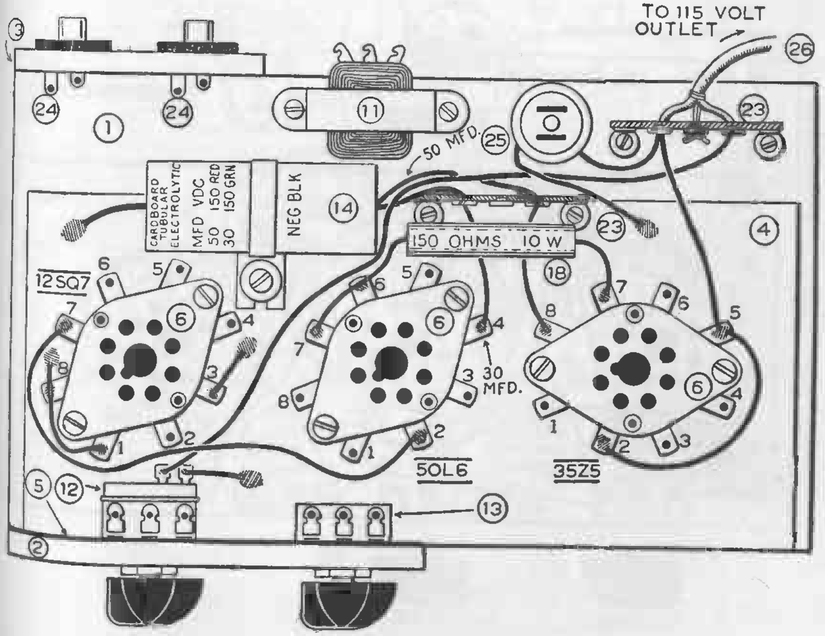
TO 115 VOLT OUTLET
26
3
24
24
1
11
25
50 MFD.
23
CARDBOARD TUBULAR ELECTROLYTIC
MFD VDC
50 150 RED
30 150 GRN
NEG BLK
14
150 OHMS 10 W
23
4
18
12SQ7
30 MFD.
50L6
35Z5
5
12
13
2

PLAN OF THE AMPLIFIER SHOWING THE POWER WIRING TO THE TUBE HEATERS AND PLATES

tap, that is, connected to the center of the primary winding. Use only the brown and green terminal wires. Connect one to the No. 3 terminal on the 50L6 tube socket. Connect the other to one of the terminals on the nearest Bakelite tie-point.

The outside terminal on the INPUT phono jack (24) is "grounded" by a short wire which connects it to the metal ground plate on the base. Remember to use a piece of shielded phono cable to connect the inner terminal of the INPUT phono jack to one of the outer terminals on the volume control. Use the wire conductor inside the cable

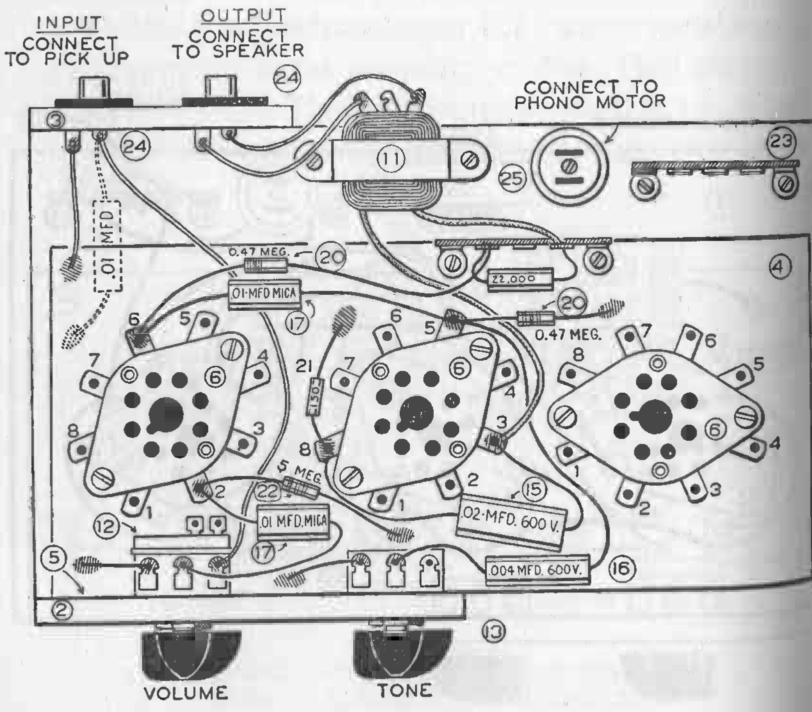
INPUT
CONNECT
TO PICK UP
OUTPUT
CONNECT
TO SPEAKER
CONNECT TO
PHONO MOTOR
③
㉔
㉔
⑪
㉓
㉕
.01 MFD
0.47 MEG.
⑳
.01 MFD MICA
22,000
⑳
0.47 MEG.
④
⑰
21
5 MEG
㉒
.01 MFD. MICA
.02 MFD. 600 V.
⑮
.004 MFD. 600V.
⑯
⑫
⑤
②
⑬
VOLUME
TONE

PLAN OF THE AMPLIFIER SHOWING THE WIRING OF THE AUDIO CIRCUIT

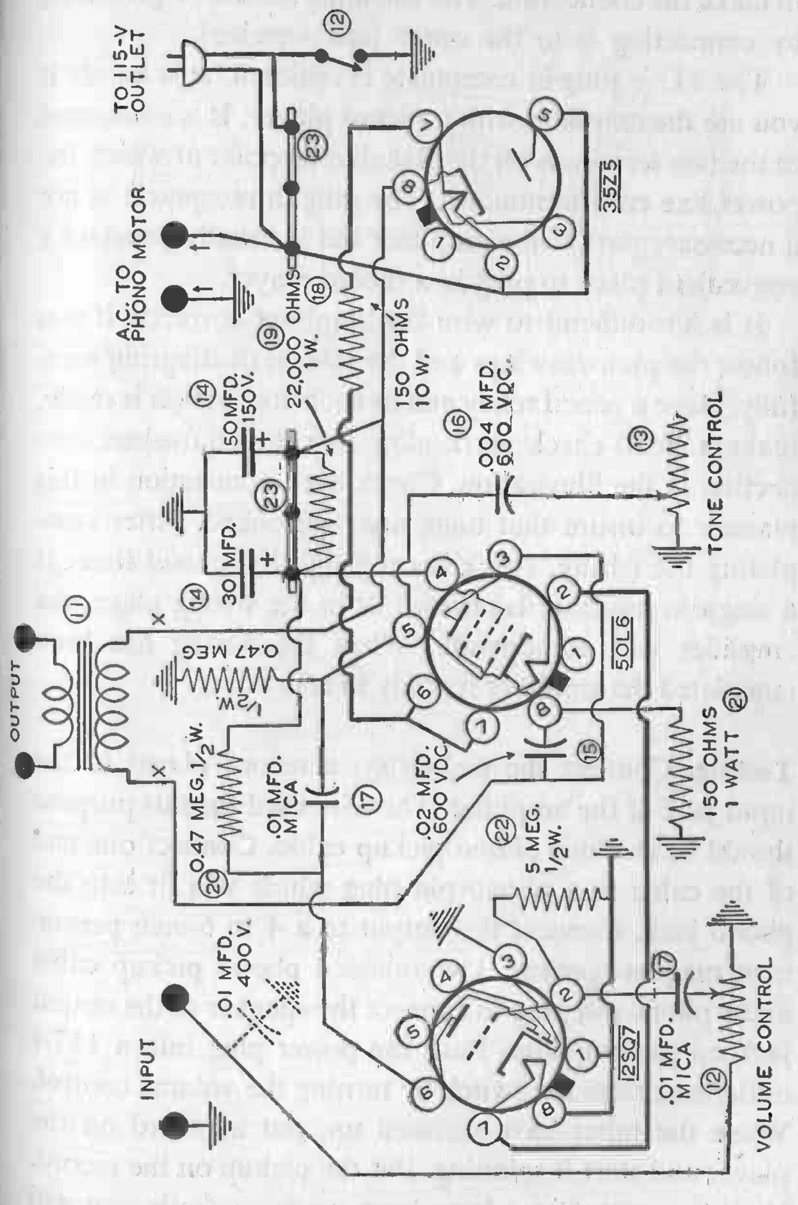
TO 115-V
OUTLET
⑫
35Z5
㉓
A.C. TO
PHONO MOTOR
⑱
⑲
22,000 OHMS
1 W.
150 OHMS
10 W.
⑭
50 MFD.
150 V.
㉓
30 MFD.
⑭
.004 MFD.
600 VDC.
⑯
⑬
TONE CONTROL
OUTPUT
⑪
X
X
0.47 MEG
1/2 W.
50L6
⑳
0.47 MEG. 1/2 W.
.01 MFD.
MICA
⑰
.02 MFD.
600 VDC.
㉒
5 MEG.
1/2 W.
⑮
150 OHMS
1 WATT
㉑
.01 MFD.
400 V.
INPUT
12SQ7
⑰
.01 MFD.
MICA
⑫
VOLUME CONTROL

SCHEMATIC CIRCUIT DIAGRAM OF TWO-STAGE AUDIO AMPLIFIER

If the amplifier is to be connected to a speaker equipped with a matching transformer, omit transformer (11) and connect the two wires marked **X** directly to the output jack.

to make the connection. The shielding should be grounded by connecting it to the outer jack terminal.

The 117-v plug-in receptacle is optional. It is handy if you use the amplifier with a record player. It is connected to the two terminals on the Bakelite tie-point at which the power line cord terminates. The plug-in receptacle is not a necessary part of the amplifier but it usually provides a convenient place to plug-in a record player.

It is not difficult to wire the amplifier correctly if you follow the plan drawings and the schematic diagram carefully. Have a pencil ready and as each connection is made, make a small check mark alongside the equivalent connection in the illustration. Check each connection in this manner to insure that none are overlooked. After completing the wiring, check it carefully the second time. If a single connection is omitted or in the wrong place, the amplifier will not operate. When the wiring has been completed the amplifier is ready to test.

Testing. Connect the pickup on a record player to the input jack of the amplifier. The wire used for this purpose should be shielded phono pickup cable. Connect one end of the cable to a phono pin plug which will fit into the phono jack. Connect the output to a 4 to 6-inch permanent magnet speaker. Use shielded phono pickup cable and a phono pin plug to connect the speaker to the output jack on the amplifier. Push the power plug into a 117-v outlet and close the switch by turning the volume control. When the tubes have warmed up, put a record on the player and start it spinning. Put the pickup on the record. If all the connections have been made correctly you will

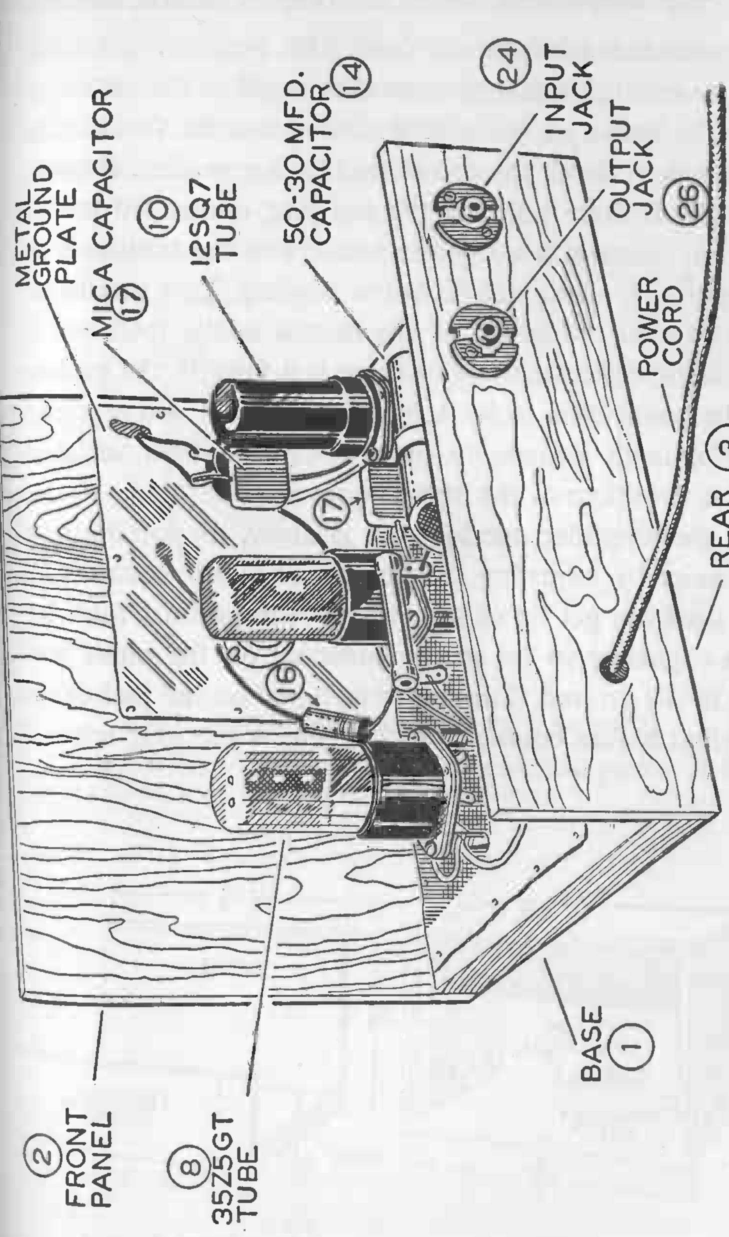

THE AMPLIFIER ON A CHASSIS WHICH CAN BE COMPLETELY ENCLOSED

Note that no matching transformer is shown. The transformer (11) is omitted when the speaker used is equipped with a transformer.

be rewarded with music from the speaker. Use the volume control and tone control to regulate the sound.

If the music of the record does not come forth from the speaker, check the connection to the speaker. If there is no fault there pull out the pin plug connected to the pickup. Connect a telephone receiver to the terminals of the pin plug while the record is playing. You should be able to hear the sound of the record in the receiver. If you cannot hear the record there is a fault in the pickup or the connection to it. A magnetic pickup will not produce enough volume to operate the amplifier satisfactorily. A pickup of the crystal type must be used.

If the amplifier operates but is noisy, try to eliminate the noise by adjusting the volume and tone controls. If this does not get rid of the noise connect a 0.01 mfd. bypass capacitor to the center terminal on the input jack and to the ground. The center terminal on the jack is the one that makes contact with the pin on the plug when it is shoved in place. This capacitor and its connection are shown by dotted lines in two of the drawings. Remember that noise can also be caused by a worn record, a worn stylus or a damaged pickup.

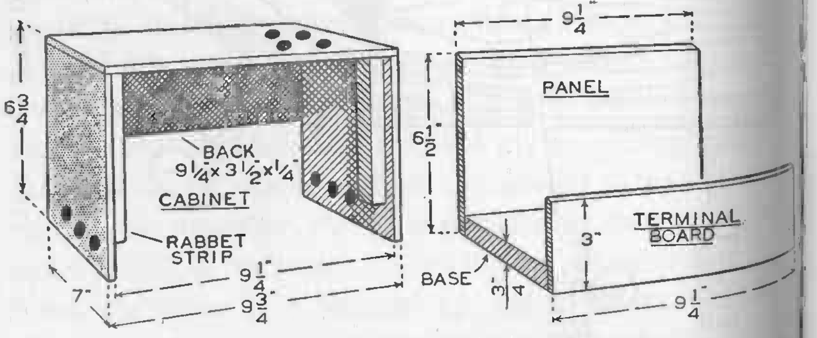
6$\frac{3}{4}$
BACK
9$\frac{1}{4}$x3$\frac{1}{2}$x$\frac{1}{4}$
CABINET
RABBET STRIP
9$\frac{1}{4}$"
9$\frac{3}{4}$
7"
6$\frac{1}{2}$"
9$\frac{1}{4}$"
PANEL
TERMINAL BOARD
3"
BASE
$\frac{3}{4}$"
9$\frac{1}{4}$"

CABINET AND CHASSIS FOR AN ENCLOSED MODEL OF THE AMPLIFIER

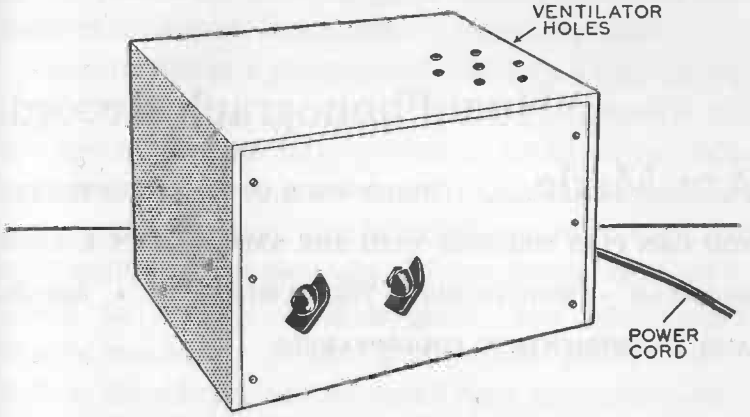
VENTILATOR HOLES
POWER CORD

THE AMPLIFIER ENCLOSED IN A CABINET

The volume and tone control knobs are on the front panel so that they can be adjusted without opening the cabinet.

How Phonograph Records Are Made . . .

BUILD YOUR OWN RECORD PLAYER. YOU CAN PLAY RECORDS WITH THE AMPLIFIER IN A RADIO RECEIVER • HOW TO BUILD THE HUSH-A-TUNE • REPAIRS AND ADJUSTMENTS TO LOUDSPEAKERS

When Thomas A. Edison invented "a machine that could talk" it brought a new form of entertainment into the world. It provided so much pleasure and happiness that for several decades little attention was paid to its principal defect, namely, the quality of its sounds. The early phonographs were purely mechanical devices and were "squawk boxes" in comparison to the realistic sounding phonographs of today. For a long time there was only one way to engrave the original wax recording from which other records were to be made. It was a purely mechanical method in which the energy used to engrave the wax was the energy of sound. Therein lay the difficulty. Some sounds do not possess enough energy to cut a faithful

recording in wax. No better method of making a recording was available until the electronic amplifier was invented.

Modern phonographs are electronic. The energy of electrons is added to the energy of sound in making an original recording. The energy of moving electrons also produces the sounds heard when a record is played.

The cartridge in a phonograph pickup is a tiny electrical generator. When the point of the stylus clamped in the cartridge follows the wavy groove in a record, the sideways motion of the stylus causes the cartridge to produce an electromotive force. This is fed into an audio amplifier which multiplies the electrons and gets enough of them in motion to operate a loudspeaker. The loudspeaker changes the energy of the electrons into sound waves. Modern records are first recorded on a magnetic tape.

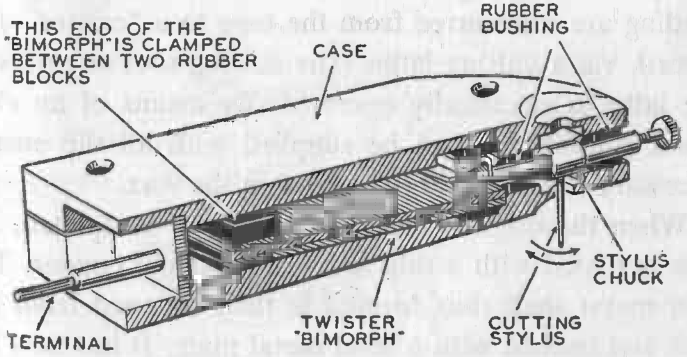

THE CARTRIDGE IN A PHONOGRAPH PICKUP IS A TINY ELECTRICAL GENERATOR

This shows a pickup cartridge of the crystal type. Part of the case is shown cut away to reveal the crystal (called a "twister bimorph"). Any sideways movement of the stylus twists the crystal and produces an electromotive force. The case of the pickup is one terminal. The other is a brass pin insulated from the case.

The first magnetic tape recorder was developed in Denmark about 1900 by Valdemar Poulsen. Incidentally, Poulsen made a great many important inventions in the field of wireless telegraphy and telephony. Poulsen's recorder, known as the Telegraphone, was sold and used for commercial purposes (dictation, etc.) in the ensuing years. The Telegraphone made its recordings on a small steel wire. When the Allies occupied Germany they found recorders which substituted a special tape for Poulsen's steel wire. The use of plastic tape, coated with a thin layer of fine particles of magnetic iron oxide, produced recordings far superior to those obtained with wire.

A tape recording can be played back without any loss of quality and artists can listen to the music immediately after it has been recorded. If no "bloopers" have occurred and the "master" tape is approved it is sent to the mold or stamper-making department. The sounds on the tape recording are transferred from the tape to a "master" disk record, via a cutting lathe. The cutting tool or stylus in the lathe is electrically operated. By means of an electronic amplifier it can be supplied with all the energy necessary to cut a faithful record in the wax.

When the original recording has been completed, the wax is plated with a thin layer of gold and copper. The thin metal shell thus formed is then stripped from the wax and backed with a solid metal plate. It has now become a "master record" used to produce a "mother." The "stompers" or molds for pressing a plastic into the records which are shipped to the music stores are made from the mother.

Assembling and installing record players, amplifiers, speakers, etc., to produce highly faithful facsimiles of the sounds which made an original recording is a widespread hobby. There are catalogs which list a great variety of components required for high fidelity sound systems. Most young experimenters must look at these with longing and realize that cost places them out of their reach.

There are also catalogs which list a great variety of popularly priced phonograph components. You do not have to win the Irish Sweepstakes before you can afford to build your own record player or phonograph. There are good phono motors priced at less than four dollars. You can buy a standard pickup and cartridge for less than the cost of a motor.

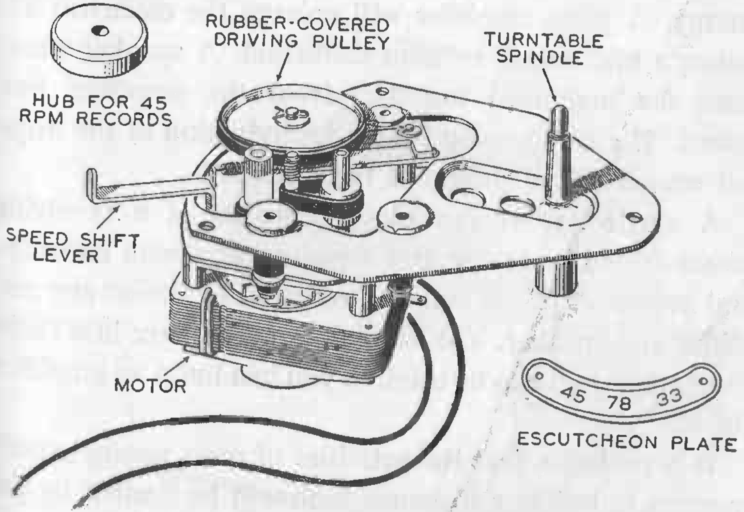

THREE-SPEED PHONOGRAPH MOTOR

The escutcheon plate is fastened to the mounting board under the speed shift lever to indicate the position of the lever which will produce the desired speed of the turntable.

BUILD YOUR OWN RECORD PLAYER

The modern phonograph consists of a revolving motor-driven turntable, a pickup arm with cartridge and needle, an audioamplifier and a loudspeaker. The needle or stylus, whichever you prefer to call it, is mechanically connected to the cartridge. As the needle follows the wavy groove in the record (the groove on one side of an LP record is ½ mile long)it vibrates from side to side. At times it vibrates 10,000 times per second, which is 50 times faster than the beating of a humming bird's wings. The vibrations of the needle create small electrical impulses in the pickup cartridge. The electrical impulses are an exact electrical replica of the sound vibrations which made the original recording. The electrical impulses are fed into an amplifier. Here they are given additional energy. A good amplifier will enlarge the electrical impulses a millionfold without distortion. A speaker translates the magnified impulses from the amplifier into sound. The result—you hear a reproduction of the original sounds which produced the recording.

A single-play record player consists of a revolving motor-driven turntable and a pickup arm with cartridge and needle. A player can be connected to almost any amplifier and speaker. The amplifier and speaker in a radio or television set can be used, or you can build an amplifier for it.

It is probable that the activities of most young experimenters in building a phonograph will be limited by the cost of the components. Why not start by building a single-play record player? It can be transformed into a complete phonograph later if you desire. When you have

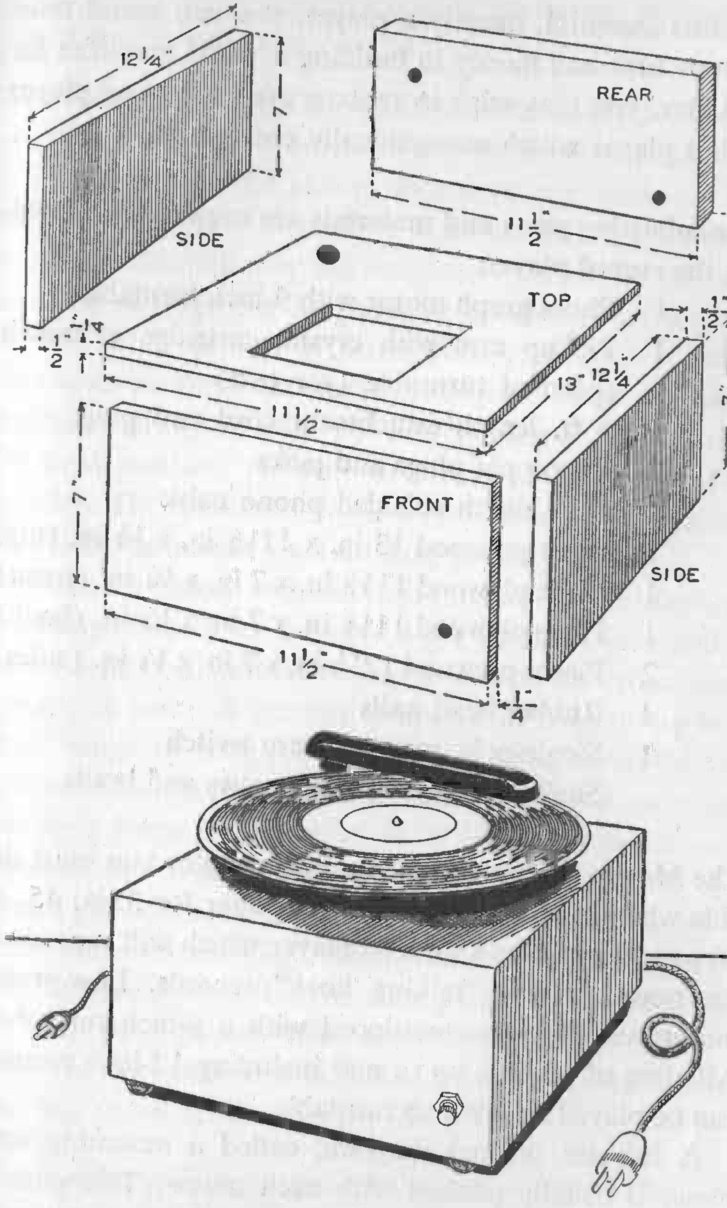

THE HOMEMADE RECORD PLAYER
AND DETAILS OF THE CABINET

the first essential, namely a player, you can spend most of your time and money in building a "hi-fi" amplifier for it. Later, you may wish to replace the single-play player with a player which automatically changes records.

The following parts and materials are required for building the record player:

1	Phonograph motor with 9-inch turntable
1	Pickup arm with crystal cartridge to match speed of turntable (*see text*)
1	6 ft. length attachment cord and plug
2	Phono pin plugs and jacks
1	3 ft. length shielded phono cable
1	Piece plywood 13 in. x 11½ in. x ¼ in. (top)
1	Piece plywood 11½ in. x 7 in. x ¼ in. (front)
1	Piece plywood 11½ in. x 7 in. x ½ in. (back)
2	Pieces plywood 12¼ in. x 7 in. x ½ in. (sides)
4	Rubber head nails
1	Single-pole rotary fixture switch
	Shellac, stain, varnish, screws and brads

The Motor. Before you buy a phono motor you must decide whether you wish to build a player for 33⅓, 45, or 78 RPM records or a 4-speed player which will reproduce the new 16 RPM "talking book" records. Low-priced motors usually come equipped with a 9-inch turntable. All sizes of records up to and including 12-inch records can be played on a 9-inch turntable.

A full-size printed diagram, called a mounting template, is usually packed with each motor. This pattern shows the size and shape of the holes which must be cut in the motor board to accommodate the motor. When you buy a motor, get a template for it.

The Pickup. The pickup which you use in building your record player should match the motor. If you decide upon a 78 RPM motor the pickup should be one made for 78 RPM records. If you have a three-speed motor, buy a pickup which will play the standard 78 RPM records and also the microgroove long-playing records (33⅓ and 45 RPM). Some models of three-speed pickups are called "turnover pickups" and are provided with a knob for turning the cartridge to either the long-playing or the standard 78 RPM position. A turnover pickup is fitted with two needles. The needle for 78 RPM records is larger in diameter than the needle used for long-playing records.

Choose a pickup with a crystal cartridge, not a magnetic cartridge. The latter requires a pre-amplifier in addition to the regular amplifier. Pickup arms and cartridges vary greatly in price. A cartridge capable of responding to a wide range of sound frequencies is most desirable but frequency range largely determines the price. A cartridge having a frequency range of 40 to 10,0000 cps will pick up recorded sounds ranging in pitch from 40 to 10,000 cycles per second. The player which you are building in this instance is not a high fidelity instrument. The speaker and amplifier do not have a frequency range of 40 to 10,000 cps. A pickup with a cartridge having a frequency range of 50-4,000 costs less than one with a wider range and will prove satisfactory for this player.

The Cabinet. This part of the player is not difficult to make. It is merely a box and a box is considered to be one of the simplest woodworking projects. The only skill necessary is that required to saw straight and accurately. It is for you to decide how carefully you wish to make the cabinet and how well you wish to finish it. The player will operate as well if the cabinet is unfinished lumber as it will if the cabinet is polished mahogany.

The cabinet for the model which was built for this book was made entirely of plywood. It was smoothed with fine sandpaper, then given a coat of light maple stain, a coat of white shellac and a coat of varnish. The shellac coat was rubbed with fine steel wool before the varnish was applied. It has been much admired.

It is not necessary to use plywood but it is difficult to find thin lumber which can be substituted for plywood. It is advisable to make the front of the cabinet and the motorboard of plywood. The motorboard is the flat top that supports the motor and pickup. These parts are only ¼-inch thick and solid lumber of that thickness would not be strong and rigid enough. The sides and back can be ½ to ¾-inch pine or any other wood.

If you do not have woodworking tools, perhaps you can make the cabinet in your school workshop. The secret of making a neat cabinet is to cut the parts accurately and smoothly. Sawing plywood often leaves a rough edge but it can be cut smoothly. It can be cut by hand satisfactorily by using a sharp fine-toothed saw. Lay the plywood on a piece of thin scrap lumber and cut through both pieces. This will prevent splintering the underside of the plywood. If a power circular saw is used, the saw blade should be one made for cutting plywood.

The location, size and shape of the holes in the motorboard depend upon the motor and pickup which is used. Use the templates which come with the motor and pickup to locate the holes. Cut the large motor hole with a power jig saw or a hand fret saw.

Use a piece of paper the same size as the top of the cabinet to make a pattern. Lay the templates for the pickup and motor on the pattern and move them into position so that:

1. When the pickup arm is swung from side to side, a needle in the pickup would move in an arc which passes through the center of the turntable.

2. The distance from the left hand edge of the cabinet top to the center of the turntable equals the distance from the right hand edge.

3. The center of the turntable will be 6 inches from the front edge.

Make all the holes in the top, front and back before assembling the cabinet. Use both glue and small brads to fasten it together. Drive the brad heads below the surface with a small nail set and fill with plastic wood. When the glue has dried smooth the outside of the cabinet with fine sandpaper. Then stain it. When the stain has dried apply a coat of white shellac. When the shellac has dried rub it smooth with No. 0 steel wool. Finish with a coat of varnish.

If you prefer, the cabinet can be painted. Apply first a coat of shellac to fill the wood. Use the variety of paint made for furniture and interior woodwork.

Drive the rubber head nails into the under edge of the front and back near the corners so that the cabinet is raised above any flat surface upon which it is placed.

This is necessary so air can move in and out of the cabinet freely and cool the motor.

Home-made rubber washers act as shock absorbers and prevent vibrations of the motor from being transmitted to the motorboard and so reaching the pickup. The washers can be cut from a piece of inner tube from an automobile tire. Six pieces ¾ inch x ¾ inch are required. Drill a hole through each with a ⅛ inch drill. The motor is fastened on top of the motorboard by three ¾ inch 8-32 R.H. machine screws. Place a rubber washer on each screw between the motor and motorboard and also between the underside of the motorboard and the washer which is slipped over the screw before placing the nut.

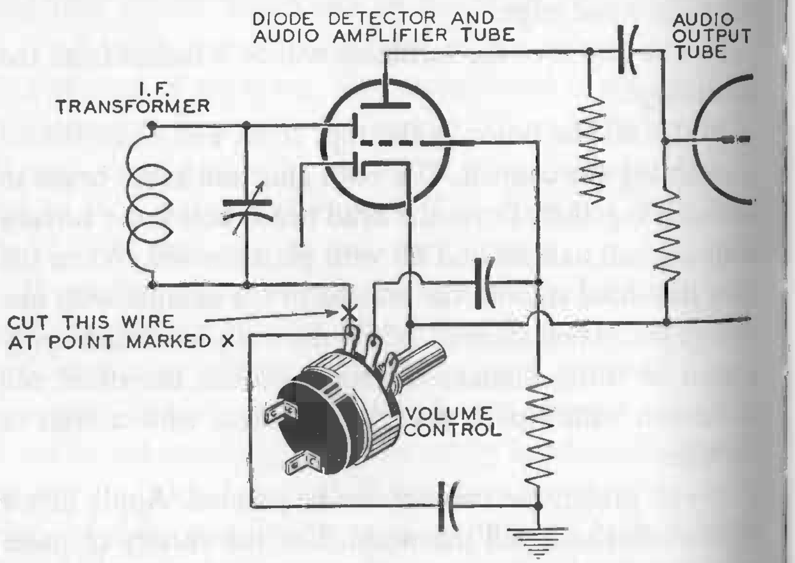
DIODE DETECTOR AND AUDIO AMPLIFIER TUBE
AUDIO OUTPUT TUBE
I.F. TRANSFORMER
CUT THIS WIRE AT POINT MARKED X
VOLUME CONTROL

HOW TO USE THE AMPLIFIER IN A RADIO RECEIVER TO PLAY PHONOGRAPH RECORDS

If the receiver is not equipped with a phono-jack, cut the wire leading from the I.F. transformer to the volume control at the point marked X in the illustration. (*See also next illustration.*)

The power cord passes through a ⅜-inch hole in the back of the cabinet. One of the power wires is connected to the motor. The other is connected to the fixture switch mounted near the front of the cabinet. Run a wire from the other terminal wire on the switch to the motor. When the plug on the power cord is pushed into an outlet the motor can be started or stopped by turning the switch.

YOU CAN PLAY PHONOGRAPH RECORDS WITH YOUR TABLE MODEL RADIO

If you have a record player but no amplifier and loudspeaker for it you can use the audio-amplifier and speaker in a broadcast receiver. Many broadcast receivers are equipped with a phono jack and switch at the factory. Push a phono pin plug connected to a record player into the phono jack, move the switch to the correct position and you have a phonograph.

If your broadcast receiver is not equipped with a phono jack you can install one at a cost of about $1.00. **You will need the following parts:**

1	Shielded phono pin-plug jack
1	Single-pole, double-throw (SPDT) toggle switch
1	12-inch length shielded phono wire
2	½ inch No. 8 self-tapping screws
1	.05 mfd., 600 DC working volts tubular by-pass capacitor
1	12-inch length of spaghetti which will slip over the shielded phono wire
1	Piece of ¼-inch plywood approximately 3 in. x 3½ in.
	Rosin-core solder

Spaghetti is the trade name for varnished cambric tubing used for insulation in radio and electronic circuits.

Mount the phono jack and the STDP switch on the plywood panel and fasten the panel to the rear of the receiver chassis. Drill two holes with a No. 28 drill through the panel and through the rear edge of the steel chassis. Use the No. 8 self-tapping screws to clamp the panel against the chassis.

The jack is to be connected to the input circuit of the audio amplifier. Most broadcast receivers employ a superheterodyne circuit and the input to the audio amplifier in that type of receiver should not be difficult to locate. Suppose that we operate on a 5-tube, table model AC-DC superheterodyne. Here is the procedure:

First, remove the control knobs so that you can take the chassis out of its cabinet after the four screws which hold it in place have been extracted. Turn the exposed chassis upside down and locate the volume control.

There is an ON-OFF switch on the back of the control. There are two terminal lugs on the switch. We are not interested in the switch except that it may be helpful to the novice to identify it and not confuse it with the control. The control has THREE terminal lugs in a row. We are not concerned with the center terminal.

Trace the wires connected to the two OUTSIDE terminals. You will find that one of the outside terminals is connected to a tube and the other is connected to a coil called an IF transformer.

Now you can start "operating." Cut this wire at the point where it is attached to the volume control and solder it to one of the contacts on the SPDT switch. If the wire is not long enough to reach the switch add a piece to it.

Connect the terminal on the volume control from which the wire to the IF transformer was removed to the center terminal on the SPDT switch.

Connect the other switch contact to the phono jack using a piece of shielded phono wire for the purpose. Connect the center wire to the center terminal of the jack. Connect the shield to the outer terminal of the jack.

The shielding on the wire must not touch any of the other terminals or components in the receiver or it will cause a short circuit. To prevent this, slip a piece of insulating spaghetti over the shield.

Connect one terminal of a .05 mfd., 600-volt by-pass capacitor to the shielding on the phono wire. Connect the other terminal of the capacitor to the steel chassis of the radio receiver.

The purpose of the by-pass capacitor is to "ground" the shield to the chassis. The shield could be grounded by a wire which connected it directly to the chassis but a direct ground is undesirable. A radio receiver of the AC-DC variety is likely to have a "hot chassis," that is a chassis that is connected to one side of the power line. If, by chance, the shielded phono wire from the record player should also be "hot" there might be a short circuit and "fireworks." The indirect ground provided by the capacitor eliminates this possibility.

All connections should be soldered with rosin-core solder. When the hookup is complete, restore the receiver to its cabinet, fasten it in and replace the control knobs.

If the SPDT switch which has been added at the back of the set is thrown so that the IF transformer is connected to the volume control, the receiver can be used to pick up broadcast programs in the usual manner. If connections have been made properly the addition of the jack, switch and capacitor will not interfere in any way with the reception of broadcast programs.

To play records through the receiver is no trouble. Push the pin plug connected to the record player into the phono jack. Throw the SPDT switch so that the volume control is connected to the phono jack, turn on the ON-OFF switch on the volume control, put a record on the player and you are "in business." The volume control will regulate the sound of records in the same way that it controls the sound of a radio program. Connect a shielded phone wire to the other switch contact. Connect the other end of this shielded wire to the phono jack. Connect the center wire to the center terminal of the jack and the shield to the outer terminal.

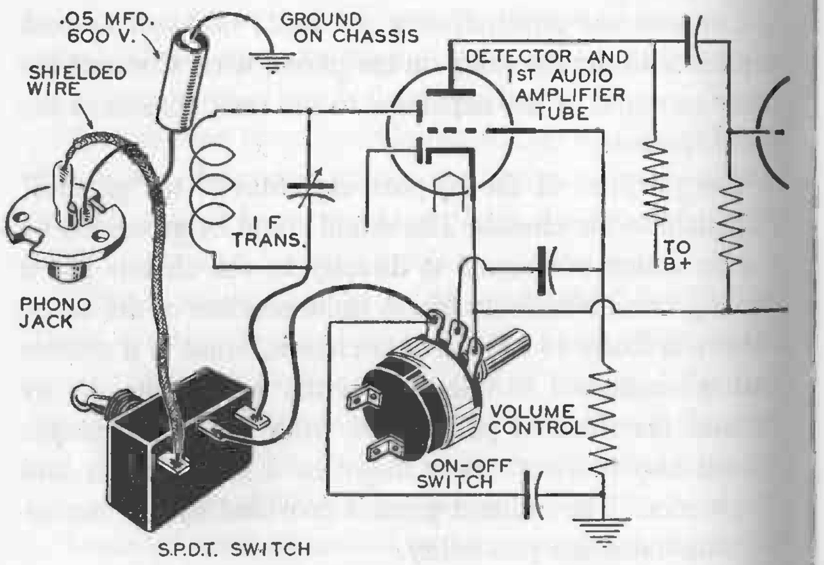

HOW TO USE THE AMPLIFIER IN A RADIO RECEIVER TO PLAY PHONOGRAPH RECORDS

When the wire from the I.F. transformer to the volume control has been cut, install a SPST switch, phono-jack and capacitor as shown above. To play a record, push the phono-plug connected to the record player into the phono-jack. One position of the switch connects the phono-jack to the amplifier. The other position eliminates the jack so that the receiver can be used for radio again.

It is best to avoid a direct ground. Connect one terminal of a .05 mfd., 600 volt by-pass capacitor to the shielding on the phono wire and the other terminal to the chassis of the radio receiver. This cap is used to avoid making a direct ground between the chassis and the shield.

The shielding on the wire must not touch any of the other terminals or components, or it will cause a short circuit. To prevent this, slip a piece of "spaghetti" over the shield.

BUILD YOUR OWN PORTABLE PHONOGRAPH

This is a complete phonograph unit. It does not have to be connected to anything but the 117-v A.C. power supply. Push the plug into an outlet, put a record on the turntable and it is ready to play. The quality of its reproduction is good—equal to that of many factory-built phonographs which cost five or six times the price you pay for parts to build this unit.

With the exception of a few additional holes, a cabinet of the same size and form is used as that described for the record player in preceding pages. The cabinet encloses a phono motor, 3-tube amplifier and 6-inch loudspeaker. There are complete 3-tube AC-DC phono amplifiers on the market, priced at less than ten dollars. These give good reproduction and you can use one of them or use the equally good 3-tube phono amplifier described in Chapter

Fourteen. The instructions which follow are based upon the use of the homemade amplifier. It does not require a great deal of ingenuity to substitute a ready-made amplifier.

The following parts and materials are required:

1	Phonograph motor with 9-inch turntable
1	Pickup arm with crystal cartridge and stylus
1	6 ft. attachment cord and plug
1	Single pole 125-v rotary fixture switch
1	6-inch permanent magnet speaker
2	1 phone pin plugs
1	2 ft. length shielded phono cable
1	Piece rubber inner tube 2 in. x 2 in.
1	8 in. x 8 in. piece of grille cloth
1	3-tube phono amplifier
1	Piece plywood 13 in. x 11½ in. x ¼ in. (top)
1	Piece plywood 11½ in. x 7 in. x ¼ in. (front)
1	Piece plywood 11½ in. x 7 in. x ½ in. (back)
2	Pieces plywood 12¼ in. x 7 in. x ½ in. (sides)
4	Rubber head nails
1	2-terminal Bakelite tie-point
	Shellac, stain, varnish, rosin-core solder, glue, rubber cement, screws and brads.

The information covering the choice of a phono motor and pickup for the record player should be used in their selection for the phonograph.

To make the cabinet, follow the instructions for making the record player cabinet. Make all holes in the top, front, sides and back before assembling the cabinet. The location, size and shape of the holes in the motorboard depend upon the motor and pickup which is used. Use the template which comes with the motor and pickup to locate the holes. Cut the large hole for the motor with a power jig saw or a hand fret saw. (*See instructions for building record player cabinet.*)

The 5⅝-inch diameter circular hole in the front of the cabinet for the loudspeaker can be cut with a power jig saw or by hand with a fret saw. After cutting, smooth the edges of the hole with sandpaper.

Grille cloth to stretch across the hole in the front of the cabinet can be purchased from almost any radio repair shop. The cloth should not be put in place until the cabinet has been stained and varnished.

Cut a paper circle the same size as the hole for the loudspeaker and place it in the center of an 8 inch x 8 inch piece of grille cloth to act as a shield when spreading the cement. Spread rubber cement on that portion of the cloth which is not covered by the template. Also spread rubber cement on the back side of the cabinet front around the edge of the hole over an area (including the hole) equal in size to the piece of grille cloth. When the cement on both the grille cloth and the cabinet has dried for about 10 minutes, put the cloth in place on the back of the cabinet front. There will be a small surplus of cloth at the lower edge of the cabinet front. Trim this off with a safety razor blade.

Another method of putting the grille cloth in place is to cement it to an 8 inch x 7 inch piece of cardboard which has a 5¾-inch hole in the center.

Drill four holes in the cabinet front which will line up with the four holes in the loudspeaker when the latter is centered over the hole. Fasten the speaker to the back of the cabinet front with four ¾-inch 8-32 machine screws. Push the screws through the holes from the front of the cabinet. Place an 8-32 hex nut on the end of each screw and tighten.

Use homemade shock-absorbing washers in mounting the motor.

The power cord passes through a hole in the back of the cabinet and terminates on a two terminal Bakelite tie-point strip fastened to the inside of the cabinet back. One of the motor terminal wires is connected to one of the tie-point terminals. A wire from the other tie-point terminal goes to the fixture switch on the cabinet front. The other switch terminal is connected to the remaining motor terminal. The fixture switch serves as an on-off switch to stop and start the motor but does not interrupt the power supply to the amplifier.

The amplifier is fastened inside the cabinet at the lower edge of the right hand side. First fasten a strip of plywood 3 inch x ¾ inch x ¼ inch to the front edge of the amplifier base so as to make it flush with the plywood panel which supports the volume and tone controls.

Drill three ⅛-inch holes ½ inch up from the bottom edge of the right hand side of the cabinet as indicated in one of the illustrations. The amplifier is held in place inside the cabinet by three 1¼-inch No. 6 round head wood screws which pass through the holes from the outside and into the edge of the amplifier base. Drill two 5⁄16 inch diameter holes in the right hand side of the cabinet for the shafts of the volume and tone controls.

Connect a phono pin plug to the end of the shielded phono wire which is connected to the crystal cartridge in

the pickup arm. Push the plug into the INPUT jack on the amplifier. Use shielded phono wire to connect the loudspeaker to the amplifier. Connect a phono pin plug to one end of the wire and push the plug into the OUTPUT jack on the amplifier. Connect the other end of the wire to the speaker.

One of the wires connected to the output jack should be permanently connected to the No. 1 terminal on the secondary winding of the 50L6 matching transformer. Determine which of the secondary terminals the other wire should be connected to by test and experiment. Connect it permanently to the terminal which produces the best tones from the speaker when a record is played.

Run a wire from each of the terminals on the tie-point, to which the power cord is connected, to the terminals on the tie-point marked **(23)** on the plan of the amplifier.

The phonograph is now ready to use.

THE HUSH-A-TUNE

Do you ever wish to play a phonograph record but must restrain yourself because the sound would annoy someone who is studying or sleeping or someone who does not approve of your taste in music?

Here is the answer to your problem. Equip your phonograph with a "Hush-a-tune." Then you can play the phonograph and you alone can hear it.

A double-pole, double-throw toggle switch and a phone jack are needed. You can use two binding posts in place of a phone jack. The jack is more convenient.

Drill three holes in the side of the phonograph cabinet so that you can fasten the switch and phone jack there.

An illustration on page 237 shows the connections. One position of the switch will connect the amplifier to the speaker voice coil so the phonograph can be played in the usual manner.

When the switch is thrown to the opposite side, the amplifier is connected to the phone jack. Push the tips on a radio headset, slip the phones on and you are equipped to listen to the phonograph music. The volume control on the amplifier will adjust the sounds in the headphones to suit you. Too much volume will produce distorted sounds.

LOUDSPEAKER REPAIRS AND ADJUSTMENTS

Loudspeakers do not require frequent repairs but they are sometimes damaged and occasionally get out of adjustment. The simplest way to remedy a defective loudspeaker is to replace it with another of the same model. The most economical method, if you have the time, is to repair it yourself.

Off-center Voice Coil. The most common defect is "off-center voice coil." This may be caused by rough handling or a bent frame but more frequently it is due to humidity changes which cause the paper cone to warp. If the voice coil is off center, it will rub against one of the magnet poles and the sound from the speaker will be distorted and "fuzzy." Radio repairmen are familiar with this sound and recognize it immediately.

Old style speakers are equipped with a "centering screw," located either behind the magnet or in the middle of the paper "spider" at the center of the cone. A centering screw makes readjustment simple. You will need a few centering shims. These are narrow strips of thin fiber or

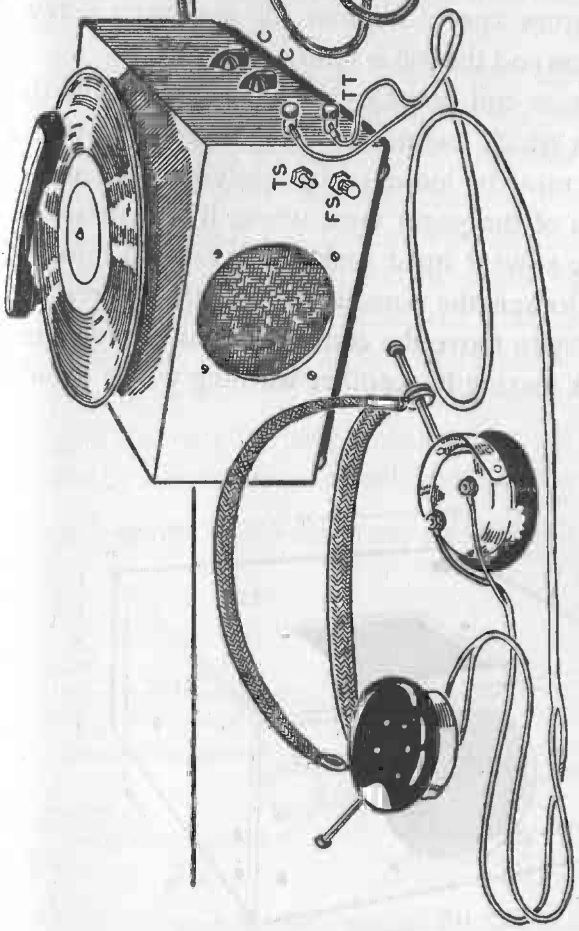

A HOMEMADE PHONOGRAPH ARRANGED SO THAT WHEN DESIRABLE ONLY ONE PERSON CAN HEAR ITS SOUND

FS is the fixture switch for starting the motor. **TT** are terminals for a pair of headphones. **TS** is a toggle switch for disconnecting the speaker and connecting the headphones. **CC** are the tone and volume controls. **PC** is the plug on the power cord.

metal of varying thickness. They are available from radio supply dealers and are priced at about forty cents for a set of twenty.

Loosen the centering screw and insert three shims between the voice coil and the pole piece. Space the shims about 120 degrees apart. Tighten the centering screw, remove the shims and the job is finished.

When the voice coil becomes off center on one of the newer speakers which has no centering screw, the whole cone assembly must be loosened by applying cement solvent to the rim of the paper cone where it is fastened to the frame. The solvent must soak in for several minutes before it will loosen the cement sufficiently to free the cone. Do not try to move the cone until it is entirely free or you will risk tearing it. Another warning which should be heeded is: Do not try to lift the cone from the frame or you will break the connection to the voice coil.

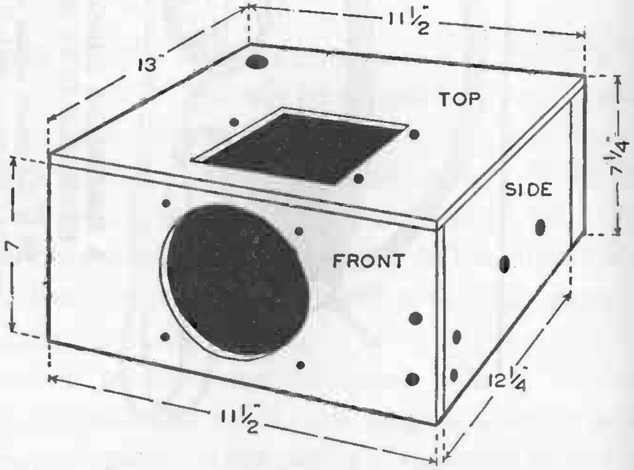

CABINET FOR THE PHONOGRAPH WITH "HUSH-A-TUNE" ATTACHMENT

Apply cement solvent to the small circular dust felt in the center of the cone and remove the felt when the cement has softened. When the felt has been removed the pole piece will be exposed.

When the cone is loose, center the voice coil by inserting three shims of the proper thickness between the voice coil and the pole pieces. Space the shims about 120 degrees apart. With the shims in position recement the cone. When the cement has thoroughly set, remove the shims. Cement and cement solvent in small bottles for loudspeaker repair can be purchased from dealers in radio supplies. The dust cap should be replaced after the shims have been removed. The dust felt prevents dirt, etc., from lodging between the pole piece and the voice coil. Foreign matter in this small space will cause distortion.

Cone Repairs. Some cones are molded and have no seam.

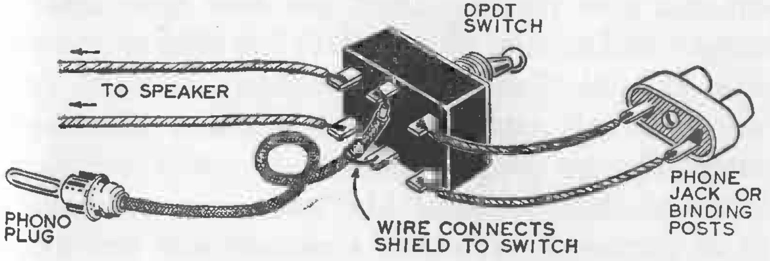

THE "HUSH-A-TUNE" ATTACHMENT

This consists of a phono-plug, double-pole, double-throw switch and a phone jack or pair of binding posts. One position of the switch connects the output of the amplifier to the speaker. The other position connects the amplifier to a radio headset.

Others, of the formed type, have an overlapping seam which is cemented. The cement sometimes dries out and the seam opens. This causes a rattling sound when the speaker is in operation. Put fresh cement along the seam and press the edges together. Use speaker cement or household Duco. Give the cement time to set thoroughly before using the speaker.

Torn cones can often be mended by pushing the torn edges together and applying cement along the tear on both the front and back of the cone.

If the tear or hole is a large one, it is advisable to have a radio repairman install a new cone in the speaker.

Broken Voice Coil Lead. This can be repaired by soldering with rosin-core solder. This defect will cause a "dead" receiver or phonograph. An "open" voice coil where the break is in the coil and not in a lead calls for a new cone and voice coil assembly.

CHAPTER SIXTEEN

Electron Tubes as Oscillators and Some of Their Uses . . .

CAPACITY RELAYS • YOU CAN BUILD A CAPACITY RELAY • FUN WITH A CAPACITY RELAY

What Is an Oscillator? To oscillate means to swing backward and forward. In an oscillating circuit, current and voltage surge back and forth, generally at a high frequency. In radio and electronics the term oscillator means any device for producing electric oscillations, or more specifically, a radio-frequency generator. Edwin H. Armstrong was first to reveal that a properly arranged circuit containing an electron tube could be made to OSCILLATE and thus become a means for converting direct current and voltage into high-frequency alternating current and voltage.

Electron-tube oscillators which generate radio-frequencies are indispensable in radio, industry and warfare and are part of the equipment of scientific, educational and commercial laboratories.

The radio-frequency currents and voltages generated by electronic oscillators are used in transmitters to produce the radio waves which carry radio-telegraph, radio-telephone, broadcast, television and radar signals. There is a small radio-frequency oscillator in every superheterodyne radio receiver.

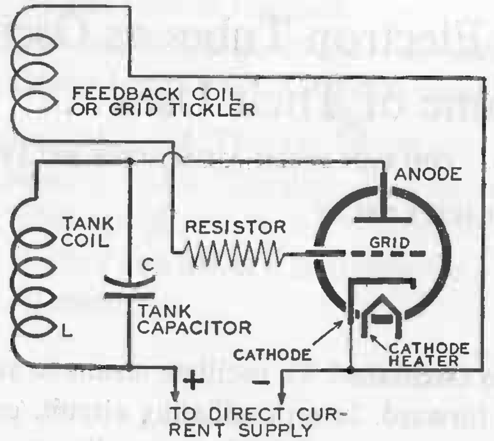

THE SIMPLEST FORM OF ELECTRONIC OSCILLATOR

"There are several types of electron tube oscillators. All provide some means in their circuit to feed part of the power passing through the plate circuit back to the grid. Most oscillators are named for the inventor who first demonstrated the particular type of feedback which is used in the oscillator. The illustration shows the circuit of the simplest type." It is called the magnetic or inductive feedback oscillator. A "tickler coil" which is part of the plate circuit feeds energy from the plate circuit to a "tank coil" which is part of the grid circuit. The energy fed from the plate circuit to the grid circuit appears again in the plate circuit in amplified form. This swinging to and fro of energy is sustained oscillation and will continue as long as power is supplied to the plate circuit. The frequency of the oscillations depends upon the amount of inductance and capacitance in the circuit. An electron tube can be made to oscillate as rapidly as 500,000,000 times per second under proper conditions.

Radio-frequency currents originated by electron-tube oscillators generate the healing heat employed in diathermy treatments. Heat produced in the same manner is used in industrial processes for curing plastics, drying lumber, gluing plywood, surface-hardening steels, to obtain a uniform thin coating of tin on the sheet iron used for cans and in many other ways.

Oscillators are no longer limited to the original type invented by Armstrong. There are several different fundamental types of oscillator circuits and many modifications of each fundamental type.

CAPACITY-OPERATED RELAYS

An electronic oscillator can be arranged and adjusted so that its oscillations cease when a metal plate is brought between its coils. If the circuit includes a relay, the relay

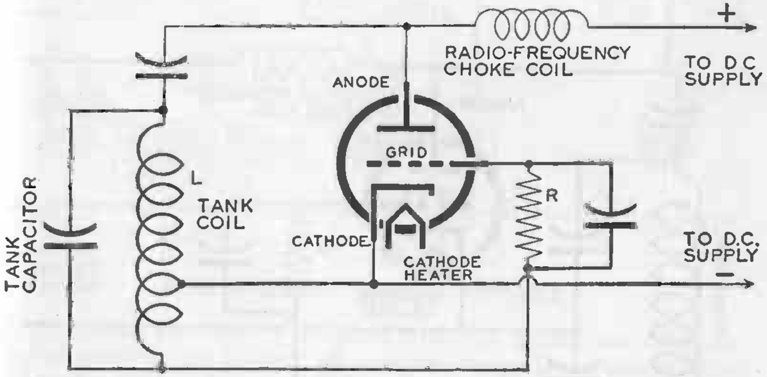

BASIC HARTLEY OSCILLATOR CIRCUIT

This is one of the circuits usually found in high-frequency heating units. A Hartley oscillator utilizes inductive coupling to feed back energy from the plate to the grid circuit. The tank coil is divided so that part of it is in the grid circuit and part is in the plate circuit. The magnetic coupling between the two sections of the tank coil furnishes the feedback which causes oscillations.

will remain open while oscillations occur and will close when the oscillations cease. This arrangement is called a capacity-operated relay. It has several practical industrial applications. It can operate a burglar alarm; it will stop and start machines and perform other useful services WITHOUT BEING TOUCHED. The next time that you ride in an elevator car it may be one which is stopped automatically by CAPACITY RELAYS.

Elevators are often controlled automatically by capacity relays. When the car approaches a floor where a stop is to be made, the elevator's slow-down relays are the first to be actuated. Then the stopping relay stops the car and levels it at the floor. Several capacity-operated relays assembled in a unit and mounted on the top of the car are required to accomplish this. Metal plates are mounted on the hoistway wall or elevator guide rails to actuate the

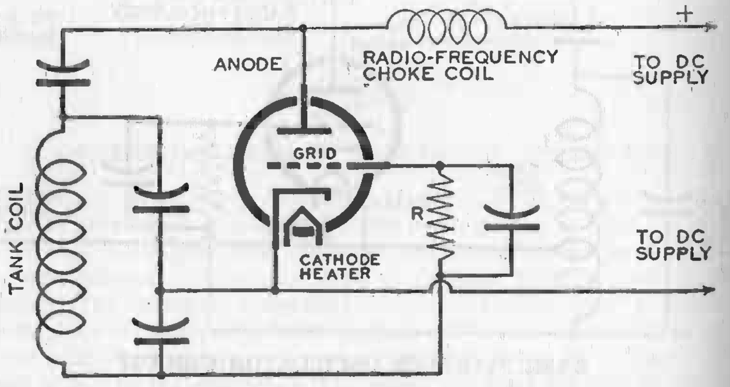

COLPITTS OSCILLATOR

The Colpitts oscillator circuit is also widely used for electronic heating units. By coupling the tank coil to an antenna, the Colpitts oscillator becomes a radio-telegraph transmitter. The Colpitts circuit usually has less frequency drift (change in frequency) with changes in load than some other oscillators.

relays. The metal plates are located so that as the elevator car approaches a floor the plates pass between but do not touch several pairs of coils. Each set of coils is part of a capacity relay circuit. When a plate passes between two coils a relay closes.

The operation of a capacity relay depends upon the action called BODY EFFECT. The effect which the presence of the hand or body near electronic apparatus has upon the circuit is called body effect.

Explanation of Body Effect. A metal plate which is insulated from the ground and from other surrounding ob-

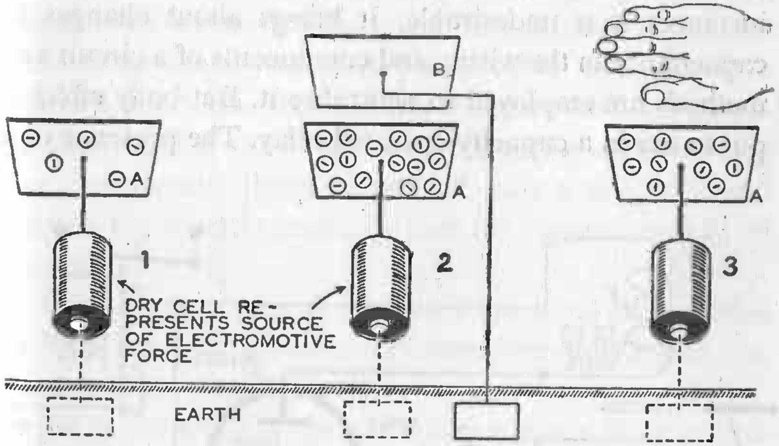

EXPLANATION OF BODY EFFECT

1. If a metal plate (A) is insulated from the ground and other surrounding objects it can be given an electrical charge. If a source of direct electromotive force is connected to the plate a few electrons will flow into the plate. In the illustration above, the source of direct current is represented by a dry cell. The circles with a negative sign represent electrons.

2. If a grounded plate (B) is brought near plate (A), a greatly increased number of electrons will flow into plate (A).

3. The presence of the human hand or body near plate (A) has much the same effect as plate (B) and increases the number of electrons which flow to plate (A).

jects can be given an electrical charge. If a source of direct electromotive force is connected to the plate a FEW electrons will flow either to the plate or away from it.

If a second plate insulated from the first plate but connected to ground is brought close to the first plate there will be a great INCREASE in the number of electrons that flow into the first plate.

If, instead of the second plate, the hand or body is brought close to the first plate, it will have much the same effect as the second plate. There will be a great increase in the number of electrons that flow into the first plate. This effect of the hand or body is BODY EFFECT. In most instances it is undesirable. It brings about changes in capacitance in the wiring and components of a circuit and methods are employed to neutralize it. But body effect is put to use in a capacity-operated relay. The presence of a person's hand or body close to a metal plate actuates the relay. It is not necessary to touch the plate. A hand held close to the plate will trigger the relay.

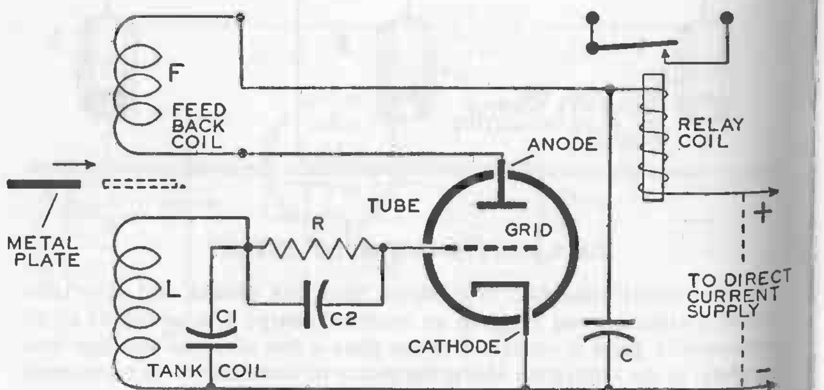

CIRCUIT OF CAPACITY RELAY USED TO STOP AND LEVEL ELEVATORS

When the metal plate is interposed between the feed-back coil and the tank coil, the relay closes.

YOU CAN BUILD A CAPACITY RELAY AND USE IT FOR INTERESTING EXPERIMENTS

You can build a capacity-operated relay with some common electronic components and an oscillator coil. You can perform some interesting and amusing demonstrations with it.

Fun with a Capacity Relay. The plate (called the capacity plate) attached to the relay may be a disk of metal foil about 6 inches in diameter glued to the inside of a show window or display case. A passer-by or customer who places a hand near the disk can set into operation an advertising display. Without actually touching anything they can turn electric lights on and off, ring a bell, stop and start a toy electric train, in fact do a great variety of things.

If the capacity plate is concealed in a book or magazine no one can pick up coins or other objects lying on the book without ringing a bell.

For a variation of the same idea place a piece of foil about 12x12 (connected to the oscillator as a capacity plate) on a dinner table underneath the tablecloth. Lay a silver spoon or a knife and fork on the tablecloth above the capacity plate. Connect an electric bell and a battery to the relay and hide them beneath the table. When the oscillator is adjusted, no one can touch the silver or pick it up without setting off the alarm bell. During dry weather the relay can be connected to a window screen so that anyone who gets close to the screen or touches it will set off an alarm such as a red light or a bell.

If the capacity plate is placed under a rug, the relay will be actuated when anyone steps on the rug over the plate. If the relay is connected to a bell, it will serve as an alarm. For amusement, the relay can be connected to a table lamp or a floor lamp and the light will flash on momentarily when a hand or foot is placed on the rug over the capacity plate. Or, the capacity plate can be placed under a slipcover on an upholstered chair located near a table lamp. If the relay is connected to the lamp, the latter will light when anyone sits in the chair and continue to burn as long as he remains.

A capacity plate larger than 6 inches x 6 inches can be used. The maximum size which is practical will have to be determined by experiment. If too large, the relay will not be sensitive.

COMPONENTS AND MATERIALS USED IN BUILDING THE CAPACITY-OPERATED RELAY

1 Shellacked wood base 6½ in. x 5 in. x ½ in. **(A)**
1 Piece plywood (shellacked) 2½ in. x 1⅜ in. x ¼ in. **(B)**
1 Piece plywood (shellacked) 2 in. x 1¼ in. x ¼ in.
For mounting relay **(D)**
1 Piece thin sheet metal 4 in. x 3 in. **(C)**
1 25A6 radio tube **(E)**
1 290-ohm line cord resistor **(F)**
1 Octal wafer socket **(G)**

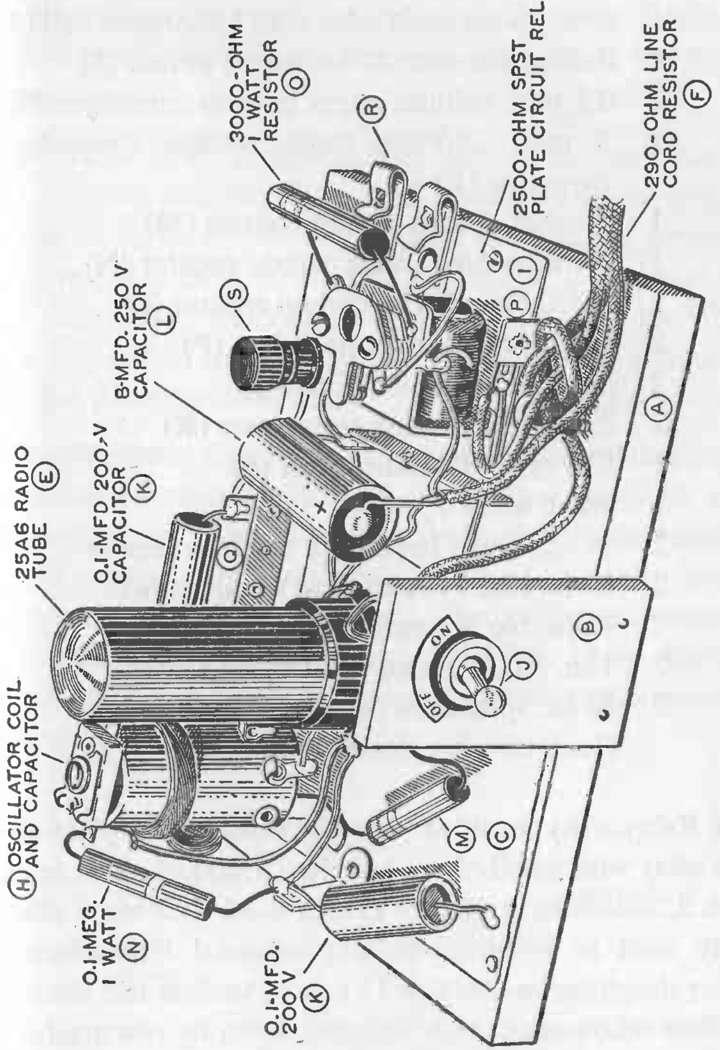
3000-OHM
1 WATT
RESISTOR
(O)
(R)
2500-OHM SPST
PLATE CIRCUIT RELAY
290-OHM LINE
CORD RESISTOR
(F)
8-MFD. 250 V
CAPACITOR
(L)
(S)
(P)
(I)
(A)
25A6 RADIO
TUBE
(E)
0.1-MFD 200-V
CAPACITOR
(K)
(O)
(B)
(J)
ON
OFF
(H) OSCILLATOR COIL
AND CAPACITOR
0.1-MEG.
1 WATT
(N)
(M)
(C)
0.1-MFD.
200 V
(K)

AN EXPERIMENTAL CAPACITY RELAY

1	Miller Cat. No. 695 oscillator coil for capacity relay (**H**)
1	SPDT plate circuit relay with 2,500-ohm coil (**1**)
1	Ball handle SPST-125-v toggle switch (**J**)
2	0.1 mfd. tubular paper by-pass capacitors (**K**)
1	8 mfd. 250-volt single section electrolytic capacitor (**L**)
1	50-ohm, 1 watt carbon resistor (**M**)
1	0.1 megohm, 1 watt carbon resistor (**N**)
1	3,000-ohm, I watt carbon resistor (**O**)
2	Single-lug Bakelite tie points (**P**)
2	5-lug Bakelite tie-points (**Q**)
2	Fahnestock spring connectors (**R**)
1	Eby type 30 binding post (**S**)
1	Small screw eye (**T**)
2	⅜-in. round head 6-32 machine screws
2	½-in. No. 4 round head wood screws
7	⅜-in. No. 4 round head wood screws
2	1-in. No. 6 round head wood screws
2	¾-in. spacers to raise socket above base
	Hook-up wire, rosin-core solder

The Relay. Any sensitive SPST electromagnetic plate circuit relay with a coil of not less than 1,000 ohms or more than 2,500 ohms resistance can be used. The same plate relay used in building the experimental Photoelectric Relay described in Chapter 12 can be used in this circuit.

Plate relays are usually fastened down by two machine screws which fit into threaded holes at the bottom. When the relay is mounted directly on the wood base it is necessary to drill two holes through the base and counterbore them from the underside so that the screw heads will not project below the surface. A more convenient method is to mount the relay on a small metal sheet or plywood sub-base of its own and fasten the sub-base to the larger base.

The Tube. 25A6 is an amplifier tube of the class known as a power pentode. It is used frequently in the output stage of AC-DC receivers. The tube has a metal envelope and requires an octal socket. The heater is designed for 25 volts and at that voltage consumes 0.3 amperes. When used alone and operated on 117 volts it must be connected to the power source with a 290-ohm line cord resistor.

The Oscillator Coil. This is an inexpensive factory-made coil wound on a plastic tube and comes complete with capacitor and radio-frequency choke coil. An instruction sheet comes with the coil showing how to use it in both a single-tube and two-tube circuit. The coil is manufactured by the J. W. Miller Company, 5917 S. Main St., Los Angeles, California. If your dealer cannot supply the coil, write direct to the manufacturer.

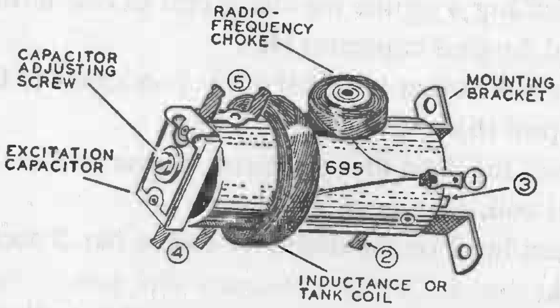

MILLER NO. 695 OSCILLATOR COIL

The Capacity Plate. This may be a square piece of WIRE screen or metal sheet about 6 inches x 6 inches or a disk of the same materials about 6 inches in diameter. Copper, brass, aluminum, iron, in fact any metal can be used as the capacity plate. A suitable plate can be made by covering a piece of cardboard with aluminum foil. Solder a flexible insulated wire to the plate. If foil is used, the wire cannot be attached by soldering, it must be clamped to the foil.

The plan of the oscillator is the best guide for assembling the components on the base.

The line cord resistor is tied to a screw eye in the forward right hand corner so that there will be no strain on the soldered connections. The Eby binding post near the back edge is the terminal to which the wire leading to the capacity plate is attached.

Use push-back wire for the connections. The ends of all wires except those which are clamped under the Fahnestock connectors and the binding post should be soldered. Use rosin-core solder. Make all wires as short and as direct as possible.

Make connections to the oscillator coil as follows:

Connect lug 4 on the oscillator coil to one terminal of a 0.1 mfd. by-pass capacitor (**K**).

Connect the other terminal of the condenser to the Eby binding post (**S**).

Connect lug 3 on the choke coil to one terminal of the relay (**H**) coil.

Connect lug 2 on the oscillator coil to No. 3 socket (**G**) terminal.

Connect lug 5 to the capacitor on the oscillator coil to No. 5 socket (**G**) terminal.

Operation. When the wiring has been completed check it carefully and if correct, the instrument is ready for a test. Connect a dry cell and an electric bell or a 1.5-v flashlight lamp to the Fahnestock connectors and push the armature of the plate relay so as to close the contacts. If the contacts are clean and touch each other the bell should ring or the lamp should light, as the case may be.

Place a 25A6 tube in the socket, push the plug on the resistor cord into a power outlet and turn the on-off switch to the ON position. Give the tube 3 or 4 minutes to warm up before making any adjustments. Connect the wire attached to the capacity plate to the binding post provided for it at the back of the base. Lay a newspaper or magazine

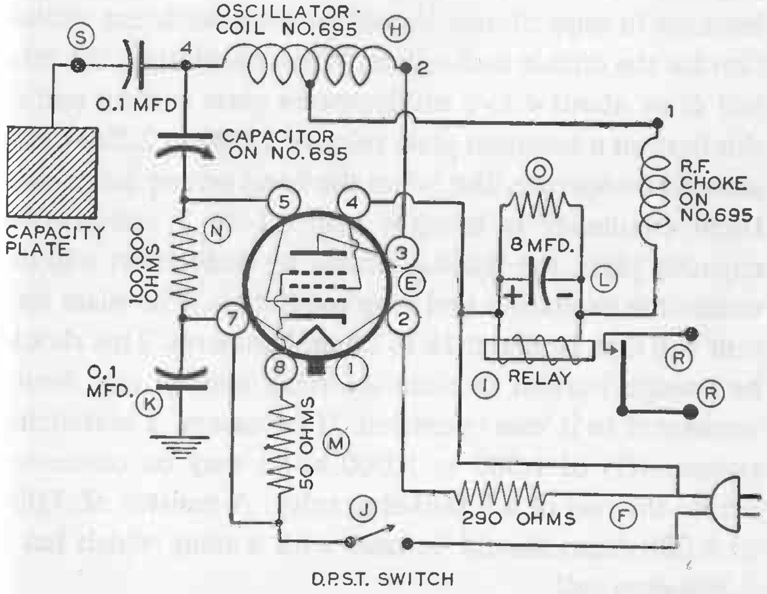

CIRCUIT FOR THE EXPERIMENTAL CAPACITY RELAY

The 290-ohm resistor is one of the conductors in the line cord resistor.

on the table and place the plate on this. The paper will insulate the plate from the table. Turn the adjustment screw on the excitation capacitor counterclockwise two or three revolutions. Use a screwdriver with a plastic handle and do not let your fingers touch the metal blade. The best screwdriver to use for this purpose is a nonmetallic screwdriver made of fiber and called an aligning tool. This tool is used by professionals for adjusting electronic equipment. Loosening the capacitor adjustment screw reduces the capacitance to a point where the circuit cannot oscillate and the bell should ring or the lamp light. If this occurs, turn the adjusting screw very slowly in a clockwise direction until the bell stops or the lamp is extinguished. Do not turn the screw beyond this point. The increase in capacitance should provide sufficient excitation for the circuit to oscillate. When oscillating, the tube will draw about 4 to 5 milliamperes plate current and at this current a common plate relay of 1,000 to 2,500 ohms should not operate. But when the hand or any large electrical conductor is brought near (1 to 3 inches) the capacity plate, the loading caused by body effect will decrease the excitation and stop oscillation. The plate current will then be about 18 to 25 milliamperes. This should be enough current to close the relay and set any device connected to it into operation. If necessary, a stabilizing resistor (**O**) of 1,500 to 5,000 ohms may be connected across the coil of a 1,000-ohm relay. A resistor of 3,000 to 5,000 ohms should be used with a relay which has a 2,500-ohm coil.

The relay and the excitation capacitor (on top of the coil) should not be adjusted until after the capacity plate

has been installed in position and connected. If the relay has a tendency to remain closed when it should open, it may be that the armature touches the end of the iron core in the relay coil. This can be remedied by slipping a piece of tissue paper between the core and the armature.

How to Build and Operate a Miniature Home Broadcasting Station

Radio signals of all sorts, telegraph code, speech, music and television, are sent out into space by radio waves appropriately called CARRIER waves. The carrier waves are produced by sending alternating currents into an antenna. The currents used for this purpose alternate at high frequency, hundreds of thousands, even millions of times per second, and are known as HIGH-FREQUENCY OR RADIO-FREQUENCY CURRENTS. They are generated by an electron tube.

A radio tube which will amplify can also be made to oscillate or generate sustained radio-frequency currents. It is merely necessary to connect the tube in a proper circuit and supply the filament and plate circuits with suffi-

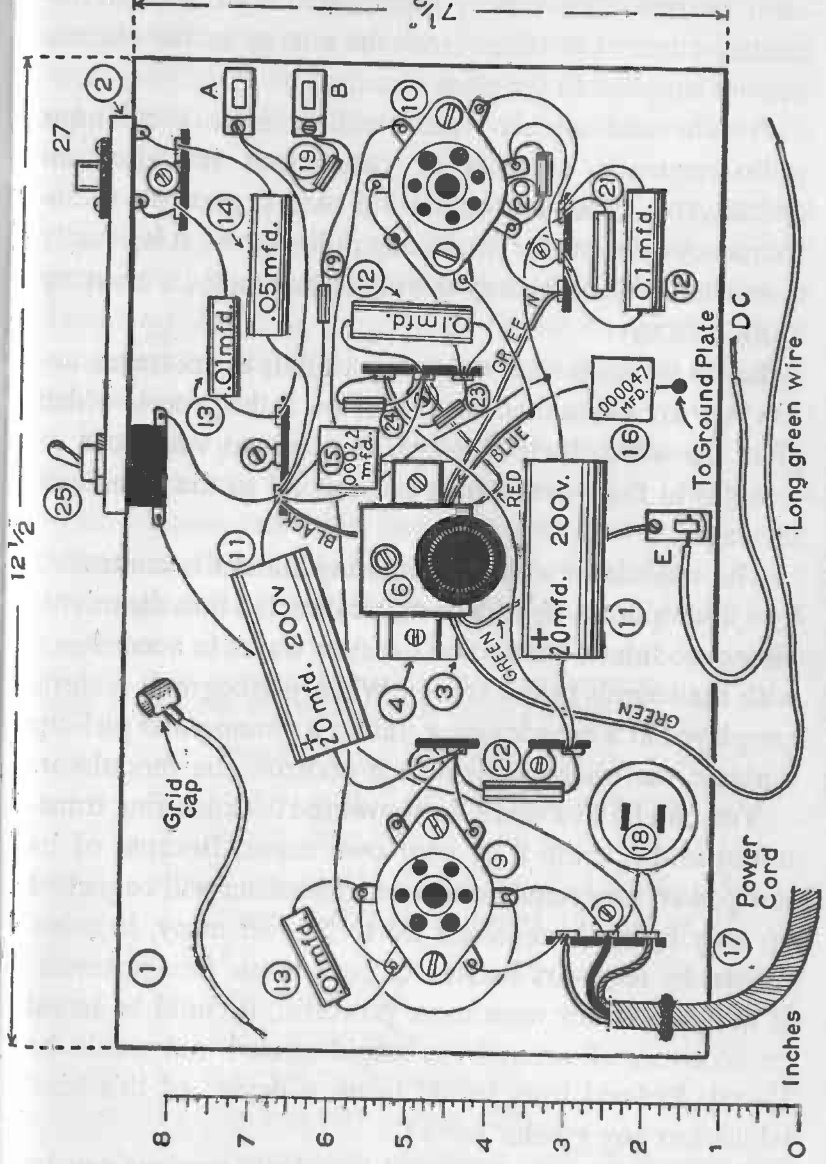
7 1/2"
12 1/2
27
A
B
(2)
(10)
(19)
(14)
.05 mfd.
(19)
(12)
.01 mfd.
0.1 mfd.
(20)
(21)
0.1 mfd.
(12)
GREEN
DC
To Ground Plate
Long green wire
.000047 MFD.
(16)
(23)
(24)
(13)
(15)
.00022 mfd.
BLUE
RED
200v.
(25)
BLACK
(11)
(6)
E
20 mfd.
(11)
20 mfd. 200v
(3)
(4)
GREEN
GREEN
(22)
Grid cap
(18)
(9)
Power Cord
(13)
.01 mfd.
(1)
(17)
Inches
0
1
2
3
4
5
6
7
8

PLAN OF THE PHONO-OSCILLATOR

cient current. The energy represented by the radio-frequency currents is taken from the energy in the electric current supplied to the plate circuit.

A tube and circuit which will generate continuous radio-frequency currents or oscillations is called an OSCILLATOR. When an oscillator is used to generate radio-frequency currents for producing radio waves, it is usually associated with an energy-storing circuit called a TANK or TANK CIRCUIT.

Radio waves do not carry any signals or messages unless they are modulated. MODULATION is the process which gives the waves their message. It produces variations or changes in the waves which correspond to the signals or messages.

The modulator at a broadcasting station is controlled by a microphone. Speech or music directed into the microphone modulates the station's carrier waves in accordance with the sounds in the studio. When phonograph records are played at a broadcasting station a phonograph pick-up replaces the microphone which controls the modulator.

You can build a small, low-powered broadcasting transmitter and operate it in your own home. Because of its low power, programs from your transmitter will be picked up only by radio receivers 20 to 50 feet away, in other words, by receivers located in your house or apartment. If the transmitter were more powerful, it could be heard on receivers all around the neighborhood and would be illegal. Federal laws forbid using a device of this kind which has any greater power.

Any first-class, factory-built broadcast receiver can be used to receive your programs. Records can be played in a room adjoining that where the receiver is located or in a room on another floor in the same building and their music will come out of the radio receiver. You can make announcements and your voice will be heard via the radio receiver. This will amaze those who do not know how it is accomplished.

The miniature broadcasting station is a device termed "wireless phonograph oscillator." It is actually a low-powered radio transmitter which sends out signals in the broadcast frequency band. The signals are picked up by a receiver nearby when the receiver is tuned to the oscillator's signals.

You can purchase a factory-built phonograph oscillator. You can build one yourself at much lower cost. With a phono oscillator, any broadcast receiver will perform as an electric phonograph without connecting it to a record player and without altering the receiver in any way.

PARTS AND MATERIALS REQUIRED TO BUILD THE WIRELESS PHONOGRAPH OSCILLATOR

1	Wood base, 12½ in. x 7½ in. x ¾ in. **(1)**
1	Wood strip, 5 in. x 2½ in. x ¼ in. **(2)**
1	Wood strip 2½ in. x ¾ in. x ¼ in. **(4)**
2	Wood blocks 1 in. x ¾ in. x ½ in. **(3)**
1	Piece of thin sheet metal 11½ in. x 6½ in. **(5)**
1	6F7 radio tube
1	25Z5 radio tube
1	6-prong tube socket **(9)**
1	7-prong tube socket **(10)**
1	Phonograph oscillator coil (Meissner No. 17-9373) **(6)**
2	20 Mfg. 200-volt electrolytic capacitors **(11)**

1	Female plug receptacle for 120 volts (**18**)
2	0.1 mfd., 200-volt tubular by-pass capacitors (**12**)
2	0.1 mfd., 400-volt tubular by-pass capacitors (**13**)
1	.05 mfd., 400-volt tubular by-pass capacitor (**14**)
1	.00022 mfd., mica capacitor (**15**)
1	.000047 mfd., mica capacitor (**16**)
1	290-ohm line cord resistor (**17**)
2	470,000-ohm, ½-watt resistors (**19**)
1	390,000-ohm, ½-watt resistor (**24**)
1	68,000-ohm, ½-watt resistor (**20**)
1	2,400-ohm, ½-watt resistor (**21**)
1	39,000-ohm, ½-watt resistor (**23**)
1	10,000-ohm, 1-watt resistor (**22**)
3	Binding posts or Fahnestock connectors
5	Solder-lug mounting strips with four lugs
1	Small single-pole, single-throw switch (**25**)
1	Phonograph jack (**27**)
4	Rubber head tacks
4	½-inch, No. 5 round head wood screws
13	⅜-inch, No. 5 round head wood screws
1	1½-inch, No. 6 round head wood screw
2	¼-inch, No. 2 round head wood screws
	Shellac, varnish, push-back wire and rosin-core solder

The Oscillator Coil. The "heart" of the oscillator is a No. 17-9373 Phono Oscillator Coil.* The coil can be ordered from the factory or from any dealer who sells

amateur and experimental radio equipment. The coil is actually three small coils and two small capacitors enclosed in a metal case or shield. One of the capacitors is a double trimmer capacitor which can be adjusted by means of a knob and a small screwdriver. Adjusting the trimmer capacitors tunes the oscillator. In other words, the adjustment changes the frequency and wavelength of the signals the oscillator sends forth.

Six wire terminal leads emerge from the bottom of the metal shield enclosing the coil. These are the terminals of the coils and colored so that they can be identified. There are four short solid wires, colored respectively black, blue, red, and green. The two other wires are both colored green, are stranded and are longer than the four short wires. One of the stranded green wires is about two feet long.

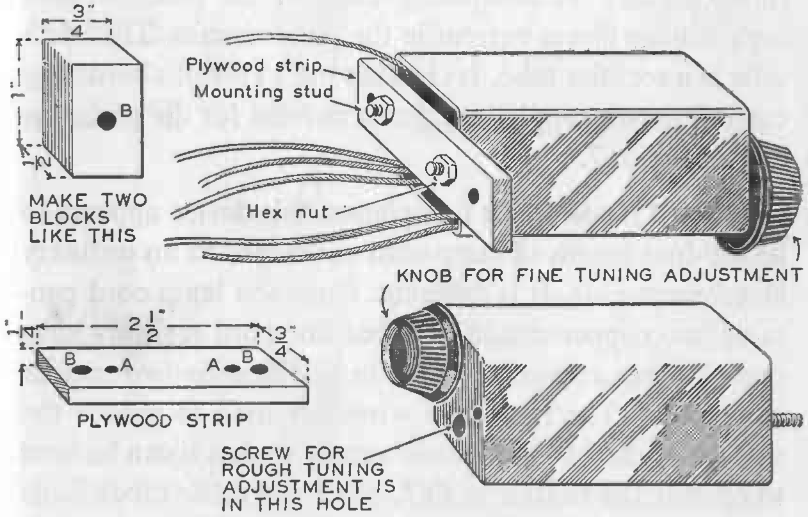
3"/4
1"
1/2"
B
MAKE TWO BLOCKS LIKE THIS
Plywood strip
Mounting stud
Hex nut
KNOB FOR FINE TUNING ADJUSTMENT
2 1/2"
3"/4
B
A
A
B
PLYWOOD STRIP
SCREW FOR ROUGH TUNING ADJUSTMENT IS IN THIS HOLE

MEISSNER NO. 17-9373 PHONO-OSCILLATOR COIL

---

* Manufactured by Thordarson-Meissner, Mt. Carmel, Illinois

The screw which adjusts the capacitor is visible through a small hole in the top of the shield and can be turned with a small screwdriver. Turning the screw gives a rough or coarse adjustment of the frequency and wavelength. A tuning knob projects through the top of the shield. Turning the knob gives a fine adjustment of the frequency and wavelength.

The Tubes. The phono oscillator employs two tubes: a 6F7 and a 25Z5. Both are the heater type and consequently can be operated on alternating current (the heater circuits only). A 6F7 tube is actually two tubes in one, a triode and a pentode enclosed in the same glass envelope. The 6F7 is the oscillator tube and generates the radio-frequency currents which produce the waves that carry the phonograph music through space from the phono oscillator to the radio receiver. Although the heater of the 6F7 tube operates on alternating current, the tube requires high voltage direct current in the plate circuits. The 25Z5 tube is a rectifier tube. It changes the 117-volt alternating current power supply into direct current for the plate circuit of the 6F7.

Line Cord Resistor. At first glance, this device appears to be a 6-foot length of lamp cord connected to an ordinary attachment plug. It is different. Common lamp cord contains two copper conductors but line cord resistors have one or more resistance wires in addition to two copper conductors. The resistance wires are used to reduce the voltage of the 117-volt power supply so that it can be used to operate the heaters in 6F7, 35Z5 and other tubes. Line cord resistors complete with attachment plug are available in several resistances ranging from 135 ohms to 960 ohms. The line cord resistor purchased for the phono oscillator should have a resistance of 290 ohms.

The Base. The first step in building the phono oscillator is to make a smooth wooden base 12 in. x 7 in. x ¾ in. and give all its surfaces a coat of white shellac varnish. When the shellac has dried, fasten a piece of thin sheet metal (11½ in. x 6½ in.) to the underside of the base. Use thin galvanized or tinned sheet iron. You can cut a suitable piece of tinned sheet iron out of the sides of a one-gallon can. Fasten the metal to the underside of the base with small brads. Do not try to drive the brads

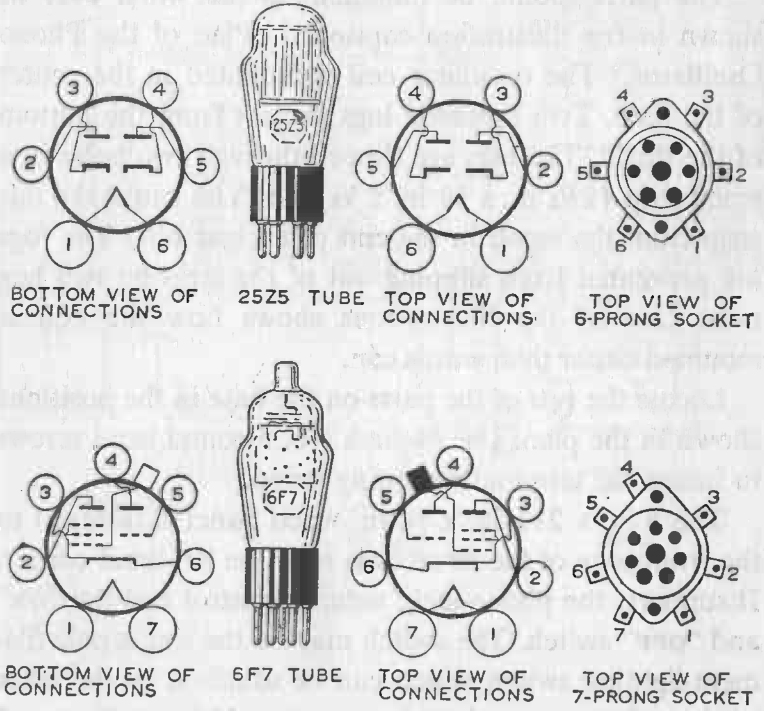
25Z5
BOTTOM VIEW OF CONNECTIONS
25Z5 TUBE
TOP VIEW OF CONNECTIONS
TOP VIEW OF 6-PRONG SOCKET
6F7
BOTTOM VIEW OF CONNECTIONS
6F7 TUBE
TOP VIEW OF CONNECTIONS
TOP VIEW OF 7-PRONG SOCKET
1
2
3
4
5
6
7

THE TUBES USED IN THE PHONO-OSCILLATOR

through the metal. Punch a small hole in the metal wherever you wish to place a brad. Make the holes with a nail. Hold the point of the nail against the metal, hit the nail head with a hammer and the hole is made. Use care not to drive the nail so hard that it makes holes larger than the heads of the brads you are using.

After fastening the metal sheet, drive a rubber headed tack into the underside of the base at each corner. They will insulate the metal sheet from any object upon which the phono oscillator rests. The metal plate must not be grounded because it is part of the antenna system from which the waves radiate.

The parts should be mounted on the wood base as shown in the illustration captioned "Plan of the Phono Oscillator." The oscillator coil is mounted in the center of the base. Two threaded lugs project from the bottom of the shield. The lugs are slipped through two holes in a wood strip (2½ in. x ¾ in. x ⅛ in.). You can make this strip from the wood in the end of a cigar box. The lugs are prevented from slipping out of the strip by two hex nuts. One of the illustrations shows how the coil is mounted better than words can.

Locate the rest of the parts on the base in the positions shown in the plan. Use ⅜-inch No. 5 round head screws to fasten the terminal mounting strips.

The 5 in. x 2½ in. x ¼ in. wood panel is fastened to the front edge of the wood base near the left hand corner. It supports the phono-jack, volume control and the "ON" and "OFF" switch. The switch may be the single-pole filament-lighting switch which can be attached to the back of the volume control or the common 120-volt "canopy" switch used on lamps and wall brackets.

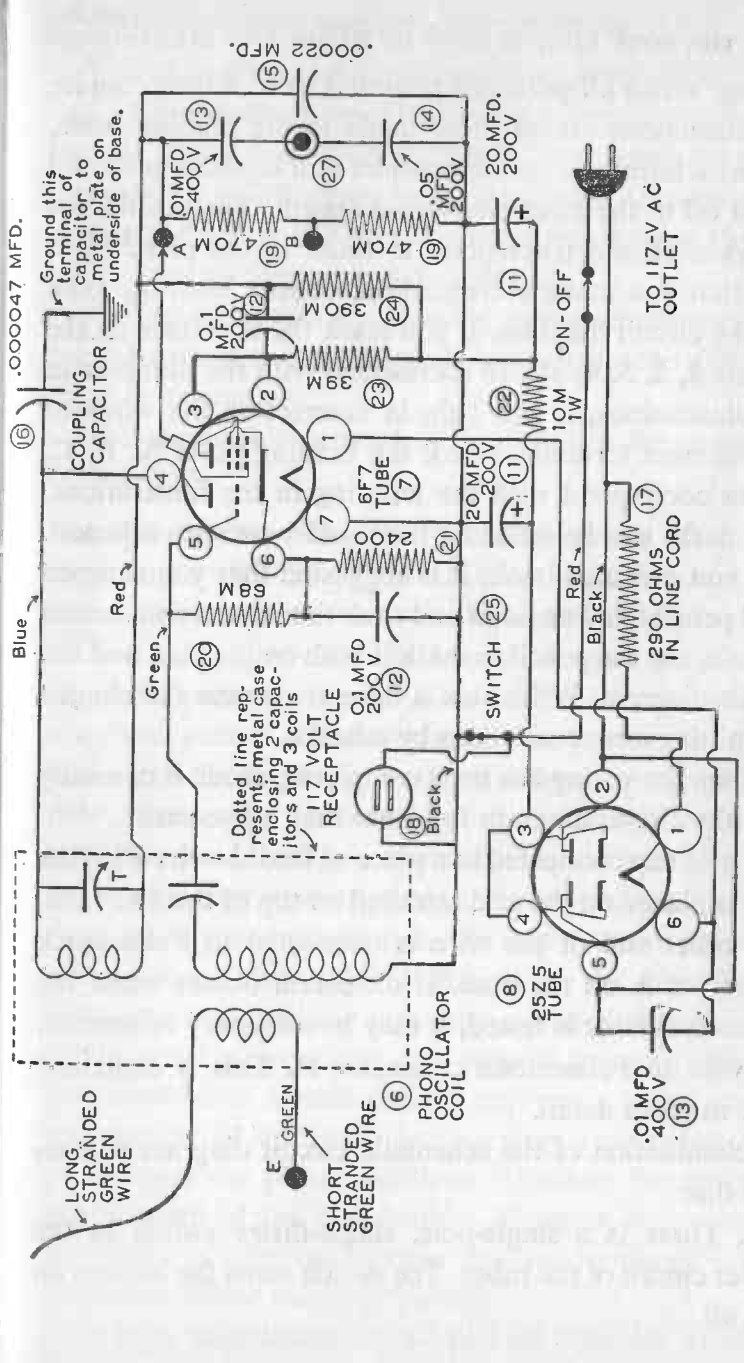

.00022 MFD.
⑮
⑬
⑭
㉗
.000047 MFD.
Ground this terminal of capacitor to metal plate on underside of base.
.01 MFD 400 V
A
470 M
⑲
B
470 M
⑲
.05 MFD 200 V
20 MFD. 200 V
+
⑯
COUPLING CAPACITOR
0.1 MFD 200 V
⑫
390 M
㉔
39 M
㉓
⑪
ON-OFF
TO 117-V AC OUTLET
4
3
2
1
6F7 TUBE
⑦
5
6
7
20 MFD 200 V
⑪
㉒
10 M 1W
Blue
Red
Green
68 M
⑳
2400
㉑
0.1 MFD 200 V
⑫
㉕
SWITCH
Red
Black
290 OHMS IN LINE CORD
⑰
Dotted line represents metal case enclosing 2 capacitors and 3 coils
117 VOLT RECEPTACLE
⑱
Black
3
2
1
4
5
6
25Z5 TUBE
⑧
PHONO OSCILLATOR COIL
⑥
.01 MFD 400 V
⑬
LONG STRANDED GREEN WIRE
E
GREEN
SHORT STRANDED GREEN WIRE

SCHEMATIC CIRCUIT DIAGRAM OF THE PHONO-OSCILLATOR

Wiring. When all parts are mounted, start wiring. Solder all connections except those made to the binding posts. The wire terminals on the resistors and capacitors should be cut off to the most convenient length. The position of the wires should correspond to those in the plan. Each time that you make a connection, consult both the plan and the circuit diagram. If you mark the terminals on the sockets **1, 2, 3, 4,** etc. to correspond with the numbers in the illustrations, it will help in connecting the wires to their correct terminal. Mark the binding posts **A, B, C,** etc. to correspond with the lettering in the illustrations. The marks can be made on the wood base with a pencil.

If you own this book, it is suggested that you sharpen a red pencil to a fine point and each time that a connection is made, use the pencil to mark it both on the plan and the circuit diagram. When this is done it reduces the chance of omitting any connections by mistake.

When the wiring has been completed, check it carefully with the circuit diagram to insure that it is correct.

A grid cap connected to a piece of flexible wire 6 inches long is placed on the grid terminal on top of the 6F7 tube. The other end of the wire is connected to Fahnestock connector **A** on the base. If distortion occurs when the phono oscillator is tested, it may be necessary to connect the wire to Fahnestock connector **B.** This is explained later in more detail.

Examination of the schematic circuit diagram will reveal that:

1. There is a single-pole, single-throw switch in the heater circuit of the tubes. The switch turns the heaters on and off.

2. The heaters of the two tubes are connected in series.

3. Current from the 117-volt A.C. circuit reaches the heaters (when the switch is closed) after passing through a resistance of 290 ohms in the line cord. This resistance prevents the full 117-volt current from flowing through the heaters and burning them out.

4. Two 470,000-ohm resistors provide a voltage dividing network to accommodate to some extent the difference there is in pickups.

The use of this network is discussed in the paragraph DISTORTION.

The Record Player for the Phono Oscillator. The record player used with the phono oscillator should be equipped with a high-impedance crystal pickup. The player described in Chapter 15 may be used. Crystal pickups and magnetic pickups may both have high impedance but a magnetic pickup will not usually give good results when connected to the phono oscillator unless a preamplifier is used. The energy produced by a magnetic pickup is less than that produced by a crystal pickup and the preamplifier is necessary to strengthen the feeble energy. The magnetic pickup is connected to the input terminals of the preamplifier. The output terminals of the amplifier are connected to the pickup terminals on the phono oscillator.

A crystal pickup does not require a preamplifier. A low impedance crystal pickup requires a small step up transformer of the audio or interstage type between the pickup and the phono oscillator. Connect the pickup to the primary of the transformer. Connect the secondary terminals of the transformer to the terminals on the phono oscillator marked "pickup."

A high impedance crystal pickup does not require a preamplifier or a transformer. The terminals of the pickup are connected to a phono-plug which is pushed into the phono-jack.

To Operate the Phono Oscillator. If a checkup of the wiring on the phono oscillator shows it to be correct, the oscillator is ready for a test. When making the first test, place the phono oscillator close to a radio receiver or close to the antenna lead-in connected to the receiver provided the lead-in is not shielded wire. The base of the phono oscillator should rest where the metal plate will not be grounded. Sometimes the phono oscillator will not operate unless it rests on a piece of window glass (7½ in. x 12½ in.).

Push the 6F7 and 25Z5 tubes into their sockets on the phono oscillator. Connect the flexible grid connection to the terminal on the top of the 6F7 tube and to the Fahnestock connector marked **A** on the base of the phono oscillator. Connect the pickup plug of a record player to the pickup jack. Push the line cord plug of the phono oscillator into a 117-volt outlet. Push the plug connected to the line cord on the record player into the 117-volt outlet on the oscillator. Turn the ON and OFF switch on the phono-oscillator to the ON position so that the tubes start to warm up. When the tubes have been warm for about one minute, start a record playing. If the record player is equipped with a volume control, turn the volume control to maximum.

Turn the volume control on the radio receiver to the approximate position used when receiving local broadcasting stations. Tune the receiver by moving the tuning control slowly through its entire range as if looking for a

distant broadcasting station. At some point on the dial, the phonograph record will be heard.

It is possible that a broadcasting station may be heard simultaneously with the record. If such interference occurs it is a simple matter to change the frequency of the oscillator so that it will be heard at a position on the dial of the radio receiver where no broadcasting is heard.

Turning the screw which can be seen through the hole in the top of the coil shield, and turning the tuning knob which projects from the shield will change the frequency of the oscillator. Turning the screw makes a much greater change in frequency than turning the knob. The knob is the fine adjustment.

Distortion. The phono oscillator can be overmodulated by a microphone or phonograph pickup which generates too great an output voltage. Overmodulation produces raspy distortion. If the reproduction of loud tones is of poor quality and sounds raspy, release the flexible grid connection on the 6F7 tube from Fahnestock connector **A** and connect it to **B.**

Trouble Shooting. Some radio receivers lack sufficient sensitivity to reproduce good signals from a phono oscillator. Weak reproduction will occur if a receiver is shielded from the oscillator by metal laths, pipes, etc., in walls and partitions. Either of these conditions may reduce the useful range of the phono oscillator to less than 20 to 50 feet.

The noise from static due to thunder storms, vacuum cleaners, motors, etc., can be greatly reduced and sometimes entirely eliminated by disconnecting the antenna from the radio receiver and connecting the long green stranded wire on the oscillator coil to the antenna post on the receiver in place of the antenna.

The terminal "E" to which a green wire from the oscillator coil is connected, is used to connect the phono-oscillator directly to a radio receiver. Disconnect the antenna from the "A" or "Ant" terminal on the receiver. Connect a wire from terminal "E" directly to the "A" or "Ant" terminal on the receiver. When the oscillator is used connected directly to the receiver, it is recommended that the .00005 mfd. coupling capacitor (16) be short-circuited by connecting a wire across its terminals. Shorting this condenser will reduce the likelihood of interference with your neighbor's radio receiver.

NOTE: Radio repair shops and service men rarely have a 6F7 or 25Z5 tube in stock. Both tubes are listed in the catalogs of Allied Radio and Lafayette Radio (*see page 269*) and may be ordered by mail from those firms.

Where to Obtain Materials

The average radio and television service shop can furnish any of the tubes, fixed capacitors, resistors, volume controls and batteries required to build the apparatus described in this book. Other components such as transistors, coils, relays, phototubes, transformers, phonomotors, etc., can be obtained from the large distributors of electronic equipment. Six reliable firms which issue mail order catalogs are listed below. They will send you a catalog if you write and request it. Neither the author nor the publisher has any interest in these business firms and does not assume any responsibilty in your dealings or transactions with them.

Write for catalog to:

Allied Radio, 100 N. Western Avenue, Chicago 80, Illinois.

Lafayette Radio, 165 Liberty Avenue, Jamaica 33, New York.

Burstein Applebee Co., 1012 McGee Street, Kansas City, Missouri.

Radio Shack Corporation, 167 Washington Street, Boston 8, Massachusetts.

Concord Radio Corporation, 45 Warren Street, New York 7, New York.

Newark Radio, 223 W. Madison Street, Chicago 6, Illinois and 4736 W. Century Blvd., Inglewood, California.

LENS FOR PHOTOTUBE RELAY

Various lenses are listed in the catalog of the Edmund Scientific Co., Barrington, New Jersey. If interested, write for catalog.

Index

Amber, 6
Amplifier, 14, 135, 203
  audio, 138
  cabinet for, 212
  cascade, 136
  chassis for, 204
  circuits, 207-209
  classification of, 140
  class A, 140
  class AB, 140
  class B, 140
  class C, 140
  experiment, 107
  how to connect, 106
  introduction to, 135
  one-stage transistor, 102
  power, 140
  radio-frequency, 139
  transistor, 100
  tubes, 157, 201
  two-stage audio, 199, 211
  video, 139
  voltage, 140
Annunciator, 169
Antenna, 9
  bedspring, 9
  insulators, 50
  loading coil, 30
  loop, 108, 112
  receiving, 48, 49
  television, 11
  vertical, 9
Array, antenna, 11
Armstrong, E. H., 46, 239, 241
Audio amplifier, 138
  circuit, 102, 103
  one-stage, 102
  two-stage, 198, 199

Baffles for loudspeaker, 86
Bardeen, Dr. John, 90
Base reflex cabinet, 87
Batteries, transistor, 97
Bedspring type antenna, 9
Bell, Alexander Graham, 145-147
Blabber stopper, 169
Body effect, 243
Bratten, Dr. Walter, 91
Bunsen, Robert W., 150
By-pass capacitors, 57, 58

Cabinet: amplifier, 212
  base reflex, 87
  geiger counter, 177
  loudspeaker, 80
  phonograph, 236
Capacitance, 40, 44, 52
  defined, 54
Capacitors: by-pass, 57, 58
  ceramic, 58
  effect of, 43
  electrolytic, 56, 59
  fixed, 53, 57

Capacitors—*Continued*
  mica, 23, 58
  padder, 63
  trimmer, 63
  variable, 22, 31, 61
Capacity operated relay, 243-247
  circuit, 244, 251
  experiments with, 245
Capacity plate, 250
Carbon microphone, 81
Carriers, 12
Cascade amplifier, 136
Cathode, 127
Ceramic capacitors, 58
Cesium, 150
Cetron phototube, 154
Chassis for amplifier, 204
Circuits, 8
  amplifier, 102, 103, 106, 209
  receiver, 16, 17, 31
  tank, 256
Code, continental, 115, 121
Code practice oscillator circuit, 118
Coils: antenna, 22
  loading, 30
  oscillator, 249, 258
  phono-oscillator, 259
  tuning, 28
  voice, 76
  winding, 31, 33
Colpitts oscillator, 242
Components, 52
Condenser, 52
Cone repairs, 237
Continental telegraph code, 115, 121
Controls, 52, 66, 159
Counter tube, 179
Coupling, 142
Crystal detector, 10-14
  circuits, 16, 17
Crystal diode, 15
  and transistor receiver, 108, 113, 114
Crystal microphones, 83
Crystal pickups, 84, 85
Crystal receiver, 10, 18, 21, 24-35
Crystal telephone receivers, 75
Crystals, piezo-electric, 84

De Forest, Dr. Lee, 14, 130
Demodulation, 14
Detection, 14
Detector: crystal, 10-17,
  triode, 131
Dielectrics, 52
  effects of, 55
Diodes, 19
  Fleming, 128
  Germanium, 22
Diode and transistor receiver, 109, 110
Distortion, 267
Dunwoody, Gen. H. H., 14

Edo fishscope, 6, 7
Electric eyes, 143
Electrodynamic microphone, 83
Electrolytic capacitors, 56, 59
Electromagnetic telephone receivers, 72
Electromagnetism, 189
Electrons, 8, 126, 152
Electron emission, 126
Electron tubes, 5
  as oscillators, 239
  as rectifiers, 138
Electronics, definition of, 4
Elevator control, 242

Emission, electron, 126
Enclosures: loudspeaker, 86
  phototube relay, 160
Energy conversions, 12
Energy, definition of, 8
Experiment with amplifier, 107

Farad, 54
Fishscope, 6, 7
Fixed capacitors, 32, 53
Fleming valve, 14, 128
Flux, magnetic, 72

Gamma snuffer, 175, 179, 180
  circuit, 189
Geiger counter, 171, 175, 191
  circuit, 193, 195
  how to build, 179
  testing, 194
Geiger-Muller counter tube, 174, 176, 179
Geigerscopes, 171, 173
Germanium, 17, 22
  diodes, 22
  semi-conductors, 15-22
Grid, 130
  suppressor, 134
Ground waves, 39

Half-wave rectifier tube, 158
Hartley oscillator, 241
Headset, 23
Heaters, 127
High-voltage for counter tubes, 186
Horizontal antenna, 9
Hush-a-tune, 233
  attachment, 237

Inductance, 29, 40, 44
  tuning, 33
Insulators, antenna, 50
Ionosphere, 41, 43
Keys, telegraph, 120
Kirchoff, Gustav R., 150

Lamp and bell for photo-relay, 168
Learning: radio telegraph code, 123
  to telegraph, 119
Light meters, 146
Lightning, 51
Light source, 161, 166
Line cord resistor, 260
Loading coil, 30
Loudspeakers, 75
  baffles for, 86
  cabinets for, 80, 86-88
  mechanism of, 77
  repairs and adjustments, 234

Magnetic flux, 72
Magnetic tape recorder, 216
Materials, where to obtain, 269
Memorizing groups, 122
Mica capacitor, 23, 58
Microfarad, 55
Micro-microfarad, 55
Microphones, 81-83
  carbon, 81
  crystal, 83
  electrodynamic, 83
  modulation, 256
Motor, phonograph, 217, 220
Multiplier phototube, 178
Multi-unit tubes, 134

National Research Council, 17
Natural frequency, 40

Operation of crystal receiver, 24
Oscillator, 239, 256
  code-practice, 115, 117
  coil, 249, 258

Oscillator—*Continued*
  Colpitts, 242
  Hartley, 241
  phonograph, 256
  simple, 240
Overmodulation, 267

Padder, capacitor, 63
Pentodes, 132
  construction of, 133
Phonograph: cabinet, 236
  cartridge, 215
  homemade, 235
  homemade portable, 229
  jack and plug, 206
  motor, 217, 220
  oscillator, 256
Phonograph Records, how made, 214
  played with table model radio, 225
Phono-motor, 220
Phono-oscillator, 255, 256
  coil, 259
  circuit, 263
Phosphor, 173
Photocells, 143
  first, 145
  selenium, 144
Photoelectric cell, selenium, 145
Photophone, 146
Photo-relay, 150, 151, 155-157
  circuit, 157
Phototube, 143, 148, 158
Phototube relay, 150-153
  cabinet, 160
Pickard, Greenleaf W., 14
Pickups, 84, 221, 265
  crystal, 84
  variable reluctance, 84
Piezo-electric crystals, 84
Plate, 128
  capacity, 250
  relay, 154
Portable phonograph, 229
Poulsen, Valdemar, 216
Power amplifier, 140
Preamplifier, 265
Projection lamp, 163
Prospecting, 197

Radar antenna, 11
Radioactivity, 171
Radio-frequency: amplifiers, 139
  currents, 12
Radio operator, the, 123
Radio receiver, building your first, 15
  (see Receiver)
Radio tubes, 125
  classification of, 125
  fundamental action, 129
  multi-unit, 134
  rectifier, 129
Radio waves, 9, 12, 39, 41, 43, 45
  from stars, 5
  skip zone, 45
Receiver: antenna for, 49
  building your first, 15
  diode and transistor, 108-110
  loop antenna, 9
  operation of, 32
  superheterodyne, 137
  telephone, 71
  tuning, 46
Receiving: antenna, 48
  tube bases and sockets, 139
Record player, 220
  how to build, 214

Rectifiers, 19
  tubes for, 158
Reflex speaker cabinet, 88
Relays: capacity-operated, 241
  phototube, 150-154
  plate, 154
Repairs to loudspeakers, 234
Repeaters, 135
Resistors, 52, 64
  color code, 65
  line cord, 260
Resonance, 42, 45
Rheostat, 68
Rotor, 62

Scintillations, 174
Scintillometers, 171, 178
Selenium, 144
  photocell, 145, 147
Semi-conductors, 18-22
Shockley, Dr. William, 91
Signal voltage, 95
Skip zone, 45
Sky waves, 39
Sonar, 6, 7
  transducer, 11
Spark gap, 184
Stacked antenna, 11
Static, 48, 51
Stator, 62
Strays, 51
Sunlight into electricity, 147
Superheterodyne: circuit, 46
  receiver, 137
Suppressor grid, 134
Switch, 69, 181
  attached to control, 67
  tuner, 123
Symbols, 92, 93

Tank circuit, 256
Taper, 66
Telescope, radio, 5
Telegraph: keys, 120
  learning to, 119
Telegraphone, 216
Telephone, receivers, 71
  crystal, 75
  electromagnetic, 72
  headset, 73
Television antenna, 11
Tenna-loop, 108
Tetrodes, 132
Timer switch, 123
Transducers, 70
  on record player, 85
  sonar, 11
Transformer, 76, 78, 187
  matching, 79
  output, 78
Transistors, 89-102
  action of, 94
  amplifier, 100-102
  batteries for, 97
  books about, 25
  characteristics of, 94
  circuits for, 102, 103, 106, 113, 114
  junction, 91
  kits, 97
  loop antenna for, 108, 112
  point contact, 91
  types of, 91
Transistor oscillator, 114-117
Transistor receiver, 108
Transmitter, 12
Trimmer capacitor, 63
Triode, 130
  as detector, 131
Tubes: amplifier, 157, 201
  detector, 131
  Geiger-Muller, 176, 179
Tuned radio-frequency amplifiers, 46

Tuning, 25, 42, 45
  coil, 31
  experiments in, 46
  inductance, 33
  taking mystery out of, 36
Two-stage amplifier, 198

Uranium prospecting, 171, 197

Variable capacitor, 23, 31, 61
Variable reluctance pickup, 84
Vertical antenna, 9
Vertical mast antenna, 9
Video amplifiers, 139
Voice coil, 76, 234
Voltage amplifier, 140

Waves: ground, 39
  radio, 39, 41, 45
  sky, 39
Whip antenna, 9
Winding coils, 31, 33
Winding machine, 33
Winkler, Clemens, 17

Understanding Electronics Components

author: Filipovic D. Miomir

This book is meant for those people who want to create electronic devices with their own hands. All components are illustrated and the circuit-symbol is explained in detail. Both simple and complex examples are provided for the beginners. These include resistors, capacitors, transformers, transistors, integrated circuits, etc and each has its own symbol to represent it in an electrical or electronic diagram - called a circuit diagram. In order to understand how a certain device functions, it is necessary to know each symbol and the characteristics of the component. These are the things we will be covering in this book.

KNJIGA 1
Komponente Elektronskih Uredaja

1. RESISTORS
1.1. Marking the resistors
1.2. Resistor power
1.3. Nonlinear resistors
1.4. Practical examples
1.5. Potentiometers
1.6. Practical examples

2. CAPACITORS
2.1. Block-capacitors
 2.1.1. Marking the clock-capacitors
2.2. Electrolytic capacitors
2.3. Variable capacitors
2.4. Practical examples

3. COILS AND TRANSFORMERS
3.1. Coils
3.2. Transformers
 3.2.1. Working principles and characteristics of transformers
3.3. Practical examples

4. TRANSISTORS
4.1. Working principles of transistors
4.2. Basic characteristics of transistors
4.3. The safest way to test transistors
4.4. TUN and TUP
4.5. Practical examples

5. DIODES
5.1. Marking the diodes
5.2. Characteristics of diodes
5.3. Practical examples

6. THYRISTORS, TRIAC, DIAC
6.1. Practical examples

7. INTEGRATED CIRCUITS
7.1. Analog integrated circuits
7.2. Digital integrated circuits
7.3. Practical examples

8. MICROPHONES, SPEAKERS, HEADPHONES
8.1. Microphones
8.2. Speakers
8.3. Headphones
8.4. Practical examples

9. OPTO-ELECTRONIC COMPONENTS
9.1. Practical examples

10. OTHER COMPONENTS
10.1. Relays
10.2. Practical examples

11. COMPONENTS CHECK
11.1. Diodes and transistors
11.2. Transformers and coils
11.3. Capacitors
11.4. Potentiometers
11.5. Speakers, headphones, microphones

12. CONDUCTIVITY PROBE
12.1. Semiconductors check
12.2. Other components check

1. Resistors

Resistors are the most commonly used component in electronics and their purpose is to create specified values of current and voltage in a circuit. A number of different resistors are shown in the photos. (The resistors are on millimeter paper, with 1cm spacing to give some idea of the dimensions). Photo 1.1a shows some low-power resistors, while photo 1.1b shows some higher-power resistors. Resistors with power dissipation below 5 watt (most commonly used types) are cylindrical in shape, with a wire protruding from each end for connecting to a circuit (photo 1.1-a). Resistors with power dissipation above 5 watt are shown below (photo 1.1-b).

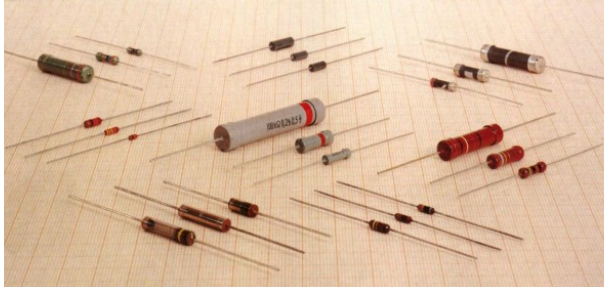

Fig. 1.1a: Some low-power resistors

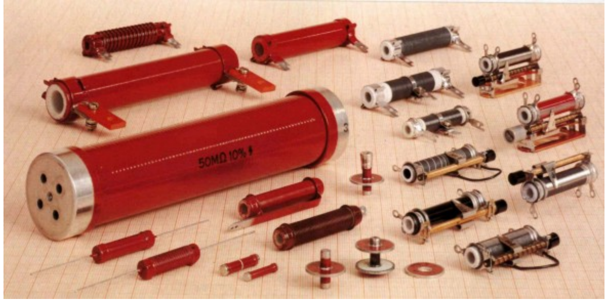

Fig. 1.1b: High-power resistors and rheostats

The symbol for a resistor is shown in the following diagram (upper: American symbol, lower: European symbol.)

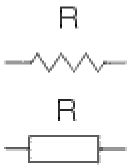

Fig. 1.2a: Resistor symbols

The unit for measuring resistance is the **OHM**. (the Greek letter Ω - called Omega). Higher resistance values are represented by "k" (kilo-ohms) and M (meg ohms). For example, 120 000 Ω is represented as 120k, while 1 200 000 Ω is represented as 1M2. The dot is generally omitted as it can easily be lost in the printing process. In some circuit diagrams, a value such as 8 or 120 represents a resistance in ohms. Another common practice is to use the letter E for resistance in ohms. The letter R can also be used. For example, 120E (120R) stands for 120 Ω, 1E2 stands for 1R2 etc.

1.1 Resistor Markings

Resistance value is marked on the resistor body. Most resistors have 4 bands. The first two bands provide the numbers for the resistance and the third band provides the number of zeros. The fourth band indicates the tolerance. Tolerance values of 5%, 2%, and 1% are most commonly available.

The following table shows the colors used to identify resistor values:

COLOR	DIGIT	MULTIPLIER	TOLERANCE	TC
Silver		x 0.01 Ω	±10%	
Gold		x 0.1 Ω	±5%	
Black	0	x 1 Ω		
Brown	1	x 10 Ω	±1%	±100*10^{-6}/K
Red	2	x 100 Ω	±2%	±50*10^{-6}/K
Orange	3	x 1 kΩ		±15*10^{-6}/K
Yellow	4	x 10 kΩ		±25*10^{-6}/K
Green	5	x 100 kΩ	±0.5%	
Blue	6	x 1 MΩ	±0.25%	±10*10^{-6}/K
Violet	7	x 10 MΩ	±0.1%	±5*10^{-6}/K
Grey	8	x 100 MΩ		
White	9	x 1 GΩ		±1*10^{-6}/K

** TC - Temp. Coefficient, only for SMD devices

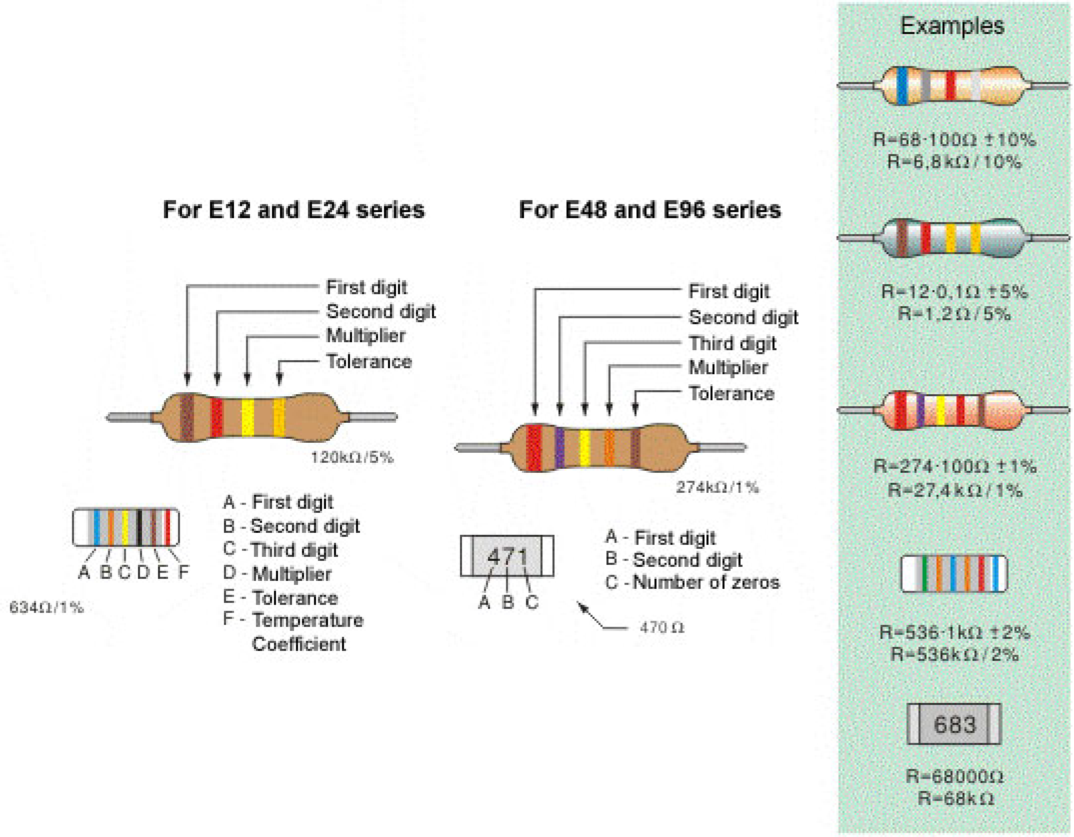
For E12 and E24 series
First digit
Second digit
Multiplier
Tolerance
120kΩ/5%
A B C D E F
634Ω/1%
A - First digit
B - Second digit
C - Third digit
D - Multiplier
E - Tolerance
F - Temperature Coefficient
For E48 and E96 series
First digit
Second digit
Third digit
Multiplier
Tolerance
274kΩ/1%
471
A B C
A - First digit
B - Second digit
C - Number of zeros
470 Ω
Examples
R=68·100Ω ±10%
R=6,8kΩ/10%
R=12·0,1Ω ±5%
R=1,2Ω/5%
R=274·100Ω ±1%
R=27,4kΩ/1%
R=536·1kΩ ±2%
R=536kΩ/2%
683
R=68000Ω
R=68kΩ

Fig. 1.2: b. Four-band resistor, c. Five-band resistor, d. Cylindrical SMD resistor, e. Flat SMD resistor

The following shows all resistors from 0R1 (one tenth of an ohm) to 22M:

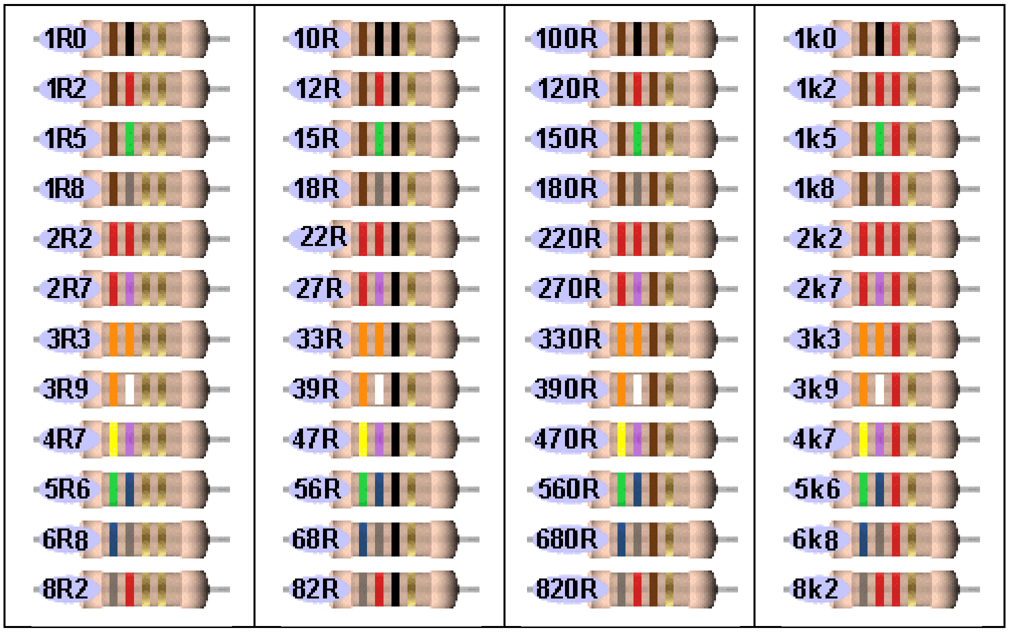
1R0
1R2
1R5
1R8
2R2
2R7
3R3
3R9
4R7
5R6
6R8
8R2
10R
12R
15R
18R
22R
27R
33R
39R
47R
56R
68R
82R
100R
120R
150R
180R
220R
270R
330R
390R
470R
560R
680R
820R
1k0
1k2
1k5
1k8
2k2
2k7
3k3
3k9
4k7
5k6
6k8
8k2

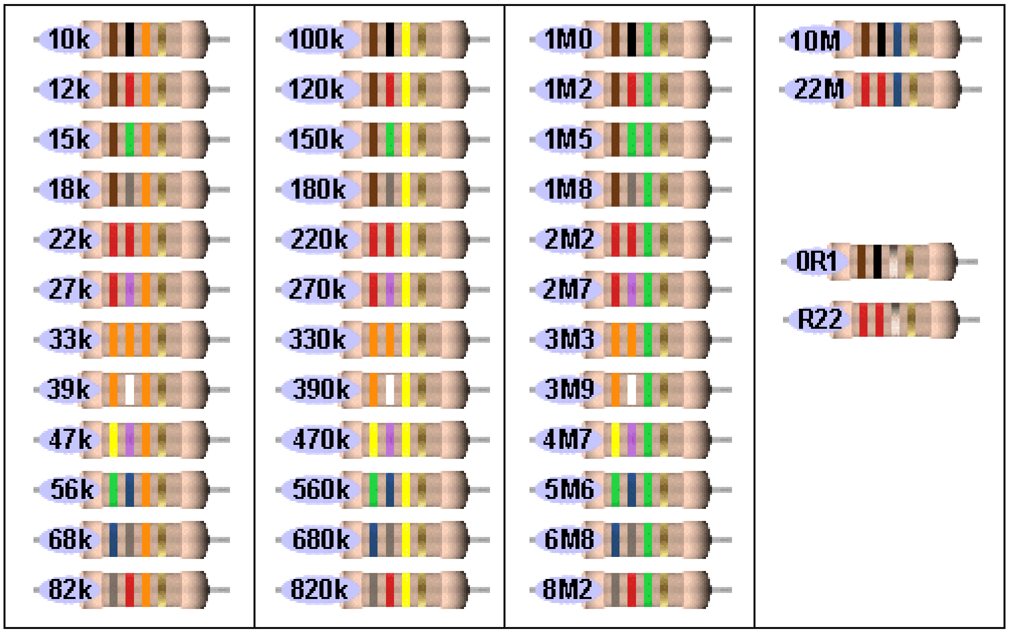
10k
12k
15k
18k
22k
27k
33k
39k
47k
56k
68k
82k
100k
120k
150k
180k
220k
270k
330k
390k
470k
560k
680k
820k
1M0
1M2
1M5
1M8
2M2
2M7
3M3
3M9
4M7
5M6
6M8
8M2
10M
22M
0R1
R22

NOTES:

The resistors above are "common value" 5% types.
The fourth band is called the "tolerance" band. Gold = 5%
(tolerance band Silver =10% but no modern resistors are 10%!!)
"common resistors" have values 10 ohms to 22M.

RESISTORS LESS THAN 10 OHMS

When the **third** band is gold, it indicates the value of the "colors" must be divided by 10.
Gold = "divide by 10" to get values 1R0 to 8R2
See 1st Column above for examples.

When the **third** band is silver, it indicates the value of the "colors" must be divided by 100.
(Remember: more letters in the word "silver" thus the divisor is "larger.")
Silver = "divide by 100" to get values 0R1 (one tenth of an ohm) to 0R82
e.g: 0R1 = 0.1 ohm   0R22 = point 22 ohms
See 4th Column above for examples.

The letters "R, k and M" take the place of a decimal point. The letter "E" is also used to indicate the word "ohm."
e.g: 1**R**0 = 1 ohm   2**R**2 = 2 point 2 ohms   22**R** = 22 ohms
2**k**2 = 2,200 ohms   100**k** = 100,000 ohms
2**M**2 = 2,200,000 ohms

Common resistors have 4 bands. These are shown above. First two bands indicate the first two digits of the resistance, third band is the multiplier (number of zeros that are to be added to the number derived from first two bands) and fourth represents the tolerance.

Marking the resistance with five bands is used for resistors with tolerance of 2%, 1% and other high-accuracy resistors. First three bands determine the first three digits, fourth is the multiplier and fifth represents the tolerance.

For SMD (Surface Mounted Device) the available space on the resistor is very small. 5% resistors use a 3 digit code, while 1% resistors use a 4 digit code.

Some SMD resistors are made in the shape of small cylinder while the most common type is flat. Cylindrical SMD resistors are marked with six bands - the first five are "read" as with common five-band resistors, while the sixth band determines the Temperature Coefficient (TC), which gives us a value of resistance change upon 1-degree temperature change.

The resistance of flat SMD resistors is marked with digits printed on their upper side. First two digits are the resistance value, while the third digit represents the number of zeros. For example, the printed number 683 stands for 68000Ω , that is 68k.

It is self-obvious that there is mass production of all types of resistors. Most commonly used are the resistors of the E12 series, and have a tolerance value of 5%. Common values for the first two digits are: 10, 12, 15, 18, 22, 27, 33, 39, 47, 56, 68 and 82.
The E24 series includes all the values above, as well as: 11, 13, 16, 20, 24, 30, 36, 43, 51, 62, 75 and 91. What do these numbers mean? It means that resistors with values for digits "39" are: 0.39Ω, 3.9Ω, 39Ω, 390Ω, 3.9kΩ, 39kΩ, etc are manufactured.
(0R39, 3R9, 39R, 390R, 3k9, 39k)

For some electrical circuits, the resistor tolerance is not important and it is not specified. In that case, resistors with 5% tolerance can be used. However, devices which require resistors to have a certain amount of accuracy, need a specified tolerance.

1.2 Resistor Dissipation

If the flow of current through a resistor increases, it heats up, and if the temperature exceeds a certain critical value, it can be damaged. The wattage rating of a resistor is the power it can dissipate over a long period of time.
Wattage rating is not identified on small resistors. The following diagrams show the size and wattage rating:

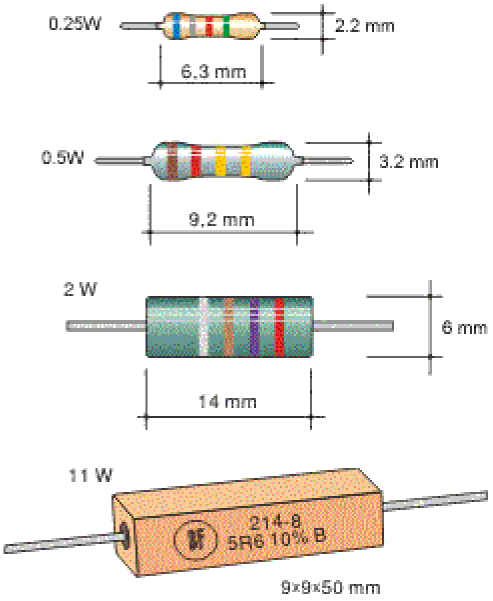

Fig. 1.3: Resistor dimensions

Most commonly used resistors in electronic circuits have a wattage rating of 1/2W or 1/4W. There are smaller resistors (1/8W and 1/16W) and higher (1W, 2W, 5W, etc).
In place of a single resistor with specified dissipation, another one with the same resistance and higher rating may be used, but its larger dimensions increase the space taken on a printed circuit board as well as the added cost.

Power (in watts) can be calculated according to one of the following formulae, where U is the symbol for Voltage across the resistor (and is in Volts), I is the symbol for Current in Amps and R is the resistance in ohms:

$$P = U \cdot I$$

$$P = R \cdot I^2$$

$$P = \frac{U^2}{R}$$

For example, if the voltage across an 820Ω resistor is 12V, the wattage dissipated by the resistors is:

$$P = \frac{U^2}{R} = \frac{12^2}{820} = 0{,}176\ \mathrm{W} = 176\,\mathrm{mW}$$

A 1/4W resistor can be used.

In many cases, it is not easy to determine the current or voltage across a resistor. In this case the wattage dissipated by the resistor is determined for the "worst" case. We should assume the highest possible voltage across a resistor, i.e. the full voltage of the power supply (battery, etc).
If we mark this voltage as V_B, the highest dissipation is:

$$P = \frac{{U_B}^2}{R}$$

For example, if V_B=9V, the dissipation of a 220Ω resistor is:

$$P = \frac{9^2}{220} = 368\,\mathrm{mW},$$

A 0.5W or higher wattage resistor should be used

1.3 Nonlinear resistors

Resistance values detailed above are a constant and do not change if the voltage or current-flow alters. But there are circuits that require resistors to change value with a change in temperate or light. This function may not be linear, hence the name **NONLINEAR RESISTORS.**

There are several types of nonlinear resistors, but the most commonly used include : *NTC* resistors (figure a) (Negative Temperature Co-efficient) - their resistance lowers with temperature rise. *PTC* resistors (figure b) (Positive Temperature Co-efficient) - their resistance increases with the temperature rise. *LDR* resistors (figure c) (Light Dependent Resistors) - their resistance lowers with the increase in light. *VDR* resistors (Voltage dependent Resistors) - their resistance critically lowers as the voltage exceeds a certain value. Symbols representing these resistors are shown below.

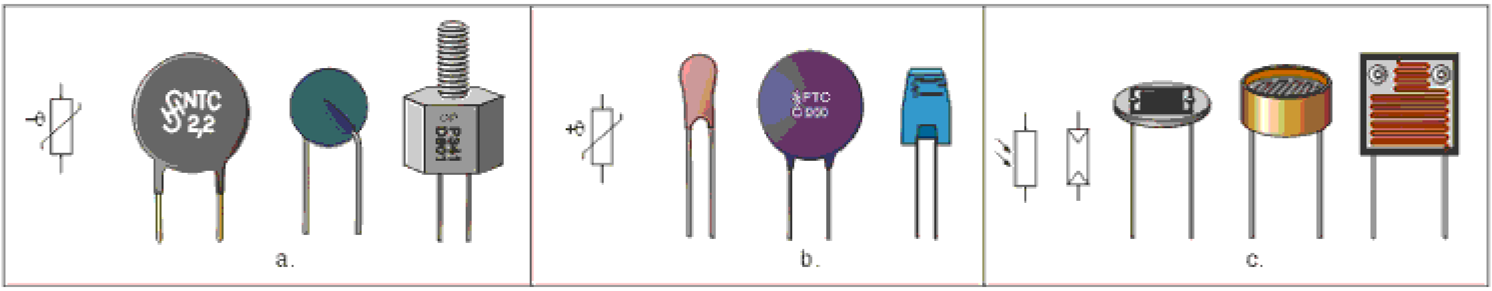

Fig. 1.4: Nonlinear resistors - a. NTC, b. PTC, c. LDR

In amateur conditions where nonlinear resistor may not be available, it can be replaced with other components. For example, *NTC* resistor may be replaced with a transistor with a trimmer potentiometer, for adjusting the required resistance value. Automobile light may play the role of *PTC* resistor, while *LDR* resistor could be replaced with an open transistor. As an example, figure on the right shows the 2N3055, with its upper part removed, so that light may fall upon the crystal inside.

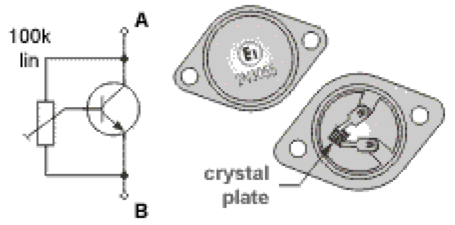

1.4 Practical examples with resistors

Figure 1.5 shows two practical examples with nonlinear and regular resistors as trimmer potentiometers, elements which will be covered in the following chapter.

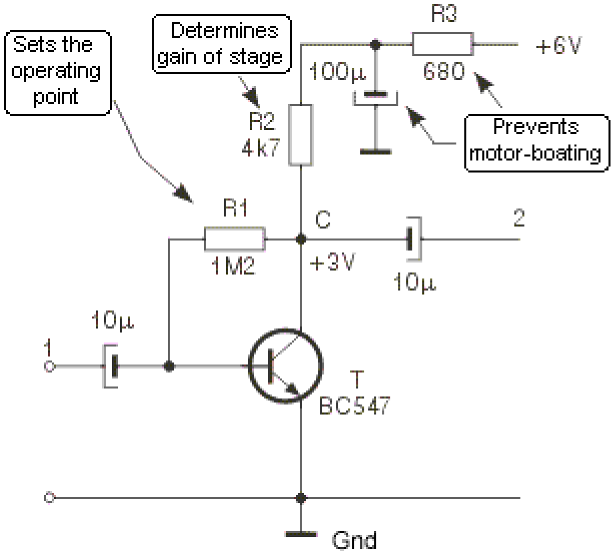

Fig. 1.5a: RC amplifier

Figure 1.5a represents an RC voltage amplifier, that can be used for amplifying low-frequency, low-amplitude audio signals, such as microphone signals. The signal to be amplified is brought between node 1 (amplifier input) and gnd, while the resulting amplified signal appears between node 2 (amplifier output) and gnd. To get the optimal performance (high amplification, low distortion, low noise, etc) , it is necessary to "set" the transistor's operating point. Details on the operating point will be provided in chapter 4; for now, let's just say that DC voltage between node C and gnd should be approximately one half of battery (power supply) voltage. Since battery voltage equals 6V, voltage in node C should be set to 3V. Adjustments are made via resistor R1.

Connect a voltmeter between node C and gnd. If voltage exceeds 3V, replace the resistor R1=1.2MΩ with a smaller resistor, say R1=1MΩ. If voltage still exceeds 3V, keep lowering the resistance until it reaches approximately 3V. If the voltage at node C is originally lower than 3V, increase the resistance of R1.

The degree of amplification of the stage depends on R2 resistance: *higher resistance* - *higher amplification*, *lower resistance* - *lower amplification*. If the value of R2 is changed, the voltage at node C should be checked and adjusted (via R1).

Resistor R3 and 100µF capacitor form a filter to prevent feedback from occurring. This feedback is called "Motor-boating" as it sounds like the noise from a motor-boat. This noise is only produced when more than one stage is employed.
As more stages are added to a circuit, the chance of feedback, in the form of instability or motor-boating, will occur.
This noise appears at the output of the amplifier, even when no signal is being delivered to the amplifier.
The instability is produced in the following manner:
Even though no signal is being delivered to the input, the output stage produces a very small background noise called "hiss. This comes from current flowing through the transistors and other components.
This puts a very small waveform on the power rails. This waveform is passed to the input of the first transistor and thus we have produced a loop for "noise-generation." The speed with which the signal can pass around the circuit determines the frequency of the instability. By adding a resistor and electrolytic to each stage, a low-frequency filter is produced and this "kills" or reduces the amplitude of the offending signal. The value of R3 can be increased if needed.

Practical examples with resistors will be covered in the following chapters as almost all circuits require resistors.

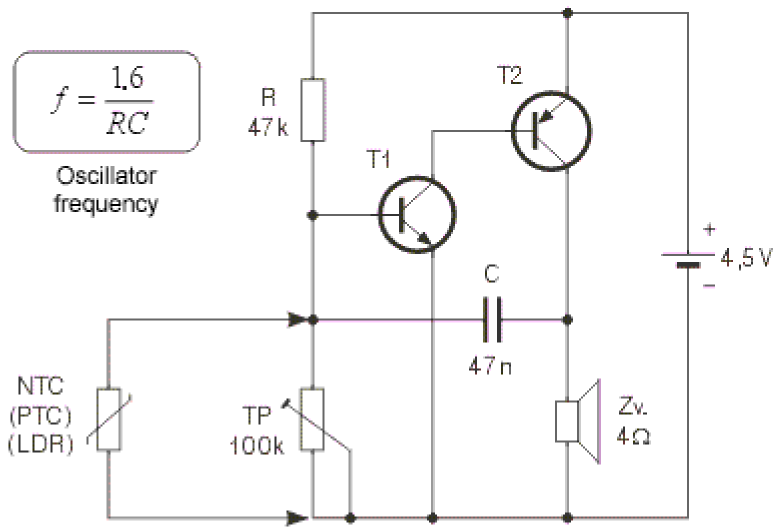

Fig. 1.5b: Sound indicator of changes in temperature or the amount of light

A practical use for nonlinear resistors is illustrated on a simple alarm device shown in figure 1.5b. Without trimmer TP and nonlinear NTC resistor it is an audio oscillator. Frequency of the sound can be calculated according to the following formula:

$$f = \frac{1.6}{RC}.$$

In our case, R=47kΩ and C=47nF, and the frequency equals:

$$f = \frac{1.6}{47 \cdot 10^3 \cdot 47 \cdot 10^{-9}} \approx 724\ Hz.$$

When, according to the figure, trim pot and NTC resistor are added, oscillator frequency increases. If the trim pot is set to minimum resistance, the oscillator stops. At the desired temperature, the resistance of the trim pot should be increased until the oscillator starts working again. For example, if these settings were made at 2°C, the oscillator remains frozen at higher temperatures, as the NTC resistor's resistance is lower than nominal. If the temperature falls the resistance increases and at 2°C the oscillator is activated.

If an NTC resistor is installed in a car, close to the road surface, the oscillator can warn driver if the road is covered with ice. Naturally, the resistor and two copper wires connecting it to the circuit should be protected from dirt and water.

If, instead of an NTC resistor, a PTC resistor is used, the oscillator will be activated when the temperature rises above a certain designated value. For example, a PTC resistor could be used for indicating the state of a refrigerator: set the oscillator to work at temperatures above 6°C via trimmer TP, and the circuit will signal if anything is wrong with the fridge.

Instead of an NTC, we could use an LDR resistor - the oscillator would be blocked as long as a certain amount of light is present. In this way, we could make a simple alarm system for rooms where a light must be always on.

The LDR can be coupled with resistor R. In that case, the oscillator works when the light is present, otherwise it is blocked. This could be an interesting alarm clock for huntsmen and fishermen who would like to get up at the crack of dawn, but only if the weather is clear. For the desired moment in the early morning, the trim pot should be set to the uppermost position. Then, the resistance should be carefully reduced, until the oscillator starts. During the night the oscillator will be blocked, since there is no light and LDR resistance is very high. As the amount of light increases in the morning, the resistance of the LDR drops and the oscillator is activated when the LDR is illuminated with the required amount of light.

The trim pot from the figure 1.5b is used for fine adjustments. Thus, TP from figure 1.5b can be used for setting the oscillator to activate under different conditions (higher or lower temperature or amount of light).

1.5 Potentiometers

Potentiometers (also called *pots*) are variable resistors, used as voltage or current regulators in electronic circuits. By means of construction, they can be divided into 2 groups: coated and wire-wound.

With coated potentiometers, (figure 1.6a), insulator body is coated with a resistive material. There is a conductive slider moving across the resistive layer, increasing the resistance between slider and one end of pot, while decreasing the resistance between slider and the other end of pot.

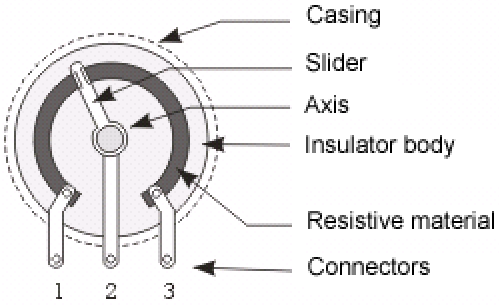

Fig. 1.6a: Coated potentiometer

Wire-wound potentiometers are made of conductor wire coiled around insulator body. There is a slider moving across the wire, increasing the resistance between slider and one end of pot, while decreasing the resistance between slider and the other end of pot.

Coated pots are much more common. With these, resistance can be linear, logarithmic, inverse-logarithmic or other, depending upon the angle or position of the slider. Most common are linear and logarithmic potentiometers, and the most common applications are radio-receivers, audio amplifiers, and similar devices where pots are used for adjusting the volume, tone, balance, etc.

Wire-wound potentiometers are used in devices which require more accuracy in control. They feature higher dissipation than coated pots, and are therefore in high current circuits.

Potentiometer resistance is commonly of E6 series, including the values: 1, 2.2 and 4.7. Standard tolerance values include 30%, 20%, 10% (and 5% for wire-wound pots).

Potentiometers come in many different shapes and sizes, with wattage ranging from 1/4W (coated pots for volume control in amps, etc) to tens of watts (for regulating high currents). Several different pots are shown in the photo 1.6b, along with the symbol for a potentiometer.

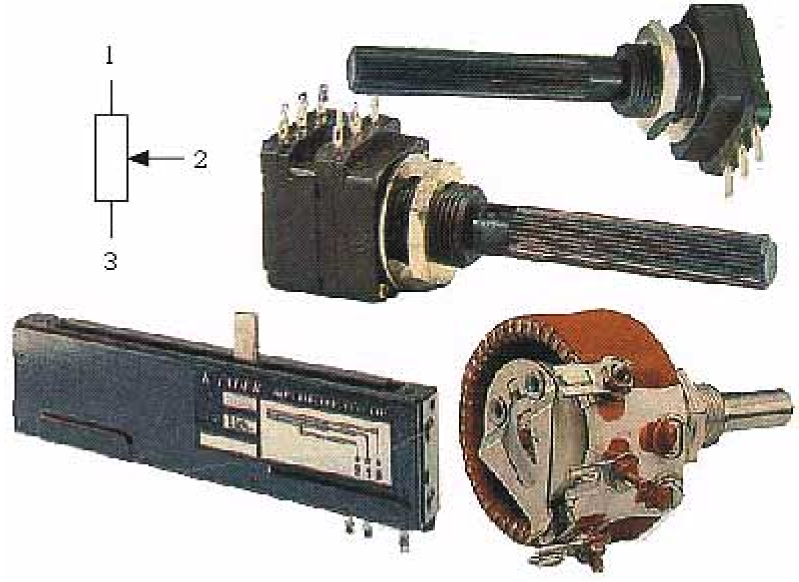

Fig. 1.6b: Potentiometers

The upper model represents a stereo potentiometer. These are actually two pots in one casing, with sliders mounted on shared axis, so they move simultaneously. These are used in stereophonic amps for simultaneous regulation of both left and right channels, etc.

Lower left is the so called slider potentiometer.

Lower right is a wire-wound pot with a wattage of 20W, commonly used as rheostat (for regulating current while charging a battery etc).

For circuits that demand very accurate voltage and current values, *trimmer potentiometers* (or just *trim pots*) are used. These are small potentiometers with a slider that is adjusted via a screwdriver.

Trim pots also come in many different shapes and sizes, with wattage ranging from 0.1W to 0.5W. Image 1.7 shows several different trim pots, along with the symbol.

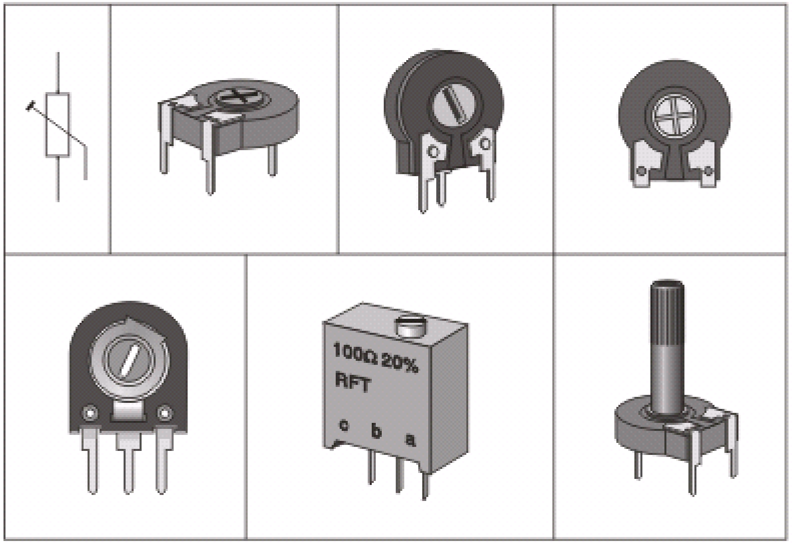

Fig. 1.7: Trim pots

Resistance adjustments are made via a screwdriver. Exception is the trim pot on the lower right, which can be adjusted via a plastic shaft. Particularly fine adjusting can be achieved with the trim pot in the plastic rectangular casing (lower middle). Its slider is moved via a screw, so that several full turns is required to move the slider from one end to the other.

1.6 Practical examples with potentiometers

As previously stated, potentiometers are most commonly used in amps, radio and TV receivers, cassette players and similar devices. They are used for adjusting volume, tone, balance, etc.

As an example, we will analyze the common circuit for tone regulation in an audio amp. It contains two pots and is shown in the figure 1.8a.

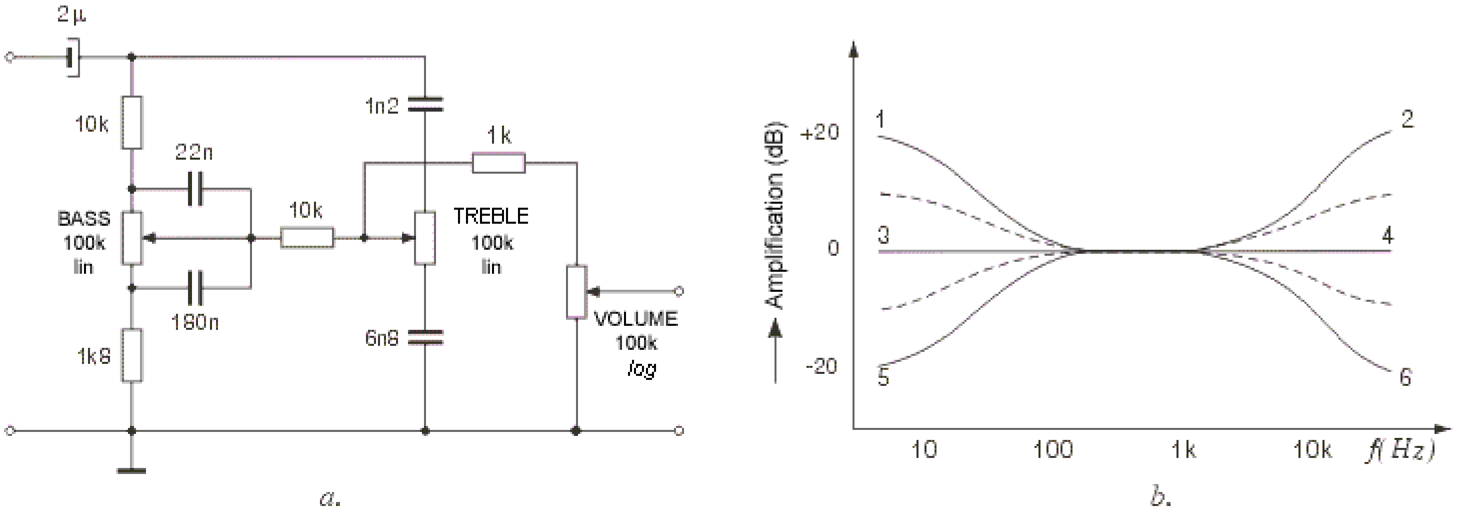

Fig. 1.8 Tone regulation circuit: a. Electrical scheme, b. Function of amplification

Potentiometer marked BASS regulates low frequency amplification. When the slider is in the lowest position, amplification of very low frequency signals (tens of Hz) is about ten times greater than the amplification of mid frequency signals (~kHz). If slider is in the uppermost position, amplification of very low frequency signals is about ten times lower than the amplification of mid frequency signals. Low frequency boost is useful when listening to music with a beat (disco, jazz, R&B...), while Low Frequency amplification should be reduced when listening to speech or classical music.

Similarly, potentiometer marked TREBLE regulates high frequency amplification. High frequency boost is useful when music consists of high-pitched tones such as chimes, while for example High Frequency amplification should be reduced when listening to an old record to reduce the background noise.

Diagram 1.8b shows the function of amplification depending upon the signal frequency. If both sliders are in their uppermost position, the result is shown with curve 1-2. If both are in mid position function is described with line 3-4, and with both sliders in the lowest position, the result is shown with curve 5-6. Setting the pair of sliders to any other possible results in curves between curves 1-2 and 5-6.

Potentiometers BASS and TREBLE are coated by construction and linear by resistance.

The third pot in the diagram is a volume control. It is coated and logarithmic by resistance (hence the mark *log*)

2. Capacitors

Capacitors are common components of electronic circuits, used almost as frequently as resistors. The basic difference between the two is the fact that capacitor resistance (called reactance) depends on the frequency of the signal passing through the item. The symbol for reactance is X_c and it can be calculated using the following formula:

$$X_C = \frac{1}{2\pi f C}$$

f representing the frequency in Hz and *C* representing the capacitance in Farads.

For example, 5nF-capacitor's reactance at *f*=125kHz equals:

$$X_C = \frac{1}{2 \cdot 3{,}14 \cdot 125000 \cdot 5 \cdot 10^{-9}} = 255\ \Omega,$$

while, at *f*=1.25MHz, it equals:

$$X_C = \frac{1}{2 \cdot 3{,}14 \cdot 1250000 \cdot 5 \cdot 10^{-9}} = 25{,}5\ \Omega \cdot$$

A capacitor has an infinitely high reactance for direct current, because *f*=0.

Capacitors are used in circuits for many different purposes. They are common components of filters, oscillators, power supplies, amplifiers, etc.

The basic characteristic of a capacitor is its capacity - the higher the capacity, the higher is the amount of electricity it can hold. Capacity is measured in Farads (F). As one Farad represents fairly high capacity, smaller values such as microfarad (µF), nanofarad (nF) and picofarad (pF) are commonly used. As a reminder, relations between units are:

$$1F=10^{6}\mu F=10^{9}nF=10^{12}pF,$$

that is 1µF=1000nF and 1nF=1000pF. It is essential to remember this notation, as same values may be marked differently in some circuits. For example, 1500pF is the same as 1.5nF, 100nF is 0.1µF.
A simpler notation system is used as with resistors. If the mark on the capacitor is 120 the value is 120pF, 1n2 stands for 1.2nF, n22 stands for 0.22nF, while .1µ (or .1u) stands for 0.1µF.

Capacitors come in various shapes and sizes, depending on their capacity, working voltage, type of insulation, temperature coefficient and other factors. All capacitors can divided in two groups: those with changeable capacity values and those with fixed capacity values. These will covered in the following chapters.

2.1 Block-capacitors

Capacitors with fixed values (the so called *block-capacitors*) consist of two thin metal plates (these are called "electrodes" or sometimes called the "foil"), separated by a thin insulating material such as plastic. The most commonly used material for the "plates" is aluminum, while the common materials used for insulator include paper, ceramic, mica, etc after which the capacitors get named. A number of different block-capacitors are shown in the photo below. A symbol for a capacitor is in the upper right corner of the image.

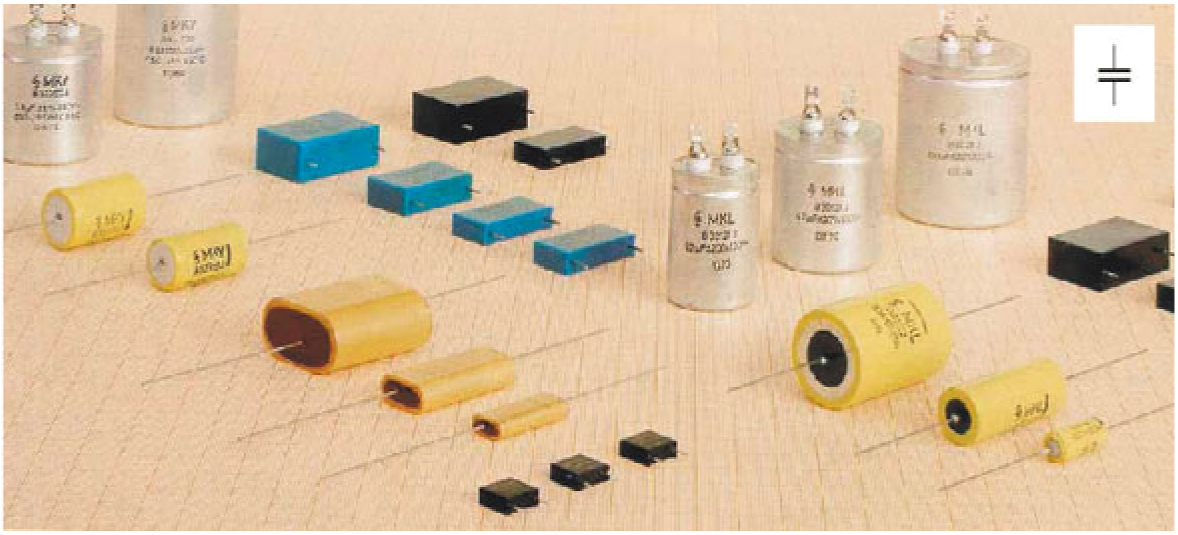

Fig. 2.1: Block capacitors

Most of the capacitors, block-capacitors included, are non-polarized components, meaning that their leads are equivalent in respect of the way the capacitor can be placed in a circuit. Electrolytic capacitors represent the exception as their polarity is important. This will be covered in the following chapters.

2.1.1 Marking the block-capacitors

Commonly, capacitors are marked by a set of numbers representing the capacity. Beside this value is another number representing the maximal working voltage, and sometimes tolerance, temperature coefficient and some other values are printed as well. But on the smallest capacitors (such as surface-mount) there are no markings at all and you must not remove them from their protective strips until they are needed. The size of a capacitor is never an indication of its value as the dielectric and the number of layers or "plates" can vary from manufacturer to manufacturer. The value of a capacitor on a circuit diagram, marked as 4n7/40V, means the capacitor is 4,700pF and its maximal working voltage is 40v. Any other 4n7 capacitor with higher maximal working voltage can be used, but they are larger and more expensive.

Sometimes, capacitors are identified with colors, similar to the 4-band system used for resistors (figure 2.2). The first two colors (A and B) represent the first two digits, third color (C) is the multiplier, fourth color (D) is the tolerance, and the fifth color (E) is the working voltage.

With disk-ceramic capacitors (figure 2.2b) and tubular capacitors (figure 2.2c) working voltage is not specified, because these are used in circuits with low DC voltage. If a tubular capacitor has five color bands on it, the first color represents the temperature coefficient, while the other four specify the capacity in the previously described way.

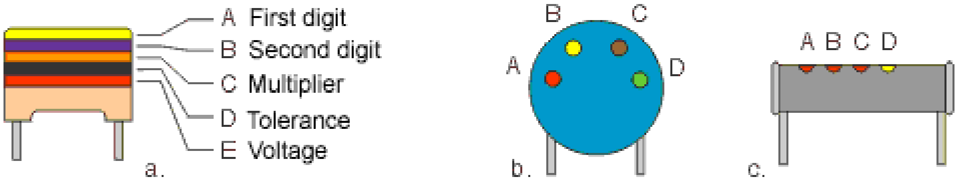

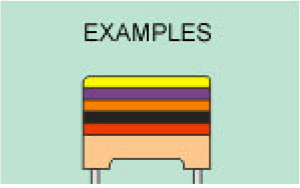

COLOR	DIGIT	MULTIPLIER	TOLERANCE	VOLTAGE
Black	0	x 1 pF	±20%	
Brown	1	x 10 pF	±1%	
Red	2	x 100 pF	±2%	250V
Orange	3	x 1 nF	±2.5%	
Yellow	4	x 10 nF		400V
Green	5	x 100 nF	±5%	
Blue	6	x 1 µF		
Violet	7	x 10 µF		
Grey	8	x 100 µF		
White	9	x 1000 µF	±10%	

Fig. 2.2: Marking the capacity using colors

The figure 2.3 shows how the capacity of miniature tantalum electrolytic capacitors are marked by colors. The first two colors represent the first two digits and have the same values as with resistors. The third color represents the multiplier, to get the capacity expressed in µF. The fourth color represents the maximal working voltage.

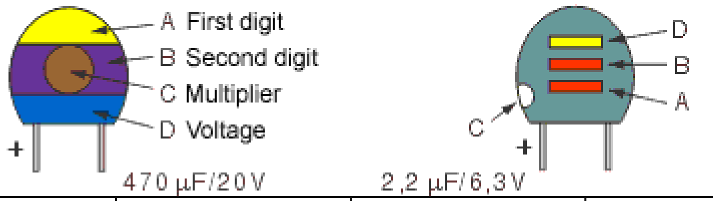

COLOR	DIGIT	MULTIPLIER	VOLTAGE
Black	0	x 1 µF	10V
Brown	1	x 10 µF	
Red	2	x 100 µF	
Orange	3		
Yellow	4		6.3V
Green	5		16V
Blue	6		20V
Violet	7		
Grey	8	x .01 µF	25V
White	9	x .1 µF	3V
Pink			35V

Fig. 2.3: Marking the tantalum electrolytic capacitors

One important note on the working voltage: The voltage across a capacitor must not exceed the maximal working voltage as the capacitor may get destroyed. In the case when the voltage is unknown, the "worst" case should be considered. There is the possibility that, due to malfunction of some other component, the voltage on capacitor equals the power supply voltage. If, for example, the supply is 12V, the maximal working voltage for the capacitor should be higher than 12V.

2.1 Electrolytic capacitors

Electrolytic capacitors represent the special type of capacitors with fixed capacity value. Thanks to special construction, they can have exceptionally high capacity, ranging from one to several thousand µF. They are most frequently used in circuits for filtering, however they also have other purposes.

Electrolytic capacitors are polarized components, meaning they have positive and negative leads, which is very important when connecting it to a circuit. The positive lead or pin has to be connected to the point with a higher positive voltage than the negative lead. If it is connected in reverse the insulating layer inside the capacitor will be "dissolved" and the capacitor will be permanently damaged.

Explosion may also occur if capacitor is connected to voltage that exceeds its working voltage. In order to prevent such instances, one of the capacitor's connectors is very clearly marked with a + or -, while the working voltage is printed on the case.

Several models of electrolytic capacitors, as well as their symbols, are shown on the picture below.

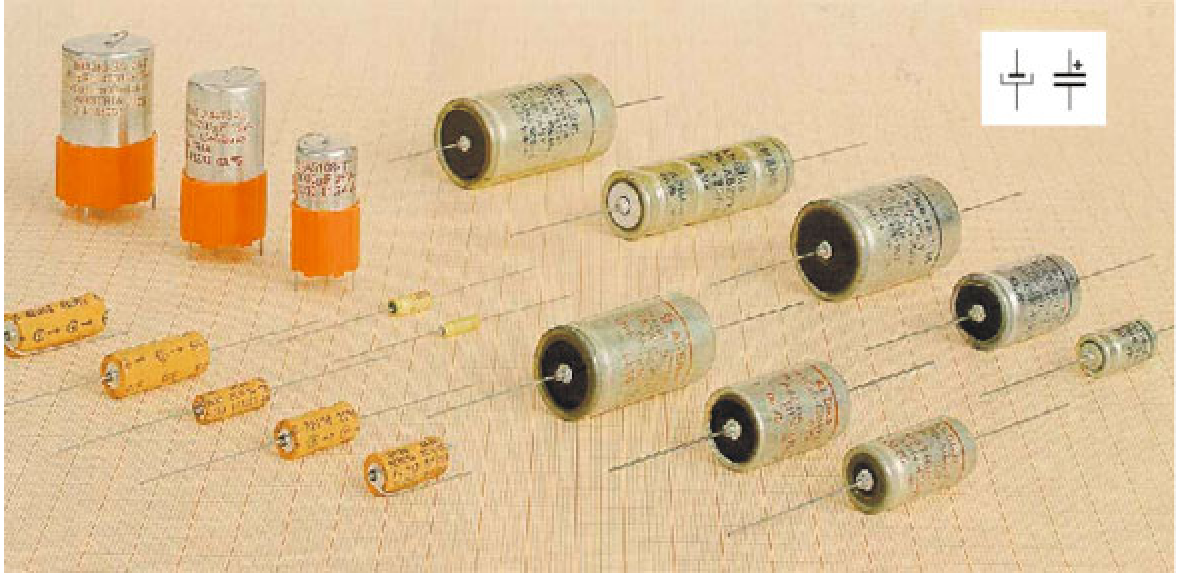

Fig. 2.4: Electrolytic capacitors

Tantalum capacitors represent a special type of electrolytic capacitor. Their parasitic inductance is much lower than standard aluminum electrolytic capacitors so that tantalum capacitors with significantly (even ten times) lower capacity can completely substitute an aluminum electrolytic capacitor.

2.3 Variable capacitors

Variable capacitors are capacitors with variable capacity. Their minimal capacity ranges from 1p and their maximum capacity goes as high as few hundred pF (500pF max). Variable capacitors are manufactured in various shapes and sizes, but common features for them is a set of fixed plates (called the stator) and a set of movable plates. These plates are fitted into each other and can be taken into and out of mesh by rotating a shaft. The insulator (dielectric) between the plates is air or a thin layer of plastic, hence the name variable capacitor. When adjusting these capacitors, it is important that the plates do not touch.

Below are photos of air-dielectric capacitors as well as mylar-insulated variable capacitors (2.5a).

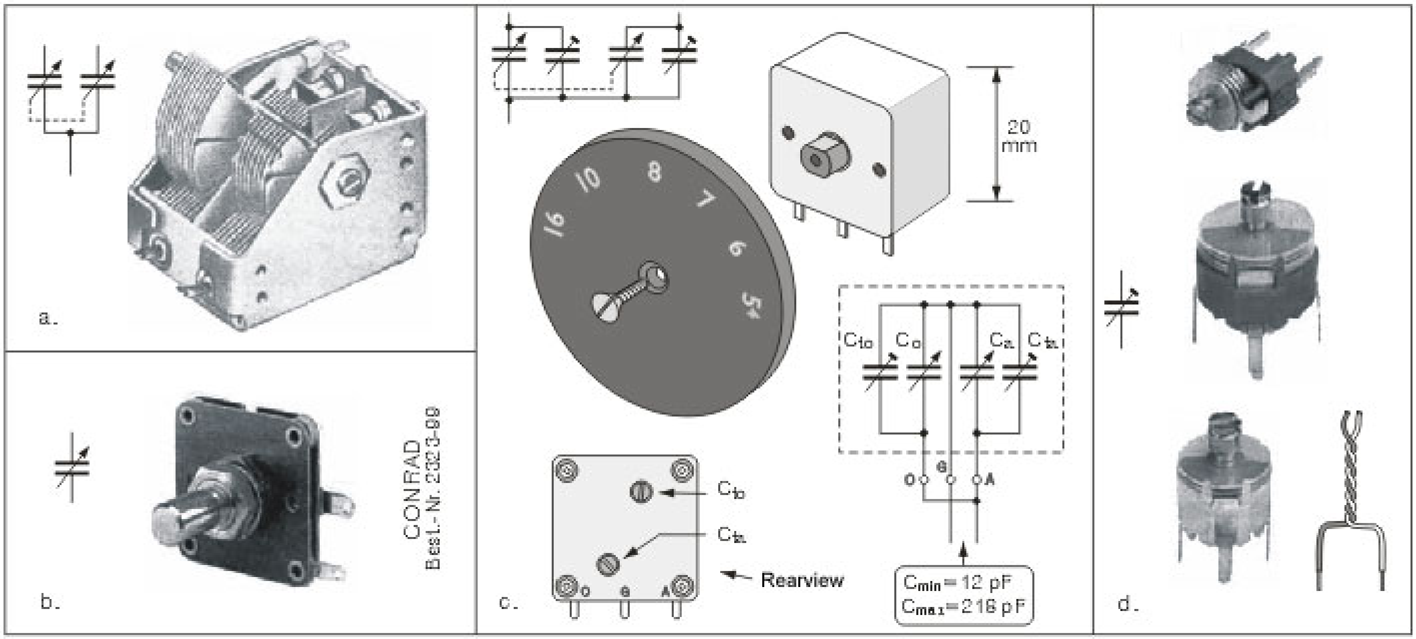

Fig. 2.5: a, b, c. Variable capacitors, d. Trimmer capacitors

The first photo shows a "ganged capacitor" in which two capacitors are rotated at the same time. This type of capacitor is used in radio receivers. The larger is used for the tuning circuit, and the smaller one in the local oscillator. The symbol for these capacitors is also shown in the photo.

Beside capacitors with air dielectric, there are also variable capacitors with solid insulator. With these, thin insulating material such as mylar occupies the space between stator and rotor. These capacitors are much more resistant to mechanical damage. They are shown in figure 2.5b.

The most common devices containing variable capacitors are radio receivers, where these are used for frequency adjustment. Semi-variable or trim capacitors are miniature capacitors, with capacity ranging from several pF to several tens of pFs. These are used for fine tuning radio receivers, radio transmitters, oscillators, etc. Three trimmers, along with their symbol, are shown on the figure 2.5d.

2.4 Practical examples

Several practical examples using capacitors are shown in figure 2.6. A 5µF electrolytic capacitor is used for DC blocking. It allows the signal to pass from one sage to the next while prevent the DC on one stage from being passed to the next stage. This occurs because the capacitor acts like a resistor of very low resistance for the signals and as a resistor of high resistance for DC.

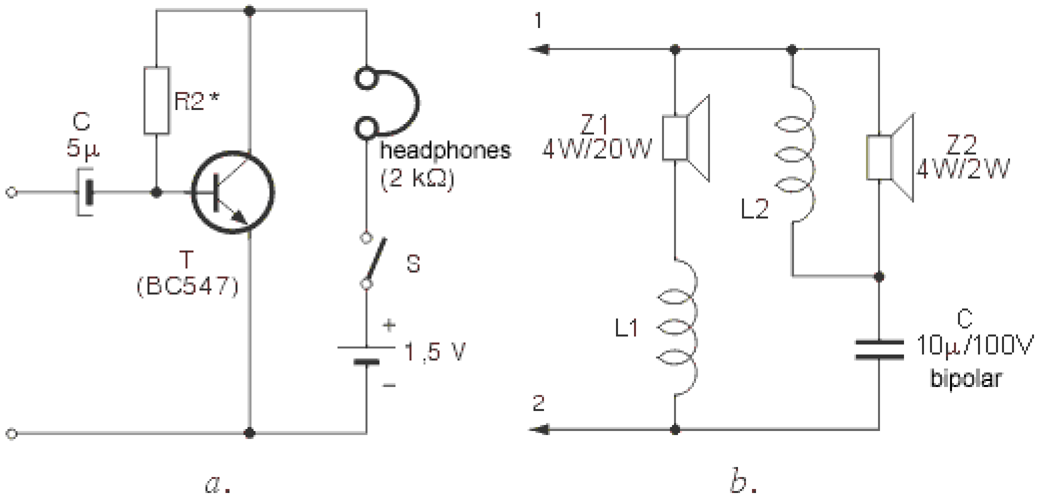
C
5μ
R2*
T
(BC547)
headphones
(2 kΩ)
S
+
1,5 V
−
a.
1
Z1
4W/20W
L2
Z2
4W/2W
L1
C
10μ/100V
bipolar
2
b.

Fig. 2.6: a. Amplifier with headphones, b. Electrical band-switch

The figure 2.6b represents a diagram of a band-switch with two speakers, with Z1 used for reproducing low and mid-frequency signals, and Z2 for high frequency signals. 1 and 2 are connected to the audio amplifier output. Coils L1 and L2 and the capacitor C ensure that low and mid-frequency currents flow to the speaker Z1, while high frequency currents flow to Z2. How this works exactly ? In the case of a high frequency current, it can flow through either Z1 and L1 or Z2 and C. Since the frequency is high, impedance (resistance) of the coils are high, while the capacitor's reactance is low. It is clear that in this case, current will flow through Z2. In similar fashion, in case of low-frequency signals, current will flow through Z1, due to high capacitor reactance and low coil impedance.

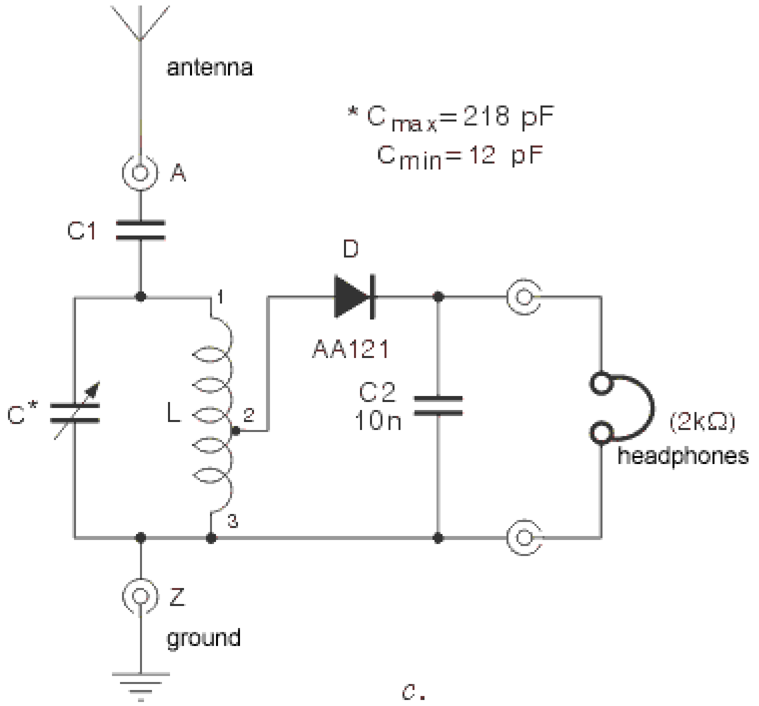
antenna
* Cmax=218 pF
Cmin=12 pF
A
C1
D
1
AA121
C*
L
2
C2
10n
(2kΩ)
headphones
3
Z
ground
c.

Fig. 2.6: c. Detector radio-receiver

The figure 2.6c represents a circuit diagram for a simple detector radio-receiver (commonly called a "crystal set"), where the variable capacitor C, forming the oscillatory circuit with the coil L, is used for frequency tuning. Turning the capacitor's rotor changes the resonating frequency of the circuit, and when matching a certain radio frequency, the station can be heard.

3. Coils and transformers

3.1 Coils

Coils are not a very common component in electronic circuits, however when they are used, they need to be understood. They are encountered in oscillators, radio-receivers, transmitter and similar devices containing oscillatory circuits. In amateur devices, coils can be made by winding one or more layers of insulated copper wire onto a former such as PVC, cardboard, etc. Factory-made coils come in different shapes and sizes, but the common feature for all is an insulated body with turns of copper wire.

The basic characteristic of every coil is its inductance. Inductance is measured in Henry (H), but more common are millihenry (mH) and microhenry (μH) as one Henry is quite a high inductance value. As a reminder:

$$1\text{H} = 1000\text{mH} = 10^6\ \mu\text{H}.$$

Coil inductance is marked by X_L, and can be calculated using the following formula:

$$X_L = 2\pi f L,$$

where f represents the frequency of the voltage in Hz and the L represents the coil inductance in H.

For example, if f equals 684 kHz, while L=0.6 mH, coil impedance will be:

$$X_L = 2 \cdot 3{,}14 \cdot 684000 \cdot 0{,}6 \cdot 10^{-3} = 2\,577\ \Omega.$$

The same coil would have three times higher impedance at three times higher frequency. As can be seen from the formula above, coil impedance is in direct proportion to frequency, so that coils, as well as capacitors, are used in circuits for filtering at specified frequencies. Note that coil impedance equals zero for DC (f=0).

Several coils are shown on the figures 3.1, 3.2, 3.3, and 3.4.

The simplest coil is a single-layer air core coil. It is made on a cylindrical insulator (PVC, cardboard, etc.), as shown in figure 3.1. In the figure 3.1a, turns have space left between them, while the common practice is to wind the wire with no space between turns. To prevent the coil unwinding, the ends should be put through small holes as shown in the figure.

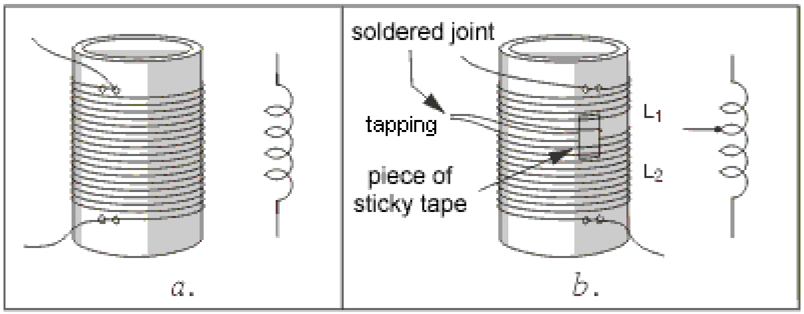

Fig. 3.1: Single-layer coil

Figure 3.1b shows how the coil is made. If the coil needs 120 turns with a tapping on the thirtieth turn, there are two coils L1 with 30 turns and L2 with 90 turns. When the end of the first and the beginning of the second coil are soldered, we get a "tapping."

A multilayered coil is shown in figure 3.2a. The inside of the plastic former has a screw-thread, so that the ferromagnetic core in the shape of a small screw can be inserted. Screwing the core moves it along the axis and into the center of the coil to increase the inductance. In this manner, fine changes to the inductance can be made.

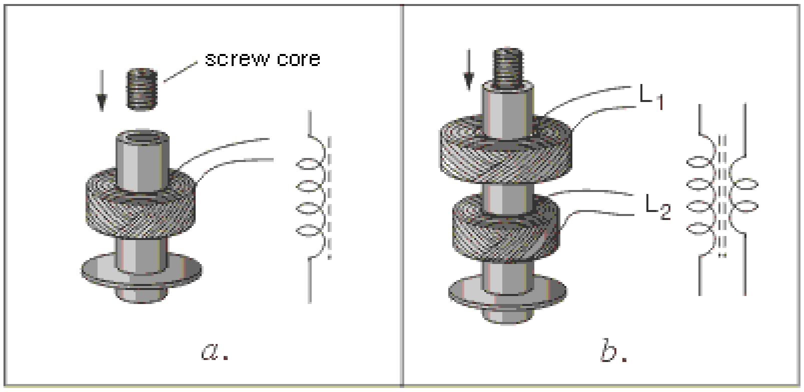

Fig. 3.2: a. Multi-layered coil with core, b. Coupled coils

Figure 3.2b shows a high-frequency transformer. As can be seen, these are two coils are coupled by magnetic induction on a shared body. When the coils are required to have exact inductance values, each coil has a ferromagnetic core that can be adjusted along the coil axis.

At very high frequencies (above 50MHz) coil inductance is small, so coils need only a few turns. These coils are made of thick copper wire (approx. 0.5mm) with no coil body, as shown on the figure 3.3a. Their inductance can be adjusted by physically stretching or squeezing the turns together.

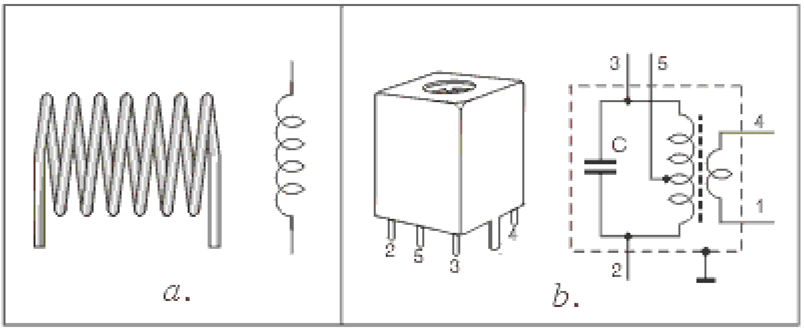

Fig. 3.3: a. High frequency coil, b. Inter-frequency transformer

Figure 3.3b shows a metal casing containing two coils, with the schematic on the right. The parallel connection of the first coil and capacitor C forms an oscillatory circuit. The second coil is used for transferring the signal to the next stage. This is used in radio-receivers and similar devices. The metal casing serves as a screen to prevent external signals affecting the coils. For the casing to be effective, it must be earthed.

Fig 3.4 shows a "pot core" inductor. The core is made in two halves and are glued together. The core is made of ferromagnetic material, commonly called "ferrite." These inductors are used at frequencies up to 100kHz. Adjustment of the inductance can be made by the brass or steel screw in the centre of the coil.

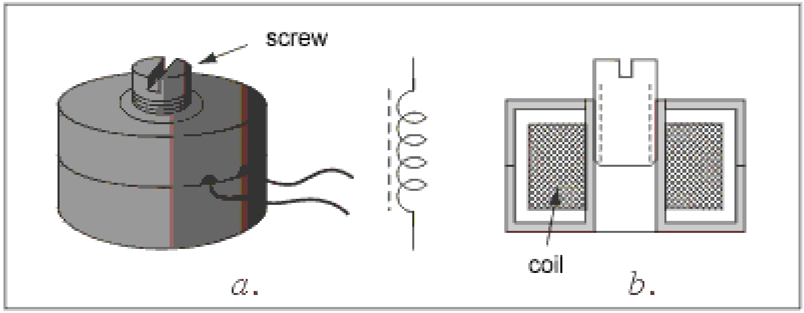

Fig. 3.4: A "pot core" inductor

3.2 Transformers

For electronic devices to function it is necessary to have a DC power supply. Batteries and rechargeable cells can fulfill the role, but a much more efficient way is to use a POWER SUPPLY. The basic component of a power supplyr is a transformer to transform the 220V "mains" to a lower value, say 12V. A common type of transformer has one primary winding which connects to the 220V and one (or several) secondary windings for the lower voltages. Most commonly, cores are made of E and I laminations, but some are made of ferromagnetic material. There are also iron core transformers used for higher frequencies. Various types of transformers are shown on the picture below.

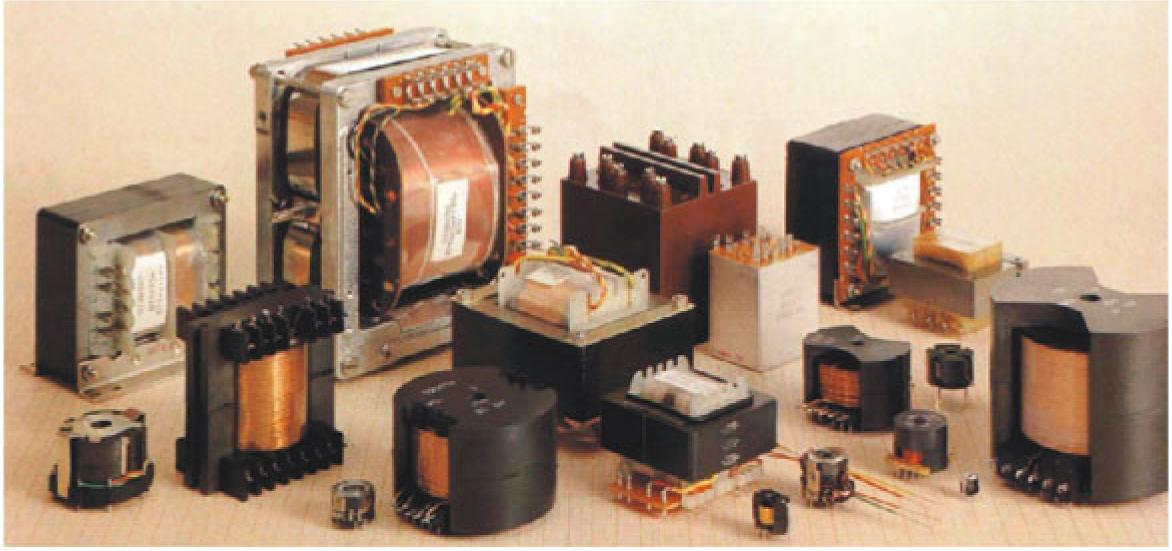

Fig. 3.5: Various types of transformers

Symbols for a transformer are shown on the figure 3.6   Two vertical lines indicate that primary and secondary windings share the same core.

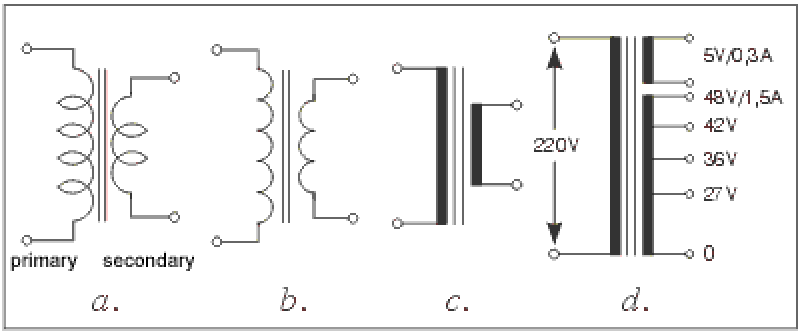

Fig. 3.6: Transformer symbols

With the transformer, manufacturers usually supply a diagram containing information about the primary and secondary windings, the voltages and maximal currents. In the case where the diagram is missing, there is a simple method for determining which winding is the primary and which is the secondary: a primary winding consists of thinner wire and more turns than the secondary. It has a higher resistance - and can be easily be tested by ohmmeter. Figure 3.6d shows the symbol for a transformer with two independent secondary windings, one of them has three tappings, giving a total of 4 different output voltages. The 5v secondary is made of thinner wire with a maximal current of 0.3A, while the other winding is made of thicker wire with a maximal current of 1.5A. Maximum voltage on the larger secondary is 48V, as shown on the figure. Note that voltages other than those marked on the diagram can be produced - for example, a voltage between tappings marked 27V and 36V equals 9V, voltage between tappings marked 27V and 42V equals 15V, etc.

3.2.1 Working principles and basic characteristics

As already stated, transformers consist of two windings, primary and the secondary (figure 3.7). When the voltage *Up* is connected to the primary winding (in our case the "mains" is 220V), AC current *Ip* flows through it. This current creates a magnetic field which passes to the secondary winding via the core of the transformer, inducing voltage *Us* (24V in our example). The "load" is connected to the secondary winding, shown in the diagram as *Rp* (30Ω in our example). A typical load could be an electric bulb working at 24V with a consumption of 19.2W.

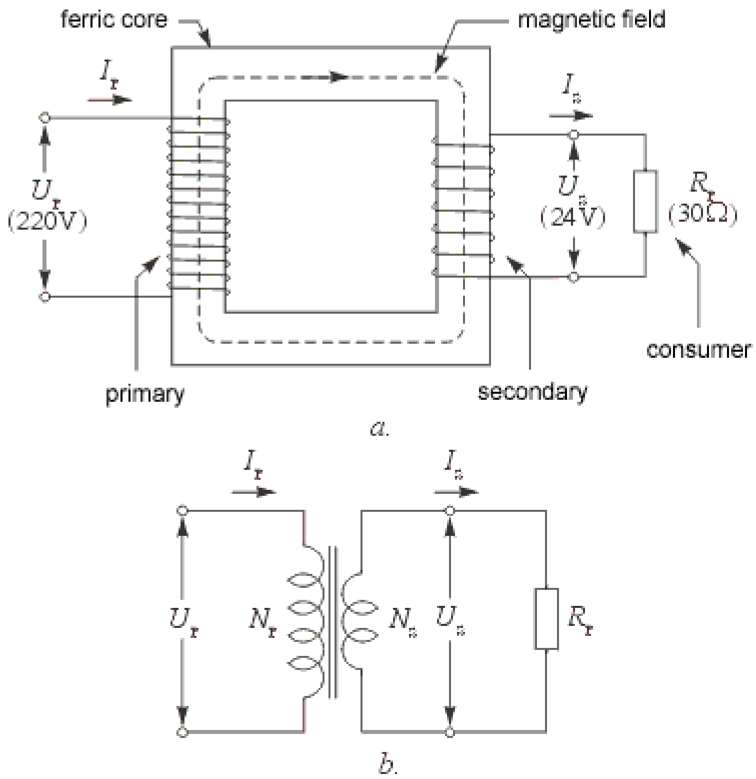

Fig. 3.7: Transformer: a. Working principles, b. Symbol

Transfer of electrical energy from the primary to the secondary is done via a magnetic field (called "flux") and a magnetic circuit called the "core of the transformer." To prevent losses, it is necessary to make sure the whole magnetic field created by the primary passes to the secondary. This is achieved by using an iron core, which has much lower magnetic resistance than air.

Primary voltage is the "mains" voltage. This value can be 220V or 110V, depending on the country. Secondary voltage is usually much lower, such as 6V, 9V, 15V, 24V, etc, but can also be higher than 220V, depending on the transformer's purpose. Relation of the primary and secondary voltage is given with the following formula:

$$\frac{U_S}{U_P} = \frac{N_S}{N_P},$$

where *Ns* and *Np* represent the number of turns on the primary and secondary winding, respectively. For instance, if *Ns* equals 80 and *Np* equals 743, secondary voltage will be:

$$U_S = U_P \cdot \frac{N_S}{N_P} = 220\,V \cdot \frac{80}{734} = 24\,V.$$

Relationship between the primary and secondary current is determined by the following formula:

$$\frac{I_P}{I_S} = \frac{N_S}{N_P}.$$

For instance, if *Rp* equals 30Ω, then the secondary current equals *Ip* = *Up*/*Rp* = 24V/30Ω = 0.8A. If *Ns* equals 80 and *Np* equals 743, primary current will be:

$$I_P = I_S \cdot \frac{N_S}{N_P} = 0{,}8\,A \cdot \frac{80}{734} = 87\,mA.$$

Transformer wattage can be calculated by the following formulae:

$$P = U_S \cdot I_S = U_P \cdot I_P.$$

In our example, the power equals:

$$P = U_S \cdot I_S = 24\,V \cdot 0{,}8\,A = 19{,}2\,W\,.$$

Everything up to this point relates to the ideal transformer. Clearly, there is no such thing as perfect, as losses are inevitable. They are present due to the fact that the windings exhibit a certain resistance value, which makes the transformer warm up during operation, and the fact that the magnetic field created by the primary does not entirely pass to the secondary. This is why the output wattage is less than the input wattage. Their ratio is called EFFICIENCY:

$$\eta = \frac{P_S}{P_P}.$$

For transformers delivering hundreds of watts, efficiency is about µ=0.85, meaning that 85% of the electrical energy taken from the mains gets to the consumer, while the 15% is lost due to previously mentioned factors in the form of heat. For example, if power required by the consumer equals *Up*Ip* = 30W, then the power which the transformer draws from the maains equals:

$$P_P = \frac{P_S}{\eta} = \frac{30\,W}{0{,}85} = 35{,}3\,W$$

To avoid any confusion here, bear in mind that manufacturers have already taken every measure in minimizing the losses of transformers and other electronic components and that, practically, this is the highest possible efficiency. When acquiring a transformer, you should only worry about the required voltage and the maximal current of the secondary. Dividing the wattage and the secondary voltage gets you the maximal current value for the consumer. Dividing the wattage and the primary voltage gets you the current that the transformer draws from network, which is important to know when buying the fuse. Anyhow, you should be able to calculate any value you might need using the appropriate formulae from above.

3.3 Practical examples with coils and transformers

On the figure 2.6b coils, along with the capacitor, form two filters for conducting the currents to the speakers.

The coil and capacitor C on figure 2.6c form a parallel oscillatory circuit for "amplifying" a particular radio signal, while rejecting all other frequencies.

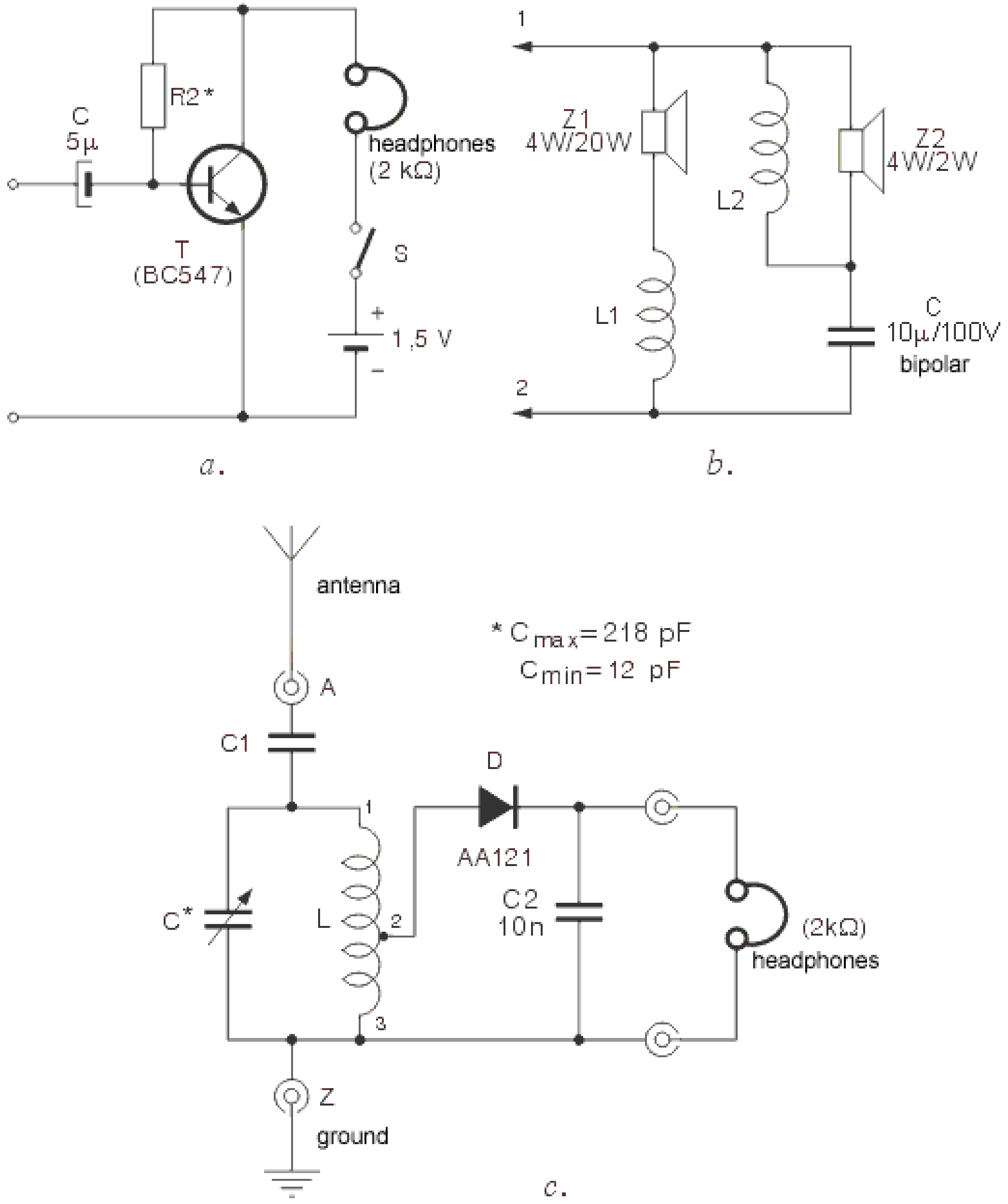
R2*
C
5µ
T
(BC547)
headphones
(2 kΩ)
S
+
1,5 V
-
a.
1
Z1
4W/20W
L2
Z2
4W/2W
L1
C
10µ/100V
bipolar
2
b.
antenna
* C_max=218 pF
C_min=12 pF
A
C1
D
1
AA121
C*
L
2
C2
10n
(2kΩ)
headphones
3
Z
ground
c.

Fig. 2.6: a. Amplifier with headphones, b. Band-switch, c. Detector radio-receiver

The most obvious application for a transformer is in a power supply. A typical transformer is shown in figure 3.8 and is used for converting 220V to 24V.

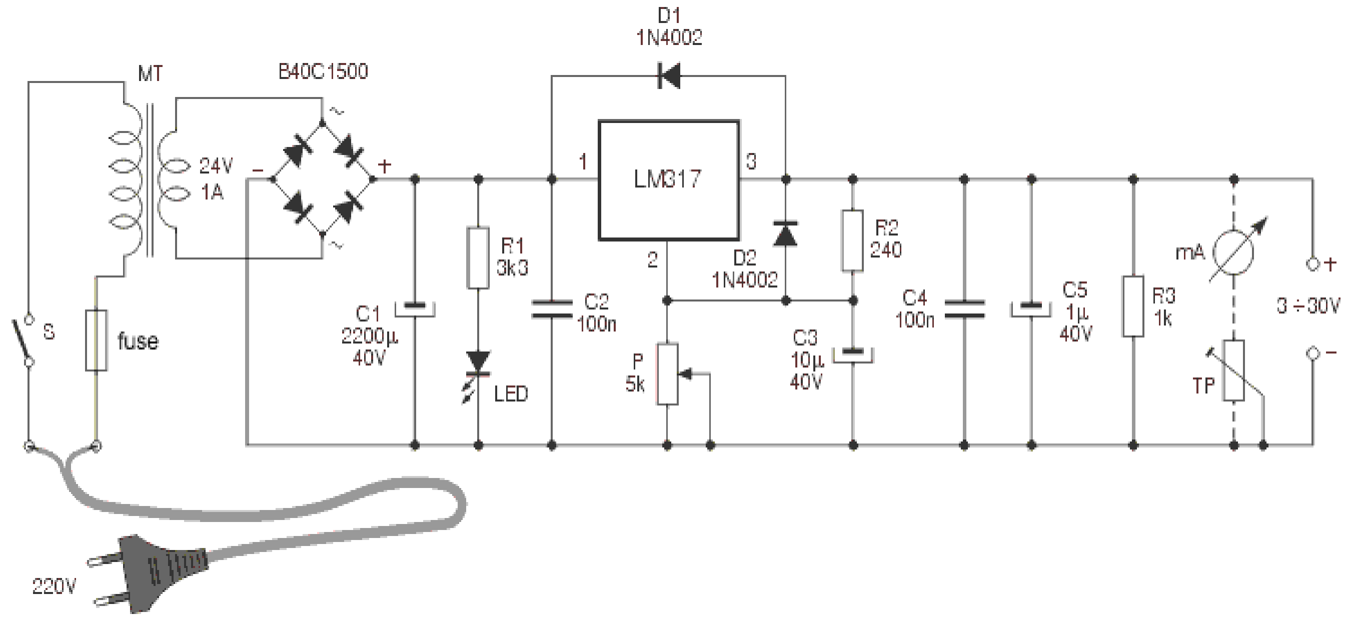
D1
1N4002
MT
B40C1500
24V
1A
LM317
1
3
2
R1
3k3
R2
240
D2
1N4002
C1
2200µ
40V
C2
100n
C3
10µ
40V
C4
100n
C5
1µ
40V
R3
1k
mA
TP
P
5k
LED
S
fuse
3 ÷ 30V
+
-
220V

Fig. 3.8: Stabilized converter with circuit LM317

Output DC voltage can be adjusted via a linear potentiometer P, in 3~30V range.

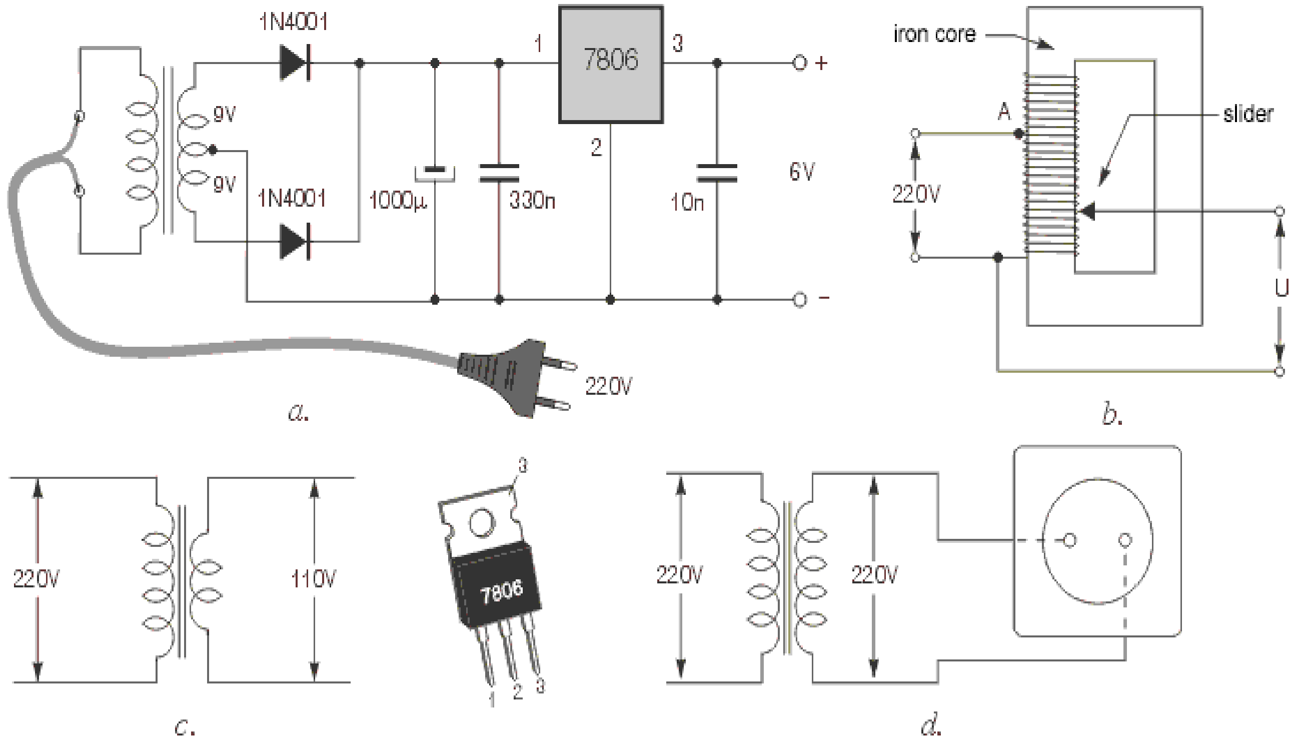
1N4001
9V
9V
1N4001
7806
1
3
2
1000µ
330n
10n
6V
+
-
220V
a.
iron core
slider
A
220V
U
b.
220V
110V
c.
7806
3
1
2
3
220V
220V
d.

Fig. 3.9: a. Stabilized converter with regulator 7806, b. auto-transformer, c. transformer for devices working at 110V, d. isolating transformer

Figure 3.9a shows a simple power supply, using a transformer with a centre-tap on the secondary winding. This makes possible the use two diodes instead of the bridge in figure 3.8.

Special types of transformers, mainly used in laboratories, are auto-transformers. The diagram for an auto-transformer is shown in figure 3.9b. It features only one winding, wound on an iron core. Voltage is taken from the transformer via a slider. When the slider is in its lowest position, voltage equals zero. Moving the slider upwards increases the voltage U, to 220V. Further moving the slider increases the voltage U above 220V.

The transformer in figure 3.9c converts 220v to 110v and is used for supplying devices designed to work on 110V.

As a final example, figure 3.9d represents an isolating transformer. This transformer features the same number of turns on primary and secondary windings. Secondary voltage is the same as the primary, 220V, but is completely isolated from the "mains," minimizing the risks of electrical shock. As a result, a person can stand on a wet floor and touch any part of the secondary without risk, which is not the case with the normal power outlet.

4. Transistors

Transistors are active components and are found everywhere in electronic circuits. They are used as amplifiers and switching devices. As amplifiers, they are used in high and low frequency stages, oscillators, modulators, detectors and in any circuit needing to perform a function. In digital circuits they are used as switches.

There is a large number of manufacturers around the world who produce semiconductors (transistors are members of this family of components), so there are literally thousands of different types. There are low, medium and high power transistors, for working with high and low frequencies, for working with very high current and/or high voltages. Several different transistors are shown on 4.1.

The most common type of transistor is called bipolar and these are divided into NPN and PNP types.
Their construction-material is most commonly silicon (their marking has the letter B) or germanium (their marking has the letter A). Original transistor were made from germanium, but they were very temperature-sensitive. Silicon transistors are much more temperature-tolerant and much cheaper to manufacture.

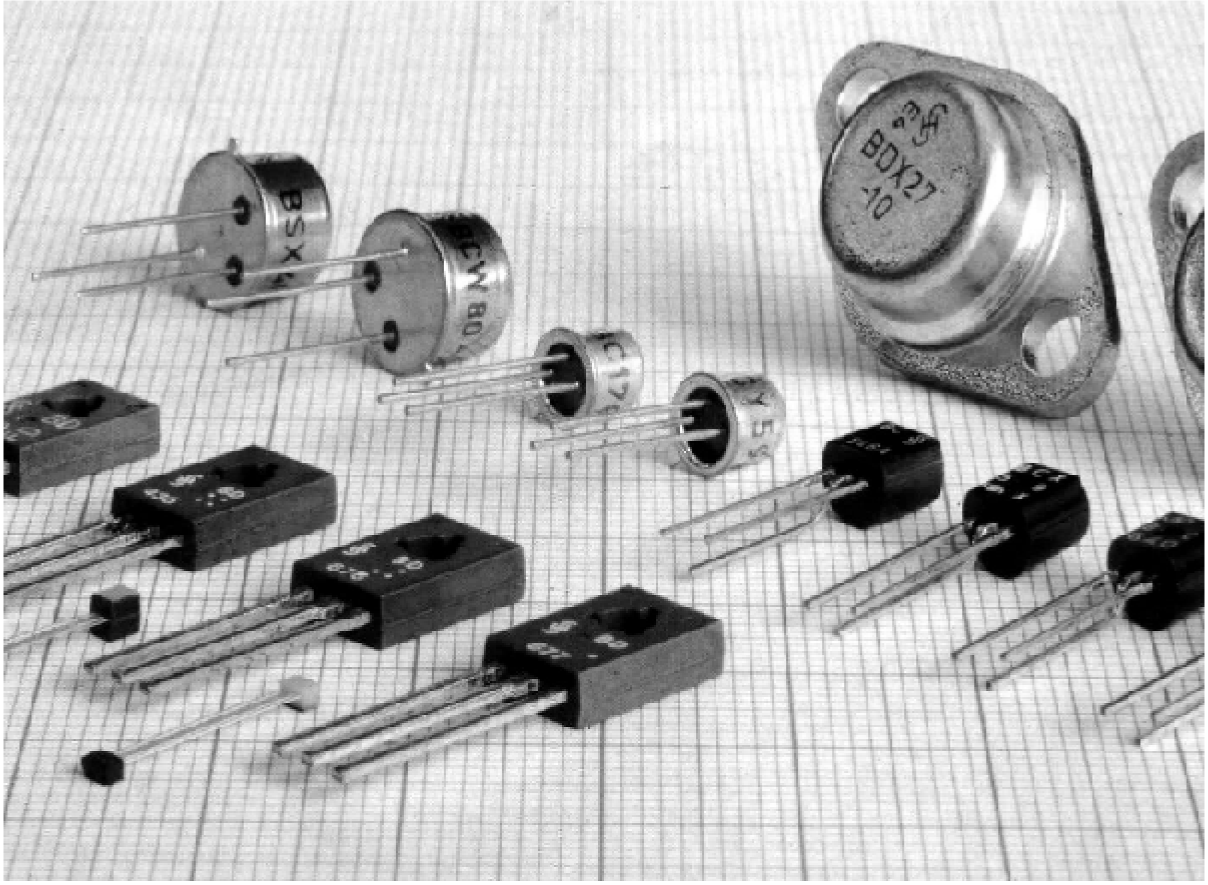
BSX
BCW 80
BDX27
10

Fig. 4.1: Different transistors

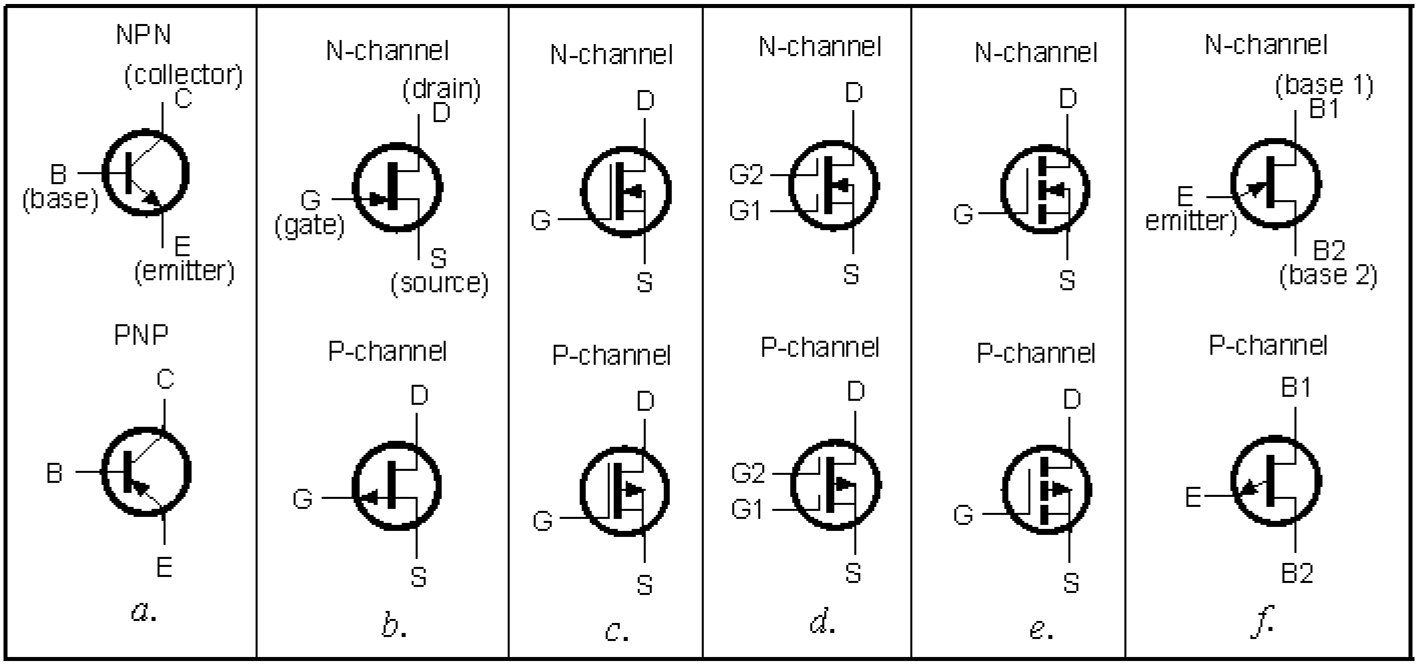
NPN
(collector)
C
B
(base)
E
(emitter)
PNP
C
B
E
a.
N-channel
(drain)
D
G
(gate)
S
(source)
P-channel
D
G
S
b.
N-channel
D
G
S
P-channel
D
G
S
c.
N-channel
D
G2
G1
S
P-channel
D
G2
G1
S
d.
N-channel
D
G
S
P-channel
D
G
S
e.
N-channel
(base 1)
B1
E
emitter)
B2
(base 2)
P-channel
B1
E
B2
f.

Fig. 4.2: Transistor symbols: a - bipolar, b - FET, c - MOSFET, d - dual gate MOSFET, e - inductive channel MOSFET, f - single connection transistor

The second letter in transistor’s marking describes its primary use:
C - low and medium power LF transistor,
D - high power LF transistor,
F - low power HF transistor,
G - other transistors,
L - high power HF transistors,
P - photo transistor,
S - switch transistor,
U - high voltage transistor.

Here are few examples:
AC540 - germanium core, LF, low power,
AF125 - germanium core, HF, low power,
BC107 - silicon, LF, low power (0.3W),
BD675 - silicon, LF, high power (40W),
BF199 - silicon, HF (to 550 MHz),
BU208 - silicon (for voltages up to 700V),
BSY54 - silicon, switching transistor.
There is a possibility of a third letter (R and Q - microwave transistors, or X - switch transistor), but these letters vary from manufacturer to manufacturer.
The number following the letter is of no importance to users.
American transistor manufacturers have different marks, with a 2N prefix followed by a number (2N3055, for example). This mark is similar to diode marks, which have a 1N prefix (e.g. 1N4004).
Japanese bipolar transistor are prefixed with a: 2SA, 2SB, 2SC or 2SD, and FET-s with 3S:
2SA - PNP, HF transistors,
2SB - PNP, LF transistors,
2SC - NPN, HF transistors,
2SD - NPN, HF transistors.

Several different transistors are shown in photo 4.1, and symbols for schematics are on 4.2. Low power transistors are housed in a small plastic or metallic cases of various shapes. Bipolar transistors have three leads: for base (B), emitter (E), and for collector (C). Sometimes, HF transistors have another lead which is connected to the metal housing. This lead is connected to the ground of the circuit, to protect the transistor from possible external electrical interference. Four leads emerge from some other types, such as two-gate FETs. High power transistors are different from low-to-medium power, both in size and in shape.

It is important to have the manufacturer’s catalog or a datasheet to know which lead is connected to what part of the transistor. These documents hold the information about the component's correct use (maximum current rating, power, amplification, etc.) as well as a diagram of the pinout. Placement of leads and different housing types for some commonly used transistors are in diagram 4.3.

Package TO5, TO18

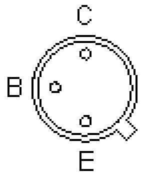

NPN	PNP
BC107	BC177
BC108	BC178
BC109	BC179
BFX84	BC186
BFY50	BC187
2N706	BFX88
2N2369	BCY71
BC286	BC287
	2N2904

TO72

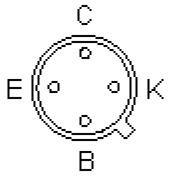

NPN	PNP
BF180	AF139
BF181	AF178
BF182	AF179
BF183	AF180
BF200	AF181

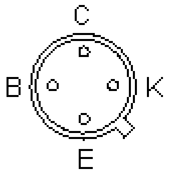

NPN	PNP
BF115	AF124
BF167	AF125
BF173	AF126
BF184	AF127
BF185	

TO3
and similar

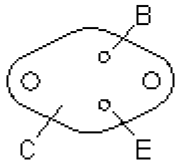

NPN	PNP
2N3055	PNP305
BDY20	BDX18
BD121	OC26
BD123	AD149
AD161	AD162

TO126

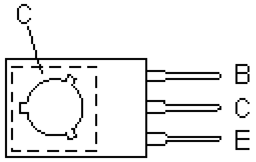

NPN	PNP
BD135	BD136
BD131	BD132
BD437	BD438
BUP41	

TO-1

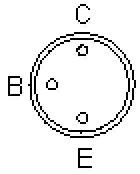

NPN	PNP
AC176	AC128
AC187	AC188

X-55

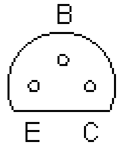

NPN	PNP
BC182	BC212
BC183	BC213
BC184	BC214
2N3707	2N3702
2N3710	2N3703

TO92

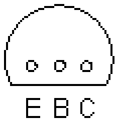

NPN	PNP
BC183L	BC213L
BC237B	BC557
BC547	BC307B
BC548	2N4402
BC546	
2N3705	
2N3903	

SOT25

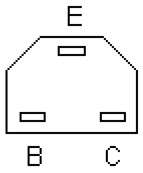

NPN
BF194
BF195
BF196
BF197

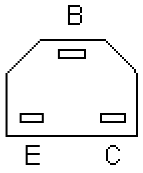

NPN	PNP
BC157	BC147
BC158	BC148
BC159	BC149
BCX35	BCX31

TO18

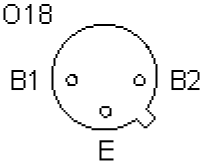

Unijunction transistor

2N2646	2N4870
2N2647	2N4871

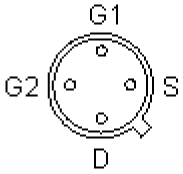

3N140
3N141
40673

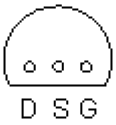

BF256
2N3819
(connection FET)

SOT103

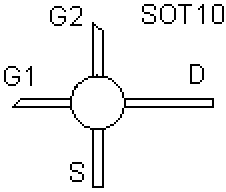

BF960	BF981
BF961	3SK81

TO220

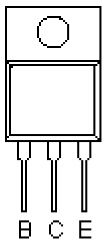

NPN	PNP
BD539	BD540
BD743	BD744
TIP29C	TIP30C
BU407	BD244C
BUP30	BD240C
2N6099	BD242C
BD243C	
D44C10	
BD241C	

X80

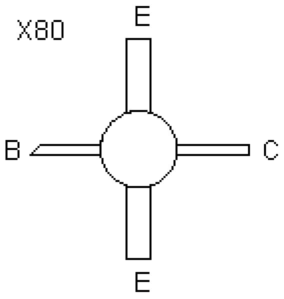

BFR14	BFR49

SMD SOT23

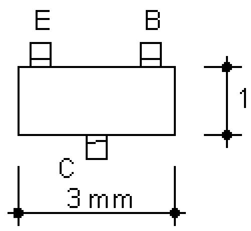

NPN	PNP
BC846B	BC856B
BC847B	BC857B
BC848B	BC858B
BC849B	BC859B

Fig. 4.3: Pinouts of some common packages

It might be useful to remember the pinout for TO-1, TO-5, TO-18 and TO-72 packages and compare them with the drawing 4.2 (a). These transistors are the ones you will come across frequently in everyday work.

The TO-3 package, which is used to house high-power transistors, has only two pins, one for base, and one for emitter. The collector is connected to the package, and this is connected to the rest of the circuit via one of the screws which fasten the transistor to the heat-sink.

Transistors used with very high frequencies (like BFR14) have pins shaped differently.
One of the breakthroughs in the field of electronic components was the invention of SMD (surface mount devices) circuits. This technology allowed manufacturers to achieve tiny components with the same properties as their larger counterparts, and therefore reduce the size and cost of the design. One of the SMD housings is the SOT23 package. There is, however, a trade-off to this, SMD components are difficult to solder to the PC board and they usually need special soldering equipment.

As we said, there are literally thousands of different transistors, many of them have similar characteristics, which makes it possible to replace a faulty transistor with a different one. The characteristics and similarities can be found in comparison charts. If you do not have one these charts, you can try some of the transistors you already have. If the circuit continues to operate correctly, everything is ok. You can only replace an NPN transistor with an NPN transistor. The same goes if the transistor is PNP or a FET. It is also necessary to make sure the pinout is correct, before you solder it in place and power up the project.
As a helpful guide, there is a chart in this chapter which shows a list of replacements for some frequently used transistors.

4.1 The working principle of a transistor

Transistors are used in analog circuits to amplify a signal. They are also used in power supplies as a regulator and you will also find them used as a switch in digital circuits.
The best way to explore the basics of transistors is by experimenting. A simple circuit is shown below. It uses a power transistor to illuminate a globe. You will also need a battery, a small light bulb (taken from a flashlight) with properties near 4.5V/0.3A, a linear potentiometer (5k) and a 470 ohm resistor. These components should be connected as shown in figure 4.4a.

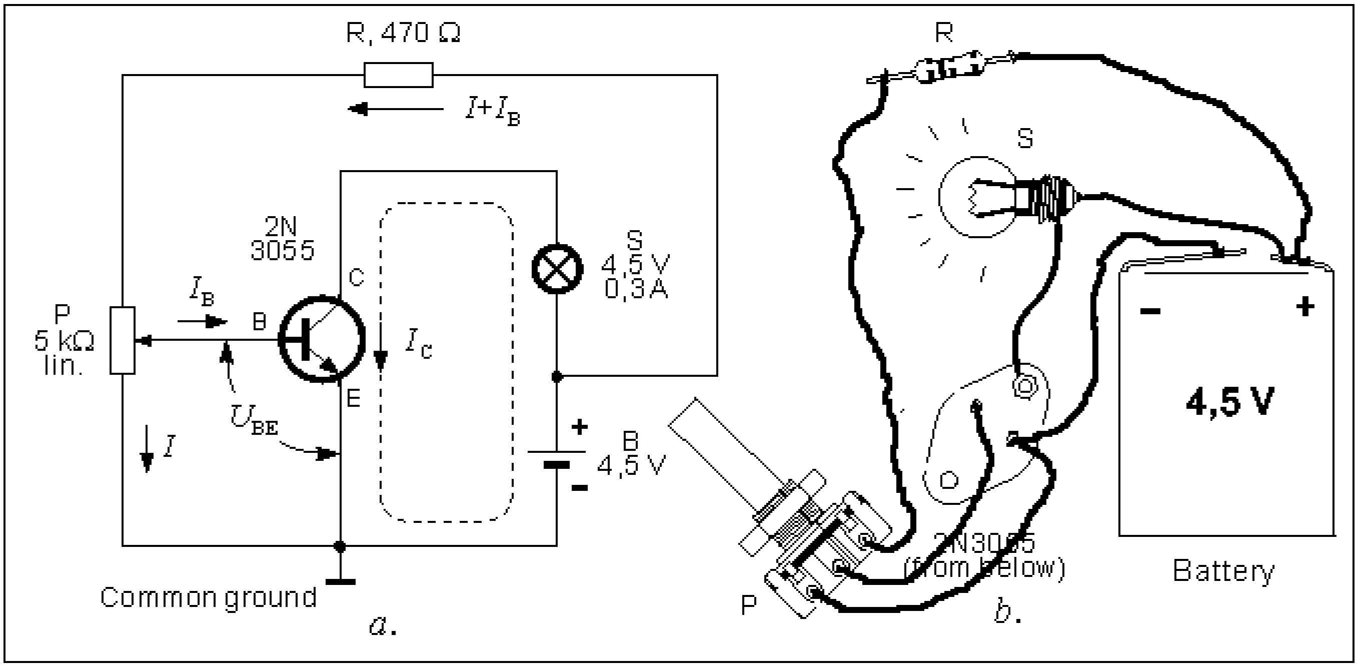

Fig. 4.4: Working principle of a transistor: potentiometer moves toward its upper position - voltage on the base increases - current through the base increases - current through the collector increases - the brightness of the globe increases.

Resistor (R) isn't really necessary, but if you don't use it, you mustn't turn the potentiometer (pot) to its high position, because that would destroy the transistor - this is because the DC voltage UBE (voltage between the base and the emitter), should not be higher than 0.6V, for silicon transistors.

Turn the potentiometer to its lowest position. This brings the voltage on the base (or more correctly between the base and ground) to zero volts (UBE = 0). The bulb doesn't light, which means there is no current passing through the transistor.

As we already mentioned, the potentiometers lowest position means that UBE is equal to zero. When we turn the knob from its lowest position UBE gradually increases. When UBE reaches 0.6v, current starts to enter the transistor and the globe starts to glow. As the pot is turned further, the voltage on the base remains at 0.6v but the current increases and this increases the current through the collector-emitter circuit. If the pot is turned fully, the base voltage will increase slightly to about 0.75v but the current will increase significantly and the globe will glow brightly.

If we connected an ammeter between the collector and the bulb (to measure IC), another ammeter between the pot and the base (for measuring IB), and a voltmeter between the ground and the base and repeat the whole experiment, we will find some interesting data. When the pot is in its low position UBE is equal to 0V, as well as currents IC and IB. When the pot is turned, these values start to rise until the bulb starts to glow when they are: UBE = 0.6V, IB = 0.8mA and IB = 36 mA (if your values differ from these values, it is because the 2N3055 the writer used

doesn't have the same specifications as the one you use, which is common when working with transistors).
The end result we get from this experiment is that when the current on the base is changed, current on the collector is changed as well.

Let's look at another experiment which will broaden our knowledge of the transistor. It requires a BC107 transistor (or any similar low power transistor), supply source (same as in previous experiment), 1M resistor, headphones and an electrolytic capacitor whose value may range between 10u to 100µF with any operating voltage.
A simple low frequency amplifier can be built from these components as shown in diagram 4.5.

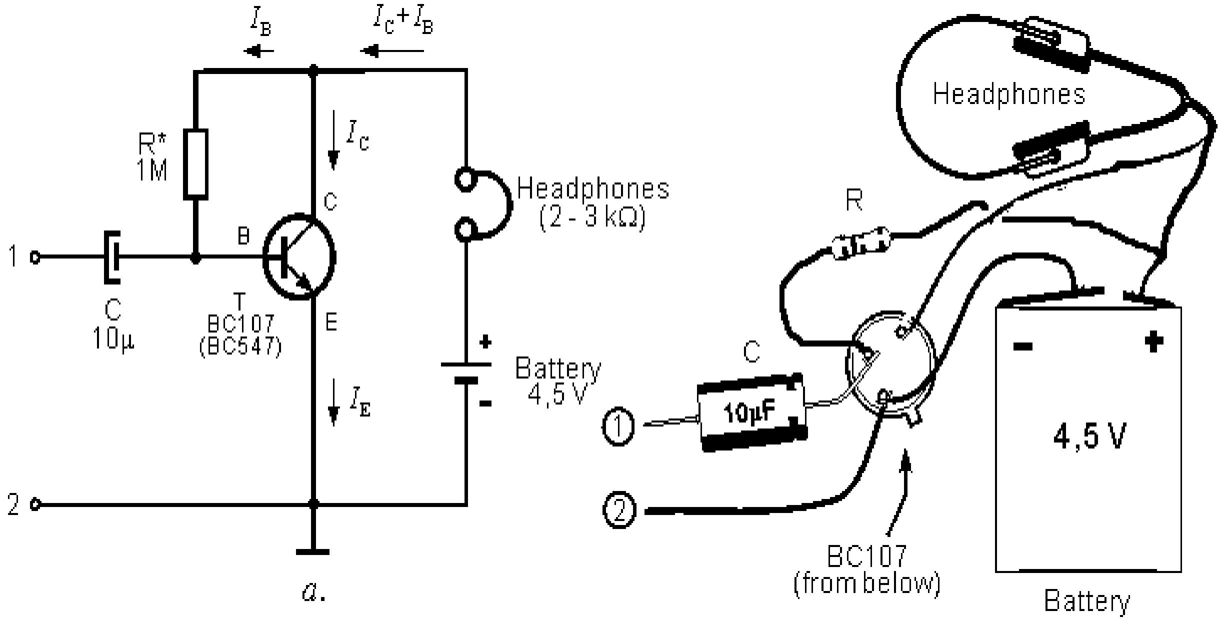

Fig. 4.5: A simple transistor amplifier

It should be noted that the schematic 4.5a is similar to the one on 4.4a. The main difference is that the collector is connected to headphones. The "turn-on" resistor - the resistor on the base, is 1M. When there is no resistor, there is no current flow IB, and no Ic current. When the resistor is connected to the circuit, base voltage is equal to 0.6V, and the base current IB = 4µA. The transistor has a gain of 250 and this means the collector current will be 1 mA. Since both of these currents enter the transistor, it is obvious that the emitter current is equal to IE = IC + IB. And since the base current is in most cases insignificant compared to the collector current, it is considered that:

$$I_C = I_E.$$

The relationship between the current flowing through the collector and the current flowing through the base is called the transistor's current amplification coefficient, and is marked as hFE. In our example, this coefficient is equal to:

$$h_{FE} = \frac{I_C}{I_B} = \frac{1mA}{4\,\mu A} = \frac{1 \cdot 10^{-3}\,A}{4 \cdot 10^{-6}\,A} = 250.$$

Put the headphones on and place a fingertip on point 1. You will hear a noise. You body picks up the 50Hz AC "mains" voltage. The noise heard from the headphones is that voltage, only amplified by the transistor. Let's explain this circuit a bit more. Ac voltage with frequency 50Hz is connected to transistor's base via the capacitor C. Voltage on the base is now equal to the sum of a DC voltage (0.6 approx.) via resistor R, and AC voltage "from" the finger. This means that this base voltage is higher than 0.6V, fifty times per second, and fifty times slightly lower than that. Because of this, current on the collector is higher than 1mA fifty times per second, and fifty times lower. This variable current is used to shift the membrane of the speakerphones forward fifty times per second and fifty times backwards, meaning that we can hear the 50Hz tone on the output.
Listening to a 50Hz noise is not very interesting, so you could connect to points 1 and 2 some low frequency signal source (CD player or a microphone).

There are literally thousands of different circuits using a transistor as an active, amplifying device. And all these transistors operate in a manner shown in our experiments, which means that by building this example, you're actually building a basic building block of electronics.

4.2 Basic characteristics of transistors

Selecting the correct transistor for a circuit is based on the following characteristics: maximum voltage rating between the collector and the emitter UCEmax, maximum collector current ICmax and the maximum power rating PCmax.
If you need to change a faulty transistor, or you feel comfortable enough to build a new circuit, pay attention to these three values. Your circuit must not exceed the maximum rating values of the transistor. If this is disregarded there are possibilities of permanent circuit damage. Beside the values we mentioned, it is sometimes important to know the current amplification, and maximum frequency of operation.
When there is a DC voltage UCE between the collector (C) and emitter (E) with a collector current, the transistor acts as a small electrical heater whose power is given with this equation:

$$P_C = U_{CE} \cdot I_C.$$

Because of that, the transistor is heating itself and everything in its proximity. When UCE or ICE rise (or both of them), the transistor may overheat and become damaged. Maximum power rating for a transistor, is PCmax (found in a datasheet). What this means is that the product of UCE and IC should should not be higher than PCmax:

$$U_{CE} \cdot I_C = P_{C\max}$$

So, if the voltage across the transistor is increased, the current must be dropped.
For example, maximum ratings for a BC107 transistor are:
ICmax=100mA,
UCEmax = 45V and
PCmax = 300mW
If we need a Ic=60mA , the maximum voltage is:

$$U_{CE} = \frac{P_{C\max}}{I_C} = \frac{300\,mW}{60\,mA} = 5\,V.$$

For UCE = 30V, the maximum current is:

$$I_C = \frac{P_{C\max}}{U_{CE}} = \frac{300\,mW}{30V} = 10\,mA.$$

Among its other characteristics, this transistor has current amplification coefficient in range between hFE= 100 to 450, and it can be used for frequencies under 300MHz. According to the recommended values given by the manufacturer, optimum results (stability, low distortion and noise, high gain, etc.) are with UCE=5V and IC=2mA.
There are occasions when the heat generated by a transistor cannot be overcome by adjusting voltages and current. In this case the transistors have a metal plate with hole, which is used to attach it to a heat-sink to allow the heat to be passed to a larger surface.

Current amplification is of importance when used in some circuits, where there is a need for equal amplification of two transistors. For example, 2N3055H transistors have hFE within range between 20 and 70, which means that there is a possibility that one of them has 20 and other 70. This means that in cases when two identical coefficients are needed, they should be measured. Some multimeters have the option for measuring this, but most don't. Because of this we have provided a simple circuit (4.6) for testing transistors. All you need is an option on your multimeter for measuring DC current up to 5mA. Both diodes (1N4001, or similar general purpose silicon diodes) and 1k resistors are used to protect the instrument if the transistor is "damaged". As we said, current gain is equal to hFE = IC / IB. In the circuit, when the switch S is pressed, current flows through the base and is approximately equal to IB=10uA, so if the collector current is displayed in milliamps. The gain is equal to:

$$h_{FE} = 100 \cdot I_C$$

For example, if the multimeter shows 2.4mA, hFE = 2.4*100 = 240.

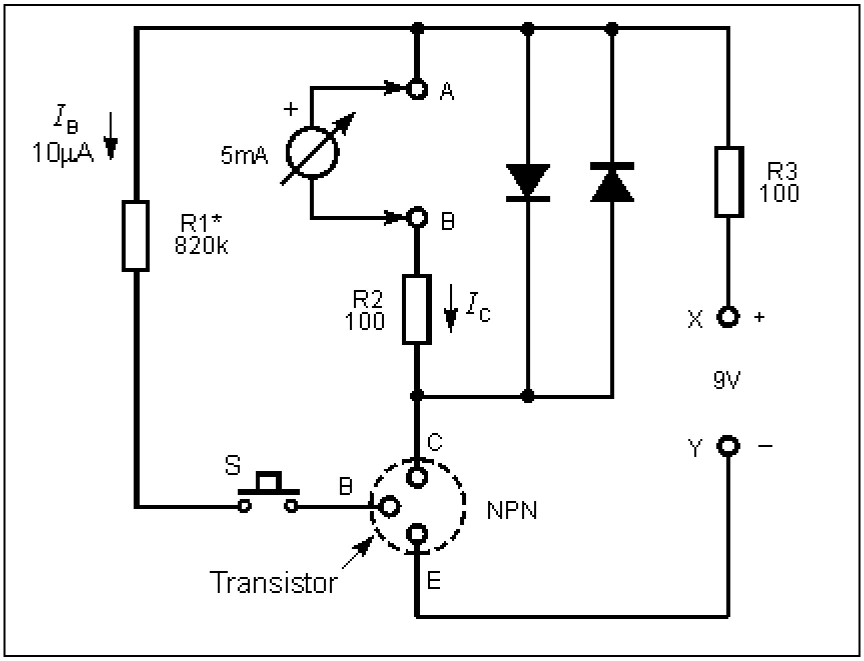

Fig. 4.6: Measuring the hFE

While measuring NPN transistors, the supply should be connected as shown in the diagram. For PNP transistors the battery is reversed. In that case, probes should be reversed as well if you're using analog instrument (one with a needle). If you are using a digital meter (highly recommended) it doesn't matter which probe goes where, but if you do it the same way as you did with NPN there would be a minus in front of the read value, which means that current flows in the opposite direction.

4.3 The safest way to test transistors

Another way to test transistor is to put it into a circuit and detect the operation. The following circuit is a multivibrator. The "test transistor" is T2. The supply voltage can be up to 12v. The LED will blink when a good transistor is fitted to the circuit.

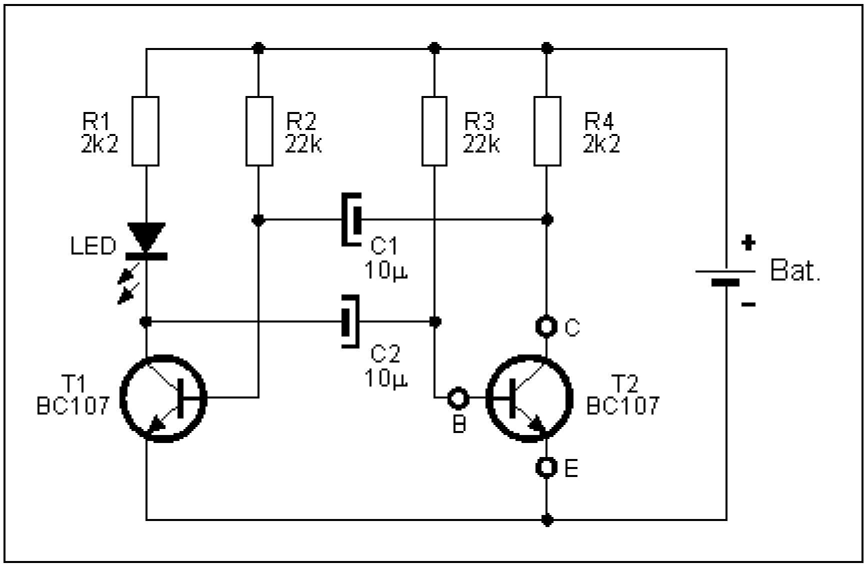

Fig. 4.7: Oscillator to test transistors

To test PNP transistors, same would go, only the transistor which would need to be replaced is the T1, and the battery, LED, C1 and C2 should be reversed.

4.4 TUN and TUP

As we previously said, many electronic devices work perfectly even if the transistor is replaced with a similar device. Because of this, many magazines use the identification TUN and TUP in their schematics. These are general purpose transistors. TUN identifies a general purpose NPN transistor, and TUP is a general purpose PNP transistor.

TUN = Transistor Universal NPN and TUP = Transistor Universal PNP.

These transistors have following characteristics:

U_{CEmax}	20V
I_{Cmax}	100mA
h_{FEmin}	110
P_{Cmax}	100mW
f_{Tmin}	100MHz

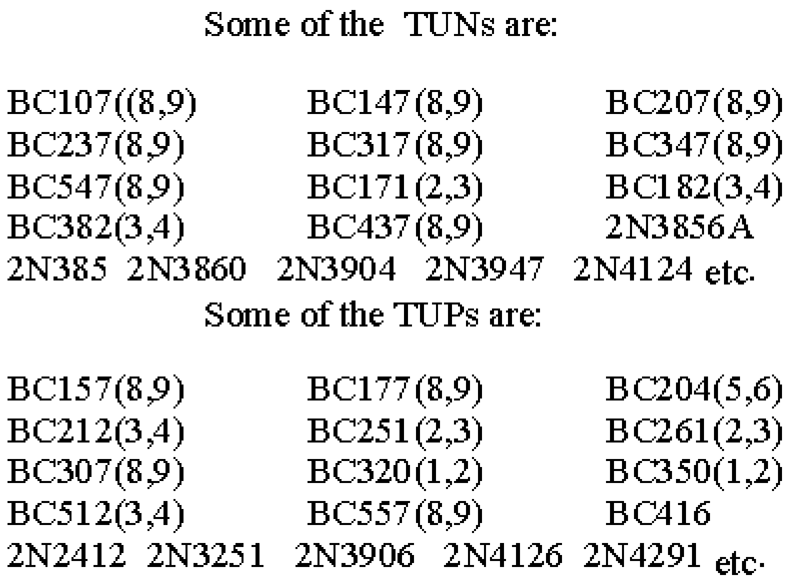

Some of the TUNs are:

BC107((8,9) BC147(8,9) BC207(8,9)
BC237(8,9) BC317(8,9) BC347(8,9)
BC547(8,9) BC171(2,3) BC182(3,4)
BC382(3,4) BC437(8,9) 2N3856A
2N385 2N3860 2N3904 2N3947 2N4124 etc.

Some of the TUPs are:

BC157(8,9) BC177(8,9) BC204(5,6)
BC212(3,4) BC251(2,3) BC261(2,3)
BC307(8,9) BC320(1,2) BC350(1,2)
BC512(3,4) BC557(8,9) BC416
2N2412 2N3251 2N3906 2N4126 2N4291 etc.

4.5 Practical example

The most common role of a transistor in an analog circuit is as an active (amplifying) component. Diagram 4.8 shows a simple radio receiver - commonly called a "Crystal Set with amplifier."

Variable capacitor C and coil L form a parallel oscillating circuit which is used to pick out the signal of a radio station out of many different signals of different frequencies. A diode, 100pF capacitor and a 470k resistor form a diode detector which is used to transform the low frequency voltage into information (music, speech). Information across the 470k resistor passes through a 1uF capacitor to the base of a transistor. The transistor and its associated components create a low frequency amplifier which amplifies the signal.
On figure 4.8 there are symbols for a common ground and grounding. Beginners usually assume these two are the same which is a mistake. On the circuit board the common ground is a copper track whose size is significantly wider than the other tracks. When this radio receiver is built on a circuit board, common ground is a copper strip connecting holes where the lower end of the capacitor C, coil L 100pF capacitor and 470k resistor are soldered. On the other hand, grounding is a metal rod stuck in a wet earth (connecting your circuits grounding point to the plumbing or heating system of your house is also a good way to ground your project).
Resistor R2 biases the transistor. This voltage should be around 0.7V, so that voltage on the collector is approximately equal to half the battery voltage.

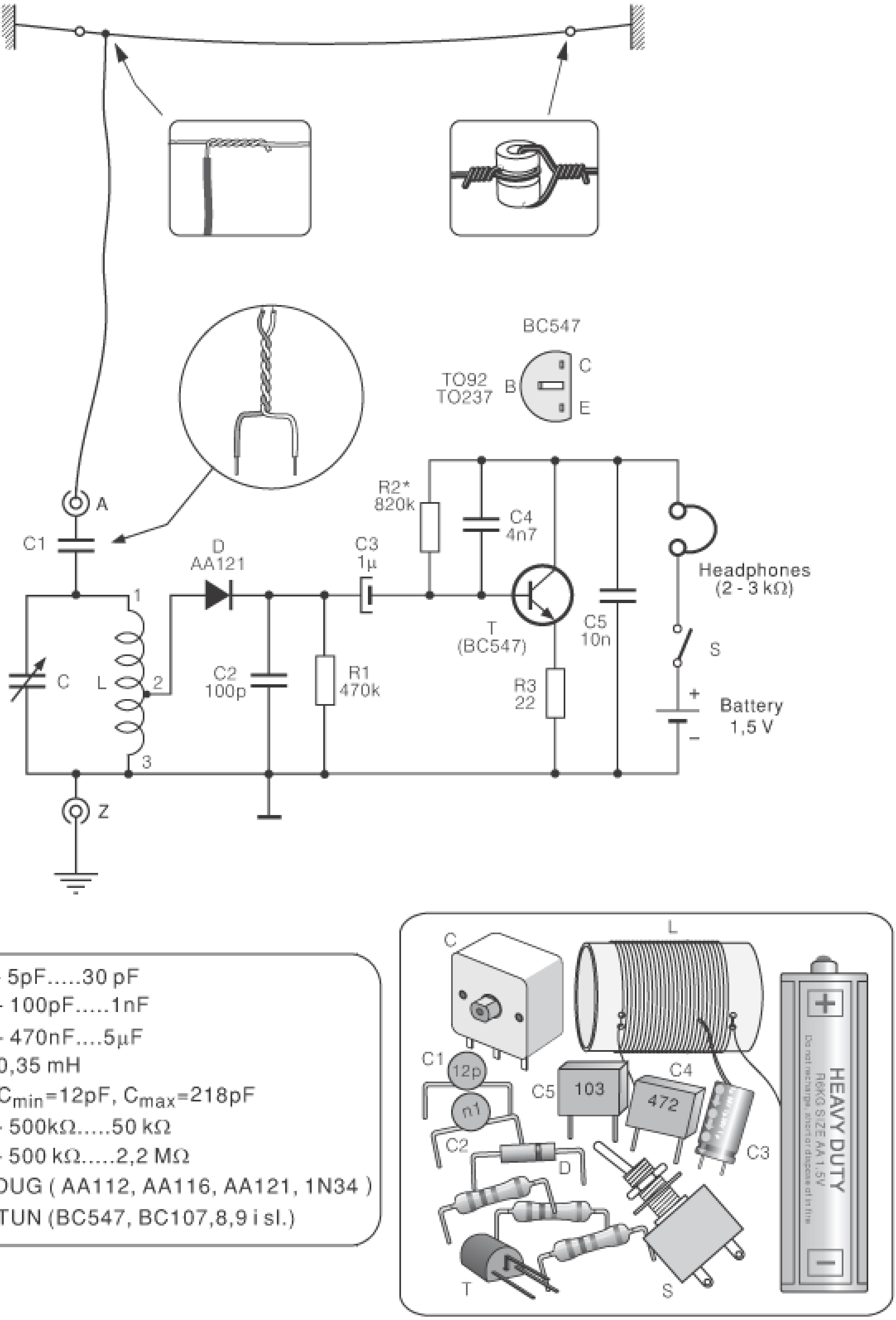
BC547
TO92
TO237
B
C
E
A
C1
D
AA121
C3
1µ
R2*
820k
C4
4n7
Headphones
(2 - 3 kΩ)
1
C
L
2
3
C2
100p
R1
470k
T
(BC547)
C5
10n
S
R3
22
+
Battery
1,5 V
−
Z
C
L
C1
12p
C5
103
C4
472
n1
C2
D
C3
HEAVY DUTY
R6KG SIZE AA 1,5V
+
−
T
S

C1 - 5pF.....30 pF
C2 - 100pF.....1nF
C3 - 470nF....5µF
L - 0,35 mH
C - C_{min}=12pF, C_{max}=218pF
R1 - 500kΩ.....50 kΩ
R2 - 500 kΩ.....2,2 MΩ
D - DUG (AA112, AA116, AA121, 1N34)
T - TUN (BC547, BC107,8,9 i sl.)

Fig. 4.8: Detector receiver with a simple amplifier

5. Diodes

As with transistors, diodes are fabricated from semi-conducting material. So, the first letter in their identification is A for germanium diode or B for silicon diode. They can be encased in glass, metal or a plastic housing. They have two leads: cathode (k) and an anode (A). The most important property of all diodes is their resistance is very low in one direction and very large in the opposite direction.

When a diode is measured with a multimeter and it reads a low value of ohms, this is not really the resistance of the diode. It represents the voltage drop across the junction of the diode. This means a multimeter can only be used to detect if the junction is not damaged. If the reading is low in one direction and very high in the other direction, the diode is operational.

When a diode is placed in a circuit and the voltage on the anode is higher than the cathode, it acts like a low value resistor and current will flow.

If it is connected in the opposite direction it acts like a large value resistor and current does not flow.

In the first case the diode is said to be "forward biased" and in the second case it is "reverse biased."

Figure 5.1 shows several different diodes:

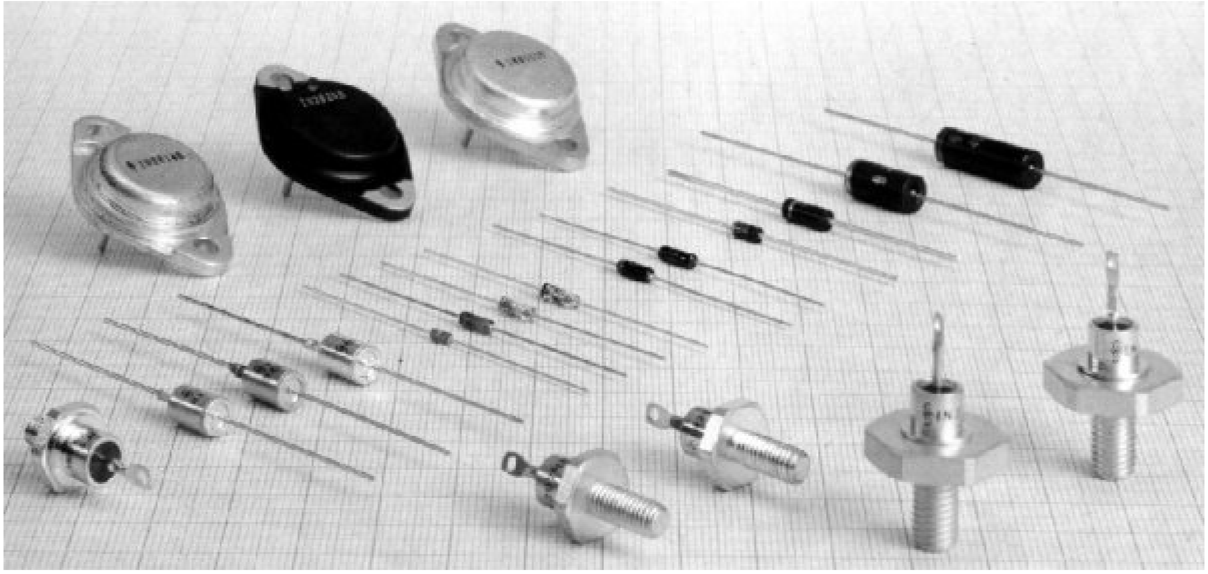

Fig. 5.1: Several different types of diodes

The diodes above are all single diodes, however 4 diodes are available in a single package. This is called a BRIDGE or BRIDGE RECTIFIER. Examples of a bridge are shown in the diagram below:

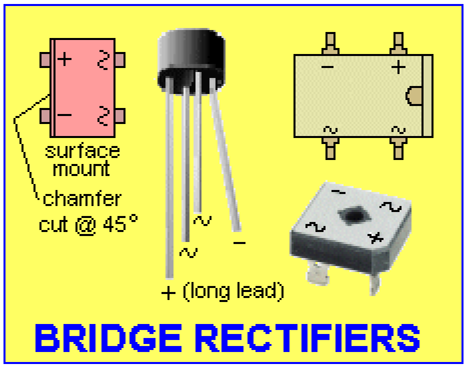
surface mount
chamfer cut @ 45°
+ (long lead)
BRIDGE RECTIFIERS

You must be able to identify each of the 4 leads on a bridge so that it can be inserted into a circuit around the correct way. The surface-mount device above is identified by a cut @ 45° along one side. The leaded bridge has one leg longer than the others and the top is marked with AC marks and "+." The high-current bridge has a corner cut off and the other surface-mount device has a cut or notch at one end.

These devices are added to a circuit as shown in the next diagram:

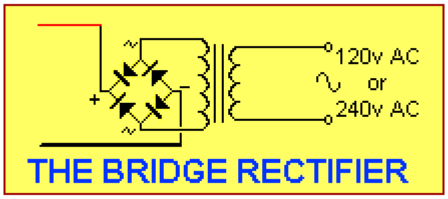
120v AC
or
240v AC
THE BRIDGE RECTIFIER

The 4 diodes face the same direction and this means a single diode can be shown on the circuit diagram:

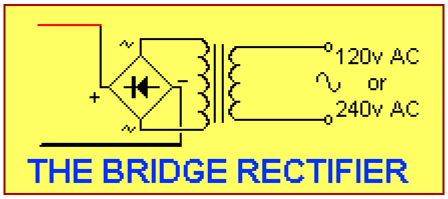
120v AC
or
240v AC
THE BRIDGE RECTIFIER

Symbols in 5.2 show a number of diodes. There are a number of specially-designed diodes: for high current, high-speed, low voltage-drop, light-detection, and varying capacitance as the voltage is altered. Most diodes are made from silicon as it will withstand high temperature, however germanium is used if a low voltage-drop is required. There is also a light emitting diode called a LED, but this is a completely different type of diode.

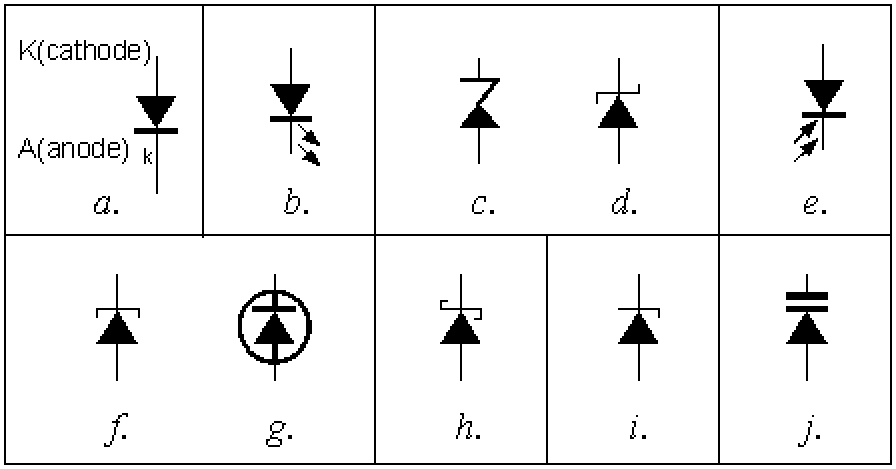

Fig. 5.2: Diode symbols: a - standard diode, b - LED, c, d - Zener, e - photo, f,g - tunnel, h - Schottky, i - breakdown, j - capacitative

LEDs (Light Emitting Diodes) are constructed from a crystalline substance that emits light when a current flows through it. Depending on the crystalline material: red, yellow, green, blue or orange light is produced. The photo below shows some of the colours hat can be produced by LEDs:

It is not possible to produce white light from any of these materials, so a triad of red, blue and green is placed inside a case and they are all illuminated at the same time to produce white light. Recently, while light has been produced from a LED by a very complex and interesting process that can be found on Wikipedia: http://en.wikipedia.org/wiki/LED

LEDs have a cathode and anode lead and must be connected to DC around the correct way. The cathode lead is identified on the body by a flat-spot on the side of the LED. The cathode lead is the shorter lead.

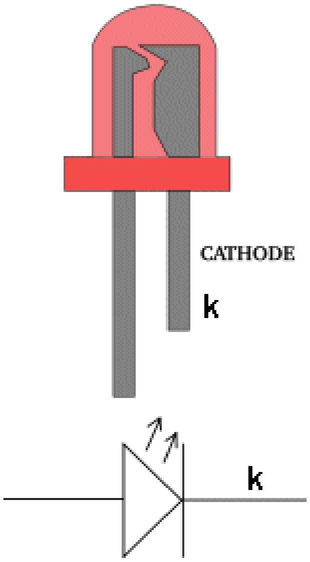

One of the most important things to remember about a LED is the characteristic voltage that appears across it when connected to a voltage. This does not change with brightness and cannot be altered.

For a red LED, this voltage is 1.7v and if you supply it with more than this voltage, it will be damaged.

The easy solution is to place a resistor on one lead as shown in the diagram below:

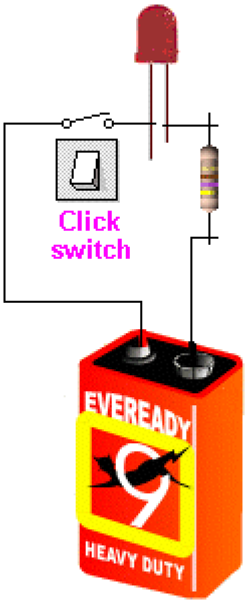

The LED will allow the exact voltage to appear across it and the brightness will depend on the value of the resistor.

Zener diodes (5.2c and 5.2d) are designed to stabilize a voltage. Diodes marked as ZPD5.6V or ZPY15V have operating voltages of 5.6V and 15V.

Photo diodes (5.2e) are constructed in a way that they allow light to fall on the P-N connection. When there is no light, a photo diode acts as a regular diode. It has high resistance in one direction, and low resistance in opposite direction. When there is light, both resistances are low. Photo diodes and LEDs are the main items in an optocoupler (to be discussed in more detail in chapter 9).

Tunnel diodes (5.2f and 5.2g) are commonly used in oscillators for very high frequencies.

Schottky diodes (5.2h) are used in high frequency circuits and for its low voltage drop in the forward direction.

Breakdown diodes (5.2i) are actually Zener diodes. They are used in various devices for protection and voltage regulation. It passes current only when voltage rises above a pre-defined value.

A Varicap diode (5.2j) is used instead of a variable capacitor in high frequency circuits. When the voltage across it is changed, the capacitance between cathode and anode is changed. This diode is commonly used in radio receivers, transceivers and oscillators.

The cathode of a low power diode is marked with a ring painted on the case, but it is worth noting that some manufacturers label the anode this way, so it is best to test it with a multimeter.

Power diodes are marked with a symbol engraved on the housing. If a diode is housed in a metal package, the case is generally the cathode and anode is the lead coming from the housing.

5.1 Diode identification

European diodes are marked using two or three letters and a number. The first letter is used to identify the material used in manufacturing the component (A - germanium, B - silicon), or, in case of letter Z, a Zener diode.
The second and third letters specify the type and usage of the diode. Some of the varities are:
A - low power diode, like the AA111, AA113, AA121, etc. - they are used in the detector of a radio receiver; BA124, BA125 : varicap diodes used instead of variable capacitors in receiving devices, oscillators, etc., BAY80, BAY93, etc. - switching diodes used in devices using logic circuits. BA157, BA158, etc. - these are switching diodes with short recovery time.
B - two capacitive (varicap) diodes in the same housing, like BB104, BB105, etc.
Y - regulation diodes, like BY240, BY243, BY244, etc. - these regulation diodes come in a plastic packaging and operate on a maximum current of 0.8A. If there is another Y, the diode is intended for higher current. For example, BYY44 is a diode whose absolute maximum current rating is 1A. When Y is the second letter in a Zener diode mark (ZY10, ZY30, etc.) it means it is intended for higher current.
G, G, PD - different tolerance marks for Zener diodes. Some of these are ZF12 (5% tolerance),

ZG18 (10% tolerance), ZPD9.1 (5% tolerance).
The third letter is used to specify a property (high current, for example).
American markings begin with 1N followed by a number, 1N4001, for example (regulating diode), 1N4449 (switching diode), etc.
Japanese style is similar to American, the main difference is that instead of N there is S, 1S241 being one of them.

5.2 Diode characteristics

The most important characteristics when using power diodes is the maximum current in the forward direction (IFmax), and maximum voltage in the reverse direction (URmax).

The important characteristics for a Zener diode are Zener voltage (UZ), Zener current (IZ) and maximum dissipation power (PD).

When working with capacitive diodes it is important to know their maximum and minimum capacitance, as well as values of DC voltage during which these capacitances occur.

With LEDs it is important to know the maximum value of current it is capable of passing. The natural characteristic voltage across a LED depends on the colour and starts at 1.7V for red to more than 2.4v for green and blue.
Current starts at 1mA for a very small glow and goes to about 40mA. High brightness LEDs and "power LEDs" require up to 1 amp and more. You must know the exact current required by the LED you are using as the wrong dropper resistor will allow too much current to flow and the LED will be damaged **instantly**.
The value of this resistors will be covered in another chapter.

Beside universal transistors TUN and TUP (mentioned in Chapter 4.4), there are universal diodes as well. They are marked with DUS (for universal silicon diode) and DUG (for germanium) on circuit diagrams.

DUS = Diode Universal Silicon   **DUG = Diode Universal Germanium**

5.3 Practical examples

The diagram of a power supply in figure (3.8) uses several diodes. The first four are in a single package, identified by B40C1500. This is a bridge rectifier.
The LED in the circuit indicates the transformer is working. Resistor R1 is used to limit the current through the LED and the brightness of the LED indicates the approximate voltage.
Diodes marked 1N4002 protect the integrated circuit.

Figure 5.3 below shows some other examples of diodes. The life of a globe can be increased by adding a diode as shown in 5.3a. By simply connecting it in series, the current passing through the globe is halved and it lasts a lot longer. However the brightness is reduced and the light becomes yellow. The Diode should have a reverse voltage of over 400V, and a current higher than the globe. A 1N4004 or BY244 is suitable.

A very simple DC voltage stabilizer for low currents can be made using 5.3c as a reference.

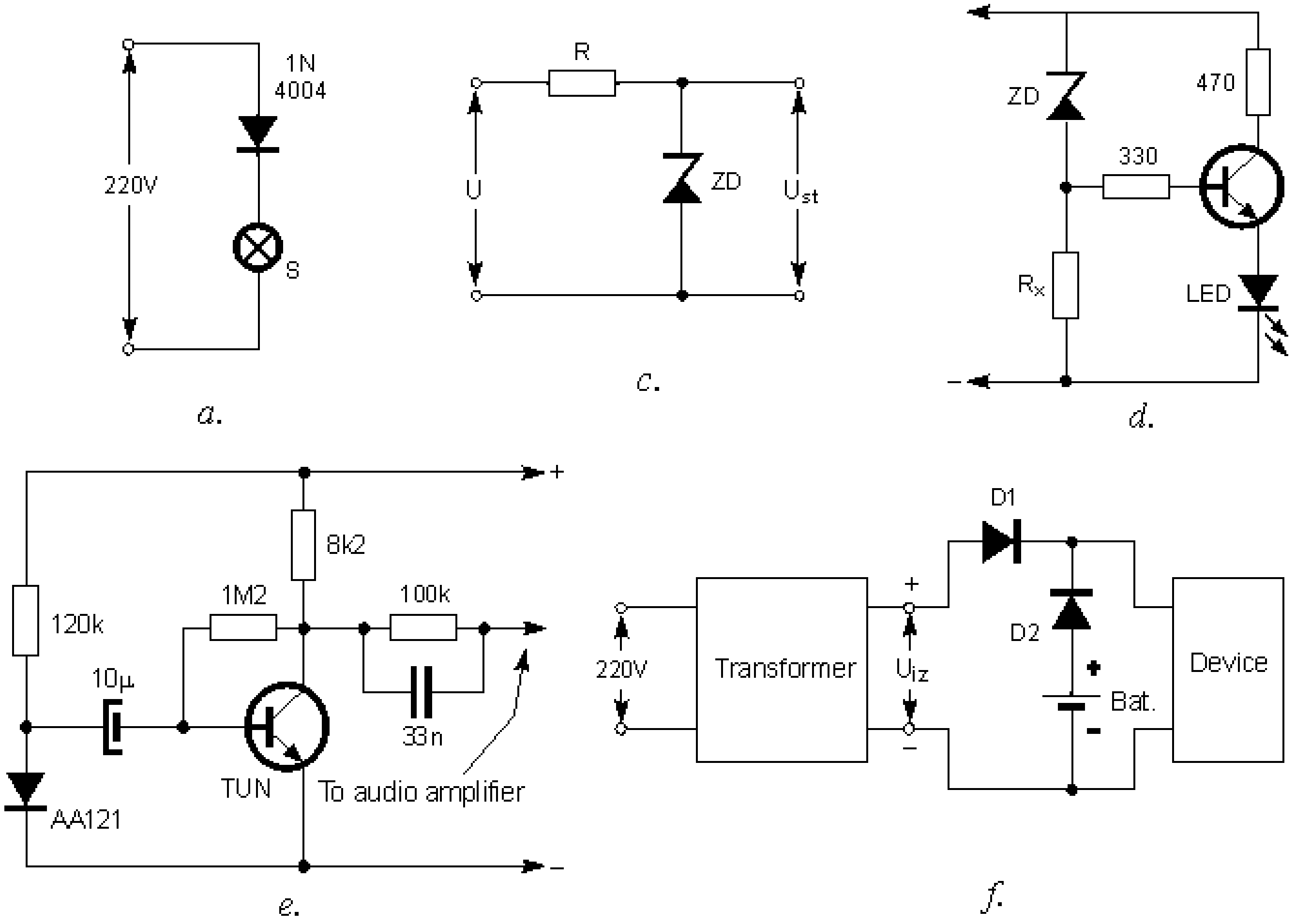

Fig. 5.3: a - using a diode to prolong the light bulb's life span, b - stair-light LED indicator, c - voltage stabilizer, d - voltage rise indicator, e - rain noise synthesizer, f - backup supply

Unstabilized voltage is marked "U", and stabilized with "UST." Voltage on the Zener diode is equal to UST, so if we want to achieve a stabilized 9V, we would use a ZPD9.1 diode. Although this stabilizer has limited use it is the basis of all designs found in power supplies.
We can also devise a voltage overload detector as sown in figure 5.3d. A LED indicates when a voltage is over a predefined value. When the voltage is lower than the operating voltage of the Zener, the zener acts as a high value resistor, so DC voltage on the base of the transistor is very low, and the transistor does not "turn on." When the voltage rises to equal the Zener voltage, its resistance is lowered, and transistor receives current on its base and it turns on to illuminate the LED. This example uses a 6V Zener diode, which means that the LED is illuminated when the voltage reaches that value. For other voltage values, different Zener diodes should be used.

Brightness and the exact moment of illuminating the LED can be set with the value of Rx.
To modify this circuit so that it signals when a voltage drops below some predefined level, the Zener diode and Rx are swapped. For example, by using a 12V Zener diode, we can make a car battery level indicator. So, when the voltage drops below 12V, the battery is ready for recharge.
Figure 5.3e shows a noise-producing circuit, which produces a rain-like sound. DC current flowing through diode AA121 isn't absolutely constant and this creates the noise which is amplified by the transistor (any NPN transistor) and passed to a filter (resistor-capacitor circuit with values 33nF and 100k).

Figure 5.3f shows a battery back-up circuit. When the "supply" fails, the battery takes over.

6. Thyristors, triacs, diacs

There are several thyristors displayed on 6.1. Triacs look the same, while diacs look like small power rectifying diodes. Their symbols, and pin-out is found in figure 6.2.

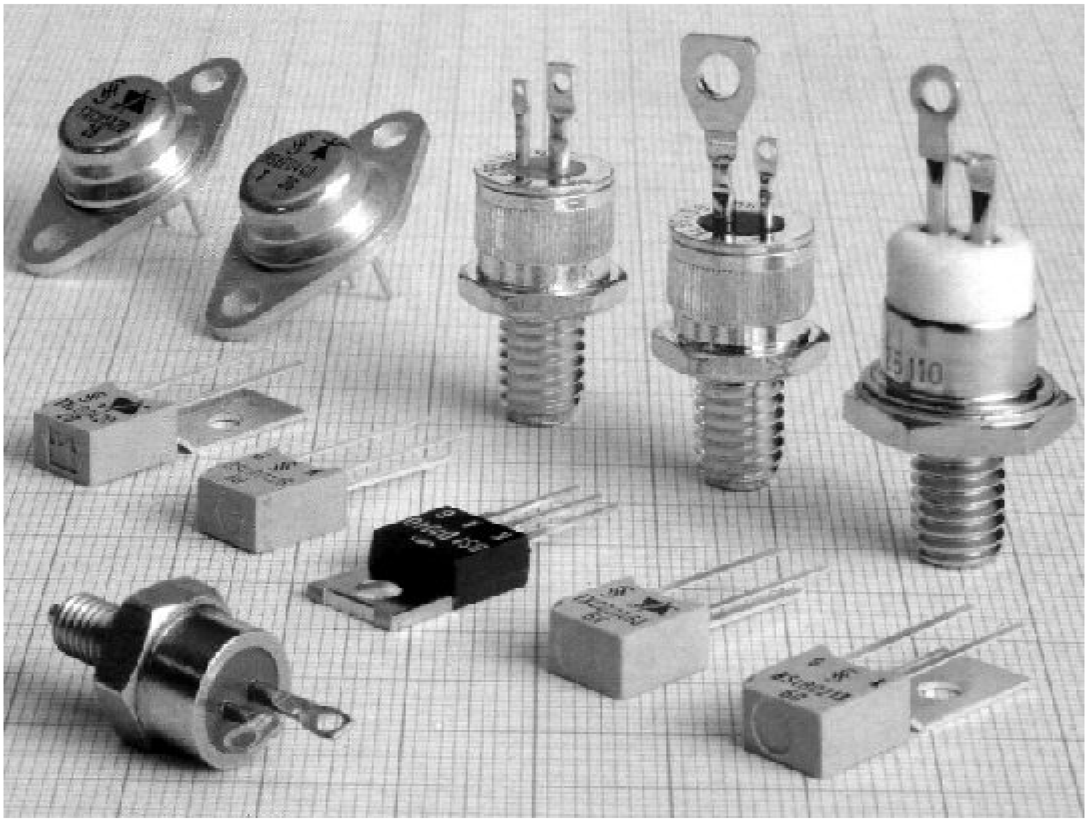

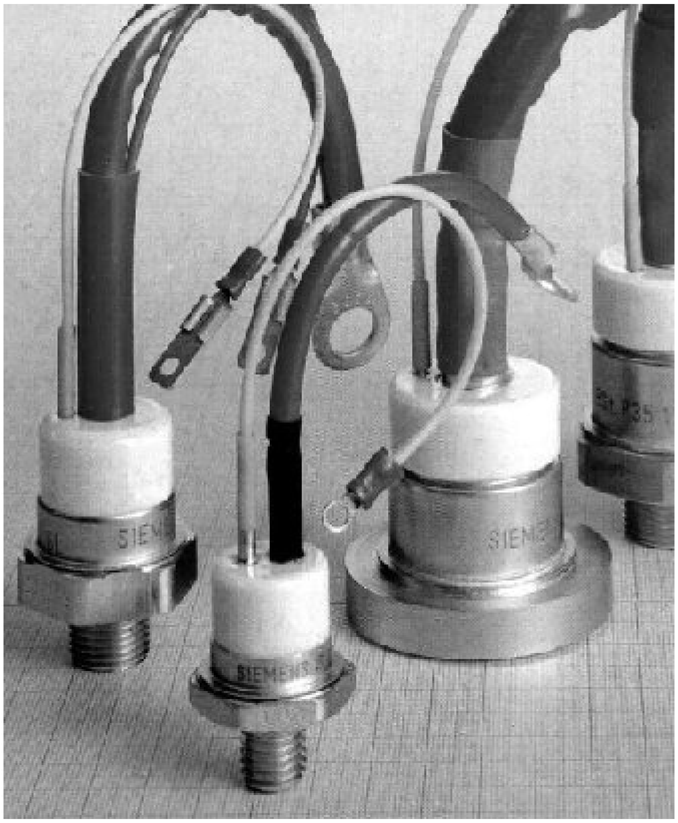

Fig. 6.1: Several thyristors and triacs

A thyristor is an improved diode. Besides anode (A) and cathode (k) it has another lead which is commonly described as a gate (G), as found on picture 6.2a. The same way a diode does, a thyristor conducts current when the anode is positive compared to the cathode, but only if the voltage on the gate is positive and sufficient current is flowing into the gate to turn on the device. When a thyristor starts conducting current into the gate is of no importance and thyristor can only be switched off by removing the current between anode and cathode. For example, see figure 6.3. If S1 is closed, the thyristor will not conduct, and the globe will not light. If S2 is closed for a very short time, the globe will illuminate. To turn off the globe, S1 must be opened. Thyristors are marked in some circuits as SCR, which is an acronym for Silicon Controlled Rectifier.

A triac is very similar to a thyristor, with the difference that it can conduct in both directions. It has three electrodes, called anode 1 (A1), anode 2 (A2), and gate (G). It is used for regulation of alternating current circuits. Devices such as hand drills or globes can be controlled with a triac.

Thyristors and triacs are marked alphanumerically, KT430, for example.

Low power thyristors and triacs are packed in same housings as transistors, but high power devices have a completely different housing. These are shown in figure 6.1. Pin-outs of some common thyristors and triacs are shown in 6.2 a and b.

Diacs (6.2c), or two-way diodes as they are often referred to, are used together with thyristors and triacs. Their main property is that their resistance is very large until voltage on their ends exceeds some predefined value. When the voltage is under this value, a diac responds as a large value resistor, and when voltage rises it acts as a low value resistor.

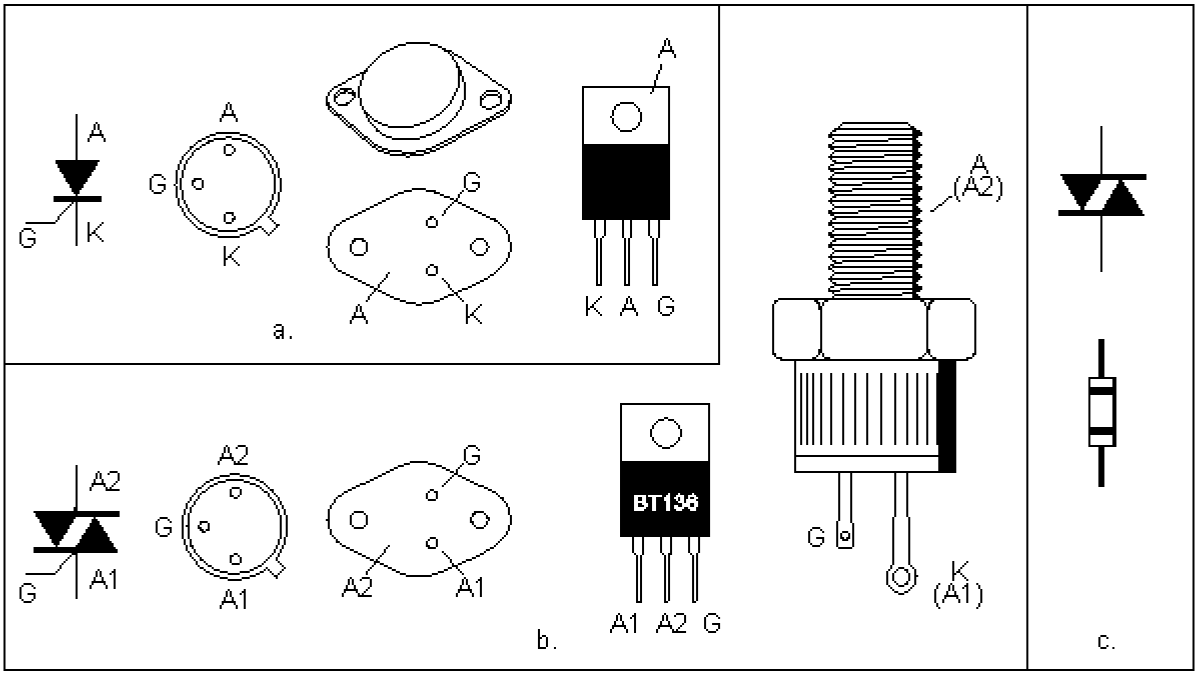
A
G
K
A
G
K
G
A
K
A
K A G
a.
A2
G
A1
A2
G
A1
G
A2
A1
BT136
A1 A2 G
b.
A
(A2)
G
K
(A1)
c.

Fig. 6.2: Symbols and pin placements for: a - thyristor, b - triac, c - diac

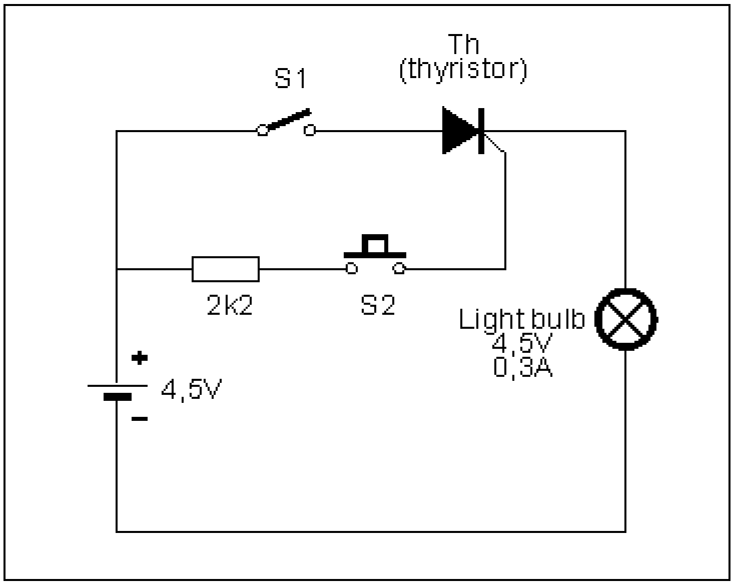
S1
Th
(thyristor)
2k2
S2
Light bulb
4,5V
0,3A
+
4,5V
-

Fig. 6.3: Thyristor principle of work

6.1 Practical examples

Picture 6.5 detects when light is present in a room. With no light, the photo-transistor does not conduct. When light is present, the photo-transistor conducts and the bell is activated. Turning off the light will not stop the alarm. The alarm is turned off via S1.

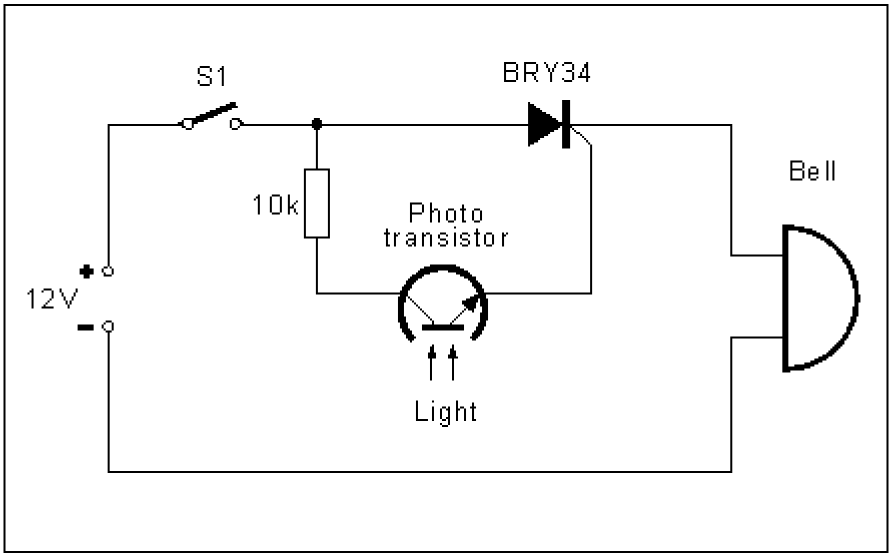

Fig. 6.5: Alarm device using a thyristor and a photo-transistor

A circuit to flash a globe is shown in figure 6.6 This circuit flashes a 40w globe several times per second. Mains voltage is regulated using the 1N4004 diode. The 220u capacitor charges and its voltage rises. When this voltage reaches the design-voltage of the the diac (20v), the capacitor discharges through the diac and into the triac. This switches the triac on and lights the bulb for a very short period of time, after a period of time (set by the 100k pot), the capacitor is charged again, and the whole cycle repeats. The 1k trim pot sets the current level which is needed to trigger the triac.

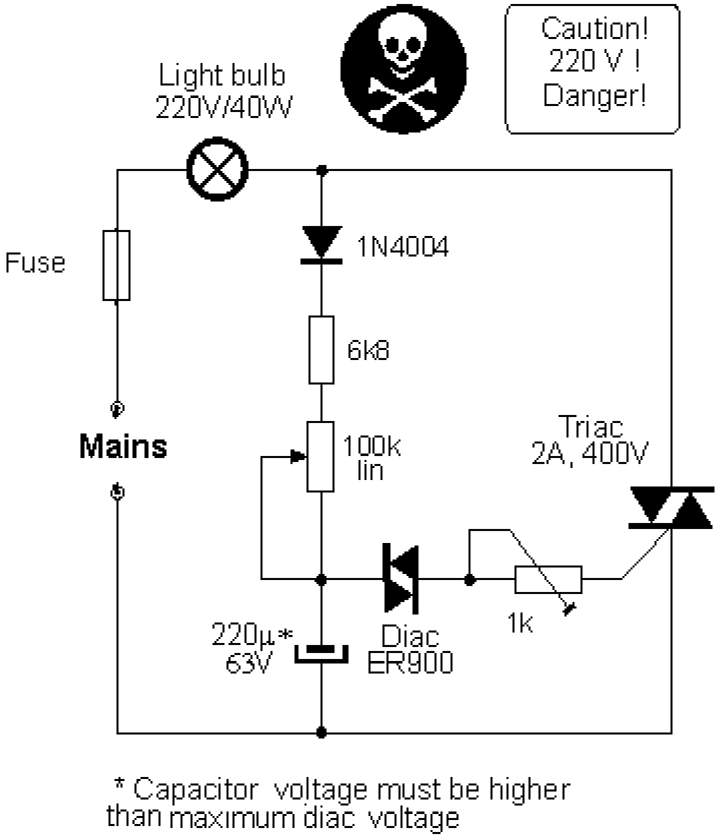
Light bulb
220V/40W
Caution!
220 V !
Danger!
Fuse
1N4004
6k8
Mains
100k
lin
Triac
2A, 400V
220µ*
63V
Diac
ER900
1k
* Capacitor voltage must be higher
than maximum diac voltage

Fig. 6.6: Flasher

A circuit to control the brightness of a globe or the speed of a motor is shown in figure 6.7

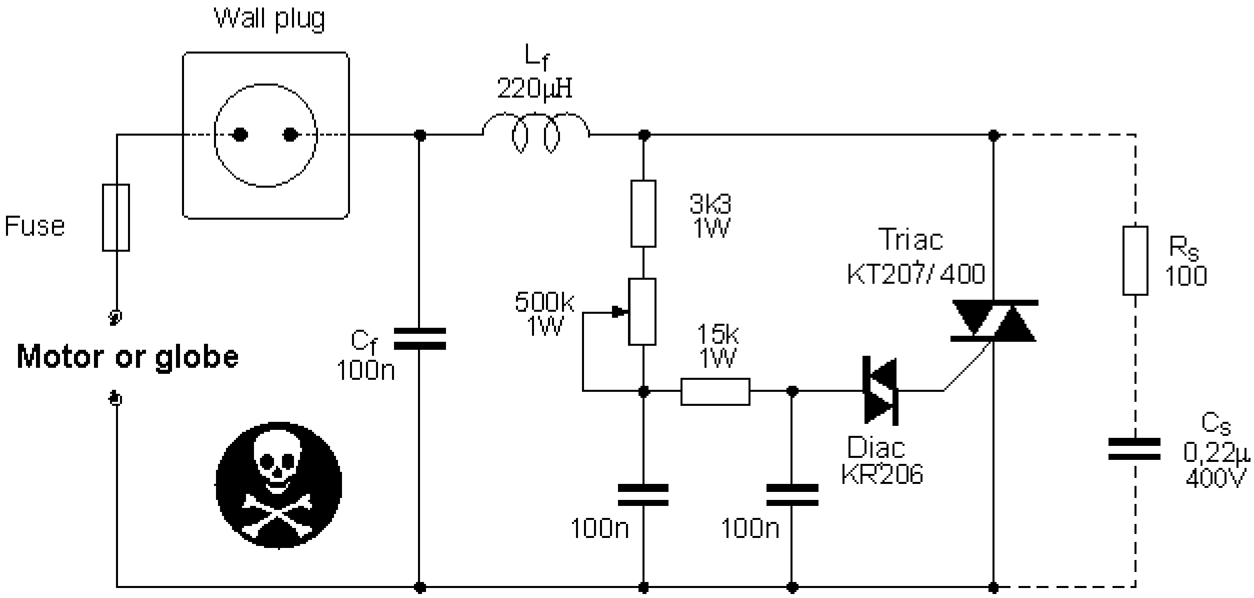
Wall plug
L_f
220µH
Fuse
3k3
1W
Triac
KT207/400
R_S
100
500k
1W
15k
1W
Motor or globe
C_f
100n
Diac
KR206
C_S
0,22µ
400V
100n
100n

Fig. 6.7: Light bulb intensity or motor speed controller

If the main use for this circuit is to control the brightness of a light bulb, RS and CS are not necessary.

7. Integrated circuits

Integrated Circuits play a very important part in electronics. Most are specially made for a specific task and contain up to thousands of transistors, diodes and resistors. Special purposes IC's such as audio-amplifiers, FM radios, logic blocks, regulators and even a whole micro computers in the form of a micro controller can be fitted inside a tiny package. Some of the simple Integrated Circuits are shown in figure 7.1.

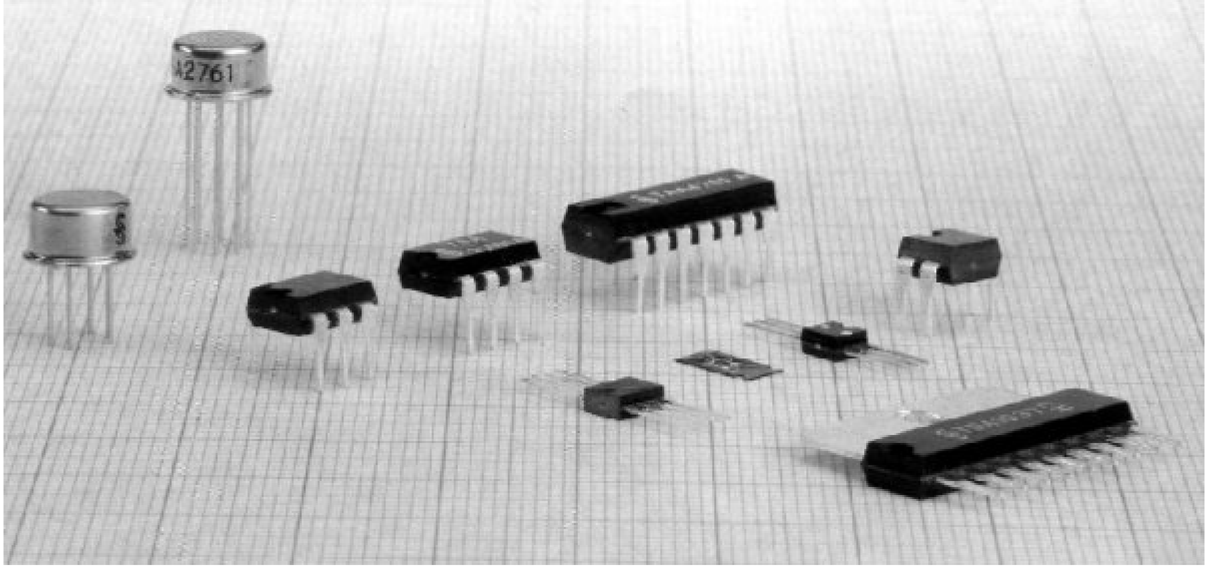

Fig. 7.1: Integrated circuits

Depending on the way they are manufactured, integrated circuits can be divided into two groups: hybrid and monolithic. Hybrid circuits have been around longer. If a transistor is opened, the crystal inside is very small. This means a transistor doesn't take up very much space and many of them can be fitted into a single Integrated Circuit.

The pin-out for some of the common packages is shown in Figure 7.2:

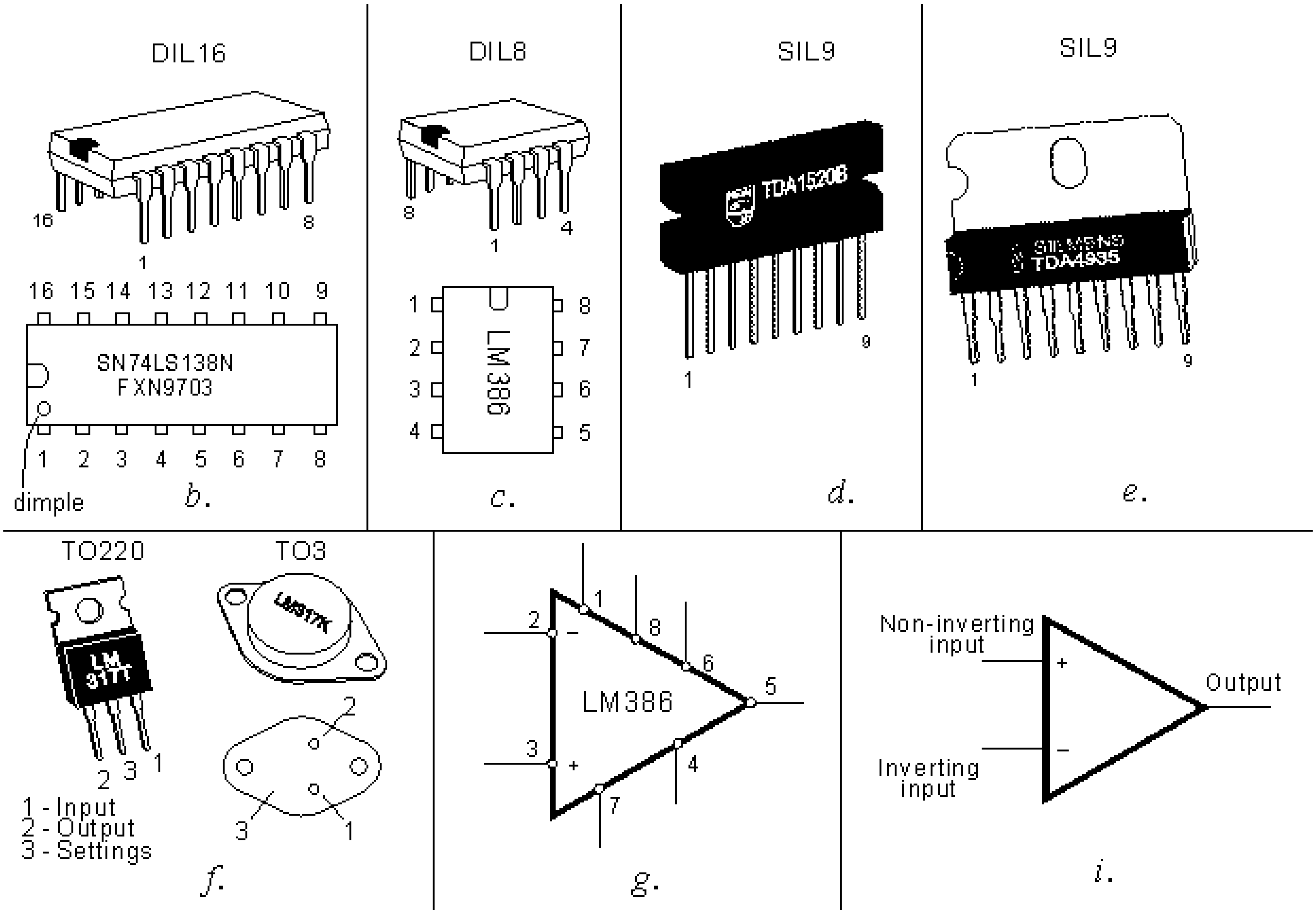
DIL16
16
8
1
16 15 14 13 12 11 10 9
SN74LS138N
FXN9703
1 2 3 4 5 6 7 8
dimple
b.
DIL8
8
1
4
1 2 3 4
LM386
8 7 6 5
c.
SIL9
TDA1520B
1
9
d.
SIL9
SIEMENS
TDA4935
1
9
e.
TO220
LM 317T
2 3 1
1 - Input
2 - Output
3 - Settings
TO3
LM317K
2
3
1
f.
LM386
1
8
6
2
5
3
4
7
g.
Non-inverting input
Output
Inverting input
i.

Fig. 7.2: Pin-out and symbols for some common integrated circuits

Most integrated Circuits are in a DIL package - Dual In Line, meaning there are two rows of pins. (DIL16 and DIL8 are shown in 7.2b and 7.2c). The device is viewed from the top and the pins are numbered in an anti-clockwise direction.

High power integrated circuits can generate a lot of heat and they have a metal tag that can be connected to a heatsink to dissipate the heat. Examples of these IC's are shown in 7.2d and 7.2e, and 7.2f.

Symbols used to represent integrated circuits are shown in 7.2g and 7.2i. Symbol 7.2g is commonly used to represent amplifiers.

Figure 7.2i shows an operational amplifier. Signs + and - represent inverting and non-inverting inputs. The signal to be amplified is applied between one of the inputs and ground (ground and supply aren't represented, but are necessary for the circuit to operate).

Integrated circuits can be divided into two further groups: **analog (linear) and digital**. The output voltage of a linear circuits is continuous, and follows changes in the input. Typical representative of a linear IC is an integrated audio amplifier. When a signal from a microphone is connected to the input the output will vary in the same way as the voltage from the microphone. If watched on an oscilloscope, the signal on the output will be the same shape as the mic's signal,

only the voltage will be higher depending on the amplification of the integrated circuit.
It is a different situation with digital IC's. Their output voltage is not continuous. It is either LOW or HIGH and it changes from one state to the other very quickly.

7.1 Analog integrated circuits

While on the topic of analog circuits, we will look at the LM386 IC. It has all the components for a complete audio-amplifier. Figure 7.3a shows an example of an amplifier made with this integrated circuit, which can be used as a complete amplifier for a walkman, interphone, cassette player or some other audio device. It can also be used as a test circuit for troubleshooting.

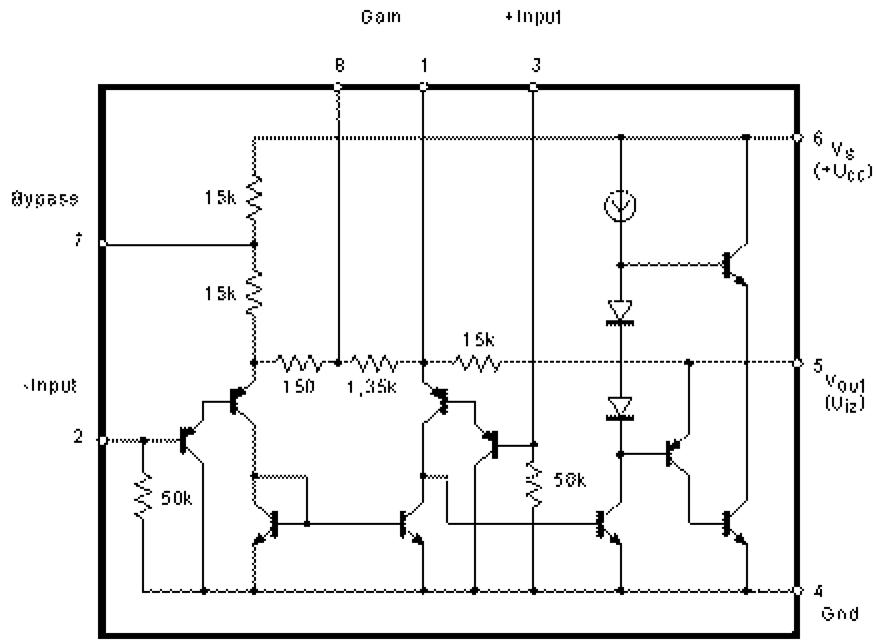
Gain
+Input
8
1
3
6 Vs
(+Ucc)
Bypass
7
15k
15k
15k
150
1,35k
-Input
2
50k
50k
5 Vout
(Viz)
4
Gnd

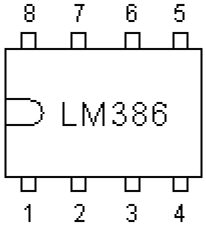
8 7 6 5
LM386
1 2 3 4

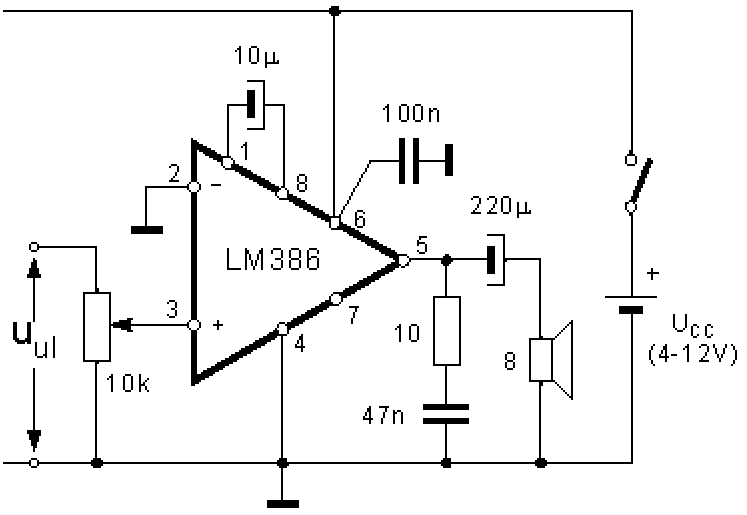
10µ
100n
220µ
1
2
8
6
5
3
7
4
−
+
LM386
u_{ul}
10k
10
8
47n
+
Ucc
(4-12V)

Fig. 7.3: a - A low frequency amplifier using the LM386

The signal is brought to the non-inverting input (between pin 3 and ground). Inverting input (pin 2) is connected to ground. If 10µF is placed between pins 1 and 8 a voltage amplification of 200 is created. If this capacitor is removed the amplification is 20. It is possible to achieve in-between amplification by adding a resistor and connecting it in series with the capacitor.

One of the essential components in this circuit is the 100nF capacitor which is placed between pin 6 (which is connected to the positive of the supply) and ground. The capacitor should be ceramic and should be mounted as close to the integrated circuit as possible. This is common practice when working with integrated circuits, even when it isn't shown in the diagram as a capacitor connected between the positive and negative stabilizes the voltage and protects the circuit from spikes and a phenomenon called instability. This is due to inductance in the power supply tracks allowing high currents taken by the IC to upset its operation.

7.2 Digital integrated circuits

The CD4011 will be our "show-and-tell" IC to cover the main characteristics of digital circuits. It is a 14 pin DIL package. The pin-out is shown in figure 7.4a. Note the small half-round slit on one end of the IC. It identifies pin 1. Pins 7 and 14 connect to a supply (battery or DC power supply). Negative is connected to pin 7. Positive is connected to pin 14.
There are four logic NAND gates in a CD4011 IC. Each has two inputs and one output. For gate N1 the inputs are pins 1 and 2, and output is pin 3. The symbol for a NAND gate is displayed in figure 7.4b. The inputs are marked A and B and output is F. The supply voltage can be up to 16v and as low as 5V. The output will deliver up to 10mA at 12v but this is reduced as the supply voltage is reduced.

Figure 7.4b shows the truth table for a NAND gate. It shows the output voltage (voltage between F and ground) with different input states. Because there are only two voltages for every pin, we call them states, with logic zero when the voltage is zero, and logic one when the voltage on the pin is the same as the supply voltage.
From this we can read the second row of the truth table: if logic zero is on both input pins, output is logic one, third row is similar: if the first input is one, and the second one is zero, output is logic one, fourth row: if the first input is zero, and the second one is one, output is logic one. Fifth row is different, since both of its inputs are one, the definition of NAND gate states that the output is zero.

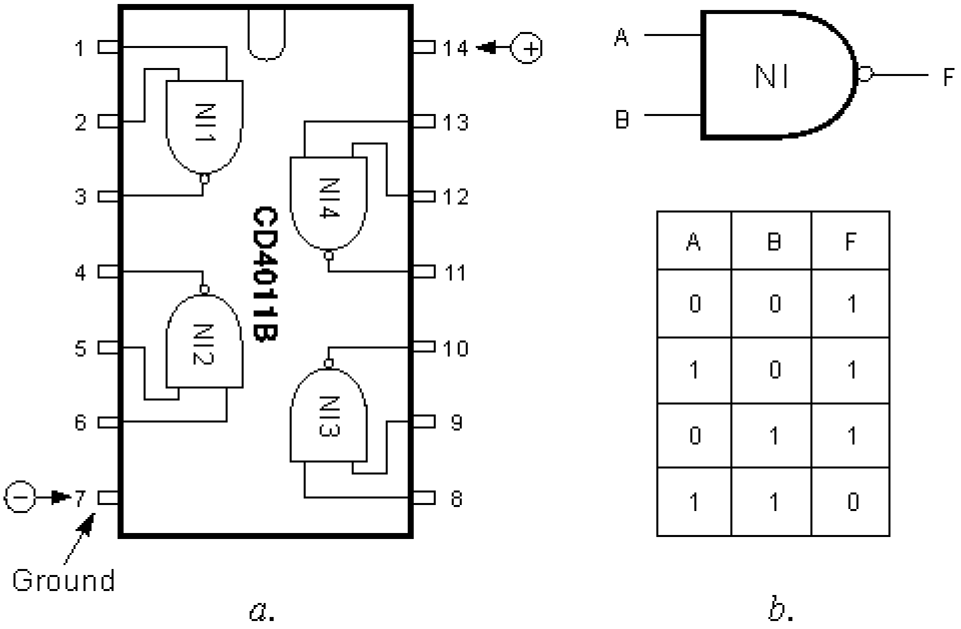
1
2
3
4
5
6
7
8
9
10
11
12
13
14
+
−
NI1
NI2
NI3
NI4
CD4011B
Ground
a.
A
B
NI
F
b.

A	B	F
0	0	1
1	0	1
0	1	1
1	1	0

Fig. 7.4: a - 4011 pin placements, b - symbol and the truth table for NAND gates,

Logic circuits have many applications, but their main use is in computer circuits.
The following circuit is a simple example to show how the gates can be connected to produce a project that turns on a globe when a finger is placed on a "touch pad."
The globe turns off after a period of time, determined by the value of the 470u and 2M2 resistor.

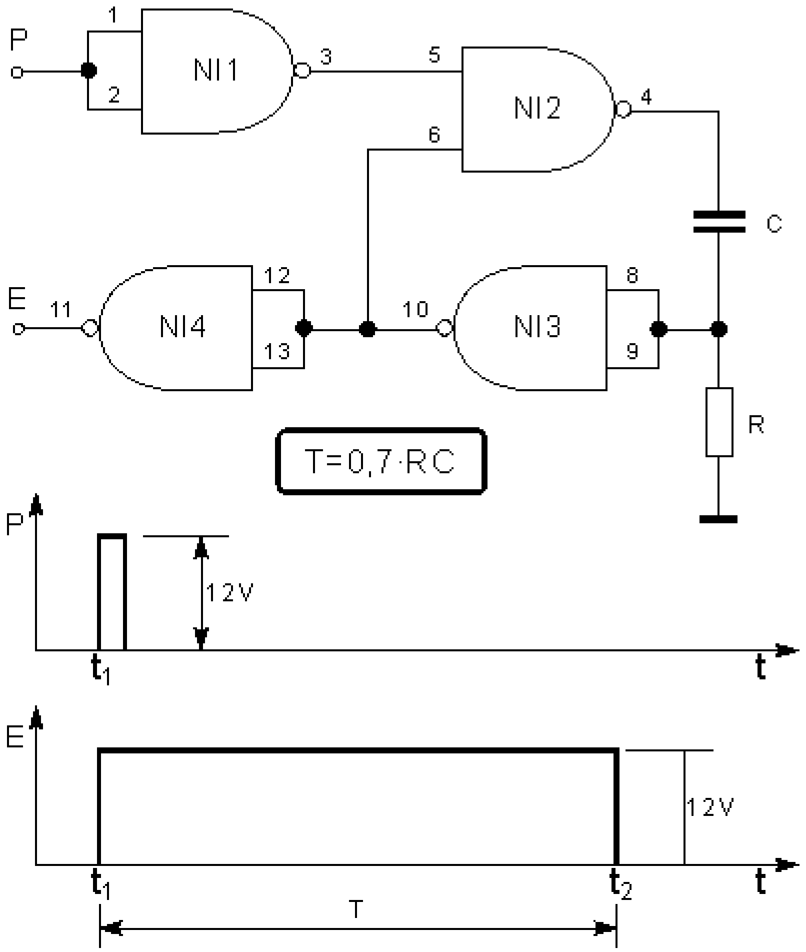
P
1
2
NI1
3
5
NI2
6
4
C
E
11
NI4
12
13
10
NI3
8
9
R
T=0,7·RC
P
12V
t_1
t
E
12V
t_1
t_2
t
T

a.

The operation of a NAND gate

Lets look at the functionalities of the following circuit. Both inputs of NI1 are connected to each other, so when input P is HIGH, output is zero. This logic zero is passed on to NI2, so no matter what is on the input 6, output 4 is logic one. This means that, between the ground and pin 4, the voltage is equal to 12V.

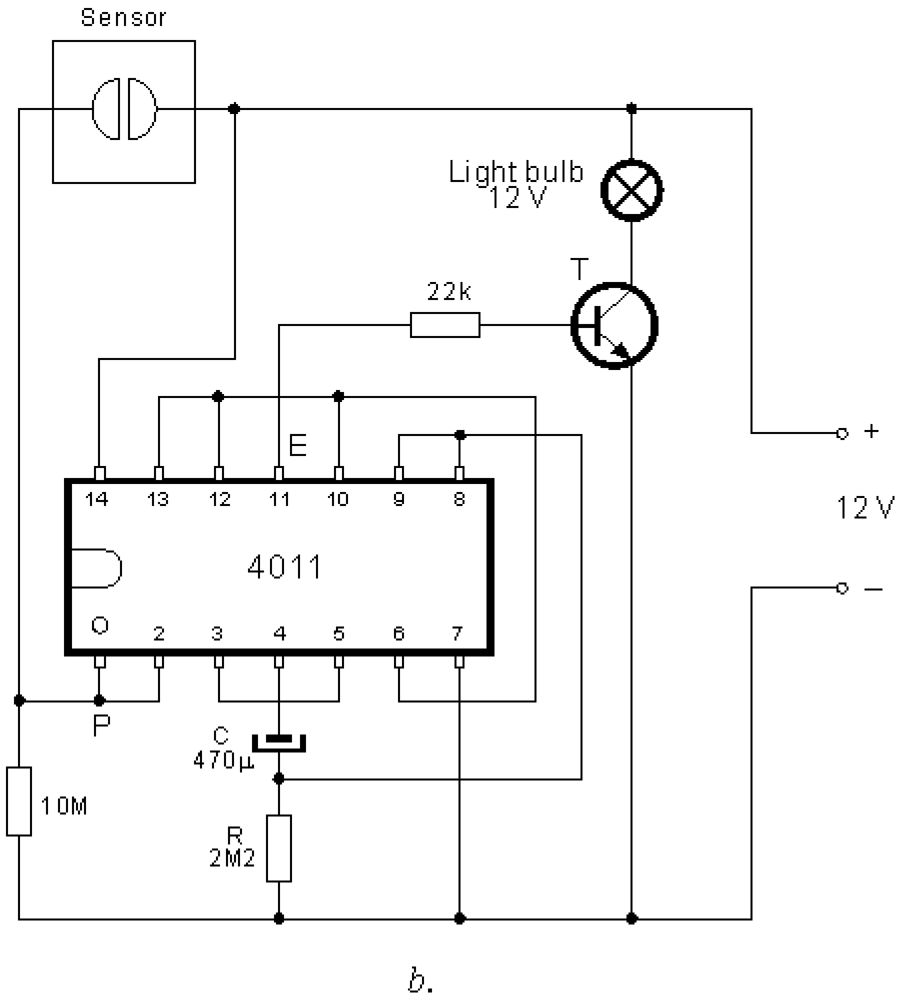

Fig. 7.5: Sensor switch using a 4011

Current flows through capacitor C and resistor R, so capacitor begins to charge. Every uncharged capacitor behaves like a short circuit. Because of that, when 12V appears on pin 4, it is also present on resistor R and also on pins 8 and 9. Pin 10 shows logic zero because of this which is connected to pin 6. From now on, logic zero on pin 5 is no longer needed because only one input needs to be zero for the output to be logic one. So input P is no longer needed. Gates NI2 and NI3 maintain logic zero on pin 4. How long will this last? It depends on the value of the capacitor and resistor. As the capacitor charges, the voltage on the resistor drops and when it falls to 1/2 of the supply voltage (6V in our case), NI3 detects a low on its inputs so logic one appears on pin 10. Since logic one is now on input 5 (no logic one present on P), and on input 6, output 4 is zero, capacitor dumps its charge via diodes on the inputs on pins 8 and 9 and the circuit starts operating again.

As we saw, for a certain period of time, which is equal to T=0.7*RC output of pin 10 was logic zero. During that time output E (pin 11) is logic one. For example, if R = 2M2 and C=47µF, for time T = 2.2*10^6*47*10^-6 = 94 sec from the moment impulse on input P subsided, voltage on

output E is 12V.
The end result of our experiment is on diagram 7.5a. Short positive pulses appearing on P in the time t1 caused a longer variable ulse on output E.
Schematic 7.5b displays this circuit which allows us to light a bulb using four NAND gates interconnected in the way shown on picture 7.5a.
The sensor is two copper (or some other conducting material) plates glued to some non-conducting material (plastics, wood, etc.) in close proximity to each other. So, when we touch the sensor with the tip of our finger, we close the circuit. 12V appears on input P, which in turn conducts the voltage to the output E, resistor R = 22k conducts base current and the bulb lights. When we remove our finger, output E will last for 94 seconds, after which it goes to logic zero and the light goes out.
Transistor T is selected so that its maximum allowed collector current is higher than the current of the globe.
(The globes current flow value is found by dividing its power by its voltage. For example, if its power is P = 6W and voltage is U = 12V, current through the globe is I = P/U = 6W/12V = 0.5A or higher. One thing you must remember with a globe is the starting or "turn-on" current. It is about six times the operating current and the transistor must be able to pass this current for the globe to illuminate.

7.3 Practical examples

Diagram 7.6 shows a circuit for a stereo audio-amp using a TDA4935 IC. It is a modern integrated circuit with two separate pre-amps and stereo outputs.
Left and right input signals are marked UL and UD, which are brought to two inputs of the amplifier. The chip also has built-in heat and overload protection. Maximum output for each amplifier is 15W, so they can be used in stereo mode of 2x15W amplifiers.

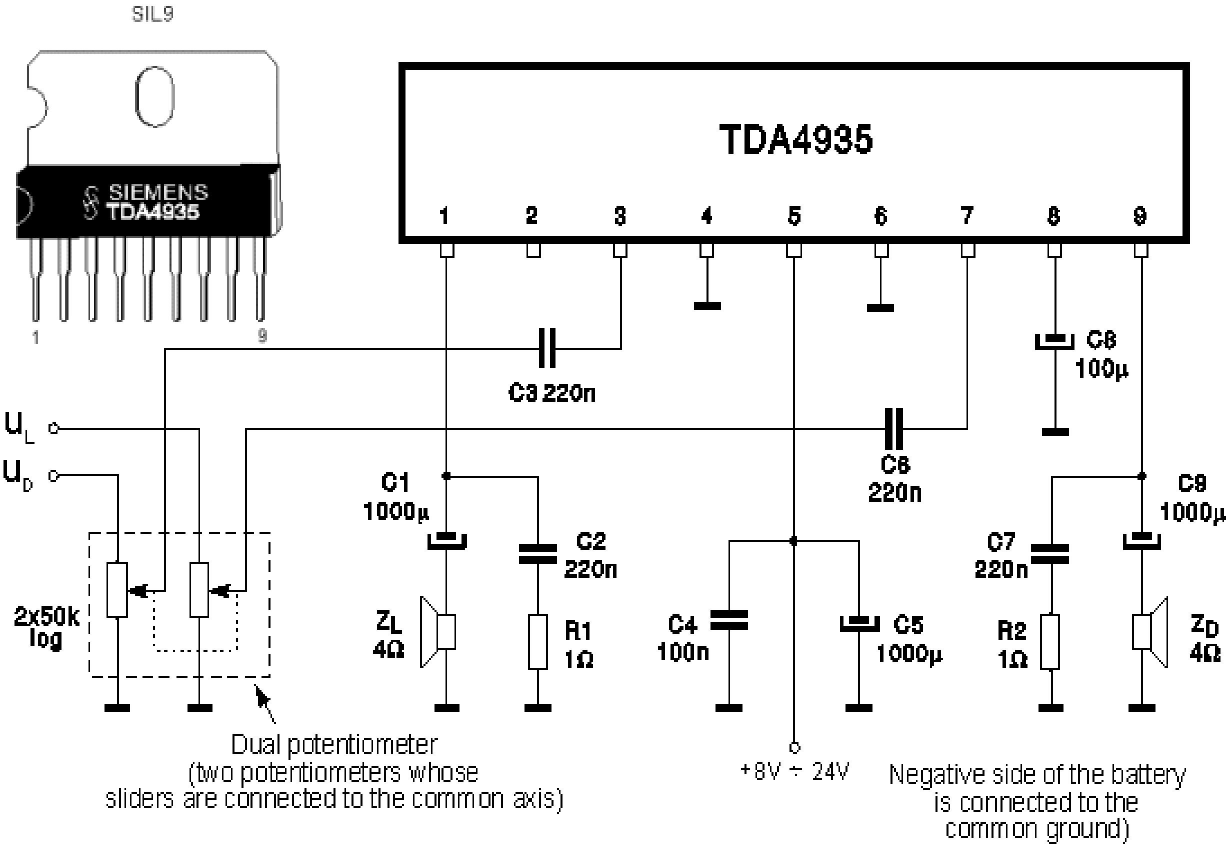
SIL9
SIEMENS
TDA4935
1
9
TDA4935
1
2
3
4
5
6
7
8
9
C6
100µ
C3 220n
U_L
U_D
C8
220n
C1
1000µ
C2
220n
C9
1000µ
C7
220n
2x50k
log
Z_L
4Ω
R1
1Ω
C4
100n
C5
1000µ
R2
1Ω
Z_D
4Ω
Dual potentiometer
(two potentiometers whose
sliders are connected to the common axis)
+8V ÷ 24V
Negative side of the battery
is connected to the
common ground)

Fig. 7.6: Stereo audio-amplifier using the TDA4935

Another example is an audio amplifier using an LM386 circuit, with a preamp using a BC107 transistor. The series connected capacitor and resistor between pins 1 and 5 produces low frequency amplification (around 100Hz) improving the characteristics of the circuit. This amplifier could be used with any low frequency source (gramophone, microphone, etc.).

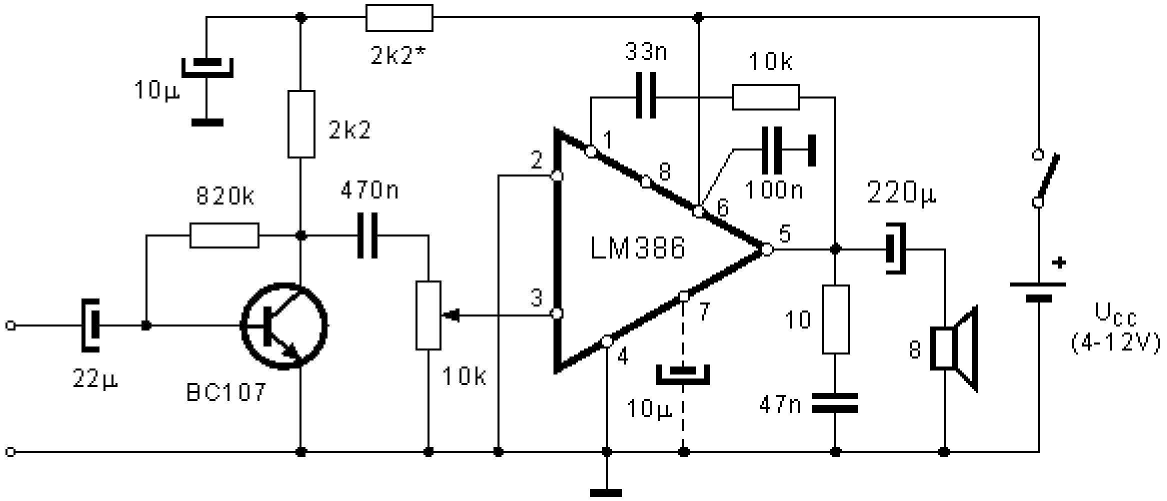
2k2*
10µ
2k2
33n
10k
820k
470n
100n
220µ
1
2
8
6
LM386
5
3
7
4
+
Ucc
(4-12V)
22µ
BC107
10k
10
8
10µ
47n

Fig. 7.7: complete audio-amplifier using the LM386

The third example is a simple alarm, shown in figure 7.8. It uses a CD4011 IC. Gates NI3 and NI4 form a 600Hz audio oscillator. This signal is amplified using an NPN transistor and passed to an 8R speaker. To hear the 600Hz tone, remove the connection to pin 8 and connect pin 8 to pin 9. This produces a constant tone. Gates NI1 and NI2 form a 4Hz oscillator, whose output is connected to pin 8. This turns the 600Hz tone on and off at 4Hz. To use this alarm in your home, on doors for example, connected pin 1 to 7 via a switch.

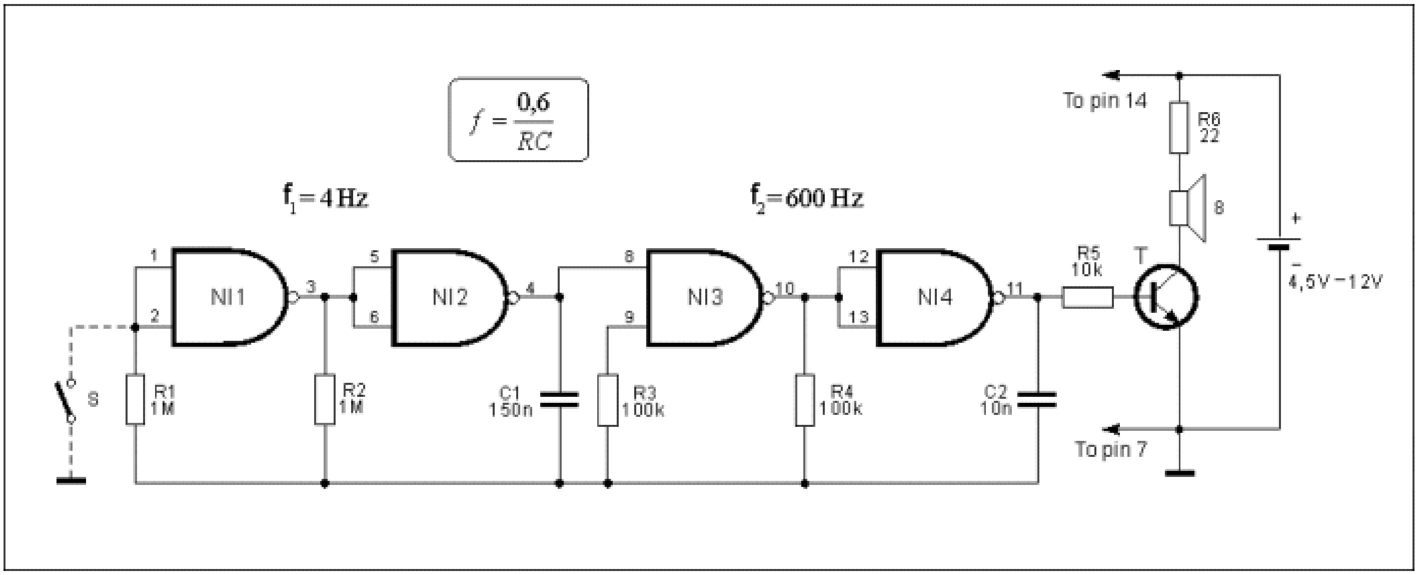
$f = \frac{0,6}{RC}$
$f_1 = 4$ Hz
$f_2 = 600$ Hz
To pin 14
R6
22
8
+
-
4,5V−12V
1
2
NI1
3
5
6
NI2
4
8
9
NI3
10
12
13
NI4
11
R5
10k
T
S
R1
1M
R2
1M
C1
150n
R3
100k
R4
100k
C2
10n
To pin 7

Fig. 7.8: Alarm using a CD4011 IC

The last circuit in this chapter is an example of a mono FM receiver using a TDA7088T IC, which can be, along with the SMD components, housed in a match-box along with two miniature watch batteries. You can purchase a ready-built scanning radio in a "junk shop" for as little as

$5.00 with stereo head-phones Always look in the toy sections of large stores for the latest technology at the cheapest price.

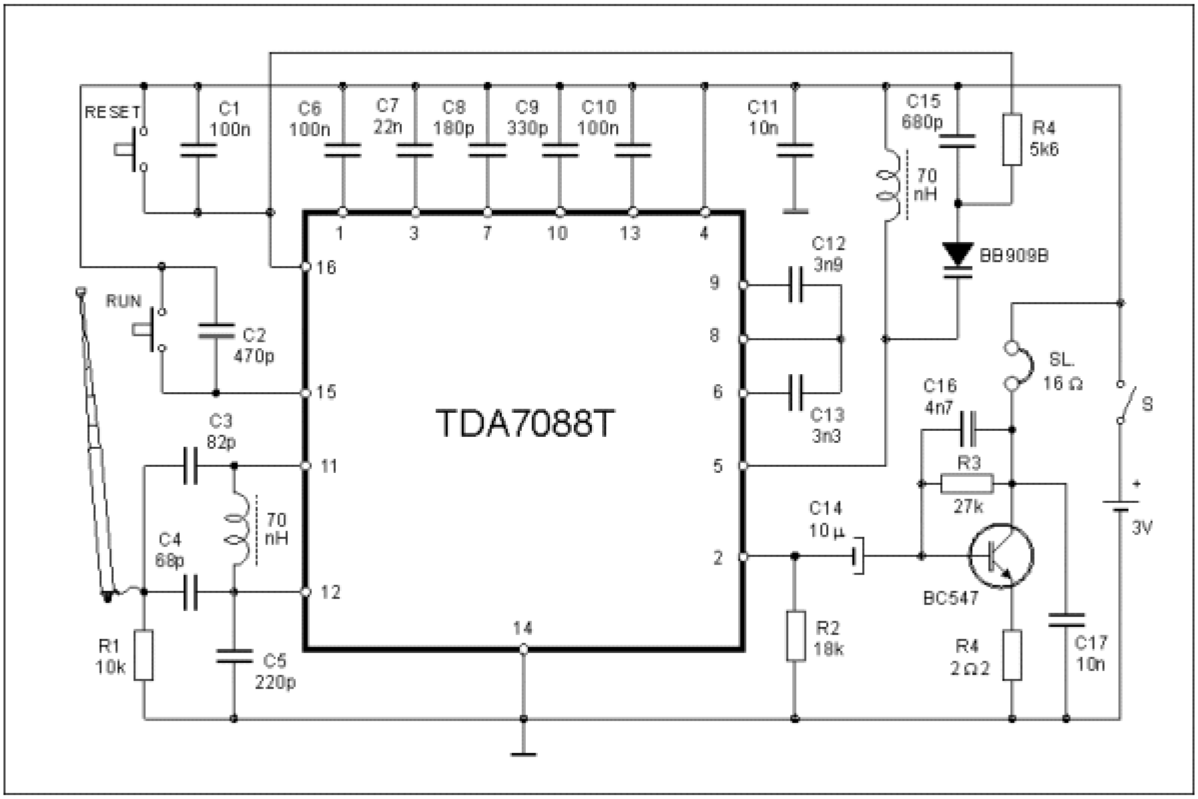
RESET
C1
100n
C6
100n
C7
22n
C8
180p
C9
330p
C10
100n
C11
10n
C15
680p
R4
5k6
70
nH
BB909B
1
3
7
10
13
4
16
9
C12
3n9
8
RUN
C2
470p
15
6
C13
3n3
SL.
16 Ω
S
C16
4n7
TDA7088T
C3
82p
11
5
R3
27k
+
3V
70
nH
C4
68p
C14
10 µ
2
12
BC547
14
R1
10k
C5
220p
R2
18k
R4
2Ω2
C17
10n

Fig. 7.9: Mono FM radio receiver with an electronic preset

Tuning to a low frequency station is done automatically by pressing the RUN button. This turns on the part of the integrated circuit which is designated for scanning over a given range. When it finds a station it locks on until the RUN button is pressed again. When it reaches 108MHz it waits for the RESET signal which returns the scan to 88MHz.

8. Microphones, speakers and headphones

Microphones, speakers and headphones are components commonly used as the input and output devices of many circuits. A microphone converts sound waves into electrical signals that closely follow the waveform of the sound being received. This signal is then amplified by the circuit and transformed into a sound by a speaker or headphone. The symbols for these components are shown on 8.1.

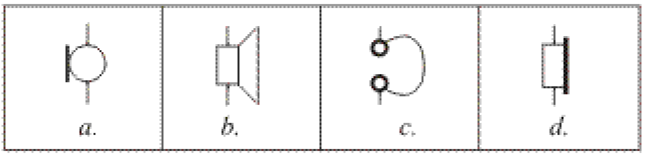

8.1. Symbols: a - microphone, b - speaker, c, d - headphones

8.1 Microphones

There are several different types of microphone: carbon, dynamic, crystal, capacitive (electret). Carbon microphones were one of the first to be invented and were used mainly in telephone applications. But they are very noisy as the carbon granules rattle when the microphone is moved and this type is being replaced by more advanced types.

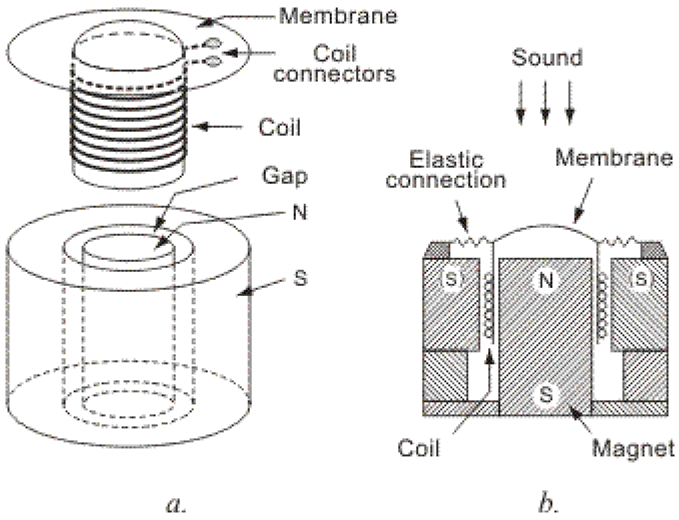

Sl. 8.2. Dynamic microphone: a - internal structure, b - cross section

Dynamic microphones are in wide use and their quality of reproduction is superb. They are used in the recording industry for music and speech where high fidelity is required. Basically they are exactly the same as a speaker, the only difference being the size. But their only limitation is the very low output. The internal structure is shown in figure 8.2. A paper cylinder, onto which fine copper wire is wound, is connected to a membrane which moves under the force of sound pressure created by the sound source. This coil is in a narrow gap with a high magnetic field created by a permanent magnet. When the coil moves in this magnetic field, it produces a voltage identical to the sound causing the movement.

Because of the low resistance (impedance) of a dynamic microphone, it usually needs a transformer so it can be connected to an amplifier (called a pre-amp). This transformer is usually built into the microphone's case, but if is absent, it is necessary to connect the microphone to a preamplifier with low input resistance.

Crystal microphones contain a crystal called a "piezo crystal" that is connected to a small diaphragm. When sound waves hit the diaphragm, the crystal changes shape and it produces a voltage. This voltage is passed to an amplifier.

Recently, electret microphones have improved in quality and taken over from nearly all other types of microphone. They are small, rugged, low in price and produce a very high quality output.
The shape, size and characteristics are shown in 8.3.
The microphone contains a Field Effect Transistor, which means it needs a DC voltage for it to operate. Figure 8.3d shows how an electret mic is connected to a circuit. It needs a "load resistor" to limit the current to the FET and the output is taken across this resistor. That's the technical way of saying the output is taken from the point where the resistor meets the FET.

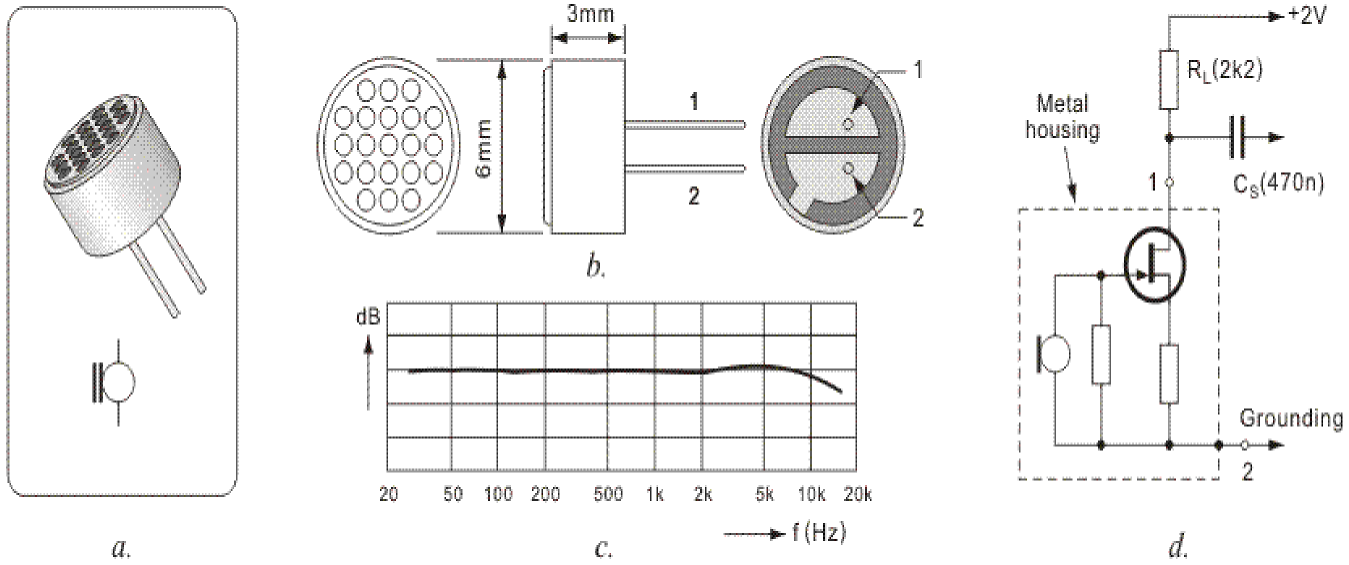
3mm
6mm
1
2
1
2
b.
dB
20
50
100
200
500
1k
2k
5k
10k
20k
f (Hz)
+2V
R_L(2k2)
Metal housing
1
C_S(470n)
Grounding
2
a.
c.
d.

8.3. Electret microphone: a - picture and schematic symbol, b - size and connectors, c - frequency characteristics, d - connection

8.2 Speakers

Speakers vary enormously in size and shape. They can be designed as crystal or capacitive, but most often they are dynamic (called electro-dynamic construction).

The cross-section of an electro-dynamic speaker is shown in 8.4. Ferrite rings (2, 3 and 4) are added to a large permanent magnet (1) which creates a strong magnetic field in the narrow gap between magnets North and South poles. A Cylindrical former is added to the gap and it holds coil (5). The ends of the coil are taken to the outside of the speaker.

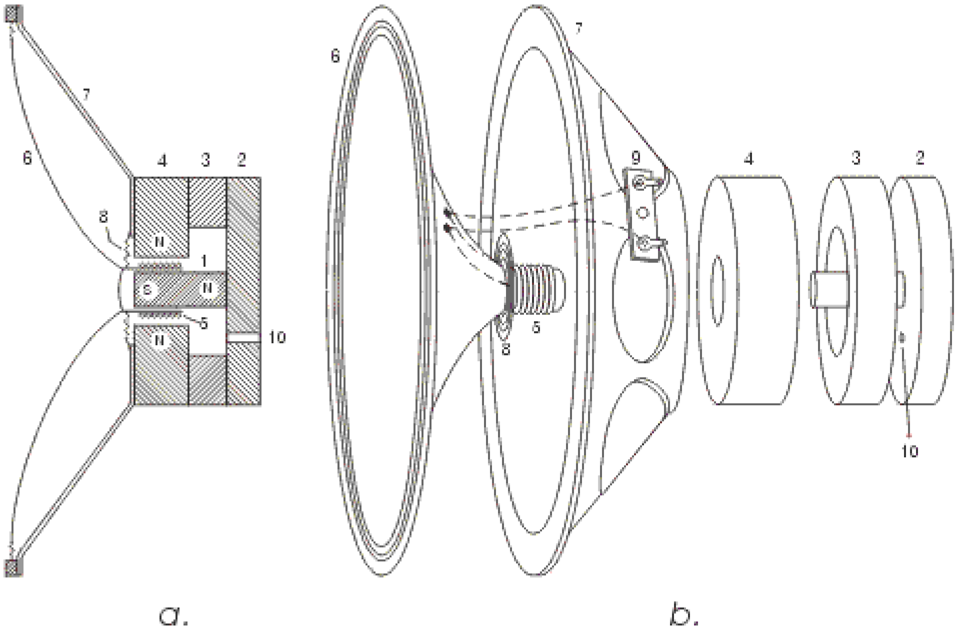
7
6
4
3
2
8
N
1
S
N
5
N
10
6
7
9
4
3
2
8
5
10
a.
b.

The two most important characteristics of a speaker are its resistance (we actually call the resistance of a speaker IMPEDANCE as the value is determined at a frequency of 1kHz and the value is higher than its actual resistance) and its wattage. Common impedances are 4, 8 and 16 ohm, but there are also 1.5, 40 and 80 ohm speakers. Speaker wattages range from a fraction of a watt to hundreds of watts.

When choosing a speaker, it is advisable to choose the largest speaker possible as they are more efficient and produce the least distortion especially in the low frequency range. Speakers should be housed in a large box since it functions as a resonating chamber and this greatly adds to the overall quality of the sound.

8.3 Headphones

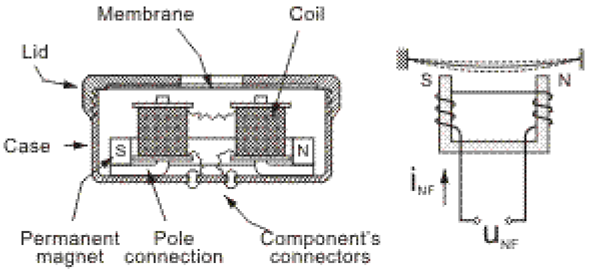

a. *b.*

8.5. Elektromagnetic headphone:

There are several types of headphone: crystal and electromagnetic. The electromagnetic type is the most commonly used. They functioning in the same way as speakers, with obvious differences in construction, since they are intended for much lower power. Their main characteristic is their resistance (impedance), from a few ohms to a few thousand ohms.

The cross section of an electromagnetic headphone is shown in 8.5. It consists of a horseshoe magnet with poles that hold two coils. These are connected in series. The diaphragm is a thin steel plate. When current flows through the coils, the diaphragm is pulled towards the coils. This moves the air and the result is heard as a faithful reproduction.

8.4 Examples

The schematic of a very simple FM radio-transmitter is shown in figure 8.6. It uses an electret microphone and transmits on a frequency between 88MHz and 108MHz.

The transistor, coil L, trimmer capacitor Ct, capacitor C3 and resistors R2, R3 and R4 creates an oscillator with a frequency determined by:

$$f_o = \frac{1}{2\pi\sqrt{L(C_t + C_{CB})}}.$$

In this equation C_{CB} represents the capacitance between the collector and the base. The value of this capacitance depends on the voltage on the base. The higher the voltage, the lower the capacitance and vice versa. The voltage on the base is constant while there is no sound, which means the frequency of the oscillator is constant. When the microphone picks up a sound, it is passed to the base of the transistor via C1. This causes the frequency of the oscillator to change and that's why the circuit is called FREQUENCY MODULATED (FM).

To transmit on a frequency away from any other radio station, a trim cap is included. The transmitter has a range up to 200 metres, depending on the length of the antenna and where it is placed. Ideally, the antenna should be vertical and as high as possible.

The antenna can be as long as 3 metres but 180cm will work very well.

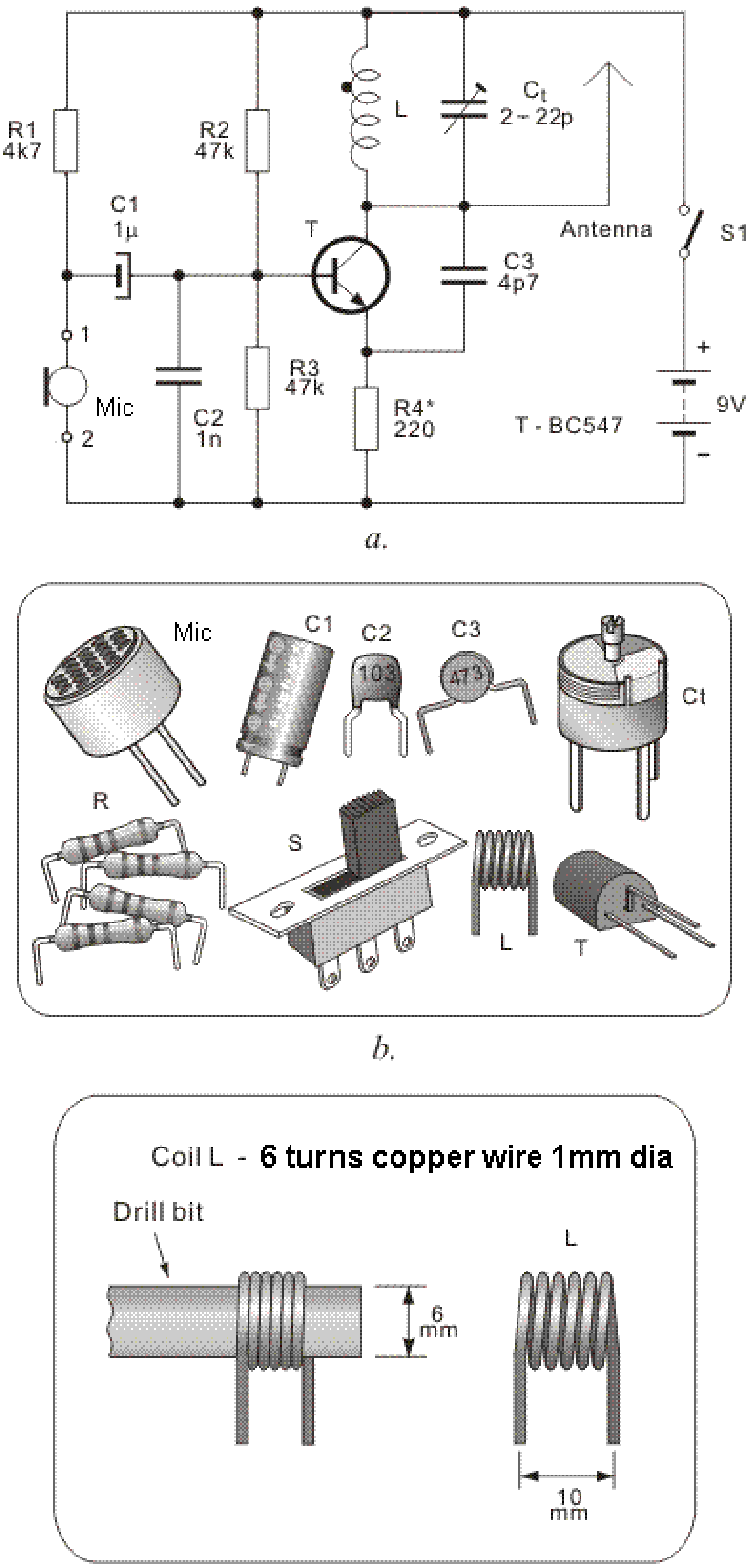
R1
4k7
R2
47k
L
Ct
2 - 22p
C1
1µ
T
Antenna
S1
C3
4p7
1
Mic
2
C2
1n
R3
47k
R4*
220
T - BC547
+
9V
-
Mic
C1
C2
C3
103
473
Ct
R
S
L
T
Coil L - 6 turns copper wire 1mm dia
Drill bit
L
6 mm
10 mm

a.

b.

c.

Sl. 8.6. Simple FM radio transmitter:
a - schematic, b - components, c -making the coil

Coil L is made by winding 6 turns of 1mm enameled wire on a 6mm dia drill bit. This coil can be stretched of squashed to adjust the operating frequency of the circuit and the trimmer will fine tune the frequency.

High Fidelity (or Hi-Fi) sound reproduction is the main purpose for using good-quality speakers. They are used in radios, TV's, cassette players, CD players, etc. The speakers are housed in speaker boxes and use at least two speakers. This is because no individual speaker is capable of reproducing the full range of frequencies. A speaker with a large cone is called a "Woofer" and will reproduce the low frequencies. A speaker with a small cone is called a "tweeter" and will reproduce the high frequencies. Together, they will reproduce the full range of between 30Hz and 15kHz.

The difficulty is now to detect the low or high frequency and divert the correct frequency to the particular speaker. This is the job of a cross-over network. In the figure 8.7 an inductor L1 passes the low frequencies to speaker Z1 and capacitor C1 passes the high frequencies to speaker Z2. Z1 reproduces frequencies from 30Hz to 800Hz and Z2 reproduces sounds with frequencies from 800Hz to 15kHz.

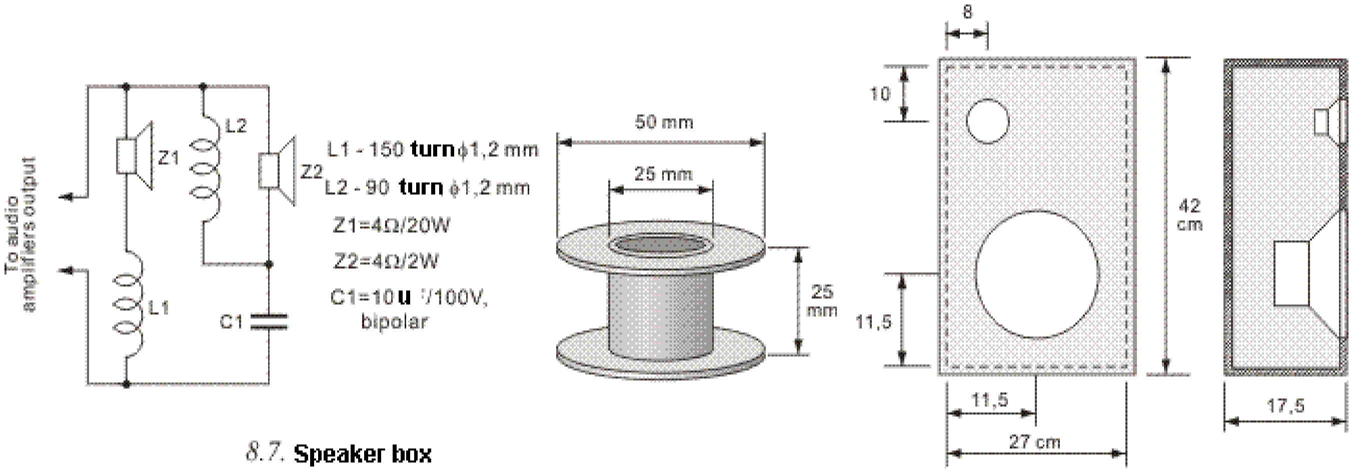
To audio amplifiers output
Z1
L2
Z2
L1
C1
L1 - 150 turn ϕ1,2 mm
L2 - 90 turn ϕ1,2 mm
Z1=4Ω/20W
Z2=4Ω/2W
C1=10u F/100V, bipolar
50 mm
25 mm
25 mm
8
10
11,5
42 cm
11,5
27 cm
17,5

8.7. **Speaker box**

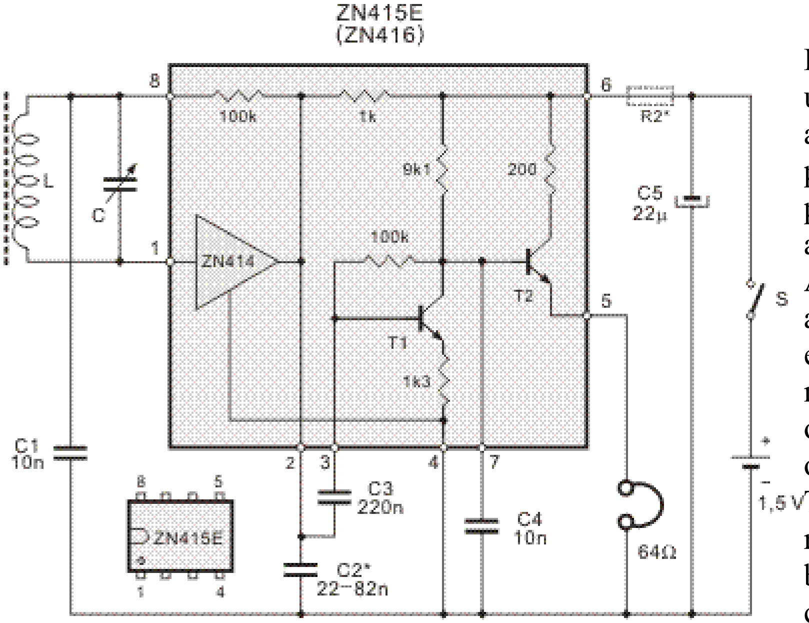
ZN415E
(ZN416)
8
100k
1k
6
R2*
L
C
9k1
200
C5
22μ
1
ZN414
100k
T2
5
S
T1
1k3
C1
10n
2
3
4
7
C3
220n
C4
10n
+
1,5 V
8
5
ZN415E
1
4
C2*
22-82n
64Ω

8.8. Mini AM radio-receiver using the ZN415E

Headphones are most commonly used with portable devices, such as radio receivers, cassette players, CD walkmans, mp3 players, etc. Headphones produce a very high quality reproduction. All modern devices have an audio-amplifier. It usually employs an integrated circuit and most of these are designed for 32 ohm headphones. There are also 8 ohm and 16 ohm headphones.

The schematic of a AM portable radio is shown in figure 8.8. It's built around the ZN416 integrated circuit. The output is connected to two serially connected 32 ohm headphones, with overall resistance of 64 ohms.

It is possible to connect the radio receiver in figure 8.8 to amplifier in figure 7.3 to produce a radio with speaker output.

9. Opto-electronic components

Optoelectronic components (or as often referred to photo-electronic components), are electronic components which produce light or react to it. Some components among them are LEDs (Light

Emitting Diodes), photo transistors, photo diodes, photo resistors (or LDR – Light Dependant Resistors), different visual indicators, light emitters and detectors, optocouplers, etc. Many of those components can be recognized easily recognized because of the “window” on the component's case which is used to pass the light. Sometimes, instead of a window, there is a small lens, which directs light to some predestined location inside of the component. Some of the most important optoelectronic components are shown on photo 9.1.

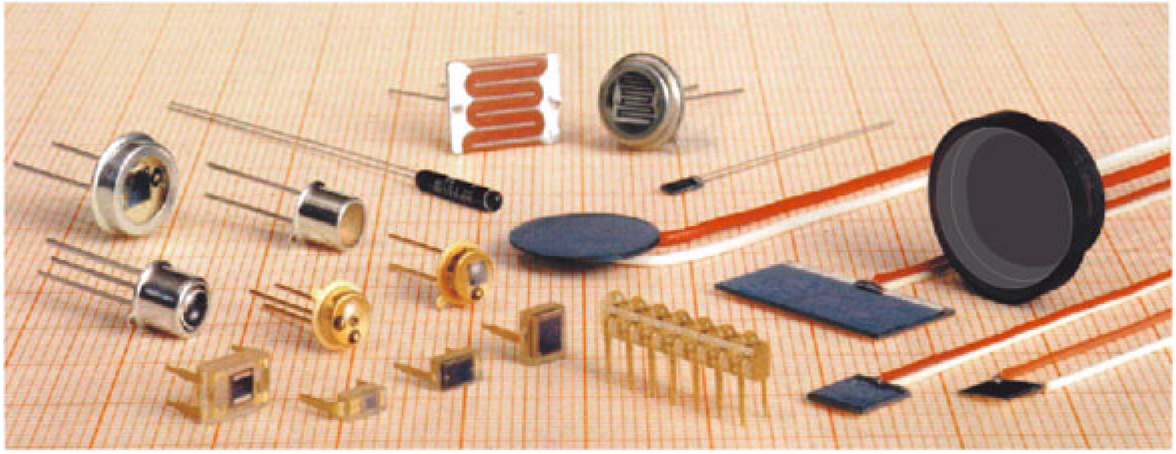

9.1 Photo-electronic components

We already mentioned the most frequently used component of them – the LED. Basic role of a LED in circuits is a visual indicator of, for example, state of the device (on/off), but is not rare in other indicator appliances, voltage stabilizers, etc. There is an abundance of colors, shapes and sizes to choose from, but most frequent ones are red, green and yellow. Because of the different and more complicated manufacturing process, blue ones cost a bit more than other ones. There are square, housed, SMD, angled, ultra bright, multicolored and many other kinds, but they all have the same principles of use.

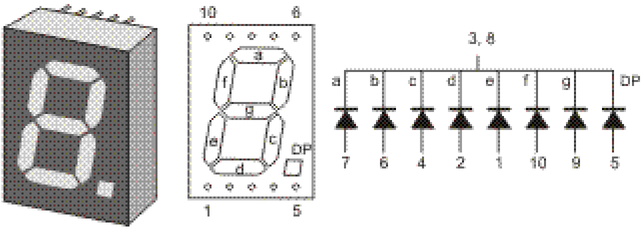

9.2 LED display

Another application of LEDs is a LED display. One display is on 9.2. It is, as shown, facilitated out of 8 diodes marked with an a,b,c,d,e,f,g and DP (DP being the Decimal Point). These devices come in two possible flavors – with a common cathode (as this display), or with a common anode. In both cases it is necessary to connect protection resistors to to all diodes (which is the same as when working with ordinary LEDs).

Photo diodes are similar to other, ordinary, diodes internally. One main difference is in that that photo diode has an exposed surface to for light to fall onto. These diodes are acting as high value resistor while in dark. It's resistance lowers as light gains in intensity. In their behavior they are similar to photo resistors, apart from that as with all diodes polarity of the component must be appropriately positioned.

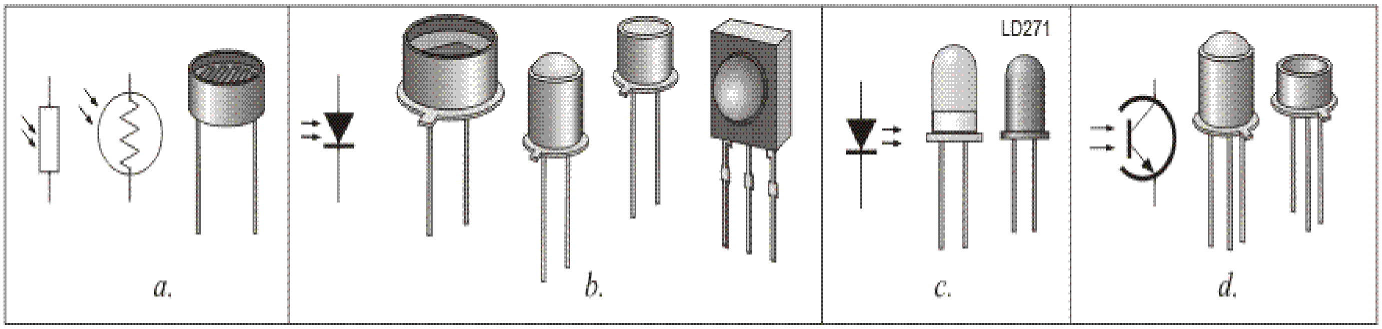

9.3. Opto-electronic components: a - resistors, b - detecting (receiving) diodes, c - emitting (transmitter) diodes, d - transistors

Emitting diodes are special kind of photo-diodes. One of them is the LED, and some of them include infra-red or ultra-violet emitting for different wireless communication purposes. Most common area of application of IR-LEDs (Infra Red) are remote controllers for TVs and other devices.

Photo diodes are usually housed in round metallic or square plastic cases with a glass window or a lens which focuses the incoming light.

Photo-transistor's internal parts are similar to internals of a regular transistor. One main difference between them is the glass window which allows light to reach the crystal plate which holds all transistor's parts. With changes of light intensity, resistance between base and the collector varies, and this influences variations of the collector current. In this component light has the same role as voltage over base of the regular transistor. When intensity rises, current through the transistor rises as well, and other way round, if intensity fades, current fades.

Photo electronic components are manufactured in an array of different case shapes and sizes. Several of them, together with their schematics symbols are displayed on 9.3.

One special group of photo-electronic components are the optocouplers. These are special integrated circuits facilitated out of an IR photo diode, and some component which is sensitive to light (photo transistor, photo thyristor). Diode is called an emitter, and “receiving” end is called the detector. This means that the only connection between the emitter and detector is through a ray of light. This is an important property of optocouplers, since it allows two different parts of the circuit which operate on different supply voltages to connect to each other without actually conducting electricity, which means that one part could operate on 9V and other on 5V without fear of burning the sensitive lower voltage components.

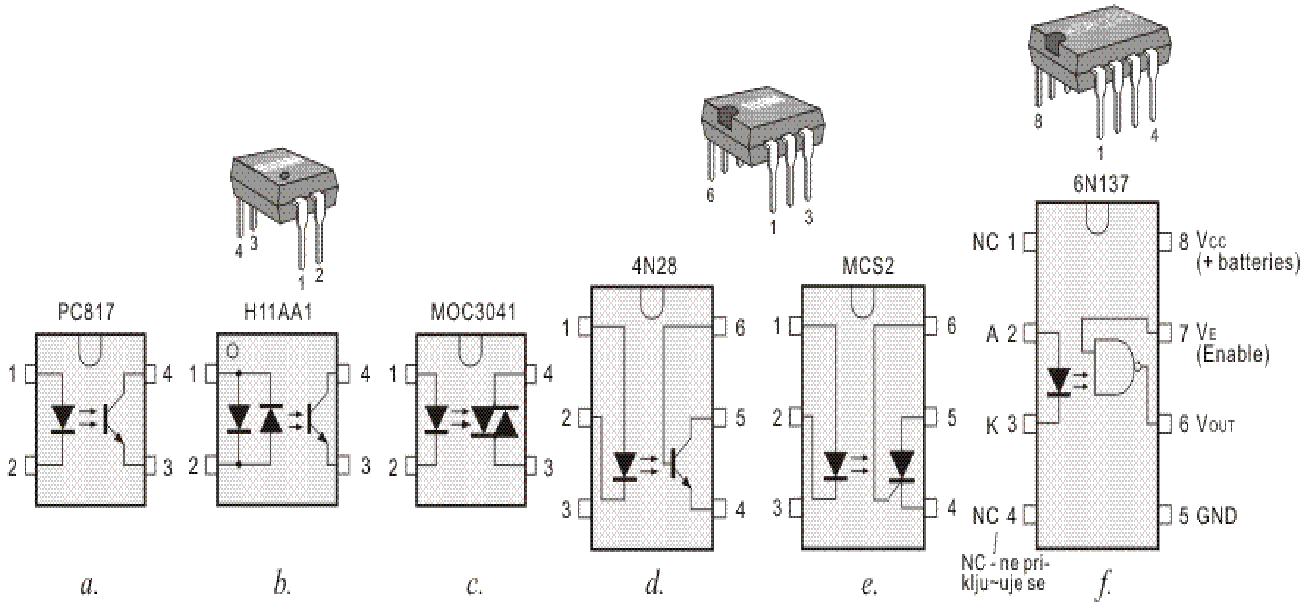

Sl. 9.4. Optocouplers

There are several optocouplers and their cases on 9.4.
Photo transistors on 9.4a are connected to other components in the same manner as ordinary transistors. Control of current which passes through it is done by light falling onto it.
Voltage to the diode on 9.4a can be variable in time, but anode must always be positive compared to the cathode. In case this component is used in an alternating current circuit, diode emits light only during one half of the interval in which anode is positive comparing to cathode. It is possible to use circuit on 9.4b in case it is needed for diode to be lit during both periods. This circuit demonstrates two diodes in anti-parallel connection, so one of the two is lit during each half of the period.
Picture 9.4c is an optocoupler using a thyristor. Thyristor is connected to other components in usual manner, and it starts conducting only upon receiving light impulse created by the diode.
Transistor on 9.4d is controlled by regulating either the light intensity of the diode or voltage over pin 6. Same goes when using a triac on 9.4e, light intensity of the diode or voltage on pin6 trigger the circuit.
Dual input NAND gate circuit is used as a detector on the 9.4f, one of those inputs controls the voltage on pin 7, and the other is controlling diode's light intensity. Logic zero on pin 6 remains only in case pin 7 has a logic one and diode is lit, any other case pin 6 has logic one.

9.1 Examples

We offer a schematic of a device which detects a certain level of intensity of ambient light, and when that level is detected, it turns on a device connected to mains grid. Data on 9.5 shows that in absence of light resistance of the LDR resistor, NORP12, is R=1MOhm, which makes both base voltage and base current very low, so there is practically no current flowing through transistor. Since there is no current flowing through the coil of the relay it's other end is in switched off position. When light intensity reaches certain point, resistance of the LDR lowers (at around 10lx resistance is approximately 9kOhm), voltages and current of the base rise, this current flows further through the relay's coil which connects pins 1 and 3 and this switches on the wanted appliance to the mains.

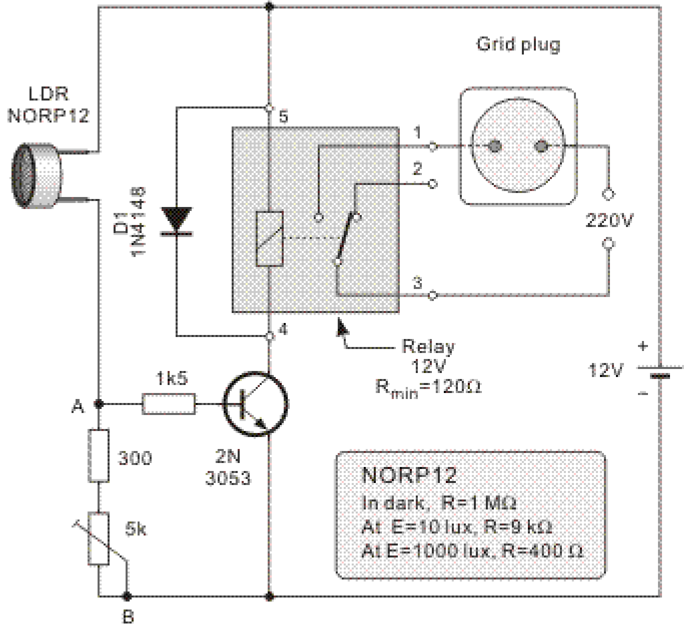

9.5. Optic relay

Slider of the 5kOhm trimmer resistor sets sensitivity of entire circuit. Lower the slider's position to lower the light level that triggers the appliance on. Greatest sensitivity is reached when trimmer is omitted from the circuit.

There is a possibility to use a photo-diode instead of a LDR (cathode goes up, to + of the battery), or a photo-transistor (collector up).

The device would be turned off when light is absent in case we placed 47kOhm regular resistor instead, and LDR between points A and B.

Each relay has a coil which accords to voltage of the battery. In our case that is 12V.Resistance of the coil is several hundreds of Ohms, and it shouldn't be lower than 120Ohm. Current rate through the relay should be equal to or greater than needed by the device plugged to mains. If, for example, we were looking at an 1kW electric heater, it's current is equal to:

I=P/U=1000W/220V=4,5 A.

Any TUN transistor whose maximum current rating is higher than current through relay's rate, is alright. This value is calculated by dividing battery voltage with relay's coil resistance.

When we want to employ remote control over some device, it is possible to utilize different technologies, but in some cases cable connection or radio wave control aren't the most appropriate ones, like the one between the TV and it's remote controller. Some IR emitting and receiving photo diodes are used specifically in low range transmitters and receivers. Block scheme on 9.6 represents usage of photo diodes between the sound source (hi-fi, radio receiver, TV) and headphones, which removes the need for long cables.

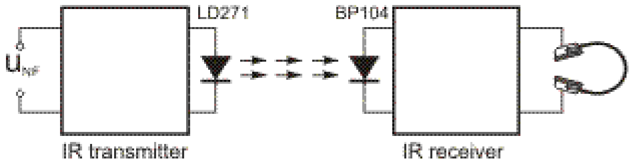

9.6. Radio transmission using the infrared light

Low frequency signal which is to be carried is marked with uLF. Based on that frequency, IR transmitter modulates the HF voltage, called the carrier. This modulated HF voltage is further sent to emitting diode LD271. Variable light emitted by this diode varies resistance of the receiving diode, and thus the HF signal created using this variations is equal to the modulated signal on the transceivers end. IR receiver is demodulating this signal, which transforms the received HF signal into the original LF signal which is equal to the original sound. This signal is further amplified and brought to headphones.

Using optical components enables safe interfacing of different devices to your home PC. There is a schematic on 9.7 which displays a simple way to interface a random device to the parallel (printer) port of the computer. For simplicity we chose to connect small portable radio receiver

supplied using a 9V battery.
Receiver, battery and the interface circuit are connected to the parallel port using the male SUB-D 25 connector. Program which is to control the circuit is easily developed in any programming language. We display a sample program written in Q-Basic, it will turn the receiver in 7am and turn it off in 7:30 am.

```
REM Wake up program
10 DO
20 LOOP UNTIL TIME$="07:00:00"
30 OUT &H378, 128
40 SLEEP 900
50 OUT &H378, 0
60 STOP
```

At 7 o'clock, voltage on pin 9 will turn to +5V, and it will remain that way for the next 900seconds.

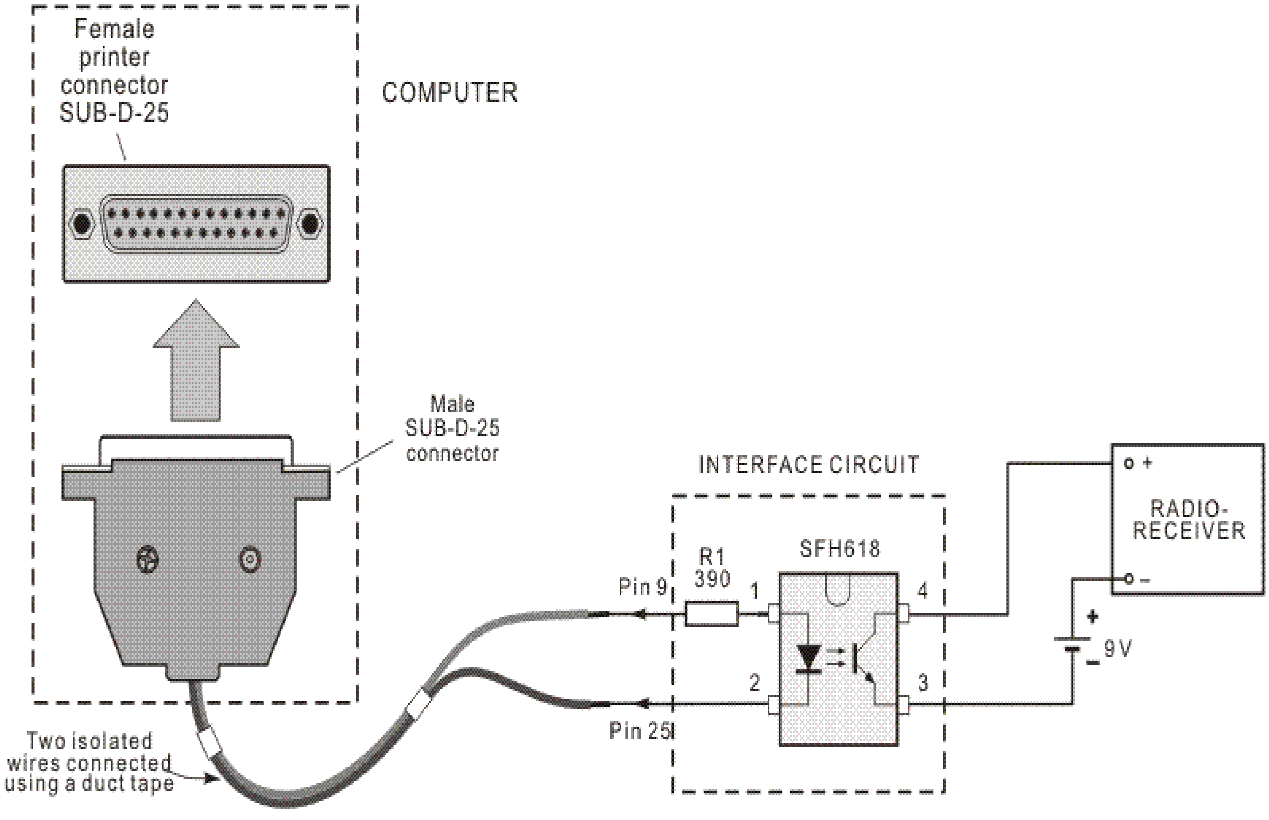
Female
printer
connector
SUB-D-25
COMPUTER
Male
SUB-D-25
connector
INTERFACE CIRCUIT
R1
390
SFH618
Pin 9
Pin 25
1
2
3
4
RADIO-
RECEIVER
+
–
9V
Two isolated
wires connected
using a duct tape

9.7. Controlling the radio receiver using a computer

A bit more modern operating systems than Windows 95 will have different ways of controlling the parallel port, and there is an extensive knowledge base on the Internet for programming this kind of operation on any operating system. Google is your friend!
Schematic of another interface circuit on 9.8 enables connection of any device plugged to the mains grid to be turned on or off. Control over this device is done in the same fashion as done in

previous program.
When, according to the program pin 9 is +5V (logic one), diode will conduct electricity. Light emitted by it switches the triac inside of the optocoupler on. This current flows through the 150Ohm resistor and creates a voltage drop which ignites the triac, which enables current flow from the mains, which powers the device.
Maximum allowed current of the BT136 triac is 4A, which means that maximum allowed power of the device is 990W. It is worth saying that optocouplers should be used only with resistance load devices (light bulbs, heaters...). When connecting inductance load devices like electro motors, transformers and such, it is advised to use the relay interfaces.

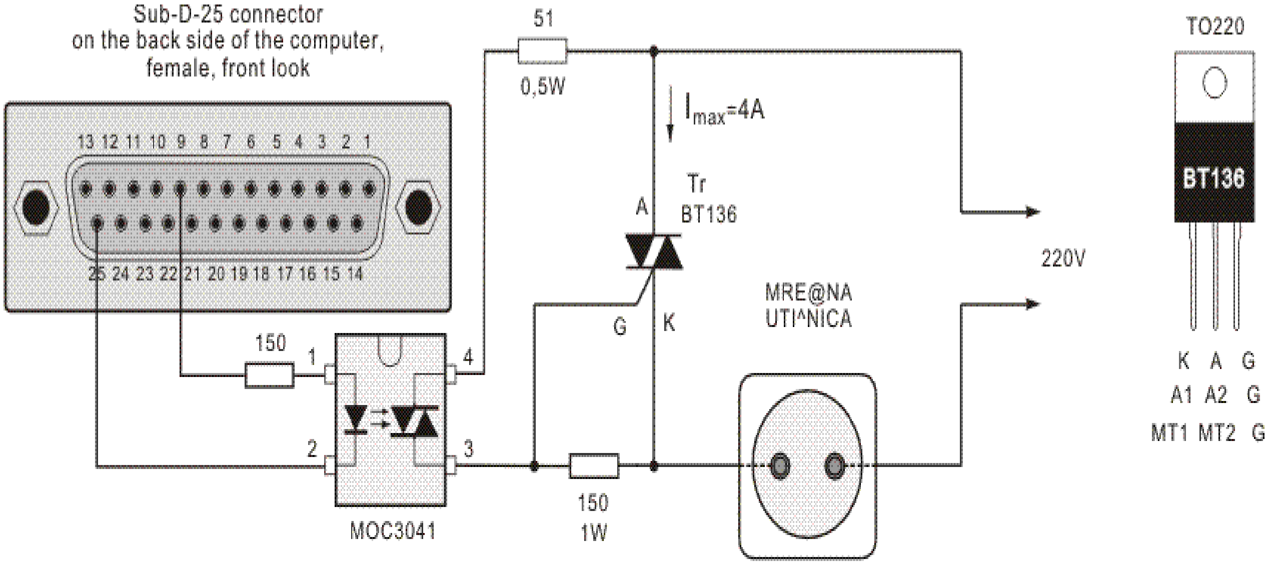
Sub-D-25 connector
on the back side of the computer,
female, front look
13 12 11 10 9 8 7 6 5 4 3 2 1
25 24 23 22 21 20 19 18 17 16 15 14
150
1
2
4
3
MOC3041
51
0,5W
I_{max}=4A
Tr
BT136
A
G
K
150
1W
MRE@NA
UTI^NICA
220V
TO220
BT136
K A G
A1 A2 G
MT1 MT2 G

9.8. Interface using an opto-coupler and triac

10. Other components

The following table covers almost every circuit symbol you will need. This is the English/American version of each symbol. The European version of some symbols is slightly different and are shown further down the page.

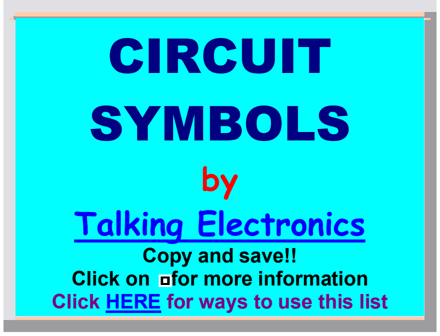

CIRCUIT
SYMBOLS
by
Talking Electronics
Copy and save!!
Click on ▫for more information
Click HERE for ways to use this list

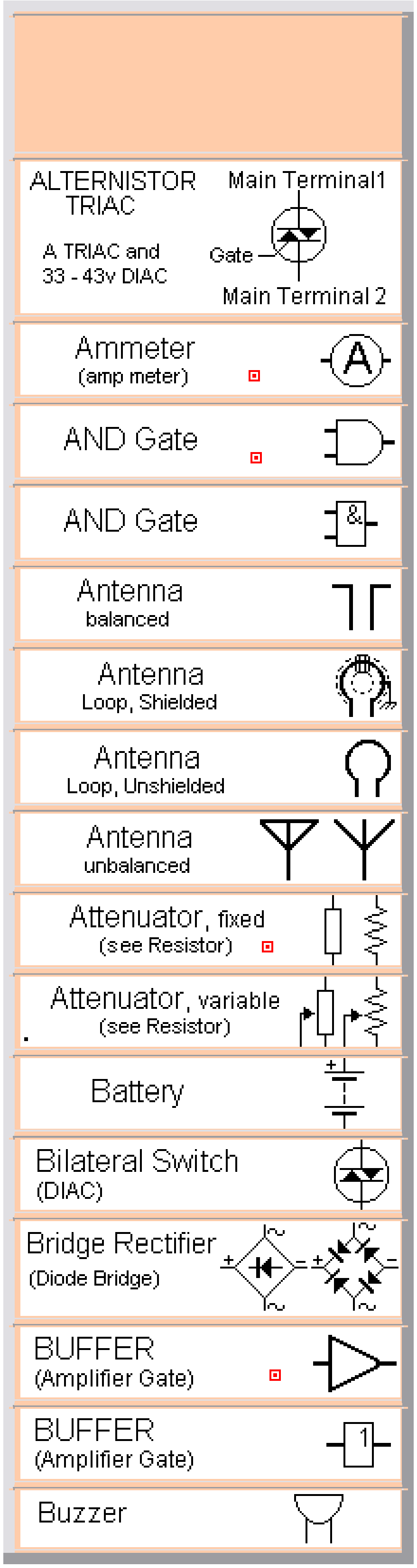

Main Terminal1
Gate
Main Terminal 2
A
&
+
~
=
-
1

ALTERNISTOR TRIAC A TRIAC and 33 - 43v DIAC
Ammeter (amp meter)
AND Gate
AND Gate
Antenna balanced
Antenna Loop, Shielded
Antenna Loop, Unshielded
Antenna unbalanced
Attenuator, fixed (see Resistor)
Attenuator, variable (see Resistor)
Battery
Bilateral Switch (DIAC)
Bridge Rectifier (Diode Bridge)
BUFFER (Amplifier Gate)
BUFFER (Amplifier Gate)
Buzzer

Capacitor feedthrough	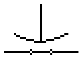
Capacitor non-polarised	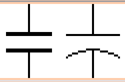
Capacitor polarised (see electrolytic)	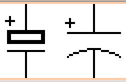
Capacitor Variable	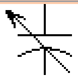
Cavity Resonator	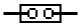
Cell	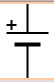
Circuit Breaker	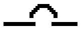
Coaxial Cable	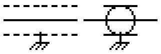
Connectors	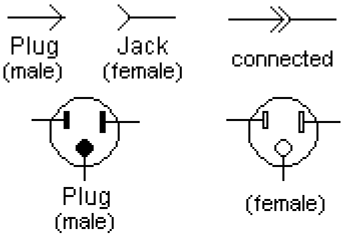 Plug (male) Jack (female) connected Plug (male) (female)
CRO – Cathode Ray Oscilloscope	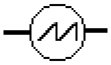
Crystal Microphone (Piezoelectric)	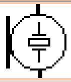
Crystal Piezoelectric	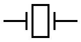
Darlington Transistor	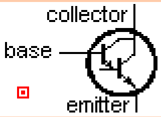 collector base emitter
Delay Line	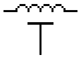

DIAC (Bilateral Switch)	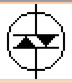
Diode	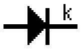
Diode - Gunn	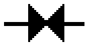
Diode - Light Emitting (LED)	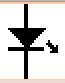
Diode Photo Sensitive	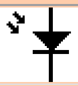
Diode Photovoltaic	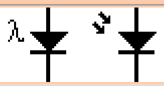
Diode Bridge (Bridge Rectifier)	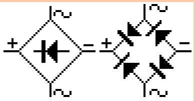
Diode - Pin	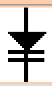
Diode - Varactor	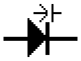
Diode - Zener	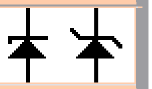
Earpiece (earphone, crystal earpiece)	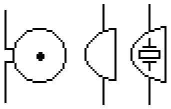
Electroluminescence	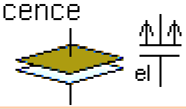
Electret Microphone (Condenser mic)	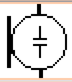
Electrolytic (Polarised Capacitor) alternate symbols: (positive on top)	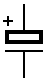 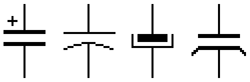
Electrolytic - Tanatalum	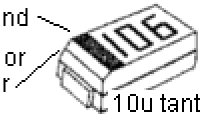 positive end black band or chamfer 106 10u tantalum 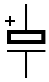

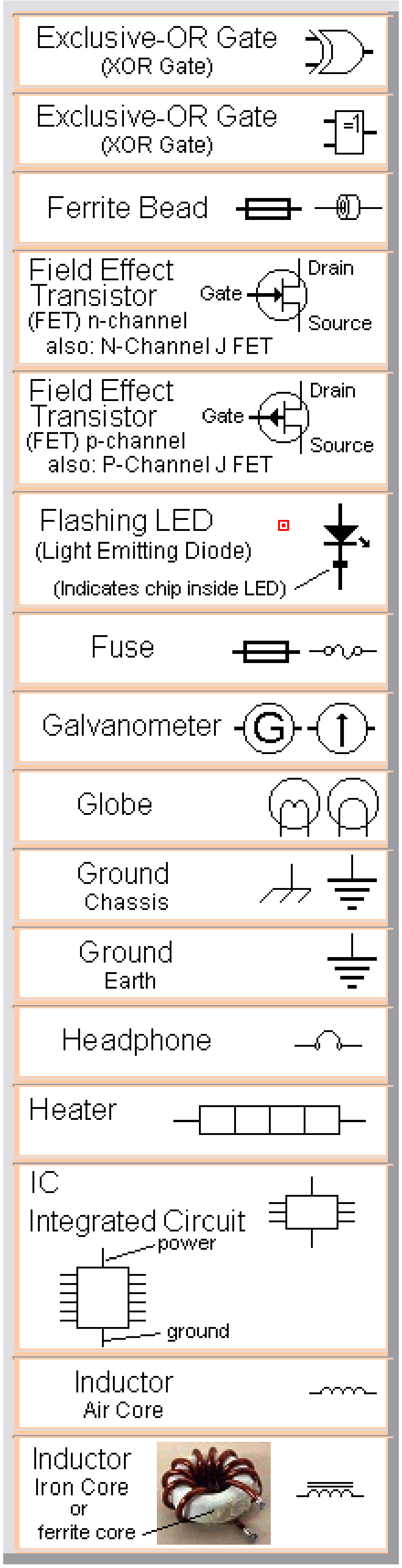

Exclusive-OR Gate
(XOR Gate)

Exclusive-OR Gate
(XOR Gate)

Ferrite Bead

Field Effect Transistor
(FET) n-channel
also: N-Channel J FET
Drain
Gate
Source

Field Effect Transistor
(FET) p-channel
also: P-Channel J FET
Drain
Gate
Source

Flashing LED
(Light Emitting Diode)
(Indicates chip inside LED)

Fuse

Galvanometer

Globe

Ground
Chassis

Ground
Earth

Headphone

Heater

IC
Integrated Circuit
power
ground

Inductor
Air Core

Inductor
Iron Core
or
ferrite core

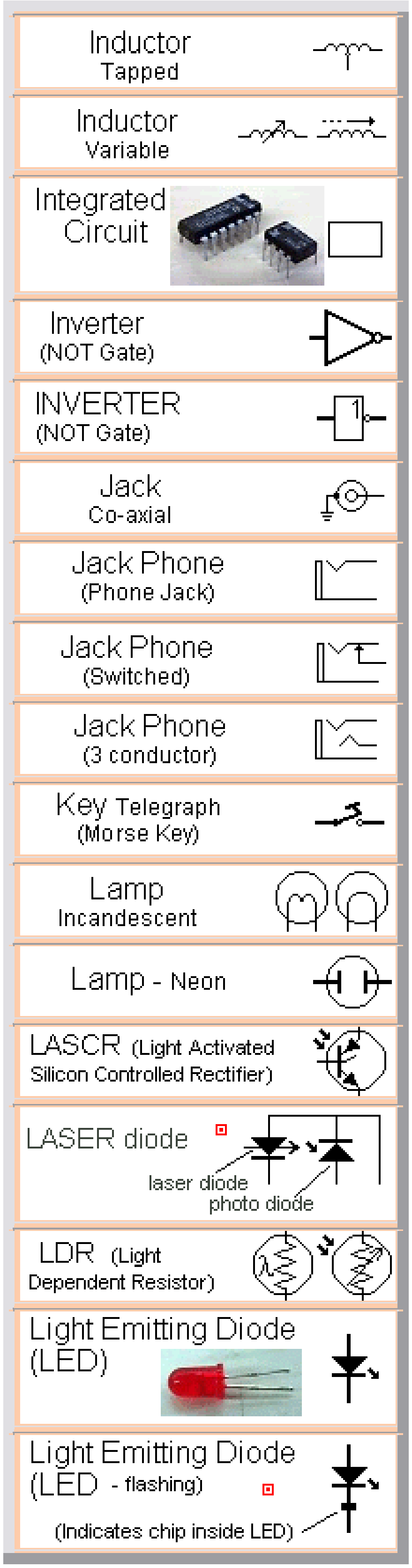

Inductor Tapped	
Inductor Variable	
Integrated Circuit	
Inverter (NOT Gate)	
INVERTER (NOT Gate)	1
Jack Co-axial	
Jack Phone (Phone Jack)	
Jack Phone (Switched)	
Jack Phone (3 conductor)	
Key Telegraph (Morse Key)	
Lamp Incandescent	
Lamp - Neon	
LASCR (Light Activated Silicon Controlled Rectifier)	
LASER diode	laser diode photo diode
LDR (Light Dependent Resistor)	
Light Emitting Diode (LED)	
Light Emitting Diode (LED - flashing) (Indicates chip inside LED)	

Mercury Switch	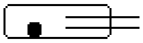
Micro-amp meter (micro-ammeter)	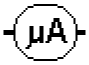
Microphone (see Electret Mic)	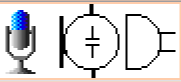
Microphone (Crystal - piezoelectric)	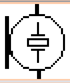
Milliamp meter (milli-ammeter)	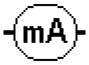
Motor	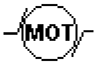
NAND Gate	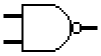
NAND Gate	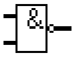
Nitinol wire "Muscle wire"	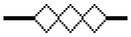
Negative Voltage Connection	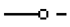
NOR Gate	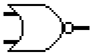
NOR Gate	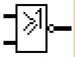
NOT Gate Inverter	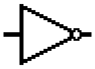
NOT Gate Inverter	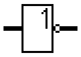
Ohm meter	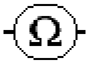
Operational Amplifier (Op Amp)	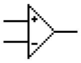
Optocoupler (Transistor output)	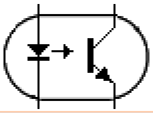
Opto Coupler (Opto-isolator)	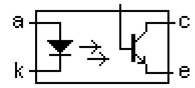 Photo-transistor output

Optocoupler
(Darlington output)

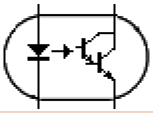

Opto Coupler
(Opto-isolator)

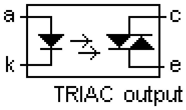

OR Gate

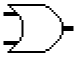

OR Gate

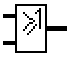

Oscilloscope
see CRO

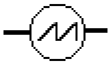

Outlet
(Power Outlet)

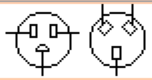

Piezo Diaphragm

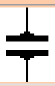

Photo Cell
(photo sensitive resistor)

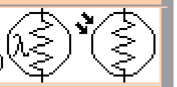

Photo Darlington
Transistor

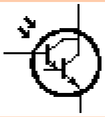

Photo Diode

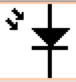

Photo FET
(Field Effect Transistor)

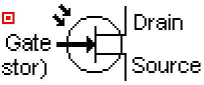

Photo Transistor

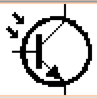

Photovoltaic Cell
(Solar Cell)

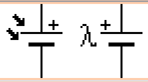

Piezo Tweeter
(Piezo Speaker)

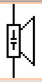

Positive Voltage
Connection

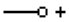

Potentiometer
(variable resistor)

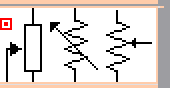

Rectifier Semiconductor	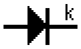
Rectifier Silicon Controlled (SCR)	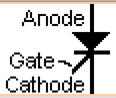
Reed Switch	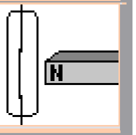
Relay - spst	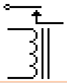
Relay - spdt	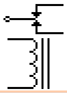
Relay - dpst	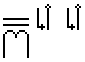
Relay - dpdt	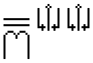
Resistor Fixed	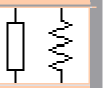
Resistor preset	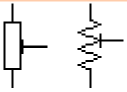
Resistor variable	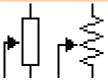
Resonator 3-pin	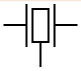
RFC Radio Frequency Choke	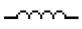
Rheostat (Variable Resistor)	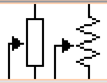
Saturable Reactor	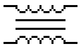
Schmitt Trigger (Inverter Gate)	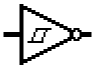
Schottky Diode (also Shottky) Low forward voltage 0.3v Fast switching also called Schottky Barrier Diode	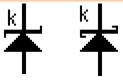

Shielding

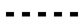

Shockley Diode

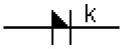

4-layer PNPN device
Remains off until forward current reaches the forward break-over voltage.

Signal Generator

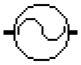

Silicon Bilateral Switch (SBS)

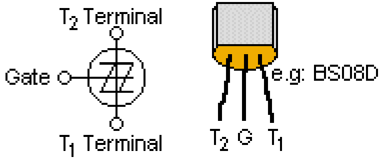
T_2 Terminal
Gate
T_1 Terminal
e.g: BS08D
T_2 G T_1

Silicon Controlled Rectifier (SCR)

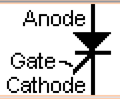
Anode
Gate
Cathode

Silicon Unilateral Switch (SUS)

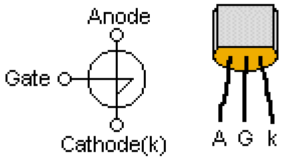
Anode
Gate
Cathode(k)
A G k

Solar Cell

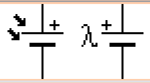

Speaker

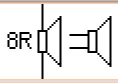

Surface Mount

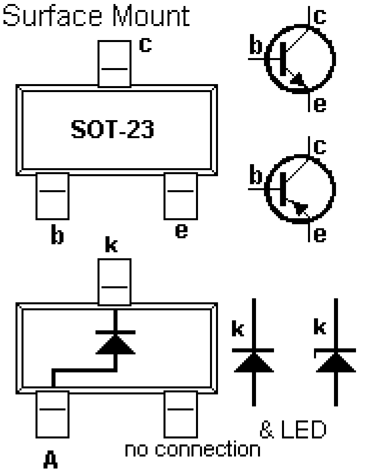
c
SOT-23
b
e
b
c
e
b
c
e
k
k
k
A
no connection
& LED

Switch - mercury
tilt switch

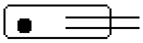

Switch - spst

Switch - spdt	
Switch - dpst	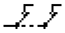
Switch - dpdt	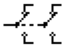
Switch - process activated normally open: normally closed:	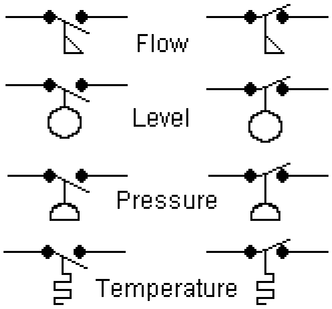 Flow Level Pressure Temperature
Switch - push (Push Button)	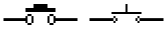
Switch - push off (used in alarms etc)	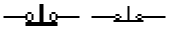
Switch - Rotary	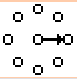
Test Point	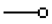
Thermal Probe **NTC:** as temp rises, resistance decreases	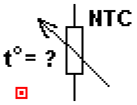
Thermocouple	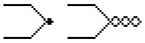
Thyristors:	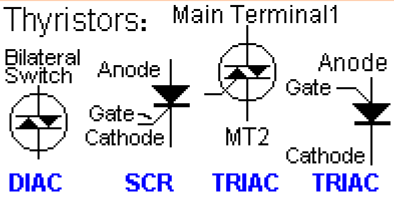 Main Terminal1 Bilateral Switch Anode Gate Cathode MT2 Anode Gate Cathode DIAC SCR TRIAC TRIAC
Tilt switch mercury	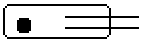
Touch Sensor	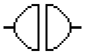
Transformer Air Core	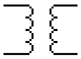

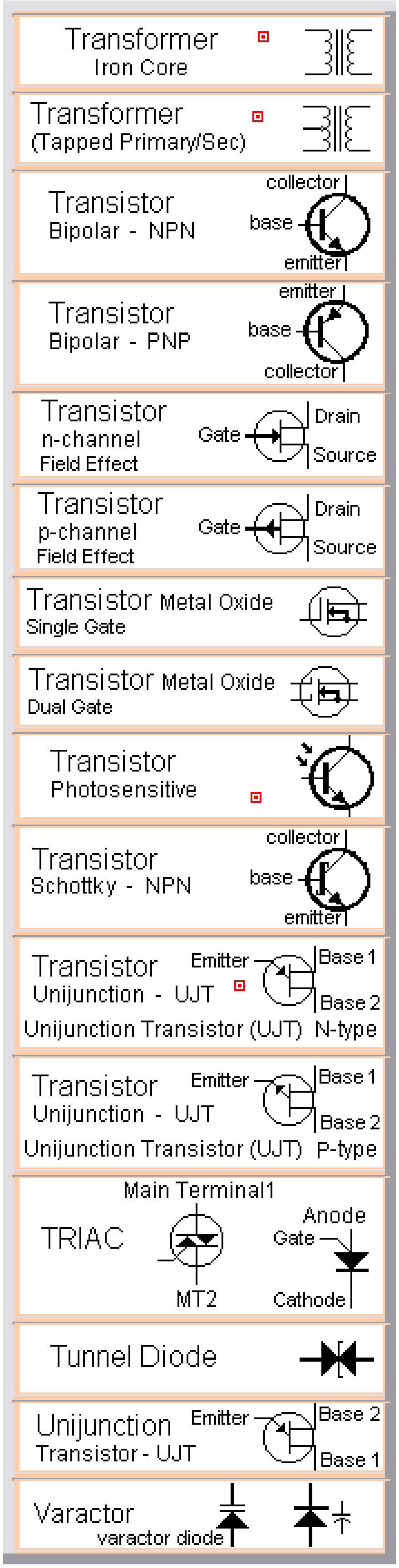

Transformer
Iron Core

Transformer
(Tapped Primary/Sec)

Transistor
Bipolar - NPN
collector
base
emitter

Transistor
Bipolar - PNP
emitter
base
collector

Transistor
n-channel
Field Effect
Gate
Drain
Source

Transistor
p-channel
Field Effect
Gate
Drain
Source

Transistor Metal Oxide
Single Gate

Transistor Metal Oxide
Dual Gate

Transistor
Photosensitive

Transistor
Schottky - NPN
collector
base
emitter

Transistor
Unijunction - UJT
Emitter
Base 1
Base 2
Unijunction Transistor (UJT) N-type

Transistor
Unijunction - UJT
Emitter
Base 1
Base 2
Unijunction Transistor (UJT) P-type

Main Terminal1
TRIAC
MT2
Anode
Gate
Cathode

Tunnel Diode

Unijunction
Transistor - UJT
Emitter
Base 2
Base 1

Varactor
varactor diode

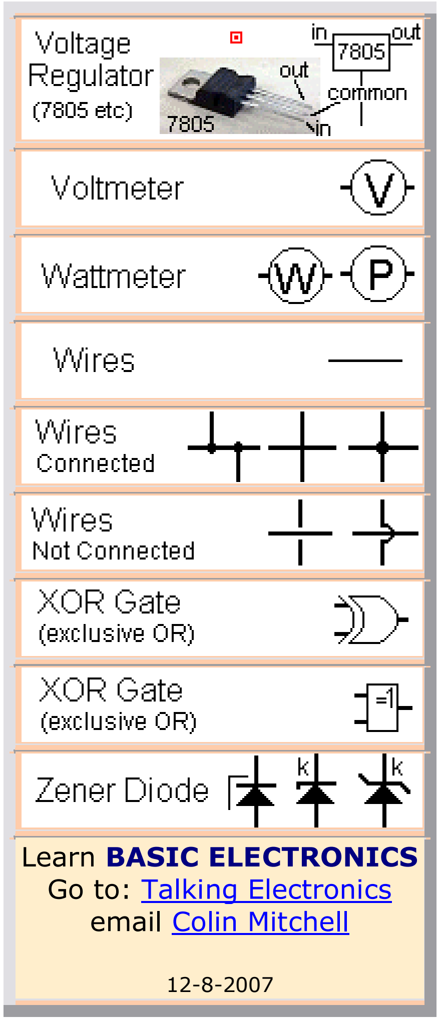
in
7805
out
out
common
7805
in
k
k

Voltage Regulator (7805 etc)

Voltmeter

Wattmeter

Wires

Wires Connected

Wires Not Connected

XOR Gate (exclusive OR)

XOR Gate (exclusive OR)

Zener Diode

Learn **BASIC ELECTRONICS**
Go to: Talking Electronics
email Colin Mitchell

12-8-2007

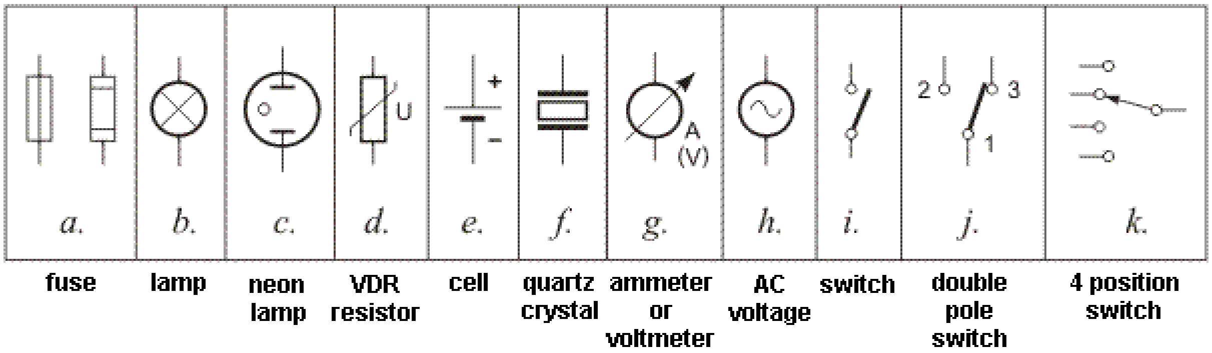
U
+
-
A
(V)
2
3
1
a.
b.
c.
d.
e.
f.
g.
h.
i.
j.
k.
fuse
lamp
neon lamp
VDR resistor
cell
quartz crystal
ammeter or voltmeter
AC voltage
switch
double pole switch
4 position switch

10.1.

NOTES

Here are a few notes on the symbols above.

Fuses (10.1a) have single role in a circuit - to detect excess current and protect the device. In most cases the excess current flows when a higher voltage is present but a fuse cannot detect the voltage - it can only detect when a higher current flows. The higher voltage causes the higher current to flow and this triggers the action of "blowing the fuse." Of course, when a component fails, a higher current can flow and this will also "blow the fuse."

Fuses come in all sizes and ratings (current flow) and it is important to know that the size of the wire inside a fuse does not necessarily indicate the current rating.

The wire inside can be made from copper and plated to protect it from oxidizing or it can be a low temperature material that needs to be a larger diameter.

The wire can also be wound in a spiral and formed into a spring. The end of the spring sits in a dob of solder and when the spring heats up, the solder melts and the spring separates from the other end.

This is called a DELAY FUSE.

Other forms of delay fuse consist of a wire joined at the centre by a dob of solder and others are made of low-temperature-melting material.

Some pieces of equipment use expensive fuses and whenever a fuse is damaged, you must decide if the problem is a major or minor fault.

Sometimes a fuse can go open-circuit for no apparent reason. It can "wear-out."

For instance, some equipment takes a very high current when it is turned on and you will see the fuse heat up and stretch and dip in the middle. This causes strain on the fuse and eventually the wire oxidizes to a point where it finally "burns out."

The equipment is not faulty and it is just a matter of replacing the fuse.

Sometimes the fuse completely explodes and the glass is thrown all over the chassis. This indicates a short-circuit in the power supply and most often one or more of the diodes must be replaced.

The fuse can also go off with a "bang" and the inside of the glass is coated with "silver." This also indicates a diode is damaged in the power supply. Generally 2 or 4 diodes are damaged.

If the fuse is damaged beyond recognition, you will not know if it is a delay fuse or a normal fuse.

The current-rating on the end-cap can sometimes help you.

For instance, if a fuse is rated at 4A, you will need to replace it with a 4 amp normal fuse or 3.15 amp slow-blow.

When fuses are rating at 100mA to 250mA, they are very delicate and will not accept the slightest overload.

When replacing this type of fuse, it is necessary to determine if the equipment is drawing a heavy current when turning on or if a fault exists in the power supply. Sometimes the switch can cause the problem if it is not making contact fast enough.

Replace the fuse and watch it as someone else turns on the equipment. If the fuse burns out immediately, a short exists. If the fuse glows red and burns out, the equipment is drawing too much current during turn-on. This may be due to devices you have added to the equipment or operation on a slightly higher voltage. You can try a fuse with a slightly higher rating to see if the fault is fixed.

Never replace a 100mA fuse with a 1 amp fuse. The 1 amp fuse will never "blow" and if the transformer is being overloaded, the transformer will simply "cook."

Lamps (10.1b) Ordinary electric light globes heat a coil of tungsten wire inside a glass bulb that

has an inert gas such as argon. The resistance of the filament depends on the temperature it is heated to. It can be ten to twenty times higher than when it is cold.
A neon lamp (10.1c) contains a gas (such as neon) and this gas gives off a glow when a high voltage is applied to two plates. This glow occurs at about 70v to 90v and a resistor must be used in series to prevent the voltage rising higher than required by the lamp. To put this more accurately, the resistance of the neon lamp reduces when it "strikes" and a high current will flow. To limit this current a "current limit" resistor is needed.

VDR (10.1d) The resistance of a VDR depends on the voltage across it. A VDR is also called a VARISTOR. Its resistance is high until a critical value of voltage and the resistance suddenly drops. They are used as voltage protection devices. If they, for example, see a voltage higher than 220V, their resistance decreases and this “soaks” the excess voltage. Their response time is only a few 10's of nanoseconds.

The symbol for a single DC cell is shown in 10.1e.
A Quartz crystal is shown in 10.1f. It is a thin sheet of quarts material between two metal plates and packaged in a metal case. Quartz crystals are commonly used as the reference for an oscillator circuit, such as a clock source in microprocessor designs.
An instrument for measuring current (A) and voltage (V) is shown in 10.1g. This symbol dates from the time when analog instruments with a needle were used. The symbol remains the same, although digital instruments have replaced analog devices.
AC voltage symbol is shown in 10.1h. The shape of the wave is shown in the symbol. It can be sine-wave or saw-tooth or square-wave.
The simplest form of switch device is displayed in 10.1i. Because of the wide range of switches, there are many different types in use. For example, a two pole switch (10.1j) has two operating positions, in one position it connects points 1 and 2, and in the other it connects points 1 and 3. There are switches with more operating positions. 10.1k is an example of a rotary switch with four positions.

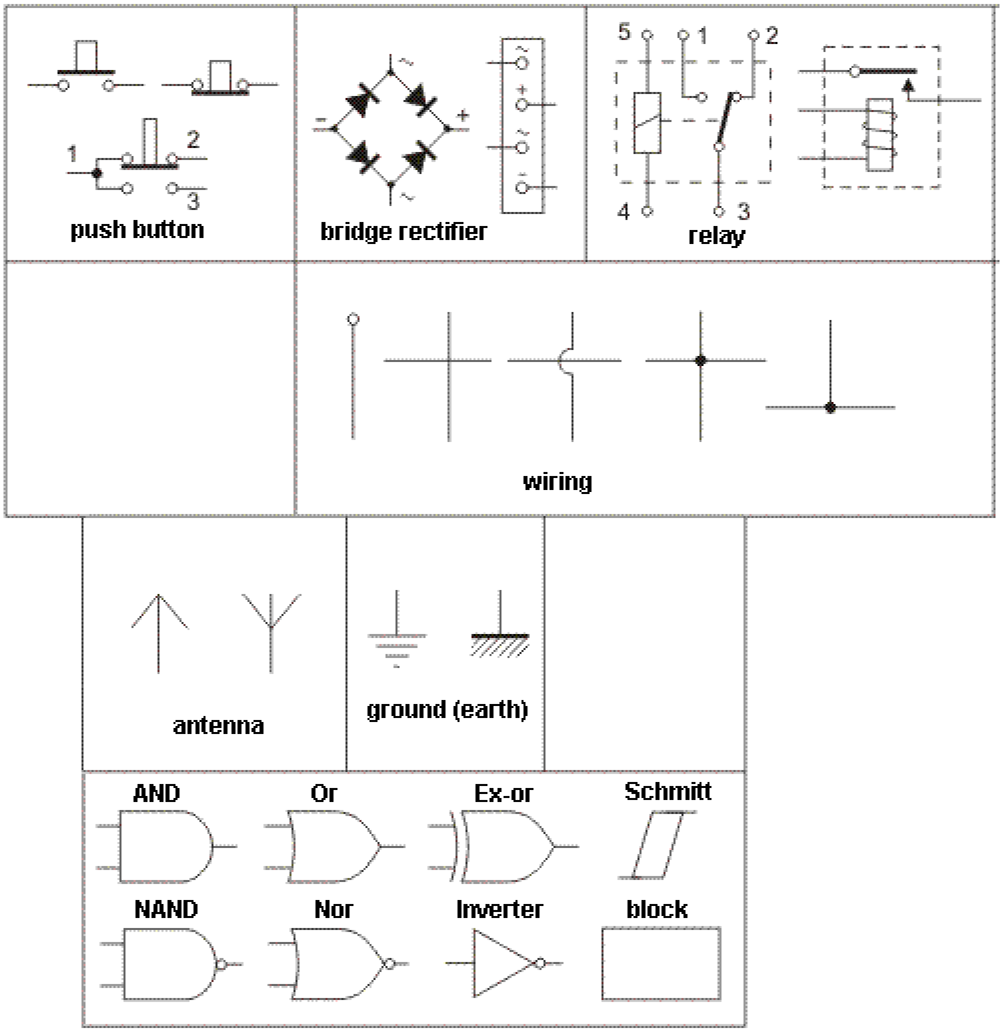

Momentary switches, or push buttons have a built-in spring, which makes the switch conduct only while it is being pressed (your standard doorbell has this kind of switch).
Four diodes in a single case is called a BRIDGE. Two pins are marked with sine waves, used to connect to the AC voltage and two marked with "+" and "-"

RELAY When an electromagnet receives sufficient voltage on points 4 and 5, connection between points 2 and 3 is opened, and at the same time points 3 and 1 are closed. A relay is actually an electromagnetic switch.

Symbols for a receiving and transmitting antenna are shown.
Grounding symbols Grounding and common ground aren't the same thing, but if both exist in a circuit, they are always connected to each other. With electronic devices housed in a metal case, grounding is connected to the metal housing.

Schematic symbols representing logic gates and different digital integrated circuits are shown above. It should be kept in mind that basic logic gates (AND, OR, XOR, Inverter, etc.) aren't manufactured as single standalone components. They are always integrated in groups in an IC, but for the sake of clarity, they are represented as separate blocks. These components require a DC voltage, which may or may not be represented on the schematic. These voltages might be different depending on the internal structure and technology used between different family types. Detailed info on this can be found in the component's datasheet provided by the manufacturer.

10.1 Relays

A relay is an electro mechanical device which is commonly used to connect two different circuits. It can connect a low voltage circuit to a high voltage circuit or a low current circuit to a high current circuit or simply to isolate two circuits.
The simplest relay has one set of contacts (commonly called "change-over" contacts). Inside the relay is a coil (called a solenoid) and when the coil is energised, the centre core of the solenoid becomes magnetised and moves an arm closer to the coil. A "contact" is connected to this arm and the contact touches another contact to complete a circuit. The contacts are labeled "common" for the moving contact, "normally open" and "normally closed." This can be seen in diagram 10.2 a:

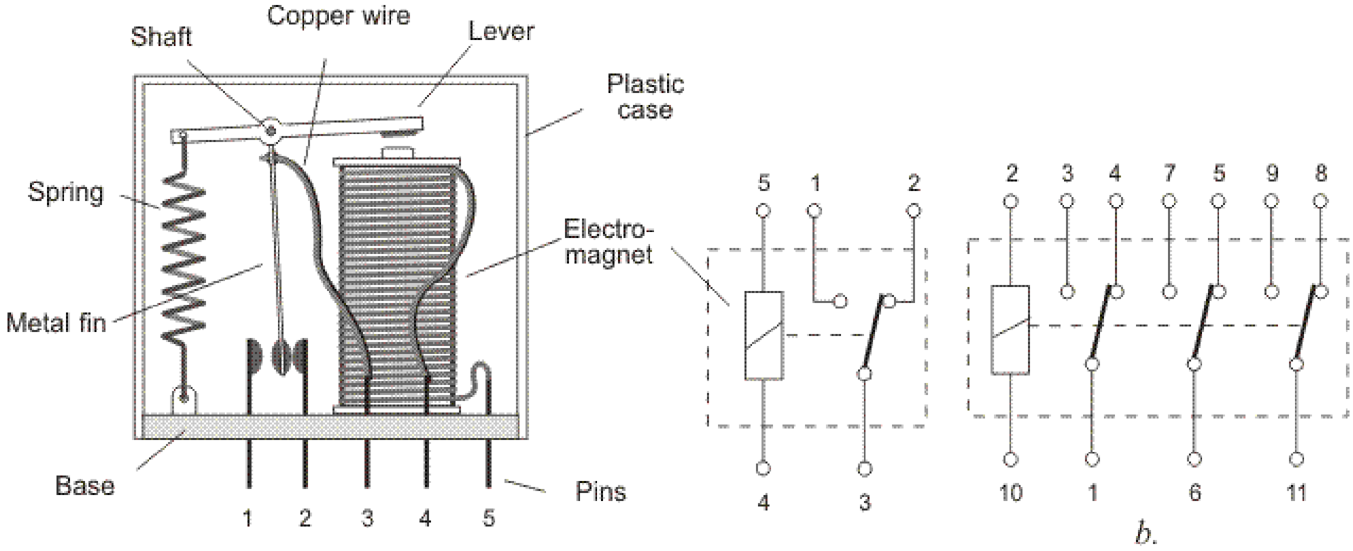

10.2. a - internal structure and relay symbol b - relay with three pairs of contacts

A relay can be connected as the collector load of a transistor, as shown on 10.3. When sufficient collector current flows in the transistor, the relay is activated and any device connected to the contacts will be operational.
Since a relay is an electro mechanic component which is consisted of moving parts, it has a limited operational life span, and cannot be used for rapid switching. It would not be very effective using it in a, for example, light show which has frequent switching frequency (several hundreds or thousands times per hour). Each opening and closing of the contact is followed by sparks which would dramatically shorten the life of such device.

Coil values are “input values” or voltage and resistance values at which relay draws the lever and switches. Usual coil voltages are 3V, 5V, 6V, 12V and 24V. They can be found printed on the relay's housing. These are all DC voltages, but there are AC voltage designed relays with 230V/250V. The current taken by the relay depends on the resistance of the coil. The coil resistance can be measured with a multimeter. Current flowing through the coil is calculated using Ohm's law, by dividing the relay's voltage by its resistance. For example a 12v relay has a coil resistance of 300 Ohm, which means the current flow is: I=U/R=12/300=40mA.

2. Voltage on relay's contacts, also marked on the housing, is the maximum value allowed. Over-voltage will cause sparks inside the relay and possibly damage the contacts.

The maximum current rating for a relay is marked on the housing with all the other information. It is usually higher than 1A.

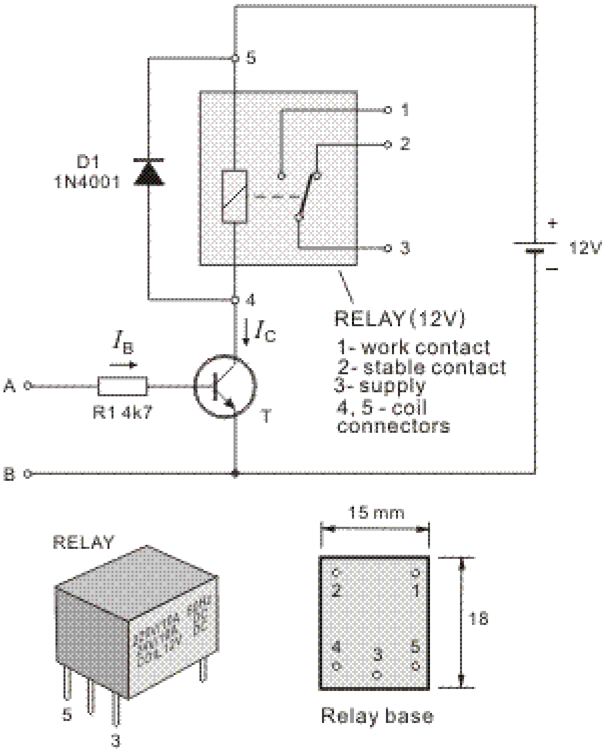

10.3. Relay as an collector transistor load

11. Checking Components

So you've put a circuit together and as far as you know everything appears to be ok, but it doesn't work as expected. Even worse, it refuses to give any signs of life. What do you do? First, check the circuit for mechanical failures, like non-connected wires, broken vias on the board (these are holes on the printed circuit board that have a metal coating down the length of the hole to connect one side of the board to the other), bad battery contacts inside the case, broken pins on a component, cold solder joints, etc.

If this doesn't come up with a result, you should compare values of components with the schematic.

You may have put a component in the wrong place, or read values the wrong way. Maybe you forgot k in front of Ohms. Maybe you connected the supply to the wrong pin of an IC.

The next step is to test each component on the board.

Start troubleshooting by measuring DC voltages at certain points of the board, and comparing these values to the schematic. So, by knowing the operation of the circuit you start the process of elimination to find the "suspect" component.

If there are several "suspects", and this is not a rare occurrence in complex devices, the testing is divided into groups of components. You start checking in reverse soldering order, this means you start with components last soldered, because those are the most sensitive components on the circuit like integrated circuits, transistors, diodes, etc.

The fastest and simplest method to troubleshoot is to use an "ohm-meter."

In most cases you don't have an ohm-meter by itself as it is usually aded to an ammeter and voltmeter in one instrument, called AVO meter or multimeter.

The safest and most accurate method is to desolder the component from the board when testing it, because other components could lead to a wrong diagnosis, so you have to be very careful when testing in-circuit.

Ok, you should know something about multimeters now. There are two kinds: analog and digital. Analog ones are items of the past, and since they use a needle to tell you values, it can be difficult determining the right value. Digital meters, on the other hand have a display. You should go for this type, although both come in different sizes and with different ranges. Their price is from several dollars, to several hundreds of dollars for really good professional types.

Two instruments are shown in 11.1.

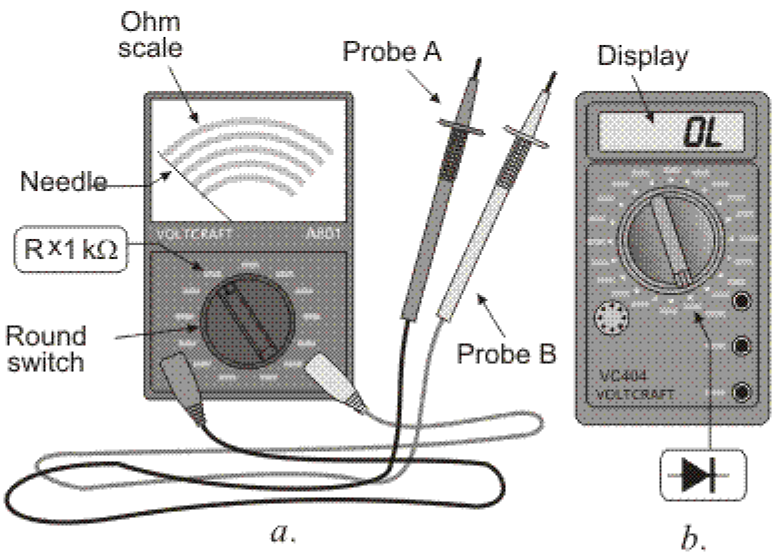

11.1. Multimeter (AVO meter):
a - analog, b - digital

11.1 Diodes and Transistors

When using an analog instrument to test a diode, the needle will swing almost fully across the scale when the diode is placed in one direction and hardly move when the diode is reversed. The needle does not measure the resistance of the diode but rather the flow of current in one direction and no current-flow in the other direction.

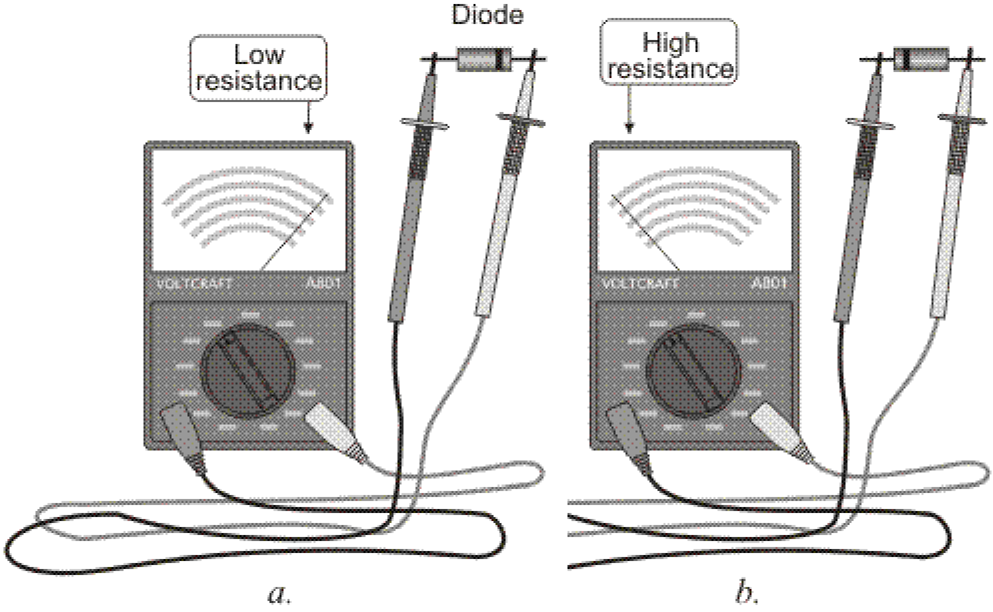
Low resistance
Diode
High resistance
VOLTCRAFT
A801
VOLTCRAFT
A801
a.
b.

11.2. Diode testing using an analog instrument

If the value is equal to or near equal, either low or high in both directions, the diode is faulty, and should be replaced.

Digital instruments have a position on the dial to measure diodes, as shown in 11.1b. When we connect probes to each other, the multimeter should buzz, which signals a short circuit, and display tells 0. When we separate the probes the buzzing stops, and a symbol for open circuit is displayed (this can be either 0L or 1). Now we connect probes to the diode (11.3a). Then we reverse the diode and connect it again (11.3b). If the measured diode was ok, one of the two measurements would have shown a value which represents a minimum voltage that could be conducted through the diode (between 400mV and

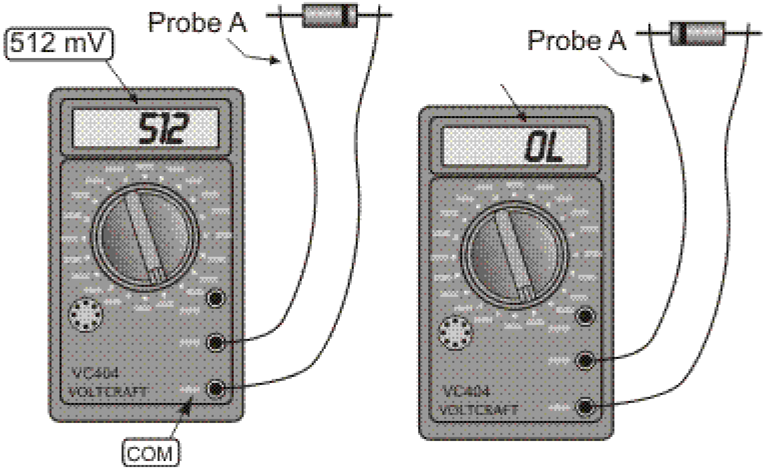
512 mV
Probe A
512
COM
VC404
VOLTCRAFT
Probe A
0L
VC404
VOLTCRAFT

11.3. Diode testing using a digital instrument

800mV), and the anode is the end of the diode which is connected to probe A (red one). The diode is faulty if you hear a buzz (closed circuit) or some value which represents infinity. Transistors are tested in a similar fashion, since they act as two connected diodes. According to 11.4b, the positive probe is connected to the base, and the negative probe is first connected to the collector and then the emitter. In both cases the resistance should be low. After that, you do the same thing, only with switched probes. The negative probe is connected to the base and you test the collector and emitter with a positive probe. Both cases should produce a high value on the meter.

When testing PNP transistors, all steps are the same, but the measurements should be opposite: on 11.4a they are high, and on 11.4c they are low.

If you test transistors using a digital instrument, the process remains similar to the one with diodes. Each diode should produce a value between 400mV and 800mV. Many modern digital multimeters have a socket for testing transistors. There is, as displayed on 11.5, a special socket where low and medium power transistors fit. If you need to test high power transistors, thin wires (0.8mm) should be soldered to transistor's pins and then plugged into the socket. As displayed on 11.5, a transistor is plugged into the socket according to its type (PNP or NPN) and the switch with a hFE marking is brought into position. If the transistor works, the display shows a value which represents the current amplification coefficient. If, for example, a transistor is tested, and the display shows 74, this means the collector current is 74 times higher than the base current.

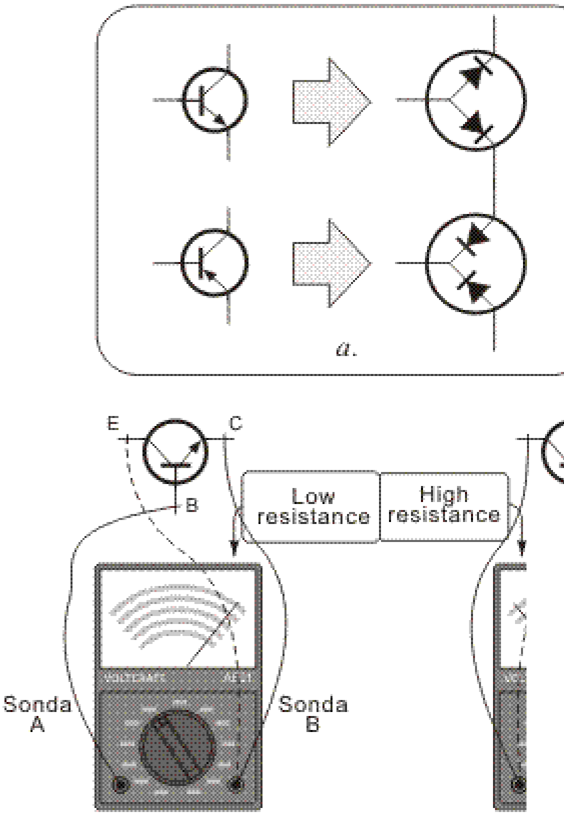

11.4. Transistor test

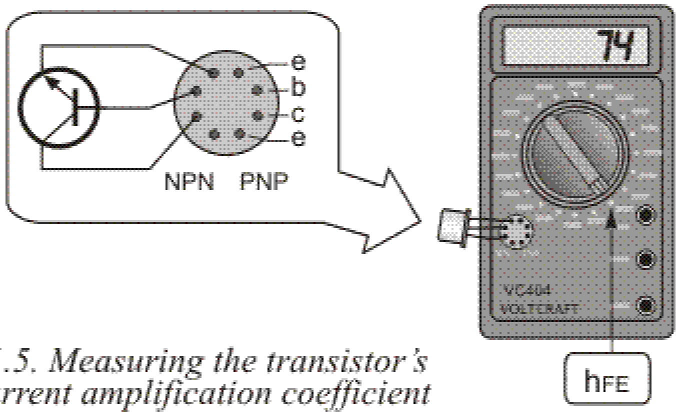

11.5. Measuring the transistor's current amplification coefficient

11.2 Transformers and coils

Transformers are tested by measuring the resistance of the copper wire on the primary and secondary. Since the primary has more turns than the secondary, and is wound using a thinner wire, its resistance is higher, and its value is in range of tens of ohms (in high power transformers) to several hundreds of ohms.

Secondary resistance is lower and is in range between several ohms to several tens of ohms, where the principle of inverse relations is still in place, high power means low resistance.

If the multimeter shows an infinite value, it means the coil is either poorly connected or the turns are disconnected at some point.

Coils can be tested in the same way as transformers – through their resistance. All principles remain the same as with transformers. Infinite resistance means an open winding.

11.3 Capacitors

Capacitors should produce an infinite reading on a multimeter. Exceptions are electrolytics and very high value block capacitors. When the positive end of an electrolytic capacitor is connected to the positive probe of an analog instrument, and a negative end to a negative probe, the needle moves slightly and gradually comes back towards infinity. This is proof the capacitor is ok, and the needle's movement is charge being stored in the capacitor. (Even small capacitors get charged while testing.)

Variable capacitors are tested by connecting an ohm-meter to them, and turning the rotor. The needle should point to infinity at all times, because any other value means the plates of the rotor and stator are touching at some point.

There are digital meters that have the ability to measure capacitance, which simplifies the process. With this said, it is worth mentioning that capacitors have considerably wider tolerance than resistors, (about 20%).

11.4 Potentiometers

To test a potentiometer, (pot), or a variable resistor, the process is rather simple – you connect the component to the probes of a meter set to ohms and turn the shaft.

(A “noisy” pot can be repaired using a special spray.)

11.5 Speakers and headphones

When testing speakers, their voice-coil can be between 1.5 and 32 Ohms. The value marked on the speaker is an impedance value and the actual DC resistance will be lower. When measuring a speaker with an analogue meter, you should hear a click when the probes are connected.

12. Conductivity probe

Conductivity tester is a simple, but very important instrument, which is able to test for faults many components like: diodes, transistors, coils, transformers, speakers and headphones, capacitors, switches, jumpers, cables and many other different electronic components. This method is a lot faster and straightforward than it is using some “off the shelf” instrument. Schematic for this device is on 12.1a. It is called a relaxation audio oscillator. When you connect points A and B using a piece of copper wire, a variable current flows through the transistors as sequences of impulses. This means that immediately upon connecting the points A and B, current level rapidly rises to some destined maximum value, and then drops to zero. For certain amount of time there is no current, after which it rises again rapidly, and whole cycle repeats itself. Since relation of times when current is flowing and when it is not is highly in favor of the later, this kind of current is called the spike impulse current. Collector current of a T2 transistor flows through the speaker which generates sound, whose base frequency could be calculated using this approximate equation.

$$f = \frac{1{,}6}{RC}.$$

In our case R=47 kOhm and C=47 nF, which means:

$$f = \frac{1{,}6}{47 \cdot 10^{3} \cdot 47 \cdot 10^{-9}} = 724 \text{ Hz.}$$

From the equation above, it is clear that varying of the frequency is possible by varying the resistor or capacitor value. Frequency rise is achieved by lowering the resistance or capacitance of the circuit, and vice versa, rising the values of the resistor or capacitor, lowers the oscillator frequency. Active variation of the frequency base is possible by replacing the resistor with a several hundred kiloohm trimmer potentiometer. If such

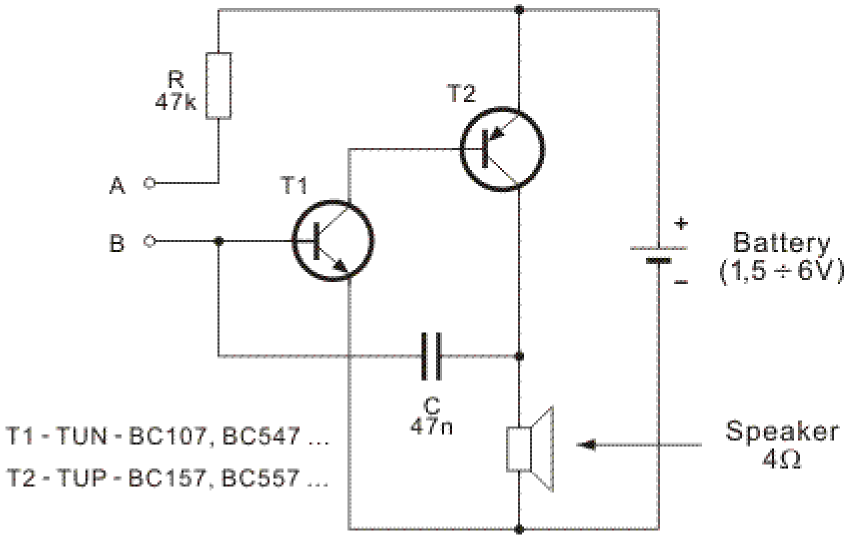

12.1. Schematic of a conductivity tester modification of the circuit was needed, special care must be taken not to set the trimmer into it's lowest position since this means zero resistance, and that could burn the transistors. To avoid unnecessary care and further complicating the operation of this straightforward device, low value resistor could be connected to the trimmer in series. This resistor would act as a protection for transistors inside of the circuit since it facilitates a minimum resistance, and thus doesn't leave transistors bare in the frying pan when the trimmer is in it's lowest position.

In this example we used an 1.5V battery for supply, but it is possible to plug this instrument on any battery between 1.5V and 9V.

Current flowing through the component that is being tested is lower than I=U/R, where U is the voltage of the supply battery, and R is the resistance of the resistor in the base circuit. In our example, these values are U=1,5 V i R=47 kW, which means that current flow is I=32 micro amperes, which is very low, so tested component is safe from harm from this device.

Oscillator's printed board design is on 12.2. This is viewed from the copper plated side of the board, components are placed on the other side, so their positions are marked in dotted lines.

Component side of the board is on 12.2.

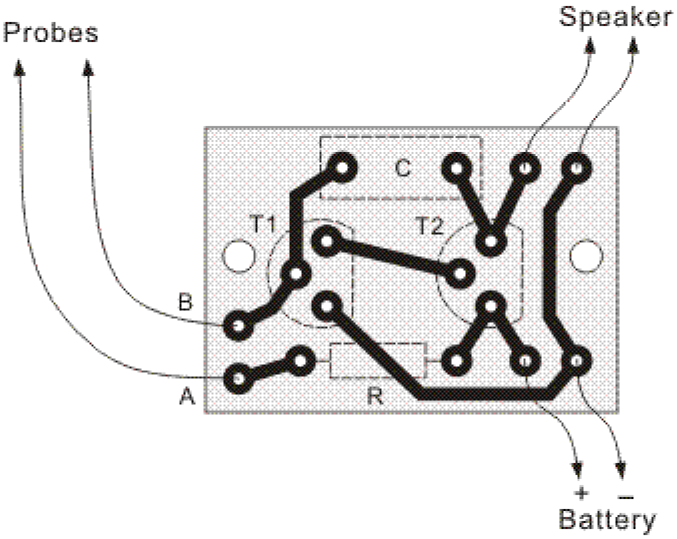

12.2. Tester's printed circuit board with components placement

Printed board, battery and the speaker are placed in a small box, as shown on 12.3. Miniature speaker is fixed to the upper pane of the box using two wood screws. It is connected to the circuit board using two threaded isolated wires. Same wires are used for all other connections as well. Battery is connected to the board using these wires, for example. In our example, wires are soldered directly to the poles of the battery, and the board fixed inside of the box using wood screws and two rectangular wooden pads glued to the bottom of the box, leaving just enough space to squeeze the battery in. These are not proper solutions, they are cheap “hacks” used when other options are limited. But these are functional for people who always have their trusty soldering iron at hand. What would be a proper solution? Buying a battery holder (with enough battery slots as needed) or battery clips (for those square 9V batteries) would simplify the process of changing the battery, although this circuit is very low in power consumption. Other thing is plastic or metal mounts for boards, these are pretty cheap and you should keep them at hand in your “junk box” when experimenting with electronics.On the front side of the box, we drilled two holes, one for the switch and the other for wires which hold probes on their ends. Probes are cheap components and come in various shapes and sizes with various purposes in mind. Since we've been applying dirty methods, like soldering the battery, there is no reason why we should back from building our own probes now. Any old marker-pen will do, just slip thicker copper wire through it's center, and sand/grind/cut protruding ends into a pointy tip. It is advisable to make probes in different colors, red and black are dominant standards for distinguishing them. Positive probe (red) is connected to point A, and negative one (black) is connected to the point B. You could use alligator clips instead of probes, for example, this would leave your hands free for other purposes, but for some precise testing of the on-board components, go with the more precise probes we already mentioned.

Give your new instrument

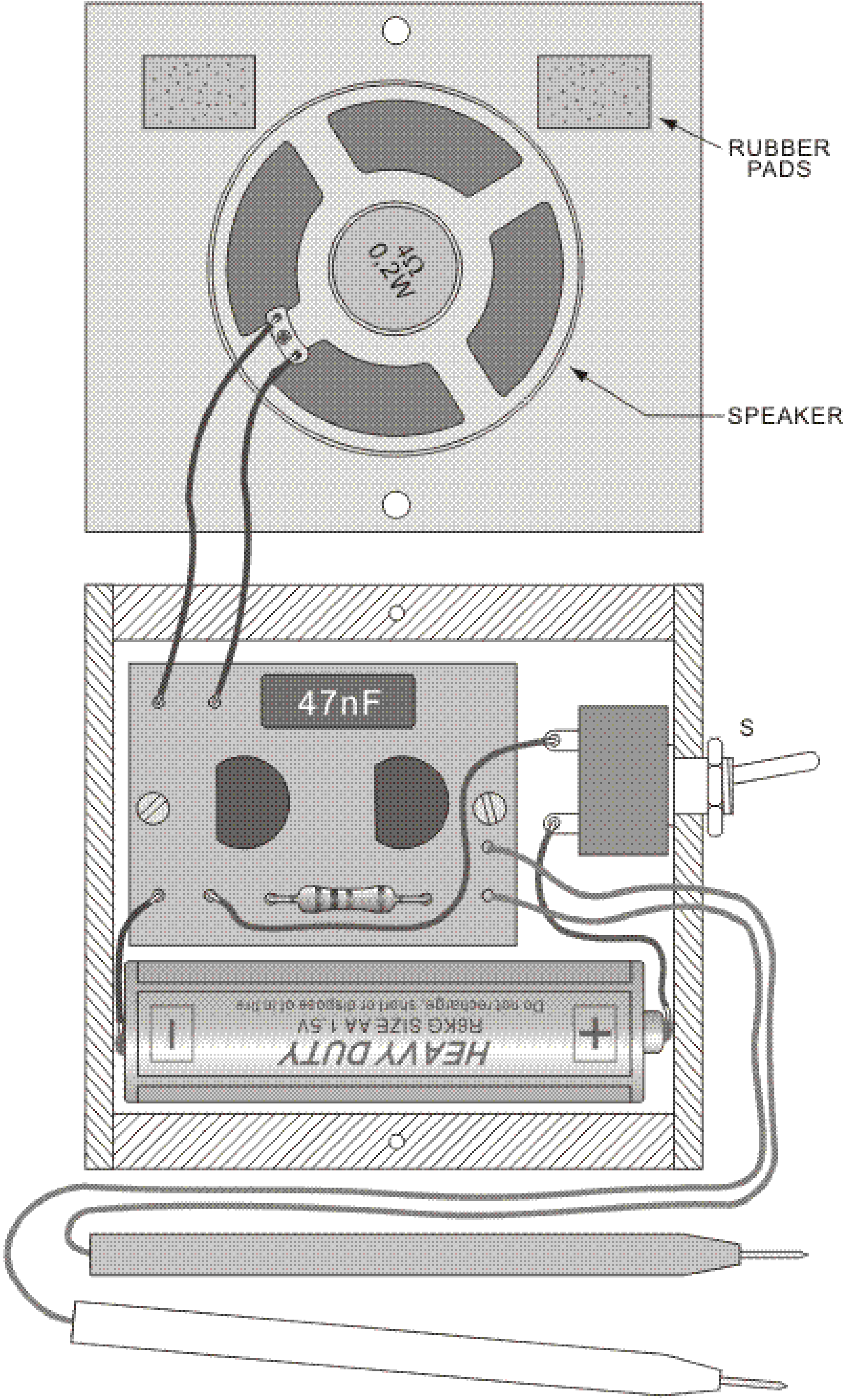

12.3. Conductivity tester

the initial self-test (battery might be empty, or some other unexpected thing happened) by connecting the probe tips together. If sound is heard from the speaker, everything is fine and ready for work.
Ok, everything is working, now you want to play with your new toy. Check, for example, conductivity of your own body. Hold probe tips between thumb and index finger of your left and right hand. What you hear is a sound whose level and especially frequency depend on your skin moistness. Wow, now instrument could be used as a very crude an inaccurate lie detector. This probably wouldn't be accepted in a court of law, but may be an interesting game you play on your friends. "Suspect" holds in his/hers hands probes which could be made of a metal pipe for this occasion. Pipe should be wide enough so that a large portion of palm surface is actually in contact with metal. When the suspect starts dodging questions or lying, his palms start sweating more than usual, and the tone produced by our "lie detector" is higher than usual.

12.1 Semiconductors check

To test diodes using this circuit, we fall back to the diode theory of operation: when anode is positive comparing to the cathode (red probe on anode, black on cathode), whole diode acts as a low value resistor, which means that speaker sound is higher than usual. On the other hand, in the opposite direction, sound is lower because in that direction diode acts as a high value resistor. Testing process is shown on 12.4.

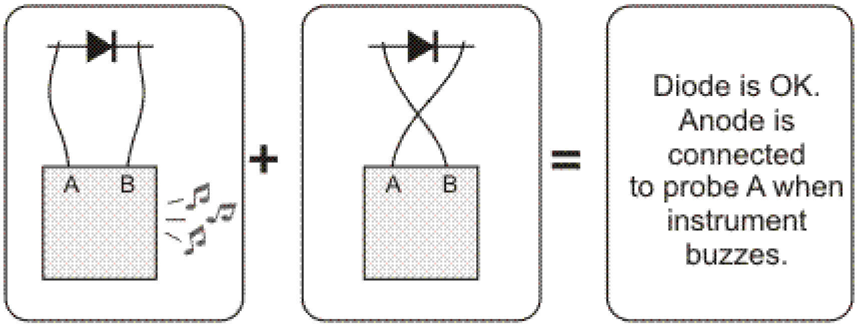

12.4. Diode testing

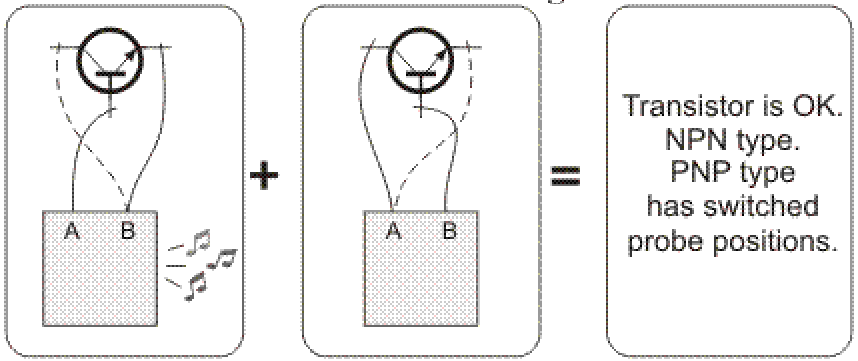

12.5. Bipolar tranistor testing

DC transistor acts in the same fashion as two connected diodes (11.4a). If both diodes are functional, transistor is functional as well as shown on 12.5. As you can see, probe A is

connected to the base, and then probe B is connected first to the emitter, and then to the collector. In both cases, if the transistor is ok, “music” would have been heard. We then switch probe connections, A goes where B was connected to and vice versa, if there is no music now, everything is in order. So, transistor is faulty if speaker remains silent in the first two measurements, or if it “plays” in one of the second two measurements.
FET testing is done in similar fashion as testing the bipolar transistors, which is shown on 12.6.

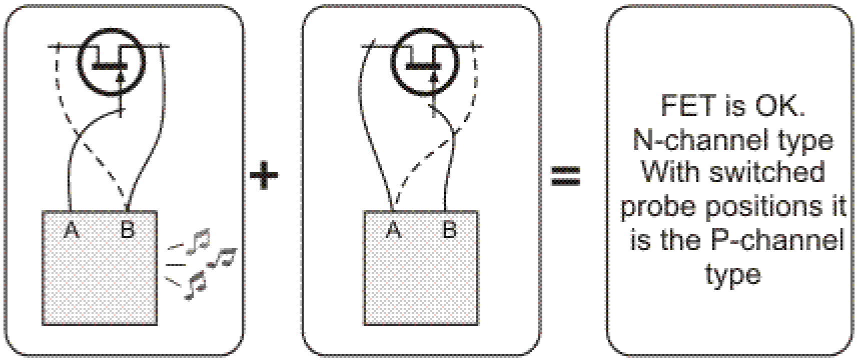

12.6. FET testing

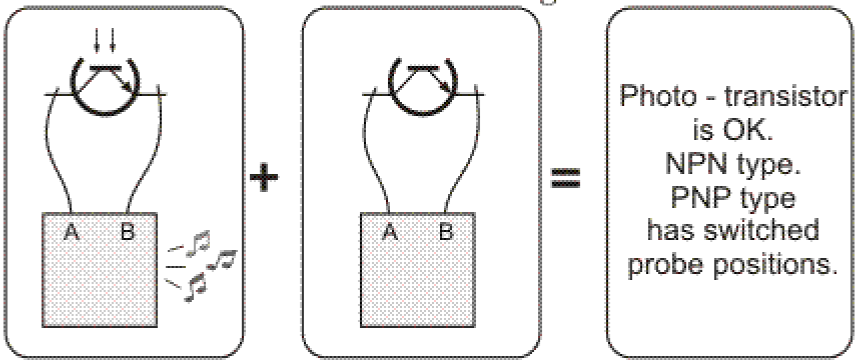

12.7. Photo transistor testing

One principle that is applicable when testing the photo resistors, photo transistors and diodes is NL-NM (or, No Light – No Music). Probe A is connected to the collector of the transistor, or diode's anode or one side of the photo resistor, and the other one is connected to transistor's emitter or diode's cathode or the other resistor's side and some kind of sound should be heard from the speaker. If this continues when the component is shadowed using your palm, everything is in functional order. We displayed graphically the method of testing photo sensitive components on 12.7.

12.2 Checking other components

Many other components may be tested using this instrument. Base rule is: if component is intended to conduct electricity, sound will be heard. This is the case with resistors, coils, transformers, fuses, closed switches. If component doesn't conduct electricity, like capacitors, or open switches, or two copper wires on the circuit board which shouldn't be connected, then music would have not been heard.

When testing different resistors, it is apparent that different resistance values give different output sound. So with some experience using this instrument on various resistors it will be possible to tell the resistance of the resistor in question from only the generated sound. This may be easier and more accurately done using regular ohmmeter on your multimeter, but your nerd level will certainly rise sky high if you are able to tell resistor's value from bare sound.
Components which have coils in them, like different electro motors, headphones, speakers, transformers and such conduct electricity, so absence of sound while testing tells of some coil connection failure. With transformers with several secondary coils there is a possibility to find beginning and the end of each of them. And from the sound frequency one is possible to tell which coil is primary and which is secondary.
Functional capacitor will generate no music. An exception are electrolithic and block capacitors, especially the larger ones. Tone generated by connecting these capacitors to the instrument will change in level and frequency and fade until completely off when capacitor is discharged. Length of playing depends on the capacitance of the component, where higher values give longer sound time, which allows for a crude approximation of the component's capacitance.

The end

Issue date:14th June 2012

FREE-ENERGY:

NIKOLA TESLA SECRETS FOR EVERYBODY

by **Vladimir Utkin** u.v@bk.ru

FIRST SECRET

All of Tesla's secrets are based on

ELECTROMAGNETIC FEEDBACK

EXPLANATION: An ordinary energy system comprises a generator and motor (common view), and can be completed with an electric current feedback as shown here in electrical circuit (a)

NO FREE-ENERGY

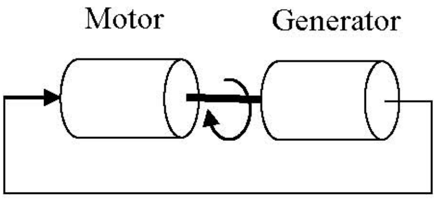

(a)

Electrical feedback

FREE ENERGY IS POSSIBLE

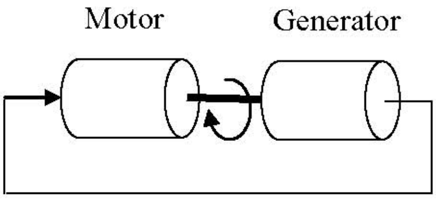

(b)

ElectroMagnetic field feedback

In case (a), the system once started, will slow down and stop because of friction, resistance and so on. Nikola Tesla arranged a feedback loop for the electromagnetic field: case (b), and he said:

ELECTROMAGNETIC FIELD FEEDBACK DESTROYS THE INTERACTION SYMMETRY

This means that an action no longer has an equal and opposite reaction

In case (b), once started, the system will accelerate in spite of friction, resistance and so on (provided that the phase of the electromagnetic feedback is positive and is sufficiently large). In order for an electromagnetic field to exist in a motor, there must be some energy input, and Tesla said:

ENERGY GENERATION BY IT'S OWN APPLICATION

QUESTION: How can you produce positive electromagnetic field feedback?

AN ANSWER: The simplest and well-known example is Michael Faraday's unipolar motor, as modified by Nikola Tesla:

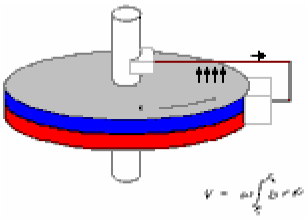

(a)

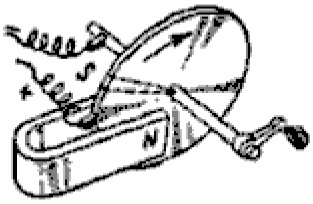

(b)

An ordinary unipolar motor consists of a magnetised disk, and a voltage applied between the axis and a point on the circumference of the disc as shown in (a) above. But an ordinary unipolar motor can also consists of an external magnet and a metal disc with a voltage applied between the axis and a peripheral point on the disc as in (b) above. Tesla decided to modify this version of the unipolar motor. He cut the metal disc into helical sections as shown here:

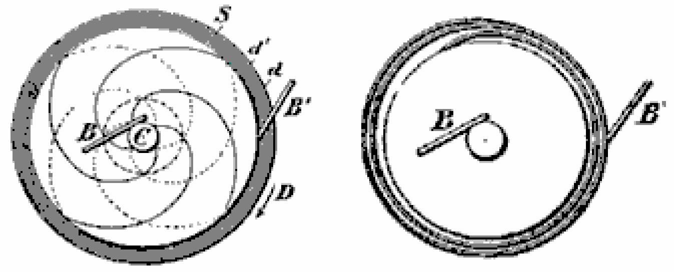

In this case, the consumption of current produces an additional magnetic field along the axis of the disc. When the current-carrying wires are tilted in one direction, their magnetic field augments the main external magnetic field. When the wires are tilted in the other direction, their magnetic field reduces the main external magnetic field. So, the current flow can increase or reduce the external magnetic field of the unipolar motor.

Amplification is not possible without applying power

If it is possible to arrange a magnetic field feedback loop for mechanical devices, then it is probably possible to arrange it for solid-state devices like coils and capacitors.

The others parts of this article are devoted to devices which use coils and capacitors. All of the examples in this article are only intended to help your understanding of the principles involved. Understanding would be made easier if we pay attention to the ferromagnetic shielding of the second coil in the transformer invented by Nikola Tesla:

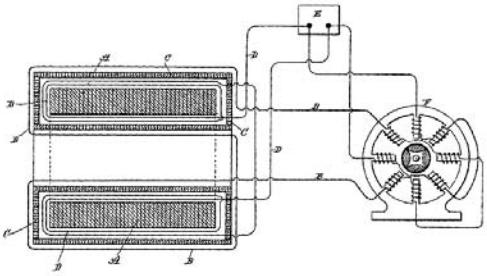

In this case, the ferromagnetic shield separates the first and second coils in the transformer from each other, and that shield can be used as magnetic field feedback loop. This fact will be useful for understanding the final part of this article. It is also helpful to consider the properties of the electrostatic field.

ELECTROSTATICS

(scalar field and the longitudinal electromagnetic waves)

Comment: Mr. Tesla said, “there is radiant energy, perpendicular to the surface of any charged conductor, produced by a scalar electromagnetic field, thus giving rise to longitudinal electromagnetic waves”.

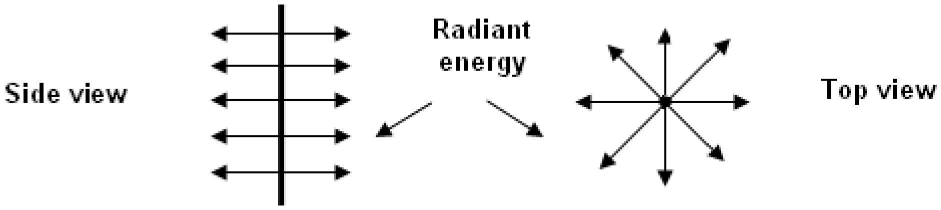

At first glance, this contradicts the age-old experience in studying the electromagnetic field (according to modern concepts, any electromagnetic field has components which are perpendicular to the direction of the propagated electromagnetic wave), also, Maxwell's equations describe an electromagnetic field as a vector. However, the first impression is erroneous, and no contradiction exists.

Definitions of Physics: Any conductor has both inductance and capacitance, that is, the ability to accumulate charge on it's surface. A charge on the surface of a conductor creates an electric field (electrostatic field). The potential (voltage) at any point of the electric field is a scalar quantity!!! (That is, it is a scalar electric field ...).

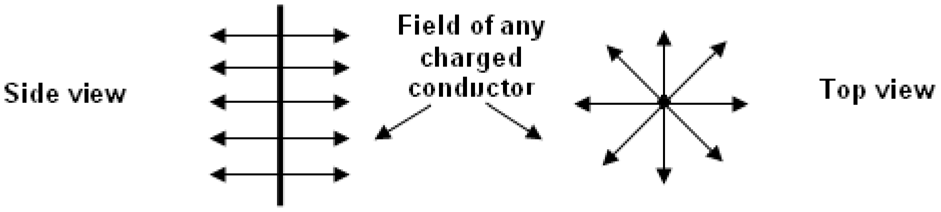

If the electric charge of the conductor varies with time, then the electrostatic field will also vary with time, resulting in the appearance of the magnetic field component:

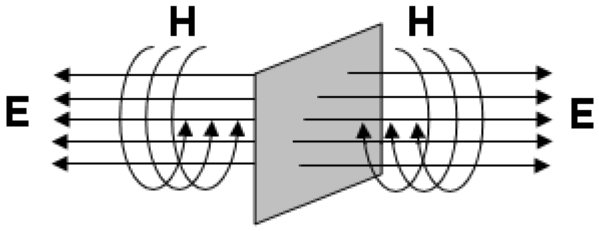

Thus, the electromagnetic wave is formed (with the longitudinal component of E ...).

REMARK: In order to understand how a longitudinal wave interacts with conductive bodies, one needs to read the section of electrostatics entitled "Electrification by Influence". Particularly interesting are Maxwell's equations where they mention the displacement current.

Now we come to the first secret:

SECRET 1

The power source in Nikola Tesla's free energy device, the amplifying transformer, is a

SELF-POWERED L-C CIRCUIT

EXPLANATIONS

An ordinary LC circuit – with decay

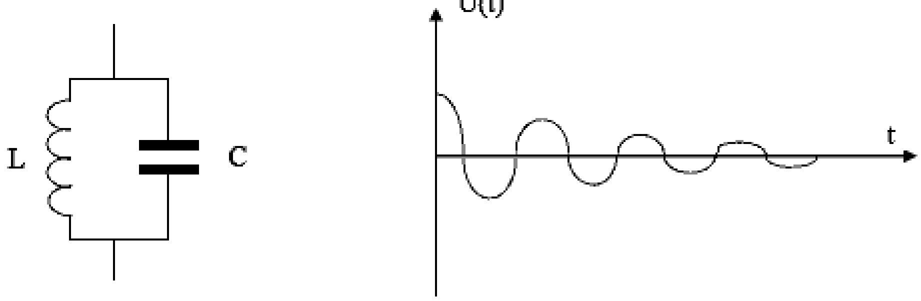

Nikola Tesla LC circuit – with amplification

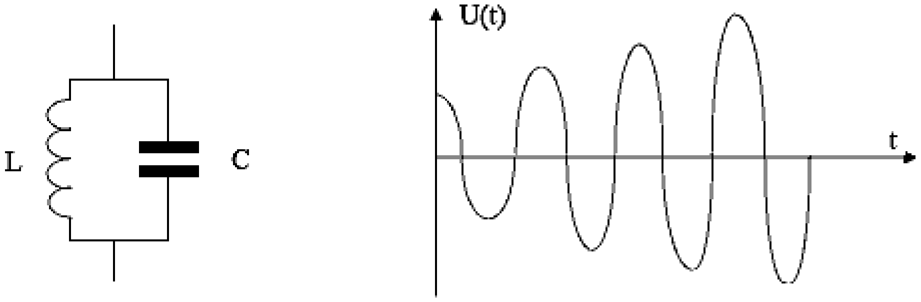

AN EXAMPLE OF UNLIMITED VOLTAGE RISE

(Based on batteries and a switch)

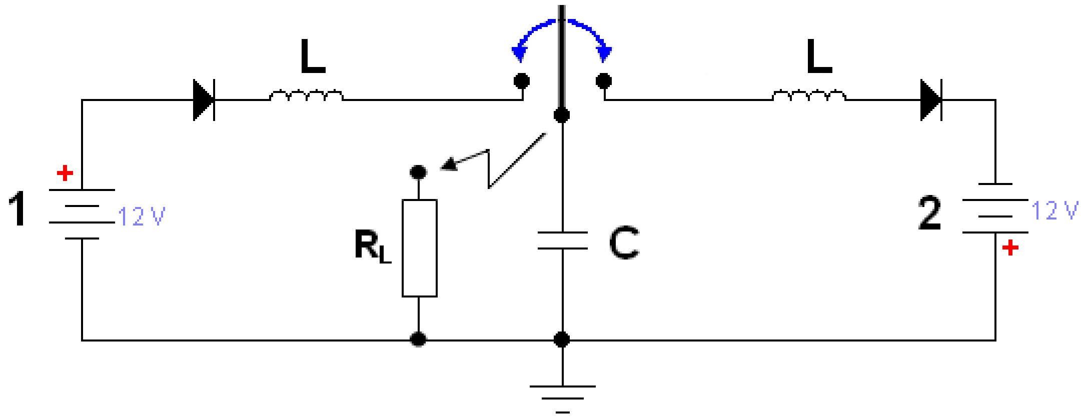

EXPLANATION: Batteries 1 and 2 are connected to the capacitor C alternately, through the inductances L. Voltage on capacitor C and the voltage from the batteries are increasing. As a result, there can be unlimited voltage rise. When the voltage on the capacitor reaches the desired level, it is connected to the load.

COMMENT: Two diodes were used to avoid synchronisation requirements. Manual or relay switching can be used. One implementation used a spark gap to connect the output load but a switch is an alternative method.

TIME LINE FOR THE PROCESS:

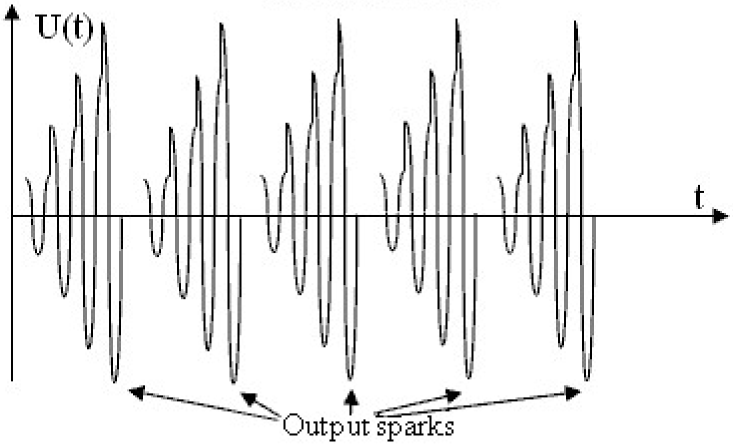

The schematics can be simplified, and only one battery used (load is connected in the same way).

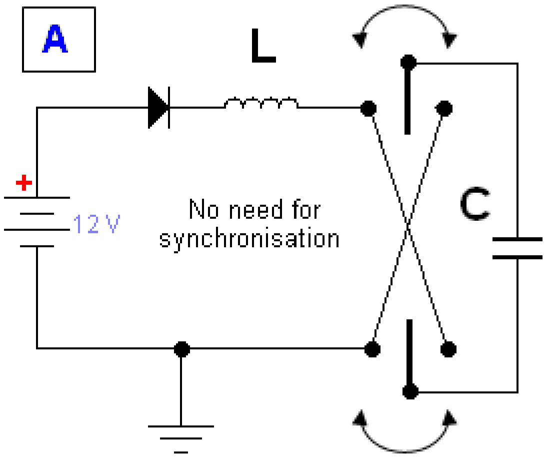
A
L
+
12 V
No need for synchronisation
C

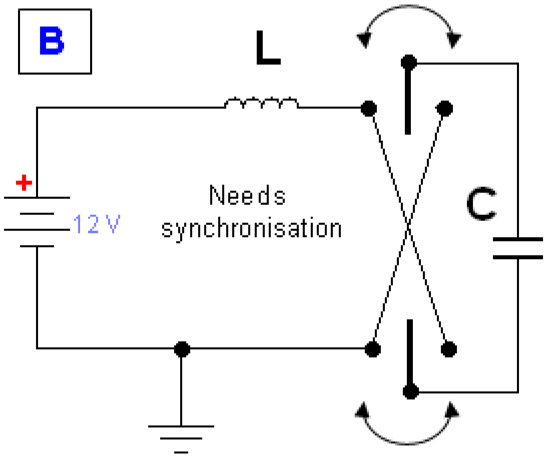
B
L
+
12 V
Needs synchronisation
C

COMMENT: Maybe Alfred Hubbard used an idea shown as option B, in some versions of his transformer

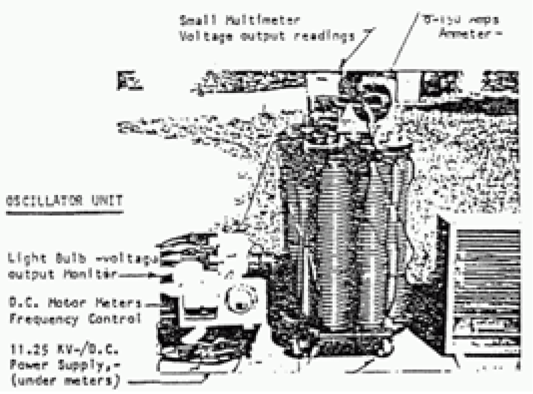
Small Multimeter
Voltage output readings
0-150 amps
Ammeter
OSCILLATOR UNIT
Light Bulb voltage output Monitor
D.C. Motor Meters Frequency Control
11.25 KV-/D.C. Power Supply,-
(under meters)

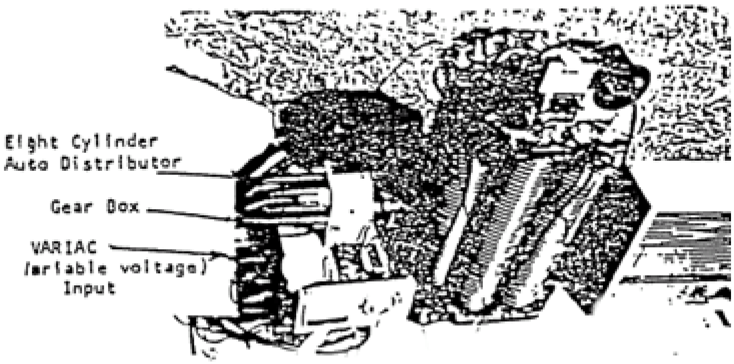
Eight Cylinder
Auto Distributor
Gear Box
VARIAC
(variable voltage)
Input

COMMENT: If you want to get a self-powered circuit, you have to arrange some kind of energy feedback to the batteries. But, is this an actual Free-Energy technology? I am not sure….

COMMENT: Is this the only possible way?
No, of course not - there are different ways of doing it. For example, you can use fields inside and outside of some LC circuits. How can we do that?

For more secrets read the following parts…

HOW DO WE GET THIS RESULT?

AN ANSWER

You need to charge the capacitor using the electric component of the electromagnetic field of the inductor (using the displacement current of Maxwell's equations)

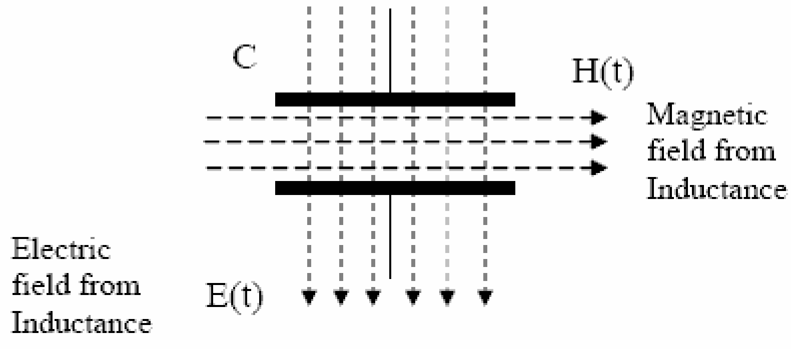
C
H(t)
Magnetic field from Inductance
Electric field from Inductance
E(t)

EXPLANATION

When the electric field in capacitor C is decaying, due to feeding electrical current into an inductor (not shown), the external electric field generated by the inductor tries to charge this capacitor with the inductor's displacement current. As a result, the capacitor draws energy in from the surrounding electromagnetic field, and the capacitor's voltage rises cycle by cycle.

IMPLEMENTATION A – a central capacitor is used:

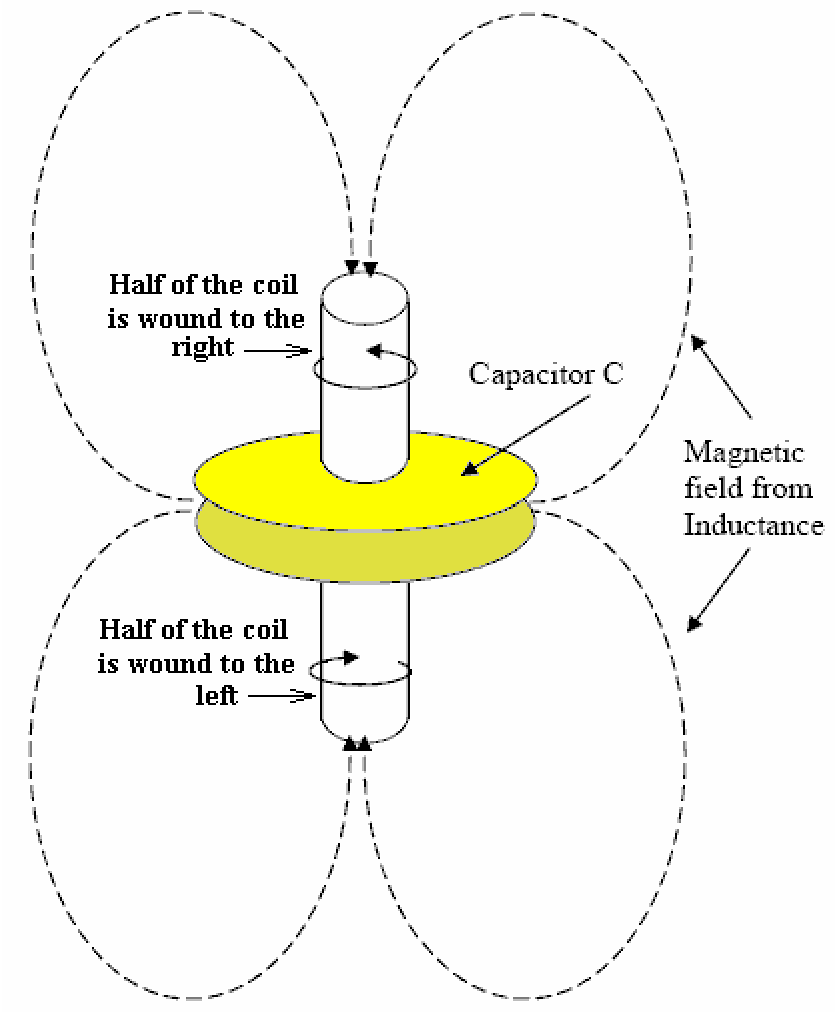
Half of the coil is wound to the right
Capacitor C
Magnetic field from Inductance
Half of the coil is wound to the left

IMPLEMENTATION B – no capacitors are used:

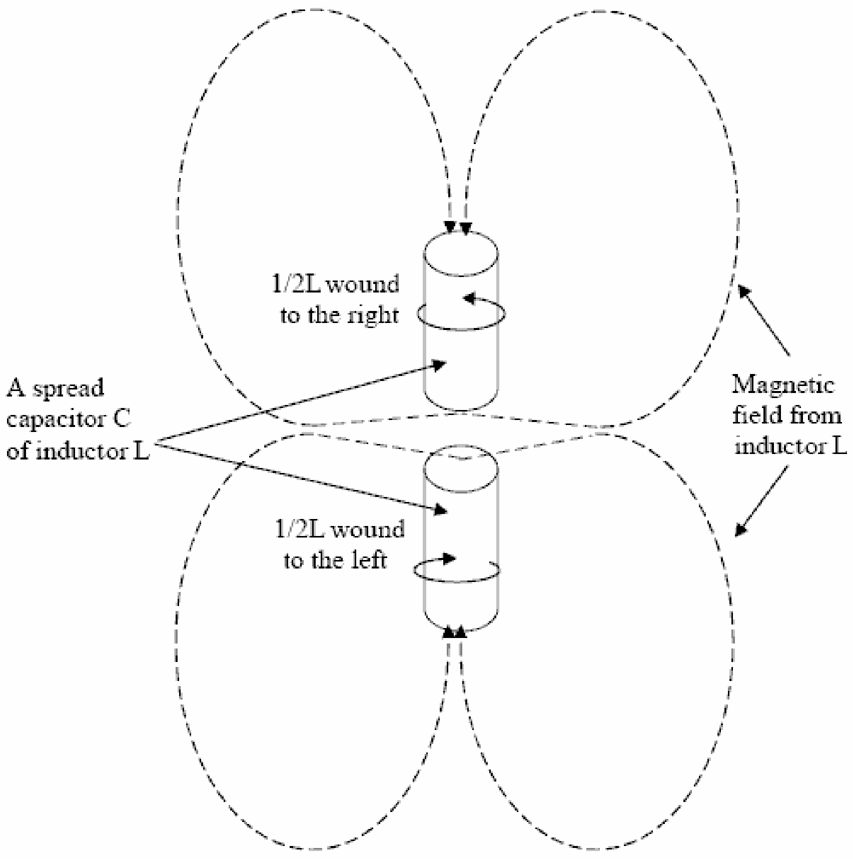

1/2L wound to the right
A spread capacitor C of inductor L
Magnetic field from inductor L
1/2L wound to the left

In this case instead of using a capacitor, the capacitance between the two sections of inductor L provides the necessary capacitance.

HOW DO WE START THE PROCESS?

In implementation A, you must charge the capacitor and connect it to the inductor to start the process.

In implementation B, you must use an additional pulsing or “kicking” coil, which starts the process by providing a pulse in either the electrical field or the magnetic field (shown later on).

HOW DO WE STOP THE PROCESS?

The process of pumping energy can continue uninterrupted for an unlimited length of time and so the question arises; how do you stop the device if you should want to?. This can be done by connecting a spark gap across the coil L and the resulting sparking will be sufficient to stop the process.

THE “KICKING” PROCESS USING AN ELECTRIC FIELD

Use an additional special “kicking” coil, which can generate short powerful magnetic pulses, and install an amplifying Tesla coil along the electrical vector of the electromagnetic field of this coil.

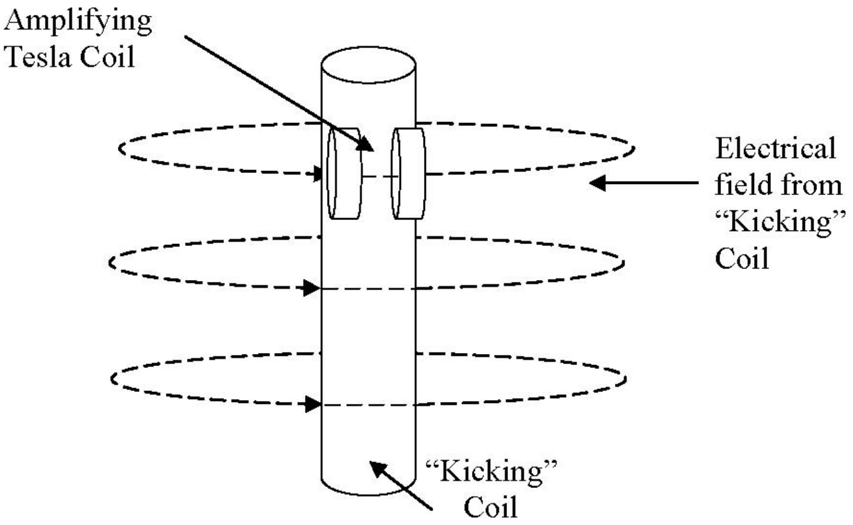
Amplifying
Tesla Coil
Electrical
field from
"Kicking"
Coil
"Kicking"
Coil

The electrical field of the driving pulse or "kicking" coil will charge the spread capacitors of the inductor, and the process will be started. Use pulses as short as possible in "kicking" coil, because the displacement current depends on the speed of the changes in the magnetic field.

THE "KICKING" PROCESS WITH A MAGNETIC FIELD

It is not possible to "kick" the process by displacement of the amplifying Tesla coil in the uniform changing magnetic field of the "kicking" coil, because the output voltage on the ends of the Tesla amplifying coil will be equal to zero in this case. So, you must use a non-uniform magnetic field. For that you must install a "kicking" coil, not in the centre of the amplifying Tesla coil, but positioned away from the centre

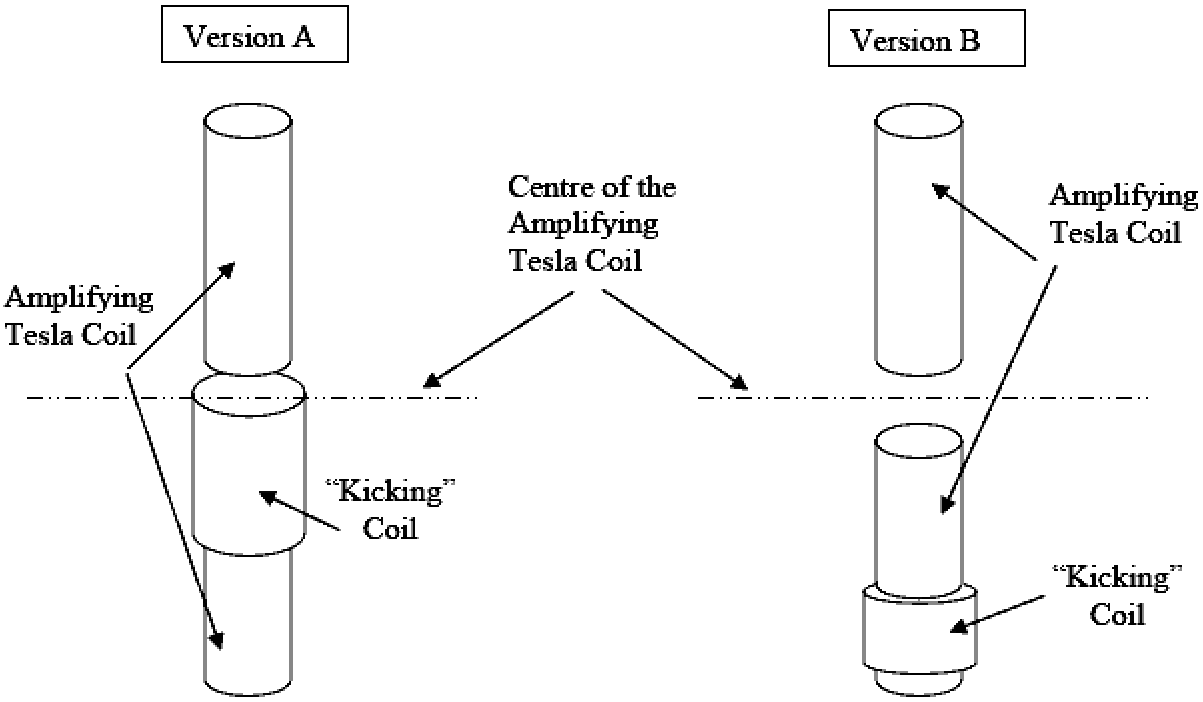
Version A
Version B
Centre of the
Amplifying
Tesla Coil
Amplifying
Tesla Coil
"Kicking"
Coil
Amplifying
Tesla Coil
"Kicking"
Coil

IS THAT ALL TRUE, AND THE BEST TECHNIQUE TO USE?

No, it is not! Nikola Tesla found more subtle and more powerful method – his bi-filar pancake coil!

BI-FILAR PANCAKE COIL – MAY BE THE BEST METHOD

The voltage between adjacent turns in an ordinary coil is very low, and so their ability to generate additional energy is not good. Consequently, you need to raise the voltage between adjacent turns in an inductor.

Method: divide the inductor into separate parts, and position the turns of the first part in between the turns of the second part, and then connect end of the first coil to the beginning of the second coil. **When you do that, the voltage between adjacent turns will be the same as the voltage between the ends of the whole coil !!!**

Next step – rearrange the position of the magnetic and electric fields in the way needed for applying amplifying energy (as described above). The method for doing this is – the flat pancake coil **where the magnetic and electric fields are arranged in exactly the way needed for amplifying energy.**

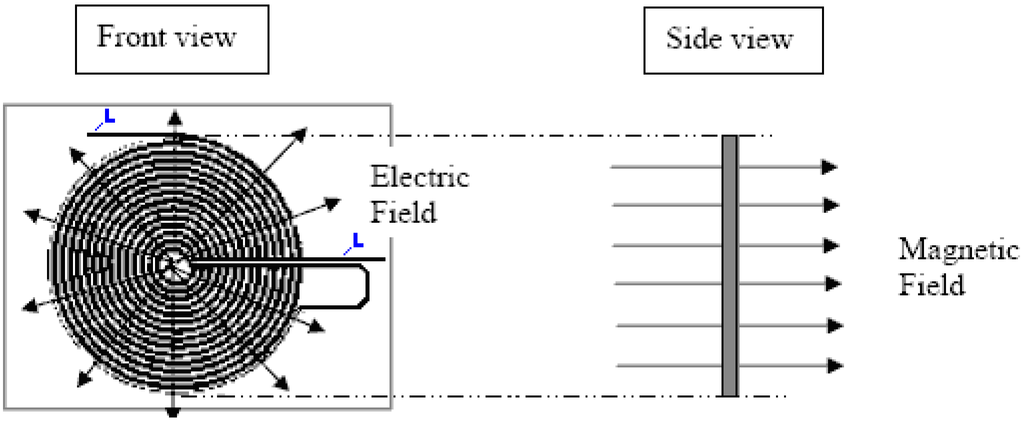

Coil for electromagnets

Now, it is clear why Tesla always said that his bi-filar pancake coil was an energy-amplifying coil !!!

REMARK: for the best charging of the natural self-capacitance of the coil, you have to use electric pulses which are as short as possible, because the displacement current as shown in Maxwell's equation, depends to a major degree on the speed of the change in the magnetic field.

THE DUAL - LAYER CYLINDRICAL BI-FILAR COIL

Instead of the standard side-by-side cylindrical bi-filar coil, the coil winding may also be arranged in two separate layers, one on top of the other:

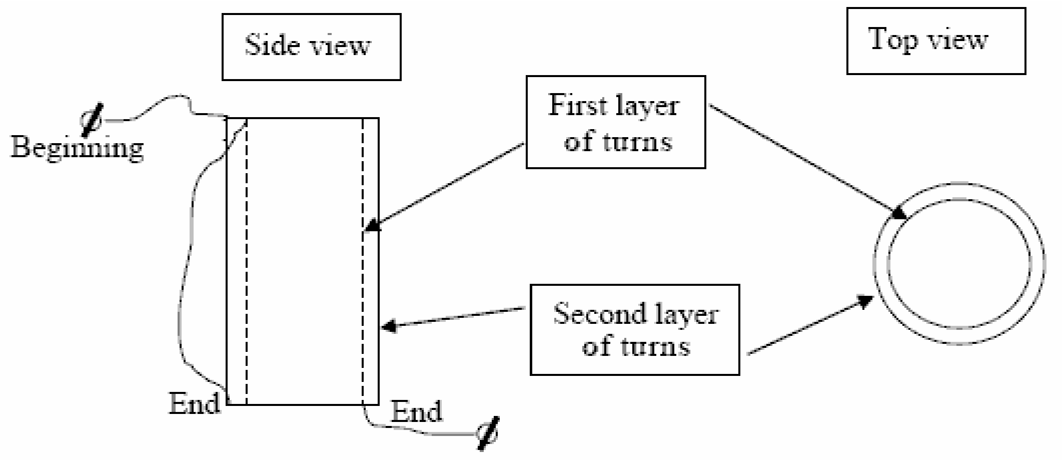

THE ELECTRO – RADIANT EFFECT

(Inductance in an electrostatic field)

EXPLANATION

The primary coil in Tesla's transformer is the first plate of the capacitor. The secondary coil - is the second plate of the capacitor.

When you charge a capacitor C from your source of energy, you charge a wire of the primary coil also. As a result, a wire of the secondary coil is charging also (as a return from ambient space).

In order to start the process, you have to remove charge from the primary coil (by arranging a jump in potential in ambient space). When this is done, a huge displacement current occurs – as a result of that potential jump. Inductance catches this magnetic flux, and you have energy amplification.

If this process is operating, then you generate a magnetic field in ambient space.

COMMENT: The capacitance of the wire of the primary coil is very low, and so it takes very little energy to charge it, and a very short spark to discharge it (without removing charge from the capacitor C).

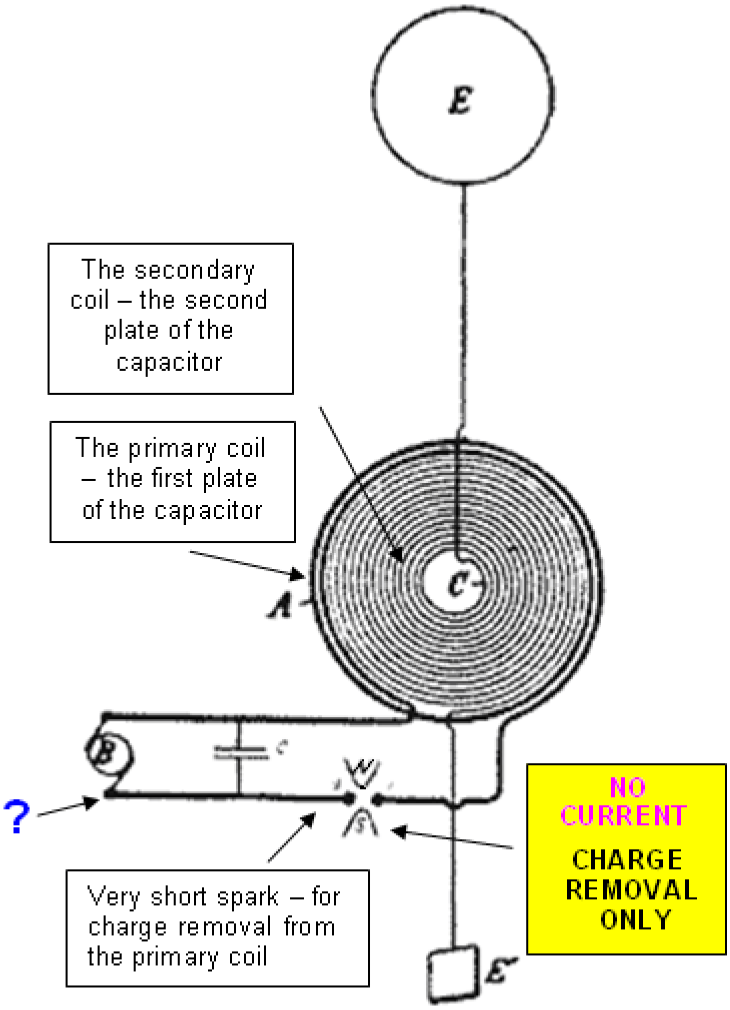

The secondary coil – the second plate of the capacitor

The primary coil – the first plate of the capacitor

Very short spark – for charge removal from the primary coil

NO CURRENT

CHARGE REMOVAL ONLY

Tesla's Magnifying Transmitter

COMMENT: Notice that the spark gap **must be** connected to the ground as, in my opinion, this is a very important feature of this process, **but Mr Tesla did not show grounding**. Perhaps this needs to be a separate grounding point.

REMARK: In my opinion, this technology was also used in Gray's device and in Smith's devices and in both cases **the spark gap was connected to the ground**.

ALSO:

Pay attention to the words used in Gray's patent "…. for inductive load".

And, pay attention to Smith's words "I can see this magnetic field, if I use a magnetometer".

MODERN IMPLEMENTATIONS

in self-powered L-C circuits

EXAMPLE 1

Using a bi-filar coil as the primary coil in a resonant Tesla transformer

By Don Smith

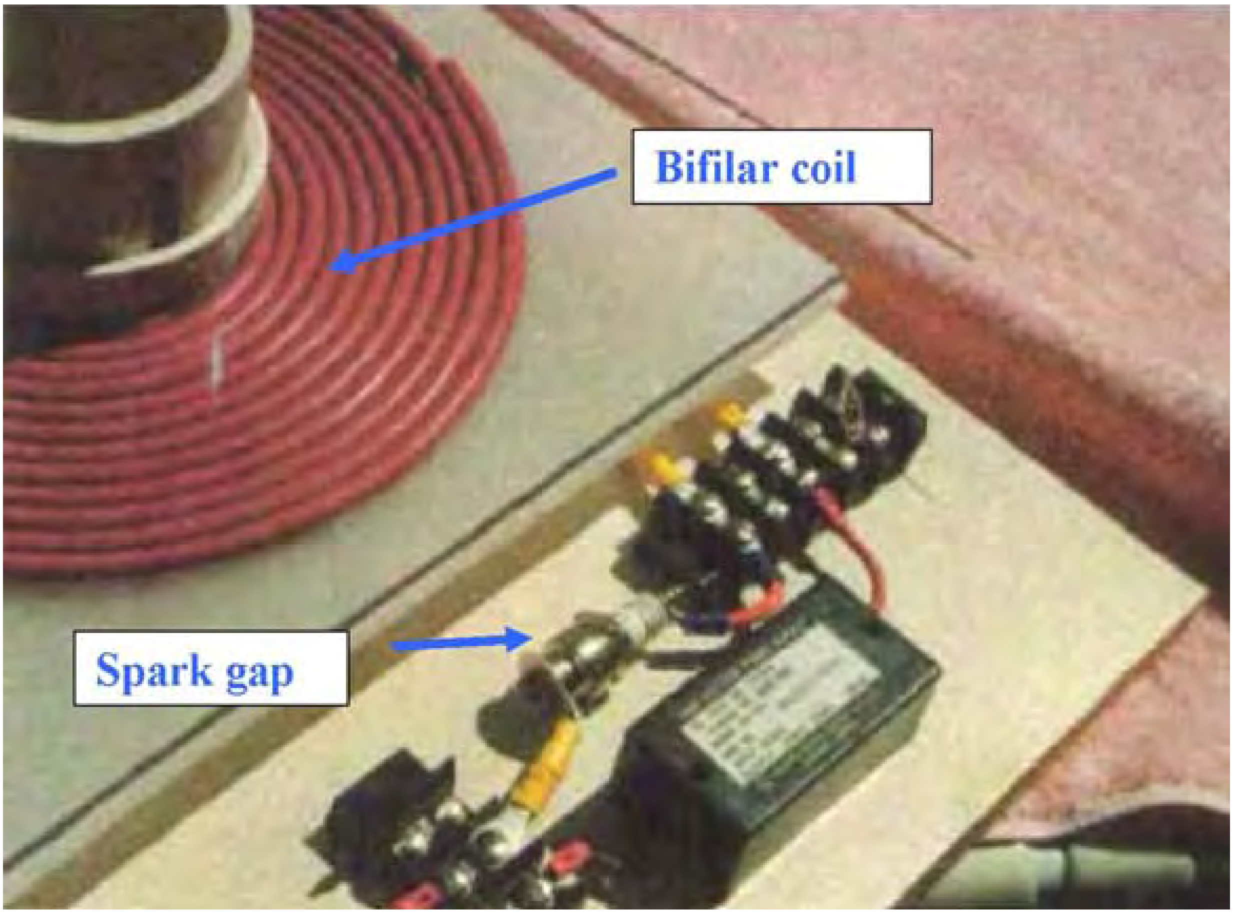
Bifilar coil
Spark gap

Explanation: The bi-filar primary coil is used as primary for energy amplification, and is pulsed through the spark gap.

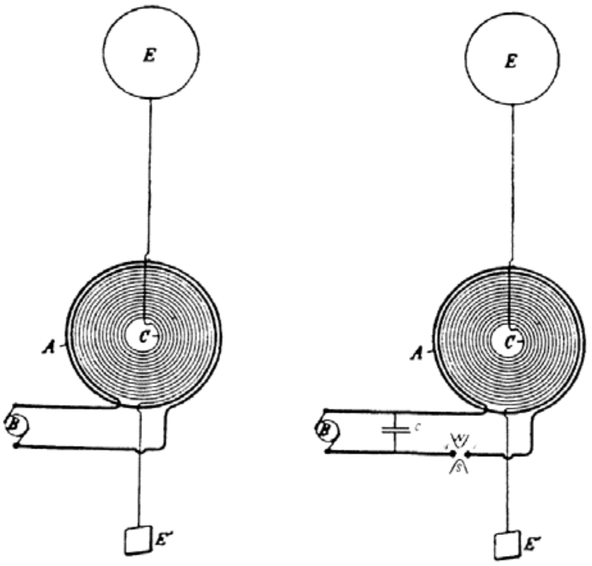
E
A
C
B
E'
E
A
C
B
c
W
S
E'

Tesla's Magnifying Transmitter

EXAMPLE 2

By Mislavskij

Is comprised of two capacitor plates sandwiching a ferrite ring core with a coil wound on it:

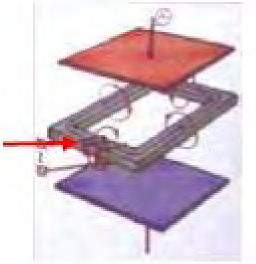

EXPLANATION

When a capacitor is charging (or discharging), this “displacement” current flow generates a magnetic field in the vacuum in a circular form (Maxwell’s equations). If a coil is wound on a ferrite toroid placed between the plates of the capacitor, then a voltage is generated in the turns of that coil:

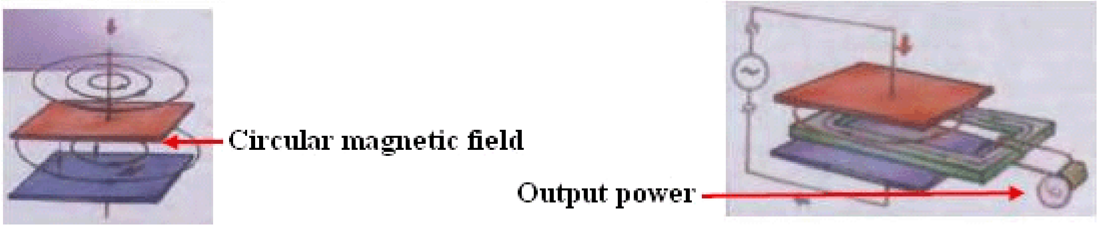

Also, if an alternating current is applied to the coil wound on the ferrite toroid, then voltage is generated on the capacitor plates.

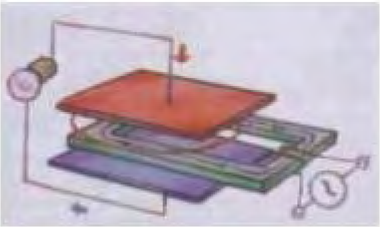

If an inductor and a capacitor are combined in an L-C circuit, then there are two cases inside such an L-C circuit:

a) energy amplification and **b) energy destruction**

The situation depends on how the coils and capacitor are connected together

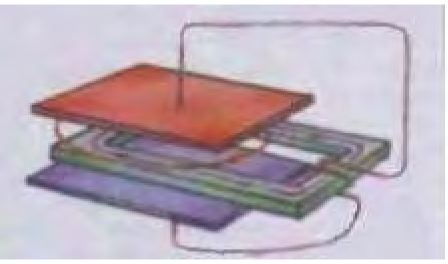

Energy Generation

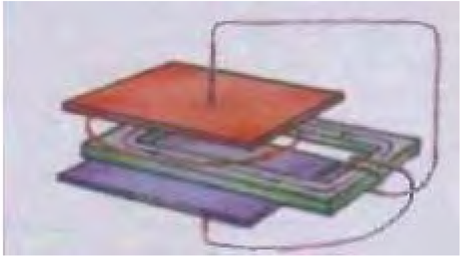

Energy Destruction

COMMENT: If the direction of the turns in the coil wound on the ferrite core is reversed, then the wires connecting the coil to the capacitor plates need to be swapped over as well.

The first experiments with a ferrite core inside a capacitor were made in 1992 by Mislavskij (a 7th-year pupil of the Moscow school), and so it is known as “Mislavskij’s transformer”.

PROTOTYPE TRANSFORMER:

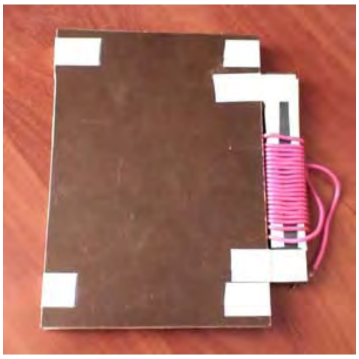

THE SAME APPROACH?

By Don Smith

In this arrangement, the capacitor is charged by sparks and powerful displacement current is produced. The transformer with the ferromagnetic core is collecting this current.

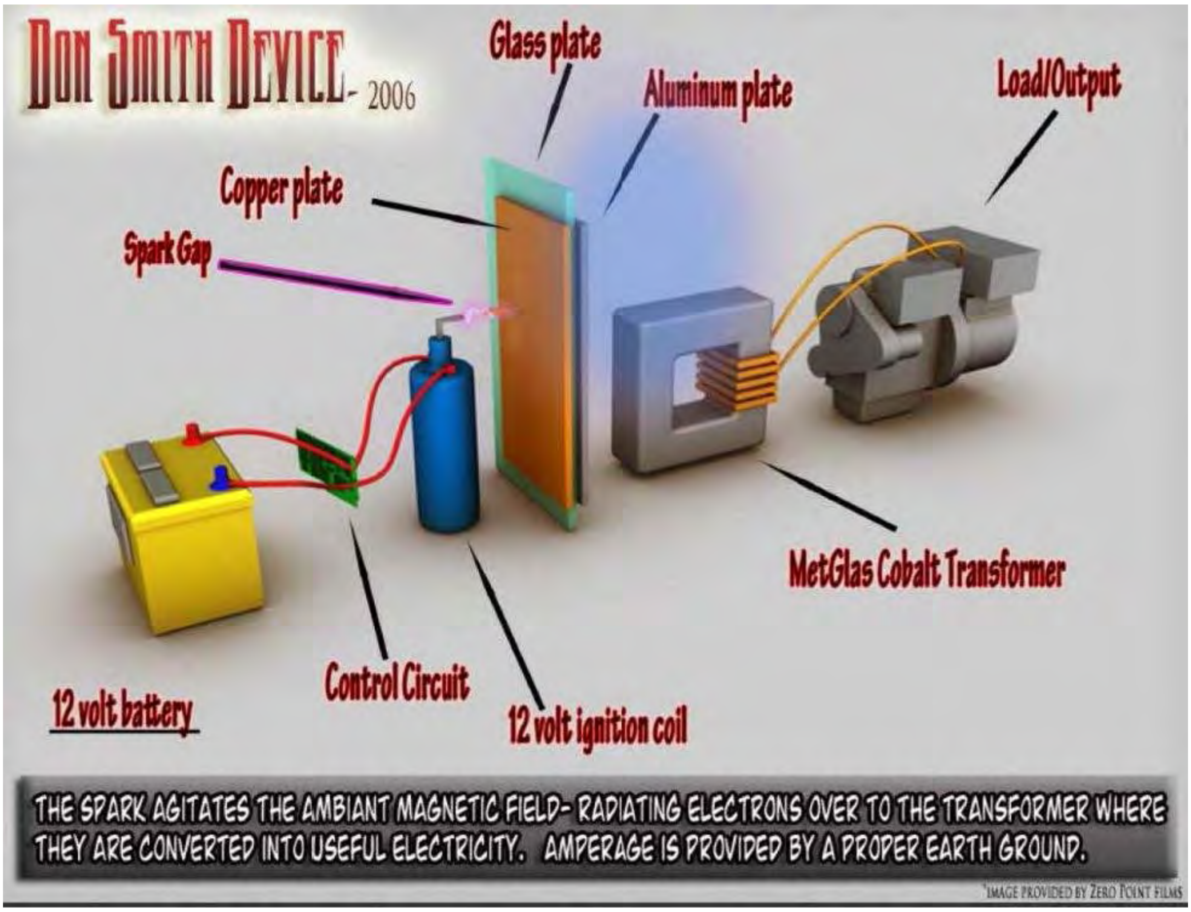
DON SMITH DEVICE - 2006
Glass plate
Aluminum plate
Load/Output
Copper plate
Spark Gap
MetGlas Cobalt Transformer
12 volt battery
Control Circuit
12 volt ignition coil

THE SPARK AGITATES THE AMBIANT MAGNETIC FIELD- RADIATING ELECTRONS OVER TO THE TRANSFORMER WHERE THEY ARE CONVERTED INTO USEFUL ELECTRICITY. AMPERAGE IS PROVIDED BY A PROPER EARTH GROUND.

*IMAGE PROVIDED BY ZERO POINT FILMS

COMMENT: This schematic diagram is very rough, and lacking in details. It will not perform correctly without back-electromagnetic force suppression of some kind (see below).

SECRET 1.1

Back-EMF suppression in a resonating Tesla coil

Version 1

The primary and secondary coils, and the ground connection in this Tesla coil are arranged in special manner:

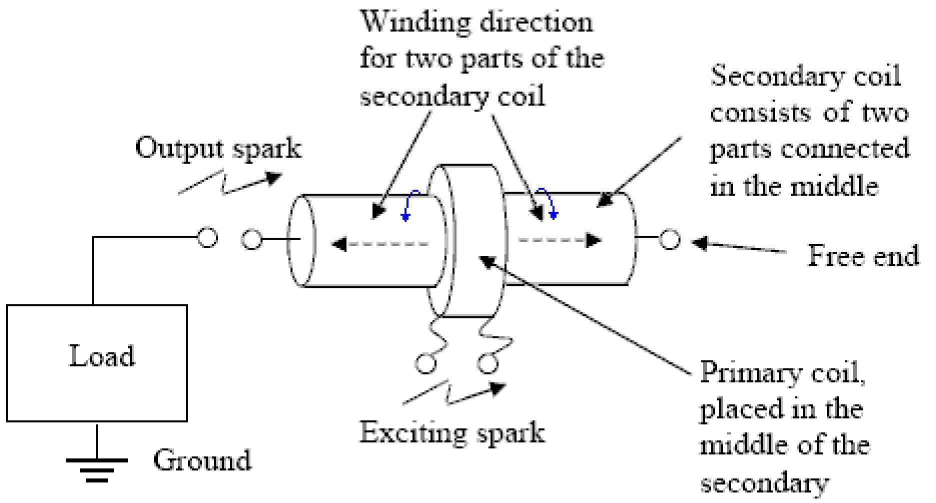
Winding direction for two parts of the secondary coil
Output spark
Secondary coil consists of two parts connected in the middle
Free end
Load
Ground
Exciting spark
Primary coil, placed in the middle of the secondary

Explanation: The exciting (driving) current and the load current in an electromagnetic field, are perpendicular to each other as shown here:

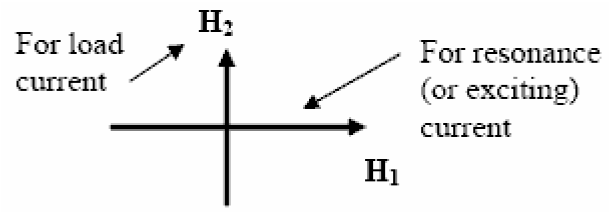
For load current
H_2
For resonance (or exciting) current
H_1

COMMENT: In order to get an energy gain, the frequency of excitation of the primary coil must be the resonant frequency of the secondary coil.

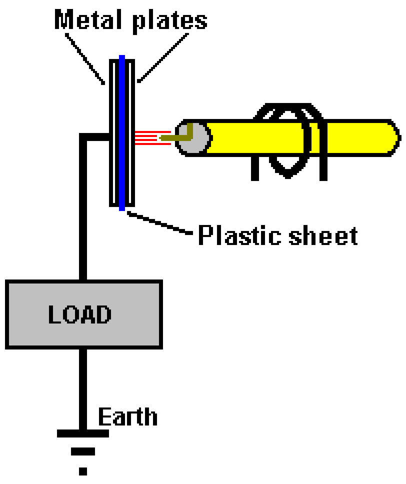
Metal plates
Plastic sheet
LOAD
Earth

Don Smith

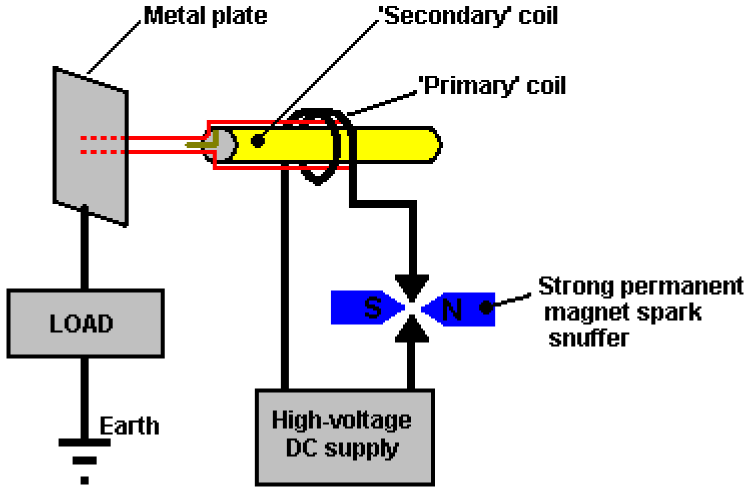
Metal plate
'Secondary' coil
'Primary' coil
S
N
Strong permanent magnet spark snuffer
LOAD
Earth
High-voltage DC supply

Nikola Tesla

COMMENT: Excitation with just a single spark is possible.

COMMENT: In Mr. Tesla's terminology, this is pumping charges or charge funneling, the charge is coming from the ground (which is a source of energy).

POTENTIAL (VOLTAGE) DISTRIBUTION ON THE COIL

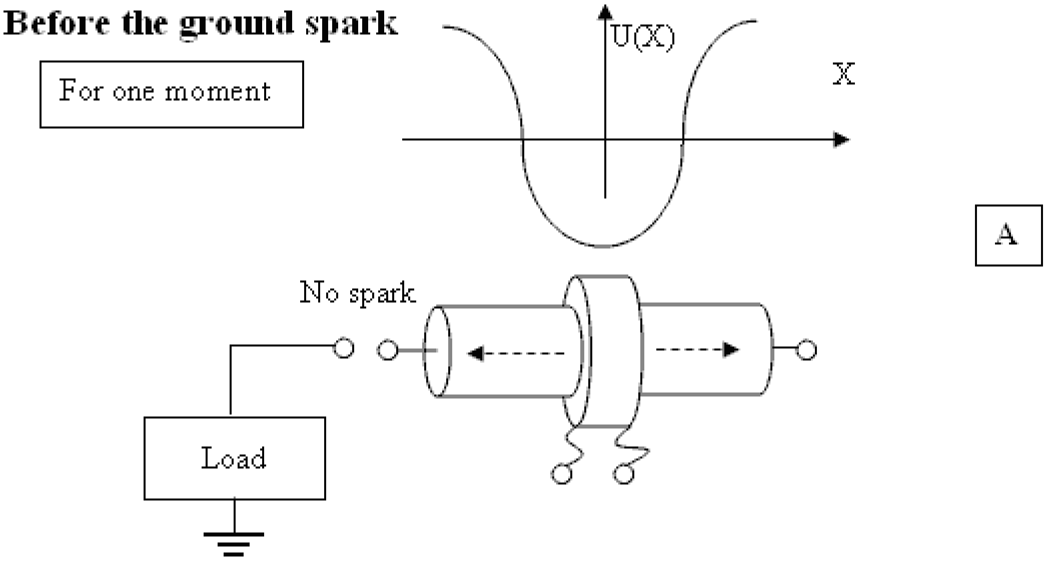

Prototype Coil

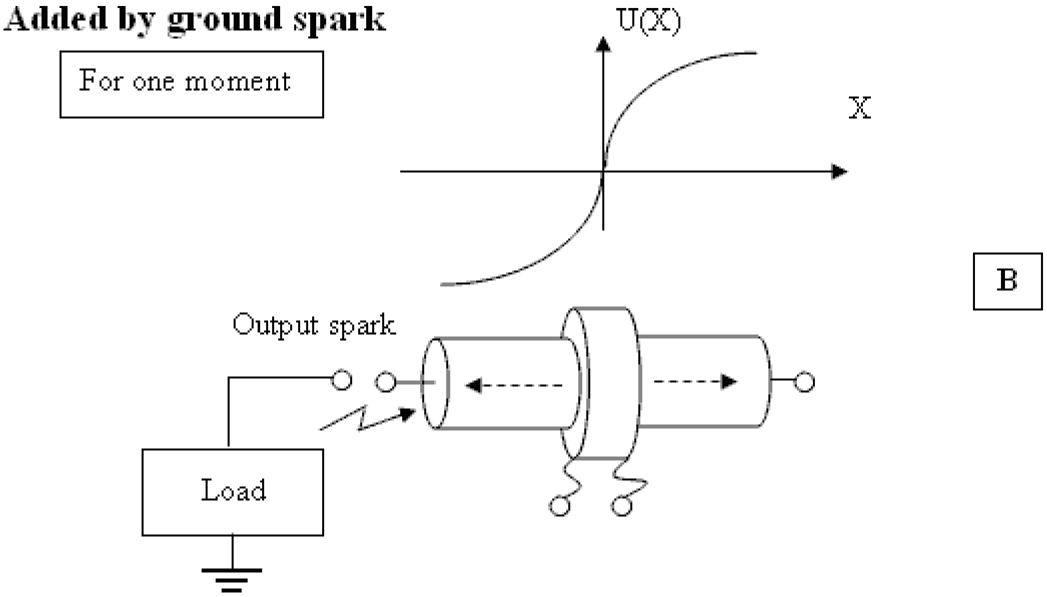

EXPLANATION The task of the oscillating circuit is to create a local electromagnetic field with a large electrical component. In theory, it would only be necessary to charge up the high voltage capacitor just once and then a lossless circuit would maintain the oscillations indefinitely without needing any further power input. In reality, there are some losses and so some additional power input is needed.

THESE OSCILLATIONS ACT AS A "BAIT", ATTRACTING CHARGE INFLOW FROM THE LOCAL ENVIRONMENT. Almost no energy is needed in order to create and maintain such a "bait"...

The next step is to move to this "bait" to one side of the circuit, close to the source of the charges which is the Ground. At this small separation, breakdown occurs and the inherent parasitic capacitance of the circuit will be instantly recharged with energy flowing into the circuit from outside.

At the ends of the circuit there will be a voltage difference, and so there will be spurious oscillations. The direction of this electromagnetic field is perpendicular to the original field of the "bait" and so it does not destroy it. This effect is due to the fact that the coil consists of two opposing halves. The parasitic oscillations gradually die out, and they do not destroy the "bait" field.

The process is repeated spark by spark for every spark which occurs. Consequently, the more often sparks occur, the greater the efficiency of the process will be. The energy in the "bait" experiences almost no dissipation, providing a much greater power output than the power needed to keep the device operating.

TESLA SCHEMATICS

COMMENT: Don Smith named this technology “Bird on the wire”. The bird is safe on the wire until a spark occurs.

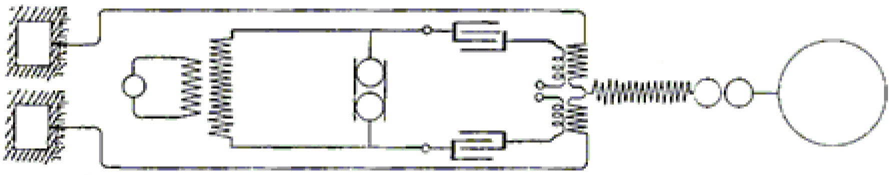

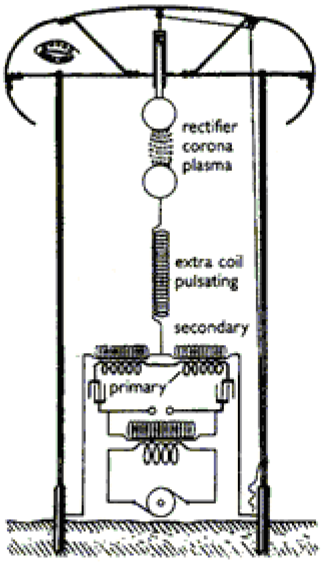
rectifier
corona
plasma
extra coil
pulsating
secondary
primary

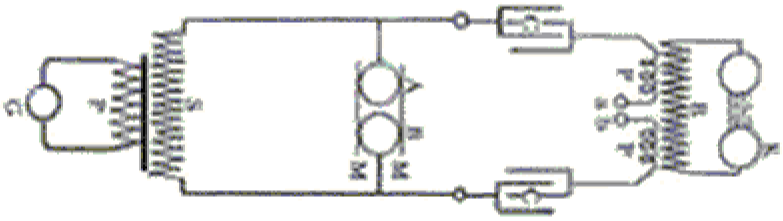

COMMENT: Mr. Tesla named this technology a “charge funnel” or “charge pump”

THE PRINCIPLE OF THE TECHNOLOGY

1. This Free-Energy device generates an AC electrical potential in ambient space (“bait” for electrons),
2. Electrons flowing through the load, flow in from the environment, attracted by this “bait” (pumped in)

NOT A SINGLE ELECTRON USED FOR EXCITING AMBIENT SPACE NEEDS TO FLOW THROUGH THE LOAD

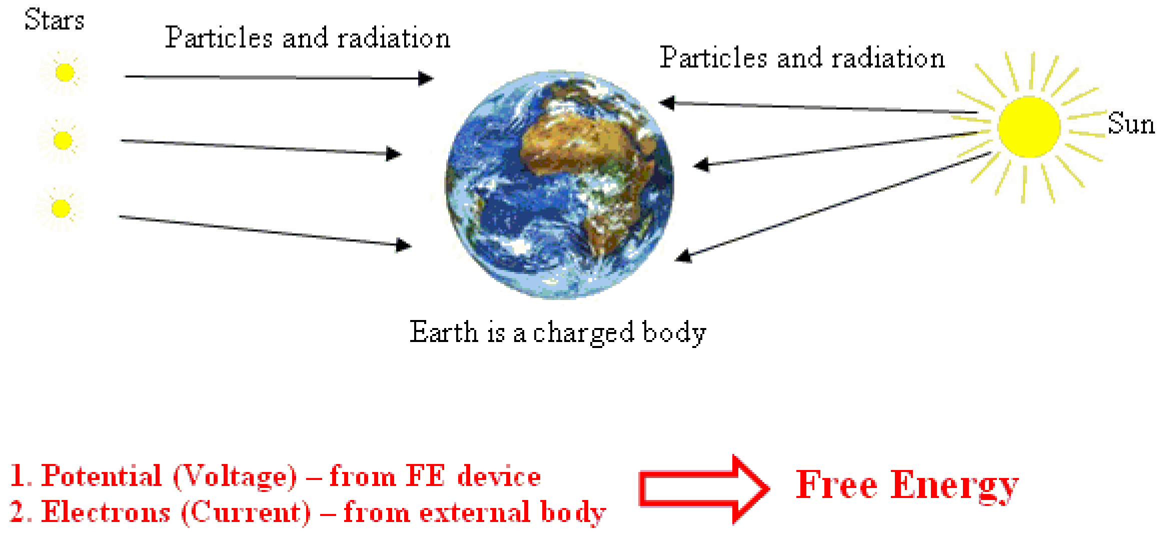
Stars
Particles and radiation
Particles and radiation
Sun
Earth is a charged body
Free Energy

1. Potential (Voltage) – from FE device
2. Electrons (Current) – from external body

POSSIBLE DESIGN FOR THE “CHARGE PUMP” OR “CHARGE FUNNEL”

By Edwin Gray

Probable Schematic for Edwin Gray’s Cold Electricity Circuit

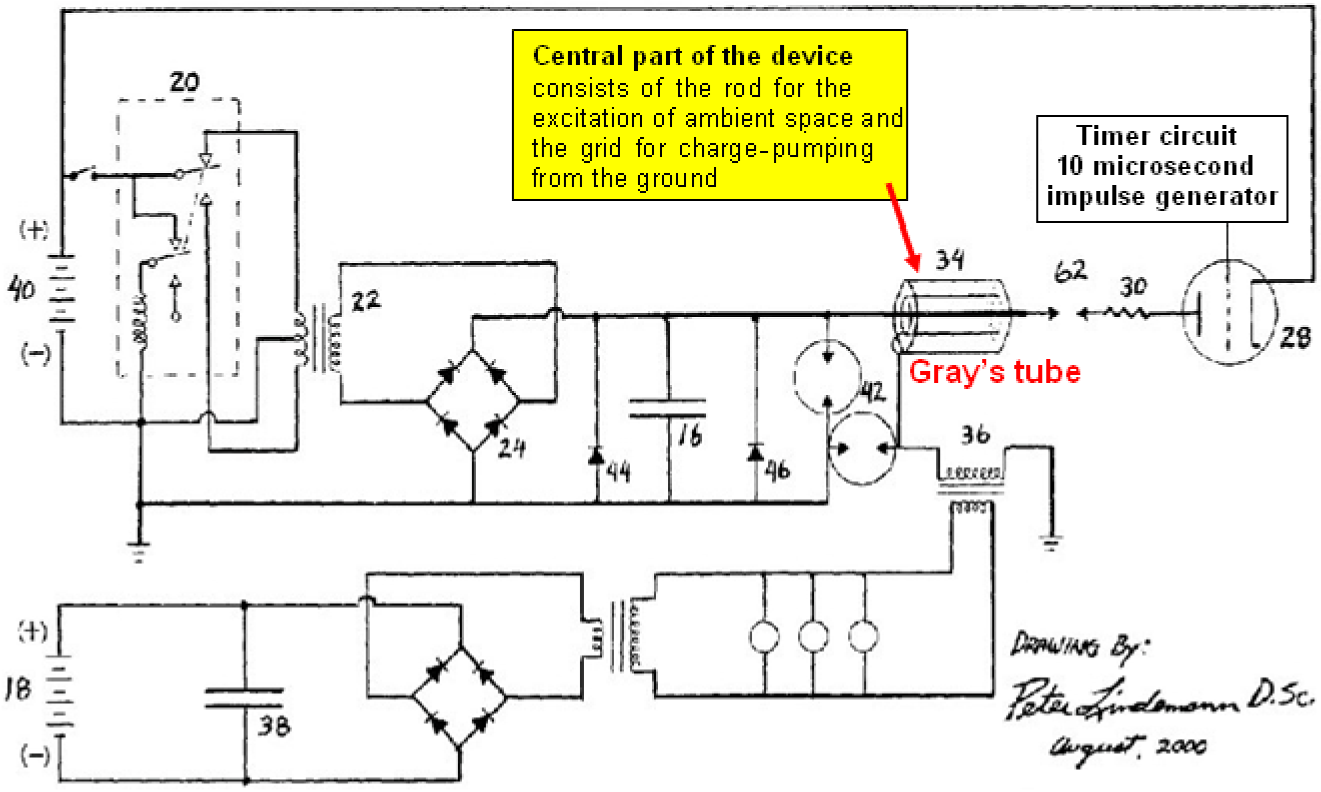
Central part of the device
consists of the rod for the
excitation of ambient space and
the grid for charge-pumping
from the ground
Timer circuit
10 microsecond
impulse generator
20
(+)
40
(-)
22
24
44
16
46
42
34
62
30
28
Gray’s tube
36
(+)
18
(-)
38
Drawing By:
Peter Lindemann D.Sc.
August 2000

EXPLANATION: This schematic is a simplification of Gray’s patent, produced by Dr. Peter Lindemann for greater clarification in his book.

FIG. 2

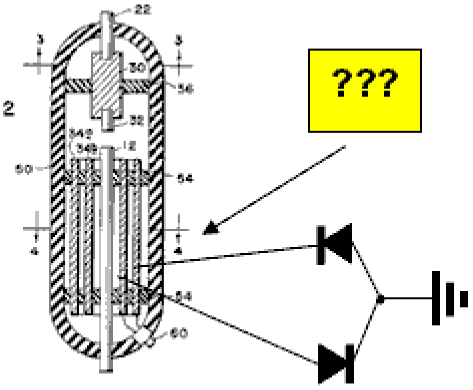
???
22
3
3
30
36
32
34e
34B
12
50
54
4
4
64
60

FIG. 3

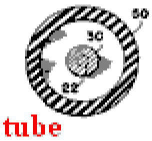

Gray’s tube

FIG. 4

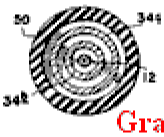
50
34e
12
34B

Gray’s original

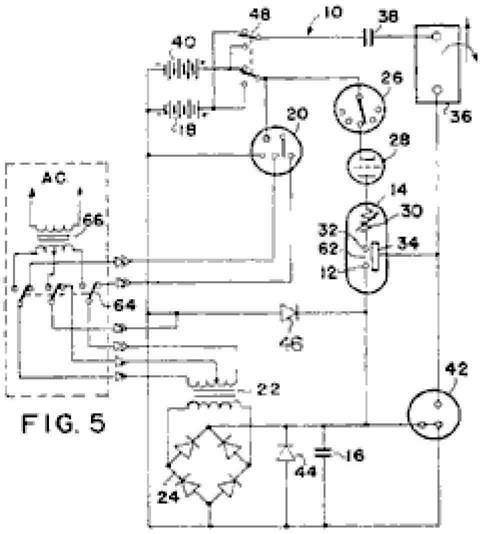
48
10
38
40
26
36
18
20
28
AC
66
14
30
32
62
34
12
64
46
22
42
16
44
24

FIG. 5

POSSIBLE DESIGN FOR THE “CHARGE PUMP” or “CHARGE FUNNEL”

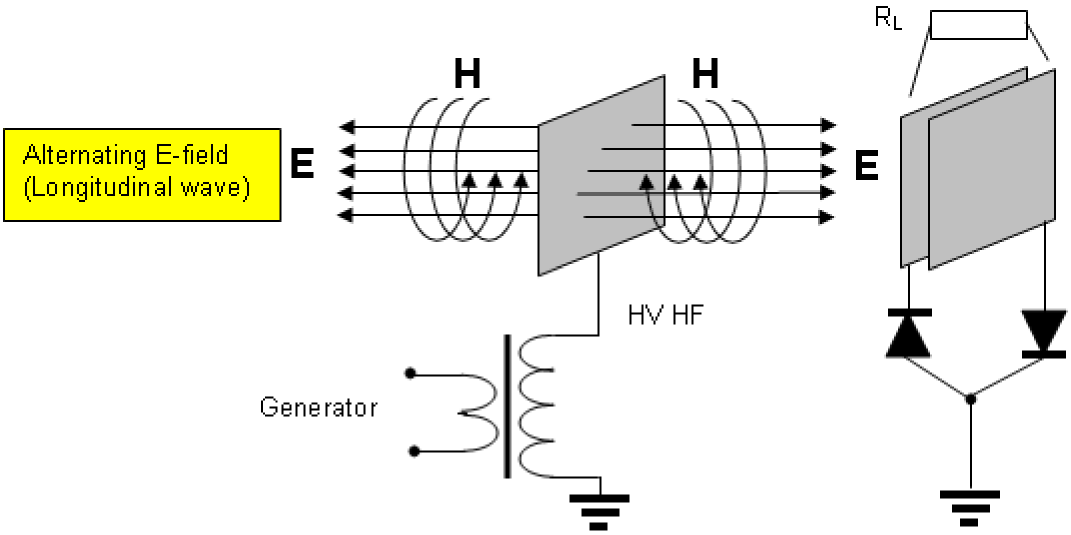

Alternating E-field
(Longitudinal wave)
E
H
H
E
R_L
HV HF
Generator

EXPLANATION: The charging system is unable to “see” the field inside a charging capacitor.

COMMON VIEW OF RESONANCE: Resonance is not destroyed if you short-circuit or open a “pumping” capacitor.

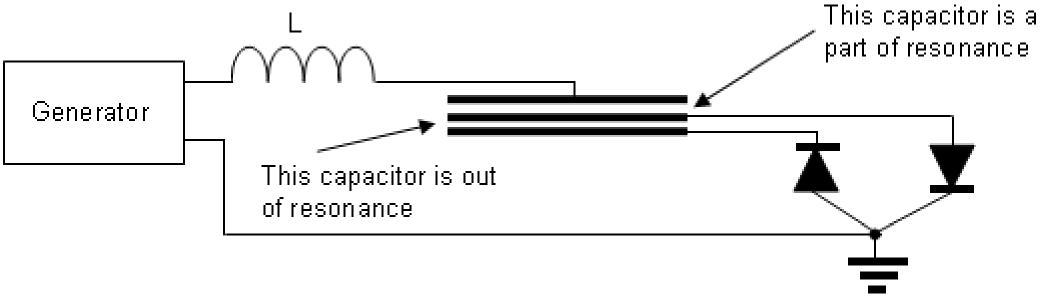

L
This capacitor is a part of resonance
Generator
This capacitor is out of resonance

COMMENT: You can add an ordinary, very large capacitor in parallel with the “pumping” capacitor for more impressive results.

Don Smith illustration

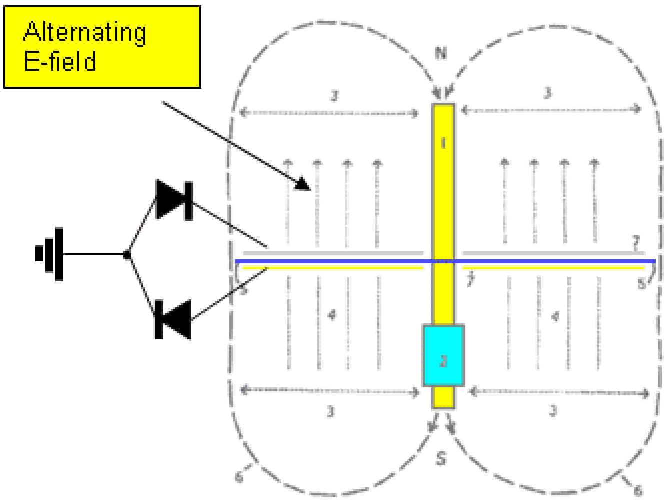

Alternating E-field
N
S

COMMENTS: You have to use an alternating E-field, in order to charge the capacitor. But, Smith marked the North and South poles in his drawing. I think that this is true for only one instant. Diodes are not shown in his drawings, which indicates that his device as shown, is to my mind not complete.

THE EXTERNAL APPEARANCE OF ED GRAY'S TUBE

EXPLANATION: Gray's tube with it's two internal grids is seen in the middle. Two diodes are underneath the acrylic sheet (???). A Leiden Jar is located on the left (???) The HF HV coil is behind Gray's tube (???)

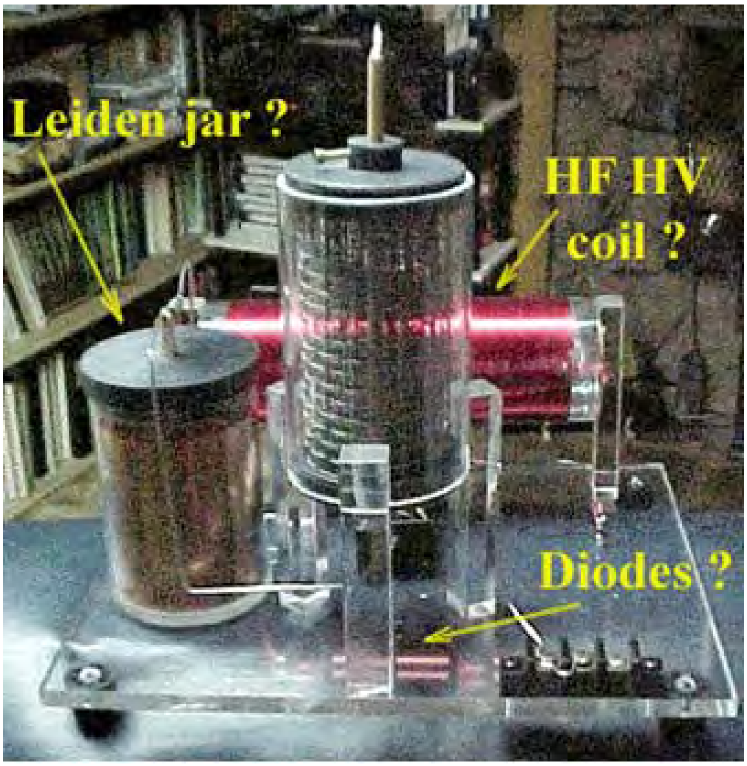
Leiden jar ?
HF HV coil ?
Diodes ?

A POSSIBLE DESIGN FOR THE "CHARGE PUMP" or "CHARGE FUNNEL"

THE TESTATIKA by Paul Bauman

EXPLANATION: The central electrode in the jars (capacitors) is for the excitation of ambient space; the two external cylinders are the plates of the charging capacitors.

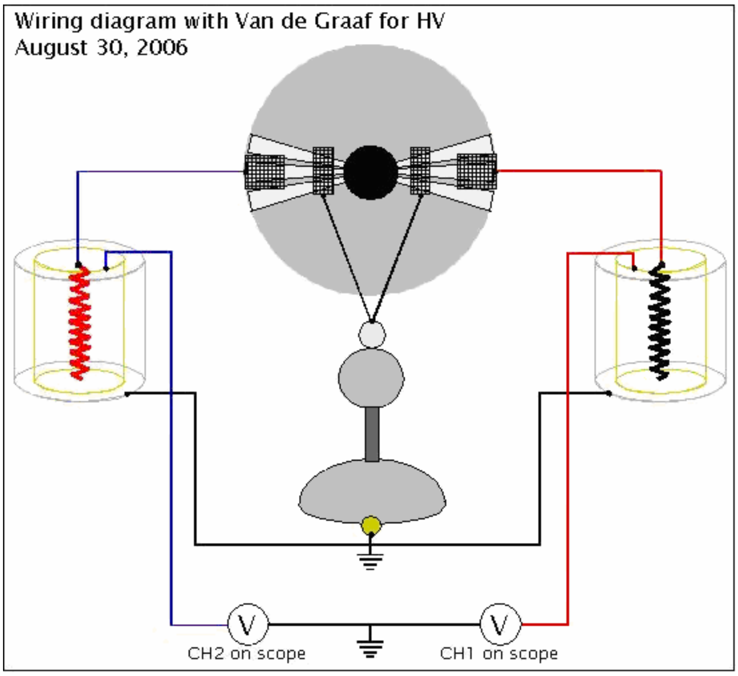
Wiring diagram with Van de Graaf for HV
August 30, 2006
V
V
CH2 on scope
CH1 on scope

EXPLANATION: The charging mechanism is unable to "see" the field inside the charging capacitors.

COMMENT: For more details read the section on asymmetrical capacitors.

A POSSIBLE DESIGN FOR THE “CHARGE PUMP” or “CHARGE FUNNEL”

COMMENT: This is based on Tesla’s schematics

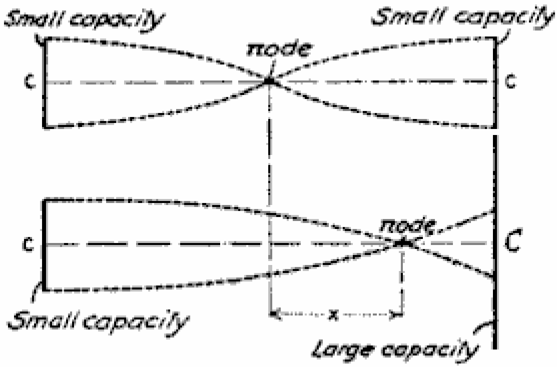

Small capacity
Small capacity
node
c
c
c
node
C
Small capacity
x
Large capacity

COMMENT: First, you need to arrange a “voltage killer” barrier on one side of the Tesla coil. This is to create a “BLIND” charging system which can’t “see” the charge on the capacitor (see below for more detail on “blindness”).

COMMENTS: ‘Huge capacitor’ means: as much ordinary capacitance as possible.
Effectiveness depends on voltage and coil frequency, and current in the node.
Effectiveness depends also on the frequency at which the excitation spark occurs.
It is very similar to Don Smith’s devices.

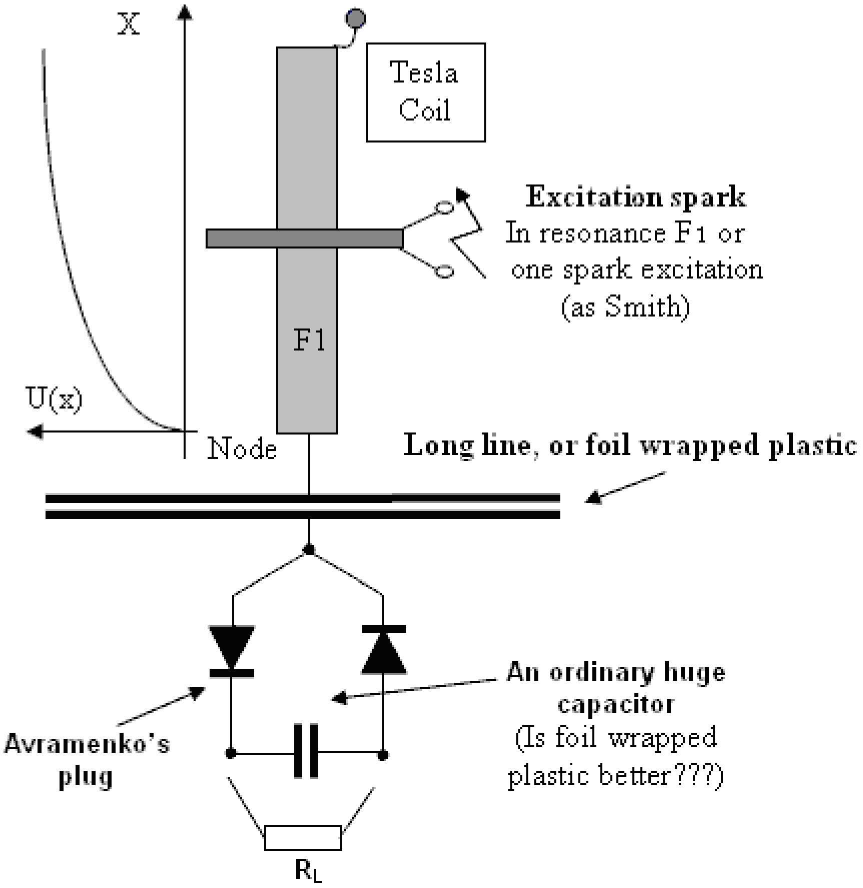

X
Tesla Coil
Excitation spark
In resonance F1 or one spark excitation (as Smith)
F1
U(x)
Node
Long line, or foil wrapped plastic
An ordinary huge capacitor (Is foil wrapped plastic better???)
Avramenko’s plug
R_L

COMMENT: For more details read part devoted to Avramenko’s plug…

POSSIBLE DESIGN FOR THE “CHARGE PUMP” or “CHARGE FUNNEL”

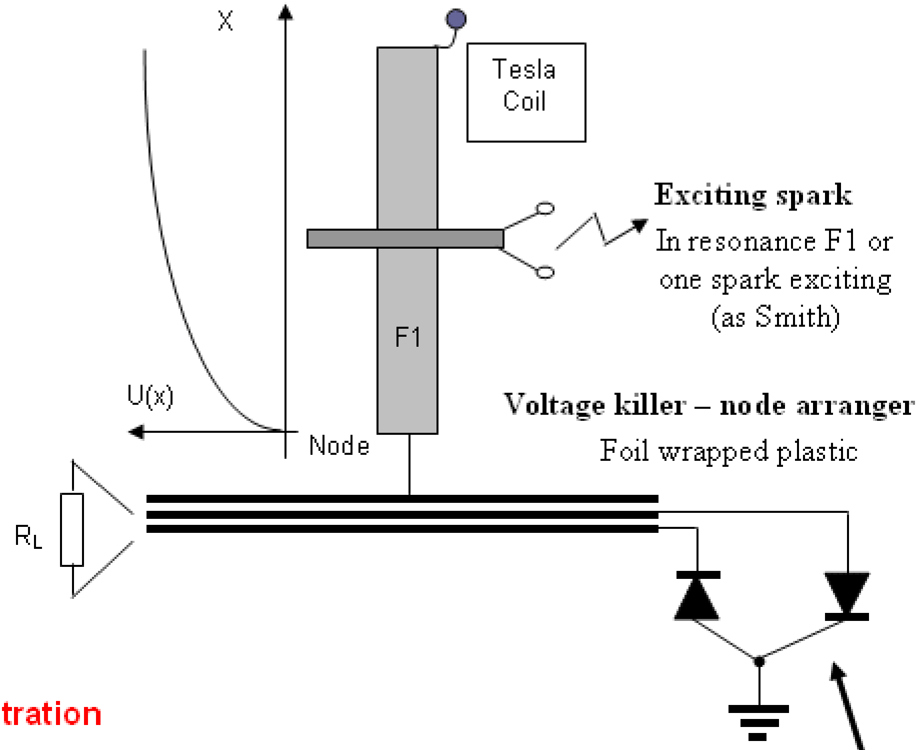
X
U(x)
Node
Tesla Coil
F1
Exciting spark
In resonance F1 or one spark exciting (as Smith)
Voltage killer – node arranger
Foil wrapped plastic
R_L

Don Smith illustration

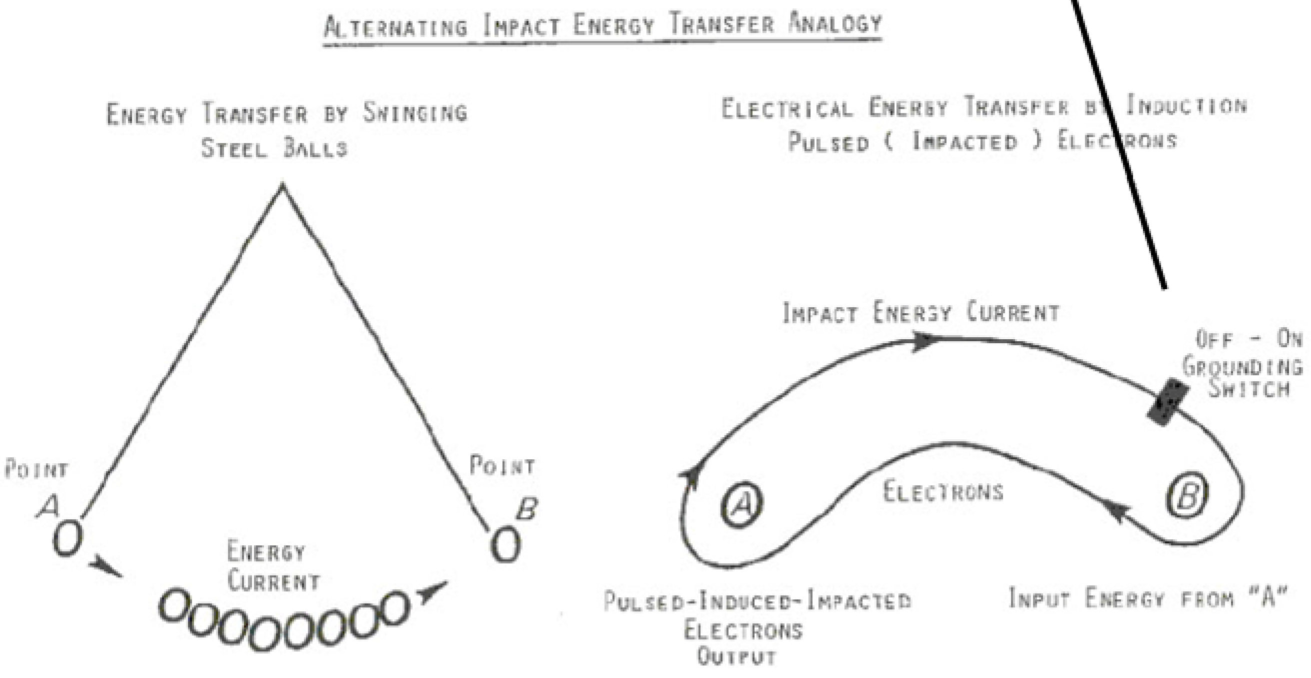
ALTERNATING IMPACT ENERGY TRANSFER ANALOGY
ENERGY TRANSFER BY SWINGING STEEL BALLS
POINT A
POINT B
ENERGY CURRENT
ELECTRICAL ENERGY TRANSFER BY INDUCTION PULSED (IMPACTED) ELECTRONS
IMPACT ENERGY CURRENT
OFF - ON GROUNDING SWITCH
A
ELECTRONS
B
PULSED-INDUCED-IMPACTED ELECTRONS OUTPUT
INPUT ENERGY FROM "A"

IN ALTERNATING ELECTRICAL ENERGY SYSTEMS, ONLY THE POTENTIAL-IMPACT LEVEL REPLICATES IT'S SELF AT POINT "B". ELECTRONS DO NOT TRAVEL FROM POINT "A" TO POINT "B".

EXPLANATION: The charging system is unable to “see” the field inside the charging capacitor.

COMMENT: For more details read the part which is devoted to Avramenko’s plug…

COMMENT: An ordinary piece of wire can be used in some versions of this gadget, see below….

ENERGY REGENERATION BY L/4 COIL

COMMENT: This system is based on wireless energy transmission through the ground

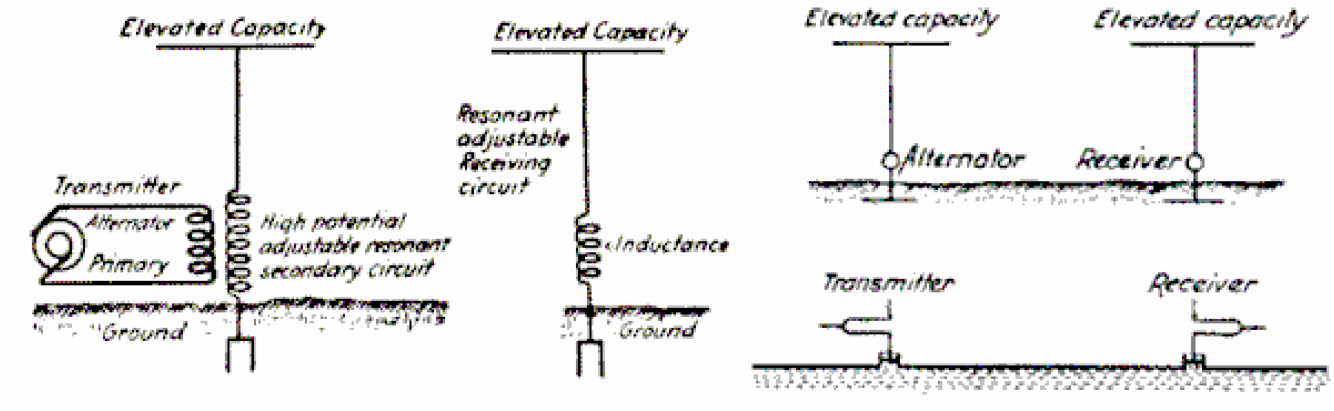
Elevated Capacity
Transmitter
Alternator
Primary
High potential adjustable resonant secondary circuit
Ground
Elevated Capacity
Resonant adjustable Receiving circuit
Inductance
Ground
Elevated capacity
Elevated capacity
Alternator
Receiver
Transmitter
Receiver

COMMENT: Energy radiated to ambient space lowers the efficiency of this process

COMMENT: The Receiver and Transmitter coils must have the same resonant frequency

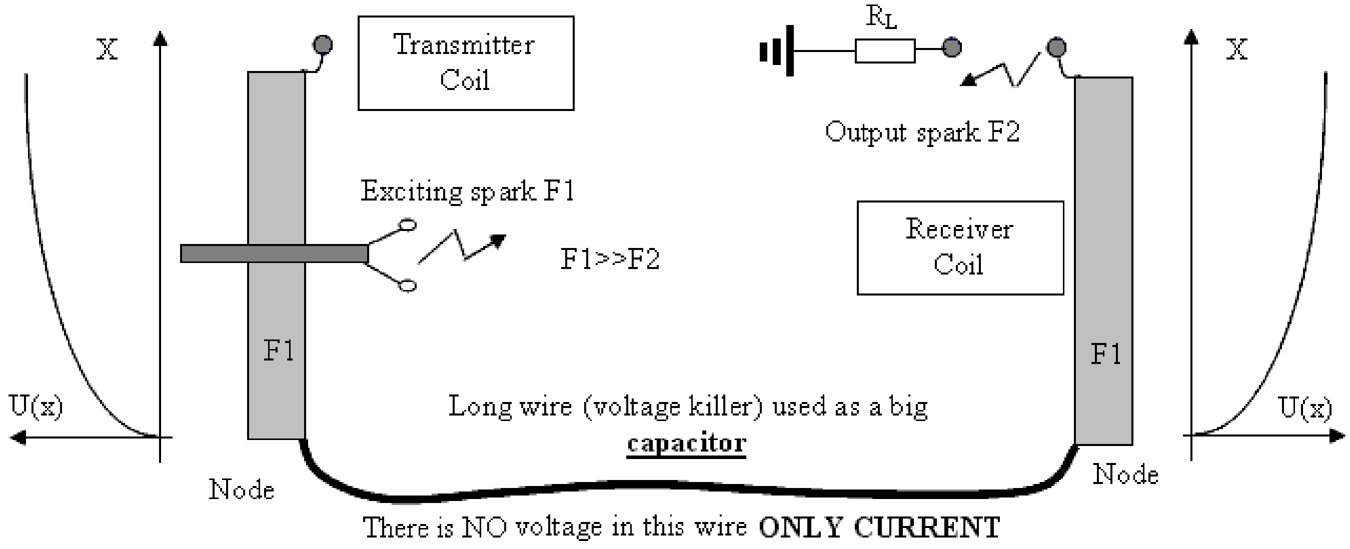
X
Transmitter Coil
R$_L$
Output spark F2
X
Exciting spark F1
F1>>F2
Receiver Coil
F1
F1
U(x)
Long wire (voltage killer) used as a big capacitor
U(x)
Node
Node
There is NO voltage in this wire ONLY CURRENT

COMMENT: Possible alternative arrangement:

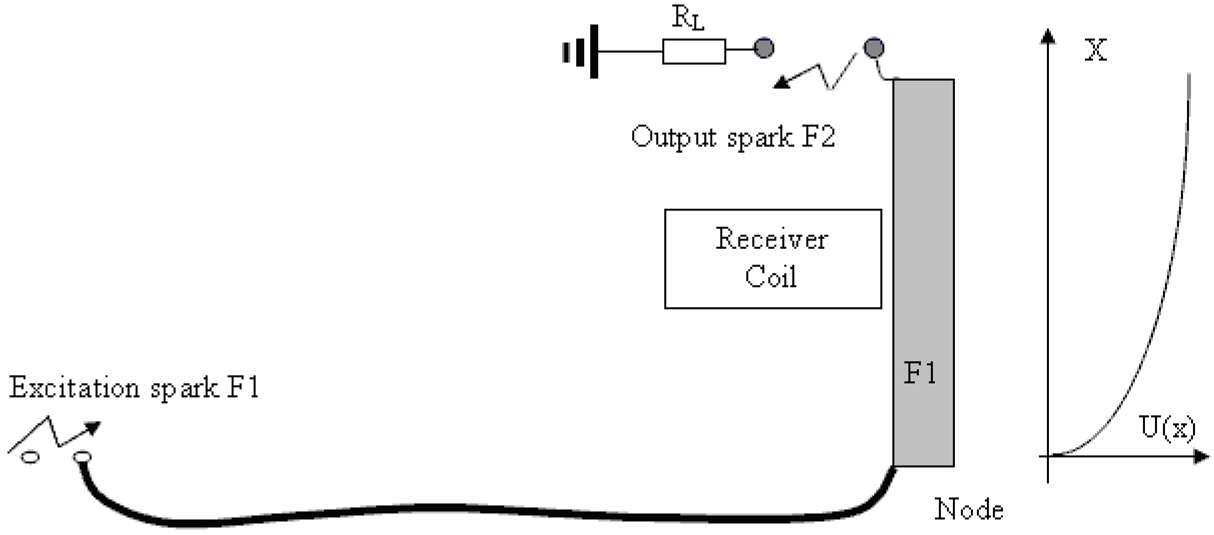
R$_L$
X
Output spark F2
Receiver Coil
Excitation spark F1
F1
U(x)
Node

COMMENT: A metal sheet can be used instead of a long wire

The “COLD” and “HOT” ends of a Tesla Coil

by Donald Smith

COMMENT: If the excitation coil L1 is positioned in the centre of coil L2, then the Tesla Coil will have a “cold” end and a “hot” end. **A spark gap can only be connected to the “hot” end**. You cannot get a good spark if the spark gap is connected to the “cold” end.

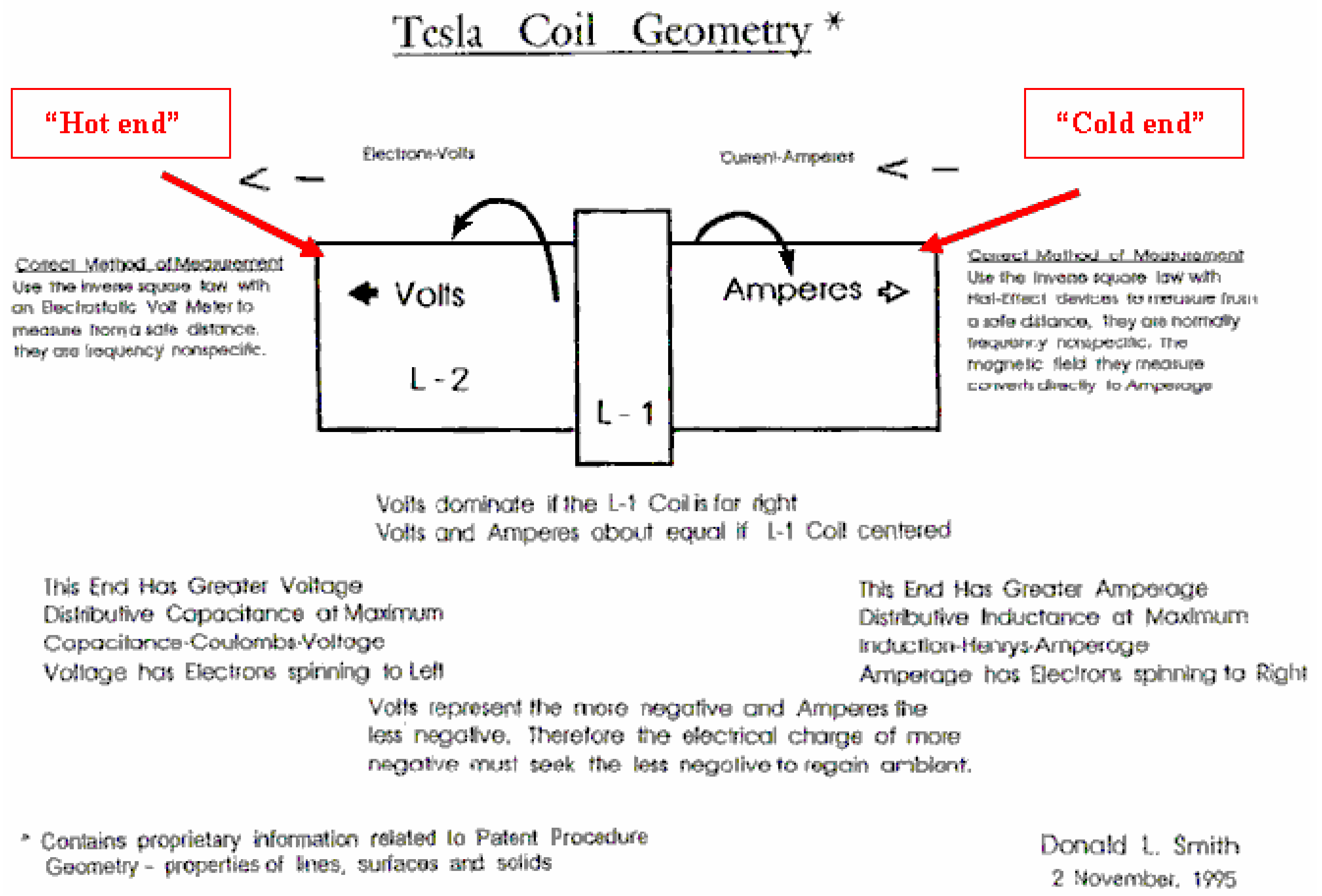
Tesla Coil Geometry *
“Hot end”
“Cold end”
Electron-Volts
< —
Current-Amperes
< —
Volts
L - 2
L - 1
Amperes

Correct Method of Measurement
Use the inverse square law with an Electrostatic Volt Meter to measure from a safe distance. they are frequency nonspecific.

Correct Method of Measurement
Use the inverse square law with Hall-Effect devices to measure from a safe distance. they are normally frequency nonspecific. The magnetic field they measure converts directly to Amperage

Volts dominate if the L-1 Coil is far right
Volts and Amperes about equal if L-1 Coil centered

This End Has Greater Voltage
Distributive Capacitance at Maximum
Capacitance-Coulombs-Voltage
Voltage has Electrons spinning to Left

This End Has Greater Amperage
Distributive Inductance at Maximum
Induction-Henrys-Amperage
Amperage has Electrons spinning to Right

Volts represent the more negative and Amperes the less negative. Therefore the electrical charge of more negative must seek the less negative to regain ambient.

* Contains proprietary information related to Patent Procedure
Geometry - properties of lines, surfaces and solids

Donald L. Smith
2 November, 1995

COMMENT: This is very important for practical applications, so read Don Smith’s documents for more details.

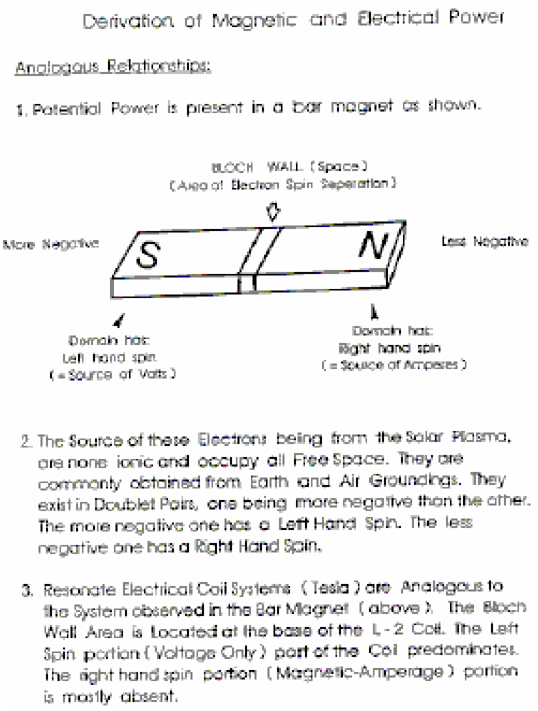

Derivation of Magnetic and Electrical Power

Analogous Relationships:

1. Potential Power is present in a bar magnet as shown.

BLOCH WALL (Space)
(Area of Electron Spin Seperation)
More Negative
S
N
Less Negative
Domain has:
Left hand spin
(= Source of Volts)
Domain has:
Right hand spin
(= Source of Amperes)

2. The Source of these Electrons being from the Solar Plasma, are none ionic and occupy all Free Space. They are commonly obtained from Earth and Air Groundings. They exist in Doublet Pairs, one being more negative than the other. The more negative one has a Left Hand Spin. The less negative one has a Right Hand Spin.

3. Resonate Electrical Coil Systems (Tesla) are Analogous to the System observed in the Bar Magnet (above). The Bloch Wall Area is Located at the base of the L-2 Coil. The Left Spin portion (Voltage Only) part of the Coil predominates. The right hand spin portion (Magnetic-Amperage) portion is mostly absent.

COMMENT: It is easy to understand the “Hot” and “Cold” ends, if one end of the Tesla Coil is grounded…

The Grounded Tesla coil – a hidden form of energy

EXPLANATION: We can look at the Tesla coil as a piece of metal. Every piece of metal can be charged. If Tesla coil is grounded, it has an extra charge delivered from the ground, and has an extra energy also. But, it can be find out only in electrostatics interactions, not in electromagnetic one.

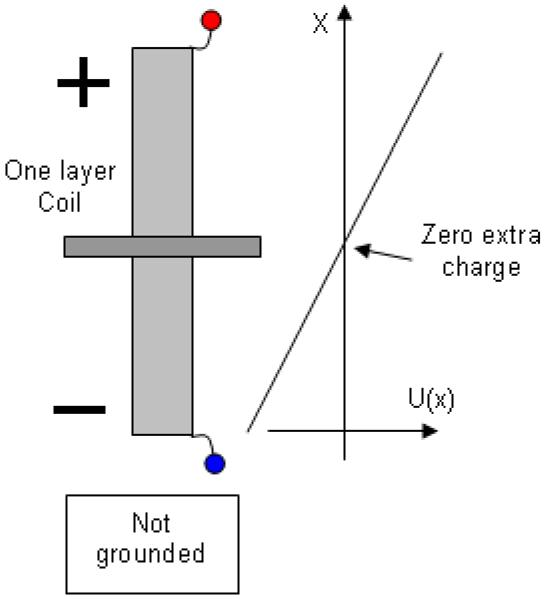
+
One layer Coil
–
Not grounded
X
Zero extra charge
U(x)

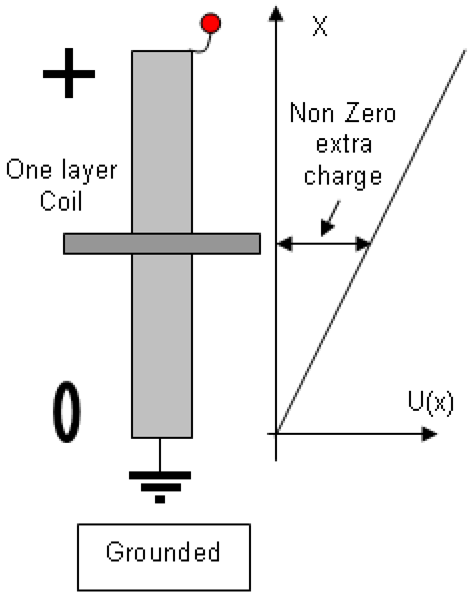
+
One layer Coil
0
Grounded
X
Non Zero extra charge
U(x)

Comment: This diagram shows only one instant, after half a cycle, the polarities will be swapped over.

Question: How can we use this fact?

Answer: We have to arrange an electrostatic interaction:

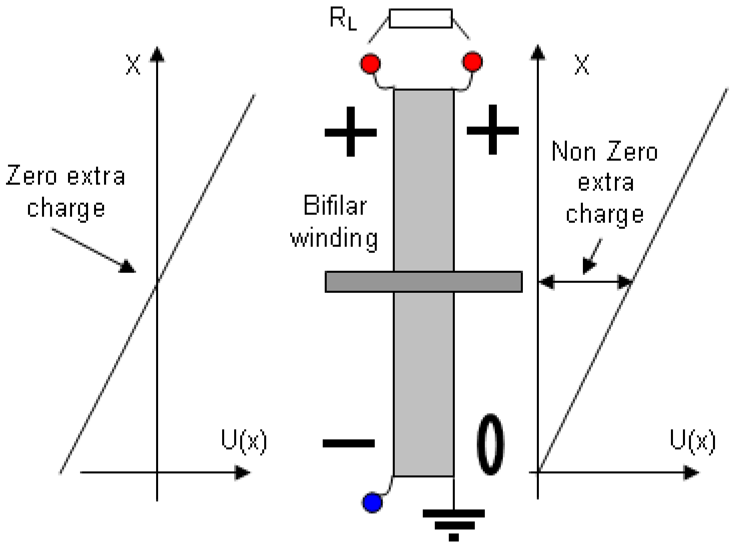
X
Zero extra charge
U(x)
R_L
+
+
Bifilar winding
–
0
X
Non Zero extra charge
U(x)

OR

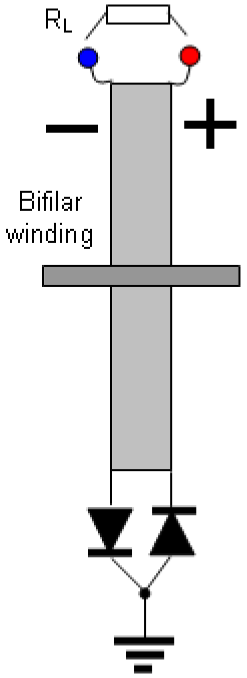
R_L
–
+
Bifilar winding

Comments: Extra capacitors can be used for charging them.

This looks like Smith’s plasma globe device. Maybe, he used this technology.

This can be used in charge pump technology for excitation by an alternating electrical field, read the section on the charge pump or charge funnel.

The wiring can be different to that shown above.

Examples of grounded bifilar (multi-strand) coils

From Tariel Kapanadze in his 100 KW device

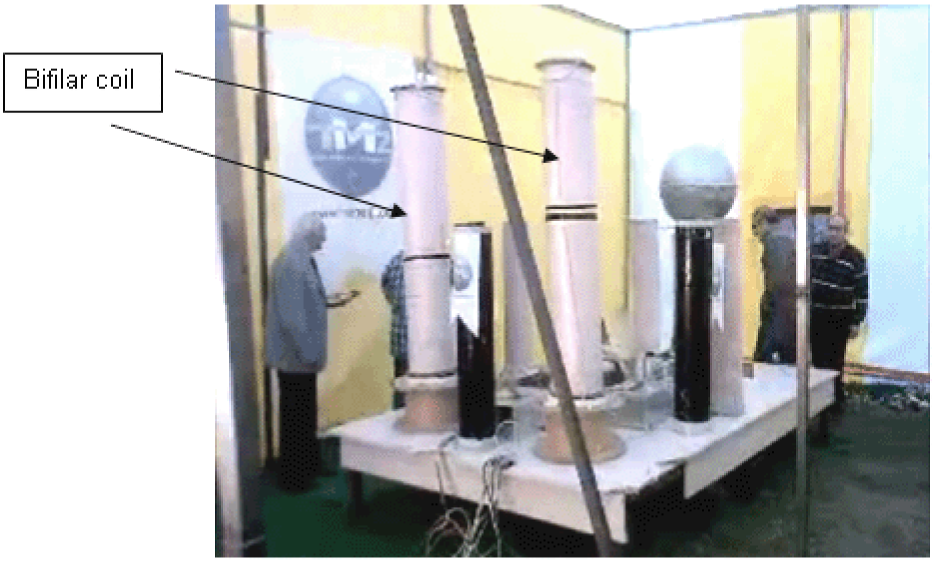
Bifilar coil

from Steven Mark in big TPU

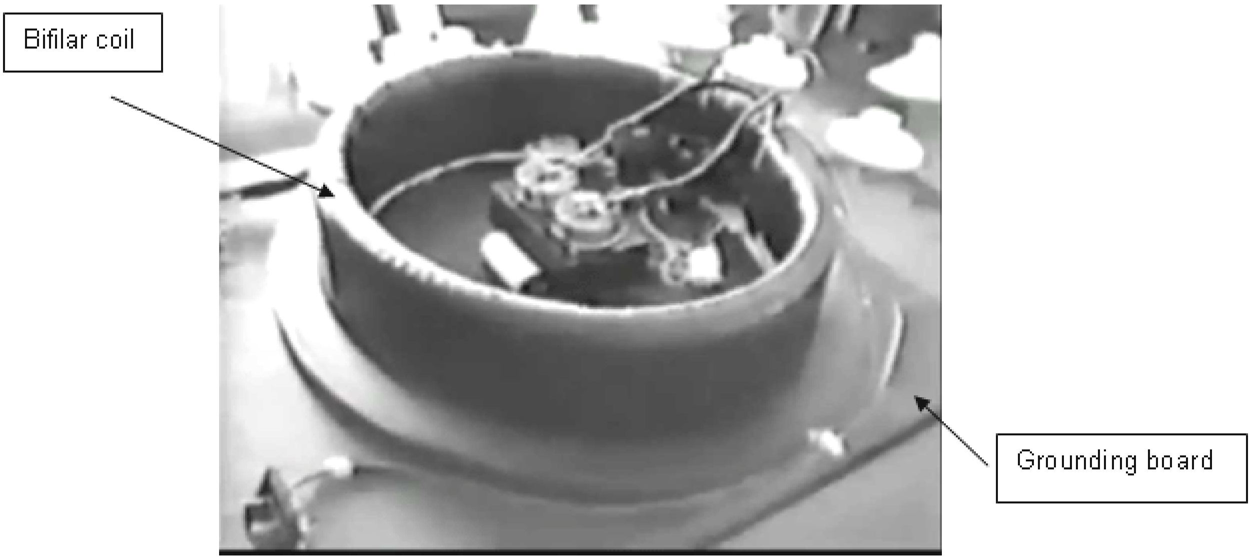
Bifilar coil
Grounding board

from Donald Smith

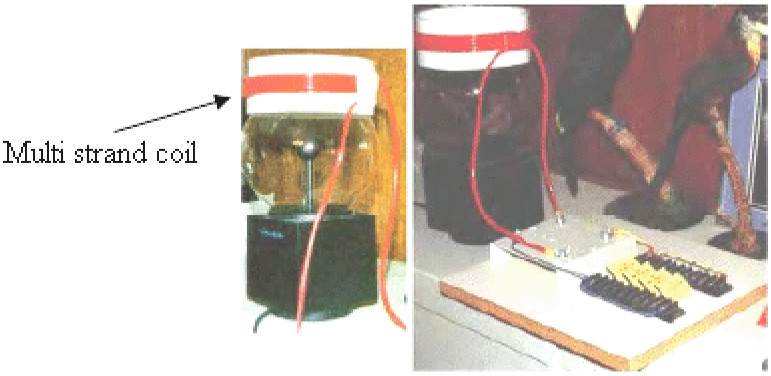
Multi strand coil

Globe device for energy generation

Possible simulation of the L/4 wave “Cold” and “Hot” ends

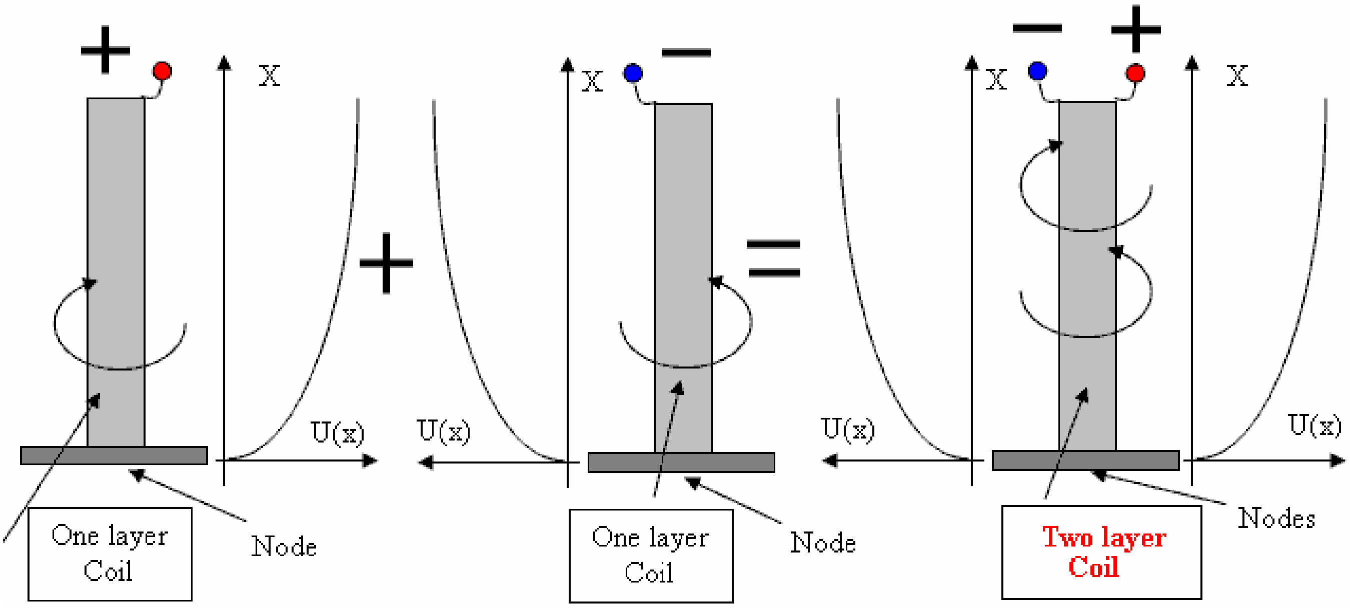
+
X
U(x)
One layer Coil
Node
+
−
X
U(x)
One layer Coil
Node
=
− +
X
X
U(x)
U(x)
Two layer Coil
Nodes

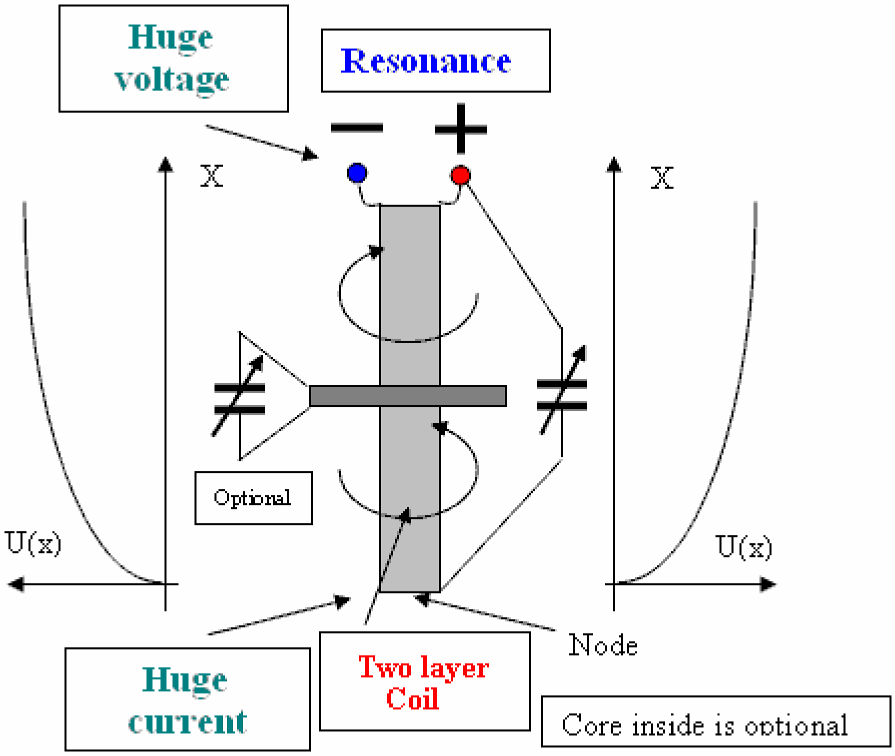
Huge voltage
Resonance
− +
X
X
U(x)
U(x)
Optional
Node
Huge current
Two layer Coil
Core inside is optional

MODERN OPTIONS

In back EMF suppression

Version 1

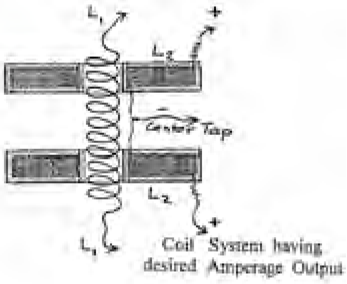
L1
L2
+
Center Tap
L2
+
L1
Coil System having desired Amperage Output

SYMMETRICAL VERSION

By Don Smith

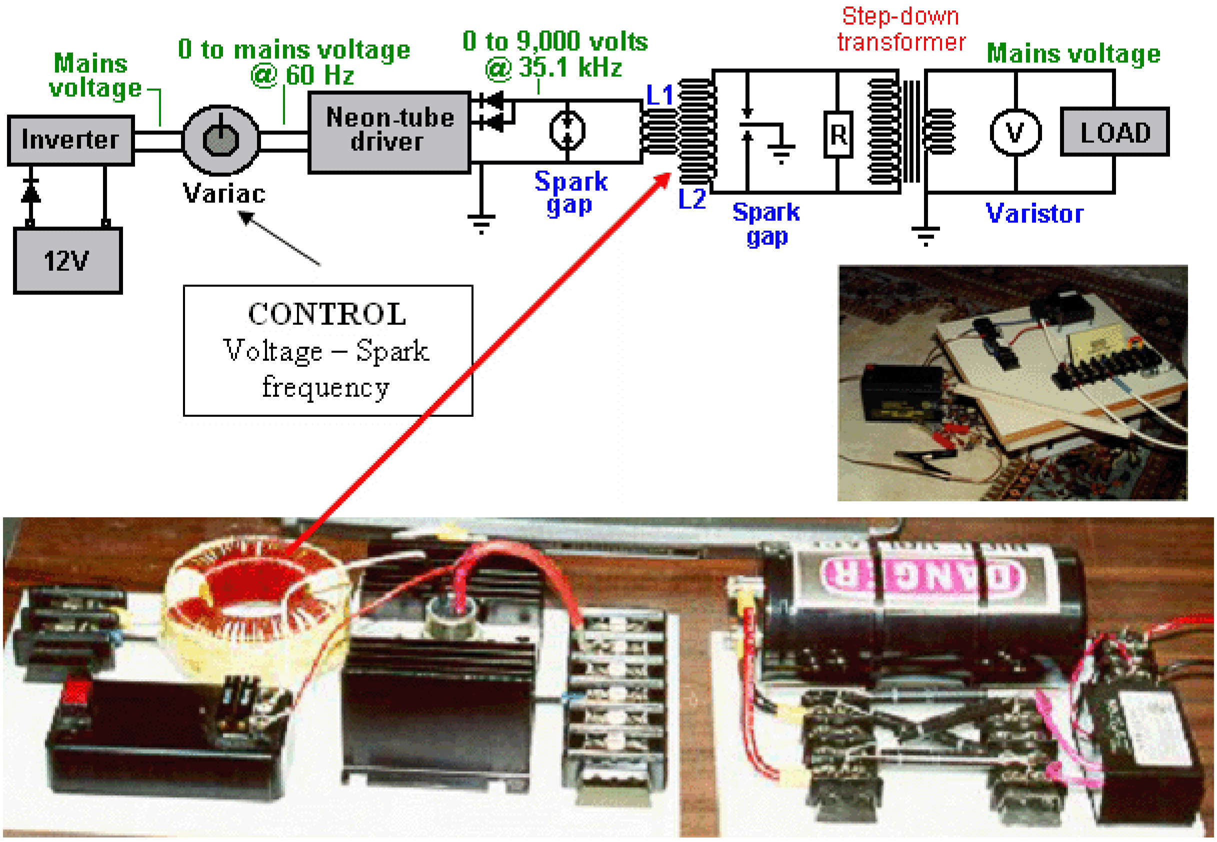
Mains voltage
0 to mains voltage @ 60 Hz
0 to 9,000 volts @ 35.1 kHz
Step-down transformer
Mains voltage
Inverter
Neon-tube driver
L1
R
V
LOAD
Spark gap
L2
Spark gap
Varistor
12V
Variac
DANGER

CONTROL
Voltage – Spark frequency

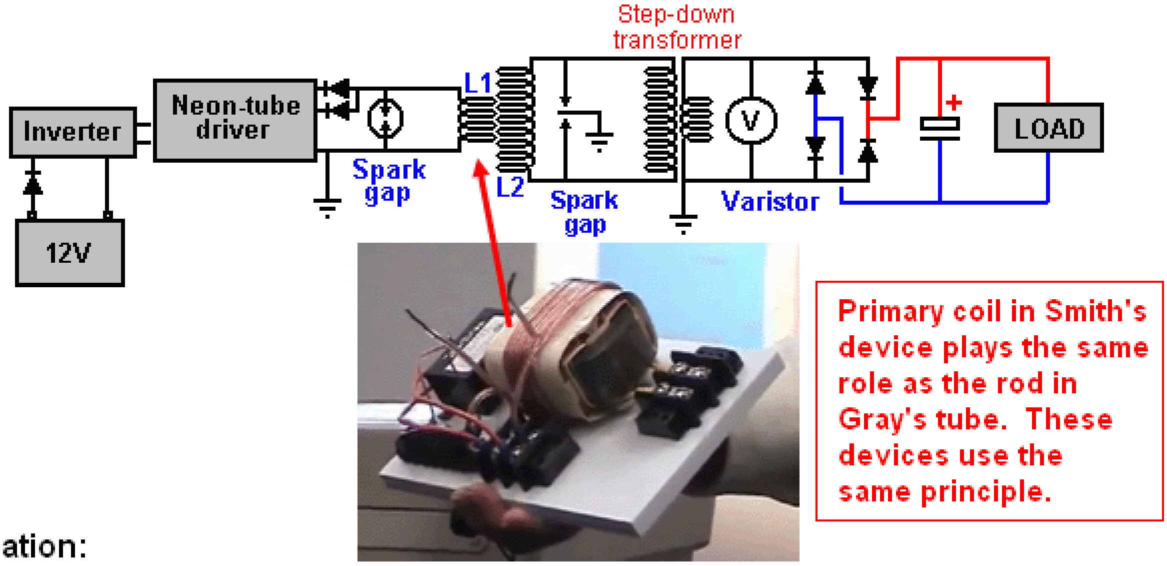
Step-down transformer
Inverter
Neon-tube driver
L1
V
LOAD
Spark gap
L2
Spark gap
Varistor
12V

Primary coil in Smith's device plays the same role as the rod in Gray's tube. These devices use the same principle.

Explanation:

Both of the two out of phase outputs were used and both connected to the step-down transformer.

1. Between sparks:

There is no current in the step-down transformer and so the two ends of L2 are at the same voltage.

2. During a spark:

Parasitic capacitors (not shown) connected across both sections of L2 are discharged to ground, and current is produced in the step-down transformer. One end of L2 is at ground potential. But, the magnetic field of this current in L2 is perpendicular to the resonating field and so has no influence on it. **As a result of this, you have power in the load, but the resonance is not destroyed.**

COMMENTS: In my opinion, these schematics have errors in the excitation section. Find those errors.

Excitation by a single spark is possible.

In the terminology of Mr. Tesla, this is a 'charge pump' or 'charge funnel'. The charges are coming from the Ground which is the source of the energy.

There are more secrets in the following parts.

SECRET 1.1

Back EMF suppression in a resonance coil

Version 2

Primary and secondary coils are placed on a rod core. All of the coils are arranged in special manner. The primary coil is placed in the middle of the core. The secondary coil is in two parts which are positioned at the ends of the rod. All of the coils are wound in the same direction.

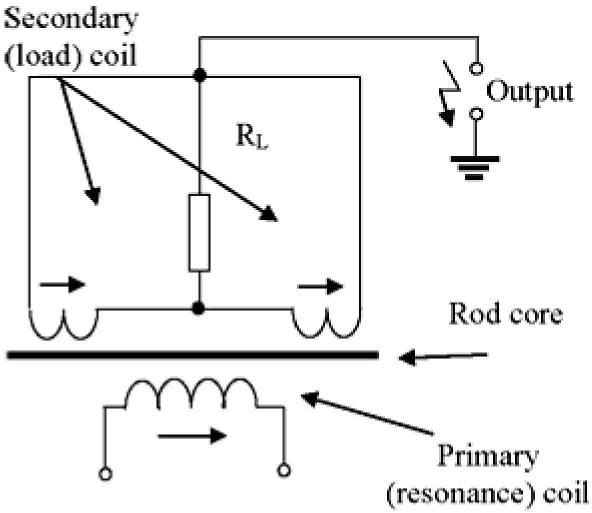
Secondary
(load) coil
R_L
Output
Rod core
Primary
(resonance) coil

Explanation:

The electromagnetic fields produced by the resonant (excitation) current and the load current are perpendicular to each other:

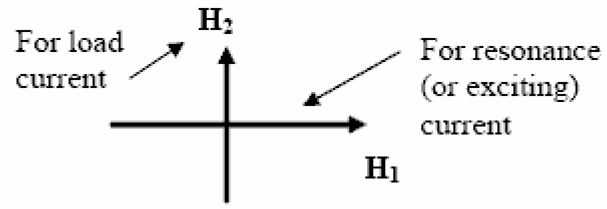
For load
current
H_2
For resonance
(or exciting)
current
H_1

So, although you have power in the load, resonance is not destroyed by that output power.

Comments: The load must be chosen so as to get the maximum amount of power flowing into it. Very low loads and very high loads will both have close to zero energy flowing in them.

The secondary coil is shunting the primary coil, and so it has a current flowing in it even if no loads are connected.

The secondary coil can be adjusted for resonance too.

The "rod" material can be air, or other materials.

SECRET 1.1

Back EMF suppression in a resonance coil

Version 3

(long wire usage – bifilar usage)

EXPLANATION: It is very much like **Version 1**, but here, the two coils are combined into a single coil.

REMARK There is no current in the load without spark

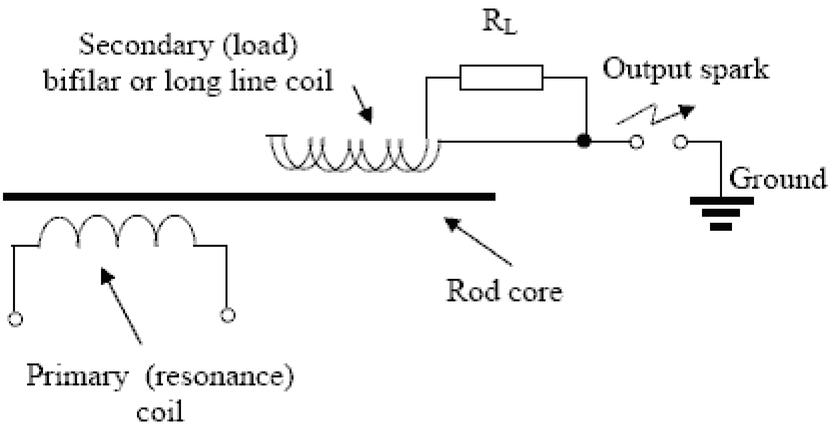
R_L
Secondary (load)
bifilar or long line coil
Output spark
Ground
Rod core
Primary (resonance)
coil

IT IS IMPOSSIBLE!

(Without back EMF suppression)

By Don Smith

DANGER
HIGH VOLTAGE

Multi-coil system for energy multiplication

COMMENT : You decide how you think it was made. Maybe short-circuited coils will be useful…

Read the following parts to discover more secrets…

MODERN OPTIONS?

For Back EMF suppression

Version 3

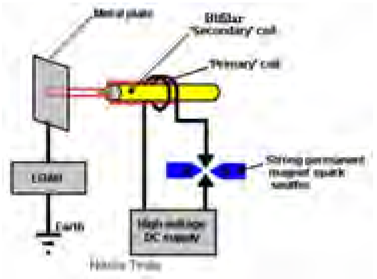
Metal plate
Bifilar 'Secondary' coil
'Primary' coil
Strong permanent magnet spark snuffer
LOAD
Earth
High voltage DC supply
Nikola Tesla

BI-FILAR USAGE

By Tariel Kapanadze

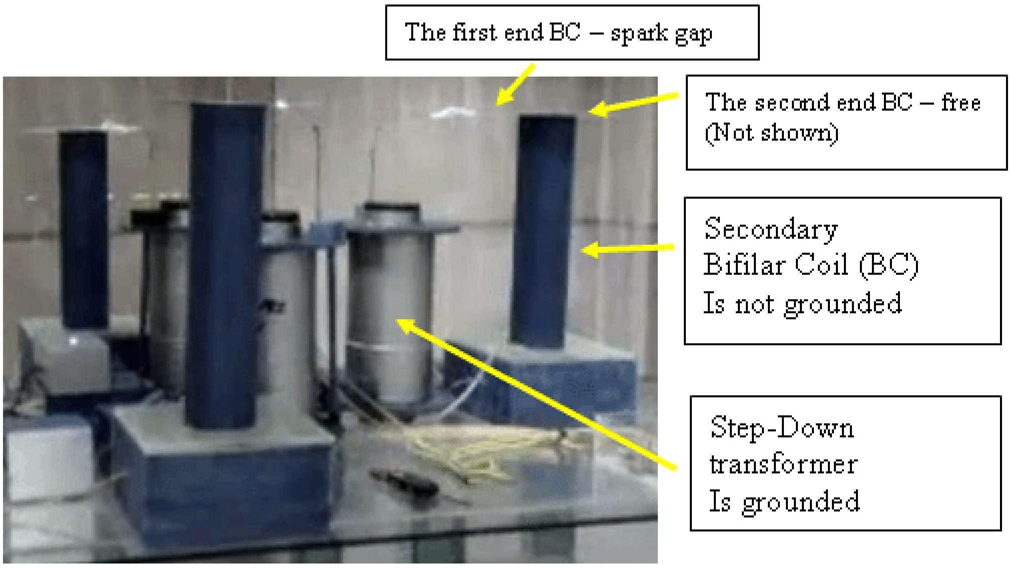
The first end BC – spark gap
The second end BC – free (Not shown)
Secondary Bifilar Coil (BC) Is not grounded
Step-Down transformer Is grounded

Possible schematics of the device

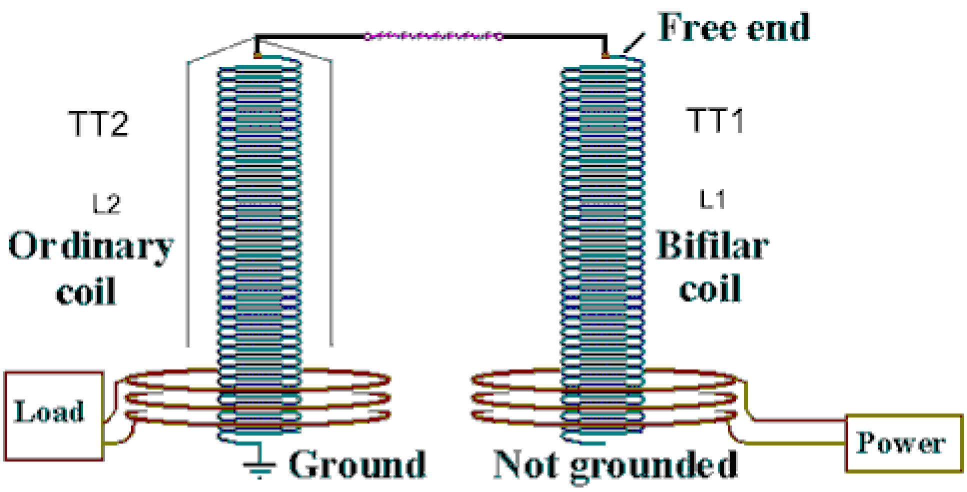
Free end
TT2
L2
Ordinary coil
TT1
L1
Bifilar coil
Load
Ground
Not grounded
Power

BIFILAR USAGE

By Timothy Trapp

COMMENT: See Trapp's sites for more details

POSSIBLE CORE CONFIGURATION
For back EMF suppression

TOROIDAL CORE

BI-FILAR WINDING

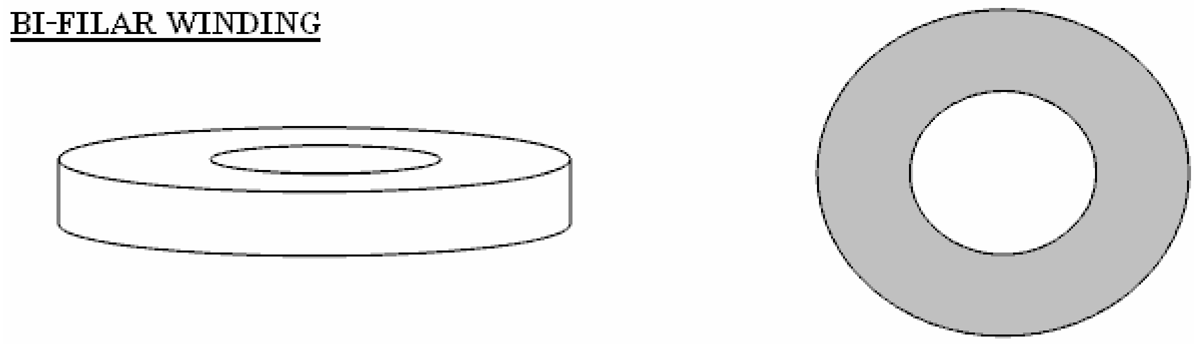

COMMENTS: An ordinary excitation winding is wound all of the way around a toroidal core.
A bi-filar output winding is wound around the whole of a toroidal core.
Remember about the "Hot" and "Cold" ends of a bi-filar coil.

ORDINARY WINDING

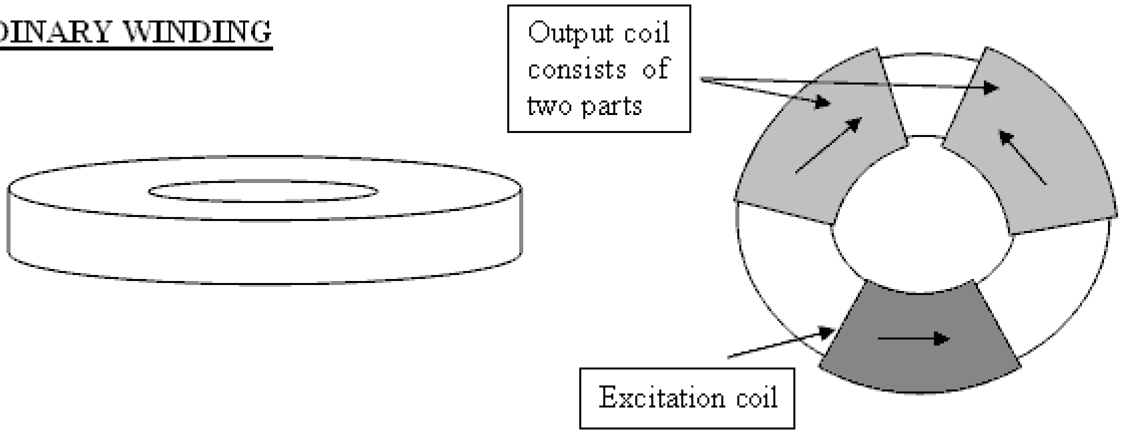
Output coil consists of two parts
Excitation coil

COMMENT: Remember about the "Hot" and "Cold" ends of the output coil

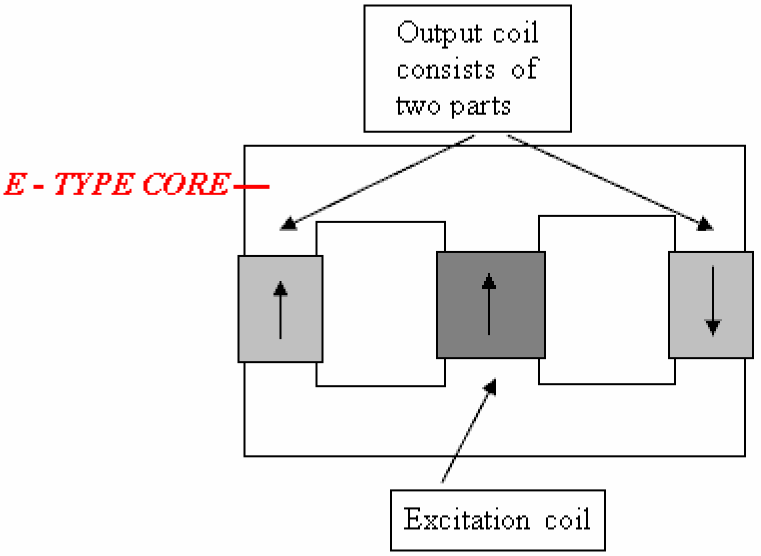
Output coil consists of two parts
E - TYPE CORE
Excitation coil

THE BASIS OF BACK EMF SUPPRESSION

(Tesla patent)

(No Model.) 2 Sheets—Sheet 1.

N. TESLA.

REGULATOR FOR DYNAMO ELECTRIC MACHINES.

No. 336,961. Patented Mar. 2, 1886.

Fig. 1.

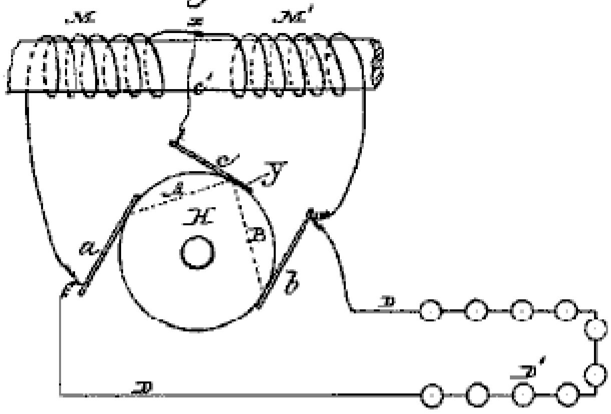
M
x
M'
c'
c
Y
A
H
B
a
b
D
D'
D

Fig. 2.

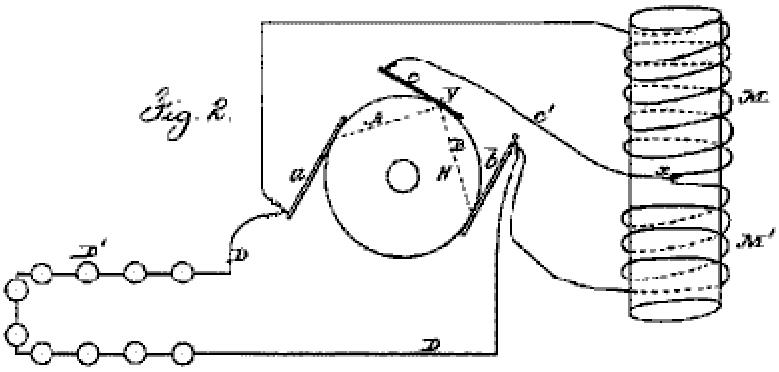
c
Y
A
B
H
a
b
c'
x
M
M'
D'
D
D

Fig. 3.

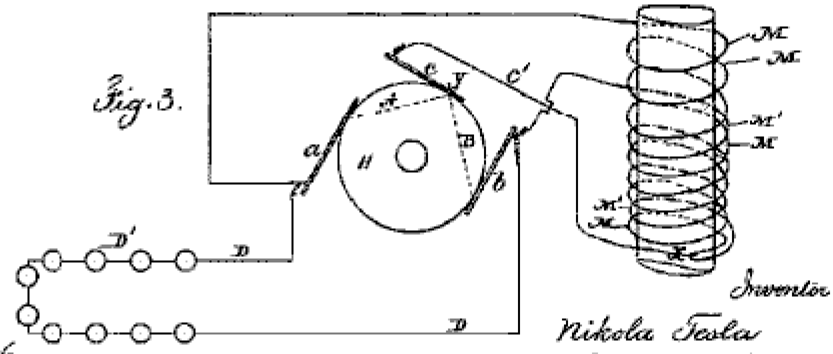
c
Y
c'
A
a
H
B
b
M
M
M'
M
M'
M
x
D'
D
D

Witnesses Chas. H. Smith
Geo. T. Pinckney

Inventor
Nikola Tesla
for Lemuel W. Serrell
atty

SECRET 1.2

The Spark-Exciting Generator (“SEG”)

(Charge delivering to LC circuit)

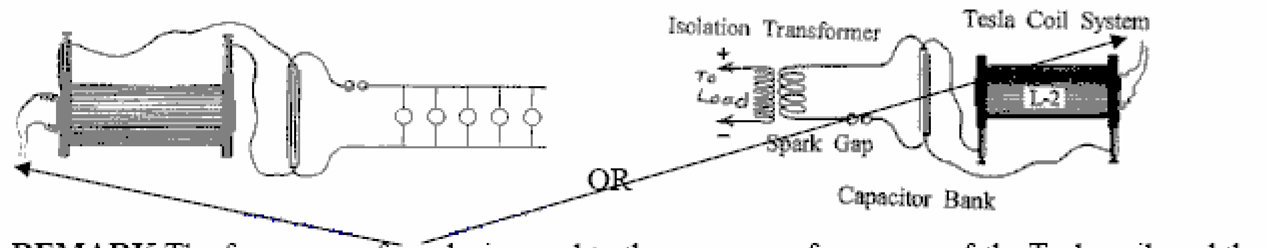
Isolation Transformer
Tesla Coil System
+
To Load
-
L-2
Spark Gap
OR
Capacitor Bank

REMARK The frequency of sparks is equal to the resonance frequency of the Tesla coil, and the moment of exciting corresponds to the maximum voltage on the Tesla coil.

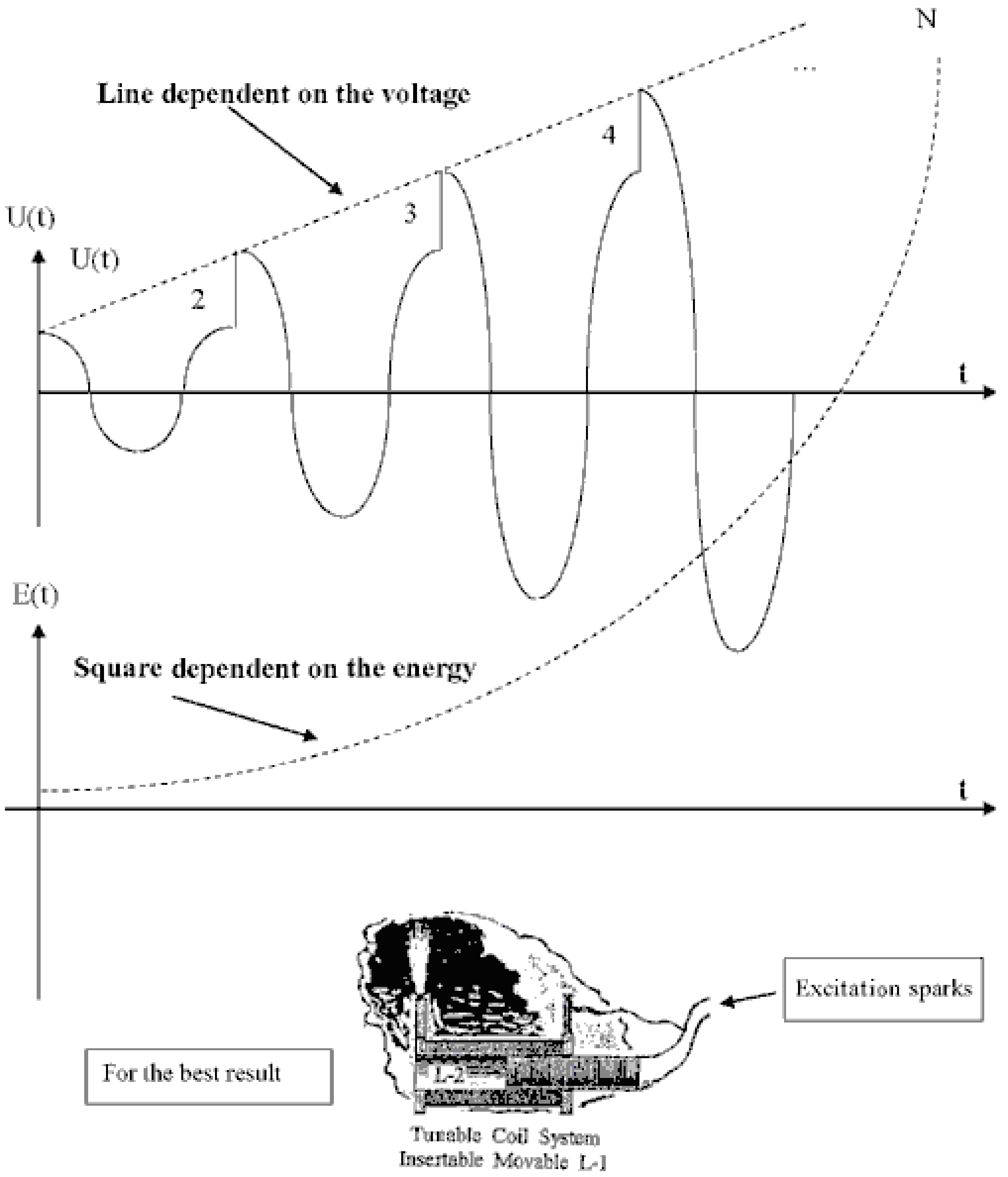
N
...
Line dependent on the voltage
4
3
U(t)
U(t)
2
t
E(t)
Square dependent on the energy
t
Excitation sparks
For the best result
L-2
Tunable Coil System
Insertable Movable L-1

EXPLANATION:

The spark delivers charge to the L-C circuit

The charge Q on a capacitor C with voltage U is: $Q = U \times C$ or $U = Q / C$

Where Q is a charge delivered by one spark.

During the excitation of the L-C circuit by the sparks, the capacitance C is constant.

After N excitations, the voltage Un on C will be Un = N x Q / C   And, energy En will be raised as N^2.

In other words, **If the L-C circuit is excited by charges, we have energy amplification**.

POSSIBLE MODIFICATION

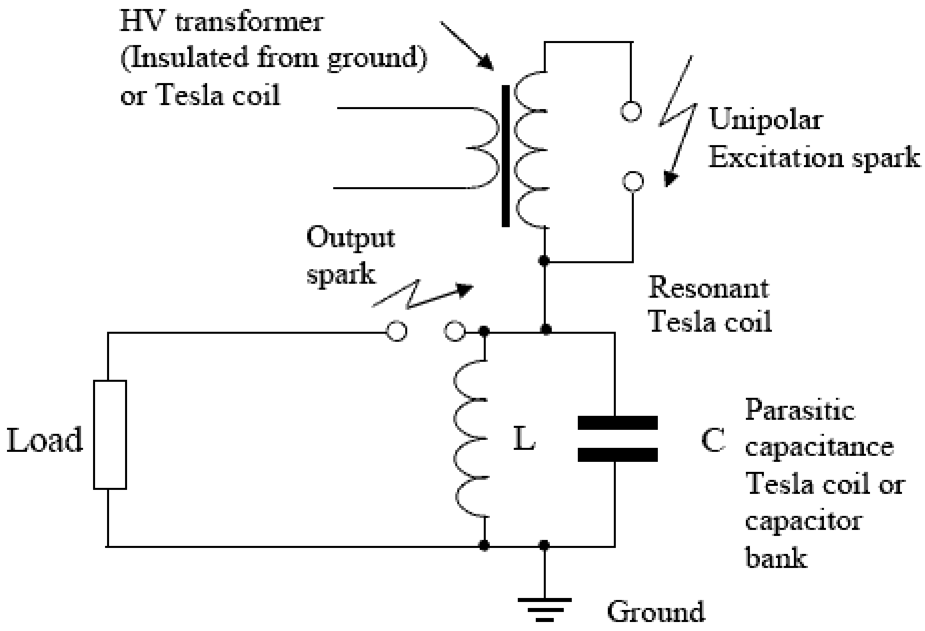
HV transformer
(Insulated from ground)
or Tesla coil
Unipolar
Excitation spark
Output
spark
Resonant
Tesla coil
Load
L
C
Parasitic
capacitance
Tesla coil or
capacitor
bank
Ground

COMMENT: You need to understand that a feedback loop in the electromagnetic field is a changing voltage level in the L-C circuit capacitor, a high-voltage transformer is connected to collect the excess energy.

WITHOUT SYNCHRONISATION

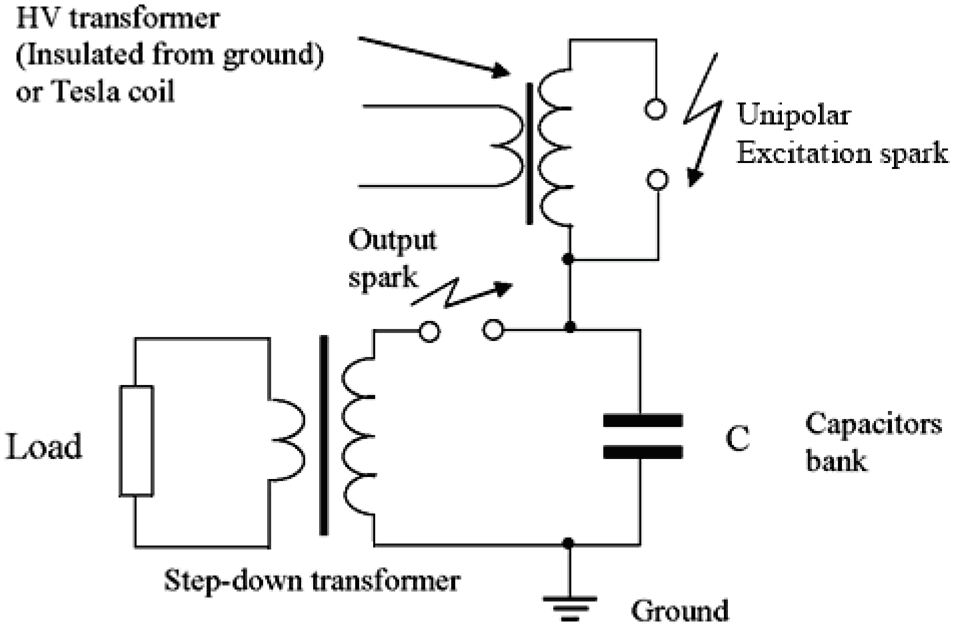
HV transformer
(Insulated from ground)
or Tesla coil
Unipolar
Excitation spark
Output
spark
Load
C
Capacitors
bank
Step-down transformer
Ground

The Spark-Exciting Generator

From Don Smith

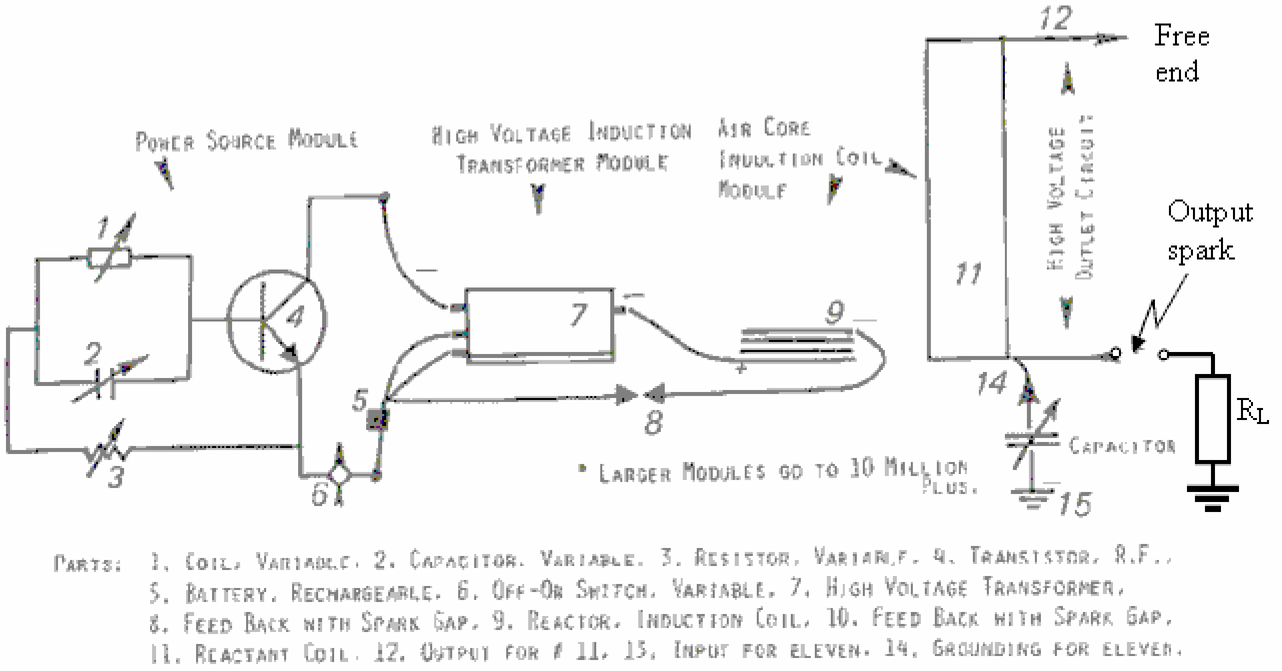
Power Source Module
High Voltage Induction Transformer Module
Air Core Induction Coil Module
12
Free end
High Voltage Outlet Circuit
11
Output spark
1
2
3
4
5
6
7
8
9
14
15
Capacitor
R_L
* Larger Modules go to 10 Million Plus.
Parts: 1. Coil, Variable. 2. Capacitor. Variable. 3. Resistor. Variable. 4. Transistor. R.F..
5. Battery. Rechargeable. 6. Off-On Switch. Variable. 7. High Voltage Transformer.
8. Feed Back with Spark Gap. 9. Reactor. Induction Coil. 10. Feed Back with Spark Gap.
11. Reactant Coil. 12. Output for # 11. 13. Input for eleven. 14. Grounding for eleven.

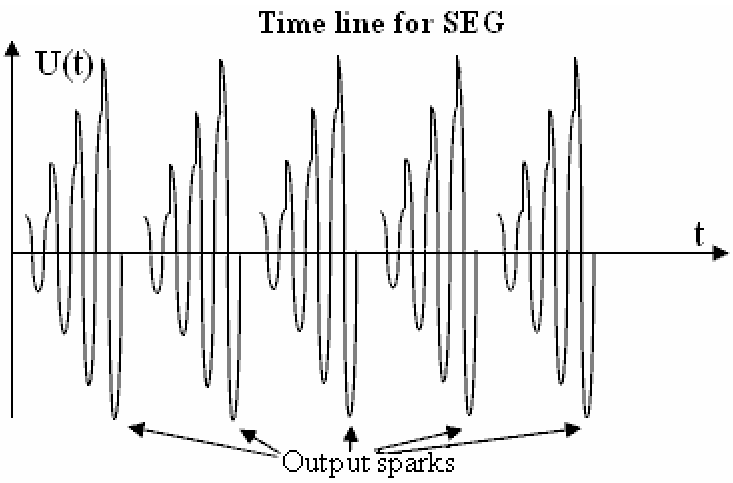
Time line for SEG

MAINTAIN RESONANCE AND GET FREE-ENERGY !!

EXPLANATION: It appears that we need to charge the capacitor circuit to an energy level which is greater than that of the source energy itself. At first glance, this appears to be an impossible task, but the problem is actually solved quite simply.

The charging system is screened, or "blinded", to use the terminology of Mr. Tesla, so that it cannot “see” the presence of the charge in the capacitor. To accomplish this, one end of a capacitor is connected to the ground and the other end is connected to the high-energy coil, the second end of which is free. After connecting to this higher energy level from the energising coil, electrons from the ground can charge a capacitor to a very high level.

In this case, the charging system does not "see" what charge is already in a capacitor. Each pulse is treated as if it were the first pulse ever generated. Thus, the capacitor can reach a higher energy level than of the source itself.

After the accumulation of the energy, it is discharged to the load through the discharge spark gap. After that, the process is repeated again and again indefinitely ...

COMMENT: The frequency of the excitation sparks, must match the resonant frequency of the output coil. (capacitors 2 and 14 are used to achieve this goal). This is multi-spark excitation.

COMMENT: Charges are pumping from the ground to 11-15 circuit, this device extracts charge from ambient space. Because of this, it will not work properly without a ground connection.

If you need Mains frequency, or don't want use an output spark, then read the following parts…
Asymmetrical transformers can be used (read the following parts)

POSSIBLE SEG ARRANGEMENT

(From Russian forum)

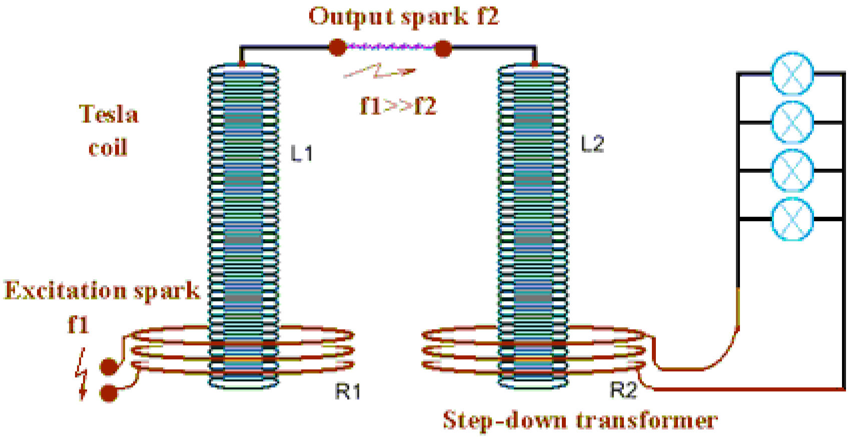
Output spark f2
f1>>f2
Tesla coil
L1
L2
Excitation spark f1
R1
R2
Step-down transformer

COMMENT: The L1 Tesla coil shown above, is energised by spark f1. Resonant, step-down transformer L2 is connected to the L1 Tesla coil by output spark f2. The frequency of f1 is much higher than that of f2**.**

SEG WITHOUT SYNCHRONISATION

From Don Smith

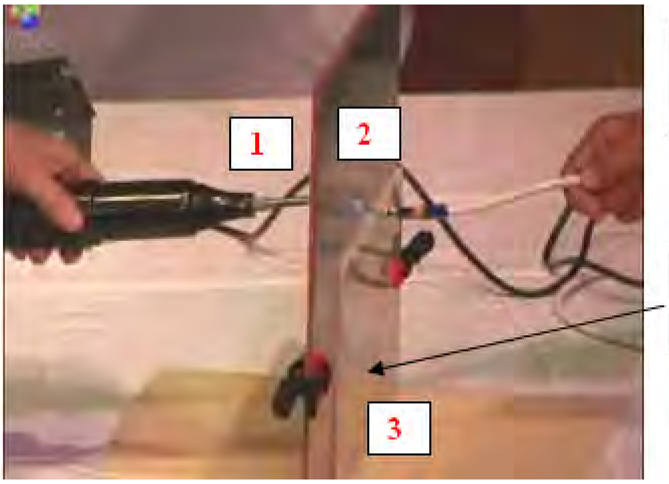

Three metal sheets are used

Pay attention to this part…

REMARK: It must be adjusted by dimensions, materials (???)

Possible schematics

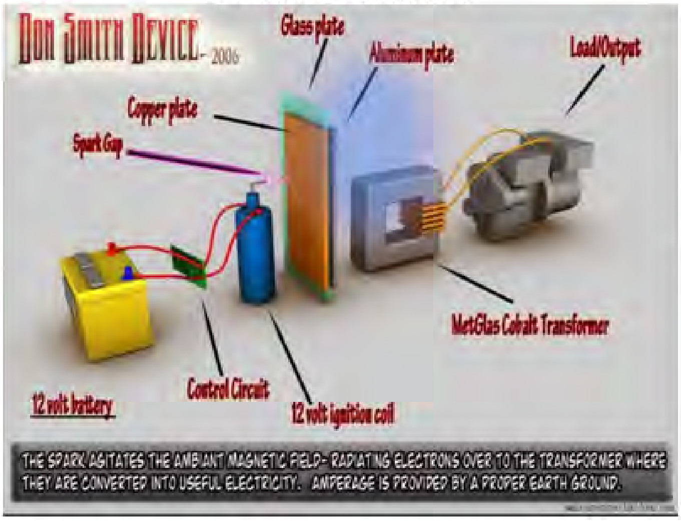
Don Smith Device - 2006
Glass plate
Aluminum plate
Load/Output
Copper plate
Spark Gap
MetGlas Cobalt Transformer
12 volt battery
Control Circuit
12 volt ignition coil
THE SPARK AGITATES THE AMBIANT MAGNETIC FIELD- RADIATING ELECTRONS OVER TO THE TRANSFORMER WHERE THEY ARE CONVERTED INTO USEFUL ELECTRICITY.   AMPERAGE IS PROVIDED BY A PROPER EARTH GROUND.

This is not completed…

EXPLANATION

REMINDER: **An ordinary capacitor is a device for separating charges on it's plates**, **The total charge inside an ordinary capacitor is zero** (read the textbooks).

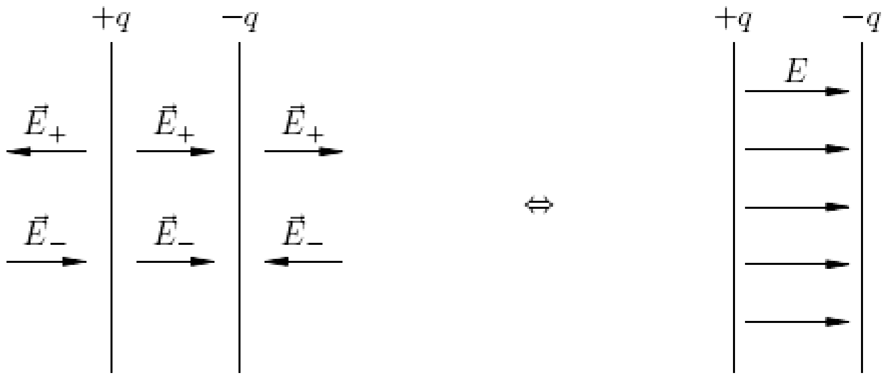

There is an electrical field only inside the capacitor. The electrical field outside the capacitor is zero (**because the fields cancel each other**).

So far, connecting one plate to the ground we will get no current flowing in this circuit:

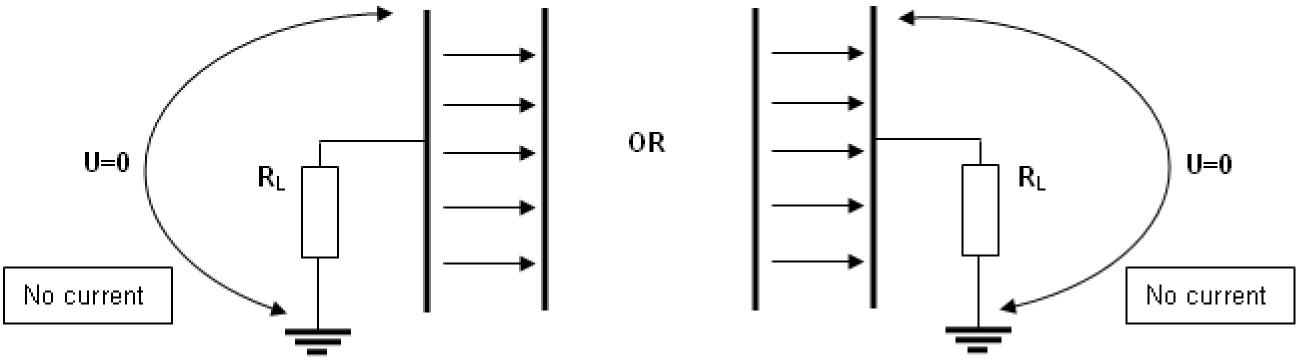

REMINDER: **A separated capacitor is a device for accumulating charges on it's plates**.

The total charge on a separated capacitor is NOT zero (read the textbooks). So far, by connecting one plate of the separated capacitor to the ground we will get a current flowing in this circuit (because there is an external field).

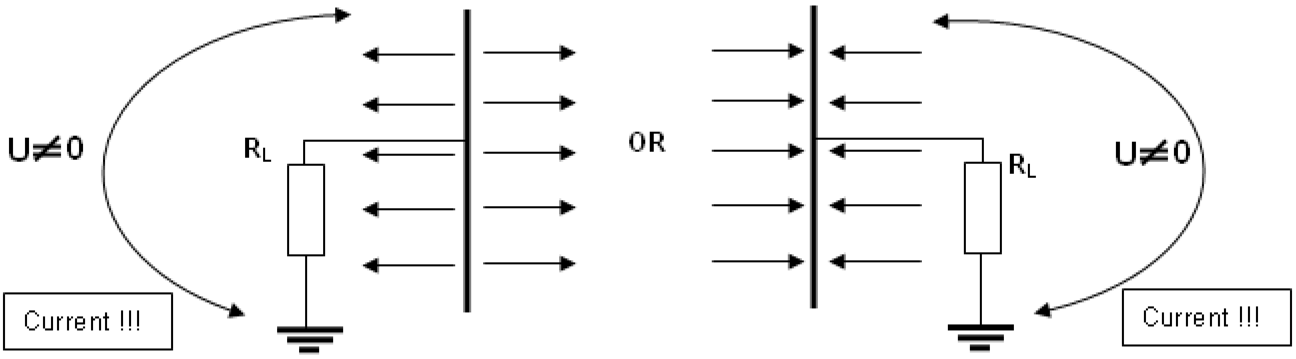

REMARK: We get the same situation, if only one plate of an ordinary capacitor is charged. So far, connecting an **uncharged plate** of an ordinary capacitor to the ground we get a current flowing in this circuit also (because there is an external field).

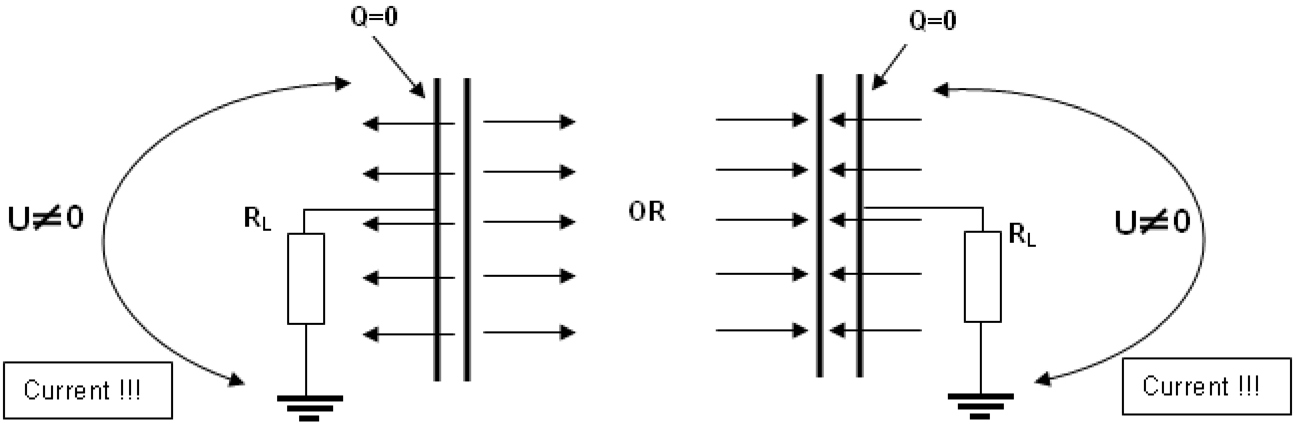
Q=0
U≠0
RL
Current !!!
OR
Q=0
RL
U≠0
Current !!!

Alternately charging a capacitor's plates

Avramenko's plug – is it a free energy device?

The principle: Each plate of a capacitor charges as a separated capacitor. Charging takes place in an alternating fashion, first one plate and then the other plate.

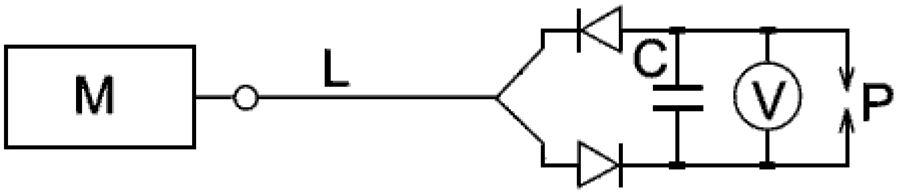

The result: The capacitor is charged to a voltage which is greater than that which the charging system delivers.

Explanation: The external field of an ordinary charged capacitor is equal to or near zero, as noted above. So, if you charge plates as a separated capacitor (upload or download charge), the charging system will not "see" the field which already exists inside the capacitor, and will charge the plates as if the field inside the capacitor is absent.

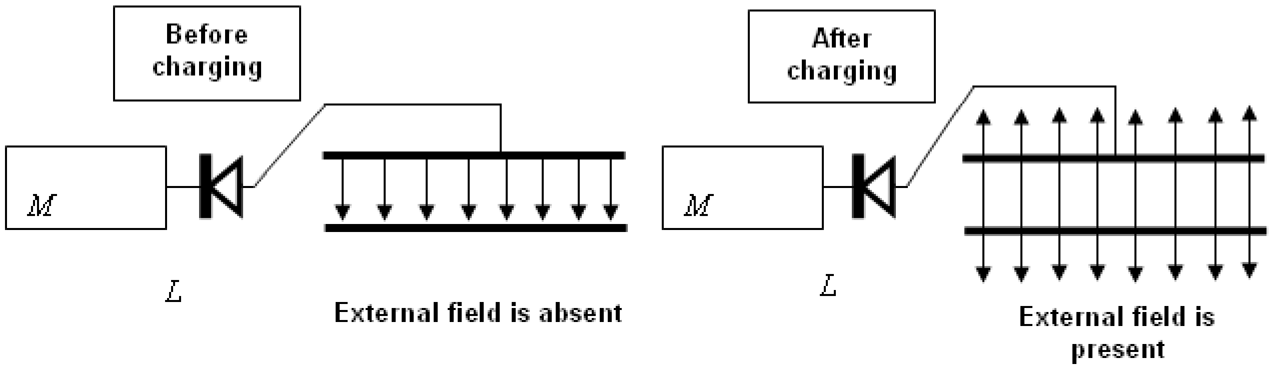
Before charging
M
L
External field is absent
After charging
M
L
External field is present

Once a plate has been charged, begin to charge another plate.

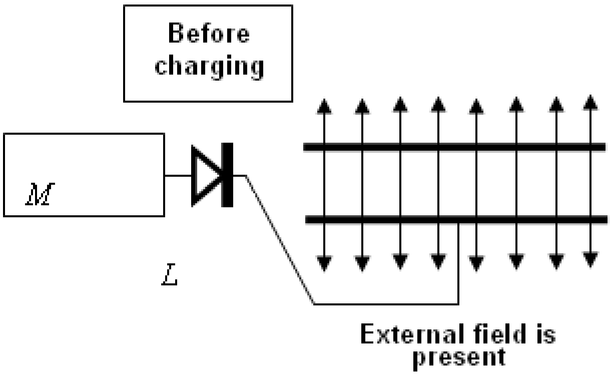
Before charging
M
L
External field is present

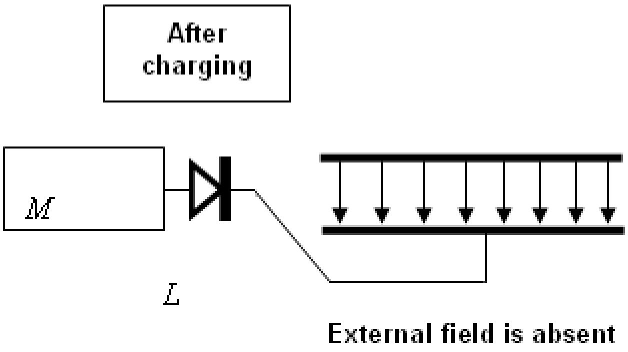
After charging
M
L
External field is absent

After the second plate of the capacitor has been charged, the external field becomes zero again. The charging system cannot "see" the field inside the capacitor once again and the process repeats again several times, raising the voltage until the spark gap connected to the output load discharges it.

REMARK: You will recall that an ordinary capacitor is a device for charge separation. The charging process of a capacitor causes electrons from on one plate to be "pumped" to another plate. After that, there is an excess of electrons on one plate, while the other one has deficit, and that creates a potential difference between them (read the textbooks). The total amount of charge inside the capacitor does not change. Thus **the task of the charging system is to move charge temporarily from one plate to another**.

The simplest Free-Energy device (???)

REMARK: The capacitance of an ordinary capacitor is much greater than the capacitance of a separated plate capacitor (provided that it's plates are close to each other).

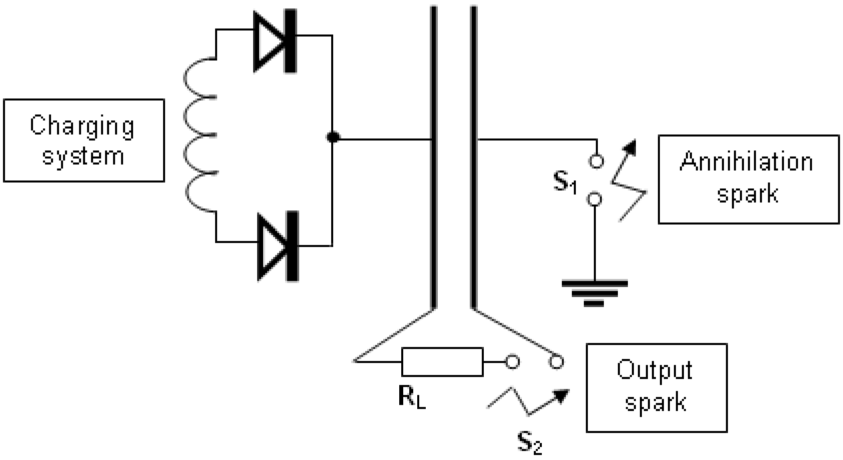
Charging system
S1
Annihilation spark
RL
S2
Output spark

COMMENT: The time between S1 and S2 is very short.

FIELD BEFORE S1

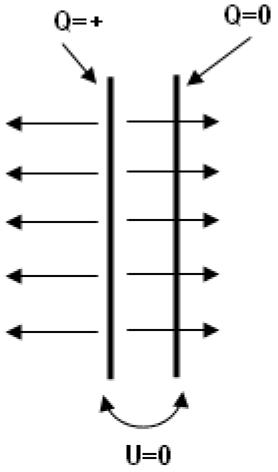

FIELD AFTER S1

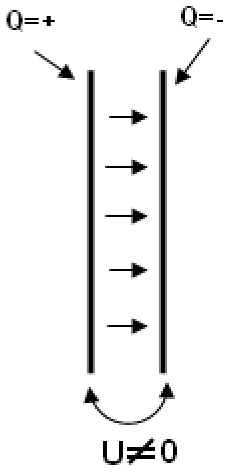

REMARKS: This is an illustration of energy-dependence in a coordinated system.
This is an illustration of the so-called Zero-Point Energy.

ASYMMETRICAL CAPACITOR

(Current amplification???)

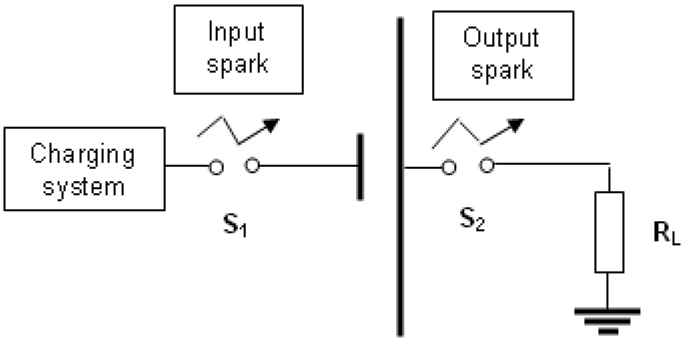
Input spark
Output spark
Charging system
S1
S2
RL

COMMENT: The capacitance (size) of the plate on the right is much greater than that of the plate on the left.

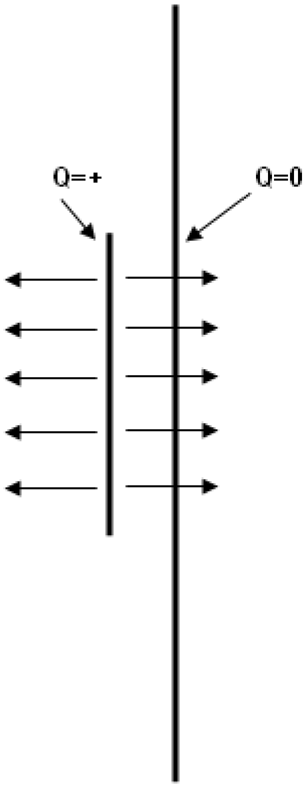

FIELD AFTER S1

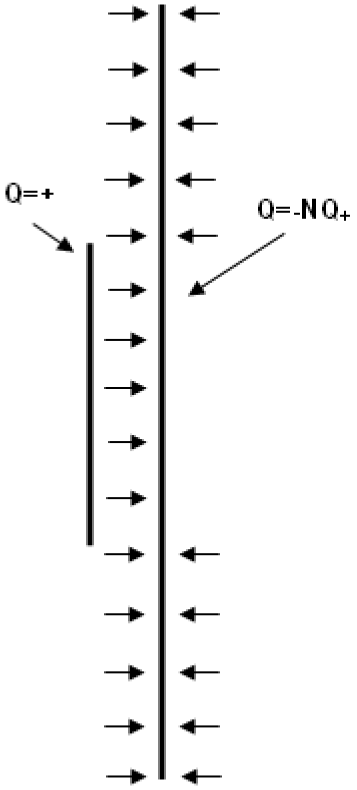

FIELD AFTER S2

COMMENT: Charges from the ground will run on to the right hand plate **UNTIL** the moment when the external field drops to zero caused by the second spark ("S2"). It takes more charges flowing from the ground to annihilate the external field at the instant of the second spark, because the capacitance of the plate on the right is far greater. 'More charge' means 'more current', so you have achieved current amplification through this arrangement.

COMMENT: The field at the terminals of the plate on the right is not zero after both sparks have occurred, this is because a field remains due to the additional charges which have flowed in ('pumped') from the ground.

THE SIMPLEST ASYMMETRICAL CAPACITORS

The most simple asymmetrical capacitors are the Leyden jar and the coaxial cable (also invented by Mr. Tesla).

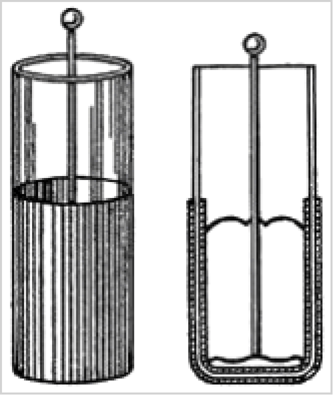

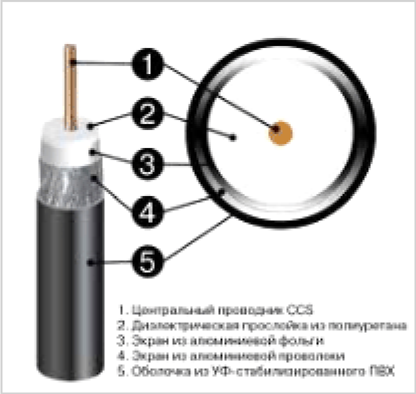
1
2
3
4
5
1. Центральный проводник CCS
2. Диэлектрическая прослойка из полиуретана
3. Экран из алюминиевой фольги
4. Экран из алюминиевой проволоки
5. Оболочка из УФ-стабилизированного ПВХ

Apart from the fact that the area (capacitance) of the plates of these capacitors is different, and they therefore are asymmetrical, they have another property:

The electrostatic field of the external electrode of these devices does not affect the internal electrode.

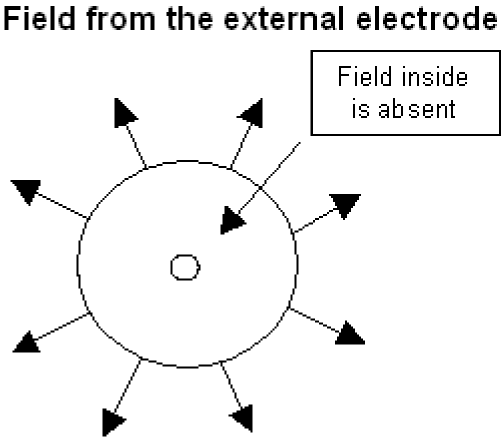
Field from the external electrode
Field inside is absent

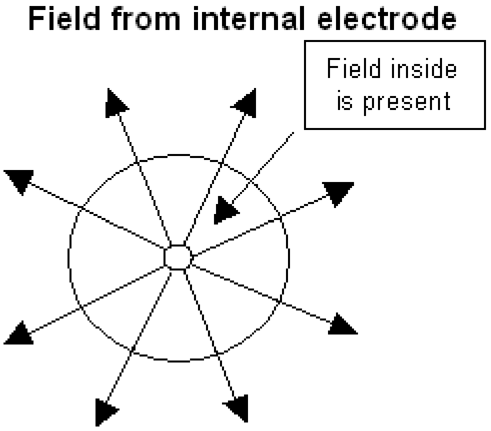
Field from internal electrode
Field inside is present

EXPLANATION: This is caused by the fact that the electrostatic field is absent inside the metal bodies (see textbooks).

REMARK: This is true provided that the plates are charged separately.

CAPACITOR - TRIODE

(The third electrode in an asymmetrical capacitor)

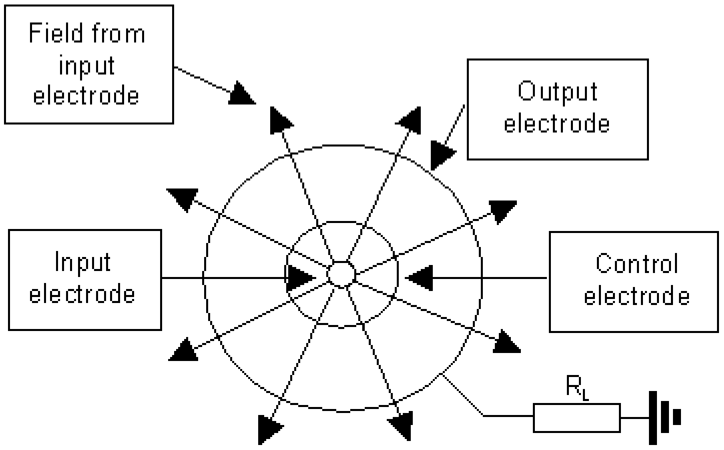
Field from input electrode
Output electrode
Input electrode
Control electrode
R_L

REMARK: Dr. Harold Aspden has pointed out the possibility of Energy Amplification when using this device.

THE PRINCIPLE OF CURRENT AMPLIFICATION IN THE CAPACITOR – TRIODE

EXPLANATION: You have to get zero potential on the inside of a small cylinder (on the input electrode). In this case, the charge on the external cylinder will be more than on the internal cylinder. More charge means more current, and so you'll have current amplification.

In detail:
The potential around any cylinder with radius 'R' is:

$$\varphi = \frac{q}{4\pi\varepsilon\varepsilon_0 R}.$$

where q is a charge on this cylinder

Potential inside this cylinder is the same, because:

$$\frac{d\varphi}{dr} = 0 \quad \text{and} \quad \varphi = const.$$

If you want to get zero potential inside two cylinders (on the input electrode) you have to deliver more charge of the opposite sign to the external one, proportionally to the radius of this cylinder. A larger radius means more charge.

OPERATIONAL SEQUENCE:

1. Charge the input electrode from your source of energy.
2. The small cylinder (control electrode) will be charged automatically, if it's connected to the ground through a diode with the properly polarity.
3. Discharge the input electrode to zero level (for example, by using a spark). As a result there will be a zero potential on it.
4. If the external cylinder is connected to the ground through a diode with the properly polarity, it will be charged automatically with the opposite sign. But the charge "pumped" from the ground will more than for small cylinder (proportional to the ratio of their radiuses).
5. As a result, there is current amplification.

COMMENT: Did Edwin Gray use this principle in his device?.

FIG. 2

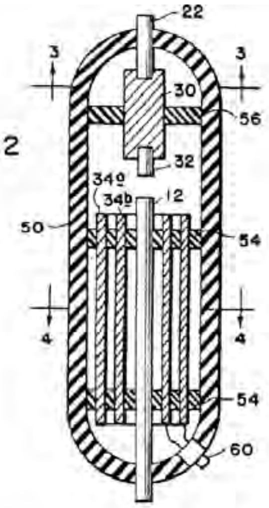

FIG. 3

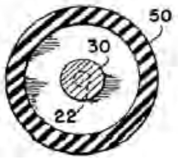

FIG. 4

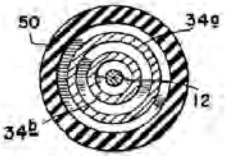

COMMENT: If so, then Gray's patent is lacking some very important details (???)

THE PRINCIPLE OF THE “BLINDNESS” CHARGING SYSTEM IN THE SEG

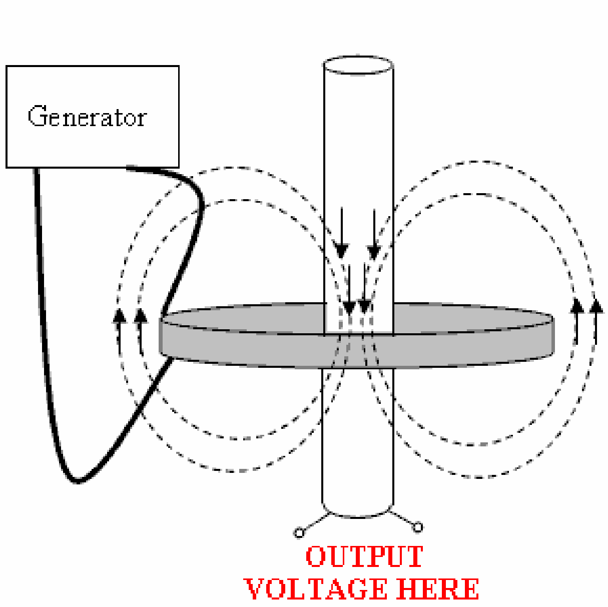
Generator
OUTPUT
VOLTAGE HERE

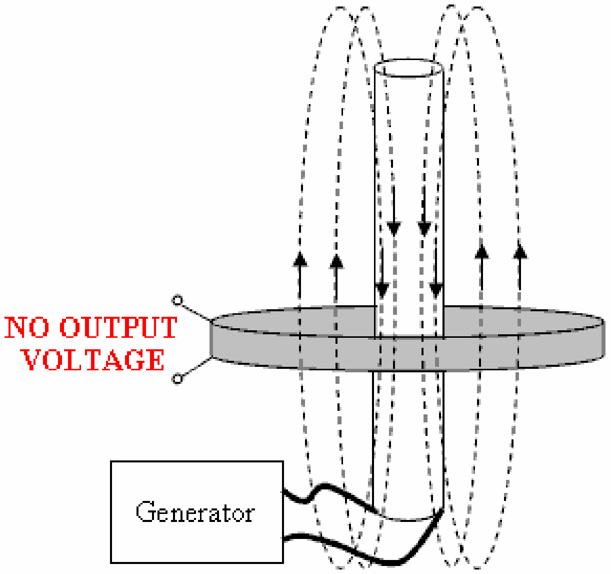
NO OUTPUT
VOLTAGE
Generator

EXPLANATION: The “short” coil is not able to see the oscillations in the “long” coil, because the total number of magnetic lines from the “long” coil which are passing through the “short” coil is close to zero (because one half is in one direction and the other half is in the opposite direction).

COMMENT: This a particular case of an asymmetrical transformer, for more details read the part about asymmetrical transformers.

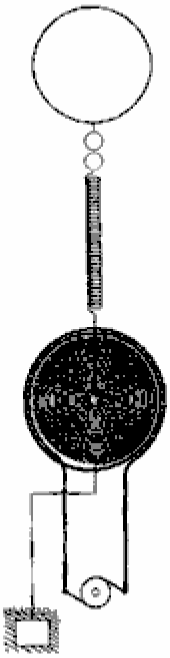

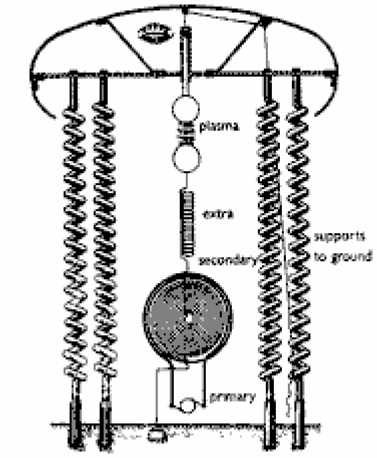
plasma
extra
secondary
supports
to ground
primary

Fig. 6 Oscillating electrostatically charged dome.

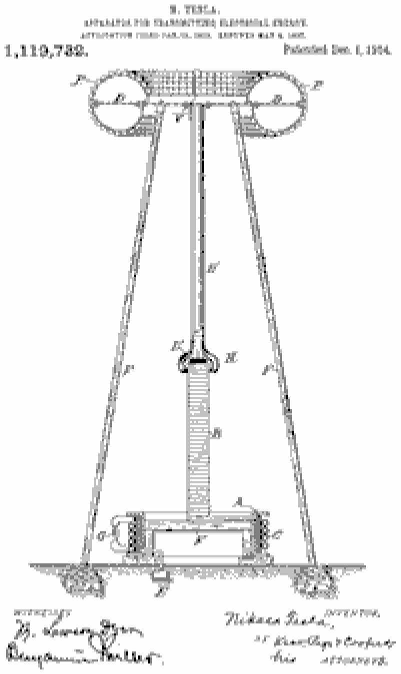
1,119,732.
Patented Dec. 1, 1914.
Nikola Tesla

COMMENTS ABOUT THE SEG:
All Back EMF schematics can be used in SEG

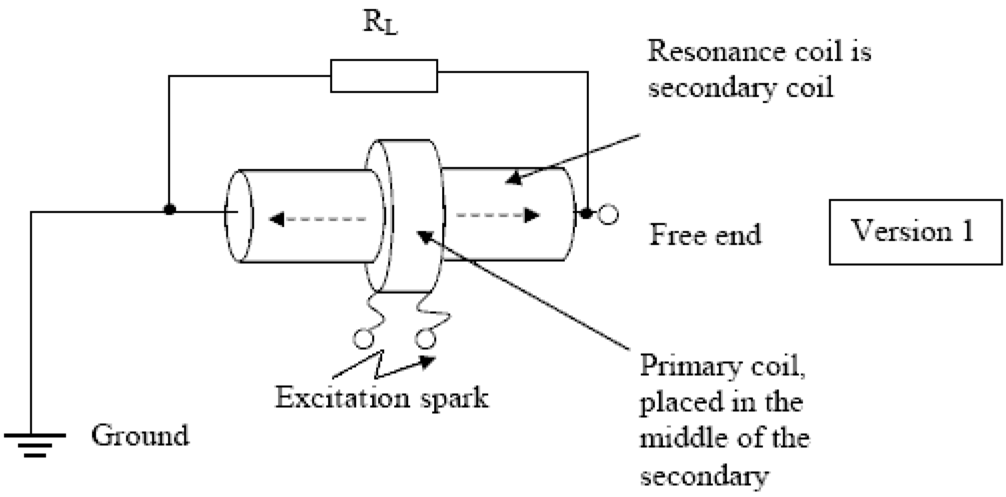

RL
Resonance coil is secondary coil
Free end
Version 1
Excitation spark
Ground
Primary coil, placed in the middle of the secondary

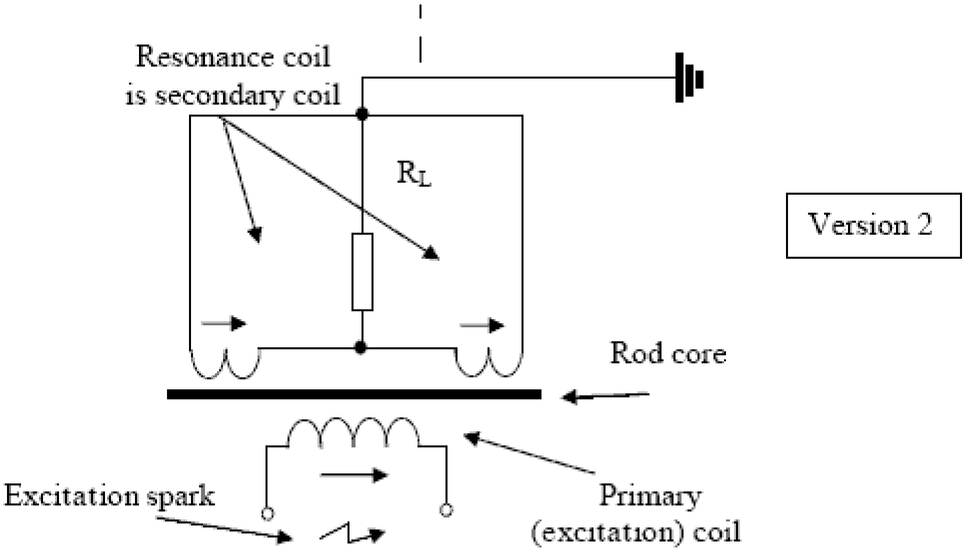

Resonance coil is secondary coil
RL
Version 2
Rod core
Excitation spark
Primary (excitation) coil

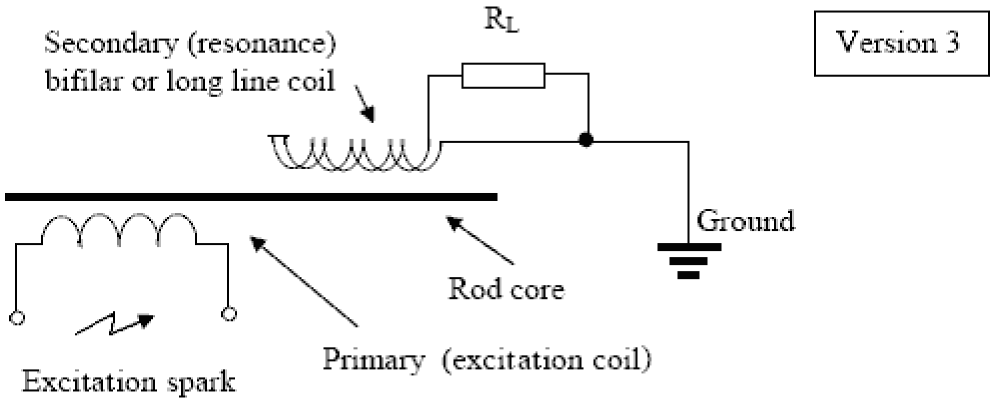

Secondary (resonance) bifilar or long line coil
RL
Version 3
Ground
Rod core
Primary (excitation coil)
Excitation spark

COMMENTS: No current will be produced in the load in any of these circuits, unless there is a ground connection. Is excitation possible with just a single spark (???)

FOR MORE ASYMMETRY IN SEG ?
FOR ONE SPARK EXCITING IN SEG ?

By Don Smith

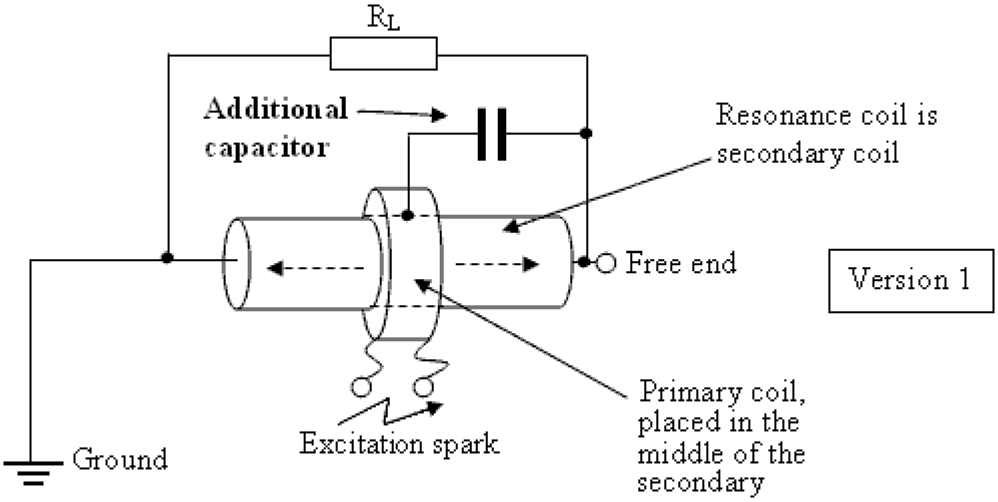
RL
Additional capacitor
Resonance coil is secondary coil
Free end
Version 1
Primary coil, placed in the middle of the secondary
Excitation spark
Ground

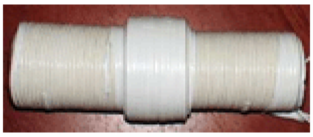

Prototype coil

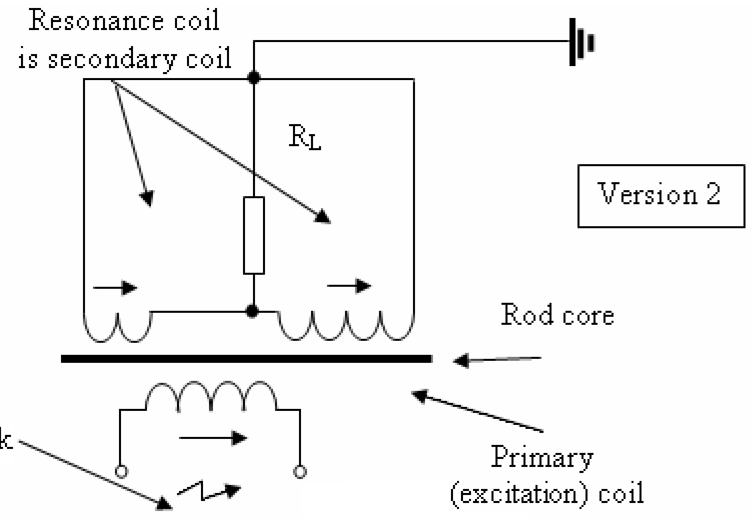
Resonance coil is secondary coil
RL
Version 2
Rod core
Excitation spark
Primary (excitation) coil

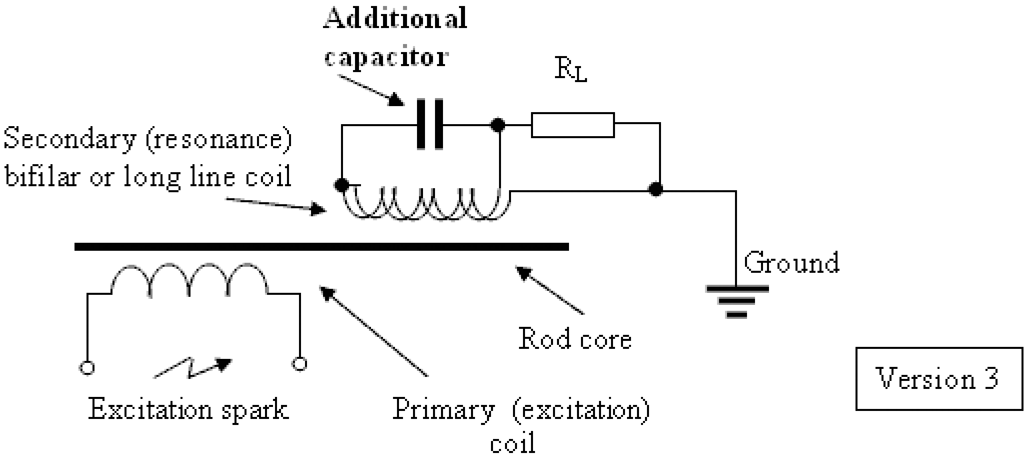
Additional capacitor
RL
Secondary (resonance) bifilar or long line coil
Ground
Rod core
Version 3
Excitation spark
Primary (excitation) coil

COMMENT: This arrangement becomes more asymmetrical after excitation

EXPLANATION

Symmetry is destroyed by a spark

If the impedances of Ra and Rc are the same at the frequency produced by signal generator F1, then the resulting voltage at points **A** and **B** will also be identical which means that there will be zero output.

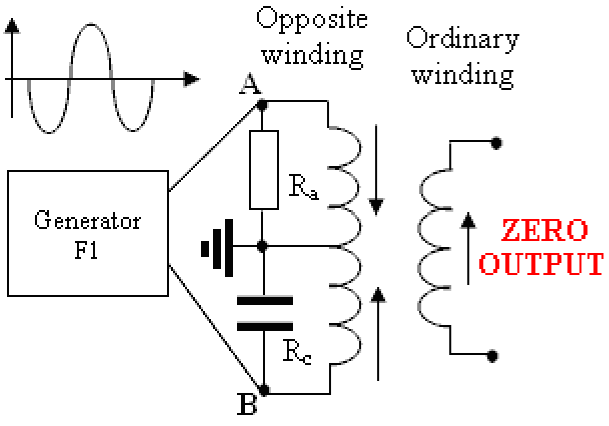

Opposite winding
Ordinary winding
A
R_a
Generator F1
R_c
B
ZERO OUTPUT

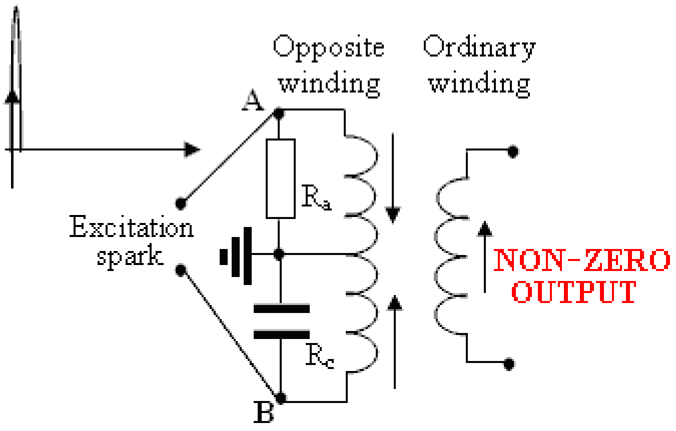

Opposite winding
Ordinary winding
A
Excitation spark
R_a
R_c
B
NON-ZERO OUTPUT

If the circuit is excited by the very sharp, positive-only, DC voltage spike produced by a spark, then the impedances of Ra and Rc are not the same and there is a non-zero output.

Here is a possible alternative. Please note that the position of the output coil must be adjusted, it's best position depending on value of resistor Rc and the frequency being produced by signal generator F1.

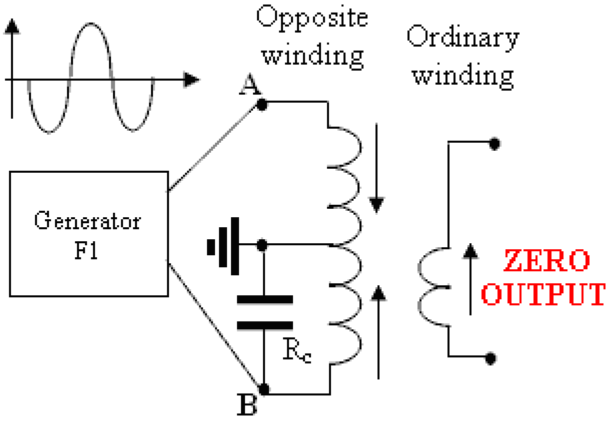

Opposite winding
Ordinary winding
A
Generator F1
R_c
B
ZERO OUTPUT

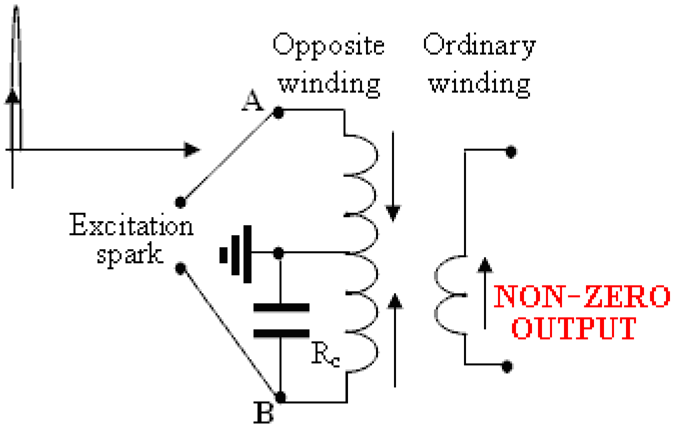

Opposite winding
Ordinary winding
A
Excitation spark
R_c
B
NON-ZERO OUTPUT

Here is another possible arrangement. Here, the position of the output coil depends on L1 and L2:

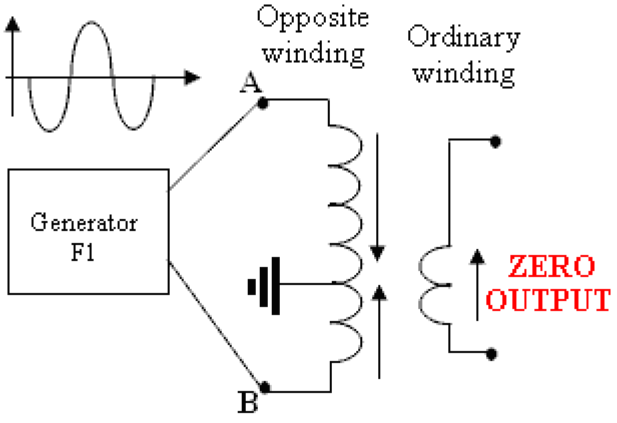

Opposite winding
Ordinary winding
A
Generator F1
B
ZERO OUTPUT

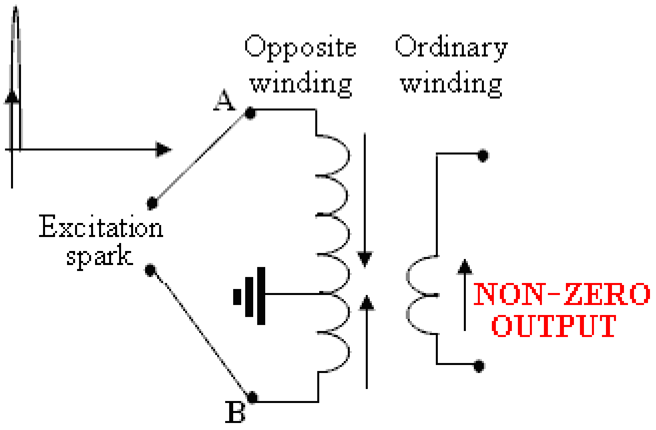

Opposite winding
Ordinary winding
A
Excitation spark
B
NON-ZERO OUTPUT

A NOMOGRAPH

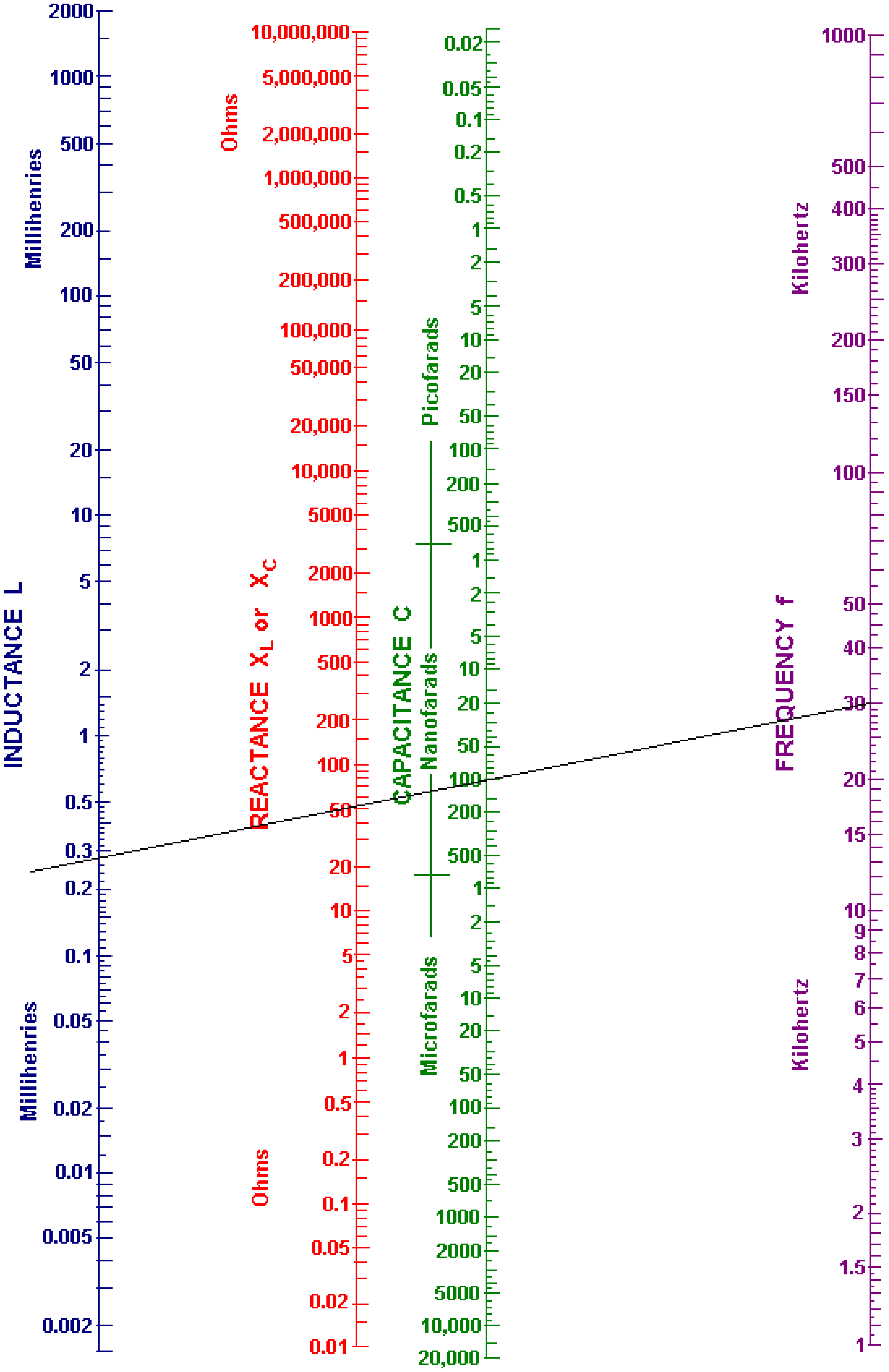

Using a nomograph: Draw a straight line from your chosen 30 kHz frequency (purple line) through your chosen 100 nanofarad capacitor value and carry the line on as far as the (blue) inductance line as shown above.

You can now read the reactance off the red line, which looks like 51 ohms to me. This means that when the circuit is running at a frequency of 30 kHz, then the current flow through your 100 nF capacitor will be the same as through a 51 ohm resistor. Reading off the blue "Inductance" line that same current flow at that frequency would occur with a coil which has an inductance of 0.28 millihenries.

MODERN OPTIONS IN SEG

Back EMF suppression in resonance coil

Version 3

By Don Smith

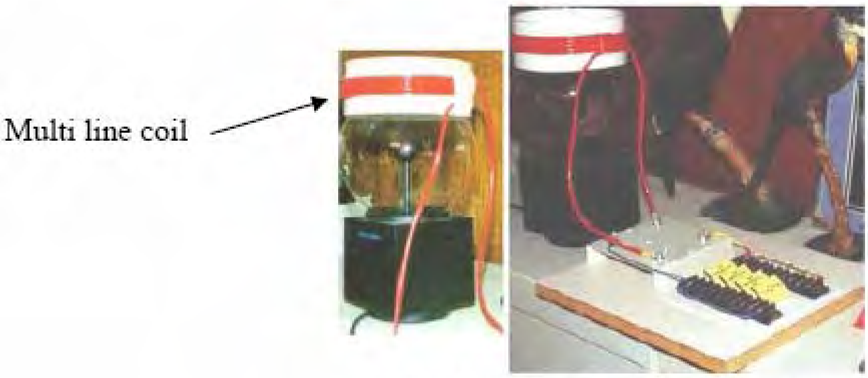

Globe device for energy generation

COMMENT: Please note that a long wire is used and one-spark excitation, where additional capacitors are used to create non-symmetry (???)

Version???

By Don Smith

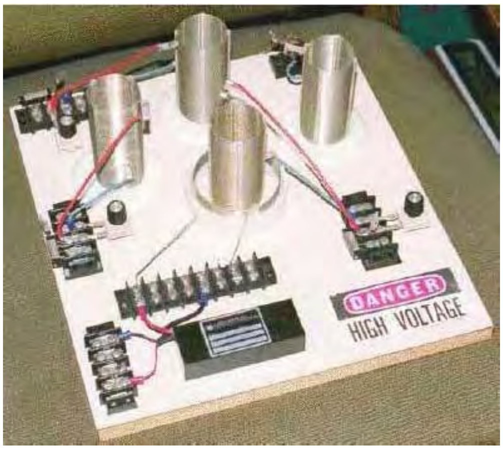

Multi coil system for energy multiplication

Version???

By Tariel Kapanadze

No description, so read the following section…

KAPANADZE PROCESS

The process requires only 4 steps:

STEP 1

An L-C (coil-capacitor) circuit is pulsed and it's resonant frequency determined (possibly by feeding it power through a spark gap and adjusting a nearby coil for maximum power collection).

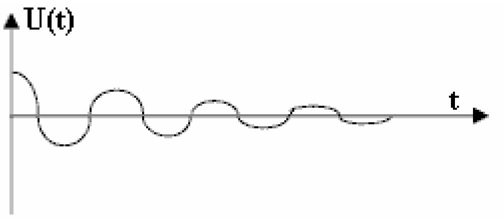

STEP 2

The SEG process causes the energy level in the L-C circuit to rise. Power is fed via a spark gap which produces a very sharp square wave signal which contains every frequency in it. The L-C circuit automatically resonates at it's own frequency in the same way that a bell always produces the same musical frequency when struck, no matter how it is struck.

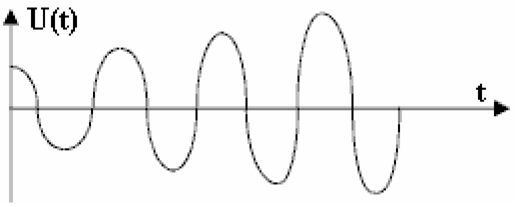

STEP 3

The output waveform from the L-C circuit is then manipulated to provide an output which oscillates at the frequency on the local mains supply (50 Hz or 60 Hz typically).

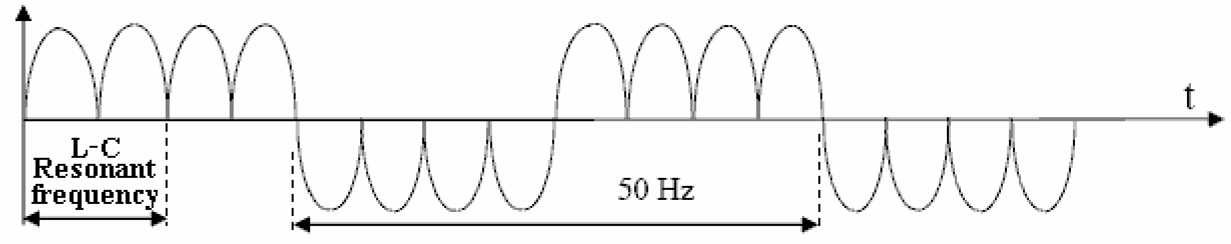

STEP 4

Finally, the oscillations are smoothed by filtering to provide mains-frequency output power.

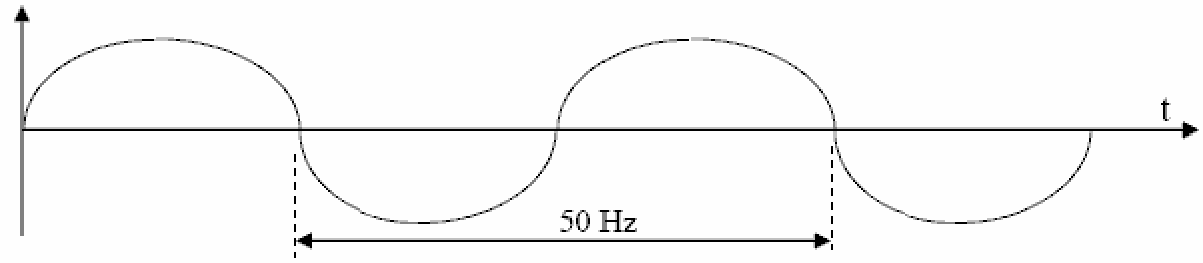

COMMENT: All of these processes are described in Kapanadze's patents and so, no state or private confidential information is shown here. Kapanadze's process is the SEG process.

COMMENT: As I see it, the main difference between the designs of Don Smith and Tariel Kapanadze is the inverter or modulator in the output circuit. At mains frequency you need a huge transformer core in a powerful inverter.

Read the following parts to discover more secrets…

MODERN OPTION

Lowering the L-C frequency to mains frequency (Modulation)

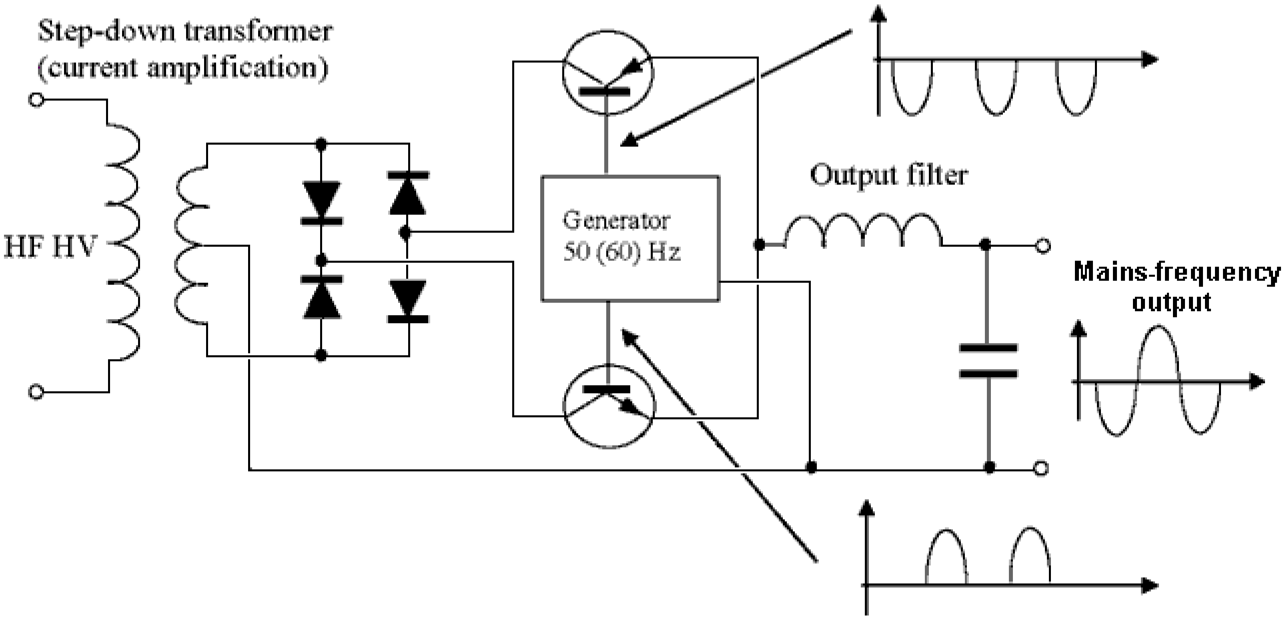
Step-down transformer
(current amplification)
HF HV
Generator
50 (60) Hz
Output filter
Mains-frequency
output

COMMENTS: It is possible to use square waves instead of sine waves to ease the loading on the transistors. This is very similar to the output sections of Tariel Kapanadze's patents. This method does not require a powerful transformer with a huge core in order to provide 50 Hz or 60 Hz.

Don Smith's option (guessed at by Patrick Kelly)

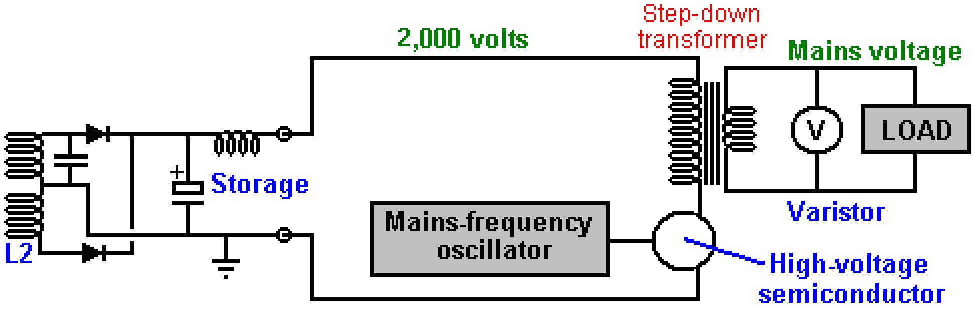
2,000 volts
Step-down
transformer
Mains voltage
V
LOAD
L2
Storage
Mains-frequency
oscillator
Varistor
High-voltage
semiconductor

COMMENT : There is no high-frequency high-voltage step-down transformer, but a step-down transformer is used for mains frequency which means that it will need a huge core.

FOR BOTH SCHEMATICS:

You must choose the load in order to get the maximum power output. Very low, and very high loads will give almost no energy in the load (because the current flowing in the output circuit is restricted by the current flowing in the resonant circuit).

ILLUSTRATIONS FOR FREQUENCY LOWERING

From Tariel Kapanadze

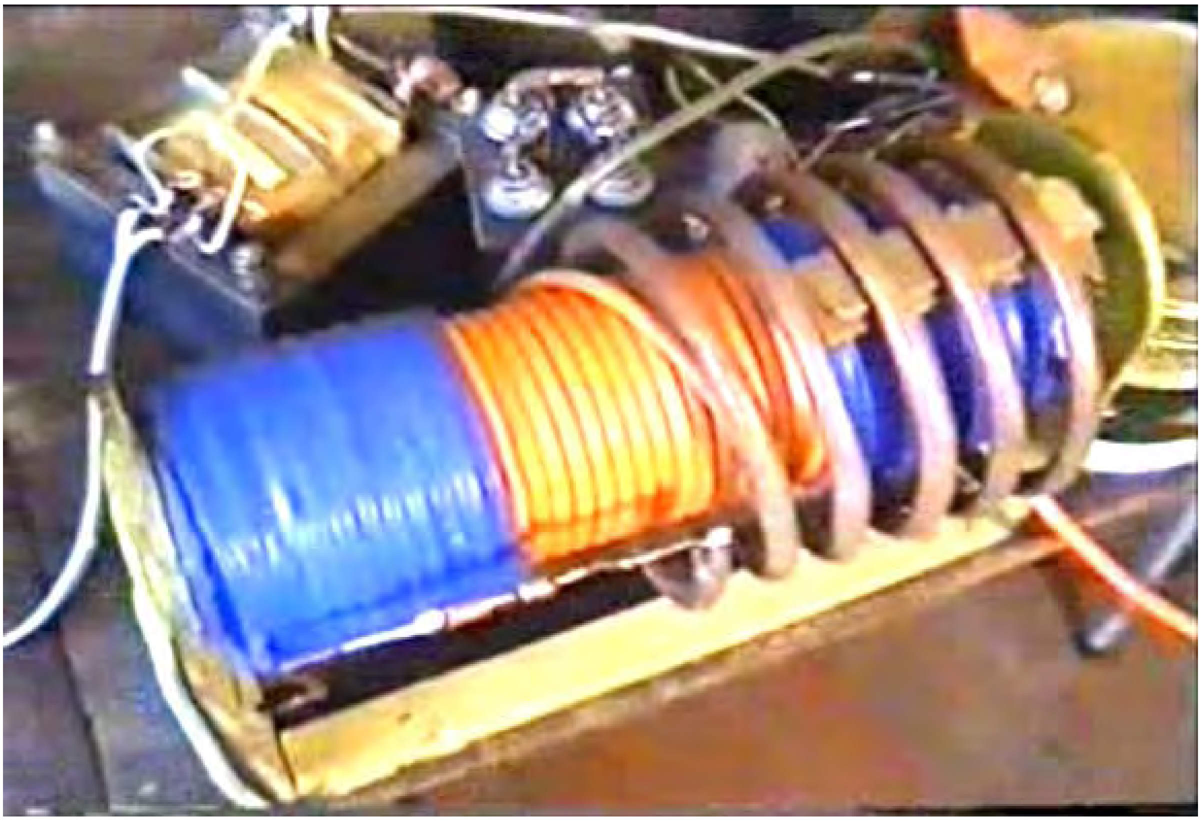

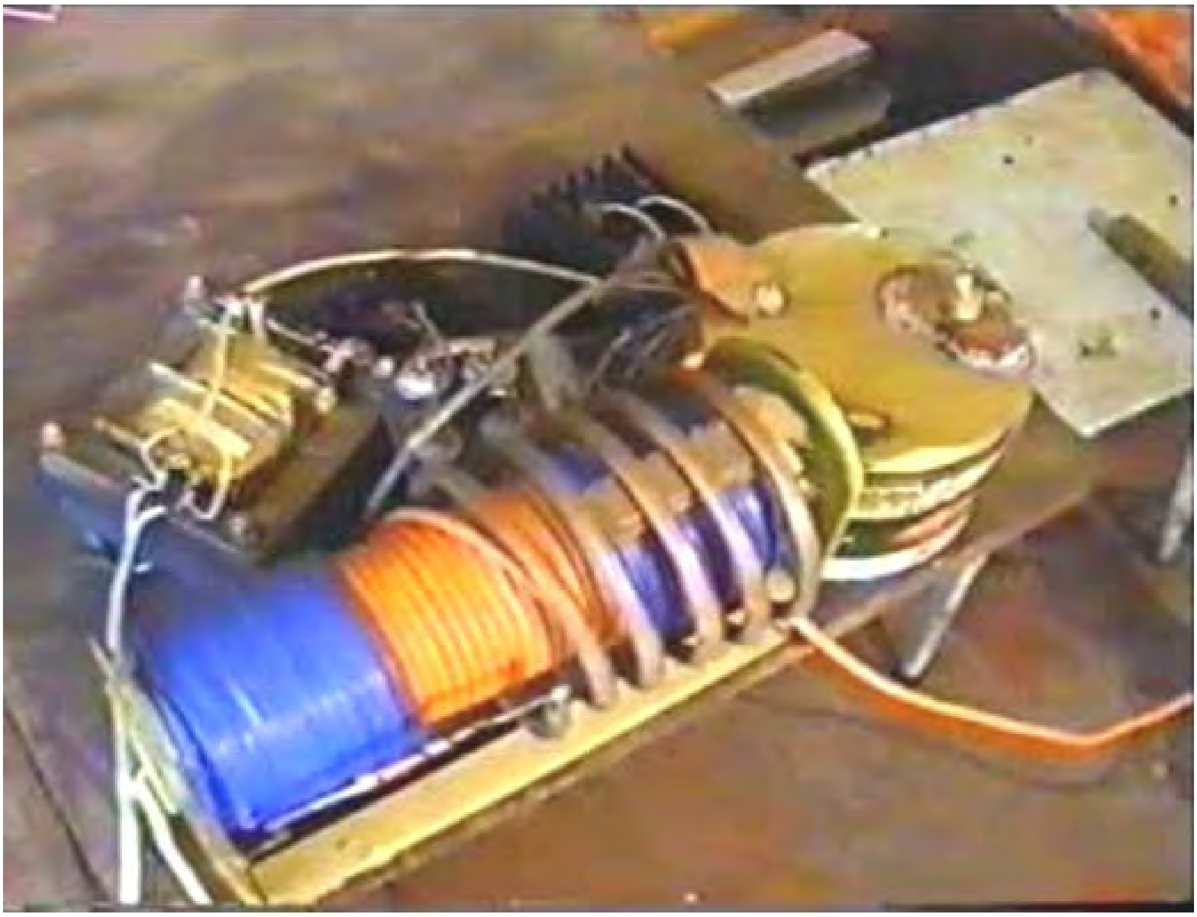

ENERGY GAIN

(REMARKS on 1.1 and 1.2 SECRETS)

We must consider two options:
1. Back-EMF suppression . . . . . (1.1)
2. Excitation by a spark . . . . (1.2).

THESE OPTIONS ARE DIFFERENT

However, in both cases, an increase of energy occurs due to the charges being pumped in from the ground. In the terminology of Mr. Tesla – "a charge funnel" or in modern terminology "a charge pump".

1. In the first case, the problem for the oscillating circuit is to "create" an electromagnetic field which has a high intensity electrical component in ambient space. (Ideally, it is only necessary for the high-voltage capacitor be fully charged once. After that, if the circuit is lossless, then oscillation will be maintained indefinitely without the need for any further input power).

THIS IS A "BAIT" TO ATTRACT CHARGES FROM THE AMBIENT SPACE.

Only a tiny amount of energy is needed to create such a "bait"...

Next, move the "bait" to one side of the circuit, the side which is the source of the charges (Ground). The separation between the "bait" and the charges is now so small that breakdown occurs. The inherent parasitic capacitance of the circuit will be instantly charged, creating a voltage difference at the opposite ends of the circuit, which in turn causes spurious oscillations. The energy contained in these oscillations is the energy gain which we want to capture and use. This energy powers the load. This very useful electromagnetic field containing our excess power oscillates in a direction which is perpendicular to the direction of oscillation of the "bait" field and because of this very important difference, the output power oscillations do not destroy it. This vital factor happens because the coil is wound with two opposing halves. The parasitic oscillations gradually die out, passing all of their energy to the load.

This energy-gaining process is repeated, spark by spark. The more often a spark occurs, the higher the excess power output will be. That is, the higher the spark frequency (caused by a higher voltage across the spark gap), the higher the power output and the greater the efficiency of the process. Hardly any additional "bait" energy is ever required.

2. In the second case we must charge the capacitor circuit to an energy level higher than that of the source energy itself. At first glance, this appears to be an impossible task, but the problem is solved quite easily.

The charging system is screened, or "blinded", to use the terminology of Mr. Tesla, so that it cannot "see" the presence of the charge in the capacitor. To accomplish this, one end of a capacitor is connected to the ground and the other end is connected to the high-energy coil, the second end of which is free. After connecting to this higher energy level from the energising coil, electrons from the ground can charge a capacitor to a very high level.

In this case, the charging system does not "see" what charge is already in a capacitor. Each pulse is treated as if it were the first pulse ever generated. Thus, the capacitor can reach a higher energy level than that of the source itself.

After the accumulation of the energy, it is discharged to the load through the discharge spark gap. After that, the process is repeated again and again indefinitely ...

THIS PROCESS DOES NOT REQUIRE THE SUPPRESSION OF BACK-EMF

3. It should be noted, that option 1 and option 2 above could be combined.

SECRET 2

SWITCHABLE INDUCTANCE

The inductance is comprised of two coils which are positioned close to each other. Their connections are shown in front.

1

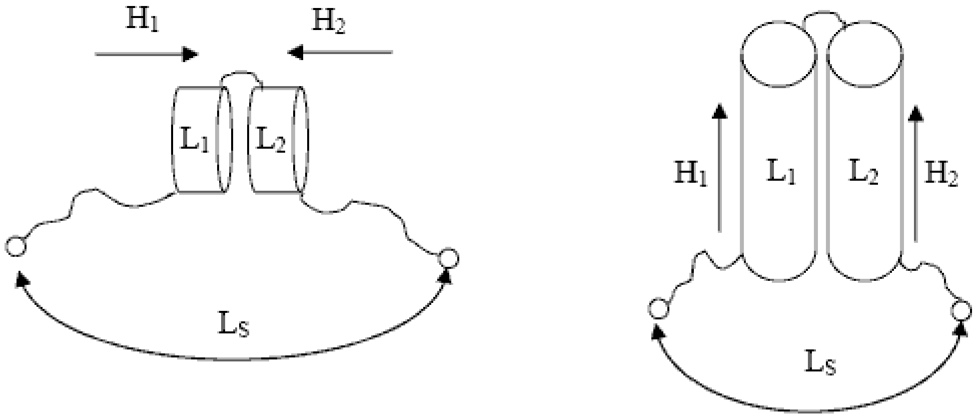

CONSTRUCTION: When constructing this arrangement there are many different options due to the various types of core which can be used for the coils:

1. Air-core

2. A ferromagnetic bar core

3. A ferromagnetic toroidal core

4. A transformer style ferromagnetic core.

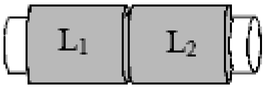

2

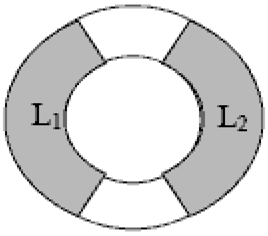

3

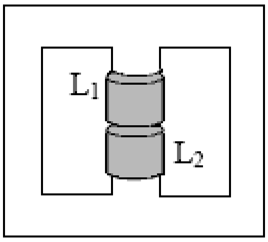

4

PROPERTIES: (tested many times with a variety of cores)

The value of the total inductance L_S does not change if you short one of the inductors L_1 or L_2

(This may have been tested for the first time by Mr. Tesla back in the 19th century).

APPLICATION TECHNIQUE:

This energy generation is based on the asymmetrical process:

1. Feed the total inductance L_S with a current I
2. Then short-circuit one of the inductors (say, L1)
3. Drain the energy from inductor L2 into a capacitor
4. After draining L2, then remove the short-circuit from L1, short-circuit L2 and then drain the energy from L1 into a capacitor

QUESTION: Is it possible, using this method, to get twice the energy amount due to the asymmetry of the process, and if not, then what is wrong?

AN ANSWER: We need to start winding coils and performing tests.

EXAMPLES OF COILS ACTUALLY CONSTRUCTED

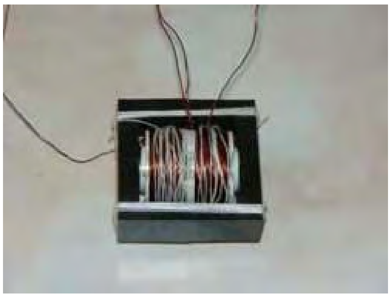

A coil was wound on a transformer ferromagnetic core (the size is not important) with permeability 2500 (not important) which was designed as a power-supply transformer. Each half-coil was 200 turns (not important), of 0.33 mm diameter wire (not important). The total inductance L_S is about 2 mH (not important).

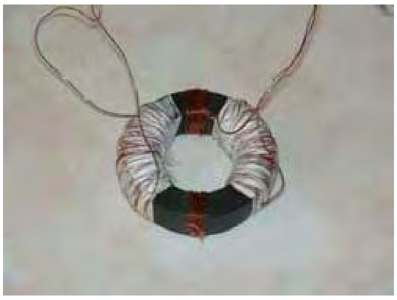

A coil was wound on a toroidal ferromagnetic core with permeability 1000 (not important). Each half-coil was 200 turns (not important), of 0.33 mm diameter wire (not important). The total inductance L_S is about 4 mH (not important).

An ordinary laminated iron core transformer intended for 50-60 Hz power supply use (size is not important) was wound with a coil placed on each of it's two halves. The total inductance L_S is about 100 mH (not important).

THE OBJECTIVE OF THE TESTS

To make tests to confirm the properties of the coils, and then make measurements of the L_S inductance both with coil L2 short-circuited and coil L2 not short-circuited, and then compare the results.

COMMENT: All of the tests can be done with just the toroidal coil as the other coils have been shown to have the same properties. You can repeat these tests and confirm this for yourself.

OPTION 1

These simple inductance measurements can be carried out with the help of an ordinary RLC (Resistance / Inductance / Capacitance) meter, such as the one shown here:

The measurements taken:

The total coil inductance L_S was measured without short-circuited coils, the figure was recorded. The L2 coil was then short-circuited and the inductance L_S measured again and the result recorded. Then, the results of the two measurements were compared.

The result: The inductance L_S was unchanged (to an accuracy of about a one percent).

OPTION 2

A special set-up was used, consisting of an analogue oscilloscope, a digital voltmeter and a signal generator, to measure a voltage on the inductance L_S without L2 being short-circuited and then with L2 short-circuited.

After the measurements were made, all of the results were compared.

Schematic of the set-up:

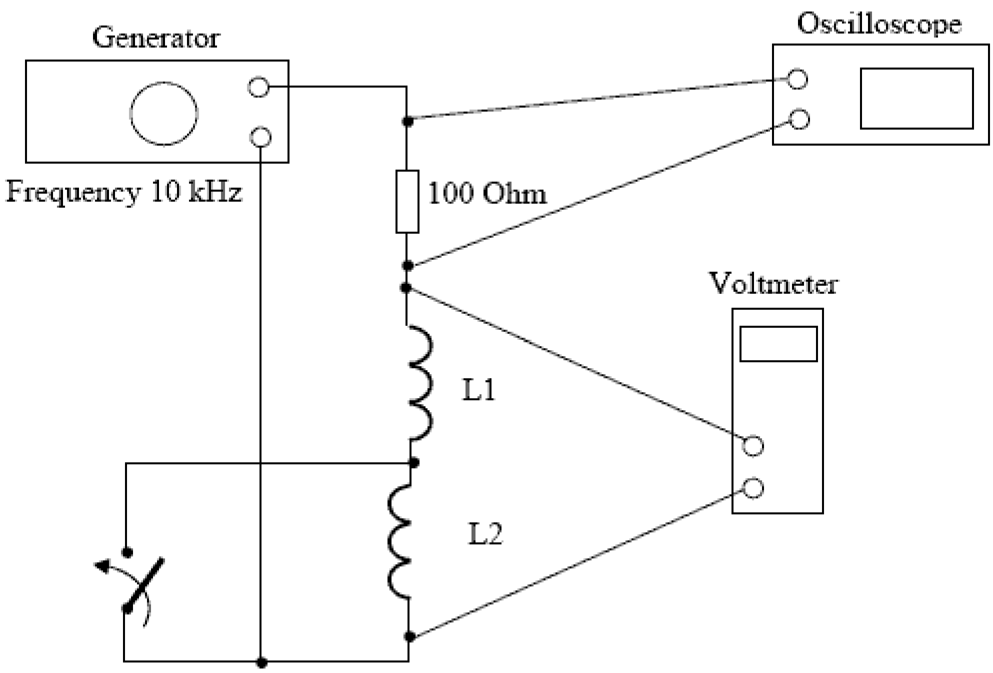

The order in which the measurements were taken

The voltage on the resistor was measured using the oscilloscope and the voltage on the inductor was measured using the voltmeter. Readings were taken before and after short-circuiting L2.

The result: The voltages remained unchanged (to an accuracy of about one percent).

Additional measurements

Before the above measurements were taken, the voltages across L1 and L2 were measured. The voltage on both halves was a half of the voltage on the total inductor L_S.

COMMENT: The frequency of about 10 kHz was chosen because the coil did not have parasitic resonances at this frequency or at low frequencies. All measurements were repeated using a coil with a ferromagnetic E-shaped transformer core. All of the results were the same.

OPTION 3

Capacitor recharging.

The objective was to match voltages on a capacitor, both before and after it being recharged by interaction with an inductor which could be connected into the circuit via a switch.

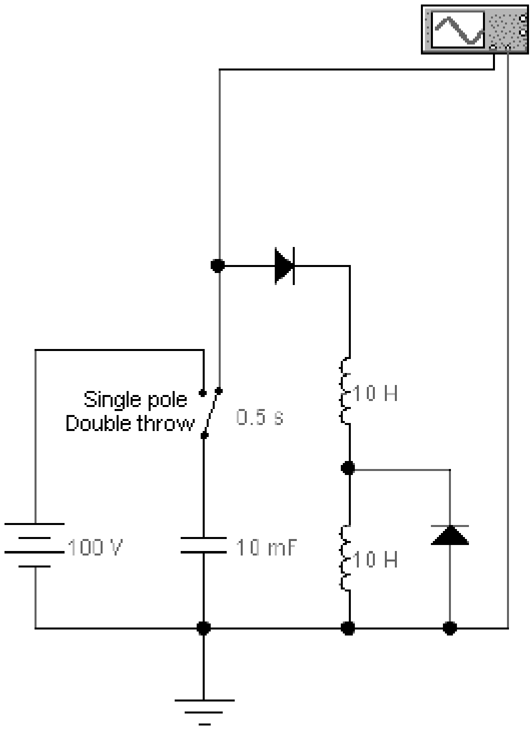

Single pole
Double throw
0.5 s
10 H
100 V
10 mF
10 H

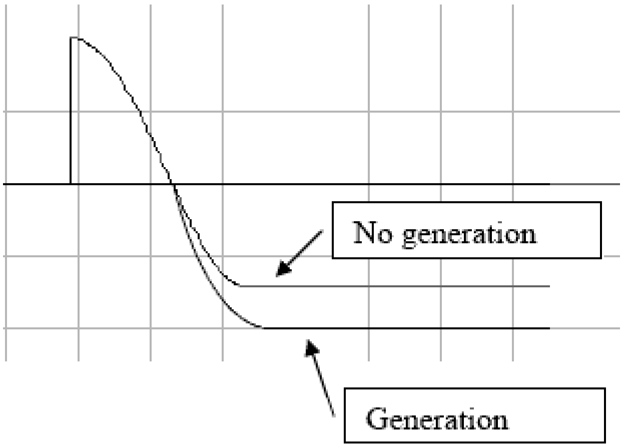

The experiment conditions

A capacitor is charged from a battery and is connected to the inductor through the first diode (included to give protection against oscillations). At the moment of feedback, half of the inductor is shunted by the second diode (due to it's polarity), while the inductance must remain unchanged. If after recharging the capacitor the capacitor voltage is the same (but with reversed polarity), then generation will have taken place (because a half of the energy remains in the shunted half of the inductor).

In theory, it is impossible, for an ordinary inductor consisting of two coils to do this.

The result :

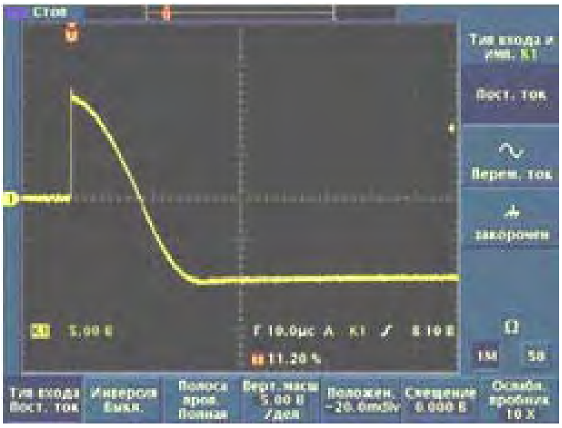

Стоп
Пост. ток
Перем. ток
закорочен
5.00 В
11.20 %
Ω
1M
50
Полоса проп. Полная
Верт. масш 5.00 В /дел
Смещение 0.000 В
Ослабл. пробника 10 X

The result confirms the prediction – the remaining energy is more than the capacitor gives to the coil (with an accuracy of 20%).

Test components: Capacitor 47 nano Farads, inductor LS is about 2 mH , Shotky silicon diodes BAT42, voltage used: 12 V.

THE RESULT VERIFICATION FOR OPTION 3

For verification of these results and in order to improve the accuracy, all measurements were repeated using alternative components.

Test components: Capacitor: 1.5 nano Farads; total inductance: 1.6 mH, germanium diodes: (Russian) D311, charging voltage: 5V.

The result: Confirmation of the previous measurements (a) shown below:

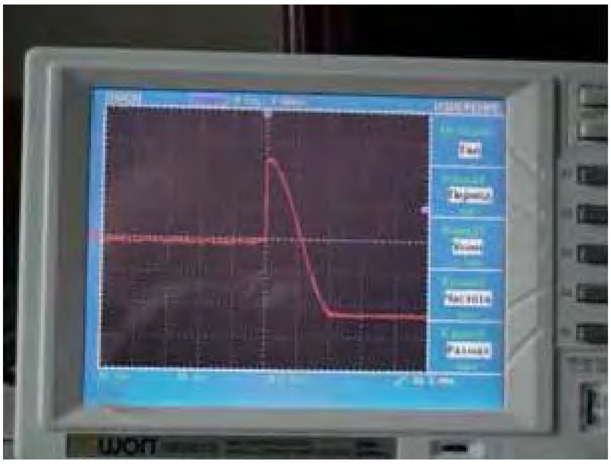

(a)

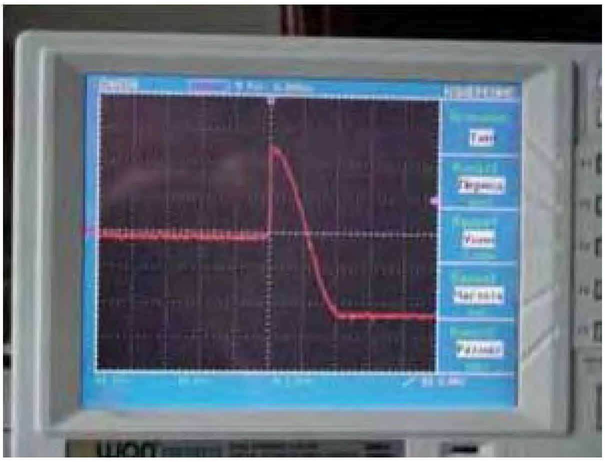

(b)

The recharging accuracy was improved to 10 percent. Also, a check measurement was made **without the second diode**. The result was essentially the same as the measurement which used the shunting diode. The missing 10 percent of the voltage can be explained as losses due to the spread capacitor's inductance and in it's resistance.

CONTINUED TESTING

The shunting diode was reversed and the test performed again:

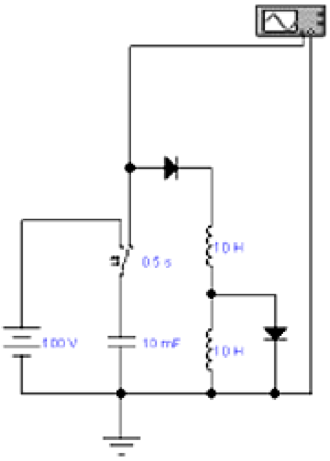

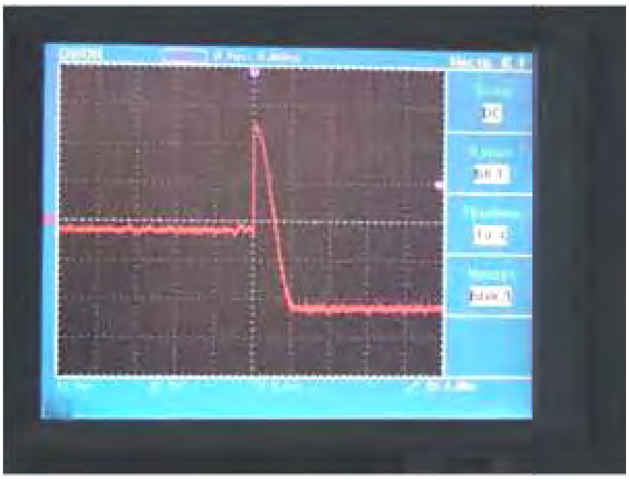

The result: It seems that the charge is spot on…

Further testing

An oscilloscope was connected to the coil instead of to the capacitor, in order to avoid influence of the first diode so the oscillations viewed were based on the inductance of the spread capacitors.

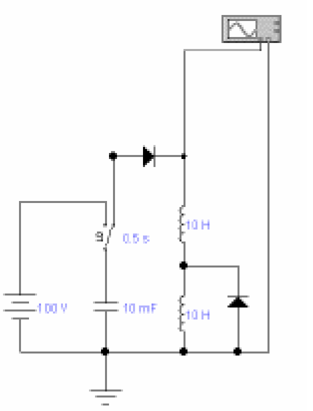

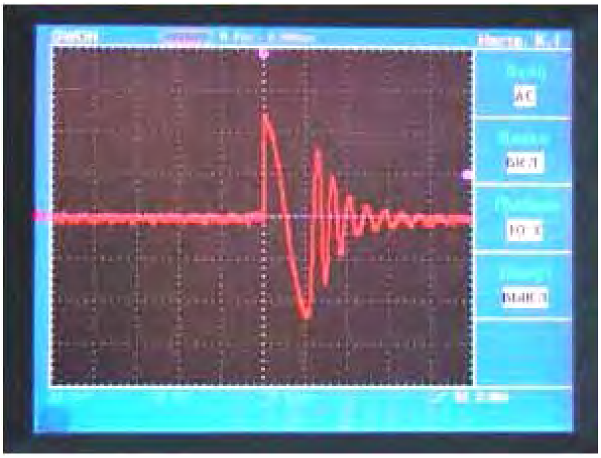

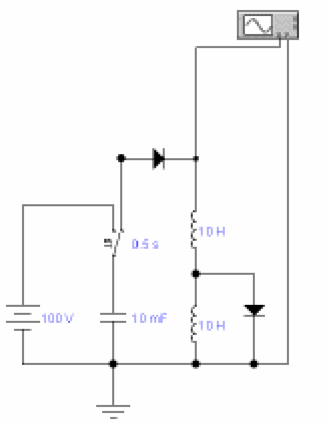

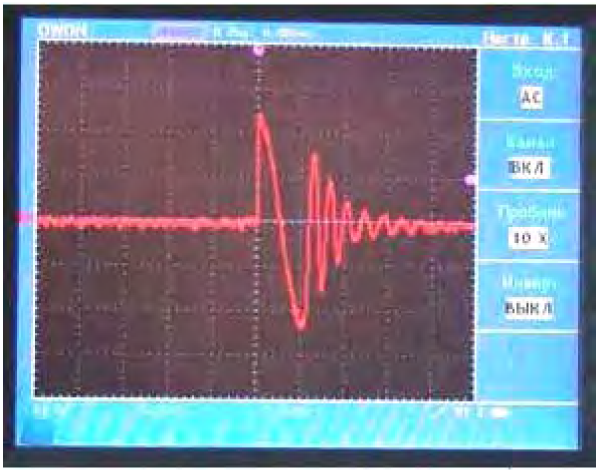

The result: The accuracy of capacitor recharging was improved to 5 percent (due to the removal of the influence of the first diode). After the main capacitor was switched off (by the diode), you can see oscillations caused by the spread capacitance of the inductors. Based on the frequency of the oscillations which were 4 to 5 times higher than that of the main capacitor, one can estimate the spread capacitance as being 16 to 25 times lower than the main capacitor.

Still further testing

Testing of the oscillation circuit shunting, with the two cases combined (and without the first diode):

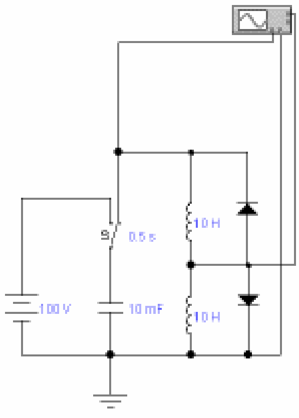

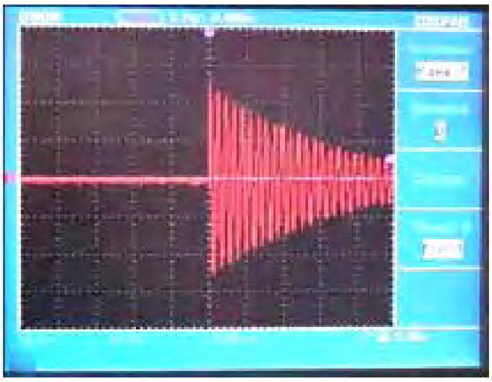

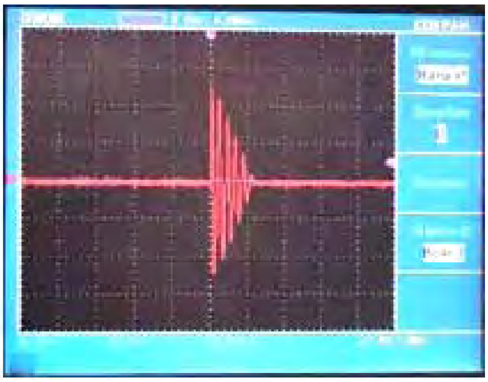

The result: A contour (oscillation circuit) is not destroyed, but it is shunted a lot. One can explain it by considering the moments when both diodes are conducting and so, shunt the circuit. As an addition, the voltage on the down diode is shown (the time scale is stretched). The negative voltage is close to maximum.

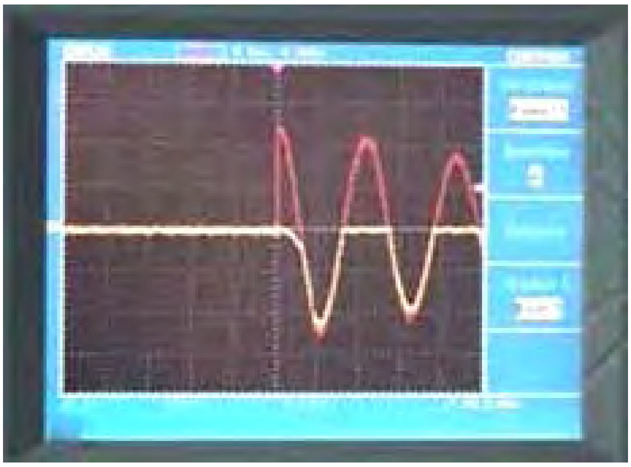

Still further testing

Charging a capacitor by shunting current in oscillation mode.

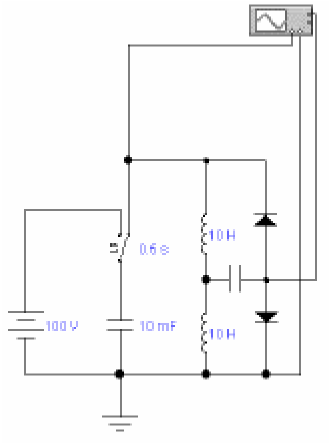

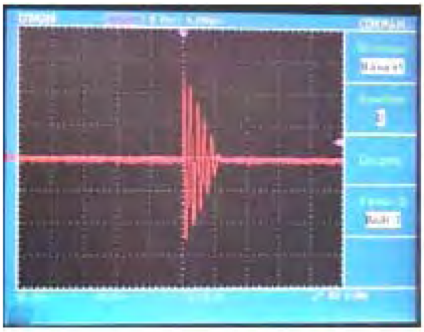

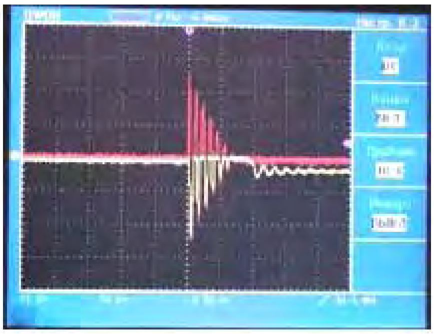

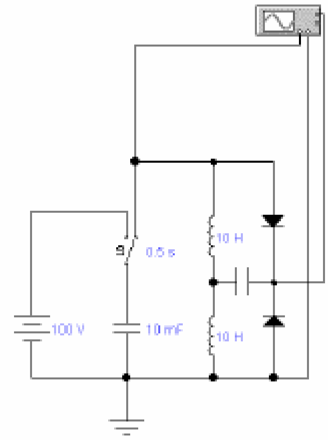

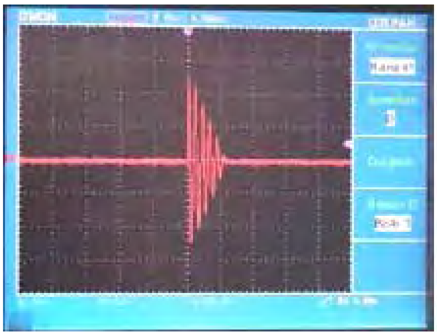

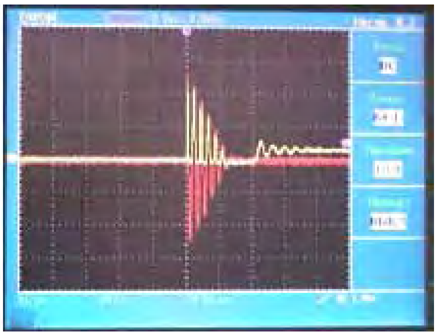

Conditions: The addition of a charging capacitor of 47 nano Farads.

The result: A capacitor is charging without shunting the circuit. The final voltage on it is 0.8 V, and rises an falls of the voltage depend on the value of the capacitor.

THE OVERALL RESULTS OF THE TESTS (OPTIONS 1, 2 and 3)

The symmetry of interaction in systems with electromagnetic field feedback (as with switched inductance) appears to be violated, and this implies that this arrangement could be used to generate energy.

COMMENT: You need to choose the load in order to get the maximum power output. **Very low, and very high loads, will send almost no energy to the load.**

ILLUSTRATION FOR SWITCHABLE INDUCTANCE

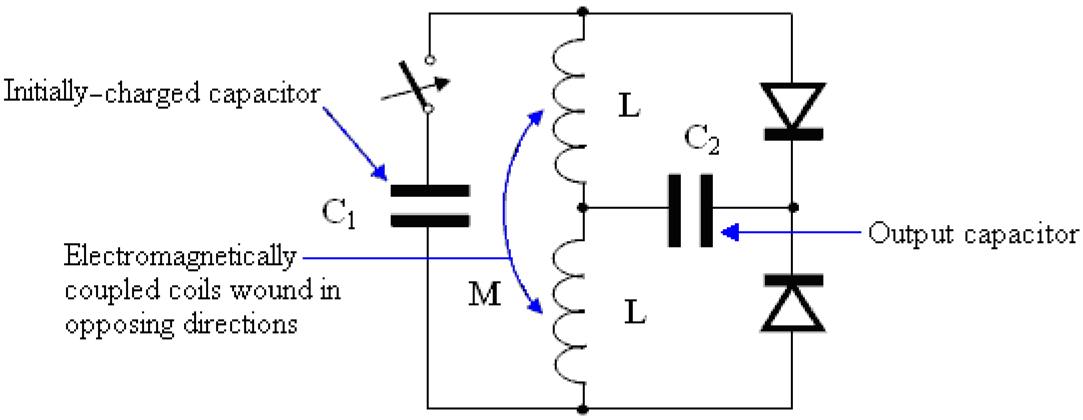
Initially-charged capacitor
L
C_2
C_1
Output capacitor
Electromagnetically coupled coils wound in opposing directions
M
L

EXPLANATION: The circuit has two kinds of currents: the main current and the shunting current.

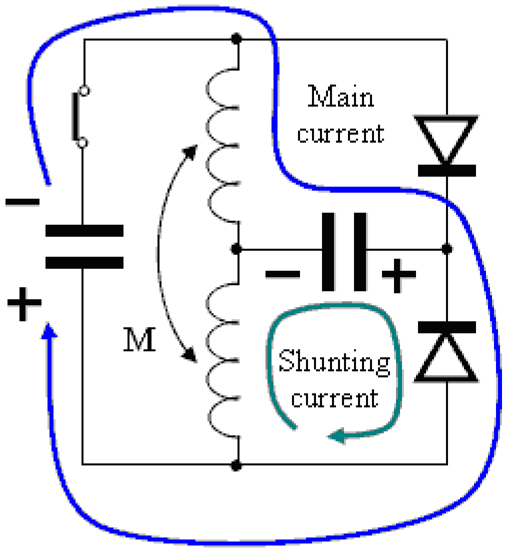
Main current
−
+
−
+
M
Shunting current

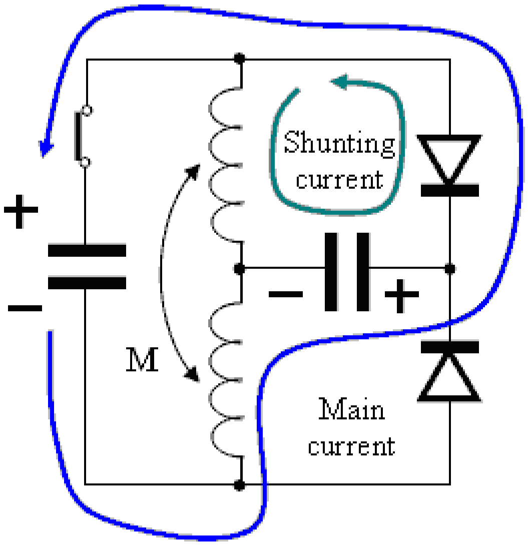
Shunting current
+
−
+
−
M
Main current

The main and the shunting currents run through the same output capacitor in one direction, if the output capacitor is discharged.

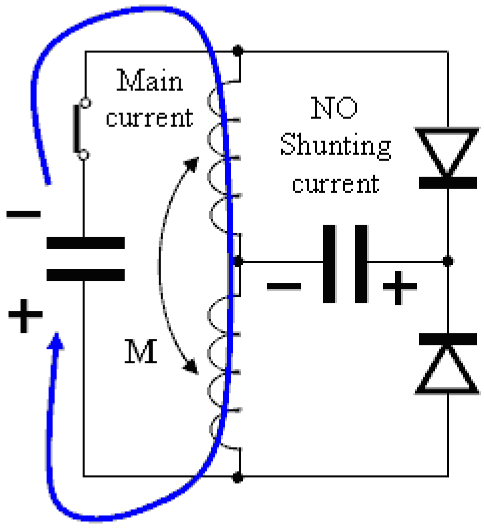
Main current
NO Shunting current
−
−
+
+
M

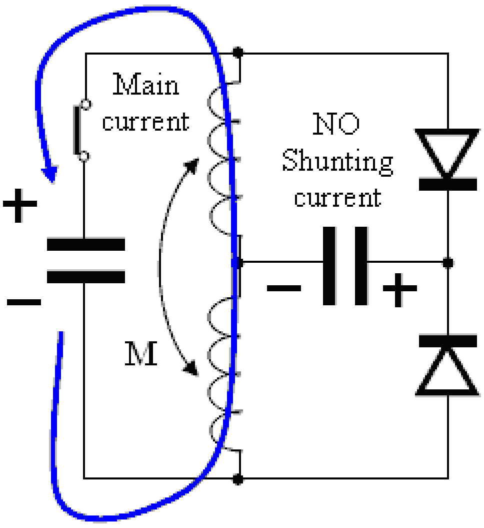
Main current
NO Shunting current
+
−
−
+
M

There is no shunting current, if the output capacitor is charged.

ILLUSTRATION FOR SWITCHABLE INDUCTANCE

From Don Smith

EXPLANATION As Don Smith said, two detector receivers were combined, and one FE device was constructed.

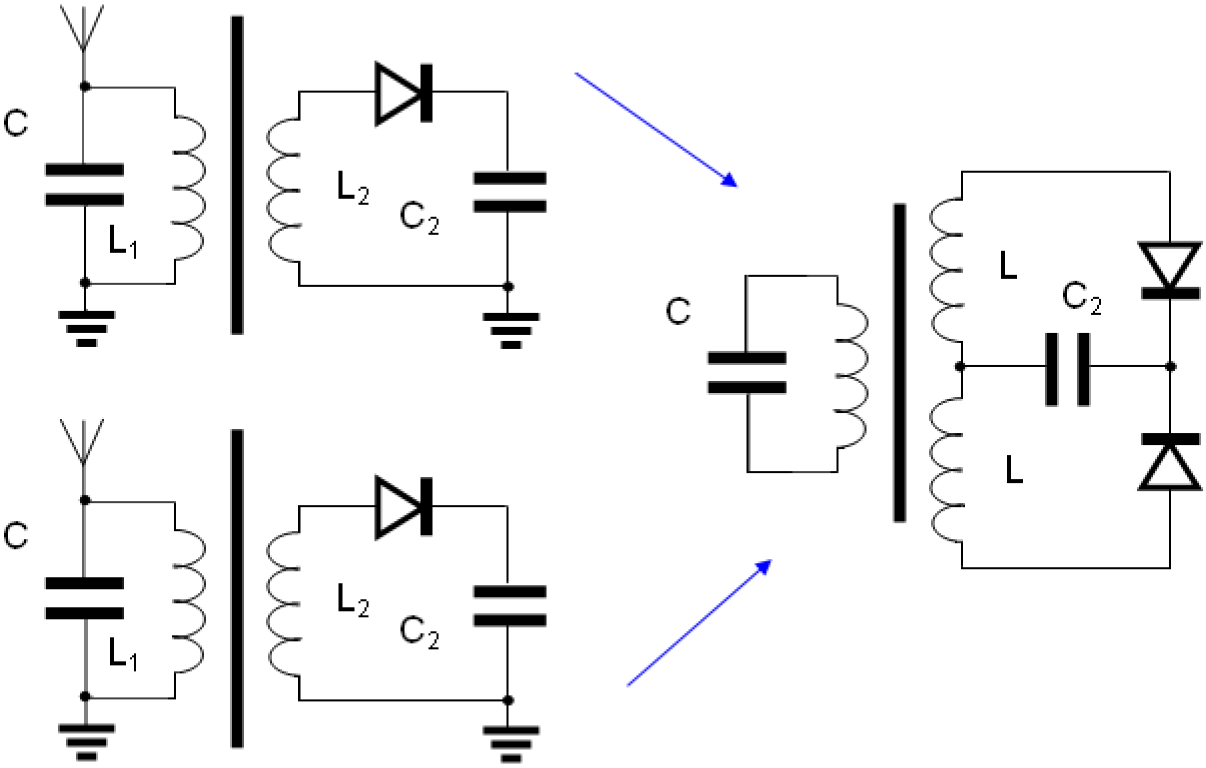

COMMENTS: Don Smith produced this explanation as a PDF file: www.free-energy-info.com/Smith.pdf

The resistance of the load must be chosen so as to get the maximum possible power in it.

The "board" does not contain an output circuit, because a couple of spark gaps and one step-down transformer can be used instead of diodes and a capacitor (this was pointed out before, so read the part which describes the suppression of back EMF).

MECHANICAL (INERTIAL) ANALOGUE OF SWITCHABLE INDUCTANCE

From Tariel Kapanadze

Version 1

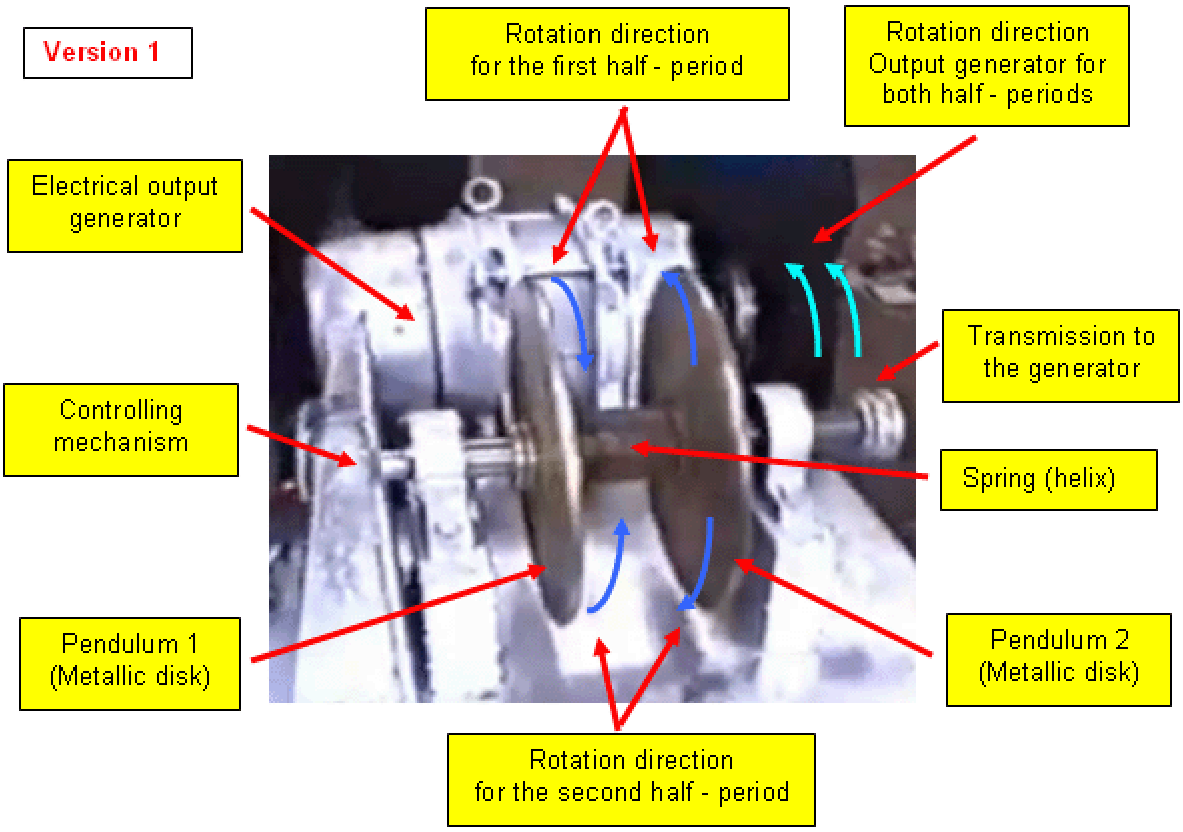
Rotation direction for the first half - period
Rotation direction Output generator for both half - periods
Electrical output generator
Transmission to the generator
Controlling mechanism
Spring (helix)
Pendulum 1 (Metallic disk)
Pendulum 2 (Metallic disk)
Rotation direction for the second half - period

Spring (helix)

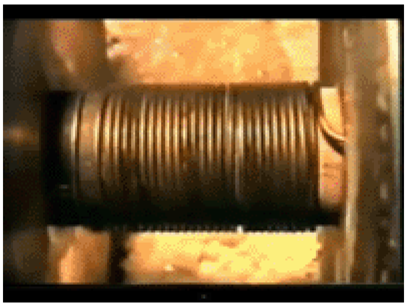

Version 2

Transmission

Controlling mechanism

EXPLANATION: When one pendulum is stopping the other is accelerating. The controlling mechanism connects the pendulums to the output generator one after the other and so maintains the oscillations.

CONNECTING EXTRA MASS TO A MECHANICAL OSCILLATOR

EXPLANATION: Mechanical energy can be stored in any spring by compressing it or stretching it (1). It corresponds to two positions in a mechanical oscillator (2), when only potential energy takes place in an oscillating process

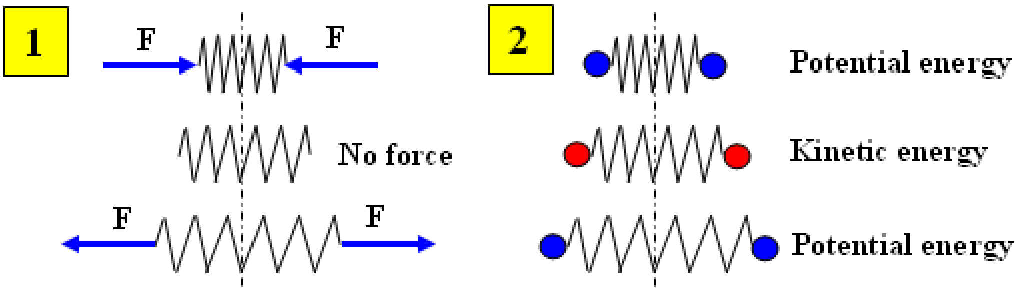
1
F
F
No force
F
F
2
Potential energy
Kinetic energy
Potential energy

EXPLANATION: If extra mass is connecting periodically to one side or the other, of a mechanical oscillator, it will be shifting without any energy loss during the oscillation process

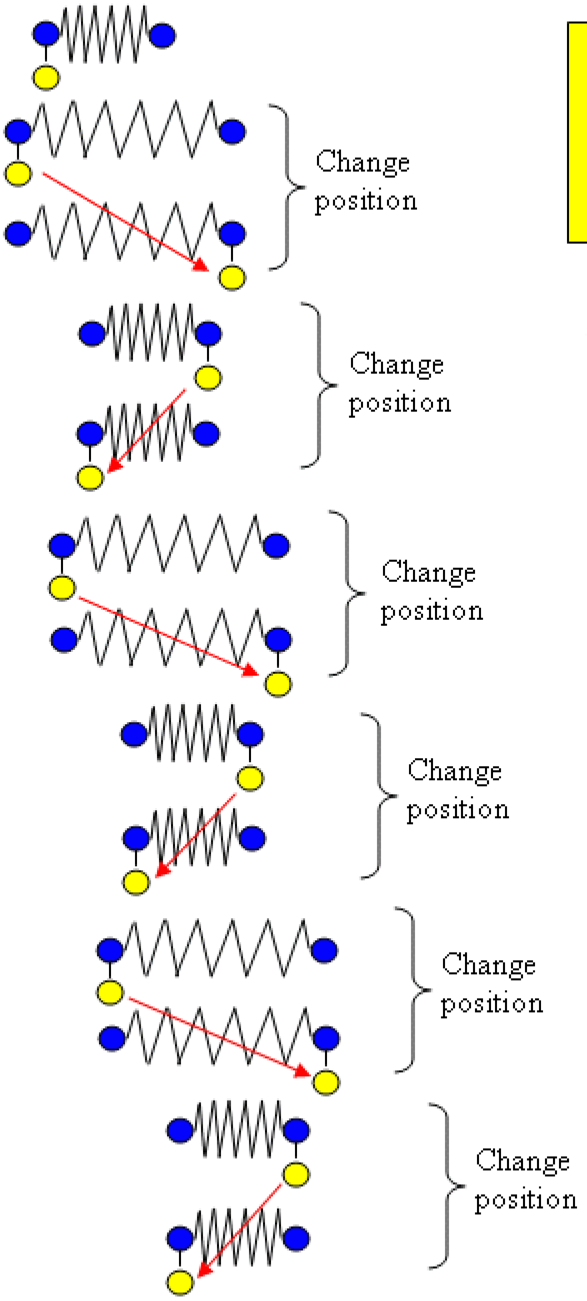
Change position
Change position
Change position
Change position
Change position
Change position

And so on …

COMMENT There is no need to transfer extra mass from one position to another in an actual device, because extra mass is a disc and masses in a mechanical oscillator are discs also. You only have to connect them in a variety of ways.

COMMENT A grounded Tesla coil is an analogue of a spring with one side fixed

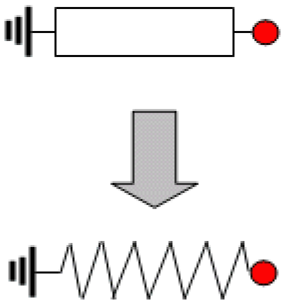

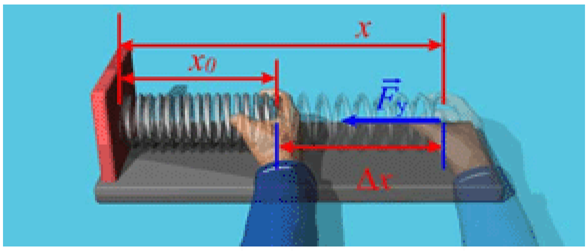

$E_p = k(dx)^2/2$ – Potential energy of deformed spring (helix)

THE PRINCIPLE OF AMPLIFICATION OF MECHANICAL ENERGY

Explanation: The principle is based on an asymmetrical flywheel (1) consisting of a small mass and a large mass. These masses are balanced across the centre of rotation, that is, are located at a distance proportional to their weights, from the center of rotation. This helps to avoid vibration when they are rotating (the same principle used when balancing a car wheel).

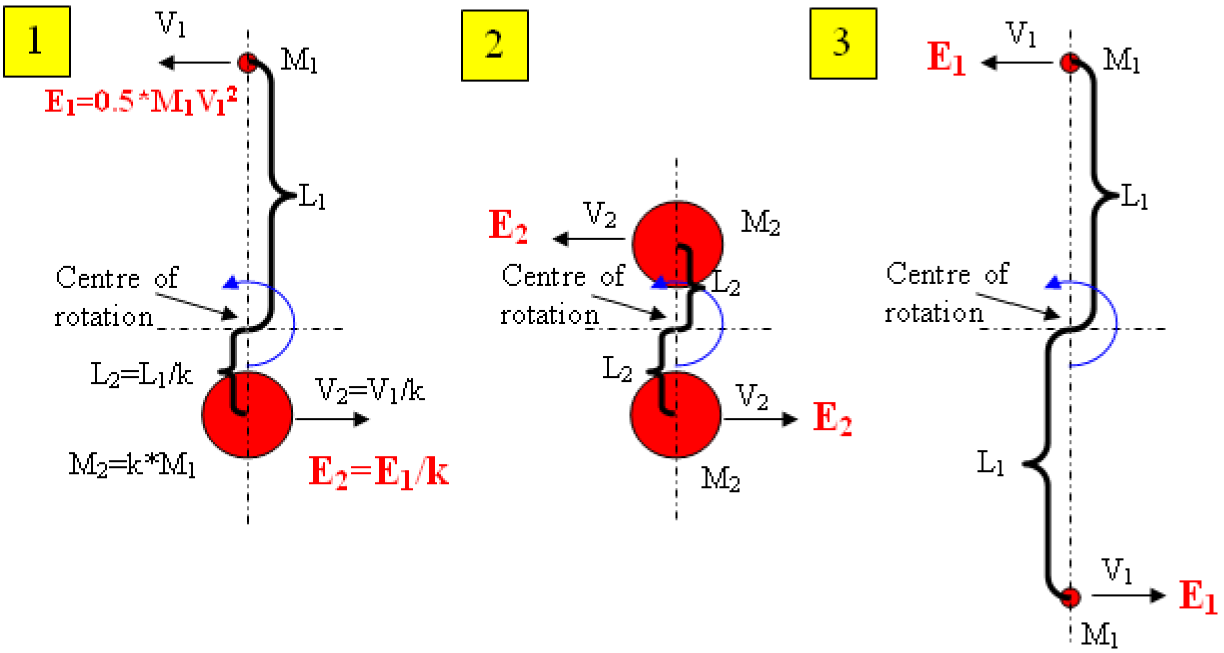
1
V_1
M_1
$E_1=0.5*M_1V_1^2$
L_1
Centre of rotation
$L_2=L_1/k$
$V_2=V_1/k$
$M_2=k*M_1$
$E_2=E_1/k$
2
E_2
V_2
M_2
Centre of rotation
L_2
L_2
V_2
E_2
M_2
3
E_1
V_1
M_1
L_1
Centre of rotation
L_1
V_1
E_1
M_1

The inertial moment of such a flywheel (1) is analogous to the inertial moments of flywheels (2) and (3), consisting only of large or small masses. However, from the point of view of kinetic energy, all of these examples, (1), (2) and (3) are different. This is because the kinetic energy of every mass depends on the direction and speed at which it moves (if is released during rotation). The highest common kinetic energy is in the masses of flywheel (3), as less energy is contained in flywheel (1) and the smallest kinetic energy is in flywheel (2). In order to get an increase in energy one needs to achieve a set-up which is based on a spring (for energy transformation from kinetic energy to potential energy and back again) and a lever of Archimedes (for changing the point where the force is applied).

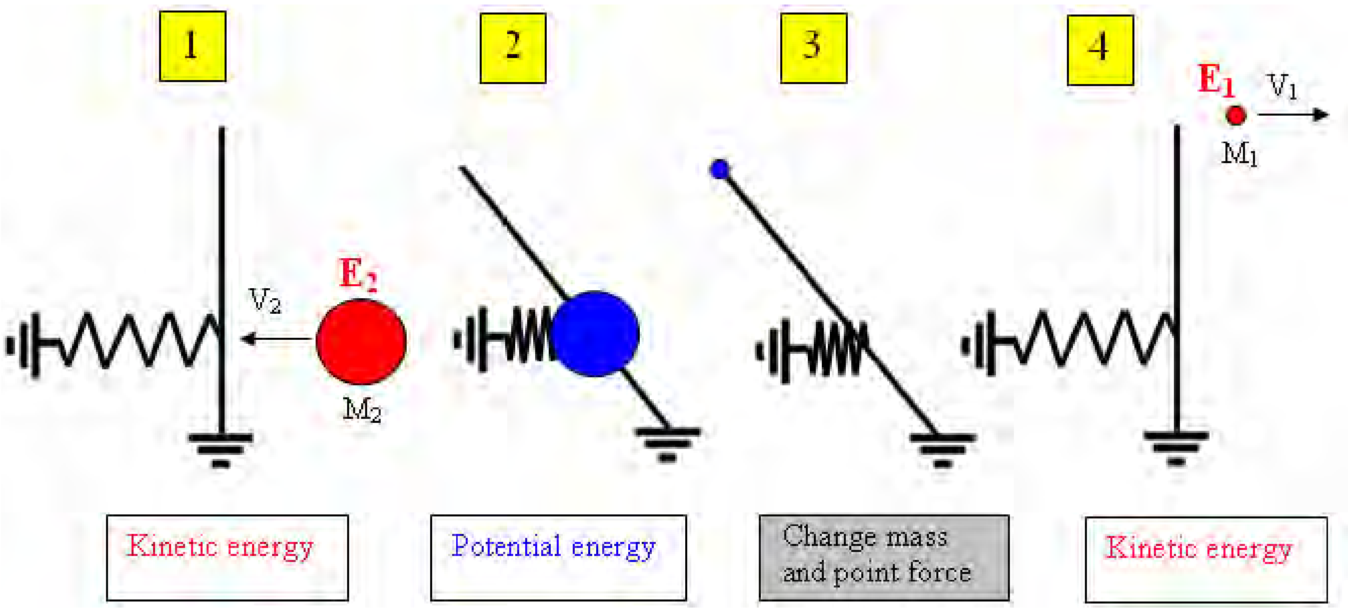
1
2
3
4
E_2
V_2
M_2
E_1
V_1
M_1
Kinetic energy
Potential energy
Change mass and point force
Kinetic energy

Comments:

1. The simplified schematic diagrams shown here are for explanation purposes only.
2. In an actual device, you can use a spring in rotation mode (as Tariel Kapanadze did).
3. You can use disks and rings as flywheel masses (as Tariel Kapanadze did).
4. Altering one mass to another is actually achieved by connecting them in various ways.

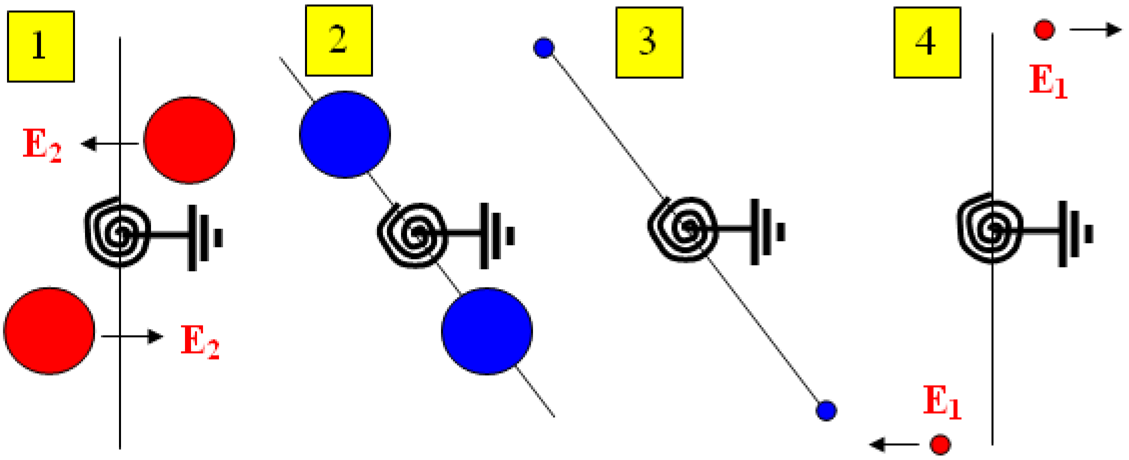
1
2
3
4
E$_2$
E$_2$
E$_1$
E$_1$

Comment: Any asymmetrical mechanical oscillator behaves as indicated above, when the potential energy of a compressed spring is transformed to the kinetic energy of moving masses.

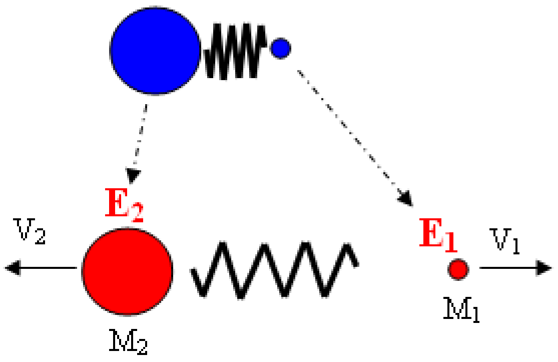
E$_2$
V$_2$
M$_2$
E$_1$
V$_1$
M$_1$

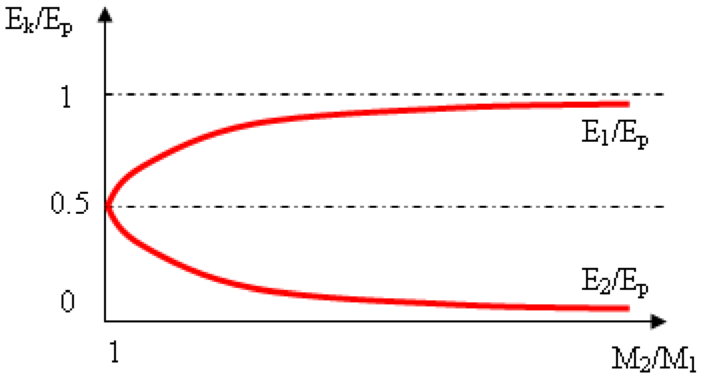

The potential energy of the spring is distributed unequally between the small and large masses. A small mass acquires more energy relative to it's size than a large mass does. The sum of the kinetic energies of both masses is equal to the potential energy of the spring.

Comment: This is based on Tesla's asymmetrical schematic:

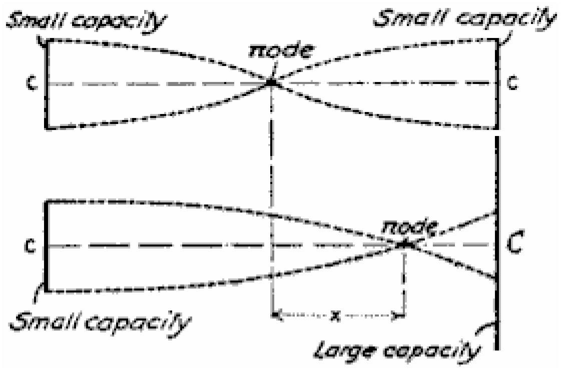
Small capacity
Small capacity
node
C
C
C
C
node
x
Small capacity
Large capacity

FLYWHEEL – A HIDDEN FORM OF ENERGY

(Clarifications on mechanical energy amplification)

EXPLANATION: If you don't want to lose mechanical energy when doing work, then this work must be done by an imaging force. This force is absent in an inertial coordinate system, but it is present in a non-inertial coordinate system. When in a rotational coordinate system this force is called 'centrifugal' force.

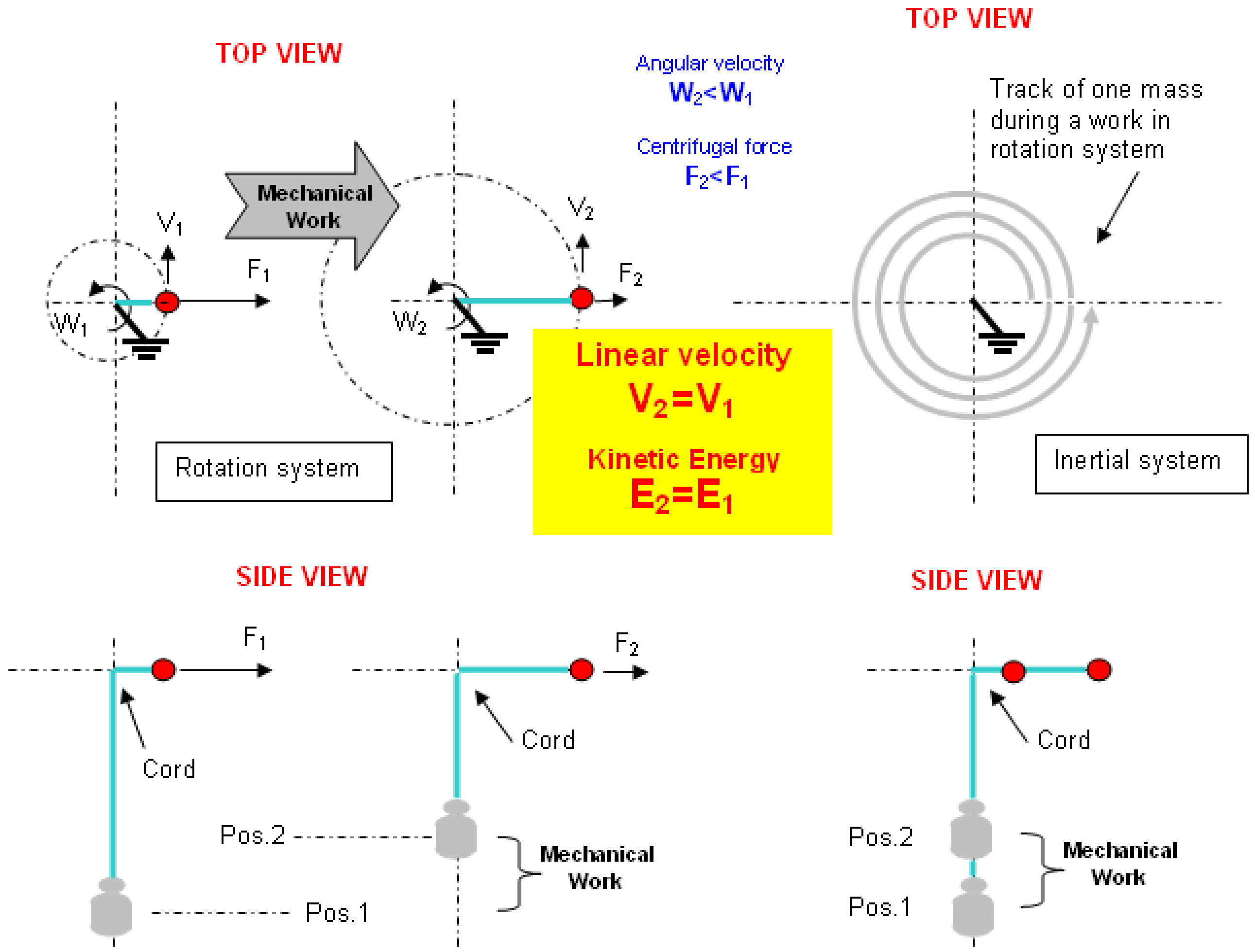
TOP VIEW
Mechanical Work
V1
F1
W1
V2
F2
W2
Rotation system
Angular velocity
W2<W1
Centrifugal force
F2<F1
Linear velocity
V2=V1
Kinetic Energy
E2=E1
TOP VIEW
Track of one mass during a work in rotation system
Inertial system
SIDE VIEW
F1
Cord
Pos.2
Pos.1
F2
Cord
Mechanical Work
SIDE VIEW
Cord
Pos.2
Pos.1
Mechanical Work

COMMENT: After the work is done, the centrifugal force is low and if you want to continue producing mechanical work, you have to use **the other** coordinate system where centrifugal force is high again. This is possible because linear velocity does not change. You have to provide **the other** support point only (and a cord) in order to produce mechanical energy again.

COMMENT: If you want to make this mechanical work continuous, then the end of the first track must also be the beginning of the second track. You have to change coordinate system periodically.

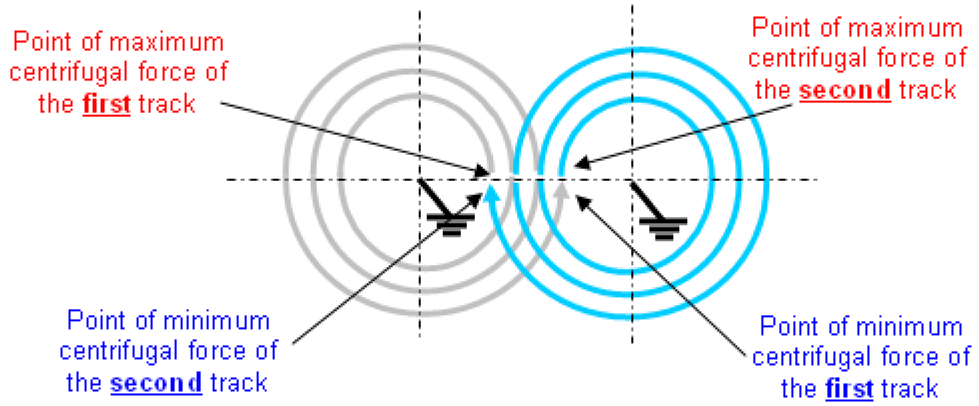
Point of maximum centrifugal force of the first track
Point of maximum centrifugal force of the second track
Point of minimum centrifugal force of the second track
Point of minimum centrifugal force of the first track

COMMENT: In a real situation, you have to compensate for energy loss due to friction and so a part of the excess energy must be used to maintain the process.

ILLUSTRATION FOR SWITCHABLE INDUCTANCE

From Alfred Hubbard

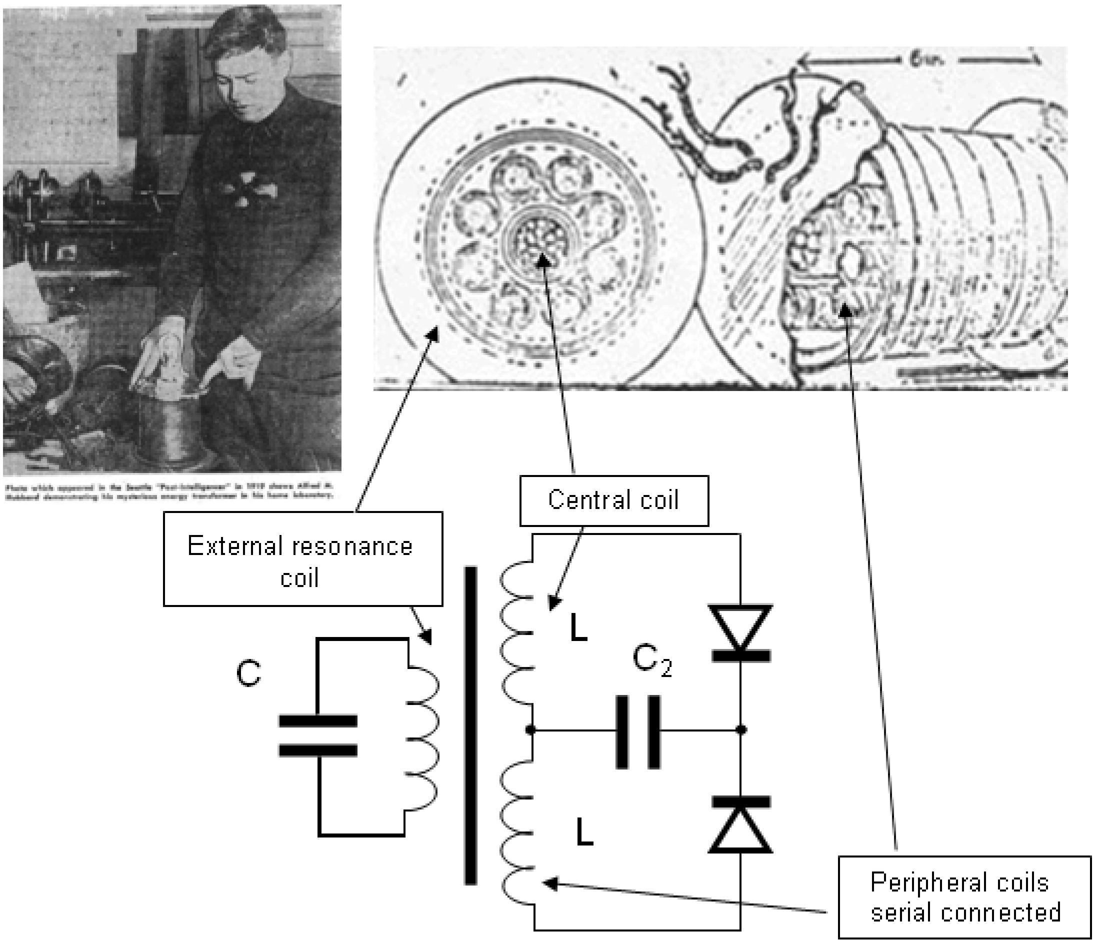
6in.
External resonance coil
Central coil
Peripheral coils serial connected
C
L
C2
L

Photo which appeared in the Seattle "Post-Intelligencer" in 1919 shows Alfred M. Hubbard demonstrating his mysterious energy transformer in his home laboratory.

EXPLANATION: The center coil and all of the peripheral coils can “grasp” the same flux coming from the resonance coil. All other details are the same as in Smith’s version.

COMMENTS: In other words, you can use rods as the coil core, instead of a closed ferromagnetic core.

But, this is not the only option in Hubbard’s device. He may have had another one, based on a different principle, perhaps the principle of energy amplification in an LC circuit as described earlier, but with switchable inductance being used.

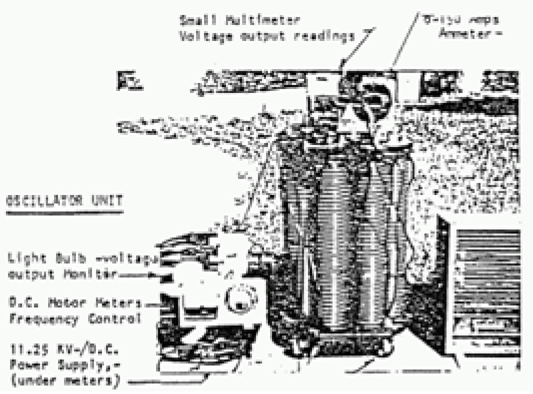
Small Multimeter
Voltage output readings
0-150 amps
Ammeter
OSCILLATOR UNIT
Light Bulb -voltage
output Monitor
D.C. Motor Meters
Frequency Control
11.25 KV-/D.C.
Power Supply,-
(under meters)

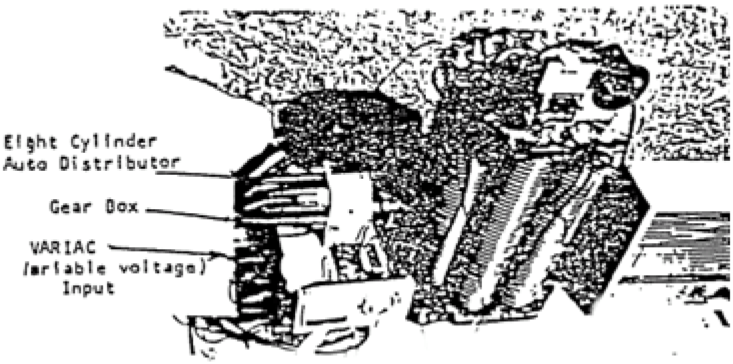
Eight Cylinder
Auto Distributor
Gear Box
VARIAC
(Variable voltage)
Input

MODERN OPTIONS?

In switchable inductance

Version 1

A coil has more inductance when some of it's parts are short-circuited:

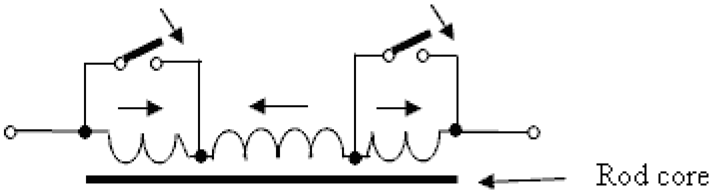

EXPLANATION: The central section of the coil and it's two end sections are wound in opposite directions.

COMMENT: The coil shown in the picture above has **twice the inductance**, when it's end sections are short-circuited (measurements made with the Chinese-built RLC test meter shown here):

Hubbard's version

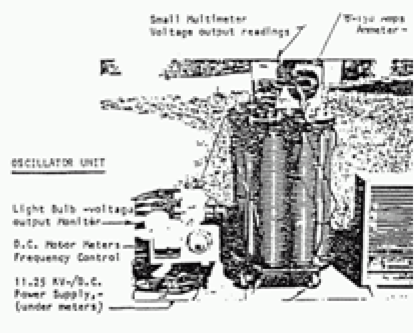
Small Multimeter
Voltage output readings
Amps
Ammeter
OSCILLATOR UNIT
Light Bulb -voltage output Monitor
D.C. Motor Meters Frequency Control
11.25 KV-/D.C. Power Supply (under meters)

Version 2

By Don Smith

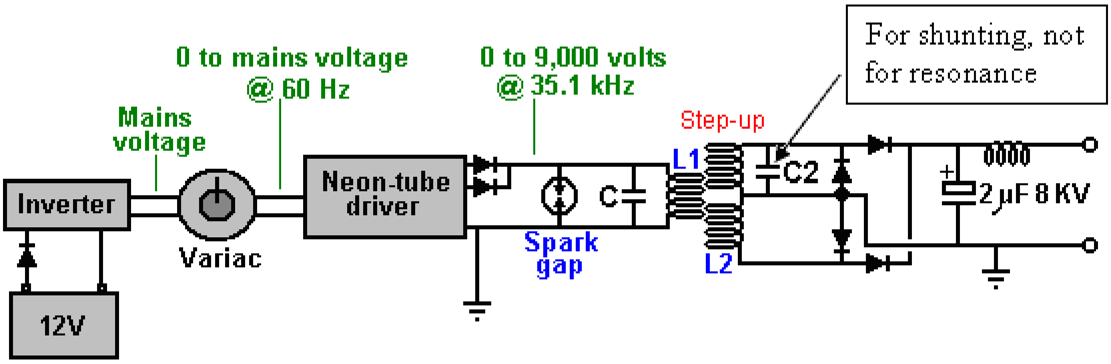
0 to mains voltage
@ 60 Hz
0 to 9,000 volts
@ 35.1 kHz
Mains voltage
Step-up
Inverter
Variac
Neon-tube driver
C
Spark gap
L1
L2
C2
2 µF 8 KV
12V

For shunting, not for resonance

But, this looks like resonance in an asymmetrical transformer **?????**

Version 3

By Tariel Kapanadze

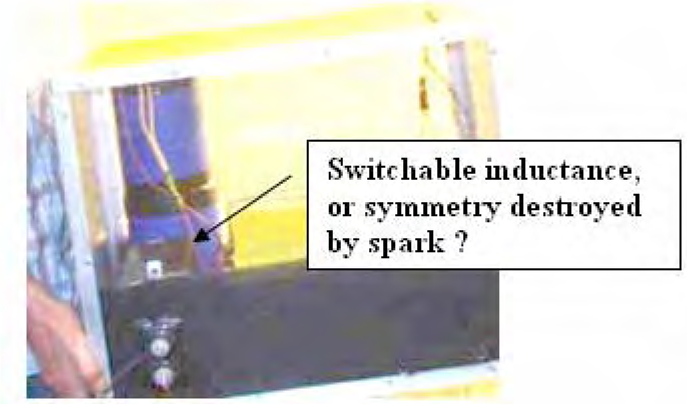

Switchable inductance, or symmetry destroyed by spark ?

No description …**???**

Read on for further details….

THE BASIS OF SWITCHABLE INDUCTANCES

(Tesla patent)

(No Model.) 2 Sheets—Sheet 1.

N. TESLA.

REGULATOR FOR DYNAMO ELECTRIC MACHINES.

No. 336,961. Patented Mar. 2, 1886.

Fig. 1.

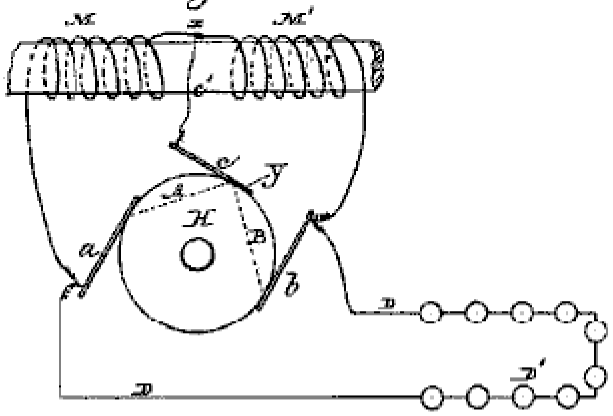
M
x
M'
c'
c
y
A
H
B
a
b
D
D'
D

Fig. 2.

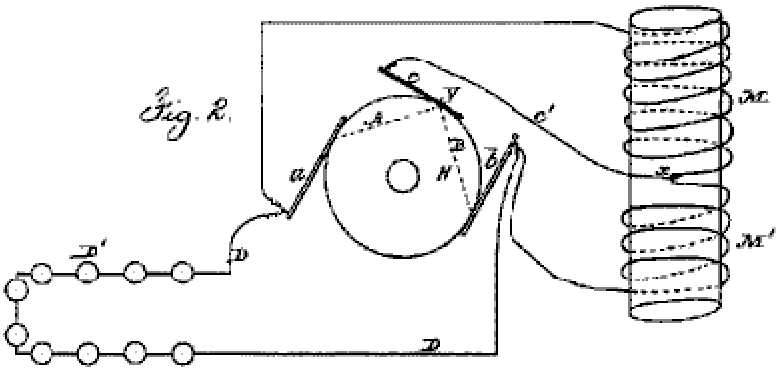
c
y
c'
M
A
B
H
a
b
x
M'
D'
D
D

Fig. 3.

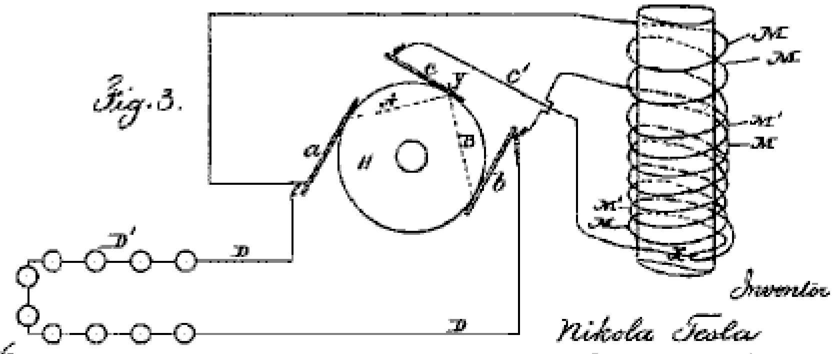
M
M
c'
c
y
A
B
H
a
b
M'
M
M'
M
x
D'
D
D

Witnesses
Chas. H. Smith
Geo. T. Pinckney

Inventor
Nikola Tesla
per Lemuel W. Serrell
atty

SECRET 3

THE ASYMMETRICAL TRANSFORMER

With a magnetic field feedback loop (evolution of the 2nd secret)

LENZ LAW IS VIOLATED IN AN ASYMMETRICAL TRANSFORMER

(Therefore it is not possible to use it as an ordinary transformer)

An asymmetrical transformer can have two coils: L_2 and L_S. Coil L_2 is wound on one side of the toroidal core while L_S is wound so that it encloses both the toroid and the coil L_2 as shown here:

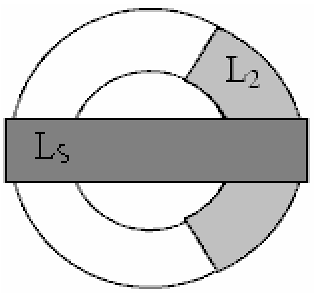

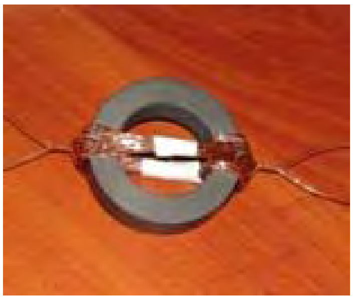

Optionally, this arrangement can be implemented with a wide range of styles of transformer core:

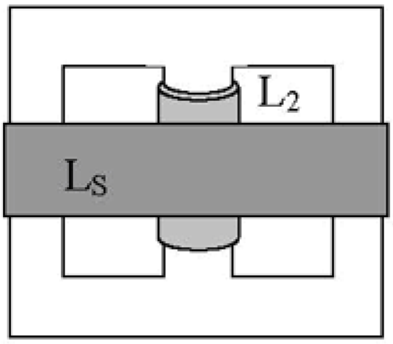

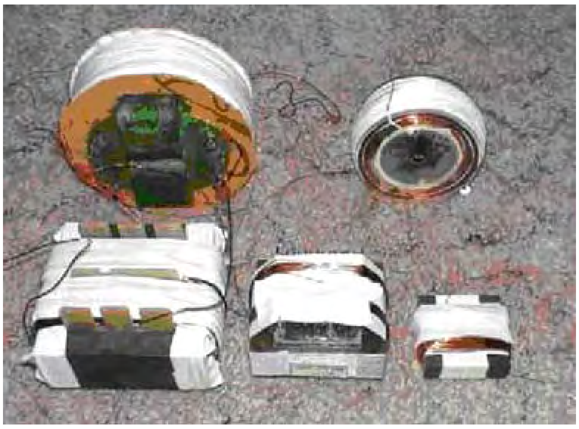

One option is to use the above (switched inductor) arrangement and add one more coil:

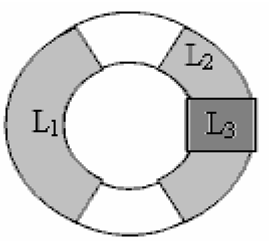

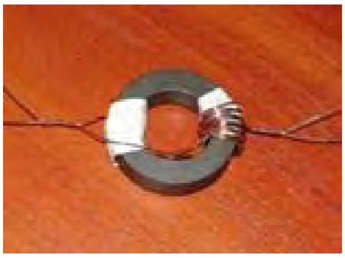

Now that you understand the operational principles of this system, you can use any configuration which you need. For example:

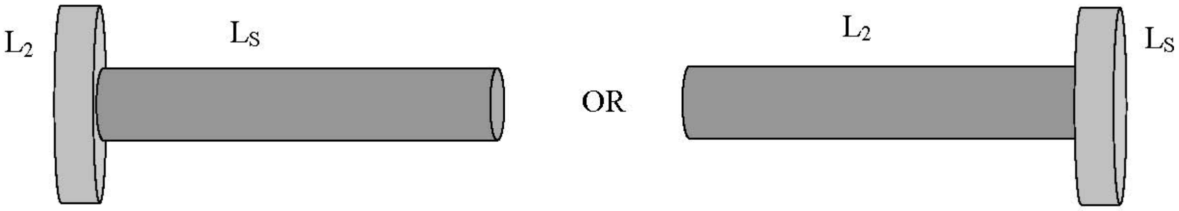

ILLUSTRATION FOR AN ASYMMETRICAL TRANSFORMER OF SOME KIND

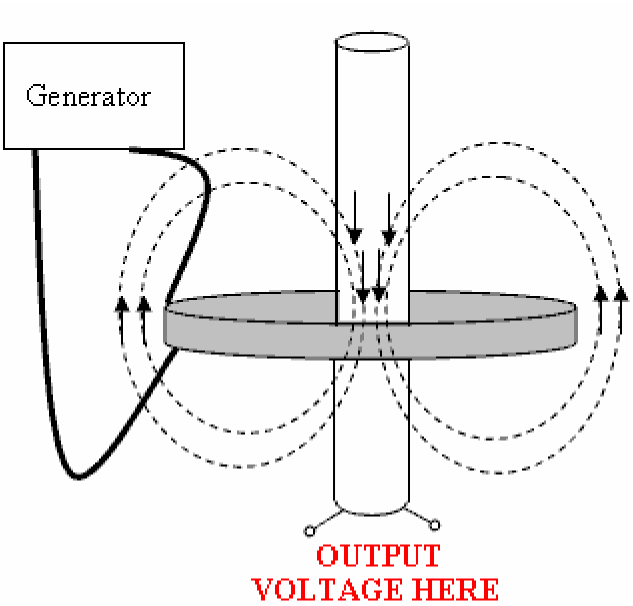

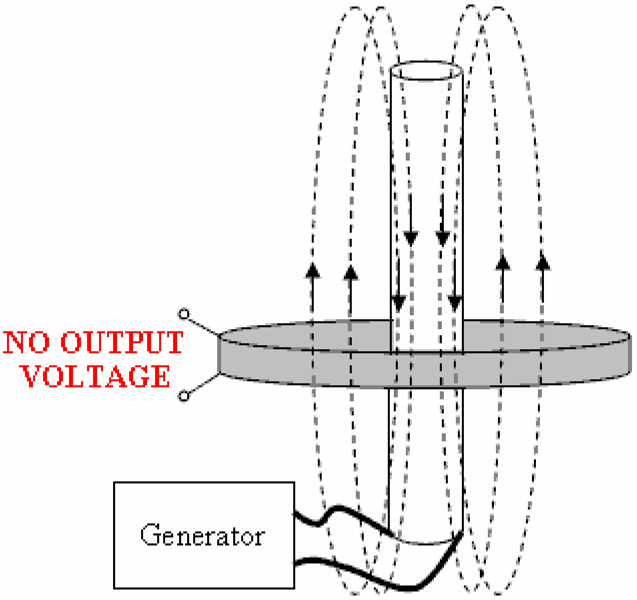

THE MECHANICAL EQUIVALENT OF AN ASYMMETRICAL TRANSFORMER

This example shows an ordinary transformer, wound on an E-core plus an external excitation magnet:

The magnet orientation may be different

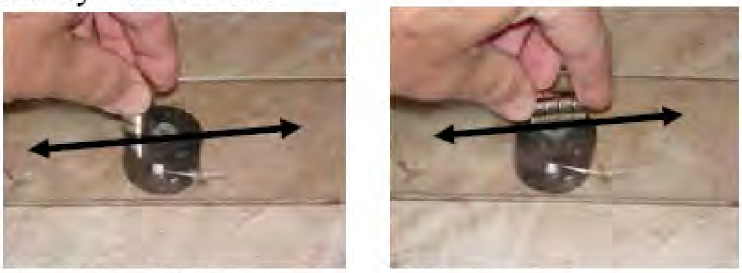

The magnet orientation may be different

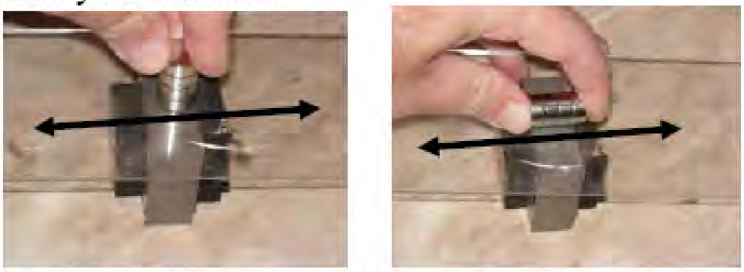

In other words: L2 is still used, but instead of L_S the exciting magnet is used.

The result:

1. The voltage developed across coil L2 depends on the number of turns in L2, but **the short-circuit current through L2 does NOT depend on the number of turns in coil L2.**
2. You need to choose the load connected to L2 in order to get the maximum power output. **Very low, and very high loads, will give almost no power output**.

RESONANCE IN AN ASYMMETRICAL TRANSFORMER

The first coil is used as a transmitter of energy, and the second coil as a receiver of energy.

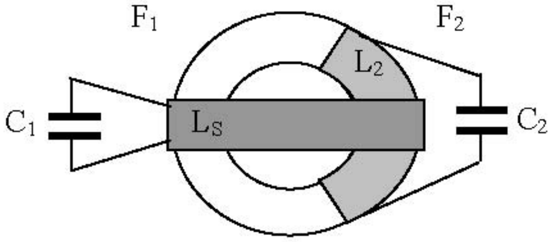
F1
F2
L2
C1
LS
C2

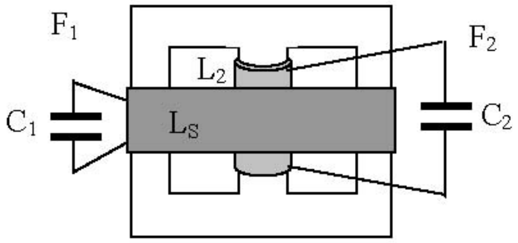
F1
F2
L2
C1
LS
C2

It is very like radio broadcasting, where the receiver is located far away from the transmitter, and has no feedback. The first coil works in **parallel resonance** and the second coil in **serial resonance** (although the two schematic diagrams look alike).

EQUIVALENT SCHEMATICS

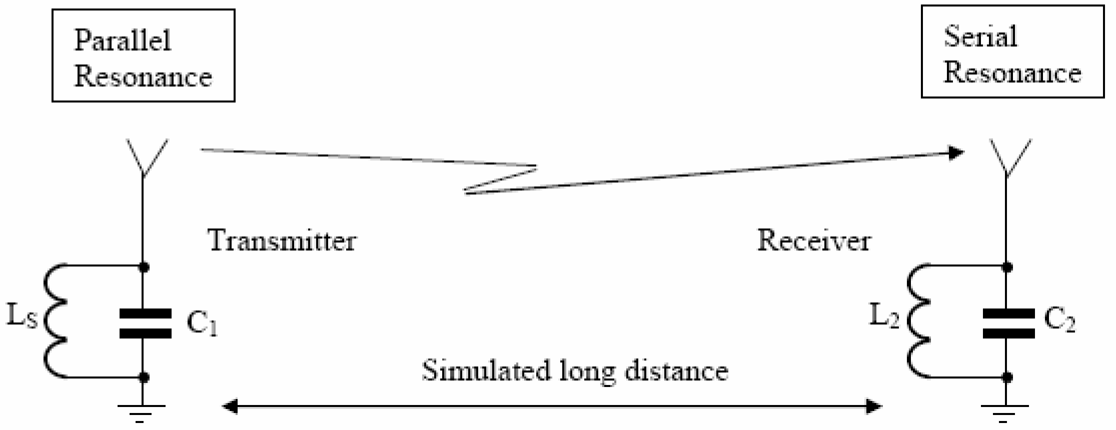
Parallel Resonance
Serial Resonance
Transmitter
Receiver
LS
C1
L2
C2
Simulated long distance

FOR TRANSMITTER

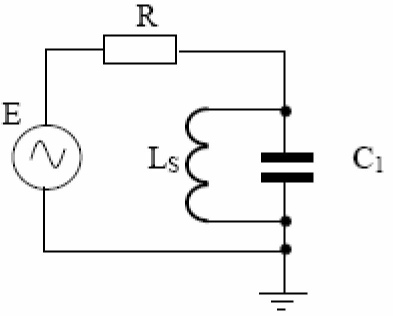

FOR RECEIVER

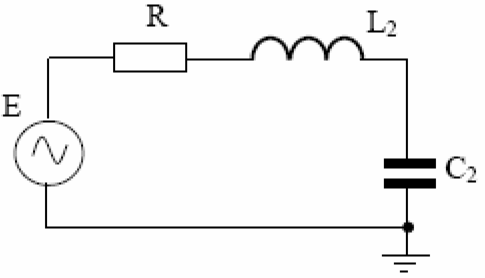

CONSEQUENTLY: You can get much more voltage on L_2 than on L_S

An experiment:

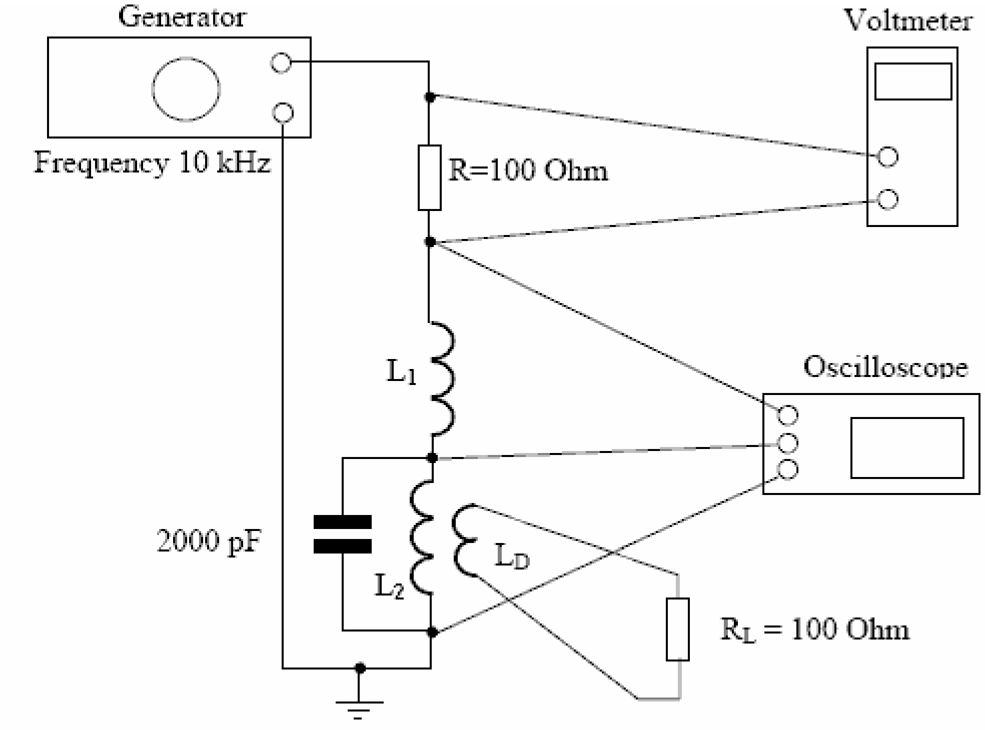
Generator
Voltmeter
Frequency 10 kHz
R=100 Ohm
L1
Oscilloscope
2000 pF
L2
LD
RL = 100 Ohm

Conditions:

The resonance frequency is about 10 kHz. The total inductance L_S is 2.2 mH, the L_2 inductance (same as the L1 inductance) is 100 mH, the ratio L_S:L_2 is 1:45 with an E-shape core, permeability is 2500.

The result:

At the resonance frequency, there can be a voltage which is **50 times more** on any parts (L1 or L2) matched with the total coil LS, and voltage changes on R are **no more than 15 percent.**

The phase shift in voltage is about 90 degrees between LS and L2.

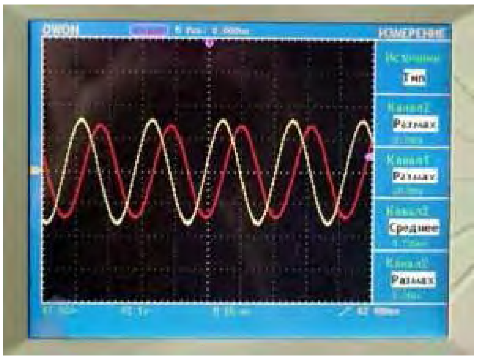

(The amplitudes were equalised)

Further

An additional step-down coil L_D was wound around L_2, turns ratio 50:1 (matched with L_2), and the load resistor R_L = 100 Ohms was connected to it.

The result

Changes in current consumption (estimated by measuring the voltage across R) are no more 15 percent.

MODERN OPTIONS IN USAGE OF AN
Asymmetrical transformer

By Don Smith

The schematic is like this:

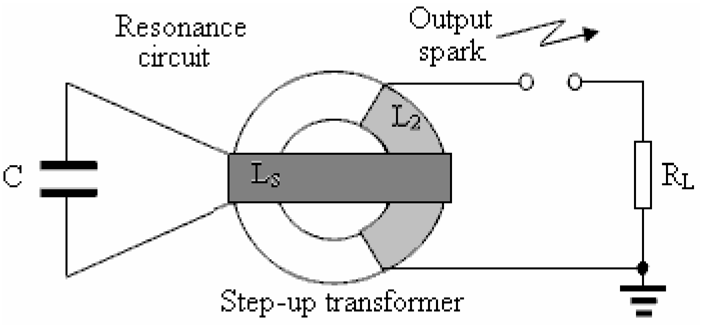

Resonance circuit
Output spark
L2
C
Ls
RL
Step-up transformer

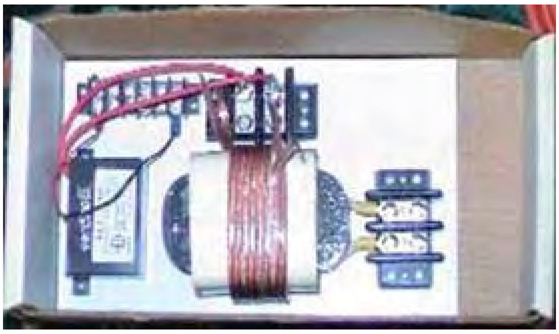

COMMENTS: Between sparks, L2 has a voltage on it's ends. If RL is connected directly to L2 then there will be no output current without resonance and there will be no output current without a spark.

MORE ACCURATE:

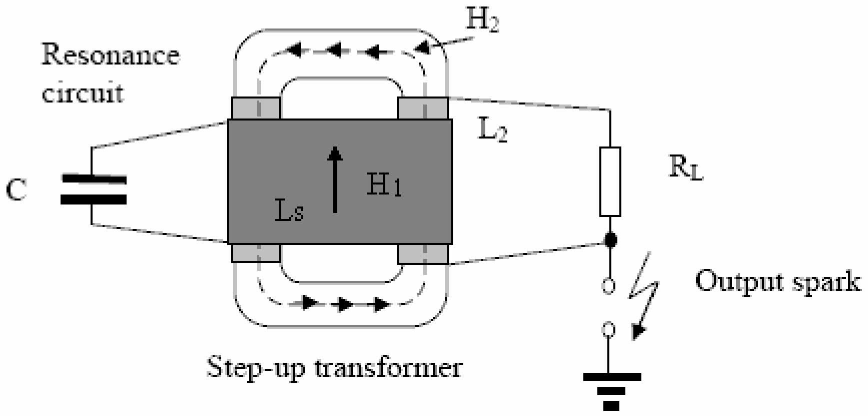

H2
Resonance circuit
L2
RL
C
H1
Ls
Output spark
Step-up transformer

COMMENTS: L2 has no voltage on it's ends (without a spark). This is ordinary back-EMF suppression, invented by Nikola Tesla.

MORE USEFUL

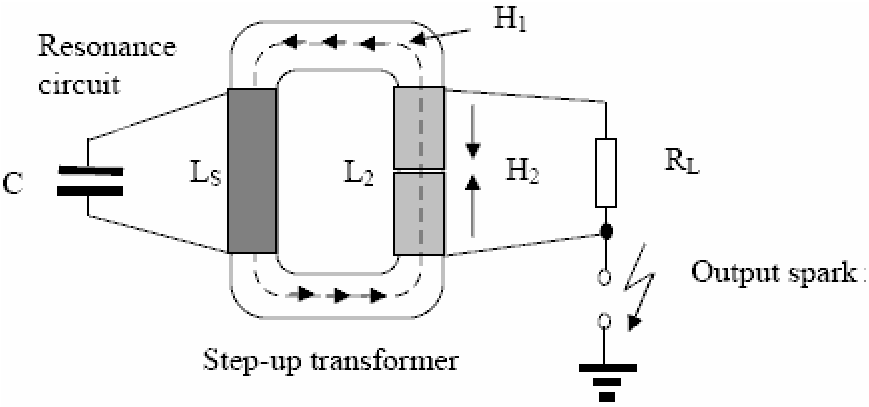

H1
Resonance circuit
RL
C
Ls
L2
H2
Output spark
Step-up transformer

COMMENT: L2 has no voltage on it's ends (without a spark).

Secret 3.1

THE ASYMMETRICAL TRANSFORMER BASED ON THE SHORT-CIRCUITED COIL

INTRODUCTION

Remark: Voltage distribution on the shorted coil depends on the position of the exciting coil.

DESCRIPTION

CASE 1 The excitation coil is centered:
Result: We have the full period of the voltage distribution on the short-circuited coil

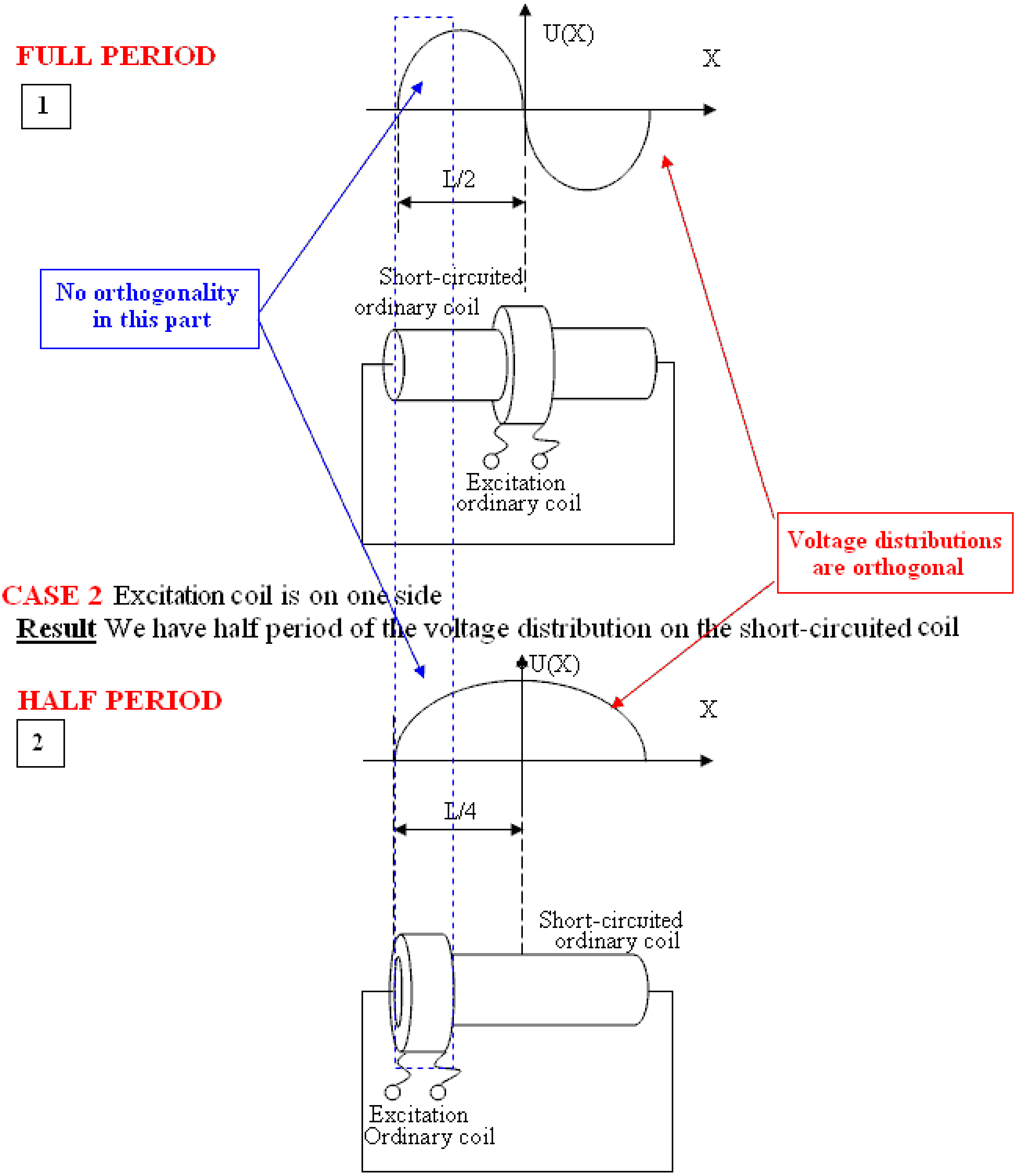
FULL PERIOD
1
U(X)
X
L/2
No orthogonality in this part
Short-circuited ordinary coil
Excitation ordinary coil
Voltage distributions are orthogonal
HALF PERIOD
2
U(X)
X
L/4
Short-circuited ordinary coil
Excitation Ordinary coil

CASE 2 Excitation coil is on one side
Result We have half period of the voltage distribution on the short-circuited coil

CONSTRUCTION OF THE ASYMMETRICAL TRANSFORMER
based on the short-circuited coil

CASE 1 The short-circuited coil is wound in one direction

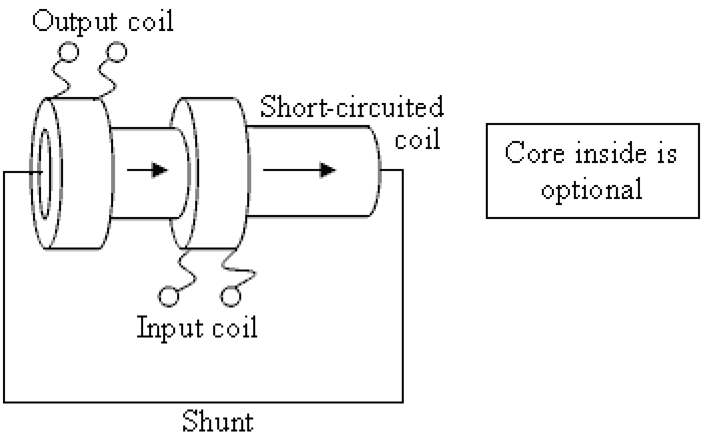

Result: The output does not influence the input in any way.

Explanation: The signal from the output coil generates zero voltage difference on the input coil.

Remark: The position of the coils should be adjusted in order to give the best result.

CASE 2: The short-circuited coil is wound in opposite directions from the centre outwards, and only half of the coil is short-circuited:

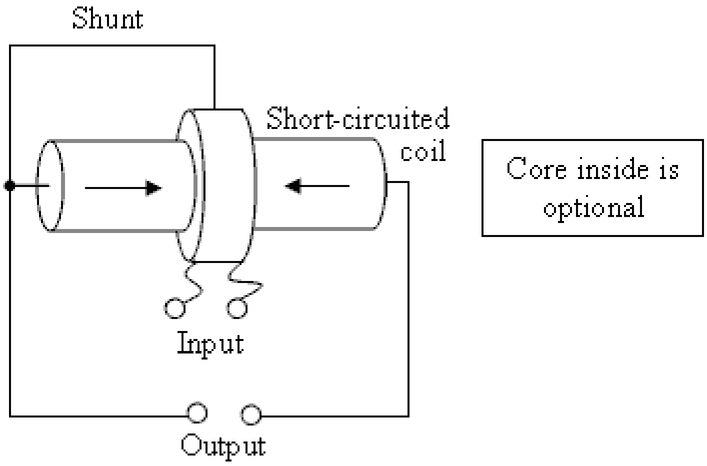

Result: The output has no influence on the input coil

Explanation: The signal from the output coil generates zero voltage difference on the input coil.

Remark: The position of the input coil needs to be adjusted to get the best result.

Remark: The coil's position depends on permeability of the core. More permeability means more alike with distribution pointed at the beginning.

Best Position: To find the best coil position, connect the signal generator to the output, and then find the coil position which shows zero at the input terminals. Alternatively, use an RLC meter connected to the input terminals and then find the coil position which gives no change in reading when the output terminals are short-circuited **(for both case 1 and case 2)**.

Comment: The length of the wire, the total length of the coil, and the diameter of the coil are not important. The number of turns in the input and output coils plays the same role as in an ordinary transformer, **for both case 1 and case 2**.

MODERN APPLICATIONS FOR SHORT-CIRCUITED COILS

By Don Smith

CASE 1

ELECTRICAL ENERGY GENERATING SYSTEM

Patent Pending 08 / 100,074

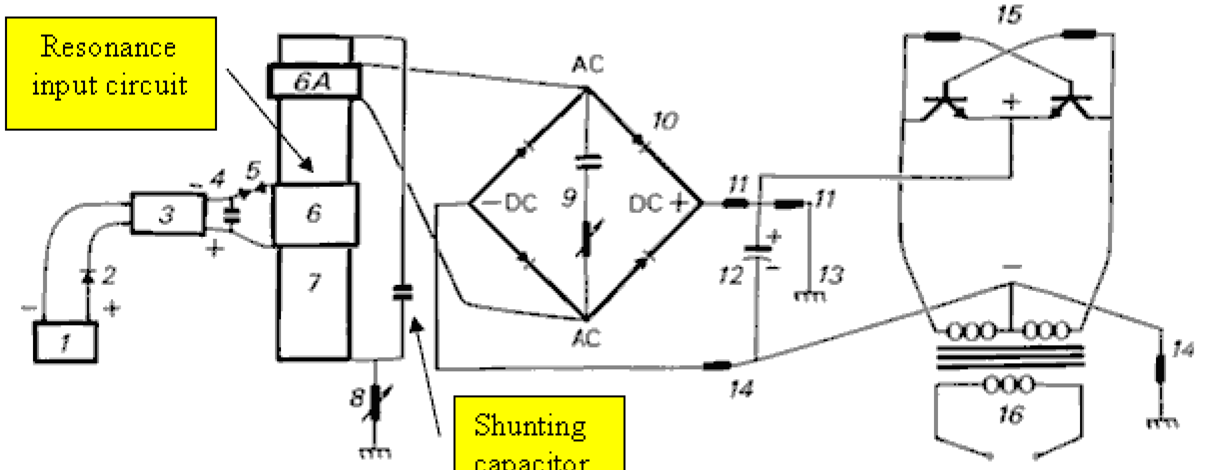

1. Geicel, 6 or 12 Volt.
2. Diode, Poss. use a Varactor.
3. High Voltage Module, Constituting the L-1 and L-2 Coils.
4. Capacitor, TDK 10.9 Pf., 30 KV.
5. Spark Gap, Small Engine Spark Plug, Gap = .0025 in.
6. Induction Transfer Coil L-3. , 6A = L-5
7. Induction Receiving Coil L-4.
8. Voltage Control Shunt.
9. Frequency Adjustor, prevents derating by Diode Bridge
10. Diode Bridge, 200 Nanosecond, R.F., > 100 KV .
11. Voltage Divider Circuit, corrects voltage for next stage.
12. Capacitor, electrolytic, smooths out DC + ripple effect.
13. Earth Ground.
14. Voltage Divider Curcit, corrects voltage for Transformer
15. Inverter Circuit, DC + in and 60 CPS to Transformer
16. Output from Transformer to Load (Work).

20 Dec., 1994

CASE 2

ELECTRICAL ENERGY GENERATING SYSTEM

Patent Pending 08 / 100,074

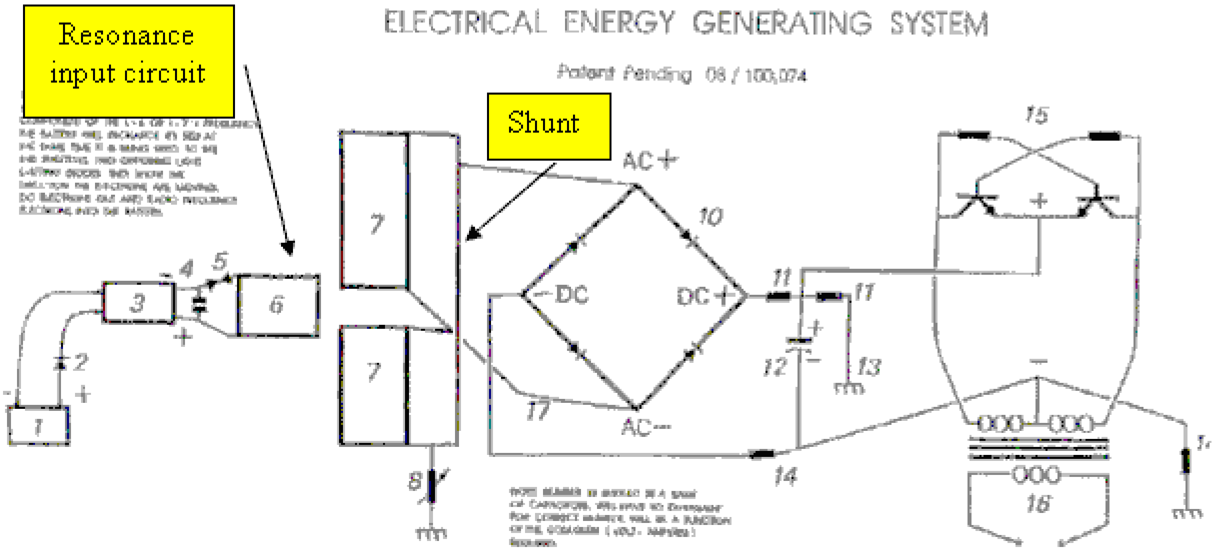

1. Geicel, 6 or 12 Volt,
2. Diode, Poss. use a Varactor.
3. High Voltage Module, Constituting the L-1 and L-2 Coils.
4. Capacitor, TDK 10.9 Pf., 30 KV.
5. Spark Gap, Small Engine Spark Plug, Gap = .0025 in.
6. Induction Transfer Coil L 3.
7. Induction Receiving Coil L-4.
8. Voltage Control Shunt.
9. Frequency Adjustor, prevents derating by Diode Bridge
10. Diode Bridge, 200 Nanosecond, R.F., > 100 KV .
11. Voltage Divider Circuit, corrects voltage for next stage.
12. Capacitor, electrolytic, smooths out DC + ripple effect.
13. Earth Ground.
14. Voltage Divider Circuit, corrects voltage for Transformer
15. Inverter Circuit, DC + in and 60 CPS to Transformer
16. Output from Transformer to Load (Work).
17. Center Tap

17 JANUARY, 1997

REMARK: The position of the coils must be adjusted until the output has zero influence on the input.

REMEMBER: None of the (input) energy used for exciting ambient space should appear in the load.

AN EXAMPLE OF CASE 2

By Don Smith

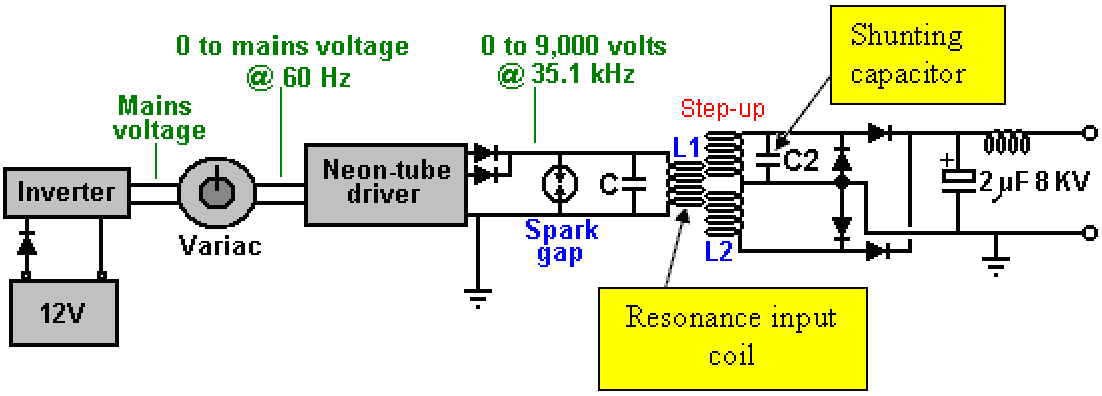
0 to mains voltage
@ 60 Hz
0 to 9,000 volts
@ 35.1 kHz
Shunting capacitor
Mains voltage
Step-up
Inverter
Variac
Neon-tube driver
C
L1
C2
2 µF 8 KV
Spark gap
L2
12V
Resonance input coil

COMMENTS: The output coil can be adjusted to resonate with the input coil, but this is not important for understanding the principle. Excitation with just one spark is possible (not in resonance), but the frequency of the sparks influences the output power directly.

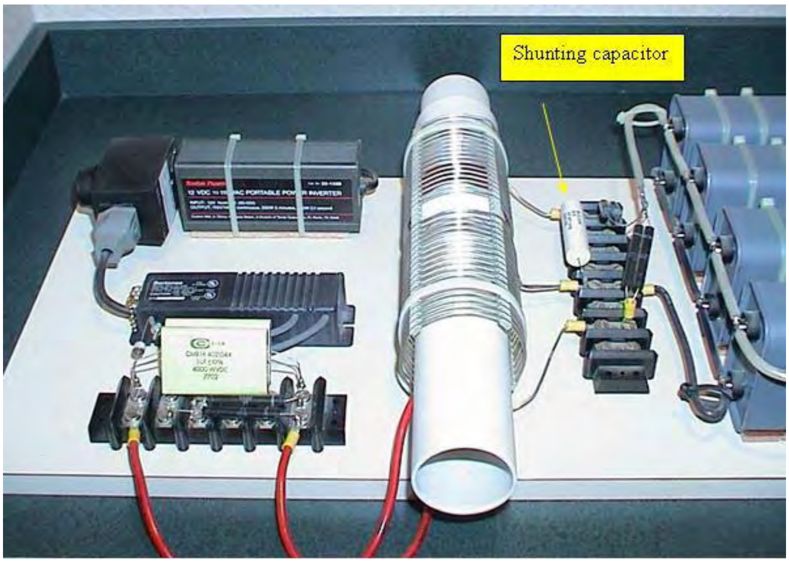
Shunting capacitor

COMMENT: Don Smith used the simplest **doubling excitation frequency**

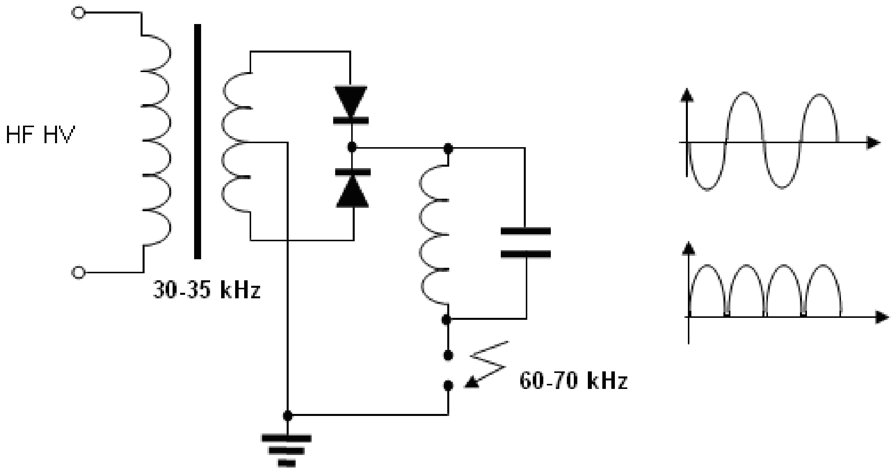
HF HV
30-35 kHz
60-70 kHz

COMMENTS: Resonance frequency of the circuit is about 60-70 kHz, but dimmer is for 30-35 kHz.

For adjusting the excitation frequency, voltage/frequency technology was used.

You have to adjust two parameters: the position of the slider and the excitation frequency.

MODERN APPLICATION FOR SHORT-CIRCUITED COILS
By William Barbat

US Patent Application number 2007/0007844

Self-Sustaining Electric-Power Generator Utilizing Electrons of Low Inertial Mass to Magnify Inductive Energy

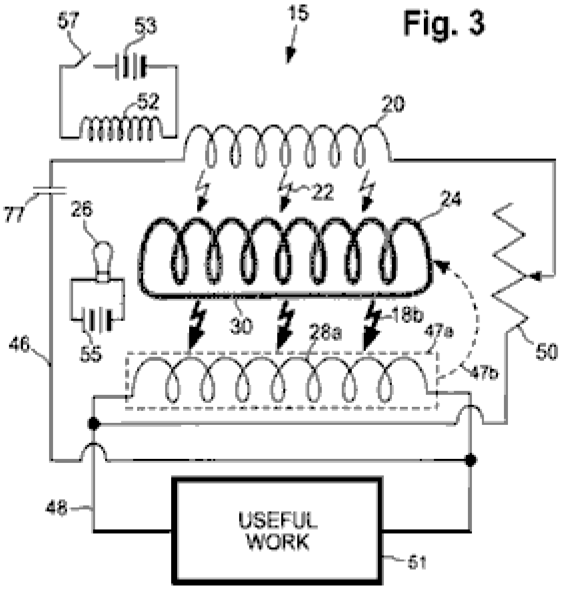
Fig. 3
57
53
15
52
20
77
26
22
24
30
28a
18b
47a
47b
50
46
55
48
USEFUL WORK
51

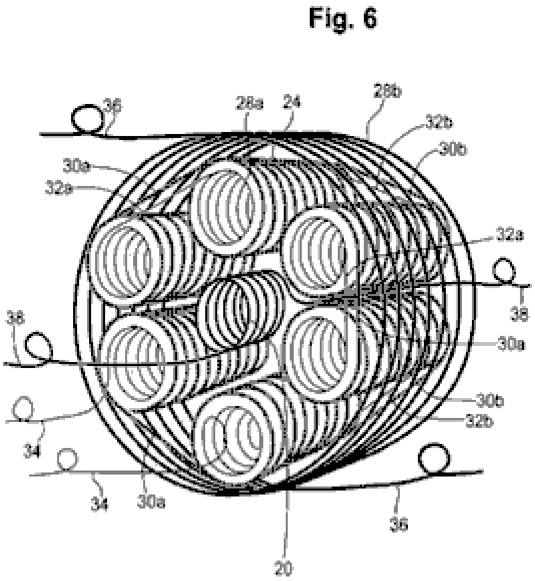
Fig. 6
36
28a
24
28b
32b
30b
30a
32a
32a
38
38
30a
30b
32b
34
34
30a
36
20

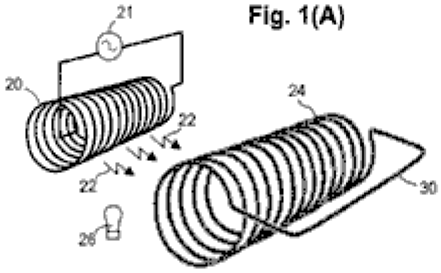
Fig. 1(A)
21
20
24
22
22
30
26

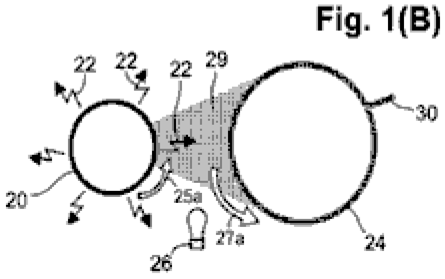
Fig. 1(B)
22
22
22
29
30
20
25a
27a
26
24

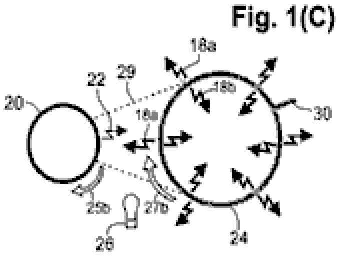
Fig. 1(C)
18a
22
29
18b
20
30
18a
25b
27b
26
24

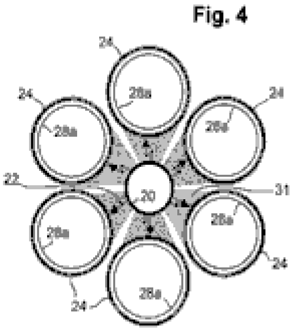
Fig. 4
24
28a
24
24
28a
28a
22
20
31
28a
28a
24
24
28a

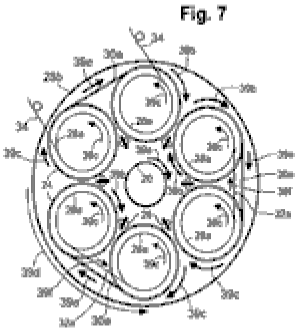
Fig. 7

COMMENT In order to understand this device, you have to read Barbat's patent application US 2007/0007844 A1: www.free-energy-info.com/PatD25.pdf

COMMENT I would like to point out that externally, it looks very much like Alfred Hubbard's device.

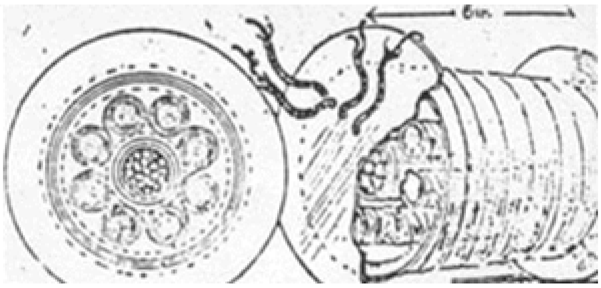

This is Hubbard's device.

AN EXAMPLE OF CASE 1

By Tariel Kapanadze

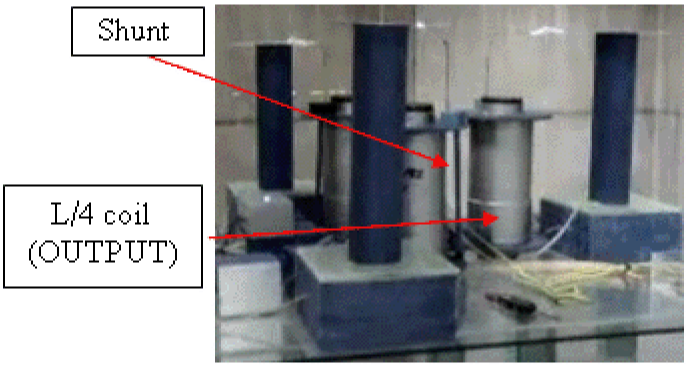
Shunt
L/4 coil (OUTPUT)

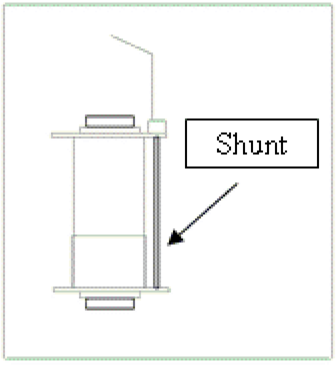

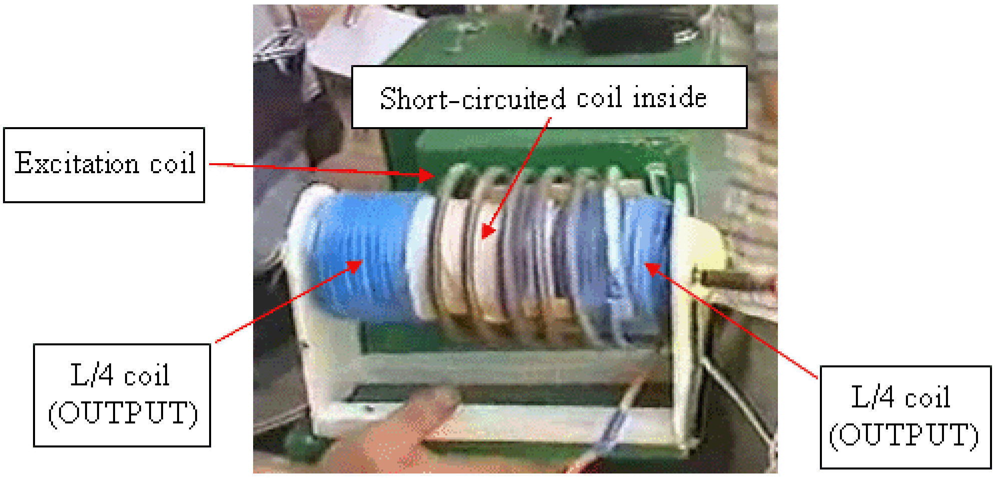
Short-circuited coil inside
Excitation coil
L/4 coil (OUTPUT)
L/4 coil (OUTPUT)

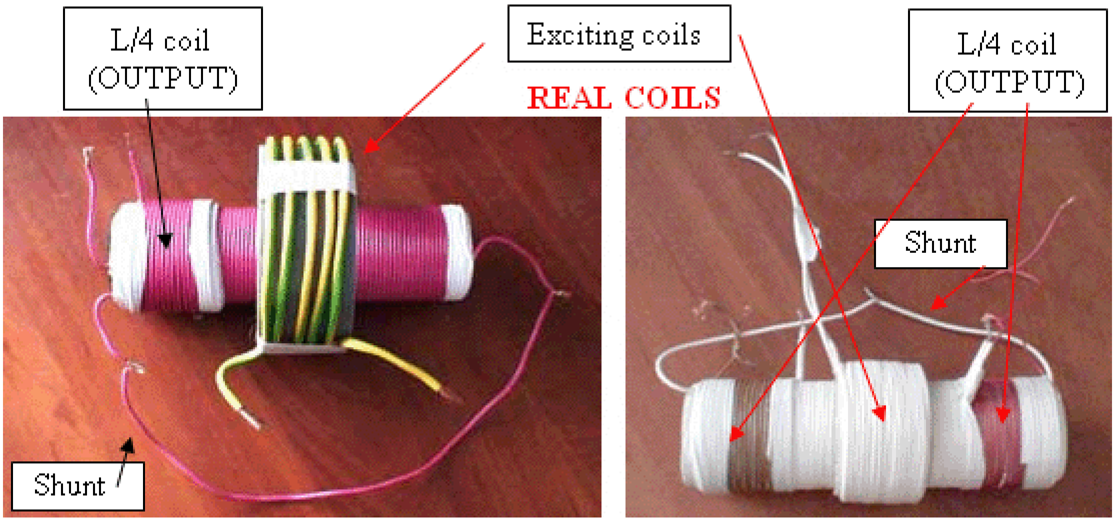
L/4 coil (OUTPUT)
Exciting coils
L/4 coil (OUTPUT)
REAL COILS
Shunt
Shunt

COMMENT: Adjust the positions of the coils to get the best result.

AN EXAMPLE OF CASE 1

By Steven Mark

TPU

REMARK: An idea – an asymmetrical transformer based on the shorted-circuited coil:

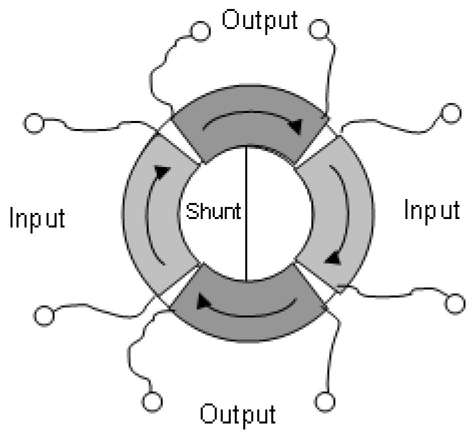
Output
Input
Shunt
Input
Output

REMARK: The positions of the coils must be properly adjusted, in order to have no transmission feedback from the output to the input. To understand this better, read the part which is devoted to switchable inductance.

EXPLANATION:

The first layer (shunted coil wound around the whole core) The second layer (coils for input and output)

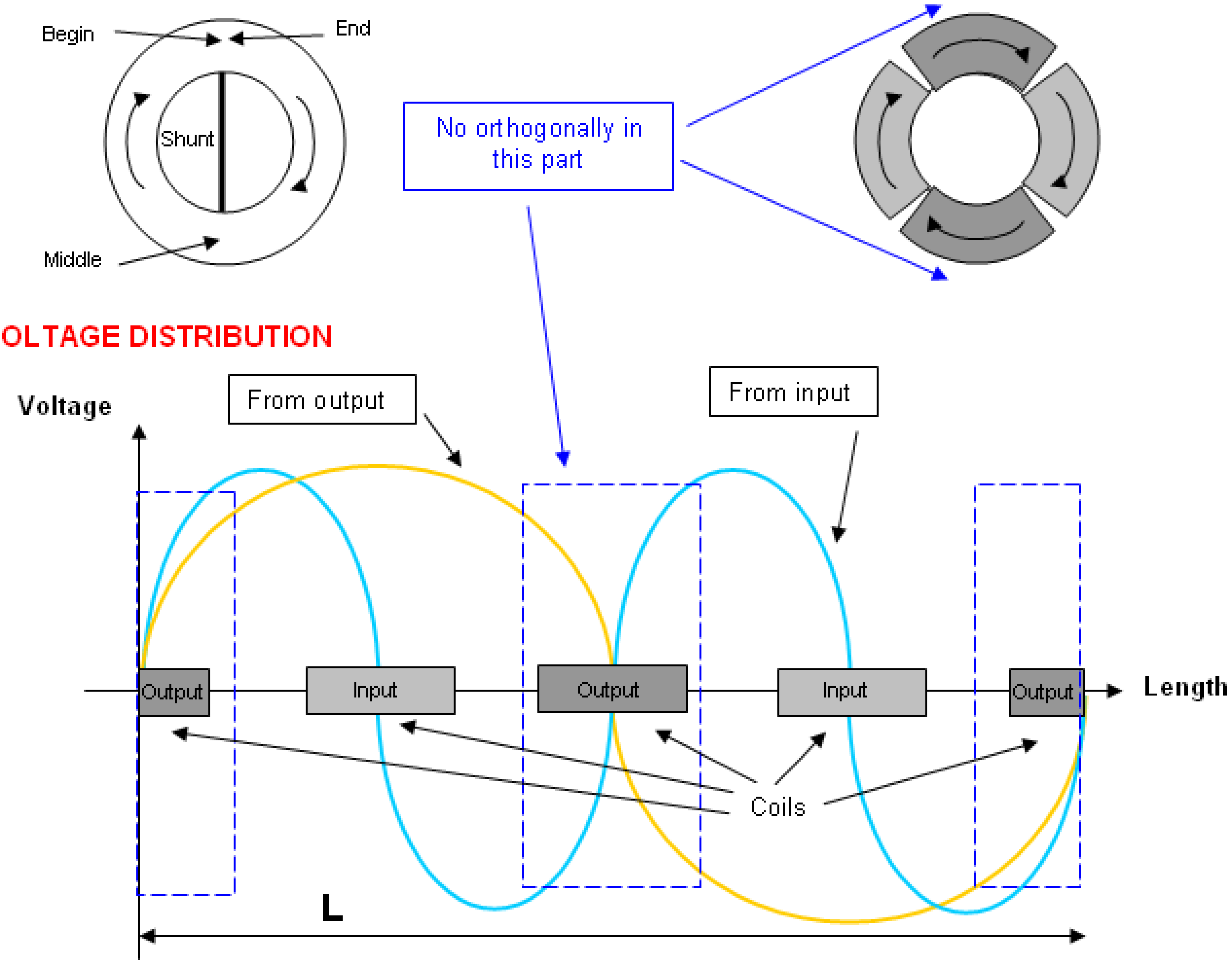
Begin
End
Shunt
Middle
No orthogonally in this part
Voltage
From output
From input
Output
Input
Output
Input
Output
Length
Coils
L

VOLTAGE DISTRIBUTION

THE BASIS OF THE TPU

(Tesla Patent)

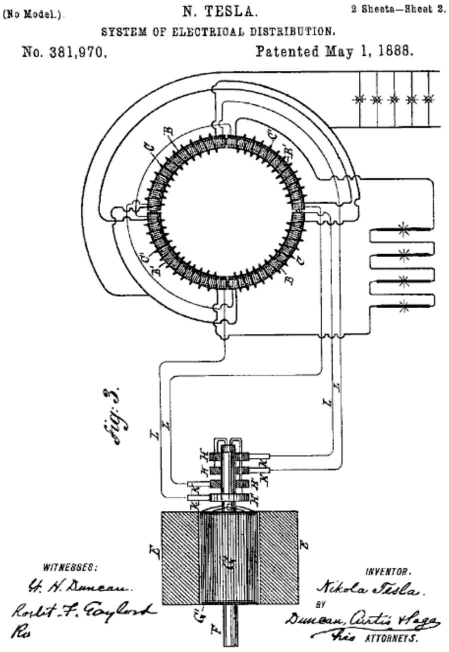
(No Model.) N. TESLA. 2 Sheets—Sheet 2.
SYSTEM OF ELECTRICAL DISTRIBUTION.
No. 381,970. Patented May 1, 1888.

Fig. 3.
WITNESSES:
H. H. Duncan.
Robt. F. Gaylord
INVENTOR.
Nikola Tesla.
BY
Duncan, Curtis & Page
His ATTORNEYS.

ACTUAL TPU

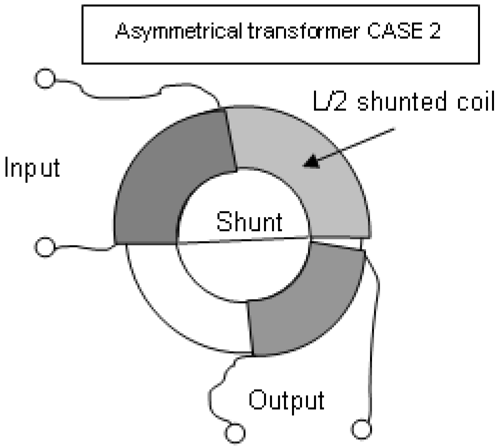
Asymmetrical transformer CASE 2
L/2 shunted coil
Input
Shunt
Output

REMEMBER:

The position of the coils must be adjusted.
The easiest way to do this is to add or remove
turns at the ends of the coils.

AN EXAMPLE OF CASE 2

By Tariel Kapanadze

Mechanical device

?????

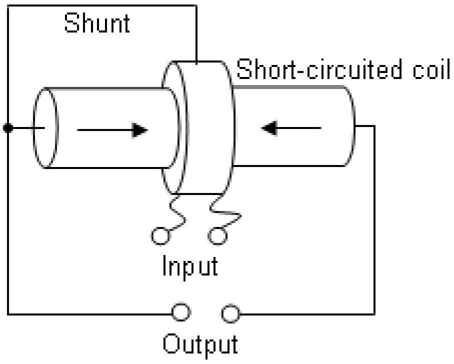
Shunt
Short-circuited coil
Input
Output

MODERN USE OF SHORT-CIRCUITED COILS

by Cherepanov Valera ('SR193' in Russian forum)

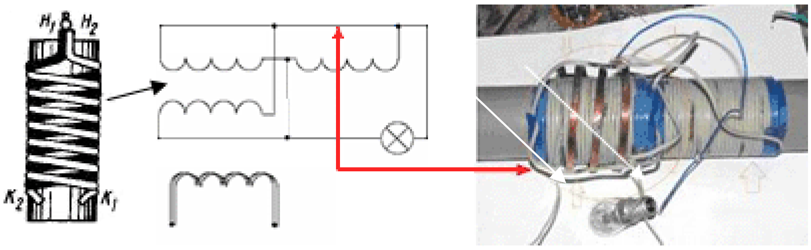

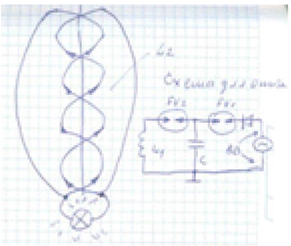

COMMENT: This arrangement can be used for back-EMF suppression in resonance (spark excited) mode to get a laser effect (very exciting summation effects).

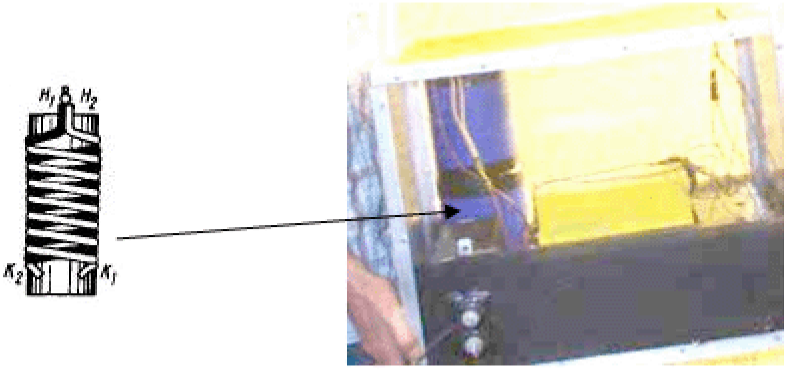

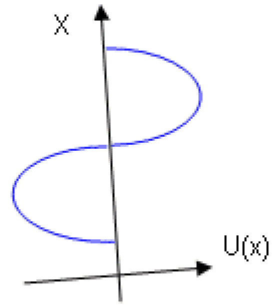

COMMENT: This was copied from this device of Tariel Kapanadze (???).

Don Smith

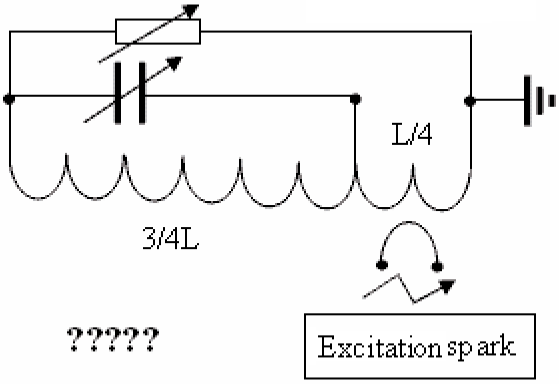

COMMENT: Mr. Tesla said: "The optimum relation for the main and additional coil is 3/4L and L/4". Is that ratio used here?

COMMENT: If you don't understand this schematic, look at simplest version of the coil.

THE SIMPLEST VERSION

where the output has zero influence on the input

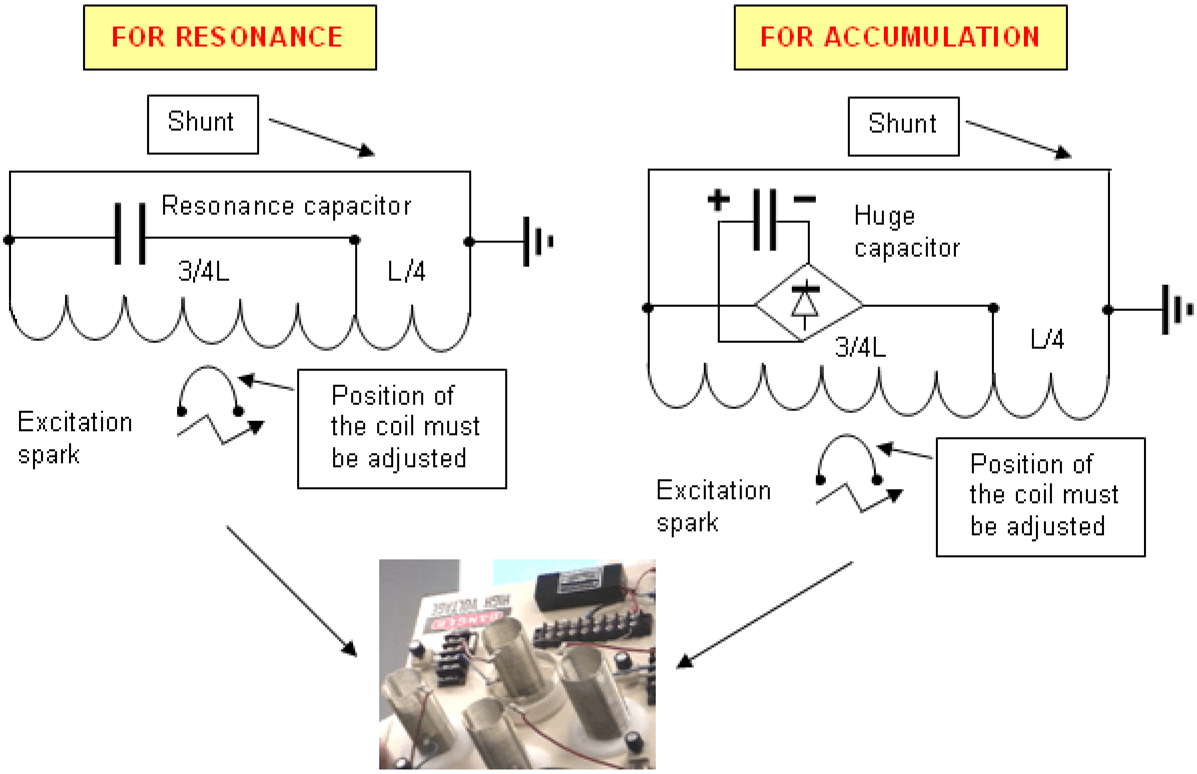
FOR RESONANCE
Shunt
Resonance capacitor
3/4L
L/4
Excitation spark
Position of the coil must be adjusted
FOR ACCUMULATION
Shunt
+
−
Huge capacitor
3/4L
L/4
Excitation spark
Position of the coil must be adjusted

It can be used like this....

Comment: This is an instance of case 1 where the output coil was removed, and some of the turns from the short-circuited coil were used instead.

THE ASYMMETRICAL TRANSFORMER (BASED ON A SHORT-CIRCUITED COIL) COMBINED WITH A STEP-DOWN TRANSFORMER?

Don Smith

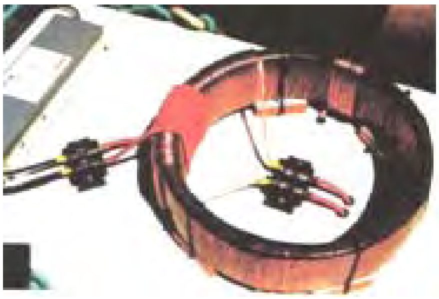

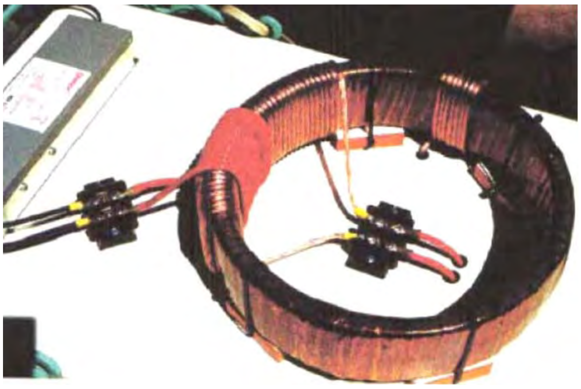

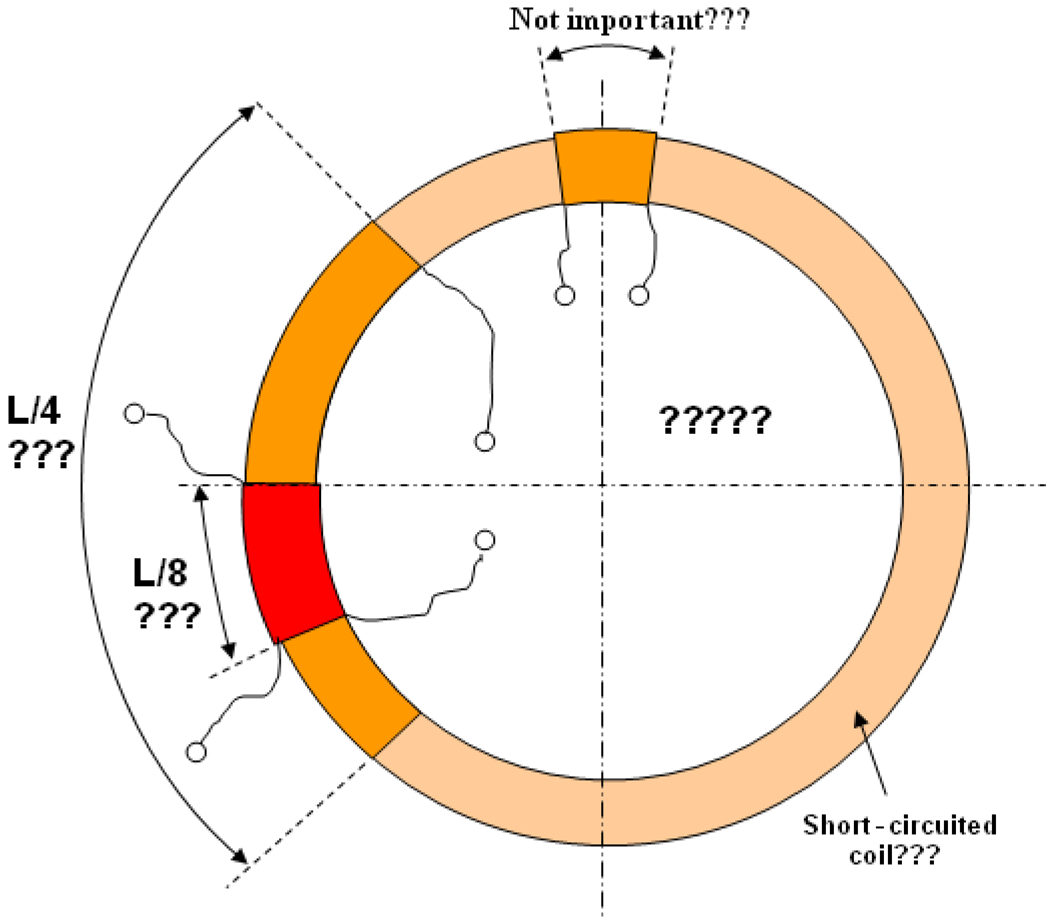
Not important???
L/4
???
L/8
???
?????
Short-circuited
coil???

THE RELATIONSHIPS of Don Smith's TPU size and position are important.

REMARK: Those relationships are used to produce an asymmetrical transformer

MECHANICAL ANALOGUE OF THE ASYMMETRICAL TRANSFORMER

CASE 2

By Don Smith

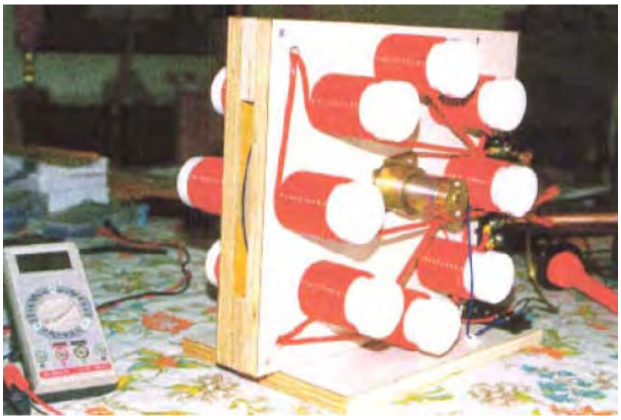

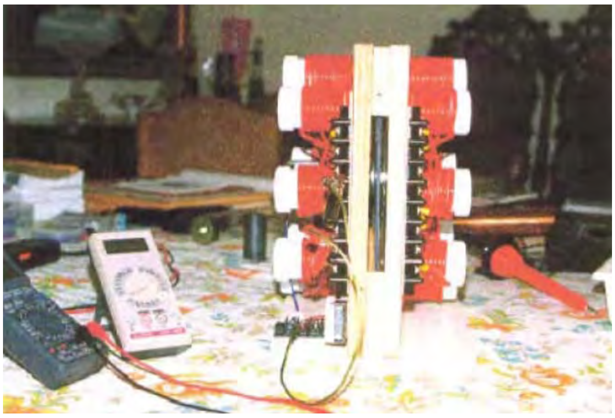

Schematic:

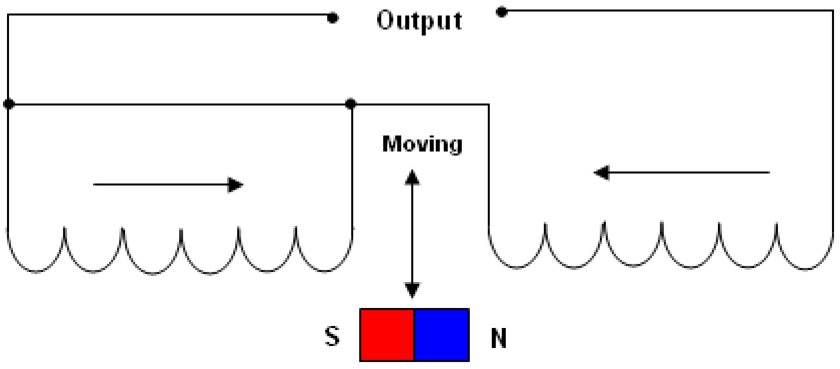
Output
Moving
S
N

REMEMBER: Any asymmetrical transformer must be adjusted.

REMARK: Don Smith placed magnets inside the coils, but that is not important for understanding the process as his device does not match the schematic.

SOME REMARKS ON ASYMMETRICAL IN-FRONT CONNECTION

(Useful remarks)

Some turns were added on one half of the coil, and some turns were removed from the other half. An additional magnetic field H_3 was created, with inductance - L_D.

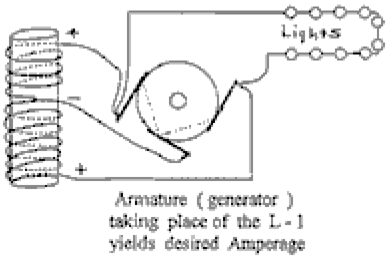
Lights
+
−
+
Armature (generator)
taking place of the L - 1
yields desired Amperage

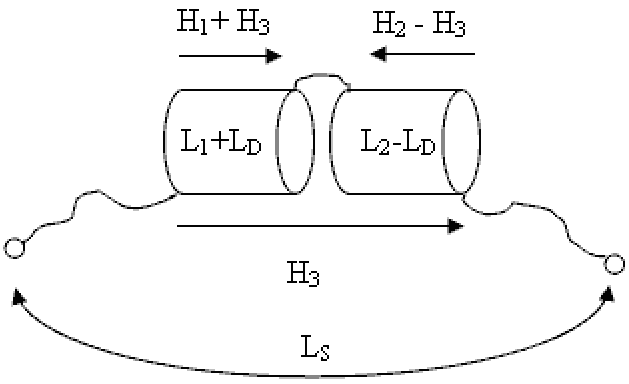
$H_1 + H_3$
$H_2 - H_3$
$L_1 + L_D$
$L_2 - L_D$
H_3
L_S

RESULT: A large part of the total inductance acts as an inductor, and a small part acts as a capacitor. **This is a well known fact (read books).** The total voltage on the coil is less than on it's halves.

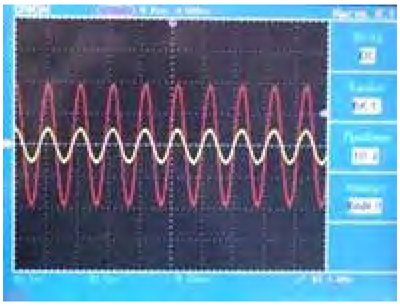

Yellow – The voltage on the total coil

Red – The voltage on the large section of that coil

RESULT: The voltage on it's halves is 4 times the voltage on the total coil

The measurements were made in the frequency band 10 kHz to 100 kHz.

Here is the result of a capacitor discharging into this coil:

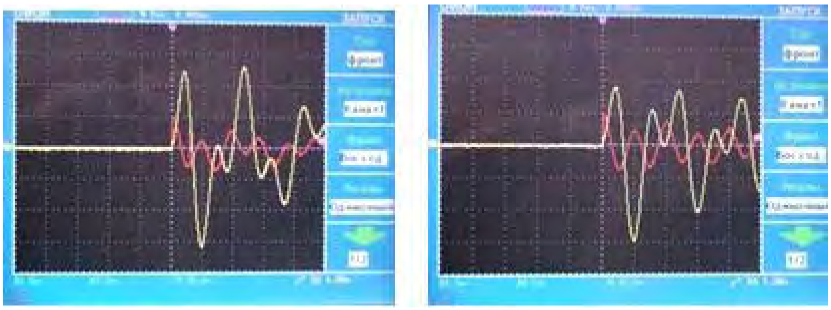

SECRET 4

CURRENT AMPLIFICATION

If a lot of asymmetric transformers are placed with a common flux flow through them, they will have no influence on this flux flow, as any one asymmetric transformer does not have any influence on the flux flow. If the secondary L2 transformer coils are then connected in parallel, this produces current amplification.

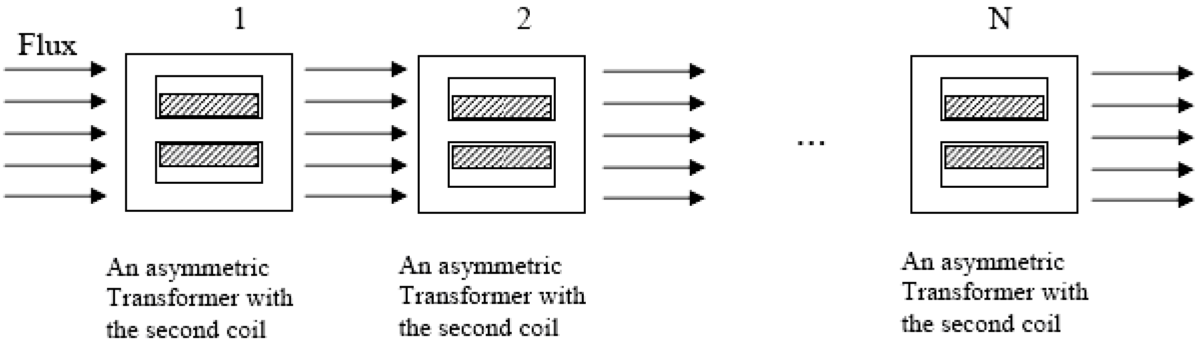
Flux
1
2
N
...
An asymmetric Transformer with the second coil
An asymmetric Transformer with the second coil
An asymmetric Transformer with the second coil

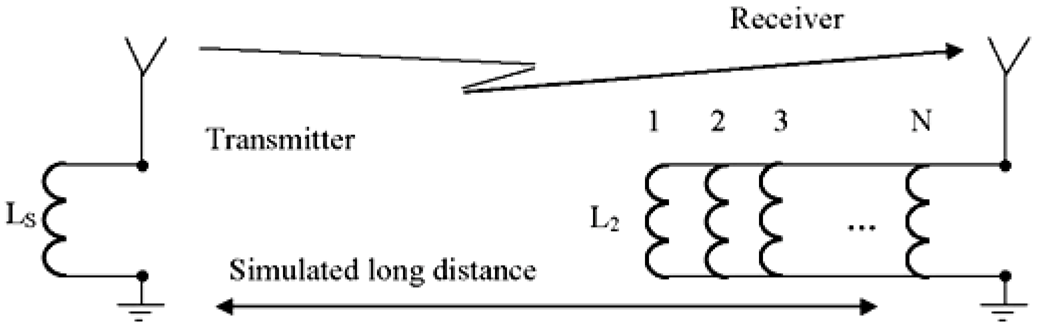
Receiver
Transmitter
1
2
3
N
Ls
L2
...
Simulated long distance

AS A RESULT

You have an asymmetric transformer arranged in a stack:

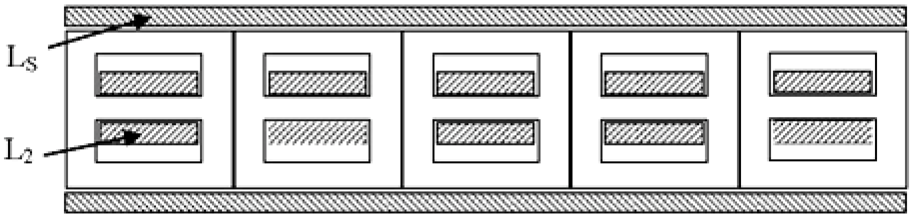

For flat (uniform) field inside of LS, it can be arranged with additional turns at it’s ends.

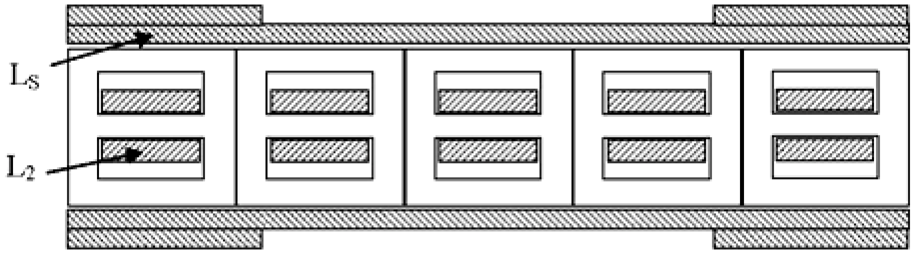

EXAMPLES OF COILS WHICH WERE ACTUALLY CONSTRUCTED

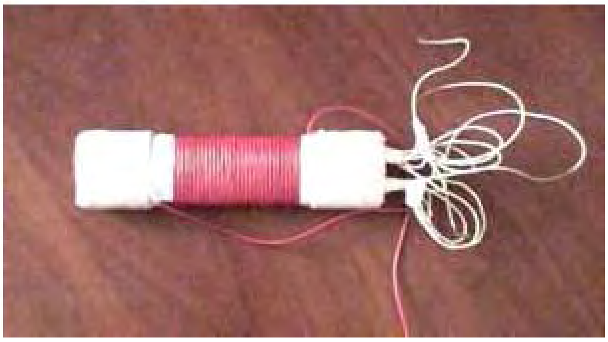

The coils are constructed from 5 sections, made from E-type ferrite core with a permeability of 2500, and wound using plastic-covered wire. The central sections L2 have 25 turns, and edge sections have 36 turns (to equalise the voltage on them). All sections are connected in parallel. The coil Ls has field-flattening turns at it's ends, and a single-layer winding Ls was used, the number of turns depending on the diameter of the wire used.

The current amplification for these particular coils is 4 times.

Changing Ls inductance is 3% (if L2 is short-circuited)

SECRET 5

The power source in Nikola Tesla car "Red arrow" is FERROMAGNETIC RESONANCE

1931 Pierce-Arrow 8 Model 43 D.C. Phaeton
Great Cars & Grand Marques Collection

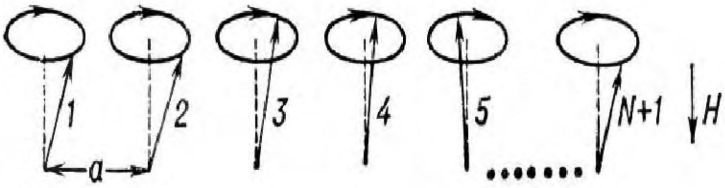

COMMENT: To understand electromagnetic feedback, you must consider the action to be like that of domains which have a group behaviour, or alternatively, spin waves (like a row of standing dominos falling over where each one is toppled by the previous one hitting it).

THE BASIS OF FERROMAGNETIC RESONANCE

When a ferromagnetic material is placed in a magnetic field, it can absorb external electromagnetic radiation in a direction perpendicular to the direction of the magnetic field, which will cause ferromagnetic resonance at the correct frequency.

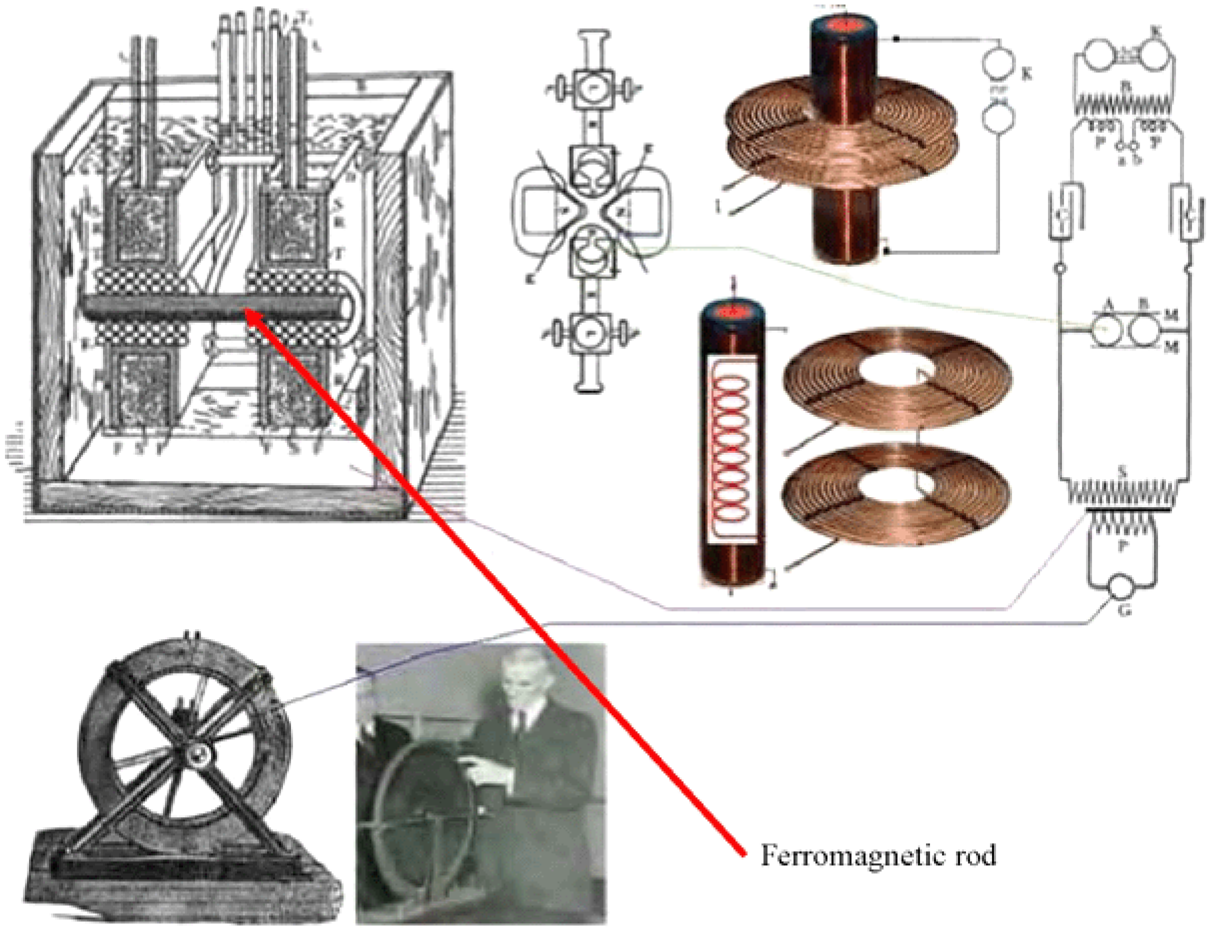

This is an energy-amplifying transformer invented by Mr. Tesla.

QUESTION: What use is a ferromagnetic rod in Free-Energy devices?

AN ANSWER: It can change magnetisation of the material along magnetic field direction without the need for a powerful external force.

QUESTION: Is it true that the resonant frequencies for ferromagnetics are in the tens of Gigahertz range?

AN ANSWER: Yes, it is true, and the frequency of ferromagnetic resonance depends on the external magnetic field (a strong magnetic field produces a high frequency). But with ferromagnetics it is possible to get resonance without applying any external magnetic field, this is the so-called "natural ferromagnetic resonance". In this case, the magnetic field is defined by the local magnetisation of the sample. Here, the absorption frequencies occur in a wide band, due to the large variations possible in the conditions of magnetisation, and so you must use a wide band of frequencies to get ferromagnetic resonance

A POSSIBLE PROCESS FOR ACQUIRING FREE-ENERGY

1. Subjecting a ferromagnetic to a short electromagnetic pulse even without an external magnetic field, causes the acquisition of spin precession (domains will have group behaviour, and so ferromagnetics can easily be magnetised).

2. Magnetisation of ferromagnetics can be by an external magnetic field.

3. Energy acquisition can be as a result of strong sample magnetisation caused by an external magnetic field of lesser strength.

COMMENT: You must use synchronisation for processes of irradiation and magnetisation of the sample.

USEFUL COMMENT: A ferromagnetic shield will not destroy the inductance of any coil placed inside it, provided that the ends of that coil are positioned on one side of the coil.

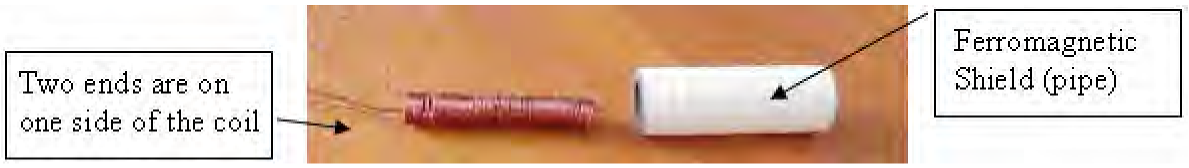
Two ends are on one side of the coil
Ferromagnetic Shield (pipe)

But, this coil can magnetise the ferromagnetic shield.

SECRET 5 CONTINUATION …

TWO PERPENDICULAR COILS ON A COMMON AXIS

(Standing waves, spin waves, domino effect, laser effect, open resonator, etc…)

EXPLANATION: Standing waves can be excited not only in Tesla's "horseshoe" magnet, but also in Tesla's ferromagnetic transformer (excited by sparks…)

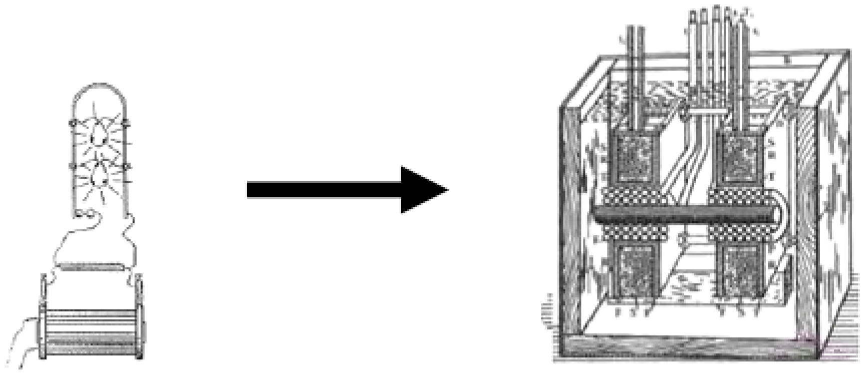

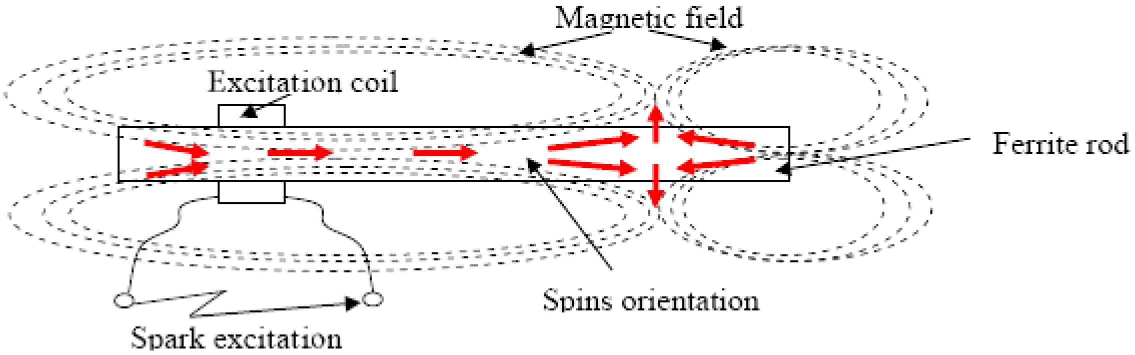
Magnetic field
Excitation coil
Ferrite rod
Spins orientation
Spark excitation

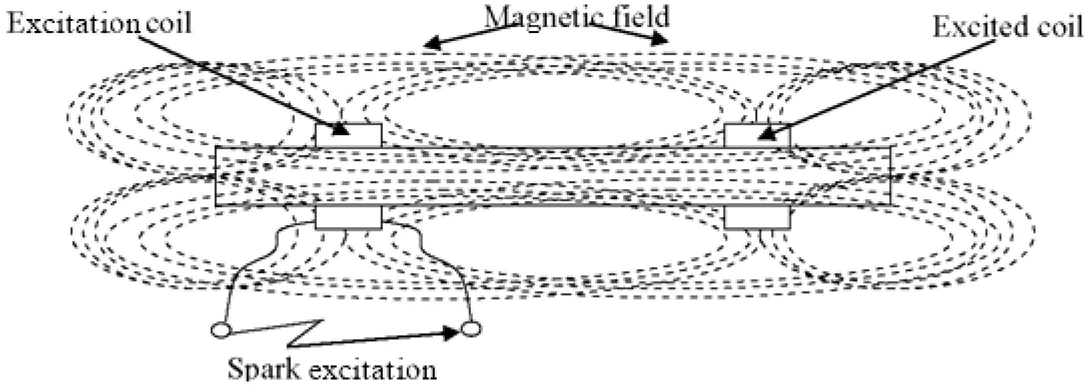
Excitation coil
Magnetic field
Excited coil
Spark excitation

COMMENT: Excitation can be arranged in different ways, by coils connection. The frequencies of oscillations in a coil depends on the number of turns in it (a big variation is possible due to this factor).

ACTUAL COILS

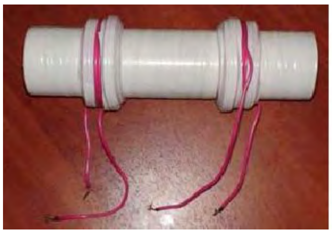

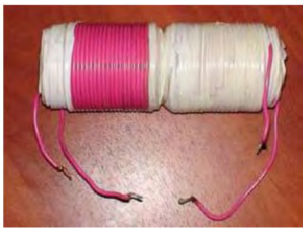

COMMENT: The positions of the coils on the rods depends on whatever ferromagnetic material is being used, and on it's size. The optimum arrangement has to be determined through experimentation.

A transformer can have two pairs of coils: exciting (tubes), resonance or load (inside) – see Tesla's picture

TOROIDAL VERSION OF AN ASYMMETRIC STACKED TRANSFORMER

An inductor L2 is placed on the central ring between the short-circuits of the core, and the coil L_S (not shown) is wound around all three rings, covering the whole of the toroid - this is an ordinary toroidal coil.

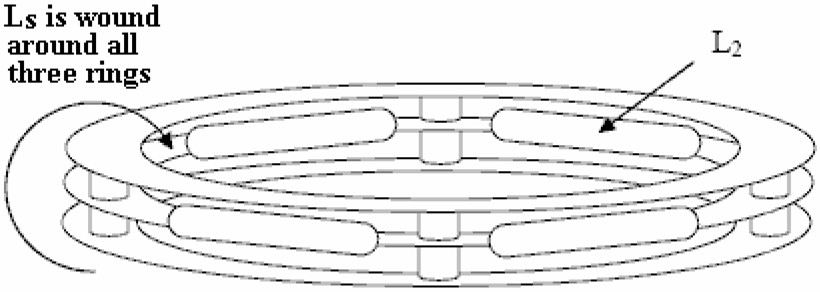

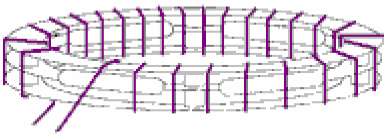

The number of short-circuits depends on your requirements, and influences on the current amplification.

THAT IS ALL, GOOD LUCK …

CONCLUSIONS

1. The Energy-Conservation Law is a result (not reason) of symmetrical interaction.
2. The simplest way to destroy symmetrical interaction is by using electromagnetic field feedback.
3. All asymmetrical systems are outside the area covered by the Energy-Conservation Law.

THE ENERGY CONSERVATION LAW CANNOT BE VIOLATED
(The field covered by this law is only symmetrical interactions)

No Private or State secrets are contained in this document.

There are no ready-to-use schematics in this document, as all diagrams are only provided as an aid to understanding the principles involved.

ASM HANDBOOK®

VOLUME
6

WELDING BRAZING AND SOLDERING

ASM
INTERNATIONAL ®

PUBLICATION INFORMATION AND CONTRIBUTORS

WELDING, BRAZING, AND SOLDERING WAS PUBLISHED IN 1993 AS VOLUME 6 OF THE ASM HANDBOOK. THE VOLUME WAS PREPARED UNDER THE DIRECTION OF THE ASM HANDBOOK COMMITTEE.

VOLUME CHAIRMEN

THE VOLUME CHAIRMEN WERE DAVID LEROY OLSON, THOMAS A. SIEWERT, STEPHEN LIU, AND GLEN R. EDWARDS.

AUTHORS

- **LAMET** UFRGS
- **BRUNO L. ALIA**
- **RICHARD L. ALLEY** AMERICAN WELDING SOCIETY
- **WILLIAM R. APBLETT, JR.**
- **WILLIAM A. BAESLACK III** THE OHIO STATE UNIVERSITY
- **WILLIAM BALLIS** COLUMBIA GAS OF OHIO
- **CLIFF C. BAMPTON** ROCKWELL INTERNATIONAL SCIENCE CENTER
- **PROBAL BANERJEE** AUBURN UNIVERSITY
- **JOHN G. BANKER** EXPLOSIVE FABRICATORS INC.
- **ROBERT G. BARTIFAY** ALUMINUM COMPANY OF AMERICA
- **ROY I. BATISTA**
- **ROY E. BEAL** AMALGAMATED TECHNOLOGIES INC.
- **RAYMOND E. BOHLMANN** MCDONNELL AIRCRAFT COMPANY
- **SÉRGIO D. BRANDI** ESCOLA POLITECNICA DA USP
- **JOHN A. BROOKS** SANDIA NATIONAL LABORATORIES
- **DONALD W. BUCHOLZ** IBM FEDERAL SYSTEMS CORPORATION
- **PAUL BURGARDT** EG&G ROCKY FLATS PLANT
- **ROGER A. BUSHEY** THE ESAB GROUP INC.
- **CHRIS CABLE** FEIN POWER TOOL
- **RICHARD D. CAMPBELL** JOINING SERVICES INC.
- **HOWARD CARY** HOBART BROTHERS COMPANY
- **HARVEY CASTNER** EDISON WELDING INSTITUTE
- **ALLEN CEDILOTE** INDUSTRIAL TESTING LABORATORY SERVICES
- **HARRY A. CHAMBERS** TRW NELSON STUD WELDING
- **C. CHRIS CHEN** MICROALLOYING INTERNATIONAL INC.
- **SHAOFENG CHEN** AUBURN UNIVERSITY
- **SHAO-PING CHEN** LOS ALAMOS NATIONAL LABORATORY
- **BRYAN A. CHIN** AUBURN UNIVERSITY
- **MICHAEL J. CIESLAK** SANDIA NATIONAL LABORATORIES
- **RODGER E. COOK** THE WILKINSON COMPANY
- **STEPHEN A. COUGHLIN** ACF INDUSTRIES INC.
- **MARK COWELL** METCAL INC.
- **RICHARD S. CREMISIO** RESCORP INTERNATIONAL INC.
- **CARL E. CROSS**
- **CRAIG DALLAM** THE LINCOLN ELECTRIC COMPANY
- **BRIAN DAMKROGER** SANDIA NATIONAL LABORATORIES
- **JOSEPH R. DAVIS** DAVIS AND ASSOCIATES

- **JANET DEVINE** SONOBOND ULTRASONICS
- **PAUL B. DICKERSON**
- **RAY DIXON** LOS ALAMOS NATIONAL LABORATORY
- **SUE DUNKERTON** THE WELDING INSTITUTE
- **KEVIN DUNN** TEXAS INSTRUMENTS INC.
- **CHUCK DVORAK** UNI-HYDRO, INC.
- **JIM DVORAK** UNI-HYDRO, INC.
- **ROBERT J. DYBAS** GENERAL ELECTRIC COMPANY
- **THOMAS W. EAGAR** MASSACHUSETTS INSTITUTE OF TECHNOLOGY
- **GLEN R. EDWARDS** COLORADO SCHOOL OF MINES
- **GRAHAM R. EDWARDS** THE WELDING INSTITUTE
- **W.H. ELLIOTT, JR.** OAK RIDGE NATIONAL LABORATORY
- **JOHN W. ELMER** LAWRENCE LIVERMORE NATIONAL LABORATORY
- **STEVEN C. ERNST** EASTMAN CHEMICAL COMPANY
- **WILLIAM FARRELL** FERRANTI-SCIAKY COMPANY
- **JOEL G. FELDSTEIN** FOSTER WHEELER ENERGY CORPORATION
- **DAVID A. FLEMING** COLORADO SCHOOL OF MINES
- **JAMES A. FORSTER** TEXAS INSTRUMENTS INC.
- **MICHAEL D. FREDERICKSON** ELECTRONICS MANUFACTURING PRODUCTIVITY FACILITY
- **EDWARD FRIEDMAN** WESTINGHOUSE ELECTRIC CORPORATION
- **R.H. FROST** COLORADO SCHOOL OF MINES
- **CHARLES E. FUERSTENAU** LUCAS-MILHAUPT INC.
- **EDWARD B. GEMPLER**
- **STANLEY S. GLICKSTEIN** WESTINGHOUSE ELECTRIC CORPORATION
- **JOHN A. GOLDAK** CARLETON UNIVERSITY
- **ROBIN GORDON** EDISON WELDING INSTITUTE
- **JERRY E. GOULD** EDISON WELDING INSTITUTE
- **JOHN B. GREAVES, JR.** ELECTRONICS MANUFACTURING PRODUCTIVITY FACILITY
- **F. JAMES GRIST**
- **JOHN F. GRUBB** ALLEGHENY LUDLUM STEEL
- **MAOSHI GU** CARLETON UNIVERSITY
- **IAN D. HARRIS** EDISON WELDING INSTITUTE
- **L.J. HART-SMITH** DOUGLAS AIRCRAFT COMPANY
- **DAN HAUSER** EDISON WELDING INSTITUTE
- **C.R. HEIPLE** METALLURGICAL CONSULTANT
- **HERBERT HERMAN** STATE UNIVERSITY OF NEW YORK
- **G. KEN HICKEN** SANDIA NATIONAL LABORATORIES
- **EVAN B. HINSHAW** INCO ALLOYS INTERNATIONAL INC.
- **D. BRUCE HOLLIDAY** WESTINGHOUSE MARINE DIVISION
- **S. IBARRA** AMOCO CORPORATION
- **J. ERNESTO INDACOCHEA** UNIVERSITY OF ILLINOIS AT CHICAGO
- **SUNIL JHA** TEXAS INSTRUMENTS INC.
- **JERALD E. JONES** COLORADO SCHOOL OF MINES
- **RAYMOND H. JUERS** NAVAL SURFACE WARFARE CENTER
- **WILLIAM R. KANNE, JR.** WESTINGHOUSE SAVANNAH RIVER COMPANY
- **MICHAEL J. KARAGOULIS** GENERAL MOTORS CORPORATION
- **MICHAEL KARAVOLIS** TEXAS INSTRUMENTS INC.
- **LENNART KARLSSON** LULEÅ UNIVERSITY OF TECHNOLOGY
- **MICHAEL E. KASSNER** OREGON STATE UNIVERSITY
- **DOUG D. KAUTZ** LAWRENCE LIVERMORE NATIONAL LABORATORY
- **W. DANIEL KAY** WALL COLMONOY CORPORATION
- **JAMES F. KEY** IDAHO NATIONAL ENGINEERING LABORATORY
- **H.-E. KIM** SEOUL NATIONAL UNIVERSITY

- **SAMUEL D. KISER** INCO ALLOYS INTERNATIONAL INC.
- **MARVIN L. KOHN** FMC CORPORATION
- **DAMIAN J. KOTECKI** THE LINCOLN ELECTRIC COMPANY
- **KENNETH KRYSIAC** HERCULES INC.
- **CHUCK LANDRY** THERMAL DYNAMICS
- **CHARLES LANE** DURALCAN
- **H.J. LATIMER** TAYLOR-WINFIELD CORPORATION
- **GLEN S. LAWRENCE** FERRANTI-SCIAKY COMPANY
- **KARL LAZAR**
- **WERNER LEHRHEUER** FORSCHUNGSZENTRUM JÜLICH GMBH
- **ALEXANDER LESNEWICH**
- **J.F. LIBSCH** LEPEL CORPORATION
- **TOM LIENERT** THE OHIO STATE UNIVERSITY
- **ALLEN C. LINGENFELTER** LAWRENCE LIVERMORE NATIONAL LABORATORY
- **DALE L. LINMAN** CENTECH CORPORATION
- **VONNE LINSE** EDISON WELDING INSTITUTE
- **JOHN C. LIPPOLD** EDISON WELDING INSTITUTE
- **JIAYAN LIU** AUBURN UNIVERSITY
- **STEPHEN LIU** COLORADO SCHOOL OF MINES
- **MATTHEW J. LUCAS, JR.** GENERAL ELECTRIC COMPANY
- **KEVIN A. LYTTLE** PRAXAIR INC.
- **KIM MAHIN** SANDIA NATIONAL LABORATORIES
- **MURRAY W. MAHONEY** ROCKWELL INTERNATIONAL SCIENCE CENTER
- **DARRELL MANENTE** VAC-AERO INTERNATIONAL INC.
- **RICHARD P. MARTUKANITZ** PENNSYLVANIA STATE UNIVERSITY
- **KOICHI MASUBUCHI** MASSACHUSETTS INSTITUTE OF TECHNOLOGY
- **DAVID K. MATLOCK** COLORADO SCHOOL OF MINES
- **R.B. MATTESON** TAYLOR-WINFIELD CORPORATION
- **STEVEN J. MATTHEWS** HAYNES INTERNATIONAL INC.
- **JYOTI MAZUMDER** UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
- **C.N. MCCOWAN** NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
- **KRIS MEEKINS** LONG MANUFACTURING LTD.
- **GREGORY MELEKIAN** GENERAL MOTORS CORPORATION
- **ANTHONY R. MELLINI, SR.** MELLINI AND ASSOCIATES INC.
- **DAVID W. MEYER** THE ESAB GROUP INC.
- **JULE MILLER**
- **HOWARD MIZUHARA** WESGO INC.
- **ARTHUR G. MOORHEAD** OAK RIDGE NATIONAL LABORATORY
- **MILO NANCE** MARTIN MARIETTA ASTRONAUTICS GROUP
- **E.D. NICHOLAS** THE WELDING INSTITUTE
- **DAVID NOBLE** ARCO EXPLORATION AND PRODUCTION TECHNOLOGY
- **THOMAS NORTH** UNIVERSITY OF TORONTO
- **DAVID B. O'DONNELL** INCO ALLOYS INTERNATIONAL INC.
- **JONATHAN S. OGBORN** THE LINCOLN ELECTRIC COMPANY
- **DAVID L. OLSON** COLORADO SCHOOL OF MINES
- **TOSHI OYAMA** WESGO INC.
- **R. ALAN PATTERSON** LOS ALAMOS NATIONAL LABORATORY
- **LARRY PERKINS** WRIGHT LABORATORY
- **DARYL PETER** DARYL PETER AND ASSOCIATES
- **MANFRED PETRI** GERHARD PETRI GMBH & CO. KG
- **DAVID H. PHILLIPS** EDISON WELDING INSTITUTE
- **ABE POLLACK** MICROALLOYING INTERNATIONAL INC.
- **BARRY POLLARD**
- **ANATOL RABINKIN** ALLIEDSIGNAL AMORPHOUS METALS

- **GEETHA RAMARATHNAM** UNIVERSITY OF TORONTO
- **EDWARD G. REINEKE** EXPLOSIVE FABRICATORS INC.
- **JULIAN ROBERTS** THERMATOOL CORPORATION
- **M. NED ROGERS** BATESVILLE CASKET COMPANY
- **J.R. ROPER** EG&G ROCKY FLATS PLANT
- **ROBERT S. ROSEN** LAWRENCE LIVERMORE NATIONAL LABORATORY
- **JAMES E. ROTH** JAMES E. ROTH INC.
- **WILLIAM J. RUPRECHT** GENERAL ELECTRIC COMPANY
- **K. SAMPATH** CONCURRENT TECHNOLOGIES CORPORATION
- **BERNARD E. SCHALTENBRAND** ALUMINUM COMPANY OF AMERICA
- **BERNARD SCHWARTZ** NORFOLK SOUTHERN CORPORATION
- **MEL M. SCHWARTZ** SIKORSKY AIRCRAFT
- **ANN SEVERIN** LUCAS-MILHAUPT INC.
- **THOMAS A. SIEWERT** NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
- **HERSCHEL SMARTT** IDAHO NATIONAL ENGINEERING LABORATORY
- **RONALD B. SMITH** ALLOY RODS CORPORATION
- **WARREN F. SMITH** THERMATOOL CORPORATION
- **LANCE R. SOISSON** WELDING CONSULTANTS INC.
- **HARVEY D. SOLOMON** GENERAL ELECTRIC COMPANY
- **BRUCE R. SOMERS** LEHIGH UNIVERSITY
- **ROBERT E. SOMERS** SOMERS CONSULTANTS
- **ROGER K. STEELE** AAR TECHNICAL CENTER
- **FRANK STEIN** TAYLOR-WINFIELD CORPORATION
- **TIM STOTLER** EDISON WELDING INSTITUTE
- **ROBERT L. STROHL** TWECO/ARCAIR
- **ROBERT A. SULIT** SULIT ENGINEERING
- **VERN SUTTER** AMERICAN WELDING INSTITUTE
- **W.T. TACK** MARTIN MARIETTA
- **R. DAVID THOMAS, JR.** R.D. THOMAS & COMPANY
- **KARL THOMAS** TECHNISCHE UNIVERSITÄT, BRAUNSCHWEIG
- **RAYMOND G. THOMPSON** UNIVERSITY OF ALABAMA AT BIRMINGHAM
- **DONALD J. TILLACK** D.J. TILLACK & ASSOCIATES
- **CHON L. TSAI** THE OHIO STATE UNIVERSITY
- **SCHILLINGS TSANG** EG&G ROCKY FLATS PLANT
- **HENDRIKUS H. VANDERVELDT** AMERICAN WELDING INSTITUTE
- **RICCARDO VANZETTI** VANZETTI SYSTEMS INC.
- **PAUL T. VIANCO** SANDIA NATIONAL LABORATORIES
- **P. RAVI VISHNU** LULEÅ UNIVERSITY OF TECHNOLOGY
- **MARY B. VOLLARO** UNIVERSITY OF CONNECTICUT
- **A. WAHID** COLORADO SCHOOL OF MINES
- **DANIEL W. WALSH** CALIFORNIA POLYTECHNIC STATE UNIVERSITY
- **R. TERRENCE WEBSTER** CONSULTANT
- **JOHN R. WHALEN** CONTOUR SAWS INC.
- **NEVILLE T. WILLIAMS** BRITISH STEEL
- **FRED J. WINSOR** WELDING CONSULTANT
- **R. XU** UNIVERSITY OF ILLINOIS AT CHICAGO
- **XIAOSHU XU** AMERICAN WELDING INSTITUTE
- **PHILIP M. ZARROW** SYNERGISTEK ASSOCIATES

REVIEWERS

- **YONI ADONYI** U.S. STEEL TECHNICAL CENTER
- **RICHARD L. ALLEY** AMERICAN WELDING SOCIETY
- **BERNARD ALTSHULLER** ALCAN INTERNATIONAL LTD.

- **TED L. ANDERSON** TEXAS A&M UNIVERSITY
- **LLOYD ANDERSON** MARION-INDRESCO INC.
- **FRANK G. ARMAO** ALCOA TECHNICAL CENTER
- **DANIEL ARTHUR** TELEDYNE MCKAY
- **RICHARD E. AVERY** NICKEL DEVELOPMENT INSTITUTE
- **R.F. BACON** TECUMSEH PRODUCTS COMPANY
- **WALLY G. BADER**
- **WILLIAM A. BAESLACK III** THE OHIO STATE UNIVERSITY
- **CLIFF C. BAMPTON** ROCKWELL INTERNATIONAL SCIENCE CENTER
- **JOHN G. BANKER** EXPLOSIVE FABRICATORS INC.
- **GEORGE C. BARNES**
- **ROBERT G. BARTIFAY** ALUMINUM COMPANY OF AMERICA
- **ROY E. BEAL** AMALGAMATED TECHNOLOGIES INC.
- **GARY BECKA** ALLIEDSIGNAL AEROSPACE COMPANY
- **DAN BEESON** EXXON PRODUCTION MALAYSIA
- **DAVID M. BENETEAU** CENTERLINE (WINDSOR) LTD.
- **CHRISTOPHER C. BERNDT** THE THERMAL SPRAY LABORATORY
- **SURENDRA BHARGAVA** GENERAL MOTORS INC.
- **NORMAN C. BINKLEY** EDISON WELDING INSTITUTE
- **ROBERT A. BISHEL** INCO ALLOYS INTERNATIONAL INC.
- **R.A. BLACK** BLACKS EQUIPMENT LTD.
- **OMER W. BLODGETT** THE LINCOLN ELECTRIC COMPANY
- **RICHARD A. BRAINARD** GENERAL DYNAMICS LAND SYSTEMS DIVISION
- **GLENN H. BRAVE** ASSOCIATION OF AMERICAN RAILROADS
- **ROBERT S. BROWN** CARPENTER TECHNOLOGY CORPORATION
- **WILLIAM A. BRUCE** EDISON WELDING INSTITUTE
- **CHUCK CADDEN** GENERAL MOTORS
- **HARVEY R. CASTNER** EDISON WELDING INSTITUTE
- **ALLEN B. CEDILOTE** INDUSTRIAL TESTING LABORATORY SERVICES CORPORATION
- **KENNETH D. CHALLENGER** SAN JOSE STATE UNIVERSITY
- **P.R. CHIDAMBARAM** COLORADO SCHOOL OF MINES
- **BOB CHRISTOFFEL**
- **ROBIN CHURCHILL** ESCO CORPORATION
- **MICHAEL J. CIESLAK** SANDIA NATIONAL LABORATORIES
- **BRADLEY A. CLEVELAND** MTS SYSTEMS CORPORATION
- **NANCY C. COLE** OAK RIDGE NATIONAL LABORATORY
- **HAROLD R. CONAWAY** ROCKWELL INTERNATIONAL
- **RICHARD B. CORBIT** GENERAL PUBLIC UTILITIES NUCLEAR CORPORATION
- **MARK COWELL** METCAL INC.
- **NORM COX** RESEARCH INC.
- **JOHN A. CRAWFORD** NAVAL SURFACE WARFARE CENTER
- **DENNIS D. CROCKETT** THE LINCOLN ELECTRIC COMPANY
- **CARL E. CROSS**
- **NARENDRA B. DAHOTRE** UNIVERSITY OF TENNESSEE SPACE INSTITUTE
- **T. DEBROY** PENNSYLVANIA STATE UNIVERSITY
- **JOSEPH DEVITO** THE ESAB GROUP INC.
- **JOHN A. DEVORE** GENERAL ELECTRIC COMPANY
- **PAUL B. DICKERSON**
- **RAY DIXON** LOS ALAMOS NATIONAL LABORATORY
- **KARL E. DORSCHU** WELDRING COMPANY INC.
- **ROBERT J. DYBAS** GENERAL ELECTRIC COMPANY
- **THOMAS W. EAGAR** MASSACHUSETTS INSTITUTE OF TECHNOLOGY
- **BRUCE J. EBERHARD** WESTINGHOUSE SAVANNAH RIVER COMPANY
- **GLEN R. EDWARDS** COLORADO SCHOOL OF MINES

- **JOHN W. ELMER** LAWRENCE LIVERMORE NATIONAL LABORATORY
- **WERNER ENGELMAIER** ENGELMAIER ASSOCIATES INC.
- **CHRIS ENGLISH** GE AIRCRAFT ENGINES
- **ROBERT G. FAIRBANKS** SCARROTT METALLURGICAL COMPANY
- **HOWARD N. FARMER** CONSULTANT
- **DAVID A. FLEMING** COLORADO SCHOOL OF MINES
- **ROBERT FOLEY** COLORADO SCHOOL OF MINES
- **BOBBY FOLKENING** FMC GROUND SYSTEMS DIVISION
- **DARREL FREAR** SANDIA NATIONAL LABORATORIES
- **MICHAEL D. FREDERICKSON** ELECTRONICS MANUFACTURING PRODUCTIVITY FACILITY
- **EUGENE R. FREULER** SOUDRONIC NEFTENBACH AG
- **STEVEN A. GEDEON** WELDING INSTITUTE OF CANADA
- **BOB GIBBONS** PLS MATERIALS INC.
- **PAUL S. GILMAN** ALLIEDSIGNAL INC.
- **STANLEY S. GLICKSTEIN** WESTINGHOUSE ELECTRIC CORPORATION
- **JOHN A. GOLDAK** CARLETON UNIVERSITY
- **CARL GRAF** EDISON WELDING INSTITUTE
- **WILLIAM L. GREEN** THE OHIO STATE UNIVERSITY
- **CHUCK GREGOIRE** NATIONAL STEEL CORPORATION
- **ROBERT A. GRIMM** EDISON WELDING INSTITUTE
- **BRIAN GRINSELL** THOMPSON WELDING INC.
- **ROBIN GROSS-GOURLEY** WESTINGHOUSE
- **JOHN F. GRUBB** ALLEGHENY LUDLUM STEEL
- **BOB GUNOW, JR.** VAC-MET INC.
- **C. HOWARD HAMILTON** WASHINGTON STATE UNIVERSITY
- **JAMES R. HANNAHS** PMI FOOD EQUIPMENT GROUP
- **FRANK HANNEY** ABCO WELDING & INDUSTRIAL SUPPLY INC.
- **DAVID E. HARDT** MASSACHUSETTS INSTITUTE OF TECHNOLOGY
- **IAN D. HARRIS** EDISON WELDING INSTITUTE
- **MARK J. HATZENBELLER** KRUEGER INTERNATIONAL
- **DAN HAUSER** EDISON WELDING INSTITUTE
- **C.R. HEIPLE** METALLURGICAL CONSULTANT
- **J.S. HETHERINGTON** HETHERINGTON INC.
- **BARRY S. HEUER** NOOTER CORPORATION
- **ROGER B. HIRSCH** UNITROL ELECTRONICS INC.
- **TIM P. HIRTHE** LUCAS-MILHAUPT
- **HUGH B. HIX** INTERNATIONAL EXPLOSIVE METALWORKING ASSOCIATION
- **F. GALEN HODGE** HAYNES INTERNATIONAL INC.
- **RICHARD L. HOLDREN** WELDING CONSULTANTS INC.
- **ALAN B. HOPPER** ROBERTSHAW TENNESSEE DIVISION
- **CHARLES HUTCHINS** C. HUTCHINS AND ASSOCIATES
- **JENNIE S. HWANG** IEM-FUSION INC.
- **S. IBARRA** AMOCO CORPORATION
- **J. ERNESTO INDACOCHEA** UNIVERSITY OF ILLINOIS AT CHICAGO
- **GARY IRONS** HOBART TAFA TECHNOLOGIES INC.
- **JAMES R. JACHNA** MODINE MANUFACTURING COMPANY
- **ROBERT G. JAITE** WOLFENDEN INDUSTRIES INC.
- **JOHN C. JENKINS** CONSULTANT
- **KATHI JOHNSON** HEXACON ELECTRIC COMPANY
- **WILLIAM R. JONES** VACUUM FURNACE SYSTEMS CORPORATION
- **ROBERT W. JUD** CHRYSLER CORPORATION
- **WILLIAM F. KAUKLER** UNIVERSITY OF ALABAMA IN HUNTSVILLE
- **DOUG D. KAUTZ** LAWRENCE LIVERMORE NATIONAL LABORATORY

- **W. DANIEL KAY** WALL COLMONOY CORPORATION
- **JACQUE KENNEDY** WESTINGHOUSE
- **JAMES F. KING** OAK RIDGE NATIONAL LABORATORY
- **ANDREW G. KIRETA** COPPER DEVELOPMENT ASSOCIATION INC.
- **SAMUEL D. KISER** INCO ALLOYS INTERNATIONAL INC.
- **JOSEPH H. KISSEL** ITT STANDARD
- **FRED KOHLER** CONSULTANT
- **M.L. KOHN** FMC CORPORATION
- **DAMIAN J. KOTECKI** THE LINCOLN ELECTRIC COMPANY
- **SINDO KOU** UNIVERSITY OF WISCONSIN-MADISON
- **CURTIS W. KOVACH** TECHNICAL MARKETING RESOURCES INC.
- **LAWRENCE S. KRAMER** MARTIN MARIETTA LABORATORIES
- **RAYMOND B. KRIEGER** AMERICAN CYANAMID COMPANY
- **KENNETH KRYSIAC** HERCULES INC.
- **DANIEL KURUZAR** MANUFACTURING TECHNOLOGY INC.
- **RICHARD A. LAFAVE** ELLIOTT COMPANY
- **FRANK B. LAKE** THE ESAB GROUP INC.
- **JOHN D. LANDES** UNIVERSITY OF TENNESSEE
- **WERNER LEHRHEUER** FORSCHUNGSZENTRUM JÜLICH GMBH
- **J.F. LIBSCH** LEPEL CORPORATION
- **VONNE LINSE** EDISON WELDING INSTITUTE
- **JOHN C. LIPPOLD** EDISON WELDING INSTITUTE
- **STEPHEN LIU** COLORADO SCHOOL OF MINES
- **RONALD LOEHMAN** ADVANCED MATERIALS LABORATORY
- **PAUL T. LOVEJOY** ALLEGHENY LUDLUM STEEL
- **GEORGE LUCEY** U.S. ARMY LABORATORY COMMAND
- **KEVIN A. LYTTLE** PRAXAIR INC.
- **COLIN MACKAY** MICROELECTRONICS AND COMPUTER TECHNOLOGY CORPORATION
- **MICHAEL C. MAGUIRE** SANDIA NATIONAL LABORATORIES
- **KIM W. MAHIN** SANDIA NATIONAL LABORATORIES
- **WILLIAM E. MANCINI** DUPONT
- **DARRELL MANENTE** VAC-AERO INTERNATIONAL INC.
- **AUGUST F. MANZ** A.F. MANZ ASSOCIATES
- **RICHARD P. MARTUKANITZ** PENNSYLVANIA STATE UNIVERSITY
- **KOICHI MASUBUCHI** MASSACHUSETTS INSTITUTE OF TECHNOLOGY
- **STEVEN J. MATTHEWS** HAYNES INTERNATIONAL
- **JYOTI MAZUMDER** UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
- **JIM MCMAHON** DOALL COMPANY
- **ALAN MEIER** COLORADO SCHOOL OF MINES
- **STANLEY MERRICK** TELEDYNE MCKAY
- **ROBERT W. MESSLER, JR.** RENSSELAER POLYTECHNIC INSTITUTE
- **E.A. METZBOWER** U.S. NAVAL RESEARCH LABORATORY
- **JOEL MILANO** DAVID TAYLOR MODEL BASIN
- **ROBERT A. MILLER** SULZER PLASMA TECHNIK INC.
- **HERBERT W. MISHLER** EDISON WELDING INSTITUTE
- **BRAJENDRA MISHRA** COLORADO SCHOOL OF MINES
- **HOWARD MIZUHARA** WESGO INC.
- **RICHARD MONTANA** MID-FLORIDA TECHNICAL INSTITUTE
- **JERRY MOODY** WORLD WIDE WELDING
- **RICHARD A. MORRIS** NAVAL SURFACE WARFARE CENTER
- **P.J. MUDGE** THE WELDING INSTITUTE
- **AMIYA MUKHERJEE** UNIVERSITY OF CALIFORNIA
- **BILL MYERS** DRESSER-RAND INC.

- **ERNEST F. NIPPES** CONSULTANT
- **DONG WON OH** COLORADO SCHOOL OF MINES
- **DAVID L. OLSON** COLORADO SCHOOL OF MINES
- **EDGAR D. OPPENHEIMER** CONSULTING ENGINEER
- **CARMEN PAPONETTI** HI TECMETAL GROUP INC.
- **MADHU PAREKH** HOBART BROTHERS COMPANY
- **SUBHASH R. PATI** INTERNATIONAL PAPER COMPANY
- **R. ALAN PATTERSON** LOS ALAMOS NATIONAL LABORATORIES
- **CHARLES C. PEASE** CP METALLURGICAL
- **ROBERT LEON PEASLEE** WALL COLMONOY CORPORATION
- **DARYL PETER** DARYL PETER & ASSOCIATES
- **LORENZ PFEIFER**
- **JOHN F. PFLZNIENSKI** KOLENE CORPORATION
- **DAVID H. PHILLIPS** EDISON WELDING INSTITUTE
- **EARL W. PICKERING, JR.** CONSULTANT
- **E.R. PIERRE** CONSULTING WELDING ADVISOR
- **JOHN PILLING** MICHIGAN TECHNOLOGICAL UNIVERSITY
- **ABE POLLACK** MICROALLOYING INTERNATIONAL INC.
- **BARRY POLLARD**
- **ALEXANDRE M. POPE** COLORADO SCHOOL OF MINES
- **JEFFREY W. POST** J.W. POST & ASSOCIATES INC.
- **TERRY PROFUGHI** HI TECMETAL GROUP INC.
- **ANATOL RABINKIN** ALLIEDSIGNAL AMORPHOUS METALS
- **JIM D. RABY** SOLDERING TECH INTERNATIONAL
- **TED RENSHAW** CONSULTANT
- **THERESA ROBERTS** SIKAMA INTERNATIONAL
- **DAVID E. ROBERTSON** PACE INC.
- **CHARLES ROBINO** SANDIA NATIONAL LABORATORIES
- **M.N. ROGERS** ABB POWER T&D COMPANY INC.
- **J.R. ROPER** EG&G ROCKY FLATS PLANT
- **N.V. ROSS** AJAX MAGNETHERMIC
- **DIETRICH K. ROTH** ROMAN MANUFACTURING INC.
- **JOHN RUFFING** 3M FLUIDS LABORATORY
- **WAYNE D. RUPERT** ENGLEHARD CORPORATION
- **J.D. RUSSELL** THE WELDING INSTITUTE
- **C.O. RUUD** PENNSYLVANIA STATE UNIVERSITY
- **EDMUND F. RYBICKI** UNIVERSITY OF TULSA
- **JONATHAN T. SALKIN** ARC APPLICATIONS INC.
- **MEL M. SCHWARTZ** SIKORSKY AIRCRAFT
- **JOE L. SCOTT** DEVASCO INTERNATIONAL INC.
- **ALAN P. SEIDLER** RMI TITANIUM COMPANY
- **OSCAR W. SETH** CHICAGO BRIDGE & IRON COMPANY
- **ANN SEVERIN** LUCAS-MILHAUPT INC.
- **LEWIS E. SHOEMAKER** INCO ALLOYS INTERNATIONAL INC.
- **LYNN E. SHOWALTER** NEWPORT NEWS SHIPBUILDING
- **THOMAS A. SIEWERT** NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
- **ALLEN W. SINDEL** BEGEMANN HEAVY INDUSTRIES INC.
- **MICHAEL H. SKILLINGBERG** REYNOLDS METALS COMPANY
- **GERALD M. SLAUGHTER** OAK RIDGE NATIONAL LABORATORY
- **HERSCHEL SMARTT** IDAHO NATIONAL ENGINEERING LABORATORY
- **JAMES P. SNYDER II** BETHLEHEM STEEL CORPORATION
- **LANCE R. SOISSON** WELDING CONSULTANTS INC.
- **HARVEY D. SOLOMON** GENERAL ELECTRIC
- **BRUCE R. SOMERS** LEHIGH UNIVERSITY

- **NARASI SRIDHAR** SOUTHWEST RESEARCH INSTITUTE
- **BOB STANLEY** NATIONAL TRAINING FUND
- **ROGER K. STEELE** ASSOCIATION OF AMERICAN RAILROADS
- **ARCHIE STEVENSON** MAGNESIUM ELEKRON INC.
- **VIJAY K. STOKES** GENERAL ELECTRIC
- **TIM STOTLER** EDISON WELDING INSTITUTE
- **M.A. STREICHER** CONSULTANT
- **ROBERT L. STROHL** TWECO/ARCAIR
- **LAWRENCE STRYKER** ALTECH INTERNATIONAL
- **MARK TARBY** WALL COLMONOY CORPORATION
- **CLAY TAYLOR** MERRICK AND COMPANY
- **J.R. TERRILL** CONSULTANT
- **RAYMOND G. THOMPSON** UNIVERSITY OF ALABAMA AT BIRMINGHAM
- **J.S. THROWER** GENERAL ELECTRIC POWER GENERATION
- **DONALD J. TILLACK** D.J. TILLACK & ASSOCIATES
- **FELIX TOMEI** TRUMPF INC.
- **CHON L. TSAI** THE OHIO STATE UNIVERSITY
- **SCHILLINGS TSANG** EG&G ROCKY FLATS PLANT
- **M. NASIM UDDIN** THYSSEN STEEL GROUP
- **ELMAR UPITIS** CBI TECHNICAL SERVICES COMPANY
- **JAMES VAN DEN AVYLE** SANDI NATIONAL LABORATORIES
- **CLARENCE VAN DYKE** LUCAS-MIHAUPT INC.
- **HENDRIKUS H. VANDERVELDT** AMERICAN WELDING INSTITUTE
- **DAVID B. VEVERKA** EDISON WELDING INSTITUTE
- **PAUL T. VIANCO** SANDIA NATIONAL LABORATORIES
- **ROBERT G. VOLLMER**
- **R. WALLACH** UNIVERSITY OF CAMBRIDGE
- **SANDRA J. WALMSLEY** WESTINGHOUSE ELECTRIC CORPORATION
- **RICHARD A. WATSON** THE P&LE CAR COMPANY
- **CHRIS WEHLUS** GENERAL MOTORS
- **C.E.T. WHITE** INDIUM CORPORATION OF AMERICA
- **ROGER N. WILD**
- **ELLIOTT WILLNER** LOCKHEED MISSILES & SPACE COMPANY
- **RICHARD WILSON** HOUSTON LIGHTING AND POWER COMPANY
- **W.L. WINTERBOTTOM** FORD MOTOR COMPANY
- **A.P. WOODFIELD** GENERAL ELECTRIC AIRCRAFT ENGINES
- **JAMES B.C. WU** STOODY COMPANY
- **THOMAS ZACHARIA** OAK RIDGE NATIONAL LABORATORY

FOREWORD

COVERAGE OF JOINING TECHNOLOGIES IN THE *ASM HANDBOOK* HAS GROWN DRAMATICALLY OVER THE YEARS. A SHORT CHAPTER ON WELDING--EQUAL IN SIZE TO ABOUT 5 PAGES OF TODAY'S *ASM HANDBOOK*--APPEARED IN THE 1933 EDITION OF THE *NATIONAL METALS HANDBOOK* PUBLISHED BY THE AMERICAN SOCIETY OF STEEL TREATERS, ASM'S PREDECESSOR. THAT MATERIAL WAS EXPANDED TO 13 PAGES IN THE CLASSIC 1948 EDITION OF *METALS HANDBOOK.* THE FIRST FULL VOLUME ON WELDING AND BRAZING IN THE SERIES APPEARED IN 1971, WITH PUBLICATION OF VOLUME 6 OF THE 8TH EDITION OF *METALS HANDBOOK*. VOLUME 6 OF THE 9TH EDITION, PUBLISHED IN 1983, WAS EXPANDED TO INCLUDE COVERAGE OF SOLDERING.

THE NEW VOLUME 6 OF THE *ASM HANDBOOK* BUILDS ON THE PROUD TRADITION ESTABLISHED BY THESE PREVIOUS VOLUMES, BUT IT ALSO REPRESENTS A BOLD NEW STEP FOR THE SERIES. THE HANDBOOK HAS NOT ONLY BEEN REVISED, BUT ALSO ENTIRELY

REFORMATTED TO MEET THE NEEDS OF TODAY'S MATERIALS COMMUNITY. OVER 90% OF THE ARTICLES IN THIS VOLUME ARE BRAND-NEW, AND THE REMAINDER HAVE BEEN SUBSTANTIALLY REVISED. MORE SPACE HAS BEEN DEVOTED TO COVERAGE OF SOLID-STATE WELDING PROCESSES, MATERIALS SELECTION FOR JOINED ASSEMBLIES, WELDING IN SPECIAL ENVIRONMENTS, QUALITY CONTROL, AND MODELING OF JOINING PROCESSES, TO NAME BUT A FEW. INFORMATION ALSO HAS BEEN ADDED FOR THE FIRST TIME ABOUT JOINING OF SELECTED NONMETALLIC MATERIALS.

WHILE A DELIBERATE ATTEMPT HAS BEEN MADE TO INCREASE THE AMOUNT OF CUTTING-EDGE INFORMATION PROVIDED, THE ORGANIZERS HAVE WORKED HARD TO ENSURE THAT THE HEART OF THE BOOK REMAINS PRACTICAL INFORMATION ABOUT JOINING PROCESSES, APPLICATIONS, AND MATERIALS WELDABILITY--THE TYPE OF INFORMATION THAT IS THE HALLMARK OF THE *ASM HANDBOOK* SERIES.

PUTTING TOGETHER A VOLUME OF THIS MAGNITUDE IS AN ENORMOUS EFFORT AND COULD NOT HAVE BEEN ACCOMPLISHED WITHOUT THE DEDICATED AND TIRELESS EFFORTS OF THE VOLUME CHAIRPERSONS: DAVID L. OLSON, THOMAS A. SIEWERT, STEPHEN LIU, AND GLEN R. EDWARDS. SPECIAL THANKS ARE ALSO DUE TO THE SECTION CHAIRPERSONS, TO THE MEMBERS OF THE ASM HANDBOOK COMMITTEE, AND TO THE ASM EDITORIAL STAFF. WE ARE ESPECIALLY GRATEFUL TO THE OVER 400 AUTHORS AND REVIEWERS WHO HAVE CONTRIBUTED THEIR TIME AND EXPERTISE IN ORDER TO MAKE THIS HANDBOOK A TRULY OUTSTANDING INFORMATION RESOURCE.

EDWARD H. KOTTCAMP, JR.
PRESIDENT
ASM INTERNATIONAL

EDWARD L. LANGER
MANAGING DIRECTOR
ASM INTERNATIONAL

PREFACE

THE *ASM HANDBOOK*, VOLUME 6, *WELDING, BRAZING, AND SOLDERING*, HAS BEEN ORGANIZED INTO A UNIQUE FORMAT THAT WE BELIEVE WILL PROVIDE HANDBOOK USERS WITH READY ACCESS TO NEEDED MATERIALS-ORIENTED JOINING INFORMATION AT A MINIMAL LEVEL OF FRUSTRATION AND STUDY TIME. WHEN WE DEVELOPED THE ORGANIZATIONAL STRUCTURE FOR THIS VOLUME, WE RECOGNIZED THAT ENGINEERS, TECHNICIANS, RESEARCHERS, DESIGNERS, STUDENTS, AND TEACHERS DO NOT SEEK OUT JOINING INFORMATION WITH THE SAME LEVEL OF UNDERSTANDING, OR WITH THE SAME NEEDS. THEREFORE, WE ESTABLISHED DISTINCT SECTIONS THAT WERE INTENDED TO MEET THE SPECIFIC NEEDS OF PARTICULAR USERS.

THE EXPERIENCED JOINING SPECIALIST CAN TURN TO THE SECTION "CONSUMABLE SELECTION, PROCEDURE DEVELOPMENT, AND PRACTICE CONSIDERATIONS" AND FIND DETAILED JOINING MATERIALS DATA ON A WELL-DEFINED PROBLEM. THIS HANDBOOK ALSO PROVIDES GUIDANCE FOR THOSE WHO NOT ONLY MUST SPECIFY THE JOINING PRACTICE, BUT ALSO THE MATERIALS TO BE JOINED. THE SECTION "MATERIALS SELECTION FOR JOINED ASSEMBLIES" CONTAINS COMPREHENSIVE INFORMATION ABOUT THE PROPERTIES, APPLICATIONS, AND WELDABILITIES OF THE MAJOR CLASSES OF STRUCTURAL MATERIALS. TOGETHER, THESE TWO MAJOR SECTIONS OF THE HANDBOOK SHOULD PROVIDE AN ENGINEER ASSIGNED A LOOSELY DEFINED DESIGN PROBLEM WITH THE MEANS TO MAKE INTELLIGENT CHOICES FOR COMPLETING AN ASSEMBLY.

FREQUENTLY, TECHNOLOGISTS ARE CALLED UPON TO INITIATE AND ADOPT WELDING PROCESSES WITHOUT IN-DEPTH KNOWLEDGE OF THESE PROCESSES OR THE SCIENTIFIC

PRINCIPLES THAT IMPACT THE PROPERTIES AND PERFORMANCE OF WELDMENTS. THE SECTIONS "FUNDAMENTALS OF JOINING" AND "JOINING PROCESSES" ARE DESIGNED TO MEET THE NEEDS OF THESE USERS, OR ANYONE WHO NEEDS BASIC BACKGROUND INFORMATION ABOUT JOINING PROCESSES AND PRINCIPLES.

WELDING, BRAZING, AND SOLDERING ARE TRULY INTERDISCIPLINARY ENTERPRISES; NO INDIVIDUAL CAN BE EXPECTED TO BE AN EXPERT IN ALL ASPECTS OF THESE TECHNOLOGIES. THEREFORE, WE HAVE ATTEMPTED TO PROVIDE A HANDBOOK THAT CAN BE USED AS A COMPREHENSIVE REFERENCE BY ANYONE NEEDING MATERIALS-RELATED JOINING INFORMATION.

MANY COLLEAGUES AND FRIENDS CONTRIBUTED THEIR TIME AND EXPERTISE TO THIS HANDBOOK, AND WE ARE VERY GRATEFUL FOR THEIR EFFORTS. WE WOULD ALSO LIKE TO EXPRESS OUR THANKS TO THE AMERICAN WELDING SOCIETY FOR THEIR COOPERATION AND ASSISTANCE IN THIS ENDEAVOR.

DAVID LEROY OLSON, COLORADO SCHOOL OF MINES

THOMAS A. SIEWERT, NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY

STEPHEN LIU, COLORADO SCHOOL OF MINES

GLEN R. EDWARDS, COLORADO SCHOOL OF MINES

OFFICERS AND TRUSTEES OF ASM INTERNATIONAL (1992-1993)

OFFICERS

- **EDWARD H. KOTTCAMP, JR.** PRESIDENT AND TRUSTEE SPS TECHNOLOGIES
- **JACK G. SIMON** VICE PRESIDENT AND TRUSTEE GENERAL MOTORS CORPORATION
- **WILLIAM P. KOSTER** IMMEDIATE PAST PRESIDENT AND TRUSTEE METCUT RESEARCH ASSOCIATES, INC.
- **EDWARD L. LANGER** SECRETARY AND MANAGING DIRECTOR ASM INTERNATIONAL
- **LEO G. THOMPSON** TREASURER LINDBERG CORPORATION

TRUSTEES

- **WILLIAM H. ERICKSON** FDP ENGINEERING
- **NORMAN A. GJOSTEIN** FORD MOTOR COMPANY
- **NICHOLAS C. JESSEN, JR.** MARTIN MARIETTA ENERGY SYSTEMS, INC.
- **E. GEORGE KENDALL** NORTHROP AIRCRAFT
- **GEORGE KRAUSS** COLORADO SCHOOL OF MINES
- **LYLE H. SCHWARTZ** NATIONAL INSTITUTE OF STANDARDS & TECHNOLOGY
- **GERNANT E. MAURER** SPECIAL METALS CORPORATION
- **ALTON D. ROMIG, JR.** SANDIA NATIONAL LABORATORIES
- **MERLE L. THORPE** HOBART TAFA TECHNOLOGIES, INC.

MEMBERS OF THE ASM HANDBOOK COMMITTEE (1992-1993)

- **ROGER J. AUSTIN** (CHAIRMAN 1992-; MEMBER 1984-) CONCEPT SUPPORT AND DEVELOPMENT CORPORATION
- **DAVID V. NEFF** (VICE CHAIRMAN 1992-; MEMBER 1986-) METAULLICS SYSTEMS
- **TED L. ANDERSON** (1991-) TEXAS A&M UNIVERSITY
- **BRUCE P. BARDES** (1993-) MIAMI UNIVERSITY

- **ROBERT J. BARNHURST** (1988-) NORANDA TECHNOLOGY CENTRE
- **TONI BRUGGER** (1993-) PHOENIX PIPE & TUBE COMPANY
- **STEPHEN J. BURDEN** (1989-)
- **CRAIG V. DARRAGH** (1989-) THE TIMKEN COMPANY
- **RUSSELL J. DIEFENDORF** (1990-) CLEMSON UNIVERSITY
- **AICHA EISHABINI-RIAD** (1990-) VIRGINIA POLYTECHNIC & STATE UNIVERSITY
- **GREGORY A. FETT** (1993-) DANA CORPORATION
- **MICHELLE M. GAUTHIER** (1990-) RAYTHEON COMPANY
- **TONI GROBSTEIN** (1990-) NASA LEWIS RESEARCH CENTER
- **SUSAN HOUSH** (1990-) DOW CHEMICAL U.S.A.
- **DENNIS D. HUFFMAN** (1982-) THE TIMKEN COMPANY
- **S. JIM LBARRA** (1991-) AMOCO RESEARCH CENTER
- **J. ERNESTO INDACOCHEA** (1987-) UNIVERSITY OF ILLINOIS AT CHICAGO
- **PETER W. LEE** (1990-) THE TIMKEN COMPANY
- **WILLIAM L. MANKINS** (1989-) INCO ALLOYS INTERNATIONAL, INC.
- **RICHARD E. ROBERTSON** (1990-) UNIVERSITY OF MICHIGAN
- **JOGENDER SINGH** (1993-) NASA GEORGE C. MARSHALL SPACE FLIGHT CENTER
- **JEREMY C. ST. PIERRE** (1990-) HAYES HEAT TREATING CORPORATION
- **EPHRAIM SUHIR** (1990-) AT&T BELL LABORATORIES
- **KENNETH TATOR** (1991-) KTA-TATOR, INC.
- **MALCOLM THOMAS** (1993-) ALLISON GAS TURBINES
- **WILLIAM B. YOUNG** (1991-) DANA CORPORATION

PREVIOUS CHAIRMEN OF THE ASM HANDBOOK COMMITTEE

- **R.S. ARCHER** (1940-1942) (MEMBER 1937-1942)
- **L.B. CASE** (1931-1933) (MEMBER 1927-1933)
- **T.D. COOPER** (1984-1986) (MEMBER 1981-1986)
- **E.O. DIXON** (1952-1954) (MEMBER 1947-1955)
- **R.L. DOWDELL** (1938-1939) (MEMBER 1935-1939)
- **J.P. GILL** (1937) (MEMBER 1934-1937)
- **J.D. GRAHAM** (1966-1968) (MEMBER 1961-1970)
- **J.F. HARPER** (1923-1926) (MEMBER 1923-1926)
- **C.H. HERTY, JR.** (1934-1936) (MEMBER 1930-1936)
- **D.D. HUFFMAN** (1986-1990) (MEMBER 1982-1990)
- **J.B. JOHNSON** (1948-1951) (MEMBER 1944-1951)
- **L.J. KORB** (1983) (MEMBER 1978-1983)
- **R.W.E. LEITER** (1962-1963) (MEMBER 1955-1958, 1960-1964)
- **G.V. LUERSSEN** (1943-1947) (MEMBER 1942-1947)
- **G.N. MANIAR** (1979-1980) (MEMBER 1974-1980)
- **J.L. MCCALL** (1982) (MEMBER 1977-1982)
- **W.J. MERTEN** (1927-1930) (MEMBER 1923-1933)
- **D.L. OLSON** (1990-1992) (MEMBER 1982-1988, 1989-1992)
- **N.E. PROMISEL** (1955-1961) (MEMBER 1954-1963)
- **G.J. SHUBAT** (1973-1975) (MEMBER 1966-1975)
- **W.A. STADTLER** (1969-1972) (MEMBER 1962-1972)
- **R. WARD** (1976-1978) (MEMBER 1972-1978)
- **M.G.H. WELLS** (1981) (MEMBER 1976-1981)
- **D.J. WRIGHT** (1964-1965) (MEMBER 1959-1967)

STAFF

ASM INTERNATIONAL STAFF WHO CONTRIBUTED TO THE DEVELOPMENT OF THE VOLUME INCLUDED WILLIAM W. SCOTT, JR., DIRECTOR OF TECHNICAL PUBLICATIONS; SCOTT D.

HENRY, MANAGER OF HANDBOOK DEVELOPMENT; SUZANNE E. HAMPSON, PRODUCTION PROJECT MANAGER; THEODORE B. ZORC, TECHNICAL EDITOR; FAITH REIDENBACH, CHIEF COPY EDITOR; LAURIE A. HARRISON, EDITORIAL ASSISTANT; NANCY M. SOBIE, PRODUCTION ASSISTANT. EDITORIAL ASSISTANCE WAS PROVIDED BY JOSEPH R. DAVIS, KELLY FERJUTZ, NIKKI D. WHEATON, AND MARA S. WOODS.

CONVERSION TO ELECTRONIC FILES

ASM HANDBOOK, VOLUME 6, WELDING, BRAZING, AND SOLDERING WAS CONVERTED TO ELECTRONIC FILES IN 1998. THE CONVERSION WAS BASED ON THE SECOND PRINTING (1994). NO SUBSTANTIVE CHANGES WERE MADE TO THE CONTENT OF THE VOLUME, BUT SOME MINOR CORRECTIONS AND CLARIFICATIONS WERE MADE AS NEEDED.

ASM INTERNATIONAL STAFF WHO CONTRIBUTED TO THE CONVERSION OF THE VOLUME INCLUDED SALLY FAHRENHOLZ-MANN, BONNIE SANDERS, SCOTT HENRY, ROBERT BRADDOCK, AND MARLENE SEUFFERT. THE ELECTRONIC VERSION WAS PREPARED UNDER THE DIRECTION OF WILLIAM W. SCOTT, JR., TECHNICAL DIRECTOR, AND MICHAEL J. DEHAEMER, MANAGING DIRECTOR.

COPYRIGHT INFORMATION (FOR PRINT VOLUME)

COPYRIGHT © 1993 BY ASM INTERNATIONAL

ALL RIGHTS RESERVED.

ASM HANDBOOK IS A COLLECTIVE EFFORT INVOLVING THOUSANDS OF TECHNICAL SPECIALISTS. IT BRINGS TOGETHER IN ONE BOOK A WEALTH OF INFORMATION FROM WORLD-WIDE SOURCES TO HELP SCIENTISTS, ENGINEERS, AND TECHNICIANS SOLVE CURRENT AND LONG-RANGE PROBLEMS.

GREAT CARE IS TAKEN IN THE COMPILATION AND PRODUCTION OF THIS VOLUME, BUT IT SHOULD BE MADE CLEAR THAT NO WARRANTIES, EXPRESS OR IMPLIED, ARE GIVEN IN CONNECTION WITH THE ACCURACY OR COMPLETENESS OF THIS PUBLICATION, AND NO RESPONSIBILITY CAN BE TAKEN FOR ANY CLAIMS THAT MAY ARISE.

NOTHING CONTAINED IN THE ASM HANDBOOK SHALL BE CONSTRUED AS A GRANT OF ANY RIGHT OF MANUFACTURE, SALE, USE, OR REPRODUCTION, IN CONNECTION WITH ANY METHOD, PROCESS, APPARATUS, PRODUCT, COMPOSITION, OR SYSTEM, WHETHER OR NOT COVERED BY LETTERS PATENT, COPYRIGHT, OR TRADEMARK, AND NOTHING CONTAINED IN THE ASM HANDBOOK SHALL BE CONSTRUED AS A DEFENSE AGAINST ANY ALLEGED INFRINGEMENT OF LETTERS PATENT, COPYRIGHT, OR TRADEMARK, OR AS A DEFENSE AGAINST LIABILITY FOR SUCH INFRINGEMENT.

COMMENTS, CRITICISMS, AND SUGGESTIONS ARE INVITED, AND SHOULD BE FORWARDED TO ASM INTERNATIONAL.

LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA (FOR PRINT VOLUME)

ASM HANDBOOK (REVISED VOL. 6) METALS HANDBOOK. VOLS. 1-2 HAVE TITLE:

METALS HANDBOOK. VOL. 4 LACKS ED. STATEMENTS. INCLUDES BIBLIOGRAPHICAL REFERENCES AND INDEXES. CONTENTS: V. 1. PROPERTIES AND SELECTION-IRONS, STEELS, AND HIGH-PERFORMANCE ALLOYS-V. 2. PROPERTIES AND SELECTION-NONFERROUS ALLOYS

AND SPECIAL-PURPOSE MATERIALS-[ETC.]-V. 6. WELDING, BRAZING, AND SOLDERING. 1. METALS-HANDBOOKS, MANUALS, ETC. 2. METAL-WORK-HANDBOOKS, MANUALS, ETC. I. ASM INTERNATIONAL. HANDBOOK COMMITTEE. II. TITLE: METALS HANDBOOK.

TA459.M43 1990 620.1'6 90-115

ISBN 0-87170-377-7(V.1)

SAN 204-7586 ISBN 0-87170-382-3

PRINTED IN THE UNITED STATES OF AMERICA

Energy Sources Used for Fusion Welding

Thomas W. Eagar, Massachusetts Institute of Technology

Introduction

WELDING AND JOINING processes are essential for the development of virtually every manufactured product. However, these processes often appear to consume greater fractions of the product cost and to create more of the production difficulties than might be expected. There are a number of reasons that explain this situation.

First, welding and joining are multifaceted, both in terms of process variations (such as fastening, adhesive bonding, soldering, brazing, arc welding, diffusion bonding, and resistance welding) and in the disciplines needed for problem solving (such as mechanics, materials science, physics, chemistry, and electronics). An engineer with unusually broad and deep training is required to bring these disciplines together and to apply them effectively to a variety of processes.

Second, welding or joining difficulties usually occur far into the manufacturing process, where the relative value of scrapped parts is high.

Third, a very large percentage of product failures occur at joints because they are usually located at the highest stress points of an assembly and are therefore the weakest parts of that assembly. Careful attention to the joining processes can produce great rewards in manufacturing economy and product reliability.

The Section "Fusion Welding Processes" in this Volume provides details about equipment and systems for the major fusion welding processes. The purpose of this Section of the Volume is to discuss the fundamentals of fusion welding processes, with an emphasis on the underlying scientific principles.

Because there are many fusion welding processes, one of the greatest difficulties for the manufacturing engineer is to determine which process will produce acceptable properties at the lowest cost. There are no simple answers. Any change in the part geometry, material, value of the end product, or size of the production run, as well as the availability of joining equipment, can influence the choice of joining method. For small lots of complex parts, fastening may be preferable to welding, whereas for long production runs, welds can be stronger and less expensive.

The perfect joint is indistinguishable from the material surrounding it. Although some processes, such as diffusion bonding, can achieve results that are very close to this ideal, they are either expensive or restricted to use with just a few materials. There is no universal process that performs adequately on all materials in all geometries. Nevertheless, virtually any material can be joined in some way, although joint properties equal to those of the bulk material cannot always be achieved.

The economics of joining a material may limit its usefulness. For example, aluminum is used extensively in aircraft manufacturing and can be joined by using adhesives or fasteners, or by welding. However, none of these processes has proven economical enough to allow the extensive replacement of steel by aluminum in the frames of automobiles. An increased use of composites in aircrafts is limited by an inability to achieve adequate joint strength.

It is essential that the manufacturing engineer work with the designer from the point of product conception to ensure that compatible materials, processes, and properties are selected for the final assembly. Often, the designer leaves the problem of joining the parts to the manufacturing engineer. This can cause an escalation in cost and a decrease in reliability. If the design has been planned carefully and the parts have been produced accurately, the joining process becomes much easier and cheaper, and both the quality and reliability of the product are enhanced.

Generally, any two solids will bond if their surfaces are brought into intimate contact. One factor that generally inhibits this contact is surface contamination. Any freshly produced surface exposed to the atmosphere will absorb oxygen, water vapor, carbon dioxide, and hydrocarbons very rapidly. If it is assumed that each molecule that hits the surface will be absorbed, then the time-pressure value to produce a monolayer of contamination is approximately 0.001 Pa · s (10^{-8} atm · s). For example, at a pressure of 1 Pa (10^{-5} atm), the contamination time is 10^{-3} s, whereas at 0.1 MPa (1 atm), it is only 10 × 10^{-9} s.

In fusion welding, intimate interfacial contact is achieved by interposing a liquid of substantially similar composition as the base metal. If the surface contamination is soluble, then it is dissolved in the liquid. If it is insoluble, then it will float away from the liquid-solid interface.

Energy Sources Used for Fusion Welding

Thomas W. Eagar, Massachusetts Institute of Technology

Energy-Source Intensity

One distinguishing feature of all fusion welding processes is the intensity of the heat source used to melt the liquid. Virtually every concentrated heat source has been applied to the welding process. However, many of the characteristics of each type of heat source are determined by its intensity. For example, when considering a planar heat source diffusing into a very thick slab, the surface temperature will be a function of both the surface power density and the time.

Figure 1 shows how this temperature will vary on steel with power densities that range from 400 to 8000 W/cm^2. At the lower value, it takes 2 min to melt the surface. If that heat source were a point on the flat surface, then the heat flow would be divergent and might not melt the steel. Rather, the solid metal would be able to conduct away the heat as fast as it was being introduced. It is generally found that heat-source power densities of approximately 1000 W/cm^2 are necessary to melt most metals.

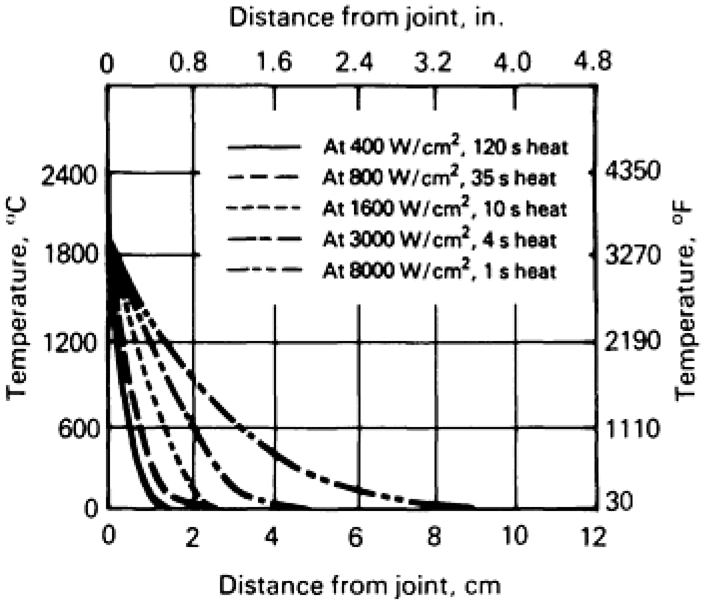

FIG. 1 TEMPERATURE DISTRIBUTION AFTER A SPECIFIC HEATING TIME IN A THICK STEEL PLATE HEATED

UNIFORMLY ON ONE SURFACE AS A FUNCTION OF APPLIED HEAT INTENSITY; INITIAL TEMPERATURE OF PLATE IS 25 °C (77 °F)

At the other end of the power-density spectrum, heat intensities of 10^6 or 10^7 W/cm^2 will vaporize most metals within a few microsecond. At levels above these values, all of the solid that interacts with the heat source will be vaporized, and no fusion welding can occur. Thus, the heat sources for all fusion welding processes should have power densities between approximately 0.001 and 1 MW/cm^2. This power-density spectrum is shown in Fig. 2, along with the points at which common joining processes are employed.

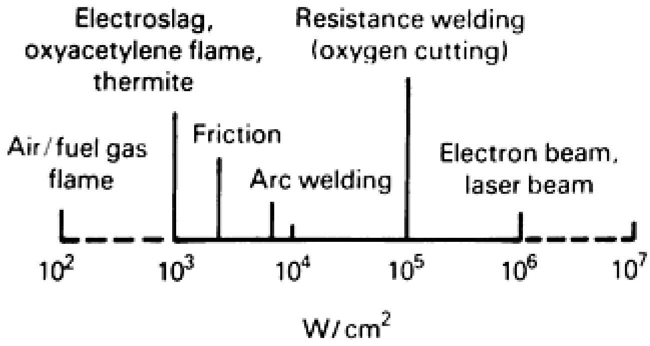
Electroslag, oxyacetylene flame, thermite
Resistance welding (oxygen cutting)
Friction
Air/fuel gas flame
Arc welding
Electron beam, laser beam
10^2
10^3
10^4
10^5
10^6
10^7
W/cm^2

FIG. 2 SPECTRUM OF PRACTICAL HEAT INTENSITIES USED FOR FUSION WELDING

The fact that power density is inversely related to the interaction time of the heat source on the material is evident in Fig. 1. Because this represents a transient heat conduction problem, one can expect the heat to diffuse into the steel to a depth that increases as the square root of time, that is, from the Einstein equation:

$$X \sim \sqrt{a t} \qquad \textbf{(EQ 1)}$$

where *x* is the distance that the heat diffuses into the solid, in centimeters: α is the thermal diffusivity of the solid, in cm^2/s; and *t* is the time in seconds. Tables 1 and 2 give the thermal diffusivities of common elements and common alloys, respectively.

TABLE 1 THERMAL DIFFUSIVITIES OF COMMON ELEMENTS FROM 20 TO 100 °C (68 TO 212 °F)

ELEMENT	DENSITY		HEAT CAPACITY		THERMAL CONDUCTIVITY		mm^2/s	THERMAL DIFFUSIVITY cm^2/s
	g/cm^3	lb/$in.^3$	j/kg · k	cal_{it}/g · °c	w/m · k	cal_{it}/cm · s · °c		
ALUMINUM	2.699	0.098	900	0.215	221	0.53	91	0.91
ANTIMONY	6.62	0.239	205	0.049	19	0.045	14	0.14
BERYLLIUM	1.848	0.067	1880	0.45	147	0.35	42	0.42
BISMUTH	9.80	0.354	123	0.0294	8	0.020	7	0.09
CADMIUM	8.65	0.313	230	0.055	92	0.22	46	0.46
CARBON	2.25	0.081	691	0.165	24	0.057	15	0.15
COBALT	8.85	0.320	414	0.099	69	0.165	19	0.188
COPPER	8.96	0.324	385	0.092	394	0.941	114	1.14
GALLIUM	5.907	0.213	331	0.079	29-38	0.07-0.09	17	0.17
GERMANIUM	5.323	0.192	306	0.073	59	0.14	36	0.36
GOLD	19.32	0.698	131	0.0312	297	0.71	118	1.178
HAFNIUM	13.09	0.472	147	0.0351	22	0.053	12	0.12

INDIUM	7.31	0.264	239	0.057	24	0.057	14	0.137
IRIDIUM	22.5	0.813	129	0.0307	59	0.14	20	0.20
IRON	7.87	0.284	460	0.11	75	0.18	21	0.208
LEAD	11.36	0.410	129	0.0309	35	0.083	24	0.236
MAGNESIUM	1.74	0.063	1025	0.245	154	0.367	86	0.86
MOLYBDENUM	10.22	0.369	276	0.066	142	0.34	50	0.50
NICKEL	8.902	0.322	440	0.105	92	0.22	23.5	0.235
NIOBIUM	8.57	0.310	268	0.064	54	0.129	23.6	0.236
PALLADIUM	12.02	0.434	244	0.0584	70	0.168	24	0.24
PLATINUM	21.45	0.775	131	0.0314	69	0.165	24.5	0.245
PLUTONIUM	19.84	0.717	138	0.033	8	0.020	3.0	0.030
RHODIUM	12.44	0.449	247	0.059	88	0.21	29	0.286
SILICON	2.33	0.084	678	0.162	84	0.20	53	0.53
SILVER	10.49	0.379	234	0.0559	418	1.0	170	1.705
SODIUM	0.9712	0.035	1235	0.295	134	0.32	112	1.12
TANTALUM	16.6	0.600	142	0.034	54	0.130	23	0.23
TIN	7.2984	0.264	226	0.054	63	0.150	38	0.38
TITANIUM	4.507	0.163	519	0.124	22	0.052	9	0.092
TUNGSTEN	19.3	0.697	138	0.033	166	0.397	62	0.62
URANIUM	19.07	0.689	117	0.0279	30	0.071	13	0.13
VANADIUM	6.1	0.22	498	0.119	31	0.074	10	0.10
ZINC	7.133	0.258	383	0.0915	113	0.27	41	0.41
ZIRCONIUM	6.489	0.234	280	0.067	21	0.050	12	0.12

TABLE 2 THERMAL DIFFUSIVITIES OF COMMON ALLOYS FROM 20 TO 100 °C (68 TO 212 °F)

ALLOYS	DENSITY		HEAT CAPACITY		THERMAL CONDUCTIVITY		THERMAL DIFFUSIVITY	
	g/cm^3	lb/in.^{3}	j/kg · k	cal$_{it}$/g · °c	w/m · k	cal$_{it}$/cm · s · °c	mm^2/s	cm^2/s
ALUMINUM ALLOYS								
1100	2.71	0.098	963	0.23	222	0.53	85	0.85
2014	2.80	0.101	963	0.23	193	0.46	71	0.71
5052	2.68	0.097	963	0.23	138	0.33	54	0.54
6061	2.70	0.098	963	0.23	172	0.41	66	0.66
7075	2.80	0.101	963	0.23	121	0.29	45	0.45
COPPER ALLOYS								
COMMERCIAL BRONZE	8.80	0.318	377	0.09	188	0.45	57	0.57
CARTRIDGE BRASS	8.53	0.308	377	0.09	121	0.29	38	0.38
NAVAL BRASS	8.41	0.303	377	0.09	117	0.28	37	0.37
BERYLLIUM COPPER	8.23	0.297	419	0.1	84	0.20	24	0.24
9% ALUMINUM BRONZE	7.58	0.273	435	0.104	60	0.144	18	0.18
MAGNESIUM ALLOYS								
AZ 31	1.78	0.064	1050	0.25	84	0.20	45	0.45
AZ 91	1.83	0.066	1005	0.24	84	0.20	46	0.46
ZW 1	1.8	0.065	1005	0.24	134	0.32	74	0.74
RZ 5	1.84	0.066	963	0.23	113	0.27	64	0.64
STAINLESS STEELS								

TYPE 301	7.9	0.285	502	0.12	16	0.039	4.1	0.041
TYPE 304	7.9	0.285	502	0.12	15.1	0.036	3.8	0.038
TYPE 316	8.0	0.289	502	0.12	15.5	0.037	3.9	0.039
TYPE 410	7.7	0.278	460	0.11	24	0.057	6.7	0.067
TYPE 430	7.7	0.278	460	0.11	26	0.062	7.3	0.073
TYPE 501	7.7	0.278	460	0.11	37	0.088	10	0.10
NICKEL-BASE ALLOYS								
NIMONIC 80A	8.19	0.296	460	0.11	11	0.027	3.0	0.030
INCONEL 600	8.42	0.304	460	0.11	15	0.035	3.8	0.038
MONEL 400	8.83	0.319	419	0.10	22	0.052	5.8	0.058
TITANIUM ALLOYS								
TI-6AL-4V	4.43	0.160	611	0.146	5.9	0.014	2.1	0.021
TI-5AL-2.5SN	4.46	0.161	460	0.11	6.3	0.015	3.1	0.031

For the planar heat source on a steel surface, as represented by Fig. 1, the time in seconds to produce melting on the surface, t_m, is given by:

$$T_M = (5000/\text{H.I.})^2 \qquad \textbf{(EQ 2)}$$

where H.I. is the net heat intensity (in W/cm^2) transferred to the workpiece.

Equation 2 provides a rough estimate of the time required to produce melting, and is based upon the thermal diffusivity of steel. Materials with higher thermal diffusivities--or the use of a local point heat source rather than a planar heat source--will increase the time to produce melting by a factor of up to two to five times. On the other hand, thin materials tend to heat more quickly.

If the time to melting is considered to be a characteristic interaction time, t_I, then the graph shown in Fig. 3 can be generated. Heat sources with power densities that are of the order of 1000 W/cm^2, such as oxyacetylene flames or electroslag welding, require interaction times of 25 s with steel, whereas laser and electron beams, at 1 MW/cm^2, need interaction times on the order of only 25 µs. If this interaction time is divided into the heat-source diameter, d_H, then a maximum travel speed, V_{max}, is obtained for the welding process (Fig. 4).

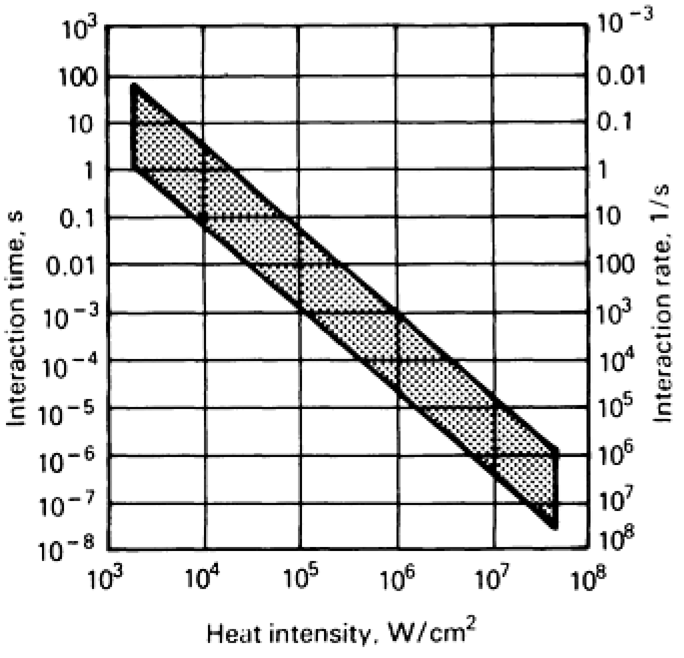

FIG. 3 TYPICAL WELD POOL-HEAT SOURCE INTERACTION TIMES AS FUNCTION OF HEAT-SOURCE INTENSITY.

MATERIALS WITH A HIGH THERMAL DIFFUSIVITY, SUCH AS COPPER OR ALUMINUM, WOULD LIE NEAR THE TOP OF THIS BAND, WHEREAS STEELS, NICKEL ALLOYS, OR TITANIUM WOULD LIE IN THE MIDDLE. URANIUM AND CERAMICS, WITH VERY LOW THERMAL DIFFUSIVITIES, WOULD LIE NEAR THE BOTTOM OF THE BAND.

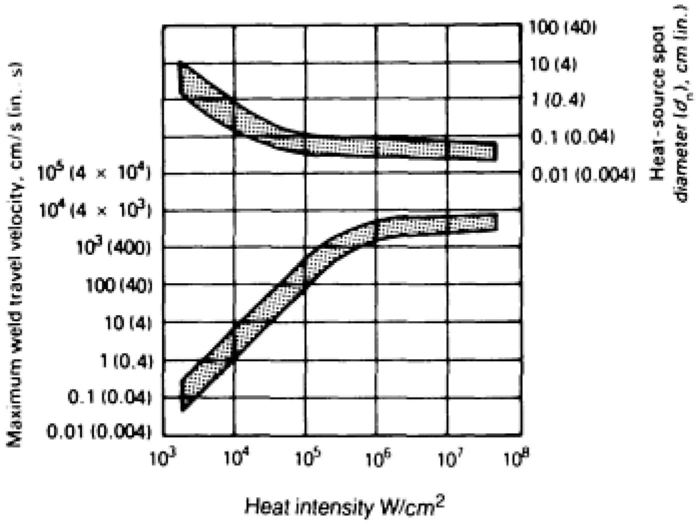

FIG. 4 MAXIMUM WELD TRAVEL VELOCITY AS A FUNCTION OF HEAT-SOURCE INTENSITY BASED ON TYPICAL HEAT-SOURCE SPOT DIAMETERS

The reason why welders begin their training with the oxyacetylene process should be clear: it is inherently slow and does not require rapid response time in order to control the size of the weld puddle. Greater skill is needed to control the more-rapid fluctuations in arc processes. The weld pool created by the high-heat-intensity processes, such as laser-beam and electron-beam welding, cannot be humanly controlled and must therefore be automated. This need to automate leads to increased capital costs. On an approximate basis, the W/cm^2 of a process can be substituted with the dollar cost of the capital equipment. With reference to Fig. 2, the cost of oxyacetylene welding equipment is nearly $1000, whereas a fully automated laser-beam or electron-beam system can cost $1 million. Note that the capital cost includes only the energy source, control system, fixturing, and materials handling equipment. It does not include operating maintenance or inspection costs, which can vary widely depending on the specific application.

For constant total power, a decrease in the spot size will produce a squared increase in the heat intensity. This is one of the reasons why the spot size decreases with increasing heat intensity (Fig. 4). It is easier to make the spot smaller than it is to increase the power rating of the equipment. In addition, only a small volume of material usually needs to be melted. If the spot size were kept constant and the input power were squared in order to obtain higher densities, then the volume of fused metal would increase dramatically, with no beneficial effect.

However, a decreasing spot size, coupled with a decreased interaction time at higher power densities, compounds the problem of controlling the higher-heat-intensity process. A shorter interaction time means that the sensors and controllers necessary for automation must operate at higher frequencies. The smaller spot size means that the positioning of the heat source must be even more precise, that is, on the order of the heat-source diameter, d_H. The control frequency must be greater than the travel velocity divided by the diameter of the heat source. For processes that operate near the maximum travel velocity, this is the inverse of the process interaction time, t_I (Fig. 3).

Thus, not only must the high-heat-intensity processes be automated because of an inherently high travel speed, but the fixturing requirements become greater, and the control systems and sensors must have ever-higher frequency responses. These factors lead to increased costs, which is one reason that the very productive laser-beam and electron-beam welding processes have not found wider use. The approximate productivity of selected welding processes, expressed as length of weld produced per second, to the relative capital cost of equipment is shown in Fig. 5.

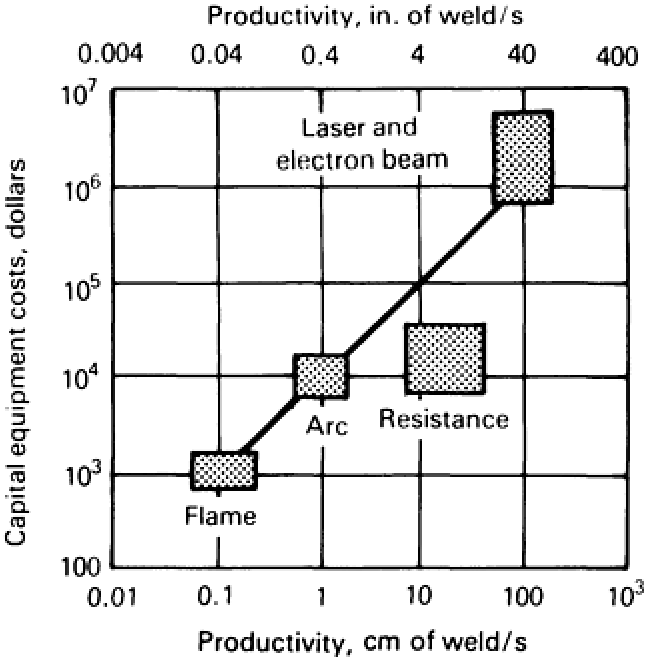

FIG. 5 APPROXIMATE RELATIONSHIP BETWEEN CAPITAL COST OF WELDING EQUIPMENT AND SPEED AT WHICH SHEET METAL JOINTS CAN BE PRODUCED

Another important welding process parameter that is related to the power density of the heat source is the width of the heat-affected zone (HAZ). This zone is adjacent to the weld metal and is not melted itself but is structurally changed because of the heat of welding. Using the Einstein equation, the HAZ width can be estimated from the process interaction time and the thermal diffusivity of the material. This is shown in Fig. 6, with one slight modification. At levels above approximately 10^4 W/cm^2, the HAZ width becomes roughly constant. This is due to the fact that the HAZ grows during the heating stage at power densities that are below 10^4 W/cm^2, but at higher power densities it grows during the cooling cycle. Thus, at low power densities, the HAZ width is controlled by the interaction time, whereas at high power densities, it is independent of the heat-source interaction time. In the latter case, the HAZ width grows during the cooling cycle as the heat of fusion is removed from the weld metal, and is proportional to the fusion zone width.

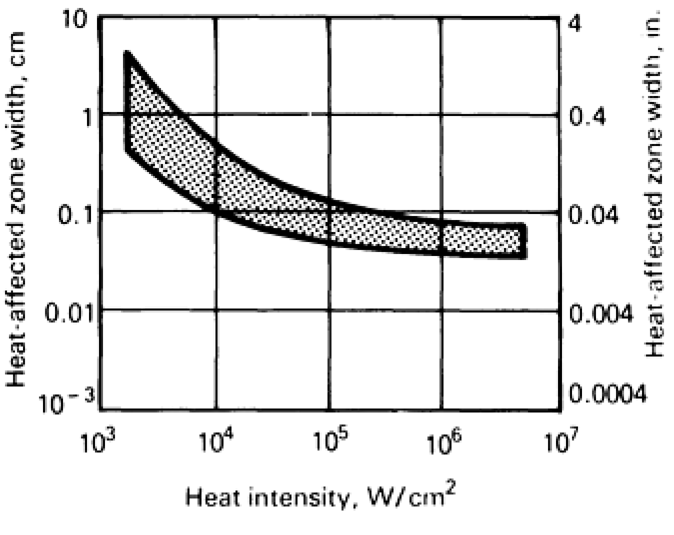

FIG. 6 RANGE OF WELD HAZ WIDTHS AS FUNCTION OF HEAT-SOURCE INTENSITY

The change of slope in Fig. 6 also represents the heat intensity at which the heat utilization efficiency of the process changes. At high heat intensities, nearly all of the heat is used to melt the material and little is wasted in preheating the surroundings. As heat intensity decreases, this efficiency is reduced. For arc welding, as little as half of the heat generated may enter the plate, and only 40% of this heat is used to fuse the metal. For oxyacetylene welding, the heat entering the metal may be 10% or less of the total heat, and the heat necessary to fuse the metal may be less than 2% of the total heat.

A final point is that the heat intensity also controls the depth-to-width ratio of the molten pool. This value can vary from 0.1 in low-heat-intensity processes to more than 10 in high-heat-intensity processes.

It should now be evident that all fusion welding processes can be characterized generally by heat-source intensity. The properties of any new heat source can be estimated readily from the figures in this article. Nonetheless, it is useful to more fully understand each of the common welding heat sources, such as flames, arcs, laser beams, electron beams, and electrical resistance. These are described in separate articles in the Section "Fusion Welding Processes" in this Volume.

Heat Flow in Fusion Welding

Chon L. Tsai and Chin M. Tso, The Ohio State University

Introduction

DURING FUSION WELDING, the thermal cycles produced by the moving heat source cause physical state changes, metallurgical phase transformation, and transient thermal stress and metal movement. After welding is completed, the finished product may contain physical discontinuities that are due to excessively rapid solidification, or adverse microstructures that are due to inappropriate cooling, or residual stress and distortion that are due to the existence of incompatible plastic strains.

In order to analyze these problems, this article presents an analysis of welding heat flow, focusing on the heat flow in the fusion welding process. The primary objective of welding heat flow modeling is to provide a mathematical tool for thermal data analysis, design iterations, or the systematic investigation of the thermal characteristics of any welding parameters. Exact comparisons with experimental measurements may not be feasible, unless some calibration through the experimental verification procedure is conducted.

Welding Thermal Process. A physical model of the welding system is shown in Fig. 1. The welding heat source moves at a constant speed along a straight path. The end result, after either initiating or terminating the heat source, is the formation of a transient thermal state in the weldment. At some point after heat-source initiation but before termination, the temperature distribution is stationary, or in thermal equilibrium, with respect to the moving coordinates. The origin of the moving coordinates coincides with the center of the heat source. The intense welding heat melts the metal and forms a molten pool. Some of the heat is conducted into the base metal and some is lost from either the arc column or the metal surface to the environment surrounding the plate. Three metallurgical zones are formed in the plate upon completion of the thermal cycle: the weld-metal (WM) zone, the heated-affected zone (HAZ), and the base-metal (BM) zone. The peak temperature and the subsequent cooling rates determine the HAZ structures, whereas the thermal gradients, the solidification rates, and the cooling rates at the liquid-solid pool boundary determine the solidification structure of the WM zone. The size and flow direction of the pool determines the amount of dilution and weld penetration. The material response in the temperature range near melting temperatures is primarily responsible for the metallurgical changes.

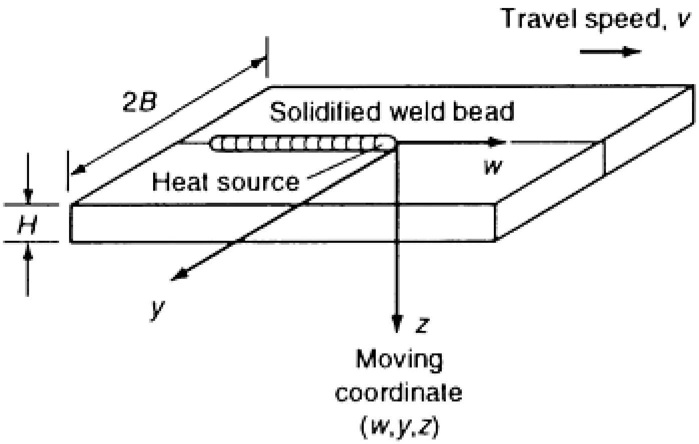

FIG. 1 SCHEMATIC OF THE WELDING THERMAL MODEL

Two thermal states, quasi-stationary and transient, are associated with the welding process. The transient thermal response occurs during the source initiation and termination stages of welding, the latter of which is of greater metallurgical interest. Hot cracking usually begins in the transient zone, because of the nonequilibrium solidification of the base material. A crack that forms in the source-initiation stage may propagate along the weld if the solidification strains sufficiently multiply in the wake of the welding heat source. During source termination, the weld pool solidifies several times faster than the weld metal in the quasi-stationary state. Cracks usually appear in the weld crater and may propagate along the weld. Another dominant transient phenomenon occurs when a short repair weld is made to a weldment. Rapid cooling results in a brittle HAZ structure and either causes cracking problems or creates a site for fatigue-crack initiation.

The quasi-stationary thermal state represents a steady thermal response of the weldment in respect to the moving heat source. The majority of the thermal expansion and shrinkage in the base material occurs during the quasi-stationary thermal cycles. Residual stress and weld distortion are the thermal stress and strain that remain in the weldment after completion of the thermal cycle.

Relation to Welding Engineering Problems. To model and analyze the thermal process, an understanding of thermally induced welding problems is important. A simplified modeling scheme, with adequate assumptions for specific problems, is possible for practical applications without using complex mathematical manipulations. The relationship between the thermal behavior of weldments and the metallurgy, control, and distortion associated with welding is summarized below.

Welding Metallurgy. As already noted, defective metallurgical structures in the HAZ and cracking in the WM usually occur under the transient thermal condition. Therefore, a transient thermal model is needed to analyze cracking and embrittlement problems.

To evaluate the various welding conditions for process qualification, the quasi-stationary thermal responses of the weld material need to be analyzed. The minimum required amount of welding heat input within the allowable welding speed range must be determined in order to avoid rapid solidification and cooling of the weldment Preheating may be necessary if the proper thermal conditions cannot be obtained under the specified welding procedure. A quasi-stationary thermal model is adequate for this type of analysis.

Hot cracking results from the combined effects of strain and metallurgy. The strain effect results from weld-metal displacement at near-melting temperatures, because of solidification shrinkage and weldment restraint. The metallurgical effect relates to the segregation of alloying elements and the formation of the eutectic during the high nonequilibrium solidification process. Using metallurgical theories, it is possible to determine the chemical segregation, the amounts and distributions of the eutectic, the magnitudes and directions of grain growth, and the weld-metal displacement at high temperatures. Using the heating and cooling rates, as well as the retention period predicted by modeling and analysis, hot-cracking tendencies can be determined. To analyze these tendencies, it is important to employ a more accurate numerical model that considers finite welding heat distribution, latent heat, and surface heat loss.

Welding Control. In-process welding control has been studied recently. Many of the investigations are aimed at developing sensing and control hardware. However, a link between weld-pool geometry and weld quality has not been fully established. A transient heat-flow analysis needs to be used to correlate the melted surface, which is considered to be the primary control variable, to the weld thermal response in a time domain.

Welding Distortion The temperature history and distortion caused by the welding thermal process creates nonlinear thermal strains in the weldment. Thermal stresses are induced if any incompatible strains exist in the weld. Plastic strains are formed when the thermal stresses are higher than the material yield stress. Incompatible plastic strains accumulate over the thermal process and result in residual stress and distortion of the final weldment. The material response in the lower temperature range during the cooling cycle is responsible for the residual stresses and weldment distortion. For this type of analysis, the temperature field away from the welding heat source is needed for the modeling of the heating and cooling cycle during and after welding. A quasi-stationary thermal model with a concentrated moving heat source can predict, with reasonable accuracy, the temperature information for the subsequent stress and distortion analysis.

Literature Review. Many investigators have analytically, numerically, and experimentally studied welding heat-flow modeling and analysis (Ref 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18). The majority of the studies were concerned with the quasi-stationary thermal state. Lance and Martin (Ref 1), Rosenthal and Schmerber (Ref 2) and Rykalin (Ref 3) independently obtained an analytical temperature solution for the quasi-stationary state using a point or line heat source moving along a straight line on a semi-infinite body. A solution for plates of finite thickness was later obtained by many investigators using the imaged heat source method (Ref 3, 4). Tsai (Ref 5) developed an analytical solution for a model that incorporated a welding heat source with a skewed Gaussian distribution and finite plate thickness. It was later called the "finite source theory" (Ref 6).

With the advancement of computer technology and the development of numerical techniques like the finite-difference and finite-element methods, more exact welding thermal models were studied and additional phenomena were considered, including nonlinear thermal properties, finite heat-source distributions, latent heat, and various joint geometries. Tsai (Ref 5), Pavelic (Ref 7), Kou (Ref 8), Kogan (Ref 9), and Brody (Ref 10) studied the simulation of the welding process using the finite-difference scheme. Hibbitt and Marcal (Ref 11), Friedman (Ref 12), and Paley (Ref 13) made some progress in welding simulation using the finite-element method.

Analytical solutions for transient welding heat flow in a plate were first studied by Naka (Ref 14), Rykalin (Ref 3), and Masubuchi and Kusuda (Ref 15) in the 1940s and 1950s. A point or line heat source, constant thermal properties, and adiabatic boundary conditions were assumed. Later, Tsai (Ref 16) extended the analytical solution to incorporate Gaussian heat distribution using the principle of superposition. The solution was used to investigate the effect of pulsed conditions on weld-pool formation and solidification without the consideration of latent heat and nonlinear thermal properties.

The analysis of the transient thermal behavior of weldments using numerical methods has been the focus of several investigations since 1980. Friedman (Ref 17) discussed the finite-element approach to the general transient thermal analysis of the welding process. Brody (Ref 10) developed a two-dimensional transient heat flow model using a finite-difference scheme and a simulated pulsed-current gas-tungsten arc welding process (GTAW). Tsai and Fan (Ref 18) modeled the two-dimensional transient welding heat flow using a finite-element scheme to study the transient welding thermal behavior of the weldment.

General Approach. The various modeling and analysis schemes summarized above can be used to investigate the thermal process of different welding applications. With adequate assumptions, analytical solutions for the simplified model can be used to analyze welding problems that show a linear response to the heat source if the solutions are properly calibrated by experimental tests. Numerical solutions that incorporate nonlinear thermal characteristics of weldments are usually required for investigating the weld-pool growth or solidification behavior. Numerical solutions can also be necessary for metallurgical studies in the weld HAZ if the rapid cooling phenomenon is significant under an adverse welding environment, such as welding under water.

Thermally related welding problems can be categorized as:

- SOLIDIFICATION RATES IN THE WELD POOL
- COOLING RATES IN THE HAZ AND ITS VICINITY
- THERMAL STRAINS IN THE GENERAL DOMAIN OF THE WELDMENT

The domain of concern in the weld pool solidification is within the molten pool area, in which the arc (or other heat source) phenomena and the liquid stirring effect are significant. A convective heat-transfer model with a moving boundary at the melting temperature is needed to study the first category, and numerical schemes are usually required, as well.

The HAZ is always bounded on one side by the liquid-solid interface during welding. This inner-boundary condition is the solidus temperature of the material. The liquid weld pool might be eliminated from thermal modeling if the interface could be identified. A conduction heat-transfer model would be sufficient for the analysis of the HAZ. Numerical methods are often employed and very accurate results can be obtained.

The thermal strains caused by welding thermal cycles are caused by the nonlinear temperature distribution in the general domain of the weldment. Because the temperature in the material near the welding heat source is high, very little stress can be accumulated from the thermal strains. This is due to low rigidity, that is, small modulus of elasticity and low yield strength. The domain for thermal strain study is less sensitive to the arc and fluid-flow phenomena and needs only a relatively simple thermal model. Analytical solutions with minor manipulations often provide satisfactory results.

In this article, only the analytical heat-flow solutions and their practical applications are addressed. The numerical conduction solutions and the convective models for fluid flow in a molten weld pool are not presented.

References

1. N.S. BOULTON AND H.E. LANCE-MARTIN, RESIDUAL STRESSES IN ARC WELDED PLATES, *PROC. INST. MECH. ENG.,* VOL 33, 1986, P 295
2. D. ROSENTHAL AND R. SCHMERBER, THERMAL STUDY OF ARC WELDING, *WELD. J.,* VOL 17 (NO. 4), 1983, P 2S
3. N.N. RYKALIN, "CALCULATIONS OF THERMAL PROCESSES IN WELDING," MASHGIZ, MOSCOW, 1951
4. K. MASUBUCHI, *ANALYSIS OF WELDED STRUCTURES,* PERGAMON PRESS, 1980
5. C.L. TSAI, "PARAMETRIC STUDY ON COOLING PHENOMENA IN UNDERWATER WELDING," PH.D THESIS, MIT, 1977
6. C.L. TSAI, FINITE SOURCE THEORY, *MODELING OF CASTING AND WELDING PROCESSES II,* ENGINEERING FOUNDATION MEETING, NEW ENGLAND COLLEGE (HENNIKER, NH), 31 JULY TO 5 AUG 1983, P 329
7. R. PAVELIC, R. TANAKUCHI, O. CZEHARA, AND P. MYERS, EXPERIMENTAL AND COMPUTED TEMPERATURE HISTORIES IN GAS TUNGSTEN ARC WELDING IN THIN PLATES, *WELD. J.,* VOL 48 (NO. 7), 1969, P 295S
8. S. KOU, 3-DIMENSIONAL HEAT FLOW DURING FUSION WELDING, *PROC. OF METALLURGICAL SOCIETY OF AIME,* AUG 1980, P 129-138
9. P.G. KOGAN, THE TEMPERATURE FIELD IN THE WELD ZONE, *AVE. SVARKA,* VOL 4 (NO. 9), 1979, P 8
10. G.M. ECER, H.D. DOWNS, H.D. BRODY, AND M.A. GOKHALE, HEAT FLOW SIMULATION OF PULSED CURRENT GAS TUNGSTEN ARC WELDING, *MODELING OF CASTING AND WELDING PROCESSES,* ENGINEERING FOUNDATION 1980 MEETING (RINDGE, NH), 3-8 AUG 1980, P 139-160
11. H. HIBBITT AND P. MARCAL, A NUMERICAL THERMOMECHANICAL MODEL FOR WELDING AND SUBSEQUENT LOADING OF A FABRICATED STRUCTURE, *COMPUT. STRUCT.,* VOL 3, 1973, P 1145
12. E. FRIEDMAN, THERMOMECHANICAL ANALYSIS OF THE WELDING PROCESS USING FINITE ELEMENT METHODS, *TRANS. ASME,* AUG 1975, P 206
13. Z. PALEY AND P. HIBBERT, COMPUTATION OF TEMPERATURE IN ACTUAL WELD DESIGN, *WELD. J.,* VOL 54 (NO. 11), 1975, P 385.S

14. T. NAKA, TEMPERATURE DISTRIBUTION DURING WELDING, *J. JPN. WELD. SOC.,* VOL 11 (NO. 1), 1941, P 4
15. K. MASUBUCHI AND T. KUSUDA, TEMPERATURE DISTRIBUTION OF WELDED PLATES, *J. JPN WELD. SOC.,* VOL 22 (NO. 5), 1953, P 14
16. C.L. TSAI AND C.A. HOU, THEORETICAL ANALYSIS OF WELD POLL BEHAVIOR IN THE PULSED CURRENT GTAW PROCESS, *TRANSPORT PHENOMENA IN MATERIALS PROCESSING,* ASME WINTER ANNUAL MEETING, 1983
17. E. FRIEDMAN, "FINITE ELEMENT ANALYSIS OF ARC WELDING," REPORT WAPD-TM-1438, DEPARTMENT OF ENERGY, 1980
18. J.S. FAN AND C.L. TSAI, "FINITE ELEMENT ANALYSIS OF WELDING THERMAL BEHAVIOR IN TRANSIENT CONDITIONS," 84-HT-80, ASME

Heat Flow in Fusion Welding

Chon L. Tsai and Chin M. Tso, The Ohio State University

Mathematical Formulations

Conduction Equation. A diagram of the welding thermal model is shown in Fig. 1. The origin of the moving coordinates (w,x,z) is fixed at the center of the welding heat source. The coordinates move with the source at the same speed. The conduction equation for heat flow in the weldments is:

$$\nabla.(l\nabla q) + rC_p v\frac{\partial q}{\partial w} + \dot{Q} = rC_p v\frac{\partial q}{\partial t} \qquad \textbf{(EQ 1)}$$

The initial condition is:

$$\theta = \theta_0,\ \text{AT } T = 0 \qquad \textbf{(EQ 2)}$$

and the general boundary condition is:

$$l\frac{\partial q}{\partial w}I_w + \frac{\partial q}{\partial y}I_y + \frac{\partial q}{\partial z}I_z - \dot{q} + h(q - q_\infty) = 0 \qquad \textbf{(EQ 3)}$$

where ∇ is a differential operator; θ is the temperature; θ_∞ is the environmental temperature; θ_0 is the initial temperature; λ is thermal conductivity; ρ is density; C_p is specific heat h is the surface heat-loss coefficient; l_w, l_y, and l_z are the direction cosines of the boundary surface; Q is the volumetric heat source, t is time, and v is welding speed.

The volumetric heat source represents the Joule heating in the weldment that is due to the electric current flow within that conducting medium. The total energy of such heating in welding is usually minimal, compared to the arc heat input. The majority of the energy is concentrated in a very small volume beneath the arc (Ref 5). In other words, a very high energy density generation exists in the weld pool, and it may have a significant effect on transient pool growth and solidification.

Heat-Source Formulation. The direction cosines on the surface that receive the heat flux from the welding source (z = 0) are $l_w = l_y = 0$ and $l_z = -1$. Within the significant heat-input area (to be defined later in this section), the heat loss coefficient, h, is zero.

The distribution of the welding heat flux on the weldment surface can be characterized, in a general form, by a skewed Gaussian function (Ref 19):

$$\dot{q}(r,w) = \dot{q}_0 \exp\left(-Cr^2 - \frac{bv}{2k}w\right) \qquad \textbf{(EQ 4)}$$

where β is a weight constant, κ is the thermal diffusivity of the base material, C is a shape constant, $\dot{q}$ is heat flux as a function of (*r, w*), $\dot{q}_0$ is heat flux at the source center, *r* is the radial coordinate from the source center, and *v* is welding speed.

The weight constant, β, indicates the significance of the welding travel speed. A normal distribution of the welding heat flux is obtained if the weight constant is zero.

In general, the total energy input to the weldment, which is a fraction of the total welding power generated by the welding machine, is the sum of the concentrated heat and the diffused heat (Ref 20). The concentrated heat is carried by the core of the energy transmission medium, for example, the arc plasma column. The diffused heat reaches the weld surface by radiation and convection energy transport from the core surface. The heat-flux distribution is a function of the proportional values between these two types of energy. The fraction of the total welding power reaching the weldment indicates the heating efficiency of the welding process, and the fraction percentage is defined as welding heat efficiency, η.

The shape constant, C, can be obtained in terms of the core diameter, *D*, and the concentration factor, *F*. The concentration factor is defined as the ratio of the concentrated heat to the net energy reaching the weldment. The core diameter can be assumed to be the diameter of the plasma column in the arc welding process. The concentration factor and welding heat efficiency are not fully understood and have been subjected to manipulation during the mathematical analyses in order to obtain a better correlation with the experimental data.

Assuming a normal heat-flux model, two concentrations are required to determine the shape constant and the heat flux at the source center, $\dot{q}_0$. By integrating Eq 4 over the core heat area and the entire heat input domain ($r = 0 \rightarrow \infty$), the shape factor can be determined by dividing the two integrals. The heat flux at the source center can then be determined from the second integral. The two constants are expressed as:

$$C = \frac{4l_{arc}[1/(1-F)]}{D^2} \qquad \textbf{(EQ 5)}$$

where l_{arc} is the length of the arc plasma column.

$$\frac{\dot{q}_0}{\dot{Q}} = \frac{C}{p} \qquad \textbf{(EQ 6)}$$

In the case of arc welding,

$$\frac{\dot{q}_0}{hEI} = \frac{C}{p} \qquad \textbf{(EQ 7)}$$

where *E* is the welding arc voltage and *I* is the welding current.

For practical purposes, the welding heat source can be considered to be restricted within a circle of radius r_a, where the heat flux drops to 1/100 of the center flux $\dot{q}_0$. The radius of the significant heat input area can be written as:

$$r_a = \left(\frac{I_{arc}100}{C}\right)^{0.5} \qquad \textbf{(EQ 8)}$$

Surface Heat Loss. The heat-loss coefficient, *h*, represents both radiation and convection heat loss from the boundary surfaces outside the significant heat input area. The formulation for both heat-loss mechanisms can be written as the radiation heat-loss coefficient (in air):

$$H_{\mathrm{RAD}} = \varepsilon\sigma(\theta_{\mathrm{w}} + \infty)(\theta_{\mathrm{W}}{}^{2} - \theta_{\infty}{}^{2})) \qquad \textbf{(EQ 9)}$$

or the natural convection heat-loss coefficient (in air):

$$H_{AC} = 0.00042\frac{q_W - q_\infty}{B} \qquad \textbf{(EQ 10)}$$

or the convection heat-loss coefficient (in water):

$$H_{\mathrm{CW}} = 0.442(\theta_{\mathrm{W}} - \theta_{\infty})^{0.25} \qquad \textbf{(EQ 11)}$$

where ε is emissivity, σ is the Stefan-Boltzmann constant, θ_{w} is the surface temperature, θ_{∞}, is the environmental temperature, and B is the characteristic surface dimension.

Natural convection is dominant at a temperature below 550 °C (1020 °F), whereas radiation becomes more important at temperatures above this level. The total heat-loss coefficient is the sum of Eq 9 and 10. The characteristic surface dimension is the effective distance from the source beyond which the temperature rises insignificantly during welding. The characteristic dimension for steel is about 150 mm (6 in.) (Ref 5). In underwater welding, heat losses are primarily due to heat transfer from the surface to the moving water environment. This motion is created by the rising gas column in the arc area (Ref 21).

For an insulated surface, no heat transfer into or out of the surface is assumed. The temperature gradient normal to the surface is zero, and can be represented by:

$$N \cdot \nabla q = 0 \qquad \textbf{(EQ 12)}$$

where n is a unit vector normal to the surface and equals $(l_{\mathrm{w}}^{2} + l_{\mathrm{y}}^{2} + l_{\mathrm{z}}^{2})^{0.5}$.

Other Boundary Conditions. There are several other possible boundary conditions in welding heat-flow modeling that depend on the assumptions used for model simplification. One is the condition at infinity:

$$q = q_\infty \text{ or } \lim_{r\to\infty}\frac{\partial q}{\partial r} = 0 \qquad \textbf{(EQ 13)}$$

Another is the condition near the heat source. In the case of a line source for a thin plate:

$$-2pl\,\mathrm{H}\lim_{r\to\infty} r\frac{\partial q}{\partial r} = hEI \qquad \textbf{(EQ 14)}$$

In the case of a point source for a thick plate:

$$-2pl\lim_{r\to\infty} r^2\frac{\partial q}{\partial r} = hEI \qquad \textbf{(EQ 15)}$$

In the case of a finite source for a thick plate:

$$N \cdot (-\lambda\nabla\theta) = \dot{q};\ R \leq R_{\mathrm{A}} \qquad \textbf{(EQ 16)}$$

Another is represented by the conditions at the solid-liquid interface:

$$\theta_1 = \theta_S = \theta_M \qquad \textbf{(EQ 17)}$$

$$\mathrm{n}.(l\nabla q)_s - n(l\nabla q)_1 = \pm r_s L \frac{ds}{dt} \qquad \textbf{(EQ 18)}$$

where + indicates the melting process and - indicates the solidification process. The subscripts s and l indicate the temperature and the properties in a solid and liquid, respectively. The *n* is a normal vector on the boundary surface or interface, r_a is the radius of the heat-input area, *L* is the latent heat of the base material, and the subscript m represents the melting temperature of the base material.

References cited in this section

5. C.L. TSAI, "PARAMETRIC STUDY ON COOLING PHENOMENA IN UNDERWATER WELDING," PH.D THESIS, MIT, 1977
19. R.L. APPS AND D.R. MILNER, HEAT FLOW IN ARGON-ARC WELDING, *BR. WELD. J.,* VOL 2 (NO. 10), 1955, P 475
20. H.S. CARSLAW AND J.C. JAEGER, *CONDUCTION OF HEAT IN SOLIDS,* OXFORD PRESS
21. C.L. TSAI AND J.H. WU, "AN INVESTIGATION OF HEAT TRANSPORT PHENOMENA IN UNDERWATER WELDING," PRESENTED AT THE ASME WINTER ANNUAL MEETING (MIAMI BEACH, FL), 1985

Heat Flow in Fusion Welding

Chon L. Tsai and Chin M. Tso, The Ohio State University

Engineering Solutions and Empirical Correlation

General Solutions. The general (analytical) heat-flow solutions for fusion welding can be categorized by those appropriate for a thick plate, a thin plate, or a plate with finite thickness. In most cases, the boundary surfaces (except for the heat-input area) are assumed to be adiabatic, and the thermal properties are independent of temperature. The various metallurgical zones in the weldment are assumed to be homogeneous, and the thermal model is linear.

The solutions give the temperature for a specific point if the welding velocity, *v*, voltage, *E*, and current, *I*, as well as the physical properties of the plate material (ρ, λ, C_p) and the welding heat efficiency, η, are known. This specific point is defined by *r* and *w* in:

$$r = \sqrt{w^2 + y^2 + z^2} \qquad \textbf{(EQ 19)}$$

where $w = x - vt$. The heat-flow solutions are not accurate at points near the welding arc, because a point source or line source is assumed for thick and thin plates, respectively.

To approximate the transient temperature changes at the start and end of a weld, Fig. 2 shows a global coordinate system (*x*,*y*,*z*), the origin of which is fixed at the source initiation, where t_0 is the welding time and t_1 is the time after the welding heat-source termination. The temperature solutions at t_0 and t_1 are the temperature changes at the start and end of the weld, respectively.

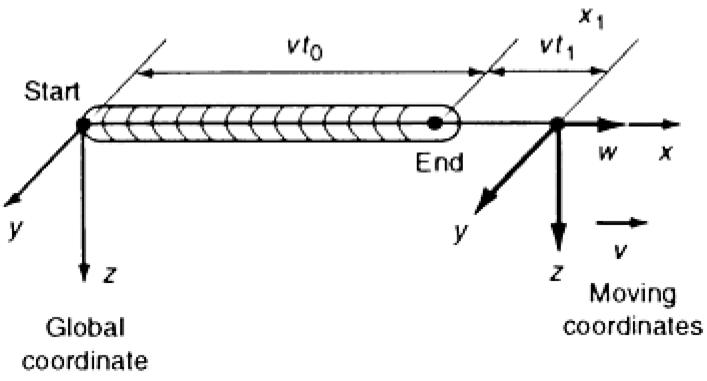

FIG. 2 GLOBAL AND MOVING COORDINATE SYSTEMS FOR WELDING HEAT CONDUCTION.

The temperature solution for thick plate at the arc start location is:

$$q - q_0 = \frac{h\,EI}{2p\,l\,vt_o} \quad \textbf{(EQ 20)}$$

The quasi-stationary temperature distribution is:

$$q - q_0 = \frac{h\,EI}{2p\,l\,r}\exp\left[\frac{-v(w+r)}{2k}\right] \quad \textbf{(EQ 21)}$$

At the arc termination location, the solution is:

$$q - q_0 = \frac{h\,EI}{2p\,l\,vt_1} \quad \textbf{(EQ 22)}$$

The temperature solution for thin plate at the arc start location is:

$$q - q_0 = \frac{h\,EI}{2p\,l\,H}\exp\frac{v^2 t_o}{2k}\,k_o\,\frac{v^2 t_o}{2k} \quad \textbf{(EQ 23)}$$

The quasi-stationary temperature distribution is:

$$q - q_0 = \frac{h\,EI}{2p\,l\,H}\exp\frac{-vw}{2k}\,k_o\,\frac{vr}{2k} \quad \textbf{(EQ 24)}$$

For the arc termination location, the solution is:

$$q - q_0 = \frac{h\,EI}{2p\,l\,H}\exp\frac{v^2 t_1}{2k}\,k_o\left(\frac{v^2 t_1}{2k}\right) \quad \textbf{(EQ 25)}$$

where K_0 is the modified Bessel function of the second kind of zeroth order and $\eta\ EI$ is the welding heat input rate.

Temperature for Plate With Finite Thickness. The image method enables the investigator to superimpose the solutions for an infinitely thick plate, the source of which is placed on imaginary surfaces until the proper boundary conditions on the plate surfaces are obtained. This method is based on the premise that if a solution satisfies the governing equation and the boundary conditions, then it must be not only a correct solution, but the only solution (that is, the uniqueness of solution premise).

Using the image method, the solution for plates of finite thickness with a adiabatic surfaces can be modified from the respective temperature solutions described previously.

Let f_0 (w,y,z,t) be the initial solution for an infinitely thick plate. The temperature solution for a finite thick plate can be obtained by super imposing the imaginary solutions f_{mn} (w,y_m, z_n, t) and f'_{mn} $(w,y'_{\mathrm{m}}, z'_n, t)$ to the initial solution, and this can be written in a general form as:

$$q - q_0 = -f_0(w, y, z, f) + \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}[f_{mn}(w, y_m, z_m, t] + f'_{mn}(w, y'_m, Z_m t)] \qquad \textbf{(EQ 26)}$$

where $y_m = 2mB - y$; $y'_m = 2mB + y$; $z_n = 2nH - z$; and $z'_n = 2nH + z$, in which B is the half width and H is the thickness of the plate. The subscripts m and n are integers that vary from zero to infinity.

For a plate with sufficient width, the subscript m is zero. The solution will coverage and reach the correct adiabatic surface condition in six to ten superposition steps, depending on the thickness of the plate. The two-dimensional solution (that is, thin plate) is generally used for any solution that requires more than ten superposition steps.

Equation 26 can be expressed as:

$$q - q_0 = -\frac{hEI\exp\left(\frac{-vw}{2k}\right)}{2pl}\left\{\frac{\exp\left(\frac{-vr}{2k}\right)}{r} + \sum_{n=1}^{\infty}\left[\frac{\exp\left(\frac{-vr_{2n}}{2k}\right)}{r_{2n}} + \right]\frac{\exp\left(\frac{-vr_{2n+1}}{2k}\right)}{r_{2n-1}}\right\} \qquad \textbf{(EQ 27)}$$

where

$$R_{2N} = W^2 + Y^2 + (2NH - Z)^2 \text{ AND}$$
$$R_{2N+1} = W^2 + Y^2 + (2NH + Z)^2 \qquad \textbf{(EQ 28)}$$

Cooling Rate. Frequently, it is desirable to know the cooling rate experienced at some location in a weldment to enable a prediction of the metallurgical structure in that area. A general methodology by which cooling rate equations are obtained from the temperature-distribution equations is discussed below.

Recall that the moving coordinate w is defined by $w = x - vt$. Using this definition, it is easily shown that:

$$\frac{\partial w}{\partial t} = -v \qquad \textbf{(EQ 29)}$$

Using the chain rule, the cooling-rate equation is:

$$\frac{\partial q}{\partial t} = -v\frac{\partial q}{\partial w} \qquad \textbf{(EQ 30)}$$

Because the temperature-distribution equations are a function of w and r, the cooling-rate equations can be obtained by differentiating the temperature-distribution equations with respect to w and multiplying by $-v$.

The cooling rate is defined as the slope of a tangent line drawn on the temperature-time curve. Because the cooling rate changes with temperature, when one speaks of a cooling rate, the specific temperature, θ_c, at which it occurs must also be given. In a weldment, the variable of interest is the cooling rate at the critical temperature that ultimately defines what type of metallurgical structure will result (if the material is heat treatable). For steels, this critical temperature is the "nose" of the continuous cooling-transformation (CCT) curve. At this temperature, the cooling rate determines if upper transformation products (pearlite, upper bainite) or lower transformation products (martensite, lower bainite) will form. For many steels, this critical temperature ranges from approximately 200 to 540 °C (400 to 1000 °F).

The cooling rate in a weldment is also a function of location. In order to find a cooling-rate equation, the particular location in the weldment that is of interest must be defined. The resulting cooling-rate equation will be applicable only to that location.

The differentiation, $\partial\theta/\partial w$, of either Eq 21 or 27, which is required to obtain the cooling-rate expression, will result in a function of w and r. The variable r can be written in terms of w if the location of interest is defined by a given set of values of y and z. This relationship for r, once formulated, can then be substituted into $\partial\theta/\partial w$, the result being a function of w alone.

To determine w corresponding to the critical temperature, θ_c, a temperature-distribution equation is required (Eq 21 or 27). The aforementioned r-w relationship and temperature distribution equation (Eq 21 or 27) where θ is equal to θ_c, critical temperature, are used to determine w. Then w is substituted into the $d\theta/dw$ expression obtained previously. The end result will be an equation that defines the cooling rate for a particular location in the weldment, and, being a function of the critical temperature, the welding conditions and thermal conductivity of the base material.

To determine the cooling rate in a thick plate along the weld centerline (that is $y = 0$) for a particular critical temperature, the cooling-rate equation can be reduced to:

$$\frac{\partial q}{\partial t} = \frac{2p\,l\,v\left(q_c - q_0\right)^2}{h\,EI} \qquad \textbf{(EQ 31)}$$

This equation has been used to predict weld cooling rates in shop practices. Cooling rate is inversely (that is, $\dot{Q}$ /V) and is proportional to thermal conductivity and the critical temperature at which the cooling rate needs to be evaluated.

On the basis of experimental results, a cooling-rate equation was developed for the HAZ of low-carbon steel weldments (Ref 25). This equation considers the combined effects of plate thickness, H, preheating temperature, θ_0, and welding conditions, and is given as:

$$\left.\frac{\partial q}{\partial t}\right|_{HAZ} = 0.35\left\{\left(\frac{q_c - q_0}{I/v}\right)^{1.7}\left[1+\frac{2}{p}\tan^{-1}\left(\frac{H-H_o}{a}\right)\right]\right\}^{0.8} \qquad \textbf{(EQ 32)}$$

The variables α and H_0 depend on the critical temperature of interest. Several values are given in Table 1.

TABLE 1 SELECTED CRITICAL TEMPERATURE AND CORRESPONDING VALUES FOR α AND H_0

CRITICAL TEMPERATURE, θ_C		**H_0**		**α**	
°C	**°F**	**mm**	**in.**	**mm**	**in.**
700	1290	9.9	0.39	2.0	0.08
540	1000	14.2	0.56	4.1	0.16
300	570	19.8	0.78	9.9	0.39

The units used in Eq 32 are important, because the same units that were used in developing the equation must be employed in its application. The plate thickness, H, must be given in inches, and the travel speed, v, must be given in in./min. The temperatures θ_c and θ_0 must be given in °C, and the welding current, I, must be given in amperes. Using the correct units, the application of Eq 32 will result in a predicted cooling rate (°C/s) for the HAZ of a low-carbon steel weldment.

For low-carbon steels welded by the shielded metal arc welding (SMAW), gas-metal arc welding (GMAW), and submerged arc welding (SAW) processes, an empirical equation has been developed that correlates the weld-metal cooling rate at 538 °C (1000 °F) with a 95 to 150 °C (200 to 300 °F) preheat, the weld nugget area (Ref 26):

$$°C/s=\left(\frac{2012}{nugget\ \text{area}}\right)^{1.119} \qquad \textbf{(EQ 33)}$$

where the nugget area is in mm^2. For low-carbon steels, an empirical chart for determining nugget area for a given welding condition (Fig. 3) also has been developed (Ref 27). The straight line drawn to connect the current and travel speed intersects the nugget area at the calculated value. Welding voltage controls weld bead shape.

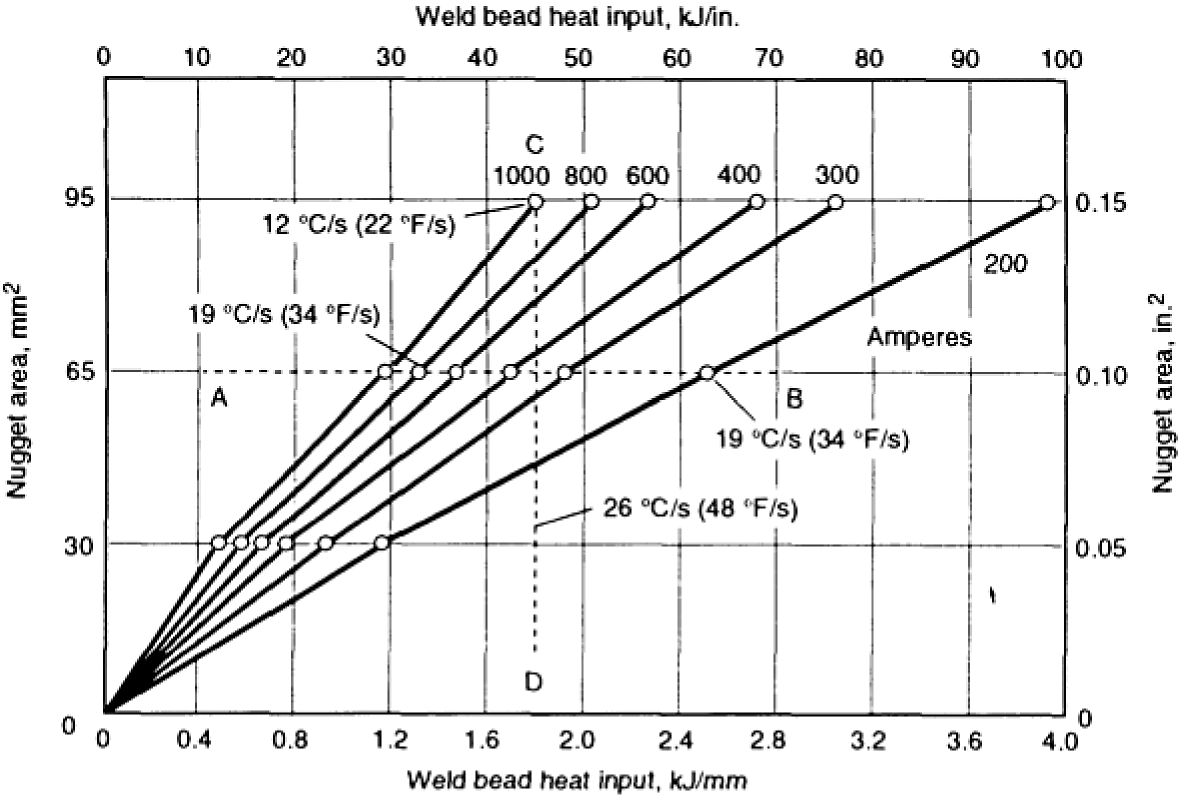

FIG. 3 RELATION BETWEEN NUGGET AREA, HEAT INPUT, AND CURRENT

Peak Temperature. An equation to determine the peak-temperature in a weldment at a given distance y from the weld centerline would enable the prediction of HAZ sizes, as well as weld bead widths. The general concept of obtaining a peak-temperature equation, as well as some results that have been obtained, are discussed below.

Consider Fig. 4 and note that the maximum, or peak, temperature is given when $\partial\theta/\partial t = 0$. For the thick-plate model, the cooling rate can be obtained by differentiating Eq 21 and multiplying by -v:

$$\frac{\partial q}{\partial t}=-v\frac{\partial q}{\partial w}=\frac{-vh\,EI}{2p\,l\,r}\exp\left[\frac{-v(w+r)}{2k}\right]x\left[\frac{-w}{r^2}-\frac{v}{2k}\left(1+\frac{w}{r}\right)\right] \qquad \textbf{(EQ 34)}$$

Clearly, the only way that $\partial\theta/\partial t$ can be equal to zero is if

$$\frac{-w}{r^2}-\frac{v}{2k}\left(1+\frac{w}{r}=0\right) \qquad \textbf{(EQ 35)}$$

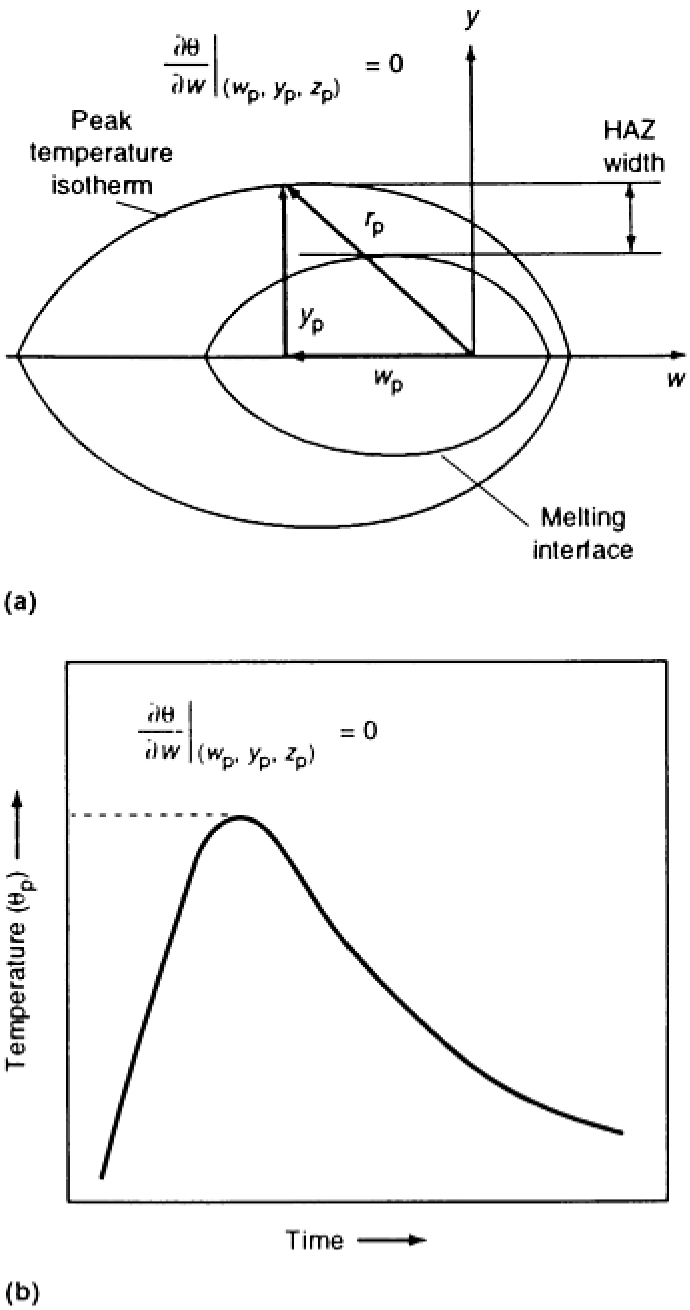

$\left.\frac{\partial\theta}{\partial w}\right|_{(w_p,\, y_p,\, z_p)} = 0$
Peak temperature isotherm
y
HAZ width
r_p
y_p
w_p
w
Melting interface
(a)
$\left.\frac{\partial\theta}{\partial w}\right|_{(w_p,\, y_p,\, z_p)} = 0$
Temperature (θ_p)
Time
(b)

FIG. 4 SCHEMATIC SHOWING PEAK TEMPERATURE AT (W_P, Y_P, Z_P) WITH W_P TO BE DETERMINED FOR A GIVEN PEAK TEMPERATURE VALUE FOR A GIVEN (Y_P, Z_P) LOCATION. (A) ISOTHERMS. (B) TEMPERATURE HISTORY

Equation 35 describes the relationship that must exist between the two location variables, *r* and *w*, for the temperature at the point to be equal to the peak temperature. If this expression were to be substituted into the temperature distribution equation for thick plates (Eq 21) and solved for *w* and *r* (two equations and two unknowns), then the location of the peak temperature could be determined in terms of *w* and *r*. The location given by *r* and *w* would be easily converted to *y* and *z* as:

$$R^2 = W^2 + Y^2 + Z^2 \qquad \textbf{(EQ 36)}$$

Such a solution for *r* and *w* is not explicitly possible, however, because the equations for *r* and *w* that result are not explicit. Consequently, iterative techniques are required, resulting in a solution that is both cumbersome and time-consuming.

One method of obtaining a simpler thick-plate peak-temperature equation is to assume that the heat input is from an instantaneous line on the surface of the plate, rather than from a moving point source (that is, $v \to \infty$). This allows the elimination of the time dependency in the peak-temperature evaluation. Using this assumption, the temperature distribution is given by:

$$q - q_0 = \frac{hEI}{2plt} \exp\left(\frac{r^2}{4kt}\right) \qquad \textbf{(EQ 37)}$$

Again, to find the peak-temperature location, $\partial \theta / \partial t$ is set equal to zero and the equation is solved for *r*. The result is:

$$\frac{r^2}{4kt} = 1 \qquad \textbf{(EQ 38)}$$

Substituting Eq 38 into Eq 37 yields the peak-temperature expression:

$$\frac{1}{q - q_0} = \left(\frac{erC_p p r^2}{2}\right) / \left(\frac{hEI}{v}\right) \qquad \textbf{(EQ 39)}$$

It has been found that Eq 39 gives results that are too high, but that the slope of $1/(\theta_p - \theta_0)$ versus r^2 is accurate. To rectify this situation, Eq 39 is forced to fit experimental results by specifying a known temperature/location condition. When this is done, Eq 39 becomes:

$$\frac{1}{q_p - q_0} = \left(\frac{[erC_p p(r^2 - r_t^2)]/2}{\left(\frac{hEI}{v}\right)}\right) + \frac{1}{q_r - q_0} \qquad \textbf{(EQ 40)}$$

where θ_r and r_r are the reference temperature and distance. If the peak temperature (θ_p) evaluation is restricted to locations on the plate surface ($z = 0$), and if the reference temperature and distance are assumed to be the melting temperature and the distance from the weld centerline to the fusion boundary (one half of the weld bead width), then Eq 40 can be written as:

$$\frac{1}{q_p - q_0} = \left(\frac{\frac{erC_p p\left(y^2 - \left(\frac{d}{2}\right)^2\right.}{2}}{\left(\frac{hEI}{v}\right)}\right) + \frac{1}{q_m - q_0} \qquad \textbf{(EQ 41)}$$

where θ_m is the melting temperature and *d* is the weld bead width. This equation gives the peak temperature θ_p in a thick plate at a distance *y* from the weld centerline.

Solidification Rate. The weld solidification structure can be determined by using the constitutional supercooling criterion. Three thermal parameters that influence the solidification structure are temperature gradient normal to the solid-liquid interface, *G* (°C/cm), solidification rate of the interface, *R* (cm/s), and cooling rate at the interface, $d\theta/dt$ at melting temperature (°C/s and equal to the product of *GR*). The microstructure may change from being planar to being cellular, a columnar dendrite, or an equiaxial structure if the *G*/*R* ratio becomes smaller. The dendrite arm spacing will decrease as the cooling rate increases. The solidification structure becomes refined at higher cooling rates.

At a quasi-steady state, the weld pool solidifies at a rate that is equal to the component of the electrode travel speed normal to the solid-liquid interface. Therefore, the solidification rate varies along the solid-liquid interface from the electrode travel speed, at the weld trailing edge, to zero, at the maximum pool width. The temperature gradient and the cooling rate at the solid-liquid interface can be determined from Eq 21, 24, and 27.

Modified Temperature Solution. The temperature solutions have a singularity at the center of the heat source. This singularity causes the predicted temperatures to be inaccurate in the area surrounding the heat source. However, a condition exists in which the peak temperature along the weld bead edge, that is, the solid-liquid interface location at the maximum pool width, is the melting temperature of the material. Using this temperature condition as a boundary condition for the temperature solutions, Eq 21 and 24 can be modified as shown below for thin plate:

$$q - q_0 = B_z \exp\left(\frac{-vw}{2k}\right) K_0\left(\frac{-vr}{2k}\right) \qquad \textbf{(EQ 42)}$$

where B_z is a heat input constant to be determined from the weld bead width, *d.*

$$B_z = \frac{q_m - q_0 \exp(vr_B / 2k)\left[\dfrac{K_0(vr_B / 2k)}{K_1(vr_B / 2k)}\right]}{K_0(vr_B / 2k)} \qquad \textbf{(EQ 43)}$$

with a bead width of:

$$d = 2r_B \left\{1 - \left[\frac{K_0(vr_B / 2k)}{K_1(vr_B / 2k)}\right]^2\right\}^{0.5} \qquad \textbf{(EQ 44)}$$

and for thick plate:

$$q_m - q_0 = B_z \exp\left[\frac{-v(w+r)}{2k}\right] / r \qquad \textbf{(EQ 45)}$$

$$B_z = (q_m - q_0) r_B \exp\left(\frac{-vr_B}{2k}\right)\left[\frac{1}{1+(2k / vr_B)} - 1\right] \qquad \textbf{(EQ 46)}$$

with a bead width of:

$$d = 2r_B \left[1 - \frac{1}{1+(2k / vr_B)}\right]^{0.5} \qquad \textbf{(EQ 47)}$$

The welding heat input, *Q,* is replaced by the weld bead width.

A Practical Application of Heat Flow Equations (Ref 22). The thermal condition in and near the weld metal must be established to control the metallurgical events in welding. The particular items of interests are the:

- DISTRIBUTION OF PEAK TEMPERATURE IN THE HAZ
- COOLING RATES IN THE WELD METAL AND IN THE HAZ
- SOLIDIFICATION RATE OF THE WELD METAL

Although the following discussion primarily focuses on manual are welding, certain general statements are applicable to all welding processes.

Peak Temperatures. The distribution of peak temperatures in the base metal adjacent to the weld is given by (Ref 23):

$$\frac{1}{T_p - T_0} = \frac{4.13 r C_p tY}{H_{met}} + \frac{1}{T_m - T_0} \qquad \textbf{(EQ 48)}$$

where T_p is the peak temperature (°C) at distance Y (mm) from the weld fusion boundary, T_0 is the initial temperature (°C), T_m is the melting temperature (°C), H_{net} is the net energy input equal to $\eta EI/v$ (J/s · mm), ρ is the density of the material (g/mm^3), C_p is the specific heat of solid metal (J/g · °C), and t is the thickness of the base metal (mm).

Equation 48 can be used in order to determine the:

- PEAK TEMPERATURES AT SPECIFIC LOCATIONS IN THE HAZ
- WIDTH OF THE HAZ
- EFFECT OF PREHEAT ON THE WIDTH OF THE HAZ

In addition, determination of the peak temperature at specific locations in the HAZ and the width of the HAZ can be obtained by the procedure described from Eq 34, 35, 36, 37, 38, 39, 40, and 41.

Cooling Rate. Because the cooling rate varies with position and time, its calculation requires the careful specification of conditions. The most useful method is to determine the cooling rate on the weld centerline at the instant when the metal passes through a particular temperature of interest, T_c. At a temperature well below melting, the cooling rate in the weld and in its immediate HAZ is substantially independent of position. For carbon and low-alloy steels, T_c is the temperature near the pearlite "nose" temperature on the time-temperature transformation (TTT) diagram. The value of T_c = 550 °C (1020 °F) is satisfactory for most steels, although not critical.

The cooling rate for thick plate (Ref 23) is:

$$R = \frac{2p\,l\,(T_c - T_0)^2}{H_{met}} \qquad \textbf{(EQ 49)}$$

Equation 49 is comparable to Eq 31, which was obtained by the procedure described in the section "Cooling Rate" in this article. Adams (Ref 24) has developed a cooling rate equation for thin plate along the centerline from Eq 24.

The cooling rate for thin plate (Ref 24) is:

$$R = 2p\,l\,r\,C_p \left(\frac{t}{H_{met}} \right)^2 + (T_c - T_0)^3 \qquad \textbf{(EQ 50)}$$

where R is the cooling rate (°C/s) at a point on the weld centerline at just that moment when the point is cooling past the T_c, and λ is the thermal conductivity of the metal (J/mm · s · °C).

The dimensionless quantity t, called the "relative plate thickness," can be used to determine whether the plate is thick or thin:

$$t = t\sqrt{\frac{r\,C_p (T_c - T_0)}{H_{met}}} \qquad \textbf{(EQ 51)}$$

The thick-plate equation applies when t is greater than 0.75, and the thin-plate equation applies when t is less than that value.

Equations 49 and 50 are used to determine the cooling rate along the centerline for thick plate and thin plate, respectively. If one is interested in the cooling rate at the location at distance *y* (in mm) from the centerline, iterative techniques should be used to solve the cooling rate. First, *w* and *r* can be obtained by iteration of the simultaneous equation, which consists of Eq 21 or 27, where θ equals θ_c and $r^2 = w^2 + y^2$, where *y* is given. Then substitute *w* and *r* into the differentiation, $\partial\theta/\partial t = -v\partial\theta/\partial w$, from the temperature from Eq 21 or 27. The result will be the cooling rate for thick or thin plate located at *y* distance from the centerline:

$$\frac{\partial q}{\partial t}_{qc}$$

In addition, the cooling rate for HAZ of low-carbon steel weldments can be obtained from Eq 32 directly.

The solidification rate can have a significant effect on metallurgical structure, properties, response to heat treatment, and soundness. The solidification time, S_t, of weld metal, measured in seconds, is:

$$S_t = \frac{LH_{net}}{2p\,l\,C_p(T_m - T_o)^2} \qquad \textbf{(EQ 52)}$$

where *L* is the heat of fusion (J/mm^3).

Example 1: Welding of 5 mm (0.2 in.) Thick Low-Carbon Steels.

The thermal properties needed for heat flow analysis are assumed to be:

MELTING TEMPERATURE (T_{M}), °C (°F)	1510 (2750)
AUSTENIZATION TEMPERATURE, °C (°F)	730 (1350)
THERMAL CONDUCTIVITY (λ), W/M · K (J/MM · S · °C)	11.7 (0.028)
VOLUMETRIC SPECIFIC HEAT (ρC_{P}), J/MM^3 · °C	0.0044
HEAT OF FUSION (*L*), J/MM^3	2

The welding condition is assumed to be:

CURRENT (*I*), A	200
ARC VOLTAGE (*E*), V	20
TRAVEL SPEED (*V*), MM/S (IN./S)	5 (0.2)
PREHEAT (T_0), °C (°F)	25 (77)
HEAT-TRANSFER EFFICIENCY (η)	0.9
NET ENERGY INPUT, H_{NET}, J/MM (KJ/IN.)	720 (18.3)

Calculation of the HAZ Width. The value of *Y* at T_p = 730 °C (1345 °F) must be determined from Eq 48:

$$\frac{1}{730-25} = \frac{4.13(0.0044)5Y_z}{720}\frac{1}{1510-25} \qquad \textbf{(EQ 53)}$$

resulting in a value for Y_z (the width of the HAZ) of 5.9 mm (0.24 in.).

In addition, the HAZ width can be obtained by the procedure described in Eq 34, 35, 36, 37, 38, 39, 40, and 41. By substituting $w = w_p$, $r = r_p$, $y = y_p$, and $z = 0$ (on surface) into Eq 35 and 36, we obtain:

$$\frac{-w_p}{r_p^2} - \frac{v}{2k}\left(1+\frac{w_p}{r_p}\right) = 0$$ **(EQ 54)**

$$r_p^2 = w_p^2 + y_p^2$$ **(EQ 55)**

The substitution of Eq 55 in Eq 54 yields:

$$\frac{-w_p}{w_p^2 + y_p^2} = \frac{v}{2k} 1 + \frac{w_p}{\sqrt{w_p^2 + y_p^2}}$$ **(EQ 56)**

Inserting $v = 5$ mm/s, $\kappa = \lambda/\rho c_p = 0.028/0.004$ into Eq 56:

$$\frac{-w_p}{w_p^2 + y_p^2} = 0.3831 + \frac{w_p}{\sqrt{w_p^2 + y_p^2}}$$ **(EQ 57)**

w_p and y_p should satisfy Eq 57.

From the temperature distribution equation (Eq 21) and $E = 10$V, $I = 200$ A, $\eta = 0.9$, $\lambda = 0.028$, $\theta_p = 730$ °C, $\theta_0 = 25$ °C, $v = 5$ mm/s, $\kappa = \lambda/\rho c_p = 0.028/0.044$:

$$730 - 25 = \frac{0.9x20x200/5}{2p\,x0.028r_p} x \exp\left[\frac{-5(w_p + r_p)}{2x0.028/0.0044}\right]$$ **(EQ 58)**

$$-1.76 + \mathrm{LN}R_P = -0.393\,(W_P + R_P)$$ **(EQ 59)**

Solving Eq 21 using the above values, by iteration:

- ASSUME $R_P = 3.4$ AND PUT THIS VALUE OF R_P INTO EQ 59 TO OBTAIN $W_P = -2.036$. SUBSTITUTE $R_P = 3.4$ AND $W_P = -2.036$ INTO EQ 55 AND SOLVE FOR Y_P: $Y_P = 2.7$. SUBSTITUTE Y_P AND W_P INTO EQ 56: THE LEFT SIDE OF EQ 56 = 0.176; THE RIGHT SIDE OF EQ 56 = 0.158. THE RESULTS DO NOT SATISFY EQ 56.
- ASSUME $R_P = 3.5$ AND PUT THIS VALUE OF R_P INTO EQ 59 TO OBTAIN $W_P = -2.2$. SUBSTITUTE $R_P = 3.5$ AND $W_P = -2.2$ INTO EQ 55 AND SOLVE FOR Y_P: $Y_P = 2.72$. SUBSTITUTE Y_P AND W_P INTO EQ 56: THE LEFT SIDE OF EQ 56 = 0.179; THE RIGHT SIDE OF EQ 56 = 0.146. THE RESULTS DO NOT SATISFY EQ 56.
- ASSUME $R_P = 3.3$ AND PUT THIS VALUE OF R_P INTO EQ 59 TO OBTAIN $W_P = -1.86$. SUBSTITUTE $R_P = 3.3$ AND $W_P = -1.86$ INTO EQ 55 AND SOLVE FOR Y_P: $Y_P = 2.726$. SUBSTITUTE Y_P AND W_P INTO EQ 56: THE LEFT SIDE OF EQ 56 = 0.170; THE RIGHT SIDE OF EQ 56 = 0.171. EQUATION 56 IS NOW SATISFIED.

The full HAZ width equals $2y_p = 2 \times 2.726 = 5.5$ mm. Comparing this result, 5.5 mm, with the result obtained with Eq 48 (that is, 5.9 mm), we know that Eq 48 can be used to obtain accurate results when calculating HAZ width and peak temperature.

Effect of Tempering Temperature on Quenched and Tempered (Q&T) Steels. If the plate had been quenched and then tempered to 430 °C (810 °F), then any region heated above that temperature will have been "over-tempered" and may exhibit modified properties. It would then be reasonable to consider the modified zone as being "heat affected," with its outer extremity located where T_p = 430 °C (810 °F):

$$\frac{1}{430-25}=\frac{4.13(0.0044)5Y_z}{720}+\frac{1}{1510-25} \qquad \textbf{(EQ 60)}$$

resulting in a value for Y_z of 14.2 mm (0.568 in.).

Effect of Preheating Temperature on Q&T Steels. Assume that the Q&T steel described above was preheated to a temperature, T_0, of 200 °C (390 °F):

$$\frac{1}{430-200}=\frac{4.13(0.0044)5Y_z}{720}+\frac{1}{1510-200} \qquad \textbf{(EQ 61)}$$

resulting in a value for Y_z of 28.4 mm > 14.2 mm (1.14 > 0.568 in.). Therefore, increasing the preheating temperature will increase the value of Y_z.

Effect of Energy Input on Q&T Steels. Assume that the energy input into the Q&T steel (without preheating) increases 50% (that is, 1.08 kJ/mm, or 27.4 kJ/in.):

$$\frac{1}{430-200}=\frac{4.13(0.0044)5Y_z}{1080}+\frac{1}{1510-25} \qquad \textbf{(EQ 62)}$$

resulting in a value for Y_z of 21.3 mm > 14.2 mm (0.839 > 0.568 in.). Therefore, increasing the energy input will increase the value of Y_z.

Example 2: Welding of 6 mm (0.24 in.) Thick Low-Carbon Steels.

The thermal properties needed for heat-flow analysis are assumed to be:

MELTING TEMPERATURE (T_M), °C (°F)	1510 (2750)
AUSTENIZATION TEMPERATURE, °C (°F)	730 (1346)
THERMAL CONDUCTIVITY (l), W/M · K (J/MM · S · °C)	11.7 (0.028)
VOLUMETRIC SPECIFIC HEAT (r C_P), J/MM^3 · °C	0.0044
HEAT OF FUSION (L), J/MM^3	2

The welding condition is assumed to be:

CURRENT (I), A	**300**
ARC VOLTAGE (E), V	**25**
PREHEAT (T_0), °C (°F)	**25 (77)**
HEAT-TRANSFER EFFICIENCY (η)	**0.9**

Critical Cooling Rate at 550 °C (1020 °F) (T_c). A critical cooling rate exists for each steel composition. If the actual cooling rate in the weld metal exceeds this critical value, then hard martensitic structures may develop in the HAZ, and there is a great risk of cracking under the influence of thermal stresses in the presence of hydrogen.

The best way to determine the critical cooling rate is to make a series of bead-on-plate weld passes in which all parameters, except the arc travel speed, are held constant. After the hardness tests on the weld passes deposited at travel speeds of 6, 7, 8, 9 and 10 mm/s (0.23, 0.28, 0.32, 0.35, and 0.39 in./s), it was found that at the latter two travel speeds, the weld HAZ had the highest hardness. Therefore, the critical cooling rate was encountered at a travel speed of approximately 8 mm/s (0.32 in.s). At this speed, the net energy input is:

$$H_{net} = \frac{25(300)0.9}{8} = 8.43.75 J/mm \qquad \textbf{(EQ 63)}$$

From Eq 51, the relative plate thickness is:

$$t = 6\sqrt{\frac{0.0044(550-25)}{843.75}} = 0.31 \qquad \textbf{(EQ 64)}$$

Because t is less than 0.75, the thin-plate equation (Eq 50) applies:

$$\frac{R}{2pl} = 0.0044\left(\frac{6}{843.75}\right)^2 (550-25)^3 = 32.2 \qquad \textbf{(EQ 65)}$$

resulting in R being equal to 2π(0.028)32.2, which is equal to 5.7 °C/s (10.3 °F/s). This value is the maximum safe cooling rate for this steel and the actual cooling rate cannot exceed this value.

Preheating Temperature Requirement. Although the critical cooling rate cannot be exceeded, in the actual welding operation a preheat can be used to reduce the cooling rate to 5.7 °C/s (10.3 °F/s).

Assume that the welding condition is:

CURRENT (*I*), A	250
ARC VOLTAGE (*E*), V	25
HEAT-TRANSFER EFFICIENCY (η)	0.9
TRAVEL VELOCITY (*V*), MM/S (IN./S)	7 (0.3)
PLATE THICKNESS (*T*), MM (IN.)	9 (0.4)

The energy heat input, H_{net}, is:

$$H_{net} = \frac{25(250)0.9}{7} = 804 J/mm \qquad \textbf{(EQ 66)}$$

Assuming that the thin-plate equation (Eq 50) applies:

$$\left(\frac{R}{2pl}\right)_{\max} = 32.2 = 0.0044\left(\frac{9}{804}\right)^2 (550-T_0)^3 \qquad \textbf{(EQ 67)}$$

resulting in a T_0 of 162 °C (325 °F).

The relative plate thickness should be checked:

$$t = 9\sqrt{\frac{0.0044(550-162)}{804}} = 0.41 \qquad \textbf{(EQ 68)}$$

Because t is less than 0.75, the thin-plate equation does apply. If the initial plate temperature is raised either to or above 162 °C (325 °F), then the cooling rate will not exceed 5.7 °C/s (10.3 °F/s).

Effect of Joint Thickness. If the plate thickness increases from 9 to 25 mm (0.36 to 1 in.), but there is the same level of energy input, then the calculation of the initial plate temperature would be as follows. First, using the thin-plate equation (Eq 50):

$$\left(\frac{R}{2pl}\right)_{\max} = 32.2 = 0.0044\left(\frac{25}{804}\right)^2 (550-T_0)^3 \qquad \textbf{(EQ 69)}$$

resulting in a value for T_0 of 354 °C (670 °F).

The relative plate thickness, t, should be checked:

$$t = 25\sqrt{\frac{0.0044(550-354)}{804}} = 0.82 \qquad \textbf{(EQ 70)}$$

Because t is greater than 0.75, the use of the thin-plate equation is inadequate. Using the thick-plate equation (Eq 49):

$$32.2 = \frac{(550-T_0)^2}{804} \qquad \textbf{(EQ 71)}$$

resulting in a value for T_0 of 389 °C (730 °F).

The relative plate thickness should be checked:

$$t = 25\sqrt{\frac{0.0044(550-389)}{804}} = 0.74 \qquad \textbf{(EQ 72)}$$

Although t is less than, but near to, 0.75, using the thin-plate equation is adequate. Therefore, the initial temperature should be raised to 389 °C (730 °F) to avoid exceeding the cooling rate of 5.7 °C/s (10.3 °F).

Now, if the plate thickness increases to 50 mm (2 in.), but there is the same level of energy input, then the thick-plate equation (Eq 49) applies and, again, the value for T_0 is 389 °C (730 °F).

The relative plate thickness should be checked:

$$t = 50\sqrt{\frac{0.0044(550-389)}{804}} = 1{,}48 \qquad \textbf{(EQ 73)}$$

Because t is greater than 0.75, the use of the thick-plate equation is adequate.

Under some welding conditions, it is not necessary to reduce the cooling rate by using a preheat. For example, if the plate thickness is 5 mm (0.2 in.) and there is the same level of energy input:

$$\left(\frac{R}{2pl}\right)_{\max} = 32.2 = 0.0044\left(\frac{5}{804}\right)^2 (550 - T_0)^3 \qquad \textbf{(EQ 74)}$$

resulting in a value for T_0 of -24 °C (-11 °F). Therefore, using a preheat is unnecessary.

Fillet-Welded "T" Joints. For a weld with a higher number of paths, as occurs in fillet-welded "T" joints, it is sometime necessary to modify the cooling-rate equation, because the cooling of a weld depends on the available paths for conducting heat into the surrounding cold base metal.

When joining 9 mm (0.35 in.) thick plate, where H_{net} = 804 J/mm (20.4 kJ/in.), and when there are three legs instead of two, the cooling-rate equation is modified by reducing the effective energy input by a factor of $\frac{2}{3}$:

$$H_{\text{NET}} = \frac{2}{3}(804) = 536 \text{ J/MM} \qquad \textbf{(EQ 75)}$$

Using the thin-plate equation (Eq 50):

$$\left(\frac{R}{2pl}\right)_{\max} = 32.2 = 0.0044\left(\frac{9}{536}\right)^2 (550 - T_0)^3 \qquad \textbf{(EQ 76)}$$

resulting in a value for T_0 of 254 °C (490 °F).

The relative plate thickness should be checked:

$$t = 9\sqrt{\frac{0.0044(550-254)}{536}} = 0{,}44 \qquad \textbf{(EQ 77)}$$

Because t is less than 0.75, using the thin-plate equation is adequate. Therefore, a higher preheat temperature is more necessary than a butt weld because of the enhanced cooling.

Example 3: Cooling Rate for the Location at Distance *y* (in cm) from the Centerline.

For a steel plate of 25 mm (1 in.) thickness, (t), the welding condition is assumed to be:

HEAT INPUT (ηEI), KW (CAL/S)	7.5 (1800)
TRAVEL SPEED (V), CM/S (IN./S)	0.1 (0.04)
PREHEAT (T_0), °C (°F)	20 (68)
NET ENERGY INPUT (H_{NET}), CAL/CM	18,000

The thermal properties needed for heat flow analysis are assumed to be:

MELTING TEMPERATURE (T_M), °C (°F)	1400 (2550)
THERMAL CONDUCTIVITY (λ), W/M · K (CAL_{IT}/CM · S · °C)	43.1 (0.103)
SPECIFIC HEAT (C_P), J/KG · C (BTU/LB · °F)	473 (0.113)
DENSITY (ρ), G/CM^3 (LB/IN.^{3})	7.8 (0.28)

Assume that one is interested in the critical cooling rate at the location on the surface (z = 0) at distance y = 2 cm from the centerline at the instant when the metal passes through the specific temperature of 615 °C (1140 °F). Initially, the relative plate thickness should be checked. From Eq 51, the relative plate thickness (using English units) is:

$$t = 25\sqrt{\frac{7.8x0.113(615-20}{18,000}} = 4,267$$

Because t is greater than 0.75, this plate can be treated as a thick plate. From Eq 78, cooling rate for thick plate at the location where the variables are w and r and at critical temperature θ= θ_c is:

$$\frac{\partial q}{\partial t} = -v\frac{\partial q}{\partial w} = \frac{-vhEI}{2p\,l\,r}\exp\left[\frac{-v(w+r)}{2k}\right]x\left[\frac{-w}{r^2} - \frac{v}{2k}\left(1+\frac{w}{r}\right)\right] \qquad \textbf{(EQ 78)}$$

To solve Eq 78, we need to calculate the value of w and r first. From Eq 21, the temperature distribution of thick plate, and Eq 36 where z = 0, we can get:

$$q - q_0 = \frac{hEI}{2p\,l\,r}\exp\left(\frac{-v(w+r)}{2k}\right) \qquad \textbf{(EQ 79)}$$
$$\textit{and} \quad \mathrm{r}=\sqrt{\mathrm{w}^2+\mathrm{y}^2}$$

By substituting the welding condition and material properties into Eq 21, Eq 36, and r = $\sqrt{w^2+y^2}$, we can obtain the following simultaneous equations:

$$615-20 = \frac{1800}{22p\,x0.103r}\exp\left[\frac{-0.1(w+r)}{2x0.117}\right]$$
$$\textit{and} \quad \mathrm{r}=\sqrt{\mathrm{w}^2+4}$$

The value of w and r can be solved by using iteration techniques to solve the above simultaneous equation. The result is that w = -3 cm and r = 3.606 cm.

Substituting w and r into Eq 78:

$$\frac{\partial q}{\partial t}_{qC} = -v\frac{\partial q}{\partial w} = \frac{-0.1x1800}{2p\,x0.103x3.606}x\exp\left[\frac{-0.1(-3+3.606}{2x0.117}\right]x\left[\frac{-(-3)}{3.606^2} - \frac{0.1}{2x0.117}\left(1+\frac{-3}{3.606}\right)\right]$$

Therefore:

$$\left.\frac{\partial q}{\partial t}\right|_{qC} = -9.78°C/s$$

From Eq 49 we can calculate the cooling rate along the centerline at the same temperature (615 °C, or 1140 °F):

$$R = \frac{2p\,x0.103(615-20)^2}{18,000} = 12.7°C/s$$

Also from Eq 32, we can calculate the cooling rate in the heat-affected zone at a temperature of 615 °C (1140 °F):

$$\frac{\partial q}{\partial t}_{qC=615} = 0.35\left\{\left[\frac{615-20}{336/2.362}\right]^{1.7} x\left[1+\frac{2}{p}\tan^{-1}\left(\frac{0.9843-0.486}{0.125}\right)\right]\right\}^{0.8} = 4°C/s$$

Therefore, at the same temperature, the cooling rate at the centerline is greater than the cooling rate at the location a distance *y* from the centerline. In addition, the cooling rate of the heat-affected zone is less than the cooling rate in the weld pool at the same temperature.

Example 4: Solidification Rate.

A weld pass of 800 J/mm (20.3 kJ/in.) in net energy input is deposited on a steel plate. The initial temperature is 25 °C (75 °F). The solidification time would be:

$$S_t = \frac{2(800)}{2p\,(0.028)(0.0044)(1510-25)^2} = 0.94(s)$$ **(EQ 80)**

References cited in this section

22. HEAT FLOW IN WELDING, CHAPTER 3, *WELDING HANDBOOK,* VOL 1, 7TH ED., AWS, 1976
23. C.M. ADAMS, JR. COOLING RATE AND PEAK TEMPERATURE IN FUSION WELDING, *WELD. J.,* VOL 37 (NO. 5), 1958, P 210S-215S
24. C.M. ADAMS, JR., COOLING RATES AND PEAK TEMPERATURES IN FUSION WELDING, *WELD. J.,* VOL 37 (NO. 5), P 210-S TO 215-S
25. H. KIHARA, H. SUZUKI, AND H. TAMURA, *RESEARCH ON WELDABLE HIGH-STRENGTH STEELS,* 60TH ANNIVERSARY SERIES, VOL 1, SOCIETY OF NAVAL ARCHITECTS OF JAPAN, TOKYO, 1957
26. C.E. JACKSON, DEPARTMENT OF WELDING ENGINEERING, THE OHIO STATE UNIVERSITY LECTURE NOTE, 1977
27. C.E. JACKSON AND W.J. GOODWIN, EFFECTS OF VARIATIONS IN WELDING TECHNIQUE ON THE TRANSITION BEHAVIOR OF WELDED SPECIMENS--PART II, *WELD. J.,* MAY 1948, P 253-S TO 266-S

Heat Flow in Fusion Welding

Chon L. Tsai and Chin M. Tso, The Ohio State University

Parametric Effects

To show the effects of material property and welding condition on the temperature distribution of weldments, the welding of 304 stainless steel, low-carbon steel, and aluminum are simulated for three different welding speeds: 1.0, 5.0, and 8.0 mm/s (0.04, 0.02, and 0.03 in./s). The thermal conductivity and thermal diffusivity of 304 stainless steel are 26 W/m · K (0.062 cal/cm · s · °C) and 4.6 mm^2/s (0.007 $in.^2/s$), respectively. For low-carbon steel, the respective values are 50 W/m K (0.12 cal/cm s °C) and 7.5 mm^2/s (0.012 $in.^2/s$), whereas for aluminum, the respective values are 347 W/m K (0.93 cal/cm · s · °C) and 80 mm^2 (0.12 $in.^2/s$). The heat input per unit weld length was kept constant, 4.2 kJ/s (1 kcal/s), for all cases. The parametric results are described below.

Effect of Material Type. Figures 5(a), 5(b), and 5(c) depict the effect of thermal properties on isotemperature contours for a heat input of 4.2 kJ/s (1 kcal/s) and travel speeds of 1.0, 5.0, and 8.0 mm/s (0.04, 0.02, and 0.3 in./s). The temperature spreads over a larger area and causes a larger weld pool (larger weld bead) for low-conductivity material. The isotemperature contours also elongate more toward the back of the arc for low-conductivity material. For aluminum, a larger heat input would be required to obtain the same weld size as the stainless steel weldment.

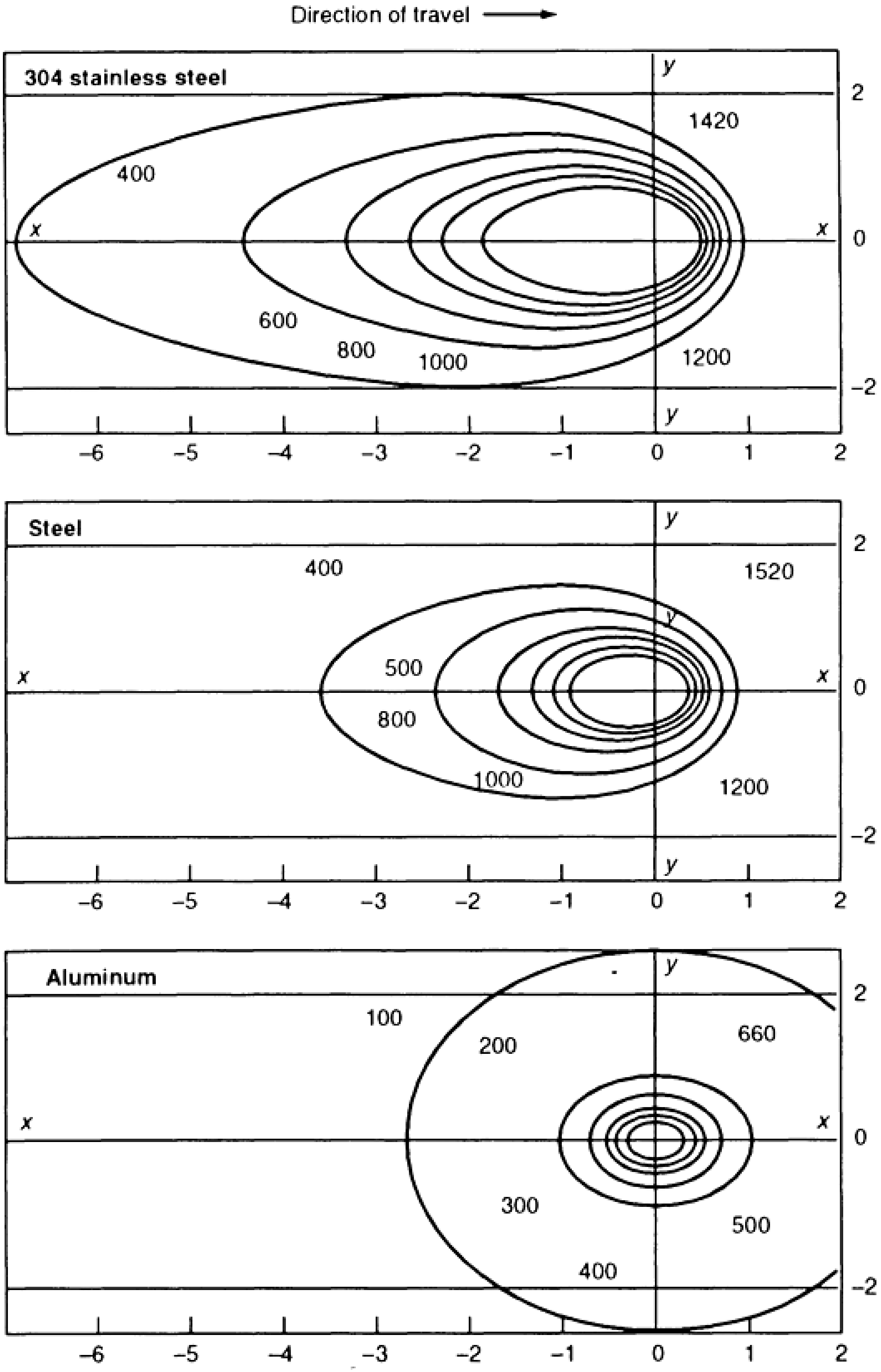

FIG. 5(A) EFFECT OF THERMAL PROPERTY ON ISOTEMPERATURE CONTOURS FOR A HEAT INPUT OF 4.2 KJ/S (1000 CAL/S) AT A WELDING SPEED, *V*, OF 1 MM/S (0.04 IN./S) AND THE RESPECTIVE THERMAL CONDUCTIVITIES OF EACH MATERIAL (REFER TO TEXT FOR VALUES). VALUES FOR *X* AND *Y* ARE GIVEN IN CM, AND TEMPERATURES ARE GIVEN IN °C.

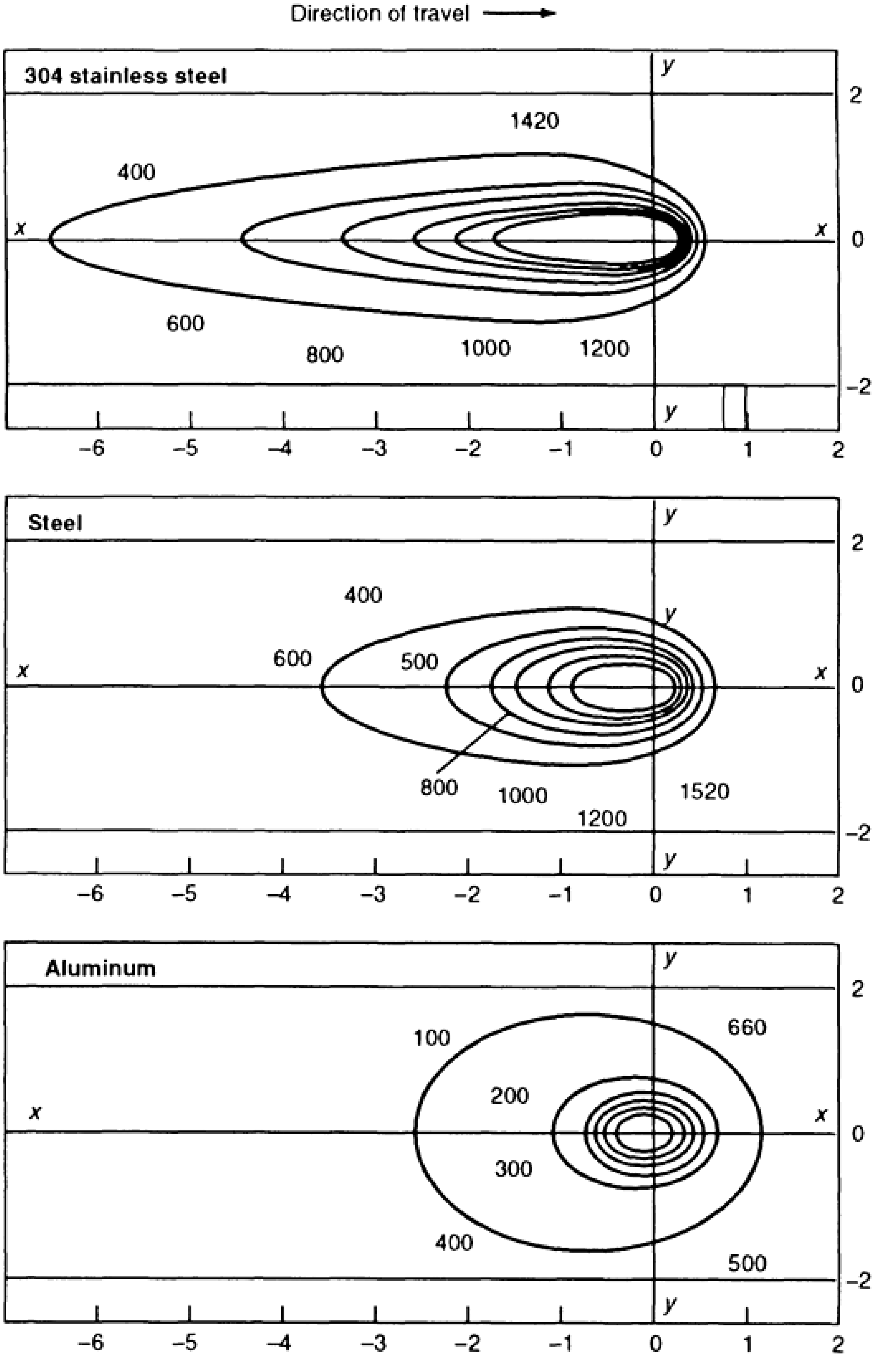

FIG. 5(B) EFFECT OF THERMAL PROPERTY ON ISOTEMPERATURE CONTOURS FOR A HEAT INPUT OF 4.2 KJ/S (1000 CAL/S) AT A WELDING SPEED, *V*, OF 5 MM/S (0.02 IN./S) AND THE RESPECTIVE THERMAL CONDUCTIVITIES OF EACH MATERIAL (REFER TO TEXT FOR VALUES). VALUES FOR *X* AND *Y* ARE GIVEN IN CM, AND TEMPERATURES ARE GIVEN IN °C.

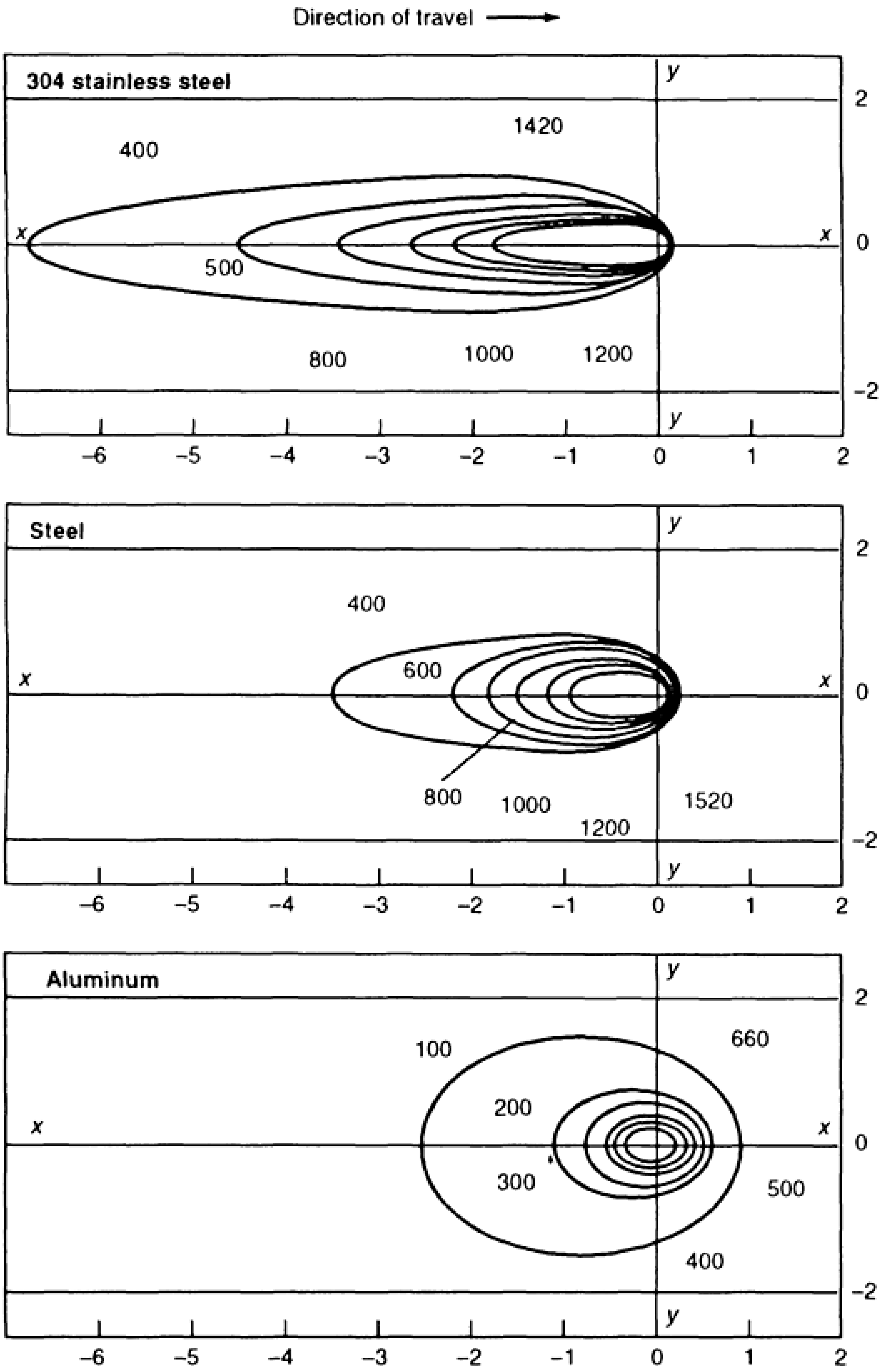

FIG. 5(C) EFFECT OF THERMAL PROPERTY ON ISOTEMPERATURE CONTOURS FOR A HEAT INPUT OF 4.2 KJ/S (1000 CAL/S) AT A WELDING SPEED, *V*, OF 8 MM/S (0.3 IN./S) AND THE RESPECTIVE THERMAL CONDUCTIVITIES OF EACH MATERIAL (REFER TO TEXT FOR VALUES). VALUES FOR *X* AND *Y* ARE GIVEN IN CM, AND TEMPERATURES ARE GIVEN IN °C.

Welding Speed. Figures 6(a), 6(b), and 6(c) show the effect of welding speed on isotemperature contours. When the travel speed increases, the weld size decreases and the isotemperature contours are more elongated toward the back of the arc. Larger heat inputs would be required for faster travel speeds in order to obtain the same weld size.

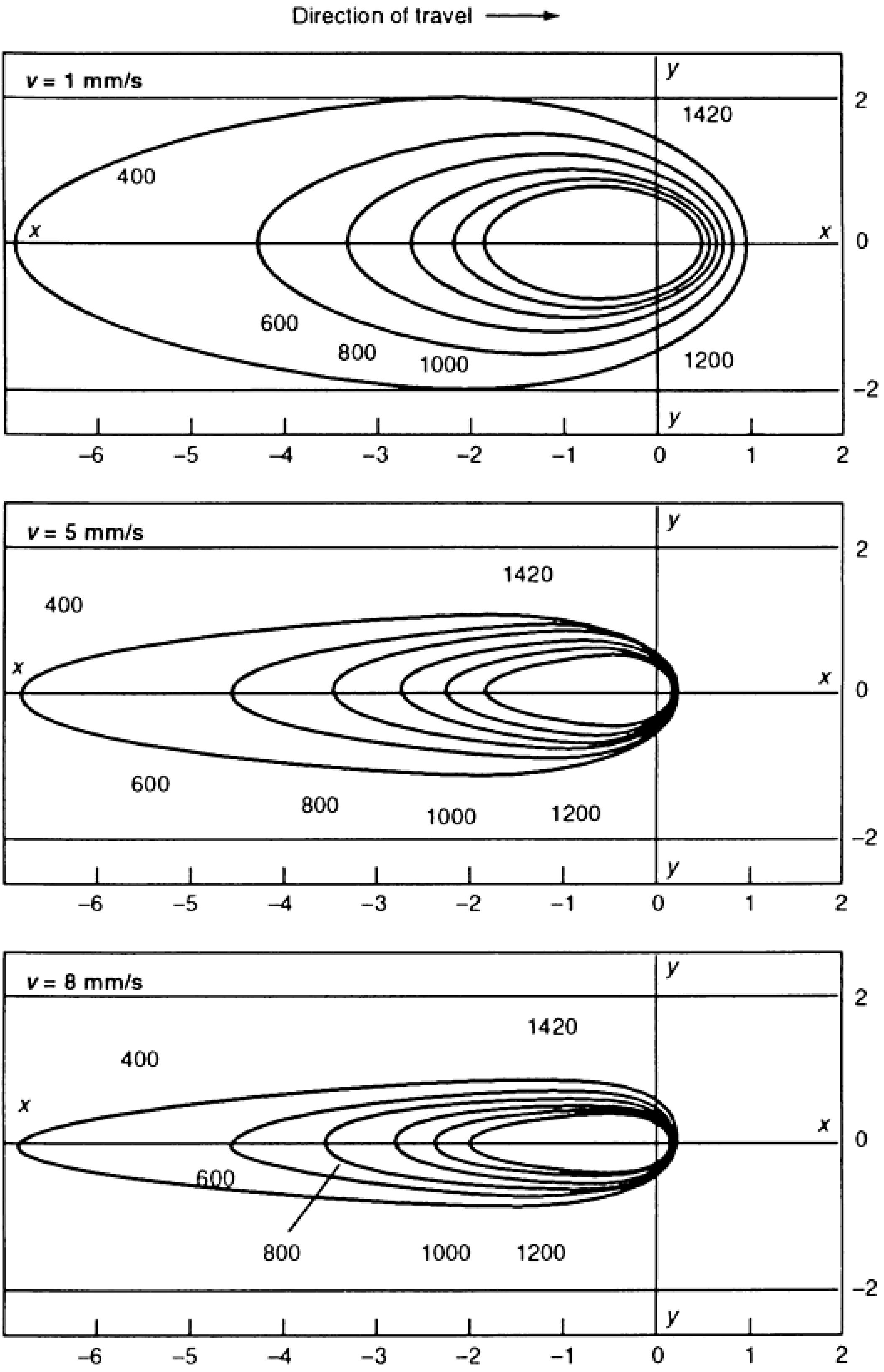

FIG. 6(A) EFFECT OF WELDING SPEED, *V*, ON ISOTEMPERATURE CONTOURS OF 304 STAINLESS STEEL FOR 4.2 KJ/S (1000 CAL/S) HEAT INPUT

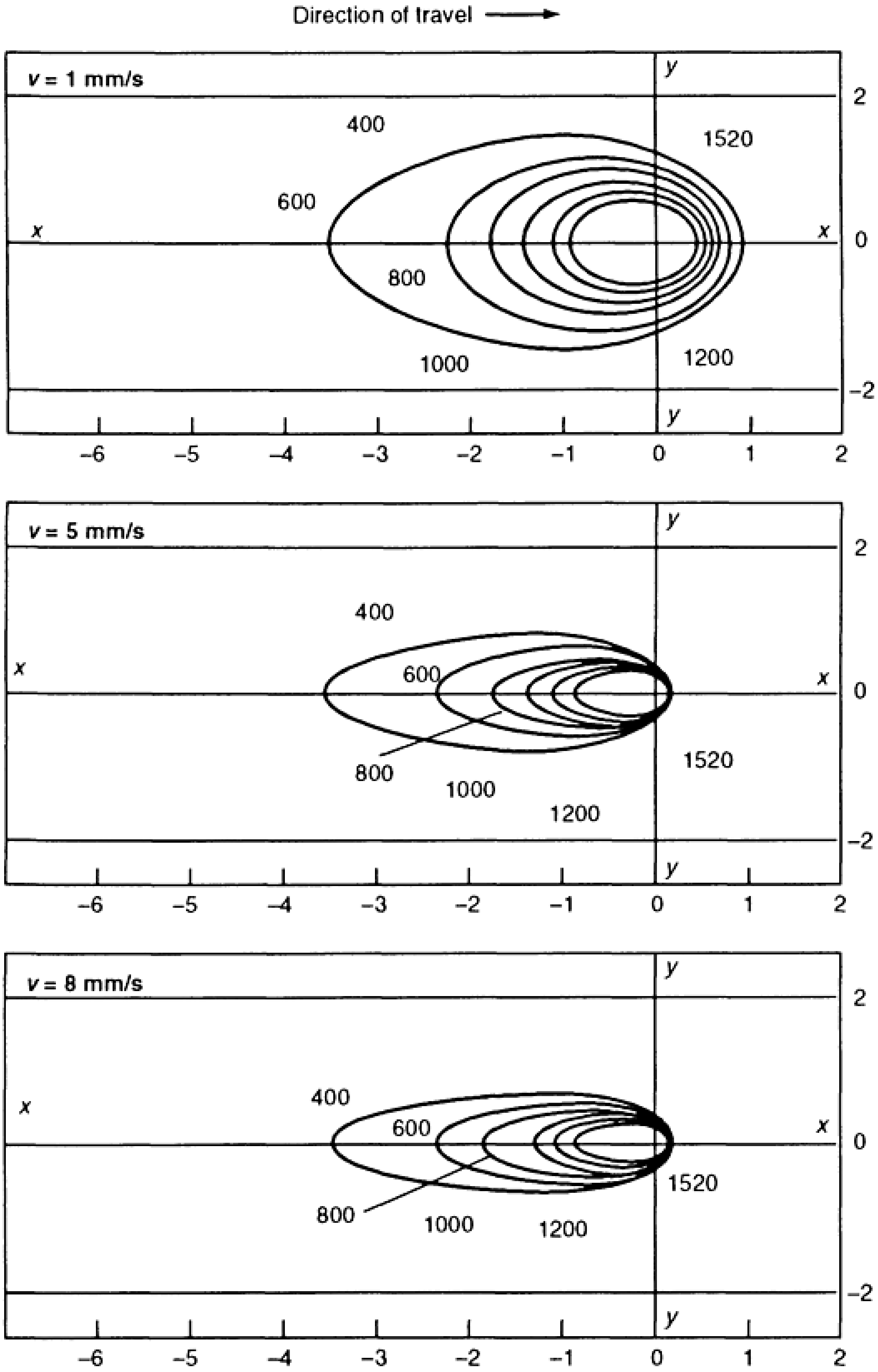

FIG. 6(B) EFFECT OF WELDING SPEED, *V*, ON ISOTEMPERATURE CONTOURS OF LOW-CARBON STEEL FOR 4.2 KJ/S (1000 CAL/S) HEAT INPUT

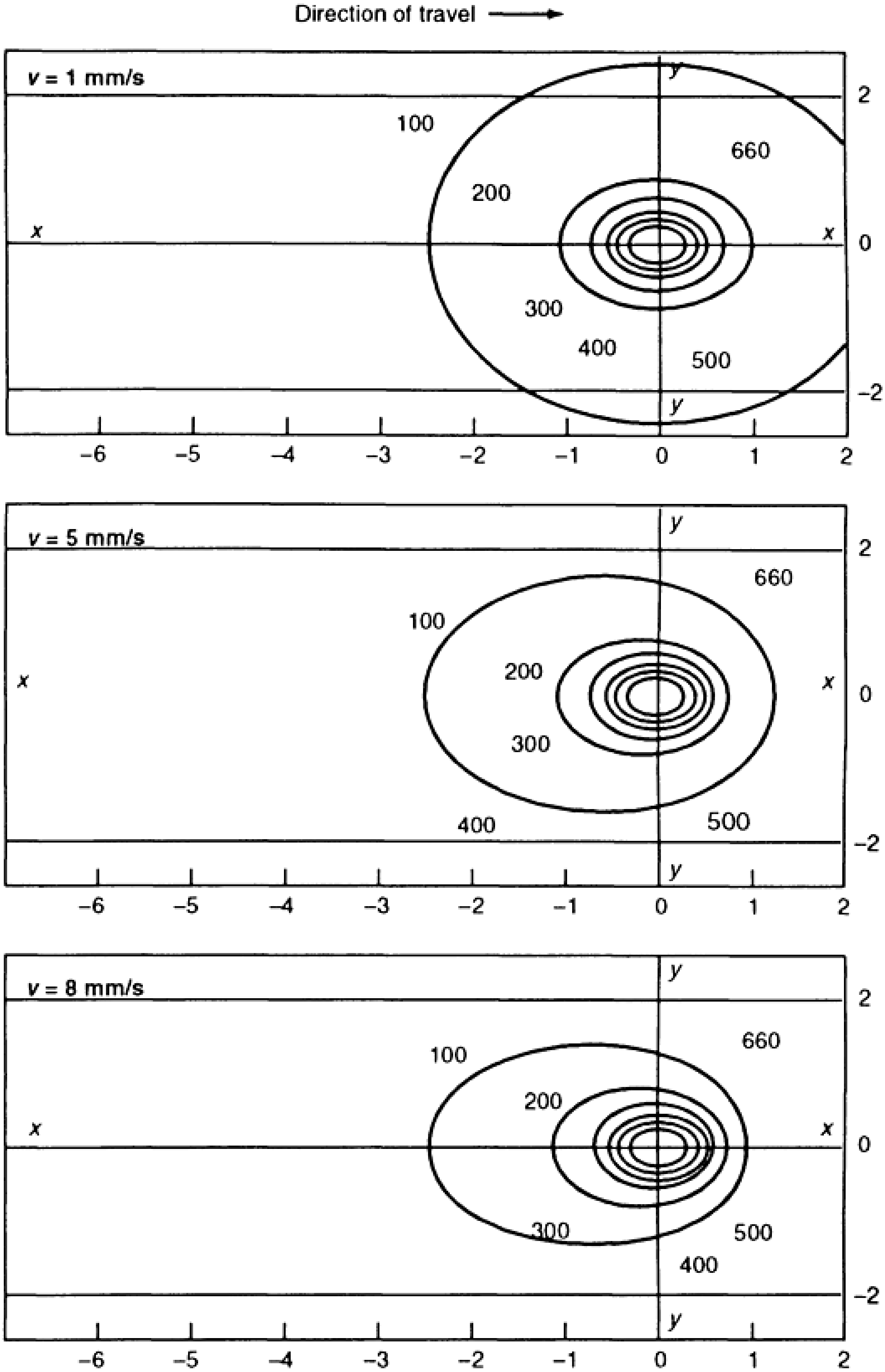

FIG. 6(C) EFFECT OF WELDING SPEED, *V*, ON ISOTEMPERATURE CONTOURS OF ALUMINUM FOR 4.2 KJ/S (1000 CAL/S) HEAT INPUT

Figure 7 shows the effect of welding heat input on the peak temperature at two locations, 6.4 mm ($\frac{1}{4}$ in.) and 13 mm ($\frac{1}{2}$ in.) from the weld centerline. (The material simulated in this illustration is low-carbon steel plate with a large thickness.) Within the practical range of welding conditions for the GMAW process of low-carbon steels, the peak temperature at both locations increases linearly, as the welding current increases, and decreases exponentially, with the travel speed. To increase welding current proportionally with travel speed for a constant heat input per unit weld length (that is. $\dot{Q}/V$ = constant), the peak temperature at two locations increases with the travel speed, which implies a larger weld size resulting from higher heat input. The welding current has a more significant effect on peak temperature than does the travel speed. However, the increase in peak temperature is more dominant at the 6.4 mm ($\frac{1}{4}$ in.) location. The influence of proportional increase in welding current and travel speed on peak temperature diminishes as the distance from the weld centerline increase and when the travel speed becomes high.

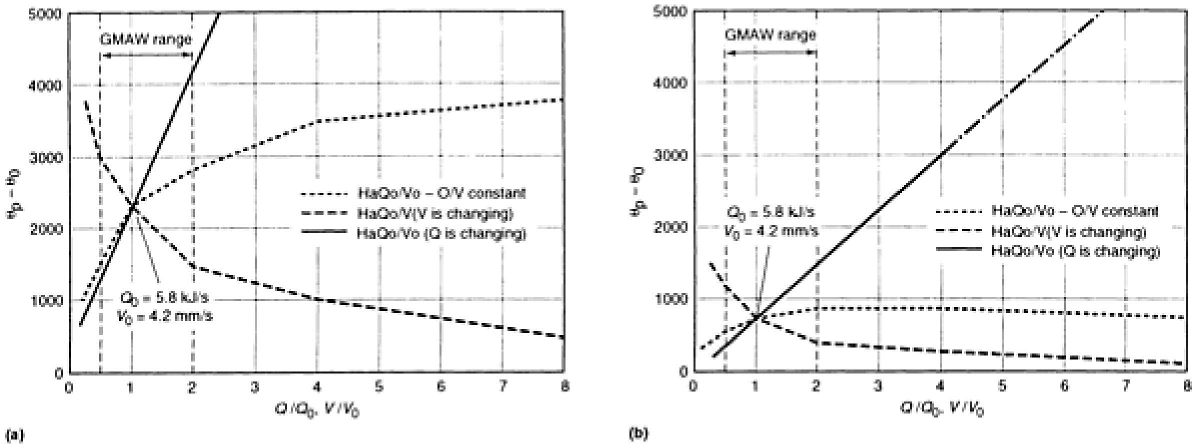

FIG. 7 PEAK TEMPERATURE DETERMINED BY THE POINT HEAT SOURCE SOLUTION. (A) AT 6.4 MM ($\frac{1}{2}$ IN.) FROM THE CENTER. (B) AT 13 MM ($\frac{1}{2}$ IN.) FROM THE CENTER

Heat Flow in Fusion Welding

Chon L. Tsai and Chin M. Tso, The Ohio State University

Thermophysical Properties of Selected Engineering Materials

For the sake of convenience when using heat-flow equations, the thermal properties of selected engineering materials are provided in Table 2.

TABLE 2 THERMAL PROPERTIES OF SELECTED ENGINEERING ALLOYS

ALLOY	DENSITY (ρ), AT 20 °C (68 °F), G/CM^3	SPECIFIC HEAT (C_P), AT 20 °C (68 °F), KJ/KG · K	THERMAL CONDUCTIVITY (κ) AT 20 °C (68 °F, W/M · K)	COEFFICIENT OF THERMAL EXPANSION (α), AT 20 °C (68 °F), 10^{-5}/°C
CARBON STEEL				
~0.5% C	7.833	0.465	54	1.474

~1.0% C	7.801	0.473	43	1.172
~1.5% C	7.753	0.486	36	0.970
ALUMINUM				
AL-CU (94-96 % AL, 3-5% CU)	2.787	0.883	164	6.676
AL-SI (86.5% AL. 1% CU)	2.659	0.867	137	5.933
AL-SI (78-80% AL, 20-22% SI)	2.627	0.854	161	7.172
AL-MG-SI (97% AL, 1% MG, 1% SI, 1% MN)	2.707	0.892	177	7.311
TITANIUM	4.500	0.52	16	0.84

ALLOY	**DENSITY (ρ), AT 100 °C (212 °F), G/CM^2**	**SPECIFIC HEAT (CP), AT 0-100 °C (32-212 °F), KJ/KG · K**	**THERMAL CONDUCTIVITY (κ), AT 100 °C (212 °F), W/M · K**	**COEFFICIENT OF THERMAL EXPANSION (α), AT 0-538 °C (32-1000 °F), 10 $^{-5}$/°C**
STAINLESS STEEL				
CHROMIUM-NICKEL AUSTENITIC	7.800-8.000	0.46-0.50	18.7-22.8	1.700-1.920
CHROMIUM FERRITIC	7.800	0.46-0.50	24.4-26.3	1.120-1.210
CHROMIUM MARTENSITIC	7.800	0.42-0.46	28.7	1.160-1.210
CARBON STEEL	7.800	0.48	60	1.17

Heat Flow in Fusion Welding

Chon L. Tsai and Chin M. Tso, The Ohio State University

References

1. N.S. BOULTON AND H.E. LANCE-MARTIN, RESIDUAL STRESSES IN ARC WELDED PLATES, *PROC. INST. MECH. ENG.,* VOL 33, 1986, P 295
2. D. ROSENTHAL AND R. SCHMERBER, THERMAL STUDY OF ARC WELDING, *WELD. J.,* VOL 17 (NO. 4), 1983, P 2S
3. N.N. RYKALIN, "CALCULATIONS OF THERMAL PROCESSES IN WELDING," MASHGIZ, MOSCOW, 1951
4. K. MASUBUCHI, *ANALYSIS OF WELDED STRUCTURES,* PERGAMON PRESS, 1980
5. C.L. TSAI, "PARAMETRIC STUDY ON COOLING PHENOMENA IN UNDERWATER WELDING," PH.D THESIS, MIT, 1977
6. C.L. TSAI, FINITE SOURCE THEORY, *MODELING OF CASTING AND WELDING PROCESSES II,* ENGINEERING FOUNDATION MEETING, NEW ENGLAND COLLEGE (HENNIKER, NH), 31 JULY TO 5 AUG 1983, P 329
7. R. PAVELIC, R. TANAKUCHI, O. CZEHARA, AND P. MYERS, EXPERIMENTAL AND COMPUTED TEMPERATURE HISTORIES IN GAS TUNGSTEN ARC WELDING IN THIN PLATES, *WELD. J.,* VOL 48 (NO. 7), 1969, P 295S
8. S. KOU, 3-DIMENSIONAL HEAT FLOW DURING FUSION WELDING, *PROC. OF METALLURGICAL SOCIETY OF AIME,* AUG 1980, P 129-138
9. P.G. KOGAN, THE TEMPERATURE FIELD IN THE WELD ZONE, *AVE. SVARKA,* VOL 4 (NO. 9), 1979, P 8
10. G.M. ECER, H.D. DOWNS, H.D. BRODY, AND M.A. GOKHALE, HEAT FLOW SIMULATION OF PULSED CURRENT GAS TUNGSTEN ARC WELDING, *MODELING OF CASTING AND WELDING PROCESSES,* ENGINEERING FOUNDATION 1980 MEETING (RINDGE, NH), 3-8 AUG 1980, P 139-160

11. H. HIBBITT AND P. MARCAL, A NUMERICAL THERMOMECHANICAL MODEL FOR WELDING AND SUBSEQUENT LOADING OF A FABRICATED STRUCTURE, *COMPUT. STRUCT.,* VOL 3, 1973, P 1145
12. E. FRIEDMAN, THERMOMECHANICAL ANALYSIS OF THE WELDING PROCESS USING FINITE ELEMENT METHODS, *TRANS. ASME,* AUG 1975, P 206
13. Z. PALEY AND P. HIBBERT, COMPUTATION OF TEMPERATURE IN ACTUAL WELD DESIGN, *WELD. J.,* VOL 54 (NO. 11), 1975, P 385.S
14. T. NAKA, TEMPERATURE DISTRIBUTION DURING WELDING, *J. JPN. WELD. SOC.,* VOL 11 (NO. 1), 1941, P 4
15. K. MASUBUCHI AND T. KUSUDA, TEMPERATURE DISTRIBUTION OF WELDED PLATES, *J. JPN WELD. SOC.,* VOL 22 (NO. 5), 1953, P 14
16. C.L. TSAI AND C.A. HOU, THEORETICAL ANALYSIS OF WELD POLL BEHAVIOR IN THE PULSED CURRENT GTAW PROCESS, *TRANSPORT PHENOMENA IN MATERIALS PROCESSING,* ASME WINTER ANNUAL MEETING, 1983
17. E. FRIEDMAN, "FINITE ELEMENT ANALYSIS OF ARC WELDING," REPORT WAPD-TM-1438, DEPARTMENT OF ENERGY, 1980
18. J.S. FAN AND C.L. TSAI, "FINITE ELEMENT ANALYSIS OF WELDING THERMAL BEHAVIOR IN TRANSIENT CONDITIONS," 84-HT-80, ASME
19. R.L. APPS AND D.R. MILNER, HEAT FLOW IN ARGON-ARC WELDING, *BR. WELD. J.,* VOL 2 (NO. 10), 1955, P 475
20. H.S. CARSLAW AND J.C. JAEGER, *CONDUCTION OF HEAT IN SOLIDS,* OXFORD PRESS
21. C.L. TSAI AND J.H. WU, "AN INVESTIGATION OF HEAT TRANSPORT PHENOMENA IN UNDERWATER WELDING," PRESENTED AT THE ASME WINTER ANNUAL MEETING (MIAMI BEACH, FL), 1985
22. HEAT FLOW IN WELDING, CHAPTER 3, *WELDING HANDBOOK,* VOL 1, 7TH ED., AWS, 1976
23. C.M. ADAMS, JR. COOLING RATE AND PEAK TEMPERATURE IN FUSION WELDING, *WELD. J.,* VOL 37 (NO. 5), 1958, P 210S-215S
24. C.M. ADAMS, JR., COOLING RATES AND PEAK TEMPERATURES IN FUSION WELDING, *WELD. J.,* VOL 37 (NO. 5), P 210-S TO 215-S
25. H. KIHARA, H. SUZUKI, AND H. TAMURA, *RESEARCH ON WELDABLE HIGH-STRENGTH STEELS,* 60TH ANNIVERSARY SERIES, VOL 1, SOCIETY OF NAVAL ARCHITECTS OF JAPAN, TOKYO, 1957
26. C.E. JACKSON, DEPARTMENT OF WELDING ENGINEERING, THE OHIO STATE UNIVERSITY LECTURE NOTE, 1977
27. C.E. JACKSON AND W.J. GOODWIN, EFFECTS OF VARIATIONS IN WELDING TECHNIQUE ON THE TRANSITION BEHAVIOR OF WELDED SPECIMENS--PART II, *WELD. J.,* MAY 1948, P 253-S TO 266-S

Fluid Flow Phenomena During Welding

C.R. Heiple and P. Burgardt, EG&G Rocky Flats

Introduction

MOLTEN WELD POOLS are dynamic. Liquid in the weld pool in acted on by several strong forces, which can result in high-velocity fluid motion. Fluid flow velocities exceeding 1 m/s (3.3 ft/s) have been observed in gas tungsten arc (GTA) welds under ordinary welding conditions, and higher velocities have been measured in submerged arc welds. Fluid flow is important because it affects weld shape and is related to the formation of a variety of weld defects. Moving liquid transports heat and often dominates heat transport in the weld pool. Because heat transport by mass flow depends on the direction and speed of fluid motion, weld pool shape can differ dramatically from that predicted by conductive heat flow. Temperature gradients are also altered by fluid flow, which can affect weld microstructure. A number of defects in GTA welds have been attributed to fluid flow or changes in fluid flow, including lack of penetration, top bead roughness, humped beads, finger penetration, and undercutting. Instabilities in the liquid film around the keyhole in electron beam and laser welds are responsible for the uneven penetration (spiking) characteristic of these types of welds.

Fluid Flow Phenomena During Welding

C.R. Heiple and P. Burgardt, EG&G Rocky Flats

Mass Transport in the Arc

High-velocity gas motion occurs in and around the arc during welding. The gas motion is partially due to cover gas flow, but, more importantly, it is driven by electromagnetic forces associated with the arc itself. In gas metal arc (GMA) welds, liquid filler metal is also being transferred through the arc from the electrode to the workpiece. Both the mode and velocity of metal transfer have major effects on weld pool shape. Mass transport in and around the arc is important in GTA welding (GTAW) and even more so in GMA welding (GMAW); a detailed description of this phenomenon can be found in the articles on arc physics and metal transfer and weld behavior in this Volume. Only the effects of this transport on the weld pool will be discussed in this article.

Fluid Flow Phenomena During Welding

C.R. Heiple and P. Burgardt, EG&G Rocky Flats

Gas Tungsten Arc Welding

Most experimental and theoretical work on weld pool fluid flow and its effects has been directed toward GTAW. The motivation for much of this work was the observation of dramatically different weld pool shapes for GTA welds made using identical welding parameters on different heats of the same material with the same nominal composition. An extreme example of weld shape variability is shown in Fig. 1. Early observations of variable weld shape (often referred to as variable penetration) were not only an intellectual puzzle but also an indication of a growing practical problem. Gas tungsten arc welding is commonly used for high-precision, high-quality automated welding applications, where reproducibility of weld shape or penetration is critical.

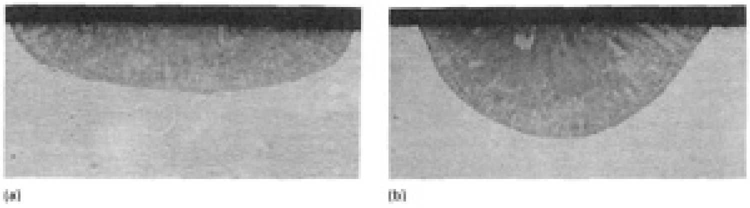

FIG. 1 PARTIAL-PENETRATION GTA WELDS MADE UNDER THE SAME WELDING CONDITIONS ON TWO HEATS OF TYPE 304L STAINLESS STEEL HAVING THE SAME NOMINAL COMPOSITION. (A) 3 PPM S, *D*/*W* = 0.2. (B) 160 PPM S, *D*/*W* = 0.40. 9×

The possibility that fluid flow in the weld pool could alter weld shape has been recognized for many years. For example, in 1965 Christensen *et al.* (Ref 1) proposed that convection is partially responsible for deviations in weld pool shapes from those predicted by conduction solutions. The forces driving fluid flow in GTA weld pools have also been long known. The four primary driving forces are surface tension gradients, electromagnetic or Lorentz forces, buoyancy forces, and aerodynamic drag forces caused by passage of the arc plasma over the weld pool surface.

Surface-Tension-Driven Fluid Flow Model. Surface tension gradients were first proposed by Ishizake *et al.* (Ref 2) as potential driving forces for weld pool fluid flow. Surface-tension-driven fluid flow was first described by Thomson (Ref 3) in 1855, but the phenomenon is commonly called Marangoni convection from the work of Carlo Marangoni (Ref 4).

In 1982, Heiple and Roper (Ref 5) proposed that surface tension gradients are commonly the dominant forces driving fluid flow in GTA welds and that these gradients could be drastically altered by very small concentrations of certain trace elements. Surface tension gradients exist on a weld pool surface because the surface tension is temperature dependent, and there are large temperature gradients on a weld pool surface. For pure metals and many alloys, the surface tension decreases as temperature increases; that is, the surface tension temperature coefficient is negative. For weld pools in such materials, the surface tension will be greatest on the coolest part of the pool surface at the edge and lowest on the hottest part under the arc near the center of the pool. Such a surface tension gradient produces outward surface fluid flow, as shown schematically in Fig. 2(a). This fluid flow pattern transfers heat efficiently from the hottest part of the weld pool (near the center) to the edge and produces a relatively wide and shallow weld.

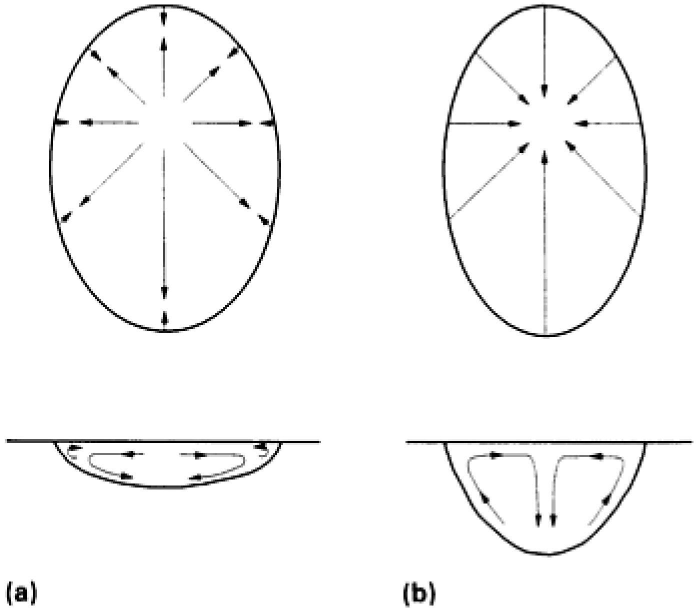

FIG. 2 SCHEMATIC SHOWING SURFACE FLUID FLOW (TOP) AND SUBSURFACE FLUID FLOW (BOTTOM) IN THE WELD POOL. (A) NEGATIVE SURFACE TENSION TEMPERATURE COEFFICIENT (PURE MATERIAL). (B) POSITIVE SURFACE TENSION TEMPERATURE COEFFICIENT (SURFACE-ACTIVE ELEMENTS PRESENT)

Certain elements are surface active in molten metals; that is, they segregate to the surface of the solvent liquid metal and lower the magnitude of the surface tension, often drastically. Small concentrations of surface-active additions can also change the temperature dependence of the surface tension of the solvent metal or alloy so that, for a limited range of temperature above the melting point, the surface tension increases with increasing temperature. With a positive surface tension temperature coefficient, the surface tension will be highest near the center of the weld pool. Such a surface tension gradient will produce fluid flow inward along the surface of the weld pool and then down, as indicated schematically in Fig. 2(b). This fluid flow pattern transfers heat efficiently to the bottom of the weld pool and produces a relatively deep, narrow weld.

Experimental Observations. This physical model was developed and verified with a series of experiments in which stainless steel base metal was doped with low concentrations of various elements and the effect of the doping on weld pool shape measured. High-speed motion pictures of the weld pool surface suggested the fluid flow patterns indicated in Fig. 2. The addition of sulfur, oxygen, selenium, and tellurium to stainless steel in low concentrations (less than 150 ppm) was shown to substantially increase GTA weld depth-to-width ratio (*d*/*w*). All these elements are known to be highly surface active in liquid iron. Measurements of the temperature dependence of the surface tension for steels with different GTA weld penetration characteristics produced an impressive correlation between a positive surface tension temperature coefficient arising from surface-active impurities and high *d*/*w* ratio welds (Fig. 3).

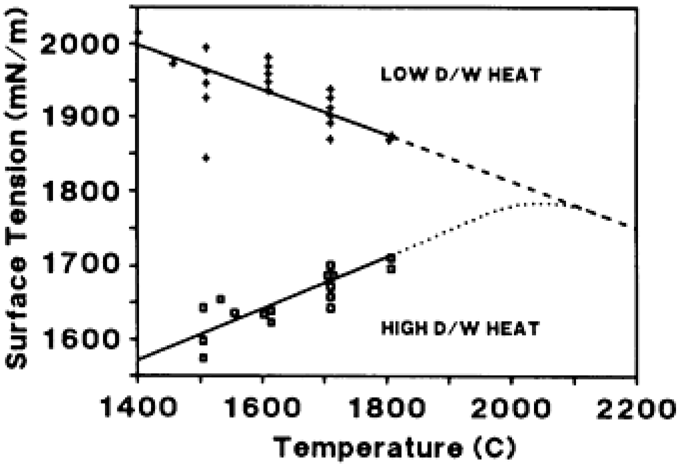

FIG. 3 PLOT OF SURFACE TENSION VERSUS TEMPERATURE FOR TWO LIQUID STEELS. THE DATA LABELED "HIGH *D*/*W* HEAT" ARE FROM MATERIAL HAVING APPROXIMATELY 160 PPM MORE SULFUR THAN THE MATERIAL LABELED "LOW *D*/*W* HEAT." THE DASHED LINES INDICATE THE EXPECTED BEHAVIOR OF THE SURFACE TENSION ABOVE THE MAXIMUM TEMPERATURE STUDIED. SOURCE: REF 6

When additions were made to the weld pool of elements known to react with surface-active elements already present in the steel to form compounds that are not surface active, the GTA weld *d*/*w* ratio decreased. Aluminum reacts with oxygen and produced wider, shallower welds in 21-6-9 stainless steel. Cerium reacts with both sulfur and oxygen and also produced lower *d*/*w* ratio welds. The effects of trace elements on weld shape have also been observed in a number of other alloys. These observations are summarized in Table 1.

TABLE 1 EFFECT OF TRACE ELEMENT IMPURITIES ON GTA WELD PENETRATION OF SELECTED ALLOYS

ALLOY SYSTEM	TRACE ELEMENT IMPURITY	
	INCREASES WELD PENETRATION	DECREASES WELD PENETRATION
ZIRCALOY-2	CHLORINE	. . .
IRON-BASE ALLOYS STAINLESS STEELS 304, 316, 21-6-9 JBK-75 AISI 8630 2.5CR-1MO	SULFUR, OXYGEN, SELENIUM$^{(A)}$, TELLURIUM$^{(A)}$	CALCIUM, ALUMINUM, CERIUM$^{(A)}$, LANTHANUM$^{(A)}$, SILICON$^{(B)}$, TITANIUM$^{(B)}$
NICKEL-BASE ALLOYS INCONEL 600, 718	SULFUR	OXYGEN

(A) UNCOMMON IMPURITIES

(B) EFFECT NEGLIGIBLE OR UNCERTAIN

The simple physical model illustrated in Fig. 2 has been remarkably successful in qualitatively explaining trace element effects. For example, changing GTA welding conditions alters the magnitude and distribution of arc energy input to the weld, which in turn changes temperature gradients on the weld pool surface. From Fig. 2, a change in welding conditions that makes the center of the weld hotter, such as increasing current, should drive the existing fluid flow pattern more strongly. As shown in Fig. 4, increasing current improves the *d/w* ratio of steel doped with surface-active elements and reduces it for high-purity base metal. If the center of the weld becomes so hot that there is a region where the temperature coefficient of the surface tension is no longer positive, then the fluid flow pattern necessary for deep penetration is disrupted and the *d/w* ratio decreases. This effect is seen at high currents in Fig. 4. Similar results have been obtained for other welding parameters. The surface temperature at which the change from positive to negative surface tension temperature coefficient occurs for stainless steel is estimated by extrapolation in Fig. 3 to be about 2050 °C (3720 °F). Detailed thermodynamic calculations of the temperature dependence of the surface tension of iron-sulfur alloys predict that the transition from positive to negative $d\gamma/dT$ will occur at 2032 °C (3690 °F) for the high-sulfur alloy in Fig. 3. Recent spectrographic weld pool temperature measurements and numerical simulations have indicated that this temperature can be exceeded in stainless steel GTA weld pools under normal welding conditions.

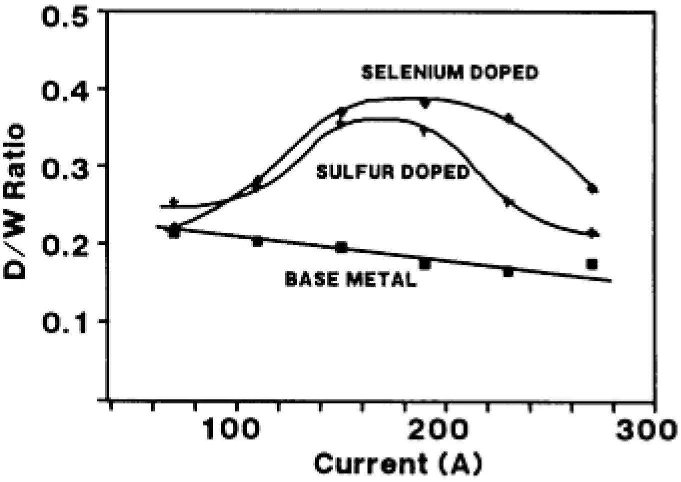

FIG. 4 PLOT OF WELD *D/W* RATIO VERSUS WELD CURRENT FOR THE STARTING BASE METAL (TYPE 304 STAINLESS STEEL WITH VERY LOW RESIDUAL IMPURITY CONTENT) AS WELL AS FOR ZONES DOPED WITH SULFUR AND SELENIUM. SOURCE: REF 7

The surface-tension-driven fluid flow model should be applicable to non-arc processes, provided the energy input distribution is similar to a GTA arc. This condition is satisfied for conduction-mode electron beam and laser welds. Dramatic increases in weld *d*/*w* ratio in selenium-doped zones in stainless steel have been observed for both traveling laser and electron beam conduction-mode welds. The weld shape changes were similar to those observed for GTA welds.

Conduction-mode electron beam welds can also be used to demonstrate that variations in weld shape with changes in welding parameters, as illustrated in Fig. 4, are not a result of some complex arc/weld pool interaction. One of the results of an investigation of the effect of changes in beam focus on weld shape in electron beam welds on low- and high-sulfur materials is shown in Fig. 5. The high-sulfur material exhibits a maximum in *d*/*w* ratio with increasing power density at a moderate power density away from sharp focus, which is analogous to that shown with increasing current in Fig. 4. (The low- and high-sulfur stainless steels have deep and nearly identical penetration near sharp focus. Near sharp focus, where the power density is high, penetration is by a keyhole mechanism in which average penetration is unrelated to surface-tension-driven fluid flow.) Measurements of the electron beam power-density distribution verified that there were no anomalous changes in the beam, such as a beam width maximum, with increasing peak power density. The *d*/*w* ratio maximum away from sharp focus is therefore proposed to originate from exactly the same mechanism as for increasing current with GTA welds.

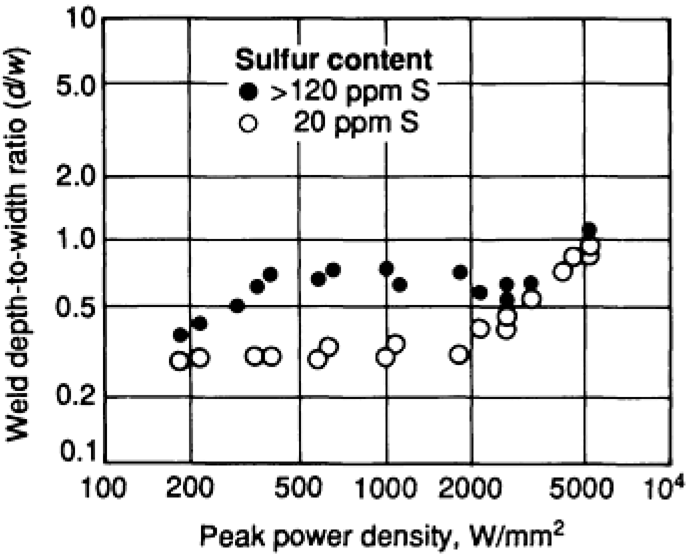

FIG. 5 PLOT OF ELECTRON BEAM WELD POOL RATIO (*D*/*W*) VERSUS ELECTRON BEAM POWER DENSITY FOR LOW-SULFUR (20 PPM) AND HIGH-SULFUR (>120 PPM) TYPE 304L STAINLESS STEEL. KEYHOLE FORMATION BEGINS AT ABOUT 2 × 10^3 W/MM^2.

Measurements of electron beam power distribution were made as a function of beam focus and were used to calculate the beam power density at the *d*/*w* maximum. The power density was also calculated at the *d*/*w* maximum for GTA welds. Calculation of weld pool surface temperatures using a traveling distributed heat source model showed the central surface temperatures to be essentially identical at the *d*/*w* maxima for the electron beam and GTA processes. The calculated surface temperatures using the traveling distributed heat source conduction approximation are much too high because, as indicated previously, most of the heat transport in the weld pool is by mass flow rather than by conduction. Nevertheless, the equality of the calculated weld pool peak temperatures at the *d*/*w* maxima provides strong confirmation that the mechanism responsible for the presence of the maximum in weld *d*/*w* with increasing input power density is independent of heat source and is not a result of an arc phenomenon.

A final example of the success of the fluid flow model in explaining GTA welding phenomena is provided by butt welding together two steels with large differences in weld penetration characteristics. The weld pool is not centered over the joint; rather, it is displaced toward the material with low *d*/*w* behavior (Ref 8), as indicated schematically in Fig. 6. The low *d*/*w* material has a low concentration of surface-active impurities and therefore a high surface tension. Thus, there is a net surface tension gradient across the weld pool toward the low *d*/*w* material, producing the fluid flow pattern and weld cross section indicated in Fig. 6. The actual fluid flow pattern is certain to be more complicated than that illustrated.

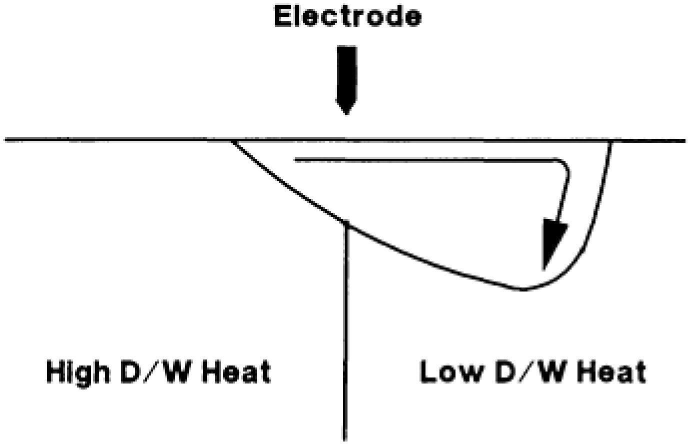

FIG. 6 SCHEMATIC SHOWING TYPICAL FLUID FLOW GENERATED WHEN BUTT WELDING TWO HEATS OF MATERIAL WITH DIFFERENT PENETRATION CHARACTERISTICS. SOURCE: REF 8

Numerical Simulations. Direct measurement of weld pool fluid flow is very difficult. Surface flow has been studied by observing the motion on the surface of slag or intentionally added particles, as has bulk flow in transparent liquid/solid systems using simulated welding heat sources. There are a number of numerical calculations of weld pool fluid flow and shape, the most comprehensive of which appears to be that of Zacharia *et al.* (Ref 9). These numerical fluid flow simulations agree in broad terms with the physical model illustrated in Fig. 2; however, they differ in detail with Fig. 2 and with one another. One result of the analysis of Zacharia is that the weld pool surface temperature reached in stationary welds exceeds that in traveling welds. Trace element effects should therefore be less pronounced in stationary welds, because a larger portion of the pool surface is above the temperature range where a positive surface tension temperature coefficient exists. This prediction is in agreement with experimental observations.

Interactions. Experiments in which low concentrations of various elements have been added to the base metal have demonstrated clear changes in weld pool shape. However, when weld *d*/*w* ratio is plotted versus chemical analysis for numerous heats of material, considerable scatter is observed. A comparison of weld *d*/*w* ratio in a standard weldability test versus sulfur content for about 200 lots of steel is shown in Fig. 7. There is a clear trend of increasing *d*/*w* with increasing sulfur, but the variability for a given sulfur content is substantial. Some of this variability is associated with imprecision in chemical analysis for sulfur and some with variations in oxygen content. However, it appears that much of the variability is related to interactions of the surface-active elements sulfur and oxygen with other components of the steel. Calcium is known to react with oxygen and, to a lesser extent, with sulfur to form stable compounds unlikely to be surface active. Aluminum and silicon also react with oxygen to form stable compounds. Thus, the amount of sulfur and oxygen available for segregation to the weld pool surface is a complicated function of the total weld pool chemistry.

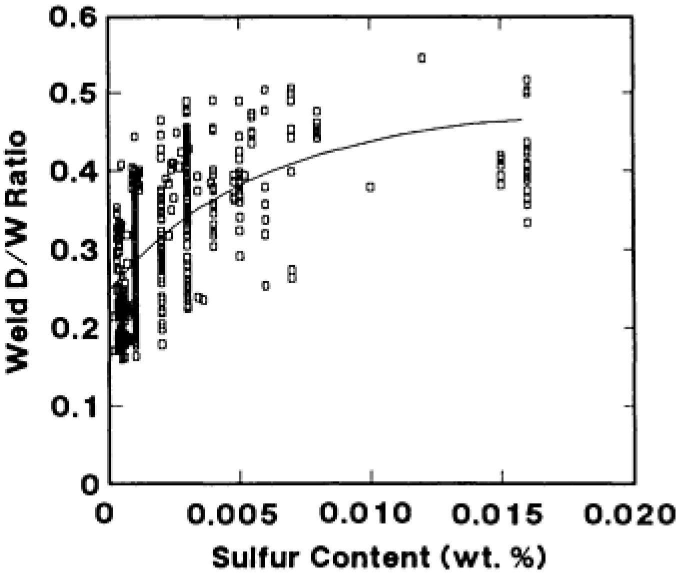

FIG. 7 PLOT OF WELD *D/W* RATIO VERSUS SULFUR CONTENT FOR APPROXIMATELY 200 HEATS OF TYPE 304L STAINLESS STEEL. EACH POINT IS AN AVERAGE OF MULTIPLE SULFUR ANALYSES AND WELD *D/W* RATIO MEASUREMENTS. IF SINGLE VALUES ARE USED, THE SCATTER IS GREATER. SOURCE: REF 10

There is an additional complication, illustrated in Fig. 8. When oxygen is added to the shielding gas, weld *d*/*w* ratio increases, passes through a maximum, and then declines with increasing oxygen content. Similar effects are seen with SO_2 additions. The surface of welds made with torch gas concentrations above the *d*/*w* maximum are heavily oxidized. A possible explanation for the decreasing *d*/*w* ratios at higher oxygen concentrations is that a liquid oxide film (slag) is formed on the weld pool surface, altering the surface tension gradients.

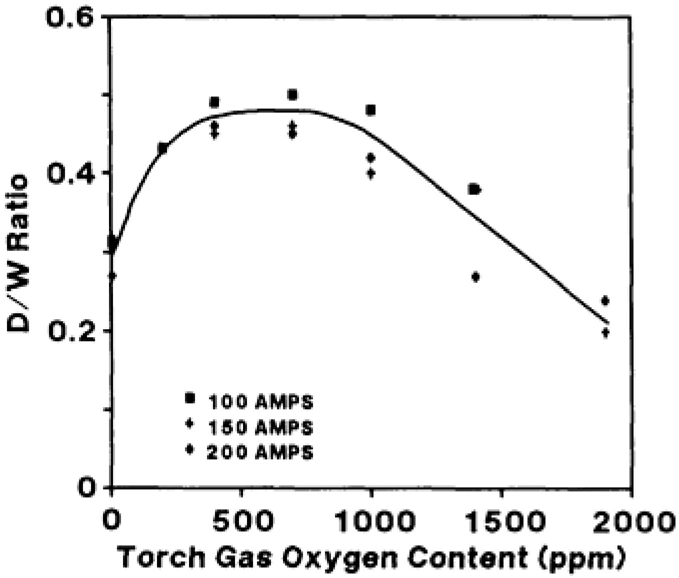

FIG. 8 PLOT OF WELD *D/W* RATIO VERSUS OXYGEN CONCENTRATION IN THE TORCH GAS FOR GTA BEAD-ON-PLATE WELDS ON 21-6-9 STAINLESS STEEL. SOURCE: REF 11

Oxygen can be added to the weld pool in other ways. Oxidizing the plate surface prior to welding adds oxygen to the weld pool and increases the weld *d*/*w* ratio. Wire brushing and grit blasting an originally clean surface increases both the surface oxide thickness and the surface area, thereby adding oxygen to the weld pool. The *d*/*w* ratio of JBK-75 stainless steel (a modification of ASTM A-286) GTA welds as a function of amount of wire brushing is similar in form to Fig. 8. Welds on plate brushed beyond the weld *d*/*w* maximum had extensive slag on the weld surface. The weld *d*/*w* ratio on JBK-75 stainless steel appears to be particularly sensitive to oxygen additions. Similar wire-brushing experiments on type 304 stainless steel showed smaller effects.

Although the origin of the difference in sensitivity between JBK-75 and type 304 is not known, one possibility is indicated by surface tension measurements on iron-silicon alloys in contact with carbon dioxide. The effect of CO_2 on the surface tension was a strong function of silicon content. For low-silicon alloys, the surface tension dropped sharply when contacted with CO_2. For alloys with more than 1.2% Si, the surface tension increased when contacted with CO_2. The different behavior was attributed to differences in slag formation on the liquid metal surface. Thus, silicon appears to interfere with the ability of oxygen to produce a positive surface tension temperature coefficient on liquid iron. The JBK-75 stainless steel used in the wire-brushing experiments contained only 0.06% Si, and the 21-6-9 used for the torch gas experiment (Fig. 8) contained only 0.16% Si. Type 304 stainless steel typically contains more than 0.5% Si. Thus, the high sensitivity of JBK-75 to oxygen additions may be related to its unusually low silicon content. Reports in the literature on the effect of oxygen on GTA weld penetration have been somewhat variable; differences in weld pool silicon content may be partially responsible for this variability.

Effects of High Current. Electromagnetic (Lorentz) stirring of the weld pool becomes more important at higher currents. The direction of the Lorentz force produces the deep penetration fluid flow pattern indicated in Fig. 2(b). At sufficiently high current, the Lorentz force dominates other forces that drive fluid flow, and the effects of trace elements become less important. In addition, the plasma jet becomes stronger with increasing current and at high enough currents produces a significant depression of the weld pool surface. The plasma jet forces are resisted by surface tension forces on the weld pool. Halmoy (Ref 12) has shown that for a traveling weld the radial pressure gradient from the plasma jet tends to transport liquid from the front to the rear of the weld (Fig. 9). For a sufficiently strong pressure gradient, p_B, the liquid level under the arc may be pushed down to the bottom of the pool, as shown in Fig. 9(c). High-speed motion pictures of the effect of sulfur additions on weld pool fluid flow in 21-6-9 stainless steel showed behavior similar to that in Fig. 9(c) after large sulfur additions. Sulfur additions substantially reduce the weld pool surface tension and thereby increase the effect of the plasma jet.

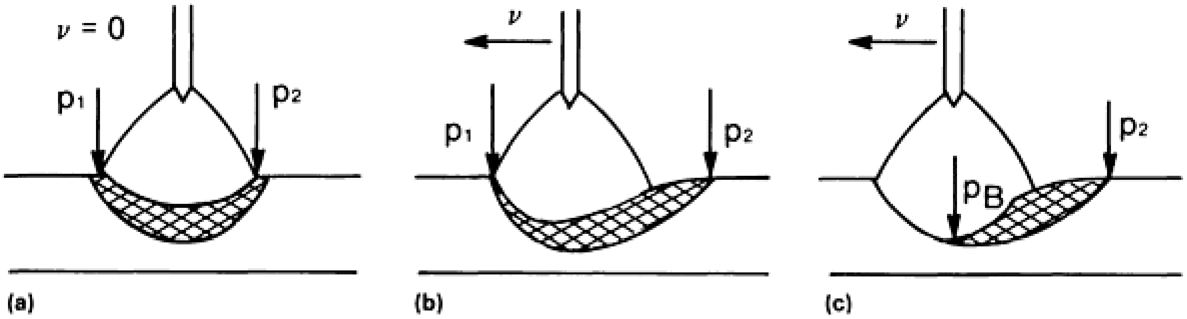

FIG. 9 EFFECT OF ARC PRESSURE ON THE WELD POOL FOR STATIONARY AND TRAVELING WELDS. (A) $V = 0$. (B) $V > 0$, WITH WEAK RADIAL PRESSURE GRADIENT, P_1. (C) $V > 0$, WITH STRONG RADIAL PRESSURE GRADIENT, P_B. SOURCE: REF 12

In addition, a surface depression almost certainly changes the energy distribution input to the weld pool by the arc compared with a flat weld pool surface. Finally, under some high-current conditions, a vortex has been observed near the center of weld pools.

Strategies for Controlling Poor and Variable Penetration. A number of techniques have been developed to reduce penetration variability and improve penetration. Sometimes the weld pool shift illustrated in Fig. 6, which occurs when heats with different penetration characteristics are welded together, can be minimized by very tight heat sinking. The sensitivity of weld shape to trace element differences is a function of welding parameters, as illustrated in Fig. 4 for current; thus, combinations of welding parameters can be chosen that minimize heat-to-heat penetration variability. For wire-fed joints, some joint designs are more tolerant of penetration variability than others. For example, joints with thinner, wider lands are less sensitive than narrower grooves with thicker lands.

Material control by selection or specification is another approach. It was originally anticipated that stainless steels with less than 20 ppm S would have poor, but consistent, penetration. However, the data illustrated in Fig. 7 demonstrate that penetration variability is substantial even at these very low sulfur levels. Stainless steels with more than about 100 ppm S generally have consistently good penetration. A related approach is to test incoming material for welding characteristics and then select material with desired welding behavior for critical applications.

Finally, the weld pool can be doped with enough surface-active elements to ensure good penetration. Sulfur and oxygen are the most practical dopants for ferrous alloys. The addition of oxygen or sulfur dioxide to the torch shielding gas has been shown to improve penetration and reduce penetration variability; however, this approach presents several practical problems. Oxygen can also be added to the weld pool by oxidizing the weld joint or by otherwise increasing the oxygen content of the weld groove surface--by wire brushing, for example. Probably the most useful approach is to dope the weld pool in the wire-fed joints by using a special filler wire. A group of defense contractors had a special heat of type 380L stainless steel produced with a sulfur content in the range of 100 to 150 ppm. The heat was then converted into weld wire. Welds made with this wire on a variety of stainless steel base metals have exhibited consistent and good penetration.

References cited in this section

1. N. CHRISTENSEN, V. DE L. DAVIES, AND K. GJERMUNDSEN, DISTRIBUTION OF TEMPERATURES IN ARC WELDING *BR. WELD.J.*, VOL 12, 1965, P 54-75
2. K. ISHIZAKE, K. MURAI, AND Y. KANBE, "PENETRATION IN ARC WELDING AND CONVECTION IN MOLTEN METAL," DOCUMENT 77-66, INTERNATIONAL INSTITUTE OF WELDING, STUDY GROUP 212, 1966
3. J. THOMSON, ON CERTAIN CURIOUS MOTIONS OBSERVABLE AT THE SURFACES OF WINE AND OTHER ALCOHOLIC LIQUORS, *PHILOS.MAG.*, VOL 10, 1855, P 330-333
4. C. MARANGONI, ÜBER DIE AUSBREITUNG DER TROPFEN EINER FLUESSIGKEIT AUF DER OBERFLAESCHE EINER ANDEREN, *ANN. PHYS. CHEM.*, VOL 143 (NO. 7), 1871, P 337-354

5. C.R. HEIPLE AND J.R. ROPER, MECHANISM FOR MINOR ELEMENT EFFECT ON GTA FUSION ZONE GEOMETRY, *WELD. J.*, VOL 61, 1982, P 97S-102S
6. B.J. KEENE, K.C. MILLS, AND R.F. BROOKS, SURFACE PROPERTIES OF LIQUID METALS AND THEIR EFFECTS ON WELDABILITY, *MATER. SCI.TECHNOL.*, VOL 1, 1985, P 568-571
7. P. BURGARDT AND C.R. HEIPLE, INTERACTION BETWEEN IMPURITIES AND WELDING PARAMETERS IN DETERMINING GTA WELD SHAPE, *WELD. J.*, VOL 65, 1986, P 150S-155S
8. M.J. TINKLER, I. GRANT, G. MIZUNO, AND C. GLUCK, WELDING 304L STAINLESS STEEL TUBING HAVING VARIABLE PENETRATION CHARACTERISTICS, *THE EFFECTS OF RESIDUAL, IMPURITY, AND MICRO-ALLOYING ELEMENTS ON WELDABILITY AND WELD PROPERTIES*, P.H.M. HART, ED., PAPER 29, THE WELDING INSTITUTE, CAMBRIDGE, 1984
9. T. ZACHARIA, S.A. DAVID, J.M. VITEK, AND T. DEBROY, WELD POOL DEVELOPMENT DURING GTA AND LASER BEAM WELDING OF TYPE 304 STAINLESS STEEL, PART 1--THEORETICAL ANALYSIS, *WELD. J.*, VOL 68, 1989, P 499S-509S
10. P. BURGARDT AND R.D. CAMPBELL, CHEMISTRY EFFECTS ON STAINLESS STEEL WELD PENETRATION, *FERROUS ALLOY WELDMENTS*, TRANS TECH PUBLICATIONS, SWITZERLAND, 1992, P 379-415
11. C.R. HEIPLE, P. BURGARDT, AND J.R. ROPER, THE EFFECT OF TRACE ELEMENTS ON GTA WELD PENETRATION, *MODELING OF CASTING AND WELDING PROCESSES II*, J.A. DANTZIG AND J.T. BERRY, ED., TMS-AIME, 1984, P 193-205
12. E. HALMOY, THE PRESSURE OF THE ARC ACTING ON THE WELD POOL, *ARC PHYSICS AND WELD POOL BEHAVIOR*, THE WELDING INSTITUTE, CAMBRIDGE, 1979, P 259-266

Fluid Flow Phenomena During Welding

C.R. Heiple and P. Burgardt, EG&G Rocky Flats

Deep-Penetration Electron Beam and Laser Welds

Keyhole Formation. A fundamental difference between arc heat sources and electron or laser beam heat sources is that electron and laser beams are capable of delivering heat over a small area at much higher power densities. As the power density of the welding heat source is increased, the peak surface temperature of the weld pool rises. For many metals, vapor pressure rises nearly exponentially with temperature and becomes appreciable (above 4000 Pa, or 30 torr, for an electron beam operating in a vacuum) near 0.8 T_b, where T_b is the material boiling point in degrees Kelvin. As the surface temperature approaches this value, the liquid surface under the power source is depressed by the vapor pressure. As the liquid moves away from the power source, the surface is depressed and a cavity is formed. This is the basic theory of keyhole formation by an electron or laser beam.

Fluid Flow in the Keyhole. The keyhole is a cavity having roughly the size and shape of the beam; that is, it is usually approximately cylindrical. Fluid flow in the thin layer around the cavity plays a major role in determining the behavior of deep-penetration, keyhole-mode electron beam and laser welds. The first flow required for a traveling weld is transport of metal melted at the front wall of the cavity to the rear (where it eventually solidifies). Because of the weld geometry, this liquid must move around the keyhole cavity as a thin, high-velocity layer on the walls of the cavity. The dominant driving force for this motion appears to be surface tension gradients. As in GTAW, these gradients arise because the surface tension is temperature dependent and there is a substantial temperature difference between the front and rear of the cavity.

Calculations of the fluid flow have been performed by Wei and Giedt (Ref 13). Their calculations for pure iron predict that liquid metal is moved from the front to the rear of a sharp focus keyhole in a film that is about 0.02 mm (0.0008 in.) thick and is moving at a velocity of about 250 mm/s (10 in./s). The fluid is driven by a temperature difference of about 400 °C (720 °F). These results are only nominal values from the calculations, but they change little with the assumptions used and are representative of electron beam welds in general.

There is an interesting potential problem with this model. As discussed previously, the surface tension temperature coefficient, $d\gamma/dT$, can be positive over a limited temperature range in steels and some other alloys if surface-active impurities are present in sufficient quantity. If $d\gamma/dT$ is positive, then it would appear that molten metal would not be transported around the cavity. The likely resolution of this difficulty is that the liquid dwells slightly longer in the front of the cavity and is heated by the beam above the temperature at which a transition to negative $d\gamma/dT$ occurs. Some perturbation of the normal fluid flow may occur under these circumstances. Effects of such a perturbation on weld characteristics appear to be uncommon. The only available evidence for such an effect comes from electron beam welds performed on a heat of type 304 stainless steel with very low residual impurity content, except for about 340 ppm S. These welds had very high porosity compared with identical welds in other heats of material.

Instability in Keyhole Fluid Flow. In addition to the steady-state flow of liquid around the keyhole, significant instability in the fluid motion has been seen. For example, Mara *et al.* (Ref 14) used side-view, self-illuminated X-ray films to show that the beam locations shifts between full penetration and nearly zero penetration in an irregular fashion. High-speed X-ray photographs show that the variable penetration is caused by a lump of metal that sags into the keyhole from high up on the cavity rear wall. The sequence of events that creates weld penetration irregularity was analyzed by Tong and Giedt (Ref 15) and is illustrated schematically in Fig. 10. First, the beam forms a keyhole and produces a lump of very hot displaced metal at the top rear of the traveling cavity (Fig. 10a). After some period of time, the weld reaches the full penetration allowed by heat flow (Fig. 10b). The lump of displaced metal is unstable and eventually falls into the cavity, partially filling it. The beam must now drill through this additional material. The sequence (a) to (c) in Fig. 10 occurs repeatedly, but irregularly, and produces weld penetration variability. When particularly severe, the irregular penetration is called spiking. Spiking also leads to a tendency for voids to become trapped in the root of the weld.

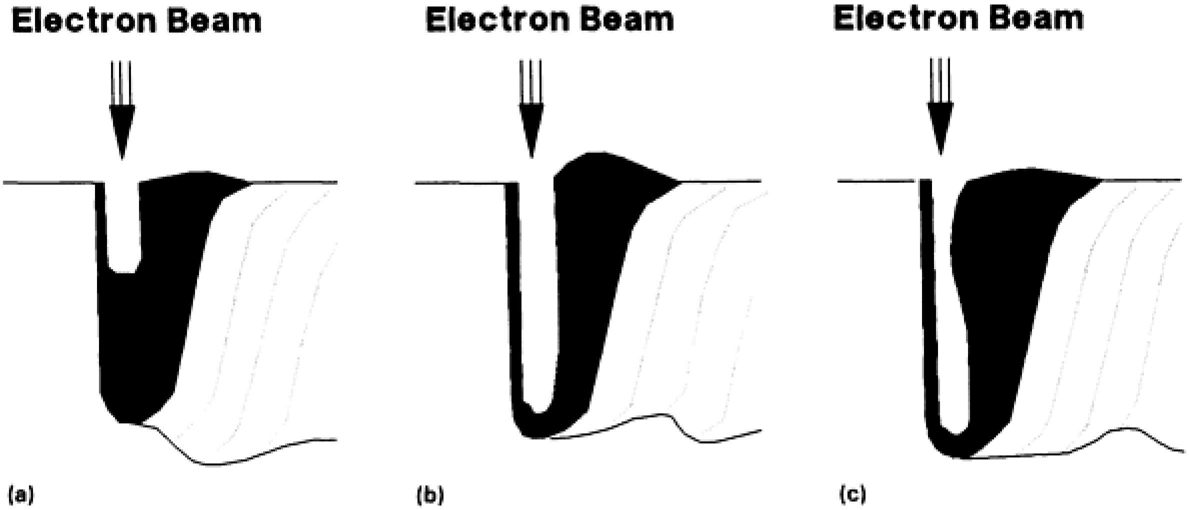

FIG. 10 SCHEMATIC SHOWING KEYHOLE INSTABILITY. (A) KEYHOLE IS FORMED BY HEAT GENERATED BY ELECTRON BEAM. (B) MAXIMUM PENETRATION THAT CAN BE PRODUCED BY HEAT FLOW. (C) LIQUID COOLS, CAUSING IMPENDING COLLAPSE OF DISPLACED METAL. THE KEYHOLE IS FILLED BY A LUMP OF COOLING MATERIAL AT THE END OF (C), WHICH RETURNS THE KEYHOLE TO CONDITION (A) TO RESTART THE SEQUENCE.

A mathematical model that is the basis for understanding this keyhole instability was developed by Giedt *et al.* (Ref 13, 15). Stability of the keyhole is basically a balance between vapor pressure, which keeps the cavity open, and surface tension, which tries to close the cavity. Vapor pressure and surface tension are both functions of temperature. Vapor pressure increases nearly exponentially with temperature for many materials, while surface tension generally decreases slowly with temperature. Because there is a substantial temperature gradient from the top to the bottom of the keyhole, both vapor pressure and surface tension vary with depth in the keyhole. Direct measurements of cavity wall temperatures for electron beam welds indicate that the temperature at the top is nearly equal to the melting temperature, while the temperature at the bottom is high enough that the vapor pressure there is large. In type 304 stainless steel, the temperature at the bottom of the cavity is about 2200 °C (1200 °F). The data indicate that the temperature profile is insensitive to welding variables and is determined primarily by material properties, that is, by the temperature where the vapor pressure exceeds about 1330 Pa (10 torr).

Based on temperature measurements of a typical keyhole, the vapor pressure decreases approximately exponentially with height from the bottom of the keyhole and approaches zero at the top of the keyhole. The vapor pressure is balanced primarily by surface tension. For pure iron and many other materials, the surface tension decreases approximately linearly with temperature. Thus, the inward surface tension pressure increases slowly from the bottom to the top of the keyhole. The balance of vapor pressure and surface tension is such that the cavity is stable near the bottom, where the vapor pressure tending to expand the cavity exceeds the surface tension tending to collapse it. At the top, the cavity is unstable because the surface tension tending to collapse it exceeds the vapor pressure tending to expand it. Thus, the cavity will always tend to be filled by liquid originating from above some height H in the keyhole, the pressure crossover height. This cavity filling produces weld penetration irregularity. Weld penetration irregularity is a consequence of the nature of the keyhole in deep-penetration welding and will not generally be solved by simple equipment or material modifications.

Penetration irregularity is usually less severe if H (as measured from the top) is a small fraction of the keyhole depth. In such cases the volume of the lump is small, and liquid tends to flow into the cavity relatively smoothly. If H is a large fraction of the keyhole depth, then the liquid lump has a relatively large volume and, when it falls into the keyhole at irregular times, it nearly fills the cavity and produces large penetration variations.

References cited in this section

13. P.S. WEI AND W.H. GIEDT, SURFACE TENSION GRADIENT-DRIVEN FLOW AROUND AN ELECTRON BEAM WELDING CAVITY, *WELD. J.*,VOL 64, 1985, P 251S-259S
14. G.L. MARA, E.R. FUNK, R.C. MCMASTER, AND P.E. PENCE, PENETRATION MECHANISMS OF ELECTRON BEAM WELDING AND THE SPIKING PHENOMENON, *WELD. J.*, VOL 53, 1974, P 246S-251S
15. H. TONG AND W.H. GIEDT, A DYNAMIC INTERPRETATION OF ELECTRON BEAM WELDING, *WELD. J.*, VOL 49, 1970, P 259S-266S

Fluid Flow Phenomena During Welding

C.R. Heiple and P. Burgardt, EG&G Rocky Flats

Gas Metal Arc Welding

Fluid flow certainly occurs in GMA weld pools, but reports on the details of its nature and effects are quite limited. In the spray transfer mode, the impact of the stream of droplets from the electrode on the weld pool forms a substantial depression or crater. This is the mechanism responsible for the typical fingerlike penetration observed with argon as the torch gas. Thus, the depth of penetration is primarily dependent on the momentum of the stream of droplets. A plot of penetration versus the momentum of the droplet stream shows an excellent correlation (Fig. 11). Because the major force driving the droplets toward the weld is drag from the plasma jet, the kinematic viscosity (density times viscosity) of the shielding gas is important. For example, argon has a substantially higher kinematic viscosity than helium, so argon drives the droplet stream more effectively, resulting in a deeper, narrower penetration finger than with helium torch gas. The strength of the plasma jet can also be altered by changing the ambient pressure. For GMA welds on aluminum. Amson and Salter (Ref 17) saw a steady decrease in penetration with pressure until at 13,300 Pa (100 torr) there was essentially no penetration at all. Another way to modify the penetration depth is to spread the stream of droplets over the surface of the weld. This can be accomplished by applying a varying transverse magnetic field.

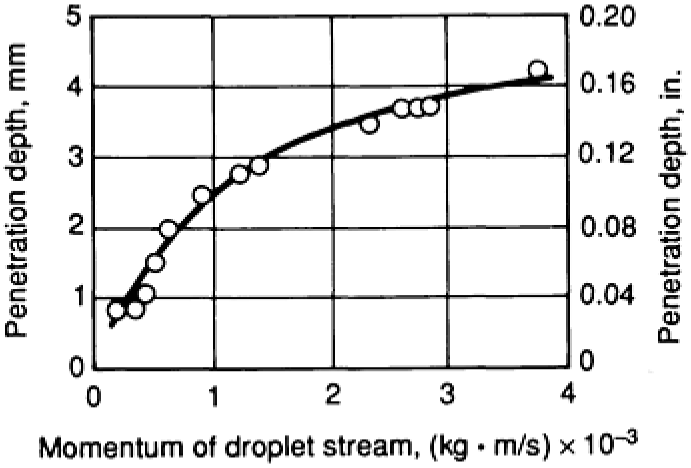

FIG. 11 PLOT OF GMA WELD PENETRATION IN MILD STEEL FOR SPRAY TRANSFER MODE VERSUS MOMENTUM OF DROPLET STREAM. SOURCE: REF 16

References cited in this section

16. W.G. ESSERS AND R. WALTER, HEAT TRANSFER AND PENETRATION MECHANISMS WITH GMA AND PLASMA-GMA WELDING, *WELD. J.*, VOL 62, 1981, P 37S-42S
17. J.C. AMSON AND G.R. SALTER, ANALYSIS OF THE GAS-SHIELDED CONSUMABLE METAL-ARC WELDING SYSTEM, *BR. WELD. J.*, VOL 10, 1963, P 472-483

Fluid Flow Phenomena During Welding

C.R. Heiple and P. Burgardt, EG&G Rocky Flats

Submerged Arc Welding

Currents commonly employed in submerged arc welding (SAW) are much higher than those used in GTAW or GMAW. Submerged arc welding currents often exceed 1000 A. Thus, the electromagnetic or Lorentz force combined with the tendency of the radial pressure gradient in the moving cavity to transport liquid to the rear of the cavity (Fig. 9) are likely the dominant forces driving fluid flow. The generally accepted flow pattern is indicated schematically in Fig. 12. A cavity is formed at the front of the moving weld pool by arc pressure and by the momentum of drops from the rapidly melting electrode. Metal that is melted at the front of the pool flows underneath and on either side of the cavity. At the rear of the pool, the flow reverses and metal flows back toward the cavity along and near the surface. Flow velocities can be very high; Eichhorn and Engel (Ref 19) measured 4 m/s (13 ft/s) at 720 A. The flow pattern was derived from observations of motion of marker elements, added to the weld pool, as determined from subsequent metallographic sections, and from the motion of radioactive tracer additions. The general features illustrated in Fig. 12 appear to be supported by X-ray fluoroscopy observations. Major perturbations in the indicated fluid flow have been demonstrated, including reversal of the flow direction. These perturbations have been associated with weld defects.

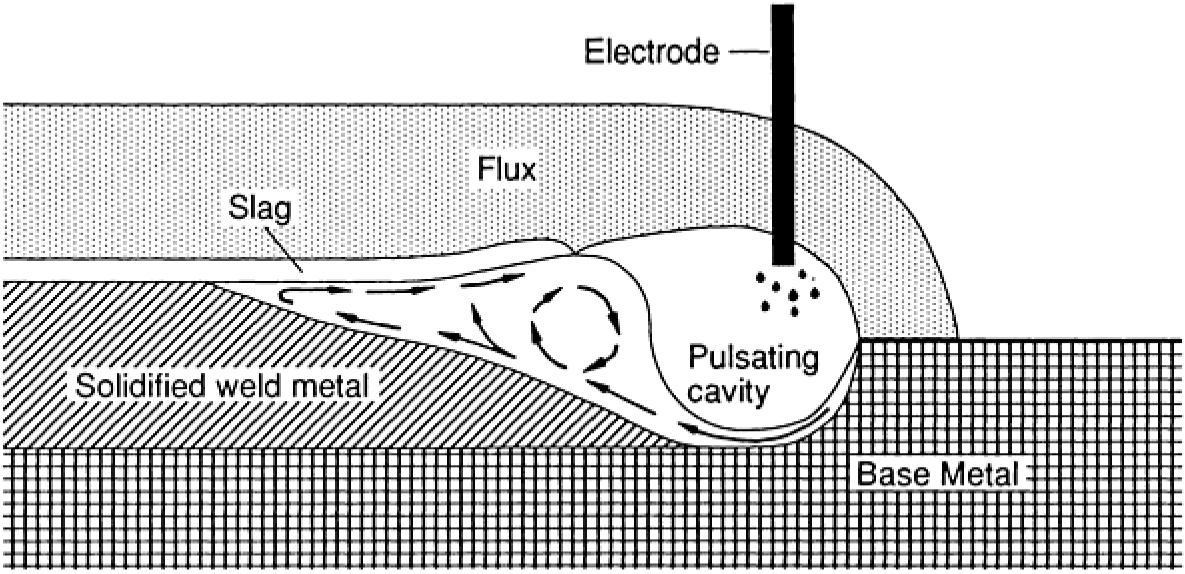

FIG. 12 SCHEMATIC SHOWING TYPICAL FLOW PATTERN IN A SUBMERGED ARC WELD POOL. SOURCE: REF 18

References cited in this section

18. J.F. LANCASTER, *THE PHYSICS OF WELDING*, INTERNATIONAL INSTITUTE OF WELDING, PERGAMON PRESS, OXFORD, 1984, P 243
19. F. EICHHORN AND A. ENGEL, "MASS TRANSFER IN THE WELD POOL," DOCUMENT 201-70, INTERNATIONAL INSTITUTE OF WELDING, STUDY GROUP 212, 1970

Fluid Flow Phenomena During Welding

C.R. Heiple and P. Burgardt, EG&G Rocky Flats

References

1. N. CHRISTENSEN, V. DE L. DAVIES, AND K. GJERMUNDSEN, DISTRIBUTION OF TEMPERATURES IN ARC WELDING *BR. WELD.J.*, VOL 12, 1965, P 54-75
2. K. ISHIZAKE, K. MURAI, AND Y. KANBE, "PENETRATION IN ARC WELDING AND CONVECTION IN MOLTEN METAL," DOCUMENT 77-66, INTERNATIONAL INSTITUTE OF WELDING, STUDY GROUP 212, 1966
3. J. THOMSON, ON CERTAIN CURIOUS MOTIONS OBSERVABLE AT THE SURFACES OF WINE AND OTHER ALCOHOLIC LIQUORS, *PHILOS.MAG.*, VOL 10, 1855, P 330-333
4. C. MARANGONI, ÜBER DIE AUSBREITUNG DER TROPFEN EINER FLUESSIGKEIT AUF DER OBERFLAESCHE EINER ANDEREN, *ANN. PHYS. CHEM.*, VOL 143 (NO. 7), 1871, P 337-354
5. C.R. HEIPLE AND J.R. ROPER, MECHANISM FOR MINOR ELEMENT EFFECT ON GTA FUSION ZONE GEOMETRY, *WELD. J.*, VOL 61, 1982, P 97S-102S
6. B.J. KEENE, K.C. MILLS, AND R.F. BROOKS, SURFACE PROPERTIES OF LIQUID METALS AND THEIR EFFECTS ON WELDABILITY, *MATER. SCI.TECHNOL.*, VOL 1, 1985, P 568-571
7. P. BURGARDT AND C.R. HEIPLE, INTERACTION BETWEEN IMPURITIES AND WELDING PARAMETERS IN DETERMINING GTA WELD SHAPE, *WELD. J.*, VOL 65, 1986, P 150S-155S
8. M.J. TINKLER, I. GRANT, G. MIZUNO, AND C. GLUCK, WELDING 304L STAINLESS STEEL

TUBING HAVING VARIABLE PENETRATION CHARACTERISTICS, *THE EFFECTS OF RESIDUAL, IMPURITY, AND MICRO-ALLOYING ELEMENTS ON WELDABILITY AND WELD PROPERTIES*, P.H.M. HART, ED., PAPER 29, THE WELDING INSTITUTE, CAMBRIDGE, 1984

9. T. ZACHARIA, S.A. DAVID, J.M. VITEK, AND T. DEBROY, WELD POOL DEVELOPMENT DURING GTA AND LASER BEAM WELDING OF TYPE 304 STAINLESS STEEL, PART 1--THEORETICAL ANALYSIS, *WELD. J.*, VOL 68, 1989, P 499S-509S
10. P. BURGARDT AND R.D. CAMPBELL, CHEMISTRY EFFECTS ON STAINLESS STEEL WELD PENETRATION, *FERROUS ALLOY WELDMENTS*, TRANS TECH PUBLICATIONS, SWITZERLAND, 1992, P 379-415
11. C.R. HEIPLE, P. BURGARDT, AND J.R. ROPER, THE EFFECT OF TRACE ELEMENTS ON GTA WELD PENETRATION, *MODELING OF CASTING AND WELDING PROCESSES II*, J.A. DANTZIG AND J.T. BERRY, ED., TMS-AIME, 1984, P 193-205
12. E. HALMOY, THE PRESSURE OF THE ARC ACTING ON THE WELD POOL, *ARC PHYSICS AND WELD POOL BEHAVIOR*, THE WELDING INSTITUTE, CAMBRIDGE, 1979, P 259-266
13. P.S. WEI AND W.H. GIEDT, SURFACE TENSION GRADIENT-DRIVEN FLOW AROUND AN ELECTRON BEAM WELDING CAVITY, *WELD. J.*,VOL 64, 1985, P 251S-259S
14. G.L. MARA, E.R. FUNK, R.C. MCMASTER, AND P.E. PENCE, PENETRATION MECHANISMS OF ELECTRON BEAM WELDING AND THE SPIKING PHENOMENON, *WELD. J.*, VOL 53, 1974, P 246S-251S
15. H. TONG AND W.H. GIEDT, A DYNAMIC INTERPRETATION OF ELECTRON BEAM WELDING, *WELD. J.*, VOL 49, 1970, P 259S-266S
16. W.G. ESSERS AND R. WALTER, HEAT TRANSFER AND PENETRATION MECHANISMS WITH GMA AND PLASMA-GMA WELDING, *WELD. J.*, VOL 62, 1981, P 37S-42S
17. J.C. AMSON AND G.R. SALTER, ANALYSIS OF THE GAS-SHIELDED CONSUMABLE METAL-ARC WELDING SYSTEM, *BR. WELD. J.*, VOL 10, 1963, P 472-483
18. J.F. LANCASTER, *THE PHYSICS OF WELDING*, INTERNATIONAL INSTITUTE OF WELDING, PERGAMON PRESS, OXFORD, 1984, P 243
19. F. EICHHORN AND A. ENGEL, "MASS TRANSFER IN THE WELD POOL," DOCUMENT 201-70, INTERNATIONAL INSTITUTE OF WELDING, STUDY GROUP 212, 1970

Fluid Flow Phenomena During Welding

C.R. Heiple and P. Burgardt, EG&G Rocky Flats

Selected References

- P. BURGARDT, WELDING VARIABLE EFFECTS ON WELD SHAPE IN ELECTRON BEAM WELDING OF STEELS, *FERROUS ALLOY WELDMENTS*, TRANS TECH PUBLICATIONS, SWITZERLAND, 1992, P 269-328
- P. BURGARDT AND R.D. CAMPBELL, CHEMISTRY EFFECTS ON STAINLESS STEEL WELD PENETRATION, *FERROUS ALLOY WELDMENTS*, TRANS TECH PUBLICATIONS, SWITZERLAND, 1992, P 379-415
- C.R. HEIPLE AND P. BURGARDT, PENETRATION IN GTA WELDING, *WELDABILITY OF MATERIALS*, R.A. PATTERSON AND K.W. MAHIN, ED., ASM INTERNATIONAL, 1990, P 73-80
- C.R. HEIPLE AND J.R. ROPER, THE GEOMETRY OF GAS TUNGSTEN ARC, GAS METAL ARC, AND SUBMERGED ARC WELD BEADS, *WELDING: THEORY AND PRACTICE*, D.L. OLSON, R.D. DIXON, AND A.L. LIBY, ED., NORTH-HOLLAND ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, 1990, 1-34
- K.C. MILLS AND B.J. KEENE, FACTORS AFFECTING VARIABLE WELD PENETRATION,

INT.MATER.REVIEW, VOL 35, 1990, P 185-216

Transfer of Heat and Mass to the Base Metal in Gas-Metal Arc Welding

Herschel B. Smartt, Idaho National Engineering Laboratory

Introduction

HEAT AND MASS TRANSFER in arc welding is normally studied from the standpoint of the weld pool and heat-affected zone (HAZ); however, it is also instructive to examine heat and mass transfer from the arc to the base metal. This article describes the latter topic in terms of the gas-metal arc welding (GMAW) process and provides practical information related to the development of welding procedures and the general operation of the process.

Welding procedures emphasize control of parameters such as electrode speed (or current), voltage, welding speed, contact tube-to-base metal distance, as well as current pulse parameters for out-of-position welding. It is therefore easy to overlook the fact that the process is simply a source of heat and mass inputs to the weldment. Melting of the base metal, dilution of the filler metal, solidification of the weld bead, microstructural development in the weld bead and HAZ, and thermomechanical distortion and residual stresses all follow from the heat and mass inputs. The conventional parameters identified above arc variables that control the heat and mass inputs. An example of the relationship between the conventional parameters of electrode speed and welding speed to heat and mass transferred to the weldment is shown in Fig. 1 for certain conditions, as applied to the welding of thick-section steel.

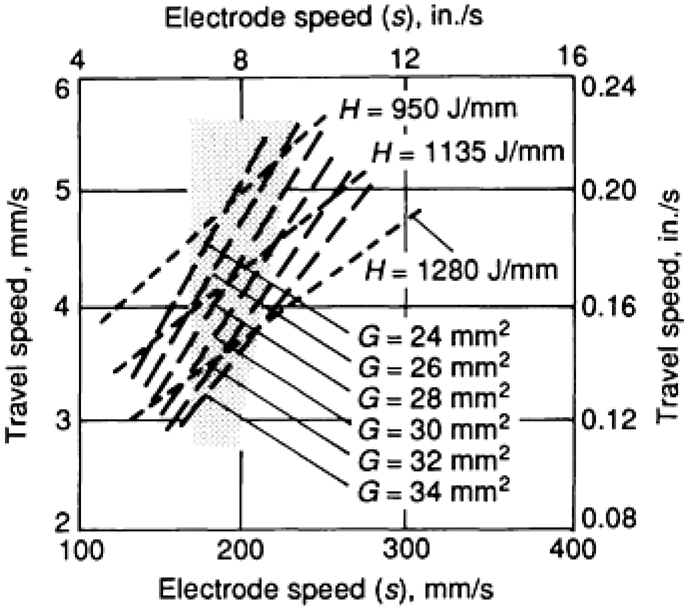

FIG. 1 PLOT OF WELDING SPEED VERSUS ELECTRODE SPEED AS FUNCTION OF HEAT TRANSFER PER LENGTH OF WELD, *H*, AND MASS TRANSFER EXPRESSED IN TERMS OF REINFORCEMENT, *G*. POWER SUPPLY OPEN-CIRCUIT VOLTAGE, E_0, IS 32 V; CONTACT TUBE-TO-BASE METAL DISTANCE, C_T, IS 15.9 MM (0.625 IN.). SHADED AREA DENOTES REGION IN WHICH SPRAY AND STREAMING TRANSFER MODES OCCUR; GLOBULAR TRANSFER OCCURS AT LOWER ELECTRODE SPEEDS, AND ELECTRODE CONTACTS THE WELD POOL AT HIGHER ELECTRODE SPEEDS. SOURCE: REF 1

The issues described in this article include the:

- TOTAL HEAT TRANSFERRED TO THE BASE METAL
- PARTITIONING OF HEAT TRANSFER BETWEEN THE ARC AND THE MOLTEN ELECTRODE

DROPLETS
- TRANSFER MODES OF THE DROPLETS
- ROLE OF THE ARC IN DROPLET TRANSFER
- SIMPLE MODEL FOR WELDING PROCEDURE DEVELOPMENT BASED ON AN UNDERSTANDING OF HEAT AND MASS TRANSFER TO THE BASE METAL

Reference

1. H.B. SMARTT, IDAHO NATIONAL ENGINEERING LABORATORY, 1992

Transfer of Heat and Mass to the Base Metal in Gas-Metal Arc Welding

Herschel B. Smartt, Idaho National Engineering Laboratory

Heat Transfer

The total transfer of heat, *H*, (neglecting preheating) from the GMAW process to the weldment per unit time is given by:

$$H = \eta EI \qquad \textbf{(EQ 1)}$$

where *E* is voltage, *I* is current, and η is the heat-transfer efficiency. The rate at which heat is transferred to the weldment per unit length of weld is given by:

$$H = \frac{hEI}{R} \qquad \textbf{(EQ 2)}$$

where *R* is welding speed. Calorimeter-based heat-transfer experiments reveal that the heat-transfer efficiency for welding thick-section steel is nominally 80 to 90%, as indicated in Fig. 2. The total heat-transfer efficiency is altered somewhat by changing other parameters. For example, it increases slightly as the power supply open-circuit voltage is decreased (for a silicon controlled rectifier regulated power supply and it increases slightly with increasing contact tube-to-base metal distance. However, 85% is a reasonable estimate for most conditions.

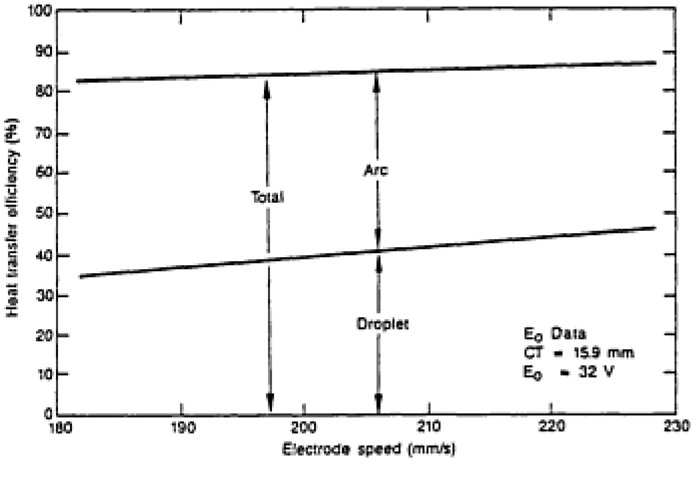

FIG. 2 PLOT OF HEAT-TRANSFER EFFICIENCY TO BASE METAL VERSUS ELECTRODE-SPEED FOR 0.89 MM (0.035 IN.) DIAMETER STEEL ELECTRODE IN AN AR-2% O_2 SHIELD GAS. TOTAL HEAT-TRANSFER EFFICIENCY

IS SHOWN PARTITIONED INTO ARC AND MOLTEN DROP COMPONENTS. POWER SUPPLY OPEN-CIRCUIT VOLTAGE, E_0, IS 32 V; CONTACT TUBE-TO-BASE METAL DISTANCE, C_T, IS 15.9 MM (0.625 IN.). SOURCE: REF 2

Partitioning of Heat Transfer. In the GMAW process, the molten droplets of electrode material carry a significant portion of the total heat transferred to the weld pool. This is seen in calorimetry experiments (Fig. 2), where the total heat-transfer efficiency of the GMAW process is partitioned into those portions associated with transfer by the arc and by the molten droplets. At low electrode speeds, about 60% of the total heat transferred is associated with the arc. As electrode speed increases, the fraction of total heat transferred associated with the droplets increases, reaching nominally 50% at current levels in excess of about 220 A (that is, at about 230 mm/s, or 9.1 in./s, electrode speed) for the conditions used.

Reference cited in this section

2. A.D. WATKINS, "HEAT TRANSFER EFFICIENCY IN GAS METAL ARC WELDING," MASTER'S THESIS, UNIVERSITY OF IDAHO, APRIL 1989

Transfer of Heat and Mass to the Base Metal in Gas-Metal Arc Welding

Herschel B. Smartt, Idaho National Engineering Laboratory

Mass Transfer

Droplet Transfer Modes. Although the International Institute of Welding (IIW) lists eight distinct metal transfer modes (Ref 3), the modes commonly used in U.S. welding practice are globular, spray, streaming, rotating, and short circuiting. The mode terms "drop" and "repelled" used by the IIW are often referred to as "globular," and the mode term "projected" is generally referred to as "spray." The globular, spray, streaming, and short-circuiting transfer modes are shown in Fig. 3, 4, 5, 6.

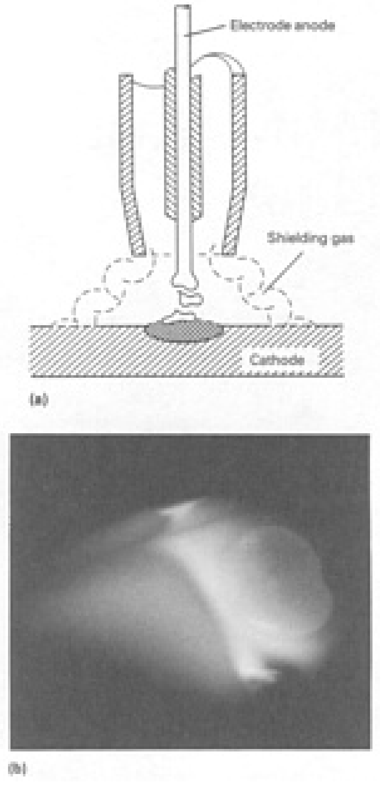

FIG. 3 GLOBULAR TRANSFER MODE IN GAS-METAL ARC WELDING OF STEEL. (A) SCHEMATIC SHOWING TRANSFER OF ELECTRODE MATERIAL GLOBULES ONTO CATHODE BASE METAL. (B) HIGH-SPEED PHOTOGRAPH OF GLOBULAR METAL TRANSFER

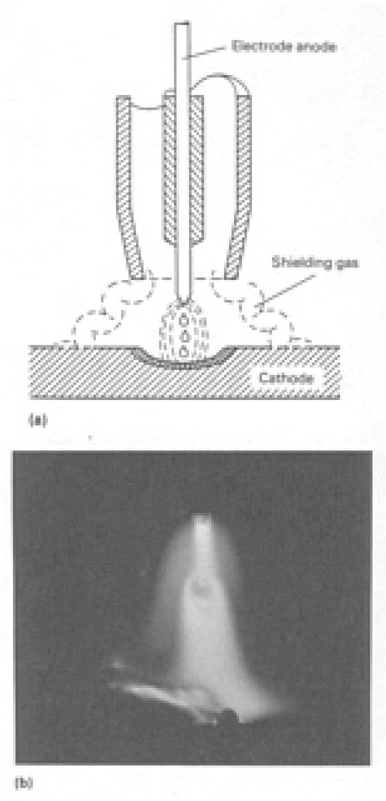

FIG. 4 SPRAY TRANSFER MODE IN GAS-METAL ARC WELDING OF STEEL. (A) SCHEMATIC SHOWING TRANSFER OF ELECTRODE MATERIAL DROPLETS ONTO CATHODE BASE METAL. (B) HIGH-SPEED PHOTOGRAPH OF SPRAY METAL TRANSFER MODE

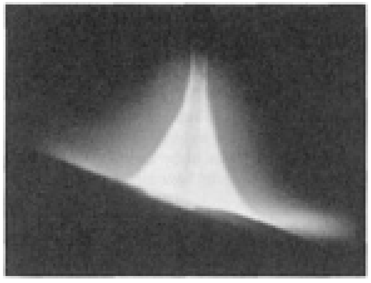

FIG. 5 HIGH-SPEED PHOTOGRAPH OF STREAMING TRANSFER MODE IN GAS-METAL ARC WELDING OF STEEL

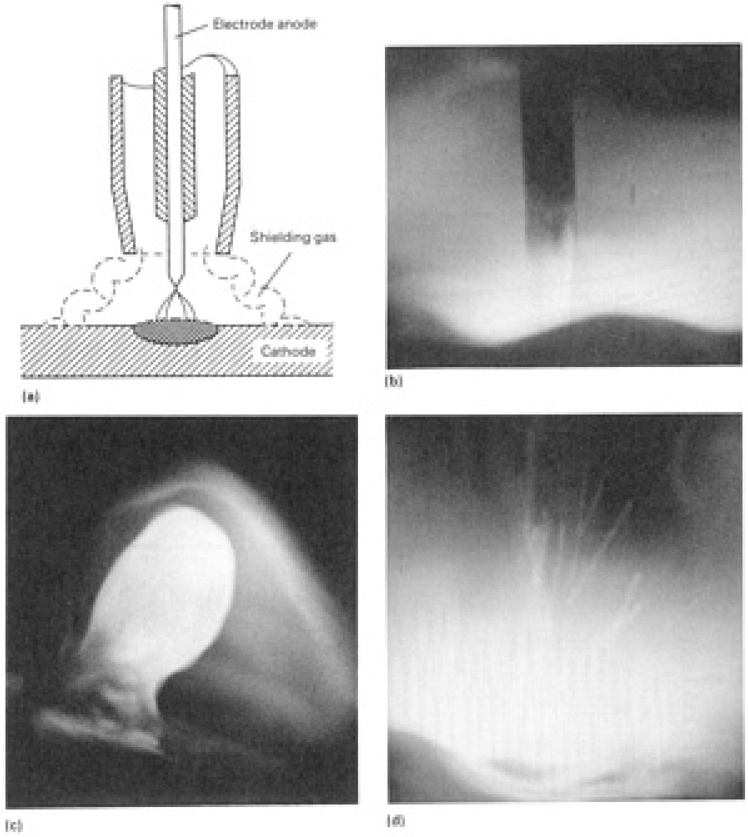

FIG. 6 SHORT-CIRCUITING TRANSFER MODE IN GAS-METAL ARC WELDING OF STEEL. (A) SCHEMATIC SHOWING TRANSFER OF ELECTRODE MATERIAL BY SURFACE TENSION OF WELD POOL ONTO CATHODE BASE METAL. (B) HIGH-SPEED PHOTOGRAPH OF MATERIAL TRANSFER WHEN ARC LENGTH IS VERY SHORT. (C) HIGH-SPEED PHOTOGRAPH OF CYCLIC SHORTING OF ARC BY THE ELECTRODE DURING METAL TRANSFER TO WELD POOL. (D) HIGH-SPEED PHOTOGRAPH OF VIOLENT ARC REIGNITION WITH ASSOCIATED SPATTER

When all other parameters are held constant, the metal transfer mode at the lowest wire feed speed (and associated current level) is globular. As wire speed (and therefore current) increases, the mode changes rapidly from globular to spray. With an additional increase in wire speed (and current), spray transfer becomes streaming transfer. Figure 7 shows the droplet sizes during the transition from globular to spray to streaming transfer for direct and pulsed current welding. If an adequately high current, contact tube-to-base metal distance, and voltage exist, then rotating transfer, wherein the lower part of the electrode becomes molten over a considerable length and rotates in a helical spiral under the influence of the magnetic field surrounding the arc, can occur. As it rotates, a controlled stream of droplets is transferred from the electrode tube to the weld pool over a relatively wide area. Additional increases in wire feed/current at low voltage shorten the arc length and, eventually, the wire stubs into the weld pool. In addition, with appropriate conditions (for example, carbon dioxide, argon-carbon dioxide mixtures, and helium-based shielding gases), droplets can be transferred directly, by surface tension forces, after contact of the drop with the weld pool, a condition called short-circuiting transfer.

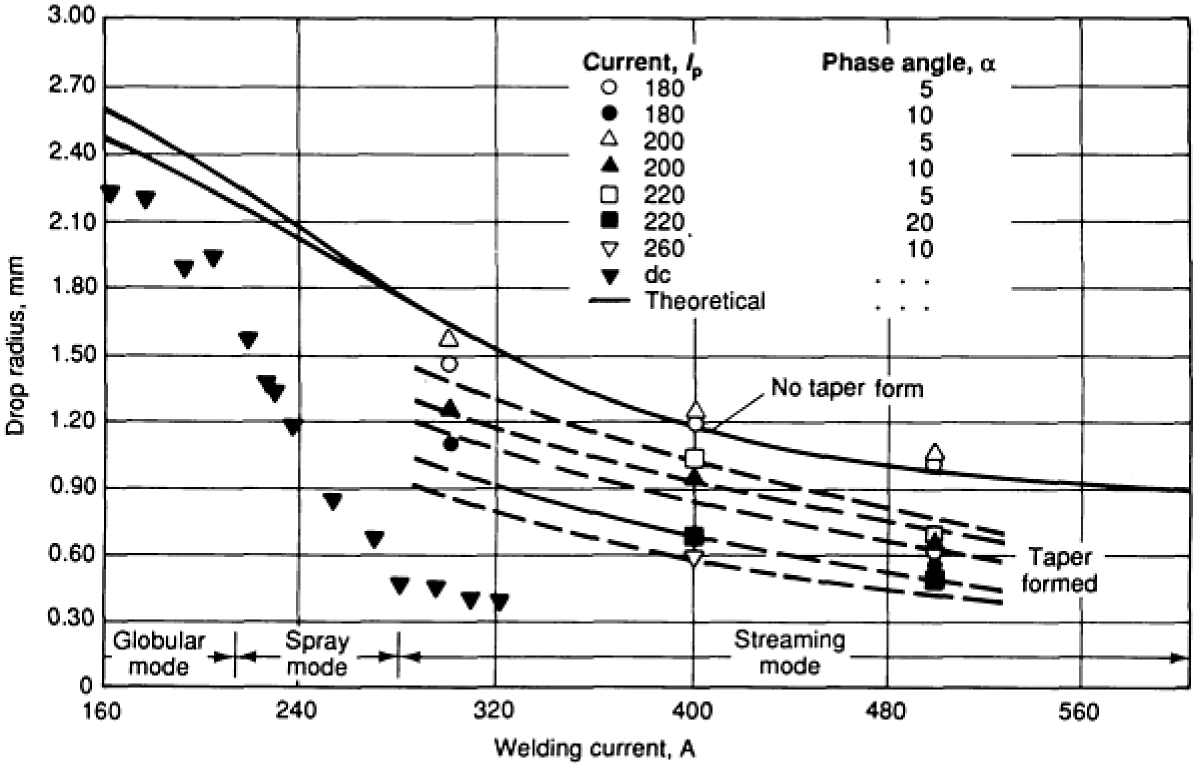

FIG. 7 DROPLET SIZES ATTAINABLE UNDER SELECTED CURRENT PULSING CONDITIONS IN GAS-METAL ARC WELDING OF STEEL. SOURCE: REF 4

In practical applications, the optimum transfer mode depends in part on the thickness of the base metal being welded. For example, very thin sections (in all positions) require the short-circuiting mode (with low current levels and appropriate settings of voltage and other operating parameters, including shielding gas composition). Thicker sections show best results with spray or streaming transfer. These transfer modes also produce high heat input, maximum penetration, and a high deposition rate. In welding steel, they are generally limited to welding that occurs in the flat position and the horizontal fillet position, except when pulsed current is used. Rotating transfer can be used in a deep groove in thick-section material. One- and two-pass heavy fillet welds are also a major area of application for this process variation.

Globular transfer (Fig. 3) involves a droplet that, generally, is much larger in diameter than the electrode wire. Globular transfer can involve a transfer rate of about 1 to 10 drops/s, and the arc has a soft, rounded appearance. Droplet detachment and transfer are mainly due to the gravitational force, which limits globular transfer to in-position welding. For spray transfer (Fig. 4), droplet and electrode diameters are roughly equivalent.

Spray transfer can involve 100 drops/s, and the arc noticeably contracts, or "stiffens." Drops usually travel in-line down the center of the arc, but several drops may be in flight at the same instant.

In streaming transfer (Fig. 5), a well-developed liquid column extends from the solid electrode down into the arc and breaks into small droplets before contacting the weld pool (see the discussion on spatter below). Streaming transfer can involve 1000 drops/s, and the arc has a characteristic "cone" shape, as shown in Fig. 5.

In short-circuiting transfer, the arc is very short (Fig. 6a). During metal transfer, the undetached molten droplet contacts the weld pool, shorting out the arc (Fig. 6b), which then extinguishes. Surface tension plays an important role in transferring the drop to the weld pool, but detachment of the drop from the electrode is due to electromagnetic pinch forces. Reignition of the arc is violent (Fig. 6c), resulting in considerable scatter.

As seen in Fig. 7, the transition from one mode to another actually involves continuous variations in droplet size. Because the product of droplet volume and transfer rate equals the electrode melting rate (in appropriate dimensions), there is also a continuous variation in droplet transfer rate. Detailed experimental studies have shown that rapid, cyclic transition from globular to spray to globular, and so on, occurs in the transition region between globular and spray transfer (Ref 5).

Although it is generally not possible for the welder to see droplet transfer events in the spray and streaming transfer modes (except under ideal conditions for spray transfer), the arc shape changes and associated changes in electrical and acoustic noise allow a trained welder to readily identify all of the transfer modes. The section "Electrical and Acoustic Signals" in this article describes this event more fully.

Weld Reinforcement. Assuming that mass is conserved as welding occurs, the transverse cross-sectional area of the weld bead added to the weldment for a single weld pass, which is called the reinforcement, *G,* is given by:

$$G = \frac{S}{R}\frac{p d^2}{4} \qquad \textbf{(EQ 3)}$$

where *S* is electrode speed, *R* is welding speed, and *d* is electrode diameter. For the spray and streaming transfer modes, the assumption that mass is conserved is reasonably good. However, because short circuiting and other conditions generate considerably spatter, the actual reinforcement will be slightly less than the value calculated by Eq 3.

Droplet Velocity and Temperature. In spray and streaming transfer, the droplets are accelerated rapidly through the arc to the weld pool. Velocities of 1 m/s (40 in./s) are typical, increasing with voltage (Fig. 8). Calorimetry-based experimental results indicate that steel droplets can reach temperatures of approximately 2600 K (4220 °F) (Ref 2). This explains why the droplets transport one-half of the total heat transferred to the base metal.

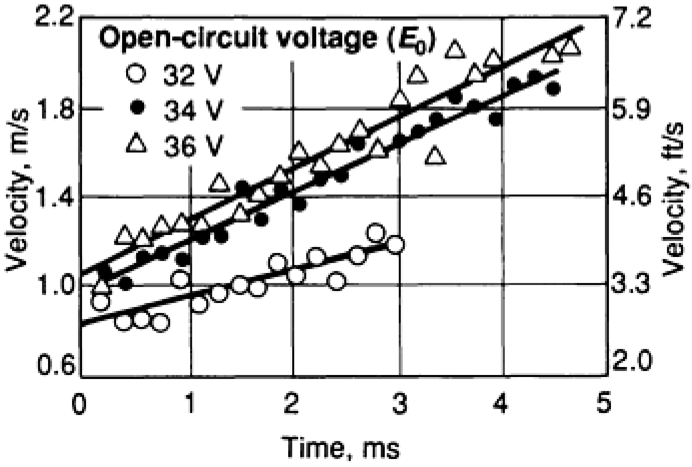

FIG. 8 DROPLET VELOCITIES FOR SPRAY TRANSFER IN GAS-METAL ARC WELDING OF STEEL AT THREE DIFFERENT OPEN-CIRCUIT VOLTAGES. SOURCE: REF 5

A result of the activity of molten electrode droplets in heat transfer to the base metal is that they also play an important role in convective heat transport in the weld pool and, thus, in weld penetration (Ref 6). This can be seen in Fig. 9, a transverse cross section of a gas-metal arc bead-on-plate weld on carbon steel. The region of deep penetration in the center of the weld bead is associated with the heat convected to the lower portion of the weld pool by the entering droplets.

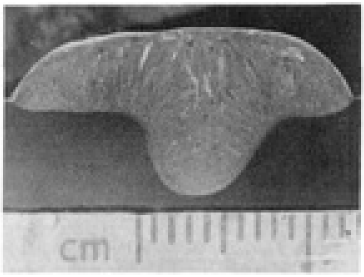

FIG. 9 TRANSVERSE CROSS SECTION OF GAS-METAL ARC BEAD-ON-PLATE WELD IN CARBON STEEL TO SHOW DEEP PENETRATION IN THE WELD BEAD CENTER GENERATED BY MOLTEN ELECTRODE DROPLETS

Electrical and Acoustic Signals. As the droplets of molten metal detach from the electrode, an almost instantaneous change occurs in the electrode extension, which results in a sudden, although small, change in the electrical resistance between the contact tube and the base metal. The results are spikes in the secondary circuit voltage and welding current, accompanied by pressure changes in the arc. Thus, both electrical and acoustic noise (Fig. 10) are generated by the GMAW process and are characteristic of the droplet transfer mode. The waveforms for globular transfer show prominent spikes associated with individual droplet transfer events. This can also be seen in Fig. 11, a plot of the computer-digitized current for a weld on carbon steel, where a change from globular to spray transfer has occurred at about 8 s into the weld. The power spectra of the current changes dramatically for the two transfer modes (Fig. 12), leading to a means of detecting transfer mode during welding.

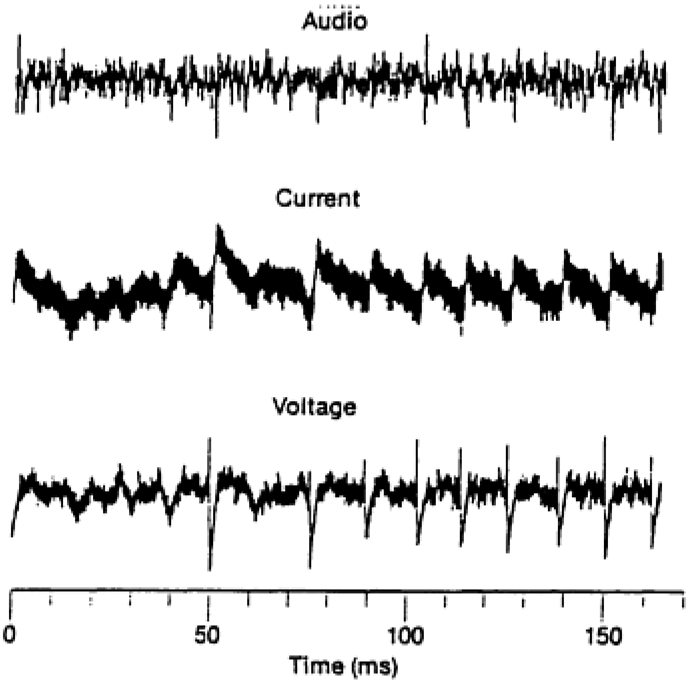

FIG. 10 AUDIO, CURRENT, AND VOLTAGE DURING GLOBULAR TRANSFER IN GAS-METAL ARC WELDING OF STEEL. SOURCE: REF 7

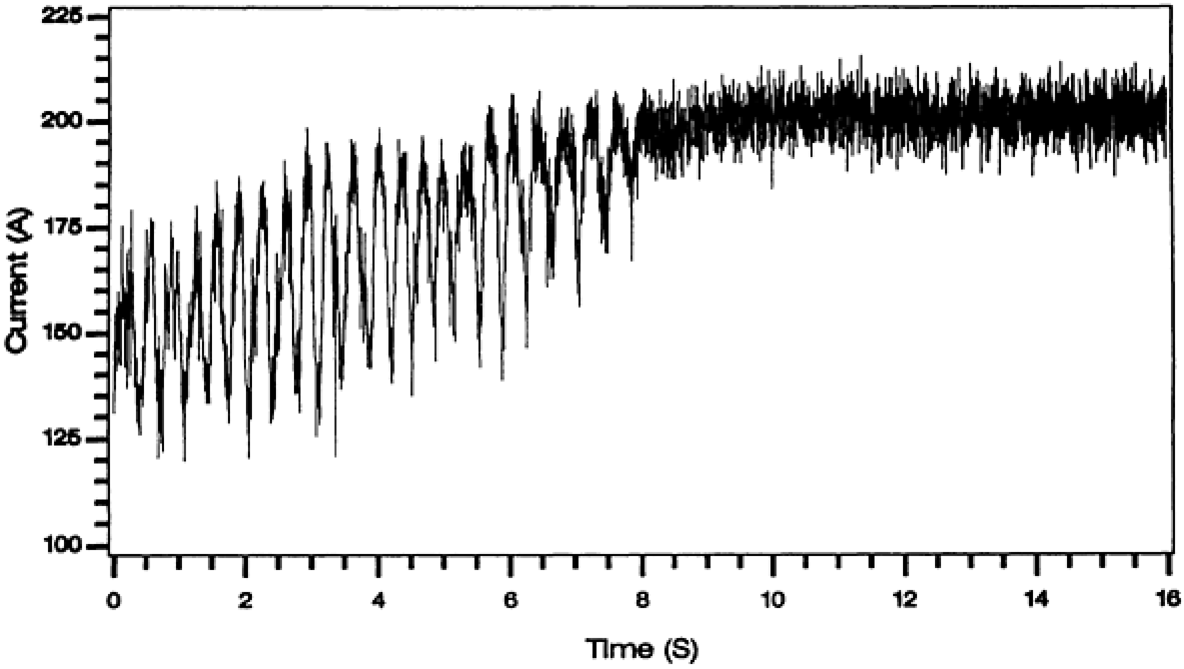

FIG. 11 PLOT OF DIGITIZED CURRENT VERSUS TIME DURING A GAS-METAL ARC WELD IN CARBON STEEL IN WHICH A TRANSITION FROM GLOBULAR TO SPRAY TRANSFER OCCURRED (AT ~*T* = 8 S) AS CURRENT WAS INCREASED. SOURCE: REF 1

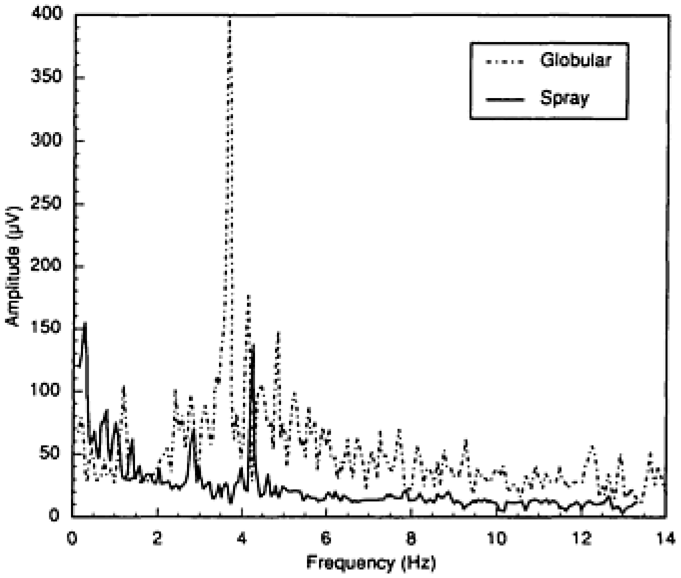

FIG. 12 WELDING CURRENT POWER SPECTRA FOR GLOBULAR TRANSFER AND SPRAY TRANSFER MODES DURING GAS-METAL ARC WELDING OF STEEL. SOURCE: REF 1

Several specific techniques have been identified (Ref 8) for detecting the mode of metal transfer: Fourier transform, standard deviation, peak ratios, and integrated amplitude of the current and voltage signal. In a similar manner, spatter, lack of shielding gas, and contact tube wear can also be detected. Similar power spectra changes also occur in the secondary circuit voltage and acoustic noise. The change in acoustic noise allows the aural detection of the transfer mode. An experienced welder can readily hear, for example, the difference between globular and spray transfer.

Contact of the molten droplet with the weld pool, while the droplet is still attached to the electrode, also results in a sudden change in electrical resistance. This is important, because spatter is generally produced when this electrode-to-weld pool contact breaks. (Spatter can also be produced by rapid expansion of gas bubbles in the electrode as it melts.) The resulting electrical and acoustic noise can be used to detect that spatter is being produced.

References cited in this section

1. H.B. SMARTT, IDAHO NATIONAL ENGINEERING LABORATORY, 1992
2. A.D. WATKINS, "HEAT TRANSFER EFFICIENCY IN GAS METAL ARC WELDING," MASTER'S THESIS, UNIVERSITY OF IDAHO, APRIL 1989
3. J.F. LANCASTER, *THE PHYSICS OF WELDING*, 2ND ED., INTERNATIONAL INSTITUTE OF WELDING, PERGAMON PRESS, 1986, P 233
4. T.W. EAGAR, D.E. HARDT, H.B. SMARTT, AND J.A. JOHNSON, PROCESS CONTROLLABILITY IN GAS METAL ARC WELDING: MICRO AND MACROSCOPIC PROBLEMS, *PROC. EIGHTH SYMPOSIUM ON ENERGY SCIENCES*, CONF-9005183, U.S. DEPARTMENT OF ENERGY, ARGONNE NATIONAL LABORATORY, MAY 1990, P 51-67
5. D.E. CLARK, C. BUHRMASTER, AND H.B. SMARTT, DROPLET TRANSFER MECHANISMS IN GMAW, *PROC. 2ND INTERNATIONAL CONFERENCE ON TRENDS IN WELDING RESEARCH* (GATLINBURG, TN), 1989
6. W.G. ESSERS AND R. WALTER, HEAT TRANSFER AND PENETRATION MECHANISMS WITH GMA AND PLASMA-GMA WELDING, *WELD.J.*, VOL 60 (NO. 2), 1981, P 37-S TO 42-S
7. J.A. JOHNSON, N.M. CARLSON, AND H.B. SMARTT, DETECTION OF METAL TRANSFER MODE IN GMAW, *PROC. 2ND INTERNATIONAL CONFERENCE ON TRENDS IN WELDING RESEARCH* (GATLINBURG, TN), 1989
8. G. ADAM AND T.A. SIEWERT, SENSING OF GMAW DROPLET TRANSFER MODES USING AN ER100S-1 ELECTRODE, *WELD.J.*, VOL 69 (NO. 3), MARCH 1990, P 103-S TO 108-S

Transfer of Heat and Mass to the Base Metal in Gas-Metal Arc Welding

Herschel B. Smartt, Idaho National Engineering Laboratory

Procedure Development

The preceding discussion leads to a logical approach toward selecting welding parameters during the development of welding procedures. First, given a nominal power supply open-circuit voltage and contact tube-to-base metal distance, the electrode speed can be set to determine the welding current. It can be shown that:

$$S = \frac{4hGI}{Hp\,d^2}(E_0 + HI) \qquad \textbf{(EQ 4)}$$

where E_0 is the power supply open-circuit voltage. It is assumed that the proper selection of electrode type and shielding gas have been made, and that the electrode diameter is appropriate for the application.

Second, the open-circuit voltage is adjusted to give either an acceptable arc length or electrode extension. The objective is to prevent either burnback (transfer of the arc to the contact tube) or spatter caused by shorting of the molten electrode droplets with the weld pool prior to detachment from the electrode. Fine adjustment of voltage may be necessary to obtain a clean start of the process, thus avoiding a short period of spatter or globular transfer following arc ignition. The combination of current and voltage for a given contact tube-to-base metal distance determines the droplet transfer mode.

Third, the welding current determines the melting rate of the electrode. The electrode melting rate (M_{rp}) is a quadratic function of current (Ref 9):

$$M_{rp} = (C_1 + C_2 A)I + C_3 \frac{LI^2}{A^2} \qquad \textbf{(EQ 5)}$$

where A is the cross-sectional area of the electrode and C_1, C_2, C_3, and i are constants.

Another researcher (Ref 10) defines M_{rp} by:

$$M_{rp} = C_4 I + C_5 \frac{LI^2}{A} \qquad \textbf{(EQ 6)}$$

where C_4 and C_5 are constants.

Given a desired melting rate, the welding travel speed is set to obtain the desired weld bead reinforcement. The current, voltage, and weld travel speed have all now been set, thus determining the heat input per length of weld ($H = \eta EI/R$).

It should be noted that it is actually possible to independently vary heat input and mass input (in terms of weld bead reinforcement) to the weld, over at least a small range. This is shown in Fig. 13, where the same data presented in Fig. 1 are replotted in terms of reinforcement as a function of heat input per length of weld.

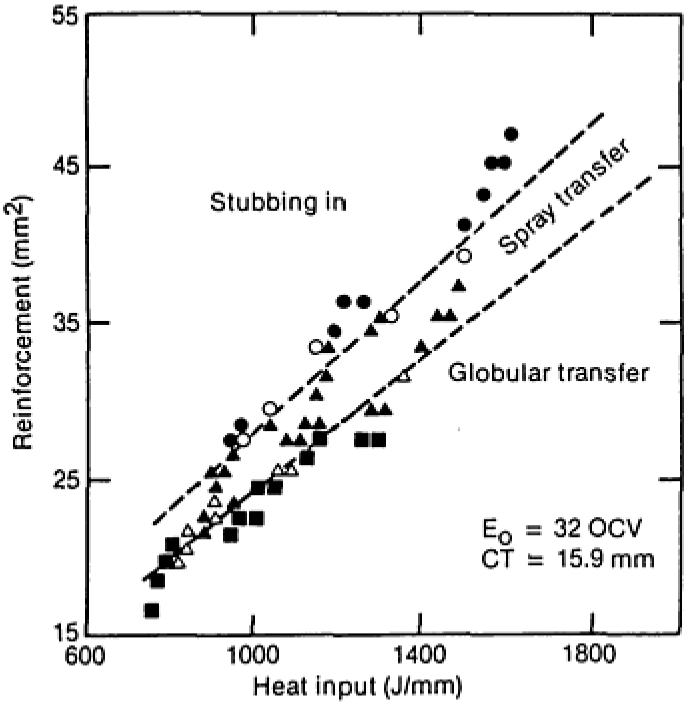

FIG. 13 PLOT OF REINFORCEMENT VERSUS HEAT INPUT TO SHOW PARAMETERS THAT FAVOR SPRAY TRANSFER MODE FOR GAS-METAL ARC WELDING OF STEEL. POWER SUPPLY OPEN-CIRCUIT VOLTAGE, E_0, IS 32 V; CONTACT TUBE-TO-BASE METAL DISTANCE, C_T, IS 15.9 MM (0.625 IN.).

The above discussion on procedure development ignores second-order effects, such as the dependence of arc length on weld travel speed. It should also be realized that several iterations through the steps defined above may be required for final parameter determination. Proper power supply settings must be used, and code requirements must be met.

References cited in this section

9. A. LESNEWICH, CONTROL OF MELTING RATE AND METAL TRANSFER IN GAS-SHIELDED

METAL-ARC WELDING, PART II--CONTROL OF METAL TRANSFER, *WELD.J.*, VOL 37 (NO. 9), SEPT 1958, P 418-S TO 425-S

10. J.H. WASZINK AND G.J.P.M. VAN DEN HEUVEL, HEAT GENERATION AND HEAT FLOW IN THE FILLER METAL IN GMA WELDING, *WELD.J.*, VOL 61 (NO. 8), AUG 1982, P 269-S TO 282-S

Transfer of Heat and Mass to the Base Metal in Gas-Metal Arc Welding

Herschel B. Smartt, Idaho National Engineering Laboratory

References

1. H.B. SMARTT, IDAHO NATIONAL ENGINEERING LABORATORY, 1992
2. A.D. WATKINS, "HEAT TRANSFER EFFICIENCY IN GAS METAL ARC WELDING," MASTER'S THESIS, UNIVERSITY OF IDAHO, APRIL 1989
3. J.F. LANCASTER, *THE PHYSICS OF WELDING*, 2ND ED., INTERNATIONAL INSTITUTE OF WELDING, PERGAMON PRESS, 1986, P 233
4. T.W. EAGAR, D.E. HARDT, H.B. SMARTT, AND J.A. JOHNSON, PROCESS CONTROLLABILITY IN GAS METAL ARC WELDING: MICRO AND MACROSCOPIC PROBLEMS, *PROC. EIGHTH SYMPOSIUM ON ENERGY SCIENCES*, CONF-9005183, U.S. DEPARTMENT OF ENERGY, ARGONNE NATIONAL LABORATORY, MAY 1990, P 51-67
5. D.E. CLARK, C. BUHRMASTER, AND H.B. SMARTT, DROPLET TRANSFER MECHANISMS IN GMAW, *PROC. 2ND INTERNATIONAL CONFERENCE ON TRENDS IN WELDING RESEARCH* (GATLINBURG, TN), 1989
6. W.G. ESSERS AND R. WALTER, HEAT TRANSFER AND PENETRATION MECHANISMS WITH GMA AND PLASMA-GMA WELDING, *WELD.J.*, VOL 60 (NO. 2), 1981, P 37-S TO 42-S
7. J.A. JOHNSON, N.M. CARLSON, AND H.B. SMARTT, DETECTION OF METAL TRANSFER MODE IN GMAW, *PROC. 2ND INTERNATIONAL CONFERENCE ON TRENDS IN WELDING RESEARCH* (GATLINBURG, TN), 1989
8. G. ADAM AND T.A. SIEWERT, SENSING OF GMAW DROPLET TRANSFER MODES USING AN ER100S-1 ELECTRODE, *WELD.J.*, VOL 69 (NO. 3), MARCH 1990, P 103-S TO 108-S
9. A. LESNEWICH, CONTROL OF MELTING RATE AND METAL TRANSFER IN GAS-SHIELDED METAL-ARC WELDING, PART II--CONTROL OF METAL TRANSFER, *WELD.J.*, VOL 37 (NO. 9), SEPT 1958, P 418-S TO 425-S
10. J.H. WASZINK AND G.J.P.M. VAN DEN HEUVEL, HEAT GENERATION AND HEAT FLOW IN THE FILLER METAL IN GMA WELDING, *WELD.J.*, VOL 61 (NO. 8), AUG 1982, P 269-S TO 282-S

Arc Physics of Gas-Tungsten Arc Welding

J.F. Key, EG&G Idaho, Inc.

Introduction

THE GAS-TUNGSTEN ARC WELDING (GTAW) process is performed using a welding arc between a nonconsumable tungsten-base electrode and the workpieces to be joined. C.E. Jackson defined a welding arc as "a sustained electrical discharge through a high-temperature conducting plasma producing sufficient thermal energy so as to be useful for the joining of metals by fusion." This definition is a good foundation for the discussion that follows.

The physics of GTAW are fundamental to all arc processes and are more straightforward, because the complications of materials (for example, filler and flux) transferred through and interacting with the arc can be avoided. Geometrically, the arc discharge in GTAW is between a rod-shaped tungsten electrode and a planar-shaped electrode, that is, the workpiece.

Pure tungsten electrodes are less expensive and, possibly, more environmentally compatible than those with rare earth or other oxide additions. They are used for lower-specification welds, where tungsten contamination that is caused by the molten electrode surface can be tolerated. They are also used for alternating current (ac) welding of aluminum, copper, magnesium, and thin sections of low-alloy and stainless steels.

Analysis of the arc discharge is separated into electrode regions and the arc column. The electrode regions are confined to very small distances from the electrode surfaces, have very high electrical and thermal fields, and have much higher current density, because of the contraction of the arc to a small spot. As a result, electrode regions for both the cathode and the anode are difficult to analyze by diagnostic measurements and theoretical computation. This situation must be remedied for a thorough understanding of the process, because the process parameters control the arc discharge at the cathode, with the anode serving as the connection to ground. The arc column, on the other hand, is relatively easy to analyze, but is important primarily as a means to deduce arc characteristics at the electrodes.

Polarity. The GTAW process generally utilizes a direct current (dc) arc, where the tungsten electrode has a negative polarity. The tungsten electrode thus becomes the cathode and the workpiece becomes the anode. The polarity is called straight polarity, or direct current electrode negative (DCEN).

Reverse polarity, or direct current electrode positive (DCEP), is literally the reverse of DCEN. The workpiece is the cathode and the tungsten electrode serves as the anode. Because most heat is generated at the anode in the GTAW process, DCEP is used for welding certain thin-section, low melting point materials when DCEN would be likely to cause excessive penetration or burn-through.

Either alternating current or DCEP is used for removing an oxide film from the surface of the weld pool or workpiece. The oxide film promotes emission during the half-cycle (ac) when the workpiece is negative polarity. As the oxide is depleted, the emission moves to a new location that has a high enough oxide content to sustain the discharge of electrons. The arc root or cathode spot where the emission occurs is highly mobile in ac or DCEP and, as a result, the arc is much less stable than in DCEN.

Gas Shielding. In all cases, the arc and both electrodes are shielding by gas, usually an inert gas or a gas mixture. Argon and argon-helium mixtures are used most often, although argon-hydrogen mixtures are used for some applications. The GTAW process may simply utilize the arc to fuse the workpieces together without the addition of filler materials (autogenous) or filler may be added to the molten pool to fill grooves in thicker weldments. A reasonable understanding of welding arc fundamentals and the GTAW process requires a more thorough discussion of the electrode regions of the arc and the arc column.

Arc Physics of Gas-Tungsten Arc Welding

J.F. Key, EG&G Idaho, Inc.

Electrode Regions and Arc Column

The cathode and anode are similar in several respects. Both exhibit a voltage drop caused by a space charge that covers a very thin region over their surfaces, and the arc is significantly contracted on the surfaces. Figure 1 shows that the total arc voltage is partitioned between the electrode drops and arc column. The relative magnitude of these drops depends on welding parameters and electrode material.

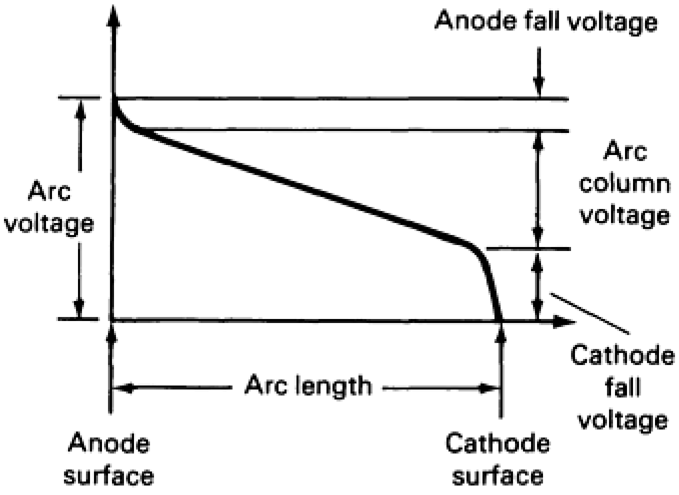

FIG. 1 PLOT OF RELATIVE ARC VOLTAGE DISTRIBUTION VERSUS RELATIVE ARC LENGTH BETWEEN ELECTRODES

The arc discharge requires a flow of electrons from the cathode through the arc column to the anode, regardless of polarity or whether ac or dc is used. Two cases of electron discharge at the cathode will be discussed: thermionic emission and nonthermionic emission, also called cold cathode, or field emission.

Thermionic emission results from joule heating (resistance) of the cathode by the imposed welding current until the electron energy at the cathode tip exceeds the work function (energy required to strip off an electron). This case applies to the general case of DCEN, where the tungsten electrode is the emitter, or cathode. Pure tungsten electrodes have to be heated to their melting point to achieve thermionic emission. Once molten, the equilibrium tip shape becomes a hemisphere, and a stable arc results from uniform emission over this surface. Thoria (ThO_2), zirconia (ZrO_2), or ceria (CeO_2) are added to pure tungsten in amounts up to 2.2 wt% ThO_2, 0.4 wt% ZrO_2, or 3.0 wt% CeO_2 to lower the work function, which results in thermionic emission at lower temperatures and avoids melting the cathode tip. These electrodes typically have a ground conical tip, and thermionic emission is localized to a cathode spot. Thermionic emission creates a cloud of electrons, called a space charge, around the cathode. If a second electrode at a higher potential is nearby (the workpiece, in this case), then the electrons will flow to it, thus establishing the arc.

Nonthermionic, or field, emission creates an electron discharge with a very high electric field, typically exceeding 10^9 V/m. This intense electric field literary pulls electrons out of a relatively cold or unheated cathode. This would not appear to be applicable to welding until one considers that for reverse polarity or DCEP, a condensation of positive ions from the arc column can build up in a very thin (1 nm, or 0.04 μin.) layer over the cathode surface, creating a very high localized electric field even though the cathode voltage drop may only amount to several volts.

An oxide layer, which is always present on the cathode surface in an actual weld, facilitates the discharge with a source of lower work function electrons. When the oxide layer is very thin (on the order of one atom layer), emission occurs via a tunneling mechanism through the film to an emitting site. Thicker oxide films exhibit locally conducting spots at the end of filamentary channels through the oxide. Large currents flow in these channels, which are on the order of 100 nm (4 μin.) in diameter and have lifetimes of 0.001 to 1 μs.

The cathode cleaning action, which is one of the principal reasons to use DCEP or ac, results from stripping away the oxide film at the emitting sites by very small and intense jets of metal vapor and debris. It becomes obvious that a practical implication of the short lifetime of these cathode spots is a generally unstable arc that is due to the necessity of continual movement of the cathode spot to undepleted regions of oxide film. Arc instability is undesirable and DCEP or ac is only used when cathodic cleaning or the minimizing of heat input to the workpiece is a higher priority than optimizing weld bead shape and location.

Anode. Welding process parameters (for example, current and voltage) control the arc discharge at the cathode. Although the electron flow enters the anode through the anode spot and constitutes 85% of the energy going into the weld pool, thus making current density the single most important welding parameter that determines pool shape, events at the anode can only be controlled indirectly by controlling the cathode. Anode spot stability does depend somewhat on shielding gas composition and the shape of the anode (that is, weld groove).

Current density and heat input measurements at the anode have been made to better understand how process parameters that are largely controlling events at the cathode will, in turn, influence the shape and melting rate of the weld pool. The relative contributions of heat transfer to the workpiece, in terms of the GTAW process, are shown in Fig. 2.

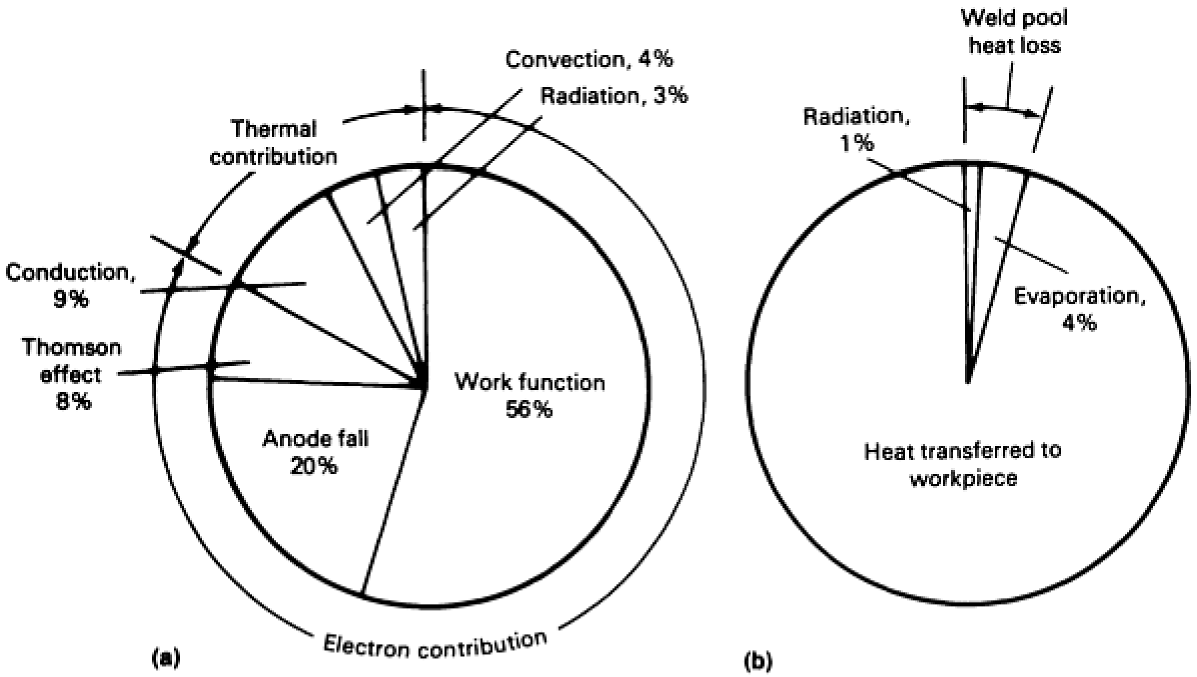

FIG. 2 RELATIVE HEAT TRANSFER CONTRIBUTIONS TO WORKPIECE WITH GTAW. (A) CONTRIBUTION OF INDIVIDUAL PARAMETERS TO ANODE HEAT INPUT. (B) HEAT OUTPUT AT CATHODE (WORKPIECE) RELATIVE TO WELD POOL HEAT LOSS

The Thomson effect represents the energy lost by electrons as they move from higher to lower temperatures. The sum of work function, Thomson effect, and anode fall gives an electron contribution to heat transfer of approximately 84%. The remaining 16% is due to thermal effects (that is, conduction, convection, and radiation). There are small heat losses from the pool that are due to evaporation of metal ions and radiation. Figure 3 shows that the energy distribution for this particular case approaches a Gaussian distribution (that is, normal distribution curve). High helium contents in the shielding gas have produced data that are typically better fit by a Lorentzian curve, indicating a narrower current density distribution (that is, a more-contracted arc at the anode spot).

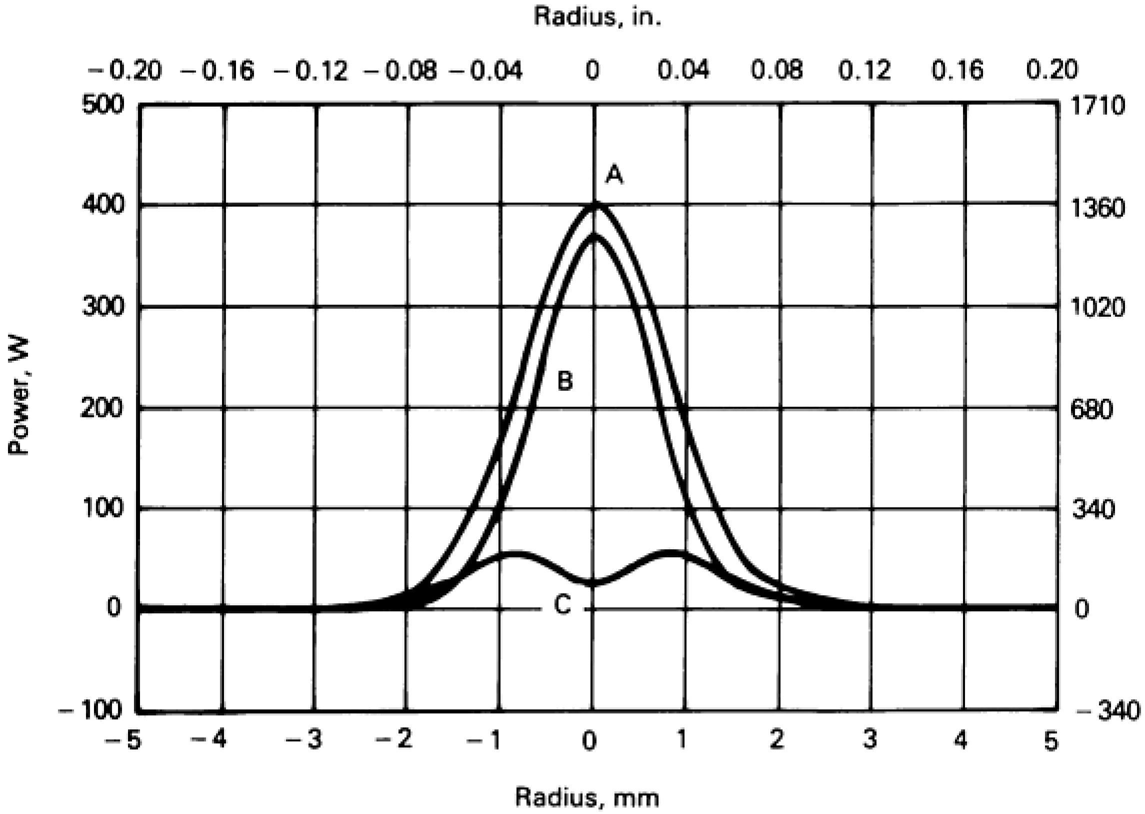

FIG. 3 PLOT OF ELECTRON AND THERMAL CONTRIBUTIONS TO HEAT TRANSFER. A, TOTAL ARC POWER (STANDARD DEVIATION, *s* , OF 0.8 MM, OR 0.031 IN.); B, ELECTRON CONTRIBUTION (σ = 0.7 MM OR 0.028 IN.); C, THERMAL CONTRIBUTION. WELD PARAMETERS: CURRENT, 10 A; VOLTAGE, 10 V; TIME, 10 S; SHIELDING GAS, ARGON; ELECTRODE ANGLE, 30°

Arc efficiency, in addition to those variables that have an effect on it, is an extremely important term in the heat transfer analysis of welding. It gives the percentage of heat dissipated in the arc that actually is captured by the workpieces and is available for melting. Arc efficiency, as a function of all GTAW welding parameters and many materials, has been determined experimentally and found to be nominally 75%. The variables having the greatest effect on arc efficiency are arc voltage and anode material. For those variables, the effect is usually no more than ±5%. Other parameters have a negligible effect.

Arc Column. The electron discharge between the electrodes partially ionizes the shielding gas in its path, thus making the arc column a conductor, or plasma. Overall, the arc column is neutral and is composed of electrons, positive gas and possibly metal ions, and neutral gas atoms.

Ironically, fundamental measurements of arc properties are most easily made in the arc column, although the actual effect of these properties on the electrode region of an actual weld must still be inferred. Nevertheless, it is useful to understand fundamentals of the arc that relate essential welding variables (for example, current, voltage, electrode gap, choice of shielding gas, and electrode shape) to arc temperature, current density distribution, and gas flow structure at the anode surface.

Effect of Cathode Tip Shape. For the general case of straight polarity, DCEN, the tip of the tungsten alloy cathode is ground to a point and then truncated somewhat to prevent the sharp tip from burning off and contaminating the weld. The included angle of the cone and the diameter of the truncation under some welding conditions have a significant effect on weld pool shape. Figure 4 shows examples of the effect of these two parameters on weld pool shape.

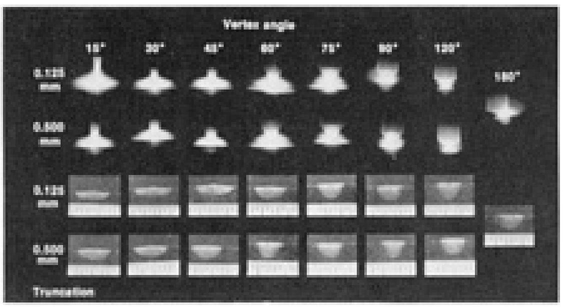

FIG. 4 FUSION ZONE PROFILE FOR SPOT-ON-PLATE WELDS AS A FUNCTION OF ELECTRODE TIP GEOMETRY USING 100% AR AS A SHIELDING GAS. WELD PARAMETERS: CURRENT, 150 A; DURATION, 2 S

For a stationary spot-on-plate weld shielded by pure argon, the weld depth-to-width ratio increased with an increasing vertex angle up to 90° and with an increasing truncation diameter. The arc became less "bell shaped" and more "ball shaped" as the vertex angle or truncation diameter increased. These results should be a valid indication of the effect of cathode tip shape for pulsed current welding, which produces a series of overlapping spot welds.

A study of bead-on-plate welds (Fig. 5) made with constant current and velocity indicated a similar but less pronounced trend. These conditions produce a tear-drop molten pool shape when viewed from above, compared to circular shape for spot or pulsed current welding. Fluid and heat flow within the pool is less uniform front-to-rear in a tear-drop-shaped pool and probably has a greater influence on pool shape than electrode tip shape.

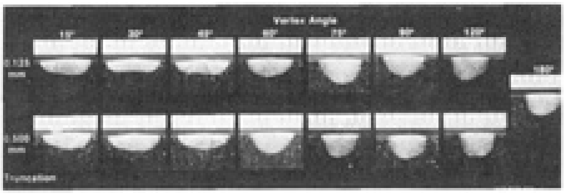

FIG. 5 FUSION ZONE PROFILE FOR BEAD-ON-PLATE WELDS AS A FUNCTION OF ELECTRODE TIP GEOMETRY USING 100% AR AS A SHIELDING GAS. WELD PARAMETERS: CURRENT, 150 A; WELDING SPEED, 3 MM/S (0.12 IN./S)

When the arc is used in a weld groove, the relative shapes of the cathode tip and the anode groove become more important. The arc discharge from the cathode will seek a path to ground with the lowest electrical resistance. For a stable arc properly centered in the groove (for example, a root pass), the shortest path to ground should be between the cathode tip and the bottom of the groove (Fig. 6). This will require the cathode vertex angle to be somewhat less than the included angle of the groove and/or a sufficiently wide groove to ensure that the shortest path to ground is from the cathode tip to the groove bottom and not, for example, from the electrode shoulder to the groove wall, as the case would be with a 90° electrode in a 10° narrow groove. Welding in a groove places a higher priority on arc stability and location than on maximum penetration.

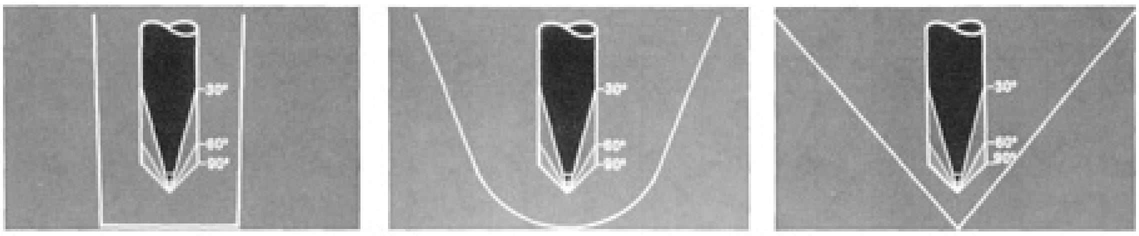

FIG. 6 EFFECTS OF ELECTRODE TIP GEOMETRY ON THE PATH LENGTH TO GROUND IN WELD GROOVES OF VARIOUS SHAPES. (A) 75° V-GROOVE. (B) 40° U-GROOVE. (C) 10° NARROW GROOVE

To understand the effects of tip shape, temperature distributions in the plasma were measured. Figure 7 shows that as the cathode vertex angle increases, the plasma radius of the arc column increases at mid-gap and becomes more constricted near the anode. The quantitative interpretation of these results requires theoretical modeling, which has yet to be completed.

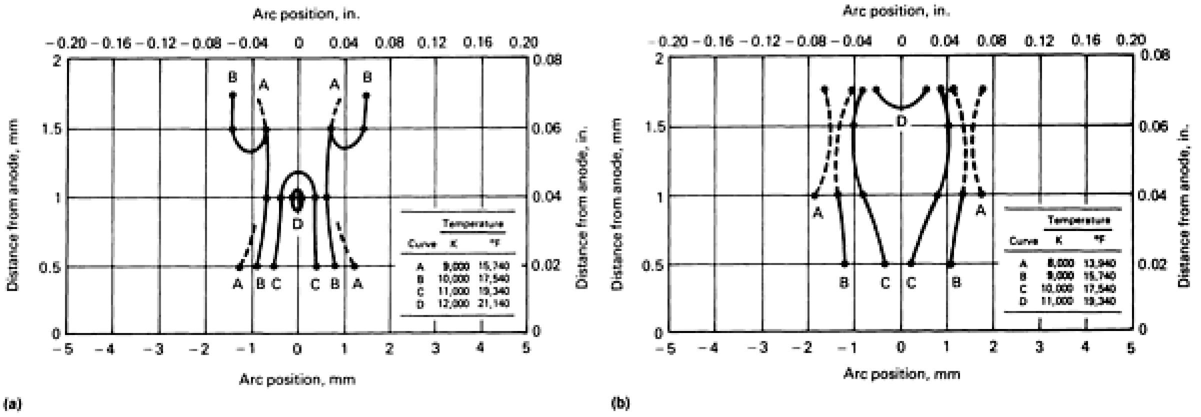

FIG. 7 EFFECT OF VERTEX ANGLE ON GTAW ARC COLUMN TEMPERATURE DISTRIBUTION WITH 100% AR USED AS SHIELDING GAS. (A) 30° ELECTRODE VERTEX ANGLE. (B) 90° ELECTRODE VERTEX ANGLE. WELDING CURRENT, 150 A

Effect of Shielding Gas Composition. The GTAW process typically uses either an inert gas, such as argon, or an inert gas mixture, such as argon and helium, to shield the arc and the weld from atmospheric contamination. Occasionally, a slightly reactive gas mixture, such as argon with up to 15 vol% H_2, is used. (The 15 vol% limit is based on safety considerations.) Shielding gas composition has a rather strong effect on arc temperature distribution and, under certain conditions, a significant effect on weld pool shape.

Figure 8 shows how shielding gas affects arc voltage. The curves would all be displaced downward for shorter arc lengths, but the relative positions would be maintained. Figure 9 shows the effects of both cathode vertex angle and shielding gas composition on weld pool shape for spot-on-plate welds. Increasing additions of helium to argon show a remarkable increase in penetration when using a 30° vertex angle. However, the effect is much less evident when using a 90° vertex angle.

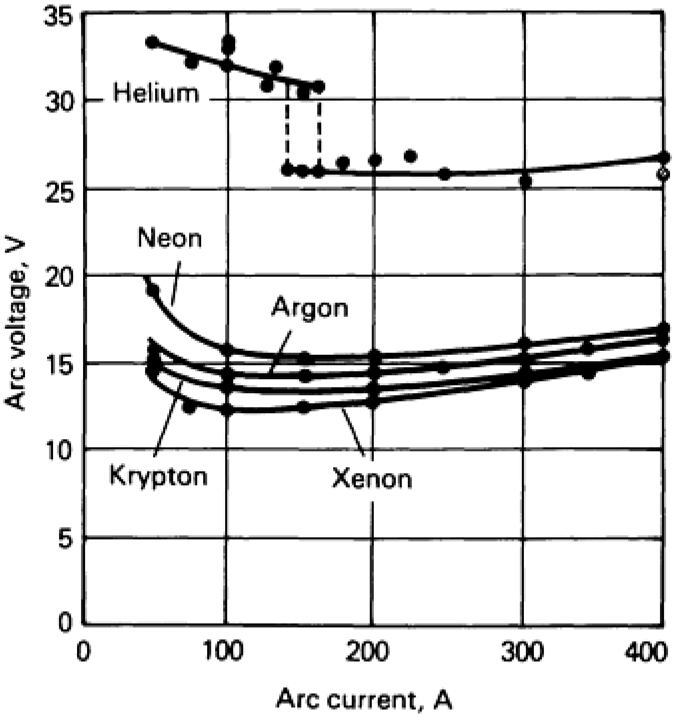

FIG. 8 PLOT OF ARC VOLTAGE VERSUS ARC CURRENT FOR SELECTED INERT SHIELDING GASES. WELDING PARAMETERS: ANODE, TITANIUM; CATHODE, TUNGSTEN; POLARITY, DCEN; ARC LENGTH, 12.7 MM (0.050 IN.)

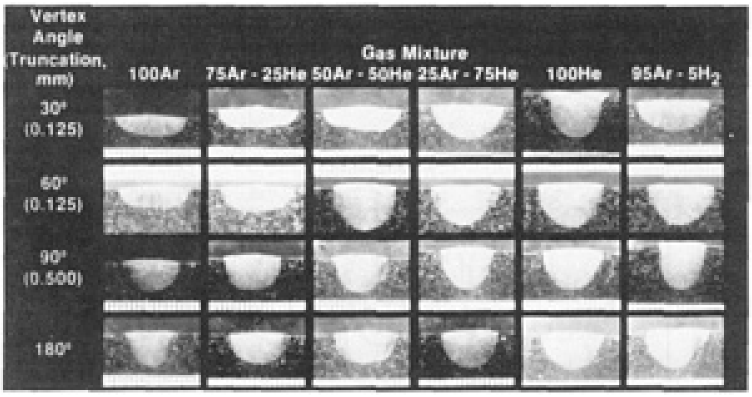

FIG. 9 EFFECT OF ELECTRODE TIP GEOMETRY AND SHIELDING GAS COMPOSITION ON WELD POOL SHAPE FOR SPOT-ON-PLATE WELDS. WELDING PARAMETERS: CURRENT, 150 A; DURATION, 2 S

To understand these phenomena, arc temperature distributions for a variety of shielding gases and mixtures, electrode shapes, current, arc voltages (electrode gaps), and anode materials have been measured in order to clarify welding arc fundamentals. Welding arcs are composed of electrons, positive gas ions, and neutral gas atoms. Some measurement techniques give the temperature of one species (electrons), whereas other techniques give the temperature of another species (neutral atoms). If the arc is in local thermodynamic equilibrium (LTE), all techniques should give the same temperature.

Although the assumption of LTE used to be considered completely valid, contemporary investigations have suggested that this is not always the case and that some of the older measurements may be somewhat in error. Absolute values of arc temperature are only needed to establish boundary conditions for detailed arc modeling of temperature-dependent properties. What is of more importance to the welding engineer or technologist is the relative effect of essential variables on heat input to the workpiece. Arc temperature measurement is one useful indication of how these variables affect the arc.

When compared to Fig. 7(a), Fig. 10 shows that large additions of helium to argon decrease peak temperatures slightly and increase the plasma diameter in the plasma column. The arc appears to become a broader and more isothermal heat source. The lower peak temperature is reasonable, because a combination of the high ionization potential of helium and relatively low currents of welding arcs gives an arc column that is only slightly ionized. Figure 7 showed that a large vertex angle had a similar, but less pronounced, effect on arc temperature when adding helium to the shielding gas (that is, the axial peak temperature decreased and the plasma diameter increased).

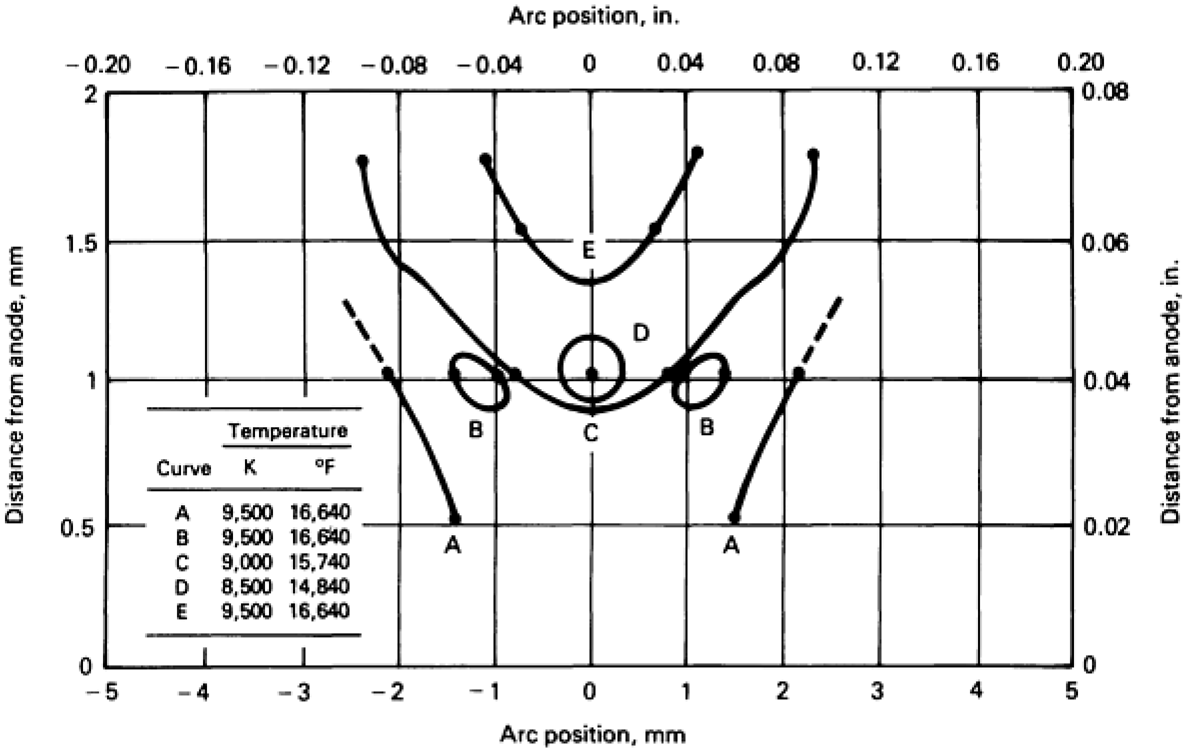

FIG. 10 PLOT OF GTAW ARC COLUMN TEMPERATURE DISTRIBUTION AS A FUNCTION OF ANODE DISTANCE AND ARC POSITION. WELDING PARAMETERS: ELECTRODE VERTEX ANGLE, 30°; CURRENT, 150 A; SHIELDING GAS, 10AR-90HE

Figure 11 shows that doubling the current from 150 to 300 A produces an increase in plasma diameter (that is, that portion of the arc above approximately 8000 K, or 13,900 °F). This effect occurs regardless of the shielding gas composition.

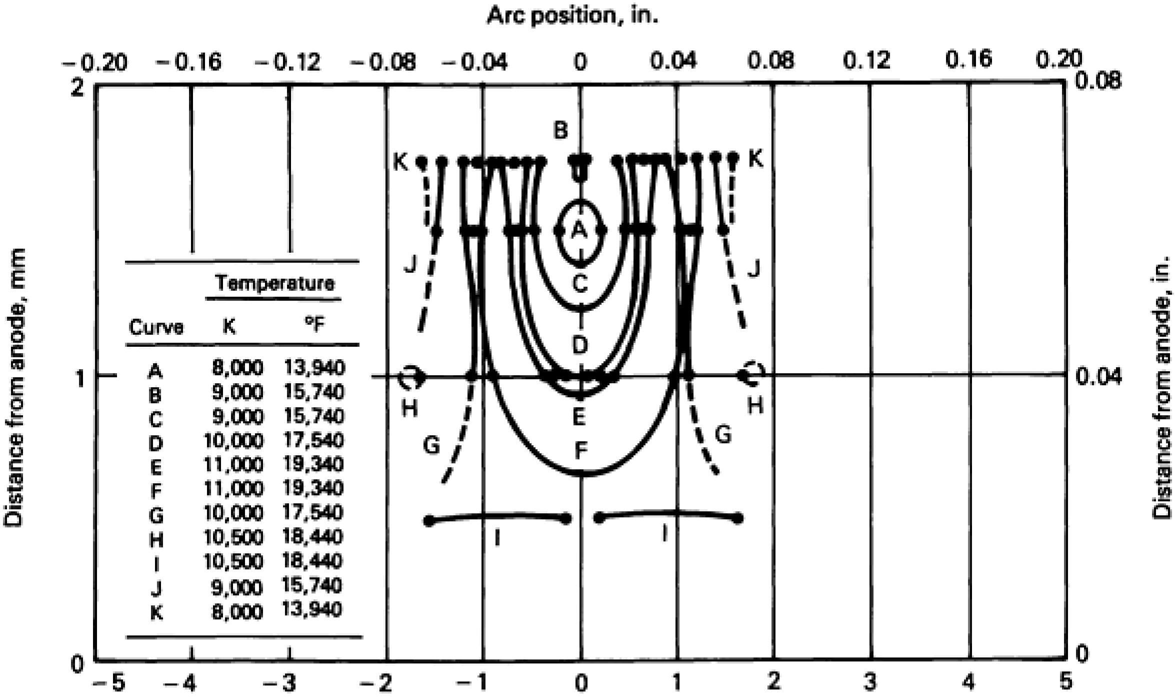

FIG. 11 PLOT OF GTAW ARC COLUMN TEMPERATURE DISTRIBUTION RELATIVE TO ANODE DISTANCE AND ARC POSITION. WELDING PARAMETERS: ELECTRODE VERTEX ANGLE, 30°; CURRENT, 300 A; SHIELDING GAS, 100% AR

The arc length or gap between the electrodes is yet another process parameter that must be considered, especially for mechanized welding, where it can be kept reasonably constant. Because arc length is proportional to arc voltage, longer arcs have higher arc voltages and consume more energy for the same current. However, this increased energy is generally lost through radiation to the environment surrounding the weld and does not effectively supply additional heat to the workpieces. Mechanized welding generally utilizes rather short arc lengths (2 to 3 mm, or 0.08 to 0.12 in.), whereas manual welding uses a longer arc length. Unfortunately, relationships that establish a direct correlation between the temperature distribution in the arc column and the weld pool shape still have not been established, primarily because weld pool shape depends on other factors, such as compositionally dependent molten metal properties.

Flow Structure. Shielding gas is used to displace reactive gases in the atmosphere from the vicinity of the weld. Inert gases are preferred for the GTAW process, because they minimize unwanted gas-metal reactions with the workpieces. A uniform laminar flow of gas from the gas cup would be ideal and, in fact, is usually achieved as long as there is no welding arc. However, the arc discharge rapidly heats the gas in the arc column, and thermal expansion causes plasma jets. The lower temperature near the cathode tip in Fig. 11 is an indication of a jet pumping in cooler gas. This becomes an important factor at high currents, because these jets can depress the surface of the weld pool and alter heat transfer to it. The rapid gas expansion can cause the flow to deviate from laminar and, in extreme cases, the flow can become turbulent. Turbulence tends to mix atmospheric contaminants into the arc, often where they can do the most harm: at the surface of the molten weld pool. Holographic interferometry and Schlieren photography have been used to characterize gas flow.

Figures 12, 13, and 14 show examples of laminar and turbulent flow. The flow from a commercial-design gas cup for three current levels is shown in Fig. 12. Laminar flow occurs where the fringes are generally straight and parallel. Turbulent flow is indicated by very wavy or circular fringes. Increasing current tends to make the laminar region somewhat broader and thicker, effectively increasing the area shielded from atmospheric contamination.

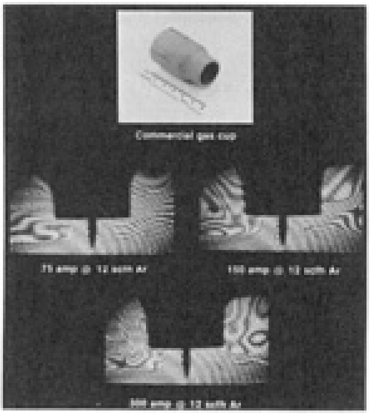
Commercial gas cup
75 amp @ 12 scfh Ar
150 amp @ 12 scfh Ar
300 amp @ 12 scfh Ar

FIG. 12 EFFECT OF GEOMETRY ON COMMERCIAL GAS CUP LAMINAR AND TURBULENT FLOW AS DETECTED BY REAL-TIME HOLOGRAPHIC INTERFEROMETRY

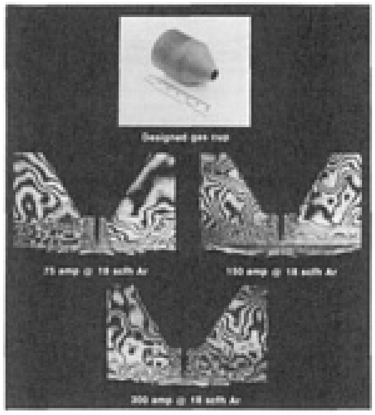
Designed gas cup
75 amp @ 18 scfh Ar
150 amp @ 18 scfh Ar
300 amp @ 18 scfh Ar

FIG. 13 EFFECT OF GEOMETRY ON CONVERGING CONE CUP LAMINAR AND TURBULENT FLOW AS DETECTED BY REAL-TIME HOLOGRAPHIC INTERFEROMETRY

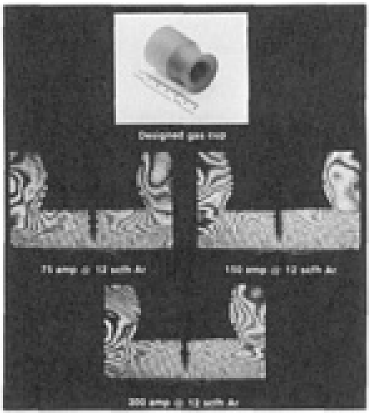
Designed gas cup
75 amp @ 12 scfh Ar
150 amp @ 12 scfh Ar
300 amp @ 12 scfh Ar

FIG. 14 EFFECT OF GEOMETRY ON VENTURI GAS CUP LAMINAR AND TURBULENT FLOW AS DETECTED BY REAL-TIME HOLOGRAPHIC INTERFEROMETRY

Figures 13 and 14 result from experiments with gas cup shapes that were designed to improve shielding. Figure 13 shows that a converging cone would be a very poor choice for the GTAW process, as indicated by the very small area of laminar flow and the extreme turbulence in the surrounding region.

Figure 14 is a venturi shape, which provides a large laminar flow region for all current levels and excellent shielding. This design may be somewhat better than commercial designs, but weld contamination studies should be conducted to verify this possibility.

Arc Physics of Gas-Tungsten Arc Welding

J.F. Key, EG&G Idaho, Inc.

Selected References

- S.S. GLICKSTEIN, "ARC-WELD POOL INTERACTIONS," DEPARTMENT OF ENERGY RESEARCH AND DEVELOPMENT REPORT WAPD-TM-1429, BETTIS ATOMIC POWER LABORATORY, AUG 1978
- C.E. JACKSON, THE SCIENCE OF ARC WELDING,*WELD. J.,* VOL 39, 1960, P 129S-140S, 177S-190S, 225S-230S
- J.F. KEY, H.B. SMARTT, J.W. CHAN, AND M.E. MCILWAIN, PROCESS PARAMETER EFFECTS ON ARC PHYSICS AND HEAT FLOW IN GTAW, *PROC. INT. CONF. ON WELDING TECHNOLOGY FOR ENERGY APPLICATIONS* (GATLINBURG, TN), 16-19 MAY, 1982, P 179-199
- J.F. KEY, J.W. CHAN, AND M.E. MCILWAIN, PROCESS VARIABLE INFLUENCE ON ARC TEMPERATURE DISTRIBUTION, *WELD. J./WELD. RES. SUPP.*, VOL 62 (NO. 7), P 179-S TO 184-S
- J.F. LANCASTER, ED., *THE PHYSICS OF WELDING,* 2ND ED., INTERNATIONAL INSTITUTE OF WELDING, PERGAMON PRESS, 1986
- H.B. SMARTT, J.A. STEWART, AND C.J. EINERSON, "HEAT TRANSFER IN GAS TUNGSTEN ARC WELDING," ASM METALS/MATERIALS TECHNOLOGY SERIES, ASM INTERNATIONAL WELDING CONGRESS (TORONTO), 14-17 OCT 1985

Power Sources for Welding

F. James Grist, Miller Electric Manufacturing Company, Inc.; William Farrell and Glen S. Lawrence, Ferranti Sciaky

---

Introduction

POWER SOURCES are apparatuses that are used to supply current and voltages that are suitable for particular welding processes. This article describes power sources for arc welding, resistance welding, and electron-beam welding.

Arc Welding Power Sources

Arc welding requires that an electric arc be established between an electrode and the workpiece to produce the heat needed for melting the base plate. Because utility energy is not delivered at the proper voltage and current, it must be converted to the required levels by the welding power source.

Arc power sources convert the customary 240 or 480 V alternating current (ac) utility power to a range from 20 to 80 V and simultaneously increase the current proportionately. Galvanic isolation is also provided from the utility line voltage to ensure operator safety. Motor- or engine-driven welding generators are wound to deliver the correct voltage and current directly; therefore, no transformer is necessary. Actual arc voltage is established by factors other than the power source (for example, electrode type, base-metal type, shielding gas, and standoff distance).

Ratings and Standards

The National Electrical Manufacturers Association (NEMA) categorizes arc welding power sources into three classes on the basis of duty cycle:

- CLASS I: RATED OUTPUT AT 60, 80, OR 100% DUTY CYCLE
- CLASS II: RATED OUTPUT AT 30, 40, OR 50% DUTY CYCLE
- CLASS III: RATED OUTPUT AT 20% DUTY CYCLE

Duty cycle is the ratio of arc time to total time based on a 10 min averaging period. A 60% machine will deliver 6 min of arc time and 4 min off time without overheating. In Fig. 1, curve A shows a NEMA class I (60%) 300 A rated machine that is capable of a maximum 375 A at reduced duty cycle (38%) and 232 A at 100% (continuous). Curve B represents a NEMA class II (50%) 250 A machine with a continuous duty of 176 A. Curve C represents an engine-driven machine rated at 225 A and 20% duty. It does not offer output in excess of its rating, because of a horsepower limitation of the engine.

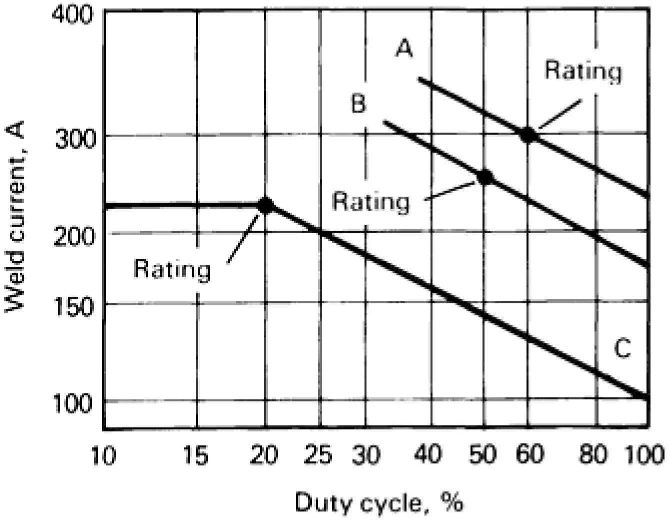

FIG. 1 SELECTED DUTY CYCLE CURVES. A, 300 A, 60% MACHINE; B, 250 A, 50% MACHINE; C, 225 A, 20% ENGINE-DRIVEN MACHINE WITH HORSEPOWER LIMITATION

NEMA requires most machines to produce a maximum of 125% of rated output current. No minimum current is specified, but 10% (of rating) is typical. Load voltage, E, for class I and II machines is defined by:

$$E = 20 + 0.04I \qquad \textbf{(EQ 1)}$$

where I is the current in amperes, below 600 A. Above 600 A, the rated load voltage remains at 44 V. Power sources intended for gas-tungsten arc welding (GTAW) service are defined by:

$$E_{\mathrm{GTAW}} = 13 + 0.012I \qquad \textbf{(EQ 2)}$$

Class III machines are constant-current types limited to 20% duty cycle.

Efficiencies (that is, ratios of power output to power input) of 68% for line-frequency machines and 88% for inverters are typical. Line current draw can be reduced in certain machines by the addition of power factor correcting capacitors connected to the primary winding.

Input (line) voltages shown on the manufacturer nameplate are often multivoltage by means of tap changing for 200, 230, 460, or 575 V. Designed primarily for 60 Hz power, most systems will operate at 50 Hz, with minor changes in output. Input voltages for 50 Hz systems are typically 220, 380, and 415 V. Usual service conditions are temperature that range from 0 to 40 °C (32 to 105 °F) and altitudes of up to 1000 m (3300 ft). (Altitude is a factor because of fuel-air mixtures.)

The power source nameplate must include the following specifications:

- MANUFACTURER
- NEMA CLASS (I, II, OR III) AND DUTY CYCLE
- MAXIMUM OPEN-CIRCUIT VOLTAGE
- NUMBER OF PHASES
- MAXIMUM SPEED (IN REVOLUTIONS PER MINUTE) AT NO LOAD (ROTATING EQUIPMENT)
- RATED LOAD (IN VOLTS)
- RATED LOAD (IN AMPERES)
- FREQUENCY (OR FREQUENCIES) IN HERTZ
- INPUT VOLTAGE(S)

The National Electric Code (NEC) requires that a main disconnect switch be provided for each installed welding machine. Manufacturer installation instructions list the kVA and kW rating and recommend the correct fuse and conductor size. Local codes may differ from NEC requirements.

Proper grounding is a necessary safety consideration. Machines equipped with high-frequency arc stabilizers must be correctly grounded to prevent electromagnetic interference (EMI) to nearby equipment. The installer certifies (to the Federal Communications Commission) that the machine has been grounded according to manufacturer instructions. Most power sources are forced-air cooled by internal fans. Thermal sensors will interrupt output if airflow is impeded.

Power Source Selection

Because no single power source is right for all welding situations, it is necessary to know the processes to be used before selecting the best power source. Table 1 gives the more-common processes that use constant-current and constant-voltage power sources.

TABLE 1 SELECTION OF POWER SOURCE RELATIVE TO WELDING PROCESS TO BE EMPLOYED

WELDING PROCESS	**POWER SOURCE**	
	CONSTANT CURRENT	**CONSTANT VOLTAGE**
SHIELDED METAL ARC WELDING (SMAW)	•	. . .
GAS-TUNGSTEN ARC WELDING (GTAW)	•	. . .
GAS-METAL ARC WELDING (GMAW)	. . .	•
PULSED GAS-METAL ARC WELDING (GMAW-P)	•	. . .
FLUX-CORED ARC WELDING (FCAW)	. . .	•
SUBMERGED ARC WELDING (SAW)	. . .	•
ELECTROGAS WELDING (EGW)	•	•
ELECTROSLAG WELDING (ESW)	. . .	•
PLASMA ARC WELDING (PAW)	•	. . .

Alternating current power sources for shielded metal arc welding (SMAW) and submerged arc welding (SAW) can be as simple as a single transformer (Fig. 2a). Low-voltage machines with open-circuit voltages of 50 V or less are transformer types. Output alternating current reflects the sinusoidal waveform of the utility input line (Fig. 3a).

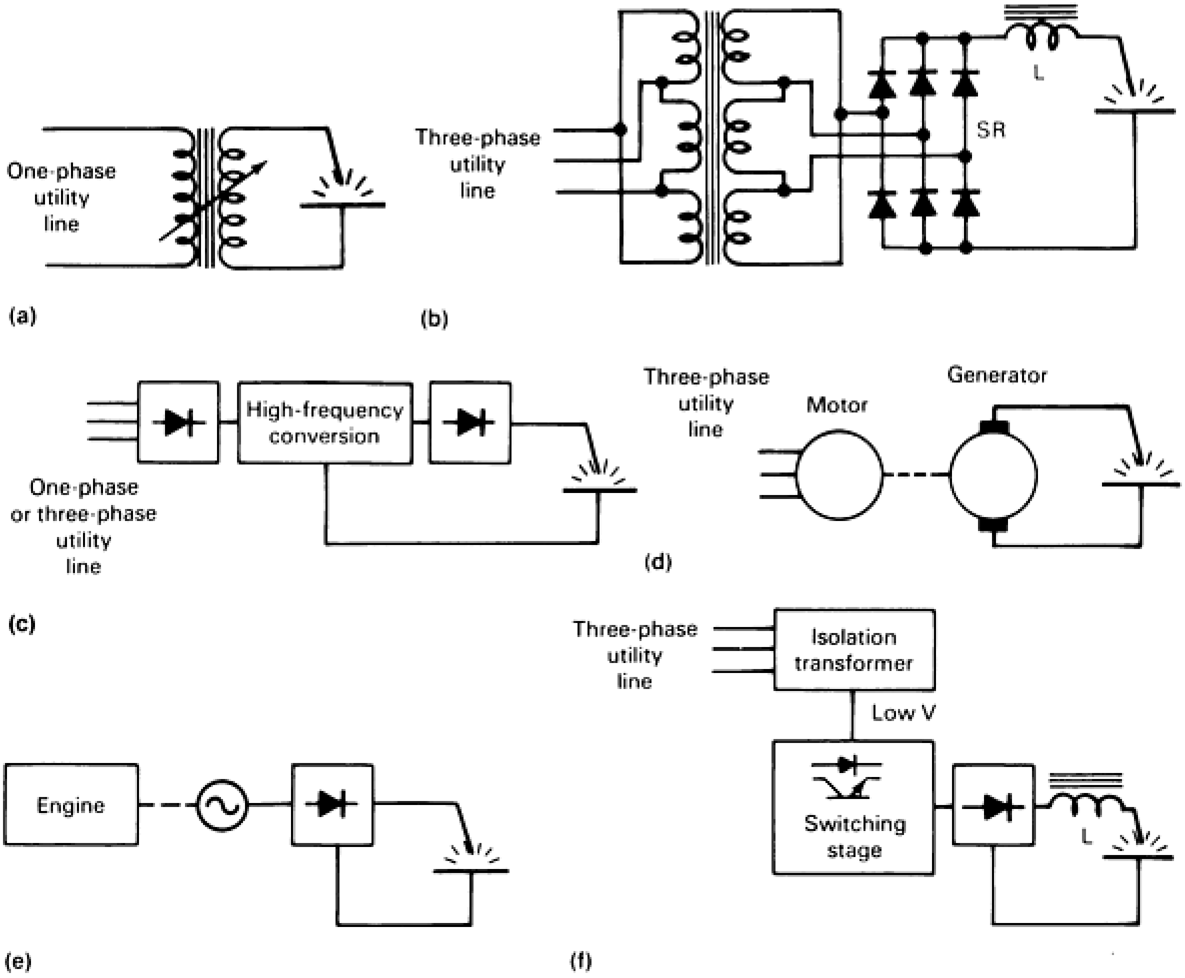
One-phase utility line
(a)
Three-phase utility line
L
SR
(b)
High-frequency conversion
One-phase or three-phase utility line
(c)
Three-phase utility line
Motor
Generator
(d)
Engine
(e)
Three-phase utility line
Isolation transformer
Low V
Switching stage
L
(f)

FIG. 2 SELECTED ALTERNATING CURRENT POWER SOURCES. (A) ALTERNATING CURRENT WELDING TRANSFORMER, ADJUSTABLE CORE OR WINDINGS. (B) THREE-PHASE SCR-CONTROLLED DC WELDING POWER SOURCE. (C) INVERTER BLOCK DIAGRAM. (D) MOTOR-GENERATOR SET. (E) ENGINE-DRIVEN ALTERNATOR WITH DC OUTPUT. (F) SECONDARY SWITCHING POWER SOURCE

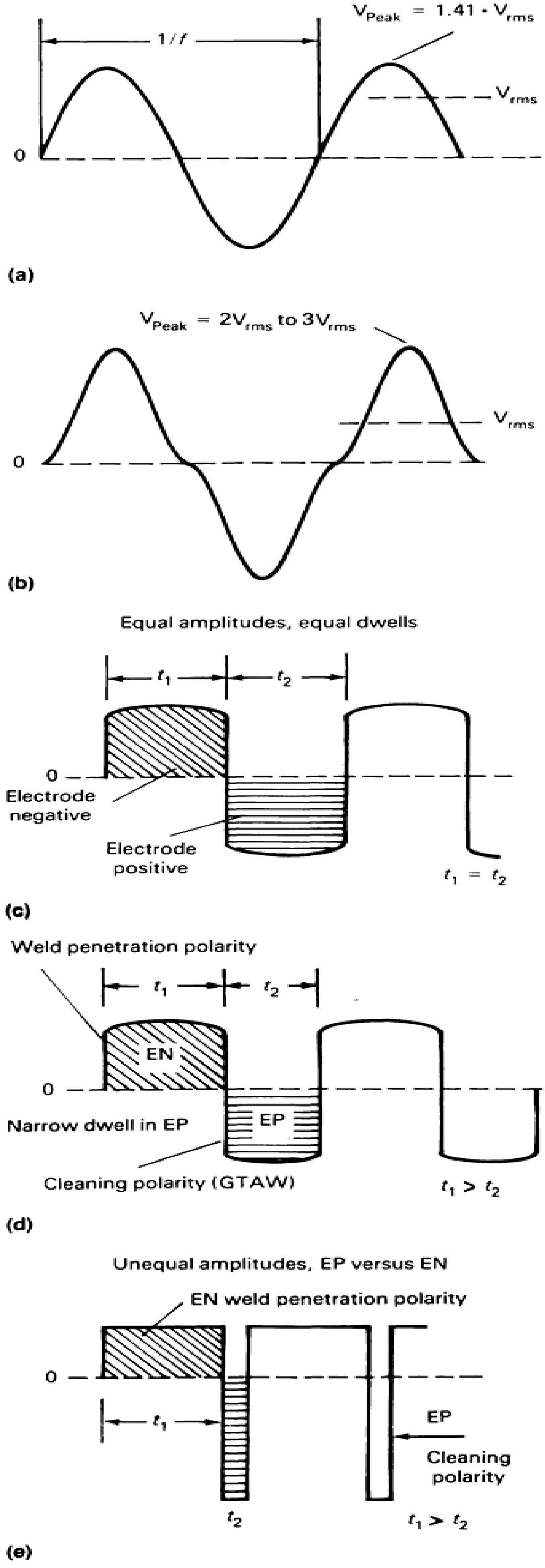

$V_{Peak} = 1.41 \cdot V_{rms}$
1/f
V_{rms}
0
(a)
$V_{Peak} = 2V_{rms}$ to $3V_{rms}$
V_{rms}
0
(b)
Equal amplitudes, equal dwells
t_1
t_2
0
Electrode negative
Electrode positive
$t_1 = t_2$
(c)
Weld penetration polarity
t_1
t_2
EN
0
EP
Narrow dwell in EP
Cleaning polarity (GTAW)
$t_1 > t_2$
(d)
Unequal amplitudes, EP versus EN
EN weld penetration polarity
0
t_1
EP
Cleaning polarity
t_2
$t_1 > t_2$
(e)

FIG. 3 TYPICAL AC WELDING CURRENT WAVEFORMS. (A) SINE WAVE ALTERNATING CURRENT, SIMPLE TRANSFORMER WELDER. (B) MODIFIED SINE WAVE OBTAINED FROM MAGNETIC AMPLIFIER-TYPE WELDER. (C) SQUARE WAVE, LINE FREQUENCY, AND BALANCED DWELL. (D) SQUARE WAVE WITH UNBALANCED DWELLS, EP VERSUS EN. (E) VARIABLE-POLARITY AC WITH DWELL, AMPLITUDE, AND FREQUENCY ALL ADJUSTABLE

Magnetically controlled equipment (that is, a saturable reactor or magnetic amplifier) used for GTAW delivers a modified sinewave output (Fig. 3b). There is a risk of electrode (tungsten) migration to the workpiece if peak current greatly exceeds the average value. This technology exhibits moderate response.

Newer technologies deliver a square wave output at line frequency (Fig. 3c). A square wave eliminates peaking and provides a rapid transition through zero, which is important to cyclic reignition of the arc. Thyristors are employed in concert with magnetic cores to generate the square current waveform. Adjustable imbalance permits the operator to control the ratio of electrode positive (EP) to electrode negative (EN) current by dwell extension (Fig. 3d).

Direct current power sources can be classified as being constant-current sources or constant-voltage sources.

Constant-Current Sources. For the SMAW process, a family of "drooping" volt-ampere (V-A) curves (A, B, and C) is provided in Fig. 4(a). Often, a current boost or "dig" characteristic is made available (curve D in Fig. 4a). A burst of extra current during the brief period of metal transfer shorting tends to alleviate sticking during keyhole welding or when weaving or whipping the electrode, especially with the more-globular transfer electrodes.

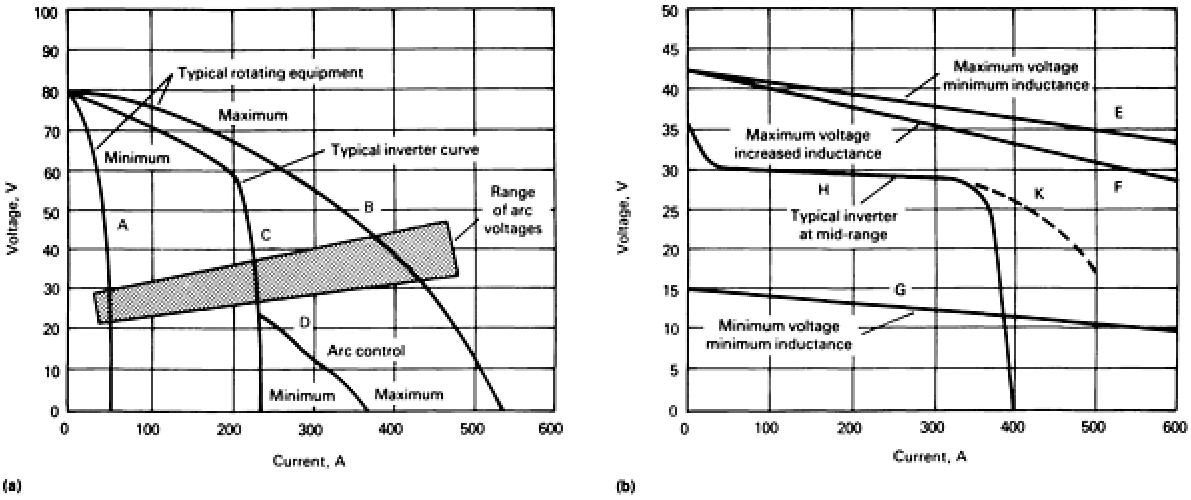

FIG. 4 VOLT-AMPERE CURVES OF TYPICAL DC POWER SOURCES. (A) CONSTANT-CURRENT SOURCE. (B) CONSTANT-VOLTAGE SOURCE

Power sources for dc GTAW processes are generally more critically constant-current than are the SMAW "stick droopers." Ripple content is lower and current regulation is tighter. Inverter types customarily regulate to ±1% or ±1 A, whichever is larger. Remote foot control is essential to ensure that both hands are free for manipulation. Response time need not be fast, except for pulsed programs. Programmable and pulse-capable units permit the automation of complex current profiles.

Power sources for GTAW processes include a high-frequency arc starting device that impresses a high radio frequency (RF) voltage on the electrode. This energy "jumps the gap" from the electrode to the workpiece, ionizing the shielding gas, and permits establishment of an arc. Thus, the electrode need not touch the workpiece.

Constant-Voltage Sources. Power sources intended for gas-metal arc welding (GMAW) exhibit a relatively flat V-A curve, as shown in Fig. 4(b) (curves E, F, and G). The self-correcting characteristic of GMAW regulates the electrode burn-off rate. As the wire feed rate is increased, additional current will be drawn, but terminal voltage will remain essentially constant. Voltage feedback techniques are used in solid-state designs to hold long-term output close to a set value.

In the short-circuiting GMAW process, for example, drops of metal are transferred to the base metal through a pinch-off routine during a period of short circuiting of the electrode to the workpiece. Shorting time and arc time are determined by power source dynamics.

Source Characteristics

Whether utility fed or motor driven, all welding power sources have a characteristic that indicates their dynamic and steady-state response. In early machines, source characteristics were optimized for a single process and were relatively unchangeable. More recent designs can be programmed for a variety of characteristics that include multiple weld processes.

The welding arc process is a dynamic, fast-changing load, exhibiting short circuits (if shorting, then no arc) at high currents and momentary voltages from near zero to over 50 V. Changes occur in milliseconds, particularly when metal transfer takes place. The ability of the power source to respond properly to rapidly changing conditions at the arc is critical to good performance. Modern technology invariably includes a closed-loop feedback circuit in the control system.

Open-Circuit Voltage (OCV). When no load is connected to the output terminals of a welding power source, the voltage that appears at the terminals is at its maximum. A high OCV value generally aids in arc starting and stability. In transformer-type power sources. OCV is established by the incoming utility line voltage and the transformer primary-to-secondary turns ratio. The output winding turns, the flux, and the angular velocity (rev/min) will establish the OCV in generators.

In manual welding, the maximum OCV is 80 V_{rms} for machines with more than 10% ripple, or 100 V OCV for a ripple content of less than 10%. Power sources intended for automatic welding only are permitted 100 V OCV. The percentage ripple voltage is the ratio of rms ripple to average pulsating dc, expressed as:

$$\%ripple = \frac{V_{\max} - V_{\min}}{V_{rms}} \qquad \textbf{(EQ 3)}$$

Static V-A Curves. Two terms are commonly used to define the characteristics of a power source for welding: static output and dynamic response. Static characteristics are set forth by the manufacturer as a series of V-A curves (Fig. 4). The slope of the static V-A curve is the ratio of load voltage change to the change in load current, expressed in V/100 A. It indicates the achievable envelope of output limits, including the highest voltage the machine will support at a given current, and the highest current that can be drawn at a given voltage.

Dynamic Response. The static V-A curves discussed above do not take into account the dynamic interplay between the power source and the process. Instantaneous response of the power source to step changes in load voltage is important to welding performance. This is especially true in the case of certain GMAW processes, where the response to a metal transfer short circuit is instrumental in bead formation and the reduction of spatter generation.

Methods of Control or Electrical Conversation. Power source control methods are a key factor in the rate of current rise (dI/dt). Either increased reactance in the ac circuitry or inductance in the dc portion will slow the rate of rise (and fall) of current during changes at the arc. Some equipment incorporates an adjustment to select the proper number of turns for a given welding condition (with silicon-controlled rectifiers [SCR] and transistor controls, these characteristics can be synthesized). Because there is no industry-accepted method to quantify the dynamics, the user must obtain a power source with a satisfactory range of adjustment.

Another consideration is the frequency response of the power source, that is, its ability to "follow" a fast command signal, such as a pulse program. Magnetically based machines are generally not fast enough to faithfully reproduce a pulsed GMAW control signal and deliver it at high currents to the weld. When a better response is required, inverter-based SCR, transistorized sources with a high-frequency response are preferred.

Droopers. A constant-current, a drooper, power source displays V-A curves such as those in Fig.(a) (curves A and C). A considerable negative slope is evident. The internal impedance of the power source is said to be high. Large variations in the arc (load) voltage (for example, below 40 V) do not appreciably affect the selected current. Droopers are noted for their use in the SMAW process, where the operator sets current to establish a suitable burn-off rate for the selected electrode. Manipulation, that is, shortening or lengthening the arc, does not greatly affect the arc current.

However, if welding is conducted on a flatter V-A curve (see curve B in Fig. 3a), then electrode manipulation may change the current considerably. Thus, the operator gains more local control of the current through eye-hand coordination. Increasing the arc length by "pulling back" on the electrode causes the operating point to shift up slightly in terms of voltage, but decrease considerably in terms of current.

Open-circuit voltage can be field adjusted in generator types, which provides the operator with a great degree of control through electrode manipulation. Skilled operators use this flexibility to their advantage when performing difficult out-of-position welds.

Constant-Voltage Sources. Equipment designed for constant-voltage processes will display a "stiff," nearly horizontal V-A curve (curve E, F, and G in Fig. 4b). The open-circuit voltage is 43 V for curve F. When serving a 300 A load, it falls to 35 V. This is expressed as:

$$\frac{(43-35)}{300A} = \frac{8V}{300A} = \frac{2.6V}{100A} \qquad \textbf{(EQ 4)}$$

Curve H of Fig. 4(b) is representative of an inverter source in the constant-voltage mode. When set to 30 V, the electronic regulation system keeps the output very constant out to 300 A under a static (resistive) load, beyond which it falls abruptly to 0 V at 400 A. Curve K in Fig. 4(b) represents an extended output that may be available on a short-term (dynamic) basis from inverters.

The term "constant" is taken to mean relatively constant. Some droop in voltage is permitted, and is even preferred in certain welding situations. Load current at a given voltage is responsive to the rate at which a consumable electrode is fed into the arc. The designer ensures that terminal voltage is maintained during heavy loads by providing tight coupling between the primary and secondary windings of the transformer, or by providing electronic regulation systems, or both.

Movable Magnetic Cores. Welding power must be delivered to the arc under complete control of the operator. In the simplest form, conversion to the lower voltage and selection of current are incorporated into a single, special welding transformer. Movable coils (varying the coupling between primary and secondary windings) and moving iron shunts provide stepless adjustment of the output (Fig. 2a).

Saturable Reactors. In heavy industrial equipment, control is imparted by saturable reactors. These reactors comprise laminated iron core assemblies fitted with power and control windings. A small control current varies the effective magnetic permeability of the iron core and controls flux to the power winding. Considerable gain is obtained. Typically, 5 A of control will command several hundred amperes of welding current. Both ac and dc outputs are available. Direct current output is obtained by the addition of secondary rectifiers. By combining rectifier diodes with self-saturating magnetic cores, a wider range is obtained, referred to as magamp control.

Alternating current and ac/dc power sources are manufactured in sizes that range from 180 A Class III "farm welders" to units designed for industrial use and rated in hundreds of amps. Principal attributes are low cost and simplicity. Alternating current is specified wherever arc blow is a concern. It is also preferred for welding aluminum. Direct current machines are favored for most SMAW and GTAW processes, and for essentially all GMAW processes.

Silicon-Controlled Rectifiers. In addition to the transformer taps, moving coils, moving shunts, and saturable reactors mentioned earlier, there are a number of other means by which output can be controlled. Thyristors, commonly known a SCRs, are frequently employed as control elements, replacing magnetic cores.

The SCR is similar to a rectifying diode, but is equipped with an additional gating element. With properly phased low-power gate signals, a configuration of SCRs will deliver hundreds of amperes of current.

The output of such a phase-controlled arrangement (Fig. 2b) can be tailored to perform as a stiff (that is, flat slope) constant-voltage source or as a drooping constant-current source through closed-loop feedback controls. When designed for three-phase operation, the dc output ripple frequency is either three times the line frequency (using three SCRs) or six times, using six SCRs. Output ripple is "smoothed" with an appropriate inductor. Single-phase versions display the ripple of twice the line frequency and require a larger output inductor, *L*.

Because welding current is analog in nature, it has traditionally been adjusted with analog circuitry via rheostats and potentiometers. Recent advances allow digital control with microprocessors and dedicated software; tight-tolerance repeatability is the driving force.

Inverters. Inverter welding power sources are smaller and more efficient, and they offer better performance than their line-frequency counterparts. The size and weight of iron-core components are reduced when operating at higher frequencies (Fig. 5).

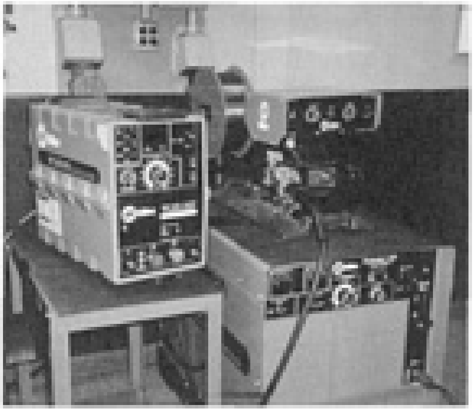

FIG. 5 300 A RATED CONVENTIONAL LINE-FREQUENCY POWER SOURCE (RIGHT) AND INVERTER POWER SOURCE WITH SAME RATING (LEFT). COURTESY OF MILLER ELECTRIC MANUFACTURING COMPANY, INC.

With reference to the inverted shown in Fig. 2(c), the incoming utility line directly feeds a rectifier for conversion to direct current, which powers an inverter stage operating in the range from 2 to 100 kHz. Various topologies include forward, buck, series capacitor, and series resonant inverters.

Control methods employ either pulse width modulation (PWM) or frequency modulation (FM). A variety of semiconductors are used as switching elements, including transistors (for example, field-effect transistors [FETs] and insulated-gate bipolar transistors [IGBTs]).

The transformer, which is often designed of ferrite core material, operates at high frequencies without introducing high losses. The turns ratio determines welding voltage, usually 80 V_{rms} ac at the secondary winding. A second rectifier converts it back to dc for welding. A small output inductor is required to smooth the high-frequency ripple.

Inverters have the greatest application where portability or advanced performance is required. Fast response permits current with fast transitions, low overshoot, and faithful reproduction of pulse programs.

Variable Polarity. High-technology power sources permit adjustment of the ac waveform to vary the ratio of EP to EN current and terms of both width and amplitude (Fig. 3e). Among these are switched dc types, wherein two dc constant-current power sources are alternately switched on and off, and arc current is of one polarity and then the other polarity. Almost unlimited frequencies can be generated, with rapid zero crossings.

Cycloconverter power sources constitute a family that uses three-phase inputs to generate low-frequency ac for the GTAW process. Through careful selection of specific portions of each line-frequency cycle, a pseudo square wave is assembled at a frequency lower than line frequency.

Rotating Equipment. For a variety of reasons, rotating generators and alternators are often used to supply welding power. In what is known as a motor-generator (M-G) set, an electric motor (Fig. 2d) drives the shaft of a generator, which also isolates weld power from the input utility power. The flywheel effect of the M-G set ensures stable output.

Where no electric utility power is readily available, engine-driven equipment can supply welding power (Fig. 2e). Gasoline and diesel fuels predominate. Both air- and water-cooled prime movers are available. Some include an auxiliary air compressor for air tool use.

Generators produce direct current. A wire-wound armature spins in a frame arranged with electromagnetic field poles. The armature windings cut through these fields cyclically, generating ac. This is collected by commutator, that is, copper segments fitted to the armature shaft and equipped with fixed carbon brushes. The commutator serves as a "mechanical rectifier," converting the armature currents to dc. Both the OCV and current can be user adjusted, providing good control of the V-A slopes. The output is relatively pure dc, with very little ripple content. Engine speed is governor controlled.

Most generators provide constant-current output for SMAW and GTAW processes, but adapters are often available for constant-voltage operation in the GMAW process. Auxiliary power for lights, tools, and so on derived from auxiliary stator windings.

Alternators (that is, ac generators) produce alternating current, as the name implies (Fig. 2e). A spinning rotor winding is charged with direct current obtained from the exciter and applied through slip rings, creating a rotating magnetic field. No commutator is required. The rotating field cuts through a surrounding stator assembly, the windings of which collectively produce alternating current, either single phase or three phase, as designed. In single-phase designs, the resulting current can be used for ac welding.

Output selection is provided in course ranges with reactor taps (Fig. 6), and fine adjustment is provided by rheostat-controlled current in the revolving field windings. Engine speeds are generally either 1800 or 3600 rev/min, corresponding with four-people and two-pole alternators, respectively. Some alternators eliminate the revolving field winding by employing rotating permanent magnets. If direct current is required, a static (silicon diode) rectifier is included. The combination is referred to as a dc generator-rectifier.

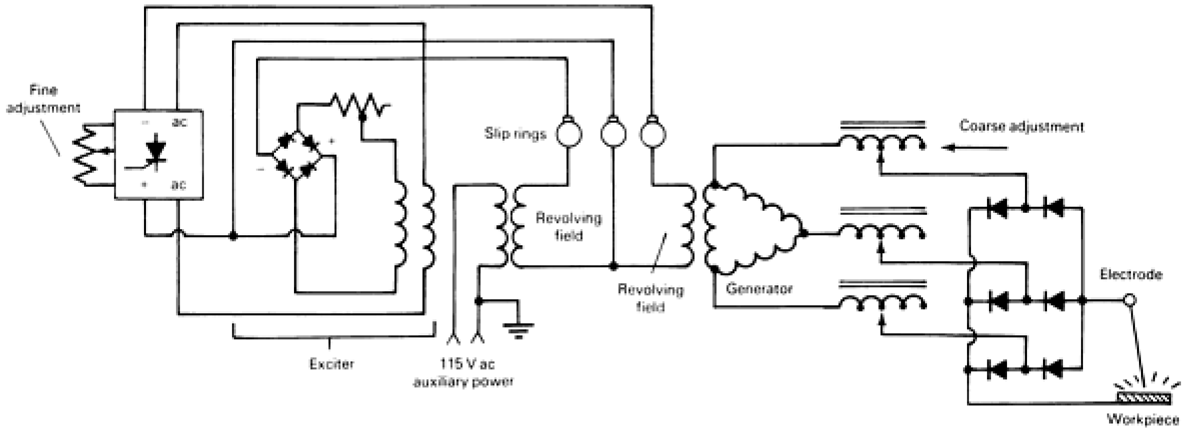

FIG. 6 SCHEMATIC SHOWING TYPICAL ENGINE-DRIVEN WELDING POWER SOURCE

The recent evolution to solid-state controls has raised the performance level of engine-driven machines. Accessories include remote controls, idle devices to reduce engine speed when not welding, battery charging capabilities, and high-frequency arc igniters.

Pulsed Power Supplies

Pulsed GTAW and GMAW processes require a power source with the ability to deliver precisely shaped current pulses superimposed on a lower background (arc-sustaining) level. In addition to thyristors in phase control, inverters are also used to generate pulses and are especially versatile for such applications. Add-on "pulse formers" have also been designed for attachment for conventional dc welding power sources.

Pulse Repetition Rates (PRR). For the GTAW process, PRRs (the number of times per second that a pulse is transmitted) are relatively low (typically, 1 to 10 PRR). Fast rise times are not essential. However, in the GMAW pulse process, repetition rates typically range from 100 to 200 PRR. Rise (and fall) times must be short (50 to 500 μs) and overshoot must be low to preserve pulse fidelity.

The mean current of a train of rectangular pulses is given by:

$$I_{AV} = FT_P(I_P - I_B) + I_B \qquad \textbf{(EQ 5)}$$

where I_{av} is the average current, I_B is the background current, I_P is pulse peak current, f is frequency, and T_P is pulse width.

From Eq. 5, with background current, pulse width, and peak current selected, the average current will be proportional to frequency. Increasing and decreasing pulse frequency in proportion to wire feed speed will balance the current with the burn-off rate of the electrode, maintaining the proper arc length for the GMAW process. Individual droplet size will remain essentially constant.

Some pulse power sources generate pulses at line frequency or multiples thereof and stabilize the process by adjusting the power envelope of each pulse. (In a refined development, designated "one-knob control," the operator need only select a wire and gas type; pulse parameters are established by algorithms in the controller. As wire feed speed is adjusted, all other functions "track.")

Synergic Pulsed GMAW. A step higher in technology results in synergic pulsed GMAW. Inverters (Fig. 2c) and secondary switchers (Fig. 2f) ensure process stability by controlling melting rate through modulation of pulse shape and frequency. The relationship between digitally monitored wire feed speed and pulsation parameters is managed by microprocessor look-up tables and similar control strategies. The power source is simply required to exhibit fast response and high peak current capability.

Multiple Operator Power Sources

On large job sites with a number of arcs, a multiple operator (MO) power source can serve several operators. Diversity of use suggests that all arcs will not be operating simultaneously and that a 0.25 factor is typical. Accordingly, a single, large, 76 V constant potential power source can supply up to 20 SMAW operators. Each operator is equipped with a resistor "grid" to adjust arc current to his particular needs.

An alternative MO concept provides individual magamp control modules (typically, eight) fed from a single, large, multiwinding transformer. Each operator can select his preferred polarity and current. Modules can be connected in parallel for greater output.

Power Sources for Welding

F. James Grist, Miller Electric Manufacturing Company, Inc.; William Farrell and Glen S. Lawrence, Ferranti Sciaky

---

Resistance Welding Power Sources

The function of the power source for resistance welding (RW) is to deliver a predetermined amplitude of current to the welding electrodes that are clamping the workpiece. Current flows from the power source to one electrode, through the workpiece to the opposing electrode, and returns to the power source. The electrodes concentrate current into a small, usually circular area. The flow of this constricted current generates heat, concentrated largely at the interface(s) of the workpiece details to be joined.

Heat Input. The rate of heating must be sufficiently intense to cause local electrode/workpiece interface melting. The opposed electrodes apply a pressure and, when sufficient melting has been achieved, the current is interrupted. Electrode force is maintained while the molten metal solidifies, producing a sound, strong weld.

If both the local resistance of the workpiece and the welding current magnitude were constant, then the total quantity of heat, Q, developed in the workpiece would be given (in joules) by:

$$Q = I^2 RT \qquad \textbf{(EQ 6)}$$

where I is the effective value of current in amperes, R is the resistance of the workpiece in ohms, and t is the duration of flow of current in seconds.

However, the resistance is not constant. Contact resistances decrease in magnitude and the bulk resistance of the work column increases as its temperature rises. Furthermore, resistance variations of the workpiece may cause variations in the current amplitude, depending on the nature of the power source. Under these circumstances, the total quantity of heat, Q, is given by:

$$Q = \sum_{0}^{t} i^2 r.dt \qquad \textbf{(EQ 7)}$$

where i is the instantaneous value of current expressed as a function of time and r is the instantaneous resistance that is changing with time.

Equipment Selection. The design of resistance welding machines resembles that of a "C" frame press (Fig. 7). The power source transformer secondary winding is connected to the electrodes on cantilevered arms, one of which provides for adjustable electrode force. Throat depth is the distance from the frame to the electrodes. Throat gap is the vertical distance between the two parallel secondary conductors. The total area of the secondary circuit (throat depth times throat gap) exerts a profound influence on output performance. Any increase in inductance, whether due to geometry or magnetic materials within the secondary, will degrade performance and lower the power factor.

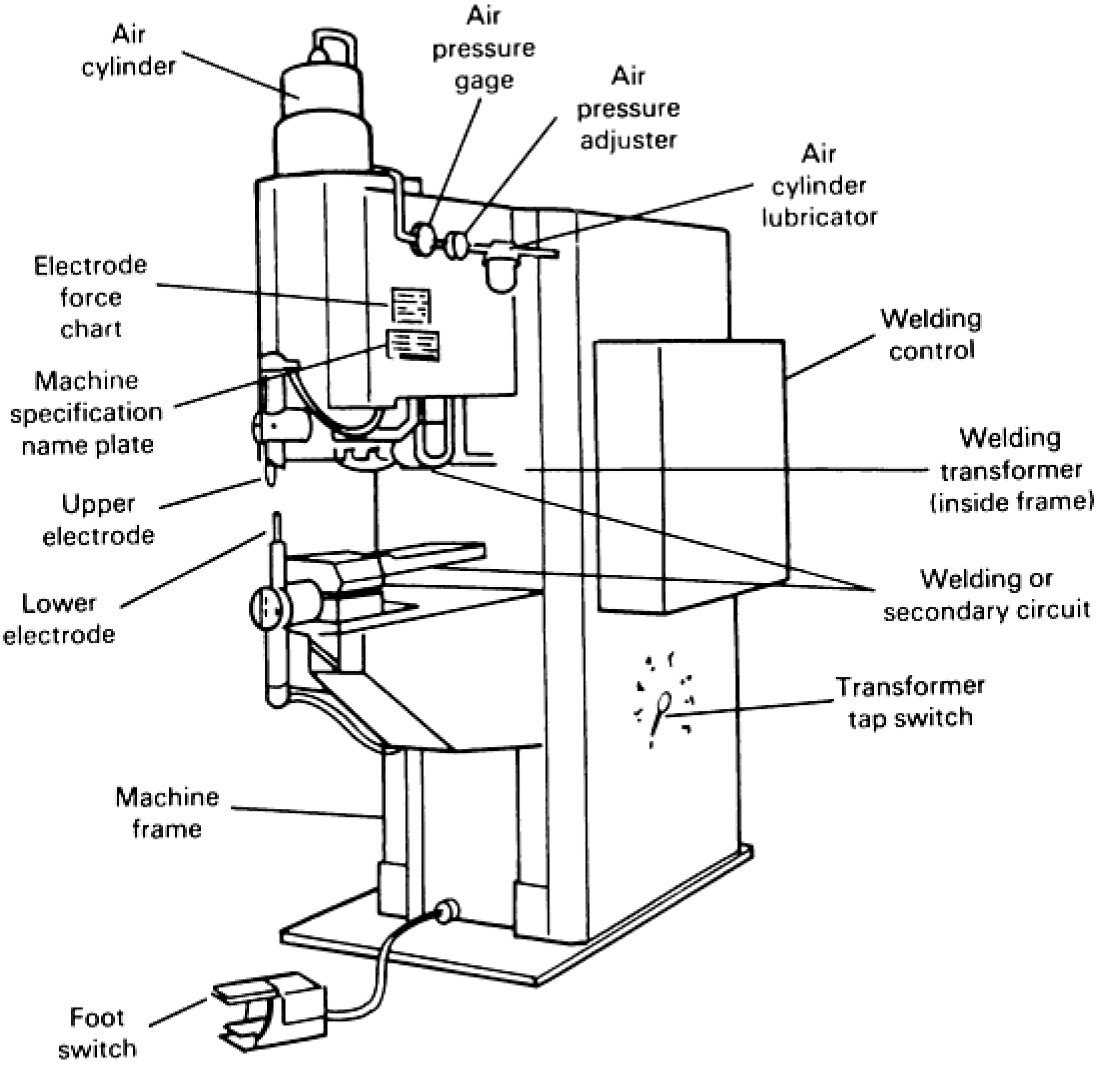

Air cylinder
Air pressure gage
Air pressure adjuster
Air cylinder lubricator
Electrode force chart
Welding control
Machine specification name plate
Welding transformer (inside frame)
Upper electrode
Lower electrode
Welding or secondary circuit
Transformer tap switch
Machine frame
Foot switch

FIG. 7 SCHEMATIC SHOWING PRIMARY COMPONENTS OF A RESISTANCE WELDING MACHINE

The input to resistance welding power sources is usually 230 or 460 V utility power delivering single-phase or three-phase alternating current at 60 Hz. All systems require transformers to step down the line voltage to a relatively low value, with a proportional increase in current. The simplest systems consist of a transformer connected to single-phase power. More advanced sources produce a dc welding current through rectification.

Rectification Systems. In primary rectification systems, a special three-phase transformer is designed with a very large iron core. Each dc pulse is limited in duration and each is reversed in polarity to prevent saturation of the core. In the majority of applications, a weld is produced by a single impulse of direct current. For heavy gages, pulsation welding is used. A series of dc impulses of alternating polarity is separated by "cool time." The result is a very low frequency alternating current. Hence, the term frequency converter is commonly applied to this type of power source. Figure 8 shows a three-phase half-wave primary rectified power source.

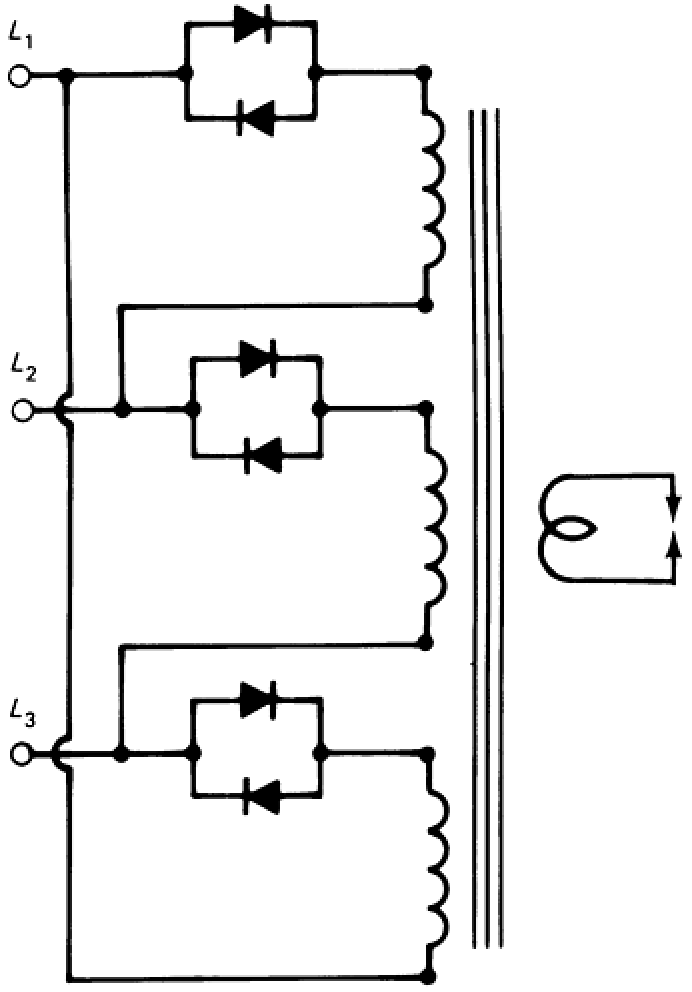

FIG. 8 CIRCUIT DIAGRAM OF THREE-PLATE HALF-WAVE RESISTANCE WELDING POWER SOURCE WITH PRIMARY RECTIFICATION

Secondary rectification systems are available in single-phase full-wave, three-phase half-wave, and three-phase full-wave forms; the latter is most commonly utilized. Three-phase full-wave systems consist of three single-phase transformers, with rectifiers connected to the secondary windings (Fig. 9). With this system, current pulse time is limited only by thermal considerations and polarity is fixed.

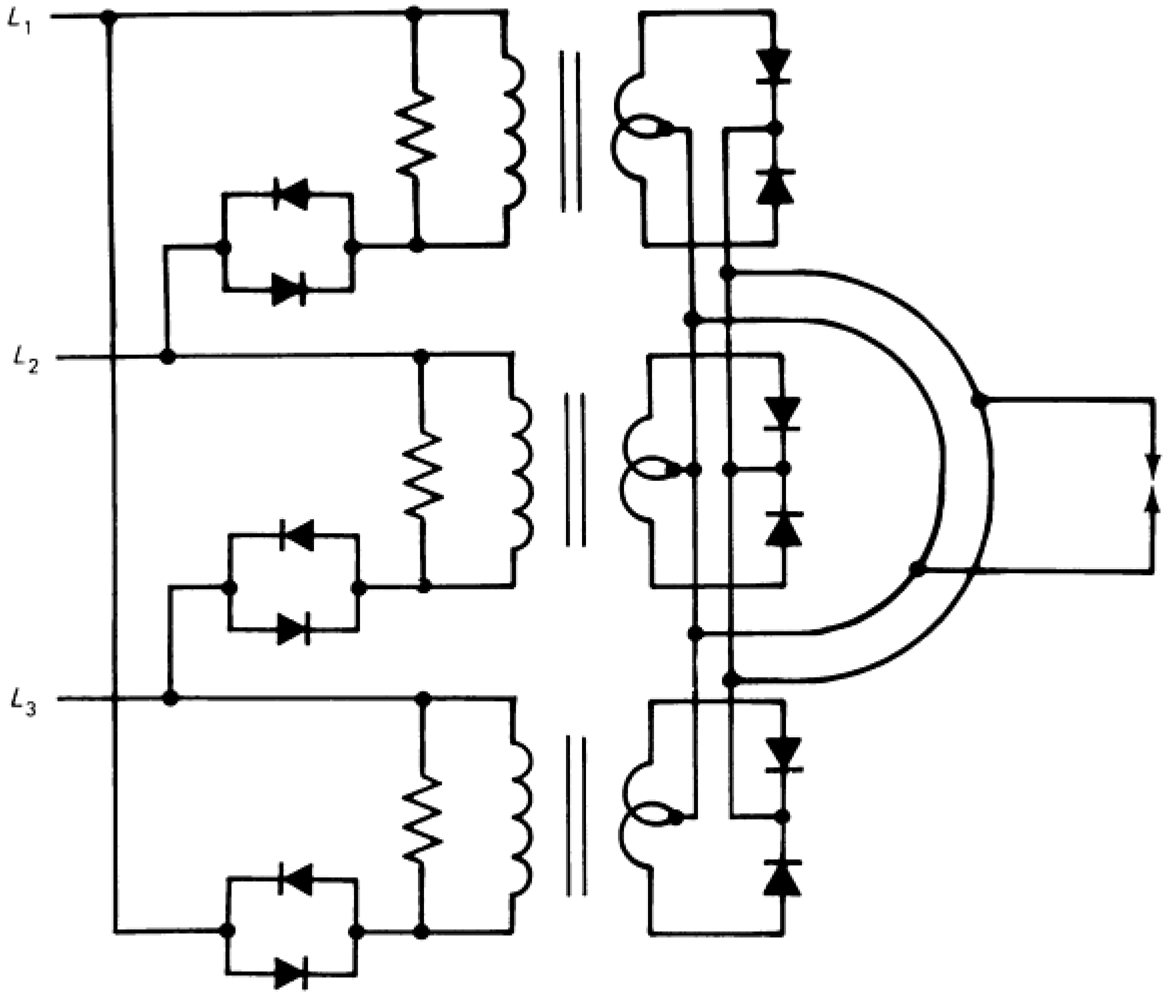

FIG. 9 CIRCUIT DIAGRAM OF THREE-PHASE FULL-WAVE RESISTANCE WELDING POWER SOURCE WITH SECONDARY RECTIFICATION

The energy efficiency of a resistance welder depends on the efficiency of delivered energy at the electrodes. Energy is dissipated in the form of heat, because of resistance in the conductors. The current, *I*, which produces the heat of welding, is simply the secondary voltage divided by the total machine impedance, *Z*, plus that of the weld:

$$i = \frac{E}{Z} \qquad \textbf{(EQ 8)}$$

where *Z* is the vector sum of the resistance and the reactance:

$$Z = \sqrt{R^2 + X^2} = \sqrt{R^2 + (wL)^2} \qquad \textbf{(EQ 9)}$$

where *R* is the total resistance of the circuit and $X = \omega L$ is the total reactance of the circuit (where ωequals $2\pi f$ and *L* is the inductance).

Resistance welders designed for single-phase utility power are lowest in investment cost, but highest in energy cost. The high-magnitude alternating current results in large inductive losses and a low power factor. Three-phase primary rectification and three-phase full-wave secondary rectification power sources reduce the power demand to about one half and operate at a higher power factor.

The rapidly pulsing weld zone temperature inherent with a single-phase source is least desirable from a welding standpoint. In addition, copper electrodes suffer from clamping impact, force, and heat, causing the occurrence of "mushrooming," which lowers current density.

Mushrooming also reduces resistance in the current path. Welding current will increase, because current is controlled indirectly in resistance welding machines by adjusting secondary voltage (that is, advancing or retarding the firing angle). From Eq 8, a decrease in impedance causes an increase in the current. Therefore, all resistance in the work path influences the current magnitude produced by a given secondary voltage. This influence is small if total system impedance is high, but significant if the system impedance is low.

A three-phase primary rectification system exhibits the greatest degree of this self-compensating effect, followed to a somewhat lesser degree by the three-phase secondary rectification category.

Finally, a thermoelectric effect is associated with passing current from a copper alloy electrode to a workpiece of a given material. This phenomenon is important when welding aluminum or magnesium alloys with a single impulse. Reversing the current polarity for each successive weld, an inherent characteristic of the three-phase primary rectification system, provides the maximum electrode life between dressings.

Duty Cycle. Resistance welding is inherently an intermittent process with a series of very short current periods followed by "off" or "cool" periods. Therefore, this type of power source is rated on a 50% duty cycle and a 1 min integrating period. The relationship between kVA rating at 50% duty cycle and kVA demand at actual duty cycle is as follows:

$$KVA_{50\%} = KVA_{Demand}\sqrt{2.(D.C.)} \qquad \textbf{(EQ 10)}$$

where D.C. is the duty cycle in %, entered in the form of a decimal. Duty cycle is defined by:

$$D.C. = \frac{N.T}{60.f} \qquad \textbf{(EQ 11)}$$

where N is the number of welds per minute, T is the weld time in cycles, and f is the frequency in hertz.

Heat Controls. Resistance welds are short and are measured in cycles. Because timing is very important, digital microprocessor controls abound. Electrochemical contactors have been replaced by SCRs or ignitrons connected in inverse parallel to handle the high currents. Coarse heat adjustment is accomplished by transformer tap switch setting, and fine adjustment is achieved through phase shifting, with digital circuitry providing the highest repeat accuracy.

Additional information on resistance welding is available in the Section "Joining Processes" in this Volume.

F. James Grist, Miller Electric Manufacturing Company, Inc.; William Farrell and Glen S. Lawrence, Ferranti Sciaky

Electron-Beam Welding Power Sources

Electron-beam welding (EBW) requires a variety of power supplies in two categories: those that generate and control beam position and those that control beam power. The outputs of various power supplies are coordinated by a machine control system.

Beam Position and Intensity. Beam position arrangements for alignment, deflection, and focus are dominated by magnetic devices in electron-beam welders. Because the magnetic field effect on the beam (the magnetomotive force) is directly proportional to the current in the windings of the device, these power supplies are typically current regulated.

$$MMF = N \cdot I \qquad \textbf{(EQ 12)}$$

where N is the number of turns in the coil and I is the current in the coil. The desired current value is programmed by potentiometer, computer, or function generator.

A second factor regarding the effectiveness of magnetic devices is the velocity of the electrons in the beam, as determined by the accelerating voltage. At the voltages used in EBW (typically, 50 to 150 kV), relativistic effects must be considered. In situations where the beam potential is changing, but it is desired to have the effect of the magnetic devices be constant, then the effective magnetic excitation, k, must be constant:

$$K^2 = \frac{(N.I)^2}{V^*} \qquad \textbf{(EQ 13)}$$

where NI is the magnetomotive force and V^* is the relativistic voltage. The relativistic voltage can be expressed in volts as:

$$V^* = V + (0.9785 \times 10^{-6})V^2 \qquad \textbf{(EQ 14)}$$

Sensors monitor the beam potential generated by the high-voltage (HV) supply and correct the control signal to the magnetic device power supply. Thus, the action of the device can be held constant as beam potential changes.

This correction can be achieved by using computer software or an analog circuit. Such correction is often called square root compensation, because of the relationship indicated in Eq 13.

Alignment coils are used for slight corrections to the x and y positions of the beam as it enters the magnetic focusing coil. Power supplies for these devices are typically potentiometer programmed by the operator in an alignment procedure. These supplies are characterized by a high degree of long-term stability, with minimal need for dynamic response. Selected installations employ alternate means for achieving alignment.

The magnetic focus coil acts very much as a glass focusing lens would for optics. It takes a divergent electron beam and converges it at a more distant point along the beam axis. The beam focus point from the magnetic lens is roughly inversely proportional to the square of the lens excitation:

$$f \propto \frac{V^*}{(N.I)^2} \qquad \textbf{(EQ 15)}$$

The magnetic lens supply is usually controlled by computer output in order to coordinate beam focus with the weld program. Manual control (operator offset) and square root compensation are also usually provided. These supplies are of high stability and moderate dynamic response.

The deflection coil moves the beam along the x and y axes to position the beam over the surface of the workpiece. Deflection rates can range from dc to high audio frequencies. Thus, the deflection supply is a high-quality current-regulated stereo amplifier.

Amplifier input controls allow computer input for coordinated movements of the part and beam, function generator inputs for beam deflection patterns (rasters, and so on), and dc offsets to patterns. These are highly stable current-regulated coils drivers with wide dynamic response.

High-Voltage Control. The electron beam is generated in an electron gun (resembling a diode or triode of cylindrical symmetry) that generates a paraxial flow of electrons. The power in an electron beam comes from the high-voltage supply that generates the large voltage difference between the anode and cathode of the gun. Electrons are emitted from the cathode and, accelerated by the voltage, fall to the anode. The beam passes through a central aperture in the anode and drifts through various alignment, focus, and deflection elements to strike the workpiece (Fig. 10). Both the workpiece and the anode are typically at ground potential, whereas the cathode is highly negative.

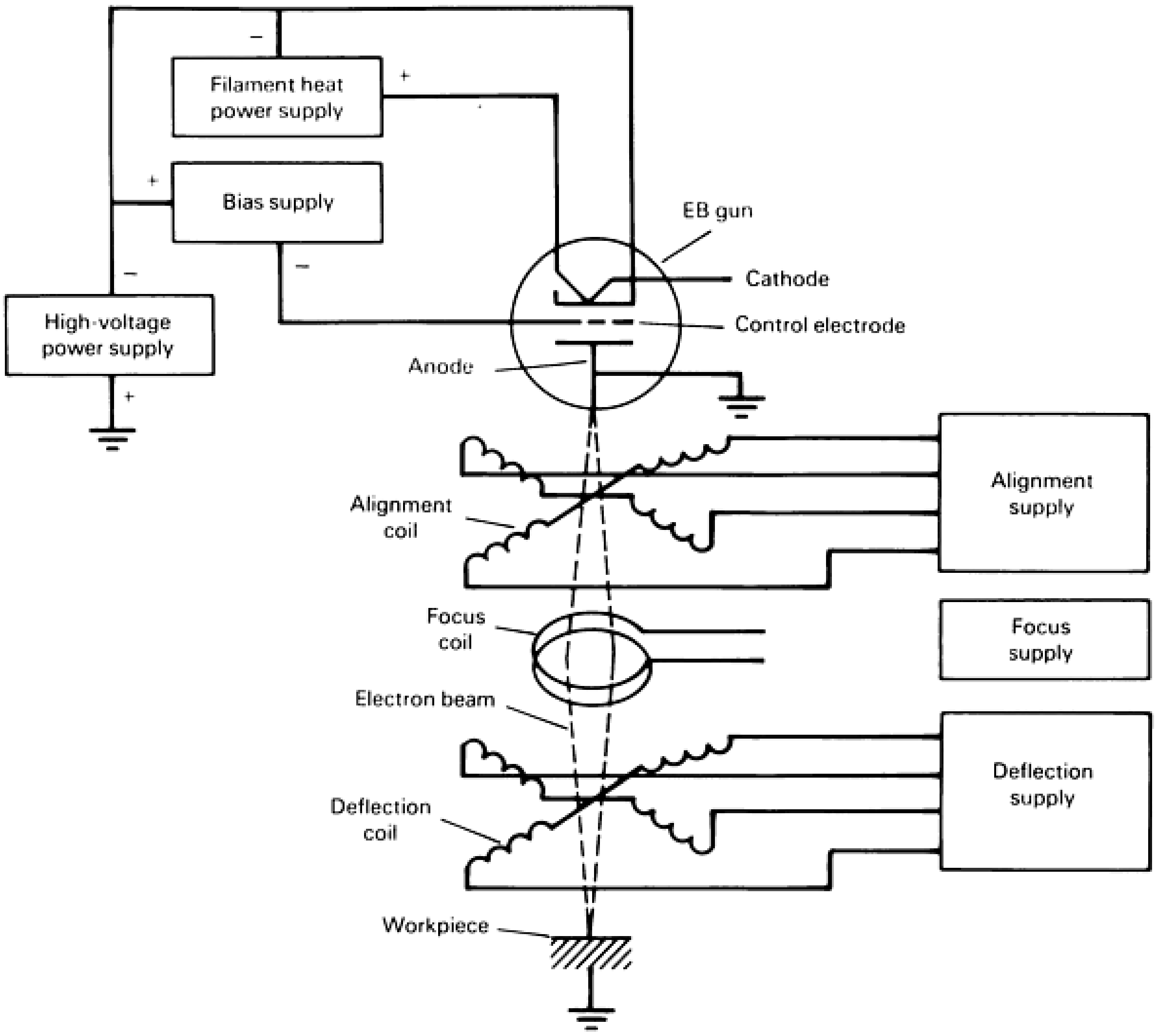

Filament heat power supply
Bias supply
High-voltage power supply
EB gun
Cathode
Control electrode
Anode
Alignment coil
Alignment supply
Focus coil
Focus supply
Electron beam
Deflection coil
Deflection supply
Workpiece

FIG. 10 SIX INDIVIDUAL POWER SUPPLIES THAT CONTROL POSITION AND INTENSITY OF THE BEAM IN EBW PROCESS

Control and regulation of the high voltage is usually done on the input to the high-voltage transformer/rectifier oil-filled tank. The input control installations range from motor-driven variable autotransformers, motor-alternators with field controls, SCR switching or phase control, to power MOSFET (metal-oxide semiconductor field-effect transistor) and IGBT devices in various switch mode regulators.

In a three-phase input unit, a standard technique is to place the switching control in the primary neutral conductor of the HV transformer (Fig. 11). The secondary circuitry is three-phase or six-phase, full-wave rectified, and filtered.

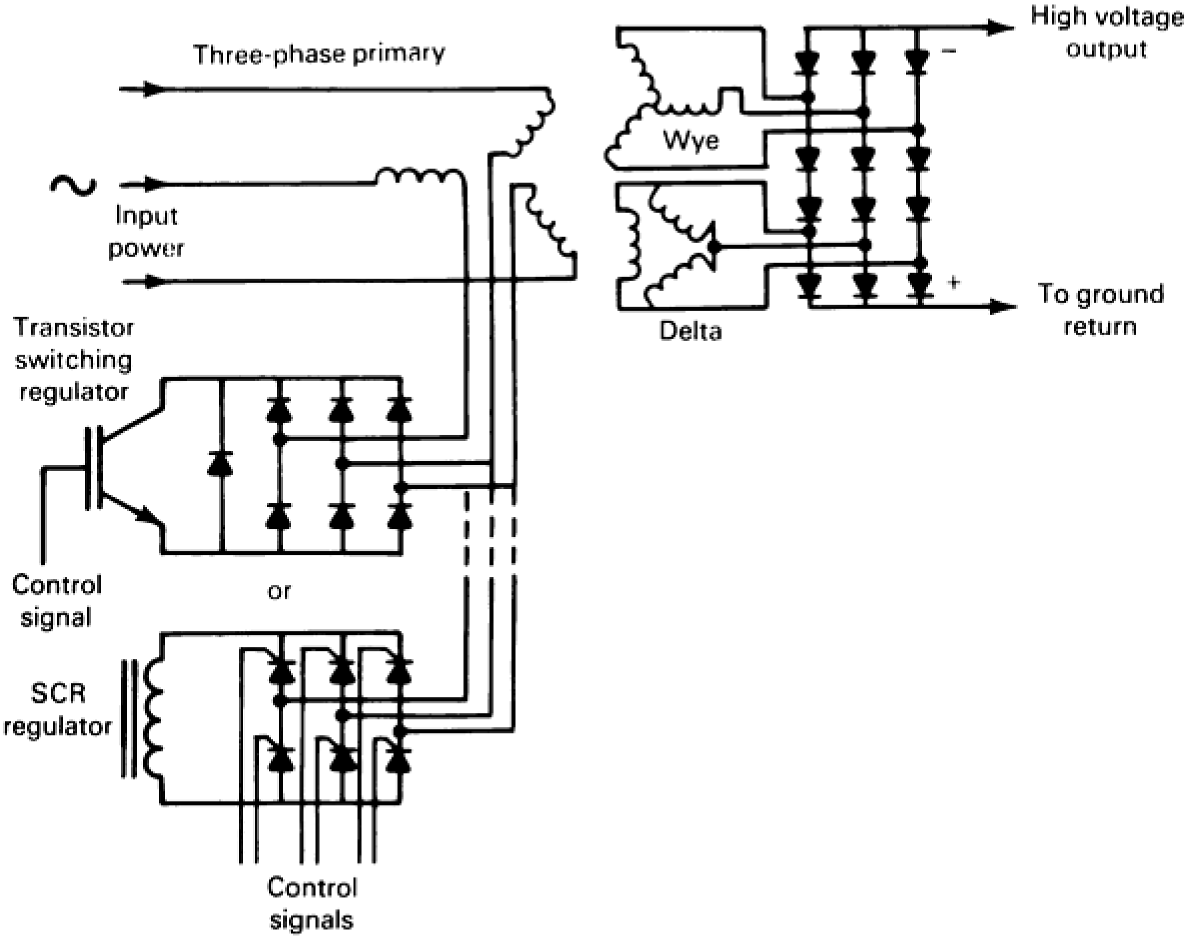

FIG. 11 CIRCUIT DIAGRAM OF A HIGH-VOLTAGE CONTROL IN NEUTRAL IN A THREE-PHASE INPUT UNIT

In an alternate arrangement, line voltage is converted to dc, which powers an inverter to produce high-frequency single-phase input to the high-voltage transformer. Control can be provided either by regulating the level of the dc supply or by controlling the inverter pulse width or frequency. The HV transformer can be a conventional step-up unit or a voltage multiplier. These supplies can be characterized as voltage regulated and input controlled, with a moderate response rate.

Thermionic Emission. The cathode heating unit provides the energy to raise the cathode to thermionic electron emission temperatures. This is usually a dc source. For beam-heated cathodes, discharge-heated cathodes, and others, it may be more complicated, to a point of supplying power to a second miniature gun.

In a space-charge limited electron beam, the cathode temperature must be held high enough that temperature-limited emission (Eq 16) is greater than space-charge limited flow (Eq 17):

$$I_P = AT^2 E^{-ef/kt} \qquad \textbf{(EQ 16)}$$

$$I_P = K(V_P + m V_G)^{3/2} \qquad \textbf{(EQ 17)}$$

In Eq 16 and 17, I_p is beam current, V_p is accelerating voltage, A is Dushman's constant, V_g is bias voltage, and T is temperature.

The bias supply provides voltage to the control electrode and regulates the magnitude of the beam current. Bias supplies can be input controlled with ground-level sensing of beam current on the return of the HV supply, or the beam can be sensed and regulated at high voltage using optical isolation (Fig. 12).

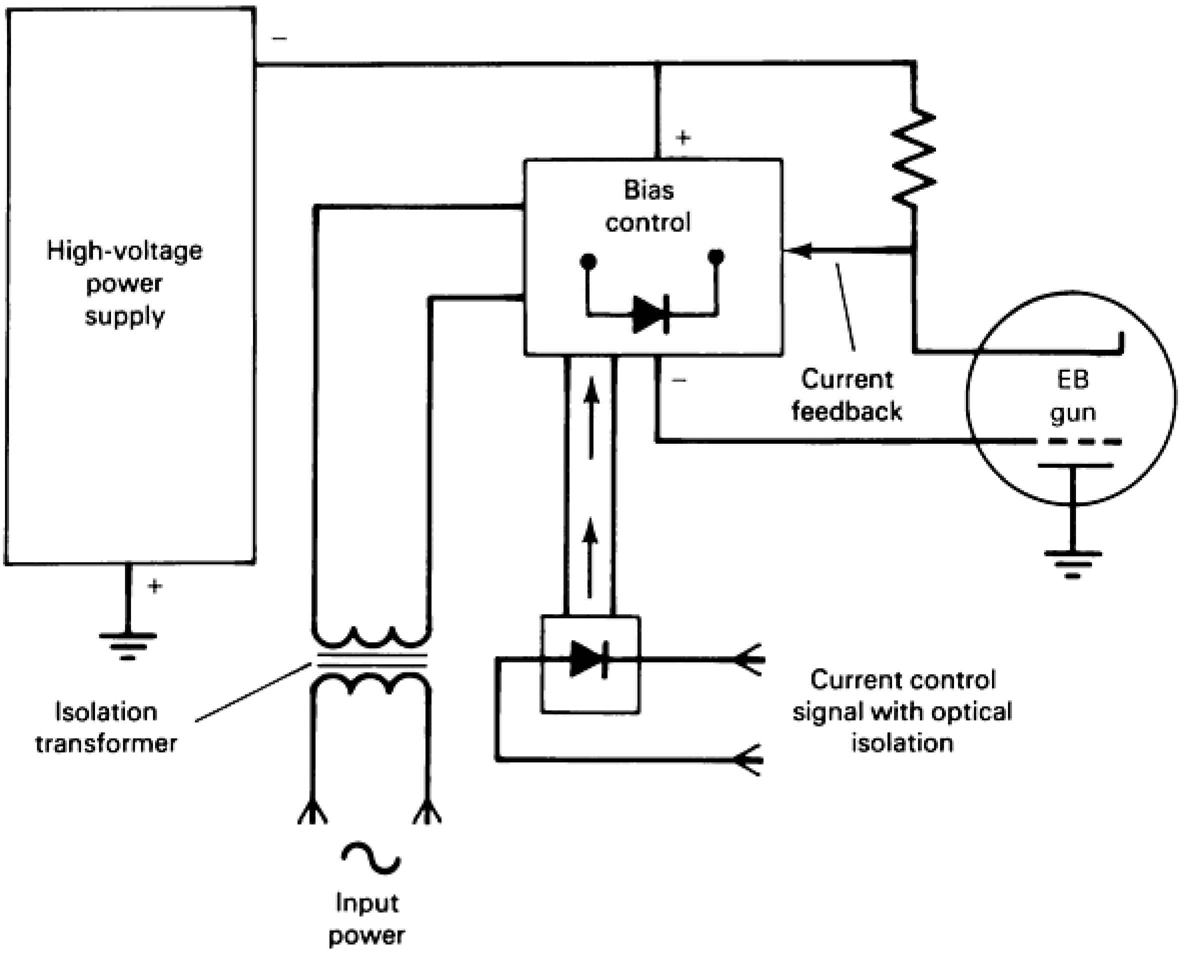

FIG. 12 CURRENT FEEDBACK AND BIAS CONTROL AT HIGH-VOLTAGE POTENTIAL

Ground-level sensing systems tend to use high-frequency inputs from linear amplifiers or switch mode inverters in order to achieve response on the order of a millisecond. Submillisecond response bias supplies are achieved by sensing and controlling the beam current at the high-voltage output. Because bias voltage is in the low kV range, vacuum tubes are used for output control.

Additional information on electron-beam welding is available in the article "Electron-Beam Welding" in this Volume.

Fundamentals of Weld Solidification

Harvey D. Solomon, General Electric Company

Introduction

OF ALL PHASE TRANSFORMATIONS, few have been more widely observed and studied than the transformation of a liquid to a solid (that is, solidification). The process of solidification is the same in all cases, whether it is the freezing of water on a windshield or in a freezer, or the solidification of metal in a casting or in the weld that joins two solids.

The process is controlled by the free energy of the liquid phase, G_l, relative to that of the solid, G_s. This is depicted in Fig. 1, which shows the behavior of a pure (single component) material. Above the freezing temperature, T_f, the liquid phase has the lower free energy and is therefore stable, but below T_f, the solid is the stable phase. At T_f, both phases are in equilibrium, that is, $G_l = G_s$.

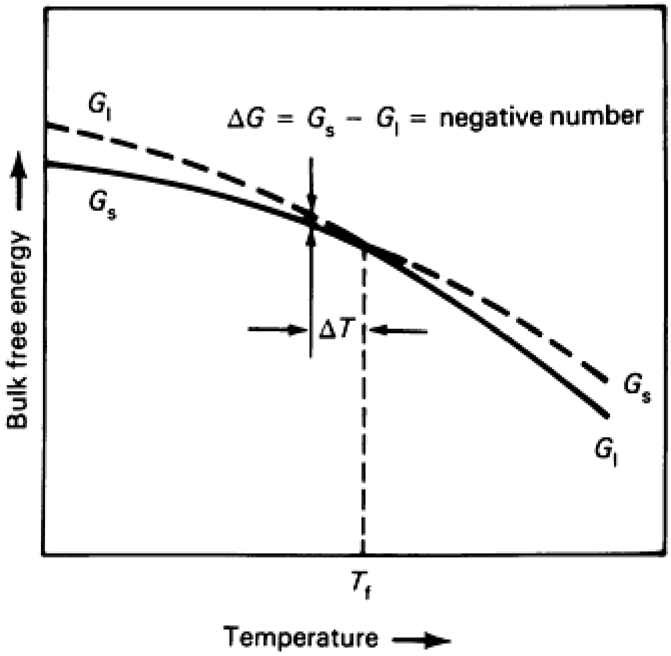

FIG. 1 TEMPERATURE DEPENDENCE OF BULK FREE ENERGY OF THE LIQUID AND SOLID PHASES IN SINGLE-COMPONENT SYSTEM. THE SOLID LINE PORTION AND THE DASHED LINE PORTION OF EACH G_L, AND G_S CURVE INDICATE THE STABLE AND UNSTABLE PHASES, RESPECTIVELY, OF THE FREE ENERGY ON EITHER SIDE OF T_F. SOURCE: REF 1

In the transition from one phase to another, the change in free energy, ΔG, is the difference in free energy of the product and the reactant. This free energy change can be expressed in terms of the enthalpy and the entropy changes, that is, for the transformation of a liquid to a solid during freezing:

$$\begin{aligned}\Delta G = G_S - G_L &= (H_S - H_L) \\ - T(S_S - S_L) &= \Delta H - T\Delta S\end{aligned} \qquad \textbf{(EQ 1)}$$

At the freezing temperature, T_f, $\Delta G = G_s - G_l = 0$, because the free energy of the two phases is the same, and $\Delta H = T_f\Delta S$, It is necessary to cool below T_f for solidification, because at T_f both the solid and liquid phases are present and in equilibrium. Below T_f, ΔG is not equal to zero (Fig. 1 shows that $G_s < G_l$) and is given by Eq 1 with $T = T'$, where $T' < T_f$.

Because the ΔH and ΔS are not strong functions of temperatures, they can be assumed to be temperature independent. Therefore, at any temperature, $\Delta H = \Delta H_f$ and $\Delta S = \Delta S_f$, where ΔH_f and ΔS_f are the values of the enthalpy and the entropy changes for the equilibrium reactions at T_f (that is, the latent heat of fusion and the entropy change on fusion, respectively). Combining these enthalpy and entropy expressions, the fact that $\Delta H = T_f\ \Delta S$, and Eq 1, then at T', one obtains:

$$\Delta G = (\Delta H_F/T_F)(T_F - T') \qquad \textbf{(EQ 2)}$$

where ΔH_f, the latent heat of fusion, is negative. Hence, in agreement with Fig. 1, ΔG is negative. The greater the amount of undercooling (supercooling) below T_f ($T_f - T'$), then the greater the thermodynamic driving force for solidification.

However, even when the conditions of Eq 2 are met, the liquid does not spontaneously transform to the solid below T_f. Rather, small amounts of solid nucleate and grow to produce complete solidification. Nucleation creates a new surface, that is, the surface between the solid and the liquid. The energy per unit are of this surface is the surface tension, γ, which is always positive. For solidification to occur, the increase in energy associated with the surface energy must be balanced by a greater decrease in the free energy of the solid relative to that of the liquid. This requires undercooling, as shown by Eq 2.

There are three ways in which a solid can form: homogeneous nucleation, heterogeneous nucleation, and epitaxial growth. Homogeneous nucleation occurs when there is no foreign body (mold wall, solid particle in the melt, etc.) on which to form the solid. Figure 2 shows the balance of the surface tension and the bulk free energy per unit volume, ΔG_v, as a function of the size of the nucleus that forms during homogeneous nucleation. The ΔG_v, value is just ΔG (as given by Eq 2) divided by the molar volume of the solid, V_s.

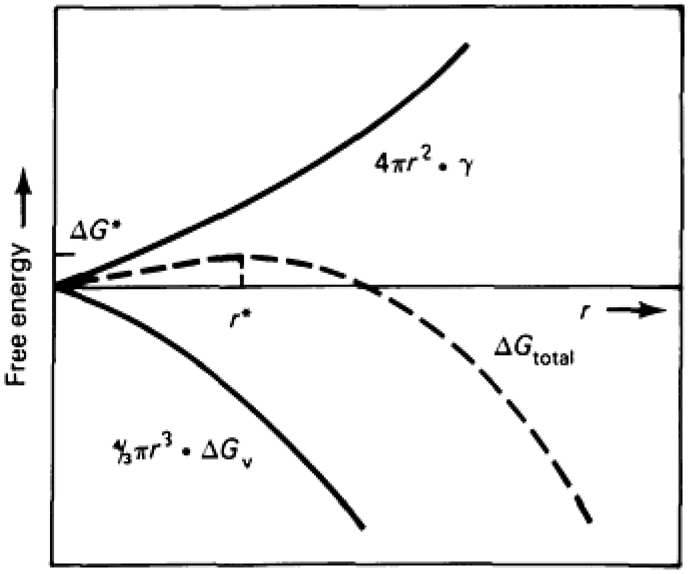

FIG. 2 FREE ENERGY OF FORMATION OF A NUCLEUS AS FUNCTION OF ITS RADIUS. SOURCE: REF 1

For a spherical nucleus of radius greater than r_c, the volume free energy decrease outweighs the increase in energy that is due to the surface energy, and the nucleus is stable. At $r = r^*$, the net energy is a maximum, but if additional atoms are added to the nucleus, then the energy decreases. The value r^* is thus the radius at which the spherical nucleus is just stable, because as additional atoms are added, the energy is decreased. The value r^* is given by:

$$R^* = -2\gamma/\Delta G_V \qquad \textbf{(EQ 3)}$$

which is a positive number, because ΔG_v is negative. Substituting into Eq 2, one gets:

$$R^* = -2\,\gamma T_F V_S/\Delta H_F(T_F - T) \qquad \textbf{(EQ 4)}$$

The r^* is positive because ΔH_f is negative and, for undercooling, $(T_f - T)$ is positive. This equation is a general expression relating the radius of curvature of a surface and the degree of undercooling that is required for solidification of that surface. The greater the degree of undercooling, the smaller the radius of curvature that is stable. This equation will be considered again when nonplanar solidification is discussed.

Nucleation requires a thermal energy either equal to or greater than ΔG^* in order for a spherical nucleus of radius r^* to form. The ΔG^* value is the activation energy, which, as Fig. 2 shows, is a positive quantity. The greater the degree of undercooling, the more negative the free energy, the less positive the ΔG^* value, and the greater the rate of nucleation of the solid. For homogeneous nucleation, ΔG^* is given by:

$$\Delta G^* = (16/3)(\pi\,\gamma^3 T_F^2 V_S^2)/(\Delta H_F^2 \Delta T^2) \qquad \textbf{(EQ 5)}$$

where $\Delta T = T_f - T$.

Heterogeneous nucleation develops when the solid forms on a foreign body. Now, the interaction of the nucleus and the foreign body must be considered. This interaction is defined in terms of the wetting angle between the nucleus and foreign body. The activation energy for heterogeneous nucleation is given by:

$$\Delta G^{*} = \frac{4p}{3}\frac{g^{3}T_{f}^{2}V_{s}^{2}}{\Delta H_{f}^{2}\Delta T^{2}}(2 - 3\cos q + \cos^{3} q) \qquad \textbf{(EQ 6)}$$

The angle is 180° if there is no wetting, and Eq 6 reduces to Eq 5, that is, with no wetting, there can only be homogeneous nucleation. With a wetting angle of 90°, ΔG^* is one half the value given by Eq 5. Instead of requiring a spherical nucleus of radius r^*, a hemispherical nucleus of the same radius is necessary, which requires one half the number of atoms and one half the activation energy. This is the sort of nucleation that develops in a casting, where the mold wall acts as the foreign body.

When the wetting angle is zero, Eq 6 shows that the activation energy is zero. This is the case for epitaxial growth on a substrate, where epitaxial derives from the Greek *epi*, upon, and *taxis*, to arrange, that is, to arrange upon. In effect, no new surface is being formed. Atoms are just being added to the substrate, thereby extending it. The important point is that no activation energy nor undercooling is required to add atoms onto an existing substrate. This is the situation that develops when a liquid solidifies on a substrate of the same material or one that is similar in composition and structure, as in the solidification of a weld.

Reference

1. J.D VERHOEVEN, *FUNDAMENTALS OF PHYSICAL METALLURGY*, JOHN WILEY & SONS, 1975

Fundamentals of Weld Solidification

Harvey D. Solomon, General Electric Company

Comparison of Casting and Welding Solidification

A weld can be thought of as a miniature casting. The fundamentals of weld solidification are the same as those of a casting, but with different boundary conditions. These differences and their effects are described below.

First, by its very nature, a sound weld *must* attach to the metals being joined, whereas a casting *must not* adhere to the mold wall. To achieve this, the mold is treated to prevent adherence, whereas the joint to be welded is prepared to promote adherence. The practical result of this preparation is that there is generally excellent heat transfer through a weld joint, but relatively poor heat transfer through the mold, because oxides are often used as mold materials or for mold coatings (the molten metal will not wet the oxide and therefore cannot adhere to it).

Second, heat is continually being added to the weld pool as it travels, whereas no heat is added to a casting after the pour, except for possibly modest heating of the mold. The practical result of this and of the first consideration described above is that the temperature of the casting is relatively uniform. In contrast, a very large temperature gradient develops in a weld. The center of a weld pool reaches a very high temperature (typically, 2000 to 2500 °C, or 3630 to 4530 °F), which is limited by the vaporization of the weld metal. At the sides of the weld pool, where solidification is taking place, the temperature is the solidification temperature, which is typically about 1000 °C (1800 °F) less than the temperature at the center on the weld pool. Thus, a positive temperature gradient is developed in a weld (that is, measured from the fusion line of the weld, the temperature increases going into the molten material). A casting will typically supercool to below the solidification temperature. Solidification will first develop by heterogeneous nucleation at the mold wall. The latent heat and the poor heat transfer through the mold wall will then cause a temperature rise at the mold wall. Thus, a small negative temperature gradient can develop briefly in a casting. The nature of this gradient is important, because it promotes nonplanar solidification (to be discussed below) in the casting, whereas the positive gradient that develops in a weld helps to limit nonplanar solidification.

Third, solidification is developed at the mold wall of a casting by heterogeneous nucleation, whereas epitaxial growth develops at the weld fusion line. Thus, some supercooling is required for the casting, but only a vanishingly small degree of supercooling is required for the weld. (As was discussed, epitaxial growth requires no supercooling, but some slight degree of undercooling is required to shift the reaction from an equilibrium between the solid and liquid, to all solid.) This undercooling in a weld would typically be *less* than 1 °C (1.8 °F). In contrast, in castings when there is a lack of effective nuclei for hetergeneous nucleation, undercoolings of 100 °C (180 °F) or more are possible. In general, however, the undercooling in castings is typically only on the order of a few degrees centigrade.

Fourth, the generally larger volume of a casting, relative to that of a weld, and the poorer heat transfer makes the cooling rate and the solidification rate much lower for castings than for welds.

Fifth, as a casting solidifies, the volume of the remaining liquid decreases. Thus, the shape of the molten pool is continually changing. In a weld, the weld pool shape is generally kept constant as it travels (if the heat input and section geometry are constant).

Sixth, because of the stirring action of the arc and the action of Marongoni surface tension gradient induced convection forces, there is good mixing of the molten weld pool. In contrast, there is comparatively little mixing of the molten material of a casting.

As discussed in the section "Solidification of Alloy Welds (Constitutional Supercooling)" in this article, all of these factors influence the nature of weld solidification, relative to what is observed in a casting.

Solidification of Alloy Welds (Constitutional Supercooling)

The solidification of an alloy is much more complex than that of a pure metal, whether in a weld pool or in a casting. Figure 3 shows a simple binary isomorphous phase diagram. The temperatures of the liquidus and solidus lines of Fig. 3 can be approximated by straight lines (for a limited temperature range). (The liquidus defines the temperature above which only liquid is present and the solidus the temperature below which only solid is present. The liquidus and solidus lines bound the two-phase liquid plus solid region.) The liquidus line is given by:

$$T_L = T_{MA} + M_L C_L \qquad \textbf{(EQ 7)}$$

where T_{mA} is the melting point of pure A, M_L is the slope of the liquidus, and C_L is the composition of the liquid at T_L. The solidus is described by a similar expression.

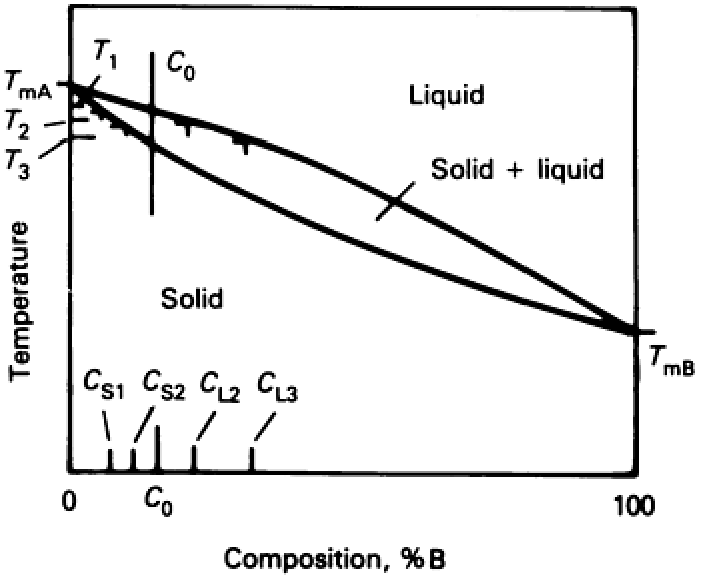

FIG. 3 BINARY ISOMORPHOUS PHASE DIAGRAM SHOWING THE LIQUID AND SOLID COMPOSITIONS AT SELECTED TEMPERATURES

Another important quantity that can be estimated from the phase diagram is the distribution coefficient, *K,* which is defined as:

$$K = C_S/C_L \qquad \textbf{(EQ 8)}$$

where C_S and C_L are the solidus and liquidus compositions, respectively, defined by equilibrium tie lines. Assuming that the solidus and liquidus lines are straight lines means that *K* is not a function of temperature. In general, the liquidus and solidus lines are curved, so that in actuality, *K* is a function of temperature. For simplicity, this curvature will be neglected and, instead, limited temperature ranges where *K* can correctly be assumed to be a constant will be considered. The use of a distribution coefficient defined by equilibrium tie lines implies that equilibrium exists between the liquid and the solid at the solid-liquid interface. This equilibrium need not exist elsewhere. Figure 3 and the following discussion are based on a *K* value of less than 1 and a negative M_L value. The following discussion is also true for diagrams where *K* is greater than 1 and M_L is positive.

The distribution coefficient defines how the solute is distributed or partitioned between the solidifying solid and the remaining liquid. The slope of the liquidus line and the distribution coefficient are important in defining the composition gradient that develops when alloys solidify. In general, the distribution coefficient, *K,* is a function of temperature, but as already noted, to simplify the analysis, it will be assumed that over a limited temperature range, it is a constant.

An equilibrium phase diagram, such as that of Fig. 3, assumes that equilibrium is maintained in all of the solid and liquid and that there are no composition gradients in the solid or liquid. It defines the equilibrium solid and liquid compositions at any temperature. Figure 3 shows that the solidification of an alloy of composition C_0 will begin when the temperature is reduced below T_1. If equilibrium is maintained, then solidification will be complete when the temperature is reduced below T_3. A two-phase mixture of solid plus liquid will be present between T_1 and T_3, with the relative amount of each phase given by the lever rule:

$$\%_{liquid} = \left(\frac{C_0 - C_{s2}}{C_{1.2} - C_{s2}} \right) 100\% \qquad \textbf{(EQ 9)}$$

at T_2. The free energy curves corresponding to this equilibrium are much more complex than those of Fig. 1, which is relevant for a pure material.

The equilibrium predicted by Fig. 3 is never realized in practice, because it would require an infinitely slow cooling rate. When considering castings, it is generally assumed that the equilibrium predicted by the phase diagram is maintained at the solid-liquid interface, but that no composition change occurs in the solid as the temperature is lowered. At T_1, a solid of composition C_{S1} forms. C_{S1} is equal to $C_{L1}K$, and because at the liquidus $C_{L1} = C_0$ the $C_{S1} = C_0K$. When the temperature drops to T_2, the solid that forms has a composition of C_{S2}, and so on. A composition gradient is thus developed in the solid. As the solid forms, solute is rejected into the liquid, and it is assumed that this solute is mixed into the liquid, raising its composition to that predicted by the phase diagram. This occurs even though the mixing action in a casting is relatively poor, because the cooling rate is generally low enough to provide enough time to allow for the required solute redistribution by diffusion in the liquid. As a result, no gradient is developed in the liquid. The composition of the solid at the solid-liquid interface is still given by the equilibrium phase diagram. However, the previously formed solid is deficient in solute. As a consequence, complete solidification does not occur at T_3. Further cooling is required, with solidification being completed only when the *average* solid composition is C_0.

The boundary conditions for a weld are somewhat different from those of a casting. The composition gradient that develops in the solid is assumed to be the same as that which forms in a casting, but the behavior of the liquid is assumed to be different. The mixing of the melt is greater than that developed in a casting, but the cooling rate is much higher. Thus, the solute that is rejected as the solid forms is not completely mixed into the liquid. Rather, the bulk liquid remains at the initial composition, C_0, with a solute gradient formed in the liquid near the solid-liquid interface. At the liquids, T_1, the solid that forms has a composition of $C_{S1} = C_0K$. At T_2, the solid composition is C_{S2}. As with castings, it is assumed that the solid-liquid equilibrium predicted by the equilibrium phase diagram is maintained at the solid-liquid interface. Therefore, the liquid in equilibrium with this solid is $C_{L2} = C_{S2}/K$. However, the lack of complete mixing means that the bulk liquid remains at C_0, and that a composition gradient forms directly ahead of the solid-liquid interface. This composition gradient is illustrated in Fig. 4, which shows the composition gradient developed by unidirectional solidification and the temperature gradient driving this solidification. The solidification is modeled by assuming that a bar is being unidirectionally solidified, but this approach is applicable to solidification in general.

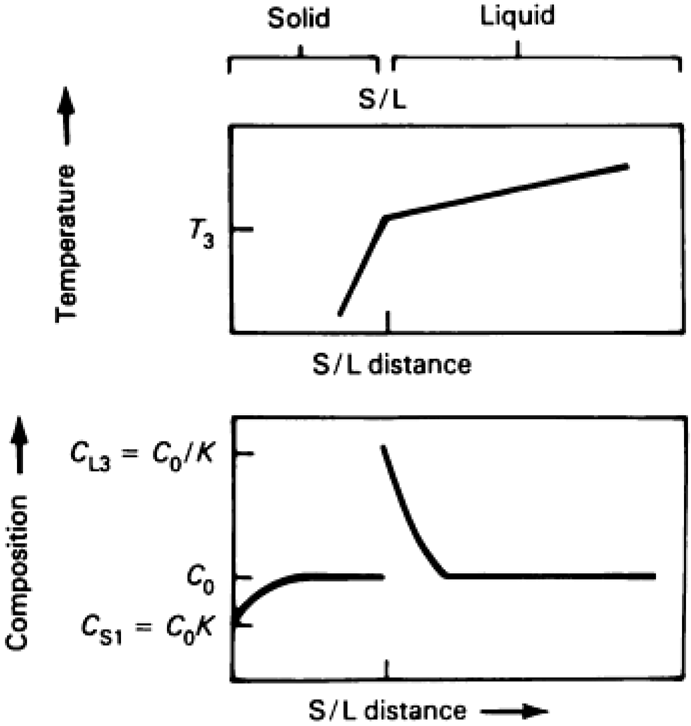

FIG. 4 INITIAL STAGE OF DIRECTIONAL SOLIDIFICATION OF A BAR, SHOWING THE TEMPERATURE PROFILE DRIVING THE SOLIDIFICATION AND THE COMPOSITION PROFILE ESTABLISHED FOR AN ALLOY WITH A SOLUTE CONTENT OF C_0. SOURCE: REF 4

At temperature T_3, the solid that forms has a composition of C_0. The solid that is forming now has the same composition as the bulk liquid. Thus, an equilibrium situation exists. Solidification can now continue *without* having to further reduce the temperature. For castings, however, the composition gradient in the solid and complete mixing of the solute into the liquid leaves the liquid still present at T_3, thereby reducing the liquidus temperature. This means that, in a casting, the temperature must be lowered even more before further solidification can take place.

Figure 4 shows the situation *during* weld solidification, whereas Fig. 5 shows the composition when solidification is complete. For most of the solidification, the solid forms at C_0. The exceptions are at the start and at the end of solidification. The first solid that forms does so with a solute composition of C_{S1}, which the phase diagram shows is less than C_0. A composition gradient develops until temperature is lowered to T_3, when the solid composition reaches C_0 and an equilibrium is reached between the solid being formed and the composition of the bulk liquid. At the end of solidification, there is an increase in the solute content. This occurs to accommodate the solute buildup at the solid-liquid interface. At the end of the solidification, there is no longer any ***bulk*** liquid at C_0. The solute that is ahead of the solid-liquid interface raises the composition of the remaining liquid as it is rejected into this final small volume of liquid. This obviates the assumed boundary condition (a lack of mixing of the solute into the bulk liquid), and the situation becomes more like that of a casting (complete mixing in this small liquid volume). The final liquid composition of this final transient is the terminal composition predicted by the phase diagram, that is, the final liquid is driven to the composition of the lowest melting point. In the case of the simple binary isomorphous diagram used for this example, this is pure B. For a simple binary eutectic, it would be the eutectic composition.

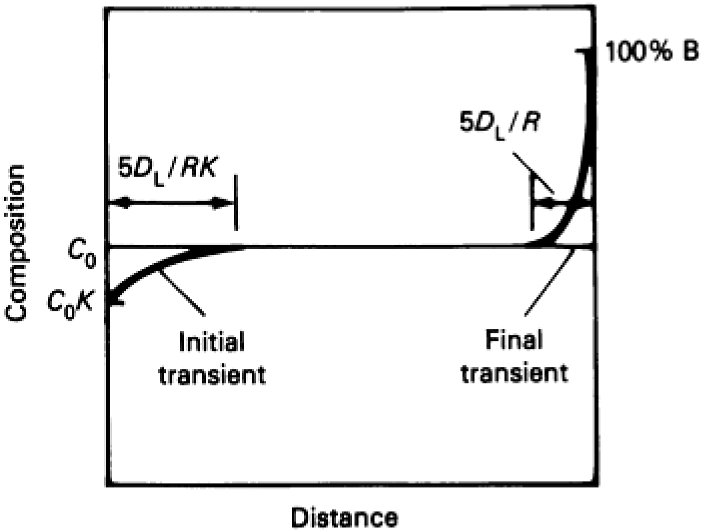

FIG. 5 FINAL COMPOSITION PROFILES FOR THE UNIDIRECTIONAL SOLIDIFIED BAR SHOWN IN FIG. 4. THE INITIAL AND FINAL TRANSIENT ARE SHOWN FOR *K* < 1. *R,* RATE OF SOLIDIFICATION; D_L, DIFFUSION RATE OF THE SOLUTE IN THE LIQUID. SOURCE: REF 4

The profile of the liquid composition gradient that is ahead of the solid-liquid interface is given by:

$$C_L = C_A e{-[RX_L / D_L]} + C_0 \qquad \textbf{(EQ 10)}$$

where R is the rate of solidification. D_L, is the diffusion rate of the solute in the liquid, C_a is the increase in the composition of the liquid at the solid-liquid interface (relative to C_0), and X_L is the distance into the liquid that is ahead of the solid-liquid interface. The liquid at the solid-liquid interface has a composition of C_{L3}, which is in equilibrium with $C_{S3} = C_0$. Therefore, $C_{L3} = C_0/K$. The enrichment of the solute in the liquid, C_a, is thus $(C_0/K) - C_0$ or $C_0(1 - K)/K$. The distribution coefficient, K, is thus an important variable in determining the composition profile in the liquid and, as will be shown, in determining the nature of the solidified microstructure.

Substituting Eq 10 into Eq 7, one gets the expression describing the temperature variation of the liquidus that is ahead of the solid-liquid interface, namely:

$$T_L = T_{MA} + M_L C_A e{-[RX_L / D_L]} + M_L C_0 \qquad \textbf{(EQ 11)}$$

Figure 6 illustrates this liquidus profile, along with three possible temperature gradients that could actually exist in the liquid. They are all positive, reflecting the positive gradient that develops in a weld. Gradient A is greater than the slope of the liquidus at the solid-liquid interface, but those of B and C are not. The slope of the liquidus can be found by differentiating Eq 11 at $X_L = 0$, that is, at the solid-liquid interface. For the temperature gradient to be less than this slope, the gradient G must be:

$$G < M_L R C_0 (K - 1)/D_L K \qquad \textbf{(EQ 12)}$$

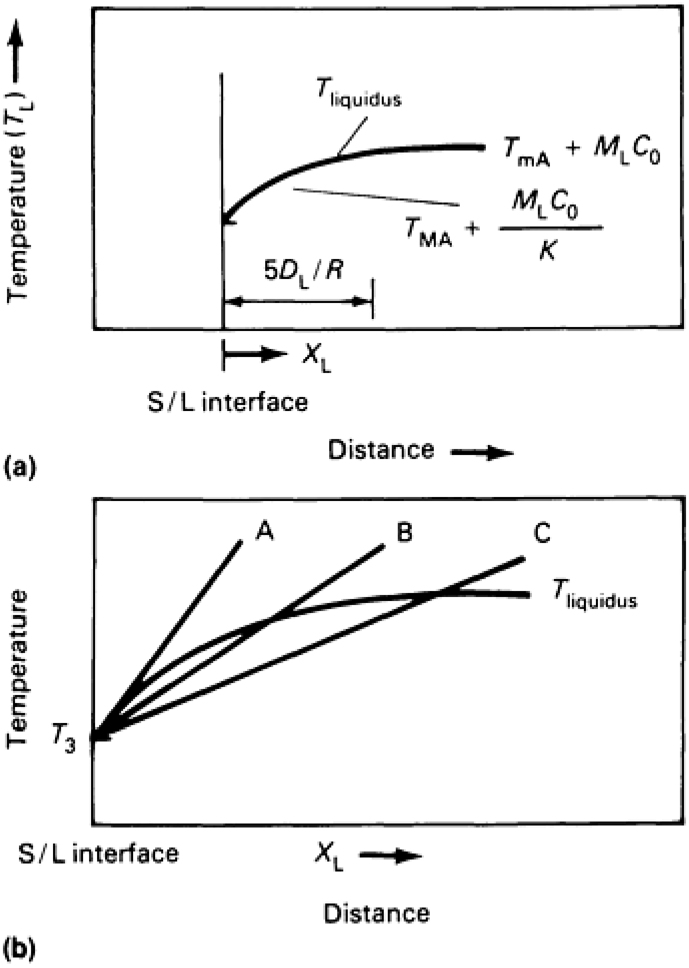

FIG. 6 LIQUIDUS PROFILE OF A SOLID-LIQUID INTERFACE. (A) LIQUIDUS PROFILE FOR STEADY-STATE CONSTITUTIONAL SUPERCOOLING. (B) THREE LIQUIDUS TEMPERATURE GRADIENTS (CURVES A, B, AND C) COMPARED TO THE LIQUIDUS PROFILE. SOURCE: REF 6

Equation 12 is important because it describes the temperature gradient required for constitutional supercooling (Ref 3, 4, 5). The establishment of the composition gradient in the liquid that is ahead of the solid-liquid interface produces a variation in the temperature of the liquidus. For curves B and C, this has allowed the actual temperature to be below the liquidus (that is, supercooling has developed, even though the temperature of the bulk liquid is above the bulk liquidus temperature). This is constitutional supercooling, and as will be shown, this promotes nonplanar solidification, even though there is a positive temperature gradient. The liquidus profile shown in Fig. 6 is correlated with the propensity for hot cracking. The greater the temperature difference between that of the solid-liquid interface and that of the bulk liquid, the greater the tendency for hot cracking. This is discussed in more detail in the article "Cracking Phenomena Associated with Welding" in this Volume.

The constitutional supercooling model for describing weld solidification is presented because it qualitatively describes the evolution of different microstructures (described in the section "Development of Weld Microstructures" in this article). Many newer nonequilibrium approaches to solidification give better quantitative results. These approaches will be discussed in the section "Nonequilibrium Effects: High-Rate Weld Solidification and Composition Banding" in this article.

References cited in this section

3. J.A. BURTON, R.C. PRIM, AND J. SLICHTER, *J. CHEM. PHYS.,* VOL 21, 1953, P 1987-1991
4. W.A. TILLER, K.A. JACKSON, J.W. RUTTER, AND B. CHALMERS, *ACTA METALL.,* VOL 1, 1953, P 428-437
5. W.A. TILLER AND J.W. RUTTER, *CAN. J. PHYS.,* VOL 34, 1956, P 96-121
6. W.F. SAVAGE, E.F. NIPPES, AND T.W. MILLER, *WELD. J.,* VOL 55, 1976, P 165S-173S

Fundamentals of Weld Solidification

Harvey D. Solomon, General Electric Company

Development of Weld Microstructures

Nonplanar solidification develops when a protrusion moves ahead of the rest of the solid-liquid interface and continues to grow in a stable manner. This increases the surface area. The stable radius is given by Eq 4, which shows that the protrusion must move into a supercooled region in order to be stable. This is possible in a casting, because a negative temperature gradient is developed (that is, the temperature decreases from the solid-liquid interface into the liquid). In a pure material (single-component system), the positive temperature gradient that is established in the liquid of a weld prevents this nonplanar growth. In an alloy, however, constitutional supercooling allows nonplanar solidification in a weld, even with a positive temperature gradient.

The lower the gradient, G, (compare curve C with curve B in Fig. 6), the greater the degree of constitutional supercooling. This is illustrated in Fig. 7, which shows several temperature gradients (Fig. 7f) compared to the variation in the liquidus and the nature of the solid-liquid interface (Ref 5, 6, 7). For a curve a, the temperature gradient is steeper than the liquidus temperature curve and only planar solidification is possible. For curve b, there is a shallower gradient and the solid-liquid interface can move by cellular growth. The gradient is successively lower for c and d, which allows larger protrusions and the development of cellular dendritic and columnar dendritic growth. At curve e, the gradient is so shallow and the resulting liquidus temperature is so far above the actual temperature in the liquid that equiaxed dendrites can form ahead of the solid-liquid interface.

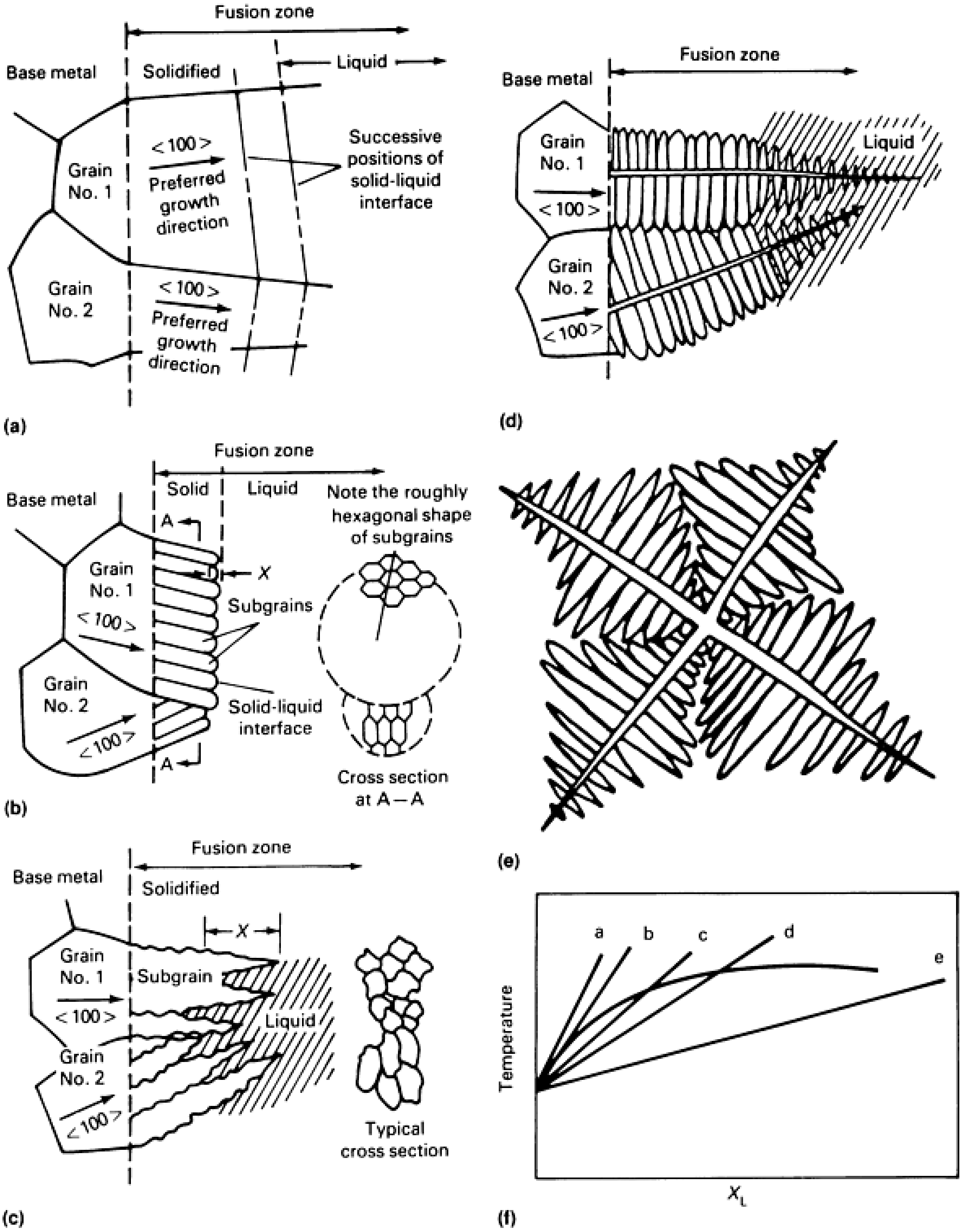
Fusion zone
Base metal
Solidified
Liquid
Grain No. 1
Grain No. 2
< 100 >
Preferred growth direction
Successive positions of solid-liquid interface
(a)
Fusion zone
Base metal
Solid
Liquid
A
X
Note the roughly hexagonal shape of subgrains
Subgrains
Solid-liquid interface
Grain No. 1
Grain No. 2
< 100 >
Cross section at A—A
(b)
Fusion zone
Base metal
Solidified
X
Grain No. 1
Subgrain
< 100 >
Liquid
Grain No. 2
Typical cross section
(c)
Fusion zone
Base metal
Grain No. 1
Liquid
< 100 >
Grain No. 2
(d)
(e)
a
b
c
d
e
Temperature
X_L
(f)

FIG. 7 SCHEMATICS SHOWING MICROSTRUCTURE OF SOLID-LIQUID INTERFACE FOR DIFFERENT MODES OF SOLIDIFICATION AND THE TEMPERATURE GRADIENTS THAT GENERATE EACH OF DIFFERENT MODES. (A) PLANAR GROWTH. (B) CELLULAR GROWTH. (C) CELLULAR DENDRITIC GROWTH. (D) COLUMNAR DENDRITIC GROWTH. (E) EQUIAXED DENDRITE. (F) FIVE TEMPERATURE GRADIENTS VERSUS CONSTITUTIONAL SUPERCOOLING. SOURCE: REF 6

Figure 8, which is based on experimental observations, relates the gradient, solidification rate, and composition to the type of structure developed (Ref 5, 6, 7). The basis for this figure lies in Eq 12. If R is moved to the left side of Eq 12, then the right side contains only material variables. For a given alloy system (where D_L, M_L, and K are fixed), the greater the C_0, the greater the tendency for constitutional super-cooling and nonplanar solidification. Equation 12 predicts that the smaller the G/R, the more nonplanar the solidification. Figure 8 shows that, experimentally, the correlation is with $G/\sqrt{R}$, rather than G/R.

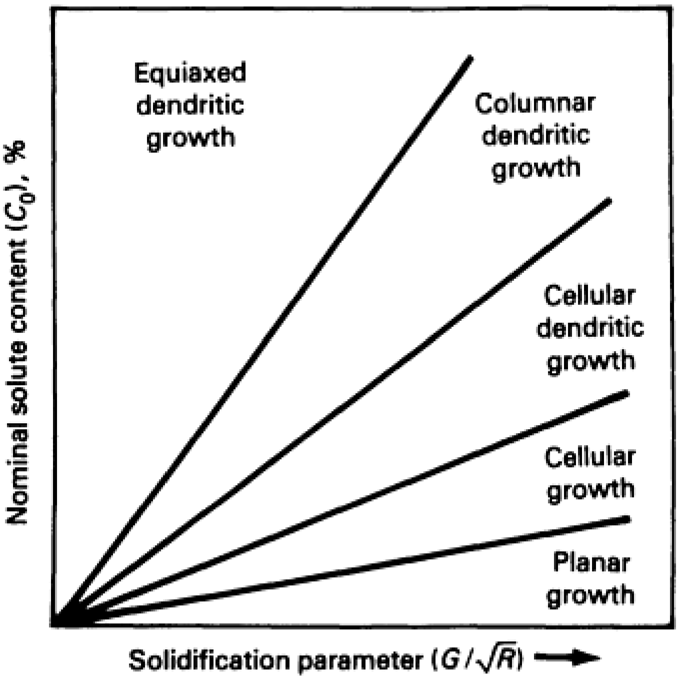

FIG. 8 PLOT OF SOLUTE CONTENT VERSUS SOLIDIFICATION PARAMETER TO SHOW THE NATURE OF SOLIDIFICATION. SOURCE: REF 6

As the dendrite, or columnar grain, grows into the liquid, it should follow the behavior shown in Fig. 5. (As will be discussed in the section "Nonequilibrium Effects: High-Rate Weld Solidification and Composition Banding" in this article, the experimentally determined compositions do not always exactly agree with that predicted by Fig. 5.) The center of each cell or dendrite behaves like the initial transient, that is, the solute content is low. The regions between the cells or dendrites behave like the final transients, that is, they are enriched in solute. This is illustrated schematically in Fig. 9. Cellular growth occurs for a smaller C_0 and/or for a larger $G/\sqrt{R}$ than that for dendritic growth (see Fig. 8).

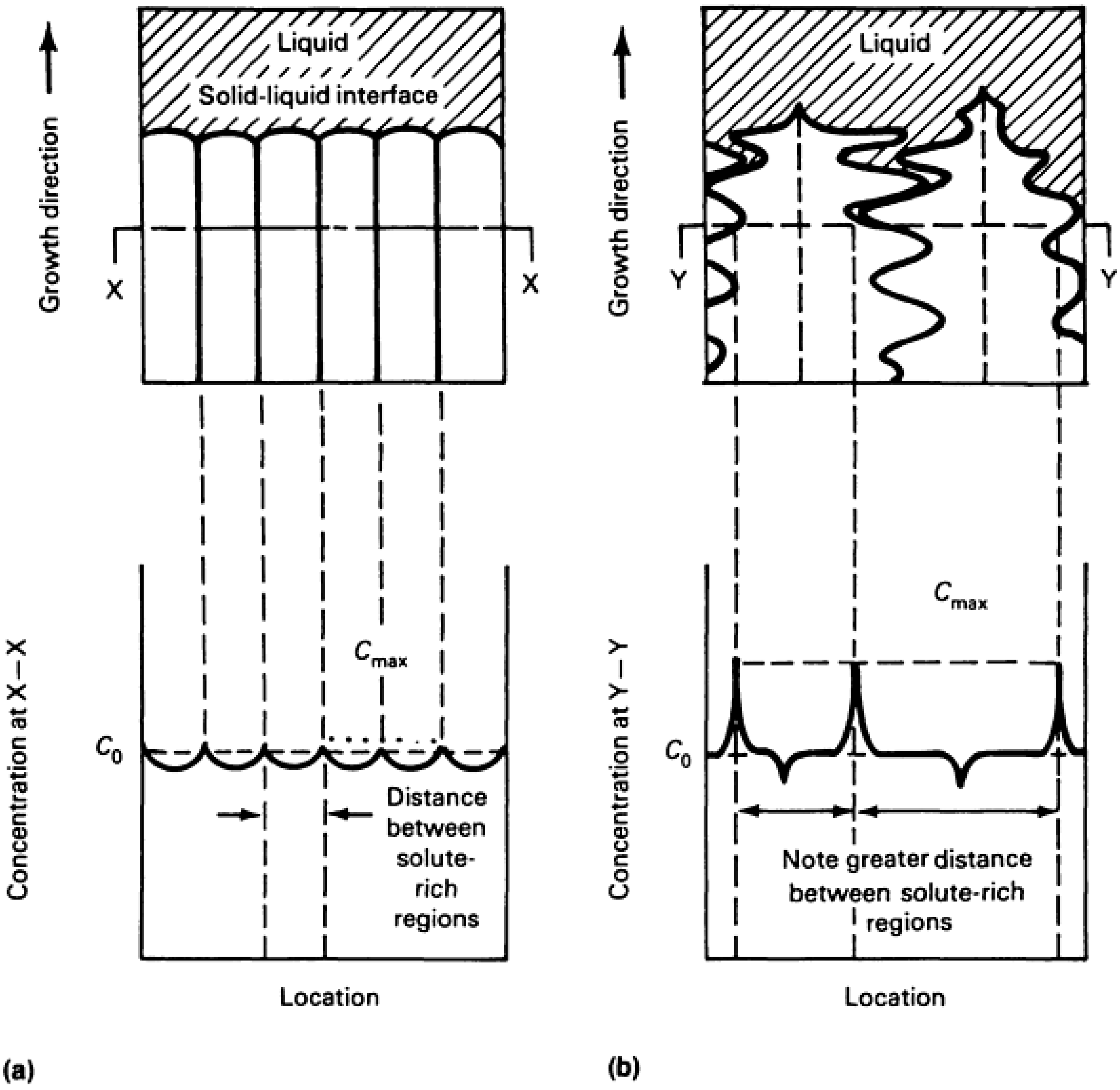
Liquid
Solid-liquid interface
Growth direction
X
X
Concentration at X—X
C_{max}
C_0
Distance between solute-rich regions
Location
(a)
Liquid
Growth direction
Y
Y
Concentration at Y—Y
C_{max}
C_0
Note greater distance between solute-rich regions
Location
(b)

FIG. 9 SCHEMATIC SHOWING SOLUTE DISTRIBUTION AT THE DENDRITE OR CELL CORE AND IN THE INTERCELLULAR OR INTERDENDRITIC REGIONS. (A) CELLULAR GROWTH. (B) DENDRITIC GROWTH. SOURCE: REF 8

The constitutional cooling described by Fig. 4, 5, 6, 7, 8, and 9 occurs on a very fine scale. Solving Eq 10 shows that the solute composition that is ahead of the solid-liquid interface will decrease to almost C_0 when $X_L = 5D_L/R$. For a typical liquid diffusion rate of 5 × 10^{-5} cm^2/s (5 × 10^{-8} ft^2/s) and a solidification rate of 5 mm/s (0.2 in./s), the liquid composition drops to C_0 within about 5 μm (200 μin.). In the initial transient, the distance over which the solid composition increases to C_0 is given by $5D_L/RK$. Thus, for $K < 1$, this distance is greater than the region in the liquid just ahead of the solid-liquid interface. This is illustrated in Fig. 5. This difference is also important in defining the extent of the final transient, which is of the order of the length of the zone in the liquid where the composition drops to C_0. The length of the initial transient is thus $1/K$ larger than that of the final transient. The length of the solute-poor cell or dendrite core is thus $1/K$ larger than that of the solute-rich intercellular or interdendritic regions.

The solute-rich region between the cells or dendrites has a lower freezing temperature than that of the cellular or dendritic core. Thus, it is possible to have liquid films between impinging cells or dendrites. These films cannot support the thermally induced shrinkage stress that develops during solidification, and hot cracking can develop (see the article "Cracking Phenomena Associated with Welding" in this Volume). The hot cracking produced by nonplanar solidification can be minimized by reducing C_0 (that is, by using purer materials) and/or by increasing $G/\sqrt{R}$ (see Fig. 8).

The temperature gradient in the liquid, G, is *inversely* proportional to the heat input, Q, which is given by:

$$Q = FEI/V \qquad \textbf{(EQ 13)}$$

where f is the efficiency of the welding process, E is the arc voltage, I is the arc current, and V is the welding torch velocity. Thus, G can be increased by decreasing either the arc voltage or arc current, or by increasing the torch velocity. The influence of the velocity is counterbalanced by the fact that the solidification rate, R, is proportional to V (see the section "Effect of Welding Rate on Weld Pool Shape and Microstructure" in this article). Thus, increasing V increases R, which tends to increase the degree of constitutional supercooling.

References cited in this section

5. W.A. TILLER AND J.W. RUTTER, *CAN. J. PHYS.*, VOL 34, 1956, P 96-121
6. W.F. SAVAGE, E.F. NIPPES, AND T.W. MILLER, *WELD. J.*, VOL 55, 1976, P 165S-173S
7. F. MATSUDA, T. HASHIMOTO, AND T. SENDA, *TRANS. NATL. RES. INST. MET. (JPN)*, VOL 11, 1969, P 43
8. W.F. SAVAGE, C.D. LUNDIN, AND A.H. ARONSON, *WELD. J.*, VOL 44, 1965, P 420S-425S

Fundamentals of Weld Solidification

Harvey D. Solomon, General Electric Company

Effect of Welding Rate on Weld Pool Shape and Microstructure

The velocity of the welding torch affects not only the rate of solidification, but the shape of the weld pool and the propensity to develop centerline hot cracks. The shape of the weld pool is dictated by the velocity, V, at which the welding torch moves and by the rate at which heat can be removed at the solid-liquid interface. To keep a constant shape, the rate of new melting must be exactly balanced by the solidification rate. On the average, this solidification occurs normal to the solid-liquid interface, because this is the direction of the maximum temperature gradient and, thus, the direction of the maximum heat removal. Simple geometry requires that for the weld pool shape to remain constant, the solidification rate, R, must be related to the torch velocity by:

$$R = V\,\mathrm{COS}\ f \qquad \textbf{(EQ 14)}$$

where f is the angle between the normal to the solid-liquid interface and the direction of motion of the torch (Fig. 10). At the back of the weld pool, where $f = 0$, $R = V$. If this were not so, then the weld pool would be elongating or shortening. At the side of the weld pool, where $f = 90°$, $R = 0$. If this were not the case, then the weld pool would be decreasing in diameter. With a constant weld pool shape, solidification at the side of the weld pool occurs as the region adjacent to where $f = 90°$ sweeps by, rather than by the motion normal to the weld pool velocity.

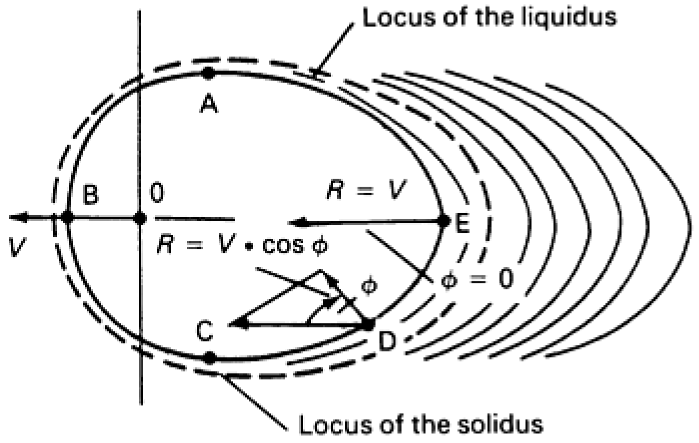
Locus of the liquidus
A
B
0
$R = V$
E
V
$R = V \cdot \cos \phi$
ϕ
$\phi = 0$
C
D
Locus of the solidus

FIG. 10 SCHEMATIC OF A MOVING WELD POOL SHOWING THE RELATIONSHIP BETWEEN VELOCITY OF TRAVEL OF WELDING TORCH, *V*, AND THE RATE OF SOLIDIFICATION, *R*, AT SELECTED POINTS ALONG WELD POOL BOUNDARY. SOURCE: REF 9

Equation 14 predicts that *R* will vary around the weld pool. Figure 8 shows that as *R* varies, so does the microstructure. The result of these two effects is a variation of the microstructure around the weld pool (Ref 7). Planar solidification is to be expected at the sides of the weld pool, with the extent of the nonplanar solidification increasing towards the center of the weld pool. This is shown schematically in Fig. 11.

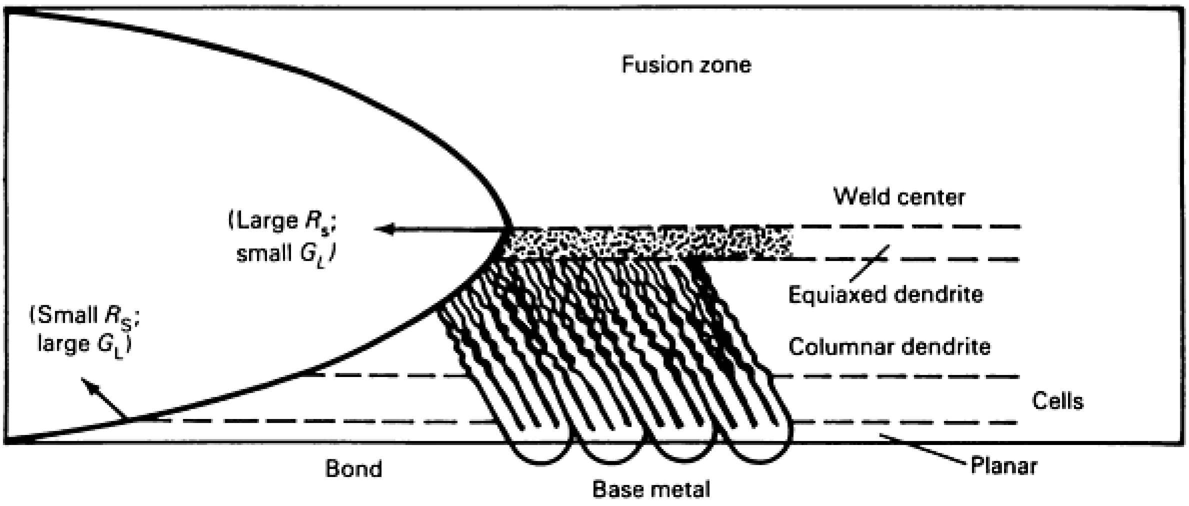
Fusion zone
Weld center
(Large R_s; small G_L)
(Small R_S; large G_L)
Equiaxed dendrite
Columnar dendrite
Cells
Bond
Base metal
Planar

FIG. 11 SCHEMATIC SHOWING VARIATION OF MICROSTRUCTURES IN RESPONSE TO VARIATION OF THE SOLIDIFICATION RATE AROUND THE WELD POOL. SOURCE: REF 7

The coarseness of the nonplanar structure is also a function of *R*. The dendrite tip radius varies inversely with the square root of *R*. Not only does the dendrite tip radius become smaller as *R* increases, the dendrite arm spacing does also, although the exact form of the dependency is more complex. The same general inverse dependency with *R* is also true for the cell spacing when the solidification is cellular.

The ability of the solid-liquid interface to move at a velocity, *V*, depends on two factors: the maximum solidification rate, which is a function of the crystallographic orientation of the solid at the point in question, and the ability of the temperature gradient in the solid to remove the heat and establish the appropriate temperature for solidification to proceed. Consider first the influence of the orientation of the solid grain at the solid-liquid interface. For body-centered cubic (bcc) and face-centered cubic (fcc) metals, the preferred growth direction for epitaxial solidification is the <100> direction. For hexagonal close-packed (hcp) metals, the preferred growth direction is <1010>.

The planar growth of grains oriented with this direction normal to the solid-liquid interface is favored. This is illustrated in Fig. 12. Grains with the preferred orientation will grow at the expense of adjacent grains that are less favorably oriented (Ref 8, 9, 10, 11, 12). Because dendrites follow the same preferred growth directions, those with the proper orientation will be favored. The requirement for a constant weld pool size becomes somewhat more complex when the effect of crystallographic orientation is considered (Fig. 12). Equation 14 can be modified to become:

$$R' = V\,\mathrm{COS}\ f\ /\mathrm{COS}\ (f' - f) \qquad \textbf{(EQ 15)}$$

where R' is the growth rate in the preferred direction required to maintain a constant weld pool shape and f' is the angle between R' and V, and $(f' - f)$ is the angle between the local preferred growth direction and R, which is normal to the solid-liquid interface.

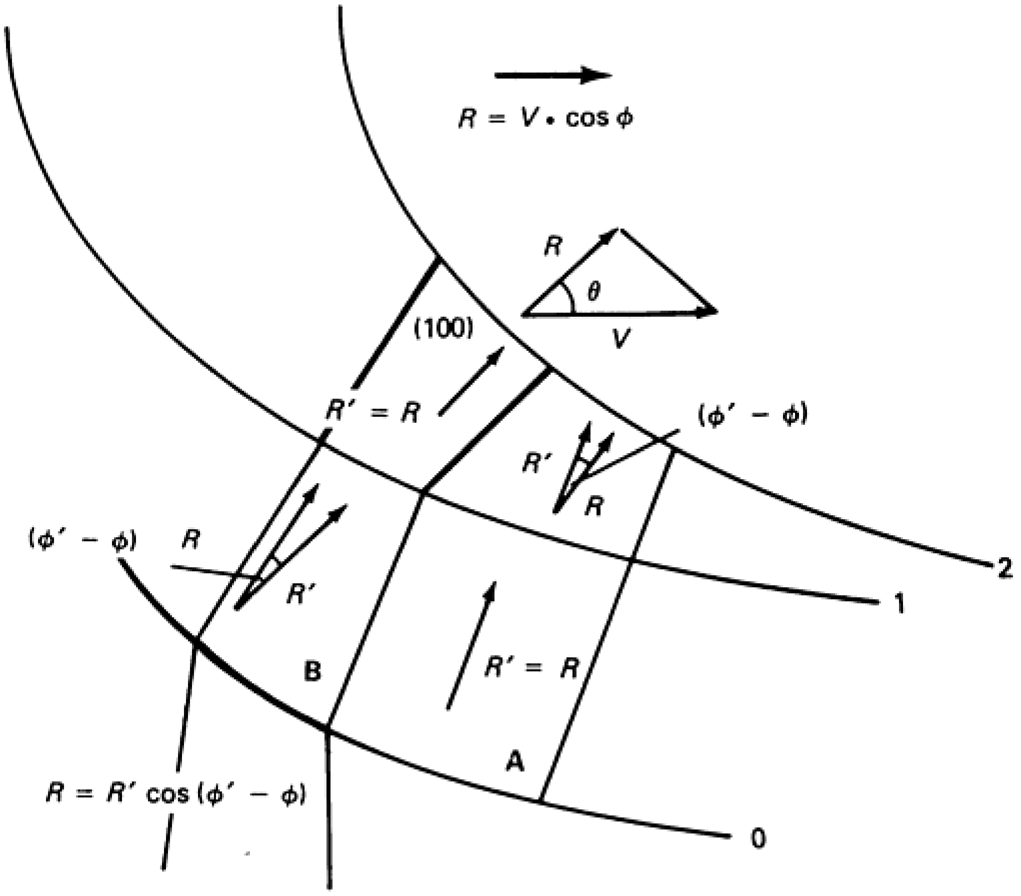

FIG. 12 SCHEMATIC SHOWING MOVEMENT OF A CURVED SOLID-LIQUID INTERFACE FOR SEVERAL GRAINS (A AND B), AND THE CHANGE IN THE RAPID GROWTH DIRECTION RELATIVE TO THE INTERFACE POSITION.

Figure 12 shows the situation at successive positions of the weld pool as it moves in the direction of V. The situation for the weld moving from position 1 to position 2 is shown. The weld pool is assumed to be elliptical in shape, so that at any point (except at the very back of the weld pool where $f = 0$) as the pool moves, the angle between R (or R') and V changes. If an adjacent grain becomes oriented such that R' becomes normal to the solid-liquid interface [that is, so that $(f' - f)$ is equal to zero], it will become most favored and will grow at the expense of adjacent grains. In Fig. 12, between positions 0 and 1, grain A is more favorably oriented than grain B. The reverse is true between positions 1 and 2. Figure 12 uses only a few widely spaced positions, but in reality there is a continuum of such positions.

Competitive growth causes grains to try to rotate into the direction of *V*. Figure 13(a) shows how the grains rotate in an elliptical weld pool. Dendrites will also try to rotate into the *R* direction, although the grains or dendrites do not actually rotate. Rather, the growth shifts to the most favored <100> (or <1010>) direction of the six present in the fcc, bcc, or hcp lattices. As grains are edged out by more favorably oriented grains, they get smaller and eventually can disappear. The solute that they were rejecting is spread out as this competitive process favors and disfavors different grains. Those grains or dendrites that eventually grow into the *V* direction will reject their solute into the weld pool, which gradually becomes solute enriched, contributing to possible crater cracking when the weld terminates. To combat this, the weld power should be gradually decreased in order to maintain a steep temperature gradient (minimizing the degree of constitutional supercooling). The use of a runoff weld tab could also be considered.

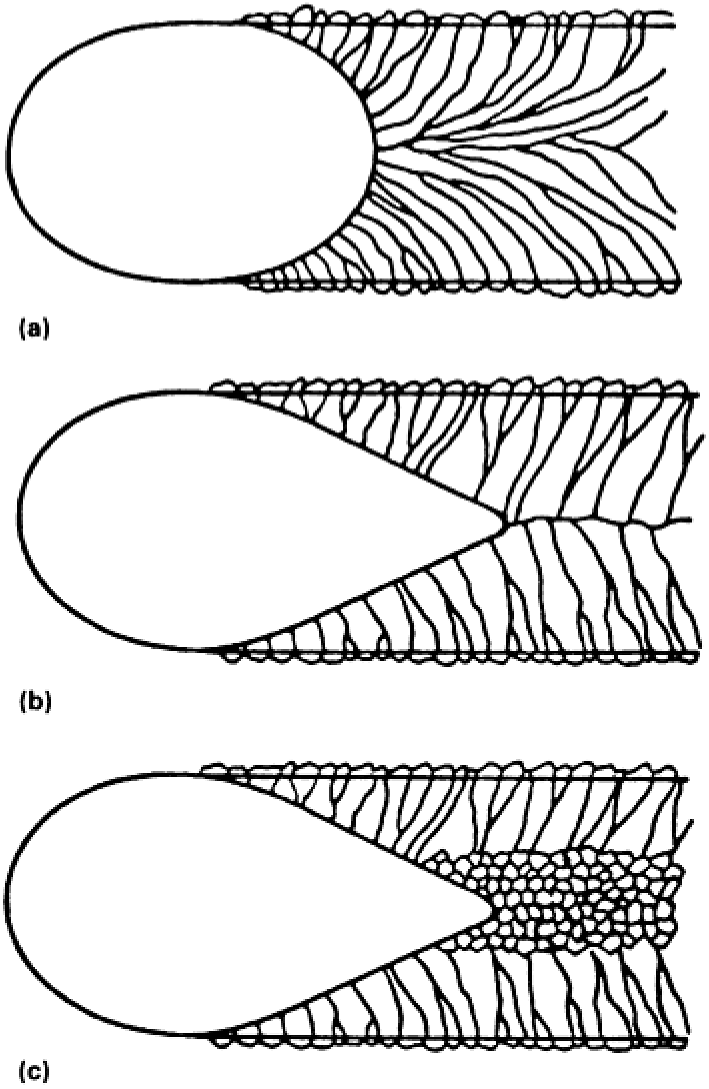

FIG. 13 SCHEMATIC SHOWING EFFECT OF HEAT INPUT AND WELDING SPEED VARIATIONS ON WELD GRAIN STRUCTURE. (A) LOW HEAT INPUT AND LOW WELDING SPEED, PRODUCING AN ELLIPTICAL WELD POOL. (B) HIGH HEAT INPUT AND HIGH WELDING SPEED, PRODUCING A TEAR-DROP-SHAPED WELD POOL. HERE, THE HEAT INPUT AND WELDING SPEED ARE NOT YET SUFFICIENT TO CAUSE HETEROGENEOUS NUCLEATION AT THE WELD POOL CENTERLINE. (C) HIGH HEAT INPUT AND HIGH WELDING SPEED, WITH HETEROGENEOUS NUCLEATION AT THE WELD POOL CENTERLINE. SOURCE: REF 13

The second factor to consider is the ability to extract heat at the solid-liquid interface. Heat is extracted by the temperature gradient of the solid. This gradient is a minimum at the back of the weld pool, where the temperature of the solid is the highest. Unfortunately, this is exactly where the solidification rate must be a maximum to keep up with *V*. The temperature gradient is a maximum at the side of the weld pool, where the least heat extraction is required, because here $R = 0$. As long as the maximum possible solidification rate, R_{max}, is equal to or greate than *V*, the weld pool can maintain an elliptical shape. If *V* is increased above R_{max}, then the solidification rate at the back of the weld pool cannot keep up, and the weld pool elongates. Increasing V decreases the heat input, which shrinks the weld pool volume. With the weld elongating, this means that the diameter must decrease as the pool elongates. This causes the weld to take on a tear-drop shape, as shown in Fig. 13(b). The angle at the back of the weld pool will decrease until f_{critical} is reached. f_{critical} is given by:

$$f_{\text{CRITICAL}} = \text{COS}^{-1} (R_{\text{MAX}}/V) \qquad \textbf{(EQ 16)}$$

R_{max} is controlled by the thermal conductivity of the solid, the geometry of the part being welded (primarily the thickness), and the preferred direction growth rate.

When there is a shift to a tear-drop weld pool, the grain (or dendrite) rotation is limited to f_{critical}. When this situation is attained (Fig. 13b), the grains (or dendrites) will all grow together at the center of the weld pool. The solute that is being rejected into the terminal transient will build up at the centerline, which can lead to centerline cracking. This cracking is caused by the lower solidification temperature of this solute-rick centerline region. As the surrounding material solidifies, stresses are developed that cannot be supported by the liquid centerline material, causing centerline cracking. The obvious remedy is to reduce the welding rate, although this has a production rate penalty.

Figure 13(b) typifies what is to be expected from a pure material or one with a low solute content. In the case of alloys or in cases when the solute content is high, Fig. 8 predicts that when R is increased, not only does the weld pool become tear-drop shaped, but the nature of the structure changes, as shown in Fig. 11. If R is large enough, equiaxed grains can develop from heterogenous nucleation in the supercooled centerline material (Fig. 13c). Such behavior has indeed been observed in aluminum alloys (Ref 13, 14, 15). Figure 14 shows actual weld microstructures corresponding to the schematic structures of Fig. 13.

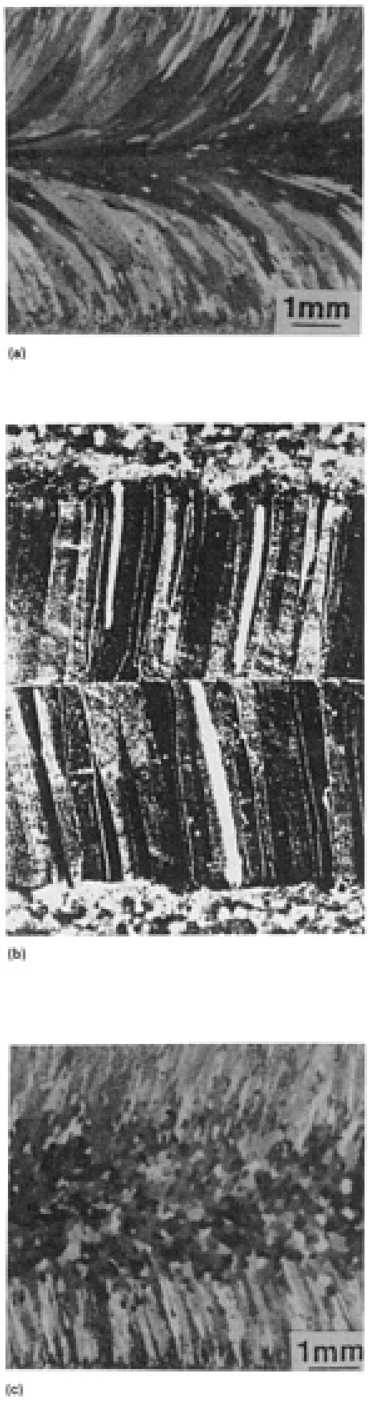

FIG. 14 ACTUAL WELD MICROSTRUCTURES CORRESPONDING TO SCHEMATICS OF FIG. 13. (A) GRAIN STRUCTURE OF A GTAW OF ALUMINUM ALLOY 6061, MADE WITH *Q* OF 700 W (200 BTU/H) AND *V* OF 5.1 MM/S (0.20 IN./S). (B) GRAIN STRUCTURE OF A GTAW OF PURE ALUMINUM, MADE WITH *V* OF 20.8 MM/S (0.819 IN./S). (C) GRAIN STRUCTURE OF A GTAW OF ALUMINUM ALLOY 6061, MADE WITH A HIGHER *Q* THAN THAT OF (A) (1320 W, OR 385 BTU/H) AND *V* OF 12.7 MM/S (0.500 IN./S). GTAW, GAS-TUNGSTEN ARC WELD. SOURCE: REF 13, 15

References cited in this section

7. F. MATSUDA, T. HASHIMOTO, AND T. SENDA, *TRANS. NATL. RES. INST. MET. (JPN),* VOL 11, 1969, P 43
8. W.F. SAVAGE, C.D. LUNDIN, AND A.H. ARONSON, *WELD. J.,* VOL 44, 1965, P 420S-425S

9. W.F. SAVAGE, *WELD. WORLD,* VOL 18, 1980, P 89-113
10. W.F. SAVAGE, C.D. LUNDIN, AND R.J. HRUBEC, *WELD. J.*, VOL 47, 1968, P 420S-425S
11. W.F. SAVAGE, C.D. LUNDIN, AND T.F. CHASE, *WELD. J.,* VOL 47, 1968, P 522S-526S
12. W.F. SAVAGE, E.F. NIPPES, AND J.S. ERICKSON, *WELD. J.*, VOL 55, 1976, P 213S-221S
13. S. KOU AND L. LE, *METALL. TRANS. A*, VOL 19, 1988, P 1075-1082
14. T. GANAHA, B.P. PEARCE, AND H.W. KERR, *METALL. TRANS. A*, VOL 11, 1980, P 1351-1359
15. H. NAKAGAWA, M. KATOH, F. MATSUDA, AND T. SENDA, *TRANS. JPN. WELD. SOC.*, VOL 4, 1973, P 11

Fundamentals of Weld Solidification

Harvey D. Solomon, General Electric Company

Nonequilibrium Effects: High-Rate Weld Solidification and Composition Banding

As has been noted, while the constitutional supercooling model (described in previous sections of this article) gives valuable insights into the establishment of nonplanar structures, it does not necessarily give an exact depiction of the composition gradient develop within the dendrites or cells that may form and within the interdendritic or intercellular regions. This has been shown by Brooks and Baskes (Ref 16) in their study of gas-tungsten arc and electron-beam aluminum-copper and iron-niobium welds. For the aluminum-copper welds, the cell core compositions were about 2 to 3 times the KC_0 value expected for the first material to solidify (see Fig. 5) and almost 4 times KC_0 for the iron-niobium welds. This discrepancy was ascribed to the undercooling because of the curvature at the cell tip (see Eq 4), which is not considered in the constitutional supercooling model. Tip undercooling decreases the temperature at which the solidification occurs and, as Fig. 3 shows, increases the composition of the solid that is forming.

No initial transient of the sort shown in Fig. 5 was observed. This was explained by using the lower solidification rate at the side of a cell, rather than that at the dendrite tip in the calculation of the size of the initial transient. This increases the length of the transient (which is inversely proportional to the solidification rate, R) and eliminates the sharp composition variation of the initial transient shown in Fig. 5. Brooks and Baskes also incorporated postsolidification solid-state diffusion of solute (following the approach of Brody and Flemings, which is given in Ref 17) to explain the composition gradients that they measured. Such diffusion is also not considered in the conventional constitutional supercooling model.

In addition, the classical theory of constitutional supercooling, as described by Eq 10, 11, and 12, does not take into account the temperature gradient that develops in the solid, the thermal conductivity of the solid and the liquid, and, most importantly, the surface tension of the solid-liquid interface. These factors are considered in interface stability models (Ref 18, 19, 20, 21, 22, 23, 24, 25, 26), which describe the breakdown of a planar solid-liquid interface in terms of the stability of nonplanar perturbations. The importance of these models is that they explain the observed transition from nonplanar solidification *back* to planar solidification when the solidification rate is increased to very high levels.

As has already been shown, at low solidification rates there can be a transition from planar to nonplanar solidification, as described by constitutional supercooling (as shown in Fig. 8). As the solidification rate is increased, the radius of curvature and spacing of the dendrite arms will decrease, as does the cell spacing. What eventually occurs, at very high solidification rates, is that the very small radius that is required is not supportable, because of the increased surface energy. The stability models predict that there will be a solidification rate (the limiting stability rate) beyond which there will be a transition *back* to planar solidification. Figure 15 shows the calculated dendrite tip radius as a function of velocity of the solid-liquid interface. The increase in R at high rates defines the limiting stability rate.

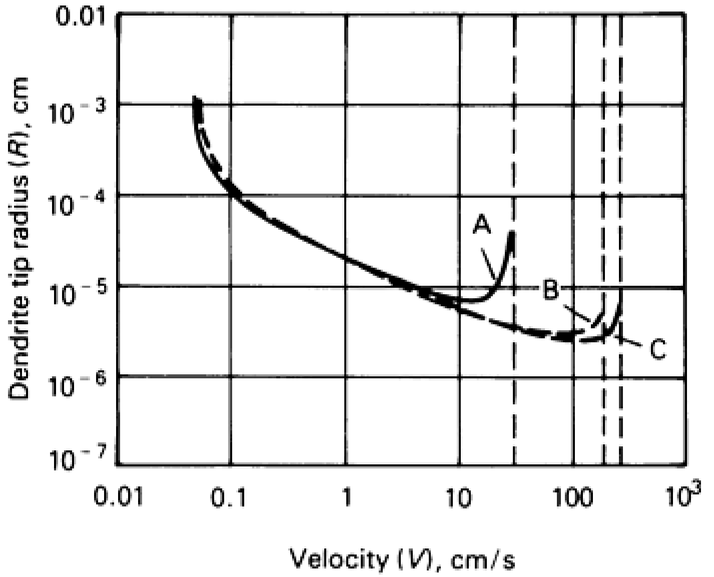

FIG. 15 PLOT OF DENDRITE TIP RADIUS, *R*, VERSUS VELOCITY, *V*, OF THE SOLID-LIQUID INTERFACE FOR AN AG-5% CU COMPOSITION WITH TEMPERATURE GRADIENT, *G*, OF 10^5 K/CM (5×10^5 °F/IN). CURVE A HAS $D_L(T)$ AND $K(V)$ WITH K_E AND M_L AS CONSTANTS; CURVE B, D_L (T) WITH K_E AND M_L AS CONSTANTS; CURVE C, WITH D_L, K_E, AND M_L AS CONSTANTS. THE VELOCITY WHERE THE TIP RADIUS INCREASES IS THE STABILITY LIMIT. THE CALCULATIONS ARE MADE USING THE CONVENTIONAL STABILITY LIMIT APPROACH, WITH THE ADDITION OF A TEMPERATURE-DEPENDENT DIFFUSION COEFFICIENT, AND WITH THE ADDITION OF BOTH A TEMPERATURE-DEPENDENT DIFFUSION COEFFICIENT AND A VELOCITY-DEPENDENT PARTITION COEFFICIENT. SOURCE: REF 21

Each of the three curves in Fig. 15 illustrates different assumptions. The right-most curve shows the result using the approach of Mullins and Sekerka (Ref 18), who developed the original interface stability model. Here, the solute diffusion rate in the liquid, D_L, the equilibrium distribution coefficient, K, and the liquidus slope, M_L, are all assumed to be constants. The middle curve takes into account the fact that the diffusion coefficient is temperature dependent. (Increasing the degree of undercooling that is due to rapid solidification decreases both the solute diffusion rate and the limiting velocity.) The left-most curve not only considers the temperature dependence of the diffusion coefficient, but also takes into account a dependency of the distribution coefficient on the solidification rate.

The preceding discussions have utilized the distribution coefficient, K, determined from equilibrium phase diagrams. At high solidification rates, this equilibrium is not achieved at the solid-liquid interface. Rather, K is a function of the solidification rate. One of several different expressions for the variation of K with V is given by (Ref 23):

$$K = [K_E + (V A_0/D_L)] / [1 + (VA_0/D_L)] \qquad \textbf{(EQ 17)}$$

where K_e is the equilibrium distribution coefficient (K in Eq 8 and 12), and a_0 is a distance that is related to the interatomic spacing. Different models use different definitions for a_0 and different functional relationships, but give similar results. When $V > D/a_0$, K will approach 1. For $D = 5 \times 10^{-5}$ cm^2/s, or 5×10^{-8} ft^2/s (a typical value for liquid diffusion), and $a_0 = 5 \times 10^{-8}$ cm, or 2.0×10^{-8} in., K will approach 1 for $V > 10^3$ cm/s, or 4×10^2 in./s. With $K = 1$, there is no partition of the solute, no constitutional supercooling, and the solidification will be planar for a positive temperature gradient. Figure 15 shows that the limiting velocity, as predicted by the stability model, is quite a bit lower than 10^3 cm/s (4×10^2 in./s). Thus, the reestablishment of planar solidification is not just the result of K approaching unity. Rather, it is due to the lack of the stability of nonplanar protrusions, which is influenced by this change in K. This development of planar solidification is important because it can arise in electron-beam and laser welds, and because the development of planar solidification lessens the tendency for hot cracking.

Figure 15 shows the results of calculations, whereas Fig. 16 shows the results of actual observations made on silver-copper electron-beam scans (passes of the electron beam). These results are in general agreement with the predictions of Fig. 15. Figure 16 shows the compositional dependence of the transition to segregation-free planar solidification. It is analogous to Fig. 8, which describes the transition from planar to nonplanar solidification that occurs at lower solidification rates. Figure 16 also shows that just below the limiting solidification rate, a banded structure can develop. This banding develops at a constant solidification rate and is due to instabilities at the solid-liquid interface.

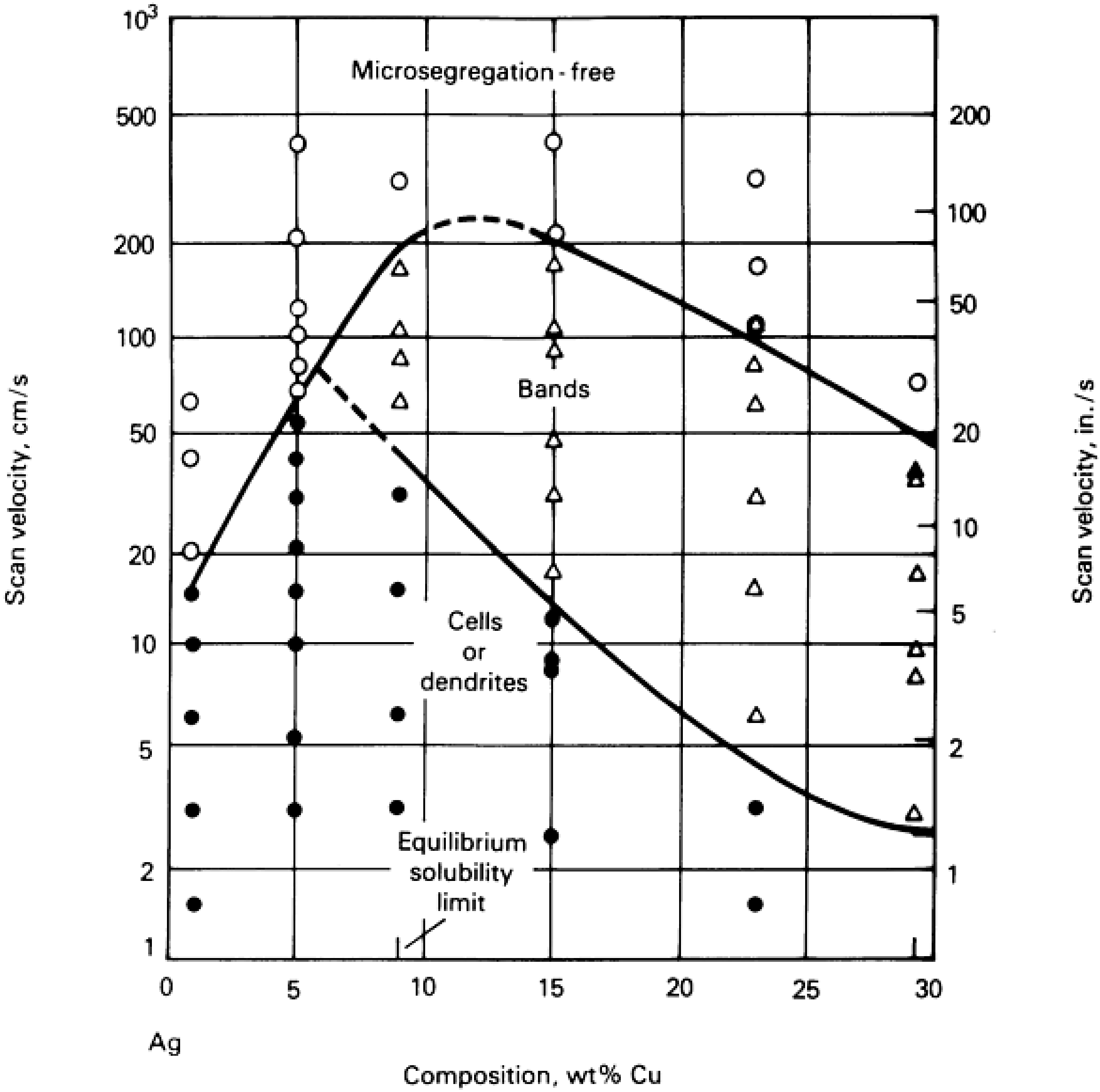

FIG. 16 OBSERVED MICROSTRUCTURES IN SILVER-COPPER ALLOYS AS A FUNCTION OF THE SCAN VELOCITY (BEAM TRANSLATION VELOCITY) FOR ELECTRON-BEAM WELDS. OPEN CIRCLES DENOTE MICROSEGREGATION-FREE STRUCTURE; CLOSED CIRCLES, A STRUCTURE CONSISTING OF CELLS OR DENDRITES; AND OPEN TRIANGLES, THE DEVELOPMENT OF A BANDED STRUCTURE. SOURCE: REF 24

A more common type of banding is associated with the ripples observed on the surface of most welds. Variations in the welding power will cause variations in the solidification rate. If the power decreases, then the solidification rate will increase. More solute will be incorporated into the solid and a solute-rich band will be created (for $K < 1$ and where M_L is negative). The reverse is true for an increase in the power (that is, the solidification rate will decrease), and this will cause a solute-poor band to form (for $K < 1$ and where M_L is negative). These power fluctuations can be deliberate, as in pulsed welding, or can result from unavoidable instabilities in the welding power supplies or fluctuations in primary voltage, mechanical (travel) motion, and so on. Solute banding and surface rippling are illustrated in Fig. 17.

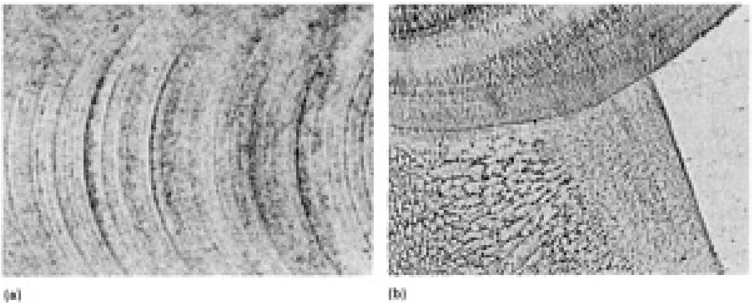

FIG. 17 SOLUTE BANDING AND SURFACE RIPPLING IN WELD BEADS. (A) SURFACE RIPPLES IN A TUNGSTEN-INERT GAS WELD MADE WITH AL-5SI FILLER. 6×. (B) SULFUR AND PHOSPHORUS SOLUTE BANDING AT THE FUSION BOUNDARY OF A MANUAL ARC WELD IN CARBON-MANGANESE STEEL. 26×. COURTESY OF THE WELDING INSTITUTE

Variations in the welding speed can also cause banding, as can the waves associated with the weld pool circulation, which can also influence the solidification rate. Special alternating magnetic fields have been used to supplement the surface tension forces, the usual electromagnetic forces, and the arc pressure and buoyancy forces, all of which affect the weld pool circulation (Ref 27). The aim is to induce interface fragmentation and cause grain refinement.

The solidified structure can also be influenced by several other factors, in addition to variations in the rate of movement of the solid-liquid interface (Ref 27). Inoculants have been used to nucleate equiaxed grains by heterogeneous nucleation on the inoculant particles. Ferro-titanium, ferro-niobium, titanium carbide, and titanium nitride have been used as grain refiners in steels. Zirconium and titanium have also been used as inoculants for aluminum. The effectiveness of inoculants depends not only on the type of inoculants used, but on the size of the inoculant particles, the method of introduction (where in the weld pool it is added), and welding conditions.

Arc motion and vibrations have also been used to suppress columnar growth and to produce a finer grain size (Ref 27). The effectiveness of arc vibration on the as-solidified grain size of an Al-2.5% Mg weld is illustrated in Fig. 18. The effectiveness of this approach depends on the frequency and amplitude of the arc motion, the heat input, arc length, and sheet thickness. In general, the greater the amplitude of vibration, the less the frequency required for grain refinement (Ref 28).

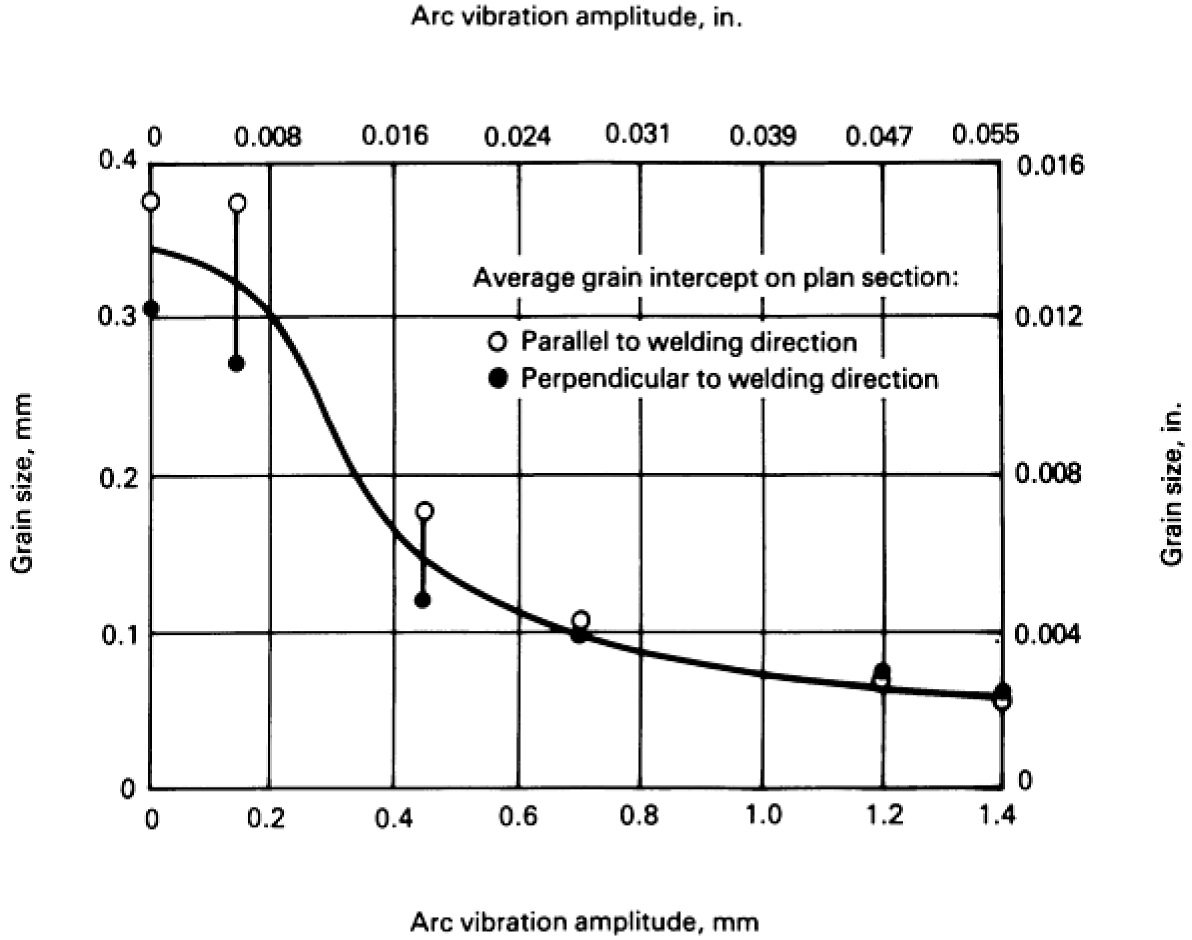

FIG. 18 PLOT OF GRAIN SIZE VERSUS ARC VIBRATION AMPLITUDE FOR AL-2.5MG WELD. WELD PARAMETERS: WELDING CURRENT, 150 A; WELDING SPEED, 200 MM/MIN (50 IN./MIN); ARC GAP, 2.4 MM (0.094 IN.); ARC VIBRATION FREQUENCY, 10 HZ. SOURCE: REF 27

Campbell (Ref 28) has developed frequency-amplitude maps that define several physical effects associated with the use of vibrations (that is, peak pressures generated, development of shearing stresses, power densities, laminar flow around dendrite arms, and the development of standing waves). These factors affect microstructural features, such as grain size, amount of microsegregation, and the development of porosity. For a given frequency of vibration, increasing the amplitude of vibration can cause a refinement of the grain size (Fig. 18), reduce microsegregation, and, unfortunately, increase the degree of porosity that is caused by vibration-induced cavitation. Vibrations can also cause splatter of the liquid weld metal; coarsen intragranular features, such as the secondary dendrite arm spacing; and promote hot tearing. Great care must therefore be taken when using vibratory motion and when determining the optimum amplitude and frequency of vibration.

Variations in the arc current can also help to refine the weld microstructure. Ultrasonic vibration of the weld pool has also been found to be effective in some cases. These vibrations have been added through the filler wire and through the use of water-cooled copper probes (in large weld pools, such as those developed in electro-slag welding). In some cases, variations in the nature of the weld pool circulation have also been found to be effective in refining the weld structure. The desire to produce a fine-grained equiaxed weld structure is driven by the improved mechanical properties and lessened hot cracking that this structure provides.

Torch weaving (Ref 29) (that is, moving the arc in a sinusoidal rather than linear fashion) has been used to break up the linear nature of the weld and to prevent hot cracks from propagating along the weld. A weld made with torch weaving is shown in Fig. 19. This weaved weld structure prevents cracks from running the length of the weld, but adds to the cost of the weld by decreasing the linear travel speed.

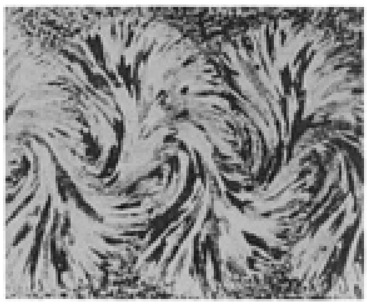

FIG. 19 TYPICAL GRAIN STRUCTURE OBTAINED WHEN TORCH WEAVING IS USED TO WELD ALUMINUM ALLOY 2014. WELD PARAMETERS: TRANSVERSE ARC FREQUENCY OSCILLATION, 1.0 HZ; AMPLITUDE OF OSCILLATION, 1.5 MM (0.059 IN.). SOURCE: REF 13, 30

Our understanding of weld solidification owes much to the more extensive study of the solidification of castings. The reader should refer to that subject area for a more in-depth treatment of solidification in general. However, it should be remembered that the boundary conditions for weld solidification and for that of a casting are not the same. Therefore, care must be taken in applying all that has been learned from casting metallurgy to the metallurgy of welds.

References cited in this section

13. S. KOU AND L. LE, *METALL. TRANS. A,* VOL 19, 1988, P 1075-1082
16. J.A. BROOKS AND M.I. BASKES, WELD MICROSTRUCTURE CHARACTERIZATION AND MODELING, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID AND J.M. VITEK, ED., ASM INTERNATIONAL, 1990, P 93-99
17. H.D. BRODY AND M.C. FLEMINGS, *TRANS. AIME,* VOL 236, 1966, P 615-624
18. W.W. MULLINS AND R.F. SEKERKA, *J. APPL. PHYS.,* VOL 35, 1964, P 444-451
19. J.S. LANGER AND H. MULLER-KRUMBHAAR, *ACTA METALL.,* VOL 26, 1978, P 1681-1687
20. W. KURZ AND D.J. FISCHER, *ACTA METALL.,* VOL 29, 1981, P 11-20
21. W. KURZ, B. GIOVANOLA, AND R. TRIVEDI, *ACTA METALL.,* VOL 34, 1986, P 823-830
22. R. TRIVEDI AND W. KURZ, *ACTA METALL.,* VOL 34, 1986, P 1663-1670
23. M.J. AZIZ, *J. APPL. PHYS.,* VOL 53, 1982, P 1158-1168
24. W.J. BOETTINGER, D. SHECHTMAN, R.J. SCHAEFFER, AND BIANCANIELLO, *METALL. TRANS. A,* VOL 15, 1984, P 55-66
25. W.J. BOETTINGER AND S.R. CORIELL, *MATER. SCI. ENG.,* VOL 65, 1984, P 27-36
26. M. CARRARD, M. GREMAUD, M. ZIMMERMANN, AND W. KURZ, *ACTA METALL.,* VOL 40, 1992, P 983-996
27. G.J. DAVIES AND J.G. GARLAND, *INT. MET. REV.,* VOL 20, 1975, P 83-106
28. J. CAMPBELL, *INT. MET. REV.,* VOL 26 (NO. 2), 1981, P 71-108
29. S. KOU, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID AND J.M. VITEK, ED., ASM INTERNATIONAL, 1989, P 137-147
30. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987, P 161-162

Fundamentals of Weld Solidification

Harvey D. Solomon, General Electric Company

References

1. J.D VERHOEVEN, *FUNDAMENTALS OF PHYSICAL METALLURGY,* JOHN WILEY & SONS, 1975
2. H.D. SOLOMON, IN *TREATISE ON MATERIALS SCIENCE AND TECHNOLOGY,* VOL 25, *EMBRITTLEMENT OF ENGINEERING ALLOYS*, C.L. BRIANT AND S.K. BANERJI, ED., ACADEMIC PRESS INC., 1983, P 525-599
3. J.A. BURTON, R.C. PRIM, AND J. SLICHTER, *J. CHEM. PHYS.,* VOL 21, 1953, P 1987-1991
4. W.A. TILLER, K.A. JACKSON, J.W. RUTTER, AND B. CHALMERS, *ACTA METALL.,* VOL 1, 1953, P 428-437
5. W.A. TILLER AND J.W. RUTTER, *CAN. J. PHYS.,* VOL 34, 1956, P 96-121
6. W.F. SAVAGE, E.F. NIPPES, AND T.W. MILLER, *WELD. J.,* VOL 55, 1976, P 165S-173S
7. F. MATSUDA, T. HASHIMOTO, AND T. SENDA, *TRANS. NATL. RES. INST. MET. (JPN),* VOL 11, 1969, P 43
8. W.F. SAVAGE, C.D. LUNDIN, AND A.H. ARONSON, *WELD. J.,* VOL 44, 1965, P 420S-425S
9. W.F. SAVAGE, *WELD. WORLD,* VOL 18, 1980, P 89-113
10. W.F. SAVAGE, C.D. LUNDIN, AND R.J. HRUBEC, *WELD. J.,* VOL 47, 1968, P 420S-425S
11. W.F. SAVAGE, C.D. LUNDIN, AND T.F. CHASE, *WELD. J.,* VOL 47, 1968, P 522S-526S
12. W.F. SAVAGE, E.F. NIPPES, AND J.S. ERICKSON, *WELD. J.,* VOL 55, 1976, P 213S-221S
13. S. KOU AND L. LE, *METALL. TRANS. A,* VOL 19, 1988, P 1075-1082
14. T. GANAHA, B.P. PEARCE, AND H.W. KERR, *METALL. TRANS. A,* VOL 11, 1980, P 1351-1359
15. H. NAKAGAWA, M. KATOH, F. MATSUDA, AND T. SENDA, *TRANS. JPN. WELD. SOC.,* VOL 4, 1973, P 11
16. J.A. BROOKS AND M.I. BASKES, WELD MICROSTRUCTURE CHARACTERIZATION AND MODELING, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID AND J.M. VITEK, ED., ASM INTERNATIONAL, 1990, P 93-99
17. H.D. BRODY AND M.C. FLEMINGS, *TRANS. AIME,* VOL 236, 1966, P 615-624
18. W.W. MULLINS AND R.F. SEKERKA, *J. APPL. PHYS.,* VOL 35, 1964, P 444-451
19. J.S. LANGER AND H. MULLER-KRUMBHAAR, *ACTA METALL.,* VOL 26, 1978, P 1681-1687
20. W. KURZ AND D.J. FISCHER, *ACTA METALL.,* VOL 29, 1981, P 11-20
21. W. KURZ, B. GIOVANOLA, AND R. TRIVEDI, *ACTA METALL.,* VOL 34, 1986, P 823-830
22. R. TRIVEDI AND W. KURZ, *ACTA METALL.,* VOL 34, 1986, P 1663-1670
23. M.J. AZIZ, *J. APPL. PHYS.,* VOL 53, 1982, P 1158-1168
24. W.J. BOETTINGER, D. SHECHTMAN, R.J. SCHAEFFER, AND BIANCANIELLO, *METALL. TRANS. A,* VOL 15, 1984, P 55-66
25. W.J. BOETTINGER AND S.R. CORIELL, *MATER. SCI. ENG.,* VOL 65, 1984, P 27-36
26. M. CARRARD, M. GREMAUD, M. ZIMMERMANN, AND W. KURZ, *ACTA METALL.,* VOL 40, 1992, P 983-996
27. G.J. DAVIES AND J.G. GARLAND, *INT. MET. REV.,* VOL 20, 1975, P 83-106
28. J. CAMPBELL, *INT. MET. REV.,* VOL 26 (NO. 2), 1981, P 71-108
29. S. KOU, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID AND J.M. VITEK, ED., ASM INTERNATIONAL, 1989, P 137-147
30. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987, P 161-162

Fundamentals of Weld Solidification

Harvey D. Solomon, General Electric Company

Selected References

- *CASTING,* VOL 15, *ASM HANDBOOK,* ASM INTERNATIONAL, 1988, P 100-181

- S.A. DAVID AND J.M. VITEK, *INT. MATER. REV.*, VOL 34, 1989, P 213-245
- J.H. DEVLETION AND W.E. WOOD, VOL 6, *METALS HANDBOOK*, 9TH ED., AMERICAN SOCIETY FOR METALS, 1983, P 21-49
- K. EASTERLING, *INTRODUCTION TO THE PHYSICAL METALLURGY OF WELDING,* BUTTERWORTHS, 1985
- M.C. FLEMINGS, *SOLIDIFICATION PROCESSING,* MCGRAW-HILL, 1974
- W. KURZ AND D.J. FISHER, *FUNDAMENTALS OF SOLIDIFICATION,* 3RD ED., TRANS TECH, 1989
- J.F. LANCASTER, *METALLURGY OF WELDING,* 3RD ED., GEORGE ALLEN & UNWIN, 1980
- M. RAPPAZ, *INT. MATER. REV.,* VOL 34, 1989, P 93-123

Nature and Behavior of Fluxes Used for Welding

D.L. Olson, S. Liu, R.H. Frost, G.R. Edwards, and D.A. Fleming, Colorado School of Mines

---

Introduction

FLUXES are added to the welding environment to improve arc stability, to provide a slag, to add alloying elements, and to refine the weld pool (Ref 1, 2). Different ingredients in the flux system will provide the process with different pyrometallurgical characteristics and thus different weld metal properties (Ref 3, 4).

The slag that forms during welding covers the hot weld metal and protects it from the atmosphere. Welding slag consists of the glass-forming components of the flux, as well as inclusions that form in the weld pool, coalesce, rise, and become incorporated into the slag.

The need to improve flux formulation to achieve optimal weld metal composition, and ultimately improve the properties of weldments, has led to fundamental studies of weld pyrochemistry. Understanding the thermodynamic and kinetic factors that are prevalent at the electrode, in the arc column, and in the weld pool, has led to more precise prediction of the final weld metal composition (Ref 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15).

References

1. G.E. LINNERT, CHAPTER 8, *WELDING METALLURGY*, VOL 1, AWS, 1965, P 367-396
2. C.E. JACKSON, FLUXES AND SLAGS IN WELDING, *WELD. RES. BULL*., NO. 190, 1973
3. T. LAU, G.C. WEATHERLY, AND A. MCLEAN, THE SOURCES OF OXYGEN AND NITROGEN CONTAMINATION IN SUBMERGED ARC WELDING USING CAO-AL_2O_3 BASED FLUXES, *WELD. J*., VOL 64 (NO. 12), 1985, P 343S-347S
4. T.H. NORTH, H.B. BELL, A. NOWICKI, AND I. CRAIG, SLAG/METAL INTERACTION, OXYGEN, AND TOUGHNESS IN SUBMERGED ARC WELDING, *WELD. J*., VOL 57 (NO. 3), 1978, P 63S-75S
5. N. CHRISTENSEN AND J. CHIPMAN, SLAG-METAL INTERACTION IN ARC WELDING, *WELD. RES. BULL*., NO. 15, JAN 1953, P 1-14
6. R.H. FROST, D.L. OLSON, AND S. LIU, PYROCHEMICAL EVALUATION OF WELD METAL INCLUSION EVOLUTION, *PROC. 3RD INT. CONF. TRENDS IN WELDING*, ASM INTERNATIONAL, JUNE 1992
7. C.A. NATALIE, D.L. OLSON, AND M. BLANDER, PHYSICAL AND CHEMICAL BEHAVIOR OF WELDING FLUXES, *ANN. REV. MATER. SCI*., VOL 16, 1986, P 389-413
8. J.E. INDACOCHEA, M. BLANDER, N. CHRISTENSEN, AND D.L. OLSON, CHEMICAL REACTIONS WITH FEO-MNO-SIO_2 *FLUXES, METALL. TRANS. B*, VOL 16, 1985, P 237-245
9. U. MITRA AND T.W. EAGAR, SLAG-METAL REACTIONS DURING WELDING, *METALL. TRANS. B*, VOL 22, 1991, P 65-100

10. C.S. CHAI AND T.W. EAGAR, SLAG-METAL EQUILIBRIUM DURING SUBMERGED ARC WELDING,*METALL. TRANS. B*, VOL 12, 1981, P 539-547
11. N. CHRISTENSEN, METALLURGICAL ASPECTS OF ARC WELDING, *WELD. J.*, VOL 27, 1949, P 373S-380S
12. O. GRONG, D.L. OLSON, AND N. CHRISTENSEN, CARBON OXIDATION IN HYPERBARIC MMA WELDING, *MET. CONSTRUCT.*, VOL 17. DEC 1985, P 810R-814R
13. U. MITRA AND T.W. EAGAR, SLAG-METAL REACTIONS DURING SUBMERGED ARC WELDING OF ALLOY STEELS, *METALL. TRANS. A*, VOL 15, 1984, P 217-227
14. T.W. EAGAR, SOURCES OF WELD METAL OXYGEN CONTAMINATION DURING SUBMERGED ARC WELDING, *WELD. J.*,VOL 57, 1978, P 76S-80S
15. N. CHRISTENSEN, *WELDING METALLURGY, LECTURE NOTES*, NTH, 1979

Nature and Behavior of Fluxes Used for Welding

D.L. Olson, S. Liu, R.H. Frost, G.R. Edwards, and D.A. Fleming, Colorado School of Mines

---

Equilibrium Parameters

Equilibrium is not achieved during welding, because of the very large temperature and density gradients, the short reaction times, and the large electric currents. Despite these expected departures from equilibrium, thermodynamic considerations can be used as a guide for constraining chemical reactions and mechanisms involved in welding (Ref 5, 7, 9). A common approach is to assume that extremely high temperatures and high surface-to-volume ratios allow thermodynamic equilibrium to be locally attained, in spite of the short reaction times available.

Further complicating the issue is the dependence on welding parameters of the chemical partitioning between the slag and the weld metal (Ref 3, 7). This dependence suggests that the pyrometallurgical reactions involved are influenced by the processes that occur at the electrode tip. As a specific example, the welding parameters affect metal-droplet size, which in turn alters the chemical kinetics. Nonetheless, thermodynamics will reveal the direction taken by the chemical reactions (but will not accurately predict weld metal composition).

The effective slag-metal reaction temperature has been estimated (by pyrochemical analysis of slag-metal compositions) to be approximately 1900 °C (3450 °F), a temperature intermediate between that of the hot spot at the arc root (2300 °C or 4170 °F) and the melting point of iron (1500 °C or 2730 °F). Thus, during the droplet lifetime, the average temperature is effectively 1900 °C (2730 °F). During this period, oxygen from the hot spot reactions is distributed throughout the droplet, and that oxygen reacts with the metallic elements to form oxides; these products of oxidation pass into the slag, and the slag-metal reactions tend toward equilibrium. Estimates of the time for which the molten slag and molten metal are in contact range from 3 to 8 s. During the process, the gaseous phase in the arc cavity contacts the metal for an estimated 0.5 to 1.0 s.

Effect of Oxygen

The most important chemical reagent in controlling weld metal composition, and thus microstructure and properties, is oxygen (Ref 3, 4, 14, 16, 17, 18, 19). Oxygen directly reacts with alloying elements to alter their effective role by:

- REDUCING HARDENABILITY
- PROMOTING POROSITY
- PRODUCING INCLUSIONS.

All three effects are significant to weld quality.

Oxygen is introduced into the weld pool at high temperatures by:

- THE PRESENCE OF OXIDE FLUXES THAT DISSOCIATE IN THE ARC
- THE SLAG-METAL REACTIONS IN THE WELD POOL
- THE OXIDES ON THE SURFACE OF BAKED METALLIC POWDERS MIXED WITH FLUX OR ON ELECTRODE
- THE ASPIRATION OF ATMOSPHERE (AIR) INTO THE ARC.

Shielding gas may contain oxidizing reagents; a common gas is 75Ar-25CO_2 or 100% CO_2. As a 100 to 1000 ppm, depending on the type of welding consumable used.

The weld metal oxygen measured directly at the molten electrode tip has been reported to be as high as 1400 ppm. Individual droplets have been found to contain as much as 2000 ppm O_2. There are two views concerning the genesis of the high oxygen content. One suggests that pyrochemical or electrochemical reactions (or both) provide oxygen to the electrode tip, and then further oxidation occurs within the droplet as it passes through the arc. The other view is that the high oxygen levels of the droplet represent the maximum buildup of oxygen in the electrode prior to detachment, with limited reaction during flight across the arc to the weld pool. In either case, high oxygen concentrations are introduced into the weld pool by the welding process.

At the melting point, the solubility of oxygen in pure liquid iron is approximately 1600 ppm at 100 kPa (1 atm) pressure. During solidification, solubility decreases to about 860 ppm at 1500 °C (2730 °F) in δ-Fe. Most of the alloying elements present in liquid steel reduce oxygen solubility through deoxidation equilibria. Steelmaking processes typically yield analytical oxygen levels ranging from 70 to 100 ppm. Welds typically pick up oxygen to levels of several hundred ppm, then deoxidize to oxygen levels of around 200 to 300 ppm. Deoxidation of the weld metal occurs in two separate steps, the first being the primary deoxidation of the weld pool (Ref 6). Secondary deoxidation occurs during solidification as solute concentrations increase within the intercellular or interdendritic regions. The secondary deoxidation will either form very small inclusions or will coat the interdendritically trapped primary inclusions.

The high oxygen concentrations added to the weld pool by the metal droplets significantly affect deoxidation. Figure 1 shows experimentally measured soluble oxygen concentrations for various deoxidants (solid lines), along with deoxidation curves predicted by the solubility products (broken lines). The experimental deviation is caused by interactions of the deoxidant with other alloy elements. If the oxygen and metallic element concentrations resulting from the welding process exceed the equilibrium concentration for a specific reaction, inclusions will result. The ability to form a specific inclusion will correspond directly to the position of the weld pool composition relative to the activity plot for this inclusion. Thus, the thermodynamic order for the formation of primary oxides would be: Al_2O_3 > Ti_2O_3 > SiO_2 > Cr_2O_4 > MnO.

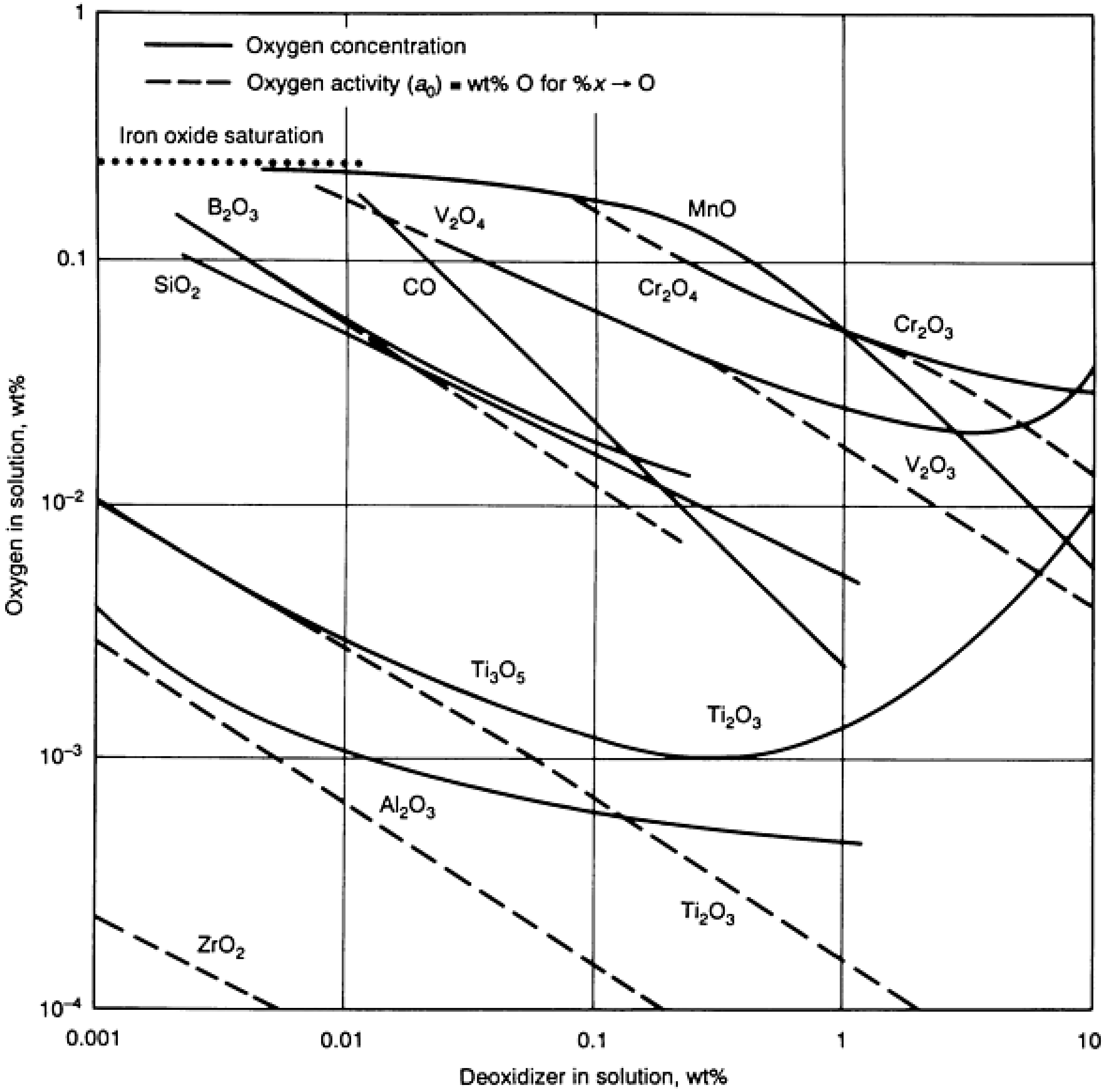

FIG. 1 DEOXIDATION EQUILIBRIA IN LIQUID IRON ALLOYS AT 1600 °C (2910 °F). THE BROKEN LINES SHOW DEOXIDATION EQUILIBRIA PREDICTED BY SOLUBILITY PRODUCT CALCULATIONS. THE SOLID LINES SHOW EXPERIMENTALLY DETERMINED SOLUBLE OXYGEN CONCENTRATIONS FOR VARIOUS DEOXIDANTS. THE EXPERIMENTAL DEVIATION IS CAUSED BY VARIATIONS IN THE ACTIVITY COEFFICIENTS WITH INCREASING DEOXIDANT CONCENTRATION. SOURCE: REF 20

Inclusion Formation. Inclusions form as a result of reactions between metallic alloy elements and nonmetallic tramp elements, or by mechanical entrapment of nonmetallic slag or refractory particles. Inclusions may include:

- OXIDES
- SULFIDES
- NITRIDES
- CARBIDES
- OTHER COMPOUNDS
- MULTIPLE PHASES

Among these, oxides and complex oxides occur most frequently in the size range known to influence weld metal microstructure.

Using only thermodynamic considerations in the analysis of slag-metal reactions, the following reactions may be considered as the ones that describe inclusion formation:

$$X\ \underline{M}\ + Y\ \underline{O}\ = (\mathrm{M}_X\mathrm{O}_Y) \qquad \textbf{(EQ 1)}$$

where the underlining of a component M ($\underline{M}$) and the component O ($\underline{O}$) indicate that M and O, respectively, are dissolved in the metal, and:

$$X\ \underline{M}\ + Y(\mathrm{FEO}) = (\mathrm{M}_X\mathrm{O}_Y) + Y\mathrm{FE} \qquad \textbf{(EQ 2)}$$

The equilibrium constants (K_1 and K_2) for the above two reactions are:

$$K_1 = \frac{(A_{M_xO_y})}{[a_M]^x[a_o]^y} \qquad \textbf{(EQ 3)}$$

where a_i is the activity, a function of the concentration, for component *i*, and:

$$K_2 = \frac{(A_{M_xO_y})}{[a_M]^x[a_{FeO}]^y} \qquad \textbf{(EQ 4)}$$

Equilibrium compositions for the weld deposit can be estimated with Eq 1, 2, 3, and 4 and can be used to predict trends for the weld pool pyrochemical reactions. The actual compositions may, however, differ from the calculated values due to alloying-element partitioning during cellular or dendritic solidification, commonly observed in steel weldments.

During solidification of a weld metal, solute elements segregate to the liquid at the solid/liquid interface, and the solute concentrations can reach high levels in the interdendritic regions, as suggested in Fig. 2. Neglecting solid diffusion, the solute composition in the liquid at the solid/liquid interface can be modeled (Ref 6) by the nonequilibrium lever rule, or Scheil Equation:

$$C_\mathrm{L} = \mathrm{C}_0 F_\mathrm{L}^{K\text{-}1} \qquad \textbf{(EQ 5)}$$

where C_0 is the bulk concentration in the weld pool, C_L is the solute concentration in the liquid at the interface, and *k* is the equilibrium partition ratio.

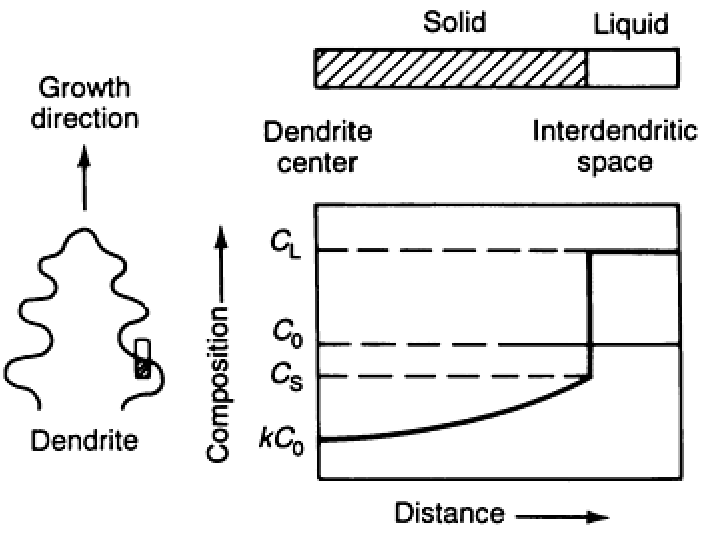

FIG. 2 SCHEMATIC SHOWING THE SOLID AND LIQUID COMPOSITION PROFILES MODELED BY EQ 5 BASED ON THE ASSUMPTION THAT THERE IS COMPLETE LIQUID DIFFUSION AND NO SOLID DIFFUSION. C_0, THE INITIAL ALLOY COMPOSITION; K, THE PARTITION RATIO OF THE SOLID TO LIQUID COMPOSITIONS ON THE EQUILIBRIUM PHASE DIAGRAM; KC_0, THE INITIAL COMPOSITION OF THE SOLID; AND C_S AND C_L, THE SOLID AND LIQUID COMPOSITIONS, RESPECTIVELY, AT THE SOLIDIFICATION INTERFACE. SOURCE: REF 4

The equilibrium partition ratio, *k*, controls the direction and the extent of segregation. For most alloy elements in steel, the partition ratio is less than one, and the element segregates to the interdendritic liquid.

Consider the deoxidation equilibrium as represented by the dissolution reaction for a complex oxide:

$$\mathrm{M}_X\mathrm{N}_Y\mathrm{O}_Z = X\ \underline{M} + Y\ \underline{N} + Z\ \underline{O} \qquad \textbf{(EQ 6)}$$

The free energy change associated with the dissolution of the $M_xN_yO_z$ can be written as:

$$\Delta G = \Delta G^\circ + RT \ln \frac{[M]^x[N]^y[O]^Z}{[M_xN_yO_z]} \qquad \textbf{(EQ 7)}$$

where [M], [N], and [O] are the solute activities in the liquid; and *x*, *y*, and *z* are the stoichiometric constants from Eq 6. The value $[M_xN_yO_z]$ is the activity for the specific inclusion and can be assumed to have the value of one.

The ratio of the activities of the reactants to that of the oxide can be termed the activity quotient (*Q*):

$$Q = \frac{[M]^x[N]^y[O]^Z}{[M_xN_yO_z]} \qquad \textbf{(EQ 8)}$$

With Eq 5 representing the extent of segregation, the solute activities in the interdendritic liquid can be written as:

$$[\mathrm{M}]=[\gamma_\mathrm{M}][\mathrm{M_O}]f_\mathrm{L}^{k_\mathrm{M}-1} \qquad \textbf{(EQ 9A)}$$

$$[\mathrm{N}]=[\gamma_\mathrm{N}][\mathrm{N_0}]f_\mathrm{L}^{k_\mathrm{N}-1} \qquad \textbf{(EQ 9B)}$$

$$[\mathrm{O}] = [\gamma_\mathrm{O}][\mathrm{O_0}]f_\mathrm{L}^{k_\mathrm{O}-1} \qquad \textbf{(EQ 9C)}$$

In Eq 9a, 9b, and 9c, $[\mathrm{M_O}]$, $[\mathrm{N_O}]$, and $[\mathrm{O_O}]$ are the bulk concentrations of M, N, and O in the melt; and $[\gamma_\mathrm{M}]$, $[\gamma_\mathrm{N}]$, and $[\gamma_\mathrm{O}]$ are the activity coefficients for the solutes.

Substituting Eq 9a, 9b, and 9c into Eq 8 gives the free energy as a function of the remaining liquid fraction:

$$\Delta G = \Delta G^\circ + RT \ln \frac{[\gamma_\mathrm{M}\mathrm{M_O}f_\mathrm{L}^{k\mathrm{M}-1}]^x[\gamma_\mathrm{N}\mathrm{N_O}f_\mathrm{L}^{k\mathrm{N}-1}]^y[\gamma_\mathrm{O}\mathrm{O_O}f_\mathrm{L}^{k\mathrm{O}-1}]^z}{[\mathrm{M}_x\mathrm{N}_y\mathrm{O}_z]} \qquad \textbf{(EQ 10)}$$

Equation 10 expresses the free energy driving force for oxide dissolution. The first term on the right side, ΔG, represents equilibrium conditions. The second term represents the departure from equilibrium caused by changes in reactant concentrations or by segregation during solidification.

At equilibrium, the free energy driving force, ΔG, is zero, and the equilibrium concentrations can be found by equating the two terms on the right-hand side of Eq 10. At equilibrium, the activity quotient becomes the equilibrium constant, K_eq:

$$\Delta G^\circ = -RT \ln \frac{[\mathrm{M}]^x[\mathrm{N}]^y[\mathrm{O}]^z}{[\mathrm{M}_x\mathrm{N}_y\mathrm{O}_z]} = -RT \ln K_{\mathrm{eq}} \qquad \textbf{(EQ 11)}$$

Inclusion precipitation is possible when the concentrations of oxygen and deoxidants exceed the equilibrium values for a particular oxide. This condition can be expressed with the ratio of the activity quotient to the equilibrium coefficient:

$$\text{Precipitation index} = \frac{Q}{K_{\mathrm{eq}}} \qquad \textbf{(EQ 12)}$$

A precipitation index less than unity indicates that the concentrations of oxygen and deoxidants are below the equilibrium value, and the precipitation of the inclusions will not occur. A value greater than 1.0 indicates that concentrations are sufficiently high for precipitation according to the methodology described above. The compounds Al_2O_3, Ti_2O_3, and SiO_2 are some of the oxides that will form in a low-carbon low-alloy steel weld. Multiple reactions can occur, and different oxides may appear in the same weld when more than one of these deoxidizers is present.

Metal Transferability During Pyrochemical Reactions. The final weld metal concentration for a particular element is made up of contributions from the filler wire, flux, and base metal; however, losses caused by the welding process vary for each element.

Delta Quantity. The nominal composition of each weld can be calculated considering just the dilution effect of the filler wire and base metal. The extent of loss or transfer of a specific element can be evaluated by a quantity, which expresses the difference between the analytical and the nominal composition. These quantities, designated delta quantity in this article, indicate the effect of the flux on element transfer during welding. A positive delta quantity indicates an elemental transfer from the flux to the weld metal. A negative delta quantity suggests an elemental loss from the weld pool to the slag. A null delta quantity for a specific element suggests an apparent equilibrium condition, in which the flux and slag content for that element are the same. Flux compositions with null delta-quantity behavior have been used to make equilibrium calculations and thus achieve a better understanding of the chemical reactions involved in welding (Ref 8). Investigators have also quantified elemental transfer by measuring similar neutral points (null delta quantity) for various flux systems and have developed a thermodynamic model capable of predicting neutral points for some slag systems (Ref 10).

Arc Stabilizers. Arc welding fluxes are compositionally more complex than fluxes used in other metallurgical processes, such as steelmaking. Many of the additions to the flux are not designed to assist weld metal refinement. Some additions are present to promote arc stability, generate plasma and protective gases, control viscosity, support out-of-position welding, and promote slag detachability.

The welding arc requires an inert or chemically reducing plasma and shielding gas that can be easily ionized. Additions must be made to the flux to achieve the necessary current-carrying capacity and to maintain arc stability. Specific additions will be necessary for the various modes of current (direct current, dc, or alternating current, ac) and polarity.

Alkali metal, zirconium, and titanium additions to the arc affect the ionization process and the ease with which the welding is reinitiated (reinitiation is required 50 to 60 times per second with ac welding). These additions come to the welding flux as feldspar (alkali aluminum silicates), rutile, lithium carbonate, titanium aluminate, and potassium oxalate, and they play an especially important role when welding either in the dc electrode negative (DCEN) mode or in the alternate current (ac) mode. Arc stabilizers are also important in high-speed welding when the cathode and anode spots are less stable. These additions are part of the AWS classification of electrodes, as seen by the fourth digit (E XXX *X*) in the classification standard for steel electrodes for shielded metal arc welding (SMAW) given in Table 1. The electrodes designed to perform with ac mode or in the DCEN mode current have been specified as containing high titania or alkali metal (potassium or sodium) additions. Special additions such as Li_2O have been used to achieve multipurpose results--for example, reducing the viscosity while increasing the arc stability.

TABLE 1 AWS CLASSIFICATION OF SELECTED SMAW ELECTRODES FOR WELDING MILD AND LOW-ALLOY STEELS

ELECTRODE DESIGNATION	**CURRENT AND POLARITY[(A)]**	**PENETRATION**	**ARC STABILIZERS**	**% FE**
EXX10	DCEP	DEEP	HIGH CELLULOSE-NA	0-10
EXXX1	AC, DCEP	DEEP	HIGH CELLULOSE-K	0
EXXX2	AC, DCEN	MEDIUM	HIGH RUTILE-NA	0-10
EXXX3	AC, DCEP, DCEN	LIGHT	HIGH RUTILE-K	0-10
EXXX4	AC, DCEP, DCEN	LIGHT	RUTILE-IRON POWDER	25-40
EXX24	AC, DCEP, DCEN	LIGHT	RUTILE-IRON POWDER	50
EXXX5	DCEP	MEDIUM	LOW H-NA	0
EXXX6	AC, DCEP	MEDIUM	LOW H-K	0
EXX27	AC, DCEN, DCEP	MEDIUM	IRON OXIDE-IRON POWDER	50
EXXX8	AC, DCEP	MEDIUM	LOW HYDROGEN-IRON POWDER	25-40
EXX28	AC, DCEP	MEDIUM	LOW HYDROGEN-IRON POWDER	50

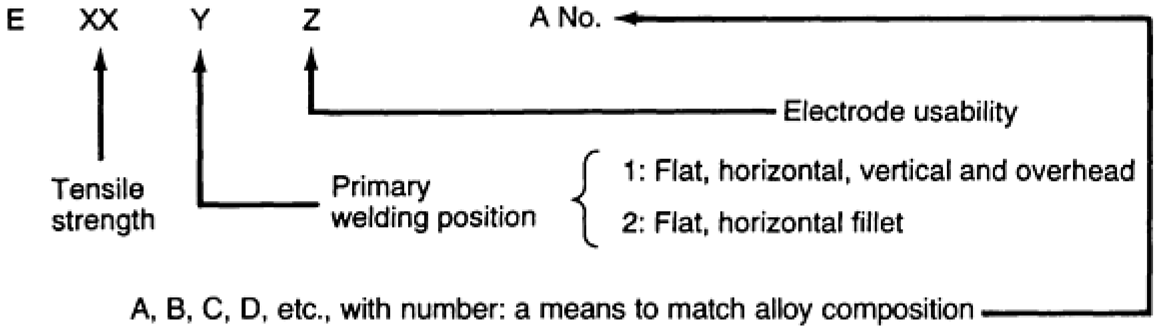
E
XX
Y
Z
A No.
Electrode usability
Tensile strength
Primary welding position
1: Flat, horizontal, vertical and overhead
2: Flat, horizontal fillet
A, B, C, D, etc., with number: a means to match alloy composition

(A) DCEP, DC ELECTRODE POSITIVE, DCEN, DC ELECTRODE NEGATIVE

Changes in Flux Composition With Delta Quantity. Figure 3 shows a transfer of manganese for welds made with SiO_2-TiO_2-CaO-1Na_2O flux at constant SiO_2 content in the flux. These values vary widely, depending on other alloy concentrations. The data imply that manganese is almost always lost to the slag, and that the activity of manganese varies considerably with changes in the amount of titania in the flux. Manganese is very important to weld metal hardenability and must be closely controlled to obtain the optimum weld microstructure. Controlling the weld metal manganese concentration in the titania-containing flux systems would require strict compositional control of the welding flux to ensure the correct manganese concentration in the weld metal.

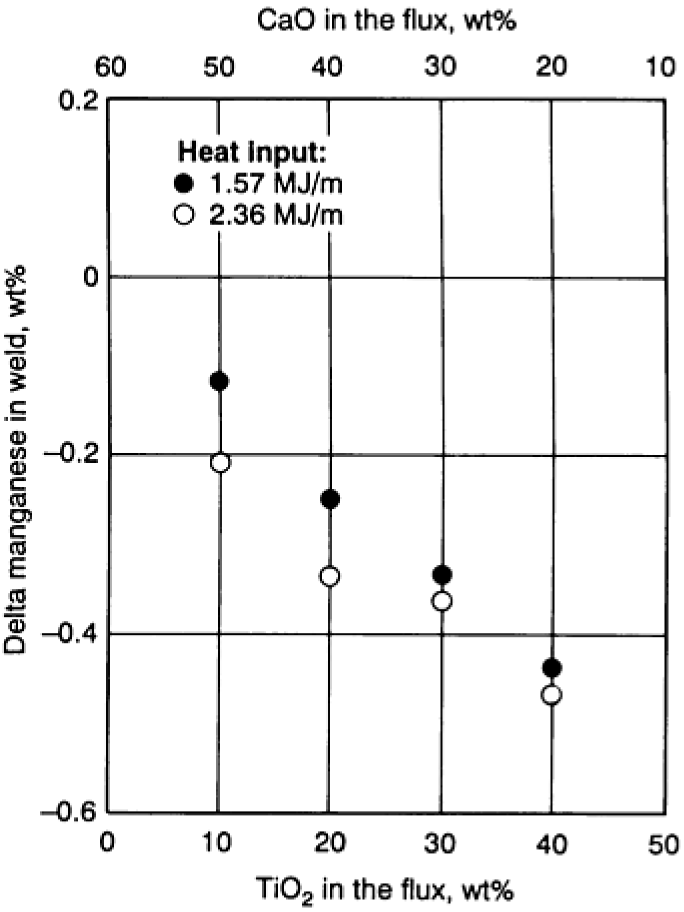

FIG. 3 PLOT OF CHANGES IN MANGANESE CONTENT IN THE WELD VERSUS THE VARIATION IN FLUX COMPOSITION AS A FUNCTION OF TWO-HEAT INPUTS. THE SIO_2 CONTENT WAS MAINTAINED AT 40% THROUGHOUT THE FLUX-CORED ARC WELDING PROCESS. SOURCE: REF 21

In the same system, the delta titanium showed a constant increase with increasing titania content of the flux (Fig. 4). Because of the large amounts of titania in the flux, it is not surprising that the welds show positive compositional deviations for titanium. These results also suggest that control of the weld metal titanium content will require tight control of the limits of the TiO_2 content of the flux.

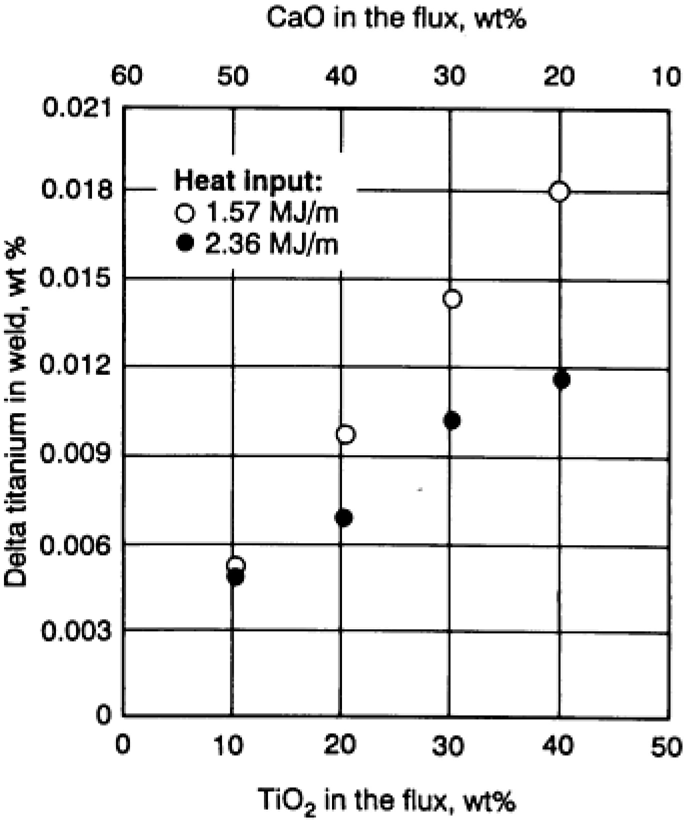

FIG. 4 PLOT OF CHANGES IN TITANIUM CONTENT IN THE WELD VERSUS THE VARIATION IN FLUX COMPOSITION AS A FUNCTION OF TWO HEAT INPUTS. THE SIO_2 CONTENT WAS MAINTAINED AT 40% THROUGHOUT THE FLUX-CORED ARC WELDING PROCESS. SOURCE: REF 21

If the delta quantity changes rapidly with flux composition, it may be difficult to maintain specific weld metal composition, microstructure, and properties with variation in flux composition. The magnitude of the delta quantity for specific elements is most often not as serious a concern as a rapid change in the delta quantity with variations in flux chemistry and welding parameters. The magnitude of the delta quantity can be adjusted by altering the alloy content of the welding wire or the amount of ferro-additions to the flux. Thus, a suitable combination of wire, flux, and welding parameters should achieve negligibly small delta quantities of major alloying elements.

Shielding Gas. When the shielding gas that protects the weld pool comes from the flux, it is necessary to understand the decomposition of specific flux components. Two common shielding gas atmospheres from flux dissociation are hydrogen and CO/CO_2. The hydrogen gas can be produced by the decomposition of cellulose (wood flour or similar hydrocarbons). A CO/CO_2 atmosphere results from the decomposition of carbonates such as limestone ($CaCO_3$) or dolomite [$CaMg(CO_3)_2$]. The CO/CO_2 atmosphere can be balanced to provide a reducing (and thus protective) atmosphere.

At high temperatures, CO_2 or CO will react with carbon to achieve equilibrium, which requires the presence of CO and CO_2. It is the relative amounts of CO and CO_2 that determine the reducing/oxidizing nature of the arc environment. The CO/CO_2 ratio also determines the recovery of specific alloying elements. Typical plasma atmospheres for both hydrogen-type and low-hydrogen-type electrodes are given in Table 2.

TABLE 2 GAS COMPOSITION OF WELDING ARC OBTAINED FROM SPECIFIC TYPES OF WELDING ELECTRODES

AWS DESIGNATION	TYPE	COMPOSITION, WT%			
		H_2O	H_2	CO_2	CO
6010	CELLULOSIC	16	41	3	40

6015	BASIC	2	2	19	77

Source: Ref 15

In submerged arc welding, the covering flux also produces the protective shielding gas and plasma. Other gaseous phases, including flourine-bearing components, are also part of the plasma.

The effect of arc environment (for example, amounts of CO and CO_2) on weld metal chemistry control for hyperbaric welding with a basic electrode is shown in Fig. 5. The high-pressure welding allows evaluation of the CO reaction. Considering the CO reaction:

$$C + O = CO \qquad \textbf{(EQ 13)}$$

the law of mass action gives:

$$k = \frac{P_{CO}}{[C][O]} \qquad \textbf{(EQ 14)}$$

where [C] and [O] are the weld metal carbon and oxygen contents, respectively. At equilibrium the partial pressure of CO is directly related to the total pressure by Dalton's Law.

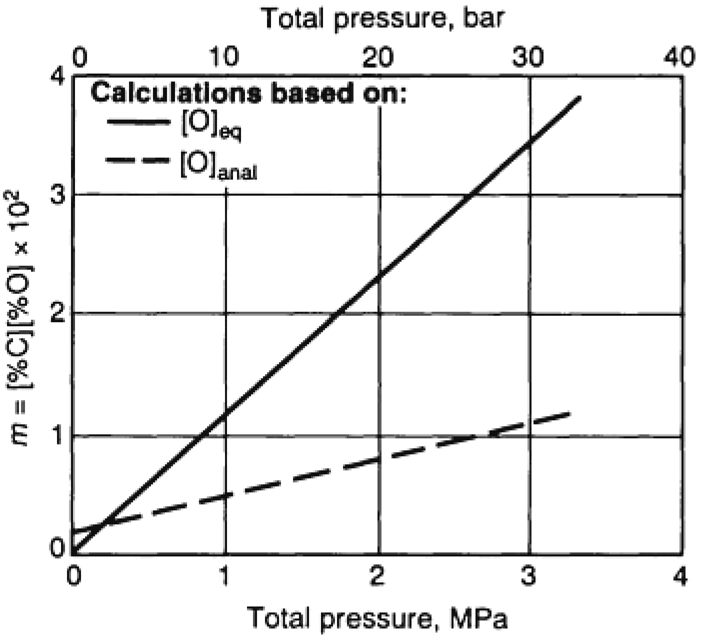

FIG. 5 EFFECT OF PRESSURE ON THE PRODUCT *M* = [%C][%O] FOR HYPERBARIC WELDING WITH A BASIC ELECTRODE. SOURCE: REF 12

In Fig. 5, two lines are indicated. The solid line plots the product of the weld metal oxygen and carbon as a function of total pressure when the analytical weld metal oxygen concentration, $[\%O]_{anal}$, was used. The broken line indicates a similar relationship; however, in this case, the weld metal oxygen concentration has been corrected for the displacement of oxygen because of the formation of manganese silicate inclusions during welding and its transport to the slag. The equilibrium oxygen content, $[\%O]_{eq}$, in the liquid steel at high temperatures is then given by the following expression:

$$[\%O]_{eq} = [\%O]_{anal} + \frac{32}{28}\Delta[\%Si] + \frac{16}{55}\Delta[\%Mn] \quad \textbf{(EQ 15)}$$

where Δ[%Si] and Δ[%Mn] represent the difference between the expected compositions of these elements (from knowledge of initial consumables and base plate compositions as well as dilution) and the actual compositions measured. The linearity of these lines in Fig. 5 is evidence of the strong influence of the CO reaction in arc welding.

The evidence of water reaction control for some electrodes (for example, cellulosic electrodes) can be seen in Fig. 6. With an increase in the degree of oxidation of the deposited metal, the weld metal hydrogen content decreases (Ref 22). The relationship between weld metal oxygen and hydrogen contents as seen in Fig. 6 is consistent with the functional form expected from the law of mass action for a H_2O reaction. It is apparent that with increasing weld metal oxygen content there is a significant reduction in weld metal hydrogen. The hydrogen content of the deposited metal can be reduced by increasing the $CaCO_3$ content of the coating. With the ever-increasing requirements of weld-deposit properties and the increasing need for higher productivity, pyrochemistry will play an important role for flux formulation and weld-property prediction.

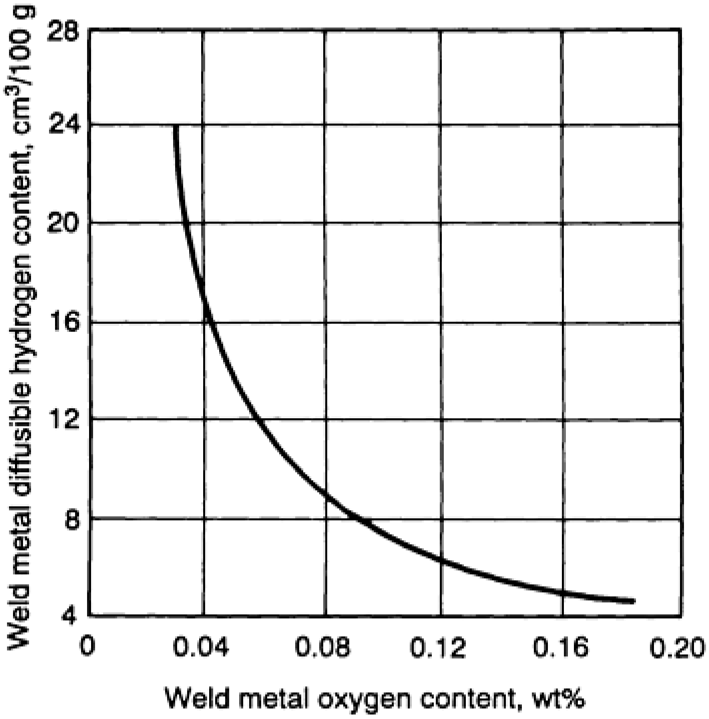

FIG. 6 PLOT OF WELD METAL OXYGEN CONTENT VERSUS WELD METAL HYDROGEN CONTENT WHEN WELDING WITH ELECTRODES THAT CONTAIN CHROMIUM AND NIOBIUM IN THEIR COATINGS. SOURCE: REF 22

Basicity Index

As indicated in the section "Metal Transferability During Pyrochemical Reactions" in this article, the transfer of alloying elements during welding depends strongly on the physical and chemical properties of the flux. The ability to correlate flux properties with weld metal chemical compositions and properties is essential to understand the interactions between weld metal and flux.

Because of the incomplete understanding of the thermodynamic properties of slags, the empirical concept of basicity has been applied to predict flux and weld properties. A basicity index, *BI,* for welding has been proposed:

$$BI = \frac{CaO + CaF_2 + MgO + K_2O + Na_2O + Li_2O + 1/2(MnO + FeO)}{SiO_2 + \frac{1}{2}(Al_2O_3 + TiO_2 + ZrO_2)} \quad \textbf{(EQ 16)}$$

where the chemical components are given in wt%. When the basicity index for a given flux is less than one, the flux is considered acidic. A flux with a *BI* between 1.0 and 1.2 can be classified as neutral. A flux with a *BI* greater than 1.2 is considered basic.

In general, the higher the basicity, the cleaner the weld with respect to nonmetallic inclusions (that is, lower weld metal oxygen content) (Ref 23). Figure 7 illustrates the correlation between weld metal oxygen and the basicity index for some flux systems. The weld metal oxygen content drops significantly as the *BI* is increased to 1.2 and then remains relatively constant at about 250 ppm O. The correlation between weld metal oxygen, which is an indication of weld metal toughness, and basicity is acceptable for some welding flux systems, especially those that are primarily based on CaO, MgO, and SiO_2. However, Eq 16 cannot be used to correlate the strength and toughness of welds made with high flux concentrations of Al_2O_3, TiO_2, ZrO_2, MnO, FeO, and CaF_2. Although numerous basicity formulas have been considered, none has been flexible enough to deal with high amphoteric oxide contents. There is still some concern over utilizing this index (which was primarily conceived by steelmakers for evaluating sulfur refinement) for predicting weld metal oxygen, or for use as a general criterion for weld metal quality (Ref 7). Additionally, the *BI* does not consider the physical properties of the fluxes, nor does it explain the kinetics of the flux-metal reactions.

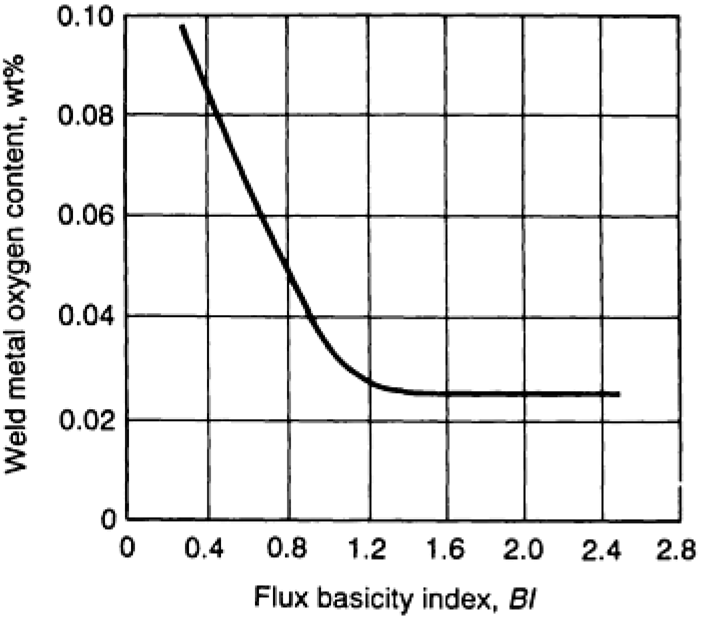

FIG. 7 EFFECT OF WELD METAL OXYGEN CONTENT ON FLUX BASICITY INDEX WHEN USING THE SUBMERGED ARC WELDING PROCESS. SOURCE: REF 23

Basicity index has been used by the welding industry as a measure of expected weld metal cleanliness and mechanical properties. Consequently, manufacturers of welding consumables have classified and advertised their fluxes with this index. It is believed that high basicity means high toughness, a quality of great interest to the engineer, while an acidic flux means excellent slag behavior, a characteristic of interest to the welder attempting to improve weld bead morphology and deposition rate. Table 3 classifies the various coating formulations for SMAW electrodes using the descriptors of cellulosic, basic, acid, and rutile (alternating current stabilizer and slag former).

TABLE 3 ELECTRODE COATING FORMULATIONS OF SELECTED SMAW ELECTRODES

Electrode		**Coating formulation(a)**											**Comments**
Type	**AWS designation**	**Rutile**	**Cellulose**	**Quartz**	**Carbonates**	**Ferromanganese**	**Organics**	**Iron ore-manganese ore**	**Calcium carbonate**	**Complex silicates**	**Fluorspar**	**Ferroalloys**	
Cellulosic	6010	20-60	10-50	15-30	0-15	5-10	. . .	. . .	. . .	. . .	. . .	. . .	Cellulose promotes gas shielding in the arc region. Hydrogen increases heat at weld. High hydrogen content (30-200 ppm). Deep penetration, fast cooling weld
Rutile	6012 6013	40-60 20-40	. . . . . .	15-25 15-25	0-15 5-25	10-12 1.4-14	2-6 0-5	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	Slags mainly for slag shielding relatively high hydrogen content (15-30 ppm). High inclusion content in weld deposit
Acid-ore(a)	6020	. . .	. . .	X	X	X	. . .	X	. . .	X	. . .	. . .	Relatively high hydrogen content. High slag content in weld metal
Basic	7015	0-10	. . .	0-5	. . .	. . .	. . .	. . .	20-50	. . .	20-40	5-10	Relatively low hydrogen levels (⩽ 10 ppm), hence commonly used in welding low-alloy construction steels. Electrodes should be kept dry. Low inclusion content in weld deposit

Source: Ref 24

(A) X, DATA UNAVAILABLE.

Pyrochemical Kinetics During Welding

The ability of a flux to refine as well as protect the weld pool is related to the mass transport processes in the flux. The flux should melt approximately 200 °C (360 °F) below that of the alloy for proper flux coverage and for protection of the weld deposit. One of the most important physical properties of a flux is its slag viscosity, which not only governs the way the slag flows and covers the molten weld pool but also strongly affects the transport processes involved in pore removal, deoxidation, and retention of alloying additions. The chemical processing and refining by the flux to achieve a weld deposit with low concentrations of oxygen and sulfur and optimal concentration of hardenability agents (carbon, manganese, chromium, molybdenum, nickel, and so on) may not be achieved unless slag viscosity is also adequate. The viscosity is strongly temperature-dependent, so the use of various heat inputs during welding may require different flux compositions to produce the matching slag viscosity. For simple ionic melts, the viscosity has been shown to follow an Arrhenius-type temperature dependence:

$$\eta = \eta_0 \exp\left(\frac{E_\eta}{RT}\right) \qquad \textbf{(EQ 17)}$$

where $\eta$is the viscosity, η_0 is a system constant, $E\eta$ is the viscous activation energy, RT is the gas constant, and T is the temperature.

For polymeric melts, Eq 17 often does not hold true. It has been shown that in certain cases a modified equation can be used:

$$\eta = \eta_0 \exp\left[\frac{E_\eta}{R(T - T_0)}\right] \qquad \textbf{(EQ 18)}$$

where T_0 is a constant for a given flux composition. The presence of entrapped particulates tends to increase the slag viscosity. Slag viscosity is also affected by composition. The compositional dependence is commonly reported by considering $E\eta$ to be a function of composition at constant temperature.

The slag must be fluid enough so that it flows and covers the molten weld pool but must be viscous enough so that it does not run away from the molten metal and flow in front of the arc, leading to possible overlapping by the weld metal. (For overhead welding, surface tension becomes a primary factor because fluidity reduces coverage [opposite gravitational vector].) It has been reported that if the manganese silicate flux viscosity at 1450 °C (2640 °F) is above 0.7 Pa · s (7 P), a definite increase in weld surface pocking will occur. Pock marks have been associated with easily reducible oxides in the flux, which contribute oxygen to the weld pool. The weld pool reacts with carbon to form carbon monoxide, which cannot be transported through a high-viscosity flux and is trapped at the liquid-metal/flux interface. The result is a weld metal surface blemished by surface defects or pocks. Because viscosity is sensitive to temperature and thus heat input, pocking can be the evidence that a flux formulated for high-current welding is being used at too low a current or too great a travel speed. The viscosity of most welding fluxes at 1400 °C (2550 °F) is in the range of 0.2 to 0.7 Pa · s (2 to 7 P).

Slag viscosity also affects the shape of the weld deposit and must be carefully controlled when covered electrodes are used out of position. The higher the slag viscosity, the greater the weld penetration in submerged arc welding. However, this benefit must be balanced, because if the viscosity is too high, the gaseous products cannot escape the weld pool, resulting in unacceptable porosity. This condition can be monitored by observing the density of pores trapped in the underside of the detached slag. Detached slags manifesting a honeycomb structure suggest a severe weld metal porosity problem. This condition usually means that a given flux has experienced an insufficient heat input for the effective transport of gas through the slag.

References cited in this section

3. T. LAU, G.C. WEATHERLY, AND A. MCLEAN, THE SOURCES OF OXYGEN AND NITROGEN CONTAMINATION IN SUBMERGED ARC WELDING USING CAO-AL_2O_3 BASED FLUXES, *WELD. J.*, VOL 64 (NO. 12), 1985, P 343S-347S

4. T.H. NORTH, H.B. BELL, A. NOWICKI, AND I. CRAIG, SLAG/METAL INTERACTION, OXYGEN, AND TOUGHNESS IN SUBMERGED ARC WELDING, *WELD. J.*, VOL 57 (NO. 3), 1978, P 63S-75S
5. N. CHRISTENSEN AND J. CHIPMAN, SLAG-METAL INTERACTION IN ARC WELDING, *WELD. RES. BULL.*, NO. 15, JAN 1953, P 1-14
6. R.H. FROST, D.L. OLSON, AND S. LIU, PYROCHEMICAL EVALUATION OF WELD METAL INCLUSION EVOLUTION, *PROC. 3RD INT. CONF. TRENDS IN WELDING*, ASM INTERNATIONAL, JUNE 1992
7. C.A. NATALIE, D.L. OLSON, AND M. BLANDER, PHYSICAL AND CHEMICAL BEHAVIOR OF WELDING FLUXES, *ANN. REV. MATER. SCI.*, VOL 16, 1986, P 389-413
8. J.E. INDACOCHEA, M. BLANDER, N. CHRISTENSEN, AND D.L. OLSON, CHEMICAL REACTIONS WITH FEO-MNO-SIO_2 *FLUXES, METALL. TRANS. B*, VOL 16, 1985, P 237-245
9. U. MITRA AND T.W. EAGAR, SLAG-METAL REACTIONS DURING WELDING, *METALL. TRANS. B*, VOL 22, 1991, P 65-100
10. C.S. CHAI AND T.W. EAGAR, SLAG-METAL EQUILIBRIUM DURING SUBMERGED ARC WELDING,*METALL. TRANS. B*, VOL 12, 1981, P 539-547
12. O. GRONG, D.L. OLSON, AND N. CHRISTENSEN, CARBON OXIDATION IN HYPERBARIC MMA WELDING, *MET. CONSTRUCT.*, VOL 17. DEC 1985, P 810R-814R
14. T.W. EAGAR, SOURCES OF WELD METAL OXYGEN CONTAMINATION DURING SUBMERGED ARC WELDING, *WELD. J.*,VOL 57, 1978, P 76S-80S
15. N. CHRISTENSEN, *WELDING METALLURGY*, *LECTURE NOTES*, NTH, 1979
16. O. GRONG, T.A. SIEWERT, T.A. MARTINS, AND D.L. OLSON, A MODEL FOR THE SILICON-MANGANESE DEOXIDATION OF STEEL WELD METAL, *METALL. TRANS. A*, VOL 17 (NO. 10), 1985, P 1797-1807
17. T.H. NORTH, THE DISTRIBUTION OF MANGANESE BETWEEN SLAG AND METAL DURING SUBMERGED ARC WELDING, *WELD. RES. ABROAD*, VOL 23 (NO. 1), 1977, P 2-40
18. T. BONISZEWSKI, BASIC FLUXES AND DEOXIDATION IN SUBMERGED ARC WELDING OF STEEL, *METAL. CONSTR. BRIT. WELD. J.*, VOL 6, 1974, P 128
19. L. DAVIS, *AN INTRODUCTION TO WELDING FLUXES FOR MILD AND LOW ALLOY STEELS*, WELDING INSTITUTE, 1981
20. *MAKING, SHAPING AND TREATING OF STEEL*, UNITED STATES STEEL CORPORATION, 1984
21. P.S. DUNN, C.A. NATALIE, AND D.L. OLSON, SOL-GEL FLUXES FOR FLUX CORED WELDING CONSUMABLES, *J. MATER. ENERGY SYSTEMS*, VOL 8 (NO. 2), 1986, P 176-184
22. L.I. SOROKIN AND Z.A. SIDLIN, THE EFFECT OF ALLOYING ELEMENTS AND OF MARBLE IN AN ELECTRODE COATING ON THE SUSCEPTIBILITY OF A DEPOSITED NICKEL CHROME METAL TO PORE FORMATION, *SVAR. PROIZVOD.*, NO. 11, 1974, P 7-9
23. S.S. TULIANI, T. BONISZEWSKI, AND N.F. EATON, NOTCH TOUGHNESS OF COMMERCIAL SUBMERGED ARC WELD METAL, *WELD. MET. FABR.*, VOL 37 (NO. 8), 1969, P 27
24. T.G.F. GRAY, J. SPENCE, AND T.H. NORTH, *RATIONAL WELDING DESIGN*, BUTTERWORTHS, 1975

Nature and Behavior of Fluxes Used for Welding

D.L. Olson, S. Liu, R.H. Frost, G.R. Edwards, and D.A. Fleming, Colorado School of Mines

Alloy Modification

Other additions to the welding consumables are required to make alloy additions to the weld pool, usually in the form of powder metal or ferro-additions (Ref 1). Often the composition of the wire that makes up the rod for the shielded metal arc electrode from a specific manufacturer is the same, regardless of the alloy to be welded. Alloying is achieved by powder metal additions to the flux coating. Manganese, silicon, chromium, niobium, and other alloying additions are adjusted in the weld pool by ferroalloy powder additions. Specially prepared alloy additions of Fe-50Si, Fe-80Mn, Fe-60Mn-30Si, and others are used. One concern in formulating electrodes using ferro-additions is the alloying element recovery (that is, the amount of the element that is transferred across the arc and into the weld deposit). Values for the recovery of typical elements in steel welding are given in Table 4. The metal losses are either to the slag or to the fume.

TABLE 4 RECOVERY OF ELEMENTS FROM SELECTED ELECTRODE COVERINGS

ALLOY ELEMENT	FORM OF MATERIAL IN ELECTRODE COVERING	APPROXIMATE RECOVERY OF ELEMENT, WT%
ALUMINUM	FERROALUMINUM	20
BORON	FERROBORON	2
CARBON	GRAPHITE	75
CHROMIUM	FERROCHROMIUM	95
NIOBIUM	FERROCOLUMBIUM	70
COPPER	COPPER METAL	100
MANGANESE	FERROMANGANESE	75
MOLYBDENUM	FERROMOLYBDENUM	97
NICKEL	ELECTROLYTIC NICKEL	100
SILICON	FERROSILICON	45
TITANIUM	FERROTITANIUM	5
VANADIUM	FERROVANADIUM	80

Source: Ref 1

Slipping and Binding Agents. In the case of SMAW electrodes, slipping agents are also added to the green flux formulation to improve the extrudability of the flux onto the rod (Ref 1). Glycerin, china clay, kaolin, talc, bentonite, and mica have all been used as slipping agents. Binding agents can be classified into two types. The first type comprises binders that bond the flux components to the rod without introducing a hydrogen source. These low-hydrogen binders include sodium silicate and potassium silicate. The second type of binding agent does function as a hydrogen source. The binders used for high-hydrogen electrodes can be organic in nature and include gum arabic, sugar, dextrine, and other specialized synthetic organic binders.

Slag Formation. Slag, a mixture of glass and crystalline structure, must solidify on the already solidified weld deposit to protect the surface from oxidation during cooling. Specific physical properties are required of the slag. It must melt below the melting temperature of steel (~ 1450 °C, or 2640 °F), must have a density significantly less than steel to reduce slag entrapment in the weld deposit, must possess the proper viscosity in the temperature range of 1450 to 1550 °C (2640 to 2820 °F), and must easily detach from the weld deposit after welding.

Silicates, aluminates, and titanates are all primary slag formers. The high-valence cations of these compounds produce a bonding network that can promote glass formation. Most electrodes produce silicate or titanate slag. Silicates of such elements as manganese produce very workable glass deposits, but also produce weld deposits that are relatively high in weld metal oxygen content. These fluxes are said to range from acidic to neutral. The titanate and aluminate fluxes produce more rigid and stable oxide covers. The result is a lower concentration of weld metal oxygen, but a higher-viscosity slag. Aluminates and calcium-bearing compounds are common additions to basic fluxes for the submerged arc welding of high-toughness linepipe steels. Minerals used for slag formation include:

- RUTILE
- POTASSIUM TITANATE
- ILMENITE
- ALUMINA
- SILICA FLOUR

- IRON POWDER
- FLUORSPAR
- FELDSPAR
- MANGANESE DIOXIDE

Asbestos (up to 50%) was used as a slag former, but has been phased out by welding consumable manufacturers.

Slag detachability is a serious productivity concern for steel fabricators. The relative ease of slag detachability influences the economic advantages of flux-related welding processes. Residual slag on the weld deposit promotes slag stringers in multipass weldments, limits the effective use of narrow-gap flux-related welding processes, and may reduce the corrosion resistance of the weldment. Welding flux formulators have modified flux compositions to alleviate or reduce this hindrance (Ref 25).

Poor slag removal has been reported to occur when the flux contains fluorite. Slags containing spinels generally have been found to attach tenaciously to the weld deposit. Slags with cordierite and $(Cr,Mn,Mg)O \cdot (Cr,Mn,Al)_2O_3$ type spinel phases have been reported to be difficult to remove from stainless steel weldments. It has also been reported to be difficult to remove from stainless steel weldments. It has also been reported that if $(CaO)_2SiO_2$, Cr_2TiO_5, and $FeTiO_5$ are present, the slag readily detaches from the weld deposit (Ref 25). The compositional range for acceptable slag detachability has been reported for the $CaO \cdot CaF_2 \cdot SiO_2$ system and the $CaO \cdot TiO_2 \cdot SiO_2$ system. Easier slag removal has also been related to deoxidation with aluminum instead of titanium. Increasing the Al_2O_3 content in the flux has demonstrated improved detachability (Ref 25).

Differences in the coefficient of thermal expansion between the slag and the weld are also important.

References cited in this section

1. G.E. LINNERT, CHAPTER 8, *WELDING METALLURGY*, VOL 1, AWS, 1965, P 367-396
25. D.L. OLSON, G.R. EDWARDS, AND S.K. MARYA, THE PHYSICAL AND CHEMICAL BEHAVIOR ASSOCIATED WITH SLAG DETACHABILITY DURING WELDING, *FERROUS ALLOY WELDMENTS*, VOL 67-70, TRANS TECH, 1992, P 253-268

Nature and Behavior of Fluxes Used for Welding

D.L. Olson, S. Liu, R.H. Frost, G.R. Edwards, and D.A. Fleming, Colorado School of Mines

Types of Fluxes

Producing a weld deposit with high mechanical integrity and with high productivity requires many physical and chemical occurrences in the welding arc. These occurrences are achieved in flux-related arc welding processes by the careful formulation of the welding flux. The welding flux must:

- STABILIZE ARC AND CONTROL ARC RESISTIVITY
- PROVIDE SLAG WITH PROPER MELTING TEMPERATURE
- PROVIDE LOW-DENSITY SLAG
- PERMIT USE OF DIFFERENT TYPES OF CURRENT AND POLARITY
- ADD ALLOYING ELEMENTS
- REFINE THE WELD POOL (DEOXIDATION AND DESULFURIZATION)
- PROVIDE PROPER VISCOSITY FOR OUT-OF-POSITION WELDING
- PROMOTE SLAG DETACHABILITY
- PRODUCE SMOOTH WELD CONTOUR

- REDUCE SPATTER AND FUME

It is apparent, after considering the large number of necessary requirements, that a welding flux must be carefully and deliberately formulated to achieve an optimized performance.

A number of different arc welding processes depend on welding fluxes. Each of these processes requires a different formulation.

SMAW Fluxes. The typical constituents and their functions in electrode coatings for the SMAW process are given in Table 5. It should be noted that the flux ingredients are based on additions of refined minerals of the earth. Natural minerals offer an economical method of keeping welding consumables at a reasonable cost. Table 6 presents the elemental content of these minerals. Table 7 gives the typical chemical compositions for flux coatings for three different SMAW electrodes.

TABLE 5 TYPICAL FUNCTIONS AND COMPOSITIONS OF CONSTITUENTS FOR SELECTED MILD STEEL SMAW ELECTRODE COATINGS

COATING CONSTITUENT	FUNCTION OF CONSTITUENT		COMPOSITION RANGE OF COATING ON ELECTRODE				
	PRIMARY	SECONDARY	E6010, E6011	E6013	E7018	E7024	E7028
CELLULOSE	SHIELDING GAS	. . .	25-40	2-12	. . .	1-5	. . .
CALCIUM CARBONATE	SHIELDING GAS	FLUXING AGENT	. . .	0-5	15-30	0-5	0-5
FLUORSPAR	SLAG FORMER	FLUXING AGENT	. . .	. . .	15-30	. . .	5-10
DOLOMITE	SHIELDING GAS	FLUXING AGENT	. . .	. . .	. . .	. . .	5-10
TITANIUM DIOXIDE (RUTILE)	SLAG FORMER	ARC STABILIZER	10-20	30-55	0-5	20-35	10-20
POTASSIUM TITANATE	ARC STABILIZER	SLAG FORMER	(A)	(A)	0-5	. . .	0-5
FELDSPAR	SLAG FORMER	STABILIZER	. . .	0-20	0-5	. . .	0-5
MICA	EXTRUSION	STABILIZER	. . .	0-15	. . .	0-15	. . .
CLAY	EXTRUSION	SLAG FORMER	. . .	0-10	. . .	. . .	. . .
SILICA	SLAG FORMER	. . .	. . .	. . .	. . .	. . .	. . .
ASBESTOS	SLAG FORMER	EXTRUSION	10-20	. . .	. . .	. . .	. . .
MANGANESE OXIDE	SLAG FORMER	ALLOYING	. . .	. . .	. . .	. . .	. . .
IRON OXIDE	SLAG FORMER	. . .	. . .	. . .	. . .	. . .	
IRON POWDER	DEPOSITION RATE	CONTACT WELDING	. . .	. . .	25-40	40-55	40-55
FERROSILICON	DEOXIDIZER	. . .	. . .	. . .	5-10	0-5	2-6
FERROMANGANESE	ALLOYING	DEOXIDIZER	5-10	5-10	2-6	5-10	2-6
SODIUM SILICATE	BINDER	FLUXING	20-30	5-10	0-5	0-10	0-5

		AGENT					
POTASSIUM SILICATE	ARC STABILIZER	BINDER	(B)	5-15(B)	5-10	0-10	0-5

(A) REPLACES TITANIUM DIOXIDE (RUTILE) TO PERMIT USE WITH ALTERNATING CURRENT.

(B) REPLACES SODIUM SILICATE TO PERMIT USE WITH ALTERNATING CURRENT

TABLE 6 TYPICAL COMPOSITION OF COMMON MINERALS USED IN SMAW ELECTRODE COATINGS

MINERAL	**CHEMICAL COMPOSITION**
ILMENITE	$FEO \cdot TIO_2$
TALC	$3MGO \cdot 4SIO_2 \cdot H_2O$
BENTONITE	COMPLEX AL, MG, CA, FE HYDROXIDES
SILICA, QUARTZ	SIO_2
CELLULOSE	$(C_6H_{10}O_5)_X$
ALUMINA	AL_2O_3
MUSCOVITE, MICA(A)	$K_2O \cdot 3AL_2 \cdot 6SIO_2 \cdot 2H_2O$
ACTINOLITE	$CAO \cdot MGO \cdot 2FEO \cdot 4SIO_2$
MAGNETITE	FE_3O_4
HEMATITE	FE_2O_3
RUTILE, TITANIA	TIO_2
DOLOMITE	$MGO \cdot CAO \cdot (CO_2)_2$
FLUORSPAR, FLUORITE	CAF_2
CRYOLITE	NA_3ALF_4
LIME	CAO
LIMESTONE, CALCITE, MARBLE	$CACO_3$
ZIRCONIA	ZRO_2
FELDSPAR(A)	$K_2O \cdot AL_2O_3 \cdot 6SIO_2$
CLAY(A)	$AL_2O_3 \cdot 2SIO_2 \cdot 2H_2O$
SODIUM SILICATE	SIO_2/NA_2O (RATIO-3.22)
POTASSIUM SILICATE	SIO_2 K_2O (RATIO-2.11)
CHROMIC OXIDE	CR_2O_3

Source: Ref 1

(A) ALTHOUGH THESE SUBSTANCES CAN HAVE SEVERAL CHEMICAL COMPOSITIONS, ONLY TYPICAL COMPOSITION IS GIVEN.

TABLE 7 CHEMICAL COMPOSITION OF COVERINGS USED IN ELECTRODES FOR SMAW WELDING OF MILD STEELS AND LOW-ALLOY STEELS

ELECTRODE		**COMPOSITION, WT%**[(A)]														
DESCRIPTION	**AWS DESIGNATION**	**CAO**	**TIO_2**	**CAF_2**	**SIO_2**	**AL_2O_3**	**MGO**	**NA_3ALF_3**	**N_3O**	**FEO**	**SI**	**MN**	**FE**	**CO + CO_2**	**VOLATILE MATTER**	**MOISTURE**
HIGH-CELLULOSE, GAS SHIELDED	E6010	. . .	10.1	. . .	47.0	. . .	3.2	. . .	5.1	1.3	1.5	2.8	. . .	. . .	25.0	4.0
HIGH-TITANIUM GAS-SLAG SHIELD	E6012	. . .	46.0	. . .	23.6	5.0	2.0	. . .	2.4	7.0	1.5	2.5	. . .	. . .	5.0	2.0
LOW-HYDROGEN IRON POWDER	E7018	14.4	. . .	11.0	20.5	2.0	1.0	5.0	1.2	. . .	2.5	1.8	28.5	12.0	. . .	0.1

Source: Ref 1

(A) AFTER BAKING.

Flux-cored arc welding (FCAW) uses a hollow wire filled with flux reagents and ferro-additions. The two types of flux-cored electrodes are carbon-dioxide-shielded flux-cored electrodes and self-shielded flux-cored electrodes (Ref 26). Table 8 gives typical compositions for the three types of carbon-dioxide-shielded cored electrodes. Table 9 gives the typical compositions for the four types of self-shielded flux-cored electrodes.

TABLE 8 TYPICAL FLUX COMPOSITIONS OF AVAILABLE CO_2 SHIELDED FCAW ELECTRODES

ELECTRODE DESCRIPTION		AWS DESIGNATION	COMPOSITION, WT%											
TYPE NUMBER	FLUX TYPE		SIO_2	AL_2O_3	TIO_2	ZRO_2	CAO	NA_2O	K_2O	CO_2 (AS CARBONATE)	C	FE	MN	CAF_2
1	TITANIA (NONBASIC)	E70T-1, E70T-2	21.0	2.1	40.5	. . .	0.7	1.6	1.4	0.5	0.6	20.1	15.8	. . .
2	LIME-TITANIA (BASIC OR NEUTRAL)	E70T-1	17.8	4.3	9.8	6.2	9.7	1.9	1.5	. . .	0.3	24.7	13.0	18.0
3	LIME (BASIC)	E70T-1, E70T-5	7.5	0.5	. . .	. . .	3.2	. . .	0.5	2.5	1.1	55.0	7.2	20.5

Source: Ref 26

TABLE 9 TYPICAL FLUX COMPOSITIONS OF AVAILABLE SELF-SHIELDED FCAW ELECTRODES

ELECTRODE DESCRIPTION		AWS DESIGNATION	COMPOSITION, WT%													
TYPE NUMBER	FLUX TYPE		SIO_2	AL	AL_2O_3	TIO_2	CAO	MGO	K_2O	NA_2O	C	CO_2 (AS CARBONATE)	FE	MN	NI	CAF_2
1	FLUORSPAR-ALUMINUM	E70T-4, E70T-7, E60T-8	0.5	15.4	. . .	. . .	. . .	12.6	0.4	0.2	1.2	0.4	0.4	3.0	. . .	63.5
2	FLUORSPAR-TITANIA	E70T-3	3.6	1.9	. . .	20.6	. . .	4.5	0.6	0.1	0.6	0.6	50.0	4.5	. . .	22.0
3	FLUORSPAR-LIME-TITANIA	E70T-6	4.2	1.4	. . .	14.7	4.0	2.2	. . .	. . .	0.6	2.1	50.5	2.0	2.4	15.3
4	FLUORSPAR-LIME	E70T-5	6.9	. . .	0.6	1.2	3.2	. . .	. . .	0.6	0.3	1.3	58.0	7.9	. . .	22.0

Source: Ref 26

Submerged Arc Welding (SAW). In the SAW process, the flux drops from a hopper onto the work such that the welding arc is submerged beneath the granular flux, producing an arc cavity that contains metal vapors, silicon monoxide, manganese oxide, gaseous fluorides, and other higher-vapor-pressure components of the flux. This arc welding process has been recognized as one that produces very little, if any, fume. Submerged arc welding is limited mainly to the flat or horizontal position, and requires significant setup time. It is very successful in manufacturing numerous similar parts, such as producing welded steel pipe.

Submerged arc fluxes are made in three different forms:

- *BONDED FLUXES* MIX NONMETALLIC AND FERRO-ADDITIONS WITH A LOW-TEMPERATURE BINDER INTO MIXTURES OF SMALL PARTICLES.
- *AGGLOMERATED FLUXES* ARE SIMILAR TO THE BONDED FLUXES EXCEPT THAT A CERAMIC GLASS BINDER, CURED AT HIGH TEMPERATURE, IS USED.
- *FUSED FLUXES* ARE MADE BY POURING A HOMOGENEOUS GLASS MIXTURE OF THE PROPER FLUX COMPOSITION INTO WATER, RESULTING IN A FRIT.

Generally speaking, changing from bonded to agglomerated to fused fluxes improves control of the weld metal composition, especially with respect to such impurities as hydrogen and oxygen.

There are seven types of submerged arc fluxes (Table 10). Flux classification according to basicity is the result of observed correlations between weld metal oxygen concentration and flux composition. Table 11 gives some typical compositions for SAW fluxes.

TABLE 10 FLUXES USED FOR SAW APPLICATIONS

FLUX TYPE	CONSTITUENTS	BASICITY	FLUX FORM	ADVANTAGES	LIMITATIONS	COMMENTS
MANGANESE SILICATE	MNO + SIO_2 > 50%	ACID	FUSED	MODERATE STRENGTH; TOLERANT TO RUST; FAST WELDING SPEEDS; HIGH HEAT INPUT; GOOD STORAGE	LIMITED USE FOR MULTIPASS WELDING; USE WHERE NO TOUGHNESS REQUIREMENT; HIGH WELD METAL OXYGEN; INCREASE IN SILICON ON WELDING; LOW IN CARBON	ASSOCIATED MANGANESE GAIN; MAXIMUM CURRENT, 1100 A; HIGHER WELDING SPEEDS
CALCIUM-HIGH SILICA	CAO + MGO + SIO_2 > 60%	ACID	AGGLOMERATED, FUSED	HIGH WELDING CURRENT; TOLERANT TO RUST	POOR WELD TOUGHNESS; USE WHERE NO TOUGHNESS REQUIREMENT; HIGH WELD METAL OXYGEN	DIFFER IN SILICON GAIN; SOME CAPABLE OF 2500 A; WIRES WITH HIGH MANGANESE
CALCIUM SILICATE-NEUTRAL	CAO + MGO + SIO_2 > 60%	NEUTRAL	AGGLOMERATED, FUSED	MODERATE STRENGTH AND TOUGHNESS; ALL CURRENT TYPES; TOLERANT TO RUST; SINGLE- OR MULTIPLE-PASS WELD	. . .	. . .
CALCIUM SILICATE-LOW SILICA	CAO + MGO + SIO_2 > 60%	BASIC	AGGLOMERATED, FUSED	GOOD TOUGHNESS WITH MEDIUM STRENGTH; FAST WELDING SPEEDS; LESS CHANGE IN COMPOSITION AND LOWER OXYGEN	NOT TOLERANT TO RUST; NOT USED FOR MULTIWIRE WELDING	. . .
ALUMINATE	AL_2O_3 + CAO +	BASIC	AGGLOMERATED	GOOD STRENGTH	NOT TOLERANT TO	USUALLY

BASIC	MGO > 45%; AL_2O_3 > 20%			AND TOUGHNESS IN MULTIPASS WELDS; NO CHANGE IN CARBON; LOSS OF SULFUR AND SILICON	RUST, LIMITED TO DC ELECTRODE POSITIVE	MANGANESE GAIN; MAXIMUM CURRENT 1200 A; GOOD MECHANICAL PROPERTIES
ALUMINA	BAUXITE BASE	NEUTRAL	AGGLOMERATED, FUSED	LESS CHANGE IN WELD COMPOSITION AND LOWER OXYGEN THAN FOR ACID TYPE; MODERATE TO FAST WELDING SPEEDS	. . .	. . .
BASIC FLUORIDE	CAO + MGO + MNO + CAF_2 > 50%; $SIO_2 \leqslant 22\%$; $CAF_2 \leqslant 15\%$	BASIC	AGGLOMERATED, FUSED	VERY LOW OXYGEN; MODERATE TO GOOD LOW-TEMPERATURE TOUGHNESS	MAY PRESENT PROBLEMS OF SLAG DETACHABILITY; MAY PRESENT PROBLEM OF MOISTURE PICKUP	CAN BE USED WITH ALL WIRES, PREFERABLE DC WELDING; VERY GOOD WELD PROPERTIES

TABLE 11 TYPICAL COMPOSITIONS OF SMAW FLUXES

FLUX	COMPOSITION, WT%									BASICITY INDEX (*BI*)
	AL_2O_3	SIO_2	TIO_2	MGO	CAF_2	CAO	MNO	NA_2O	K_2O	
A	49.9	13.7	10.1	2.9	5.7	. . .	15.1	1.6	0.2	0.4
B	24.9	18.4	0.2	28.9	24.2	. . .	1.8	2.1	0.07	1.8
C	19.3	16.3	0.8	27.2	23.6	9.8	0.08	0.9	1.1	2.4
D	18.1	13.2	0.5	28.2	31.8	4.5	0.1	0.9	0.9	3.0
E	17.0	12.2	0.7	36.8	29.2	0.7	8.9	1.6	0.1	3.5

References cited in this section

1. G.E. LINNERT, CHAPTER 8, *WELDING METALLURGY*, VOL 1, AWS, 1965, P 367-396
26. *FUMES AND GASES IN THE WELDING ENVIRONMENT*, AWS, 1979

Nature and Behavior of Fluxes Used for Welding

D.L. Olson, S. Liu, R.H. Frost, G.R. Edwards, and D.A. Fleming, Colorado School of Mines

References

1. G.E. LINNERT, CHAPTER 8, *WELDING METALLURGY*, VOL 1, AWS, 1965, P 367-396
2. C.E. JACKSON, FLUXES AND SLAGS IN WELDING, *WELD. RES. BULL.*, NO. 190, 1973
3. T. LAU, G.C. WEATHERLY, AND A. MCLEAN, THE SOURCES OF OXYGEN AND NITROGEN CONTAMINATION IN SUBMERGED ARC WELDING USING CAO-AL_2O_3 BASED FLUXES, *WELD. J.*, VOL 64 (NO. 12), 1985, P 343S-347S
4. T.H. NORTH, H.B. BELL, A. NOWICKI, AND I. CRAIG, SLAG/METAL INTERACTION, OXYGEN, AND TOUGHNESS IN SUBMERGED ARC WELDING, *WELD. J.*, VOL 57 (NO. 3), 1978, P 63S-75S
5. N. CHRISTENSEN AND J. CHIPMAN, SLAG-METAL INTERACTION IN ARC WELDING, *WELD. RES. BULL.*, NO. 15, JAN 1953, P 1-14
6. R.H. FROST, D.L. OLSON, AND S. LIU, PYROCHEMICAL EVALUATION OF WELD METAL INCLUSION EVOLUTION, *PROC. 3RD INT. CONF. TRENDS IN WELDING*, ASM INTERNATIONAL, JUNE 1992
7. C.A. NATALIE, D.L. OLSON, AND M. BLANDER, PHYSICAL AND CHEMICAL BEHAVIOR OF WELDING FLUXES, *ANN. REV. MATER. SCI.*, VOL 16, 1986, P 389-413
8. J.E. INDACOCHEA, M. BLANDER, N. CHRISTENSEN, AND D.L. OLSON, CHEMICAL REACTIONS WITH FEO-MNO-SIO_2 *FLUXES, METALL. TRANS. B*, VOL 16, 1985, P 237-245
9. U. MITRA AND T.W. EAGAR, SLAG-METAL REACTIONS DURING WELDING, *METALL. TRANS. B*, VOL 22, 1991, P 65-100
10. C.S. CHAI AND T.W. EAGAR, SLAG-METAL EQUILIBRIUM DURING SUBMERGED ARC WELDING,*METALL. TRANS. B*, VOL 12, 1981, P 539-547
11. N. CHRISTENSEN, METALLURGICAL ASPECTS OF ARC WELDING, *WELD. J.*, VOL 27, 1949, P 373S-380S
12. O. GRONG, D.L. OLSON, AND N. CHRISTENSEN, CARBON OXIDATION IN HYPERBARIC MMA WELDING, *MET. CONSTRUCT.*, VOL 17. DEC 1985, P 810R-814R
13. U. MITRA AND T.W. EAGAR, SLAG-METAL REACTIONS DURING SUBMERGED ARC WELDING OF ALLOY STEELS, *METALL. TRANS. A*, VOL 15, 1984, P 217-227

14. T.W. EAGAR, SOURCES OF WELD METAL OXYGEN CONTAMINATION DURING SUBMERGED ARC WELDING, *WELD. J.*,VOL 57, 1978, P 76S-80S
15. N. CHRISTENSEN, *WELDING METALLURGY, LECTURE NOTES*, NTH, 1979
16. O. GRONG, T.A. SIEWERT, T.A. MARTINS, AND D.L. OLSON, A MODEL FOR THE SILICON-MANGANESE DEOXIDATION OF STEEL WELD METAL, *METALL. TRANS. A*, VOL 17 (NO. 10), 1985, P 1797-1807
17. T.H. NORTH, THE DISTRIBUTION OF MANGANESE BETWEEN SLAG AND METAL DURING SUBMERGED ARC WELDING, *WELD. RES. ABROAD*, VOL 23 (NO. 1), 1977, P 2-40
18. T. BONISZEWSKI, BASIC FLUXES AND DEOXIDATION IN SUBMERGED ARC WELDING OF STEEL, *METAL. CONSTR. BRIT. WELD. J.*, VOL 6, 1974, P 128
19. L. DAVIS, *AN INTRODUCTION TO WELDING FLUXES FOR MILD AND LOW ALLOY STEELS*, WELDING INSTITUTE, 1981
20. *MAKING, SHAPING AND TREATING OF STEEL*, UNITED STATES STEEL CORPORATION, 1984
21. P.S. DUNN, C.A. NATALIE, AND D.L. OLSON, SOL-GEL FLUXES FOR FLUX CORED WELDING CONSUMABLES, *J. MATER. ENERGY SYSTEMS*, VOL 8 (NO. 2), 1986, P 176-184
22. L.I. SOROKIN AND Z.A. SIDLIN, THE EFFECT OF ALLOYING ELEMENTS AND OF MARBLE IN AN ELECTRODE COATING ON THE SUSCEPTIBILITY OF A DEPOSITED NICKEL CHROME METAL TO PORE FORMATION, *SVAR. PROIZVOD.*, NO. 11, 1974, P 7-9
23. S.S. TULIANI, T. BONISZEWSKI, AND N.F. EATON, NOTCH TOUGHNESS OF COMMERCIAL SUBMERGED ARC WELD METAL, *WELD. MET. FABR.*, VOL 37 (NO. 8), 1969, P 27
24. T.G.F. GRAY, J. SPENCE, AND T.H. NORTH, *RATIONAL WELDING DESIGN*, BUTTERWORTHS, 1975
25. D.L. OLSON, G.R. EDWARDS, AND S.K. MARYA, THE PHYSICAL AND CHEMICAL BEHAVIOR ASSOCIATED WITH SLAG DETACHABILITY DURING WELDING, *FERROUS ALLOY WELDMENTS*, VOL 67-70, TRANS TECH, 1992, P 253-268
26. *FUMES AND GASES IN THE WELDING ENVIRONMENT*, AWS, 1979

Shielding Gases for Welding

Kevin A. Lyttle, Praxair, Inc.

Introduction

THE SHIELDING GAS used in a welding process has a significant influence on the overall performance of the welding system. Its primary function is to protect the molten metal from atmospheric nitrogen and oxygen as the weld pool is being formed. The shielding gas also promotes a stable arc and uniform metal transfer. In gas-metal arc welding (GMAW) and flux-cored arc welding (FCAW), the gas used has a substantial influence on the form of metal transfer during welding. This, in turn, affects the efficiency, quality, and overall operator acceptance of the welding operation.

The shielding gas interacts with the base material and with the filler material, if any, to produce the basic strength, toughness, and corrosion resistance of the weld. It can also affect the weld bead shape and the penetration pattern.

Understanding the basic properties of a shielding gas will aid in the selection of the right shielding gas or gases for a welding application. Use of the best gas blend will improve the quality and may reduce the overall cost of the welding operation as well.

Shielding Gases for Welding

Kevin A. Lyttle, Praxair, Inc.

Basic Properties of a Shielding Gas

The "controlled electrical discharge" known as the welding arc is formed and sustained by the establishment of a conductive medium called the arc plasma. This plasma consists of ionized gas, molten metals, slags, vapors, and gaseous atoms and molecules. The formation and structure of the arc plasma is dependent on the properties of the shielding gases used for welding. Table 1 lists the basic properties of gases used for welding (Ref 1).

TABLE 1 PROPERTIES OF SHIELDING GASES USED FOR WELDING

GAS	CHEMICAL SYMBOL	MOLECULAR WEIGHT	SPECIFIC GRAVITY(A)	DENSITY		IONIZATION POTENTIAL	
				g/ft^3	g/l	aj(b)	eV
ARGON	AR	39.95	1.38	0.1114	1.784	2.52	15.7
CARBON DIOXIDE	CO$_2$	44.01	1.53	0.1235	1.978	2.26	14.4
HELIUM	HE	4.00	0.1368	0.0111	0.178	3.92	24.5
HYDROGEN	H$_2$	2.016	0.0695	0.0056	0.090	2.16	13.5
NITROGEN	N$_2$	28.01	0.967	0.782	12.5	2.32	14.5
OXYGEN	O$_2$	32.00	1.105	0.0892	1.43	2.11	13.2

Source: Ref 1

(A) AT 100 KPA (1 ATM) AND 0 °C (32 °F); AIR = 1.

(B) 10^{-18} J.

The ionization potential is the energy, expressed in electron volts, necessary to remove an electron from a gas atom--making it an ion, or an electrically charged gas atom. All other factors held constant, the value of the ionization potential decreases as the molecular weight of the gas increases. Arc starting and arc stability are greatly influenced by the ionization potentials of the component shielding gases used in welding process. A gas with a low ionization potential, such as argon, can atoms into ions easily. Helium, with its significantly higher ionization potential, produces a harder to start, less stable arc.

Although other factors are involved in sustaining the plasma, the respective energy levels required to ionize these gases must be maintained; as a consequence, the arc voltage is directly influenced. For equivalent arc lengths and welding currents, the voltage obtained with helium is appreciably higher than is with argon. This translates into more available heat input to the base material with helium than with argon.

The thermal conductivity of a gas is a measure of how well it is able to conduct heat. It influences the radial heat loss from the center to the periphery of the arc column as well as heat transfer between the plasma and the liquid metal. Argon, which has a low thermal conductivity, produces an arc that has two zones: a narrow hot core and a considerably cooler outer zone. The penetration profile of the weld fusion area then exhibits a narrow "finger" at the root and a wider top. A gas with a high thermal conductivity conducts heat outward from the core; this results in a wider, hotter arc core. This type of heat distribution occurs with helium, argon-hydrogen, and argon-carbon dioxide blends; it gives a more even distribution of heat to the work surface and produces a wider fusion area.

Dissociation and Recombination. Shielding gases such as carbon dioxide, hydrogen, and oxygen are multiatom molecules. When heated to high temperatures within the arc plasma, these gases break down, or dissociate, into their component atoms. They are then at least partially ionized, producing free electrons and current flow. As the dissociated gas comes into contact with the relatively cool work surface, the atoms recombine and release heat at that point. This heat of recombination causes multiatomic gases to behave as if they have a higher thermal conductivity, similar to that of helium. Dissociation and recombination do not occur with gases, such as argon, that consist of a single atom. Thus, at the same arc temperature, the heat generated at the work surface can be considerably greater with gases such as carbon dioxide and hydrogen.

Reactivity/Oxidation Potential. The oxidizing nature of the shielding gas affects both welding performance and the properties of the resultant weld deposit. Argon and helium are completely nonreactive, or inert, and thus have no direct chemical affect on the weld metal. Oxidizing or active gases, such as CO_2 and oxygen, will react with elements in the filler metal or baseplate and will form a slag on the surface of the weld deposit. The loss of elements, such as manganese and silicon, from steel can affect the quality and cost of the weldment produced. Both weld strength and toughness generally decline as the oxidizing nature of the shielding gas increases.

Additions of reactive gases such as oxygen or carbon dioxide enhance the stability of the arc and affect the type of metal transfer obtained. Metal droplet size is decreased, and the number of droplets transferred per unit time increases as the level of oxygen in the shielding gas increases. Oxygen reduces the molten weld bead surface tension, promoting better bead wetting and higher welding travel speeds. Small additions of CO_2 work in a similar manner.

The surface tension between the molten metal and its surrounding atmosphere has a pronounced influence on bead shape. If the surface energy is high, a convex, irregular bead will result. Low values promote flatter beads with minimal susceptibility for undercutting.

Pure argon is generally associated with high interfacial energy, producing a sluggish weld puddle and high, crowned bead. The addition of a small amount of a reactive gas, such as oxygen, lowers this surface tension and promote fluidity and better wetting of the base material; it does this without creating excessive oxidation of the weld metal.

Gas Purity. Some metals, such as carbon steel and copper, have a relatively high tolerance for contaminants in the shielding gas; others, such as aluminum and magnesium, are fairly sensitive to particular contaminants. Still others, such as titanium and zirconium, have an extremely low tolerance for any foreign constituent in the shielding gas.

Depending on the metal being welded and the welding process used, very small quantities of gas impurities can significantly affect welding speed, weld surface appearance, weld bead solidification, and porosity levels. The effects of any given impurity are wide ranging, but weld quality and eventual fitness for purpose are major areas of concern.

There is always a possibility that the gas, as delivered, is contaminated; however, it is far more likely that impurities will enters somewhere between the supply and the end-use points. For this reason, property designed piping systems and high-quality hose are recommended for use with welding shielding gases. Typical industry minimum purity levels for welding gases are listed in Table 2 (Ref 2).

TABLE 2 TYPICAL GASES PURITY AND MOISTURE CONTENT OF SHIELDING

GAS	PRODUCT STATE	MINIMUM PURITY, %	MAXIMUM MOISTURE, PPM[(A)]	APPROXIMATE DEW POINT AT MAXIMUM MOISTURE CONTENT	
				°C	°F
ARGON	GAS	99.995	10	-60	-77
	LIQUID	99.997	6	-64	-83
CARBON DIOXIDE	GAS	99.5	19	-51	-60
	LIQUID	99.8	50	-58	-73
HELIUM	GAS	99.95	32	-51	-61
	LIQUID	99.995	3	-69	-92
HYDROGEN	GAS	99.95	8	-63	-80
	LIQUID	99.995	5	-65	-86
NITROGEN	GAS	99.7	32	-51	-61
	LIQUID	99.997	5	-65	-86
OXYGEN	INDUSTRIAL	99.5	50	-48	-54
	LIQUID	99.5	6	-64	-83

Source: Ref 2

(A) MOISTURE SPECIFICATIONS ARE MEASURED AT FULL CYLINDER PRESSURE, THE PRESSURE AT WHICH THE CYLINDER IS ANALYZED.

Gas density is the weight of the gas per unit volume. Density is one of the chief factors that influence shielding gas effectiveness. Basically, gases heavier than air, such as argon and carbon dioxide, require lower flow rates in use than do the lighter gases, such as helium, to ensure adequate protection of the weld puddle.

References cited in this section

1. N.E. LARSON AND W.F. MEREDITH, *SHIELDING GAS SELECTION MANUAL*, UNION CARBIDE INDUSTRIAL GASES TECHNOLOGY CORP., 1990, P 10
2. N.E. LARSON AND W.F. MEREDITH, *SHIELDING GAS SELECTION MANUAL*, UNION CARBIDE INDUSTRIAL GASES TECHNOLOGY CORP., 1990, P 11

Shielding Gases for Welding

Kevin A. Lyttle, Praxair, Inc.

Characteristics of the Components of a Shielding Gas Blend

To obtain a shielding gas that is suited to a specific application, a mix of gases is generally needed. Each basic gas contributes certain characteristics to the performance of the overall mix. Some gas blends have relatively specific areas of application and limited operating ranges; others can be used on many materials under a variety of welding conditions. Each component of the blend brings with it properties that are supplemented by the others to produce an enhanced level of performance.

Argon is inert or unreactive with respect to the materials present in the welding electrode. With its low ionization potential, argon promotes easy arc starting and stable arc operation. Its lower thermal conductivity promotes the development of axial "spray" transfer in certain forms of GMAW. It is also used in applications where base material distortion must be controlled or where good gap-bridging ability is required.

Helium. Unlike argon, helium is lighter than air and has a low density. Like argon, it is chemically inert and does not react with other elements or compounds. Because of its high thermal conductivity and high ionization potential, more heat is transferred to the base material, thus enhancing the penetration characteristics of the arc. In many applications, it also allows higher weld travel speeds to be obtained. Because of its higher cost, helium is frequently combined with argon or argon mixtures to enhance the overall performance of the blend while minimizing its cost.

Oxygen combines with almost all known elements except rare and inert gases; it vigorously support combustion. Small amounts of oxygen are added to some inert mixtures to improve the stability of the welding arc developed as well as to increase the fluidity of the weld puddle.

In the spray-transfer mode of GMAW, small additions of oxygen enhance the range over which this spatterless form of welding can be performed. The droplet size decreases and the number of drops transferred per unit time increases as oxygen is added to the blend.

Carbon dioxide is a reactive gas that is commonly used alone in certain types of GMAW. Oxidation of the base material and any filler electrode occurs readily. Carbon dioxide is added to argon blends to improve are stability, enhance penetration, and improve weld puddle flow characteristics. The higher thermal conductivity of carbon dioxide (because of the dissociation and recombination of its component parts) transfer more heat to the base material than does argon alone. A broader penetration pattern versus argon is obtained; however, base material distortion and lack of gap-bridging ability are possible problems.

Hydrogen is the lightest known element and is a flammable gas. Explosive mixtures can be formed when certain concentration of hydrogen are mixed with oxygen or air. It is added to inert gases to increase the heat input to the base material or for operations involving cutting and gouging. Because some materials are especially sensitive to hydrogen-related contamination, its use is generally limited to special applications, such as the joining of stainless steels, and to plasma are cutting and gouging.

Nitrogen is generally considered to be inert except at high temperatures. At arc welding temperatures, it will react with some metals (e.g., aluminum, magnesium, steel, and titanium), so it is not used as a primary shielding gas. It can be used with other gases for some welding applications (e.g., copper) and is also widely used in plasma cutting.

Shielding Gases for Welding

Kevin A. Lyttle, Praxair, Inc.

Shielding Gas Selection

In most welding applications, more than one shielding gas or gas blend can be used successfully. For example, there is no one optimal gas blend for joining carbon steels, but a considerable array of mixes are available depending on the specific requirements of the application. For some processes, such as gas-tungsten arc welding (GTAW) and plasma arc welding (PAW), the choices may be somewhat limited by the nature of the electrodes used and the materials being welded. However, for applications involving GMAW, a multitude of blends can be selected from when carbon steels are being joined. Determination of the best blend depends on a number of specific job-related needs.

Accuracy of Gas Blends

The accuracy with which gases are blended is a function of the way in which they are supplied. If the source of the gas is a high-pressure cylinder, the following generally applies:

1. ±10% RELATIVE, MINOR COMPONENT (REF 3)

2. ±0.5% ABSOLUTE FOR CONCENTRATIONS UP TO 5%; ±10% RELATIVE FOR CONCENTRATIONS BETWEEN 5 AND 50% (REF 4)

For example, the mix accuracy of an Ar-2O_2 blend would be Ar/1.8-2.2O_2 (method 1) or Ar/1.5-2.5O_2 (method 2). A blend of Ar-25CO_2 would yield Ar/22.5-27.5CO_2 by their method of calculation.

When gas cylinders are properly filled with the appropriate blend, the components of that mixture will not separate unless the temperature of the environment is reduced far below normal working temperatures. If the gases are supplied from a liquid source, such as a bulk tank, the accuracy of the blend is a function of the mixing equipment used, but most likely falls within the ±10% minor component range.

Shielding Gases for GMAW

By far, the largest number of gas blends have been developed for GMAW, especially for joining carbon steel. These can be roughly divided into four categories: pure gases, argon-oxygen mixes, argon/carbon dioxide mixes, and three-part gas blends composed of either argon, helium, oxygen, carbon dioxide, or hydrogen. Table 3 contains suggestions for shielding gas selection based on material type, thickness, and mode of metal transfer.

TABLE 3 RECOMMENDED SHIELDING GAS SELECTION FOR GMAW

MATERIAL	THICKNESS		TRANSFER MODE	RECOMMENDED SHIELDING GAS	ADVANTAGES AND LIMITATIONS
	MM	IN.			
CARBON STEEL	<2.0	<0.080	SHORT CIRCUITING	AR-25CO_2 AR-15CO_2 AR-8CO_2	GOOD PENETRATION AND DISTORTION CONTROL TO REDUCE POTENTIAL BURNTHROUGH
	2.0-3.2	0.080-0.125	SHORT CIRCUITING	AR-8CO_2 AR-15CO_2 AR-25CO_2	HIGHER DEPOSITION RATES WITHOUT BURNTHROUGH: MINIMUM DISTORTION AND SPATTER; GOOD PUDDLE CONTROL FOR OUT-OF-POSITION WELDING
	>3.2	>0.125	SHORT CIRCUITING	AR-15CO_2 AR-25CO_2 CO_2	HIGH WELDING SPEEDS, GOOD PENETRATION AND PUDDLE CONTROL; APPLICABLE LOT OUT-OF-POSITION WELDS
			GLOBULAR	AR-25CO_2 CO_2	SUITABLE FOR HIGH-CURRENT AND HIGH-SPEED WELDING; DEEP PENETRATION AND FAST TRAVEL SPEEDS, BUT WITH GREATER BURNTHROUGH POTENTIAL
			CONVENTIONAL SPRAY ARC	AR-1O_2 AR-2O_2 AR-5CO_2	GOOD ARC STABILITY; PRODUCES A MORE FLUID PUDDLE AS O_2

			AR-8CO_2 AR-10CO_2 AR-15CO_2 AR-CO_2-O_2 BLENDS	INCREASES; GOOD COALESCENCE AND BEAD CONTOUR, GOOD WELD APPEARANCE AND PUDDLE CONTROL
		PULSED SPRAY	ARGON-5CO_2 AR-HE-CO_2 BLENDS AR-CO_2-O_2 BLENDS	USED FOR BOTH GAGE AND OUT-OF-POSITION WELDMENTS; ACHIEVES GOOD PULSED SPRAY STABILITY OVER A WIDE RANGE OF ARC CHARACTERISTICS AND DEPOSITION RANGES
ALLOY STEEL	ALL SIZES	SHORT CIRCUITING	AR-8CO_2 AR-15CO_2 AR-CO_2-O_2 BLEND	HIGH WELDING SPEEDS; GOOD PENETRATION AND PUDDLE CONTROL; APPLICABLE FOR OUT-OF-POSITION WELDS; SUITABLE FOR HIGH-CURRENT AND HIGH-SPEED WELDING
		SPRAY ARC (HIGH-CURRENT DENSITY AND ROTATIONAL)	AR-2O_2 AR-5O_2 AR-CO_2-O_2 BLENDS AR-HE-CO_2 BLENDS	REDUCES UNDERCUTTING; HIGHER DEPOSITION RATES AND IMPROVED BEAD WETTING; DEEP PENETRATION AND GOOD MECHANICAL PROPERTIES
		PULSED SPRAY	AR-5CO_2 AR-8CO_2 AR-2O_2	USED FOR BOTH LIGHT-GAGE AND HEAVY OUT-OF-POSITION WELDMENTS; ACHIEVES GOOD PULSED SPRAY STABILITY OVER A WIDE RANGE OF ARC CHARACTERISTICS AND DEPOSITION RANGES
STAINLESS STEEL, COPPER, NICKEL, AND CU-NI ALLOYS	ALL SIZES	SHORT-CIRCUITING TRANSFER	AR-HE-CO_2 BLENDS HE-AR-CO_2 BLENDS AR-1O_2 AR-2O_2	LOW CO_2 CONTENTS IN HELIUM MIX MINIMIZE CARBON PICKUP, WHICH CAN CAUSE INTERGRANULAR CORROSION WITH SOME ALLOYS; HELIUM IMPROVES WETTING ACTION; CO_2 CONTENTS >5% SHOULD BE USED WITH CAUTION ON

					SOME ALLOYS; APPLICABLE FOR ALL POSITION WELDING
			SPRAY ARC	AR-HE-CO_2 BLENDS AR-1O_2 AR-2O_2	GOOD ARC STABILITY; PRODUCES A FLUID BUT CONTROLLABLE WELD PUDDLE; GOOD COALESCENCE AND BEAD CONTOUR; MINIMIZES UNDERCUTTING ON HEAVIER THICKNESSES
			PULSED SPRAY	AR-HE-CO_2BLENDS AR-1O_2 AR-2O_2	USED FOR BOTH LIGHT-GAGE AND HEAVY OUT-OF-POSITION WELDMENTS; ACHIEVES GOOD PULSED SPRAY STABILITY OVER A WIDE RANGE OF ARC CHARACTERISTICS AND DEPOSITION RANGES
ALUMINUM, TITANIUM, AND OTHER REACTIVE METALS	≤ 13	$\le \frac{1}{2}$	SPRAY ARC	ARGON	BEST METAL TRANSFER, ARC STABILITY, AND PLATE CLEANING; LITTLE OR NO SPATTER; REMOVES OXIDES WHEN USED WITH DCEP (REVERSE POLARITY)
			SPRAY ARC	75HE-25AR 50HE-50AR	HIGH HEAT INPUT; PRODUCES FLUID PUDDLE, FLAT BEAD CONTOUR, AND DEEP PENETRATION; MINIMIZES POROSITY
	>13	$> \frac{1}{2}$	SPRAY ARC	HELIUM 50HE-25AR	HIGH HEAT INPUT; GOOD FOR MECHANIZED WELDING AND OVERHEAD; APPLICABLE TO HEAVY SECTION WELDING
			PULSED SPRAY	ARGON	GOOD WETTING; GOOD PUDDLE CONTROL

Argon. Pure argon is generally used on nonferrous base metal, such as aluminum, nickel, copper, and magnesium alloys, and on reactive metals, such as titanium. Argon provides excellent arc stability, penetration, and bead profile when joining these materials. Its low ionization potential results in easy arc starting. Argon produces a constricted arc column with high current density, which concentrates the arc energy over a small area; deep, fingerlike penetration results.

Carbon Dioxide. A reactive gas, carbon dioxide is generally used only for joining carbon steel. It is readily available and relatively inexpensive. Because CO_2 will not support spray transfer, deposition efficiency is lower and spatter and fume levels are higher than with argon blends. Weld bead surfaces are more oxidized and irregular in shape. The higher ionization potential of CO_2 and its characteristic dissociation upon heating provide greater weld fusion and penetration while still achieving acceptable mechanical properties.

Helium. Because of its higher thermal conductivity, helium can provide additional heat input to the base material while still maintaining an inert atmosphere. Wetting action, depth of fusion, and travel speed can be improved over comparable argon levels. This advantage is most frequently utilized in the welding of heavier sections of aluminum, magnesium, and copper alloys.

Argon-Oxygen. The addition of a small amount of oxygen to a argon greatly stabilizes the welding arc, increases the filler metal droplet rate, lowers the spray transition current, and influences bead shape. The weld pool is more fluid and stays molten longer, allowing the metal to flow out toward the edges of the weld (Fig. 1a).

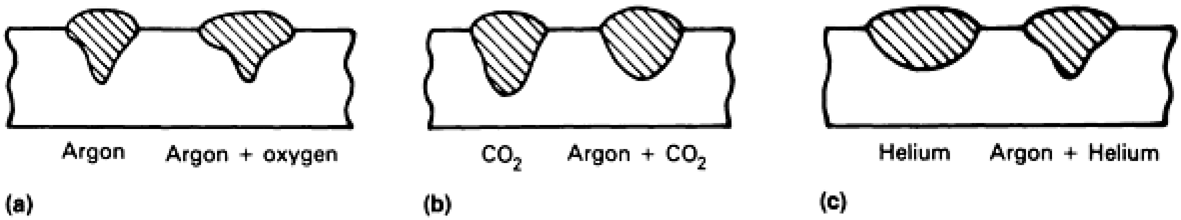

FIG. 1 EFFECT OF SHIELDING GAS BLENDS ON WELD PROFILE USING DIRECT CURRENT ELECTRODE POSITIVE (DCEP). (A) ARGON VERSUS ARGON-OXYGEN. (B) CARBON DIOXIDE VERSUS ARGON/CARBON DIOXIDE. (C) HELIUM VERSUS ARGON-HELIUM. SOURCE: REF 5

The most common blends contain 1, 2, 5, or 8% O_2 in argon. Increasing oxygen improves arc stability and makes higher travel speeds possible by enhanced puddle fluidity. Some increased alloy loss and a greater chance of undercut occur as the oxygen level is increased, especially beyond 5%.

Argon/carbon dioxide blends are primarily used for carbon and low-alloy steels and have limited use for stainless steels. The addition of CO_2 to argon produces results similar to the addition of oxygen, but is also broadens the penetration pattern as the CO_2 content is increased (Fig. 1b). Above a range of 18 to 20% CO_2, spraylike transfer can no longer be obtained; short-circuiting/globular transfer with somewhat increased spatter levels is found from this point up to approximately 50% CO_2 in argon.

The most common blends for spray transfer are argon plus 5, 8, 10, or 13 to 18% CO_2. With increased CO_2 content, the more fluid weld puddle permits higher weld travel speeds.

Mixes with higher carbon dioxide levels can also be used for short-circuiting transfer--commonly, argon plus 20 or 25% CO_2. Mixtures in this range provide an optimum droplet frequency for minimum spatter when small-diameter (0.9 and 1.2 mm, or 0.035 and 0.045 in.) wire is used.

Argon-Helium. Helium is often mixed with argon to obtain the advantages of both gases. These blends are primarily used for nonferrous base materials, such as aluminum, copper, and nickel alloys. Helium increases the heat input to the base material and thus is used for joining thick, thermally conductive plates. As the helium percentage increases, the arc voltage, spatter, and weld width-to-depth ratio increase (Fig. 1c).

The most common blends contain 25, 50, or 75% He in argon. The highest percentage of helium is used for joining thick (>50 mm, or 2 in.) plate material, especially aluminum and copper. Higher travel speeds can be obtained using helium-enhanced blends.

Argon/Oxygen/Carbon Dioxide. Mixtures containing three gas components are versatile because of their ability to operate in short-circuiting, globular, and spray-transfer modes. These blends are generally proprietary, and manufacturers' recommendations should be followed for their proper use. Blends of argon, carbon dioxide, and oxygen are generally used to join carbon and alloy steels.

Argon/Helium/Carbon Dioxide. Helium and carbon dioxide additions to argon increase the heat input to the base metal, which improves wetting, puddle fluidity, and weld bead profile. With helium plus CO_2 additions less than 40%, good spray transfer is obtained for carbon and low-alloy steel welding. Some increase in tolerance of base material surface contamination is also noted.

When the helium content exceeds 50 to 60%, transfer is restricted to short-circuiting and globular. Blends in which the CO_2 content is relatively low ($\leq 5\%$) are generally used for the joining of stainless steels without any loss in corrosion resistance.

Argon/Carbon Dioxide/Hydrogen. Three-part blends of this design are intended for the joining of austenitic stainless in the spray or short-circuiting transfer modes. Because of the addition of hydrogen, these blends should not be used for carbon steel. The carbon dioxide and hydrogen increase the heat input to the base material and improve bead shape characteristics, as well as promote higher welding travel speeds.

Shielding Gas Selection for FCAW

Carbon Dioxide. The majority of large-diameter (>1.6 mm, or $\frac{1}{16}$ in.) wires that use a shielding gas use carbon dioxide. Some smaller diameter wires are formulated to operate in 100% CO_2. The arcs are generally stable and provide a globular transfer over the usable operating range. Good performance over rust and mill scale on the plate surface is obtained with these large-diameter wires and CO_2 shielding.

Argon/Carbon Dioxide. A significant number of small-diameter (≤ 1.6 mm, or $\frac{1}{16}$ in.) cored wires are shielded with the blends of argon with 15 to 50% CO_2. These blends provide better out-of-position weld puddle control versus that of CO_2. To obtain the best performance from a particular cored wire, check the manufacturer's product literature for the recommended gas blend.

Shielding Gas Selection for GTAW

Argon. The most commonly used gas for GTAW, argon exhibits low thermal conductivity, which produces a narrow, constricted arc column; this allows greater variations in arc length with minimal influence on arc power or weld bead shape. Its low ionization potential provides good arc starting characteristics and good arc stability using the direct current electrode negative (DCEN) power connection plus superior arc cleaning action and bead appearance when ac power is used. Argon is the most commonly selected gas for DCEN welding of most materials and ac manual welding of aluminum.

Helium. The high thermal conductivity and ionization potential of helium make it suitable for the high-current joining of heavy sections of heat-conductive materials such as aluminum. Helium increases the penetration of the weld as well as its width. It also allows the use of higher weld travel speeds.

Argon-Helium. Blends of argon and helium are selected to increase the heat input to the base material while maintaining favorable arc stability and superior arc starting characteristics. Blends of 25, 50, and 75% He in argon are commonly used.

Argon-Hydrogen. Hydrogen is added to argon to enhance its thermal properties. The slightly reducing atmosphere improves weld puddle wetting and reduces some surface oxides to produce a cleaner weld surface. To minimize problems associated with arc starting, additions of hydrogen are generally limited to 5 to 15%. These blends are primarily used to join some stainless steels, nickel, and nickel alloys. These mixtures should not be used to join alloy steels; delayed weld cracking may result.

Argon/2-5H_2 is used in manual welding applications on materials thicker than 1.6 mm ($\frac{1}{16}$ in.). Additions of 10 to 15% H_2 are used in mechanized applications, such as those found in the manufacture of stainless steel tubing.

Caution: Special safety precautions are required when mixing argon and hydrogen. Do not attempt to mix argon and hydrogen from separate cylinders. Always purchase ready-mixed hydrogen blends from a qualified supplier.

Shielding Gas Selection for PAW

The physical configuration of the PAW system requires the use of two gases: a "plasma" or orifice gas and a shielding gas. The primary role of the plasma gas, which exits the torch through the center orifice, is to control arc characteristics and shield the electrode. The shielding gas, introduced around the periphery of the arc, shields or protects the weld area. In many applications, the shielding gas is also partially ionized to enhance the performance of the plasma gas.

Low-Current (<100 A) PAW. Argon is the preferred plasma gas for low-current PAW because of its low ionization potential, which ensures easy and reliable arc starting. Argon-hellium mixtures are used for some applications requiring higher heat inputs. The choice of the shielding gas depends on the type and thickness of the base material. Recommendations can be found in Table 4.

TABLE 4 RECOMMENDED GUIDELINES FOR SELECTING PAW SHIELDING GASES

MATERIAL	THICKNESS		MODE OF PENETRATION	
	MM	IN.	KEYHOLE TECHNIQUE(C)	MELT-IN TECHNIQUE(D)
LOW-CURRENT PAW(A)				
ALUMINUM COPPER	<1.6	$<\frac{1}{16}$	NOT RECOMMENDED	ARGON HELIUM
	[GES]1.6	[GES] $\frac{1}{16}$	HELIUM	HELIUM
CARBON STEEL	<1.6	$<\frac{1}{16}$	NOT RECOMMENDED	ARGON HELIUM 75AR-25HE
	[GES]1.6	[GES] $\frac{1}{16}$	ARGON 75HE-25AR	ARGON 75HE-25AR
LOW-ALLOY STEEL	<1.6	$<\frac{1}{16}$	NOT RECOMMENDED	ARGON HELIUM AR-1.5H_2
	[GES]1.6	[GES] $\frac{1}{16}$	ARGON 75HE-25AR AR-1.5H_2	ARGON HELIUM AR-1.5H_2
STAINLESS STEEL, NICKEL ALLOYS	<1.6	$<\frac{1}{16}$	ARGON 75HE-25AR AR-1.5H_2	ARGON HELIUM AR-1.5H_2
	[GES]1.6	[GES] $\frac{1}{16}$	ARGON 75HE-25AR AR-1.5H_2	ARGON HELIUM AR-1.5H_2
HIGH-CURRENT PAW(B)				
ALUMINUM	<6.4	$<\frac{1}{4}$	ARGON	ARGON 75-25AR
	[GES]6.4	[GES] $\frac{1}{4}$	HELIUM	HELIUM 75-25AR
COPPER	<2.4	$<\frac{3}{32}$	NOT RECOMMENDED	HELIUM
CARBON STEEL; LOW-ALLOY STEEL	<3.2	$<\frac{1}{8}$	ARGON	ARGON
	[GES]3.2	[GES]	ARGON	75HE-25AR

		$\frac{1}{8}$		
STAINLESS STEEL; NICKEL ALLOYS	<3.2	$<\frac{1}{8}$	ARGON AR-5H_2	ARGON
	[GES]3.2	[GES] $\frac{1}{8}$	ARGON AR-5H_2	75-HE-25AR

Source: Ref 6, 7

(A) GAS SELECTIONS SHOWN ARE FOR SHIELDING GAS ONLY. ORIFICE GAS IN ALL CASES IS ARGON.

(B) GAS SELECTIONS SHOWN ARE FOR BOTH THE ORIFICE AND SHIELDING GAS.

(C) PROPERLY BALANCED GAS FLOW RATES PRODUCE COMPLETE JOINT PENETRATION BY FORMING A SMALL WELD POOL WITH A HOLE PENETRATING COMPLETELY THROUGH THE BASE METAL. AS THE PLASMA TORCH IS MOVED, METAL MELTED BY THE ARC IS FORCED TO FLOW AROUND THE PLASMA STEAM PRODUCING THE "KEYHOLE" TO THE REAR WHERE THE WELD POOL IS FORMED AND SOLIDIFIED.

(D) CONVENTIONAL FUSION WELDING SIMILAR TO THAT DONE WITH GTAW.

High-Current ([ges]100 A) PAW. The choice of gas used for high-current PAW also depends on the material to be welded. In almost all cases, the shielding gas is the same as the orifice gas. Again, argon is suitable for welding all metals, but it does not necessarily produce optimum results. Depending on the welding mode used (keyhole or melt-in), the optimum gas blend will vary. Table 4 lists gases recommended on the basis of materials joined.

References cited in this section

3. "SPECIFICATION FOR SHIELDING GASES," AWS A5.32-9X, JULY 1991 (DRAFT)
4. "GASES FOR GAS-SHIELDED ARC WELDING AND CUTTING," EUROPEAN STANDARD EN 439, DEC 1990 (DRAFT)
5. H.B. CARY, *MODERN WELDING TECHNOLOGY*, 2ND ED., PRENTICE-HALL, 1989
6. N.E. LARSON AND W.F. MEREDITH, *SHIELDING GAS SELECTION MANUAL*, UNION CARBIDE INDUSTRIAL GASES TECHNOLOGY CORP., 1990, P 30-32
7. N.E. LARSON AND W.F. MEREDITH, *SHIELDING GAS SELECTION MANUAL*, UNION CARBIDE INDUSTRIAL GASES TECHNOLOGY CORP., 1990, P 18-19

Shielding Gases for Welding

Kevin A. Lyttle, Praxair, Inc.

Influence of Shielding Gas on Weld Mechanical Properties

The mechanical properties of a weld are dependent on specific characteristics of the shielding gas. When the gas blend is totally inert, the effects are less pronounced and are derived more indirectly. In this case, the shielding gas affects penetration and solidification, which can influence the microstructure of the resulting weld.

When the shielding gas contains active components, such as oxygen or carbon dioxide, the influence is direct and more substantial. The oxygen potential of the shielding gas influences the amount of surface slag, the fume emission rate, the fluidity of the weld puddle, and the mechanical properties (both strength and toughness) of the weld metal. A number of empirical formulas have been developed to estimate the oxygen potential of a gas blend (Ref 8); the differences lie in their treatment of the CO_2 component of the gas mix. Because parameters other than gas composition can affect the oxygen level of the weld metal (e.g., welding speed), most of the recently developed formulas are also welding parameter and material specific. Their importance centers on how the oxidation potential is linked to elemental loss of silicon and manganese in the weld metal, to weld metal oxygen content, and to weld mechanical properties.

Because a relatively complex relationship exists between the loss of alloying elements, the composition of the shielding gas, and the mechanical properties of the resulting weld metal it is difficult to select an optimum gas blend that will work well with all types of wires. This influence is most noticeable in the gas metal-arc welding of carbon steel. Figure 2 (Ref 8) and Table 5 (Ref 9) illustrate the effect of the shielding gas on impact strength and tensile strength. In general, as the oxidation potential of the shielding gas increases, the toughness and the tensile strength of the weld deposit decrease. Because of their lower oxidizing gas content, argon blends will generally produce weld properties superior to those obtained by shielding with CO_2 only. There appears to be an "optimum" oxygen content, as too low an oxygen level can also be detrimental to toughness.

TABLE 5 EFFECT OF SHIELDING GAS ON CARBON, MANGANESE, AND SILICON LOSSES AND ON WELD STRENGTH VALUES

SHIELDING GAS(A)	**ULTIMATE TENSILE STRENGTH**		**YIELD STRENGTH**		**ELONGATION, %**	**WELD METAL COMPOSITION, %(B)**		
	MPA	**KSI**	**MPA**	**KSI**		**C**	**MN**	**SI**
AR-10CO_2	640	92.9	544	79.0	25.7	0.09	1.43	0.72
AR-18CO_2	620	90.0	522	75.8	26.8	0.09	1.37	0.70
AR-5CO_2-4O_2	610	88.5	472	68.5	28.1	0.08	1.32	0.67
AR-25CO_2	601	87.2	505	73.3	29.3	0.09	1.30	0.65
AR-12O_2	591	85.8	510	74.0	27.5	0.06	1.20	0.60
CO_2	594	86.2	487	70.7	27.8	0.10	1.21	0.62

Source: Ref 9

(A) GASES LISTED IN ORDER OF INCREASING OXIDATION POTENTIAL.

(B) BASE WIRE COMPOSITION: 0.115% C, 1.53% MN, 0.98% SI.

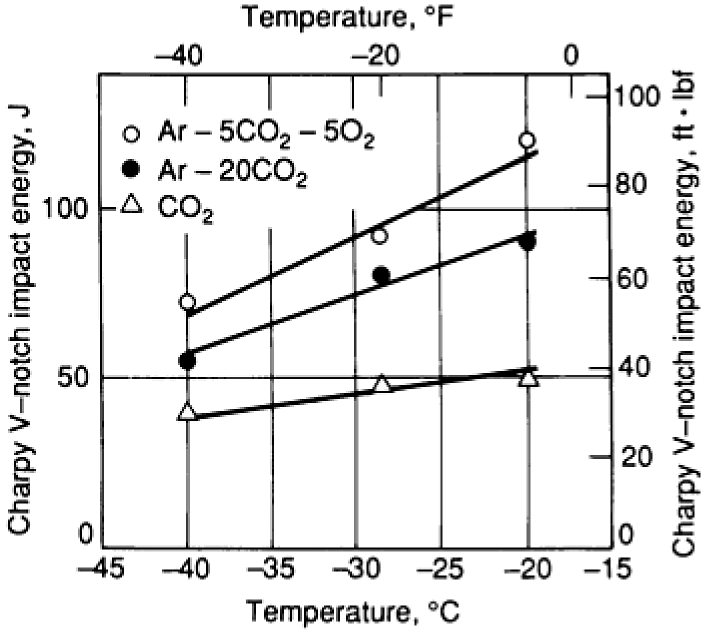

FIG. 2 PLOT OF WELD METAL IMPACT ENERGY VERSUS TEST TEMPERATURE AS A FUNCTION OF SHIELDING GAS COMPOSITION. SOURCE: REF 8

References cited in this section

8. N. STENBACKA AND K.A. PERSSON, SHIELDING GASES FOR GAS METAL ARC WELDING, *WELD.J.*, VOL 68, NOV 1989, P 41-47
9. H.U. POMASKA, "SHIELDING GASES FOR ARC WELDING AND CUTTING," SONDERDRUCK 48, LINDE AG, LINDE TECHNISCHE GASES, HOLRIEGELSKREUTH, GERMANY

Shielding Gases for Welding

Kevin A. Lyttle, Praxair, Inc.

Shielding Gas and Fume Generation

The shielding gas used in solid and flux-cored wire welding affects the rate at which fumes are produced during welding, as well as the composition of the fumes. The type of shielding gas also affects the composition of the pollutant gases in the welding area. The fumes and gases generated in some welding applications impact on health and safety; therefore, it is important for the user to review the Material Safety Data Sheets and precautionary labeling provided by the manufacturers and suppliers of welding consumables. Welding safety and health information can be also be found in this Handbook and in ANSI Z49.1, "Safety in Welding and Cutting."

Fume generation rates are generally highest when CO_2 shielding is used. Inert gas blends containing CO_2 produce higher fume levels than those containing only oxygen additions. For solid wire welding at the same weld metal deposition rate, argon blends typically generate significantly less fumes than when CO_2 only is used (Ref 10, 11, 12). Figure 3 (Ref 13) illustrates the dependence of the fume generation rate on shielding gas composition as a function of current level in GMAW. Similar trends can be seen when using gas-shielded flux-cored wires.

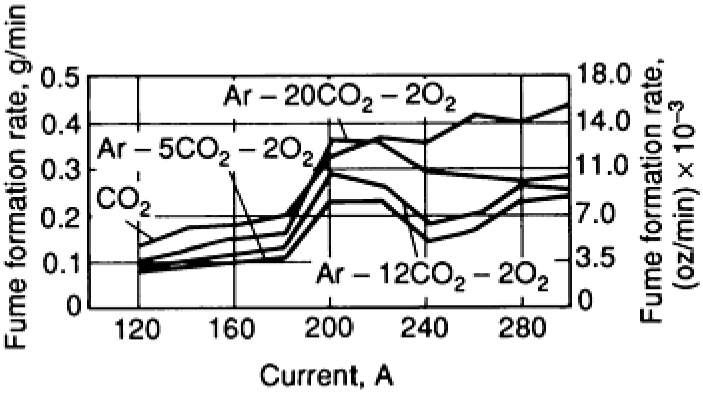

FIG. 3 PLOT OF FUME FORMATION RATE VERSUS CURRENT FOR MILD STEEL SOLID WIRE USING SELECTED SHIELDING GASES. SOURCE: REF 13

Pollutant gases must be considered when arc welding. The gases of greatest interest arc carbon monoxide, ozone, and nitrogen oxides. When CO_2 is used for shielding in confined spaces, CO can be a potential problem. Ozone can be a concern when high-energy gas-shielded welding is conducted, particularly on aluminum and stainless steel plate. It is produced in the immediate arc area as well as in the surrounding environment. Oxides of nitrogen can be present in some plasma welding and cutting applications (Ref 14).

References cited in this section

10. R.F. HEILE AND D.C. HILL, PARTICULATE FUME GENERATION IN ARC WELDING PROCESSES, *WELD.J.*, VOL 54, JULY 1975, P 201S-210S
11. H. PRESS AND W. FLORIAN, CONTRIBUTION ON THE FORMATION OF TOXIC SUBSTANCES IN GAS SHIELDED ARC WELDING AND DETERMINATION OF DESIRED EXHAUSTION CAPACITY, *IIW PROCEEDINGS OF COLLOQUIUM ON WELDING AND HEALTH*, PORTUGAL, JULY 1980
12. B. HAAS, INFLUENCE OF PROCESS SPECIFIC WELDING PARAMETERS IN FUME GENERATION IN SOLID WIRE, GMA/MAG WELDING, *IIW PROCEEDINGS OF COLLOQUIUM ON WELDING SAFETY AND HEALTH*, PORTUGAL, JULY 1980
13. D.E. HILTON AND P.N. PLUMRIDGE, PARTICULATE FUME GENERATION DURING GMAW AND GTAW, *WELD. MET. FABR.*, VOL 59 (NO. 10), DEC 1991, P 555-560
14. N. JENKINS, J. MORETON, P.J. OAKLEY, AND S.M. STEVENS, *WELDING FUMES SOURCES, CHARACTERISTICS, CONTROL*, VOL 2, THE WELDING INSTITUTE, CAMBRIDGE, UNITED KINGDOM, 1981, P 272

Self-Shielded Flux-Cored Arc Welding

The self-shielded flux-cored wire welding process employs a continuous wire electrode that requires no external shielding. These cored wires generate protective shielding gases from components in the core material similar to those found in coated electrodes. The unique feature of this class of consumables is that they rely only partly on the exclusion of air from the molten electrode and weld pool to produce a quality deposit. In addition to deoxidizers, they contain denitriders to react with nitrogen that may be entrained in the molten metal. The less a given consumable relies on shielding and the more it relies on "killing" to control nitrogen and produce sound weld metal, the less that consumable will be affected by cross air currents and side winds.

Through careful control of welding parameters and the proper balance of core constituents, good-quality weld metal can be obtained under poor welding conditions. Self-shielded flux-cored wires are especially suited to welding outdoors, where it is difficult to provide acceptable external gas shielding. Both light- and heavy-gage materials have been successfully joined with these electrodes.

References

1. N.E. LARSON AND W.F. MEREDITH, *SHIELDING GAS SELECTION MANUAL*, UNION CARBIDE INDUSTRIAL GASES TECHNOLOGY CORP., 1990, P 10
2. N.E. LARSON AND W.F. MEREDITH, *SHIELDING GAS SELECTION MANUAL*, UNION CARBIDE INDUSTRIAL GASES TECHNOLOGY CORP., 1990, P 11
3. "SPECIFICATION FOR SHIELDING GASES," AWS A5.32-9X, JULY 1991 (DRAFT)
4. "GASES FOR GAS-SHIELDED ARC WELDING AND CUTTING," EUROPEAN STANDARD EN 439, DEC 1990 (DRAFT)

5. H.B. CARY, *MODERN WELDING TECHNOLOGY*, 2ND ED., PRENTICE-HALL, 1989
6. N.E. LARSON AND W.F. MEREDITH, *SHIELDING GAS SELECTION MANUAL*, UNION CARBIDE INDUSTRIAL GASES TECHNOLOGY CORP., 1990, P 30-32
7. N.E. LARSON AND W.F. MEREDITH, *SHIELDING GAS SELECTION MANUAL*, UNION CARBIDE INDUSTRIAL GASES TECHNOLOGY CORP., 1990, P 18-19
8. N. STENBACKA AND K.A. PERSSON, SHIELDING GASES FOR GAS METAL ARC WELDING, *WELD.J.*, VOL 68, NOV 1989, P 41-47
9. H.U. POMASKA, "SHIELDING GASES FOR ARC WELDING AND CUTTING," SONDERDRUCK 48, LINDE AG, LINDE TECHNISCHE GASES, HOLRIEGELSKREUTH, GERMANY
10. R.F. HEILE AND D.C. HILL, PARTICULATE FUME GENERATION IN ARC WELDING PROCESSES, *WELD.J.*, VOL 54, JULY 1975, P 201S-210S
11. H. PRESS AND W. FLORIAN, CONTRIBUTION ON THE FORMATION OF TOXIC SUBSTANCES IN GAS SHIELDED ARC WELDING AND DETERMINATION OF DESIRED EXHAUSTION CAPACITY, *IIW PROCEEDINGS OF COLLOQUIUM ON WELDING AND HEALTH*, PORTUGAL, JULY 1980
12. B. HAAS, INFLUENCE OF PROCESS SPECIFIC WELDING PARAMETERS IN FUME GENERATION IN SOLID WIRE, GMA/MAG WELDING, *IIW PROCEEDINGS OF COLLOQUIUM ON WELDING SAFETY AND HEALTH*, PORTUGAL, JULY 1980
13. D.E. HILTON AND P.N. PLUMRIDGE, PARTICULATE FUME GENERATION DURING GMAW AND GTAW, *WELD. MET. FABR.*, VOL 59 (NO. 10), DEC 1991, P 555-560
14. N. JENKINS, J. MORETON, P.J. OAKLEY, AND S.M. STEVENS, *WELDING FUMES SOURCES, CHARACTERISTICS, CONTROL*, VOL 2, THE WELDING INSTITUTE, CAMBRIDGE, UNITED KINGDOM, 1981, P 272

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

Introduction

SOLID-STATE TRANSFORMATIONS occurring in a weld are highly nonequilibrium in nature and differ distinctly from those experienced during casting, thermomechanical processing, and heat treatment. This discussion will primarily focus on the welding metallurgy of fusion welding of steels and attempt to highlight the fundamental principles that form the basis of many of the recent developments in steels and consumables for welding. Accordingly, examples are largely drawn from the well-known and relatively well-studied case of ferritic steel weldments to illustrate the special physical metallurgical considerations brought about by the weld thermal cycles and by the welding environment. Because of space limitations, only a very brief discussion is included on welds in other alloy systems such as stainless steels and aluminum-base, nickel-base, and titanium-base alloys. For detailed information on how the principles explained in the first part of this article (using steel weldments as an example) are applicable to other materials, the reader is referred to the Sections where individual alloy systems are discussed (see "Selection of Stainless Steels," "Selection of Nonferrous Low-Temperature Materials," "Selection of Nonferrous High-Temperature Materials," and "Selection of Nonferrous Corrosion-Resistant Materials" in this Volume).

A concise method of describing the transformation behavior of a steel is by a continuous cooling transformation (CCT) diagram (Fig. 1). However, a conventional CCT diagram such as the one shown in Fig. 1 cannot be used to accurately describe the transformation behavior in a weldment of the same material because weld thermal cycles are very different from those used for generating conventional CCT diagrams.

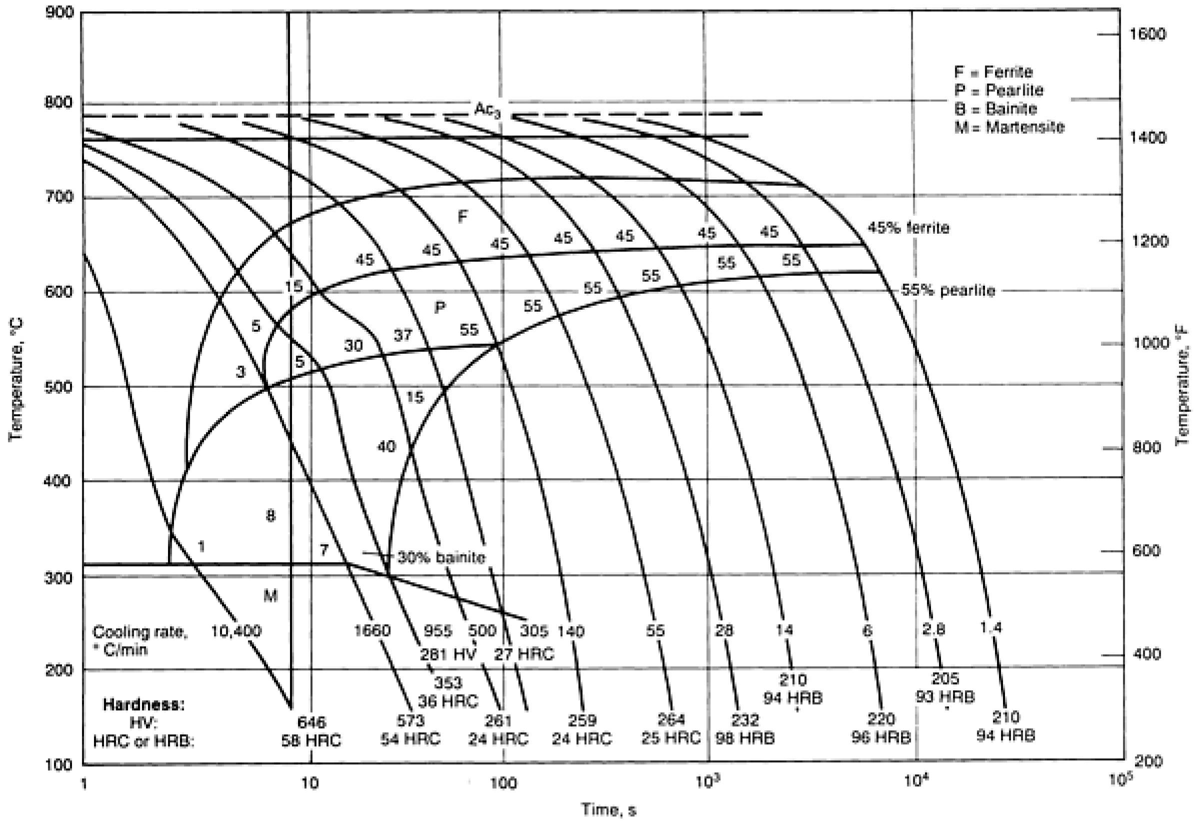

FIG. 1 CONVENTIONAL CCT DIAGRAM FOR AISI 1541 (0.39C-1.56MN-0.21SI-0.24S-0.010P) PLAIN CARBON STEEL WITH ASTM NUMBER 8 GRAIN SIZE AND AUSTENITIZATION AT 980 °C (1800 °F). FOR EACH OF THE COOLING CURVES IN THE PLOT (GIVEN IN TERMS OF °C/MIN), THE TRANSFORMATION START AND END TEMPERATURES GIVEN BY THE CCT CURVES, THE AMOUNT OF EACH TRANSFORMATION PRODUCT, AND THE HARDNESS OF THE FINAL STRUCTURE ARE SHOWN. AC_3, 788 °C (1450 °F); AC_1, 716 °C (1321 °F). F, FERRITE; P, PEARLITE; B, BAINITE; M, MARTENSITE. SOURCE: REF 1

Acknowledgements

The author would like to convey his immense gratitude to the late Professor K. Easterling, University of Exeter, U.K., for giving him the opportunity to write this article and for his constant support and encouragement. The author thanks Dr. Bengt Loberg and Professor Richard Warren of the Department of Engineering Materials, University of Luleå, for their active encouragement. The author would also like to thank Dr. H.K.D.H. Bhadeshia, University of Cambridge, U.K., and Dr. K. Sampath, Concurrent Technologies Corporation, Johnstown, PA, for valuable technical consultation.

Reference

1. *ATLAS OF ISOTHERMAL TRANSFORMATION AND COOLING TRANSFORMATION DIAGRAMS* AMERICAN SOCIETY FOR METALS, 1977

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

Special Factors Affecting Transformation Behavior in a Weldment

Several aspects of the weld thermal cycle and weld segregation should be considered because of their effect on the transformation upon cooling:

- PEAK TEMPERATURES REACHED IN THE HEAT-AFFECTED ZONE (HAZ) CAN BE VERY MUCH HIGHER THAN THE AC_3 TEMPERATURE (THAT IS, THE TEMPERATURE AT WHICH TRANSFORMATION OF FERRITE TO AUSTENITE IS COMPLETED DURING HEATING). THE HEATING RATES ARE VERY HIGH, AND THE TIMES SPENT AT HIGH TEMPERATURE ARE ONLY OF THE ORDER OF A FEW SECONDS.
- THE TEMPERATURE GRADIENT IN THE HAZ IS VERY STEEP, AND THIS COMPLICATES THE PROBLEM OF STUDYING IN *IN SITU* TRANSFORMATIONS IN THE HAZ DURING WELDING (REF 2).
- DURING SOLIDIFICATION OF THE WELD METAL, ALLOYING AND IMPURITY ELEMENTS TEND TO SEGREGATE EXTENSIVELY TO THE INTERDENDRITIC OR INTERCELLULAR REGIONS UNDER THE CONDITIONS OF RAPID COOLING. ALSO, THE PICKUP OF ELEMENTS LIKE OXYGEN BY THE MOLTEN WELD POOL LEADS TO THE ENTRAPMENT OF OXIDE INCLUSIONS IN THE SOLIDIFIED WELD. THESE INCLUSIONS THEN SERVE AS HETEROGENEOUS NUCLEATION SITES AND CAN SUBSTANTIALLY INFLUENCE THE KINETICS OF SUBSEQUENT SOLID STATE TRANSFORMATIONS. ACCORDINGLY, THE WELD METAL TRANSFORMATION BEHAVIOR IS QUITE DIFFERENT FROM THAT OF THE BASE METAL, EVEN THOUGH THE NOMINAL CHEMICAL COMPOSITION HAS NOT BEEN SIGNIFICANTLY CHANGED BY THE WELDING PROCESS (REF 3). MOST OF THE CCT DIAGRAMS (APPLICABLE TO THE WELD METAL) HAVE BEEN GENERATED BY REHEATING THE AS-DEPOSITED WELD METAL (REF 4). ONE OF THE LIMITATIONS OF THESE DIAGRAMS IS THAT THEY ARE STRICTLY APPLICABLE ONLY TO THE HIGH-TEMPERATURE REHEATED ZONE OF MULTIPASS WELDS, BECAUSE THE INITIAL MICROSTRUCTURE AT HIGH TEMPERATURES IS NOT CHARACTERISTIC OF THAT DEVELOPED FROM THE LIQUID PHASE.
- WELDING MAY BE CARRIED OUT IN SEVERAL PASSES, AND THIS MAY RESULT IN THE SUPERPOSITION OF SEVERAL DIFFERENT HEATING AND COOLING CYCLES AT ONE POINT, EACH OF THESE CYCLES HAVING THE CHARACTERISTICS NOTED ABOVE.
- SOLIDIFICATION OF THE WELD METAL IS ACCOMPANIED BY SHRINKAGE, AND THE ANISOTHERMAL CONDITIONS ALREADY EMPHASIZED CAUSE DEFORMATION. THE THERMAL CYCLES ARE THEREFORE ACTING ON METAL THAT IS SUBJECTED TO MECHANICAL STRESSES AT THE SAME TIME.

The essential differences between weld thermal cycles and then thermal cycles used for generating a conventional CCT diagram are summarized in Fig. 2. Figure 2(a) shows thermal cycles which involve a slow heating rate, soak at a temperature just above the Ac_3 temperature, and various constant cooling rates. Instead of a constant cooling rate, some investigators (Ref 5) have used cooling curves according to Newton's law of cooling:

$$T \propto E^{-T} \qquad \textbf{(EQ 1)}$$

where T is the temperature and t is the time. Others have used cooling curves corresponding to the mid-radial position of cylindrical bars of different diameters when cooled in air or when quenched in different media. The weld thermal cycles shown in Fig. 2(b) are very different, and this is why a conventional CCT diagram can give only an approximate idea of the transformation behavior in the HAZ of a weldment.

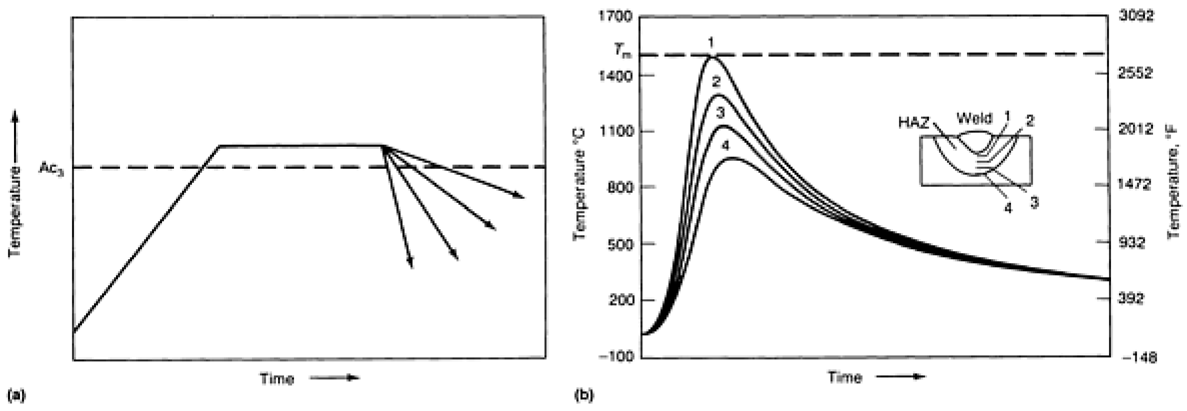

FIG. 2 GRAPHS TO SHOW DIFFERENCES IN THERMAL CYCLES. (A) THERMAL CYCLES USED TO GENERATE A CONVENTIONAL CCT DIAGRAM. (B) WELD THERMAL CYCLES. THE NUMBERS IN (B) CORRESPOND TO LOCATIONS INDICATED IN THE HAZ. NOTE THE CORRESPONDENCE BETWEEN THE THERMAL CYCLES IN (A) AND THOSE IN FIG. 1

References cited in this section

2. O.M. AKSELSEN AND T. SIMONSEN, TECHNIQUES FOR EXAMINING TRANSFORMATION BEHAVIOUR IN WELD METAL AND HAZ--A STATE OF ART REVIEW, *WELD. WORLD,* VOL 25 (NO. 1/2), 1987, P 26-34
3. O. GRONG AND D.K. MATLOCK, MICROSTRUCTURAL DEVELOPMENT IN MILD AND LOW ALLOY STEEL WELD METALS, *INT. MET. REV.,* VOL 31 (NO. 1), 1986, P 27-48
4. P.L. HARRISON AND R.A. FARRAR, APPLICATION OF CONTINUOUS COOLING TRANSFORMATION DIAGRAMS FOR WELDING OF STEELS, *INT. MATER. REV.,* VOL 34 (NO. 1), 1989, P 35-51
5. G.T. ELDIS, A CRITICAL REVIEW OF DATA SOURCES FOR ISOTHERMAL TRANSFORMATION AND CONTINUOUS COOLING TRANSFORMATION DIAGRAMS, *HARDENABILITY CONCEPTS WITH APPLICATION TO STEEL,* D.V. DOANE AND J.S. KIRKALDY, ED., TMS-AIME, 1978, P 126-157

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

HAZ of a Single-Pass Weld

Peak Temperature-Cooling Time Diagrams. The gradient in microstructure than can be obtained in a single-pass weld is shown in Fig. 3. High peak temperatures in the HAZ just adjacent to the fusion line cause coarsening of the austenite (γ) grains, and this in turn increases the hardenability of this region compared to other regions. Because each of the subzones shown in Fig. 3 occurs in a small volume, it is difficult to study the transformation behavior of individual regions by *in situ* methods (Ref 2). It is more convenient to obtain information about the microstructural and property changes in the HAZ by weld simulation (Ref 6, 7). A thermal cycle simulator (TCS) is used to reproduce the thermal cycle corresponding to a point in the HAZ in a large volume (usually a specimen has the Charpy test specimen dimensions). It is possible to program the required thermal cycle in the TCS so that the peak temperature and the cooling rate can be varied independently. It is thus possible to obtain information about the microstructural changes in the HAZ for a wide range of welding parameters.

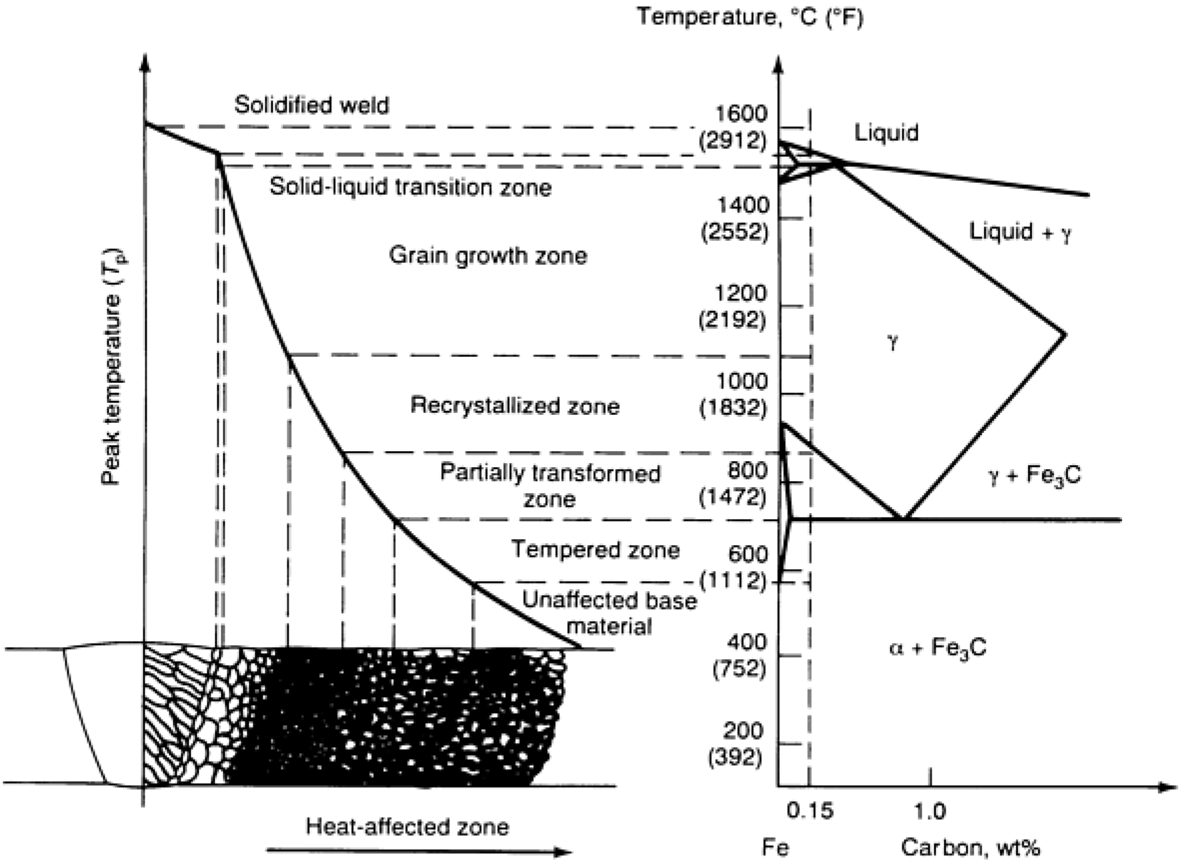
Temperature, °C (°F)
Peak temperature (T_p)
Solidified weld
Solid-liquid transition zone
Grain growth zone
Recrystallized zone
Partially transformed zone
Tempered zone
Unaffected base material
Heat-affected zone
1600 (2912)
1400 (2552)
1200 (2192)
1000 (1832)
800 (1472)
600 (1112)
400 (752)
200 (392)
Liquid
Liquid + γ
γ
γ + Fe₃C
α + Fe₃C
0.15
1.0
Fe
Carbon, wt%

FIG. 3 SCHEMATIC SHOWING VARIOUS SUBZONES THAT CAN FORM IN THE HAZ OF A CARBON STEEL CONTAINING 0.15% C. SOURCE: REF 6

Figure 4 shows how a change in the peak temperature of the thermal cycle affects the CCT characteristics of a steel. The well-known effect of a larger γ grain size (due to a higher peak temperature) in increasing the hardenability of the steel is seen. To present the information about the CCT behavior for a number of peak temperatures (see Fig. 2b and 3), it is more convenient to adopt the scheme shown in Fig. 5. In this peak temperature-cooling time (PTCT) diagram (Ref 8, 9), each point represents a weld thermal cycle with a peak temperature, T_p, given by the ordinate and the cooling time, $\Delta t_{8\text{-}5}$ (that is, required for cooling from 800 to 500 °C, or 1470 to 930 °F), given by the abscissa. A microstructural constituent or a combination of two or more constituents is shown to occur over an area in the diagram. The upward slope in the boundary between two areas is consistent with the information presented earlier in Fig. 4 that the hardenability increases with an increase in the peak temperature of the thermal cycle. Hardness and C_v transition temperatures are also shown in the diagram corresponding to different thermal cycles.

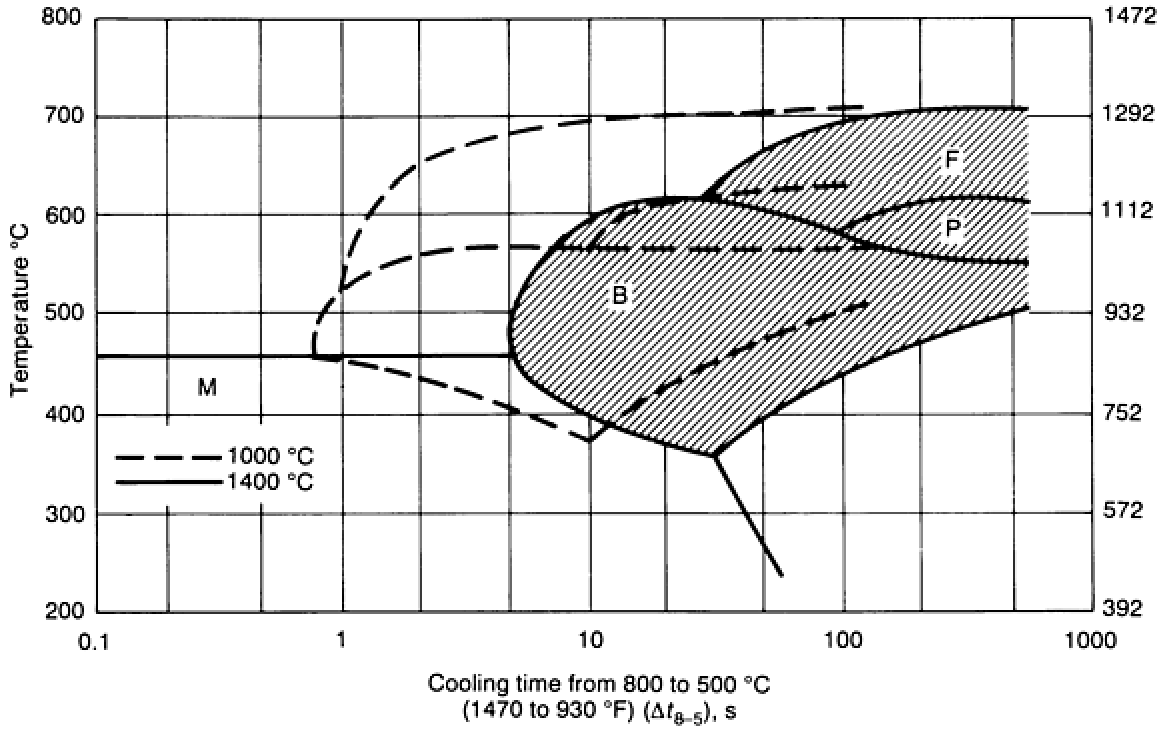

FIG. 4 EFFECT OF A CHANGE IN THE PEAK TEMPERATURE OF THE WELD THERMAL CYCLE (FROM 1000 TO 1400 °C, OR 1830 TO 2550 °F) ON THE CCT CHARACTERISTICS. M, MARTENSITE; B, BAINITE; P, PEARLITE; F, FERRITE. SOURCE: REF 8

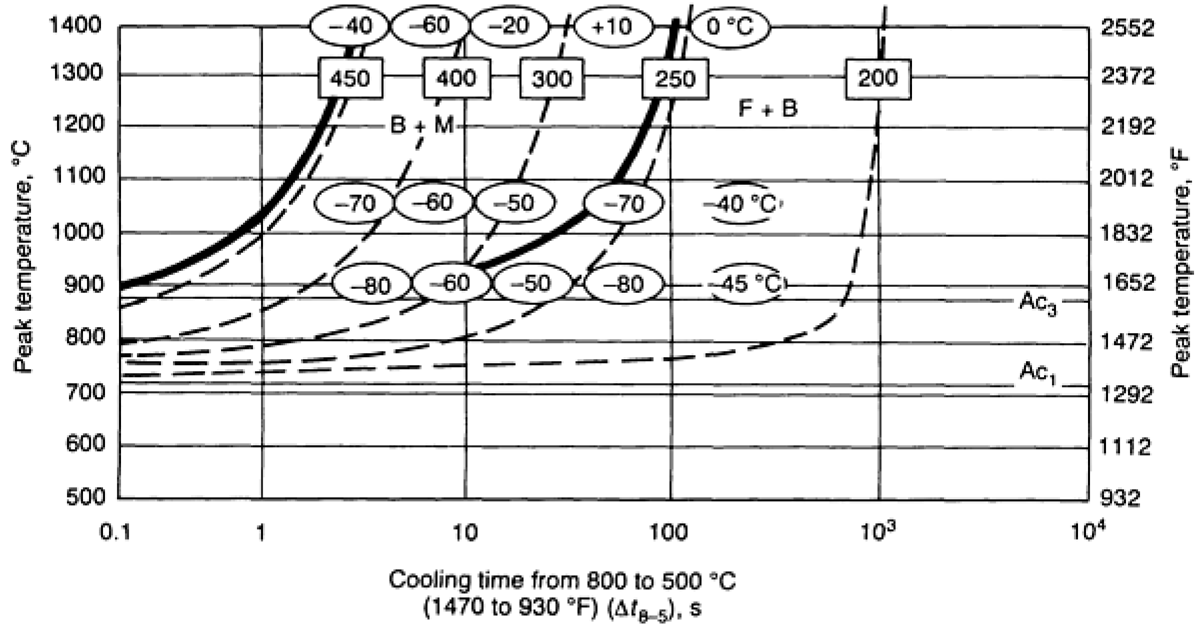

FIG. 5 TYPICAL PEAK TEMPERATURE VERSUS COOLING TIME DIAGRAM, SHOWING THE EFFECTS OF THESE PARAMETERS OF A WELD THERMAL CYCLE ON THE FINAL TRANSFORMATION PRODUCTS, ON HARDNESS, AND CHARPY V-NOTCH IMPACT ENERGY. M, MARTENSITE; B, BAINITE; F, FERRITE. FIGURES IN SQUARES INDICATE THE HARDNESS (30 HV); FIGURES IN OVALS INDICATE THE 21 J (16 FT · LBF) TRANSITION TEMPERATURES FOR CHARPY V-NOTCH IMPACT SPECIMENS SUBJECTED TO A SPECIFIC THERMAL CYCLE. SOURCE: REF 9

It may appear strange that the CCT diagram shown in Fig. 4 is plotted with Δt_{8-5} as the abscissa instead of time as in Fig. 1. Historically, the practice of thinking in terms of a cooling time began when a need was recognized (Ref 5) to compare the CCT behavior of specimens subjected to different types of cooling curves--Jominy specimens; cylindrical bars cooled in air, water, or oil; dilatometer specimens cooled at a constant or a Newtonian rate; and so on. Initially, the cooling rate at 704 °C (1300 °F) was used as a criterion to judge the equivalency of these specimens in terms of microstructure and hardness. Because transformation takes place at lower temperatures in many steels, the half cooling time (Ref 5) was found to be a better equivalency criterion. The half cooling time was defined as the time to cool from Ac_3 to a temperature that lies midway between Ac_3 and room temperature. For the sake of general applicability, this was later modified to the Δt_{8-5} criterion, and CCT diagrams began to be plotted with Δt_{8-5} as the abscissa. It must be noted that using Δt_{8-5} as an equivalency criterion is just an expedient solution that does not have strict theoretical justification (refer, for example, to the additivity principle in Ref 5).

A special significance of using Δt_{8-5} as the abscissa in Fig. 5 is that it is almost constant (for T_p > 900 °C, or 1650 °F) in the whole of the HAZ. This can be seen from Ref 6, where the expressions given for Δt_{8-5} do not contain the distance from the weld centerline as a factor. The derivation of these expressions is based on the assumption that the time to cool to 800 °C (1470 °F) is far greater than the time to reach the peak temperature, so the Δt_{8-5} can be taken to be the same for the whole of the weld and the HAZ. This can be intuitively rationalized by observing that the thermal cycles in Fig. 2(b) are such that the curves are approximately parallel below 800 °C (1470 °F). The constancy of Δt_{8-5} in the HAZ means that the gradient in microstructure (in terms of the final transformation products from austenite) is mainly due to a variation in the peak temperature. By drawing a vertical line in Fig. 5 at a value of Δt_{8-5} corresponding to a given heat input and preheating temperature, it is possible to get information about the type of microstructural gradient in the HAZ. An idea of property changes like hardness and toughness can also be obtained from Fig. 5.

A fairly satisfactory correlation between real welds and simulation studies has been observed with respect to microstructure and property measurements. However, it has frequently been observed that the maximum prior to austenite grain size in real welds is less than that in corresponding simulation specimens (Ref 2, 6, 7). It is believed that this is due to the growth of large grains in the HAZ of real welds being hindered by the adjacent smaller grains. The constraint arises from the fact that grain growth in real welds is in some sense directional, occurring from the fusion line toward lower temperatures, while under the simulation condition the austenite grains can grow in all directions.

The above discussion shows that the PTCT diagram (along with property measurements, as in Fig. 5) can alert the user to the possibility of local brittle zones in actual weldments, and that any error in the property being assessed will be on the conservative side.

While the expressions for Δt_{8-5} given Ref 6 (to determine the point at which a vertical line is to be drawn in the PTCT diagram) are reasonably accurate, those for T_p may be grossly in error because they are derived from a model with many simplistic assumptions (Ref 10, 11). For this reason, although an idea of the type of microstructural gradient in the HAZ can be obtained from the PTCT diagram, it is not possible to calculate the width of each region using simple closed form solutions of the type given in Ref 6 (derived from Rosenthal's equations). Moreover, T_p and Δt_{8-5} are alone not enough to completely characterize a weld thermal cycle. To improve the accuracy of simulation, it would be necessary to program the thermal cycle with heating rates and the shape of the thermal cycle at higher temperatures, as in the HAZ of real welds. The significance of these additional factors is discussed below.

Continuous Heating Transformation Diagrams. The heating rate in many welding processes can be very high. The implication is that considerable superheating may be required for the transformation to austenite on heating. That is, the Ac_1 and Ac_3 temperatures will be raised with an increase in the heating rate (Ref 11). This is seen in Fig. 6, a continuous heating transformation (CHT) diagram, which is analogous to a CCT diagram. It is seen that grain growth begins only after the carbides have dissolved and after a homogeneous austenite (with respect to the distribution of carbon, at least) is formed.

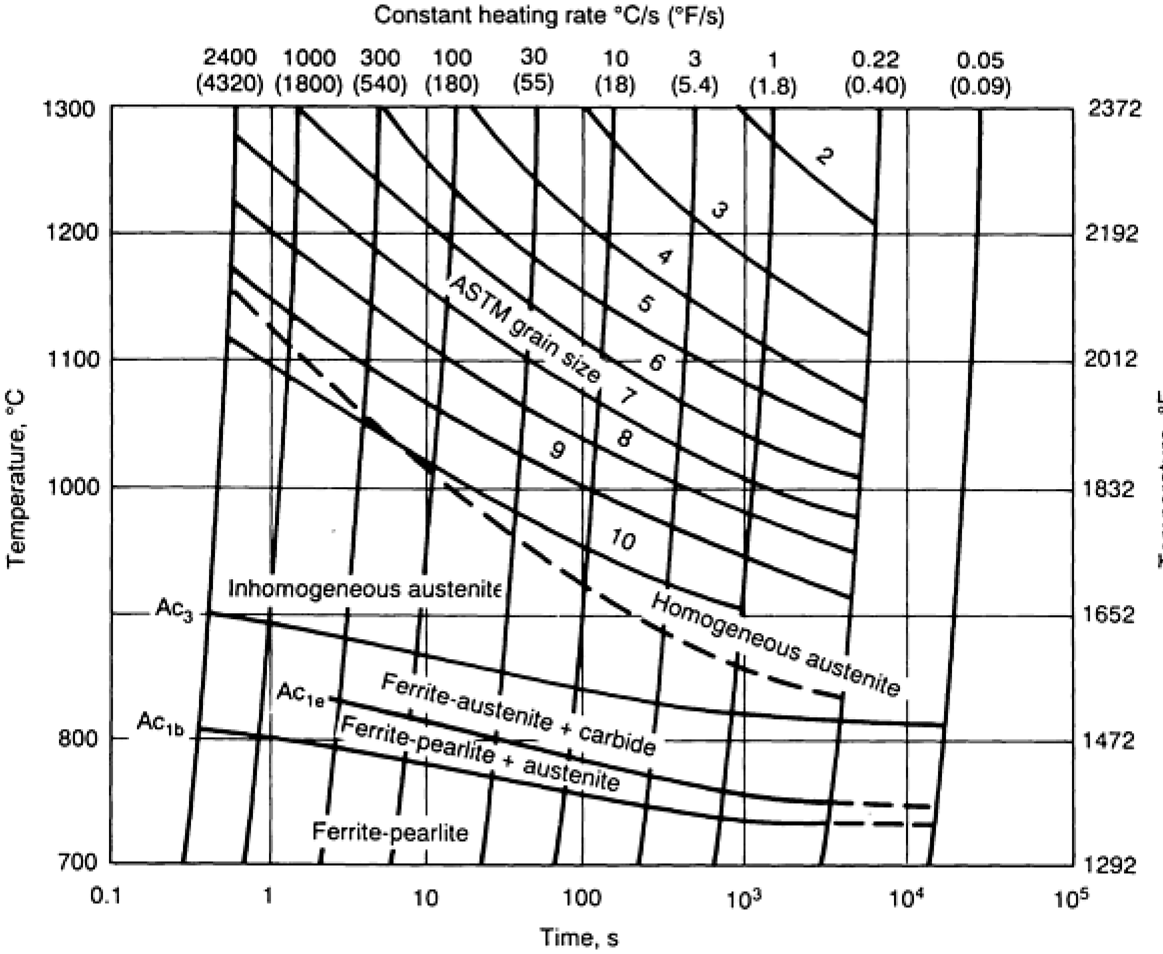

FIG. 6 CONTINUOUS HEATING TRANSFORMATION DIAGRAM FOR 34CRMO4STEEL. IN THE REGION OF HOMOGENEOUS AUSTENITE THERE ARE LINES OF CONSTANT AUSTENITE GRAIN SIZE (ASTM NUMBERS). BECAUSE OF THE MEASURING PROCEDURE, THE DIAGRAMS CAN ONLY BE INTERPRETED ALONG LINES OF CONSTANT HEATING RATE. TO SHOW THE HEATING TIME MORE CLEARLY, A TIME SCALE IS ADDED. SOURCE: REF 12

Precipitate Stability and Grain Boundary Pinning. In the last three decades, there has been a surge in the development of microalloyed or high-strength low-alloy (HSLA) steels (base metals). Almost the entire thrust behind this effort has been to tailor the thermomechanical processing of these steels to produce a fine grain size (Ref 13). Previously, higher strength was obtained by alloying the steel sufficiently to increase the hardenability and then heat treating to obtain a tempered martensitic structure. The higher carbon equivalent (Ref 6, 14) in these alloy steels led to an increased susceptibility to cold cracking or hydrogen-induced cracking. In the microalloyed steels, however, alloying additions are kept to a minimum, and higher strength is achieved primarily by a reduction in the grain size and by precipitation strengthening. A reduction in grain size is the only known method of increasing the strength and toughness at the same time. Because strengthening is obtained by other means, both carbon and carbon equivalent can be decreased to significantly reduce the susceptibility to cold cracking in these new HSLA steels (Fig. 7).

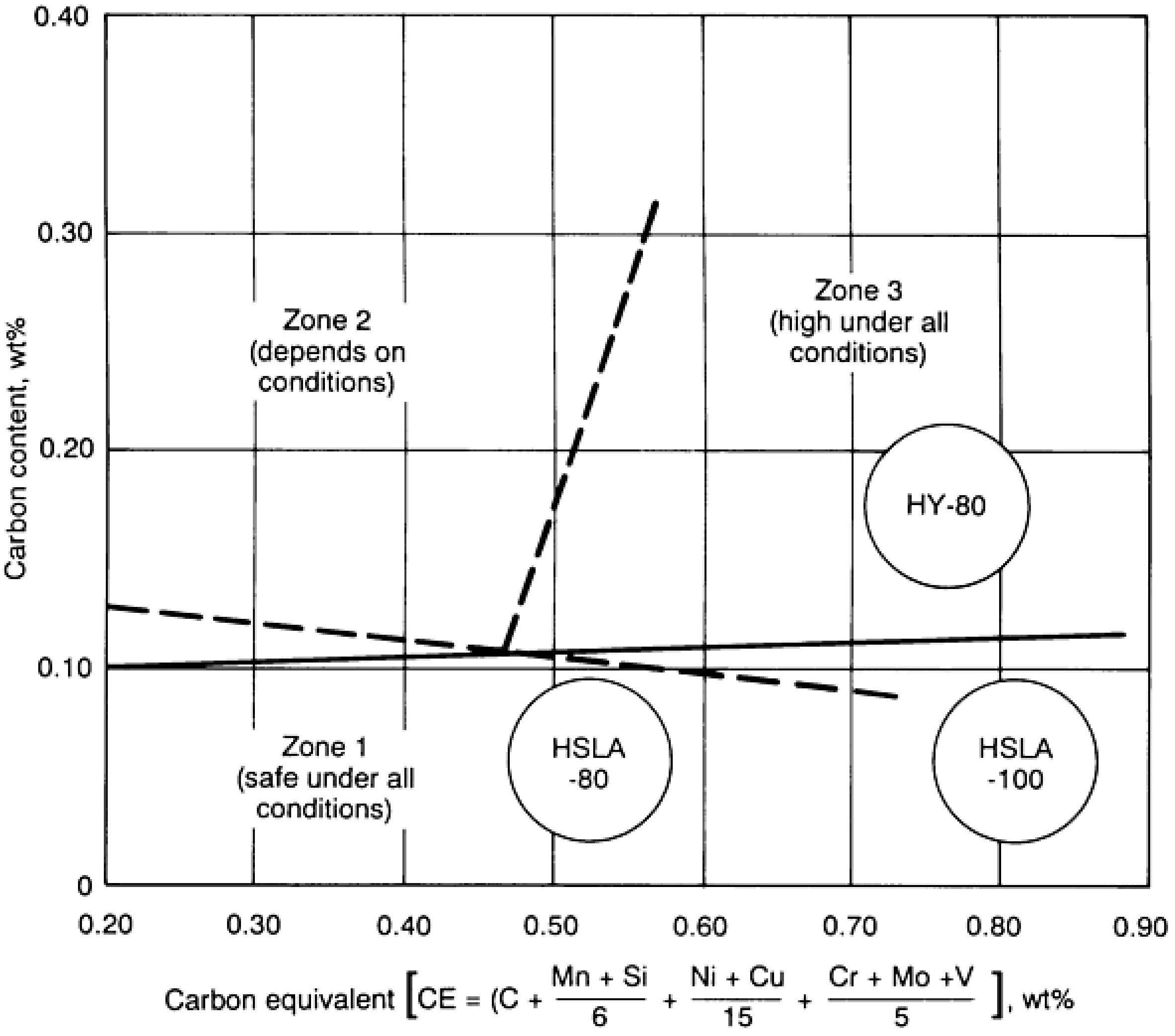

FIG. 7 EFFECT OF STEEL COMPOSITION (HY-80, HSLA-80, HSLA-100) ON THE SUSCEPTIBILITY TO COLD CRACKING IN THE HAZ. SOURCE: REF 13

A fine ferrite grain size in these steels is best achieved (after thermomechanical processing) by obtaining a fine austenite grain size before the γ-α transformation (Ref 13). To realize this goal, it is essential that microalloy additions such as niobium, vanadium, titanium, and aluminum are present so that the fine precipitate particles they form with nitrogen, carbon, or oxygen can "pin" the austenite grain boundaries and inhibit grain coarsening. The pinning effect arises from the fact that when a short length of grain boundary is replaced by a precipitate particle, the effective grain boundary energy is lowered. When the grain boundary attempts to migrate away from the particles, the local energy increases and a drag is exerted on the boundary by the particles.

In the HAZs of welds in these microalloyed steels, it is not possible to get the same optimum microstructure and microalloy precipitation obtained in the parent material by controlled thermomechanical processing. Because peak temperatures are much higher in the HAZ, the precipitate particles coarsen and dissolve, resulting in reduced pinning forces and therefore coarser austenite grains. This effect can be minimized by having precipitate particles that do not dissolve, even at higher temperatures. An idea of the stability of precipitate particles can be obtained from the plot shown in Fig. 8. It is seen that TiN has maximum stability, and this has been used to advantage in many steel. The other carbides and nitrides are not as effective as titanium nitride in limiting the extent of grain coarsening; they play a bigger role during thermomechanical processing (Ref 13).

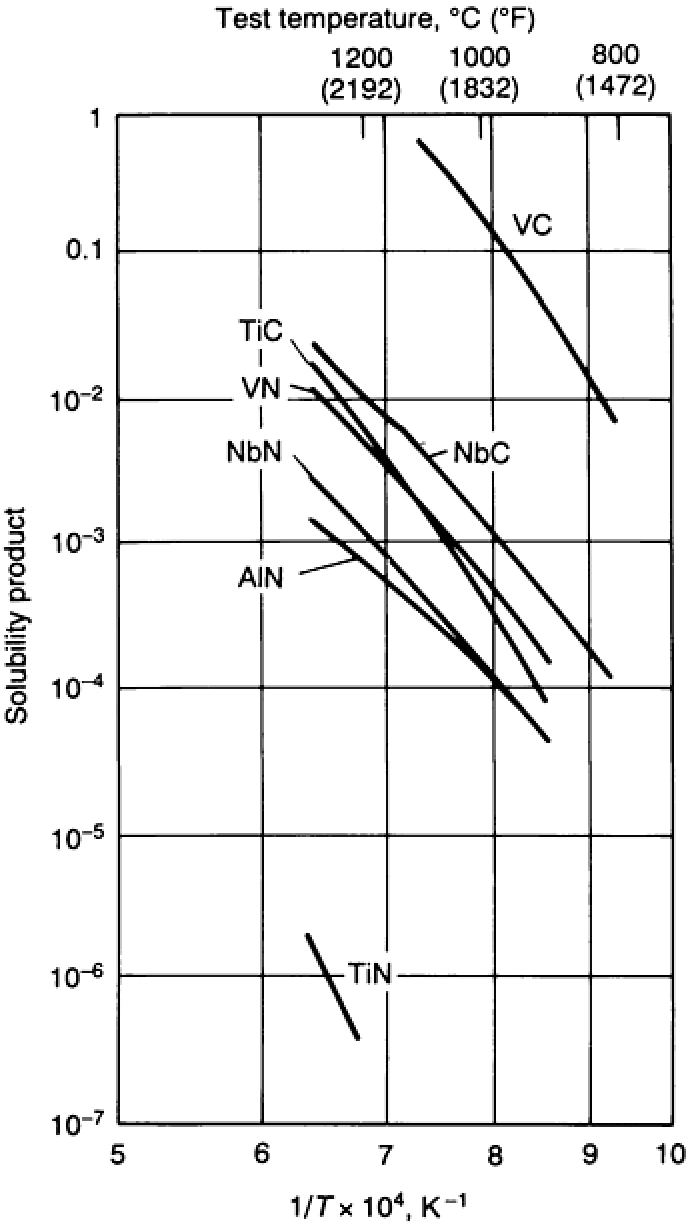

FIG. 8 SOLUBILITY PRODUCTS OF CARBIDES AND NITRIDES AS A FUNCTION OF TEMPERATURE. SOURCE: REF 15

Figure 9 shows how a high volume fraction of fine precipitate particles is needed to maximize the pinning effect. It also shows that heating to high temperatures increases the size of the particles, by a process known as Ostwald ripening (Ref 6), and decreases their volume fraction due to dissolution. This means that it is only possible to limit, and not to totally stop, grain coarsening in the HAZ, especially in high heat input welds. This inevitable grain coarsening is actually used to advantage in titanium, oxide steels (see the section "Titanium Oxide Steels" in this article).

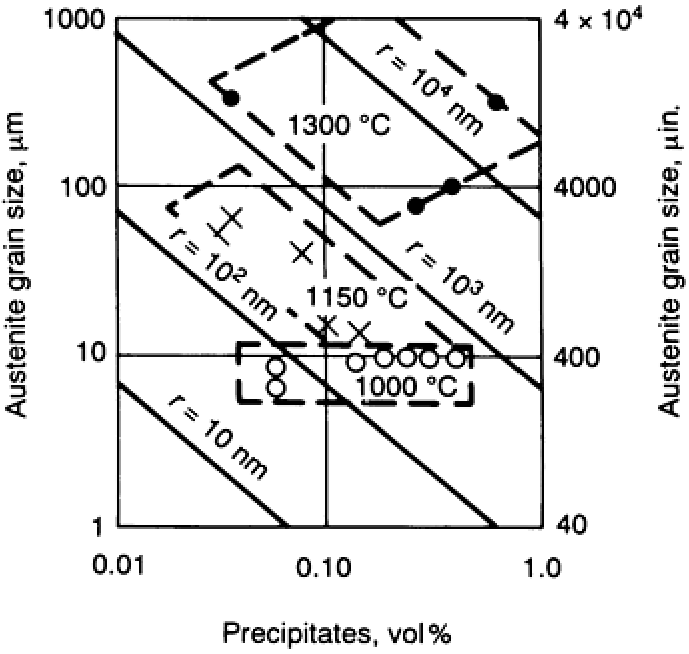

FIG. 9 EFFECT OF PARTICLE RADIUS, *R,* AND AMOUNT OF PRECIPITATE ON AUSTENITE GRAIN SIZE. SOURCE: REF 16

Unmixed and Partially Melted Zones in a Weldment. It is common to think of a single-pass weld as consisting of two zones: weld metal and HAZ. Careful metallographic examination has shown that a weld can in fact be divided into four regions (Fig. 10):

- *THE COMPOSITE ZONE,* IN WHICH A VOLUME OF BASE METAL MELTED BY THE SUPERHEATED FILLER METAL EXPERIENCES COMPLETE MIXING TO PRODUCE AN ALLOY WITH NOMINAL COMPOSITION INTERMEDIATE BETWEEN THAT OF THE BASE METAL AND THAT OF THE FILLER METAL.
- *THE UNMIXED ZONE,* WHICH FORMS FROM THE STAGNANT MOLTEN BOUNDARY LAYER REGION (ABOUT 100 TO 1000 µM, OR 0.004 TO 0.040 IN., THICK) AT THE OUTER EXTREMITIES OF THE COMPOSITE REGION. BECAUSE NO MECHANICAL MIXING WITH THE FILLER METAL OCCURS HERE, THE COMPOSITION OF THE METAL IN THIS REGION IS IDENTICAL TO THAT OF THE BASE METAL, EXCEPT FOR MINOR CHANGES PRODUCED BY DIFFUSION.
- *THE PARTIALLY MELTED ZONE,* WHICH IS A REGION AT THE FUSION BOUNDARY WHERE THE PEAK TEMPERATURES FALL BETWEEN THE LIQUIDUS AND SOLIDUS SO THAT MELTING IS INCOMPLETE.
- *THE TRUE HEAT-AFFECTED ZONE,* WHICH IS THAT PORTION OF THE BASE METAL WHERE ALL MICROSTRUCTURAL CHANGES INDUCED BY WELDING OCCUR IN THE SOLID STATE.

Special metallographic procedures are needed to reveal the existence of these special zones, as shown in Fig. 11.

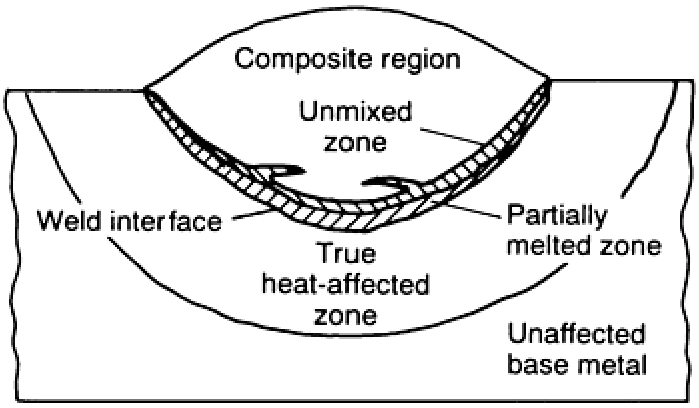
Composite region
Unmixed zone
Weld interface
Partially melted zone
True heat-affected zone
Unaffected base metal

FIG. 10 SCHEMATIC SHOWING THE DIFFERENT DISCRETE REGIONS PRESENT IN A SINGLE-PASS WELD. SOURCE: REF 17

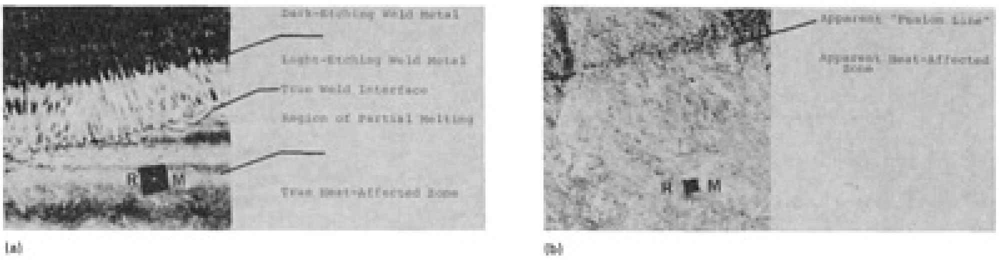
Dark-Etching Weld Metal
Light-Etching Weld Metal
True Weld Interface
Region of Partial Melting
True Heat-Affected Zone
R M
(a)
Apparent "Fusion Line"
Apparent Heat-Affected Zone
R M
(b)

FIG. 11 SPECIAL METALLOGRAPHIC PROCEDURES SHOW THE MICROSTRUCTURE OF DISCRETE REGIONS (SEE FIG. 10) IN AN HY-80 WELD. (A) MICROSTRUCTURE OBTAINED NEAR A REFERENCE MARK (INDICATED BY "RM") USING A TWO-STAGE NITAL AND SODIUM BISULFITE PROCEDURE. (B) MICROSTRUCTURE AT IDENTICAL LOCATION REPOLISHED AND ETCHED WITH NITAL ONLY. TWO-STAGE PROCEDURE SHOWS MUCH GREATER DETAIL. SOURCE: REF 17

The width of the partially melted zone can be extended by a phenomenon known as constitutional liquation (Ref 18), whereby melting can occur even when the peak temperature is less than the solidus temperature. This can be understood by referring to a phase diagram for a simple binary system A and B (Fig. 12). Consider an alloy of composition C. Its equilibrium structures at temperatures T_1, T_2, T_3 and T_4 are α+ β, α, α, and L + α, respectively (Fig. 12a). On heating from room temperature to T_3 and holding at this temperature for a long time, the βparticles will dissolve and give a homogeneous α of composition C. However, under conditions of rapid heating in the HAZ, the dissolution of β will give rise to a solute concentration gradient around each particle, as shown in Fig. 12(b). In the region surrounding each particle, the concentration of the solute B will correspond to that of liquid because, as the phase diagram shows, a liquid phase must exist between α and β at T_3. This phenomenon of localized melting due to a nonequilibrium distribution of phases during rapid heating is called constitutional liquation.

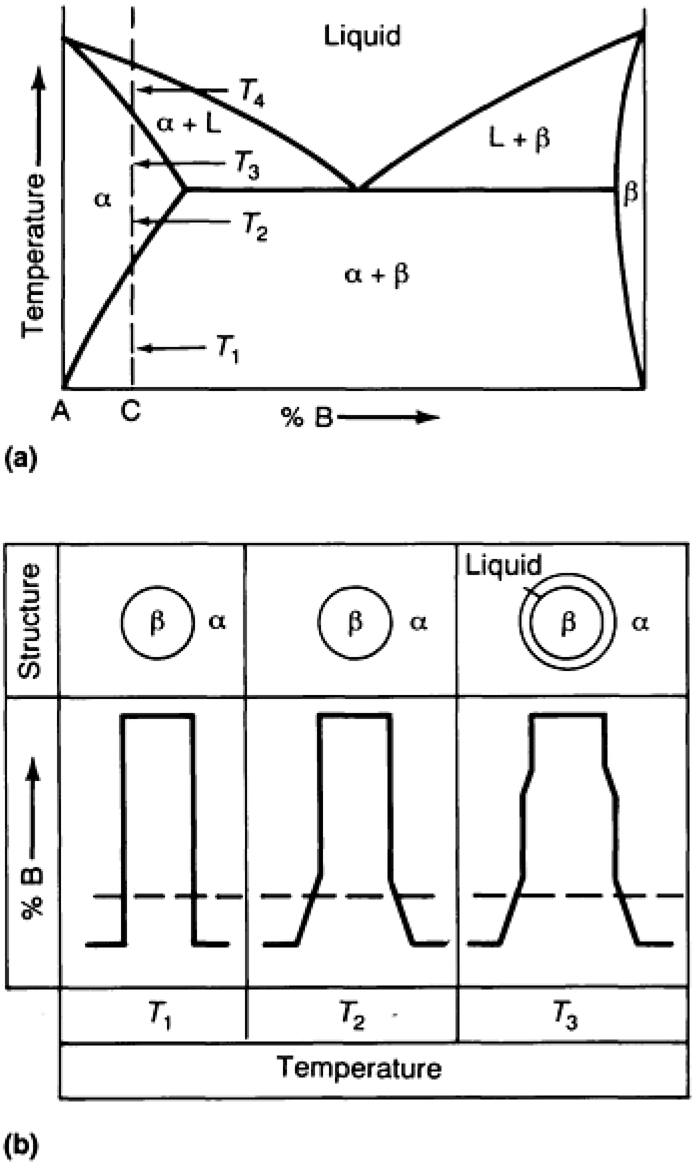

FIG. 12 EXTENSION OF PARTIALLY MELTED ZONE BY CONSTITUTIONAL LIQUATION. (A) PHASE DIAGRAM FOR A SIMPLE BINARY ALLOY. (B) STRUCTURES AND DISTRIBUTION OF CONSTITUENT B IN ALLOY C AT THREE TEMPERATURES SHOWN IN (A) WHEN RAPID HEATING IS APPLIED. SOURCE: REF 19

The partially melted zone is the region where liquation cracks have been known to occur in maraging steels, austenitic stainless steels, heat-treatable aluminum alloys, and nickel-base superalloys. It can also be a site where hydrogen-induced cracking is initiated (Ref 17), because it can act as a pipeline for the diffusion of hydrogen picked up by the molten weld metal and also because of the higher hardenability at these segregated regions.

While the phenomenon of constitutional liquation is usually discussed in connection with liquation cracking of aluminum- and nickel-base alloys (Ref 14, 18, 19), a more dramatic example can actually be found in the case of cast iron welds (Ref 20). Figure 13 shows the micro structure of the HAZ in a cast iron weld deposited using a "quench welding" technique (Ref 21). This procedure involves welding without any preheat by intermittently depositing a series of small stringer beads and strictly maintaining the interpass temperature below about 80 °C (175 °F). The idea is to limit the size of the hard and brittle white iron colonies forming by constitutional liquation around the graphite nodules and to see that they do not interconnect. By contrast, a procedure involving a preheat of about 200 °C (390 °F) can prevent the formation of martensite but not the formation of ledeburite (the structure of white iron). Because ledeburite forms at higher temperatures than martensite ($\approx$1100 °C, or 2010 °F), a much higher preheat temperature of about 600 °C (1110 °F) is required to significantly reduce the cooling rate at the higher temperatures, thereby preventing the formation of white iron. This high level of preheat is impractical. While cast iron appears to have been successfully welded with preheats of about 200 °C (390 °F), it has sometimes been found that weld repair is most effective using the quench welding technique, by which it is possible to limit both the amount of ledeburite formed and the width of the HAZ.

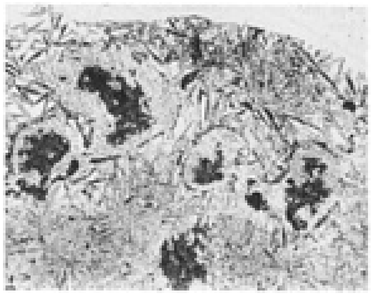

FIG. 13 DISCONTINUOUS WHITE IRON COLONIES OBTAINED IN THE HAZ OF BLACKHEART MALLEABLE IRON WELDED USING A "QUENCH WELDING" TECHNIQUE WITH A NICKEL-BASE ELECTRODE. 180×. SOURCE: REF 21

References cited in this section

2. O.M. AKSELSEN AND T. SIMONSEN, TECHNIQUES FOR EXAMINING TRANSFORMATION BEHAVIOUR IN WELD METAL AND HAZ--A STATE OF ART REVIEW, *WELD. WORLD,* VOL 25 (NO. 1/2), 1987, P 26-34
5. G.T. ELDIS, A CRITICAL REVIEW OF DATA SOURCES FOR ISOTHERMAL TRANSFORMATION AND CONTINUOUS COOLING TRANSFORMATION DIAGRAMS, *HARDENABILITY CONCEPTS WITH APPLICATION TO STEEL,* D.V. DOANE AND J.S. KIRKALDY, ED., TMS-AIME, 1978, P 126-157
6. K. EASTERLING, *INTRODUCTION TO THE PHYSICAL METALLURGY OF WELDING,* BUTTERWORTHS, 1983
7. *WELD THERMAL SIMULATORS FOR RESEARCH AND PROBLEM SOLVING,* THE WELDING INSTITUTE, U.K., 1972
8. C.F. BERKHOUT AND P.H. VAN LENT, THE USE OF PEAK TEMPERATURE-COOLING TIME DIAGRAMS IN THE WELDING OF HIGH STRENGTH STEELS (IN GERMAN), *SCHWEISS. SCHNEID.,* VOL 20 (NO. 6), 1968, P 256-260
9. TH.J. VAN ADRICHEM AND J. KAS, *CALCULATION, MEASUREMENT AND SIMULATION OF WELD THERMAL CYCLES,* SMITWELD N.V., NIJMEGEN, THE NETHERLANDS
10. P. RAVI VISHNU, W.B. LI, AND K. EASTERLING, HEAT FLOW MODEL FOR PULSED WELDING, *MATER. SCI. TECH.,* VOL 7 (NO. 7), 1991, P 649-659
11. P. RAVI VISHNU AND K. EASTERLING, PHENOMENOLOGICAL MODELLING OF HEAT FLOW AND MICROSTRUCTURAL CHANGES IN PULSED GTAW WELDS OF A QT STEEL, *MATHEMATICAL MODELING OF WELD PHENOMENA,* H. CERJAK AND K. EASTERLING, ED., THE INSTITUTE OF METALS, 1993
12. VEREIN DEUTSCHER EISENHÜTTENLEUTE, ED., *STEEL--A HANDBOOK FOR MATERIALS RESEARCH AND ENGINEERING, VOLUME 1: FUNDAMENTALS,* SPRINGER VERLAG, 1992, P 175
13. A.J. DEARDO, C.I. GARCIA, AND E.J. PALMIERE, THERMOMECHANICAL PROCESSING OF STEELS, *HEAT TREATING,* VOL 4, *ASM HANDBOOK,* 1991, P 237-255
14. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987
15. K. NARITA, PHYSICAL CHEMISTRY OF THE GROUPS IVA (TI, ZR), VA (V, NB, TA) AND THE RARE EARTH ELEMENTS IN STEEL, *TRANS. ISIJ,* VOL 15, 1975, P 145-152
16. R.K. AMIN AND F.B. PICKERING, AUSTENITE GRAIN COARSENING AND THE EFFECT OF

THERMOMECHANICAL PROCESSING ON AUSTENITE RECRYSTALLIZATION, *THERMOMECHANICAL PROCESSING OF MICROALLOYED AUSTENITE,* A.J. DEARDO, G.A. RATZ, AND P.J. WRAY, ED., TMS-AIME, 1982, P 14

17. W.F. SAVAGE, E.F. NIPPES, AND E.S SZEKERES, A STUDY OF WELD INTERFACE PHENOMENA IN A LOW ALLOY STEEL, *WELD. J.,* SEPT 1976, P 260S-268S, AND HYDROGEN INDUCED COLD CRACKING IN A LOW ALLOY STEEL, *WELD. J.*, SEPT 1976, P 276S-283S
18. J.J. PEPE AND W.F. SAVAGE, EFFECTS OF CONSTITUTIONAL LIQUATION IN 18-NI MARAGING STEEL WELDMENTS, *WELD J.,* SEPT 1967, P 411S-422S
19. W. YENISCAVITCH, JOINING, *SUPERALLOYS II*, C.T. SIMS, N.S. STOLOFF, AND W.C. HAGEL, ED., JOHN WILEY & SONS, 1987, P 495-516
20. E.E. HUCKE AND H. UDIN, WELDING METALLURGY OF NODULAR CAST IRON, *WELD. J.,* AUG 1953, P 378S-385S
21. E.N. GREGORY AND S.B. JONES, WELDING CAST IRONS, *WELDING OF CASTINGS*, THE WELDING INSTITUTE, U.K., 1977, P 145-156

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

Fusion Zone of a Single-Pass Weld

Transformations in Single-Pass Weld Metal. It is usually not necessary to select a filler metal that has exactly the same composition as the base metal; it is more important that the weld metal has the same strength and other properties (such as toughness or corrosion resistance). Because these properties are governed by the microstructure, it is important to understand the influence of different factors on phase transformations in the weld metal. But first, for a meaningful communication of the different features in a microstructure, the various phases and microconstituents must be identified using a system of nomenclature that is both widely accepted and well understood. In wrought steels, this need has been satisfied to a large degree by the Dubé scheme (Ref 22) for classifying the different morphologies of ferrite, as shown in Fig. 14. Similarly, confusion and controversy in the terminology for describing the microstructures in ferritic steel weld metals have been largely resolved by the classification scheme shown in Fig. 15. This scheme (Ref 23) was the result of several collaborative exercises undertaken under the auspices of the International Institute of Welding (IIW). Typical micrographs illustrating some of the microstructural constituents are shown in Fig. 16.

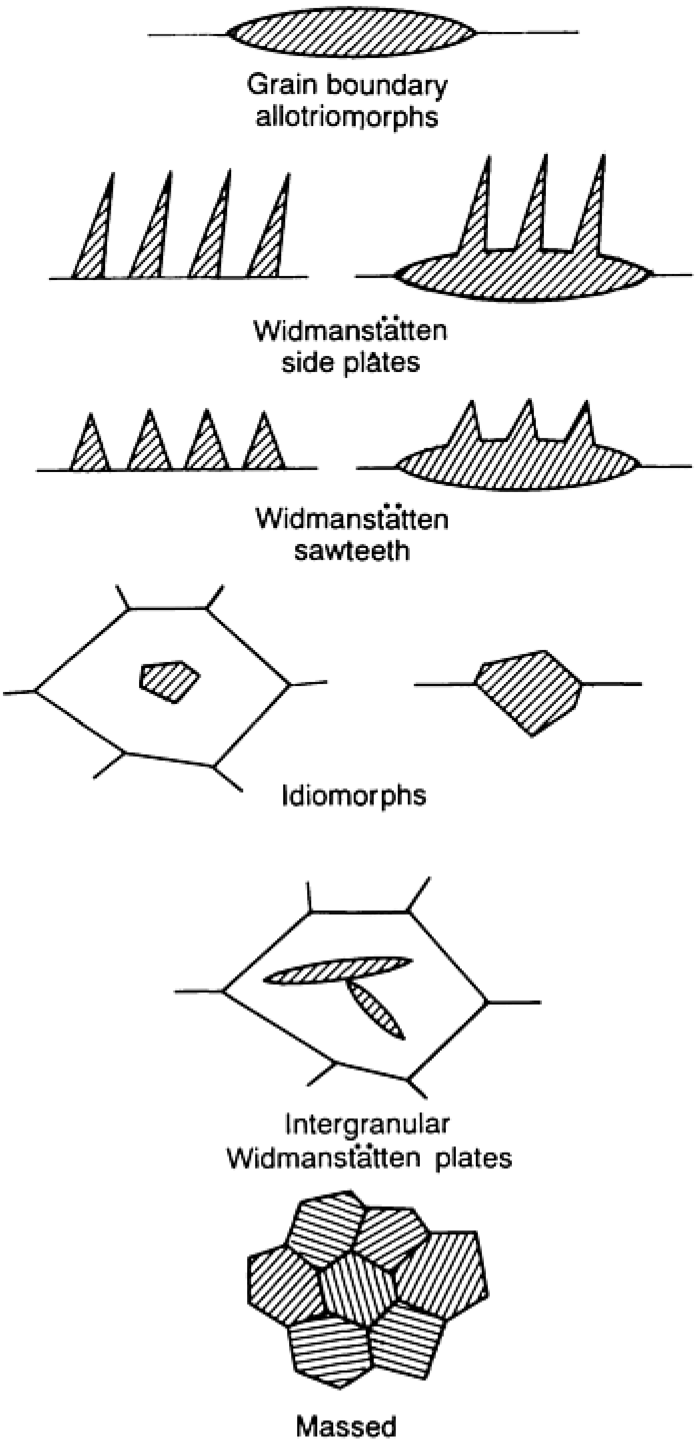
Grain boundary allotriomorphs
Widmanstätten side plates
Widmanstätten sawteeth
Idiomorphs
Intergranular Widmanstätten plates
Massed

FIG. 14 SCHEMATICS SHOWING THE DUBÉ CLASSIFICATION OF FERRITE MORPHOLOGIES. SOURCE: REF 22

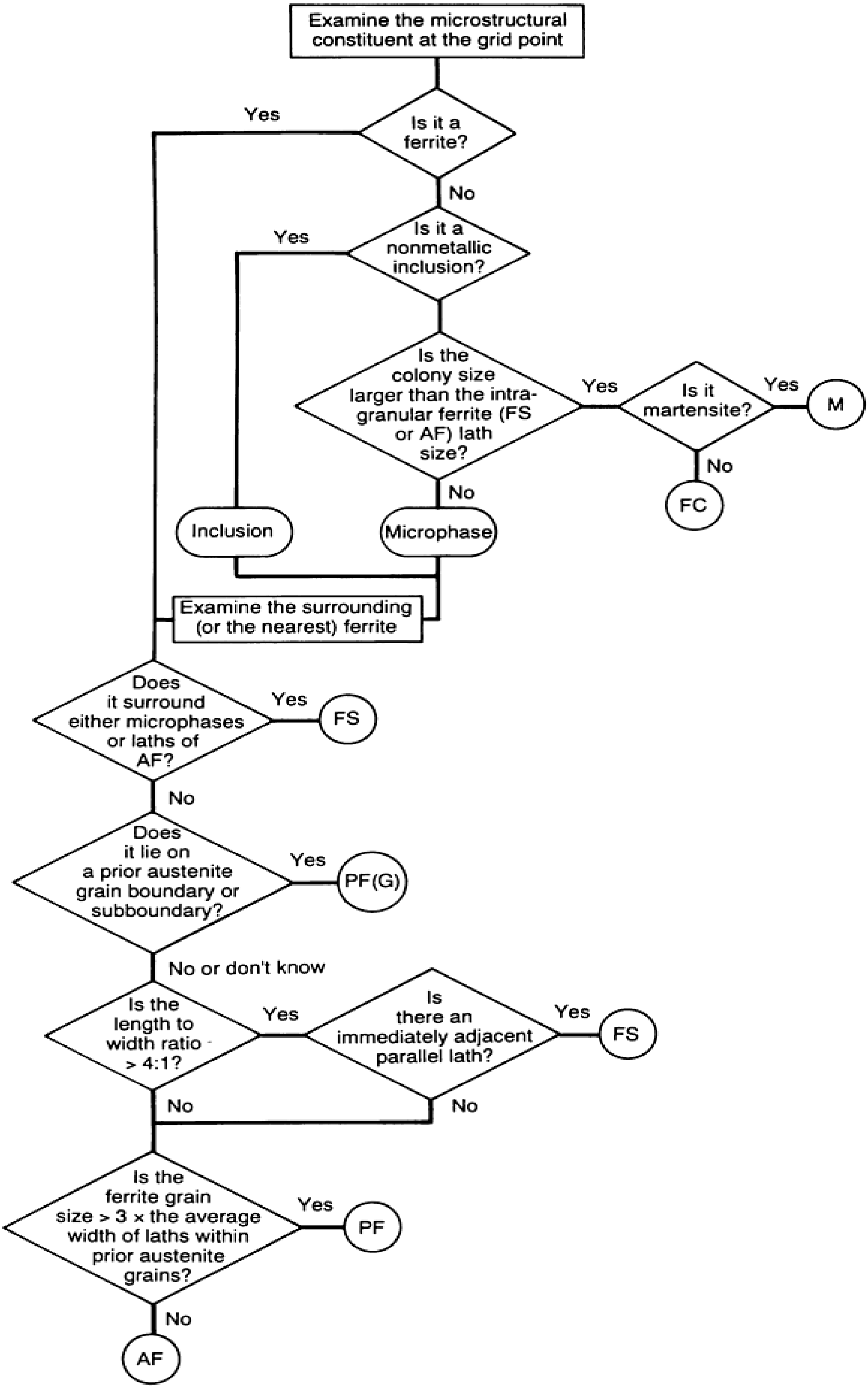
Examine the microstructural constituent at the grid point
Is it a ferrite?
Yes
No
Is it a nonmetallic inclusion?
Yes
Is the colony size larger than the intra-granular ferrite (FS or AF) lath size?
Yes
Is it martensite?
Yes
M
No
FC
No
Inclusion
Microphase
Examine the surrounding (or the nearest) ferrite
Does it surround either microphases or laths of AF?
Yes
FS
No
Does it lie on a prior austenite grain boundary or subboundary?
Yes
PF(G)
No or don't know
Is the length to width ratio > 4:1?
Yes
Is there an immediately adjacent parallel lath?
Yes
FS
No
No
Is the ferrite grain size > 3 × the average width of laths within prior austenite grains?
Yes
PF
No
AF

CATEGORY	ABBREVIATION
PRIMARY FERRITE	PF
GRAIN BOUNDARY FERRITE	PF(G)
INTRAGRANULAR POLYGONAL FERRITE	PF(I)
FERRITE WITH SECOND PHASE	FS
FERRITE WITH NONALIGNED SECOND PHASE	FS(NA)
FERRITE WITH ALIGNED SECOND PHASE	FS(A)
FERRITE SIDE PLATES	FS(SP)
BAINITE	FS(B)
UPPER BAINITE	FS(UB)
LOWER BAINITE	FS(LB)
ACICULAR FERRITE	AF
FERRITE-CARBIDE AGGREGATE	FC
PEARLITE	FC(P)
MARTENSITE	M
LATH MARTENSITE	M(L)
TWIN MARTENSITE	M(T)

FIG. 15 INTERNATIONAL INSTITUTE OF WELDING SCHEME FOR CLASSIFYING MICROSTRUCTURAL CONSTITUENTS IN FERRITIC STEEL WELD METALS WITH THE OPTICAL MICROSCOPE. SOURCE: REF 23

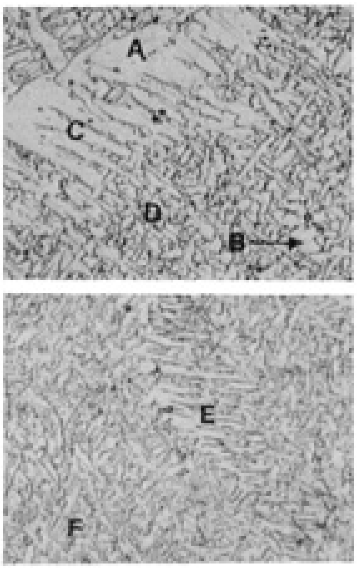

FIG. 16 MICROGRAPHS SHOWING TYPICAL MICROSTRUCTURES IN LOW-CARBON STEEL WELD METAL (NITAL ETCH). A, GRAIN BOUNDARY FERRITE [PF(G)]; B, POLYGONAL FERRITE [PF(I)]; C, WIDMANSTÄTTEN FERRITE [FS(SP)]; D, ACICULAR FERRITE (AF); E, UPPER BAINITE [FS(UB)]; F, LOWER BAINITE [FS(LB)] AND/OR MARTENSITE (M). SOURCE: REF 3

The transformation behavior in ferritic steel weld metals is best understood by first noting that ferrite is nucleated heterogeneously and that for all practical purposes, it is necessary to consider the competitive nucleation behavior only at grain boundaries and at inclusions. Figure 17 shows that inclusions have to be larger than a certain size (0.2 to 0.5 μm, or 8 to 20 μin., the typical size range of most weld metal inclusions) for their potency as nucleation sites to reach a maximum. It also shows that nucleation of ferrite is always energetically more favorable at grain boundaries than at inclusions. Based on additional considerations, like thermal contraction strains and lattice matching at inclusion-austenite-ferrite interfaces (not taken into account to get the results in Fig. 17), a discussion of why certain inclusions are more potent nucleation sites than others is found in Ref 3, 24, and 25.

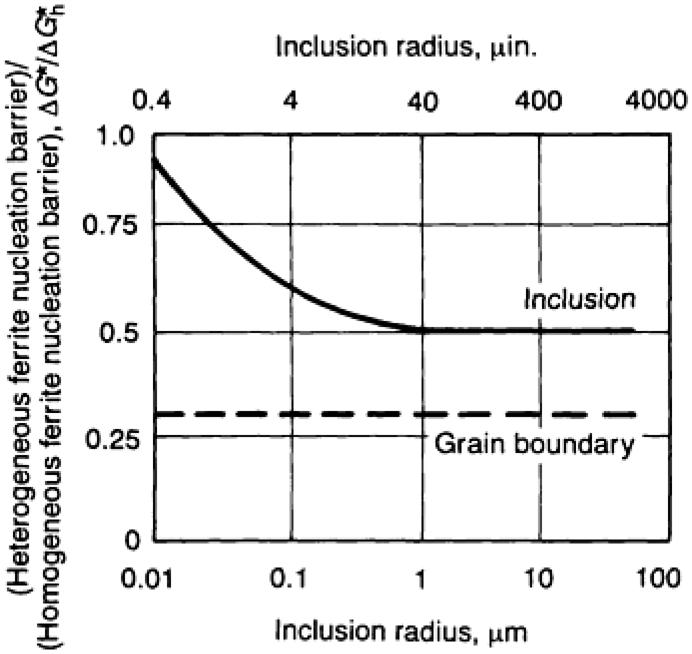

FIG. 17 EFFECT OF PARTICLE RADIUS ON ENERGY BARRIER TO FERRITE NUCLEATION AT INCLUSIONS, Δ*G (HETEROGENEOUS), NORMALIZED RELATIVE TO THE HOMOGENEOUS NUCLEATION BARRIER, Δ*G**$_H$ (HOMOGENEOUS). CORRESPONDING ENERGY BARRIER TO NUCLEATION OF FERRITE AT AUSTENITE GRAIN BOUNDARIES IS INDICATED BY HORIZONTAL BROKEN LINE. SOURCE: REF 6**

Broadly, the major factors affecting transformation behavior in ferritic steel weld metals are alloy composition, weld heat input (by its effect on γ grain size and Δt_{8-5}), oxygen content (that is, the inclusion content), and the nature of segregation in the weld metal. A typical weld CCT diagram is shown in Fig. 18. For the cooling curve shown in the figure, the first phase that forms is allotriomorphic ferrite, or grain boundary ferrite in the IIW scheme. The term "allotriomorphic" means a "particle of a phase that does not have a regular external shape" and in the present context means that ferrite grows on the grain boundary surfaces and does not have a regular faceted shape reflecting the symmetry of its internal crystalline structure. At lower temperatures, the mobility of the curved or random γ/α-allotriomorph boundaries decreases, and Widmanstätten side plates (ferrite side plates in the IIW scheme) form. Growth of these side plates is rapid because carbon is efficiently redistributed to the sides of the growing tips, thus avoiding solute pile-up problems. In addition, substitutional atoms do not diffuse during the growth of Widmanstätten ferrite.

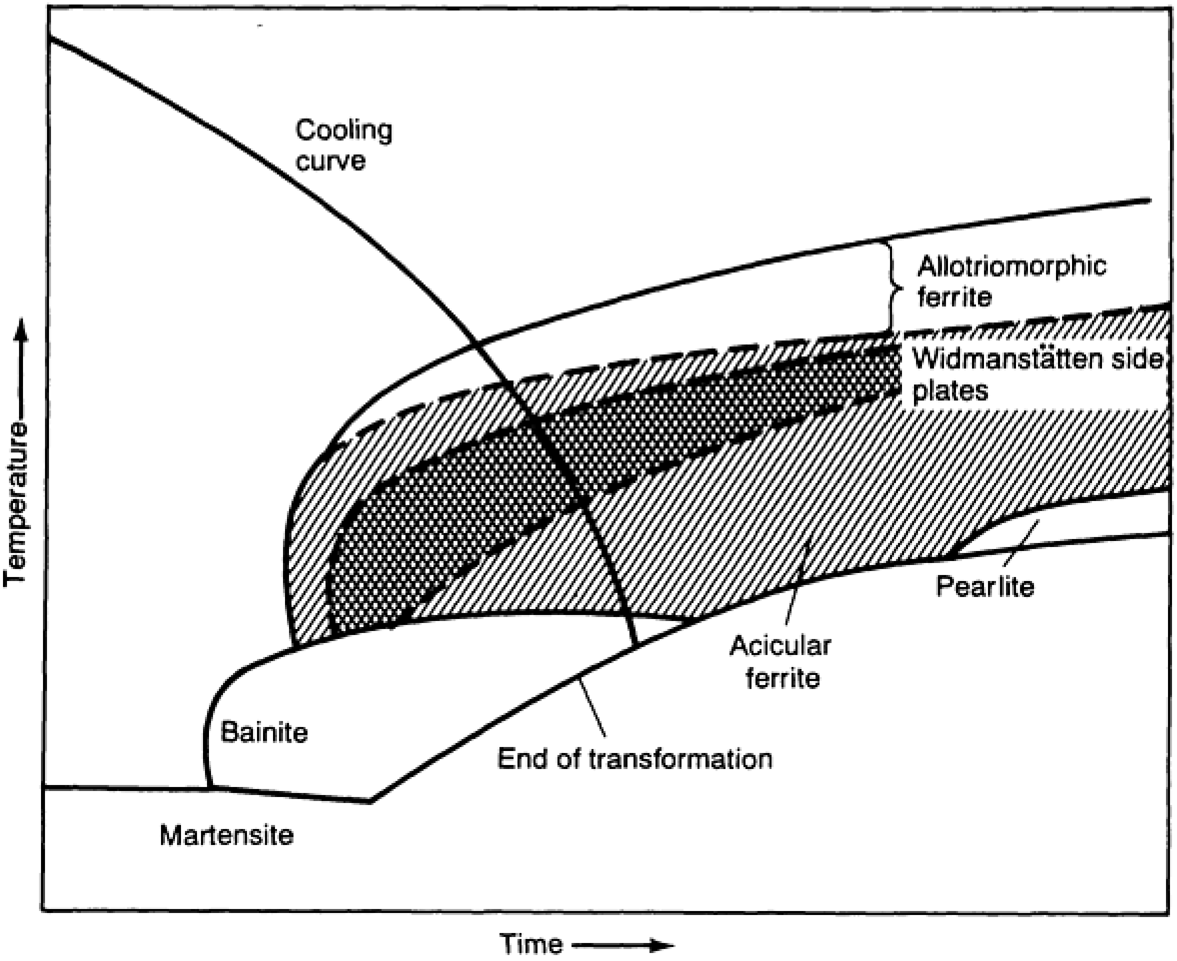

FIG. 18 SCHEMATIC OF WELD CCT DIAGRAM SHOWING SELECTED MICROSTRUCTURES

After all the grain boundary sites are saturated with allotriomorphic or Widmanstätten ferrite and their growth rate is not sufficient to extend to the interior of the grains, the nucleation of ferrite at inclusions within the γ grains becomes competitive. Acicular ferrite forms, a structure resulting from ferrite laths growing in different directions from inclusions and from laths already nucleated. Upon impingement, high-angle grain boundaries and a very fine dispersion of microphases are obtained between ferrite laths. "Microphases" in this context means the transformation structures resulting from the carbon-enriched regions between the ferrite laths and could be the martensite-austenite constituent, bainite, or pearlite (see the section "HAZ in Multipass Weldments" in this article). An example of the microstructure of acicular ferrite is shown in Fig. 19. Acicular ferrite does not figure in the Dubé scheme because it is rarely observed in wrought steels. When the cooling rate is higher or when the inclusion content is very low, bainite can be nucleated directly at the γ grain boundaries. Bainite can form as upper or lower bainite; the difference between the two is illustrated in Fig. 20. Examination with transmission electron microscopy (TEM) is usually required for a firm identification of the type of bainite formed. In the case of the highest cooling rates, martensite is obtained.

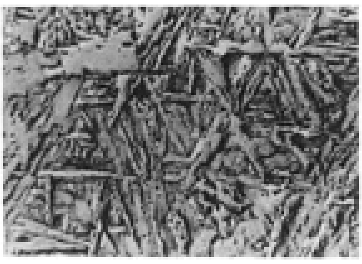

FIG. 19 SCANNING ELECTRON MICROGRAPH SHOWING MORPHOLOGY OF ACICULAR FERRITE. SOURCE: REF 24

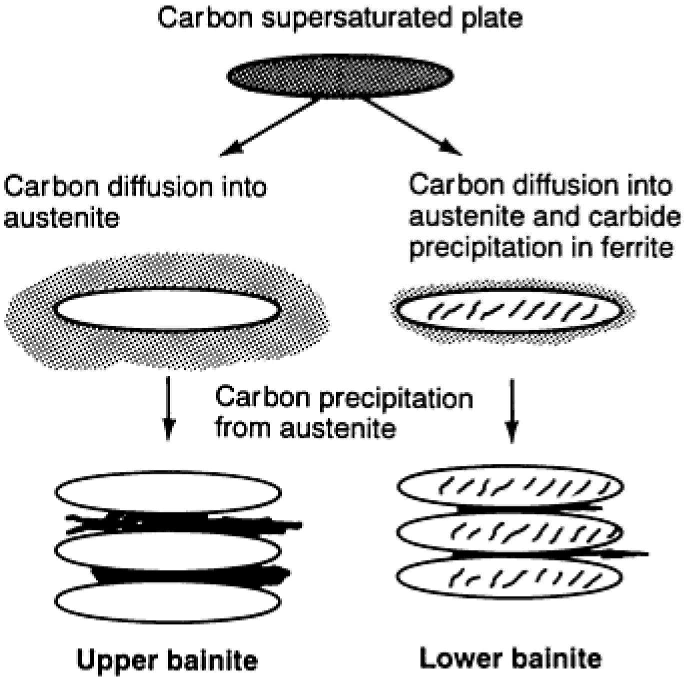
Carbon supersaturated plate
Carbon diffusion into austenite
Carbon diffusion into austenite and carbide precipitation in ferrite
Carbon precipitation from austenite
Upper bainite
Lower bainite

FIG. 20 SCHEMATIC SHOWING THE DIFFERENCES IN THE TRANSFORMATION MECHANISM FOR UPPER AND LOWER BAINITE AND THE EFFECT OF THESE MECHANISMS ON THE FINAL MORPHOLOGY. SOURCE: REF 24

Figure 21 shows the effect of alloying additions, Δt_{8-5}, oxygen content, and the γ grain size on the sequence and relative amounts of allotriomorphic ferrite, Widmanstätten ferrite, and bainite in the weld metal. Broadly, the trends shown can be understood in terms of the wellknown observation that an increase in γ grain size, a decrease in inclusion content, and an increase in alloying additions will shift the CCT curves to lower temperatures and to longer times. While Fig. 21 can generally be interpreted by assuming that the individual factors act independently of each other, the transition from (b) to (a) with an increase in the inclusion content must be understood in terms of its effect of the γ grain size. It has been observed that the γ grain size decreases with an increase in the inclusion content, and this has been explained by the higher magnitude of pinning forces at higher inclusion contents (Ref 25). The smaller grain size implies a greater grain boundary surface area, and this increases the amount of grain boundary ferrite and sideplate structures formed at the expense of acicular ferrite in Fig. 21(a). The role of the initial formation or non-formation of allotriomorphic ferrite at γ grain boundaries in bringing about the transition from Fig. 21(b) to Fig. 21(c) in Fig. 21 is discussed in the next section of this article.

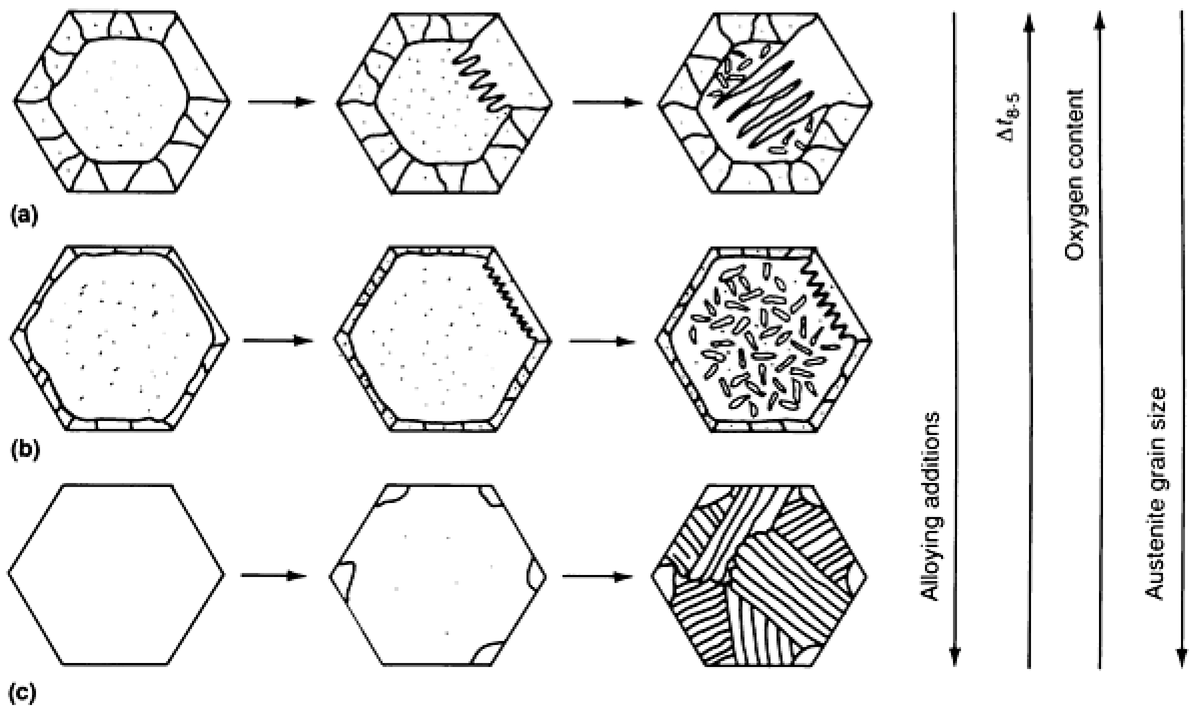

FIG. 21 SCHEMATIC SHOWING THE EFFECT OF ALLOY COMPOSITION, $\Delta T_{8\text{-}5}$, OXYGEN CONTENT, AND γ GRAIN SIZE ON THE DEVELOPMENT OF MICROSTRUCTURE IN FERRITIC STEEL WELD METALS. THE HEXAGONS REPRESENT CROSS SECTIONS OF COLUMNAR γ GRAINS. (A) THE GRAIN BOUNDARIES BECOME DECORATED FIRST WITH A UNIFORM, POLYCRYSTALLINE LAYER OF ALLOTRIOMORPHIC FERRITE, FOLLOWED BY FORMATION OF WIDMANSTÄTTEN FERRITE, AND THEN BY FORMATION OF ACICULAR FERRITE. (B) THE GROWTH RATE OF WIDMANSTÄTTEN FERRITE IS NOT SUFFICIENTLY HIGH TO EXTEND ENTIRELY ACROSS γ GRAINS. NUCLEATION OF FERRITE AT INCLUSIONS WITHIN THE γ GRAINS LEADS TO AN INCREASE IN THE AMOUNT OF ACICULAR FERRITE WHEN COMPARED WITH CASE (A). (C) THE HIGHER ALLOY CONTENT OR THE HIGHER COOLING RATE SUPPRESSES THE FORMATION OF ALLOTRIOMORPHIC FERRITE. THIS LEAVES THE γ GRAIN BOUNDARIES FREE TO NUCLEATE UPPER BAINITE. SOURCE: REF 24

Although the IIW classification is the result of a consensus viewpoint, some researchers (Ref 24) still feel that it places an exaggerated emphasis on morphological features observed with the light microscope. They believe that it is more important that the terminology reflect the mechanism of transformation. For example, it can be difficult to differentiate between Widmanstätten ferrite and upper bainite when using the light microscope, and both are identified as FS in the IIW classification. For a more fundamental understanding and to clearly interpret the trends in microstructural development, it is essential to make a distinction between the two. In addition, the subclassification of grain boundary ferrite in the IIW classification is not unequivocal in its meaning--it could refer to ferrite that forms by diffusion, Widmanstätten ferrite, or the ferrite in upper bainite. By contrast, the term "allotriomorphic ferrite" clearly identifies the mechanism by which it is formed. On the other hand, the IIW classification is easy to use and allows a more detailed identification, either by judgment or by further TEM examination, for example, to determine whether FS is FS(SP) or FS(UB). At least, it decreases the tendency for a wrong identification of the transformation mechanism with insufficient data. For a microstructure-toughness correlation, it is sufficient to characterize the morphological features.

Relating Weld Metal Toughness to the Microstructure. A good insight into the subject of toughness in ferrite steel weld metal is obtained by examining the data shown in Fig. 22. It is seen from Fig. 22(a) that the upper-shelf Charpy V-notch (CVN) energy monotonically decreases with an increase in oxygen content. In the upper-shelf temperature region, ductile fracture occurs by a process of microvoid coalescence, and because microvoids are initiated at the inclusion-matrix interface, the upper-shelf impact energy decreases with an increase in the inclusion content. By contrast, from Fig. 22(b) it is seen that an optimum inclusion content is required to obtain the lowest CVN transition temperature. The transition temperature is mainly governed by cleavage fracture, and this in turn depends on how effectively a propagating cleavage crack is forced to change direction as it traverses the microstructure. At low inclusion contents, an upper bainitic structure is obtained, and this consists of parallel platelets of ferrite (in a single packet) growing from the grain boundary surfaces. With an optimum inclusion content, a predominantly acicular ferrite structure is obtained, and here the adjacent ferrite platelets tend to radiate in many different directions from inclusion nucleation site. At higher inclusion contents, the amount of ferrite sideplate structures increases, again having nearly parallel ferrite platelets. The highest toughness (that is, the lowest transition temperature) is obtained only in the "chaotic" microstructure of acicular ferrite because it has the smallest effective ferrite grain size.

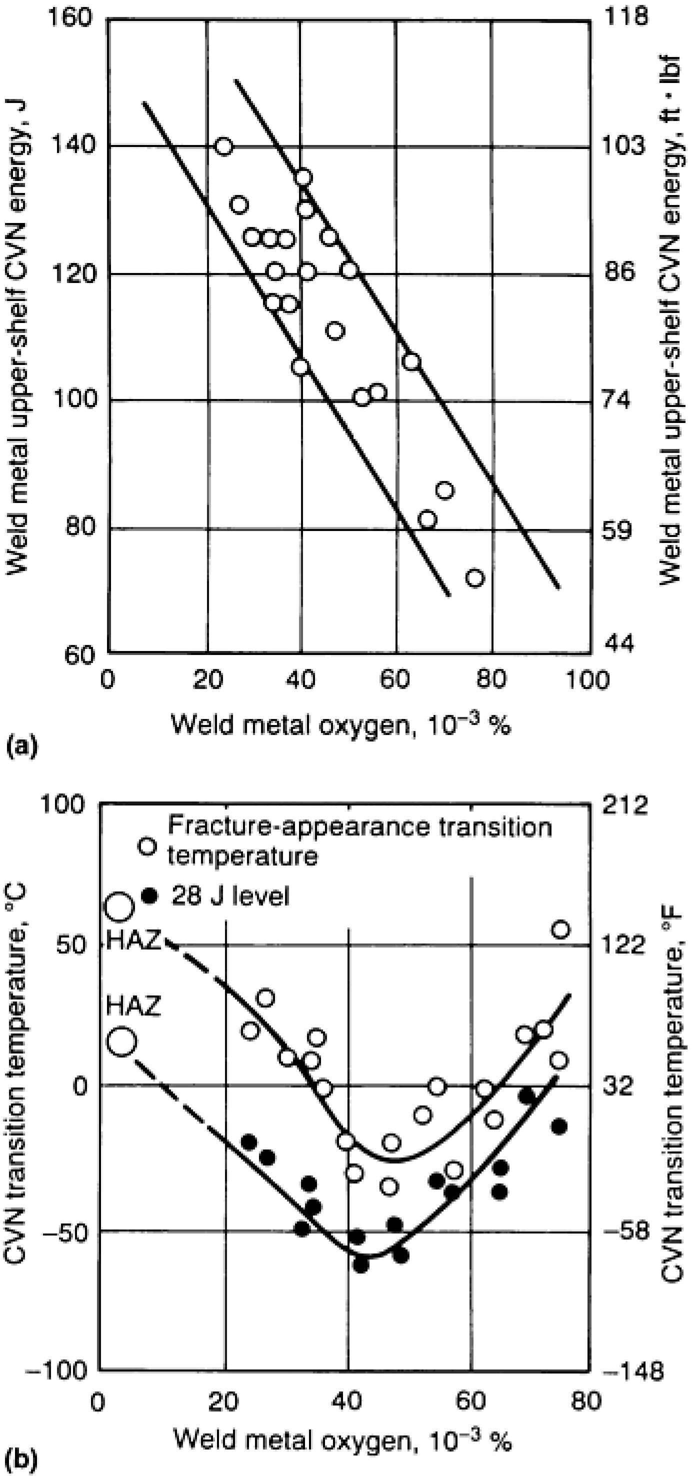

FIG. 22 TYPICAL VARIATION OF SELECTED WELD METAL PROPERTIES WITH OXYGEN CONTENT IN FERRITIC STEEL WELD METALS. (A) PLOT OF CHARPY V-NOTCH (CVN) IMPACT TEST UPPER-SHELF ENERGY VERSUS OXYGEN CONTENT. (B) PLOT OF CVN TRANSITION TEMPERATURE VERSUS OXYGEN CONTENT. SOURCE: REF 26

In recognition of the latter fact, most of the work in consumables development in the recent past has concentrated on increasing the amount of acicular ferrite in the microstructure. In line with the trends shown in Fig. 21(a) and 21(b), it has been found that decreasing the amount of grain boundary ferrite and ferrite sideplate structures increases the acicular ferrite content. This is shown more clearly in Fig. 23.

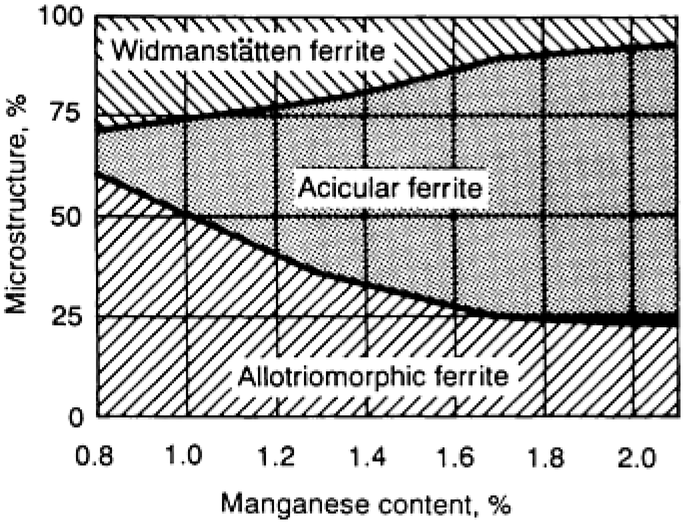

FIG. 23 EFFECT OF MANGANESE CONTENT OF WELD METAL ON THE RELATIVE AMOUNTS OF THE MICROSTRUCTURAL CONSTITUENTS PRESENT. CARBON CONTENT MAINTAINED AT 0.03%. SOURCE: REF 27

Recent work (Ref 24, 25, 27), however, has shown that a 100% acicular ferritic structure is neither desirable nor, in most cases, achievable. Figure 24 shows that the impact toughness peaks at about 70% acicular ferrite and then falls. This was attributed to the formation of segregated bands of brittle microphases in the alloy compositions giving the highest acicular ferrite contents. Moreover, when the alloying additions are increased still further, the transformation behavior becomes such that allotriomorphic ferrite formation is totally suppressed, leaving the γ grain boundaries free to nucleate upper bainite at lower temperatures (see Fig. 21c). It appears that a small layer of allotriomorphic ferrite is essential to obtain high acicular ferrite contents. When it forms, it saturates the γ grain boundary sites, and bainite nucleation at the γ/α-allotriomorphic interface is inhibited by the high carbon content there as a result of its rejection from the ferrite (Ref 24). Conditions are then favorable for intragranular nucleation at inclusion sites and thus the formation of acicular ferrite.

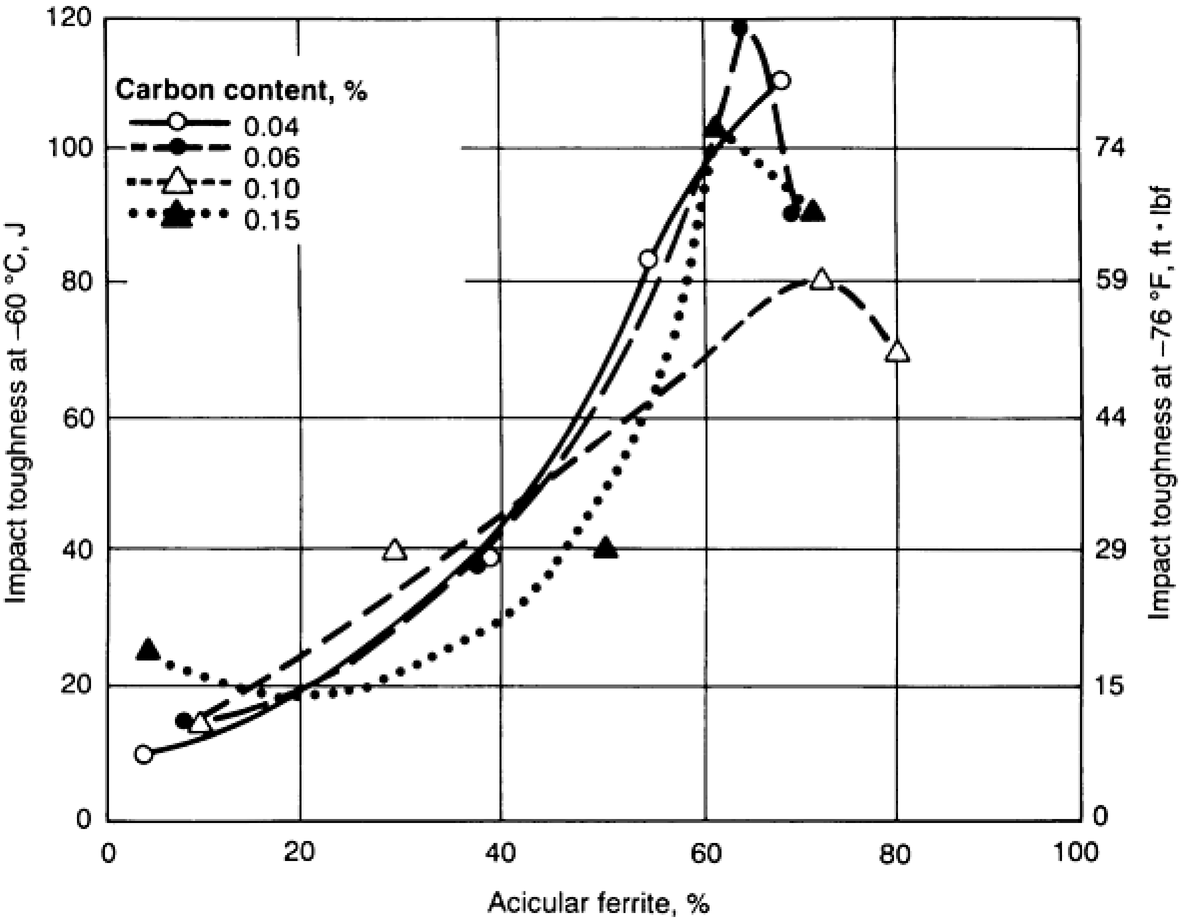

FIG. 24 PLOT OF IMPACT ENERGY VERSUS ACICULAR FERRITE CONTENT FOR SELECTED CARBON CONTENTS AT -60 °C (-78 °F). SOURCE: REF 27

For the welding of steels with yield strengths less than 600 MPa (85 ksi), it has been possible to develop weld metals with matching strength and toughness by having high acicular ferrite contents in them. A yield strength of about 600 MPa (85 ksi) seems to be the ceiling for weld metals based on acicular ferrite. Attempting to obtain additional strength by additional alloying to increase the amount of solid solution hardening and precipitation hardening in fact only results in increasing amounts of martensite and bainite. These structures are associated with a lower toughness. This has been observed in the case of HSLA-100 steels, for example, where attempts to get a matching strength in the weld metal have resulted in a drop in toughness (Ref 25, 28). Work on consumables development for high-strength steels is still in progress to match the base metal in strength and toughness (Ref 25, 28). However, current thinking appears to favor resorting to a slightly undermatching weld metal (in terms of strength) so that the already available lower-strength, high-toughness, consumables having high acicular ferrite contents (Ref 29) can be used. These weld metal compositions have the additional benefit of requiring less stringent measures for avoiding hydrogen-induced cracking.

Titanium Oxide Steels. In the discussion of the role of pinning by precipitate particles (for example, TiN) in limiting grain growth in the HAZ (in the section "Precipitate Stability and Grain Boundary Pinning" in this article), it was pointed out that at high heat inputs the particles dissolve and are not able to prevent coarse γ grains from forming. The low toughness in these coarse-grained regions is of some concern. The fact that a coarse γ grain size will lead to a higher acicular ferrite content (see Fig. 21b) with improved toughness has been used to advantage in the development of titanium-oxide-containing steels (Ref 30, 31). Ti_2O_3 is more stable than titanium nitride and does not dissolve, even at the highest heat inputs. The undissolved, Ti_2O_3 particles do not stop grain growth, but their survival after the heating cycle means that they can be effective in nucleating acicular ferrite within the coarse γ grains. The precise reasons why they are more effective than other undissolved inclusions, such as Al_2O_3 in nucleating acicular ferrite are not well understood (Ref 24). The improved HAZ microstructure and toughness obtained with titanium oxide steels over titanium nitride steels at high heat inputs are shown in Fig. 25 and 26, respectively.

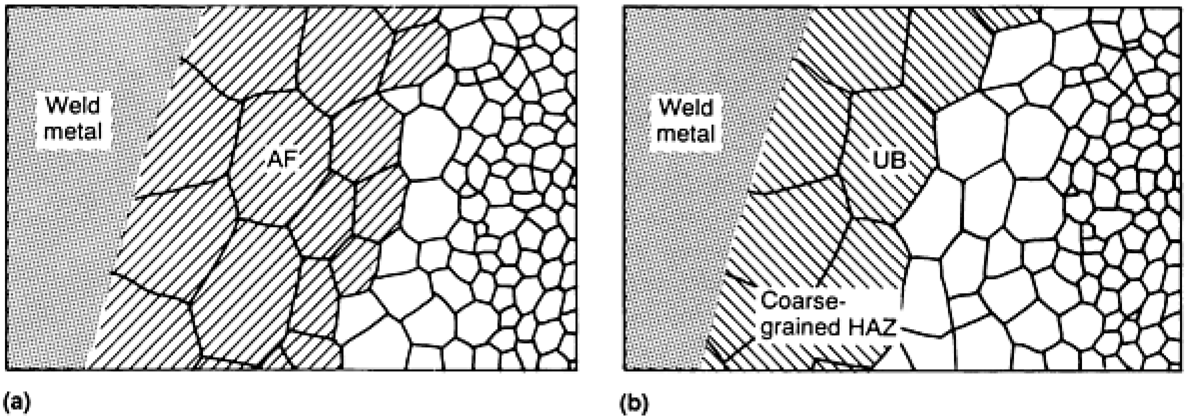

FIG. 25 SCHEMATIC SHOWING HAZ MICROSTRUCTURE IN SELECTED HIGH HEAT INPUT WELDS. (A) TITANIUM OXIDE STEEL. (B) TITANIUM NITRIDE STEEL. AF, ACICULAR FERRITE; UB, UPPER BAINITE. SOURCE: REF 30

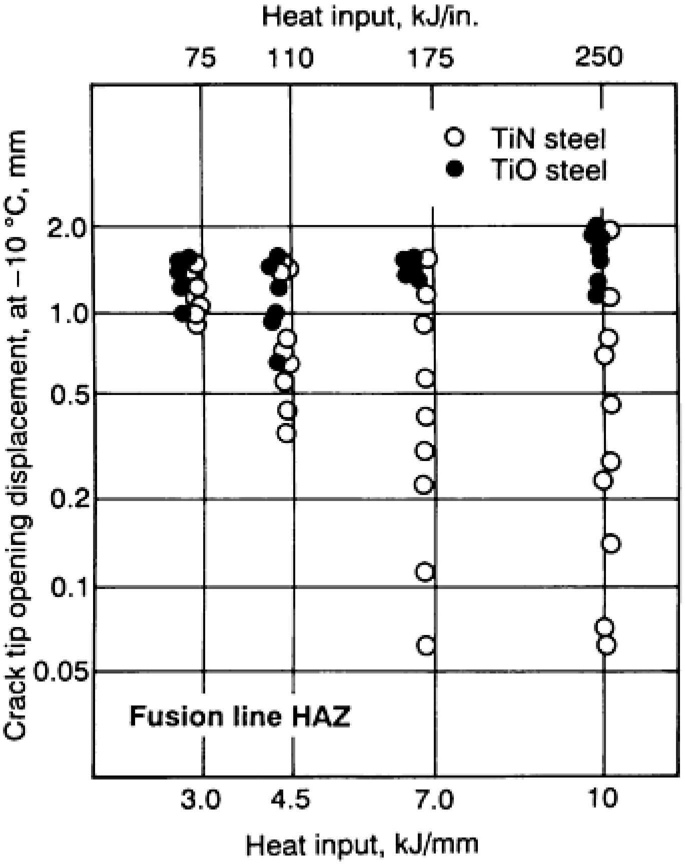

FIG. 26 HAZ TOUGHNESS OF TITANIUM NITRIDE AND TITANIUM OXIDE STEELS WITH 420 MPA (60 KSI) YIELD STRENGTH. SOURCE: REF 31

Effect of Transformations on Transient Weld Stresses. The effect of transformations on weld stresses (and vice versa) is best illustrated by introducing an apparently little-known proposal (Ref 32) about the use of a duplex austenite-martensite weld metal for welding high-strength steels. First, to prepare the background, the way in which transient stresses develop in a weld and the effect of transformations on weld stresses will be discussed.

Four uniaxial specimens using an HAZ simulation technique, the manner in which residual stress is accumulated during a weld thermal cycle was investigated (Ref 33), and the results are shown in Fig. 27. Initially, when a material is austenite at high temperature, its yield strength is low and only a small tensile stress can develop. On cooling, the tensile stress increases. When a phase transformation occurs, the resulting volume expansion opposes the contraction due to cooling and decreases the stress. Once the transformation is complete, the tensile stress increases again. As seen in Fig. 27, if the transformation occurs at a lower temperature (for example, in the case of 9Cr1Mo or 12CrMo steels), the magnitude of the final residual stress is less. The dotted line in the plot shows the expected variation for a material with an austenite-martensite structure. The greater the amount of martensite, the lower the final residual stress.

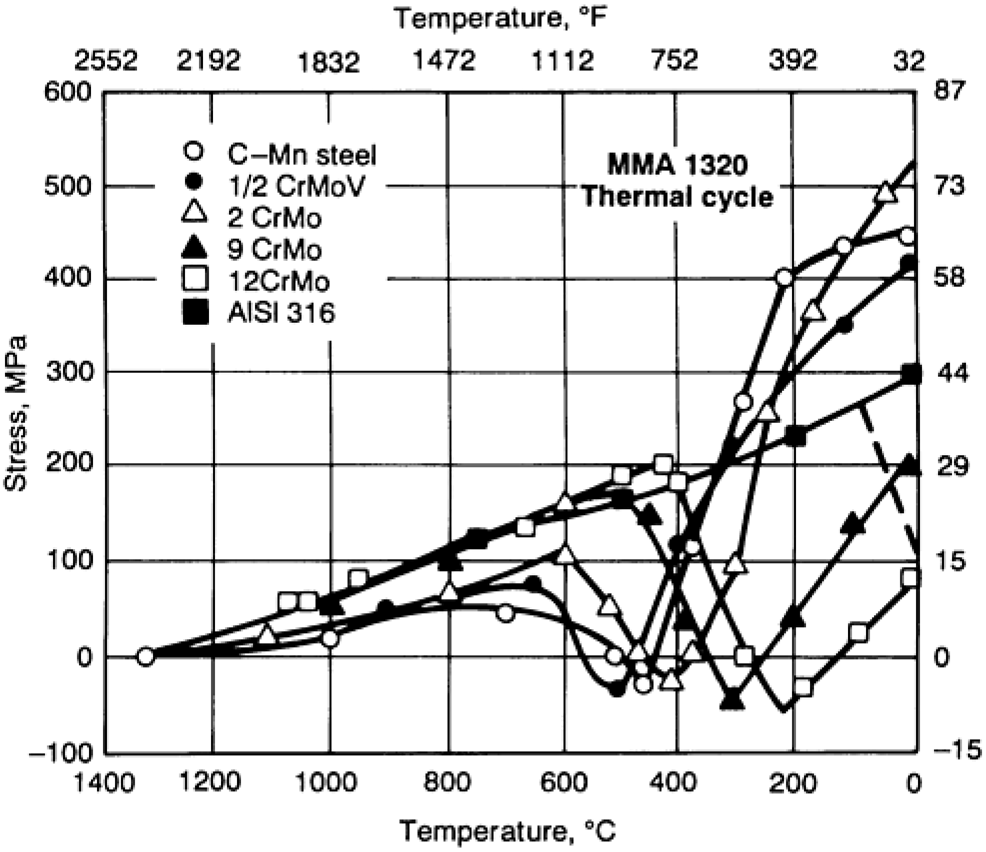

FIG. 27 STRESS ACCUMULATED DURING COOLING UNDER RESTRAINT FOR SELECTED STEELS. THE DOTTED LINE SHOWS THE EXPECTED VARIATION IN STRESS FOR A DUPLEX AUSTENITE-MARTENSITE WELD METAL WHERE TRANSFORMATION BELOW ABOUT 100 °C (212 °F) GREATLY REDUCES THE FINAL STRESS. (MMA 1320 INDICATES THE PEAK TEMPERATURE OF THE SIMULATED WELD THERMAL CYCLE AND CORRESPONDS TO THE TEMPERATURE AT THE START OF THE COOLING CYCLE.) SOURCE: REF 33

As discussed in the section "Weld Metal Toughness Relative to the Microstructure" in this article, the main problems in the welding of high-strength steels (with yield strengths of 600 to 1000 MPa, or 85 to 145 ksi) are hydrogen-induced cracking (HIC) and low toughness in the weld metal. Both problems can be addressed (in principle at least) by the use of a duplex austenite-martensite weld metal. Austenite can contribute by increasing the toughness and by minimizing the risk for HIC (Ref 14). The role of martensite will be to increase the strength and to decrease the final residual tensile stress, which also makes conditions less favorable for HIC. An additional benefit will be that the higher tensile residual stress in the weld metal before martensite forms will induce the transformations in the HAZ to occur at a slightly higher temperature, by Le Chatelier's principle (Ref 34), and thereby cause a microstructure less susceptible to HIC to form in the HAZ. The results of this effect are shown in Fig. 28, where it is seen that a tensile stress in the 500 to 300 °C (930 to 570 °F) temperature range on a ferrite-pearlite weld increased the stresses for failure (Ref 35).

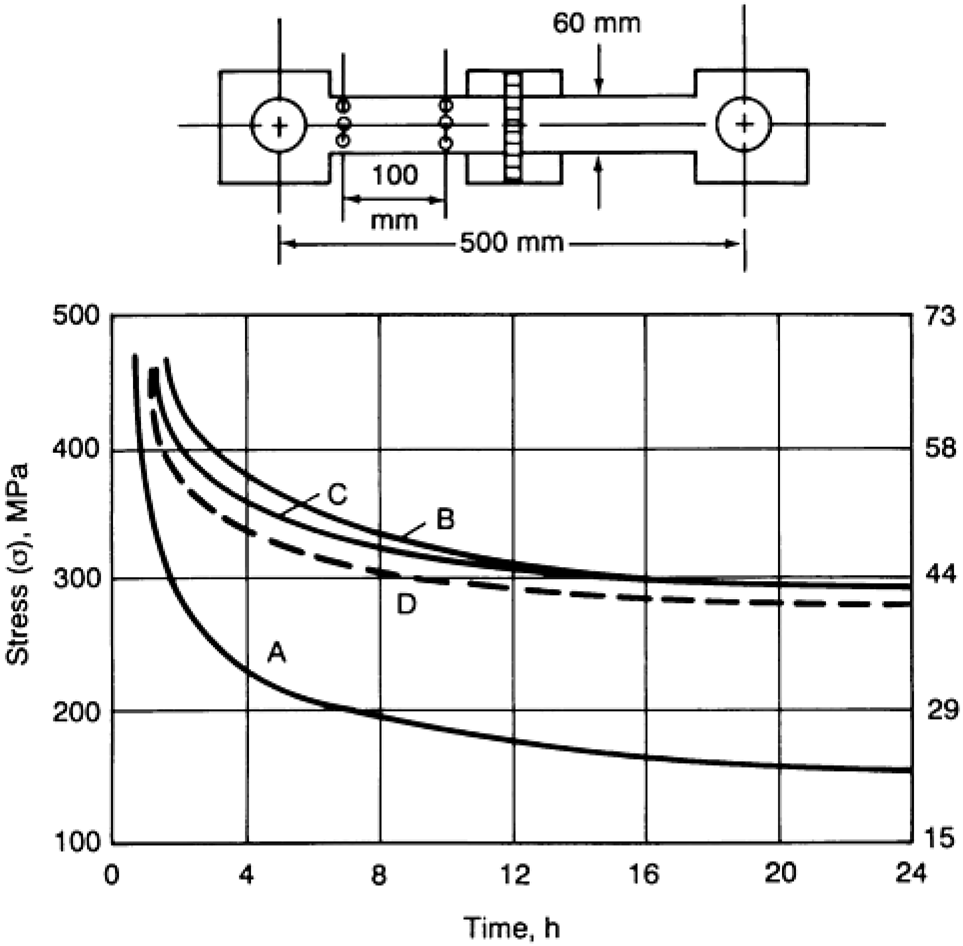

FIG. 28 DELAYED FRACTURE CURVES FOR TENSILE TESTS ON WELDED JOINTS IN 35KH3N3M (A RUSSIAN SPECIFICATION) STEEL. A, FERRITIC-PEARLITIC WELD; B, BAINITIC-MARTENSITIC WELD; C, AUSTENITIC WELD; D, FERRITIC-PEARLITIC WELD PLUS DEFORMATION AT 500 TO 300 °C (930 TO 570 °F). THE SCHEMATIC SHOWS THE SPECIMEN DIMENSIONS FOR THE DELAYED FRACTURE TENSILE TESTS ON WELDED JOINTS. SOURCE: REF 35

It must be mentioned that the idea of using an austenite-martensite weld metal represents a considerable departure from conventional thinking. For example, for certain overlaying applications of stainless steel deposits on low-alloy steel, an overalloyed electrode like E310 or E312 is preferred to E309. Even though the weld metal composition with any of the three electrodes would be acceptable, the use of overalloyed electrodes will minimize the width of the martensitic layer in the unmixed zone at the weld-HAZ interface (Ref 36). By contrast, an austenite-martensite weld metal implies that the choice of consumable will have to be such that the width of the martensitic layer will increase. However, because the proposal is meant for the recently developed high-strength steels that have a low carbon content, only a low-carbon martensite will form, which will not be that harmful. The benefits already discussed could override conventional concerns about having martensite in the weld metal in this case.

To sum up, even though the idea of using an austenite-martensite weld metal for high-strength steels does not seem to have been taken up for serious investigation so far, it has been included here because it is a nice way of illustrating the principles discussed. In addition, it shows that a fundamental knowledge of different effects on phase transformations can lead to unconventional solutions to practical problems.

References cited in this section

3. O. GRONG AND D.K. MATLOCK, MICROSTRUCTURAL DEVELOPMENT IN MILD AND LOW ALLOY STEEL WELD METALS, *INT. MET. REV.*, VOL 31 (NO. 1), 1986, P 27-48

6. K. EASTERLING, *INTRODUCTION TO THE PHYSICAL METALLURGY OF WELDING,* BUTTERWORTHS, 1983
14. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987
22. L.E. SAMUELS, *OPTICAL MICROSCOPY OF CARBON STEELS,* AMERICAN SOCIETY FOR METALS, 1980
23. "GUIDE TO THE LIGHT MICROSCOPE EXAMINATION OF FERRITIC STEEL WELD METALS," DOC. NO. IX-1533-88, INTERNATIONAL INSTITUTE OF WELDING, 1988
24. H.K.D.H. BHADESHIA AND L.E. SVENSSON, MODELLING THE EVOLUTION OF MICROSTRUCTURE OF STEEL WELD METAL, *MATHEMATICAL MODELLING OF WELD PHENOMENA,* H. CERJAK AND K. EASTERLING, ED., THE INSTITUTE OF MATERIALS, 1993
25. G.R. EDWARDS AND S. LIU, RECENT DEVELOPMENTS IN HSLA STEEL WELDING, *ADVANCES IN WELDING METALLURGY,* FIRST U.S.-JAPAN SYMPOSIUM, AMERICAN WELDING SOCIETY AND OTHERS, JUNE 1990, P 215-293
26. B. AHLBLOM, "OXYGEN AND ITS ROLE IN DETERMINING WELD METAL MICROSTRUCTURE AND TOUGHNESS--A STATE OF THE ART REVIEW," DOC. NO. IX-1322-84, INTERNATIONAL INSTITUTE OF WELDING, 1984
27. L.-E. SVENSSON AND B. GRETOFT, MICROSTRUCTURE AND IMPACT TOUGHNESS OF C-MN WELD METALS. *WELD. J.,* DEC 1990, P 454S-461S
28. P.W. HOLSBERG AND R.J. WONG, WELDING OF HSLA-100 STEEL FOR NAVAL APPLICATIONS, *WELDABILITY OF MATERIALS,* R.A. PATTERSON AND K.W. MAHIN, ED., ASM INTERNATIONAL, 1990
29. J.M.B. LOSZ AND K.D. CHALLENGER, MICROSTRUCTURE AND PROPERTIES OF A COPPER PRECIPITATION STRENGTHENED HSLA STEEL WELDMENT, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID AND J.M. VITEK, ED., ASM INTERNATIONAL, 1990
30. H. HOMMA, S. OHKITA, S. MATSUDA, AND K. YAMAMOTO, IMPROVEMENTS OF HAZ TOUGHNESS IN HSLA STEEL BY INTRODUCING FINELY DISPERSED TI-OXIDE, *WELD. J.,* OCT 1987, P 301S-309S
31. N. YURIOKA, "MODERN HIGH STRENGTH STEEL IN JAPAN," FIFTH INTERNATIONAL SYMPOSIUM, JAPAN WELDING SOCIETY, APRIL 1990 (TOKYO)
32. YU.N. GOTALSKII, THE PROBLEM OF WELDING HIGH STRENGTH STEELS, *AUTOM. WELD.,* JUNE 1984, P 37-40
33. W.K.C. JONES AND P.J. ALBERRY, THE ROLE OF PHASE TRANSFORMATIONS IN THE DEVELOPMENT OF RESIDUAL STRESSES DURING THE WELDING OF SOME FAST REACTOR STEELS, *PROC. CONF. FERRITIC STEELS FOR FAST REACTOR STEAM GENERATORS,* BRITISH NUCLEAR ENERGY SOCIETY, 1977
34. *MCGRAW-HILL ENCYCLOPEDIA OF SCIENCE AND TECHNOLOGY,* VOL 9, P 685
35. A.M. MAKARA AND N.A. MOSENDZ, EFFECTS OF THE WELD METAL ON CRACKING IN HAZ, *WELD. RES. ABROAD,* NOV 1965, P 78-86
36. H. IKAWA, S. SHIN, M. INUI, Y. TAKEDA, AND A. NAKANO, "ON THE MARTENSITE-LIKE STRUCTURE AT WELD BOND AND THE MACROSCOPIC SEGREGATION IN WELD METAL IN THE WELDED DISSIMILAR METALS OF α-STEELS AND γ-STEELS," DOC. NO. IX-785-72, INTERNATIONAL INSTITUTE OF WELDING, 1972

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

HAZ in Multipass Weldments

In the HAZ of a single-pass weld, the grain-coarsened zone (GC HAZ) is normally the region having the lowest toughness. Turning to a multipass weld, Fig. 29 (compare with Fig. 3) shows how the GC HAZ can be modified by subsequent passes and can be categorized into four regions, depending on the reheating temperature (Ref 37):

- SUBCRITICALLY REHEATED GRAIN-COARSENED (SCGC) ZONE, THE ZONE REHEATED BELOW AC_1
- INTERCRITICALLY REHEATED GRAIN-COARSENED (ICGC) ZONE, THE ZONE REHEATED BETWEEN AC_1 AND AC_3
- SUPERCRITICALLY REHEATED GRAIN-REFINED (SCGR) ZONE, THE ZONE REHEATED ABOVE AC_3 AND BELOW ABOUT 1200 °C (2190 °F)
- UNALTERED GRAIN-COARSENED (UAGC) ZONE, THE ZONE THAT IS NOT REHEATED ABOVE ABOUT 200 °C (390 °F) OR THE ZONE THAT IS AGAIN REHEATED ABOVE ABOUT 1200 °C (2190 °F)

Figure 30 shows how the crack tip opening displacement (CTOD) value of simulated specimens varies with the peak temperature of the second thermal cycle, T_{p2} (T_{p1} = 1400 °C, or 2550 °F). It is seen that the ICGC, UAGC, and SCGC regions have CTOD values less than about 0.1 mm (0.004 in.). Similar low values have been obtained by locating the crack tip in the CTOD tests at corresponding locations in the HAZ of actual multipass weldments. These low toughness regions are commonly known as local brittle zones (LBZs).

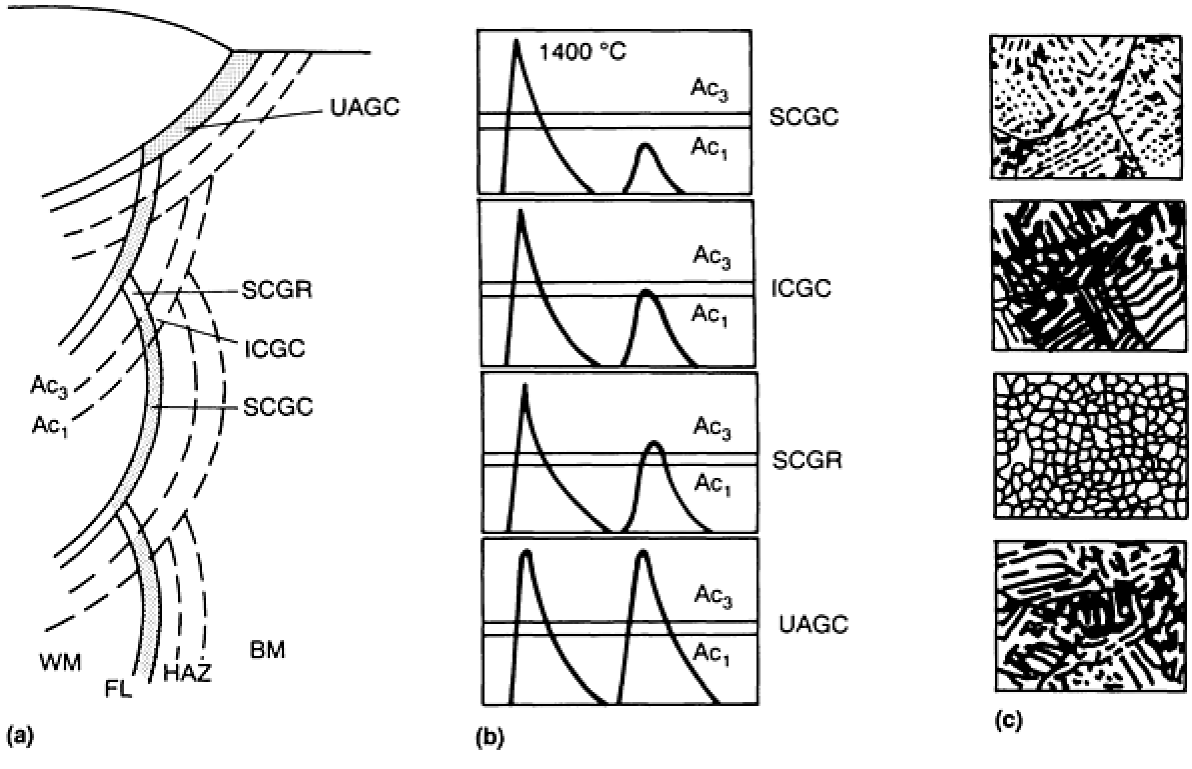

FIG. 29 SCHEMATIC SHOWING THE DIFFERENT SUBZONES THAT CAN FORM IN THE COARSE-GRAINED REGION OF THE HAZ IN A MULTIPASS WELD. (A) POSITION OF SUBZONES RELATIVE TO BASE METAL (BM) AND WELD METAL (WM). (B) PLOT OF THERMAL CYCLES RELATIVE TO AC_3 AND AC_1. (C) MICROSTRUCTURES AT THE DIFFERENT ZONES. FL REFERS TO THE FUSION LINE. SEE TEXT FOR DETAILS. SOURCE: REF 37

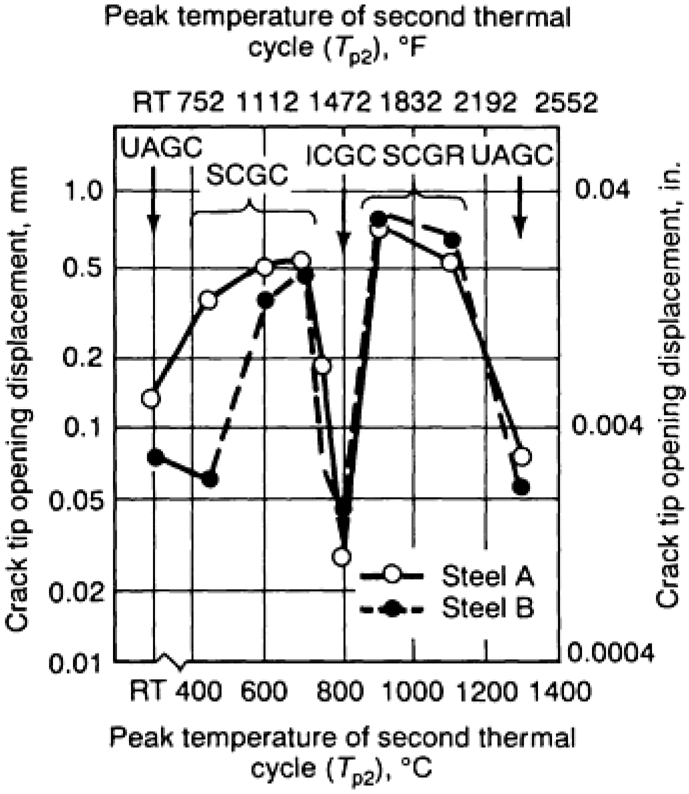

FIG. 30 PLOT OF CRACK TIP OPENING DISPLACEMENT VERSUS PEAK TEMPERATURE OF THE SECOND THERMAL CYCLE FOR SIMULATED SPECIMENS SUBJECTED TO A DOUBLE THERMAL CYCLE. (T_{P1} = 1400 °C = 2550 °F; ΔT_{8-5} = 20 S). THE HAZ ZONES (FIG. 29) CORRESPONDING TO THE SIMULATION TRIALS ARE ALSO SHOWN IN THE FIGURE. SOURCE: REF 37

The low CTOD values have been obtained only in tests on modern structural steels, whereas no actual structural failure attributable to the LBZs has been reported so far (Ref 38). Because the structural significance of the low toughness test results is still largely unresolved, the present conservative strategies are to choose a steel whose tendency to form LBZs is less or to choose a welding procedure by which the size of the LBZs can be reduced.

The ICGC HAZ usually has a lower toughness than the SCGC or UAGC regions, even though all of them have nearly the same γ grain size. This is due to the higher amount of highcarbon martensite-austenite (M-A) constituent in the ICGC HAZ. When the GC HAZ is reheated to a temperature between Ac_1 and Ac_3, austenite is nucleated at the high carbon areas. Upon cooling, these local high-carbon areas can transform to give twinned martensite with every thin regions of retained austenite in between. (See Ref 39 for detailed metallographic procedures for identifying the M-A constituent.) The carbon content in the M-A islands can range from about 0.3 to 0.5% C. The significance of this is that, for a given nominal carbon content in the steel, the volume fraction of the M-A constituent will be much higher than if most of the carbon formed carbides (by the lever rule). This will increase the number of crack nucleation sites and thereby contribute to the inferior toughness of ICGC HAZ. As is to be expected, the volume fraction of the M-A constituent also depends on the hardenability of the steel, which in turn depends on the alloying content. It has been shown that if the development of pearlitic microphases could be promoted instead of the development of M-A, by decreasing the alloy content, the toughness of the ICGC HAZ could be improved (Ref 40). However, this would have penalties in terms of achievement of parent plate strength. A more feasible solution would be to inhibit grain growth in the HAZ.

The multipass welding procedure can alternatively be controlled to limit the size of the LBZs. Figure 31 shows how this can be done with a tandem three-wire high current gas-metal arc welding procedure by adjusting the distance between the three arcs. Special "temper-bead" procedures (Ref 42) have been developed for controlling the microstructure in the HAZ, and a need for these procedures arises in the following way. Low-alloy steel weldments for critical applications (for example, pressure vessels) require a postweld heat treatment (PWHT) in a furnace. This is done to temper the hard regions in the HAZ and to relieve residual stresses. If repairs become necessary on site after the component has been in service, PWHT is usually not feasible. The heat of the arc can then be used to achieve the tempering function of PWHT by suitable spatial positioning and sequencing of the individual passes. Grain refinement in he HAZ is also sought to increase the toughness and thereby offset the harmful effects of residual stresses that would remain in the absence of PWHT.

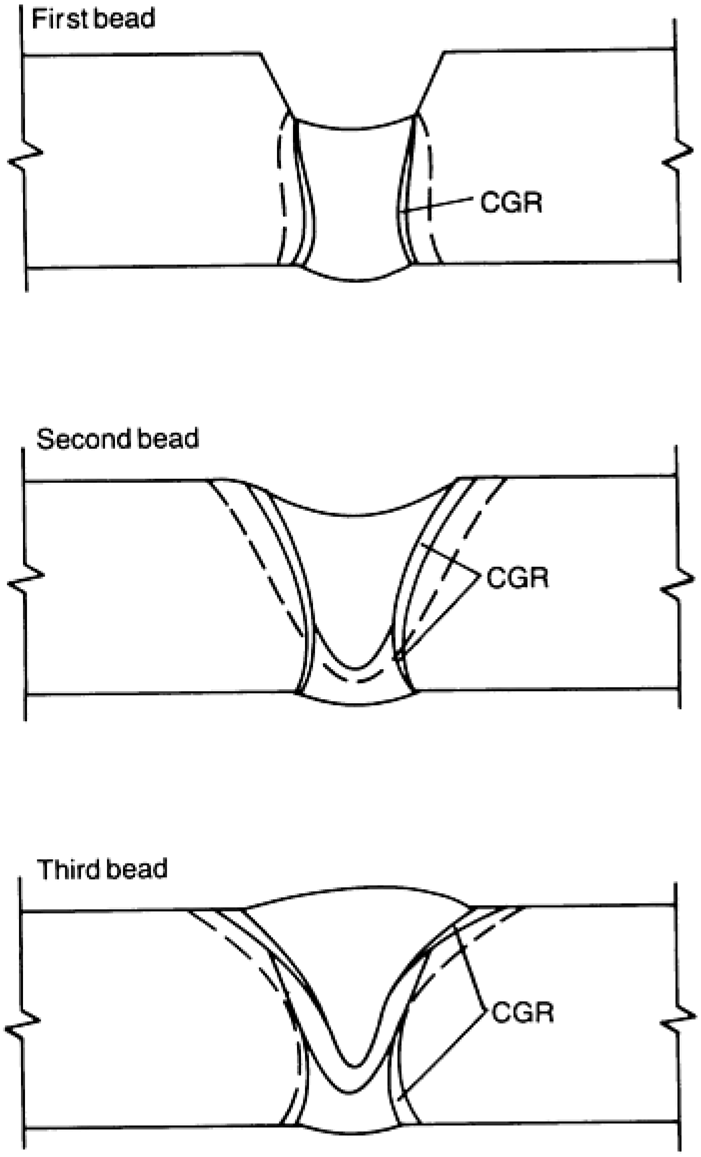

FIG. 31 SCHEMATIC SHOWING THAT THE HAZ ISOTHERMS AND THE SIZE AND LOCATION OF THE COARSE-GRAINED REGION (CGR) CAN BE CONTROLLED IN A TANDEM THREE-WIRE HIGH CURRENT GAS-METAL ARC WELDING PROCEDURE. GRAIN REFINEMENT OF COARSE-GRAINED REGIONS INITIALLY FORMED IS OBTAINED BY OPTIMIZING THE DISTANCES BETWEEN THE THREE ARCS. SOURCE: REF 41

An example of a two-layer temper bead procedure is shown in Fig. 32. The heat inputs of the first and second layer are carefully controlled, so that the heat from the second layer is used to refine the coarse-grained region in the HAZ of the base metal due to the first layer. This ideas can be extended one step further by using a pulsed gas-tungsten arc welding procedure whereby a given pulse in a pulsed weldment can be used to successively refine and temper the preceding pulse pitch regions (Ref 11). The degree of microstructural refinement depends primarily on the welding speed and pulsing frequency, and these parameters can be controlled with great precision. By contrast, an important variable to be controlled in the temper-bead procedure is the weld deposit height, and it is difficult to exercise the same degree of control on this. For this reason, it is argued that the maximum possible control on the microstructural changes in the repair weldment is possible with the pulsing procedure.

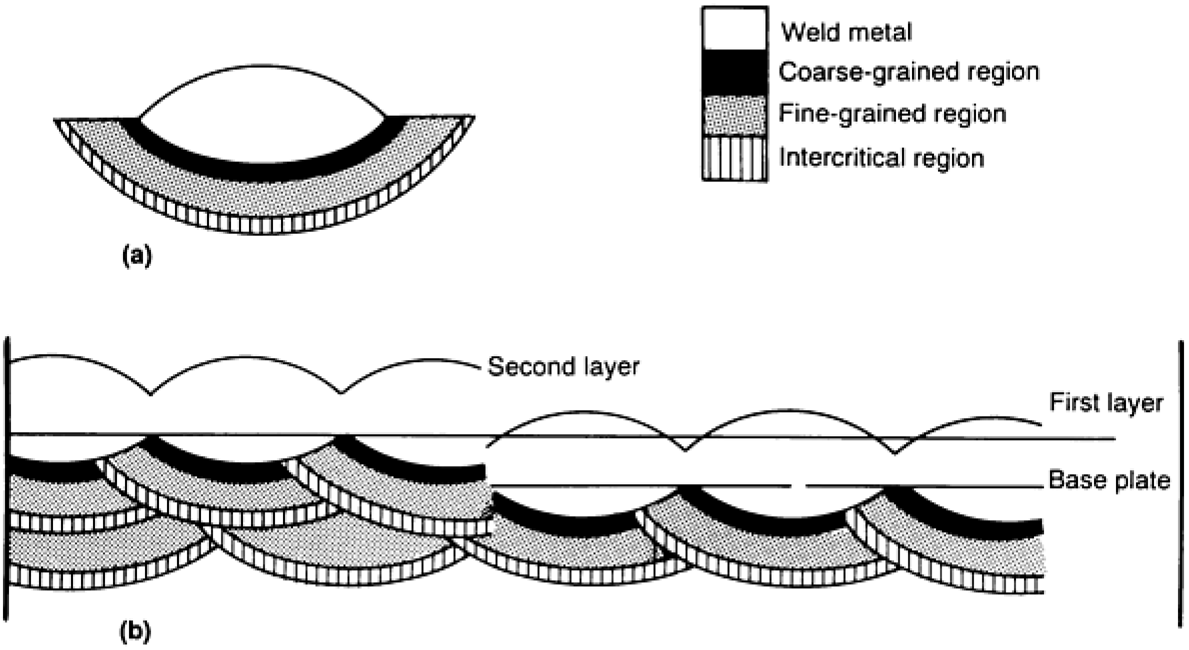

FIG. 32 SCHEMATIC SHOWING A TWO-LAYER REPAIR PROCEDURE. (A) HEAT-AFFECTED AREAS OF A SINGLE WELD BEAD. (B) FIRST LAYER CAUSES COARSE-GRAINED REGIONS TO FORM IN THE HAZ OF THE BASE METAL (RIGHT PORTION OF THE FIGURE). DEPOSITION OF THE SECOND LAYER REFINES THESE INITIAL COARSE-GRAINED REGIONS (LEFT PORTION OF FIGURE). SOURCE: REF 42

References cited in this section

11. P. RAVI VISHNU AND K. EASTERLING, PHENOMENOLOGICAL MODELLING OF HEAT FLOW AND MICROSTRUCTURAL CHANGES IN PULSED GTAW WELDS OF A QT STEEL, *MATHEMATICAL MODELING OF WELD PHENOMENA,* H. CERJAK AND K. EASTERLING, ED., THE INSTITUTE OF METALS, 1993
37. T. HAZE AND S. AIHARA, "INFLUENCE OF TOUGHNESS AND SIZE OF LOCAL BRITTLE ZONES ON HAZ TOUGHNESS OF HSLA STEELS," SEVENTH INTERNATIONAL CONFERENCE ON OFFSHORE MECHANICS AND ARCTIC ENGINEERING, HOUSTON, 1988
38. C.P. ROYER, A USER'S PERSPECTIVE ON HEAT AFFECTED ZONE TOUGHNESS, *WELDING METALLURGY OF STRUCTURAL STEELS,* J.Y. KOO, ED., TMS-AIME, 1987
39. F. MATSUDA, Z. LI, P. BERNASOVSKY, K. ISHIHARA, AND H. OKADA, "AN INVESTIGATION OF THE BEHAVIOUR OF M-A CONSTITUENT IN SIMULATED HAZ OF HSLA STEELS," DOC. NO. IX-B-1591-90, INTERNATIONAL INSTITUTE OF WELDING, 1990
40. P.L. HARRISON AND P.H.M. HART, RELATIONSHIPS BETWEEN HAZ MICROSTRUCTURE AND CTOD TRANSITION BEHAVIOUR IN MULTIPASS C-MN STEEL WELDS, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID AND J.M. VITEK, ED., ASM INTERNATIONAL, 1990
41. H. ONOE, J. TANAKA, AND I. WATANABE, JAPANESE LNG TANKER CONSTRUCTED USING A NEW WELDING PROCESS AND IMPROVED AL-KILLED STEELS, *MET. CONSTR.,* JAN 1979, P 26-31
42. P.J. ALBERRY, SENSITIVITY ANALYSIS OF HALF-BEAD AND ALTERNATIVE GTAW TECHNIQUES, *WELD. J.,* NOV 1989, P 442S-451S

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

Fusion Zone in Multipass Weldments

In the weld metal of a multipass weld, reheating effects will lead to a gradient in microstructure similar to the case of the HAZ. However, instead of a detailed classification, the multipass weld metal is usually considered to consist of just two regions (Fig. 33):

- *AS-DEPOSITED OR PRIMARY REGION,* WHERE THE MICROSTRUCTURE DEVELOPS AS THE WELD COOLS FROM THE LIQUID PHASE TO AMBIENT TEMPERATURE
- *REHEATED OR SECONDARY REGION,* WHERE REGIONS WITH THE ORIGINAL PRIMARY MICROSTRUCTURE ARE REHEATED TO TEMPERATURES ABOVE THE AC_1 TEMPERATURE. THE TEMPERED REGIONS WHICH ARE REHEATED TO SLIGHTLY LOWER TEMPERATURES ARE ALSO GENERALLY CONSIDERED TO BELONG TO THIS CATEGORY

The properties of the weld metal depend on the relative area or volume fractions of the two regions, which in turn depend on the welding procedure, so the properties are procedure-specific. Hence arises the need for welding procedure qualification as per codes and standards in addition to the qualification of consumables (for which is standard procedure is specified).

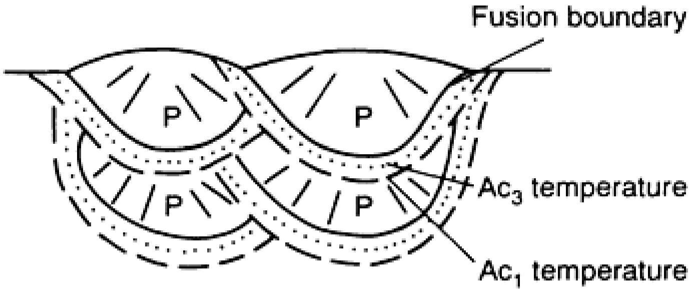

FIG. 33 PRIMARY (P) AND REAUSTENITIZED REGIONS IN THE WELD METAL REGION OF A MULTIPASS WELD. THE REAUSTENITIZED REGION IS WHERE THE COLUMNAR STRUCTURE IS NOT CLEARLY DETECTED. SOURCE: REF 23

Reference cited in this section

23. "GUIDE TO THE LIGHT MICROSCOPE EXAMINATION OF FERRITIC STEEL WELD METALS," DOC. NO. IX-1533-88, INTERNATIONAL INSTITUTE OF WELDING, 1988

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

Stainless Steels

It has long been known that solidification cracking can be avoided in austenitic stainless steel welds by having a small concentration of ferrite in them. Recent work has shown, however, that residual ferrite content at room temperature is no more than a symptom and that it is really the solidification mode (whether the weld metal solidifies as primary austenite or ferrite) that is the deciding criterion (Ref 43). It has been found that susceptibility to solidification cracking is least for a primary ferrite solidification mode (more specifically, when the solidification mode corresponds to the types shown in Fig. 34c and 34d). It is believed that low-melting-point liquid phases (formed by the segregation of impurities like sulfur and phosphorus, for example) solidifying in the intercellular regions do not wet the δ-γ interphase boundaries as easily as they would δ-δ or γ-γ boundaries. In the ferritic-austenitic solidification mode (Fig. 34c and 34d), the δ-γ interphase boundary area is greater at temperatures just below the nominal solidus temperature, and this is the reason for a greater resistance to solidification cracking.

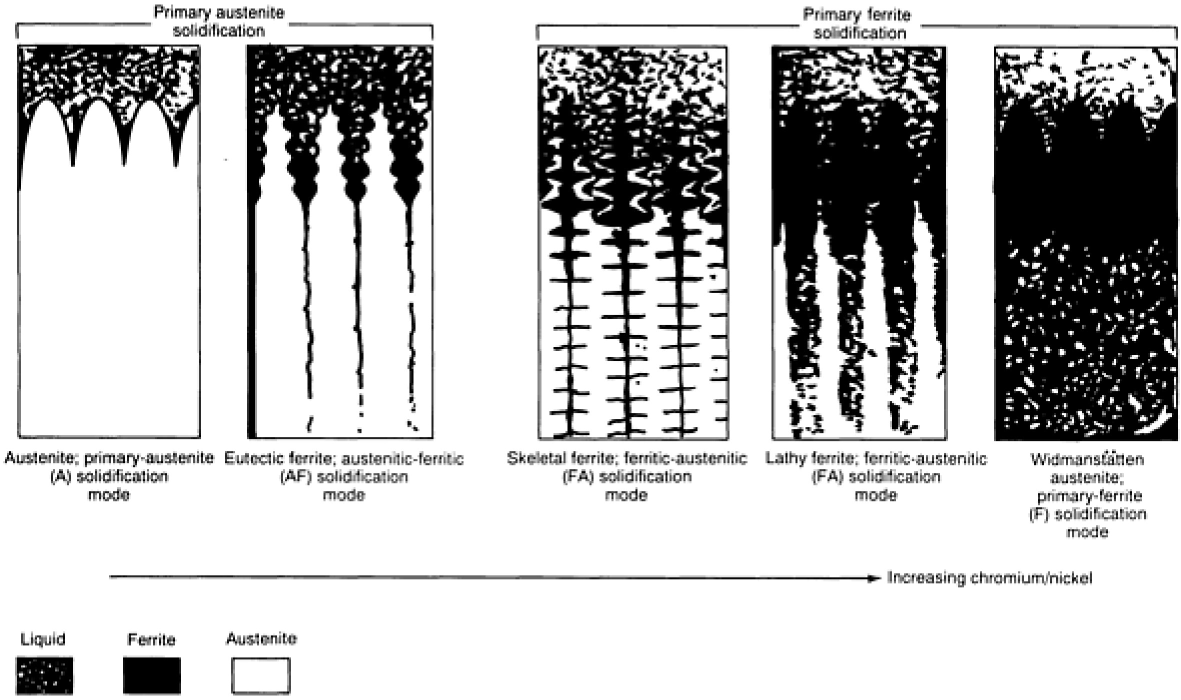
Primary austenite solidification
Primary ferrite solidification
Austenite; primary-austenite (A) solidification mode
Eutectic ferrite; austenitic-ferritic (AF) solidification mode
Skeletal ferrite; ferritic-austenitic (FA) solidification mode
Lathy ferrite; ferritic-austenitic (FA) solidification mode
Widmanstätten austenite; primary-ferrite (F) solidification mode
Increasing chromium/nickel
Liquid
Ferrite
Austenite

FIG. 34 SCHEMATIC SHOWING SOLIDIFICATION AND TRANSFORMATION BEHAVIOR RESULTING IN A RANGE OF FERRITE MORPHOLOGIES IN AUSTENITIC STAINLESS STEEL WELDS. SOURCE: REF 43

Information about the solidification mode can be obtained from the chemical composition of the weld metal by referring to the WRC-1992 diagram (Fig. 35). It has been found that this diagram is more accurate than the Schaeffler (for duplex stainless steel welds) and Delong (for 300 series austenitic steel welds) diagrams developed earlier. All of these diagrams are strictly applicable only for the cooling rates obtained in manual metal arc (MMA) welding because the data for establishing the statistical correlations have been taken from MMA welds. The dotted lines in Fig. 35 are the demarcation lines between the different solidification modes shown (that is, A, AF, FA, and F) corresponding to the types in Fig. 34(a), 34(b), 34(c), 34(d), and 34(e), respectively. The AF-FA boundary in Fig. 35 (which, as discussed earlier, corresponds to the onset of cracking) intersects the iso-ferrite number lines at an angle. This is consistent with the findings of several investigators that the minimum ferrite content necessary to avoid hot cracking is different for different weld metal compositions (that is, for weld metals deposited with E 316, E 308L, E 309, and so on).

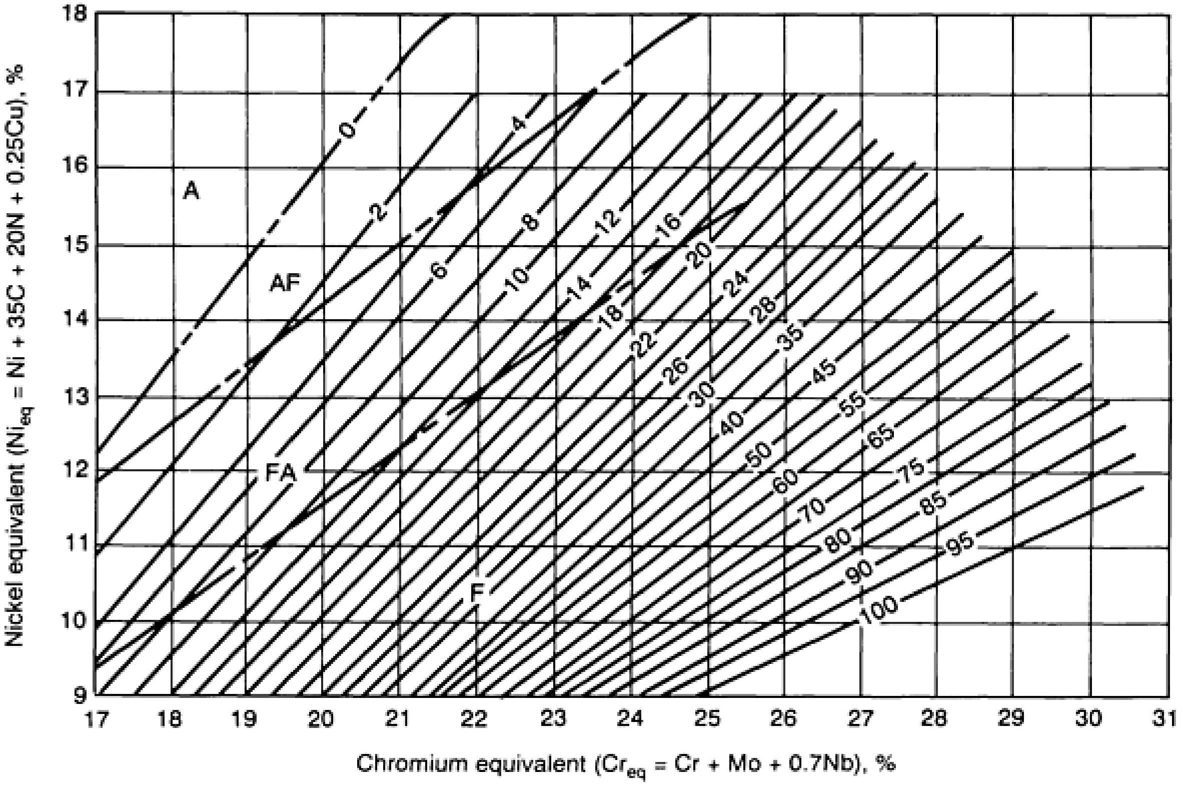

FIG. 35 WRC-1992 DIAGRAM PREDICTING FERRITE CONTENT IN STAINLESS STEELS. FERRITE CONTENT IS GIVEN BY THE FERRITE NUMBER (FN), WHERE 100 FN IS APPROXIMATELY EQUAL TO 65 VOL% FERRITE. BOUNDARIES DENOTING A CHANGE IN SOLIDIFICATION MODE (A, AF, FA AND F PER FIG. 34) ARE ALSO SHOWN (INDICATED BY DOTTED LINES). SOURCE: REF 45

It is a fortunate coincidence that the coefficients in the Cr_{eq} and Ni_{eq} expressions (see Fig. 35), which have been obtained by statistical fitting methods, result in satisfactory correlations with both the solidification mode and the ferrite number (a precisely measurable quantity that is roughly proportional to the actual volume percentage of ferrite at room temperature) (Ref 44). Indeed, it has been found that the same Cr_{eq} and Ni_{eq} expressions (that is, with the same elements and coefficients) are not able to satisfactorily fix the locations of the martensite and martensite-austenite phase fields seen in the lower left region of the Schaeffler diagram. For this reason, such regions of Cr_{eq} and Ni_{eq} are excluded in the WRC-1992 diagram.

A discussion of weld decay due to sensitization and knife line attack in the HAZ of austenitic stainless steels can be found in Ref 14 and 46.

References cited in this section

14. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987
43. J.A. BROOKS AND A.W. THOMPSON, MICROSTRUCTURAL DEVELOPMENT AND SOLIDIFICATION CRACKING SUSCEPTIBILITY OF AUSTENITIC STAINLESS STEEL WELDS, *INT. MATER. REV.,* VOL 36 (NO. 1), 1991
44. T.A. SIEWERT, C.N. MCCOWAN, AND D.L. OLSON, FERRITE NUMBER PREDICTION TO 100 FN IN STAINLESS STEEL WELD METAL, *WELD. J.,* DEC 1988, P 289S-298S
45. D.J. KOTECKI AND T.A. SIEWERT, WRC-1992 CONSTITUTION DIAGRAM FOR STAINLESS STEEL WELD METALS: A MODIFICATION OF THE WRC-1988 DIAGRAM, *WELD. J.,* MAY 1992, P 171S-178S
46. T.G. GOOCH AND D.C. WILLINGHAM, *WELD DECAY IN AUSTENITIC STAINLESS STEELS,* THE

WELDING INSTITUTE, UK, 1975

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

Aluminum Alloys

The main problems in the welding of heat-treatable aluminum alloys are liquation cracking in the partially melted zone (see the section "Unmixed and Partially Melted Zones in a Weldment" in this article) and softening in the HAZ (Ref 14, 47). The latter means that the weld joint strength can be reduced. Reasons for the softening can be understood by referring to Fig. 36. The parent material is assumed to be a 2000- or 6000-series aluminum alloy, artificially aged to contain the metastable phase of θ' (in aluminum-copper alloys), S' (in aluminum-copper-magnesium alloys), or β' (in aluminum-magnesium-silicon alloys). A very fine dispersion of precipitate particles is obtained in the parent material by a solution heat treatment followed by aging, and this is the reason for the high strength in the materials. Figure 36(b) shows that a gradient in hardness is obtained in the HAZ immediately after welding. At location 1, the high peak temperature causes the precipitate particles to go into solution, and the cooling rate is too high for reprecipitation. At locations 2 and 3, the precipitate particles partially dissolve and coarsen. After postweld artificial aging, fine precipitate particles are again formed at location 1, causing the hardness to increase to the parent material level. However, at locations 2 and 3, only a lower hardness can be obtained because of the formation of coarse precipitate particles. For increasing weld heat inputs, the width of the softened zone increases. A good discussion of how the above phenomenon can be modeled is found in Ref 48.

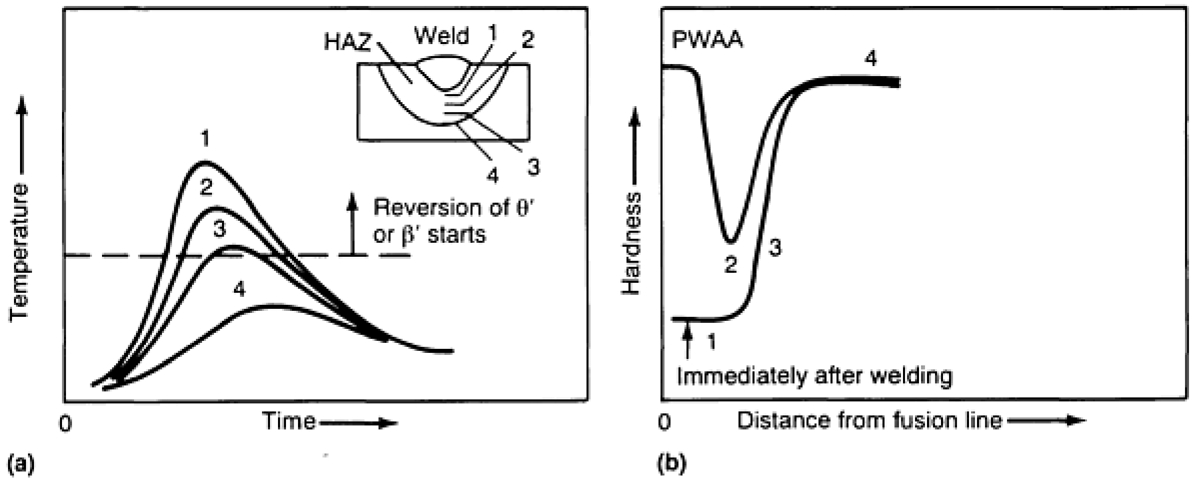

FIG. 36 SCHEMATIC SHOWING THE EFFECT OF WELD THERMAL CYCLES ON THE SOFTENING IN THE HAZ IN AGE-HARDENABLE ALUMINUM ALLOYS. (A) THERMAL CYCLES IN THE HAZ (FOR CORRESPONDING LOCATIONS IN WELD, SEE INSET. (B) HAZ HARDNESS PROFILES BEFORE AND AFTER AGING. PWAA, POSTWELD ARTIFICIAL AGING. SOURCE: REF 14

References cited in this section

14. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987
47. S. KOU, WELDING METALLURGY AND WELDABILITY OF HIGH STRENGTH ALUMINUM ALLOYS, *WELD. RES. COUNC. BULL.,* NO. 320, DEC 1986

48. H. SHERCLIFF AND M.F. ASHBY, A PROCESS MODEL FOR AGE HARDENING OF AL ALLOYS--I. MODEL AND II. APPLICATIONS OF THE MODEL, *ACTA METALL. MATER.*, VOL 38 (NO. 10), 1990, P 1789-1812

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

Nickel-Base Superalloys

Nickel-base superalloys have a very high creep resistance, making them attractive for critical high-temperature applications such as gas turbines. Yet it is this very characteristic that gives rise to the problem of strain age cracking in the weld HAZ (Ref 14, 19, 49). The high-temperature creep strength in the parent materials is obtained by forming fine precipitates of γ' ($Ni_3(Al, Ti)$) or γ" (Ni_3Nb) in them. In the region of the weld HAZ subjected to high peak temperatures, these precipitates dissolve, similar to the case of aluminum alloys. After welding, the weldments are again solution heat-treated and aged (note the difference between aluminum alloys and the nickel-base superalloys in this respect). The purposes of this treatment are to obtain the same uniform and fine precipitation in the weld as in the parent material and to relieve the weld residual stresses.

Normally (for example, in the case of low-alloy or carbon steel welds), PWHT in a furnace relieves the residual stresses because the yield stress at the PWHT temperature is very low. A residual stress greater than the yield stress cannot be supported, and stress relief is brought about by a process of creep and plastic deformation. This process is difficult in nickel-base superalloy weldments, however, because of their high creep strength and low creep ductility. Reprecipitation in the solutionized regions of the HAZ (during the heating-up to the PWHT temperature) strengthens the matrix and does not allow any creep or plastic deformation (necessary for the relief of residual stress) to take place. Cracking can occur in this condition at the grain boundaries in the grain-coarsened region of the HAZ. Cracks initiated here have been known to run far into the base material. This problem is very difficult to solve, and the best solution has been found to be to switch over to materials such as Inconel 718 alloy that are inherently less susceptible to cracking, even though their creep strength is somewhat lower. Inconel 718 alloy contains γ" precipitate, and the slow kinetics of precipitation in this alloy allow substantial stress relief to be achieved before precipitation increases the creep strength in the HAZ.

References cited in this section

14. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987
19. W. YENISCAVITCH, JOINING, *SUPERALLOYS II,* C.T. SIMS, N.S. STOLOFF, AND W.C. HAGEL, ED., JOHN WILEY & SONS, 1987, P 495-516
49. W.A. OWCZARSKI, PROCESS AND METALLURGICAL FACTORS IN JOINING SUPERALLOYS AND OTHER HIGH SERVICE TEMPERATURE MATERIALS, *PHYSICAL METALLURGY OF JOINING,* R. KOSSOWSKY AND M.E. GLICKSMAN, ED., TMS-AIME, 1980

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

Titanium Alloys

Titanium alloys find use in aerospace applications, pressure vessels, and so on because of their high strength-to-weight ratio, high corrosion resistance, and high fracture toughness. Pure titanium has two allotropic forms: low-temperature hexagonal close-packed (hcp) α-phase and elevated-temperature body-centered cubic (bcc) β-phase. Various alloying elements tend to preferentially stabilize one or the other of these phases. As a result, titanium alloys are generally classified as α, α+ β, metastable β, and β alloys (Fig. 37).

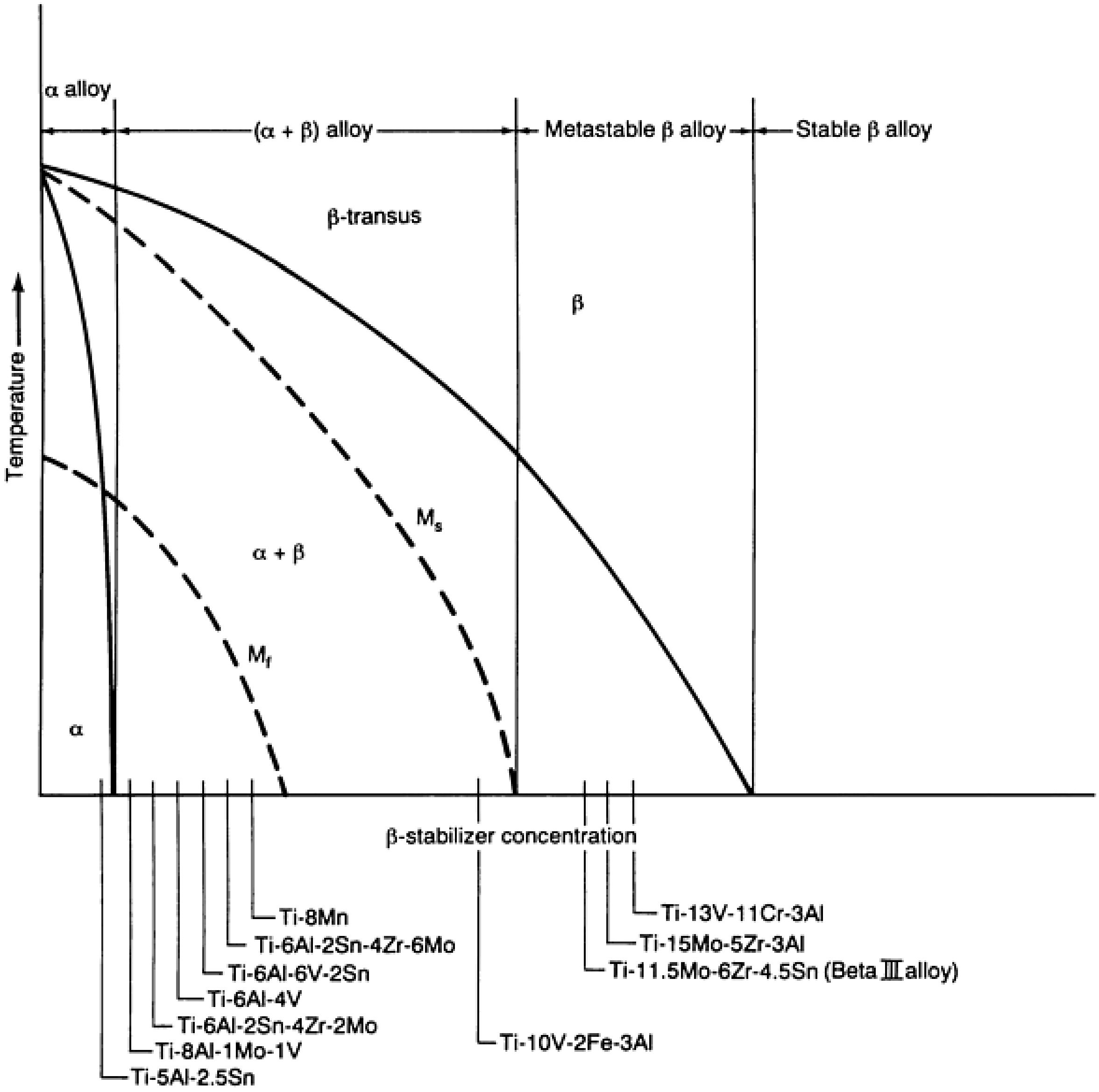
α alloy
(α + β) alloy
Metastable β alloy
Stable β alloy
β-transus
β
Temperature
M_s
α + β
M_f
α
β-stabilizer concentration
Ti-8Mn
Ti-6Al-2Sn-4Zr-6Mo
Ti-6Al-6V-2Sn
Ti-6Al-4V
Ti-6Al-2Sn-4Zr-2Mo
Ti-8Al-1Mo-1V
Ti-5Al-2.5Sn
Ti-10V-2Fe-3Al
Ti-13V-11Cr-3Al
Ti-15Mo-5Zr-3Al
Ti-11.5Mo-6Zr-4.5Sn (Beta III alloy)

FIG. 37 HYPOTHETICAL β ISOMORPHOUS PHASE DIAGRAM THAT SHOWS RELATION BETWEEN β -STABILIZER SOLUTE CONTENT AND TITANIUM ALLOY FAMILY DESCRIPTION. SELECTED COMMON ALLOYS ARE LISTED ALONG THE *X*-AXIS OF THE DIAGRAM BASED ON THEIR COMPOSITIONS. SOURCE: REF 50

The α-alloys are not heat-treatable (that is, a significant increase in strength and toughness cannot be obtained by heat treatment), and they have the lowest strength of the four titanium alloy families. Interestingly enough, these alloys are thus considered to have excellent weldability. What this means in practice is that there is no significant degradation in the properties of the weld and HAZ with regard to the parent material due to the weld thermal cycles. Also, the β ↔ α transformation in the weld and near HAZ means that the extent of grain coarsening will not be so high as in other single-phase materials in which no phase transformation occurs upon heating or cooling, such as stabilized ferritic stainless steels (Ref 51), pure aluminum, and so on.

In the α+ β alloys, a higher strength is obtained with a higher volume fraction of the β phase, which in turn is obtained with a higher β -stabilizing alloy content (Fig. 37). The effect of cooling rate on the phase transformations in two α+ β alloys is shown schematically in Fig. 38. As in the case of steels, it is seen that the β ↔ αtransformation can occur either by nucleation and growth at slower cooling rates or martensitically at higher cooling rates. The martensitic product in the lean α+ β alloys has an hcp crystal structure (designated as α') or an orthorhombic one (designated as α'') in the β-stabilized α+ β alloys. It is seen that the C-curve for the β ↔ intergranular α transformation is shifted to the right for alloys with increasing β-stabilizing alloy content, while that for the allotriomorphic α (more commonly referred to as GB

α) is relatively insensitive to alloy concentration (Ref 52). As a result, alloys that are fairly rich in β-stabilizing additions exhibit a stronger tendency to form a continuous network of GB α than do leaner α + β alloys.

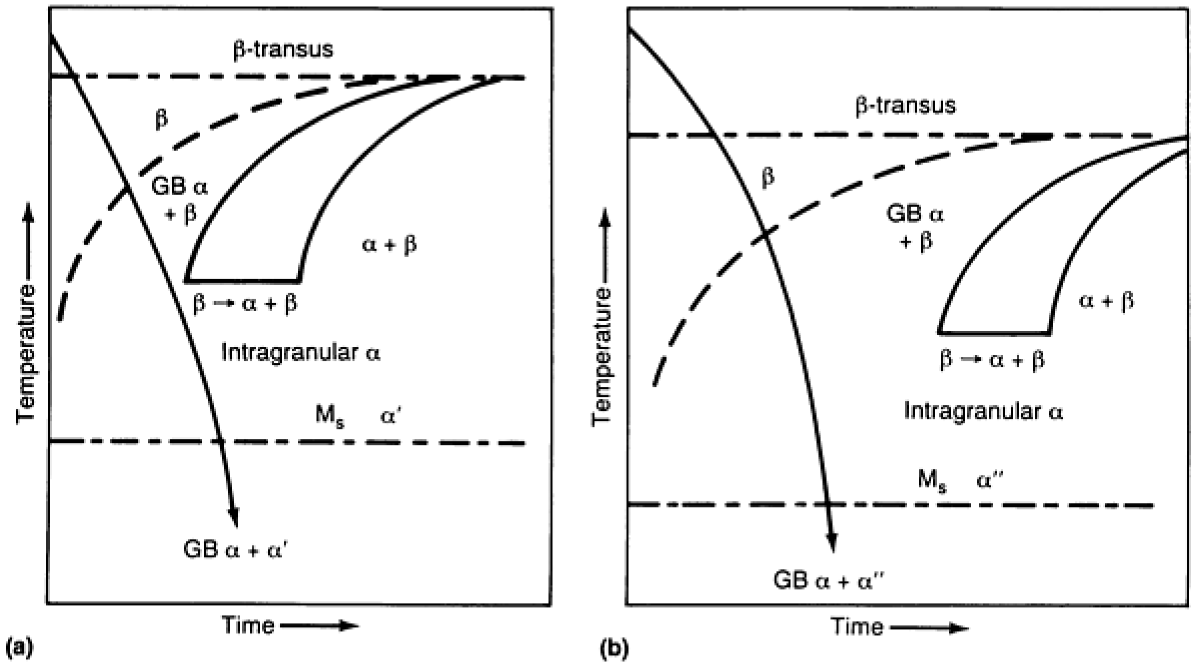

FIG. 38 SCHEMATIC CCT DIAGRAMS FOR TWO α + β TITANIUM ALLOYS WITH DIFFERENT β -STABILIZER SOLUTE CONTENTS. (A) LEAN α + β ALLOY (FOR EXAMPLE, TI-6AL-4V). (B) α + β ALLOY (FOR EXAMPLE, CORONA 5 [TI-4.5AL-5MO-1.5CR]) RICHER IN β -STABILIZING SOLUTE CONTENT

The morphology of the α and β phases in the α + β base materials is strongly dependent on thermomechanical processing (TMP) and heat treatment, which are performed either in the β-phase field or in the two-phase α + β phase field. When TMP is done in the α+ β phase field, the α that forms on cooling is continuously deformed. After recrystallization, a near-equiaxed α structure is obtained after this α + β processing. Very little GB α is formed because deformation introduces sufficient alternate heterogeneous nucleation sites in the form of dislocations. On the other hand, if TMP is done above the β-transus temperature (β-processing) and cooled, α forms at grain boundaries (giving GB α) and in the interior of grains as a Widmanstätten structure (also referred to as an acicular, lenticular, or basketweave structure). The microstructure in the weld and the near HAZ is closer to that obtained after β-processing.

Figure 39 shows the effect of α + β processing and β-processing on the strength and toughness of α + β alloys with increasing β-stabilizing alloy content. At the lower strength levels obtained with a lower β-stabilizing alloy content (because of a lower volume fraction of β-phase present), the finer Widmanstätten structure obtained after β-processing has a higher toughness than the coarser structure obtained by α + β processing. At higher strength levels, the trend reverses with the α + β processed structure having a higher toughness. This result can be understood in terms of the strength differences between GB α and the interior of the prior β grains (Fig. 40). At lower strengths, when the grain interiors are relatively weak, the crack tip plastic zone is distributed between the GB α and the grain interiors. However, as the α-phase precipitates are refined by lower aging temperatures, as in the case of the metastable β alloys, the grain interiors are strengthened considerably. Under such circumstances, the cracks are essentially constrained to follow the GB α layer. This results in a contraction of the plastic zone, a reduction in the crack tip opening displacement for crack extension, and an attendant drop in the toughness.

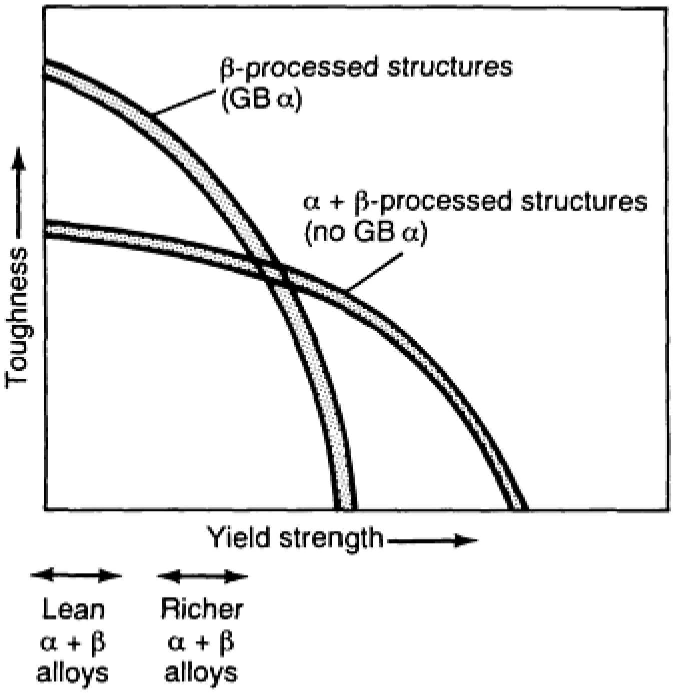

FIG. 39 SCHEMATIC PLOT OF TOUGHNESS VERSUS YIELD STRENGTH FOR α + β AND β-PROCESSED STRUCTURES. APPROXIMATE LOCATIONS OF LEAN (FOR EXAMPLE, TI-6AL-4V) AND RICHER β STABILIZED α + β ALLOYS (FOR EXAMPLE, CORONA 5 [TI-4.5AL-5MO-1.5CR]) ARE SHOWN. SOURCE: REF 52

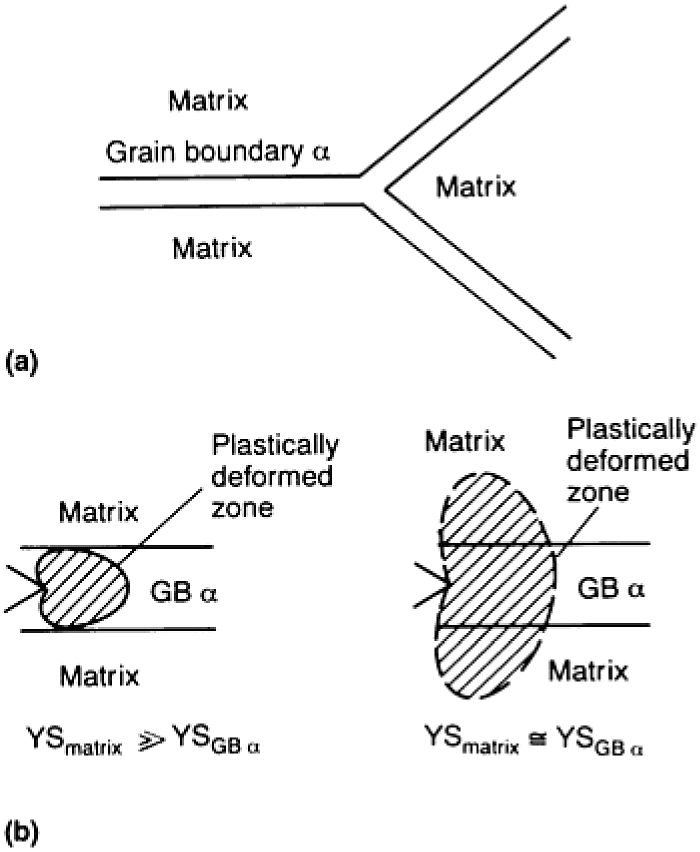
Matrix
Grain boundary α
Matrix
Matrix
(a)
Plastically deformed zone
Matrix
GB α
Matrix
YS_matrix ≫ YS_GB α
Matrix
Plastically deformed zone
GB α
Matrix
YS_matrix ≅ YS_GB α
(b)

FIG. 40 SCHEMATIC SHOWING THE EFFECT OF THE RELATIVE STRENGTHS OF GB α AND MATRIX ON THE

PLASTIC ZONE SIZE. (A) GB α LOCATION RELATIVE TO MATRIX. (B) EFFECT OF PLASTIC ZONE SIZE ON YIELD STRENGTHS OF MATRIX AND GB α. SOURCE: REF 52

The microstructures in the weld metal and the near HAZ (the region where the peak temperature exceeds the β-transus temperature) resemble those obtained after β-processing. This means that it is easier to get a matching toughness in the weld metal at the lower strength levels referred to in Fig. 39. Even this is made somewhat difficult by the coarser β grain size in the weld and HAZ. Grain coarsening in the HAZ induces coarse β grains to form in the weld also, because solidification in the weld metal occurs by epitaxial growth from the HAZ. The extent of grain coarsening is such that it is not uncommon to find a single columnar β grain traversing the entire thickness in gas-tungsten arc welds on thin sheet (Ref 53). If high energy density welding processes are used (for example, electron-beam welding or laser-beam welding), it is possible to limit the extent of β grain coarsening. However, the high cooling rates in the weld and HAZ will mean that a martensitic structure will be obtained.

To obtain adequate ductility and toughness, PWHT will have to be done at temperatures close to the β-transus temperature. This can be a problem because the low strength at these temperatures can cause sagging in large welded structures, and complex fixturing will be necessary to maintain dimensional tolerances. It can also be expensive, because inert gas shielding is required for almost the entire time of PWHT (it is assumed in the present discussion that adequate care is taken to maintain proper shielding during welding). Moreover, it can be difficult to increase the ductility in the weld metal and HAZ to a level equal to that in an α+ β processed base material (Fig. 41). The trends seen in Fig. 39 and 41 can be understood by noting that crack nucleation is more difficult in an equaixed structure (because the strain concentration effects are less) and that the plastic zone sizes are bigger (Table 1). This results in a higher tensile ductility. However, because crack propagation follows a less tortuous path, the toughness is lower. The reverse is true for a Widmanstätten structure (referred to as a lenticular structure in Table 1). Table 1 explains the apparently strange observation in titanium alloys that a structure having a high strength and toughness is not necessarily the one having a high tensile ductility. The implication of this for weldments is that for many applications it may be more important to optimize the welding and PWHT procedure with respect to the toughness than with respect to the ductility.

TABLE 1 EFFECT OF MORPHOLOGICAL FEATURES OF THE MICROSTRUCTURE ON CRACK NUCLEATION AND PROPAGATION IN α + β TITANIUM ALLOYS

Note: Voids initiated at the α -β interfaces are shown as black regions. Source: Ref 54

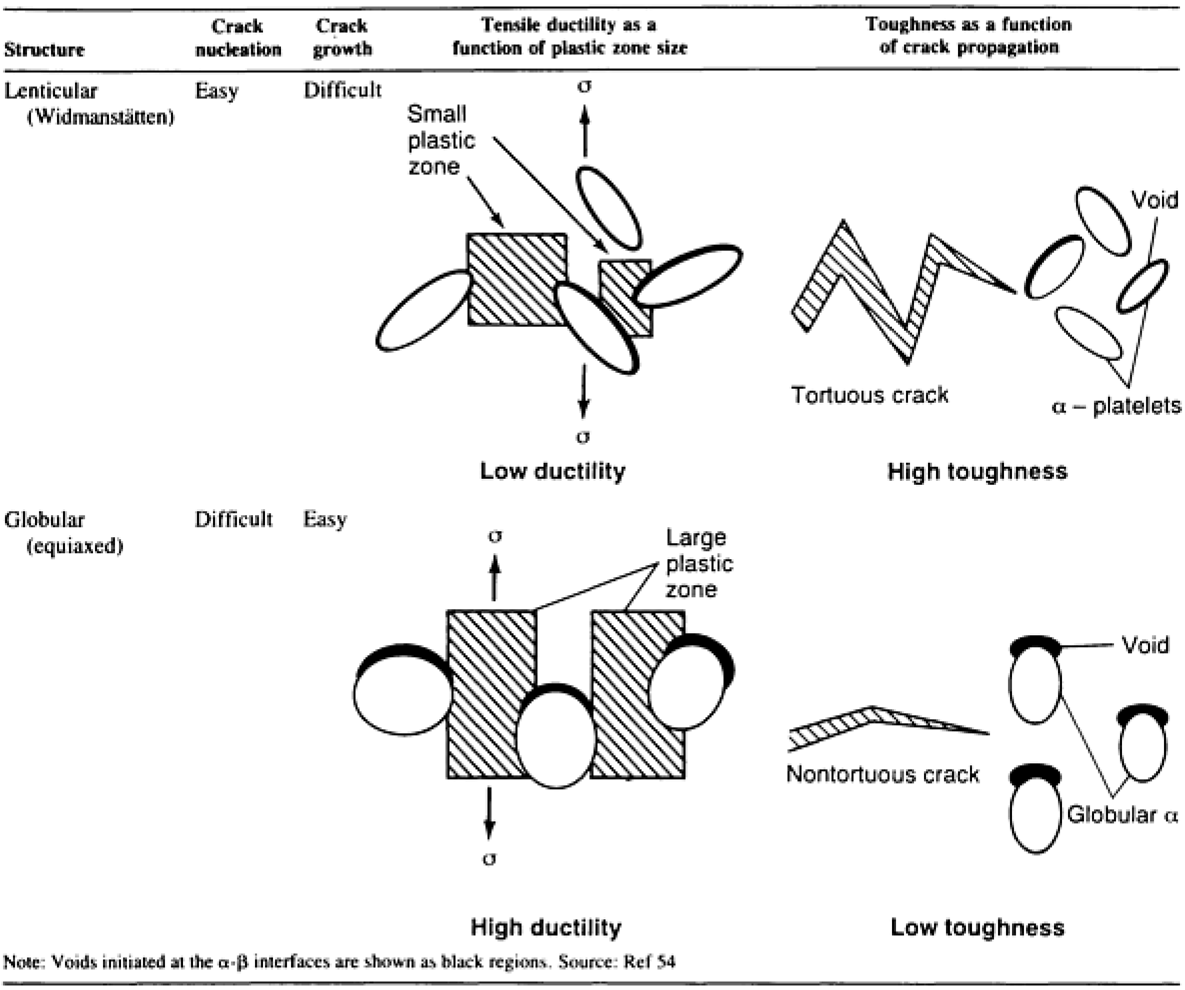

Structure	Crack nucleation	Crack growth	Tensile ductility as a function of plastic zone size	Toughness as a function of crack propagation
Lenticular (Widmanstätten)	Easy	Difficult	Low ductility	High toughness
Globular (equiaxed)	Difficult	Easy	High ductility	Low toughness

Note: Voids initiated at the α-β interfaces are shown as black regions. Source: Ref 54

σ
Small plastic zone
σ
Void
Tortuous crack
α – platelets
σ
Large plastic zone
σ
Void
Nontortuous crack
Globular α

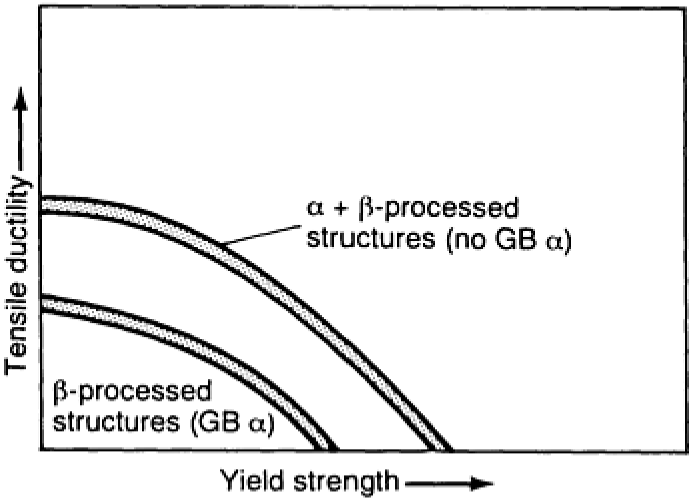

FIG. 41 SCHEMATIC PLOT OF DUCTILITY VERSUS STRENGTH FOR α + β AND β -PROCESSED STRUCTURES. SOURCE: REF 52

A more detailed discussion of the welding metallurgy of titanium alloys can be found in Ref 53, 55, and 56.

References cited in this section

50. E.W. COLLINGS, INTRODUCTION TO TITANIUM ALLOY DESIGN, *ALLOYING,* J.L. WALTER,

M.R. JACKSON, AND C.T. SIMS, ED., ASM INTERNATIONAL, 1988, P 267

51. C.R. THOMAS AND F.P.A. ROBINSON, KINETICS AND MECHANISM OF GRAIN GROWTH DURING WELDING IN NIOBIUM STABILISED 17% CHROMIUM STAINLESS STEELS, *MET. TECHNOL.,* APRIL 1978, P 133-138
52. J.C. WILLIAMS AND E.A. STARKE, JR., THE ROLE OF THERMOMECHANICAL PROCESSING IN TAILORING THE PROPERTIES OF ALUMINUM AND TITANIUM ALLOYS, *DEFORMATION, PROCESSING AND STRUCTURE,* G. KRAUSS, ED., AMERICAN SOCIETY FOR METALS, 1984, P 279-354
53. W.A. BAESLACK III, D.W. BECKER, AND F.H. FROES, ADVANCES IN TITANIUM ALLOY WELDING METALLURGY, *J. MET.,* MAY 1984, P 46-58
54. J.P. HIRTH AND F.H. FROES, INTERRELATIONS BETWEEN FRACTURE TOUGHNESS AND OTHER MECHANICAL PROPERTIES IN TITANIUM ALLOYS, *METALL. TRANS. A,* VOL 8, JULY 1977, P 1165-1176
55. C.G. RHODES, MICROSCOPY AND TITANIUM ALLOY DEVELOPMENT, *APPLIED METALLOGRAPHY,* G.F. VANDER VOORT, ED., VAN NOSTRAND REINHOLD, 1986
56. W.A. BAESLACK III, METALLOGRAPHY OF TITANIUM ALLOY WELDMENTS, *METALLOGRAPHY AND INTERPRETATION OF WELD MICROSTRUCTURES,* J.L. MCCALL, D.L. OLSON, AND I. LEMAY, ED., ASM INTERNATIONAL, 1987, P 23-60

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

Titanium Alloys

Titanium alloys find use in aerospace applications, pressure vessels, and so on because of their high strength-to-weight ratio, high corrosion resistance, and high fracture toughness. Pure titanium has two allotropic forms: low-temperature hexagonal close-packed (hcp) α-phase and elevated-temperature body-centered cubic (bcc) β-phase. Various alloying elements tend to preferentially stabilize one or the other of these phases. As a result, titanium alloys are generally classified as α, α+ β, metastable β, and βalloys (Fig. 37).

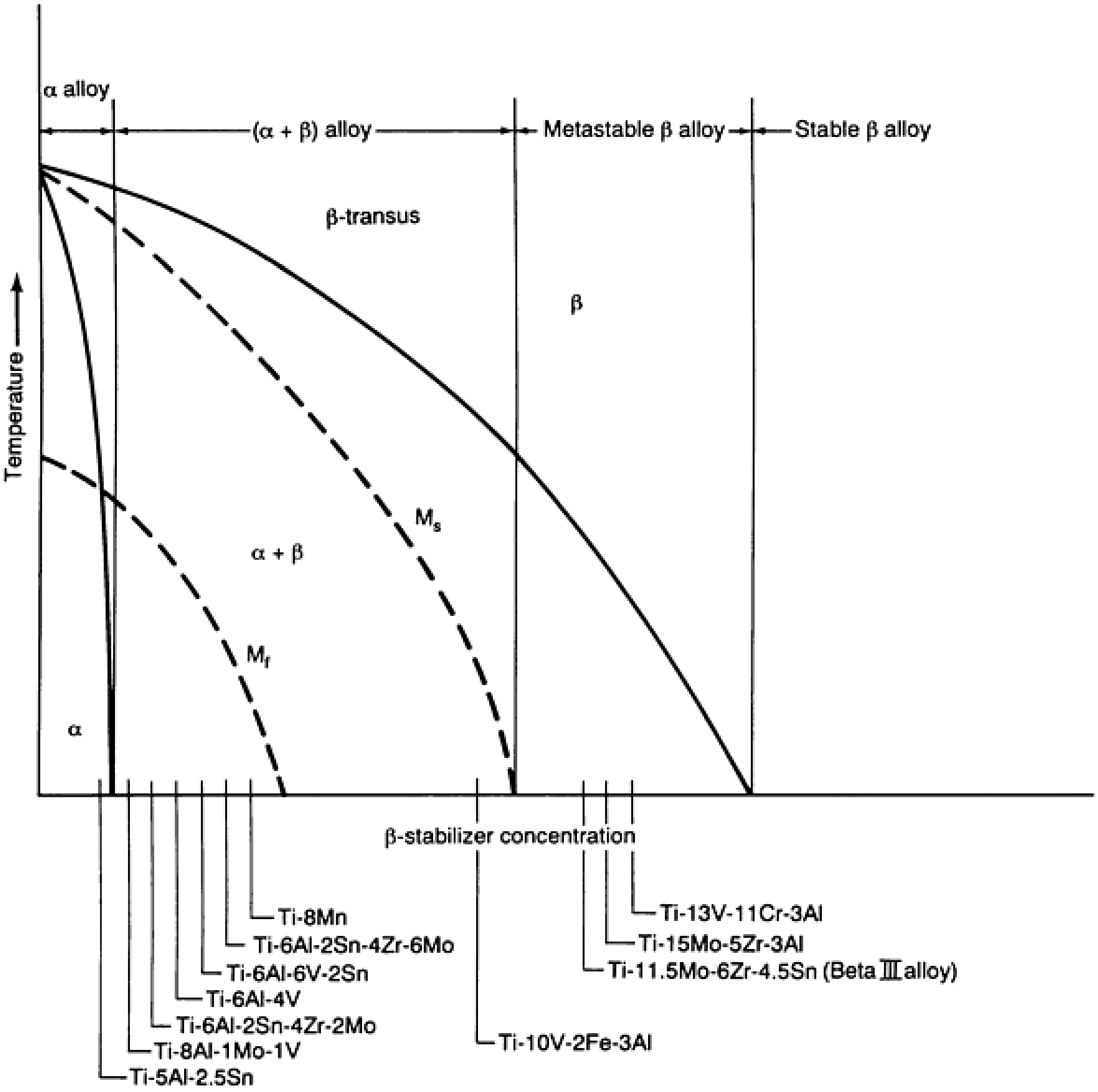
α alloy
(α + β) alloy
Metastable β alloy
Stable β alloy
β-transus
β
Temperature
M_s
α + β
M_f
α
β-stabilizer concentration
Ti-8Mn
Ti-6Al-2Sn-4Zr-6Mo
Ti-6Al-6V-2Sn
Ti-6Al-4V
Ti-6Al-2Sn-4Zr-2Mo
Ti-8Al-1Mo-1V
Ti-5Al-2.5Sn
Ti-10V-2Fe-3Al
Ti-13V-11Cr-3Al
Ti-15Mo-5Zr-3Al
Ti-11.5Mo-6Zr-4.5Sn (Beta III alloy)

FIG. 37 HYPOTHETICAL β ISOMORPHOUS PHASE DIAGRAM THAT SHOWS RELATION BETWEEN β -STABILIZER SOLUTE CONTENT AND TITANIUM ALLOY FAMILY DESCRIPTION. SELECTED COMMON ALLOYS ARE LISTED ALONG THE *X*-AXIS OF THE DIAGRAM BASED ON THEIR COMPOSITIONS. SOURCE: REF 50

The α-alloys are not heat-treatable (that is, a significant increase in strength and toughness cannot be obtained by heat treatment), and they have the lowest strength of the four titanium alloy families. Interestingly enough, these alloys are thus considered to have excellent weldability. What this means in practice is that there is no significant degradation in the properties of the weld and HAZ with regard to the parent material due to the weld thermal cycles. Also, the β ↔ α transformation in the weld and near HAZ means that the extent of grain coarsening will not be so high as in other single-phase materials in which no phase transformation occurs upon heating or cooling, such as stabilized ferritic stainless steels (Ref 51), pure aluminum, and so on.

In the α+ β alloys, a higher strength is obtained with a higher volume fraction of the β phase, which in turn is obtained with a higher β-stabilizing alloy content (Fig. 37). The effect of cooling rate on the phase transformations in two α + β alloys is shown schematically in Fig. 38. As in the case of steels, it is seen that the β ↔ α transformation can occur either by nucleation and growth at slower cooling rates or martensitically at higher cooling rates. The martensitic product in the lean α + β alloys has an hcp crystal structure (designated as α') or an orthorhombic one (designated as α ") in the β-stabilized α + β alloys. It is seen that the C-curve for the β ↔ intergranular α transformation is shifted to the right for alloys with increasing β-stabilizing alloy content, while that for the allotriomorphic α(more commonly referred to as GB

α) is relatively insensitive to alloy concentration (Ref 52). As a result, alloys that are fairly rich in β-stabilizing additions exhibit a stronger tendency to form a continuous network of GB α than do leaner α + β alloys.

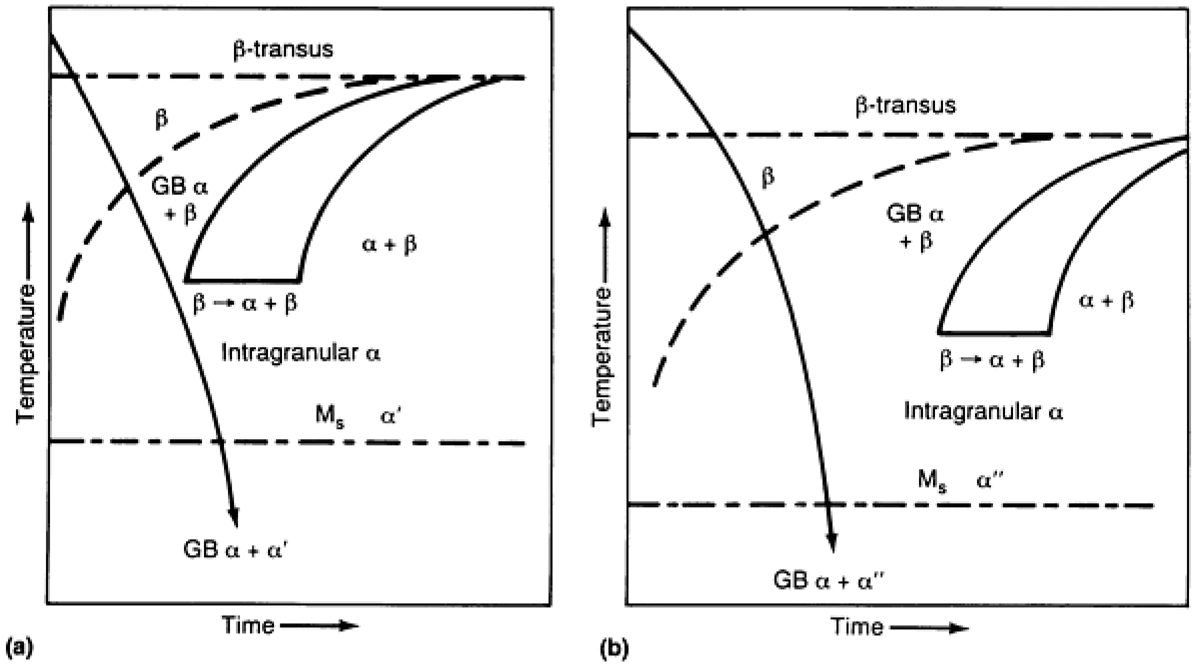

FIG. 38 SCHEMATIC CCT DIAGRAMS FOR TWO α + β TITANIUM ALLOYS WITH DIFFERENT β -STABILIZER SOLUTE CONTENTS. (A) LEAN α + β ALLOY (FOR EXAMPLE, TI-6AL-4V). (B) α + β ALLOY (FOR EXAMPLE, CORONA 5 [TI-4.5AL-5MO-1.5CR]) RICHER IN β -STABILIZING SOLUTE CONTENT

The morphology of the α and β phases in the α + β base materials is strongly dependent on thermomechanical processing (TMP) and heat treatment, which are performed either in the β-phase field or in the two-phase α+ β phase field. When TMP is done in the α+ β phase field, the α that forms on cooling is continuously deformed. After recrystallization, a near-equiaxed α structure is obtained after this α + β processing. Very little GB α is formed because deformation introduces sufficient alternate heterogeneous nucleation sites in the form of dislocations. On the other hand, if TMP is done above the β-transus temperature (β-processing) and cooled, α forms at grain boundaries (giving GB α) and in the interior of grains as a Widmanstätten structure (also referred to as an acicular, lenticular, or basketweave structure). The microstructure in the weld and the near HAZ is closer to that obtained after β-processing.

Figure 39 shows the effect of α + β processing and β-processing on the strength and toughness of α + β alloys with increasing β-stabilizing alloy content. At the lower strength levels obtained with a lower β-stabilizing alloy content (because of a lower volume fraction of β-phase present), the finer Widmanstätten structure obtained after β-processing has a higher toughness than the coarser structure obtained by α + β processing. At higher strength levels, the trend reverses with the α + β processed structure having a higher toughness. This result can be understood in terms of the strength differences between GB α and the interior of the prior β grains (Fig. 40). At lower strengths, when the grain interiors are relatively weak, the crack tip plastic zone is distributed between the GB α and the grain interiors. However, as the α-phase precipitates are refined by lower aging temperatures, as in the case of the metastable β alloys, the grain interiors are strengthened considerably. Under such circumstances, the cracks are essentially constrained to follow the GB α layer. This results in a contraction of the plastic zone, a reduction in the crack tip opening displacement for crack extension, and an attendant drop in the toughness.

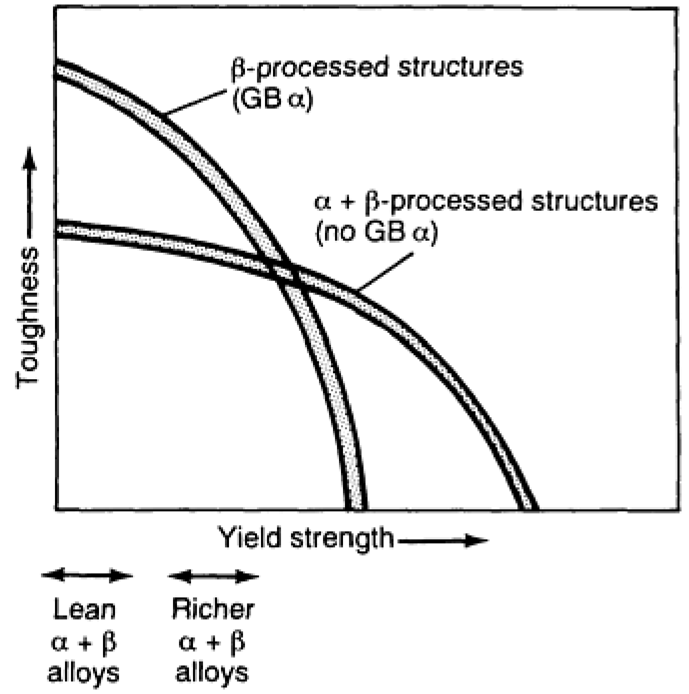

FIG. 39 SCHEMATIC PLOT OF TOUGHNESS VERSUS YIELD STRENGTH FOR α + β AND β -PROCESSED STRUCTURES. APPROXIMATE LOCATIONS OF LEAN (FOR EXAMPLE, TI-6AL-4V) AND RICHER β STABILIZED α + β ALLOYS (FOR EXAMPLE, CORONA 5 [TI-4.5AL-5MO-1.5CR]) ARE SHOWN. SOURCE: REF 52

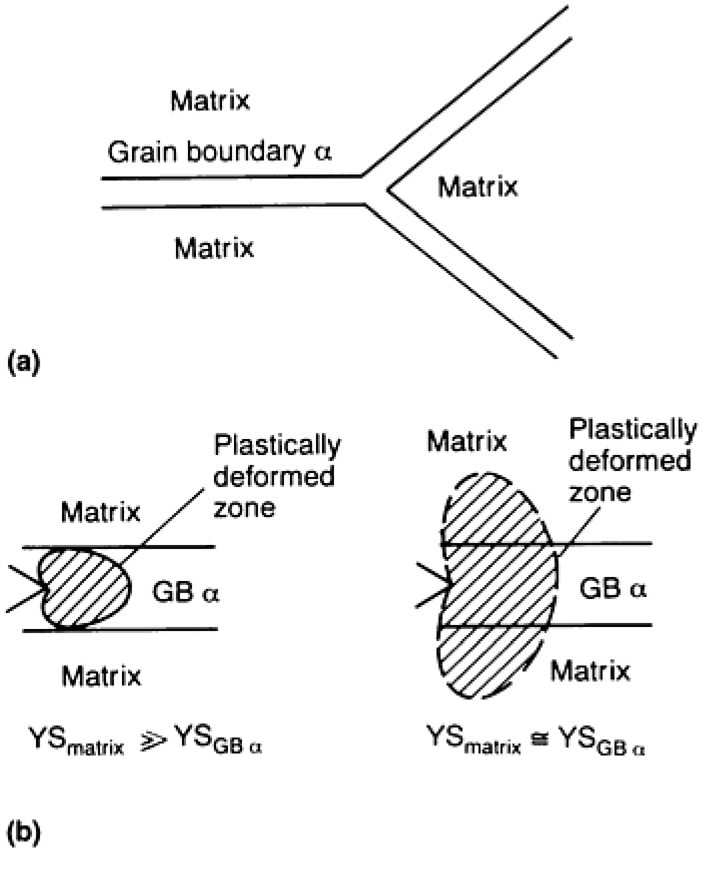

Matrix
Grain boundary α
Matrix
Matrix
(a)
Plastically
deformed
zone
Matrix
GB α
Matrix
$YS_{matrix} \gg YS_{GB\ \alpha}$
Matrix
Plastically
deformed
zone
GB α
Matrix
$YS_{matrix} \cong YS_{GB\ \alpha}$
(b)

FIG. 40 SCHEMATIC SHOWING THE EFFECT OF THE RELATIVE STRENGTHS OF GB α AND MATRIX ON THE PLASTIC ZONE SIZE. (A) GB α LOCATION RELATIVE TO MATRIX. (B) EFFECT OF PLASTIC ZONE SIZE ON YIELD

STRENGTHS OF MATRIX AND GB α. SOURCE: REF 52

The microstructures in the weld metal and the near HAZ (the region where the peak temperature exceeds the β-transus temperature) resemble those obtained after β-processing. This means that it is easier to get a matching toughness in the weld metal at the lower strength levels referred to in Fig. 39. Even this is made somewhat difficult by the coarser β grain size in the weld and HAZ. Grain coarsening in the HAZ induces coarse β grains to form in the weld also, because solidification in the weld metal occurs by epitaxial growth from the HAZ. The extent of grain coarsening is such that it is not uncommon to find a single columnar β grain traversing the entire thickness in gas-tungsten arc welds on thin sheet (Ref 53). If high energy density welding processes are used (for example, electron-beam welding or laser-beam welding), it is possible to limit the extent of β grain coarsening. However, the high cooling rates in the weld and HAZ will mean that a martensitic structure will be obtained.

To obtain adequate ductility and toughness, PWHT will have to be done at temperatures close to the β-transus temperature. This can be a problem because the low strength at these temperatures can cause sagging in large welded structures, and complex fixturing will be necessary to maintain dimensional tolerances. It can also be expensive, because inert gas shielding is required for almost the entire time of PWHT (it is assumed in the present discussion that adequate care is taken to maintain proper shielding during welding). Moreover, it can be difficult to increase the ductility in the weld metal and HAZ to a level equal to that in an α + β processed base material (Fig. 41). The trends seen in Fig. 39 and 41 can be understood by noting that crack nucleation is more difficult in an equaixed structure (because the strain concentration effects are less) and that the plastic zone sizes are bigger (Table 1). This results in a higher tensile ductility. However, because crack propagation follows a less tortuous path, the toughness is lower. The reverse is true for a Widmanstätten structure (referred to as a lenticular structure in Table 1). Table 1 explains the apparently strange observation in titanium alloys that a structure having a high strength and toughness is not necessarily the one having a high tensile ductility. The implication of this for weldments is that for many applications it may be more important to optimize the welding and PWHT procedure with respect to the toughness than with respect to the ductility.

TABLE 1 EFFECT OF MORPHOLOGICAL FEATURES OF THE MICROSTRUCTURE ON CRACK NUCLEATION AND PROPAGATION IN α+ β TITANIUM ALLOYS

Note: Voids initiated at the α-β interfaces are shown as black regions. Source: Ref 54

Structure	Crack nucleation	Crack growth	Tensile ductility as a function of plastic zone size	Toughness as a function of crack propagation
Lenticular (Widmanstätten)	Easy	Difficult	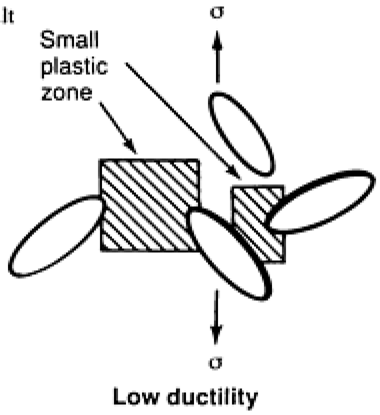 σ Small plastic zone σ Low ductility	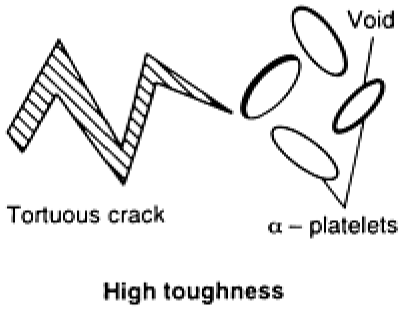 Void Tortuous crack α – platelets High toughness
Globular (equiaxed)	Difficult	Easy	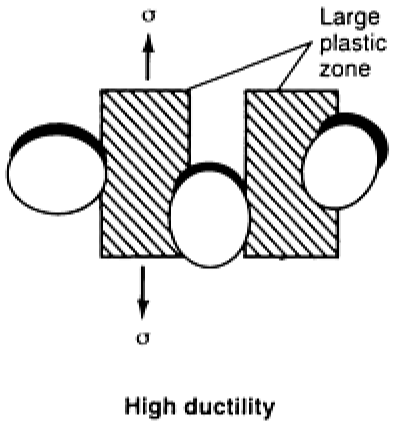 σ Large plastic zone σ High ductility	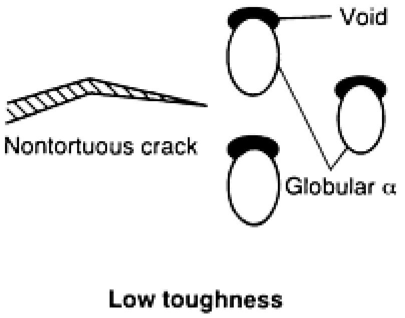 Void Nontortuous crack Globular α Low toughness

Note: Voids initiated at the α-β interfaces are shown as black regions. Source: Ref 54

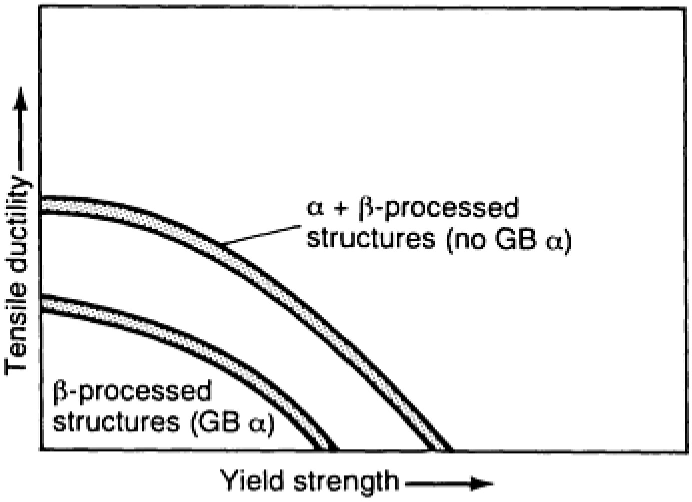

FIG. 41 SCHEMATIC PLOT OF DUCTILITY VERSUS STRENGTH FOR α+β AND β -PROCESSED STRUCTURES. SOURCE: REF 52

A more detailed discussion of the welding metallurgy of titanium alloys can be found in Ref 53, 55, and 56.

References cited in this section

50. E.W. COLLINGS, INTRODUCTION TO TITANIUM ALLOY DESIGN, *ALLOYING,* J.L. WALTER,

M.R. JACKSON, AND C.T. SIMS, ED., ASM INTERNATIONAL, 1988, P 267

51. C.R. THOMAS AND F.P.A. ROBINSON, KINETICS AND MECHANISM OF GRAIN GROWTH DURING WELDING IN NIOBIUM STABILISED 17% CHROMIUM STAINLESS STEELS, *MET. TECHNOL.,* APRIL 1978, P 133-138
52. J.C. WILLIAMS AND E.A. STARKE, JR., THE ROLE OF THERMOMECHANICAL PROCESSING IN TAILORING THE PROPERTIES OF ALUMINUM AND TITANIUM ALLOYS, *DEFORMATION, PROCESSING AND STRUCTURE,* G. KRAUSS, ED., AMERICAN SOCIETY FOR METALS, 1984, P 279-354
53. W.A. BAESLACK III, D.W. BECKER, AND F.H. FROES, ADVANCES IN TITANIUM ALLOY WELDING METALLURGY, *J. MET.,* MAY 1984, P 46-58
54. J.P. HIRTH AND F.H. FROES, INTERRELATIONS BETWEEN FRACTURE TOUGHNESS AND OTHER MECHANICAL PROPERTIES IN TITANIUM ALLOYS, *METALL. TRANS. A,* VOL 8, JULY 1977, P 1165-1176
55. C.G. RHODES, MICROSCOPY AND TITANIUM ALLOY DEVELOPMENT, *APPLIED METALLOGRAPHY,* G.F. VANDER VOORT, ED., VAN NOSTRAND REINHOLD, 1986
56. W.A. BAESLACK III, METALLOGRAPHY OF TITANIUM ALLOY WELDMENTS, *METALLOGRAPHY AND INTERPRETATION OF WELD MICROSTRUCTURES,* J.L. MCCALL, D.L. OLSON, AND I. LEMAY, ED., ASM INTERNATIONAL, 1987, P 23-60

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

References

1. *ATLAS OF ISOTHERMAL TRANSFORMATION AND COOLING TRANSFORMATION DIAGRAMS* AMERICAN SOCIETY FOR METALS, 1977
2. O.M. AKSELSEN AND T. SIMONSEN, TECHNIQUES FOR EXAMINING TRANSFORMATION BEHAVIOUR IN WELD METAL AND HAZ--A STATE OF ART REVIEW, *WELD. WORLD,* VOL 25 (NO. 1/2), 1987, P 26-34
3. O. GRONG AND D.K. MATLOCK, MICROSTRUCTURAL DEVELOPMENT IN MILD AND LOW ALLOY STEEL WELD METALS, *INT. MET. REV.,* VOL 31 (NO. 1), 1986, P 27-48
4. P.L. HARRISON AND R.A. FARRAR, APPLICATION OF CONTINUOUS COOLING TRANSFORMATION DIAGRAMS FOR WELDING OF STEELS, *INT. MATER. REV.,* VOL 34 (NO. 1), 1989, P 35-51
5. G.T. ELDIS, A CRITICAL REVIEW OF DATA SOURCES FOR ISOTHERMAL TRANSFORMATION AND CONTINUOUS COOLING TRANSFORMATION DIAGRAMS, *HARDENABILITY CONCEPTS WITH APPLICATION TO STEEL,* D.V. DOANE AND J.S. KIRKALDY, ED., TMS-AIME, 1978, P 126-157
6. K. EASTERLING, *INTRODUCTION TO THE PHYSICAL METALLURGY OF WELDING,* BUTTERWORTHS, 1983
7. *WELD THERMAL SIMULATORS FOR RESEARCH AND PROBLEM SOLVING,* THE WELDING INSTITUTE, U.K., 1972
8. C.F. BERKHOUT AND P.H. VAN LENT, THE USE OF PEAK TEMPERATURE-COOLING TIME DIAGRAMS IN THE WELDING OF HIGH STRENGTH STEELS (IN GERMAN), *SCHWEISS. SCHNEID.,* VOL 20 (NO. 6), 1968, P 256-260
9. TH.J. VAN ADRICHEM AND J. KAS, *CALCULATION, MEASUREMENT AND SIMULATION OF WELD THERMAL CYCLES,* SMITWELD N.V., NIJMEGEN, THE NETHERLANDS

10. P. RAVI VISHNU, W.B. LI, AND K. EASTERLING, HEAT FLOW MODEL FOR PULSED WELDING, *MATER. SCI. TECH.,* VOL 7 (NO. 7), 1991, P 649-659
11. P. RAVI VISHNU AND K. EASTERLING, PHENOMENOLOGICAL MODELLING OF HEAT FLOW AND MICROSTRUCTURAL CHANGES IN PULSED GTAW WELDS OF A QT STEEL, *MATHEMATICAL MODELING OF WELD PHENOMENA,* H. CERJAK AND K. EASTERLING, ED., THE INSTITUTE OF METALS, 1993
12. VEREIN DEUTSCHER EISENHÜTTENLEUTE, ED., *STEEL--A HANDBOOK FOR MATERIALS RESEARCH AND ENGINEERING, VOLUME 1: FUNDAMENTALS,* SPRINGER VERLAG, 1992, P 175
13. A.J. DEARDO, C.I. GARCIA, AND E.J. PALMIERE, THERMOMECHANICAL PROCESSING OF STEELS, *HEAT TREATING,* VOL 4, *ASM HANDBOOK,* 1991, P 237-255
14. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987
15. K. NARITA, PHYSICAL CHEMISTRY OF THE GROUPS IVA (TI, ZR), VA (V, NB, TA) AND THE RARE EARTH ELEMENTS IN STEEL, *TRANS. ISIJ,* VOL 15, 1975, P 145-152
16. R.K. AMIN AND F.B. PICKERING, AUSTENITE GRAIN COARSENING AND THE EFFECT OF THERMOMECHANICAL PROCESSING ON AUSTENITE RECRYSTALLIZATION, *THERMOMECHANICAL PROCESSING OF MICROALLOYED AUSTENITE,* A.J. DEARDO, G.A. RATZ, AND P.J. WRAY, ED., TMS-AIME, 1982, P 14
17. W.F. SAVAGE, E.F. NIPPES, AND E.S SZEKERES, A STUDY OF WELD INTERFACE PHENOMENA IN A LOW ALLOY STEEL, *WELD. J.,* SEPT 1976, P 260S-268S, AND HYDROGEN INDUCED COLD CRACKING IN A LOW ALLOY STEEL, *WELD. J.,* SEPT 1976, P 276S-283S
18. J.J. PEPE AND W.F. SAVAGE, EFFECTS OF CONSTITUTIONAL LIQUATION IN 18-NI MARAGING STEEL WELDMENTS, *WELD J.,* SEPT 1967, P 411S-422S
19. W. YENISCAVITCH, JOINING, *SUPERALLOYS II,* C.T. SIMS, N.S. STOLOFF, AND W.C. HAGEL, ED., JOHN WILEY & SONS, 1987, P 495-516
20. E.E. HUCKE AND H. UDIN, WELDING METALLURGY OF NODULAR CAST IRON, *WELD. J.,* AUG 1953, P 378S-385S
21. E.N. GREGORY AND S.B. JONES, WELDING CAST IRONS, *WELDING OF CASTINGS,* THE WELDING INSTITUTE, U.K., 1977, P 145-156
22. L.E. SAMUELS, *OPTICAL MICROSCOPY OF CARBON STEELS,* AMERICAN SOCIETY FOR METALS, 1980
23. "GUIDE TO THE LIGHT MICROSCOPE EXAMINATION OF FERRITIC STEEL WELD METALS," DOC. NO. IX-1533-88, INTERNATIONAL INSTITUTE OF WELDING, 1988
24. H.K.D.H. BHADESHIA AND L.E. SVENSSON, MODELLING THE EVOLUTION OF MICROSTRUCTURE OF STEEL WELD METAL, *MATHEMATICAL MODELLING OF WELD PHENOMENA,* H. CERJAK AND K. EASTERLING, ED., THE INSTITUTE OF MATERIALS, 1993
25. G.R. EDWARDS AND S. LIU, RECENT DEVELOPMENTS IN HSLA STEEL WELDING, *ADVANCES IN WELDING METALLURGY,* FIRST U.S.-JAPAN SYMPOSIUM, AMERICAN WELDING SOCIETY AND OTHERS, JUNE 1990, P 215-293
26. B. AHLBLOM, "OXYGEN AND ITS ROLE IN DETERMINING WELD METAL MICROSTRUCTURE AND TOUGHNESS--A STATE OF THE ART REVIEW," DOC. NO. IX-1322-84, INTERNATIONAL INSTITUTE OF WELDING, 1984
27. L.-E. SVENSSON AND B. GRETOFT, MICROSTRUCTURE AND IMPACT TOUGHNESS OF C-MN WELD METALS. *WELD. J.,* DEC 1990, P 454S-461S
28. P.W. HOLSBERG AND R.J. WONG, WELDING OF HSLA-100 STEEL FOR NAVAL APPLICATIONS, *WELDABILITY OF MATERIALS,* R.A. PATTERSON AND K.W. MAHIN, ED., ASM INTERNATIONAL, 1990
29. J.M.B. LOSZ AND K.D. CHALLENGER, MICROSTRUCTURE AND PROPERTIES OF A COPPER PRECIPITATION STRENGTHENED HSLA STEEL WELDMENT, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID AND J.M. VITEK, ED., ASM INTERNATIONAL, 1990
30. H. HOMMA, S. OHKITA, S. MATSUDA, AND K. YAMAMOTO, IMPROVEMENTS OF HAZ

TOUGHNESS IN HSLA STEEL BY INTRODUCING FINELY DISPERSED TI-OXIDE, *WELD. J.,* OCT 1987, P 301S-309S

31. N. YURIOKA, "MODERN HIGH STRENGTH STEEL IN JAPAN," FIFTH INTERNATIONAL SYMPOSIUM, JAPAN WELDING SOCIETY, APRIL 1990 (TOKYO)
32. YU.N. GOTALSKII, THE PROBLEM OF WELDING HIGH STRENGTH STEELS, *AUTOM. WELD.*, JUNE 1984, P 37-40
33. W.K.C. JONES AND P.J. ALBERRY, THE ROLE OF PHASE TRANSFORMATIONS IN THE DEVELOPMENT OF RESIDUAL STRESSES DURING THE WELDING OF SOME FAST REACTOR STEELS, *PROC. CONF. FERRITIC STEELS FOR FAST REACTOR STEAM GENERATORS,* BRITISH NUCLEAR ENERGY SOCIETY, 1977
34. *MCGRAW-HILL ENCYCLOPEDIA OF SCIENCE AND TECHNOLOGY,* VOL 9, P 685
35. A.M. MAKARA AND N.A. MOSENDZ, EFFECTS OF THE WELD METAL ON CRACKING IN HAZ, *WELD. RES. ABROAD,* NOV 1965, P 78-86
36. H. IKAWA, S. SHIN, M. INUI, Y. TAKEDA, AND A. NAKANO, "ON THE MARTENSITE-LIKE STRUCTURE AT WELD BOND AND THE MACROSCOPIC SEGREGATION IN WELD METAL IN THE WELDED DISSIMILAR METALS OF α-STEELS AND γ-STEELS," DOC. NO. IX-785-72, INTERNATIONAL INSTITUTE OF WELDING, 1972
37. T. HAZE AND S. AIHARA, "INFLUENCE OF TOUGHNESS AND SIZE OF LOCAL BRITTLE ZONES ON HAZ TOUGHNESS OF HSLA STEELS," SEVENTH INTERNATIONAL CONFERENCE ON OFFSHORE MECHANICS AND ARCTIC ENGINEERING, HOUSTON, 1988
38. C.P. ROYER, A USER'S PERSPECTIVE ON HEAT AFFECTED ZONE TOUGHNESS, *WELDING METALLURGY OF STRUCTURAL STEELS,* J.Y. KOO, ED., TMS-AIME, 1987
39. F. MATSUDA, Z. LI, P. BERNASOVSKY, K. ISHIHARA, AND H. OKADA, "AN INVESTIGATION OF THE BEHAVIOUR OF M-A CONSTITUENT IN SIMULATED HAZ OF HSLA STEELS," DOC. NO. IX-B-1591-90, INTERNATIONAL INSTITUTE OF WELDING, 1990
40. P.L. HARRISON AND P.H.M. HART, RELATIONSHIPS BETWEEN HAZ MICROSTRUCTURE AND CTOD TRANSITION BEHAVIOUR IN MULTIPASS C-MN STEEL WELDS, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID AND J.M. VITEK, ED., ASM INTERNATIONAL, 1990
41. H. ONOE, J. TANAKA, AND I. WATANABE, JAPANESE LNG TANKER CONSTRUCTED USING A NEW WELDING PROCESS AND IMPROVED AL-KILLED STEELS, *MET. CONSTR.,* JAN 1979, P 26-31
42. P.J. ALBERRY, SENSITIVITY ANALYSIS OF HALF-BEAD AND ALTERNATIVE GTAW TECHNIQUES, *WELD. J.,* NOV 1989, P 442S-451S
43. J.A. BROOKS AND A.W. THOMPSON, MICROSTRUCTURAL DEVELOPMENT AND SOLIDIFICATION CRACKING SUSCEPTIBILITY OF AUSTENITIC STAINLESS STEEL WELDS, *INT. MATER. REV.,* VOL 36 (NO. 1), 1991
44. T.A. SIEWERT, C.N. MCCOWAN, AND D.L. OLSON, FERRITE NUMBER PREDICTION TO 100 FN IN STAINLESS STEEL WELD METAL, *WELD. J.,* DEC 1988, P 289S-298S
45. D.J. KOTECKI AND T.A. SIEWERT, WRC-1992 CONSTITUTION DIAGRAM FOR STAINLESS STEEL WELD METALS: A MODIFICATION OF THE WRC-1988 DIAGRAM, *WELD. J.,* MAY 1992, P 171S-178S
46. T.G. GOOCH AND D.C. WILLINGHAM, *WELD DECAY IN AUSTENITIC STAINLESS STEELS,* THE WELDING INSTITUTE, UK, 1975
47. S. KOU, WELDING METALLURGY AND WELDABILITY OF HIGH STRENGTH ALUMINUM ALLOYS, *WELD. RES. COUNC. BULL.,* NO. 320, DEC 1986
48. H. SHERCLIFF AND M.F. ASHBY, A PROCESS MODEL FOR AGE HARDENING OF AL ALLOYS--I. MODEL AND II. APPLICATIONS OF THE MODEL, *ACTA METALL. MATER.*, VOL 38 (NO. 10), 1990, P 1789-1812
49. W.A. OWCZARSKI, PROCESS AND METALLURGICAL FACTORS IN JOINING SUPERALLOYS

AND OTHER HIGH SERVICE TEMPERATURE MATERIALS, *PHYSICAL METALLURGY OF JOINING,* R. KOSSOWSKY AND M.E. GLICKSMAN, ED., TMS-AIME, 1980

50. E.W. COLLINGS, INTRODUCTION TO TITANIUM ALLOY DESIGN, *ALLOYING,* J.L. WALTER, M.R. JACKSON, AND C.T. SIMS, ED., ASM INTERNATIONAL, 1988, P 267
51. C.R. THOMAS AND F.P.A. ROBINSON, KINETICS AND MECHANISM OF GRAIN GROWTH DURING WELDING IN NIOBIUM STABILISED 17% CHROMIUM STAINLESS STEELS, *MET. TECHNOL.,* APRIL 1978, P 133-138
52. J.C. WILLIAMS AND E.A. STARKE, JR., THE ROLE OF THERMOMECHANICAL PROCESSING IN TAILORING THE PROPERTIES OF ALUMINUM AND TITANIUM ALLOYS, *DEFORMATION, PROCESSING AND STRUCTURE,* G. KRAUSS, ED., AMERICAN SOCIETY FOR METALS, 1984, P 279-354
53. W.A. BAESLACK III, D.W. BECKER, AND F.H. FROES, ADVANCES IN TITANIUM ALLOY WELDING METALLURGY, *J. MET.,* MAY 1984, P 46-58
54. J.P. HIRTH AND F.H. FROES, INTERRELATIONS BETWEEN FRACTURE TOUGHNESS AND OTHER MECHANICAL PROPERTIES IN TITANIUM ALLOYS, *METALL. TRANS. A,* VOL 8, JULY 1977, P 1165-1176
55. C.G. RHODES, MICROSCOPY AND TITANIUM ALLOY DEVELOPMENT, *APPLIED METALLOGRAPHY,* G.F. VANDER VOORT, ED., VAN NOSTRAND REINHOLD, 1986
56. W.A. BAESLACK III, METALLOGRAPHY OF TITANIUM ALLOY WELDMENTS, *METALLOGRAPHY AND INTERPRETATION OF WELD MICROSTRUCTURES,* J.L. MCCALL, D.L. OLSON, AND I. LEMAY, ED., ASM INTERNATIONAL, 1987, P 23-60

Cracking Phenomena Associated With Welding

Michael J. Cieslak, Physical and Joining Metallurgy Department, Sandia National Laboratories

Introduction

THE FORMATION OF DEFECTS in materials that have been fusion welded is a major concern in the design of welded assemblies. Four types of defects in particular have been the focus of much attention because of the magnitude of their impact on product quality. These defects, all of which manifest themselves as cracks, are characteristic of phenomena that occur at certain temperature intervals specific to a given alloy. Colloquially, these four defect types are known as hot cracks, heat-affected zone (HAZ) microfissures, cold cracks, and lamellar tearing.

Acknowledgements

This work was performed at Sandia National Laboratories, supported by the U.S. Department of Energy under Contract No. DE-AC04-76DP00789.

Michael J. Cieslak, Physical and Joining Metallurgy Department, Sandia National Laboratories

Solidification Cracking (Hot Cracking)

Hot cracks are solidification cracks that occur in the fusion zone near the end of solidification. Simplistically, they result from the inability of the semisolid material to accommodate the thermal shrinkage strains associated with weld solidification and cooling. Cracks then form at susceptible sites to relieves the accumulating strain. Susceptible sites are interfaces, such as solidification grain boundaries and interdendritic regions, that are at least partially wetted.

Figure 1 is a photomicrograph of a metallographic section showing the appearance of a typical solidification crack in a fusion weld. Note the intergranular nature of the crack path. Figure 2 shows the surface morphology of a solidification crack, reflecting the nearly solidified nature of the interface at the time of failure. Note the smooth dendritic appearance of this surface.

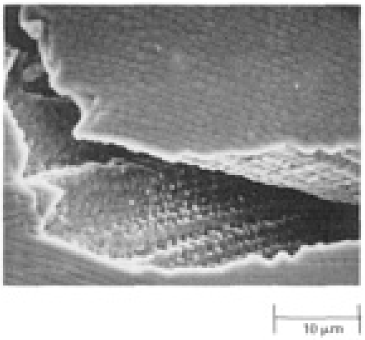

FIG. 1 SOLIDIFICATION CRACK IN A PULSED ND:YAG LASER WELD JOINING HASTELLOY C-276 TO 17-4 PH STAINLESS STEEL. YAG, YTTRIUM-ALUMINUM-GARNET

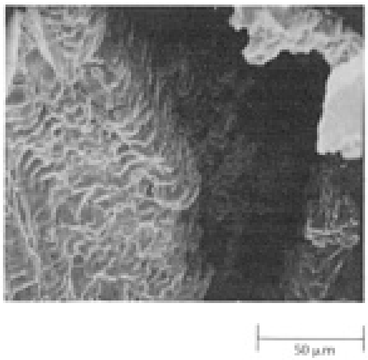

FIG. 2 SURFACE OF A SOLIDIFICATION CRACK IN AN ALLOY 214 VARESTRAINT TEST SPECIMEN

Solidification cracking requires both a sufficient amount of mechanical restraint (strain) and a susceptible microstructure. Knowledge regarding the development of shrinkage strains at the microstructural scale of a solidifying dendrite is still quite limited, making it nearly impossible to quantify *a priori* the amount of local strain required to initiate solidification cracks. Furthermore, the rate at which the strain is applied can influence the ability of an alloy to survive the solidification process without cracking. At temperatures near their melting points, metal alloys can readily dissipate applied loads by time-dependent deformation processes (i.e., creep processes). At high rates of applied stress, both the fusion zone and the HAZ have limited time to accommodate this mechanical imposition. Under conditions of rapid solidification and cooling, the rate of strain accumulation is rapid, leading to an increased cracking susceptibility. Inherently, then, requisite strains for solidification cracking are more likely to be experienced with welding processes that promote rapid solidification and cooling.

Because of the limited quantitative understanding of strain development in a solidifying weld, the practical approach taken to minimize the mechanical factor is to reduce the overall weld restraint through judicious joint design and appropriate choice of welding parameters. A simple way to minimize the restraint on a solidifying weld joint is to keep the joint gap to a minimum by designing hardware with good fit-up. Another approach, particularly attractive for small parts, is to design the weld joint as a standing edge. Welding parameters can have a profound influence on the occurrence of solidification cracking. The natural tendency to use high-speed welding to improve productivity can have detrimental effects. Formation of a teardrop-shaped weld pool, which may occur as the weld travel speed increases, can result in centerline solidification cracks. The solidification pattern associated with this type of weld pool is such that solidifying grains meet at the weld centerline, forming a particularly susceptible site for solidification crack initiation.

The manner in which a fusion weld is terminated can also influence cracking behavior. In a closure weld, simply extinguishing the power can result in the formation of "crater cracks" as the final weld pool rapidly solidifies in on itself. A common method used to minimize this problem is to ramp down the power slowly, either while the workpiece remains stationary or as it continues to travel a short length under the heat source. In some cases, the use of a runoff tab as a site for weld termination may be appropriate.

The extent to which special joint designs or precisely controlled welding parameters can be used to produce welds without solidification cracks is specific for a given alloy. Alloys that are microstructurally susceptible to cracking will necessarily have a more limited set of joint designs and processing parameters around which defect-free welds can be produced.

Alloys with a wide solidification temperature range are more susceptible to solidification cracking than alloys that solidify over a narrow temperature range. Simply described, this is because accumulated thermal strain is proportional to the temperature range over which a material solidifies. This temperature range is determined primarily by chemical composition. For many binary alloy systems and a few ternary alloy systems, the equilibrium solidification (or melting) temperature range has been established in phase diagram studies. For the vast majority of commercial alloys, however, these data are not precisely known.

Borland (Ref 1, 2) attempted to establish a general quantitative method for comparing the effects of various alloying elements on the melting temperature range of a given solvent element. Figure 3 shows a schematic binary alloy phase diagram having linear liquidus and solidus boundaries. For an alloy of composition C_0, T_L is the liquidus composition and T_S is the solidus composition, where the alloy would be completely solidified under conditions of equilibrium cooling. T_M is the invariant melting temperature of the solvent element. The slope of the liquidus is m_L, and the slope of the solidus is m_S. The equilibrium distribution coefficient, k, is defined as the ratio of the composition of the solid, C_S, to the composition of the liquid, C_L, at a given temperature (i.e., $k = C_S/C_L|_T$). For a linear liquidus and solidus, this value is constant and also equal to m_L/m_S.

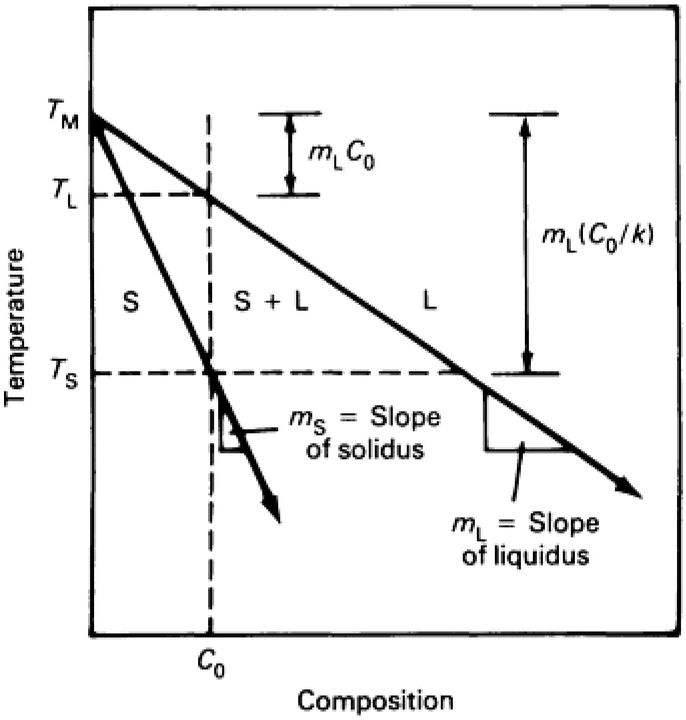

FIG. 3 HYPOTHETICAL BINARY ALLOY PHASE DIAGRAM SHOWING KEY PARAMETERS REQUIRED TO DERIVE THE RELATIVE POTENCY FACTOR (RPF)

With these terms thus defined, a derivation of the relative potency factor (RPF), after Borland (Ref 2), follows:

$$T_L = T_M - M_L C_0 \qquad \textbf{(EQ 1)}$$

$$T_S = T_M - M_S C_0 \qquad \textbf{(EQ 2)}$$

However, $m_S = m_L/k$:

$$T_s = T_M - \frac{M_L}{K} C_o \qquad \textbf{(EQ 3)}$$

Then:

$$T_L - T_S = \frac{m_L C_o (K-1)}{K} \qquad \textbf{(EQ 4)}$$

The RPF, always a positive number, is defined as the change in melting temperature ranger per weight percent of solute added, or:

$$RPF = \frac{T1 - Ts}{C_o} = \frac{M_L (k-1)}{K} \qquad \textbf{(EQ 5)}$$

where for $k < 1$, $m_L < 0$; and for $k > 1$,$m_L > 0$. The larger the RPF, the wider the melting temperature range and the higher the solidification cracking susceptibility. The final form of Eq 5 indicates that as k becomes very small (indicating a high level of microsegregation), the solidification cracking sensitivity increases. Similarly, as the magnitude of m_L increases (that is, becomes more highly negative), indicating a rapid decrease in liquidus with alloying concentration, the cracking propensity also increases.

Table 1 lists RPF values for various alloying elements in binary iron (Ref 2), nickel-base alloys (Ref 3), and aluminum alloys. It is clear that certain alloying elements, such as sulfur, phosphorus, and boron, are detrimental to the weldability of these alloy systems. This has been borne out by field experience.

TABLE 1 RELATIVE POTENCY FACTORS FOR SELECTED ALLOYING ELEMENTS IN BINARY IRON, NICKEL, AND ALUMINUM SYSTEMS

SYSTEM	**RPF FOR:**										
	Al	**B**	**C**	**Co**	**Mn**	**Nb**	**Ni**	**P**	**S**	**Si**	**Ti**
BINARY IRON	1.5	. . .	322	0.3	26.2	28.8	2.9	121	925	1.8	13.8
BINARY NICKEL	1.7	22700	330	. . .	0.7	. . .	. . .	556	244000	11	5.4

	Cu	**Li**	**Mg**	**Si**	**Zn**
BINARY ALUMINUM	16.6	9.0	8.0	42.4	2.3

Source: Ref 2, 3

Although the exact derivation of the RPF is specific to a very simple binary system solidifying under equilibrium conditions, the methodology provides general insight that can be applied to commercial alloys. The association of cracking susceptibility with solidification temperature range has implications for alloy design and primary melt processing. An extended melting/solidification temperature range can be the result of either intentional alloying or impurities present after refining.

An example (Ref 4) of the application of Borland's concepts to the solidification cracking propensity of a commercial alloy system involves alloy 625. This alloy is derived from the ternary Ni-Cr-Mo system, a very common base system for alloys designed to survive aggressive corrosive or high-temperature environments. A factorially designed set of alloys was produced to identify the effects of the alloying elements niobium, silicon, and carbon on the temperature range of two-phase coexistence of liquid and solid (equilibrium melting/solidification temperature range) and on solidification cracking susceptibility.

Solidification cracking susceptibility was measured with the varestraint test, using the maximum crack length (MCL) parameter as the quantitative measurement (Ref 5). The melting temperature ranges of these same alloys were measured using differential thermal analysis (DTA). Table 2 lists the compositions of the alloys examined.

TABLE 2 SELECTED ALLOY 625 COMPOSITIONS USED IN WELDABILITY STUDY

ALLOY NO.	COMPOSITION, %										
	C	Cr	Fe	Mn	Mo	Nb	Ni	P	S	Si	Ti
NIOBIUM-FREE											
1	0.006	22.10	2.56	0.02	9.54	0.01	BAL	0.005	0.002	0.03	0.06
2	0.031	21.95	2.55	0.02	9.61	0.01	BAL	0.005	0.002	0.03	0.06
3	0.006	21.63	2.18	0.02	9.60	0.02	BAL	0.005	0.002	0.35	0.06
4	0.036	21.57	2.59	0.03	9.63	0.02	BAL	0.005	0.003	0.39	0.06
NIOBIUM ALLOYED											
5	0.009	21.81	2.30	0.03	9.81	3.61	BAL	0.006	0.003	0.03	0.06
6	0.038	21.83	2.31	0.03	9.81	3.60	BAL	0.006	0.003	0.03	0.06
7	0.008	21.65	2.26	0.03	9.68	3.57	BAL	0.006	0.004	0.38	0.06
8	0.035	21.68	2.29	0.03	9.67	3.53	BAL	0.006	0.003	0.46	0.06

The results of the varestraint testing are shown in Fig. 4. These data clearly show that the alloys that contain intentionally alloyed niobium (alloys 5 to 8) are more susceptible to solidification cracking (larger MCL) than niobium-free alloys (1 to 4). Furthermore, the alloy with no intentionally alloyed carbon, silicon, or niobium (alloy 1) is the least susceptible to cracking.

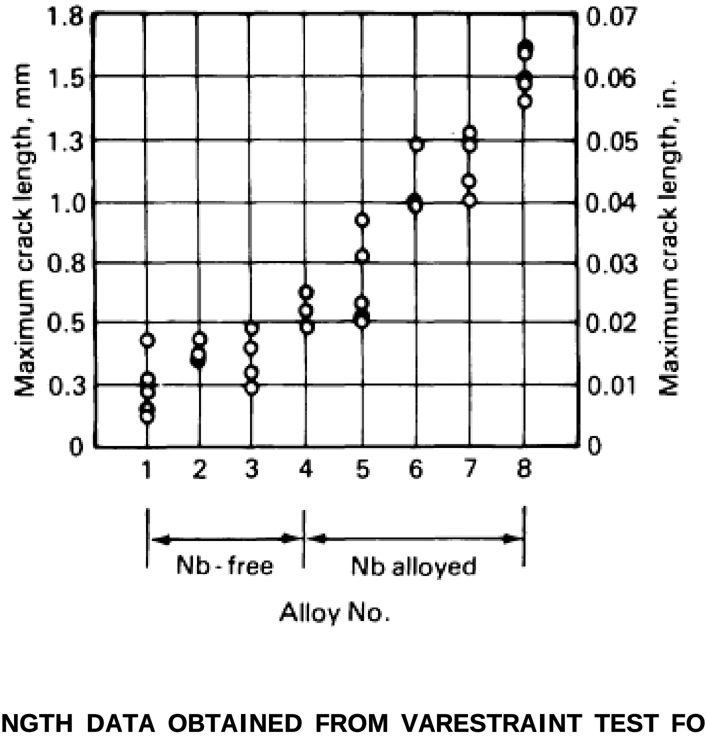

FIG. 4 MAXIMUM CRACK LENGTH DATA OBTAINED FROM VARESTRAINT TEST FOR ALLOY 625 WELDABILITY STUDY

The results of the DTA testing are given in Table 3. Listed are the liquidus (T_L), solidus (T_S), and melting temperature range (ΔT) data obtained from multiple tests. The baseline ΔT of an alloy without intentionally alloyed carbon, silicon, and niobium (alloy 1) is 26.5 °C (79.7 °F). Increases in Δ*T* are observed when carbon or silicon is added (alloys 2 and 3) at levels that would be considered high for commercial alloys. When niobium is added (alloy 5) at a level that would be considered nominal for alloy 625 (~3.6 wt%), the melting temperature range effectively doubles to 55.3 °C (131.5 °F). Additions of carbon and silicon to the high-niobium alloys (alloys 6 to 8) further increase this value.

TABLE 3 MELTING TEMPERATURE DATA FOR ALLOY 625 WELDABILITY STUDY

σ is the standard deviation.

ALLOY NO.	**ALLOY ADDITION**			**LIQUIDUS, $T_L \pm \sigma T_L$**				**SOLIDUS, $T_S \pm \sigma T_S$**				**MELTING TEMPERATURE RANGE, $\Delta T \pm \sigma \Delta T$**			
				T_L		**σT_L**		**T_S**		**σT_S**		**ΔT**		**$\sigma \Delta T$**	
	C	**NB**	**SI**	**°C**	**°F**	**°C**	**°F**	**°C**	**°F**	**°C**	**°F**	**°C**	**°F**	**°C**	**°F**
1				1406.5	2563.7	1.3	2.3	1380.0	2516.0	1.4	2.5	26.5	47.7	0.6	1.1
2	X			1403.0	2557.4	1.0	1.8	1369.3	249.7	2.5	4.5	33.7	60.7	1.5	2.7
3			X	1395.5	2543.9	1.3	2.3	1366.5	2491.7	1.3	2.3	29.0	52.2	1.6	2.9
4	X		X	1390.8	2535.4	0.5	0.9	1347.5	2457.5	1.7	3.1	43.3	77.9	1.3	2.3
5		X		1363.3	2485.9	0.5	0.9	1308.0	2386.4	2.2	4.0	55.3	99.5	2.4	4.3
6	X	X		1362.0	2483.6	0.8	1.4	1289.3	2352.7	2.6	4.7	72.8	131.0	3.2	5.8
7		X	X	1356.0	2472.8	0.8	1.4	1287.8	2350.0	2.4	4.3	68.3	122.9	2.5	4.5
8	X	X	X	1352.0	2465.6	0.0	0.0	1275.7	2328.3	0.6	1.1	76.3	137.3	0.6	1.1

The correlation of cracking susceptibility with melting temperature range is shown in Fig. 5. Linear regression analysis performed on the data set resulted in the following functional relationship:

$$\mathrm{MCL} = 0.00087(\Delta T) - 0.015 \qquad \textbf{(EQ 6)}$$

The incomplete statistical correlation ($R^2 = 0.86$) suggests that other factors are necessarily involved.

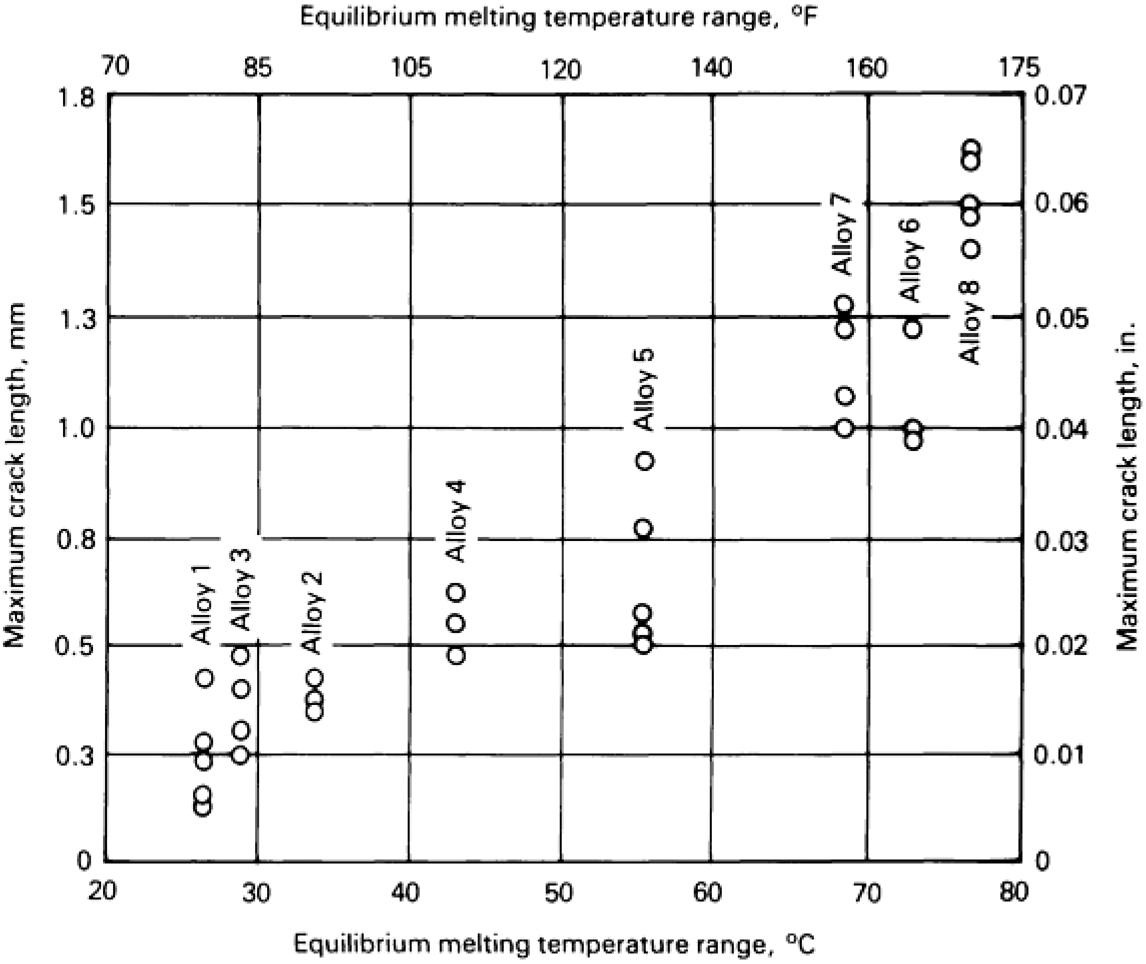

FIG. 5 RELATIONSHIP BETWEEN MAXIMUM CRACK LENGTH AND EQUILIBRIUM MELTING TEMPERATURE RANGE IN ALLOY 625 WELDABILITY STUDY

The solidification behavior of fusion welds is nonequilibrium in nature and may result in an extension of the solidification temperature range significantly beyond that measured under "equilibrium" conditions. Indeed, the niobium-bearing alloys terminate weld solidification below the equilibrium solidus with the formation of nonequilibrium eutecticlike constituents involving niobium carbides and Laves-type phases. Nonetheless, it is clear from an engineering perspective that Borland's concepts are valid enough to allow the establishment of a useful alloy ranking system from a solidification cracking perspective.

In addition to the effect of melting temperature range. Borland realized that the distribution of liquid along solidifying grain boundaries plays an important role in determining the solidification cracking susceptibility on an alloy (Ref 1). If the terminal solidification liquid easily wets the solidifying grain boundary (that is, has a low solid/liquid surface tension), it will tend to be spread out over a wider area, easing separation of the grains under a mechanical imposition. Conversely, if the liquid does not readily wet these boundaries, it will be more localized, creating a smaller area of liquid-solid contact and more solid-solid bridging, resulting in a more robust mechanical structure with increased cracking resistance. Elements such as sulfur and boron have been suggested as providing this surfactant property to terminal solidification liquids in iron and nickel alloys. Manganese may serve to increase the surface tension of terminal solidification fluids.

Figure 6 shows the results (Ref 6) of varestraint testing on two heats of Cabot alloy 214, a Ni-16Cr-4.5Al-2.5Fe alloy designed for use in high-temperature environments. The two alloys differ primarily in their boron content. With no boron intentionally added (0.0002 wt%), the alloy is extremely resistant to solidification cracking. With boron intentionally added at a level of 0.003 wt%, the solidification cracking propensity is much greater. Figures 7 and 8 are surface images of pulsed Nd:YAG laser welds made on the two alloys. Extensive cracking is evident in the high-boron alloy. Figure 7(b) is a good example of the extensive terminal "crater cracking" discussed above.

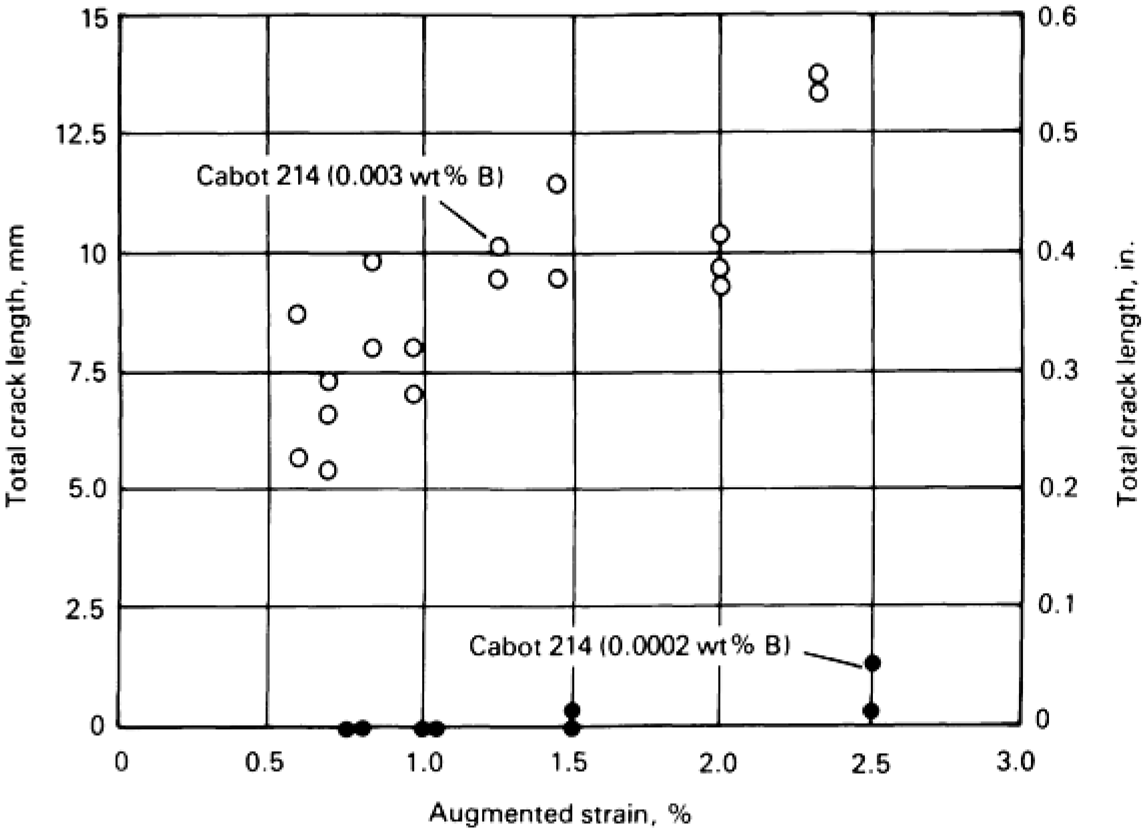

FIG. 6 VARESTRAINT TEST WELDABILITY DATA FOR CABOT ALLOY 214 WITH VARYING BORON CONCENTRATIONS

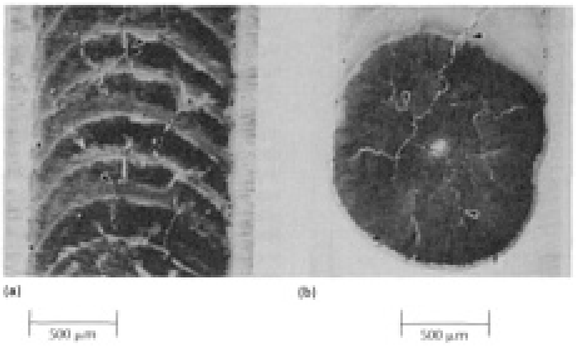

(a)
(b)
500 µm
500 µm

FIG. 7 SEM SURFACE IMAGE OF PULSED ND:YAG LASER WELDS IN HIGH-BORON ALLOY 214. (A) SOLIDIFICATION CRACKING. (B) CRATER CRACKING

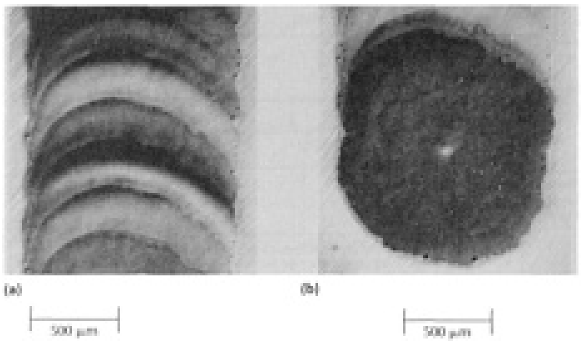

(a)
(b)
500 µm
500 µm

FIG. 8 SEM SURFACE IMAGE OF PULSED ND:YAG LASER WELDS IN LOW-BORON CONTENT ALLOY 214. (A) SOLIDIFICATION CRACKING. (B) CRATER CRACKING

Figures 9 and 10 summarize parameters that affect hot cracking in the weld metal and in the base metal HAZ, respectively. For most alloys that are highly susceptible to solidification cracking, filler metals of different composition have been developed to minimize the cracking that accompanies fusion welding.

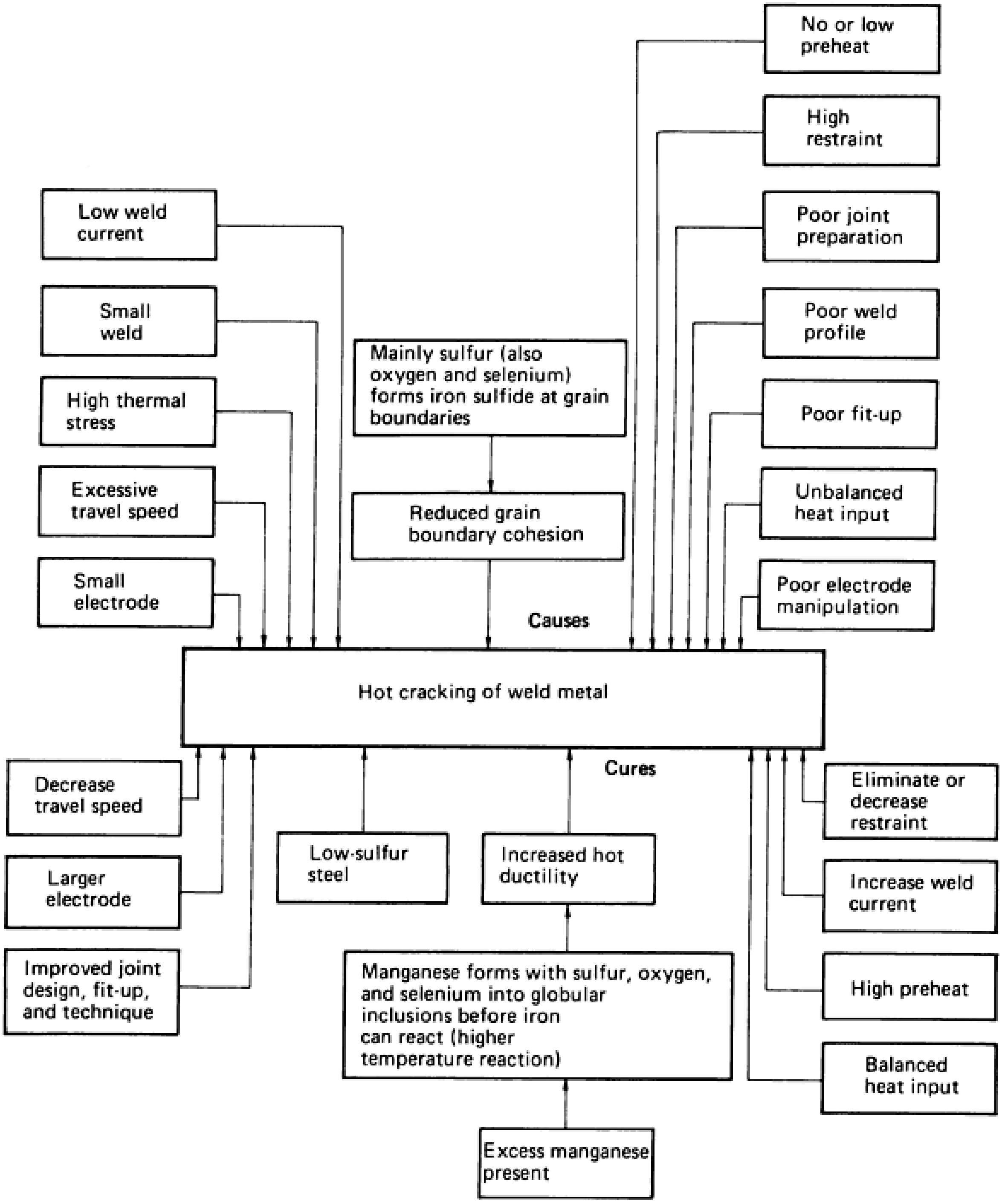
No or low preheat
High restraint
Low weld current
Poor joint preparation
Small weld
Poor weld profile
Mainly sulfur (also oxygen and selenium) forms iron sulfide at grain boundaries
High thermal stress
Poor fit-up
Excessive travel speed
Reduced grain boundary cohesion
Unbalanced heat input
Small electrode
Poor electrode manipulation
Causes
Hot cracking of weld metal
Cures
Decrease travel speed
Eliminate or decrease restraint
Low-sulfur steel
Increased hot ductility
Larger electrode
Increase weld current
Improved joint design, fit-up, and technique
Manganese forms with sulfur, oxygen, and selenium into globular inclusions before iron can react (higher temperature reaction)
High preheat
Balanced heat input
Excess manganese present

FIG. 9 FACTORS AFFECTING HOT CRACKING IN WELD METAL

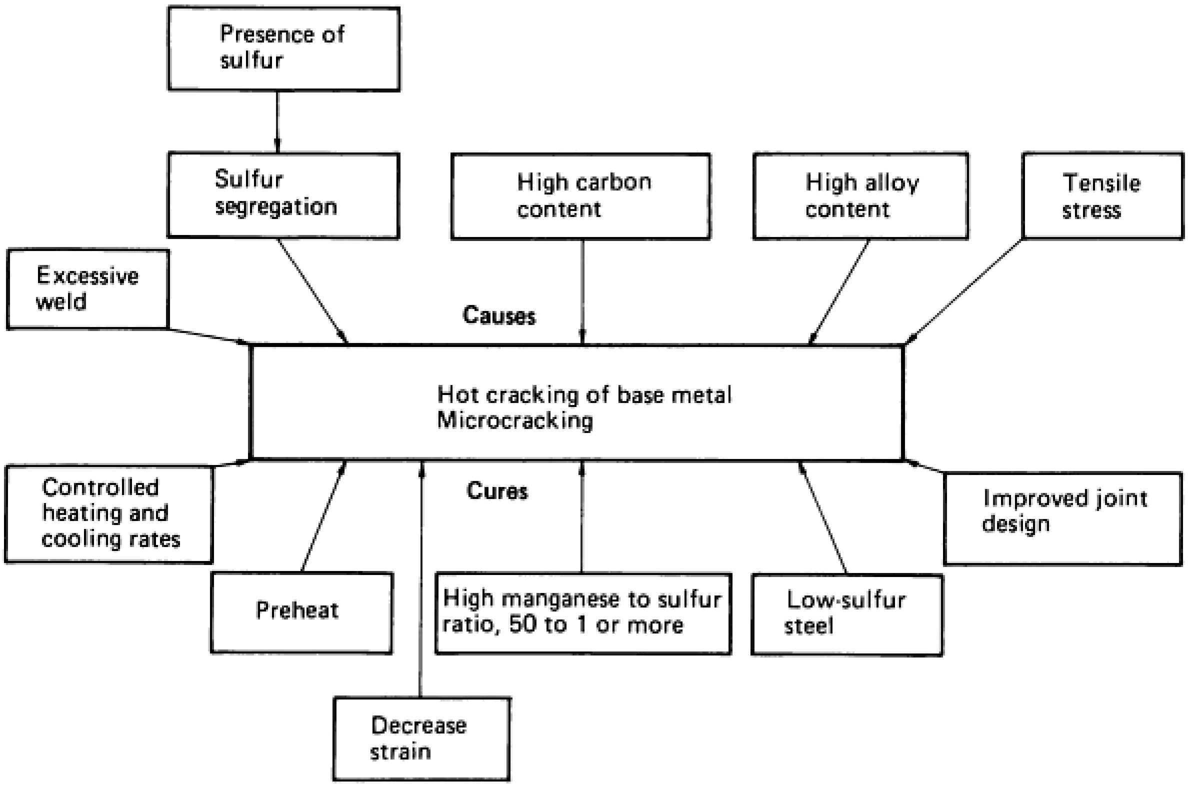
Presence of sulfur
Sulfur segregation
High carbon content
High alloy content
Tensile stress
Excessive weld
Causes
Hot cracking of base metal Microcracking
Cures
Controlled heating and cooling rates
Preheat
High manganese to sulfur ratio, 50 to 1 or more
Low-sulfur steel
Improved joint design
Decrease strain

FIG. 10 FACTORS AFFECTING HOT CRACKING IN BASE METAL HAZ

References cited in this section

1. J.C. BORLAND, GENERALIZED THEORY OF SUPER-SOLIDUS CRACKING IN WELDS (AND CASTINGS), *BR. WELD. J.,* VOL 7 (NO. 8), 1960, P 508-512
2. J.C. BORLAND, HOT CRACKING IN WELDS, *BR. WELD. J.,* VOL 7 (NO. 9), 1960, P 558-559
3. D.A. CANONICO, W.F. SAVAGE, W.J. WERNER, AND G.M. GOODWIN, EFFECTS OF MINOR ADDITIONS ON THE WELDABILITY OF INCOLOY 800, *EFFECTS OF MINOR ELEMENTS ON THE WELDABILITY OF HIGH-NICKEL ALLOYS,* WRC, JULY 1969, P 68-92
4. M.J. CIESLAK, THE WELDING AND SOLIDIFICATION METALLURGY OF ALLOY 625, *WELD. J.,* FEB 1991, P 49S-56S
5. W.F. SAVAGE AND C.D. LUNDIN, APPLICATION OF THE VARESTRAINT TECHNIQUE TO THE STUDY OF WELDABILITY, *WELD. J.,* NOV 1966, P 497S-503S
6. M.J. CIESLAK, J.J. STEPHENS, AND M.J. CARR, A STUDY OF THE WELDABILITY AND WELD RELATED MICROSTRUCTURE OF CABOT ALLOY 214, *METALL. TRANS. A,* MARCH 1988, P 657-667

Cracking Phenomena Associated With Welding

Michael J. Cieslak, Physical and Joining Metallurgy Department, Sandia National Laboratories

Heat-Affected-Zone Cracks

Microfissures are cracks that occur in the area of partial melting and the HAZ adjacent to the fusion line. Because no material can be purified to the state where it solidifies truly as an invariant, all materials have a temperature region of stable two-phase coexistence of solid and liquid. In fusion welding, this manifests itself in the formation of a zone of partial melting at temperatures below the alloy liquidus. The extent of this zone may be enlarged by the presence of chemical inhomogeneity in a material. Local chemical variations will result in local variations in the melting point. Segregation of specific alloying elements to grain boundaries may cause a reduction in the melting temperatures of these areas.

Many microfissures form in a manner somewhat analogous to the formation of solidification cracks in that susceptible sites, generally grain boundaries that intersect the fusion zone, are wetted with liquid from one of various sources. shrinkage strains accumulating as the weld pool advances past the liquated boundary can develop to a level sufficient to cause boundary separation (i.e., cracking). Figure 11 is a photomicrograph of an HAZ microfissure in a cast austenitic stainless steel (this crack also extends into the fusion zone as a solidification crack). Cast microstructures are highly inhomogeneous, raising the possibility of extensive partial melting. Note the melting associated with interdendritic constituents in Fig. 11.

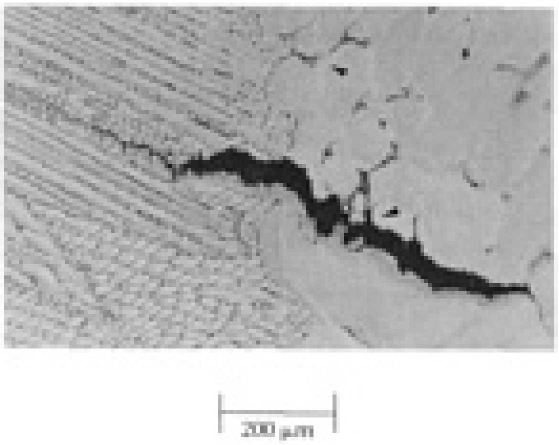

FIG. 11 HAZ MICROFISSURE IN A CAST STAINLESS STEEL. NOTE EXTENSIVE REGION OF PARTIAL MELTING.

Castings represent a special case of inhomogeneity on a continuous scale. Chemical inhomogeneities in wrought products are more discrete, generally involving minor phase particles. In many cases, these minor phases arise from thermomechanical processing of alloys well below the solidus, where the solubility for specific alloying elements can be lower. Because of the potential for extremely rapid rates of heating and cooling (especially heating) in the HAZ of fusion welds, it is entirely possible that minor phases not thermodynamically stable under equilibrium conditions may continue to exist at near-solidus temperatures. Under these conditions, the minor phases may interact with the matrix to form a liquid product. This reaction, often referred to as constitutional liquation (Ref 7), can provide the liquid required to wet grain boundaries producing a microstructure that may be susceptible to cracking.

Hot ductility testing is a common method of assessing the susceptibility of an alloy to this form of cracking. In this test, a laboratory-scale specimen is subjected to a thermal cycle representative of what the material would experience in the HAZ during fusion welding. At various temperatures during both the heating and cooling portions of the cycle, samples are strained to failure. The ductility response (usually expressed in terms of reduction of area) as the alloy progresses through the thermal cycle simulates the response of the alloy to an actual welding situation. In particular, the ability of the material to reestablish ductility after experiencing the peak temperature is a critical measure of sensitivity to this form of cracking.

Figures 12(a) and 12(b) show the hot ductility response of two heats of Cabot alloy 214 containing two different concentrations of boron. The return of ductility in the low-boron alloy after cooling from a peak temperature of 1345 °C (2453 °F) indicates that this alloy should be virtually immune to this form of cracking, as weldability tests confirmed (Ref 6). The high-boron heat showed no appreciable recovery of ductility until a temperature approximately 250 °C (450 °F) below the thermal peak temperature. This alloy should be quite susceptible to this form of cracking, as weldability tests also confirmed (Ref 6). Figure 13 is a fractograph of a hot ductility test specimen from the high-boron alloy taken from the sample strained at 1150 °C (2102 °F) on cooling from the peak temperature. The smooth, intergranular appearance is characteristic of this type of cracking. Although the presence of a liquid phase along this fracture surface could not be unequivocally established by fractographic techniques, it was noted that the failure occurred at an exceedingly low load (Ref 6), consistent with the presence of a liquid film.

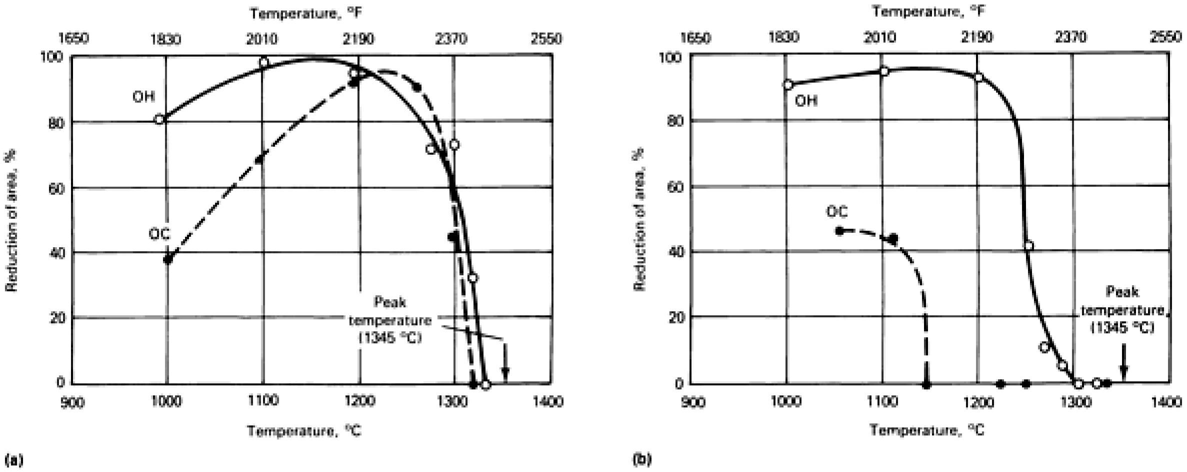

FIG. 12 HOT DUCTILITY RESPONSE OF CABOT ALLOY 214 WITH VARYING BORON CONCENTRATIONS. (A) LOW-BORON CONTENT (0.0002 WT% B). (B) HIGH-BORON CONTENT (0.003 WT% B). (OH, TESTING DONE ON HEATING; OC, TESTING DONE ON COOLING FROM 1345 °C (2455 °F)

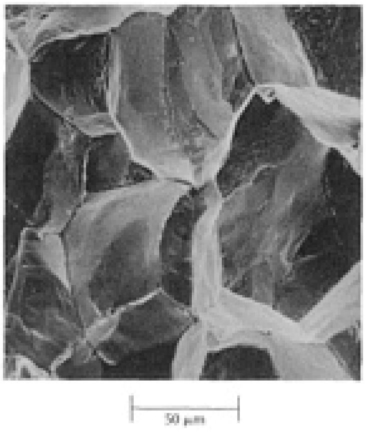

FIG. 13 FRACTURE SURFACE OF A HOT DUCTILITY SAMPLE OF HIGH-BORON ALLOY 214 TESTED AT 1150 °C (2100 °F) ON COOLING FROM 1345 °C (2455 °F)

The presence of a liquid is not essential to the formation of HAZ cracks, as they may occur at temperatures well below the solidus. Poor ductility inherent in certain materials, such as intermetallics, can lead to solid-state failure during welding if the thermal stresses associated with the weld thermal cycle exceed the local tensile strength.

Cracks may also occur in alloys that undergo polymorphic transformations during the weld thermal cycle. The act of fusion welding subjects the fusion zone and the HAZ to temperatures above those normally encountered during prior deformation processing. The resultant HAZ microstructure may be substantially different from that of the base material. In the case of martensitic-type transformations, there is often a relatively large, discontinuous change in the molar volume occurring very rapidly with decreasing temperature below the M_s. If that event causes a local shrinkage, then large, localized tensile stresses can develop. If the newly transformations microstructure has limited ductility (not uncommon in shear transformations or in transformations that result in ordered structures), the likelihood of cracks forming at susceptible sites, such as grain boundaries, matrix/minor phase interfaces, and slip band intersections, is increased.

Reheat cracking, also referred to as stress-relief of strain-age cracking, is another defect type observed in certain alloys that undergo precipitation reactions. Historically observed in several nickel-base superalloys and creep-resistant steels, this type of defect manifests itself during postweld heat treatment, generally as intergranular HAZ cracks.

Postweld heat treatment is often a recommended practice for high-strength thick-section welded steels, where reduction of residual stresses developed as a result of welding is desired. In the case of welded nickel-base superalloys, postweld heat treatment is used both to relieve residual stresses and to achieve optimum mechanical properties through precipitation-hardening reactions.

In alloys that undergo precipitation reactions, the rate at which the alloy strengthens may greatly exceed the rate at which residual stresses are thermally diminished. This can be especially true in the case of heavy-section welds or in alloys with relatively poor thermal conductivity. (That is, before the alloy can reach the temperature at which residual stresses begin to be eliminated, it has aged and lost ductility relative to its as-welded state.) At this new lower level of ductility, the residual stresses present may induce cracks to eliminate the accumulated strain energy.

In principle, the problem of reheat cracking could be largely eliminated if the rate of heating through the precipitation temperature range was rapid enough to prevent precipitate formation. However, extremely rapid rates of heating in welded hardware would likely lead to problems such as excessive distortion; therefore, improved alloys (e.g., alloy 718 among the nickel-base superalloys) have been developed in which the kinetics of precipitate formation have been sufficiently retarded to allow for successful postweld heat treatment.

Dealing with HAZ cracks, however they might form, is generally more difficult than attending to the problem of solidification cracking. Modifying the composition in the HAZ through the addition of consumables is not a viable option. Prior heat treatment may allow the production of a less susceptible starting microstructure, but this option generally is quite limited. Optimization of weld process parameters to minimize the extent of tensile stress and strain development should always be considered, as should optimizing weld joint design as discussed above.

References cited in this section

6. M.J. CIESLAK, J.J. STEPHENS, AND M.J. CARR, A STUDY OF THE WELDABILITY AND WELD RELATED MICROSTRUCTURE OF CABOT ALLOY 214, *METALL. TRANS. A,* MARCH 1988, P 657-667
7. J.J. PEPE AND W.F. SAVAGE, EFFECTS OF CONSTITUTIONAL LIQUATION IN 18-NI MARAGING STEEL WELDMENTS, *WELD. J.,* SEPT 1967, P 411S-422S

Hydrogen-Induced Cracking (Cold Cracking)

Cold cracks are defects that form as the result of the contamination of the weld microstructure by hydrogen. Whereas solidification cracking and HAZ cracking occur during or soon after the actual welding process, hydrogen-induced cracking is usually a delayed phenomenon, occurring possibly weeks or even months after the welding operation. The temperature at which these defects tend to form ranges from -50 to 150 °C (-60 to 300 °F) in steels. The fracture is either intergranular or transgranular cleavage.

As with other forms of cracking, hydrogen-induced cracking involves both a requisite microstructure and a threshold level of stress. It also involves a critical level of hydrogen, which is alloy and microstructure dependent.

In the case of ideal weld processing, hydrogen-induced cracking would be at most a minor welding engineering concern. However, excluding hydrogen from structures during welding is exceedingly difficult. Although the primary source of hydrogen in weld metal is considered to be the disassociation of water vapor in the arc and absorption of gaseous or ionized hydrogen into the liquid, other sources are also available. All organic compounds contain hydrogen in their molecular structure, and all may be broken down in the intense thermal environment of a welding heat source. Organic compounds are ubiquitous in the welding environment, from lubricants in assembly areas to body oils on the hands of welding operators. Plated hardware may also contain high levels of residual hydrogen.

The mechanism of hydrogen-induced crack formation is still being investigated. An early hypothesis, involving the buildup of hydrogen gas pressure in voids, is now generally discredited. Currently, the most widely accepted model involves the presence of preexisting defect sites in the material--small cracks or discontinuities caused by minor phase particles or inclusions. In the presence of existing stress, these sites may develop high local areas of biaxial or triaxial tensile stress. Hydrogen diffuses preferentially to these sites of dilated lattice structure. As the local hydrogen concentration increases, the cohesive energy and stress of the lattice decrease. When the cohesive stress falls below the local intensified stress level, fracture occurs spontaneously. Hydrogen then evolves in the crack volume, and the process is repeated. This model of hydrogen-induced cracking is consistent with the relatively slow and discontinuous nature of the process.

In steels, where the problem of hydrogen-induced cracking is extremely significant, cracking susceptibility has been correlated both with material hardness and strength, and with specific microstructure. Higher-strength steels are more susceptible to hydrogen-induced cracking than low-strength steels. Steels that transform martensitically are particularly susceptible, especially the higher-carbon alloys with twinned martensitic structures. The desire to avoid martensite formation has driven the development of high-strength structural steels for welded applications.

Production of the newer high-strength low-alloy (HSLA) steels uses a variety of precisely controlled alloying additions (for example, aluminum, titanium, vanadium, and niobium) along with meticulous thermomechanical processing to develop a very fine-grained ferrite microstructure possessing substantial strength and fracture toughness with a high degree of resistance to hydrogen-induced cracking. The form of the ferrite produced during transformation on cooling from the austenite phase field is also critical, with acicular ferrite resulting in improved properties compared with grain-boundary, polygonal, or Widmanstätten ferrite. Acicular ferrite is often nucleated on minor phase particles, such as specific oxides or borides. Acknowledgement of the beneficial effect of certain oxides in providing sites for acicular ferrite nucleation in modern steels is in sharp contrast to steel design of earlier decades, when elimination of oxygen to the greatest extent possible was considered essential to the development of optimal fracture behavior.

A useful concept for understanding the susceptibility of carbon and alloy steels to hydrogen-induced cracking is the carbon equivalent (CE), an empirical relationship that attempts to reduce the number of significant compositional variables affecting the weldability of steels into a single quantity. Several carbon equivalent relationships have been developed for different classes of steels. An example is:

$$CE = \%C + \frac{\%Mn}{6} + \frac{\%Cr + \%Mo + \%V}{5} + \frac{\%Si + \%Ni + \%Cu}{15} \qquad \textbf{(EQ 7)}$$

From a metallurgical perspective, the carbon equivalent can be related to the development of hydrogen-sensitive microstructures. That is, as the carbon equivalent increases, microstructures are evolved during cooling through the transformation temperature range that are increasingly more susceptible to hydrogen-induced cracking. At high carbon equivalent values, martensitic structures can be expected.

Carbon equivalent formulas are usually developed from large databases of critical hydrogen concentrations and weld joint restraints that will result in hydrogen-induced cracking in the steels under consideration. At various levels of carbon equivalent, certain weld preheat requirements are often established. In the example of Eq 7, when the carbon equivalent exceeds 0.35%, preheats are recommended to minimize susceptibility to hydrogen cracking. At higher levels of carbon equivalent, both preheats and postheats may be required.

Most structural carbon and low-alloy steels that may be susceptible to hydrogen-induced cracking transform from austenite during cooling through the temperature range of 800 to 500 °C (1470 to 930 °F). The length of time a steel spends in this temperature range during cooling will establish its microstructure and hence its cracking sensitivity. This time segment is generally referred to as $t_{8/5}$ (in seconds). To maximize cracking resistance, a microstructure free of untempered martensite is desired.

Using a specifically developed carbon equivalent, Yurioka (Ref 8) formulated a relationship to establish the critical $t_{8/5}$ for a martensite-free HAZ in low-carbon alloy steels. The carbon equivalent is defined as:

$$CE = \%C^{*} + \frac{\%Mn}{3.6} + \frac{\%Cu}{20} + \frac{\%Ni}{9} + \frac{\%Cr}{5} + \frac{\%Mo}{4} \qquad \textbf{(EQ 8)}$$

where %C* = %C for C $\leq$0.3% and %C* = %C/6 + 0.25 for %C > 0.3%. The critical time length in seconds, $t_{8/5}$, for the avoidance of martensitic transformation is given as:

$$\text{LOG } t_{8/5} = 2.69 \text{ CE}^{*} + 0.321 \qquad \textbf{(EQ 9)}$$

When this quantity is known, welding parameters and preheat temperature for the given thickness of material being welded can be established to produce cooling rates that avoid formation of a martensitic constituent in the HAZ.

The use of preheating and postheating to minimize the susceptibility to hydrogen-induced cracking is an accepted welding procedure for many steels. Preheating controls the cooling rate through the transformation temperature range; a higher base metal temperature in a slower cooling rate through the transformation temperature range and thus a larger $\tau_{8/5}$. Thicker sections often require preheating because of the greater heat-sinking capability of a thick section, for which a given set of welding parameters will produce a faster cooling rate and a smaller $t_{8/5}$. Preheating may also reduce the level of residual stress in the welded assembly.

There are, of course, practical limits to the use of preheating. Excessively high base metal temperatures can make welding difficult, especially manual welding. A high preheat temperature will result in the flattening of temperature gradients, increasing the spatial extent of and the time spent in the austenite phase field. This may result in excessive grain growth in the HAZ, which can lead to a loss of fracture toughness independent of hydrogen content. Preheat temperatures up to 150 °C (300 °F) are not uncommon for low-alloy steels and may increase up to 425 °C (800 °F) for higher-carbon equivalent martensitic steels.

Several benefits can be realized by postheating. At high enough temperatures (550 to 600 °C, or 1020 to 1110 °F), postheating can diminish residual stresses present in the as-welded structure. Microstructural modifications such as tempering can also occur, producing a microstructure less susceptible to hydrogen-induced cracking. Finally, the higher temperature may allow hydrogen to diffuse into the bulk of the structure or even to the surface, where it may recombine and exit as a gas. In either case, the level of hydrogen concentration may be reduced below that required to initiate cracks.

The following guidelines should be followed to minimize the occurrence of hydrogen-induced cracking. For a given level of required strength, the steel with the lowest carbon equivalent should be considered. Low-hydrogen welding practice should be followed. This involves elimination of possible sources of hydrogen by using ultrahigh-purity gases and moisture-free gas lines and by baking coated electrodes following the manufacturer's recommendations to ensure removal of nascent water. Finally, preheating and postheating requirements should be followed assiduously. Figures 14 and 15 summarize the causes and cures of hydrogen-induced cracking in the weld metal and in the base metal, respectively.

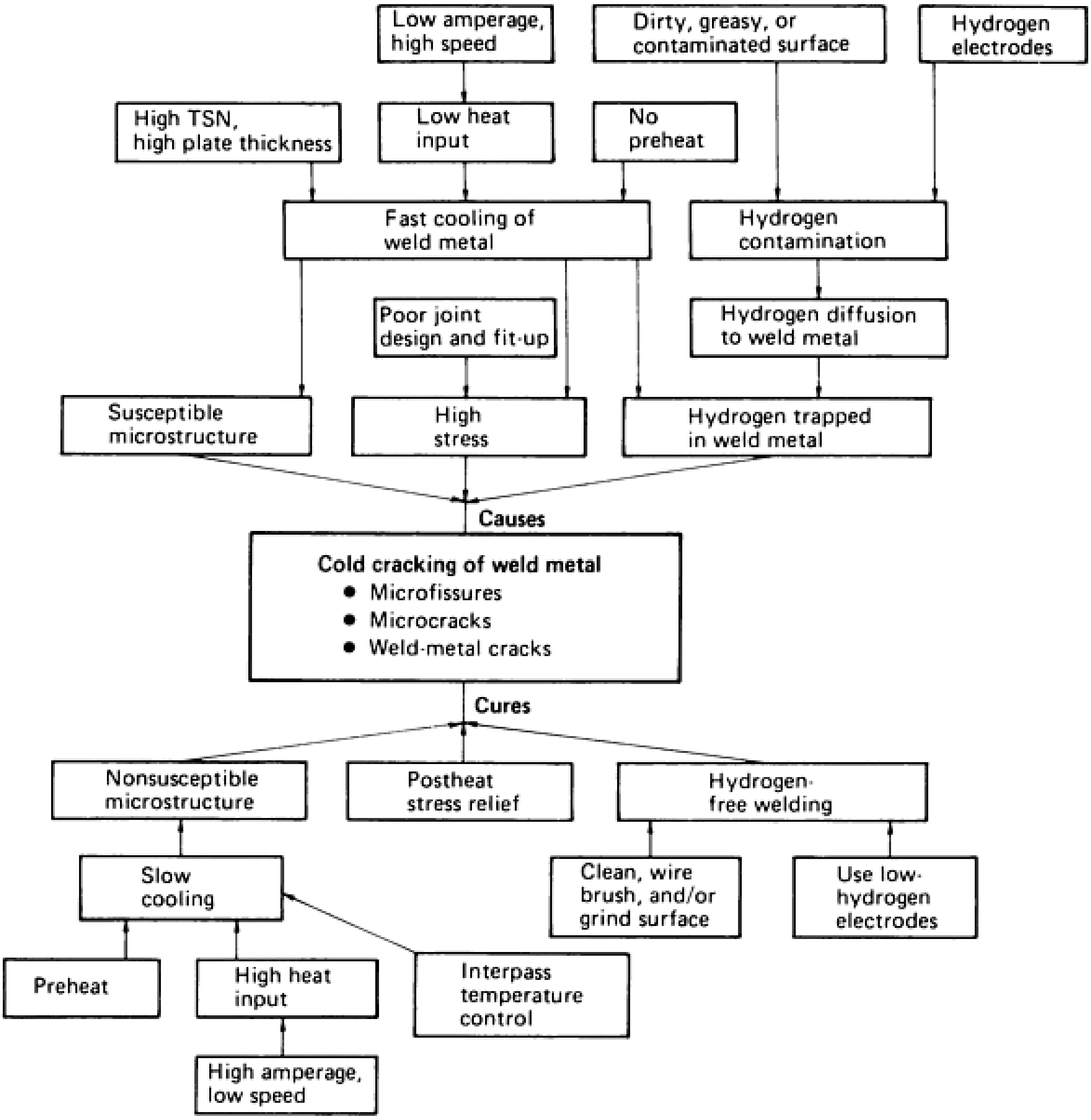
Low amperage, high speed
Dirty, greasy, or contaminated surface
Hydrogen electrodes
High TSN, high plate thickness
Low heat input
No preheat
Fast cooling of weld metal
Hydrogen contamination
Poor joint design and fit-up
Hydrogen diffusion to weld metal
Susceptible microstructure
High stress
Hydrogen trapped in weld metal
Causes
Cold cracking of weld metal
● Microfissures
● Microcracks
● Weld-metal cracks
Cures
Nonsusceptible microstructure
Postheat stress relief
Hydrogen-free welding
Slow cooling
Clean, wire brush, and/or grind surface
Use low-hydrogen electrodes
Preheat
High heat input
Interpass temperature control
High amperage, low speed

FIG. 14 CAUSES AND CURES OF HYDROGEN-INDUCED CRACKING IN WELD METAL. THERMAL SEVERITY NUMBER (TSN), WHICH IS FOUR TIMES THE TOTAL PLATE THICKNESS CAPABLE OF REMOVING HEAT FROM THE JOINT, IS A MEASURE OF THE ABILITY OF THE MEMBER TO SERVE AS A HEAT SINK.

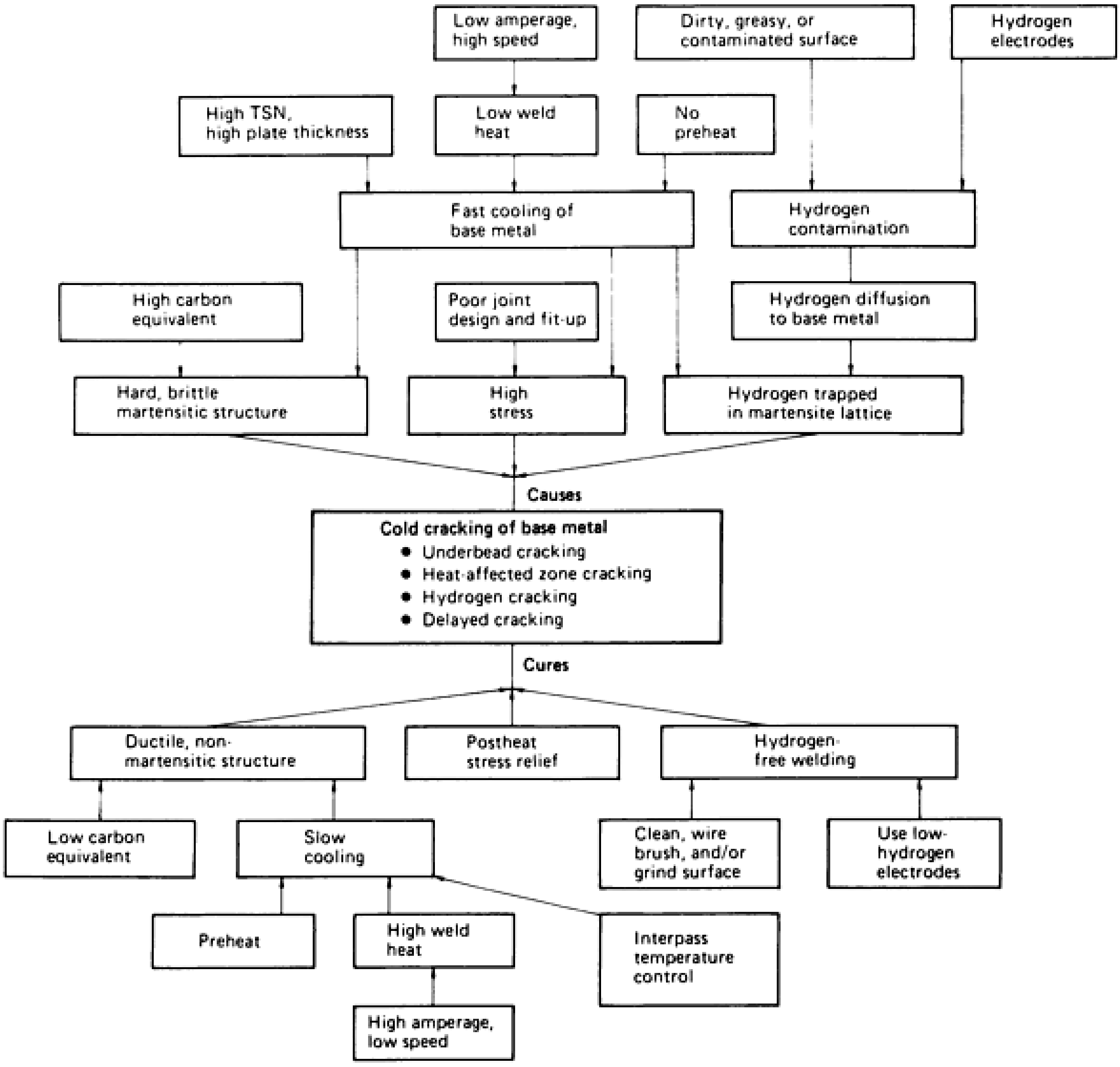
Low amperage, high speed
Dirty, greasy, or contaminated surface
Hydrogen electrodes
High TSN, high plate thickness
Low weld heat
No preheat
Fast cooling of base metal
Hydrogen contamination
High carbon equivalent
Poor joint design and fit-up
Hydrogen diffusion to base metal
Hard, brittle martensitic structure
High stress
Hydrogen trapped in martensite lattice
Causes
Cold cracking of base metal
● Underbead cracking
● Heat-affected zone cracking
● Hydrogen cracking
● Delayed cracking
Cures
Ductile, non-martensitic structure
Postheat stress relief
Hydrogen-free welding
Low carbon equivalent
Slow cooling
Clean, wire brush, and/or grind surface
Use low-hydrogen electrodes
Preheat
High weld heat
Interpass temperature control
High amperage, low speed

FIG. 15 CAUSES AND CURES OF HYDROGEN-INDUCED CRACKING IN BASE METAL

Reference cited in this section

8. N. YURIOKA, WELDABILITY OF MODERN HIGH STRENGTH STEELS, *1ST U.S.-JAPAN SYMP. ADVANCES IN WELDING METALLURGY,* AWS, 1990, P 79-100

Cracking Phenomena Associated With Welding

Michael J. Cieslak, Physical and Joining Metallurgy Department, Sandia National Laboratories

Lamellar Tearing (Ref 9)

Lamellar tearing is cracking that occurs beneath welds. It is found in rolled steel plate weldments. The tearing always lies within the base metal, generally outside the HAZ and parallel to the weld fusion boundary. The problem is caused by welds that subject the base metal to tensile loads in the *z*, or through, direction of the rolled steel. Occasionally the tearing comes to the surface of the metal, but more commonly remains under the weld (Fig. 16) and is detectable only by ultrasonic testing. Lamellar tearing occurs when three conditions are simultaneously present:

- STRAINS DEVELOP IN THE THROUGH DIRECTION OF THE PLATE. THEY ARE CAUSED BY WELD METAL SHRINKAGE IN THE JOINT AND CAN BE INCREASED BY RESIDUAL STRESSES AND BY LOADING.
- THE WELD ORIENTATION IS SUCH THAT THE STRESS ACTS THROUGH THE JOINT ACROSS THE PLATE THICKNESS (THE *Z* DIRECTION). THE FUSION LINE BENEATH THE WELD IS ROUGHLY PARALLEL TO THE LAMELLAR SEPARATION.
- THE MATERIAL HAS POOR DUCTILITY IN THE *Z* DIRECTION.

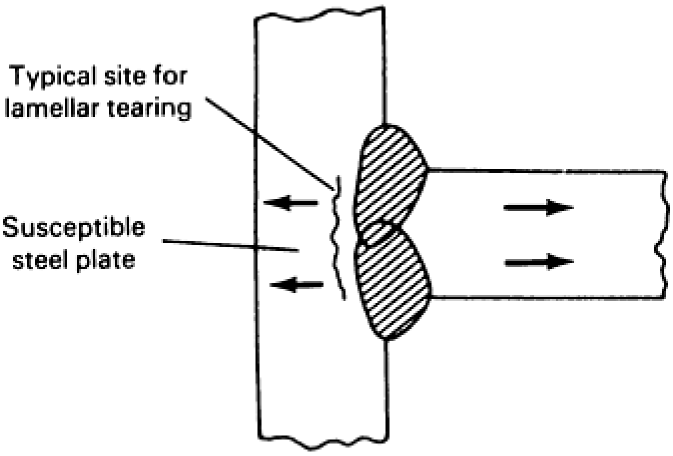

FIG. 16 TYPICAL LOCATION FOR LAMELLAR TEARING IN A T-JOINT

Lamellar tearing can occur during flame cutting and cold shearing operations. Low material strength in the *z* direction is the primary cause, with stress in that direction initiating the tearing. Thermal heating and stresses from weld shrinking create the fracture. Lamellar tearing can take place shortly after welding or occasionally months later. Thicker, higher-strength materials appear to be more susceptible. However, the phenomenon affects only a very small percentage of steel plates, as all three of the above conditions rarely occur in combination.

The problem can be avoided by proper attention to joint details. In T-joints (Fig. 16), double-fillet welds appear to be less susceptible than full-penetration welds. Also, balanced welds on both sides of the joint appear to present less risk than large, single-sided welds. In corner joints, common in box columns, lamellar tearing can be readily detected on the exposed edge of the plate (Fig. 17a). The problem can be overcome by placing the bevel for the joint on the edge of the plate that would exhibit the tearing rather than on the other plate (Fig. 17b). Butt joints rarely exhibit lamellar tearing, because weld shrinkage does not set up a tensile stress in the thickness direction of the plates.

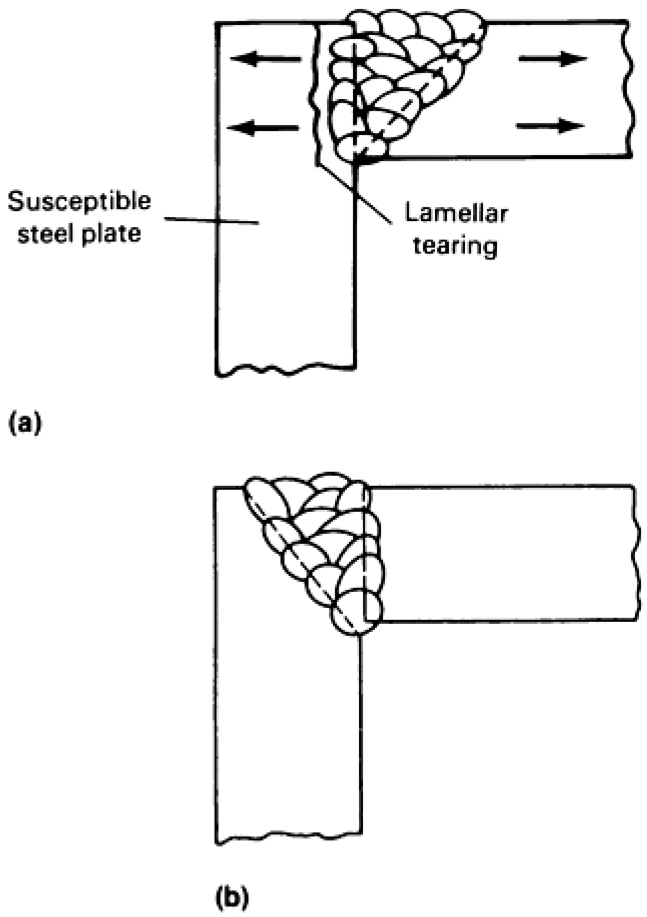

FIG. 17 CORNER JOINT. (A) LAMELLAR TEARING SURFACES AT THE EXPOSED PLATE EDGE. (B) REDESIGNED JOINT

Arc welding processes with high heat input are less likely to create lamellar tearing, perhaps because of the fewer number of applications of heat and the lesser number of shrinkage cycles involved in making a weld. Also, filler metal with lower yield strength and high ductility appears to lessen the possibility of tearing. Preheating does not appear to be particularly advantageous, nor does stress relieving. The "buttering" technique of laying one or more layers of low-strength, high-ductility weld metal deposit on the surface of the plate stressed in the *z* direction will reduce the possibility of lamellar tearing. However, this is an extreme solution and should be used only as a last resort.

Reference cited in this section

9. H.B. CARY, *MODERN WELDING TECHNOLOGY,* 2ND ED., PRENTICE-HALL, 1989, P 702-703

Cracking Phenomena Associated With Welding

Michael J. Cieslak, Physical and Joining Metallurgy Department, Sandia National Laboratories

---

References

1. J.C. BORLAND, GENERALIZED THEORY OF SUPER-SOLIDUS CRACKING IN WELDS (AND CASTINGS), *BR. WELD. J.,* VOL 7 (NO. 8), 1960, P 508-512

2. J.C. BORLAND, HOT CRACKING IN WELDS, *BR. WELD. J.,* VOL 7 (NO. 9), 1960, P 558-559
3. D.A. CANONICO, W.F. SAVAGE, W.J. WERNER, AND G.M. GOODWIN, EFFECTS OF MINOR ADDITIONS ON THE WELDABILITY OF INCOLOY 800, *EFFECTS OF MINOR ELEMENTS ON THE WELDABILITY OF HIGH-NICKEL ALLOYS,* WRC, JULY 1969, P 68-92
4. M.J. CIESLAK, THE WELDING AND SOLIDIFICATION METALLURGY OF ALLOY 625, *WELD. J.,* FEB 1991, P 49S-56S
5. W.F. SAVAGE AND C.D. LUNDIN, APPLICATION OF THE VARESTRAINT TECHNIQUE TO THE STUDY OF WELDABILITY, *WELD. J.,* NOV 1966, P 497S-503S
6. M.J. CIESLAK, J.J. STEPHENS, AND M.J. CARR, A STUDY OF THE WELDABILITY AND WELD RELATED MICROSTRUCTURE OF CABOT ALLOY 214, *METALL. TRANS. A,* MARCH 1988, P 657-667
7. J.J. PEPE AND W.F. SAVAGE, EFFECTS OF CONSTITUTIONAL LIQUATION IN 18-NI MARAGING STEEL WELDMENTS, *WELD. J.,* SEPT 1967, P 411S-422S
8. N. YURIOKA, WELDABILITY OF MODERN HIGH STRENGTH STEELS, *1ST U.S.-JAPAN SYMP. ADVANCES IN WELDING METALLURGY,* AWS, 1990, P 79-100
9. H.B. CARY, *MODERN WELDING TECHNOLOGY,* 2ND ED., PRENTICE-HALL, 1989, P 702-703

Characterization of Welds

Craig B. Dallam, The Lincoln Electric Company; Brian K. Damkroger, Sandia National Laboratories

---

Introduction

WELDS CAN BE CHARACTERIZED according to a number of criteria, including the welding process used, size, shape, mechanical properties, chemical composition, and a number of others. The appropriate methods of characterization depend on the weld's function and the particular set of properties required for the application. In some instances, the ability of a weld to function successfully can be addressed by characterizing the size or shape of the weld. An example of this is where factors related to the welding procedure, such as inadequate weld size, convexity of the bead, or lack of penetration, may cause a weld to fail. In other cases, it is important to characterize metallurgical factors such as weld metal composition and microstructure. Examples might include welds for which the goal is to avoid failures due to inadequate strength, ductility, toughness, or corrosion resistance. In general, the goals of weld characterization are to assess the ability of a weld to successfully perform its function, to thoroughly document a weld and welding procedure that have been demonstrated to be adequate, or to determine why a weld failed.

In the first part of this article, characterization of welds will be treated as a sequence of procedures, each more involved than the last and concerned with a finer scale of detail. Initially, non-destructive characterization procedures will be the focus. The first level of characterization involves information that may be obtained by direct visual inspection and measurement of the weld. A discussion of nondestructive evaluation follows. This encompasses techniques used to characterize the locations and structure of internal and surface defects, including radiography, ultrasonic testing, and liquid penetrant inspection.

The next group of characterization procedures discussed are destructive, requiring the removal of specimens from the weld. The first of these is macrostructural characterization of a sectioned weld, including features such as number of passes; weld bead size, shape, and homogeneity; and the orientation of beads in a multipass weld. Macroscopic characterization is followed by microstructural analysis, including microsegregation, grain size and structure, the phase makeup of the weld, and compositional analysis.

The third component of weld characterization is the measurement of mechanical and corrosion properties. The goal of any weld is to create a structure that can meet all the demands of its service environment. In many cases, the best way of assessing the performance of a weld is to establish its mechanical properties. In addition to a number of standard material tests, many mechanical tests are directed specifically at determining a weld's capabilities. Examples of mechanical properties typically characterized for welds include yield and tensile strength, ductility, hardness, and impact or fracture toughness. Corrosion testing is often employed in situations where a welding operation is performed on a corrosion-resistant material, or in a structure exposed to a hostile environment. Although absolute corrosion performance is important, a major concern is to ensure that a weld and its heat-affected zone (HAZ) are cathodic to the surrounding metal.

Following the discussion of the characterization procedures, the second part of this article will give examples of how two particular welds were characterized according to these procedures.

Acknowledgement

Support for the portion of this work performed at Sandia National Laboratories was provided by the U.S. Department of Energy under contract number DE-AC0476DP00789.

Characterization of Welds

Craig B. Dallam, The Lincoln Electric Company; Brian K. Damkroger, Sandia National Laboratories

Nondestructive Characterization Techniques

Externally Observed Macroscopic Features. Several factors associated with the production and performance of welds are macroscopic and easily observed. The most obvious of these are the size, shape, and general appearance of the weld. To a large extent, these parameters depend on the geometry of the weld joint and the welding process selected. Figure 1 shows schematic representations of fillet, lap, butt, and groove welds, in which a number of features (defined in Table 1) are labeled.

TABLE 1 NONMENCLATURE FOR FILLET, LAP, BUTT, AND GROOVE WELDS

FEATURE	DEFINITION
FACE	EXPOSED SURFACE OF A WELD ON THE SIDE FROM WHICH THE WELDING WAS PERFORMED
ROOT	POINTS, AS SHOWN IN CROSS SECTION, AT WHICH THE BACK OF THE WELD INTERSECTS THE BASE METAL SURFACE
LEG	SHORTEST DISTANCE FROM ROOT TO TOE IN A FILLET WELD
TOE	JUNCTION BETWEEN THE WELD FACE AND THE BASE METAL
THROAT	SHORTEST DISTANCE FROM ROOT TO FACE IN A FILLET WELD
PENETRATION	DEPTH A GROOVE WELD EXTENDS INTO THE ROOT OF A JOINT, MEASURED ON THE CENTERLINE OF A ROOT CROSS SECTION
REINFORCEMENT	WELD METAL IN EXCESS OF THE SPECIFIED WELD SIZE
FACE REINFORCEMENT	REINFORCEMENT AT THE SIDE OF THE WELD FROM WHICH WELDING WAS PERFORMED
SIZE (GROOVE WELD)	JOINT PENETRATION (DEPTH OF ROOT PREPARATION) PLUS THE ROOT PENETRATION (WHEN SPECIFIED)
SIZE (FILLET WELD)	LEG OF THE LARGEST ISOSCELES RIGHT TRIANGLE THAT CAN BE INSCRIBED WITHIN THE FILLET WELD CROSS SECTION

Source: Ref 1

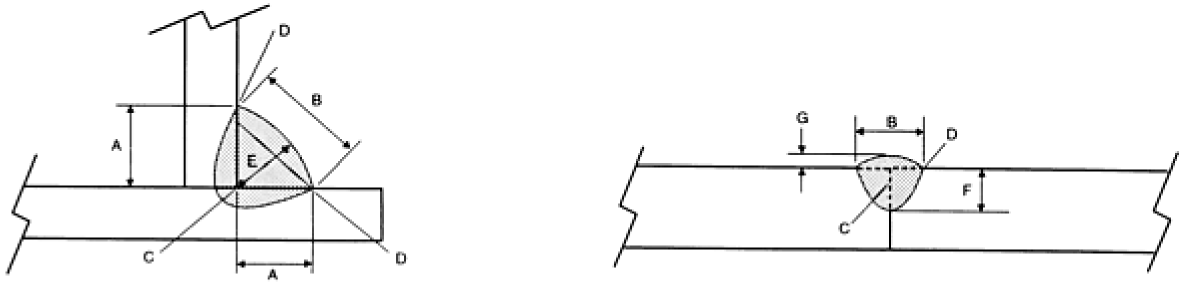

Fillet weld

Butt weld

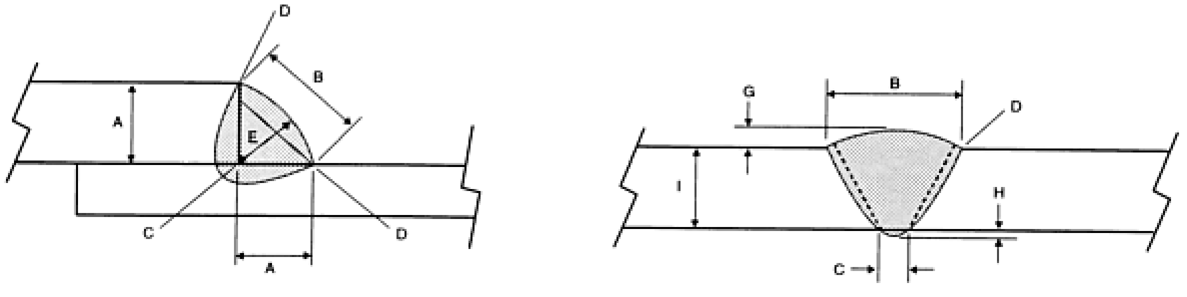

Lap weld

Groove weld

FIG. 1 SCHEMATICS SHOWING PARAMETERS OF COMMONLY USED JOINT GEOMETRIES. A, LEG; A', SMALLER LEG SIZE (EQUAL TO SIZE OF WELD); B, FACE, C, ROOT; D, TOE; E, THROAT; F, PENETRATION (EQUAL TO SIZE OF WELD); G, FACE REINFORCEMENT; H, ROOT REINFORCEMENT; I, SIZE OF WELD

Methods of Characterization. The primary tools used in the external macrocharacterization of a weld are the unaided eye and a hand lens. A number of macrocharacterization parameters (for example, weld location, size, shape, and general uniformity) can be assessed by visual inspection. Figure 2 shows top-view macrographs of two welds that illustrate several of these considerations. In many cases, the presence of gross defects such as hot cracking or porosity may also be detected by visual inspection and a top-view photograph. Additional examples of factors included in overall weld bead appearance include distortion, discoloration due to inadequate shielding or excessive heat, undercut, excessive crater size, and uneven bead width. In fillet welds, a number of performance aspects can be affected by the relative geometry of the joint. Fillet gages can be used to determine the size of a weld, the curvature of the face, and the length of the weld legs.

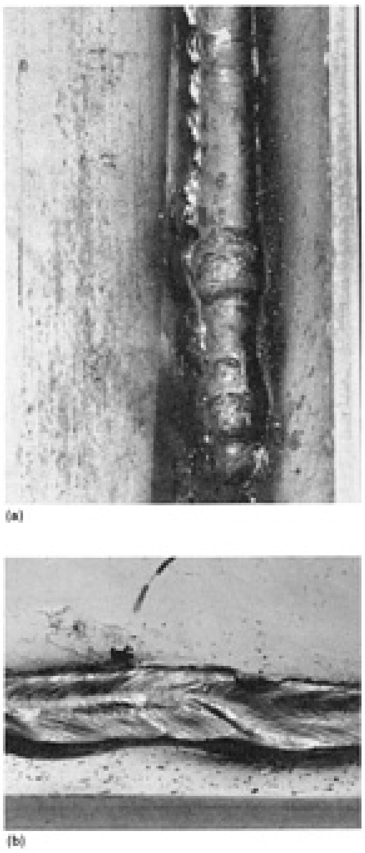

FIG. 2 MACROGRAPHS SHOWING TOP VIEW OF FILLET WELDS. (A) VERTICAL WELD IN TYPE 304L STAINLESS STEEL (NOTE CONVEXITY OF BEAD AND NONUNIFORM REGIONS). (B) HORIZONTAL E 7024 WELD (NOTE WELD PLACEMENT AND LACK OF UNIFORMITY)

General Considerations. Most types of welds will be evaluated with respect to specific macrocharacteristics deemed to be either desirable or undesirable. However, a number of general considerations apply to most types of welds. These include:

- *SIZE:* THE SIZE OF THE WELD SHOULD BE APPROPRIATE FOR THE PART (AS AN EXAMPLE, A GENERAL RULE FOR FILLET WELDS IS THAT THE RATIO OF LEG SIZE/PLATE THICKNESS SHOULD BE BETWEEN 3:4 AND 1:1.
- *LOCATION:* AN INCORRECTLY LOCATED BUTT WELD MAY NOT ALLOW THE PART TO FUNCTION CORRECTLY. A LESS EXTREME EXAMPLE IS A FILLET WELD WITH UNEQUAL LEG LENGTHS, LEADING TO AN UNEVEN STRESS DISTRIBUTION AND PERHAPS LAMELLAR TEARING.
- *UNIFORMITY:* DISTORTION, THE PROBABILITY OF SLAG ENTRAPMENT ON MULTIPASS WELDS, AND UNIFORMITY OF LOAD-CARRYING ABILITY ALL DEPEND ON THE RELATIVE UNIFORMITY OF THE WELD.
- *DEFECTS:* IDEALLY, A WELD SHOULD BE FREE OF MACROSCOPIC DEFECTS. COMMON DEFECTS INCLUDE UNDERCUTTING, LACK OF FUSION, PINHOLE POROSITY, AND SLAG ENTRAPMENT.
- *FACE SHAPE:* A WELD SHOULD HAVE A RELATIVELY FLAT FACE. IF THE WELD IS TOO CONVEX, STRESS WILL BE CONCENTRATED ALONG THE TOE OF THE WELD. CONVERSELY, A HIGHLY CONCAVE WELD FACE WILL RESULT IN LOCALLY HIGH STRESSES IN THE THROAT REGION OF THE WELD.

Defects. The presence of surface or internal defects can degrade the performance of a weld that may appear to be sound in a cursory examination. The causes and means of preventing defects are covered in Ref 2, 3, and 4. Typical defects found in welds include:

- POROSITY
- UNDERCUT
- OVERLAP
- LACK OF PENETRATION
- LACK OF FUSION
- SHRINKAGE VOIDS
- CRATER CRACKING
- MELT-THROUGH OR BURN-THROUGH
- SUBSOLIDUS OR HOT CRACKING
- INCLUSIONS DUE TO ENTRAPPED SLAG, TUNGSTEN ELECTRODE PIECES, OR DEFECTS PRESENT IN THE STARTING MATERIAL

A number of techniques are widely used to assess the presence of surface and subsurface defects in welds. The most common of these are liquid penetrant inspection for surface cracks, magnetic particle inspection, x-ray radiography, and ultrasonic inspection. Penetrant inspection involves the application of an indicator fluid that has a surface tension sufficiently low to be drawn into surface cracks too small to be detected visually. The excess dye or fluorescent material is then removed from the surface, but it remains in and highlights the cracks when a developer is applied. Figure 3 shows cracks in a flux-cored are weld highlighted by dye penetrant inspection. Table 2 lists the characteristics of several types of nondestructive inspection techniques. Additional information is available in the article "Inspection of Welded Joints" in this Volume.

TABLE 2 GUIDELINES FOR SELECTING TECHNIQUES

Technique	Equipment Requirements	Defects Detected	Advantages	Limitations	Comments
liquid-penetrant or fluorescent-penetrant inspection	fluorescent or visible penetration liquids and developers; ultraviolet light for fluorescent dyes	defects open to the surface only; good for leak detection	detects small surface imperfections; easy application; inexpensive; use on magnetic or nonmagnetic material; low cost	time-consuming; not permanent	used on root pass of highly critical pipe welds; indications may be misleading on poorly prepared surfaces
magnetic particle inspection	wet or dry iron particles, or fluorescent; special power source; ultraviolet light for fluorescent dyes	surface and near-surface discontinuities: cracks, porosity, slag	indicates discontinuities not visible to the naked eye; useful for checking edges before welding; no size limitations	for magnetic materials; surface roughness may distort magnetic field; not permanent	test from two perpendicular directions to detect any indications parallel to one set of magnetic lines
radiographic inspection	x-ray or γ ray; film processing and viewing equipment	most internal discontinuities and flaws; limited by direction of discontinuity	provides permanent record of surface and internal flaws; applicable to any alloy	usually not suitable for fillet weld inspection; film exposure and processing	popular technique for sub-surface inspection

				critical; slow and expensive	
ultrasonic inspection	ultrasonic units and probes; reference patterns	can locate all internal flaws located by other methods, as well as small flaws	extremely sensitive; complex weldments restrict usage; can be used on all materials	highly skilled interpretor required; not permanent	required by some specifications and codes

Source: Ref 5

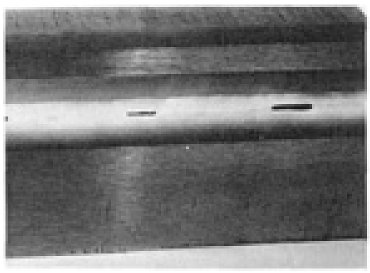

FIG. 3 SURFACE CRACKS IN A FLUX-CORED ARC WELD HIGHLIGHTED BY DYE PENETRANT INSPECTION

X-ray radiography is used to detect internal defects such as porosity or inclusions: these defects show up because of the difference in x-ray absorption between the matrix and defect materials. Although a number of factors affect the resolution level of radiographic techniques (Ref 5, 6), the minimum size of defects considered in AWS specifications is 0.4 mm (0.0156 in.). In practice, this refers to slag entrapment and large inclusions that were present in the starting material. Defect structures are usually quantified by comparison with an existing standard, of which several exist. An example of such a standard is American Petroleum Institute (API) 1104 (Ref 7), a pipeline welding specification. The API 1104 standard includes criteria for acceptance of steel welds based on a number of defects, including inadequate penetration, incomplete fusion, internal concavity, burn-through, slag inclusions, and porosity. For example, one section of API 1104 can be summarized as follows:

> Slag inclusions--50 mm (2 in.) maximum length for elongated slag inclusions, 50 mm (2 in.) limit per 305 mm (12 in.) of weld, 1.6 mm ($\frac{1}{16}$ in.) width limit per elongated slag inclusion. Isolated slag inclusions, 13 mm ($\frac{1}{2}$ in.) per 305 mm (12 in.) of weld, width greater than 3.2 mm ($\frac{1}{8}$ in.), 105 mm ($4\frac{1}{8}$ in.) isolated slag inclusions per 305 mm (12 in.) of weld. Aggregate length of isolated slag inclusion indications may not exceed two times the thinner of the nominal wall thickness and the width may not exceed one-half the thinner of the wall thicknesses joined.

Other examples include ASTM E 390 (Ref 8), which contains reference radiographs for steel welds and the *Structural Welding Code* (Ref 9), which includes a group of radiographic plates showing defect patterns for rounded inclusions corresponding to grade 1 and grade 2 quality levels. Several industries and product categories have applicable specifications and standards. A number of these are given in Ref 5.

Ultrasonic testing can also be used to locate internal defects, including porosity and inclusions. Ultrasonic testing involves transmitting mechanical vibrations through a piece of metal and analyzing both reflected and transmitted vibrations. Vibrations interact with discontinuities in the media through which they are passing, so the operator can detect voids, inclusions, and other internal interfaces. Table 2 includes ultrasonic testing, and Ref 5 lists a number of example specifications and detailed references.

References cited in this section

1. STANDARD WELDING TERMS AND DEFINITIONS, ANSI/AWS A3.0-89, AWS, 1989
2. WELD DISCONTINUITIES, *METALS HANDBOOK*, 9TH ED., VOL 6, AMERICAN SOCIETY FOR METALS, 1983, P 829-855
3. *WELDING HANDBOOK*, 8TH ED., VOL 2, AWS, 1987
4. *PROCEDURE HANDBOOK OF ARC WELDING*, 12TH ED., LINCOLN ELECTRIC CO., 1973
5. CODES, STANDARDS, AND INSPECTION, *METALS HANDBOOK*, 9TH ED., VOL 6, AMERICAN SOCIETY FOR METALS, 1983, P 823-838
6. *RADIOGRAPHY IN MODERN INDUSTRY*, 4TH ED., EASTMAN KODAK CO., 1980
7. WELDING OF PIPELINES AND RELATED FACILITIES, API STANDARD 1104, 17TH ED., AMERICAN PETROLEUM INSTITUTE, 1988
8. REFERENCE RADIOGRAPHS FOR STEEL WELDMENTS, ASTM E 390, ASTM
9. *STRUCTURAL WELDING CODE--STEEL*, ANSI/AWS D1.1-92, AWS, 1992

Characterization of Welds

Craig B. Dallam, The Lincoln Electric Company; Brian K. Damkroger, Sandia National Laboratories

Internal Characterization Requiring Destructive Procedures

This section briefly covers macrostructural and microstructural characterization and compositional analysis. These procedures are usually based on a cross section of the weld, referred to as a transverse section. The transverse section may be supplemented by the top-view photograph, the longitudinal section, and the normal section (Fig. 4). In particular, the top-view photograph will depict the general appearance of the weld, illustrating surface irregularities, spatter, or macroscopic defects such as hot cracking or porosity.

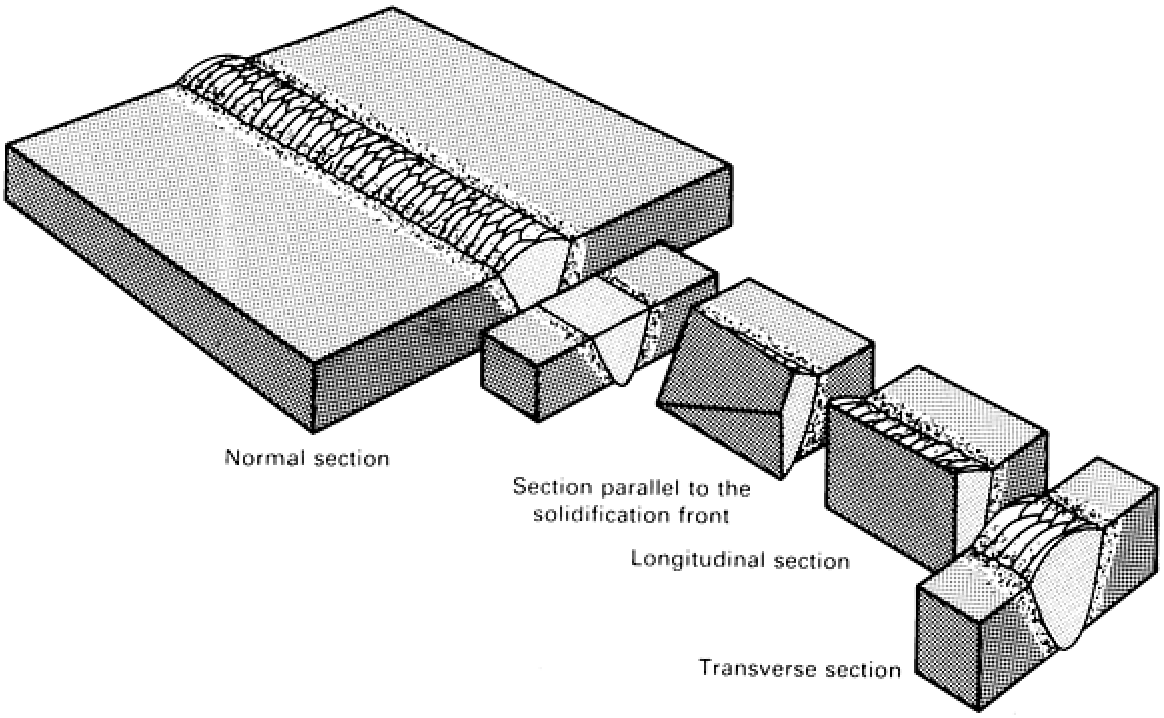

FIG. 4 TYPICAL SECTIONS USED IN THE METALLOGRAPHIC EXAMINATION OF WELDED JOINTS

Weld Macrostructure

Figures 5(a) and 5(b) show transverse sections of two welds, a submerged arc weld made on a 25 mm (1 in.) thick ASTM A 36 steel and a flux-cored weld made on a 50 mm (2 in.) thick ASTM A 537 steel. Readily apparent features include the number of passes and number of layers, fusion zone area, weld aspect ratio, extent of penetration, face width, and the reinforcement and curvature of the top bead. A transverse macrosection will also show any gross porosity or large inclusions present in a weld and the extent of the HAZ. It is important to note that quantitative measurements of weld features made based on a transverse section will be accurate only if the section is properly done. Potential errors, sectioning procedures, and sample preparation techniques are discussed in Ref 10.

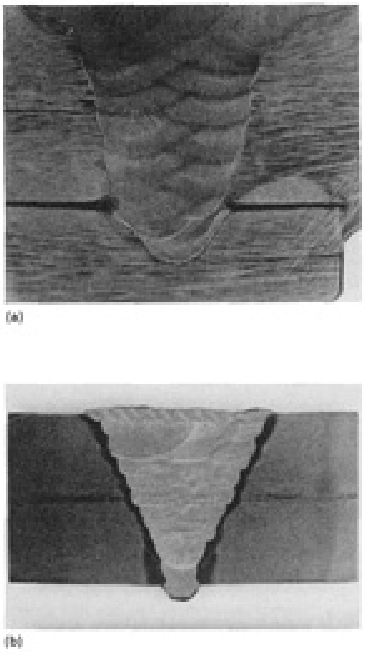

FIG. 5 TYPICAL MULTIPASS ARC WELDS IN STEELS USED IN STRUCTURAL APPLICATIONS. (A) SUBMERGED ARC WELD ON A 25 MM (1 IN.) THICK A 36 STRUCTURAL STEEL; THE MUSHROOM SHAPE OF THE LAST BEAD IS TYPICAL OF WELDS PRODUCED BY THIS PROCESS. (B) FLUX-CORED ARC WELD ON A 50 MM (2 IN.) THICK A 537 STEEL USED IN PRESSURE VESSEL AND STRUCTURE APPLICATIONS (THE LAST LAYER WAS MADE WITH SEVERAL SMALL PASSES TO IMPROVE MECHANICAL PROPERTIES)

Weld Microstructure

In many cases, it is important to examine and characterize the weldment microstructure and to understand its formation and effects on properties. This is the case when the materials and processes involved are not well characterized, so specifications have not been established. Other problems arise where the potential exists for the formation of detrimental microconstituents and/or where the consequences of weld failure are severe. Welding of high-strength or brittle materials are typical examples.

The microstructure of a weld consists of three regions: a fusion zone (material that has been melted); a heat affected zone (material that was not melted, but whose microstructure has been altered); and the base metal. These three regions can be seen in the transverse section shown in Fig. 5(b). Weld microstructures are examined using standard specimen removal and preparation techniques, with some concessions made for their inhomogeneous nature (Ref 10). Similarly, the parameters used to characterize the weld microstructures, such as grain size, grain morphology, and the amount of the various phases or microconstituents present, are those used to characterize monolithic materials. The equipment, procedures, and techniques used are well described in Ref 10.

Special importance may be attached to one or more specific features of a weld microstructure. An example is the classification of ferrite and carbide microconstituents in a low-ally steel weld. The various morphologies included in a characterization of a low-alloy steel weld are listed in Table 3. In these welds, a large amount of acicular ferrite is associated with high toughness levels, whereas martensitic or bainitic microstructures have much lower toughness levels. Bainite and martensite are also associated with higher effective cooling rates, so decreasing the weld metal manganese content, or reducing the cooling rate with increased heat input or preheat, will increase the amount of acicular ferrite and improve weld metal toughness.

TABLE 3 DESCRIPTION OF FERRITE/CARBIDE MICROCONSTITUENTS IN LOW-CARBON STEEL WELDS

Microconstituent	Description
Primary ferrite	
Grain boundry ferrite	Proeutectoid ferrite that grows along the prior austenite grain boundaries. It is equiaxed or polygonal, and it may occur in veins
Intragranular polygonal ferrite	Polygonal ferrite that is not associated with the prior austenite grain boundaries. Much larger than the average width of the surrounding acicular ferrite laths
Ferrite with second phase	
With aligned second phase	Parallel ferrite laths. Further classifications are widmanstätten ferrite and upper/lower bainite
With nonaligned second phase	Ferrite that completely surrounds either equiaxed, randomly distributed microphases, or solitary ferrite laths
Ferrite/carbide aggregate	Fine ferrite/carbide structures, including pearlite
Acicular ferrite	Small nonaligned ferrite grains found within prior austenite grains
Martensite	Martensite colony larger than adjacent ferrite laths

Microstructural characterization of welds has two purposes: to evaluate the microstructure with respect to properties and to relate the microstructure to the process used. The ultimate goal is to optimize the process to produce the most desirable microstructure. In general, the effects of a process and parameters on microstructure are due to the compositional and thermal effects. The compositional effects are largely limited to the fusion zone and will be discussed in the next section. Thermal cycles affect both the fusion zone and HAZ. Microstructural development in the HAZ has been the subject of much study, and a number of computational tools have been developed specifically for the prediction of HAZ microstructures (Ref 11).

Composition of a Weld

The composition of a weld will have a significant effect on its performance, contributing to both the mechanical and corrosion properties of the weld. In some instances, the effects can be drastic, such as sensitization of stainless steel or changing the ductile-to-brittle transition temperature of ferritic steels by an amount sufficient to cause brittle failure under normal use. In a number of cases, particular elements will govern the propensity toward certain defects or behavior. For example, a low-alloy steel weld with a carbon equivalent of over 0.4% is considered to be a high-carbon deposit, and it is treated differently than a low-carbon weld. Similarly, deposit silicon levels in excess of approximately 0.4% are often associated with a defect called "hollow bead," where the root pass in a pipe weld made with a cellulosic electrode has a continuous pore. In general, care must be taken to either minimize the compositional changes associated with welding or to factor the composition and property changes into the design.

Factors Affecting Weld Composition. The composition of a weld is affected by the base metal composition, the composition of any filler metal used (and dilution between the two), reactions with flux or shielding gas, and any material losses associated with the process. These factors are, in turn, largely controlled by the welding setup, process selection, parameters, and stability. The Schaeffler diagram (Ref 12) for stainless steels is an example of composition being controlled by dilution. The weld metal composition and microstructure are predicted based on the "ferrite stabilizer" and the "austenite stabilizer" content of the base metal, the nickel equivalent and chromium equivalent of the filler metal, and the level of dilution. The microstructure will be predicted as martensitic, ferritic, or austenitic, and the level of ferrite in the weldment will be estimated. In these welds, ferrite numbers below 4 indicate microstructures in which low-melting-point compounds may be formed at the grain boundaries, and the welds can be prone to hot cracking. Conversely, welds with ferrite numbers above 10 have reduced corrosion resistance and can be susceptible to the formation of σ phase at high temperatures (Ref 13). The Schaeffler diagram has been revised many times for adaptation to different alloy modifications. An example is the inclusion of nitrogen in the nickel equivalent. Many stainless steels often contain small amounts of nitrogen, and in fact some stainless steels are deliberately strengthened with nitrogen additions. Other modifications include copper in the nickel equivalent and vanadium and aluminum in the chromium equivalent (Ref 14). Additional information is available in the Section "Selection of Stainless Steels" and in the article "Welding of Stainless Steels" in this Volume.

An example of the effect of flux on weld metal composition is seen when low-carbon steel is welded with E6010 and E7018 electrodes, using identical parameters. Although the core wire composition of the two electrodes is identical, the E6010 electrode will produce a deposit with approximately 0.15 to 0.25% Si, and the E7018 electrode will produce a deposit with 0.5 to 0.6% Si. An example of a procedure-related effect is seen when the arc length is changed for welds made with an E7018 electrode. Arc length changes can alter the silicon content from 0.3 to 0.6%, and they can alter the manganese content from 0.8 to 1.3%.

Measurement Techniques and Procedures. The usual techniques for compositional analysis of metallic samples can be used on welds and are covered in Ref 15. However, some considerations must be made when dealing with welds, because of their inhomogeneity. The most commonly used method for compositional analysis of welds is optical emission spectroscopy, where a spot of material (typically 6 mm, or $\frac{1}{4}$ in., in diameter) is ablated off the surface of the specimen and the light emissions are analyzed. In many welds, a spot this size will encompass several passes in a multipass weld, and it will perhaps encompass weld metal, HAZ, and base metal in a single-pass weld. An average composition over this large an area may not address local effects, for example, sensitization of stainless steel.

Welds are prone to both macrosegregation and microsegregation, and the properties may be governed by the composition in a very local region. A thorough characterization of welds requires techniques with sufficient spatial resolution to characterize their inhomogeneity. Scanning electron microscopy with wavelength or energy dispersive x-ray analysis systems, electron microprobes, and x-ray fluorescence techniques are often employed. Often, a number of closely spaced analyses are made along a line traversing a weld. This data may then be coupled with a similar microhardness traverse and microstructural analyses to characterize the variation across the weld. Thorough treatments of a number of microanalytical techniques are contained in Ref 16.

Mechanical Testing

A number of mechanical properties are used to characterize welds, including strength, ductility, hardness, and toughness. In general, the same samples and procedures are used in other areas of metallurgy (Ref 17). However, a prominent concern regarding the mechanical performance of welds is the direct comparison with base material. The goal is to ensure that the weld is not the weakest component of a structure, or if it is, to compensate for this in the design.

Strength. Yield and tensile strength are measured for all-weld-metal specimens using a standard tensile test (Ref 18), but with specimens removed from test plates welded according to AWS-specified procedures (Ref 19). These tests form the basis for the assignment of yield and ultimate strength values to welds made using a specific electrode and according to a set procedure. Additional tests are sometimes performed to compare the base metal and weld metal strengths. An example of this type of test is the transverse tensile test, in which the specimen is removed from the weld so that the loading axis is perpendicular to the weld bead and the weld reinforcement is left intact. The goal of this test is to verify that overload failure will occur in the base metal rather than in the weld metal or HAZ.

Ductility is another critical weld property. In addition to defects, many welding processes can produce hard, brittle microstructures. The standard measures of ductility--percent reduction in area and percent elongation--are obtained in a uniaxial tensile test. Another test often specified for welds is a bend test (face bends, roof bends, and side bends). In this test, a strip of material containing a weld is deformed around a specified radius and its surface is examined. The criteria for success or failure are the number and size of defects seen on the outer surface of the bend. An example of bend test criteria is the AWS *Structural Welding Code* (Ref 9), which calls for bending around a 19 mm (0.75 in.) radius for materials with yield strengths less than or equal to 345 MPa (50 ksi), a 25 mm (1 in.) radius for 345 to 620 MPa (50 to 90 ksi) materials, and a 32 mm (1.25 in.) radius for materials with yield strengths greater than or equal to 620 MPa (90 ksi).

Hardness. One common use of hardness values in weld specifications is as a check for the formation of microstructures that might have low ductility and toughness and thus are prone to cracking. For example, in pipeline steels, the formation of martensite in the HAZ is a cause for concern because of the potential for cracking. This is addressed by specifying maximum values for microhardness traverses across several sections of the weld. Hardness values are also used as an indicator of susceptibility to some forms of stress-corrosion cracking.

Toughness is the ability of a material to absorb energy during fracture. There are two approaches to toughness testing: impact toughness testing and fracture mechanics testing.

Impact Toughness Testing. To test impact toughness, a sample of specified geometry is subjected to an impact load, and the amount of energy absorbed during fracture is recorded. Usually the specimen is oriented so that the notch and expected plane of fracture run longitudinally through the weld metal. Charpy tests do not measure an inherent material property, but they result in a relative measure of impact toughness between materials. A very common use of the Charpy test is to determine a material's ductile-to-brittle transition temperature by performing tests at several different temperatures. AWS A5.1 (Ref 19) gives minimum Charpy impact values, at several temperatures, for welds made on carbon steel using a number of different electrodes.

Fracture Mechanics Testing. The second type of toughness testing is based on fracture mechanics, and it can use either linear elastic or elastic-plastic methodologies. Although elastic-plastic behavior (J_{Ic}) is becoming of interest in some cases, the bulk of fracture mechanics testing is based on linear elastic considerations. These tests, using specimens and procedures given in ASTM E 399 (Ref 20), are used to measure a material's fracture toughness (K_{Ic}), which is a material property. In the case of welding, fracture toughness is usually expressed using a value for crack tip opening displacement (Ref 21). Fracture toughness testing has only recently begun gaining acceptance as applicable to welds. The major shortcomings of this approach include the complexity and cost of testing and the wide variability in fracture toughness values for weld metal, due largely to the inhomogeneous nature of welds and residual stress effects.

References cited in this section

9. *STRUCTURAL WELDING CODE--STEEL,* ANSI/AWS D1.1-92, AWS, 1992
10. METALLOGRAPHY OF WELDMENTS, *METALS HANDBOOK,* 9TH ED., VOL 9, AMERICAN SOCIETY FOR METALS, 1985, P 577-586
11. K.E. EASTERLING, MICROSTRUCTURE AND PROPERTIES OF THE HAZ, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* S. DAVID AND J. VITEK, ED., ASM INTERNATIONAL, 1990, P 177-188
12. WRC 1992 DIAGRAM, *WELD. J.,* MAY 1992, P 171
13. T.G.F. GRAY, J. SPENCE, AND T.H. NORTH, *RATIONAL WELDING DESIGN,* NEWNES-BUTTERWORTHS, 1975, P 111
14. R.H. ESPY, WELDABILITY OF NITROGEN-STRENGTHENED STAINLESS STEELS, *WELD. J. RES. SUPPL.,* VOL 61 (NO. 5), MAY 1982, P 149S-156S
15. METAL TEST METHODS AND ANALYTICAL PROCEDURES, *1991 ANNUAL BOOK OF ASTM STANDARDS,* VOL 03.05 AND 03.06, ASTM, 1991
16. *MATERIALS CHARACTERIZATION,* 9TH ED., VOL 10, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1986
17. *MECHANICAL TESTING,* 9TH ED., VOL 8, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1985
18. STANDARD METHODS OF TENSION TESTING OF METALLIC MATERIALS, ASTM E 8, *ANNUAL BOOK OF ASTM STANDARDS,* VOL 03.01, ASTM, 1984
19. *SPECIFICATION FOR CARBON STEEL ELECTRODES FOR SHIELDED METAL ARC WELDING,* ANSI/AWS A5.1-91, AWS, 1991
20. STANDARD TEST METHOD FOR PLANE-STRAIN FRACTURE TOUGHNESS OF METALLIC MATERIALS, ASTM E 399, *ANNUAL BOOK OF ASTM STANDARDS,* ASTM, 1984
21. STANDARD TEST METHOD FOR CRACK TIP OPENING DISPLACEMENT (CTOD) FRACTURE

Characterization of Welds

Craig B. Dallam, The Lincoln Electric Company; Brian K. Damkroger, Sandia National Laboratories

Weld Characterization Examples

In this section, two weld will be analyzed to serve as examples of techniques discussed thus far in this article. The first of these welds is a multipass shielded metal arc weld/flux-cored arc weld made on pipe steel for an industrial application. This type of weld would typically be made in great quantity. In practice, it would be characterized largely on the basis of its external appearance, perhaps supplemented by nondestructive testing. The initial characterization of a weld and procedure, however, would be based on chemical composition and mechanical properties. Primary concerns are whether the weld being examined is representative of a normal weld of this type and can thus be reasonably well described based on the materials, process, and specifications involved in its production. The bulk of the actual characterization of such a weld would be done during the development and certification of the initial process to determine its suitability for the intended application.

The second example is a weld used for a very specific application. This weld is a butt weld made on Ti-6Al-4V, with a tantalum shim inserted between the base metal sections. Among the requirements for this weld are that it have microstructural and property homogeneity and stability at elevated temperatures and that it be cathodic to the base metal under unusual and extremely severe environmental conditions. In this case, the concerns center less on the eventual application and more on building a thorough database regarding the metallurgical and microstructural characteristics of the weld.

Example 1: Multipass Weld in a 1.07 m (42 in.) Diameter X-65 Pipe.

This weld is a circumferential weld on a 12.7 mm (0.50 in.) thick, 1.07 m (42 in.) diameter X-65 pipe. The X-65 materials is designated by API as a pipe steel with at least 448 MPa (65 ksi) yield strength and 566 MPa (82 ksi) tensile strength. This type of weld is typical for a large oil or gas pipeline. However, for this example, procedures were deliberately selected to produce a weld with a range of defects. These procedures, given by Table 4, are not representative of standard practice.

TABLE 4 PARAMETERS (NOT STANDARD PROCEDURE) USED TO OBTAIN MULTIPASS WELD IN 1.07 M (42 IN.) DIAMETER X-65 STEEL PIPE

PROCESS	PASS NO.	ELECTRODE	DIAMETER		WIRE FEED		CURRENT, A	VOLTAGE, V	TRAVEL SPEED	
			mm	in.	mm/s	in./s			mm/s	in./s
SMAW	1	E8010-G	4.0	0.16	. . .	. . .	150	26-28	5.9-7.2	2.3-2.8
	2	E8010-G	4.0	0.16	. . .	. . .	180	26-28	5.9-7.2	2.3-2.8
FCAW	3	E71T8-K6	1.73	0.068	47	1.9	200	19	4.5	1.8
	4	E71T8-K6	1.73	0.068	47	1.9	200	19	4.0	1.6
	5	E71T8-K6	1.73	0.068	47	1.9	200	19	4.0	1.6
	6	E71T8-K6	1.73	0.068	47	1.9	200	19	4.7	1.9
	7	E71T8-K6	1.73	0.068	47	1.9	200	19	5.9	2.3
	8	E71T8-K6	1.73	0.068	42	1.7	180	18	4.2	1.7
	9	E71T8-K6	1.73	0.068	42	1.7	180	18	4.5	1.8

Additional specifications include: 12.7 mm (0.50 in.) wall thickness; 1.59 mm (0.0625 in.) land; 1.59 mm (0.0625 in.) gap; 60° bevel in groove; vertical down position; preheat and interpass temperature of 150 °C (300 °F)

Visual Observation. The weld was initially characterized by visual observation. From the top of the joint, the size and uniformity of the weld and undercut can be observed (Fig. 6a). From the underside, spatter, lack of penetration, and internal undercut can be seen (Fig. 6b).

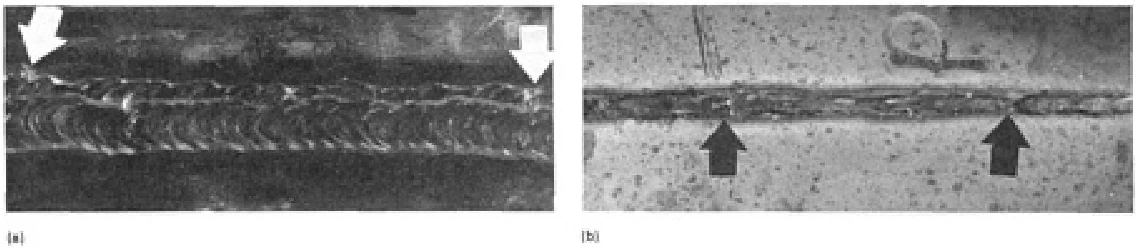

FIG. 6 DEFECTS IN A MULTIPASS WELD MADE ON A 1.07 M (42 IN.) DIAMETER X-65 STEEL PIPE. (A) EXTERIOR VIEW SHOWING LACK OF UNIFORMITY (RIGHT ARROW) AND UNDERCUT (LEFT ARROW). (B) INTERIOR VIEW SHOWING LACK OF PENETRATION (RIGHT ARROW) AND BURN-THROUGH (LEFT ARROW)

Nondestructive evaluation was then used to examine the weld for internal and surface defects. Slag entrapment and lack of fusion are typical internal defects for this type of weld, and they can be seen in the radiograph shown in Fig. 7. Figure 7 also shows the external defects noted during visual observation. In addition to nondestructive evaluation, other tests that highlight defects (for example, "nick-break" tests and face or root bend tests) can be run, but these were not used in this case. Similarly, ultrasonic testing, magnetic particle testing, and penetrant testing were not used in this characterization.

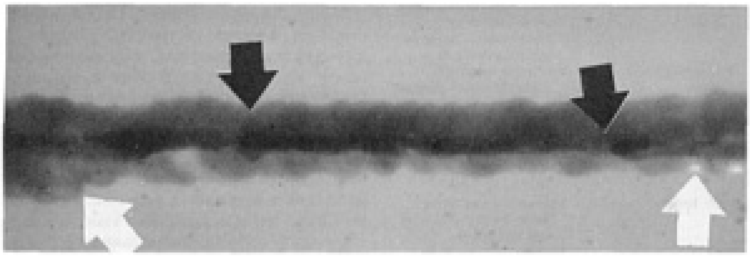

FIG. 7 RADIOGRAPH SHOWING POSITIVE IMAGE OF MACROSCOPICALLY VISIBLE DEFECTS IN THE X-65 PIPE STEEL OF FIG. 6. SEE FIG. 6 FOR DESIGNATION OF SPECIFIC DEFECTS.

Weld Macrostructure. Destructive evaluation begins by removing a transverse section from the weld and preparing it for metallographic examination (Fig. 8). The welding pass sequence (that is, number and size of passes, number of passes per layer, amount of penetration, and extent of HAZ) is apparent. In addition, any gross defects present in the sectioning plane are visible. In this example, nine passes were made, three layers with one pass and three with two passes. Figure 8 also shows four small pores, one each in the third, fourth, sixth, and ninth passes. In this case, the porosity was induced by deliberately varying the electrode stickout.

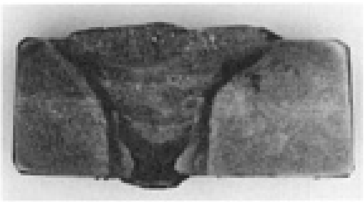

FIG. 8 TRAVERSE SECTION OF THE X-65 PIPE STEEL SHOWN IN FIG. 6 AND 7. PORES IN PASSES 3, 4, 6, AND

8 WERE CAUSED BY INTENTIONALLY VARYING THE ELECTRODE STICKOUT.

Weld Metal Composition. Compositional analysis of the weld was performed at this point. For this application, the bulk, or average, composition of the weld is of the most interest. The bulk composition of the weld was measured using an emission spectrometer, and the results are given in Table 5. The interstitial content of the weld is also of interest, and the levels of oxygen and nitrogen were measured with dedicated instruments. The oxygen level of the weld was found to be 110 ppm and the nitrogen level, 330 ppm. For some welds, techniques can be used to map compositional variations within the weld. In this case, the bulk composition, measured at the weld center, is considered adequate to characterize the weld.

TABLE 5 BULK COMPOSITION OF WELD IN 1.07 M (42 IN.) DIAMETER X-65 STEEL PIPE

SAMPLE	COMPOSITION, WT. %									
	C	Mn	Si	Al	Ni	Cr	Mo	Ti	V	Nb
WELD METAL	0.062	0.85	0.22	0.76	0.65	0.02	0.02	0.011	0.008	0.003
BASE METAL	0.093	1.35	0.21	0.02	0.02	0.04	0.00	0.002	0.069	0.033
ELECTRODES:										
E8010-G	0.14	0.70	0.20	0.01	0.73	0.02	0.07	0.008	0	0
E71T8-K6	0.056	0.91	0.24	0.95	0.97	0.02	0.01	0.002	0	0

Weld Microstructure. The transverse section was also used for examination and characterization of the microstructure of the weld. Figure 9 shows micrographs of different regions of the weld. Figure 9(a) shows weld metal that has not been reheated by successive passes. Figure 9(b) shows weld metal that has been reheated and has a finer structure than the non-reheated weld metal. Figure 9(c) shows a location in the HAZ that closely resembles the structure seen in the reheat zone. The important parameters to characterize for this type of weld are the percentage of reheat zone, the relative amount of various microconstituents, and the average grain size. In the example weld, the weld metal is approximately 80% reheat zone. In the non-reheated zone (Fig. 9a), quantitative microstructural analysis showed that the structure consisted of 20% grain boundary ferrite, 20% intergranular polygonal ferrite, 16% acicular ferrite, 32% ferrite with an aligned second phase, and 12% ferrite with a nonaligned second phase. No martensite or ferrite-carbide aggregate was found.

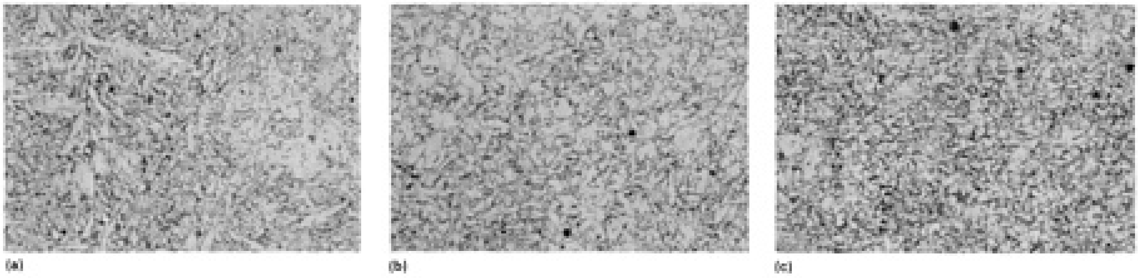

FIG. 9 MICROSTRUCTURE OBTAINED IN SELECTED REGIONS OF THE TRAVERSE SECTION OF THE X-65 PIPE STEEL OF FIG. 8. (A) NON-REHEATED WELD METAL. (B) REHEATED WELD METAL. (C) HEAT-AFFECTED ZONE. ALL 250×

Mechanical Testing. Following the characterization of the microstructure, specimens were removed and used for mechanical testing. For this application, the pertinent mechanical tests are tension tests (transverse and all-weld-metal), Charpy impact tests, fracture toughness tests, and a microhardness traverse. A transverse tensile test is primarily used to ensure that the welded joint is not the weak link in the final structure. In this case, a 305 × 32 mm (12.0 × 1.25 in.) specimen was removed from the pipe, the long axis perpendicular to the weld bead. The specimen was loaded to 247 kN (27.8 tonf) and failed in the base metal. Based on a weld cross-sectional area of 5.08 cm^2 (0.787 $in.^2$), a minimum weld tensile strength was calculated as 487 MPa (70.6 ksi).

An all-weld-metal tensile test was run using a specimen removed from the bead so that its loading axis was parallel to the weld direction. In this case, the weld was aged prior to testing, with the heat treatment being 48 h at 104 °C (220 °F). The results of the tensile test are shown in Fig. 10. The yield strength was found to be 520 MPa (75.4 ksi), the ultimate strength 594 MPa (86.1 ksi), and the percent elongation 30%. With respect to a typical vendor specification for this type of weld, these values indicate that the weld is adequate.

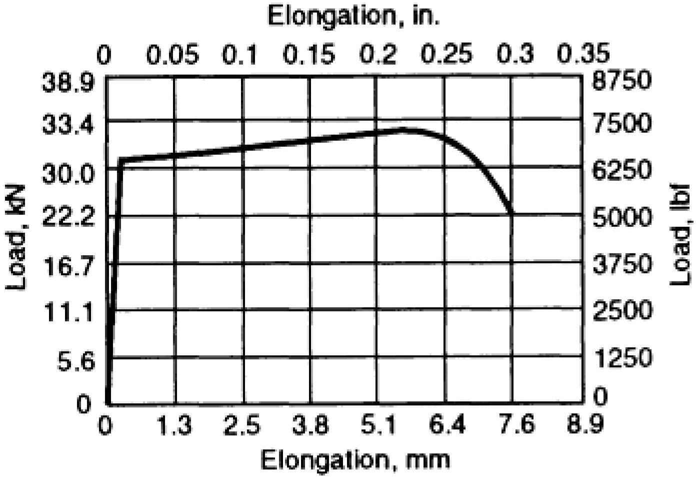

FIG. 10 PLOT OF LOAD VERSUS ELONGATION FOR ALL-WELD-METAL TENSILE TEST OF A 6.4 MM (0.25 IN.) DIAMETER SPECIMEN TAKEN FROM THE 1.07 M (42 IN.) DIAMETER X-65 STEEL PIPE

Charpy V-notch impact specimens were removed from the weld as shown in Fig. 11. For this type of application, tests are usually run at only one temperature, dictated by a specification and/or past experience. For a full characterization, however, tests will be run over a broad enough range to clearly establish the upper- and lower-shelf energy levels and to identify the ductile-to-brittle transition temperature (DBTT). For this example, tests were run at several temperatures, and the results are shown in Fig. 12. This figure clearly shows that the upper shelf energy is approximately 200 J (150 ft · lbf), the lower shelf 2 J (1.5 ft · lbf), and the DBTT -10 °C (14 °F).

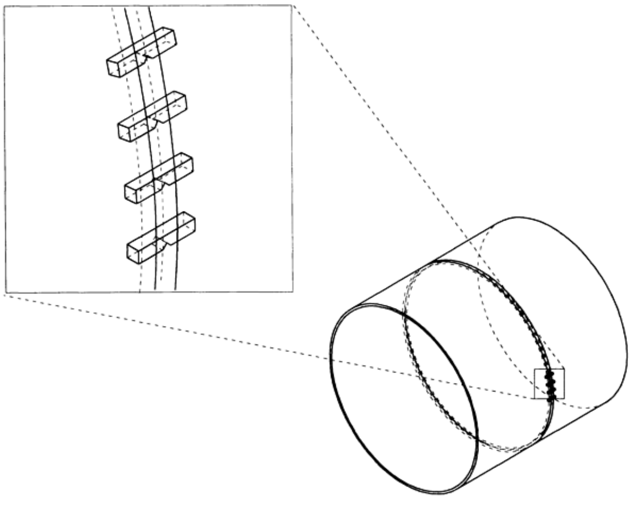

FIG. 11 SCHEMATIC SHOWING LOCATION AND ORIENTATION OF CHARPY V-NOTCH SPECIMENS TAKEN FROM THE 1.07 M (42 IN.) DIAMETER X-65 STEEL PIPE. ALTHOUGH INSET SHOWS WIDELY SPACED SAMPLES FOR THE SAKE OF CLARITY, ACTUAL SAMPLES REMOVED ARE TYPICALLY ADJACENT TO ONE ANOTHER.

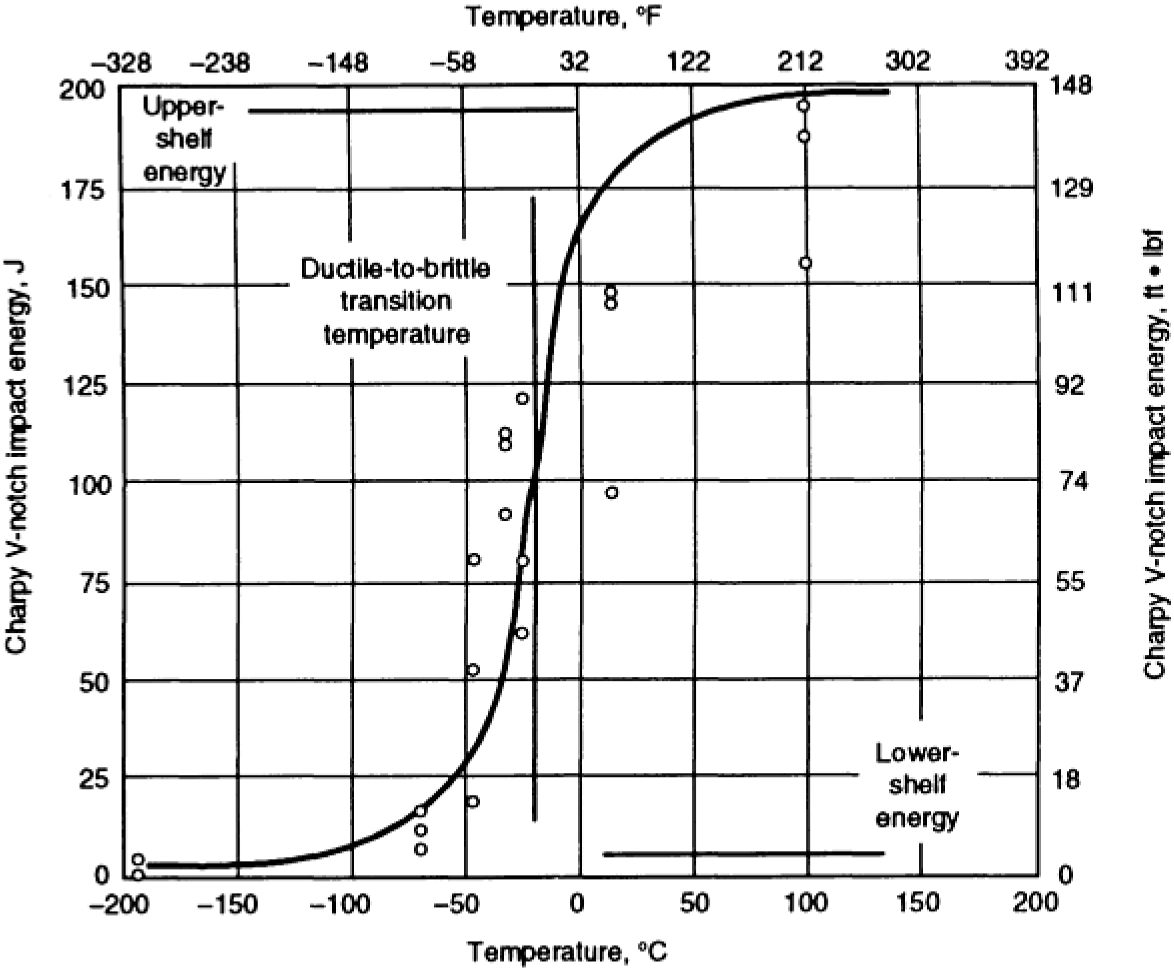

FIG. 12 CHARPY V-NOTCH IMPACT ENERGY TEST RESULTS USED TO OBTAIN DUCTILE-TO-BRITTLE TRANSITION TEMPERATURE (-10 °C, OR 14 °F) FOR THE 1.07 M (42 IN.) DIAMETER X-65 STEEL PIPE

Fracture toughness specimens were removed from the weld in an orientation similar to that of the Charpy specimens. As with Charpy tests, fracture toughness specifications usually require a certain value at a given temperature, usually the minimum service temperature. Crack tip opening displacement tests were conducted at -45 °C (-50 °F) on three specimens removed from the example weld. The results were found to vary from 0.137 to 0.322 mm (0.005 to 0.013 in.), with the average being 0.229 mm (0.009 in.). Figure 13 shows the fracture surface of one of the test specimens, on which the ductile fracture mode and overload zone can be clearly seen.

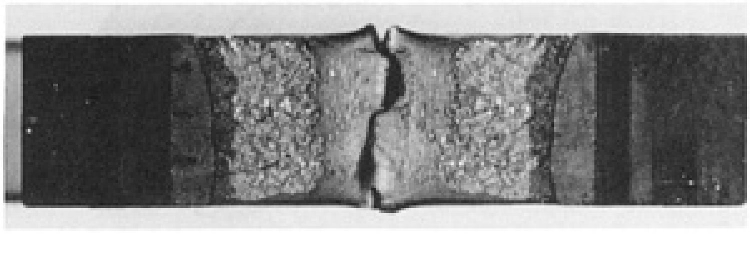

FIG. 13 MACROGRAPH OF THE FRACTURE SURFACE OF A CRACK TIP OPENING DISPLACEMENT TEST SPECIMEN REMOVED FROM THE 1.07 M (42 IN.) X-65 STEEL PIPE. THE REGIONS OF THE SURFACE SHOWN CORRESPOND WITH THE NOTCH, PRECRACK, STABLE CRACK GROWTH, FAST FRACTURE, AND OVERLOAD ZONES.

In addition to the previous mechanical tests, microhardness testing is often used for this type of weld. The primary purpose of microhardness testing is to check for the presence of martensite in the HAZ. Figure 14 shows a schematic of the example weld and the location and values of several microhardness tests. In this case, none of the values exceeded 240 HV, the level associated with the presence of martensite.

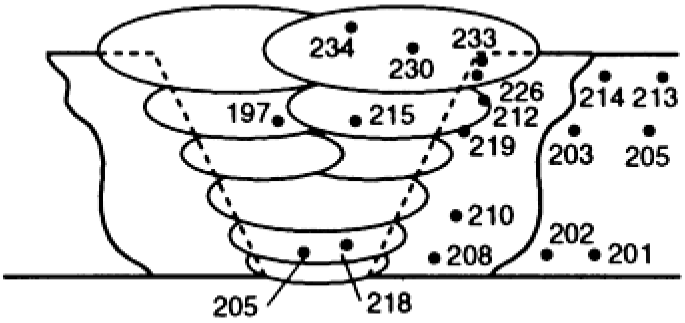

FIG. 14 MICROHARDNESS READINGS (LOCATION IN MULTI-PASS WELD INDICATED BY DOTS) BELOW 240 HV INDICATING THAT NO MARTENSITE STRUCTURE IS PRESENT IN THE 1.07 M (42 IN.) X-65 STEEL PIPE

Typically, this type of pipe weld is characterized with respect to stress-corrosion cracking resistance. In addition, testing is often performed to characterize the HAZ toughness.

Example 2: Welding of 2.5 mm (0.100 in.) Thick Ti-6Al-4V Sheet using a 0.127 mm (0.005 in.) Thick Tantalum Shim.

Figure 15 shows a transverse section of a butt weld, where the base metal is 2.5 mm (0.100 in.) thick Ti-6Al-4V and a 0.127 mm (0.005 in.) thick tantalum shim has been placed in the joint. This weld geometry was used in a study of the effect of dilution on the microstructures and properties of dissimilar metal welds of titanium to refractory metals. The questions to be answered in the characterization of this weld were:

- HOW UNIFORMLY WOULD THE TANTALUM BE DISTRIBUTED IN THE WELD POOL?
- HOW ACCURATELY COULD THE COMPOSITION OF THE WELD POOL BE PREDICTED BASED ON DILUTION CALCULATIONS?
- WHAT PHASES AND MICROCONSTITUENTS ARE PRESENT IN THE WELD METAL?
- WHAT ARE THE MECHANICAL PROPERTIES OF THE WELD METAL?

In addition, a number of very application-specific corrosion and elevated-temperature tests were performed on the dissimilar metal welds. In contrast with the pipe weld of the first example, the primary tools used in this characterization were macrostructural examination, compositional mapping, x-ray diffraction, tension testing of all weld metal specimens, and microhardness traverses.

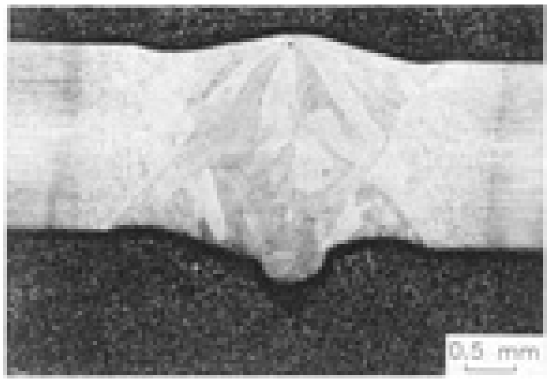

FIG. 15 MICROGRAPH OF TRANSVERSE SECTION OF AN ELECTRON-BEAM WELDED BUTT WELD JOINING 2.5 MM (0.100 IN.) THICK TI-6AL-4V SHEET USING A 0.127 MM (0.005 IN.) THICK TANTALUM SHIM PLACED IN THE JOINT. KROLL'S REAGENT WAS USED AS ETCHANT.

Weld Macrostructure. The weld shown in Fig. 15 was made with the electron-beam welding process, using sharp focus, 100 kV power, 16 to 25 mA current, and a travel speed of 12.7 mm/s (30 in./min). As can be seen in the transverse section, a full penetration weld was achieved using one pass from each side of the weld. The transverse section also shows that no obvious discontinuities exist in the weld pool and that the weld metal microstructure is fairly uniform. Figure 15 also suggests that the tantalum shim was completely melted.

Weld Metal Composition. An initial examination of the weld metal composition was performed to help assess the uniformity of the weld macrostructure. Figure 16 shows tantalum, titanium, vanadium, and aluminum elemental scan lines traversing the weld. These data were generated using a scanning electron microscope with energy-dispersive spectroscopy capability, and they confirm that the elements in question are uniformly distributed throughout the weld pool. Based on these analyses and the macrostructural characterization, it was determined that this combination of welding processes and parameters was successful in producing a full-penetration weld and a macroscopically homogeneous and defect-free weld pool.

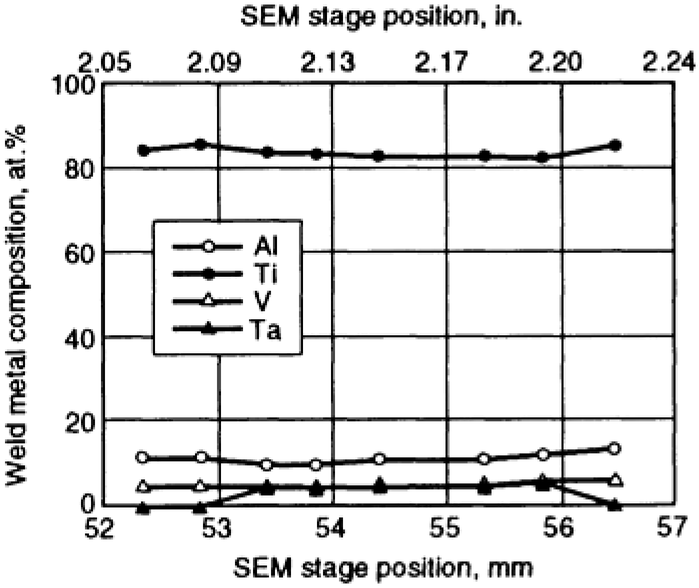

FIG. 16 ELEMENTAL LINE SCANS OF TITANIUM, ALUMINUM, VANADIUM, AND TANTALUM SPANNING THE HAZ AND FUSION ZONE OF THE 2.5 MM (0.100 IN.) THICK TI-6AL-4V SHEET WELDED USING A 0.127 MM (0.005 IN.) THICK TANTALUM SHIM.

The properties of titanium alloys can be strongly affected by the phases present in the microstructure, which are largely controlled by the weld metal composition. In this example, aluminum stabilizes the hexagonal close-packed (hcp) α and/or α' phases, and vanadium and tantalum, the body-centered cubic (bcc) β phase. In titanium alloys, the composition can be expressed as an electron/atom (*e*/*a*) ratio, which allows the phase stabilization effects of different elements to be combined. Table 6 lists the calculated (based on dilution) and measured compositions of the weld. This table shows that the measured and calculated compositions agree reasonably well. The measured aluminum values are somewhat lower than the calculations predict, perhaps due to some loss of aluminum during the welding process. Of most interest is that the measured and calculated *e*/*a* ratios are 3.96 to 3.98. This compositional range indicates that the weld metal microstructure should be α', a hexagonal martensite structure associated with good mechanical properties (Ref 20).

TABLE 6 COMPOSITION OF TI-6AL-4V RELATIVE TO THAT OF TI-6AL-4V PLUS TANTALUM WELD METAL

SAMPLE	COMPOSITION, AT.%				
	Ti	V	Al	Ta	*E/A* RATIO
TI-6AL-4V	86.2	3.6	10.2	. . .	3.93
TI-6AL-4V WELDED TO TA SHIM:					
CALCULATED	83.9	3.5	9.9	2.7	3.96
MEASURED	83.0	3.9	9.5	3.6	3.98

Weld Microstructure. The microstructure of the weld was analyzed using optical microscopy, transmission electron microscopy, and x-ray diffraction. Of particular interest were the structure of any martensitic phase present and the possible existence of ω, an ordered phase that can severely embrittle the material. The composition of the weld metal suggests that the structure will be α' hexagonal martensite. Figure 17 shows an optical micrograph of the fusion zone. Examination of Fig. 17 and comparison with Ti-6Al-4V (no tantalum shim) welds further suggest that the weld metal microstructure is α'. This was further supported by x-ray diffraction and TEM data. Figure 18 shows an x-ray diffraction trace for the weld metal. Aside from peaks from the specimen-mounting material, only peaks associated with a hexagonal structure were detected.

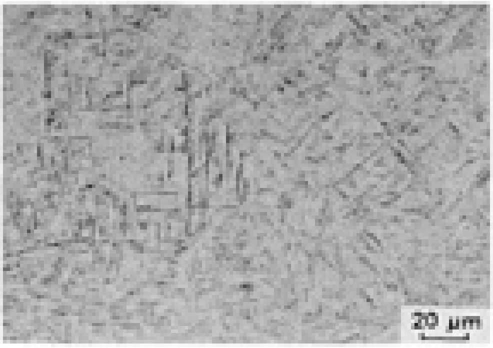

FIG. 17 OPTICAL MICROGRAPH OF FUSION ZONE OBTAINED WITH THE TI-6AL-4V SHEET WELDED USING A TANTALUM SHIM. KROLL'S REAGENT WAS USED AS THE ETCHANT. MICROSTRUCTURE CONSISTS OF WIDMANSTÄTTEN α+ β AND α' MARTENSITE.

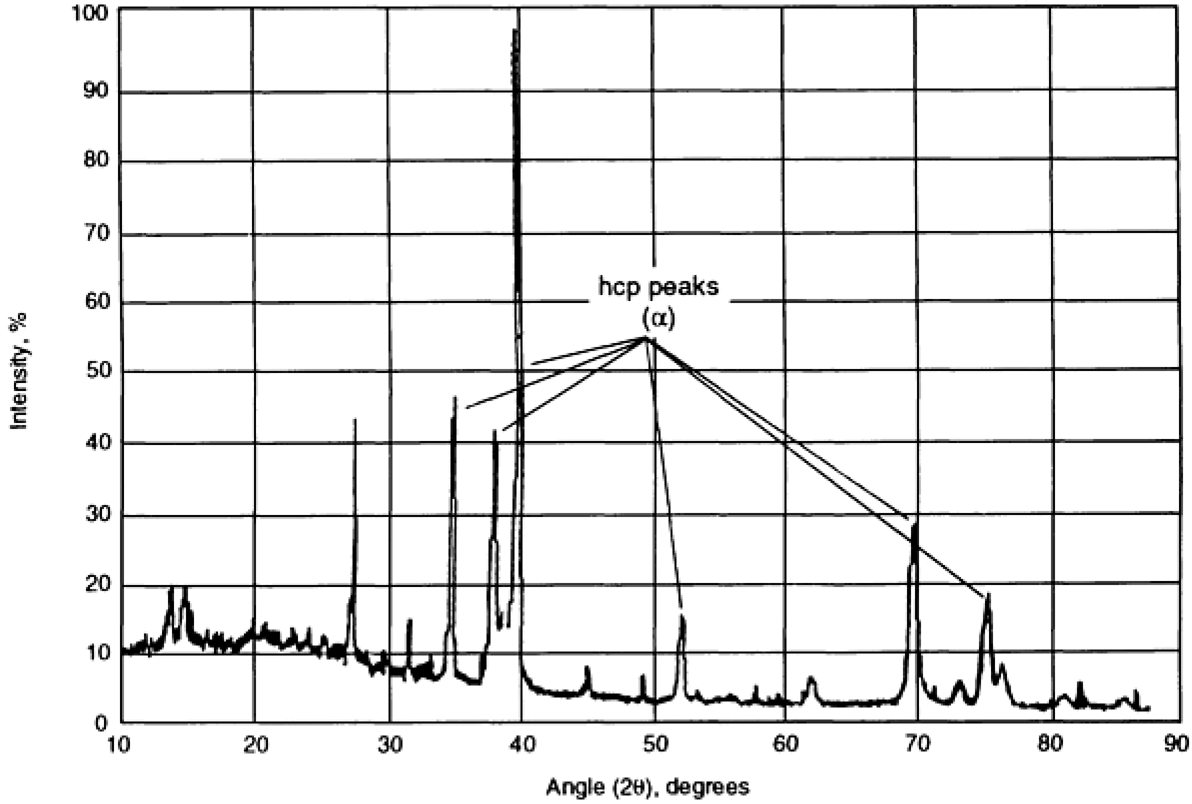

FIG. 18 X-RAY DIFFRACTION TRACE OF FUSION ZONE IN TI-6AL-4V SHEET WELDED USING A TANTALUM SHIM, SHOWING PEAKS THAT IDENTIFY HEXAGONAL STRUCTURES

Based on the compositional analysis, diffraction results, and microstructural analysis, the weld metal can be characterized as α' hexagonal martensite. By comparison with Ti-6Al-4V base metal and (no shim) weld metal, it was determined that the weld metal could be treated as a slightly more heavily stabilized variant of Ti-6Al-4V. This analysis allowed its properties and performance to be reasonably predicted.

Mechanical Testing. Microhardness traverses and all-weld-metal tension tests were performed on this weld. Figure 19 shows the results of two microhardness traverses, one across the second pass only, and one across the middle of the weld where the two passes overlap. Both traces are very similar, and the hardness is fairly constant across the fusion zone. These data confirm that the weld metal is relatively homogeneous. The hardness of the fusion zone is slightly higher than that of the base metal, as would be expected because of the martensitic structure and slightly higher β-stabilizer content.

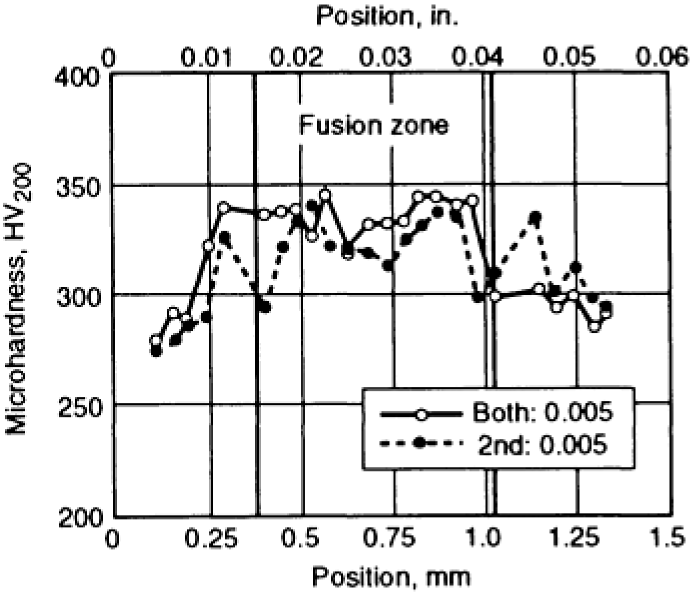

FIG. 19 MICROHARDNESS TRAVERSE DATA IN TWO LOCATIONS FOR THE TI-6AL-4V SHEET WELDED USING A TANTALUM SHIM. DATA LABELED "2ND" REFER TO TRAVERSE OVER SECOND PASS OF WELD ONLY. DATA LABELED "BOTH" INDICATE TRAVERSE ACROSS CENTER OF WELD WHERE TWO PASSES OVERLAP (SEE FIG. 15).

The tensile testing further confirmed that the weld metal could be considered to be a more heavily stabilized variant of Ti-6Al-4V. Table 7 compares the results of tensile tests performed on a baseline Ti-6Al-4V (no shim) weld with those of tests performed on the example weld. As the data show, the tantalum-containing weld had slightly higher strength and lower ductility than the baseline weld. Again these data are consistent with a slightly higher β-stabilizer content. However, the differences were not as great as in the microhardness testing, because in the case of the tensile tests, both structures were α' martensite.

TABLE 7 TENSILE PROPERTIES OF BASELINE TI-6AL-4V WELD COMPARED TO TI-6AL-4V WELDED USING TANTALUM SHIM

SAMPLE	YIELD STRENGTH		ULTIMATE TENSILE STRENGTH		YOUNG'S MODULUS		MAXIMUM STRAIN, %
	MPA	KSI	MPA	KSI	GPA	10^6 PSI	
BASELINE 1	828	120	970	141	114	16.5	12.7
BASELINE 2	787	114	960	139	127	18.4	12.3
BASELINE AVERAGE	807	117	965	140	121	17.5	12.5
TA WELD 1(A)	839	121.7	983	143	105	15.2	4.7
TA WELD 2(A)	838	121.5	971	141	103	14.9	10.5
TA WELD 3(A)	837	121.4	981	142	107	15.5	14.3
TA WELD AVERAGE	838	121.5	978	142	105	15.2	9.8

(A) REPLICANTS OF TI-6AL-4V WELDED TO TANTALUM SHIM

Test Results. The goal of the Ti-6Al-4V/tantalum dissimilar metal weld study was to thoroughly characterize the homogeneity, structure, and properties of the weld, relative to a baseline Ti-6Al-4V weld, and to determine the effect of the tantalum additions. The study showed that the tantalum shim could be completely melted using the chosen welding procedure and that a uniform weld bead could be produced. The compositional and microstructural analyses showed that the weld was very similar to the baseline weld. The tantalum-alloyed weld metal appeared to be consistent with standard principles regarding titanium alloy compositions and structure, and it was what would be expected in a weld made on slightly more heavily alloyed material. The mechanical test results confirmed this interpretation as well. In total, the characterization showed that this weld could be made and that its properties could be predicted and/or interpreted accurately as an extrapolation of the baseline weld.

Characterization Objectives. A number of characterization techniques have been briefly discussed with respect to their application to welds. Examples 1 and 2 show that the choice and emphasis of the characterization procedure are largely determined by the intended application of the weld and the goals to be accomplished by the characterization. Each of the procedures in Examples 1 and 2 are discussed in great detail in the cited references. Additional information is available in the Section "Joint Evaluation and Quality Control" in this Volume and in Ref 16, 17, and 22

References cited in this section

16. *MATERIALS CHARACTERIZATION,* 9TH ED., VOL 10, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1986
17. *MECHANICAL TESTING,* 9TH ED., VOL 8, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1985
20. STANDARD TEST METHOD FOR PLANE-STRAIN FRACTURE TOUGHNESS OF METALLIC MATERIALS, ASTM E 399, *ANNUAL BOOK OF ASTM STANDARDS,* ASTM, 1984
22. *WELDING, BRAZING, AND SOLDERING,* 9TH ED., VOL 6, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1983

Characterization of Welds

Craig B. Dallam, The Lincoln Electric Company; Brian K. Damkroger, Sandia National Laboratories

---

References

1. STANDARD WELDING TERMS AND DEFINITIONS, ANSI/AWS A3.0-89, AWS, 1989
2. WELD DISCONTINUITIES, *METALS HANDBOOK,* 9TH ED., VOL 6, AMERICAN SOCIETY FOR METALS, 1983, P 829-855
3. *WELDING HANDBOOK,* 8TH ED., VOL 2, AWS, 1987
4. *PROCEDURE HANDBOOK OF ARC WELDING,* 12TH ED., LINCOLN ELECTRIC CO., 1973
5. CODES, STANDARDS, AND INSPECTION, *METALS HANDBOOK,* 9TH ED., VOL 6, AMERICAN SOCIETY FOR METALS, 1983, P 823-838
6. *RADIOGRAPHY IN MODERN INDUSTRY,* 4TH ED., EASTMAN KODAK CO., 1980
7. WELDING OF PIPELINES AND RELATED FACILITIES, API STANDARD 1104, 17TH ED., AMERICAN PETROLEUM INSTITUTE, 1988
8. REFERENCE RADIOGRAPHS FOR STEEL WELDMENTS, ASTM E 390, ASTM
9. *STRUCTURAL WELDING CODE--STEEL,* ANSI/AWS D1.1-92, AWS, 1992
10. METALLOGRAPHY OF WELDMENTS, *METALS HANDBOOK,* 9TH ED., VOL 9, AMERICAN SOCIETY FOR METALS, 1985, P 577-586
11. K.E. EASTERLING, MICROSTRUCTURE AND PROPERTIES OF THE HAZ, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* S. DAVID AND J. VITEK, ED., ASM INTERNATIONAL, 1990, P 177-188
12. WRC 1992 DIAGRAM, *WELD. J.,* MAY 1992, P 171
13. T.G.F. GRAY, J. SPENCE, AND T.H. NORTH, *RATIONAL WELDING DESIGN,* NEWNES-BUTTERWORTHS, 1975, P 111
14. R.H. ESPY, WELDABILITY OF NITROGEN-STRENGTHENED STAINLESS STEELS, *WELD. J. RES. SUPPL.,* VOL 61 (NO. 5), MAY 1982, P 149S-156S

15. METAL TEST METHODS AND ANALYTICAL PROCEDURES, *1991 ANNUAL BOOK OF ASTM STANDARDS,* VOL 03.05 AND 03.06, ASTM, 1991
16. *MATERIALS CHARACTERIZATION,* 9TH ED., VOL 10, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1986
17. *MECHANICAL TESTING,* 9TH ED., VOL 8, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1985
18. STANDARD METHODS OF TENSION TESTING OF METALLIC MATERIALS, ASTM E 8, *ANNUAL BOOK OF ASTM STANDARDS,* VOL 03.01, ASTM, 1984
19. *SPECIFICATION FOR CARBON STEEL ELECTRODES FOR SHIELDED METAL ARC WELDING,* ANSI/AWS A5.1-91, AWS, 1991
20. STANDARD TEST METHOD FOR PLANE-STRAIN FRACTURE TOUGHNESS OF METALLIC MATERIALS, ASTM E 399, *ANNUAL BOOK OF ASTM STANDARDS,* ASTM, 1984
21. STANDARD TEST METHOD FOR CRACK TIP OPENING DISPLACEMENT (CTOD) FRACTURE TOUGHNESS MEASUREMENT, ASTM E 1290, VOL 03.01, ASTM, 1989
22. *WELDING, BRAZING, AND SOLDERING,* 9TH ED., VOL 6, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1983

Introduction to Brazing and Soldering

Melvin M. Schwartz, Sikorsky Aircraft

Introduction

BRAZING AND SOLDERING require the application of a number of scientific and engineering skills to produce joints of satisfactory quality and reliability. Brazing employs higher temperatures than soldering, but the fundamental concepts are similar, particularly with respect to metallurgy and surface chemistry (Table 1). However, joint design, materials to be joined, filler metal and flux selection, heating methods, and joint preparation can vary widely between the two processes. Economic considerations involving filler metal and process technology are also varied, particularly in relation to automated techniques and inspection and testing. Brazing and soldering are performed in many industries, from exotic applications in the electronics and aerospace field to everyday plumbing applications.

TABLE 1 COMPARISON OF SOLDERING, BRAZING, AND WELDING

Parameter	**Process**		
	Soldering	**Brazing**	**Welding**
Joint formed	Mechanical	Metallurgical	Metallurgical
Filler metal melt temperature, °c (°f)	<450 (<840)	>450 (>840)(a)	>450 (>840)(b)
Base metal	Does not melt	Does not melt	. . .
Fluxes used to protect and to assist in wetting of base-metal surfaces	Required	Optional	Optional
Typical heat sources	Soldering iron; ultrasonics; resistance; oven	Furnace; chemical reaction; induction; torch; infared	Plasma; electron beam; tungsten and submerged arc; resistance; laser
Tendency to warp or burn	Atypical	Atypical	Potential distortion and warpage of base-metal likely
Residual stresses	. . .	. . .	Likely around weld area

(A) LESS THAN MELTING POINT OF BASE METAL.

(B) LESS THAN OR EQUAL TO MELTING POINT OF BASE METAL

Introduction to Brazing and Soldering

Melvin M. Schwartz, Sikorsky Aircraft

Brazing

Many products are assembled from two or more individual components that are often permanently joined to produce structurally sound assemblies. Joining methods include various fasteners, interference-type joints, adhesives, and, for the highest-integrity joints, the many techniques classified under welding. Brazing, although fundamentally indifferent from welding, is one such technique.

Early metalworkers, stimulated by a desire to produce structures that were difficult or impossible to build using methods then in existence, realized that it was possible to fill the joint between two metal pieces with molten metal and allow it to solidify. These artisans soon learned by experience that, in order to achieve adherence, the metals to be joined and the brazing filler metal had to be kept free of oxides and the filler metal had to have a lower melting point, and, furthermore, that a given filler metal would not necessarily adhere to all metals. From these basic requirements, brazing grew into a craft whose practitioners were well versed in what to do and what not to do in order to produce sound joints.

Just as the technique of brazing developed empirically, so did the lower-melting point filler metals. Workers first used lead and tin solders as well as silver and copper-arsenic ores, which were readily available and had low melting points. Later, the alloy brass was developed and found to be more desirable for joining copper, silver, and steel structures, because it provided higher-strength joints and could withstand higher temperatures.

Early silversmiths, probably wanting to produce white solder joints for aesthetic reasons, melted brass and silver together and found it to have an even lower melting point than brass, good adherence, and better corrosion resistance. Although innumerable combinations of silver, copper, and zinc subsequently evolved, primarily to meet melting-point requirements, these silver-brass, brass, and lead-tin alloys were essentially the only brazing and soldering filler materials available for generations.

Advantages of Brazing. Brazing has many distinct advantages, including the following:

- ECONOMICAL FABRICATION OF COMPLEX AND MULTI-COMPONENT ASSEMBLIES
- SIMPLE METHOD TO OBTAIN EXTENSIVE JOINT AREA OR JOINT LENGTH
- JOINT TEMPERATURE CAPABILITY APPROACHING THAT OF BASE METAL
- EXCELLENT STRESS DISTRIBUTION AND HEAT-TRANSFER PROPERTIES
- ABILITY TO PRESERVE PROTECTIVE METAL COATING OR CLADDING
- ABILITY TO JOIN CAST MATERIALS TO WROUGHT METALS
- ABILITY TO JOIN NONMETALS TO METALS
- ABILITY TO JOIN METAL THICKNESSES THAT VARY WIDELY IN SIZE
- ABILITY TO JOIN DISSIMILAR METALS
- ABILITY TO JOIN POROUS METAL COMPONENTS
- ABILITY TO FABRICATE LARGE ASSEMBLIES IN A STRESS-FREE CONDITION
- ABILITY TO PRESERVE SPECIAL METALLURGICAL CHARACTERISTICS OF METALS
- ABILITY TO JOIN FIBER- AND DISPERSION-STRENGTHENED COMPOSITES
- CAPABILITY FOR PRECISION PRODUCTION TOLERANCE
- REPRODUCIBLE AND RELIABLE QUALITY CONTROL TECHNIQUES

Strong, uniform, leakproof joints can be made rapidly, inexpensively, and even simultaneously. Joints that are inaccessible and parts that may not be joinable at all by other methods often can be joined by brazing. Complicated assemblies comprising thick and thin sections, odd shapes, and differing wrought and cast alloys can be turned into integral components by a single trip through a brazing furnace or a dip pot. Metal as thin as 0.01 mm (0.0004 in.) and as thick as 150 mm (6 in.) can be brazed.

Brazed joint strength is high. The nature of the interatomic (metallic) bond is such that even a simple joint, when properly designed and made, will have strength equal to or greater than that of the as-brazed parent metal. The natural shapes of brazing fillets are excellent. The meniscus surface formed by the fillet metal as it curves across corners and adjoining sections is ideally shaped to resist fatigue. (It should be noted that in brazed joints using eutectic-type filler metal, fillets often contain an excessive amount of brittle intermetallic compounds. In fact, fillets are 5 to 10 times thicker than the joint and thus have a much higher volume of the liquid phase from which these brittle phases crystallize. Therefore, crack nucleation often originates in fillets.)

Complex shapes with greatly varied sections can be brazed with little distortion, and precise joining is comparatively simple. Unlike welding, in which the application of intense heat to small areas acts to move the parts out of alignment and introduces residual stresses, brazing involves fairly even heating and thus part alignment is easier.

Limitations of Brazing. A brazed joint is not a homologous body but rather is heterogeneous, composed of different phases with differing physical and chemical properties. In the simplest case, it consists of the base metal parts to be joined and the added filler metal. However, partial dissolution of the base metal, combined with diffusion processes, can change the composition and therefore the chemical and physical properties of the boundary zone formed at the interface between base metal and filler metal and often of the entire joint. Thus, in addition to the two different materials present in the simplest example given above, a complicated transitional or even completely different zone must be considered.

In determining the strength of such heterogeneous joints, the simplified concepts of elasticity and plasticity theory--valid for a homogeneous metallic body where imposed stresses are uniformly transmitted from one surface or space element to the adjacent ones--no longer apply. In a brazed joint formed of several materials with different characteristics of deformation resistance and deformation speed, the stresses caused by externally applied loads are nonuniformly distributed.

Mechanics of Brazing. Brazing involves a limited dissolution or plastic deformation of the base metal (Ref 1). Brazing comprises a group of joining processes in which coalescence is produced by heating to a suitable temperature above 450 °C (840 °F) and by using a ferrous or nonferrous filler metal that must have a liquidus temperature above 450 °C (840 °F) and below the solidus temperature(s) of the base metal(s). The filler metal is distributed between the closely fitted surfaces of the joint. Brazing is distinguished from soldering in that soldering employs a filler metal having a liquidus below 450 °C (840 °F) (Table 1).

Brazing proceeds through four distinct steps:

- THE ASSEMBLY OR THE REGION OF THE PARTS TO BE JOINED IS HEATED TO A TEMPERATURE OF AT LEAST 450 °C (840 °F).
- THE ASSEMBLED PARTS AND BRAZING FILLER METAL REACH A TEMPERATURE HIGH ENOUGH TO MELT THE FILLER METAL (FOIL, WIRE, PASTE, PLATINGS, ETC.) BUT NOT THE PARTS.
- THE MOLTEN FILLER METAL, HELD IN THE JOINT BY SURFACE TENSION, SPREADS INTO THE JOINT AND WETS THE BASE METAL SURFACES.
- THE PARTS ARE COOLED TO SOLIDITY, OR "FREEZE," THE FILLER METAL, WHICH IS HELD IN THE JOINT BY CAPILLARY ATTRACTION AND ANCHORS THE PARTS TOGETHER BY METALLURGICAL REACTION AND ATOMIC BONDING.

Brazing Versus Other Welding Processes. The mere fact that brazing does not involve any substantial melting of the base metals offers several advantages over other welding processes. It is generally possible to maintain closer assembly tolerances and to produce a cosmetically neater joint without costly secondary operations. Even more important, however, is that brazing makes it possible to join dissimilar metals (or metals to ceramics) that, because of metallurgical incompatibilities, cannot be joined by traditional fusion welding processes. If the base metals do not have to be melted to be joined, it does not matter that they have widely different melting points. Therefore, steel can be brazed to copper as easily as to another steel.

Brazing also generally produces less thermally induced distortion, or warping, than fusion welding. An entire part can be brought up to the same brazing temperature, thereby preventing the kind of localized heating that causes distortion in welding.

Finally, and perhaps most important to the manufacturing engineer, brazing readily lends itself to mass production techniques. It is relatively easy to automate, because the application of heat does not have to be localized, as in fusion welding, and the application of filler metal is less critical. In fact, given the proper clearance conditions and heat, a brazed joint tends to "make itself" and is not dependent on operator skill, as are most fusion welding processes.

Automation is also simplified by the fact that there are many means of applying heat to the joint, including torches, furnaces, induction coils, electrical resistance, and dipping. Several joints in one assembly often can be produced in one multiple-braze operation during one heating cycle, further enhancing production automation.

As noted in Table 1, essentially no melting of the base metal occurs in brazing; however, the temperatures involved can affect the properties of the metals being joined. For example, base metals whose mechanical properties were obtained by cold working may soften or undergo grain growth if the brazing temperature is above their recrystallization temperatures. Mechanical properties obtained by heat treatment may be altered by the heat of brazing. On the other hand, materials in the annealed condition are usually not altered by brazing.

As with other welding processes, brazing produces a heat-affected zone (HAZ) with a strongly altered microstructure due to intensive mutual mass transfer between base metal and filler metal. The width of this zone varies with the heating process used. In torch and induction brazing, for example, only a localized zone is heated; in france and dip brazing, the entire part is subjected to the brazing temperature. As a rule, the HAZ produced during brazing is wider and less sharply defined than those resulting from other fusion-related processes.

Several elements of the brazing process must be understood in order to produce satisfactory brazed joints:

- FILLER METAL FLOW
- BASE METAL CHARACTERISTICS
- FILLER METAL CHARACTERISTICS
- SURFACE PREPARATION
- JOINT DESIGN AND CLEARANCE
- TEMPERATURE AND TIME
- RATE AND SOURCE OF HEATING

Each of these topics is discussed in detail in the article "Fundamentals of Brazing" in this Volume.

Reference cited in this section

1. E. LIEBERMAN, *MODERN SOLDERING AND BRAZING TECHNIQUES,* BUSINESS NEWS PUBLICATIONS, 1988

Soldering

Like brazing and other joining processes, soldering involves several fields of science, including mechanics, chemistry, and metallurgy. Soldering is a simple operation (Ref 2), consisting of the relative placement of the parts to be joined, wetting the surfaces with molten solder, and allowing the solder to cool until it has solidified.

Soldering in the field of electronics is in many respects different from soldering in other branches of industry. Although the physical principles of all soldering (and brazing) processes are the same, the features specific to their use in electronics are so numerous that it is possible to speak of soldering in electronics as a separate subject.

Soldering Versus Other Joining Technologies. Soldering has several clear advantages over competitive joining techniques, such as welding or bonding with conductive adhesives:

- THE SOLDER JOINT FORMS ITSELF BY THE NATURE OF THE FLOW, WETTING, AND SUBSEQUENT CRYSTALLIZATION PROCESS, EVEN WHEN THE HEAT AND THE SOLDER ARE NOT DIRECTED PRECISELY TO THE PLACES TO BE SOLDERED. BECAUSE SOLDER DOES NOT ADHERE TO INSULATING MATERIALS, IT OFTEN CAN BE APPLIED IN EXCESS QUANTITIES, IN CONTRAST TO CONDUCTIVE ADHESIVES. THE SOLDERING TEMPERATURE IS RELATIVELY LOW, SO THERE IS NO NEED FOR THE HEAT TO BE APPLIED LOCALLY AS IN WELDING.
- SOLDERING ALLOWS CONSIDERABLE FREEDOM IN THE DIMENSIONING OF JOINTS, SO THAT IT IS POSSIBLE TO OBTAIN GOOD RESULTS EVEN IF A VARIETY OF COMPONENTS ARE USED ON THE SAME PRODUCT.
- THE SOLDERED CONNECTIONS CAN BE DISCONNECTED IF NECESSARY, THUS FACILITATING REPAIR.
- THE EQUIPMENT FOR BOTH MANUAL SOLDERING AND MACHINE SOLDERING IS RELATIVELY SIMPLE.
- THE SOLDERING PROCESS CAN BE EASILY AUTOMATED, OFFERING THE POSSIBILITY OF IN-LINE ARRANGEMENTS OF SOLDERING MACHINES WITH OTHER EQUIPMENT.

Mass soldering by wave, drag, or dip machines has been the preferred method for making high-quality, reliable connections for many decades. Despite the appearance of new connecting systems, it still retains this position. Correctly controlled, soldering is one of the least expensive methods for fabricating electrical connections. Incorrectly controlled, it can be one of the most costly processes--not because of the initial cost, but because of the many far-reaching effects of poor workmanship.

Chronology of Soldering. Soldering is not a new process (Ref 2). It probably dates to the Bronze Age, when some metalworker discovered the affinity of a tin-lead mixture for a clean copper surface. As recorded in the writings of Pliny, the Romans used 60Sn-40Pb mixture to solder their lead water pipes.

During the late 19th and early 20th centuries, a great deal of work was devoted to the development and improvement of the soldering process and to the determination of the fundamental principles involved. Examination of the books, papers, and especially the patents of the era shows that soldering as a science was clearly understood long before the start of the electronics industry. There has certainly been a great improvement in the materials and machines used in the process, but the basics have changed little since the early part of this century.

Soldering in the electronics field has undergone many dramatic changes since the first radios were built. Between the 1920s and the 1940s, all interconnections were made using the point-to-point wiring method, and the solder joints were made using soldering irons. After World War II (Ref 3, 4), the demand for consumer electronics began to increase, and by the 1950s a mass market had developed. To meet this demand, the first mass soldering technique, dip soldering, was used with boards that were the precursors of printed wiring boards (PWBs). By the late 1950s, wave soldering had been developed in England and soon thereafter reached the United States. This development was necessitated by the introduction of PWBs. Wave soldering continued to be the method of choice through the 1980s and is still widely used.

The development and maturation of the multi-layer PWB and the first major effort toward miniaturization occurred in the 1970s. The dual inline integrated circuit package replaced many discrete components. By the early 1980s, the pressure to increase component densities had grown to such a degree that a new technique for attachment was needed, and surface mounting was developed. Surface mount technology in turn required new ways to make solder joints, prompting the development of vapor phase, infrared, hot gas, and other reflow soldering techniques.

Soldering remains the attachment method of choice in the evolving electronics industry. It is the only metallurgical joining technique suitable for the mass manufacture of electronic and electrical interconnections within the temperature restrictions of electronic components and substrates.

Solder Metals. The solder metal business has changed dramatically, having moved from merely producing wire to producing bulk, preform, and paste in order to meet the needs of soldering techniques. Solder can be applied in several ways (Ref 2, 3):

- A MIXTURE OF SOLDER POWDER AND FLUX, KNOWN AS SOLDER PASTE (REF 5), CAN BE APPLIED TO THE JOINTS BY SCREENING, STENCILING, OR USING ONE OF THE MACHINES THAT APPLIES DOTS OF THE PASTE BY A SIMPLE PNEUMATIC SYSTEM. THIS METHOD IS ESPECIALLY USEFUL WHERE SURFACE-MOUNTED COMPONENTS ARE TO BE SOLDERED AND CANNOT BE PASSED THROUGH THE SOLDER WAVE.
- PARTS CAN BE PRETINNED, THAT IS, COATED WITH A LAYER OF SOLDER BY FLUXING AND DIPPING INTO A BATH OF THE MOLTEN METAL.
- SOLDER CAN BE OBTAINED IN THE SHAPE OF RINGS, WASHERS, OR TUBES, WHICH ARE PLACED ON OR ADJACENT TO THE PARTS TO BE JOINED. THESE SOLDER PREFORMS CAN BE OBTAINED WITH OR WITHOUT A FLUX COATING.

The two latter methods of providing solder for the joint are used in many areas of the electronics industry where machine soldering is not possible. Tiny solder rings are placed over long wire wrap pins on a mother board, where the gold plating of the pins in the wrapping area must not be coated with solder. Solder washers are fitted under bolts that must be soldered to a ground plane. Wires and tinned before soldering into a pretinned wiring lug. The possibilities are seemingly limitless (Ref 6).

Solder filler metals have also changed. The first electronic soldering was done with the 50Sn-50Pb solder material used by plumbers. By the 1950s, the choice of materials had switched to eutectic tin-lead solders to counter the problems created by a vibrating wave solder machine conveyor. Today a wide range of filler materials is used in the electronics industry. With the spread of electronics technology into every facet of life, the environmental and materials demands have grown substantially, requiring the use of such specialized materials as indium and gold-tin fillers.

Flux technology has also evolved (Ref 3, 5). As a result of the restraints of military and government specifications, the emphasis has been on the addition of new flux types rather than on the replacement of old ones. Extensive work has been done with organic acids, which enlarge the processing windows. Other categories include synthetic activated materials, which are non-rosin fluorocarbon solvent-removable types, and low solids fluxes for surface mounting that can be used instead of the classic fully activated rosin formulations.

Other materials, such as adhesives for the surface mounting of components intended for wave soldering, are new to soldering technology. These materials must be specially designed to avoid interference with the soldering process, yet must remain intact through the solvent action of fluxes and, often, flux removal.

Mass Soldering. With the introduction of mass soldering techniques (Ref 1, 6), the issue of solderability has become one of the major interest. Mass production processes are not as tolerant as the hand soldering process. In addition, modern components cannot withstand as much heat, as for example, a tube socket. Lower soldering temperatures and less mechanical working have made it necessary to attain the highest solderability to ensure a metallurgically sound and reliably solder joint.

Machine or mass soldering (Ref 2, 4, 6) can make thousands of joints in a few seconds, providing the electrical connections and simultaneously mechanically fastening the components. No other process has yet been developed that can so economically interconnect electronic circuitry.

Wave Soldering. From the original concept of hand dipping boards came the idea of pumping solder through a slot or nozzle to form a wave, or mound, of constantly moving, dross-free metal through which the board could be passed. This offered several advantages: the surface of the solder was always clean and did not require skimming, the pumping action maintained an even temperature, and the raised wave provided an easy form for adding a conveyor system for automatic movement of the board over the solder.

Once the idea of the solder wave had been born, it was logical to use a similar method to apply the flux, and the basic soldering machine appeared. These original machines were quite crude, consisting of individual modules with a freestanding conveyor (often added by the user) placed over them. There was little attempt to integrate the parts into one system, and the smoke and fumes were left to escape into the factory exhaust, aided by a makeshift canopy hung over the machine.

Much work was done on the development of the actual wave in an attempt to improve the speed of soldering and to reduce the incidence of shorts and icicles. The solder weir, the double wave, and the cascade all experienced brief popularity. The cascade wave, which consists of several waves in series, was one of the more interesting of these developments. It was used with an inclined conveyor and was the basis of machines built for in-house use by several manufacturers. However, it never achieved great popularity.

Another development was the use of ultrasonics to form a wave, with the intention of causing cavitation at the joint of avoiding the use of flux, or at least reducing the activity required. Some experimental machines were built, but the cost was excessive, the soldering results were no better than those produced with the conventional pumped wave, and the power of the ultrasonic energy had to be so high that the matching horns used to energize the pot were quickly eroded away. With the advent of the semiconductor, ulrasonics were found to cause occasional damage to the devices. Today the main use for ultrasonics in the soldering industry is to be found in the pots used for tinning components prior to assembly. This era also saw the introduction of oil into the solder wave. This idea worked extremely well and has been a feature of some soldering machines for many years (Fig. 1) (Ref 2, 3, 4, 5, 6, 7).

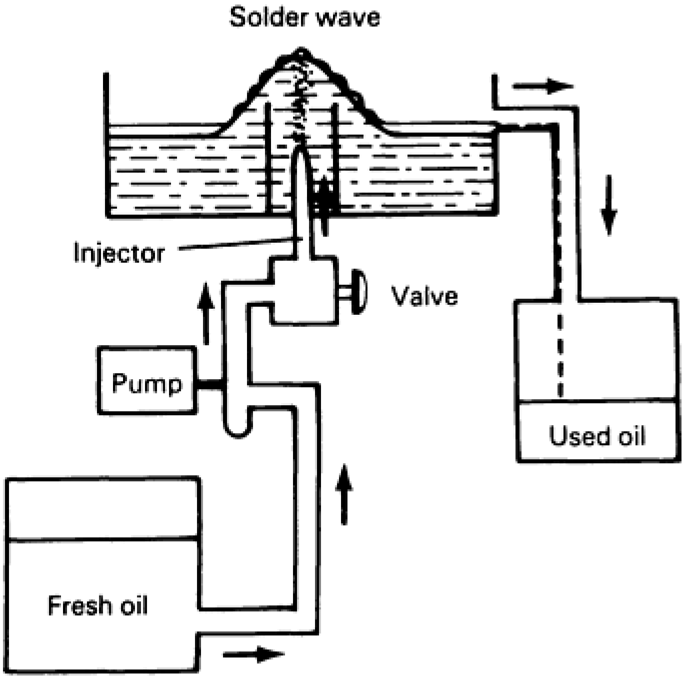

FIG. 1 SCHEMATIC SHOWING SIMPLE OIL INJECTION OF A SOLDER WAVE

Drag Soldering. In a totally different direction, the static solder bath soldering process, or drag soldering, was developed, and machines were designed using this principle (Fig. 2). These machines have been developed to a high level of sophistication, initially in Europe and Japan and later in the United States. As has often been the case in the growth of the soldering machine industry, initial developments were prompted because of patent restrictions that at the time prevented the free use of the more conventional solder wave. These restrictions have often proved beneficial to the industry by forcing creative thinking that has eventually resulted in improved products (Ref 2, 3, 4, 5, 6).

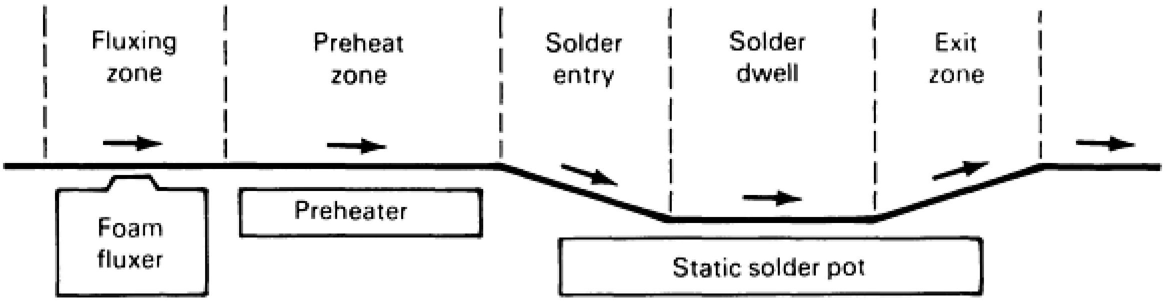

FIG. 2 SCHEMATIC SHOWING SEQUENCE OF OPERATIONS FOR A DRAG SOLDERING MACHINE

Current Technology. Today machines are available from many different manufacturers, each with its own particular features. These machines have various levels of sophistication and capability, ranging from small laboratory units to high-production systems, often complete with in-line cleaning and assembly and return conveyors. Machines in use today are generally self-contained, with a high degree of control, optional auxiliary systems, and exhaust venting. Coming onto the market are systems with computer-controlled operation and setup, which practically eliminate human error from the machine soldering process.

Soldering processes have evolved dramatically since the introduction of mass soldering techniques. Wave machines have become highly specialized, especially since the introduction of surface-mount technology. Reflow techniques (Ref 8), which were once limited to hybrid technology and leaded flatpacks, have grown remarkably. Vapor phase, infrared, and hot gas soldering equipment have initiated a new process of evolution.

The performance of a solder joint in its service environment is a strong consideration in modern electronics. Many applications, such as jet engines, dishwashers, and automobiles, impose a repeated cyclic thermal cycle on the equipment. This produces thermomechanical strains, which result in fatigue or creep failure if the joint design is poor or if life predictions have been incorrectly calculated. Each situation must be carefully analyzed and the correct design employed for the equipment to fulfill its intended service life.

Thermally induced fatigue and creep failures, virtually unknown in the 1950s but common today, are illustrative of the added stresses and strains placed on joints. It should be emphasized that miniaturization of electronic parts, particularly current paths, results in a dramatic increase in current density, which generates strong thermal and mechanical shocks upon switching. Proper design can minimize these strains and make the joints more highly producible. Design has become even more critical with the introduction of surface-mount technology.

Quality Control. Once a joint is made, its quality must be assessed through inspection. Advanced technology has produced some electronic assembly designs that are impossible to inspect visually, and an entire group of automated inspection tools has thus evolved. These tools include radiography, infrared signature, and two- or three-dimensional vision systems. Most of these are aided with very sophisticated computer software. The role of inspection, which historically was to locate those joints that needed rework, is changing along with the technology. Inspection is now used as a process-control audit to locate those portions of the process that need improvement or should returned to a position within the control limits. Therefore, the emphasis has shifted from merely shipping an acceptable product to manufacturing it correctly the first time.

Future Trends. Several other technologies and materials for soldering are being developed and tested (Ref 2, 3, 4, 5). One process that is becoming widely used is condensation, or vapor phase, soldering. In this process, the vapor from a liquid with a high-temperature boiling point is allowed to condense on the items to be soldered and, in giving up the latent heat of vaporization, quickly and evenly raises their temperature and produces a sound soldered joint. The method is simple, clean, and extremely well controlled. It requires-special equipment, and, of course, the parts are all exposed to the same temperature.

In another method, the parts are dipped into an oil heated above the melting point of the solder. Peanut oil can be used, although several excellent synthetic products are available. Here again, the component parts are subjected to the same temperature as the joint. This is a messy process, and the parts must be cleaned after soldering to remove the oil.

Infrared heating is a useful method for producing the temperature necessary for soldering. The shorter wavelengths can be focused onto very small areas, allowing selective heating of the joints to be soldered, without subjecting nearby components to the soldering temperature. Equipment and tooling are expensive if complex assemblies have to be soldered, and the relative emissivity of the joint metals and the adjacent materials can cause processing problems. For example, the base laminate of a PWB can be burned before the joint to be soldered has even reached the soldering temperature. This process thus requires very careful control.

Resistance heating can be used for soldering in two ways. The first method passes a current through the parts of the joint, causing their temperature to rise by virtue of the resistance existing in the joint structure. This is not a very controllable process and is primarily used for less critical applications on large connections. The second method passes a current through a wire that is shaped to touch the parts to be heated. Despite the lack of precise control, the process is often used to solder the leads on surface-mounted components.

Lasers have been used to provide the heat for soldering, although this is not yet a common method. A laser can provide a very high-energy, small-area beam, and heating is therefore extremely fast, localized, and limited only by the ability of the joint materials to absorb this energy. However, the shiny metal surfaces in the joint reflect most of the energy of the commonly used lasers, and coupling the laser and the joint is a major problem. The soldering flux is often used as a coupling medium. When correctly set up, laser soldering offers a fast, clean method of making joints. The cost of the equipment is quite high.

Since 1980, significant improvements and innovations in the soldering arena with respect to materials, equipment, and processes have occurred. However, two main tasks remain: developing environmentally friendly solder materials and establishing failure-free solder joint technology (Ref 2, 3, 4, 6, 7, 8).

For solders and soldering, two issues that are directly related to environmental concerns are the replacement of chlorofluorocarbon (CFCs) as the primary cleaning agent for electronic assemblies and the feasibility of using alternatives to lead-containing solders. For electronic assembly solders (Ref 4), in addition to using CFC substitutes and other solvents (e.g., hydrocarbon-based systems), aqueous cleaning and "no-clean" setups are two viable solutions.

The other task to be accomplished is the development of failure-free solder joint technology. Strengthened solder materials are expected to possess the properties that impede the occurrence of undesirable metallurgical phenomena during their service life under various in-circuit and external conditions, particularly temperature and induced stresses. Three approaches are being examined: macroscopic blends of selected fillers in a tin-rich matrix (composite solder), microscopic incorporation of doping additives to tin-rich base material, and atomic-level interaction. Preliminary experiments have been carried out to accomplish the first approach. Figure 3 demonstrates the increased yield strength of the 63Sn-37Pb composite solder with various amounts of filler (Ref 9).

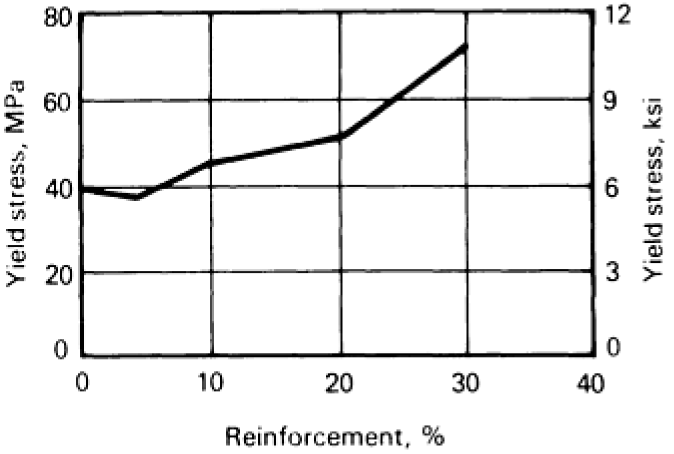

FIG. 3 EFFECT OF FILLER REINFORCEMENT ON THE YIELD STRENGTH OF 63SN-37PB COMPOSITE SOLDER. SOURCE: REF 9

The intrinsic properties of the new materials are expected to possess the following characteristics, while maintaining thermal and electrical conductivities:

- STABLE MICROSTRUCTURE
- INCREASED THERMAL AND MECHANICAL FATIGUE RESISTANCE
- HIGH CREEP RESISTANCE
- DECREASED THERMAL EXPANSION COEFFICIENT

These aggregate characteristics constitute the "strengthened" solder materials that are able to withstand harsher application conditions. The reduced thermal expansion coefficient of these new solders is expected to be more compatible with that of state-of-the-art electronic components and with ceramic substrates.

With respect to quality, solder joints have historically been judged by visual criteria. However, due to the nature of solder materials and the effect of the solder joint-forming process on appearance, many serviceable solder joints are often rejected, while other potentially questionable joints are accepted to carry out critical missions. This not only adds to manufacturing costs, but can also jeopardize service performance. Strengthened solder joints are expected to relieve the necessity of on-line inspection and rework while also reducing manufacturing costs and enhancing life span and tolerance of harsh service conditions in many military applications.

References cited in this section

1. E. LIEBERMAN, *MODERN SOLDERING AND BRAZING TECHNIQUES,* BUSINESS NEWS PUBLICATIONS, 1988
2. R.J. KLEIN WASSINK, *SOLDERING IN ELECTRONICS,* ELECTROCHEMICAL PUBLICATIONS, 1989
3. H.H. MANKO, *SOLDERS AND SOLDERING,* 2ND ED., MCGRAW-HILL 1964
4. L.P. LAMBERT, *SOLDERING FOR ELECTRONIC ASSEMBLIES,* MARCEL DEKKER, 1987
5. H.H. MANKO, SOLDERING FLUXES--PAST AND PRESENT *WELD. J.,* VOL 52 (NO. 3), 1973, P 163-166
6. R.W. WOODGATE, *THE HANDBOOK OF MACHINE SOLDERING,* 2ND ED., JOHN WILEY & SONS, 1988
7. T.J. NEILLO, "AN APPROACH TO SOLDERING IN THE AUTOMATED FACTORY," SME SOLDERING TECH. CONF., SME MS 91-288, MAY 1991
8. T. ROBERTS, "CONDUCTIVE REFLOW SOLDERING," SME SOLDERING TECH. CONF., SME MS 91-290, MAY 1991
9. J.S. HWANG, *ET AL.*, FUTURISTIC SOLDERS, *SURF. MOUNT. TECHNOL.*, SEPT 1991, P 40-43

Introduction to Brazing and Soldering

Melvin M. Schwartz, Sikorsky Aircraft

References

1. E. LIEBERMAN, *MODERN SOLDERING AND BRAZING TECHNIQUES,* BUSINESS NEWS PUBLICATIONS, 1988
2. R.J. KLEIN WASSINK, *SOLDERING IN ELECTRONICS,* ELECTROCHEMICAL PUBLICATIONS, 1989
3. H.H. MANKO, *SOLDERS AND SOLDERING,* 2ND ED., MCGRAW-HILL 1964
4. L.P. LAMBERT, *SOLDERING FOR ELECTRONIC ASSEMBLIES,* MARCEL DEKKER, 1987
5. H.H. MANKO, SOLDERING FLUXES--PAST AND PRESENT *WELD. J.,* VOL 52 (NO. 3), 1973, P 163-166
6. R.W. WOODGATE, *THE HANDBOOK OF MACHINE SOLDERING*, 2ND ED., JOHN WILEY & SONS,

1988
7. T.J. NEILLO, "AN APPROACH TO SOLDERING IN THE AUTOMATED FACTORY," SME SOLDERING TECH. CONF., SME MS 91-288, MAY 1991
8. T. ROBERTS, "CONDUCTIVE REFLOW SOLDERING," SME SOLDERING TECH. CONF., SME MS 91-290, MAY 1991
9. J.S. HWANG, *ET AL.*, FUTURISTIC SOLDERS, *SURF. MOUNT. TECHNOL.*, SEPT 1991, P 40-43

Fundamentals of Brazing

Mel M. Schwartz, Sikorsky Aircraft

Introduction

BRAZING is a process for joining solid metals in close proximity by introducing a liquid metal that melts above 450 °C (840 °F). A sound brazed joint generally results when an appropriate filler alloy is selected, the parent metal surfaces are clean and remain clean during heating to the flow temperature of the brazing alloy, and a suitable joint design is used. Like the other joining processes, brazing encompasses a variety of scientific disciplines (for example, mechanics, physics, and chemistry).

Recently, the demands of more-sophisticated structures have forced technicians and engineers to encourage metal producers to apply their metallurgical knowledge to produce brazing filler metals that meet more-specific needs. To ensure the production of good brazed joints, the technicians and engineers also had to appeal to mechanical engineers for improved joint design, to chemical engineers for solutions to corrosion problems, and to metallurgists and ceramists for proper material selection.

Brazing has been embraced by the engineering community and has now reached a very successful plateau within the joining field. This has come about because of the:

- DEVELOPMENT OF NEW TYPES OF BRAZING FILLER METALS (RAPID SOLIDIFICATION AMORPHOUS FOILS AND TITANIUM-ADDED FILLER METALS FOR CERAMIC JOINING
- AVAILABILITY OF NEW FORMS AND SHAPES OF FILLER METALS
- INTRODUCTION OF AUTOMATION THAT HAS BROUGHT BRAZING PROCESSES TO THE FOREFRONT IN HIGH-PRODUCTION SITUATIONS
- INCREASED USE OF FURNACE BRAZING IN A VACUUM, AS WELL AS ACTIVE AND INERT-GAS ATMOSPHERES

Since the early 1980s, other developments, such as aluminum-clad foils for fluxless aluminum brazing, copper-nickel-titanium filler metals for brazing titanium and some of its alloys, cadmium-free silver filler metals, and vacuum-grade metal brazing foils, have evolved to be used in production applications.

Physical Principles

Capillary flow is the dominant physical phenomenon that ensures good brazements when both faying surfaces to be joined are wet by the molten filter metal. The joint must be properly spaced to permit efficient capillary action and coalescence. More specifically, capillarity is a result of the relative attraction of the molecules of the liquid to each other and to those of the solid. In actual practice, brazing filler metal flow characteristics are also influenced by dynamic considerations involving fluidity, viscosity, vapor pressure, gravity, and, especially, by the effects of any metallurgical reasons between the filler metal and the base metal.

Capillary attraction makes the brazing of leak-tight joints a simple proposition. In a properly designed joint, the molten brazing filler metal is normally drawn completely through the joint area without any voids gaps when processed in a protective atmosphere. Solidified joints will remain intact and gas will remain tight under heavy pressures, even when the joint is subjected to shock or vibrational type of loading. Capillary attraction is also the physical force that governs the action of a liquid against solid surfaces in small, confined areas.

The phenomena of wetting and spreading are very important to the formation of brazed joints. Other significant factors that also must be considered include the condition of the solid surface in terms of the presence of oxide films and their effects on wetting and spreading, surface roughness, alloying with between the brazing filler metal and base metal, and the extent to which alloying is affected by the thermodynamic properties of the brazing atmosphere. A number of studies (Ref 1, 2) have been conducted on surface activation, contact angle, equilibrium, and surface energies. Some of these alloy systems have a finite contact angle that is thermodynamically unstable, because the solid-vapor surface energy exceeds the sum of the liquid-solid surface energies, that is, an alloy system in which thermodynamics would predict complete spreading. In actual fact, spreading may or may not occur in this type of alloy system, and the rate of spreading can be markedly dependent on surface chemistry.

Wetting is, perhaps, best explained by the following example. If a solid is immersed in a liquid bath and wetting occurs, then a thin continuous layer of liquid adheres to the solid when it is removed from the liquid (Fig. 1). Technically, the force of adhesion between the solid and liquid (during the wetting process) is greater than the cohesive force of the liquid. In practical terms, with respect to brazing, wetting implies that the liquid brazing filler metal spreads on the solid base metal, instead of balling up on its surface. It has been demonstrated that wetting actually depends on a slight surface alloying of the base metal with the brazing filler metal. Lead, for example, does not alloy with iron and will not wet it. Tin, on the other hand, does form an alloy with iron, and, therefore, a tin-lead solder will wet steel.

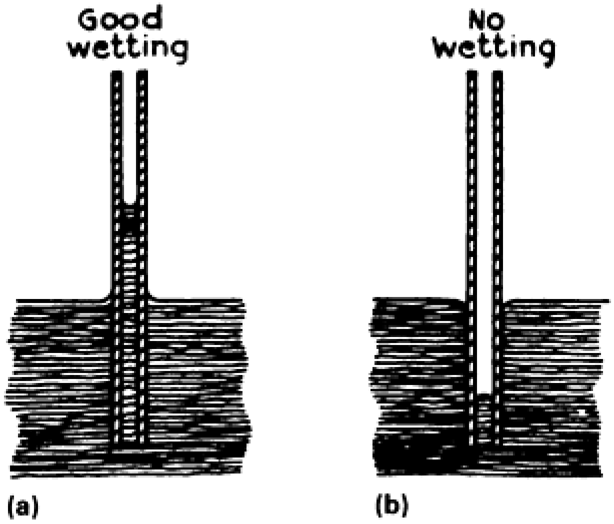

FIG. 1 SCHEMATIC SHOWING PRINCIPLE OF CAPILLARY ATTRACTION FOR SELECTED LIQUIDS WHEN THE LIQUID IS SANDWICHED BETWEEN TWO CLEAN GLASS PLATES. (A) WHEN IMMERSED IN WATER OR INK, A COLUMN WILL RISE BETWEEN THE PLATES BECAUSE OF WETTING. (B) WHEN THE PLATES ARE IMMERSED IN MERCURY, NO WETTING OCCURS AND THE COLUMN IS DEPRESSED. IF PARAFFIN PLATES ARE USED IN PLACE OF THE GLASS PLATES, IMMERSION OF THE PLATES IN MERCURY WILL PRODUCE WETTING.

A comprehensive theory of the wetting or spreading of liquids on solid surfaces and a complete and detailed derivation of the quantitative relationships are presented by Harkins (Ref 3) and Schwartz (Ref 4).

Experimentally, it has been observed that liquids placed on solid surfaces usually do not completely wet, but, rather, remain as a drop that has a definite contact angle between the liquid and solid phases (Ref 5). This condition is illustrated in Fig. 2. The Young and Dupré equation (Eq 1) permits the determination of change in surface free energy, ΔF, accompanying a small change in solid surface covered, ΔA (Ref 5). Thus,

$$\frac{\partial F}{\partial A}_{p,T} = g_{LV} + g_{SL} - g_{SV} \qquad \textbf{(EQ 1)}$$

where γ_{LV} is the liquid-vapor surface energy, γ_{SL} is the solid-liquid surface energy, and γ_{SV} is the solid-vapor surface energy.

$$\Delta F = \Delta A\ (\gamma_{SL} - \gamma_{SV}) + \Delta A \gamma_{LV}\ \mathrm{COS}\ (\theta - \Delta\theta) \qquad \textbf{(EQ 2)}$$

At equilibrium,

$$\lim_{\Delta A \to \Delta 0} \frac{\Delta F}{\Delta A} = 0 \qquad \textbf{(EQ 3)}$$

and

$$\gamma_{SL} - \gamma_{SV} + \gamma_{LV}\ \mathrm{COS}\ \theta = 0 \qquad \textbf{(EQ 4)}$$

or

$$\gamma_{SL} = \gamma_{SV} - \gamma_{LV}\ \mathrm{COS}\ \theta \qquad \textbf{(EQ 5)}$$

In Eq 4 and 5, it can be seen that θ is greater than 90° (1.6 rad) when γ_{SL} is larger than γ_{SV}, as shown in Fig. 3, and the liquid drop tends to spheroidize.

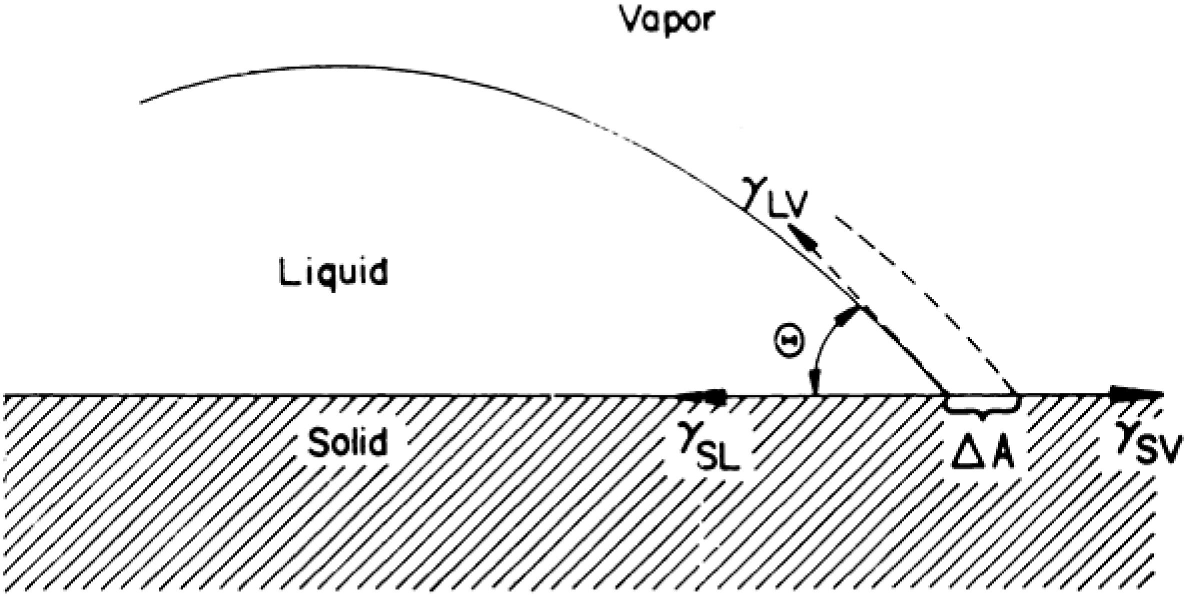

FIG. 2 SESSILE DROP LOCATED ON A SURFACE AND DIRECTIONS OF VECTORS OF THE SURFACE ENERGIES FOR THE SYSTEM

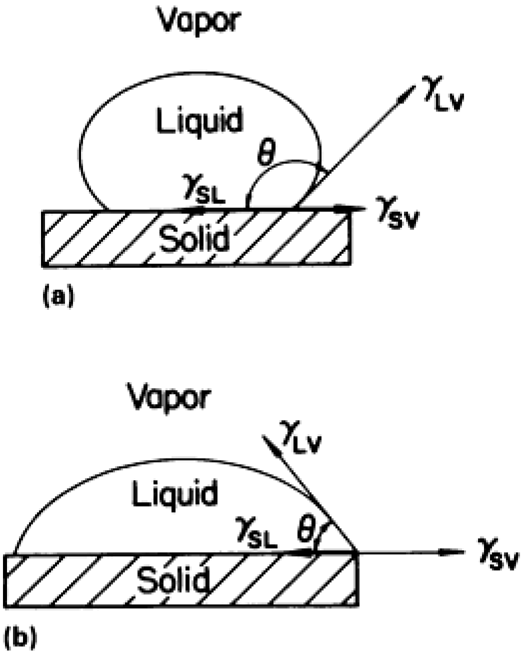

FIG. 3 EFFECT OF INTERFACIAL ENERGIES ON SESSILE DROPS. (A) NO WETTING TAKES PLACE WHEN CONTACT ANGLE IS > 90°. (B) WETTING OCCURS WHEN CONTACT ANGLE IS < 90°.

The contact angle, θ, is less than 90° (1.6 rad) when the reverse is true, as shown at right in Fig. 3, and the liquid drop flattens out and wets the solid. If the balance in such that θ is zero (0 rad) and greater wetting is desired, θ should be as small as possible so that cos θ approaches unity and the liquid spreads over the solid surface.

These considerations show the importance of surface energies in brazing. If brazing filler metal is to successfully form a joint, it must wet the solid material. The surface-energy balance must be such that the contact angle is less than 90° (1.6 rad). The energy equation shows that if θ is to be less than 90° (1.6 rad) ($\cos \theta > 0$), then γ_{SV} must be greater than γ_{SL}. This is demonstrated in Table 1, which gives contact-angle measurements of several liquid metals on beryllium at various test temperatures in argon and vacuum.

TABLE 1 CONTACT ANGLES OF SELECTED LIQUID METALS ON BERYLLIUM AT VARIOUS TEST TEMPERATURES IN ARGON AND VACUUM ATMOSPHERES

SYSTEM	ATMOSPHERE	TEMPERATURE		CONTACT ANGLE, DEGREES, AT INDICATED ELAPSED TIME				
		°C	°F	0 s	15 s	45 s	90 s	150 s
AG	VACUUM	1010	1850	142.2	141.1	140.6	141.2	140.2
		1060	1940	135.4	135.6	131.4	133.2	127.9
	ARGON	1010	1850	148.1	148.9	150.7	147.3	143.3
		1060	1940	61.0	59.9	57.2	54.4	52.6
AU	VACUUM	1070	1958	58.6	58.0	57.2	55.7	54.6
	ARGON	1070	1958	57.3	56.6	57.9	56.9	55.9
CU	VACUUM	1108	2026	129.0	128.5	124.2	124.1	123.6
		1133	2071	126.6	122.5	121.8	120.0	120.0
	ARGON	1108	2026	EXTREME BASE-METAL SOLUTION				
		1133	2071	EXTREME BASE-METAL SOLUTION				
GE	VACUUM	1000	1832	145.5	145.3	144.3	143.7	143.6
		1050	1922	102.9	100.0	100.9	96.1	95.4
	ARGON	1000	1832	99.3	99.6	98.8	98.2	97.1

		1050	1922	89.2	89.2	88.9	87.1	87.1
PD-2.1BE	VACUUM	980	1796	77.0	74.0	73.2	70.8	69.9
		1030	1886	74.2	73.2	68.7	66.5	65.6
	ARGON	980	1796	77.2	76.5	76.0	75.0	72.7
		1030	1886	70.2	63.1	61.7	57.2	57.2
AL	VACUUM	710	1310	155.1	154.1	151.3	146.6	142.0
		760	1400	140.0	138.8	136.6	134.1	132.0
	ARGON	710	1310	146.2	139.6	130.5	120.1	116.6
		760	1400	108.2	106.6	102.3	100.0	99.9
ZR-5BE	VACUUM	1030	1886	88.0	85.6	83.9	81.8	80.9
		1080	1976	69.8	67.7	67.3	66.5	64.7
	ARGON	1030	1886	101.4	98.8	99.2	98.3	96.2
		1080	1976	83.6	84.2	84.9	82.4	79.7
TI-6BE	VACUUM	1075	1967	15.8	14.8	14.5	14.0	13.2
		1125	2057	14.2	13.4	12.1	11.2	10.7
	ARGON	1075	1967	13.4	12.0	11.9	9.7	8.6
		1125	2057	10.0	7.9	7.8	6.8	5.6

Although during the testing period minor fluctuations in angle were noted, the overall tendency was for the contact angles to decrease with increasing time at test temperature. This behavior is reflected in the data presented in Table 2. The fluctuations can be attributed to alloying effects, because as alloying progresses, composition and temperature variations change the interfacial energies.

TABLE 2 EFFECT OF TEST CONDITIONS ON INTERFACIAL ENERGIES

SYSTEM	ATMOSPHERE	TEMPERATURE		CONTACT ANGLE(A)		γ_{LV}, DYNES/CM	γ_{LS}, DYNES/CM	FINAL SPREADING COEFFICIENT S_{LS}, DYNES/CM	WORK OF ADHESION (W_{AD}), DYNES/CM
		°C	°F	DEGREES	RADIANS				
AL	VACUUM	710	1310	142.5	2.5	831	2520	-1490	176
		760	1400	132.0	2.3	726	2350	-1210	239
	ARGON	710	1310	116.7	2.0	787	2220	-1140	434
		760	1400	99.9	1.7	561	1960	-658	464
ZR-5BE	VACUUM	1030	1886	79.8	1.4	625	1940	-702	549
		1080	1976	64.6	1.1	520	1770	-427	614
	ARGON	1030	1886	97.0	1.7	488	1780	-405	576
		1080	1976	79.8	1.4	172	1690	-225	569
TI-6BE	VACUUM	1075	1967	13.2	0.2	492(A)	1380	-9.00	970
		1125	2057	10.8	0.2	277	1590	-4.00	548
	ARGON	1075	1967	8.6	0.2	424	1440	-1.00	844
		1125	2057	5.6	0.01	329(B)	1530	+3.00	660

(A) AVERAGE OF TWO TESTS.

(B) ONE TEST ONLY

These data provide a means of comparing the wettability of beryllium by various liquids. If one assumes wettability for contact angles less than 90° and nonwetting for angles greater than 90°, then these data permit qualitative separation of the systems into two classes: those with contact angles less than 90°, and those with contact angles greater than 90°. Gold, Pd-2.1Be, and Ti-6Be were found to wet beryllium at both temperatures. The latter two brazing filler metals, as well as silver and gold, did not wet when tested in argon atmospheres at the lower temperature. However, it was found that an increase of 50 °C (90 °F) resulted in wetting. Aluminum exhibited a nonwetting behavior at both temperatures and in both atmospheres. Copper was nonwetting at both temperatures in vacuum, but underwent extensive alloying with the solid beryllium at these temperatures in argon. From these data it appears that the binary alloys have the greatest tendency to wet beryllium (Ref 6).

Surface-energy calculations have been performed on beryllium-base metals brazed with Al, Zr-5Be, and Ti-6Be filler metals. The calculated spreading coefficient were negative for all system test conditions, except with Ti-6Be. Negative spreading coefficients indicated a very definite trend. The greater the system wettability, as indicated by lower contact angles and liquid-vapor surface tensions, the greater the numerical magnitude of the spreading coefficient.

Of the three systems used, Ti-6Be was observed to provide the best wetting of beryllium, followed by Zr-5Be. Although no basic significance can be given to the absolute values of the calculated surface-tension properties, they are useful in indicating the relative tendencies for these liquids to wet beryllium.

The above observations indicate that some liquid-metal systems will wet beryllium, whereas some system that should wet it do not, in actual practice. The problem is more complex than it seems at first.

Recent work in Japan (Ref 7) and the United States (Ref 8) has shown that the effects of the base-metal/filler-metal reaction on the wetting of silicon nitride (Si_3N_4) and titanium-added active filler metals vary with reaction-layer thickness and morphology, as well as with the composition of the reaction layer. Several results of thermodynamic studies on the formation of the compound by the interfacial reaction are reported (Ref 7, 9). Moreover, kinetic studies (Ref 7) were also performed to achieve a quantitative understanding of the formation of the reaction layer.

It can be concluded that wetting is the ability of the molten brazing filler metal to adhere to the surface of a metal in the solid state and, when cooled below its solidus temperature, to make a strong bond with that metal. Wetting is a function not only of the nature of the brazing filler metal (Table 3), but of the degree of interaction between materials (ceramics, and so on) to be joined. There is considerable evidence that in order to wet well, a molten metal must be capable of dissolving, or alloying with, some of the metal on which it flows.

TABLE 3 COMPOSITION OF SELECTED FILLER METALS USED FOR BRAZING APPLICATIONS

FILLER METAL	COMPOSITION	LIQUIDUS		SOLIDUS	
		°C	°F	°C	°F
SILVER	99.99AG	961	1762	961	1762
CUSIL	72AG-28CU	780	1436	780	1436
PALCUSIL 5	68AG-27CU-5PD	810	1490	807	1485
PALCUSIL 10	58AG-32CU-10PD	852	1566	824	1515
PALCUSIL 15	65AG-20CU-15PD	900	1650	850	1560
PALCUSIL 25	54AG-21CU-25PD	950	1740	900	1650
GAPASIL 9	82AG-9GA-9PD	880	1615	845	1555
NICUSIL 3	71.5AG-28.1CU-0.75NI	795	1465	780	1435
NICUSIL 8	56AG-42CU-2NI	893	1639	771	1420
T50	62.5AG-32.5CU-5NI	866	1591	780	1435
T51	75AG-24.5CU-0.5NI	802	1476	780	1435
T52	77AG-21CU-2NI	830	1525	780	1435
CUSILTIN 5	68AG-27CU-5SN	760	1400	743	1369
CUSILTIN 10	60AG-30CU-10SN	718	1324	602	1116
BRAZE 630	63AG-28CU-6SN-3NI	800	1472	690	1275
BRAZE 580	57AG-33CU-7SN-3MN	730	1345	605	1120
BRAZE 655	65AG-28CU-5MN-2NI	850	1560	750	1380
SILCORO 60	60AU-20AG-20CU	845	1550	835	1535
NIORO	82AU-18NI	950	1740	950	1740
PALNIRO 7	70AU-22NI-8PD	1037	1899	1005	1840
INCURO 60	60AU-37CU-3IN	900	1650	860	1580
SILCORO 75	75AU-20CU-5AG	895	1645	885	1625

NICORO 80	81.5AU-16.5CU-2NI	925	1695	910	1670
PALCUSIL 20	52AU-28CU-20PD	925	1695	875	1605
GOLD	99.99AU	1064	1947	1064	1947
PALNIRO 4	30AU-36NI-34PD	1169	2136	1135	2075
PALNIRO 1	50AU-25NI-25PD	1121	2050	1102	2016
TICUSIL	68.8AG-26.7CU-4.5TI	850	1560	830	1525
PALNICUSIL	48AG-18.9CU-10NI-22.5PD	1179	2154	910	1670
PALCO	65PD-35CO	1235	2255	1230	2245
INCUSIL15	62AG-24CU-15IN	705	1300	630	1165
INCUSIL 10	63AG-27CU-10IN	730	1345	685	1265
BAG-8A	71.8AG-28CU-0.2LI	760	1400	760	1400
BAG-19	92.5AG-7.3CU-0.2LI	890	1635	760	1400
BRAZE 071	85CU-7AG-8SN	986	1807	665	1230
BRAZE 852	85AG-15MN	970	1780	960	1760
NIORONI	73.8AU-26.2NI	1010	1850	980	1795
NICUMAN 23	67.5CU-23.5MN-9NI	955	1750	925	1695
PALSIL 10	90AG-10PD	1065	1950	1002	1836
PALNI	60PD-40NI	1238	2260	1238	2260

Wetting is only one facet of the brazing process. Another factor that affects wetting is the cleanliness of the surface to be wetted. Oxide layers inhibit wetting and spreading, as do grease, dirt, and other contaminants that prevent good contact between the brazing filler metal and the base metal. One of the functions of a flux the oxide layer on the joint area and thereby expose clean base metal.

Good wetting and spreading of the liquid filler metal on the base metal are necessary in brazing because the mechanics of the process demand that the filler metal be brought smoothly, rapidly, and continuously to the joint opening. If the conditions within the capillary space of the joint do not promote good wetting, then the filler metal will not be drawn into the space by capillary attraction.

Kim *et al.* (Ref 10) investigated the wettability and reactivity of pressureless sintered Si_3N_4 with powdered copper-titanium filler metal using sessile drop tests conducted in a vacuum where Si_3N_4 was bonded to itself and joint strength was evaluated by shear testing. The wettability of Si_3N_4 on copper-titanium base filler metals was improved greatly by adding up to 50 wt% titanium. However, the reaction-layer thickness was increased up to 10% and, thereafter, decreased up to 50 wt%.

Wettability improved with the increment of titanium concentration, and no linear proportionality was found between the reaction and the wettability. As titanium concentration increased, a continuous thin layer tended to form at the interface and markedly improved wettability.

The thickness of the reaction layer was not in proportion to the titanium concentration. For a copper-titanium filler metal with a low titanium concentration (less than 5 wt%), the initial reaction layer tended to form discontinuously, whereas the reaction layer thickness increased with titanium concentration. On the contrary, in the filler metal with a high titanium concentration (15 wt%) the reaction-layer thickness decreased as titanium concentration increased because a continuous thin reaction layer tended to form (Ref 10).

The shear strength of the Si_3N_4-Si_3N_4 joint using copper-titanium filler metal is influenced by reaction-layer thickness and morphology. Strength increased with the increment of titanium concentration, and was accompanied by improved wettability, because the tendency of the continuous thin reaction layer to form increased.

With the development of new steels, Keller *et al.* (Ref 11) undertook an extensive study of the wettability of commercial braze filler metals on type 304 stainless steel, the most common of the austenitic stainless steels; type 316 stainless steel, which is often substituted for type 304 when increased corrosion resistance is desired; and 21-6-9, a new high-manganese stainless steel that is an attractive alternative to types 304 or 316 for certain applications. Because the 21-6-9 steel has a very stable austenitic structure, it is attractive for cryogenic applications (Ref 12).

Because high wettability means that the thermocapillary attraction that fills the braze joint is strong, wettability is an important component of braze performance. The wettability index (WI) developed by Feduska (Ref 13) is used as the measure of wettability. The WI is defined as the area covered by the braze metal filler times the cosine of the contact angle between the braze and the base metal. Therefore, the higher the WI, the better the braze filler metal wets the base metal. It should be emphasized that the WI, as defined, is a relative measure, and depends on the volume of filler metal used (Table 4).

TABLE 4 WETTABILITY INDICES OF BRAZE FILLER METALS ON SELECTED STAINLESS STEEL BASE METALS

FILLER METAL	TEST TEMPERATURE		WETTABILITY INDEX FOR INDICATED STAINLESS STEEL BASE METAL		
	°C	°F	TYPE 316	21-6-9	TYPE 304
SILVER	975	1785	. . .	. . .	. . .
	1000	1830	0.008	. . .	0.015
	1050	1920	0.041	. . .	0.023
	1100	2010	0.053	. . .	0.020
	1150	2100	. . .	. . .	. . .
CUSIL	800	1470	. . .	0.013	. . .
	850	1560	. . .	0.007	. . .
	900	1650	0.022	0.016	0.003
	925	1695	0.032	. . .	. . .
	950	1740	0.037	. . .	. . .
PALCUSIL 5	800	1470	0.027	. . .	. . .
	850	1560	. . .	0.014	. . .
	900	1650	0.047	0.020	0.011
	950	1740	0.080	0.061	0.035
PALCUSIL 10	850	1560	0.035	0.020	0.015
	900	1650	0.057	0.050	0.025
	950	1740	0.107	0.101	0.062
PALCUSIL 15	900	1650	0.107	0.092	0.068
	950	1740	0.152	0.170	0.119
	1000	1830	0.754	0.263	0.212
PALCUSIL 25	950	1740	0.104	0.107	0.096
	975	1785	0.225	. . .	. . .
	1000	1830	. . .	0.283	0.226
GAPASIL 9	900	1650	0.023	. . .	. . .
	950	1740	0.068	. . .	. . .
	1000	1830	0.269	. . .	. . .
NICUSIL 3	800	1470	0.006	0.000	0.012
	850	1560	0.027	0.017	0.008
	900	1650	0.041	0.039	0.021
	925	1695	0.051	. . .	. . .
	950	1740	0.064	. . .	. . .
NICUSIL 8	850	1560	0.024	. . .	. . .
	900	1650	0.052	. . .	0.033
	950	1740	0.085	. . .	0.064
	1000	1830	. . .	. . .	0.229
T-50	850	1560	0.024	. . .	. . .
	900	1650	0.038	. . .	0.026

FILLER METAL	TEST TEMPERATURE		WETTABILITY INDEX FOR INDICATED STAINLESS STEEL BASE METAL		
	°C	°F	TYPE 316	21-6-9	TYPE 304
	950	1740	0.062	. . .	0.057
	1000	1830	. . .	. . .	. . .
T-51	800	1470	0.007	. . .	. . .
	850	1560	0.029	. . .	. . .
	900	1650	0.039	. . .	. . .
	950	1740	. . .	. . .	0.029
	1000	1830	. . .	. . .	0.082
T-52	850	1560	0.027	. . .	0.001
	900	1650	0.045	. . .	0.014
	950	1740	0.063	. . .	0.045
	1000	1830	. . .	. . .	0.090
CUSILTIN 5	850	1560	0.015	. . .	0.000
	875	1605	0.043	. . .	. . .
	900	1650	0.047	. . .	0.013
	950	1740	. . .	. . .	0.033
	1000	1830	. . .	. . .	. . .
CUSILTIN 10	750	1380	. . .	0.025	0.000
	800	1470	. . .	0.017	0.023
	825	1515	0.021	. . .	. . .
	850	1560	0.034	0.051	0.050
	875	1605	0.043	. . .	. . .
BRAZE 630	800	1470	0.014	0.037	0.023
	850	1560	0.046	0.065	0.024
	900	1650	0.064	. . .	. . .
BRAZE 580	750	1380	0.020	0.060	0.060
	800	1470	0.051	0.056	0.089
	850	1560	0.073	0.120	0.102
	875	1605	0.078	. . .	. . .
BRAZE 655	825	1515	0.080	. . .	. . .
	850	1560	0.110	0.037	0.074
	875	1605	0.116	. . .	. . .
	900	1650	. . .	0.137	0.124
SILCORO 60	850	1560	. . .	0.005	0.004
	900	1650	0.039	0.007	0.011
	925	1695	0.051	. . .	. . .
	950	1740	0.073	0.025	0.016
	1000	1830	. . .	0.055	0.037
NIORO	950	1740	. . .	0.000	0.000
	975	1785	0.049	. . .	. . .
	1000	1830	0.061	0.060	0.065
PALNIRO 7	1025	1875	. . .	. . .	. . .
	1050	1920	0.053	. . .	. . .
	1075	1965	0.073	. . .	. . .
INCURO 60	900	1650	0.010	. . .	. . .
	950	1740	0.025	. . .	. . .
	1000	1830	0.091	. . .	. . .
SILCORO 75	900	1650	0.006	. . .	. . .

FILLER METAL	TEST TEMPERATURE		WETTABILITY INDEX FOR INDICATED STAINLESS STEEL BASE METAL		
	°C	°F	TYPE 316	21-6-9	TYPE 304
	950	1740	0.057	. . .	. . .
	1000	1830	0.170	. . .	. . .
NICORO 80	950	1740	0.041	0.019	0.016
	1000	1830	0.163	0.126	0.070
	1050	1920	0.413	0.190	0.084
PALCUSIL 20	875	1610	0.061	. . .	. . .
	900	1650	0.110	. . .	. . .
	925	1695	0.122	. . .	. . .
GOLD	1070	1960	. . .	. . .	0.088
	1075	1965	0.087	. . .	. . .
	1100	2010	0.358	. . .	0.238
	1150	2100	. . .	. . .	0.355
PALNIRO 4	1175	2145	0.036	. . .	. . .
	1200	2190	0.061	. . .	. . .
	1225	2235	0.078	. . .	. . .
PALNIRO 1	1125	2055	0.041	. . .	. . .
	1150	2100	0.063	. . .	. . .
	1175	2145	0.075	. . .	. . .
TICUSIL	875	1605	0.032	. . .	. . .
	900	1650	0.057	. . .	. . .
	950	1740	0.083	. . .	. . .
PALNICUSIL	950	1740	0.025	. . .	0.038
	975	1785	0.213	. . .	. . .
	1000	1830	. . .	. . .	. . .
	1025	1875	0.362	. . .	. . .
	1050	1920	. . .	. . .	. . .
	1075	1965	0.336	. . .	. . .
	1100	2010	. . .	. . .	. . .
PALCO	1250	2280	0.000	. . .	. . .
	1275	2325	0.073	. . .	. . .
	1300	2370	0.159	. . .	. . .
INCUSIL 15	750	1380	. . .	. . .	. . .
	800	1470	. . .	. . .	. . .
	850	1560	. . .	. . .	. . .
INCUSIL 10	750	1380	. . .	. . .	. . .
	800	1470	. . .	. . .	. . .
	850	1560	0.008	. . .	. . .
BAG-8A	800	1470	0.005	. . .	. . .
	850	1560	0.008	. . .	. . .
	900	1650	0.032	. . .	. . .
	950	1740	0.045	. . .	. . .
BAG-19	950	1740	0.016	. . .	. . .
	1000	1830	0.016	. . .	. . .
	1050	1920	0.034	. . .	. . .
BRAZE 071	1000	1830	0.066	0.106	0.101
	1050	1920	. . .	0.125	0.101
BRAZE 852	1000	1830	0.039	0.037	0.038

FILLER METAL	**TEST TEMPERATURE**		**WETTABILITY INDEX FOR INDICATED STAINLESS STEEL BASE METAL**		
	°C	**°F**	**TYPE 316**	**21-6-9**	**TYPE 304**
	1050	1920	0.029	0.038	0.042
	1100	2010	0.043	0.037	0.020
NIORONI	1000	1830	0.068	. . .	. . .
	1025	1875	0.070	. . .	. . .
	1100	2010	0.116	. . .	. . .
NICUMAN 23	950	1740	0.026	. . .	. . .
	975	1785	0.099	. . .	. . .
	1000	1830	0.091	. . .	. . .
	1025	1875	0.091	. . .	. . .
PALSIL 10	1050	1920	0.113	. . .	. . .
	1075	1965	0.113	. . .	. . .
	1100	2010	0.126	. . .	. . .
PALNI	1225	2235	. . .	. . .	. . .
	1250	2280	0.068	. . .	. . .
	1275	2325	0.078	. . .	. . .
	1300	2370	0.107	. . .	. . .
PALMANSIL 5	950	1740	. . .	. . .	0.057
	1000	1830	. . .	. . .	. . .
	1050	1920	. . .	. . .	. . .
	1100	2010	. . .	. . .	. . .

The WIs for each braze filler metal at each temperature on type 316 stainless steel, type 304L stainless steel, and 21-6-9 stainless steel are provided in Table 4. Wetting indices greater than 0.05 are indicative of good performance during brazing, whereas WI values greater than 0.10 are indicative of excellent performance during brazing (Ref 13).

A comparison of the WI of the braze filler metals on the three stainless steels revealed these trends:

- FILLER METALS GENERALLY WET TYPE 316 STAINLESS STEEL BETTER THAN TYPE 304 STAINLESS STEEL AND GENERALLY WET TYPE 316 STAINLESS BETTER THAN 21-6-9 STAINLESS.
- THE DEGREE OF WETTING OF MOST BRAZE FILLER METALS ON 21-6-9 STAINLESS WAS EQUAL TO OR BETTER THAN IT WAS ON TYPE 304 STAINLESS (REF 12).
- THE IMPROVED WETTABILITY OF BRAZE FILLER METALS ON TYPE 316 STAINLESS IS BELIEVED TO BE DUE TO THE PRESENCE OF MOLYBDENUM IN THE SURFACE OXIDE, WHEREAS THE ENHANCED WETTING ON 21-6-9 STAINLESS VERSUS TYPE 304 STAINLESS IS BELIEVED TO BE DUE TO THE HIGHER MANGANESE CONTENT OF THE SURFACE OXIDE.

Essentially, the successful joining of components by the brazing process depends on the selected brazing filler metal having a melting point above 450 °C (840 °F) and wetting the base metal without melting it. Furthermore, the joint must be designed to ensure that the mating surfaces of the components are parallel and close enough together to cause capillary attraction.

Fundamentals of Brazing

Mel M. Schwartz, Sikorsky Aircraft

Elements of the Brazing Process

An engineer must consider reliability and cost when designing the braze joint. Joint strength, fatigue resistance, corrosion susceptibility, and high-temperature stability are additional concerns that determine the selection of joint design, braze filler materials, and processing parameters.

A careful and intelligent appraisal of the following elements is required in order to produce satisfactory brazed joints:

- FILLER-METAL FLOW
- BASE-METAL CHARACTERISTICS
- FILLER-METAL CHARACTERISTICS
- SURFACE PREPARATION
- JOINT DESIGN AND CLEARANCE
- TEMPERATURE AND TIME
- RATE AND SOURCE OF HEATING
- PROTECTION BY AN ATMOSPHERE OR FLUX

Filler-Metal Flow. As mentioned previously, wetting is only one important facet of the brazing process. A low contact angle, which implies wetting, is also necessary, but is not a sufficient condition itself for flow. Viscosity is also important. Brazing filler metals with narrow melting ranges that are close to the eutectic composition generally have lower viscosities than those with wide melting ranges. Thus, a high surface tension of liquid filler metal, a low contact angle, and low viscosity are all desirable.

Flowability is the property of a brazing filler metal that determines the distance it will travel away from its original position, because of the action of capillary forces. To flow well, a filler metal must not gain an appreciable increase in its liquidus temperature even though its composition is altered by the addition of the metal it has dissolved. This is important because the brazing operation is carried out at temperatures just above the liquidus of the filler metal.

The composition and surface energy of liquids and solids are assumed to remain constant. In real systems, however, these interactions occur (Ref 14):

- ALLOY FORMATION BETWEEN LIQUID AND BASE METAL
- DIFFUSION OF BASE METAL INTO BRAZING FILLER METAL
- DIFFUSION OF FILLER METAL INTO GRAINS OF BASE METAL
- PENETRATION OF FILLER METAL ALONG GRAIN BOUNDARIES
- FORMATION OF INTERMETALLIC COMPOUNDS

In practice, interactions are usually minimized by selecting the proper brazing filler metal; keeping the brazing temperature as low as possible, but high enough to produce flow; and keeping the time of brazing temperature short and cooling the brazed joint as quickly as possible without causing cracking or distortion. When diffusion brazing is desired, higher brazing temperatures and longer times at brazing temperatures are employed.

Base-Metal Characteristics. The base metal has a prime effect on joint strength. A high-strength base metal produces joints of greater strength than those made with softer base metals (other factors being equal). When hardenable metals are brazed, joint strength becomes less predictable. This is because more-complex metallurgical reactions between hardenable base metals and the brazing filler metals are involved. These reactions can cause changes in the base-metal hardenability and can create residual stresses (Ref 15).

In cases where different materials make up the assembly and gaps may open or close as heating proceeds to the joining temperature, the coefficient of thermal expansion becomes vitally important.

Several metallurgical phenomena influence the behavior of brazed joints and, in some instances, necessitate special procedures. These base-metal effects include alloying by brazing filler; carbide precipitation; stress cracking; hydrogen, sulfur, and phosphorus embrittlement; and oxidation stability.

The extent of interaction varies greatly, depending on compositions (base metal and brazing filler metal) and thermal cycles. There is always some interaction, except in cases of mutual insolubility.

The strength of the base metal has a profound effect on the strength of the brazed joint. Therefore, this property must be carefully considered when designing the joint to have specific properties. Some base metals also are easier to braze than others, particularly by specific brazing processes.

Filler-Metal Characteristics. The second material involved in joint structures is the brazing filler metal. Unfortunately, it cannot be chosen to provide a specific joint strength. Actually, strong joints can be brazed with almost any good commercial brazing filler metal if correct brazing methods and joint design are implemented.

Necessary characteristics of brazing filler metals are:

- PROPER FLUIDITY AT BRAZING TEMPERATURES TO ENSURE FLOW BY CAPILLARY ACTION AND TO PROVIDE FULL ALLOY DISTRIBUTION
- STABILITY TO AVOID PREMATURE RELEASE OF LOW-MELTING-POINT ELEMENTS IN THE BRAZING FILLER METAL
- ABILITY TO WET THE BASE-METAL JOINT SURFACE
- LOW VOLATILIZATION OF ALLOYING ELEMENTS OF THE BRAZING FILLER METAL AT BRAZING TEMPERATURES
- ABILITY TO ALLOY OR COMBINE WITH THE BASE METAL TO FORM AN ALLOY WITH A HIGHER MELTING TEMPERATURE
- CONTROL OF WASHING OR CONTROL OF EROSION BETWEEN THE BRAZING FILLER METAL AND THE BASE METAL WITHIN THE LIMITS REQUIRED FOR THE BRAZING OPERATION

It should be noted that the strength of the brazed joint is not directly related to the method of filler-metal melting. For example, if constructional metals are produced by vacuum melting, then there is a definite relationship between the vacuum-melting practice and the final strength of the ingot, bar, or rolled sheet. With a brazing filler metal, however, joint strength is dependent on joint design, brazing temperature, amount of brazing filler metal applied, location and method of application, heating rate, and many other factors that constitute the brazing technique.

The degree to which brazing filler metal interacts with and penetrates the base metal during brazing depends on the intensity of mutual diffusion processes that occur between both those materials. In applications that require strong joints for high-temperature, high-stress service conditions (such as turbine rotor assemblies and jet-engine components), it is generally wise to specify a brazing filler metal that has high diffusion and solution properties with the base metal. When the assembly is constructed of extremely thin base metals (as in honeycomb structures and some heat exchangers), good practice entails specifying a brazing filler metal that contains elements with a low-diffusion characteristic relative to the base metal being used. Diffusion, a normal part of the metallurgical process, can contribute to good brazed joints when brazing, for example, high-temperature metals with nickel-base brazing filler metals.

Brazing Temperature. When choosing a brazing filler metal, the first selection criterion is the brazing temperature. Some brazing-temperature ranges are given in Table 5. Very few brazing filler metals possess narrow melting ranges. Brazing filler metals in which the solidus and liquidus temperatures are close together do not usually exhibit a strong tendency to coexist as a mixture of liquid and solid phases or to liquate. They flow readily and should be used with small joint clearances. As the solidus and liquidus temperatures diverge, the tendency to liquate increases, requiring greater precautions in brazing filler metal application. The mixture of solid and liquid metal can aid gap filling.

TABLE 5 JOINING TEMPERATURES FOR FILLER METALS USED IN BRAZING APPLICATIONS

BRAZING FILLER-METAL GROUPS	JOINING TEMPERATURE	
	°C	°F
NI, CO, AND PD ALLOYS	[GES]1100	[GES]2012
CU, NI, AND AU ALLOYS	[GES]1100	[GES]2012
CU-ZN, CU-SN, NI-P, NI-CR-P, PD-AG-CU	[GES]900	[GES]1652
CU-P, CU-AG-P	600-800	1112-1472
AG-CU-ZN, AG-CU-ZN-CD	600-800	1112-1472
AL-SI	580-600	1076-1112
MG-AL-ZN	585-615	1085-1139

The necessity for the brazing filler metal to melt below the solidus of the base metal is just one of several factors that affect its selection. It may be necessary for the brazing filler metal to melt below the temperature at which parts to be brazed lose strength or above the temperature at which oxides are reduced or dissociated. Joining may have to be carried out above the solution-treatment temperature of the base metal, or at a structure refining temperature, or below the remelt temperature of a brazing filler metal used previously in the production of a brazed subassembly.

Liquidation. During melting, the composition of the liquid and solid filler metal phases changes as the temperature increases from the solidus to the liquidus point. If the portion that melts first is allowed to flow out, then the remaining solid phases have higher melting points than the original composition and may never melt, remaining behind as a residue, or "skull." Filler metals with narrow melting ranges do not tend to separate, but flow quite freely in joints of extremely narrow clearance as long as the solution and diffusion rates of the filler metal with the base metal are low (as in aluminum brazing, the use of silver filler metal on copper, and so on). The rapid heating of filler metals with wide melting ranges or their application to the joint after the base metal reaches brazing temperature will minimize the separation, or liquation. However, liquation cannot be entirely eliminated and wide-melting-range filler metals, which tend to have more sluggish flow, will require wider joint clearances and will form larger fillets at the joint extremities. A few brazing filler metals become sufficiently fluid below the actual liquidus temperature, and satisfactory joints are achieved, even though the liquidus temperature has not been reached.

When sluggish behavior is needed, such as in filling large gaps, brazing can be accomplished within the melting range of the filler metal. However, the brazing temperature is usually 10 to 93 °C (20 to 170 °F) above the liquidus of the filler metal. The actual temperature required to produce a good joint filling is influenced by factors such as heating rate, brazing environment (atmosphere or flux), thickness of parts, thermal conductivity of the metals being joined, and type of joint to be made.

Large-scale mechanical properties of brazing filler metals can be a guideline for their suitability (in terms of strength, oxidation resistance, and so on) for use in different capillary joining applications. However, designers cannot use the mechanical properties of brazed assemblies that are related to different joint configurations brazed at given cycles of time and temperatures.

The placement of the brazing filler metal is an important design consideration, not only because the joint must be accessible to the method chosen, but because, in automatic heating setups, the filler metal must be retained in its location until molten. Brazing filler metals are available in different forms (Table 6) and filler-metal selection may depend on which form is suitable for a particular joint design.

TABLE 6 AVAILABLE PRODUCT FORMS OF BRAZING FILLER METALS CATEGORIZED BY GROUP

BRAZING FILLER-METAL GROUPS	AVAILABLE FORMS(A)											
	AD	CL	FL	FO	PA	PB	PO	PR	SH	RD	SP	WI
NI AND CO ALLOYS	X	. . .	X	(B)	X	X	X	(B)	X	(B)	. . .	(C)
PD ALLOYS	. . .	. . .	X	X	. . .	. . .	X	X	X	. . .	X	X
CU AND AU ALLOYS	. . .	. . .	X	X	X	. . .	X	X	X	. . .	X	X

NI ALLOYS	X	. . .	X	(B)	X	X	X	(B)	X	(B)	. . .	(C)
CU-SN, CU-ZN, PD-AG-CU	. . .	. . .	X	X	. . .	. . .	X	X	X	. . .	X	X
NI-P, NI-CR-P	X	. . .	. . .	. . .	X	X	X	. . .	. . .	. . .	. . .	. . .
CU-P, CU-AG-P	. . .	. . .	. . .	. . .	X	. . .	X	. . .	. . .	. . .	X	X
AG-CU-ZN, AG-CU-ZN-CD	. . .	. . .	X	X	. . .	. . .	X	X	X	. . .	X	X
AL-SI	. . .	X	. . .	. . .	. . .	. . .	X	. . .	X	. . .	. . .	X
MG-AL-ZN	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	X	. . .	X	X

(A) AD, ADHESIVE SHEET; CL, CLADDING; FL, FLUX PASTE; FO, FOIL; PA, PASTE (NONFLUXING); PB, PLASTIC-BONDED SHEET; PO, POWDER; PR, PREFORM; SH, SHIM; RD, ROD; SP, STRIP; WI, WIRE.

(B) A FEW ALLOYS ONLY.

(C) PLASTIC BOND

Several general rules apply in the filler-metal placement. Wherever possible, the filler metal should be placed on the most slowly heated part of the assembly in order to ensure complete melting. Although brazing is independent of gravity, gravity can be used to assist filler-metal flow, particularly for those filler metals having wide ranges between their solidus and liquidus temperatures. Brazing filler metals can be chosen to fill wide gaps or to flow through joint configurations where the gap may vary, for example, around a corner. Unless movement between the components being joined is unimportant or can be corrected manually (through self-jigging or by using fixtures after the filler metal is molten), the filler metal should be placed outside the joint and allowed to flow into it. It should not be placed between the joint members. If erosion of thin members is possible, then the brazing filler metal should be placed on the heavier sections, which heat up more slowly, so that flow proceeds toward the thin sections. Apart from suiting the placement method selected for the joint, the form chosen for the brazing filler metal may be needed to accurately gage the amount applied, not just for economy and reproducibility, but to regulate and maintain joint properties and configuration.

Surface Preparation. A clean, oxide-free surface is imperative to ensure uniform quality and sound brazed joints. All grease, oil, dirt, and oxides must be carefully removed from the base and filler metals before brazing, because only then can uniform capillary attraction be obtained. Brazing should be done as soon as possible after the material has been cleaned. The length of time that the cleaning remains effective depends on the metals involved, atmospheric conditions, storage and handling practices, and other factors. Cleaning operations are commonly categorized as being either chemical or mechanical. Chemical cleaning is the most effective means of removing all traces of oil or grease. Trichloroethylene and trisodium phosphate are the usual cleaning agents employed. Oxides and scale that cannot be eliminated by these cleaners should be removed by other chemical means.

The selection of the chemical cleaning agent depends on the nature of the contaminant, the base metal, the surface condition, and the joint design. Regardless of the nature of the cleaning agent or the cleaning method used, it is important that all residue or surface film be adequately rinsed from the cleaned parts to prevent the formation of other equally undesirable films on the faying surfaces.

Objectionable surface conditions can be removed by mechanical means, such as grinding, filing, wire brushing, or any form of machining, provided that joint clearances are not disturbed. When grinding the surfaces of the parts to be brazed, care should be exercised to ensure that the coolant is clean and free from impurities to avoid grinding these impurities into the finished surfaces.

When faying surfaces of parts to be brazed are prepared by blasting techniques, several factors should be considered. There are two purposes behind the blasting of parts to be brazed. One is to remove any oxide film and the other is to roughen the mating surfaces in order to increase capillary attraction of the brazing filler metal. The blasting material must be clean and must not leave any deposit on the surfaces to be joined that restricts filler-metal flow or impairs brazing. The particles of the blasting material should be angular rather than spherical, so that the blasted parts are lightly roughened, rather than peened, after the scale is removed. The operation should not distort or otherwise harm delicate parts. Vapor blasting and similar wet blasting methods require care, because of the possibility of surface contamination.

Mechanical cleaning may be adequate, in which case it should be permitted (by the design) during manufacture. In those cases that require chemical cleaning, the cleaning operation may be followed by protective electroplating, which necessitates access to the faying surface by the liquids involved.

Another surface protection technique is the use of solid and liquid brazing fluxes. At temperatures up to about 1000 °C (1830 °F), fluxes often provide the easiest method of maintaining or producing surface cleanliness. In such cases, the design must not only permit easy ingress of the flux, but should allow the filler metal to wash it through the joint. Above 1000 °C (1830 °F), the flux residues can be difficult to remove, and surface cleaning can be accompanied by brazing in furnaces with protective atmospheres. However, the design must permit the gas to penetrate the joint.

In addition to cleanliness and freedom from oxides, another important factor in determining the ease and evenness of brazing filler metal flow is surface roughness. Generally, a liquid that wets a smooth surface will wet a rough one even more. A rough surface will modify filler-metal flow from laminar to turbulent, prolonging flow time and increasing the possibility of alloying and other interactions. Surfaces often are not truly planar, and, in some instances, surface roughening will improve the uniformity of the joint clearance. A rough surface features a series of crests and troughs that produce turbulent flow. These irregularities on the surface simulate grips to prevent the flow of the filler metal.

Conversely, there may be a requirement that brazing filler metal not flow onto some surfaces. Stop-off materials will often avoid flow, but the design must permit easy application of the stop-off material without contaminating the surfaces to be joined.

Self-fluxing filler metals, in a suitably protective environment such as vacuum, may provide the essential surface wetting.

Joint Design and Clearance. A brazed joint is not a homogeneous body. Rather, it is a heterogeneous area that is composed of different phases with different physical and chemical properties. In the simplest case, it consists of the base-metal parts to be joined and the added brazing filler metal. However, dissimilar materials must also be considered.

Small clearances are used because the smaller the clearance, the easier it is for capillarity to distribute the brazing filler metal throughout the joint area and the less likely it is that voids or shrinkage cavities will form as the brazing filler metal solidifies. The optimum joints are those in which the entire joint area is wetted and filled by the brazing filler metal. Typically, brazing clearances that range from 0.03 to 0.08 mm (0.001 to 0.003 in.) are designed for the best capillary action and greatest joint strength.

Because brazing relies on capillary attraction, the design of a point must provide an unobstructed and unbroken capillary path to enable the escape of flux, if used, as well as allow the brazing filler metal into the joint. In cases where filler metal is added to a joint by hand, such as by feeding in a rod or wire, the joint entry must be visible and accessible.

Some of the more-important factors influencing joint design are the required strength and corrosion resistance, the necessary electrical and thermal conductivity, the materials to be joined, the mode of application of the brazing filler metal, and the postjoining inspection needs.

Consideration also should be given to the ductility of the base metal, the stress distribution in the joint, and the relative movements of the two surfaces during joining, which may introduce distortion in dimensions of the work to be brazed.

Viscosity, surface tension, and specific gravity of the brazing filler metal are not the only factors that determine the gap-filling capability of a given filler metal. Joint strength increases as joint gap decreases, down to a minimum. Table 7 gives the allowable joint clearances for various filler-metal systems. Other factors that influence optimum joint gap with a specific brazing filler metal are joint length, brazing temperature, and base-metal/filler-metal metallurgical reactions.

TABLE 7 RECOMMENDED GAP FOR SELECTED BRAZE FILLER METALS

BRAZING FILLER-METAL SYSTEM	**JOINT CLEARANCE**	
	mm	**in.**
AL-SI ALLOYS$^{(A)}$	0.15-0.61	0.006-0.024
MG ALLOYS	0.10-0.25	0.004-0.010
CU	0.00-0.05	0.000-0.002
CU-P	0.03-0.13	0.001-0.005
CU-ZN	0.05-0.13	0.002-0.005
AG ALLOYS	0.05-0.13	0.002-0.005

AU ALLOYS	0.03-0.13	0.001-0.005
NI-P ALLOYS	0.00-0.03	0.000-0.001
NI-CR ALLOYS(B)	0.03-0.61	0.001-0.024
PD ALLOYS	0.03-0.10	0.001-0.004

(A) IF JOINT LENGTH IS LESS THAN 6 MM (0.240 IN.), GAP IS 0.12 TO 0.75 MM (0.005 TO 0.030 IN.). IF JOINT LENGTH EXCEEDS 6 MM (0.240 IN.), GAP IS 0.25 TO 0.60 MM (0.010 TO 0.24 IN.).

(B) MANY DIFFERENT NICKEL BRAZING FILLER METALS ARE AVAILABLE, AND JOINT GAP REQUIREMENTS MAY VARY GREATLY FROM ONE FILLER METAL TO ANOTHER

It is important to remember that an assembly expands during heating and that the joint gap can either widen or close by the time the brazing filler metal starts to melt and move. It is desirable to design the joint to expose the solidifying filler metal to compressive, rather than tensile, stress. This is much more important in brazing than in soldering, because brazing temperatures are higher, increasing the total thermal expansion. With cylindrical joints, the components with the larger coefficient of expansion should be on the outside, whenever possible.

Joint clearance is probably one of the most significant factors in vacuum brazing operations. Naturally, it receives special consideration when joint are designed at room temperature. Actually, joint clearance is not the same at all phases of brazing. It will have one value before brazing, another value at the brazing temperature, and still another value after brazing, especially if there has been a substantial interaction between the brazing filler metal and the base metal. To avoid confusion, it has become general practice to specify joint clearance as the value at room temperature before brazing.

The recommended joint clearances given in Table 7 are based on joints that have members of similar metals and equal mass. When dissimilar metals and/or metals of widely differing masses are joined by brazing, special problems arise. These problems necessitate specialized selection of brazing filler metals. Furthermore, the most suitable joint clearance for the specific job must be carefully determined. Thus, clearances given in Table 7 are the clearances that must be attained at the brazing temperature.

Although there are many kinds of brazed joints, the selection of joint type is not as complicated as it may seem, because butt and lap joints are the two fundamental types used. All other types, such as the scarf joint, are modifications of these two. The scarf joint of the scarf angle. It approaches the lap joint at the other extreme of the scarf angle.

Selection of joint type is influenced by the configuration of the parts, as well as by joint strength and other service requirements, such as electrical conductivity, pressure tightness, and appearance. Also influential in the selection of joint type are fabrication techniques, production quantities, and methods of feeding brazing filler metal. Lap joints are generally preferred for brazing operations, particularly when it is important that the joints be at least as strong as the weaker member. For maximum strength, lap-joint length should equal three to four times the thickness of the thinner member.

Temperature and Time. The temperature of the brazing filler metal has an important effect on the wetting and alloying action, which increases with increasing temperature. The temperature must be above the melting point of the brazing filler metal and below the melting point of the parent metal. Within this range, a brazing temperature that is most satisfactory overall is generally selected.

Usually, low brazing temperatures are preferred in order to economize on the heat energy required, minimize the heat effect on the base metal (annealing, grain growth, or warpage, for example), minimize base-metal/filler-metal interactions, and increase the life of fixtures, jigs, or other tools.

Higher brazing temperatures may be desirable so as to:

- ENABLE THE USE OF A HIGHER-MELTING, BUT MORE ECONOMICAL, BRAZING FILLER METAL
- COMBINE ANNEALING, STRESS RELIEF, OR HEAT TREATMENT OF THE BASE METAL WITH BRAZING
- PERMIT SUBSEQUENT PROCESSING AT ELEVATED TEMPERATURES
- PROMOTE BASE-METAL/FILLER-METAL INTERACTIONS IN ORDER TO CHANGE THE

COMPOSITION OF THE BRAZING FILLER METAL (THIS TECHNIQUE IS USUALLY USED TO INCREASE THE REMELT TEMPERATURE AND DUCTILITY OF THE JOINT)

- EFFECTIVELY REMOVE SURFACE CONTAMINANTS AND OXIDES WITHIN PROTECTIVE ATMOSPHERES BRAZING (ALSO APPLIES TO PURE DRY HYDROGEN, TO ARGON, AND TO VACUUM)
- AVOID STRESS CRACKING

The time at brazing temperature also affects the wetting action. If the brazing filler metal has a tendency to creep, the distance generally increases with time. The alloying action between filler metal and parent metal is, of course, a function of temperature, time, and quantity of filler metal. For production work, temperature, time, and quantity of filler metal are generally kept at a minimum, consistent with good quality, where diffusion is not required.

References cited in this section

14. W.H. KOHL, SOLDERING AND BRAZING, *VACUUM,* VOL 14, 1964, P 175-198

15. *METALS HANDBOOK,* 9TH ED., *WELDING, BRAZING, AND SOLDERING,* VOL 6, AMERICAN SOCIETY FOR METALS, 1983, P 929-995

Heating Methods

The numerous heating methods available for brazing often represent constraints on the designer or engineer when selecting the best type of capillary joint (Table 8). However, because effective capillary joining requires the efficient transfer of heat from the heat source into the joint, one cannot braze a 0.025 mm (0.001 in.) diam wire to a 2.3 kg (5.1 lb) lump of copper with a small torch.

TABLE 8 RELATIVE RATING OF SELECTED BRAZING PROCESS HEATING METHODS

METHOD	CHARACTERISTICS[(A)]					
	CAPITAL COST	RUNNING COST	BASIC OUTPUT	FLUX REQUIRED	VERSATILITY	OPERATOR SKILL REQUIRED
Torch (flame)	L/M	M/H	L	YES	H	YES
Electrical resistance	M	M	M/H	YES	L	NO
Induction	M/H	M	M/H	Y/N	M	NO
Furnace (atmosphere)	M/H	M/H	H	Y/N	M	NO
Furnace (vacuum)	H	L	H	NO	M	NO
Dip (flux bath)	L/M	M/H	L/M	YES	L	YES
Infrared	M	L	M	Y/N	L	NO

(A) H, HIGH; M, MEDIUM; L, LOW

The size and value of individual assemblies, the numbers required, and the required rate of production will influence the selection of heating method. Other factors that also must be considered include the rate of heating, differential thermal gradients, and both external and internal cooling rates. These factors vary tremendously with different methods of heating, and their effects on dimensional stability, distortion, and joint structure must be considered.

Manual torch brazing is the method most frequently used for repairs, one-of-a-kind brazing jobs, and short production runs, as an alternative to fusion welding. Any joint that can be accessed by a torch and brought to brazing temperature (by the torch alone or in conjunction with auxiliary heating) can be readily brazed by this technique.

Although any flame-producing device can be used for torch brazing, commercial applications are accomplished with the same type of torch, controls, and gases used for torch fusion welding. Conversion to brazing merely requires changes in torch nozzles and goggle lenses.

The torch brazing technique is relatively simple and can be mastered by the mechanically adept in a short time. Those already experienced in torch welding and the brazing of other metals generally encounter little difficulty learning torch brazing.

Depending on the temperature and heat required, all commercial gas mixtures can be used to fuel the torch: oxyacetylene, oxyhydrogen, oxy-natural gas, acetylene and air, hydrogen and air, propane, methane, and natural gas and air. Oxyacetylene and oxy-natural gas are the mixtures most often used commercially and are preferred in that order. Flame adjustment is very important. Generally, a slightly reducing flame is desirable in order to preserve part surfaces from oxidation.

The oxyacetylene combination produces the highest temperature. The other gases are cooler and their flames have relatively lower intensity. Thus, they are easier to use and are applied advantageously to light-gage material.

Manual torch brazing is particularly useful on assemblies with sections of unequal mass. As warranted by the rate of production, machine operations can be set up using one or more torches equipped with single or multiple flame tips. The machine can be designed to move either the work or the torches.

Torch brazing can be rather easily automated with appropriate gas supplies, indexing fixtures, and cycle controls. Usually, such systems involve multiple-station rotary indexing tables. The part is fed into a holding fixture at the first station and is then indexed to one or more preheating stations, depending on the heating time required. A brazing station is next, followed by a cooling station and an ejection station (see Fig. 4).

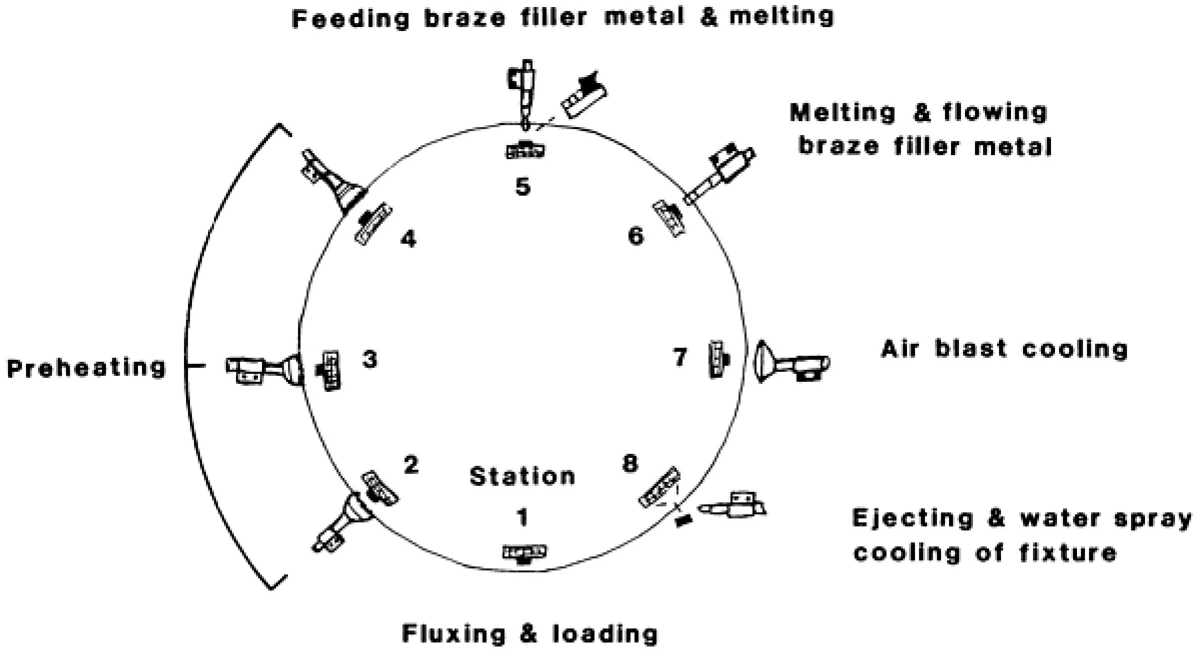
Feeding braze filler metal & melting
Melting & flowing braze filler metal
5
4
6
Preheating
3
7
Air blast cooling
2
Station
8
1
Ejecting & water spray cooling of fixture
Fluxing & loading

FIG. 4 SCHEMATIC OF AN EIGHT-STATION AUTOMATIC TORCH BRAZING OPERATION THAT PRODUCES MAGNET ARMATURE ASSEMBLIES (230 PIECES/H) USED AS STRIKING MEMBERS OF A PRINTING MACHINE

Additional information is available in the article "Torch Brazing" in this Volume.

Furnace brazing is popular because of its comparatively low equipment cost, furnace adaptability, and minimal required jigging. Furnace brazing is a low-cost process relative to other processes such as torch brazing, induction brazing, or salt-bath brazing when high-volume production output is a primary factor. Secondary factors include tooling requirements, fluxing, and cleaning requirements. With many brazing assemblies, the weight of the parts alone is sufficient to hold them together. Other configurations require only one or two rectangular fixturing blocks of metal.

Four basic types of furnaces are used:

- BATCH, WITH EITHER AIR OR CONTROLLED ATMOSPHERE (FIG. 5)
- CONTINUOUS, WITH EITHER AIR OR CONTROLLED ATMOSPHERE (FIG. 6)
- RETORT, WITH CONTROLLED ATMOSPHERE
- VACUUM

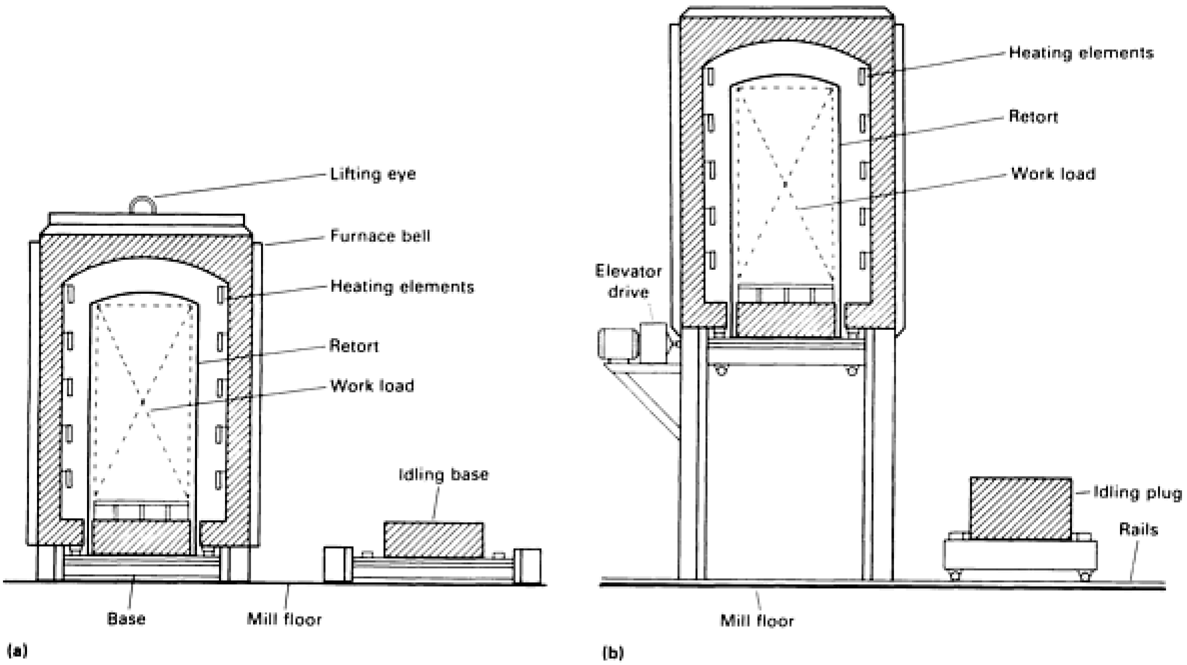
Lifting eye
Furnace bell
Heating elements
Retort
Work load
Idling base
Base
Mill floor
(a)
Heating elements
Retort
Work load
Elevator drive
Idling plug
Rails
Mill floor
(b)

FIG. 5 BATCH-TYPE BRAZING FURNACES. (A) BELL FURNACE. (B) ELEVATOR FURNACE

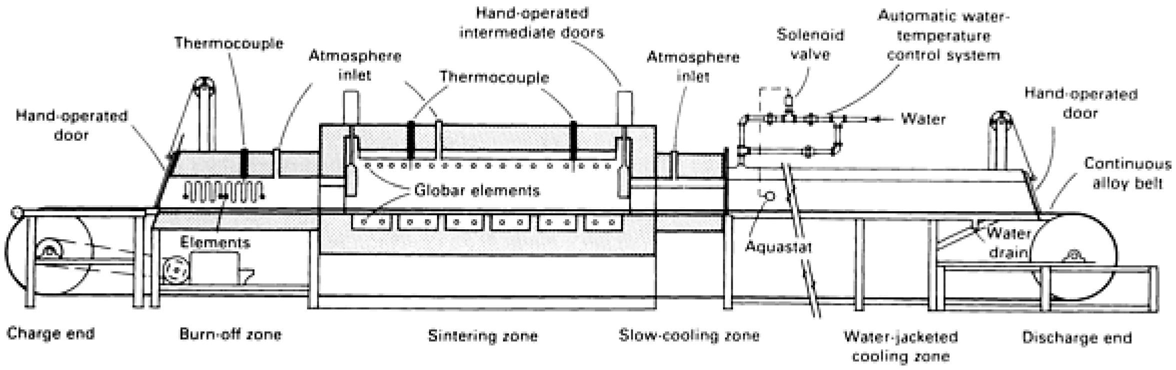
Hand-operated intermediate doors
Thermocouple
Atmosphere inlet
Thermocouple
Atmosphere inlet
Solenoid valve
Automatic water-temperature control system
Hand-operated door
Water
Hand-operated door
Continuous alloy belt
Globar elements
Elements
Aquastat
Water drain
Charge end
Burn-off zone
Sintering zone
Slow-cooling zone
Water-jacketed cooling zone
Discharge end

FIG. 6 CONTINUOUS-TYPE BRAZING FURNACE WITH WATER-JACKETED COOLING CHAMBER FOR USE IN AIR OR CONTROLLED ATMOSPHERES

Furnace brazing is used extensively where the parts to be brazed can be assembled with the brazing filler metal preplaced near or in the joint. Furnace brazing is particularly applicable for high-production applications in which continuous conveyor-type furnaces are used. For medium-production work, batch-type furnaces are best. Regardless of furnace type, heating is usually produced by electrical resistance, although other types of fuel can be used in muffle-type furnaces.

The parts should be self jigged or fixtured and assembled, with brazing filler metal preplaced near or in the joint. Fluxing is used, except in cases where a reducing atmosphere, such as hydrogen, and either exothermic or endothermic combusted gas can be introduced into the furnace. Sometimes, both flux and a reducing atmosphere are necessary. Pure and dry inert gases, such as argon and helium, are used to obtain special atmosphere properties. Fluxes should never be used when furnace brazing with a vacuum atmosphere.

When continuous-type furnaces are used (Ref 16, 17), several different temperature zones can be set up to provide the proper preheating, brazing, and cooling temperatures. The speed through a conveyor-type furnace must be controlled to provide the appropriate time at the brazing temperature. It is also necessary to properly support the assembly so that it does not move while traveling on the belt (Ref 18).

A large volume of furnace brazing is performed in a vacuum atmosphere, which prevents oxidation during the heating cycle and eliminates the need for flux (Ref 19) (Fig. 7). Vacuum brazing has found wide application in the aerospace and nuclear fields, where reactive metals are joined or where entrapped fluxes would not be tolerable. Vacuum brazing does not allow as wide a choice of brazing filler metals as does atmosphere brazing (Ref 20, 21).

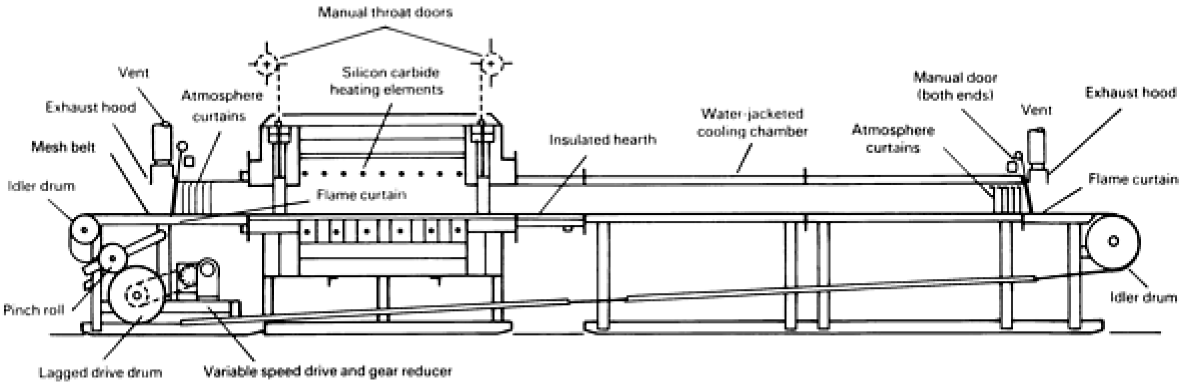
Manual throat doors
Vent
Silicon carbide heating elements
Exhaust hood
Atmosphere curtains
Mesh belt
Idler drum
Flame curtain
Insulated hearth
Water-jacketed cooling chamber
Manual door (both ends)
Atmosphere curtains
Vent
Exhaust hood
Flame curtain
Idler drum
Pinch roll
Lagged drive drum
Variable speed drive and gear reducer

FIG. 7 HIGH-VACUUM BOTTOM-LOADING VACUUM FURNACE FOR MEDIUM-PRODUCTION APPLICATIONS

In most cases, vacuum furnaces are heated by electrical resistance heaters in a numbers of different forms. The overall time of a vacuum brazing operation is usually much longer than that of other batch brazing techniques and involves typing up expensive equipment for long periods to process comparatively small workloads. However, most of the metals on which vacuum brazing excels are also costly, the use of a process that permits material economics through fabrication while ensuring the necessary joint properties often can be justified (Ref 22, 23, 24).

Detailed information is available in the article "Furnace Brazing" in this Volume.

Induction Brazing. The high-frequency induction heating method for brazing is clean and rapid, lends itself to close control of temperature and location, and requires little operator skill.

The workpiece is placed in or near a coil carrying alternating current (ac), which induces the heating current in the desired area (Fig. 8). The coils, which are water cooled, are designed specifically for each part. Therefore, heating efficiency relies on establishing the best coil design and power frequency for each application. In most cases, the coils provide heat only to the joint area.

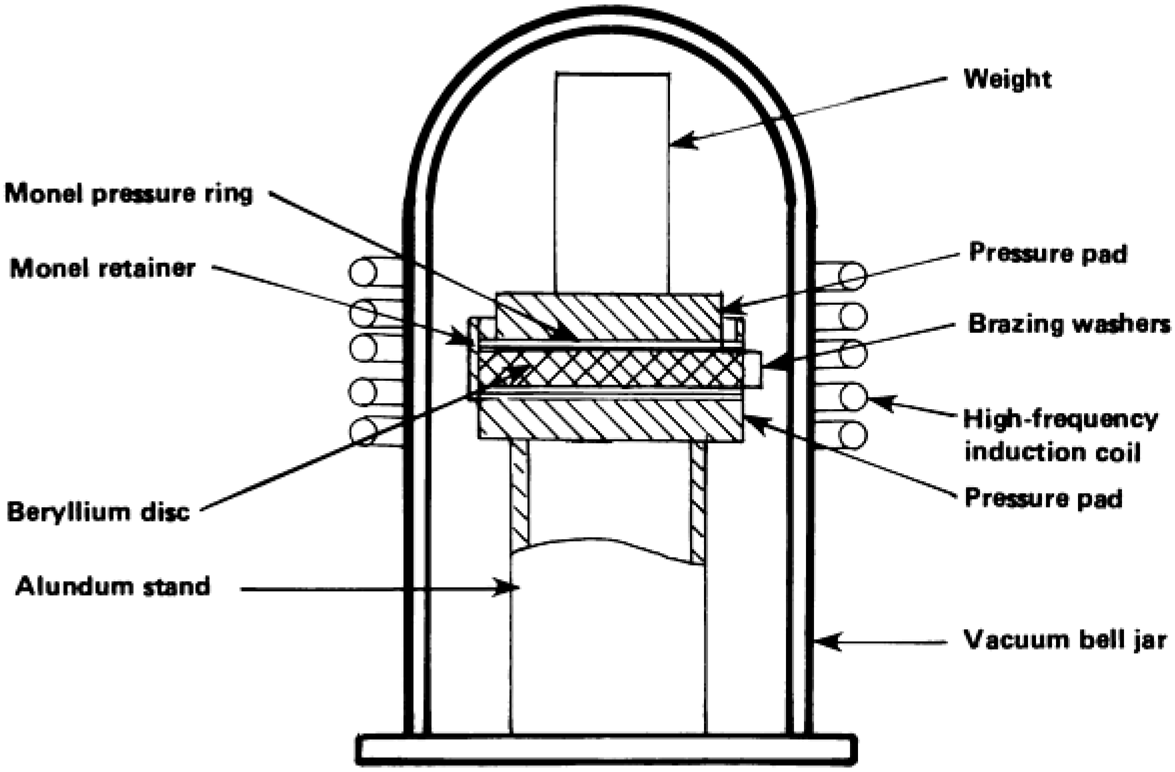

FIG. 8 INDUCTION HEATING APPARATUS USED TO VACUUM BRAZE BERYLLIUM AND MONEL COMPONENTS

The capability to heat selectively enables the induction method to be used when the nature of the brazement demands localized brazing.

There are several types of high-frequency sources to supply power to induction coils, each with a different range of frequencies:

- MOTOR GENERATORS (5 TO 10 KHZ)
- SPARK-GAP OSCILLATORS (20 TO 30 KHZ)
- VACUUM-TUBE OSCILLATORS (20 TO 5000 KHZ)
- SOLID-STATE POWER SUPPLIES (VARIABLE HZ)

The frequency of the power source determines the type of heat that will be induced in the part: high-frequency sources produce "skin" heating, whereas lower-frequency sources provide heating in thicker areas. The brazing heat usually develops within 10 to 60 s.

Induction brazing is well suited for mass production. Mechanized brazing lines for moving assemblies to and from the coils are very common. Brazing filler metals is normally preplaced in the joint, and the brazing can be done in air with the use of a flux, in an inert atmosphere, or in a vacuum atmosphere. The rapid heating rates available with induction heating are a major advantage when using brazing filler metals that tend to vaporize or segregate. The heating cycle for induction brazing is invariably automated, even when manual loading of assemblies is used.

Additional information is available in the article "Induction Brazing" in this Volume.

Dip brazing involves immersion of assembled parts into a suitable molten bath to effect brazing. The bath can be molten brazing filler metal, molten chemical flux, or molten chemical salts. The molten material is contained in a "pot" furnace heated by oil, gas, or electricity. In some instances, electrical-resistance heaters are used in the bath (Fig. 9). In the first type of bath, the parts being joined are held together and immersed in flux-covered molten filler metal, which flows into the joints when the parts reach the bath temperature. The flux cleans the workpiece as it is introduced and protects the brazing filler metal by preventing oxidation and the loss of volatile elements from the bath.

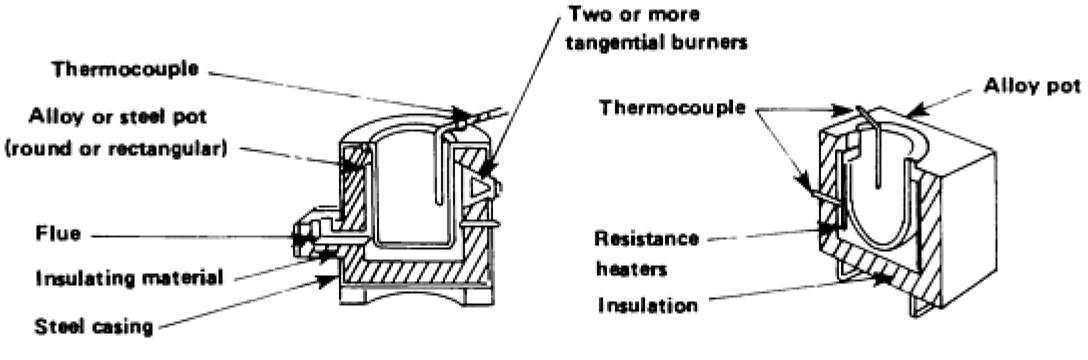
Two or more tangential burners
Thermocouple
Alloy or steel pot (round or rectangular)
Flue
Insulating material
Steel casing
Thermocouple
Alloy pot
Resistance heaters
Insulation

(a) Gas fired or oil fired

(b) Resistance heated

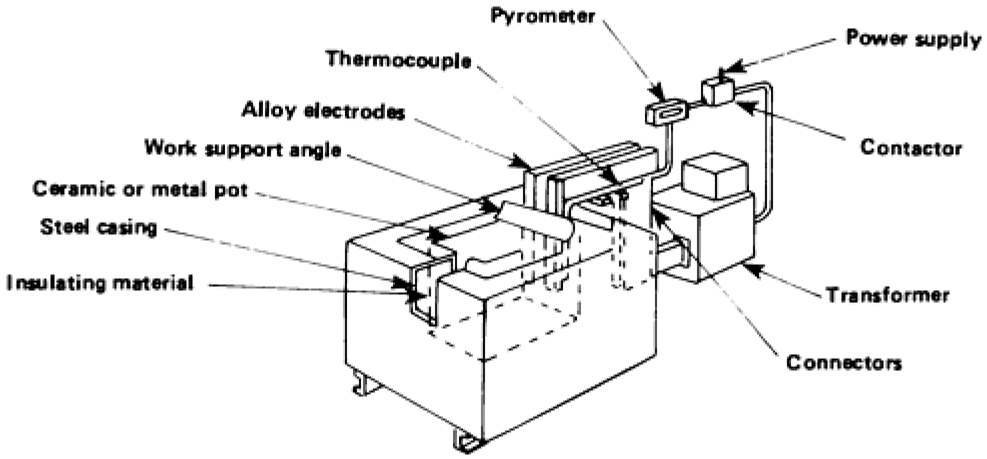
Pyrometer
Power supply
Thermocouple
Alloy electrodes
Contactor
Work support angle
Ceramic or metal pot
Steel casing
Insulating material
Transformer
Connectors

(c) Immersed electrodes

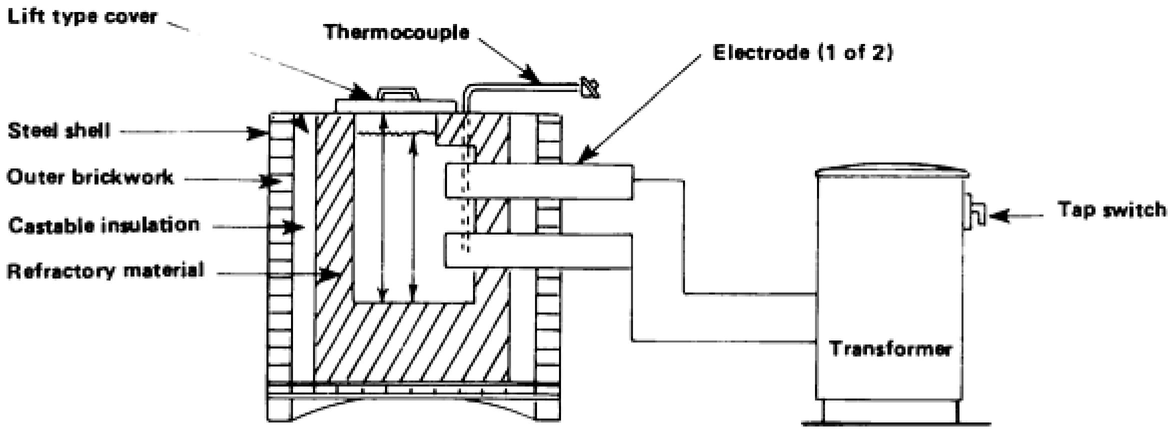
Lift type cover
Thermocouple
Electrode (1 of 2)
Steel shell
Outer brickwork
Castable insulation
Refractory material
Tap switch
Transformer

(d) Electrically heated (submerged electrodes)

FIG. 9 EXTERNALLY AND INTERNALLY HEATED FURNACES USED FOR DIP BRAZING APPLICATIONS

The pot, or crucible, used in the molten-metal method is usually made of graphite. Jigging to maintain alignment is generally necessary. Because of the difficulties of heating and containing metals at high temperatures, filler metals that require a brazing temperature above 1000 °C (1830 °F) are rarely used. Therefore, the choice of brazing filler metal is restricted to straight brasses and silver-base filler metals.

In the molten-flux method, the brazing filler metal is located in or near the joints and is heated to the required temperature by immersion in a bath of molten flux. Molten salt bath brazing has a greater scope than any other single brazing process. It can be used on as wide a range of parent metals as torch brazing, but is not subject to the same maximum temperature limitations. Unfortunately, it is an inflexible process. The type of salt used for a particular application depends on the ease with which the parent-metal surface oxides can be removed and on the temperature required for brazing.

The molten-flux method is used extensively for brazing aluminum and its alloys. The flux bath provides excellent protection against metal reoxidation, which can occur quite easily with aluminum.

The dip brazing method generally causes less distortion than torch brazing, because of its uniform heating. However, it may require relatively complex tooling and is therefore best used in medium- to high-production runs. This process is particularly well suited for small- to medium-sized parts with multiple hidden joints.

Detailed information is available in the article "Dip Brazing" in this Volume.

Resistance brazing is most applicable to relatively simple joints in metals that have high electrical conductivity. In this process, the workpieces are heated locally. Brazing filler metal that is preplaced between the workpieces is melted by the heat obtained from its resistance to the flow of electric current through the electrodes and the work. Usually, the heating current, which is normally ac, is passed through the joint itself. The joint becomes part of an electrical circuit, and the brazing heat is generated by the resistance at the joint.

Equipment is the same as that used for resistance welding, and the pressure needed for establishing electrical contact across the joint is ordinarily applied through the electrodes. The electrode pressure also in the usual means for providing the tight fit needed for capillary behavior in the joint. The component parts are generally held between copper or carbon-graphite electrodes.

The flux that is used must be conductive. Normally, fluxes are insulators when cool and dry, but are conductive when wet. The process is generally used for low-volume production in joining electrical contacts, related electrical elements, copper commutator segments, stainless steel tube to fittings, and so on.

Additional information is available in the article "Resistance Brazing" in this Volume.

Infrared (Quartz) Brazing. The development of high-intensity quartz lamps and the availability of suitable reflectors have made the infrared heat technique commercially important for brazing. Infrared heat is radiant heat that is obtained with the sources having energy frequency below the red rays in the light spectrum. Although there is some visible light with every "black" source, heating is principally done by the invisible radiation. Heat sources (lamps) capable of delivering up to 5000 W of radiant energy are commercially available. The lamps do not need to follow the contour of the part in order to hear the joint area, even though the heat input varies inversely as the square of the distance from the source, unless reflectors are used to concentrate the heat.

Lamps are often arranged in a toaster-like configuration, with parts traveling between two banks of lamps (Fig. 10). Infrared brazing can concentrate large amounts of heat in small areas, which can be advantageous in certain applications. Infrared brazing setups are generally not as fast as induction brazing, but the equipment is less expensive.

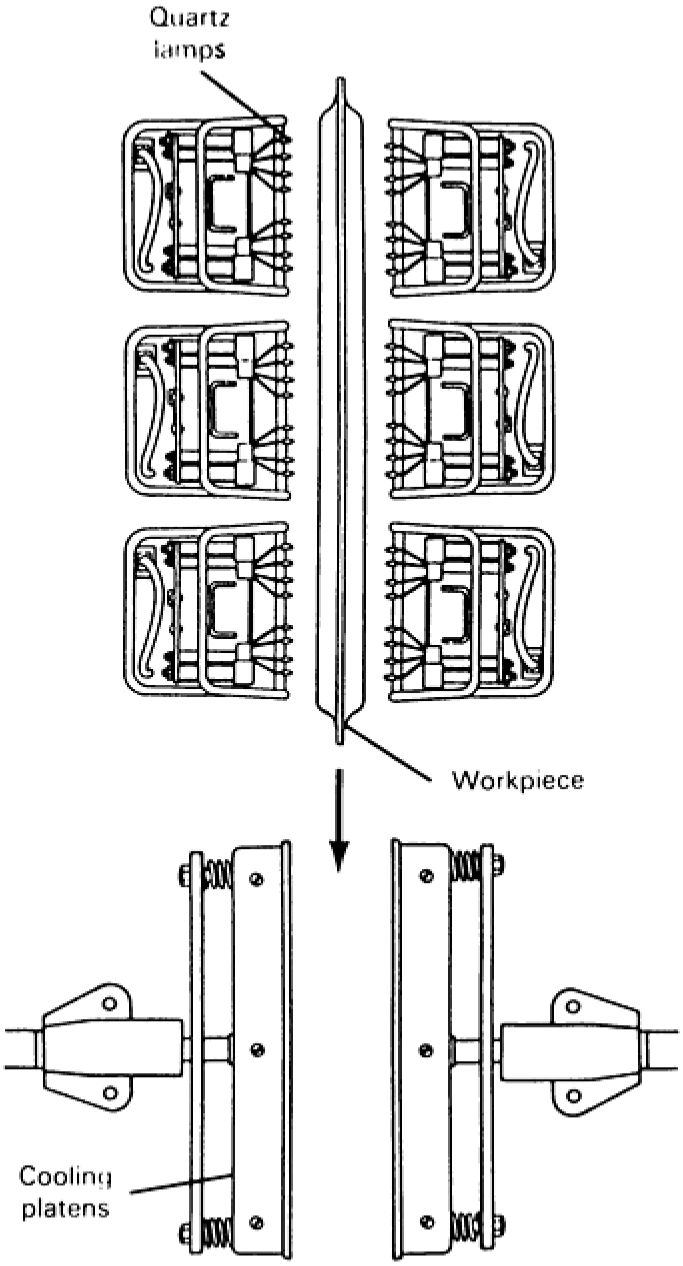

FIG. 10 WORKPIECE SUBJECTED TO INFRARED BRAZING PROCESS PRIOR TO BEING MOVED TO COOLING PLATENS

Exothermic brazing is a special process, but is rarely used because there are more-economical production methods. In this process, the heat required to melt and flow the brazing filler metal is produced by a solid-state exothermic chemical reaction, which is defined as any reaction between two or more reactants that involves the free energy of the system. Although nature provides countless numbers of these reactions, only the solid-state or nearly solid-state metal/metal-oxide reactions are suitable for use in exothermic brazing units.

Additional information is available in the article "Exothermic Brazing" in this Volume.

Electron-beam and laser brazing have been used in a limited number of applications. A laser can be used when a very small localized area of heat is required, such as for brazing small carbide tips on printer heads for electronic printers. The electron beam can be used for brazing by defocusing the beam to provide a wider area of heating. Because this is done in a vacuum, fluxes cannot be used and the brazing filler metal must be selected so that there is little or no vaporization during brazing.

Microwave Brazing. One of the newest heating methods to be developed is the use of microwaves. This technique is being used to join ceramics in high-temperature, corrosion-resistant, and high-performance applications. These applications require complex ceramics parts with strong, durable joints (Ref 25, 26).

The technique uses a single-mode microwave cavity. An iris controls the percent of microwaves reflected in the cavity, and a plunger adjusts the frequency. Together, they focus microwaves on the joint. The major advantage of this technique in terms of ceramic parts is that the entire part does not have to be heated in order to make a joint. Reheating finished ceramic parts can subject them to thermal stresses and cause cracking, which weakens the part. With microwave joining, only the interface between the pieces is heated. The technique is faster than conventional joining, which requires long times to heat ceramics uniformly. Large pieces, can, in principle, be joined at a lower cost.

One disadvantage of microwave joining is that commercial equipment is only produced by a few companies, mostly for research and development purposes.

Some microwave units feature 2.45 GHz single-mode or controlled multimode operation. Their cavity is designed to provide controlled-mode patterns that can focus high-energy microwave fields to heat a part only where it is needed, such as at a joint. The joint or part can be heated up to 2200 °C (3990 °F). Researchers have joined 92% Al_2O_3 ceramic parts in a 6 GHz single-mode microwave cavity at various temperatures and times. The bending strength of this joint gradually increased from the solidus temperature of 1400 °C (2550 °F) and reached a peak of 420 MPa (60 ksi) at about 1750 °C (3180 °F).

Braze welding is a joining process in which a filler metal is melted and deposited in a specific joint configuration, and a metallurgical bond is obtained by a wetting action that is often accompanied by some degree of diffusion with the base metals. Braze welding requires heating and melting of the filler metal that has a melting (liquidus) temperature above 450 °C (840 °F).

Stringent fit-up is not critical, because the filler metal is deposited in grooves and spaces and flows into gaps wider than those used for brazing. Fabricators use braze welding as a low-temperature substitute for oxyfuel welding or as a low-cost substitute for brazing. Joint designs for braze welding are the same as for oxyfuel welding (OFW). Braze welding has been used to join cast iron, steels, copper, nickel, and nickel alloys.

Compared to conventional fusion-welding processes, braze welding requires less heat input, permits higher travel speeds, and causes less distortion. Deposited filler metal is soft and ductile, providing good machinability, and residual stresses are low. The process joins brittle cast irons without extensive preheating.

Although most braze welding initially used an OFW torch, copper filler metal brazing rod, and a suitable flux, present applications use carbon arc welding (CAW), gas-tungsten arc welding (GTAW), gas-metal arc welding (GMAW), or plasma arc welding (PAW) without flux in the manual, semiautomatic, or automatic modes to economically bond and deposit the filler metal in the braze weld joints. OFW is, however, still widely used in machinery repair applications. Filler metal selection, proper wetting and compatibility with the base metals, and shielding from air are important to the effective use of the process with any suitable heating method.

A wide variety of parts can be braze welded using typical weld joint designs. Groove, fillet, and edge welds can be used to join simple and complex assemblies made from sheet, plate, pipe, tubing, rods, bars, castings, and forgings. Sharp corners that are easily overheated and may become points of stress concentrations should be avoided. To obtain good strength, an adequate bond area is required between filler metal and the base metal. Weld groove geometry should provide an adequate groove face area, so that the joint will not fail along the interfaces. Proper joint design selection will produce deposited filler-metal strengths that may meet or exceed the minimum base-metal tensile strengths. Because of the inert shielding gas, electrical arc methods have fewer included flux compounds and oxides at the faying surfaces. The result is higher joint strength and improved corrosion resistance. Original surfaces are restored by overlayments and subsequent machining.

References cited in this section

16. W.S. STEFFAN, NITROGEN-BASED ATMOSPHERE FURNACE BRAZING SYSTEM PROVIDES STRENGTH AND LEAK-PROOF PERFORMANCE OF NEWLY DESIGNED FUEL INJECTION RAIL,

IND. HEAT., AUG 1989, P 21-23

17. *METALS HANDBOOK,* 9TH ED., *POWDER METALLURGY,* VOL 7, 1984, P 358
18. *METALS HANDBOOK,* 8TH ED., VOL 6, 1971, P 595
19. PRODUCT LITERATURE, SECO-WARWICK CORP. (SUNBEAM), SEPT 1977
20. A. SAKAMOTO, STUDY OF FURNACE ATMOSPHERE FOR VACUUM-INERT GAS PARTIAL-PRESSURE BRAZING, *WELD. J.,* NOV 1991, P 311S-320S
21. NEW GENERATION OF VACUUM HEAT-TREATMENT FURNACES FEATURE HIGH GAS PRESSURE QUENCHING AND OIL QUENCH VACUUM FURNACE, *METALLURGIA,* MAY 1989, P 205
22. W.R. JONES, VACUUM--ANOTHER ATMOSPHERE?, *HEAT TREAT.,* OCT 1986, P 39-41
23. W.T. HOOVEN AND K.W. NOKES, MAKE YOUR VACUUM BRAZING FURNACES USER FRIENDLY, *WELD. J.,* OCT 1990, P 25-29
24. S. DRESSLER, ROLE OF HOT WALL VACUUM FURNACES IN PLASMA-ASSISTED SURFACE TREATMENT: PART II, OPERATING INFLUENCES, *IND. HEAT.,* SEPT 1989, P 31-33
25. "ADVANCED STRUCTURAL CERAMICS: TECHNOLOGIES, ECONOMICS AND MARKET OPPORTUNITIES," GB-107, BCC INC., DEC 1987
26. E.J. KUBEL, JR., STRUCTURAL CERAMICS; MATERIALS OF THE FUTURE, *ADV. MATER. PROCESS.,* VOL 134, AUG 1988, P 25-33

Fundamentals of Brazing

Mel M. Schwartz, Sikorsky Aircraft

References

1. C.M. ADAMS, JR., "DYNAMICS OF WETTING IN BRAZING AND SOLDERING," TECHNICAL REPORT WAL TR 650/1, ARMY MATERIALS RESEARCH AGENCY, WATERTOWN ARSENAL, JULY 1962
2. S. WEISS AND C.M. ADAMS, JR., THE PROMOTION OF WETTING, *WELD. J.,* VOL 46 (NO. 2), FEB 1967, P 49S-57S
3. W.D. HARKINS, *PHYSICAL CHEMISTRY OF SURFACE FILMS,* REINHOLD, 1952
4. M.M. SCHWARTZ, *MODERN METAL JOINING TECHNIQUES,* WILEY, 1969
5. A. BONDI, SPREADING OF LIQUID METALS ON SOLID SURFACE CHEMISTRY OF HIGH ENERGY SUBSTANCES, *CHEM. REV.,* VOL 52 (NO. 2), 1953, P 417-458
6. R.G. GILLILAND, WETTING OF BERYLLIUM BY VARIOUS PURE METALS AND ALLOYS, *WELD. J.,* VOL 43, JUNE 1964, P 248S-258S
7. M. NAKA, T. TANAKA, AND I. OKAMOTO, *Q. J. JPN. WELD. SOC.,* VOL 4, 1986, P 597
8. R.R. KAPOOR AND T.W. EAGER, *J. AM. CERAM. SOC.,* VOL 72, 1989, P 448
9. R.E. LOCHMAN, *CERAM. BULL.,* VOL 68, 1989, P 891
10. D.H. KIM, S.H. HWANG, AND S.S. CHUN, THE WETTING, REACTION, AND BONDING OF SILICON NITRIDE BY CU-TI ALLOYS, *J. MATER. SCI.,* VOL 26, 1991, P 3223-3234
11. D.L. KELLER *ET AL.,* WETTABILITY OF BRAZING FILLER METALS, *WELD. J.,* VOL 69 (NO. 10), 1990, P 31-34
12. W.S. BENNETT *ET AL.,* VACUUM BRAZING STUDIES ON HIGH-MANGANESE STAINLESS STEEL, *WELD. J.,* VOL 53, 1974, P 510S-516S
13. W. FEDUSKA, HIGH-TEMPERATURE BRAZING ALLOY--BASE METAL WETTING REACTIONS, *WELD. J.,* VOL 38 (NO. 3) 1959, P 122S-130S

14. W.H. KOHL, SOLDERING AND BRAZING, *VACUUM,* VOL 14, 1964, P 175-198
15. *METALS HANDBOOK,* 9TH ED., *WELDING, BRAZING, AND SOLDERING,* VOL 6, AMERICAN SOCIETY FOR METALS, 1983, P 929-995
16. W.S. STEFFAN, NITROGEN-BASED ATMOSPHERE FURNACE BRAZING SYSTEM PROVIDES STRENGTH AND LEAK-PROOF PERFORMANCE OF NEWLY DESIGNED FUEL INJECTION RAIL, *IND. HEAT.,* AUG 1989, P 21-23
17. *METALS HANDBOOK,* 9TH ED., *POWDER METALLURGY,* VOL 7, 1984, P 358
18. *METALS HANDBOOK,* 8TH ED., VOL 6, 1971, P 595
19. PRODUCT LITERATURE, SECO-WARWICK CORP. (SUNBEAM), SEPT 1977
20. A. SAKAMOTO, STUDY OF FURNACE ATMOSPHERE FOR VACUUM-INERT GAS PARTIAL-PRESSURE BRAZING, *WELD. J.,* NOV 1991, P 311S-320S
21. NEW GENERATION OF VACUUM HEAT-TREATMENT FURNACES FEATURE HIGH GAS PRESSURE QUENCHING AND OIL QUENCH VACUUM FURNACE, *METALLURGIA,* MAY 1989, P 205
22. W.R. JONES, VACUUM--ANOTHER ATMOSPHERE?, *HEAT TREAT.,* OCT 1986, P 39-41
23. W.T. HOOVEN AND K.W. NOKES, MAKE YOUR VACUUM BRAZING FURNACES USER FRIENDLY, *WELD. J.,* OCT 1990, P 25-29
24. S. DRESSLER, ROLE OF HOT WALL VACUUM FURNACES IN PLASMA-ASSISTED SURFACE TREATMENT: PART II, OPERATING INFLUENCES, *IND. HEAT.,* SEPT 1989, P 31-33
25. "ADVANCED STRUCTURAL CERAMICS: TECHNOLOGIES, ECONOMICS AND MARKET OPPORTUNITIES," GB-107, BCC INC., DEC 1987
26. E.J. KUBEL, JR., STRUCTURAL CERAMICS; MATERIALS OF THE FUTURE, *ADV. MATER. PROCESS.,* VOL 134, AUG 1988, P 25-33

Fundamentals of Brazing

Mel M. Schwartz, Sikorsky Aircraft

Selected References

- *BRAZING HANDBOOK,* 4TH ED., AWS, 1991
- M.M. SCHWARTZ, *BRAZING,* ASM INTERNATIONAL, 1987
- M.M. SCHWARTZ, *CERAMIC JOINING,* ASM INTERNATIONAL, 1990

Shielding Gases for Welding

Kevin A. Lyttle, Praxair, Inc.

Introduction

THE SHIELDING GAS used in a welding process has a significant influence on the overall performance of the welding system. Its primary function is to protect the molten metal from atmospheric nitrogen and oxygen as the weld pool is being formed. The shielding gas also promotes a stable arc and uniform metal transfer. In gas-metal arc welding (GMAW) and flux-cored arc welding (FCAW), the gas used has a substantial influence on the form of metal transfer during welding. This, in turn, affects the efficiency, quality, and overall operator acceptance of the welding operation.

The shielding gas interacts with the base material and with the filler material, if any, to produce the basic strength, toughness, and corrosion resistance of the weld. It can also affect the weld bead shape and the penetration pattern.

Understanding the basic properties of a shielding gas will aid in the selection of the right shielding gas or gases for a welding application. Use of the best gas blend will improve the quality and may reduce the overall cost of the welding operation as well.

Shielding Gases for Welding

Kevin A. Lyttle, Praxair, Inc.

Basic Properties of a Shielding Gas

The "controlled electrical discharge" known as the welding arc is formed and sustained by the establishment of a conductive medium called the arc plasma. This plasma consists of ionized gas, molten metals, slags, vapors, and gaseous atoms and molecules. The formation and structure of the arc plasma is dependent on the properties of the shielding gases used for welding. Table 1 lists the basic properties of gases used for welding (Ref 1).

TABLE 1 PROPERTIES OF SHIELDING GASES USED FOR WELDING

GAS	CHEMICAL SYMBOL	MOLECULAR WEIGHT	SPECIFIC GRAVITY(A)	DENSITY		IONIZATION POTENTIAL	
				g/ft^3	g/l	aj(b)	eV
ARGON	AR	39.95	1.38	0.1114	1.784	2.52	15.7
CARBON DIOXIDE	CO_2	44.01	1.53	0.1235	1.978	2.26	14.4
HELIUM	HE	4.00	0.1368	0.0111	0.178	3.92	24.5
HYDROGEN	H_2	2.016	0.0695	0.0056	0.090	2.16	13.5
NITROGEN	N_2	28.01	0.967	0.782	12.5	2.32	14.5
OXYGEN	O_2	32.00	1.105	0.0892	1.43	2.11	13.2

Source: Ref 1

(A) AT 100 KPA (1 ATM) AND 0 °C (32 °F); AIR = 1.

(B) 10^{-18} J.

The ionization potential is the energy, expressed in electron volts, necessary to remove an electron from a gas atom--making it an ion, or an electrically charged gas atom. All other factors held constant, the value of the ionization potential decreases as the molecular weight of the gas increases. Arc starting and arc stability are greatly influenced by the ionization potentials of the component shielding gases used in welding process. A gas with a low ionization potential, such as argon, can atoms into ions easily. Helium, with its significantly higher ionization potential, produces a harder to start, less stable arc.

Although other factors are involved in sustaining the plasma, the respective energy levels required to ionize these gases must be maintained; as a consequence, the arc voltage is directly influenced. For equivalent arc lengths and welding currents, the voltage obtained with helium is appreciably higher than is with argon. This translates into more available heat input to the base material with helium than with argon.

The thermal conductivity of a gas is a measure of how well it is able to conduct heat. It influences the radial heat loss from the center to the periphery of the arc column as well as heat transfer between the plasma and the liquid metal. Argon, which has a low thermal conductivity, produces an arc that has two zones: a narrow hot core and a considerably cooler outer zone. The penetration profile of the weld fusion area then exhibits a narrow "finger" at the root and a wider top. A gas with a high thermal conductivity conducts heat outward from the core; this results in a wider, hotter arc core. This type of heat distribution occurs with helium, argon-hydrogen, and argon-carbon dioxide blends; it gives a more even distribution of heat to the work surface and produces a wider fusion area.

Dissociation and Recombination. Shielding gases such as carbon dioxide, hydrogen, and oxygen are multiatom molecules. When heated to high temperatures within the arc plasma, these gases break down, or dissociate, into their component atoms. They are then at least partially ionized, producing free electrons and current flow. As the dissociated gas comes into contact with the relatively cool work surface, the atoms recombine and release heat at that point. This heat of recombination causes multiatomic gases to behave as if they have a higher thermal conductivity, similar to that of helium. Dissociation and recombination do not occur with gases, such as argon, that consist of a single atom. Thus, at the same arc temperature, the heat generated at the work surface can be considerably greater with gases such as carbon dioxide and hydrogen.

Reactivity/Oxidation Potential. The oxidizing nature of the shielding gas affects both welding performance and the properties of the resultant weld deposit. Argon and helium are completely nonreactive, or inert, and thus have no direct chemical affect on the weld metal. Oxidizing or active gases, such as CO_2 and oxygen, will react with elements in the filler metal or baseplate and will form a slag on the surface of the weld deposit. The loss of elements, such as manganese and silicon, from steel can affect the quality and cost of the weldment produced. Both weld strength and toughness generally decline as the oxidizing nature of the shielding gas increases.

Additions of reactive gases such as oxygen or carbon dioxide enhance the stability of the arc and affect the type of metal transfer obtained. Metal droplet size is decreased, and the number of droplets transferred per unit time increases as the level of oxygen in the shielding gas increases. Oxygen reduces the molten weld bead surface tension, promoting better bead wetting and higher welding travel speeds. Small additions of CO_2 work in a similar manner.

The surface tension between the molten metal and its surrounding atmosphere has a pronounced influence on bead shape. If the surface energy is high, a convex, irregular bead will result. Low values promote flatter beads with minimal susceptibility for undercutting.

Pure argon is generally associated with high interfacial energy, producing a sluggish weld puddle and high, crowned bead. The addition of a small amount of a reactive gas, such as oxygen, lowers this surface tension and promote fluidity and better wetting of the base material; it does this without creating excessive oxidation of the weld metal.

Gas Purity. Some metals, such as carbon steel and copper, have a relatively high tolerance for contaminants in the shielding gas; others, such as aluminum and magnesium, are fairly sensitive to particular contaminants. Still others, such as titanium and zirconium, have an extremely low tolerance for any foreign constituent in the shielding gas.

Depending on the metal being welded and the welding process used, very small quantities of gas impurities can significantly affect welding speed, weld surface appearance, weld bead solidification, and porosity levels. The effects of any given impurity are wide ranging, but weld quality and eventual fitness for purpose are major areas of concern.

There is always a possibility that the gas, as delivered, is contaminated; however, it is far more likely that impurities will enters somewhere between the supply and the end-use points. For this reason, property designed piping systems and high-quality hose are recommended for use with welding shielding gases. Typical industry minimum purity levels for welding gases are listed in Table 2 (Ref 2).

TABLE 2 TYPICAL GASES PURITY AND MOISTURE CONTENT OF SHIELDING

GAS	PRODUCT STATE	MINIMUM PURITY, %	MAXIMUM MOISTURE, PPM$^{(A)}$	APPROXIMATE DEW POINT AT MAXIMUM MOISTURE CONTENT	
				°C	°F
ARGON	GAS	99.995	10	-60	-77
	LIQUID	99.997	6	-64	-83
CARBON DIOXIDE	GAS	99.5	19	-51	-60
	LIQUID	99.8	50	-58	-73
HELIUM	GAS	99.95	32	-51	-61
	LIQUID	99.995	3	-69	-92
HYDROGEN	GAS	99.95	8	-63	-80
	LIQUID	99.995	5	-65	-86
NITROGEN	GAS	99.7	32	-51	-61
	LIQUID	99.997	5	-65	-86
OXYGEN	INDUSTRIAL	99.5	50	-48	-54
	LIQUID	99.5	6	-64	-83

Source: Ref 2

(A) MOISTURE SPECIFICATIONS ARE MEASURED AT FULL CYLINDER PRESSURE, THE PRESSURE AT WHICH THE CYLINDER IS ANALYZED.

Gas density is the weight of the gas per unit volume. Density is one of the chief factors that influence shielding gas effectiveness. Basically, gases heavier than air, such as argon and carbon dioxide, require lower flow rates in use than do the lighter gases, such as helium, to ensure adequate protection of the weld puddle.

References cited in this section

1. N.E. LARSON AND W.F. MEREDITH, *SHIELDING GAS SELECTION MANUAL*, UNION CARBIDE INDUSTRIAL GASES TECHNOLOGY CORP., 1990, P 10
2. N.E. LARSON AND W.F. MEREDITH, *SHIELDING GAS SELECTION MANUAL*, UNION CARBIDE INDUSTRIAL GASES TECHNOLOGY CORP., 1990, P 11

Shielding Gases for Welding

Kevin A. Lyttle, Praxair, Inc.

Characteristics of the Components of a Shielding Gas Blend

To obtain a shielding gas that is suited to a specific application, a mix of gases is generally needed. Each basic gas contributes certain characteristics to the performance of the overall mix. Some gas blends have relatively specific areas of application and limited operating ranges; others can be used on many materials under a variety of welding conditions. Each component of the blend brings with it properties that are supplemented by the others to produce an enhanced level of performance.

Argon is inert or unreactive with respect to the materials present in the welding electrode. With its low ionization potential, argon promotes easy arc starting and stable arc operation. Its lower thermal conductivity promotes the development of axial "spray" transfer in certain forms of GMAW. It is also used in applications where base material distortion must be controlled or where good gap-bridging ability is required.

Helium. Unlike argon, helium is lighter than air and has a low density. Like argon, it is chemically inert and does not react with other elements or compounds. Because of its high thermal conductivity and high ionization potential, more heat is transferred to the base material, thus enhancing the penetration characteristics of the arc. In many applications, it also allows higher weld travel speeds to be obtained. Because of its higher cost, helium is frequently combined with argon or argon mixtures to enhance the overall performance of the blend while minimizing its cost.

Oxygen combines with almost all known elements except rare and inert gases; it vigorously support combustion. Small amounts of oxygen are added to some inert mixtures to improve the stability of the welding arc developed as well as to increase the fluidity of the weld puddle.

In the spray-transfer mode of GMAW, small additions of oxygen enhance the range over which this spatterless form of welding can be performed. The droplet size decreases and the number of drops transferred per unit time increases as oxygen is added to the blend.

Carbon dioxide is a reactive gas that is commonly used alone in certain types of GMAW. Oxidation of the base material and any filler electrode occurs readily. Carbon dioxide is added to argon blends to improve are stability, enhance penetration, and improve weld puddle flow characteristics. The higher thermal conductivity of carbon dioxide (because of the dissociation and recombination of its component parts) transfer more heat to the base material than does argon alone. A broader penetration pattern versus argon is obtained; however, base material distortion and lack of gap-bridging ability are possible problems.

Hydrogen is the lightest known element and is a flammable gas. Explosive mixtures can be formed when certain concentration of hydrogen are mixed with oxygen or air. It is added to inert gases to increase the heat input to the base material or for operations involving cutting and gouging. Because some materials are especially sensitive to hydrogen-related contamination, its use is generally limited to special applications, such as the joining of stainless steels, and to plasma are cutting and gouging.

Nitrogen is generally considered to be inert except at high temperatures. At arc welding temperatures, it will react with some metals (e.g., aluminum, magnesium, steel, and titanium), so it is not used as a primary shielding gas. It can be used with other gases for some welding applications (e.g., copper) and is also widely used in plasma cutting.

Shielding Gases for Welding

Kevin A. Lyttle, Praxair, Inc.

Shielding Gas Selection

In most welding applications, more than one shielding gas or gas blend can be used successfully. For example, there is no one optimal gas blend for joining carbon steels, but a considerable array of mixes are available depending on the specific requirements of the application. For some processes, such as gas-tungsten arc welding (GTAW) and plasma arc welding (PAW), the choices may be somewhat limited by the nature of the electrodes used and the materials being welded. However, for applications involving GMAW, a multitude of blends can be selected from when carbon steels are being joined. Determination of the best blend depends on a number of specific job-related needs.

Accuracy of Gas Blends

The accuracy with which gases are blended is a function of the way in which they are supplied. If the source of the gas is a high-pressure cylinder, the following generally applies:

1. ±10% RELATIVE, MINOR COMPONENT (REF 3)

2. ±0.5% ABSOLUTE FOR CONCENTRATIONS UP TO 5%; ±10% RELATIVE FOR CONCENTRATIONS BETWEEN 5 AND 50% (REF 4)

For example, the mix accuracy of an Ar-2O_2 blend would be Ar/1.8-2.2O_2 (method 1) or Ar/1.5-2.5O_2 (method 2). A blend of Ar-25CO_2 would yield Ar/22.5-27.5CO_2 by their method of calculation.

When gas cylinders are properly filled with the appropriate blend, the components of that mixture will not separate unless the temperature of the environment is reduced far below normal working temperatures. If the gases are supplied from a liquid source, such as a bulk tank, the accuracy of the blend is a function of the mixing equipment used, but most likely falls within the ±10% minor component range.

Shielding Gases for GMAW

By far, the largest number of gas blends have been developed for GMAW, especially for joining carbon steel. These can be roughly divided into four categories: pure gases, argon-oxygen mixes, argon/carbon dioxide mixes, and three-part gas blends composed of either argon, helium, oxygen, carbon dioxide, or hydrogen. Table 3 contains suggestions for shielding gas selection based on material type, thickness, and mode of metal transfer.

TABLE 3 RECOMMENDED SHIELDING GAS SELECTION FOR GMAW

MATERIAL	THICKNESS		TRANSFER MODE	RECOMMENDED SHIELDING GAS	ADVANTAGES AND LIMITATIONS
	MM	IN.			
CARBON STEEL	<2.0	<0.080	SHORT CIRCUITING	AR-25CO_2 AR-15CO_2 AR-8CO_2	GOOD PENETRATION AND DISTORTION CONTROL TO REDUCE POTENTIAL BURNTHROUGH
	2.0-3.2	0.080-0.125	SHORT CIRCUITING	AR-8CO_2 AR-15CO_2 AR-25CO_2	HIGHER DEPOSITION RATES WITHOUT BURNTHROUGH: MINIMUM DISTORTION AND SPATTER; GOOD PUDDLE CONTROL FOR OUT-OF-POSITION WELDING
	>3.2	>0.125	SHORT CIRCUITING	AR-15CO_2 AR-25CO_2 CO_2	HIGH WELDING SPEEDS, GOOD PENETRATION AND PUDDLE CONTROL; APPLICABLE LOT OUT-OF-POSITION WELDS
			GLOBULAR	AR-25CO_2 CO_2	SUITABLE FOR HIGH-CURRENT AND HIGH-SPEED WELDING; DEEP PENETRATION AND FAST TRAVEL SPEEDS, BUT WITH GREATER BURNTHROUGH POTENTIAL
			CONVENTIONAL SPRAY ARC	AR-1O_2 AR-2O_2 AR-5CO_2	GOOD ARC STABILITY; PRODUCES A MORE FLUID PUDDLE AS O_2

			AR-8CO_2 AR-10CO_2 AR-15CO_2 AR-CO_2-O_2 BLENDS	INCREASES; GOOD COALESCENCE AND BEAD CONTOUR, GOOD WELD APPEARANCE AND PUDDLE CONTROL
		PULSED SPRAY	ARGON-5CO_2 AR-HE-CO_2 BLENDS AR-CO_2-O_2 BLENDS	USED FOR BOTH GAGE AND OUT-OF-POSITION WELDMENTS; ACHIEVES GOOD PULSED SPRAY STABILITY OVER A WIDE RANGE OF ARC CHARACTERISTICS AND DEPOSITION RANGES
ALLOY STEEL	ALL SIZES	SHORT CIRCUITING	AR-8CO_2 AR-15CO_2 AR-CO_2-O_2 BLEND	HIGH WELDING SPEEDS; GOOD PENETRATION AND PUDDLE CONTROL; APPLICABLE FOR OUT-OF-POSITION WELDS; SUITABLE FOR HIGH-CURRENT AND HIGH-SPEED WELDING
		SPRAY ARC (HIGH-CURRENT DENSITY AND ROTATIONAL)	AR-2O_2 AR-5O_2 AR-CO_2-O_2 BLENDS AR-HE-CO_2 BLENDS	REDUCES UNDERCUTTING; HIGHER DEPOSITION RATES AND IMPROVED BEAD WETTING; DEEP PENETRATION AND GOOD MECHANICAL PROPERTIES
		PULSED SPRAY	AR-5CO_2 AR-8CO_2 AR-2O_2	USED FOR BOTH LIGHT-GAGE AND HEAVY OUT-OF-POSITION WELDMENTS; ACHIEVES GOOD PULSED SPRAY STABILITY OVER A WIDE RANGE OF ARC CHARACTERISTICS AND DEPOSITION RANGES
STAINLESS STEEL, COPPER, NICKEL, AND CU-NI ALLOYS	ALL SIZES	SHORT-CIRCUITING TRANSFER	AR-HE-CO_2 BLENDS HE-AR-CO_2 BLENDS AR-1O_2 AR-2O_2	LOW CO_2 CONTENTS IN HELIUM MIX MINIMIZE CARBON PICKUP, WHICH CAN CAUSE INTERGRANULAR CORROSION WITH SOME ALLOYS; HELIUM IMPROVES WETTING ACTION; CO_2 CONTENTS >5% SHOULD BE USED WITH CAUTION ON

					SOME ALLOYS; APPLICABLE FOR ALL POSITION WELDING
			SPRAY ARC	AR-HE-CO_2 BLENDS AR-1O_2 AR-2O_2	GOOD ARC STABILITY; PRODUCES A FLUID BUT CONTROLLABLE WELD PUDDLE; GOOD COALESCENCE AND BEAD CONTOUR; MINIMIZES UNDERCUTTING ON HEAVIER THICKNESSES
			PULSED SPRAY	AR-HE-CO_2BLENDS AR-1O_2 AR-2O_2	USED FOR BOTH LIGHT-GAGE AND HEAVY OUT-OF-POSITION WELDMENTS; ACHIEVES GOOD PULSED SPRAY STABILITY OVER A WIDE RANGE OF ARC CHARACTERISTICS AND DEPOSITION RANGES
ALUMINUM, TITANIUM, AND OTHER REACTIVE METALS	≤ 13	$\leq \frac{1}{2}$	SPRAY ARC	ARGON	BEST METAL TRANSFER, ARC STABILITY, AND PLATE CLEANING; LITTLE OR NO SPATTER; REMOVES OXIDES WHEN USED WITH DCEP (REVERSE POLARITY)
			SPRAY ARC	75HE-25AR 50HE-50AR	HIGH HEAT INPUT; PRODUCES FLUID PUDDLE, FLAT BEAD CONTOUR, AND DEEP PENETRATION; MINIMIZES POROSITY
	>13	$> \frac{1}{2}$	SPRAY ARC	HELIUM 50HE-25AR	HIGH HEAT INPUT; GOOD FOR MECHANIZED WELDING AND OVERHEAD; APPLICABLE TO HEAVY SECTION WELDING
			PULSED SPRAY	ARGON	GOOD WETTING; GOOD PUDDLE CONTROL

Argon. Pure argon is generally used on nonferrous base metal, such as aluminum, nickel, copper, and magnesium alloys, and on reactive metals, such as titanium. Argon provides excellent arc stability, penetration, and bead profile when joining these materials. Its low ionization potential results in easy arc starting. Argon produces a constricted arc column with high current density, which concentrates the arc energy over a small area; deep, fingerlike penetration results.

Carbon Dioxide. A reactive gas, carbon dioxide is generally used only for joining carbon steel. It is readily available and relatively inexpensive. Because CO_2 will not support spray transfer, deposition efficiency is lower and spatter and fume levels are higher than with argon blends. Weld bead surfaces are more oxidized and irregular in shape. The higher ionization potential of CO_2 and its characteristic dissociation upon heating provide greater weld fusion and penetration while still achieving acceptable mechanical properties.

Helium. Because of its higher thermal conductivity, helium can provide additional heat input to the base material while still maintaining an inert atmosphere. Wetting action, depth of fusion, and travel speed can be improved over comparable argon levels. This advantage is most frequently utilized in the welding of heavier sections of aluminum, magnesium, and copper alloys.

Argon-Oxygen. The addition of a small amount of oxygen to a argon greatly stabilizes the welding arc, increases the filler metal droplet rate, lowers the spray transition current, and influences bead shape. The weld pool is more fluid and stays molten longer, allowing the metal to flow out toward the edges of the weld (Fig. 1a).

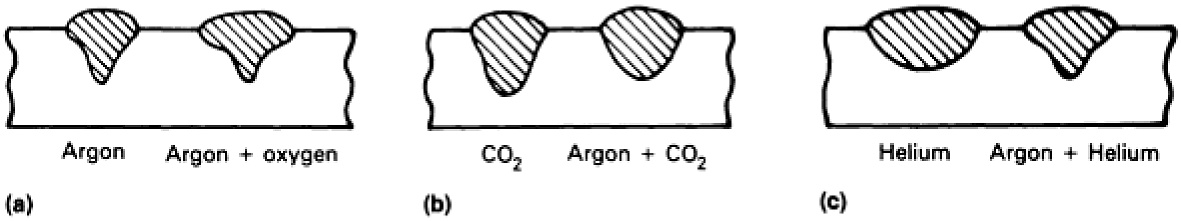

FIG. 1 EFFECT OF SHIELDING GAS BLENDS ON WELD PROFILE USING DIRECT CURRENT ELECTRODE POSITIVE (DCEP). (A) ARGON VERSUS ARGON-OXYGEN. (B) CARBON DIOXIDE VERSUS ARGON/CARBON DIOXIDE. (C) HELIUM VERSUS ARGON-HELIUM. SOURCE: REF 5

The most common blends contain 1, 2, 5, or 8% O_2 in argon. Increasing oxygen improves arc stability and makes higher travel speeds possible by enhanced puddle fluidity. Some increased alloy loss and a greater chance of undercut occur as the oxygen level is increased, especially beyond 5%.

Argon/carbon dioxide blends are primarily used for carbon and low-alloy steels and have limited use for stainless steels. The addition of CO_2 to argon produces results similar to the addition of oxygen, but is also broadens the penetration pattern as the CO_2 content is increased (Fig. 1b). Above a range of 18 to 20% CO_2, spraylike transfer can no longer be obtained; short-circuiting/globular transfer with somewhat increased spatter levels is found from this point up to approximately 50% CO_2 in argon.

The most common blends for spray transfer are argon plus 5, 8, 10, or 13 to 18% CO_2. With increased CO_2 content, the more fluid weld puddle permits higher weld travel speeds.

Mixes with higher carbon dioxide levels can also be used for short-circuiting transfer--commonly, argon plus 20 or 25% CO_2. Mixtures in this range provide an optimum droplet frequency for minimum spatter when small-diameter (0.9 and 1.2 mm, or 0.035 and 0.045 in.) wire is used.

Argon-Helium. Helium is often mixed with argon to obtain the advantages of both gases. These blends are primarily used for nonferrous base materials, such as aluminum, copper, and nickel alloys. Helium increases the heat input to the base material and thus is used for joining thick, thermally conductive plates. As the helium percentage increases, the arc voltage, spatter, and weld width-to-depth ratio increase (Fig. 1c).

The most common blends contain 25, 50, or 75% He in argon. The highest percentage of helium is used for joining thick (>50 mm, or 2 in.) plate material, especially aluminum and copper. Higher travel speeds can be obtained using helium-enhanced blends.

Argon/Oxygen/Carbon Dioxide. Mixtures containing three gas components are versatile because of their ability to operate in short-circuiting, globular, and spray-transfer modes. These blends are generally proprietary, and manufacturers' recommendations should be followed for their proper use. Blends of argon, carbon dioxide, and oxygen are generally used to join carbon and alloy steels.

Argon/Helium/Carbon Dioxide. Helium and carbon dioxide additions to argon increase the heat input to the base metal, which improves wetting, puddle fluidity, and weld bead profile. With helium plus CO_2 additions less than 40%, good spray transfer is obtained for carbon and low-alloy steel welding. Some increase in tolerance of base material surface contamination is also noted.

When the helium content exceeds 50 to 60%, transfer is restricted to short-circuiting and globular. Blends in which the CO_2 content is relatively low ($\leq$5%) are generally used for the joining of stainless steels without any loss in corrosion resistance.

Argon/Carbon Dioxide/Hydrogen. Three-part blends of this design are intended for the joining of austenitic stainless in the spray or short-circuiting transfer modes. Because of the addition of hydrogen, these blends should not be used for carbon steel. The carbon dioxide and hydrogen increase the heat input to the base material and improve bead shape characteristics, as well as promote higher welding travel speeds.

Shielding Gas Selection for FCAW

Carbon Dioxide. The majority of large-diameter (>1.6 mm, or $\frac{1}{16}$ in.) wires that use a shielding gas use carbon dioxide. Some smaller diameter wires are formulated to operate in 100% CO_2. The arcs are generally stable and provide a globular transfer over the usable operating range. Good performance over rust and mill scale on the plate surface is obtained with these large-diameter wires and CO_2 shielding.

Argon/Carbon Dioxide. A significant number of small-diameter ($\leq$1.6 mm, or $\frac{1}{16}$ in.) cored wires are shielded with the blends of argon with 15 to 50% CO_2. These blends provide better out-of-position weld puddle control versus that of CO_2. To obtain the best performance from a particular cored wire, check the manufacturer's product literature for the recommended gas blend.

Shielding Gas Selection for GTAW

Argon. The most commonly used gas for GTAW, argon exhibits low thermal conductivity, which produces a narrow, constricted arc column; this allows greater variations in arc length with minimal influence on arc power or weld bead shape. Its low ionization potential provides good arc starting characteristics and good arc stability using the direct current electrode negative (DCEN) power connection plus superior arc cleaning action and bead appearance when ac power is used. Argon is the most commonly selected gas for DCEN welding of most materials and ac manual welding of aluminum.

Helium. The high thermal conductivity and ionization potential of helium make it suitable for the high-current joining of heavy sections of heat-conductive materials such as aluminum. Helium increases the penetration of the weld as well as its width. It also allows the use of higher weld travel speeds.

Argon-Helium. Blends of argon and helium are selected to increase the heat input to the base material while maintaining favorable arc stability and superior arc starting characteristics. Blends of 25, 50, and 75% He in argon are commonly used.

Argon-Hydrogen. Hydrogen is added to argon to enhance its thermal properties. The slightly reducing atmosphere improves weld puddle wetting and reduces some surface oxides to produce a cleaner weld surface. To minimize problems associated with arc starting, additions of hydrogen are generally limited to 5 to 15%. These blends are primarily used to join some stainless steels, nickel, and nickel alloys. These mixtures should not be used to join alloy steels; delayed weld cracking may result.

Argon/2-5H_2 is used in manual welding applications on materials thicker than 1.6 mm ($\frac{1}{16}$ in.). Additions of 10 to 15% H_2 are used in mechanized applications, such as those found in the manufacture of stainless steel tubing.

Caution: Special safety precautions are required when mixing argon and hydrogen. Do not attempt to mix argon and hydrogen from separate cylinders. Always purchase ready-mixed hydrogen blends from a qualified supplier.

Shielding Gas Selection for PAW

The physical configuration of the PAW system requires the use of two gases: a "plasma" or orifice gas and a shielding gas. The primary role of the plasma gas, which exits the torch through the center orifice, is to control arc characteristics and shield the electrode. The shielding gas, introduced around the periphery of the arc, shields or protects the weld area. In many applications, the shielding gas is also partially ionized to enhance the performance of the plasma gas.

Low-Current (<100 A) PAW. Argon is the preferred plasma gas for low-current PAW because of its low ionization potential, which ensures easy and reliable arc starting. Argon-hellium mixtures are used for some applications requiring higher heat inputs. The choice of the shielding gas depends on the type and thickness of the base material. Recommendations can be found in Table 4.

TABLE 4 RECOMMENDED GUIDELINES FOR SELECTING PAW SHIELDING GASES

MATERIAL	**THICKNESS**		**MODE OF PENETRATION**	
	MM	**IN.**	**KEYHOLE TECHNIQUE(C)**	**MELT-IN TECHNIQUE(D)**
LOW-CURRENT PAW(A)				
ALUMINUM COPPER	<1.6	$<\frac{1}{16}$	NOT RECOMMENDED	ARGON HELIUM
	[GES]1.6	[GES] $\frac{1}{16}$	HELIUM	HELIUM
CARBON STEEL	<1.6	$<\frac{1}{16}$	NOT RECOMMENDED	ARGON HELIUM 75AR-25HE
	[GES]1.6	[GES] $\frac{1}{16}$	ARGON 75HE-25AR	ARGON 75HE-25AR
LOW-ALLOY STEEL	<1.6	$<\frac{1}{16}$	NOT RECOMMENDED	ARGON HELIUM AR-1.5H_2
	[GES]1.6	[GES] $\frac{1}{16}$	ARGON 75HE-25AR AR-1.5H_2	ARGON HELIUM AR-1.5H_2
STAINLESS STEEL, NICKEL ALLOYS	<1.6	$<\frac{1}{16}$	ARGON 75HE-25AR AR-1.5H_2	ARGON HELIUM AR-1.5H_2
	[GES]1.6	[GES] $\frac{1}{16}$	ARGON 75HE-25AR AR-1.5H_2	ARGON HELIUM AR-1.5H_2
HIGH-CURRENT PAW(B)				
ALUMINUM	<6.4	$<\frac{1}{4}$	ARGON	ARGON 75-25AR
	[GES]6.4	[GES] $\frac{1}{4}$	HELIUM	HELIUM 75-25AR
COPPER	<2.4	$<\frac{3}{32}$	NOT RECOMMENDED	HELIUM
CARBON STEEL; LOW-ALLOY STEEL	<3.2	$<\frac{1}{8}$	ARGON	ARGON
	[GES]3.2	[GES]	ARGON	75HE-25AR

		$\frac{1}{8}$		
STAINLESS STEEL; NICKEL ALLOYS	<3.2	$<\frac{1}{8}$	ARGON AR-5H_2	ARGON
	[GES]3.2	[GES] $\frac{1}{8}$	ARGON AR-5H_2	75-HE-25AR

Source: Ref 6, 7

(A) GAS SELECTIONS SHOWN ARE FOR SHIELDING GAS ONLY. ORIFICE GAS IN ALL CASES IS ARGON.

(B) GAS SELECTIONS SHOWN ARE FOR BOTH THE ORIFICE AND SHIELDING GAS.

(C) PROPERLY BALANCED GAS FLOW RATES PRODUCE COMPLETE JOINT PENETRATION BY FORMING A SMALL WELD POOL WITH A HOLE PENETRATING COMPLETELY THROUGH THE BASE METAL. AS THE PLASMA TORCH IS MOVED, METAL MELTED BY THE ARC IS FORCED TO FLOW AROUND THE PLASMA STEAM PRODUCING THE "KEYHOLE" TO THE REAR WHERE THE WELD POOL IS FORMED AND SOLIDIFIED.

(D) CONVENTIONAL FUSION WELDING SIMILAR TO THAT DONE WITH GTAW.

High-Current ([ges]100 A) PAW. The choice of gas used for high-current PAW also depends on the material to be welded. In almost all cases, the shielding gas is the same as the orifice gas. Again, argon is suitable for welding all metals, but it does not necessarily produce optimum results. Depending on the welding mode used (keyhole or melt-in), the optimum gas blend will vary. Table 4 lists gases recommended on the basis of materials joined.

References cited in this section

3. "SPECIFICATION FOR SHIELDING GASES," AWS A5.32-9X, JULY 1991 (DRAFT)
4. "GASES FOR GAS-SHIELDED ARC WELDING AND CUTTING," EUROPEAN STANDARD EN 439, DEC 1990 (DRAFT)
5. H.B. CARY, *MODERN WELDING TECHNOLOGY*, 2ND ED., PRENTICE-HALL, 1989
6. N.E. LARSON AND W.F. MEREDITH, *SHIELDING GAS SELECTION MANUAL*, UNION CARBIDE INDUSTRIAL GASES TECHNOLOGY CORP., 1990, P 30-32
7. N.E. LARSON AND W.F. MEREDITH, *SHIELDING GAS SELECTION MANUAL*, UNION CARBIDE INDUSTRIAL GASES TECHNOLOGY CORP., 1990, P 18-19

Shielding Gases for Welding

Kevin A. Lyttle, Praxair, Inc.

Influence of Shielding Gas on Weld Mechanical Properties

The mechanical properties of a weld are dependent on specific characteristics of the shielding gas. When the gas blend is totally inert, the effects are less pronounced and are derived more indirectly. In this case, the shielding gas affects penetration and solidification, which can influence the microstructure of the resulting weld.

When the shielding gas contains active components, such as oxygen or carbon dioxide, the influence is direct and more substantial. The oxygen potential of the shielding gas influences the amount of surface slag, the fume emission rate, the fluidity of the weld puddle, and the mechanical properties (both strength and toughness) of the weld metal. A number of empirical formulas have been developed to estimate the oxygen potential of a gas blend (Ref 8); the differences lie in their treatment of the CO_2 component of the gas mix. Because parameters other than gas composition can affect the oxygen level of the weld metal (e.g., welding speed), most of the recently developed formulas are also welding parameter and material specific. Their importance centers on how the oxidation potential is linked to elemental loss of silicon and manganese in the weld metal, to weld metal oxygen content, and to weld mechanical properties.

Because a relatively complex relationship exists between the loss of alloying elements, the composition of the shielding gas, and the mechanical properties of the resulting weld metal it is difficult to select an optimum gas blend that will work well with all types of wires. This influence is most noticeable in the gas metal-arc welding of carbon steel. Figure 2 (Ref 8) and Table 5 (Ref 9) illustrate the effect of the shielding gas on impact strength and tensile strength. In general, as the oxidation potential of the shielding gas increases, the toughness and the tensile strength of the weld deposit decrease. Because of their lower oxidizing gas content, argon blends will generally produce weld properties superior to those obtained by shielding with CO_2 only. There appears to be an "optimum" oxygen content, as too low an oxygen level can also be detrimental to toughness.

TABLE 5 EFFECT OF SHIELDING GAS ON CARBON, MANGANESE, AND SILICON LOSSES AND ON WELD STRENGTH VALUES

SHIELDING GAS[(A)]	**ULTIMATE TENSILE STRENGTH**		**YIELD STRENGTH**		**ELONGATION, %**	**WELD METAL COMPOSITION, %[(B)]**		
	MPA	**KSI**	**MPA**	**KSI**		**C**	**MN**	**SI**
AR-10CO_2	640	92.9	544	79.0	25.7	0.09	1.43	0.72
AR-18CO_2	620	90.0	522	75.8	26.8	0.09	1.37	0.70
AR-5CO_2-4O_2	610	88.5	472	68.5	28.1	0.08	1.32	0.67
AR-25CO_2	601	87.2	505	73.3	29.3	0.09	1.30	0.65
AR-12O_2	591	85.8	510	74.0	27.5	0.06	1.20	0.60
CO_2	594	86.2	487	70.7	27.8	0.10	1.21	0.62

Source: Ref 9

(A) GASES LISTED IN ORDER OF INCREASING OXIDATION POTENTIAL.

(B) BASE WIRE COMPOSITION: 0.115% C, 1.53% MN, 0.98% SI.

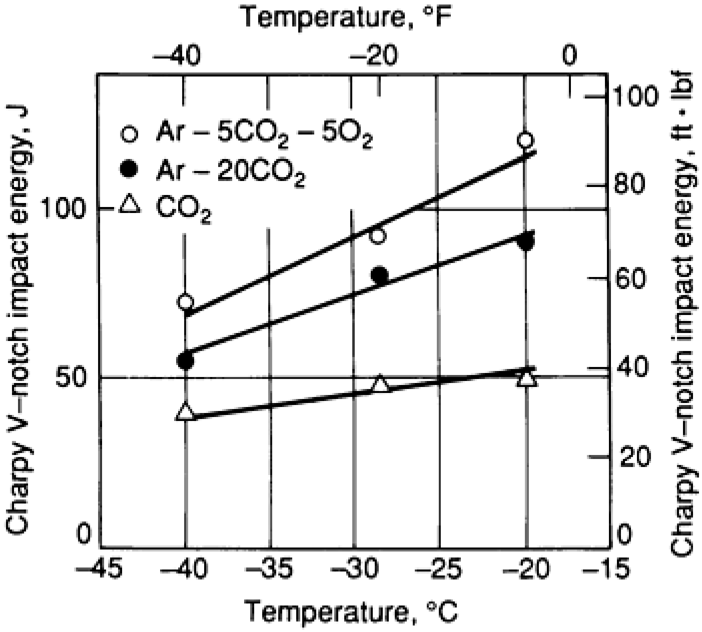

FIG. 2 PLOT OF WELD METAL IMPACT ENERGY VERSUS TEST TEMPERATURE AS A FUNCTION OF SHIELDING GAS COMPOSITION. SOURCE: REF 8

References cited in this section

8. N. STENBACKA AND K.A. PERSSON, SHIELDING GASES FOR GAS METAL ARC WELDING, *WELD.J.*, VOL 68, NOV 1989, P 41-47
9. H.U. POMASKA, "SHIELDING GASES FOR ARC WELDING AND CUTTING," SONDERDRUCK 48, LINDE AG, LINDE TECHNISCHE GASES, HOLRIEGELSKREUTH, GERMANY

Shielding Gases for Welding

Kevin A. Lyttle, Praxair, Inc.

Shielding Gas and Fume Generation

The shielding gas used in solid and flux-cored wire welding affects the rate at which fumes are produced during welding, as well as the composition of the fumes. The type of shielding gas also affects the composition of the pollutant gases in the welding area. The fumes and gases generated in some welding applications impact on health and safety; therefore, it is important for the user to review the Material Safety Data Sheets and precautionary labeling provided by the manufacturers and suppliers of welding consumables. Welding safety and health information can be also be found in this Handbook and in ANSI Z49.1, "Safety in Welding and Cutting."

Fume generation rates are generally highest when CO_2 shielding is used. Inert gas blends containing CO_2 produce higher fume levels than those containing only oxygen additions. For solid wire welding at the same weld metal deposition rate, argon blends typically generate significantly less fumes than when CO_2 only is used (Ref 10, 11, 12). Figure 3 (Ref 13) illustrates the dependence of the fume generation rate on shielding gas composition as a function of current level in GMAW. Similar trends can be seen when using gas-shielded flux-cored wires.

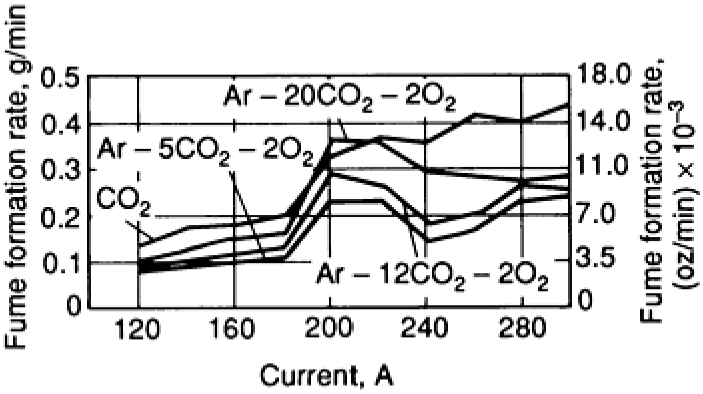

FIG. 3 PLOT OF FUME FORMATION RATE VERSUS CURRENT FOR MILD STEEL SOLID WIRE USING SELECTED SHIELDING GASES. SOURCE: REF 13

Pollutant gases must be considered when arc welding. The gases of greatest interest arc carbon monoxide, ozone, and nitrogen oxides. When CO_2 is used for shielding in confined spaces, CO can be a potential problem. Ozone can be a concern when high-energy gas-shielded welding is conducted, particularly on aluminum and stainless steel plate. It is produced in the immediate arc area as well as in the surrounding environment. Oxides of nitrogen can be present in some plasma welding and cutting applications (Ref 14).

References cited in this section

10. R.F. HEILE AND D.C. HILL, PARTICULATE FUME GENERATION IN ARC WELDING PROCESSES, *WELD.J.*, VOL 54, JULY 1975, P 201S-210S
11. H. PRESS AND W. FLORIAN, CONTRIBUTION ON THE FORMATION OF TOXIC SUBSTANCES IN GAS SHIELDED ARC WELDING AND DETERMINATION OF DESIRED EXHAUSTION CAPACITY, *IIW PROCEEDINGS OF COLLOQUIUM ON WELDING AND HEALTH*, PORTUGAL, JULY 1980
12. B. HAAS, INFLUENCE OF PROCESS SPECIFIC WELDING PARAMETERS IN FUME GENERATION IN SOLID WIRE, GMA/MAG WELDING, *IIW PROCEEDINGS OF COLLOQUIUM ON WELDING SAFETY AND HEALTH*, PORTUGAL, JULY 1980
13. D.E. HILTON AND P.N. PLUMRIDGE, PARTICULATE FUME GENERATION DURING GMAW AND GTAW, *WELD. MET. FABR.*, VOL 59 (NO. 10), DEC 1991, P 555-560
14. N. JENKINS, J. MORETON, P.J. OAKLEY, AND S.M. STEVENS, *WELDING FUMES SOURCES, CHARACTERISTICS, CONTROL*, VOL 2, THE WELDING INSTITUTE, CAMBRIDGE, UNITED KINGDOM, 1981, P 272

Self-Shielded Flux-Cored Arc Welding

The self-shielded flux-cored wire welding process employs a continuous wire electrode that requires no external shielding. These cored wires generate protective shielding gases from components in the core material similar to those found in coated electrodes. The unique feature of this class of consumables is that they rely only partly on the exclusion of air from the molten electrode and weld pool to produce a quality deposit. In addition to deoxidizers, they contain denitriders to react with nitrogen that may be entrained in the molten metal. The less a given consumable relies on shielding and the more it relies on "killing" to control nitrogen and produce sound weld metal, the less that consumable will be affected by cross air currents and side winds.

Through careful control of welding parameters and the proper balance of core constituents, good-quality weld metal can be obtained under poor welding conditions. Self-shielded flux-cored wires are especially suited to welding outdoors, where it is difficult to provide acceptable external gas shielding. Both light- and heavy-gage materials have been successfully joined with these electrodes.

References

1. N.E. LARSON AND W.F. MEREDITH, *SHIELDING GAS SELECTION MANUAL*, UNION CARBIDE INDUSTRIAL GASES TECHNOLOGY CORP., 1990, P 10
2. N.E. LARSON AND W.F. MEREDITH, *SHIELDING GAS SELECTION MANUAL*, UNION CARBIDE INDUSTRIAL GASES TECHNOLOGY CORP., 1990, P 11
3. "SPECIFICATION FOR SHIELDING GASES," AWS A5.32-9X, JULY 1991 (DRAFT)
4. "GASES FOR GAS-SHIELDED ARC WELDING AND CUTTING," EUROPEAN STANDARD EN 439, DEC 1990 (DRAFT)

5. H.B. CARY, *MODERN WELDING TECHNOLOGY*, 2ND ED., PRENTICE-HALL, 1989
6. N.E. LARSON AND W.F. MEREDITH, *SHIELDING GAS SELECTION MANUAL*, UNION CARBIDE INDUSTRIAL GASES TECHNOLOGY CORP., 1990, P 30-32
7. N.E. LARSON AND W.F. MEREDITH, *SHIELDING GAS SELECTION MANUAL*, UNION CARBIDE INDUSTRIAL GASES TECHNOLOGY CORP., 1990, P 18-19
8. N. STENBACKA AND K.A. PERSSON, SHIELDING GASES FOR GAS METAL ARC WELDING, *WELD.J.*, VOL 68, NOV 1989, P 41-47
9. H.U. POMASKA, "SHIELDING GASES FOR ARC WELDING AND CUTTING," SONDERDRUCK 48, LINDE AG, LINDE TECHNISCHE GASES, HOLRIEGELSKREUTH, GERMANY
10. R.F. HEILE AND D.C. HILL, PARTICULATE FUME GENERATION IN ARC WELDING PROCESSES, *WELD.J.*, VOL 54, JULY 1975, P 201S-210S
11. H. PRESS AND W. FLORIAN, CONTRIBUTION ON THE FORMATION OF TOXIC SUBSTANCES IN GAS SHIELDED ARC WELDING AND DETERMINATION OF DESIRED EXHAUSTION CAPACITY, *IIW PROCEEDINGS OF COLLOQUIUM ON WELDING AND HEALTH*, PORTUGAL, JULY 1980
12. B. HAAS, INFLUENCE OF PROCESS SPECIFIC WELDING PARAMETERS IN FUME GENERATION IN SOLID WIRE, GMA/MAG WELDING, *IIW PROCEEDINGS OF COLLOQUIUM ON WELDING SAFETY AND HEALTH*, PORTUGAL, JULY 1980
13. D.E. HILTON AND P.N. PLUMRIDGE, PARTICULATE FUME GENERATION DURING GMAW AND GTAW, *WELD. MET. FABR.*, VOL 59 (NO. 10), DEC 1991, P 555-560
14. N. JENKINS, J. MORETON, P.J. OAKLEY, AND S.M. STEVENS, *WELDING FUMES SOURCES, CHARACTERISTICS, CONTROL*, VOL 2, THE WELDING INSTITUTE, CAMBRIDGE, UNITED KINGDOM, 1981, P 272

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

Introduction

SOLID-STATE TRANSFORMATIONS occurring in a weld are highly nonequilibrium in nature and differ distinctly from those experienced during casting, thermomechanical processing, and heat treatment. This discussion will primarily focus on the welding metallurgy of fusion welding of steels and attempt to highlight the fundamental principles that form the basis of many of the recent developments in steels and consumables for welding. Accordingly, examples are largely drawn from the well-known and relatively well-studied case of ferritic steel weldments to illustrate the special physical metallurgical considerations brought about by the weld thermal cycles and by the welding environment. Because of space limitations, only a very brief discussion is included on welds in other alloy systems such as stainless steels and aluminum-base, nickel-base, and titanium-base alloys. For detailed information on how the principles explained in the first part of this article (using steel weldments as an example) are applicable to other materials, the reader is referred to the Sections where individual alloy systems are discussed (see "Selection of Stainless Steels," "Selection of Nonferrous Low-Temperature Materials," "Selection of Nonferrous High-Temperature Materials," and "Selection of Nonferrous Corrosion-Resistant Materials" in this Volume).

A concise method of describing the transformation behavior of a steel is by a continuous cooling transformation (CCT) diagram (Fig. 1). However, a conventional CCT diagram such as the one shown in Fig. 1 cannot be used to accurately describe the transformation behavior in a weldment of the same material because weld thermal cycles are very different from those used for generating conventional CCT diagrams.

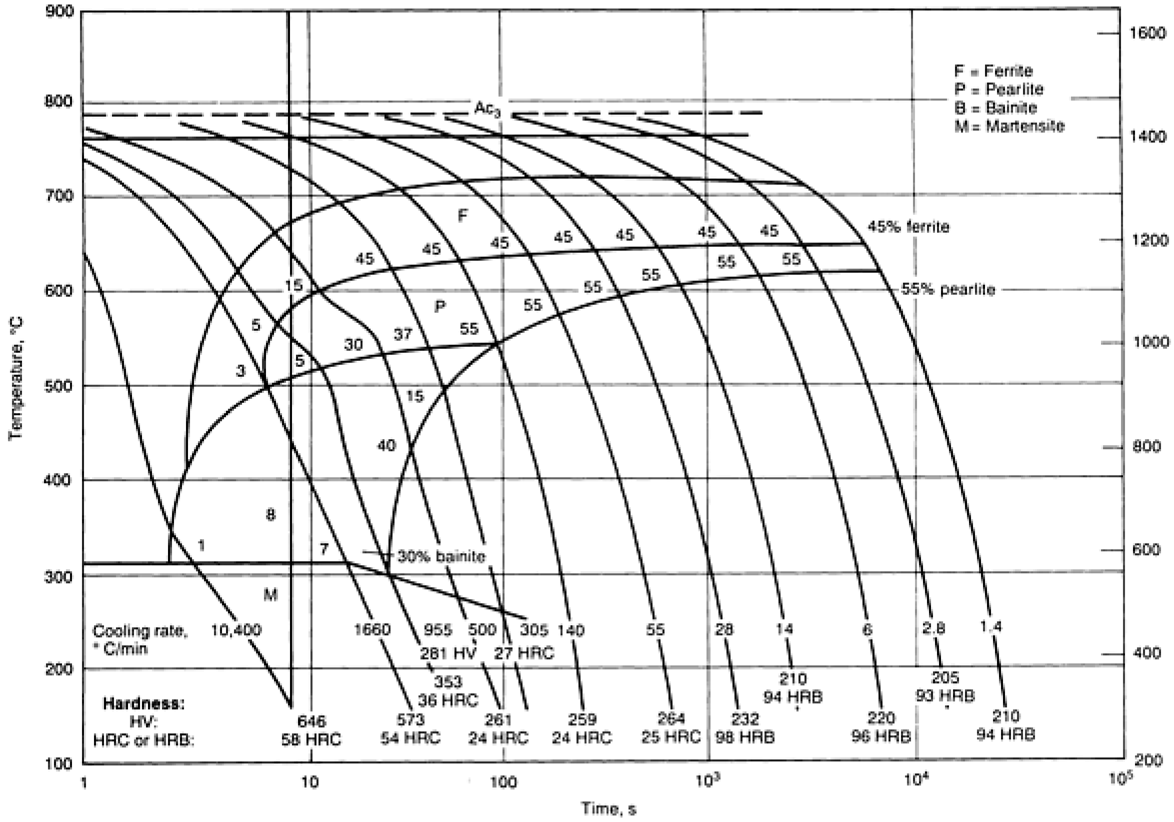

FIG. 1 CONVENTIONAL CCT DIAGRAM FOR AISI 1541 (0.39C-1.56MN-0.21SI-0.24S-0.010P) PLAIN CARBON STEEL WITH ASTM NUMBER 8 GRAIN SIZE AND AUSTENITIZATION AT 980 °C (1800 °F). FOR EACH OF THE COOLING CURVES IN THE PLOT (GIVEN IN TERMS OF °C/MIN), THE TRANSFORMATION START AND END TEMPERATURES GIVEN BY THE CCT CURVES, THE AMOUNT OF EACH TRANSFORMATION PRODUCT, AND THE HARDNESS OF THE FINAL STRUCTURE ARE SHOWN. AC_3, 788 °C (1450 °F); AC_1, 716 °C (1321 °F). F, FERRITE; P, PEARLITE; B, BAINITE; M, MARTENSITE. SOURCE: REF 1

Acknowledgements

The author would like to convey his immense gratitude to the late Professor K. Easterling, University of Exeter, U.K., for giving him the opportunity to write this article and for his constant support and encouragement. The author thanks Dr. Bengt Loberg and Professor Richard Warren of the Department of Engineering Materials, University of Luleå, for their active encouragement. The author would also like to thank Dr. H.K.D.H. Bhadeshia, University of Cambridge, U.K., and Dr. K. Sampath, Concurrent Technologies Corporation, Johnstown, PA, for valuable technical consultation.

Reference

1. *ATLAS OF ISOTHERMAL TRANSFORMATION AND COOLING TRANSFORMATION DIAGRAMS* AMERICAN SOCIETY FOR METALS, 1977

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

Special Factors Affecting Transformation Behavior in a Weldment

Several aspects of the weld thermal cycle and weld segregation should be considered because of their effect on the transformation upon cooling:

- PEAK TEMPERATURES REACHED IN THE HEAT-AFFECTED ZONE (HAZ) CAN BE VERY MUCH HIGHER THAN THE AC_3 TEMPERATURE (THAT IS, THE TEMPERATURE AT WHICH TRANSFORMATION OF FERRITE TO AUSTENITE IS COMPLETED DURING HEATING). THE HEATING RATES ARE VERY HIGH, AND THE TIMES SPENT AT HIGH TEMPERATURE ARE ONLY OF THE ORDER OF A FEW SECONDS.
- THE TEMPERATURE GRADIENT IN THE HAZ IS VERY STEEP, AND THIS COMPLICATES THE PROBLEM OF STUDYING IN *IN SITU* TRANSFORMATIONS IN THE HAZ DURING WELDING (REF 2).
- DURING SOLIDIFICATION OF THE WELD METAL, ALLOYING AND IMPURITY ELEMENTS TEND TO SEGREGATE EXTENSIVELY TO THE INTERDENDRITIC OR INTERCELLULAR REGIONS UNDER THE CONDITIONS OF RAPID COOLING. ALSO, THE PICKUP OF ELEMENTS LIKE OXYGEN BY THE MOLTEN WELD POOL LEADS TO THE ENTRAPMENT OF OXIDE INCLUSIONS IN THE SOLIDIFIED WELD. THESE INCLUSIONS THEN SERVE AS HETEROGENEOUS NUCLEATION SITES AND CAN SUBSTANTIALLY INFLUENCE THE KINETICS OF SUBSEQUENT SOLID STATE TRANSFORMATIONS. ACCORDINGLY, THE WELD METAL TRANSFORMATION BEHAVIOR IS QUITE DIFFERENT FROM THAT OF THE BASE METAL, EVEN THOUGH THE NOMINAL CHEMICAL COMPOSITION HAS NOT BEEN SIGNIFICANTLY CHANGED BY THE WELDING PROCESS (REF 3). MOST OF THE CCT DIAGRAMS (APPLICABLE TO THE WELD METAL) HAVE BEEN GENERATED BY REHEATING THE AS-DEPOSITED WELD METAL (REF 4). ONE OF THE LIMITATIONS OF THESE DIAGRAMS IS THAT THEY ARE STRICTLY APPLICABLE ONLY TO THE HIGH-TEMPERATURE REHEATED ZONE OF MULTIPASS WELDS, BECAUSE THE INITIAL MICROSTRUCTURE AT HIGH TEMPERATURES IS NOT CHARACTERISTIC OF THAT DEVELOPED FROM THE LIQUID PHASE.
- WELDING MAY BE CARRIED OUT IN SEVERAL PASSES, AND THIS MAY RESULT IN THE SUPERPOSITION OF SEVERAL DIFFERENT HEATING AND COOLING CYCLES AT ONE POINT, EACH OF THESE CYCLES HAVING THE CHARACTERISTICS NOTED ABOVE.
- SOLIDIFICATION OF THE WELD METAL IS ACCOMPANIED BY SHRINKAGE, AND THE ANISOTHERMAL CONDITIONS ALREADY EMPHASIZED CAUSE DEFORMATION. THE THERMAL CYCLES ARE THEREFORE ACTING ON METAL THAT IS SUBJECTED TO MECHANICAL STRESSES AT THE SAME TIME.

The essential differences between weld thermal cycles and then thermal cycles used for generating a conventional CCT diagram are summarized in Fig. 2. Figure 2(a) shows thermal cycles which involve a slow heating rate, soak at a temperature just above the Ac_3 temperature, and various constant cooling rates. Instead of a constant cooling rate, some investigators (Ref 5) have used cooling curves according to Newton's law of cooling:

$$T \propto E^{-T} \qquad \textbf{(EQ 1)}$$

where T is the temperature and t is the time. Others have used cooling curves corresponding to the mid-radial position of cylindrical bars of different diameters when cooled in air or when quenched in different media. The weld thermal cycles shown in Fig. 2(b) are very different, and this is why a conventional CCT diagram can give only an approximate idea of the transformation behavior in the HAZ of a weldment.

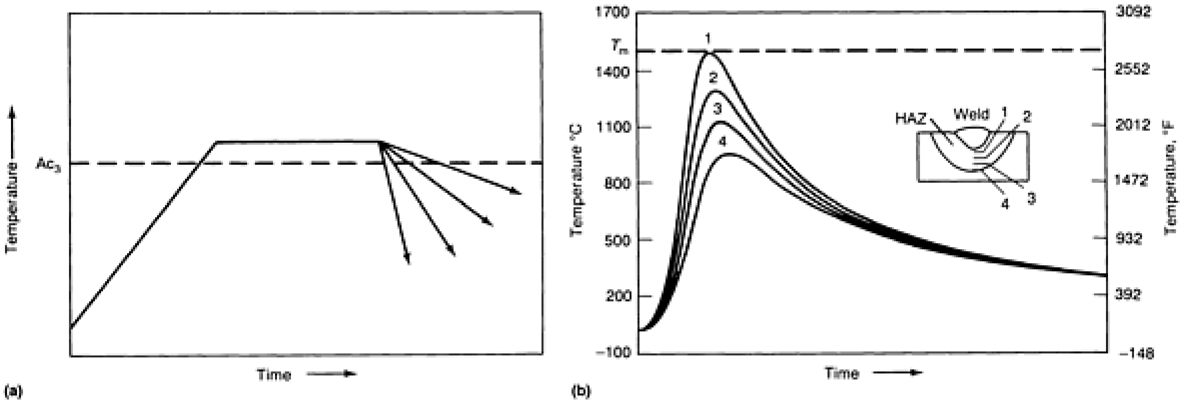

FIG. 2 GRAPHS TO SHOW DIFFERENCES IN THERMAL CYCLES. (A) THERMAL CYCLES USED TO GENERATE A CONVENTIONAL CCT DIAGRAM. (B) WELD THERMAL CYCLES. THE NUMBERS IN (B) CORRESPOND TO LOCATIONS INDICATED IN THE HAZ. NOTE THE CORRESPONDENCE BETWEEN THE THERMAL CYCLES IN (A) AND THOSE IN FIG. 1

References cited in this section

2. O.M. AKSELSEN AND T. SIMONSEN, TECHNIQUES FOR EXAMINING TRANSFORMATION BEHAVIOUR IN WELD METAL AND HAZ--A STATE OF ART REVIEW, *WELD. WORLD,* VOL 25 (NO. 1/2), 1987, P 26-34
3. O. GRONG AND D.K. MATLOCK, MICROSTRUCTURAL DEVELOPMENT IN MILD AND LOW ALLOY STEEL WELD METALS, *INT. MET. REV.,* VOL 31 (NO. 1), 1986, P 27-48
4. P.L. HARRISON AND R.A. FARRAR, APPLICATION OF CONTINUOUS COOLING TRANSFORMATION DIAGRAMS FOR WELDING OF STEELS, *INT. MATER. REV.,* VOL 34 (NO. 1), 1989, P 35-51
5. G.T. ELDIS, A CRITICAL REVIEW OF DATA SOURCES FOR ISOTHERMAL TRANSFORMATION AND CONTINUOUS COOLING TRANSFORMATION DIAGRAMS, *HARDENABILITY CONCEPTS WITH APPLICATION TO STEEL,* D.V. DOANE AND J.S. KIRKALDY, ED., TMS-AIME, 1978, P 126-157

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

HAZ of a Single-Pass Weld

Peak Temperature-Cooling Time Diagrams. The gradient in microstructure than can be obtained in a single-pass weld is shown in Fig. 3. High peak temperatures in the HAZ just adjacent to the fusion line cause coarsening of the austenite (γ) grains, and this in turn increases the hardenability of this region compared to other regions. Because each of the subzones shown in Fig. 3 occurs in a small volume, it is difficult to study the transformation behavior of individual regions by *in situ* methods (Ref 2). It is more convenient to obtain information about the microstructural and property changes in the HAZ by weld simulation (Ref 6, 7). A thermal cycle simulator (TCS) is used to reproduce the thermal cycle corresponding to a point in the HAZ in a large volume (usually a specimen has the Charpy test specimen dimensions). It is possible to program the required thermal cycle in the TCS so that the peak temperature and the cooling rate can be varied independently. It is thus possible to obtain information about the microstructural changes in the HAZ for a wide range of welding parameters.

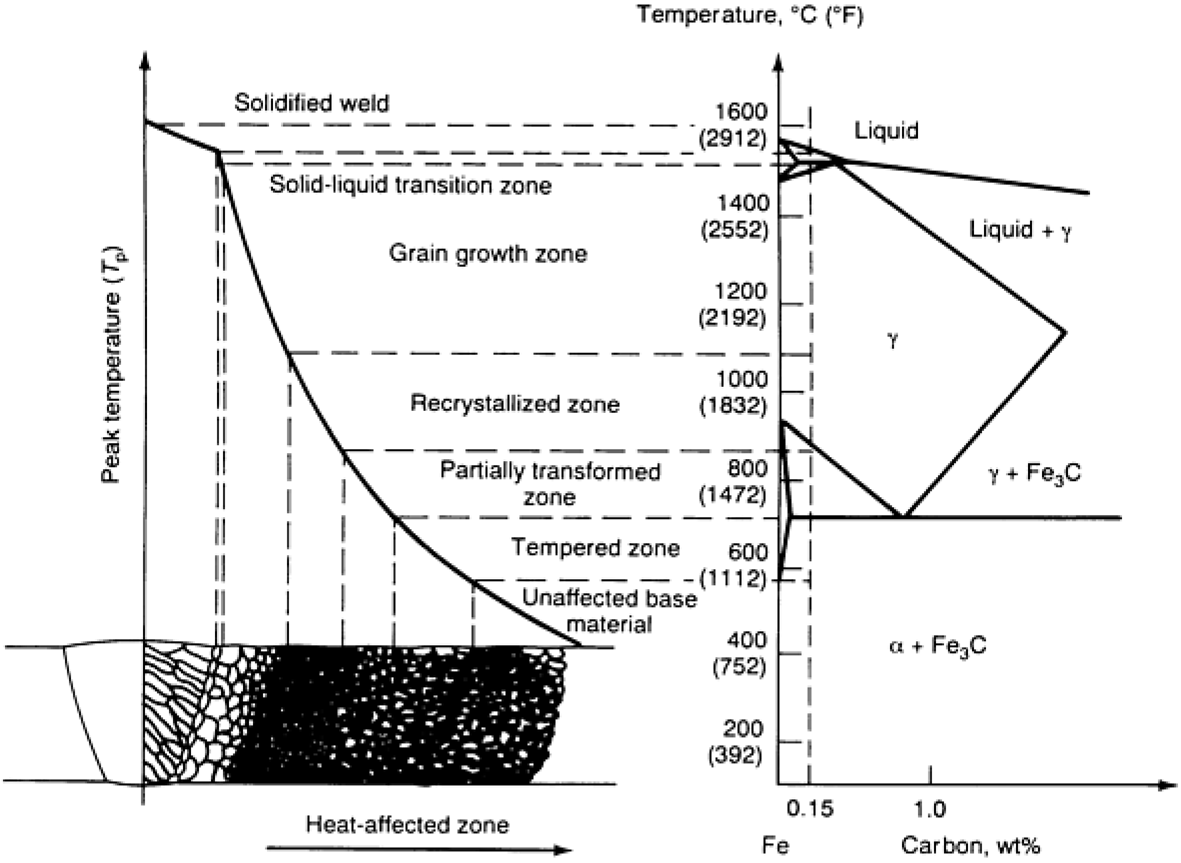
Temperature, °C (°F)
Peak temperature (T_p)
Solidified weld
Solid-liquid transition zone
Grain growth zone
Recrystallized zone
Partially transformed zone
Tempered zone
Unaffected base material
Heat-affected zone
1600 (2912)
1400 (2552)
1200 (2192)
1000 (1832)
800 (1472)
600 (1112)
400 (752)
200 (392)
Liquid
Liquid + γ
γ
γ + Fe$_3$C
α + Fe$_3$C
0.15
1.0
Fe
Carbon, wt%

FIG. 3 SCHEMATIC SHOWING VARIOUS SUBZONES THAT CAN FORM IN THE HAZ OF A CARBON STEEL CONTAINING 0.15% C. SOURCE: REF 6

Figure 4 shows how a change in the peak temperature of the thermal cycle affects the CCT characteristics of a steel. The well-known effect of a larger γ grain size (due to a higher peak temperature) in increasing the hardenability of the steel is seen. To present the information about the CCT behavior for a number of peak temperatures (see Fig. 2b and 3), it is more convenient to adopt the scheme shown in Fig. 5. In this peak temperature-cooling time (PTCT) diagram (Ref 8, 9), each point represents a weld thermal cycle with a peak temperature, T_p, given by the ordinate and the cooling time, $\Delta t_{8\text{-}5}$ (that is, required for cooling from 800 to 500 °C, or 1470 to 930 °F), given by the abscissa. A microstructural constituent or a combination of two or more constituents is shown to occur over an area in the diagram. The upward slope in the boundary between two areas is consistent with the information presented earlier in Fig. 4 that the hardenability increases with an increase in the peak temperature of the thermal cycle. Hardness and C_v transition temperatures are also shown in the diagram corresponding to different thermal cycles.

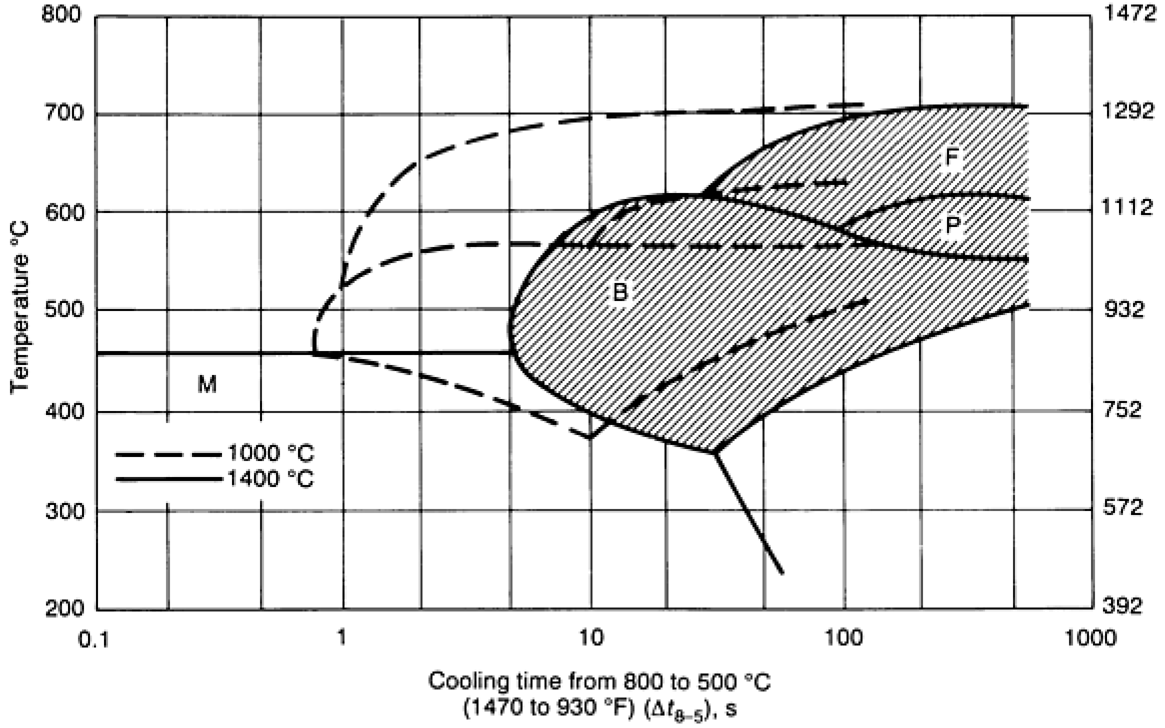

FIG. 4 EFFECT OF A CHANGE IN THE PEAK TEMPERATURE OF THE WELD THERMAL CYCLE (FROM 1000 TO 1400 °C, OR 1830 TO 2550 °F) ON THE CCT CHARACTERISTICS. M, MARTENSITE; B, BAINITE; P, PEARLITE; F, FERRITE. SOURCE: REF 8

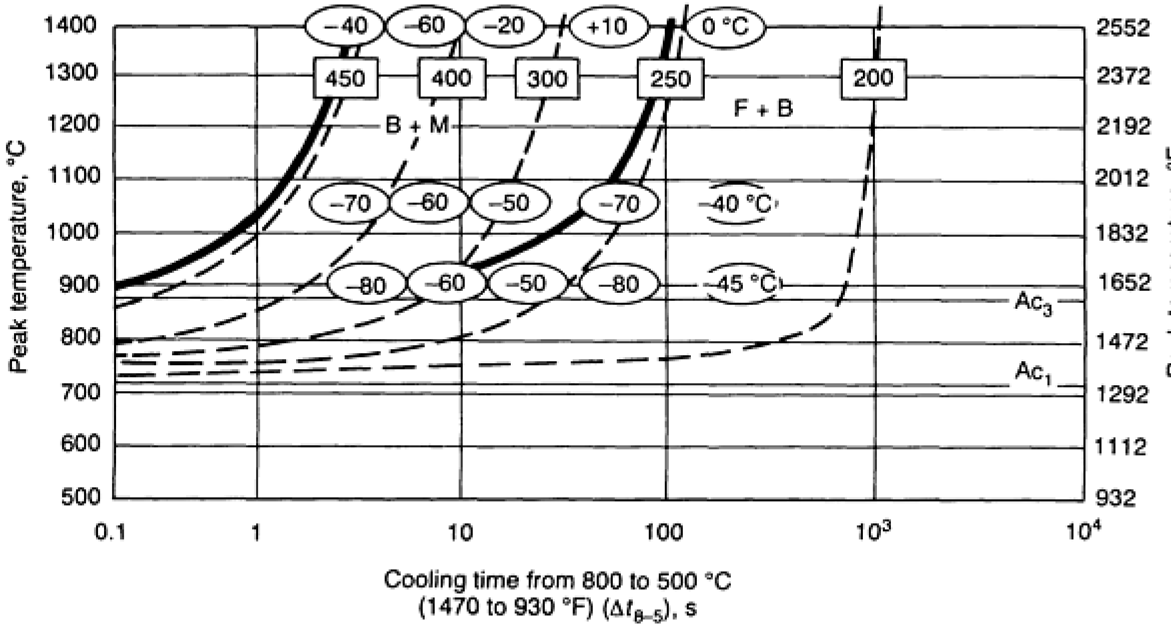

FIG. 5 TYPICAL PEAK TEMPERATURE VERSUS COOLING TIME DIAGRAM, SHOWING THE EFFECTS OF THESE PARAMETERS OF A WELD THERMAL CYCLE ON THE FINAL TRANSFORMATION PRODUCTS, ON HARDNESS, AND CHARPY V-NOTCH IMPACT ENERGY. M, MARTENSITE; B, BAINITE; F, FERRITE. FIGURES IN SQUARES INDICATE THE HARDNESS (30 HV); FIGURES IN OVALS INDICATE THE 21 J (16 FT · LBF) TRANSITION TEMPERATURES FOR CHARPY V-NOTCH IMPACT SPECIMENS SUBJECTED TO A SPECIFIC THERMAL CYCLE. SOURCE: REF 9

It may appear strange that the CCT diagram shown in Fig. 4 is plotted with Δt_{8-5} as the abscissa instead of time as in Fig. 1. Historically, the practice of thinking in terms of a cooling time began when a need was recognized (Ref 5) to compare the CCT behavior of specimens subjected to different types of cooling curves--Jominy specimens; cylindrical bars cooled in air, water, or oil; dilatometer specimens cooled at a constant or a Newtonian rate; and so on. Initially, the cooling rate at 704 °C (1300 °F) was used as a criterion to judge the equivalency of these specimens in terms of microstructure and hardness. Because transformation takes place at lower temperatures in many steels, the half cooling time (Ref 5) was found to be a better equivalency criterion. The half cooling time was defined as the time to cool from Ac_3 to a temperature that lies midway between Ac_3 and room temperature. For the sake of general applicability, this was later modified to the Δt_{8-5} criterion, and CCT diagrams began to be plotted with Δt_{8-5} as the abscissa. It must be noted that using Δt_{8-5} as an equivalency criterion is just an expedient solution that does not have strict theoretical justification (refer, for example, to the additivity principle in Ref 5).

A special significance of using Δt_{8-5} as the abscissa in Fig. 5 is that it is almost constant (for $T_p > 900$ °C, or 1650 °F) in the whole of the HAZ. This can be seen from Ref 6, where the expressions given for Δt_{8-5} do not contain the distance from the weld centerline as a factor. The derivation of these expressions is based on the assumption that the time to cool to 800 °C (1470 °F) is far greater than the time to reach the peak temperature, so the Δt_{8-5} can be taken to be the same for the whole of the weld and the HAZ. This can be intuitively rationalized by observing that the thermal cycles in Fig. 2(b) are such that the curves are approximately parallel below 800 °C (1470 °F). The constancy of Δt_{8-5} in the HAZ means that the gradient in microstructure (in terms of the final transformation products from austenite) is mainly due to a variation in the peak temperature. By drawing a vertical line in Fig. 5 at a value of Δt_{8-5} corresponding to a given heat input and preheating temperature, it is possible to get information about the type of microstructural gradient in the HAZ. An idea of property changes like hardness and toughness can also be obtained from Fig. 5.

A fairly satisfactory correlation between real welds and simulation studies has been observed with respect to microstructure and property measurements. However, it has frequently been observed that the maximum prior to austenite grain size in real welds is less than that in corresponding simulation specimens (Ref 2, 6, 7). It is believed that this is due to the growth of large grains in the HAZ of real welds being hindered by the adjacent smaller grains. The constraint arises from the fact that grain growth in real welds is in some sense directional, occurring from the fusion line toward lower temperatures, while under the simulation condition the austenite grains can grow in all directions.

The above discussion shows that the PTCT diagram (along with property measurements, as in Fig. 5) can alert the user to the possibility of local brittle zones in actual weldments, and that any error in the property being assessed will be on the conservative side.

While the expressions for Δt_{8-5} given Ref 6 (to determine the point at which a vertical line is to be drawn in the PTCT diagram) are reasonably accurate, those for T_p may be grossly in error because they are derived from a model with many simplistic assumptions (Ref 10, 11). For this reason, although an idea of the type of microstructural gradient in the HAZ can be obtained from the PTCT diagram, it is not possible to calculate the width of each region using simple closed form solutions of the type given in Ref 6 (derived from Rosenthal's equations). Moreover, T_p and Δt_{8-5} are alone not enough to completely characterize a weld thermal cycle. To improve the accuracy of simulation, it would be necessary to program the thermal cycle with heating rates and the shape of the thermal cycle at higher temperatures, as in the HAZ of real welds. The significance of these additional factors is discussed below.

Continuous Heating Transformation Diagrams. The heating rate in many welding processes can be very high. The implication is that considerable superheating may be required for the transformation to austenite on heating. That is, the Ac_1 and Ac_3 temperatures will be raised with an increase in the heating rate (Ref 11). This is seen in Fig. 6, a continuous heating transformation (CHT) diagram, which is analogous to a CCT diagram. It is seen that grain growth begins only after the carbides have dissolved and after a homogeneous austenite (with respect to the distribution of carbon, at least) is formed.

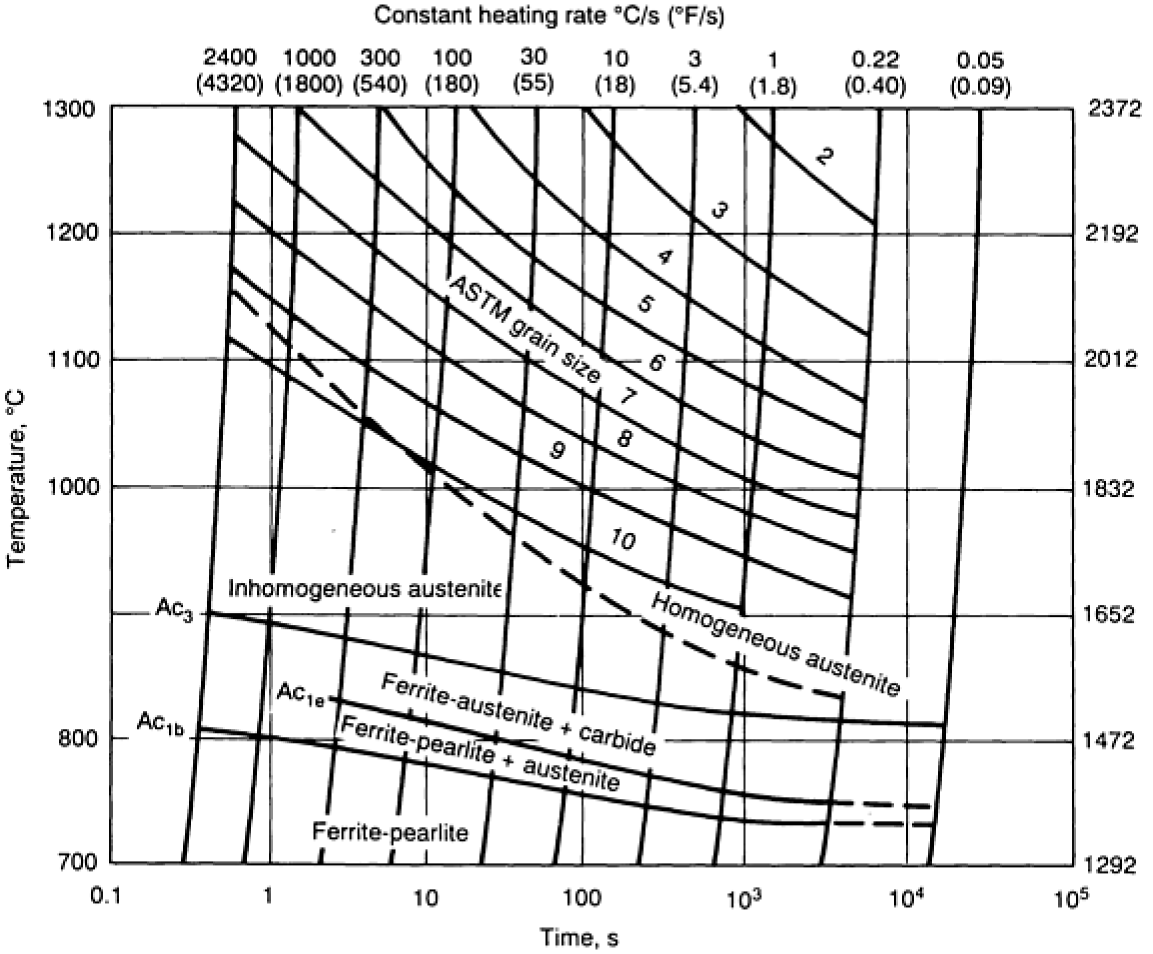

FIG. 6 CONTINUOUS HEATING TRANSFORMATION DIAGRAM FOR 34CRMO4STEEL. IN THE REGION OF HOMOGENEOUS AUSTENITE THERE ARE LINES OF CONSTANT AUSTENITE GRAIN SIZE (ASTM NUMBERS). BECAUSE OF THE MEASURING PROCEDURE, THE DIAGRAMS CAN ONLY BE INTERPRETED ALONG LINES OF CONSTANT HEATING RATE. TO SHOW THE HEATING TIME MORE CLEARLY, A TIME SCALE IS ADDED. SOURCE: REF 12

Precipitate Stability and Grain Boundary Pinning. In the last three decades, there has been a surge in the development of microalloyed or high-strength low-alloy (HSLA) steels (base metals). Almost the entire thrust behind this effort has been to tailor the thermomechanical processing of these steels to produce a fine grain size (Ref 13). Previously, higher strength was obtained by alloying the steel sufficiently to increase the hardenability and then heat treating to obtain a tempered martensitic structure. The higher carbon equivalent (Ref 6, 14) in these alloy steels led to an increased susceptibility to cold cracking or hydrogen-induced cracking. In the microalloyed steels, however, alloying additions are kept to a minimum, and higher strength is achieved primarily by a reduction in the grain size and by precipitation strengthening. A reduction in grain size is the only known method of increasing the strength and toughness at the same time. Because strengthening is obtained by other means, both carbon and carbon equivalent can be decreased to significantly reduce the susceptibility to cold cracking in these new HSLA steels (Fig. 7).

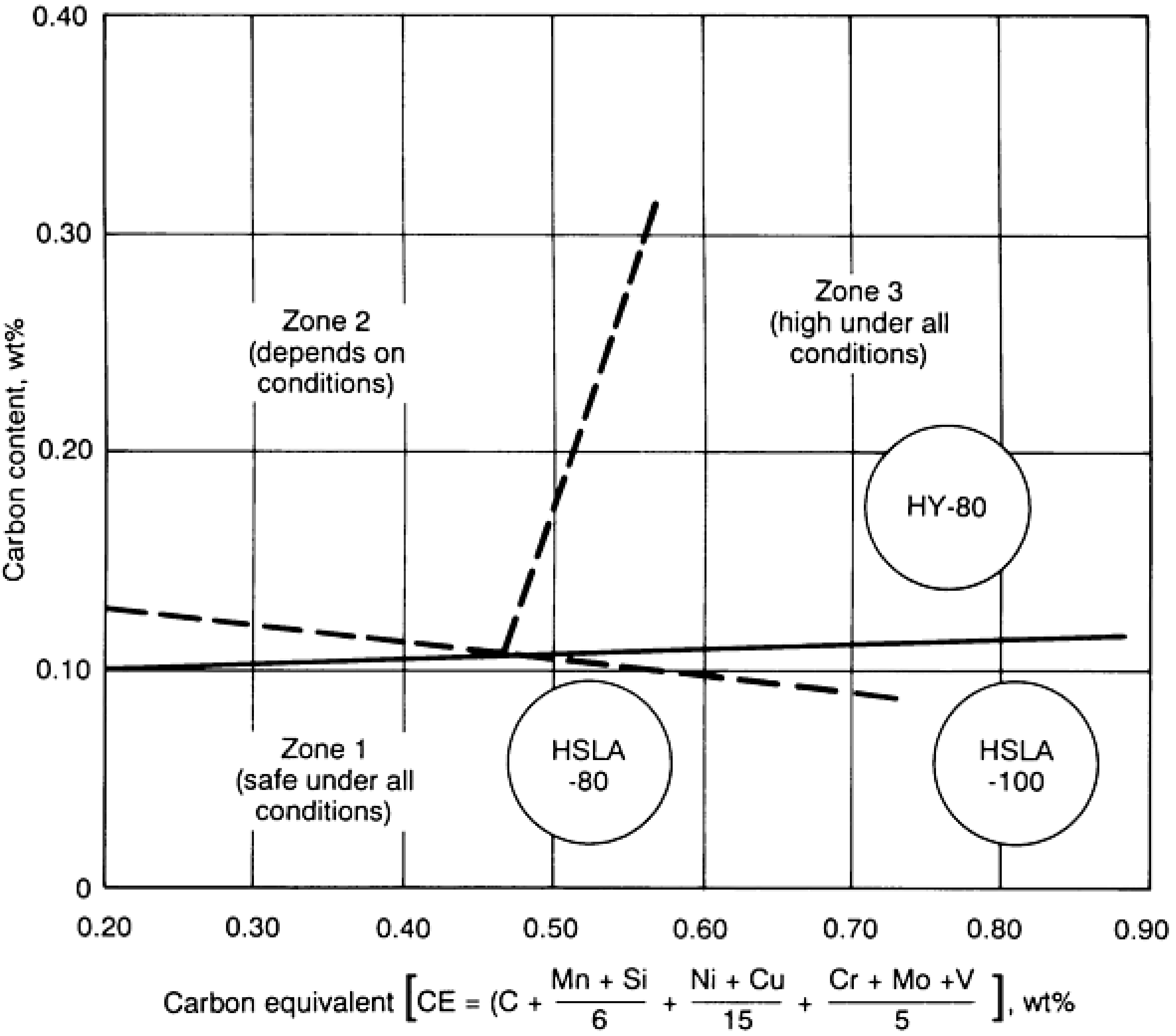

FIG. 7 EFFECT OF STEEL COMPOSITION (HY-80, HSLA-80, HSLA-100) ON THE SUSCEPTIBILITY TO COLD CRACKING IN THE HAZ. SOURCE: REF 13

A fine ferrite grain size in these steels is best achieved (after thermomechanical processing) by obtaining a fine austenite grain size before the γ-α transformation (Ref 13). To realize this goal, it is essential that microalloy additions such as niobium, vanadium, titanium, and aluminum are present so that the fine precipitate particles they form with nitrogen, carbon, or oxygen can "pin" the austenite grain boundaries and inhibit grain coarsening. The pinning effect arises from the fact that when a short length of grain boundary is replaced by a precipitate particle, the effective grain boundary energy is lowered. When the grain boundary attempts to migrate away from the particles, the local energy increases and a drag is exerted on the boundary by the particles.

In the HAZs of welds in these microalloyed steels, it is not possible to get the same optimum microstructure and microalloy precipitation obtained in the parent material by controlled thermomechanical processing. Because peak temperatures are much higher in the HAZ, the precipitate particles coarsen and dissolve, resulting in reduced pinning forces and therefore coarser austenite grains. This effect can be minimized by having precipitate particles that do not dissolve, even at higher temperatures. An idea of the stability of precipitate particles can be obtained from the plot shown in Fig. 8. It is seen that TiN has maximum stability, and this has been used to advantage in many steel. The other carbides and nitrides are not as effective as titanium nitride in limiting the extent of grain coarsening; they play a bigger role during thermomechanical processing (Ref 13).

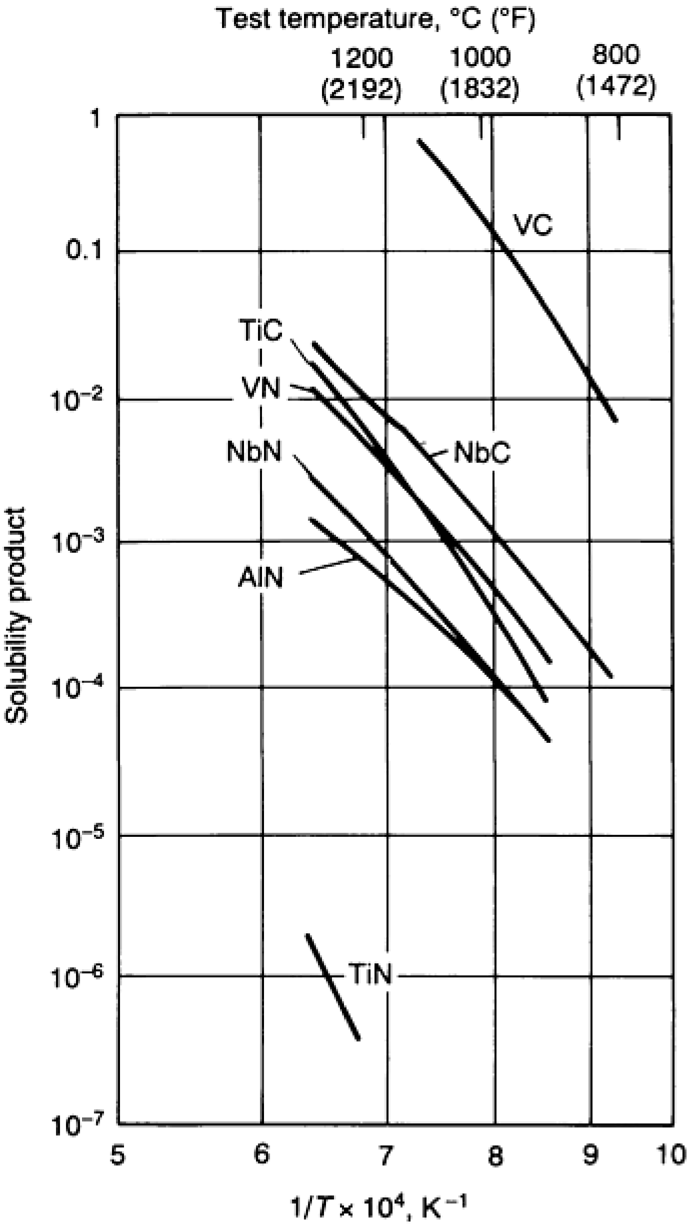

FIG. 8 SOLUBILITY PRODUCTS OF CARBIDES AND NITRIDES AS A FUNCTION OF TEMPERATURE. SOURCE: REF 15

Figure 9 shows how a high volume fraction of fine precipitate particles is needed to maximize the pinning effect. It also shows that heating to high temperatures increases the size of the particles, by a process known as Ostwald ripening (Ref 6), and decreases their volume fraction due to dissolution. This means that it is only possible to limit, and not to totally stop, grain coarsening in the HAZ, especially in high heat input welds. This inevitable grain coarsening is actually used to advantage in titanium, oxide steels (see the section "Titanium Oxide Steels" in this article).

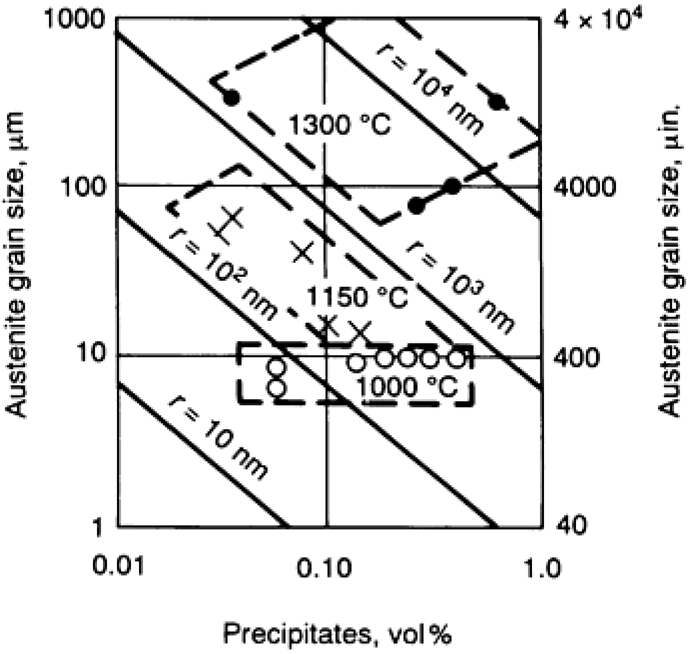

FIG. 9 EFFECT OF PARTICLE RADIUS, *R,* AND AMOUNT OF PRECIPITATE ON AUSTENITE GRAIN SIZE. SOURCE: REF 16

Unmixed and Partially Melted Zones in a Weldment. It is common to think of a single-pass weld as consisting of two zones: weld metal and HAZ. Careful metallographic examination has shown that a weld can in fact be divided into four regions (Fig. 10):

- *THE COMPOSITE ZONE,* IN WHICH A VOLUME OF BASE METAL MELTED BY THE SUPERHEATED FILLER METAL EXPERIENCES COMPLETE MIXING TO PRODUCE AN ALLOY WITH NOMINAL COMPOSITION INTERMEDIATE BETWEEN THAT OF THE BASE METAL AND THAT OF THE FILLER METAL.
- *THE UNMIXED ZONE,* WHICH FORMS FROM THE STAGNANT MOLTEN BOUNDARY LAYER REGION (ABOUT 100 TO 1000 µM, OR 0.004 TO 0.040 IN., THICK) AT THE OUTER EXTREMITIES OF THE COMPOSITE REGION. BECAUSE NO MECHANICAL MIXING WITH THE FILLER METAL OCCURS HERE, THE COMPOSITION OF THE METAL IN THIS REGION IS IDENTICAL TO THAT OF THE BASE METAL, EXCEPT FOR MINOR CHANGES PRODUCED BY DIFFUSION.
- *THE PARTIALLY MELTED ZONE,* WHICH IS A REGION AT THE FUSION BOUNDARY WHERE THE PEAK TEMPERATURES FALL BETWEEN THE LIQUIDUS AND SOLIDUS SO THAT MELTING IS INCOMPLETE.
- *THE TRUE HEAT-AFFECTED ZONE,* WHICH IS THAT PORTION OF THE BASE METAL WHERE ALL MICROSTRUCTURAL CHANGES INDUCED BY WELDING OCCUR IN THE SOLID STATE.

Special metallographic procedures are needed to reveal the existence of these special zones, as shown in Fig. 11.

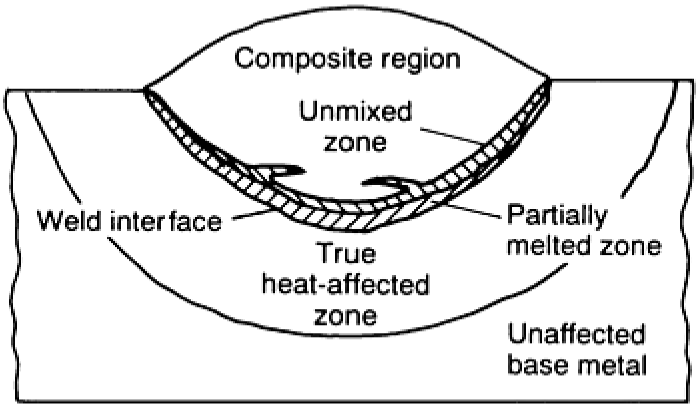
Composite region
Unmixed zone
Weld interface
Partially melted zone
True heat-affected zone
Unaffected base metal

FIG. 10 SCHEMATIC SHOWING THE DIFFERENT DISCRETE REGIONS PRESENT IN A SINGLE-PASS WELD. SOURCE: REF 17

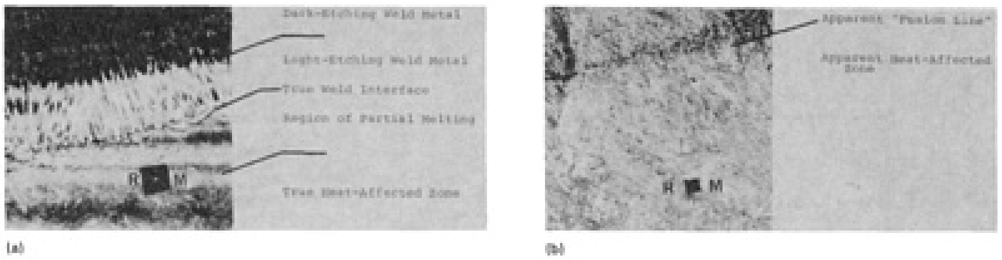
Dark-Etching Weld Metal
Light-Etching Weld Metal
True Weld Interface
Region of Partial Melting
True Heat-Affected Zone
R M
(a)
Apparent "Fusion Line"
Apparent Heat-Affected Zone
R M
(b)

FIG. 11 SPECIAL METALLOGRAPHIC PROCEDURES SHOW THE MICROSTRUCTURE OF DISCRETE REGIONS (SEE FIG. 10) IN AN HY-80 WELD. (A) MICROSTRUCTURE OBTAINED NEAR A REFERENCE MARK (INDICATED BY "RM") USING A TWO-STAGE NITAL AND SODIUM BISULFITE PROCEDURE. (B) MICROSTRUCTURE AT IDENTICAL LOCATION REPOLISHED AND ETCHED WITH NITAL ONLY. TWO-STAGE PROCEDURE SHOWS MUCH GREATER DETAIL. SOURCE: REF 17

The width of the partially melted zone can be extended by a phenomenon known as constitutional liquation (Ref 18), whereby melting can occur even when the peak temperature is less than the solidus temperature. This can be understood by referring to a phase diagram for a simple binary system A and B (Fig. 12). Consider an alloy of composition C. Its equilibrium structures at temperatures T_1, T_2, T_3 and T_4 are $\alpha+\beta$, α, α, and $L + \alpha$, respectively (Fig. 12a). On heating from room temperature to T_3 and holding at this temperature for a long time, the βparticles will dissolve and give a homogeneous α of composition C. However, under conditions of rapid heating in the HAZ, the dissolution of β will give rise to a solute concentration gradient around each particle, as shown in Fig. 12(b). In the region surrounding each particle, the concentration of the solute B will correspond to that of liquid because, as the phase diagram shows, a liquid phase must exist between α and β at T_3. This phenomenon of localized melting due to a nonequilibrium distribution of phases during rapid heating is called constitutional liquation.

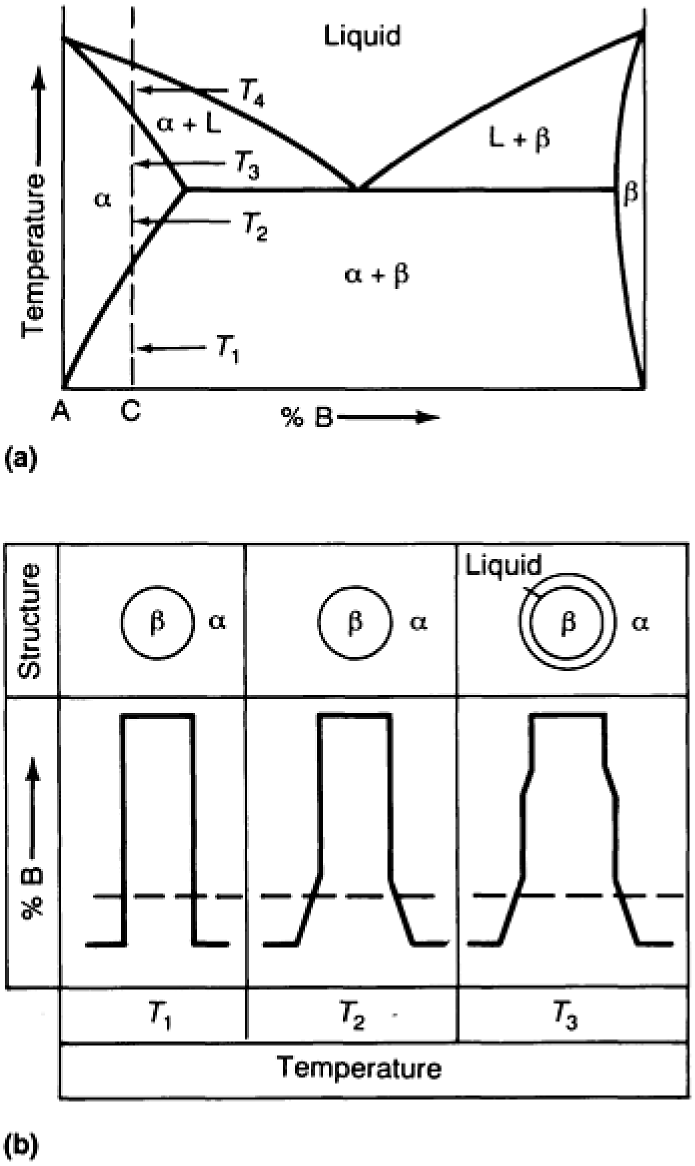

FIG. 12 EXTENSION OF PARTIALLY MELTED ZONE BY CONSTITUTIONAL LIQUATION. (A) PHASE DIAGRAM FOR A SIMPLE BINARY ALLOY. (B) STRUCTURES AND DISTRIBUTION OF CONSTITUENT B IN ALLOY C AT THREE TEMPERATURES SHOWN IN (A) WHEN RAPID HEATING IS APPLIED. SOURCE: REF 19

The partially melted zone is the region where liquation cracks have been known to occur in maraging steels, austenitic stainless steels, heat-treatable aluminum alloys, and nickel-base superalloys. It can also be a site where hydrogen-induced cracking is initiated (Ref 17), because it can act as a pipeline for the diffusion of hydrogen picked up by the molten weld metal and also because of the higher hardenability at these segregated regions.

While the phenomenon of constitutional liquation is usually discussed in connection with liquation cracking of aluminum- and nickel-base alloys (Ref 14, 18, 19), a more dramatic example can actually be found in the case of cast iron welds (Ref 20). Figure 13 shows the micro structure of the HAZ in a cast iron weld deposited using a "quench welding" technique (Ref 21). This procedure involves welding without any preheat by intermittently depositing a series of small stringer beads and strictly maintaining the interpass temperature below about 80 °C (175 °F). The idea is to limit the size of the hard and brittle white iron colonies forming by constitutional liquation around the graphite nodules and to see that they do not interconnect. By contrast, a procedure involving a preheat of about 200 °C (390 °F) can prevent the formation of martensite but not the formation of ledeburite (the structure of white iron). Because ledeburite forms at higher temperatures than martensite (≈ 1100 °C, or 2010 °F), a much higher preheat temperature of about 600 °C (1110 °F) is required to significantly reduce the cooling rate at the higher temperatures, thereby preventing the formation of white iron. This high level of preheat is impractical. While cast iron appears to have been successfully welded with preheats of about 200 °C (390 °F), it has sometimes been found that weld repair is most effective using the quench welding technique, by which it is possible to limit both the amount of ledeburite formed and the width of the HAZ.

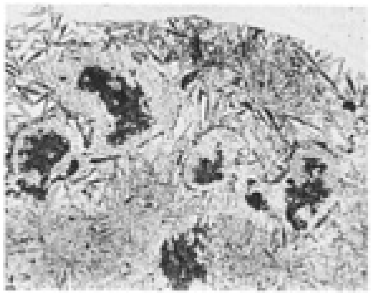

FIG. 13 DISCONTINUOUS WHITE IRON COLONIES OBTAINED IN THE HAZ OF BLACKHEART MALLEABLE IRON WELDED USING A "QUENCH WELDING" TECHNIQUE WITH A NICKEL-BASE ELECTRODE. 180×. SOURCE: REF 21

References cited in this section

2. O.M. AKSELSEN AND T. SIMONSEN, TECHNIQUES FOR EXAMINING TRANSFORMATION BEHAVIOUR IN WELD METAL AND HAZ--A STATE OF ART REVIEW, *WELD. WORLD,* VOL 25 (NO. 1/2), 1987, P 26-34
5. G.T. ELDIS, A CRITICAL REVIEW OF DATA SOURCES FOR ISOTHERMAL TRANSFORMATION AND CONTINUOUS COOLING TRANSFORMATION DIAGRAMS, *HARDENABILITY CONCEPTS WITH APPLICATION TO STEEL,* D.V. DOANE AND J.S. KIRKALDY, ED., TMS-AIME, 1978, P 126-157
6. K. EASTERLING, *INTRODUCTION TO THE PHYSICAL METALLURGY OF WELDING,* BUTTERWORTHS, 1983
7. *WELD THERMAL SIMULATORS FOR RESEARCH AND PROBLEM SOLVING,* THE WELDING INSTITUTE, U.K., 1972
8. C.F. BERKHOUT AND P.H. VAN LENT, THE USE OF PEAK TEMPERATURE-COOLING TIME DIAGRAMS IN THE WELDING OF HIGH STRENGTH STEELS (IN GERMAN), *SCHWEISS. SCHNEID.,* VOL 20 (NO. 6), 1968, P 256-260
9. TH.J. VAN ADRICHEM AND J. KAS, *CALCULATION, MEASUREMENT AND SIMULATION OF WELD THERMAL CYCLES,* SMITWELD N.V., NIJMEGEN, THE NETHERLANDS
10. P. RAVI VISHNU, W.B. LI, AND K. EASTERLING, HEAT FLOW MODEL FOR PULSED WELDING, *MATER. SCI. TECH.,* VOL 7 (NO. 7), 1991, P 649-659
11. P. RAVI VISHNU AND K. EASTERLING, PHENOMENOLOGICAL MODELLING OF HEAT FLOW AND MICROSTRUCTURAL CHANGES IN PULSED GTAW WELDS OF A QT STEEL, *MATHEMATICAL MODELING OF WELD PHENOMENA,* H. CERJAK AND K. EASTERLING, ED., THE INSTITUTE OF METALS, 1993
12. VEREIN DEUTSCHER EISENHÜTTENLEUTE, ED., *STEEL--A HANDBOOK FOR MATERIALS RESEARCH AND ENGINEERING, VOLUME 1: FUNDAMENTALS,* SPRINGER VERLAG, 1992, P 175
13. A.J. DEARDO, C.I. GARCIA, AND E.J. PALMIERE, THERMOMECHANICAL PROCESSING OF STEELS, *HEAT TREATING,* VOL 4, *ASM HANDBOOK,* 1991, P 237-255
14. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987
15. K. NARITA, PHYSICAL CHEMISTRY OF THE GROUPS IVA (TI, ZR), VA (V, NB, TA) AND THE RARE EARTH ELEMENTS IN STEEL, *TRANS. ISIJ,* VOL 15, 1975, P 145-152
16. R.K. AMIN AND F.B. PICKERING, AUSTENITE GRAIN COARSENING AND THE EFFECT OF

THERMOMECHANICAL PROCESSING ON AUSTENITE RECRYSTALLIZATION, *THERMOMECHANICAL PROCESSING OF MICROALLOYED AUSTENITE,* A.J. DEARDO, G.A. RATZ, AND P.J. WRAY, ED., TMS-AIME, 1982, P 14

17. W.F. SAVAGE, E.F. NIPPES, AND E.S SZEKERES, A STUDY OF WELD INTERFACE PHENOMENA IN A LOW ALLOY STEEL, *WELD. J.,* SEPT 1976, P 260S-268S, AND HYDROGEN INDUCED COLD CRACKING IN A LOW ALLOY STEEL, *WELD. J.,* SEPT 1976, P 276S-283S
18. J.J. PEPE AND W.F. SAVAGE, EFFECTS OF CONSTITUTIONAL LIQUATION IN 18-NI MARAGING STEEL WELDMENTS, *WELD J.,* SEPT 1967, P 411S-422S
19. W. YENISCAVITCH, JOINING, *SUPERALLOYS II,* C.T. SIMS, N.S. STOLOFF, AND W.C. HAGEL, ED., JOHN WILEY & SONS, 1987, P 495-516
20. E.E. HUCKE AND H. UDIN, WELDING METALLURGY OF NODULAR CAST IRON, *WELD. J.,* AUG 1953, P 378S-385S
21. E.N. GREGORY AND S.B. JONES, WELDING CAST IRONS, *WELDING OF CASTINGS*, THE WELDING INSTITUTE, U.K., 1977, P 145-156

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

Fusion Zone of a Single-Pass Weld

Transformations in Single-Pass Weld Metal. It is usually not necessary to select a filler metal that has exactly the same composition as the base metal; it is more important that the weld metal has the same strength and other properties (such as toughness or corrosion resistance). Because these properties are governed by the microstructure, it is important to understand the influence of different factors on phase transformations in the weld metal. But first, for a meaningful communication of the different features in a microstructure, the various phases and microconstituents must be identified using a system of nomenclature that is both widely accepted and well understood. In wrought steels, this need has been satisfied to a large degree by the Dubé scheme (Ref 22) for classifying the different morphologies of ferrite, as shown in Fig. 14. Similarly, confusion and controversy in the terminology for describing the microstructures in ferritic steel weld metals have been largely resolved by the classification scheme shown in Fig. 15. This scheme (Ref 23) was the result of several collaborative exercises undertaken under the auspices of the International Institute of Welding (IIW). Typical micrographs illustrating some of the microstructural constituents are shown in Fig. 16.

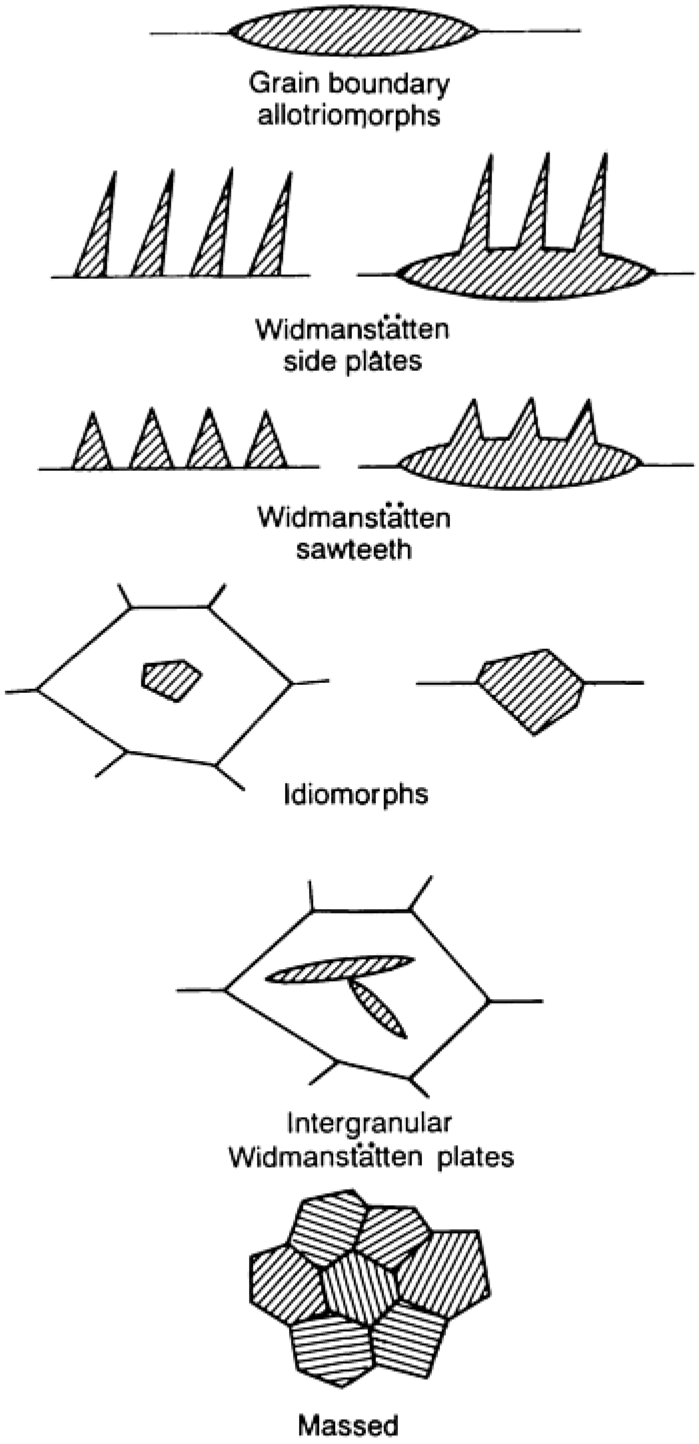
Grain boundary allotriomorphs
Widmanstätten side plates
Widmanstätten sawteeth
Idiomorphs
Intergranular Widmanstätten plates
Massed

FIG. 14 SCHEMATICS SHOWING THE DUBÉ CLASSIFICATION OF FERRITE MORPHOLOGIES. SOURCE: REF 22

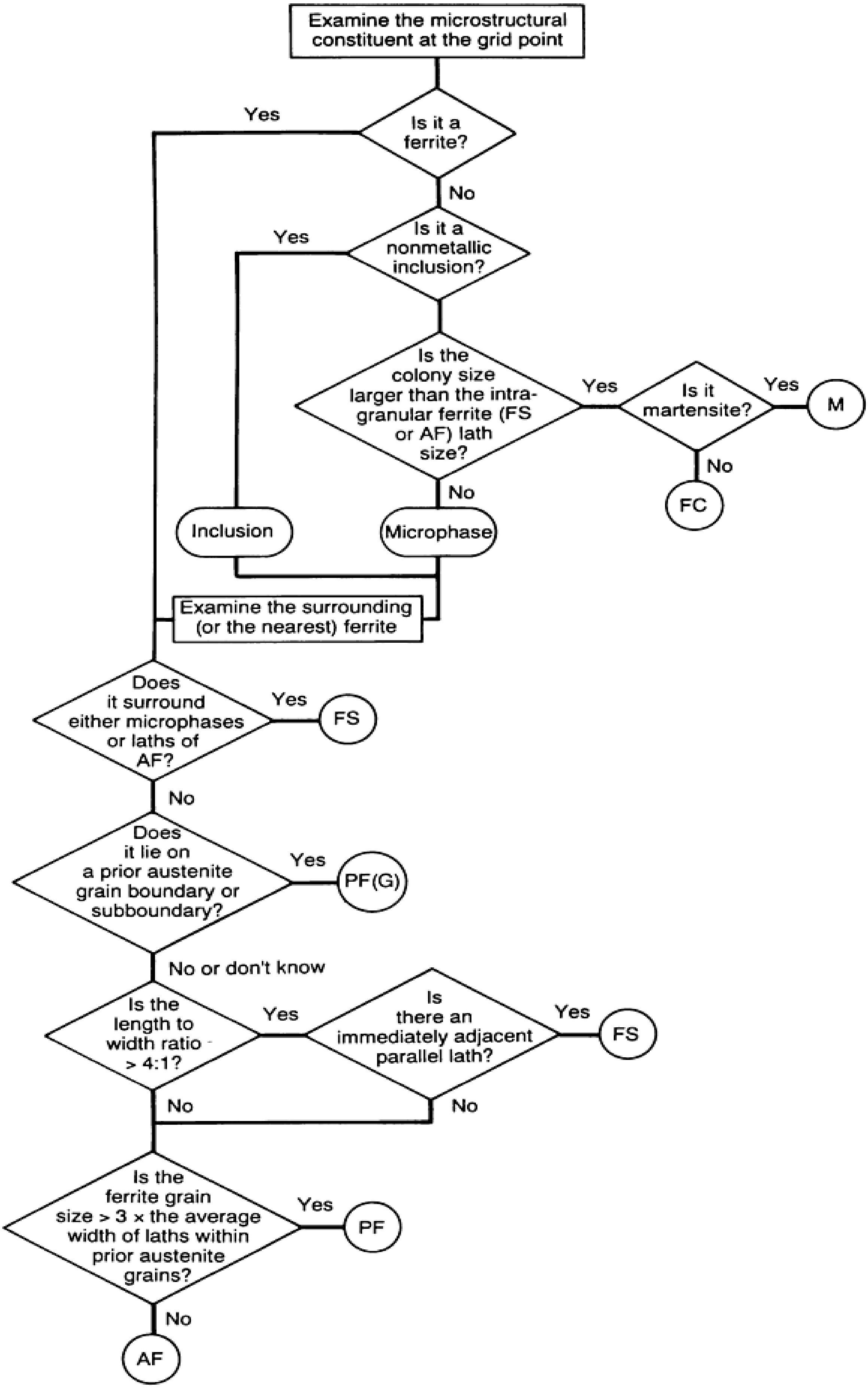
Examine the microstructural constituent at the grid point
Is it a ferrite?
Yes
No
Is it a nonmetallic inclusion?
Yes
Is the colony size larger than the intra-granular ferrite (FS or AF) lath size?
Yes
No
Is it martensite?
Yes
No
M
FC
Inclusion
Microphase
Examine the surrounding (or the nearest) ferrite
Does it surround either microphases or laths of AF?
Yes
FS
No
Does it lie on a prior austenite grain boundary or subboundary?
Yes
PF(G)
No or don't know
Is the length to width ratio > 4:1?
Yes
Is there an immediately adjacent parallel lath?
Yes
FS
No
No
Is the ferrite grain size > 3 × the average width of laths within prior austenite grains?
Yes
PF
No
AF

CATEGORY	ABBREVIATION
PRIMARY FERRITE	PF
GRAIN BOUNDARY FERRITE	PF(G)
INTRAGRANULAR POLYGONAL FERRITE	PF(I)
FERRITE WITH SECOND PHASE	FS
FERRITE WITH NONALIGNED SECOND PHASE	FS(NA)
FERRITE WITH ALIGNED SECOND PHASE	FS(A)
FERRITE SIDE PLATES	FS(SP)
BAINITE	FS(B)
UPPER BAINITE	FS(UB)
LOWER BAINITE	FS(LB)
ACICULAR FERRITE	AF
FERRITE-CARBIDE AGGREGATE	FC
PEARLITE	FC(P)
MARTENSITE	M
LATH MARTENSITE	M(L)
TWIN MARTENSITE	M(T)

FIG. 15 INTERNATIONAL INSTITUTE OF WELDING SCHEME FOR CLASSIFYING MICROSTRUCTURAL CONSTITUENTS IN FERRITIC STEEL WELD METALS WITH THE OPTICAL MICROSCOPE. SOURCE: REF 23

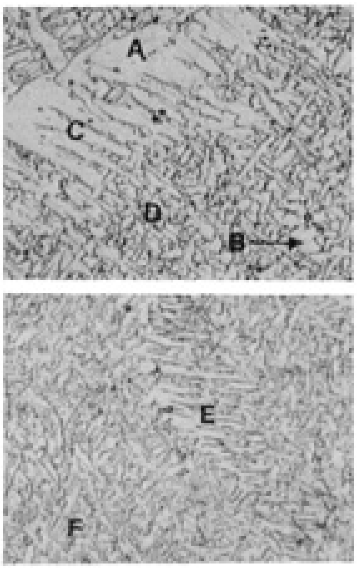

FIG. 16 MICROGRAPHS SHOWING TYPICAL MICROSTRUCTURES IN LOW-CARBON STEEL WELD METAL (NITAL ETCH). A, GRAIN BOUNDARY FERRITE [PF(G)]; B, POLYGONAL FERRITE [PF(I)]; C, WIDMANSTÄTTEN FERRITE [FS(SP)]; D, ACICULAR FERRITE (AF); E, UPPER BAINITE [FS(UB)]; F, LOWER BAINITE [FS(LB)] AND/OR MARTENSITE (M). SOURCE: REF 3

The transformation behavior in ferritic steel weld metals is best understood by first noting that ferrite is nucleated heterogeneously and that for all practical purposes, it is necessary to consider the competitive nucleation behavior only at grain boundaries and at inclusions. Figure 17 shows that inclusions have to be larger than a certain size (0.2 to 0.5 μm, or 8 to 20 μin., the typical size range of most weld metal inclusions) for their potency as nucleation sites to reach a maximum. It also shows that nucleation of ferrite is always energetically more favorable at grain boundaries than at inclusions. Based on additional considerations, like thermal contraction strains and lattice matching at inclusion-austenite-ferrite interfaces (not taken into account to get the results in Fig. 17), a discussion of why certain inclusions are more potent nucleation sites than others is found in Ref 3, 24, and 25.

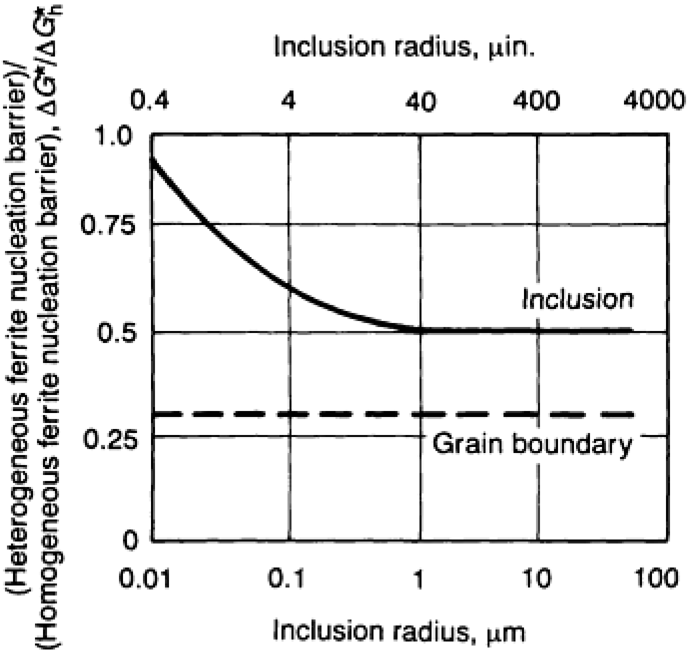

FIG. 17 EFFECT OF PARTICLE RADIUS ON ENERGY BARRIER TO FERRITE NUCLEATION AT INCLUSIONS, ΔG^* (HETEROGENEOUS), NORMALIZED RELATIVE TO THE HOMOGENEOUS NUCLEATION BARRIER, ΔG^*_H (HOMOGENEOUS). CORRESPONDING ENERGY BARRIER TO NUCLEATION OF FERRITE AT AUSTENITE GRAIN BOUNDARIES IS INDICATED BY HORIZONTAL BROKEN LINE. SOURCE: REF 6

Broadly, the major factors affecting transformation behavior in ferritic steel weld metals are alloy composition, weld heat input (by its effect on γ grain size and Δt_{8-5}), oxygen content (that is, the inclusion content), and the nature of segregation in the weld metal. A typical weld CCT diagram is shown in Fig. 18. For the cooling curve shown in the figure, the first phase that forms is allotriomorphic ferrite, or grain boundary ferrite in the IIW scheme. The term "allotriomorphic" means a "particle of a phase that does not have a regular external shape" and in the present context means that ferrite grows on the grain boundary surfaces and does not have a regular faceted shape reflecting the symmetry of its internal crystalline structure. At lower temperatures, the mobility of the curved or random γ/α-allotriomorph boundaries decreases, and Widmanstätten side plates (ferrite side plates in the IIW scheme) form. Growth of these side plates is rapid because carbon is efficiently redistributed to the sides of the growing tips, thus avoiding solute pile-up problems. In addition, substitutional atoms do not diffuse during the growth of Widmanstätten ferrite.

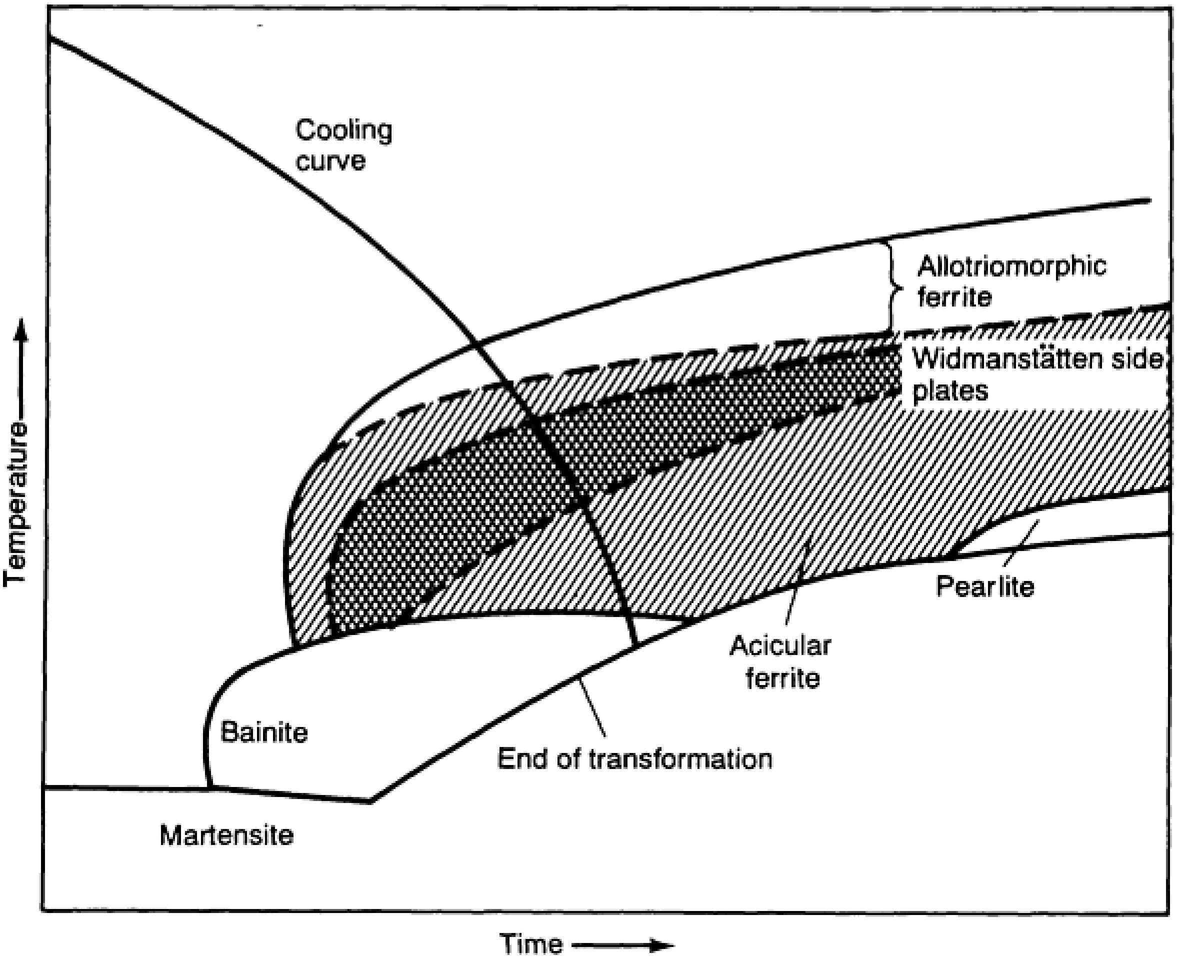

FIG. 18 SCHEMATIC OF WELD CCT DIAGRAM SHOWING SELECTED MICROSTRUCTURES

After all the grain boundary sites are saturated with allotriomorphic or Widmanstätten ferrite and their growth rate is not sufficient to extend to the interior of the grains, the nucleation of ferrite at inclusions within the γ grains becomes competitive. Acicular ferrite forms, a structure resulting from ferrite laths growing in different directions from inclusions and from laths already nucleated. Upon impingement, high-angle grain boundaries and a very fine dispersion of microphases are obtained between ferrite laths. "Microphases" in this context means the transformation structures resulting from the carbon-enriched regions between the ferrite laths and could be the martensite-austenite constituent, bainite, or pearlite (see the section "HAZ in Multipass Weldments" in this article). An example of the microstructure of acicular ferrite is shown in Fig. 19. Acicular ferrite does not figure in the Dubé scheme because it is rarely observed in wrought steels. When the cooling rate is higher or when the inclusion content is very low, bainite can be nucleated directly at the γ grain boundaries. Bainite can form as upper or lower bainite; the difference between the two is illustrated in Fig. 20. Examination with transmission electron microscopy (TEM) is usually required for a firm identification of the type of bainite formed. In the case of the highest cooling rates, martensite is obtained.

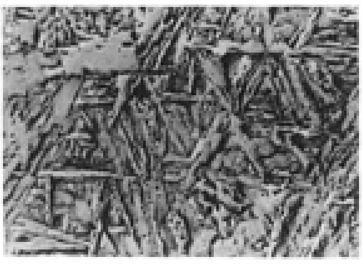

FIG. 19 SCANNING ELECTRON MICROGRAPH SHOWING MORPHOLOGY OF ACICULAR FERRITE. SOURCE: REF 24

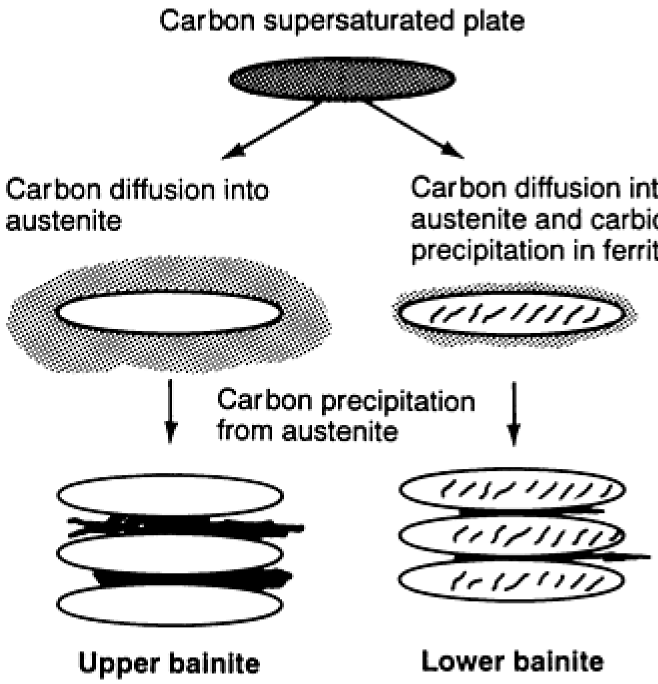
Carbon supersaturated plate
Carbon diffusion into austenite
Carbon diffusion into austenite and carbide precipitation in ferrite
Carbon precipitation from austenite
Upper bainite
Lower bainite

FIG. 20 SCHEMATIC SHOWING THE DIFFERENCES IN THE TRANSFORMATION MECHANISM FOR UPPER AND LOWER BAINITE AND THE EFFECT OF THESE MECHANISMS ON THE FINAL MORPHOLOGY. SOURCE: REF 24

Figure 21 shows the effect of alloying additions, $\Delta t_{8\text{-}5}$, oxygen content, and the γ grain size on the sequence and relative amounts of allotriomorphic ferrite, Widmanstätten ferrite, and bainite in the weld metal. Broadly, the trends shown can be understood in terms of the wellknown observation that an increase in γ grain size, a decrease in inclusion content, and an increase in alloying additions will shift the CCT curves to lower temperatures and to longer times. While Fig. 21 can generally be interpreted by assuming that the individual factors act independently of each other, the transition from (b) to (a) with an increase in the inclusion content must be understood in terms of its effect of the γ grain size. It has been observed that the γ grain size decreases with an increase in the inclusion content, and this has been explained by the higher magnitude of pinning forces at higher inclusion contents (Ref 25). The smaller grain size implies a greater grain boundary surface area, and this increases the amount of grain boundary ferrite and sideplate structures formed at the expense of acicular ferrite in Fig. 21(a). The role of the initial formation or non-formation of allotriomorphic ferrite at γ grain boundaries in bringing about the transition from Fig. 21(b) to Fig. 21(c) in Fig. 21 is discussed in the next section of this article.

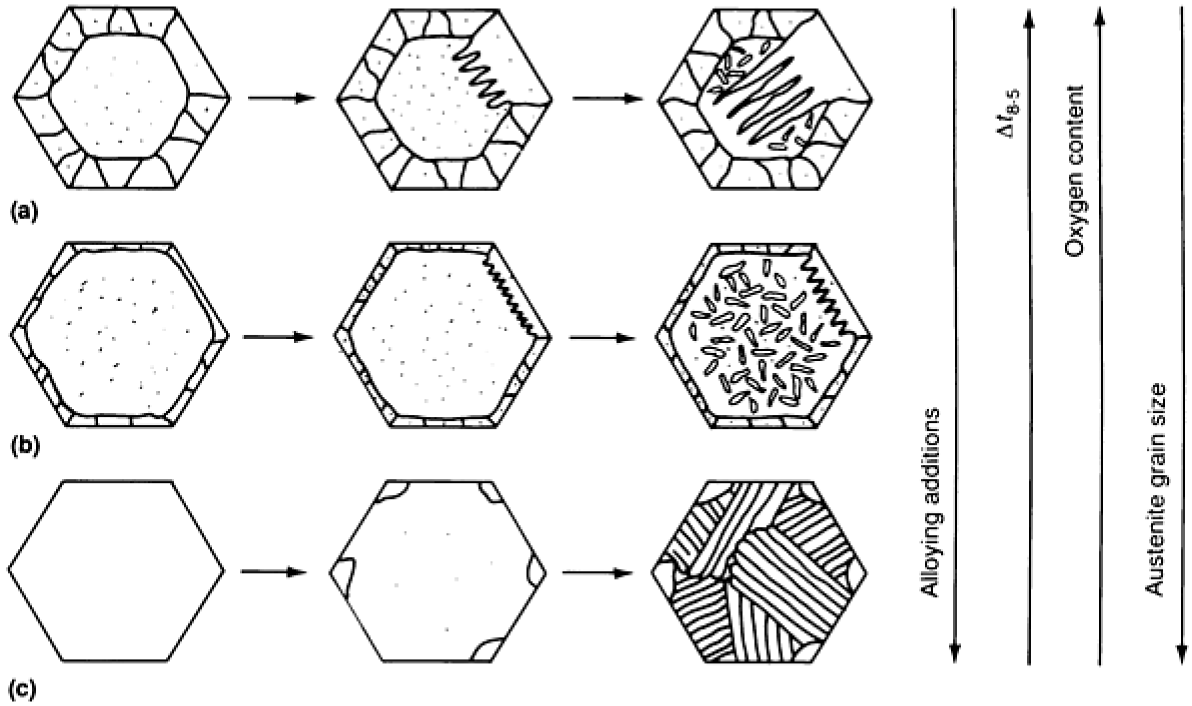

FIG. 21 SCHEMATIC SHOWING THE EFFECT OF ALLOY COMPOSITION, $\Delta T_{8\text{-}5}$, OXYGEN CONTENT, AND γ GRAIN SIZE ON THE DEVELOPMENT OF MICROSTRUCTURE IN FERRITIC STEEL WELD METALS. THE HEXAGONS REPRESENT CROSS SECTIONS OF COLUMNAR γ GRAINS. (A) THE GRAIN BOUNDARIES BECOME DECORATED FIRST WITH A UNIFORM, POLYCRYSTALLINE LAYER OF ALLOTRIOMORPHIC FERRITE, FOLLOWED BY FORMATION OF WIDMANSTÄTTEN FERRITE, AND THEN BY FORMATION OF ACICULAR FERRITE. (B) THE GROWTH RATE OF WIDMANSTÄTTEN FERRITE IS NOT SUFFICIENTLY HIGH TO EXTEND ENTIRELY ACROSS γ GRAINS. NUCLEATION OF FERRITE AT INCLUSIONS WITHIN THE γ GRAINS LEADS TO AN INCREASE IN THE AMOUNT OF ACICULAR FERRITE WHEN COMPARED WITH CASE (A). (C) THE HIGHER ALLOY CONTENT OR THE HIGHER COOLING RATE SUPPRESSES THE FORMATION OF ALLOTRIOMORPHIC FERRITE. THIS LEAVES THE γ GRAIN BOUNDARIES FREE TO NUCLEATE UPPER BAINITE. SOURCE: REF 24

Although the IIW classification is the result of a consensus viewpoint, some researchers (Ref 24) still feel that it places an exaggerated emphasis on morphological features observed with the light microscope. They believe that it is more important that the terminology reflect the mechanism of transformation. For example, it can be difficult to differentiate between Widmanstätten ferrite and upper bainite when using the light microscope, and both are identified as FS in the IIW classification. For a more fundamental understanding and to clearly interpret the trends in microstructural development, it is essential to make a distinction between the two. In addition, the subclassification of grain boundary ferrite in the IIW classification is not unequivocal in its meaning--it could refer to ferrite that forms by diffusion, Widmanstätten ferrite, or the ferrite in upper bainite. By contrast, the term "allotriomorphic ferrite" clearly identifies the mechanism by which it is formed. On the other hand, the IIW classification is easy to use and allows a more detailed identification, either by judgment or by further TEM examination, for example, to determine whether FS is FS(SP) or FS(UB). At least, it decreases the tendency for a wrong identification of the transformation mechanism with insufficient data. For a microstructure-toughness correlation, it is sufficient to characterize the morphological features.

Relating Weld Metal Toughness to the Microstructure. A good insight into the subject of toughness in ferrite steel weld metal is obtained by examining the data shown in Fig. 22. It is seen from Fig. 22(a) that the upper-shelf Charpy V-notch (CVN) energy monotonically decreases with an increase in oxygen content. In the upper-shelf temperature region, ductile fracture occurs by a process of microvoid coalescence, and because microvoids are initiated at the inclusion-matrix interface, the upper-shelf impact energy decreases with an increase in the inclusion content. By contrast, from Fig. 22(b) it is seen that an optimum inclusion content is required to obtain the lowest CVN transition temperature. The transition temperature is mainly governed by cleavage fracture, and this in turn depends on how effectively a propagating cleavage crack is forced to change direction as it traverses the microstructure. At low inclusion contents, an upper bainitic structure is obtained, and this consists of parallel platelets of ferrite (in a single packet) growing from the grain boundary surfaces. With an optimum inclusion content, a predominantly acicular ferrite structure is obtained, and here the adjacent ferrite platelets tend to radiate in many different directions from inclusion nucleation site. At higher inclusion contents, the amount of ferrite sideplate structures increases, again having nearly parallel ferrite platelets. The highest toughness (that is, the lowest transition temperature) is obtained only in the "chaotic" microstructure of acicular ferrite because it has the smallest effective ferrite grain size.

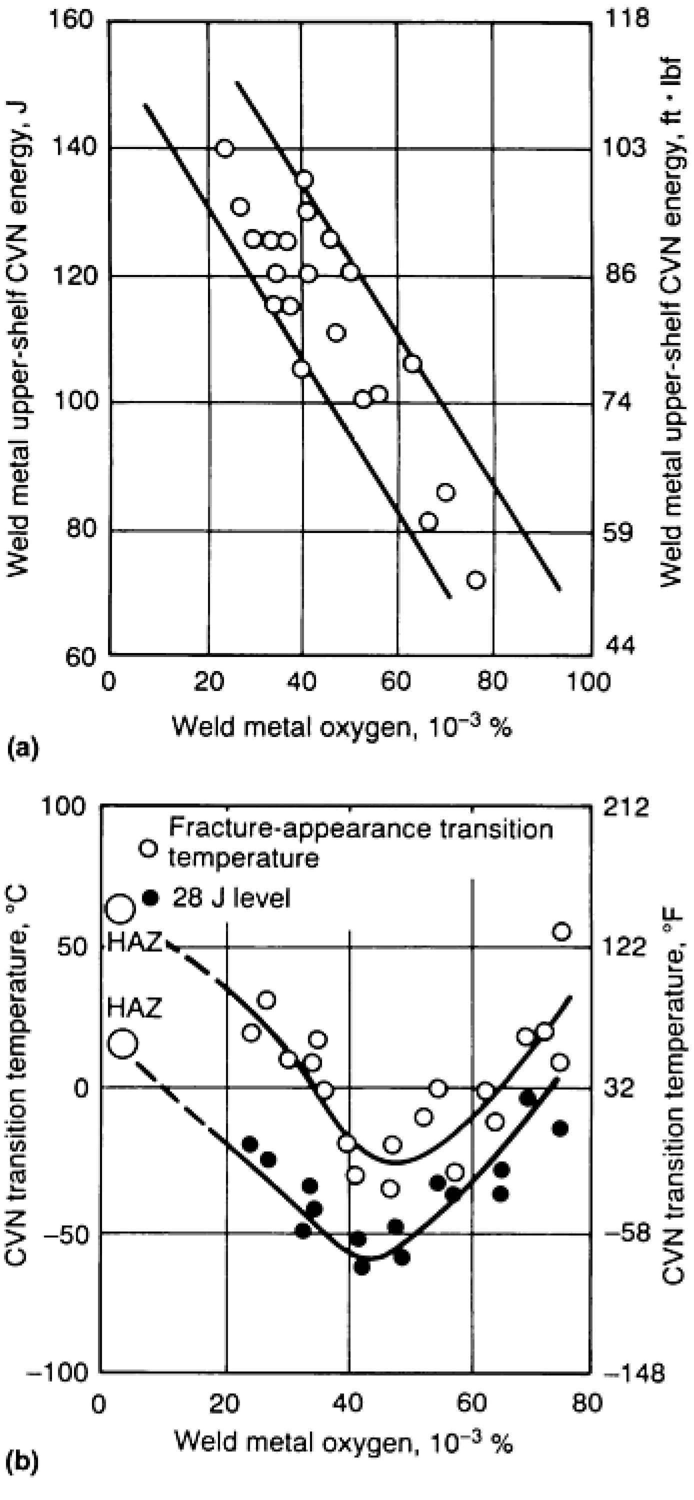

FIG. 22 TYPICAL VARIATION OF SELECTED WELD METAL PROPERTIES WITH OXYGEN CONTENT IN FERRITIC STEEL WELD METALS. (A) PLOT OF CHARPY V-NOTCH (CVN) IMPACT TEST UPPER-SHELF ENERGY VERSUS OXYGEN CONTENT. (B) PLOT OF CVN TRANSITION TEMPERATURE VERSUS OXYGEN CONTENT. SOURCE: REF 26

In recognition of the latter fact, most of the work in consumables development in the recent past has concentrated on increasing the amount of acicular ferrite in the microstructure. In line with the trends shown in Fig. 21(a) and 21(b), it has been found that decreasing the amount of grain boundary ferrite and ferrite sideplate structures increases the acicular ferrite content. This is shown more clearly in Fig. 23.

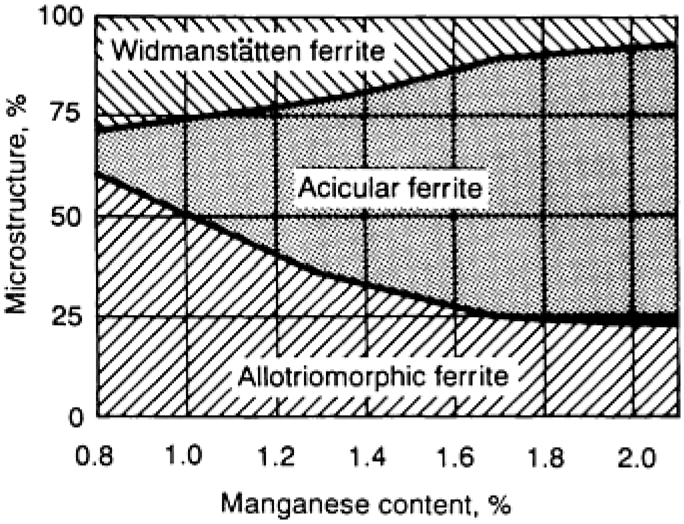

FIG. 23 EFFECT OF MANGANESE CONTENT OF WELD METAL ON THE RELATIVE AMOUNTS OF THE MICROSTRUCTURAL CONSTITUENTS PRESENT. CARBON CONTENT MAINTAINED AT 0.03%. SOURCE: REF 27

Recent work (Ref 24, 25, 27), however, has shown that a 100% acicular ferritic structure is neither desirable nor, in most cases, achievable. Figure 24 shows that the impact toughness peaks at about 70% acicular ferrite and then falls. This was attributed to the formation of segregated bands of brittle microphases in the alloy compositions giving the highest acicular ferrite contents. Moreover, when the alloying additions are increased still further, the transformation behavior becomes such that allotriomorphic ferrite formation is totally suppressed, leaving the γ grain boundaries free to nucleate upper bainite at lower temperatures (see Fig. 21c). It appears that a small layer of allotriomorphic ferrite is essential to obtain high acicular ferrite contents. When it forms, it saturates the γ grain boundary sites, and bainite nucleation at the γ/α-allotriomorphic interface is inhibited by the high carbon content there as a result of its rejection from the ferrite (Ref 24). Conditions are then favorable for intragranular nucleation at inclusion sites and thus the formation of acicular ferrite.

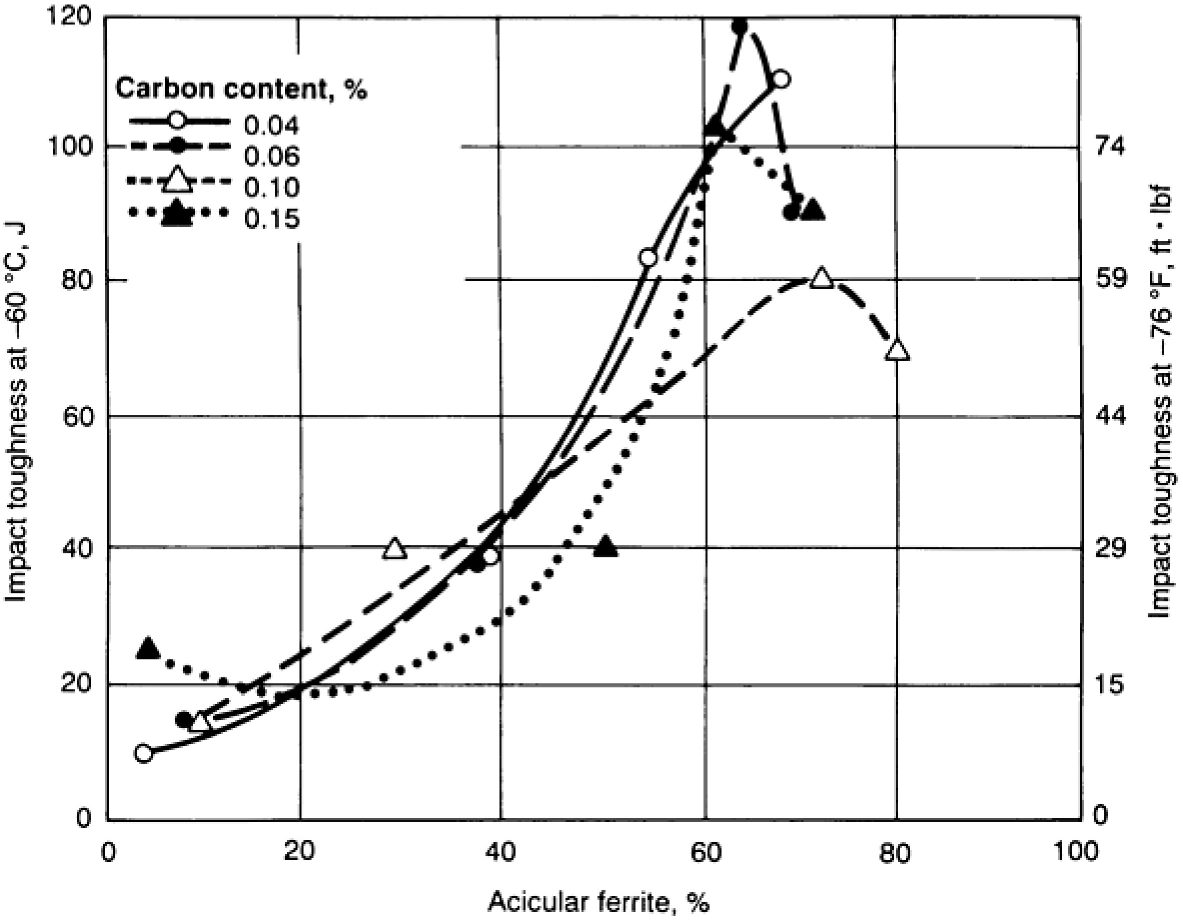

FIG. 24 PLOT OF IMPACT ENERGY VERSUS ACICULAR FERRITE CONTENT FOR SELECTED CARBON CONTENTS AT -60 °C (-78 °F). SOURCE: REF 27

For the welding of steels with yield strengths less than 600 MPa (85 ksi), it has been possible to develop weld metals with matching strength and toughness by having high acicular ferrite contents in them. A yield strength of about 600 MPa (85 ksi) seems to be the ceiling for weld metals based on acicular ferrite. Attempting to obtain additional strength by additional alloying to increase the amount of solid solution hardening and precipitation hardening in fact only results in increasing amounts of martensite and bainite. These structures are associated with a lower toughness. This has been observed in the case of HSLA-100 steels, for example, where attempts to get a matching strength in the weld metal have resulted in a drop in toughness (Ref 25, 28). Work on consumables development for high-strength steels is still in progress to match the base metal in strength and toughness (Ref 25, 28). However, current thinking appears to favor resorting to a slightly undermatching weld metal (in terms of strength) so that the already available lower-strength, high-toughness, consumables having high acicular ferrite contents (Ref 29) can be used. These weld metal compositions have the additional benefit of requiring less stringent measures for avoiding hydrogen-induced cracking.

Titanium Oxide Steels. In the discussion of the role of pinning by precipitate particles (for example, TiN) in limiting grain growth in the HAZ (in the section "Precipitate Stability and Grain Boundary Pinning" in this article), it was pointed out that at high heat inputs the particles dissolve and are not able to prevent coarse γ grains from forming. The low toughness in these coarse-grained regions is of some concern. The fact that a coarse γ grain size will lead to a higher acicular ferrite content (see Fig. 21b) with improved toughness has been used to advantage in the development of titanium-oxide-containing steels (Ref 30, 31). Ti_2O_3 is more stable than titanium nitride and does not dissolve, even at the highest heat inputs. The undissolved, Ti_2O_3 particles do not stop grain growth, but their survival after the heating cycle means that they can be effective in nucleating acicular ferrite within the coarse γ grains. The precise reasons why they are more effective than other undissolved inclusions, such as Al_2O_3 in nucleating acicular ferrite are not well understood (Ref 24). The improved HAZ microstructure and toughness obtained with titanium oxide steels over titanium nitride steels at high heat inputs are shown in Fig. 25 and 26, respectively.

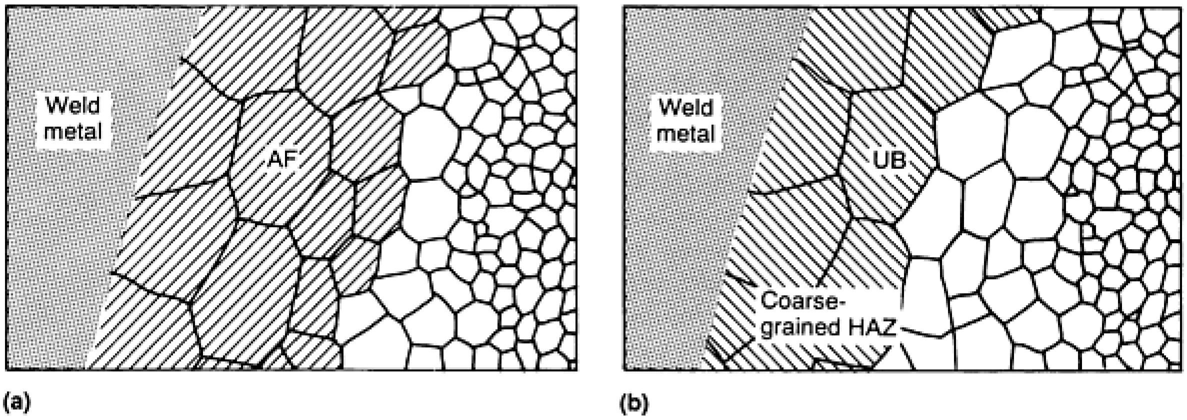

FIG. 25 SCHEMATIC SHOWING HAZ MICROSTRUCTURE IN SELECTED HIGH HEAT INPUT WELDS. (A) TITANIUM OXIDE STEEL. (B) TITANIUM NITRIDE STEEL. AF, ACICULAR FERRITE; UB, UPPER BAINITE. SOURCE: REF 30

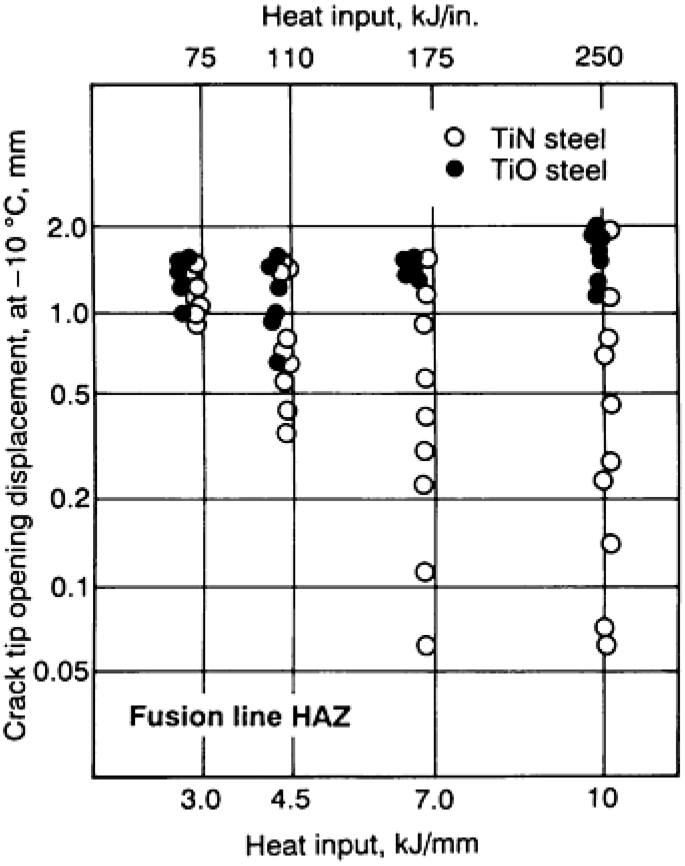

FIG. 26 HAZ TOUGHNESS OF TITANIUM NITRIDE AND TITANIUM OXIDE STEELS WITH 420 MPA (60 KSI) YIELD STRENGTH. SOURCE: REF 31

Effect of Transformations on Transient Weld Stresses. The effect of transformations on weld stresses (and vice versa) is best illustrated by introducing an apparently little-known proposal (Ref 32) about the use of a duplex austenite-martensite weld metal for welding high-strength steels. First, to prepare the background, the way in which transient stresses develop in a weld and the effect of transformations on weld stresses will be discussed.

Four uniaxial specimens using an HAZ simulation technique, the manner in which residual stress is accumulated during a weld thermal cycle was investigated (Ref 33), and the results are shown in Fig. 27. Initially, when a material is austenite at high temperature, its yield strength is low and only a small tensile stress can develop. On cooling, the tensile stress increases. When a phase transformation occurs, the resulting volume expansion opposes the contraction due to cooling and decreases the stress. Once the transformation is complete, the tensile stress increases again. As seen in Fig. 27, if the transformation occurs at a lower temperature (for example, in the case of 9Cr1Mo or 12CrMo steels), the magnitude of the final residual stress is less. The dotted line in the plot shows the expected variation for a material with an austenite-martensite structure. The greater the amount of martensite, the lower the final residual stress.

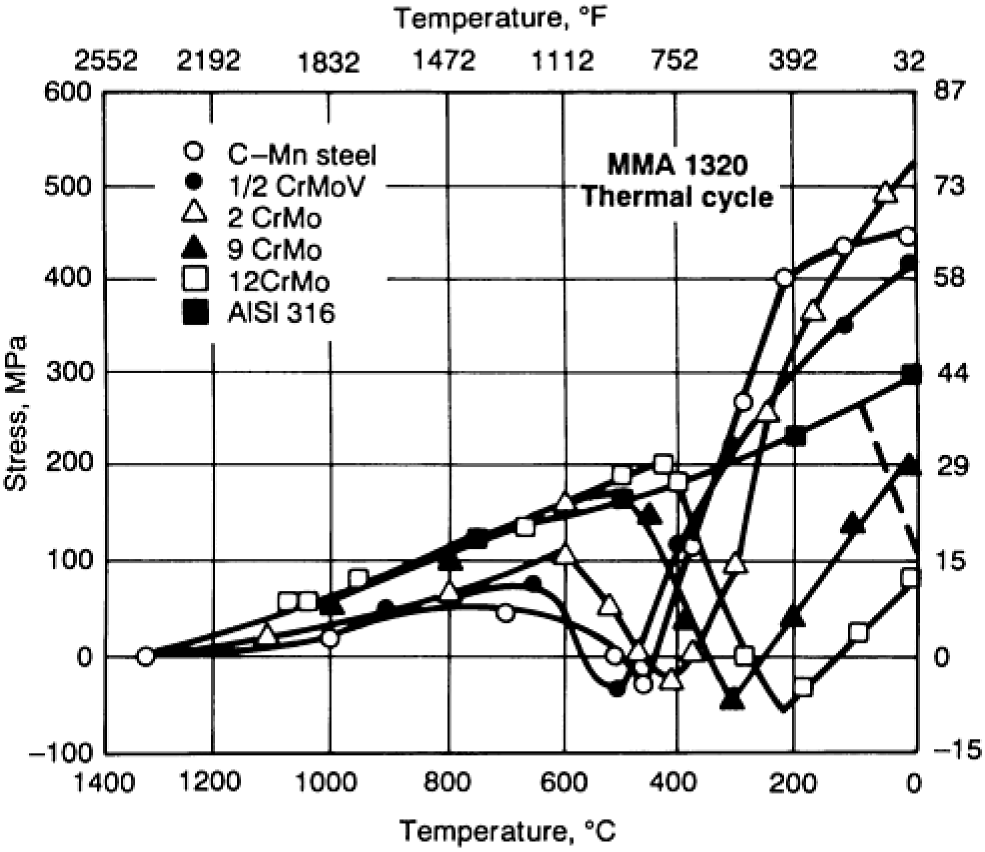

FIG. 27 STRESS ACCUMULATED DURING COOLING UNDER RESTRAINT FOR SELECTED STEELS. THE DOTTED LINE SHOWS THE EXPECTED VARIATION IN STRESS FOR A DUPLEX AUSTENITE-MARTENSITE WELD METAL WHERE TRANSFORMATION BELOW ABOUT 100 °C (212 °F) GREATLY REDUCES THE FINAL STRESS. (MMA 1320 INDICATES THE PEAK TEMPERATURE OF THE SIMULATED WELD THERMAL CYCLE AND CORRESPONDS TO THE TEMPERATURE AT THE START OF THE COOLING CYCLE.) SOURCE: REF 33

As discussed in the section "Weld Metal Toughness Relative to the Microstructure" in this article, the main problems in the welding of high-strength steels (with yield strengths of 600 to 1000 MPa, or 85 to 145 ksi) are hydrogen-induced cracking (HIC) and low toughness in the weld metal. Both problems can be addressed (in principle at least) by the use of a duplex austenite-martensite weld metal. Austenite can contribute by increasing the toughness and by minimizing the risk for HIC (Ref 14). The role of martensite will be to increase the strength and to decrease the final residual tensile stress, which also makes conditions less favorable for HIC. An additional benefit will be that the higher tensile residual stress in the weld metal before martensite forms will induce the transformations in the HAZ to occur at a slightly higher temperature, by Le Chatelier's principle (Ref 34), and thereby cause a microstructure less susceptible to HIC to form in the HAZ. The results of this effect are shown in Fig. 28, where it is seen that a tensile stress in the 500 to 300 °C (930 to 570 °F) temperature range on a ferrite-pearlite weld increased the stresses for failure (Ref 35).

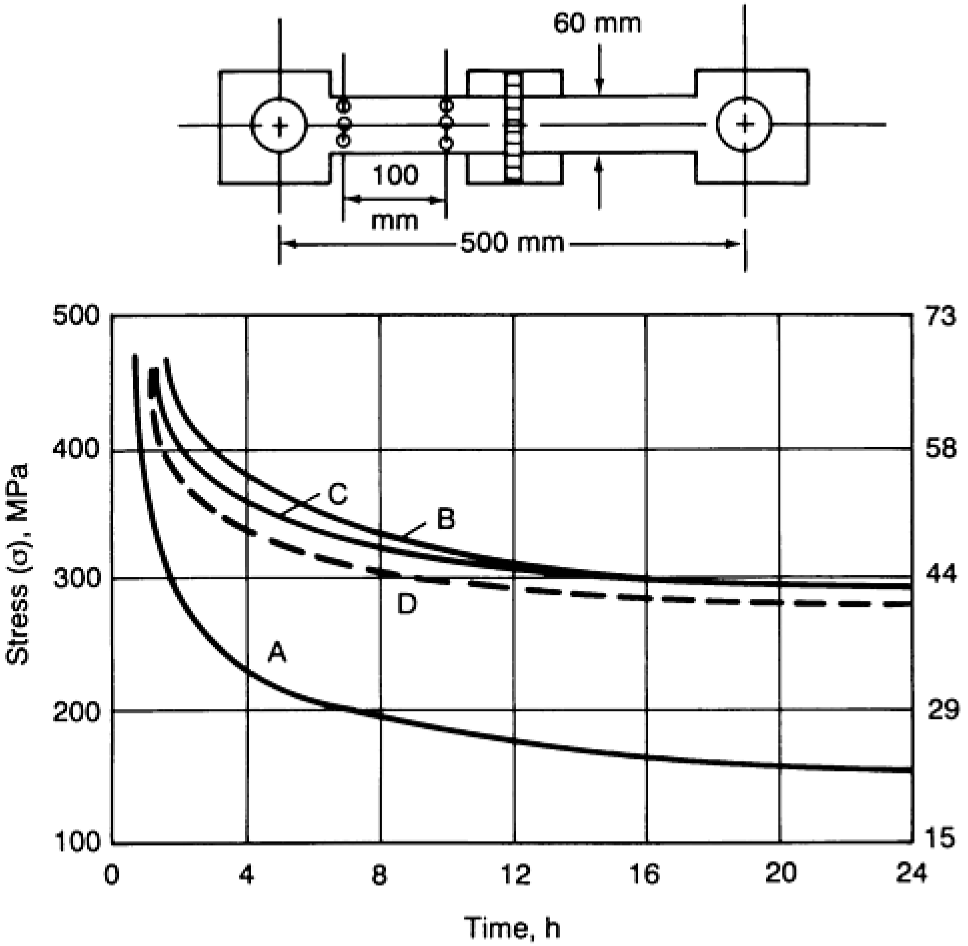

FIG. 28 DELAYED FRACTURE CURVES FOR TENSILE TESTS ON WELDED JOINTS IN 35KH3N3M (A RUSSIAN SPECIFICATION) STEEL. A, FERRITIC-PEARLITIC WELD; B, BAINITIC-MARTENSITIC WELD; C, AUSTENITIC WELD; D, FERRITIC-PEARLITIC WELD PLUS DEFORMATION AT 500 TO 300 °C (930 TO 570 °F). THE SCHEMATIC SHOWS THE SPECIMEN DIMENSIONS FOR THE DELAYED FRACTURE TENSILE TESTS ON WELDED JOINTS. SOURCE: REF 35

It must be mentioned that the idea of using an austenite-martensite weld metal represents a considerable departure from conventional thinking. For example, for certain overlaying applications of stainless steel deposits on low-alloy steel, an overalloyed electrode like E310 or E312 is preferred to E309. Even though the weld metal composition with any of the three electrodes would be acceptable, the use of overalloyed electrodes will minimize the width of the martensitic layer in the unmixed zone at the weld-HAZ interface (Ref 36). By contrast, an austenite-martensite weld metal implies that the choice of consumable will have to be such that the width of the martensitic layer will increase. However, because the proposal is meant for the recently developed high-strength steels that have a low carbon content, only a low-carbon martensite will form, which will not be that harmful. The benefits already discussed could override conventional concerns about having martensite in the weld metal in this case.

To sum up, even though the idea of using an austenite-martensite weld metal for high-strength steels does not seem to have been taken up for serious investigation so far, it has been included here because it is a nice way of illustrating the principles discussed. In addition, it shows that a fundamental knowledge of different effects on phase transformations can lead to unconventional solutions to practical problems.

References cited in this section

3. O. GRONG AND D.K. MATLOCK, MICROSTRUCTURAL DEVELOPMENT IN MILD AND LOW ALLOY STEEL WELD METALS, *INT. MET. REV.*, VOL 31 (NO. 1), 1986, P 27-48

6. K. EASTERLING, *INTRODUCTION TO THE PHYSICAL METALLURGY OF WELDING*, BUTTERWORTHS, 1983
14. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987
22. L.E. SAMUELS, *OPTICAL MICROSCOPY OF CARBON STEELS*, AMERICAN SOCIETY FOR METALS, 1980
23. "GUIDE TO THE LIGHT MICROSCOPE EXAMINATION OF FERRITIC STEEL WELD METALS," DOC. NO. IX-1533-88, INTERNATIONAL INSTITUTE OF WELDING, 1988
24. H.K.D.H. BHADESHIA AND L.E. SVENSSON, MODELLING THE EVOLUTION OF MICROSTRUCTURE OF STEEL WELD METAL, *MATHEMATICAL MODELLING OF WELD PHENOMENA,* H. CERJAK AND K. EASTERLING, ED., THE INSTITUTE OF MATERIALS, 1993
25. G.R. EDWARDS AND S. LIU, RECENT DEVELOPMENTS IN HSLA STEEL WELDING, *ADVANCES IN WELDING METALLURGY,* FIRST U.S.-JAPAN SYMPOSIUM, AMERICAN WELDING SOCIETY AND OTHERS, JUNE 1990, P 215-293
26. B. AHLBLOM, "OXYGEN AND ITS ROLE IN DETERMINING WELD METAL MICROSTRUCTURE AND TOUGHNESS--A STATE OF THE ART REVIEW," DOC. NO. IX-1322-84, INTERNATIONAL INSTITUTE OF WELDING, 1984
27. L.-E. SVENSSON AND B. GRETOFT, MICROSTRUCTURE AND IMPACT TOUGHNESS OF C-MN WELD METALS. *WELD. J.,* DEC 1990, P 454S-461S
28. P.W. HOLSBERG AND R.J. WONG, WELDING OF HSLA-100 STEEL FOR NAVAL APPLICATIONS, *WELDABILITY OF MATERIALS,* R.A. PATTERSON AND K.W. MAHIN, ED., ASM INTERNATIONAL, 1990
29. J.M.B. LOSZ AND K.D. CHALLENGER, MICROSTRUCTURE AND PROPERTIES OF A COPPER PRECIPITATION STRENGTHENED HSLA STEEL WELDMENT, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID AND J.M. VITEK, ED., ASM INTERNATIONAL, 1990
30. H. HOMMA, S. OHKITA, S. MATSUDA, AND K. YAMAMOTO, IMPROVEMENTS OF HAZ TOUGHNESS IN HSLA STEEL BY INTRODUCING FINELY DISPERSED TI-OXIDE, *WELD. J.,* OCT 1987, P 301S-309S
31. N. YURIOKA, "MODERN HIGH STRENGTH STEEL IN JAPAN," FIFTH INTERNATIONAL SYMPOSIUM, JAPAN WELDING SOCIETY, APRIL 1990 (TOKYO)
32. YU.N. GOTALSKII, THE PROBLEM OF WELDING HIGH STRENGTH STEELS, *AUTOM. WELD.*, JUNE 1984, P 37-40
33. W.K.C. JONES AND P.J. ALBERRY, THE ROLE OF PHASE TRANSFORMATIONS IN THE DEVELOPMENT OF RESIDUAL STRESSES DURING THE WELDING OF SOME FAST REACTOR STEELS, *PROC. CONF. FERRITIC STEELS FOR FAST REACTOR STEAM GENERATORS,* BRITISH NUCLEAR ENERGY SOCIETY, 1977
34. *MCGRAW-HILL ENCYCLOPEDIA OF SCIENCE AND TECHNOLOGY,* VOL 9, P 685
35. A.M. MAKARA AND N.A. MOSENDZ, EFFECTS OF THE WELD METAL ON CRACKING IN HAZ, *WELD. RES. ABROAD,* NOV 1965, P 78-86
36. H. IKAWA, S. SHIN, M. INUI, Y. TAKEDA, AND A. NAKANO, "ON THE MARTENSITE-LIKE STRUCTURE AT WELD BOND AND THE MACROSCOPIC SEGREGATION IN WELD METAL IN THE WELDED DISSIMILAR METALS OF α-STEELS AND γ-STEELS," DOC. NO. IX-785-72, INTERNATIONAL INSTITUTE OF WELDING, 1972

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

HAZ in Multipass Weldments

In the HAZ of a single-pass weld, the grain-coarsened zone (GC HAZ) is normally the region having the lowest toughness. Turning to a multipass weld, Fig. 29 (compare with Fig. 3) shows how the GC HAZ can be modified by subsequent passes and can be categorized into four regions, depending on the reheating temperature (Ref 37):

- SUBCRITICALLY REHEATED GRAIN-COARSENED (SCGC) ZONE, THE ZONE REHEATED BELOW AC$_1$
- INTERCRITICALLY REHEATED GRAIN-COARSENED (ICGC) ZONE, THE ZONE REHEATED BETWEEN AC$_1$ AND AC$_3$
- SUPERCRITICALLY REHEATED GRAIN-REFINED (SCGR) ZONE, THE ZONE REHEATED ABOVE AC$_3$ AND BELOW ABOUT 1200 °C (2190 °F)
- UNALTERED GRAIN-COARSENED (UAGC) ZONE, THE ZONE THAT IS NOT REHEATED ABOVE ABOUT 200 °C (390 °F) OR THE ZONE THAT IS AGAIN REHEATED ABOVE ABOUT 1200 °C (2190 °F)

Figure 30 shows how the crack tip opening displacement (CTOD) value of simulated specimens varies with the peak temperature of the second thermal cycle, T_{p2} (T_{p1} = 1400 °C, or 2550 °F). It is seen that the ICGC, UAGC, and SCGC regions have CTOD values less than about 0.1 mm (0.004 in.). Similar low values have been obtained by locating the crack tip in the CTOD tests at corresponding locations in the HAZ of actual multipass weldments. These low toughness regions are commonly known as local brittle zones (LBZs).

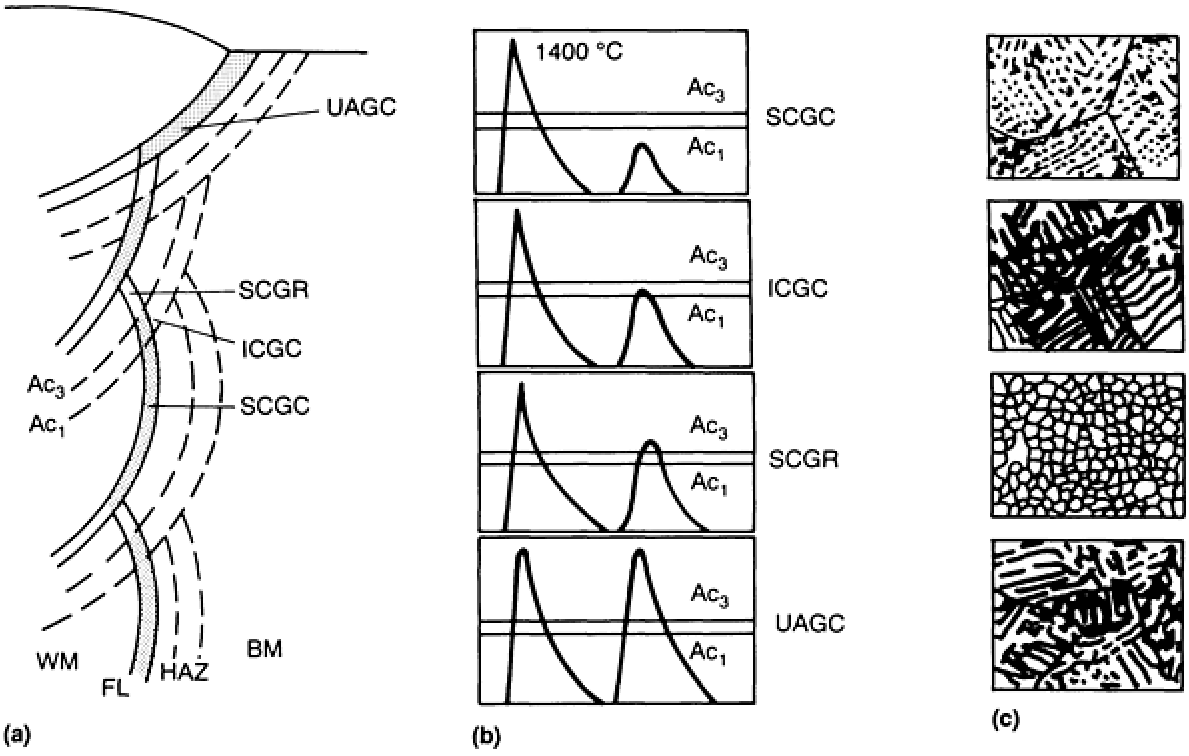

FIG. 29 SCHEMATIC SHOWING THE DIFFERENT SUBZONES THAT CAN FORM IN THE COARSE-GRAINED REGION OF THE HAZ IN A MULTIPASS WELD. (A) POSITION OF SUBZONES RELATIVE TO BASE METAL (BM) AND WELD METAL (WM). (B) PLOT OF THERMAL CYCLES RELATIVE TO AC$_3$ AND AC$_1$. (C) MICROSTRUCTURES AT THE DIFFERENT ZONES. FL REFERS TO THE FUSION LINE. SEE TEXT FOR DETAILS. SOURCE: REF 37

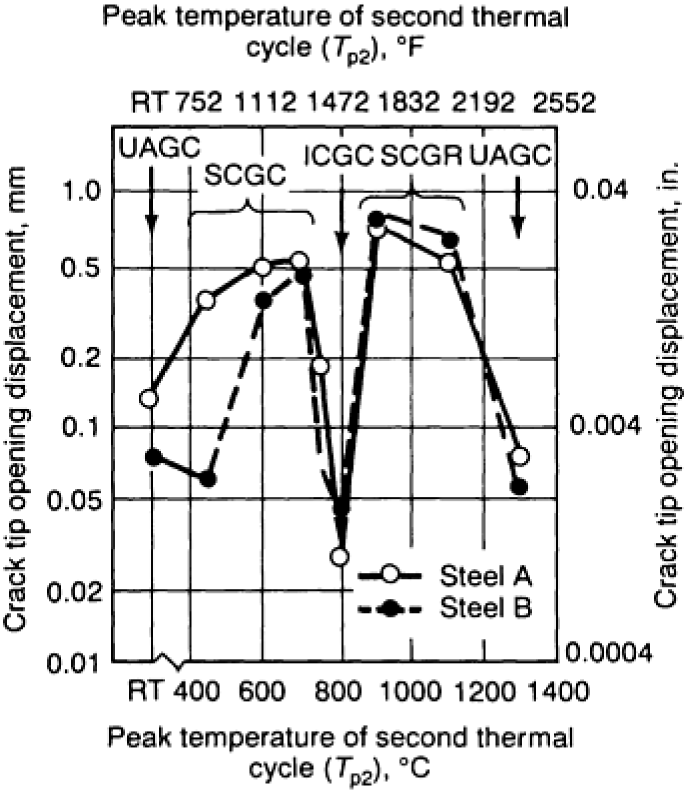

FIG. 30 PLOT OF CRACK TIP OPENING DISPLACEMENT VERSUS PEAK TEMPERATURE OF THE SECOND THERMAL CYCLE FOR SIMULATED SPECIMENS SUBJECTED TO A DOUBLE THERMAL CYCLE. (T_{P1} = 1400 °C = 2550 °F; $\Delta T_{8\text{-}5}$ = 20 S). THE HAZ ZONES (FIG. 29) CORRESPONDING TO THE SIMULATION TRIALS ARE ALSO SHOWN IN THE FIGURE. SOURCE: REF 37

The low CTOD values have been obtained only in tests on modern structural steels, whereas no actual structural failure attributable to the LBZs has been reported so far (Ref 38). Because the structural significance of the low toughness test results is still largely unresolved, the present conservative strategies are to choose a steel whose tendency to form LBZs is less or to choose a welding procedure by which the size of the LBZs can be reduced.

The ICGC HAZ usually has a lower toughness than the SCGC or UAGC regions, even though all of them have nearly the same γ grain size. This is due to the higher amount of highcarbon martensite-austenite (M-A) constituent in the ICGC HAZ. When the GC HAZ is reheated to a temperature between Ac_1 and Ac_3, austenite is nucleated at the high carbon areas. Upon cooling, these local high-carbon areas can transform to give twinned martensite with every thin regions of retained austenite in between. (See Ref 39 for detailed metallographic procedures for identifying the M-A constituent.) The carbon content in the M-A islands can range from about 0.3 to 0.5% C. The significance of this is that, for a given nominal carbon content in the steel, the volume fraction of the M-A constituent will be much higher than if most of the carbon formed carbides (by the lever rule). This will increase the number of crack nucleation sites and thereby contribute to the inferior toughness of ICGC HAZ. As is to be expected, the volume fraction of the M-A constituent also depends on the hardenability of the steel, which in turn depends on the alloying content. It has been shown that if the development of pearlitic microphases could be promoted instead of the development of M-A, by decreasing the alloy content, the toughness of the ICGC HAZ could be improved (Ref 40). However, this would have penalties in terms of achievement of parent plate strength. A more feasible solution would be to inhibit grain growth in the HAZ.

The multipass welding procedure can alternatively be controlled to limit the size of the LBZs. Figure 31 shows how this can be done with a tandem three-wire high current gas-metal arc welding procedure by adjusting the distance between the three arcs. Special "temper-bead" procedures (Ref 42) have been developed for controlling the microstructure in the HAZ, and a need for these procedures arises in the following way. Low-alloy steel weldments for critical applications (for example, pressure vessels) require a postweld heat treatment (PWHT) in a furnace. This is done to temper the hard regions in the HAZ and to relieve residual stresses. If repairs become necessary on site after the component has been in service, PWHT is usually not feasible. The heat of the arc can then be used to achieve the tempering function of PWHT by suitable spatial positioning and sequencing of the individual passes. Grain refinement in he HAZ is also sought to increase the toughness and thereby offset the harmful effects of residual stresses that would remain in the absence of PWHT.

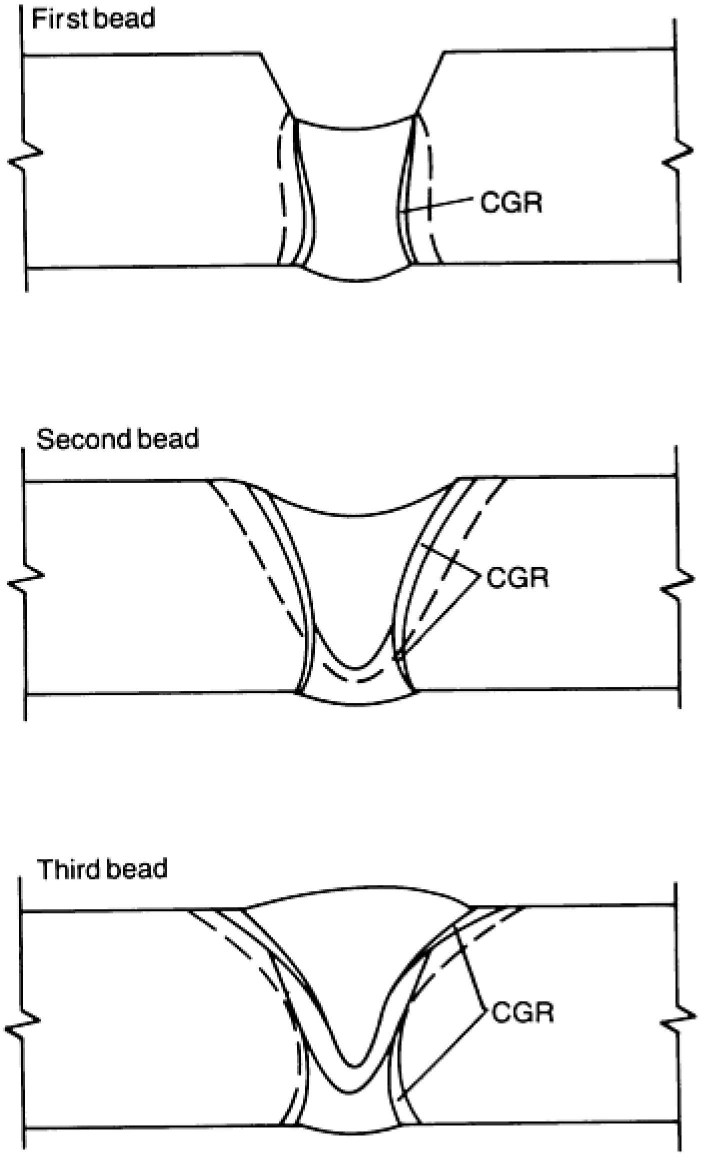

FIG. 31 SCHEMATIC SHOWING THAT THE HAZ ISOTHERMS AND THE SIZE AND LOCATION OF THE COARSE-GRAINED REGION (CGR) CAN BE CONTROLLED IN A TANDEM THREE-WIRE HIGH CURRENT GAS-METAL ARC WELDING PROCEDURE. GRAIN REFINEMENT OF COARSE-GRAINED REGIONS INITIALLY FORMED IS OBTAINED BY OPTIMIZING THE DISTANCES BETWEEN THE THREE ARCS. SOURCE: REF 41

An example of a two-layer temper bead procedure is shown in Fig. 32. The heat inputs of the first and second layer are carefully controlled, so that the heat from the second layer is used to refine the coarse-grained region in the HAZ of the base metal due to the first layer. This ideas can be extended one step further by using a pulsed gas-tungsten arc welding procedure whereby a given pulse in a pulsed weldment can be used to successively refine and temper the preceding pulse pitch regions (Ref 11). The degree of microstructural refinement depends primarily on the welding speed and pulsing frequency, and these parameters can be controlled with great precision. By contrast, an important variable to be controlled in the temper-bead procedure is the weld deposit height, and it is difficult to exercise the same degree of control on this. For this reason, it is argued that the maximum possible control on the microstructural changes in the repair weldment is possible with the pulsing procedure.

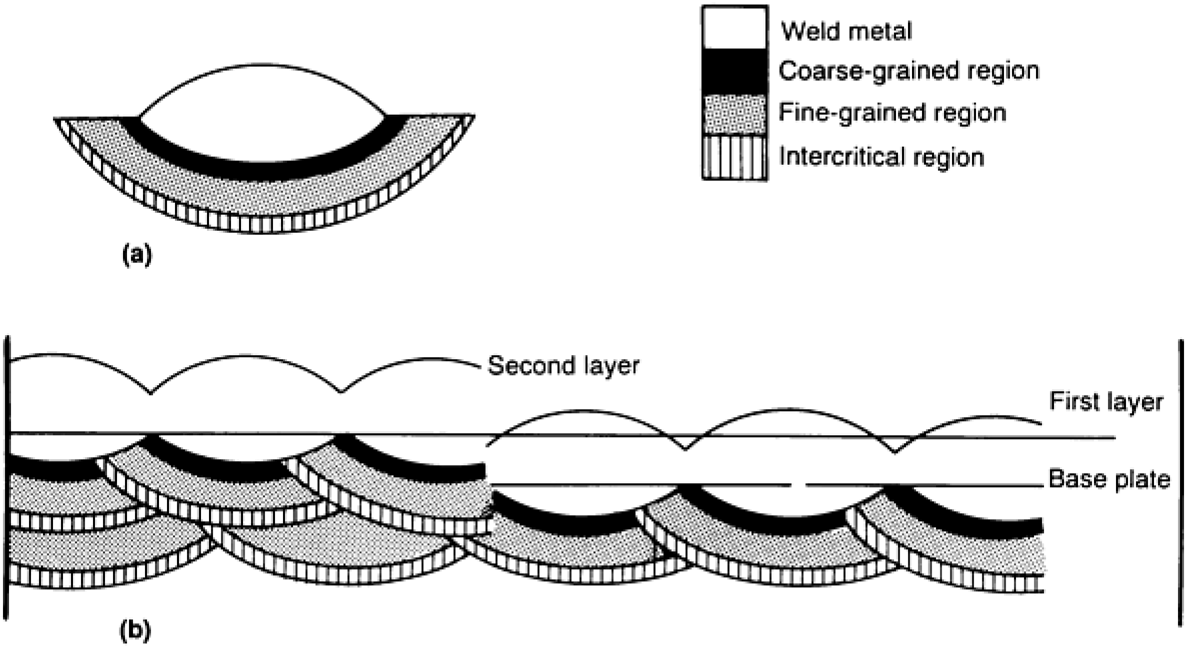

FIG. 32 SCHEMATIC SHOWING A TWO-LAYER REPAIR PROCEDURE. (A) HEAT-AFFECTED AREAS OF A SINGLE WELD BEAD. (B) FIRST LAYER CAUSES COARSE-GRAINED REGIONS TO FORM IN THE HAZ OF THE BASE METAL (RIGHT PORTION OF THE FIGURE). DEPOSITION OF THE SECOND LAYER REFINES THESE INITIAL COARSE-GRAINED REGIONS (LEFT PORTION OF FIGURE). SOURCE: REF 42

References cited in this section

11. P. RAVI VISHNU AND K. EASTERLING, PHENOMENOLOGICAL MODELLING OF HEAT FLOW AND MICROSTRUCTURAL CHANGES IN PULSED GTAW WELDS OF A QT STEEL, *MATHEMATICAL MODELING OF WELD PHENOMENA,* H. CERJAK AND K. EASTERLING, ED., THE INSTITUTE OF METALS, 1993
37. T. HAZE AND S. AIHARA, "INFLUENCE OF TOUGHNESS AND SIZE OF LOCAL BRITTLE ZONES ON HAZ TOUGHNESS OF HSLA STEELS," SEVENTH INTERNATIONAL CONFERENCE ON OFFSHORE MECHANICS AND ARCTIC ENGINEERING, HOUSTON, 1988
38. C.P. ROYER, A USER'S PERSPECTIVE ON HEAT AFFECTED ZONE TOUGHNESS, *WELDING METALLURGY OF STRUCTURAL STEELS,* J.Y. KOO, ED., TMS-AIME, 1987
39. F. MATSUDA, Z. LI, P. BERNASOVSKY, K. ISHIHARA, AND H. OKADA, "AN INVESTIGATION OF THE BEHAVIOUR OF M-A CONSTITUENT IN SIMULATED HAZ OF HSLA STEELS," DOC. NO. IX-B-1591-90, INTERNATIONAL INSTITUTE OF WELDING, 1990
40. P.L. HARRISON AND P.H.M. HART, RELATIONSHIPS BETWEEN HAZ MICROSTRUCTURE AND CTOD TRANSITION BEHAVIOUR IN MULTIPASS C-MN STEEL WELDS, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID AND J.M. VITEK, ED., ASM INTERNATIONAL, 1990
41. H. ONOE, J. TANAKA, AND I. WATANABE, JAPANESE LNG TANKER CONSTRUCTED USING A NEW WELDING PROCESS AND IMPROVED AL-KILLED STEELS, *MET. CONSTR.,* JAN 1979, P 26-31
42. P.J. ALBERRY, SENSITIVITY ANALYSIS OF HALF-BEAD AND ALTERNATIVE GTAW TECHNIQUES, *WELD. J.,* NOV 1989, P 442S-451S

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

Fusion Zone in Multipass Weldments

In the weld metal of a multipass weld, reheating effects will lead to a gradient in microstructure similar to the case of the HAZ. However, instead of a detailed classification, the multipass weld metal is usually considered to consist of just two regions (Fig. 33):

- *AS-DEPOSITED OR PRIMARY REGION,* WHERE THE MICROSTRUCTURE DEVELOPS AS THE WELD COOLS FROM THE LIQUID PHASE TO AMBIENT TEMPERATURE
- *REHEATED OR SECONDARY REGION,* WHERE REGIONS WITH THE ORIGINAL PRIMARY MICROSTRUCTURE ARE REHEATED TO TEMPERATURES ABOVE THE AC_1 TEMPERATURE. THE TEMPERED REGIONS WHICH ARE REHEATED TO SLIGHTLY LOWER TEMPERATURES ARE ALSO GENERALLY CONSIDERED TO BELONG TO THIS CATEGORY

The properties of the weld metal depend on the relative area or volume fractions of the two regions, which in turn depend on the welding procedure, so the properties are procedure-specific. Hence arises the need for welding procedure qualification as per codes and standards in addition to the qualification of consumables (for which is standard procedure is specified).

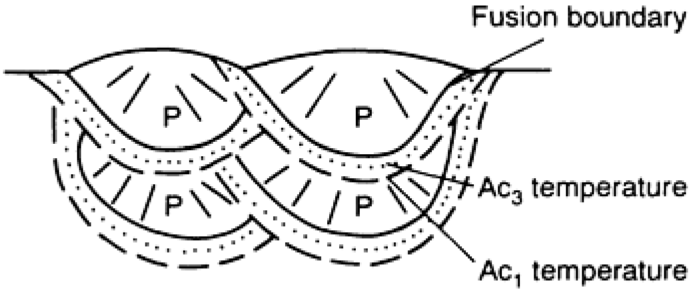

FIG. 33 PRIMARY (P) AND REAUSTENITIZED REGIONS IN THE WELD METAL REGION OF A MULTIPASS WELD. THE REAUSTENITIZED REGION IS WHERE THE COLUMNAR STRUCTURE IS NOT CLEARLY DETECTED. SOURCE: REF 23

Reference cited in this section

23. "GUIDE TO THE LIGHT MICROSCOPE EXAMINATION OF FERRITIC STEEL WELD METALS," DOC. NO. IX-1533-88, INTERNATIONAL INSTITUTE OF WELDING, 1988

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

Stainless Steels

It has long been known that solidification cracking can be avoided in austenitic stainless steel welds by having a small concentration of ferrite in them. Recent work has shown, however, that residual ferrite content at room temperature is no more than a symptom and that it is really the solidification mode (whether the weld metal solidifies as primary austenite or ferrite) that is the deciding criterion (Ref 43). It has been found that susceptibility to solidification cracking is least for a primary ferrite solidification mode (more specifically, when the solidification mode corresponds to the types shown in Fig. 34c and 34d). It is believed that low-melting-point liquid phases (formed by the segregation of impurities like sulfur and phosphorus, for example) solidifying in the intercellular regions do not wet the δ-γ interphase boundaries as easily as they would δ-δ or γ-γ boundaries. In the ferritic-austenitic solidification mode (Fig. 34c and 34d), the δ-γ interphase boundary area is greater at temperatures just below the nominal solidus temperature, and this is the reason for a greater resistance to solidification cracking.

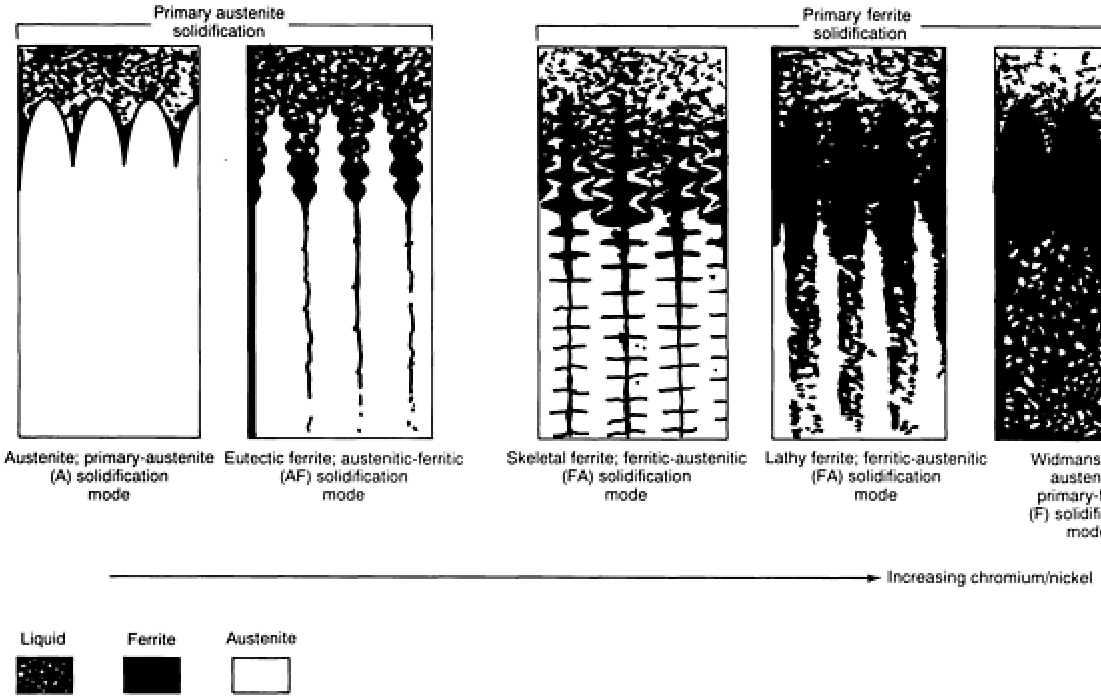
Primary austenite solidification
Primary ferrite solidification
Austenite; primary-austenite (A) solidification mode
Eutectic ferrite; austenitic-ferritic (AF) solidification mode
Skeletal ferrite; ferritic-austenitic (FA) solidification mode
Lathy ferrite; ferritic-austenitic (FA) solidification mode
Widmanstätten austenite; primary-ferrite (F) solidification mode
Increasing chromium/nickel
Liquid
Ferrite
Austenite

FIG. 34 SCHEMATIC SHOWING SOLIDIFICATION AND TRANSFORMATION BEHAVIOR RESULTING IN A RANGE OF FERRITE MORPHOLOGIES IN AUSTENITIC STAINLESS STEEL WELDS. SOURCE: REF 43

Information about the solidification mode can be obtained from the chemical composition of the weld metal by referring to the WRC-1992 diagram (Fig. 35). It has been found that this diagram is more accurate than the Schaeffler (for duplex stainless steel welds) and Delong (for 300 series austenitic steel welds) diagrams developed earlier. All of these diagrams are strictly applicable only for the cooling rates obtained in manual metal arc (MMA) welding because the data for establishing the statistical correlations have been taken from MMA welds. The dotted lines in Fig. 35 are the demarcation lines between the different solidification modes shown (that is, A, AF, FA, and F) corresponding to the types in Fig. 34(a), 34(b), 34(c), 34(d), and 34(e), respectively. The AF-FA boundary in Fig. 35 (which, as discussed earlier, corresponds to the onset of cracking) intersects the iso-ferrite number lines at an angle. This is consistent with the findings of several investigators that the minimum ferrite content necessary to avoid hot cracking is different for different weld metal compositions (that is, for weld metals deposited with E 316, E 308L, E 309, and so on).

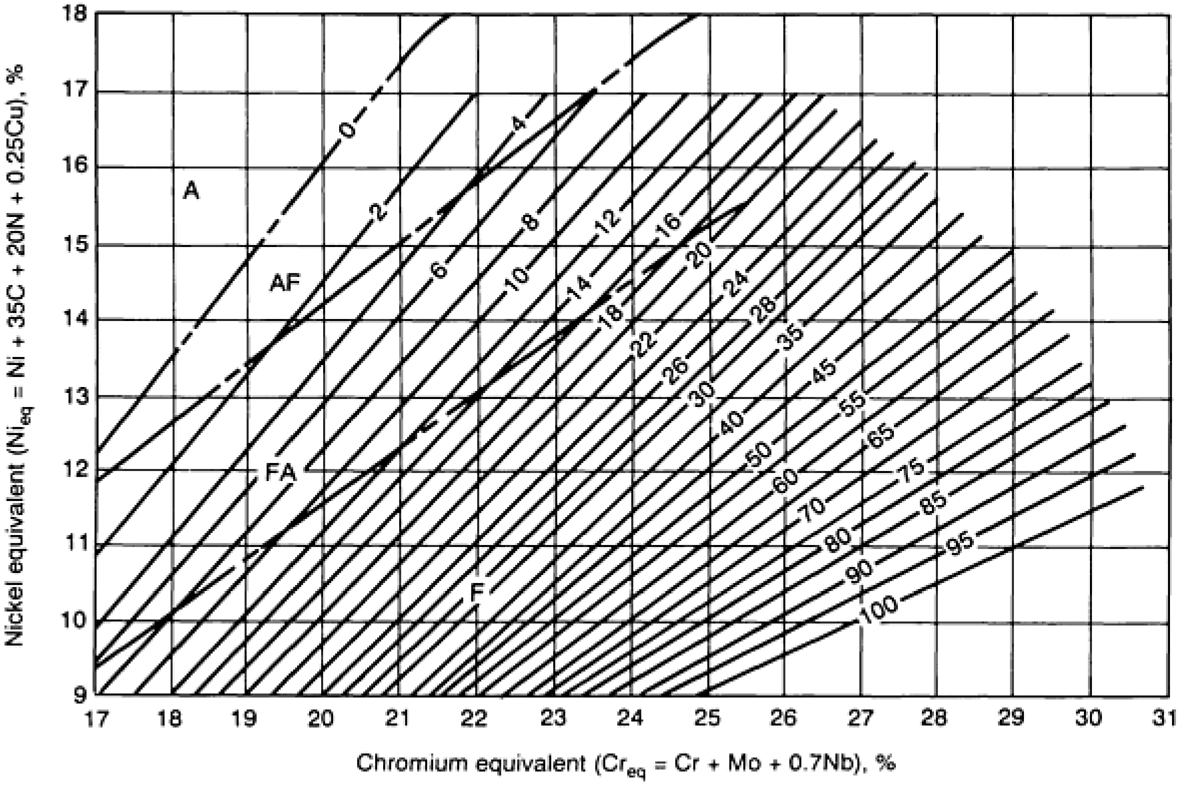

FIG. 35 WRC-1992 DIAGRAM PREDICTING FERRITE CONTENT IN STAINLESS STEELS. FERRITE CONTENT IS GIVEN BY THE FERRITE NUMBER (FN), WHERE 100 FN IS APPROXIMATELY EQUAL TO 65 VOL% FERRITE. BOUNDARIES DENOTING A CHANGE IN SOLIDIFICATION MODE (A, AF, FA AND F PER FIG. 34) ARE ALSO SHOWN (INDICATED BY DOTTED LINES). SOURCE: REF 45

It is a fortunate coincidence that the coefficients in the Cr_{eq} and Ni_{eq} expressions (see Fig. 35), which have been obtained by statistical fitting methods, result in satisfactory correlations with both the solidification mode and the ferrite number (a precisely measurable quantity that is roughly proportional to the actual volume percentage of ferrite at room temperature) (Ref 44). Indeed, it has been found that the same Cr_{eq} and Ni_{eq} expressions (that is, with the same elements and coefficients) are not able to satisfactorily fix the locations of the martensite and martensite-austenite phase fields seen in the lower left region of the Schaeffler diagram. For this reason, such regions of Cr_{eq} and Ni_{eq} are excluded in the WRC-1992 diagram.

A discussion of weld decay due to sensitization and knife line attack in the HAZ of austenitic stainless steels can be found in Ref 14 and 46.

References cited in this section

14. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987
43. J.A. BROOKS AND A.W. THOMPSON, MICROSTRUCTURAL DEVELOPMENT AND SOLIDIFICATION CRACKING SUSCEPTIBILITY OF AUSTENITIC STAINLESS STEEL WELDS, *INT. MATER. REV.,* VOL 36 (NO. 1), 1991
44. T.A. SIEWERT, C.N. MCCOWAN, AND D.L. OLSON, FERRITE NUMBER PREDICTION TO 100 FN IN STAINLESS STEEL WELD METAL, *WELD. J.,* DEC 1988, P 289S-298S
45. D.J. KOTECKI AND T.A. SIEWERT, WRC-1992 CONSTITUTION DIAGRAM FOR STAINLESS STEEL WELD METALS: A MODIFICATION OF THE WRC-1988 DIAGRAM, *WELD. J.,* MAY 1992, P 171S-178S
46. T.G. GOOCH AND D.C. WILLINGHAM, *WELD DECAY IN AUSTENITIC STAINLESS STEELS,* THE

WELDING INSTITUTE, UK, 1975

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

Aluminum Alloys

The main problems in the welding of heat-treatable aluminum alloys are liquation cracking in the partially melted zone (see the section "Unmixed and Partially Melted Zones in a Weldment" in this article) and softening in the HAZ (Ref 14, 47). The latter means that the weld joint strength can be reduced. Reasons for the softening can be understood by referring to Fig. 36. The parent material is assumed to be a 2000- or 6000-series aluminum alloy, artificially aged to contain the metastable phase of θ' (in aluminum-copper alloys), S' (in aluminum-copper-magnesium alloys), or β' (in aluminum-magnesium-silicon alloys). A very fine dispersion of precipitate particles is obtained in the parent material by a solution heat treatment followed by aging, and this is the reason for the high strength in the materials. Figure 36(b) shows that a gradient in hardness is obtained in the HAZ immediately after welding. At location 1, the high peak temperature causes the precipitate particles to go into solution, and the cooling rate is too high for reprecipitation. At locations 2 and 3, the precipitate particles partially dissolve and coarsen. After postweld artificial aging, fine precipitate particles are again formed at location 1, causing the hardness to increase to the parent material level. However, at locations 2 and 3, only a lower hardness can be obtained because of the formation of coarse precipitate particles. For increasing weld heat inputs, the width of the softened zone increases. A good discussion of how the above phenomenon can be modeled is found in Ref 48.

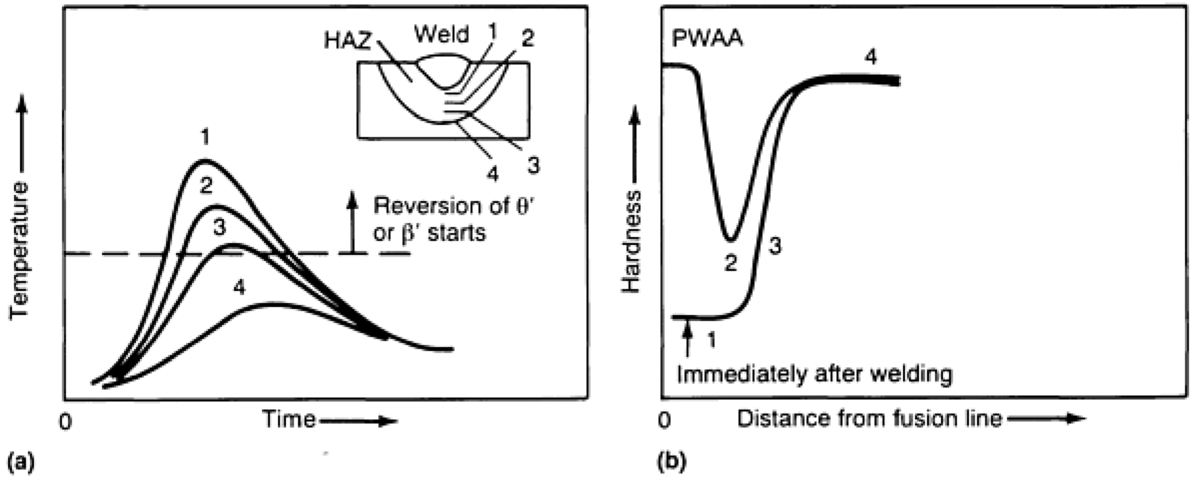

FIG. 36 SCHEMATIC SHOWING THE EFFECT OF WELD THERMAL CYCLES ON THE SOFTENING IN THE HAZ IN AGE-HARDENABLE ALUMINUM ALLOYS. (A) THERMAL CYCLES IN THE HAZ (FOR CORRESPONDING LOCATIONS IN WELD, SEE INSET. (B) HAZ HARDNESS PROFILES BEFORE AND AFTER AGING. PWAA, POSTWELD ARTIFICIAL AGING. SOURCE: REF 14

References cited in this section

14. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987
47. S. KOU, WELDING METALLURGY AND WELDABILITY OF HIGH STRENGTH ALUMINUM ALLOYS, *WELD. RES. COUNC. BULL.,* NO. 320, DEC 1986

48. H. SHERCLIFF AND M.F. ASHBY, A PROCESS MODEL FOR AGE HARDENING OF AL ALLOYS--I. MODEL AND II. APPLICATIONS OF THE MODEL, *ACTA METALL. MATER.*, VOL 38 (NO. 10), 1990, P 1789-1812

P. Ravi Vishnu, Luleå University of Technology, Sweden

Nickel-Base Superalloys

Nickel-base superalloys have a very high creep resistance, making them attractive for critical high-temperature applications such as gas turbines. Yet it is this very characteristic that gives rise to the problem of strain age cracking in the weld HAZ (Ref 14, 19, 49). The high-temperature creep strength in the parent materials is obtained by forming fine precipitates of γ' (Ni_3(Al, Ti)) or γ" (Ni_3Nb) in them. In the region of the weld HAZ subjected to high peak temperatures, these precipitates dissolve, similar to the case of aluminum alloys. After welding, the weldments are again solution heat-treated and aged (note the difference between aluminum alloys and the nickel-base superalloys in this respect). The purposes of this treatment are to obtain the same uniform and fine precipitation in the weld as in the parent material and to relieve the weld residual stresses.

Normally (for example, in the case of low-alloy or carbon steel welds), PWHT in a furnace relieves the residual stresses because the yield stress at the PWHT temperature is very low. A residual stress greater than the yield stress cannot be supported, and stress relief is brought about by a process of creep and plastic deformation. This process is difficult in nickel-base superalloy weldments, however, because of their high creep strength and low creep ductility. Reprecipitation in the solutionized regions of the HAZ (during the heating-up to the PWHT temperature) strengthens the matrix and does not allow any creep or plastic deformation (necessary for the relief of residual stress) to take place. Cracking can occur in this condition at the grain boundaries in the grain-coarsened region of the HAZ. Cracks initiated here have been known to run far into the base material. This problem is very difficult to solve, and the best solution has been found to be to switch over to materials such as Inconel 718 alloy that are inherently less susceptible to cracking, even though their creep strength is somewhat lower. Inconel 718 alloy contains γ" precipitate, and the slow kinetics of precipitation in this alloy allow substantial stress relief to be achieved before precipitation increases the creep strength in the HAZ.

References cited in this section

14. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987
19. W. YENISCAVITCH, JOINING, *SUPERALLOYS II,* C.T. SIMS, N.S. STOLOFF, AND W.C. HAGEL, ED., JOHN WILEY & SONS, 1987, P 495-516
49. W.A. OWCZARSKI, PROCESS AND METALLURGICAL FACTORS IN JOINING SUPERALLOYS AND OTHER HIGH SERVICE TEMPERATURE MATERIALS, *PHYSICAL METALLURGY OF JOINING,* R. KOSSOWSKY AND M.E. GLICKSMAN, ED., TMS-AIME, 1980

P. Ravi Vishnu, Luleå University of Technology, Sweden

Titanium Alloys

Titanium alloys find use in aerospace applications, pressure vessels, and so on because of their high strength-to-weight ratio, high corrosion resistance, and high fracture toughness. Pure titanium has two allotropic forms: low-temperature hexagonal close-packed (hcp) α-phase and elevated-temperature body-centered cubic (bcc) β-phase. Various alloying elements tend to preferentially stabilize one or the other of these phases. As a result, titanium alloys are generally classified as α, α+ β, metastable β, and β alloys (Fig. 37).

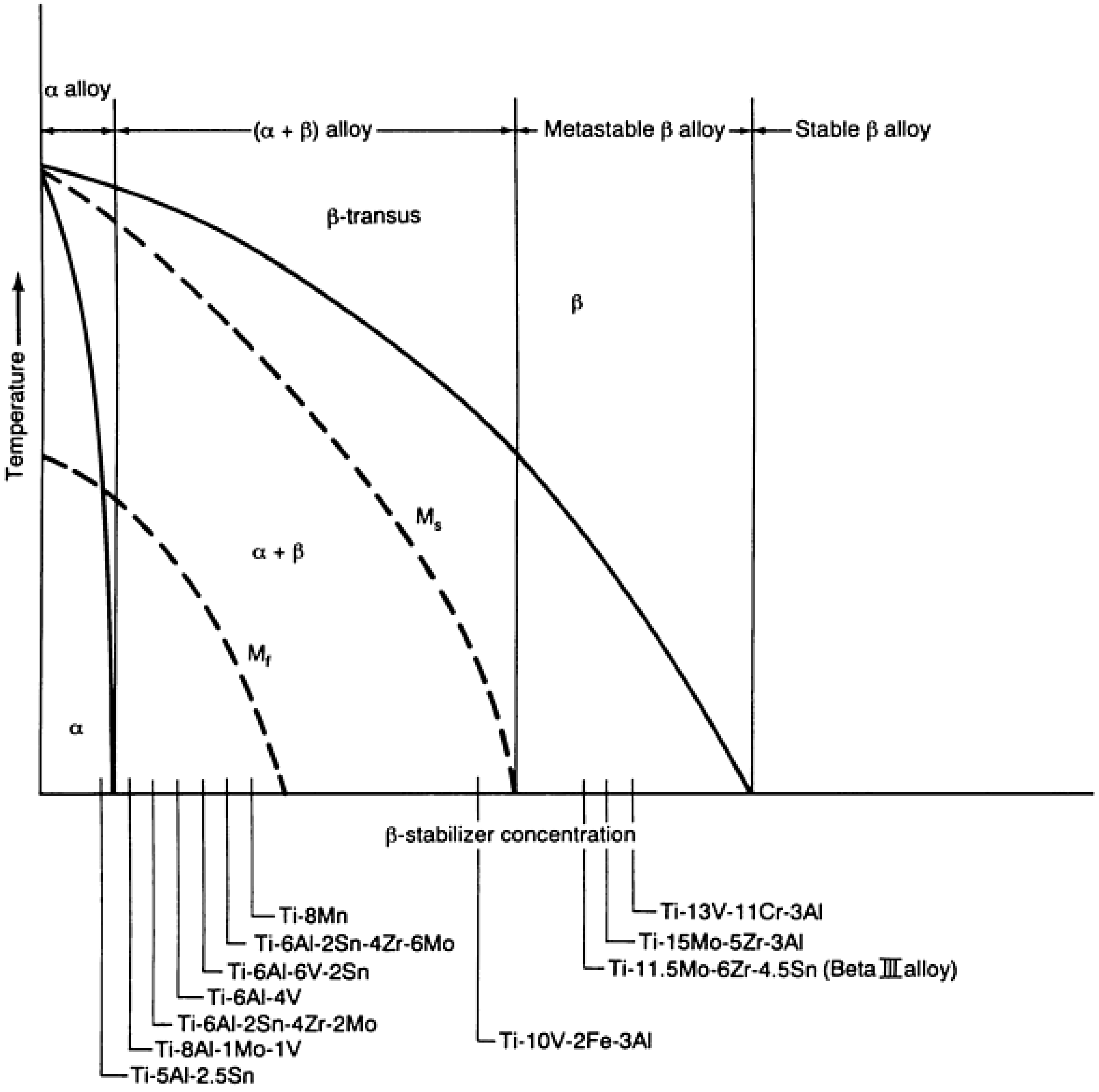
α alloy
(α + β) alloy
Metastable β alloy
Stable β alloy
β-transus
β
Temperature
M_s
α + β
M_f
α
β-stabilizer concentration
Ti-8Mn
Ti-6Al-2Sn-4Zr-6Mo
Ti-6Al-6V-2Sn
Ti-6Al-4V
Ti-6Al-2Sn-4Zr-2Mo
Ti-8Al-1Mo-1V
Ti-5Al-2.5Sn
Ti-10V-2Fe-3Al
Ti-13V-11Cr-3Al
Ti-15Mo-5Zr-3Al
Ti-11.5Mo-6Zr-4.5Sn (Beta III alloy)

FIG. 37 HYPOTHETICAL β ISOMORPHOUS PHASE DIAGRAM THAT SHOWS RELATION BETWEEN β -STABILIZER SOLUTE CONTENT AND TITANIUM ALLOY FAMILY DESCRIPTION. SELECTED COMMON ALLOYS ARE LISTED ALONG THE *X*-AXIS OF THE DIAGRAM BASED ON THEIR COMPOSITIONS. SOURCE: REF 50

The α-alloys are not heat-treatable (that is, a significant increase in strength and toughness cannot be obtained by heat treatment), and they have the lowest strength of the four titanium alloy families. Interestingly enough, these alloys are thus considered to have excellent weldability. What this means in practice is that there is no significant degradation in the properties of the weld and HAZ with regard to the parent material due to the weld thermal cycles. Also, the β ↔ α transformation in the weld and near HAZ means that the extent of grain coarsening will not be so high as in other single-phase materials in which no phase transformation occurs upon heating or cooling, such as stabilized ferritic stainless steels (Ref 51), pure aluminum, and so on.

In the α+ β alloys, a higher strength is obtained with a higher volume fraction of the β phase, which in turn is obtained with a higher β -stabilizing alloy content (Fig. 37). The effect of cooling rate on the phase transformations in two α+ β alloys is shown schematically in Fig. 38. As in the case of steels, it is seen that the β ↔ αtransformation can occur either by nucleation and growth at slower cooling rates or martensitically at higher cooling rates. The martensitic product in the lean α+ β alloys has an hcp crystal structure (designated as α') or an orthorhombic one (designated as α") in the β-stabilized α+ β alloys. It is seen that the C-curve for the β ↔ intergranular α transformation is shifted to the right for alloys with increasing β-stabilizing alloy content, while that for the allotriomorphic α (more commonly referred to as GB

α) is relatively insensitive to alloy concentration (Ref 52). As a result, alloys that are fairly rich in β-stabilizing additions exhibit a stronger tendency to form a continuous network of GB α than do leaner α + β alloys.

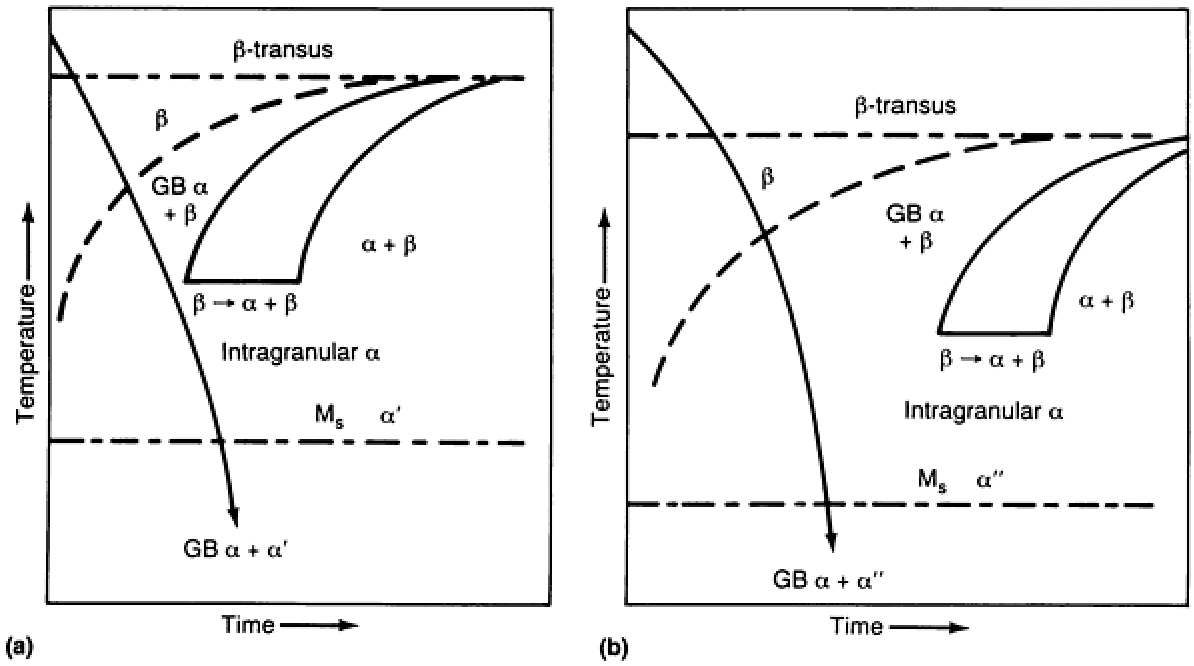

FIG. 38 SCHEMATIC CCT DIAGRAMS FOR TWO α + β TITANIUM ALLOYS WITH DIFFERENT β -STABILIZER SOLUTE CONTENTS. (A) LEAN α + β ALLOY (FOR EXAMPLE, TI-6AL-4V). (B) α + β ALLOY (FOR EXAMPLE, CORONA 5 [TI-4.5AL-5MO-1.5CR]) RICHER IN β -STABILIZING SOLUTE CONTENT

The morphology of the α and β phases in the α + β base materials is strongly dependent on thermomechanical processing (TMP) and heat treatment, which are performed either in the β-phase field or in the two-phase α + β phase field. When TMP is done in the α+ β phase field, the α that forms on cooling is continuously deformed. After recrystallization, a near-equiaxed α structure is obtained after this α + β processing. Very little GB α is formed because deformation introduces sufficient alternate heterogeneous nucleation sites in the form of dislocations. On the other hand, if TMP is done above the β-transus temperature (β-processing) and cooled, α forms at grain boundaries (giving GB α) and in the interior of grains as a Widmanstätten structure (also referred to as an acicular, lenticular, or basketweave structure). The microstructure in the weld and the near HAZ is closer to that obtained after β-processing.

Figure 39 shows the effect of α + β processing and β-processing on the strength and toughness of α + β alloys with increasing β-stabilizing alloy content. At the lower strength levels obtained with a lower β-stabilizing alloy content (because of a lower volume fraction of β-phase present), the finer Widmanstätten structure obtained after β-processing has a higher toughness than the coarser structure obtained by α + β processing. At higher strength levels, the trend reverses with the α + β processed structure having a higher toughness. This result can be understood in terms of the strength differences between GB α and the interior of the prior β grains (Fig. 40). At lower strengths, when the grain interiors are relatively weak, the crack tip plastic zone is distributed between the GB α and the grain interiors. However, as the α-phase precipitates are refined by lower aging temperatures, as in the case of the metastable β alloys, the grain interiors are strengthened considerably. Under such circumstances, the cracks are essentially constrained to follow the GB α layer. This results in a contraction of the plastic zone, a reduction in the crack tip opening displacement for crack extension, and an attendant drop in the toughness.

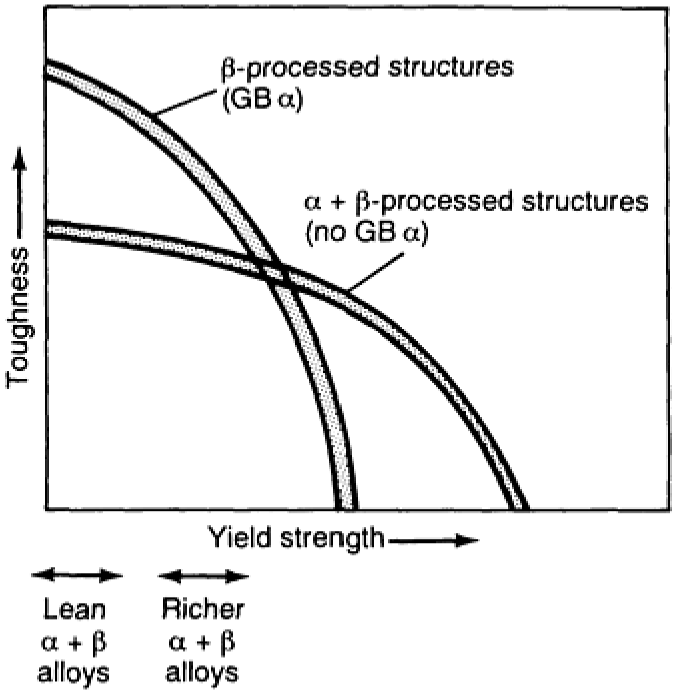

FIG. 39 SCHEMATIC PLOT OF TOUGHNESS VERSUS YIELD STRENGTH FOR α + β AND β-PROCESSED STRUCTURES. APPROXIMATE LOCATIONS OF LEAN (FOR EXAMPLE, TI-6AL-4V) AND RICHER β STABILIZED α + β ALLOYS (FOR EXAMPLE, CORONA 5 [TI-4.5AL-5MO-1.5CR]) ARE SHOWN. SOURCE: REF 52

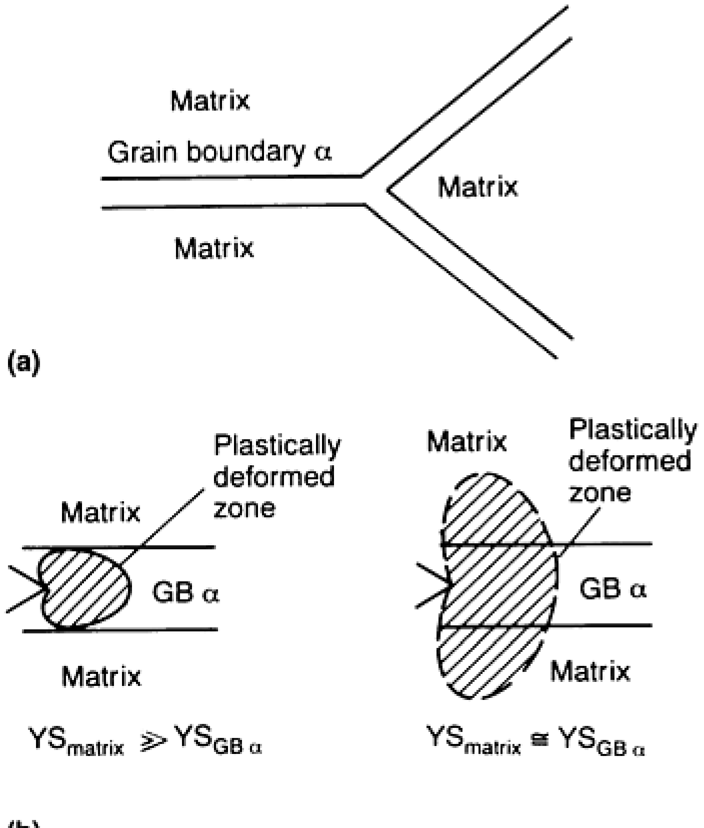
Matrix
Grain boundary α
Matrix
Matrix
(a)
Plastically deformed zone
Matrix
GB α
Matrix
$YS_{matrix} \gg YS_{GB\ \alpha}$
Matrix
Plastically deformed zone
GB α
Matrix
$YS_{matrix} \cong YS_{GB\ \alpha}$
(b)

FIG. 40 SCHEMATIC SHOWING THE EFFECT OF THE RELATIVE STRENGTHS OF GB α AND MATRIX ON THE

PLASTIC ZONE SIZE. (A) GB α LOCATION RELATIVE TO MATRIX. (B) EFFECT OF PLASTIC ZONE SIZE ON YIELD STRENGTHS OF MATRIX AND GB α. SOURCE: REF 52

The microstructures in the weld metal and the near HAZ (the region where the peak temperature exceeds the β-transus temperature) resemble those obtained after β-processing. This means that it is easier to get a matching toughness in the weld metal at the lower strength levels referred to in Fig. 39. Even this is made somewhat difficult by the coarser β grain size in the weld and HAZ. Grain coarsening in the HAZ induces coarse β grains to form in the weld also, because solidification in the weld metal occurs by epitaxial growth from the HAZ. The extent of grain coarsening is such that it is not uncommon to find a single columnar β grain traversing the entire thickness in gas-tungsten arc welds on thin sheet (Ref 53). If high energy density welding processes are used (for example, electron-beam welding or laser-beam welding), it is possible to limit the extent of β grain coarsening. However, the high cooling rates in the weld and HAZ will mean that a martensitic structure will be obtained.

To obtain adequate ductility and toughness, PWHT will have to be done at temperatures close to the β-transus temperature. This can be a problem because the low strength at these temperatures can cause sagging in large welded structures, and complex fixturing will be necessary to maintain dimensional tolerances. It can also be expensive, because inert gas shielding is required for almost the entire time of PWHT (it is assumed in the present discussion that adequate care is taken to maintain proper shielding during welding). Moreover, it can be difficult to increase the ductility in the weld metal and HAZ to a level equal to that in an α+ β processed base material (Fig. 41). The trends seen in Fig. 39 and 41 can be understood by noting that crack nucleation is more difficult in an equaixed structure (because the strain concentration effects are less) and that the plastic zone sizes are bigger (Table 1). This results in a higher tensile ductility. However, because crack propagation follows a less tortuous path, the toughness is lower. The reverse is true for a Widmanstätten structure (referred to as a lenticular structure in Table 1). Table 1 explains the apparently strange observation in titanium alloys that a structure having a high strength and toughness is not necessarily the one having a high tensile ductility. The implication of this for weldments is that for many applications it may be more important to optimize the welding and PWHT procedure with respect to the toughness than with respect to the ductility.

TABLE 1 EFFECT OF MORPHOLOGICAL FEATURES OF THE MICROSTRUCTURE ON CRACK NUCLEATION AND PROPAGATION IN α + β TITANIUM ALLOYS

Note: Voids initiated at the α -β interfaces are shown as black regions. Source: Ref 54

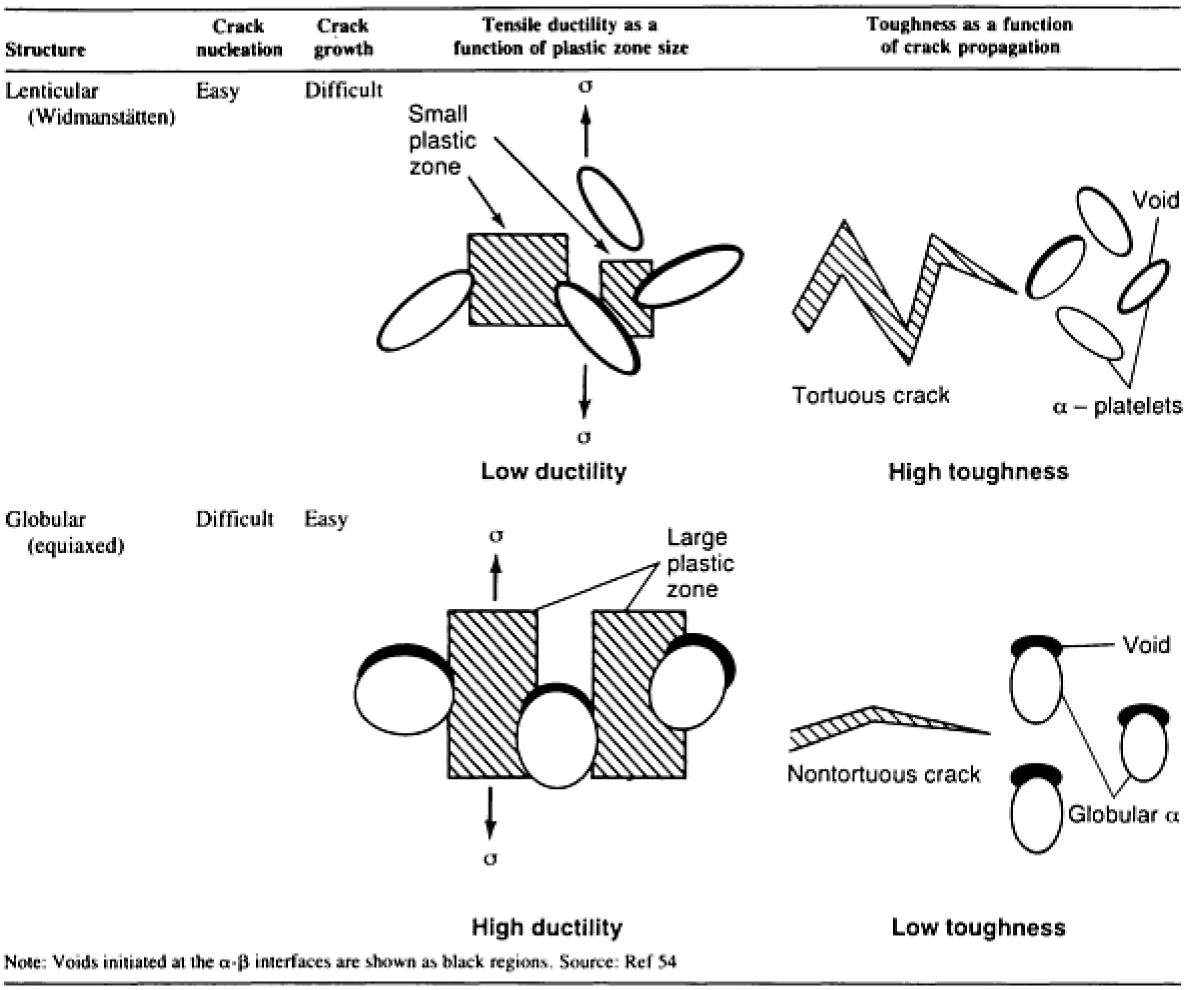

Structure	Crack nucleation	Crack growth	Tensile ductility as a function of plastic zone size	Toughness as a function of crack propagation
Lenticular (Widmanstätten)	Easy	Difficult	Low ductility	High toughness
Globular (equiaxed)	Difficult	Easy	High ductility	Low toughness

Note: Voids initiated at the α-β interfaces are shown as black regions. Source: Ref 54

σ
Small plastic zone
σ
Void
Tortuous crack
α – platelets
σ
Large plastic zone
σ
Void
Nontortuous crack
Globular α

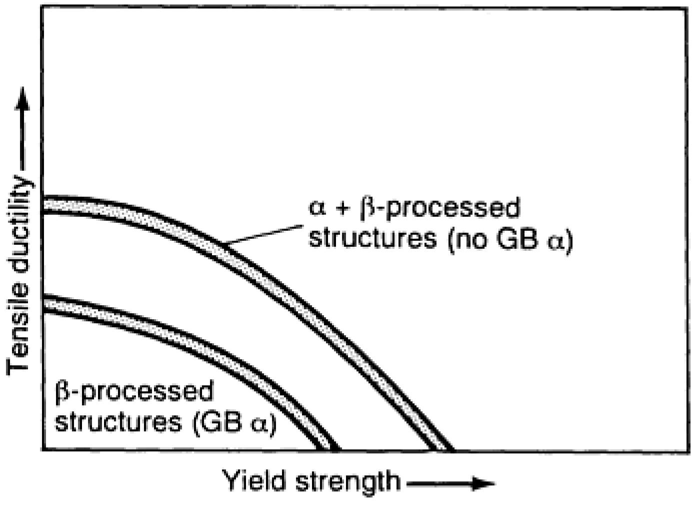

FIG. 41 SCHEMATIC PLOT OF DUCTILITY VERSUS STRENGTH FOR α + β AND β -PROCESSED STRUCTURES. SOURCE: REF 52

A more detailed discussion of the welding metallurgy of titanium alloys can be found in Ref 53, 55, and 56.

References cited in this section

50. E.W. COLLINGS, INTRODUCTION TO TITANIUM ALLOY DESIGN, *ALLOYING,* J.L. WALTER,

M.R. JACKSON, AND C.T. SIMS, ED., ASM INTERNATIONAL, 1988, P 267

51. C.R. THOMAS AND F.P.A. ROBINSON, KINETICS AND MECHANISM OF GRAIN GROWTH DURING WELDING IN NIOBIUM STABILISED 17% CHROMIUM STAINLESS STEELS, *MET. TECHNOL.,* APRIL 1978, P 133-138
52. J.C. WILLIAMS AND E.A. STARKE, JR., THE ROLE OF THERMOMECHANICAL PROCESSING IN TAILORING THE PROPERTIES OF ALUMINUM AND TITANIUM ALLOYS, *DEFORMATION, PROCESSING AND STRUCTURE,* G. KRAUSS, ED., AMERICAN SOCIETY FOR METALS, 1984, P 279-354
53. W.A. BAESLACK III, D.W. BECKER, AND F.H. FROES, ADVANCES IN TITANIUM ALLOY WELDING METALLURGY, *J. MET.,* MAY 1984, P 46-58
54. J.P. HIRTH AND F.H. FROES, INTERRELATIONS BETWEEN FRACTURE TOUGHNESS AND OTHER MECHANICAL PROPERTIES IN TITANIUM ALLOYS, *METALL. TRANS. A,* VOL 8, JULY 1977, P 1165-1176
55. C.G. RHODES, MICROSCOPY AND TITANIUM ALLOY DEVELOPMENT, *APPLIED METALLOGRAPHY,* G.F. VANDER VOORT, ED., VAN NOSTRAND REINHOLD, 1986
56. W.A. BAESLACK III, METALLOGRAPHY OF TITANIUM ALLOY WELDMENTS, *METALLOGRAPHY AND INTERPRETATION OF WELD MICROSTRUCTURES,* J.L. MCCALL, D.L. OLSON, AND I. LEMAY, ED., ASM INTERNATIONAL, 1987, P 23-60

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

Titanium Alloys

Titanium alloys find use in aerospace applications, pressure vessels, and so on because of their high strength-to-weight ratio, high corrosion resistance, and high fracture toughness. Pure titanium has two allotropic forms: low-temperature hexagonal close-packed (hcp) α-phase and elevated-temperature body-centered cubic (bcc) β-phase. Various alloying elements tend to preferentially stabilize one or the other of these phases. As a result, titanium alloys are generally classified as α, α+ β, metastable β, and βalloys (Fig. 37).

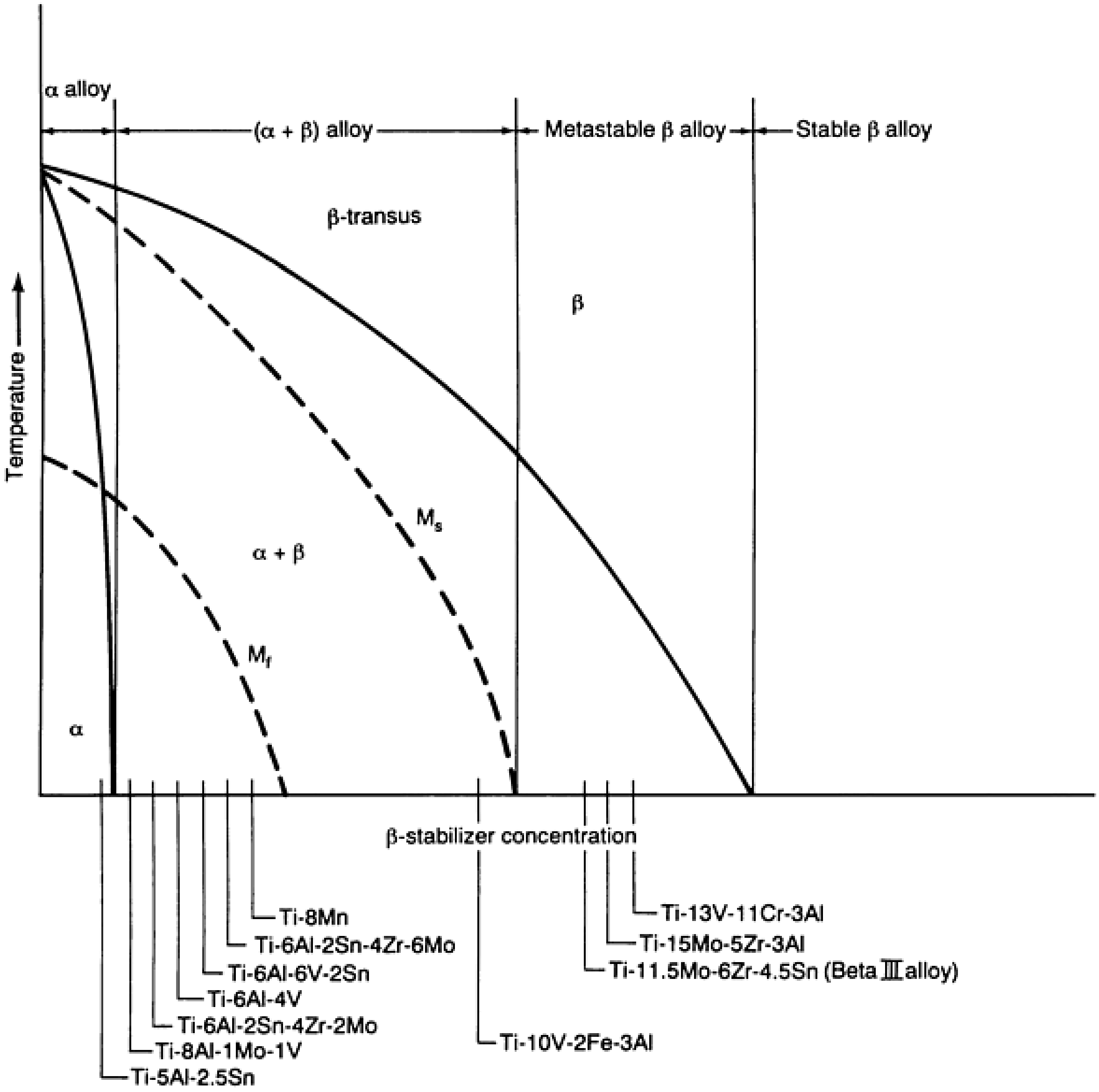
α alloy
(α + β) alloy
Metastable β alloy
Stable β alloy
β-transus
β
Temperature
Ms
α + β
Mf
α
β-stabilizer concentration
Ti-8Mn
Ti-6Al-2Sn-4Zr-6Mo
Ti-6Al-6V-2Sn
Ti-6Al-4V
Ti-6Al-2Sn-4Zr-2Mo
Ti-8Al-1Mo-1V
Ti-5Al-2.5Sn
Ti-10V-2Fe-3Al
Ti-13V-11Cr-3Al
Ti-15Mo-5Zr-3Al
Ti-11.5Mo-6Zr-4.5Sn (Beta III alloy)

FIG. 37 HYPOTHETICAL β ISOMORPHOUS PHASE DIAGRAM THAT SHOWS RELATION BETWEEN β -STABILIZER SOLUTE CONTENT AND TITANIUM ALLOY FAMILY DESCRIPTION. SELECTED COMMON ALLOYS ARE LISTED ALONG THE *X*-AXIS OF THE DIAGRAM BASED ON THEIR COMPOSITIONS. SOURCE: REF 50

The α-alloys are not heat-treatable (that is, a significant increase in strength and toughness cannot be obtained by heat treatment), and they have the lowest strength of the four titanium alloy families. Interestingly enough, these alloys are thus considered to have excellent weldability. What this means in practice is that there is no significant degradation in the properties of the weld and HAZ with regard to the parent material due to the weld thermal cycles. Also, the β ↔ α transformation in the weld and near HAZ means that the extent of grain coarsening will not be so high as in other single-phase materials in which no phase transformation occurs upon heating or cooling, such as stabilized ferritic stainless steels (Ref 51), pure aluminum, and so on.

In the α+ β alloys, a higher strength is obtained with a higher volume fraction of the β phase, which in turn is obtained with a higher β-stabilizing alloy content (Fig. 37). The effect of cooling rate on the phase transformations in two α + β alloys is shown schematically in Fig. 38. As in the case of steels, it is seen that the β ↔ α transformation can occur either by nucleation and growth at slower cooling rates or martensitically at higher cooling rates. The martensitic product in the lean α + β alloys has an hcp crystal structure (designated as α') or an orthorhombic one (designated as α ") in the β-stabilized α + β alloys. It is seen that the C-curve for the β ↔ intergranular α transformation is shifted to the right for alloys with increasing β-stabilizing alloy content, while that for the allotriomorphic α(more commonly referred to as GB

α) is relatively insensitive to alloy concentration (Ref 52). As a result, alloys that are fairly rich in β-stabilizing additions exhibit a stronger tendency to form a continuous network of GB α than do leaner α + β alloys.

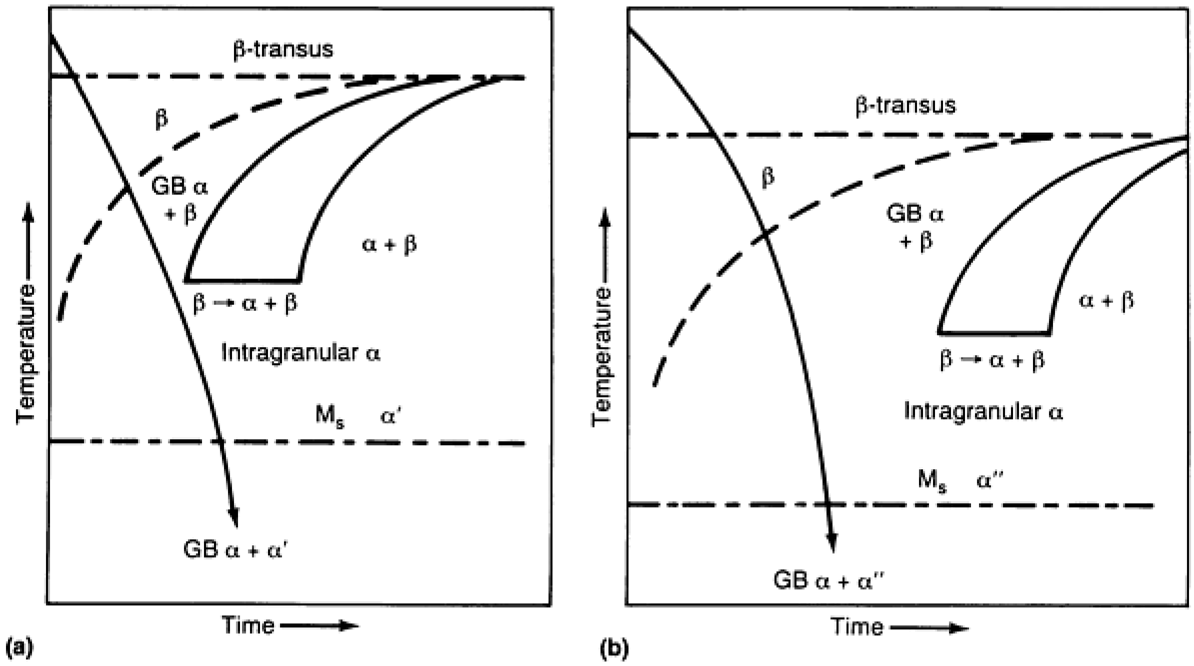

FIG. 38 SCHEMATIC CCT DIAGRAMS FOR TWO α + β TITANIUM ALLOYS WITH DIFFERENT β -STABILIZER SOLUTE CONTENTS. (A) LEAN α + β ALLOY (FOR EXAMPLE, TI-6AL-4V). (B) α + β ALLOY (FOR EXAMPLE, CORONA 5 [TI-4.5AL-5MO-1.5CR]) RICHER IN β -STABILIZING SOLUTE CONTENT

The morphology of the α and β phases in the α + β base materials is strongly dependent on thermomechanical processing (TMP) and heat treatment, which are performed either in the β-phase field or in the two-phase α+ β phase field. When TMP is done in the α+ β phase field, the α that forms on cooling is continuously deformed. After recrystallization, a near-equiaxed α structure is obtained after this α + β processing. Very little GB α is formed because deformation introduces sufficient alternate heterogeneous nucleation sites in the form of dislocations. On the other hand, if TMP is done above the β-transus temperature (β-processing) and cooled, α forms at grain boundaries (giving GB α) and in the interior of grains as a Widmanstätten structure (also referred to as an acicular, lenticular, or basketweave structure). The microstructure in the weld and the near HAZ is closer to that obtained after β-processing.

Figure 39 shows the effect of α + β processing and β-processing on the strength and toughness of α + β alloys with increasing β-stabilizing alloy content. At the lower strength levels obtained with a lower β-stabilizing alloy content (because of a lower volume fraction of β-phase present), the finer Widmanstätten structure obtained after β-processing has a higher toughness than the coarser structure obtained by α + β processing. At higher strength levels, the trend reverses with the α + β processed structure having a higher toughness. This result can be understood in terms of the strength differences between GB α and the interior of the prior β grains (Fig. 40). At lower strengths, when the grain interiors are relatively weak, the crack tip plastic zone is distributed between the GB α and the grain interiors. However, as the α-phase precipitates are refined by lower aging temperatures, as in the case of the metastable β alloys, the grain interiors are strengthened considerably. Under such circumstances, the cracks are essentially constrained to follow the GB α layer. This results in a contraction of the plastic zone, a reduction in the crack tip opening displacement for crack extension, and an attendant drop in the toughness.

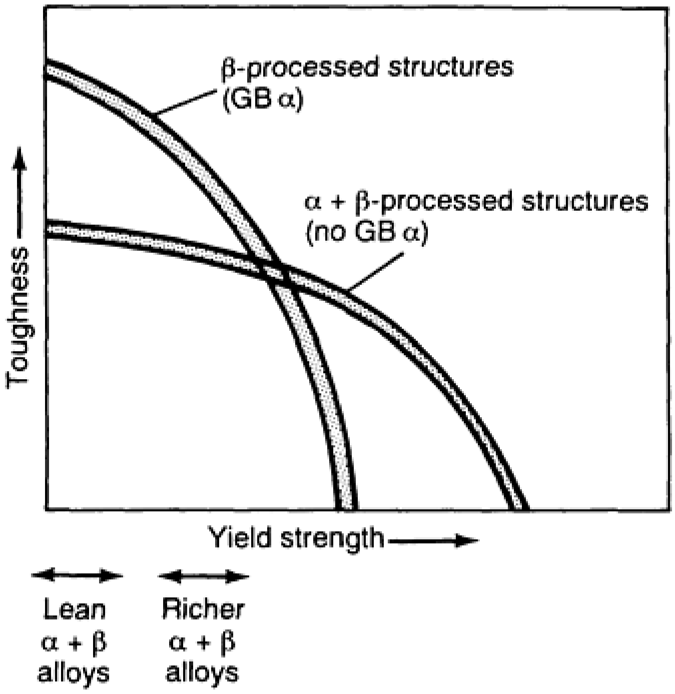

FIG. 39 SCHEMATIC PLOT OF TOUGHNESS VERSUS YIELD STRENGTH FOR α + β AND β -PROCESSED STRUCTURES. APPROXIMATE LOCATIONS OF LEAN (FOR EXAMPLE, TI-6AL-4V) AND RICHER β STABILIZED α + β ALLOYS (FOR EXAMPLE, CORONA 5 [TI-4.5AL-5MO-1.5CR]) ARE SHOWN. SOURCE: REF 52

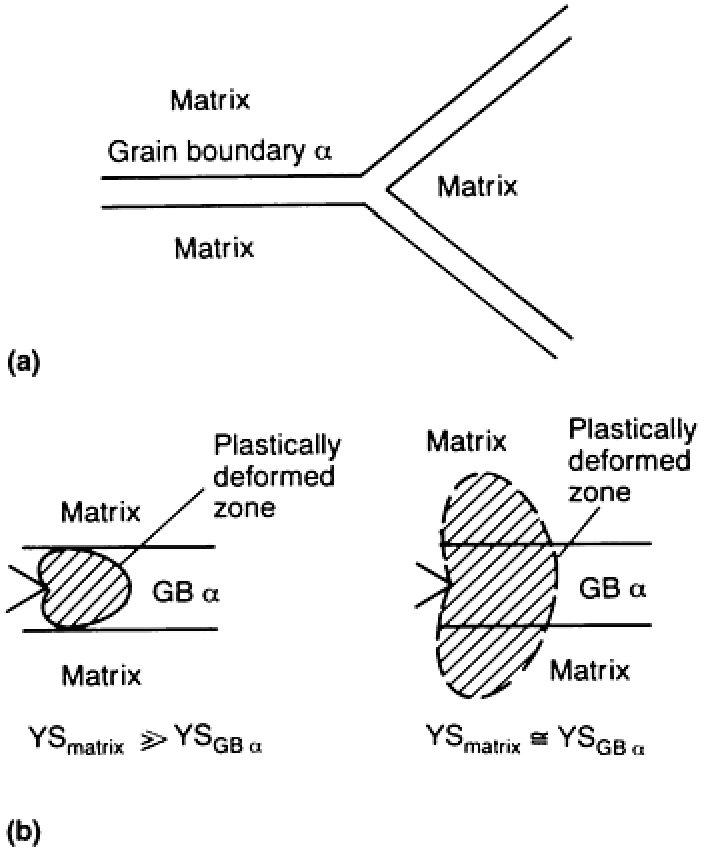
Matrix
Grain boundary α
Matrix
Matrix
(a)
Plastically deformed zone
Matrix
GB α
Matrix
$YS_{matrix} \gg YS_{GB\ \alpha}$
Matrix
Plastically deformed zone
GB α
Matrix
$YS_{matrix} \cong YS_{GB\ \alpha}$
(b)

FIG. 40 SCHEMATIC SHOWING THE EFFECT OF THE RELATIVE STRENGTHS OF GB α AND MATRIX ON THE PLASTIC ZONE SIZE. (A) GB α LOCATION RELATIVE TO MATRIX. (B) EFFECT OF PLASTIC ZONE SIZE ON YIELD

STRENGTHS OF MATRIX AND GB α. SOURCE: REF 52

The microstructures in the weld metal and the near HAZ (the region where the peak temperature exceeds the β-transus temperature) resemble those obtained after β-processing. This means that it is easier to get a matching toughness in the weld metal at the lower strength levels referred to in Fig. 39. Even this is made somewhat difficult by the coarser β grain size in the weld and HAZ. Grain coarsening in the HAZ induces coarse β grains to form in the weld also, because solidification in the weld metal occurs by epitaxial growth from the HAZ. The extent of grain coarsening is such that it is not uncommon to find a single columnar β grain traversing the entire thickness in gas-tungsten arc welds on thin sheet (Ref 53). If high energy density welding processes are used (for example, electron-beam welding or laser-beam welding), it is possible to limit the extent of β grain coarsening. However, the high cooling rates in the weld and HAZ will mean that a martensitic structure will be obtained.

To obtain adequate ductility and toughness, PWHT will have to be done at temperatures close to the β-transus temperature. This can be a problem because the low strength at these temperatures can cause sagging in large welded structures, and complex fixturing will be necessary to maintain dimensional tolerances. It can also be expensive, because inert gas shielding is required for almost the entire time of PWHT (it is assumed in the present discussion that adequate care is taken to maintain proper shielding during welding). Moreover, it can be difficult to increase the ductility in the weld metal and HAZ to a level equal to that in an α + β processed base material (Fig. 41). The trends seen in Fig. 39 and 41 can be understood by noting that crack nucleation is more difficult in an equaixed structure (because the strain concentration effects are less) and that the plastic zone sizes are bigger (Table 1). This results in a higher tensile ductility. However, because crack propagation follows a less tortuous path, the toughness is lower. The reverse is true for a Widmanstätten structure (referred to as a lenticular structure in Table 1). Table 1 explains the apparently strange observation in titanium alloys that a structure having a high strength and toughness is not necessarily the one having a high tensile ductility. The implication of this for weldments is that for many applications it may be more important to optimize the welding and PWHT procedure with respect to the toughness than with respect to the ductility.

TABLE 1 EFFECT OF MORPHOLOGICAL FEATURES OF THE MICROSTRUCTURE ON CRACK NUCLEATION AND PROPAGATION IN α+ β TITANIUM ALLOYS

Note: Voids initiated at the α-β interfaces are shown as black regions. Source: Ref 54

Structure	Crack nucleation	Crack growth	Tensile ductility as a function of plastic zone size	Toughness as a function of crack propagation
Lenticular (Widmanstätten)	Easy	Difficult	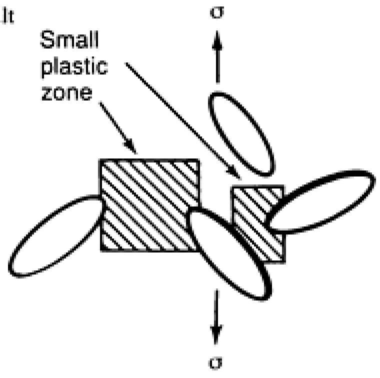 σ Small plastic zone σ **Low ductility**	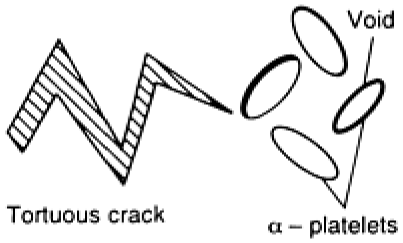 Void Tortuous crack α – platelets **High toughness**
Globular (equiaxed)	Difficult	Easy	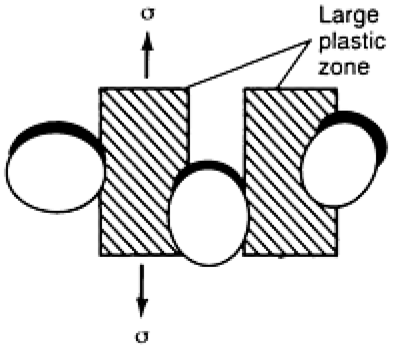 σ Large plastic zone σ **High ductility**	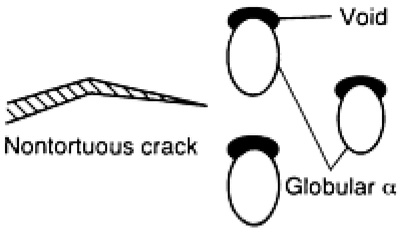 Void Nontortuous crack Globular α **Low toughness**

Note: Voids initiated at the α-β interfaces are shown as black regions. Source: Ref 54

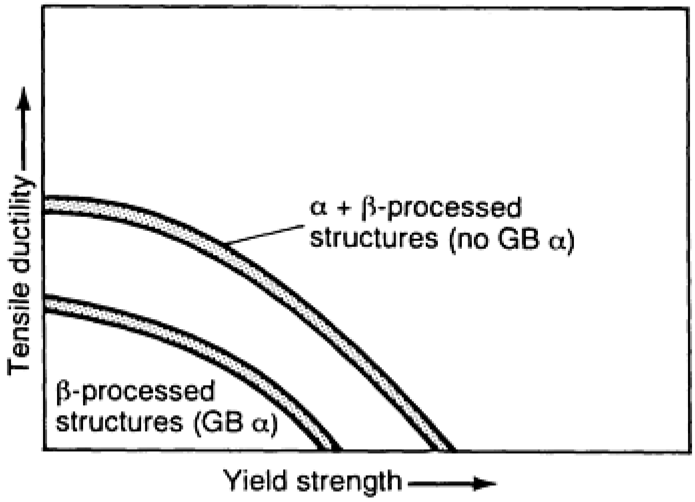

FIG. 41 SCHEMATIC PLOT OF DUCTILITY VERSUS STRENGTH FOR α+β AND β -PROCESSED STRUCTURES. SOURCE: REF 52

A more detailed discussion of the welding metallurgy of titanium alloys can be found in Ref 53, 55, and 56.

References cited in this section

50. E.W. COLLINGS, INTRODUCTION TO TITANIUM ALLOY DESIGN, *ALLOYING,* J.L. WALTER,

M.R. JACKSON, AND C.T. SIMS, ED., ASM INTERNATIONAL, 1988, P 267

51. C.R. THOMAS AND F.P.A. ROBINSON, KINETICS AND MECHANISM OF GRAIN GROWTH DURING WELDING IN NIOBIUM STABILISED 17% CHROMIUM STAINLESS STEELS, *MET. TECHNOL.,* APRIL 1978, P 133-138
52. J.C. WILLIAMS AND E.A. STARKE, JR., THE ROLE OF THERMOMECHANICAL PROCESSING IN TAILORING THE PROPERTIES OF ALUMINUM AND TITANIUM ALLOYS, *DEFORMATION, PROCESSING AND STRUCTURE,* G. KRAUSS, ED., AMERICAN SOCIETY FOR METALS, 1984, P 279-354
53. W.A. BAESLACK III, D.W. BECKER, AND F.H. FROES, ADVANCES IN TITANIUM ALLOY WELDING METALLURGY, *J. MET.,* MAY 1984, P 46-58
54. J.P. HIRTH AND F.H. FROES, INTERRELATIONS BETWEEN FRACTURE TOUGHNESS AND OTHER MECHANICAL PROPERTIES IN TITANIUM ALLOYS, *METALL. TRANS. A,* VOL 8, JULY 1977, P 1165-1176
55. C.G. RHODES, MICROSCOPY AND TITANIUM ALLOY DEVELOPMENT, *APPLIED METALLOGRAPHY,* G.F. VANDER VOORT, ED., VAN NOSTRAND REINHOLD, 1986
56. W.A. BAESLACK III, METALLOGRAPHY OF TITANIUM ALLOY WELDMENTS, *METALLOGRAPHY AND INTERPRETATION OF WELD MICROSTRUCTURES,* J.L. MCCALL, D.L. OLSON, AND I. LEMAY, ED., ASM INTERNATIONAL, 1987, P 23-60

Solid-State Transformations in Weldments

P. Ravi Vishnu, Luleå University of Technology, Sweden

References

1. *ATLAS OF ISOTHERMAL TRANSFORMATION AND COOLING TRANSFORMATION DIAGRAMS* AMERICAN SOCIETY FOR METALS, 1977
2. O.M. AKSELSEN AND T. SIMONSEN, TECHNIQUES FOR EXAMINING TRANSFORMATION BEHAVIOUR IN WELD METAL AND HAZ--A STATE OF ART REVIEW, *WELD. WORLD,* VOL 25 (NO. 1/2), 1987, P 26-34
3. O. GRONG AND D.K. MATLOCK, MICROSTRUCTURAL DEVELOPMENT IN MILD AND LOW ALLOY STEEL WELD METALS, *INT. MET. REV.,* VOL 31 (NO. 1), 1986, P 27-48
4. P.L. HARRISON AND R.A. FARRAR, APPLICATION OF CONTINUOUS COOLING TRANSFORMATION DIAGRAMS FOR WELDING OF STEELS, *INT. MATER. REV.,* VOL 34 (NO. 1), 1989, P 35-51
5. G.T. ELDIS, A CRITICAL REVIEW OF DATA SOURCES FOR ISOTHERMAL TRANSFORMATION AND CONTINUOUS COOLING TRANSFORMATION DIAGRAMS, *HARDENABILITY CONCEPTS WITH APPLICATION TO STEEL,* D.V. DOANE AND J.S. KIRKALDY, ED., TMS-AIME, 1978, P 126-157
6. K. EASTERLING, *INTRODUCTION TO THE PHYSICAL METALLURGY OF WELDING,* BUTTERWORTHS, 1983
7. *WELD THERMAL SIMULATORS FOR RESEARCH AND PROBLEM SOLVING,* THE WELDING INSTITUTE, U.K., 1972
8. C.F. BERKHOUT AND P.H. VAN LENT, THE USE OF PEAK TEMPERATURE-COOLING TIME DIAGRAMS IN THE WELDING OF HIGH STRENGTH STEELS (IN GERMAN), *SCHWEISS. SCHNEID.,* VOL 20 (NO. 6), 1968, P 256-260
9. TH.J. VAN ADRICHEM AND J. KAS, *CALCULATION, MEASUREMENT AND SIMULATION OF WELD THERMAL CYCLES,* SMITWELD N.V., NIJMEGEN, THE NETHERLANDS

10. P. RAVI VISHNU, W.B. LI, AND K. EASTERLING, HEAT FLOW MODEL FOR PULSED WELDING, *MATER. SCI. TECH.*, VOL 7 (NO. 7), 1991, P 649-659
11. P. RAVI VISHNU AND K. EASTERLING, PHENOMENOLOGICAL MODELLING OF HEAT FLOW AND MICROSTRUCTURAL CHANGES IN PULSED GTAW WELDS OF A QT STEEL, *MATHEMATICAL MODELING OF WELD PHENOMENA,* H. CERJAK AND K. EASTERLING, ED., THE INSTITUTE OF METALS, 1993
12. VEREIN DEUTSCHER EISENHÜTTENLEUTE, ED., *STEEL--A HANDBOOK FOR MATERIALS RESEARCH AND ENGINEERING, VOLUME 1: FUNDAMENTALS,* SPRINGER VERLAG, 1992, P 175
13. A.J. DEARDO, C.I. GARCIA, AND E.J. PALMIERE, THERMOMECHANICAL PROCESSING OF STEELS, *HEAT TREATING,* VOL 4, *ASM HANDBOOK,* 1991, P 237-255
14. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987
15. K. NARITA, PHYSICAL CHEMISTRY OF THE GROUPS IVA (TI, ZR), VA (V, NB, TA) AND THE RARE EARTH ELEMENTS IN STEEL, *TRANS. ISIJ,* VOL 15, 1975, P 145-152
16. R.K. AMIN AND F.B. PICKERING, AUSTENITE GRAIN COARSENING AND THE EFFECT OF THERMOMECHANICAL PROCESSING ON AUSTENITE RECRYSTALLIZATION, *THERMOMECHANICAL PROCESSING OF MICROALLOYED AUSTENITE,* A.J. DEARDO, G.A. RATZ, AND P.J. WRAY, ED., TMS-AIME, 1982, P 14
17. W.F. SAVAGE, E.F. NIPPES, AND E.S SZEKERES, A STUDY OF WELD INTERFACE PHENOMENA IN A LOW ALLOY STEEL, *WELD. J.,* SEPT 1976, P 260S-268S, AND HYDROGEN INDUCED COLD CRACKING IN A LOW ALLOY STEEL, *WELD. J.*, SEPT 1976, P 276S-283S
18. J.J. PEPE AND W.F. SAVAGE, EFFECTS OF CONSTITUTIONAL LIQUATION IN 18-NI MARAGING STEEL WELDMENTS, *WELD J.,* SEPT 1967, P 411S-422S
19. W. YENISCAVITCH, JOINING, *SUPERALLOYS II,* C.T. SIMS, N.S. STOLOFF, AND W.C. HAGEL, ED., JOHN WILEY & SONS, 1987, P 495-516
20. E.E. HUCKE AND H. UDIN, WELDING METALLURGY OF NODULAR CAST IRON, *WELD. J.,* AUG 1953, P 378S-385S
21. E.N. GREGORY AND S.B. JONES, WELDING CAST IRONS, *WELDING OF CASTINGS,* THE WELDING INSTITUTE, U.K., 1977, P 145-156
22. L.E. SAMUELS, *OPTICAL MICROSCOPY OF CARBON STEELS,* AMERICAN SOCIETY FOR METALS, 1980
23. "GUIDE TO THE LIGHT MICROSCOPE EXAMINATION OF FERRITIC STEEL WELD METALS," DOC. NO. IX-1533-88, INTERNATIONAL INSTITUTE OF WELDING, 1988
24. H.K.D.H. BHADESHIA AND L.E. SVENSSON, MODELLING THE EVOLUTION OF MICROSTRUCTURE OF STEEL WELD METAL, *MATHEMATICAL MODELLING OF WELD PHENOMENA,* H. CERJAK AND K. EASTERLING, ED., THE INSTITUTE OF MATERIALS, 1993
25. G.R. EDWARDS AND S. LIU, RECENT DEVELOPMENTS IN HSLA STEEL WELDING, *ADVANCES IN WELDING METALLURGY,* FIRST U.S.-JAPAN SYMPOSIUM, AMERICAN WELDING SOCIETY AND OTHERS, JUNE 1990, P 215-293
26. B. AHLBLOM, "OXYGEN AND ITS ROLE IN DETERMINING WELD METAL MICROSTRUCTURE AND TOUGHNESS--A STATE OF THE ART REVIEW," DOC. NO. IX-1322-84, INTERNATIONAL INSTITUTE OF WELDING, 1984
27. L.-E. SVENSSON AND B. GRETOFT, MICROSTRUCTURE AND IMPACT TOUGHNESS OF C-MN WELD METALS. *WELD. J.,* DEC 1990, P 454S-461S
28. P.W. HOLSBERG AND R.J. WONG, WELDING OF HSLA-100 STEEL FOR NAVAL APPLICATIONS, *WELDABILITY OF MATERIALS,* R.A. PATTERSON AND K.W. MAHIN, ED., ASM INTERNATIONAL, 1990
29. J.M.B. LOSZ AND K.D. CHALLENGER, MICROSTRUCTURE AND PROPERTIES OF A COPPER PRECIPITATION STRENGTHENED HSLA STEEL WELDMENT, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID AND J.M. VITEK, ED., ASM INTERNATIONAL, 1990
30. H. HOMMA, S. OHKITA, S. MATSUDA, AND K. YAMAMOTO, IMPROVEMENTS OF HAZ

TOUGHNESS IN HSLA STEEL BY INTRODUCING FINELY DISPERSED TI-OXIDE, *WELD. J.,* OCT 1987, P 301S-309S

31. N. YURIOKA, "MODERN HIGH STRENGTH STEEL IN JAPAN," FIFTH INTERNATIONAL SYMPOSIUM, JAPAN WELDING SOCIETY, APRIL 1990 (TOKYO)
32. YU.N. GOTALSKII, THE PROBLEM OF WELDING HIGH STRENGTH STEELS, *AUTOM. WELD.*, JUNE 1984, P 37-40
33. W.K.C. JONES AND P.J. ALBERRY, THE ROLE OF PHASE TRANSFORMATIONS IN THE DEVELOPMENT OF RESIDUAL STRESSES DURING THE WELDING OF SOME FAST REACTOR STEELS, *PROC. CONF. FERRITIC STEELS FOR FAST REACTOR STEAM GENERATORS,* BRITISH NUCLEAR ENERGY SOCIETY, 1977
34. *MCGRAW-HILL ENCYCLOPEDIA OF SCIENCE AND TECHNOLOGY,* VOL 9, P 685
35. A.M. MAKARA AND N.A. MOSENDZ, EFFECTS OF THE WELD METAL ON CRACKING IN HAZ, *WELD. RES. ABROAD,* NOV 1965, P 78-86
36. H. IKAWA, S. SHIN, M. INUI, Y. TAKEDA, AND A. NAKANO, "ON THE MARTENSITE-LIKE STRUCTURE AT WELD BOND AND THE MACROSCOPIC SEGREGATION IN WELD METAL IN THE WELDED DISSIMILAR METALS OF α-STEELS AND γ-STEELS," DOC. NO. IX-785-72, INTERNATIONAL INSTITUTE OF WELDING, 1972
37. T. HAZE AND S. AIHARA, "INFLUENCE OF TOUGHNESS AND SIZE OF LOCAL BRITTLE ZONES ON HAZ TOUGHNESS OF HSLA STEELS," SEVENTH INTERNATIONAL CONFERENCE ON OFFSHORE MECHANICS AND ARCTIC ENGINEERING, HOUSTON, 1988
38. C.P. ROYER, A USER'S PERSPECTIVE ON HEAT AFFECTED ZONE TOUGHNESS, *WELDING METALLURGY OF STRUCTURAL STEELS,* J.Y. KOO, ED., TMS-AIME, 1987
39. F. MATSUDA, Z. LI, P. BERNASOVSKY, K. ISHIHARA, AND H. OKADA, "AN INVESTIGATION OF THE BEHAVIOUR OF M-A CONSTITUENT IN SIMULATED HAZ OF HSLA STEELS," DOC. NO. IX-B-1591-90, INTERNATIONAL INSTITUTE OF WELDING, 1990
40. P.L. HARRISON AND P.H.M. HART, RELATIONSHIPS BETWEEN HAZ MICROSTRUCTURE AND CTOD TRANSITION BEHAVIOUR IN MULTIPASS C-MN STEEL WELDS, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID AND J.M. VITEK, ED., ASM INTERNATIONAL, 1990
41. H. ONOE, J. TANAKA, AND I. WATANABE, JAPANESE LNG TANKER CONSTRUCTED USING A NEW WELDING PROCESS AND IMPROVED AL-KILLED STEELS, *MET. CONSTR.,* JAN 1979, P 26-31
42. P.J. ALBERRY, SENSITIVITY ANALYSIS OF HALF-BEAD AND ALTERNATIVE GTAW TECHNIQUES, *WELD. J.,* NOV 1989, P 442S-451S
43. J.A. BROOKS AND A.W. THOMPSON, MICROSTRUCTURAL DEVELOPMENT AND SOLIDIFICATION CRACKING SUSCEPTIBILITY OF AUSTENITIC STAINLESS STEEL WELDS, *INT. MATER. REV.,* VOL 36 (NO. 1), 1991
44. T.A. SIEWERT, C.N. MCCOWAN, AND D.L. OLSON, FERRITE NUMBER PREDICTION TO 100 FN IN STAINLESS STEEL WELD METAL, *WELD. J.,* DEC 1988, P 289S-298S
45. D.J. KOTECKI AND T.A. SIEWERT, WRC-1992 CONSTITUTION DIAGRAM FOR STAINLESS STEEL WELD METALS: A MODIFICATION OF THE WRC-1988 DIAGRAM, *WELD. J.,* MAY 1992, P 171S-178S
46. T.G. GOOCH AND D.C. WILLINGHAM, *WELD DECAY IN AUSTENITIC STAINLESS STEELS,* THE WELDING INSTITUTE, UK, 1975
47. S. KOU, WELDING METALLURGY AND WELDABILITY OF HIGH STRENGTH ALUMINUM ALLOYS, *WELD. RES. COUNC. BULL.,* NO. 320, DEC 1986
48. H. SHERCLIFF AND M.F. ASHBY, A PROCESS MODEL FOR AGE HARDENING OF AL ALLOYS--I. MODEL AND II. APPLICATIONS OF THE MODEL, *ACTA METALL. MATER.*, VOL 38 (NO. 10), 1990, P 1789-1812
49. W.A. OWCZARSKI, PROCESS AND METALLURGICAL FACTORS IN JOINING SUPERALLOYS

AND OTHER HIGH SERVICE TEMPERATURE MATERIALS, *PHYSICAL METALLURGY OF JOINING,* R. KOSSOWSKY AND M.E. GLICKSMAN, ED., TMS-AIME, 1980

50. E.W. COLLINGS, INTRODUCTION TO TITANIUM ALLOY DESIGN, *ALLOYING,* J.L. WALTER, M.R. JACKSON, AND C.T. SIMS, ED., ASM INTERNATIONAL, 1988, P 267
51. C.R. THOMAS AND F.P.A. ROBINSON, KINETICS AND MECHANISM OF GRAIN GROWTH DURING WELDING IN NIOBIUM STABILISED 17% CHROMIUM STAINLESS STEELS, *MET. TECHNOL.,* APRIL 1978, P 133-138
52. J.C. WILLIAMS AND E.A. STARKE, JR., THE ROLE OF THERMOMECHANICAL PROCESSING IN TAILORING THE PROPERTIES OF ALUMINUM AND TITANIUM ALLOYS, *DEFORMATION, PROCESSING AND STRUCTURE,* G. KRAUSS, ED., AMERICAN SOCIETY FOR METALS, 1984, P 279-354
53. W.A. BAESLACK III, D.W. BECKER, AND F.H. FROES, ADVANCES IN TITANIUM ALLOY WELDING METALLURGY, *J. MET.,* MAY 1984, P 46-58
54. J.P. HIRTH AND F.H. FROES, INTERRELATIONS BETWEEN FRACTURE TOUGHNESS AND OTHER MECHANICAL PROPERTIES IN TITANIUM ALLOYS, *METALL. TRANS. A,* VOL 8, JULY 1977, P 1165-1176
55. C.G. RHODES, MICROSCOPY AND TITANIUM ALLOY DEVELOPMENT, *APPLIED METALLOGRAPHY,* G.F. VANDER VOORT, ED., VAN NOSTRAND REINHOLD, 1986
56. W.A. BAESLACK III, METALLOGRAPHY OF TITANIUM ALLOY WELDMENTS, *METALLOGRAPHY AND INTERPRETATION OF WELD MICROSTRUCTURES,* J.L. MCCALL, D.L. OLSON, AND I. LEMAY, ED., ASM INTERNATIONAL, 1987, P 23-60

Cracking Phenomena Associated With Welding

Michael J. Cieslak, Physical and Joining Metallurgy Department, Sandia National Laboratories

Introduction

THE FORMATION OF DEFECTS in materials that have been fusion welded is a major concern in the design of welded assemblies. Four types of defects in particular have been the focus of much attention because of the magnitude of their impact on product quality. These defects, all of which manifest themselves as cracks, are characteristic of phenomena that occur at certain temperature intervals specific to a given alloy. Colloquially, these four defect types are known as hot cracks, heat-affected zone (HAZ) microfissures, cold cracks, and lamellar tearing.

Acknowledgements

This work was performed at Sandia National Laboratories, supported by the U.S. Department of Energy under Contract No. DE-AC04-76DP00789.

Solidification Cracking (Hot Cracking)

Hot cracks are solidification cracks that occur in the fusion zone near the end of solidification. Simplistically, they result from the inability of the semisolid material to accommodate the thermal shrinkage strains associated with weld solidification and cooling. Cracks then form at susceptible sites to relieves the accumulating strain. Susceptible sites are interfaces, such as solidification grain boundaries and interdendritic regions, that are at least partially wetted.

Figure 1 is a photomicrograph of a metallographic section showing the appearance of a typical solidification crack in a fusion weld. Note the intergranular nature of the crack path. Figure 2 shows the surface morphology of a solidification crack, reflecting the nearly solidified nature of the interface at the time of failure. Note the smooth dendritic appearance of this surface.

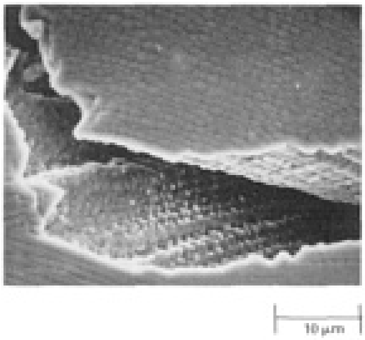

FIG. 1 SOLIDIFICATION CRACK IN A PULSED ND:YAG LASER WELD JOINING HASTELLOY C-276 TO 17-4 PH STAINLESS STEEL. YAG, YTTRIUM-ALUMINUM-GARNET

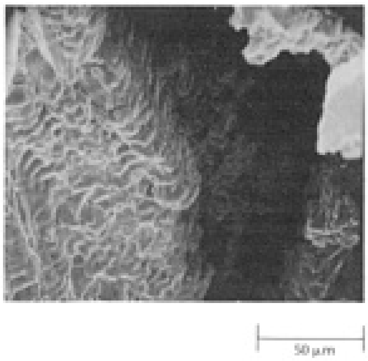

FIG. 2 SURFACE OF A SOLIDIFICATION CRACK IN AN ALLOY 214 VARESTRAINT TEST SPECIMEN

Solidification cracking requires both a sufficient amount of mechanical restraint (strain) and a susceptible microstructure. Knowledge regarding the development of shrinkage strains at the microstructural scale of a solidifying dendrite is still quite limited, making it nearly impossible to quantify *a priori* the amount of local strain required to initiate solidification cracks. Furthermore, the rate at which the strain is applied can influence the ability of an alloy to survive the solidification process without cracking. At temperatures near their melting points, metal alloys can readily dissipate applied loads by time-dependent deformation processes (i.e., creep processes). At high rates of applied stress, both the fusion zone and the HAZ have limited time to accommodate this mechanical imposition. Under conditions of rapid solidification and cooling, the rate of strain accumulation is rapid, leading to an increased cracking susceptibility. Inherently, then, requisite strains for solidification cracking are more likely to be experienced with welding processes that promote rapid solidification and cooling.

Because of the limited quantitative understanding of strain development in a solidifying weld, the practical approach taken to minimize the mechanical factor is to reduce the overall weld restraint through judicious joint design and appropriate choice of welding parameters. A simple way to minimize the restraint on a solidifying weld joint is to keep the joint gap to a minimum by designing hardware with good fit-up. Another approach, particularly attractive for small parts, is to design the weld joint as a standing edge. Welding parameters can have a profound influence on the occurrence of solidification cracking. The natural tendency to use high-speed welding to improve productivity can have detrimental effects. Formation of a teardrop-shaped weld pool, which may occur as the weld travel speed increases, can result in centerline solidification cracks. The solidification pattern associated with this type of weld pool is such that solidifying grains meet at the weld centerline, forming a particularly susceptible site for solidification crack initiation.

The manner in which a fusion weld is terminated can also influence cracking behavior. In a closure weld, simply extinguishing the power can result in the formation of "crater cracks" as the final weld pool rapidly solidifies in on itself. A common method used to minimize this problem is to ramp down the power slowly, either while the workpiece remains stationary or as it continues to travel a short length under the heat source. In some cases, the use of a runoff tab as a site for weld termination may be appropriate.

The extent to which special joint designs or precisely controlled welding parameters can be used to produce welds without solidification cracks is specific for a given alloy. Alloys that are microstructurally susceptible to cracking will necessarily have a more limited set of joint designs and processing parameters around which defect-free welds can be produced.

Alloys with a wide solidification temperature range are more susceptible to solidification cracking than alloys that solidify over a narrow temperature range. Simply described, this is because accumulated thermal strain is proportional to the temperature range over which a material solidifies. This temperature range is determined primarily by chemical composition. For many binary alloy systems and a few ternary alloy systems, the equilibrium solidification (or melting) temperature range has been established in phase diagram studies. For the vast majority of commercial alloys, however, these data are not precisely known.

Borland (Ref 1, 2) attempted to establish a general quantitative method for comparing the effects of various alloying elements on the melting temperature range of a given solvent element. Figure 3 shows a schematic binary alloy phase diagram having linear liquidus and solidus boundaries. For an alloy of composition C_0, T_L is the liquidus composition and T_S is the solidus composition, where the alloy would be completely solidified under conditions of equilibrium cooling. T_M is the invariant melting temperature of the solvent element. The slope of the liquidus is m_L, and the slope of the solidus is m_S. The equilibrium distribution coefficient, k, is defined as the ratio of the composition of the solid, C_S, to the composition of the liquid, C_L, at a given temperature (i.e., $k = C_S/C_L|_T$). For a linear liquidus and solidus, this value is constant and also equal to m_L/m_S.

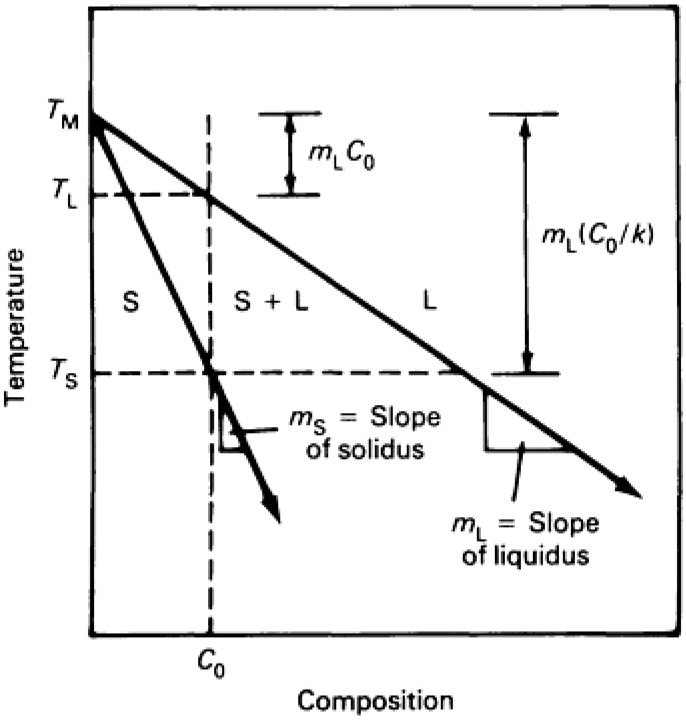

FIG. 3 HYPOTHETICAL BINARY ALLOY PHASE DIAGRAM SHOWING KEY PARAMETERS REQUIRED TO DERIVE THE RELATIVE POTENCY FACTOR (RPF)

With these terms thus defined, a derivation of the relative potency factor (RPF), after Borland (Ref 2), follows:

$$T_L = T_M - M_L C_0 \qquad \textbf{(EQ 1)}$$

$$T_S = T_M - M_S C_0 \qquad \textbf{(EQ 2)}$$

However, $m_S = m_L/k$:

$$T_s = T_M - \frac{M_L}{K} C_o \qquad \textbf{(EQ 3)}$$

Then:

$$T_L - T_S = \frac{m_L C_o (K-1)}{K} \qquad \textbf{(EQ 4)}$$

The RPF, always a positive number, is defined as the change in melting temperature ranger per weight percent of solute added, or:

$$RPF = \frac{T1 - Ts}{C_o} = \frac{M_L (k-1)}{K} \qquad \textbf{(EQ 5)}$$

where for $k < 1$, $m_L < 0$; and for $k > 1$,$m_L > 0$. The larger the RPF, the wider the melting temperature range and the higher the solidification cracking susceptibility. The final form of Eq 5 indicates that as k becomes very small (indicating a high level of microsegregation), the solidification cracking sensitivity increases. Similarly, as the magnitude of m_L increases (that is, becomes more highly negative), indicating a rapid decrease in liquidus with alloying concentration, the cracking propensity also increases.

Table 1 lists RPF values for various alloying elements in binary iron (Ref 2), nickel-base alloys (Ref 3), and aluminum alloys. It is clear that certain alloying elements, such as sulfur, phosphorus, and boron, are detrimental to the weldability of these alloy systems. This has been borne out by field experience.

TABLE 1 RELATIVE POTENCY FACTORS FOR SELECTED ALLOYING ELEMENTS IN BINARY IRON, NICKEL, AND ALUMINUM SYSTEMS

SYSTEM	**RPF FOR:**										
	Al	**B**	**C**	**Co**	**Mn**	**Nb**	**Ni**	**P**	**S**	**Si**	**Ti**
BINARY IRON	1.5	. . .	322	0.3	26.2	28.8	2.9	121	925	1.8	13.8
BINARY NICKEL	1.7	22700	330	. . .	0.7	. . .	. . .	556	244000	11	5.4

	Cu	**Li**	**Mg**	**Si**	**Zn**
BINARY ALUMINUM	16.6	9.0	8.0	42.4	2.3

Source: Ref 2, 3

Although the exact derivation of the RPF is specific to a very simple binary system solidifying under equilibrium conditions, the methodology provides general insight that can be applied to commercial alloys. The association of cracking susceptibility with solidification temperature range has implications for alloy design and primary melt processing. An extended melting/solidification temperature range can be the result of either intentional alloying or impurities present after refining.

An example (Ref 4) of the application of Borland's concepts to the solidification cracking propensity of a commercial alloy system involves alloy 625. This alloy is derived from the ternary Ni-Cr-Mo system, a very common base system for alloys designed to survive aggressive corrosive or high-temperature environments. A factorially designed set of alloys was produced to identify the effects of the alloying elements niobium, silicon, and carbon on the temperature range of two-phase coexistence of liquid and solid (equilibrium melting/solidification temperature range) and on solidification cracking susceptibility.

Solidification cracking susceptibility was measured with the varestraint test, using the maximum crack length (MCL) parameter as the quantitative measurement (Ref 5). The melting temperature ranges of these same alloys were measured using differential thermal analysis (DTA). Table 2 lists the compositions of the alloys examined.

TABLE 2 SELECTED ALLOY 625 COMPOSITIONS USED IN WELDABILITY STUDY

ALLOY NO.	COMPOSITION, %										
	C	Cr	Fe	Mn	Mo	Nb	Ni	P	S	Si	Ti
NIOBIUM-FREE											
1	0.006	22.10	2.56	0.02	9.54	0.01	BAL	0.005	0.002	0.03	0.06
2	0.031	21.95	2.55	0.02	9.61	0.01	BAL	0.005	0.002	0.03	0.06
3	0.006	21.63	2.18	0.02	9.60	0.02	BAL	0.005	0.002	0.35	0.06
4	0.036	21.57	2.59	0.03	9.63	0.02	BAL	0.005	0.003	0.39	0.06
NIOBIUM ALLOYED											
5	0.009	21.81	2.30	0.03	9.81	3.61	BAL	0.006	0.003	0.03	0.06
6	0.038	21.83	2.31	0.03	9.81	3.60	BAL	0.006	0.003	0.03	0.06
7	0.008	21.65	2.26	0.03	9.68	3.57	BAL	0.006	0.004	0.38	0.06
8	0.035	21.68	2.29	0.03	9.67	3.53	BAL	0.006	0.003	0.46	0.06

The results of the varestraint testing are shown in Fig. 4. These data clearly show that the alloys that contain intentionally alloyed niobium (alloys 5 to 8) are more susceptible to solidification cracking (larger MCL) than niobium-free alloys (1 to 4). Furthermore, the alloy with no intentionally alloyed carbon, silicon, or niobium (alloy 1) is the least susceptible to cracking.

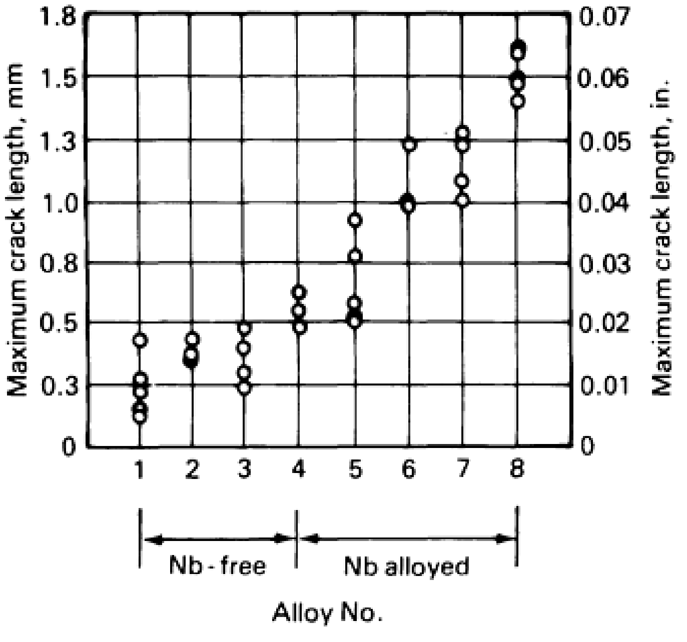

FIG. 4 MAXIMUM CRACK LENGTH DATA OBTAINED FROM VARESTRAINT TEST FOR ALLOY 625 WELDABILITY STUDY

The results of the DTA testing are given in Table 3. Listed are the liquidus (T_L), solidus (T_S), and melting temperature range (ΔT) data obtained from multiple tests. The baseline ΔT of an alloy without intentionally alloyed carbon, silicon, and niobium (alloy 1) is 26.5 °C (79.7 °F). Increases in Δ*T* are observed when carbon or silicon is added (alloys 2 and 3) at levels that would be considered high for commercial alloys. When niobium is added (alloy 5) at a level that would be considered nominal for alloy 625 (~3.6 wt%), the melting temperature range effectively doubles to 55.3 °C (131.5 °F). Additions of carbon and silicon to the high-niobium alloys (alloys 6 to 8) further increase this value.

TABLE 3 MELTING TEMPERATURE DATA FOR ALLOY 625 WELDABILITY STUDY

σ is the standard deviation.

ALLOY NO.	**ALLOY ADDITION**			**LIQUIDUS, $T_L \pm \sigma T_L$**				**SOLIDUS, $T_S \pm \sigma T_S$**				**MELTING TEMPERATURE RANGE, $\Delta T \pm \sigma \Delta T$**			
				T_L		**σT_L**		**T_S**		**σT_S**		**ΔT**		**$\sigma \Delta T$**	
	C	**NB**	**SI**	**°C**	**°F**	**°C**	**°F**	**°C**	**°F**	**°C**	**°F**	**°C**	**°F**	**°C**	**°F**
1				1406.5	2563.7	1.3	2.3	1380.0	2516.0	1.4	2.5	26.5	47.7	0.6	1.1
2	X			1403.0	2557.4	1.0	1.8	1369.3	249.7	2.5	4.5	33.7	60.7	1.5	2.7
3			X	1395.5	2543.9	1.3	2.3	1366.5	2491.7	1.3	2.3	29.0	52.2	1.6	2.9
4	X		X	1390.8	2535.4	0.5	0.9	1347.5	2457.5	1.7	3.1	43.3	77.9	1.3	2.3
5		X		1363.3	2485.9	0.5	0.9	1308.0	2386.4	2.2	4.0	55.3	99.5	2.4	4.3
6	X	X		1362.0	2483.6	0.8	1.4	1289.3	2352.7	2.6	4.7	72.8	131.0	3.2	5.8
7		X	X	1356.0	2472.8	0.8	1.4	1287.8	2350.0	2.4	4.3	68.3	122.9	2.5	4.5
8	X	X	X	1352.0	2465.6	0.0	0.0	1275.7	2328.3	0.6	1.1	76.3	137.3	0.6	1.1

The correlation of cracking susceptibility with melting temperature range is shown in Fig. 5. Linear regression analysis performed on the data set resulted in the following functional relationship:

$$\mathrm{MCL} = 0.00087(\Delta T) - 0.015 \qquad \textbf{(EQ 6)}$$

The incomplete statistical correlation ($R^2 = 0.86$) suggests that other factors are necessarily involved.

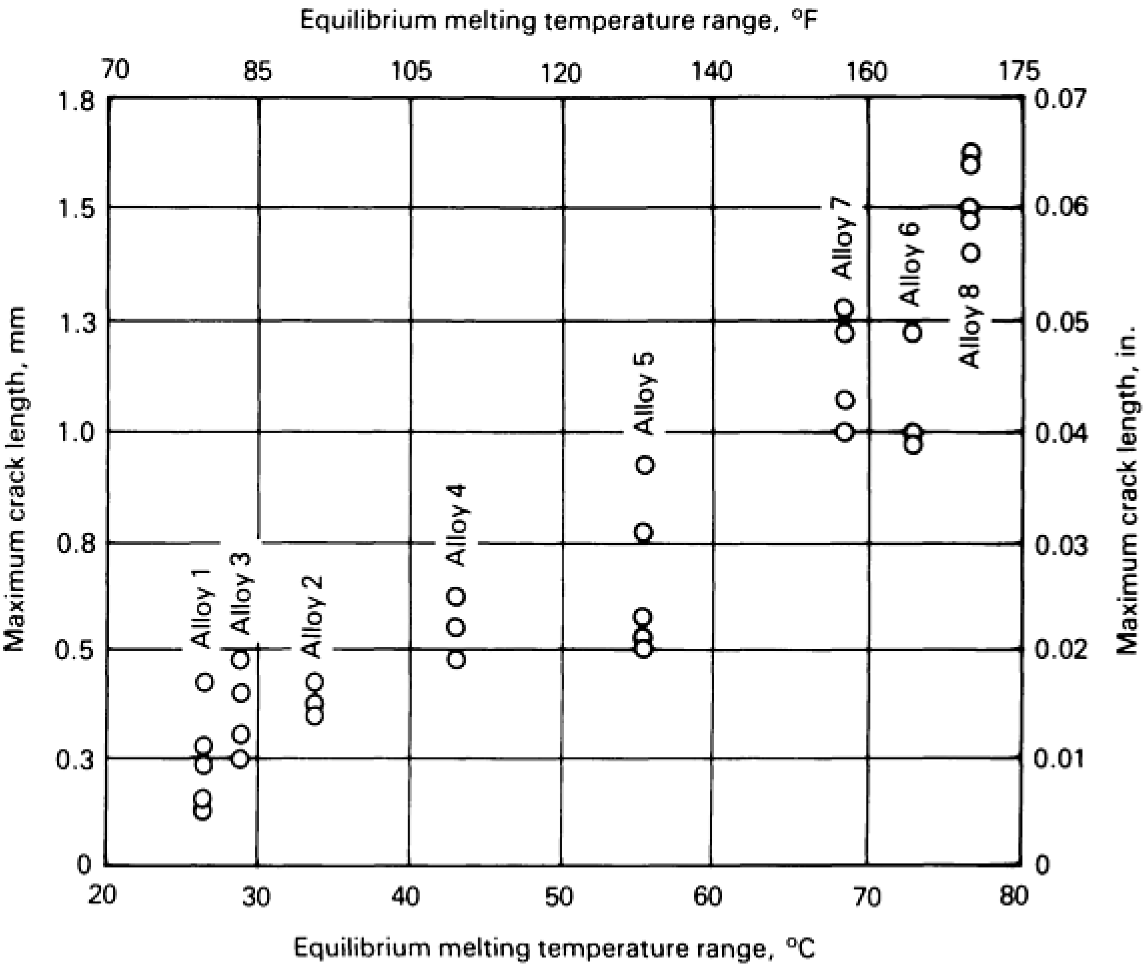

FIG. 5 RELATIONSHIP BETWEEN MAXIMUM CRACK LENGTH AND EQUILIBRIUM MELTING TEMPERATURE RANGE IN ALLOY 625 WELDABILITY STUDY

The solidification behavior of fusion welds is nonequilibrium in nature and may result in an extension of the solidification temperature range significantly beyond that measured under "equilibrium" conditions. Indeed, the niobium-bearing alloys terminate weld solidification below the equilibrium solidus with the formation of nonequilibrium eutecticlike constituents involving niobium carbides and Laves-type phases. Nonetheless, it is clear from an engineering perspective that Borland's concepts are valid enough to allow the establishment of a useful alloy ranking system from a solidification cracking perspective.

In addition to the effect of melting temperature range. Borland realized that the distribution of liquid along solidifying grain boundaries plays an important role in determining the solidification cracking susceptibility on an alloy (Ref 1). If the terminal solidification liquid easily wets the solidifying grain boundary (that is, has a low solid/liquid surface tension), it will tend to be spread out over a wider area, easing separation of the grains under a mechanical imposition. Conversely, if the liquid does not readily wet these boundaries, it will be more localized, creating a smaller area of liquid-solid contact and more solid-solid bridging, resulting in a more robust mechanical structure with increased cracking resistance. Elements such as sulfur and boron have been suggested as providing this surfactant property to terminal solidification liquids in iron and nickel alloys. Manganese may serve to increase the surface tension of terminal solidification fluids.

Figure 6 shows the results (Ref 6) of varestraint testing on two heats of Cabot alloy 214, a Ni-16Cr-4.5Al-2.5Fe alloy designed for use in high-temperature environments. The two alloys differ primarily in their boron content. With no boron intentionally added (0.0002 wt%), the alloy is extremely resistant to solidification cracking. With boron intentionally added at a level of 0.003 wt%, the solidification cracking propensity is much greater. Figures 7 and 8 are surface images of pulsed Nd:YAG laser welds made on the two alloys. Extensive cracking is evident in the high-boron alloy. Figure 7(b) is a good example of the extensive terminal "crater cracking" discussed above.

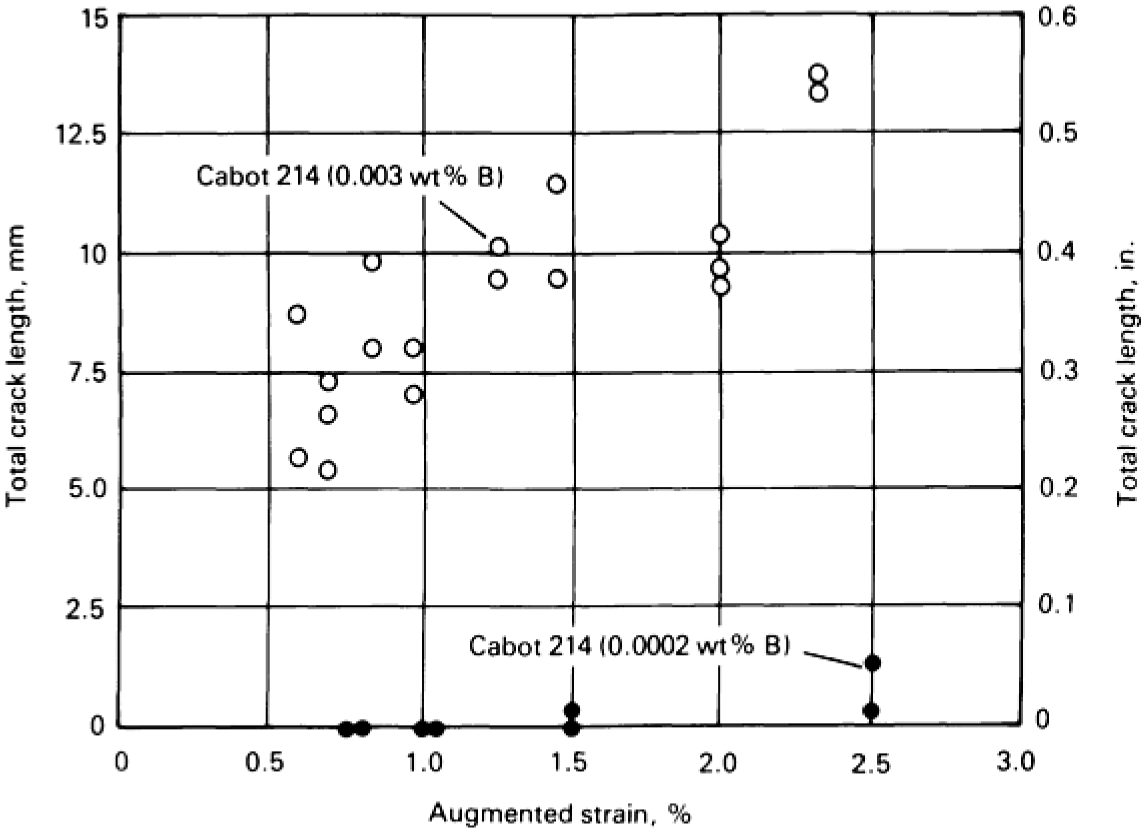

FIG. 6 VARESTRAINT TEST WELDABILITY DATA FOR CABOT ALLOY 214 WITH VARYING BORON CONCENTRATIONS

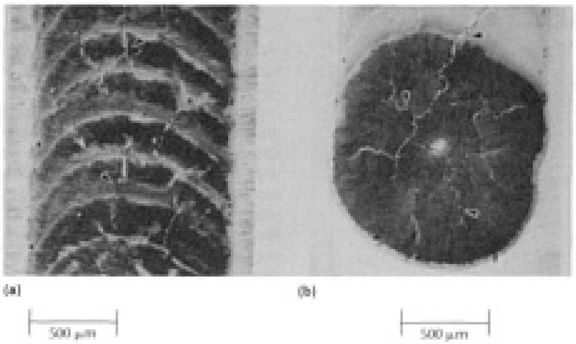
(a)
500 μm
(b)
500 μm

FIG. 7 SEM SURFACE IMAGE OF PULSED ND:YAG LASER WELDS IN HIGH-BORON ALLOY 214. (A) SOLIDIFICATION CRACKING. (B) CRATER CRACKING

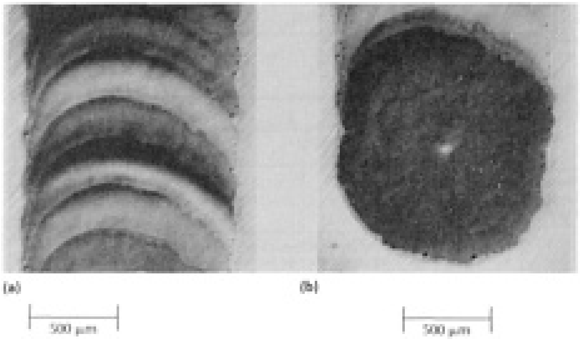
(a)
500 μm
(b)
500 μm

FIG. 8 SEM SURFACE IMAGE OF PULSED ND:YAG LASER WELDS IN LOW-BORON CONTENT ALLOY 214. (A) SOLIDIFICATION CRACKING. (B) CRATER CRACKING

Figures 9 and 10 summarize parameters that affect hot cracking in the weld metal and in the base metal HAZ, respectively. For most alloys that are highly susceptible to solidification cracking, filler metals of different composition have been developed to minimize the cracking that accompanies fusion welding.

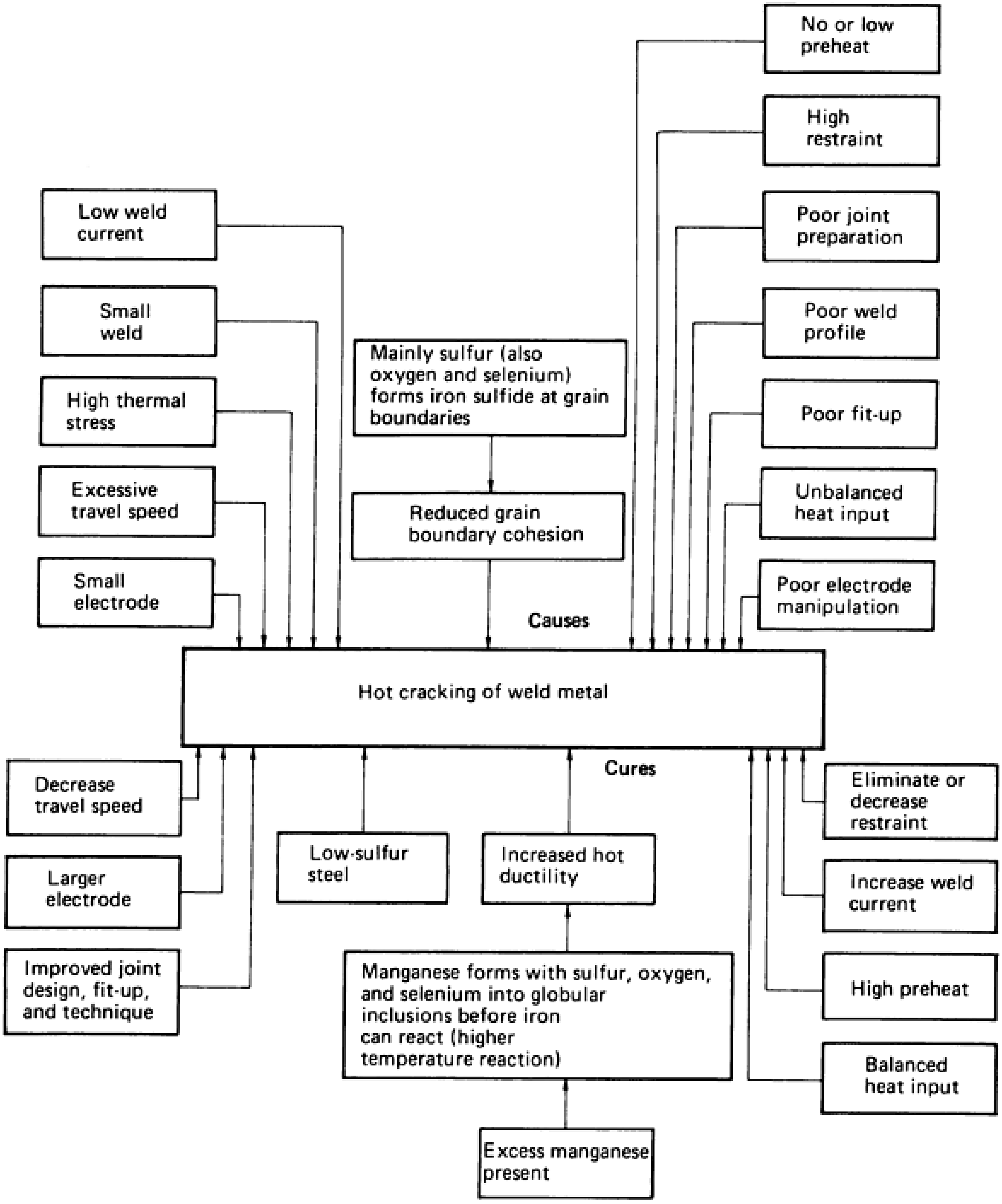
No or low preheat
High restraint
Poor joint preparation
Poor weld profile
Poor fit-up
Unbalanced heat input
Poor electrode manipulation
Low weld current
Small weld
High thermal stress
Excessive travel speed
Small electrode
Mainly sulfur (also oxygen and selenium) forms iron sulfide at grain boundaries
Reduced grain boundary cohesion
Causes
Hot cracking of weld metal
Cures
Decrease travel speed
Larger electrode
Improved joint design, fit-up, and technique
Low-sulfur steel
Increased hot ductility
Manganese forms with sulfur, oxygen, and selenium into globular inclusions before iron can react (higher temperature reaction)
Excess manganese present
Eliminate or decrease restraint
Increase weld current
High preheat
Balanced heat input

FIG. 9 FACTORS AFFECTING HOT CRACKING IN WELD METAL

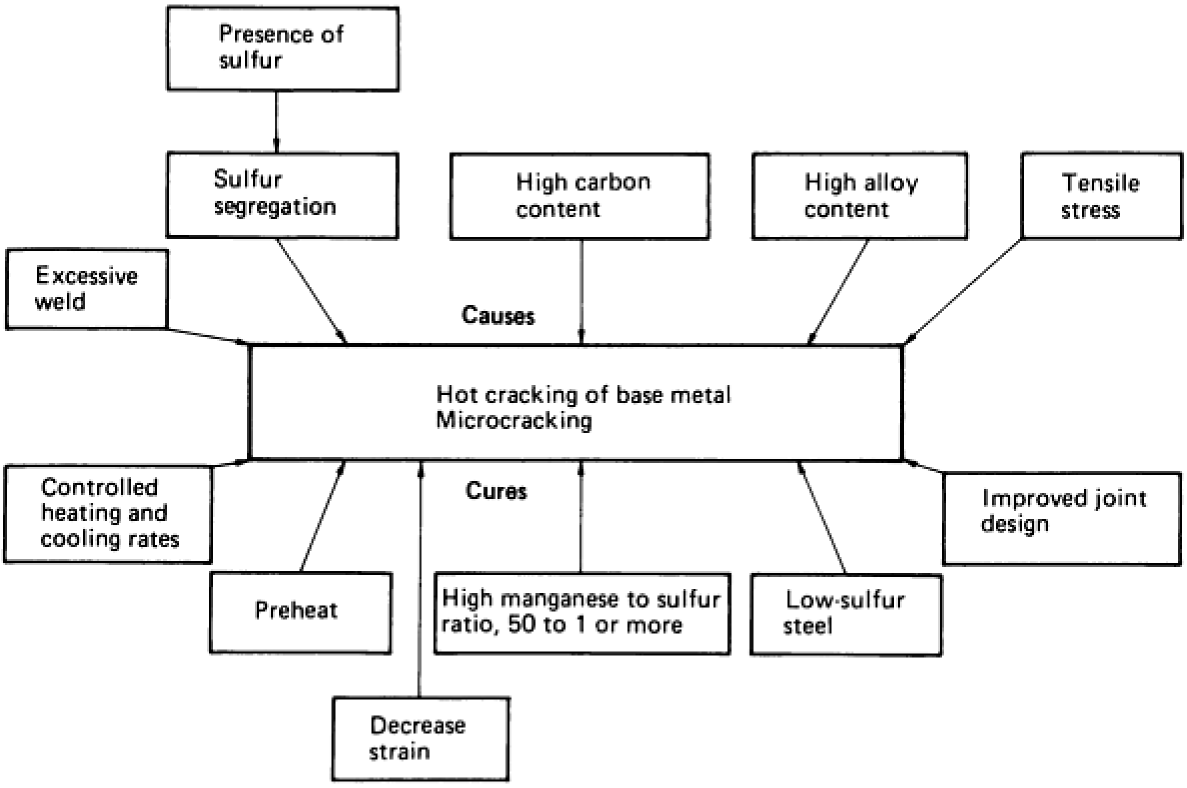
Presence of sulfur
Sulfur segregation
High carbon content
High alloy content
Tensile stress
Excessive weld
Causes
Hot cracking of base metal Microcracking
Cures
Controlled heating and cooling rates
Improved joint design
Preheat
High manganese to sulfur ratio, 50 to 1 or more
Low-sulfur steel
Decrease strain

FIG. 10 FACTORS AFFECTING HOT CRACKING IN BASE METAL HAZ

References cited in this section

1. J.C. BORLAND, GENERALIZED THEORY OF SUPER-SOLIDUS CRACKING IN WELDS (AND CASTINGS), *BR. WELD. J.,* VOL 7 (NO. 8), 1960, P 508-512
2. J.C. BORLAND, HOT CRACKING IN WELDS, *BR. WELD. J.,* VOL 7 (NO. 9), 1960, P 558-559
3. D.A. CANONICO, W.F. SAVAGE, W.J. WERNER, AND G.M. GOODWIN, EFFECTS OF MINOR ADDITIONS ON THE WELDABILITY OF INCOLOY 800, *EFFECTS OF MINOR ELEMENTS ON THE WELDABILITY OF HIGH-NICKEL ALLOYS,* WRC, JULY 1969, P 68-92
4. M.J. CIESLAK, THE WELDING AND SOLIDIFICATION METALLURGY OF ALLOY 625, *WELD. J.,* FEB 1991, P 49S-56S
5. W.F. SAVAGE AND C.D. LUNDIN, APPLICATION OF THE VARESTRAINT TECHNIQUE TO THE STUDY OF WELDABILITY, *WELD. J.,* NOV 1966, P 497S-503S
6. M.J. CIESLAK, J.J. STEPHENS, AND M.J. CARR, A STUDY OF THE WELDABILITY AND WELD RELATED MICROSTRUCTURE OF CABOT ALLOY 214, *METALL. TRANS. A,* MARCH 1988, P 657-667

Cracking Phenomena Associated With Welding

Michael J. Cieslak, Physical and Joining Metallurgy Department, Sandia National Laboratories

---

Heat-Affected-Zone Cracks

Microfissures are cracks that occur in the area of partial melting and the HAZ adjacent to the fusion line. Because no material can be purified to the state where it solidifies truly as an invariant, all materials have a temperature region of stable two-phase coexistence of solid and liquid. In fusion welding, this manifests itself in the formation of a zone of partial melting at temperatures below the alloy liquidus. The extent of this zone may be enlarged by the presence of chemical inhomogeneity in a material. Local chemical variations will result in local variations in the melting point. Segregation of specific alloying elements to grain boundaries may cause a reduction in the melting temperatures of these areas.

Many microfissures form in a manner somewhat analogous to the formation of solidification cracks in that susceptible sites, generally grain boundaries that intersect the fusion zone, are wetted with liquid from one of various sources. shrinkage strains accumulating as the weld pool advances past the liquated boundary can develop to a level sufficient to cause boundary separation (i.e., cracking). Figure 11 is a photomicrograph of an HAZ microfissure in a cast austenitic stainless steel (this crack also extends into the fusion zone as a solidification crack). Cast microstructures are highly inhomogeneous, raising the possibility of extensive partial melting. Note the melting associated with interdendritic constituents in Fig. 11.

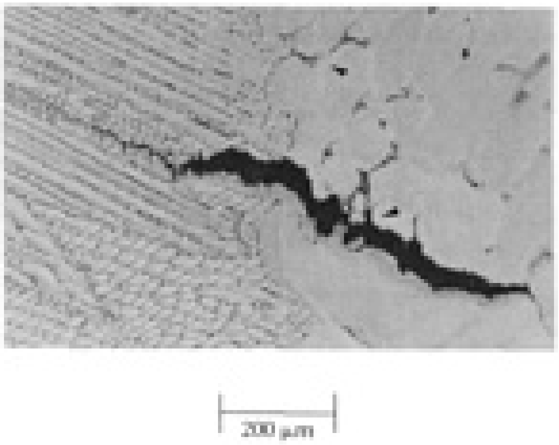

FIG. 11 HAZ MICROFISSURE IN A CAST STAINLESS STEEL. NOTE EXTENSIVE REGION OF PARTIAL MELTING.

Castings represent a special case of inhomogeneity on a continuous scale. Chemical inhomogeneities in wrought products are more discrete, generally involving minor phase particles. In many cases, these minor phases arise from thermomechanical processing of alloys well below the solidus, where the solubility for specific alloying elements can be lower. Because of the potential for extremely rapid rates of heating and cooling (especially heating) in the HAZ of fusion welds, it is entirely possible that minor phases not thermodynamically stable under equilibrium conditions may continue to exist at near-solidus temperatures. Under these conditions, the minor phases may interact with the matrix to form a liquid product. This reaction, often referred to as constitutional liquation (Ref 7), can provide the liquid required to wet grain boundaries producing a microstructure that may be susceptible to cracking.

Hot ductility testing is a common method of assessing the susceptibility of an alloy to this form of cracking. In this test, a laboratory-scale specimen is subjected to a thermal cycle representative of what the material would experience in the HAZ during fusion welding. At various temperatures during both the heating and cooling portions of the cycle, samples are strained to failure. The ductility response (usually expressed in terms of reduction of area) as the alloy progresses through the thermal cycle simulates the response of the alloy to an actual welding situation. In particular, the ability of the material to reestablish ductility after experiencing the peak temperature is a critical measure of sensitivity to this form of cracking.

Figures 12(a) and 12(b) show the hot ductility response of two heats of Cabot alloy 214 containing two different concentrations of boron. The return of ductility in the low-boron alloy after cooling from a peak temperature of 1345 °C (2453 °F) indicates that this alloy should be virtually immune to this form of cracking, as weldability tests confirmed (Ref 6). The high-boron heat showed no appreciable recovery of ductility until a temperature approximately 250 °C (450 °F) below the thermal peak temperature. This alloy should be quite susceptible to this form of cracking, as weldability tests also confirmed (Ref 6). Figure 13 is a fractograph of a hot ductility test specimen from the high-boron alloy taken from the sample strained at 1150 °C (2102 °F) on cooling from the peak temperature. The smooth, intergranular appearance is characteristic of this type of cracking. Although the presence of a liquid phase along this fracture surface could not be unequivocally established by fractographic techniques, it was noted that the failure occurred at an exceedingly low load (Ref 6), consistent with the presence of a liquid film.

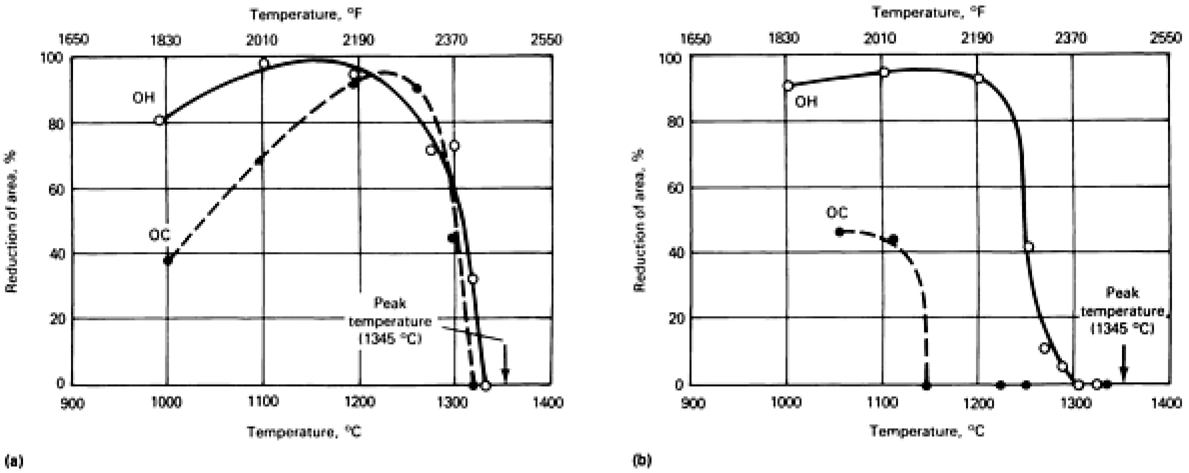

FIG. 12 HOT DUCTILITY RESPONSE OF CABOT ALLOY 214 WITH VARYING BORON CONCENTRATIONS. (A) LOW-BORON CONTENT (0.0002 WT% B). (B) HIGH-BORON CONTENT (0.003 WT% B). (OH, TESTING DONE ON HEATING; OC, TESTING DONE ON COOLING FROM 1345 °C (2455 °F)

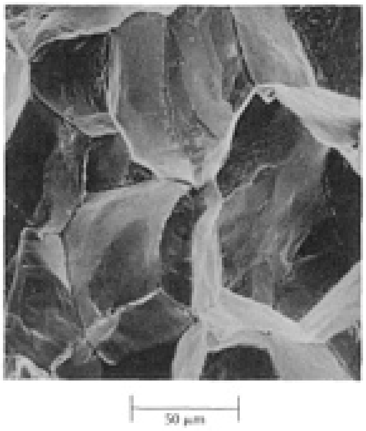

FIG. 13 FRACTURE SURFACE OF A HOT DUCTILITY SAMPLE OF HIGH-BORON ALLOY 214 TESTED AT 1150 °C (2100 °F) ON COOLING FROM 1345 °C (2455 °F)

The presence of a liquid is not essential to the formation of HAZ cracks, as they may occur at temperatures well below the solidus. Poor ductility inherent in certain materials, such as intermetallics, can lead to solid-state failure during welding if the thermal stresses associated with the weld thermal cycle exceed the local tensile strength.

Cracks may also occur in alloys that undergo polymorphic transformations during the weld thermal cycle. The act of fusion welding subjects the fusion zone and the HAZ to temperatures above those normally encountered during prior deformation processing. The resultant HAZ microstructure may be substantially different from that of the base material. In the case of martensitic-type transformations, there is often a relatively large, discontinuous change in the molar volume occurring very rapidly with decreasing temperature below the M_s. If that event causes a local shrinkage, then large, localized tensile stresses can develop. If the newly transformations microstructure has limited ductility (not uncommon in shear transformations or in transformations that result in ordered structures), the likelihood of cracks forming at susceptible sites, such as grain boundaries, matrix/minor phase interfaces, and slip band intersections, is increased.

Reheat cracking, also referred to as stress-relief of strain-age cracking, is another defect type observed in certain alloys that undergo precipitation reactions. Historically observed in several nickel-base superalloys and creep-resistant steels, this type of defect manifests itself during postweld heat treatment, generally as intergranular HAZ cracks.

Postweld heat treatment is often a recommended practice for high-strength thick-section welded steels, where reduction of residual stresses developed as a result of welding is desired. In the case of welded nickel-base superalloys, postweld heat treatment is used both to relieve residual stresses and to achieve optimum mechanical properties through precipitation-hardening reactions.

In alloys that undergo precipitation reactions, the rate at which the alloy strengthens may greatly exceed the rate at which residual stresses are thermally diminished. This can be especially true in the case of heavy-section welds or in alloys with relatively poor thermal conductivity. (That is, before the alloy can reach the temperature at which residual stresses begin to be eliminated, it has aged and lost ductility relative to its as-welded state.) At this new lower level of ductility, the residual stresses present may induce cracks to eliminate the accumulated strain energy.

In principle, the problem of reheat cracking could be largely eliminated if the rate of heating through the precipitation temperature range was rapid enough to prevent precipitate formation. However, extremely rapid rates of heating in welded hardware would likely lead to problems such as excessive distortion; therefore, improved alloys (e.g., alloy 718 among the nickel-base superalloys) have been developed in which the kinetics of precipitate formation have been sufficiently retarded to allow for successful postweld heat treatment.

Dealing with HAZ cracks, however they might form, is generally more difficult than attending to the problem of solidification cracking. Modifying the composition in the HAZ through the addition of consumables is not a viable option. Prior heat treatment may allow the production of a less susceptible starting microstructure, but this option generally is quite limited. Optimization of weld process parameters to minimize the extent of tensile stress and strain development should always be considered, as should optimizing weld joint design as discussed above.

References cited in this section

6. M.J. CIESLAK, J.J. STEPHENS, AND M.J. CARR, A STUDY OF THE WELDABILITY AND WELD RELATED MICROSTRUCTURE OF CABOT ALLOY 214, *METALL. TRANS. A,* MARCH 1988, P 657-667
7. J.J. PEPE AND W.F. SAVAGE, EFFECTS OF CONSTITUTIONAL LIQUATION IN 18-NI MARAGING STEEL WELDMENTS, *WELD. J.,* SEPT 1967, P 411S-422S

Hydrogen-Induced Cracking (Cold Cracking)

Cold cracks are defects that form as the result of the contamination of the weld microstructure by hydrogen. Whereas solidification cracking and HAZ cracking occur during or soon after the actual welding process, hydrogen-induced cracking is usually a delayed phenomenon, occurring possibly weeks or even months after the welding operation. The temperature at which these defects tend to form ranges from -50 to 150 °C (-60 to 300 °F) in steels. The fracture is either intergranular or transgranular cleavage.

As with other forms of cracking, hydrogen-induced cracking involves both a requisite microstructure and a threshold level of stress. It also involves a critical level of hydrogen, which is alloy and microstructure dependent.

In the case of ideal weld processing, hydrogen-induced cracking would be at most a minor welding engineering concern. However, excluding hydrogen from structures during welding is exceedingly difficult. Although the primary source of hydrogen in weld metal is considered to be the disassociation of water vapor in the arc and absorption of gaseous or ionized hydrogen into the liquid, other sources are also available. All organic compounds contain hydrogen in their molecular structure, and all may be broken down in the intense thermal environment of a welding heat source. Organic compounds are ubiquitous in the welding environment, from lubricants in assembly areas to body oils on the hands of welding operators. Plated hardware may also contain high levels of residual hydrogen.

The mechanism of hydrogen-induced crack formation is still being investigated. An early hypothesis, involving the buildup of hydrogen gas pressure in voids, is now generally discredited. Currently, the most widely accepted model involves the presence of preexisting defect sites in the material--small cracks or discontinuities caused by minor phase particles or inclusions. In the presence of existing stress, these sites may develop high local areas of biaxial or triaxial tensile stress. Hydrogen diffuses preferentially to these sites of dilated lattice structure. As the local hydrogen concentration increases, the cohesive energy and stress of the lattice decrease. When the cohesive stress falls below the local intensified stress level, fracture occurs spontaneously. Hydrogen then evolves in the crack volume, and the process is repeated. This model of hydrogen-induced cracking is consistent with the relatively slow and discontinuous nature of the process.

In steels, where the problem of hydrogen-induced cracking is extremely significant, cracking susceptibility has been correlated both with material hardness and strength, and with specific microstructure. Higher-strength steels are more susceptible to hydrogen-induced cracking than low-strength steels. Steels that transform martensitically are particularly susceptible, especially the higher-carbon alloys with twinned martensitic structures. The desire to avoid martensite formation has driven the development of high-strength structural steels for welded applications.

Production of the newer high-strength low-alloy (HSLA) steels uses a variety of precisely controlled alloying additions (for example, aluminum, titanium, vanadium, and niobium) along with meticulous thermomechanical processing to develop a very fine-grained ferrite microstructure possessing substantial strength and fracture toughness with a high degree of resistance to hydrogen-induced cracking. The form of the ferrite produced during transformation on cooling from the austenite phase field is also critical, with acicular ferrite resulting in improved properties compared with grain-boundary, polygonal, or Widmanstätten ferrite. Acicular ferrite is often nucleated on minor phase particles, such as specific oxides or borides. Acknowledgement of the beneficial effect of certain oxides in providing sites for acicular ferrite nucleation in modern steels is in sharp contrast to steel design of earlier decades, when elimination of oxygen to the greatest extent possible was considered essential to the development of optimal fracture behavior.

A useful concept for understanding the susceptibility of carbon and alloy steels to hydrogen-induced cracking is the carbon equivalent (CE), an empirical relationship that attempts to reduce the number of significant compositional variables affecting the weldability of steels into a single quantity. Several carbon equivalent relationships have been developed for different classes of steels. An example is:

$$CE = \%C + \frac{\%Mn}{6} + \frac{\%Cr + \%Mo + \%V}{5} + \frac{\%Si + \%Ni + \%Cu}{15} \qquad \textbf{(EQ 7)}$$

From a metallurgical perspective, the carbon equivalent can be related to the development of hydrogen-sensitive microstructures. That is, as the carbon equivalent increases, microstructures are evolved during cooling through the transformation temperature range that are increasingly more susceptible to hydrogen-induced cracking. At high carbon equivalent values, martensitic structures can be expected.

Carbon equivalent formulas are usually developed from large databases of critical hydrogen concentrations and weld joint restraints that will result in hydrogen-induced cracking in the steels under consideration. At various levels of carbon equivalent, certain weld preheat requirements are often established. In the example of Eq 7, when the carbon equivalent exceeds 0.35%, preheats are recommended to minimize susceptibility to hydrogen cracking. At higher levels of carbon equivalent, both preheats and postheats may be required.

Most structural carbon and low-alloy steels that may be susceptible to hydrogen-induced cracking transform from austenite during cooling through the temperature range of 800 to 500 °C (1470 to 930 °F). The length of time a steel spends in this temperature range during cooling will establish its microstructure and hence its cracking sensitivity. This time segment is generally referred to as $t_{8/5}$ (in seconds). To maximize cracking resistance, a microstructure free of untempered martensite is desired.

Using a specifically developed carbon equivalent, Yurioka (Ref 8) formulated a relationship to establish the critical $t_{8/5}$ for a martensite-free HAZ in low-carbon alloy steels. The carbon equivalent is defined as:

$$CE = \%C^{*} + \frac{\%Mn}{3.6} + \frac{\%Cu}{20} + \frac{\%Ni}{9} + \frac{\%Cr}{5} + \frac{\%Mo}{4} \qquad \textbf{(EQ 8)}$$

where %C* = %C for C $\leq$0.3% and %C* = %C/6 + 0.25 for %C > 0.3%. The critical time length in seconds, $t_{8/5}$, for the avoidance of martensitic transformation is given as:

$$\text{LOG } t_{8/5} = 2.69 \text{ CE}^{*} + 0.321 \qquad \textbf{(EQ 9)}$$

When this quantity is known, welding parameters and preheat temperature for the given thickness of material being welded can be established to produce cooling rates that avoid formation of a martensitic constituent in the HAZ.

The use of preheating and postheating to minimize the susceptibility to hydrogen-induced cracking is an accepted welding procedure for many steels. Preheating controls the cooling rate through the transformation temperature range; a higher base metal temperature in a slower cooling rate through the transformation temperature range and thus a larger $\tau_{8/5}$. Thicker sections often require preheating because of the greater heat-sinking capability of a thick section, for which a given set of welding parameters will produce a faster cooling rate and a smaller $t_{8/5}$. Preheating may also reduce the level of residual stress in the welded assembly.

There are, of course, practical limits to the use of preheating. Excessively high base metal temperatures can make welding difficult, especially manual welding. A high preheat temperature will result in the flattening of temperature gradients, increasing the spatial extent of and the time spent in the austenite phase field. This may result in excessive grain growth in the HAZ, which can lead to a loss of fracture toughness independent of hydrogen content. Preheat temperatures up to 150 °C (300 °F) are not uncommon for low-alloy steels and may increase up to 425 °C (800 °F) for higher-carbon equivalent martensitic steels.

Several benefits can be realized by postheating. At high enough temperatures (550 to 600 °C, or 1020 to 1110 °F), postheating can diminish residual stresses present in the as-welded structure. Microstructural modifications such as tempering can also occur, producing a microstructure less susceptible to hydrogen-induced cracking. Finally, the higher temperature may allow hydrogen to diffuse into the bulk of the structure or even to the surface, where it may recombine and exit as a gas. In either case, the level of hydrogen concentration may be reduced below that required to initiate cracks.

The following guidelines should be followed to minimize the occurrence of hydrogen-induced cracking. For a given level of required strength, the steel with the lowest carbon equivalent should be considered. Low-hydrogen welding practice should be followed. This involves elimination of possible sources of hydrogen by using ultrahigh-purity gases and moisture-free gas lines and by baking coated electrodes following the manufacturer's recommendations to ensure removal of nascent water. Finally, preheating and postheating requirements should be followed assiduously. Figures 14 and 15 summarize the causes and cures of hydrogen-induced cracking in the weld metal and in the base metal, respectively.

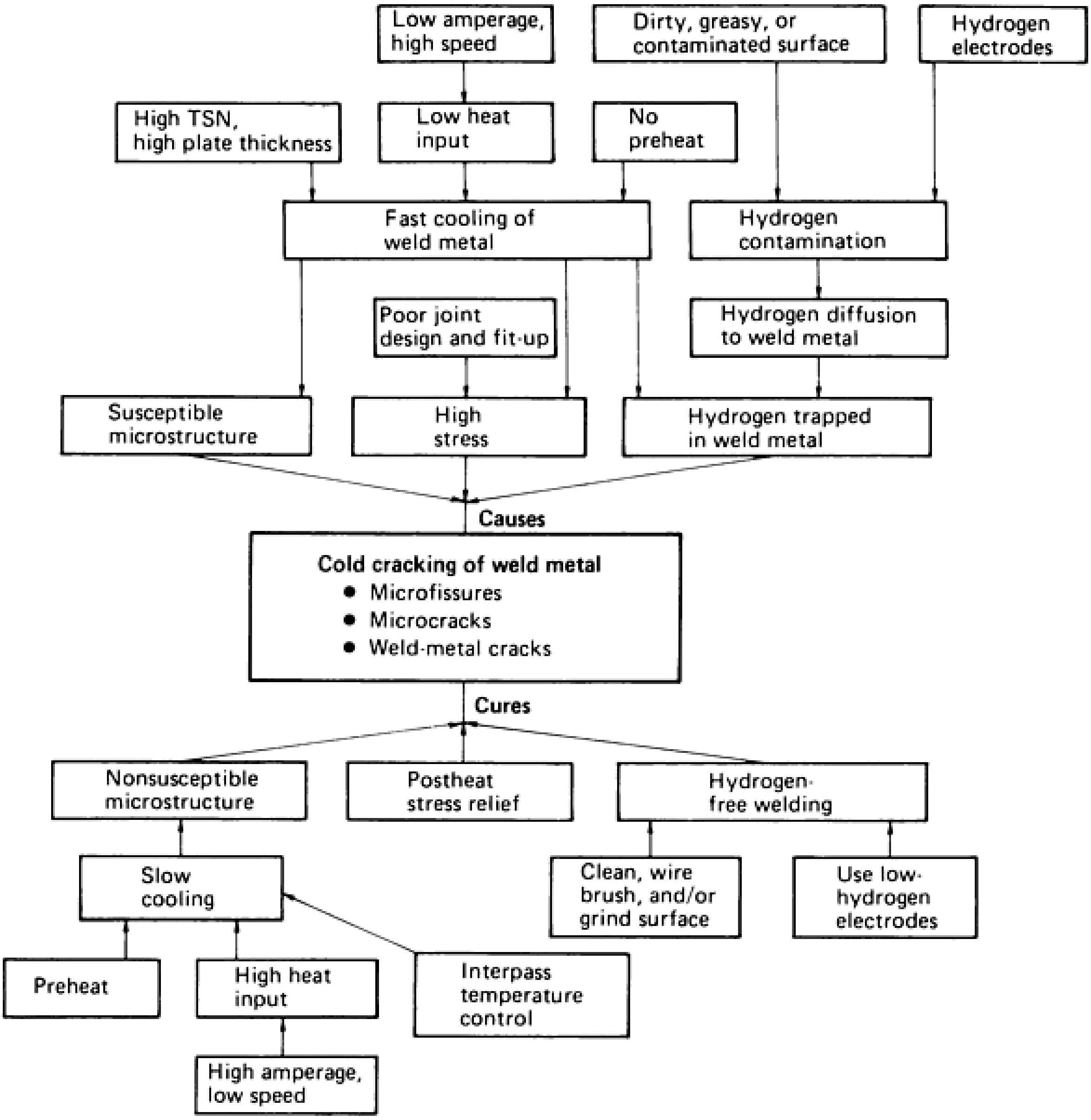

Low amperage, high speed
Dirty, greasy, or contaminated surface
Hydrogen electrodes
High TSN, high plate thickness
Low heat input
No preheat
Fast cooling of weld metal
Hydrogen contamination
Poor joint design and fit-up
Hydrogen diffusion to weld metal
Susceptible microstructure
High stress
Hydrogen trapped in weld metal
Causes
Cold cracking of weld metal
● Microfissures
● Microcracks
● Weld-metal cracks
Cures
Nonsusceptible microstructure
Postheat stress relief
Hydrogen-free welding
Slow cooling
Clean, wire brush, and/or grind surface
Use low-hydrogen electrodes
Preheat
High heat input
Interpass temperature control
High amperage, low speed

FIG. 14 CAUSES AND CURES OF HYDROGEN-INDUCED CRACKING IN WELD METAL. THERMAL SEVERITY NUMBER (TSN), WHICH IS FOUR TIMES THE TOTAL PLATE THICKNESS CAPABLE OF REMOVING HEAT FROM THE JOINT, IS A MEASURE OF THE ABILITY OF THE MEMBER TO SERVE AS A HEAT SINK.

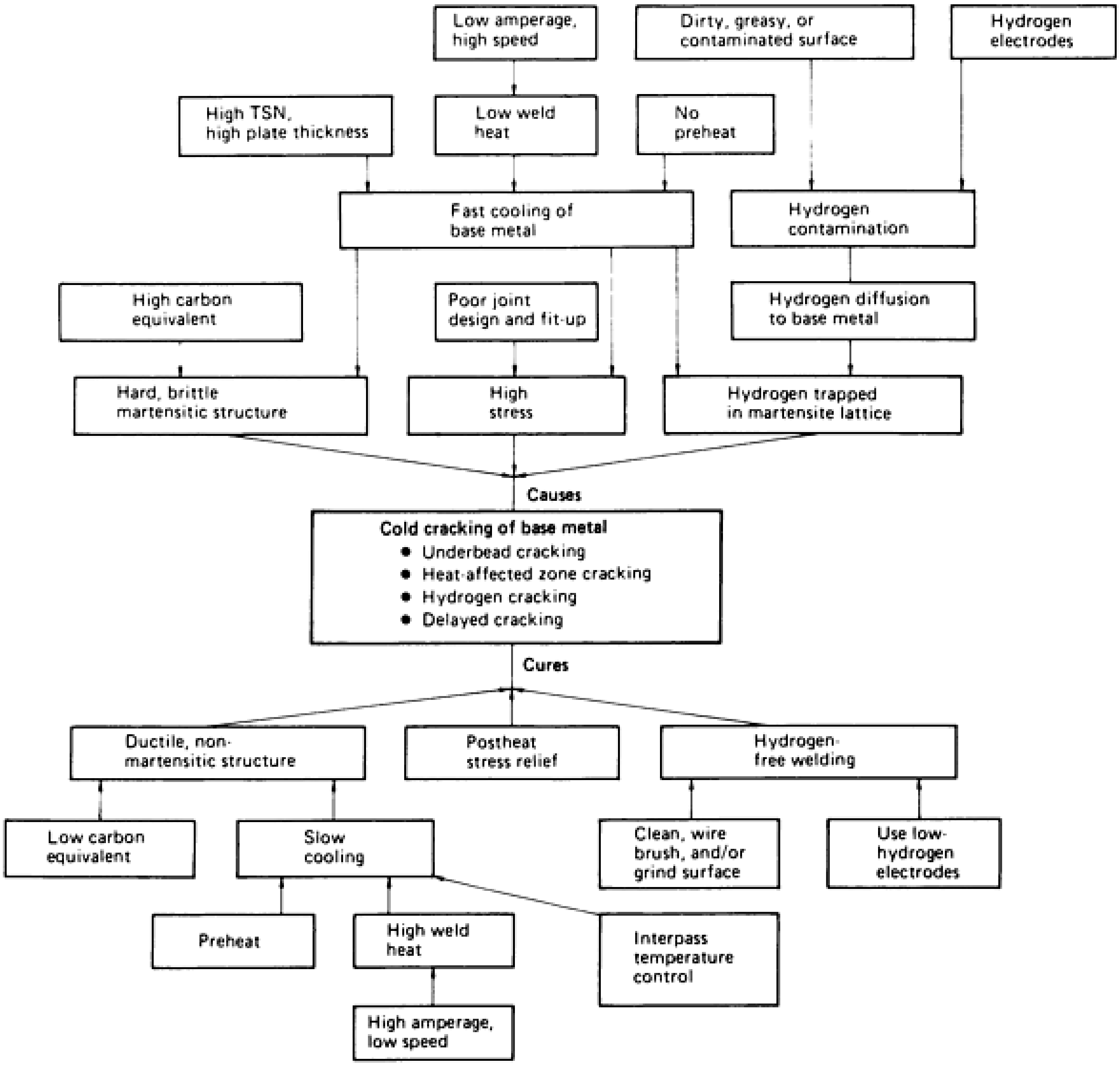
Low amperage, high speed
Dirty, greasy, or contaminated surface
Hydrogen electrodes
High TSN, high plate thickness
Low weld heat
No preheat
Fast cooling of base metal
Hydrogen contamination
High carbon equivalent
Poor joint design and fit-up
Hydrogen diffusion to base metal
Hard, brittle martensitic structure
High stress
Hydrogen trapped in martensite lattice
Causes
Cold cracking of base metal
● Underbead cracking
● Heat-affected zone cracking
● Hydrogen cracking
● Delayed cracking
Cures
Ductile, non-martensitic structure
Postheat stress relief
Hydrogen-free welding
Low carbon equivalent
Slow cooling
Clean, wire brush, and/or grind surface
Use low-hydrogen electrodes
Preheat
High weld heat
Interpass temperature control
High amperage, low speed

FIG. 15 CAUSES AND CURES OF HYDROGEN-INDUCED CRACKING IN BASE METAL

Reference cited in this section

8. N. YURIOKA, WELDABILITY OF MODERN HIGH STRENGTH STEELS, *1ST U.S.-JAPAN SYMP. ADVANCES IN WELDING METALLURGY,* AWS, 1990, P 79-100

Cracking Phenomena Associated With Welding

Michael J. Cieslak, Physical and Joining Metallurgy Department, Sandia National Laboratories

---

Lamellar Tearing (Ref 9)

Lamellar tearing is cracking that occurs beneath welds. It is found in rolled steel plate weldments. The tearing always lies within the base metal, generally outside the HAZ and parallel to the weld fusion boundary. The problem is caused by welds that subject the base metal to tensile loads in the *z*, or through, direction of the rolled steel. Occasionally the tearing comes to the surface of the metal, but more commonly remains under the weld (Fig. 16) and is detectable only by ultrasonic testing. Lamellar tearing occurs when three conditions are simultaneously present:

- STRAINS DEVELOP IN THE THROUGH DIRECTION OF THE PLATE. THEY ARE CAUSED BY WELD METAL SHRINKAGE IN THE JOINT AND CAN BE INCREASED BY RESIDUAL STRESSES AND BY LOADING.
- THE WELD ORIENTATION IS SUCH THAT THE STRESS ACTS THROUGH THE JOINT ACROSS THE PLATE THICKNESS (THE *Z* DIRECTION). THE FUSION LINE BENEATH THE WELD IS ROUGHLY PARALLEL TO THE LAMELLAR SEPARATION.
- THE MATERIAL HAS POOR DUCTILITY IN THE *Z* DIRECTION.

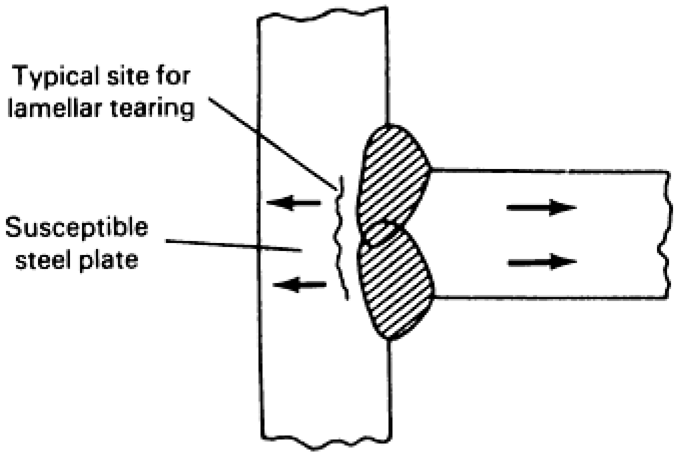

FIG. 16 TYPICAL LOCATION FOR LAMELLAR TEARING IN A T-JOINT

Lamellar tearing can occur during flame cutting and cold shearing operations. Low material strength in the *z* direction is the primary cause, with stress in that direction initiating the tearing. Thermal heating and stresses from weld shrinking create the fracture. Lamellar tearing can take place shortly after welding or occasionally months later. Thicker, higher-strength materials appear to be more susceptible. However, the phenomenon affects only a very small percentage of steel plates, as all three of the above conditions rarely occur in combination.

The problem can be avoided by proper attention to joint details. In T-joints (Fig. 16), double-fillet welds appear to be less susceptible than full-penetration welds. Also, balanced welds on both sides of the joint appear to present less risk than large, single-sided welds. In corner joints, common in box columns, lamellar tearing can be readily detected on the exposed edge of the plate (Fig. 17a). The problem can be overcome by placing the bevel for the joint on the edge of the plate that would exhibit the tearing rather than on the other plate (Fig. 17b). Butt joints rarely exhibit lamellar tearing, because weld shrinkage does not set up a tensile stress in the thickness direction of the plates.

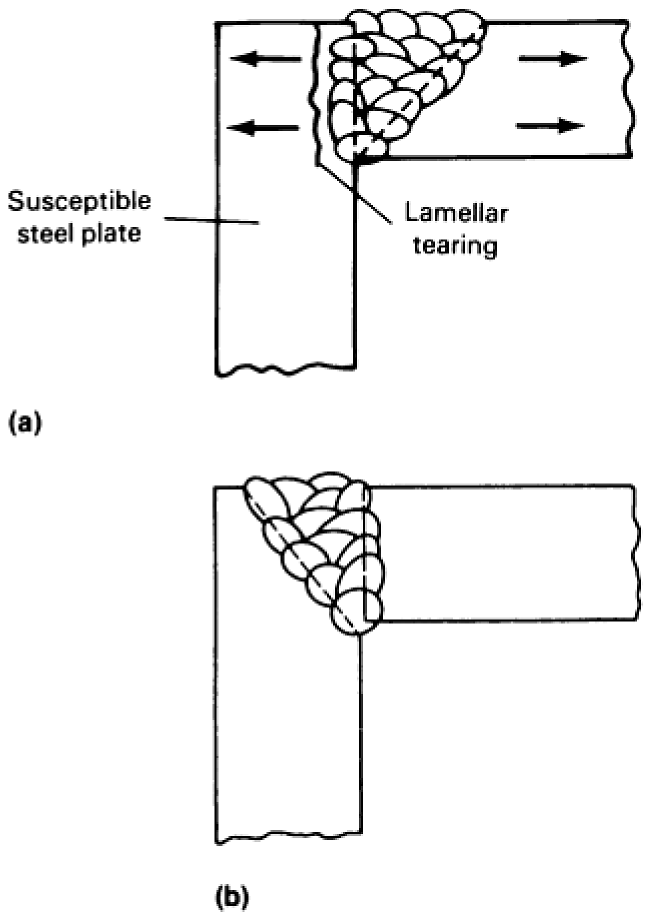

FIG. 17 CORNER JOINT. (A) LAMELLAR TEARING SURFACES AT THE EXPOSED PLATE EDGE. (B) REDESIGNED JOINT

Arc welding processes with high heat input are less likely to create lamellar tearing, perhaps because of the fewer number of applications of heat and the lesser number of shrinkage cycles involved in making a weld. Also, filler metal with lower yield strength and high ductility appears to lessen the possibility of tearing. Preheating does not appear to be particularly advantageous, nor does stress relieving. The "buttering" technique of laying one or more layers of low-strength, high-ductility weld metal deposit on the surface of the plate stressed in the *z* direction will reduce the possibility of lamellar tearing. However, this is an extreme solution and should be used only as a last resort.

Reference cited in this section

9. H.B. CARY, *MODERN WELDING TECHNOLOGY,* 2ND ED., PRENTICE-HALL, 1989, P 702-703

Cracking Phenomena Associated With Welding

Michael J. Cieslak, Physical and Joining Metallurgy Department, Sandia National Laboratories

---

References

1. J.C. BORLAND, GENERALIZED THEORY OF SUPER-SOLIDUS CRACKING IN WELDS (AND CASTINGS), *BR. WELD. J.,* VOL 7 (NO. 8), 1960, P 508-512

2. J.C. BORLAND, HOT CRACKING IN WELDS, *BR. WELD. J.,* VOL 7 (NO. 9), 1960, P 558-559
3. D.A. CANONICO, W.F. SAVAGE, W.J. WERNER, AND G.M. GOODWIN, EFFECTS OF MINOR ADDITIONS ON THE WELDABILITY OF INCOLOY 800, *EFFECTS OF MINOR ELEMENTS ON THE WELDABILITY OF HIGH-NICKEL ALLOYS,* WRC, JULY 1969, P 68-92
4. M.J. CIESLAK, THE WELDING AND SOLIDIFICATION METALLURGY OF ALLOY 625, *WELD. J.,* FEB 1991, P 49S-56S
5. W.F. SAVAGE AND C.D. LUNDIN, APPLICATION OF THE VARESTRAINT TECHNIQUE TO THE STUDY OF WELDABILITY, *WELD. J.,* NOV 1966, P 497S-503S
6. M.J. CIESLAK, J.J. STEPHENS, AND M.J. CARR, A STUDY OF THE WELDABILITY AND WELD RELATED MICROSTRUCTURE OF CABOT ALLOY 214, *METALL. TRANS. A,* MARCH 1988, P 657-667
7. J.J. PEPE AND W.F. SAVAGE, EFFECTS OF CONSTITUTIONAL LIQUATION IN 18-NI MARAGING STEEL WELDMENTS, *WELD. J.,* SEPT 1967, P 411S-422S
8. N. YURIOKA, WELDABILITY OF MODERN HIGH STRENGTH STEELS, *1ST U.S.-JAPAN SYMP. ADVANCES IN WELDING METALLURGY,* AWS, 1990, P 79-100
9. H.B. CARY, *MODERN WELDING TECHNOLOGY,* 2ND ED., PRENTICE-HALL, 1989, P 702-703

Characterization of Welds

Craig B. Dallam, The Lincoln Electric Company; Brian K. Damkroger, Sandia National Laboratories

---

Introduction

WELDS CAN BE CHARACTERIZED according to a number of criteria, including the welding process used, size, shape, mechanical properties, chemical composition, and a number of others. The appropriate methods of characterization depend on the weld's function and the particular set of properties required for the application. In some instances, the ability of a weld to function successfully can be addressed by characterizing the size or shape of the weld. An example of this is where factors related to the welding procedure, such as inadequate weld size, convexity of the bead, or lack of penetration, may cause a weld to fail. In other cases, it is important to characterize metallurgical factors such as weld metal composition and microstructure. Examples might include welds for which the goal is to avoid failures due to inadequate strength, ductility, toughness, or corrosion resistance. In general, the goals of weld characterization are to assess the ability of a weld to successfully perform its function, to thoroughly document a weld and welding procedure that have been demonstrated to be adequate, or to determine why a weld failed.

In the first part of this article, characterization of welds will be treated as a sequence of procedures, each more involved than the last and concerned with a finer scale of detail. Initially, non-destructive characterization procedures will be the focus. The first level of characterization involves information that may be obtained by direct visual inspection and measurement of the weld. A discussion of nondestructive evaluation follows. This encompasses techniques used to characterize the locations and structure of internal and surface defects, including radiography, ultrasonic testing, and liquid penetrant inspection.

The next group of characterization procedures discussed are destructive, requiring the removal of specimens from the weld. The first of these is macrostructural characterization of a sectioned weld, including features such as number of passes; weld bead size, shape, and homogeneity; and the orientation of beads in a multipass weld. Macroscopic characterization is followed by microstructural analysis, including microsegregation, grain size and structure, the phase makeup of the weld, and compositional analysis.

The third component of weld characterization is the measurement of mechanical and corrosion properties. The goal of any weld is to create a structure that can meet all the demands of its service environment. In many cases, the best way of assessing the performance of a weld is to establish its mechanical properties. In addition to a number of standard material tests, many mechanical tests are directed specifically at determining a weld's capabilities. Examples of mechanical properties typically characterized for welds include yield and tensile strength, ductility, hardness, and impact or fracture toughness. Corrosion testing is often employed in situations where a welding operation is performed on a corrosion-resistant material, or in a structure exposed to a hostile environment. Although absolute corrosion performance is important, a major concern is to ensure that a weld and its heat-affected zone (HAZ) are cathodic to the surrounding metal.

Following the discussion of the characterization procedures, the second part of this article will give examples of how two particular welds were characterized according to these procedures.

Acknowledgement

Support for the portion of this work performed at Sandia National Laboratories was provided by the U.S. Department of Energy under contract number DE-AC0476DP00789.

Characterization of Welds

Craig B. Dallam, The Lincoln Electric Company; Brian K. Damkroger, Sandia National Laboratories

Nondestructive Characterization Techniques

Externally Observed Macroscopic Features. Several factors associated with the production and performance of welds are macroscopic and easily observed. The most obvious of these are the size, shape, and general appearance of the weld. To a large extent, these parameters depend on the geometry of the weld joint and the welding process selected. Figure 1 shows schematic representations of fillet, lap, butt, and groove welds, in which a number of features (defined in Table 1) are labeled.

TABLE 1 NONMENCLATURE FOR FILLET, LAP, BUTT, AND GROOVE WELDS

FEATURE	DEFINITION
FACE	EXPOSED SURFACE OF A WELD ON THE SIDE FROM WHICH THE WELDING WAS PERFORMED
ROOT	POINTS, AS SHOWN IN CROSS SECTION, AT WHICH THE BACK OF THE WELD INTERSECTS THE BASE METAL SURFACE
LEG	SHORTEST DISTANCE FROM ROOT TO TOE IN A FILLET WELD
TOE	JUNCTION BETWEEN THE WELD FACE AND THE BASE METAL
THROAT	SHORTEST DISTANCE FROM ROOT TO FACE IN A FILLET WELD
PENETRATION	DEPTH A GROOVE WELD EXTENDS INTO THE ROOT OF A JOINT, MEASURED ON THE CENTERLINE OF A ROOT CROSS SECTION
REINFORCEMENT	WELD METAL IN EXCESS OF THE SPECIFIED WELD SIZE
FACE REINFORCEMENT	REINFORCEMENT AT THE SIDE OF THE WELD FROM WHICH WELDING WAS PERFORMED
SIZE (GROOVE WELD)	JOINT PENETRATION (DEPTH OF ROOT PREPARATION) PLUS THE ROOT PENETRATION (WHEN SPECIFIED)
SIZE (FILLET WELD)	LEG OF THE LARGEST ISOSCELES RIGHT TRIANGLE THAT CAN BE INSCRIBED WITHIN THE FILLET WELD CROSS SECTION

Source: Ref 1

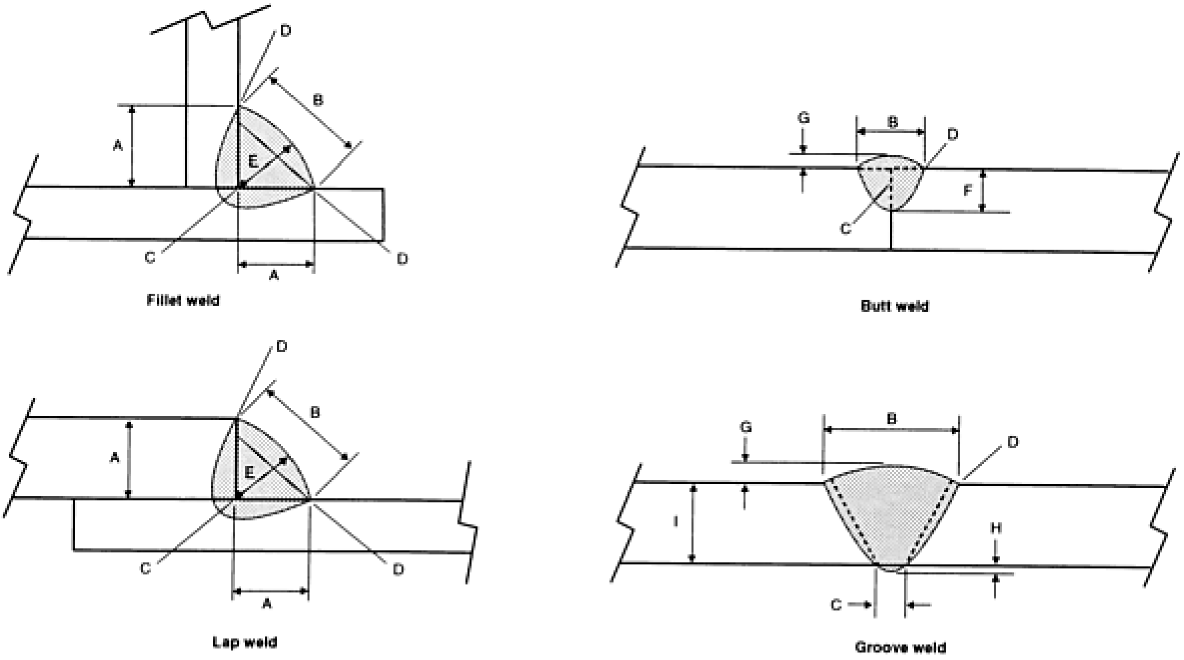

FIG. 1 SCHEMATICS SHOWING PARAMETERS OF COMMONLY USED JOINT GEOMETRIES. A, LEG; A', SMALLER LEG SIZE (EQUAL TO SIZE OF WELD); B, FACE, C, ROOT; D, TOE; E, THROAT; F, PENETRATION (EQUAL TO SIZE OF WELD); G, FACE REINFORCEMENT; H, ROOT REINFORCEMENT; I, SIZE OF WELD

Methods of Characterization. The primary tools used in the external macrocharacterization of a weld are the unaided eye and a hand lens. A number of macrocharacterization parameters (for example, weld location, size, shape, and general uniformity) can be assessed by visual inspection. Figure 2 shows top-view macrographs of two welds that illustrate several of these considerations. In many cases, the presence of gross defects such as hot cracking or porosity may also be detected by visual inspection and a top-view photograph. Additional examples of factors included in overall weld bead appearance include distortion, discoloration due to inadequate shielding or excessive heat, undercut, excessive crater size, and uneven bead width. In fillet welds, a number of performance aspects can be affected by the relative geometry of the joint. Fillet gages can be used to determine the size of a weld, the curvature of the face, and the length of the weld legs.

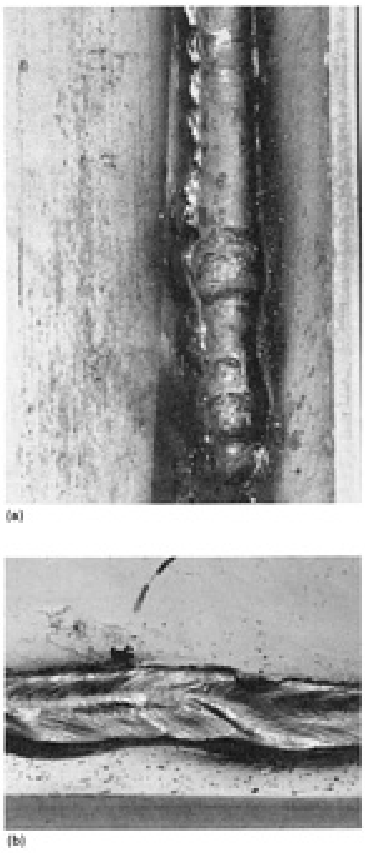

FIG. 2 MACROGRAPHS SHOWING TOP VIEW OF FILLET WELDS. (A) VERTICAL WELD IN TYPE 304L STAINLESS STEEL (NOTE CONVEXITY OF BEAD AND NONUNIFORM REGIONS). (B) HORIZONTAL E 7024 WELD (NOTE WELD PLACEMENT AND LACK OF UNIFORMITY)

General Considerations. Most types of welds will be evaluated with respect to specific macrocharacteristics deemed to be either desirable or undesirable. However, a number of general considerations apply to most types of welds. These include:

- *SIZE:* THE SIZE OF THE WELD SHOULD BE APPROPRIATE FOR THE PART (AS AN EXAMPLE, A GENERAL RULE FOR FILLET WELDS IS THAT THE RATIO OF LEG SIZE/PLATE THICKNESS SHOULD BE BETWEEN 3:4 AND 1:1.
- *LOCATION:* AN INCORRECTLY LOCATED BUTT WELD MAY NOT ALLOW THE PART TO FUNCTION CORRECTLY. A LESS EXTREME EXAMPLE IS A FILLET WELD WITH UNEQUAL LEG LENGTHS, LEADING TO AN UNEVEN STRESS DISTRIBUTION AND PERHAPS LAMELLAR TEARING.
- *UNIFORMITY:* DISTORTION, THE PROBABILITY OF SLAG ENTRAPMENT ON MULTIPASS WELDS, AND UNIFORMITY OF LOAD-CARRYING ABILITY ALL DEPEND ON THE RELATIVE UNIFORMITY OF THE WELD.
- *DEFECTS:* IDEALLY, A WELD SHOULD BE FREE OF MACROSCOPIC DEFECTS. COMMON DEFECTS INCLUDE UNDERCUTTING, LACK OF FUSION, PINHOLE POROSITY, AND SLAG ENTRAPMENT.
- *FACE SHAPE:* A WELD SHOULD HAVE A RELATIVELY FLAT FACE. IF THE WELD IS TOO CONVEX, STRESS WILL BE CONCENTRATED ALONG THE TOE OF THE WELD. CONVERSELY, A HIGHLY CONCAVE WELD FACE WILL RESULT IN LOCALLY HIGH STRESSES IN THE THROAT REGION OF THE WELD.

Defects. The presence of surface or internal defects can degrade the performance of a weld that may appear to be sound in a cursory examination. The causes and means of preventing defects are covered in Ref 2, 3, and 4. Typical defects found in welds include:

- POROSITY
- UNDERCUT
- OVERLAP
- LACK OF PENETRATION
- LACK OF FUSION
- SHRINKAGE VOIDS
- CRATER CRACKING
- MELT-THROUGH OR BURN-THROUGH
- SUBSOLIDUS OR HOT CRACKING
- INCLUSIONS DUE TO ENTRAPPED SLAG, TUNGSTEN ELECTRODE PIECES, OR DEFECTS PRESENT IN THE STARTING MATERIAL

A number of techniques are widely used to assess the presence of surface and subsurface defects in welds. The most common of these are liquid penetrant inspection for surface cracks, magnetic particle inspection, x-ray radiography, and ultrasonic inspection. Penetrant inspection involves the application of an indicator fluid that has a surface tension sufficiently low to be drawn into surface cracks too small to be detected visually. The excess dye or fluorescent material is then removed from the surface, but it remains in and highlights the cracks when a developer is applied. Figure 3 shows cracks in a flux-cored are weld highlighted by dye penetrant inspection. Table 2 lists the characteristics of several types of nondestructive inspection techniques. Additional information is available in the article "Inspection of Welded Joints" in this Volume.

TABLE 2 GUIDELINES FOR SELECTING TECHNIQUES

Technique	Equipment Requirements	Defects Detected	Advantages	Limitations	Comments
liquid-penetrant or fluorescent-penetrant inspection	fluorescent or visible penetration liquids and developers; ultraviolet light for fluorescent dyes	defects open to the surface only; good for leak detection	detects small surface imperfections; easy application; inexpensive; use on magnetic or nonmagnetic material; low cost	time-consuming; not permanent	used on root pass of highly critical pipe welds; indications may be misleading on poorly prepared surfaces
magnetic particle inspection	wet or dry iron particles, or fluorescent; special power source; ultraviolet light for fluorescent dyes	surface and near-surface discontinuities: cracks, porosity, slag	indicates discontinuities not visible to the naked eye; useful for checking edges before welding; no size limitations	for magnetic materials; surface roughness may distort magnetic field; not permanent	test from two perpendicular directions to detect any indications parallel to one set of magnetic lines
radiographic inspection	x-ray or γ ray; film processing and viewing equipment	most internal discontinuities and flaws; limited by direction of discontinuity	provides permanent record of surface and internal flaws; applicable to any alloy	usually not suitable for fillet weld inspection; film exposure and processing	popular technique for sub-surface inspection

				critical; slow and expensive	
ultrasonic inspection	ultrasonic units and probes; reference patterns	can locate all internal flaws located by other methods, as well as small flaws	extremely sensitive; complex weldments restrict usage; can be used on all materials	highly skilled interpretor required; not permanent	required by some specifications and codes

Source: Ref 5

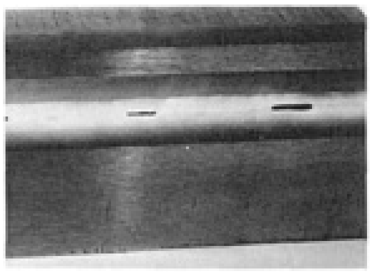

FIG. 3 SURFACE CRACKS IN A FLUX-CORED ARC WELD HIGHLIGHTED BY DYE PENETRANT INSPECTION

X-ray radiography is used to detect internal defects such as porosity or inclusions: these defects show up because of the difference in x-ray absorption between the matrix and defect materials. Although a number of factors affect the resolution level of radiographic techniques (Ref 5, 6), the minimum size of defects considered in AWS specifications is 0.4 mm (0.0156 in.). In practice, this refers to slag entrapment and large inclusions that were present in the starting material. Defect structures are usually quantified by comparison with an existing standard, of which several exist. An example of such a standard is American Petroleum Institute (API) 1104 (Ref 7), a pipeline welding specification. The API 1104 standard includes criteria for acceptance of steel welds based on a number of defects, including inadequate penetration, incomplete fusion, internal concavity, burn-through, slag inclusions, and porosity. For example, one section of API 1104 can be summarized as follows:

> Slag inclusions--50 mm (2 in.) maximum length for elongated slag inclusions, 50 mm (2 in.) limit per 305 mm (12 in.) of weld, 1.6 mm ($\frac{1}{16}$ in.) width limit per elongated slag inclusion. Isolated slag inclusions, 13 mm ($\frac{1}{2}$ in.) per 305 mm (12 in.) of weld, width greater than 3.2 mm ($\frac{1}{8}$ in.), 105 mm ($4\frac{1}{8}$ in.) isolated slag inclusions per 305 mm (12 in.) of weld. Aggregate length of isolated slag inclusion indications may not exceed two times the thinner of the nominal wall thickness and the width may not exceed one-half the thinner of the wall thicknesses joined.

Other examples include ASTM E 390 (Ref 8), which contains reference radiographs for steel welds and the *Structural Welding Code* (Ref 9), which includes a group of radiographic plates showing defect patterns for rounded inclusions corresponding to grade 1 and grade 2 quality levels. Several industries and product categories have applicable specifications and standards. A number of these are given in Ref 5.

Ultrasonic testing can also be used to locate internal defects, including porosity and inclusions. Ultrasonic testing involves transmitting mechanical vibrations through a piece of metal and analyzing both reflected and transmitted vibrations. Vibrations interact with discontinuities in the media through which they are passing, so the operator can detect voids, inclusions, and other internal interfaces. Table 2 includes ultrasonic testing, and Ref 5 lists a number of example specifications and detailed references.

References cited in this section

1. STANDARD WELDING TERMS AND DEFINITIONS, ANSI/AWS A3.0-89, AWS, 1989
2. WELD DISCONTINUITIES, *METALS HANDBOOK,* 9TH ED., VOL 6, AMERICAN SOCIETY FOR METALS, 1983, P 829-855
3. *WELDING HANDBOOK,* 8TH ED., VOL 2, AWS, 1987
4. *PROCEDURE HANDBOOK OF ARC WELDING,* 12TH ED., LINCOLN ELECTRIC CO., 1973
5. CODES, STANDARDS, AND INSPECTION, *METALS HANDBOOK,* 9TH ED., VOL 6, AMERICAN SOCIETY FOR METALS, 1983, P 823-838
6. *RADIOGRAPHY IN MODERN INDUSTRY,* 4TH ED., EASTMAN KODAK CO., 1980
7. WELDING OF PIPELINES AND RELATED FACILITIES, API STANDARD 1104, 17TH ED., AMERICAN PETROLEUM INSTITUTE, 1988
8. REFERENCE RADIOGRAPHS FOR STEEL WELDMENTS, ASTM E 390, ASTM
9. *STRUCTURAL WELDING CODE--STEEL,* ANSI/AWS D1.1-92, AWS, 1992

Characterization of Welds

Craig B. Dallam, The Lincoln Electric Company; Brian K. Damkroger, Sandia National Laboratories

Internal Characterization Requiring Destructive Procedures

This section briefly covers macrostructural and microstructural characterization and compositional analysis. These procedures are usually based on a cross section of the weld, referred to as a transverse section. The transverse section may be supplemented by the top-view photograph, the longitudinal section, and the normal section (Fig. 4). In particular, the top-view photograph will depict the general appearance of the weld, illustrating surface irregularities, spatter, or macroscopic defects such as hot cracking or porosity.

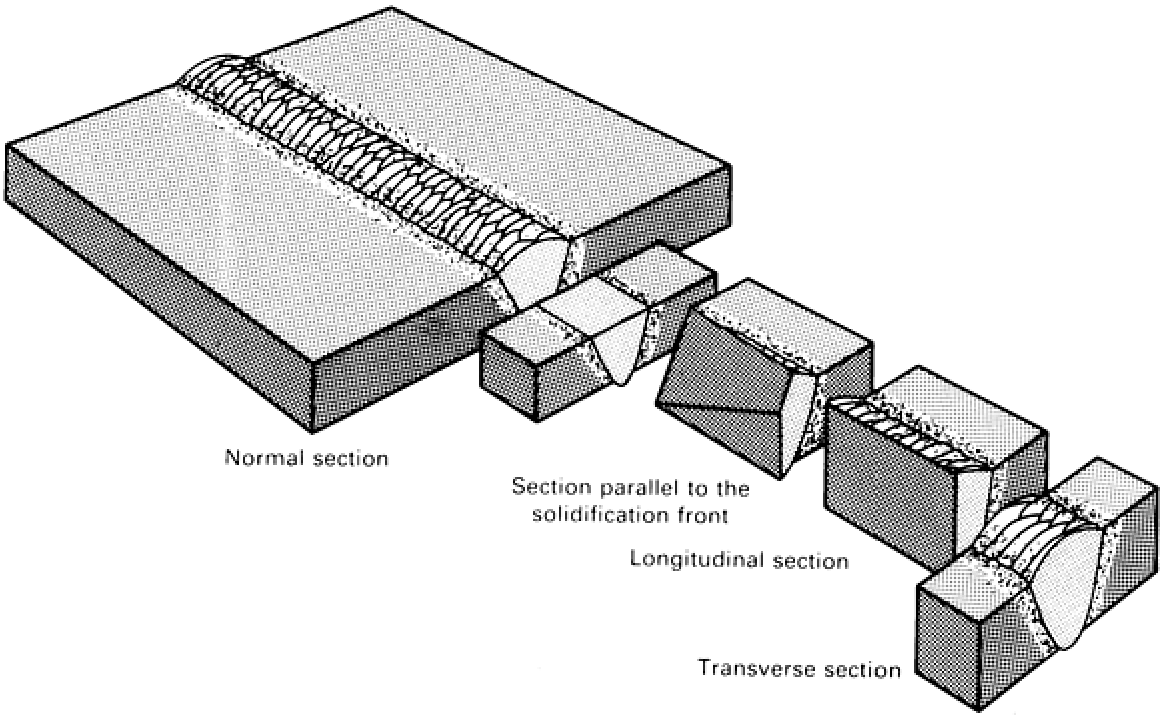

FIG. 4 TYPICAL SECTIONS USED IN THE METALLOGRAPHIC EXAMINATION OF WELDED JOINTS

Weld Macrostructure

Figures 5(a) and 5(b) show transverse sections of two welds, a submerged arc weld made on a 25 mm (1 in.) thick ASTM A 36 steel and a flux-cored weld made on a 50 mm (2 in.) thick ASTM A 537 steel. Readily apparent features include the number of passes and number of layers, fusion zone area, weld aspect ratio, extent of penetration, face width, and the reinforcement and curvature of the top bead. A transverse macrosection will also show any gross porosity or large inclusions present in a weld and the extent of the HAZ. It is important to note that quantitative measurements of weld features made based on a transverse section will be accurate only if the section is properly done. Potential errors, sectioning procedures, and sample preparation techniques are discussed in Ref 10.

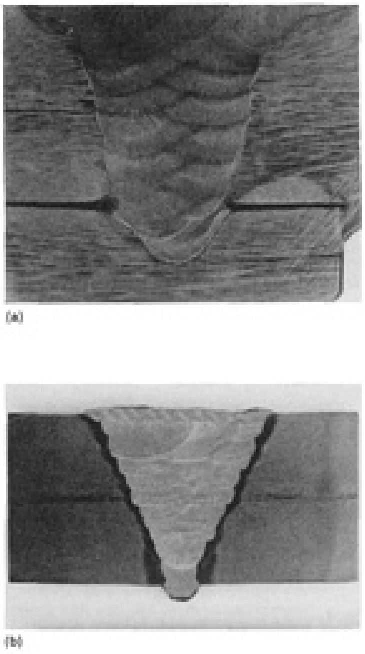

FIG. 5 TYPICAL MULTIPASS ARC WELDS IN STEELS USED IN STRUCTURAL APPLICATIONS. (A) SUBMERGED ARC WELD ON A 25 MM (1 IN.) THICK A 36 STRUCTURAL STEEL; THE MUSHROOM SHAPE OF THE LAST BEAD IS TYPICAL OF WELDS PRODUCED BY THIS PROCESS. (B) FLUX-CORED ARC WELD ON A 50 MM (2 IN.) THICK A 537 STEEL USED IN PRESSURE VESSEL AND STRUCTURE APPLICATIONS (THE LAST LAYER WAS MADE WITH SEVERAL SMALL PASSES TO IMPROVE MECHANICAL PROPERTIES)

Weld Microstructure

In many cases, it is important to examine and characterize the weldment microstructure and to understand its formation and effects on properties. This is the case when the materials and processes involved are not well characterized, so specifications have not been established. Other problems arise where the potential exists for the formation of detrimental microconstituents and/or where the consequences of weld failure are severe. Welding of high-strength or brittle materials are typical examples.

The microstructure of a weld consists of three regions: a fusion zone (material that has been melted); a heat affected zone (material that was not melted, but whose microstructure has been altered); and the base metal. These three regions can be seen in the transverse section shown in Fig. 5(b). Weld microstructures are examined using standard specimen removal and preparation techniques, with some concessions made for their inhomogeneous nature (Ref 10). Similarly, the parameters used to characterize the weld microstructures, such as grain size, grain morphology, and the amount of the various phases or microconstituents present, are those used to characterize monolithic materials. The equipment, procedures, and techniques used are well described in Ref 10.

Special importance may be attached to one or more specific features of a weld microstructure. An example is the classification of ferrite and carbide microconstituents in a low-ally steel weld. The various morphologies included in a characterization of a low-alloy steel weld are listed in Table 3. In these welds, a large amount of acicular ferrite is associated with high toughness levels, whereas martensitic or bainitic microstructures have much lower toughness levels. Bainite and martensite are also associated with higher effective cooling rates, so decreasing the weld metal manganese content, or reducing the cooling rate with increased heat input or preheat, will increase the amount of acicular ferrite and improve weld metal toughness.

TABLE 3 DESCRIPTION OF FERRITE/CARBIDE MICROCONSTITUENTS IN LOW-CARBON STEEL WELDS

Microconstituent	Description
Primary ferrite	
Grain boundry ferrite	Proeutectoid ferrite that grows along the prior austenite grain boundaries. It is equiaxed or polygonal, and it may occur in veins
Intragranular polygonal ferrite	Polygonal ferrite that is not associated with the prior austenite grain boundaries. Much larger than the average width of the surrounding acicular ferrite laths
Ferrite with second phase	
With aligned second phase	Parallel ferrite laths. Further classifications are widmanstätten ferrite and upper/lower bainite
With nonaligned second phase	Ferrite that completely surrounds either equiaxed, randomly distributed microphases, or solitary ferrite laths
Ferrite/carbide aggregate	Fine ferrite/carbide structures, including pearlite
Acicular ferrite	Small nonaligned ferrite grains found within prior austenite grains
Martensite	Martensite colony larger than adjacent ferrite laths

Microstructural characterization of welds has two purposes: to evaluate the microstructure with respect to properties and to relate the microstructure to the process used. The ultimate goal is to optimize the process to produce the most desirable microstructure. In general, the effects of a process and parameters on microstructure are due to the compositional and thermal effects. The compositional effects are largely limited to the fusion zone and will be discussed in the next section. Thermal cycles affect both the fusion zone and HAZ. Microstructural development in the HAZ has been the subject of much study, and a number of computational tools have been developed specifically for the prediction of HAZ microstructures (Ref 11).

Composition of a Weld

The composition of a weld will have a significant effect on its performance, contributing to both the mechanical and corrosion properties of the weld. In some instances, the effects can be drastic, such as sensitization of stainless steel or changing the ductile-to-brittle transition temperature of ferritic steels by an amount sufficient to cause brittle failure under normal use. In a number of cases, particular elements will govern the propensity toward certain defects or behavior. For example, a low-alloy steel weld with a carbon equivalent of over 0.4% is considered to be a high-carbon deposit, and it is treated differently than a low-carbon weld. Similarly, deposit silicon levels in excess of approximately 0.4% are often associated with a defect called "hollow bead," where the root pass in a pipe weld made with a cellulosic electrode has a continuous pore. In general, care must be taken to either minimize the compositional changes associated with welding or to factor the composition and property changes into the design.

Factors Affecting Weld Composition. The composition of a weld is affected by the base metal composition, the composition of any filler metal used (and dilution between the two), reactions with flux or shielding gas, and any material losses associated with the process. These factors are, in turn, largely controlled by the welding setup, process selection, parameters, and stability. The Schaeffler diagram (Ref 12) for stainless steels is an example of composition being controlled by dilution. The weld metal composition and microstructure are predicted based on the "ferrite stabilizer" and the "austenite stabilizer" content of the base metal, the nickel equivalent and chromium equivalent of the filler metal, and the level of dilution. The microstructure will be predicted as martensitic, ferritic, or austenitic, and the level of ferrite in the weldment will be estimated. In these welds, ferrite numbers below 4 indicate microstructures in which low-melting-point compounds may be formed at the grain boundaries, and the welds can be prone to hot cracking. Conversely, welds with ferrite numbers above 10 have reduced corrosion resistance and can be susceptible to the formation of σ phase at high temperatures (Ref 13). The Schaeffler diagram has been revised many times for adaptation to different alloy modifications. An example is the inclusion of nitrogen in the nickel equivalent. Many stainless steels often contain small amounts of nitrogen, and in fact some stainless steels are deliberately strengthened with nitrogen additions. Other modifications include copper in the nickel equivalent and vanadium and aluminum in the chromium equivalent (Ref 14). Additional information is available in the Section "Selection of Stainless Steels" and in the article "Welding of Stainless Steels" in this Volume.

An example of the effect of flux on weld metal composition is seen when low-carbon steel is welded with E6010 and E7018 electrodes, using identical parameters. Although the core wire composition of the two electrodes is identical, the E6010 electrode will produce a deposit with approximately 0.15 to 0.25% Si, and the E7018 electrode will produce a deposit with 0.5 to 0.6% Si. An example of a procedure-related effect is seen when the arc length is changed for welds made with an E7018 electrode. Arc length changes can alter the silicon content from 0.3 to 0.6%, and they can alter the manganese content from 0.8 to 1.3%.

Measurement Techniques and Procedures. The usual techniques for compositional analysis of metallic samples can be used on welds and are covered in Ref 15. However, some considerations must be made when dealing with welds, because of their inhomogeneity. The most commonly used method for compositional analysis of welds is optical emission spectroscopy, where a spot of material (typically 6 mm, or $\frac{1}{4}$ in., in diameter) is ablated off the surface of the specimen and the light emissions are analyzed. In many welds, a spot this size will encompass several passes in a multipass weld, and it will perhaps encompass weld metal, HAZ, and base metal in a single-pass weld. An average composition over this large an area may not address local effects, for example, sensitization of stainless steel.

Welds are prone to both macrosegregation and microsegregation, and the properties may be governed by the composition in a very local region. A thorough characterization of welds requires techniques with sufficient spatial resolution to characterize their inhomogeneity. Scanning electron microscopy with wavelength or energy dispersive x-ray analysis systems, electron microprobes, and x-ray fluorescence techniques are often employed. Often, a number of closely spaced analyses are made along a line traversing a weld. This data may then be coupled with a similar microhardness traverse and microstructural analyses to characterize the variation across the weld. Thorough treatments of a number of microanalytical techniques are contained in Ref 16.

Mechanical Testing

A number of mechanical properties are used to characterize welds, including strength, ductility, hardness, and toughness. In general, the same samples and procedures are used in other areas of metallurgy (Ref 17). However, a prominent concern regarding the mechanical performance of welds is the direct comparison with base material. The goal is to ensure that the weld is not the weakest component of a structure, or if it is, to compensate for this in the design.

Strength. Yield and tensile strength are measured for all-weld-metal specimens using a standard tensile test (Ref 18), but with specimens removed from test plates welded according to AWS-specified procedures (Ref 19). These tests form the basis for the assignment of yield and ultimate strength values to welds made using a specific electrode and according to a set procedure. Additional tests are sometimes performed to compare the base metal and weld metal strengths. An example of this type of test is the transverse tensile test, in which the specimen is removed from the weld so that the loading axis is perpendicular to the weld bead and the weld reinforcement is left intact. The goal of this test is to verify that overload failure will occur in the base metal rather than in the weld metal or HAZ.

Ductility is another critical weld property. In addition to defects, many welding processes can produce hard, brittle microstructures. The standard measures of ductility--percent reduction in area and percent elongation--are obtained in a uniaxial tensile test. Another test often specified for welds is a bend test (face bends, roof bends, and side bends). In this test, a strip of material containing a weld is deformed around a specified radius and its surface is examined. The criteria for success or failure are the number and size of defects seen on the outer surface of the bend. An example of bend test criteria is the AWS *Structural Welding Code* (Ref 9), which calls for bending around a 19 mm (0.75 in.) radius for materials with yield strengths less than or equal to 345 MPa (50 ksi), a 25 mm (1 in.) radius for 345 to 620 MPa (50 to 90 ksi) materials, and a 32 mm (1.25 in.) radius for materials with yield strengths greater than or equal to 620 MPa (90 ksi).

Hardness. One common use of hardness values in weld specifications is as a check for the formation of microstructures that might have low ductility and toughness and thus are prone to cracking. For example, in pipeline steels, the formation of martensite in the HAZ is a cause for concern because of the potential for cracking. This is addressed by specifying maximum values for microhardness traverses across several sections of the weld. Hardness values are also used as an indicator of susceptibility to some forms of stress-corrosion cracking.

Toughness is the ability of a material to absorb energy during fracture. There are two approaches to toughness testing: impact toughness testing and fracture mechanics testing.

Impact Toughness Testing. To test impact toughness, a sample of specified geometry is subjected to an impact load, and the amount of energy absorbed during fracture is recorded. Usually the specimen is oriented so that the notch and expected plane of fracture run longitudinally through the weld metal. Charpy tests do not measure an inherent material property, but they result in a relative measure of impact toughness between materials. A very common use of the Charpy test is to determine a material's ductile-to-brittle transition temperature by performing tests at several different temperatures. AWS A5.1 (Ref 19) gives minimum Charpy impact values, at several temperatures, for welds made on carbon steel using a number of different electrodes.

Fracture Mechanics Testing. The second type of toughness testing is based on fracture mechanics, and it can use either linear elastic or elastic-plastic methodologies. Although elastic-plastic behavior (J_{Ic}) is becoming of interest in some cases, the bulk of fracture mechanics testing is based on linear elastic considerations. These tests, using specimens and procedures given in ASTM E 399 (Ref 20), are used to measure a material's fracture toughness (K_{Ic}), which is a material property. In the case of welding, fracture toughness is usually expressed using a value for crack tip opening displacement (Ref 21). Fracture toughness testing has only recently begun gaining acceptance as applicable to welds. The major shortcomings of this approach include the complexity and cost of testing and the wide variability in fracture toughness values for weld metal, due largely to the inhomogeneous nature of welds and residual stress effects.

References cited in this section

9. *STRUCTURAL WELDING CODE--STEEL,* ANSI/AWS D1.1-92, AWS, 1992
10. METALLOGRAPHY OF WELDMENTS, *METALS HANDBOOK,* 9TH ED., VOL 9, AMERICAN SOCIETY FOR METALS, 1985, P 577-586
11. K.E. EASTERLING, MICROSTRUCTURE AND PROPERTIES OF THE HAZ, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* S. DAVID AND J. VITEK, ED., ASM INTERNATIONAL, 1990, P 177-188
12. WRC 1992 DIAGRAM, *WELD. J.,* MAY 1992, P 171
13. T.G.F. GRAY, J. SPENCE, AND T.H. NORTH, *RATIONAL WELDING DESIGN,* NEWNES-BUTTERWORTHS, 1975, P 111
14. R.H. ESPY, WELDABILITY OF NITROGEN-STRENGTHENED STAINLESS STEELS, *WELD. J. RES. SUPPL.,* VOL 61 (NO. 5), MAY 1982, P 149S-156S
15. METAL TEST METHODS AND ANALYTICAL PROCEDURES, *1991 ANNUAL BOOK OF ASTM STANDARDS,* VOL 03.05 AND 03.06, ASTM, 1991
16. *MATERIALS CHARACTERIZATION,* 9TH ED., VOL 10, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1986
17. *MECHANICAL TESTING,* 9TH ED., VOL 8, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1985
18. STANDARD METHODS OF TENSION TESTING OF METALLIC MATERIALS, ASTM E 8, *ANNUAL BOOK OF ASTM STANDARDS,* VOL 03.01, ASTM, 1984
19. *SPECIFICATION FOR CARBON STEEL ELECTRODES FOR SHIELDED METAL ARC WELDING,* ANSI/AWS A5.1-91, AWS, 1991
20. STANDARD TEST METHOD FOR PLANE-STRAIN FRACTURE TOUGHNESS OF METALLIC MATERIALS, ASTM E 399, *ANNUAL BOOK OF ASTM STANDARDS,* ASTM, 1984
21. STANDARD TEST METHOD FOR CRACK TIP OPENING DISPLACEMENT (CTOD) FRACTURE

Characterization of Welds

Craig B. Dallam, The Lincoln Electric Company; Brian K. Damkroger, Sandia National Laboratories

Weld Characterization Examples

In this section, two weld will be analyzed to serve as examples of techniques discussed thus far in this article. The first of these welds is a multipass shielded metal arc weld/flux-cored arc weld made on pipe steel for an industrial application. This type of weld would typically be made in great quantity. In practice, it would be characterized largely on the basis of its external appearance, perhaps supplemented by nondestructive testing. The initial characterization of a weld and procedure, however, would be based on chemical composition and mechanical properties. Primary concerns are whether the weld being examined is representative of a normal weld of this type and can thus be reasonably well described based on the materials, process, and specifications involved in its production. The bulk of the actual characterization of such a weld would be done during the development and certification of the initial process to determine its suitability for the intended application.

The second example is a weld used for a very specific application. This weld is a butt weld made on Ti-6Al-4V, with a tantalum shim inserted between the base metal sections. Among the requirements for this weld are that it have microstructural and property homogeneity and stability at elevated temperatures and that it be cathodic to the base metal under unusual and extremely severe environmental conditions. In this case, the concerns center less on the eventual application and more on building a thorough database regarding the metallurgical and microstructural characteristics of the weld.

Example 1: Multipass Weld in a 1.07 m (42 in.) Diameter X-65 Pipe.

This weld is a circumferential weld on a 12.7 mm (0.50 in.) thick, 1.07 m (42 in.) diameter X-65 pipe. The X-65 materials is designated by API as a pipe steel with at least 448 MPa (65 ksi) yield strength and 566 MPa (82 ksi) tensile strength. This type of weld is typical for a large oil or gas pipeline. However, for this example, procedures were deliberately selected to produce a weld with a range of defects. These procedures, given by Table 4, are not representative of standard practice.

TABLE 4 PARAMETERS (NOT STANDARD PROCEDURE) USED TO OBTAIN MULTIPASS WELD IN 1.07 M (42 IN.) DIAMETER X-65 STEEL PIPE

PROCESS	PASS NO.	ELECTRODE	DIAMETER		WIRE FEED		CURRENT, A	VOLTAGE, V	TRAVEL SPEED	
			mm	in.	mm/s	in./s			mm/s	in./s
SMAW	1	E8010-G	4.0	0.16	. . .	. . .	150	26-28	5.9-7.2	2.3-2.8
	2	E8010-G	4.0	0.16	. . .	. . .	180	26-28	5.9-7.2	2.3-2.8
FCAW	3	E71T8-K6	1.73	0.068	47	1.9	200	19	4.5	1.8
	4	E71T8-K6	1.73	0.068	47	1.9	200	19	4.0	1.6
	5	E71T8-K6	1.73	0.068	47	1.9	200	19	4.0	1.6
	6	E71T8-K6	1.73	0.068	47	1.9	200	19	4.7	1.9
	7	E71T8-K6	1.73	0.068	47	1.9	200	19	5.9	2.3
	8	E71T8-K6	1.73	0.068	42	1.7	180	18	4.2	1.7
	9	E71T8-K6	1.73	0.068	42	1.7	180	18	4.5	1.8

Additional specifications include: 12.7 mm (0.50 in.) wall thickness; 1.59 mm (0.0625 in.) land; 1.59 mm (0.0625 in.) gap; 60° bevel in groove; vertical down position; preheat and interpass temperature of 150 °C (300 °F)

Visual Observation. The weld was initially characterized by visual observation. From the top of the joint, the size and uniformity of the weld and undercut can be observed (Fig. 6a). From the underside, spatter, lack of penetration, and internal undercut can be seen (Fig. 6b).

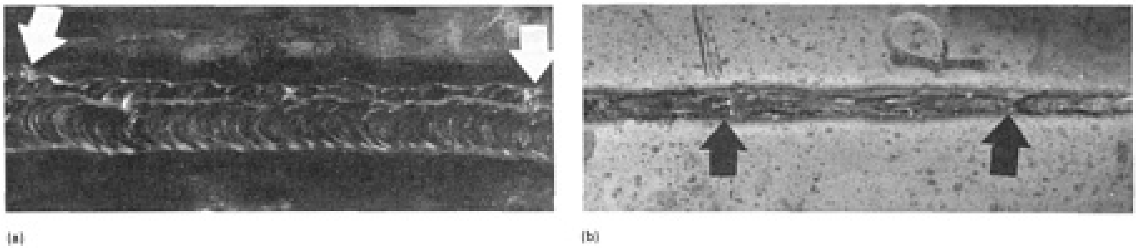
(a)
(b)

FIG. 6 DEFECTS IN A MULTIPASS WELD MADE ON A 1.07 M (42 IN.) DIAMETER X-65 STEEL PIPE. (A) EXTERIOR VIEW SHOWING LACK OF UNIFORMITY (RIGHT ARROW) AND UNDERCUT (LEFT ARROW). (B) INTERIOR VIEW SHOWING LACK OF PENETRATION (RIGHT ARROW) AND BURN-THROUGH (LEFT ARROW)

Nondestructive evaluation was then used to examine the weld for internal and surface defects. Slag entrapment and lack of fusion are typical internal defects for this type of weld, and they can be seen in the radiograph shown in Fig. 7. Figure 7 also shows the external defects noted during visual observation. In addition to nondestructive evaluation, other tests that highlight defects (for example, "nick-break" tests and face or root bend tests) can be run, but these were not used in this case. Similarly, ultrasonic testing, magnetic particle testing, and penetrant testing were not used in this characterization.

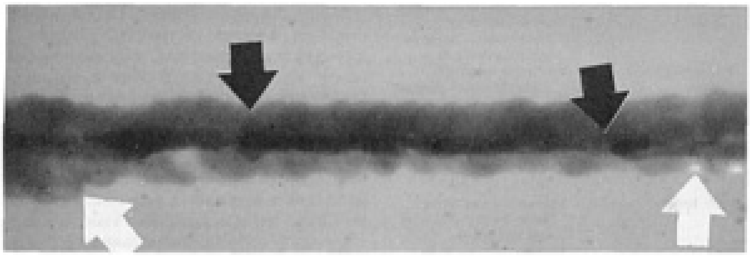

FIG. 7 RADIOGRAPH SHOWING POSITIVE IMAGE OF MACROSCOPICALLY VISIBLE DEFECTS IN THE X-65 PIPE STEEL OF FIG. 6. SEE FIG. 6 FOR DESIGNATION OF SPECIFIC DEFECTS.

Weld Macrostructure. Destructive evaluation begins by removing a transverse section from the weld and preparing it for metallographic examination (Fig. 8). The welding pass sequence (that is, number and size of passes, number of passes per layer, amount of penetration, and extent of HAZ) is apparent. In addition, any gross defects present in the sectioning plane are visible. In this example, nine passes were made, three layers with one pass and three with two passes. Figure 8 also shows four small pores, one each in the third, fourth, sixth, and ninth passes. In this case, the porosity was induced by deliberately varying the electrode stickout.

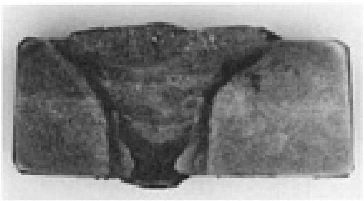

FIG. 8 TRAVERSE SECTION OF THE X-65 PIPE STEEL SHOWN IN FIG. 6 AND 7. PORES IN PASSES 3, 4, 6, AND

8 WERE CAUSED BY INTENTIONALLY VARYING THE ELECTRODE STICKOUT.

Weld Metal Composition. Compositional analysis of the weld was performed at this point. For this application, the bulk, or average, composition of the weld is of the most interest. The bulk composition of the weld was measured using an emission spectrometer, and the results are given in Table 5. The interstitial content of the weld is also of interest, and the levels of oxygen and nitrogen were measured with dedicated instruments. The oxygen level of the weld was found to be 110 ppm and the nitrogen level, 330 ppm. For some welds, techniques can be used to map compositional variations within the weld. In this case, the bulk composition, measured at the weld center, is considered adequate to characterize the weld.

TABLE 5 BULK COMPOSITION OF WELD IN 1.07 M (42 IN.) DIAMETER X-65 STEEL PIPE

SAMPLE	COMPOSITION, WT. %									
	C	Mn	Si	Al	Ni	Cr	Mo	Ti	V	Nb
WELD METAL	0.062	0.85	0.22	0.76	0.65	0.02	0.02	0.011	0.008	0.003
BASE METAL	0.093	1.35	0.21	0.02	0.02	0.04	0.00	0.002	0.069	0.033
ELECTRODES:										
E8010-G	0.14	0.70	0.20	0.01	0.73	0.02	0.07	0.008	0	0
E71T8-K6	0.056	0.91	0.24	0.95	0.97	0.02	0.01	0.002	0	0

Weld Microstructure. The transverse section was also used for examination and characterization of the microstructure of the weld. Figure 9 shows micrographs of different regions of the weld. Figure 9(a) shows weld metal that has not been reheated by successive passes. Figure 9(b) shows weld metal that has been reheated and has a finer structure than the non-reheated weld metal. Figure 9(c) shows a location in the HAZ that closely resembles the structure seen in the reheat zone. The important parameters to characterize for this type of weld are the percentage of reheat zone, the relative amount of various microconstituents, and the average grain size. In the example weld, the weld metal is approximately 80% reheat zone. In the non-reheated zone (Fig. 9a), quantitative microstructural analysis showed that the structure consisted of 20% grain boundary ferrite, 20% intergranular polygonal ferrite, 16% acicular ferrite, 32% ferrite with an aligned second phase, and 12% ferrite with a nonaligned second phase. No martensite or ferrite-carbide aggregate was found.

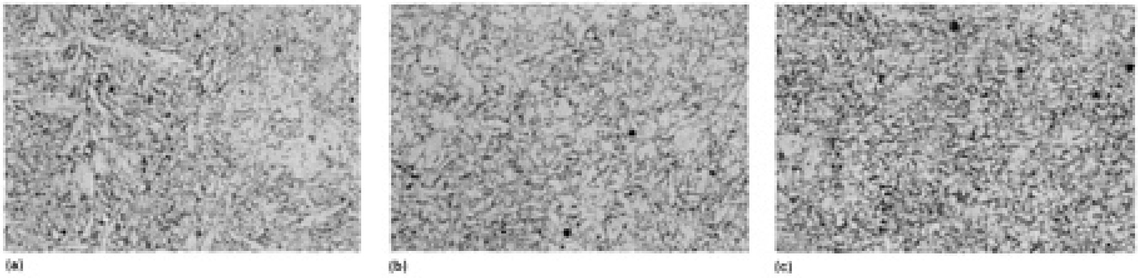

FIG. 9 MICROSTRUCTURE OBTAINED IN SELECTED REGIONS OF THE TRAVERSE SECTION OF THE X-65 PIPE STEEL OF FIG. 8. (A) NON-REHEATED WELD METAL. (B) REHEATED WELD METAL. (C) HEAT-AFFECTED ZONE. ALL 250×

Mechanical Testing. Following the characterization of the microstructure, specimens were removed and used for mechanical testing. For this application, the pertinent mechanical tests are tension tests (transverse and all-weld-metal), Charpy impact tests, fracture toughness tests, and a microhardness traverse. A transverse tensile test is primarily used to ensure that the welded joint is not the weak link in the final structure. In this case, a 305 × 32 mm (12.0 × 1.25 in.) specimen was removed from the pipe, the long axis perpendicular to the weld bead. The specimen was loaded to 247 kN (27.8 tonf) and failed in the base metal. Based on a weld cross-sectional area of 5.08 cm^2 (0.787 $in.^2$), a minimum weld tensile strength was calculated as 487 MPa (70.6 ksi).

An all-weld-metal tensile test was run using a specimen removed from the bead so that its loading axis was parallel to the weld direction. In this case, the weld was aged prior to testing, with the heat treatment being 48 h at 104 °C (220 °F). The results of the tensile test are shown in Fig. 10. The yield strength was found to be 520 MPa (75.4 ksi), the ultimate strength 594 MPa (86.1 ksi), and the percent elongation 30%. With respect to a typical vendor specification for this type of weld, these values indicate that the weld is adequate.

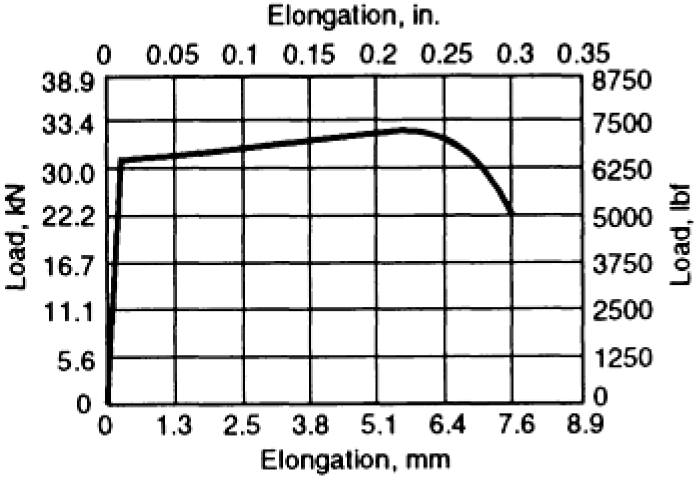

FIG. 10 PLOT OF LOAD VERSUS ELONGATION FOR ALL-WELD-METAL TENSILE TEST OF A 6.4 MM (0.25 IN.) DIAMETER SPECIMEN TAKEN FROM THE 1.07 M (42 IN.) DIAMETER X-65 STEEL PIPE

Charpy V-notch impact specimens were removed from the weld as shown in Fig. 11. For this type of application, tests are usually run at only one temperature, dictated by a specification and/or past experience. For a full characterization, however, tests will be run over a broad enough range to clearly establish the upper- and lower-shelf energy levels and to identify the ductile-to-brittle transition temperature (DBTT). For this example, tests were run at several temperatures, and the results are shown in Fig. 12. This figure clearly shows that the upper shelf energy is approximately 200 J (150 ft · lbf), the lower shelf 2 J (1.5 ft · lbf), and the DBTT -10 °C (14 °F).

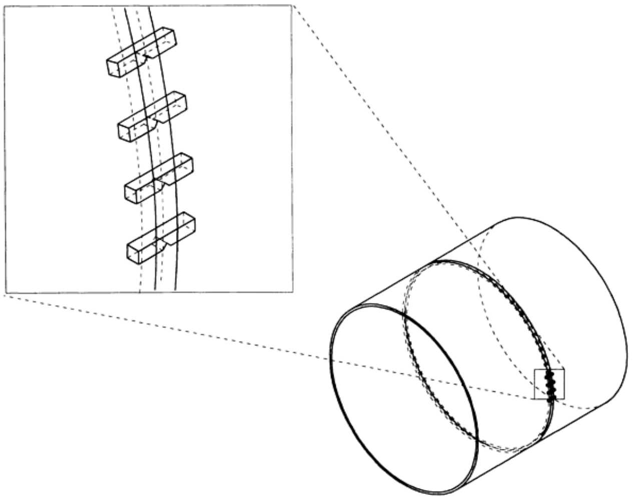

FIG. 11 SCHEMATIC SHOWING LOCATION AND ORIENTATION OF CHARPY V-NOTCH SPECIMENS TAKEN FROM THE 1.07 M (42 IN.) DIAMETER X-65 STEEL PIPE. ALTHOUGH INSET SHOWS WIDELY SPACED SAMPLES FOR THE SAKE OF CLARITY, ACTUAL SAMPLES REMOVED ARE TYPICALLY ADJACENT TO ONE ANOTHER.

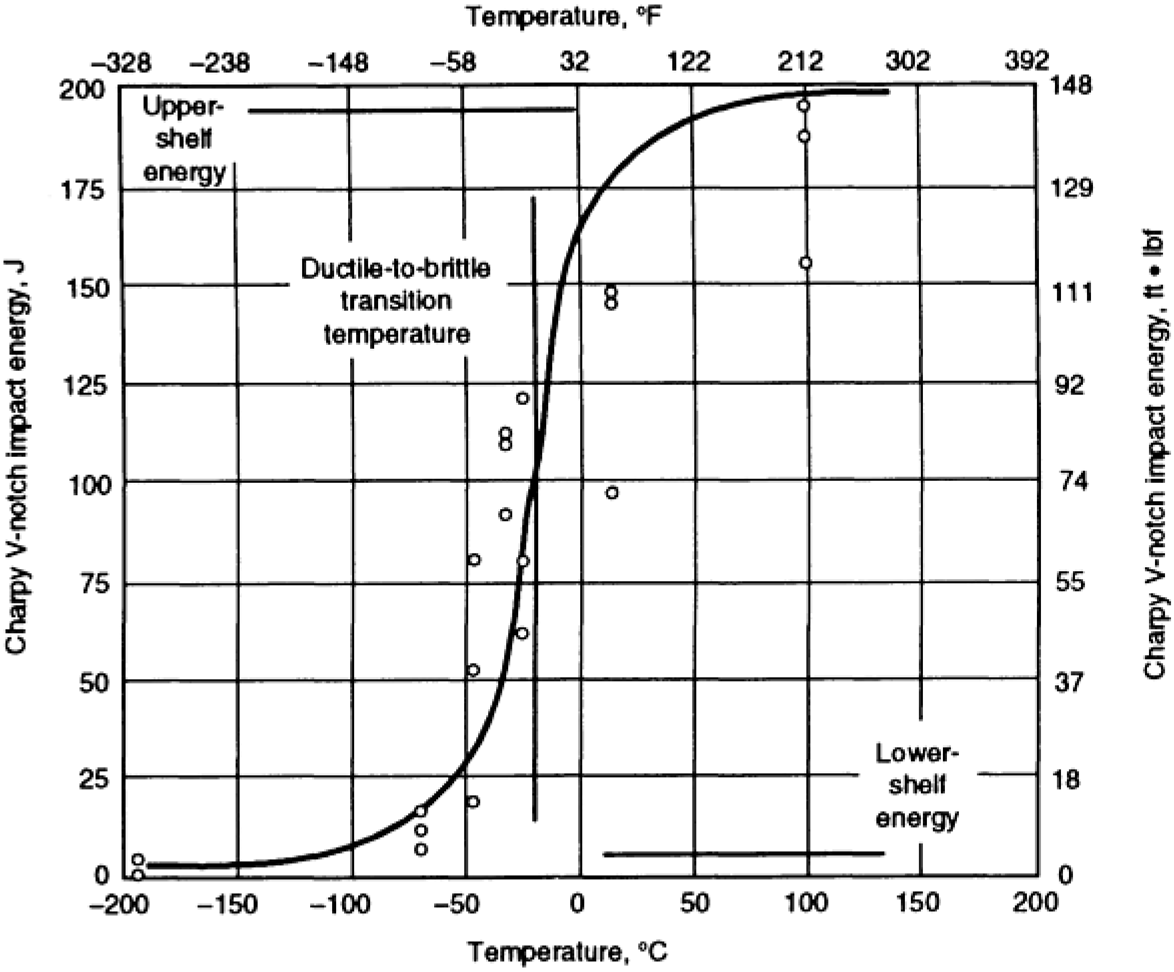

FIG. 12 CHARPY V-NOTCH IMPACT ENERGY TEST RESULTS USED TO OBTAIN DUCTILE-TO-BRITTLE TRANSITION TEMPERATURE (-10 °C, OR 14 °F) FOR THE 1.07 M (42 IN.) DIAMETER X-65 STEEL PIPE

Fracture toughness specimens were removed from the weld in an orientation similar to that of the Charpy specimens. As with Charpy tests, fracture toughness specifications usually require a certain value at a given temperature, usually the minimum service temperature. Crack tip opening displacement tests were conducted at -45 °C (-50 °F) on three specimens removed from the example weld. The results were found to vary from 0.137 to 0.322 mm (0.005 to 0.013 in.), with the average being 0.229 mm (0.009 in.). Figure 13 shows the fracture surface of one of the test specimens, on which the ductile fracture mode and overload zone can be clearly seen.

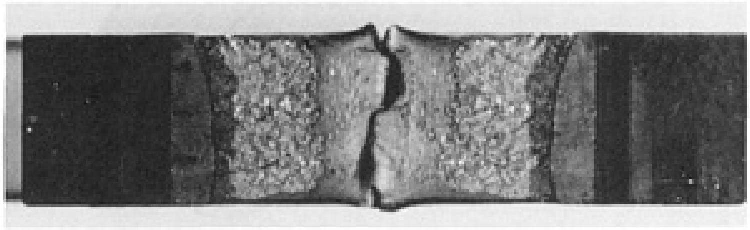

FIG. 13 MACROGRAPH OF THE FRACTURE SURFACE OF A CRACK TIP OPENING DISPLACEMENT TEST SPECIMEN REMOVED FROM THE 1.07 M (42 IN.) X-65 STEEL PIPE. THE REGIONS OF THE SURFACE SHOWN CORRESPOND WITH THE NOTCH, PRECRACK, STABLE CRACK GROWTH, FAST FRACTURE, AND OVERLOAD ZONES.

In addition to the previous mechanical tests, microhardness testing is often used for this type of weld. The primary purpose of microhardness testing is to check for the presence of martensite in the HAZ. Figure 14 shows a schematic of the example weld and the location and values of several microhardness tests. In this case, none of the values exceeded 240 HV, the level associated with the presence of martensite.

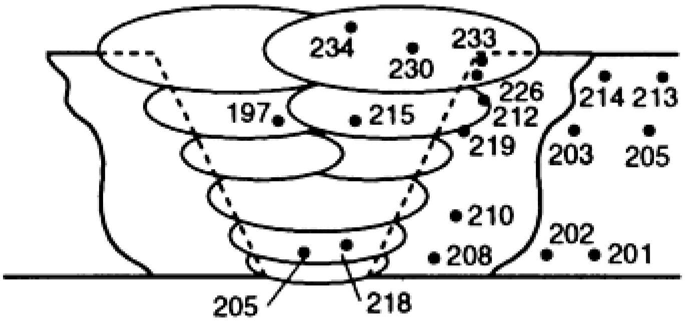

FIG. 14 MICROHARDNESS READINGS (LOCATION IN MULTI-PASS WELD INDICATED BY DOTS) BELOW 240 HV INDICATING THAT NO MARTENSITE STRUCTURE IS PRESENT IN THE 1.07 M (42 IN.) X-65 STEEL PIPE

Typically, this type of pipe weld is characterized with respect to stress-corrosion cracking resistance. In addition, testing is often performed to characterize the HAZ toughness.

Example 2: Welding of 2.5 mm (0.100 in.) Thick Ti-6Al-4V Sheet using a 0.127 mm (0.005 in.) Thick Tantalum Shim.

Figure 15 shows a transverse section of a butt weld, where the base metal is 2.5 mm (0.100 in.) thick Ti-6Al-4V and a 0.127 mm (0.005 in.) thick tantalum shim has been placed in the joint. This weld geometry was used in a study of the effect of dilution on the microstructures and properties of dissimilar metal welds of titanium to refractory metals. The questions to be answered in the characterization of this weld were:

- HOW UNIFORMLY WOULD THE TANTALUM BE DISTRIBUTED IN THE WELD POOL?
- HOW ACCURATELY COULD THE COMPOSITION OF THE WELD POOL BE PREDICTED BASED ON DILUTION CALCULATIONS?
- WHAT PHASES AND MICROCONSTITUENTS ARE PRESENT IN THE WELD METAL?
- WHAT ARE THE MECHANICAL PROPERTIES OF THE WELD METAL?

In addition, a number of very application-specific corrosion and elevated-temperature tests were performed on the dissimilar metal welds. In contrast with the pipe weld of the first example, the primary tools used in this characterization were macrostructural examination, compositional mapping, x-ray diffraction, tension testing of all weld metal specimens, and microhardness traverses.

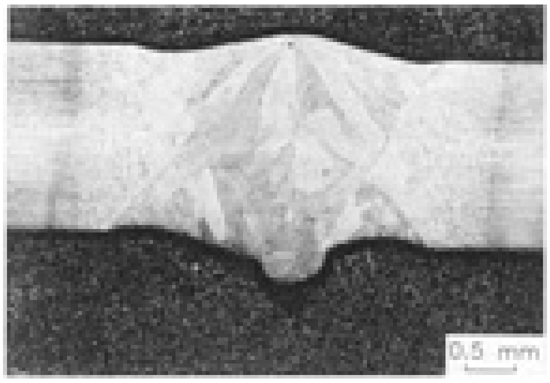

FIG. 15 MICROGRAPH OF TRANSVERSE SECTION OF AN ELECTRON-BEAM WELDED BUTT WELD JOINING 2.5 MM (0.100 IN.) THICK TI-6AL-4V SHEET USING A 0.127 MM (0.005 IN.) THICK TANTALUM SHIM PLACED IN THE JOINT. KROLL'S REAGENT WAS USED AS ETCHANT.

Weld Macrostructure. The weld shown in Fig. 15 was made with the electron-beam welding process, using sharp focus, 100 kV power, 16 to 25 mA current, and a travel speed of 12.7 mm/s (30 in./min). As can be seen in the transverse section, a full penetration weld was achieved using one pass from each side of the weld. The transverse section also shows that no obvious discontinuities exist in the weld pool and that the weld metal microstructure is fairly uniform. Figure 15 also suggests that the tantalum shim was completely melted.

Weld Metal Composition. An initial examination of the weld metal composition was performed to help assess the uniformity of the weld macrostructure. Figure 16 shows tantalum, titanium, vanadium, and aluminum elemental scan lines traversing the weld. These data were generated using a scanning electron microscope with energy-dispersive spectroscopy capability, and they confirm that the elements in question are uniformly distributed throughout the weld pool. Based on these analyses and the macrostructural characterization, it was determined that this combination of welding processes and parameters was successful in producing a full-penetration weld and a macroscopically homogeneous and defect-free weld pool.

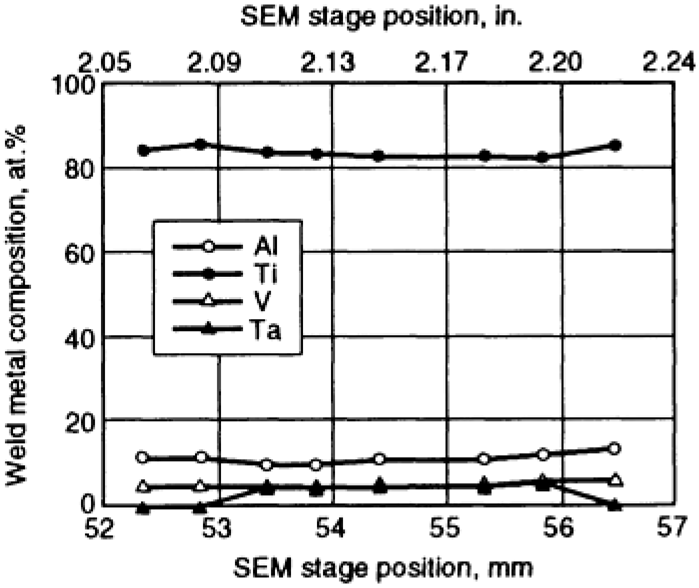

FIG. 16 ELEMENTAL LINE SCANS OF TITANIUM, ALUMINUM, VANADIUM, AND TANTALUM SPANNING THE HAZ AND FUSION ZONE OF THE 2.5 MM (0.100 IN.) THICK TI-6AL-4V SHEET WELDED USING A 0.127 MM (0.005 IN.) THICK TANTALUM SHIM.

The properties of titanium alloys can be strongly affected by the phases present in the microstructure, which are largely controlled by the weld metal composition. In this example, aluminum stabilizes the hexagonal close-packed (hcp) α and/or α' phases, and vanadium and tantalum, the body-centered cubic (bcc) β phase. In titanium alloys, the composition can be expressed as an electron/atom (*e*/*a*) ratio, which allows the phase stabilization effects of different elements to be combined. Table 6 lists the calculated (based on dilution) and measured compositions of the weld. This table shows that the measured and calculated compositions agree reasonably well. The measured aluminum values are somewhat lower than the calculations predict, perhaps due to some loss of aluminum during the welding process. Of most interest is that the measured and calculated *e*/*a* ratios are 3.96 to 3.98. This compositional range indicates that the weld metal microstructure should be α', a hexagonal martensite structure associated with good mechanical properties (Ref 20).

TABLE 6 COMPOSITION OF TI-6AL-4V RELATIVE TO THAT OF TI-6AL-4V PLUS TANTALUM WELD METAL

SAMPLE	COMPOSITION, AT.%				
	Ti	V	Al	Ta	*E/A* RATIO
TI-6AL-4V	86.2	3.6	10.2	. . .	3.93
TI-6AL-4V WELDED TO TA SHIM:					
CALCULATED	83.9	3.5	9.9	2.7	3.96
MEASURED	83.0	3.9	9.5	3.6	3.98

Weld Microstructure. The microstructure of the weld was analyzed using optical microscopy, transmission electron microscopy, and x-ray diffraction. Of particular interest were the structure of any martensitic phase present and the possible existence of ω, an ordered phase that can severely embrittle the material. The composition of the weld metal suggests that the structure will be α' hexagonal martensite. Figure 17 shows an optical micrograph of the fusion zone. Examination of Fig. 17 and comparison with Ti-6Al-4V (no tantalum shim) welds further suggest that the weld metal microstructure is α'. This was further supported by x-ray diffraction and TEM data. Figure 18 shows an x-ray diffraction trace for the weld metal. Aside from peaks from the specimen-mounting material, only peaks associated with a hexagonal structure were detected.

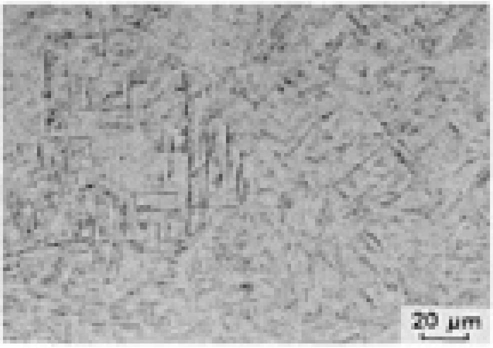

FIG. 17 OPTICAL MICROGRAPH OF FUSION ZONE OBTAINED WITH THE TI-6AL-4V SHEET WELDED USING A TANTALUM SHIM. KROLL'S REAGENT WAS USED AS THE ETCHANT. MICROSTRUCTURE CONSISTS OF WIDMANSTÄTTEN α+ β AND α' MARTENSITE.

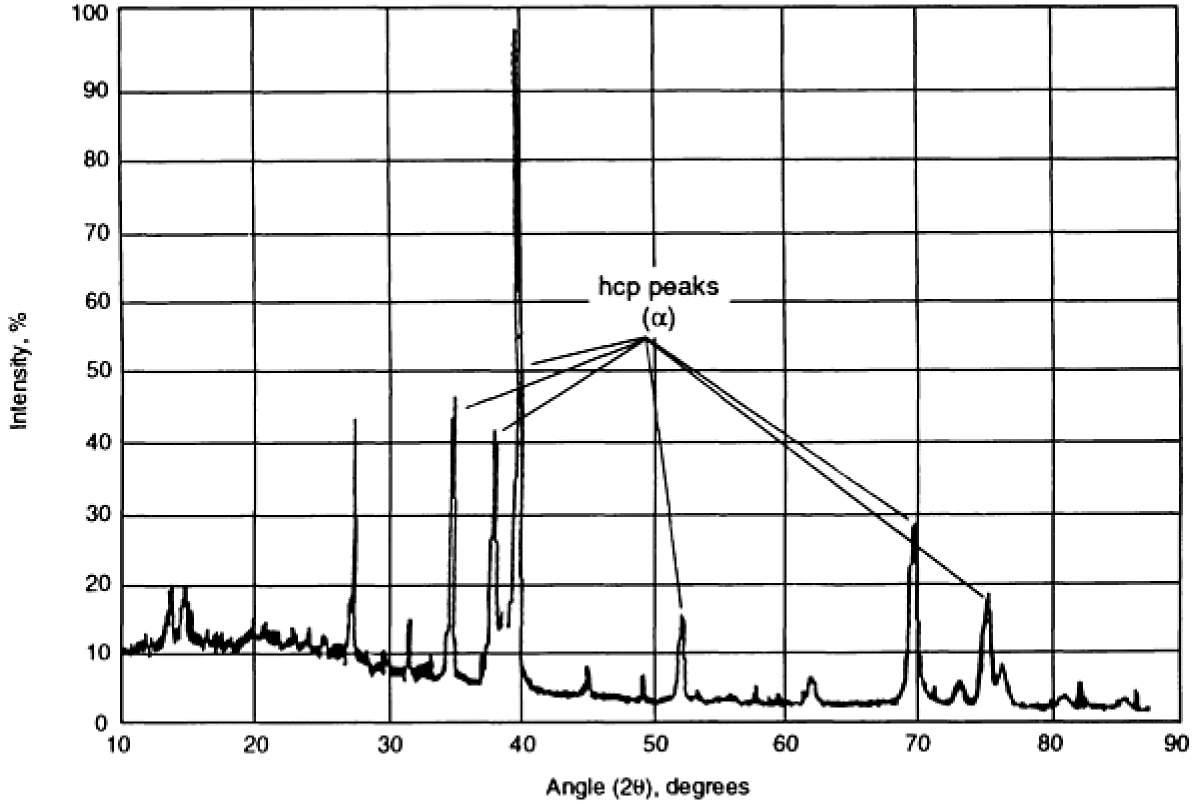

FIG. 18 X-RAY DIFFRACTION TRACE OF FUSION ZONE IN TI-6AL-4V SHEET WELDED USING A TANTALUM SHIM, SHOWING PEAKS THAT IDENTIFY HEXAGONAL STRUCTURES

Based on the compositional analysis, diffraction results, and microstructural analysis, the weld metal can be characterized as α' hexagonal martensite. By comparison with Ti-6Al-4V base metal and (no shim) weld metal, it was determined that the weld metal could be treated as a slightly more heavily stabilized variant of Ti-6Al-4V. This analysis allowed its properties and performance to be reasonably predicted.

Mechanical Testing. Microhardness traverses and all-weld-metal tension tests were performed on this weld. Figure 19 shows the results of two microhardness traverses, one across the second pass only, and one across the middle of the weld where the two passes overlap. Both traces are very similar, and the hardness is fairly constant across the fusion zone. These data confirm that the weld metal is relatively homogeneous. The hardness of the fusion zone is slightly higher than that of the base metal, as would be expected because of the martensitic structure and slightly higher β-stabilizer content.

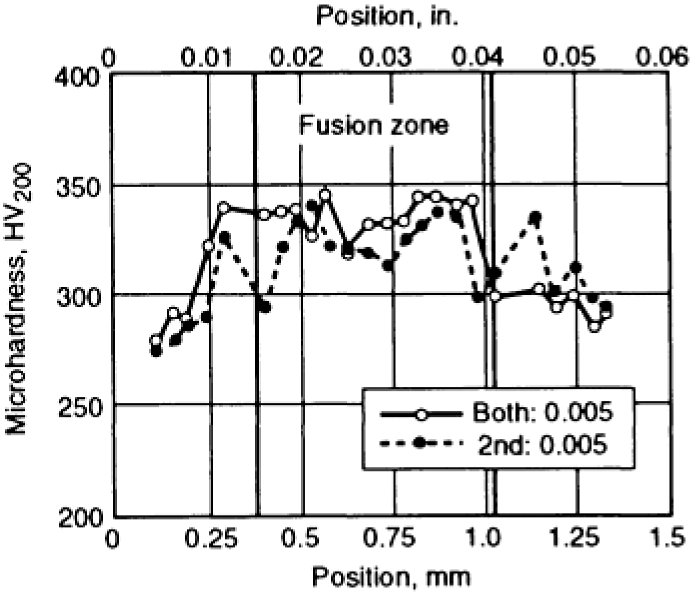

FIG. 19 MICROHARDNESS TRAVERSE DATA IN TWO LOCATIONS FOR THE TI-6AL-4V SHEET WELDED USING A TANTALUM SHIM. DATA LABELED "2ND" REFER TO TRAVERSE OVER SECOND PASS OF WELD ONLY. DATA LABELED "BOTH" INDICATE TRAVERSE ACROSS CENTER OF WELD WHERE TWO PASSES OVERLAP (SEE FIG. 15).

The tensile testing further confirmed that the weld metal could be considered to be a more heavily stabilized variant of Ti-6Al-4V. Table 7 compares the results of tensile tests performed on a baseline Ti-6Al-4V (no shim) weld with those of tests performed on the example weld. As the data show, the tantalum-containing weld had slightly higher strength and lower ductility than the baseline weld. Again these data are consistent with a slightly higher β-stabilizer content. However, the differences were not as great as in the microhardness testing, because in the case of the tensile tests, both structures were α' martensite.

TABLE 7 TENSILE PROPERTIES OF BASELINE TI-6AL-4V WELD COMPARED TO TI-6AL-4V WELDED USING TANTALUM SHIM

SAMPLE	YIELD STRENGTH		ULTIMATE TENSILE STRENGTH		YOUNG'S MODULUS		MAXIMUM STRAIN, %
	MPA	KSI	MPA	KSI	GPA	10^6 PSI	
BASELINE 1	828	120	970	141	114	16.5	12.7
BASELINE 2	787	114	960	139	127	18.4	12.3
BASELINE AVERAGE	807	117	965	140	121	17.5	12.5
TA WELD 1(A)	839	121.7	983	143	105	15.2	4.7
TA WELD 2(A)	838	121.5	971	141	103	14.9	10.5
TA WELD 3(A)	837	121.4	981	142	107	15.5	14.3
TA WELD AVERAGE	838	121.5	978	142	105	15.2	9.8

(A) REPLICANTS OF TI-6AL-4V WELDED TO TANTALUM SHIM

Test Results. The goal of the Ti-6Al-4V/tantalum dissimilar metal weld study was to thoroughly characterize the homogeneity, structure, and properties of the weld, relative to a baseline Ti-6Al-4V weld, and to determine the effect of the tantalum additions. The study showed that the tantalum shim could be completely melted using the chosen welding procedure and that a uniform weld bead could be produced. The compositional and microstructural analyses showed that the weld was very similar to the baseline weld. The tantalum-alloyed weld metal appeared to be consistent with standard principles regarding titanium alloy compositions and structure, and it was what would be expected in a weld made on slightly more heavily alloyed material. The mechanical test results confirmed this interpretation as well. In total, the characterization showed that this weld could be made and that its properties could be predicted and/or interpreted accurately as an extrapolation of the baseline weld.

Characterization Objectives. A number of characterization techniques have been briefly discussed with respect to their application to welds. Examples 1 and 2 show that the choice and emphasis of the characterization procedure are largely determined by the intended application of the weld and the goals to be accomplished by the characterization. Each of the procedures in Examples 1 and 2 are discussed in great detail in the cited references. Additional information is available in the Section "Joint Evaluation and Quality Control" in this Volume and in Ref 16, 17, and 22

References cited in this section

16. *MATERIALS CHARACTERIZATION,* 9TH ED., VOL 10, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1986
17. *MECHANICAL TESTING,* 9TH ED., VOL 8, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1985
20. STANDARD TEST METHOD FOR PLANE-STRAIN FRACTURE TOUGHNESS OF METALLIC MATERIALS, ASTM E 399, *ANNUAL BOOK OF ASTM STANDARDS,* ASTM, 1984
22. *WELDING, BRAZING, AND SOLDERING,* 9TH ED., VOL 6, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1983

References

1. STANDARD WELDING TERMS AND DEFINITIONS, ANSI/AWS A3.0-89, AWS, 1989
2. WELD DISCONTINUITIES, *METALS HANDBOOK,* 9TH ED., VOL 6, AMERICAN SOCIETY FOR METALS, 1983, P 829-855
3. *WELDING HANDBOOK,* 8TH ED., VOL 2, AWS, 1987
4. *PROCEDURE HANDBOOK OF ARC WELDING,* 12TH ED., LINCOLN ELECTRIC CO., 1973
5. CODES, STANDARDS, AND INSPECTION, *METALS HANDBOOK,* 9TH ED., VOL 6, AMERICAN SOCIETY FOR METALS, 1983, P 823-838
6. *RADIOGRAPHY IN MODERN INDUSTRY,* 4TH ED., EASTMAN KODAK CO., 1980
7. WELDING OF PIPELINES AND RELATED FACILITIES, API STANDARD 1104, 17TH ED., AMERICAN PETROLEUM INSTITUTE, 1988
8. REFERENCE RADIOGRAPHS FOR STEEL WELDMENTS, ASTM E 390, ASTM
9. *STRUCTURAL WELDING CODE--STEEL,* ANSI/AWS D1.1-92, AWS, 1992
10. METALLOGRAPHY OF WELDMENTS, *METALS HANDBOOK,* 9TH ED., VOL 9, AMERICAN SOCIETY FOR METALS, 1985, P 577-586
11. K.E. EASTERLING, MICROSTRUCTURE AND PROPERTIES OF THE HAZ, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* S. DAVID AND J. VITEK, ED., ASM INTERNATIONAL, 1990, P 177-188
12. WRC 1992 DIAGRAM, *WELD. J.,* MAY 1992, P 171
13. T.G.F. GRAY, J. SPENCE, AND T.H. NORTH, *RATIONAL WELDING DESIGN,* NEWNES-BUTTERWORTHS, 1975, P 111
14. R.H. ESPY, WELDABILITY OF NITROGEN-STRENGTHENED STAINLESS STEELS, *WELD. J. RES. SUPPL.,* VOL 61 (NO. 5), MAY 1982, P 149S-156S

15. METAL TEST METHODS AND ANALYTICAL PROCEDURES, *1991 ANNUAL BOOK OF ASTM STANDARDS,* VOL 03.05 AND 03.06, ASTM, 1991
16. *MATERIALS CHARACTERIZATION,* 9TH ED., VOL 10, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1986
17. *MECHANICAL TESTING,* 9TH ED., VOL 8, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1985
18. STANDARD METHODS OF TENSION TESTING OF METALLIC MATERIALS, ASTM E 8, *ANNUAL BOOK OF ASTM STANDARDS,* VOL 03.01, ASTM, 1984
19. *SPECIFICATION FOR CARBON STEEL ELECTRODES FOR SHIELDED METAL ARC WELDING,* ANSI/AWS A5.1-91, AWS, 1991
20. STANDARD TEST METHOD FOR PLANE-STRAIN FRACTURE TOUGHNESS OF METALLIC MATERIALS, ASTM E 399, *ANNUAL BOOK OF ASTM STANDARDS,* ASTM, 1984
21. STANDARD TEST METHOD FOR CRACK TIP OPENING DISPLACEMENT (CTOD) FRACTURE TOUGHNESS MEASUREMENT, ASTM E 1290, VOL 03.01, ASTM, 1989
22. *WELDING, BRAZING, AND SOLDERING,* 9TH ED., VOL 6, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1983

Introduction to Brazing and Soldering

Melvin M. Schwartz, Sikorsky Aircraft

Introduction

BRAZING AND SOLDERING require the application of a number of scientific and engineering skills to produce joints of satisfactory quality and reliability. Brazing employs higher temperatures than soldering, but the fundamental concepts are similar, particularly with respect to metallurgy and surface chemistry (Table 1). However, joint design, materials to be joined, filler metal and flux selection, heating methods, and joint preparation can vary widely between the two processes. Economic considerations involving filler metal and process technology are also varied, particularly in relation to automated techniques and inspection and testing. Brazing and soldering are performed in many industries, from exotic applications in the electronics and aerospace field to everyday plumbing applications.

TABLE 1 COMPARISON OF SOLDERING, BRAZING, AND WELDING

Parameter	**Process**		
	Soldering	**Brazing**	**Welding**
Joint formed	Mechanical	Metallurgical	Metallurgical
Filler metal melt temperature, °c (°f)	<450 (<840)	>450 (>840)(a)	>450 (>840)(b)
Base metal	Does not melt	Does not melt	. . .
Fluxes used to protect and to assist in wetting of base-metal surfaces	Required	Optional	Optional
Typical heat sources	Soldering iron; ultrasonics; resistance; oven	Furnace; chemical reaction; induction; torch; infared	Plasma; electron beam; tungsten and submerged arc; resistance; laser
Tendency to warp or burn	Atypical	Atypical	Potential distortion and warpage of base-metal likely
Residual stresses	. . .	. . .	Likely around weld area

(A) LESS THAN MELTING POINT OF BASE METAL.

(B) LESS THAN OR EQUAL TO MELTING POINT OF BASE METAL

Introduction to Brazing and Soldering

Melvin M. Schwartz, Sikorsky Aircraft

Brazing

Many products are assembled from two or more individual components that are often permanently joined to produce structurally sound assemblies. Joining methods include various fasteners, interference-type joints, adhesives, and, for the highest-integrity joints, the many techniques classified under welding. Brazing, although fundamentally indifferent from welding, is one such technique.

Early metalworkers, stimulated by a desire to produce structures that were difficult or impossible to build using methods then in existence, realized that it was possible to fill the joint between two metal pieces with molten metal and allow it to solidify. These artisans soon learned by experience that, in order to achieve adherence, the metals to be joined and the brazing filler metal had to be kept free of oxides and the filler metal had to have a lower melting point, and, furthermore, that a given filler metal would not necessarily adhere to all metals. From these basic requirements, brazing grew into a craft whose practitioners were well versed in what to do and what not to do in order to produce sound joints.

Just as the technique of brazing developed empirically, so did the lower-melting point filler metals. Workers first used lead and tin solders as well as silver and copper-arsenic ores, which were readily available and had low melting points. Later, the alloy brass was developed and found to be more desirable for joining copper, silver, and steel structures, because it provided higher-strength joints and could withstand higher temperatures.

Early silversmiths, probably wanting to produce white solder joints for aesthetic reasons, melted brass and silver together and found it to have an even lower melting point than brass, good adherence, and better corrosion resistance. Although innumerable combinations of silver, copper, and zinc subsequently evolved, primarily to meet melting-point requirements, these silver-brass, brass, and lead-tin alloys were essentially the only brazing and soldering filler materials available for generations.

Advantages of Brazing. Brazing has many distinct advantages, including the following:

- ECONOMICAL FABRICATION OF COMPLEX AND MULTI-COMPONENT ASSEMBLIES
- SIMPLE METHOD TO OBTAIN EXTENSIVE JOINT AREA OR JOINT LENGTH
- JOINT TEMPERATURE CAPABILITY APPROACHING THAT OF BASE METAL
- EXCELLENT STRESS DISTRIBUTION AND HEAT-TRANSFER PROPERTIES
- ABILITY TO PRESERVE PROTECTIVE METAL COATING OR CLADDING
- ABILITY TO JOIN CAST MATERIALS TO WROUGHT METALS
- ABILITY TO JOIN NONMETALS TO METALS
- ABILITY TO JOIN METAL THICKNESSES THAT VARY WIDELY IN SIZE
- ABILITY TO JOIN DISSIMILAR METALS
- ABILITY TO JOIN POROUS METAL COMPONENTS
- ABILITY TO FABRICATE LARGE ASSEMBLIES IN A STRESS-FREE CONDITION
- ABILITY TO PRESERVE SPECIAL METALLURGICAL CHARACTERISTICS OF METALS
- ABILITY TO JOIN FIBER- AND DISPERSION-STRENGTHENED COMPOSITES
- CAPABILITY FOR PRECISION PRODUCTION TOLERANCE
- REPRODUCIBLE AND RELIABLE QUALITY CONTROL TECHNIQUES

Strong, uniform, leakproof joints can be made rapidly, inexpensively, and even simultaneously. Joints that are inaccessible and parts that may not be joinable at all by other methods often can be joined by brazing. Complicated assemblies comprising thick and thin sections, odd shapes, and differing wrought and cast alloys can be turned into integral components by a single trip through a brazing furnace or a dip pot. Metal as thin as 0.01 mm (0.0004 in.) and as thick as 150 mm (6 in.) can be brazed.

Brazed joint strength is high. The nature of the interatomic (metallic) bond is such that even a simple joint, when properly designed and made, will have strength equal to or greater than that of the as-brazed parent metal. The natural shapes of brazing fillets are excellent. The meniscus surface formed by the fillet metal as it curves across corners and adjoining sections is ideally shaped to resist fatigue. (It should be noted that in brazed joints using eutectic-type filler metal, fillets often contain an excessive amount of brittle intermetallic compounds. In fact, fillets are 5 to 10 times thicker than the joint and thus have a much higher volume of the liquid phase from which these brittle phases crystallize. Therefore, crack nucleation often originates in fillets.)

Complex shapes with greatly varied sections can be brazed with little distortion, and precise joining is comparatively simple. Unlike welding, in which the application of intense heat to small areas acts to move the parts out of alignment and introduces residual stresses, brazing involves fairly even heating and thus part alignment is easier.

Limitations of Brazing. A brazed joint is not a homologous body but rather is heterogeneous, composed of different phases with differing physical and chemical properties. In the simplest case, it consists of the base metal parts to be joined and the added filler metal. However, partial dissolution of the base metal, combined with diffusion processes, can change the composition and therefore the chemical and physical properties of the boundary zone formed at the interface between base metal and filler metal and often of the entire joint. Thus, in addition to the two different materials present in the simplest example given above, a complicated transitional or even completely different zone must be considered.

In determining the strength of such heterogeneous joints, the simplified concepts of elasticity and plasticity theory--valid for a homogeneous metallic body where imposed stresses are uniformly transmitted from one surface or space element to the adjacent ones--no longer apply. In a brazed joint formed of several materials with different characteristics of deformation resistance and deformation speed, the stresses caused by externally applied loads are nonuniformly distributed.

Mechanics of Brazing. Brazing involves a limited dissolution or plastic deformation of the base metal (Ref 1). Brazing comprises a group of joining processes in which coalescence is produced by heating to a suitable temperature above 450 °C (840 °F) and by using a ferrous or nonferrous filler metal that must have a liquidus temperature above 450 °C (840 °F) and below the solidus temperature(s) of the base metal(s). The filler metal is distributed between the closely fitted surfaces of the joint. Brazing is distinguished from soldering in that soldering employs a filler metal having a liquidus below 450 °C (840 °F) (Table 1).

Brazing proceeds through four distinct steps:

- THE ASSEMBLY OR THE REGION OF THE PARTS TO BE JOINED IS HEATED TO A TEMPERATURE OF AT LEAST 450 °C (840 °F).
- THE ASSEMBLED PARTS AND BRAZING FILLER METAL REACH A TEMPERATURE HIGH ENOUGH TO MELT THE FILLER METAL (FOIL, WIRE, PASTE, PLATINGS, ETC.) BUT NOT THE PARTS.
- THE MOLTEN FILLER METAL, HELD IN THE JOINT BY SURFACE TENSION, SPREADS INTO THE JOINT AND WETS THE BASE METAL SURFACES.
- THE PARTS ARE COOLED TO SOLIDITY, OR "FREEZE," THE FILLER METAL, WHICH IS HELD IN THE JOINT BY CAPILLARY ATTRACTION AND ANCHORS THE PARTS TOGETHER BY METALLURGICAL REACTION AND ATOMIC BONDING.

Brazing Versus Other Welding Processes. The mere fact that brazing does not involve any substantial melting of the base metals offers several advantages over other welding processes. It is generally possible to maintain closer assembly tolerances and to produce a cosmetically neater joint without costly secondary operations. Even more important, however, is that brazing makes it possible to join dissimilar metals (or metals to ceramics) that, because of metallurgical incompatibilities, cannot be joined by traditional fusion welding processes. If the base metals do not have to be melted to be joined, it does not matter that they have widely different melting points. Therefore, steel can be brazed to copper as easily as to another steel.

Brazing also generally produces less thermally induced distortion, or warping, than fusion welding. An entire part can be brought up to the same brazing temperature, thereby preventing the kind of localized heating that causes distortion in welding.

Finally, and perhaps most important to the manufacturing engineer, brazing readily lends itself to mass production techniques. It is relatively easy to automate, because the application of heat does not have to be localized, as in fusion welding, and the application of filler metal is less critical. In fact, given the proper clearance conditions and heat, a brazed joint tends to "make itself" and is not dependent on operator skill, as are most fusion welding processes.

Automation is also simplified by the fact that there are many means of applying heat to the joint, including torches, furnaces, induction coils, electrical resistance, and dipping. Several joints in one assembly often can be produced in one multiple-braze operation during one heating cycle, further enhancing production automation.

As noted in Table 1, essentially no melting of the base metal occurs in brazing; however, the temperatures involved can affect the properties of the metals being joined. For example, base metals whose mechanical properties were obtained by cold working may soften or undergo grain growth if the brazing temperature is above their recrystallization temperatures. Mechanical properties obtained by heat treatment may be altered by the heat of brazing. On the other hand, materials in the annealed condition are usually not altered by brazing.

As with other welding processes, brazing produces a heat-affected zone (HAZ) with a strongly altered microstructure due to intensive mutual mass transfer between base metal and filler metal. The width of this zone varies with the heating process used. In torch and induction brazing, for example, only a localized zone is heated; in france and dip brazing, the entire part is subjected to the brazing temperature. As a rule, the HAZ produced during brazing is wider and less sharply defined than those resulting from other fusion-related processes.

Several elements of the brazing process must be understood in order to produce satisfactory brazed joints:

- FILLER METAL FLOW
- BASE METAL CHARACTERISTICS
- FILLER METAL CHARACTERISTICS
- SURFACE PREPARATION
- JOINT DESIGN AND CLEARANCE
- TEMPERATURE AND TIME
- RATE AND SOURCE OF HEATING

Each of these topics is discussed in detail in the article "Fundamentals of Brazing" in this Volume.

Reference cited in this section

1. E. LIEBERMAN, *MODERN SOLDERING AND BRAZING TECHNIQUES,* BUSINESS NEWS PUBLICATIONS, 1988

Melvin M. Schwartz, Sikorsky Aircraft

Soldering

Like brazing and other joining processes, soldering involves several fields of science, including mechanics, chemistry, and metallurgy. Soldering is a simple operation (Ref 2), consisting of the relative placement of the parts to be joined, wetting the surfaces with molten solder, and allowing the solder to cool until it has solidified.

Soldering in the field of electronics is in many respects different from soldering in other branches of industry. Although the physical principles of all soldering (and brazing) processes are the same, the features specific to their use in electronics are so numerous that it is possible to speak of soldering in electronics as a separate subject.

Soldering Versus Other Joining Technologies. Soldering has several clear advantages over competitive joining techniques, such as welding or bonding with conductive adhesives:

- THE SOLDER JOINT FORMS ITSELF BY THE NATURE OF THE FLOW, WETTING, AND SUBSEQUENT CRYSTALLIZATION PROCESS, EVEN WHEN THE HEAT AND THE SOLDER ARE NOT DIRECTED PRECISELY TO THE PLACES TO BE SOLDERED. BECAUSE SOLDER DOES NOT ADHERE TO INSULATING MATERIALS, IT OFTEN CAN BE APPLIED IN EXCESS QUANTITIES, IN CONTRAST TO CONDUCTIVE ADHESIVES. THE SOLDERING TEMPERATURE IS RELATIVELY LOW, SO THERE IS NO NEED FOR THE HEAT TO BE APPLIED LOCALLY AS IN WELDING.
- SOLDERING ALLOWS CONSIDERABLE FREEDOM IN THE DIMENSIONING OF JOINTS, SO THAT IT IS POSSIBLE TO OBTAIN GOOD RESULTS EVEN IF A VARIETY OF COMPONENTS ARE USED ON THE SAME PRODUCT.
- THE SOLDERED CONNECTIONS CAN BE DISCONNECTED IF NECESSARY, THUS FACILITATING REPAIR.
- THE EQUIPMENT FOR BOTH MANUAL SOLDERING AND MACHINE SOLDERING IS RELATIVELY SIMPLE.
- THE SOLDERING PROCESS CAN BE EASILY AUTOMATED, OFFERING THE POSSIBILITY OF IN-LINE ARRANGEMENTS OF SOLDERING MACHINES WITH OTHER EQUIPMENT.

Mass soldering by wave, drag, or dip machines has been the preferred method for making high-quality, reliable connections for many decades. Despite the appearance of new connecting systems, it still retains this position. Correctly controlled, soldering is one of the least expensive methods for fabricating electrical connections. Incorrectly controlled, it can be one of the most costly processes--not because of the initial cost, but because of the many far-reaching effects of poor workmanship.

Chronology of Soldering. Soldering is not a new process (Ref 2). It probably dates to the Bronze Age, when some metalworker discovered the affinity of a tin-lead mixture for a clean copper surface. As recorded in the writings of Pliny, the Romans used 60Sn-40Pb mixture to solder their lead water pipes.

During the late 19th and early 20th centuries, a great deal of work was devoted to the development and improvement of the soldering process and to the determination of the fundamental principles involved. Examination of the books, papers, and especially the patents of the era shows that soldering as a science was clearly understood long before the start of the electronics industry. There has certainly been a great improvement in the materials and machines used in the process, but the basics have changed little since the early part of this century.

Soldering in the electronics field has undergone many dramatic changes since the first radios were built. Between the 1920s and the 1940s, all interconnections were made using the point-to-point wiring method, and the solder joints were made using soldering irons. After World War II (Ref 3, 4), the demand for consumer electronics began to increase, and by the 1950s a mass market had developed. To meet this demand, the first mass soldering technique, dip soldering, was used with boards that were the precursors of printed wiring boards (PWBs). By the late 1950s, wave soldering had been developed in England and soon thereafter reached the United States. This development was necessitated by the introduction of PWBs. Wave soldering continued to be the method of choice through the 1980s and is still widely used.

The development and maturation of the multi-layer PWB and the first major effort toward miniaturization occurred in the 1970s. The dual inline integrated circuit package replaced many discrete components. By the early 1980s, the pressure to increase component densities had grown to such a degree that a new technique for attachment was needed, and surface mounting was developed. Surface mount technology in turn required new ways to make solder joints, prompting the development of vapor phase, infrared, hot gas, and other reflow soldering techniques.

Soldering remains the attachment method of choice in the evolving electronics industry. It is the only metallurgical joining technique suitable for the mass manufacture of electronic and electrical interconnections within the temperature restrictions of electronic components and substrates.

Solder Metals. The solder metal business has changed dramatically, having moved from merely producing wire to producing bulk, preform, and paste in order to meet the needs of soldering techniques. Solder can be applied in several ways (Ref 2, 3):

- A MIXTURE OF SOLDER POWDER AND FLUX, KNOWN AS SOLDER PASTE (REF 5), CAN BE APPLIED TO THE JOINTS BY SCREENING, STENCILING, OR USING ONE OF THE MACHINES THAT APPLIES DOTS OF THE PASTE BY A SIMPLE PNEUMATIC SYSTEM. THIS METHOD IS ESPECIALLY USEFUL WHERE SURFACE-MOUNTED COMPONENTS ARE TO BE SOLDERED AND CANNOT BE PASSED THROUGH THE SOLDER WAVE.
- PARTS CAN BE PRETINNED, THAT IS, COATED WITH A LAYER OF SOLDER BY FLUXING AND DIPPING INTO A BATH OF THE MOLTEN METAL.
- SOLDER CAN BE OBTAINED IN THE SHAPE OF RINGS, WASHERS, OR TUBES, WHICH ARE PLACED ON OR ADJACENT TO THE PARTS TO BE JOINED. THESE SOLDER PREFORMS CAN BE OBTAINED WITH OR WITHOUT A FLUX COATING.

The two latter methods of providing solder for the joint are used in many areas of the electronics industry where machine soldering is not possible. Tiny solder rings are placed over long wire wrap pins on a mother board, where the gold plating of the pins in the wrapping area must not be coated with solder. Solder washers are fitted under bolts that must be soldered to a ground plane. Wires and tinned before soldering into a pretinned wiring lug. The possibilities are seemingly limitless (Ref 6).

Solder filler metals have also changed. The first electronic soldering was done with the 50Sn-50Pb solder material used by plumbers. By the 1950s, the choice of materials had switched to eutectic tin-lead solders to counter the problems created by a vibrating wave solder machine conveyor. Today a wide range of filler materials is used in the electronics industry. With the spread of electronics technology into every facet of life, the environmental and materials demands have grown substantially, requiring the use of such specialized materials as indium and gold-tin fillers.

Flux technology has also evolved (Ref 3, 5). As a result of the restraints of military and government specifications, the emphasis has been on the addition of new flux types rather than on the replacement of old ones. Extensive work has been done with organic acids, which enlarge the processing windows. Other categories include synthetic activated materials, which are non-rosin fluorocarbon solvent-removable types, and low solids fluxes for surface mounting that can be used instead of the classic fully activated rosin formulations.

Other materials, such as adhesives for the surface mounting of components intended for wave soldering, are new to soldering technology. These materials must be specially designed to avoid interference with the soldering process, yet must remain intact through the solvent action of fluxes and, often, flux removal.

Mass Soldering. With the introduction of mass soldering techniques (Ref 1, 6), the issue of solderability has become one of the major interest. Mass production processes are not as tolerant as the hand soldering process. In addition, modern components cannot withstand as much heat, as for example, a tube socket. Lower soldering temperatures and less mechanical working have made it necessary to attain the highest solderability to ensure a metallurgically sound and reliably solder joint.

Machine or mass soldering (Ref 2, 4, 6) can make thousands of joints in a few seconds, providing the electrical connections and simultaneously mechanically fastening the components. No other process has yet been developed that can so economically interconnect electronic circuitry.

Wave Soldering. From the original concept of hand dipping boards came the idea of pumping solder through a slot or nozzle to form a wave, or mound, of constantly moving, dross-free metal through which the board could be passed. This offered several advantages: the surface of the solder was always clean and did not require skimming, the pumping action maintained an even temperature, and the raised wave provided an easy form for adding a conveyor system for automatic movement of the board over the solder.

Once the idea of the solder wave had been born, it was logical to use a similar method to apply the flux, and the basic soldering machine appeared. These original machines were quite crude, consisting of individual modules with a freestanding conveyor (often added by the user) placed over them. There was little attempt to integrate the parts into one system, and the smoke and fumes were left to escape into the factory exhaust, aided by a makeshift canopy hung over the machine.

Much work was done on the development of the actual wave in an attempt to improve the speed of soldering and to reduce the incidence of shorts and icicles. The solder weir, the double wave, and the cascade all experienced brief popularity. The cascade wave, which consists of several waves in series, was one of the more interesting of these developments. It was used with an inclined conveyor and was the basis of machines built for in-house use by several manufacturers. However, it never achieved great popularity.

Another development was the use of ultrasonics to form a wave, with the intention of causing cavitation at the joint of avoiding the use of flux, or at least reducing the activity required. Some experimental machines were built, but the cost was excessive, the soldering results were no better than those produced with the conventional pumped wave, and the power of the ultrasonic energy had to be so high that the matching horns used to energize the pot were quickly eroded away. With the advent of the semiconductor, ulrasonics were found to cause occasional damage to the devices. Today the main use for ultrasonics in the soldering industry is to be found in the pots used for tinning components prior to assembly. This era also saw the introduction of oil into the solder wave. This idea worked extremely well and has been a feature of some soldering machines for many years (Fig. 1) (Ref 2, 3, 4, 5, 6, 7).

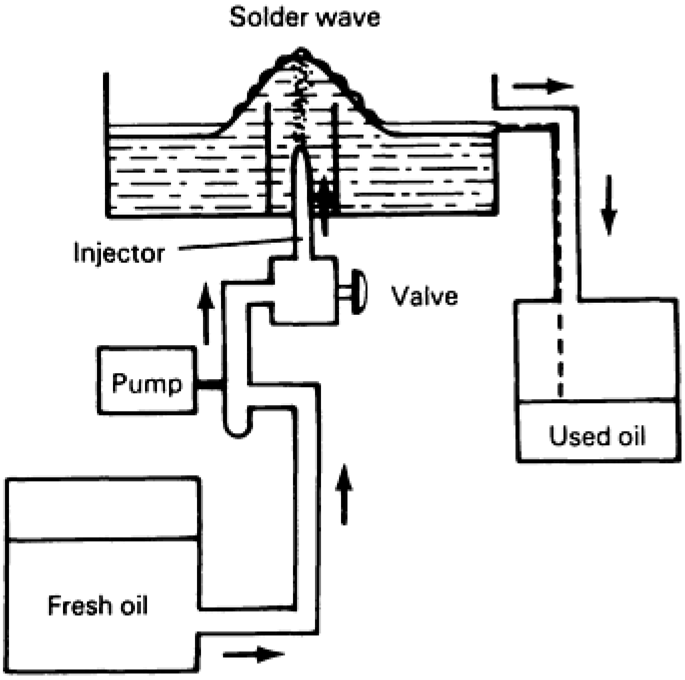

FIG. 1 SCHEMATIC SHOWING SIMPLE OIL INJECTION OF A SOLDER WAVE

Drag Soldering. In a totally different direction, the static solder bath soldering process, or drag soldering, was developed, and machines were designed using this principle (Fig. 2). These machines have been developed to a high level of sophistication, initially in Europe and Japan and later in the United States. As has often been the case in the growth of the soldering machine industry, initial developments were prompted because of patent restrictions that at the time prevented the free use of the more conventional solder wave. These restrictions have often proved beneficial to the industry by forcing creative thinking that has eventually resulted in improved products (Ref 2, 3, 4, 5, 6).

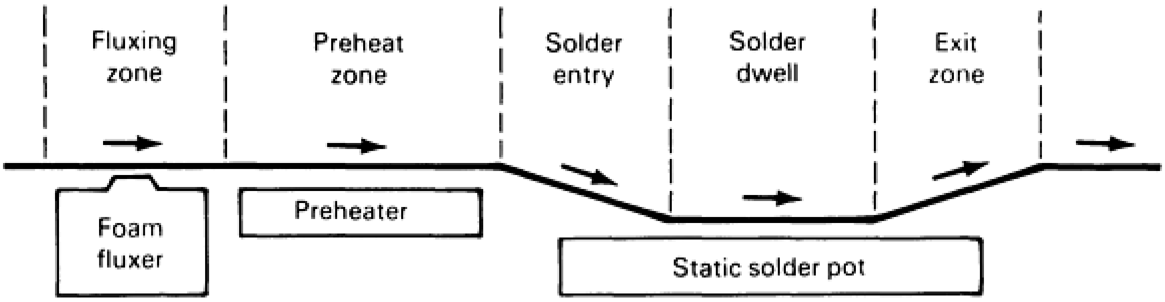

FIG. 2 SCHEMATIC SHOWING SEQUENCE OF OPERATIONS FOR A DRAG SOLDERING MACHINE

Current Technology. Today machines are available from many different manufacturers, each with its own particular features. These machines have various levels of sophistication and capability, ranging from small laboratory units to high-production systems, often complete with in-line cleaning and assembly and return conveyors. Machines in use today are generally self-contained, with a high degree of control, optional auxiliary systems, and exhaust venting. Coming onto the market are systems with computer-controlled operation and setup, which practically eliminate human error from the machine soldering process.

Soldering processes have evolved dramatically since the introduction of mass soldering techniques. Wave machines have become highly specialized, especially since the introduction of surface-mount technology. Reflow techniques (Ref 8), which were once limited to hybrid technology and leaded flatpacks, have grown remarkably. Vapor phase, infrared, and hot gas soldering equipment have initiated a new process of evolution.

The performance of a solder joint in its service environment is a strong consideration in modern electronics. Many applications, such as jet engines, dishwashers, and automobiles, impose a repeated cyclic thermal cycle on the equipment. This produces thermomechanical strains, which result in fatigue or creep failure if the joint design is poor or if life predictions have been incorrectly calculated. Each situation must be carefully analyzed and the correct design employed for the equipment to fulfill its intended service life.

Thermally induced fatigue and creep failures, virtually unknown in the 1950s but common today, are illustrative of the added stresses and strains placed on joints. It should be emphasized that miniaturization of electronic parts, particularly current paths, results in a dramatic increase in current density, which generates strong thermal and mechanical shocks upon switching. Proper design can minimize these strains and make the joints more highly producible. Design has become even more critical with the introduction of surface-mount technology.

Quality Control. Once a joint is made, its quality must be assessed through inspection. Advanced technology has produced some electronic assembly designs that are impossible to inspect visually, and an entire group of automated inspection tools has thus evolved. These tools include radiography, infrared signature, and two- or three-dimensional vision systems. Most of these are aided with very sophisticated computer software. The role of inspection, which historically was to locate those joints that needed rework, is changing along with the technology. Inspection is now used as a process-control audit to locate those portions of the process that need improvement or should returned to a position within the control limits. Therefore, the emphasis has shifted from merely shipping an acceptable product to manufacturing it correctly the first time.

Future Trends. Several other technologies and materials for soldering are being developed and tested (Ref 2, 3, 4, 5). One process that is becoming widely used is condensation, or vapor phase, soldering. In this process, the vapor from a liquid with a high-temperature boiling point is allowed to condense on the items to be soldered and, in giving up the latent heat of vaporization, quickly and evenly raises their temperature and produces a sound soldered joint. The method is simple, clean, and extremely well controlled. It requires-special equipment, and, of course, the parts are all exposed to the same temperature.

In another method, the parts are dipped into an oil heated above the melting point of the solder. Peanut oil can be used, although several excellent synthetic products are available. Here again, the component parts are subjected to the same temperature as the joint. This is a messy process, and the parts must be cleaned after soldering to remove the oil.

Infrared heating is a useful method for producing the temperature necessary for soldering. The shorter wavelengths can be focused onto very small areas, allowing selective heating of the joints to be soldered, without subjecting nearby components to the soldering temperature. Equipment and tooling are expensive if complex assemblies have to be soldered, and the relative emissivity of the joint metals and the adjacent materials can cause processing problems. For example, the base laminate of a PWB can be burned before the joint to be soldered has even reached the soldering temperature. This process thus requires very careful control.

Resistance heating can be used for soldering in two ways. The first method passes a current through the parts of the joint, causing their temperature to rise by virtue of the resistance existing in the joint structure. This is not a very controllable process and is primarily used for less critical applications on large connections. The second method passes a current through a wire that is shaped to touch the parts to be heated. Despite the lack of precise control, the process is often used to solder the leads on surface-mounted components.

Lasers have been used to provide the heat for soldering, although this is not yet a common method. A laser can provide a very high-energy, small-area beam, and heating is therefore extremely fast, localized, and limited only by the ability of the joint materials to absorb this energy. However, the shiny metal surfaces in the joint reflect most of the energy of the commonly used lasers, and coupling the laser and the joint is a major problem. The soldering flux is often used as a coupling medium. When correctly set up, laser soldering offers a fast, clean method of making joints. The cost of the equipment is quite high.

Since 1980, significant improvements and innovations in the soldering arena with respect to materials, equipment, and processes have occurred. However, two main tasks remain: developing environmentally friendly solder materials and establishing failure-free solder joint technology (Ref 2, 3, 4, 6, 7, 8).

For solders and soldering, two issues that are directly related to environmental concerns are the replacement of chlorofluorocarbon (CFCs) as the primary cleaning agent for electronic assemblies and the feasibility of using alternatives to lead-containing solders. For electronic assembly solders (Ref 4), in addition to using CFC substitutes and other solvents (e.g., hydrocarbon-based systems), aqueous cleaning and "no-clean" setups are two viable solutions.

The other task to be accomplished is the development of failure-free solder joint technology. Strengthened solder materials are expected to possess the properties that impede the occurrence of undesirable metallurgical phenomena during their service life under various in-circuit and external conditions, particularly temperature and induced stresses. Three approaches are being examined: macroscopic blends of selected fillers in a tin-rich matrix (composite solder), microscopic incorporation of doping additives to tin-rich base material, and atomic-level interaction. Preliminary experiments have been carried out to accomplish the first approach. Figure 3 demonstrates the increased yield strength of the 63Sn-37Pb composite solder with various amounts of filler (Ref 9).

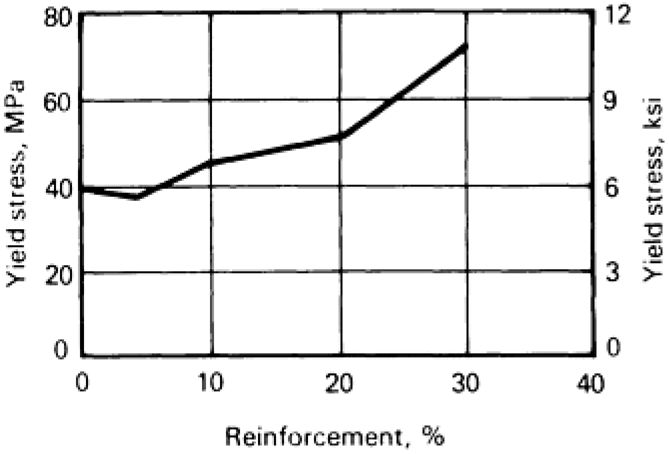

FIG. 3 EFFECT OF FILLER REINFORCEMENT ON THE YIELD STRENGTH OF 63SN-37PB COMPOSITE SOLDER. SOURCE: REF 9

The intrinsic properties of the new materials are expected to possess the following characteristics, while maintaining thermal and electrical conductivities:

- STABLE MICROSTRUCTURE
- INCREASED THERMAL AND MECHANICAL FATIGUE RESISTANCE
- HIGH CREEP RESISTANCE
- DECREASED THERMAL EXPANSION COEFFICIENT

These aggregate characteristics constitute the "strengthened" solder materials that are able to withstand harsher application conditions. The reduced thermal expansion coefficient of these new solders is expected to be more compatible with that of state-of-the-art electronic components and with ceramic substrates.

With respect to quality, solder joints have historically been judged by visual criteria. However, due to the nature of solder materials and the effect of the solder joint-forming process on appearance, many serviceable solder joints are often rejected, while other potentially questionable joints are accepted to carry out critical missions. This not only adds to manufacturing costs, but can also jeopardize service performance. Strengthened solder joints are expected to relieve the necessity of on-line inspection and rework while also reducing manufacturing costs and enhancing life span and tolerance of harsh service conditions in many military applications.

References cited in this section

1. E. LIEBERMAN, *MODERN SOLDERING AND BRAZING TECHNIQUES,* BUSINESS NEWS PUBLICATIONS, 1988
2. R.J. KLEIN WASSINK, *SOLDERING IN ELECTRONICS,* ELECTROCHEMICAL PUBLICATIONS, 1989
3. H.H. MANKO, *SOLDERS AND SOLDERING,* 2ND ED., MCGRAW-HILL 1964
4. L.P. LAMBERT, *SOLDERING FOR ELECTRONIC ASSEMBLIES,* MARCEL DEKKER, 1987
5. H.H. MANKO, SOLDERING FLUXES--PAST AND PRESENT *WELD. J.,* VOL 52 (NO. 3), 1973, P 163-166
6. R.W. WOODGATE, *THE HANDBOOK OF MACHINE SOLDERING,* 2ND ED., JOHN WILEY & SONS, 1988
7. T.J. NEILLO, "AN APPROACH TO SOLDERING IN THE AUTOMATED FACTORY," SME SOLDERING TECH. CONF., SME MS 91-288, MAY 1991
8. T. ROBERTS, "CONDUCTIVE REFLOW SOLDERING," SME SOLDERING TECH. CONF., SME MS 91-290, MAY 1991
9. J.S. HWANG, *ET AL.,* FUTURISTIC SOLDERS, *SURF. MOUNT. TECHNOL.,* SEPT 1991, P 40-43

Introduction to Brazing and Soldering

Melvin M. Schwartz, Sikorsky Aircraft

References

1. E. LIEBERMAN, *MODERN SOLDERING AND BRAZING TECHNIQUES,* BUSINESS NEWS PUBLICATIONS, 1988
2. R.J. KLEIN WASSINK, *SOLDERING IN ELECTRONICS,* ELECTROCHEMICAL PUBLICATIONS, 1989
3. H.H. MANKO, *SOLDERS AND SOLDERING,* 2ND ED., MCGRAW-HILL 1964
4. L.P. LAMBERT, *SOLDERING FOR ELECTRONIC ASSEMBLIES,* MARCEL DEKKER, 1987
5. H.H. MANKO, SOLDERING FLUXES--PAST AND PRESENT *WELD. J.,* VOL 52 (NO. 3), 1973, P 163-166
6. R.W. WOODGATE, *THE HANDBOOK OF MACHINE SOLDERING,* 2ND ED., JOHN WILEY & SONS,

1988

7. T.J. NEILLO, "AN APPROACH TO SOLDERING IN THE AUTOMATED FACTORY," SME SOLDERING TECH. CONF., SME MS 91-288, MAY 1991
8. T. ROBERTS, "CONDUCTIVE REFLOW SOLDERING," SME SOLDERING TECH. CONF., SME MS 91-290, MAY 1991
9. J.S. HWANG, *ET AL.*, FUTURISTIC SOLDERS, *SURF. MOUNT. TECHNOL.*, SEPT 1991, P 40-43

Fundamentals of Brazing

Mel M. Schwartz, Sikorsky Aircraft

Introduction

BRAZING is a process for joining solid metals in close proximity by introducing a liquid metal that melts above 450 °C (840 °F). A sound brazed joint generally results when an appropriate filler alloy is selected, the parent metal surfaces are clean and remain clean during heating to the flow temperature of the brazing alloy, and a suitable joint design is used. Like the other joining processes, brazing encompasses a variety of scientific disciplines (for example, mechanics, physics, and chemistry).

Recently, the demands of more-sophisticated structures have forced technicians and engineers to encourage metal producers to apply their metallurgical knowledge to produce brazing filler metals that meet more-specific needs. To ensure the production of good brazed joints, the technicians and engineers also had to appeal to mechanical engineers for improved joint design, to chemical engineers for solutions to corrosion problems, and to metallurgists and ceramists for proper material selection.

Brazing has been embraced by the engineering community and has now reached a very successful plateau within the joining field. This has come about because of the:

- DEVELOPMENT OF NEW TYPES OF BRAZING FILLER METALS (RAPID SOLIDIFICATION AMORPHOUS FOILS AND TITANIUM-ADDED FILLER METALS FOR CERAMIC JOINING
- AVAILABILITY OF NEW FORMS AND SHAPES OF FILLER METALS
- INTRODUCTION OF AUTOMATION THAT HAS BROUGHT BRAZING PROCESSES TO THE FOREFRONT IN HIGH-PRODUCTION SITUATIONS
- INCREASED USE OF FURNACE BRAZING IN A VACUUM, AS WELL AS ACTIVE AND INERT-GAS ATMOSPHERES

Since the early 1980s, other developments, such as aluminum-clad foils for fluxless aluminum brazing, copper-nickel-titanium filler metals for brazing titanium and some of its alloys, cadmium-free silver filler metals, and vacuum-grade metal brazing foils, have evolved to be used in production applications.

Physical Principles

Capillary flow is the dominant physical phenomenon that ensures good brazements when both faying surfaces to be joined are wet by the molten filter metal. The joint must be properly spaced to permit efficient capillary action and coalescence. More specifically, capillarity is a result of the relative attraction of the molecules of the liquid to each other and to those of the solid. In actual practice, brazing filler metal flow characteristics are also influenced by dynamic considerations involving fluidity, viscosity, vapor pressure, gravity, and, especially, by the effects of any metallurgical reasons between the filler metal and the base metal.

Capillary attraction makes the brazing of leak-tight joints a simple proposition. In a properly designed joint, the molten brazing filler metal is normally drawn completely through the joint area without any voids gaps when processed in a protective atmosphere. Solidified joints will remain intact and gas will remain tight under heavy pressures, even when the joint is subjected to shock or vibrational type of loading. Capillary attraction is also the physical force that governs the action of a liquid against solid surfaces in small, confined areas.

The phenomena of wetting and spreading are very important to the formation of brazed joints. Other significant factors that also must be considered include the condition of the solid surface in terms of the presence of oxide films and their effects on wetting and spreading, surface roughness, alloying with between the brazing filler metal and base metal, and the extent to which alloying is affected by the thermodynamic properties of the brazing atmosphere. A number of studies (Ref 1, 2) have been conducted on surface activation, contact angle, equilibrium, and surface energies. Some of these alloy systems have a finite contact angle that is thermodynamically unstable, because the solid-vapor surface energy exceeds the sum of the liquid-solid surface energies, that is, an alloy system in which thermodynamics would predict complete spreading. In actual fact, spreading may or may not occur in this type of alloy system, and the rate of spreading can be markedly dependent on surface chemistry.

Wetting is, perhaps, best explained by the following example. If a solid is immersed in a liquid bath and wetting occurs, then a thin continuous layer of liquid adheres to the solid when it is removed from the liquid (Fig. 1). Technically, the force of adhesion between the solid and liquid (during the wetting process) is greater than the cohesive force of the liquid. In practical terms, with respect to brazing, wetting implies that the liquid brazing filler metal spreads on the solid base metal, instead of balling up on its surface. It has been demonstrated that wetting actually depends on a slight surface alloying of the base metal with the brazing filler metal. Lead, for example, does not alloy with iron and will not wet it. Tin, on the other hand, does form an alloy with iron, and, therefore, a tin-lead solder will wet steel.

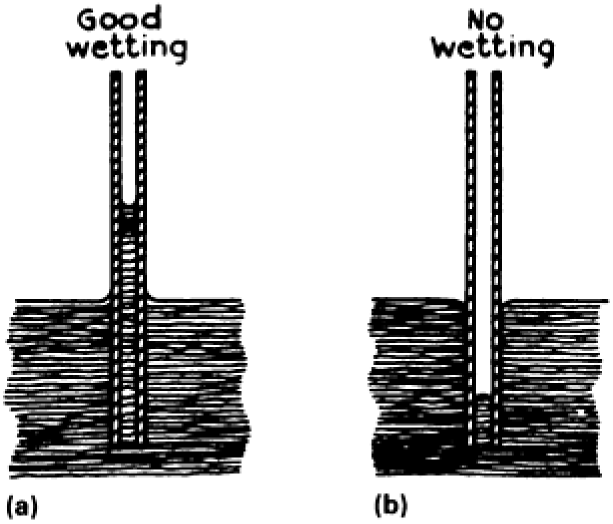

FIG. 1 SCHEMATIC SHOWING PRINCIPLE OF CAPILLARY ATTRACTION FOR SELECTED LIQUIDS WHEN THE LIQUID IS SANDWICHED BETWEEN TWO CLEAN GLASS PLATES. (A) WHEN IMMERSED IN WATER OR INK, A COLUMN WILL RISE BETWEEN THE PLATES BECAUSE OF WETTING. (B) WHEN THE PLATES ARE IMMERSED IN MERCURY, NO WETTING OCCURS AND THE COLUMN IS DEPRESSED. IF PARAFFIN PLATES ARE USED IN PLACE OF THE GLASS PLATES, IMMERSION OF THE PLATES IN MERCURY WILL PRODUCE WETTING.

A comprehensive theory of the wetting or spreading of liquids on solid surfaces and a complete and detailed derivation of the quantitative relationships are presented by Harkins (Ref 3) and Schwartz (Ref 4).

Experimentally, it has been observed that liquids placed on solid surfaces usually do not completely wet, but, rather, remain as a drop that has a definite contact angle between the liquid and solid phases (Ref 5). This condition is illustrated in Fig. 2. The Young and Dupré equation (Eq 1) permits the determination of change in surface free energy, ΔF, accompanying a small change in solid surface covered, ΔA (Ref 5). Thus,

$$\frac{\partial F}{\partial A}_{p,T} = g_{LV} + g_{SL} - g_{SV} \qquad \textbf{(EQ 1)}$$

where γ_{LV} is the liquid-vapor surface energy, γ_{SL} is the solid-liquid surface energy, and γ_{SV} is the solid-vapor surface energy.

$$\Delta F = \Delta A\,(\gamma_{SL} - \gamma_{SV}) + \Delta A\gamma_{LV}\,\mathrm{COS}\,(\theta - \Delta\theta) \qquad \textbf{(EQ 2)}$$

At equilibrium,

$$\lim_{\Delta A \to \Delta 0} \frac{\Delta F}{\Delta A} = 0 \qquad \textbf{(EQ 3)}$$

and

$$\gamma_{SL} - \gamma_{SV} + \gamma_{LV}\,\mathrm{COS}\,\theta = 0 \qquad \textbf{(EQ 4)}$$

or

$$\gamma_{SL} = \gamma_{SV} - \gamma_{LV}\,\mathrm{COS}\,\theta \qquad \textbf{(EQ 5)}$$

In Eq 4 and 5, it can be seen that θ is greater than 90° (1.6 rad) when γ_{SL} is larger than γ_{SV}, as shown in Fig. 3, and the liquid drop tends to spheroidize.

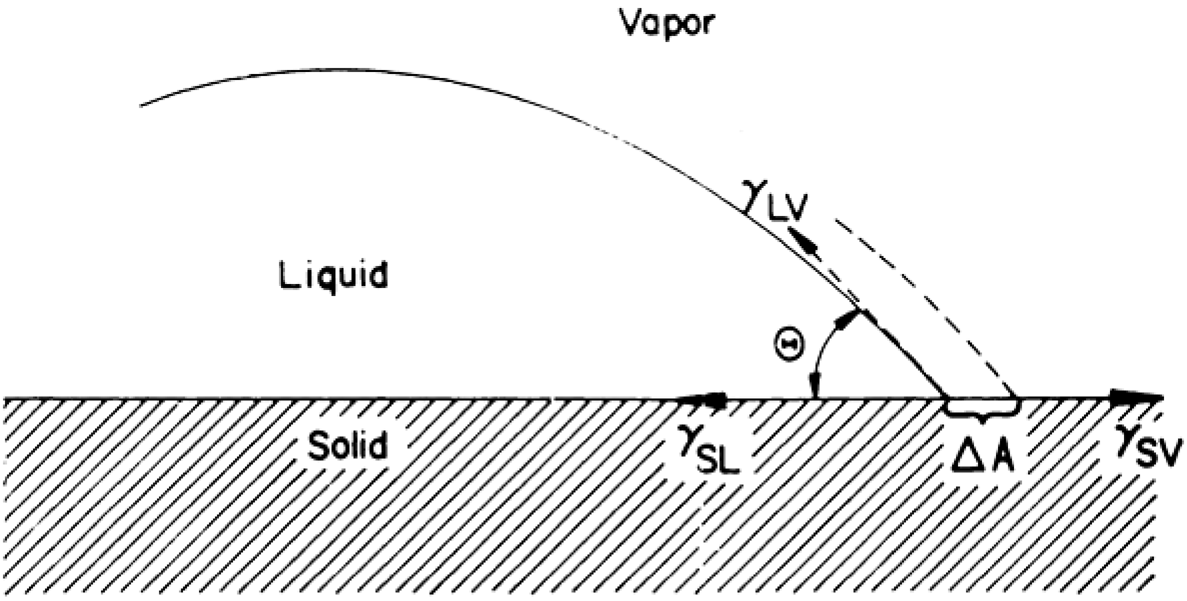

FIG. 2 SESSILE DROP LOCATED ON A SURFACE AND DIRECTIONS OF VECTORS OF THE SURFACE ENERGIES FOR THE SYSTEM

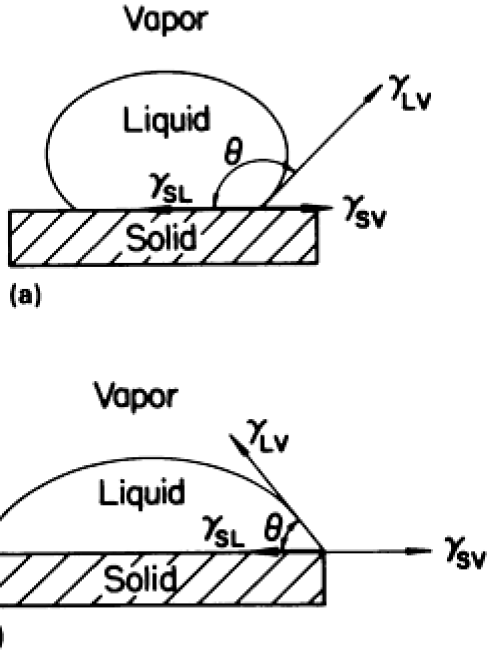

FIG. 3 EFFECT OF INTERFACIAL ENERGIES ON SESSILE DROPS. (A) NO WETTING TAKES PLACE WHEN CONTACT ANGLE IS > 90°. (B) WETTING OCCURS WHEN CONTACT ANGLE IS < 90°.

The contact angle, θ, is less than 90° (1.6 rad) when the reverse is true, as shown at right in Fig. 3, and the liquid drop flattens out and wets the solid. If the balance in such that θ is zero (0 rad) and greater wetting is desired, θ should be as small as possible so that cos θ approaches unity and the liquid spreads over the solid surface.

These considerations show the importance of surface energies in brazing. If brazing filler metal is to successfully form a joint, it must wet the solid material. The surface-energy balance must be such that the contact angle is less than 90° (1.6 rad). The energy equation shows that if θ is to be less than 90° (1.6 rad) (cos θ > 0), then γ_{SV} must be greater than γ_{SL}. This is demonstrated in Table 1, which gives contact-angle measurements of several liquid metals on beryllium at various test temperatures in argon and vacuum.

TABLE 1 CONTACT ANGLES OF SELECTED LIQUID METALS ON BERYLLIUM AT VARIOUS TEST TEMPERATURES IN ARGON AND VACUUM ATMOSPHERES

SYSTEM	ATMOSPHERE	TEMPERATURE		CONTACT ANGLE, DEGREES, AT INDICATED ELAPSED TIME				
		°C	°F	0 s	15 s	45 s	90 s	150 s
AG	VACUUM	1010	1850	142.2	141.1	140.6	141.2	140.2
		1060	1940	135.4	135.6	131.4	133.2	127.9
	ARGON	1010	1850	148.1	148.9	150.7	147.3	143.3
		1060	1940	61.0	59.9	57.2	54.4	52.6
AU	VACUUM	1070	1958	58.6	58.0	57.2	55.7	54.6
	ARGON	1070	1958	57.3	56.6	57.9	56.9	55.9
CU	VACUUM	1108	2026	129.0	128.5	124.2	124.1	123.6
		1133	2071	126.6	122.5	121.8	120.0	120.0
	ARGON	1108	2026	EXTREME BASE-METAL SOLUTION				
		1133	2071	EXTREME BASE-METAL SOLUTION				
GE	VACUUM	1000	1832	145.5	145.3	144.3	143.7	143.6
		1050	1922	102.9	100.0	100.9	96.1	95.4
	ARGON	1000	1832	99.3	99.6	98.8	98.2	97.1

		1050	1922	89.2	89.2	88.9	87.1	87.1
PD-2.1BE	VACUUM	980	1796	77.0	74.0	73.2	70.8	69.9
		1030	1886	74.2	73.2	68.7	66.5	65.6
	ARGON	980	1796	77.2	76.5	76.0	75.0	72.7
		1030	1886	70.2	63.1	61.7	57.2	57.2
AL	VACUUM	710	1310	155.1	154.1	151.3	146.6	142.0
		760	1400	140.0	138.8	136.6	134.1	132.0
	ARGON	710	1310	146.2	139.6	130.5	120.1	116.6
		760	1400	108.2	106.6	102.3	100.0	99.9
ZR-5BE	VACUUM	1030	1886	88.0	85.6	83.9	81.8	80.9
		1080	1976	69.8	67.7	67.3	66.5	64.7
	ARGON	1030	1886	101.4	98.8	99.2	98.3	96.2
		1080	1976	83.6	84.2	84.9	82.4	79.7
TI-6BE	VACUUM	1075	1967	15.8	14.8	14.5	14.0	13.2
		1125	2057	14.2	13.4	12.1	11.2	10.7
	ARGON	1075	1967	13.4	12.0	11.9	9.7	8.6
		1125	2057	10.0	7.9	7.8	6.8	5.6

Although during the testing period minor fluctuations in angle were noted, the overall tendency was for the contact angles to decrease with increasing time at test temperature. This behavior is reflected in the data presented in Table 2. The fluctuations can be attributed to alloying effects, because as alloying progresses, composition and temperature variations change the interfacial energies.

TABLE 2 EFFECT OF TEST CONDITIONS ON INTERFACIAL ENERGIES

SYSTEM	ATMOSPHERE	TEMPERATURE		CONTACT ANGLE(A)		γ_{LV}, DYNES/CM	γ_{LS}, DYNES/CM	FINAL SPREADING COEFFICIENT S_{LS}, DYNES/CM	WORK OF ADHESION (W_{AD}), DYNES/CM
		°C	°F	DEGREES	RADIANS				
AL	VACUUM	710	1310	142.5	2.5	831	2520	-1490	176
		760	1400	132.0	2.3	726	2350	-1210	239
	ARGON	710	1310	116.7	2.0	787	2220	-1140	434
		760	1400	99.9	1.7	561	1960	-658	464
ZR-5BE	VACUUM	1030	1886	79.8	1.4	625	1940	-702	549
		1080	1976	64.6	1.1	520	1770	-427	614
	ARGON	1030	1886	97.0	1.7	488	1780	-405	576
		1080	1976	79.8	1.4	172	1690	-225	569
TI-6BE	VACUUM	1075	1967	13.2	0.2	492(A)	1380	-9.00	970
		1125	2057	10.8	0.2	277	1590	-4.00	548
	ARGON	1075	1967	8.6	0.2	424	1440	-1.00	844
		1125	2057	5.6	0.01	329(B)	1530	+3.00	660

(A) AVERAGE OF TWO TESTS.
(B) ONE TEST ONLY

These data provide a means of comparing the wettability of beryllium by various liquids. If one assumes wettability for contact angles less than 90° and nonwetting for angles greater than 90°, then these data permit qualitative separation of the systems into two classes: those with contact angles less than 90°, and those with contact angles greater than 90°. Gold, Pd-2.1Be, and Ti-6Be were found to wet beryllium at both temperatures. The latter two brazing filler metals, as well as silver and gold, did not wet when tested in argon atmospheres at the lower temperature. However, it was found that an increase of 50 °C (90 °F) resulted in wetting. Aluminum exhibited a nonwetting behavior at both temperatures and in both atmospheres. Copper was nonwetting at both temperatures in vacuum, but underwent extensive alloying with the solid beryllium at these temperatures in argon. From these data it appears that the binary alloys have the greatest tendency to wet beryllium (Ref 6).

Surface-energy calculations have been performed on beryllium-base metals brazed with Al, Zr-5Be, and Ti-6Be filler metals. The calculated spreading coefficient were negative for all system test conditions, except with Ti-6Be. Negative spreading coefficients indicated a very definite trend. The greater the system wettability, as indicated by lower contact angles and liquid-vapor surface tensions, the greater the numerical magnitude of the spreading coefficient.

Of the three systems used, Ti-6Be was observed to provide the best wetting of beryllium, followed by Zr-5Be. Although no basic significance can be given to the absolute values of the calculated surface-tension properties, they are useful in indicating the relative tendencies for these liquids to wet beryllium.

The above observations indicate that some liquid-metal systems will wet beryllium, whereas some system that should wet it do not, in actual practice. The problem is more complex than it seems at first.

Recent work in Japan (Ref 7) and the United States (Ref 8) has shown that the effects of the base-metal/filler-metal reaction on the wetting of silicon nitride (Si_3N_4) and titanium-added active filler metals vary with reaction-layer thickness and morphology, as well as with the composition of the reaction layer. Several results of thermodynamic studies on the formation of the compound by the interfacial reaction are reported (Ref 7, 9). Moreover, kinetic studies (Ref 7) were also performed to achieve a quantitative understanding of the formation of the reaction layer.

It can be concluded that wetting is the ability of the molten brazing filler metal to adhere to the surface of a metal in the solid state and, when cooled below its solidus temperature, to make a strong bond with that metal. Wetting is a function not only of the nature of the brazing filler metal (Table 3), but of the degree of interaction between materials (ceramics, and so on) to be joined. There is considerable evidence that in order to wet well, a molten metal must be capable of dissolving, or alloying with, some of the metal on which it flows.

TABLE 3 COMPOSITION OF SELECTED FILLER METALS USED FOR BRAZING APPLICATIONS

FILLER METAL	COMPOSITION	LIQUIDUS		SOLIDUS	
		°C	°F	°C	°F
SILVER	99.99AG	961	1762	961	1762
CUSIL	72AG-28CU	780	1436	780	1436
PALCUSIL 5	68AG-27CU-5PD	810	1490	807	1485
PALCUSIL 10	58AG-32CU-10PD	852	1566	824	1515
PALCUSIL 15	65AG-20CU-15PD	900	1650	850	1560
PALCUSIL 25	54AG-21CU-25PD	950	1740	900	1650
GAPASIL 9	82AG-9GA-9PD	880	1615	845	1555
NICUSIL 3	71.5AG-28.1CU-0.75NI	795	1465	780	1435
NICUSIL 8	56AG-42CU-2NI	893	1639	771	1420
T50	62.5AG-32.5CU-5NI	866	1591	780	1435
T51	75AG-24.5CU-0.5NI	802	1476	780	1435
T52	77AG-21CU-2NI	830	1525	780	1435
CUSILTIN 5	68AG-27CU-5SN	760	1400	743	1369
CUSILTIN 10	60AG-30CU-10SN	718	1324	602	1116
BRAZE 630	63AG-28CU-6SN-3NI	800	1472	690	1275
BRAZE 580	57AG-33CU-7SN-3MN	730	1345	605	1120
BRAZE 655	65AG-28CU-5MN-2NI	850	1560	750	1380
SILCORO 60	60AU-20AG-20CU	845	1550	835	1535
NIORO	82AU-18NI	950	1740	950	1740
PALNIRO 7	70AU-22NI-8PD	1037	1899	1005	1840
INCURO 60	60AU-37CU-3IN	900	1650	860	1580
SILCORO 75	75AU-20CU-5AG	895	1645	885	1625

NICORO 80	81.5AU-16.5CU-2NI	925	1695	910	1670
PALCUSIL 20	52AU-28CU-20PD	925	1695	875	1605
GOLD	99.99AU	1064	1947	1064	1947
PALNIRO 4	30AU-36NI-34PD	1169	2136	1135	2075
PALNIRO 1	50AU-25NI-25PD	1121	2050	1102	2016
TICUSIL	68.8AG-26.7CU-4.5TI	850	1560	830	1525
PALNICUSIL	48AG-18.9CU-10NI-22.5PD	1179	2154	910	1670
PALCO	65PD-35CO	1235	2255	1230	2245
INCUSIL15	62AG-24CU-15IN	705	1300	630	1165
INCUSIL 10	63AG-27CU-10IN	730	1345	685	1265
BAG-8A	71.8AG-28CU-0.2LI	760	1400	760	1400
BAG-19	92.5AG-7.3CU-0.2LI	890	1635	760	1400
BRAZE 071	85CU-7AG-8SN	986	1807	665	1230
BRAZE 852	85AG-15MN	970	1780	960	1760
NIORONI	73.8AU-26.2NI	1010	1850	980	1795
NICUMAN 23	67.5CU-23.5MN-9NI	955	1750	925	1695
PALSIL 10	90AG-10PD	1065	1950	1002	1836
PALNI	60PD-40NI	1238	2260	1238	2260

Wetting is only one facet of the brazing process. Another factor that affects wetting is the cleanliness of the surface to be wetted. Oxide layers inhibit wetting and spreading, as do grease, dirt, and other contaminants that prevent good contact between the brazing filler metal and the base metal. One of the functions of a flux the oxide layer on the joint area and thereby expose clean base metal.

Good wetting and spreading of the liquid filler metal on the base metal are necessary in brazing because the mechanics of the process demand that the filler metal be brought smoothly, rapidly, and continuously to the joint opening. If the conditions within the capillary space of the joint do not promote good wetting, then the filler metal will not be drawn into the space by capillary attraction.

Kim *et al.* (Ref 10) investigated the wettability and reactivity of pressureless sintered Si_3N_4 with powdered copper-titanium filler metal using sessile drop tests conducted in a vacuum where Si_3N_4 was bonded to itself and joint strength was evaluated by shear testing. The wettability of Si_3N_4 on copper-titanium base filler metals was improved greatly by adding up to 50 wt% titanium. However, the reaction-layer thickness was increased up to 10% and, thereafter, decreased up to 50 wt%.

Wettability improved with the increment of titanium concentration, and no linear proportionality was found between the reaction and the wettability. As titanium concentration increased, a continuous thin layer tended to form at the interface and markedly improved wettability.

The thickness of the reaction layer was not in proportion to the titanium concentration. For a copper-titanium filler metal with a low titanium concentration (less than 5 wt%), the initial reaction layer tended to form discontinuously, whereas the reaction layer thickness increased with titanium concentration. On the contrary, in the filler metal with a high titanium concentration (15 wt%) the reaction-layer thickness decreased as titanium concentration increased because a continuous thin reaction layer tended to form (Ref 10).

The shear strength of the Si_3N_4-Si_3N_4 joint using copper-titanium filler metal is influenced by reaction-layer thickness and morphology. Strength increased with the increment of titanium concentration, and was accompanied by improved wettability, because the tendency of the continuous thin reaction layer to form increased.

With the development of new steels, Keller *et al.* (Ref 11) undertook an extensive study of the wettability of commercial braze filler metals on type 304 stainless steel, the most common of the austenitic stainless steels; type 316 stainless steel, which is often substituted for type 304 when increased corrosion resistance is desired; and 21-6-9, a new high-manganese stainless steel that is an attractive alternative to types 304 or 316 for certain applications. Because the 21-6-9 steel has a very stable austenitic structure, it is attractive for cryogenic applications (Ref 12).

Because high wettability means that the thermocapillary attraction that fills the braze joint is strong, wettability is an important component of braze performance. The wettability index (WI) developed by Feduska (Ref 13) is used as the measure of wettability. The WI is defined as the area covered by the braze metal filler times the cosine of the contact angle between the braze and the base metal. Therefore, the higher the WI, the better the braze filler metal wets the base metal. It should be emphasized that the WI, as defined, is a relative measure, and depends on the volume of filler metal used (Table 4).

TABLE 4 WETTABILITY INDICES OF BRAZE FILLER METALS ON SELECTED STAINLESS STEEL BASE METALS

FILLER METAL	**TEST TEMPERATURE**		**WETTABILITY INDEX FOR INDICATED STAINLESS STEEL BASE METAL**		
	°C	**°F**	**TYPE 316**	**21-6-9**	**TYPE 304**
SILVER	975	1785	. . .	. . .	. . .
	1000	1830	0.008	. . .	0.015
	1050	1920	0.041	. . .	0.023
	1100	2010	0.053	. . .	0.020
	1150	2100	. . .	. . .	. . .
CUSIL	800	1470	. . .	0.013	. . .
	850	1560	. . .	0.007	. . .
	900	1650	0.022	0.016	0.003
	925	1695	0.032	. . .	. . .
	950	1740	0.037	. . .	. . .
PALCUSIL 5	800	1470	0.027	. . .	. . .
	850	1560	. . .	0.014	. . .
	900	1650	0.047	0.020	0.011
	950	1740	0.080	0.061	0.035
PALCUSIL 10	850	1560	0.035	0.020	0.015
	900	1650	0.057	0.050	0.025
	950	1740	0.107	0.101	0.062
PALCUSIL 15	900	1650	0.107	0.092	0.068
	950	1740	0.152	0.170	0.119
	1000	1830	0.754	0.263	0.212
PALCUSIL 25	950	1740	0.104	0.107	0.096
	975	1785	0.225	. . .	. . .
	1000	1830	. . .	0.283	0.226
GAPASIL 9	900	1650	0.023	. . .	. . .
	950	1740	0.068	. . .	. . .
	1000	1830	0.269	. . .	. . .
NICUSIL 3	800	1470	0.006	0.000	0.012
	850	1560	0.027	0.017	0.008
	900	1650	0.041	0.039	0.021
	925	1695	0.051	. . .	. . .
	950	1740	0.064	. . .	. . .
NICUSIL 8	850	1560	0.024	. . .	. . .
	900	1650	0.052	. . .	0.033
	950	1740	0.085	. . .	0.064
	1000	1830	. . .	. . .	0.229
T-50	850	1560	0.024	. . .	. . .
	900	1650	0.038	. . .	0.026

FILLER METAL	TEST TEMPERATURE		WETTABILITY INDEX FOR INDICATED STAINLESS STEEL BASE METAL		
	°C	°F	TYPE 316	21-6-9	TYPE 304
	950	1740	0.062	. . .	0.057
	1000	1830	. . .	. . .	. . .
T-51	800	1470	0.007	. . .	. . .
	850	1560	0.029	. . .	. . .
	900	1650	0.039	. . .	. . .
	950	1740	. . .	. . .	0.029
	1000	1830	. . .	. . .	0.082
T-52	850	1560	0.027	. . .	0.001
	900	1650	0.045	. . .	0.014
	950	1740	0.063	. . .	0.045
	1000	1830	. . .	. . .	0.090
CUSILTIN 5	850	1560	0.015	. . .	0.000
	875	1605	0.043	. . .	. . .
	900	1650	0.047	. . .	0.013
	950	1740	. . .	. . .	0.033
	1000	1830	. . .	. . .	. . .
CUSILTIN 10	750	1380	. . .	0.025	0.000
	800	1470	. . .	0.017	0.023
	825	1515	0.021	. . .	. . .
	850	1560	0.034	0.051	0.050
	875	1605	0.043	. . .	. . .
BRAZE 630	800	1470	0.014	0.037	0.023
	850	1560	0.046	0.065	0.024
	900	1650	0.064	. . .	. . .
BRAZE 580	750	1380	0.020	0.060	0.060
	800	1470	0.051	0.056	0.089
	850	1560	0.073	0.120	0.102
	875	1605	0.078	. . .	. . .
BRAZE 655	825	1515	0.080	. . .	. . .
	850	1560	0.110	0.037	0.074
	875	1605	0.116	. . .	. . .
	900	1650	. . .	0.137	0.124
SILCORO 60	850	1560	. . .	0.005	0.004
	900	1650	0.039	0.007	0.011
	925	1695	0.051	. . .	. . .
	950	1740	0.073	0.025	0.016
	1000	1830	. . .	0.055	0.037
NIORO	950	1740	. . .	0.000	0.000
	975	1785	0.049	. . .	. . .
	1000	1830	0.061	0.060	0.065
PALNIRO 7	1025	1875	. . .	. . .	. . .
	1050	1920	0.053	. . .	. . .
	1075	1965	0.073	. . .	. . .
INCURO 60	900	1650	0.010	. . .	. . .
	950	1740	0.025	. . .	. . .
	1000	1830	0.091	. . .	. . .
SILCORO 75	900	1650	0.006	. . .	. . .

FILLER METAL	**TEST TEMPERATURE**		**WETTABILITY INDEX FOR INDICATED STAINLESS STEEL BASE METAL**		
	°C	**°F**	**TYPE 316**	**21-6-9**	**TYPE 304**
	950	1740	0.057	. . .	. . .
	1000	1830	0.170	. . .	. . .
NICORO 80	950	1740	0.041	0.019	0.016
	1000	1830	0.163	0.126	0.070
	1050	1920	0.413	0.190	0.084
PALCUSIL 20	875	1610	0.061	. . .	. . .
	900	1650	0.110	. . .	. . .
	925	1695	0.122	. . .	. . .
GOLD	1070	1960	. . .	. . .	0.088
	1075	1965	0.087	. . .	. . .
	1100	2010	0.358	. . .	0.238
	1150	2100	. . .	. . .	0.355
PALNIRO 4	1175	2145	0.036	. . .	. . .
	1200	2190	0.061	. . .	. . .
	1225	2235	0.078	. . .	. . .
PALNIRO 1	1125	2055	0.041	. . .	. . .
	1150	2100	0.063	. . .	. . .
	1175	2145	0.075	. . .	. . .
TICUSIL	875	1605	0.032	. . .	. . .
	900	1650	0.057	. . .	. . .
	950	1740	0.083	. . .	. . .
PALNICUSIL	950	1740	0.025	. . .	0.038
	975	1785	0.213	. . .	. . .
	1000	1830	. . .	. . .	. . .
	1025	1875	0.362	. . .	. . .
	1050	1920	. . .	. . .	. . .
	1075	1965	0.336	. . .	. . .
	1100	2010	. . .	. . .	. . .
PALCO	1250	2280	0.000	. . .	. . .
	1275	2325	0.073	. . .	. . .
	1300	2370	0.159	. . .	. . .
INCUSIL 15	750	1380	. . .	. . .	. . .
	800	1470	. . .	. . .	. . .
	850	1560	. . .	. . .	. . .
INCUSIL 10	750	1380	. . .	. . .	. . .
	800	1470	. . .	. . .	. . .
	850	1560	0.008	. . .	. . .
BAG-8A	800	1470	0.005	. . .	. . .
	850	1560	0.008	. . .	. . .
	900	1650	0.032	. . .	. . .
	950	1740	0.045	. . .	. . .
BAG-19	950	1740	0.016	. . .	. . .
	1000	1830	0.016	. . .	. . .
	1050	1920	0.034	. . .	. . .
BRAZE 071	1000	1830	0.066	0.106	0.101
	1050	1920	. . .	0.125	0.101
BRAZE 852	1000	1830	0.039	0.037	0.038

FILLER METAL	TEST TEMPERATURE		WETTABILITY INDEX FOR INDICATED STAINLESS STEEL BASE METAL		
	°C	**°F**	**TYPE 316**	**21-6-9**	**TYPE 304**
	1050	1920	0.029	0.038	0.042
	1100	2010	0.043	0.037	0.020
NIORONI	1000	1830	0.068	. . .	. . .
	1025	1875	0.070	. . .	. . .
	1100	2010	0.116	. . .	. . .
NICUMAN 23	950	1740	0.026	. . .	. . .
	975	1785	0.099	. . .	. . .
	1000	1830	0.091	. . .	. . .
	1025	1875	0.091	. . .	. . .
PALSIL 10	1050	1920	0.113	. . .	. . .
	1075	1965	0.113	. . .	. . .
	1100	2010	0.126	. . .	. . .
PALNI	1225	2235	. . .	. . .	. . .
	1250	2280	0.068	. . .	. . .
	1275	2325	0.078	. . .	. . .
	1300	2370	0.107	. . .	. . .
PALMANSIL 5	950	1740	. . .	. . .	0.057
	1000	1830	. . .	. . .	. . .
	1050	1920	. . .	. . .	. . .
	1100	2010	. . .	. . .	. . .

The WIs for each braze filler metal at each temperature on type 316 stainless steel, type 304L stainless steel, and 21-6-9 stainless steel are provided in Table 4. Wetting indices greater than 0.05 are indicative of good performance during brazing, whereas WI values greater than 0.10 are indicative of excellent performance during brazing (Ref 13).

A comparison of the WI of the braze filler metals on the three stainless steels revealed these trends:

- FILLER METALS GENERALLY WET TYPE 316 STAINLESS STEEL BETTER THAN TYPE 304 STAINLESS STEEL AND GENERALLY WET TYPE 316 STAINLESS BETTER THAN 21-6-9 STAINLESS.
- THE DEGREE OF WETTING OF MOST BRAZE FILLER METALS ON 21-6-9 STAINLESS WAS EQUAL TO OR BETTER THAN IT WAS ON TYPE 304 STAINLESS (REF 12).
- THE IMPROVED WETTABILITY OF BRAZE FILLER METALS ON TYPE 316 STAINLESS IS BELIEVED TO BE DUE TO THE PRESENCE OF MOLYBDENUM IN THE SURFACE OXIDE, WHEREAS THE ENHANCED WETTING ON 21-6-9 STAINLESS VERSUS TYPE 304 STAINLESS IS BELIEVED TO BE DUE TO THE HIGHER MANGANESE CONTENT OF THE SURFACE OXIDE.

Essentially, the successful joining of components by the brazing process depends on the selected brazing filler metal having a melting point above 450 °C (840 °F) and wetting the base metal without melting it. Furthermore, the joint must be designed to ensure that the mating surfaces of the components are parallel and close enough together to cause capillary attraction.

Fundamentals of Brazing

Mel M. Schwartz, Sikorsky Aircraft

Elements of the Brazing Process

An engineer must consider reliability and cost when designing the braze joint. Joint strength, fatigue resistance, corrosion susceptibility, and high-temperature stability are additional concerns that determine the selection of joint design, braze filler materials, and processing parameters.

A careful and intelligent appraisal of the following elements is required in order to produce satisfactory brazed joints:

- FILLER-METAL FLOW
- BASE-METAL CHARACTERISTICS
- FILLER-METAL CHARACTERISTICS
- SURFACE PREPARATION
- JOINT DESIGN AND CLEARANCE
- TEMPERATURE AND TIME
- RATE AND SOURCE OF HEATING
- PROTECTION BY AN ATMOSPHERE OR FLUX

Filler-Metal Flow. As mentioned previously, wetting is only one important facet of the brazing process. A low contact angle, which implies wetting, is also necessary, but is not a sufficient condition itself for flow. Viscosity is also important. Brazing filler metals with narrow melting ranges that are close to the eutectic composition generally have lower viscosities than those with wide melting ranges. Thus, a high surface tension of liquid filler metal, a low contact angle, and low viscosity are all desirable.

Flowability is the property of a brazing filler metal that determines the distance it will travel away from its original position, because of the action of capillary forces. To flow well, a filler metal must not gain an appreciable increase in its liquidus temperature even though its composition is altered by the addition of the metal it has dissolved. This is important because the brazing operation is carried out at temperatures just above the liquidus of the filler metal.

The composition and surface energy of liquids and solids are assumed to remain constant. In real systems, however, these interactions occur (Ref 14):

- ALLOY FORMATION BETWEEN LIQUID AND BASE METAL
- DIFFUSION OF BASE METAL INTO BRAZING FILLER METAL
- DIFFUSION OF FILLER METAL INTO GRAINS OF BASE METAL
- PENETRATION OF FILLER METAL ALONG GRAIN BOUNDARIES
- FORMATION OF INTERMETALLIC COMPOUNDS

In practice, interactions are usually minimized by selecting the proper brazing filler metal; keeping the brazing temperature as low as possible, but high enough to produce flow; and keeping the time of brazing temperature short and cooling the brazed joint as quickly as possible without causing cracking or distortion. When diffusion brazing is desired, higher brazing temperatures and longer times at brazing temperatures are employed.

Base-Metal Characteristics. The base metal has a prime effect on joint strength. A high-strength base metal produces joints of greater strength than those made with softer base metals (other factors being equal). When hardenable metals are brazed, joint strength becomes less predictable. This is because more-complex metallurgical reactions between hardenable base metals and the brazing filler metals are involved. These reactions can cause changes in the base-metal hardenability and can create residual stresses (Ref 15).

In cases where different materials make up the assembly and gaps may open or close as heating proceeds to the joining temperature, the coefficient of thermal expansion becomes vitally important.

Several metallurgical phenomena influence the behavior of brazed joints and, in some instances, necessitate special procedures. These base-metal effects include alloying by brazing filler; carbide precipitation; stress cracking; hydrogen, sulfur, and phosphorus embrittlement; and oxidation stability.

The extent of interaction varies greatly, depending on compositions (base metal and brazing filler metal) and thermal cycles. There is always some interaction, except in cases of mutual insolubility.

The strength of the base metal has a profound effect on the strength of the brazed joint. Therefore, this property must be carefully considered when designing the joint to have specific properties. Some base metals also are easier to braze than others, particularly by specific brazing processes.

Filler-Metal Characteristics. The second material involved in joint structures is the brazing filler metal. Unfortunately, it cannot be chosen to provide a specific joint strength. Actually, strong joints can be brazed with almost any good commercial brazing filler metal if correct brazing methods and joint design are implemented.

Necessary characteristics of brazing filler metals are:

- PROPER FLUIDITY AT BRAZING TEMPERATURES TO ENSURE FLOW BY CAPILLARY ACTION AND TO PROVIDE FULL ALLOY DISTRIBUTION
- STABILITY TO AVOID PREMATURE RELEASE OF LOW-MELTING-POINT ELEMENTS IN THE BRAZING FILLER METAL
- ABILITY TO WET THE BASE-METAL JOINT SURFACE
- LOW VOLATILIZATION OF ALLOYING ELEMENTS OF THE BRAZING FILLER METAL AT BRAZING TEMPERATURES
- ABILITY TO ALLOY OR COMBINE WITH THE BASE METAL TO FORM AN ALLOY WITH A HIGHER MELTING TEMPERATURE
- CONTROL OF WASHING OR CONTROL OF EROSION BETWEEN THE BRAZING FILLER METAL AND THE BASE METAL WITHIN THE LIMITS REQUIRED FOR THE BRAZING OPERATION

It should be noted that the strength of the brazed joint is not directly related to the method of filler-metal melting. For example, if constructional metals are produced by vacuum melting, then there is a definite relationship between the vacuum-melting practice and the final strength of the ingot, bar, or rolled sheet. With a brazing filler metal, however, joint strength is dependent on joint design, brazing temperature, amount of brazing filler metal applied, location and method of application, heating rate, and many other factors that constitute the brazing technique.

The degree to which brazing filler metal interacts with and penetrates the base metal during brazing depends on the intensity of mutual diffusion processes that occur between both those materials. In applications that require strong joints for high-temperature, high-stress service conditions (such as turbine rotor assemblies and jet-engine components), it is generally wise to specify a brazing filler metal that has high diffusion and solution properties with the base metal. When the assembly is constructed of extremely thin base metals (as in honeycomb structures and some heat exchangers), good practice entails specifying a brazing filler metal that contains elements with a low-diffusion characteristic relative to the base metal being used. Diffusion, a normal part of the metallurgical process, can contribute to good brazed joints when brazing, for example, high-temperature metals with nickel-base brazing filler metals.

Brazing Temperature. When choosing a brazing filler metal, the first selection criterion is the brazing temperature. Some brazing-temperature ranges are given in Table 5. Very few brazing filler metals possess narrow melting ranges. Brazing filler metals in which the solidus and liquidus temperatures are close together do not usually exhibit a strong tendency to coexist as a mixture of liquid and solid phases or to liquate. They flow readily and should be used with small joint clearances. As the solidus and liquidus temperatures diverge, the tendency to liquate increases, requiring greater precautions in brazing filler metal application. The mixture of solid and liquid metal can aid gap filling.

TABLE 5 JOINING TEMPERATURES FOR FILLER METALS USED IN BRAZING APPLICATIONS

BRAZING FILLER-METAL GROUPS	JOINING TEMPERATURE	
	°C	°F
NI, CO, AND PD ALLOYS	[GES]1100	[GES]2012
CU, NI, AND AU ALLOYS	[GES]1100	[GES]2012
CU-ZN, CU-SN, NI-P, NI-CR-P, PD-AG-CU	[GES]900	[GES]1652
CU-P, CU-AG-P	600-800	1112-1472
AG-CU-ZN, AG-CU-ZN-CD	600-800	1112-1472
AL-SI	580-600	1076-1112
MG-AL-ZN	585-615	1085-1139

The necessity for the brazing filler metal to melt below the solidus of the base metal is just one of several factors that affect its selection. It may be necessary for the brazing filler metal to melt below the temperature at which parts to be brazed lose strength or above the temperature at which oxides are reduced or dissociated. Joining may have to be carried out above the solution-treatment temperature of the base metal, or at a structure refining temperature, or below the remelt temperature of a brazing filler metal used previously in the production of a brazed subassembly.

Liquidation. During melting, the composition of the liquid and solid filler metal phases changes as the temperature increases from the solidus to the liquidus point. If the portion that melts first is allowed to flow out, then the remaining solid phases have higher melting points than the original composition and may never melt, remaining behind as a residue, or "skull." Filler metals with narrow melting ranges do not tend to separate, but flow quite freely in joints of extremely narrow clearance as long as the solution and diffusion rates of the filler metal with the base metal are low (as in aluminum brazing, the use of silver filler metal on copper, and so on). The rapid heating of filler metals with wide melting ranges or their application to the joint after the base metal reaches brazing temperature will minimize the separation, or liquation. However, liquation cannot be entirely eliminated and wide-melting-range filler metals, which tend to have more sluggish flow, will require wider joint clearances and will form larger fillets at the joint extremities. A few brazing filler metals become sufficiently fluid below the actual liquidus temperature, and satisfactory joints are achieved, even though the liquidus temperature has not been reached.

When sluggish behavior is needed, such as in filling large gaps, brazing can be accomplished within the melting range of the filler metal. However, the brazing temperature is usually 10 to 93 °C (20 to 170 °F) above the liquidus of the filler metal. The actual temperature required to produce a good joint filling is influenced by factors such as heating rate, brazing environment (atmosphere or flux), thickness of parts, thermal conductivity of the metals being joined, and type of joint to be made.

Large-scale mechanical properties of brazing filler metals can be a guideline for their suitability (in terms of strength, oxidation resistance, and so on) for use in different capillary joining applications. However, designers cannot use the mechanical properties of brazed assemblies that are related to different joint configurations brazed at given cycles of time and temperatures.

The placement of the brazing filler metal is an important design consideration, not only because the joint must be accessible to the method chosen, but because, in automatic heating setups, the filler metal must be retained in its location until molten. Brazing filler metals are available in different forms (Table 6) and filler-metal selection may depend on which form is suitable for a particular joint design.

TABLE 6 AVAILABLE PRODUCT FORMS OF BRAZING FILLER METALS CATEGORIZED BY GROUP

BRAZING FILLER-METAL GROUPS	AVAILABLE FORMS[(A)]											
	AD	CL	FL	FO	PA	PB	PO	PR	SH	RD	SP	WI
NI AND CO ALLOYS	X	. . .	X	(B)	X	X	X	(B)	X	(B)	. . .	(C)
PD ALLOYS	. . .	. . .	X	X	. . .	. . .	X	X	X	. . .	X	X
CU AND AU ALLOYS	. . .	. . .	X	X	X	. . .	X	X	X	. . .	X	X

NI ALLOYS	X	. . .	X	(B)	X	X	X	(B)	X	(B)	. . .	(C)
CU-SN, CU-ZN, PD-AG-CU	. . .	. . .	X	X	. . .	. . .	X	X	X	. . .	X	X
NI-P, NI-CR-P	X	. . .	. . .	. . .	X	X	X	. . .	. . .	. . .	. . .	. . .
CU-P, CU-AG-P	. . .	. . .	. . .	. . .	X	. . .	X	. . .	. . .	. . .	X	X
AG-CU-ZN, AG-CU-ZN-CD	. . .	. . .	X	X	. . .	. . .	X	X	X	. . .	X	X
AL-SI	. . .	X	. . .	. . .	. . .	. . .	X	. . .	X	. . .	. . .	X
MG-AL-ZN	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	X	. . .	X	X

(A) AD, ADHESIVE SHEET; CL, CLADDING; FL, FLUX PASTE; FO, FOIL; PA, PASTE (NONFLUXING); PB, PLASTIC-BONDED SHEET; PO, POWDER; PR, PREFORM; SH, SHIM; RD, ROD; SP, STRIP; WI, WIRE.

(B) A FEW ALLOYS ONLY.

(C) PLASTIC BOND

Several general rules apply in the filler-metal placement. Wherever possible, the filler metal should be placed on the most slowly heated part of the assembly in order to ensure complete melting. Although brazing is independent of gravity, gravity can be used to assist filler-metal flow, particularly for those filler metals having wide ranges between their solidus and liquidus temperatures. Brazing filler metals can be chosen to fill wide gaps or to flow through joint configurations where the gap may vary, for example, around a corner. Unless movement between the components being joined is unimportant or can be corrected manually (through self-jigging or by using fixtures after the filler metal is molten), the filler metal should be placed outside the joint and allowed to flow into it. It should not be placed between the joint members. If erosion of thin members is possible, then the brazing filler metal should be placed on the heavier sections, which heat up more slowly, so that flow proceeds toward the thin sections. Apart from suiting the placement method selected for the joint, the form chosen for the brazing filler metal may be needed to accurately gage the amount applied, not just for economy and reproducibility, but to regulate and maintain joint properties and configuration.

Surface Preparation. A clean, oxide-free surface is imperative to ensure uniform quality and sound brazed joints. All grease, oil, dirt, and oxides must be carefully removed from the base and filler metals before brazing, because only then can uniform capillary attraction be obtained. Brazing should be done as soon as possible after the material has been cleaned. The length of time that the cleaning remains effective depends on the metals involved, atmospheric conditions, storage and handling practices, and other factors. Cleaning operations are commonly categorized as being either chemical or mechanical. Chemical cleaning is the most effective means of removing all traces of oil or grease. Trichloroethylene and trisodium phosphate are the usual cleaning agents employed. Oxides and scale that cannot be eliminated by these cleaners should be removed by other chemical means.

The selection of the chemical cleaning agent depends on the nature of the contaminant, the base metal, the surface condition, and the joint design. Regardless of the nature of the cleaning agent or the cleaning method used, it is important that all residue or surface film be adequately rinsed from the cleaned parts to prevent the formation of other equally undesirable films on the faying surfaces.

Objectionable surface conditions can be removed by mechanical means, such as grinding, filing, wire brushing, or any form of machining, provided that joint clearances are not disturbed. When grinding the surfaces of the parts to be brazed, care should be exercised to ensure that the coolant is clean and free from impurities to avoid grinding these impurities into the finished surfaces.

When faying surfaces of parts to be brazed are prepared by blasting techniques, several factors should be considered. There are two purposes behind the blasting of parts to be brazed. One is to remove any oxide film and the other is to roughen the mating surfaces in order to increase capillary attraction of the brazing filler metal. The blasting material must be clean and must not leave any deposit on the surfaces to be joined that restricts filler-metal flow or impairs brazing. The particles of the blasting material should be angular rather than spherical, so that the blasted parts are lightly roughened, rather than peened, after the scale is removed. The operation should not distort or otherwise harm delicate parts. Vapor blasting and similar wet blasting methods require care, because of the possibility of surface contamination.

Mechanical cleaning may be adequate, in which case it should be permitted (by the design) during manufacture. In those cases that require chemical cleaning, the cleaning operation may be followed by protective electroplating, which necessitates access to the faying surface by the liquids involved.

Another surface protection technique is the use of solid and liquid brazing fluxes. At temperatures up to about 1000 °C (1830 °F), fluxes often provide the easiest method of maintaining or producing surface cleanliness. In such cases, the design must not only permit easy ingress of the flux, but should allow the filler metal to wash it through the joint. Above 1000 °C (1830 °F), the flux residues can be difficult to remove, and surface cleaning can be accompanied by brazing in furnaces with protective atmospheres. However, the design must permit the gas to penetrate the joint.

In addition to cleanliness and freedom from oxides, another important factor in determining the ease and evenness of brazing filler metal flow is surface roughness. Generally, a liquid that wets a smooth surface will wet a rough one even more. A rough surface will modify filler-metal flow from laminar to turbulent, prolonging flow time and increasing the possibility of alloying and other interactions. Surfaces often are not truly planar, and, in some instances, surface roughening will improve the uniformity of the joint clearance. A rough surface features a series of crests and troughs that produce turbulent flow. These irregularities on the surface simulate grips to prevent the flow of the filler metal.

Conversely, there may be a requirement that brazing filler metal not flow onto some surfaces. Stop-off materials will often avoid flow, but the design must permit easy application of the stop-off material without contaminating the surfaces to be joined.

Self-fluxing filler metals, in a suitably protective environment such as vacuum, may provide the essential surface wetting.

Joint Design and Clearance. A brazed joint is not a homogeneous body. Rather, it is a heterogeneous area that is composed of different phases with different physical and chemical properties. In the simplest case, it consists of the base-metal parts to be joined and the added brazing filler metal. However, dissimilar materials must also be considered.

Small clearances are used because the smaller the clearance, the easier it is for capillarity to distribute the brazing filler metal throughout the joint area and the less likely it is that voids or shrinkage cavities will form as the brazing filler metal solidifies. The optimum joints are those in which the entire joint area is wetted and filled by the brazing filler metal. Typically, brazing clearances that range from 0.03 to 0.08 mm (0.001 to 0.003 in.) are designed for the best capillary action and greatest joint strength.

Because brazing relies on capillary attraction, the design of a point must provide an unobstructed and unbroken capillary path to enable the escape of flux, if used, as well as allow the brazing filler metal into the joint. In cases where filler metal is added to a joint by hand, such as by feeding in a rod or wire, the joint entry must be visible and accessible.

Some of the more-important factors influencing joint design are the required strength and corrosion resistance, the necessary electrical and thermal conductivity, the materials to be joined, the mode of application of the brazing filler metal, and the postjoining inspection needs.

Consideration also should be given to the ductility of the base metal, the stress distribution in the joint, and the relative movements of the two surfaces during joining, which may introduce distortion in dimensions of the work to be brazed.

Viscosity, surface tension, and specific gravity of the brazing filler metal are not the only factors that determine the gap-filling capability of a given filler metal. Joint strength increases as joint gap decreases, down to a minimum. Table 7 gives the allowable joint clearances for various filler-metal systems. Other factors that influence optimum joint gap with a specific brazing filler metal are joint length, brazing temperature, and base-metal/filler-metal metallurgical reactions.

TABLE 7 RECOMMENDED GAP FOR SELECTED BRAZE FILLER METALS

BRAZING FILLER-METAL SYSTEM	JOINT CLEARANCE	
	mm	in.
AL-SI ALLOYS[(A)]	0.15-0.61	0.006-0.024
MG ALLOYS	0.10-0.25	0.004-0.010
CU	0.00-0.05	0.000-0.002
CU-P	0.03-0.13	0.001-0.005
CU-ZN	0.05-0.13	0.002-0.005
AG ALLOYS	0.05-0.13	0.002-0.005

AU ALLOYS	0.03-0.13	0.001-0.005
NI-P ALLOYS	0.00-0.03	0.000-0.001
NI-CR ALLOYS(B)	0.03-0.61	0.001-0.024
PD ALLOYS	0.03-0.10	0.001-0.004

(A) IF JOINT LENGTH IS LESS THAN 6 MM (0.240 IN.), GAP IS 0.12 TO 0.75 MM (0.005 TO 0.030 IN.). IF JOINT LENGTH EXCEEDS 6 MM (0.240 IN.), GAP IS 0.25 TO 0.60 MM (0.010 TO 0.24 IN.).

(B) MANY DIFFERENT NICKEL BRAZING FILLER METALS ARE AVAILABLE, AND JOINT GAP REQUIREMENTS MAY VARY GREATLY FROM ONE FILLER METAL TO ANOTHER

It is important to remember that an assembly expands during heating and that the joint gap can either widen or close by the time the brazing filler metal starts to melt and move. It is desirable to design the joint to expose the solidifying filler metal to compressive, rather than tensile, stress. This is much more important in brazing than in soldering, because brazing temperatures are higher, increasing the total thermal expansion. With cylindrical joints, the components with the larger coefficient of expansion should be on the outside, whenever possible.

Joint clearance is probably one of the most significant factors in vacuum brazing operations. Naturally, it receives special consideration when joint are designed at room temperature. Actually, joint clearance is not the same at all phases of brazing. It will have one value before brazing, another value at the brazing temperature, and still another value after brazing, especially if there has been a substantial interaction between the brazing filler metal and the base metal. To avoid confusion, it has become general practice to specify joint clearance as the value at room temperature before brazing.

The recommended joint clearances given in Table 7 are based on joints that have members of similar metals and equal mass. When dissimilar metals and/or metals of widely differing masses are joined by brazing, special problems arise. These problems necessitate specialized selection of brazing filler metals. Furthermore, the most suitable joint clearance for the specific job must be carefully determined. Thus, clearances given in Table 7 are the clearances that must be attained at the brazing temperature.

Although there are many kinds of brazed joints, the selection of joint type is not as complicated as it may seem, because butt and lap joints are the two fundamental types used. All other types, such as the scarf joint, are modifications of these two. The scarf joint of the scarf angle. It approaches the lap joint at the other extreme of the scarf angle.

Selection of joint type is influenced by the configuration of the parts, as well as by joint strength and other service requirements, such as electrical conductivity, pressure tightness, and appearance. Also influential in the selection of joint type are fabrication techniques, production quantities, and methods of feeding brazing filler metal. Lap joints are generally preferred for brazing operations, particularly when it is important that the joints be at least as strong as the weaker member. For maximum strength, lap-joint length should equal three to four times the thickness of the thinner member.

Temperature and Time. The temperature of the brazing filler metal has an important effect on the wetting and alloying action, which increases with increasing temperature. The temperature must be above the melting point of the brazing filler metal and below the melting point of the parent metal. Within this range, a brazing temperature that is most satisfactory overall is generally selected.

Usually, low brazing temperatures are preferred in order to economize on the heat energy required, minimize the heat effect on the base metal (annealing, grain growth, or warpage, for example), minimize base-metal/filler-metal interactions, and increase the life of fixtures, jigs, or other tools.

Higher brazing temperatures may be desirable so as to:

- ENABLE THE USE OF A HIGHER-MELTING, BUT MORE ECONOMICAL, BRAZING FILLER METAL
- COMBINE ANNEALING, STRESS RELIEF, OR HEAT TREATMENT OF THE BASE METAL WITH BRAZING
- PERMIT SUBSEQUENT PROCESSING AT ELEVATED TEMPERATURES
- PROMOTE BASE-METAL/FILLER-METAL INTERACTIONS IN ORDER TO CHANGE THE

COMPOSITION OF THE BRAZING FILLER METAL (THIS TECHNIQUE IS USUALLY USED TO INCREASE THE REMELT TEMPERATURE AND DUCTILITY OF THE JOINT)
- EFFECTIVELY REMOVE SURFACE CONTAMINANTS AND OXIDES WITHIN PROTECTIVE ATMOSPHERES BRAZING (ALSO APPLIES TO PURE DRY HYDROGEN, TO ARGON, AND TO VACUUM)
- AVOID STRESS CRACKING

The time at brazing temperature also affects the wetting action. If the brazing filler metal has a tendency to creep, the distance generally increases with time. The alloying action between filler metal and parent metal is, of course, a function of temperature, time, and quantity of filler metal. For production work, temperature, time, and quantity of filler metal are generally kept at a minimum, consistent with good quality, where diffusion is not required.

References cited in this section

14. W.H. KOHL, SOLDERING AND BRAZING, *VACUUM,* VOL 14, 1964, P 175-198
15. *METALS HANDBOOK,* 9TH ED., *WELDING, BRAZING, AND SOLDERING,* VOL 6, AMERICAN SOCIETY FOR METALS, 1983, P 929-995

Fundamentals of Brazing

Mel M. Schwartz, Sikorsky Aircraft

Heating Methods

The numerous heating methods available for brazing often represent constraints on the designer or engineer when selecting the best type of capillary joint (Table 8). However, because effective capillary joining requires the efficient transfer of heat from the heat source into the joint, one cannot braze a 0.025 mm (0.001 in.) diam wire to a 2.3 kg (5.1 lb) lump of copper with a small torch.

TABLE 8 RELATIVE RATING OF SELECTED BRAZING PROCESS HEATING METHODS

METHOD	**CHARACTERISTICS**[(A)]					
	CAPITAL COST	**RUNNING COST**	**BASIC OUTPUT**	**FLUX REQUIRED**	**VERSATILITY**	**OPERATOR SKILL REQUIRED**
Torch (flame)	L/M	M/H	L	YES	H	YES
Electrical resistance	M	M	M/H	YES	L	NO
Induction	M/H	M	M/H	Y/N	M	NO
Furnace (atmosphere)	M/H	M/H	H	Y/N	M	NO
Furnace (vacuum)	H	L	H	NO	M	NO
Dip (flux bath)	L/M	M/H	L/M	YES	L	YES
Infrared	M	L	M	Y/N	L	NO

(A) H, HIGH; M, MEDIUM; L, LOW

The size and value of individual assemblies, the numbers required, and the required rate of production will influence the selection of heating method. Other factors that also must be considered include the rate of heating, differential thermal gradients, and both external and internal cooling rates. These factors vary tremendously with different methods of heating, and their effects on dimensional stability, distortion, and joint structure must be considered.

Manual torch brazing is the method most frequently used for repairs, one-of-a-kind brazing jobs, and short production runs, as an alternative to fusion welding. Any joint that can be accessed by a torch and brought to brazing temperature (by the torch alone or in conjunction with auxiliary heating) can be readily brazed by this technique.

Although any flame-producing device can be used for torch brazing, commercial applications are accomplished with the same type of torch, controls, and gases used for torch fusion welding. Conversion to brazing merely requires changes in torch nozzles and goggle lenses.

The torch brazing technique is relatively simple and can be mastered by the mechanically adept in a short time. Those already experienced in torch welding and the brazing of other metals generally encounter little difficulty learning torch brazing.

Depending on the temperature and heat required, all commercial gas mixtures can be used to fuel the torch: oxyacetylene, oxyhydrogen, oxy-natural gas, acetylene and air, hydrogen and air, propane, methane, and natural gas and air. Oxyacetylene and oxy-natural gas are the mixtures most often used commercially and are preferred in that order. Flame adjustment is very important. Generally, a slightly reducing flame is desirable in order to preserve part surfaces from oxidation.

The oxyacetylene combination produces the highest temperature. The other gases are cooler and their flames have relatively lower intensity. Thus, they are easier to use and are applied advantageously to light-gage material.

Manual torch brazing is particularly useful on assemblies with sections of unequal mass. As warranted by the rate of production, machine operations can be set up using one or more torches equipped with single or multiple flame tips. The machine can be designed to move either the work or the torches.

Torch brazing can be rather easily automated with appropriate gas supplies, indexing fixtures, and cycle controls. Usually, such systems involve multiple-station rotary indexing tables. The part is fed into a holding fixture at the first station and is then indexed to one or more preheating stations, depending on the heating time required. A brazing station is next, followed by a cooling station and an ejection station (see Fig. 4).

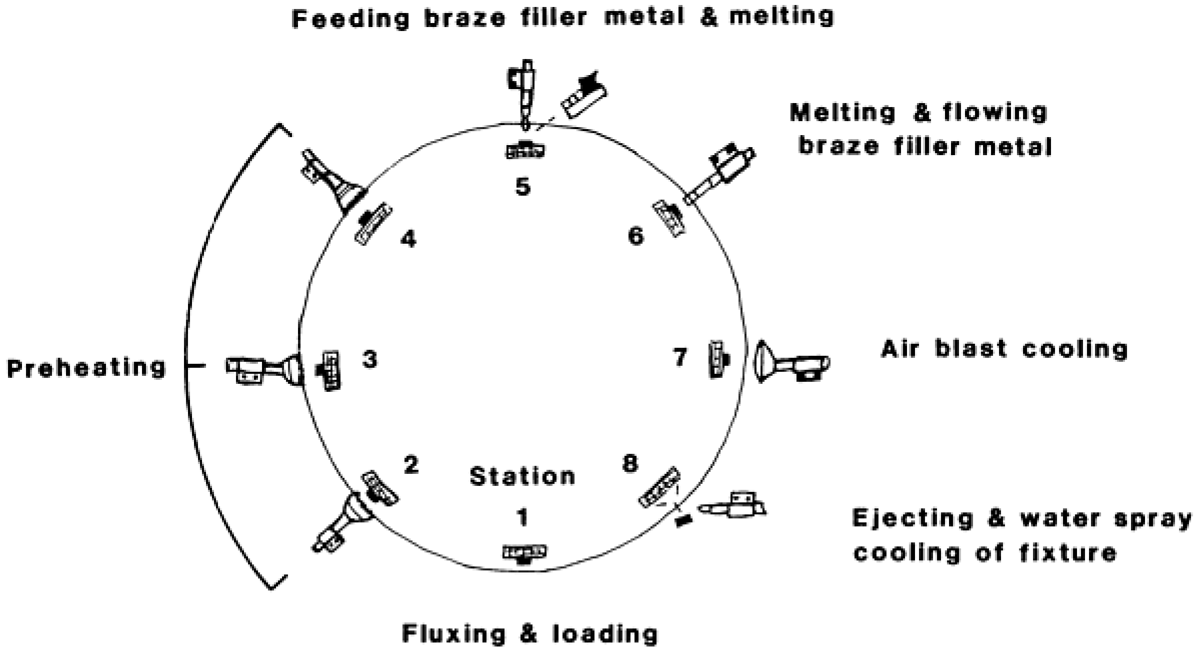
Feeding braze filler metal & melting
Melting & flowing braze filler metal
Preheating
Air blast cooling
Ejecting & water spray cooling of fixture
Fluxing & loading
Station
1
2
3
4
5
6
7
8

FIG. 4 SCHEMATIC OF AN EIGHT-STATION AUTOMATIC TORCH BRAZING OPERATION THAT PRODUCES MAGNET ARMATURE ASSEMBLIES (230 PIECES/H) USED AS STRIKING MEMBERS OF A PRINTING MACHINE

Additional information is available in the article "Torch Brazing" in this Volume.

Furnace brazing is popular because of its comparatively low equipment cost, furnace adaptability, and minimal required jigging. Furnace brazing is a low-cost process relative to other processes such as torch brazing, induction brazing, or salt-bath brazing when high-volume production output is a primary factor. Secondary factors include tooling requirements, fluxing, and cleaning requirements. With many brazing assemblies, the weight of the parts alone is sufficient to hold them together. Other configurations require only one or two rectangular fixturing blocks of metal.

Four basic types of furnaces are used:

- BATCH, WITH EITHER AIR OR CONTROLLED ATMOSPHERE (FIG. 5)
- CONTINUOUS, WITH EITHER AIR OR CONTROLLED ATMOSPHERE (FIG. 6)
- RETORT, WITH CONTROLLED ATMOSPHERE
- VACUUM

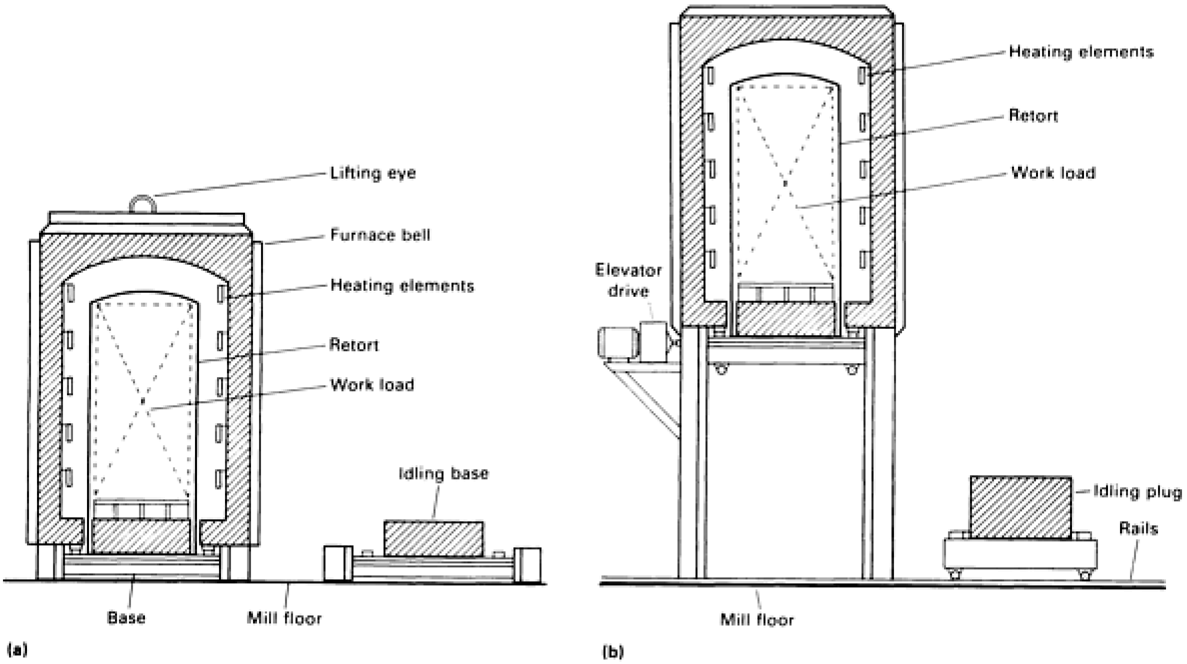
Lifting eye
Furnace bell
Heating elements
Retort
Work load
Idling base
Base
Mill floor
(a)
Heating elements
Retort
Work load
Elevator drive
Idling plug
Rails
Mill floor
(b)

FIG. 5 BATCH-TYPE BRAZING FURNACES. (A) BELL FURNACE. (B) ELEVATOR FURNACE

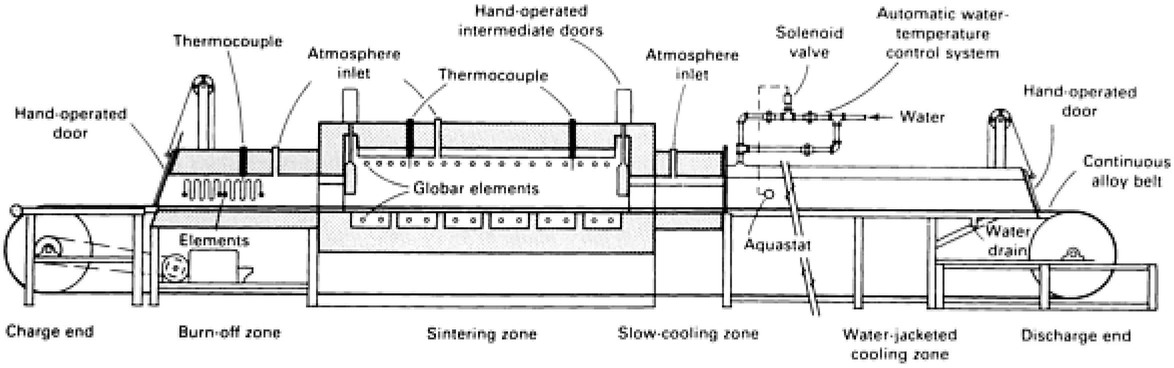
Hand-operated door
Thermocouple
Atmosphere inlet
Hand-operated intermediate doors
Thermocouple
Atmosphere inlet
Solenoid valve
Automatic water-temperature control system
Water
Hand-operated door
Continuous alloy belt
Globar elements
Elements
Aquastat
Water drain
Charge end
Burn-off zone
Sintering zone
Slow-cooling zone
Water-jacketed cooling zone
Discharge end

FIG. 6 CONTINUOUS-TYPE BRAZING FURNACE WITH WATER-JACKETED COOLING CHAMBER FOR USE IN AIR OR CONTROLLED ATMOSPHERES

Furnace brazing is used extensively where the parts to be brazed can be assembled with the brazing filler metal preplaced near or in the joint. Furnace brazing is particularly applicable for high-production applications in which continuous conveyor-type furnaces are used. For medium-production work, batch-type furnaces are best. Regardless of furnace type, heating is usually produced by electrical resistance, although other types of fuel can be used in muffle-type furnaces.

The parts should be self jigged or fixtured and assembled, with brazing filler metal preplaced near or in the joint. Fluxing is used, except in cases where a reducing atmosphere, such as hydrogen, and either exothermic or endothermic combusted gas can be introduced into the furnace. Sometimes, both flux and a reducing atmosphere are necessary. Pure and dry inert gases, such as argon and helium, are used to obtain special atmosphere properties. Fluxes should never be used when furnace brazing with a vacuum atmosphere.

When continuous-type furnaces are used (Ref 16, 17), several different temperature zones can be set up to provide the proper preheating, brazing, and cooling temperatures. The speed through a conveyor-type furnace must be controlled to provide the appropriate time at the brazing temperature. It is also necessary to properly support the assembly so that it does not move while traveling on the belt (Ref 18).

A large volume of furnace brazing is performed in a vacuum atmosphere, which prevents oxidation during the heating cycle and eliminates the need for flux (Ref 19) (Fig. 7). Vacuum brazing has found wide application in the aerospace and nuclear fields, where reactive metals are joined or where entrapped fluxes would not be tolerable. Vacuum brazing does not allow as wide a choice of brazing filler metals as does atmosphere brazing (Ref 20, 21).

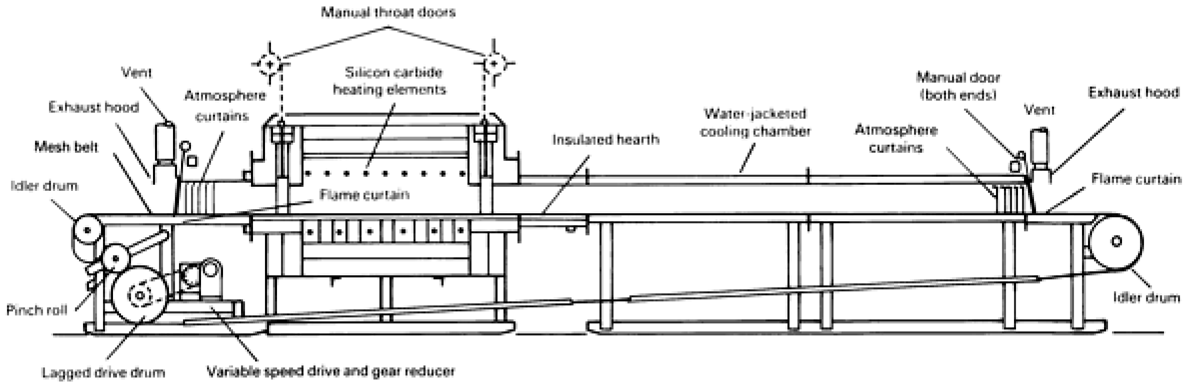
Manual throat doors
Vent
Silicon carbide heating elements
Exhaust hood
Atmosphere curtains
Mesh belt
Idler drum
Flame curtain
Insulated hearth
Water-jacketed cooling chamber
Manual door (both ends)
Vent
Exhaust hood
Atmosphere curtains
Flame curtain
Idler drum
Pinch roll
Lagged drive drum
Variable speed drive and gear reducer

FIG. 7 HIGH-VACUUM BOTTOM-LOADING VACUUM FURNACE FOR MEDIUM-PRODUCTION APPLICATIONS

In most cases, vacuum furnaces are heated by electrical resistance heaters in a numbers of different forms. The overall time of a vacuum brazing operation is usually much longer than that of other batch brazing techniques and involves typing up expensive equipment for long periods to process comparatively small workloads. However, most of the metals on which vacuum brazing excels are also costly, the use of a process that permits material economics through fabrication while ensuring the necessary joint properties often can be justified (Ref 22, 23, 24).

Detailed information is available in the article "Furnace Brazing" in this Volume.

Induction Brazing. The high-frequency induction heating method for brazing is clean and rapid, lends itself to close control of temperature and location, and requires little operator skill.

The workpiece is placed in or near a coil carrying alternating current (ac), which induces the heating current in the desired area (Fig. 8). The coils, which are water cooled, are designed specifically for each part. Therefore, heating efficiency relies on establishing the best coil design and power frequency for each application. In most cases, the coils provide heat only to the joint area.

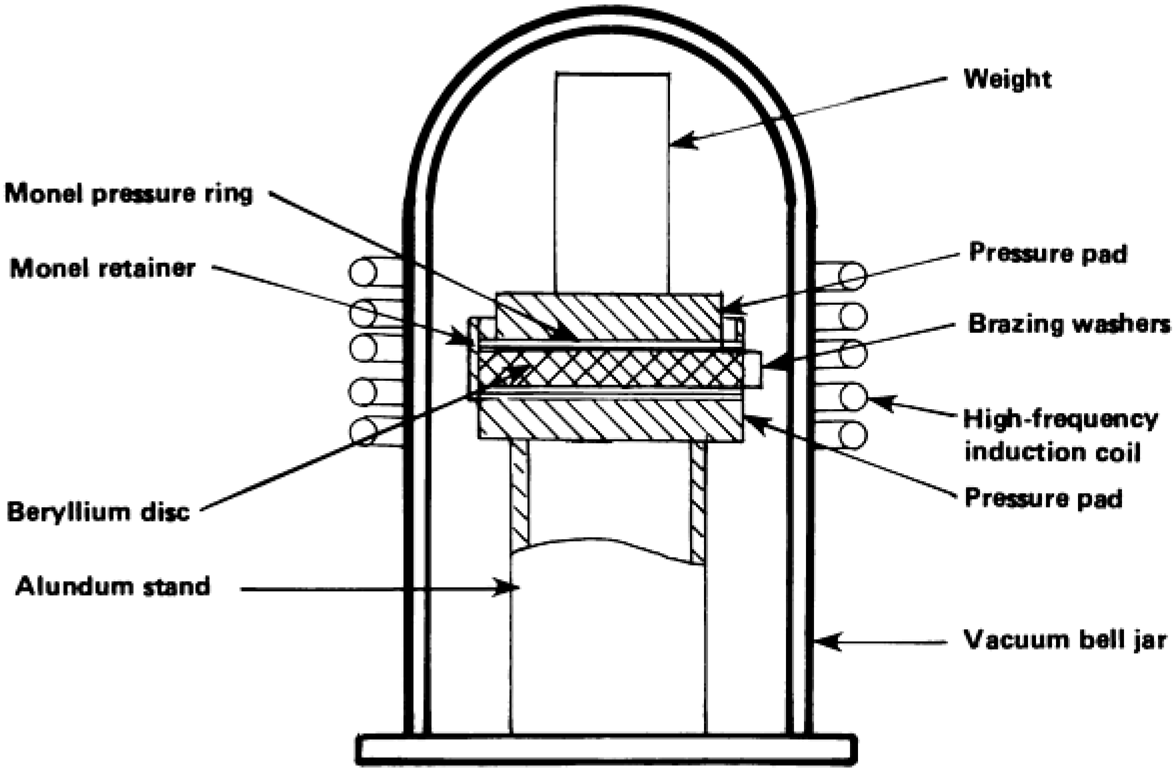

FIG. 8 INDUCTION HEATING APPARATUS USED TO VACUUM BRAZE BERYLLIUM AND MONEL COMPONENTS

The capability to heat selectively enables the induction method to be used when the nature of the brazement demands localized brazing.

There are several types of high-frequency sources to supply power to induction coils, each with a different range of frequencies:

- MOTOR GENERATORS (5 TO 10 KHZ)
- SPARK-GAP OSCILLATORS (20 TO 30 KHZ)
- VACUUM-TUBE OSCILLATORS (20 TO 5000 KHZ)
- SOLID-STATE POWER SUPPLIES (VARIABLE HZ)

The frequency of the power source determines the type of heat that will be induced in the part: high-frequency sources produce "skin" heating, whereas lower-frequency sources provide heating in thicker areas. The brazing heat usually develops within 10 to 60 s.

Induction brazing is well suited for mass production. Mechanized brazing lines for moving assemblies to and from the coils are very common. Brazing filler metals is normally preplaced in the joint, and the brazing can be done in air with the use of a flux, in an inert atmosphere, or in a vacuum atmosphere. The rapid heating rates available with induction heating are a major advantage when using brazing filler metals that tend to vaporize or segregate. The heating cycle for induction brazing is invariably automated, even when manual loading of assemblies is used.

Additional information is available in the article "Induction Brazing" in this Volume.

Dip brazing involves immersion of assembled parts into a suitable molten bath to effect brazing. The bath can be molten brazing filler metal, molten chemical flux, or molten chemical salts. The molten material is contained in a "pot" furnace heated by oil, gas, or electricity. In some instances, electrical-resistance heaters are used in the bath (Fig. 9). In the first type of bath, the parts being joined are held together and immersed in flux-covered molten filler metal, which flows into the joints when the parts reach the bath temperature. The flux cleans the workpiece as it is introduced and protects the brazing filler metal by preventing oxidation and the loss of volatile elements from the bath.

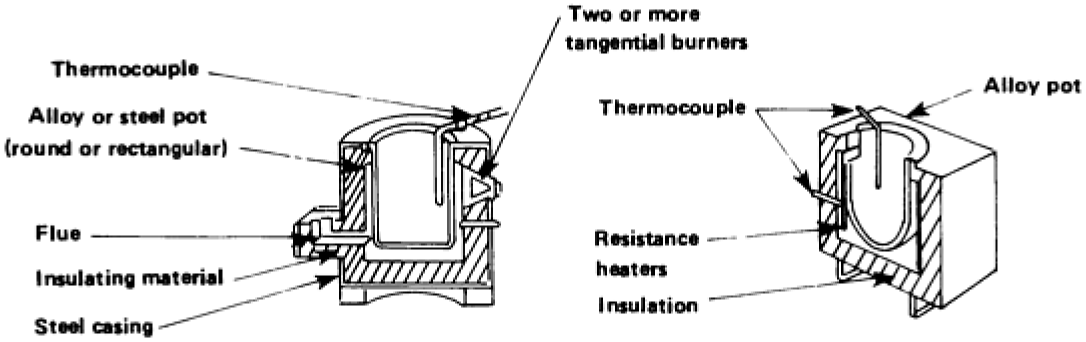
Thermocouple
Two or more tangential burners
Alloy or steel pot (round or rectangular)
Flue
Insulating material
Steel casing
Thermocouple
Alloy pot
Resistance heaters
Insulation

(a) Gas fired or oil fired

(b) Resistance heated

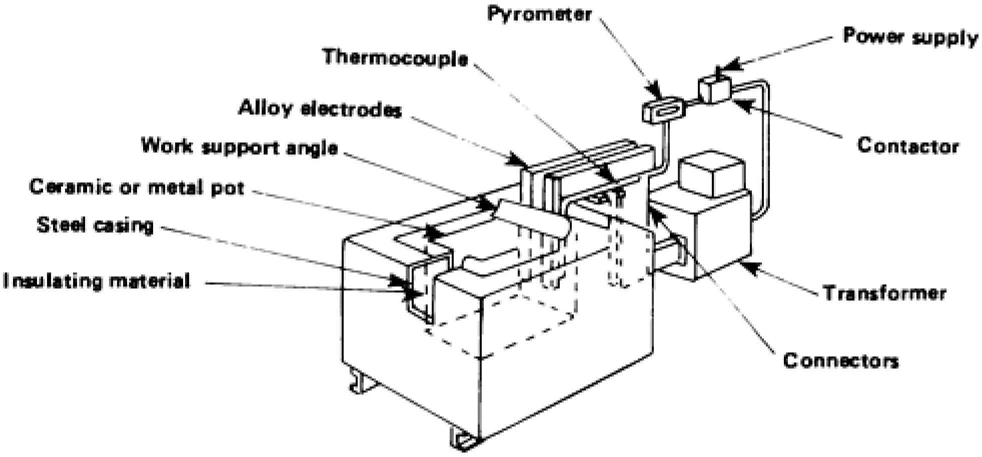
Pyrometer
Power supply
Thermocouple
Alloy electrodes
Contactor
Work support angle
Ceramic or metal pot
Steel casing
Insulating material
Transformer
Connectors

(c) Immersed electrodes

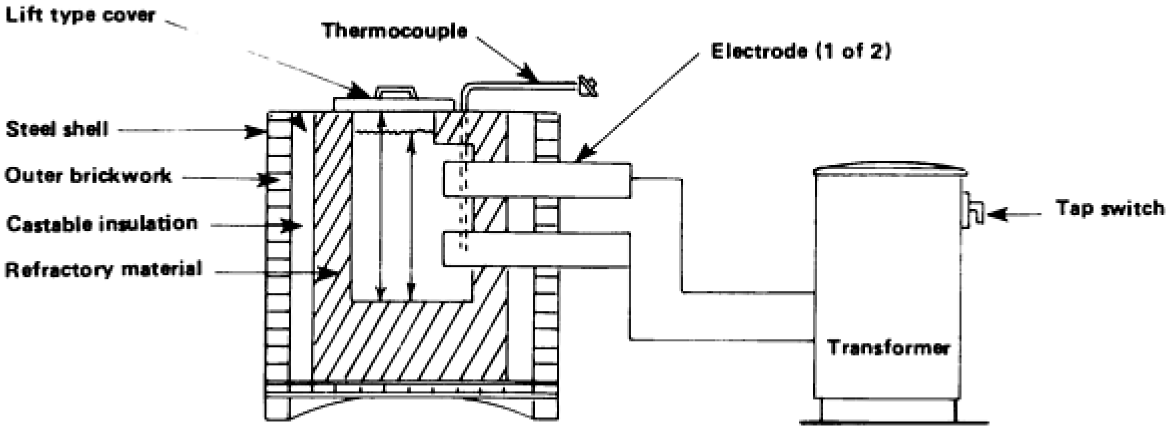
Lift type cover
Thermocouple
Electrode (1 of 2)
Steel shell
Outer brickwork
Castable insulation
Refractory material
Tap switch
Transformer

(d) Electrically heated (submerged electrodes)

FIG. 9 EXTERNALLY AND INTERNALLY HEATED FURNACES USED FOR DIP BRAZING APPLICATIONS

The pot, or crucible, used in the molten-metal method is usually made of graphite. Jigging to maintain alignment is generally necessary. Because of the difficulties of heating and containing metals at high temperatures, filler metals that require a brazing temperature above 1000 °C (1830 °F) are rarely used. Therefore, the choice of brazing filler metal is restricted to straight brasses and silver-base filler metals.

In the molten-flux method, the brazing filler metal is located in or near the joints and is heated to the required temperature by immersion in a bath of molten flux. Molten salt bath brazing has a greater scope than any other single brazing process. It can be used on as wide a range of parent metals as torch brazing, but is not subject to the same maximum temperature limitations. Unfortunately, it is an inflexible process. The type of salt used for a particular application depends on the ease with which the parent-metal surface oxides can be removed and on the temperature required for brazing.

The molten-flux method is used extensively for brazing aluminum and its alloys. The flux bath provides excellent protection against metal reoxidation, which can occur quite easily with aluminum.

The dip brazing method generally causes less distortion than torch brazing, because of its uniform heating. However, it may require relatively complex tooling and is therefore best used in medium- to high-production runs. This process is particularly well suited for small- to medium-sized parts with multiple hidden joints.

Detailed information is available in the article "Dip Brazing" in this Volume.

Resistance brazing is most applicable to relatively simple joints in metals that have high electrical conductivity. In this process, the workpieces are heated locally. Brazing filler metal that is preplaced between the workpieces is melted by the heat obtained from its resistance to the flow of electric current through the electrodes and the work. Usually, the heating current, which is normally ac, is passed through the joint itself. The joint becomes part of an electrical circuit, and the brazing heat is generated by the resistance at the joint.

Equipment is the same as that used for resistance welding, and the pressure needed for establishing electrical contact across the joint is ordinarily applied through the electrodes. The electrode pressure also in the usual means for providing the tight fit needed for capillary behavior in the joint. The component parts are generally held between copper or carbon-graphite electrodes.

The flux that is used must be conductive. Normally, fluxes are insulators when cool and dry, but are conductive when wet. The process is generally used for low-volume production in joining electrical contacts, related electrical elements, copper commutator segments, stainless steel tube to fittings, and so on.

Additional information is available in the article "Resistance Brazing" in this Volume.

Infrared (Quartz) Brazing. The development of high-intensity quartz lamps and the availability of suitable reflectors have made the infrared heat technique commercially important for brazing. Infrared heat is radiant heat that is obtained with the sources having energy frequency below the red rays in the light spectrum. Although there is some visible light with every "black" source, heating is principally done by the invisible radiation. Heat sources (lamps) capable of delivering up to 5000 W of radiant energy are commercially available. The lamps do not need to follow the contour of the part in order to hear the joint area, even though the heat input varies inversely as the square of the distance from the source, unless reflectors are used to concentrate the heat.

Lamps are often arranged in a toaster-like configuration, with parts traveling between two banks of lamps (Fig. 10). Infrared brazing can concentrate large amounts of heat in small areas, which can be advantageous in certain applications. Infrared brazing setups are generally not as fast as induction brazing, but the equipment is less expensive.

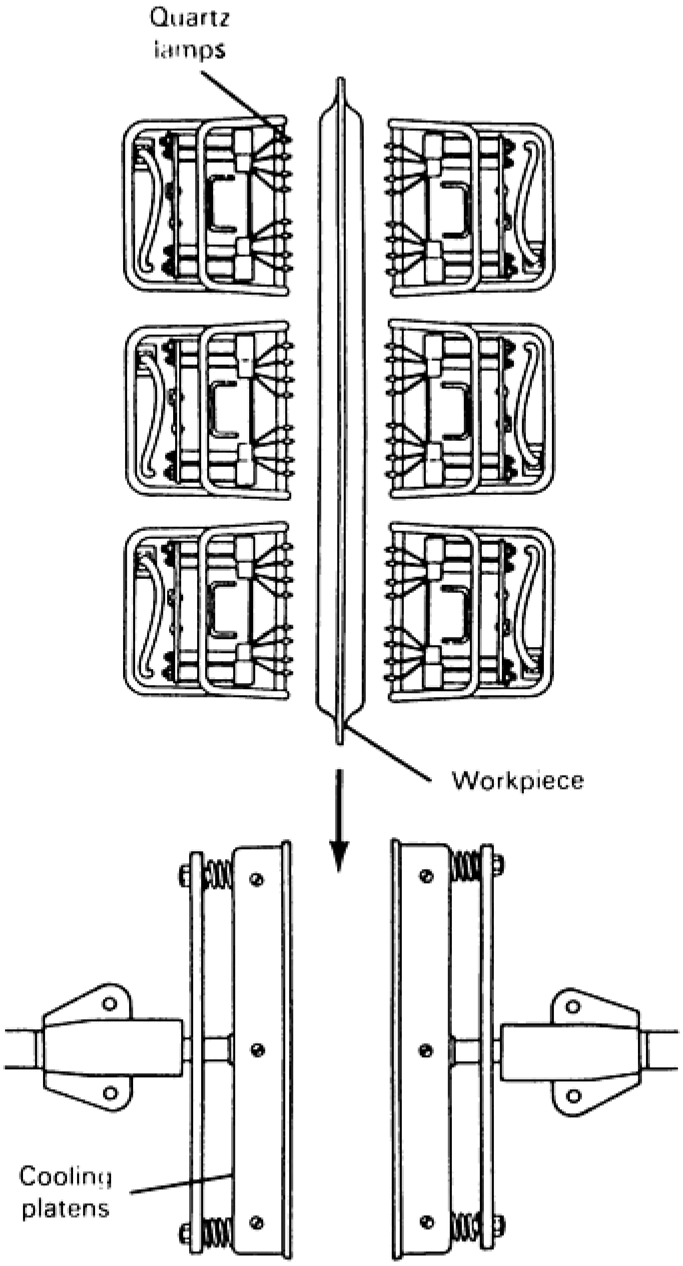

FIG. 10 WORKPIECE SUBJECTED TO INFRARED BRAZING PROCESS PRIOR TO BEING MOVED TO COOLING PLATENS

Exothermic brazing is a special process, but is rarely used because there are more-economical production methods. In this process, the heat required to melt and flow the brazing filler metal is produced by a solid-state exothermic chemical reaction, which is defined as any reaction between two or more reactants that involves the free energy of the system. Although nature provides countless numbers of these reactions, only the solid-state or nearly solid-state metal/metal-oxide reactions are suitable for use in exothermic brazing units.

Additional information is available in the article "Exothermic Brazing" in this Volume.

Electron-beam and laser brazing have been used in a limited number of applications. A laser can be used when a very small localized area of heat is required, such as for brazing small carbide tips on printer heads for electronic printers. The electron beam can be used for brazing by defocusing the beam to provide a wider area of heating. Because this is done in a vacuum, fluxes cannot be used and the brazing filler metal must be selected so that there is little or no vaporization during brazing.

Microwave Brazing. One of the newest heating methods to be developed is the use of microwaves. This technique is being used to join ceramics in high-temperature, corrosion-resistant, and high-performance applications. These applications require complex ceramics parts with strong, durable joints (Ref 25, 26).

The technique uses a single-mode microwave cavity. An iris controls the percent of microwaves reflected in the cavity, and a plunger adjusts the frequency. Together, they focus microwaves on the joint. The major advantage of this technique in terms of ceramic parts is that the entire part does not have to be heated in order to make a joint. Reheating finished ceramic parts can subject them to thermal stresses and cause cracking, which weakens the part. With microwave joining, only the interface between the pieces is heated. The technique is faster than conventional joining, which requires long times to heat ceramics uniformly. Large pieces, can, in principle, be joined at a lower cost.

One disadvantage of microwave joining is that commercial equipment is only produced by a few companies, mostly for research and development purposes.

Some microwave units feature 2.45 GHz single-mode or controlled multimode operation. Their cavity is designed to provide controlled-mode patterns that can focus high-energy microwave fields to heat a part only where it is needed, such as at a joint. The joint or part can be heated up to 2200 °C (3990 °F). Researchers have joined 92% Al_2O_3 ceramic parts in a 6 GHz single-mode microwave cavity at various temperatures and times. The bending strength of this joint gradually increased from the solidus temperature of 1400 °C (2550 °F) and reached a peak of 420 MPa (60 ksi) at about 1750 °C (3180 °F).

Braze welding is a joining process in which a filler metal is melted and deposited in a specific joint configuration, and a metallurgical bond is obtained by a wetting action that is often accompanied by some degree of diffusion with the base metals. Braze welding requires heating and melting of the filler metal that has a melting (liquidus) temperature above 450 °C (840 °F).

Stringent fit-up is not critical, because the filler metal is deposited in grooves and spaces and flows into gaps wider than those used for brazing. Fabricators use braze welding as a low-temperature substitute for oxyfuel welding or as a low-cost substitute for brazing. Joint designs for braze welding are the same as for oxyfuel welding (OFW). Braze welding has been used to join cast iron, steels, copper, nickel, and nickel alloys.

Compared to conventional fusion-welding processes, braze welding requires less heat input, permits higher travel speeds, and causes less distortion. Deposited filler metal is soft and ductile, providing good machinability, and residual stresses are low. The process joins brittle cast irons without extensive preheating.

Although most braze welding initially used an OFW torch, copper filler metal brazing rod, and a suitable flux, present applications use carbon arc welding (CAW), gas-tungsten arc welding (GTAW), gas-metal arc welding (GMAW), or plasma arc welding (PAW) without flux in the manual, semiautomatic, or automatic modes to economically bond and deposit the filler metal in the braze weld joints. OFW is, however, still widely used in machinery repair applications. Filler metal selection, proper wetting and compatibility with the base metals, and shielding from air are important to the effective use of the process with any suitable heating method.

A wide variety of parts can be braze welded using typical weld joint designs. Groove, fillet, and edge welds can be used to join simple and complex assemblies made from sheet, plate, pipe, tubing, rods, bars, castings, and forgings. Sharp corners that are easily overheated and may become points of stress concentrations should be avoided. To obtain good strength, an adequate bond area is required between filler metal and the base metal. Weld groove geometry should provide an adequate groove face area, so that the joint will not fail along the interfaces. Proper joint design selection will produce deposited filler-metal strengths that may meet or exceed the minimum base-metal tensile strengths. Because of the inert shielding gas, electrical arc methods have fewer included flux compounds and oxides at the faying surfaces. The result is higher joint strength and improved corrosion resistance. Original surfaces are restored by overlayments and subsequent machining.

References cited in this section

16. W.S. STEFFAN, NITROGEN-BASED ATMOSPHERE FURNACE BRAZING SYSTEM PROVIDES STRENGTH AND LEAK-PROOF PERFORMANCE OF NEWLY DESIGNED FUEL INJECTION RAIL,

IND. HEAT., AUG 1989, P 21-23

17. *METALS HANDBOOK*, 9TH ED., *POWDER METALLURGY*, VOL 7, 1984, P 358
18. *METALS HANDBOOK*, 8TH ED., VOL 6, 1971, P 595
19. PRODUCT LITERATURE, SECO-WARWICK CORP. (SUNBEAM), SEPT 1977
20. A. SAKAMOTO, STUDY OF FURNACE ATMOSPHERE FOR VACUUM-INERT GAS PARTIAL-PRESSURE BRAZING, *WELD. J.*, NOV 1991, P 311S-320S
21. NEW GENERATION OF VACUUM HEAT-TREATMENT FURNACES FEATURE HIGH GAS PRESSURE QUENCHING AND OIL QUENCH VACUUM FURNACE, *METALLURGIA*, MAY 1989, P 205
22. W.R. JONES, VACUUM--ANOTHER ATMOSPHERE?, *HEAT TREAT.*, OCT 1986, P 39-41
23. W.T. HOOVEN AND K.W. NOKES, MAKE YOUR VACUUM BRAZING FURNACES USER FRIENDLY, *WELD. J.*, OCT 1990, P 25-29
24. S. DRESSLER, ROLE OF HOT WALL VACUUM FURNACES IN PLASMA-ASSISTED SURFACE TREATMENT: PART II, OPERATING INFLUENCES, *IND. HEAT.*, SEPT 1989, P 31-33
25. "ADVANCED STRUCTURAL CERAMICS: TECHNOLOGIES, ECONOMICS AND MARKET OPPORTUNITIES," GB-107, BCC INC., DEC 1987
26. E.J. KUBEL, JR., STRUCTURAL CERAMICS; MATERIALS OF THE FUTURE, *ADV. MATER. PROCESS.*, VOL 134, AUG 1988, P 25-33

Fundamentals of Brazing

Mel M. Schwartz, Sikorsky Aircraft

References

1. C.M. ADAMS, JR., "DYNAMICS OF WETTING IN BRAZING AND SOLDERING," TECHNICAL REPORT WAL TR 650/1, ARMY MATERIALS RESEARCH AGENCY, WATERTOWN ARSENAL, JULY 1962
2. S. WEISS AND C.M. ADAMS, JR., THE PROMOTION OF WETTING, *WELD. J.*, VOL 46 (NO. 2), FEB 1967, P 49S-57S
3. W.D. HARKINS, *PHYSICAL CHEMISTRY OF SURFACE FILMS*, REINHOLD, 1952
4. M.M. SCHWARTZ, *MODERN METAL JOINING TECHNIQUES*, WILEY, 1969
5. A. BONDI, SPREADING OF LIQUID METALS ON SOLID SURFACE CHEMISTRY OF HIGH ENERGY SUBSTANCES, *CHEM. REV.*, VOL 52 (NO. 2), 1953, P 417-458
6. R.G. GILLILAND, WETTING OF BERYLLIUM BY VARIOUS PURE METALS AND ALLOYS, *WELD. J.*, VOL 43, JUNE 1964, P 248S-258S
7. M. NAKA, T. TANAKA, AND I. OKAMOTO, *Q. J. JPN. WELD. SOC.*, VOL 4, 1986, P 597
8. R.R. KAPOOR AND T.W. EAGER, *J. AM. CERAM. SOC.*, VOL 72, 1989, P 448
9. R.E. LOCHMAN, *CERAM. BULL.*, VOL 68, 1989, P 891
10. D.H. KIM, S.H. HWANG, AND S.S. CHUN, THE WETTING, REACTION, AND BONDING OF SILICON NITRIDE BY CU-TI ALLOYS, *J. MATER. SCI.*, VOL 26, 1991, P 3223-3234
11. D.L. KELLER *ET AL.*, WETTABILITY OF BRAZING FILLER METALS, *WELD. J.*, VOL 69 (NO. 10), 1990, P 31-34
12. W.S. BENNETT *ET AL.*, VACUUM BRAZING STUDIES ON HIGH-MANGANESE STAINLESS STEEL, *WELD. J.*, VOL 53, 1974, P 510S-516S
13. W. FEDUSKA, HIGH-TEMPERATURE BRAZING ALLOY--BASE METAL WETTING REACTIONS, *WELD. J.*, VOL 38 (NO. 3) 1959, P 122S-130S

14. W.H. KOHL, SOLDERING AND BRAZING, *VACUUM,* VOL 14, 1964, P 175-198
15. *METALS HANDBOOK,* 9TH ED., *WELDING, BRAZING, AND SOLDERING,* VOL 6, AMERICAN SOCIETY FOR METALS, 1983, P 929-995
16. W.S. STEFFAN, NITROGEN-BASED ATMOSPHERE FURNACE BRAZING SYSTEM PROVIDES STRENGTH AND LEAK-PROOF PERFORMANCE OF NEWLY DESIGNED FUEL INJECTION RAIL, *IND. HEAT.,* AUG 1989, P 21-23
17. *METALS HANDBOOK,* 9TH ED., *POWDER METALLURGY,* VOL 7, 1984, P 358
18. *METALS HANDBOOK,* 8TH ED., VOL 6, 1971, P 595
19. PRODUCT LITERATURE, SECO-WARWICK CORP. (SUNBEAM), SEPT 1977
20. A. SAKAMOTO, STUDY OF FURNACE ATMOSPHERE FOR VACUUM-INERT GAS PARTIAL-PRESSURE BRAZING, *WELD. J.,* NOV 1991, P 311S-320S
21. NEW GENERATION OF VACUUM HEAT-TREATMENT FURNACES FEATURE HIGH GAS PRESSURE QUENCHING AND OIL QUENCH VACUUM FURNACE, *METALLURGIA,* MAY 1989, P 205
22. W.R. JONES, VACUUM--ANOTHER ATMOSPHERE?, *HEAT TREAT.,* OCT 1986, P 39-41
23. W.T. HOOVEN AND K.W. NOKES, MAKE YOUR VACUUM BRAZING FURNACES USER FRIENDLY, *WELD. J.,* OCT 1990, P 25-29
24. S. DRESSLER, ROLE OF HOT WALL VACUUM FURNACES IN PLASMA-ASSISTED SURFACE TREATMENT: PART II, OPERATING INFLUENCES, *IND. HEAT.,* SEPT 1989, P 31-33
25. "ADVANCED STRUCTURAL CERAMICS: TECHNOLOGIES, ECONOMICS AND MARKET OPPORTUNITIES," GB-107, BCC INC., DEC 1987
26. E.J. KUBEL, JR., STRUCTURAL CERAMICS; MATERIALS OF THE FUTURE, *ADV. MATER. PROCESS.,* VOL 134, AUG 1988, P 25-33

Fundamentals of Brazing

Mel M. Schwartz, Sikorsky Aircraft

Selected References

- *BRAZING HANDBOOK,* 4TH ED., AWS, 1991
- M.M. SCHWARTZ, *BRAZING,* ASM INTERNATIONAL, 1987
- M.M. SCHWARTZ, *CERAMIC JOINING,* ASM INTERNATIONAL, 1990

Fundamentals of Soldering

Mel M. Schwartz, Sikorsky Aircraft

Introduction

SOLDERING is an ancient joining method. It is mentioned in the Bible (Isaiah, 41:7), and there is evidence of its use in Mesopotamia 5000 years ago, well before the time of Cleopatra, as well as later, in Egypt, Greece, and Rome (Ref 1). Pliny the Elder, in his *Historia Naturalis,* written 2000 years ago, mentions (in Chapter XLVIII of Book XXXIV) that the soldered connections of the pipes of the Roman aqueducts were made with a so-called "tertiarium" mixture, an alloy of two parts lead and one part tin. The earliest solders were alloys found in nature, which meant that only a few solders, with a severely limited range of properties, were available. Because these early solders were generally used to join jewelry parts or to attach handles to decorative vessels, the primary concerns were appearance, melting point, and, to a lesser degree, strength. The materials were rare and costly, and the work was done by highly skilled artisans. Therefore, only the wealthy could hope to own such articles.

It has only been in the last two centuries that some metals have become cheap and that strictly utilitarian parts are soldered. In the 1800s, low-cost steels made household tinware practical for most families, and the "tinner" became the common solder practitioner. The emergence of electrical technology required the attachment of electrical leads, which became a most common use of solders. This meant that the electrical properties of solders became a consideration, and the electrical continuity of the joint was of paramount importance, rather than strength or appearance, as had been the case with jewelry or utensils. The use of solders for electrical attachments also involved simple mechanical joints and repetitive operations. The cost of the materials was generally much less, because lead-tin solders were usually used. Even today, it is customary to give the tin content first when designating a solder (for example, 40/60 solder refers to a solder that is 40% Sn and 60% Pb). These new criteria promoted a new type of artisan whose skills were more oriented to engineering and production than to aesthetics, although most of the work was still done manually.

By the 20th century, metallurgical science had developed to the point that new solders could be tailored specifically for electrical, plumbing, or structural applications. The emerging electronics industry required solders having the following properties:

- COMPATIBILITY WITH COPPER (ESPECIALLY WITH RESPECT TO ALLOYING BEHAVIOR AND MELTING TEMPERATURES)
- GOOD ELECTRICAL CONDUCTIVITY
- WORKABILITY TO ENABLE FACTORY WORKERS TO RAPIDLY FORM LOW-COST, RELIABLE SOLDER JOINTS

Engineers now had formal rules to govern the design of joints used in various applications to ensure the required levels of strength. The materials and processes involved in soldering were now becoming established in engineering practice (Ref 2, 3, 4, 5).

References

1. J. WOLTERS, *ZUR GESEHICHTE DER LOTTECHNIK,* DEGUSSA, WEST GERMANY, 1977
2. H.H. MANKO, *SOLDERS AND SOLDERING,* 2ND ED., MCGRAW-HILL, 1964
3. R.J. KLEIN WASSINK, *SOLDERING IN ELECTRONICS,* ELECTROCHEMICAL PUBLICATIONS LTD., 1989
4. L.P. LAMBERT, *SOLDERING FOR ELECTRONIC ASSEMBLIES,* MARCEL DEKKER, 1987
5. *SOLDERING MANUAL,* AWS, 1978

Mel M. Schwartz, Sikorsky Aircraft

Process Overview

Soldering is defined as a joining process by which two substrates are bonded together using a filler metal (solder) with a liquidus temperature that does not exceed 450 °C (840 °F). The substrate materials remain solid during the bonding process. The solder is usually distributed between the properly fitted surfaces of the joint by capillary attraction.

The bond between solder and base metal is more than adhesion or mechanical attachment, although these do contribute to bond strength. Rather, the essential feature of the soldered joint is that a metallurgical bond is produced at the filler-metal/base-metal interface. The solder reacts with a small amount of the base metal and wets the metal by intermetallic compound formation. Upon solidification, the joint is held together by the same attraction, between adjacent atoms, that holds a piece of solid metal together. The ease of wetting is related to the ease with which this solvent action occurs. The presence of the base-metal/filler-metal reaction is one factor in the wetting action of the solder. Other factors include surface cleanliness and solder surface tension (that is, capillary flow).

Fundamentals of Soldering

Mel M. Schwartz, Sikorsky Aircraft

Fundamentals of Soldering

Pure metals have a crystalline state and a well-defined melting temperature. They also possess reasonable ductility and strength. Metals are excellent electrical conductors, when compared with ceramics or plastic polymer materials. Thermal conductivity generally follows electrical conductivity (that is, a good electrical conductor also is a good thermal conductor).

Melting and Solidification of Pure Metals. The melting behavior of pure lead is plotted in Fig. 1. From point A, which is room temperature, heat is applied to raise the temperature to 327 °C (621 °F), the melting point of lead, represented as point B, whereupon the heat of fusion is absorbed and the temperature remains constant until all the lead becomes liquid. At point C, when all of the solid is molten, the temperature of the liquid will again rise (to point D) upon the application of heat.

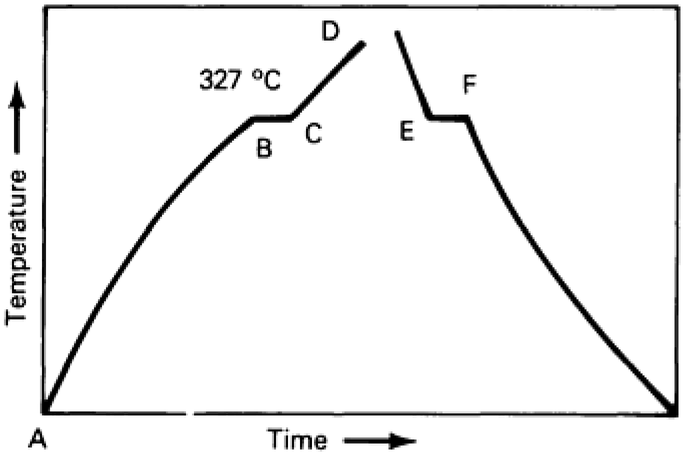

FIG. 1 PLOT OF THE HEATING AND COOLING CURVES OF PURE LEAD. SEE TEXT FOR DETAILS.

If the process is reversed and the heat source is turned off, then the temperature will fall from point D to E, where the liquid will begin to give off its heat of fusion, as atoms assume a solid crystalline form. A nucleus is formed and gradually grows as the atoms attach themselves in a regular crystalline pattern, forming dendritic crystals, which have a main trunk and branches. As the solid forms, the release of the heat of fusion causes the temperature to remain constant (point E to F) at 327 °C (621 °F), until all the lead has solidified. Then, the solid cools back to room temperature. Under equilibrium conditions, lead will always be in the liquid phase at temperatures above 327 °C (621 °F), and will always be in the solid phase below that temperature. This behavior is observed for all pure metals, although each has a specific melting temperature. The solidification and cooling curve for pure tin is shown in Fig. 2. The cooling arrest at the solidification temperature is typical of all pure metals (Ref 3).

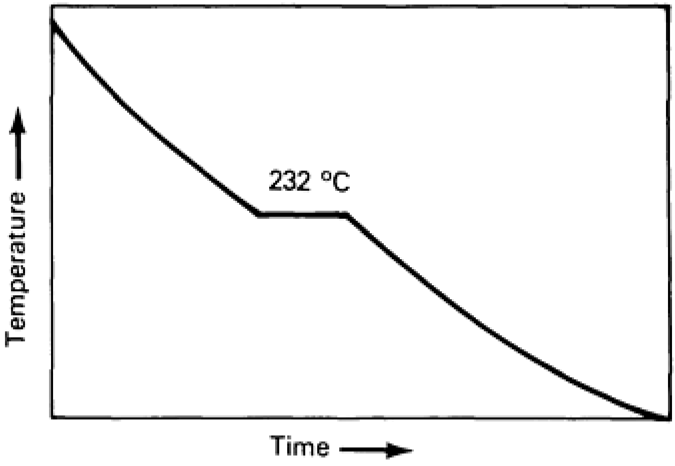

FIG. 2 COOLING CURVE FOR PURE TIN

Alloy Formation and Phase Diagrams. The best vehicle for portraying the behavior of alloys is the equilibrium phase diagram (also called equilibrium or constitution diagram), which represents the phases present under equilibrium conditions in a given alloy system. Phase diagrams are plots of alloy equilibrium composition versus temperature. The diagrams are generally worked out by measuring the melting and solidification behaviors of alloys of different compositions and then observing the resultant microstuctures. Understanding phase diagrams is essential to working effectively with solders, because they are used to predict solder microstructure and behavior.

When two or more metals are melted together, they form an alloy that behaves as a unique material with specific properties that can significantly differ from the properties of the individual pure metals. Alloy properties depend on the atomic structure and thermal-physical properties of its constituent elements. For example, the addition of tin to lead will result in an alloy that has a lower melting point than either tin or lead. Perhaps more surprising is that the addition of lead to tin will also result in an alloy with a lower melting point. The physical and mechanical properties of the alloys will also differ from the properties of the pure metals.

In Fig. 3, 327 °C (621 °F) is the melting point of pure lead (point A) and 232 °C (450 °F) is the melting point of pure tin (point B). The line AEB, which separates the liquid field from the rest of the diagram, represents the "liquidus." For any given composition in this system, the alloy is completely molten at any temperature above that line. The line below which any lead-tin alloy will be solid is called the "solidus" (represented by ACEDB).

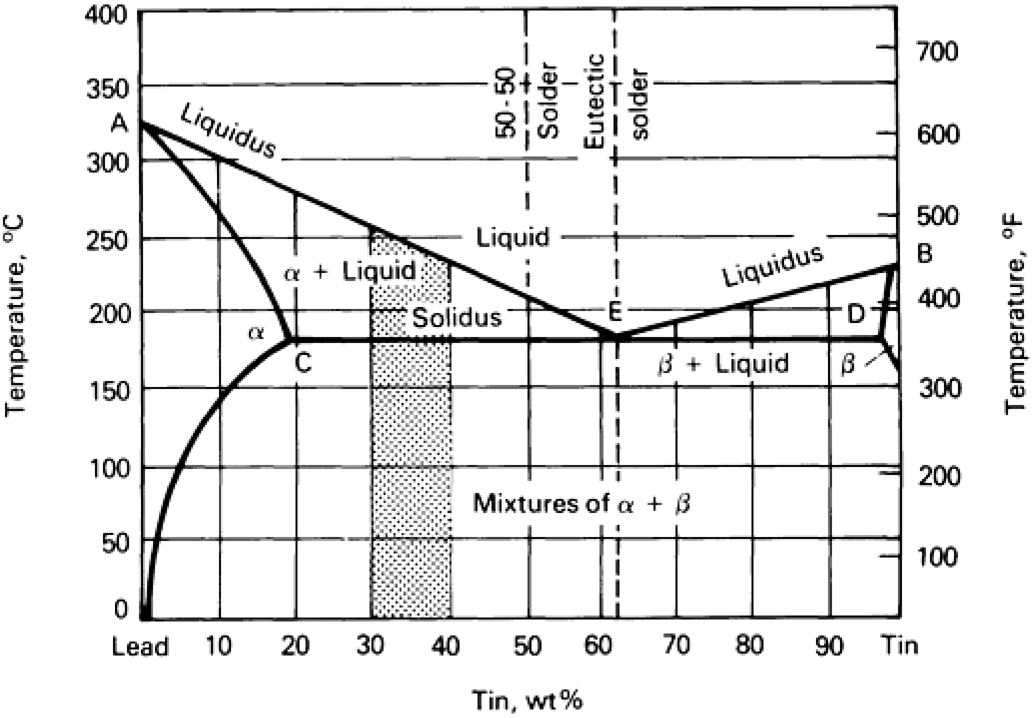

FIG. 3 CONSTITUTIONAL DIAGRAM FOR THE TIN-LEAD ALLOY SYSTEM

The CEB segment of this line is a special case known as the "eutectic" temperature. It should be noted that the fields on the phase diagram are labeled to indicate the phases existing in those regions, which enables the quick determination of the state of an alloy or the phases present at a temperature in that region.

Some metals, such as gold and silver, form alloys simply because of their mutual solubility. Such alloy formation is called a solid solution.

Two metals that have somewhat greater differences in their basic properties may not dissolve completely, that is, they have limited solubility in one another. The combination of silver and copper, however, causes a large drop in the melting temperatures, forming a eutectic (low melting) alloy system. In addition, the eutectic alloy composition itself exhibits a unique type of melting and solidification behavior, melting and solidifying at a single, specific temperature just as a pure metal does. The solid eutectic alloy is a mixture comprising two phases, rather than a solid solution. Eutectic alloy structures and behavior are of critical importance in soldering.

Wettability of Metals by Solder. Solder wetting necessarily involves the metallurgical reactions between the filler metal and the base metal. This interaction at the solder/base-metal interface can result in a covalently bonded layer of material, referred to as an intermetallic compound. Examples include tin and gold ($AuSn_4$), copper and tin (Cu_6Sn_5 and Cu_3Sn), or nickel and tin (Ni_3Sn_4). As opposed to the metallic bond of metals and alloys, the covalently bonded structure causes the intermetallic to be hard and brittle, to have a high melting point, and to be resistant to chemical attack. Excessively hard intermetallic layers can jeopardize the physical and mechanical integrity of solder joints.

All electrical/electronic applications require soldering to a base metal. This base metal can be the metal that makes up the physical structure of the electrical/electronic lead or it can constitute an underplate or a barrier metallization. In all cases, the metal must be metallurgically wettable by the solder.

Intermetallic compounds are more similar to chemical compounds than they are to metallic alloys. In soldering, intermetallic compounds in thin and uniform layers represent the "glue" that forms the structural bond between the solder and the metal being soldered. Therefore, the base metal must be metallurgically compatible with at least one of the metallic components in the solder.

This intermetallic compound formation can occur by a solid/liquid reaction (molten solder against a solid base metal) or a solid-state diffusion reaction (solder plate against a solid base metal). In either case, the solder and base metal react to form a film of intermetallic compound between the two metals, which holds them together. When tin-lead solders are used with copper, a widely used base metal, the tin reacts with the copper.

Two intermetallics discussed previously in this article can form in this metallurgical system. The compound Cu_6Sn_5 will primarily form during liquid/solid reactions. In addition, Cu_6Sn_5 continues to grow in the solid state at elevated temperatures. The compound Cu_3Sn will be present during any solid-state reactions. All base-metal solder systems form one or more such compounds. For example, soldering to nickel instead of copper will result in tin-nickel intermetallics (Ni_3Sn_4) at the interface. The Ni_3Sn_4 layer grows very slowly in the solid state and is typically difficult to observe.

The ability of base metal to form a stable intermetallic compound layer with a solder limits the number of base-metal/solder combinations that are suitable for the majority of applications. The formation kinetics of intermetallic layers must permit the thorough wetting of the solder for efficient processing. Therefore, some metals are more readily soldered than others.

Another factor is the passivation characteristic of the base metal. Passivation layers form a physical barrier to metallurgical wetting. Before the base metal can be metallurgically wet by the solder, the solder must make intimate contact, on an atomic scale, with clean base metal. Passivation layers are normally oxide films that may be only a few atomic layers thick. Even so, they affect the ability of the solder to contact the base metals underneath them. Many metals will passivate extremely rapidly. Stainless steel becomes unwettable, even with the use of very active fluxes after a short exposure in the atmosphere at room temperature.

Another issue that relates to passivation layers is their chemical and mechanical stability. Because the passivation layer that forms on solder is chemically weak, it can be reduced by weak fluxes. Nickel, on the other hand, forms a passivation film that is strong both chemically and mechanically. It is adherent to the base metal and requires much stronger fluxes to remove chemically.

Figure 4 shows the phase diagram for lead-tin. The solid solution of tin in lead is designated by α, whereas the solution of lead in tin at the other extreme of the diagram is designated by β. The line ADF, which separates the α from the rest of the diagram, is the solvus, as is the similar line (CEG) isolating the β phase.

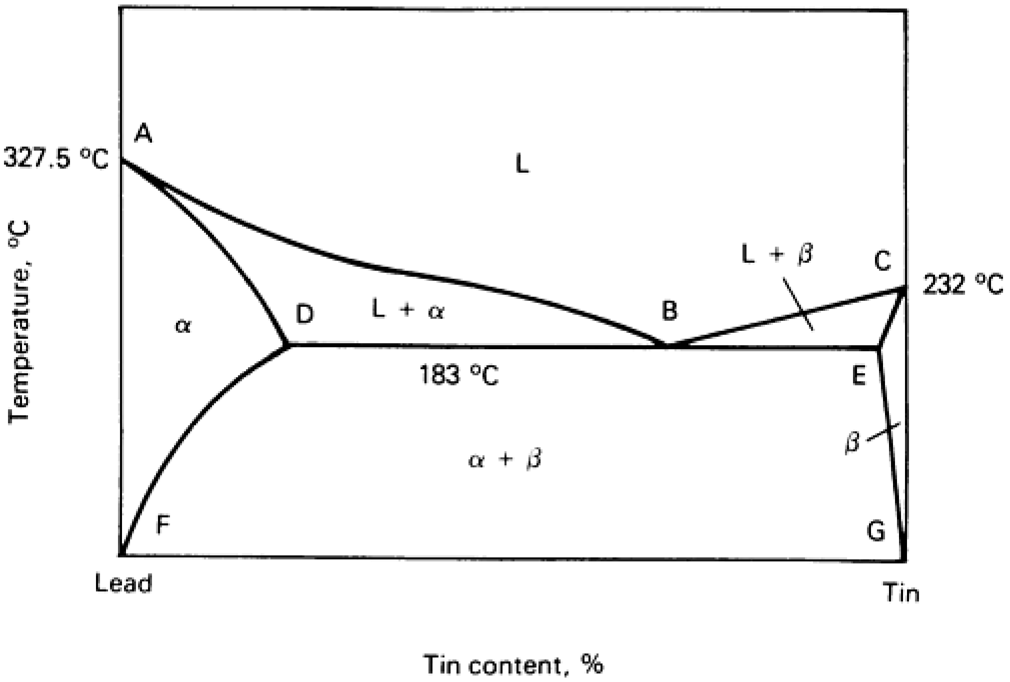

FIG. 4 LEAD-TIN PHASE DIAGRAM SHOWING EUTECTIC TEMPERATURE AT POINT B WITH 63SN-37PB

The segments D-F and E-G denote the limits of solid solubility of tin or lead in the α or the β solid solution, respectively. If the amount of tin in an alloy exceeds the solid solubility as shown by D-F, then a second phase, β, will precipitate. The amount of precipitation depends on the amount by which the Sn exceeds the solubility limit at that temperature.

Effect of Wetting Phenomena on Solderability. Solderability is a property of electronic and other components that is crucial both to the efficiency of manufacture and the reliability of the product (Ref 6). Good component solderability is important for three basic reasons:

- IT ALLOWS THE USE OF LESS-ACTIVE FLUXES, THEREBY REDUCING THE REQUIREMENT FOR CLEANING FLUX RESIDUES. (THIS IS AN ENVIRONMENTAL BENEFIT. IN ADDITION, THE REMOVAL OF FLUX RESIDUE ELIMINATES POTENTIAL CORROSION PROBLEMS THAT ARE DUE TO THAT RESIDUE.)
- IT PRODUCES GREATER "FIRST-PASS" SOLDERING YIELDS AND, CONSEQUENTLY, LESS HAND WORKING OF THE SOLDER JOINTS (IMPORTANT BECAUSE REWORKING CAN REDUCE THE FATIGUE PERFORMANCE OF A JOINT, COMPARED WITH A SUCCESSFUL FIRST-PASS JOINT).
- IT RESULTS IN A GREATER UNIFORMITY OF SOLDER FILLET, WITH A GEOMETRY THAT IS CLOSE TO THE IDEAL FOR MAXIMUM FATIGUE PERFORMANCE.

Solderability depends on the wettability of the two surfaces being joined. Poor component solderability can, to some extent, be overcome by using more-active fluxes, but the trend toward denser component packing makes it difficult to remove flux residue after assembly. Therefore, less-active fluxes are preferred. Environmental concerns (such as those about the use of chlorofluorinated carbon solvents) are also becoming important. The cost of inspection, testing, and component replacement or solder fillet rework increases the desirability of defect-free soldering. Therefore, solderability is an increasingly important manufacturing issue.

The solderability of a component, defined as its suitability to be soldered by a given method, is a complex processing parameter. The relationship between wettability and solderability represents a major unresolved gap in the understanding of the soldering process. As described in the section "Physics of Wetting," wettability can be measured directly. Solderability, however, cannot be directly measured in a quantifiable way, and its relationship to wettability has not yet been explicitly defined.

There are three important aspects to solderability:

- THERMAL DEMAND
- WETTABILITY
- RESISTANCE TO SOLDERING HEAT

The thermal characteristics of the component must allow the joint to be heated to the soldering temperature within the specified time. The solderable surfaces must allow the molten solder to wet and spread during the available time without subsequent dewetting. The soldering heat and the induced thermal stresses associated with it must not affect the functioning of the components. Each of these solderability aspects can be engineered to fit a particular application by the suitable choice of solder alloys and process control parameters. The most restraining of the solderability aspects, in regard to design for performance, is component wettability.

Wettability. When discussing the wetting characteristics of a surface by molten solder, there are two important factors to consider: the extent of wetting and the rate of wetting (Ref 6). The extent of wetting (as indicated by the contact angle) is an equilibrium governed by the laws of thermodynamics and depends on the surface and interfacial energies involved at the liquid/solid interface (Fig. 5). The rate of wetting (that is, how rapidly the solder wets and spreads) is governed by the thermal demand of the system, the ability of the heat source to supply heat, the efficacy of the flux, the viscosity of the solder, and the chemical reactions that occur at the interfaces.

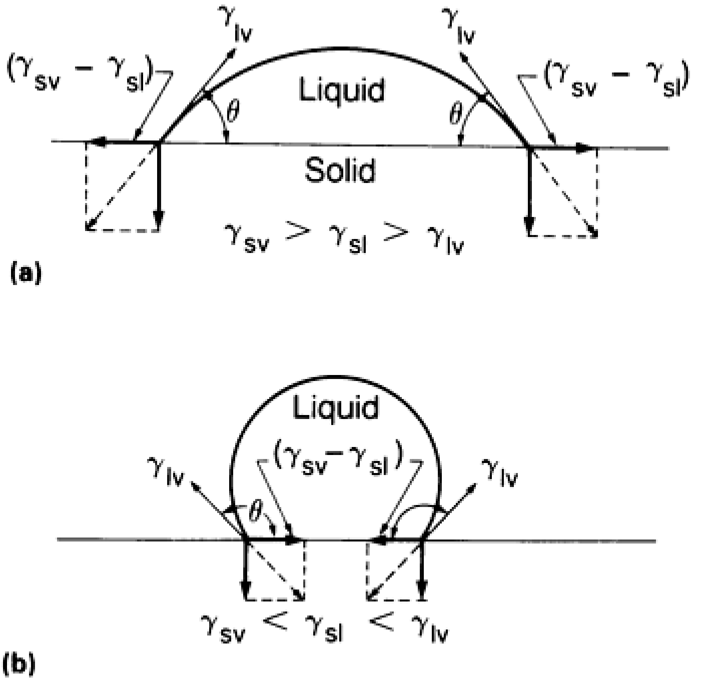
γ_{lv}
γ_{lv}
$(\gamma_{sv} - \gamma_{sl})$
$(\gamma_{sv} - \gamma_{sl})$
Liquid
θ
θ
Solid
$\gamma_{sv} > \gamma_{sl} > \gamma_{lv}$
(a)
Liquid
$(\gamma_{sv} - \gamma_{sl})$
γ_{lv}
γ_{lv}
θ
$\gamma_{sv} < \gamma_{sl} < \gamma_{lv}$
(b)

FIG. 5 LIQUID SOLDER DROPLETS ON A SOLID SURFACE UNDER TWO CONDITIONS. (A) WETTING. (B) NONWETTING. ARROWS INDICATE SURFACE TENSION.

The physics of wetting is the governed by Young's equation (also known as Young-Depre equation because Young described the problem quantitatively in 1805 and Depre put it in mathematical terms in 1869):

$$\gamma_{SV} = \gamma_{SL} - \gamma_{LV} \text{ COS } \theta \qquad \textbf{(EQ 1)}$$

where γ_{sv}, γ_{sl}, and γ_{lv} refer to the solid-vapor, solid-liquid, and liquid-vapor surface tensions, respectively. Relative to each other, these parameters can be defined by:

$$\gamma_{SV} > \gamma_{SL} > \gamma_{LV} \qquad \textbf{(EQ 2)}$$

For soldering, the vapor phase will be replaced in nearly all processes by flux (that is, γ_{sv} is replaced by γ_{sf} and γ_{lv} is replaced by γ_{lf}).

For optimum wetting, the contact angle, θ, must be minimized and, therefore, γ_{sv} must be maximized. Oxides and contamination lower the value of γ_{sv} (or γ_{sf}). A function of the flux is to remove or disperse oxides and contaminants, thereby increasingly γ_{sv}. An important function of flux is to lower the value of γ_{lf} (or γ_{lv}). The flux also maintains a local environment around the surfaces being joined, protecting them from reoxidation during soldering, as well as enhancing heat transfer from the heating source to the substrate and the solder.

Equation 1 shows that wetting is also enhanced by minimizing γ_{lv}, a value commonly referred to as the surface tension of the solder. Although γ_{lv} is a function of the solder composition, the flux covering the liquid solder, and the temperature, it is normally close to 0.4 J/m^2 (4×10^{-5} Btu/ft^2).

A typical example of nonwetting occurs when a liquid metal droplet is placed on a nonmetallic surface, such as copper oxide. Because there are no metallic bonds at the oxide surface, the liquid metal will have little tendency to interact with that surface. The metal droplet will behave as though it is repelled, and will try to ball up to minimize the area of contact with the nonmetallic surface. One method for quantifying wettability is to observe the behavior of solder droplets on the metal surface. The contact angle formed between the surface of the liquid solder and the surface of the solid can be measured (Fig. 5). If the solder droplet forms an angle, θ, of less than 90°, then the solder is said to wet the metal. If the angle is greater than 90°, the solder is nonwetting. This is one method used to qualitatively assess solderability.

The formation of intermetallic compounds can affect Young's equation because it alters the value of γ_{sl}. The influence of this term has been recognized (Ref 6), but the magnitude has not yet been calculated quantitatively.

Surfaces of a metal crystal have a higher free energy than the bulk of the crystal because of the existence of unsatisfied metallic bonds. When a liquid metal or alloy is brought into contact with this solid surface, the liquid will proceed to interact with the solid to satisfy those dangling surface bonds and thus reduce the surface free energy of the system. To the extent that a liquid solder can satisfy these surface bonds, it will wet the solid metal and spread across its surface. If no exchange of bonding energy takes place, then wetting will not occur and the solder will tend to ball up on the surface of the solid, minimizing the area of contact between liquid and solid.

Information about selecting solders for specific applications is provided in the article "General Soldering" in this Volume.

References cited in this section

3. R.J. KLEIN WASSINK, *SOLDERING IN ELECTRONICS,* ELECTROCHEMICAL PUBLICATIONS LTD., 1989
6. A.D. ROMIG, JR., Y.A. CHANG, J.J. STEPHENS, *ET AL.*, PHYSICAL METALLURGY OF SOLDER-SUBSTRATE REACTIONS, *SOLDER MECHANICS,* THE MINERALS, METALS, & MATERIALS SOCIETY, 1991, P 30-32

Mel M. Schwartz, Sikorsky Aircraft

Guidelines for Flux Selection

A flux promotes solder wetting of the base materials by:

- REMOVING TARNISH FILMS FROM PRECLEANED SURFACES
- PREVENTING OXIDATION DURING THE SOLDERING OPERATION
- LOWERING THE SURFACE TENSION OF THE SOLDER

Successful establishment of a solder joint requires that the liquid solder make contact with the metal to which it is to be joined, so that wetting can be initiated. Unfortunately, almost all of the metals involved in soldering are oxidized during exposure to elevated temperatures in air. This prevents metal-to-metal contact, as well as the wetting and formation of a metallurgical bond, unless the oxides are removed. Fluxes are used in soldering to remove such films and to protect the surfaces against reoxidation during soldering.

In addition to cleaning metal surfaces, a flux also:

- AIDS IN THE WETTING PROCESS
- IMPROVES HEAT TRANSFER (ESPECIALLY IN SOLDER PASTES)
- CARRIES OXIDES AND SOLID DEBRIS AWAY FROM THE JOINT

Fluxes are chemical (liquid, solid, or gaseous materials) that remove oxide layers from the base metal and solder. When heated, fluxes either promote or accelerate the wetting of metals by solder. Flux selection usually depends on the ease with which a material can be soldered. Rosin fluxes are used with solderable base metals or with metals that are precoated with a solderable base metal or with metals that are precoated with a solderable finish. Inorganic fluxes are often used on metals such as stainless steel. Table 1 indicates the relative ease with which a number of alloys and metals can be soldered, based on flux requirements.

TABLE 1 RELATIVE SOLDERABILITY OF SELECTED METALS AND ALLOYS AS A FUNCTION OF FLUX TYPE USED

Base Metal, Alloy, Or Applied Finish	Flux Type				Soldering Not Recommended[(A)]
	ROSIN	ORGANIC	INORGANIC	SPECIAL FLUX AND SOLDER	
Aluminum	. . .	. . .	. . .	X	. . .
Aluminum-Bronze	. . .	. . .	. . .	X	. . .
Beryllium	. . .	. . .	. . .	. . .	X
Beryllium-Copper	X	X	X	. . .	. . .
Brass	X	X	X	. . .	. . .
Cadmium	X	X	X	. . .	. . .
Cast Iron	. . .	. . .	. . .	X	. . .
Chromium	. . .	. . .	. . .	. . .	X
Copper	X	X	X	. . .	. . .
Copper-Chromium	. . .	. . .	X	. . .	. . .
Copper-Nickel	X	X	X	. . .	. . .
Copper-Silicon	. . .	. . .	X	. . .	. . .
Gold	X	X	X	. . .	. . .
Inconel	. . .	. . .	. . .	X	. . .
Lead	X	X	X	. . .	. . .
Magnesium	. . .	. . .	. . .	. . .	X
Manganese-Bronze (High Tensile)	. . .	. . .	. . .	. . .	X
Monel	. . .	X	X	. . .	. . .
Nickel	. . .	X	X	. . .	. . .
Nickel-Iron	. . .	X	X	. . .	. . .
Nichrome	. . .	. . .	. . .	X	. . .
Palladium	X	X	X	. . .	. . .
Platinum	X	X	X	. . .	. . .
Rhodium	. . .	. . .	X	. . .	. . .
Silver	X	X	X	. . .	. . .
Stainless Steel	. . .	X	X	. . .	. . .
Steel	. . .	. . .	X	. . .	. . .
Tin	X	X	X	. . .	. . .
Tin-Bronze	X	X	X	. . .	. . .
Tin-Lead	X	X	X	. . .	. . .
Tin-Nickel	. . .	X	X	. . .	. . .
Tin-Zinc	X	X	X	. . .	. . .
Titanium	. . .	. . .	. . .	. . .	X
Zinc	. . .	X	X	. . .	. . .
Zinc Die Castings	. . .	. . .	. . .	. . .	X

(A) WITH PROPER PROCEDURES, SUCH AS PRECOATING, MOST METALS CAN BE SOLDERED

Types of Fluxes

Fluxes can be categorized by their chemical makeup (Ref 2, 3, 9, 10, 11):

- ROSIN-BASE FLUXES
- ORGANIC FLUXES
- INORGANIC FLUXES
- SYNTHETICALLY ACTIVATED FLUXES (RESINS)

These materials vary in their activity (that is, aggressiveness). Rosin-base fluxes are very mild, whereas inorganic fluxes can be extremely active and corrosive. Because of the potential for corrosion by inorganic fluxes in electronic assemblies, the rosin-base fluxes (with their low flux activity) are favored by the electronics industry.

Rosin Fluxes. Water-white rosin dissolved in a suitable organic solvent is the closest model of a flux with noncorrosive residue. Rosin fluxes possess important physical and chemical properties that make them particularly suitable for use in the electrical industry. They are solid and inactive chemically at room temperature, but at soldering temperatures they gain sufficient activity to remove weakly adherent oxides from the noble metals gold and silver, as well as copper.

Nonactivated Rosin Flux. Although its noncorrosive nature has led to its widespread use in microelectronics, rosin-base flux is lacking in chemical activity, which essentially limits its use to precleaned parts and to only a few metals that do not have adherent oxides. The active constituent, abietic acid ($C_{20}H_{30}O_2$), with a melting point of 173 °C (343 °F), becomes mildly active at soldering temperatures ranging from 177 to 316 °C (350 to 601 °F). The residue is hard, nonhygroscopic, electrically nonconductive, and noncorrosive.

Mildly Activated Rosin Flux. The very low chemical activity of rosin-base flux can be countered by the addition of activators, usually organic acids or amines. The end product, rosin mildly activated (RMA) flux, is still essentially noncorrosive, but is sufficiently active to remove oxides more reliably than the rosin-base flux. In addition, the residues are noncorrosive. However, care must be exercised in cases of high-reliability applications. It is advisable to specify that these fluxes be halide-free, in order to ensure that specific safety requirements are met in critical components. Even so, cleaning is now required to remove the flux residue. Both polar and nonpolar solvents have been used. These RMA fluxes are preferred for military, telephone, and other high-reliability electronic products.

Activated rosin fluxes were developed to provide more chemically active fluxes for mass-produced electronics, such as packaged components. Most mass-produced electronics are manufactured using RMA fluxes. The use of chlorides in these fluxes requires effective cleaning after soldering to prevent corrosion and electrical leakage, because the presence of chloride ions in flux residue makes it conductive. Because of the chemistry of the activators, a double solvent cleaning is normally required. Activated rosin fluxes are widely used in commercial electronics and in high-reliability applications where the residue can be completely removed after soldering.

Inorganic Fluxes. Inorganic acids and salts that are highly corrosive and extremely active compose this class of fluxes. In many instances, they are not acceptable for use on electronics, but are suitable for plumbing and industrial applications. Inorganic fluxes do enable the soldering of ferrous alloys and high-nickel alloys used in electronic packages and hermetic enclosures. The difficulty of chloride ion removal has led to the gradual abandonment of inorganic fluxes, even for tinning purposes.

Inorganic fluxes are used to optimum advantage where conditions require rapid and highly active fluxing action. They can be applied as solutions, pastes, or dry salts in many general soldering applications. They function equally well with torch, oven, resistance, or induction soldering methods, because they neither char nor burn. These fluxes can be formulated to provide stability over the entire soldering temperature range.

One distinct disadvantage of inorganic fluxes is that their residue remains chemically active after soldering. If this residue is not removed, then severe corrosion can occur at the joint. Adjoining areas can also be attacked by residues from the spray of flux and from flux vapors. Fluxes that contain ammonium salts can cause stress-corrosion cracking (SCC) when soldering brass, as well as most iron-nickel alloys.

Organic acid fluxes, although less active than inorganic fluxes, are effective at soldering temperatures ranging from 90 to 320 °C (195 to 610 °F). They consist of organic acids and bases and, often, hydrohalides. They are active at soldering temperatures, but the period of activity is short because of their susceptibility to thermal decomposition. Their tendency to volatilize, char, or burn when heated limits their use with torch or flame heating. When these fluxes are properly used, their residues are relatively inert and can be removed with water.

Organic acid fluxes are particularly useful in applications where controlled quantities of flux can be applied and where sufficient heat can be used to fully decompose or volatilize the corrosive constituents. Caution is necessary to prevent undecomposed flux from spreading to insulating sleeving. Care must also be taken when soldering in closed systems where corrosive fumes can condense on critical parts of the assembly.

References cited in this section

2. H.H. MANKO, *SOLDERS AND SOLDERING,* 2ND ED., MCGRAW-HILL, 1964
3. R.J. KLEIN WASSINK, *SOLDERING IN ELECTRONICS*, ELECTROCHEMICAL PUBLICATIONS LTD., 1989
9. J.F. SHIPLEY, INFLUENCE OF FLUX, SUBSTRATE AND SOLDER COMPOSITION ON SOLDER WETTING, *WELD. J.,* VOL 54 (NO. 10), OCT 1975, P 357S-362S
10. H.H. MANKO, SOLDERING FLUXES--PAST AND PRESENT, *WELD. J.,* VOL 52 (NO. 3), MARCH 1973, P 163-166
11. R.W. WOODGATE, *THE HANDBOOK OF MACHINE SOLDERING,* 2ND ED., JOHN WILEY & SONS, 1988

Flux Evaluation

Several tests are used to evaluate the relative activity of fluxes. The most common are the copper mirror test, the halide test, a surface conductivity test, and a test based on the resistivity of a water extract present in the flux. In all cases, a measure of the chemical activity of the flux is obtained.

Because the fluxes all depend on other properties (for example, viscosity, ease of removal, and thermal behavior), chemical activity is only one factor in flux selection for a given application. Information about selecting fluxes for specific applications is provided in the article "General Soldering" in this Volume.

Joint Design

Joints should be designed to fulfill the requirements of the finished assembly, as well as to permit the application of the flux and solder. Joint design should maintain proper clearance during heating and upon solidification of the filler metal. Special fixtures may be necessary or the units can be crimped, clinched, wrapped, or otherwise held together.

The selection of a joint design for a specific application will primarily depend on the service requirements of the assembly. It may also depend on such factors as the heating method to be used, the fabrication techniques utilized prior to soldering, the number of items to be soldered, and the method used to apply the solder. In general, solders have low strength when compared with the metals for which they are used to join. Therefore, the soldered joint should be designed to avoid dependence on solder strength. The necessary strength can be provided by shaping the parts to be joined so that they engage or interlock, requiring the solder only to bond, seal, and stiffen the assembly.

Figure 6 shows joint designs commonly used for soldering applications. The lap joint and the lock seam joint are used when soldering sheets. Lap joints are also applied to join pipes. Actually, the lap type of joint should be used whenever possible because it offers the best chance to obtain joints with maximum strength. It should also be used whenever a seal is required. Butt joints should be avoided whenever possible.

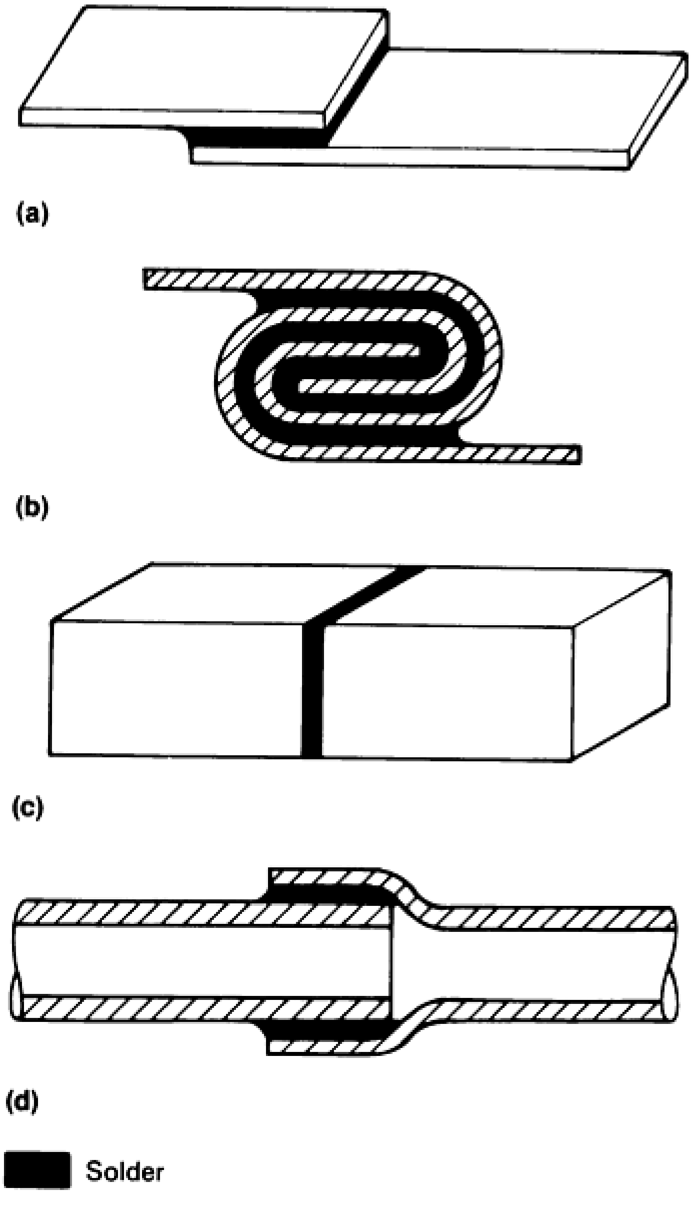

FIG. 6 BASIC JOINT CONFIGURATIONS USED FOR SOLDERING APPLICATIONS. (A) LAP JOINT. (B) LOCK SEAM JOINT. (C) BUTT JOINT. (D) PIPE JOINT

An important factor in joint design is the manner in which the solder will be applied to the joint. The designer must consider the number of joints per assembly and the number of assemblies to be manufactured. For limited production using a manual soldering process, the solder can be face-fed into the joint with few problems. However, for a large production lot of assemblies containing multiple joints, an automated process such as condensation/vapor-phase reflow soldering can be advantageous (Ref 12, 13, 14). In this case, the design must provide for accessible joints that are suitable for automated fluxing, soldering, and cleaning.

There should be enough clearance between the parts being joined to allow the solder to be drawn into the space by capillary action, but not so much that the solder is unable to fill the gap. Joint clearances ranging from 0.075 to 0.150 mm (0.0030 to 0.059 in.) are preferred for optimum strength, but variations are allowed in specific cases. It is often necessary to fabricate sample parts and to test the joints to ensure their producibility and capability regarding strength properties (Ref 2, 3, 10, 15, 16, 17).

Other commonly used soldered joints are shown in Fig. 7; self-jigging joints are shown in Fig. 8. The most definitive work available on the loading of soldered joints is that used for the load-carrying capabilities of copper tube with sleeve-type joints or fittings. Conservative joint design requires that only 50% of the joint be considered filled. Under normal conditions, it would be considered good soldering practice if 70% or more of the joint consists of sound joint material (Ref 2, 3, 4, 5, 6, 11, 15, 18). Additional information about joint configurations is provided in the article "General Soldering" in this Volume.

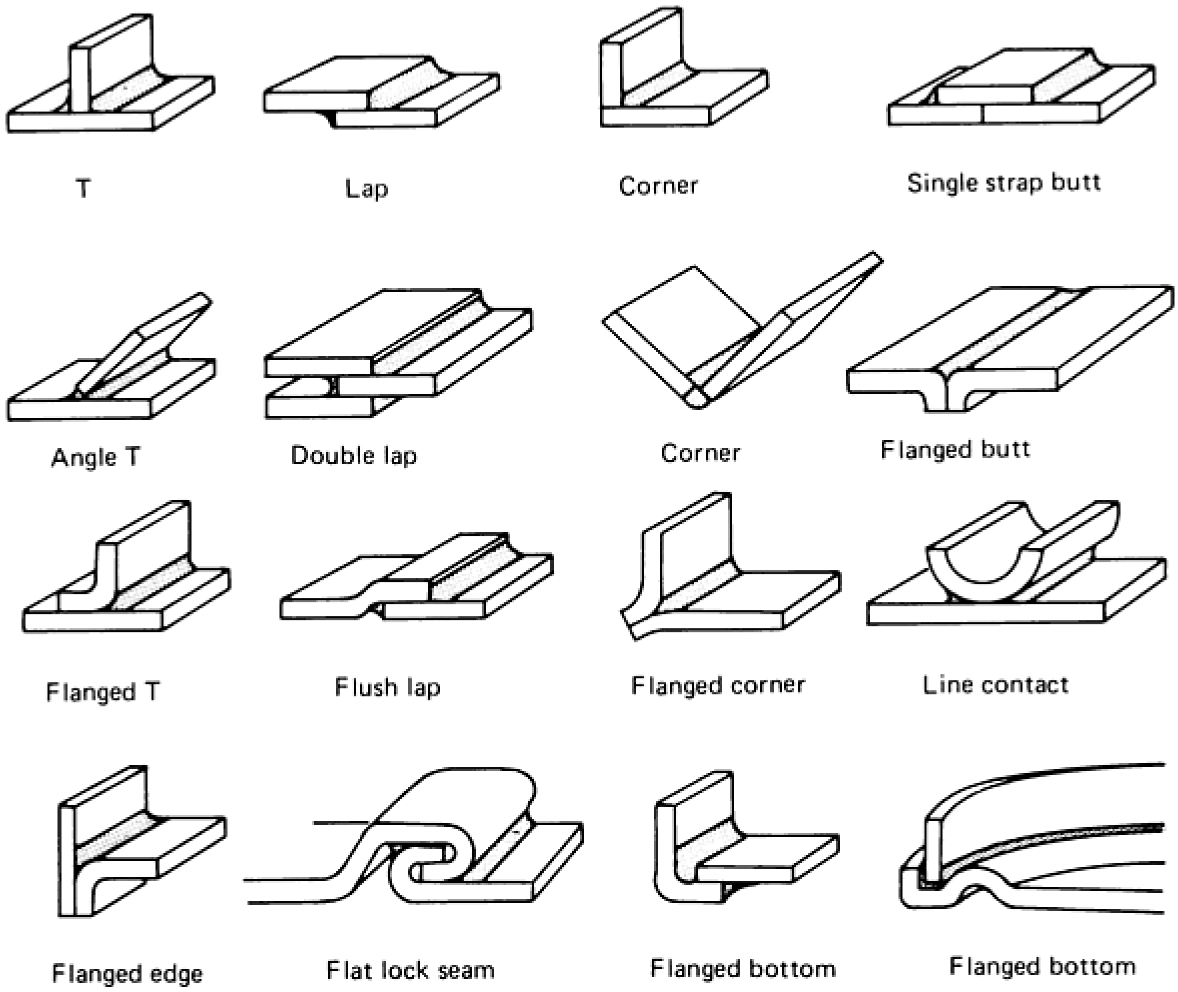
T

Lap

Corner

Single strap butt

Angle T

Double lap

Corner

Flanged butt

Flanged T

Flush lap

Flanged corner

Line contact

Flanged edge

Flat lock seam

Flanged bottom

Flanged bottom

FIG. 7 JOINT DESIGNS FREQUENTLY USED IN SOLDERING OPERATIONS. SOURCE: REF 5

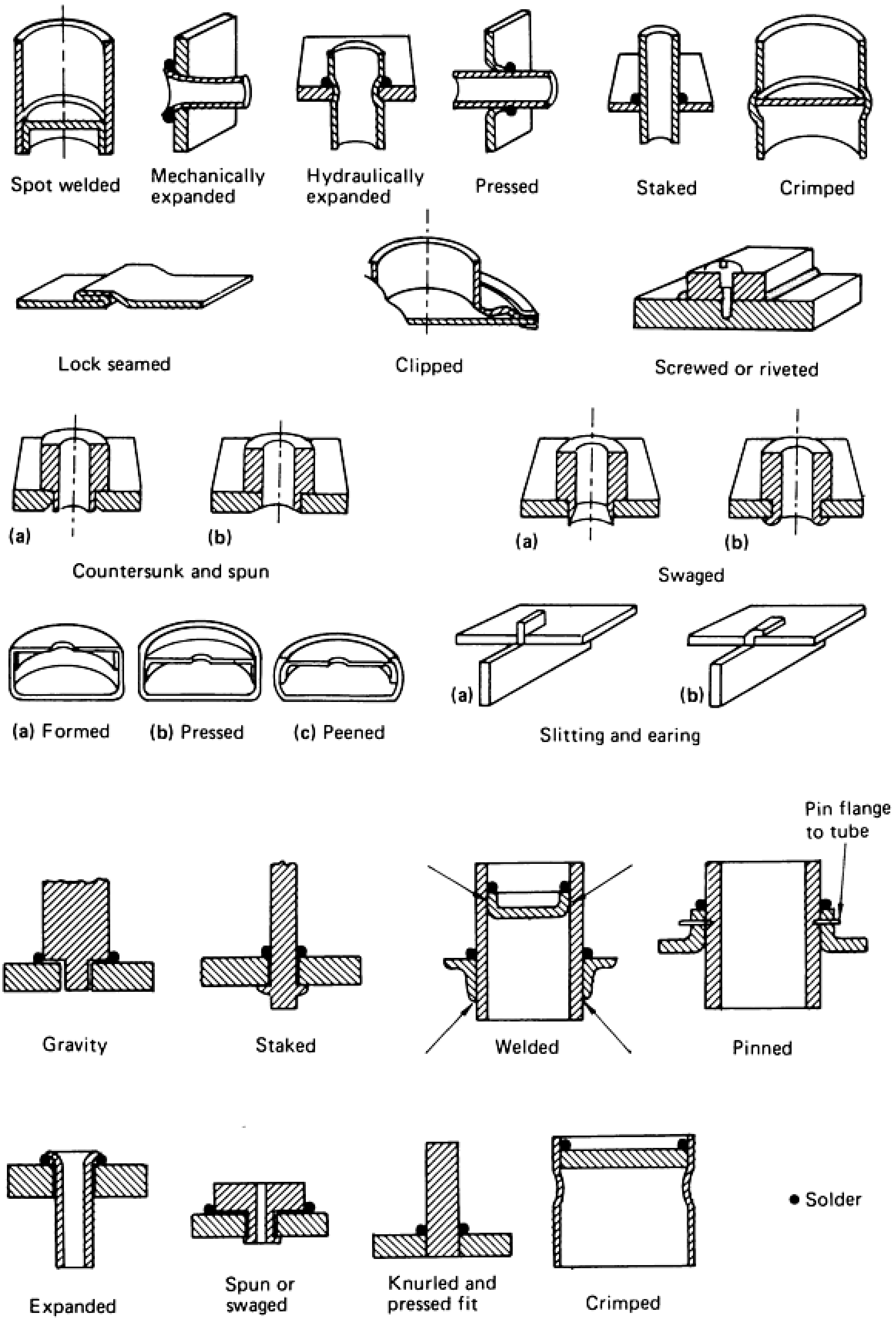
Spot welded
Mechanically expanded
Hydraulically expanded
Pressed
Staked
Crimped
Lock seamed
Clipped
Screwed or riveted
(a)
(b)
Countersunk and spun
(a)
(b)
Swaged
(a) Formed
(b) Pressed
(c) Peened
(a)
(b)
Slitting and earing
Pin flange to tube
Gravity
Staked
Welded
Pinned
Expanded
Spun or swaged
Knurled and pressed fit
Crimped
● Solder

FIG. 8 METHODS THAT CAN BE USED TO MAKE SOLDER JOINTS SELF-JIGGING. SOURCE: REF 5

References cited in this section

2. H.H. MANKO, *SOLDERS AND SOLDERING,* 2ND ED., MCGRAW-HILL, 1964
3. R.J. KLEIN WASSINK, *SOLDERING IN ELECTRONICS,* ELECTROCHEMICAL PUBLICATIONS LTD., 1989
4. L.P. LAMBERT, *SOLDERING FOR ELECTRONIC ASSEMBLIES,* MARCEL DEKKER, 1987
5. *SOLDERING MANUAL,* AWS, 1978
6. A.D. ROMIG, JR., Y.A. CHANG, J.J. STEPHENS, *ET AL.,* PHYSICAL METALLURGY OF SOLDER-SUBSTRATE REACTIONS, *SOLDER MECHANICS,* THE MINERALS, METALS, & MATERIALS SOCIETY, 1991, P 30-32
10. H.H. MANKO, SOLDERING FLUXES--PAST AND PRESENT, *WELD. J.,* VOL 52 (NO. 3), MARCH 1973, P 163-166
11. R.W. WOODGATE, *THE HANDBOOK OF MACHINE SOLDERING,* 2ND ED., JOHN WILEY & SONS, 1988
12. T. THOMPSON, CONDENSATION/VAPOR-PHASE REFLOW SOLDERING, *ASS. ENG.,* JUNE 1977, P 44-47
13. T.Y. CHU, A GENERAL REVIEW OF MASS SOLDERING METHODS, *INS. CIRC.,* NOV 1976
14. R.C. PFAHL, JR., J.C. MOLLENDORF, AND T.Y. CHU, CONDENSATION SOLDERING, *WELD. J.,* VOL 54 (NO. 1), 1975
15. E. LIEBERMAN, *MODERN SOLDERING AND BRAZING TECHNIQUES,* BUSINESS NEWS PUBLISHING CO., 1988
16. "SOLDER METAL SPECIFICATION," ASTM B 32, *ANNUAL BOOK OF ASTM STANDARDS,* PART B, ASTM
17. D.R. FREAR, W.B. JONES, AND K.R. KINSMAN, ED., *SOLDER MECHANICS,* TMS, 1991, P 30-41
18. *PROC. IEEE 41ST ELECT. COMP. & TECH. CONF.,* IEEE, 1991

Fundamentals of Soldering

Mel M. Schwartz, Sikorsky Aircraft

Precleaning and Surface Preparation

An unclean surface will not permit the solder to flow, which makes soldering either difficult or impossible and contributes to the formation of a poor joint. Materials such as oil, grease, paint, pencil markings, drawing, and cutting lubricants, general atmospheric dirt, oxide, and rust films must be removed before soldering. The importance of cleanliness cannot be overemphasized in ensuring sound soldered joints. Fluxing alone cannot substitute for adequate precleaning (Ref 19). Because cleaning methods are often designed for a specific soldering operation, their suitability for a particular application should be investigated thoroughly.

Degreasing. Either solvent or alkaline degreasing is recommended for the cleaning of oily or greasy surfaces. Of the solvent degreasing methods, the vapor condensation of the trichloroethylene-type solvents probably minimizes residual film on the surface. In the absence of vapor degreasing apparatus, immersion in liquid solvents or in detergent solutions is a suitable procedure. Hot alkali detergents are widely used for degreasing. All cleaning solutions must be thoroughly removed before soldering. Residues from hard-water rinses can interfere with soldering. Material compatibility is a prime consideration.

Pickling. The purpose of pickling, or acid cleaning, is to remove rust, scale and oxides, or sulfides from the metal in order to provide a clean surface for soldering. All of the inorganic acids (that is, hydrochloric, sulfuric, phosphoric, nitric, and hydrofluoric acids), singly or mixed, will fulfill this function, although hydrochloric and sulfuric acids are the most widely used. The pieces should be washed thoroughly in hot water after pickling and then dried as quickly as possible.

Mechanical cleaning includes the following methods:

- GRIT AND SHOTBLASTING
- MECHANICAL SANDING AND GRINDING
- FILING AND HAND SANDING
- CLEANING WITH STEEL WOOL
- WIRE BRUSHING AND SCRAPING

For best results, cleaning should extend beyond the joint area. Shot or steel gritblasting is often effective and is preferable to sandblasting because it avoids the embedding of silica particles on the surface, which interferes with the flow of solder. Mechanical cleaning is not recommended for softer metals, such as copper.

Precoating. The coating of the base metal surfaces with a more-solderable metal or alloy prior to the soldering operation can facilitate soldering. Coatings of tin, copper, silver, cadmium, iron, nickel, and the alloys of tin-lead, tin-zinc, tin-copper, and tin-nickel are used for this purpose. The advantages of precoating include the following:

- SOLDERING IS MORE RAPID AND UNIFORM.
- STRONG ACIDIC FLUXES CAN BE AVOIDED DURING SOLDERING.

The precoating of metals that have tenacious oxide films (for example, aluminum, aluminum bronzes, highly alloyed steels, and cast iron) is almost mandatory. The precoating of steel, brass, and copper, although not entirely essential, is of great value in some applications.

A number of different methods are used to precoat the metal surfaces. Solder or tin can be applied using a soldering iron or an abrasive wheel, ultrasonic soldering, immersion in molten metal, electrodeposition, or chemical displacement.

Hot dipping can be accomplished by fluxing and dipping the parts in molten tin or solder. Often, small parts are initially placed in wire baskets and then cleaned, fluxed, and dipped in the molten metal. Finally, they are centrifuged to remove the excess metal. Coating by hot dipping is applicable to carbon steel, alloy steel, cast iron, copper, and certain copper alloys. Prolonged immersion in molten tin or solder should be avoided to prevent the excessive formation of intermetallic compounds at the coating/base-metal interface.

Electrodeposition. Precoating by electrodeposition can be accomplished in stationary tanks, in conveyorized plating units, or in barrels. This method is applicable to all steels, copper alloys, and nickel alloys. The coating metals are not limited to tin and solder. Copper, cadmium, silver, precious metals, nickel, iron, and alloy platings such as tin-copper, tin-zinc, and tin-nickel are also commonly used.

Certain combinations of electrodeposited metals (that is, duplex coatings), where one metal is plated over another, are becoming more popular as an aid to soldering. A coating of 0.005 mm (0.0002 in.) of copper plus 0.0075 mm (0.0003 in.) of tin is particularly useful for brass. The solderability of aluminum is enhanced by a coating of 0.013 mm (0.0005 in.) of nickel, followed by 0.008 mm (0.0003 in.) of tin or by a combination of zincate (zinc), 0.005 mm (0.0002 in.) copper, and tin. An iron plating followed by tin plating is extremely useful over a cast iron surface.

Immersion coatings (that is, chemical displacement coatings) of tin, silver, or nickel can be applied to some of the common base metals. These coatings are usually very thin and generally have a poor shelf life.

The shelf life of a coating is defined as its ability to withstand storage conditions without impairing solderability. Hot tinned and brightened electrotin coatings have an excellent shelf life. Inadequate thicknesses of electrotinned or immersion-tinned coatings have a limited shelf life. Tin or solder coating thicknesses ranging from 0.003 to 0.008 (0.0001 to 0.0003 in.) are recommended to ensure maximum solderability after prolonged storage.

Reference cited in this section

19. J. BROUS, EVALUATION OF POST-SOLDER FLUX REMOVAL, *WELD. J.,* VOL 54 (NO. 12), 1975, P 444S-448S

Fundamentals of Soldering

Mel M. Schwartz, Sikorsky Aircraft

Soldering in Electronic Applications

This section addresses some general considerations in the soldering of electronic devices. More detailed information is provided in the article "Soldering in Electronic Applications," in this Volume.

Substrates for Electronic Component Applications

A substrate is the platform on which the electrical/electronic circuitry is built. Many different types of substrates are used in the electronics field. Early substrates were sheets of material to which insulated, stand-off terminals were attached by bolting or riveting. The circuitry was then obtained by point-to-point wiring between these terminals. Although this approach is still used, the most common substrate in use today has circuitry that is laid on it in planar form, with the components already attached. Attachment is normally accomplished through the soldering process. The most common form of this substrate is the printed wiring board (PWB).

A substrate can actually have numerous forms:

- CIRCUITRY CAN BE LAID ON ONE SIDE OF THE SUBSTRATE OR ON BOTH SIDES.
- CIRCUITRY CAN BE NOT ONLY ON THE OUTER SURFACES OF THE SUBSTRATE, BUT IN LAYERS WITHIN THE SUBSTRATE MATERIAL (MULTILAYERING).
- THE SUBSTRATE CAN BE RIGID OR FLEXIBLE.

There are many methods of applying the conductive patterns that make up the circuitry and each has its own specific soldering solutions.

Many different materials are used for substrates. Substrates used for electrical/electronic applications must be insulators to allow the conductive circuitry to function in the manner of a network made from separate insulated wires. Common substrate materials for this application include ceramics and organic laminates. Metals can also be used as substrates if they are coated with a substantial insulating material that can withstand the rigors of assembly.

Ceramic Substrates. Many types of ceramics can be used as substrates, but the most common is a high-fired alumina that is either 96 or 99% Al_2O_3. The 96% Al_2O_3 material is usually employed when thick-film materials are placed on the surface to form the circuitry. The 99% Al_2O_3 material is most often used for the evaporation or sputtering of thin materials. The materials that are laid down for terminal and component bonding pads are usually gold (or alloys high in gold), silver, and copper. Because these metals are laid down in thin layers, a soldering problem called leaching occurs. Leaching is the migration of the components of a substrate into a working solution. The process must be controlled because soldering results in bonding between the solder and the metals being attached. If the soldering process is not controlled, it can literally dissolve the metallization from the surface of the substrate. A gold-platinum alloy is more resistant to leaching than pure gold.

Thin-film (10 to 500 nm, or 100 to 5000 Å) deposits are more susceptible to leaching than are thick-film (10 to 25 μm, or 400 to 1000 μin.) deposits, because the thin layer is more susceptible to being consumed by the leaching. Therefore, the time at temperature and the maximum temperature become very important. In leaching, temperature is more important than time. Significantly, more leaching will be generated by a small increase in temperature than would be produced with even a moderate increase in time.

Composite Laminates. The most popular composites are the copper-clad laminates used for PWB substrates. The metal cladding is usually copper, but could be nickel or some other metal. Copper cladding can be either a plated material or wrought rolled foil. From a soldering standpoint, the rolled material is preferable, but it is more costly and more difficult to bond to the laminate material. Many PWBs are passed through numerous heat cycles during the manufacturing process. In addition, the use of multilayer designs is increasing dramatically. Plated foil, contaminated with codeposited organics, can become brittle at elevated temperatures. Hence, foil coppers are available in several classes, depending on their application.

Printed Wiring Laminate Materials. The laminate upon which the metal cladding is applied is available in many composite forms. Originally, layers of paper or cloth impregnated with a phenolic resin and designated XP, XXP, or XXXP were used. The latter became the standard in the mid-1950s.

The selection of organic-based laminate materials as electronic substrates has certain limitations, the most common of which is temperature. In addition, soldering must not degrade the material by burning or charring. Laminate selection should depend on the application of the electronic hardware. A glass-epoxy laminate material known as FR-4 is commonly used in consumer and industrial electronics. The "FR" designation indicates that the material is fire resistant.

Metal Substrates. There are some applications in which metals can be used as base substrates. Because the presence of an insulating material is required, the metal must be coated with a suitable material (for example, porcelain enamel).

Solder Application

Because of the small distances between components, when soldering electronics, it is often necessary to measure carefully the amount of solder that must be applied to form a satisfactory joint. It is not only a matter of providing a sufficient amount of solder for a good fillet, but it is also important to not have excess solder in the joint. This can sometimes be accomplished by plating the area to be soldered to a predetermined solder thickness, so that the solder forms a good fillet when it is reflowed. This is often the method used to accomplish "flip-chip" bonding to produce tape automated bonding (TAB) assemblies (Ref 20).

Solder can also be apportioned in proper amounts to individual joints by using preforms. In the case of surface-mount technology (SMT), the solder is often screened onto the substrate in the form of solder paste or it is dispersed in small quantities in precise locations by using tiny nozzles or tubes.

Preforms. When a square or round lid is attached to a device, but solder is not desired in the enclosed area, the solder application is often accomplished by using squares or annular rings that are punched from solder sheet. The solder quantity is usually sufficient to fill the asperities in the surfaces. The thickness and the lateral dimensions are determined by the location and the amount of solder needed to make a satisfactory joint. The solder preform is placed in the space between the parts to be joined, and then this assembly is subjected to sufficient heat to fuse the solder and hermetically seal the assembly. The entire body is then cooled. If the parts were clean and the proper amount of solder was supplied, a good solder joint results.

Heat for soldering can be applied in a number of different ways. In automated processing, the parts to be soldered are often passed through the hot zone of a furnace on a belt. This makes it possible to provide adequate preheating of the parts, the proper maximum temperature, and a gentle cooling cycle to minimize stresses. This method is fast and allows easy control of the measured amounts of solder to precise locations in a variety of geometric shapes. However, the cost of the preforms is a drawback.

Solder Pastes. The rapid development of SMT processing, which can involve very dense populations of devices and leads, necessitated a method to precisely place controlled amounts of solder on a substrate, either for lead attachment or for bonding of the device to attain more-effective heat transfer. Solder pastes that are applied by screening or printing through a mask or by application through a fine tube can fulfill these needs.

The screening method, although ancient, has been modernized by the development of new high-precision screens, solder pastes, and equipment for screening and curing the screened materials.

Solder paste also can be applied through a metal mask, or stencil. Although the location of the solder is not as precise as that obtained with screening, the metal masks are more durable and provide better control of solder thickness as it is applied.

Another paste placement method is to use fine tubes or nozzles to deliver the paste to precise locations. This method is ideal for placing discrete dots of paste, rather than covering large areas.

Paste Composition. A solder paste must contain solder particles of the desired composition to yield the melting and joining requirements of the application. The metal particles must be sized to provide good screenability and to avoid excessive oxide content. The shape of the particles is important. Spherical particles are lower in oxide content because there is less surface area and they are able to flow more freely in screening applications.

A paste must also contain a vehicle to control its rheological, or flow, properties; its drying behavior; its surface activity (wetting behavior); and the solubility of the residues left after reflow. The vehicle usually consists of a rosin (pine derivative) or a synthetic resin, solvents to adjust drying time and temperature, modifiers to control rheological behavior, and an activator, or flux, to remove oxides from the solder and to protect the liquefied solder from oxidation.

Paste Properties. The paste must set and hold its shape and location after screening so that subsequent processing maintains the solder in the desired locations and amounts. The assembly is then subjected to heat and the solder is melted and joined to the part being soldered. This reflow can be conducted either as the final step in soldering or as a separate step. The parts must generally be cleaned after soldering.

Solder resists enable control over the PWB areas that will be wet by the solder. The resist, or mask, is an organic coating placed selectively on the board. Only the uncoated areas are exposed to and are wettable by the solder. The higher interconnect densities now being attained by SMT processing make solder resists a valuable method.

Solder resists can be either permanent or temporary, depending on the application. The permanent resists must resist not only the solder, but any fluxes or solvents applied later in the manufacturing sequence. Compatibility with the process and the circuit materials is a consideration, as is the ability to withstand soldering temperatures. In addition, the solder resist offers the circuits some protection against dirt and contamination. Commonly used resist materials are polystyrene, epoxides, and acrylics, although more-costly materials, such as polyimides, are sometimes used because of their greater thermal resistance.

Screened Materials. Solder masks are commonly applied by screening. The films are subsequently dried and cured thermally. The elevated temperatures involved in the curing cycle can cause oxidation of the conductors on the PWB, resulting in a loss of solderability. Lower curing temperatures are desirable. Some of the resists are cured by exposure to ultraviolet light, rather than by heating. This approach is very fast, involves less emission of volatile pollutants, and does not cause oxidation of the metal surfaces.

Photographic Films. A second method of applying solder resists is to use sheets of photosensitive material laminated to the surface of the circuit. Although these are more expensive than screened materials, the photographic films enable much higher pattern resolution. Dry film resists are much stronger and more resistant to damage during soldering and cleaning operations. The use of these photosensitive materials increases the complexity of the process.

Temporary solder masks are often used as production aids, offering local protection from solder at locations such as:

- CONTACT PADS THAT HAVE BEEN GOLD PLATED
- PLATED-THROUGH HOLES THAT MUST BE SOLDERED IN A LATER OPERATION
- PADS THAT MUST BE KEPT FREE OF SOLDER

Adhesive solder-mask tapes that specifically protect edge connectors on PWBs are also available, and numerous mechanical approaches to limiting solder contact exists (Ref 18).

References cited in this section

18. *PROC. IEEE 41ST ELECT. COMP. & TECH. CONF.,* IEEE, 1991
20. J.H. LAU, *HANDBOOK OF TAPE AUTOMATED BONDING,* VAN NOSTRAND REINHOLD, 1992

Fundamentals of Soldering

Mel M. Schwartz, Sikorsky Aircraft

Soldering Process Parameters

The parameters that affect wetting and spreading phenomena include:

- TEMPERATURE
- TIME
- VAPOR PRESSURE
- METALLURGICAL AND CHEMICAL NATURE OF THE SURFACES
- GEOMETRY OF THE SOLID

The manipulation of each parameter can result in some control of the wetting and soldering processes.

Temperature is important to wetting and spreading for several reasons. Higher temperatures result in greater atomic activity and provide some of the energy needed to overcome surface barriers. Higher temperatures also increase reaction rates exponentially. Temperature determines phase relationships for a given alloy composition, whether for the formation of a solid solution or for that of an intermetallic compound. The fluidity of the liquid solder is also increased with increasing temperature. In addition, higher temperatures promote the formation of oxides, which can be troublesome when soldering.

Time. Wetting is time dependent because of the rate of kinetics of the solder-substrate interaction, the flux action, and the heat conduction of the substrate. The soldering process is time sensitive, because time is needed to provide sufficient opportunity for the solder to wet, penetrate, or "wick up" into the various areas of the substrate that must be incorporated into the solder joint. It is clear that measures of the time to initiate wetting and of the time it takes for spreading to occur are needed.

Vapor pressure is considered to be negligible for lead-tin solders, but it can be a significant factor when working with certain solders. Caution must be exercised when selecting and working with alloys that have appreciable vapor pressures.

The chemical and metallurgical nature of a surface affects wetting and spreading in several ways. Alloy content, as well as the fluidity of the liquid, both affect wetting and, thus, the rate of spreading. Grain boundaries are wet differently than the bulk of the metal. The formation of solid solutions or intermetallic compounds also can have a considerable effect on wetting and spreading.

The presence of oxide skins will prevent wetting and spreading, because the solder generally will not wet oxides. Likewise, films or particles of organic matter will interfere with wetting. Oils and silicones are frequent offenders and usually require different types of cleaning materials for their removal than the types used for oxides.

Surface Geometry. Because contact angle is a critical measure of wetting and spreading behavior, surface geometry is an important factor in the control of spreading. Soldering artisans have long used a wire brush not only to clean the surface to be soldered, but also to aid the spreading of the solder by providing sharp grooves that act as small capillaries to assist wetting. Isotropic spreading is actually what is needed.

Another condition of surface interaction that is sometimes encountered in soldering is dewetting, where a metal surface is initially wet by solder, but as the solder cools, its cohesive forces exceed the forces involved in wetting, and the solder balls up on the metal surface. This usually results from surface contamination of the metal or from the entrapment of foreign particles in the metal surface, leading to weak wetting or, possibly, only local wetting.

Base-Metal Selection. A sound soldered joint is achieved by selecting and using the proper materials and processes. Base metals are usually selected to achieve the specific property requirements of a component. These properties can include strength, ductility, electrical conductivity, weight, corrosion resistance, and others. When soldering is required, the solderability of the base materials should also be a selection factor. Both flux selection and surface preparation will be affected by the solderability of the base materials to be joined.

The solderability of metals and alloys is not simply a matter of chemical nobility, as might be supposed when regarding the good solderability of the noble metals, which do not readily form oxide or tarnish films. Although cadmium and tin both form oxides readily, they are considered easy to solder. On the other hand, chromium, nickel, and aluminum also form oxide films readily, but are difficult to solder. The difference is in the extremely adherent, protective nature of the oxides formed on chromium, nickel, and aluminum, compared with the oxides that form on tin and cadmium (Table 2).

TABLE 2 RELATIVE SOLDERABILITY OF SELECTED METALS AND ALLOYS

EASY TO SOLDER • PLATINUM • GOLD • COPPER • SILVER • CADMIUM PLATE • TIN • SOLDER PLATE
LESS EASY TO SOLDER • LEAD • NICKEL PLATE • BRASS • BRONZE • RHODIUM • BERYLLIUM COPPER
DIFFICULT TO SOLDER • GLAVANIZED IRON • TIN-NICKEL • NICKEL-IRON • MILD STEEL
VERY DIFFICULT TO SOLDER • CHROMIUM • NICKEL-CHROMIUM • NICKEL-COPPER • STAINLESS STEEL
MOST DIFFICULT TO SOLDER • ALUMINUM • ALUMINUM BRONZE
NOT SOLDERABLE • BERYLLIUM • TITANIUM

It should be noted that chromium, nickel, and aluminum are all soldered regularly with good results, but that special attention must be given to the selection of fluxes, which must be very active. In many cases, the use of active fluxes is either restricted or not allowed. Therefore, these hard-to-solder metals and alloys always require special consideration in order to provide reproducible soldering.

Numerous tests and methods to measure solderability are available. The tendencies of solder to either wet or spread on a given material are critical when evaluating candidate soldering systems. When hand soldering on single, large-dimension bodies prevailed, the skilled artisan could adjust parameters while watching the results develop. In most cases, visual determinations of wetting and spreading characteristics provided a sufficient basis for process control.

Today, operating parameters must be controlled very carefully, because the soldering of components and/or electronic assemblies can involve many joints that are soldered at one time, or joints that are often hidden, or dimensions that are extremely small, or situations where joint-to-joint accommodations cannot be made. This necessitates the use of automatic soldering systems. Moreover, conditions must be uniform over the whole structure being soldered. In this environment, it is essential that the processes be based on measurements of wetting and spreading behavior, so that the controlling parameters can be accurately set.

Additional information is available in the articles "Brazeability and Solderability of Engineered Materials," "General Soldering," and "Soldering in Electronic Applications" in this Volume.

Fundamentals of Soldering

Mel M. Schwartz, Sikorsky Aircraft

Soldering Equipment

The proper application of heat is of paramount importance in any soldering operation. The solder should melt while the surface is heated to permit the molten solder to wet and flow over the surface. The best heating method is therefore another important consideration.

Soldering Irons. The traditional soldering tool is the soldering iron, or bit, with a copper tip than can be heated electrically, by direct flame, or in an oven. Because soldering is a heat-transfer process, the maximum surface area of the heated tip should contact the base metal. The solder itself should not be melted upon the tip of the iron when a joint is being made. To lengthen the usable life of a copper tip, a coating of solder-wettable metal, such as iron with or without additional coatings, is applied to the surface of the copper. The rate of dissolution of the iron coating in molten solder is substantially less than the rate for copper. The iron coating also shows less wear, oxidation, and pitting than uncoated copper.

Soldering irons are available in a large variety of sizes and designs, ranging from a small pencil to special irons or bits that weigh 2 kg (5 lb) or more. The selection of the iron depends on the application and the quantity of heat needed at the joint. The heat recovery time of the iron should be fast enough to keep up with the job.

Regardless of the heating method, the tip performs the following functions:

- STORING AND CONDUCTING HEAT FROM THE HEAT SOURCE TO THE PARTS BEING SOLDERED
- STORING MOLTEN SOLDER
- CONVEYING MOLTEN SOLDER
- WITHDRAWING SURPLUS MOLTEN SOLDER
- BRINGING THE WORKPIECE JOINT AREA TO THE SOLDERING TEMPERATURE

The angle at which the copper tip is applied to the work is important in terms of delivering maximum heat. The flat side of the tip should be applied to the work to obtain the maximum area of contact. Flux-cored solders should not be melted on the soldering tip, because this destroys the effectiveness of the flux. The cored solder should be touched to the soldering tip to initiate good heat transfer, and then the solder should be melted on the work parts to complete the solder joint (Ref 11).

Additional information is available in the article "Iron Soldering" in this Volume.

Torch Soldering (TS). The selection of a gas torch is controlled by the size, mass, and configuration of the assembly to be soldered. When fast soldering is necessary, a flame is frequently used. The flame temperature is controlled by the fuel mixture used. Fuel gas burned with oxygen gives the highest flame temperature possible with that gas. The highest flame temperatures are attained with acetylene. Lower temperatures are obtained with propane, butane, natural gas, and manufactured gas.

Multiple flame tips, or burners, often have shapes that are suitable to the work. They can be designed to operate on oxygen and fuel gas, compressed air and fuel gas, or Bunsen-type torches.

When adjusting tips or torches, care should be taken to avoid adjustments that result in a "sooty" flame, because the carbon deposited on the work will prevent the flow of solder (Ref 11, 15).

Detailed information is available in the article "Torch Soldering" in this Volume.

Dip soldering (DS) utilizes a molten bath of solder to supply both the heat and solder required to produce the joints (Fig. 9a). This method is useful and economical, because an entire unit comprising any number of joints can be soldered in one operation after proper cleaning and fluxing. Fixtures are required to contain the unit and maintain proper joint clearances during solder solidification. The soldering pot should be large enough so that, at a given rate of production, the units being dipped will not appreciably lower the temperature of the solder bath.

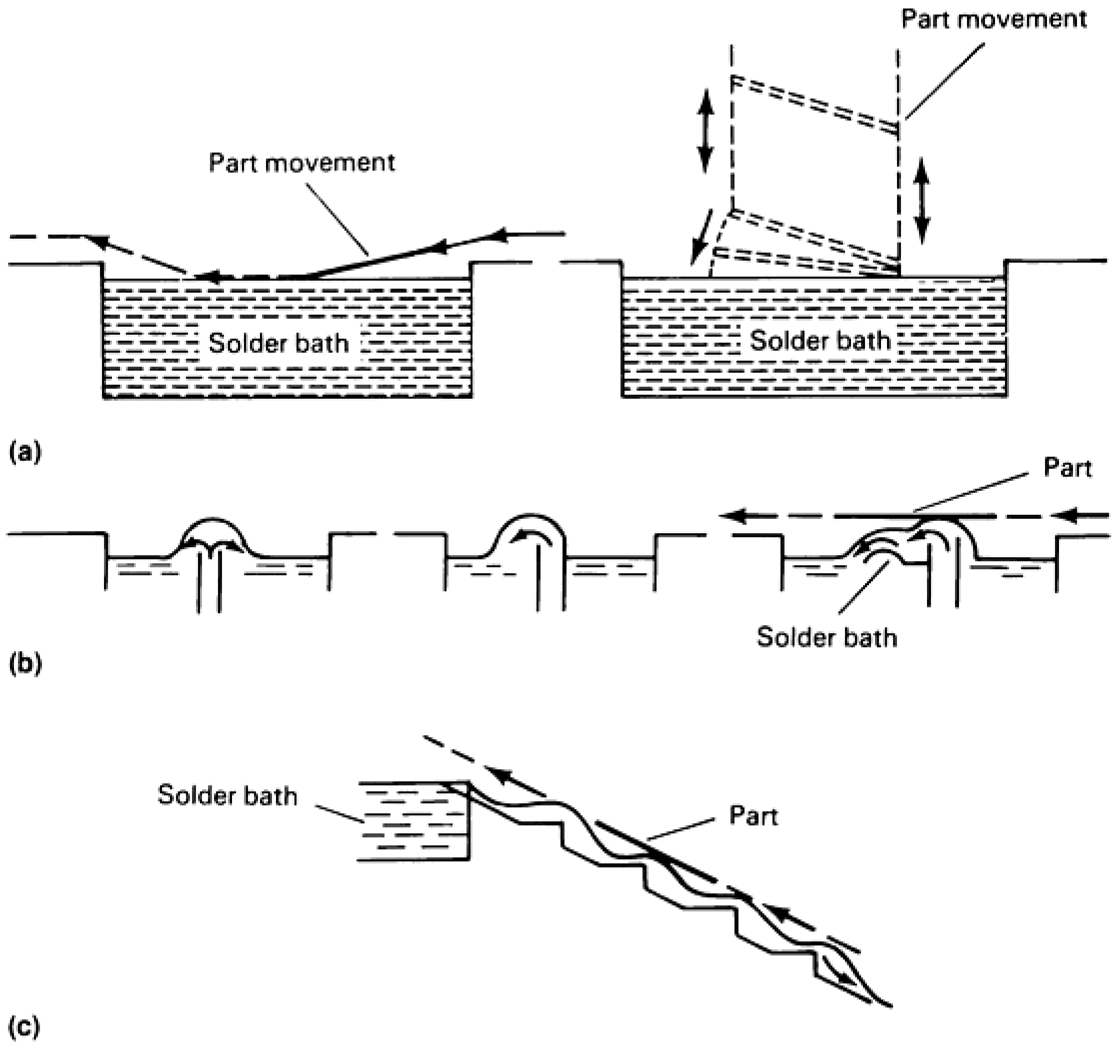

FIG. 9 SELECTED SOLDERING TECHNIQUES USED FOR HIGH-VOLUME PRODUCTION APPLICATIONS. (A) DIPPING ON A STATIC BATH. (B) WAVE SOLDERING SYSTEM. (C) CASCADE SOLDERING SYSTEM

Additional information is available in the article "Dip Soldering" in this Volume.

Wave Soldering (WS). Electronic components are commonly soldered to PWBs by wave soldering (Fig. 9b). The solder is pumped out of a narrow slot to produce a wave or series of waves. This approach, which lends itself to automatic soldering, involves a continuously pumped crest of solder that peaks at the PWB. Another method is "drag," or planar, soldering. This method uses a bath, into which the board is automatically dipped with controlled conveyorization (Ref 21, 22).

In cascade soldering (Fig. 9c), the solder flows down a trough because of gravity and is returned by pump to the upper reservoir. Both of these wave solder systems are excellent in that a virtually oxide-free solder surface is presented to the part, and the flux and vapors are dislodged by the flow of the solder.

Integrated wave soldering systems for printed wiring assemblies provide units that can apply the flux, dry and preheat the board, solder components, and clean the completed assembly. Some systems have special features that apply the flux by passing the assemblies through a wave, by spraying, by rolling, or by dipping. Several systems use oil mixed with the solder to aid in eliminating icicles and bridging between conductor paths. Another system features dual waves that flow in the opposite direction to the travel of the board.

Additional information is available in the article "Wave Soldering" in this Volume.

Induction Soldering (IS). The only requirement for a material that is to be induction soldered is that it be an electrical conductor. The rate of heating depends on the induced current flow. The distribution of heat obtained with induction heating is a function of the induced current frequency. The higher frequencies concentrate the heat at the surface of the workpiece.

Several types of equipment are available for induction heating: the vacuum tube oscillator, the resonant spark gap, the motor-generator unit, and solid-state units.

Induction heating is generally applicable for soldering operations, with the following requirements and attributes:

- LARGE-SCALE PRODUCTION
- APPLICATION OF HEAT TO A LOCALIZED AREA
- MINIMUM OXIDATION OF SURFACE ADJACENT TO THE JOINT
- GOOD APPEARANCE AND CONSISTENTLY HIGH JOINT QUALITY
- SIMPLE JOINT DESIGN THAT LENDS ITSELF TO MECHANIZATION

The induction technique requires that the parts being joined have clean surfaces and that joint clearances be maintained accurately. High-grade solders are generally required to obtain rapid spreading and good capillary flow. Preforms often afford the best means of supplying the correct amount of solder and flux to the joint.

When induction soldering dissimilar metals (particularly joints composed of both magnetic and nonmagnetic components), attention must be given to the design of the induction coil in order to bring both parts to approximately the same temperature.

Additional information is available in the article "Induction Soldering" in this Volume.

Furnace Soldering. There are many applications, especially in high-production soldering, where a furnace will produce consistent and satisfactory soldering. Although this method is not widely used, furnace heating should be considered under the following circumstances:

- WHEN ENTIRE ASSEMBLIES CAN BE BROUGHT TO THE SOLDERING TEMPERATURE WITHOUT DAMAGE TO ANY OF THE COMPONENTS
- WHEN PRODUCTION IS SUFFICIENTLY LARGE TO ALLOW EXPENDITURES FOR JIGS AND FIXTURES TO HOLD THE PARTS DURING THE SOLDERING PROCESS
- WHEN THE ASSEMBLY IS COMPLICATED IN NATURE, MAKING OTHER HEATING METHODS IMPRACTICAL

The proper clamping fixtures are very important during furnace soldering. Movement of the joint during solder solidification can result in a poor joint.

Another important consideration in furnace soldering is flux selection. Rosin and organic fluxes are subject to decomposition when maintained at elevated temperatures for an extended period of time. When either rosin or organic flux is used, the part must be brought rapidly to the liquidus temperature of the solder. It is sometimes beneficial to dip the parts in a hot flux solution before placing them in the oven. When using rosin-base flux, it is generally necessary to use a solder with a tin content of 50% or more.

The reducing atmosphere used in the furnace does not allow joints to be made without flux, because the temperatures at which these atmospheres become reducing are far above the liquidus temperature of the solders. The use of inert atmospheres will prevent further oxidation of the parts, but flux must be used to remove the oxide that is already present.

The furnaces should be equipped with adequate temperature controls because solder flow begins at approximately 45 to 50 °C (80 to 90 °F) above the solder liquidus temperature. The optimum condition exists when the heating capacity of the furnace is sufficient to heat the parts rapidly to the liquidus temperature of the solder.

Detailed information is available in the article "Furnace and Infrared Soldering" in this Volume.

Resistance soldering (RS) involves placing the workpiece either between a ground and a movable electrode or between two movable electrodes as part of an electrical circuit. Heat is applied to the joint both by the electrical resistance of the metal being soldered and by conduction from the electrode, which is usually carbon. Production assemblies can utilize multiple electrodes, rolling electrodes, or special electrodes, depending on which setup offers the most advantageous soldering speed, localized heating, and power consumption. Resistance soldering electrode bits generally cannot be tinned, and the solder must be fed directly into the joint.

Additional information is available in the article "Resistance Soldering" in this Volume.

Infrared Soldering (IRS). Optical soldering systems that focus infrared light (radiant energy) on the joint by means of a lens are available. Lamps with power ratings that range from 45 to 1500 W (140 to 4700 Btu/h) can be used for different application requirements. The devices can be programmed through a silicon-controlled rectifier (SCR) power supply with an internal timer.

The most common sources of infrared heating for soldering applications are heated filaments. The quartz-iodine tungsten-filament lamp is widely used because it is very stable and reliable over a wide range of temperatures.

In general, infrared soldering systems are simple and inexpensive to operate. One of the most critical operating parameters is surface condition. Variations in the condition of the solder surface can be compensated for, to some extent, by adjusting the heating power. Advantages are process repeatability, ability to concentrate or focus the energy with reflectors and lenses, economy of operation, and absence of contact with the workpiece.

Detailed information is available in the article "Furnace and Infrared Soldering" in this Volume.

Ultrasonic Soldering. This process uses a transducer as the source of ultrasonic energy. The transducer is energized in a bath of molten solder, and sound waves are coupled between the transducer and the workpiece, allowing the oxides in the base metal to be disrupted so that the solder melts the base metal. Sound waves are transmitted throughout the base metal, permitting wetting to occur on surfaces that are "blind" to the source. Ultrasonic soldering is also used to apply solderable coatings on difficult-to-solder metals.

Additional Soldering Processes. In addition to the above-mentioned processes, hot-gas soldering, spray-gun soldering, and condensation soldering are also used.

Hot gas soldering utilizes a fine jet of inert gas, heated to above the liquidus of the solder. The gas acts as a heat-transfer medium and as a blanket to reduce access of air at the joint.

Additional information is available in the article "Hot Gas Soldering" in this Volume.

Spray gun soldering is a heating method used when the contour of the part to be soldered is difficult to follow with either a wiping or drop method or when the part is placed in the assembly in such a way that the solder cannot be applied after the parts are assembled.

Gas-fired or electrically heated guns are available. Each type is designed to spray molten or semimolten solder on the work from a continuously fed solid solder wire. Soldering guns use either propane mixed with oxygen or natural gas mixed with air to heat and to spray a continuously fed solid solder wire of approximately 3.2 mm (0.12 in.) diameter. About 90% of the solder wire is melted by the flame of the gun. The solder contacts the workpiece in a semiliquid form. The workpiece then supplies the balance of the heat required to melt and flow the solder. Adjustments can be made within the spray gun to control the solder spray.

Vapor-phase soldering (also known as condensation soldering) utilizes the latent heat of vaporization of a condensing saturated liquid to provide the heat required for soldering. A reservoir of saturated vapor over a boiling liquid provides a constant controlled temperature with rapid heat transfer. This method is useful for large assemblies, as well as for temperature-sensitive parts.

Detailed information is available in the article "Vapor-Phase Soldering" in this Volume.

References cited in this section

11. R.W. WOODGATE, *THE HANDBOOK OF MACHINE SOLDERING,* 2ND ED., JOHN WILEY & SONS, 1988
15. E. LIEBERMAN, *MODERN SOLDERING AND BRAZING TECHNIQUES,* BUSINESS NEWS PUBLISHING CO., 1988
21. P.J. BUD, MASS PRODUCTION TECHNIQUES USING THE PRINCIPLE OF WAVE SOLDERING, *WELD. J.,* VOL 52 (NO. 7), JULY 1973, P 431-439
22. P.J. BUD AND D.A. ELLIOT, WAVESOLDERING AT ELEVATED TEMPERATURES, *WELD. J.,* VOL 53 (NO. 2), FEB 1974, P 79-87

Fundamentals of Soldering

Mel M. Schwartz, Sikorsky Aircraft

Quality Control

Table 3 lists that are commonly used to evaluate the solderability properties of selected soldered components.

TABLE 3 TEST STANDARDS USED TO EVALUATE SOLDERABILITY

BASIC METHOD	TEST STANDARD
DIP AND LOOK	ANSI/IPC-S-804 (BOARDS)
	ANSI/IPC-S-805 (COMPONENTS)

	IEC 68-2-20
	MIL-STD-202, 750, 883
	EIA RS-186-9E
ROTARY DIP	ANSI-IPC-S-804 (BOARDS)
	IEC 68-2-20
	BS4025 (BRITISH STANDARD)
WETTING BALANCE	ANSI-IPC-S-805 (COMPONENTS)
	IEC-68-2-20
	MIL-STD-883
GLOBULE TEST	ANSI-IPC-S-805 (COMPONENTS)
	IEC 68-2-20
	BS 2011 (BRITISH STANDARD)
	DIN 40046 (GERMAN STANDARD)
MENISCUS RISE	NO STANDARDS TO DATE
TIMED SOLDER RISE	ANSI/IPC-S-804 (BOARDS)

Dip Test. The most useful test for assessing solderability is the dip test, because of its accuracy and simplicity. It is conducted with the additions of cams and timers so that standardized conditions are established. The dip test is a reasonable simulation of practical soldering conditions and starts with a test specimen that has been dipped in a mild flux, as is usually the case with soldering tests. The dip test involves immersing the test specimen into a clean solder bath, waiting an appropriate length of time, and then withdrawing the specimen from the bath. In most cases, the test is automated to ensure reproducibility. The results are evaluated by means of visual inspection and comparison with a set of standards. Thus, the dip test is operator sensitive, somewhat subjective, and of limited value, because it is not quantitative.

Rotary Dip Test. This test was devised to provide a better simulation of the dynamic conditions that exist in machine soldering operations. It is used to evaluate the solderability of PWBs, and involves bringing the specimen into moving contact with the molten solder for a specified period of time. This is a "go/no-go" test that requires complete wetting of the specimen surface within a set time, and is thus dependent on visual inspection of the test specimen.

Wetting Balance (Meniscograph). Another common test employs the wetting balance, which gives a numerical result. The equipment consists of a balance that records the forces exerted on a fluxed test specimen during a controlled dipping cycle in a clean solder bath. From the values obtained, it is possible to compare the rate of wetting, as well as the forces involved for various flux formulations.

The speed of wetting is an especially important factor in the wave soldering process, where automation requires a time-based response of flux and solder as the work moves through the process machine at a constant speed.

The wetting balance test cannot be used on finished, etched printed circuits or on single-sided circuits. It is often used on specimens cut from double-sided clad materials prior to etching.

Although this test gives numerical results, they are not strictly quantitative. However, the results do provide the basis for a meaningful comparison among fluxes.

Additional Tests. There are several other tests that provide practical information for special purposes, such as the spreading test and the globule test. These tests are sometimes used to evaluate incoming materials or to aid in the selection of materials. They are not universally employed.

The spreading test is used to provide a numerical rating for both fluxes and solders. It is described in detail in MIL-F-14256, "Flux, Soldering, Liquid (Rosin Base)," and in QQ-S-571, "Solder, Electronic," which are national-level specifications.

The globule test was once widely applied, but is now limited to being used to determine the wetting of component leads and wires (Ref 24).

Additional information is available in the article "Evaluation and Quality Control of Soldered Joints" in this Volume.

Reference cited in this section

24. IEC-68-2-20, INTERNATIONAL ELECTROCHEMICAL COMMISSION, GENEVA

Fundamentals of Soldering

Mel M. Schwartz, Sikorsky Aircraft

Future Outlook

The quality soldered joints obtained by automation have resulted in the widespread use of automated equipment throughout the electronics industry, as well as throughout the world (Ref 25).

When compared to processes such as adhesive bonding, welding, brazing, or mechanical joining, soldering offers many advantages. Automated soldering equipment for wave or planar soldering can be easily and economically installed. The high reliability that can be achieved with soldering results in a quality end product. Soldering offers the additional benefit of providing sequential assembly, a common practice in the electronics industry.

Reference cited in this section

25. REFLOW SOLDERING, *WELD. J.*, VOL 52 (NO. 1), JAN 1973, P 22-30

Fundamentals of Soldering

Mel M. Schwartz, Sikorsky Aircraft

References

1. J. WOLTERS, *ZUR GESEHICHTE DER LOTTECHNIK,* DEGUSSA, WEST GERMANY, 1977
2. H.H. MANKO, *SOLDERS AND SOLDERING,* 2ND ED., MCGRAW-HILL, 1964
3. R.J. KLEIN WASSINK, *SOLDERING IN ELECTRONICS,* ELECTROCHEMICAL PUBLICATIONS LTD., 1989
4. L.P. LAMBERT, *SOLDERING FOR ELECTRONIC ASSEMBLIES,* MARCEL DEKKER, 1987
5. *SOLDERING MANUAL,* AWS, 1978
6. A.D. ROMIG, JR., Y.A. CHANG, J.J. STEPHENS, *ET AL.*, PHYSICAL METALLURGY OF SOLDER-SUBSTRATE REACTIONS, *SOLDER MECHANICS,* THE MINERALS, METALS, & MATERIALS SOCIETY, 1991, P 30-32
7. "SOFT SOLDER ALLOYS--CHEMICAL COMPOSITIONS AND FORMS," ISO 9453, 1ST ED., 1990
8. "SOFT SOLDERING FLUXES--CLASSIFICATION AND REQUIREMENTS--PART I: CLASSIFICATION, LABELING AND PACKAGING," ISO 9454-1, 1990
9. J.F. SHIPLEY, INFLUENCE OF FLUX, SUBSTRATE AND SOLDER COMPOSITION ON SOLDER WETTING, *WELD. J.,* VOL 54 (NO. 10), OCT 1975, P 357S-362S
10. H.H. MANKO, SOLDERING FLUXES--PAST AND PRESENT, *WELD. J.*, VOL 52 (NO. 3), MARCH 1973, P 163-166
11. R.W. WOODGATE, *THE HANDBOOK OF MACHINE SOLDERING,* 2ND ED., JOHN WILEY & SONS, 1988

12. T. THOMPSON, CONDENSATION/VAPOR-PHASE REFLOW SOLDERING, *ASS. ENG.*, JUNE 1977, P 44-47
13. T.Y. CHU, A GENERAL REVIEW OF MASS SOLDERING METHODS, *INS. CIRC.*, NOV 1976
14. R.C. PFAHL, JR., J.C. MOLLENDORF, AND T.Y. CHU, CONDENSATION SOLDERING, *WELD. J.*, VOL 54 (NO. 1), 1975
15. E. LIEBERMAN, *MODERN SOLDERING AND BRAZING TECHNIQUES*, BUSINESS NEWS PUBLISHING CO., 1988
16. "SOLDER METAL SPECIFICATION," ASTM B 32, *ANNUAL BOOK OF ASTM STANDARDS*, PART B, ASTM
17. D.R. FREAR, W.B. JONES, AND K.R. KINSMAN, ED., *SOLDER MECHANICS*, TMS, 1991, P 30-41
18. *PROC. IEEE 41ST ELECT. COMP. & TECH. CONF.*, IEEE, 1991
19. J. BROUS, EVALUATION OF POST-SOLDER FLUX REMOVAL, *WELD. J.*, VOL 54 (NO. 12), 1975, P 444S-448S
20. J.H. LAU, *HANDBOOK OF TAPE AUTOMATED BONDING*, VAN NOSTRAND REINHOLD, 1992
21. P.J. BUD, MASS PRODUCTION TECHNIQUES USING THE PRINCIPLE OF WAVE SOLDERING, *WELD. J.*, VOL 52 (NO. 7), JULY 1973, P 431-439
22. P.J. BUD AND D.A. ELLIOT, WAVESOLDERING AT ELEVATED TEMPERATURES, *WELD. J.*, VOL 53 (NO. 2), FEB 1974, P 79-87
23. *ELECTRONIC MATERIALS HANDBOOK*, VOL 1, *PACKAGING*, ASM INTERNATIONAL, 1989
24. IEC-68-2-20, INTERNATIONAL ELECTROCHEMICAL COMMISSION, GENEVA
25. REFLOW SOLDERING, *WELD. J.*, VOL 52 (NO. 1), JAN 1973, P 22-30

Introduction to Solid-State Welding

Ray Dixon, Los Alamos National Laboratory

Solid-State Welding

(SSW) processes are those that produce coalescence of the faying surfaces at temperatures below the melting point of the base metal being joined without the addition of brazing or solder filler metal. Pressure may or may not be applied. These processes involve either the use of deformation or of diffusion and limited deformation in order to produce high-quality joints between both similar and dissimilar materials.

Dissimilar metal joints are necessary in applications that require a variety of material properties within the same component. For example, heat exchangers often require different types of stainless steels at each end, because of temperature-induced corrosion. Under laboratory conditions, dissimilar materials can be chosen based on physical or material properties that influence the phenomenon being studied. For whatever reason, and appropriate method of producing dissimilar metal joints can usually be determined (assuming it is even possible) by examining the phase diagram. If the diagram indicates difficulty in joining the materials (intermetallics, and so on), then a solid-state (nonmelting) process may be applicable. When a nonmelting process is chosen, it is only successful if a relatively strong joint is produced.

There is considerable interest is quantifying joint strength, and extensive literature (Ref 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14) exists to help design and evaluate various specimen geometries. This experience includes not only techniques for conducting the tests, but analysis and interpretation of the data with respect to the physical properties of individual materials.

The study of dissimilar welds necessarily leads to studies of the interfacial region, because it is the most likely site of part failures. In a two-component system (materials A and B), interfacial failure can be a result of the individual properties of either A and B, but they are by necessity synergistic. These synergisms and the resulting complex interactions that they produce are what is significant. From an engineering and processing standpoint, an interfacial "cookbook" that provides all the answers would be ideal. However, this capability is not yet available, and, as explained in this Section, researchers are still far from reaching that goal. This is the case even for simple systems, such as metal pair that form solid solutions.

The opportunities for understanding interfacial behavior come from studies conducted at the atomic level (for example, using scanning tunneling microscopy, transmission electron microscopy, first principles energy band calculations, and embedded atom potential calculations) with appropriate correlations to both micromechanics and a macroscopic level, and vice versa. The computational and experimental tools that begin to provide insight at these levels have only recently been developed.

Because the interest in interfaces is so diverse (all combinations of metals, ceramics, semiconductors, glasses, and conducting polymers), the advances in interfacial science encompass the creative efforts of a large number of people and organizations. Thus, the quantity of available literature seems overwhelming, and perhaps, disjointed. It will become apparent that there is no preferred approach to the treatment of dissimilar welds or welding. It really depends on the interests and objectives of each engineer.

For example, if you know that what you are doing actually works (that is, produces a joint) and are searching for understanding, then this Section of the Handbook will guide you to the appropriate technical literature. If you have an idea and need guidance for bonding, testing, processing, or interpreting the data, then this Section should also provide that direction. Lastly, if the reader wants state-of-the-art research and development topics, then those can be found in this Section, as well. This should provide the reader with a comprehensive overview of the welding of dissimilar metals.

Therefore, this Section is concerned with the fundamentals of welding and joining materials via the application of a nonmelting process. The specific processes usually associated with this technology are also discussed. Before considering nonmelting processes in detail, however, a brief discussion of adhesion is presented. This is followed by articles that discuss the fundamentals of the major solid-state welding techniques.

References

1. P.G. CHARALAMBIDES, J. LUND, A.G. EVANS, AND R.M. MCMEEKING, "A TEST FOR DETERMINING THE FRACTURE RESISTANCE OF BIMATERIAL INTERFACES," REPORT M87-4, MATERIALS DEPARTMENT, COLLEGE OF ENGINEERING, UC SANTA BARBARA
2. Z. SUO AND J.W. HUTCHINSON, "INTERFACE CRACK BETWEEN TWO ELASTIC LAYERS," REPORT MECH-118, DIVISION OF APPLIED SCIENCES, HARVARD UNIVERSITY, MARCH 1988
3. N.P. O'DOWD, C.F. SHIH, AND M.G. STOUT, TEST GEOMETRIES FOR MEASURING INTERFACIAL FRACTURE TOUGHNESS, FOR *INT. J. SOLIDS STRUCT.,* VOL 29, 1992, P 571-589
4. J.W. HUTCHINSON AND Z. SUO, MIXED MODE CRACKING IN LAYERED MATERIALS, *ADV. IN APPL. MECH.,* VOL 28, J.W. HUTCHINSON AND T.Y. WU, ED., ACADEMIC PRESS, 1990
5. H.M. JENSEN, MIXED MODE INTERFACE FRACTURE CRITERIA, *ACTA METALL. MATER.,* VOL 38 (NO. 12), 1990, P 2637-2644
6. R.L. BRADY, R.S. PORTER, AND J.A. DONOVAN, A BUCKLED PLATE TEST TO MEASURE INTERFACIAL TOUGHNESS IN COMPOSITES, *J. MATER. SCI.,* VOL 24, 1989, P 4138-4143
7. M.-Y. HE AND J.W. HUTCHINSON, "KINKING OF A CRACK OUT OF AN INTERFACE," REPORT MECH-113, DIVISION OF APPLIED SCIENCE, HARVARD UNIVERSITY, FEB 1988
8. J.J. MECHOLSKY AND L.M. BAKER, A CHEVRON-NOTCHED SPECIMEN FOR FRACTURE TOUGHNESS MEASUREMENTS OF CERAMIC-METAL INTERFACES, *CHEVRON-NOTCHED SPECIMENS: TESTING AND STRESS ANALYSIS,* STP 855, J.H. UNDERWOOD, S.W. FREIMAN, AND F.I. BARATTA, ED., ASTM, 1984, P 324-336
9. W.W. GERBERICH, D.L. DAVIDSON, AND M. KACZOROWSKI, EXPERIMENTAL AND THEORETICAL STRAIN DISTRIBUTIONS FOR STATIONARY AND GROWING CRACKS, *J. MECH. PHYS. SOLIDS,* VOL 38 (NO. 1), 1990, P 87-113
10. J. QU AND Q. LI, INTERFACIAL DISLOCATION AND ITS APPLICATIONS TO INTERFACE CRACKS IN ANISOTROPIC BIMATERIALS, *J. ELAST.,* VOL 26, 1991, P 169-195

11. A.N. STROH, DISLOCATIONS AND CRACKS IN ANISOTROPIC ELASTICITY, *PHILOS. MAG.,* VOL 7, 1958, P 625-646
12. Q.Q. LI AND T.C.T. TING, LINE INCLUSION IN ANISOTROPIC ELASTIC SOLIDS, *J. APPL. MECH.,* VOL 56, 1989, P 556-563
13. T.C.T. TING, EXPLICIT SOLUTION AND INVARIANCE OF THE SINGULARITIES AT AN INTERFACE CRACK IN ANISOTROPIC COMPOSITES, *INT. J. SOLIDS STRUCT.,* VOL 22, 1986, P 965-983
14. M.D. PASHLEY AND D. TABOR, ADHESION AND DEFORMATION PROPERTIES OF CLEAN AND CHARACTERIZED METAL MICRO-CONTACTS, *VACUUM,* VOL 10-12, 1981, P 619-623

Introduction to Solid-State Welding

Ray Dixon, Los Alamos National Laboratory

References

1. P.G. CHARALAMBIDES, J. LUND, A.G. EVANS, AND R.M. MCMEEKING, "A TEST FOR DETERMINING THE FRACTURE RESISTANCE OF BIMATERIAL INTERFACES," REPORT M87-4, MATERIALS DEPARTMENT, COLLEGE OF ENGINEERING, UC SANTA BARBARA
2. Z. SUO AND J.W. HUTCHINSON, "INTERFACE CRACK BETWEEN TWO ELASTIC LAYERS," REPORT MECH-118, DIVISION OF APPLIED SCIENCES, HARVARD UNIVERSITY, MARCH 1988
3. N.P. O'DOWD, C.F. SHIH, AND M.G. STOUT, TEST GEOMETRIES FOR MEASURING INTERFACIAL FRACTURE TOUGHNESS, FOR *INT. J. SOLIDS STRUCT.,* VOL 29, 1992, P 571-589
4. J.W. HUTCHINSON AND Z. SUO, MIXED MODE CRACKING IN LAYERED MATERIALS, *ADV. IN APPL. MECH.,* VOL 28, J.W. HUTCHINSON AND T.Y. WU, ED., ACADEMIC PRESS, 1990
5. H.M. JENSEN, MIXED MODE INTERFACE FRACTURE CRITERIA, *ACTA METALL. MATER.,* VOL 38 (NO. 12), 1990, P 2637-2644
6. R.L. BRADY, R.S. PORTER, AND J.A. DONOVAN, A BUCKLED PLATE TEST TO MEASURE INTERFACIAL TOUGHNESS IN COMPOSITES, *J. MATER. SCI.,* VOL 24, 1989, P 4138-4143
7. M.-Y. HE AND J.W. HUTCHINSON, "KINKING OF A CRACK OUT OF AN INTERFACE," REPORT MECH-113, DIVISION OF APPLIED SCIENCE, HARVARD UNIVERSITY, FEB 1988
8. J.J. MECHOLSKY AND L.M. BAKER, A CHEVRON-NOTCHED SPECIMEN FOR FRACTURE TOUGHNESS MEASUREMENTS OF CERAMIC-METAL INTERFACES, *CHEVRON-NOTCHED SPECIMENS: TESTING AND STRESS ANALYSIS,* STP 855, J.H. UNDERWOOD, S.W. FREIMAN, AND F.I. BARATTA, ED., ASTM, 1984, P 324-336
9. W.W. GERBERICH, D.L. DAVIDSON, AND M. KACZOROWSKI, EXPERIMENTAL AND THEORETICAL STRAIN DISTRIBUTIONS FOR STATIONARY AND GROWING CRACKS, *J. MECH. PHYS. SOLIDS,* VOL 38 (NO. 1), 1990, P 87-113
10. J. QU AND Q. LI, INTERFACIAL DISLOCATION AND ITS APPLICATIONS TO INTERFACE CRACKS IN ANISOTROPIC BIMATERIALS, *J. ELAST.,* VOL 26, 1991, P 169-195
11. A.N. STROH, DISLOCATIONS AND CRACKS IN ANISOTROPIC ELASTICITY, *PHILOS. MAG.,* VOL 7, 1958, P 625-646
12. Q.Q. LI AND T.C.T. TING, LINE INCLUSION IN ANISOTROPIC ELASTIC SOLIDS, *J. APPL. MECH.,* VOL 56, 1989, P 556-563
13. T.C.T. TING, EXPLICIT SOLUTION AND INVARIANCE OF THE SINGULARITIES AT AN INTERFACE CRACK IN ANISOTROPIC COMPOSITES, *INT. J. SOLIDS STRUCT.,* VOL 22, 1986, P 965-983
14. M.D. PASHLEY AND D. TABOR, ADHESION AND DEFORMATION PROPERTIES OF CLEAN AND CHARACTERIZED METAL MICRO-CONTACTS, *VACUUM,* VOL 10-12, 1981, P 619-623

Fundamentals of Metal and Metal-to-Ceramic Adhesion

Ray Dixon and S.P. Chen, Los Alamos National Laboratory

Introduction

IN ANY SOLID-STATE (NONMELTING) WELDING PROCESS, there are two primary areas of concern: Will the material bond and how strong is the bond? Literature that deals with the science and technology of material compatibility and bond strength is readily available. This article will review recent work in quantifying adhesion, bonding, and interfacial characterization and strength. Subsequent articles in this Section of the Handbook focus on the engineering principles associated with the formation of welds using these techniques. This discussion concentrates on metal-metal configurations, with only passing comments on metal-ceramic and metal/ceramic-semiconductor interfaces. This restriction is necessary to limit the length of this article. Wherever appropriate, applications of thin films on substrates will also be used.

Adhesion Energy

Surface energy, Γ_{surf}, is defined as the increase in energy, ΔE, of a system divided by the area, A, caused by the creation of the surface:

$$\Gamma_{SURF} = \Sigma \Delta E_I / A \qquad \textbf{(EQ 1)}$$

The atoms that contribute to the extra energy, ΔE_i, are usually within 10 lattice parameters or less of the surface (or interface) (Ref 1, 2, 3). A freshly created surface has a high surface energy, whereas a relaxed or reconstructed one is lower, by about 3 to 5%. At 0 K (-459 °F), surface energies vary about 15 to 25% for planes of various indices. The variation is less at higher temperatures, because of entropy. For example, the (111) surface of a face-centered cubic (fcc) structure has the lowest energy, whereas the (210) surface is about 15% higher (Ref 1, 2, 3). The ideal fracture energy, Γ_{coh}, which is used in estimating the Griffith criterion, is twice the surface energy:

$$\Gamma_{COH}\ (HKL) = 2 \cdot \Gamma_{SURF}\ (HKL) \qquad \textbf{(EQ 2)}$$

where (hkl) are the indices of the plane. The actual energy consumed in fracture or crack propagation is composed of two terms: one is the Γ_{coh} and the other is the plastic work consumed per unit area $\Gamma_{plastic}$ (Ref 4, 5, 6, 7, 8). For ionic crystals, $\Gamma_{plastic}$ is usually a few times Γ_{coh}; for metals, Γ_{coh} is about 1 to 4 J/m^2 (2×10^{-5} to 1×10^{-4} cal/cm^2); and $\Gamma_{plastic}$ is about 1000 J/m^2 (2×10^{-2} cal/cm^2). For ceramic systems, Γ_{coh} is about the same, but $\Gamma_{plastic}$ is about 10 to 100 J/m^2 (2×10^{-4} to 2×10^{-3} cal/cm^2) (Ref 7). $\Gamma_{plastic}$ and Γ_{coh} are related, because when Γ_{coh} is zero, no plastic work ($\Gamma_{plastic}$) will be consumed. The exact form of this dependence is not known, although some attempts have been made (Ref 4, 5, 6) to correlate the two terms.

References cited in this section

1. S.P. CHEN, D.J. SROLOVITZ, AND A.F. VOTER, COMPUTER SIMULATION ON SURFACES AND [001] TILT GRAIN BOUNDARIES IN NI, AL, AND NI$_3$AL, *J. MATER. RES.*, VOL 4, 1989, P 62-77
2. S.P. CHEN AND A.F. VOTER, RECONSTRUCTION OF THE (310), (210), AND (110) SURFACES IN FCC METALS, *SURF. SCI. LETT.*, VOL 244, 1991, P L107-L112
3. S.P. CHEN, THEORETICAL STUDIES OF METALLIC INTERFACES, *MATER. SCI. ENG. B*, VOL 6, 1990, P 113-121

4. M.L. JOKL, V. VITEK, AND C.J. MCMAHON, JR., A MICROSCOPIC THEORY OF BRITTLE FRACTURE IN DEFORMABLE SOLIDS: A RELATION BETWEEN IDEAL WORK TO FRACTURE AND PLASTIC WORK, *ACTA METALL.*, VOL 28, 1980, P 1479-1488
5. M.L. JOKL, V. VITEK, AND C.J. MCMAHON, JR., ON THE MICROMECHANICS OF BRITTLE FRACTURE: EXISTING VS INJECTED CRACKS, *ACTA METALL.*, VOL 37, 1989, P 87-97
6. J. HACK, S.P. CHEN, AND D. SROLOVITZ, A KINETIC CRITERION FOR QUASI-BRITTLE FRACTURE, *ACTA METALL.*, VOL 37, 1989, P 1957-1970
7. A. KELLY AND N.H. MACMILLAN, *STRONG SOLIDS*, 3RD ED., OXFORD SCIENCE PUBLICATIONS, 1986
8. V. VITEK, MICROMECHANISMS OF INTERGRANULAR BRITTLE FRACTURE IN INTERMETALLIC COMPOUND, *J. PHYS. III*, VOL 1, 1991, P 1085-1097

Fundamentals of Metal and Metal-to-Ceramic Adhesion

Ray Dixon and S.P. Chen, Los Alamos National Laboratory

Grain Boundary Energy

The grain boundary energy, Γ_{gb}, is defined the same way as surface energy is, but it depends on the relative orientation of the adjacent grains (Ref 1):

$$\Gamma_{GB} = \Sigma \Delta E_I / A_I \ (GB) \qquad \textbf{(EQ 3)}$$

At misorientation angles of greater than 25°, Γ_{gb} usually reaches a plateau that characterizes random grain boundaries. Special grain boundaries, like $\Sigma 5$ (210) or (310), tend to have slightly lower energies (Ref 1, 2, 3, 9). Twin boundary energies are even lower than special grain boundaries. For metals, Γ_{gb} usually ranges from 40 to 60% of the surface energy (Ref 10), whereas the twin energy is only about 5 to 15%. Grain boundary cohesive energy is defined as (Ref 1):

$$\Gamma_{COH} \ (GB) = \Gamma_{SURF} \ (GRAIN\ A) + \Gamma_{SURF} \ (GRAIN\ B) - \Gamma_{GB} \qquad \textbf{(EQ 4)}$$

For general and random grain boundaries, grain boundary cohesive energy is smaller. General grain boundaries tend to be weaker than the small angle boundaries.

References cited in this section

1. S.P. CHEN, D.J. SROLOVITZ, AND A.F. VOTER, COMPUTER SIMULATION ON SURFACES AND [001] TILT GRAIN BOUNDARIES IN NI, AL, AND NI_3AL, *J. MATER. RES.*, VOL 4, 1989, P 62-77
2. S.P. CHEN AND A.F. VOTER, RECONSTRUCTION OF THE (310), (210), AND (110) SURFACES IN FCC METALS, *SURF. SCI. LETT.*, VOL 244, 1991, P L107-L112
3. S.P. CHEN, THEORETICAL STUDIES OF METALLIC INTERFACES, *MATER. SCI. ENG. B*, VOL 6, 1990, P 113-121
9. S.P. CHEN, STUDIES OF IRIDIUM SURFACES AND GRAIN BOUNDARIES, *PHIL. MAG. A*, VOL 65, 1992, TO BE PUBLISHED
10. L.E. MURR, *INTERFACIAL PHENOMENA IN METALS AND ALLOYS*, ADDISON-WESLEY, 1975

Fundamentals of Metal and Metal-to-Ceramic Adhesion

Ray Dixon and S.P. Chen, Los Alamos National Laboratory

Interfacial Energy

By definition, the interfacial energy (Ref 2, 3) between two bulk materials (A and B) is the same as the grain boundary energy, except that there are two reference states:

$$\Gamma_{\text{INTERFACE}} = [\Sigma\Delta E_I\,(\text{A}) + \Sigma\Delta E_J(\text{B})]/A \qquad \textbf{(EQ 5)}$$

where $\Delta E_i(\text{A}) \cdot [(\Delta E_j(\text{B})]$ is the extra energy of atom *i* of A (B) type, compared to its bulk state of A (B) materials. The interfacial energy between similar metals tends to be small. For dissimilar materials, however, interfacial energies tend toward large positive values if they do not form compounds and large negative values if they do form compounds. The ideal adhesion energy is similar to grain boundary cohesion and is the difference between the two surfaces and the interfacial energies:

$$\Gamma_{\text{ADHESION}} = \Gamma_{\text{SURFACE}}\,(\text{A}) + \Gamma_{\text{SURFACE}}\,(\text{B}) - \Gamma_{\text{INTERFACE}}\,(\text{A/B}) \qquad \textbf{(EQ 6)}$$

References cited in this section

2. S.P. CHEN AND A.F. VOTER, RECONSTRUCTION OF THE (310), (210), AND (110) SURFACES IN FCC METALS, *SURF. SCI. LETT.*, VOL 244, 1991, P L107-L112
3. S.P. CHEN, THEORETICAL STUDIES OF METALLIC INTERFACES, *MATER. SCI. ENG. B*, VOL 6, 1990, P 113-121

Fundamentals of Metal and Metal-to-Ceramic Adhesion

Ray Dixon and S.P. Chen, Los Alamos National Laboratory

Theory of Adhesion

Rose *et al.* (Ref 11) have calculated the bimaterial interfacial cohesion for many combinations of elements using band structure methods. They found a universal scaling relationship for the adhesion of different materials that extends beyond metal-metal interfaces to nonmetal systems. A few calculations are necessary to identify the adhesion behavior of the system and set the appropriate energy, length, and elastic constant scales. Plastic contributions are not treated in this formalism. The plastic work is more likely to be solved by incorporating the atomistic simulation and continuum analysis of Jokl *et al.* (Ref 4, 5) and Hack *et al.* (Ref 6).

The tight binding method give qualitative trends as the energy band is filled for calculations using the elements in the periodic table (Ref 12, 13, 14). The correct trend seems to be captured by the band-filling argument pioneered in the original work of Friedel (Ref 15). For noble metals, a rigid band picture is preferred (Ref 14). Another useful tool is the Pettifor structure map. By using the Mendelev number for each of the 103 elements (Ref 13, 16), Pettifor found that alloying tendencies can be grouped into separate regions. If the interfacial chemistry and average Mendelev number are known, then the interfacial structure can be found from the Pettifor map.

Extensive atomistic simulation work, using effective medium theory (Ref 17, 18), embedded atom potentials (Ref 19), or local volume potentials (Ref 1, 2, 3, 8, 9, 20, 21, 22, 23) had been very useful in describing the properties of many metals and compounds and their interactions with impurities such as boron and sulfur in Ni_3Al and NiAl (Ref 20, 21). These atomic interaction descriptions include pair potential and many body interactions and have been obtained for systems that include Ni_3Al, Cu_3Au, NiAl, nickel, aluminum, iridium, gold, platinum, silver, copper, molybdenum, tantalum, iron, tungsten, niobium, beryllium, titanium, hafnium, zirconium, and rhenium (Ref 1, 2, 3, 9, 20, 21, 22, 23, 24, 25, 26, 27). These studies indicate that stoichiometry (Ref 1, 20, 21) and impurities (Ref 20, 21) can be crucial in determining ideal and realistic adhesion, including effects induced by boron and sulfur. They represent the fundamentals for future work, which is expected to have significant impact on the theory using atomistic simulation under well-controlled situations for all conditions of stress, temperature, and chemical environments. These calculations should enable direct linking of theoretical studies and materials by design concepts, as applied in NiAl microalloyed with boron or iron (Ref 24, 25, 26, 27).

DeBoer *et al.* (Ref 28) have developed excellent tools for screening bonding materials, based on two properties of the elements: electronegativity and electron density. One can use the Miedema formula for the adhesion of a bimetallic interface, based on the work function (closely related to the Pauling electronegativity), Φ, and the electron density, *n*, at the Wigner-Seitz radius, as described in Ref 28.

References cited in this section

1. S.P. CHEN, D.J. SROLOVITZ, AND A.F. VOTER, COMPUTER SIMULATION ON SURFACES AND [001] TILT GRAIN BOUNDARIES IN NI, AL, AND NI_3AL, *J. MATER. RES.,* VOL 4, 1989, P 62-77
2. S.P. CHEN AND A.F. VOTER, RECONSTRUCTION OF THE (310), (210), AND (110) SURFACES IN FCC METALS, *SURF. SCI. LETT.,* VOL 244, 1991, P L107-L112
3. S.P. CHEN, THEORETICAL STUDIES OF METALLIC INTERFACES, *MATER. SCI. ENG. B,* VOL 6, 1990, P 113-121
4. M.L. JOKL, V. VITEK, AND C.J. MCMAHON, JR., A MICROSCOPIC THEORY OF BRITTLE FRACTURE IN DEFORMABLE SOLIDS: A RELATION BETWEEN IDEAL WORK TO FRACTURE AND PLASTIC WORK, *ACTA METALL.,* VOL 28, 1980, P 1479-1488
5. M.L. JOKL, V. VITEK, AND C.J. MCMAHON, JR., ON THE MICROMECHANICS OF BRITTLE FRACTURE: EXISTING VS INJECTED CRACKS, *ACTA METALL.,* VOL 37, 1989, P 87-97
6. J. HACK, S.P. CHEN, AND D. SROLOVITZ, A KINETIC CRITERION FOR QUASI-BRITTLE FRACTURE, *ACTA METALL.,* VOL 37, 1989, P 1957-1970
8. V. VITEK, MICROMECHANISMS OF INTERGRANULAR BRITTLE FRACTURE IN INTERMETALLIC COMPOUND, *J. PHYS. III,* VOL 1, 1991, P 1085-1097
9. S.P. CHEN, STUDIES OF IRIDIUM SURFACES AND GRAIN BOUNDARIES, *PHIL. MAG. A,* VOL 65, 1992, TO BE PUBLISHED
11. J.H. ROSE, J.R. SMITH, F. GUINEA, AND J. FERRANTE, UNIVERSAL FEATURES OF THE EQUATION OF STATE OF METALS, *PHYS. REV. B,* VOL 29, 1984, P 2963-2969
12. T.B. MASSALSKI, ED., *BINARY ALLOYS PHASE DIAGRAMS,* ASM, 1986, P 2176-2181
13. D.G. PETTIFOR, ELECTRON THEORY OF METALS, *PHYSICAL METALLURGY,* R.W. CAHN AND P. HAASEN, ED., NORTH-HOLLAND, AMSTERDAM, 1983, P 73-152
14. J.S. FAULKNER, THE MODERN THEORY OF ALLOYS, *PROG. MATER. SCI.,* VOL 27, 1982, P 1-187
15. J. FRIEDEL, TRANSITION METALS, ELECTRONIC STRUCTURE OF THE D-BAND, ITS ROLE IN THE CRYSTALLINE AND MAGNETIC STRUCTURES, *THE PHYSICS OF METALS,* J.M. ZIMAN, ED., CAMBRIDGE UNIVERSITY PRESS, 1969, P 340-408
16. D.G. PETTIFOR, NEW MANY-BODY POTENTIAL FOR THE BOND ORDER, *PHYS. REV. LETT.,* VOL 63, 1989, P 2480-2483
17. J.K. NORSKOV AND N.D. LANG, EFFECTIVE-MEDIUM THEORY OF CHEMICAL BINDING: APPLICATION TO CHEMISORPTION, *PHYS. REV. B,* VOL 21, 1980, P 2131-2136
18. M.J. STOTT AND E. ZAREMBA, QUASIATOMS: AN APPROACH TO ATOMS IN NONUNIFORM ELECTRONIC SYSTEMS, *PHYS. REV. B,* VOL 22, 1980, P 1564-1583
19. M.S. DAW AND M.I. BASKES, EMBEDDED-ATOM METHOD: DERIVATION AND APPLICATION TO IMPURITIES, SURFACES, AND OTHER DEFECTS IN METALS, *PHYS. REV. B,* VOL 29, 1984, P 6443-6453
20. S.P. CHEN, A.F. VOTER, R.C. ALBERS, A.M. BORING, AND P.J. HAY, INVESTIGATION OF BORON EFFECT ON NI AND NI_3AL GRAIN BOUNDARIES, *J. MATER. RES.,* VOL 5, 1990, P 955-970

21. S.P. CHEN, A.F. VOTER, R.C. ALBERS, A.M. BORING, AND P.J. HAY, THEORETICAL STUDIES OF GRAIN BOUNDARIES IN NI_3AL WITH BORON OR SULFUR, *SCR. METALL.*, VOL 23, 1989, P 217-222
22. A.F. VOTER AND S.P. CHEN, ACCURATE INTERATOMIC POTENTIALS FOR NI, AL, AND NI_3AL, *MATER. RES. SOC. PROC.*, VOL 82, 1987, P 175-180
23. S.P. CHEN, ANOMALOUS RELAXATIONS OF (0001) AND $(10\bar{1}0)$ SURFACES IN HCP METALS, *SURF. SCI. LETT.*, VOL 264, 1992, P L162-L168
24. R.J. HARRISON, A.F. VOTER, AND S.P. CHEN, *ATOMISTIC MODELING OF MATERIALS: BEYOND PAIR POTENTIALS,* V. VITEK AND D.J. SROLOVITZ, ED., PLENUM PRESS, 1989, P 219-222
25. R.J. HARRISON, F. SPAEPEN, A.F. VOTER, AND S P. CHEN, STRUCTURE OF GRAIN BOUNDARIES IN IRON, *PROC. 34TH SAGAMORE ARMY MATERIALS RESEARCH CONFERENCE* (LAKE GEORGE, NY), 1990, P 651-675 (FOR FE)
26. S.P. CHEN, LOCAL VOLUME POTENTIALS FOR BCC METALS, TO BE PUBLISHED (FOR FE, NB, CR, W, AND MO)
27. S.P. CHEN, A.M. BORING, R.C. ALBER, AND P.J. HAY, THEORETICAL STUDIES OF NI_3AL AND NIAL WITH IMPURITIES, *HIGH TEMPERATURE ORDERED INTERMETALLIC ALLOYS III,* C.C. KOCH, C.T. LIU, N.S. STOLOFF, AND A.I. TAUB, ED., *PROC. MATER. RES. SOC.*, VOL 133, 1989, P 149-154
28. F.R. DEBOER, R. BOOM, W.C.M. MATTENS, A.R. MIEDEMA, AND A.K. NIESSEN, *COHESION IN METALS: TRANSITION METAL ALLOYS,* NORTH-HOLLAND, 1989

Fundamentals of Metal and Metal-to-Ceramic Adhesion

Ray Dixon and S.P. Chen, Los Alamos National Laboratory

Properties Affecting Adhesion

Large impurities, such as bismuth, sulfur, phosphorus, and so on, tend to lower grain boundary adhesion, whereas small interstitial atoms, such as boron, carbon, and beryllium tend to increase cohesion in nickel, Ni_3Al, and Ni_3Si (Ref 1, 20, 21). Deviations from stoichiometry can either increase or decrease cohesion, depending on whether amounts of strong bonding elements, like nickel in Ni_3Al, are increased or decreased. Although stoichiometry may slightly affect the ideal cohesion energy, its effect on the consumption of plastic work is potentially higher, as demonstrated in Co_3Ti. Researchers (Ref 29) found that excess cobalt suppressed intergranular brittleness and rendered the material ductile, without adding strengtheners such as boron or beryllium. Sometimes impurity effects only work in off-stoichiometric compounds, such as boron in Ni_3Al. Quantitative analysis of these kinds of effects have been attempted for impurity-controlled fracture properties in iron, which has been studied extensively by Hondros and Seah (Ref 30). A simple linear formula linking cohesion and level of impurities has been extracted from many experiments (Ref 30).

References cited in this section

1. S.P. CHEN, D.J. SROLOVITZ, AND A.F. VOTER, COMPUTER SIMULATION ON SURFACES AND [001] TILT GRAIN BOUNDARIES IN NI, AL, AND NI_3AL, *J. MATER. RES.*, VOL 4, 1989, P 62-77
20. S.P. CHEN, A.F. VOTER, R.C. ALBERS, A.M. BORING, AND P.J. HAY, INVESTIGATION OF BORON EFFECT ON NI AND NI_3AL GRAIN BOUNDARIES, *J. MATER. RES.*, VOL 5, 1990, P 955-970
21. S.P. CHEN, A.F. VOTER, R.C. ALBERS, A.M. BORING, AND P.J. HAY, THEORETICAL STUDIES OF GRAIN BOUNDARIES IN NI_3AL WITH BORON OR SULFUR, *SCR. METALL.*, VOL 23, 1989, P 217-222
29. J.E. HACK, D.J. STROLOVITZ, AND S.P. CHEN, A MODEL FOR THE FRACTURE BEHAVIOR OF POLYCRYSTALLINE NI_3AL, *SCR. METALL.*, VOL 20, 1986, P 1699-1704

30. E.D. HONDROS AND M.P. SEAH, INTERFACIAL AND SURFACE MICROCHEMISTRY, *PHYSICAL METALLURGY*, 3RD ED., R.W. CAHN AND P. HAASEN, ED., ELSEVIER SCIENCE PUBLISHER, 1983, P 855-931

Fundamentals of Metal and Metal-to-Ceramic Adhesion

Ray Dixon and S.P. Chen, Los Alamos National Laboratory

Adhesion Measurement

Sikorski (Ref 31) describes two modifications of the twist-compression bonding method for adhesion testing and concludes that the measured coefficient of friction supports the adhesion theory of friction. This correlation implies that a large coefficient of adhesion (ratio of the force necessary to break a bond to the normal loading force used to produce the bond) corresponds to a large coefficient of friction.

Experimentally, he showed that the coefficient of adhesion was inversely proportional to the melting point, which meant that higher melting point materials had lower adhesion. This agress with the observation that higher melting point materials have higher hardness, which also corresponds to lower adhesion coefficients. Gerkema and Miedema (Ref 32) considered lubrication of iron substrates by thin films of lead and silver with copper, platinum, or molybdenum additions by calculating surface adhesion using heats of solution and surface energies. Their predictions agree well with sliding wear experiments.

Pashley and Tabor (Ref 33) measured adhesion between tungsten and nickel using techniques similar to those used in atomic force microscopy (AFM). They were able to verify surface cleanliness and ensure that the adhesion measured was between clean tungsten and nickel by using Auger spectroscopy (AES). In a subsequent publication, Pashley and Pethia again measured the surface energy of adhesion between tungsten and nickel (Ref 34). This work was done between a tungsten tip and a nickel flat in high vacuum and showed that surface forces alone could cause plastic flow. The loading/unloading curves provide information on the elastic/plastic response of the material. This in turn determines the interpretation of the adhesion.

Johnson and Keller (Ref 35) used a torsion balance with a 3 mm (0.12 in.) diameter ball against a flat plate to measure the force of adhesion between silver-silver, copper-nickel, and silver-nickel couples. They concluded that mutual solubility did not affect the tendency for adhesion. However, surface contaminants did have a significant effect on the adhesion (joint strength) between silver-silver couples when the vacuum varied. The difference in joint strength for a clean surface at 130 μPa (5×10^{-10} torr) and for the same surface after a few microns of argon was 50 GPa (5×10^{8} dynes/cm^2).

Conrad and Rice (Ref 36) studied the cohesion of fcc metals (silver, aluminum, copper, and nickel) by fracturing the metals in high vacuum, rewelding the metals by compressing the fractured surfaces, and retesting the metals by fracturing. They found that the cohesive strength of the weld increased with compressive load when both were normalized to the initial fracture strength. Results for all materials fell on a single curve. They proposed that the essentially constant cohesion coefficient was due to the fact that the contact area produced by a given compressive load was inversely proportional to the fracture stress of the metal, whereas the cohesive strength was directly proportional to the fracture stress.

Acoustic emission was used to measure and monitor cohesion. In one case, this technique was used to signal the onset of fiber debonding and to determine crack initiation locations in fiber-reinforced ceramic-matrix composites (Ref 37). This approach is a valuable and viable technique for monitoring debonding, but cannot be used to provide quantitative information on either bond strength or fundamental material parameters. It can be used to locate the origin of the debonding, but the error is on the order of the dimensions used for continuum calculations, which are much larger than the atomic dimensions.

In another study (Ref 38), acoustic measurements were used to monitor fatigue cracks and interfaces by defining a spring stiffness (for continuous and discontinuous spring models) that represented the degree of contact across an interface. A measured acoustic reflection coefficient was found to be in good agreement with one predicted for the same geometry. In addition, a single stress-intensity factor was found to determine bond strength.

In a similar vein, Neid (Ref 39) developed a technique for measuring bond integrity by propagating a planar shock were through the bond region. Depending on the shock energy, defects developed in the bond. The change in density was used to detect defects created at the interface, because other nondestructive test (NDT) methods were not sensitive enough to detect the small defects created. The defects were subsequently verified using optical microscopy.

Itou (Ref 40) used electrical resistance to monitor crack tip contact in order to verify crack tip shielding in aluminum, steel, and cork. The results were inconclusive, although he found that there was no oscillating stress singularity around the crack tip. From this discussion, one can see that no generally quick, accurate, and easy-to-use technique has been developed to measure adhesion.

References cited in this section

31. M.E. SIKORSKI, THE ADHESION OF METALS AND FACTORS THAT INFLUENCE IT, *WEAR,* VOL 7, 1964, P 144-162
32. J. GERKEMA AND A.R. MIEDEMA, ADHESION BETWEEN SOLID METALS: OBSERVATIONS OF INTERFACIAL SEGREGATION EFFECTS IN METAL FILM LUBRICATION EXPERIMENTS, *SURF. SCI.,* VOL 124, 1983, P 351-371
33. M.D. PASHLEY AND D. TABOR, ADHESION AND DEFORMATION PROPERTIES OF CLEAN AND CHARACTERIZED METAL MICRO-CONTACTS, *VACUUM,* VOL 10-12, 1981, P 619-623
34. M.D. PASHLEY AND J.B. PETHICA, THE ROLE OF SURFACE FORCES IN METAL-METAL CONTACTS, *J. VAC. SCI. TECHNOL. A,* VOL 3 (NO. 3), MAY/JUNE 1985, P 757-761
35. K.I. JOHNSON AND D.V. KELLER, JR., EFFECT OF CONTAMINATION ON THE ADHESION OF METALLIC COUPLES IN ULTRA-HIGH VACUUM, *J. APPL. PHYS.,* VOL 38 (NO. 4), JAN 1967, P 1896-1904
36. H. CONRAD AND L. RICE, THE COHESION OF PREVIOUSLY FRACTURED FCC METALS IN ULTRA-HIGH VACUUM, *METALL. TRANS.,* VOL 1, NOV 1970, P 3019-3029
37. T. KISHI, M. ENOKI, AND H. TSUDA, INTERFACE AND STRENGTH IN CERAMIC MATRIX COMPOSITES, *MATER. SCI. ENG.,* VOL A143, 1991, P 103-110
38. O. BUCK, R.B. THOMPSON, D.K. REHBEIN, D.D. PALMER, AND L.J.H. BRASCHE, CONTACTING SURFACES: A PROBLEM IN FATIGUE AND DIFFUSION BONDING, *METALL. TRANS. A,* VOL 20, APRIL 1989, P 627-635
39. H.A. NIED, "DETERMINATION OF THE STRUCTURAL INTEGRITY OF METALLURGICALLY BONDED PLATES BY HIGH ENERGY IMPULSE TEST," *SIXTH INT. CONF. ON EXP. STRESS ANALYSIS* (MUNICH), SEPT 1978
40. S. ITOU, AN EXPERIMENT ON THE BEHAVIOUR AROUND THE TIPS OF AN INTERFACE CRACK, *ENG. FRACT. MECH.,* VOL 37 (NO. 1), 1990, P 145-150

Fundamentals of Metal and Metal-to-Ceramic Adhesion

Ray Dixon and S.P. Chen, Los Alamos National Laboratory

Interfacial Analysis

Electron spectroscopies are extremely valuable for the quantitative and qualitative analysis of surfaces. Fundamental information about the atomic forces and energies of the regions being probed, as well as their geometry, can be obtained. The shortcoming of these techniques in analyzing interfaces is that the interfacial volume is a small fraction of the probed volume and the effects of the interface cannot be separated from the signal. This condition can be overcome by using scanning tunneling microscopy (STM) or AFM. However, preparation of the surface is extremely difficult, the equipment requires a skilled and trained operator, the data are fundamental, and engineering information is not obtained. The crux of the problem is that there is no simple connection between condensed matter theory, micromechanics, and continuum mechanics. The following discussion briefly reviews some of the available technical literature that describes the use of these spectroscopies and their application to bimaterial systems.

Photoelectron spectroscopy has been used to study electron core-level shifts for ytterbium, aluminum, and silicon grown on molybdenum (110) (Ref 41, 42). The core-level shifts between the film and the substrate are expressed in terms of adhesion energy. In this case, the adhesion energy is the difference between the energy of the neutral and excited atoms per atom, which is readily found from the spectra. The measured adhesion energy is in good agreement with the adhesion energy determined using the Miedema technique (Ref 28). High-resolution electron microscopy (HREM), electron energy loss spectroscopy (EELS), microspectroscopy, and microdiffraction have been used to analyzed the chemical distributions on or near interface planes (Ref 43). This information can be used to support models requiring detailed stoichiometry, but does not yield adhesion values.

Gerberich *et al.* (Ref 44) used selected-area-electron channeling patterns (SACP) and stereoimaging to quantify the strain in grains. The work was carried out under plane strain and plane stress conditions. The researchers determined strain distributions within 1 μm (40 μin.) of a stationary or growing crack in Fe-3Si single crystals with a {100}<010> crack. Although this arrangement is not a bimaterial interface, the results reflect the effort needed to obtain fundamental materials information.

Optical techniques, such as reflection (Ref 45), have also been used to probe interfaces and to aid in determining interfacial electronic structure. The data, in turn, are related to adhesion and segregation. For example, Lanning *et al.* (Ref 46) recently correlated reflectance spectra with surface free energy. This was then used as a measure of surface energy and the work of decohesion was calculated.

The most promising technology for understanding interfaces is probably STM and associated techniques. Advances in the science and technology of scanning tunneling microscopy/spectroscopy (STM/S), AFM, photon scanning tunneling microscopy (PSTM), scanning chemical potential microscopy (SCPM), and related spectroscopies/microscopies offer new insight into interfaces (Ref 47). Low-energy electron diffraction (LEED), LEED-spot profile analysis (SPA-LEED), scanning low-energy electron microscopy (SLEEM), and photoemission electron microscopy (PEEM) are also being applied to surface analysis and adhesion. These technologies, although currently applied to the vacuum-surface interface, are appropriate tools for the study of grain boundaries and dissimilar metal surfaces. They provide excellent methods for measuring the electronic behavior at interfaces and, thus, the fundamental interactions that occur at these interfaces. This information must then be correlated to engineering properties predicted by micromechanical and continuum mechanics methods, as well as experiments.

References cited in this section

28. F.R. DEBOER, R. BOOM, W.C.M. MATTENS, A.R. MIEDEMA, AND A.K. NIESSEN, *COHESION IN METALS: TRANSITION METAL ALLOYS*, NORTH-HOLLAND, 1989
41. N. MARTENSSON, A. STENBORG, O. BJORNEHOLM, A. NILSSON, AND J.N. ANDERSON, QUANTITATIVE STUDIES OF METAL-METAL ADHESION AND INTERFACE SEGREGATION ENERGIES USING PHOTOELECTRON SPECTROSCOPY, *PHYS. REV. LETT.*, VOL 60 (NO. 17), APRIL 1988, P 1731-1734
42. J.N. ANDERSON, O. BJORNEHOLM, A. STENBORG, A. NILSSON, C. WIGREN, AND N. MARTENSSON, MEASUREMENT OF METAL-METAL ADHESION AND INTERFACE SEGREGATION ENERGIES BY CORE-LEVEL PHOTOELECTRON SPECTROSCOPY. AL AND SI ON MO(110), *J. PHYS., CONDENS. MATTER*, VOL 1, 1989, P 7309-7313
43. R.W. CARPENTER, HIGH RESOLUTION INTERFACE ANALYSIS, *MATER. SCI. ENG.*, VOL A107, 1989, P 207-216
44. W.W. GERBERICH, D.L. DAVIDSON, AND M. KACZOROWSKI, EXPERIMENTAL AND THEORETICAL STRAIN DISTRIBUTIONS FOR STATIONARY AND GROWING CRACKS, *J. MECH. PHYS. SOLIDS*, VOL 38 (NO. 1), 1990
45. A. JOHNER AND P. SCHAAF, CALCULATION OF THE REFLECTION COEFFICIENTS OF INTERFACES: A SCATTERING APPROACH, *PHYS. REV. B*, VOL 42 (NO. 9), SEPT 1990, P 5516-5526

46. B.R. LANNING, T. FURTAK, AND G.R. EDWARDS, A METHOD FOR PREDICTION OF METAL-CERAMIC INTERFACIAL BOND ENERGIES, SUBMITTED FOR PUBLICATION IN *J. MATER. RES.*
47. R.J. COLTON, C.R.K. MARRIAN, AND J.A. STROSCIO, ED., *PROC. FIFTH INT. CONF. SCANNING TUNNELING MICROSCOPY/SPECTROSCOPY (SAM '90)* AND *THE FIRST INT. CONF. NANOMETER SCALE SCIENCE AND TECHNOLOGY (NANO I)* (BALTIMORE, MD), JUNE 1990, AND *J. VAC. SCI. TECH. B,* SECOND SERIES, VOL 9 (NO. 2), PART II, MARCH/APRIL 1991

Fundamentals of Metal and Metal-to-Ceramic Adhesion

Ray Dixon and S.P. Chen, Los Alamos National Laboratory

Welding (Bonding)

An extensive review of diffusion bonding is given by Gerken and Owczarski (Ref 48). This treatise reviewed the status of theory at that time (1965) and cited numerous examples. Eleven years later, Williams and Crossland (Ref 49, 50) highlighted cold and diffusion welding. They pointed out that the major technical concerns for these processes had not significantly changed in the interim. As recently as 1986, there was concern over the knowledge base of bonding and adhesion at interfaces (Ref 51). This group felt that detailed mechanisms of adhesion were largely missing, but that first-principle theories were beginning to appear. A strong need for collaboration between theorists and experimentalists was indicated. Furthermore, much information was still unknown about interfaces, and the field was ripe for expansion. This situation still applies. Described below are examples of some of the proposed models for bonding.

Several mechanisms have been proposed to explain the formation of solid-state welds (Ref 52, 53). Mohamed and Washburn (Ref 52) proposed a two-stage mechanism that depends on breaking the surface oxide. Their work is supported by the work of Bay. (Ref 54), who developed a theoretical model for oxide breakage and bond formation on aluminum. Kawakatsu and Kitayama (Ref 53) describe bonding as being controlled by three metallurgical factors, which have been identified as bonds between solid-solution, two-phase, or intermetallic materials. Their studies involved bond-strength measurements and microstructural examination as a function of bonding conditions for copper-nickel, copper-silver, and iron-aluminum couples. They concluded that a minimum bonding temperature occurs at about the softening temperature of one of the materials. They also found that bond strength depends on recrystallization temperature. In the case of solid-solution materials, the interfaces tend to disappear after sufficient time above the recrystallization temperature, whereas this does not occur in two-phase and intermetallic materials.

Excellent treatises on surface and interfacial segregation are provided by Dowben and Miller (Ref 55) and Hondros and Seah (Ref 30). The concept of migration or segregation of one component of a system to a surface or interface is considered in detail. This work provides insight into the behavior that must be considered when determining the possibility of adhesion between two dissimilar materials.

Garmong *et al.* (Ref 56) describe the rate-limiting step in bonding as the complete elimination of porosity from the bond line and have developed a time-prediction model to support it. The model proceeds in two stages. First, long-wavelength asperities (roughness) are eliminated by plastic flow. Second, voids are eliminated by plastic flow and vacancy diffusion. This model is similar to that proposed by Hill and Wallach (Ref 57), who consider void elimination at the bond line. In this case, voids are created by surface asperities on the mating surfaces and pressure sintering is invoked for its elimination. The effect of grain size is also considered.

Cahn (Ref 58) derived interfacial energy from thermodynamic considerations. He considered how the order parameter changes surface tension, elastic energy, and excess free energy when it affects the lattice parameter of an ordered material at an antiphase boundary. The work is not only appropriate for grain boundary interfaces in ordered materials, but is also relevant for dissimilar materials.

Cline (Ref 59) reviews the difference between deformation and diffusion bonding. He defines diffusion bonding as not requiring significant deformation for bond formation. With respect to deformation bonding, Cline discusses the "film" and "energy barrier" theories and reviews the factors affecting the resultant bond. With respect to diffusion bonding, he defines two stages for bonding: microscopic deformation and diffusion. He then reviews the results for bonding aluminum and titanium and shows that both can be readily deformation bonded. The diffusion bonding of aluminum was simplified by using a silver interlayer, whereas titanium required smooth, clean surfaces and protection from atmospheric contaminants.

Derby and Wallach (Ref 60, 61) developed a thorough model for diffusion welding which accounts for:

- SURFACE DIFFUSION FROM SURFACE SOURCES TO A NECK
- VOLUME DIFFUSION FROM SURFACE SOURCES TO A NECK
- DIFFUSION ALONG THE BOND INTERFACE FROM INTERFACIAL SOURCES TO A NECK
- VOLUME DIFFUSION FROM INTERFACIAL SOURCES TO A NECK
- POWER-LAW CREEP DEFORMING THE RIDGE (SURFACE ROUGHNESS)
- PLASTIC YIELDING DEFORMING THE RIDGE
- VAPOR PHASE MASS TRANSPORT

Mass transport rate equations were developed in terms of the process variables (time, temperature, pressure, and material properties). The model is in reasonable agreement with experiment (Ref 62, 63).

Guo and Ridley (Ref 64) consider the rate-limiting step in bonding to be the removal of interfacial voids that are due to the surface roughness and have developed a model to support this position. The main difference between their model and that of Derby and Wallach (Ref 60) is the void geometry used. Guo and Ridley use a lenticular-shaped cavity and have developed a model that agrees reasonably with the experimental data.

References cited in this section

30. E.D. HONDROS AND M.P. SEAH, INTERFACIAL AND SURFACE MICROCHEMISTRY, *PHYSICAL METALLURGY,* 3RD ED., R.W. CAHN AND P. HAASEN, ED., ELSEVIER SCIENCE PUBLISHER, 1983, P 855-931
48. J.M. GERKEN AND W.A. OWCZARSKI, "A REVIEW OF DIFFUSION WELDING," WELDING RESEARCH COUNCIL BULLETIN NO. 109, OCT 1965
49. J.D. WILLIAMS AND B. CROSSLAND, PART 1 COLD PRESSURE WELDING, *CME,* MARCH 1976, P 65-67
50. J.D. WILLIAMS AND B. CROSSLAND, PART 2 DIFFUSION BONDING, *CME,* APRIL 1976, P 73-75
51. J.R. SMITH, PANEL REPORT ON INTERFACIAL BONDING AND ADHESION, *MATER. SCI. ENG.,* VOL 83, 1986
52. H.A. MOHAMED AND J. WASHBURN, MECHANISM OF SOLID STATE PRESSURE WELDING, *WELD. J. RES. SUPP.,* SEPT 1975, P 302S-310S
53. I. KAWAKATSU AND S. KITAYAMA, STUDY ON DIFFUSION BONDING OF METALS, *TRANS. JIM,* VOL 18, 1977, P 455-465
54. N. BAY, MECHANISMS PRODUCING METALLIC BONDS IN COLD WELDING, *WELD. J. RES. SUPP.,* MAY 1983, P 137S-142S
55. P.A. DOWBEN AND A. MILLER, ED., *SURFACE SEGREGATION PHENOMENA,* CRC PRESS, 1990
56. G. GARMONG, N.E. PATON, AND A.S. ARGON, ATTAINMENT OF FULL INTERFACIAL CONTACT DURING DIFFUSION BONDING, *METALL. TRANS. A,* VOL 6, JUNE 1975, P 1269-1279
57. A. HILL AND E.R. WALLACH, MODELLING SOLID-STATE DIFFUSION BONDING, *ACTA METALL.,* VOL 37 (NO. 9), 1989, P 2425-2437
58. J.W. CAHN, INTERFACIAL FREE ENERGY AND INTERFACIAL STRESS: THE CASE OF AN INTERNAL INTERFACE IN A SOLID, *ACTA METALL.,* VOL 37 (NO. 3), 1989, P 773-776
59. C.L. CLINE, AN ANALYTICAL AND EXPERIMENTAL STUDY OF DIFFUSION BONDING, *WELD. J.,* NOV 1966, P 481-S TO 489-S
60. D. DERBY AND E.R. WALLACH, THEORETICAL MODEL FOR DIFFUSION BONDING, *MET. SCI.,*

VOL 16, JAN 1982, P 49-56

61. D. DERBY AND E.R. WALLACH, DIFFUSION BONDING: DEVELOPMENT OF THEORETICAL MODEL, *MET. SCI.,* VOL 18, SEPT 1984, P 427-431
62. D. DERBY AND E.R. WALLACH, DIFFUSION BONDS IN IRON AND A LOW-ALLOY STEEL, *J. MATER. SCI.,* VOL 19, 1984, P 3149-3158
63. D. DERBY AND E.R. WALLACH, DIFFUSION BONDS IN COPPER, *J. MATER. SCI.,* VOL 19, 1984, P 3140-3148
64. Z.X. GUO AND N. RIDLEY, MODELLING OF DIFFUSION BONDING OF METALS, *MATER. SCI. TECH.,* VOL 3, NOV 1987, P 945-953

Fundamentals of Metal and Metal-to-Ceramic Adhesion

Ray Dixon and S.P. Chen, Los Alamos National Laboratory

Interface Formation

When considering the compression needed to bond dissimilar materials, the deformation-mechanism map work of Frost and Ashby (Ref 65) is helpful. The maps are in a stress/temperature space in which a single deformation mechanism is dominant. The stress used is a normalized tensile stress (tensile stress/shear modulus), and the temperature is the homologous temperature T/T_m, where T is the test temperature and T_m is the melting point of the material being tested, where both temperatures are expressed in degrees K. From these maps and the corresponding constitutive equations, contour plots of constant strain rate can be plotted. At this point, knowing any two values of strain rate, stress, or temperature, a point on the map will give the third variable and thee dominant deformation mechanism. This information can be a guide to determining the bonding mechanism when considering one of the previously discussed mechanisms that requires elimination of defects at the interface.

Dammer *et al.* (Ref 66) describe a study where type 347 stainless steel was bonded to itself with several different interlayers. The interlayers were electrolytic-grade copper foil, electrolytic tough pitch (ETP) copper sheet, 80Au-20Cu foil, 72Ag-28Cu foil, and several combinations of silver, copper, and gold electroplates. They successfully bonded the specimens at temperatures below 540 °C (1005 °F), but found that the intermediate materials were essential to forming the weld.

Bachin (Ref 67) discusses the interactions of metals and oxides when diffusion welding metals to metals and to ceramics. He discusses the reactions and exchanges at the interface by considering the formation of spinels and oxides. This leads to the conditions under which each is known to be operative and to the possibility of diffusion across the resulting interface. For example, Al + SiO_2 proceeds by exchange:

$$4AL + 3SIO_2 \rightarrow 2AL_2O_3 + 3SI \qquad \textbf{(EQ 7)}$$

When thin silicon dioxide is joined through a thin intermediate interlayer of aluminum, new phases precipitate at the dividing boundaries and the solid solution is decomposed when the limit of solubility is reached for the reaction products. This behavior results in high internal stresses and resultant microcracks. Caution is therefore required when welding these systems. A thorough review of the thermodynamics involved for the systems being considered is essential.

Munir (Ref 68) developed equations that predict the amount of time required to dissolve oxides into their host metals for systems that have unstable oxides with respect to dissolution. He considered niobium, zirconium, titanium, copper, tantalum, cobalt, nickel, and iron and found reasonable agreement with the theory.

Wang and Vehoff (Ref 69) discuss the segregation of atoms to interfaces. Their work deals with grain boundaries as interfaces, and they conclude that the cohesive energy of the interface at a fixed chemical potential is always less than the cohesive energy at a fixed concentration. The concept of segregation to the interface is probably important for bimaterial interfaces because of the different chemical potentials, diffusion rates, and, thus, concentrations that can exist. Because these parameters are all temperature dependent, the bonding process and the service temperature will affect the interface segregation and, thus, its performance. This work provides a way to consider these effects.

References cited in this section

65. H.J. FROST AND M.F. ASHBY, *DEFORMATION MECHANISM MAPS,* PERGAMON PRESS, 1982
66. P.A. DAMMER, R.E. MONROE, AND D.C. MARTIN, FURTHER STUDIES OF DIFFUSION BONDING BELOW 1000 °F, *WELD. J.,* MARCH 1969, P 116-S TO 124-S
67. V.A. BACHIN, INTERACTION BETWEEN METALS AND THE OXIDES OF CERAMIC MATERIALS DURING DIFFUSION WELDING*, AVT. SVARKA,* NO. 10, 1982, P 15-17
68. Z.A. MUNIR, A THEORETICAL ANALYSIS OF THE STABILITY OF SURFACE OXIDES DURING DIFFUSION WELDING OF METALS, *WELD. J.,* DEC 1983, P 333-S TO 336-S
69. J.-S. WANG AND H. VEHOFF, THE EFFECT OF THE MOBILITY OF SEGREGATED ATOMS ON INTERFACIAL EMBRITTLEMENT, *SCR. METALL.,* VOL 25, 1991, P 1339-1344

Fundamentals of Metal and Metal-to-Ceramic Adhesion

Ray Dixon and S.P. Chen, Los Alamos National Laboratory

Strength of Interfaces

One of the most important and probably the most frequent questions: What is the strength of the interfacial bond? This question has often been answered by tensile testing. The difficulty is that the failure is often not at the interface and that it must be forced there. Additionally, there is concern that the strain is not one-dimensional across the bond and that the bond experiences triaxial strain in service. For these reasons, there has been considerable interests in determining fracture toughness and in detecting the failure of interfaces.

Several reviews on the importance and extent of interfacial science were recently presented in one technical journal (Ref 70). Three reviews are noteworthy because they discuss the mechanical properties of interfaces (Ref 71, 72, 73). The efforts of Hutchinson and Suo (Ref 71) bring together much of the recent work on elastic fracture phenomena for bimetallic systems and builds on it to support work being done to understand mixed-mode failure in systems with interfaces. Shih (Ref 72) summarizes progress on interfacial mechanics and points out that fracture toughness, at various mode mixities, is needed to characterize fracture resistance of a weak plane in a composite. An application of this analysis has been demonstrated on the alumina/niobium metal-ceramic system by Stout *et al.* (Ref 74). Comninou (Ref 73) reviews the interface crack problem from the standpoint of open and closed crack tips and describes preliminary experimental results on fatigue and fracture of interface cracks in aluminum bonded by epoxy.

Interface Strength Models. There has been considerable progress in analyzing the fracture behavior of interfaces with various configurations. This work has led to the development of specimen geometries for systems with differing physical and mechanical properties. The analysis work is reviewed in this section and the experimental work is described in the section "Interface Strength Measurements" of this article.

Suo and Hutchinson (Ref 75) have provided a complete solution to the semi-infinite crack lying along the interface of two infinite elastic layers under edge loads. This solution allows the determination of material properties for complex geometries, as well as the effect of welding two dissimilar materials. Jensen (Ref 76) demonstrates that shielding the full effect of mode II and III loading is important for interface cracks in brittle materials and that a proper interface fracture criteria needs to be modeled. He discusses a method for comparing different phenomenological fracture criteria and sets up micromechanical models describing shielding for interface geometries with kinks.

The models of He and Hutchinson (Ref 77) explain conditions for interfacial crack propagation (kinking) into material on either side of the interface. Their models suggest that a crack propagates along an interface as long as the compliant material is tough and that the stiff material is at least as tough as the interface. They also show that the energy release rate is increased if the crack kinks into the more-compliant material and is diminished if it kinks into the stiff material.

Qu and Li (Ref 78) also analyzed cracks and dislocations at the interface between two anisotropic elastic half spaces. They obtained an explicit and real solution to the dislocation problem in the Stroh (Ref 79) formalism and showed that a planar interface between dissimilar anisotropic solids is completely characterized by no more than nine independent parameters. The work by Li and Ting (Ref 80, 81) is also relevant.

Evans *et al.* (Ref 82, 83) concluded that fracture energy is not unique and is usually greater than the thermodynamic work of adhesion. The nonuniqueness is related to the mixity experienced by interface cracks. As a result, the fracture energy is influenced by the choice of test specimen. This result implies that when considering fracture energy for specific material combinations, the test geometry must be known, reported, and maintained. This also means that it is now possible to systematically study interfacial mechanical behavior and produce meaningful data on interface fracture toughness.

Fischmeister *et al.* (Ref 84) modeled simple crack propagation, dislocation emission from crack tips, and twinning using atomistic calculations (molecular dynamics) embedded in a finite-element continuum. They also studied the behavior of cracks propagating in tungsten and iron on different planes and in different directions within these planes. Their results confirm that {100} should be the easiest cleavage plane in these body-centered cubic (bcc) tungsten and iron crystals. They further studied the propagation direction restrictions for cracks on particular planes. The modeling work was supported by experiments to verify the crack propagation behavior in WC-Co alloy, high-speed steel consolidated with nickel particles, and nickel coated on high-speed steel.

Chen *et al.* (Ref 1, 9, 20, 21) studied crack growth in nickel, Ni_3Al, Ni_3Al plus boron, Ni_3Al plus sulfur, and iridium grain boundaries using local volume potentials. They found that cohesive strength is important in determining grain boundary fracture and can be strongly affected by impurities and stoichiometry.

Another approach to modeling fracture is given by Needleman (Ref 85), who used a cohesive zone-type interface model. His model takes full account of finite geometry changes and is used to study decohesion of a viscoplastic block from a rigid substrate. He found that the normal traction versus separation relation had an exponential form consistent with the universal binding energy relationship of Rose *et al.* (Ref 86) and of Ferrante and Smith (Ref 87).

Ting (Ref 88) showed that oscillatory displacement at the interfacial crack surface in a bimaterial under a two-dimensional deformation depends only on the material properties and is independent of specimen orientation. The exception is that bimaterials always produce oscillation and that there is an exception for one choice of relative orientation in one material combination. He also shows that for two different orthotropic materials, the oscillation depends on whether or not the Voigt elastic constants combinations, $\sqrt{C_{11}+C_{22}+C_{12}}$, for the two materials are different.

Cook (Ref 89) explains crack propagation threshold in brittle materials through the work done by the applied loading and that which is needed to create new surfaces during crack propagation. He showed that the mechanical energy release rate was a measure of the equilibrium surface energy of the material and the chemical environment. The latter affects the surface tension of both the fresh and aged crack.

Reynolds *et al.* (Ref 90) investigated the difference between the concepts of cracks with zero normal and shear traction over the entire face of the crack and cracks with normal and shear traction specified at zero on the crack surface, except near the tip, where zero shear fraction and continuity of displacements normal to the crack were required. Of these two approaches, the first leads to oscillatory stress singularities and interpenetration of the material near the crack tip, whereas the second shows that a small region of contact exists at the tip, regardless, of the loading. The two methods require different treatments in terms of stress intensity, but predict the same strain-energy release rate. The work was done using the finite-element method (FEM) and, thus, also served to verify the formulation of each problem by FEM analysis. The results showed that the two formulations could be modeled by the absence or presence of interface elements that prevent interpenetration of the crack surfaces. Additionally, fracture parameters were calculated and compared well with previous analytic solutions.

Interface Strength Measurements. Based on the theoretical analysis described above, as well as their own models, the authors identified below have carried out excellent experimental work in measuring the mechanical properties of interfaces.

Charalambides *et al.* (Ref 91) developed a test procedure to measure interface fracture resistance under mixed-mode conditions. Strain-energy release rates and stress-intensity factors can be obtained using this test. O'Dowd *et al.* (Ref 92) developed FEM model solutions and tested several geometries to determine interfacial fracture toughness as a function of mixity. They were able to explain near-tip contact and variation of mixity with distance. They also developed a procedure for determining the effect of residual stress on the stress-intensity factor. Brady *et al.* (Ref 93) developed a buckled plate test to determine the transverse fracture toughness of unidirectional continuous-fiber composites. Their test was found to be independent of sample geometry and testing conditions. Mecholsky and Baker (Ref 94) developed a double cantilever beam specimen with a chevron notch to measure fracture toughness of ceramic-metal interfaces. The specimen is easy to make and introduces its own crack, and the test is easy to perform. Zhu and Byrne (Ref 95) looked at the effect of interfacial defects by bonding and subsequent tensile, fatigue, and photomicroscopy tests. They measured void ratios at the bond line and found the defects smaller than those detectable by current nondestructive test methods (~0.2 to 0.5 mm^2 by 1 to 3 μm, or 3×10^{-4} to 8×10^{-4} $in.^2$, or 40 to 120 μin., thick) could still significantly affect bond strength. They concluded that fatigue testing is a better test method than tensile testing because it is sensitive to a void ratio of approximately 27%, whereas tensile tests were not.

Thouless (Ref 96) used epoxy-bonded glass slides to investigate interfacial failure under mixed-mode loading. He found that when the scale is sufficiently large, the sample could be considered homogeneous with a weak planar interface. This result produced a correlation between apparent fracture resistance and the degree of mixity. At a very detailed scale, the lack of homogeneity required that the crack-tip stress be determined by the position of the crack within the interface (epoxy). This condition leads to a change in failure mechanisms, whereby under pure mode-I conditions, the crack propagates entirely within the interfacial layer, whereas mode-II effects drive the crack toward the glass-epoxy interface.

Nickel specimens with known surface roughnesses were diffusion welded under controlled conditions of temperature and pressure (Ref 97). The welds were interrupted at several different time intervals and the resulting interface was examined. The defect sizes were successfully related to surface tension and diffusion. It was also found that temperature increases produced a greater effect on reducing interfacial defects than did pressure. Meriin and Sliozberg (Ref 98) make the point that sound welds (bonds) cannot be obtained unless the voids are eliminated at the interface, which requires high temperatures and pressures. It further requires that mechanical tests be used as the principal detection method.

This article has shown that there has been and continues to be active interest in interfacial behavior. The science is sufficiently complex that numerous approaches are useful to evaluate interfacial behavior, depending on the specific goal. However, it is obvious that the seemingly unrelated approaches all seek to understand the interface. It is therefore incumbent upon scientists in this field to be familiar with all developments concerning adhesion so that they can contribute to its understanding and future advancement.

References cited in this section

1. S.P. CHEN, D.J. SROLOVITZ, AND A.F. VOTER, COMPUTER SIMULATION ON SURFACES AND [001] TILT GRAIN BOUNDARIES IN NI, AL, AND NI_3AL, *J. MATER. RES.,* VOL 4, 1989, P 62-77
9. S.P. CHEN, STUDIES OF IRIDIUM SURFACES AND GRAIN BOUNDARIES, *PHIL. MAG. A,* VOL 65, 1992, TO BE PUBLISHED
20. S.P. CHEN, A.F. VOTER, R.C. ALBERS, A.M. BORING, AND P.J. HAY, INVESTIGATION OF BORON EFFECT ON NI AND NI_3AL GRAIN BOUNDARIES, *J. MATER. RES.,* VOL 5, 1990, P 955-970
21. S.P. CHEN, A.F. VOTER, R.C. ALBERS, A.M. BORING, AND P.J. HAY, THEORETICAL STUDIES OF GRAIN BOUNDARIES IN NI_3AL WITH BORON OR SULFUR, *SCR. METALL.,* VOL 23, 1989, P 217-222
70. T. TSAKALAKOS, ED., *MATER. SCI. ENG. B,* VOL 6, 1990, P 69-210
71. J.W. HUTCHINSON AND Z. SUO, MIXED MODE CRACKING IN LAYERED MATERIALS, *ADVANCES IN APPLIED MECHANICS,* VOL 28, J.W. HUTCHINSON AND T.Y. WU, ED., ACADEMIC PRESS, 1990
72. C.F. SHIH, CRACKS ON BIMATERIAL INTERFACES: ELASTICITY AND PLASTICITY ASPECTS, *MATER. SCI. ENG. A,* VOL 143, 1991, P 77-90
73. M. COMNINOU, AN OVERVIEW OF INTERFACE CRACKS, *ENG. FRACT. MECH.,* VOL 37 (NO. 1), 1990, P 197-208
74. M.G. STOUT, N.P. O'DOWD, AND C.F. SHIH, "INTERFACIAL FRACTURE TOUGHNESS OF ALUMINA/NIOBIUM SYSTEMS," SYMPOSIUM ON EXPERIMENTS IN MICROMECHANICS OF

FAILURE RESISTANT MATERIALS, ASME WINTER ANNUAL MEETING (ATLANTA, GA), DEC 1991

75. Z. SUO AND J.W. HUTCHINSON, "INTERFACE CRACK BETWEEN TWO ELASTIC LAYERS," REPORT MECH-118, HARVARD UNIVERSITY, DIVISION OF APPLIED SCIENCES, MARCH 1988
76. H.M. JENSEN, MIXED MODE INTERFACE FRACTURE CRITERIA, *ACTA METALL. MATER.,* VOL 38 (NO. 12), 1990, P 2637-2644
77. M.-Y. HE AND J.W. HUTCHINSON, "KINKING OF A CRACK OUT OF AN INTERFACE," REPORT MECH-113, HARVARD UNIVERSITY, DIVISION OF APPLIED SCIENCES, FEB 1988
78. J. QU AND Q. LI, INTERFACIAL DISLOCATION AND ITS APPLICATIONS TO INTERFACE CRACKS IN ANISOTROPIC BIMATERIALS, *J. ELASTICITY,* VOL 26, 1991, P 169-195
79. A.N. STROH, DISLOCATIONS AND CRACKS IN ANISOTROPIC ELASTICITY, *PHILOS. MAG.,* VOL 7, 1958, P 625-646
80. Q.Q. LI AND T.C.T. TING, LINE INCLUSION IN ANISOTROPIC ELASTIC SOLIDS, *J. APPL. MECH.,* VOL 56, 1989, P 556-563
81. T.C.T. TING, EXPLICIT SOLUTION AND INVARIANCE OF THE SINGULARITIES AT AN INTERFACE CRACK IN ANISOTROPIC COMPOSITES, *INT. J. SOLIDS STRUCT.,* VOL 22, 1986, P 965-983
82. A.G. EVANS, M. RUHLE, B.J. DALGLEISH, AND P.G. CHARALAMBIDES, THE FRACTURE ENERGY OF BIMATERIAL INTERFACES, *MATER. SCI. ENG.,* VOL A 126, 1990, P 53-64
83. A.G. EVANS, M. RUHLE, B.J. DALGLEISH, AND P.G. CHARALAMBIDES, THE FRACTURE ENERGY OF BIMATERIAL INTERFACES, *METALL. TRANS. A,* VOL 21, SEPT 1990, P 2419-2429
84. H.F. FISCHMEISTER, H.E. EXNER, M.-H. POECH, S. KOHLHOFF, P. GUMBSCH, S. SCHMAUDER, L.S. SIGL, AND R. SPEIGLER, MODELING FRACTURE PROCESSES IN METALS AND COMPOSITE MATERIALS, *Z. METALLKDE,* BD 80, 1989, P 839-846
85. A. NEEDLEMAN, AN ANALYSIS OF TENSILE DECOHESION ALONG AN INTERFACE, *J. MECH. PHYS. SOLIDS,* VOL 38 (NO. 3), 1990, P 289-324
86. J.H. ROSE, J. FERRANTE, AND J.R. SMITH, UNIVERSAL BINDING ENERGY CURVES FOR METALS AND BIMETALLIC INTERFACES, *PHYS. REV. LETT.,* VOL 47 (NO. 9), AUG 1981, P 675-678
87. J. FERRANTE AND J.R. SMITH, THEORY OF THE BIMETALLIC INTERFACE, *PHYS. REV. B,* VOL 31 (NO. 6), MARCH 1985, P 3427-3434
88. T.C.T. TING, INTERFACE CRACKS IN ANISOTROPIC BIMATERIALS, *J. MECH. PHYS. SOLIDS,* VOL 38 (NO. 4), 1990, P 505-513
89. R.F. COOK, CRACK PROPAGATION THRESHOLDS: A MEASURE OF SURFACE ENERGY, *J. MATER. RES.,* VOL 1 (NO. 6), NOV/DEC 1986, P 852-860
90. R.R. REYNOLDS, K. KOKINI, AND G. CHEN, THE MECHANICS OF THE INTERFACE CRACK USING THE FINITE ELEMENT METHOD, *J. ENG. MATER. TECHNOL.,* VOL 112, JAN 1990, P 38-43
91. P.G. CHARALAMBIDES, J. LUND, A.G. EVANS, AND R.M. MCMEEKING, "A TEST FOR DETERMINING THE FRACTURE RESISTANCE OF BIMATERIAL INTERFACES," REPORT M87-4, MATERIALS DEPARTMENT, COLLEGE OF ENGINEERING, UC SANTA BARBARA
92. N.P. O'DOWD, C.F. SHIH, AND M.G. STOUT, TEST GEOMETRIES FOR MEASURING INTERFACIAL FRACTURE TOUGHNESS, *INT. J. SOLIDS AND STRUCT.,* VOL 29, 1992, P 571-589
93. R.L. BRADY, R.S. PORTER, AND J.A. DONOVAN, A BUCKLED PLATE TEST TO MEASURE INTERFACIAL TOUGHNESS IN COMPOSITES, *J. MATER. SCI.,* VOL 24, 1989, P 4138-4143
94. J.J. MECHOLSKY AND L.M. BAKER, A CHEVRON-NOTCHED SPECIMEN FOR FRACTURE TOUGHNESS MEASUREMENTS OF CERAMIC-METAL INTERFACES, *CHEVRON-NOTCHED SPECIMENS: TESTING AND STRESS ANALYSIS,* STP 855, J.H. UNDERWOOD, S.W. FREIMAN, AND F.I. BARATTA, ED., ASTM, 1984, P 324-336
95. X. ZHU AND J.G. BYRNE, EFFECT OF INTERFACE DEFECTS ON THE MECHANICAL PROPERTIES OF AUTOGENOUS DIFFUSION BONDED JOINTS, *J. MATER. ENG.,* VOL 13, 1991, P

207-211

96. M.D. THOULESS, FRACTURE OF A MODEL INTERFACE UNDER MIXED-MODE LOADING, *ACTA METALL. MATER.,* VOL 38 (NO. 6), 1990, P 1135-1140
97. V.S. GOSTOMEL'SKII, E.S. KARAKOZOV, AND A.P. TERNOVSKII, THE IMPORTANCE OF DIFFUSION AND SURFACE TENSION IN THE FORMATION OF CONTACT DURING DIFFUSION WELDING, *AVT. SVARKA,* NO. 4, 1980, P 28-31
98. A.V. MERIIN AND S.K. SLIOZBERG, SOME CHARACTERISTICS OF JOINTS IN THE SOLID-PHASE WELDING OF METALS, *SVAR. PROIZ.,* NO. 5, 1973, P 22-23

Fundamentals of Metal and Metal-to-Ceramic Adhesion

Ray Dixon and S.P. Chen, Los Alamos National Laboratory

References

1. S.P. CHEN, D.J. SROLOVITZ, AND A.F. VOTER, COMPUTER SIMULATION ON SURFACES AND [001] TILT GRAIN BOUNDARIES IN NI, AL, AND NI$_3$AL, *J. MATER. RES.,* VOL 4, 1989, P 62-77
2. S.P. CHEN AND A.F. VOTER, RECONSTRUCTION OF THE (310), (210), AND (110) SURFACES IN FCC METALS, *SURF. SCI. LETT.,* VOL 244, 1991, P L107-L112
3. S.P. CHEN, THEORETICAL STUDIES OF METALLIC INTERFACES, *MATER. SCI. ENG. B,* VOL 6, 1990, P 113-121
4. M.L. JOKL, V. VITEK, AND C.J. MCMAHON, JR., A MICROSCOPIC THEORY OF BRITTLE FRACTURE IN DEFORMABLE SOLIDS: A RELATION BETWEEN IDEAL WORK TO FRACTURE AND PLASTIC WORK, *ACTA METALL.,* VOL 28, 1980, P 1479-1488
5. M.L. JOKL, V. VITEK, AND C.J. MCMAHON, JR., ON THE MICROMECHANICS OF BRITTLE FRACTURE: EXISTING VS INJECTED CRACKS, *ACTA METALL.,* VOL 37, 1989, P 87-97
6. J. HACK, S.P. CHEN, AND D. SROLOVITZ, A KINETIC CRITERION FOR QUASI-BRITTLE FRACTURE, *ACTA METALL.,* VOL 37, 1989, P 1957-1970
7. A. KELLY AND N.H. MACMILLAN, *STRONG SOLIDS,* 3RD ED., OXFORD SCIENCE PUBLICATIONS, 1986
8. V. VITEK, MICROMECHANISMS OF INTERGRANULAR BRITTLE FRACTURE IN INTERMETALLIC COMPOUND, *J. PHYS. III,* VOL 1, 1991, P 1085-1097
9. S.P. CHEN, STUDIES OF IRIDIUM SURFACES AND GRAIN BOUNDARIES, *PHIL. MAG. A,* VOL 65, 1992, TO BE PUBLISHED
10. L.E. MURR, *INTERFACIAL PHENOMENA IN METALS AND ALLOYS,* ADDISON-WESLEY, 1975
11. J.H. ROSE, J.R. SMITH, F. GUINEA, AND J. FERRANTE, UNIVERSAL FEATURES OF THE EQUATION OF STATE OF METALS, *PHYS. REV. B,* VOL 29, 1984, P 2963-2969
12. T.B. MASSALSKI, ED., *BINARY ALLOYS PHASE DIAGRAMS,* ASM, 1986, P 2176-2181
13. D.G. PETTIFOR, ELECTRON THEORY OF METALS, *PHYSICAL METALLURGY,* R.W. CAHN AND P. HAASEN, ED., NORTH-HOLLAND, AMSTERDAM, 1983, P 73-152
14. J.S. FAULKNER, THE MODERN THEORY OF ALLOYS, *PROG. MATER. SCI.,* VOL 27, 1982, P 1-187
15. J. FRIEDEL, TRANSITION METALS, ELECTRONIC STRUCTURE OF THE D-BAND, ITS ROLE IN THE CRYSTALLINE AND MAGNETIC STRUCTURES, *THE PHYSICS OF METALS,* J.M. ZIMAN, ED., CAMBRIDGE UNIVERSITY PRESS, 1969, P 340-408
16. D.G. PETTIFOR, NEW MANY-BODY POTENTIAL FOR THE BOND ORDER, *PHYS. REV. LETT.,* VOL 63, 1989, P 2480-2483
17. J.K. NORSKOV AND N.D. LANG, EFFECTIVE-MEDIUM THEORY OF CHEMICAL BINDING: APPLICATION TO CHEMISORPTION, *PHYS. REV. B,* VOL 21, 1980, P 2131-2136

18. M.J. STOTT AND E. ZAREMBA, QUASIATOMS: AN APPROACH TO ATOMS IN NONUNIFORM ELECTRONIC SYSTEMS, *PHYS. REV. B,* VOL 22, 1980, P 1564-1583
19. M.S. DAW AND M.I. BASKES, EMBEDDED-ATOM METHOD: DERIVATION AND APPLICATION TO IMPURITIES, SURFACES, AND OTHER DEFECTS IN METALS, *PHYS. REV. B,* VOL 29, 1984, P 6443-6453
20. S.P. CHEN, A.F. VOTER, R.C. ALBERS, A.M. BORING, AND P.J. HAY, INVESTIGATION OF BORON EFFECT ON NI AND NI$_3$AL GRAIN BOUNDARIES, *J. MATER. RES.,* VOL 5, 1990, P 955-970
21. S.P. CHEN, A.F. VOTER, R.C. ALBERS, A.M. BORING, AND P.J. HAY, THEORETICAL STUDIES OF GRAIN BOUNDARIES IN NI$_3$AL WITH BORON OR SULFUR, *SCR. METALL.,* VOL 23, 1989, P 217-222
22. A.F. VOTER AND S.P. CHEN, ACCURATE INTERATOMIC POTENTIALS FOR NI, AL, AND NI$_3$AL, *MATER. RES. SOC. PROC.,* VOL 82, 1987, P 175-180
23. S.P. CHEN, ANOMALOUS RELAXATIONS OF (0001) AND $(10\bar{1}0)$ SURFACES IN HCP METALS, *SURF. SCI. LETT.,* VOL 264, 1992, P L162-L168
24. R.J. HARRISON, A.F. VOTER, AND S.P. CHEN, *ATOMISTIC MODELING OF MATERIALS: BEYOND PAIR POTENTIALS,* V. VITEK AND D.J. SROLOVITZ, ED., PLENUM PRESS, 1989, P 219-222
25. R.J. HARRISON, F. SPAEPEN, A.F. VOTER, AND S P. CHEN, STRUCTURE OF GRAIN BOUNDARIES IN IRON, *PROC. 34TH SAGAMORE ARMY MATERIALS RESEARCH CONFERENCE* (LAKE GEORGE, NY), 1990, P 651-675 (FOR FE)
26. S.P. CHEN, LOCAL VOLUME POTENTIALS FOR BCC METALS, TO BE PUBLISHED (FOR FE, NB, CR, W, AND MO)
27. S.P. CHEN, A.M. BORING, R.C. ALBER, AND P.J. HAY, THEORETICAL STUDIES OF NI$_3$AL AND NIAL WITH IMPURITIES, *HIGH TEMPERATURE ORDERED INTERMETALLIC ALLOYS III,* C.C. KOCH, C.T. LIU, N.S. STOLOFF, AND A.I. TAUB, ED., *PROC. MATER. RES. SOC.,* VOL 133, 1989, P 149-154
28. F.R. DEBOER, R. BOOM, W.C.M. MATTENS, A.R. MIEDEMA, AND A.K. NIESSEN, *COHESION IN METALS: TRANSITION METAL ALLOYS,* NORTH-HOLLAND, 1989
29. J.E. HACK, D.J. STROLOVITZ, AND S.P. CHEN, A MODEL FOR THE FRACTURE BEHAVIOR OF POLYCRYSTALLINE NI$_3$AL, *SCR. METALL.,* VOL 20, 1986, P 1699-1704
30. E.D. HONDROS AND M.P. SEAH, INTERFACIAL AND SURFACE MICROCHEMISTRY, *PHYSICAL METALLURGY,* 3RD ED., R.W. CAHN AND P. HAASEN, ED., ELSEVIER SCIENCE PUBLISHER, 1983, P 855-931
31. M.E. SIKORSKI, THE ADHESION OF METALS AND FACTORS THAT INFLUENCE IT, *WEAR,* VOL 7, 1964, P 144-162
32. J. GERKEMA AND A.R. MIEDEMA, ADHESION BETWEEN SOLID METALS: OBSERVATIONS OF INTERFACIAL SEGREGATION EFFECTS IN METAL FILM LUBRICATION EXPERIMENTS, *SURF. SCI.,* VOL 124, 1983, P 351-371
33. M.D. PASHLEY AND D. TABOR, ADHESION AND DEFORMATION PROPERTIES OF CLEAN AND CHARACTERIZED METAL MICRO-CONTACTS, *VACUUM,* VOL 10-12, 1981, P 619-623
34. M.D. PASHLEY AND J.B. PETHICA, THE ROLE OF SURFACE FORCES IN METAL-METAL CONTACTS, *J. VAC. SCI. TECHNOL. A,* VOL 3 (NO. 3), MAY/JUNE 1985, P 757-761
35. K.I. JOHNSON AND D.V. KELLER, JR., EFFECT OF CONTAMINATION ON THE ADHESION OF METALLIC COUPLES IN ULTRA-HIGH VACUUM, *J. APPL. PHYS.,* VOL 38 (NO. 4), JAN 1967, P 1896-1904
36. H. CONRAD AND L. RICE, THE COHESION OF PREVIOUSLY FRACTURED FCC METALS IN ULTRA-HIGH VACUUM, *METALL. TRANS.,* VOL 1, NOV 1970, P 3019-3029
37. T. KISHI, M. ENOKI, AND H. TSUDA, INTERFACE AND STRENGTH IN CERAMIC MATRIX COMPOSITES, *MATER. SCI. ENG.,* VOL A143, 1991, P 103-110
38. O. BUCK, R.B. THOMPSON, D.K. REHBEIN, D.D. PALMER, AND L.J.H. BRASCHE, CONTACTING

SURFACES: A PROBLEM IN FATIGUE AND DIFFUSION BONDING, *METALL. TRANS. A,* VOL 20, APRIL 1989, P 627-635

39. H.A. NIED, "DETERMINATION OF THE STRUCTURAL INTEGRITY OF METALLURGICALLY BONDED PLATES BY HIGH ENERGY IMPULSE TEST," *SIXTH INT. CONF. ON EXP. STRESS ANALYSIS* (MUNICH), SEPT 1978
40. S. ITOU, AN EXPERIMENT ON THE BEHAVIOUR AROUND THE TIPS OF AN INTERFACE CRACK, *ENG. FRACT. MECH.,* VOL 37 (NO. 1), 1990, P 145-150
41. N. MARTENSSON, A. STENBORG, O. BJORNEHOLM, A. NILSSON, AND J.N. ANDERSON, QUANTITATIVE STUDIES OF METAL-METAL ADHESION AND INTERFACE SEGREGATION ENERGIES USING PHOTOELECTRON SPECTROSCOPY, *PHYS. REV. LETT.,* VOL 60 (NO. 17), APRIL 1988, P 1731-1734
42. J.N. ANDERSON, O. BJORNEHOLM, A. STENBORG, A. NILSSON, C. WIGREN, AND N. MARTENSSON, MEASUREMENT OF METAL-METAL ADHESION AND INTERFACE SEGREGATION ENERGIES BY CORE-LEVEL PHOTOELECTRON SPECTROSCOPY. AL AND SI ON MO(110), *J. PHYS., CONDENS. MATTER,* VOL 1, 1989, P 7309-7313
43. R.W. CARPENTER, HIGH RESOLUTION INTERFACE ANALYSIS, *MATER. SCI. ENG.,* VOL A107, 1989, P 207-216
44. W.W. GERBERICH, D.L. DAVIDSON, AND M. KACZOROWSKI, EXPERIMENTAL AND THEORETICAL STRAIN DISTRIBUTIONS FOR STATIONARY AND GROWING CRACKS, *J. MECH. PHYS. SOLIDS,* VOL 38 (NO. 1), 1990
45. A. JOHNER AND P. SCHAAF, CALCULATION OF THE REFLECTION COEFFICIENTS OF INTERFACES: A SCATTERING APPROACH, *PHYS. REV. B,* VOL 42 (NO. 9), SEPT 1990, P 5516-5526
46. B.R. LANNING, T. FURTAK, AND G.R. EDWARDS, A METHOD FOR PREDICTION OF METAL-CERAMIC INTERFACIAL BOND ENERGIES, SUBMITTED FOR PUBLICATION IN *J. MATER. RES.*
47. R.J. COLTON, C.R.K. MARRIAN, AND J.A. STROSCIO, ED., *PROC. FIFTH INT. CONF. SCANNING TUNNELING MICROSCOPY/SPECTROSCOPY (SAM '90)* AND *THE FIRST INT. CONF. NANOMETER SCALE SCIENCE AND TECHNOLOGY (NANO I)* (BALTIMORE, MD), JUNE 1990, AND *J. VAC. SCI. TECH. B,* SECOND SERIES, VOL 9 (NO. 2), PART II, MARCH/APRIL 1991
48. J.M. GERKEN AND W.A. OWCZARSKI, "A REVIEW OF DIFFUSION WELDING," WELDING RESEARCH COUNCIL BULLETIN NO. 109, OCT 1965
49. J.D. WILLIAMS AND B. CROSSLAND, PART 1 COLD PRESSURE WELDING, *CME,* MARCH 1976, P 65-67
50. J.D. WILLIAMS AND B. CROSSLAND, PART 2 DIFFUSION BONDING, *CME,* APRIL 1976, P 73-75
51. J.R. SMITH, PANEL REPORT ON INTERFACIAL BONDING AND ADHESION, *MATER. SCI. ENG.,* VOL 83, 1986
52. H.A. MOHAMED AND J. WASHBURN, MECHANISM OF SOLID STATE PRESSURE WELDING, *WELD. J. RES. SUPP.,* SEPT 1975, P 302S-310S
53. I. KAWAKATSU AND S. KITAYAMA, STUDY ON DIFFUSION BONDING OF METALS, *TRANS. JIM,* VOL 18, 1977, P 455-465
54. N. BAY, MECHANISMS PRODUCING METALLIC BONDS IN COLD WELDING, *WELD. J. RES. SUPP.,* MAY 1983, P 137S-142S
55. P.A. DOWBEN AND A. MILLER, ED., *SURFACE SEGREGATION PHENOMENA,* CRC PRESS, 1990
56. G. GARMONG, N.E. PATON, AND A.S. ARGON, ATTAINMENT OF FULL INTERFACIAL CONTACT DURING DIFFUSION BONDING, *METALL. TRANS. A,* VOL 6, JUNE 1975, P 1269-1279
57. A. HILL AND E.R. WALLACH, MODELLING SOLID-STATE DIFFUSION BONDING, *ACTA METALL.,* VOL 37 (NO. 9), 1989, P 2425-2437
58. J.W. CAHN, INTERFACIAL FREE ENERGY AND INTERFACIAL STRESS: THE CASE OF AN INTERNAL INTERFACE IN A SOLID, *ACTA METALL.,* VOL 37 (NO. 3), 1989, P 773-776
59. C.L. CLINE, AN ANALYTICAL AND EXPERIMENTAL STUDY OF DIFFUSION BONDING, *WELD.*

J., NOV 1966, P 481-S TO 489-S

60. D. DERBY AND E.R. WALLACH, THEORETICAL MODEL FOR DIFFUSION BONDING, *MET. SCI.,* VOL 16, JAN 1982, P 49-56
61. D. DERBY AND E.R. WALLACH, DIFFUSION BONDING: DEVELOPMENT OF THEORETICAL MODEL, *MET. SCI.,* VOL 18, SEPT 1984, P 427-431
62. D. DERBY AND E.R. WALLACH, DIFFUSION BONDS IN IRON AND A LOW-ALLOY STEEL, *J. MATER. SCI.,* VOL 19, 1984, P 3149-3158
63. D. DERBY AND E.R. WALLACH, DIFFUSION BONDS IN COPPER, *J. MATER. SCI.,* VOL 19, 1984, P 3140-3148
64. Z.X. GUO AND N. RIDLEY, MODELLING OF DIFFUSION BONDING OF METALS, *MATER. SCI. TECH.,* VOL 3, NOV 1987, P 945-953
65. H.J. FROST AND M.F. ASHBY, *DEFORMATION MECHANISM MAPS,* PERGAMON PRESS, 1982
66. P.A. DAMMER, R.E. MONROE, AND D.C. MARTIN, FURTHER STUDIES OF DIFFUSION BONDING BELOW 1000 °F, *WELD. J.,* MARCH 1969, P 116-S TO 124-S
67. V.A. BACHIN, INTERACTION BETWEEN METALS AND THE OXIDES OF CERAMIC MATERIALS DURING DIFFUSION WELDING, *AVT. SVARKA,* NO. 10, 1982, P 15-17
68. Z.A. MUNIR, A THEORETICAL ANALYSIS OF THE STABILITY OF SURFACE OXIDES DURING DIFFUSION WELDING OF METALS, *WELD. J.,* DEC 1983, P 333-S TO 336-S
69. J.-S. WANG AND H. VEHOFF, THE EFFECT OF THE MOBILITY OF SEGREGATED ATOMS ON INTERFACIAL EMBRITTLEMENT, *SCR. METALL.,* VOL 25, 1991, P 1339-1344
70. T. TSAKALAKOS, ED., *MATER. SCI. ENG. B,* VOL 6, 1990, P 69-210
71. J.W. HUTCHINSON AND Z. SUO, MIXED MODE CRACKING IN LAYERED MATERIALS, *ADVANCES IN APPLIED MECHANICS,* VOL 28, J.W. HUTCHINSON AND T.Y. WU, ED., ACADEMIC PRESS, 1990
72. C.F. SHIH, CRACKS ON BIMATERIAL INTERFACES: ELASTICITY AND PLASTICITY ASPECTS, *MATER. SCI. ENG. A,* VOL 143, 1991, P 77-90
73. M. COMNINOU, AN OVERVIEW OF INTERFACE CRACKS, *ENG. FRACT. MECH.,* VOL 37 (NO. 1), 1990, P 197-208
74. M.G. STOUT, N.P. O'DOWD, AND C.F. SHIH, "INTERFACIAL FRACTURE TOUGHNESS OF ALUMINA/NIOBIUM SYSTEMS," SYMPOSIUM ON EXPERIMENTS IN MICROMECHANICS OF FAILURE RESISTANT MATERIALS, ASME WINTER ANNUAL MEETING (ATLANTA, GA), DEC 1991
75. Z. SUO AND J.W. HUTCHINSON, "INTERFACE CRACK BETWEEN TWO ELASTIC LAYERS," REPORT MECH-118, HARVARD UNIVERSITY, DIVISION OF APPLIED SCIENCES, MARCH 1988
76. H.M. JENSEN, MIXED MODE INTERFACE FRACTURE CRITERIA, *ACTA METALL. MATER.,* VOL 38 (NO. 12), 1990, P 2637-2644
77. M.-Y. HE AND J.W. HUTCHINSON, "KINKING OF A CRACK OUT OF AN INTERFACE," REPORT MECH-113, HARVARD UNIVERSITY, DIVISION OF APPLIED SCIENCES, FEB 1988
78. J. QU AND Q. LI, INTERFACIAL DISLOCATION AND ITS APPLICATIONS TO INTERFACE CRACKS IN ANISOTROPIC BIMATERIALS, *J. ELASTICITY,* VOL 26, 1991, P 169-195
79. A.N. STROH, DISLOCATIONS AND CRACKS IN ANISOTROPIC ELASTICITY, *PHILOS. MAG.,* VOL 7, 1958, P 625-646
80. Q.Q. LI AND T.C.T. TING, LINE INCLUSION IN ANISOTROPIC ELASTIC SOLIDS, *J. APPL. MECH.,* VOL 56, 1989, P 556-563
81. T.C.T. TING, EXPLICIT SOLUTION AND INVARIANCE OF THE SINGULARITIES AT AN INTERFACE CRACK IN ANISOTROPIC COMPOSITES, *INT. J. SOLIDS STRUCT.,* VOL 22, 1986, P 965-983
82. A.G. EVANS, M. RUHLE, B.J. DALGLEISH, AND P.G. CHARALAMBIDES, THE FRACTURE ENERGY OF BIMATERIAL INTERFACES, *MATER. SCI. ENG.,* VOL A 126, 1990, P 53-64

83. A.G. EVANS, M. RUHLE, B.J. DALGLEISH, AND P.G. CHARALAMBIDES, THE FRACTURE ENERGY OF BIMATERIAL INTERFACES, *METALL. TRANS. A,* VOL 21, SEPT 1990, P 2419-2429
84. H.F. FISCHMEISTER, H.E. EXNER, M.-H. POECH, S. KOHLHOFF, P. GUMBSCH, S. SCHMAUDER, L.S. SIGL, AND R. SPEIGLER, MODELING FRACTURE PROCESSES IN METALS AND COMPOSITE MATERIALS, *Z. METALLKDE,* BD 80, 1989, P 839-846
85. A. NEEDLEMAN, AN ANALYSIS OF TENSILE DECOHESION ALONG AN INTERFACE, *J. MECH. PHYS. SOLIDS,* VOL 38 (NO. 3), 1990, P 289-324
86. J.H. ROSE, J. FERRANTE, AND J.R. SMITH, UNIVERSAL BINDING ENERGY CURVES FOR METALS AND BIMETALLIC INTERFACES, *PHYS. REV. LETT.,* VOL 47 (NO. 9), AUG 1981, P 675-678
87. J. FERRANTE AND J.R. SMITH, THEORY OF THE BIMETALLIC INTERFACE, *PHYS. REV. B,* VOL 31 (NO. 6), MARCH 1985, P 3427-3434
88. T.C.T. TING, INTERFACE CRACKS IN ANISOTROPIC BIMATERIALS, *J. MECH. PHYS. SOLIDS,* VOL 38 (NO. 4), 1990, P 505-513
89. R.F. COOK, CRACK PROPAGATION THRESHOLDS: A MEASURE OF SURFACE ENERGY, *J. MATER. RES.,* VOL 1 (NO. 6), NOV/DEC 1986, P 852-860
90. R.R. REYNOLDS, K. KOKINI, AND G. CHEN, THE MECHANICS OF THE INTERFACE CRACK USING THE FINITE ELEMENT METHOD, *J. ENG. MATER. TECHNOL.,* VOL 112, JAN 1990, P 38-43
91. P.G. CHARALAMBIDES, J. LUND, A.G. EVANS, AND R.M. MCMEEKING, "A TEST FOR DETERMINING THE FRACTURE RESISTANCE OF BIMATERIAL INTERFACES," REPORT M87-4, MATERIALS DEPARTMENT, COLLEGE OF ENGINEERING, UC SANTA BARBARA
92. N.P. O'DOWD, C.F. SHIH, AND M.G. STOUT, TEST GEOMETRIES FOR MEASURING INTERFACIAL FRACTURE TOUGHNESS, *INT. J. SOLIDS AND STRUCT.,* VOL 29, 1992, P 571-589
93. R.L. BRADY, R.S. PORTER, AND J.A. DONOVAN, A BUCKLED PLATE TEST TO MEASURE INTERFACIAL TOUGHNESS IN COMPOSITES, *J. MATER. SCI.,* VOL 24, 1989, P 4138-4143
94. J.J. MECHOLSKY AND L.M. BAKER, A CHEVRON-NOTCHED SPECIMEN FOR FRACTURE TOUGHNESS MEASUREMENTS OF CERAMIC-METAL INTERFACES, *CHEVRON-NOTCHED SPECIMENS: TESTING AND STRESS ANALYSIS,* STP 855, J.H. UNDERWOOD, S.W. FREIMAN, AND F.I. BARATTA, ED., ASTM, 1984, P 324-336
95. X. ZHU AND J.G. BYRNE, EFFECT OF INTERFACE DEFECTS ON THE MECHANICAL PROPERTIES OF AUTOGENOUS DIFFUSION BONDED JOINTS, *J. MATER. ENG.,* VOL 13, 1991, P 207-211
96. M.D. THOULESS, FRACTURE OF A MODEL INTERFACE UNDER MIXED-MODE LOADING, *ACTA METALL. MATER.,* VOL 38 (NO. 6), 1990, P 1135-1140
97. V.S. GOSTOMEL'SKII, E.S. KARAKOZOV, AND A.P. TERNOVSKII, THE IMPORTANCE OF DIFFUSION AND SURFACE TENSION IN THE FORMATION OF CONTACT DURING DIFFUSION WELDING, *AVT. SVARKA,* NO. 4, 1980, P 28-31
98. A.V. MERIIN AND S.K. SLIOZBERG, SOME CHARACTERISTICS OF JOINTS IN THE SOLID-PHASE WELDING OF METALS, *SVAR. PROIZ.,* NO. 5, 1973, P 22-23

Fundamentals of Friction Welding

J.W. Elmer and D.D. Kautz, Lawrence Livermore National Laboratory

Introduction

FRICTION WELDING (FRW) is a solid-state welding process in which the heat for welding is produced by the relative motion of the two interfaces being joined. This method relies on the direct conversion of mechanical energy to thermal energy to form the weld, without the application of heat from any other source. Under normal conditions no melting occurs at the interface. Figure 1 shows a typical friction weld, in which a nonrotating workpiece is held in contract with a rotating workpiece under constant or gradually increasing pressure until the interface reaches the welding temperature. The rotational speed, the axial pressure, and the welding time are the principal variables that are controlled in order to provide the necessary combination of heat and pressure to form the weld. These parameters are adjusted so that the interface is heated into the plastic temperature range where welding can take place. Once the interface is heated, axial pressure is used to bring the weld interfaces into intimate contact. During this last stage of the welding process, atomic diffusion occurs while the interfaces are in contact, allowing a metallurgical bond to form between the two materials.

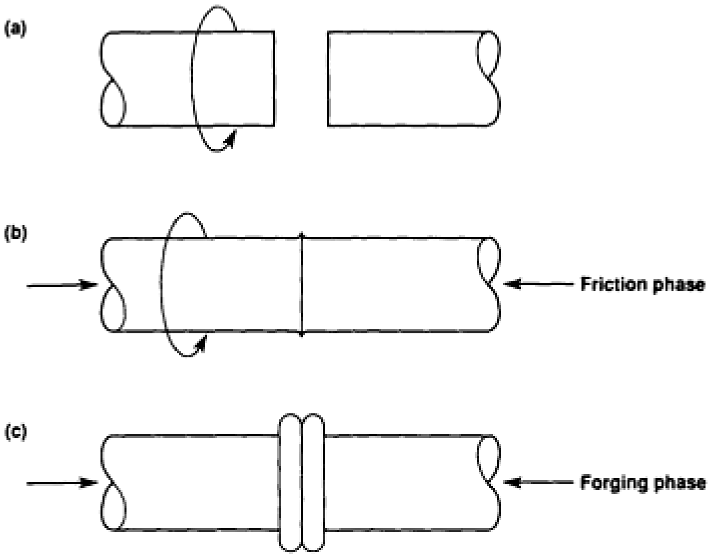

FIG. 1 SCHEMATIC SHOWING FUNDAMENTAL STEPS IN THE FRICTION WELDING PROCESS. (A) ONE WORKPIECE IS ROTATED, AND THE OTHER WORKPIECE IS HELD STATIONARY. (B) BOTH WORKPIECES ARE BROUGHT TOGETHER, AND AXIAL FORCE IS APPLIED TO BEGIN THE UPSETTING PROCESS. (C) WORKPIECE ROTATION IS STOPPED, AND THE UPSETTING PROCESS IS COMPLETED

Friction welding involves heat generation through friction abrasion (Ref 1, 2, 3), heat dissipation, plastic deformation, and chemical interdiffusion. The interrelation among these factors during FRW to complications when trying to develop predictive models of the friction welding process. However, from a qualitative standpoint the process is well understood through empirical FRW studies that have been performed on a wide variety of materials. Five qualitative factors influence the quality of a friction weld (Ref 4):

- RELATIVE VELOCITY OF THE SURFACES
- APPLIED PRESSURE
- SURFACE TEMPERATURE
- BULK MATERIAL PROPERTIES
- SURFACE CONDITION AND PRESENCE OF SURFACE FILMS

The first three factors are related to FRW, while the last two are related to the properties of the materials being joined.

During FRW, the relative velocity, the applied pressure, and the duration of the force are the three variables that are controlled. The effect of these variables on weld quality will be discussed for the two basic friction welding processes: direct-drive welding and inertia-drive welding. The surface temperature is the critical parameter for ensuring good welds and is dependent on the processing conditions and the materials being joined. Although the surface temperature is not measured or directly controlled the effects of insufficient or excessive temperature are generally apparent through visual examination of the finished weld. The bulk material properties and the condition of the surfaces being joined affect both the frictional forces and the forging characteristics of the materials being joined. These factors will be discussed for the friction welding of both similar-material and dissimilar-material combinations.

Commercial FRW applications employ a number of variations on the basic FRW concepts. These variations were developed to accommodate different part geometries and to produce different metallurgical effects (Ref 5) and will not be discussed in this article.

References

1. T.H. HAZLET, FUNDAMENTALS OF FRICTION WELDING, *SOURCE BOOK ON INNOVATIVE WELDING PROCESSES,* AMERICAN SOCIETY FOR METALS, 1981, P 11-36
2. F.P. BOWDEN AND D. TABOR, *THE FRICTION AND LUBRICATION OF SOLIDS,* PART I, OXFORD UNIVERSITY PRESS, 1954
3. J. GODDARD AND H. WILMAN, A THEORY OF FRICTION AND WEAR DURING THE ABRASION OF METALS, *WEAR,* VOL 5, 1962
4. V.I. VILL, *FRICTION WELDING OF METALS,* TRANSLATED FROM THE RUSSIAN, AMERICAN WELDING SOCIETY, 1962
5. *WELDING HANDBOOK,* 8TH ED., VOL 2, AMERICAN WELDING SOCIETY, 1991, P 739-763

Fundamentals of Friction Welding

J.W. Elmer and D.D. Kautz, Lawrence Livermore National Laboratory

Friction Welding Technology

There are two principal FRW methods: direct-drive welding and inertia-drive welding. Direct-drive FRW, sometimes called conventional friction welding, uses a motor running at constant speed to input energy to the weld. Inertia-drive friction welding, sometimes called flywheel friction welding, uses the energy stored in a flywheel to input energy to the weld. These two FRW technologies produce inherently different metallurgical effects at the joint interface.

Both FRW technologies can be applied through different types of relative motion in order to generate the friction necessary to form the weld. The most common FRW geometry is that shown in Fig. 1, in which one cylindrical component is held stationary and the other is rotated. However, in other methods, both components are rotated in opposite directions, or two-stationary components are pushed against a rotating piece positioned between them. Additional forms of FRW, such as radial, orbital, and linear reciprocating motions, have been developed for special part geometries. These alternate methods are discussed elsewhere (Ref 5, 6).

Direct Drive Welding

In direct-drive FRW, a machine resembling a lathe is equipped with a brake and clutch, a means of applying and controlling axial pressure, and a weld-cycle timer and displacement controller. The operation of a direct-drive machine consists of a friction phase where heat is generated, a stopping phase where the rotation is terminated, and a forging phase where the pressure is applied to join the pieces. The relationships among the process variables are shown in Fig. 2, which plots the rotational speed and the axial pressure as a function of time for typical weld. The time required to stop the spindle is also an important variable because it affects the weld temperature and the timing of the forging force.

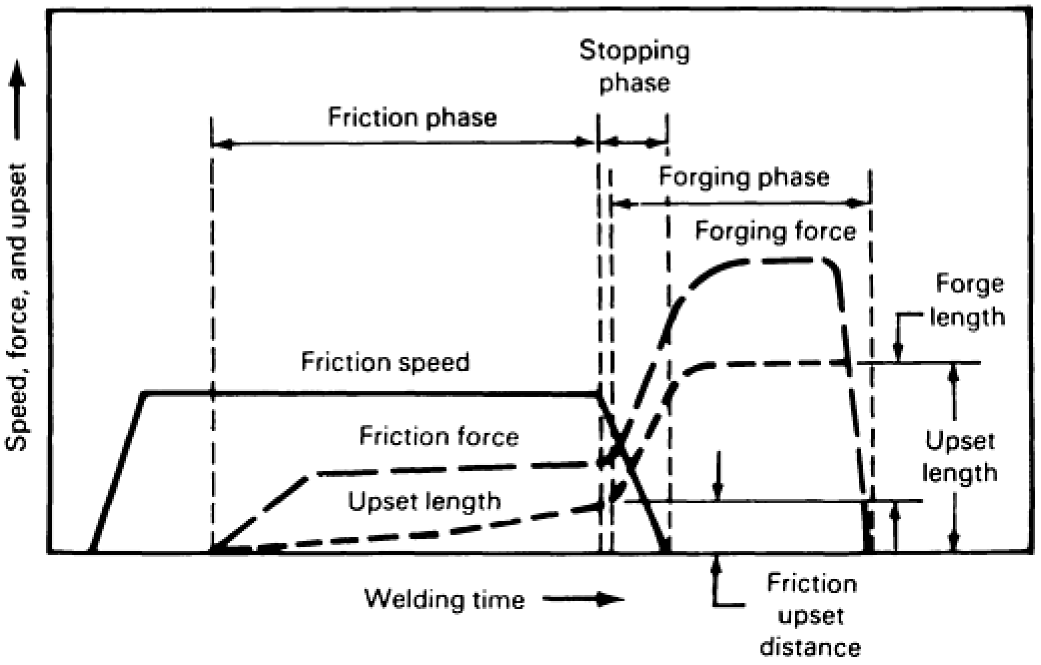

FIG. 2 PLOT OF SELECTED PARAMETERS VERSUS TIME RELATIVE TO THE THREE PHASES OF THE DIRECT-DRIVE FRW PROCESS

The forging phase starts at the instant when higher pressure (that is, a larger forging force) is applied in the weld cycle. Thus, the forging phase actually starts somewhere in the stopping phase. In general, the larger forging force can be applied (case 1) while the spindle is decelerating in the stopping phase (Fig. 2) or (case 2) after the spindle has stopped rotating at the end of the stopping phase. The difference in the two applications of the larger forging force is the presence of a second friction peak. In case 1, the torque will rise again to reach a second peak before dropping. This produces a torsional force. In case 2, especially when the stopping phase is very short due to rapid braking, frictional torque does not rise but actually starts to decrease at the onset of the forging phase. In this case, there is no torsional force, and forging is affected only by the upsetting force.

The speed of rotation is the least sensitive process variable in that it can be varied over a wide range if heating time and pressure are properly controlled. For steels, the recommended peripheral velocity varies from 75 to 215 m/min (250 to 700 ft/min) (Ref 7). In general, higher speeds correspond to low weld heat inputs and are used to weld heat-sensitive materials such as hardenable steels.

The friction force is generally applied gradually to the weld to help overcome the initial contact-torque peak. For carbon steels, a friction pressure of about 70 MPa (10 ksi) at the interface area is required to form a good joint. After the drive motor is disengaged from the workpiece, the forging force is applied to complete the weld. Typical forging forces for carbon steel are of 140 MPa (20 ksi) at the weld interface.

Inertia-Drive Welding

The inertia-drive FRW method uses a similar type of machine except that the spindle holding the rotating piece is attached to a flywheel. The flywheel controls the energy input to the weld. The moment of inertia of the flywheel is an important variable that is adjusted by adding or removing flywheels. The amount of energy stored in the flywheel is controlled by its speed. Once the spindle is at the correct speed, the drive system is disengaged, leaving a rotating flywheel mass. Axial pressure is then applied and held constant throughout the welding process. The applied pressure results in a decrease in the rotational speed, typically referred to as deceleration. In some cases, when the spindle has either nearly stopped or come to a complete stop, a higher forging force may be used.

Figure 3 illustrates the inertia-drive FRW process, which is similar to the direct-drive method in that the weld typically takes place in two stages: friction and forging. However, some weld schedules do not require a forging stage.

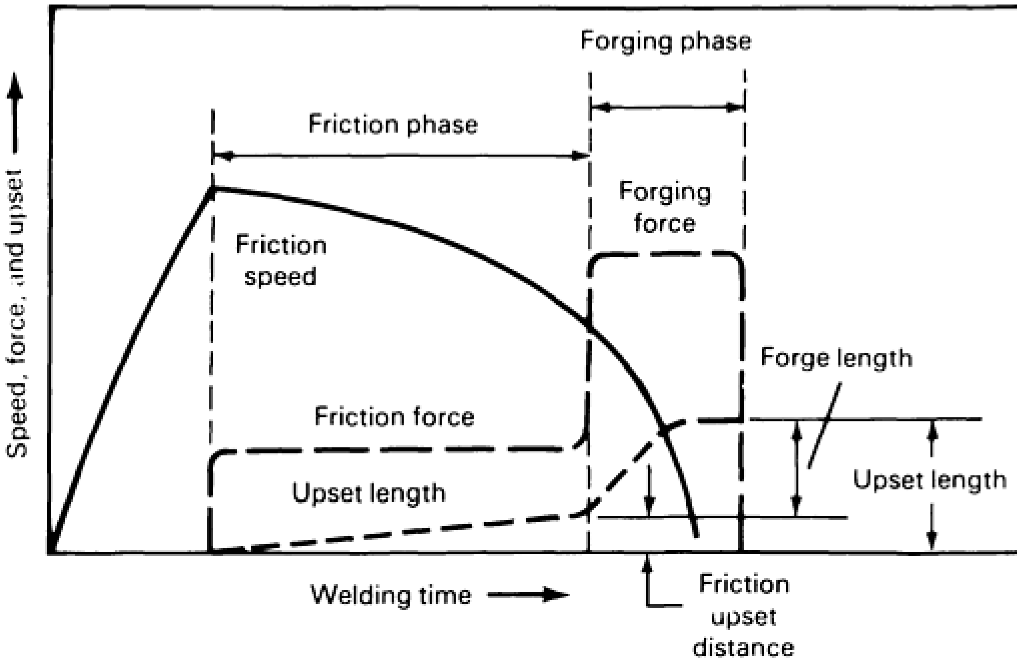
Forging phase
Friction phase
Forging force
Friction speed
Forge length
Friction force
Upset length
Upset length
Speed, force, and upset
Welding time
Friction upset distance

FIG. 3 PLOT OF SELECTED PARAMETERS VERSUS TIME RELATIVE TO THE TWO PHASES OF THE INERTIA-DRIVE FRW PROCESS

The major difference between the direct-drive and inertia-drive methods is the friction speed. In inertia welding, the friction speed continuously decreases during the friction stage, while in direct welding the friction speed stays constant.

The heat generated by the plastic deformation of materials at the faying surfaces, not the heat generated by friction in the friction phase, is the primary source in the forging phase of preventing rapid decrease of the temperature at the interface.

The process variables that control the characteristics of an inertia weld are the flywheel size (moment of inertia), the flywheel speed, and the axial pressure. The weld energy is related to the first two variables and is a fixed quantity once they have been determined. The kinetic energy in the flywheel at any time during the weld is given by:

$$E = \frac{S^2.1}{C}$$

where E is the energy (ft · lbf, or J), I is the moment of inertia (lb/ft^2, or kg/m^2), S is the flywheel speed (rev/min), and C is a conversion constant that is equal to 5873 for English units or 182.4 for metric units.

The constant C is derived from:

$$E = \frac{1}{2} mv^2$$

where $v = \omega r$. Because $\omega = 2\pi s$, $v = 2\pi sr$:

$$E = \frac{1}{2} m(2p)^2 s^2 r^2$$

$$E = \frac{1}{2} mr^2 (2p)^2 s^2$$

Making $mr^2 = I$, the previous equation becomes:

$$E = \frac{1}{2} l(2p)^2 s^2$$

In SI units, the previous equation becomes:

$$E = \left[\frac{m^2.Kg}{s^2}\right] = \frac{1}{2}(2p)^2 s^2 \left[\frac{1}{\min^2}\right] l \left[m^2.Kg\right] \left[\frac{\min}{60s}\right]^2$$

$$E = \frac{1}{2}\frac{(2p)^2}{60^2} s^2 l$$

$$E = \frac{Is^2}{C}$$

Solving for C in the previous equation:

$$C = \frac{2 * 60^2}{(2p)^2} = \frac{60^2}{2p^2} = 182.378$$

The energy stored in the flywheel is proportional to its speed of rotation squared, S^2. Therefore, a wide range of energy levels can be obtained without changing the flywheel to accommodate changes in part geometries. For large changes in the parts being joined, the capacity of an inertia welding machine can be modified by changing the flywheel moment of inertia if necessary.

Flywheel Energy. The moment of inertia of the flywheel is selected to produce both the desired amount of kinetic energy and the desired amount of forging. Forging results from the characteristic increase in torque that occurs at the weld interface as the flywheel slows and comes to rest. This increased torque, in combination with the axial pressure, produces forging. Because forging begins at some critical velocity, the amount of forging depends on the amount of energy remaining in the flywheel, which is a linear function of the flywheel moment of inertia. Large, low-speed flywheels produce greater forging force than small, high-speed flywheels even though they contain the same amount of kinetic energy. Although small, medium, and large amounts of flywheel energy produce similar heating patterns, the amount of energy greatly affects the size and shape of the weld upset, as shown in Fig. 4(a) for similar-metal joints.

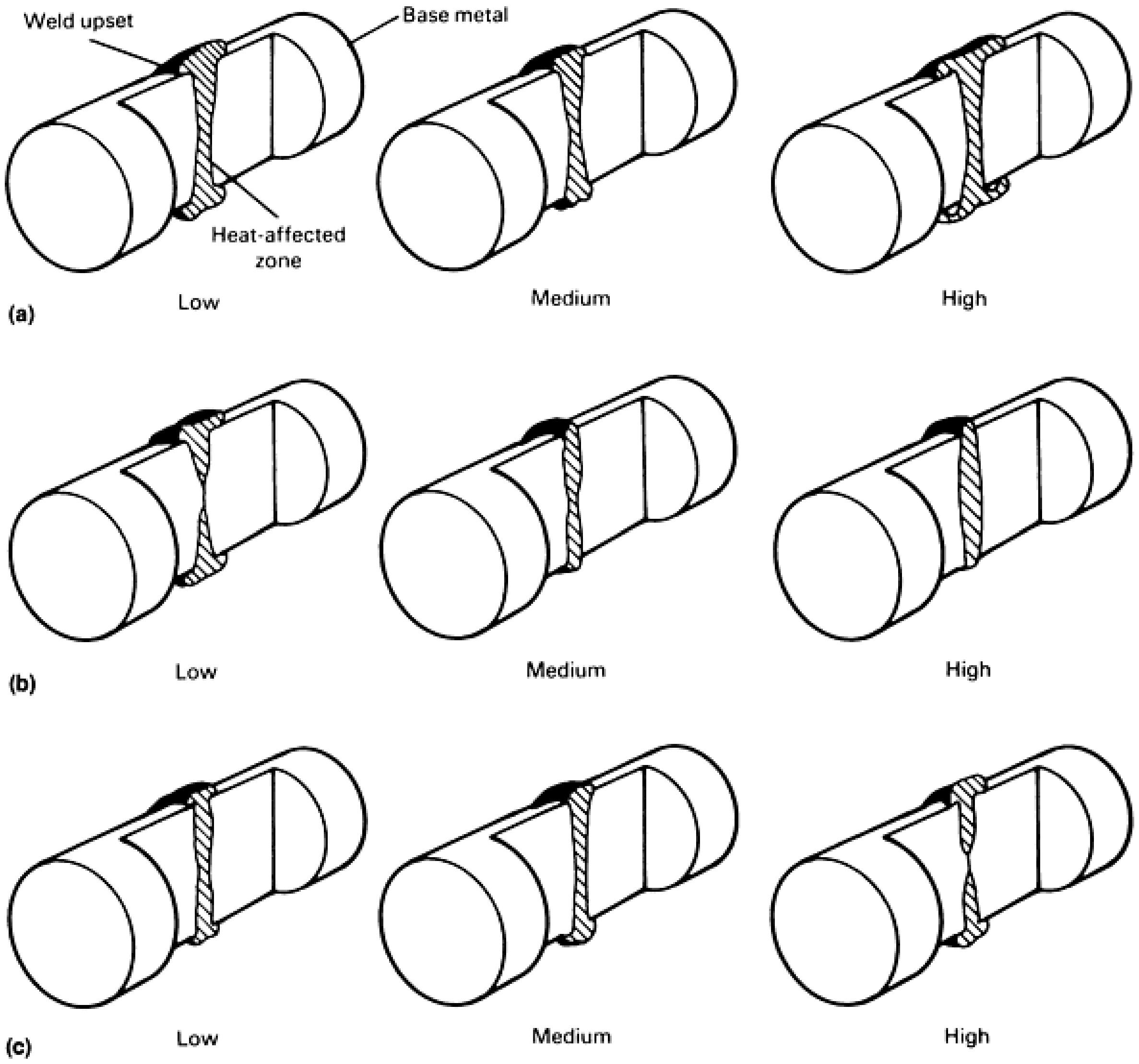

FIG. 4 SCHEMATIC SHOWING EFFECT OF WELDING PARAMETERS ON THE FINISHED WELD NUGGET OBTAINED WHEN SIMILAR METALS ARE WELDED USING INERTIA-DRIVE FRW EQUIPMENT. (A) FLYWHEEL ENERGY. (B) INITIAL PERIPHERAL VELOCITY OF WORKPIECE. (C) AXIAL PRESSURE. SOURCE: REF 7

Peripheral Velocities. There is an optimum range of peripheral or linear surface velocities for each combination of metals being joined. For welding steel to steel, the recommended initial peripheral velocity ranges from 90 to 460 m/min (300 to 1500 ft/min). However, welds can be made at velocities as low as 85 m/min (275 ft/min). Figure 4(b) shows the effect of initial peripheral velocity on weld shape for similar-metal welds.

Axial Pressure Versus Peripheral Velocities. The effect of varying the axial pressure is opposite to the effect of varying the velocity. Welds made at low axial pressure resemble welds made at medium velocity relative to the formation of weld upset and heat-affected zones (HAZ), as shown in Fig. 4(c). Excessive pressure produces a weld that has poor quality at the center and has a large amount of weld upset, similar to a weld made at a low velocity.

References cited in this section

5. *WELDING HANDBOOK,* 8TH ED., VOL 2, AMERICAN WELDING SOCIETY, 1991, P 739-763
6. *RECOMMENDED PRACTICES FOR FRICTION WELDING,* ANSI/AWS C6. 1-89, AMERICAN WELDING SOCIETY, 1989

7. *METALS HANDBOOK*, 9TH ED., VOL 6, AMERICAN SOCIETY FOR METALS, 1983

Metallurgical Parameters

Friction welding can be used to join a wide range of similar and dissimilar materials. Metals, ceramics, metal-matrix composites (MMC), and polymers have all been joined by FRW, and many of the dissimilar-metal combinations that cannot be joined by conventional fusion welding techniques are readily joined by FRW methods. This section summarizes some of the metals that have been joined by FRW and discusses the metallurgical considerations that govern the properties of the resulting weld.

Joining of Similar Metals

The two general requirements for forming good friction welds are, first, that the materials to be joined can be forged and, second, that the materials can generate friction at the weld interface. The first requirement eliminates similar-metal welds in brittle materials such as ceramics, cast irons, and cemented carbides. However, ductile materials can sometimes be joined to these materials. The second requirements eliminates materials that contain alloying that provide dry lubrication. Free-machining additives to steel, graphite-containing alloys such as cast iron, and lead alloys may suffer from this requirement.

Almost all other metal alloys can be welded to themselves by FRW techniques. Table 1 summarizes a number of common similar-metal weld joints that have been made using inertia-drive FRW. Metallic alloys known to form high-quality FRW joints include alloys based on aluminum, copper (copper-nickel, brass, bronze), iron (low-alloy steel, tool steel, stainless steel, maraging steel), nickel, titanium, tantalum, and many others (Ref 5, 6). Near full-strength metallurgical bonds can be produced for a very wide range of similar-metal alloy friction welds. The microstructure and mechanical properties of inertia-welded similar-metal joints for the following alloys can be found in the sources listed:

TABLE 1 PARAMETERS FOR INERTIA-DRIVE FRICTION WELDING OF TWO 25 MM (1 IN.) DIAM BARS MADE OF SIMILAR METALS

WORK METAL	SPINDLE SPEED, REV/MIN	AXIAL FORCE		FLYWHEEL SIZE(A)		WELD ENERGY		METAL LOSS(B)		TOTAL TIME(C), S
		KN	LBF × 10^3	KG · M^2	LB · FT^2	KJ	FT · LBF × 10^3	MM	IN.	
1018 STEEL	4600	53	12	0.28	6.7	33	24	2.5	0.10	2.0
1045 STEEL	4600	62	14	0.33	7.8	38	28	2.5	0.10	2.0
4140 STEEL	4600	67	15	0.35	8.3	41	30	2.5	0.10	2.0
INCONEL 718	1500	220	50	5.48	130.0	68	50	3.8	0.15	3.0
MARAGING STEEL	3000	90	20	0.84	20.0	41	30	2.5	0.10	2.5
TYPE 410 STAINLESS	3000	80	18	0.84	20.0	41	30	2.5	0.10	2.5
TYPE 302 STAINLESS	3500	80	18	0.59	14.0	41	30	2.5	0.10	2.5
COPPER (COMMERCIALLY PURE)	8000	22	5	0.04	1.0	14	10	3.8	0.15	0.5
COPPER ALLOY 260 (CARTRIDGE BRASS, 70%)	7000	22	5	0.05	1.2	14	10	3.8	0.15	0.7
TITANIUM ALLOY, TI-6AL-4V	6000	36	8	0.07	1.7	22	16	2.5	0.10	2.0
ALUMINUM ALLOY 1100	5700	27	6	0.11	2.7	20	15	3.8	0.15	1.0

ALUMINUM ALLOY 6061	5700	31	7	0.13	3.0	23	17	3.8	0.15	1.0

Source: Ref 7

(A) MOMENT OF INERTIA OF THE FLYWHEEL.
(B) TOTAL SHORTENING OF THE WORKPIECES DURING WELDING.
(C) SUM OF HEATING TIME PLUS WELDING TIME.

The relative ease of friction welding metals to themselves is related to the matching properties at the weld interface. Because the materials properties are matched, heat is distributed uniformly on both sides of the joint, and the deformation characteristics are identical on both sides of the joint. This results in symmetric welds with good properties. In general, the process variables do not vary significantly for different alloys within a given class of materials. However, there can be a significant variation in processing variables between different classes of materials (Table 1).

Because FRW generates localized heating at the interface, the HAZ is subject to rapid cooling due to heat transfer to the cold base metal. This rapid quenching may sufficiently alter the mechanical properties of the base metal in the HAZ region to require postweld heat treatment. For example, in order to restore ductility, stress relieving or tempering may be required to friction weld steels with hardenability greater than that of AISI 1035 (Ref 5, 6). In addition, age-hardenable alloys will lose strength in the HAZ during welding and may require postweld solution heat treating and/or postweld aging to restore their strength. Other alloys, such as those that obtain their strength from cold working, will lose strength in the HAZ of the weld, and their properties cannot be restored with postweld treatments.

Joining of Dissimilar Metals

While many similar-metal FRW joints are produced because of economic considerations, many dissimilar-metal FRW joints are produced because there are no alternative welding methods that can be used. Examples of these types of joints include dissimilar-metal combinations with widely different melting points and dissimilar-metal combinations that form incompatible phases when fusion welded. Table 2 gives parameters used for inertia welding several common dissimilar-material combinations.

TABLE 2 PARAMETERS FOR INERTIA-DRIVE FRICTION WELDING OF TWO 25-MM (1-IN.) DIAM BARS MADE OF DISSIMILAR METALS

WORK METAL	**SPINDLE SPEED, REV/MIN**	**AXIAL FORCE**		**FLYWHEEL SIZE(A)**		**WELD ENERGY**		**METAL LOSS(B)**		**TOTAL TIME(C), S**
		KN	**LBF × 10^3**	**KG · M^2**	**LB · FT^2**	**KJ**	**FT · LBF × 10^3**	**MM**	**IN.**	
Copper to 1018 steel	8000	22	5	0.06	1.4	20	15	3.8	0.15	1.0
M2 tool steel to 1045 steel	3000	180	40	1.14	27.0	54	40	2.5	0.10	3.0
Nickel alloy 718 to 1045 steel	1500	180	40	5.48	130.0	68	50	3.8	0.15	2.5
Type 302 stainless to 1020 steel	3000	80	18	0.84	20.0	41	30	2.5	0.10	2.5
Sintered high-carbon steel to 1018	4600	53	12	0.35	8.3	41	30	2.5	0.10	2.5
Aluminum 6061 to type 302 stainless	5500	22 67(D)	5 15(D)	0.16 . . .	3.9 . . .	27 . . .	20 . . .	5.1 . . .	0.20 . . .	3.0 . . .
Copper to aluminum alloy 1100	2000	33	7.5	0.46	11.0	10	7.5	5.1	0.20	1.0

(A) MOMENT OF INERTIA OF THE FLYWHEEL.
(B) TOTAL SHORTENING OF THE WORKPIECES DURING WELDING.

(C) SUM OF HEATING TIME PLUS WELDING TIME.

(D) THE 22 KN (5000 LBF) FORCE IS APPLIED DURING THE HEATING STAGE OF THE WELD; FORCE IS SUBSEQUENTLY INCREASED TO 67 KN (15000 LBF) NEAR THE END OF THE WELD.

Low-Carbon Steels to Medium-Carbon Steels. In general, low- and medium-carbon steels are joined to each other under a wide range of conditions, and high-carbon steels are readily joined to alloy steels (Ref 7) using friction welding. High-speed tool steels are welded to alloy steel shanks for numerous machine-tool applications. Steel with carbon contents as high as 1.0%, such as 52100 steel, can be joined to lower-carbon alloys. Preweld heat treating may be required in some cases to better match the properties at the interface, and postweld heat treatment may be required in some cases to temper the interface region of the high-carbon steel grades.

Stainless Steels to Other Selected Metals. Stainless steel alloys are comparatively easy to friction weld to other metals. For example, austenitic stainless steel to low-alloy steel (Ref 15), titanium and copper to stainless steel (Ref 16), and 1100 aluminum to stainless steel (Ref 17) are examples of transition joints that are made by FRW.

Titanium can be welded to stainless steel with extreme care (Ref 18), and other incompatible dissimilar combinations may be successfully welded using interlayer techniques (Ref 19). Figure 5 shows a micrograph of the interfacial region of an inertia weld between Monel 400 and 21-6-9 stainless steel. The joint properties are excellent, with plastic flow occurring in Monel 400 before joint failure during bend testing.

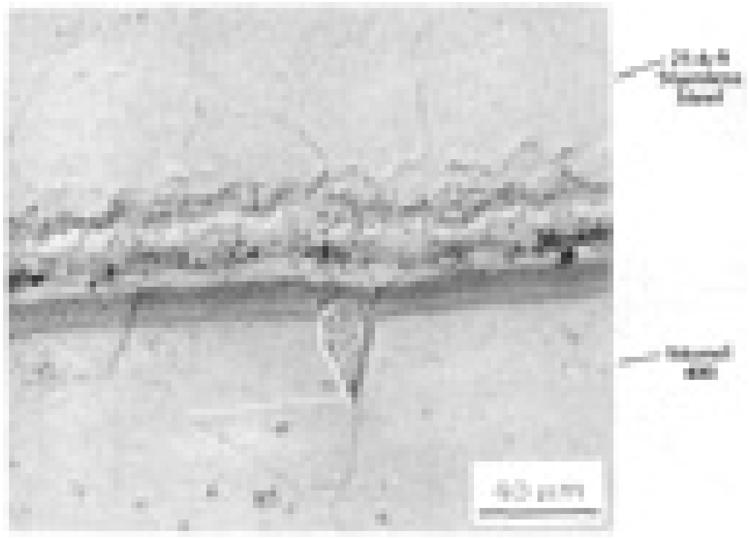

FIG. 5 METALLOGRAPHIC CROSS SECTION OF THE INTERFACE OF A MONEL 400 TO 21-6-9 STAINLESS STEEL WELD PRODUCED BY INERTIA-DRIVE FRW. NOTE THE FINE GRAIN SIZE PRESENT AT THE INTERFACE.

Such transition joints can often be used as interlayers for the friction welding of incompatible materials. For example, it is difficult to weld 5083 aluminum directly to stainless steel. However, by first friction welding aluminum alloy 1100 to the stainless steel, and machining the 1100 aluminum alloy back to an interlayer thickness of about 1 mm (0.04 in.), the 5083 aluminum alloy can be joined to the stainless steel via this 1100 interlayer with high joint efficiencies (Ref 20).

Figure 6(a) shows an example of an aluminum-base MMC that was friction welded to 1100 aluminum. The MMC is a 2024 aluminum alloy with 15 vol% Al_2O_3 particles. The interface region between these materials is shown at higher magnification in Fig. 6(b), where intermixing of both the materials is shown to occur.

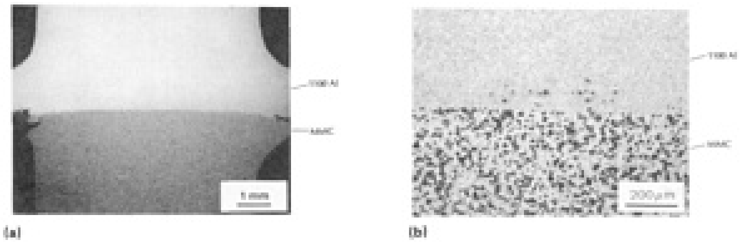

FIG. 6 CROSS SECTION OF THE INTERFACE OF A DIRECT-DRIVE FRICTION WELD JOINING 1100 ALUMINUM TO A 2024 ALUMINUM ALLOY WITH 15 VOL% AL_2O_3 PARTICLES. (B) HIGHER MAGNIFICATION OF THE SAME WELD SHOWING THE EXCELLENT WELD FORMED AT THE INTERFACE

Problems Common to Welding of Dissimilar Materials. In general, the same problems encountered when welding similar materials must be addressed when welding dissimilar materials. However, some problems are associated only with the welding of dissimilar materials or are greatly magnified during the welding of dissimilar materials. These factors include joint interfaces, low-melting phases, brittle phases, and different thermal expansions.

Joint Interfaces. While most similar-material welds are made with little concern for surface preparation, highly dissimilar-metal combinations are more sensitive. This happens for various reasons. In stainless steel to aluminum alloy welds, the oxide surface that forms on the aluminum picks up contaminants such as water and hydrocarbons, forming extremely tenacious surface layers (Ref 21). If this layer is not removed prior to welding, poor structural welds may occur. In stainless steel to refractory metal alloy welds, the oxide on the faying surfaces again contain contaminants such as water and hydrocarbons. The contaminants in this case are likely to alloy into the finished weldment. This alloying causes a reduction of structural integrity through the formation of low-melting or brittle phases at the weld interface.

Surface-treated interfaces frequently cause problems during FRW. Steels that have been carburized or nitrided, titanium alloys that have been nitrided, and other hardfaced alloys are difficult to friction weld due to the inherently low friction coefficient and low forgeability. The repeatability of welds made on materials with hard surface layers is difficult to characterize due to several factors, including coating thickness, coating quality, and physical properties of the coating. In most instances, weldability is improved if the surface-treated area is removed from the faying surface before welding.

Low-Melting Phase Formation. Some material combinations have very low melting point phases associated with mixing of constituents at the weld interface. The formation of these phases during the welding cycle is deleterious to the finished weld properties. Examples of combinations that fall into this category include iron-base alloys to titanium alloys and aluminum alloys to magnesium alloys. Low melting point eutectics are found in both of these metallurgical systems, and great care must be exercised during parameter development to prevent the formation of liquid phases during the completion of successful welds.

Other weld combinations may be affected by contaminants at the weld interface. Examples include sulfur and phosphorus in iron-base alloys and bismuth in copper alloys. These contaminants may cause problems with hot shortness in very low concentrations. It is imperative that good cleaning practices be implemented when materials may have been contaminated with these elements or with material containing these elements.

Brittle Phase Formation. Many materials, when combined, are susceptible to the formation of brittle phases. In some combinations, this occurs during the welding cycle; in others, service conditions after welding cause the problem. Two main reasons exist for brittle phase formation in friction welds:

- SURFACE CONTAMINANTS THAT EMBRITTLE THE WELD INTERFACE (SEE THE SECTION "JOINT INTERFACES" IN THIS ARTICLE)
- FORMATION OF INTERMETALLIC PHASES BETWEEN NORMAL CONSTITUENTS OF THE ALLOYS BEING WELDED

Intermetallic phase formation is common when welding refractory metal alloys to stainless steel alloys and in several other systems. In the case of stainless steels to refractory metals, σ phase or similar phases may occur upon welding at the interface. Proper weld procedure development reduces the amount of brittle phases that are formed, but typically does not eliminate their formation completely. Properly developed welds have satisfactory structural properties, because only small, noncontinuous areas of the brittle phase are present at the weld interface. Figure 7(a) shows an inertia-drive welded joint between vanadium and 21-6-9 stainless steel. The interface is smooth and shows no areas of brittle phases. Electron microscopy techniques are needed to find the small areas of σ phase present at the weld interfaces.

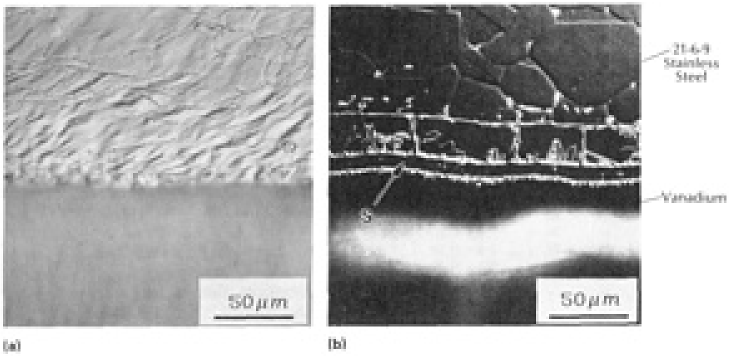

FIG. 7 METALLOGRAPHIC CROSS SECTION OF AN INERTIA-DRIVE FRW JOINT BETWEEN VANADIUM AND A 21-6-9 STAINLESS STEEL. NOTE THE EXCELLENT WELD QUALITY AT THE INTERFACE. (A) WELD INTERFACE WITH NO σ-PHASE GROWTH. (B) WELD INTERFACE WITH σ-PHASE GROWTH (INDICATED BY S) AND A SOLID-SOLUTION MIXING CAUSED BY CHEMICAL DIFFUSION AFTER EXPOSURE TO A 1000 °C (1830 °F) THERMAL CYCLE

Caution must be used when designing components for use at elevated-temperature extremes. In many instances, material combinations in which no brittle phases form during welding are susceptible to brittle phase formation at the interface during high-temperature use. This is not a design issue when the welds are used for near-room-temperature applications. Figure 7(b) shows the vanadium to 21-6-9 stainless steel inertia-drive weld after a severe thermal cycle of 1000 °C (1830 °F) for 2 h with σ-phase and solid-solution growth at the weld interface. The thick layer next to the stainless steel is a solid solution of iron and vanadium. The thin layer next to the vanadium is σphase and forms a continuous brittle fracture path across the weld interface.

Differential Thermal Expansion. Some material combinations are difficult to weld because of the large differences in thermal expansion. Low-expansion materials such as refractory metals, ceramics, and low-expansion iron-nickel and iron-nickel-cobalt alloys may fail or be highly stressed during cooling when welded to high-expansion material such as austenitic stainless steels and nickel-base and cobalt-base superalloys. Use of these combinations requires the designer to consider the large stresses developed within the fabricated structure if the welds are restrained when exposed to large temperature changes. Intermediate expansion materials and multiple friction welds may be required to allow for the transition from high to low thermal expansion materials.

References cited in this section

5. *WELDING HANDBOOK,* 8TH ED., VOL 2, AMERICAN WELDING SOCIETY, 1991, P 739-763
6. *RECOMMENDED PRACTICES FOR FRICTION WELDING,* ANSI/AWS C6. 1-89, AMERICAN WELDING SOCIETY, 1989
7. *METALS HANDBOOK,* 9TH ED., VOL 6, AMERICAN SOCIETY FOR METALS, 1983
8. T.N. HAZLET, PROPERTIES OF FRICTION WELDED PLAIN CARBON AND LOW ALLOY STEELS, *WELD. J.,* VOL 41, 1962, P 49-S TO 52-S
9. S.B. DUNKERTON, TOUGHNESS PROPERTIES OF FRICTION WELDS IN STEELS, *WELD. J.,* VOL 65, 1986, P 193-S TO 202-S
10. J.C. LIPPOLD AND B.C. ODEGARD, TECHNICAL NOTE: MICROSTRUCTURAL EVOLUTION

DURING INERTIA FRICTION WELDING OF AUSTENITIC STAINLESS STEELS, *WELD. J.,* VOL 63, 1984, P 35-S TO 38-S

11. W.A. BAESLACK III AND K.S. HAGEY, INERTIA FRICTION WELDING OF RAPIDLY SOLIDIFIED POWDER METALLURGY ALUMINUM, *WELD. J.,* VOL 67, 1988, P 139-S TO 149-S
12. H.H. KOO AND W.A. BAESLACK III, FRICTION WELDING OF A RAPIDLY SOLIDIFIED AL-FE-V-SI ALLOY, *WELD. J.,* VOL 71, 1992, P 147-S TO 169-S
13. K. OGAWA, H. YAMAGUCHI, T. MORIMOTO, K. TAKEMATA, H. SUDO, AND A. HIRATSUKA, SHEAR STRENGTH CHARACTERISTICS OF ALUMINUM ALLOY FRICTION WELDS, *WELD. J.,* VOL 5, 1991, P 860-866
14. C.G. NESSLAR *ET AL.,* FRICTION WELDING OF TITANIUM ALLOYS, *WELD. J.,* VOL 50, 1971, P 379-S TO 385-S
15. K.G.K. MURTI AND S. SANDARESAN, THERMAL BEHAVIOR OF AUSTENITIC-FERRITIC JOINTS MADE BY FRICTION WELDING, *WELD. J.,* VOL 64, 1985, P 327-S TO 334-S
16. R.A. BELL, J.C. LIPPOLD, AND D.R. ADOLPHSON, AN EVALUATION OF COPPER-STAINLESS STEEL FRICTION WELDS, *WELD. J.,* VOL 63, 1984, P 325-2 TO 332-S
17. D. YASHAN, S. TSANG, W.L. JOHNS, AND M.W. DOUGHTY, INERTIA FRICTION WELDING OF 1100 ALUMINUM TO TYPE 316 STAINLESS STEEL, *WELD. J.,* VOL 66, 1987, P 27-37
18. M. FUTAMATA AND A. FUJI, FRICTION WELDING OF TITANIUM AND SUS 304L AUSTENITIC STAINLESS STEEL, *WELD. INT.,* VOL 4, 1990, P 768-774
19. F. SASSANI AND J.R. NEELAN, FRICTION WELDING OF INCOMPATIBLE MATERIALS, *WELD. J.,* VOL 67, 1988, P 264-S TO 270-S
20. R. ARMSTRONG, PRIVATE COMMUNICATION AND INTERNAL REPORTS, LAWRENCE LIVERMORE NATIONAL LABORATORY, 1991
21. *METALS HANDBOOK,* 9TH ED., VOL 2, AMERICAN SOCIETY FOR METALS, 1979, P 204-209

Fundamentals of Friction Welding

J.W. Elmer and D.D. Kautz, Lawrence Livermore National Laboratory

---

References

1. T.H. HAZLET, FUNDAMENTALS OF FRICTION WELDING, *SOURCE BOOK ON INNOVATIVE WELDING PROCESSES,* AMERICAN SOCIETY FOR METALS, 1981, P 11-36
2. F.P. BOWDEN AND D. TABOR, *THE FRICTION AND LUBRICATION OF SOLIDS,* PART I, OXFORD UNIVERSITY PRESS, 1954
3. J. GODDARD AND H. WILMAN, A THEORY OF FRICTION AND WEAR DURING THE ABRASION OF METALS, *WEAR,* VOL 5, 1962
4. V.I. VILL, *FRICTION WELDING OF METALS,* TRANSLATED FROM THE RUSSIAN, AMERICAN WELDING SOCIETY, 1962
5. *WELDING HANDBOOK,* 8TH ED., VOL 2, AMERICAN WELDING SOCIETY, 1991, P 739-763
6. *RECOMMENDED PRACTICES FOR FRICTION WELDING,* ANSI/AWS C6. 1-89, AMERICAN WELDING SOCIETY, 1989
7. *METALS HANDBOOK,* 9TH ED., VOL 6, AMERICAN SOCIETY FOR METALS, 1983
8. T.N. HAZLET, PROPERTIES OF FRICTION WELDED PLAIN CARBON AND LOW ALLOY STEELS, *WELD. J.,* VOL 41, 1962, P 49-S TO 52-S
9. S.B. DUNKERTON, TOUGHNESS PROPERTIES OF FRICTION WELDS IN STEELS, *WELD. J.,* VOL 65, 1986, P 193-S TO 202-S
10. J.C. LIPPOLD AND B.C. ODEGARD, TECHNICAL NOTE: MICROSTRUCTURAL EVOLUTION

DURING INERTIA FRICTION WELDING OF AUSTENITIC STAINLESS STEELS, *WELD. J.,* VOL 63, 1984, P 35-S TO 38-S

11. W.A. BAESLACK III AND K.S. HAGEY, INERTIA FRICTION WELDING OF RAPIDLY SOLIDIFIED POWDER METALLURGY ALUMINUM, *WELD. J.,* VOL 67, 1988, P 139-S TO 149-S
12. H.H. KOO AND W.A. BAESLACK III, FRICTION WELDING OF A RAPIDLY SOLIDIFIED AL-FE-V-SI ALLOY, *WELD. J.,* VOL 71, 1992, P 147-S TO 169-S
13. K. OGAWA, H. YAMAGUCHI, T. MORIMOTO, K. TAKEMATA, H. SUDO, AND A. HIRATSUKA, SHEAR STRENGTH CHARACTERISTICS OF ALUMINUM ALLOY FRICTION WELDS, *WELD. J.,* VOL 5, 1991, P 860-866
14. C.G. NESSLAR *ET AL.,* FRICTION WELDING OF TITANIUM ALLOYS, *WELD. J.,* VOL 50, 1971, P 379-S TO 385-S
15. K.G.K. MURTI AND S. SANDARESAN, THERMAL BEHAVIOR OF AUSTENITIC-FERRITIC JOINTS MADE BY FRICTION WELDING, *WELD. J.,* VOL 64, 1985, P 327-S TO 334-S
16. R.A. BELL, J.C. LIPPOLD, AND D.R. ADOLPHSON, AN EVALUATION OF COPPER-STAINLESS STEEL FRICTION WELDS, *WELD. J.,* VOL 63, 1984, P 325-2 TO 332-S
17. D. YASHAN, S. TSANG, W.L. JOHNS, AND M.W. DOUGHTY, INERTIA FRICTION WELDING OF 1100 ALUMINUM TO TYPE 316 STAINLESS STEEL, *WELD. J.,* VOL 66, 1987, P 27-37
18. M. FUTAMATA AND A. FUJI, FRICTION WELDING OF TITANIUM AND SUS 304L AUSTENITIC STAINLESS STEEL, *WELD. INT.,* VOL 4, 1990, P 768-774
19. F. SASSANI AND J.R. NEELAN, FRICTION WELDING OF INCOMPATIBLE MATERIALS, *WELD. J.,* VOL 67, 1988, P 264-S TO 270-S
20. R. ARMSTRONG, PRIVATE COMMUNICATION AND INTERNAL REPORTS, LAWRENCE LIVERMORE NATIONAL LABORATORY, 1991
21. *METALS HANDBOOK,* 9TH ED., VOL 2, AMERICAN SOCIETY FOR METALS, 1979, P 204-209

Fundamentals of Diffusion Bonding

Murray W. Mahoney and Cliff C. Bampton, Rockwell International Science Center

Introduction

DIFFUSION BONDING is only one of many solid-state joining processes wherein joining is accomplished without the need for a liquid interface (brazing) or the creation of a cast product via melting and resolidification (welding). In its most narrow definition, which is used to differentiate it from other joining processes such as deformation bonding or transient liquid phase joining, diffusion bonding (DB) is a process that produces solid-state coalescence between two materials under the following conditions:

- JOINING OCCURS AT A TEMPERATURE BELOW THE MELTING POINT, T_M, OF THE MATERIALS TO BE JOINED (USUALLY $>1/2T_M$).
- COALESCENCE OF CONTACTING SURFACES IS PRODUCED WITH LOADS BELOW THOSE THAT WOULD CAUSE MACROSCOPIC DEFORMATION TO THE PART.
- A BONDING AID CAN BE USED, SUCH AS AN INTERFACE FOIL OR COATING, TO EITHER FACILITATE BONDING OR PREVENT THE CREATION OF BRITTLE PHASES BETWEEN DISSIMILAR MATERIALS, BUT THE MATERIAL SHOULD NOT PRODUCE A LOW-TEMPERATURE LIQUID EUTECTIC UPON REACTION WITH THE MATERIALS TO BE JOINED.

Thus, diffusion bonding facilitates the joining of materials to produce components with no abrupt discontinuity in the microstructure and with a minimum of deformation.

Within the confines of this definition, the DB process, in practice, is limited to either press or gas pressure or bonding approaches. Illustrations and discussions of equipment and systems are presented elsewhere within this Handbook (see the Section "Solid-State Welding, Brazing, and Soldering Processes"). This article offers a qualitative summary of the theory of diffusion bonding. For those who require a more quantitative assessment, refer to the Selected References at the end of the article.

It should be noted that the preferred term for this process, according to the American Welding Society, is diffusion *welding.* However, because diffusion *bonding* is used more commonly in industry, it is the term that will be used in this article.

Fundamentals of Diffusion Bonding

Murray W. Mahoney and Cliff C. Bampton, Rockwell International Science Center

---

Diffusion Bonding Process

The DB process, that is, the application of pressure and temperature to an interface for a prescribed period of time, is generally considered complete when cavities fully close at the faying surfaces. Relative agreement is found for the mechanisms and sequence of events that lead to the collapse of interface voids, and the discussion below describes these metallurgical processes. Although this theoretical understanding of the DB process is universally applicable, it should be understood that parent metal strength is only approached for materials with surface conditions that do not have barriers to impede atomic bonding such as the absence of surface oxides or absorbed gases at the bonding interface.

In practice, oxide-free conditions exist only for a limited number of materials. Accordingly, the properties of real surfaces limit and impede the extent of diffusion bonding. The most notable exception is titanium alloys, which, at DB temperatures greater than 850 °C (1560 °F), can readily dissolve minor amounts of adsorbed gases and thin surface oxide films and diffuse them away from the bonding surfaces, so that they will not impede the formation of the required metallic bonds across the bond interface, as shown in Fig. 1(a). An example of a successful diffusion bond in a titanium alloy is illustrated in Fig. 1(b), where the integrity of the bonded interfaces was demonstrated with subsequent superplastic expansion without interface failure.

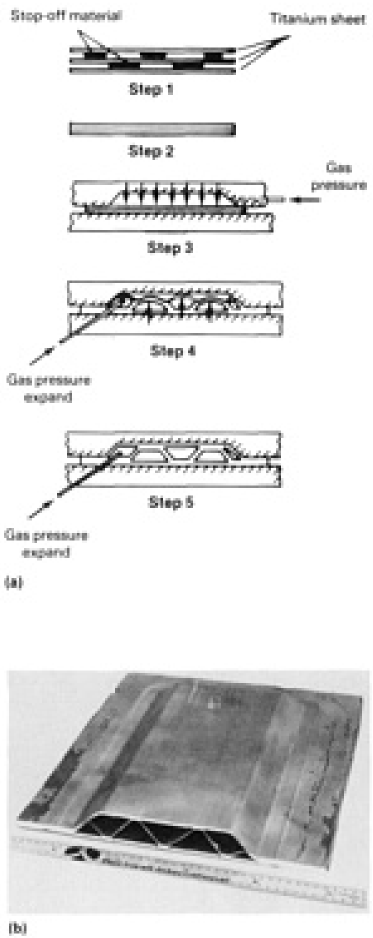

FIG. 1 SUPERPLASTIC FORMING/DIFFUSION BONDING (SPF/DB) OF TITANIUM SHEET. (A) SEQUENCE OF OPERATIONS REQUIRED TO JOIN THREE SHEETS OF SUPERPLASTIC TITANIUM ALLOY USING SPF/DB PROCESS. (B) TYPICAL THREE-SHEET TITANIUM ALLOY COMPONENT SUPERPLASTICALLY FORMED FOLLOWING DIFFUSION BONDING.

Similarly, the joining of silver at 200 °C (390 °F) requires no deformation to break up and disperse oxides, because silver oxide dissociates completely at 190 °C (375 °F). Above this temperature, silver dissolves its oxide and also scavenges many surface contaminants. Other examples of metals that have a high solubility for interstitial contaminants include tantalum, tungsten, copper, iron, zirconium, and niobium. Accordingly, this class of alloy is easiest to diffusion bond.

A second class of material, that is, metals and alloys that exhibit very low solubility for interstitials (such as aluminum-, iron-, nickel-, and cobalt-base alloys) are not readily diffusion bondable. Special consideration must be given to remove surface barriers to atomic diffusion prior to joining and subsequently prevent their reformation during the joining process. This is not an easy processing matter. Accordingly, the potential for high-strength bond interfaces for alloys with low interstitial solubility should be considered on an individual alloy basis.

Fundamentals of Diffusion Bonding

Murray W. Mahoney and Cliff C. Bampton, Rockwell International Science Center

---

Bonding Surfaces Containing Oxides

Diffusion bonding can be achieved for materials with adherent surface oxides, but the resultant interface strengths of these materials are considerably less than that measured for the parent material. Aluminum alloys are prime examples of this class of material. Research since 1960 has demonstrated only limited diffusion bond properties. Although interface strength can be increased for oxide-bearing materials, it requires considerable surface extension of the faying interfaces to create localized plastic flow of the metal and concurrent oxide breakage. This introduces an increased number of locations for metal-to-metal contact via plastic flow around or microextrusion through the broken oxide. In general, the oxide is not removed, but is simply dispersed over a greater surface area in an enclosed environment, in which oxidation cannot recur. Thus, even with significant surface deformation, only a fraction of the interface area contributes to the strength of the bond. The proportion of oxide-free metallic area revealed is dependent on the relative hardness of the metal and its oxide film, as well as on the mechanical properties of the oxide. This type of bonding, although often considered as diffusion bonding, is better described as deformation bonding and does not fit within the strict definition of the low deformation associated with diffusion bonding.

Factors that affect the relative difficulty of diffusion bonding oxide-bearing surfaces include:

- *SURFACE ROUGHNESS PRIOR TO WELDING.* A ROUGHER SURFACE WILL RESULT IN GREATER SHEAR DEFORMATION.
- *MECHANICAL PROPERTIES OF THE OXIDE.* THE MORE BRITTLE THE OXIDE, THE GREATER THE DISPERSION FOR A GIVEN LEVEL OF DEFORMATION.
- *RELATIVE HARDNESS OF THE METAL AND ITS OXIDE FILM.* BECAUSE PLASTIC FLOW CONTROLS THE AMOUNT OF BONDING AREA, LARGE DIFFERENCES IN THEIR HARDNESS SHOULD FACILITATE BONDING.
- *PRESTRAINING OR WORK HARDENING OF THE MATERIAL.* INITIATION OF BONDING WILL OCCUR AT LOWER DEFORMATIONS FOR PRESTRAINED OR WORK-HARDENED MATERIALS, AND THE DEGREE OF SURFACE EXTENSION IN THE CENTRAL REGION OF THE INTERFACE IS CONSIDERABLY GREATER FOR COLD-WORKED MATERIAL. THUS, ANNEALED MATERIAL REQUIRES A LARGER TOTAL DEFORMATION BEFORE BONDING WILL INITIATE.

It is clear that with the appropriate information, sufficient experiments can be performed to determine the diffusion bondability of most materials. Parent metal strength will not always be attained using the DB approach, particularly for materials with adherent oxides, but interface strength can be maximized if the fundamentals of the process are understood.

Fundamentals of Diffusion Bonding

Murray W. Mahoney and Cliff C. Bampton, Rockwell International Science Center

Mechanism of Diffusion Bonding

In diffusion bonding, the nature of the joining process is essentially the coalescence of two atomically clean solid surfaces. Complete coalescence comes about through a three-stage metallurgical sequence of events. Each stage, as shown in Fig. 2, is associated with a particular metallurgical mechanism that makes the dominant contribution to the bonding process. Consequently, the stages are not discretely defined, but begin and end gradually, because the metallurgical mechanisms overlap in time. During the first stage, the contact area grows to a large fraction of the joint area by localized deformation of the contacting surface asperities. Factors such as surface roughness, yield strength, work hardening, temperature, and pressure are of primary importance during this stage of bonding. At the completion of this stage, the interface boundary is no longer a planar interface, but consists of voids separated by areas of intimate contact. In these areas of contact, the joint becomes equivalent to a grain boundary between the grains on each surface. The first stage is usually of short duration for the common case of relatively high-pressure diffusion bonding.

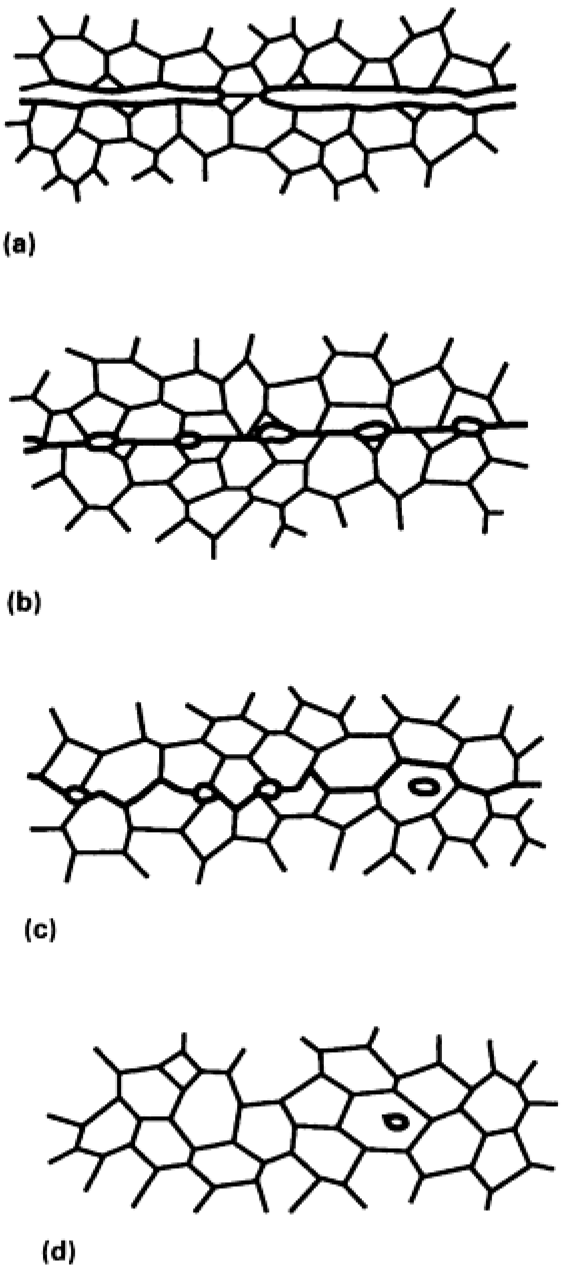

FIG. 2 SEQUENCE OF METALLURGICAL STAGES IN DIFFUSION BONDING PROCESS. (A) INITIAL CONTACT: LIMITED TO A FEW ASPERITIES (ROOM TEMPERATURE). (B) FIRST STAGE: DEFORMATION OF SURFACE ASPERITIES BY PLASTIC FLOW AND CREEP. (C) SECOND STAGE: GRAIN BOUNDARY DIFFUSION OF ATOMS TO THE VOIDS AND GRAIN BOUNDARY MIGRATION. (D) THIRD STAGE: VOLUME DIFFUSION OF ATOMS TO THE VOIDS

During the second stage of joint formation, two changes occur simultaneously. All of the voids in the joints shrink, and most are eliminated. In addition, the interfacial grain boundary migrates out of the plane of the joint to a lower-energy equilibrium. Creep and diffusion mechanisms are important during the second stage of bonding and for most, if not all, practical applications, bonding would be considered essentially complete following this stage. As the boundary moves, any remaining voids are engulfed within grains where they are no longer in contact with a grain boundary. During this third stage of bonding, the voids are very small and very likely have no impact on interface strength. Again, diffusional processes cause the shrinkage and elimination of voids, but the only possible diffusion path is now through the volume of the grains themselves.

Stage I: Microasperity Deformation

The nature of the starting surface is of considerable importance, because of the small macroscopic deformation allowed during diffusion bonding. A real surface is never perfectly clean or perfectly smooth, and the area of metal-to-metal contact between faying surfaces is a very small fraction of the area of joint contact. Contact is limited to a relatively few microasperities. At room temperature and under load, these asperities deform as long as the surface area of contact is such that the yield strength of the material is exceeded. The extent of this deformation is limited at room temperature, and is even more limited for work-hardenable materials. As temperature increases to the diffusion bonding temperature, the flow stress of the material decreases and additional asperity deformation occurs through plastic flow. Again, flow occurs until the area of contact increases to an extent that the yield strength of the material is exceeded. If the temperature is above the recrystalization temperature of the material, then work hardening is no longer a consideration. With time at temperature, creep mechanisms now control the rate of asperity deformation, and the area of contact or bond continues to grow. As the area of contract grows, the stress acting on the surface asperities decreases. Consequently, creep deformation progressively slows and diminishes in significance.

The contributions of temperature and pressure to both plastic and creep deformation during this initial stage of diffusion bonding are synergistic, that is, at higher temperatures, less pressure is required and vice versa. However, for any combination of temperature and pressure, bulk deformation to the part is limited to a small percentage (<2 to 3%). Ideally, at the completion of the first stage, the extent of asperity collapse should result in a planar area of contacting surfaces with individually dispersed voids. It is necessary to achieve this extend of contact in order to complete the final stages of diffusion bonding in a reasonable period of time.

For example, Fig. 3 illustrates the influence of pressure on the bond line morphology for a titanium alloy. At lower pressures, where surface asperity deformation is less during the first stage of bonding, large voids remain at the interface, even after a reasonable time at temperature. Conversely with a higher applied pressure and a correspondingly greater initial interface deformation, the bond interface becomes indistinguishable from the matrix alloy. It should be remembered that, by definition, very little bulk deformation is allowed and that only interface microdeformation contributes to the growth in contact area of the faying surfaces.

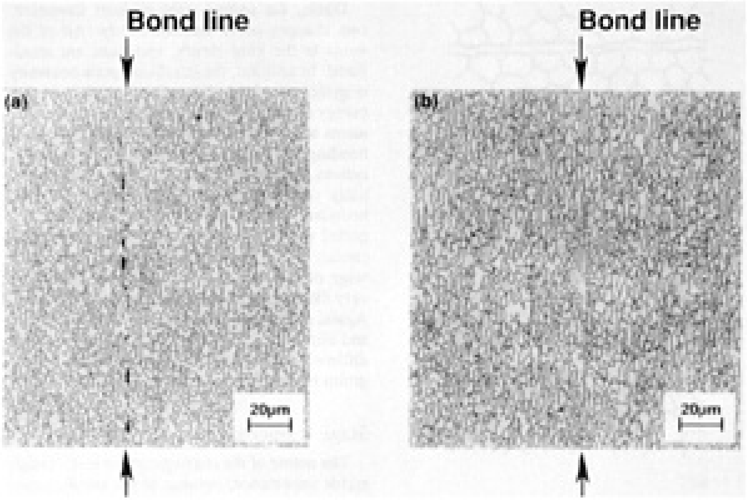

FIG. 3 EFFECT OF PRESSURE ON PRESENCE OF VOIDS AT BAND INTERFACE OF TITANIUM ALLOY DIFFUSION BONDED AT TEMPERATURES OF 980 °C (1795 °F) FOR 2 H. (A) INCOMPLETE BOND AT 7.0 MPA (1.0 KSI). (B) COMPLETE BOND AT 10.0 MPA (1.5 KSI)

Surface Roughness. When considering the sequences in the stages of diffusion bonding, it is clear that the original surface finish plays a significant role in the time dependency for the completion of the different stages. The surfaces that are mated at the bond line are usually rather irregular, as a result of the machining or other surface preparation steps. As shown in Fig. 4, the surface roughness can be viewed as a bimodal distribution of asperities, that is, small, short-wavelength asperities (surface roughness) arrange on larger, long-wavelength asperities (surfaces waviness).

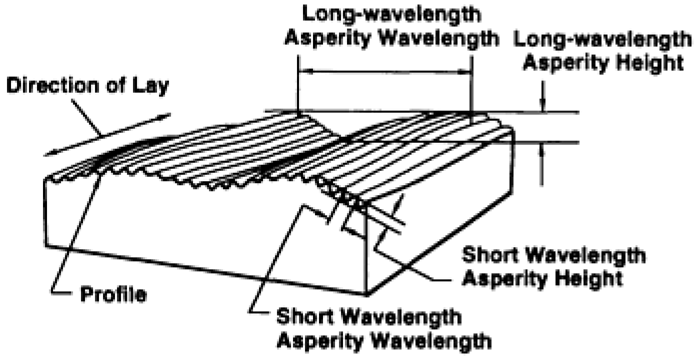

FIG. 4 BIMODAL TOPOGRAPHY OF MECHANICAL SURFACES

It is expected that during the first bonding stage, a more uneven initial surface that is due to either roughness or waviness would produce more large voids than would a smooth initial surface. For example, as load is applied, asperities on rough surfaces will experience higher stresses on the points of contact. Plastic flow will occur at a lower load, and greater interface shear will occur. Although first-stage plastic flow and creep will create more localized interface deformation for a rough surface, fewer but larger voids will remain at the interface (the contact area should approach an equilibrium value independent of the initial surface roughness). In addition to size, the shape and curvature of the voids would be considerably different, with higher aspect ratio voids for the rougher starting surfaces. For the subsequent diffusion-controlled collapse of interface voids during second-stage bonding, an initially rough surface will require a longer time or higher temperatures for the mass-transport processes to shrink and eliminate the voids.

Conversely, because real surfaces are not perfectly clean, some limited degree of interface deformation is necessary for the diffusion bonding of even oxide-free materials. Experimental evidence has demonstrated that even gold is not joined unless some shear displacement occurs as the two faying surfaces come into contact. It seems that the importance of the shear displacement, besides increasing the contact area, is that it destroys the continuity of any adsorbed oxygen layer that can contaminate the oxide-free area because of trapped air at the interface. Procedures as simple as wire brushing have been shown to be a particularly effective surface preparation. This can be partly attributed to the creation of rough layers on the surface.

Stage II: Diffusion-Controlled Mass Transport

The densification or collapse of interface cavities during the second stage of diffusion bonding is attributable to the lowering of the surface free energy by the decrease in surface area. This takes place with the formation of new, but lower-energy, solid-solid interfaces. Because the driving force is the same (reduction of surface energy) for all systems, the considerable differences in behavior in various types of systems are related to the different mechanisms of material transfer.

In diffusion bonding, a number of mass-transport processes are operative simultaneously, including time-dependent plastic flow, diffusion from the interface to the cavity via the lattice and the interface and grain boundaries, and diffusive flow around the surface of the void via the lattice and vapor phase. These different paths of mass transfer are illustrated in Fig. 5 and include:

- PLASTIC YIELDING THAT DEFORMS AN ORIGINAL CONTACTING ASPERITY
- SURFACE DIFFUSION FROM A SURFACE SOURCE TO A NECK
- VOLUME DIFFUSION FROM A SURFACE SOURCE TO A NECK
- EVAPORATION FROM A SURFACE SOURCE TO CONDENSATION AT A NECK
- GRAIN BOUNDARY DIFFUSION FROM AN INTERFACIAL SOURCE TO A NECK
- VOLUME DIFFUSION FROM AN INTERFACIAL SOURCE TO A NECK
- POWER-LAW CREEP

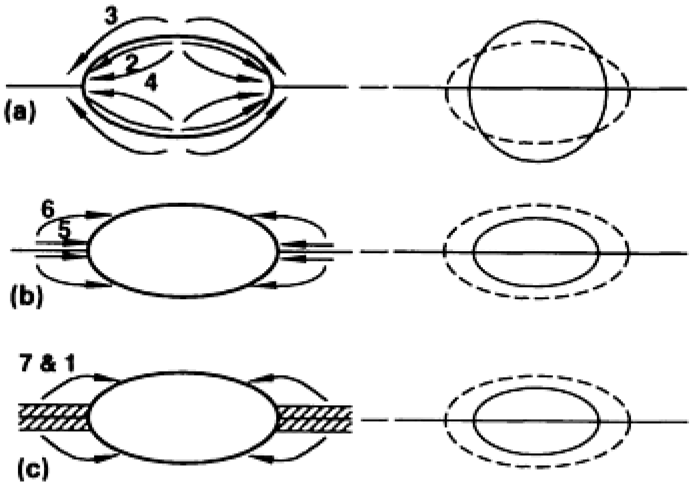

FIG. 5 SCHEMATIC OF NUMEROUS PATHS OF MATERIAL TRANSFER GENERATED DURING DIFFUSION BONDING PROCESS. (A) SURFACE SOURCE MECHANISMS. (B) INTERFACE SOURCE MECHANISMS. (C) BULK DEFORMATION MECHANISMS. SEE TEXT FOR SPECIFIC MECHANISMS INDICATED BY NUMBERS SHOWN IN SCHEMATIC.

Although vapor transport and surface transport of matter change the shape of the void, they do not directly alter the volume. Thus, these transport paths have only a secondary influence on the rate of void collapse indirectly, through a change in void curvature. Only the transfer of matter from the matrix volume or from grain boundaries causes shrinkage and pore elimination. Thus, only the first two mechanisms, plastic flow and interface diffusion, actually reduce the volume of the interfacial voids. The dominance of either of these two processes is dependent on many inseparable factors, including material system, void geometry, microstructure, and bonding parameters.

When considering diffusion, evidence seems to suggest that the most efficient path for atom flow and counterdiffusing vacancies is along the interfacial grain boundary and along grain boundaries that intersect the void when the grain size is less than the pore size. Although the area of a grain boundary is small in comparison to the void surface, this high-diffusivity path has been shown to dominate. For example, bonding has been shown to be enhanced by first blasting the faying surfaces with chilled iron grit. This treatment promoted recrystallization and, thus a finer grain size across the interface, creating additional high-diffusivity paths. In a related process (the sintering of Al_2O_3), porosity was shown to be eliminated next to grain boundaries, with residual porosity remaining at grain centers. Similarly, fine grain size, as exists in superplastic alloys, should create higher rates of void coalescence via grain boundary diffusion. However, because the chemical potential driving force for grain boundary diffusion mechanisms is partly dependent on the angle between the applied pressure and a particular grain boundary, its actual contribution will be dependent on its precise orientation with respect to the applied pressure. The driving force will be maximal when the two are perpendicular and minimal when they are parallel.

Although there is insufficient evidence for generalization, evidence also exists to illustrate that diffusion is, at times, the rate-controlling factor for void coalescence. For example, maintaining temperature while removing pressure following the first bonding stage was shown to not significantly impede the joining process. The rate of void elimination was only slightly slower without the applied pressure. However, this change in rate could be due to either the lost contribution that deformation makes to void coalescence or a reduction in efficiency of the stress-enhanced grain boundary contribution.

Stage III: Interface Migration

During the second stage of diffusion bonding, voids become much smaller and many are eliminated. Their grain boundary pinning influence decreases, so that interfacial grain boundary migrates toward an equilibrium configuration, which is indistinguishable from the other grain boundaries in the microstructure. The driving force for the boundary migration is the reduction in grain boundary area. The initially straight interfacial bond line becomes distorted with local penetrations of a few microns of one material into the other at triple points. As the boundary moves, any remaining voids become enclosed within grains where they are no longer in contact with a grain boundary. Diffusional processes continue to shrink and eliminate these cavities, but the diffusion path is now restricted to volume diffusion through the matrix lattice.

Accordingly, the elimination of this final small volume of porosity would be likely to contribute an inordinate time to the bonding process for any interface strength benefits that might be achieved.

Diffusion Bonding with Interface Aids

Additional layers of material in the form of coatings or foils are often used as bonding aids for a variety of reasons. For example, an intermediate material can be used when joining dissimilar materials where a brittle intermetallic would otherwise form. In this case, the interfacial material would be selected for its compatibility with each of the materials to be joined and for its ability to prevent the creation of a brittle reaction layer. To promote diffusion in materials that contain elements with low diffusivities, the interfacial material should contain an element with a higher mobility than elements found in the joined materials. A common material for such applications is electroless nickel, which contains phosphorus (phosphorus has been shown to have a high diffusivity in other metallic systems). However, caution should be exercised when considering the addition of high-diffusivity elements, because of their potential for accumulation at grain boundaries and their resultant influence on mechanical properties.

Another approach is to add an interfacial material that will scavenge impurity elements at the interface and thus produce clean surfaces *in situ.* Materials with high solubilities for interstitial elements, such as titanium alloys, can be appropriate for this purpose. Because of the importance of localized plastic flow at the interface, a soft material addition can also be of benefit to maximize interfacial contact during the first bonding stage, where deformation mechanisms dominate.

With the addition of interfacial materials, geometric as well as metallurgical considerations become important. The mechanical strength of solid-state metallic bonds achieved with a thin interfacial layer of bonding material goes through a maximum, with decreasing joint thickness. For thick joints, tensile strength is directly related to the bulk properties of the interfacial layer material. As joint thickness decreases, the tensile strength of these joints increases, because of the matrix material restraint on the plastic flow of the interfacial layer. However, for very thin joints, the problems of surface roughness and cleanliness start to diminish the contact area and, thus, effectively reduce the joint tensile strength. experimental studies should be performed for individual materials. In general, thicknesses of approximately 0.025 mm (0.00098 in.) yield maximum interface strengths.

Selected References

- B.M. AGERS, THE MECHANISM OF SMALL TOOL PRESSURE WELDING, *BRIT. WELD. J.,* JULY 1964, P 313-319
- N. BREZDS, INVESTIGATION OF THE FACTORS DETERMINING THE TENSILE STRENGTH OF BRAZED JOINTS, *WELD. RES. SUPP.,* NOV 1954, P 545S-563S
- C.L. CLINE, AN ANALYTICAL AND EXPERIMENTAL STUDY OF DIFFUSION BONDING, *WELD. RES. SUPP.,* NOV 1966, P 481S-489S
- R.L. COBLE, DIFFUSION MODELS FOR HOT PRESSING WITH SURFACE ENERGY AND PRESSURE EFFECTS AS DRIVING FORCE, *J. APPL. PHYS.,* VOL 41.12, 1970, P 4798-4807
- B. DERBY AND E.R. WALLACH, THEORETICAL MODEL FOR DIFFUSION BONDING, *MET. SCI.,* VOL 16, JAN 1982, P 49-56
- B. DERBY AND E.R. WALLACH, DIFFUSION BONDING: DEVELOPMENT OF THEORETICAL MODEL, *MET. SCI.,* VOL 18, 1984, P 427-431

- J.W. DINI, USE OF ELECTRODEPOSITION TO PROVIDE COATINGS FOR SOLID STATE BONDING, *WELD. J.,* VOL 61 (NO. 11), NOV 1982, P 33-39
- G. GARMONG, N.E. PATON, AND A.S. ARGON, ATTAINMENT OF FULL INTERFACIAL CONTACT DURING DIFFUSION BONDING, *MET. TRANS. A,* VOL 6A, JUNE 1975, P 1269-1279
- C.H. HAMILTON, PRESSURE REQUIREMENTS FOR DIFFUSION BONDING TITANIUM, *TITANIUM SCIENCE AND TECHNOLOGY,* R.I. JAFFEE AND H.M. BURTE, ED., PLENUM, 1973, P 625-647
- A. HILL AND E.R. WALLACH, MODELING SOLID STATE DIFFUSION BONDING, *ACTA. METALL.,* VOL 37 (NO. 9), SEPT 1989, P 2425-2437
- W.H. KING AND W.A. OWCZARSKI, ADDITIONAL STUDIES ON THE DIFFUSION WELDING OF TITANIUM, *WELD. RES. SUPP.,* OCT 1968, P 444S-450S
- H.A. MOHAMED AND J. WASHBURN, MECHANISM OF SOLID STATE PRESSURE WELDING, *WELD. RES. SUPP.,* SEPT 1975, P 302S-310S
- R. PEARCE, DIFFUSION BONDING, *PROC. INTERNATIONAL CONFERENCE ON DIFFUSION BONDING* (CRANFIELD, SIS), 1987
- J. PILLING, THE KINETICS OF ISOSTATIC DIFFUSION BONDING IN SUPERPLASTIC MATERIALS, *MATER. SCI. ENG.,* VOL 100, APRIL 1988, P 137-144
- E.R. WALLACH, SOLID-STATE DIFFUSION BONDING OF METALS, *TRANS. JWRI,* VOL 17.1, 1988, P 135-148
- D.S. WILKINSON AND M.F. ASHBY, PRESSURE SINTERING BY POWER LAW CREEP, *ACTA. METALL.,* VOL 23, 1975, P 1277-1285

Fundamentals of Explosion Welding

R. Alan Patterson, Los Alamos National Laboratory

Introduction

EXPLOSION WELDING (EXW), also known as explosive bonding, is accomplished by a high-velocity oblique impact between two metals. The impact must have sufficient energy to cause the colliding metal surfaces to flow hydrodynamically when they intimately contact one another in order to promote solid-state bonding. Oblique impact is important because conservation of momentum allows for a reentrant jetting action that is due to hydrodynamic flow of the faying metal surfaces. The jet is ejected outward from the collision apex between the metals and produces a cleaning action by scarfing or effacing the metal surfaces. The resulting virgin metal surfaces are then compressed together under high pressure from the explosion, which promotes atomistic bonding.

This phenomenon was first observed during World War I, when fragments from bombs or shells were often found to be welded to surrounding metal structures (Ref 1). The first publication to recognize the potential for generating metallic welding by using explosively driven plates appeared in 1944 (Ref 2). The first explosive bonding patent associated with the commercial application of the technique was issued in 1962 (Ref 3).

The most beneficial attribute of the technique is that dissimilar metal systems can be bonded even when conventional fusion welding techniques are metallurgically inappropriate, because of the formation of intermetallic compounds. Applications of explosive bonding are diverse and include the production of sandwiched metal for coinage (1965 to 1971), the more-sophisticated use of titanium-to-stainless steel transition joints in the Apollo spacecraft (Ref 4), and the use of aluminum-to-steel transition joints in ships. The most common utilization of explosive bonding is the production of clad metals for the purpose of corrosion resistance and for transition joints that are used to aid dissimilar metal welding. Practical applications of explosive bonding are addressed in greater detail in the article "Explosion Welding" and the Section "Procedure Development and Practice Considerations for Solid-State Welding" in this Volume.

It should be noted that the discussion in this article is restricted to the "parallel gap" explosive bonding technique. Several other bonding geometries that are fundamentally similar to the geometry of the parallel gap technique are utilized in a variety of other techniques. For example, the "preset angle" bonding technique is used to produce flat-plate clads.

Circumferential techniques are used to produce tube-to-tube welds or tubular clads. Spot welding and seam welding techniques, as well as others, are also used in certain situations. All of these techniques require the jet to travel in a straight line between the two metal surfaces (see the section "Flyer Plate Acceleration" in this article).

Several very good reviews of the explosive bonding process have been published (Ref 4, 5, 6, 7, 8). These articles should be used to augment the information presented here. This article attempts to describe the practice of producing an explosive bond/weld and draws on many previous research results in order to explain the mechanisms involved.

References

1. B. CROSSLAND, *EXPLOSIVE WELDING OF METALS AND ITS APPLICATION*, CLARENDON PRESS, 1982
2. L.R. CARL, *MET. PROGR.*, VOL 46, 1944, P 102
3. V. PHILIPCHUK AND F. LEROY BOIS, U.S. PATENT 3,024,526, 1962
4. V.D. LINSE, R.H. WITTMAN, AND R.J. CARLSON, "EXPLOSION BONDING," REPORT X68-10247, BATTELLE MEMORIAL INSTITUTE
5. J.S. RINEHART AND J. PEARSON, *EXPLOSIVE WORKING OF METALS*, PERGAMON PRESS, 1963
6. M.C. NOLAND, H.M. GADBERRY, J.B. LOSER, AND E.C. SNEEGAS, "HIGH VELOCITY METAL-WORKING: A SURVEY," NASA REPORT N67-26560, SP 5062, NASA, 1967
7. A.A. EZRA, *PRINCIPLES AND PRACTICE OF EXPLOSIVE METAL WORKING*, INDUSTRIAL NEWSPAPERS LIMITED, JOHN ADAMS HOUSE, LONDON, 1973
8. T.Z. BLAZYNSKI, *EXPLOSIVE WELDING, FORMING AND COMPACTION*, APPLIED SCIENCE PUBLISHERS, U.K., 1983

Fundamentals of Explosion Welding

R. Alan Patterson, Los Alamos National Laboratory

Bonding Practice

Figure 1 shows the arrangement used in the parallel gap explosive bonding process. The explosive charge is placed in contact with the top plate, hereafter called the flyer plate. The explosive detonation is initiated from one end to generate a linear detonation front, which runs along the flyer plate length. Pressure generated by the expansion of the explosive products accelerates the flyer plate downward at the point where the explosive is reacting and results in the desired impact on the stationary base plate. The principal parameters that affect bond success with this arrangement are:

- DETONATION VELOCITY AND ENERGY, WHICH ARE CHARACTERISTICS OF THE EXPLOSIVE USED AND RESULT IN FLYER PLATE ACCELERATION
- IMPACT ENERGY, WHICH IS A FUNCTION OF THE FLYER PLATE STANDOFF DISTANCE, THE DETONATION ENERGY, AND THE ENERGY TRANSFER CHARACTERISTICS
- DYNAMIC BEND ANGLE AND THE PHYSICAL AND MECHANICAL PROPERTIES OF THE METAL CONSTITUENTS

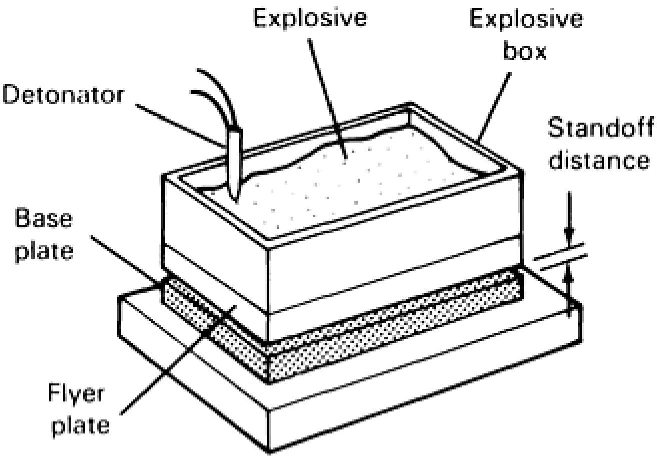
Explosive
Explosive box
Detonator
Standoff distance
Base plate
Flyer plate

FIG. 1 SCHEMATIC SHOWING KEY COMPONENTS USED IN THE PARALLEL GAP EXPLOSIVE WELDING PROCESS

Flyer Plate Acceleration

Figure 2, an intermediate view of the explosive bonding process, shows an idealization of the metal deformation that results after explosive detonation. Flash radiography performed during the bonding process has been used to develop this pictorial representation (Ref 6, 7). The energy produced by detonating the explosive results from a very rapid expansion of detonation products, which impacts an acceleration to the flyer plate and maintains a high-pressure region in the atmosphere behind the detonation front (Ref 1). Rapid flyer plate acceleration results in a dynamic bending action (Fig. 2). The dynamic bend angle, β, results in an oblique impact between the flyer and base plates, which promotes the hydrodynamic flow of the metal surfaces and a resulting jetting actions.

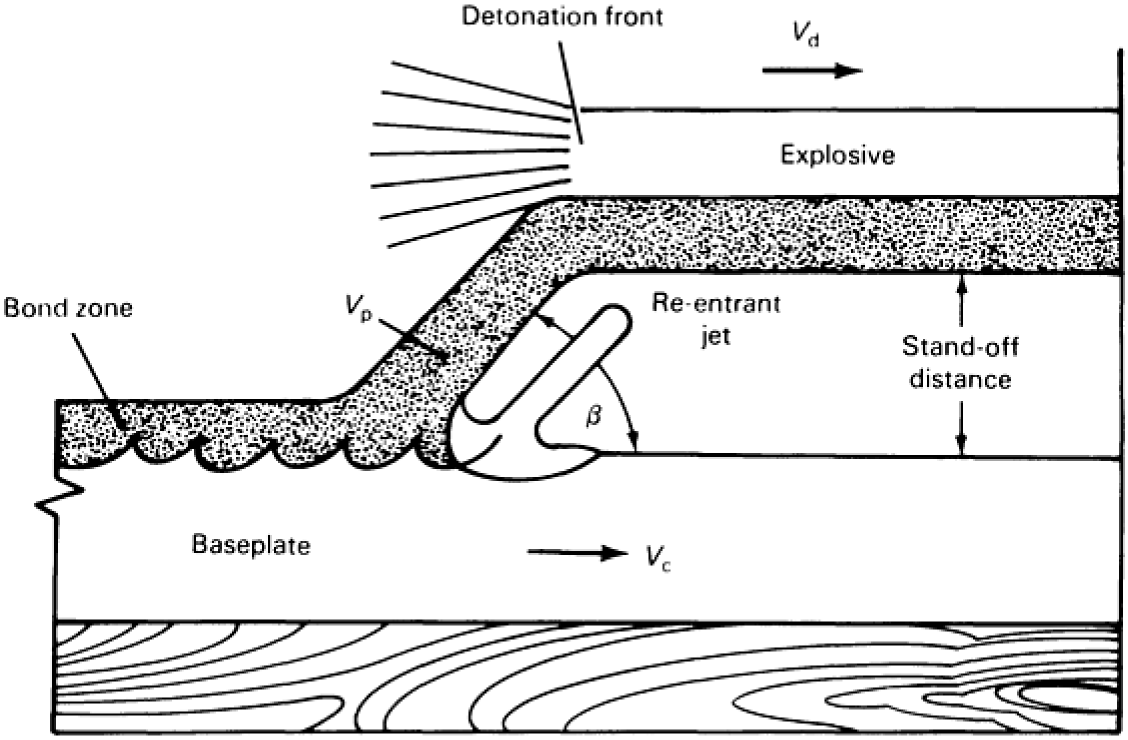
Detonation front
V_d
Explosive
Bond zone
V_p
Re-entrant jet
Stand-off distance
β
Baseplate
V_c

FIG. 2 SCHEMATIC SHOWING AN IDEALIZED VIEW OF THE FLYER PLATE DEFORMATION AND IMPACT RESULTING FROM A MOVING EXPLOSIVE DETONATION FRONT

Conservation of momentum is used to describe the bonding mechanism in which three velocity vectors are of primary importance. The detonation velocity, V_d, is a characteristic of the explosive and is generally directly proportional to the explosive density. From the geometry of the system, it is obvious that the collision point between the impacting metals propagates at the same speed as the explosive detonation front. The collision-point velocity is represented by V_c in Fig. 2. The flyer plate velocity, V_p, can be calculated based on an energy balance and conservation of momentum (Ref 9).

Impact Energy

Impact energy is generally assumed to equal the maximum kinetic energy that is due to the acceleration of the flyer plate. In simpler terms, it is assumed that the flyer plate has been accelerated to the terminal velocity as calculated from models like that proposed by Gurney (Ref 9), and that no deceleration occurs because of atmospheric interference. Furthermore, energy loss that is due to the bending strain energy in the flyer plate has been ignored in much of the earlier literature, because of its complexity. Recent estimates of energy partitioning suggest that only 15 to 20% of the flyer plate energy is used for cladding, whereas 66% is absorbed in flyer plate forming and 15% is dissipated as excess energy in the base plate (Ref 10, 11). The flyer plate standoff distance is therefore a very important factor, because it physically limits the flyer plate acceleration and dynamic bend angle.

Simple determinations of flyer plate standoff distance resulting from empirical testing are often used. Carpenter *et al.* (Ref 10) state that the standoff distance should be small enough to allow continuous acceleration of the flyer plate. Standoff distances that are twice the flyer plate thickness for thin components (up to 6.5 mm, or 0.26 in.) and are equal to the flyer plate thickness for thicker components (up to 13 mm, or 0.5 in.) are generally used (Ref 12). Other studies have indicated that a uniform flyer plate velocity is obtained with a separation distance equal to only a few multiples of the flyer plate thickness (Ref 13, 14, 15). Wolf *et al.* (Ref 16) performed a systematic evaluation of standoff distance and dynamic bend angle, which indicated that:

- A LAMINAR BOND WAS ACHIEVED AT DISTANCES EQUAL TO APPROXIMATELY HALF THE FLYER PLATE THICKNESS
- A STABLE WAVY BOND ZONE WAS PRODUCED AT INTERMEDIATE STANDOFF DISTANCES EQUAL TO 1.0 TO 1.5 TIMES THE FLYER PLATE THICKNESS
- DELAMINATION OCCURRED AT DISTANCES ABOVE TWO TIMES THE FLYER PLATE THICKNESS

Dynamic Bend Angle and Material Properties

Attempts to correlate bonding success to starting material properties have been extremely difficult, because most material property compilations are determined at ambient conditions and do not account for the high strain rates associated with explosive conditions. Rules of thumb often dominate. Carpenter *et al.* (Ref 10) indicated that materials exhibiting tensile ductilities greater than 5% will survive the bending associated with the flyer plate acceleration. It is also generally recognized that the material with the lowest density and tensile yield strength should be used as the flyer plate material. However, brittle materials (for example, refractory metals such as molybdenum and tungsten) have been explosively bonded by elevating the metal temperature above the ductile-to-brittle transition temperature (DBTT) (Ref 17).

Some correlations between initial material properties and the minimum dynamic bend angle, $\beta_{\min}$, that is required to achieve the desired shear deformation to promote jetting and bonding success have been reported (Ref 18, 19). The suggested relationship between flyer plate hardness, H_v; density, ρ; and the collision-point velocity, V_c, is given by:

$$b_{\min} = k\left(\frac{H_v}{rV^2{}_c}\right) \qquad \textbf{(EQ 1)}$$

References cited in this section

1. B. CROSSLAND, *EXPLOSIVE WELDING OF METALS AND ITS APPLICATION,* CLARENDON PRESS, 1982
6. M.C. NOLAND, H.M. GADBERRY, J.B. LOSER, AND E.C. SNEEGAS, "HIGH VELOCITY METAL-

WORKING: A SURVEY," NASA REPORT N67-26560, SP 5062, NASA, 1967

7. A.A. EZRA, *PRINCIPLES AND PRACTICE OF EXPLOSIVE METAL WORKING,* INDUSTRIAL NEWSPAPERS LIMITED, JOHN ADAMS HOUSE, LONDON, 1973
9. R.W. GURNEY, "THE INITIAL VELOCITIES OF FRAGMENTS FROM BOMBS, SHELLS, AND GRENADES," REPORT 405, BALLISTIC RESEARCH LABORATORY, 1943
10. S. CARPENTER, R.H. WITTMAN, AND R.J. CARLSON, RELATIONSHIPS OF EXPLOSIVE WELDING PARAMETERS TO MATERIAL PROPERTIES AND GEOMETRY FACTORS, *PROC. FIRST INT. CONF. OF THE CENTER FOR HIGH ENERGY FORMING,* UNIVERSITY OF DENVER, JUNE 1967, P 124
11. B. CROSSLAND AND J.D. WILLIAMS, *METALL. REV.,* NO. 144, 1970
12. H. WOLF, ON THE ENERGY BALANCE OF PLANE AND CYLINDRICAL EXPLOSIVE CLADDING, *PROC. INT. SYMPOSIUM ON INTENSE DYNAMIC LOADING AND ITS EFFECTS,* PERGAMON PRESS, 1988
13. J. GROSCHOPP, V. HEYNE, AND B. HOFFMAN, EXPLOSIVELY CLAD TITANIUM STEEL COMPOSITE, *WELD. INT.,* NO. 9, 1987, P 879-883
14. R.A. PATTERSON, EXPLOSION BONDING: ALUMINUM-MAGNESIUM ALLOYS BONDED TO AUSTENITIC STAINLESS STEEL, *HIGH ENERGY RATE FABRICATION,* M.A. MEYERS AND J.W. SCHROEDER, ED., VOL 70, ASME PRESSURE VESSELS AND PIPING DIV., 1982, P 15-27
15. M.D. CHADWICK, D. HOWD, G. WILDSMITH, AND J.H. CAIRNS, *BR. WELD. J.,* VOL 15, 1968, P 480
16. H. WOLF AND M. KOLBE, EXPERIMENTAL METHODS FOR DETERMINING THE OPTIMUM PROCESS CHARACTERISTICS FOR FLAT EXPLOSIVE CLADDING, *WELD. INT.,* VOL 4 (NO. 7), 1990, P 535-538
17. A.A. POPOFF AND H. CASEY, TECHNICAL PAPER AD 77-36, SME, 1977
18. V.G. PETUSHKOV, THE STRENGTH OF BODIES STRIKING ONE ANOTHER DURING EXPLOSIVE WELDING, *AUTO. WELD.,* VOL 39 (NO. 10), 1986, P 35-38
19. V.M. KUDINOV AND I.D. ZAKHAZENKO, CRITERIA FOR SELECTING THE PARAMETERS OF EXPLOSIVE WELDING, *WELD. PROD.,* VOL 32 (NO. 9), 1985, P 3-6

Fundamentals of Explosion Welding

R. Alan Patterson, Los Alamos National Laboratory

Fundamentals of Bonding

Several important concepts pertaining to the explosive parameters, hydrodynamic flow, jetting, and metal properties were discussed above. The following discussion summarizes the criteria used to model the explosive bonding process.

Explosive Parameters and Shock Effects

Although many different types of explosives have been used for explosive bonding, those that have a broad range of low-detonation velocities are the most appropriate, because fo the deleterious effects of shock rarefaction. Shock fronts are described by a sharp discontinuity in pressure on one side of the transmitting medium where the violent motion that is due to high pressure is separated from a near-quiescent state in the unaffected material (Ref 1). This condition can produce tensile stresses of sufficient magnitude to fracture the explosive bond. Shock waves are not pronounced unless the velocity of the shock front exceeds the sonic velocity of the transmitting medium. Sonic velocity is defined as:

$$U_s = \sqrt{\frac{E'}{r}} \qquad \textbf{(EQ 2)}$$

where U_s is the sonic velocity, E is the appropriate elastic modulus, and ρ is the material density. If the shock wave propagation does not exceed the sonic velocity, then a broadening of the shock front occurs and the magnitude of resultant stress and stress gradients is greatly reduced. Because most metals have a sonic velocity that ranges from 2000 to 6000 m/s (6600 to 20,000 ft/s), it is desirable to use explosives with detonation velocities within this range. Carpenter *et al.* (Ref 10) indicate that explosives with detonation velocities greater than 120% of the sonic velocity of the metal should not be used.

Detonation velocity is a characteristic of the type of explosive and has been shown to be directly proportional to the explosive density. This proportionally is shown in Eq 3, as derived empirically for nitroguanidine explosive (Ref 13):

$$V_D = 1440 + 4020\rho_E \qquad \textbf{(EQ 3)}$$

where V_d equals the detonation velocity and ρ_e is the explosive density. The typical operating range for nitroguanidine is from 2000 to 5000 m/s (6600 to 16,000 ft/s), which corresponds to explosive densities that range from 0.14 to 0.9 g/cm^3 (0.0051 to 0.033 lb/in.^{3}).

The explosive pressure, P, is also proportional to the explosive density, ρ_e, as shown by the Bernoulli equation:

$$P \propto V_d^2 \rho_E \qquad \textbf{(EQ 4)}$$

Flyer plate velocities can be calculated from models like that presented by Gurney (Ref 9), who assumed that the driven metal was accelerated perpendicular to the direction of the detonation propagation and used a specific energy with a characteristic value for each explosive (the Gurney energy). Kennedy (Ref 20) gave a very good explanation of the Gurney model, along with several reconstructions of the Gurney equation, to account for changes in geometry. The equation that is applicable to the parallel gap explosive bonding technique is:

$$V_p^2 = 2E'\frac{3}{\left[1+5+\left(\frac{m}{c}\right)+4\left(\frac{m^2}{c^2}\right)\right]} \qquad \textbf{(EQ 5)}$$

where V_p is the flyer plate velocity, E' is the Gurney energy, m is the flyer plate mass, c is the explosive mass, and m/c is the explosive load factor.

Once the flyer plate and detonation velocities are known, it is a simple matter to define the dynamic nature of the flyer plate bend angle through geometry. The following equation holds for the parallel gap geometry illustrated in Fig. 1 and 2:

$$V_p = 2V_d \sin\left(\frac{b}{2}\right) \qquad \textbf{(EQ 6)}$$

where V_p and V_d are the velocities defined in Eq 5 and 6, respectively, and β is the flyer plate dynamic bend angle.

Equations 1, 2, 3, 4, 5, and 6, as well as an understanding of the metal physical-mechanical properties, are the basis for most models developed for explosive bonding. These models are discussed below in general terms, beginning with a description of jetting phenomena, followed by a description of bonding and the methods used to define explosive parameter boundaries that will produce acceptable bonds.

Jet Formation

During World War II, it was discovered that a hollowed explosive charge lined with metal could be used to produce a high-energy jet to perforate armor plate. Several investigators (Ref 21, 22, 23) have described the limiting conditions for the formation of jets in high-velocity collisions. Walsh *et al.* (Ref 22) indicated that a critical impact angle must be exceeded before jetting will occur. On the other hand, Cowan and Holtzman (Ref 23) indicated that for subsonic collisions, the elastic strength of the metal must be exceeded before jetting will occur. In both cases, jetting is described by a hydrodynamic flow of the metal surfaces. Walsh *et al.* (Ref 22) presented equations based on this principle and the "equations of state," which enable a calculation of the critical angle for jetting.

Kowalick and Hay (Ref 24) reported an experimental technique to evaluate both the critical angle for jetting and the morphological changes in explosion bonds resulting from changes in collision angle. The experimental setup consisted of the normal flyer plate arrangement used in the parallel gap technique, but with the base plate replaced by a half cylinder. Each bonding trial produced continuously changing flayer plate collision angles that resulted in surface deformation and bond morphology variations. This analysis expands upon the work of previous researchers by evaluating the geometry of fluid flow, jetting, and their effects on bond morphology.

Laminar fluid flow (low collision-point velocity and low Reynolds numbers) was correlated with direct bonding and a very flat bond interface. Turbulent fluid flow was separated into two regimes:

- AT INTERMEDIATE COLLISION-POINT VELOCITIES, THE FLOW WAS DESCRIBED BY VORTICE FORMATION INTO A VON KARMAN VORTEX STREET AND THE PRODUCTION OF A WAVY BOND INTERFACE
- AT HIGH COLLISION-POINT VELOCITIES, THE FLOW WAS DISTURBED BY A DISORDERING OF THE JET WAKE, WHICH LED TO A MOLTEN LAYER BOND INTERFACE.

Selection of Bonding Parameters

The preceding discussion introduced the primary variables used to predict explosive bonding parameters. The following four conditions are used to define a window of acceptable bonding parameters:

- CRITICAL ANGLE FOR JETTING (REF 25)
- CRITICAL FLOW TRANSITION VELOCITY (REF 23, 24)
- MINIMUM FLYER PLATE IMPACT FOR JETTING, BASED ON ANALYSIS OF THE STRESS REQUIRED TO CAUSE FLOW
- MAXIMUM FLYER PLATE IMPACT FOR THE LIMITATION OF WELD DEFECTS, BASED ON JET ENTRAPMENT (REF 23)

The critical angle for jetting is used to define the upper boundary for collision-point velocities. The lower limit in collision velocity is determined by the transition from laminar to turbulent flow, where turbulent flow is preferred (Ref 23). The other two limiting conditions are specified by using the physical and mechanical properties of the material to determine the maximum amount of energy required to either produce hydrodynamic flow (that is, overcome a critical shear stress) or avoid extensive melting. Once the correct set of explosive parameters has been determined by utilizing either predictive models or an empirical approach, the production of explosively welded metal is a relatively simple task.

References cited in this section

1. B. CROSSLAND, *EXPLOSIVE WELDING OF METALS AND ITS APPLICATION,* CLARENDON PRESS, 1982
9. R.W. GURNEY, "THE INITIAL VELOCITIES OF FRAGMENTS FROM BOMBS, SHELLS, AND GRENADES," REPORT 405, BALLISTIC RESEARCH LABORATORY, 1943
10. S. CARPENTER, R.H. WITTMAN, AND R.J. CARLSON, RELATIONSHIPS OF EXPLOSIVE WELDING PARAMETERS TO MATERIAL PROPERTIES AND GEOMETRY FACTORS, *PROC. FIRST INT. CONF. OF THE CENTER FOR HIGH ENERGY FORMING,* UNIVERSITY OF DENVER, JUNE 1967, P 124
13. J. GROSCHOPP, V. HEYNE, AND B. HOFFMAN, EXPLOSIVELY CLAD TITANIUM STEEL COMPOSITE, *WELD. INT.,* NO. 9, 1987, P 879-883
20. J.E. KENNEDY, EXPLOSIVE OUTPUT FOR DRIVING METAL, *PROC. BEHAVIOR AND UTILIZATION OF EXPLOSIVES IN ENGINEERING DESIGN CONF.,* L. DAVIDSON, J.E. KENNEDY,

AND F. COFFEY, ED., ASME (NEW MEXICO SECTION), 1972, P 109
21. G. BIRKOFF, D.P. MACDOUGALL, E.M. PUGH, AND SIR G. TAYLOR, *J. APPL. PHYS.*, VOL 19, 1948, P 563
22. J.M. WALSH, R.G. SHEFFLER, AND F.J. WILLIG, *J. APPL. PHYS.*, VOL 24 (NO. 3), 1953 P 349
23. G.R. COWAN AND A.H. HOLTZMAN, *J. APPL. PHYS.*, VOL 34 (NO. 4), 1963, P 928
24. J.F. KOWALICK AND D.R. HAY, *METALL. TRANS.*, VOL 2, 1971, P 1953
25. R.H. WITTMAN, "AN EXPERIMENTALLY VERIFIED MODEL PREDICTING IMPACT WELDING PARAMETERS," THESIS, UNIVERSITY OF DENVER, 1975

Fundamentals of Explosion Welding

R. Alan Patterson, Los Alamos National Laboratory

Bond Morphology and Properties

In the discussion of the Kowalick and Hay (Ref 24) experiment, reference was made to three types of bond morphologies:

- LAMINAR INTERFACE, OR FLAT BOND
- WAVY INTERFACE
- MOLTEN LAYER INTERFACE

A typical clad trilayer that contains a mixture of wavy interface morphology (with melt pockets decorating each wave crest) and laminar interface (with fairly uniform melt entrapment) is shown in Fig. 3.

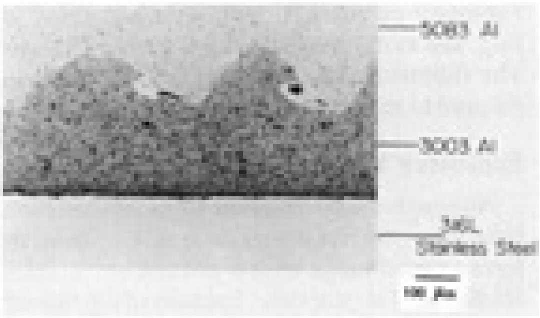

FIG. 3 CROSS-SECTIONAL PHOTOMICROGRAPH OF A 5083 AL-3003 AL-TYPE 316L STAINLESS STEEL CLAD TRI-LAYER TAKEN PARALLEL TO THE BONDING DIRECTION. NOTICE THE WELL-DEVELOPED WAVE MORPHOLOGY AND DARK ETCHING MELT POCKETS IN THE ALUMINUM-TO-ALUMINUM BOND. IN ADDITION, THE LAMINAR ALUMINUM-TO-STAINLESS STEEL BOND INTERFACE INDICATES INTERMITTENT SECOND-PHASE ENTRAPMENT.

Wave Formation. Abrahamson (Ref 26) examined the formation of surface deformations that were caused by a traveling jet using projectiles shot at thin metal targets, resulting in oblique impact. Two primary conclusions drawn from this work were that the target metal deformation was a horizontal shear in the direction of impact and that the resultant wave peaks were deformed in such a pattern that they turned back in the direction opposite to impact. When the angle of impact was varied, two important responses occurred. First, below an impact angle of 5°, no wave formation was observed. Second, as the angle of impact was increased above 5°, the amplitude and wavelength increased.

The proposed mechanism for wave formation in high-velocity oblique impacts was based on subsonic or supersonic metal flow. Good bonding was achieved with low-impact angles, because of a need to prevent jet instabilities.

Baharani *et al.* (Ref 27) and Hunt (Ref 28) drew upon this hypothesis and proposed a physical description of wave formation in explosive bonding, which included the following sequence of events:

- *STEP 1:* IMPACT PRODUCES SHEAR DEFORMATION IN THE STATIONARY BASE PLATE, WHICH RESULTS IN A DEPRESSION
- *STEP 2:* CONSERVATION OF VOLUME CRITERIA PRODUCES AN UPHEAVAL OF METAL (A HUMP) AHEAD OF THE IMPACT APEX
- *STEP 3:* HUMP INTERFERES WITH THE JET FLOW AND PRODUCES AN EDDY IN THE JET, WHICH TRAPS MOLTEN MATERIAL IN THE REAR VORTEX
- *STEP 4:* ALLOWING FOR COLLISION-POINT VELOCITY CAUSES A FORWARD DEFORMATION OF THE HUMP AND FURTHER JET TURBULENCE, WHICH AGAIN CAUSES JET ENTRAPMENT IN THE FRONT VORTEX
- *STEP 5:* PROCESS IS THEN REPEATED, STARTING WITH STEP 1

Lucas *et al.* (Ref 29) reported that when the flyer and base plates of a copper-to-copper explosive bond were electroplated with 1 μm (40 μin.) of nickel, the wave vortices were found to contain free nickel, which was assumed to result from jet entrapment caused by turbulence. Trueb (Ref 30, 31) indicated a similar phenomenon when bonding titanium to steel. The vortices appeared to contain a mix of both metals arranged in an eddy pattern that was "frozen in" by rapid solidification.

Bond Microstructure. Microstructural analysis of explosive bond zones has revealed several interesting results. Numerous metallurgical reactions have been reported:

- EXTENSIVE GRAIN ELONGATION IN THE DIRECTION OF BONDING, WITHIN SEVERAL HUNDREDS OF MICRONS OF THE BOND INTERFACE, AND MEASURED ELONGATIONS OF SEVERAL THOUSAND PERCENT (REF 31)
- RESIDUAL DISLOCATION DENSITIES ON THE ORDER OF $10^{11}/\text{CM}^2$ NEAR THE BOND INTERFACE (REF 31)
- MICROTWINNING HAS BEEN OBSERVED (REF 31, 32) IN LOW STACKING FAULT ENERGY METALS (FOR EXAMPLE, COPPER)
- RECRYSTALLIZATION AND RECOVERY HAVE BEEN OBSERVED IN BANDS ON BOTH SIDES OF HIGH-DISLOCATION-DENSITY REGIONS (REF 29)
- VERY LITTLE DIFFUSION OCCURS ACROSS THE BOND INTERFACE (REF 29)

These observations support the hypothesis that says that wave formation is a result of extreme metal deformation and that bonding is a result of solid-state processes with very little diffusion.

However, Hammerschmidt and Kreye (Ref 33) have presented an alternative mechanism for bonding based on the formation of a submicroscopic layer of molten metal, rather than solid-state bonding. Their investigation revealed that in bonds between copper and aluminum, a layer up to 5 μm (200 μin.) wide was obviously molten and resolidified as a fine-grained equiaxed structure. This conclusion raises several questions as to the appropriateness of dissimilar metal systems for explosive bonding, especially when brittle intermetallic compounds can result. Similarly, jet entrapment must be considered deleterious with respect to intermetallic compound formation. Therefore, close attention must be given to the metallurgy of the dissimilar metal couple and bonding parameters when high-strength explosive bonds are required.

The preceding discussion indicates that a very complex combination of metallurgical reactions occur during explosive bonding. These effects have been researched to a limited degree and the general results appear to be understood, but controversy over actual bonding mechanisms still remains.

Oberg *et al.* (Ref 34) used the Miedema model for alloy formation to explain the metallurgical details of explosion bonds. In their analysis, they suggest strong correlations between interfacial bond strength and the formation energies between binary alloy constituents. If the theoretical formation energy, γ_{AB}, is high, then a low bonding success would be predicted, indicating that low formation energies are preferred. Also, if the chemical portion of the interfacial free energy, g_{AB}^{chem}, deviates from theoretical mixing in a negative fashion, then the formation of intermetallic compounds would be favored. Positive deviations indicate the tendency for solid-solution formation and strong bonding.

Explosive bond strength is often negatively affected by the formation of intermetallics and defects that result from rapid quenching of melt regions. Counteracting these effects is the localized hardening and high dislocation densities that occur in a heavily worked structure adjacent to the bond interface. In fact, bonds produced by extreme metal deformations that result in a wavy interface have been reported to be optimum, with respect to strength (Ref 35).

Bond Strength Determination. Bond strength can be measured by a variety of mechanical testing techniques, including the simple bend and shear test identified in the ANSI/ASTM standard specifications A 263, A 264, A 265, and B 432. A special "top hat" tensile test has also been utilized (Fig. 4).

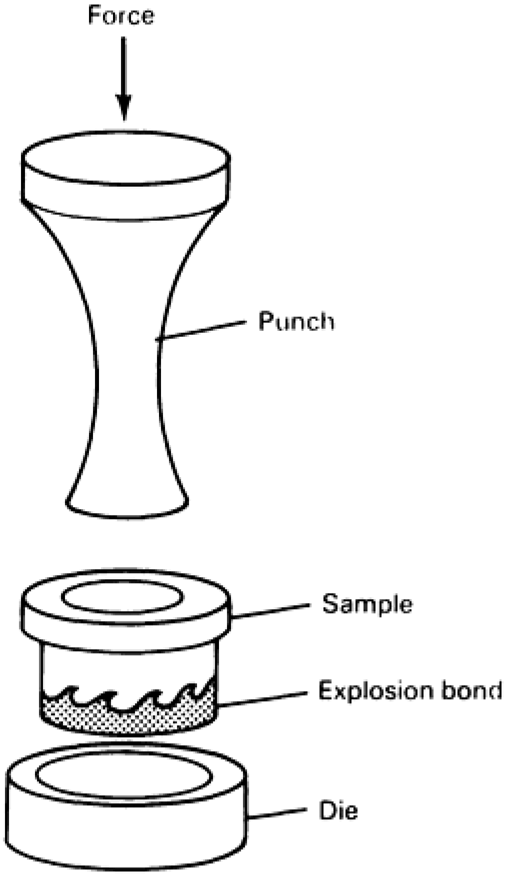

FIG. 4 SCHEMATIC SHOWING TOP-HAT TENSILE TEST APPARATUS AND RELATION OF SAMPLE TO PUNCH-DIE TOOLING COMPONENTS

Figure 5 is a cross-sectional view of a copper-to-stainless steel clad tensile specimen, where failure occurred in the copper base metal (Ref 13). Note the wavy bond interface to the right of the necked fracture zone (Fig. 5a) and the normal dimpled appearance of the copper fracture surface as revealed by the scanning electron microscope (Fig. 5b). This particular sample agrees with accepted bond strength trends, because the well-developed wavy bond performs better than the weaker of the two parent materials.

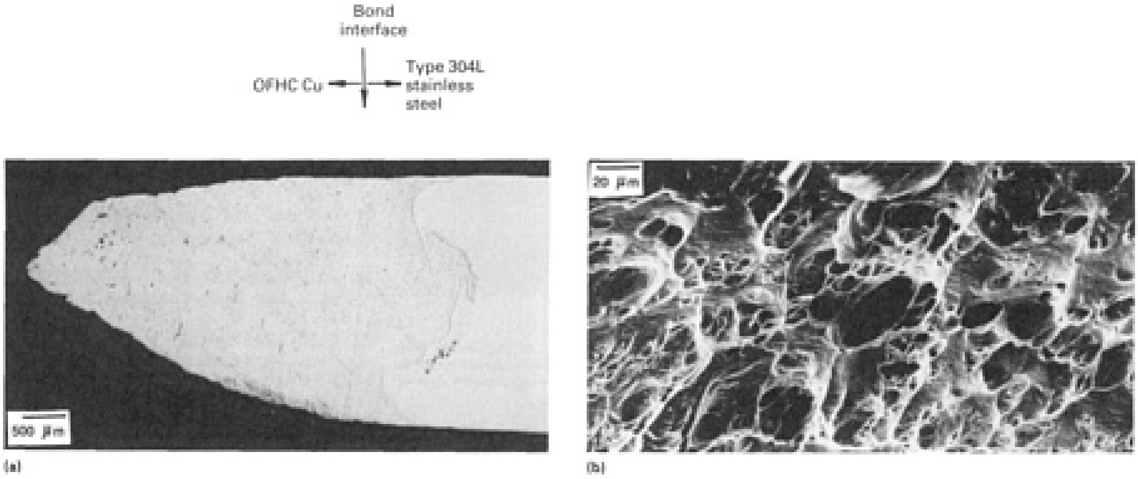

FIG. 5 FAILURE IN COPPER-TO-STAINLESS STEEL TENSILE SPECIMEN THAT WAS EXPLOSION WELDED. (A) PHOTOMICROGRAPH OF FAILED SPECIMEN SHOWING WAVY BOND LINE AND DUCTILE FRACTURE IN PARENT COPPER METAL. 16×. (B) SCANNING ELECTRON FRACTOGRAPH OF COPPER FRACTURE SHOWING DIMPLED APPEARANCE OF TYPICAL DUCTILE TENSILE OVERLOAD FAILURE. 395×

Figure 6, on the other hand, shows scanning electron fractographs of a tensile specimen that failed in the explosive bond interface. This tri-layer specimen (a bond between 5083 A 1-5052 Al-type 304L stainless steel) illustrates one important fact: wavy interface morphologies are not always the strongest (Ref 13). The principal fracture occurred in the wavy interface between the 5083 Al and the 5052 Al layers, whereas the laminar interface between the 5052 Al and the type 304L stainless steel exhibited an incipient fracture appearance with unconnected cracking along the interface. This is a direct contradiction to the previous statements that a wavy bond increases strength and further illustrates the need to evaluate each clad specimen using destructive test methods.

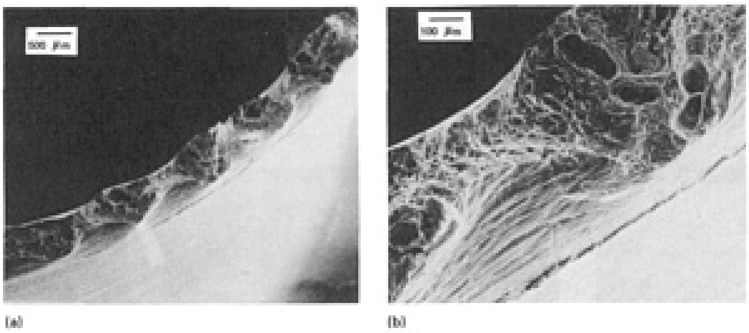

FIG. 6 SCANNING ELECTRON FRACTOGRAPHS OF FRACTURE SPECIMEN FROM CLAD TRILAYER OF 5083 AL-5052 AL-TYPE 304L STAINLESS STEEL. (A) FRACTURE OF 5083 AL-5052 AL BOND INTERFACE, SHOWING DUCTILE TENSILE OVERLOAD CHARACTERISTICS AND INCIPIENT FRACTURE IN THE 5052 AL-TYPE 304L STAINLESS STEEL BOND. 65×. (B) LOW-MAGNIFICATION VIEW OF INTERFACE SHOWING THAT FRACTURE PATH FOLLOWS WAVE MORPHOLOGY. 16×

References cited in this section

13. J. GROSCHOPP, V. HEYNE, AND B. HOFFMAN, EXPLOSIVELY CLAD TITANIUM STEEL COMPOSITE, *WELD. INT.*, NO. 9, 1987, P 879-883
24. J.F. KOWALICK AND D.R. HAY, *METALL. TRANS.*, VOL 2, 1971, P 1953

26. G.R. ABRAHAMSON, *J. APPL. PHYS.,* VOL 34 (NO. 4), 1961, P 519
27. A.S. BAHARANI, T.J. BLACK, AND B. CROSSLAND, *PROC. R. SOC. (LONDON) A,* VOL 296, 1967, P 123
28. J.N. HUNT, *PHILOS. MAG.,* VOL 18, 1968, P 669
29. W. LUCAS, J.D. WILLIAMS, AND B. CROSSLAND, SOME METALLURGICAL OBSERVATIONS ON EXPLOSIVE WELDING, *PROC. SECOND INT. CONF. OF CENTER FOR HIGH ENERGY FORMING,* 1969, P 8.1.1
30. L.F. TRUEB, THE MICROSTRUCTURE OF EXPLOSION-BONDED METALS, *TRANS. AIME,* VOL 242, 1966, P 1057
31. L.F. TRUEB, *J. APPL. PHYS.,* VOL 34 (NO. 4), 1968, P 928
32. M.P. BONDAR AND V.M. OGOLIKHIN, PLASTIC DEFORMATION IN BONDING ZONE UNDER EXPLOSIVE WELDING AND ITS ROLE IN BONDING FORMATION, *J. PHYS.,* C5 (NO. 8-46), 1985, P C5-379 TO C5-384
33. M. HAMMERSCHMIDT AND H. KREYE, MICROSTRUCTURE AND BONDING MECHANISMS IN EXPLOSION WELDING, *SHOCK WAVES AND HIGH-STRAIN RATE PHENOMENA IN METALS,* M.A. MEYERS AND L.E. MURR, ED., PLENUM PRESS, 1981, P 961
34. A. OBERG, N. MARTENSSON, AND J.A. SCHWEITZ, FUNDAMENTAL ASPECTS OF FORMATION AND STABILITY OF EXPLOSIVE WELDS, *METALL. TRANS. A,* VOL 16, MAY 1985, P 841-852
35. H.K. BALAKRISCHMA, V.C. VENTATESH, AND P.K. PHILIP, INFLUENCE OF COLLISION PARAMETERS ON THE MORPHOLOGY OF INTERFACE IN ALUMINUM-STEEL EXPLOSION WELDS, *SHOCK WAVES AND HIGH-STRAIN RATE PHENOMENA IN METALS,* M.A. MEYERS AND L.E. MURR, ED., PLENUM PRESS, 1981, P 975

Fundamentals of Explosion Welding

R. Alan Patterson, Los Alamos National Laboratory

Future Outlook

In this discussion, brief descriptions of bonding techniques and bonding parameters have been correlated with current experimental results. As shown, attempts to gain a fundamental understanding of explosive bonding involve a large number of highly interactive variables (for example, explosive detonation dynamics, dynamic materials deformation, and solid-state bonding). May of these interactive variables have been defined in this discussion, and the references cited offer additional detail and explanation. However, much of the available knowledge relating to explosive bonding is still based on experimental observation. Models and predictive algorithms have been published, but to a large extent bonding is still considered an art practiced by a rather select group of experienced technologists. Insight into the practice of producing explosive bonds and an introduction to the specific applications is available in the articles "Explosion Welding" and "Procedure Development and Practice Considerations for Explosion Welding," in this Volume.

Fundamentals of Explosion Welding

R. Alan Patterson, Los Alamos National Laboratory

References

1. B. CROSSLAND, *EXPLOSIVE WELDING OF METALS AND ITS APPLICATION,* CLARENDON PRESS, 1982
2. L.R. CARL, *MET. PROGR.,* VOL 46, 1944, P 102
3. V. PHILIPCHUK AND F. LEROY BOIS, U.S. PATENT 3,024,526, 1962

4. V.D. LINSE, R.H. WITTMAN, AND R.J. CARLSON, "EXPLOSION BONDING," REPORT X68-10247, BATTELLE MEMORIAL INSTITUTE
5. J.S. RINEHART AND J. PEARSON, *EXPLOSIVE WORKING OF METALS,* PERGAMON PRESS, 1963
6. M.C. NOLAND, H.M. GADBERRY, J.B. LOSER, AND E.C. SNEEGAS, "HIGH VELOCITY METAL-WORKING: A SURVEY," NASA REPORT N67-26560, SP 5062, NASA, 1967
7. A.A. EZRA, *PRINCIPLES AND PRACTICE OF EXPLOSIVE METAL WORKING,* INDUSTRIAL NEWSPAPERS LIMITED, JOHN ADAMS HOUSE, LONDON, 1973
8. T.Z. BLAZYNSKI, *EXPLOSIVE WELDING, FORMING AND COMPACTION,* APPLIED SCIENCE PUBLISHERS, U.K., 1983
9. R.W. GURNEY, "THE INITIAL VELOCITIES OF FRAGMENTS FROM BOMBS, SHELLS, AND GRENADES," REPORT 405, BALLISTIC RESEARCH LABORATORY, 1943
10. S. CARPENTER, R.H. WITTMAN, AND R.J. CARLSON, RELATIONSHIPS OF EXPLOSIVE WELDING PARAMETERS TO MATERIAL PROPERTIES AND GEOMETRY FACTORS, *PROC. FIRST INT. CONF. OF THE CENTER FOR HIGH ENERGY FORMING,* UNIVERSITY OF DENVER, JUNE 1967, P 124
11. B. CROSSLAND AND J.D. WILLIAMS, *METALL. REV.,* NO. 144, 1970
12. H. WOLF, ON THE ENERGY BALANCE OF PLANE AND CYLINDRICAL EXPLOSIVE CLADDING, *PROC. INT. SYMPOSIUM ON INTENSE DYNAMIC LOADING AND ITS EFFECTS,* PERGAMON PRESS, 1988
13. J. GROSCHOPP, V. HEYNE, AND B. HOFFMAN, EXPLOSIVELY CLAD TITANIUM STEEL COMPOSITE, *WELD. INT.,* NO. 9, 1987, P 879-883
14. R.A. PATTERSON, EXPLOSION BONDING: ALUMINUM-MAGNESIUM ALLOYS BONDED TO AUSTENITIC STAINLESS STEEL, *HIGH ENERGY RATE FABRICATION,* M.A. MEYERS AND J.W. SCHROEDER, ED., VOL 70, ASME PRESSURE VESSELS AND PIPING DIV., 1982, P 15-27
15. M.D. CHADWICK, D. HOWD, G. WILDSMITH, AND J.H. CAIRNS, *BR. WELD. J.,* VOL 15, 1968, P 480
16. H. WOLF AND M. KOLBE, EXPERIMENTAL METHODS FOR DETERMINING THE OPTIMUM PROCESS CHARACTERISTICS FOR FLAT EXPLOSIVE CLADDING, *WELD. INT.,* VOL 4 (NO. 7), 1990, P 535-538
17. A.A. POPOFF AND H. CASEY, TECHNICAL PAPER AD 77-36, SME, 1977
18. V.G. PETUSHKOV, THE STRENGTH OF BODIES STRIKING ONE ANOTHER DURING EXPLOSIVE WELDING, *AUTO. WELD.,* VOL 39 (NO. 10), 1986, P 35-38
19. V.M. KUDINOV AND I.D. ZAKHAZENKO, CRITERIA FOR SELECTING THE PARAMETERS OF EXPLOSIVE WELDING, *WELD. PROD.,* VOL 32 (NO. 9), 1985, P 3-6
20. J.E. KENNEDY, EXPLOSIVE OUTPUT FOR DRIVING METAL, *PROC. BEHAVIOR AND UTILIZATION OF EXPLOSIVES IN ENGINEERING DESIGN CONF.,* L. DAVIDSON, J.E. KENNEDY, AND F. COFFEY, ED., ASME (NEW MEXICO SECTION), 1972, P 109
21. G. BIRKOFF, D.P. MACDOUGALL, E.M. PUGH, AND SIR G. TAYLOR, *J. APPL. PHYS.,* VOL 19, 1948, P 563
22. J.M. WALSH, R.G. SHEFFLER, AND F.J. WILLIG, *J. APPL. PHYS.,* VOL 24 (NO. 3), 1953 P 349
23. G.R. COWAN AND A.H. HOLTZMAN, *J. APPL. PHYS.,* VOL 34 (NO. 4), 1963, P 928
24. J.F. KOWALICK AND D.R. HAY, *METALL. TRANS.,* VOL 2, 1971, P 1953
25. R.H. WITTMAN, "AN EXPERIMENTALLY VERIFIED MODEL PREDICTING IMPACT WELDING PARAMETERS," THESIS, UNIVERSITY OF DENVER, 1975
26. G.R. ABRAHAMSON, *J. APPL. PHYS.,* VOL 34 (NO. 4), 1961, P 519
27. A.S. BAHARANI, T.J. BLACK, AND B. CROSSLAND, *PROC. R. SOC. (LONDON) A,* VOL 296, 1967, P 123
28. J.N. HUNT, *PHILOS. MAG.,* VOL 18, 1968, P 669
29. W. LUCAS, J.D. WILLIAMS, AND B. CROSSLAND, SOME METALLURGICAL OBSERVATIONS

ON EXPLOSIVE WELDING, *PROC. SECOND INT. CONF. OF CENTER FOR HIGH ENERGY FORMING,* 1969, P 8.1.1

30. L.F. TRUEB, THE MICROSTRUCTURE OF EXPLOSION-BONDED METALS, *TRANS. AIME,* VOL 242, 1966, P 1057
31. L.F. TRUEB, *J. APPL. PHYS.,* VOL 34 (NO. 4), 1968, P 928
32. M.P. BONDAR AND V.M. OGOLIKHIN, PLASTIC DEFORMATION IN BONDING ZONE UNDER EXPLOSIVE WELDING AND ITS ROLE IN BONDING FORMATION, *J. PHYS.,* C5 (NO. 8-46), 1985, P C5-379 TO C5-384
33. M. HAMMERSCHMIDT AND H. KREYE, MICROSTRUCTURE AND BONDING MECHANISMS IN EXPLOSION WELDING, *SHOCK WAVES AND HIGH-STRAIN RATE PHENOMENA IN METALS,* M.A. MEYERS AND L.E. MURR, ED., PLENUM PRESS, 1981, P 961
34. A. OBERG, N. MARTENSSON, AND J.A. SCHWEITZ, FUNDAMENTAL ASPECTS OF FORMATION AND STABILITY OF EXPLOSIVE WELDS, *METALL. TRANS. A,* VOL 16, MAY 1985, P 841-852
35. H.K. BALAKRISCHMA, V.C. VENTATESH, AND P.K. PHILIP, INFLUENCE OF COLLISION PARAMETERS ON THE MORPHOLOGY OF INTERFACE IN ALUMINUM-STEEL EXPLOSION WELDS, *SHOCK WAVES AND HIGH-STRAIN RATE PHENOMENA IN METALS,* M.A. MEYERS AND L.E. MURR, ED., PLENUM PRESS, 1981, P 975

Mechanical Properties of Soft-Interlayer Solid-State Welds

R.S. Rosen, Lawrence Livermore National Laboratory; M.E. Kassner, Oregon State University

Introduction

THE UTILIZATION OF METAL INTERLAYERS for solid-state welding can be an appropriate joining method when brittle intermetallic compound formation, differential thermal contraction during cooling, or oxide dissociation and dissolution temperature preclude the use of conventional methods, such as fusion welding, brazing, and direct solid-state welding of the base metals. An interlayer can have the form of a foil (Ref 1, 2, 3, 4, 5, 6, 7, 8), or it can be applied to one or both of the base-metal surfaces by various coating methods, such as electrodeposition (Ref 1, 2, 4, 9), plasma spraying (Ref 2), or vapor-deposition methods.

Methods that utilize coated interlayers require a two-step joining procedure (coating and welding), as opposed to joints fabricated with foil interlayers, which may only require a single solid-state welding step. However, methods that utilize coated interlayers have the advantage of not requiring solid-state welding at the interface of the interlayer and base metal, as necessitated by the use of foil interlayers. The solid-state welding of some foil interlayer and base-metal combinations may require high temperatures and/or pressures, because of the high dissociation temperatures of the surface oxides, which may preclude or restrict their use (Ref 10, 11, 12).

Various base-metal combinations have been solid-state welded using coated silver as the interlayer metal. The principal advantage of silver as an interlayer for solid-state welding is that high-strength joints can be fabricated at relatively low temperature (473 to 673 K) and pressures (100 to 200 MPa, or 15 to 30 ksi), because of the low dissociation temperature (<460 K) for silver oxide. Depending on the strength of the base metals and the method (hydrostatic or uniaxial stress) by which pressure is applied to the silver/silver interface, the joining process may result in variable deformation in the base metals. Because some plasticity in the base metal may be important to assist with intimate atomic contact across the interface, "soft" interlayers are often preferred.

Silver interlayers for solid-state welding are usually applied to one or both of the base metals by deposition using either electrodeposition (Ref 1, 13, 14, 15, 16, 17) or vacuum-coating methods. The vacuum-coating methods have included electron-beam evaporation (Ref 18), hot-hollow cathode (HHC) evaporation (Ref 15, 19, 20, 21, 22, 23, 24, 25, 26), ion-plating using sputtering (Ref 27, 28, 29), and planar-magnetron (PM) sputtering (Ref 30, 31, 32, 33).

Prior to deposition, the surface oxide layer may require removal from the base metal in order to achieve adequate adhesion of the silver. The surface oxide layer is usually removed by chemical- or sputter-etching methods. One advantage of vacuum coating over electrodeposition is the ability to sputter-etch the base metal *in situ.* This permits silver deposition with minimal recontamination of the oxygen on the base-metal surface (Ref 34).

Soft-interlayer solid-state welds that join stronger base metals have unique mechanical properties that are of fundamental interest and may be of critical importance to designers. These assemblies can be subjected to various applied stress states, such as tensile, where the load is applied perpendicular to the interlayer/base-metal interfaces, or shear, where the applied shear stress lies within the plane of the interlayer. Of course, in application, the interlayer may be subjected to a combination of applied tensile and shear loads. This is important because these types of welds are, mechanically, quite anisotropic. This includes the case where interlayers may be subjected to residual stresses from fabrication. The loading may be continually increasing, as with a conventional tensile test, or static, as with creep tests. Again, these behaviors are quite different.

Additionally, some interlayer/base-metal interfaces have been shown to be susceptible to stress-corrosion cracking when exposed to critical environments (Ref 35). The mechanical properties of soft-interlayer solid-state welds and the implications of these behaviors to service stress states and environments are discussed in this article. It should be noted that although silver interlayers are emphasized, primarily because of their selection by previous researchers and industrial fabricators, the described trends appear to be completely consistent with joints that use other soft-interlayer metals.

References

1. J.T. NIEMANN, R.P. SOPHER, AND P.J. RIEPPEL, DIFFUSION BONDING BELOW 1000 °F, *WELD. J.,* VOL 37, 1958, P 337S-342S
2. I.M. BARTA, LOW TEMPERATURE DIFFUSION BONDING OF ALUMINUM ALLOYS, *WELD. J.,* VOL 43, 1964, P 241S-247S
3. D. HAUSER, P.A. KAMMER, AND J.H. DEDRICK, SOLID-STATE WELDING OF ALUMINUM, *WELD. J.,* VOL 46, 1967, P 11S-22S
4. P.A. KAMMER, R.E. MONROE, AND D.C. MARTIN, FURTHER STUDIES OF DIFFUSION BONDING BELOW 1000 °F, *WELD. J.,* VOL 48, 1969, P 116S-124S
5. Y. IINO AND N. TAGAUCHI, INTERDIFFUSING METALS LAYER TECHNIQUE OF CERAMIC-METAL BONDING, *J. MATER. SCI. LETT.,* VOL 7 (NO. 9), 1988, P 981-982
6. A. URENA, J.M.G.D. SALAZAR, AND J. QUINONES, DIFFUSION BONDING OF ALUMINA TO STEEL USING SOFT COPPER INTERLAYER, *J. MATER. SCI.,* VOL 27, 1992, P 599-606
7. T. YAMADA K, YOKOI, AND A. KOHNO, EFFECT OF RESIDUAL STRESS ON THE STRENGTH OF ALUMINA-STEEL JOINT WITH AL-SI INTERLAYER, *J. MATER. SCI.,* VOL 25, 1990, P 2188-2192
8. M.G. NICHOLAS AND R.M. CRISPIN, DIFFUSION BONDING STAINLESS STEEL TO ALUMINA USING ALUMINUM INTERLAYERS, *J. MATER. SCI.,* VOL 17, 1982, P 3347-3360
9. C.L. CLINE, AN ANALYTICAL AND EXPERIMENTAL STUDY OF DIFFUSION BONDING, *WELD. J.,* VOL 45, 1966, P 481S-489S
10. A.T. D'ANNESSA, THE SOLID-STATE BONDING OF REFRACTORY METALS, *WELD. J.,* VOL 43, 1964, P 232S-240S
11. M.G. NICHOLAS AND R.M. CRISPIN, DIFFUSION BONDING CERAMICS WITH DUCTILE METAL INTERLAYERS, *PROC. INT. DIFFUSION BONDING* (CRANFIELD, U.K.), 1987, P 173-182
12. B. DERBY, ZIRCONIA/METAL DIFFUSION BONDS, *PROC. INT. CONF. DIFFUSION BONDING* (CRANFIELD, U.K.), 1987, P 195-201
13. R.A. MORLEY AND J. CARUSO, THE DIFFUSION WELDING OF 390 ALUMINUM ALLOY HYDRAULIC VALVE BODIES, *WELD. J.,* VOL 59 (NO. 8), 1980, P 29-34
14. C.H. CRANE, D.T. LOVELL, W.A. BAGINSKY, AND M.G. OLSEN, DIFFUSION WELDING OF DISSIMILAR METALS, *WELD. J.,* VOL 46, 1967, P 23S-31S
15. M. O'BRIEN, C.R. RICE, AND D.L. OLSON, HIGH STRENGTH DIFFUSION WELDING OF SILVER COATED BASE METALS, *WELD. J.,* VOL 55 (NO. 1), 1976, P 25-27
16. J.W. DINI, W.K. KELLEY, W.C. COWDEN, AND E.M. LOPEZ, USE OF ELECTRODEPOSITED

SILVER AS AN AID IN DIFFUSION WELDING, *WELD. J.,* VOL 63 (NO. 1), 1983, P 26S-34S

17. R.S. ROSEN, "TIME-DEPENDENT FAILURE OF SILVER INTERLAYER WELDS," PH.D.THESIS, UNIVERSITY OF CALIFORNIA, DAVIS, 1990
18. J.L. KNOWLES AND T.H. HAZLETT, HIGH-STRENGTH LOW-TEMPERATURE BONDING OF BERYLLIUM AND OTHER METALS, *WELD. J.,* VOL 49 (NO. 7), 1970, P 301S-310S
19. P.S. MCLEOD AND G. MAH, THE EFFECT OF SUBSTRATE BIAS VOLTAGE ON THE BONDING OF EVAPORATED SILVER COATINGS, *J. VAC. SCI. TECHNOL.,* VOL 11 (NO. 1), 1974, P 119-121
20. G. MAH, P.S. MCLEOD, AND D.G. WILLIAMS, CHARACTERIZATION OF SILVER COATINGS DEPOSITED FROM A HOLLOW CATHODE SOURCE, *J. VAC. SCI. TECHNOL.,* VOL 11 (NO. 4), 1974, P 663-665
21. D.G. WILLIAMS, VACUUM COATING WITH A HOLLOW CATHODE SOURCE, *J. VAC. SCI. TECHNOL.,* VOL 11 (NO. 1), 1974, P 374-377
22. E.R. NAIMON, D. VIGIL, J.P. VILLEGAS, AND L. WILLIAMS, ADHESION STUDY OF SILVER FILMS DEPOSITED FROM A HOY HOLLOW-CATHODE SOURCE, *J. VAC. SCI TECHNOL.,* VOL 13 (NO. 6), 1976, P 1131-1135
23. E.R. NAIMON, J.H. DOYLE, C.R. RICE, D. VIGIL, AND D.R. WALMSLEY, DIFFUSION WELDING OF ALUMINUM TO STAINLESS STEEL, *WELD. J.,* VOL 60 (NO. 1), 1981, P 17-20
24. D.T. LARSON AND H.L. DRAPER, CHARACTERIZATION OF THE BE-AG INTERFACIAL REGION OF SILVER FILMS DEPOSITED ONTO BERYLLIUM USING A HOT HOLLOW CATHODE DISCHARGE, *THIN SOLID FILMS,* VOL 107, 1983, P 327-334
25. R.S. ROSEN, D.R. WALMSLEY, AND Z.A. MUNIR, THE PROPERTIES OF SILVER-AIDED DIFFUSION WELDS BETWEEN URANIUM AND STAINLESS STEEL, *WELD. J.,* VOL 65 (NO. 4), 1986, P 83S-92S
26. J.W. ELMER, M.E. KASSNER, AND R.S. ROSEN, THE BEHAVIOR OF SILVER-AIDED DIFFUSION-WELDED JOINTS UNDER TENSILE AND TORSIONAL LOADS, *WELD. J.,* VOL 67 (NO. 7), 1988, P 157S-162S
27. J. HARVEY, P.G. PATRIDGE, AND A.M. LURSHEY, FACTORS AFFECTING THE SHEAR STRENGTH OF SOLID STATE DIFFUSION BONDS BETWEEN SILVER-COATED CLAD AL-ZN-MG ALLOY (ALUMINUM ALLOY 70100, *MATER. SCI. ENG.,* VOL 79, 1986, P 191-199
28. P.G. PARTRIDGE AND D.V. DUNFORD, ON THE TESTING OF DIFFUSION-BONDED OVERLAP JOINTS BETWEEN CLAD AL-ZN-MG ALLOY (7010) SHEET, *J. MATER. SCI.* VOL 22, 1987 P 1597-1608
29. D.V. DUNFORD AND P.G. PARTRIDGE, THE PEEL STRENGTHS OF DIFFUSION BONDED. JOINTS BETWEEN CLAD AL-ALLOY SHEETS, *J. MATER. SCI.,* VOL 22, 1987, P 1790-1798
30. R.S. ROSEN AND M.E. KASSNER, DIFFUSION WELDING OF SILVER INTERLAYERS COATED ONTO BASE METALS BY PLANAR-MAGNETRON SPUTTERING, *J. VAC. SCI. TECHNOL. A,* VOL 8 (NO. 1), 1900, P 19-29
31. M.E. KASSNER, R.S. ROSEN, G.A. HENSHALL, AND W.E. KING, DELAYED FAILURE OF SILVER AIDED DIFFUSION WELDS BETWEEN STEEL, *PROC. 2ND INT. CONF. BRAZING, HIGH TEMPERATURE BRAZING, AND DIFFUSION WELDING,* GERMAN WELDING SOCIETY, 1989, P 47-52
32. M.E. KASSNER, R.S. ROSEN, G.A. HENSHALL, AND K.D. CHALLENGER, TIME-DEPENDENT FAILURE OF SILVER-INTERLAYER DIFFUSION WELDS BETWEEN ELASTICALLY-DEFORMING BASE METALS, *SCR. METALL. MATER.,* VOL 24, 1990, P 587-592
33. M.E. KASSNER, R.S. ROSEN, AND G.A. HENSHALL, DELAYED MECHANICAL FAILURE OF SILVER-INTERLAYER DIFFUSION BONDS, *METALL. TRANS. A,* VOL 21, 1990, P 3085-3100
34. P.G. PARTRIDGE AND J. HARVEY, "PROCESS FOR THE DIFFUSION BONDING OF ALUMINUM MATERIALS," U.K. PATENT APPLICATIONS GB 2117691 A, 1983
35. R.S. ROSEN, S. BEITSCHER, AND M.E. KASSNER, STRESS CORROSION CRACKING OF URANIUM-SILVER INTERFACES IN SILVER-AIDED DIFFUSION WELDS, *INT. CONF.*

ENVIRONMENT-INDUCED CRACKING OF METALS, VOL 10, NACE, 1989, P 429-433

Mechanical Properties of Soft-Interlayer Solid-State Welds

R.S. Rosen, Lawrence Livermore National Laboratory; M.E. Kassner, Oregon State University

Microstructure of Interlayer Welds

As-Deposited Coatings. A fine-grained (typically <1 µm, or 40 µin., in diameter) columnar structure commonly forms in coatings that are vacuum deposited onto low-temperature substrates (~0.3 T_m) using a low working-gas pressure (~1 Pa, or 1.5 × 10^{-4} psi). This type of structure has been reported previously in silver coatings fabricated for interlayer welding by HHC evaporation (Ref 20) and PM sputter deposition (Ref 30). The columns are perpendicular to the base-metal surface, and the axes of the columns are oriented perpendicular to a close-packed crystallographic direction. The columnar grains resulting from the PM sputter-deposition of silver often contain a high density of growth twins that are approximately 15 nm (0.6 µin.) thick (Ref 30).

Solid-State-Welded Interlayers. Figure 1 shows a solid-state-welded silver interlayers that is 150 µm (6 mils) thick between steel base metals after solid-state welding at 673 K (Ref 17). The silver interlayer typically consists of two types of structures: nonrecrystallized silver that has retained the columnar grains and fine microtwins of the as-deposited coating and silver that has recrystallized into large grains (>1 mm, or 0.04 in., in diameter) containing a high density of annealing twins. The solid-state-welded interfaces are generally free of any large voids, or nonbonded regions. In the case of rough base-metal surfaces, large voids can remain along the silver/silver interface.

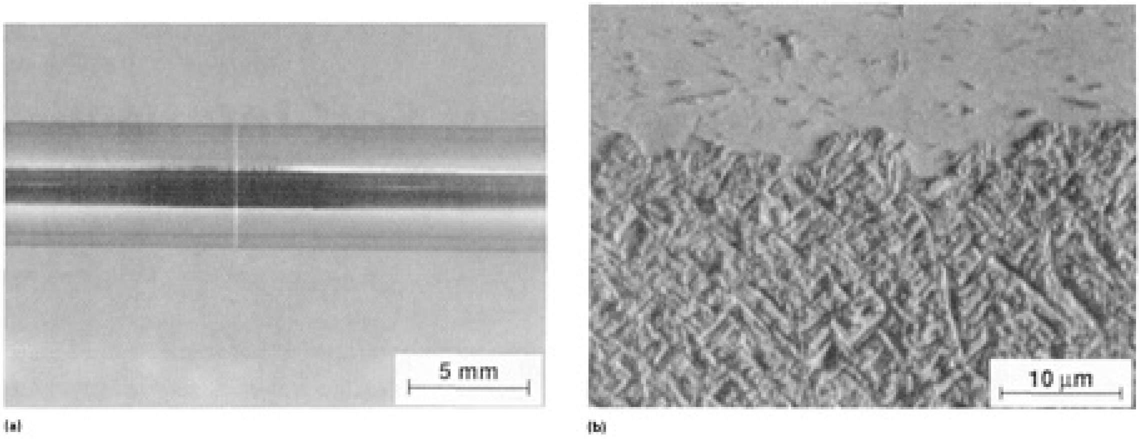

FIG. 1 OPTICAL PHOTOGRAPH SHOWING A 150 µM (6 MIL) THICK SILVER INTERLAYER JOINING BASE METALS (A) AND OPTICAL MICROGRAPH OF THE SOLID-STATE-INTERFACE (B). SOURCE: REF 17

References cited in this section

17. R.S. ROSEN, "TIME-DEPENDENT FAILURE OF SILVER INTERLAYER WELDS," PH.D.THESIS, UNIVERSITY OF CALIFORNIA, DAVIS, 1990
20. G. MAH, P.S. MCLEOD, AND D.G. WILLIAMS, CHARACTERIZATION OF SILVER COATINGS DEPOSITED FROM A HOLLOW CATHODE SOURCE, *J. VAC. SCI. TECHNOL.*, VOL 11 (NO. 4), 1974, P 663-665
30. R.S. ROSEN AND M.E. KASSNER, DIFFUSION WELDING OF SILVER INTERLAYERS COATED ONTO BASE METALS BY PLANAR-MAGNETRON SPUTTERING, *J. VAC. SCI. TECHNOL. A,* VOL 8 (NO. 1), 1900, P 19-29

Tensile Loading of Soft-Interlayer Welds

Despite the relatively low strength of the interlayer material, it has long been known that thin (1 µm to 1 mm, or 40 µin. to 0.04 in.) interlayer welds or brazes between stronger base material can have ultimate tensile strengths (UTS) that are much higher than that of the bulk interlayer material (Ref 36, 37). For example, the UTS of silver-interlayer solid-state welds between maraging steel base metals has been reported to be nearly 800 MPa (120 ksi) (Ref 32, 3), whereas the UTS of the bulk silver is only about 250 MPa (35 ksi) (Ref 38). The high tensile strength of the joint is due to the mechanical constraint provided by the stronger base metal, which restricts transverse contraction of the interlayer (Ref 36, 37, 39, 40, 41, 42, 43, 44, 45). The constraint establishes radial (triaxial) stresses that inhibit the development of shear stresses within the joint, thereby decreasing the effective, or von Mises, stress in the interlayer. This inhibits the plastic flow that leads to ductile fracture.

Effect of Interlayer Thickness on Stress. The degree of mechanical constraint in the soft interlayer and the joint strength tend to increase with a decreasing thickness-to-diameter ratio, *t/d*, of the interlayer (Ref 36, 37, 39 , 40, 41, 42, 44, 46, 47, 48, 49, 50, 51, 52, 53). Figure 2 summarizes the effect of *t/d* on the UTS of solid-state-welded silver interlayers fabricated using PM sputter deposition (Ref 30) and electrodeposition (Ref 15), compared with Ag-4Pd brazed interlayers (no hardening, because of the addition of palladium) (Ref 41). The UTS is shown to increase with a decreasing *t/d* ratio to about 0.002, below which only small strength increases can be expected. The higher UTS of sputter-deposited silver interlayer, compared with those utilizing electrodeposition or brazed-silver alloy, are discussed in the section "Effect of Interlayer Fabrication Method" in this article.

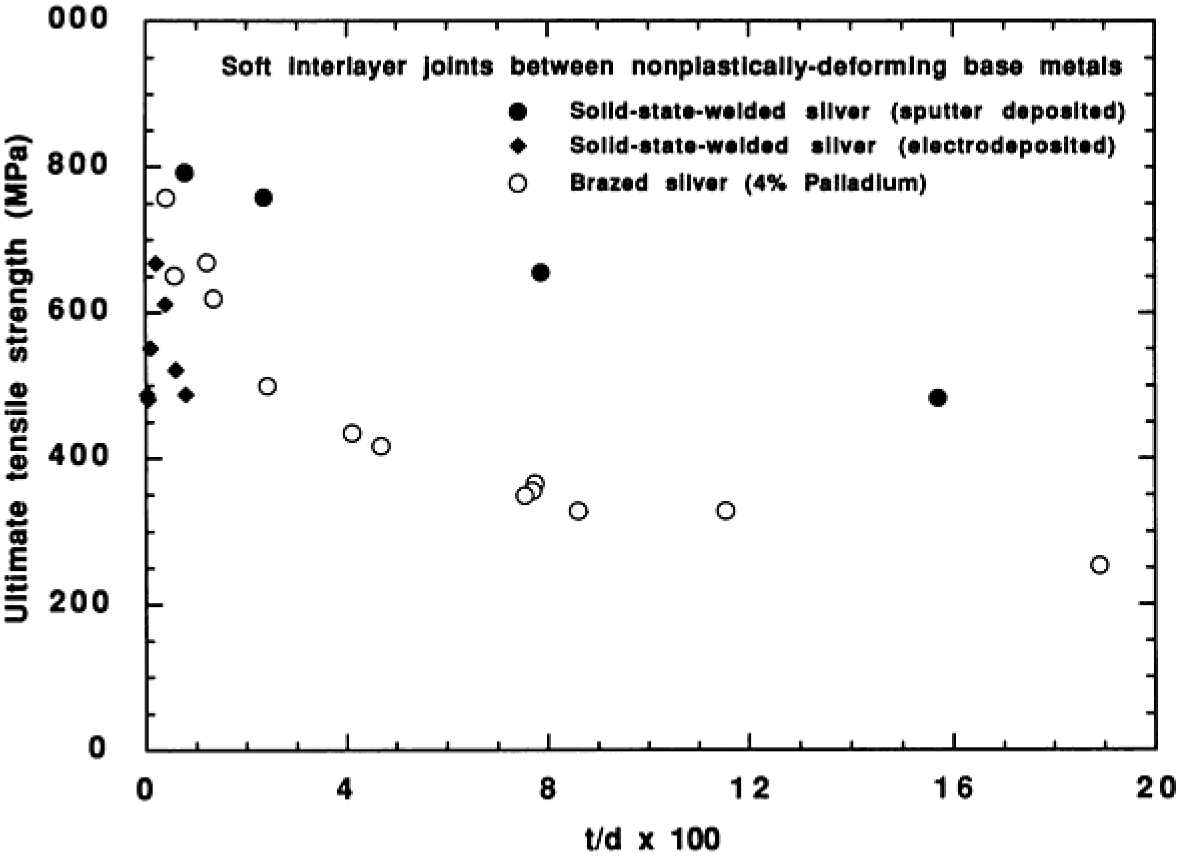

FIG. 2 EFFECT OF INTERLAYER THICKNESS TENSILE STRENGTH OF SOLID-STATE-WELDED SILVER INTERLAYERS FABRICATED USING PM SPUTTER DEPOSITION AND ELECTRODEPOSITION, COMPARED WITH

AG-4PD BRAZED INTERLAYERS

Although the tensile behavior of interlayer joints has been reasonably well understood qualitatively, particularly for thick interlayers, quantitative prediction has only recently been accomplished with finite-element method (FEM) codes. Orowan (Ref 36, 37) was perhaps the first to derive an analytical expression for the radial (constraining) stress within the interlayer. It was proposed that

$$s_r = \frac{s_y}{t}\left(\frac{d}{2}-r\right) \qquad \textbf{(EQ 1)}$$

where σ_r is the radial stress, *t* is the thickness, *d* is the diameter, *r* is the radius, and σ_y is the yield stress of the interlayer. This equation was originally derived by assuming ideal plastic behavior (that is, no strain hardening) of the interlayer and uniform plastic in the interlayer. Analytical attempts to account for strain hardening within the interlayer have led to modifications of Eq 1 that have, experimentally at least, provided reasonable correlation with upper bounds for UTS of soft-interlayer joints (Ref 41, 43, 48, 49).

More recently, FEM analyses (REf 53, 54) reveal significant deviation from q 1. Figure 3(a) illustrates the effect of *t*/*d* on th radial stress distribution at the center plane for silver interlayer welds between nonplastically deformation base metals at a fixed applied tensile stress. These were determined from a FEM analysis that included the observed strain hardening of the interlayer metal and are plotted along with results predicted by Eq 1. This comparison shows that for large *t*/*d* ratios ("thick" interlayers), Eq 1 predicts qualitatively correct behavior for the radial stress distribution. However, for small *t*/*d* ratios ("thin" interlayers), the FEM results indicate a relatively uniform and substantially lower stress distribution over most of the central portion of the interlayer than predicted by Eq 1. This is partly due to the simplifying assumptions in Eq 1, particularly that of a radially uniform strain state.

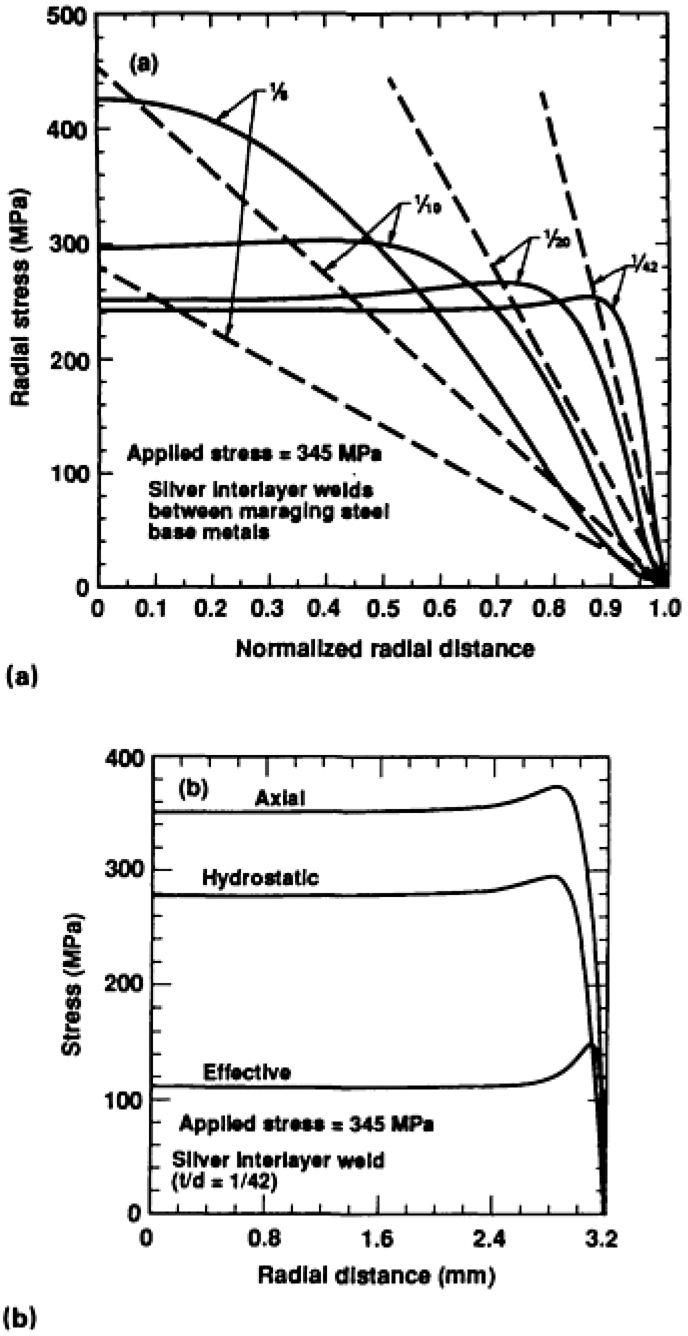

FIG. 3 FINITE-ELEMENT METHOD ANALYSIS OF THE EFFECT OF *T/D* ON THE RADIAL STRESS DISTRIBUTION AT THE CENTER PLANE FOR SILVER INTERLAYER WELDS BETWEEN NONPLASTICALLY DEFORMING BASE METALS AT A FIXED APPLIED TENSILE STRESS. (A) PREDICTIONS OF EQ 1 ARE GIVEN BY DASHED LINES. (B) THE STRESS STATE IN A STRAIN-HARDENING SILVER INTERLAYER AT A SMALL FIXED *T/D* RATIO

This effect is further illustrated in Fig. 3(b), which shows the stress state in a strain-hardening silver interlayer at a small fixed *t/d* ratio. Although it is not shown, the effective (von Mises) stress, which measures the tendency to plastically deform, decreases with decreasing *t/d*, down to approximately 0.025. Hence, a decreasing *t/d* ratio increases the mechanical constraint in the interlayer such that a larger applied stress is needed to produce a given amount of plastic strain in the interlayer. Interestingly, and in contrast with Eq 1, the effective stress does not appear to substantially changes with decreases in *t/d* that are less than approximately 0.025 (Ref 54). This is qualitatively consistent with the extrapolations of the solid-state-welded silver data in Fig. 2 (but not the braze data).

Early investigators (Ref 15, 40, 44, 45, 55) believed that this phenomenon was due only to discontinuities (solidification shrinkage cavities, surface flatness, and such) in very thin brazes and solid-state welds. Recent results (Ref 17, 30) of solid-state-welded interlayers fabricated using sputter-deposited silver onto lapped base metals (surface roughness of 0.03 μm, or 1.2 μin.) indicate that the modest increase (or plateau) in strength occurs even in the absence of preexisting voids that are common in thin brazes.

Interlayer Strain. Associated with the increased UTS of the constrained interlayers, compared to the UTS of the bulk or unconstrained interlayer metal, is a decreased plastic strain-to-failure, ε_f, of the interlayer. Early results (Ref 42) found that for brazed-silver interlayers joining elastically deforming base metals with a *t*/*d* ratio of 0.016, the ε_f is about 0.01. The results from specimens fabricated using solid-state welding of PM sputter-deposited silver interlayers, in which the base metals did not plastically deform, indicated plastic strains of only abut 0.001 for a comparable *t*/*d* ratio of 0.024 (Ref 17, 33). The reason for the disparity is not clear. Perhaps the relatively large number of interfaces in the PM sputter-deposited interlayers in the PM sputter-deposited interlayers (Ref 30, 33) led to higher cavity-nucleation rates. Other works do not appear to have made reliable attempts at determining the plastic strain-to-failure value. It should be emphasized that despite the relatively low plastic strain-to-failure, the fracture surface is that of a classic, ductile, microvoid-coalescence type of failure (Ref 33).

Time-Dependent Failure. It has been observed that at stresses substantially less than the UTS, delayed or time-dependent ductile fracture may occur. For at least the case of silver interlayers, decreases in the applied stress are associated with increased times-to-rupture,f as indicated in Fig. 4. This figure shows that decreasing the *t*/*d* ratio (to about 0.02) and/or improving the base-metal surface finish (lapping) results in longer tensile creep-rupture times. Time-dependent failure is believed to be a consequence of cavity nucleation and/or growth (Ref 33), as is observed in short-term tensile tests that determine the UTS. For solid-state welds in which two coatings are joined, nucleation occurs principally at the silver/silver welded interface or the base-metal/silver coated interface. The cavities appear to nucleate as a consequence of time-dependent plasticity within the interlayer. The cavities expand in diameter and final interlinkage and failure appears to occur at relatively low "macroscopic" strains within the interlayer (for example, $\varepsilon_f < 10^{-3}$).

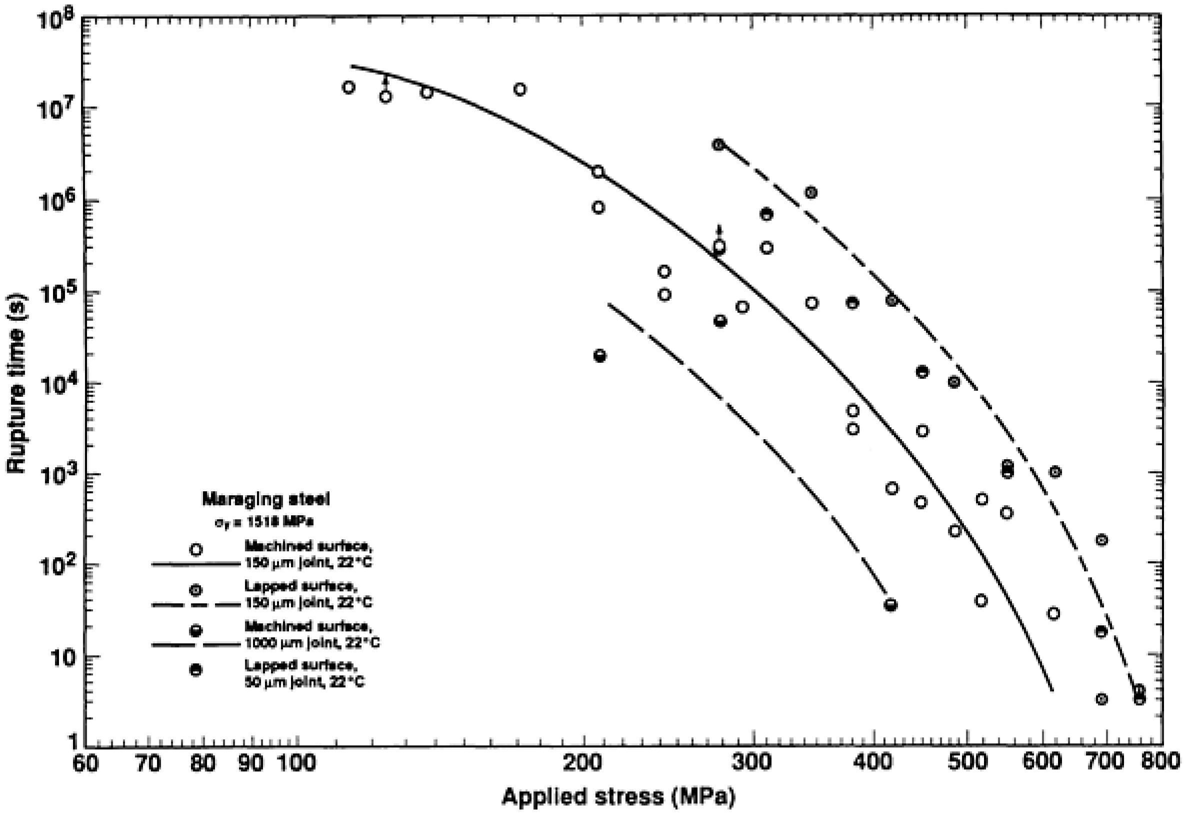

FIG. 4 EFFECT OF INTERLAYER THICKNESS AND BASE-METAL SURFACE FINISH ON CREEP RUPTURE OF SOLID-STATE-WELDED SILVER JOINTS BETWEEN NONPLASTICALLY DEFORMING BASE METALS

Figure 5 shows scanning electron micrographs of a solid-state-welded silver interlayer joining steel base metals, loaded at the relatively low stress level of 207 MPa (30 ksi) (albeit, for a substantial fraction of the expected creep-rupture time).

Some of the larger cavities consist of clusters of smaller (<0.5 μm, or 200 μin., in diameter) microvoids. A typical fracture surface of a creep-rupture specimen is illustrated in Fig. 6. Naturally, the fracture surfaces appear identical to those of the short-term tests in which UTS values were measured. It should be emphasized that, consistent with Fig. 2, as the interlayer metal strength increases, the strength of the weld joint increases. Similarly, as the interlayer metal creep resistance increases, the rupture times of the weld joint should increase. Hence, as reported for gold-nickel high-strength/creep-resistant interlayer brazes (Ref 56), time-dependent failure has not been observed over periods of about 1 y. Also, it appears that for silver brazes (Ref 33), time-dependent failure is observed, but at increasingly longer times at decreasing stresses, perhaps because of differences in substructures, compared with solid-state-welded silver interlayers.

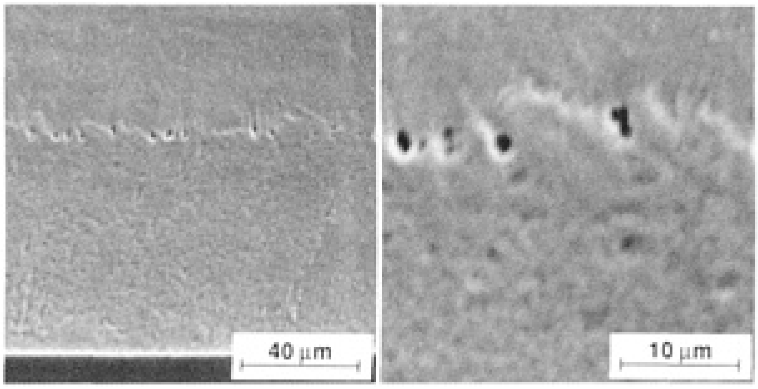

FIG. 5 SCANNING ELECTRON MICROGRAPHS SHOWING SOLID-STATE-WELDED SILVER INTERLAYER JOINING NONPLASTICALLY DEFORMING BASE METALS, LOADED AT 207 MPA (30 KSI) (~$\frac{1}{3}$ ULTIMATE TENSILE STRENGTH) FOR A SUBSTANTIAL FRACTION OF THE EXPECTED CREEP-RUPTURE TIME

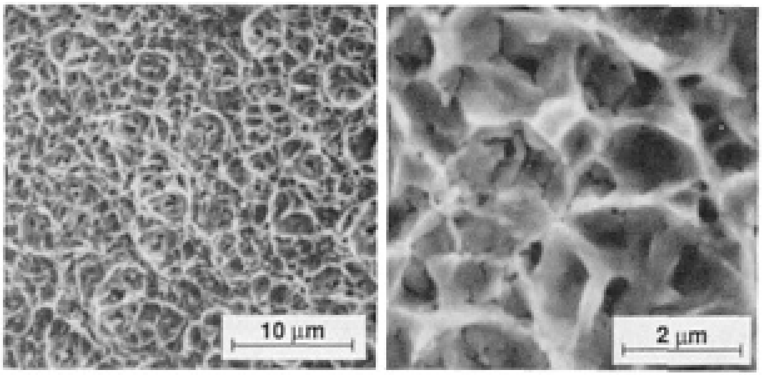

FIG. 6 SOLID-STATE-WELDED SILVER INTERLAYER FRACTURE SURFACE BETWEEN LAPPED MARAGING STEEL BASE METALS

A nucleation-based theory has been proposed to explain the low plastic strain, time-dependent failures of soft-interlayer solid-state welds in which preexisting voids do not exist, and the base metal does not plastically deform (Ref 17, 33, 57). Plastic deformation of the interlayer, acting under the effective stress, causes cavity nucleation by dislocation pile-ups at interfaces. Nucleation continues with further creep of the interlayer, until the concentration of nuclei is sufficiently high to lead to instability in the material between the cavities. Coalescence by interlinkage or by uniform wall expansion ("growth") then occurs. The point of instability between two small nanometer-sized cavities or nuclei separated by a few diameters or so may lead to subsequent rapid nucleation between the cavities, resulting in the formation of a single, larger cavity (as shown in Fig. 5). This process is repeated and eventually leads to the onset of final rupture, and gives the appearance of a classic, ductile, microvoid-coalescence type of fracture.

In contrast with this nucleation-based theory, other investigators (Ref 42) suggest that the macroscopic strain-to-failure is determined by the details of the uniform cavity expansion (growth). Here, once the cavities expand and come in proximity to other cavities, coalescence and ductile fracture occurs. In a creep experiment, according to the usual cavity-growth models, the rupture time is determined by the cavity growth rate or the speed of the expanding cavity wall. In the absence of vacancy-controlled growth, this speed is controlled by the macroscopic plastic strain rate. Therefore, the static (or creep) and dynamic (or quasi-static) failures of the thin interlayer welds should be, according to cavity-growth-based theories, related to the macroscopic strain in the interlayer. If steady-state plasticity is achieved throughout the silver interlayer material surrounding the cavities, then unstable cavity wall expansion may occur. However, since the macroscopic strains-to-failure are small ($\varepsilon_f < 10^{-3}$, the time or strain up to the onset of final catastrophic failure (where a large ductile-dimple fracture surface begins to form) is not easily explained by a cavity-growth model in a strain-hardening material.

The silver-plasticity concept for rationalizing creep rupture is consistent with the temperature dependence of the time-to-rupture. The activation energy for the creep of silver at ambient and near-ambient temperatures for various substructures falls within the range of 50 to 71 kJ/mole (Ref 38, 58). The activation energy for the creep rupture of the solid-state-welded silver joints of Fig. 2, equal to 65 kJ/mole (Ref 33), falls within the range of activation energies for silver plasticity.

Effect of Base-Metal Properties. Time-dependent tensile fracture can be dramatically accelerated if there is creep, or time-dependent plasticity, of the base materials. Figure 7 shows the creep-rupture behavior of solid-state-welded silver interlayers (*t*/*d* = 0.024) joining one of two different strengths of type 304 stainless steel base metals: annealed or cold-worked (Ref 17, 33). Interestingly, these base metals undergo time-dependent plasticity, or creep, at ambient temperatures at stresses that are less than the conventionally yield stress of the metal. Although the accumulated plastic strain in the creeping stainless steel is only about 0.001 to 0.01 at low stress levels, this nonetheless induces concomitant plastic strain within the interlayer that exceeds that which would occur within the interlayer in the absence of base-metal plasticity, resulting in shorter creep-rupture times. This behavior is consistent with previous reported data for creep ruptures of interlayers that join plastically deforming base metals (Ref 25, 26, 59, 60).

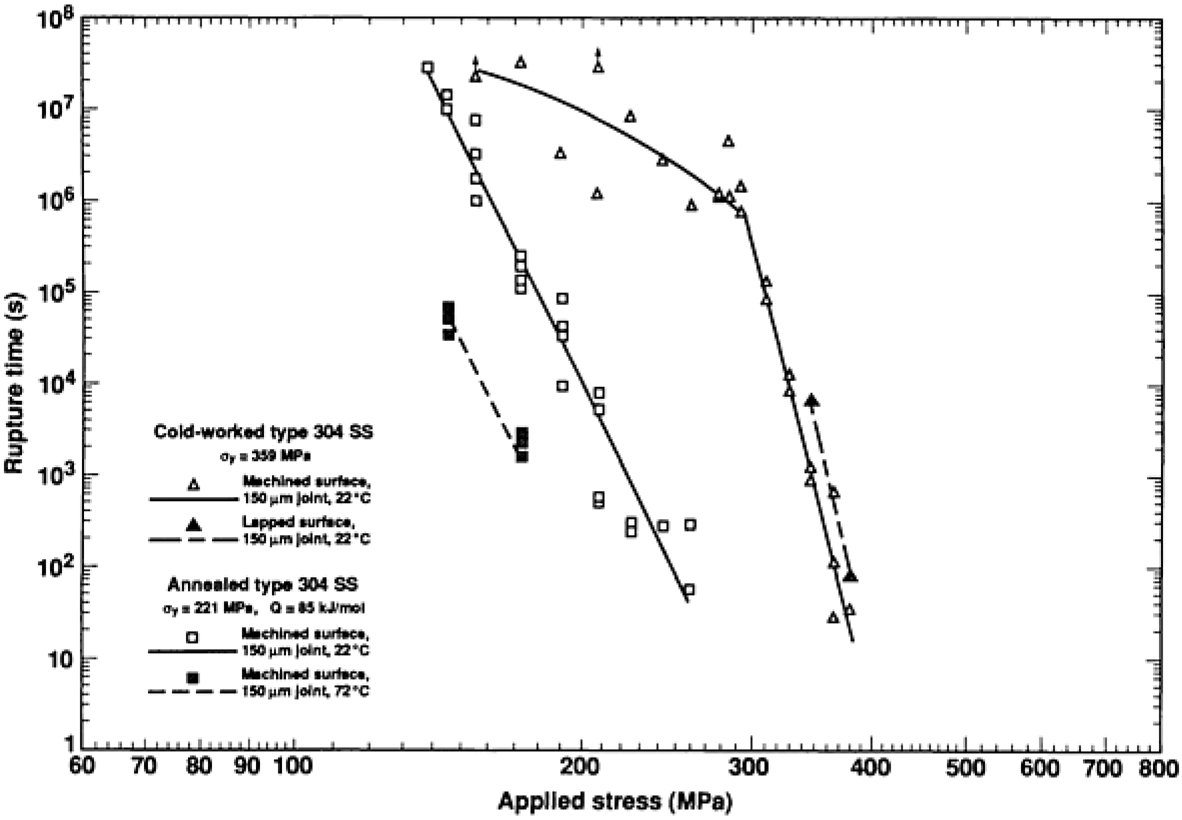

FIG. 7 CREEP-RUPTURE BEHAVIOR OF SOLID-STATE-WELDED SILVER INTERLAYERS (*T*/*D* = 0.024) JOINING ONE OF TWO DIFFERENT STRENGTHS OF TYPE 304 STAINLESS STEEL BASE METALS (ANNEALED OR COLD-WORKED), BOTH OF WHICH UNDERGO TIME-DEPENDENT PLASTICITY OR CREEP AT AMBIENT TEMPERATURES AT STRESSES LESS THAN THE CONVENTIONAL YIELD STRESS OF THE METAL

Finite-element analyses have confirmed that plasticity in the base metal allows the interlayer to plastically deform to a much greater extent than if the base metal deforms only elastically (Ref 54). This observation is important, because these small levels of time-dependent plasticity (0.001 to 0.01) are often observed, over time, in other structural metals and alloys at ambient temperature at stresses significantly below the conventionally define yield stress. With an increase in test temperature, the stainless steel base-metal creep rate significantly increases and rupture times decrease. The activation energies for the creep rupture of silver interlayers between stainless steel base metals and for plastic flow in stainless steel are comparable, further suggesting that the temperature dependence of creep rupture is determined by creep plasticity in the stainless steel base metal. The transition from base-metal to interlayer-controlled creep rupture occurs at the "knee" of the cold-worked stainless steel data plot, at stress levels below which base-metal creep is negligible.

Figure 7 also shows the effect of base-metal surface finish, which is eventually coated, on the creep-rupture behavior of joints utilizing cold-worked type 304 stainless steel. Lapped cold-worked stainless steel base-metal specimens have longer rupture times than those of machined cold-worked stainless steel, as was the case with the maraging steel specimens. However, the increase in creep-rupture times of specimens utilizing lapped versus machined plastic base metals is only a factor of 5, which is considerably less than the increase of approximately 50 for specimens utilizing lapped versus machined elastic base metals. The difference between the increases in creep-rupture times for lapped plastic versus lapped elastic base-metal specimens may be due to the difference in failure processes. For the case of a plastic base metal, the strain in the silver is externally and uniformly imposed by the deformation of the base metal. For elastic base metals, plasticity is caused by the effective stress within the interlayer. It is believed that the heterogeneities along the solid-state-welded silver interface caused during the joining of the comparatively rough machined surfaces could result in regions that may be more resistant to plasticity than other regions. The heterogeneities at the solid-state-welded silver interface may be less significant in increasing the nucleation and coalescence rate of cavities in specimens utilizing plastic base metals than those utilizing elastic base metals, where the strain is indigenous to the interlayer, rather than being imposed by base-metal deformation.

Effect of Interlayer Fabrication Method. In general, the tensile strengths of PM sputter-deposited silver joints between various base metals have been reported to be equal to or greater than those utilizing other silver-coating methods and similar solid-state-welding parameters. For elastic base metals with lapped surfaces, PM sputter-deposited silver joint tensile strengths (Ref 30) of 758 and 793 MPa (110 and 115 ksi) using interlayers that are 150 and 50 μm (6 and 2 units thick, respectively, are comparable with HHC-deposited silver joints (758 MPa, or 110 ksi, using 10 μm, or 0.4 mil, thick interlayers) and greater than those of electrodeposited silver joints (669 MPa, or 97 ksi, using 25 μm, or 1 mil, thick interlayers) (Ref 15). However, the tensile strength values for silver-interlayer welds fabricated using these other two coating methods are achieved with considerably thinner interlayers, which have the advantage of greater plastic constraint produced by the base metals. For type 304 stainless steel (plastically deforming) base metals with machined surfaces and 150 μm (6 mil) thick joints, PM sputter-deposited silver joint tensile strengths (Ref 30) of 304 MPa (44 ksi) ($\sigma_{y,SS}$ = 221 MPa, or 32 ksi) and 400 MPa (60 ksi) ($\sigma_{y,SS}$ = 359 MPa, or 52 ksi) are comparable with electrodeposited silver joint strengths of 321 to 367 MPa (47 to 53 ksi) ($\sigma_{y,SS}$ = 296 MPa, or 43 ksi) (Ref 16). For uranium (plastic) base metals with machined surface, the PM sputter-deposited silver joint tensile strength of 442 MPa (64 ksi) $\sigma_{y,U}$ =242 MPa, or 35 ksi) (Ref 30) exceed electrodeposited silver joint strengths of approximately 255 MPa (37 ksi) (despite the higher base-metal yield stress; $\sigma_{y,U}$ = 345 MPa, or 50 ksi) and HHC-deposited silver joint strengths of 345 to 373 MPa (50 to 54 ksi) (Ref 16, 61). However, the lower joint strength using HHC-deposited silver (Ref 61) may have resulted from a lower solid-state welding pressure (138 MPa, or 20 ksi) and temperature (589 K) than those using PM sputter-deposited silver. The tensile strength of solid-state-welded silver joints between dissimilar metals, such as uranium (σ_y = 345 MPa, or 50 ksi) and type 304 stainless steel (σ_y = 296 MPa, or 45 ksi) fabricated using HHC deposition has been reported to average 442 MPa (64 ksi) (Ref 25). The tensile strength of electrodeposited silver joints between the same uranium and stainless steel base metals that were solid-state welded at the same pressure (207 MPa, or 30 ksi), temperature (873 K), and time (2 h) averaged only 345 MPa (50 ksi) (Ref 17, 30). Therefore, solid-state-welded silver joints fabricated using electrodeposition appear to be weaker than those fabricated using HHC deposition or PM sputter deposition.

Although there is only a limited amount of published test results of interlayer joint strengths fabricated by low-temperature solid-state welding of foils, it appears that interlayer/base-metal adhesion is difficult to achieve at temperatures below approximately 0.7 T_m of the interlayer. Tensile strengths of silver joints between alumina and type 321 stainless steel were reported to be only 40 MPa (5.8 ksi) when solid-state-welding parameters were 1073 K and 50 MPa (7.3 ksi) for 0.5 h (Ref 11). The investigators reported threshold solid-state-welding temperatures below which no measurable bonding was produced for each of the foils utilized (silver, aluminum, copper, and nickel). The threshold temperature to achieve solid-state welding of silver foils to alumina and type 321 stainless steel was reported to be 873 K (0.7 T_m of the silver), considerably higher than the temperature required to achieve solid-state welding of silver to itself (0.35 T_m) (Ref 18). Joint strengths of the previous interlayer fabrication processes were reported to be comparable (40) to 50 MPa, or 5.8 to 7.3 ksi) to those of silver foils (Ref 11). Tensile tests reported for silver-foil interlayers between type

304L stainless steel and Ti-6Al-4V alloy suggest that higher strengths may be achieved with coated interlayers than with foil interlayers (Ref 62). These joints were fabricated by solid-state welding a 125 μm (5 mil) thick silver foil between the base metals at 1153 K (0.94 T_m). The investigators (Ref 62) reported average joint strengths of 280 MPa (40 ksi). For annealed type 304 stainless steel and 150 μm (6 mil) thick joints fabricated using PM sputter deposition, the reported (Ref 30) tensile strength of 304 MPa (44 ksi) is higher, despite the higher base-metal yield stress of Ti-6Al-4V alloy used in the silver-foil study.

References cited in this section

3. D. HAUSER, P.A. KAMMER, AND J.H. DEDRICK, SOLID-STATE WELDING OF ALUMINUM, *WELD. J.,* VOL 46, 1967, P 11S-22S
11. M.G. NICHOLAS AND R.M. CRISPIN, DIFFUSION BONDING CERAMICS WITH DUCTILE METAL INTERLAYERS, *PROC. INT. DIFFUSION BONDING* (CRANFIELD, U.K.), 1987, P 173-182
15. M. O'BRIEN, C.R. RICE, AND D.L. OLSON, HIGH STRENGTH DIFFUSION WELDING OF SILVER COATED BASE METALS, *WELD. J.,* VOL 55 (NO. 1), 1976, P 25-27
16. J.W. DINI, W.K. KELLEY, W.C. COWDEN, AND E.M. LOPEZ, USE OF ELECTRODEPOSITED SILVER AS AN AID IN DIFFUSION WELDING, *WELD. J.,* VOL 63 (NO. 1), 1983, P 26S-34S
17. R.S. ROSEN, "TIME-DEPENDENT FAILURE OF SILVER INTERLAYER WELDS," PH.D.THESIS, UNIVERSITY OF CALIFORNIA, DAVIS, 1990
18. J.L. KNOWLES AND T.H. HAZLETT, HIGH-STRENGTH LOW-TEMPERATURE BONDING OF BERYLLIUM AND OTHER METALS, *WELD. J.,* VOL 49 (NO. 7), 1970, P 301S-310S
25. R.S. ROSEN, D.R. WALMSLEY, AND Z.A. MUNIR, THE PROPERTIES OF SILVER-AIDED DIFFUSION WELDS BETWEEN URANIUM AND STAINLESS STEEL, *WELD. J.,* VOL 65 (NO. 4), 1986, P 83S-92S
26. J.W. ELMER, M.E. KASSNER, AND R.S. ROSEN, THE BEHAVIOR OF SILVER-AIDED DIFFUSION-WELDED JOINTS UNDER TENSILE AND TORSIONAL LOADS, *WELD. J.,* VOL 67 (NO. 7), 1988, P 157S-162S
30. R.S. ROSEN AND M.E. KASSNER, DIFFUSION WELDING OF SILVER INTERLAYERS COATED ONTO BASE METALS BY PLANAR-MAGNETRON SPUTTERING, *J. VAC. SCI. TECHNOL. A,* VOL 8 (NO. 1), 1900, P 19-29
32. M.E. KASSNER, R.S. ROSEN, G.A. HENSHALL, AND K.D. CHALLENGER, TIME-DEPENDENT FAILURE OF SILVER-INTERLAYER DIFFUSION WELDS BETWEEN ELASTICALLY-DEFORMING BASE METALS, *SCR. METALL. MATER.,* VOL 24, 1990, P 587-592
33. M.E. KASSNER, R.S. ROSEN, AND G.A. HENSHALL, DELAYED MECHANICAL FAILURE OF SILVER-INTERLAYER DIFFUSION BONDS, *METALL. TRANS. A,* VOL 21, 1990, P 3085-3100
36. E. OROWAN, FRACTURE AND STRENGTH OF SOLIDS, *REP. PROG. PHYS.,* VOL 12, 1948, P 185-231
37. E. OROWAN, FUNDAMENTALS OF BRITTLE BEHAVIOR OF METALS, *FATIGUE AND FRACTURE OF METALS,* JOHN WILEY & SONS, 1952
38. M.E. KASSNER, THE RATE DEPENDENCE AND MICROSTRUCTURE OF HIGH-PURITY SILVER DEFORMED TO LARGE STRAINS BETWEEN 0.16 AND 0.30 TM, *METALL. TRANS. A,* VOL 20, 1989, P 2001-2010
39. N. BREDZ, INVESTIGATION OF FACTORS DETERMINING THE TENSILE STRENGTH OF BRAZED JOINTS, *WELD. J.,* VOL 33, 1954, P 545S-563S
40. W.G. MOFFATT AND J. WULFF, TENSILE DEFORMATION AND FRACTURE OF BRAZED JOINTS, *WELD. J.,* VOL 42, 1963, P 115S-125S
41. H.J. SAXTON, A.J. WEST, AND C.R. BARRETT, DEFORMATION AND FAILURE OF BRAZED JOINTS--MACROSCOPIC CONSIDERATIONS, *METALL. TRANS.,* VOL 2, 1971, P 999-1007
42. A.J. WEST, H.J. SAXTON, A.S. TETELMAN, AND C.R. BARRETT, DEFORMATION AND FAILURE OF THIN BRAZED JOINTS--MICROSCOPIC CONSIDERATIONS, *METALL. TRANS.,* VOL 2, 1971, P

1009-1017

43. E.P. LAUTENSCHLAGER, B.C. MARKER, B.K. MOORE, AND R. WILDES, STRENGTH MECHANISMS OF DENTAL SOLDER JOINTS, *J. DENT. RES.,* VOL. 53, 1974, P 1361-1367
44. R.A. MUSIN, V.A. ANTSIVEROV, Y.A. BELIKOV, Y.V. LYAMIN, AND A.N. SOLOKOV, THE EFFECTS ON THE STRENGTH OF THE JOINT OF THE THICKNESS OF THE SOFT INTERLAYER IN DIFFUSION WELDING, *AUTO. WELD.,* VOL 32, 1979, P 38-40
45. R.M. TRIMMER AND A.T. KUHN, THE STRENGTH OF SILVER-BRAZED STEEL JOINTS--A REVIEW, *BRAZING SOLDERING,* VOL 2, 1982 P 6-13
46. N. BREDZ AND H. SCHWARTZBART, TRIAXIAL TENSION TESTING AND THE BRITTLE FRACTURE STRENGTH OF METALS, *WELD. J.,* VOL 53 (NO. 12), 1956, P 610S-615S
47. W.G. MOFFATT AND J. WULFF, STRENGTH OF SILVER BRAZED JOINTS IN MILD STEEL, *TRANS. AIME,* VOL 209 (NO. 4), 1957, P 442-445
48. R.Z. SHRON AND O.A. BAKSHI, THE PROBLEMS OF GAUGING THE STRENGTHS OF WELDED JOINTS IN WHICH THERE IS A SOFT INTERLAYER, *WELD. PROD.,* VOL 9, 1962, P 19-23
49. O.A. BAKSHI AND A.A. SHATOV, THE STRESSED STATE IN WELDED JOINTS WITH HARD AND SOFT INTERLAYERS, *WELD PROD.,* VOL 13, 1966, P 13-19
50. V.S. GOLOVCHENKO AND B.A. GOLOBOV, THE USE OF SILVER AS AN INTERMEDIATE LAYER FOR JOINING TITANIUM TO OTHER METALS, *WELD. PROD.,* VOL 18, 1971, P 55-57
51. G.K. KHARCHENKO AND A.I. IGNATENKO, THE STRENGTH OF JOINTS WITH A THIN SOFT INTERLAYER, *AUTO. WELD.,* VOL 21, 1968, P 33-35
52. W.M. LEHER AND H. SCHWARTZBART, STATIC AND FATIGUE STRENGTHS OF METALS SUBJECTED TO TRIAXIAL STRESSES, VOL 60, 1960, P 610-626
53. E.A. ALMOND, D.K. BROWN, G.J. DAVIES, AND A.M. COTTENDED, THEORETICAL AND EXPERIMENTAL INTERLAYED BUTT JOINTS TESTED IN TENSION, *INT. J. MECH. SCI.,* VOL 25 (NO. 3), 1983, P 175-189
54. G.A. HENSHALL, R.S. ROSEN, M.E. KASSNER, AND R.G. WHIRLEY, FINITE ELEMENT ANALYSIS OF INTERLAYER WELDS LOADED IN TENSION, *WELD. J.,* VOL 69 (NO. 9) 1990, P 337S-345S
55. H.J. SAXTON, A.J. WEST, AND C.R. BARRETT, THE EFFECT OF COOLING RATE ON THE STRENGTH OF BRAZED JOINTS, *METALL. TRANS.,* VOL 2, 1971, P 1019-1028
56. M.C. TOLLE AND M.E. KASSNER, TENSILE PROPERTIES OF THIN AU-NI BRAZES BETWEEN STRONG BASE METALS, *SCR. METALL. MATER.,* VOL 26, 1992, P 1281-1284
57. M.E. KASSNER, M.C. TOLLE. R.S. ROSEN, G.A. HENSHALL, AND J.W. ELMER, RECENT ADVANCES IN UNDERSTANDING THE MECHANICAL BEHAVIOR OF CONSTRAINED THIN METALS IN BRAZES AND SOLID-STATE BONDS, *THE METAL SCIENCE OF JOINING,* M.J. CIESLAK, ED., TMS, 1992, P 223-232
58. R.W. LOGAN, R.G. CASTRO, AND A.K. MUKHERJEE, MECHANICAL PROPERTIES OF SILVER AT LOW TEMPERATURES, *SCR. METALL.,* VOL 17, 1983, P 63-66
59. T.J. MOORE AND K.H. HOLKO, SOLID-STATE WELDING OF TD-NICKEL BAR, *WELD. J.,* VOL 49 (NO. 9), 1970, P 395S-409S
60. R.S. ROSEN, "AN INVESTIGATION OF THE PROPERTIES OF A SILVER-AIDED SOLID-STATE BOND BETWEEN URANIUM AND STAINLESS STEEL," REPORT UCRL-53458, LAWRENCE LIVERMORE NATIONAL LABORATORY, 1983
61. E.R. NAIMON, R.G. KURZ, D. VIGIL, AND L. WILLIAMS, "SILVER FILMS FOR SOLID STATE BONDING,"REPORT RFP-3125, ROCKWELL INTERNATIONAL, 1981
62. Z. NISENHOLTZ, J. MIRONI, AND N. NIR, DIFFUSION BONDING OF STAINLESS STEEL 304L TO TI-6AL-4V ALLOY, *PROC. 3RD INT. CONF. ISOSTATIC PRESSING* (LONDON, U.K.), 1986

Shear Loading

Effect of Microstructure. The base metals provide relatively little mechanical constraint of the interlayer during shear loading (shear stress is parallel to the plane of the interlayer) and, therefore, reveals the mechanical behavior of the interlayer material itself. The behavior of the interlayer to shear can be converted to equivalent uniaxial tensile behavior using the von Mises stress and strain criteria, $\bar{s} = t\sqrt{3}$, and $\bar{e} = \gamma \sqrt{3}$, where t and γ are shear stress and strain, respectively. Effective (von Mises) stress-strain ($\bar{s}$ -ε) behavior for a 150 μm (6 mil) thick (t/d = 0.024) silver joint, fabricated using PM sputter-deposition (between maraging steel) and tested in torsion is shown in Fig. 8 (Ref 33), along with the reported results for bulk polycrystalline (annealed) silver (Ref 38). The yield stress ($\bar{e}_p$ = 0.2%) of the solid-state-welded joint is six times greater than the bulk-silver value, presumably because of the much more refined, twinned substructure. The strain-hardening rate, $d\sigma_{eff}/d\varepsilon_{eff}$, at an effective stress of 100 MPa (15 ksi) is over nine times greater than bulk polycrystalline silver (Ref 38). The maximum interlayer stress is attained after a strain of only 0.3 (99% of the maximum value is attained after a strain of only 0.1). Therefore, although the interlayer silver has a higher yield stress and strain hardens at a higher rate, it reaches essentially the same maximum stress as bulk silver at a substantially lower strain.

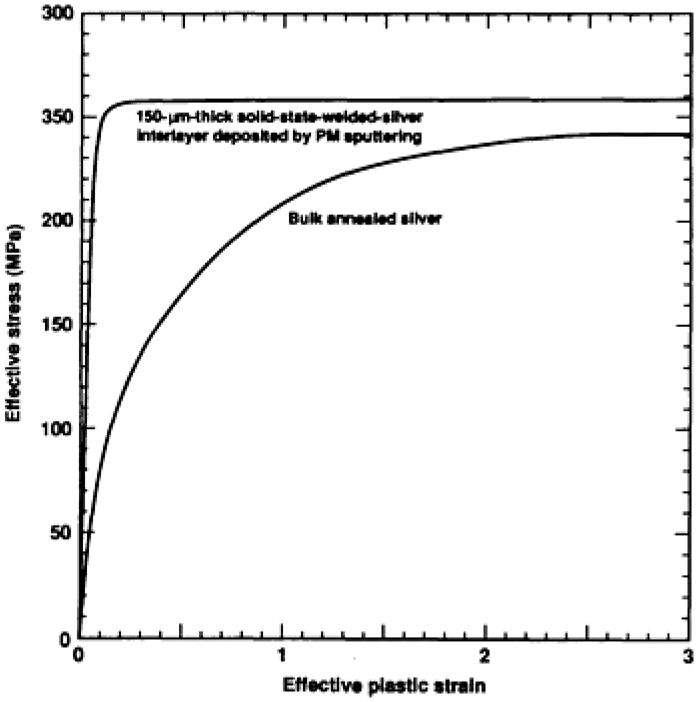

FIG. 8 EFFECTIVE (VON MISES) STRESS-STRAIN BEHAVIOR FOR A 150 μM (6 MIL) THICK SILVER INTERLAYER (*T*/*D* = 0.024), FABRICATED USING PM SPUTTER DEPOSITION, TESTED IN TORSION, ALONG WITH RESULTS REPORTED FOR BULK POLYCRYSTALLINE (ANNEALED) SILVER

Differences in the maximum (steady-state) stress may result from two factors. First, the interlayer microstructure may not be identical to that of the bulk material. The second explanation is that slip (plasticity) is not isotropic. The plasticity is not isotropic. The plasticity is the result of dislocation activity on limited slip planes with specific Burgers vectors. The shear displacements of individual grains in the interlayer must generally by accommodated by neighboring grains. However, near the base metal, the accommodation is not possible, and additional slip systems may operate in the interlayer grains near the base metal, leading to greater local hardening and higher flow stresses, Torsion tests of solid-state-welded silver interlayers coated by HHC evaporation, using uranium-to-stainless steel base metals (Ref 26), showed similar increases in the yield stress and strain-hardening rates, compared with those of bulk-silver values. However, the steady-state stress of silver joints was approximately 25% higher than that of bulk silver, compared with 105 higher values for interlayers using PM sputter-deposited silver. One explanation for this difference may be the presence of HHC posits, which may further raise the maximum, or steady-state, stress.

Time-Dependent Failure. Although not shown in Fig. 8, the plastic strain-to-failure in pure shear is very high ($\bar{e}_f = 5$ to 10) for thin, silver-interlayer specimens, just as for macroscopic, or bulk, silver specimens (Ref 33). This is nearly four orders of magnitude higher than for the case of tension. As previously discussed, cavity nucleation is promoted by the triaxial stress state established with simple uniaxial loading of constrained interlayers. The fact that torque saturates is consistent with the observation that some face-centered cubic (fcc) metals, such as aluminum, copper, and silver (Ref 38), show a saturation in the flow stress. This fact is important because such saturations imply constant torque, or time-dependent, shear-stress failures. Figure 9 shows a comparison between the rupture time versus effective (von Mises) stress behavior in shear (Ref 26) (at torsional stresses less than the maximum of HHC-deposited-silver interlayers), compared with that of tension (Ref 17) (calculated using FEM analyses of PM sputter-deposited silver interlayers). The shear rupture times are longer than tensile rupture times at equivalent uniaxial stresses (calculated using the von Mises relation), consistent with the higher plastic strains-to-failure that are much higher for the torsion specimens ($\bar{e}_f > 5$) than for the tensile specimens (ε_f ~0.001). The shorter rupture times for tensile joints are a direct result of the hydrostatic tensile stresses superimposed on the effective stresses caused by the mechanical constraint imposed by the base metal.

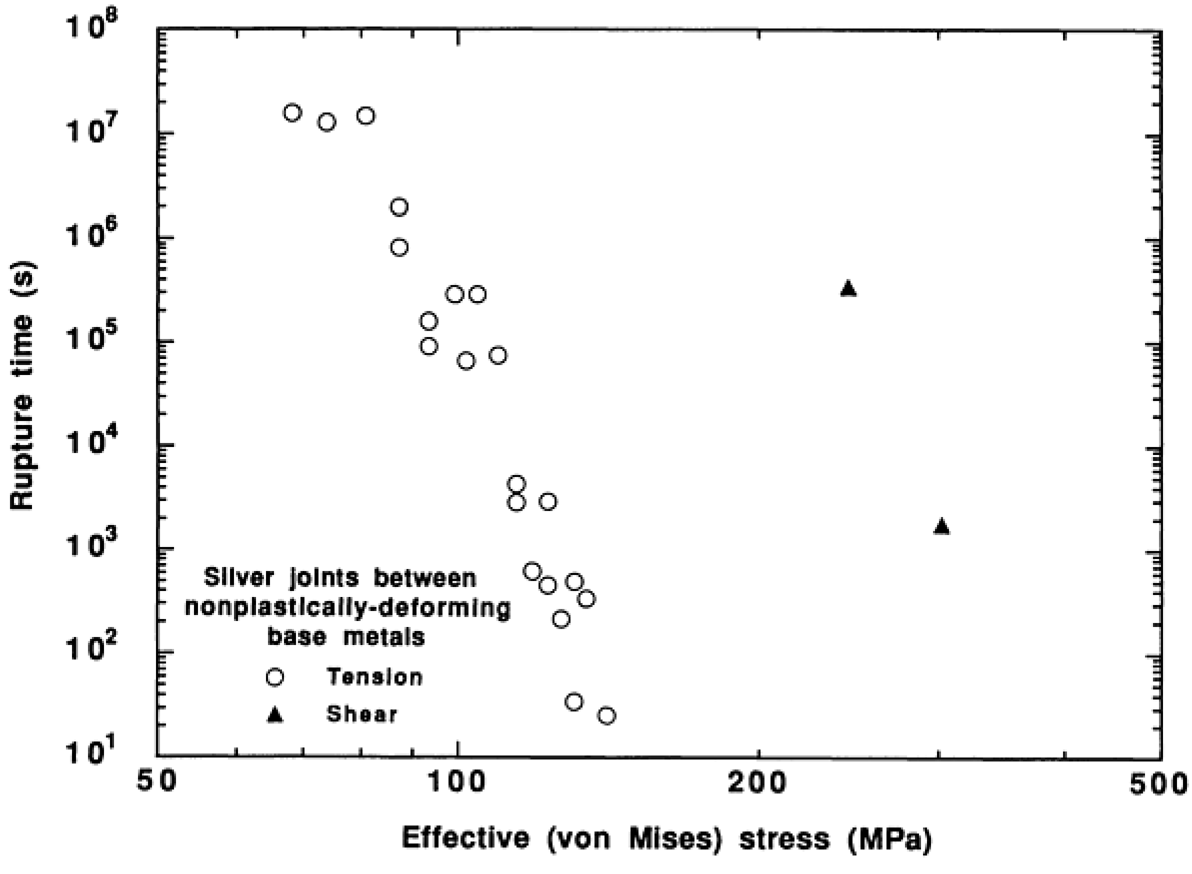

FIG. 9 COMPARISON BETWEEN THE RUPTURE TIME VERSUS EFFECTIVE (VON MISES) STRESS BEHAVIOR IN SHEAR, AT TORSIONAL STRESSES LESS THAN THE MAXIMUM OF HHC-DEPOSITED SILVER INTERLAYERS, COMPARED WITH THAT OF TENSION, CALCULATED USING FEM ANALYSES OF PM SPUTTER-DEPOSITED SILVER INTERLAYERS

The change in steady-state strain-rate (or time-to-rupture, according to the Monkman-Grant relationship) with applied stress is predictable according to the steady-state stress exponent, *n*,

$$n = \left(\frac{\partial \ln e}{\partial \ln s} \right) \quad \textbf{(EQ 2)}$$

where *n* is about 30 for silver at ambient temperature (Ref 38). Naturally, the steady-state strain rate of the torsional specimens shown in Fig. 9 is less than the constant applied strain rate of the torsional specimens shown in Fig. 8. If the steady-state rate if not achievable in the interlayer metal, then time-dependent failure in shear is not expected to occur.

It is important to realize the consequence of the difference in plastic strains-to-failure between the constant shear-stress case and the constant tensile-stress case. For example, residual shear or tensile stresses may develop during fabrication of interlayer brazes or solid-state welds. This is particularly true in cases where dissimilar base materials are joined, but have substantially different coefficients of thermal expansion. For thin interlayers subjected to residual shear stresses, the dimensions of the base metal and interlayer may be such that the stress is relaxed at strains less than the shear strain-to-failure. However, because the observed tensile strains-to-failure are small, creep rupture by residual tensile stresses is possible, provided a significant stress level exists. If the interlayer metal is resistant to tim-dependent failure, then only the indirect implications of residual stress are important (for example, environmentally induced failure).

References cited in this section

17. R.S. ROSEN, "TIME-DEPENDENT FAILURE OF SILVER INTERLAYER WELDS," PH.D.THESIS, UNIVERSITY OF CALIFORNIA, DAVIS, 1990
26. J.W. ELMER, M.E. KASSNER, AND R.S. ROSEN, THE BEHAVIOR OF SILVER-AIDED DIFFUSION-WELDED JOINTS UNDER TENSILE AND TORSIONAL LOADS, *WELD. J.*, VOL 67 (NO. 7), 1988, P 157S-162S
33. M.E. KASSNER, R.S. ROSEN, AND G.A. HENSHALL, DELAYED MECHANICAL FAILURE OF SILVER-INTERLAYER DIFFUSION BONDS, *METALL. TRANS. A*, VOL 21, 1990, P 3085-3100
38. M.E. KASSNER, THE RATE DEPENDENCE AND MICROSTRUCTURE OF HIGH-PURITY SILVER DEFORMED TO LARGE STRAINS BETWEEN 0.16 AND 0.30 TM, *METALL. TRANS. A*, VOL 20, 1989, P 2001-2010

R.S. Rosen, Lawrence Livermore National Laboratory; M.E. Kassner, Oregon State University

Multiaxial Loading

As discussed earlier, in tension there is significant mechanical constraint by the base material, which tends to reduce the effective stress. In torsion, where the shear stress is parallel to the plane of the interlayer, there is not constraint, other than, perhaps, the relatively minor constraint associated with plastic strain incompatibilities between the interlayer and the base metal. It was also mentioned earlier that the equivalent uniaxial strain-to-failure for torsional deformation of soft-interlayer joints is roughly four orders of magnitude higher than the values for uniaxial tension, and the (effective) shear stress-to-failure is only about one-half the UTS. This suggests that the interlayers will behave anisotropically to imposed stresses. The mechanical properties of the intermediate cases of biaxial deformation are not readily interpolated between these limits. The authors have preliminary experimental evidence, for example, that low-strain ductile failures in thin interlayers may occur at particularly low stress levels and strains for cases where tensile loads are accompanied by in-plane shear stresses (Ref 63). This implies that some multiaxial residual stress states may render solid-state-welded interlayer joints substantially more vulnerable to failure than simple tensile-stress states, even though the "macroscopic" effective stress may be comparable, when calculated on the basis of the applied stress state in the base material away from the interlayer.

Reference cited in this section

63. M.E. KASSNER, OREGON STATE UNIVERSITY, UNPUBLISHED RESEARCH, 1989

R.S. Rosen, Lawrence Livermore National Laboratory; M.E. Kassner, Oregon State University

Environmentally Induced Failure of Interlayers

The interlayer/base-metal interfaces of some soft-interlayer welds appear particularly vulnerable to failure when exposed to external (corrosive) environments. For example, it is known that brazed silver interlayers between stainless steel are subject to galvanic corrosion in aqueous NaCl media (Ref 64, 65). When dissimilar base materials are joined, the potential for stress-corrosion cracking (SCC) exists if the joints are exposed simultaneously to some external (corrosive) environments and either shear or tensile stresses. If an interlayer is subjected to stress, then a determination of the vulnerability to SCC is appropriate. It was mentioned earlier that elevated-temperature joining using either solid-state or brazing processes may leave the joint under substantial residual stress, particularly in the case of dissimilar metals for which the coefficients of thermal expansion are significantly different. The authors have reported (Ref 35) that the uranium to type 304 stainless steel solid-state welds, fabricated using HHC-deposited silver interlayers (Ref 25, 26), are very susceptible to SCC. These silver solid-state welds were joined at 873 K, and the difference in thermal expansion coefficients leaves the joints of cylindrical specimens under significant residual shear stresses after cooling to ambient temperature. It was found that these residual stresses led to SCC at the uranium/silver interface in air saturated with water vapor (100% relative humidity) (Ref 35). This cracking resulted in reduced tensile strengths and rupture times, when compared with those specimens tested in laboratory air (~40% relative humidity). Specimens subjected to the same 100% relative humidity in air in an unstressed state exhibited no loss in strength, even after lengthy exposure. This, or course, emphasizes the importance of considering the residual stresses as not only rendering the interlayer joints vulnerable to "pure mechanical" rupture, but also to environmentally induced failures such as SCC, hydrogen embrittlement, and metal-induced embrittlement, if exposed to critical environments.

References cited in this section

25. R.S. ROSEN, D.R. WALMSLEY, AND Z.A. MUNIR, THE PROPERTIES OF SILVER-AIDED DIFFUSION WELDS BETWEEN URANIUM AND STAINLESS STEEL, *WELD. J.,* VOL 65 (NO. 4), 1986, P 83S-92S
26. J.W. ELMER, M.E. KASSNER, AND R.S. ROSEN, THE BEHAVIOR OF SILVER-AIDED DIFFUSION-WELDED JOINTS UNDER TENSILE AND TORSIONAL LOADS, *WELD. J.,* VOL 67 (NO. 7), 1988, P 157S-162S
35. R.S. ROSEN, S. BEITSCHER, AND M.E. KASSNER, STRESS CORROSION CRACKING OF URANIUM-SILVER INTERFACES IN SILVER-AIDED DIFFUSION WELDS, *INT. CONF. ENVIRONMENT-INDUCED CRACKING OF METALS,* VOL 10, NACE, 1989, P 429-433
64. A.T. KUHN AND R.M. TRIMMER, REVIEW OF THE AQUEOUS CORROSION OF STAINLESS STEEL-SILVER BRAZED JOINTS, *BR. CORROS. J.,* VOL 17 (NO. 1), 1982, P 4-8
65. T. TAKEMOTO AND I. OKAMOTO, EFFECT OF COMPOSITION ON THE CORROSION BEHAVIOR OF STAINLESS STEELS BRAZED WITH SILVER-BASE FILLER METALS, *WELD. J.,* VOL 64, 1984, P 300S-307S

R.S. Rosen, Lawrence Livermore National Laboratory; M.E. Kassner, Oregon State University

References

1. J.T. NIEMANN, R.P. SOPHER, AND P.J. RIEPPEL, DIFFUSION BONDING BELOW 1000 °F, *WELD. J.,* VOL 37, 1958, P 337S-342S
2. I.M. BARTA, LOW TEMPERATURE DIFFUSION BONDING OF ALUMINUM ALLOYS, *WELD. J.,* VOL 43, 1964, P 241S-247S

3. D. HAUSER, P.A. KAMMER, AND J.H. DEDRICK, SOLID-STATE WELDING OF ALUMINUM, *WELD. J.,* VOL 46, 1967, P 11S-22S
4. P.A. KAMMER, R.E. MONROE, AND D.C. MARTIN, FURTHER STUDIES OF DIFFUSION BONDING BELOW 1000 °F, *WELD. J.,* VOL 48, 1969, P 116S-124S
5. Y. IINO AND N. TAGAUCHI, INTERDIFFUSING METALS LAYER TECHNIQUE OF CERAMIC-METAL BONDING, *J. MATER. SCI. LETT.,* VOL 7 (NO. 9), 1988, P 981-982
6. A. URENA, J.M.G.D. SALAZAR, AND J. QUINONES, DIFFUSION BONDING OF ALUMINA TO STEEL USING SOFT COPPER INTERLAYER, *J. MATER. SCI.,* VOL 27, 1992, P 599-606
7. T. YAMADA K, YOKOI, AND A. KOHNO, EFFECT OF RESIDUAL STRESS ON THE STRENGTH OF ALUMINA-STEEL JOINT WITH AL-SI INTERLAYER, *J. MATER. SCI.,* VOL 25, 1990, P 2188-2192
8. M.G. NICHOLAS AND R.M. CRISPIN, DIFFUSION BONDING STAINLESS STEEL TO ALUMINA USING ALUMINUM INTERLAYERS, *J. MATER. SCI.,* VOL 17, 1982, P 3347-3360
9. C.L. CLINE, AN ANALYTICAL AND EXPERIMENTAL STUDY OF DIFFUSION BONDING, *WELD. J.,* VOL 45, 1966, P 481S-489S
10. A.T. D'ANNESSA, THE SOLID-STATE BONDING OF REFRACTORY METALS, *WELD. J.,* VOL 43, 1964, P 232S-240S
11. M.G. NICHOLAS AND R.M. CRISPIN, DIFFUSION BONDING CERAMICS WITH DUCTILE METAL INTERLAYERS, *PROC. INT. DIFFUSION BONDING* (CRANFIELD, U.K.), 1987, P 173-182
12. B. DERBY, ZIRCONIA/METAL DIFFUSION BONDS, *PROC. INT. CONF. DIFFUSION BONDING* (CRANFIELD, U.K.), 1987, P 195-201
13. R.A. MORLEY AND J. CARUSO, THE DIFFUSION WELDING OF 390 ALUMINUM ALLOY HYDRAULIC VALVE BODIES, *WELD. J.,* VOL 59 (NO. 8), 1980, P 29-34
14. C.H. CRANE, D.T. LOVELL, W.A. BAGINSKY, AND M.G. OLSEN, DIFFUSION WELDING OF DISSIMILAR METALS, *WELD. J.,* VOL 46, 1967, P 23S-31S
15. M. O'BRIEN, C.R. RICE, AND D.L. OLSON, HIGH STRENGTH DIFFUSION WELDING OF SILVER COATED BASE METALS, *WELD. J.,* VOL 55 (NO. 1), 1976, P 25-27
16. J.W. DINI, W.K. KELLEY, W.C. COWDEN, AND E.M. LOPEZ, USE OF ELECTRODEPOSITED SILVER AS AN AID IN DIFFUSION WELDING, *WELD. J.,* VOL 63 (NO. 1), 1983, P 26S-34S
17. R.S. ROSEN, "TIME-DEPENDENT FAILURE OF SILVER INTERLAYER WELDS," PH.D.THESIS, UNIVERSITY OF CALIFORNIA, DAVIS, 1990
18. J.L. KNOWLES AND T.H. HAZLETT, HIGH-STRENGTH LOW-TEMPERATURE BONDING OF BERYLLIUM AND OTHER METALS, *WELD. J.,* VOL 49 (NO. 7), 1970, P 301S-310S
19. P.S. MCLEOD AND G. MAH, THE EFFECT OF SUBSTRATE BIAS VOLTAGE ON THE BONDING OF EVAPORATED SILVER COATINGS, *J. VAC. SCI. TECHNOL.,* VOL 11 (NO. 1), 1974, P 119-121
20. G. MAH, P.S. MCLEOD, AND D.G. WILLIAMS, CHARACTERIZATION OF SILVER COATINGS DEPOSITED FROM A HOLLOW CATHODE SOURCE, *J. VAC. SCI. TECHNOL.,* VOL 11 (NO. 4), 1974, P 663-665
21. D.G. WILLIAMS, VACUUM COATING WITH A HOLLOW CATHODE SOURCE, *J. VAC. SCI. TECHNOL.,* VOL 11 (NO. 1), 1974, P 374-377
22. E.R. NAIMON, D. VIGIL, J.P. VILLEGAS, AND L. WILLIAMS, ADHESION STUDY OF SILVER FILMS DEPOSITED FROM A HOY HOLLOW-CATHODE SOURCE, *J. VAC. SCI TECHNOL.,* VOL 13 (NO. 6), 1976, P 1131-1135
23. E.R. NAIMON, J.H. DOYLE, C.R. RICE, D. VIGIL, AND D.R. WALMSLEY, DIFFUSION WELDING OF ALUMINUM TO STAINLESS STEEL, *WELD. J.,* VOL 60 (NO. 1), 1981, P 17-20
24. D.T. LARSON AND H.L. DRAPER, CHARACTERIZATION OF THE BE-AG INTERFACIAL REGION OF SILVER FILMS DEPOSITED ONTO BERYLLIUM USING A HOT HOLLOW CATHODE DISCHARGE, *THIN SOLID FILMS,* VOL 107, 1983, P 327-334
25. R.S. ROSEN, D.R. WALMSLEY, AND Z.A. MUNIR, THE PROPERTIES OF SILVER-AIDED

DIFFUSION WELDS BETWEEN URANIUM AND STAINLESS STEEL, *WELD. J.,* VOL 65 (NO. 4), 1986, P 83S-92S

26. J.W. ELMER, M.E. KASSNER, AND R.S. ROSEN, THE BEHAVIOR OF SILVER-AIDED DIFFUSION-WELDED JOINTS UNDER TENSILE AND TORSIONAL LOADS, *WELD. J.,* VOL 67 (NO. 7), 1988, P 157S-162S
27. J. HARVEY, P.G. PATRIDGE, AND A.M. LURSHEY, FACTORS AFFECTING THE SHEAR STRENGTH OF SOLID STATE DIFFUSION BONDS BETWEEN SILVER-COATED CLAD AL-ZN-MG ALLOY (ALUMINUM ALLOY 70100, *MATER. SCI. ENG.,* VOL 79, 1986, P 191-199
28. P.G. PARTRIDGE AND D.V. DUNFORD, ON THE TESTING OF DIFFUSION-BONDED OVERLAP JOINTS BETWEEN CLAD AL-ZN-MG ALLOY (7010) SHEET, *J. MATER. SCI.* VOL 22, 1987 P 1597-1608
29. D.V. DUNFORD AND P.G. PARTRIDGE, THE PEEL STRENGTHS OF DIFFUSION BONDED. JOINTS BETWEEN CLAD AL-ALLOY SHEETS, *J. MATER. SCI.,* VOL 22, 1987, P 1790-1798
30. R.S. ROSEN AND M.E. KASSNER, DIFFUSION WELDING OF SILVER INTERLAYERS COATED ONTO BASE METALS BY PLANAR-MAGNETRON SPUTTERING, *J. VAC. SCI. TECHNOL. A,* VOL 8 (NO. 1), 1900, P 19-29
31. M.E. KASSNER, R.S. ROSEN, G.A. HENSHALL, AND W.E. KING, DELAYED FAILURE OF SILVER AIDED DIFFUSION WELDS BETWEEN STEEL, *PROC. 2ND INT. CONF. BRAZING, HIGH TEMPERATURE BRAZING, AND DIFFUSION WELDING,* GERMAN WELDING SOCIETY, 1989, P 47-52
32. M.E. KASSNER, R.S. ROSEN, G.A. HENSHALL, AND K.D. CHALLENGER, TIME-DEPENDENT FAILURE OF SILVER-INTERLAYER DIFFUSION WELDS BETWEEN ELASTICALLY-DEFORMING BASE METALS, *SCR. METALL. MATER.,* VOL 24, 1990, P 587-592
33. M.E. KASSNER, R.S. ROSEN, AND G.A. HENSHALL, DELAYED MECHANICAL FAILURE OF SILVER-INTERLAYER DIFFUSION BONDS, *METALL. TRANS. A,* VOL 21, 1990, P 3085-3100
34. P.G. PARTRIDGE AND J. HARVEY, "PROCESS FOR THE DIFFUSION BONDING OF ALUMINUM MATERIALS," U.K. PATENT APPLICATIONS GB 2117691 A, 1983
35. R.S. ROSEN, S. BEITSCHER, AND M.E. KASSNER, STRESS CORROSION CRACKING OF URANIUM-SILVER INTERFACES IN SILVER-AIDED DIFFUSION WELDS, *INT. CONF. ENVIRONMENT-INDUCED CRACKING OF METALS,* VOL 10, NACE, 1989, P 429-433
36. E. OROWAN, FRACTURE AND STRENGTH OF SOLIDS, *REP. PROG. PHYS.,* VOL 12, 1948, P 185-231
37. E. OROWAN, FUNDAMENTALS OF BRITTLE BEHAVIOR OF METALS, *FATIGUE AND FRACTURE OF METALS,* JOHN WILEY & SONS, 1952
38. M.E. KASSNER, THE RATE DEPENDENCE AND MICROSTRUCTURE OF HIGH-PURITY SILVER DEFORMED TO LARGE STRAINS BETWEEN 0.16 AND 0.30 TM, *METALL. TRANS. A,* VOL 20, 1989, P 2001-2010
39. N. BREDZ, INVESTIGATION OF FACTORS DETERMINING THE TENSILE STRENGTH OF BRAZED JOINTS, *WELD. J.,* VOL 33, 1954, P 545S-563S
40. W.G. MOFFATT AND J. WULFF, TENSILE DEFORMATION AND FRACTURE OF BRAZED JOINTS, *WELD. J.,* VOL 42, 1963, P 115S-125S
41. H.J. SAXTON, A.J. WEST, AND C.R. BARRETT, DEFORMATION AND FAILURE OF BRAZED JOINTS--MACROSCOPIC CONSIDERATIONS, *METALL. TRANS.,* VOL 2, 1971, P 999-1007
42. A.J. WEST, H.J. SAXTON, A.S. TETELMAN, AND C.R. BARRETT, DEFORMATION AND FAILURE OF THIN BRAZED JOINTS--MICROSCOPIC CONSIDERATIONS, *METALL. TRANS.,* VOL 2, 1971, P 1009-1017
43. E.P. LAUTENSCHLAGER, B.C. MARKER, B.K. MOORE, AND R. WILDES, STRENGTH MECHANISMS OF DENTAL SOLDER JOINTS, *J. DENT. RES.,* VOL. 53, 1974, P 1361-1367
44. R.A. MUSIN, V.A. ANTSIVEROV, Y.A. BELIKOV, Y.V. LYAMIN, AND A.N. SOLOKOV, THE EFFECTS ON THE STRENGTH OF THE JOINT OF THE THICKNESS OF THE SOFT INTERLAYER

IN DIFFUSION WELDING, *AUTO. WELD.,* VOL 32, 1979, P 38-40

45. R.M. TRIMMER AND A.T. KUHN, THE STRENGTH OF SILVER-BRAZED STEEL JOINTS--A REVIEW, *BRAZING SOLDERING,* VOL 2, 1982 P 6-13
46. N. BREDZ AND H. SCHWARTZBART, TRIAXIAL TENSION TESTING AND THE BRITTLE FRACTURE STRENGTH OF METALS, *WELD. J.,* VOL 53 (NO. 12), 1956, P 610S-615S
47. W.G. MOFFATT AND J. WULFF, STRENGTH OF SILVER BRAZED JOINTS IN MILD STEEL, *TRANS. AIME,* VOL 209 (NO. 4), 1957, P 442-445
48. R.Z. SHRON AND O.A. BAKSHI, THE PROBLEMS OF GAUGING THE STRENGTHS OF WELDED JOINTS IN WHICH THERE IS A SOFT INTERLAYER, *WELD. PROD.,* VOL 9, 1962, P 19-23
49. O.A. BAKSHI AND A.A. SHATOV, THE STRESSED STATE IN WELDED JOINTS WITH HARD AND SOFT INTERLAYERS, *WELD PROD.,* VOL 13, 1966, P 13-19
50. V.S. GOLOVCHENKO AND B.A. GOLOBOV, THE USE OF SILVER AS AN INTERMEDIATE LAYER FOR JOINING TITANIUM TO OTHER METALS, *WELD. PROD.,* VOL 18, 1971, P 55-57
51. G.K. KHARCHENKO AND A.I. IGNATENKO, THE STRENGTH OF JOINTS WITH A THIN SOFT INTERLAYER, *AUTO. WELD.,* VOL 21, 1968, P 33-35
52. W.M. LEHER AND H. SCHWARTZBART, STATIC AND FATIGUE STRENGTHS OF METALS SUBJECTED TO TRIAXIAL STRESSES, VOL 60, 1960, P 610-626
53. E.A. ALMOND, D.K. BROWN, G.J. DAVIES, AND A.M. COTTENDED, THEORETICAL AND EXPERIMENTAL INTERLAYED BUTT JOINTS TESTED IN TENSION, *INT. J. MECH. SCI.,* VOL 25 (NO. 3), 1983, P 175-189
54. G.A. HENSHALL, R.S. ROSEN, M.E. KASSNER, AND R.G. WHIRLEY, FINITE ELEMENT ANALYSIS OF INTERLAYER WELDS LOADED IN TENSION, *WELD. J.,* VOL 69 (NO. 9) 1990, P 337S-345S
55. H.J. SAXTON, A.J. WEST, AND C.R. BARRETT, THE EFFECT OF COOLING RATE ON THE STRENGTH OF BRAZED JOINTS, *METALL. TRANS.,* VOL 2, 1971, P 1019-1028
56. M.C. TOLLE AND M.E. KASSNER, TENSILE PROPERTIES OF THIN AU-NI BRAZES BETWEEN STRONG BASE METALS, *SCR. METALL. MATER.,* VOL 26, 1992, P 1281-1284
57. M.E. KASSNER, M.C. TOLLE. R.S. ROSEN, G.A. HENSHALL, AND J.W. ELMER, RECENT ADVANCES IN UNDERSTANDING THE MECHANICAL BEHAVIOR OF CONSTRAINED THIN METALS IN BRAZES AND SOLID-STATE BONDS, *THE METAL SCIENCE OF JOINING,* M.J. CIESLAK, ED., TMS, 1992, P 223-232
58. R.W. LOGAN, R.G. CASTRO, AND A.K. MUKHERJEE, MECHANICAL PROPERTIES OF SILVER AT LOW TEMPERATURES, *SCR. METALL.,* VOL 17, 1983, P 63-66
59. T.J. MOORE AND K.H. HOLKO, SOLID-STATE WELDING OF TD-NICKEL BAR, *WELD. J.,* VOL 49 (NO. 9), 1970, P 395S-409S
60. R.S. ROSEN, "AN INVESTIGATION OF THE PROPERTIES OF A SILVER-AIDED SOLID-STATE BOND BETWEEN URANIUM AND STAINLESS STEEL," REPORT UCRL-53458, LAWRENCE LIVERMORE NATIONAL LABORATORY, 1983
61. E.R. NAIMON, R.G. KURZ, D. VIGIL, AND L. WILLIAMS, "SILVER FILMS FOR SOLID STATE BONDING,"REPORT RFP-3125, ROCKWELL INTERNATIONAL, 1981
62. Z. NISENHOLTZ, J. MIRONI, AND N. NIR, DIFFUSION BONDING OF STAINLESS STEEL 304L TO TI-6AL-4V ALLOY, *PROC. 3RD INT. CONF. ISOSTATIC PRESSING* (LONDON, U.K.), 1986
63. M.E. KASSNER, OREGON STATE UNIVERSITY, UNPUBLISHED RESEARCH, 1989
64. A.T. KUHN AND R.M. TRIMMER, REVIEW OF THE AQUEOUS CORROSION OF STAINLESS STEEL-SILVER BRAZED JOINTS, *BR. CORROS. J.,* VOL 17 (NO. 1), 1982, P 4-8
65. T. TAKEMOTO AND I. OKAMOTO, EFFECT OF COMPOSITION ON THE CORROSION BEHAVIOR OF STAINLESS STEELS BRAZED WITH SILVER-BASE FILLER METALS, *WELD. J.,* VOL 64, 1984, P 300S-307S

Shielded Metal Arc Welding

Raymond H. Juers, Naval Surface Warfare Center

Introduction

SHIELDED METAL ARC WELDING (SMAW), commonly called stick, or covered electrode, welding, is a manual welding process whereby an arc is generated between a flux-covered consumable electrode and the workpiece. The process uses the decomposition of the flux covering to generate a shielding gas and to provide fluxing elements to protect the molten weld-metal droplets and the weld pool.

Shielded Metal Arc Welding

Raymond H. Juers, Naval Surface Warfare Center

The SMAW Process

The important features of the SMAW process are shown in Fig. 1. The arc is initiated by momentarily touching or "scratching" the electrode on the base metal. The resulting arc melts both the base metal and the tip of the welding electrode. The molten electrode metal/flux is transferred across the arc (by arc forces) to the base-metal pool, where it becomes the weld deposit covered by the protective, less-dense slag from the electrode covering.

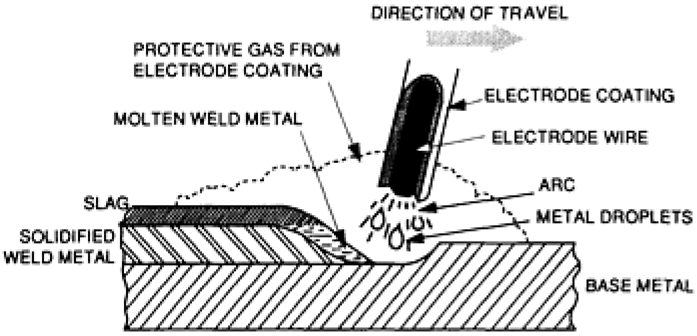

FIG. 1 SMAW PROCESS

Advantages and Limitations. The SMAW process is the most widely used welding process. It is the simplest, in terms of equipment requirements, but it is, perhaps, the most difficult in terms of welder training and skill-level requirements. Although welder skill level is a concern, most welders entering the field start as "stick welders" and develop the necessary skills through training and experience. The equipment investment is relatively small, and welding electrodes (except the very reactive metals, such as titanium, magnesium, and others) are available for virtually all manufacturing, construction, or maintenance applications. Shielded metal arc welding has the greatest flexibility of all the welding processes, because it can be used in all positions (flat, vertical, horizontal, and overhead), with virtually all base-metal thicknesses (1.6 mm, or $\frac{1}{16}$ in., and greater), and in areas of limited accessibility, which is a very important capability.

Because the SMAW process is basically a manual process, the skill level of the welder is of paramount importance in obtaining an acceptable weld. The welder duty cycle is generally low, because of the built-in work break, which occurs after each electrode is consumed and requires replacement. In addition to replacing the electrode when the arc is stopped (broken), the welder may "chip" or remove slag and clean it away from the starting and welding area with a wire brush to allow the proper deposition of the subsequent weld. This electrode replacement and cleaning operation occurs many times during the work day (about every two minutes, or the time it generally takes to consume an electrode). This stopping, chipping, wire brushing, and electrode replacement prevents the welder from attaining an operator factor, or duty cycle, that is much greater than 25%.

Weld Quality. The quality of the weld depends on the design and accessibility of the joint, as well as on the electrode, the technique, and the skill of the welder. If joint details vary greatly from established design details, then a lower-quality weld can result. Other factors that also reduce quality are improper interbead cleaning, poor location of individual weld beads within the joint, and various problems with individual electrodes, including partially missing flux and core wires that are not centered within the flux covering. Overall, welds of excellent quality can be obtained with the SMAW process, as demonstrated by its use in joining submarine pressure hull sections and high-pressure oil/gas pipe lines.

The base-metal thicknesses. that can be welded using the SMAW process generally range from 1.6 mm ($\frac{1}{16}$ in.) to an unlimited thickness. The thinner materials require a skilled welder, tight fitup, and the proper small-diameter welding electrode. Welding position also is important when determining the minimum plate thicknesses that can be welded. Flat-position butt welds and horizontal fillet welds are generally considered the easiest to weld. Out-of-position welding (vertical, overhead) requires greater skill.

Welding Circuit. The circuit diagram for the SMAW process is shown in Fig. 2. The equipment consists of a power source, electrode holder, and welding cables that connect the power source to the electrode holder and the workpiece. Alternating current (ac), or direct current, electrode negative (DCEN), or direct current, electrode positive (DCEP) can be used, depending on the electrode coating characteristics. The DCEN source is also called dc straight polarity, whereas the DCEP source is also called dc reverse polarity.

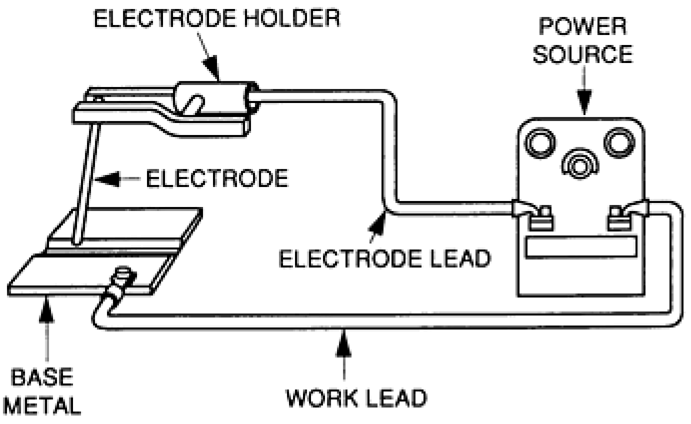

FIG. 2 SHIELDED METAL ARC CIRCUIT DIAGRAM

Equipment. The welding machine, or power source, is the crux of the SMAW process. Its primary purpose is to provide electrical power of the proper current and voltage to maintain a controllable and stable welding arc. Its output characteristics must be of the constant current (CC) type. SMAW electrodes operate within the range from 25 to 500 A. The electrode producer should suggest a narrow optimum range for each size and type of electrode. Operating arc voltage varies between 15 and 35 V.

The electrode holder, which is held by the welder, firmly grips the electrode and transmits the welding current to it. Electrode holders are available in several designs, such as the pincher type and the collet, or twist, type, shown in Fig. 3. Each style has its proponents and the selection is usually a personal preference. Electrode holders are designated by their current capacity. Selection factors, such as the current rating, duty cycle, maximum electrode size, and cable size, are shown in Table 1. The most lightweight holder that will accommodate the required electrode size is usually desired.

TABLE 1 SIZE AND CAPACITY OF ELECTRODE HOLDERS

ELECTRODE HOLDER	RATING	MAXIMUM	MAXIMUM

			ELECTRODE SIZE		CABLE SIZE
	MAXIMUM CURRENT, A	DUTY CYCLE, %	mm	in.	
SMALL	100	50	3.2	$\frac{1}{8}$	1
	200	50	4.0	$\frac{5}{32}$	1/0
MEDIUM	300	60	5.5	$\frac{7}{32}$	2/0
LARGE	400	60	6.4	$\frac{1}{4}$	3/0
EXTRA LARGE	500	75	7.9	$\frac{5}{16}$	4/0
	600	75	9.5	$\frac{3}{8}$	4/0

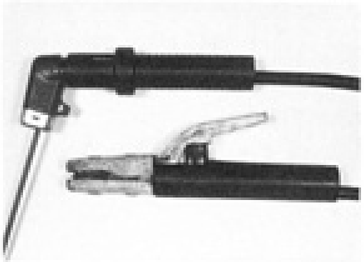

FIG. 3 SMAW ELECTRODE HOLDERS

All electrode holders should be fully insulated. Because they are used in proximity to the arc and are exposed to high heat, they will deteriorate rapidly. It is extremely important to maintain electrode holders to ensure that they retain their current-carrying efficiency, their insulating qualities, and their electrode gripping action. Manufacturers supply spare parts so that the holders can be rebuilt and maintained for safe and efficient operation.

Certain pieces of auxiliary equipment can be used with the SMAW process, such as low-voltage control circuits, which enable the relatively high open-circuit voltage to be cut off until the electrode touches the workpiece. Other items include remote-control switches for the contactors, remote-control current adjusting devices, and engine idling controllers for engine-driven power sources.

Shielded Metal Arc Welding

Raymond H. Juers, Naval Surface Warfare Center

Applications

Most manufacturing operations that require welding will strive to utilize the mechanized processes that offer greater productivity, higher quality, and, therefore, more cost-effective production. For these reasons, the SMAW process has been replaced where possible. However, the simplicity and ability of the SMAW process to achieve welds in areas of restricted accessibility means that it still finds considerable use in certain situations and applications. Heavy construction, such as shipbuilding, and welding "in the field," away from many support services that would provide shielding gas, cooling water, and other necessities, rely on the SMAW process to a great extent.

Although the SMAW process finds wide application for welding virtually all steels and many of the nonferrous alloys, it is primarily used to join steels. This family of materials includes low-carbon or mild steels, low-alloy steels, high-strength steels, quenched and tempered steels, high-alloy steels, stainless steels, and many of the cast irons. The SMAW process is also used to join nickel and its alloys and, to a lesser degree, copper and its alloys. It can be, but rarely is, used for welding aluminum.

In addition to joining metals, the SMAW process is frequently used for the protective surfacing of base metals. The surfacing deposit can be applied for the purpose of corrosion control or wear resistance (hard surfacing).

Shielded Metal Arc Welding

Raymond H. Juers, Naval Surface Warfare Center

Electrodes

The electrodes used in the SMAW process have many different compositions of core wire and a wide variety of flux-covering types and weights. Standard electrode diameters of the core wire range from 1.6 to 8 mm ($\frac{1}{16}$ to $\frac{5}{16}$ in.). Electrode length usually ranges from 230 to 455 mm (9 to 18 in.); the shorter lengths are associated with the smaller-diameter electrodes. A bare, uncoated end of the electrode (the grip end) is provided for making electrical contact in the electrode holder.

The coating on the electrode has numerous functions. It provides:

- GAS (NORMALLY, CARBON DIOXIDE), FROM THE DECOMPOSITION OF CERTAIN COATING INGREDIENTS TO SHIELD THE ARC AND WELD ZONE FROM THE ATMOSPHERE
- DEOXIDIZERS, FOR SCAVENGING AND PURIFYING THE DEPOSITED WELD METAL
- SLAG FORMERS, TO PROTECT THE DEPOSITED WELD METAL FROM ATMOSPHERIC OXIDATION AND TO HELP SHAPE THE WELD BEAD
- IONIZING ELEMENTS, TO MAKE THE ARC MORE STABLE AND TO OPERATE WITH ALTERNATING CURRENT
- ALLOYING ELEMENTS, TO PROVIDE SPECIAL CHARACTERISTICS TO THE WELD DEPOSIT
- IRON POWDER, IN CERTAIN ELECTRODES, TO INCREASE PRODUCTIVITY FOR WELDING FERROUS METALS

The American Welding Society (AWS) has established a system for identifying and classifying the different types of welding electrodes. All SMAW electrodes have the prefix letter E to indicate welding electrode. The symbols that follow the prefix are based on criteria that best describe the welding capabilities of the electrode metal. These criteria include chemical composition of the deposited weld metal, weld-metal mechanical properties, certain process parameters, or combinations of all factors.

Mild and Low-Alloy Steel-Covered Electrodes. The prefix used to identify these electrodes is followed by a number series that indicates minimum strength level, position capability, and type of covering and welding current. Table 2 explains how the number series is used in AWS A5.1, the specification for carbon steel electrodes for shielded metal arc welding and AWS A5.5, the specification for low-alloy steel electrodes. The first two digits after the E in the E6010 electrode designate a tensile strength of at least 430 MPa (62 ksi) for the deposited metal in the as-welded condition. The third digit indicates the position in which satisfactory welds can be made with the electrode. Thus, the 1 in E6010, for example, means that the electrode is satisfactory for use in all positions (flat, vertical, horizontal, and overhead). The 2 in E6020 indicates that the electrode is suitable for the flat position and horizontal fillets. The last digit or last two digits, taken together, indicate the applicable current type to be used and the type of covering on the electrode.

TABLE 2 CARBON AND LOW-ALLOY STEEL COVERED ELECTRODE IDENTIFICATION SYSTEM

AWS CLASSIFICATION[(A)]	MINIMUM TENSILE STRENGTH		MINIMUM YIELD STRENGTH		MINIMUM ELONGATION, %
	MPA	KSI	MPA	KSI	

E70XX	480-500	70-72	390-420	57-60	17-25
E80XX	550	80	460-550	67-80	16-24
E90XX	620	90	530-620	77-90	14-24
E100XX	690	100	600	87	13-20
E110XX	760	110	670-760	97-110	15-20
E120XX	830	120	740-830	107-120	14-18

CLASSIFICATION[(B)]	**FLAT POSITION**	**HORIZONTAL POSITION**	**VERTICAL POSITION**	**OVERHEAD POSITION**
EXXIX	YES	YES	YES	YES
EXX2X	YES	FILLET	NO	NO
EXX4X	YES	YES	DOWN	YES

CLASSIFICATION[(C)]	**CURRENT**	**ARC**	**PENETRATIONN**	**COVERING/SLAG**	**APPROXIMATE IRON POWDER[(D)], %**
EXX10	DCEP	DIGGING	DEEP	CELLULOSE/SODIUM	0-10
EXXX1	AC AND DCEP	DIGGING	DEEP	CELLULOSE/POTASSIUM	0
EXXX2	AC AND DCEN	MEDIUM	MEDIUM	RUTILE/SODIUM	0-10
EXXX3	AC AND DC	LIGHT	LIGHT	RUTILE/POTASSIUM	0-10
EXXX4	AC AND DC	LIGHT	LIGHT	RUTILE/IRON POWDER	25-40
EXXX5	DCEP	MEDIUM	MEDIUM	LOW HYDROGEN/SODIUM	0
EXXX6	AC OR DCEP	MEDIUM	MEDIUM	LOW HYDROGEN/POTASSIUM	0
EXXX8	AC OR DCEP	MEDIUM	MEDIUM	LOW HYDROGEN/IRON POWDER	25-40
EXX20	AC OR DC	MEDIUM	MEDIUM	IRON OXIDE/SODIUM	0
EXX24	AC OR DC	LIGHT	LIGHT	RUTILE/IRON POWDER	50
EXX27	AC OR DC	MEDIUM	MEDIUM	IRON OXIDE/IRON POWDER	50
EXX28	AC OR	MEDIUM	MEDIUM	LOW HYDROGEN/IRON	50

(A) FIRST TWO OR THREE DIGITS INDICATE TENSILE STRENGTH IN UNITS OF KSI AND OTHER MECHANICAL PROPERTIES (MECHANICAL PROPERTY REQUIREMENTS VARY WITHIN EACH CLASSIFICATION).

(B) THIRD (OR) FOURTH (SECOND TO LAST) DIGIT INDICATES THE WELDING POSITION THAT CAN BE USED.

(C) LAST DIGIT INDICATES USABILITY OF THE ELECTRODE.

(D) IRON POWDER PERCENTAGE BASED ON WEIGHT OF THE COVERING

Stainless Steel Covered Electrodes. The three-digit number that follows the prefix E indicates the chemical composition. In addition, letters or numbers can be used to indicate composition modifications or position usability. The specification AWS A5.4 identifies and classifies covered corrosion-resisting chromium and chromium-nickel steel welding electrodes.

Nickel and Copper Alloys. The designations for nonferrous product classifications, such as nickel and nickel alloys in AWS A5.11 and copper and copper alloys in AWS A5.6, follow the prefix with a list of chemical element abbreviations that are significant in identifying product composition, such as ENiCu, ENiCrFe, ECuSi, and ECuNi.

Surfacing Welding Electrodes. The designations for these products are contained in specifications AWS A5.13 and A5.21. They are very similar to the system used to identify nonferrous electrodes.

Aluminum and Aluminum Alloys. The specification for aluminum and aluminum alloy arc welding electrodes, AWS A5.3, uses the E prefix to indicate a covered electrode, followed by a series of numbers that identify the chemical composition that is equivalent to Aluminum Association alloy designations (for example, E1100, E3003, and E4043).

Suffix symbols are used in various classifications. The AWS A5.5 specification for low-alloy filler metals uses suffixes such as -A1, -B2, -B2L, and -C1 to indicate chemical compositions. Table 3 identifies the weld deposit chemical composition associated with a number of suffixes found on low-alloy electrodes. The classifications for nonferrous products in the AWS A5.6 specification for copper alloys and in the AWS A5.11 specification for nickel alloys list a letter or number suffix that indicates position in a series of similar alloy groupings. A similar suffix pattern is also used in the AWS A5.13 and AWS A5.21 specifications for surfacing welding electrodes. Covered stainless steel electrodes employ a number, -15 or -16, as a suffix to identify usability. The -15 suffix indicates that the electrode is designed for all-position operation using DCEP electrical current. The -16 suffix indicates all-position operation with either ac or DCEP.

TABLE 3 SUFFIX SYMBOLS AND CORRESPONDING COMPOSITIONS FOR LOW-ALLOY STEEL COVERED ELECTRODES

SUFFIX(A)	COMPOSITION, %						
	C	Mn	Si	Ni	Cr	Mo	V
A1	0.12	0.6-1.0(B)	0.40-0.80(B)	. . .	. . .	0.40-0.65	
B1	0.12	0.90	0.60-0.80(B)	. . .	0.40-0.65	0.40-0.65	. . .
B2L	0.05	0.90	0.8-1.0(B)	. . .	1.00-1.50	0.40-0.65	. . .
B2	0.12	0.90	0.60-0.80(B)	. . .	1.00-1.50	0.40-0.65	. . .
B3L	0.05	0.90	0.8-1.0(B)	. . .	2.00-2.50	0.90-1.20	. . .
B3	0.12	0.90	0.60-0.80(B)	. . .	2.00-2.50	0.90-1.20	. . .
B4L	0.05	0.90	1.00	. . .	1.75-2.25	0.40-0.65	. . .
B5	0.07-0.15	0.40-0.70	0.30-0.60	. . .	0.50-0.60	1.00-1.25	0.05
C1	0.12	1.20	0.6-0.8(B)	1.00-2.75	. . .	. . .	. . .
C2	0.12	1.20	0.6-0.8(B)	3.00-3.75	. . .	. . .	. . .
C3	0.12	0.40-1.25	0.80	0.80-1.10	0.15	0.35	0.05
D1	0.12	1.25-1.75	0.6-0.8(B)	. . .	. . .	0.25-0.45	. . .
D2	0.15	1.65-2.00	0.6-0.8(B)	. . .	. . .	0.25-0.45	. . .
G	. . .	1.0 MIN	0.80 MIN	0.50 MIN	0.30 MIN	0.20 MIN	0.10 MIN
M(C)	0.10	0.6-2.25(B)	0.6-0.8(B)	1.4-2.5(B)	0.15-1.5(B)	0.25-0.55(B)	0.05

(A) THE SUFFIX INDICATES THE CHEMICAL COMPOSITION OF THE WELD-METAL DEPOSIT.

(B) AMOUNT DEPENDS ON ELECTRODE CLASSIFICATION. SINGLE VALUES INDICATE MAXIMUM, CHECK AWS A5.5 FOR THE DIFFERENT ELECTRODE CLASSES.

(C) THERE ARE SEVERAL DIFFERENT M CLASSES; M CLASSIFICATIONS ARE INTENDED TO CONFORM TO MILITARY SPECIFICATIONS.

Deposition Rates. The melting rate of the electrode is directly related to the welding current. The current density in the electrode increases with higher current, which increases the melting rate, which, in turn, increases the deposition rate.

The electrode coating also affects deposition rate. The iron powder types are designed to have higher deposition rates and, therefore, greater productivity. Figure 4 shows the expected deposition rate versus amperage for various electrodes at a 100% duty cycle. (The actual deposition rate will be considerably less. Deposition rate is a function of the duty cycle, which is affected by the time spent changing the electrodes, cleaning slag off of the weld, etc.) Electrode size and, therefore, usable current range are determined by the base-metal thickness, welding position, welder skill level, and joint details.

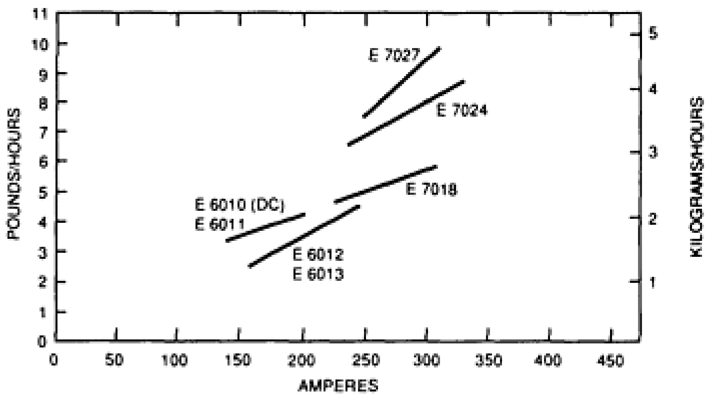

FIG. 4 DEPOSITION RATE VERSUS AMPERES FOR VARIOUS ELECTRODES

Shielded Metal Arc Welding

Raymond H. Juers, Naval Surface Warfare Center

Weld Schedules and Procedures

Welding schedules are tables of operating parameters that will provide high-quality welds under normal conditions. Strict welding schedules are not as important for the manual SMAW process as they are for semiautomatic and automatic welding for several reasons. First, in a manual welding process, the welder controls conditions by arc manipulation, which achieves better control than any of the other arc welding processes. The welder also directly controls the arc voltage and travel speed and, indirectly, the welding current.

Second, meter readings are rarely used in the SMAW process for the duplication of jobs. It is generally considered that the recommended welding current ranges given in Table 4 for the different types of electrodes are sufficient for most operations. The settings provide a good starting point when first welding on a new application, although they are not necessarily the only welding settings that can be used under every condition. For example, for high-production work, the current settings could be increased considerably over those shown. Factors such as weld appearance, welding position, and welder skill also allow variations from the settings.

TABLE 4 TYPICAL AMPERAGE RANGES FOR SELECTED SMAW ELECTRODES

ELECTRODE DIAMETER		E6010 AND E6011	E6012	E6013	E6020	E6022	E6027 AND E7027	E7014	E7015, E7016, AND E7016-1	E7018 AND E7018-1	E7024-1, E7024, AND E7028	E7048
mm	in.											
1.6	$\frac{1}{16}$	. . .	20-40	20-40	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
2.0	$\frac{5}{64}$	. . .	25-60	25-60	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
2.4(A)	$\frac{3}{32}$	40-80	35-85	45-90	. . .	. . .	. . .	80-125	65-110	70-100	100-145	. . .
3.2	$\frac{1}{8}$	75-125	80-140	80-130	100-150	110-160	125-185	110-160	100-150	115-165	140-190	80-140
4.0	$\frac{5}{32}$	110-170	110-190	105-180	130-190	140-190	160-240	150-210	140-200	150-220	180-250	150-220
4.8	$\frac{3}{16}$	140-215	140-240	150-230	175-250	170-400	210-300	200-275	180-255	200-275	230-305	210-270

5.6	$\frac{7}{32}$	170-250	200-320	210-300	225-310	370-520	250-350	260-340	240-320	260-340	275-365	. . .
6.4	$\frac{1}{4}$	210-320	250-400	250-350	275-375	. . .	300-420	330-415	300-390	315-400	335-430	. . .
8.0(A)	$\frac{5}{16}$ (A)	275-425	300-500	320-430	340-450	. . .	375-475	390-500	375-475	375-470	400-525	. . .

(A) THESE DIAMETERS ARE NOT MANUFACTURED IN THE E7028 CLASSIFICATION.

Welder Training. The SMAW process generally requires a high degree of welder skill to consistently produce quality welds. As a result, many training programs emphasize the SMAW process because of the arc manipulation skills developed by the welder. This acquired skill level makes the training on other processes much easier.

The exact content of a training program will vary, depending on the specific application of the process. The complexity of the parts to be welded and the governing codes or specifications involved also dictate the length of the training program. For example, because a pipe welder would need more skill than a tack welder, the length of his training program would be greater.

The job title of arc welder (DOT 810.384-014) describes a person whose responsibility is to weld together many components. This job includes setting up the machine and part to be welded, striking the arc and guiding it along the joint, and performing such duties as chipping, grinding, and slag removal. The welder should be able to:

- WELD IN ALL POSITIONS
- PASS EMPLOYER PERFORMANCE TESTS
- MEET CERTIFICATION STANDARDS OF GOVERNMENTAL AGENCIES OR PROFESSIONAL AND TECHNICAL ASSOCIATIONS

A tack welder (DOT 810.684-010) makes short beads at specific points to hold the parts in place for final welding. The tack welder also performs the duties of fitter helper.

A combination welder (DOT 819-384-010) welds metal parts together to either fabricate or repair assemblies. The combination welder is able to create gas welds and electric arc welds, and to perform flame-cutting operations.

The welder portion of the pipefitter description (DOT 862.381-018) must weld the pipe joint after it has been assembled and tacked in place.

Procedures. There is a definite relationship between the welding current, the size of the welding electrodes, and the welding position. These parameters must be selected to ensure that the welder has the molten weld-metal pool under complete control at all times. If the pool becomes too large, then it becomes unmanageable, perhaps allowing molten metal to run out, particularly during out-of-position welding. The welder should maintain the steady frying and crackling sound that comes with the use of correct procedures. The shape of the molten pool and the movement of the metal at the rear of the pool both serve as guides in checking weld quality. The ripples produced on the bead should be uniform and have good side tie-in with no undercut or excessive reinforcement. The factors described below are essential for maintaining high-quality welding.

Correct Electrode Type. It is important to select the proper electrode for each job. The selection should be based on the type of base metal, expected service, and mechanical properties required.

Correct Electrode Size. The choice of electrode size should depend on the type of electrode, welding position, joint preparation, base-metal thickness, and welder skill.

Correct Current. If the current is too high, then the electrode melts too fast and the molten pool becomes large, irregular, and difficult to control. Current that is too low will not provide enough heat to melt the base metal, causing the molten pool to be sluggish, with a high, irregular, ropey weld bead. It should also be noted that the electrode has inherent current limits. If the current is too high, then the core wire overheats and the coating cracks. If the current is too low, then there is insufficient heat to maintain the arc and form the protective gas shield.

Correct Arc Length. If an arc is too long, then the metal melts off the electrode in large globules that wobble from side to side, resulting in a wide and irregular weld bead with considerable spatter, and, possibly, porosity and mechanical property degradation. If the arc is too short, then it has insufficient heat to melt the base metal and electrode, which often results in the electrode sticking to the work.

Correct Travel Speed. A speed that is too fast allows the weld pool to freeze before impurities and gases can escape, and the bead will be narrow and inadequate in size. When the speed is too slow, the metal piles up and the bead is larger than required.

Correct Electrode Angle. In fillet welding and deep groove welding, the electrode angle is particularly important. When making a fillet weld, the electrode should be held so that it bisects the angle between the plates and is perpendicular to the line of the weld. When undercut occurs in the vertical member, the angle should be lowered and the arc directed toward the vertical member.

Correct Arc Manipulation. When weaving is used, the width of the weave and the pause at the ends of the weave become important. The welder must pause at each end of the weave to allow adequate fill buildup and fusion to occur. The welder should also quickly move across the center of the weld, because heating is more concentrated in the center than at the edges.

Breaking the Arc. Before an arc is broken, it is important to know whether it will be reestablished with the next electrode and the weld continued or whether it is the end of a weld pass. If welding is to continue, then the crater should remain and the arc quickly broken. If it is the end of a weld pass, then the arc should not be broken until the crater has filled.

Correct Interbead Cleaning. Proper interbead cleaning to remove slag and any spatter is essential to the production of high-quality welds. Proper cleaning prevents slag inclusions, "lack of fusion" defects, and porosity.

Shielded Metal Arc Welding

Raymond H. Juers, Naval Surface Warfare Center

Variations of the SMAW Process

Gravity Welding. Gravity feed is considered to be an automatic method of applying the SMAW process. It utilizes a relatively low-cost mechanism that includes an electrode holder attached to a bracket, which slides down an inclined bar arranged along the line of weld. Special electrodes with a heavy coating are maintained in contact with the workpiece by the weight of the electrode holder and electrode. Once the process is started, it continues automatically until the electrode has burned to a short stub, whereupon the bracket and electrode holder are automatically kicked up to break the arc. One welder can operate several gravity feeders at the same time (Fig. 5). This increases productivity, reduces welder fatigue, and requires less-skilled welders, all of which result in substantial savings in welder labor costs. Table 5 compares the deposition rate, based on pounds per hour, when using one electrode manually versus two, three, four, or five gravity feeders.

TABLE 5 COMPARISON OF DEPOSITION RATES FOR CONVENTIONAL SMAW AND MULTI-ARC GRAVITY-FED WELDING

Data are for making 7.9 mm ($\frac{5}{16}$ in.) fillet welds using E6027 electrodes.

METHOD	DEPOSITION RATE, LB/H
MANUAL, ONE ARC	9
GRAVITY, TWO ARCS	17
GRAVITY, THREE ARCS	26

GRAVITY, FOUR ARCS	34
GRAVITY, FIVE ARCS	43

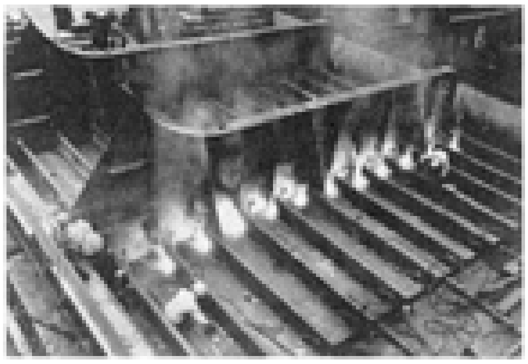

FIG. 5 GRAVITY FEEDERS BEING USED ON SHIP SUBASSEMBLIES

Although gravity-fed SMAW was investigated in the United States, England, and the Scandinavian countries in the late 1940s and early 1950s, credit must be given to the Japanese shipbuilders for perfecting and utilizing the process on a large scale in the early 1960s. The gravity welding process is being used in shipyards, railroad car shops, and barge yards throughout the world. It has reasonable acceptance in applications where large amounts of horizontal fillet welds must be made in a relatively small area.

Firecracker welding is an automatic method by which shielded metal arc welds are made using a long electrode with an electrically nonconductive coating. Human involvement is not required after the arc is initiated. A firecracker weld is generally positioned in the flat position. The welding electrode is placed in the joint and a retaining bar is placed over it. The arc is started by shorting the end of the electrode to the workpiece. The arc length is controlled by the coating thickness. As the arc travels along the stationary electrode, the electrode melts and makes a deposit on the metal immediately underneath. Once the arc is started, the process proceeds to completion automatically. Electrodes that are up to 1 m (39 in.) long and have a core diameter of 5, 6, or 8 mm (0.20, 0.24, or 0.32 in.) have been used. Both alternating and direct current have been applied, and the former may be preferred, because of arc blow problems associated with direct current.

Shielded Metal Arc Welding

Raymond H. Juers, Naval Surface Warfare Center

Special Applications of the SMAW Process

Underwater welding began during World War I when the British naval force used it to make temporary repairs of leaking rivets on ship hulls. The introduction of covered electrodes enabled succesful underwater welding and the production of welds having approximately 80% of the strength and 40% of the ductility of similar welds made in air. Because of the somewhat diminished weld properties, this SMAW application is generally restricted to salvage operations or underwater repair work.

Underwater welding can be subdivided into two major categories: welding in a wet environment and welding in a dry environment.

Wet Welding. The relatively poor quality of welds made in a wet environment is due primarily to the problem of heat transfer, welder visibility, and the presence of hydrogen in the arc atmosphere during the welding operation. When the base metal and the arc area are surrounded entirely by water, there is no temperature or heat buildup of the base metal at the weld, nor is preheating possible. This creates a weld-metal quench effect, which traps damaging amounts of hydrogen and also produces a weld solidification structure with reduced toughness and ductility. Both conditions contribute to the weld-metal cracking problems experienced when welding steels underwater. Another disadvantage is the restricted visibility, which is due to the equipment and the existing local contaminants in the water, as well as those generated by the welding arc. Under the most ideal conditions, welds produced in wet environments using covered electrodes are marginal, at best. They can be placed in service for short periods of time under reduced operating conditions, but should be replaced with quality welds as quickly as possible.

The covered electrodes used for wet welding must be waterproofed prior to underwater use. This can be done by wrapping them with waterproof tape or by dipping them in special sodium silicate mixes and allowing them to dry.

Dry Welding. The dry environment enables the production of high-quality welds that meet all code quality requirements. The SMAW process is not very popular for welding in the dry environment, because large amounts of smoke and fumes are produced. An extensive air moving, filtering, and refrigeration system must be employed when the SMAW process is used, because a dry-environment area will quickly fill with the welding fumes, making it impossible for the welder to see that weld area and to function. For this reason, the gas-tungsten arc welding and gas-metal arc welding processes have broader use in dry welding applications.

The article "Underwater Welding" in this Volume explores the topic in greater detail.

Shielded Metal Arc Welding

Raymond H. Juers, Naval Surface Warfare Center

Safety Considerations

The SMAW process, like all open arc welding processes, has a number of potential hazards. To alert welders about these safety concerns, all SMAW electrode containers carry a warning label that identifies the three most common safety hazards in these terms:

- ELECTRIC SHOCK CAN KILLDO NOT PERMIT ELECTRICALLY LIVE PARTS OR ELECTRODES TO CONTACT SKIN . . . OR YOUR CLOTHING OR GLOVES IF THEY ARE WETINSULATE YOURSELF FROM WORK AND GROUND
- FUMES AND GASES CAN BE DANGEROUS TO YOUR HEALTHKEEP FUMES AND GASES FROM YOUR BREATHING ZONE AND GENERAL AREAKEEP YOUR HEAD OUT OF FUMESUSE ENOUGH VENTILATION OR EXHAUST AT THE ARC, OR BOTH
- ARC RAYS CAN INJURE EYES AND BURN SKINWEAR CORRECT EYE, EAR, AND BODY PROTECTION

In addition to the general warnings on the container, further detailed information relating to safety is contained in ANSI/AWS Z49.1, "Safety in Welding and Cutting." The information described below is intended to expand on the general warnings, and should be particularly useful for SMAW welders.

First, the filter plates installed within welding shields must be capable of stopping the harmful levels of infrared, ultraviolet, and visible light rays originating in the arc. Filter plates are now able to absorb 99% or more of the infrared and ultraviolet rays from the arc. The shade of the filter plate suggested for use with SMAW electrodes is given in Table 6.

TABLE 6 RECOMMENDED FILTER LENS SHADES USED IN SHIELDED METAL ARC WELDING

ELECTRODE DIAMETER		LENS SHADE NO.
mm	in.	
1.6	$\frac{1}{16}$	10
2.4	$\frac{3}{32}$	
3.2	$\frac{1}{8}$	

4.0	$\frac{5}{32}$	
4.8	$\frac{3}{16}$	12
5.6	$\frac{7}{32}$	
6.4	$\frac{1}{4}$	
7.9	$\frac{5}{16}$	14
9.5	$\frac{3}{8}$	

Source: ANSI/AWS Z49.1

Second, in addition to protecting himself, the welder must also be aware of others in the area who need protection, which can usually be provided by portable screens. The failure of those working around the arc to use adequate protection can result in eye burn, which is similar to sunburn and is extremely painful for a period of up to 48 h. Usually, eye burn does not permanently injure the eyes, but it can cause intense pain. A physician should be consulted in the case of severe arc burn, regardless of whether it involves the skin or the eyes.

Third, the welding area must be adequately ventilated because of the heavy concentrations of smoke and fumes generated in the SMAW process. If welding is being performed in confined spaces with poor ventilation, such as in a tank, an external air supply in the form of a mask or special helmet may be required. In addition, a second person should be stationed at the tank manhole to provide any necessary assistance. Special ventilation is required when welding stainless steels or metals coated with copper, zinc, lead, or cadmium, because of the toxic nature of these fumes.

Fourth, cables with frayed or cracked insulation and faulty or badly worn connections can cause electrical short circuits and electrical shocks to personnel. When it is necessary to weld in a damp or wet area, the welder should wear rubber boots and stand on a dry, insulated platform.

Shielded Metal Arc Welding

Raymond H. Juers, Naval Surface Warfare Center

Selected References

- H.B. CARY, *MODERN WELDING TECHNOLOGY*, 2ND ED., PRENTICE-HALL, 1989
- "FILLER METAL COMPARISON CHARTS," AWS FMC-89, AMERICAN WELDING SOCIETY
- "SAFETY IN WELDING AND CUTTING," ANSI/AWS Z49.1, AMERICAN WELDING SOCIETY
- "TECHNICAL GUIDE FOR SHIELDED METAL ARC WELDING," HOBART BROTHERS CO.

Gas-Metal Arc Welding

D.B. Holliday, Westinghouse Electric Corporation

Introduction

GAS-METAL ARC WELDING (GMAW) is an arc welding process that joins metals together by heating them with an electric arc that is established between a consumable electrode (wire) and the workpiece. An externally supplied gas or gas mixture acts to shield the arc and molten weld pool.

Although the basic GMAW concept was introduced in the 1920s, it was not commercially available until 1948. At first, it was considered to be fundamentally a high-current-density, small-diameter, bare-metal electrode process using an inert gas for arc shielding. Its primary application was aluminum welding. As a result, it became known as metal-inert gas (MIG) welding, which is still common nomenclature.

Subsequent process developments included operation at low current densities and pulsed direct current, application to a broader range of materials, and the use of reactive gases (particularly carbon dioxide) and gas mixtures. The latter development, in which both inert and reactive gases are used, led to the formal acceptance of the term gas-metal arc welding.

The GMAW process can be operated in semi-automatic and automatic modes. All commercially important metals, such as carbon steel, high-strength low-alloy steel, stainless steel, aluminum, copper, and nickel alloys can be welded in all positions by this process if appropriate shielding gases, electrodes, and welding parameters are chosen.

Advantages. The applications of the process are dictated by its advantages, the most important of which are:

- ELECTRODE LENGTH DOES NOT FACE THE RESTRICTIONS ENCOUNTERED WITH SHIELDED-METAL ARC WELDING (SMAW).
- WELDING CAN BE ACCOMPLISHED IN ALL POSITIONS, WHEN THE PROPER PARAMETERS ARE USED, A FEATURE NOT FOUND IN SUBMERGED ARC WELDING.
- WELDING SPEEDS ARE HIGHER THAN THOSE OF THE SMAW PROCESS.
- DEPOSITION RATES ARE SIGNIFICANTLY HIGHER THAN THOSE OBTAINED BY THE SMAW PROCESS.
- CONTINUOUS WIRE FEED ENABLES LONG WELDS TO BE DEPOSITED WITHOUT STOPS AND STARTS.
- PENETRATION THAT IS DEEPER THAN THAT OF THE SMAW PROCESS IS POSSIBLE, WHICH MAY PERMIT THE USE OF SMALLER-SIZED FILLET WELDS FOR EQUIVALENT STRENGTHS.
- LESS OPERATOR SKILL IS REQUIRED THAN FOR OTHER CONVENTIONAL PROCESSES, BECAUSE THE ARC LENGTH IS MAINTAINED CONSTANT WITH REASONABLE VARIATIONS IN THE DISTANCE BETWEEN THE CONTACT TIP AND THE WORKPIECE.
- MINIMAL POSTWELD CLEANING IS REQUIRED BECAUSE OF THE ABSENCE OF A HEAVY SLAG.

These advantages make the process particularly well suited to high-production and automated welding applications. With the advent of robotics, gas-metal arc welding has become the predominant process choice.

Limitations. The GMAW process, like any welding process, has certain limitations that restrict its use:

- THE WELDING EQUIPMENT IS MORE COMPLEX, USUALLY MORE COSTLY, AND LESS PORTABLE THAN SMAW EQUIPMENT.
- THE PROCESS IS MORE DIFFICULT TO APPLY IN HARD-TO-REACH PLACES BECAUSE THE WELDING GUN IS LARGER THAN A SMAW HOLDER AND MUST BE HELD CLOSE TO THE JOINT (WITHIN 10 TO 19 MM, OR $\frac{3}{8}$ TO $\frac{3}{4}$ IN.) TO ENSURE THAT THE WELD METAL IS PROPERLY SHIELDED.
- THE WELDING ARC MUST BE PROTECTED AGAINST AIR DRAFTS THAT CAN DISPERSE THE SHIELDING GAS, WHICH LIMITS OUTDOOR APPLICATIONS UNLESS PROTECTIVE SHIELDS ARE PLACED AROUND THE WELDING AREA.
- RELATIVELY HIGH LEVELS OF RADIATED HEAT AND ARC INTENSITY CAN HINDER OPERATOR ACCEPTANCE OF THE PROCESS.

Process Fundamentals

Principles of Operation. In the GMAW process (Fig. 1), an arc is established between a continuously fed electrode of filler metal and the workpiece. After proper settings are made by the operator, the arc length is maintained at the set value, despite the reasonable changes that would be expected in the gun-to-work distance during normal operation. This automatic arc regulation is achieved in one of two ways. The most common method is to utilize a constant-speed (but adjustable) electrode feed unit with a variable-current (constant-voltage) power source. As the gun-to-work relationship changes, which instantaneously alters the arc length, the power source delivers either more current (if the arc length is decreased) or less current (if the arc length is increased). This change in current will cause a corresponding change in the electrode melt-off rate, thus maintaining the desired arc length.

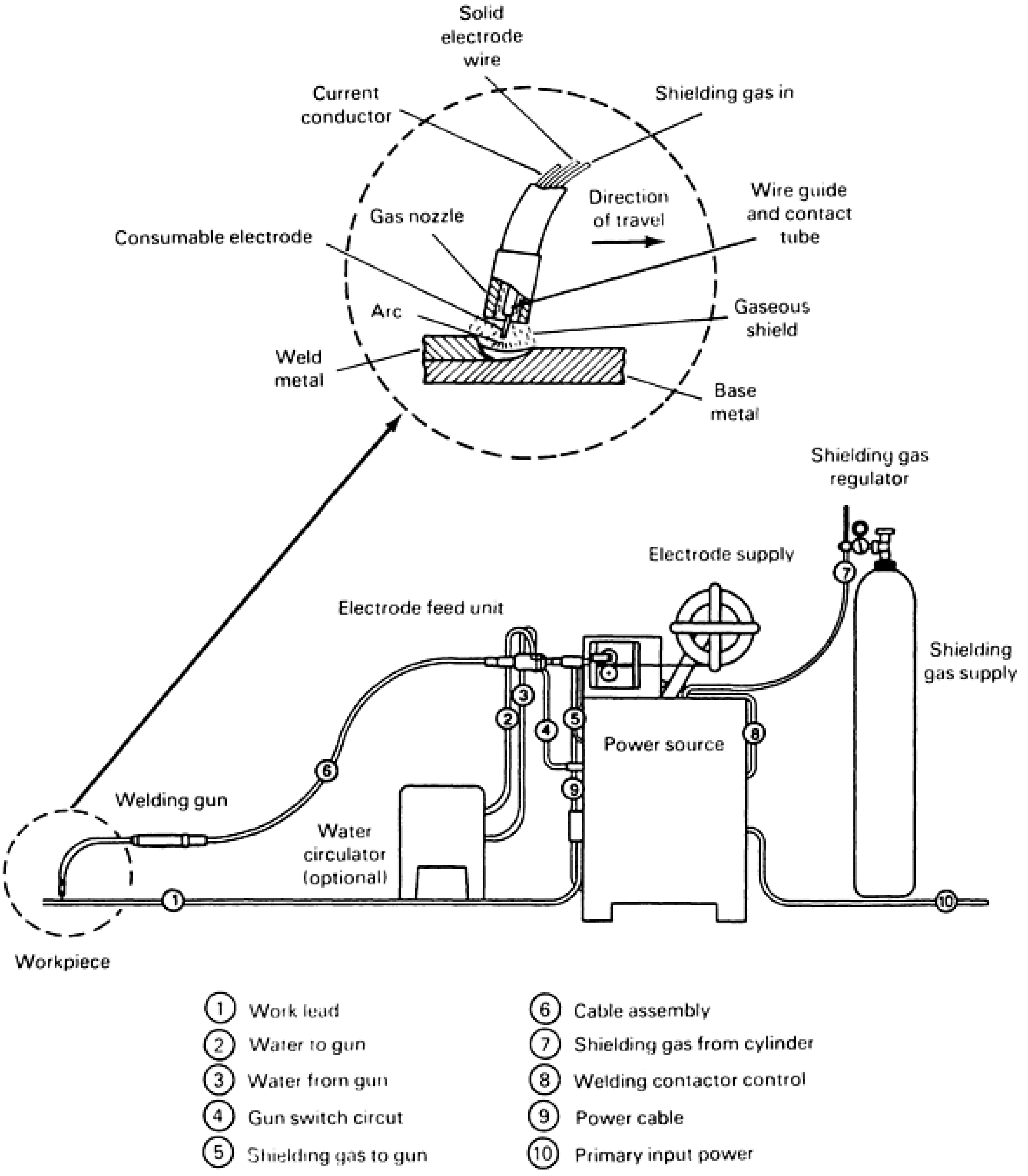
Solid electrode wire
Current conductor
Shielding gas in
Direction of travel
Wire guide and contact tube
Gas nozzle
Consumable electrode
Arc
Gaseous shield
Weld metal
Base metal
Shielding gas regulator
Electrode supply
Electrode feed unit
Shielding gas supply
Power source
Welding gun
Water circulator (optional)
Workpiece
1
2
3
4
5
6
7
8
9
10
① Work lead
② Water to gun
③ Water from gun
④ Gun switch circut
⑤ Shielding gas to gun
⑥ Cable assembly
⑦ Shielding gas from cylinder
⑧ Welding contactor control
⑨ Power cable
⑩ Primary input power

FIG. 1 SCHEMATIC OF GMAW PROCESS

The second method of arc regulation utilizes a constant-current power source and a variable-speed, voltage-sensing electrode feeder. In this case, as the arc length changes, there is a corresponding change in the voltage across the arc. As this voltage change is detected, the speed of the electrode feed unit will change to provide either more or less electrode per unit of time. This method of regulation is usually limited to larger electrodes with lower feed speeds.

Metal Transfer Mechanisms. The characteristics of the GMAW process are best described by reviewing the three basic means by which metal is transferred from the electrode to the work: short-circuiting transfer, globular transfer, or spray transfer. The type of transfer is determined by a number of factors, the most influential of which are:

- MAGNITUDE AND TYPE OF WELDING CURRENT
- ELECTRODE DIAMETER
- ELECTRODE COMPOSITION
- ELECTRODE EXTENSION BEYOND THE CONTACT TIP OR TUBE
- SHIELDING GAS
- POWER SUPPLY OUTPUT

Short-circuiting transfer encompasses the lowest range of welding currents and electrode diameters associated with the GMAW process. This type of transfer produces a small, fast-freezing weld pool that is generally suited for joining thin sections, for out-of-position welding, and for bridging of large root openings. Metal is transferred from the electrode to the workpiece only during a period when the electrode is in contact with the weld pool, and there is no metal transfer across the arc gap (Fig. 2).

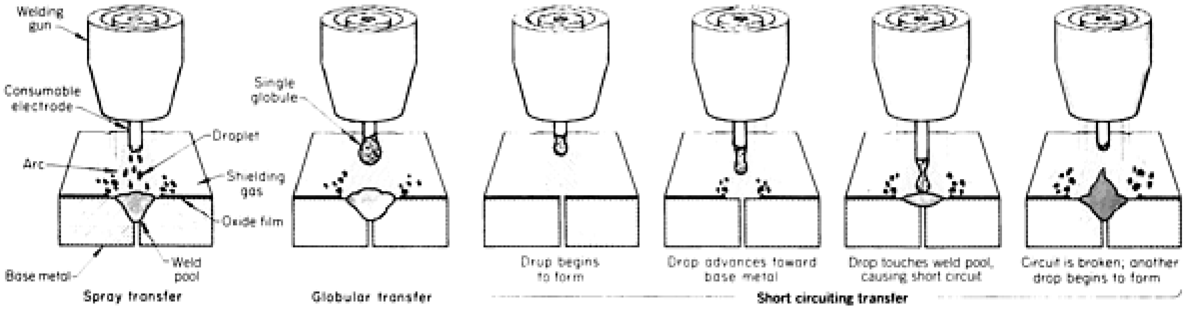
Welding gun
Consumable electrode
Single globule
Droplet
Arc
Shielding gas
Oxide film
Base metal
Weld pool
Spray transfer
Globular transfer
Drop begins to form
Drop advances toward base metal
Drop touches weld pool, causing short circuit
Circuit is broken; another drop begins to form
Short circuiting transfer

FIG. 2 TRANSFER MODES IN GMAW PROCESS

The electrode contacts the molten weld pool at a steady rate that can range from 20 to over 200 times per second. As the wire touches the weld metal, the current increases and the liquid metal at the wire tip is pinched off, initiating an arc. The rate of current increase must be high enough to heat the electrode and promote metal transfer, yet low enough to minimize spatter caused by violent separation of the molten drop. The rate of current increase is controlled by adjusting the power source inductance. The optimum setting depends on the electrical resistance of the welding circuit and the melting temperature of the electrode. When the arc is initiated, the wire melts at the tip as it is fed forward toward the next short circuit. The open-circuit voltage of the power source must be low enough so that the drop of molten metal cannot transfer until it contacts the weld metal.

Because metal transfer only occurs during short circuiting, the shielding gas has very little effect on the transfer itself. However, the gas does influence the operating characteristics of the arc and the base-metal penetration. The use of carbon dioxide generally produces high spatter levels, when compared with inert gases, but it allows deeper penetration when welding steels. To achieve a good compromise between spatter and penetration, mixtures of carbon dioxide and argon are often used. With nonferrous metals, argon-helium mixtures are used to achieve this compromise.

Globular Transfer. With a positive electrode, globular transfer takes place when the current density is relatively low, regardless of the type of shielding gas. However, the use of carbon dioxide or helium results in this type of transfer at all usable welding currents. Globular transfer is characterized by a drop size with a diameter that is greater than that of the electrode. This large drop is easily acted upon by gravity, which limits successful transfer to the flat position.

At average currents that are slightly higher than those used in short-circuiting transfer, axially directed globular transfer can be achieved in a substantially inert gas shield. However, if the arc length is too short, then the enlarging drop can short to the workpiece, become superheated, and disintegrate, producing considerable spatter. Therefore, the arc length must be long enough to ensure that the drop detaches before it contacts the weld pool. However, when higher voltage values are used, the weld is likely to be unacceptable, because of a lack of fusion, insufficient penetration, and excessive reinforcement. This limits the use of this transfer mode to very few production applications.

Carbon dioxide shielding produces a randomly directed globular transfer when the welding current and voltage values are significantly higher than the range used for short-circuiting transfer. Although severe spatter conditions result when conventional techniques are used, carbon dioxide is still the most commonly used shielding gas for welding mild steel when the quality requirements are not too rigorous. The spatter problem is controlled by "burying" the arc below the weld/base-metal surface. The resulting arc forces are adequate enough to produce a depression that contains the spatter. This technique requires relatively high currents and results in very deep penetration. Good operator setup skills are required. However, poor wetting action can result in an excessive weld reinforcement.

Spray Transfer. A very stable, spatter-free "spray" transfer mode can be produced when argon-rich shielding is used. This type of transfer requires the use of direct current with the electrode positive and a current level that is above a critical value called the "transition current." Below this current level, transfer occurs in the globular mode at the rate of a few drops per second. At values above the transition current, transfer occurs in the form of very small drops that are formed and detached at the rate of hundreds per second and are accelerated axially across the arc gap.

The transition current is proportional to the electrode diameter, and, to a lesser extent, to the electrode extension. It also has a direct relationship to the filler metal melting temperature. Transition currents for various materials and electrode diameters are shown in Table 1.

TABLE 1 GLOBULAR-TO-SPRAY TRANSITION CURRENTS FOR SELECTED ELECTRODES

WIRE ELECTRODE TYPE	SHIELDING GAS	WIRE DIAMETER		SPRAY ARC CURRENT, A
		mm	in.	
LOW-CARBON STEEL	98AR-2O$_2$	0.58	0.023	135
		0.76	0.030	150
		0.89	0.035	165
		1.14	0.045	220
		1.57	0.062	275
	95AR-5O$_2$	0.89	0.035	155
		1.14	0.045	200
		1.57	0.062	265
	92AR-8CO$_2$	0.89	0.035	175
		1.14	0.045	225
		1.57	0.062	290
	85AR-15CO$_2$	0.89	0.035	180
		1.14	0.045	240
		1.57	0.062	295
	80AR-20CO$_2$	0.89	0.035	195
		1.14	0.045	255
		1.57	0.062	345
STAINLESS STEEL	99AR-1O$_2$	0.89	0.035	150
		1.14	0.045	195
		1.57	0.062	265
	AR-HE-CO$_2$	0.89	0.035	160
		1.14	0.045	205
		1.57	0.062	280
	AR-H$_2$-CO$_2$	0.89	0.035	145

		1.14	0.045	185
		1.57	0.062	255
ALUMINUM	ARGON	0.76	0.030	95
		1.19	0.047	135
		1.57	0.062	180
DEOXIDIZED COPPER	ARGON	0.89	0.035	180
		1.14	0.045	210
		1.57	0.062	310
SILICON BRONZE	ARGON	0.89	0.035	165
		1.14	0.045	205
		1.57	0.062	270

Source: Union Carbide Industrial Gases

The spray transfer mode results in a highly directed stream of discrete drops that are accelerated by arc forces to velocities that overcome the effects of gravity. This enables the process to be used in any position, under certain conditions. Because the drops are separated, short circuits do not occur, and the spatter level is negligible, if not totally eliminated.

Another characteristic of spray transfer is the "finger" penetration pattern that it produces directly below the electrode tip. Although the penetration can be deep, it can be affected by magnetic fields that must be controlled to ensure that it is always located at the center of the weld penetration profile. Otherwise, a lack of fusion and an irregular bead surface profile can result.

The spray transfer mode can be used to weld almost any metal or alloy, because of the inert characteristics of the argon shield. Sometimes, thickness can be a factor, because of the relatively high current levels required. The resultant arc forces can cut through, rather than weld, thin sheets. In addition, high deposition rates can result in a weld pool size that cannot be supported by surface tension in the vertical and overhead positions. However, the thickness and position limitations of spray transfer have been largely overcome by specially designed power supplies. These machines produce carefully controlled current outputs that "pulse" the welding current from levels below the transition current to levels above it.

Figure 3 shows the two levels of current provided by these machines. One is a constant, low-background current that sustains the arc without providing enough energy to cause the formation of drops on the wire tip. The other is a superimposed pulsing current with an amplitude that is greater than the transition current necessary for spray transfer. During this pulse, one or more drops are formed and transferred. The frequency and amplitude of the pulses control the energy level of the arc and, therefore, the rate at which the wire melts. By reducing the arc energy and the wire melting rate, it is possible to retain many of the desirable features of spray transfer while joining sheet metals and welding thick metals in all positions.

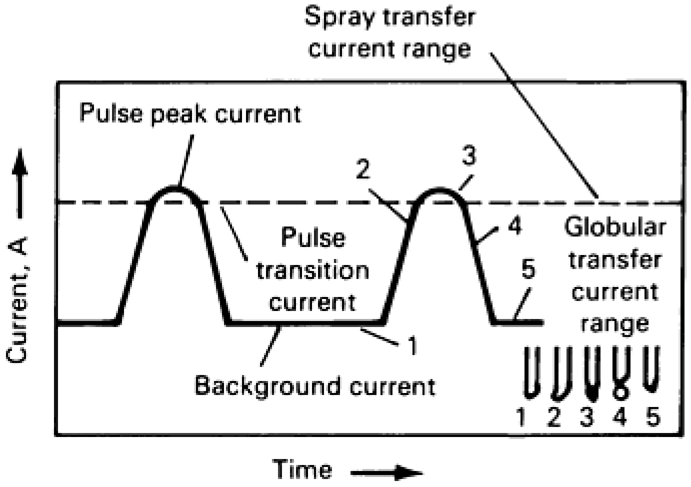

FIG. 3 CHARACTERISTIC CURRENT WAVEFORM FOR A "PULSING" POWER SUPPLY

Many variations of such machines are available. The simplest provide a single frequency of pulsing (60 or 120 pulses/s) and independent control of the background and pulsing current levels. Synergic machines, which are sophisticated, automatically provide the optimum combination of background and pulsing current levels for any given setting of wire feed speed. Normally, these settings are specific to an electrode/shielding gas combination and must be changed or reprogrammed when the combination is changed.

Process Variables. The important variables of the GMAW process that affect weld penetration, bead geometry, and overall weld quality are:

- WELDING CURRENT (ELECTRODE FEED SPEED)
- POLARITY
- ARC VOLTAGE (ARC LENGTH)
- TRAVEL SPEED
- ELECTRODE EXTENSION
- ELECTRODE ORIENTATION (GUN ANGLE)
- ELECTRODE DIAMETER

Knowledge and control of these variables are essential to consistently produce welds of satisfactory quality. Because they are not completely independent of one another, changing one variable generally requires changing one or more of the others to produce the desired results. The effects of these variables on deposit attributes are shown in Table 2.

TABLE 2 EFFECT OF CHANGES IN PROCESS VARIABLES ON WELD ATTRIBUTES

WELDING VARIABLES TO CHANGE	DESIRED CHANGES							
	PENETRATION		DEPOSITION RATE		BEAD SIZE		BEAD WIDTH	
	INCREASE	DECREASE	INCREASE	DECREASE	INCREASE	DECREASE	INCREASE	DECREASE
CURRENT AND WIRE FEED SPEED	INCREASE	DECREASE	INCREASE	DECREASE	INCREASE	DECREASE	LITTLE EFFECT	LITTLE EFFECT
VOLTAGE	NO EFFECT	NO EFFECT	LITTLE EFFECT	LITTLE EFFECT	LITTLE EFFECT	LITTLE EFFECT	INCREASE	DECREASE
TRAVEL SPEED	NO EFFECT	NO EFFECT	LITTLE EFFECT	LITTLE EFFECT	DECREASE	INCREASE	DECREASE	INCREASE
ELECTRODE EXTENSION	DECREASE	INCREASE	INCREASE(A)	DECREASE(A)	INCREASE	DECREASE	DECREASE	INCREASE
WIRE DIAMETER	DECREASE	INCREASE	DECREASE	INCREASE	LITTLE EFFECT	LITTLE EFFECT	LITTLE EFFECT	LITTLE EFFECT
SHIELD GAS %	INCREASE	DECREASE	LITTLE EFFECT	LITTLE EFFECT	LITTLE EFFECT	LITTLE EFFECT	INCREASE	DECREASE
GUN ANGLE	DRAG	PUSH	LITTLE EFFECT	LITTLE EFFECT	LITTLE EFFECT	LITTLE EFFECT	PUSH	DRAG

(A) WILL RESULT IN DESIRED CHANGE IF CURRENT LEVELS ARE MAINTAINED BY ADJUSTMENT OF WIRE FEED SPEED

Considerable skill and experience are necessary to select the optimal combination for each application. This selection is further complicated by the fact that the optimal settings for the variables are also affected by the type of base metal, the electrode composition, the welding position, quality requirements, and the number of completed weldments required. Thus, no single set of parameters provides optimal results in every case.

Welding Current. As the electrode feed speed is varied, the welding current varies in a like manner when a constant-voltage power source is used. This occurs because the current output of the power source varies dramatically with the slight changes in the arc voltage (arc length) that result when changes are made in the electrode feed speed. When all other variables are held constant, an increase in welding current results in an increase in the depth and width of penetration, deposition rate, and weld bead size.

Polarity is the term used to described the electrical connection of the welding gun in relation to the terminals of a direct-current (dc) power source. When the gun power lead is connected to the positive terminal, the polarity is designated as direct current, electrode positive (DCEP). Alternatively, a connection to the negative terminal is designated as direct current, electrode negative (DCEN). The vast majority of GMAW applications utilize DCEP, because it provides for a stable arc, low spatter, a good weld bead profile, and the greatest depth of penetration.

Arc voltage and arc length are related terms that are often used interchangeably. However, they are different. Arc voltage is an approximate means of stating the physical arc length in electrical terms. The same physical arc length, however, could yield different arc voltage readings, depending on factors such as shielding gas, current, and electrode extension. When all variables are held constant, a reliable relationship exists between the two: an increase in voltage setting will result in longer arc length. Although the arc length is the variable of interest and the one that should be controlled, arc voltage is more easily monitored. Because of this fact, and because the arc voltage is normally required to be specified in welding procedures, it is the term that is more commonly used. From any specific value of arc voltage, an increase tends to flatten the weld bead and increase the width of the fusion zone. Excessively high voltage can cause porosity, spatter, and undercut. A reduction in voltage results in a narrower weld bead with a higher crown.

Travel speed is the linear rate at which the arc is moved along the weld joint. When all other conditions are held constant, weld penetration is a maximum at an intermediate travel speed. When travel speed is decreased, the filler metal deposition per unit length increases. At very slow speeds, the welding arc impinges on the molten weld pool, rather than the base metal, thereby reducing the effective penetration.

As the travel speed is increased, the thermal energy transmitted to the base metal from the arc increases, because the arc acts more directly on the base metal. However, further increases in travel speed impart less thermal energy to the base metal. Thus, melting of the base metal first increases and then decreases with increasing travel speed. As travel speed is increased further, there is a tendency toward undercutting along the edges of the weld bead, because there is insufficient deposition of filler metal to fill the path melted by the arc.

Electrode orientation is described in two ways: by the relationship of the electrode axis with respect to the direction of travel (the travel angle) and by the angle between the electrode axis and the adjacent work surface (work angle). When the electrode points in a direction opposite to the travel direction, it results in a trail angle and is known as the backhand welding technique. When the electrode points in the direction of travel, it results in a lead angle and is called the forehand welding technique.

For all positions, a trailing travel angle that ranges from 5 to 15° (from perpendicular) provides a weld with maximum penetration and a narrow, convex surface configuration. It also provides for maximum shielding of the molten weld pool. However, the common technique utilizes a leading travel angle, which provides better visibility for the operator and a weld with a flatter surface profile. For some materials, such as aluminum, a leading angle is preferred, because it provides a "cleaning action" ahead of the molten weld metal, which promotes wetting and reduces base-material oxidation.

When producing fillet welds in the horizontal position, the work angle should be about 45° to the vertical member.

The electrode extension is the distance between the last point of electrical contact (usually the gun contact tip or tube) and the end of the electrode. An increase in the amount of this extension causes an increase in electrical resistance. This, in turn, generates additional heat in the electrode, which contributes to greater electrode melting rates. Without an increase in arc voltage, the additional metal will be deposited as a narrow, high-crowned weld bead. The optimum electrode extension generally ranges from 6.4 to 13 mm ($\frac{1}{4}$ to $\frac{1}{2}$ in.) for short-circuiting transfer and from 13 to 25 mm ($\frac{1}{2}$ to 1 in.) for spray and globular transfers.

The electrode diameter influences the weld bead configuration. A larger electrode requires a higher minimum current than a smaller electrode does to achieve the same metal transfer characteristics.

Higher currents, in turn, produce additional electrode melting and larger, more-fluid weld deposits. Higher currents also result in higher deposition rates and greater penetration, but may prevent the use of some electrodes in the vertical and overhead positions.

Gas-Metal Arc Welding

D.B. Holliday, Westinghouse Electric Corporation

Equipment

The basic equipment for a typical GMAW installation is shown in Fig. 1. The major components are discussed below.

A welding gun provides electrical current to the electrode, directs it to the workpiece, and provides a vehicle for directing shielding gas to the weld area. Different types of guns have been designed for many varied applications, ranging from heavy-duty guns for high-current, high-volume production to lightweight guns for low-current or out-of-position welding. The most commonly used guns are designed to be cooled by the surrounding air (Fig. 4). However, as amperage requirements increase, a water-cooled gun may be required. Guns are rated based on their current-carrying capacity, generally with a CO_2 shielding gas. If inert gases are used, these gun ratings must be reduced significantly. Guns can also be equipped with their own integral electrode feed units.

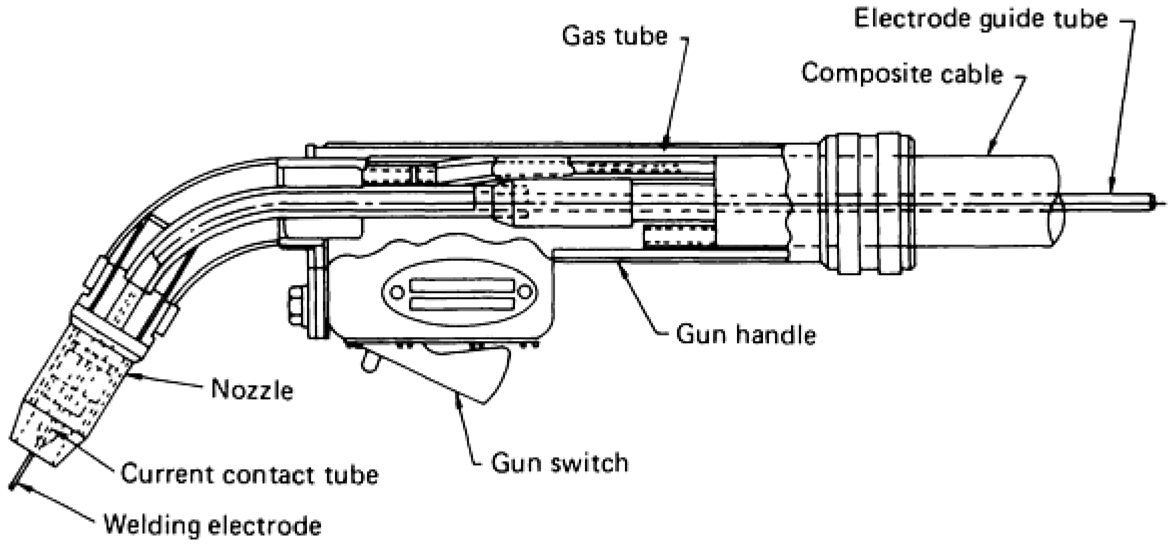

FIG. 4 CROSS SECTION OF AN AIR-COOLED GMAW GUN

The contact tube, usually made of copper or a copper alloy, is used to transmit welding current to the electrode, as well as to direct the electrode toward the work. The contact tube is connected electrically to the welding power supply by the power cable. The inner surface of the contact tube is very important, because the electrode must feed easily through this tube while making a good electrical contact. Generally, the hole in the contact tube would be from 0.13 to 0.25 mm (0.005 to 0.010 in.) larger than the wire being used, although larger sizes may be required for materials such as aluminum. The hole should be checked periodically and replaced if it has become elongated because of excessive wear. If a tip in this condition is used, it can result in poor electrical contact and erratic arc characteristics.

The nozzle directs an even-flowing column of shielding gas into the welding zone. It is extremely important to maintain an even flow in order to adequately protect the molten weld metal from atmospheric contamination. Different-sized nozzles are available and should be chosen according to the application, that is, larger nozzles for high-current work where the puddle is large and smaller nozzles for low-current and short-circuiting welding.

The electrode conduit and liner are connected to a bracket adjacent to the feed rolls on the electrode feed motor. The conduit and liner support both protect and direct the electrode from the feed rolls to the gun and contact tube. Uninterrupted electrode feeding is necessary to ensure good arc stability. A steel liner is recommended when using hard electrode materials, such as steel and copper, whereas nylon liners should be used for soft electrode materials, such as aluminum and magnesium.

The electrode feed unit, or wire feeder, consists of an electric motor, output shaft, drive rolls, and accessories for maintaining electrode alignment and pressure (Fig. 5). These units can be separate or integrated with the speed control or located remotely from it. The electrode feed motor is usually a direct-current type and provides the mechanical energy for pushing the electrode through the gun and to the work. It has a control circuit that varies the motor speed over a broad range. The feed unit can be an integral component of the gun (Fig. 6) or dual-feed units, one in the gun and one in a separate feeder, can be electrically coupled together to provide a "push-pull" system. The spool-gun and push-pull systems are often used on aluminum, where difficulty can be encountered in trying to push the wire through a conduit to the gun.

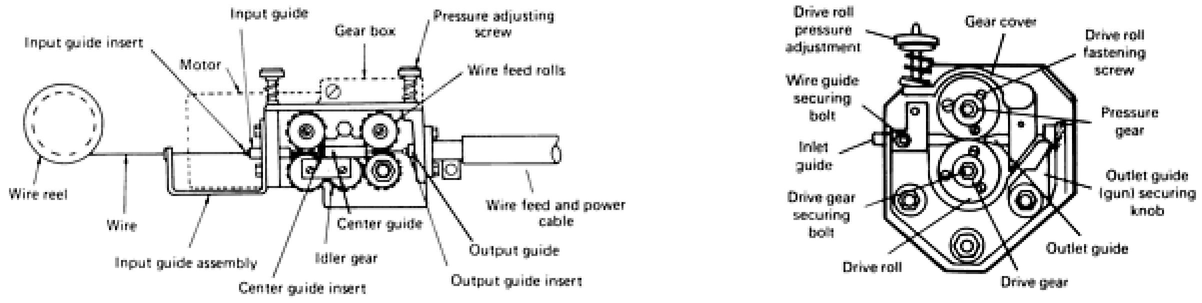
Input guide
Input guide insert
Gear box
Pressure adjusting screw
Motor
Wire feed rolls
Wire reel
Wire
Input guide assembly
Center guide insert
Idler gear
Center guide
Output guide
Output guide insert
Wire feed and power cable
Drive roll pressure adjustment
Gear cover
Drive roll fastening screw
Wire guide securing bolt
Pressure gear
Inlet guide
Outlet guide (gun) securing knob
Drive gear securing bolt
Outlet guide
Drive roll
Drive gear

FIG. 5 TYPICAL REMOTE-WIRE FEED UNITS. (A) SCHEMATIC OF A TYPICAL FOUR-ROLL SYSTEM MECHANISM. (B) CLOSE-UP VIEW OF A TYPICAL TWO-ROLL SYSTEM GEARBOX

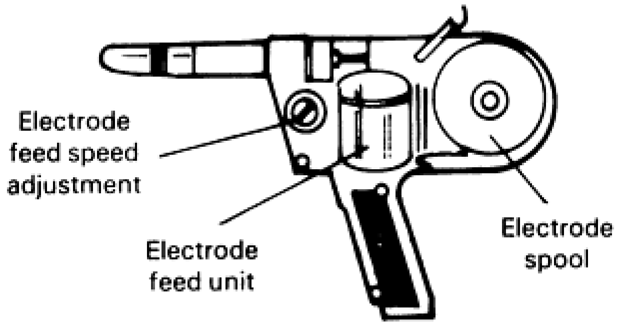
Electrode feed speed adjustment
Electrode feed unit
Electrode spool

FIG. 6 SPOOL-ON-GUN TYPE TORCH WITH COVER REMOVED TO SHOW SELF-CONTAINED PULL-TYPE FEED-UNIT SYSTEM

The feed motor is connected through a gear reducer to a set of wire feed rolls that transmit mechanical energy to the electrode, pulling it from the source and pushing it through the welding gun. Various types of rolls are available, including knurled, "U" groove, "V" groove, and flat. The knurled design is used for harder wires, such as steel, and allows maximum frictional force to be transmitted to the wire with a minimum of drive roll pressure. These types of rolls are not recommended for soft wires, such as aluminum, because they tend to cause the wire to flake, which can eventually clog the gun or liner. For these softer wires, the "U" groove or "V" groove type will allow the application of uniform pressure around the wire without deforming it. The flat rolls can be used with the smaller-diameter wires and in combination with a "U" or "V" groove.

The welding control mechanism and the electrode feed motor for semiautomatic operation are usually provided in one integrated package. The main function of the welding control mechanism is to regulate the speed of the electrode feed motor, usually through an electronic governor. The control also regulates the starting and stopping of the electrode feed through a signal received from the gun switch.

Normally, shielding gas, water (when used), and welding power are also delivered to the gun through the control mechanism, which requires direct connection to these facilities and the power supply. Gas and water flow are regulated to coincide with the weld start and stop by using solenoid valves. The control mechanism can also sequence gas flow starts and stops and energize the power source contractor. The control mechanism may allow some gas to flow before welding starts (preflow) and after welding stops (postflow) to protect the molten weld puddle. The control mechanism is usually independently powered by 115 V ac.

The welding power source provides suitable electrical power (generally 20 to 80 V) that is delivered to the electrode and workpiece to produce the arc. Because the vast majority of GMAW applications utilize DCEP, the positive lead is connected to the gun and the negative lead, to the workpiece. The power source can be the "static" type in which incoming utility power (120 to 480 V) is reduced to welding voltage by a transformer or solid-state inverter. It could also be the "rotating" type in which the welding power is provided by a rotating generator driven by a motor or internal combustion engine. The static type is normally used in shops where there is an available source of power. It has advantages over the rotating type in that it can respond more rapidly to varying arc conditions. The rotating type is generally used at field sites where external power in unavailable. Both types of power sources can be designed and built to provide either a constant current or constant potential (cp) output, the latter of which is the most common by far. On newer power sources, this cp output can be pulsed at either a constant or variable frequency. With the advent of solid-state electronic power sources, such as inverters, even further control over the pulsing variables (for example, frequency, pulse width, and so on) can be obtained.

When used in conjunction with a constant-speed wire feeder, the constant-voltage power source compensates for the variations in the contact-tip-work-distance that can occur during normal welding operations. It does this by instantaneously increasing or decreasing welding current to increase or decrease the electrode burnoff rate. The initial arc length is established by adjusting the voltage at the power source. Once this is set, no other changes are required during welding. The wire feed speed, which is also the current control, is then set by the operator and adjusted as necessary. In addition to this self-regulating feature of the CP power source, control over slope and inductance is included on those machines intended for short-circuiting transfer. Additional controls are also provided when using power sources that have pulsing capabilities.

Additional information is available in the article "Power Sources" in this Volume.

Electrode Source. The GMAW process uses a continuously fed electrode that is consumed at relatively high speeds. Therefore, the electrode source must provide a large volume of material that can readily be fed to the gun to ensure maximum process efficiency. This source is usually in the form of a spool or coil that can hold from 7 to 27 kg (15 to 60 lb) of wire that has been wound to allow free feeding without kinks or tangles. Larger spools of up to 115 kg (250 lb) are also available, and material can be provided in drums of 340 to 455 kg (750 to 1000 lb). Small spools of 0.45 to 0.9 kg (1 to 2 lb) are used for spool-on-gun equipment.

Regulated Shielding Gas Supply. A system is required to provide constant shielding gas pressure and flow rate during welding. This system consists of a regulator connected to a supply of "welding grade" shielding gas, as well as the necessary hoses or piping. The regulator is a device that reduces the source gas pressure to a constant working pressure, regardless of variations at the source. It can be a single-stage or dual-stage type and may have a built-in flowmeter. The shielding gas source can be a high-pressure cylinder, a liquid-filled cylinder, or a bulk-liquid tank. Gas mixtures are available in a single cylinder. Mixing devices can also be used to obtain the correct proportions when two or more gases or liquids are used. The type and size of the gas storage source depend on economic considerations that are based on the volume of shielding gas consumed per unit of time.

Gas-Metal Arc Welding

D.B. Holliday, Westinghouse Electric Corporation

Consumables

The two consumable, but essential, elements of the GMAW process are the electrode and the shielding gas, each of which is described below.

The chemical composition of the electrode must be selected to achieve the desired properties in the weld metal. The composition is designed with extra deoxidizers or other scavenging agents to compensate for reactions with the atmosphere and the base metal. The deoxidizers most commonly used in steel electrodes are silicon and manganese. Silicon can also be added in all transfer modes to increase weld metal fluidity or it can be added when a 300 series stainless steel electrode is used.

The physical characteristics (finish, straightness, and others) of electrodes used in the GMAW process are important to successful welding. The material specifications for these electrodes establish manufacturing requirements to ensure that users receive a uniform product that feeds smoothly through the equipment and has these characteristics, as well:

- UNIFORM WINDING ON THE SPOOL OR COIL WITH NO KINKS OR BENDS
- SMOOTH SURFACE FINISH FREE OF SLIVERS, SCRATCHES, OR SCALE
- PRESCRIBED CAST AND HELIX
- UNIFORM DIAMETER

Cast and helix refer to dimensions of a single coil of wire removed from a spool or coil and layed (that is, cast) on a flat surface. If this coil is too small in diameter (cast) or shows an excessive lift from the flat surface (helix) wire, feeding problems during welding can be anticipated.

Shielding Gas

The primary function of the shielding gas in most of the welding processes is to protect the surrounding atmosphere from contact with molten metal. In the GMAW process, this gas plays an additional role in that it has a pronounced effect on arc characteristics, mode of metal transfer, depth of fusion, weld bead profile, welding speed, and cleaning action.

Inert gases, such as argon and helium, are commonly used, as is the active gas, CO_2. It is also common to use mixtures of these gases and to employ small additions of oxygen.

Information about shielding gas compositions and about which gases to use for specific joining applications is provided in the article "Shielding Gases" in this Volume.

Gas-Metal Arc Welding

D.B. Holliday, Westinghouse Electric Corporation

Safety

The major hazards of concern during GMAW are: the fumes and gases, which can harm health; the high-voltage electricity, which can injure and kill; the arc rays, which can injure eyes and burn skin; and the noise which may be present that can damage hearing.

The type and amount of fumes and gas present during welding depend on the electrode being used, the alloy being welded, and the presence of any coatings on the base metal. To guard against potential hazards, a welder should keep his head out of the fume plume and avoid breathing the fumes and gases caused by the arc. Ventilation is always required.

Electrode shock can result from exposure to the high open-circuit voltages associated with welding power supplies. All electrical equipment and the workpiece must be connected to an approved electrical ground. Cables should be of sufficient size to carry the maximum current required. Insulation should be protected from cuts and abrasion, and the cable should not come into contact with oils, paints, or other fluids which may cause deterioration. Work areas, equipment, and clothing must be kept dry at all times. The welder should be well insulated, wearing dry gloves, rubber-soled shoes, and standing on a dry board or platform while welding.

Radiant energy, especially in the ultraviolet range, is intense during GMAW. To protect the eyes from injury, the proper filter shade for the welding-current level selected should be used. These greater intensities of ultraviolet radiation can cause rapid disintegration of cotton clothing. Leather, wool, and aluminum-coated cloth will better withstand exposure to arc radiation and better protect exposed skin surfaces.

When noise has been determined to be excessive in the work area, ear protection should be used. This can also be used to prevent spatter from entering the ear.

Conventional fire prevention requirements, such as removal of combustibles from the work area, should be followed. Sparks, slag, and spatter can travel long distances, so care must be taken to minimize the start of a fire at locations removed from the welding operation. For further information, see the guidelines set forth in the National Fire Protection Association Standard NFPA No. 51B, "Fire Protection in Use of Cutting and Welding Processes."

Care should be exercised in the handling, storage, and use of cylinders containing high-pressure and liquefied gases. Cylinders should be secured by chains or straps during handling or use. Approved pressure-reducing regulators should be used to provide a constant, controllable working pressure for the equipment in use. Lubricants or pipe fitting compounds should not be used for making any connections, as they can interfere with the regulating equipment, and in the case of oxygen service, they can contribute to a catastrophic fire.

For further safety information, see ANSI/AWS Z49.1, "Safety in Welding and Cutting."

Flux-Cored Arc Welding

David W. Meyer, The Esab Group, Inc.

Introduction

IN THE FLUX-CORED ARC WELDING (FCAW) process, the heat for welding is produced by an electric arc between a continuous filler metal electrode and the workpiece. A tubular, flux-cored electrode makes this welding process unique. The flux contained within the electrode can make the electrode self-shielding. Alternatively, an external shielding gas may be required.

Flux-Cored Arc Welding

David W. Meyer, The Esab Group, Inc.

Process Features

Flux-cored arc welding has two major variations. The gas-shielded FCAW process (Fig. 1) uses an externally supplied gas to assist in shielding the arc from nitrogen and oxygen in the atmosphere. Generally, the core ingredients in gas-shielded electrodes are slag formers, deoxidizers, arc stabilizers, and alloying elements.

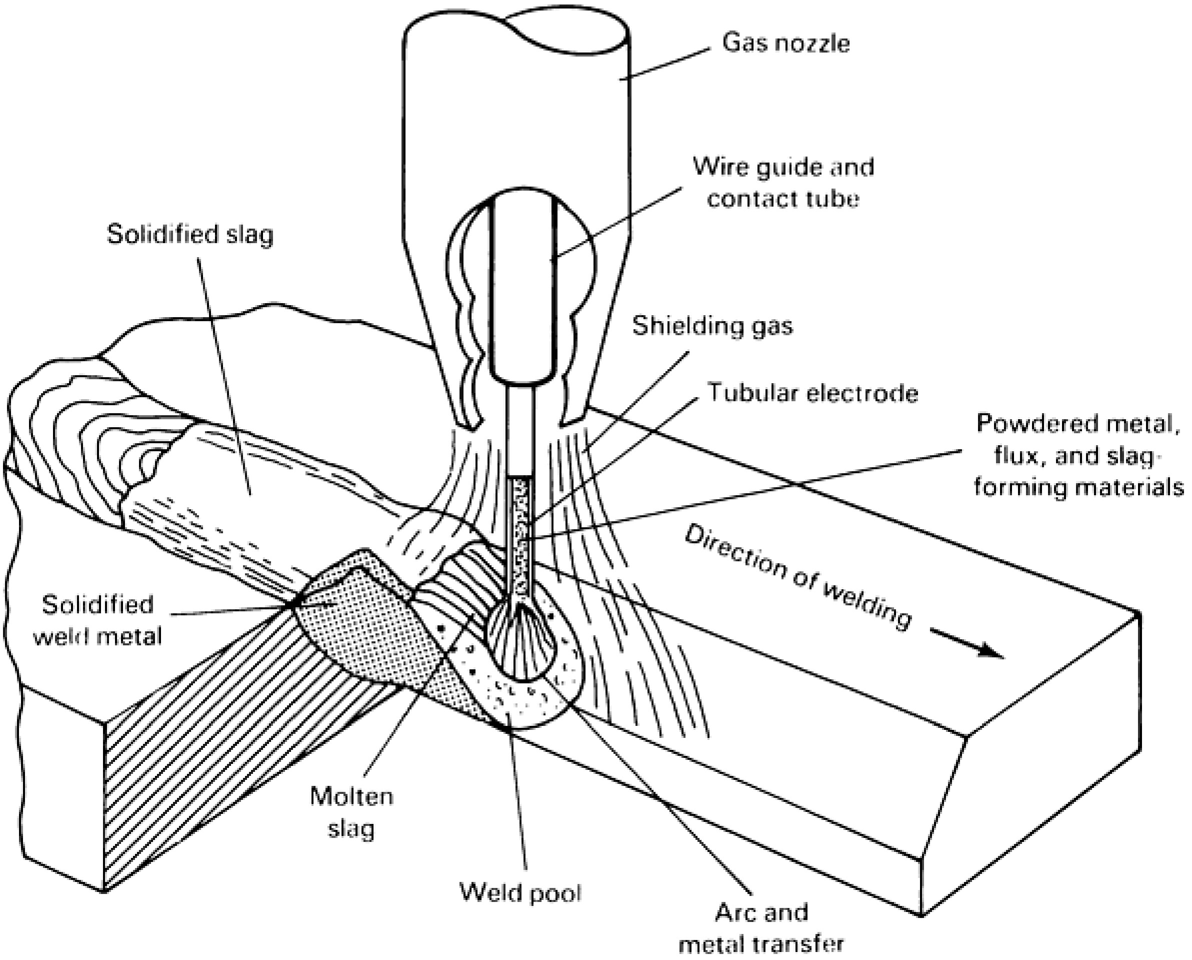
Gas nozzle
Wire guide and contact tube
Solidified slag
Shielding gas
Tubular electrode
Powdered metal, flux, and slag-forming materials
Direction of welding
Solidified weld metal
Molten slag
Weld pool
Arc and metal transfer

FIG. 1 GAS-SHIELDED FLUX-CORED ARC WELDING. SOURCE: REF 1

In the self-shielded FCAW process, the core ingredients protect the weld metal from the atmosphere without external shielding. Some self-shielded electrodes provide their own shielding gas through the decomposition of core ingredients. Others rely on slag shielding, where the metal drops being transferred across the arc and the molten weld pool are protected from the atmosphere by a slag covering. Many self-shielded electrodes also contain substantial amounts of deoxidizing and denitrifying ingredients to help achieve sound weld metal. Self-shielded electrodes can also contain arc stabilizers and alloying elements.

Advantages. Because it combines the productivity of continuous welding with the benefits of having a flux present, the FCAW process has several advantages relative to other welding processes. These advantages include:

- HIGH DEPOSITION RATES, ESPECIALLY FOR OUT-OF-POSITION WELDING
- LESS OPERATOR SKILL REQUIRED THAN FOR GAS-METAL ARC WELDING (GMAW)
- SIMPLER AND MORE ADAPTABLE THAN SUBMERGED ARC WELDING (SAW)
- DEEPER PENETRATION THAN SHIELDED METAL ARC WELDING (SMAW)
- MORE TOLERANT OF RUST AND MILL SCALE THAN GMAW

Disadvantages of the FCAW process include:

- SLAG MUST BE REMOVED FROM THE WELD AND DISPOSED OF
- MORE SMOKE AND FUME ARE PRODUCED IN FCAW THAN IN THE GMAW AND SAW PROCESSES

- FUME EXTRACTION IS GENERALLY REQUIRED
- EQUIPMENT IS MORE COMPLEX AND MUCH LESS PORTABLE THAN SMAW EQUIPMENT

Reference cited in this section

1. *WELDING HANDBOOK,* 8TH ED., VOL 2, AWS, 1991

Applications

Flux-cored arc welding enjoys widespread use in many industries. Both the gas-shielded and self-shielded FCAW processes are used to fabricate structures from carbon and low-alloy steels. Both process variants are used for shop fabrication, but the self-shielded FCAW process is preferred for field use. The acceptability of the FCAW process for structural use is illustrated by the fact that prequalified joints are included in the structural welding code of the American Welding Society (AWS).

Gas-shielded flux-cored electrodes are commonly used to weld carbon, low-alloy steel, and stainless steels in the construction of pressure vessels and piping for the chemical processing, petroleum refining, and power-generation industries. In addition, flux-cored electrodes are used to weld some nickel-base alloys.

Flux-cored electrodes are also used in the automotive and heavy-equipment industries in the fabrication of frame members, axle housings, wheel rims, suspension components, and other parts. Small-diameter flux-cored electrodes are used for automotive body repair.

Equipment

The FCAW process utilizes semiautomatic, mechanized, and fully automatic welding systems. The basic equipment includes a power supply, wire feed system, and welding gun. The required auxiliary equipment, such as shielding gas, depends on the process variant used and the degree of automation. Fume removal equipment must also be considered in most applications of the FCAW process.

Typical semiautomatic equipment is shown in Fig. 2. The equipment used in the gas-shielded FCAW process is typically identical to GMAW equipment.

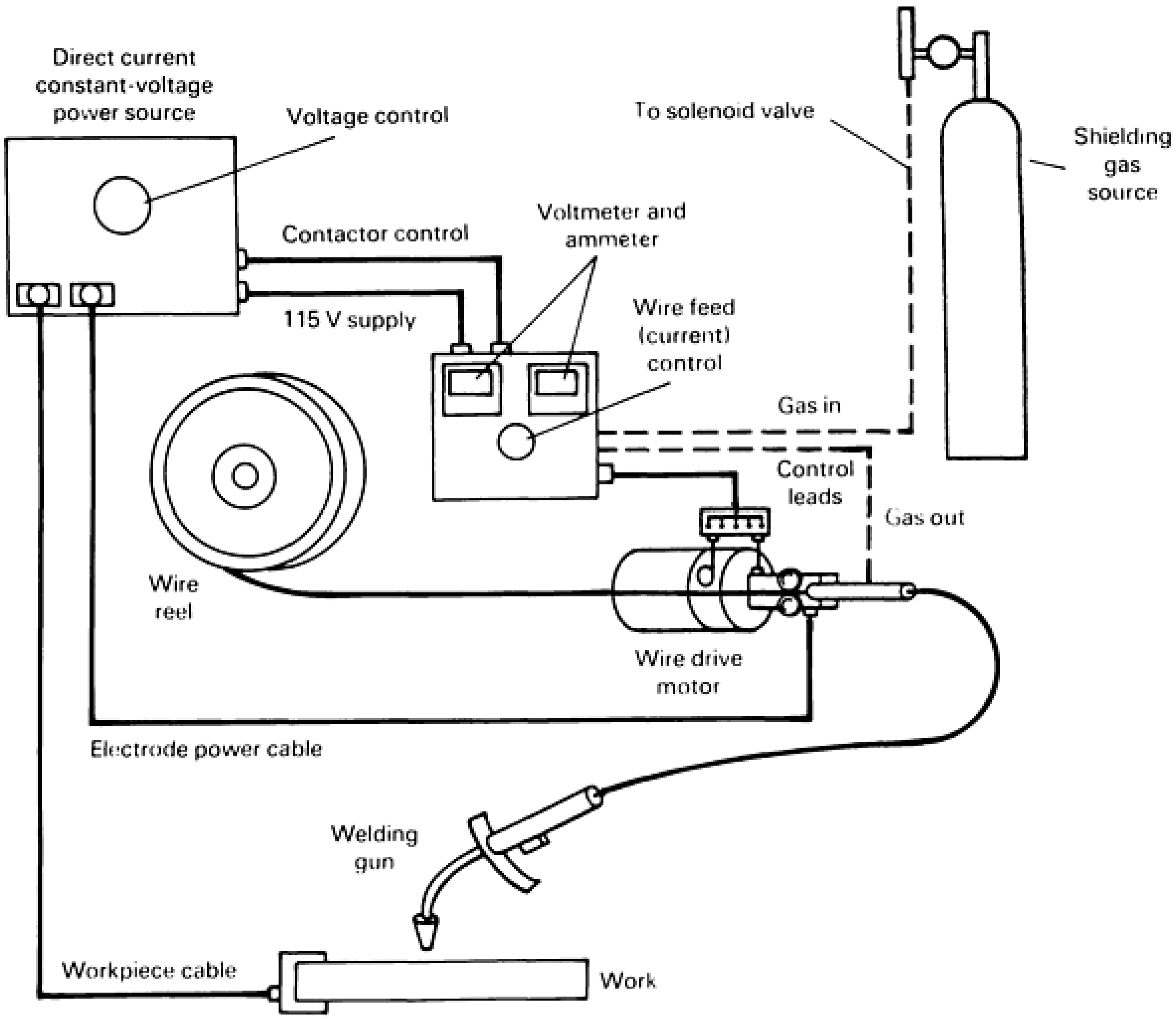

Direct current constant-voltage power source
Voltage control
To solenoid valve
Shielding gas source
Contactor control
Voltmeter and ammeter
115 V supply
Wire feed (current) control
Gas in
Control leads
Gas out
Wire reel
Wire drive motor
Electrode power cable
Welding gun
Workpiece cable
Work

FIG. 2 SEMIAUTOMATIC FCAW EQUIPMENT. SOURCE: REF 1

The recommended power supply for the semi-automatic FCAW process is a constant-voltage direct current (dc) machine. Most power supplies used for semiautomatic FCAW have output ratings of 600 A or less. A power supply rated at 60% or more duty cycle is the best choice for most industrial applications, whereas a duty-cycle rating as low as 20% may be sufficient for maintenance and repair applications.

Constant-current power supplies are used in certain situations, such as field welding applications, where portable constant-current SMAW power supplies are readily available. The addition of a contactor and a voltage-sensing wire feeder makes this an adequate welding system. However, such a system is only recommended when the use of a constant-voltage system is not feasible, because constant-current systems produce an inherently less-stable welding arc than constant-voltage systems.

Wire feeders for constant-voltage FCAW systems are generally simple devices that provide a constant wire feed speed. The power supply provides sufficient current to maintain an arc at the voltage that is preset at the power supply. A change in wire feed speed results in a change in the welding current.

In a constant-current system, the wire feeder is somewhat more complex. The welding current is preset at the power supply. The wire feeder has a voltage-sensing feedback loop that allows it to adjust the wire feed speed to maintain the desired welding voltage. The wire feeder generally contains systems to close the contactor and open the shielding gas solenoid valve (gas-shielded FCAW process only) when welding is started.

Because flux-cored wires are easily deformed by excessive feed roll pressure, knurled feed rolls are generally used in the FCAW process. Some wire feeders use a single drive roll paired with an undriven pressure roll. Others have one or two pairs of drive rolls. It is generally believed that systems having two pairs of drive rolls require the least drive roll pressure to provide dependable feeding.

Both air-cooled and water-cooled welding guns are used in the semiautomatic FCAW process. Air-cooled guns are generally preferred, because they are simpler to maintain, lighter in weight, and less bulky. Water-cooled guns may be required when welding currents over 500 A are used, especially when the shielding gas is rich in argon.

Air-cooled guns designed for gas-shielded welding should not be used for self-shielded welding, because the gun depends on the flow of shielding gas for proper cooling. Although curved-neck guns are the most common, straight guns are used to a limited extent.

A trigger switch on the welding gun is closed to initiate wire feeding, welding current flow, and shielding gas flow. The electrode is delivered from the wire feeder to the gun through a flexible conduit. Standard conduit lengths are 3, 3.7, 4.6, and 6 m (10, 12, 15, and 20 ft). Other lengths may also be available.

Mechanized and automatic FCAW equipment is not substantially different from that used in the semiautomatic FCAW process. The power supply should be rated for 100% duty cycle. Power supplies capable of outputs up to 1000 A may be required for some applications. Constant-current systems are very seldom used for mechanized and automatic welding.

The wire feed system is separated into a drive motor assembly and a welding control device, the latter of which often has a system to automatically start the travel mechanism when wire feed, current flow, and shielding gas flow are initiated. The welding control device is often equipped with a voltmeter and ammeter, as well.

The welding gun in mechanized and automatic systems is often mounted directly to the drive motor assembly, eliminating the need for a conduit. The gun is usually straight, but curved-neck guns are also used. Both water- and air-cooled guns are used, depending on the welding current level and shielding gas.

Most of the commonly available air-cooled guns can be used at levels up to 500 A with CO_2 shielding gas. When argon-rich shielding gas is used, the same gun may only be suitable for use at levels up to 300 A. Water-cooled guns are generally used at higher current levels.

Various travel mechanisms are used, depending on the applications. These mechanisms include side-beam carriages, tractor-type carriages, and robots.

Fume-Removal Equipment. In many cases, the amount of fumes generated by the FCAW process is sufficient to require fume-removal equipment. Although such equipment can be as simple as exhaust fans in the shop roof, local fume collection is often necessary. These systems can be either collection hoods (located above the welding gun) or fume-extractor guns. These guns are more efficient at collecting fumes, but are heavier and more bulky than standard welding guns. Fume collection hoods must be repositioned each time the welding location is moved in order to be effective.

Reference cited in this section

1. *WELDING HANDBOOK*, 8TH ED., VOL 2, AWS, 1991

Flux-Cored Arc Welding

David W. Meyer, The Esab Group, Inc.

Base Metals

Most weldable carbon and low-alloy steels can be welded by the FCAW process if suitable electrodes are available. Steels for which electrodes are available include:

- MILD STEELS, SUCH AS AISI 1010 TO 1030, ASTM A 36, A 285, A 515, AND A 516

- WEATHERING STRUCTURAL STEELS, SUCH AS ASTM A 588
- HIGH-STRENGTH, LOW-ALLOY STEELS, SUCH AS ASTM A 710
- HIGH-TEMPERATURE CHROMIUM-MOLYBDENUM STEELS, SUCH AS ASTM A 387 GRADES 12 (1CR-0.5MO) AND 22 (2.25CR-1MO)
- NICKEL-BASE STEELS, SUCH AS ASTM A 203
- HIGH-STRENGTH QUENCHED AND TEMPERED STEELS, SUCH AS HY-80 AND ASTM A 514 AND A 517
- MEDIUM-CARBON, HEAT-TREATABLE, LOW-ALLOY STEELS, SUCH AS AISI 4130

Several grades of stainless steel are welded with both gas-shielded and self-shielded flux-cored electrodes. Electrodes designed for welding AISI types 304, 316, 347, and others are available. Also available are electrodes suitable for joining stainless steels to carbon and low-alloy steels.

Electrodes that can be used to weld some nickel-base alloys have also been introduced. Among the nickel-base alloys that have been joined using the FCAW process are alloy 600 (UNS N06600) and alloy 625 (UNS N06625). Some cast irons are also welded using nickel-base flux-cored electrodes designed specifically for this purpose.

Flux-Cored Arc Welding

David W. Meyer, The Esab Group, Inc.

Electrode Manufacture

Flux-cored electrodes are generally manufactured using the process shown in Fig. 3. A flat sheath material is first formed into a "U" shape. The core ingredients are poured into this "U" at the desired rate. The sheath is then closed around the core materials to form a round tube. The diameter of this tube is then reduced, generally by drawing or rolling operations, to compress the core materials and bring the electrodes to a size that is usable for welding. The finished wire is than wound on spools, coils, or other packages. Other manufacturing methods, which are generally considered proprietary, are also used.

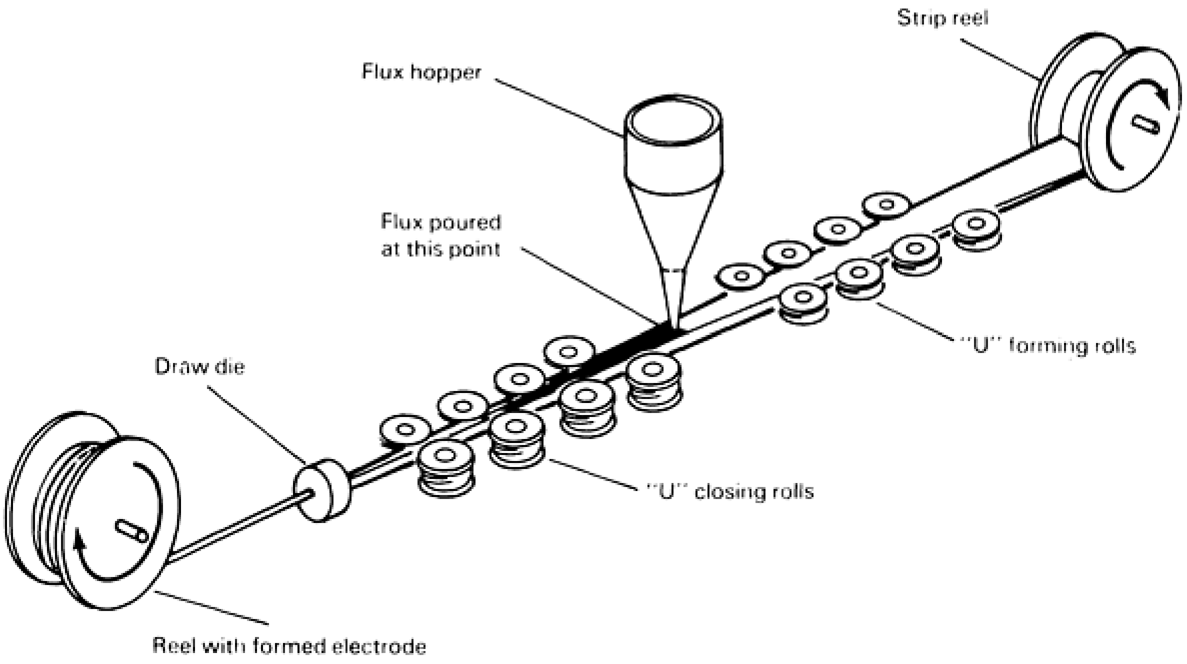

FIG. 3 FLUX-CORED ELECTRODE MANUFACTURING PROCESS. SOURCE: REF 2

The usability of a flux-cored electrode, as well as the properties and chemistry of the weld it deposits, primarily depends on the ingredients in the electrode core. Common core ingredients and their functions are listed in Table 1.

TABLE 1 FUNCTIONS OF COMMON CORE INGREDIENTS IN FCAW ELECTRODES

CORE INGREDIENT	GAS FORMER	DEOXIDIZER	DENITRIFIER	SLAG FORMER	VISCOSITY CONTROL	ARC STABILIZER	ALLOY
RUTILE (TIO_2)				X	X		
FLUORSPAR (CAF_2)				X	X		
LIME ($CACO_3$)	X			X		X	
FELDSPAR				X		X	
SYNTHETIC FRITS				X	X	X	
MANGANESE		X					X
SILICON		X					X
TITANIUM		X	X				
ALUMINUM		X	X				
CHROMIUM, NICKEL, MOLYBDENUM							X

Reference cited in this section

2. C.E. JACKSON, "FLUXES AND SLAGS IN WELDING," WRC BULLETIN 190, WELDING RESEARCH COUNCIL

Flux-Cored Arc Welding

David W. Meyer, The Esab Group, Inc.

Electrode Diameters

Flux-cored electrodes are produced in diameters ranging from 0.8 to 3.2 mm (0.030 to $\frac{1}{8}$ in.). Electrodes for all-position welding may be available in 0.8 mm (0.030 in.), 0.9 mm (0.035 in.), 1.2 mm (0.045 in.), 1.4 mm (0.052 in.), and 1.6 mm ($\frac{1}{16}$ in.) diameters. Electrodes for flat- and horizontal-position welding may be available in 1.6 mm ($\frac{1}{16}$ in.), 2.0 mm ($\frac{5}{64}$ in.), 2.4 mm ($\frac{3}{32}$ in.), 2.8 mm ($\frac{7}{64}$ in.), and 3.2 mm ($\frac{1}{8}$ in.) diameters. Other diameters may be made available by agreement between the manufacturer and the purchaser.

Electrode Classification

Carbon and Low-Alloy Steel Electrodes. Carbon steel flux-cored electrodes are classified by AWS A5.20, "Specification for Carbon Steel Electrodes for Flux Cored Arc Welding." The layout of this classification system is defined in Fig. 4; an "X" indicates a position where a designator would be. For example, the designator indicating minimum tensile strength can be a 6, to denote 62 ksi (430 MPa), or a 7, to denote 72 ksi (495 MPa). The assignment of the final designator depends on the shielding requirements and features of the electrode, as well as the recommended welding polarity. These are described in Table 2. The designations EXXT-G and EXXT-GS are designed to allow the classification of electrodes that are not covered by any of the other classifications.

TABLE 2 USABILITY TYPE DESIGNATORS FOR FLUX-CORED ELECTRODES

TYPE	SHIELDING(A)	SINGLE OR MULTIPASS(B)	TRANSFER	IMPACT TOUGHNESS REQUIREMENT	POLARITY(C)	SPECIAL CHARACTERISTICS
T-1	GAS	MULTIPASS	SPRAY-LIKE	27 J AT -18 °C (20 FT · LBF AT 0 °F)	DCEP	LOW SPATTER, FULL SLAG COVERAGE
T-2	GAS	SINGLE	SPRAY-LIKE	NO REQUIREMENT	DCEP	LOW SPATTER, FULL SLAG COVERAGE, HIGH DEOXIDIZERS
T-3	SELF	SINGLE	SPRAY-LIKE	NO REQUIREMENT	DCEP	HIGH SPEED
T-4	SELF	MULTIPASS	GLOBULAR	NO REQUIREMENT	DCEP	HIGH DEPOSITION, LOW PENETRATION, CRACK RESISTANT
T-5	GAS	MULTIPASS	GLOBULAR	27 J AT -30 °C (20 FT · LBF AT -20 °F)	DCEP(D)	IMPROVED TOUGHNESS, CRACK RESISTANT, THIN SLAG
T-6	SELF	MULTIPASS	SPRAY-LIKE	27 J AT -30 °C (20 FT · LBF AT -20 °F)	DCEP	IMPROVED TOUGHNESS, DEEP PENETRATION
T-7	SELF	MULTIPASS	GLOBULAR	NO REQUIREMENT	DCEN	CRACK RESISTANT, GOOD SLAG REMOVAL
T-8	SELF	MULTIPASS	GLOBULAR	27 J AT -30 °C (20 FT · LBF AT -20 °F)	DCEN	IMPROVED TOUGHNESS, CRACK RESISTANT
T-10	SELF	SINGLE	GLOBULAR	NO REQUIREMENT	DCEN	HIGH SPEED
T-11	SELF	MULTIPASS	SPRAY-LIKE	NO REQUIREMENT	DCEN	GENERAL PURPOSE
T-G	NOT SPECIFIED	MULTIPASS	NOT SPECIFIED	NO REQUIREMENT	NOT SPECIFIED	NOT SPECIFIED
T-GS	NOT SPECIFIED	SINGLE	NOT SPECIFIED	NO REQUIREMENT	NOT SPECIFIED	NOT SPECIFIED

Source: AWS A5.20

(A) GAS-SHIELDED ELECTRODES ARE CLASSIFIED USING CO_2 SHIELDING; HOWEVER, AR-CO_2 MIXTURES ARE COMMONLY USED IN PRACTICE.

(B) MULTIPASS ELECTRODES ARE SUITABLE FOR SINGLE-PASS OR MULTIPASS WELDING; SINGLE-PASS ELECTRODES ARE SUITABLE FOR SINGLE-PASS WELDING ONLY.

(C) DCEP, DIRECT-CURRENT ELECTRODE POSITIVE; DCEN, DIRECT-CURRENT ELECTRODE NEGATIVE.

(D) T-5 ELECTRODES ARE CLASSIFIED USING DCEP; HOWEVER, THEY ARE SOMETIMES USED WITH DCEN IN PRACTICE.

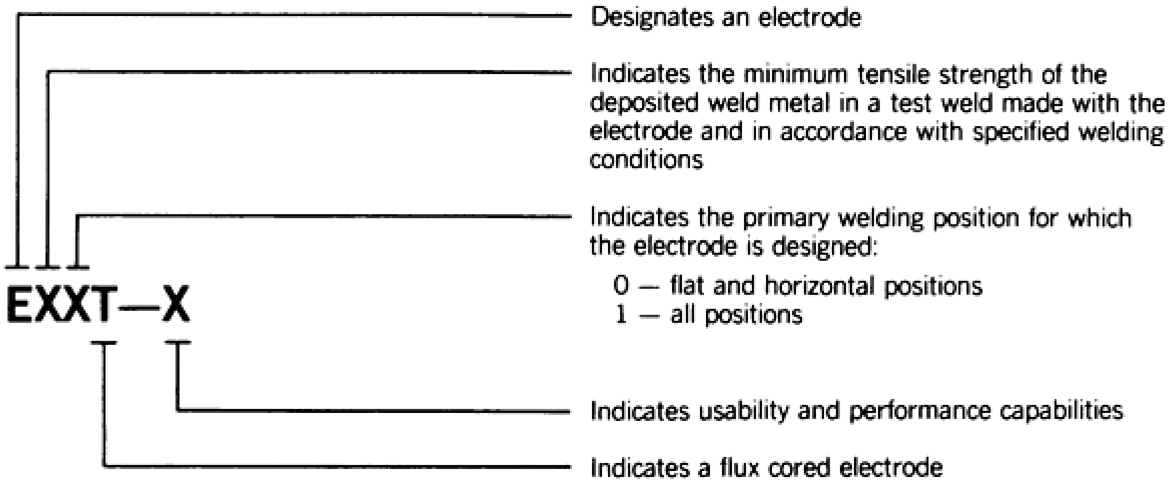
Designates an electrode
Indicates the minimum tensile strength of the deposited weld metal in a test weld made with the electrode and in accordance with specified welding conditions
Indicates the primary welding position for which the electrode is designed:
0 — flat and horizontal positions
1 — all positions
EXXT—X
Indicates usability and performance capabilities
Indicates a flux cored electrode

FIG. 4 CLASSIFICATION SYSTEM FOR CARBON STEEL FLUX-CORED ELECTRODES. THE LETTER "X" AS USED IN THIS FIGURE AND IN ELECTRODE CLASSIFICATION DESIGNATIONS IN AWS SPECIFICATION A5.20-79 SUBSTITUTES FOR SPECIFIC DESIGNATIONS INDICATED BY THIS FIGURE. SOURCE: REF 3

It should be noted that the Canadian Standards Association specification W48.5-M1990, "Carbon Steel Electrodes for Flux- and Metal-Cored Arc Welding," includes two additional electrode types that are planned for inclusion in the next revision of the AWS A5.20 specification. These classifications are EXXT-9 and EXXT-12. T-9 electrodes are generally similar to T-1 electrodes, but with improved toughness. T-12 electrodes are generally similar to T-9 electrodes, but with a maximum tensile strength also specified.

Low-alloy steel electrodes are classified under AWS A5.29, "Specification for Low Alloy Steel Electrodes for Flux Cored Arc Welding." This specification uses a classification system very similar to that used in specification AWS A5.20, except that a chemical composition designator is added to the designation. An electrode classified according to this specification will have the form EXXTX-X. All of the positions before the dash have the same meaning as those in specification AWS A5.20. The position behind the dash is the chemical composition designator, which consists of a letter and a number. The letter denotes the alloy type of the electrode as follows:

DESIGNATOR	ALLOY TYPE
A	CARBON-MOLYBDENUM STEEL
B	CHROMIUM-MOLYBDENUM STEEL
NI	NICKEL STEEL
D	MANGANESE-MOLYBDENUM STEEL
W	WEATHERING STEEL
K	OTHER LOW-ALLOY STEELS
G	NOT SPECIFIED

Specification AWS A5.29 classifies only the EXXT1-X, EXXT4-X, EXXT5-X, and EXXT8-X electrode types, and the usability designator has the name meaning as that in specification A5.20. Minimum tensile strengths of up to 830 MPa (120 ksi) are included in specification AWS A5.29. Impact toughness requirements are based on the strength, usability, and chemical composition requirements of the electrode.

Stainless steel electrodes are classified under AWS A5.22, "Specification for Flux Cored Corrosion Resisting Chromium and Chromium-Nickel Steel Electrodes." Classifications to this specification have the form EXXXT-X. The first three positions are the chemical composition designator, which corresponds to the American Iron and Steel Institute (AISI) designations (such as 308, 316, and 410) of steels having a similar composition. The final position is the shielding-type designator. T-1 types are designed for use with CO_2 or Ar-CO_2 shielding gases (classification requires the use of CO_2). T-2 types are designed for use with Ar-2O_2 shielding gas. T-3 types are self-shielded. A T-G type is included for electrodes that are not covered by the other shielding-type designators.

Nickel-Base Electrodes. A specification for nickel-base flux-cored electrodes has not yet been published. The manufacturers of these electrodes should be consulted for information on their usability, composition, and properties.

Reference cited in this section

3. "SPECIFICATION FOR CARBON STEEL ELECTRODES FOR FLUX CORED ARC WELDING," AWS A5.20-79, AWS, 1979

Flux-Cored Arc Welding

David W. Meyer, The Esab Group, Inc.

References

1. *WELDING HANDBOOK,* 8TH ED., VOL 2, AWS, 1991
2. C.E. JACKSON, "FLUXES AND SLAGS IN WELDING," WRC BULLETIN 190, WELDING RESEARCH COUNCIL
3. "SPECIFICATION FOR CARBON STEEL ELECTRODES FOR FLUX CORED ARC WELDING," AWS A5.20-79, AWS, 1979

Gas-Tungsten Arc Welding

Grant Ken-Hicken, Sandia National Laboratory

Introduction

GAS-TUNGSTEN ARC WELDING (GTAW), also known as HeliArc, tungsten inert gas (TIG), and tungsten arc welding, was developed in the late 1930s when a need to weld magnesium became apparent. Russell Meredith (Ref 1) developed a welding process using the inert gas helium and a tungsten electrode to fuse magnesium. This joining method replaced riveting as a method of building aircraft with aluminum and magnesium components. The HeliArc welding has continued to this day with many refinements and name changes, but with no change in the fundamentals demonstrated by Meredith (Ref 1).

The melting temperature necessary to weld materials in the GTAW process is obtained by maintaining an arc between a tungsten alloy electrode and the workpiece (Fig. 1). Weld pool temperatures can approach 2500 °C (4530 °F). An inert gas sustains the arc and protects the molten metal from atmospheric contamination. The inert gas is normally argon, helium, or a mixture of helium and argon.

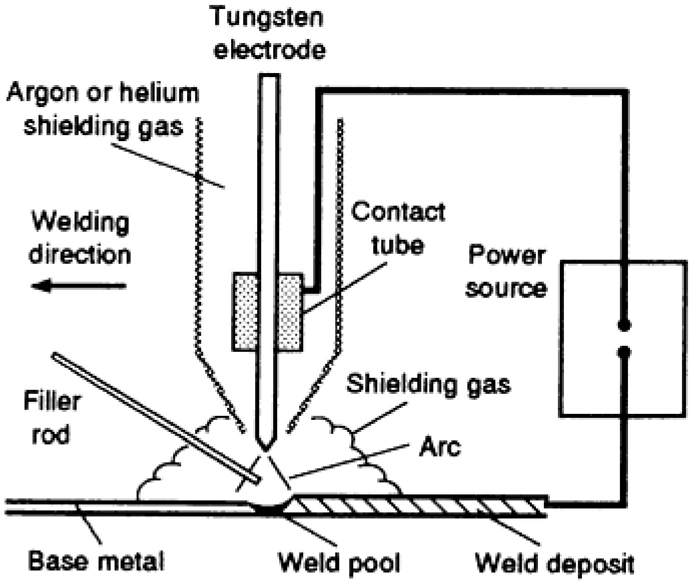
Tungsten electrode
Argon or helium shielding gas
Welding direction
Contact tube
Power source
Filler rod
Shielding gas
Arc
Base metal
Weld pool
Weld deposit

FIG. 1 SCHEMATIC SHOWING KEY COMPONENTS AND PARAMETERS OF THE GTAW PROCESS. SOURCE: REF 2

References

1. R. MEREDITH, U.S. PATENT 2,274,631
2. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987, P 10

Gas-Tungsten Arc Welding

Grant Ken-Hicken, Sandia National Laboratory

Applications

Gas-tungsten arc welding is used extensively for welding stainless steel, aluminum, magnesium, copper, and reactive materials (for example, titanium and tantalum). The process can also be used to join carbon and alloy steels. In carbon steels, it is primarily used for root-pass welding with the application of consumable inserts or open-root techniques on pipe. The materials welded range from a few thousandths of an inch to several inches in thickness.

Advantages and Limitations

Advantages of GTAW include (Ref 3):

- PRODUCES HIGH-QUALITY, LOW-DISTORTION WELDS
- FREE OF THE SPATTER ASSOCIATED WITH OTHER METHODS
- CAN BE USED WITH OR WITHOUT FILLER WIRE
- CAN BE USED WITH A RANGE OF POWER SUPPLIES
- WELDS ALMOST ALL METALS, INCLUDING DISSIMILAR ONES
- GIVES PRECISE CONTROL OF WELDING HEAT

The GTAW process is applicable when the highest weld quality is required. It can be used to weld almost all types of metals. The operator has excellent control of heat input, and vision is not limited by fumes or smoke from the process.

Limitations of GTAW include (Ref 4):

- PRODUCES LOWER DEPOSITION RATES THAN CONSUMABLE ELECTRODE ARC WELDING PROCESSES
- REQUIRES SLIGHTLY MORE DEXTERITY AND WELDER COORDINATION THAN GAS METAL ARC WELDING (GMAW) OR SHIELDED METAL ARC WELDING (SMAW) FOR MANUAL WELDING
- LESS ECONOMICAL THAN CONSUMABLE ELECTRODE ARC WELDING FOR THICK SECTIONS GREATER THAN 9.5 MM ($\frac{3}{8}$ IN.)
- PROBLEMATIC IN DRAFTY ENVIRONMENTS BECAUSE OF DIFFICULTY IN SHIELDING THE WELD ZONE PROPERLY

Additional problems with the process may include:

- TUNGSTEN INCLUSIONS IF THE ELECTRODE IS ALLOWED TO CONTACT THE WELD POOL
- CONTAMINATION OF THE WELD METAL, IF PROPER SHIELDING OF THE FILLER METAL BY THE GAS STREAM IS NOT MAINTAINED
- LOW TOLERANCE FOR CONTAMINANTS ON FILLER OR BASE METALS
- CONTAMINATION OR POROSITY, CAUSED BY COOLANT LEAKAGE FROM WATER-COOLED TORCHES
- ARC BLOW OR ARC DEFLECTION, AS WITH OTHER PROCESSES

Power supplies for GTAW are usually the constant-current type with a drooping (negative) volt-ampere (V-A) curve. Saturable reactors and thyristor-controlled units are the most common. Advances in the electronics industry have readily been accepted in the welding community, resulting in sophisticated, lightweight power supplies. Transistorized direct current (dc) power supplies are becoming common, and the newer rectifier-inverter supplies are very compact and versatile. The inverter power supply consists of three converters:

- 60 HZ PRIMARY ALTERNATING CURRENT (AC) IS RECTIFIED TO DC
- DIRECT CURRENT IS INVERTED TO HIGH-FREQUENCY AC
- ALTERNATING CURRENT IS RECTIFIED TO DC (REF 5)

The inverter supplies can be switched from constant current to constant voltage for GMAW, resulting in a very versatile piece of equipment. The inverter-controlled power supplies are more stable and have faster response times than conventional silicon-controlled rectifier (SCR) power supplies. Figure 2 compares the response of an inverter-controlled arc welding machine and a thyristor-controlled welding machine.

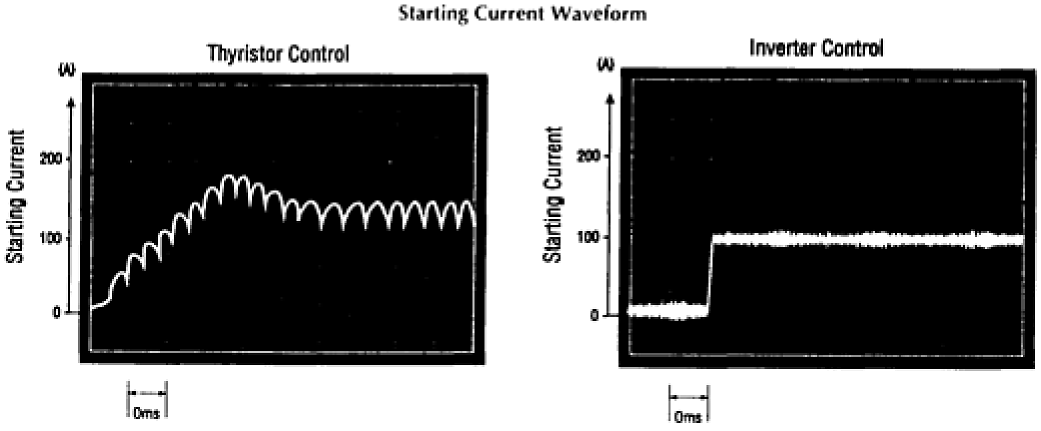

FIG. 2 STARTING CURRENT WAVEFORMS OF TWO POWER SOURCES TO SHOW RELATIVE RESPONSE TIMES OF EACH SOURCE. (A) THYRISTOR-CONTROLLED SOURCE. (B) INVERTER-CONTROLLED SOURCE. FASTER RESPONSE OF INVERTER-CONTROLLED ARC WELDING MACHINE (2 MS TO GO FROM 0 TO 100 A) INDICATES A MORE STABLE ARC.

Torch Construction. The welding torch holds the tungsten electrode that conducts the current to the arc, and it provides a means of shielding the arc and molten metal. The major components of a typical welding torch are shown in Fig. 3.

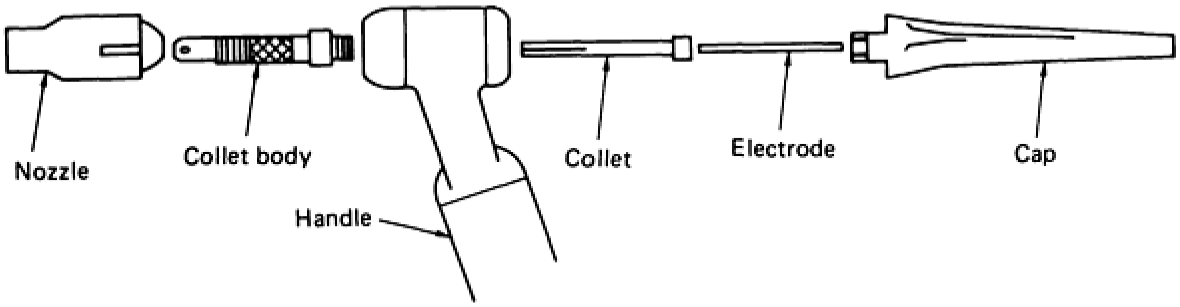

FIG. 3 SCHEMATIC SHOWING EXPLODED VIEW OF KEY COMPONENTS COMPRISING A GTAW MANUAL TORCH

Welding torches rated at less than 200 A are normally gas-cooled (that is, the shielding gas flows around the conductor cable, providing the necessary cooling). Water-cooled torches are used for continuous operation or at higher welding currents and are common for mechanized or automatic welding (see the section "GTAW Process Variations" in this article). The cooling water may be supplied to the torch from a recirculating tank that uses a radiator or chiller to cool the water.

Electrodes. The nonconsumable electrodes used in GTAW are composed of tungsten or alloys of tungsten. The most common electrode is a 2% ThO_2-W alloy (EWTh-2). This material has excellent operating characteristics and good stability. Thoria is radioactive, so care must be taken when sharpening electrodes not to inhale metal dust. The grindings are considered hazardous waste in some states, and disposal may be subject to environmental regulations. Lanthaniated (EWLa-1) and yttriated tungsten electrodes have the best starting characteristics in that an arc can be started and maintained at a lower voltage. Ceriated tungsten (EWCe-2) is only slightly better than the thoriated tungsten with respect to arc starting and melt-off rate. Any of the aforementioned electrodes produce acceptable welds. The easy starting of the lanthaniated electrode is a result of the lower work function (Ref 6) which allows it to emit electrodes readily at a lower voltage.

Pure tungsten is used primarily in ac welding and has the highest consumption rate. Alloys of zirconium are also used. Tungsten electrodes are classified on the basis of their chemical composition (Table 1). Requirements for tungsten electrodes are given in the latest edition of ANSI/AWS A5.12 ("Specification for Tungsten and Tungsten Alloy Electrodes for Arc Welding and Cutting"). The shape of the electrode tip can affect the resulting weld shape. Electrodes with included angles from 60 to 120° are stable and give good weld penetration depth-to-width ratios (Ref 7). Electrodes with smaller included angles (5 to 30°) are used for grooved weld joints to eliminate arcing to the part side walls.

TABLE 1 CLASSIFICATION OF ALLOYING ELEMENTS IN SELECTED TUNGSTEN ALLOY ELECTRODES FOR GTAW APPLICATIONS

AWS CLASSIFICATION	COLOR(A)	ALLOYING ELEMENT	ALLOYING OXIDE	ALLOYING OXIDE, WT%
EWP	GREEN	. . .	. . .	
EWCE-2	ORANGE	CERIUM	CEO_2	2
EWLA-1	BLACK	LANTHANUM	LA_2O_3	1
EWTH-1	YELLOW	THORIUM	THO_2	1
EWTH-2	RED	THORIUM	THO_2	2
EWZR-1	BROWN	ZIRCONIUM	ZRO_2	0.25
EWG	GRAY	NOT SPECIFIED(B)	. . .	. . .

Source: Ref 4

(A) COLOR MAY BE APPLIED IN THE FORM OF BANDS, DOTS, AND SO ON, AT ANY POINT ON THE SURFACE OF THE ELECTRODE.

(B) MANUFACTURER MUST IDENTIFY THE TYPE AND NOMINAL CONTENT OF THE RARE-EARTH OXIDE ADDITION.

Wire feed systems are made from a number of components and vary from simple to complex. The basic system consists of a means of gripping the wire sufficiently to pull it from the spool and push it through the guide tube to the point of welding. Electronic switches and controls are necessary for the electric drive motor. The wire will be fed into the leading edge for cold wire feeds and into the trailing edge for hot wire feeds.

Cables, hoses, and gas regulators are necessary to deliver the process consumable of electricity, water, and inert gas to the welding torch.

Arc oscillation is used in both manual and mechanized welding (see the section "GTAW Process Variations" in this article). The benefits in manual welding are basic to the control of the weld when adapting to changes in the weld joint and gap. In mechanized welding, the oscillation is typically produced by moving the entire welding torch mechanically or by moving the arc plasma with the aid of an externally applied magnetic field. Oscillation allows the welding heat to be placed at precise locations. This is advantageous when welding irregularly shaped parts. The number of welding passes and total heat input can be decreased when arc oscillation is used, because it reduces the cost as well as the weld shrinkage and upsetting.

Figure 4 shows the effect of magnetic oscillation on distortion. Some alloys need the tempering produced by multipass stringer bead welding techniques, and arc oscillation should not be used. Externally applied magnetic fields can be used to stabilize the arc, minimize arc blow, and displace the arc plasma forward of the welding torch. This results in improved weld appearances and increased welding speeds in tube mills and other high-speed applications (Ref 9).

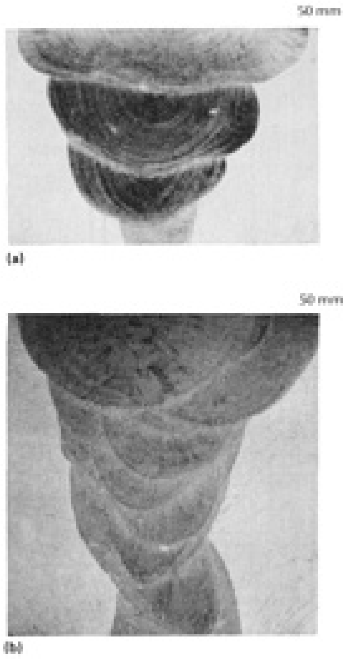

FIG. 4 EFFECT OF MAGNETIC OSCILLATION ON WELD DISTORTION IN 25 MM ($\frac{1}{2}$ IN.) THICK STAINLESS STEEL. (A) WELD PRODUCED WITH ARC OSCILLATION. (B) STRINGER BEADS PRODUCED IN WELD WITHOUT ARC OSCILLATION. NOTE MINIMAL UPSETTING WITH ARC OSCILLATION. SOURCE: REF 8

References cited in this section

3. T. MYERS, WHY THE GROWING INTEREST IN GAS TUNGSTEN ARC WELDING? *THE FABRICATOR,* VOL 22 (NO. 9), NOV 1992, P 38 -39
4. *WELDING HANDBOOK,* 8TH ED., VOL 2, AMERICAN WELDING SOCIETY, 1991, P 74-107
5. T. BYRD, INVERTER POWER SOURCES: AN EFFICIENT ALTERNATIVE, *WELD. J.,* VOL 72 (NO. 1), JAN 1993, P 37-40
6. A.A. SADEK, M. USHIO, AND F. MATSUDA, EFFECTS OF RARE EARTH METAL OXIDE ADDITIONS TO TUNGSTEN ELECTRODES, *METALL. TRANS. A,* VOL 21, 1990, P 3221-3234
7. J.R. KEY, ANODE/CATHODE GEOMETRY AND SHIELDING GAS INTERRELATIONSHIPS IN GTAW, *WELD. J.,* VOL 59 (NO. 12), DEC 1980, P 364-S TO 370-S
8. G.K. HICKEN, N.D. STUCK, AND H.W. RANDAL, APPLICATIONS OF MAGNETICALLY CONTROLLED ARCS, *WELD. J.,* VOL 55 (NO. 4), 1976, P 264-267
9. G.K. HICKEN, C.E. JACKSON, THE EFFECTS OF APPLIED OF MAGNETIC FIELDS ON WELDING ARCS, *WELD. J.,* VOL 45 (NO. 11), NOV 1966, P 515-S TO 524-S

Gas-Tungsten Arc Welding

Grant Ken-Hicken, Sandia National Laboratory

Process Parameters

Welding Current. Current is one of the most important operating conditions to control in any welding operation, because it is related to the depth of penetration, welding speed, deposition rate, and quality of the weld.

Fundamentally, there are but three choices of welding current:

- DIRECT CURRENT ELECTRODE NEGATIVE (DCEN)
- DIRECT CURRENT ELECTRODE POSITIVE (DCEP)
- ALTERNATING CURRENT

Figures 5 and 6 show the effect of dc and ac on weld shape. Table 2 gives the recommended current relative to workpiece material.

TABLE 2 SUITABILITY OF TYPES OF CURRENT FOR GTAW OF SELECTED METALS

METAL WELDED	**ALTERNATING CURRENT(A)**	**DCEN**	**DCEP**
LOW CARBON STEEL:			
0.38-0.76 MM (0.015-0.030 IN.)(A)	G(B)	E	NR
0.76-3.18 MM (0.030-0.125 IN.)	NR	E	NR
HIGH-CARBON STEEL	G(B)	E	NR
CAST IRON	G(B)	E	NR
STAINLESS STEEL	G(B)	E	NR
HEAT-RESISTANT ALLOYS	G(B)	E	NR
REFRACTORY METALS	NR	E	NR
ALUMINUM ALLOYS:			
⩽0.64 MM (0.025 IN.)	E	NR(C)	G
>0.64 MM (0.025 IN.)	E	NR(C)	NR
CASTINGS	E	NR(C)	NR
BERYLLIUM	G(B)	E	NR
COPPER AND ALLOYS:			
BRASS	G(B)	E	NR
DEOXIDIZED COPPER	NR	E	NR
SILICON BRONZE	NR	E	NR
MAGNESIUM ALLOYS:			
≤ 3.2 MM ($\frac{1}{8}$ IN.)	E	NR(C)	G
>4.8 MM ($\frac{3}{16}$ IN.)	E	NR(C)	NR
CASTINGS	E	NR(C)	NR
SILVER	G(B)	E	NR
TITANIUM ALLOYS	NR	E	NR

Note: E, excellent; G, good; NR, not recommended.

(A) STABILIZED. DO NOT USE ALTERNATING CURRENT ON TIGHTLY JIGGED ASSEMBLIES.

(B) AMPERAGE SHOULD BE ABOUT 25% HIGHER THAN WHEN DCEN IS USED.

(C) UNLESS WORK IS MECHANICALLY OR CHEMICALLY CLEANED IN THE AREAS TO BE WELDED

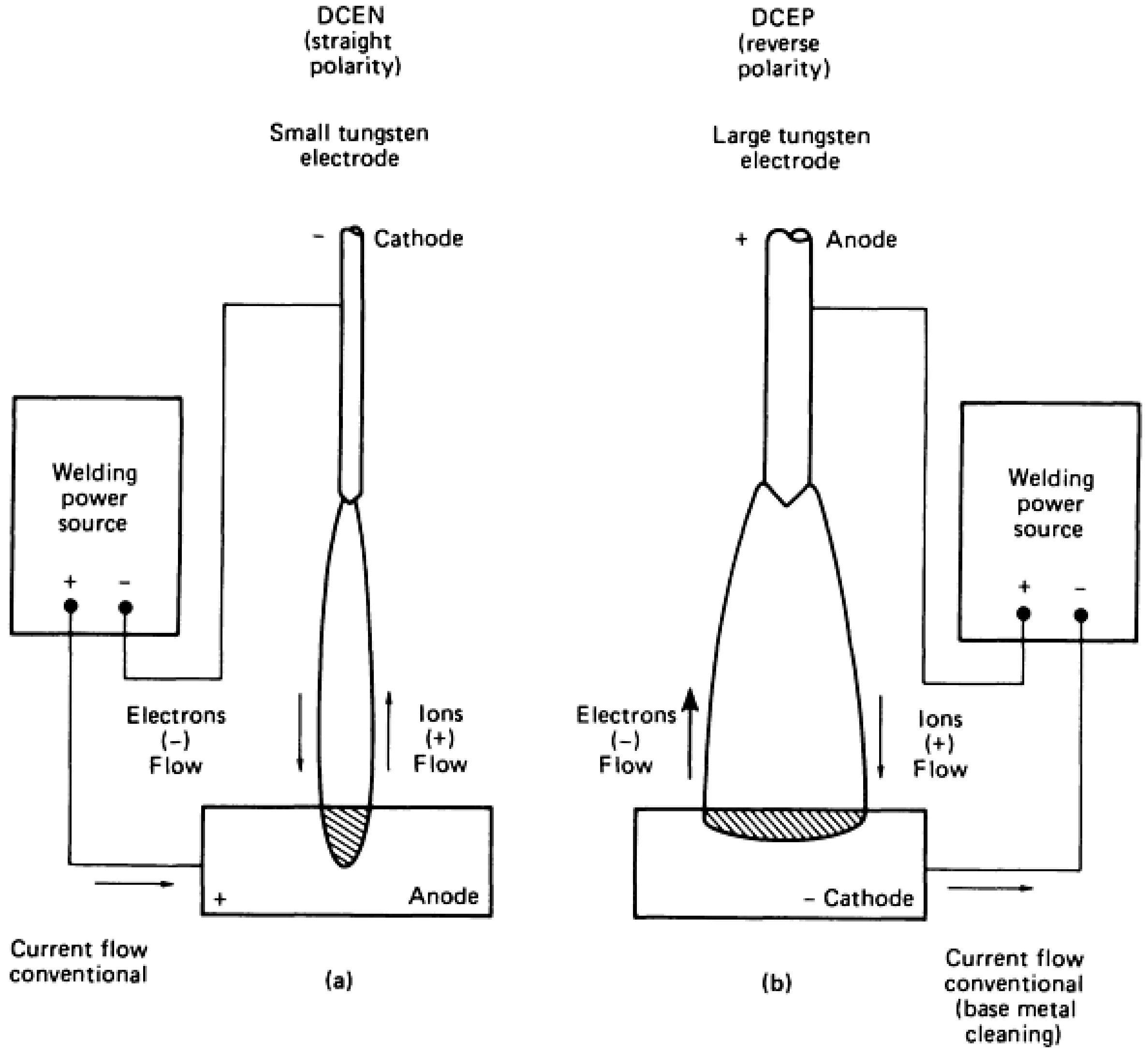
DCEN
(straight polarity)
Small tungsten electrode
– Cathode
Welding power source
+ –
Electrons (–) Flow
Ions (+) Flow
+ Anode
Current flow conventional
(a)
DCEP
(reverse polarity)
Large tungsten electrode
+ Anode
Welding power source
+ –
Electrons (–) Flow
Ions (+) Flow
– Cathode
(b)
Current flow conventional (base metal cleaning)

FIG. 5 EFFECT OF POLARITY ON GTAW WELD CONFIGURATION WHEN USING DIRECT CURRENT. (A) DCEN. DEEP PENETRATION, NARROW MELTED AREA, APPROXIMATE 30% HEAT IN ELECTRODE AND 70% HEAT IN BASE METAL. (B) DCEP. SHALLOW PENETRATION, WIDE MELTED AREA, APPROXIMATE 70% HEAT IN ELECTRODE AND 30% HEAT IN BASE METAL

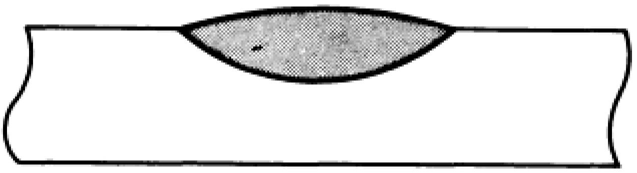

Direct current electrode positive
Shallow penetration, wide bead

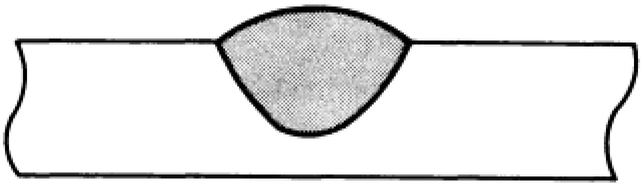

Alternating current
Medium penetration and bead width

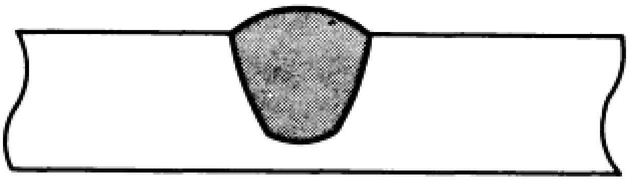

Direct current electrode negative
Deep penetration, narrow bead

FIG. 6 WELD CONFIGURATION AS A FUNCTION OF TYPE OF CURRENT (AC OR DC) USED

Alternating current is characterized as reversing the polarity of the work and electrode at 60 Hz. The rapidly changing polarity gives a cathodic cleaning action that is beneficial for oxide removal when welding aluminum and magnesium. The alternating currents result in electrode heating during the DCEP portion of each cycle. This necessitates the use of larger-diameter electrodes, normally made of pure tungsten. Variable polarity welding allows the frequency of polarity switching to be preset. This can produce the cleaning effects to ac welding and the high efficiency of dc welding. Direct current electrode negative is most often used in the GTAW process. This results in maximum application of heat to the work and maximum melting of the workpiece.

Pulsed versus Nonpulsed Current. Nonpulsed or continuous current is the standard for GTAW. However, there are several advantages to using pulsed current. Pulsing produces the maximum amount of penetration while minimizing the total heat applied to the part. Pulsing also aids in timing the motion necessary in manual welding and allows the weld pool to cool between pulses.

Microwelding refers to a class of weldments that are made at welding currents from 1 to 20 A. In most cases, the welding is used for electronic applications, bellows, wires, and other components where heat input must be precisely controlled.

Shielding Gases. The original GTAW process used helium as the shielding gas for welding magnesium and aluminum. Today, argon is the predominant shielding gas.

Argon is the least expensive of the inert gases used for shielding gas-tungsten arc welds, which is only partially responsible for its widespread use. Argon has a low ionization potential (2.52×10^{-18} J, or 15.7 eV), making it easier to form an arc plasma than with other shielding gases. Argon is approximately 1.4 times heavier than air, so it displaces air, resulting in excellent shielding of the molten weld pool.

Helium has an ionization potential of 3.92×10^{-18} J (24.5 eV), which results in more difficult arc initiation and operation at a higher arc voltage. The higher arc voltage, V, results in a higher heat input, Q, for a given arc length and current, I:

$$Q = IVT \qquad \textbf{(EQ 1)}$$

where Q is in Joules, I is in amperes; and t is in seconds. This high heat input can be very beneficial when welding copper, aluminum, and other high-conductivity materials. Helium shielding used with DCEN is very effective for welding thick aluminum.

Gas Purity. Most materials can be welded using a welding grade torch gas with a purity of 99.995% or 50 ppm impurities. However, some reactive materials (for example, titanium, molybdenum, and tantalum) require that the contaminant level be less than 50 ppm, which may require certified purity or the use of gas filters and purifiers (Ref 10).

Gas Flow Rates. Helium, because of its low density, must be used at higher flow rates than argon. Typical flow rates for argon are 7 L/min (15 ft^3/h) and 14 L/min (30 ft^3/h) for helium.

Backup Purge. Protecting the molten weld pool from the atmosphere is very important in GTAW. Atmospheric contamination can result in weld cracks, porosity, scaling, and an unacceptable granular appearance. The gas cup on the welding torch is the primary outlet of shielding gas for most GTAW applications. Back side shielding is important because the presence of oxygen can reduce weld metal penetration and result in the effects mentioned above (Ref 11). Balloon and water-soluble paper dams are sometimes used to minimize the volume to be purged. Copper backing bars and ceramics are sometimes used to hold shielding gas against the back surface of the molten weld and support the molten underbead. Reactive materials and special applications may require more elaborate shielding. This can be in the form of a simple trailing device providing the inert shielding gas or may be as elaborate as a special welding chamber equipped with gas purifiers and analyzers. Specially constructed plastic bags have been used successfully to weld large, irregularly shaped components (Fig. 7).

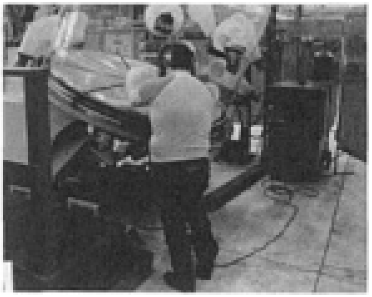

FIG. 7 PLASTIC BAG ENCLOSURE THAT SIMULATES A GLOVE BOX, USED TO PURGE IRREGULARLY SHAPED COMPONENTS IN GTAW OPERATIONS

Filler Metals. The thickness of the part to be welded will determine the need for filler metal additions. Material thinner than 3.2 mm (0.125 in.) can be successfully welded without filler metal additions. Filler metal, when needed, can be added manually in straight length or automatically from a roll or coil. The filler metal is normally added cold; hot wire can be used for automatic applications (Fig. 8). A welding insert is preplaced filler material of several possible configurations to aid in root-pass welding.

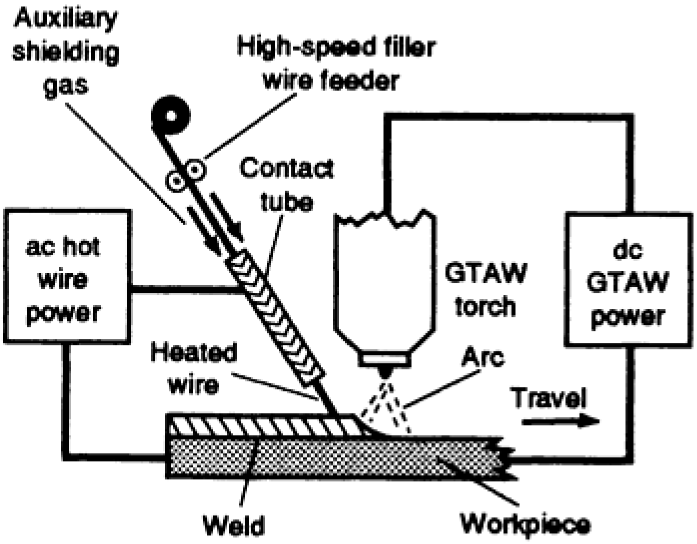

FIG. 8 SCHEMATIC SHOWING KEY COMPONENTS AND PARAMETERS OF A GTAW HOT WIRE SYSTEM. SOURCE: REF 4

Rods. Straight lengths of filler wire, typically 915 mm (36 in.) in length, are used for manual welding. Most straight lengths are round in cross section, but some aluminum fillers are somewhat rectangular.

Cold Wire. Coiled wire may be acquired in small 100 mm (4 in.) spools. Larger 305 mm (12 in.) spools or large coils can weigh over 225 kg (500 lb). The larger coils are normally used with GMAW because it requires larger quantities of filler metal. The filler wire is fed into the leading edge of the weld pool during cold-wire welding.

Hot wire GTAW utilizes a heated filler metal to increase the deposition rate of the process (Fig. 9). The wire is resistance heated to near the melting temperature and fed into the trailing edge of the weld pool. Deposition rates to 29 kg/h (65 lb/h) are achievable. The higher deposition rates obtained with hot wire make the process competitive for welds and overlays and improve productivity. Table 3 lists typical hot wire parameters.

TABLE 3 TYPICAL PARAMETERS FOR AUTOMATIC HOT WIRE GTAW

ARC CURRENT, A	ARC VOLTAGE, V	TORCH TRAVEL SPEED		WIRE FEED RATE[(A)]		DEPOSITION RATE	
		mm/min	in./min	mm/min	in./min	kg/h	lb/h
300	10-12	100-255	4-10	2790-9400	110-370	1.4-4.5	3-10
400	11-13	150-355	6-14	4700-11300	185-445	2.3-5.4	5-12
500	12-15	205-510	8-20	7490-16900	292-665	3.6-8.2	8-18

Note: Using a 4.0 to 4.8 mm ($\frac{5}{32}$ to $\frac{3}{16}$ in.) diameter 2% Th tungsten electrode with a 75He-25Ar shielding gas.

(A) WIRE DIAMETER: 1.14 MM (0.045 IN.)

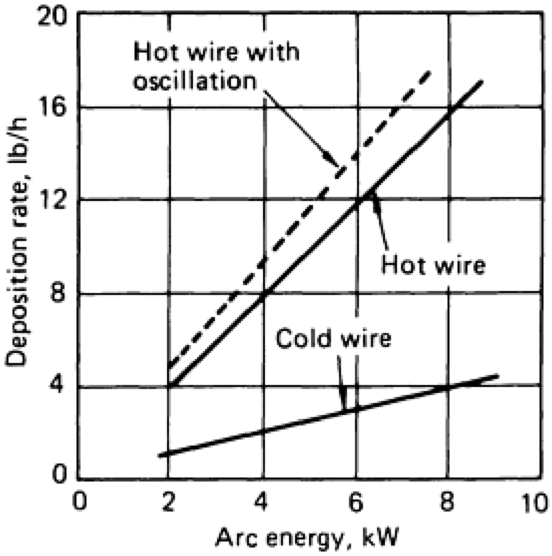

FIG. 9 DEPOSITION RATES FOR GTAW WITH COLD AND HOT FILLER WIRE ON A STEEL WORKPIECE

Welding inserts are used to produce a smooth uniform underbead. The insert is normally a separate piece of material, although integral inserts are sometimes used. Smooth underbead can be produced in a mechanized system or in manual systems when highly skilled welders are available. The insert can be obtained in several different configurations (Fig. 10). The purpose of the insert is to preplace the filler metal at the joint root. Inserts with compositions that differ from that of the base metal can be used to improve the weldability of some materials.

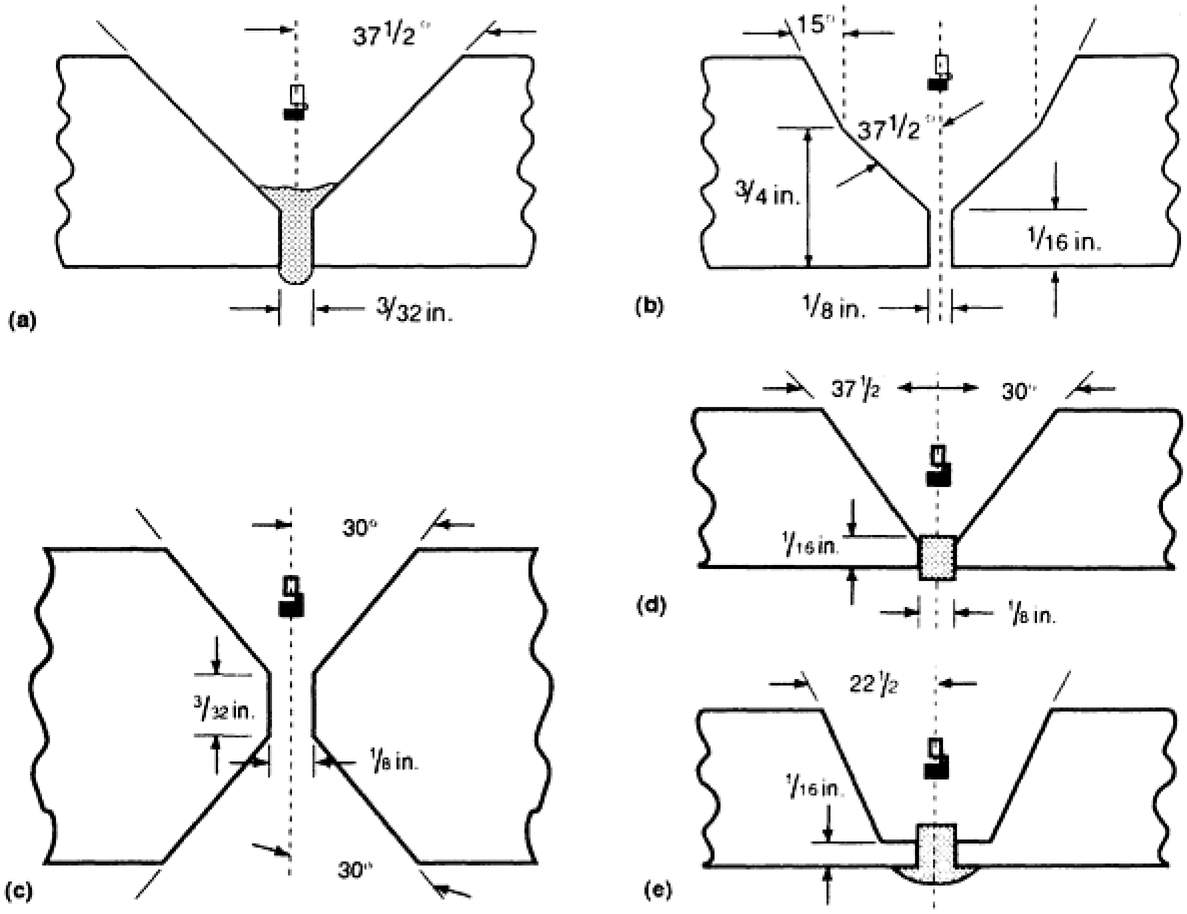
(a)
37 1/2°
3/32 in.
(b)
15°
37 1/2°
3/4 in.
1/16 in.
1/8 in.
(c)
30°
3/32 in.
1/8 in.
30°
(d)
37 1/2
30°
1/16 in.
1/8 in.
(e)
22 1/2
1/16 in.

FIG. 10 SELECTED JOINT CONFIGURATIONS AND WELDING INSERTS USED IN GTAW PROCESSES. (A) Y-TYPE INSERT, $\frac{5}{32}$ IN., FOR INDEXING WHEN ACCESSIBILITY IS LIMITED; NO LAND REQUIRED. (B) COMPOUND BEVEL WITH OPEN ROOT, $\frac{1}{8}$ IN. GAP; LAND, $\frac{1}{16}$ IN. THICK. WASHER-TYPE INSERT ($\frac{1}{8}$ × $\frac{5}{32}$ IN.) SHOULD BE USED IF INSERT REQUIRED. (C) INSIDE DIAMETER/OUTSIDE DIAMETER PREPARATION WITH OPEN ROOT, $\frac{1}{8}$ IN. GAP; LAND, $\frac{3}{32}$ IN. THICK. WASHER-TYPE INSERT ($\frac{1}{8}$ × $\frac{5}{32}$ IN.) SHOULD BE USED IF INSERT REQUIRED. (D) WASHER-TYPE INSERT ($\frac{1}{8}$ × $\frac{5}{32}$ IN.); LAND, $\frac{1}{16}$ IN. THICK. (E) J PREP MUSHROOM INSERT; LAND $\frac{1}{16}$ IN. THICK; EXTENDED LAND, $\frac{3}{32}$ IN. LONG

References cited in this section

4. *WELDING HANDBOOK,* 8TH ED., VOL 2, AMERICAN WELDING SOCIETY, 1991, P 74-107
10. K.F. KRYSIAK AND P.M. BHADHA, SHIELDING GAS PURIFICATION IMPROVES WELD QUALITY, *WELD. J.,* NOV 1990, P 47-49
11. H. GIEPL, ORBITAL TIG (GTA) WELDING, *STAINL. STEEL LEUR.,* MAY 1992, P 48-55

Gas-Tungsten Arc Welding

Grant Ken-Hicken, Sandia National Laboratory

GTAW Process Variations

Manual welding refers to the GTAW process in which the welder manipulates the welding torch by hand. If a motorized wire feeder is attached to the torch, the process is classified as semiautomatic welding.

Products generated by skilled manual welders account for a large proportion of GTAW applications. The equipment can be quite inexpensive, and properly trained welders can join a wide variety of materials. Manual welding is used extensively in stainless steel piping as well as for the root pass in carbon steel pipe welds.

Mechanized welding may require adjustment to welding parameters in response to visual observation of the weld. Machine or mechanized welding requires some specialized accessories. The basic system contains a means for holding and moving the welding torch as well as the workpiece. Because arc voltage is an essential variable in GTAW and is proportional to arc length, voltage feedback devices are often used with motorized torch holders to control the arc length.

Narrow groove welding makes use of the GTAW cold wire welding process with a narrowed weld joint. Figure 11 shows a typical weld. Narrow groove welding is limited to mechanized welding applications where precise torch location can be maintained.

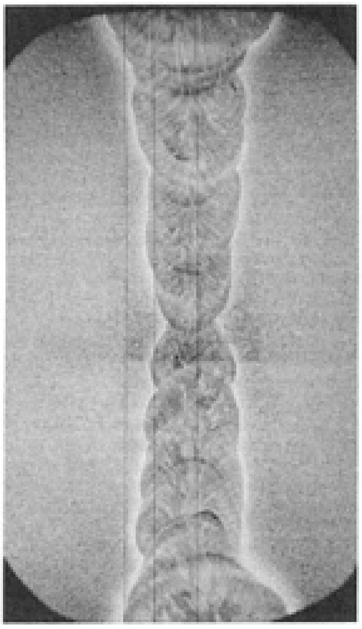

FIG. 11 TYPICAL NARROW GROOVE WELD PRODUCED IN MECHANIZED WELDING APPLICATIONS

Automatic welding does not require manual 0parameter adjustment or observation of the weld during the welding process. The most common application of automatic welding is associated with orbiting weld heads used to weld pipe and tubing. These devices attach to the workpieces and move around the circumference, fusing the metal. Most systems perform an autogenous weld; others have wire feed and oscillation capabilities. These systems are often used in conjunction with a computer to control the welding variables. Automatic controls utilizing microprocessors and computer numerical control servodrives make it possible to use one welding system to weld a variety of materials and shapes. Figure 12 shows 14 different parts welded with an automatic GTAW system.

WELDING SPECIFICATIONS FOR COMPONENTS SHOWN

Welds were made with a 250 A water-cooled GTAW torch that used 1.6 mm (0.062 in.) diameter electrode

	6061 ALUMINUM WORKPIECE$^{(A)}$			TYPE 304 STAINLESS STEEL WORKPIECE$^{(B)}$	
	COMPONENT NO. 1	COMPONENT NO. 2	COMPONENT NO. 3	COMPONENT NO. 4	COMPONENT NO. 5
WORKPIECE					
TUBE DIAMETER, MM (IN.)	22.22 (0.875)	28.58 (1.125)	26.82 (1.056)	12.7 (0.5)	19.1 (0.75)
WALL THICKNESS,	1.57 (0.062)	0.88 (0.035)	1.65 (0.065)	0.63 (0.025)	0.63 (0.025)

MM (IN.)					
POWER SUPPLY					
PEAK CURRENT, A	50-100	52-60	70-85	20-45	35-45
BACKGROUND, A	15	15	15	15	15
PULSE RATE, PPS(C)	5	6	6	8	8
WIDTH, %	65	65	65	75	75
TORCH					
WELD TRAVEL SPEED, MM/MIN	405 (16)	405 (16)	405 (16)	380 (15)	380 (15)
WELDING CYCLE TIME, S	8.0	4.5	6.5	6.0	10.0
SHIELDING GAS					
FLOW RATE, L/MIN (FT^{3}H)	8.5 (18)	8.5 (18)	8.5 (18)	7.1 (15)	7.1 (15)
FILLER METAL					
DIAMETER, MM (IN.)	0.8 (0.03)	0.8 (0.03)	0.8 (0.03)	. . .	. . .
FEED RATE, MM/MIN (IN./MIN)	3810 (150)	3560 (140)	3810 (140)	. . .	. . .
JOINT TYPE	FILLET/BUTT	BUTT	SADDLE	FILLET	FILLET/BUTT

(A) 4043 ALUMINUM FILLER METAL; AC/DC SQUARE WAVE POWER SUPPLY; 2% CERIATED ELECTRODE MATERIAL; 50% HE-50% AR SHIELDING GAS; AC PROCESS

(B) NO FILLER METAL USED; DC PRECISION POWER SUPPLY; 2% THORIATED ELECTRODE MATERIAL; ARGON SHIELDING GAS; DC PROCESS

(C) PULSES PER SECOND

FIG. 12 TYPICAL COMPONENTS PRODUCED BY AUTOMATIC GTAW PROCESS. SOURCE: REF 12

Reference cited in this section

12. K.J. PFAHL, AUTOMATIC CONTOUR WELDING OF TUBE AND PIPE, *TUBE PIPE Q.,* VOL 3 (NO. 4), 1992, P 58-62

Gas-Tungsten Arc Welding

Grant Ken-Hicken, Sandia National Laboratory

References

1. R. MEREDITH, U.S. PATENT 2,274,631
2. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987, P 10
3. T. MYERS, WHY THE GROWING INTEREST IN GAS TUNGSTEN ARC WELDING? *THE FABRICATOR,* VOL 22 (NO. 9), NOV 1992, P 38 -39
4. *WELDING HANDBOOK,* 8TH ED., VOL 2, AMERICAN WELDING SOCIETY, 1991, P 74-107
5. T. BYRD, INVERTER POWER SOURCES: AN EFFICIENT ALTERNATIVE, *WELD. J.,* VOL 72 (NO. 1), JAN 1993, P 37-40
6. A.A. SADEK, M. USHIO, AND F. MATSUDA, EFFECTS OF RARE EARTH METAL OXIDE ADDITIONS TO TUNGSTEN ELECTRODES, *METALL. TRANS. A,* VOL 21, 1990, P 3221-3234

7. J.R. KEY, ANODE/CATHODE GEOMETRY AND SHIELDING GAS INTERRELATIONSHIPS IN GTAW, *WELD. J.,* VOL 59 (NO. 12), DEC 1980, P 364-S TO 370-S
8. G.K. HICKEN, N.D. STUCK, AND H.W. RANDAL, APPLICATIONS OF MAGNETICALLY CONTROLLED ARCS, *WELD. J.,* VOL 55 (NO. 4), 1976, P 264-267
9. G.K. HICKEN, C.E. JACKSON, THE EFFECTS OF APPLIED OF MAGNETIC FIELDS ON WELDING ARCS, *WELD. J.,* VOL 45 (NO. 11), NOV 1966, P 515-S TO 524-S
10. K.F. KRYSIAK AND P.M. BHADHA, SHIELDING GAS PURIFICATION IMPROVES WELD QUALITY, *WELD. J.,* NOV 1990, P 47-49
11. H. GIEPL, ORBITAL TIG (GTA) WELDING, *STAINL. STEEL LEUR.,* MAY 1992, P 48-55
12. K.J. PFAHL, AUTOMATIC CONTOUR WELDING OF TUBE AND PIPE, *TUBE PIPE Q.,* VOL 3 (NO. 4), 1992, P 58-62

Plasma Arc Welding

Ian D. Harris, Edison Welding Institute

Introduction

PLASMA ARC WELDING (PAW) can be defined as a gas-shielded arc welding process where the coalescence of metals is achieved via the heat transferred by an arc that is created between a tungsten electrode and a workpiece. The arc is constricted by a copper alloy nozzle orifice to form a highly collimated arc column (Fig. 1). The plasma is formed through the ionization of a portion of the plasma (orifice) gas. The process can be operated with or without a filler wire addition.

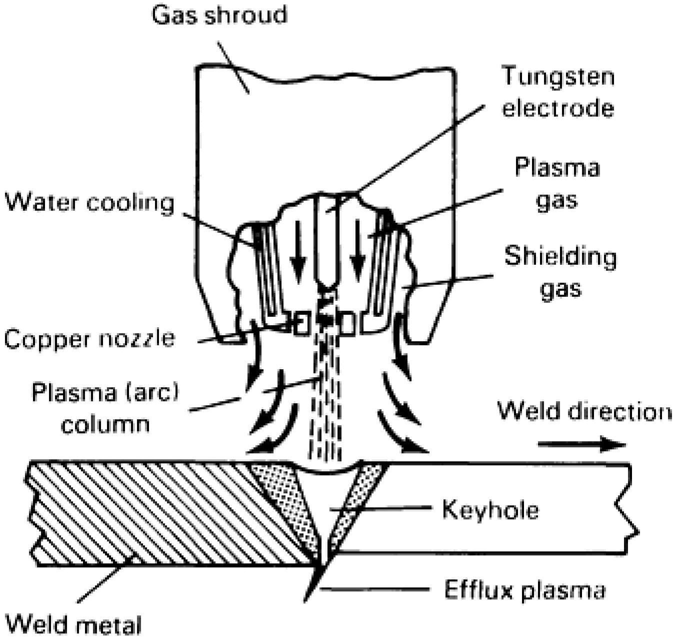

FIG. 1 PLASMA ARC WELDING PROCESS, SHOWING CONSTRICTION OF THE ARC BY A COPPER NOZZLE AND A KEYHOLE THROUGH THE PLATE

Principles of Operation. Once the equipment is set up and the welding sequence is initiated, the plasma and shielding gases are switched on. A pilot arc is then struck between a tungsten alloy electrode and the copper alloy nozzle within the torch (nontransferred arc mode), usually by applying a high-frequency open-circuit voltage. When the torch is brought in close proximity to the workpiece or when the selected welding current is initiated, the arc is transferred from the electrode to the workpiece through the orifice in the copper alloy nozzle (transferred arc mode), at which point a weld pool is formed (Fig. 1).

The PAW process can be used in two distinct operating modes, often described as the melt-in mode and the keyhole mode.

The melt-in-mode refers to a weld pool similar to that which typically forms in the gas-tungsten arc welding (GTAW) process, where a bowl-shaped portion of the workpiece material that is under the arc is melted.

In the keyhole mode, the arc fully penetrates the workpiece material, forming a nominally concentric hole, or keyhole, through the thickness. The molten weld metal flows around the arc and resolidifies behind the keyhole as the torch traverses the workpiece.

Current and Operating Modes. The PAW process uses three current modes: microplasma (melt-in mode), medium-current plasma (melt-in mode), and keyhole plasma (keyhole mode). This categorization is primarily based on the level of welding current. The microplasma mode is usually defined in the current range from 0.1 to 15 A. The medium-current plasma mode ranges from 15 to 100 A. The keyhole plasma mode is above 100 A. There is a certain degree of overlap between these current ranges. For example, keyholing can be achieved at 70 A on a 2 mm (0.08 in.) sheet. Equipment is available for welding currents up to 500 A, although a 300 A maximum is typical. Microplasma and medium-current melt-in modes are used for material up to 3 mm (0.12, or $\frac{1}{8}$, in.) thick, whereas the keyhole plasma mode is used for greater thicknesses and higher travel speeds.

In addition to operating in a continuous and steady direct current electrode negative (DCEN) mode, the PAW process can be carried out using DCEN pulsed current, as well as in the variable polarity mode, which uses both direct current electrode positive (DCEP) and electrode negative polarity switching. The pulsed current mode (both DCEN and DCEN/DCEP) is most often used when current levels (typically, above 100 A) are employed for keyhole plasma welding. Pulsing the current widens the tolerance region of acceptance welding parameters, primarily by further stabilizing the formation of the keyhole itself (Fig. 2).

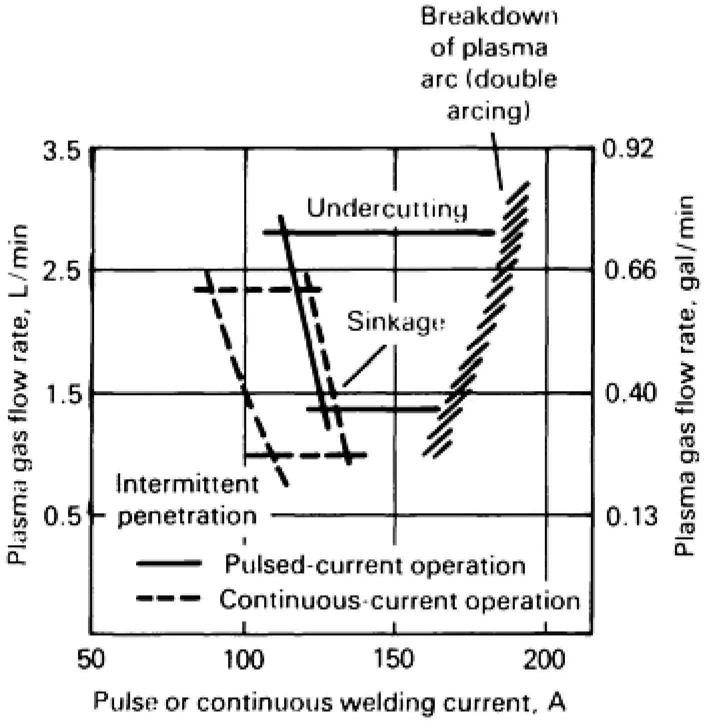

FIG. 2 TOLERANCE TO VARIATION IN WELDING CURRENT AND PLASMA GAS FLOW RATE IN PULSED AND CONTINUOUS CURRENT KEYHOLE WELDING; BOUNDARIES SHOW THE WELDING PARAMETER COMBINATIONS AT WHICH SPECIFIC DEFECTS ARE LIKELY TO OCCUR. WELDING PARAMETERS; NOZZLE BORE, 2.36 MM (0.0929 IN.); ELECTRODE DIAMETER, 4.8 MM (0.19 IN.)

The electrode positive component of the variable polarity plasma arc (VPPA) welding process promotes cathode etching of the tenacious surface oxide film when welding aluminum alloys, allowing good flow characteristics and consistent bead shape. Pulsing times are typically 20 ms for the electrode negative component and 3 ms for the electrode positive polarity. The VPPA welding process is used very effectively in specialized aerospace applications.

The PAW process is generally applied when the high penetration of the keyhole welding mode can be exploited to minimize the number of welding passes and, hence, welding time. The time saved can reduce the direct labor element of the welding operation. At the other end of the scale, the microplasma operating mode is used to weld small, thin-section components (as low as 0.025 mm, or 1 mil, thick), where the high arc constriction and low welding current can be beneficial in controlling heat input and distortion.

Advantages and Disadvantages. The advantages of the PAW process are primarily intrinsic to the keyhole mode of operation, because greater thicknesses of metal can be penetrated in a single pass, compared with other processes, such as GTAW. This greater amount of penetration allows a reduced amount of joint preparation. In some materials, for example, a square-grooved butt joint preparation can be used for thicknesses up to 12 mm (0.5 in.). The process can produce high weld integrity (similar to GTAW) while minimizing weld passes and, hence, welding times and labor costs. The columnar shape of the arc results in a greater tolerance to variations in torch standoff distance, when compared with the conical arc shape of a GTAW arc. The tungsten electrode used in the PAW process is protected from contamination by the constricting nozzle (Fig. 1). The longer arc length allows better viewing of the weld pool, which is important in manual welding.

Disadvantages include the greater capital equipment cost, when compared with its main rival, the GTAW process. Although high arc constriction achieves higher penetration, it also reduces the tolerance of the process to joint gaps and misalignment, when compared with the broader, conical arc of the GTAW process. The greater complexity of the PAW torch design and the greater number of parts requires more scheduled maintenance. The accurate set-back of the electrode tip, with respect to the nozzle orifice, is required to maintain consistent results. However, this task is facilitated by a general-purpose tool designed for nozzle removal and replacement and for electrode set-back adjustment.

Plasma Arc Welding

Ian D. Harris, Edison Welding Institute

Equipment

A basic PAW system consists of a power source, a plasma control console, a water cooler, a welding torch, and a gas supply system for the plasma and shielding gases (Fig. 3).

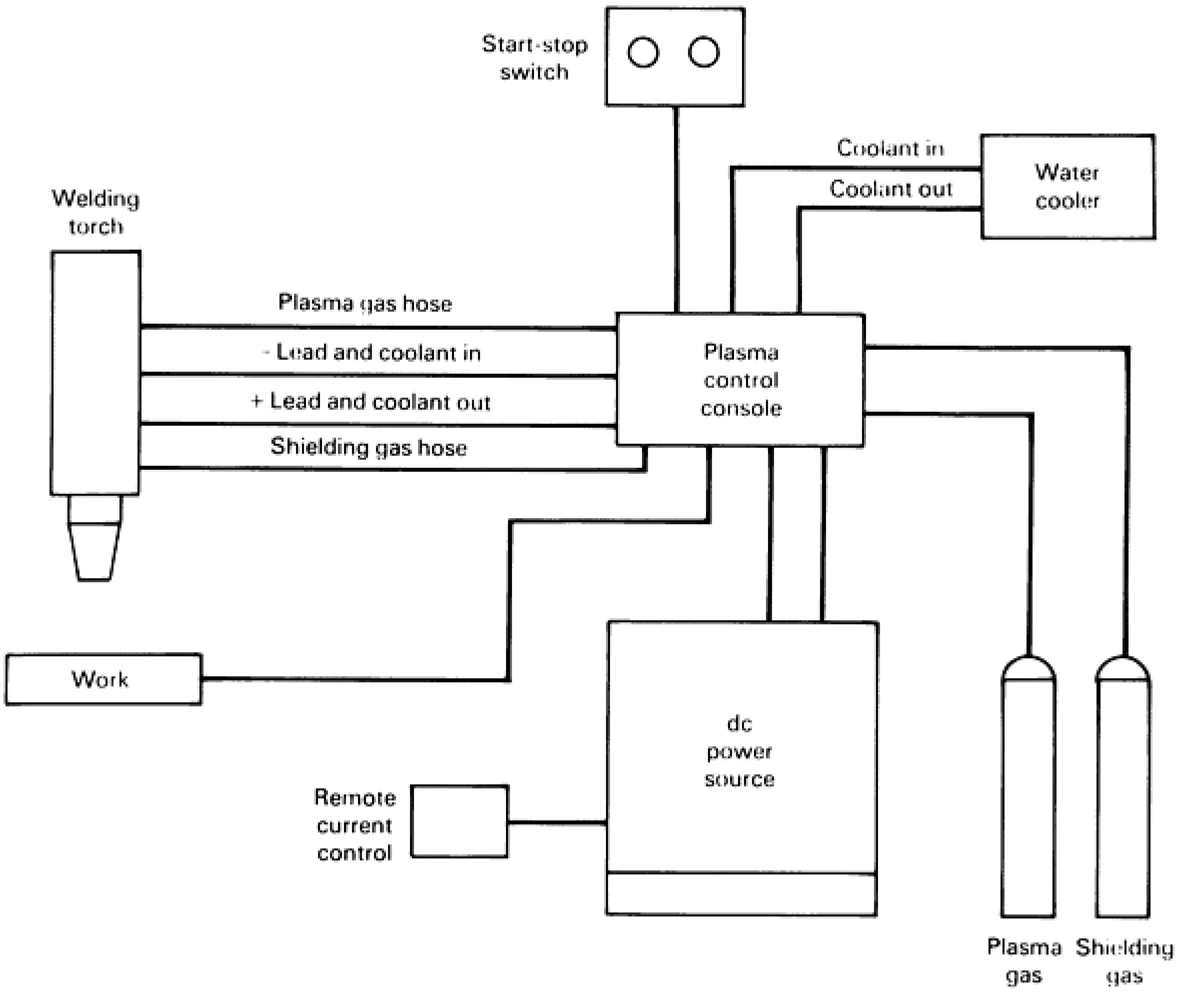
Start-stop switch
Coolant in
Coolant out
Water cooler
Welding torch
Plasma gas hose
- Lead and coolant in
+ Lead and coolant out
Shielding gas hose
Plasma control console
Work
dc power source
Remote current control
Plasma gas
Shielding gas

FIG. 3 TYPICAL EQUIPMENT FOR PLASMA ARC WELDING

The power source, which supplies the main power for the welding system, is usually supplemented with a sequence controller and control console. The sequence controller sequences the timing of gas flow, arc initiation, main welding current control, and any up-slope and downslope parameters. In its simplest form, the plasma control console controls the gas flow for plasma and shielding gases from separate flow-meters and incorporates the high-frequency pilot arc initiation circuit. The welding torch can be manual or mechanized and is water cooled to avoid torch overheating and to maximize component life.

In most PAW installations, plasma and shielding gases are supplied from separate gas cylinders, although bulk gas can readily be used. The gas supply is usually routed through the plasma control console, where the individual flow rates are set by the operator.

The power source should be of a constant-current design. Transistorized power sources are most common, although inverter power supplies are also available. It should have a minimum open-circuit voltage of 80 V to ensure the reliable initiation and transfer of the main arc current. The power source can be adjusted for welding current and it should have the capability to adjust the up slope and down slope of the current. It may be equipped with thumbwheels or potentiometers to select the parameters for pulse current operation, that is, peak and background current levels, as well as peak and background times.

Welding Torches. Like those of the other arc welding processes, PAW torches are available in a range of sizes for different power ratings and in manual and mechanized versions. The design principles are the same in each case. A tungsten alloy electrode is held in a collet within the torch body. To avoid one of the most common defects in plasma torches, it is critical to hold concentricity between the tungsten electrode and the orifice in the design and manufacture of the torch. The electrode assembly is set inside a plenum chamber and the plasma gas is supplied to this chamber. A threaded copper alloy nozzle forms the front of this chamber and contains the nozzle orifice that is used to constrict the plasma arc. A shielding gas nozzle, usually of an insulating ceramic material, is threaded onto the front end of the torch and surrounds the constricting nozzle, creating an annulus through which the shielding gas is supplied. The torch is connected electrically to the power source and the electrode forms the negative pole of the circuit for dc welding. The gas hoses that supply the plasma and shielding gases and the water hoses that supply and remove water from the torch are all connected to the torch body or handle. These hoses are enclosed in a flexible sheath that extends from the torch to the components of the welding system.

Most constricting nozzles have a single orifice in the center. However, multiple-nozzle orifices can be used with higher-power torches to achieve further arc constriction. The most common version of this type of nozzle has a central orifice flanked by a smaller orifice on each side. The common centerline of the three orifices is arranged at 90° to the weld line during operation.

Electrodes. The nonconsummable electrode employed is usually a 2% thoriated tungsten electrode, that is, tungsten with 2% thorium oxide. The electrode specification is covered in AWS A5.12-92. The electrode size is selected according to the welding current level that will be used. The electrode is ground with a tapered point on the end, the angle of which depends on the selected welding current level. Electrode sizes and tip angles for a range of welding current levels are shown in Table 1.

TABLE 1 MAXIMUM CURRENT FOR PLASMA WELDING WITH SELECTED ELECTRODE DIAMETER, VERTEX ANGLE, AND NOZZLE BORE DIAMETER

MAXIMUM CURRENT, A	ELECTRODE DIAM		VERTEX ANGLE, DEGREES	PLASMA(A)				SHIELDING(B)			
				NOZZLE BORE DIAM		FLOW RATE		SHROUD DIAM		FLOW RATE	
	mm	in.		mm	in.	l/min	gal/min	mm	in.	l/min	gal/min
MICROPLASMA											
TORCH RATING, 20 A 5											
5	1.0	0.04	15	0.8	0.03	0.2	0.05	8	0.32	4-7	1-1.8
10				0.8	0.03	0.3	0.08				
20				1.0	0.04	0.5	0.13				
MEDIUM CURRENT											
TORCH RATING, 100 A											
30	2.4	0.10	30	0.79	0.03	0.47	0.12	12	0.48	4-7	1-1.8
50				1.17	0.05	0.71	0.19				
75				1.57	0.06	0.94	0.25				
100				2.06	0.08	1.18	0.31				
TORCH RATING, 200 A											
50	4.8	0.19	30	1.17	0.05	0.71	0.19	17	0.68	4-12	1-3.2
100				1.57	0.06	0.94	0.25				
160				2.36	0.09	1.42	0.38				
200				3.20	0.13	1.65	0.44				
TORCH RATING, 400 A											
180	3.2	0.13	60(C)	2.82	0.11	2.4	0.63	18	0.72	20-35	5.3-9.2
200				2.82(C)	0.11(C)	2.5	0.66				

HIGH CURRENT											
TORCH RATING, 400 A											
250	4.8	0.19	60(C)	3.45(D)	0.14	3.0	0.79	. . .	. . .	20-35	5.3-9.2
300				3.45(D)	0.14(D)	3.5	0.92				
350				3.96(D)	0.16(D)	4.1	1.08				

(A) ARGON PLASMA GAS.

(B) ARGON AND AR-5H_2 SHIELDING GAS.

(C) ELECTRODE TIP BLUNTED TO 1 MM (0.04 IN.) DIAMETER.

(D) MULTIPORT NOZZLE

The new types of tungsten electrodes, which contain oxides of rare-earth elements in place of the thorium oxide, can also be used. These electrodes have been shown to have greater tip life. However, they are more expensive and their usefulness in plasma arc welding may be limited because of the high level of protection provided by the nozzle. In low-current microplasma welding applications, their better emissivity provides easier arc transfer and better overall performance.

If a finer wire is required, to fit the joint or to avoid undercut at the weld toes, then wires suitable for gas-shielded arc welding (for example, those used for GTAW) should be employed. The appropriate AWS specification series is AWS A5.XX.

Plasma (Orifice) and Shielding Gases. The plasma gas is used to generate the arc, whereas the shielding gas is used to provide the weld pool with supplementary shielding from atmospheric contamination while it solidifies and cools. The plasma gas is almost always argon. Gas properties affect both weld shape and quality. Flow rates, particularly of the plasma gas, are also important, because they control the extent of plasma constriction. The flow rate can vary from 0.1 L/min (0.026 gal/min) for microplasma welding up to 10 L/min (2.6 gal/min) for keyhole plasma welding. Considerable care is needed to regulate the gas flow rate if keyhole closure is required, because the flow rate must be sloped out to 1 to 2 L/min (0.26 to 0.52 gal/min) within about 1 s. Gas flow control is best achieved by electronic means.

The design and current rating of welding torches are based on argon plasma gas. Argon provides effective shielding, being heavier than air, and is cheaper than helium. Shielding gas selection is based on the type of base metal (Table 2).

TABLE 2 PLASMA AND SHIELDING GAS COMPOSITIONS

MATERIAL	PLASMA GAS	SHIELDING GAS
MILD STEEL	ARGON	ARGON ARGON-2-5% H_2(A)
LOW-ALLOY STEELS	ARGON	ARGON
AUSTENITIC STAINLESS STEEL	ARGON	ARGON-2-5% H_2 HELIUM(A)
NICKEL AND NICKEL ALLOY	ARGON	ARGON ARGON-2-5% H_2(A)
TITANIUM	ARGON	ARGON 75HE-25AR(A)
ALUMINUM AND ALUMINUM ALLOYS	ARGON	ARGON HELIUM(A)
COPPER AND COPPER ALLOYS	ARGON	ARGON 75HE-25AR(A)

(A) ALSO USED

Helium and argon-helium mixtures can be used as shielding gases to increase the thermal conductivity of the gas and, hence, the heating effect on the weld pool. Helium results in wider weld pools than argon, because it produces a higher arc voltage. Hydrogen additions to argon shielding gas tends to promote slightly narrower weld pools through arc constriction and achieves a very clean weld pool appearance, because it is a reducing gas. Although helium and hydrogen can be added to argon in the shielding gas to give higher heat input, the use of gases with higher heat contents in the plasma gas can result in torch overheating and potential damage.

Plasma Arc Welding

Ian D. Harris, Edison Welding Institute

Applications

Material Types. The PAW process is commonly used to weld stainless steels in a wide range of thicknesses. The process can also be used with carbon and alloy steels, aluminum alloys, titanium alloys, copper and nickel alloys, and more specialized materials, such as zirconium and tantalum. The thicknesses that can be welded in a single pass range from 0.025 mm (1 mil) for microplasma applications to 12.5 mm (0.5 in.) for the VPPA welding of aluminum. Direct-current pulsing can be used on most materials.

The PAW process is often carried out in an autogenous mode, that is, without filler wire. When edge beveling is used, a filler wire is required to complete the joint. A filler wire can also be used with the keyhole mode of operation to avoid undercut at high welding speeds. Wire composition depends on that of the parent materials in the joint. The same continuous-wound wire that is used in GTAW operations is suitable.

The industries that use the PAW process can be categorized as those that weld thin-section sheet using microplasma or medium-current plasma welding and those that weld plate using keyhole plasma welding.

A wide range of small devices and assemblies made from thin stainless steel sheet, including bellows assemblies and associated fittings, are welded using the microplasma operating mode. The narrow weld bead that can be produced provides sheet-metal fabrications with a good cosmetic appearance. Furthermore, the high welding speed that can be achieved, coupled with the good tolerance to stand-off variations resulting from the columnar nature of the arc, makes the process attractive for high-volume production work.

Microplasma, as well as medium-current plasma modes, can be used to spot weld guide wires and lamp filaments, as well as in other applications that require highly repetitive autogenous welds. This type of application allows a user to limit the number of high-frequency arc starts that would be required with tungsten-inert gas welding.

Keyhole plasma welding is extensively used to weld stainless steel pipe and tankage. The process is applied to individual strakes from plate to make stainless steel vessels in the food and chemical processing industries. Circumferential welding of strakes also can be used to create these products. Longitudinal seam welds in stainless steel pipe with wall thickness of [ges]3 mm ([ges]0.12 in.) are ideally suited to keyhole plasma welding, because joint preparation is minimized and single-pass welding can be consistently achieved without the use of weld-backing devices. Pipes with wall thickness above 5 to 6 mm (0.20 to 0.24 in.) employ the keyhole mode of operation for the root pass. Depending on the material type and wall thickness, melt-mode PAW, GTAW, gas-metal arc welding (GMAW), or submerged arc welding (SAW) are used to complete the joint. High-alloy composition piping is similarly manufactured.

The manufacturing of stainless steel tube from strip was one of the first applications of the PAW process. Because the process can reliably produce full-penetration welds without the use of backing, it is extensively used on tube mills, because a lack of access precludes welding from the inside.

The VPPA welding process has been very effective for welding large space shuttle orbiter sections of aluminum alloys, particularly the external fuel tank. The cathodic cleaning action of the electrode positive portion of the current cycle breaks up tenacious surface oxide film on the aluminum. The process can be used on thicknesses up to 12.5 mm (0.5 in.) with square-grooved butt joints, and it can be used for the root pass on thicker sections with edge beveling. It is anticipated that this technique will be extensively employed in building aluminum alloys structures in a modular space station.

Plasma Arc Welding

Typical Components and Joints

The most common joint configuration used with the PAW process is a butt joint. The microplasma mode is used with overlapped butt (micro-lap) joints and with joints that have integral weld metal as a result of flanged, butted edges on very thin metals (Table 3). Corner joints with edge welds are also commonly welded using the microplasma and medium-current modes.

TABLE 3 JOINT CONFIGURATIONS FOR PLASMA WELDING SHEET AND TUBULAR COMPONENTS

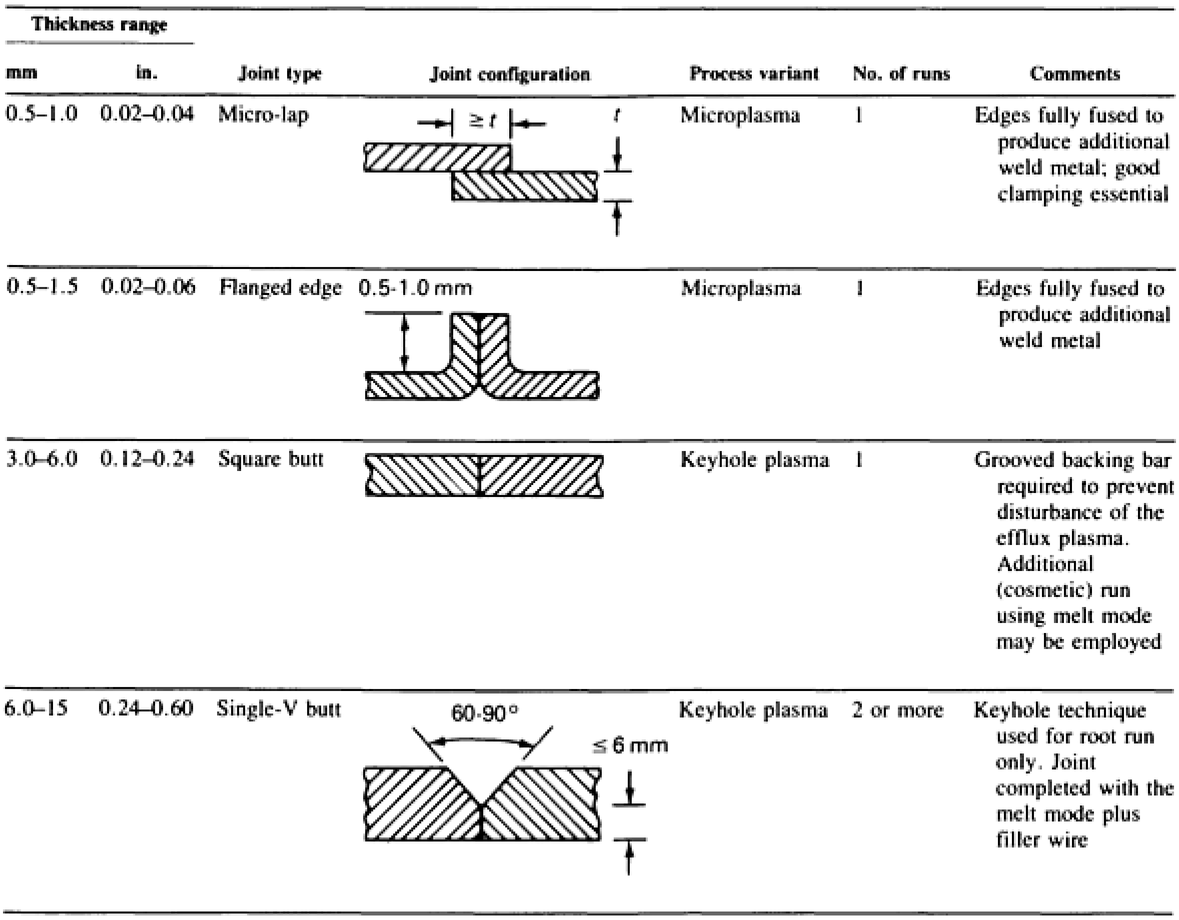
≥ t
t
0.5-1.0 mm
60-90°
≤ 6 mm

Thickness range						
mm	in.	Joint type	Joint configuration	Process variant	No. of runs	Comments
0.5–1.0	0.02–0.04	Micro-lap		Microplasma	1	Edges fully fused to produce additional weld metal; good clamping essential
0.5–1.5	0.02–0.06	Flanged edge		Microplasma	1	Edges fully fused to produce additional weld metal
3.0–6.0	0.12–0.24	Square butt		Keyhole plasma	1	Grooved backing bar required to prevent disturbance of the efflux plasma. Additional (cosmetic) run using melt mode may be employed
6.0–15	0.24–0.60	Single-V butt		Keyhole plasma	2 or more	Keyhole technique used for root run only. Joint completed with the melt mode plus filler wire

Because the keyhole operating mode fully penetrates the workpiece, it is used exclusively on square-grooved butt joints and single-V joints with a root face (Table 3). For square-grooved butt joint preparation, the thickness that can be welded in a single pass depends on the fluid flow characteristics of the workpiece material for a given heat input from the plasma torch. Thus, alloys of titanium and zirconium can be butt welded with square-grooved preparation at greater thicknesses than steels and stainless steels. Generally, it is an industry-accepted practice to use square-grooved preparation without edge beveling for stainless steels up to 6 mm (0.24 in.) in thickness.

A 4 or 5 mm (0.16 or 0.20 in.) root face and a single-V preparation, usually with a 60° included angle, are used on steels thicker than 6 mm (0.24 in.) (Table 4). The root is welded in the keyhole mode and the rest of the joint is completed with a filler wire addition by melt-mode PAW, GTAW, GMAW, or SAW, depending on the material type, joint volume to be filled, and the mechanical property requirements of the joint.

TABLE 4 KEYHOLE PLASMA WELDING CONDITIONS FOR PIPE PREFABRICATION IN TYPE 304L/316L STAINLESS STEEL

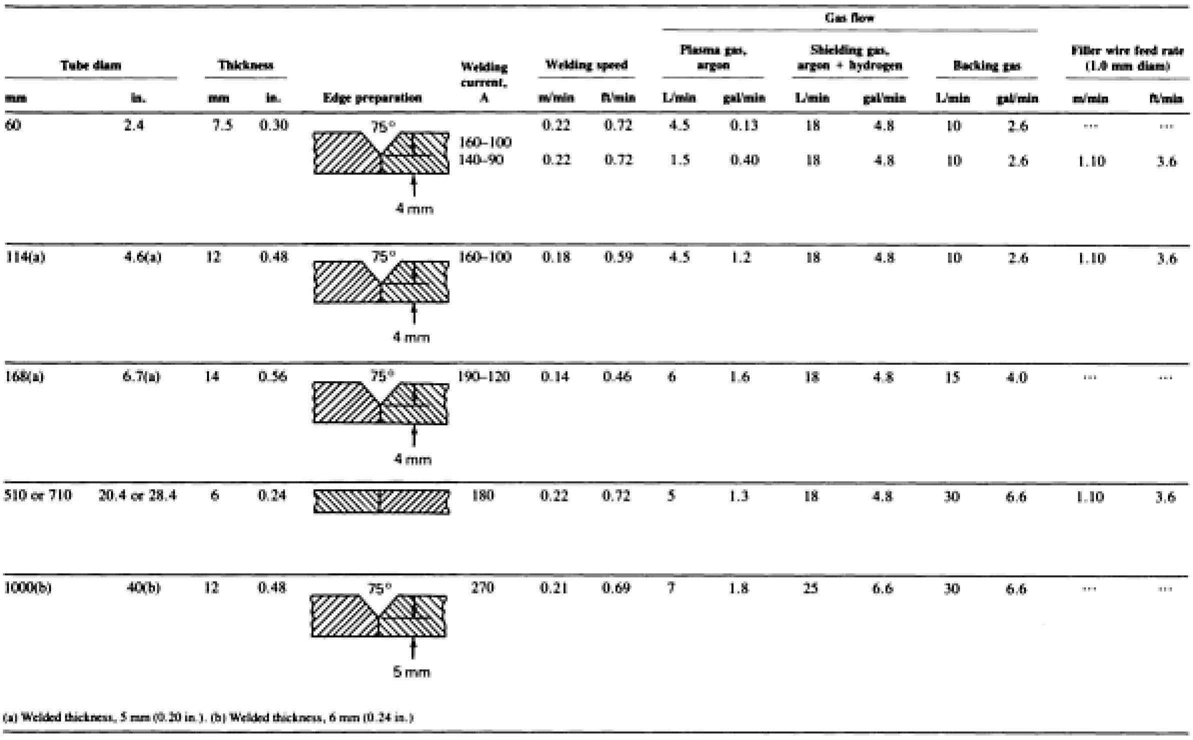

Tube diam		Thickness		Edge preparation	Welding current, A	Welding speed		Gas flow						Filler wire feed rate (1.0 mm diam)	
								Plasma gas, argon		Shielding gas, argon + hydrogen		Backing gas			
mm	in.	mm	in.			m/min	ft/min	L/min	gal/min	L/min	gal/min	L/min	gal/min	m/min	ft/min
60	2.4	7.5	0.30	75° 4 mm	160–100	0.22	0.72	4.5	0.13	18	4.8	10	2.6	...	...
					140–90	0.22	0.72	1.5	0.40	18	4.8	10	2.6	1.10	3.6
114(a)	4.6(a)	12	0.48	75° 4 mm	160–100	0.18	0.59	4.5	1.2	18	4.8	10	2.6	1.10	3.6
168(a)	6.7(a)	14	0.56	75° 4 mm	190–120	0.14	0.46	6	1.6	18	4.8	15	4.0	...	...
510 or 710	20.4 or 28.4	6	0.24		180	0.22	0.72	5	1.3	18	4.8	30	6.6	1.10	3.6
1000(b)	40(b)	12	0.48	75° 5 mm	270	0.21	0.69	7	1.8	25	6.6	30	6.6	...	...

(a) Welded thickness, 5 mm (0.20 in.). (b) Welded thickness, 6 mm (0.24 in.)

(a) Welded thickness, 5 mm (0.20 in.). (b) Welded thickness, 6 mm (0.24 in.)

In the PAW process, fixturing for part fit-up is more critical than it is for the GTAW process, primarily because of the narrower, more-constricted arc in PAW. The backing bar design and shielding gas techniques employed for GTAW are appropriate for microplasma and medium-current modes. During keyhole welding, however, the arc passes right through the workpiece, forming an efflux plasma (Fig. 1) that normally extends 10 mm (0.4 in.) below the back face of the joint. This characteristic must be accommodated when designing backing systems. Failure to allow sufficient gap can result in turbulence in the efflux plasma, causing weld pool disturbance and porosity.

When welding longitudinal seams in tube or pipe, the weld can be started and stopped on run-on and run-off tabs, precluding the need for keyhole closure. However, when using the keyhole mode on circumferential seams of vessels or on butt joints in tubes or pipes, the keyhole will overlap the start of the weld to produce a complete joint. Keyhole closure without porosity has been an area of difficulty, particularly when butt welding tubes, and it represented an important concern for more-widespread use of the process. Modern electronic controls for simultaneous reduction in gas flow and current slope-out provide a more reliable method for keyhole closure.

Plasma Arc Welding

Ian D. Harris, Edison Welding Institute

Procedures

Process Operating Procedure. Welding parameters, such as welding current, arc voltage, travel speed (for mechanized/automatic operation), and plasma and shielding gas flow rates, are set by the procedure and implemented by the welder. Torch parameters include the correct electrode vertex angle and set-back distance, as well as the correct orifice diameter for the welding current level.

The operating sequence for the PAW process was described in the section "Principles of Operation" at the beginning of this article. After the weld pool or keyhole is formed, the torch is traversed across the workpiece at the preset welding speed. Welding is terminated by the down slope of the welding current, with simultaneous sloping of the plasma gas flow rate if keyhole welding and keyhole closure are desired. Slope control for the keyhole mode is shown in Fig. 4.

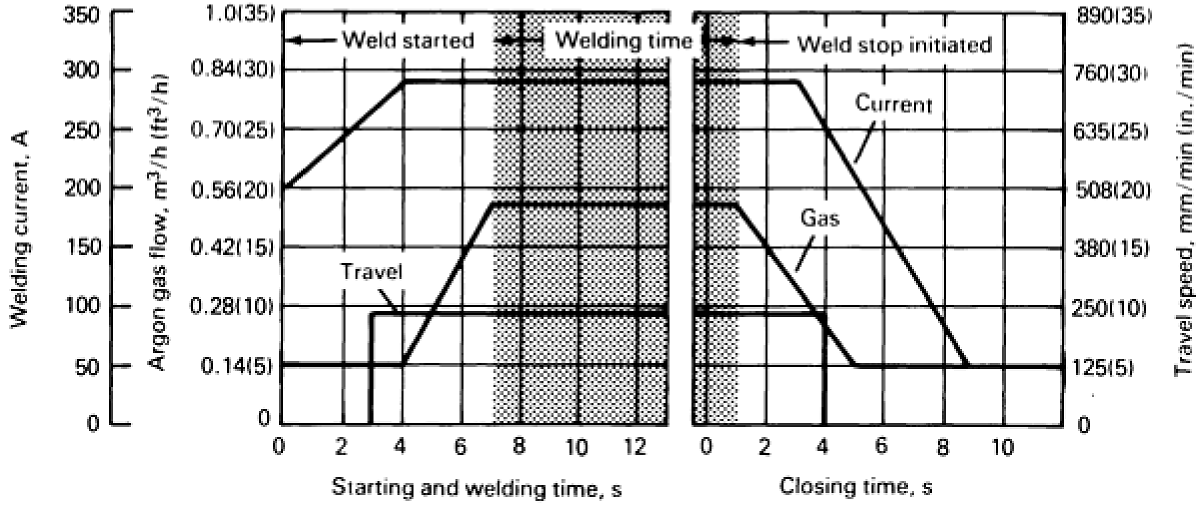

FIG. 4 TYPICAL SLOPE CONTROL PATTERN FOR WELDING CURRENT AND PLASMA GAS FLOW WHEN STARTING AND CLOSING A KEYHOLE; EXAMPLE IS FOR 9.5 MM ($\frac{3}{8}$ IN.) THICK STEEL

Most plasma arc welding is done in a mechanized or automated operation that does not require much operator intervention, except initially to set the parameters and to position the workpiece and torch. When current pulsing and keyhole closure operations are involved, numerous parameters must be set, which requires strict attention to detail, accurate part fit-up, and careful alignment of the torch relative to the joint.

Inspection and Weld Quality Control. All of the common nondestructive evaluation (NDE) techniques are applicable to plasma arc welds. Radiography and ultrasonic inspection are the most common techniques used. The fact that most plasma arc welding is carried out in the keyhole mode means that there is a penetration bead on the root side of the joint. Visual inspection for full penetration, as well as for correct width and profile of both the penetration bead and the weld surface profile, can readily be accomplished. Smooth, even underbeads can be achieved through good procedure development with tolerances improved through current pulsing.

Troubleshooting. When troubleshooting a PAW operation, all of the previously defined basic parameters of the welding schedule should be checked. Welding discontinuities can result from incorrect electrode set-back or from a worn or damaged nozzle orifice. The concentricity of the diameter of the nozzle orifice and the alignment of the electrode and nozzle orifice are very important. Worn nozzles should be replaced. The condition of the tungsten electrode tip should also be checked.

An undercut is one of the most common defects in the PAW process. Depending on material type, thickness, and preset welding parameters, a two-sided undercut can occur above certain threshold welding speeds. This can only be eliminated by increasing welding current. A one-sided undercut can result from misalignment of the torch, the electrode/orifice and multiport nozzles, or from the mismatched fit-up of the workpieces. Correct alignment is essential.

Plasma Arc Welding

Ian D. Harris, Edison Welding Institute

Personnel Requirements

The skill level and training requirements for the PAW process are generally similar to those for the GTAW process. When manual welding, the columnar nature of the plasma arc allows greater variation of the torch-to-workpiece distance without altering the size and shape of the weld pool. The fact that the electrode is protected from contamination by the nozzle means that tungsten contamination and tungsten inclusions in the weld metal are more readily avoided than they are in the GTAW process. From the viewpoint of an operator, disadvantages include the greater complexity of the torch, in terms of nozzle orifice and electrode set-back considerations, and the correct setting of two gases (plasma and shielding gas).

When mechanized welding is carried out, which is usually the case for the keyhole mode, the operator is required to set a greater number of parameters (compared to GTAW), particularly if current pulsing and keyhole closure operations are required. Specific operator training is required for pulsed keyhole operation, and particularly for VPPA operations, which are usually carried out in the vertical position, welding upward, and thus require a high level of operator competence. Computerization of welding parameter selection is the trend for VPPA welding and other mechanized or automated PAW operations.

The health and safety issues related to the PAW process are very similar to those of other gas-shielded arc welding processes, especially GTAW. These include electrical shock; electromagnetic radiation hazards, particularly ultraviolet radiation; burns from hot metal parts; and welding fumes and gases, including ozone. The volume of welding fumes produced is low and is similar to that produced by the GTAW process. Hexavalent chromium and ozone are concerns when welding stainless steels and aluminum alloys, respectively. Like the GTAW process, the low level of general welding fumes associated with PAW results in a comparatively higher level of ozone. However, the rapid decay of ozone within a short distance of the arc, coupled with the high degree of mechanization and automation typical of the keyhole plasma welding operation, means that operator exposure is generally very low. The microplasma mode, which is more commonly used manually, employs such low current levels that fume and ozone levels are very low.

Carbon Arc Welding

Lance Soisson, Welding Consultants, Inc.

Introduction

CARBON ARC WELDING (CAW) utilizes what is considered to be a nonconsumable electrode, made of carbon or graphite, to establish an arc between itself and either the workpiece or another carbon electrode. However, this electrode erodes fairly quickly and generates carbon monoxide (CO) gas that partially replaces the air around the arc, thereby providing the molten weld with some protection.

The CAW process, which uses either single or twin electrodes, most closely resembles gas-tungsten arc welding (GTAW), where the arc is used only as a source of heat. The single-electrode arrangement usually operates with direct current (dc), electrode negative (straight polarity), using most dc power supplies. The twin-electrode arrangement usually operates with alternating current (ac), generally with small ac power supplies. Figure 1 shows typical configurations for both procedures.

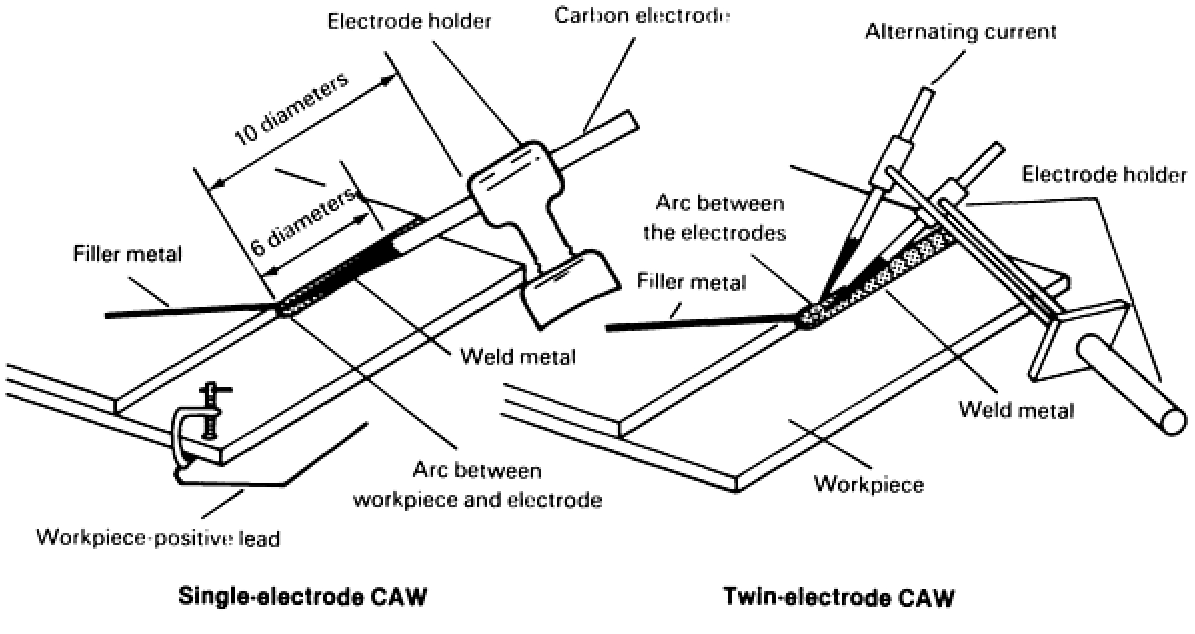

Single-electrode CAW **Twin-electrode CAW**

FIG. 1 TYPICAL ARRANGEMENTS FOR SINGLE-ELECTRODE AND TWIN-ELECTRODE CARBON ARC WELDING

Although the CAW process has been almost completely replaced by the newer processes used in the welding industry, it is still used in certain applications. The process does produce adequate welds in thin-sheet steel, but should be used with caution in any critical application because it provides only limited shielding from the atmosphere.

Carbon Arc Welding

Lance Soisson, Welding Consultants, Inc.

Operation

CAW electrodes can be either carbon or graphite, although baked carbon electrodes are most commonly used because they last longer and are more readily available. Electrode sizes range from 3.2 to 22 mm ($\frac{1}{8}$ to $\frac{7}{8}$ in.). The typical torch is air cooled for the smaller electrodes, but water cooled for the larger ones. The recommended current ranges for both electrode types are given in Table 1.

TABLE 1 RECOMMENDED CURRENT RANGES FOR CARBON AND GRAPHITE ELECTRODES

ELECTRODE DIAMETER		CURRENT, A	
mm	in.	CARBON	GRAPHITE
SINGLE ELECTRODE			
3.2	$\frac{1}{8}$	15-30	15-35
4.8	$\frac{3}{16}$	25-55	25-60
6.4	$\frac{1}{4}$	50-85	50-90
7.9	$\frac{5}{16}$	75-115	80-125
9.5	$\frac{3}{8}$	100-150	110-165
11.1	$\frac{7}{16}$	125-185	140-210
12.7	$\frac{1}{2}$	150-225	170-260
15.9	$\frac{5}{8}$	200-310	230-370
19	$\frac{3}{4}$	250-400	290-490
22.2	$\frac{7}{8}$	300-500	400-750
TWIN ELECTRODES			
6.4	$\frac{1}{4}$	55	. . .
7.9	$\frac{5}{16}$	75	. . .
9.5	$\frac{3}{8}$	95-120	. . .

In the single-electrode operation, the carbon electrode is prepared to a long tapered point that is one half the electrode diameter (see Fig. 1). The arc is established by "scratch starting," that is, by bringing the electrode into contact with the workpiece and immediately withdrawing it to the correct length for welding. The generally accepted arc length is maintained between 6.4 and 9.5 mm ($\frac{1}{4}$ and $\frac{3}{8}$ in.). Holding the length too short can result in carburization of the base metal, creating potentially brittle welds. This is particularly true when welding without filler metal.

When the joint requires filler metal, the same technique that is used in GTAW or oxyfuel gas welding (OFW) should be employed. The welding rod should be fed into the weld pool with one hand while manipulating the arc with the other.

If the arc is broken for any reason, then restart should not occur directly at the hot metal where welding stopped, because a hard spot could result. Rather, a restart should occur in front or to one side of the weld, followed by a quick return to the point where welding stopped. Care should be taken to either incorporate the restart area into the weld or to grind the arc strike after welding is completed. The arc strike areas can be very hard and can cause cracking if they are not fused into the weld or ground out.

In the twin-electrode operation, the workpiece is not part of the electrical circuit. The heat for welding is produced by creating an arc between the two electrodes. The electrode holder has two adjustable arms that grip the carbon electrodes at the proper length to maintain the arc. The operator maintains a constant distance between the carbon electrodes by adjusting them as they are consumed. An experienced welder can make these adjustments while still welding or brazing.

Small ac machines are normally used in the twin-electrode process because the electrodes will wear at an equal rate. Direct current can be used as long as the positive-side electrode is one size larger than the negative electrode. The larger electrode compensates for the faster consumption rate on the positive side of the direct current.

Electrodes range in size from 6.4 to 9.5 mm ($\frac{1}{4}$ to $\frac{3}{8}$ in.) and are copper coated for better current flow. The heat must be maintained so that the copper does not melt more than 13 mm ($\frac{1}{2}$ in.) back from the electrode tip. Only enough heat to cause the filler metal to flow easily should be used (Table 1). High heat causes excessive electrode consumption and could cause carburization of the weld metal.

Carbon Arc Welding

Lance Soisson, Welding Consultants, Inc.

Applications

The single-electrode carbon arc process is primarily used on steel, cast iron, and copper, although it can be used on most ferrous and nonferrous materials. With steels, it is principally applied to make outside corner welds on thinner-gage materials, where no filler metal is used. A good fit-up is required, and a fluxing agent is often used to promote better welds. The resulting welds arc smoother and, frequently, more economical than similar welds created by the shielded metal arc welding (SMAW) process.

Another application is the welding of galvanized steel using a silicon bronze filler metal. In this instance, it is important to maintain a short arc and to direct the heat to the filler metal to minimize damage to the galvanic coating.

Cast iron can be welded using the carbon arc process, but it is more often used in brazing. The casting must be preheated to approximately 650 °C (1200 °F) and then slowly cooled to reduce the possibility of cracking and to produce machinable welds.

The process is also used on copper, because of the high heat capabilities of the larger electrodes. Direct current, electrode negative, should always be used when welding or brazing copper. The workpiece should be preheated to a temperature between 150 and 650 °C (300 and 1200 °F), or at least locally preheated in the weld area. Holding a long arc will allow the carbon to combine with oxygen to form the CO that somewhat shields the weld.

Twin carbon arc welding is used primarily as a maintenance tool in small shops, farms, or homes. It is most commonly applied in brazing or soldering operations, although it can be used to weld, preheat, or postheat smaller parts.

Carbon Arc Welding

Lance Soisson, Welding Consultants, Inc.

Selected References

- H.B. CARY, *MODERN WELDING TECHNOLOGY,* 2ND ED., PRENTICE-HALL, 1989
- *WELDING HANDBOOK,* 8TH ED., VOL 2, *WELDING PROCESSES,* AWS, 1991

Submerged Arc Welding

Jonathan S. Ogborn, The Lincoln Electric Company

Introduction

SUBMERGED ARC WELDING (SAW) is an arc welding process in which the arc is concealed by a blanket of granular and fusible flux (Ref 1, 2, 3). Heat for SAW is generated by an arc between a bare, solid-metal (or cored) consumable-wire or strip electrode and the workpiece. The arc is maintained in a cavity of molten flux or slag, which refines the weld metal and protects it from atmospheric contamination. Alloy ingredients in the flux may be present to enhance the mechanical properties and crack resistance of the weld deposit.

References

1. R.L. O'BRIEN, *WELDING HANDBOOK,* VOL 11, AWS, 1991, P 191-232
2. D.L. OLSON *ET AL.,* SUBMERGED ARC WELDING, VOL 6, 9TH ED., *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1983, P 114-152
3. *THE PROCEDURE HANDBOOK OF ARC WELDING,* THE LINCOLN ELECTRIC CO., 1973, P 6.3-1 TO 6.3-24

Submerged Arc Welding

Jonathan S. Ogborn, The Lincoln Electric Company

Principles of Operation

Figure 1 shows a typical setup for automatic SAW. A continuous electrode is being fed into the joint by mechanically powered drive rolls. A layer of granular flux, just deep enough to prevent flash through, is being deposited in front of the arc. Electrical current, which produces the arc, is supplied to the electrode through the contact tube. The current can be direct current (dc) with electrode positive (reverse polarity), with electrode negative (straight polarity), or alternating current (ac). Figure 2 shows the melting and solidification sequence of SAW. After welding is completed and the weld metal has solidified, the unfused flux and slag are removed. The unfused flux may be screened and reused. The solidified slag may be collected, crushed, resized, and blended back into new flux. Recrushed slag and blends of recrushed slag with unused (virgin) flux are chemically different from new flux. Blends of recrushed slag may be classified as a welding flux, but cannot be considered the same as the original virgin flux (Ref 4).

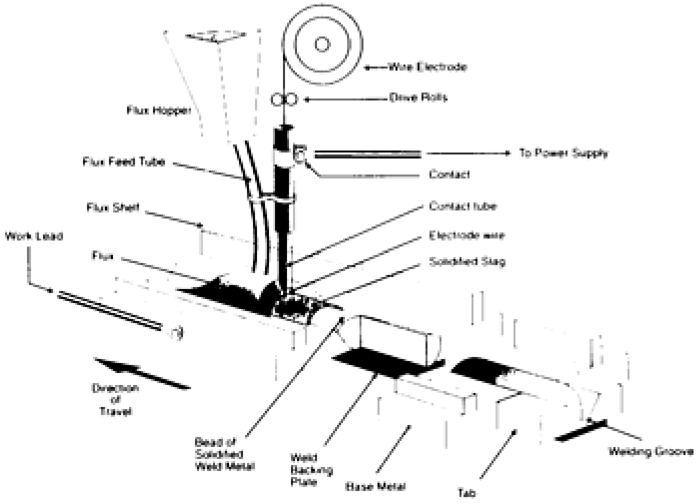
Wire Electrode
Drive Rolls
Flux Hopper
To Power Supply
Flux Feed Tube
Contact
Flux Shelf
Contact tube
Work Lead
Electrode wire
Flux
Solidified Slag
Direction of Travel
Bead of Solidified Weld Metal
Weld Backing Plate
Base Metal
Tab
Welding Groove

FIG. 1 SCHEMATIC SHOWING KEY COMPONENTS OF AUTOMATIC SUBMERGED ARC WELDING SETUP

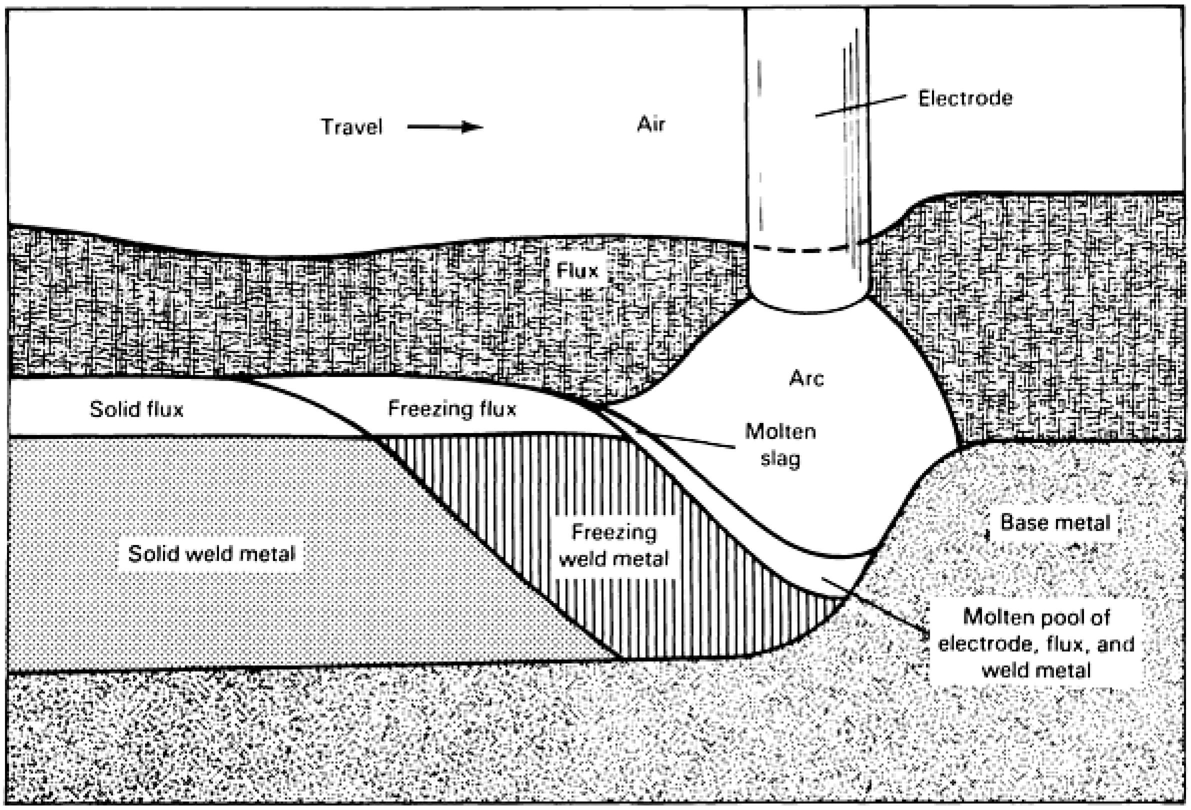
Travel
Air
Electrode
Flux
Arc
Solid flux
Freezing flux
Molten slag
Base metal
Solid weld metal
Freezing weld metal
Molten pool of electrode, flux, and weld metal

FIG. 2 SCHEMATIC SHOWING TYPICAL WELD POOL DYNAMICS OF A SUBMERGED ARC WELD

Submerged arc welding is adaptable to both semiautomatic and fully automatic operation, although the latter, because of its inherent advantages, is more popular (see the section "Advantages of Submerged Arc Welding" in this article). In semiautomatic welding, the welder controls the travel speed, direction, and placement of the weld. A semiautomatic welding gun is designed to transport the flux and wire to the operator, who welds by dragging the gun along the weld joint. Semiautomatic electrode diameters are usually less than 2.4 mm ($\frac{3}{32}$ in.) to provide sufficient flexibility and feedability in the gun assembly. Manually guiding the gun over the joint requires skill because the joint is obscured from view by the flux layer. In automatic SAW, travel speed and direction are controlled mechanically. Flux may be automatically deposited in front of the arc, while the unfused flux may be picked by a vacuum recovery system behind the arc.

To increase deposition rate or welding speed, more than one wire can be fed simultaneously into the same weld pool. For example, in the twin arc process, two electrodes are fed into the same weld pool while sharing a common power source and contact tip. In tandem arc SAW, multiple electrodes are arranged with one in front of the other. Each electrode has an independent power supply and contact tip. The spacing, configuration, and electrical nature of the electrodes may be arranged to optimize welding speed and bead shape (Ref 5).

Advantages of submerged arc welding include the following:

- THE ARC IS UNDER A BLANKET OF FLUX, WHICH VIRTUALLY ELIMINATES ARC FLASH, SPATTER, AND FUME (THUS MAKING THE PROCESS ATTRACTIVE FROM AN ENVIRONMENTAL STANDPOINT).
- HIGH CURRENT DENSITIES INCREASE PENETRATION AND DECREASE THE NEED FOR EDGE PREPARATION.
- HIGH DEPOSITION RATES AND WELDING SPEEDS ARE POSSIBLE.
- COST PER UNIT LENGTH OF JOINT IS RELATIVELY LOW.
- THE FLUX ACTS AS A SCAVENGER AND DEOXIDIZER TO REMOVE CONTAMINANTS SUCH AS OXYGEN, NITROGEN, AND SULFUR FROM THE MOLTEN WELD POOL. THIS HELPS TO PRODUCE SOUND WELDS WITH EXCELLENT MECHANICAL PROPERTIES.
- LOW-HYDROGEN WELD DEPOSITS CAN BE PRODUCED.
- THE SHIELDING PROVIDED BY THE FLUX IS SUBSTANTIAL AND IS NOT SENSITIVE TO WIND AS IN SHIELDED METAL ARC WELDING AND GAS METAL ARC WELDING.
- MINIMAL WELDER TRAINING IS REQUIRED (THUS, RELATIVELY UNSKILLED WELDERS CAN BE EMPLOYED).
- THE SLAG CAN BE COLLECTED, REGROUND, AND SIZED FOR MIXING BACK INTO NEW FLUX AS PRESCRIBED BY MANUFACTURERS AND QUALIFIED PROCEDURES.

Limitations of submerged arc welding include the following:

- THE INITIAL COST OF WIRE FEEDER, POWER SUPPLY, CONTROLS, AND FLUX-HANDLING EQUIPMENT IS HIGH.
- THE WELD JOINT NEEDS TO BE PLACED IN THE FLAT OR HORIZONTAL POSITION TO KEEP THE FLUX POSITIONED IN THE JOINT.
- THE SLAG MUST BE REMOVED BEFORE SUBSEQUENT PASSES CAN BE DEPOSITED.
- BECAUSE OF THE HIGH HEAT INPUT, SAW IS MOST COMMONLY USED TO JOIN STEELS MORE THAN 6.4 MM ($\frac{1}{4}$ IN.) THICK.

References cited in this section

4. "SPECIFICATION FOR CARBON STEEL ELECTRODES AND FLUXES FOR SUBMERGED ARC WELDING," A5.17-89, AWS, 1989, P 18
5. *THE PROCEDURE HANDBOOK OF ARC WELDING,* THE LINCOLN ELECTRIC CO., 1973, P 3.2-3, 6.3-21 TO 6.3-22, 6.3-49 TO 6.3-58

Submerged Arc Welding

Jonathan S. Ogborn, The Lincoln Electric Company

Process Applications

If a steel is suitable for welding with gas-metal arc welding, flux-cored arc welding, shielded metal arc welding, or gas-tungsten arc welding, procedures can be developed to weld the steel with SAW (Ref 6). The main limitations of SAW are plate thickness and position. Because SAW is a high heat input and high deposition rate process, it is generally used to weld thicker-section steels. Although the welding of 1.6 mm ($\frac{1}{16}$ in.) thick steel is possible, most SAW is done on plate over 6.4 mm ($\frac{1}{4}$ in.) in thickness.

Types of Metals. Submerged arc welding is most commonly used to join plain carbon steels. Alloy steels can be readily welded with SAW if care is taken to limit the heat input as required to prevent damage to the heat-affected zone (HAZ). Low-heat-input procedures are available for welding alloy steels and heat-treated steels to prevent grain coarsening and cracking in the HAZ. Maintaining proper preheat and interpass temperature is also important when welding alloy steels to prevent weld metal and HAZ cracking and to develop the required mechanical properties in the weld deposit. Submerged arc welding can be used to join stainless steels and nonferrous alloys. It is also commonly used to produce a stainless or nonferrous overlay on top of a base metal.

Joint Configurations. The most common weld deposits made with SAW are groove, fillet, and plug welds, and surfacing deposits. For groove welds, the characteristic deep penetration of SAW plays a role in joint selection. Plate up to 15.9 mm ($\frac{5}{8}$ in.) thick can be completely welded from one side using a square butt joint with a 0.8 mm ($\frac{1}{32}$ in.) root opening. Beveled joints should be used for multiple-pass welds on plate thicknesses greater than 15.9 mm ($\frac{5}{8}$ in.). Backing bars should be used to prevent loss of flux, to prevent melt through, and to ensure full-penetration weldments on one-sided joints. Although groove welds are generally made in the flat position, it is possible to develop special procedures to weld horizontally.

Single-pass fillet welds with throat sizes up to 7.9 mm ($\frac{5}{16}$ in.) thick and multiple-pass fillet welds are usually made in the horizontal position. Single-pass fillet welds over 7.9 mm (5/16 in.) thick are usually made in the flat position. Submerged arc plug welds are made in the flat position by puddling the electrode into the center of the hole until the weld is complete. Surfacing welds are made with both wire and strip electrode. Surfacing is done to rebuild worn parts with a wear-resistant material or to overlay a plain steel part with stainless steel or other alloys (Ref 7, 8, 9).

To prevent porosity, the surface to be welded should be clean and free of grease, oil, paints, moisture, and oxides. All slag from tack welds or previous layers should be removed. Tack welds should be positioned so that the submerged arc weld completely melts out the tack. The workpiece should be firmly clamped to minimize distortion and the need for tack welds.

Industrial Uses for SAW. Because SAW is used to join thick steel sections, it is primarily used for shipbuilding, pipe fabrication, pressure vessels, and structural components for bridges and buildings. Other than joining, SAW is used to build up parts and overlay with stainless or wear-resistant steel (for example, rolls for continuous casting steel, pressure vessels, rail car wheels, and equipment for mining, mineral processing, construction, and agriculture).

References cited in this section

6. *THE PROCEDURE HANDBOOK OF ARC WELDING,* THE LINCOLN ELECTRIC CO., 1973, P 6.3-26 TO 6.3-73
7. R.S. BROWN *ET AL.,* ARC WELDING OF STAINLESS STEELS, VOL 6, 9TH ED., *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1983, P 327-329
8. K.C. ANTONY *ET AL.,* HARDFACING, VOL 6, 9TH ED., *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1983, P 784
9. D.R. THOMAS *ET AL.,* WELD OVERLAYS, VOL 6, 9TH ED., *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1983, P 804-819

Submerged Arc Welding

Jonathan S. Ogborn, The Lincoln Electric Company

Equipment and Supplies

The power supplied to the contact tip may be direct current or alternating current. In the United States, dc power is most common for welding applications that require less than 1000 A. In Europe and the Far East, however, ac is used extensively, even in applications requiring less than 1000 A. Currents over 1000 A dc tend to create arc blow problems. Alternating current is most commonly used for high-current applications, for applications where arc blow may be a problem, and in multiwire applications. Direct-current power supplies may be either transformer-rectifiers or motor/engine generators that produce either constant voltage or constant-current power. Alternating-current power sources are most commonly transformers.

To produce a submerged arc weld, both flux and electrode are consumed. Each flux and electrode combination, along with the variation of base material and process parameters, will produce a unique weld deposit. Because the integrity of the weld deposit depends on these parameters, specific fluxes and electrodes must be used in combination to optimize the weld metal properties.

Power Sources

A constant-voltage power supply is self-regulating, so it can be used with a constant-speed wire feeder. No voltage or current sensing is required. The current is controlled by the wire diameter, the electrical stickout, and the wire-feed speed, while the voltage is controlled by the power supply. Constant-voltage dc power is the best choice for the high-speed welding of thin steel.

Unlike constant voltage, constant-current power supplies are not self-regulating, so they must be used with voltage-sensing variable-wire-feed speed controls. A constant-current wire feeder monitors arc voltage and adjusts the wire-feed speed in response to changes in the arc voltage. The wire-feed speed control attempts to maintain a constant arc length, while the power supply controls the arc current.

The constant-current output of a conventional ac machine varies with time like a sine wave dropping through zero with each polarity reversal. The voltage associated with the current is approximately a square wave. Because the power output is zero with each polarity change, an open-circuit voltage greater than 80 V may be required to ensure arc initiation. The constant-current ac machine requires voltage-sensing variable-wire-feed speed controls. On newer, solid-state power supplies, the current and voltage output both approximate square waves, with the instantaneous polarity reversal reducing arc initiation problems. The solid-state power supplies have constant-voltage characteristics that may be used with constant-speed wire-feed controls.

Flux Classification

Fluxes can be categorized depending on the method of manufacture, the extent to which they can affect the alloy content of the weld deposit, and the effect on weld deposit properties.

Classification Relative to Production Method. Based on the manufacturing process, there are two different types of fluxes: fused and bonded.

Fused Fluxes. The raw materials for a fused flux are dry mixed and melted in a furnace. The molten mixture is then rapidly solidified, crushed, screened, and packaged. Because of their method of manufacture, fused fluxes typically do not contain ferroalloys and deoxidizers.

Bonded Fluxes. The powdered ingredients of a bonded flux are dry blended, and then mixed with a binder, usually potassium or sodium silicate. After pelletizing, the wet flux is dried in an oven or kiln, sized appropriately, and then packaged. The relatively low baking temperatures allow bonded fluxes to contain deoxidizers and ferroalloys.

Classification Relative to Effect on Alloy Content of Weld Deposit. Independent of manufacturing method, a given flux may be described as an active, neutral, or alloy flux, depending on its ability to change the alloy content of the weld deposit. With all submerged arc fluxes, variations in arc voltage and other welding variables will change the ratio of flux consumed to electrode or weld metal deposited. This ratio is often referred to as the flux-to-wire ratio. Normal flux-to-wire ratios for SAW are 0.7 to 0.9. An increase in the flux-to-wire ratio may be caused by either an increase in arc voltage or a decrease in the welding current. Likewise, a decrease in the flux-to-wire ratio may be caused by a decrease in arc voltage or an increase in the welding current. How the weld deposit composition changes with voltage (flux-to-wire ratio) provides an additional means of describing a flux.

Active fluxes contain controlled amounts of manganese and/or silicon. These alloys are added as ingredients in the flux to provide improved resistance to porosity and weld cracking caused by contaminants such as oxygen, nitrogen, and sulfur on the plate or in the plate composition itself. Active fluxes are primarily used to make single-pass welds. Because active fluxes contain deoxidizers such as manganese and silicon, the alloy in the weld metal will change with the flux-to-wire ratio. Changes in manganese and silicon content in the weld deposit will affect the strength and impact properties of the weld metal; therefore, the arc voltage must be more tightly controlled when welding with active fluxes than with neutral fluxes.

Neutral fluxes contain little or no deoxidizers, and by definition, will not produce any significant change in the weld metal composition as a result of a large change in arc voltage. Because neutral fluxes contain little or no alloy, they must rely on alloy in the electrode to provide deoxidation. Single-pass welds on oxidized plate may be prone to cracking or porosity due to insufficient alloy content and are usually best welded with active fluxes. Neutral fluxes are primarily used for welding multiple passes on plate exceeding 25 mm (1 in.) in thickness.

To quantify the active/neutral behavior of a flux, the Wall neutrality number (N) was developed (Ref 10) and is defined as:

$$N = 100\,[|\Delta\%\mathrm{SI}| + |\Delta\%\mathrm{MN}|]$$

(EQ 1)

where $\Delta\%\mathrm{Si}$ is the change in the weight percent of silicon, and $\Delta\%\mathrm{Mn}$ is the change in the weight percent of manganese. To determine the Wall neutrality number, two different weld deposit pads must be made. One weld pad is welded with the same parameters as those specified for the weld test plate for the flux and electrode being used. The other weld pad is welded with the same parameters except that the voltage is increased by 8 V. The $\Delta\%\mathrm{Si}$ and $\Delta\%\mathrm{Mn}$ quantities in Eq 1 are then determined by the difference in the amount of silicon and manganese in the two deposits. A flux electrode combination that produces a Wall neutrality number of 40 or less is considered neutral, while a number of 40 indicates an active flux.

Alloy fluxes contain enough alloy as a flux ingredient to produce an alloy weld metal with a carbon steel electrode. They are also used with alloy and stainless steel wire and strip electrodes. Alloy fluxes find application primarily in welding alloy steels and hardfacing. Because the alloy in the weld deposit is a function of arc voltage (flux-to-wire ratio), it is important to follow the manufacturer's recommended procedure for obtaining proper weld deposit composition.

Classification Relative to Basicity Index. Aside from the manufacturing method and the active, neutral, or alloy behavior of flux types, another common method used to describe submerged arc fluxes is the basicity index, *BI*. The basicity index is the ratio of strongly bound metallic oxides to weakly bound metallic oxides (Ref 11). The Boniszewski basicity index is defined by:

$$BI = \frac{\begin{array}{c}\mathrm{CaO} + \mathrm{CaF_2} + \mathrm{MgO} \\ + \mathrm{K_2O} + \mathrm{Na_2O} + \mathrm{Li_2O} + \\ \mathrm{BaO} + \mathrm{SrO} + {}^1\!/_2(\mathrm{MnO} + \mathrm{FeO})\end{array}}{\mathrm{SiO_2} + {}^1\!/_2(\mathrm{Al_2O_3} + \mathrm{TiO_2} + \mathrm{ZrO_2})}$$

(EQ 2)

The basicity index is an estimate of the oxygen content in the weld metal and is therefore used to predict weld metal properties. Basic fluxes tend to have lower weld metal oxygen content with good weld metal toughness, while acidic fluxes tend to produce higher weld metal oxygen content and coarser microstructure with a lower resistance to cleavage. Fluxes with a basicity index greater than 1.5 are considered basic; fluxes below 1.0 are considered acidic.

Acid fluxes are typically preferred for single-pass welding because of their superior operating and bead-wetting characteristics. In addition, acid fluxes usually have more resistance than basic fluxes to porosity caused by plate contamination by oil, rust, and mill scale.

Basic fluxes tend to have better impact properties than acid fluxes. This advantage is particularly evident on large multiple-pass welds. Highly basic fluxes generally produce weld metal with very good impact properties in large multiple-pass weldments. Basic fluxes tend to exhibit poorer welding characteristics than acid fluxes on single-pass welds. Their use should be limited to large multipass weldments where good weld metal notch toughness is required.

Selection of Electrodes Relative to Steel and Alloy Type

There are three different types of consumable electrodes: solid wire, cored wire, and strip. Electrodes for SAW are available for welding carbon steel, low-alloy steel, stainless steel, and nickel-base alloys. Wire diameters vary from 1.6 to 6.4 mm ($\frac{1}{16}$ to $\frac{1}{4}$ in.). Carbon steel wire usually has a light copper coating, which protects the wire from corrosion and provides good electrical contact in the welding tip.

Cored electrodes were developed to provide a low-cost method of increasing the number of alloys that can be welded with SAW. Cored electrodes typically consist of a mild steel tube with alloying elements in the center. Compositions of cored electrodes vary from mild and low-alloy steels through tool steels to stainless steels and high-alloy austenitic steels.

Strip electrodes are generally used for welding overlays and for hardfacing applications. The advantages of strip are wide weld beads, low penetration, low dilution, and high deposition rates. The equipment is very similar to SAW with wire except that special drive rolls are needed to feed the strip, and sometimes auxiliary magnetic fields are set up to improve bead shape and tie-in. Strip thicknesses vary from 0.5 to 1.0 mm (0.02 to 0.04 in.), and widths vary from about 25 to 100 mm (1 to 4 in.).

Plain Carbon Steels. The requirements for the classification of electrodes and fluxes for welding carbon steels are detailed in Ref 12. Compositional requirements for solid electrodes are given in Table 1. Requirements for composite electrodes are defined in Table 2. Classification of solid electrodes is based on the chemical composition of the electrode. Classification of composite electrodes is based on weld deposit composition. Flux classification is based on the soundness and mechanical properties of the weld metal produced with a particular flux and electrode combination. The classification system used to describe flux-electrode combinations for mild steel welding is defined in Fig. 3. Interpretation of the classification system code is illustrated in the flux-electrode combination examples that follow.

TABLE 1 COMPOSITION OF SOLID-CARBON STEEL ELECTRODES

ELECTRODE		COMPOSITION, WT%$^{(A)(B)}$						
AWS CLASSIFICATION	UNS NO.	C	Mn	Si	S	P	Cu$^{(c)}$	Ti
LOW-MANGANESE ELECTRODES								
EL8	K01008	0.10	0.25/0.60	0.07	0.030	0.030	0.35	. . .
EL8K	K01009	0.10	0.25/0.60	0.10/0.25	0.030	0.030	0.35	. . .
EL12	K01012	0.04/0.14	0.25/0.60	0.10	0.030	0.030	0.35	. . .
MEDIUM-MANGANESE ELECTRODES								
EM12	K01112	0.06/0.15	0.80/1.25	0.10	0.030	0.030	0.35	. . .
EM12K	K11113	0.05/0.15	0.80/1.25	0.10/0.35	0.030	0.030	0.35	. . .
EM13K	K01313	0.06/0.16	0.90/1.40	0.35/0.75	0.030	0.030	0.35	. . .
EM14K	K01314	0.06/0.19	0.90/1.40	0.35/0.75	0.025	0.025	0.35	0.03-0.17
EM15K	K01515	0.10/0.20	0.80/1.25	0.10/0.35	0.030	0.030	0.35	. . .
HIGH-MANGANESE ELECTRODES								
EH11K	K11140	0.07/0.15	1.40/1.85	0.80/1.15	0.030	0.030	0.35	. . .
EH12K	K01213	0.06/0.15	1.50/2.00	0.25/0.65	0.025	0.025	0.35	. . .
EH14	K11585	0.10/0.20	1.70/2.20	0.10	0.030	0.030	0.35	. . .

Source: Ref 12

(A) THE FILLER METAL SHALL BE ANALYZED FOR THE SPECIFIC ELEMENTS FOR WHICH VALUES ARE SHOWN IN THIS TABLE. IF THE PRESENCE OF OTHER ELEMENTS IS INDICATED, IN THE COURSE OF THIS WORK, THE AMOUNT OF THOSE ELEMENTS SHALL

BE DETERMINED TO ENSURE THAT THEIR TOTAL (EXCLUDING IRON) DOES NOT EXCEED 0.50%.

(B) SINGLE VALUES ARE MAXIMUM.

(C) COPPER LIMIT INCLUDES ANY COPPER COATING THAT MAY BE APPLIED TO THE ELECTRODE.

TABLE 2 COMPOSITION OF CARBON STEEL WELD METAL MADE WITH COMPOSITE ELECTRODE

ELECTRODE AWS CLASSIFICATION	COMPOSITION, WT%(A)(B)(C)					
	C	Mn	Si	S	P	Cu
EC1	0.15	1.80	0.90	0.035	0.035	0.35

Source: Ref 12

(A) WELD METAL SHALL BE ANALYZED FOR THE SPECIFIC ELEMENTS FOR WHICH VALUES ARE SHOWN IN THIS TABLE. IF THE PRESENCE OF OTHER ELEMENTS IS INDICATED, IN THE COURSE OF THIS WORK, THE AMOUNT OF THOSE ELEMENTS SHALL BE DETERMINED TO ENSURE THAT THEIR TOTAL (EXCLUDING IRON) DOES NOT EXCEED 0.50%.

(B) SINGLE VALUES ARE MAXIMUM.

(C) A LOW-DILUTION AREA OF A GROOVE WELD OR A FRACTURED TENSION TEST SPECIMEN AS DESCRIBED IN AWS A5.17 (SEE REF 10) MAY BE SUBSTITUTED FOR THE WELD PAD, AND SHALL MEET THE ABOVE REQUIREMENTS. IN CASE OF DISPUTE, THE WELD PAD SHALL BE THE REFEREE METHOD.

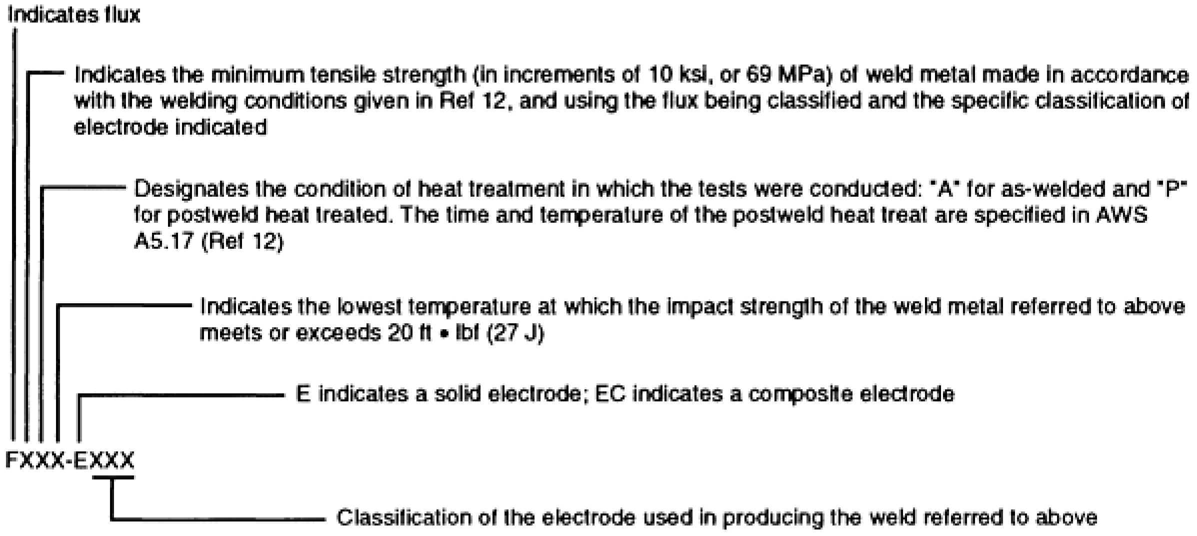
Indicates flux
Indicates the minimum tensile strength (in increments of 10 ksi, or 69 MPa) of weld metal made in accordance with the welding conditions given in Ref 12, and using the flux being classified and the specific classification of electrode indicated
Designates the condition of heat treatment in which the tests were conducted: "A" for as-welded and "P" for postweld heat treated. The time and temperature of the postweld heat treat are specified in AWS A5.17 (Ref 12)
Indicates the lowest temperature at which the impact strength of the weld metal referred to above meets or exceeds 20 ft • lbf (27 J)
E indicates a solid electrode; EC indicates a composite electrode
FXXX-EXXX
Classification of the electrode used in producing the weld referred to above

FIG. 3 CLASSIFICATION SYSTEM FOR CARBON STEEL ELECTRODES AND FLUXES USED IN SAW APPLICATIONS. SOURCE: REF 12

F7A6-EM12K is a complete designation. It refers to a flux that will produce weld metal that, in the as-welded condition, will have a tensile strength no lower than 485 MPa (70 ksi) and Charpy V-notch impact strength of at least 27 J (20 ft · lbf) at -51 °C (-60 °F) when produced with an EM12K electrode under the conditions called for in this specification.

F7A4-EC1 is a complete designation for a flux when the trade name of the electrode used in classification is indicated as well. It refers to a flux that will produce weld metal with that electrode, which in the as-welded condition, will have a tensile strength no lower than 485 MPa (70 ksi) and Charpy V-notch energy of at least 27 J (20 ft · lbf) at -40 °C (-40 °F) under the conditions called for in this specification.

Low-Alloy Steels. The requirements for low-alloy steel welding electrodes are detailed in Ref 13. The classification system for flux-electrode combinations of low-alloy steel is shown in Fig. 4. Table 3 shows the AWS electrode classification as determined by the composition of the solid electrode composition, while Table 4 is the flux-electrode classification requirement for the weld deposit composition of both solid and composite flux-electrode combinations. Interpretation of the classification system code is illustrated in the flux-electrode combination examples that follow.

TABLE 3 COMPOSITION OF LOW-ALLOY STEEL SOLID ELECTRODES

ELECTRODE		COMPOSITION, WT%(A)(B)(C)												
AWS CLASSIFICATION	**UNS NO.**	**C**	**Mn**	**Si**	**S**	**P**	**Cr**	**Ni**	**Mo**	**Cu(d)**	**V**	**Al**	**Ti**	**Zr**
EL12(E)	K01012	0.04-0.14	0.25-0.60	0.10	0.030	0.030	. . .	. . .	. . .	0.35	. . .	. . .	. . .	. . .
EM12K(E)	K01113	0.05-0.15	0.80-1.25	0.10-0.35	0.030	0.030	. . .	. . .	. . .	0.35	. . .	. . .	. . .	. . .
EA1	K11222	0.07-0.17	0.65-1.00	0.20	0.030	0.025	. . .	. . .	0.45-0.65	0.35	. . .	. . .	. . .	. . .
EA2	K11223	0.07-0.17	0.95-1.35	0.20	0.030	0.025	. . .	. . .	0.45-0.65	0.35	. . .	. . .	. . .	. . .
EA3	K11423	0.07-0.17	1.65-2.20	0.20	0.030	0.025	. . .	. . .	0.45-0.65	0.35	. . .	. . .	. . .	. . .
EA3K	K21451	0.07-0.12	1.60-2.10	0.50-0.80	0.025	0.025	. . .	. . .	0.40-0.60	0.35	. . .	. . .	. . .	. . .
EA4	K11424	0.07-0.17	1.20-1.70	0.20	0.030	0.025	. . .	. . .	0.45-0.65	0.35	. . .	. . .	. . .	. . .
EB1	K11043	0.10	0.40-0.80	0.05-0.30	0.025	0.025	0.40-0.75	. . .	0.45-0.65	0.35	. . .	. . .	. . .	. . .
EB2	K11172	0.07-0.15	0.45-1.00	0.05-0.30	0.030	0.025	1.00-1.75	. . .	0.45-0.65	0.35	. . .	. . .	. . .	. . .
EB2H	K23016	0.28-0.33	0.45-0.65	0.55-0.75	0.015	0.015	1.00-1.50	. . .	0.40-0.65	0.30	0.20-0.30	. . .	. . .	. . .
EB3	K31115	0.05-0.15	0.40-0.80	0.05-0.30	0.025	0.025	2.25-3.00	. . .	0.90-1.10	0.35	. . .	. . .	. . .	. . .
EB5	K12187	0.18-0.23	0.40-0.70	0.40-0.60	0.025	0.025	0.45-0.65	. . .	0.90-1.20	0.30	. . .	. . .	. . .	. . .
EB6(F)	S50280	0.10	0.35-0.70	0.05-0.50	0.025	0.025	4.50-6.50	. . .	0.45-0.70	0.35	. . .	. . .	. . .	. . .
EB6H	S50180	0.25-0.40	0.75-1.00	0.25-0.50	0.030	0.025	4.80-6.00	. . .	0.45-0.65	0.35	. . .	. . .	. . .	. . .

ELECTRODE		**COMPOSITION, WT%(A)(B)(C)**												
AWS CLASSIFICATION	**UNS NO.**	**C**	**Mn**	**Si**	**S**	**P**	**Cr**	**Ni**	**Mo**	**Cu(d)**	**V**	**Al**	**Ti**	**Zr**
EB8(F)	S50480	0.10	0.30-0.65	0.05-0.50	0.030	0.040	8.00-10.50	. . .	0.80-1.20	0.35	. . .	. . .	. . .	. . .
ENI1	K11040	0.12	0.75-1.25	0.05-0.30	0.020	0.020	0.15	0.75-1.25	0.30	0.35	. . .	. . .	. . .	. . .
ENI2	K21010	0.12	0.75-1.25	0.05-0.30	0.020	0.020	. . .	2.10-2.90	. . .	0.35	. . .	. . .	. . .	. . .
ENI3	K31310	0.13	0.60-1.20	0.05-0.30	0.020	0.020	0.15	3.10-3.80	. . .	0.35	. . .	. . .	. . .	. . .
ENI4	K11485	0.12-0.19	0.60-1.00	0.10-0.30	0.020	0.015	. . .	1.60-2.10	0.10-0.30	0.35	. . .	. . .	. . .	. . .
ENIIK	K11058	0.12	0.80-1.40	0.40-0.80	0.020	0.020	. . .	0.75-1.25	. . .	0.35	. . .	. . .	. . .	. . .
EF1	K11160	0.07-0.15	0.90-1.70	0.15-0.35	0.025	0.025	. . .	0.95-1.60	0.25-0.55	0.35	. . .	. . .	. . .	. . .
EF2	K21450	0.10-0.18	1.70-2.40	0.20	0.025	0.025	. . .	0.40-0.80	0.40-0.65	0.35	. . .	. . .	. . .	. . .
EF3	K21485	0.10-0.18	1.70-2.40	0.30	0.025	0.025	. . .	0.70-1.10	0.40-0.65	0.35	. . .	. . .	. . .	. . .
EF4	K12048	0.16-0.23	0.60-0.90	0.15-0.35	0.035	0.025	0.40-.60	0.40-0.80	0.15-0.30	0.35	. . .	. . .	. . .	. . .
EF5	K41370	0.10-0.17	1.70-2.20	0.20	0.010	0.010	0.25-0.50	2.30-2.80	0.45-0.65	0.50	. . .	. . .	. . .	. . .
EF6	K21135	0.07-0.15	1.45-1.90	0.10-0.30	0.015	0.015	0.20-0.55	1.75-2.25	0.40-0.65	0.35	. . .	. . .	. . .	. . .
EM2(G)	K10882	0.10	1.25-1.80	0.20-0.60	0.010	0.010	0.30	1.40-2.10	0.25-0.55	0.25	0.05	0.10	0.10	0.10
EM3(G)	K21015	0.10	1.40-1.80	0.20-0.60	0.010	0.010	0.55	1.90-2.60	0.25-0.65	0.25	0.04	0.10	0.10	0.10
EM4(G)	K21030	0.10	1.40-1.80	0.20-0.60	0.010	0.010	0.60	2.00-2.80	0.30-0.65	0.25	0.03	0.10	0.10	0.10
EW	K11245	0.12	0.35-0.65	0.20-0.35	0.040	0.030	0.50-0.80	0.40-0.80	. . .	0.30-0.80	. . .	. . .	. . .	. . .
EG	. . .	NOT SPECIFIED												

Source: Ref 13

(A) THE FILLER METAL SHALL BE ANALYZED FOR THE SPECIFIC ELEMENTS FOR WHICH VALUES ARE SHOWN IN THIS TABLE. IF THE PRESENCE OF OTHER ELEMENTS IS INDICATED, IN THE COURSE OF THIS WORK, THE AMOUNT OF THOSE ELEMENTS SHALL BE DETERMINED TO ENSURE THAT THEIR TOTAL (EXCLUDING IRON) DOES NOT EXCEED 0.50 WT%.

(B) SINGLE VALUES ARE MAXIMUM.

(C) THE LETTER "N" AS A SUFFIX TO A CLASSIFICATION INDICATES THAT THE ELECTRODE IS INTENDED FOR WELDS IN THE CORE BELT REGION OF NUCLEAR REACTOR VESSELS. THIS SUFFIX CHANGES THE LIMITS ON THE PHOSPHORUS, VANADIUM, AND COPPER AS FOLLOWS: P = 0.012% MAXIMUM: V = 0.05% MAXIMUM; CU = 0.08% MAXIMUM. "N" ELECTRODES SHALL NOT BE COATED WITH COPPER OR ANY MATERIAL CONTAINING COPPER. THE "EF5" AND "EW" ELECTRODES SHALL NOT BE DESIGNATED AS "N" ELECTRODES.

(D) THE COPPER LIMIT INCLUDES ANY COPPER COATING THAT MAY BE APPLIED TO THE ELECTRODE.

(E) THE EL12 AND EM12K CLASSIFICATIONS ARE IDENTICAL TO THOSE SAME CLASSIFICATIONS IN ANSI/AWS A5.17-89 (REF. 10). THEY ARE INCLUDED IN THIS SPECIFICATION BECAUSE THEY ARE SOMETIMES USED WITH AN ALLOY FLUX TO DEPOSIT SOME OF THE WELD METALS CLASSIFIED IN TABLE 4.

(F) THE EB6 AND EB8 CLASSIFICATIONS ARE SIMILAR, BUT NOT IDENTICAL, TO THE ER502 AND ER505 CLASSIFICATIONS, RESPECTIVELY, IN ANSI/AWS A5.9-80, "SPECIFICATION FOR CORROSION-RESISTING CHROMIUM AND CHROMIUM-NICKEL STEEL BARE AND COMPOSITE METAL CORED AND STRANDED ARC WELDING ELECTRODES AND WELDING RODS." THESE CLASSIFICATIONS WILL BE DROPPED FROM THE NEXT REVISION OF A5.9 (SEE REF 14).

(G) THE COMPOSITION RANGES OF CLASSIFICATIONS WITH THE "EM" PREFIX ARE INTENDED TO CONFORM TO THE RANGES FOR SIMILAR ELECTRODES IN THE MILITARY SPECIFICATIONS.

TABLE 4 COMPOSITION OF LOW-ALLOY STEEL WELD METAL (BOTH SOLID FLUX-ELECTRODE AND COMPOSITE FLUX-ELECTRODE COMBINATIONS)

AWS WELD METAL CLASSIFICATION(E)	**UNS NO.**		**COMPOSITION, WT%(A)(B)(C)(D)**											
	SOLID	**COMPOSITE**	**C**	**Mn**	**Si**	**S**	**P**	**Cr**	**Ni**	**Mo**	**Cu**	**V**	**Ti**	**Zr**
A1	K11222	W17041	0.12	1.00	0.80	0.040	0.030	. . .	. . .	0.40-0.65	0.35	. . .	. . .	. . .
A2	K11223	W17042	0.12	1.40	0.80	0.040	0.030	. . .	. . .	0.40-0.65	0.35	. . .	. . .	. . .
A3	K11423	W17043	0.15	2.10	0.80	0.040	0.030	. . .	. . .	0.40-0.65	0.35	. . .	. . .	. . .
A4	K11424	W17044	0.15	1.60	0.80	0.040	0.030	. . .	. . .	0.40-0.65	0.35	. . .	. . .	. . .

B1	K11043	W51040	0.12	1.60	0.80	0.040	0.030	0.40-0.65	. . .	0.40-0.65	0.35	. . .	. . .	. . .
B2	K11172	W52040	0.15	1.20	0.80	0.040	0.030	1.00-1.50	. . .	0.40-0.65	0.35	. . .	. . .	. . .
B2H	K23016	W52240	0.10-0.25	1.20	0.80	0.040	0.030	1.00-1.50	. . .	0.40-0.65	0.35	0.30	. . .	. . .
B3	K31115	W53040	0.15	1.20	0.80	0.040	0.030	2.00-2.50	. . .	0.90-1.20	0.35	. . .	. . .	. . .
B4	. . .	W53346	0.12	1.20	0.80	0.040	0.030	1.75-2.25	. . .	0.40-0.65	0.35	. . .	. . .	. . .
B5	K12187	W51348	0.18	1.20	0.80	0.040	0.030	0.40-0.65	. . .	0.90-1.20	0.30	. . .	. . .	. . .
B6	S50280	W50240	0.12	1.20	0.80	0.040	0.030	4.50-6.00	. . .	0.40-0.65	0.35	. . .	. . .	. . .
B6H	S50180	W50140	0.10-0.25	1.20	0.80	0.040	0.030	4.50-6.00	. . .	0.40-0.65	0.35	. . .	. . .	. . .
B8	S50180	W50440	0.12	1.20	0.80	0.040	0.030	8.00-10.00	. . .	0.80-1.20	0.35	. . .	. . .	. . .
NIL(F)	K11040	W21048	0.12	1.60	0.80	0.030	0.030	0.15	0.75-1.10	0.35	0.35	. . .	0.05(G)	. . .
NI2(F)	K21010	W22040	0.12	1.60	0.80	0.030	0.030	. . .	2.00-2.90	. . .	0.35	. . .	. . .	. . .
NI3	K31310	W23040	0.12	1.60	0.80	0.030	0.030	0.15	2.80-3.80	. . .	0.35	. . .	. . .	. . .
NI4	K11485	W21250	0.14	1.60	0.80	0.030	0.030	. . .	1.40-2.10	0.35	0.35	. . .	. . .	. . .
F1	K111	W21150	0.1	0.7	0.8	0.04	0.03	0.15	0.9	0.5	0.3	. . .	. . .	.

	60		2	0-1.50	0	0	0		0-1.70	5	5			..
F2	K21450	W20240	0.17	1.25-2.25	0.80	0.040	0.030	. . .	0.40-0.80	0.40-0.65	0.35	. . .	. . .	...
F3	K21485	W21140	0.17	1.25-2.25	0.80	0.040	0.030	. . .	0.70-1.10	0.40-0.65	0.35	. . .	. . .	...
F4	K12048	W20440	0.17	1.60	0.80	0.040	0.030	0.60	0.40-0.80	0.25	0.35	. . .	0.03(G)	...
F5	K41370	W22640	0.17	1.20-1.80	0.80	0.030	0.030	0.65	2.00-2.80	0.30-0.80	0.50	. . .	. . .	...
F6	K21135	W21040	0.14	0.80-1.85	0.80	0.030	0.030	0.65	1.50-2.25	0.60	0.40	. . .	. . .	...
M1	. . .	W21240	0.10	0.60-1.60	0.80	0.040	0.030	0.15	1.25-2.00	0.35	0.30	. . .	0.03(G)	...
M2	. . .	W21340	0.10	0.90-1.80	0.80	0.040	0.030	0.35	1.40-2.10	0.25-0.65	0.30	. . .	0.03(G)	...
M3	. . .	W22240	0.10	0.90-1.80	0.80	0.030	0.030	0.65	1.80-2.60	0.20-0.70	0.30	. . .	0.03(G)	...
M4	. . .	W22440	0.10	1.30-2.25	0.80	0.030	0.030	0.80	2.00-2.80	0.30-0.80	0.30	. . .	0.03(G)	
W	. . .	W21040	0.12	0.50-1.60	0.80	0.040	0.030	0.45-0.70	0.40-0.80	. . .	0.30-0.75	. . .	. . .	...
G	NOT SPECIFIED													

Source: Ref 13

(A) THE WELD METAL SHALL BE ANALYZED FOR THE SPECIFIC ELEMENTS FOR WHICH VALUES ARE SHOWN IN THIS TABLE. IF THE PRESENCE OF OTHER ELEMENTS IS INDICATED, IN THE COURSE OF THIS WORK, THE AMOUNT OF THOSE ELEMENTS SHALL BE DETERMINED TO ENSURE THAT THEIR TOTAL (EXCLUDING IRON) DOES NOT EXCEED 0.50%.

(B) SINGLE VALUES ARE MAXIMUM.

(C) THE LETTER "N" AS A SUFFIX TO CLASSIFICATION INDICATES THAT THE WELD METAL IS INTENDED FOR WELDS IN THE CORE BELT REGION OF NUCLEAR REACTOR VESSELS. THIS SUFFIX CHANGES THE LIMITS ON PHOSPHORUS, VANADIUM, AND COPPER AS FOLLOWS: P = 0.012% MAXIMUM; V = 0.05% MAXIMUM; CU = 0.08% MAXIMUM. "N" ELECRODES SHALL NOT BE COATED WITH COPPER OR ANY MATERIAL CONTAINING COPPER. THE F5 AND W CLASSIFICATIONS SHALL NOT BE DESIGNATED AS "N" WELD METALS.

(D) A LOW-DILUTION AREA OF THE GROOVE WELD OR THE REDUCED SECTION OF THE FRACTURED TENSION TEST SPECIMEN MAY BE SUBSTITUTED FOR THE WELD PAD, AND SHALL MEET THE ABOVE REQUIREMENTS. IN CASE OF DISPUTE, THE WELD PAD SHALL BE THE REFEREE METHOD.

(E) THE ELECRODE DESIGNATION FOR COMPOSITE ELECTRODES IS OBTAINED BY PLACING AN "EC" BEFORE THE APPROPRIATE WELD METAL CLASSIFICATION.

(F) MANGANESE IN NI1 AND NI2 CLASSIFICATIONS MAY BE 1.80% MAXIMUM WHEN CARBON IS RESTRICTED TO 0.10% MAXIMUM.

(G) A SINGLE VALUE SPANNING THE COLUMNS FOR VANADIUM, TITANIUM, AND ZIRCONIUM IS THE TOTAL OF THESE ELEMENTS.

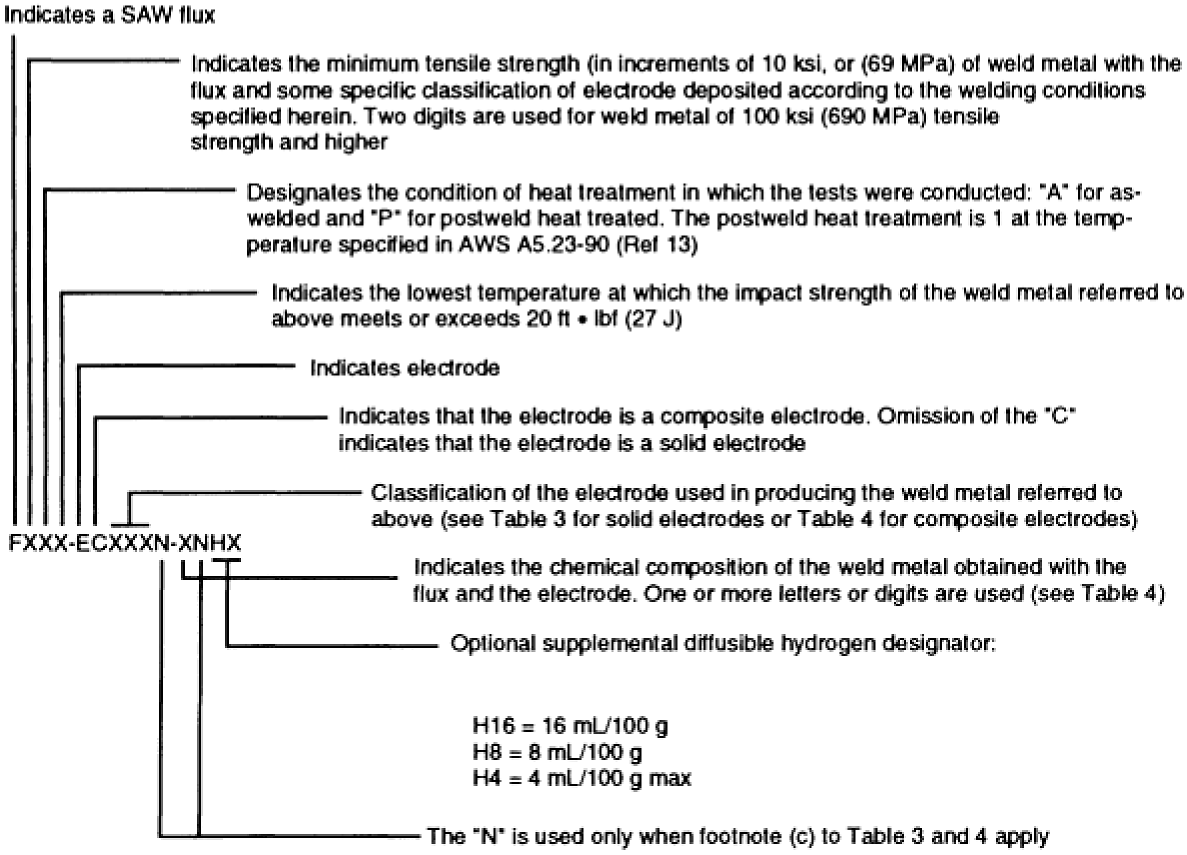
Indicates a SAW flux
Indicates the minimum tensile strength (in increments of 10 ksi, or (69 MPa) of weld metal with the flux and some specific classification of electrode deposited according to the welding conditions specified herein. Two digits are used for weld metal of 100 ksi (690 MPa) tensile strength and higher
Designates the condition of heat treatment in which the tests were conducted: "A" for as-welded and "P" for postweld heat treated. The postweld heat treatment is 1 at the temperature specified in AWS A5.23-90 (Ref 13)
Indicates the lowest temperature at which the impact strength of the weld metal referred to above meets or exceeds 20 ft • lbf (27 J)
Indicates electrode
Indicates that the electrode is a composite electrode. Omission of the "C" indicates that the electrode is a solid electrode
Classification of the electrode used in producing the weld metal referred to above (see Table 3 for solid electrodes or Table 4 for composite electrodes)
FXXX-ECXXXN-XNHX
Indicates the chemical composition of the weld metal obtained with the flux and the electrode. One or more letters or digits are used (see Table 4)
Optional supplemental diffusible hydrogen designator:
H16 = 16 mL/100 g
H8 = 8 mL/100 g
H4 = 4 mL/100 g max
The "N" is used only when footnote (c) to Table 3 and 4 apply

FIG. 4 CLASSIFICATION SYSTEM FOR LOW-ALLOY STEEL ELECTRODES AND FLUXES USED IN SAW APPLICATIONS

F9PO-EB3-B3 is a complete designation for a flux-electrode combination. It refers to a flux that will produce weld metal that, in the postweld heat-treated condition, will have a tensile strength of 620 to 760 MPa (90 to 110 ksi) and Charpy V-notch impact strength of at least 27 J (20 ft · lbf) at - 18 °C (0 °F) when made with an EB3 electrode under the conditions called for in this specification. The composition of the weld metal will be B3 (Table 4).

F9A2-ECM1-M1 is a complete designation for a flux when the trade name of the composite electrode used in the classification tests is indicated as well. The designation refers to a flux that will produce weld metal that, with the M1 electrode, in the as-welded condition, will have a tensile strength of 620 to 760 MPa (90 to 110 ksi) and Charpy V-notch impact strength of at least 27 J (20 ft · lbf) at -29 °C (-20 °F) under the conditions called for in this specification. The composition of the weld metal will be M1 (Table 4).

Stainless and Nickel-Base Alloys. For welding with corrosion-resistant chromium and chromium-nickel bare and composite metal cored and stranded electrodes, the composition of the filler metal is classified according to AWS A5.9 (Table 5). There is no AWS classification of fluxes or of flux-electrode combinations for stainless steel SAW.

TABLE 5 COMPOSITION OF CORROSION-RESISTANT CHROMIUM AND CHROMIUM-NICKEL BARE AND COMPOSITE METAL CORED AND STRANDED ELECTRODES

AWS ELECTRODE CLASSIFICATION	COMPOSITION, WT%(A)(B)										
	C	Cr	Ni	Mo	(nb + ta)	Mn	Si	P	S	N	Cu
ER209(D)	0.05	20.5-24.0	9.5-12.0	1.5-3.0	. . .	4.0-7.0	0.90	0.03	0.03	0.10-0.30	0.75
ER218	0.10	16.0-18.0	8.0-9.0	0.75	. . .	7.0-9.0	3.5-4.5	0.03	0.03	0.08-0.18	0.75
ER219	0.05	19.0-21.5	5.5-7.0	0.75	. . .	8.0-10.0	1.00	0.03	0.03	0.10-0.30	0.75
ER240	0.05	17.0-19.0	4.0-6.0	0.75	. . .	10.5-13.5	1.00	0.03	0.03	0.10-0.20	0.75
ER307	0.04-0.14	19.5-22.0	8.0-10.7	0.5-1.5	. . .	3.3-4.75	0.30-0.65	0.03	0.03	. . .	0.75
ER308(C)	0.08	19.5-22.0	9.0-11.0	0.75	. . .	1.0-2.5	0.30-0.65	0.03	0.03	. . .	0.75
ER308H	0.04-0.08	19.5-22.0	9.0-11.0	0.75	. . .	1.0-2.5	0.30-0.65	0.03	0.03	. . .	0.75
ER308L(C)	0.03	19.5-22.0	9.0-11.0	0.75	. . .	1.0-2.5	0.30-0.65	0.03	0.03	. . .	0.75
ER308MO	0.08	18.0-21.0	9.0-12.0	2.0-3.0	. . .	1.0-2.5	0.30-0.65	0.03	0.03	. . .	0.75
ER308MOL	0.04	18.0-21.0	9.0-12.0	2.0-3.0	. . .	1.0-2.5	0.30-0.65	0.03	0.03	. . .	0.75
ER309(D)	0.12	23.0-25.0	12.0-14.0	0.75	. . .	1.0-2.5	0.30-0.65	0.03	0.03	. . .	0.75
ER309L	0.03	23.0-25.0	12.0-14.0	0.75	. . .	1.0-2.5	0.30-0.65	0.03	0.03	. . .	0.75
ER310	0.08-0.15	25.0-28.0	20.0-22.5	0.75	. . .	1.0-2.5	0.30-0.65	0.03	0.03	. . .	0.75
ER312	0.15	28.0-32.0	8.0-10.5	0.75	. . .	1.0-2.5	0.30-065	0.03	0.03	. . .	0.75
ER316(C)	0.08	18.0-20.0	11.0-14.0	2.0-3.0	. . .	1.0-2.5	0.30-0.65	0.03	0.03	. . .	0.75
ER316H	0.04-0.08	18.0-20.0	11.0-14.0	2.0-3.0	. . .	1.0-2.5	0.30-0.65	0.03	0.03	. . .	0.75
ER316L(D)	0.03	18.0-20.0	11.0-14.0	2.0-3.0	. . .	1.0-2.5	0.30-0.65	0.03	0.03	. . .	0.75
ER317	0.08	18.5-20.5	13.0-15.0	3.0-4.0	. . .	1.0-2.5	0.30-0.65	0.03	0.03	. . .	0.75
ER317L	0.03	18.5-20.5	13.0-15.0	3.0-4.0	. . .	1.0-2.5	0.30-0.65	0.03	0.03	. . .	0.75
ER318	0.08	18.0-	11.0-	2.0-	8 × C	1.0-	0.30-	0.03	0.03	. . .	0.75

		20.0	14.0	3.0	MIN TO 1.0 MAX	2.5	0.65				
ER320	0.07	19.0-21.0	32.0-36.0	2.0-3.0	8 × C MIN TO 1.0 MAX	2.5	0.60	0.03	0.03	. . .	3.0-4.0
ER320LR(D)	0.025	19.0-21.0	32.0-36.0	2.0-3.0	8 × C MIN TO 0.40 MAX	1.5-2.0	0.15	0.015	0.020	. . .	3.0-4.0
ER321(E)	0.08	18.5-20.5	9.0-10.5	0.75	. . .	1.0-2.5	0.30-0.65	0.03	0.03	. . .	0.75
ER330	0.18-0.25	15.0-17.0	34.0-37.0	0.75	. . .	1.0-2.5	0.30-0.65	0.03	0.03	. . .	0.75
ER347(C)	0.08	19.0-21.5	9.0-11.0	0.75	10 × C MIN TO 1.0 MAX	1.0-2.5	0.30-0.65	0.03	0.03	. . .	0.75
ER349(F)	0.07-0.13	19.0-21.5	8.0-9.5	0.35-0.65	1.0-1.4	1.0-2.5	0.30-0.65	0.03	0.03	. . .	0.75
ER410	0.12	11.5-13.5	0.6	0.75	. . .	0.6	0.50	0.03	0.03	. . .	0.75
ER410NIMO	0.06	11.0-12.5	4.0-5.0	0.4.-0.7	. . .	0.6	0.50	0.03	0.03	. . .	0.75
ER420	0.25-0.40	12.0-14.0	0.6	0.75	. . .	0.6	0.50	0.03	0.03	. . .	0.75
ER430	0.10	15.5-17.0	0.6	0.75	. . .	0.6	0.50	0.03	0.03	. . .	0.75
ER502	0.10	4.6-6.0	0.6	0.45-0.65	. . .	0.6	0.50	0.03	0.03	. . .	0.75
ER505	0.10	8.0-10.5	0.5	0.8-1.2	. . .	0.6	0.50	0.04	0.03	. . .	0.75
ER630	0.05	16.0-16.75	4.5-5.0	0.75	0.15-0.30	0.25-0.75	0.75	0.04	0.03	. . .	3.25-4.00
ER26-1	0.01	25.0-27.5	(G)	0.75-1.50	. . .	0.40	0.40	0.02	0.02	0.015	0.20(G)
ER16-8-2	0.10	14.5-16.5	7.5-9.5	1.0-2.0	. . .	1.0-2.5	0.30-0.65	0.03	0.03	. . .	0.75

Source: Ref 14

(A) ANALYSIS SHALL BE MADE FOR THE ELEMENTS FOR WHICH SPECIFIC VALUES ARE SHOWN IN THIS TABLE. IF, HOWEVER, THE PRESENCE OF OTHER ELEMENTS IS INDICATED IN THE COURSE OF ROUTINE ANALYSIS, FURTHER ANALYSIS SHALL BE MADE TO DETERMINE THAT THE TOTAL OF THESE OTHER ELEMENTS, EXCEPT IRON, IS NOT PRESENT IN EXCESS OF 0.50 %.

(B) SINGLE VALUES SHOWN ARE MAXIMUM PERCENTAGES EXCEPT WHERE OTHERWISE SPECIFIED.

(C) THESE GRADES ARE AVAILABLE IN HIGH-SILICON CLASSIFICATIONS THAT SHALL HAVE THE SAME CHEMICAL COMPOSITION REQUIREMENTS AS GIVEN BELOW WITH THE EXCEPTION THAT THE SILICON CONTENT SHALL BE 0.65 TO 1.00%. THESE HIGH-SILICON CLASSIFICATIONS SHALL BE DESIGNATED BY THE ADDITION `SI' TO THE STANDARD

CLASSIFICATION DESIGNATIONS INDICATED BELOW. THE FABRICATOR SHOULD CONSIDER CAREFULLY THE USE OF HIGH-SILICON FILLER METALS IN HIGHLY RESTRAINED FULLY AUSTENITIC WELDS. A DISCUSSION OF THE PROBLEM IS PRESENTED IN REF 14.

(D) CARBON SHALL BE REPORTED TO THE NEAREST 0.01% EXCEPT FOR THE CLASSIFICATION E320LR, FOR WHICH CARBON SHALL BE REPORTED TO THE NEAREST 0.005%.

(E) TITANIUM: 9 × C MIN TO 1.0 MAX.

(F) TITANIUM: 0.10 TO 0.30. TUNGSTEN IS 1.25 TO 1.75%.

(G) NICKEL, MAX: 0.5 MINUS THE COPPER CONTENT.

Like stainless steels, nickel and nickel alloys are classified according to the composition of the filler metal only (Ref 15). Classifications of filler metal are shown in Table 6. There is no AWS classification of fluxes or of flux-electrode combinations for nickel alloy SAW.

TABLE 6 COMPOSITION OF NICKEL AND NICKEL-ALLOY BARE ELECTRODES AND RODS

AWS CLASSIFICATION	**UNS NO.**	**COMPOSITION, WT%(A)(B)**																
		C	**MN**	**FE**	**P**	**S**	**SI**	**CU**	**N(C)**	**CO**	**AL**	**TI**	**CR**	**(NB + TA)**	**MO**	**V**	**W**	**OTHER ELEMENTS, TOTAL**
ERNI-1	N02061	0.15	1.0	1.0	0.03	0.015	0.75	0.25	93.0 MIN	. . .	1.5	2.0-3.5	. . .	. . .	. . .	. . .	. . .	0.50
ERNICU-7	N04060	0.15	4.0	2.5	0.02	0.015	1.25	BAL	62.0-69.0	. . .	1.25	1.5-3.0	. . .	. . .	. . .	. . .	. . .	0.50
ERNICR-3	N06082	0.10	2.5-3.5	3.0	0.03	0.015	0.50	0.50	67.0 MIN	(D)	. . .	0.75	18.0-22.0	2.0-3.0(E)	. . .	. . .	. . .	0.50
ERNLCRFE-5	N06062	0.08	1.0	6.0-10.0	0.03	0.015	0.35	0.50	70.0 MIN	(D)	. . .	. . .	14.0-17.0	1.5-3.0(E)	. . .	. . .	. . .	0.50
ERNLCRFE-6	N07092	0.08	2.0-2.7	8.0	0.03	0.015	0.35	0.50	67.0 MIN	. . .	. . .	2.0-3.5	14.0-17.0	. . .	. . .	. . .	. . .	0.50
ERNIFECR-1	N08065	0.05	1.0	22.0 MIN	0.03	0.03	0.50	1.50-3.0	38.0-46.0	. . .	0.20	0.60-1.2	19.5 23.5	. . .	2.5-3.5	. . .	. . .	0.50
ERNIFECR-2(F)	N07718	0.08	0.35	BAL	0.015	0.015	0.35	0.30	50.0-55.0	. . .	0.20-0.80	0.65-1.15	17.0-21.0	4.75-5.50	2.80-3.30	. . .	. . .	0.50
ERNIMO-1	N10001	0.08	1.0	4.0-7.0	0.025	0.03	1.0	0.50	BAL	2.5	. . .	. . .	1.0	. . .	26.0-30.0	0.20-0.40	1.0	0.50
ERNIMO-2	N10003	0.04-0.08	1.0	5.0	0.015	0.02	1.0	0.50	BAL	0.20	. . .	. . .	6.0-8.0		15.0-18.0	0.50	0.50	0.50
ERNIMO-3	N10004	0.12	1.0	4.0-7.0	0.04	0.03	1.0	0.50	BAL	2.5	. . .	. . .	4.0-6.0	. . .	23.0-26.0	0.60	1.0	0.50
ERNIMO-7	N1066	0.02	1.0	2.0	0.04	0.03	0.1	0.50	BA	1.0	. . .	. . .	1.0	. . .	26.0	. . .	1.0	0.50

	5						0		L						-30.0			
ERNICRMO-1	N06007	0.05	1.0-2.0	18.0-21.0	0.04	0.03	1.0	1.5-2.5	BAL	2.5	. . .	. . .	21.0-23.5	1.75-2.50	5.5-7.5	. . .	1.0	0.50
ERNICRMO-2	N06002	0.05-0.15	1.0	17.0-20.0	0.04	0.03	1.0	0.50	BAL	0.50-2.5	. . .	. . .	20.5-23.0	. . .	8.0-10.0	. . .	0.20-1.0	0.50
ERNICRMO-3	N06625	0.10	0.50	5.0	0.02	0.015	0.50	0.50	58.0 MIN	. . .	0.40	0.40	20.0-23.0	3.15-4.15	8.0-10.0	. . .	. . .	0.50
ERNICRMO-4	N10276	0.02	1.0	4.0-7.0	0.04	0.03	0.08	0.50	BAL	2.5	. . .	. . .	14.5-16.5	. . .	15.0-17.0	0.35	3.0-4.5	0.50
ERNICRMO-7	N06455	0.015	1.0	3.0	0.04	0.03	0.08	0.50	BAL	2.0	. . .	0.70	14.0-18.0	. . .	14.0-18.0	. . .	0.50	0.50
ERNICRMO-8	N06975	0.03	1.0	BAL	0.03	0.03	1.0	0.7-1.20	47.0-52.0	. . .	. . .	0.70-1.50	23.0-26.0	. . .	5.0-7.0	. . .	. . .	0.50
ERNICRMO-9	N06985	0.015	1.0	18.0-21.0	0.04	0.03	1.0	1.5-2.5	BAL	5.0	. . .	. . .	21.0-23.5	0.50	6.0-8.0	. . .	1.5	0.50
ERNICRMO-10	N06022	0.015	0.50	2.0-6.0	0.02	0.010	0.08	0.50	BAL	2.5	. . .	. . .	20.0-22.5	. . .	12.5-14.5	0.35	2.5-4.5	0.50
ERNICRMO-11	N06030	0.03	1.5	13.0-17.0	0.04	0.02	0.80	1.0-2.4	BAL	5.0	. . .	. . .	28.0-31.5	0.30-1.50	4.0-6.0	. . .	1.5-4.0	0.50
ERNICRCOMO-1	N06617	0.05-0.15	1.0	3.0	0.03	0.015	1.0	0.50	BAL	10.0-15.0	0.80-1.50	0.60	20.0-24.0	. . .	8.0-10.0	. . .	. . .	0.50

Source: Ref 15

(A) THE FILLER METAL SHALL BE ANALYZED FOR THE SPECIFIC ELEMENTS FOR WHICH VALUES AVE SHOWN IN THIS TABLE. IN THE COURSE OF THIS WORK, IF THE PRESENCE OF OTHER ELEMENTS IS INDICATED, THE AMOUNT OF THOSE ELEMENTS SHALL HE DETERMINED TO ENSURE THAT THEIR TOTAL DOES NOT EXCEED THE LIMIT SPECIFIED FOR "OTHER ELEMENTS, TOTAL" IN

THE LAST COLUMN OF THE TABLE.

(B) SINGLE VALUES ARE MAXIMUM, EXCEPT WHERE OTHERWISE SPECIFIED.

(C) INCLUDES INCIDENTAL COBALT.

(D) COBALT: 0.12 MAXIMUM, WHEN SPECIFIED.

(E) TANTALUM: 0.30 MAXIMUM, WHEN SPECIFIED.

(F) BORON IS 0.006% MAXIMUM.

Submerged Arc Welding

Jonathan S. Ogborn, The Lincoln Electric Company

Personnel Considerations

Employee Training. Submerged arc welding is an automatic or semiautomatic process carried out primarily in the flat or horizontal position with the arc obscured from view and does not require a high level of skill. The primary expenditures both in cost and time are consumed in designing jigs and fixtures to hold and position the workpiece so that SAW can be performed. While the actual welding process does not require a high degree of skill, the setup procedure requires skilled personnel with an aptitude for fabricating the fixtures and clamps necessary to position the workpiece in an acceptable position (usually flat).

Health and Safety in the Workplace. The main safety concerns in welding are electric shock, fume inhalation, and burns from both the arc ultraviolet source and the hot-metal infrared source (Ref 16). To reduce the risk of electrical shock, all equipment must be properly grounded and welding cables must be in good condition. The fume is nearly eliminated in SAW. However, certain vaporized elements (for example, chromium, cobalt, manganese, nickel, and vanadium) can be potentially dangerous. To prevent exposure to these elements, the welding area should have adequate ventilation. In SAW, the arcflash and its accompanying spatter are nearly eliminated because of the slag and the depth of the flux pile. However, eye protection should always be worn.

Additional information is available in the article "Safe Practices" in this Volume.

Reference cited in this section

16. *SAFETY IN WELDING AND CUTTING*, AWS, 1988

Submerged Arc Welding

Jonathan S. Ogborn, The Lincoln Electric Company

SAW Parameters

While SAW is the most inexpensive and efficient process for making large, long, and repetitive welds, much time and energy are required to prepare the joint. Care must be taken to line up all joints to have a consistent gap in groove welds and to provide backing plates and flux dams to prevent spillage of flux and molten metal. Once all the pieces are clamped or tacked in place, welding procedures and specifications should be consulted before welding begins.

Procedural Variations and Effect on Weld Bead Characteristics

Procedural variations in SAW include current, voltage, electrical stickout (distance from last electrical contact to plate), travel speed, and flux depth. Variation in any of these parameters will affect the shape and penetration of the weld, as well as the integrity of the weld deposit.

Weld Current. Because the welding current controls such parameters as deposition rate, penetration, and dilution, it is the most important welding variable. An increase in welding current at a constant voltage will decrease the flux-to-wire ratio, while a decrease in current will increase the flux-to-wire ratio. The effect of current variation on weld bead profile is shown in Fig. 5. Welds made at excessively low current will tend to have little penetration and higher width-to-depth ratios. Welds made at an excessively high current will have deep penetration, high dilution, more shrinkage, and excess buildup. Low current will also produce a less stable arc than higher currents.

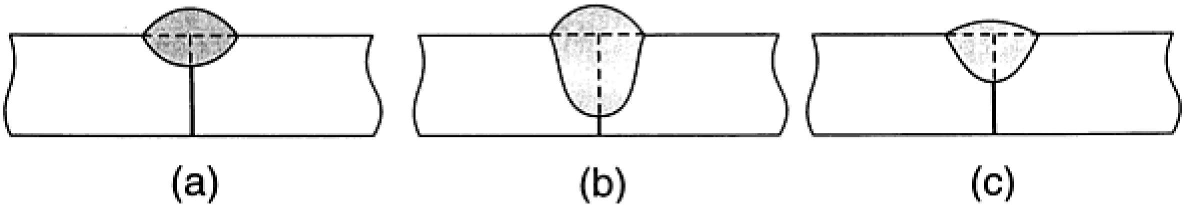

FIG. 5 EFFECT OF VARIATION IN WELDING CURRENT ON WELD BEAD PROFILE. (A) EXCESSIVELY LOW CURRENT. (B) EXCESSIVELY HIGH CURRENT. (C) RECOMMENDED CURRENT

The direction of current flow will also affect the weld bead profile. The current may be direct with the electrode positive (reverse polarity), electrode negative (straight polarity), or alternating. Reverse polarity is most commonly used. For a given set of welding conditions, reverse polarity will produce wider beads with more penetration at a lower deposition rate than straight polarity. Straight-polarity welding will contribute to narrower beads with less penetration and more buildup. Because straight polarity reduces base plate dilution, it is frequently used in surfacing applications. For the same welding current, the deposition rate with straight polarity is higher than with reverse polarity. Straight polarity is preferred for poor fitup. The bead shape, penetration, and deposition rate for alternating current fall between those of straight and reverse polarity. Alternating Current is used when welding current exceeds 1000 A and on multiple-wire applications to reduce arc blow and arc interaction.

In SAW, the current density in the electrode also plays a role in bead shape and penetration. Smaller-diameter electrodes with a high current density will produce narrower beads with deeper penetration than larger-diameter electrodes. Larger-diameter electrodes are able to bridge larger root openings. In cases where a given current can be achieved with two different electrode diameters, the smaller electrode will produce the higher deposition rate.

Weld Voltage. Like current, welding voltage will affect the bead shape and the weld deposit composition. Increasing the arc voltage at a constant current will increase the flux-to-wire electrode ratio, while decreasing the voltage will reduce the flux-to-electrode ratio. The effect of the magnitude of arc voltage on bead shape is shown in Fig. 6. Increasing the arc voltage will produce a longer arc length and a correspondingly wider, flatter bead with less penetration. Higher voltage will increase flux consumption, which could then change deposit composition and properties. Slightly increasing the arc voltage will help the weld to bridge gaps when welding in grooves. Excessively high voltage will produce a hat-shaped concave weld, which has low resistance to cracking and a tendency to undercut. Lower voltages will shorten the arc length and increase penetration. Excessively low voltage will produce an unstable arc and a crowned bead, which has an uneven contour where it meets the plate.

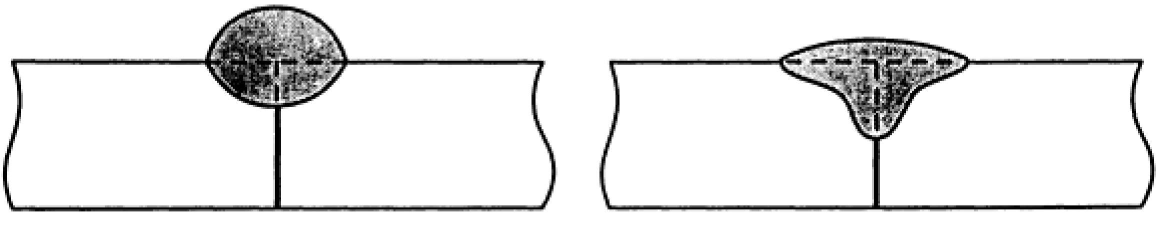

FIG. 6 EFFECT OF VARIATION IN WELDING VOLTAGE AT CONSTANT CURRENT ON WELD BEAD PROFILE. (A) EXCESSIVELY LOW VOLTAGE. (B) EXCESSIVELY HIGH VOLTAGE

Electrical Stickout. In SAW, the current flowing in the electrode between the contact tube and the arc (electrode extension) will cause some I^2R heating, resulting in a voltage drop across that length of electrode. This resistance heating and subsequent voltage drop can be used to obtain higher deposition rates. Normal electrode extension for solid SAW wire is approximately 8 to 12 times the electrode diameter. As this length increases at a constant current, so does the resistance heating and the melt-off rate. To compensate for the voltage drop and the increase in wire-feed speed, the voltage must be increased to obtain a properly shaped bead. Extending the electrode 20 to 40 times the diameter can increase deposition rates by more than 50% (Ref 17). Although higher deposition rates can be achieved by extending the electrode, keeping the electrode aligned with the joint becomes increasingly difficult as the extension increases.

Travel Speed. Variations in travel speed at a set current and voltage also affect bead shape. As welding speed is decreased, heat input per length of joint increases, and the penetration and bead width increase. The penetration will increase until molten metal begins to flow under the arc and interfere with heat flow at excessively slow speeds. Excessively high travel speeds will promote a crowned bead as well as the tendency for undercut and porosity.

Flux layer depth is another variable that will alter the appearance, penetration, and quality of a submerged arc weld. If the flux layer is too deep, a greater-than-normal amount of flux will be melted, resulting in weld beads that are narrower than normal. Some surface imperfections may also appear because gases may be trapped by the deep flux layer. If the flux layer is too shallow, the arc will flash through, and the bead will have a rough appearance or porosity due to lack of shielding from the atmosphere. The correct depth of flux is just enough to prevent flash through. This will allow welding gases to escape while providing adequate protection.

Sources of Defects in SAW

The fact that SAW is a high heat input process under a protective blanket of flux greatly decreases the chance of weld defects. However, defects such as lack of fusion, slag entrapment, solidification cracking, hydrogen cracking, or porosity occasionally occur.

Insufficient Fusion and Slag Entrapment. Lack-of-fusion defects and slag entrapment are most commonly caused by improper bead placement or procedure. Improper placement can cause the weld metal to roll over and trap slag underneath, or if the weld bead is placed away from the edge to be joined, the liquid metal may not fuse to the base material. A crown-shaped bead caused by low welding voltage may also contribute to slag entrapment and lack of fusion by not allowing the liquid metal to spread out evenly.

Solidification cracking of SAW along the center of the bead is usually due to bead shape, joint design, or incorrect choice of welding consumables. A convex bead shape with a bead width-to-depth ratio greater than one will decrease solidification cracking tendencies. If weld penetration is too deep, the shrinkage stresses may cause centerline cracking. Joint design may also contribute to excessive shrinkage stresses, again increasing the risk of solidification cracking. Because cracking is related to stresses in the weld, high-strength materials will have a greater tendency to crack. Therefore, special care must be taken to generate proper bead shape, preheat temperatures, and interpass temperatures, in addition to correct electrode and flux combinations, when welding these materials.

Hydrogen Cracking. Unlike solidification cracking, which appears immediately after welding, hydrogen cracking is a delayed process and may occur from several hours to several days after welding has been completed. To minimize hydrogen cracking, all possible sources of hydrogen (for example, water, oil, grease, and dirt) present in the flux, electrode, or joint should be eliminated. The flux, electrode, and plate should be clean and dry.

To prevent moisture pickup, fluxes and electrodes should be stored in moisture-resistant containers in dry areas. If a flux or electrode becomes contaminated with moisture, it should be dried according to manufacturer recommendations.

Selection of the consumable, especially when welding high-strength steels that are more susceptible to hydrogen cracking, can also play a role in hydrogen cracking. Special consumables have recently been developed that produce weld deposits that are very low in diffusible hydrogen.

To further reduce hydrogen-related cracking, the joint to be welded should be preheated. Because hydrogen is fairly mobile in steel at temperatures above 95 °C (200 °F), the recommended preheat temperatures should be followed to allow most of the hydrogen to escape and to reduce the risk of hydrogen damage (Ref 18). In thick weldments, maintaining preheat for several hours after welding is completed will also reduce the risk of hydrogen cracking.

Porosity caused by trapped gas is uncommon in SAW because of the protection provided by the flux. When porosity does occur, it may be in the form of internal porosity or as depressions on the weld bead surface. The gas bubbles that cause porosity originate either from a lack of protection from the atmosphere or from contaminants such as water, oil, grease, and dirt. To reduce porosity in SAW, the weld should have sufficient flux coverage, and all water, grease, and dirt should be removed from the plate, electrode, and flux. Another cause of porosity in SAW is excessive travel speed. Travel at excessively high speeds will not allow the gas bubbles to escape from the weld, and the bubbles may become trapped in the weld metal at the slag-to-metal interface.

References cited in this section

17. *HOW TO MAKE SINGLE ELECTRODE SUBMERGED ARC WELDS,* PUBLICATION S604, THE LINCOLN ELECTRIC CO., 1991, P 11
18. F.R. COE, *WELD. WORLD,* VOL 14 (NO. 1/2), 1976, P 1-7

Submerged Arc Welding

Jonathan S. Ogborn, The Lincoln Electric Company

References

1. R.L. O'BRIEN, *WELDING HANDBOOK,* VOL 11, AWS, 1991, P 191-232
2. D.L. OLSON *ET AL.,* SUBMERGED ARC WELDING, VOL 6, 9TH ED., *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1983, P 114-152
3. *THE PROCEDURE HANDBOOK OF ARC WELDING,* THE LINCOLN ELECTRIC CO., 1973, P 6.3-1 TO 6.3-24
4. "SPECIFICATION FOR CARBON STEEL ELECTRODES AND FLUXES FOR SUBMERGED ARC WELDING," A5.17-89, AWS, 1989, P 18
5. *THE PROCEDURE HANDBOOK OF ARC WELDING,* THE LINCOLN ELECTRIC CO., 1973, P 3.2-3, 6.3-21 TO 6.3-22, 6.3-49 TO 6.3-58
6. *THE PROCEDURE HANDBOOK OF ARC WELDING,* THE LINCOLN ELECTRIC CO., 1973, P 6.3-26 TO 6.3-73
7. R.S. BROWN *ET AL.,* ARC WELDING OF STAINLESS STEELS, VOL 6, 9TH ED., *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1983, P 327-329
8. K.C. ANTONY *ET AL.,* HARDFACING, VOL 6, 9TH ED., *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1983, P 784
9. D.R. THOMAS *ET AL.,* WELD OVERLAYS, VOL 6, 9TH ED., *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1983, P 804-819
10. "SPECIFICATION FOR CARBON STEEL ELECTRODES AND FLUXES FOR SUBMERGED ARC WELDING," A5.17-89, AWS, 1989, P 17-18
11. S.S. TULIANI, T. BONISZEWSKI, AND N.F. EATON, NOTCH TOUGHNESS OF COMMERCIAL SUBMERGED ARC WELD METAL, *WELD. MET. FABR.,* VOL 8, 1969, P 327-339
12. "SPECIFICATION FOR CARBON STEEL ELECTRODES AND FLUXES FOR SUBMERGED ARC WELDING," A5.17-89, AWS, 1989, P 2-3
13. "SPECIFICATION FOR LOW ALLOY STEEL ELECTRODES AND FLUXES FOR SUBMERGED ARC WELDING," A5.23-90, AWS, 1990, P 2-6
14. "SPECIFICATION FOR CORROSION RESISTING CHROMIUM AND CHROMIUM-NICKEL STEEL BARE AND COMPOSITE METAL CORED AND STRANDED WELDING ELECTRODES AND WELDING RODS," A5.9-81, AWS, 1981, P 3, 4
15. "SPECIFICATION FOR NICKEL AND NICKEL ALLOY BARE WELDING ELECTRODES AND RODS," A5.14-89, AWS, 1989, P 2, 3
16. *SAFETY IN WELDING AND CUTTING,* AWS, 1988
17. *HOW TO MAKE SINGLE ELECTRODE SUBMERGED ARC WELDS,* PUBLICATION S604, THE LINCOLN ELECTRIC CO., 1991, P 11
18. F.R. COE, *WELD. WORLD,* VOL 14 (NO. 1/2), 1976, P 1-7

Stud Arc Welding

Harry A. Chambers, TRW Nelson Stud Welding Division

Introduction

STUD ARC WELDING (SW), also known as arc stud welding, is a commonly used method for joining a metal stud, or fastener, to a metal workpiece. The process has been used as an alternative metal-fastening method since the 1940s. Millions of specially designed and manufactured metal studs are welded by this process every week in such diverse industries as construction, shipbuilding, automotive, and hard goods, as well as in miscellaneous industrial applications.

This article will serve as a basic information source for those interested in accomplishing one-sided, no-hole attachment of metal fasteners. The SW process represents an alternative to other welding processes, and is also a substitute for other fastening procedures, such as drilling and tapping, bolting, and self-tapping screws.

Stud Arc Welding

Harry A. Chambers, TRW Nelson Stud Welding Division

Process Overview

Stud arc welding is similar to many other welding processes, including arc and percussion welding, in that the base (weld end) of a specifically designed stud is joined to a base material by heating both parts with an arc that is drawn between the two. Equipment that is unique to this process regulates the arc length and arc dwell time. After an arc is struck, the stud weld end and the workpiece surface are brought to the proper temperature for joining and, after a controlled period of time, the two heated surfaces are brought together under pressure, creating a metallurgical bond capable of developing the full strength of the stud.

There are two basic types of stud arc welding, which are differentiated by the source of welding power. One type uses direct current (dc) power provided by a transformer/rectifier or a motor generator similar to that used in the shielded metal arc welding (SMAW) process. The second type uses power discharged from a capacitor storage bank. The process based on a dc power source is known as stud arc welding, whereas the process that utilizes capacitors is known as capacitor discharge stud welding (CDSW).

Both the SW and CDSW processes overlap in some areas of application. Generally, the SW process is used in applications that require similar stud and workpiece metals, the workpiece thickness is greater in relation to the stud diameter, and an accommodation must be made for the stud flash (fillet). The term "flash" is preferred to the term "fillet," because the metal that forms the flash during the stud arc welding process is expelled, rather than added, as occurs with other welding processes.

In contrast, the CDSW process is used extensively when welding to thin sheet metal, and is used frequently with dissimilar workpiece and stud alloys. It is also used in cases where marks on the opposite side of the workpiece must be avoided or minimized. With this process, the stud diameter is limited to smaller sizes. The CDSW process is more fully described in the article "Capacitor Discharge Stud Welding" in this Volume. The factors on which process selection should be based are fastener size, base-metal thickness, base-metal composition, and reverse-side marking requirements (Table 1).

TABLE 1 STUD-WELDING PROCESS SELECTION

PARAMETERS	STUD ARC WELDING	CAPACITOR DISCHARGE STUD WELDING	
		GAP AND CONTACT METHODS	DRAWN ARC METHOD
STUD SHAPE			
ROUND	A	A	A
SQUARE	A	A	A

RECTANGULAR	A	A	A
IRREGULAR	A	A	A
STUD DIAMETER OR AREA			
1.6 TO 3.2 MM ($\frac{1}{16}$ TO $\frac{1}{8}$ IN.)	D	A	A
3.2 TO 6.4 MM ($\frac{1}{8}$ TO $\frac{1}{4}$ IN.)	C	A	A
6.4 TO 12.7 MM ($\frac{1}{4}$ TO $\frac{1}{2}$ IN.)	A	B	B
12.7 TO 25.4 MM ($\frac{1}{2}$ TO 1 IN.)	A	D	D
UP TO 32.3 MM^2 (0.05 $IN.^2$)	C	A	A
OVER 32.3 MM^2 (0.05 $IN.^2$)	A	D	D
STUD METAL			
CARBON STEEL	A	A	A
STAINLESS STEEL	A	A	A
ALLOY STEEL	B	C	C
ALUMINUM	B	A	A
BRASS	C	A	D
BASE METAL			
CARBON STEEL	A	A	A
STAINLESS STEEL	A	A	A
ALLOY STEEL	B	A	C
ALUMINUM	B	A	A
BRASS	C	A	D
BASE-METAL THICKNESS			
UNDER 0.4 MM (0.015 IN.)	D	A	B
0.4 TO 1.6 MM (0.015 TO 0.062 IN.)	C	A	A
1.6 TO 3.2 MM (0.062 TO 0.125 IN.)	B	A	A
OVER 3.2 MM (0.125 IN.)	A	A	A
STRENGTH CRITERIA			
HEAT EFFECT ON EXPOSED SURFACES	B	A	A
WELD FILLET CLEARANCE	B	A	A
STRENGTH OF STUD GOVERNS	A	A	A
STRENGTH OF BASE METAL GOVERNS	A	A	A

A. applicable without special procedures or equipment; B. applicable with special techniques or for specific applications that justify preliminary trials or testing to develop welding procedure and technique; C, limited application; D, not recommended-welding methods not developed at this time.

Source: Ref 1

Equipment. As shown in Fig. 1, the basic equipment used for stud arc welding consists of a control system, which regulates the arc time and controls gun movement; a fixed or portable stud-welding gun, which holds the stud in position during the welding process to create the proper arc length and joining pressure; and connecting cables, which must be connected to a separate source of dc power. The other items that are needed to weld the workpiece are the studs themselves and ceramic arc shields, or ferrules. The equipment used for stud welding is comparable, in terms of size, portability, and ease of operation, to the equipment used in the SMAW process. Although the initial cost of stud-welding systems varies with the method selected, stud size, and productivity requirements, it is generally competitive with other fastening methods on an in-place or finished part cost basis.

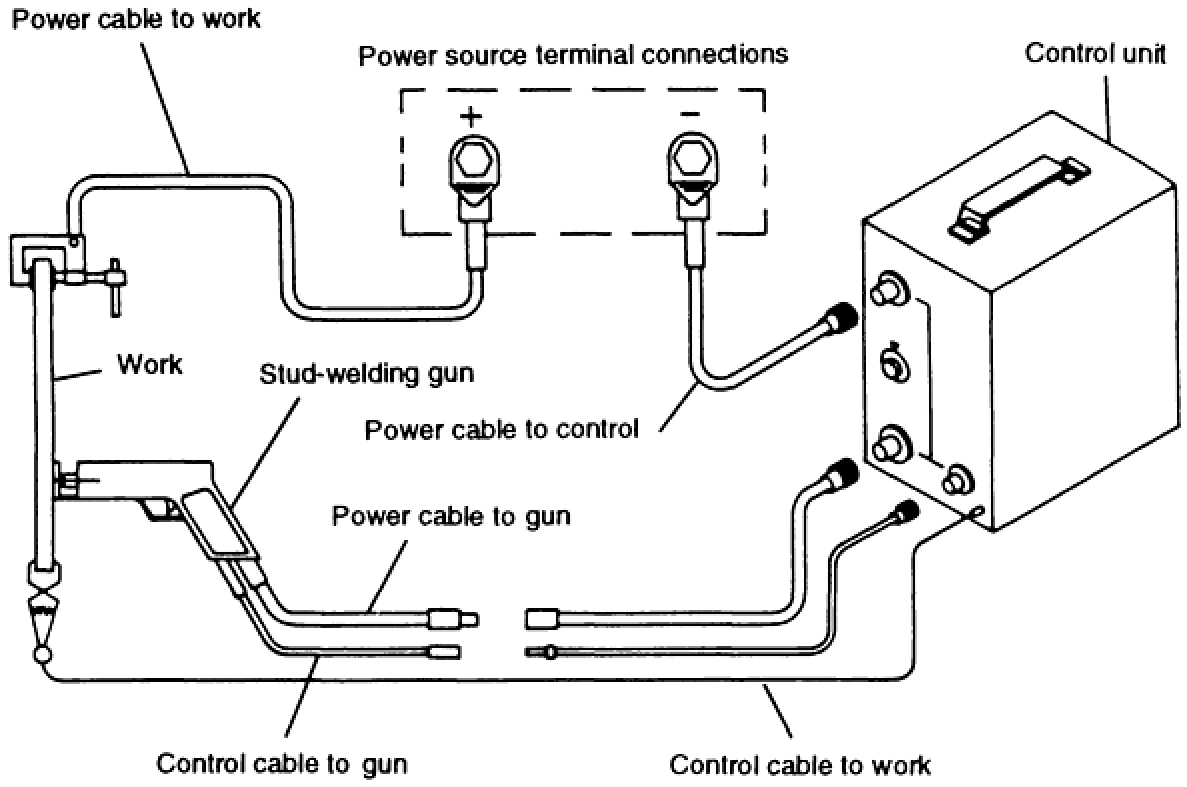
Power cable to work
Power source terminal connections
+
−
Control unit
Work
Stud-welding gun
Power cable to control
Power cable to gun
Control cable to gun
Control cable to work

FIG. 1 STUD ARC WELDING CONTROL SYSTEM. THE CONTROL SYSTEM MUST BE CONNECTED TO A DC POWER SOURCE FOR WELDING. IN MOST APPLICATIONS, THE STUD (ELECTRODE) SHOULD BE NEGATIVE.

A stud-welding control system that has been integrated in a transformer/rectifier power source is shown in Fig. 2. This type of equipment, which is the most widely used, is called a power/control system. It can weld studs with diameters up to 28.5 mm ($1\frac{1}{8}$ in.). Either a single gun or dual guns can be used. Although a light-duty control system can weigh 11.3 kg (25 lb), the system shown in Fig. 2 weighs approximately 450 kg (1000 lb) and can be put on a wheeled cart for mobility.

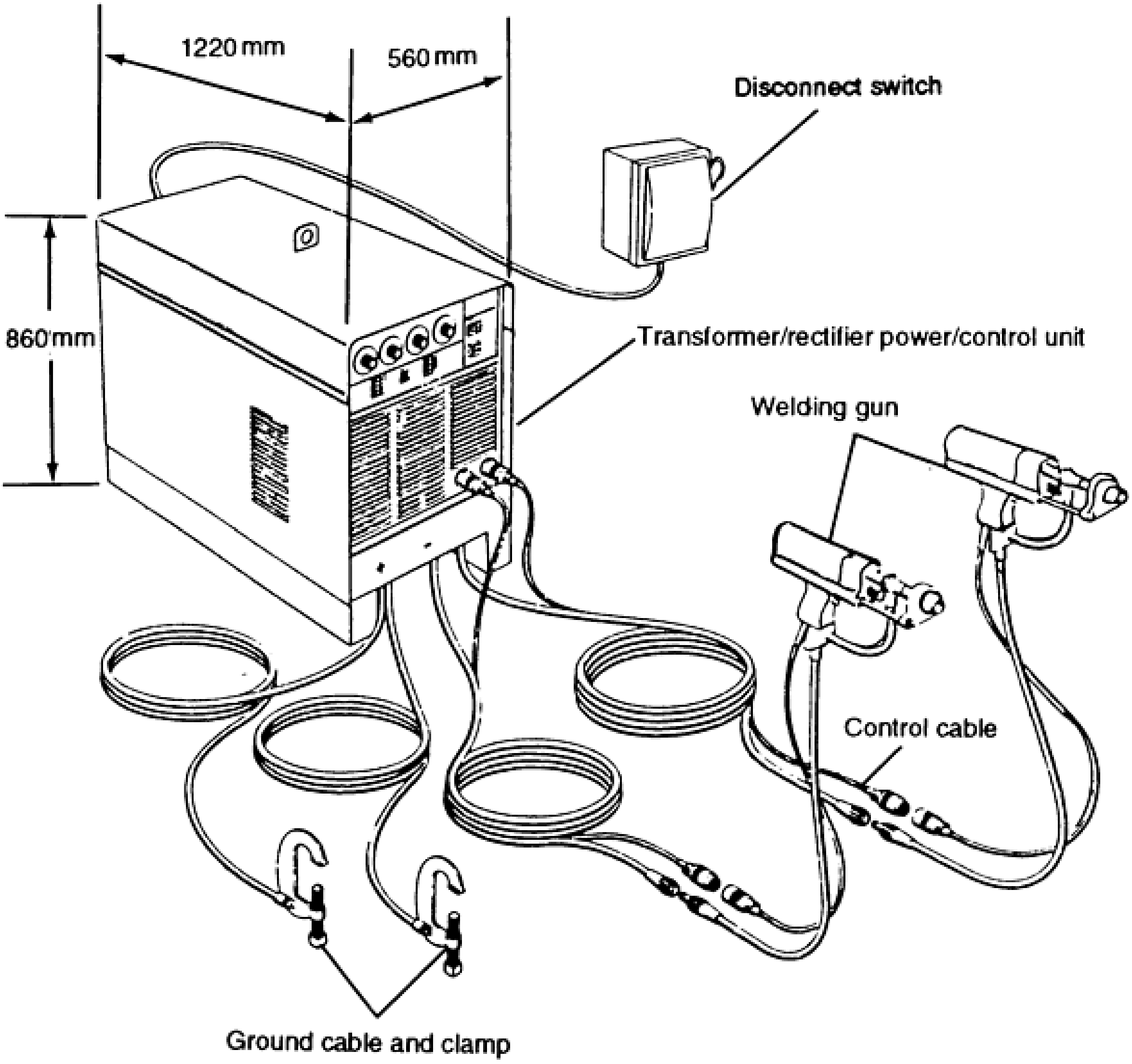
1220 mm
560 mm
Disconnect switch
860 mm
Transformer/rectifier power/control unit
Welding gun
Control cable
Ground cable and clamp

FIG. 2 TYPICAL INTEGRATED POWER/CONTROL SYSTEM FOR STUD ARC WELDING

A typical stud-welding gun (Fig. 3) is usually available in at least two sizes: standard duty, for use with studs up to 16 mm ($\frac{5}{8}$ in.) in diameter, and heavy duty, for use on larger-diameter studs. The weight of the gun can vary from approximately 1.8 kg (4 lb), for a standard-duty gun, to 6.8 kg (15 lb), for heavy-duty models.

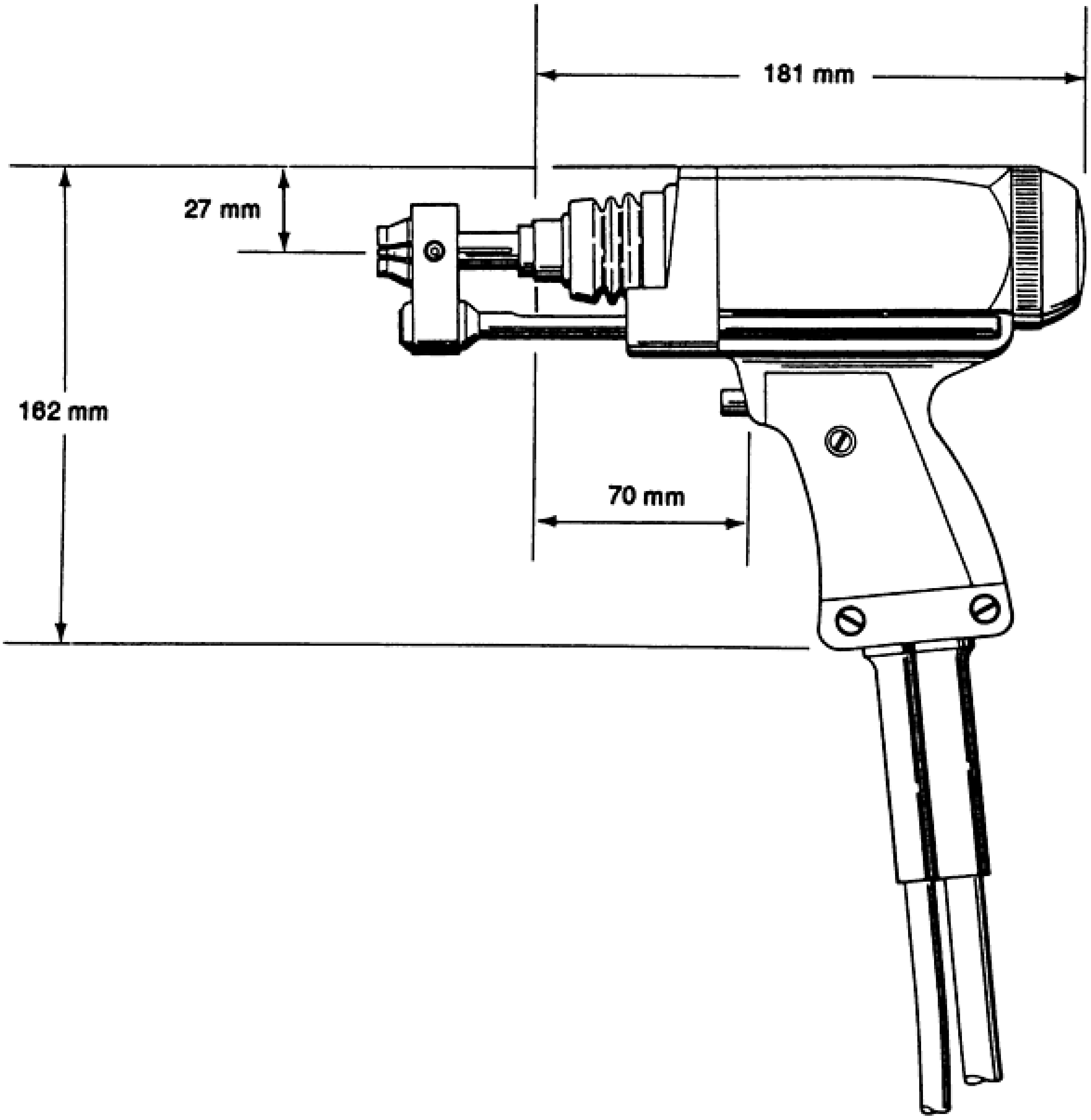

FIG. 3 TYPICAL STANDARD-DUTY GUN FOR STUD ARC WELDING

Operation. The stud arc welding process utilizes the same principles as any other arc welding procedure. First, the stud, which acts as an electrode, is inserted into a chuck on the end of the gun, surrounded by a ceramic ferrule, and positioned against the workpiece (Fig. 4a). Next, the gun trigger is depressed, which starts the automatic weld cycle by energizing a solenoid coil within the gun body that lifts the stud off the work and draws an arc. The arc melts the end of the stud and a portion of the workpiece (Fig. 4b). After a preset arc time (set on the control unit), the welding current is shut off and the solenoid is de-energized (Fig. 4c). A mainspring in the gun forces the stud into the molten pool of metal, producing a full-strength weld, which is shown in Fig. 4(d) after the gun has been lifted off the stud and the ceramic ferrule removed. The result is a full-penetration, full-strength stud-to-workpiece weld, as shown in Fig. 5. This weld is similar to that obtained with other types of arc-welding processes.

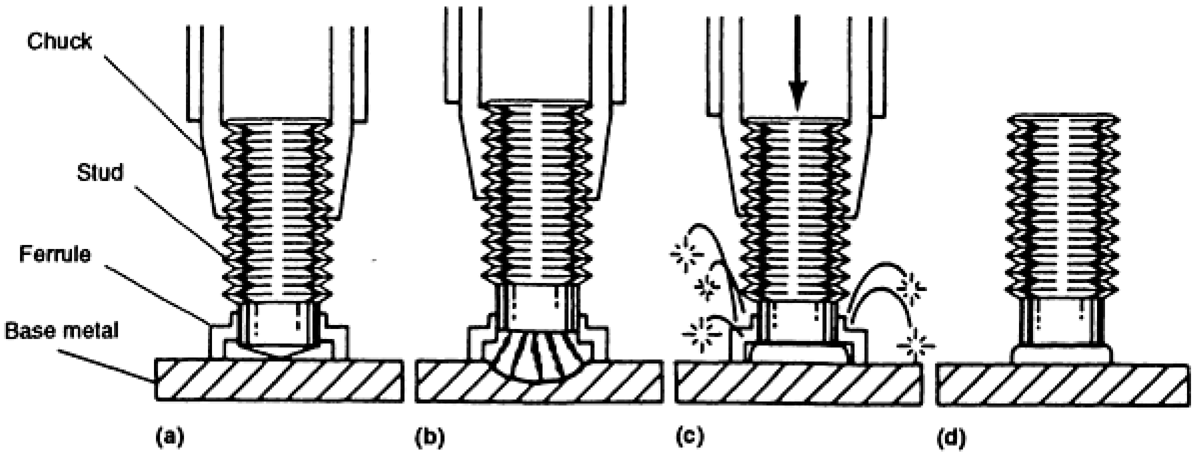

FIG. 4 STUD ARC WELDING PROCESS. (A) GUN IS PROPERLY POSITIONED. (B) TRIGGER IS DEPRESSED AND STUD IS LIFTED, CREATING AN ARC. (C) ARCING PERIOD IS COMPLETED AND STUD IS PLUNGED INTO MOLTEN POOL OF METAL ON BASE MATERIAL. (D) GUN IS WITHDRAWN FROM WELDED STUD AND FERRULE IS REMOVED.

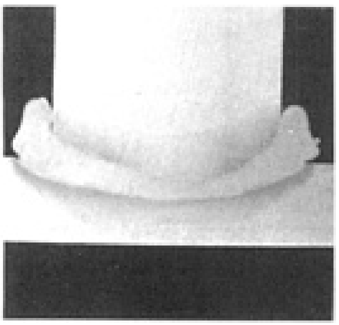

FIG. 5 MACROSECTION OF LOW-CARBON STEEL STUD WELD

Studs and Ferrules. With the exception of special processes, the stud arc welding of steel, stainless steel, and aluminum studs requires the use of a specifically designed weld stud and ceramic ferrule.

Studs. A wide range of studs are made by stud-welding manufacturers for the SW process. They vary in size and weld-base configuration, depending on the application requirements. The typical stud styles that are available include threaded, unthreaded, headed, and rectangular studs (Fig. 6). The steel stud diameters that are commonly welded using the SW process range from 2.7 to 25.4 mm (0.105 to 1.00 in.).

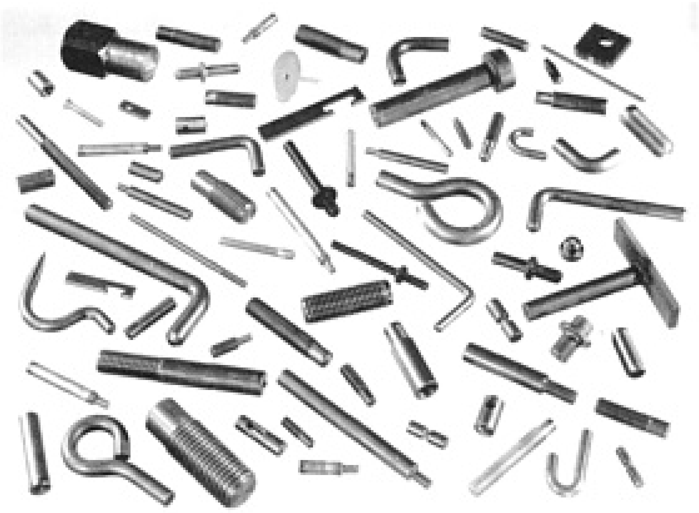

FIG. 6 COMMON STUD CONFIGURATIONS FOR STUD ARC WELDING

Full-strength welds result when using studs made from low-carbon steel that conforms to ASTM A-108 grades C-1010 through C-1020, although other steel grades can be used on a special-application qualified basis (Ref 2). Stainless steel studs are typically made from the austenitic series, including AISI 302, 302 HQ, 304, 305, 308, 309, 310, 316, 321, and 347, in both normal-carbon and low-carbon varieties. Stainless steel 303 is not an acceptable stud alloy. Aluminum studs are usually made from aluminum-magnesium alloys, including 5183, 5356, 5556, 5086, and 5456 (Ref 3).

Steel and stainless steel arc welded studs that are larger than 6.4 mm (0.250 in.) in diameter use a welding flux to stabilize the welding arc and to deoxidize the weld area. The most commonly used flux is commercially pure aluminum, which is installed onto the weld end of the stud by thermal spraying, staking, or pressing a slug into a drilled hole in the end of the stud (Fig. 7). Studs that are 6.4 mm (0.250 in.) and under usually do not require fluxing.

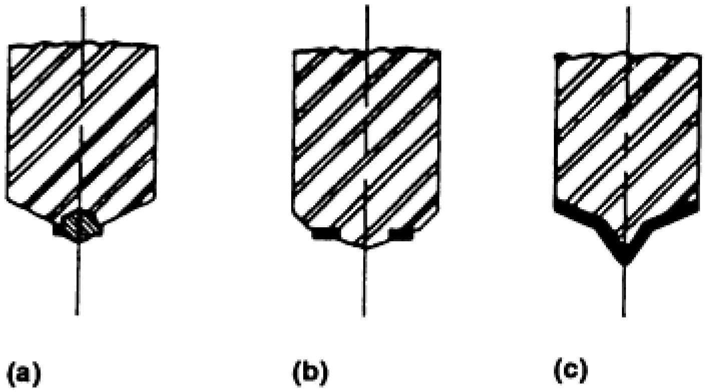

FIG. 7 METHODS OF FLUX LOADING STUDS. (A) PRESSED IN SLUG. (B) STAKED ON WASHER. (C) COATING

Aluminum studs do not have to be flux loaded. Instead, the weld end of the stud has a conical or cylindrical tip that is shaped into the parent metal to help initiate the arc and control the arc length. To prevent the oxidation and embrittlement of the stud/plate weld zone during aluminum stud arc welding, the weld area is purged with an inert gas during the weld cycle. Argon gas is commonly used for this purpose, although helium also can be used. The typical gas flow rates for aluminum studs of various diameters are shown in Table 2.

TABLE 2 GAS FLOW RATES FOR ALUMINUM STUD ARC WELDING

STUD WELD BASE DIAMETER		**SHIELDING GAS FLOW(A)**	
mm	**in.**	**l/min**	**ft^3/h**
6.4	$\frac{1}{4}$	7.1	15
7.9	$\frac{5}{16}$	7.1	15
9.5	$\frac{3}{8}$	9.4	20
11.1	$\frac{7}{16}$	9.4	20
12.7	$\frac{1}{2}$	9.4	20

(A) SHIELDING GAS, 99.95% PURE ARGON

When stud arc welding low-carbon and stainless steel studs, the stud (electrode) should be negative. Aluminum stud arc welding should be set up so that the stud is positive.

Obtaining a full-quality stud arc weld requires a sufficient amount of total energy input to the weld joint in order to produce melting and complete metallurgical bonding of the stud and workpiece. The energy input or weld current that is necessary depends on the stud diameter. Other parameters in the stud arc welding process are: arc voltage, arc time, and plunge. Arc voltage is a function of the arc length, which is set as "lift," or the length the stud is drawn away from the workpiece during weld-cycle initiation. Plunge is the length of stud that extends past the end of the ceramic ferrule and is available for melt off during the weld cycle.

Studs are reduced in length during the weld cycle. Table 3 shows the typical reductions for steel and aluminum studs as a result of welding.

TABLE 3 TYPICAL STUD WELDING SETTINGS

STUD BASE DIAMETER		LENGTH REDUCTION		DOWNHAND WELDING							OVERHEAD WELDING							VERTICAL WELDING						
				CURRENT, A	TIME		LIFT		PLUNGE		CURRENT, A	TIME		LIFT		PLUNGE		CURRENT, A	TIME		LIFT		PLUNGE	
mm	in.	mm	in.		s	cycles	mm	in.	mm	in.		s	cycles	mm	in.	mm	in.		s	cycles	mm	in.	mm	in.
STEEL AND STAINLESS STEEL																								
4.8	0.187	3.2	0.125	300	0.15	8-9	1.6	0.062	3.2	0.125	300	0.15	8-9	1.6	0.062	3.2	0.125	300	0.15	8-9	1.6	0.062	3.2	0.125
6.4	0.250	3.2	0.125	450	0.17	10-12	1.6	0.062	3.2	0.125	450	0.15	10-12	1.6	0.062	3.2	0.125	450	0.17	10-12	1.6	0.062	3.2	0.125
7.9	0.312	3.2	0.125	500	0.25	15	1.6	0.062	3.2	0.125	500	0.25	15	1.6	0.062	3.2	0.125	500	0.25	15	1.6	0.062	3.2	0.125
9.5	0.375	3.2	0.125	550	0.33	20	1.6	0.062	3.2	0.125	550	0.33	20	1.6	0.062	3.2	0.125	600	0.33	20	1.6	0.062	3.2	0.125
11.1	0.437	3.2	0.125	675	0.42	25	1.6	0.062	3.2	0.125	675	0.42	25	1.6	0.062	3.2	0.125	750	0.33	20	1.6	0.062	3.2	0.125
12.7	0.500	4.8	0.125	800	0.55	33	1.6	0.062	3.2	0.125	800	0.55	33	1.6	0.062	3.2	0.125	875	0.47	28	1.6	0.062	3.2	0.125
15.9	0.625	4.8	0.187	1200	0.67	40	2.4	0.093	4.7	0.187	1200	0.67	40	2.4	0.093	4.7	0.187	1275	0.60	36	1.6	0.062	4.7	0.187 (A)
19.1	0.750	4.8	0.187	1500	0.84	50-55	2.4	0.093	4.7	0.187	1500	0.84	50-55	2.4	0.093	4.7	0.187	1700	0.73	50	2.4	0.093	4.7	0.187 (A)
22.2	0.875	4.8	0.187	1700	1.00	60-65	3.2	0.125	6.4	0.250	1700	1.00	65	3.2	0.125	6.4	0.250	NOT RECOMMENDED						
25.4	1.000	6.4	0.250	1900	1.40	85	3.2	0.125	6.4	0.250	2050	1.40	72	3.2	0.125	6.4	0.250	NOT RECOMMENDED						
ALUMINUM (A)																								
4.8	0.187	3.2	0.125	150	0.25	15	2.4	0.093	3.2	0.125	150	0.25	15	2.4	0.093	3.2	0.125	180	0.20	12	2.4	0.093	3.2	0.125
6.4	0.250	3.2	0.125	200	0.40	24	2.4	0.093	3.2	0.125	200	0.40	24	2.4	0.093	3.2	0.125	225	0.30	18	2.4	0.093	3.2	0.125
7.9	0.312	3.2	0.125	250	0.50	30	2.4	0.093	3.2	0.125	250	0.50	30	2.4	0.093	3.2	0.125	275	0.40	24	2.4	0.093	3.2	0.125
9.5	0.375	3.2	0.125	325	0.65	39	3.2	0.125	3.2	0.125	325	0.65	39	2.4	0.093	3.2	0.125	350	0.60	36	3.2	0.125	3.2	0.125

11.1	0.437	3.2	0.125	400	0.80	48	3.2	0.125	3.9	0.156	400	0.80	48	3.9	0.156	3.9	0.156	430	0.70	42	3.2	0.125	3.9	0.156
12.7	0.500	3.2	0.125	460	0.90	54	3.2	0.125	4.7	0.187	460	0.90	54	4.7	0.187	4.7	0.187	475	0.80	48	3.2	0.125	4.7	0.187

(A) STUD ARC WELDED ALUMINUM STUDS REQUIRE INERT GAS SHIELDING: SEE TABLE 2.

There is a range of setting combinations for all stud diameters. The same total energy input can be obtained by varying current input and arc time. For example, a low current input can be compensated for, to some extent, by increasing weld time. Table 3 also shows typical weld settings for various SW stud diameters. The ideal settings in each weld position for a particular stud and workpiece should be established by beginning with typical settings and then varying them within the allowable range to meet the required conditions.

Typical settings are based on good weld conditions, such as a clean workpiece, good ground, and others. Also necessary is a proper dc power source, which should have the following characteristics:

- HIGH OPEN-CIRCUIT VOLTAGE (70 TO 100 V)
- RAPID OUTPUT CURRENT RISE
- DROOPING OUTPUT VOLT-AMPERE CURVE
- HIGH CURRENT OUTPUT OVER A SHORT TIME

Stud arc welding requires a very short weld time (Table 3), but a very high current input, when compared with other types of arc welding. Usually, the duty cycle for this process is much lower than that of other welding processes. Constant-voltage types of dc power sources are not suitable for stud welding. Many of the available dc power sources are acceptable for stud welding, including those supplied as integrated power/control systems by the stud-welding manufacturer. These integrated systems use both three-phase and single-phase incoming alternating current (ac) power. Both standard and special voltage units are available. Single-phase units are relatively low cost, portable systems that are usually suitable for smaller-diameter (≤ 12.7 mm, or $\frac{1}{2}$ in.) stud arc welding. Three-phase units are preferred for larger-diameter studs, because they provide a balanced load on incoming power lines and have smoother arc characteristics. A transformer/rectifier power source, combined with an integral stud-welding control system, is shown in Fig. 2.

Weld cable length and cable size also can affect the total output of a power source. Basically, a larger welding cable size minimizes current loss that is due to cable length, resistance, and heating. Thus, when the distance from the power source to the gun increases significantly, or when the number and size of studs applied per minute are large, larger welding cables should be used. Smaller dc power sources suitable for stud arc welding can be wired in parallel to produce the necessary current range.

Ferrules, or ceramic arc shields, are used in most stud arc weld applications. They are available in a wide variety of sizes and shapes to fit specific stud base designs and applications. For example, there are specific designs for welding studs through metal deck, to the fillet or heel of an angle, to round pipes or bars, or in a vertical position. A ferrule is placed over every stud at the weld end, where it is held in place by a grip on the stud-welding gun. The ferrule is used only once and must be removed from the stud after welding is completed to allow the inspection of the completed weld. Typical ferrule configurations are shown in Fig. 8.

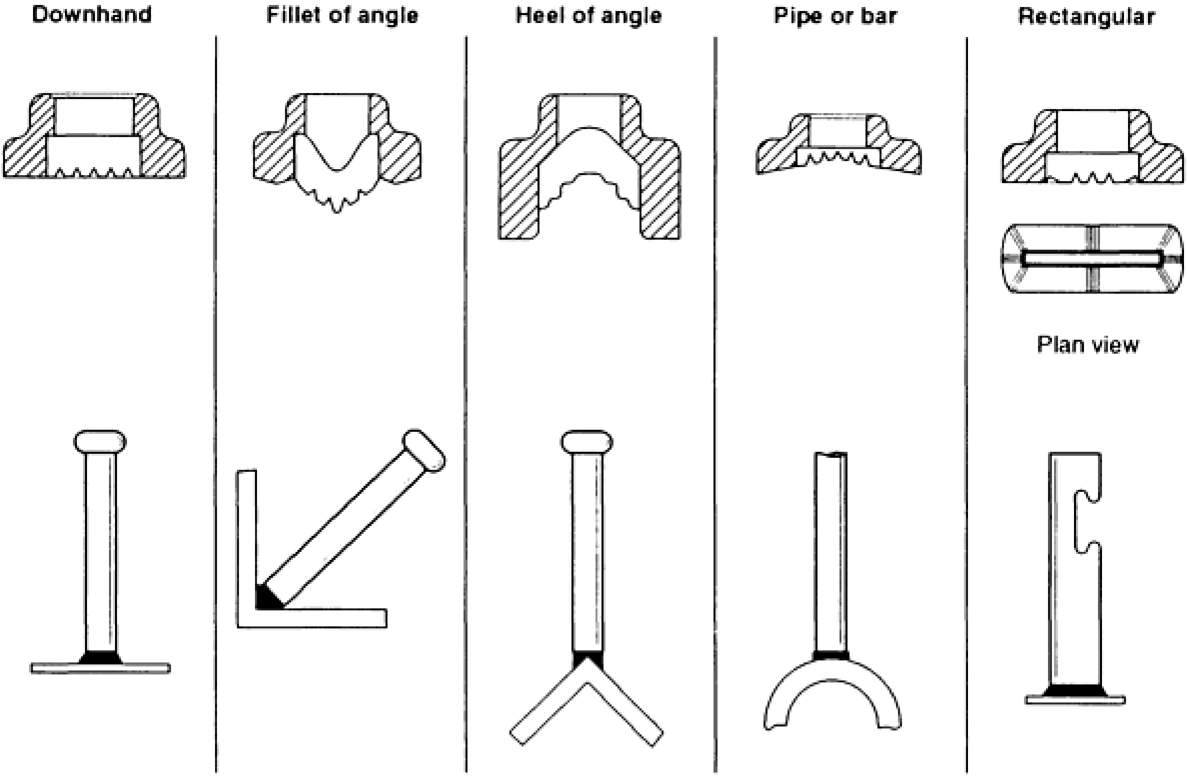

FIG. 8 CROSS SECTIONS OF FERRULES SHOWING SOME OF THE MANY VARIETIES AVAILABLE FOR DIFFERENT STUD BASE GEOMETRIES, WELD POSITIONS, AND APPLICATIONS

Ferrule design is important, because it controls certain functions during welding. For example:

- VENTS ON THE BOTTOM OF THE FERRULE ALLOW WELD GASES TO ESCAPE FROM THE WELD AREA AND RESTRICT THE INFLOW OF AIR TO MINIMIZE WELD POROSITY AND OXIDATION.
- THE INTERNAL CAVITY OF THE FERRULE CONFINES AND SHAPES THE MOLTEN METAL INTO A FLASH (FILLET) AROUND THE STUD PERIPHERY.
- HEAT IN THE WELD AREA IS CONCENTRATED AND CONTAINED.
- FLASH AND WELD SPATTER ARE MINIMIZED.

Although the dimensions of the flash are controlled by the ferrule configuration, the flash diameter and height must be taken into consideration when designing mating parts. Manufacturer specifications on finished stud flash dimensions should be followed, and test welds should be made to establish part fit. Flash dimensions can be accommodated by any of the five methods shown in Fig. 9.

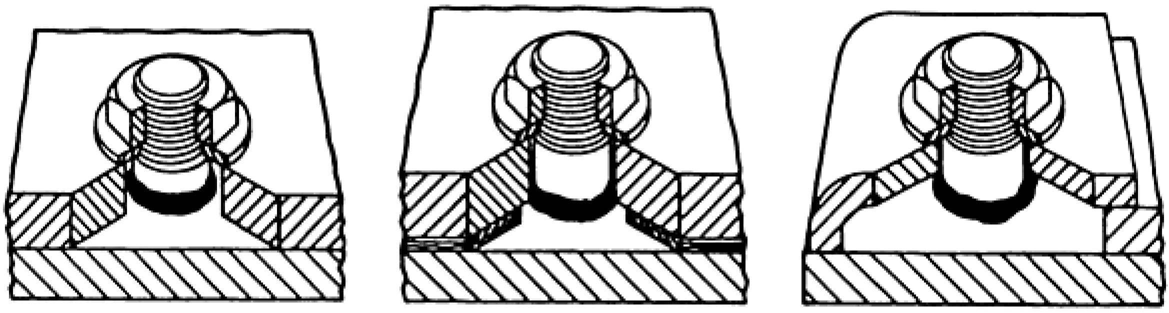

Oversize hole Gasket or filler plate Dog clamp

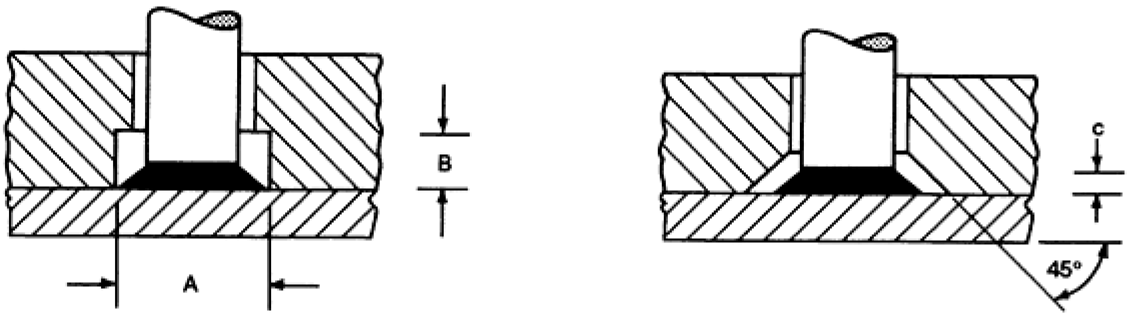

Counterbore Countersink

FIG. 9 COMMON METHODS FOR ACCOMMODATING WELD FLASH (FILLET) WITH STUD ARC WELDS. TYPICAL DIMENSIONS FOR A, B, AND C ARE GIVEN IN TABLE 4.

The typical dimensions for flash clearance are shown in Table 4. It should be noted that the flash (fillet) of a stud weld is not the same as the fillet that is produced by conventional welding techniques. It may have areas of nonfusion on its vertical leg or shrink fissures, which have no adverse affect on weld strength or ductility (Ref 4). Consequently, it is not subject to the profile and inspection criteria used for a conventional weld fillet.

TABLE 4 DIMENSIONS FOR COUNTERBORE AND COUNTERSINK WELD FLASH CLEARANCE

A, B, and C are shown in Fig. 9.

STUD BASE DIAMETER		COUNTERBORE				90° COUNTERSINK	
		A		B		C	
mm	in.	mm	in.	mm	in.	mm	in.
6.4	1/4	11.1	0.437	3.2	0.125	3.2	0.125
7.9	5/16	12.7	0.500	3.2	0.125	3.2	0.125
9.5	3/8	15.1	0.593	3.2	0.125	3.2	0.125
11.1	7/16	16.7	0.656	4.7	0.187	3.2	0.125
12.7	1/2	19.1	0.750	4.7	0.187	4.7	0.187
15.9	5/8	22.2	0.875	5.5	0.218	4.7	0.187

19.1	3/4	28.6	1.125	7.9	0.312	4.7	0.187

Note: Dimensions can vary, depending on stud style and ceramic arc shield selected. Consult manufacturer for details.

Base Plate Material and Thickness. To obtain full strength, the stud arc welding of low-carbon steel, stainless steel, or aluminum studs should be made to a base material of sufficient thickness so that the stud fails in tension or torque, rather than by pulling a hole in the base-metal workpiece. In the case of stainless or low-carbon steel studs, the base-metal thickness should be at least one-third that of the stud diameter. For aluminum studs, the aluminum base plate material should be at least half that of the stud diameter. Thinner base plates can be used if strength is not the primary design characteristic that is required. On thinner materials, the studs will pull a hole in the plate at strength levels below their maximum. If the plate is too thin, then the stud will bum a hole through it. The recommended base plate thickness for full-strength development and the minimum thicknesses required to prevent burning through the base plate are given in Table 5.

TABLE 5 RECOMMENDED BASE METAL THICKNESSES FOR STUD ARC WELDING

STUD BASE DIAMETER		STEEL BASE METAL				ALUMINUM BASE METAL			
		FOR FULL STRENGTH(A)		MINIMUM(B)		WITHOUT BACKUP		MINIMUM WITH BACKUP(C)	
mm	**in.**	**mm**	**in.**	**mm**	**in.**	**mm**	**in.**	**mm**	**in.**
4.8	3/16	1.59	0.062	0.91	0.036	3.18	0.125	3.18	0.125
6.4	1/4	3.18	0.125	1.21	0.048	3.18	0.125	3.18	0.125
7.9	5/16	3.18	0.125	1.52	0.060	4.76	0.187	3.18	0.125
9.5	3/8	4.76	0.187	1.90	0.075	4.76	0.187	4.76	0.187
11.1	7/16	4.76	0.187	2.28	0.090	6.35	0.250	4.76	0.187
12.7	1/2	4.76	0.187	3.04	0.120	6.35	0.250	6.35	0.250
15.9	5/8	6.35	0.250	3.68	0.145	. . .	. . .	. . .	. . .
19.1	3/4	7.94	0.318	4.70	0.185	. . .	. . .	. . .	. . .
22.2	7/8	9.53	0.375	6.35	0.250	. . .	. . .	. . .	. . .
25.4	1	9.53	0.375	9.53	0.375	. . .	. . .	. . .	. . .

(A) NEAREST COMMONLY AVAILABLE BASE METAL THICKNESS.

(B) STEEL WITHOUT BACKUP--WILL NOT DEVELOP FULL STUD STRENGTH.

(C) METAL BACKUP REQUIRED TO PREVENT MELT-THROUGH OF BASE METAL

The most widely used combinations of base plate and stud materials are shown in Table 6. With low-carbon steel studs, no preheat or postheat treatments are needed when welding to low-carbon steel or austenitic stainless steel base plate materials. As carbon content in the base plate increases into the medium range, heat treatment for stud welding may be required. Welding to high-carbon base plate is not suggested. When base plate/stud combinations are questionable, an application qualification test is suggested (Ref 5). Details on this procedure are provided in the section "Stud-Welding Quality Control, Qualification, and Inspection" in this article. For best results when welding stainless steel studs to low-carbon steel plate (especially when working stress levels are very high or there are repetitive loading cycles), the studs should either be manufactured from annealed-in-process materials or postmanufacture annealed before use.

TABLE 6 TYPICAL COMBINATIONS OF BASE AND STUD METALS FOR STUD ARC WELDING

BASE METAL	STUD METAL
LOW-CARBON STEEL, AISI 1006 TO 1022(A)	LOW-CARBON STEEL, AISI 1006 TO 1020; STAINLESS STEEL, 300 SERIES
STAINLESS STEEL, 300 SERIES(B), 405, 410, AND 430	LOW-CARBON STEEL, AISI 1006 TO 1020; STAINLESS STEEL, 300 SERIES
ALUMINUM ALLOYS, 5000 SERIES(C)	ALUMINUM ALLOYS, 5000 SERIES(C)

Source: Ref 1

(A) REFER TO ANSI/AWS D1.1-92, TABLE 4.1 (GROUPS I AND II) FOR APPROVED STEELS.
(B) EXCEPT FOR THE FREE-MACHINING TYPE 303 STAINLESS STEEL.
(C) REFER TO ANSI/AWS D1.2-89, TABLES 7.1 AND 7.4.

Good welding practice requires that the base plate material at the weld spot be cleaned. Common contaminants that can result in unsatisfactory welds if not removed include paint, heavy mill scale, heavy rust, oxidation, oil or grease, and plating, such as galvanizing and anodizing on aluminum. Removal methods vary according to the contaminant, the base material, and the end use. Wire brushing and grinding may be satisfactory on heavy structural steel but can be destructive to thin-gage metals or anodized aluminum, because they may reduce the thickness to a point where bum through or less-than-full-strength welds result.

Suggested cleaning methods include grinding, wire brushing, or needle scaling on heavy materials; using solvents to clean grease and oil; and using a noncontaminating stainless wire brush or a nonmetallic foam disk to clean aluminum and milling. Because anodizing is difficult to remove from aluminum without reducing the metal thickness, consideration should be given to anodizing the base plate and stud after the welding operation.

Similarly, the weld end of the stud should be free from materials that would contaminate the weld, including paint, rust, galvanizing, and others. Studs cannot be coated with any nonconductive material that would interfere with the flow of welding current. Light copper flashing and nickel or chrome plating usually do not cause welding problems.

Because a high-amperage, short-time current is typical in the stud arc welding process, it is also important that the grounding spot(s) on the base material be clean, that the ground be tight, and that all cables be in good condition and have tight connections.

The position of the ground(s) also can influence stud-welding quality. This depends on the geometry of the base-metal shape and is the result of electromagnetic arc deflection, or arc blow (Ref 6). Occurrences are more common when studs are very near a free edge or are on long and narrow base plates, hollow pipe sections, or irregular peripheries, and can be corrected by moving the ground to a central position or using multiple grounds. Because current flow is usually away from the ground toward the heavier or larger area of the base plate, arc blow is characterized by a lack of flash or fillet on the side of the stud nearer the edge, end, or smaller area.

Stud Strength. Fasteners will develop full material strength when they have been stud arc welded to compatible base plate alloys. Tables 7, 8, and 9 show the tension and torque loads for various threaded stud diameters, based on minimum specified stud strengths (Ref 2, 3, and 6). For unthreaded stud fasteners, the yield strength and tensile or ultimate strength is calculated by:

$$\text{YIELD STRENGTH} = A_S F_Y$$
$$\text{TENSILE STRENGTH} = A_S F_U$$

where A_s is the area of the stud shank (in.^{2}), F_y is the specified stud material yield strength (minimum psi) and F_u is the specified stud material ultimate strength (minimum psi).

TABLE 7 MECHANICAL PROPERTIES OF LOW-CARBON STEEL STUD ARC WELDED FASTENERS

Fasteners have 380 MPa (55 ksi) minimum ultimate strength and 345 MPa (50 ksi) minimum yield strength.

STUD THREAD DIAMETER(A)	**MEAN EFFECTIVE THREAD AREA(B)**		**YIELD TENSILE LOAD(C)**		**ULTIMATE TENSILE LOAD**		**YIELD TORQUE(D)**		**ULTIMATE TORQUE**		**ULTIMATE SHEAR LOAD(E)**	
	mm^2	**$in.^2$**	**kn**	**lbf**	**kn**	**lbf**	**j**	**ft · lbf**	**j**	**ft · lbf**	**kn**	**lbf**
10-24 UNC	11	0.017	3.8	850	4.2	935	3.6	32(F)	4.0	35.1(F)	3.1	701
10-32 UNF	13	0.020	4.4	1,000	4.9	1,100	4.3	38(F)	4.7	41.3(F)	3.7	825
$\frac{1}{4}$-20 UNC	21	0.032	7.1	1,600	7.8	1,760	9.1	6.7	9.9	7.3	5.9	1,320
$\frac{1}{4}$-28 UNF	23	0.036	8.0	1,800	8.8	1,980	10.2	7.5	11.3	8.3	6.6	1,485
$\frac{5}{16}$-18 UNC	34	0.052	11.6	2,600	12.7	2,860	18.4	13.6	20.2	14.9	9.5	2,145
$\frac{5}{16}$-24 UNF	37	0.058	12.9	2,900	14.2	3,190	20.5	15.1	22.5	16.6	10.6	2,393
$\frac{3}{8}$-16 UNC	50	0.078	17.3	3,900	19.1	4,290	33.1	24.4	36.3	26.8	14.3	3,218
$\frac{3}{8}$-24 UNF	57	0.088	19.6	4,400	21.5	4,840	37.3	27.5	41.1	30.3	16.1	3,630
$\frac{7}{16}$-14 UNC	68	0.106	23.6	5,300	25.9	5,830	52.3	38.6	57.6	42.5	19.5	4,373
$\frac{7}{16}$-20 UNF	76	0.118	26.2	5,900	28.9	6,490	58.3	43.0	64.1	47.3	21.7	4,868
$\frac{1}{2}$-13 UNC	92	0.142	31.6	7,100	34.7	7,810	80.3	59.2	88.3	65.1	26.0	5,856
$\frac{1}{2}$-20 UNF	103	0.160	35.6	8,000	39.1	8,800	90.4	66.7	99.4	73.3	29.4	6,600
$\frac{5}{8}$-11	146	0.226	50.3	11,300	55.3	12,430	159.6	117.7	175.5	129.5	41.5	9,323

UNC											
$\frac{5}{8}$-18 UNF	165	0.255	56.7	12,750	62.4	14,025	180.1	132.8	198.1	146.1	46.8	10,519
$\frac{3}{4}$-10 UNC	215	0.334	74.3	16,700	81.7	18,370	283.1	208.8	311.3	229.6	61.3	13,778
$\frac{3}{4}$-16 UNF	240	0.372	82.7	18,600	91.0	20,460	315.2	232.5	346.8	255.8	68.3	15,345
$\frac{7}{8}$-9 UNC	298	0.462	102.8	23,100	112.8	25,355	456.8	336.9	501.4	369.8	84.6	19,017
$\frac{7}{8}$-14 UNF	328	0.509	113.2	25,450	124.5	27,995	503.1	371.1	553.6	408.3	93.4	20,996
1-8 UNC	91	0.606	134.8	30,300	148.0	33,275	684.7	505.0	751.9	554.6	111.0	24,956
1-14 UNF	437	0.678	150.8	33,900	165.9	37,290	766.0	565.0	842.6	621.5	124.4	27,967

(A) UNC, UNIFIED COARSE THREAD SERIES; UNF, UNIFIED FINE THREAD SERIES.

(B) MEAN EFFECTIVE THREAD AREA IS BASED ON A MEAN FULL DIAMETER MIDWAY BETWEEN MINOR AND PITCH THREAD DIAMETERS.

(C) IN PRACTICE, A STUD SHOULD NOT BE USED AT OR HIGHER THAN YIELD LOAD. A FACTOR OF SAFETY SHOULD BE APPLIED, AND 60% OF YIELD IS COMMONLY USED, ALTHOUGH OTHER VALUES MAY BE USED AT DISCRETION OF USER.

(D) TORQUE FIGURES ARE BASED ON THE ASSUMPTION THAT EXCESSIVE THREAD DEFORMATION HAS NOT AFFECTED THE PROPORTIONAL RANGE OF TORQUE/TENSION RELATIONSHIP. AN AVERAGE TORQUE COEFFICIENT OF 0.20 WAS USED IN THESE CALCULATIONS.

(E) SHEAR LOAD IS BASED UPON 0.75 TIMES ULTIMATE LOAD. THE USER SHOULD APPLY AN APPROPRIATE SAFETY FACTOR TO THESE FIGURES.

(F) VALUE GIVEN IN IN. · LBF.

TABLE 8 MECHANICAL PROPERTIES OF STAINLESS STEEL STUD ARC WELDED FASTENERS

Fasteners have 520 MPa (75 ksi) minimum ultimate strength and 205 MPa (30 ksi) minimum yield strength.

STUD THREAD DIAMETER(A)	**MEAN EFFECTIVE THREAD AREA(B)**		**YIELD TENSILE LOAD(C)**		**ULTIMATE TENSILE LOAD**		**YIELD TORQUE(D)**		**ULTIMATE TORQUE**		**ULTIMATE SHEAR LOAD(E)**	
	mm^2	**in.^{2}**	**kn**	**lbf**	**kn**	**lbf**	**j**	**ft · lbf**	**j**	**ft · lbf**	**kn**	**lbf**
10-24 UNC	11	0.017	2.3	510	5.7	1,275	2.2	19.1(F)	5.4	47.8(F)	4.3	956
10-32 UNF	13	0.020	2.7	600	6.7	1,500	2.5	22.5(F)	6.4	56.3(F)	5.0	1,125
$\frac{1}{4}$-20 UNC	21	0.032	4.3	960	10.7	2,400	5.4	4.0	13.6	10.0	8.0	1,800
$\frac{1}{4}$-28 UNF	23	0.036	4.8	1,080	12.0	2,700	6.1	4.5	15.3	11.3	9.0	2,025
$\frac{5}{16}$-18 UNC	34	0.052	6.9	1,560	17.3	3,900	11.0	8.1	27.5	20.3	13.0	2,925

$\frac{5}{16}$-24 UNF	37	0.058	7.7	1,740	19.3	4,350	12.3	9.1	30.6	22.6	14.5	3,263
$\frac{3}{8}$-16 UNC	50	0.078	10.4	2,340	26.0	5,850	19.8	14.6	49.6	36.6	19.5	4,388
$\frac{3}{8}$-24 UNF	57	0.088	11.7	2,640	29.4	6,600	22.4	16.5	56.0	41.3	22.0	4,950
$\frac{7}{16}$-14 UNC	68	0.106	14.1	3,180	35.4	7,950	31.5	23.2	78.6	58.0	26.5	5,963
$\frac{7}{16}$-20 UNF	76	0.118	15.7	3,540	39.4	8,850	35.0	25.8	87.4	64.5	29.5	6,638
$\frac{1}{2}$-13 UNC	92	0.142	18.9	4,260	47.4	10,650	48.1	35.5	120.4	88.8	35.5	7,988
$\frac{1}{2}$-20 UNF	103	0.160	21.4	4,800	53.4	12,000	54.2	40.0	135.6	100.0	40.0	9,000
$\frac{5}{8}$-11 UNC	146	0.226	30.2	6,780	75.4	16,950	95.7	70.6	239.4	176.6	56.5	12,713
$\frac{5}{8}$-18 UNF	165	0.255	34.0	7,650	85.1	19,125	108.1	79.7	270.1	199.2	63.8	14,344
$\frac{3}{4}$-10 UNC	215	0.334	44.6	10,020	111.4	25,050	169.9	125.3	424.50	313.1	83.6	18,788
$\frac{3}{4}$-16 UNF	240	0.372	49.6	11,160	124.1	27,900	189.1	139.5	472.9	348.8	93.1	20,925
$\frac{7}{8}$-9 UNC	298	0.462	61.7	13,860	154.1	34,650	274.0	202.1	685.1	505.3	115.6	25,988
$\frac{7}{8}$-14 UNF	328	0.509	67.9	15,270	169.8	38,175	301.9	222.7	754.8	556.7	127.4	28,631
1-8 UNC	391	0.606	80.9	18,180	202.2	45,450	410.8	303.0	1027.0	757.5	151.6	34,088
1-14 UNF	437	0.678	90.5	20,340	226.2	50,850	459.6	339.0	1149.0	847.5	169.6	38,138

(A) UNC, UNIFIED COARSE THREAD SERIES; UNF, UNIFIED FINE THREAD SERIES.

(B) META, MEAN EFFECTIVE THREAD AREA, IS BASED ON A MEAN FULL DIAMETER MIDWAY BETWEEN MINOR AND PITCH THREAD DIAMETERS.

(C) IN PRACTICE, A STUD SHOULD NOT BE USED AT OR HIGHER THAN YIELD LOAD. A FACTOR OF SAFETY SHOULD BE APPLIED, AND 60% OF YIELD IS COMMONLY USED, ALTHOUGH OTHER VALUES MAY BE USED AT THE DISCRETION OF USER.

(D) TORQUE FIGURES ARE BASED ON THE ASSUMPTION THAT EXCESSIVE THREAD DEFORMATION HAS NOT AFFECTED THE PROPORTIONAL RANGE OF TORQUE/TENSION RELATIONSHIP. AN AVERAGE TORQUE COEFFICIENT OF 0.20 WAS USED IN THESE CALCULATIONS.

(E) SHEAR LOAD IS BASED UPON 0.75 TIMES ULTIMATE LOAD. THE USER SHOULD APPLY AN APPROPRIATE SAFETY FACTOR TO THESE FIGURES.

(F) VALUE GIVEN IN IN. · LBF.

TABLE 9 MECHANICAL PROPERTIES OF ALUMINUM STUD ARC WELDED FASTENERS

Fasteners have 290 MPa (42 ksi) minimum ultimate strength and 205 MPa (30 ksi) minimum yield strength.

STUD THREAD DIAMETER(A)	**MEAN EFFECTIVE THREAD AREA(B)**		**YIELD TENSILE LOAD(C)**		**ULTIMATE TENSILE LOAD**		**YIELD TORQUE(D)**		**ULTIMATE TORQUE**		**ULTIMATE SHEAR LOAD(E)**	
	mm^2	**in.^{2}**	**kn**	**lbf**	**kn**	**lbf**	**j**	**ft · lbf**	**j**	**ft · lbf**	**kn**	**lbf**
10-24 UNC	11	0.017	2.3	510	3.2	714	2.2	19.1(F)	3.0	26.8(F)	1.9	428
10-32 UNF	13	0.020	2.7	600	3.9	840	2.5	22.5(F)	3.6	31.5(F)	2.2	504
$\frac{1}{4}$-20 UNC	21	0.032	4.3	960	6.0	1344	5.4	4.0	7.6	5.6	3.6	806
$\frac{1}{4}$-28 UNF	23	0.036	4.8	1080	6.7	1512	6.1	4.5	8.5	6.3	4.0	907
$\frac{5}{16}$-18 UNC	34	0.052	6.9	1560	9.7	2184	11.0	8.1	15.5	11.4	5.8	1310
$\frac{5}{16}$-24 UNF	37	0.058	7.7	1740	10.8	2436	12.3	9.1	17.2	12.7	6.5	1462
$\frac{3}{8}$-16 UNC	50	0.078	10.4	2340	14.6	3276	19.8	14.6	27.8	20.5	8.7	1966
$\frac{3}{8}$-24 UNF	57	0.088	11.7	2640	16.4	3696	22.3	16.5	31.3	23.1	9.6	2218
$\frac{7}{16}$-14 UNC	68	0.106	14.1	3180	19.8	4452	31.5	23.2	44.1	32.5	11.9	2671
$\frac{7}{16}$-20 UNF	76	0.118	15.7	3540	22.0	4956	35.0	25.8	48.9	36.1	13.2	2974
$\frac{1}{2}$-13 UNC	92	0.142	18.9	4260	26.5	5964	48.1	35.5	67.4	49.7	15.9	3578
$\frac{1}{2}$-20 UNF	103	0.160	21.4	4800	29.9	6720	54.2	40.0	75.9	56.0	17.9	4032

(A) UNC, UNIFIED COARSE THREAD SERIES; UNF, UNIFIED FINE THREAD SERIES.

(B) MEAN EFFECTIVE THREAD AREA IS BASED ON A MEAN FULL DIAMETER MIDWAY BETWEEN MINOR AND PITCH THREAD DIAMETERS.

(C) IN PRACTICE, A STUD SHOULD NOT BE USED AT OR HIGHER THAN YIELD LOAD. A FACTOR OF SAFETY SHOULD BE APPLIED, AND 60% OF YIELD IS COMMONLY USED, ALTHOUGH OTHER VALUES MAY BE USED AT DISCRETION OF USER.

(D) TORQUE FIGURES ARE BASED ON THE ASSUMPTION THAT EXCESSIVE THREAD DEFORMATION HAS NOT AFFECTED THE PROPORTIONAL RANGE OF TORQUE/TENSION RELATIONSHIP. AN AVERAGE TORQUE COEFFICIENT OF 0.20 WAS USED IN THESE CALCULATIONS.

(E) SHEAR LOAD IS BASED ON 0.60 TIMES ULTIMATE LOAD. THE USER SHOULD APPLY AN APPROPRIATE SAFETY FACTOR TO THESE FIGURES.

(F) VALUE GIVEN IN IN. · LBF.

For threaded studs, the area is based on the mean effective thread area, which is calculated by:

$$A_S = 0.7854\,[D- (0.9743/N)]^2$$

where *D* is the nominal diameter of the stud and *N* is the number of threads per inch.

Process Variations and Special Equipment. There are several specific applications that lend themselves to special variations of the stud arc welding technique. One application is the welding of studs to thin base materials that are less than the minimum thicknesses listed in Table 5. Although full weld base strength is usually not achieved, the resulting strength level is suitable for the application loadings involved. This process variation, which is called short-cycle stud arc welding, does not use a ceramic ferrule. Instead, higher weld current is used with very short times, which minimizes penetration of the stud into the base plate. Normally, this variation is used with stud diameters ≤9.5 mm (≤0.375 in.) in situations where backside marking is not considered detrimental.

The second special process is gas-shielded arc welding, which also does not use a ceramic ferrule. Instead, the stud-welding area is shielded by an inert gas, usually argon. This process can be used with both steel and aluminum, but is more widely used with the latter material. The welding variables fall into a very narrow range, and conditions for application usually include tight tolerances and a tightly controlled setup. Consequently, the stud arc welding equipment most often consists of a fixed-gun production unit. A typical application is the welding of aluminum studs with special end configurations to aluminum kitchen utensils.

Special equipment is frequently used with the short-cycle and/or gas-shielded arc welding processes, because the application parameters must be tightly controlled and involve large quantities of studs applied in a production environment. Equipment manufacturers assemble many types of special equipment for production-line use. The equipment can be used with automatic stud feed, automatic stud and ferrule feed, computer-programmed indexing, robotic stud-welding guns, and other mechanisms. A typical automated stud arc welding production unit is shown in Fig. 10. A unit can involve simple, column-mounted, single-gun systems or sophisticated, multigun, multiple feed units that cost thousands of dollars.

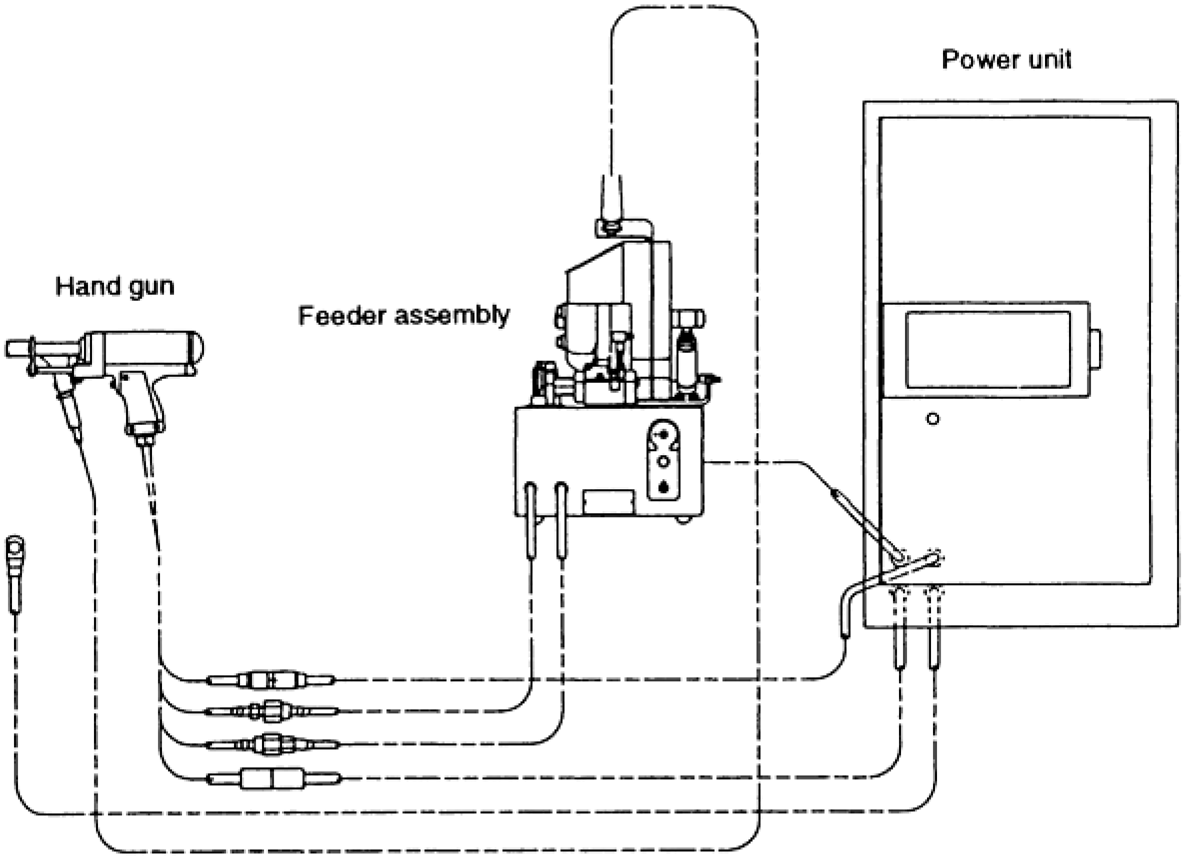

FIG. 10 TYPICAL AUTOMATIC STUD-FEED HAND-GUN SYSTEM FOR PORTABLE OPERATION USING TEMPLATES OR FIXTURES ON THE WORKPIECE

References cited in this section

1. "RECOMMENDED PRACTICES FOR STUD WELDING," ANSI/AWS C5.4, AWS
2. "STRUCTURAL WELDING CODE--STEEL," ANSI/AWS D1.1, SECTION 7, "STUD WELDING," AWS
3. "STRUCTURAL WELDING CODE--ALUMINUM," ANSI/AWS D1.2, AWS
4. "STRUCTURAL WELDING CODE--STEEL," ANSI/AWS D1.1, SECTION 7, "STUD WELDING," FOOTNOTE 27, AWS
5. "STRUCTURAL WELDING CODE--STEEL," ANSI/AWS D1.1, SECTION 7, "STUD WELDING," PARAGRAPH 7.6, AWS
6. W.A. BAESLACK, G. FAYER, S. REAM, AND C.E. JACKSON, QUALITY CONTROL IN ARC STUD WELDING, *WELD. J.,* NOV 1975, P 789-798

Stud Arc Welding

Harry A. Chambers, TRW Nelson Stud Welding Division

Fixturing and Tooling for Stud Arc Welding

Regardless of the stud-welding method employed, certain "expendable" accessories are required. These include such items as a chuck or collet to hold the particular stud being welded, a ferrule grip for the fastener, a foot that holds the ferrule grip, legs that attach to the gun and adjust to accommodate various stud lengths, and other minor items. The operating life of these items is variable, depending on the care and maintenance they receive. For example, a chuck may last for 5000 to 25,000 welds. Legs last indefinitely, whereas ferrule grips and feet usually have a shorter usage life. Accessories are relatively inexpensive items that should be planned for when preparing budgetary estimates on any given project.

The extent and sophistication of tooling for stud welding reflects the required production rate and the total number of studs to be welded. Locating the stud arc welding centers can be as simple as laying out the workpiece and center punching the locations either directly or through a template, as shown in Fig. 11. The studs are then placed in the punch marks, the stud-welding gun is held vertically, and the weld is initiated. Although operator skill is a factor, careful welders can achieve a perpendicularity of ±5° and a location tolerance of approximately 1.2 mm (0.046 in.). Often, the cover plate in a base plate/cover plate assembly can be used as the template for marking.

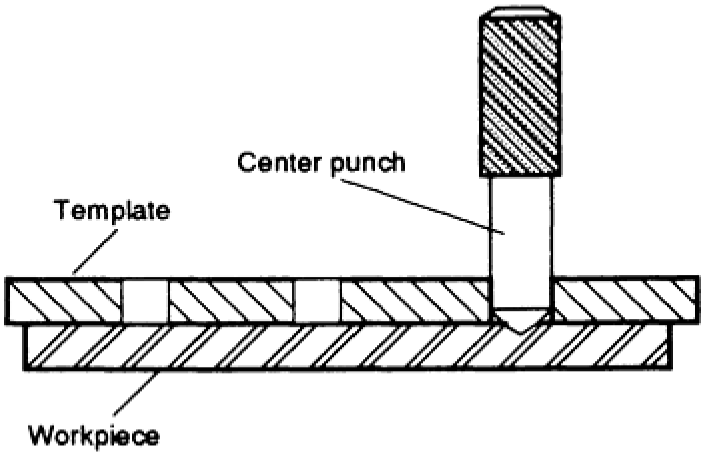

FIG. 11 CENTER PUNCHING BASE MATERIAL THROUGH A TEMPLATE

A permanent template becomes more practical when an increasing number of studs are to be welded on repetitive locations. Stud arc welded fasteners can be welded directly through temperature-resistant material which should be spaced off the work by 1.65 to 6.35 mm (0.065 to 0.250 in.) to allow expelled weld gases and weld spatter to escape without restriction, which could adversely affect weld quality. This type of template arrangement is shown in Fig. 12. Note that the template holes are drilled slightly larger than the outside dimension of the ceramic ferrule. Before the template is prepared, it is good practice to consult the manufacturer specifications for the ceramic ferrule dimensions to be used with the stud being welded. Location accuracy, in this case, can be as tight as ±0.78 mm (±0.031 in.).

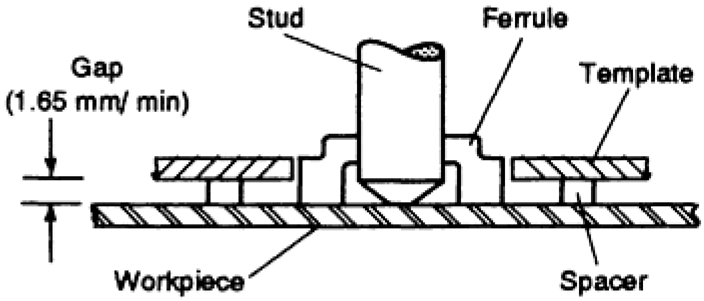
Stud
Ferrule
Gap
(1.65 mm/ min)
Template
Workpiece
Spacer

FIG. 12 WELDING THROUGH A TEMPLATE USING A FERRULE WITH STUD ARC WELDING FASTENERS

Even greater accuracy can be obtained, both in the vertical and horizontal directions, by adapting the template hole to fit a drill jig bushing tightly and by inserting the template adapter (or a ferrule tube adapter) through the bushing to the workpiece. The ferrule tube adapter also can be used for welding down into deeply drilled holes or into tight-fit areas (Fig. 13).

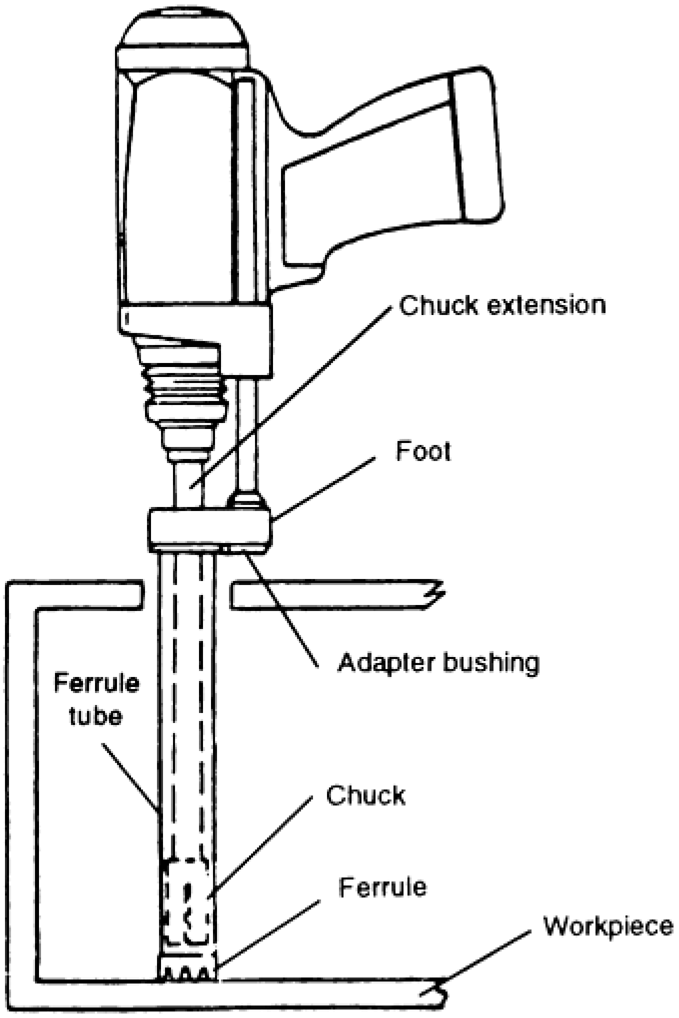
Chuck extension
Foot
Adapter bushing
Ferrule
tube
Chuck
Ferrule
Workpiece

FIG. 13 PORTABLE STUD ARC WELDING GUN EQUIPPED WITH ACCESSORIES FOR WELDING STUDS THROUGH

OR INTO A DRILLED OR FORMED HOLE

With closely spaced studs, the foot grip is sometimes modified so that after welding the initial stud at a center-punched location, the foot can be held against the reference stud to weld the next or adjacent stud on the required spacing.

The verticality of the weld stud can be ensured by several methods. The most accurate is obviously a fixed gun on a slide mechanism of the type used for automatic-feed, high-production work. This type is also the most expensive. A portable gun can be mounted on a machine slide or drill press by fabricating a bracket (Fig. 14) in cases where production quantities do not justify the expense of automatic equipment.

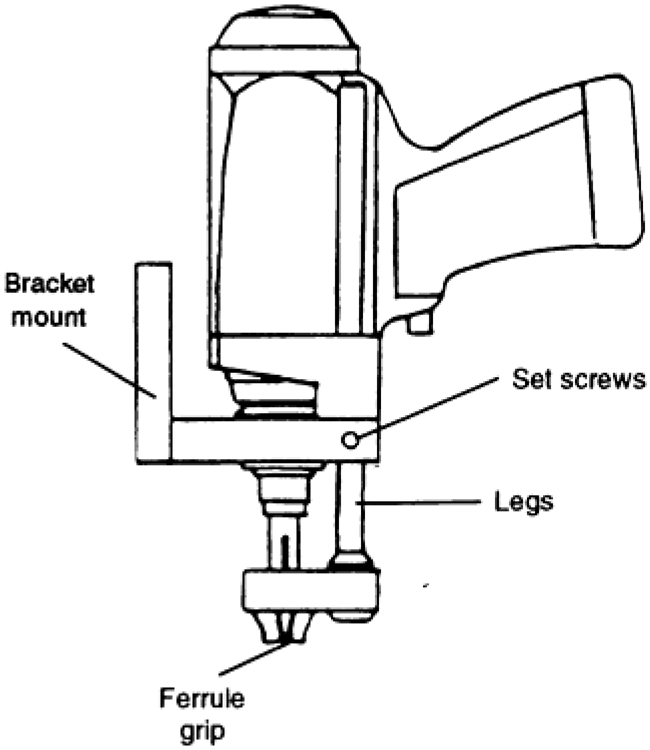

FIG. 14 INEXPENSIVE METHOD OF MOUNTING PORTABLE HAND GUN TO MACHINE SLIDE OR DRILL PRESS TO OBTAIN FIXED-POSITION ACCURACY

Other methods of ensuring verticality are to use a bushing-type template, a bubble level mounted on the rear of the gun, or a bipod foot arrangement, as shown in Fig. 15. With the stud and ferrule acting as one fixed point, the two bipod screws are adjusted to obtain perpendicularity. Vertical variations that are less than ±1° can be achieved with proper accessories and adjustments.

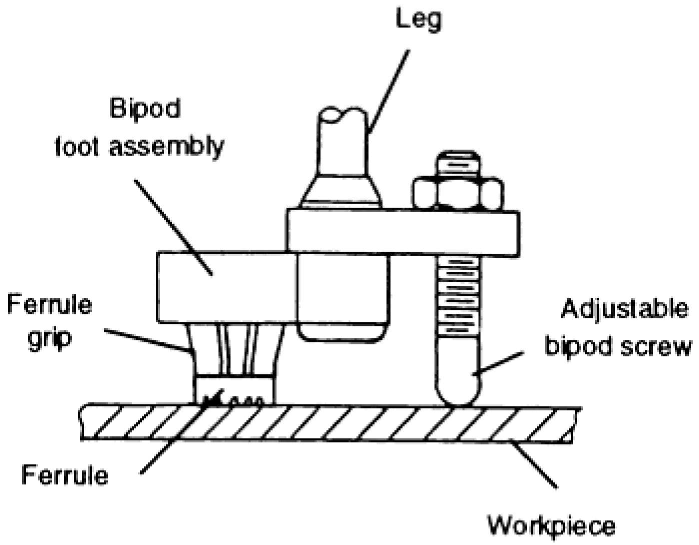
Leg
Bipod foot assembly
Ferrule grip
Adjustable bipod screw
Ferrule
Workpiece

FIG. 15 USE OF AN ADJUSTABLE BIPOD FOOT ASSEMBLY TO PROVIDE THREE-POINT CONTACT, ENSURING PERPENDICULARITY OF STUD TO WORKPIECE

A wide range of accessories is available for welding studs of different lengths or for welding studs into areas with limited access. Two widely used special accessories are shown in Fig. 16 and 17. For unusual situations, the manufacturer should be consulted to suggest or, possibly, design stud-welding accessories.

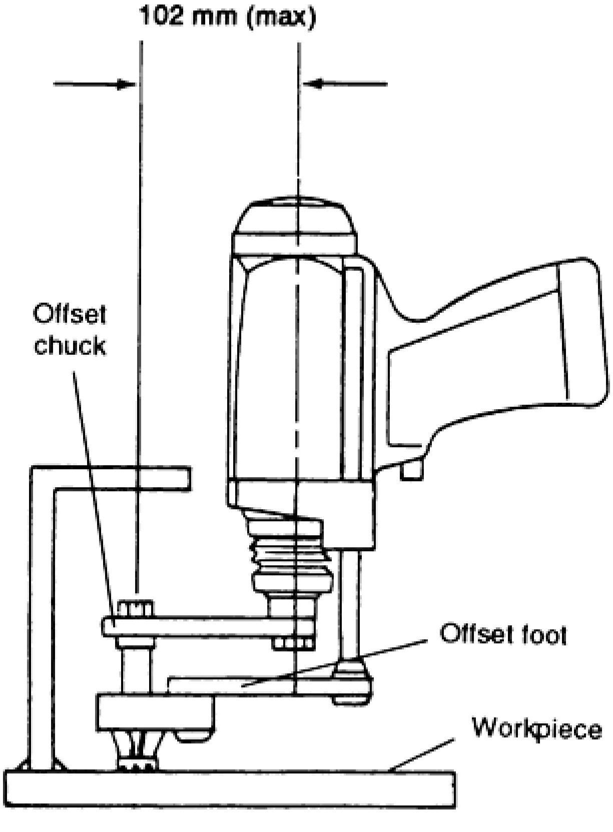
102 mm (max)
Offset chuck
Offset foot
Workpiece

FIG. 16 TYPICAL OFFSET ASSEMBLY USED WHEN STUD ARC WELDED FASTENERS MUST BE WELDED BELOW AN OBSTRUCTION

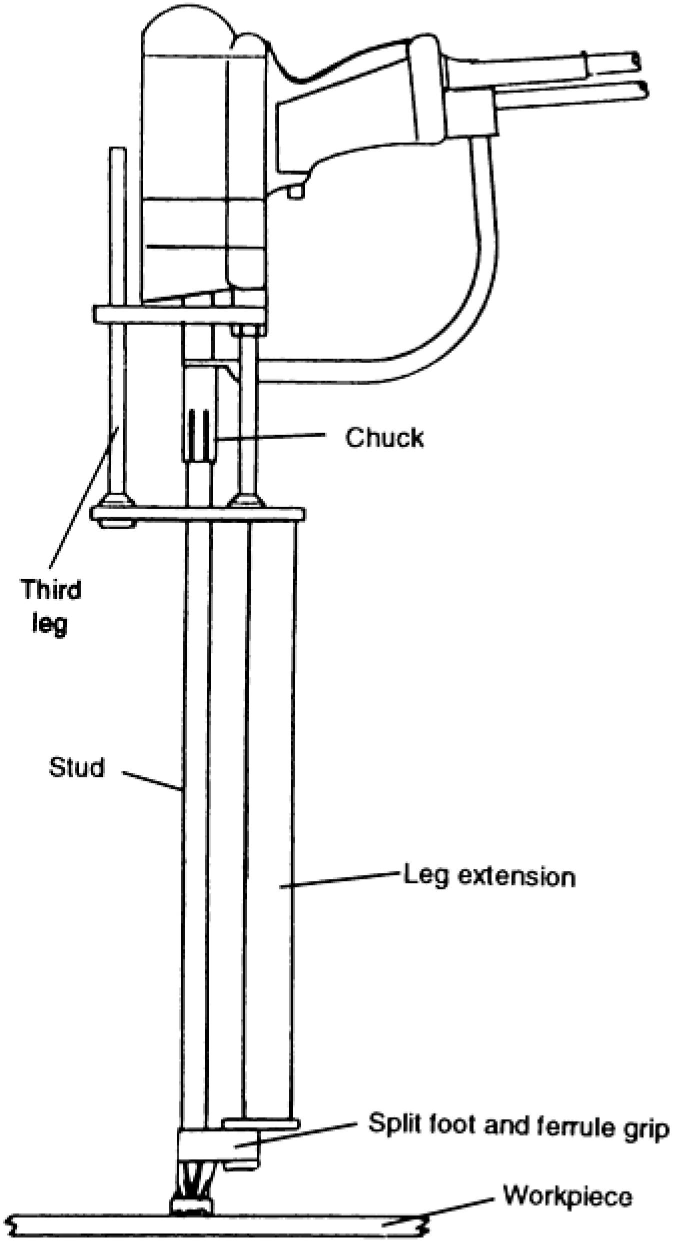

FIG. 17 TYPICAL LEG-EXTENSION ACCESSORY FOR WELDING LONG STUDS (UP TO 1.83 M, OR 6 FT). NOTE THE USE OF A SPLIT FOOT AND FERRULE GRIP; THE GUN DOES NOT HAVE TO BE STRIPPED THE FULL LENGTH OF THE STUD WHEN THE STUD IS WELDED.

Stud Arc Welding

Harry A. Chambers, TRW Nelson Stud Welding Division

Stud-Welding Quality Control, Qualification, and Inspection

Because of its broad use for many years and the relative simplicity of its applications, stud welding in the downhand position is considered to be a prequalified procedure, that is, it is only necessary to weld two studs satisfactorily in the downhand position to qualify both the process and the operator. The two qualification studs can be bend tested, torque tested, or tension tested. The physical test is satisfactory if no failure occurs in the weld or the heat-affected zone.

Stud-welding code requirements (for example, Ref 2 and 7) define the necessary tests for prequalified stud welding to the flat or down-hand position, as well as other positions. The limit on the flat position is defined as 0 to 15° slope on the surface to which the stud is applied. Beyond 15°, a ten-stud test is required. Again, each of the ten studs in the nonflat position must be bend, tensile, or torque tested. Bend testing is 90° from the original axis, whereas torque and tension testing are conducted until failure. Failure must occur in the stud shank or in the workpiece material, not in the weld or the heat-affected zone. All failures require retesting. Satisfactory testing qualifies the welder and the process.

Typical bend, torque, and tensile testing setups are shown in Fig. 18, 19, and 20. Qualification records, including stud drawings, ceramic arc shield drawings (if used), weld settings, and position should be kept.

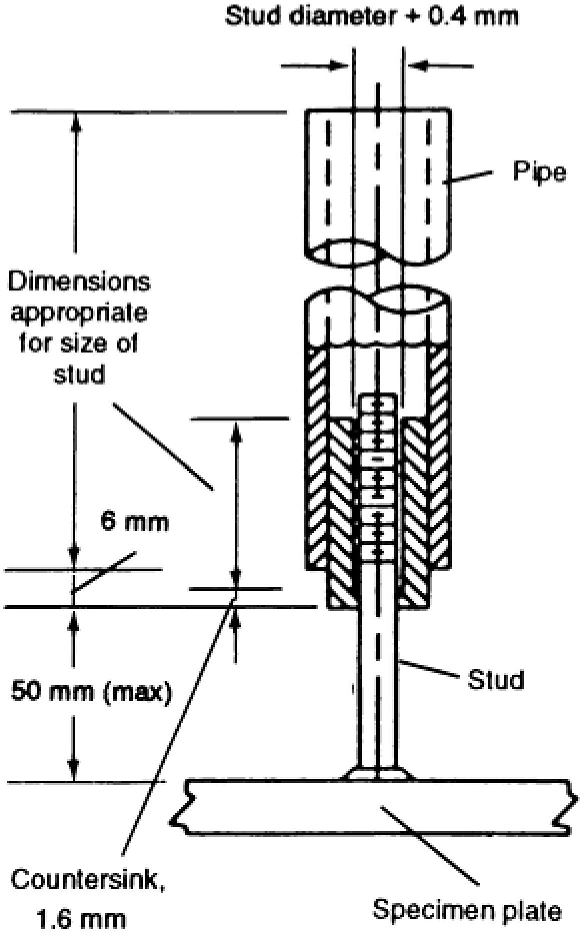
Stud diameter + 0.4 mm
Pipe
Dimensions appropriate for size of stud
6 mm
Stud
50 mm (max)
Countersink, 1.6 mm
Specimen plate

FIG. 18 TYPICAL BEND TESTING SETUP. STUDS THAT MUST UNDERGO APPLICATION QUALIFICATION SHOULD BE BENT 90° WITHOUT WELD FAILURE (15° FOR ALUMINUM STUDS). TESTING FOR PREPRODUCTION OR TESTING DURING PRODUCTION FOR INSPECTION PURPOSES REQUIRES BEND TESTING TO 30° FROM ORIGINAL POSITION (15° FOR ALUMINUM STUDS). FOR THREADED STUDS, TORQUE TESTING PER FIG. 19 SHOULD BE USED. SOURCE: REF 2

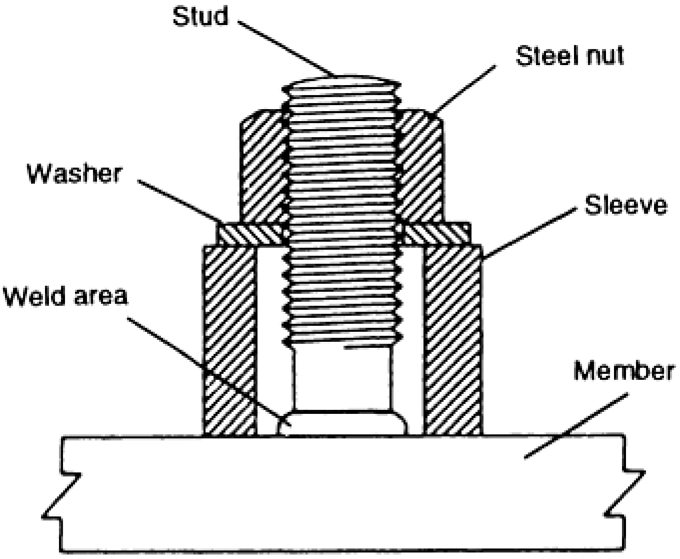
Stud
Steel nut
Washer
Sleeve
Weld area
Member

FIG. 19 TYPICAL TORQUE TESTING ARRANGEMENT. STUDS THAT MUST UNDERGO APPLICATION QUALIFICATION OR PREPRODUCTION TESTING SHOULD BE TORQUED TO DESTRUCTION WITHOUT FAILURE IN THE WELD. PROOF TORQUE TESTING DURING INSPECTION OF THREADED STUDS SHOULD BE APPLIED AT 60 TO 66% OF THE YIELD TORQUE LOAD FOR THE STUD SIZE TESTING. REFER TO TABLES 7, 8, AND 9. SOURCE: REF 2

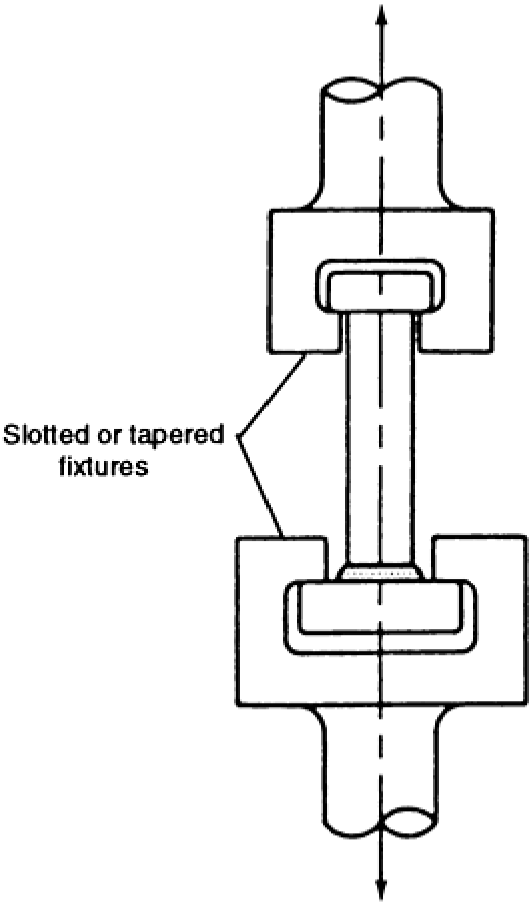
Slotted or tapered fixtures

FIG. 20 TYPICAL TENSION TEST FIXTURE. FOR QUALIFICATION OF APPLICATION OR PREPRODUCTION TESTING, EITHER BEND OR TORQUE TESTING CAN BE SUBSTITUTED FOR TENSION TESTING.

Either two-stud or ten-stud testing should be conducted on a production piece or on a separate piece of the same type and thickness (±25%) of material that will be used in the application, in the same position in which the studs will actually be applied.

In addition to physical testing, the studs that are welded for qualification also must be visually inspected, whether there are two or ten. Visual inspection consists of verifying that a full 360 ° weld flash (fillet) is formed around the stud base. There should be no gaps or undercuts in the flash or in the wetted area. As mentioned previously, flash can be irregular in both height and evenness and it can contain shrinkage fissures or cracks. Because flash is expelled, not deposited, metal, it should not be subject to the usual inspection criteria for a weld fillet.

The final visual inspection should be a measurement of the after-weld length. With stud arc welding, the after-weld length of the stud will be shorter by the lengths shown in Table 3. A consistent after-weld length, combined with acceptable visual inspection and physical testing, are assurances of satisfactory welds. Figure 21 shows typical stud arc weld appearances for both acceptable and problematic welds.

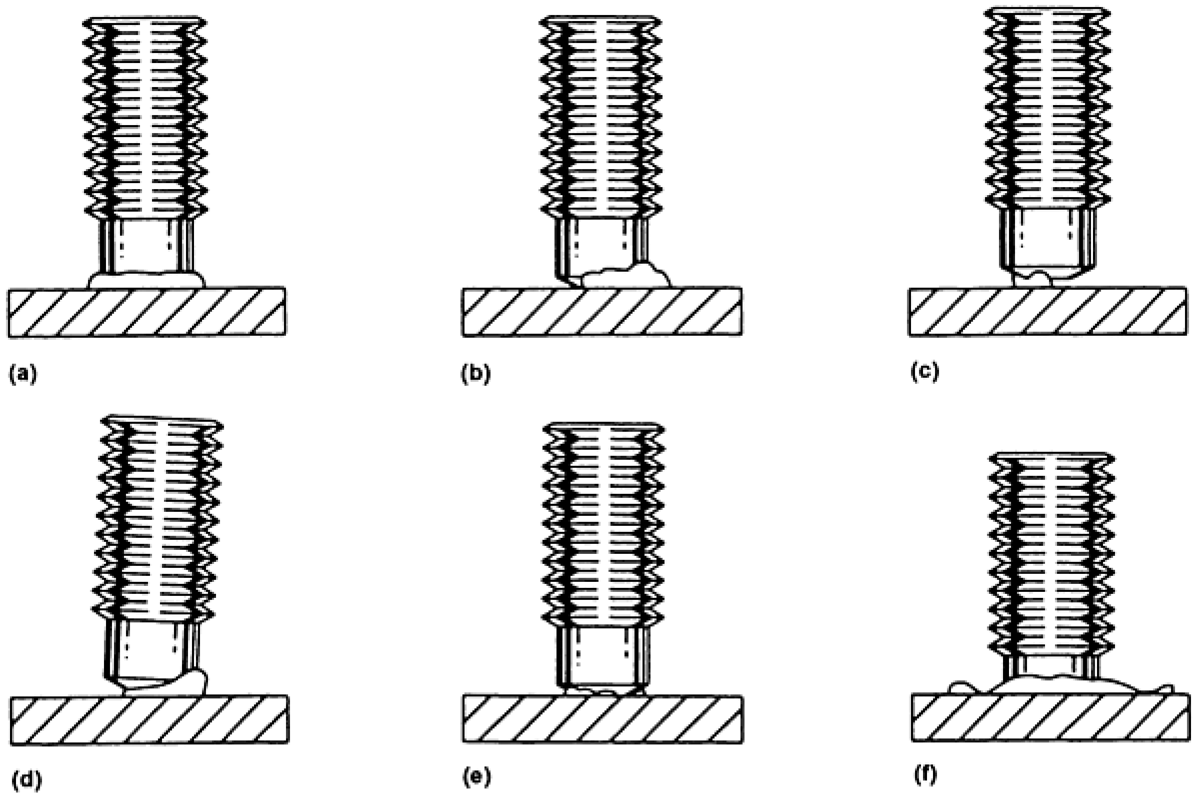

FIG. 21 STUD ARC WELD CHARACTERISTICS FOR SATISFACTORY AND UNSATISFACTORY WELDS. (A) SATISFACTORY; GOOD FLASH FORMATION, WETTING ENTIRE PERIPHERY OF STUD BASE. (B) UNSATISFACTORY; PLUNGE TOO SHORT, NOT ENOUGH MATERIAL ALLOWED FOR BURN-OFF. STUD MUST PROJECT AN ADEQUATE LENGTH BEYOND THE FERRULE. THIS CONDITION CAN ALSO BE CAUSED BY BAD GROUNDING PRACTICE. (C) UNSATISFACTORY; HANG UP CAUSED BY INADEQUATE ARC LENGTH LIFT OR BY MISALIGNMENT OF STUD, RESTRICTING FREE MOVEMENT. (D) UNSATISFACTORY; MISALIGNMENT OF STUD BY POOR GUN POSITION. (E) UNSATISFACTORY; COLD WELD, RESULTING IN NONUNIFORM OR INCOMPLETE FLASH AROUND STUD PERIPHERY. (F) UNSATISFACTORY; HOT WELD, WHERE MOLTEN MATERIAL IS EXPELLED VIOLENTLY FROM WELD, RESULTING IN POOR FLASH FORMATION.

Passing the two-stud preproduction or the ten-stud application qualification test approves the procedure and operator. Production stud welding is then allowed to proceed. However, at the start of each new production period, a two-stud test is again required before production can begin. Similarly, a change in stud diameter, equipment, settings, or operator requires requalification.

During production, welding inspection and testing should be used on a continuous basis to confirm acceptable quality. This can be done on a reasonable number of studs, depending on the welding conditions and stud appearance. After the visual inspection of all stud welds, physical testing of at least one weld per hundred is usually suggested. Any weld that does not show a full 360° flash or wetness should be tested. Welds without a full flash can be repair welded, but repaired welds are also required to be physically tested.

In the case of unthreaded studs, the quality inspection test can be to bend 15°. The studs can be straightened after testing, if required. Threaded studs can be subjected to a torque test to approximately 80% of the yield torque load, rather than a bend test. Failure in the weld or heat-affected zone is cause to reject the weld being tested. Failure also requires that the quantity of studs welded from the previous test to the present test be carefully inspected for weld deficiencies. Additional tests may be necessary to confirm weld quality.

If the quality inspection test studs must be straightened for end use, then they should be returned to the vertical position by using a bending tool. Continuous slow pressure should be applied during the straightening process.

Temperature at the time of stud welding and during stud testing is an influential parameter. Stud welding should not be conducted at temperatures below -18 °C (0 °F). When the temperature is below 0 °C (32 °F), one additional stud per hundred welded should be tested to verify weld quality. Below 10 °C (50 °F), welded studs should be tested using the bending tool, rather than by hammer testing. In colder welding temperature conditions, threaded studs should always be tested by torque, rather than by bending. Impact testing at low temperatures produces brittle failures. This is not a function of weld quality, but of the ambient temperature at the time of welding and testing. Data resulting from extensive tests conducted on studs welded and tested at various temperatures have shown that it is the temperature and manner of testing that influences the failure mode, rather than the temperature at the time of welding (Ref 8).

Another factor, besides temperature, that can affect weld quality is cleanliness of the weld area. Although stud arc welding does provide some cleaning action during welding, it is not sufficient to remove heavy contaminants. Coatings or rust on studs also can cause poor welds. Plating, except for a copper flash, nickel plating, or chromium plating, should be removed from the stud end prior to welding. This includes such materials as paint, zinc, and cadmium. Stud arc welded fasteners are usually capped during plating so that the weld end is left clean.

The plating or contaminant removal methods should be compatible with the base-material thickness and composition. For example, a heavy paint or rust cover on thick structural steel can be removed by grinding, needle scaling, sand blasting, or chipping, and the weld surface will still be acceptable for stud welding. Stainless steel, brass, aluminum, or other polished surfaces should be cleaned with appropriate solvents or other cleaning methods that will not destroy the material finish or contaminate the weld area. For example, a carbon-steel wire wheel or brush should not be used to clean stainless steel or aluminum base material.

Studs and ferrules should be kept clean and dry prior to welding. Although cold conditions will not adversely affect either, they should be protected from exposure to adverse weather conditions, high humidity, marine atmospheres, and other extremes. When ferrules absorb water, the steam generated during the welding process can cause very poor welds and excessive weld spatter. Ferrules exposed to moisture should be oven dried at a temperature between 101 and 120 °C (215 and 250 °F) for 2 h before use.

The condition of weld cables, connections, and ground clamps can affect weld quality. These items should be checked frequently and either repaired or replaced, so that the cables are not frayed, connections are tight and clean, and ground clamps are properly functioning. Ground clamps should be tightly connected to the workpiece or work platen and attached to areas that have been cleaned of all deleterious materials that would prevent good current flow.

References cited in this section

2. "STRUCTURAL WELDING CODE--STEEL," ANSI/AWS D1.1, SECTION 7, "STUD WELDING," AWS
7. "WELDED STEEL CONSTRUCTION (METAL ARC WELDING), STANDARD W59," CANADIAN STANDARDS ASSOCIATION
8. KENNEDY AND D.J. LAURIE, STUD WELDING AT LOW TEMPERATURES, *CAN. J. CIVIL ENG.*, VOL 7, 1980, P 442-455

Harry A. Chambers, TRW Nelson Stud Welding Division

Stud-Welding Safety Precautions

Any welding process can be dangerous if proper safety precautions are not followed. Equipment should be properly installed according to the directions of the manufacturer and maintained in good condition. Operators should be thoroughly trained in the use of the stud-welding process. They also should be familiar with the installation, operation, and maintenance procedures for the equipment being used.

Electrical shock is a potential cause of injury or death. Stud-welding systems should be properly installed and grounded using electrical connections made according to national (Ref 9) and local code guidelines.

Neither operators nor anyone else in the welding area should ever touch live electrical parts, or weld in wet areas, or wear wet gloves or clothing. In addition, all workers should ensure that they are insulated from shock.

All electric cables and connections should be kept in good condition. They should be inspected on a regular basis and frayed sections, broken insulation, or broken connectors should be repaired or replaced at once. Stud welding frequently uses long cable lengths, particularly on construction sites, where the cables are subject to abuse by towmotors, trucks, material storage, or foot traffic. It should be ensured that these conditions do not result in cable damage.

Although weld spatter and arc flash are minimal with the stud-welding process, they do occur and precautions should be taken. All combustible or volatile materials should be removed from the weld area so that sparks or spatter cannot reach them. Gas cylinders should be stored and secured properly and checked to ensure that they do not contact any electrical cables in the welding circuit. Proper protective clothing should be worn, including boots, aprons, and gloves, as necessary. Eye protection is always necessary. Eye glasses with spectacle frames, side shields, and lenses with shade number three absorption filters should be worn by the operator at all times (Ref 10). Helpers or workers within 1.5 m (5 ft) of the weld area also should wear clear safety glasses with side shields. Fire-suppression equipment should be available in or adjacent to the weld area for immediate use in emergencies.

Ventilation of the welding area is necessary. Fumes from welding and cleaning materials, such as solvents, as well as paints, epoxies, and galvanizing or other coatings, can be harmful. Ventilation can be either forced or natural, depending on the job conditions. Material suppliers should provide material safety data sheets on any items being used in the welding area, and proper precautions for potentially dangerous contents should be followed.

Pinch points where fingers or hands can be caught should be avoided during material handling or where moving parts are involved in the welding process.

The instructions of the manufacturer should be followed when maintaining and servicing the stud-welding equipment. If possible, all electric supplies should be disconnected and locked out during maintenance or troubleshooting work.

Finally, certain welding operations produce elevated noise levels during the weld cycle. Operators and other workers near the equipment should use hearing protection that can adequately protect them, in accordance with Ref 11.

References cited in this section

9. "NATIONAL ELECTRICAL CODE" NFPA-70, NATIONAL FIRE PROTECTION ASSOCIATION
10. "SAFE PRACTICE FOR OCCUPATION AND EDUCATIONAL EYE AND FACE PROTECTION," Z87.1, ANSI
11. "WELDING, CUTTING, AND BRAZING," STANDARD 29CFR, PART 1910, OCCUPATIONAL SAFETY AND HEALTH ADMINISTRATION

Harry A. Chambers, TRW Nelson Stud Welding Division

References

1. "RECOMMENDED PRACTICES FOR STUD WELDING," ANSI/AWS C5.4, AWS
2. "STRUCTURAL WELDING CODE--STEEL," ANSI/AWS D1.1, SECTION 7, "STUD WELDING," AWS
3. "STRUCTURAL WELDING CODE--ALUMINUM," ANSI/AWS D1.2, AWS
4. "STRUCTURAL WELDING CODE--STEEL," ANSI/AWS D1.1, SECTION 7, "STUD WELDING," FOOTNOTE 27, AWS
5. "STRUCTURAL WELDING CODE--STEEL," ANSI/AWS D1.1, SECTION 7, "STUD WELDING," PARAGRAPH 7.6, AWS
6. W.A. BAESLACK, G. FAYER, S. REAM, AND C.E. JACKSON, QUALITY CONTROL IN ARC STUD WELDING, *WELD. J.,* NOV 1975, P 789-798
7. "WELDED STEEL CONSTRUCTION (METAL ARC WELDING), STANDARD W59," CANADIAN STANDARDS ASSOCIATION
8. KENNEDY AND D.J. LAURIE, STUD WELDING AT LOW TEMPERATURES, *CAN. J. CIVIL ENG.*, VOL 7, 1980, P 442-455
9. "NATIONAL ELECTRICAL CODE" NFPA-70, NATIONAL FIRE PROTECTION ASSOCIATION
10. "SAFE PRACTICE FOR OCCUPATION AND EDUCATIONAL EYE AND FACE PROTECTION," Z87.1, ANSI
11. "WELDING, CUTTING, AND BRAZING," STANDARD 29CFR, PART 1910, OCCUPATIONAL SAFETY AND HEALTH ADMINISTRATION

Stud Arc Welding

Harry A. Chambers, TRW Nelson Stud Welding Division

Selected References

- "DESIGN AND APPLICATION HAND BOOK FOR LIGHT GAUGE METALWORKING," PUBLICATION CD-92, TRW NELSON STUD WELDING DIVISION, 1992
- "INDUSTRIAL DESIGN DATA--NELSON STUD WELDING PROCESS," PUBLICATION MB-12-86, TRW NELSON STUD WELDING DIVISION
- "SAFETY IN WELDING AND CUTTING," ANSI/AWS Z49.1, AWS
- T.E. SHOUP, "STUD WELDING," BULLETIN 214, WELDING RESEARCH COUNCIL, APRIL 1976
- "TOOLING TECHNIQUES FOR STUD WELDING," TRW NELSON STUD WELDING DIVISION
- STUD WELDING, CHAPTER 9, *WELDING HANDBOOK,* 8TH ED., AWS, 1990

Capacitor Discharge Stud Welding

Richard L. Alley, American Welding Society

Introduction

CAPACITOR DISCHARGE STUD WELDING is a stud arc welding process in which the tip of the stud melts almost instantly when energy stored in capacitors is discharged through it. The three basic modes of capacitor discharge (CD) stud welding are initial-gap welding, initial-contact welding, and drawn-arc welding. Figure 1 shows current-versus-time curves for the three process variations.

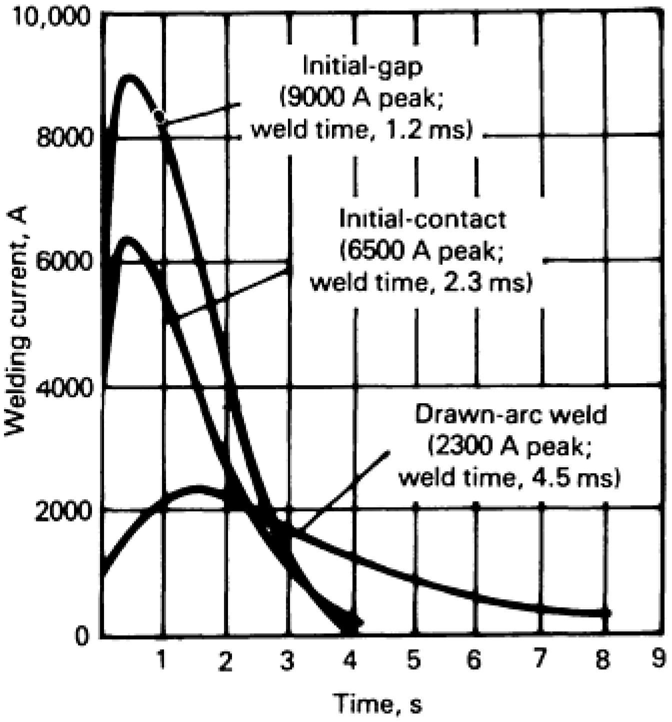

FIG. 1 TYPICAL CURRENT-VERSUS-TIME CURVES FOR THE THREE CD STUD WELDING METHODS. SOURCE: REF 1

The initial-gap mode (Fig. 2) is begun with the stud held away from the work surface by the welding head. At the beginning of the weld cycle, the stud is forced against the work surface, melting the tip, the face of the stud, and the adjoining work surface upon contact with the work surface. The weld is completed using the gun forces (i.e., spring pressure or air pressure) to plunge the stud into the molten materials, forming a strong welded bond between the stud and the work surface. The weld cycle time (Fig. 1) for this process is from 4 to 6 ms, and the penetration of the weld zone into the work surface is normally from 0.10 to 0.15 mm (0.004 to 0.006 in.).

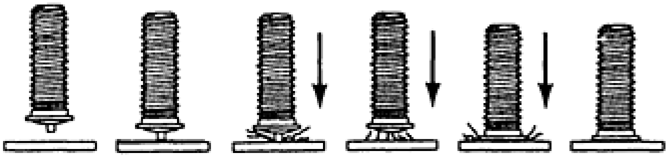

FIG. 2 INITIAL-GAP CD STUD WELDING. SEE TEXT FOR EXPLANATION. SOURCE: REF 1

The initial-contact mode (Fig. 3) begins with the weld stud in contact with the work surface. The weld cycle is initiated with a surge of current that disintegrates the weld tip, thus melting the stud face area and the work surface that it immediately contacts. The stud is forced into the molten material, forming a strong homogeneous weld. This process has a weld cycle time of approximately 6 ms, much like the gap process.

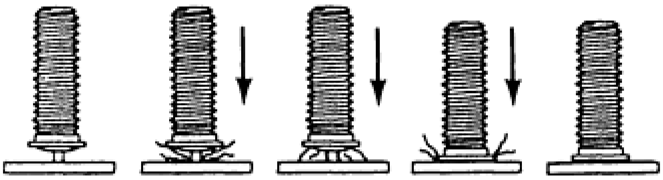

FIG. 3 INITIAL-CONTACT CD STUD WELDING. SEE TEXT FOR EXPLANATION. SOURCE: REF 1

The drawn-arc mode (Fig. 4) begins with the stud in contact with the work surface. When the weld cycle is initiated, a current surge is applied to the weld tip and the stud is retracted from the work surface, drawing an arc that melts the face of the stud and the work surface directly beneath it. The stud is then plunged into the molten pool of material, forming a welded connection. The weld cycle time for this process is longer than for the other two processes, and the heat-affected zone (HAZ) is thicker than it is in the preceding two processes.

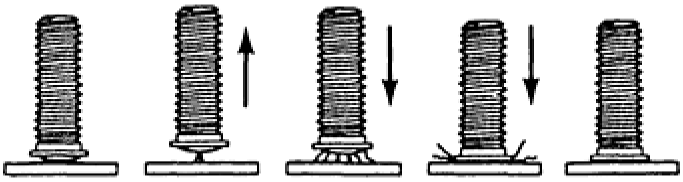

FIG. 4 DRAWN-ARC CD STUD WELDING. SEE TEXT FOR EXPLANATION. SOURCE: REF 1

Advantages and Disadvantages. A major reason for using the CD stud welding process is that it provides a strong welded fastener with either a minimum or no reverse side marking of the work material. Another reason is that it is cost effective, especially for small-diameter fasteners. This process also allows fasteners to be welded to very thin sections of work material, as well as to thick sections (as thick as necessary), with reliable results. Furthermore, the process allows the welding of many dissimilar material combinations, such as aluminum studs to zinc die castings. The disadvantages of using the CD process are the limited stud diameters available and the fact that the work surface must be clean of mill scale, dirt, oxidation products, and oil.

Reference

1. R.L. O'BRIEN, ED., *WELDING HANDBOOK,* 8TH ED., VOL 2, AWS, 1991, P 317-323

Capacitor Discharge Stud Welding

Richard L. Alley, American Welding Society

Applications

The CD process is normally used for fastener sizes of 4-40 through $\frac{1}{4}$-20, with lengths from 6.4 to 38.1 mm ($\frac{1}{4}$ to 1.5 in.). Although there are limited uses of $\frac{5}{16}$-18 and $\frac{3}{8}$-16 diameter studs in special applications, only the more-common stud diameters will be considered in this article. The work material thickness can typically range from 0.25 mm (0.010 in.) to as thick as the application requires. The determining factors for making successful welds are the stud diameter, the materials being used, the strength required, and the amount of allowable reverse side marking. See Table 1 for guidance on the expected strength parameters.

TABLE 1 STANDARD CAPACITOR DISCHARGE STUD FASTENER LOAD STRENGTHS

STUD MATERIAL	STUD SIZE	MAXIMUM FASTENING TORQUE$^{(A)}$		ULTIMATE TENSILE LOAD		MAXIMUM SHEAR LOAD	
		n · m	**lbf · in.**	**kn**	**lbf**	**kn**	**lbf**
LOW-CARBON COPPER-FLASHED STEEL	6-32	0.7	6	2.2	500	1.7	375
	8-32	1.4	12	3.4	765	2.6	575
	10-24	1.6	14	4.3	960	3.2	720
	$\frac{1}{4}$-20	4.9	43	7.8	1750	5.8	1300
	$\frac{5}{16}$-18	8.1	72	12.9	2900	9.8	2200
	$\frac{3}{8}$-16	11.9	106	19.1	4300	14.6	3250
STAINLESS STEEL (TYPE 304)	6-32	1.1	10	3.5	790	2.6	590
	8-32	2.3	20	5.6	1260	4.1	940
	10-24	2.6	23	6.8	1530	5.2	1150
	$\frac{1}{4}$-20	8.5	75	12.8	2880	9.6	2160
	$\frac{5}{16}$-18	14.2	126	16.7	3750	12.5	2800
	$\frac{3}{8}$-16	20.0	186	21.6	4850	16.0	3600
ALUMINUM ALLOY (5000 SERIES)	6-32	0.40	3.5	1.7	375	1.0	235
	8-32	0.85	7.5	2.6	585	1.6	365
	10-24	1.13	10	3.3	735	2.0	460
	$\frac{1}{4}$-20	4.5	40	6.0	1360	3.8	850
	$\frac{5}{16}$-18	7.9	70	10.2	2300	6.2	1400
	$\frac{3}{8}$-16	9.1	81	15.1	3400	9.3	2100
	$\frac{1}{2}$-13	15.8	140	25.0	5500	13.4	3000
BRASS (70-30 AND 65-35)	6-32	0.9	8	2.7	600	1.7	390
	8-32	1.8	16	3.8	860	2.5	560
	10-24	2.09	18.5	4.6	1040	3.0	680
	$\frac{1}{4}$-20	6.9	61	8.7	1950	5.7	1275
	$\frac{5}{16}$-18	11.5	102	14.6	3280	9.5	2140
	$\frac{3}{8}$-16	16.9	150	21.4	4800	14.1	3160

(A) THESE VALUES SHOULD DEVELOP FASTENER TENSION TO SLIGHTLY LESS THAN YIELD POINT AND SHOULD BE USED ONLY AS A GUIDE; BASIC SPECIFICATIONS SHOWN ABOVE COVER TYPICAL MECHANICAL PROPERTY VALUES.

The stud and workpiece materials can be common low-carbon steel, stainless steel, or aluminum. Also used are medium-carbon steel, lead-free brass and copper, Inconel, titanium, René 41, zirconium, gold, silver, and platinum. Table 2 shows some of the more popular combinations of materials. Various studs and plate materials can be combined, except when aluminum-base material is used.

TABLE 2 WELDING CAPABILITIES OF CAPACITOR DISCHARGE STUD FASTENERS

BASE MATERIAL	STUD MATERIAL			
	MILD STEEL (TYPES 1008, 1010)	STAINLESS STEEL (TYPES 304, 305)	ALUMINUM (5356, 6061)	BRASS (70-30, 65-35)
MILD STEEL (1006 THROUGH 1030)	EXCELLENT	EXCELLENT	. . .	EXCELLENT
MEDIUM-CARBON STEEL (1030 THROUGH 1050)	GOOD(A)	GOOD(A)	. . .	GOOD(A)
GALVANIZED SHEET DUCT OR DECKING	GOOD(A)	GOOD(A)	. . .	. . .
STRUCTURAL STEEL	GOOD(A)	GOOD(A)	. . .	EXCELLENT
STAINLESS STEEL (TYPES 405, 410, 430, AND 300 SERIES, EXCEPT 303)	EXCELLENT	EXCELLENT	. . .	EXCELLENT
LEAD-FREE BRASS, ELECTROLYTIC COPPER, LEAD-FREE ROLLED COPPER	EXCELLENT	EXCELLENT	. . .	EXCELLENT
MOST ALUMINUM ALLOYS OF THE 1000, 3000, 5000, AND 6000 SERIES(B)	. . .	. . .	EXCELLENT	. . .
DIE-CAST ZINC ALLOYS	GOOD(A)	GOOD(A)	EXCELLENT	GOOD(A)

(A) GENERALLY FULL-STRENGTH RESULTS, DEPENDING ON THE COMBINATION OF STUD SIZE AND BASE METAL.

(B) OTHER MATERIALS, SUCH AS 7000 SERIES ALUMINUM, TITANIUM ALLOYS, INCONEL, AND SO ON, CAN BE WELDED UNDER SPECIFIED CONDITIONS.

CD stud welding is used in many industries in a large variety of applications because it is one of the most versatile and reliable processes available. The food processing industry utilizes this process on stainless steel utensils and machinery, aluminum cookware, and appliances. The aerospace industry applies the process to aircraft engines, aircraft parts and instrument areas, and components for outer space vehicles. The appliance industry uses CD welding for its cabinets, trims, and other components. The building industry uses it for doors, facia, subassemblies, and other exposed finished areas. The insulation industry uses it in securing insulation to heating and air conditioning duct work, as well as on apparatus at power plants, schools, and other major buildings. The automotive industry uses the process to secure trim and various components to cars and trucks. The ship-building industry uses it in fabricating components in ship kitchens, in insulation applications aboard ship and in other areas where components and instruments must be secured. The electrical industry uses this process when securing components inside various controllers and equipment, as well as in trim parts for cabinets. The sign industry uses it on both structural and component members of large and small signs.

Capacitor Discharge Stud Welding

Richard L. Alley, American Welding Society

Equipment

The equipment used for CD stud welding can be either portable or stationary production-type units. The portable units consist of the basic controller that utilizes standard 110 V alternating current (ac) power input to charge the capacitors and a lightweight gun-shaped tool used for placing and welding the fasteners. The production-type units typically require

240/480 V ac three-phase incoming power and a compressed air supply. The production units are normally used for higher rates of productivity, close-tolerance work (down to ±0.13 mm, or 0.005 in.), critical reverse side marking requirements, automatic stud feeding, automatic feeding of the part to be welded, automatic location of the stud on the part, and exotic materials.

The production rate of portable equipment normally ranges from 4 to 6 studs/min, depending on the application. Material handling and other factors could cause this rate to be either slower or faster. The production rates for the permanent production equipment can range from 3 to 20 studs/min, depending on the application, special requirements, and tooling.

When the initial-gap or initial-contact modes of stud welding are used, welds are made on all materials without the use of shielding gases or ceramic arc shields. When using the drawn-arc mode while working with aluminum, a shielding gas would be necessary. This mode is also less flexible when working with exotic materials.

Personnel Responsibilities

The CD stud welding process is easy to operate and maintain. The operator must first ensure that his unit is connected to the proper incoming voltage. The operator then determines whether the weld gun is connected to the negative and the ground is connected to the positive. Straight polarity is used for welding in nearly all applications. The operator places a given diameter of fastener into the stud gun. The gun is then placed on the work surface. The operator closes the trigger of the gun and the welding process begins. After the welding process is completed, the operator lifts the gun off the welded stud.

The operator needs to inspect the welds to ensure that the settings were correct. The controller voltage, which is set according to the recommendations of the equipment manufacturer, depends on the material and diameter of the fastener to be welded, as well as on the thickness and type of base material being used.

There are also other parameters that should be considered. If it is desirable to reduce spatter and smoky conditions in the weld area, then wetting the weld area with a wetting agent just prior to making the weld is recommended. If a wetting agent is used, then it is best to add a few drops of liquid hand soap to a pint of plain water and to spray or brush the solution into the weld area. This will produce a clean weld with reduced, if not eliminated, spatter in the weld area. However, the noise of welding increases when a wet weld is made.

Weld inspection is normally done by bending the fastener by applying either a torque or tensile force to it. If a tensile force is applied, then care should be given to the diameter of the hole used to secure the plate material. A hole that is as small as possible should be used to accommodate the flange of the stud. If a large hole is used, the tensile factors may be reduced, because the plate material could yield prematurely. Prior to production welding, it is recommended that sample welds be made on similar materials and tested to destruction. During production, operators should visually inspect the welds periodically. If weld inconsistency occurs, the process should cease, and the cause should be determined and eliminated.

A common occurrence that contributes to inconsistency is movement of the plate material during the short weld cycle. This can occur when working with materials that are less than 3.2 mm (0.125 in.) thick. There should not be a gap between the workpiece and the backup material of the fixture or table being used. Proper fixtures and attention to their clamping system can alleviate this problem.

Reference

1. R.L. O'BRIEN, ED., *WELDING HANDBOOK,* 8TH ED., VOL 2, AWS, 1991, P 317-323

Capacitor Discharge Stud Welding

Richard L. Alley, American Welding Society

Selected References

- *RECOMMENDED PRACTICES FOR STUD WELDING,* C5.4-84, AWS, 1984

Plasma-MIG Welding

Ian D. Harris, Edison Welding Institute

Introduction

PLASMA-MIG WELDING can be defined as a combination of plasma arc welding (PAW) and gas-metal arc welding (GMAW) within a single torch, where a filler wire is fed through the plasma nozzle orifice. The process can be used for both welding and surfacing. The plasma-metal inert gas (MIG) welding process was invented at the Philips Research Laboratories in Eindhoven, Netherlands, around 1969 (Ref 1).

For descriptions of the PAW and GMAW processes, see the articles "Plasma Arc Welding," and "Gas-Metal Arc Welding" in this Section of the Handbook.

The principles of operation, in terms of equipment, are illustrated in Fig. 1. Separate power supplies are used for the PAW and the GMAW elements of the equipment. An arc is struck between the tungsten electrode and the workpiece in a similar fashion to that of a PAW system. The filler wire can be fed to the plasma arc, either with or without the GMAW arc established. Without power supplied to the filler wire, the system can be operated as a PAW system with concentric feed of filler wire. Later versions of the system incorporated an annular electrode to replace the offset tungsten electrode in the welding torch.

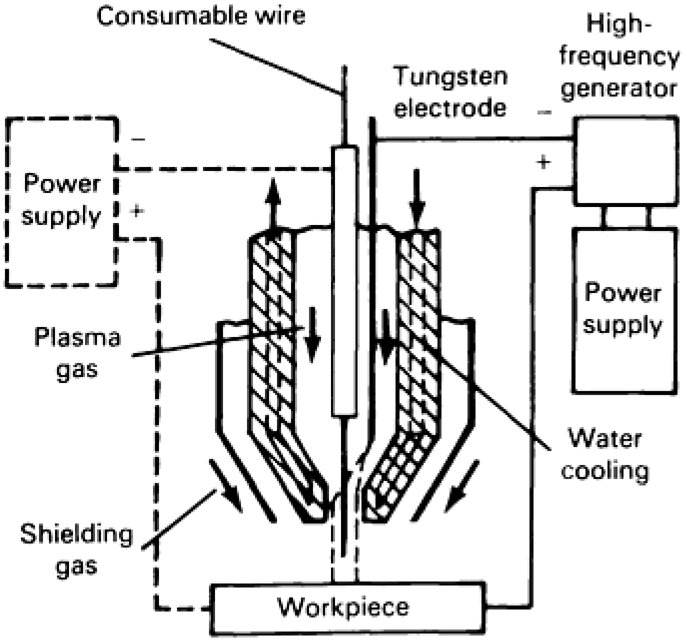

FIG. 1 SCHEMATIC OF PLASMA-MIG WELDING EQUIPMENT. SOURCE: REF 1

Current and Operating Modes. The equipment can be operated either with a single power source, effectively as a PAW system with concentric filler wire feed, or with two power sources, for the plasma-MIG operation.

The polarity of the tungsten electrode is direct current, electrode negative (DCEN), as is that of the GMAW part of the system. The heat of the plasma arc is sufficient to achieve good metal transfer stability for the GMAW element, despite the fact that when this process is used separately, it is almost always used in a direct-current, electrode positive (DCEP) mode. The filler wire is heated by the constricted plasma arc, as well as by the cathode heating of its own arc, and by resistance heating along the wire extension. Therefore, the melting and deposition rates of the wire are higher than the rates achieved by heating with either arc alone.

Metal transfer is governed not only by plasma streaming, but also by arc forces between the wire tip and the workpiece. Because the metal droplets are totally enclosed by the plasma stream, spray transfer takes place even though the GMAW element operates on negative polarity.

Advantages and Disadvantages. The advantages of the plasma-MIG process include deposition rates and joint completion rates that are higher than those of the conventional GMAW process. The independent control of the plasma arc and current to the filler wire leads to more control of metal deposition. This capability can yield improved productivity and good flexibility for controlling heat input and arc characteristics in both welding and surfacing operations. Good control of dilution is achieved by running the system without any power applied to the filler wire. Metal transfer stability is increased, compared to that of the conventional GMAW process, and results in lower spatter levels. The cleaning action of the plasma arc results in lower porosity in aluminum alloys, compared to that of the conventional GMAW process.

Disadvantages include the capital cost of two power sources (although there are systems that are designed to operate with one), the greater complexity of the torch, and the increased maintenance time and cost associated with this complexity. With two power sources, more welding parameters need to be set up, compared to the conventional GMAW process.

Equipment

As noted earlier, the basic equipment includes a power source for the plasma arc and a power source for the GMAW part of the system. A special torch incorporating both a contact tip for the GMAW element and a cathode for the PAW element is required. The initial design incorporated an offset tungsten electrode, as well as a concentric conduit and contact tip for the delivery of the consumable wire (Fig. 1). A later design incorporated a concentric cathode for the plasma arc (Fig. 2).

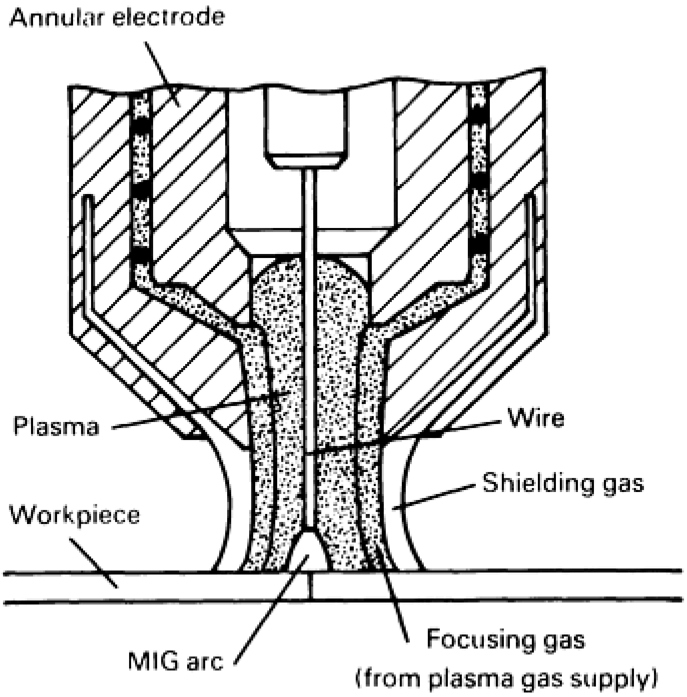

FIG. 2 SCHEMATIC OF MODERN PLASMA-MIG TORCH WITH ANNULAR PAW ELECTRODE AND ADDITIONAL (FOCUSING) GAS STREAM. SOURCE: REF 2

The plasma-MIG torch can be readily fitted to existing welding equipmentsuch as side beams and welding carriages, to replace the GMAW process in mechanized welding operations.

Power Sources. A constant-current power source with a high-frequency circuit to initiate the pilot arc is used for the plasma arc component of the system. The power source for the GMAW component can be used as a constant-voltage or a constant-current rectifier. Power sources have welding currents that typically range from 40 to 200 A for the plasma arc and from 60 to 300 A for the GMAW element at 100% duty cycle. However, equipment with welding currents up to 800 A is available and can be used for surfacing applications.

Welding Torches. A special torch with a concentric cathode for the plasma arc and a concentric conduit and contact tip for the delivery of the consumable wire is required (Fig. 2). A water-cooled copper alloy nozzle is used to constrict the arc and to form a collimated plasma jet that exits the nozzle orifice. A plasma orifice gas and a focusing gas from the same supply are used; the latter is delivered via channels between the plasma welding electrode and the constricting nozzle. The focusing gas results in greater arc constriction and arc stability and prolongs the life of the constricting nozzle by creating a boundary gas layer between the nozzle orifice and the plasma arc. The contact tip, plasma cathode, and the constricting nozzle are all directly water cooled to provide 100% duty cycle at the typical welding currents identified in the preceding section, "Power Sources."

Shielding Gases. Three shielding gases are utilized: one for the plasma (orifice) gas, one to provide additional arc constriction and arc stability, and one for supplementary shielding. The plasma gas and the focusing gas are usually argon, because an inert gas is required to prevent oxidation of the PAW electrode. The supplementary shielding gas can be argon, argon-oxygen, argon-carbon dioxide, or argon-hydrogen, depending on the nature of the workpiece being welded or, in the case of a surfacing operation, on the material being deposited. Argon is used when welding aluminum alloys, whereas argon-oxygen and argon-carbon dioxide are used when welding steels. Argon-hydrogen is used when welding stainless steels or when surfacing with them.

References

1. W.G. ESSERS, A.C.H.G. LIEFKENS, AND G.W. TICHELAAR, PLASMA-MIG WELDING, *PROC. CONF. ADVANCES IN WELDING PROCESSES,* THE WELDING INSTITUTE, 1971, P 216-219
2. J.D. SWART, PLASMA-MIG BOOSTS TANK TRAILER OUTPUT, *WELD. DES. FABR.,* FEB 1983, P 54-55, 59

Plasma-MIG Welding

Ian D. Harris, Edison Welding Institute

Procedure

Process Operating Procedure. The plasma arc is ignited using a pilot arc in a fashion similar to that of a PAW system. The main arc is transferred from the electrode to the workpiece and the plasma jet passes through the nozzle orifice. The system can be operated in this way, with the concentric cold wire being fed through the axis of the torch. A higher melting rate is achieved when power is applied to the wire through the contact tip and when both arcs are run simultaneously. The higher energy imparted to the wire by the plasma arc results in an increased wire deposition rate (Table 1). In this operating mode, deposition rates higher than those typical of the GMAW process can be achieved.

TABLE 1 COMPARISON OF FILLER WIRE MELTING RATES FOR PLASMA-MIG WELDING WITH AND WITHOUT THE GMAW ARC

Mild steel filler wire; plasma gas, argon; shielding gas, 89% Ar + 6% CO_2 + 5% O_2

PLASMA ARC CURRENT, A	PLASMA ARC VOLTAGE, V	FILLER WIRE CURRENT, A	MIG-ARC VOLTAGE, V	DIAMETER OF FILLER WIRE		MELTING RATE OF FILLER WIRE	
				mm	in.	g/min	oz/min
110	29	. . .	. . .	0.9	0.04	22	0.78
135	30	. . .	. . .	0.9	0.04	28	0.99
160	32	. . .	. . .	0.9	0.04	33	1.16
190	34	. . .	. . .	0.9	0.04	40	1.41
190	37	100	31	1.2	0.05	85	3.0
190	38	150	32	1.2	0.05	130	4.6

Source: Ref 1

Inspection and Weld Quality Control. Inspection requirements are similar to those of other arc welding or surfacing operations. Visual, ultrasonic, and radiographic inspection techniques are most appropriate. The dual action of the GMAW and plasma arcs results in weld quality that is sometimes higher than that achieved by the GMAW process alone. This is particularly true for aluminum alloys, because the cleaning action of the plasma arc often results in reduced porosity.

Quality control requires monitoring the welding parameters for both power sources, as well as monitoring the wire feed. In addition, the condition of the nozzle orifice (that is, the wear and concentricity of the orifice) should be monitored.

Troubleshooting. The relatively complex nature of the welding torch involves increased maintenance time. Erosion of the copper alloy nozzle orifice will cause a change in the arc shape and will affect the weld profile. Therefore, the nozzle should be checked periodically.

Reference cited in this section

1. W.G. ESSERS, A.C.H.G. LIEFKENS, AND G.W. TICHELAAR, PLASMA-MIG WELDING, *PROC. CONF. ADVANCES IN WELDING PROCESSES,* THE WELDING INSTITUTE, 1971, P 216-219

Plasma-MIG Welding

Ian D. Harris, Edison Welding Institute

Applications

Material Types. The plasma-MIG process is suitable for welding a wide variety of materials. The high heat energy supplied by the plasma and gas-metal arcs makes the process suitable for high-melting-point materials, such as tungsten and molybdenum. The most common application is welding aluminum sheet and plate. Wear-resistant steels are used with the process in hardfacing applications. Austenitic stainless steels, such as types 308, 309, and 347, as well as nickel alloys, such as alloy 625, are used in cladding applications. Both solid and flux-cored wires (Ref 3) can be employed for welding and surfacing, although most applications involve solid wires.

Industries. The plasma-MIG welding process has been used for the deposition of corrosion-resisting stainless steel and nickel-base alloys in the offshore industry, for the general fabrication of silos and tank trailers made from aluminum alloys (Ref 2), and for hardfacing applications in the excavation equipment industry, as well as the dredging and offshore industries.

Typical Components and Joints. The plasma-MIG process is suitable for welding of joints and for surfacing operations. The wide range of heat inputs available by choosing how to apply current to the consumable electrode provides additional flexibility for surfacing operations, compared to the range of heat inputs available with an external wire feed using the GTAW/PAW or the conventional GMAW process.

Single-V butt joints are commonly used for the plasma-MIG welding of plate. A full-penetration weld can be made in a single pass on a 9.5 mm ($\frac{3}{8}$ in.) thick mild steel plate when operating the plasma and gas-metal arcs simultaneously. This compares to three passes when just the PAW process is used. One would expect the same joint to require two or three passes for the conventional GMAW process.

When butt welding aluminum alloy sheet and plate in thicknesses ranging from 4.6 to 6.1 mm (0.18 to 0.24 in.), single-pass welds can be made at travel speeds that are more than twice as fast as when the GMAW process is used alone (Table 2). Lower porosity levels are achieved with the plasma-MIG process because of the cleaning action of the plasma arc. When using mild steel and stainless steel, the plasma-MIG process can make butt welds comparable to those made with the GMAW process alone at almost twice the travel speed (Table 2).

TABLE 2 COMPARISON OF WELDING SPEEDS FOR GMAW AND PLASMA-MIG WELDING OF ALUMINUM AND STEEL

PARAMETER	MATERIAL				
	ALMG4.5MN			STEEL	
	GMAW	PLASMA-MIG	PLASMA-MIG	GMAW	PLASMA-MIG
SHEET THICKNESS, MM (IN.)	4.6 (0.18)	4.6 (0.18)	5.1 (0.20)	4.1 (0.16)	4.1 (0.16)
WIRE DIAMETER, MM (IN.)	1.6 (0.063)	1.6 (0.063)	1.6 (0.063)	1.2 (0.047)	1.2 (0.047)
GMAW CURRENT, A	280-300	200	280	250	280
PLASMA CURRENT, A	. . .	200	200	. . .	200
WIRE FEED SPEED, M/MIN (IN./MIN)	8.1 (320)	10.9 (430)	10.9 (430)	5.1 (200)	11.4 (450)
WELDING SPEED, M/MIN	0.4 (16)	1.1 (43)	0.97 (38)	0.51 (20)	0.97(38)

(IN./MIN)						

Source: Ref 2

The hardfacing of dredging equipment with martensitic steel was made more productive by using the plasma-MIG process, rather than the conventional GMAW process. Higher deposition rates were achieved, resulting in reduced labor costs for the semiautomatic welding operation. Deposition rates of up to 20 kg/h (44 lb/h) can be achieved by plasma-MIG surfacing of mild steel with stainless steel flux-cored wires (Table 3).

TABLE 3 WELD SURFACING PARAMETERS FOR STAINLESS STEEL DEPOSITS ON MILD STEEL USING PLASMA-MIG WITH FLUX-CORED WIRES

	MIG WELDING CURRENT, A	MIG WELDING VOLTAGE, V	PLASMA WELDING CURRENT, A	PLASMA WELDING VOLTAGE, V	WELDING SPEED		DEPOSITION RATE		INNER WIRE STICK-OUT		DILUTION, %	FERRITE NO.
					mm/min	in./min	kg/h	lb/h	mm	in.		
UNDILUTED WELD METAL	300	39	150	44	110	4.4	10	22	35	1.4	. . .	15, 12, 13
PZ 6400 (347) SURFACING	200	31	150	40	80	3.2	6.6	14.6	55	2.2	8.6	12
PZ 6410 (308) SURFACING	300	44	150	46	140	5.6	13	28.7	55	2.2	13.8	5.5
PZ 6415 (309) SURFACING	400	50	150	51	190	7.6	20	44	55	2.2	9.7	6.5

Source: Ref 3

References cited in this section

2. J.D. SWART, PLASMA-MIG BOOSTS TANK TRAILER OUTPUT, *WELD. DES. FABR.,* FEB 1983, P 54-55, 59
3. F. EICHHORN AND E. VAN GAEVER, ADVANTAGEOUS SURFACING WITH PLASMA-MIG USING CORED WIRES, *PROC. 1ST INT. CONF. SURFACE ENGINEERING,* THE WELDING INSTITUTE, 1986, P 99-110

Plasma-MIG Welding

Ian D. Harris, Edison Welding Institute

Personnel

Skill Level and Training. The plasma-MIG process can be used for semiautomatic, mechanized, or automated operation. It requires a skill level comparable to that of the conventional GMAW process, although additional parameters need to be set for the plasma arc. Most welding and surfacing activities are carried out with mechanized equipment. When welding parameters are set, operation is similar to that of the GMAW process.

Health and safety issues are similar to those of other arc welding processes. Electric shock, eye protection, burns, ultraviolet radiation, and fume exposure are typical concerns. The plasma-MIG process generates more radiant heat than the conventional GMAW process.

Ian D. Harris, Edison Welding Institute

References

1. W.G. ESSERS, A.C.H.G. LIEFKENS, AND G.W. TICHELAAR, PLASMA-MIG WELDING, *PROC. CONF. ADVANCES IN WELDING PROCESSES,* THE WELDING INSTITUTE, 1971, P 216-219
2. J.D. SWART, PLASMA-MIG BOOSTS TANK TRAILER OUTPUT, *WELD. DES. FABR.,* FEB 1983, P 54-55, 59
3. F. EICHHORN AND E. VAN GAEVER, ADVANTAGEOUS SURFACING WITH PLASMA-MIG USING CORED WIRES, *PROC. 1ST INT. CONF. SURFACE ENGINEERING,* THE WELDING INSTITUTE, 1986, P 99-110

Resistance Spot Welding

Neville T. Williams, British Steel

Introduction

RESISTANCE SPOT WELDING (RSW) is a process in which faying surfaces are joined in one or more spots by the heat generated by resistance to the flow of electric current through workpieces that are held together under force by electrodes. The contacting surfaces in the region of current concentration are heated by a short-time pulse of low-voltage, high-amperage current to form a fused nugget of weld metal. When the flow of current ceases, the electrode force is maintained while the weld metal rapidly cools and solidifies. The electrodes are retracted after each weld, which usually is completed in a fraction of a second.

Spot welding is the most widely used joining technique for the assembly of sheet metal products such as automotive body-in-white assemblies, domestic appliances, furniture, building products, enclosures and, to a limited extent, aircraft components.

Many assemblies of two or more sheet metal stampings that do not require gas-tight or liquid-tight joints can be more economically joined by high-speed RSW than by mechanical methods. Containers frequently are spot welded. The attachment of braces, brackets, pads, or clips to formed sheet-metal parts such as cases, covers, bases, or trays is another common application of RSW.

Major advantages of spot welding include high operating speeds and suitability for automation or robotization and inclusion in high-production assembly lines together with other fabricating operations. With automatic control of current, timing, and electrode force, sound spot welds can be produced consistently at high production rates and low unit labor costs using semiskilled operators.

Most metals can be resistance spot welded if the appropriate equipment is used coupled with suitable welding conditions. This is particularly true for thin sheet or strip steel products, whether uncoated or coated.

Neville T. Williams, British Steel

Equipment

The equipment needed for RSW can be simple and inexpensive or complex and costly, depending on the degree of automation. Spot welding machines are composed of three principal elements:

- *ELECTRICAL CIRCUIT*, WHICH CONSISTS OF A WELDING TRANSFORMER, TAP SWITCH, AND A SECONDARY CIRCUIT
- *CONTROL CIRCUIT*, WHICH INITIATES AND TIMES THE DURATION OF CURRENT FLOW AND REGULATES THE WELDING CURRENT
- *MECHANICAL SYSTEM*, WHICH CONSISTS OF THE FRAME, FIXTURES, AND OTHER DEVICES THAT HOLD AND CLAMP THE WORKPIECE AND APPLY THE WELDING FORCE

Specifications for resistance welding equipment have been standardized by the Resistance Welder Manufacturers Association (RWMA), and specifications for controls are issued by the National Electrical Manufacturers Association (NEMA).

Resistance Spot Welding

Neville T. Williams, British Steel

Electrical Circuit

The transformer used in a direct energy resistance delivers the power to the workpiece by changing the input high-voltage, low-amperage, alternating current (ac) in the primary winding to a low-voltage, high-amperage current in the secondary winding. This system forms the basis of pedestal, or gun-type, welding machines where the output of the secondary transformer is applied directly to the welding electrodes. More complex secondary circuits are used in multiwelders.

To minimize the size and cost of multiwelders, transformers can be designed with two or more secondary circuits supplied from the same primary circuit. This design is achieved by using transformers in which the primary current is supplied to two or more separate secondary windings. Split or multiple secondary transformers are generally used to produce four or more welds at the same time. One controller is used for control of the timing and current supplied to the electrode, which means that it is not possible, under these circumstances, to achieve individual control or setting of the welding conditions for each pair of welding electrodes. Difficulties can arise in optimizing the necessary welding conditions to maximize the electrode life for each welding station under these conditions. In addition, the possibility exists for developing shunt currents, or an out-of-balance welding current, between each secondary circuit. Multiwelders make large use of single-welding and multiple-welding configurations based on direct, indirect, series, and push-pull secondary circuits (see Fig. 7 of the article "Procedure Development and Practice Considerations for Resistance Welding" in this Volume).

Single-Weld Configurations. Most pedestal and portable gun-welding machines employ direct welding, where the current flows directly through the welding electrodes and the workpiece to produce a single weld. In areas of difficult access, indirect welding may be necessary. Indirect welding is generally used when welding is carried out using either stationary or indexing welding guns and when access to the sheets being welded is limited to one side. In this case, the backing electrode generally consists of a large bar or platen, and the welding current flows from one side of the transformer secondary to the active electrode, through the welded area, and then along the platen/bar electrode or connecting jumper cable to the dummy electrode, which is connected to the other leg of the transformer secondary. Higher welding currents are generally necessary when using dummy welding guns, which can lead to a shorter electrode life. The extent to which larger currents are required depends largely on the distance between the welding electrode and the dummy return electrode. In a multiwelder, this distance can be either fixed or variable, depending on whether a stationary or indexing head is used. Both types of welding heads can lead to losses due to stray current paths and extensive shunting of the current through jigs and fixtures. These factors need to be considered in any design concept for a multiwelder station.

Two-Weld Configurations. Because of the need to minimize the number of transformers and the overall size and cost of a multiwelder, it is necessary to utilize configurations whereby two welds are made in one stroke of an electrode. Parallel, series, or push-pull welding configurations are used for this purpose.

In parallel welding, two or more direct welds, usually closely spaced, are made simultaneously with all upper electrodes connected to one side of the transformer, and all bottom electrodes connected to the other side of the transformer. Parallel welding has the advantage of lower shunt currents, thereby allowing shorter interweld spacings to be used. However, if two or more circuits are fed from a common transformer secondary winding, the current in each circuit will not be equal unless the impedances of each circuit are identical. This is often difficult to achieve in practice, and the electrode life obtained can, therefore, be different for each of the electrode pairs connected to the same transformer.

Series and Push-Pull Welding. Two simultaneous welds can also be produced by means of series welding in which the secondary circuit is connected to two contoured electrodes that contact the workpiece from the same side, and a current backup bar/platen is used on the reverse side. Push-pull welding is also employed whereby either one or two welds are made, depending on the electrode configuration. In push-pull RSW, two transformers are used to form the circuit between the welding electrodes; such a configuration has advantages, particularly where a deep-throated machine is required. Such machines ordinarily require a large secondary loop because of the need for conducting the welding current around the large working area. Using a push-pull configuration, where the upper set of transformers matches the lower set, eliminates the need for long cable loops to the electrodes, resulting in a lower secondary impedance. A longer electrode life can be obtained in push-pull welding compared to series welding.

It should also be noted that current shunting between the electrodes occurs in both series and push-pull welding configurations, but to a lower extent in the latter. The magnitude of the shunt currents depends on interelectrode distance and the type of material being welded. In all cases, current shunting will lead to a lower electrode life when welding zinc-coated steels due to the need to use higher welding currents to compensate for the shunt currents (this can also lead to additional electrode sticking problems) and to the rapid falloff in weld size that occurs under conditions where current compensation is insufficient or nonexistent. To some extent, the effects of both high secondary circuit impedance and current shunting can be overcome by using a constant-current facility fitted into the welding machine control circuit, but there is a limit to the effectiveness of these controls in a production situation. In addition, direct current (dc) welding systems can be used in which secondary inductance/reactance effects are minimized. However, the effectiveness of using dc depends on the method used to develop the dc waveform and on the extent of the superimposed ac ripple on the dc waveform.

Secondary Impedance. The electrical characteristics of a spot welding machine are best defined in terms of their effect on the impedance of the secondary circuit. The latter is composed of two components:

- *REACTANCE*, WHICH IS DETERMINED PRIMARILY BY THE DEPTH AND HEIGHT OF THE THROAT OF THE MACHINE AND THE AMOUNT OF STEEL IN THE THROAT. ANOTHER CRITICAL FACTOR IS THE FREQUENCY OF THE APPLIED CURRENT.
- *RESISTANCE*, WHICH DEPENDS ON THE LENGTH AND CROSS SECTION OF THE CONDUCTORS. CONDUCTOR MATERIAL AND OPERATING TEMPERATURE NEED ALSO TO BE CONSIDERED.

Cables can be classified into three types: air-cooled jumper cables, water-cooled jumper cables, and kickless cables. In any application, it is necessary to ensure that the type and dimensions of cables are chosen to achieve the best balance between the total circuit resistance and the electrode life to be obtained with a multiple spot welder.

Machine Construction

Welding operations in highly automated production lines are based primarily on multiple spot welders and robotic cells. In addition, manual welding operations can be used to manufacture either subassemblies, which are fed into the main production/assembly lines, or, in many instances, finished products. These differing end uses require welding machines of varying designs and characteristics. RSW machines can be divided into three basic types:

- PEDESTAL-TYPE WELDING MACHINES
- PORTABLE WELDING GUNS
- MULTIPLE SPOT WELDING MACHINES INCORPORATING LIGHTWEIGHT GUN WELDING UNITS

Pedestal-type welding machines form the basis of most resistance spot or projection welding operations. Standard machines are either of the rocker arm type, as determined by the rocker action of the moving upper electrode arm, or stationary, direct-acting machines. The latter outnumber the former because rocker arm machines are difficult to align due to the movement of the electrode in an arc motion. As the electrode wears, various skidding actions can be created between the upper and lower welding electrodes, depending on the extent of the electrode wear. This can change the point at which the electrodes meet relative to the fulcrum point of the moving arm. If the electrodes meet at a point above the fulcrum point, they have a tendency to slide outward from the machine. On the other hand, if the electrodes meet on a line below the fulcrum of the moving arm, the electrodes show a tendency to slide inward toward the machine. These relative movements can be different between the top and bottom electrodes, a consequence of which is a reduction in electrode life, particularly if welding coated steels. Other advantages of direct-acting machines include a lower electrode assembly mass, faster "follow up" during welding, and a degree of rigidity that is impossible to obtain with an ordinary rocker arm machine. Press-type direct-acting machines have an upper electrode and welding head that move vertically in a straight line. The welding head is guided in bearings or guideways of sufficient proportion to withstand the offset loads imposed upon them. All of these factors tend to increase the electrode life obtained in a particular welding cell.

Portable welding guns are used in RSW when it is impractical or inconvenient to transport the work to the machine. The portable welding gun consists of water-cooled electrode holders, an air or hydraulic actuating cylinder, hand grips, and an initiating switch. The gun usually is suspended from an adjustable balancing unit, and the welding force is supplied by an air cylinder. Because of the high secondary losses of portable machines, transformers used in these machines have a secondary voltage two to four times as great as the voltage of transformers used in stationary machines of equal rating.

Welding current is transmitted between the transformer terminals and gun terminals through a secondary cable, usually of the low-impedance or kickless type. The reactance of this type of cable is near zero, which results in a high power factor and reduced kilovolt-ampere (kVA) demand. For safety reasons, the current initiating switch is usually operated at low voltage. Recent developments in encapsulated transformers and high-speed fail-safe devices enable the transformer to be integral with the welding gun. This results in lower secondary losses and is particularly preferred in robotic applications.

Gun welders typical of portable manually operated and robotic welding cells can be classified into two basic types:

- S-TYPE (SCISSORS)
- J-TYPE (DIRECT ACTION OR C/J TYPE)

Two important features that influence the performance of these types of welding guns are the rate of force buildup and the extent to which the arms of the gun deflect under load. These features characterize the welding gun in terms of rigidity and the mechanics involved in their operation. An S-type gun utilizes a lever action to transmit the force applied by the cylinder head to the welding electrodes. However, due to an unbalanced pivot point, the force measured at the electrodes is less than that applied at the cylinder. With J-type welding guns, the force applied by the pneumatic cylinder is transmitted directly to the welding electrode. The rigidity of the framework of the welding gun is characterized by the extent to which the electrode arms deflect on application of the electrode force. This is an important factor in determining weld quality and the amount of skidding that can occur between the two welding electrodes. It should be noted that extensive skidding can lead to elliptical welds and/or decreased electrode life.

Welding Electrodes

Materials for RSW electrodes should have sufficiently high thermal and electrical conductivities and sufficiently low contact resistance to prevent burning of the workpiece surface or alloying at the electrode face. In addition, the electrode should have adequate strength to resist deformation at operating pressures and temperatures. Because the part of the electrode that contacts the workpiece becomes heated to high temperatures during welding, hardness and annealing temperatures must also be considered. Electrode materials for RSW have been classified by RWMA and in International Standards Organization (ISO) standard ISO 5182.

When welding two sheets of thickness up to 3 mm ($\frac{1}{8}$ in.) using truncated cone-type electrodes, the electrode tip diameter should be determined from:

$$D_2 = \sqrt{t} \qquad \textbf{(EQ 1)}$$

where d_2 is the initial tip diameter (in mm) and t is the thickness (in mm) of the sheet in contact with the electrode. When using truncated cone electrodes, the initial or set-up weld diameter should be equal to the diameter of the electrode tip:

$$D = D_2 = 5\sqrt{t} \qquad \textbf{(EQ 2)}$$

where d is the weld diameter (in mm). Whenever welding two sheets of dissimilar thicknesses, the electrode dimensions and the required weld size should be specified appropriate to the thinner sheet thickness. In the case of three thicknesses, the second thinnest sheet should be used.

When using domed or pointed electrodes, Eq 1 may not apply and the electrode dimensions will depend on accessibility and flange width. In this case, the electrode tip dimensions and welding conditions should be chosen to give an initial weld diameter as specified in Eq 2. Where a pad or mandrel is used as the second electrode, its surface must be maintained to match the profile of the work.

During normal production, electrodes tend to mushroom leading to an increase in electrode tip size. According to trends in automotive practice and international standards, the diameter of at least one of the electrode tips should not normally be allowed to increase above a value that results in a reduction greater than 30% from the starting weld diameter during a production run. This is equivalent to an increase in the electrode diameter to 1.3d. When this diameter has been reached, the electrode should be replaced or redressed to its original size and contour.

Electrode shapes have been standardized by RWMA. Figure 1 shows the six standard face or nose shapes, identified by letters A through F. Electrodes with type A (pointed) tips are used in applications for which full diameter tips are too wide. Type D (eccentric, formerly called offset) faces are used in comers or close to upturned flanges. Special tools are available for dressing the electrode faces, either in or out of the welding machine.

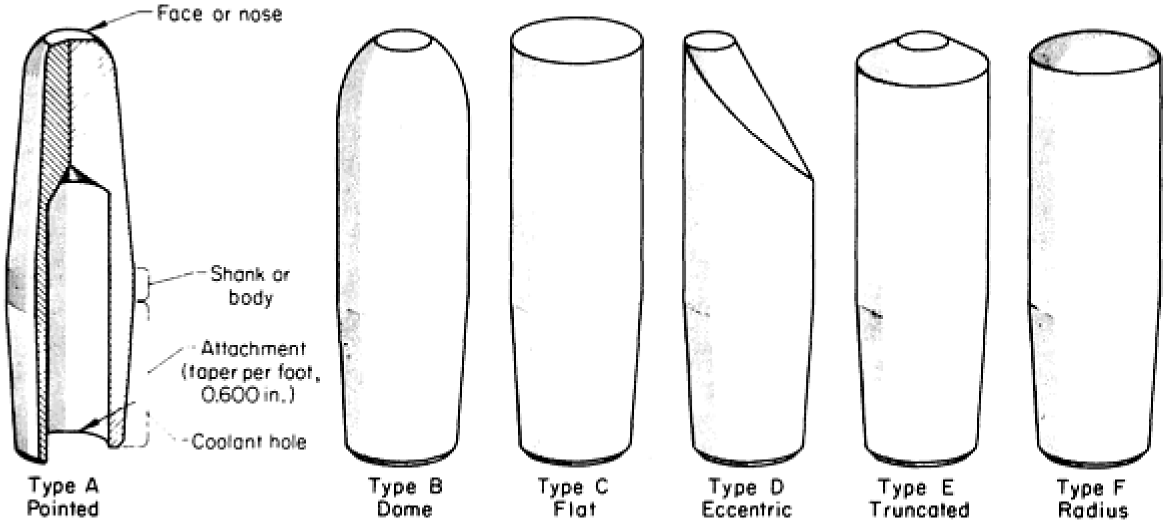

FIG. 1 STANDARD-RESISTANCE SPOT WELDING ELECTRODE FACE AND NOSE CONFIGURATIONS

Electrode Coolant Parameters. Generally, it is recommended that the water flow to the electrodes should be a minimum of 4 L/min (1.1 gal/min) for welding two uncoated steel sheets of thickness up to and including 3.0 mm ($\frac{1}{8}$ in.). Higher flow rates are recommended when welding coated steels. The internal water cooling feed tube should be arranged to ensure that the water impinges onto the back working face of the electrode. The distance between the back and the working face of the electrode should not exceed the values given in the appropriate ISO or RWMA standard. The inlet water temperature should not exceed 20 °C (70 °F), and the outlet temperature should not exceed 30 °C (85 °F). To maintain these temperature levels, the electrode cooling water supply should be independent of transformer and thyristor water cooling circuits. Separate water circuits should be used for both top and bottom electrodes.

Resistance Spot Welding

Neville T. Williams, British Steel

Weldability of Steel

Uncoated Steels. Low-carbon steel can be satisfactorily resistance spot welded using a wide range of time, current, and electrode force parameters. Practices recommended for spot welding SAW 1010 steel have been issued by the RWMA (Ref 1) and the American Welding Society (Ref 2). These data can serve as starting points for the establishment of optimum settings but may require modification, depending on the type of welding machine, the dynamic properties of the welding machine, the pneumatic characteristics, the characteristics of the secondary circuit, the electrode shape, and the material. Sufficient squeeze time should be chosen to enable the electrode force to build up to its preset value. For lightweight gun welders with limited force capability, the electrode force values are reduced up to 30% for sheet thicknesses greater than 1.6 mm ($\frac{1}{16}$ in.). The welding current needs to be adjusted accordingly. When welding sheets of dissimilar thickness, welding conditions may be based on the thinner sheet, or on the second thinnest sheet when welding three thicknesses.

In the case of high-strength low-alloy (HSLA) steels, up to 20% higher electrode force may be necessary. Welding currents may be reduced up to 20%, depending on the type of high-strength steel being welded. Pulse welding schedules may be necessary at the higher carbon equivalent levels and at sheet thicknesses greater than 2.0 mm ($\frac{5}{64}$ in.). The welding conditions required may depend greatly on the analysis of the steel because of differences in the resistivity of the various types of high-strength steel. For example, carbon content has the greatest effect on weldability of steels; weld hardness increases rapidly with small increases in carbon content. This high hardness causes nugget interfacial tears and nugget deterioration. To obtain acceptable weld performance, carbon content should be kept below 0.10% + 0.3*t*, where *t* is the sheet thickness in inches. For materials above this range, postweld tempering may be necessary.

Low-carbon steels are generally considered to be spot weldable. These steels can be obtained from a number of suppliers with reasonable certainty of only minor variations in weldability. On the other hand, HSLA steels are sold on the basis of strength, with each steelmaker producing its own composition. Therefore, this variation in composition may cause variations in weldability, which the steel user must consider.

Specific alloying elements or combinations of elements can impart desired nugget properties in low-carbon or HSLA steels. The effects of particular elements are discussed below.

Phosphorus and Sulfur. Phosphorus and, to some extent, sulfur, are generally considered to promote nugget interface tearing. When the total content of phosphorus, sulfur, and carbon exceeds a critical value, spot weld interfacial failure is observed. This interfacial tearing during peel testing results in reduced current ranges and minimal lobe curve areas.

Titanium. Data on the effect of titanium on both hot-rolled and cold-rolled steel indicate that the spot weldability of these steels is not as good as that of the HSLA steels containing niobium and/or vanadium. Increasing titanium content generally reduces maximum button diameter, shear and cross-weld tensile strength, and welding current range. Titanium content should not exceed 0.18%, and the use of oversized electrodes and higher force may be required when welding titanium-bearing steels.

Nitrogen. In HSLA steels, nitrogen promotes interfacial nugget failure. Nitrogen appears to be more critical in unkilled cold-rolled steel. However, the sensitivity can be reduced by decreasing the nitrogen content or by combining the nitrogen with aluminum as in aluminum-killed steels.

Zinc-Coated Steels. The present trend in the automotive industry toward the use of larger amounts of zinc-coated steels in automotive assemblies demands that certain strict guidelines regarding the selection of equipment, choice of welding schedules, and maintenance procedures are rigidly followed.

Two basic types of alloy coating are available:

- ZINC-IRON (BOTH HOT-DIP AND ELECTRODEPOSITED)
- ELECTRODEPOSITED ZINC-NICKEL (SUPPLIED WITH OR WITHOUT AN ORGANIC PRIMER TOP COAT)

The available welding ranges for zinc-iron and zinc-nickel coated steels are similar to those for uncoated mild steel although displaced toward slightly higher currents (Fig. 2). When these steels are coated, it is essential that the film thickness not exceed 1 to 1.5 μm (40 to 60 μin.) in order to facilitate breakthrough of the film to enable current flow between the welding electrodes at the low secondary voltages typical of RSW machines.

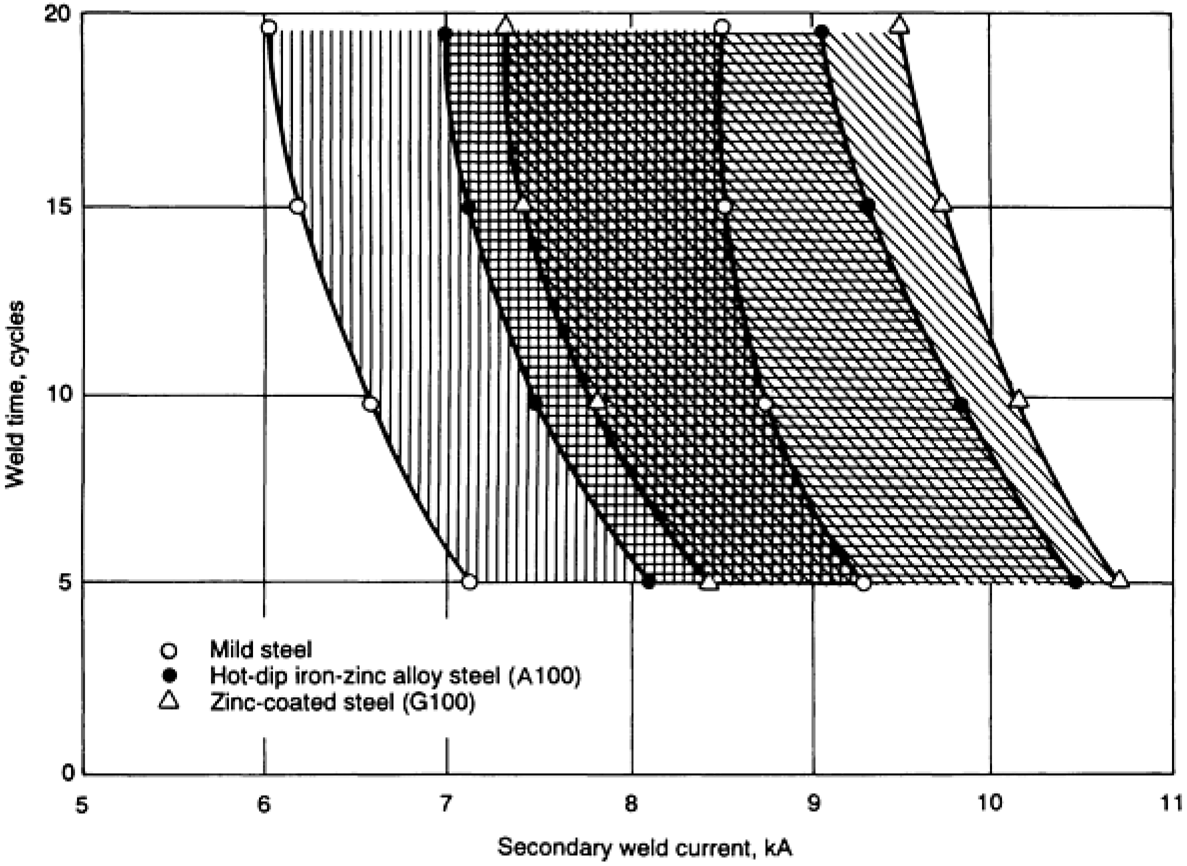

FIG. 2 PLOT OF WELD TIME VERSUS SECONDARY WELD CURRENT TO OBTAIN WELDABILITY LOBES FOR SELECTED 0.8 MM (0.03 IN.) THICK STEELS. ELECTRODE PARAMETERS: FORCE, 1.8 KN (0.20 TONF); TIP DIAMETER, 5.0 MM (0.20 IN.)

Welding conditions can be optimized easily when using pedestal-type welding machines; however, this may not be the case in body-in-white assembly where a combination of multiple spot welders and robotic cells are more commonly used. Such operations are susceptible to poor fit-up of parts, access problems, and electrode configurations that give rise to electrode skidding. Consequently, the welding parameters may need to be adjusted to compensate for such situations. In many instances, one welding gun is called on to weld a range of sheet thickness combinations or even a variety of material combinations in a particular part. Although a multi-program welding control can be used to adjust for the different welding currents necessary for the various combinations, the actual current values may not be optimum because, more often than not, a single compromise electrode force value has to be used.

In these circumstances, the electrode life obtained may be lower than that derived using optimized welding conditions. For example, the electrode life obtained can depend greatly on the type of welding gun used (Fig. 3) or the angle of approach of the electrode. For multiple spot welders, it is essential that all welding stations are balanced electrically and that the same air pressure is supplied to each station. Electrode shape and configuration, water cooling arrangements, and electrode dressing schedules need to be optimized. The last item is most important because satisfactory weld quality can only be provided when the necessary electrode condition and shape are maintained over a production run.

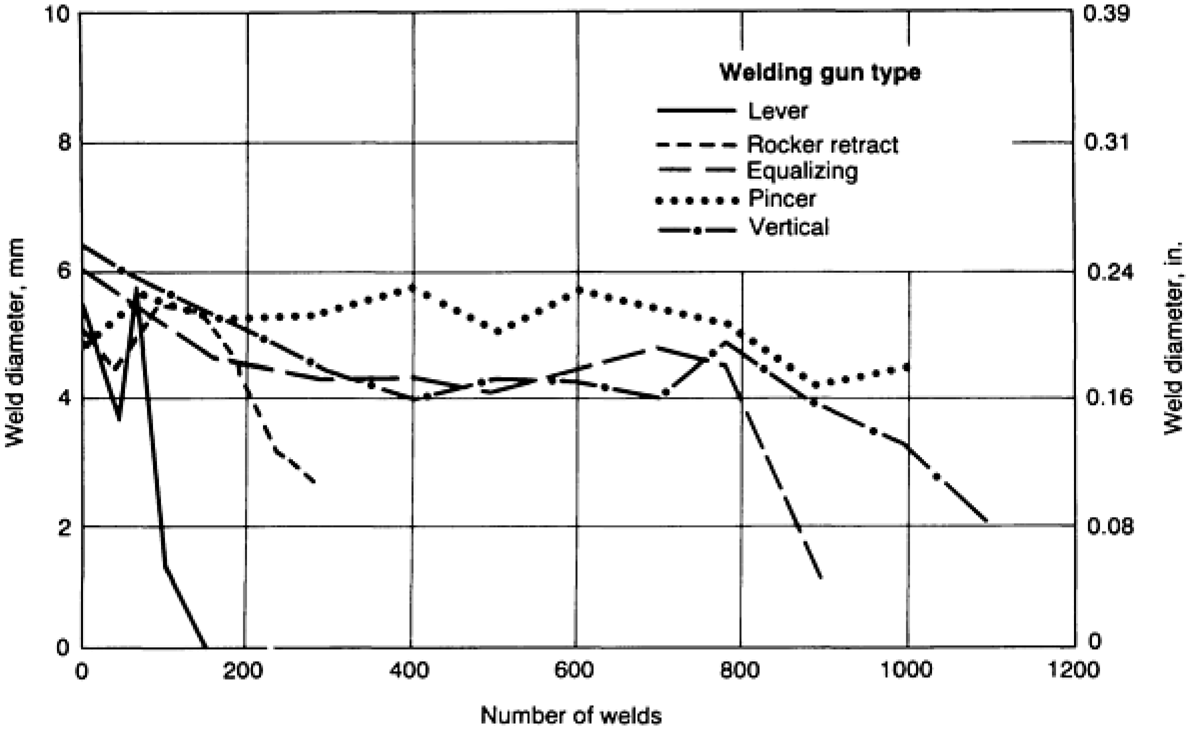

FIG. 3 EFFECT OF RSW ELECTRODE GUN TYPE ON ELECTRODE LIFE FOR SELECTED STEELS AND ALLOYS

References cited in this section

1. *RESISTANCE WELDING MANUAL,* 4TH ED., RESISTANCE WELDER MANUFACTURERS ASSOCIATION, 1989
2. *RECOMMENDED PRACTICES FOR AUTOMOTIVE WELD QUALITY--RESISTANCE SPOT WELDING,* ANSI/AWS D8.7-88, AMERICAN WELDING SOCIETY, 1988

Resistance Spot Welding

Neville T. Williams, British Steel

Weldability of Aluminum Alloys

Most of the commercial aluminum alloys produced in sheet form can be spot welded provided that suitable welding equipment is used. Because of the low electrical resistance of aluminum alloys, considerably higher welding currents are required relative to welding mild steel. In addition, the narrow plastic range between softening and melting means that welding forces, time, and current need to be closely controlled. For consistent weld quality, it is essential that the tenacious surface oxide films are removed by mechanical or chemical techniques prior to welding.

To minimize the likelihood of cracking or porosity in the weld, careful control of the electrode force is necessary. If the force is applied too soon, the welding current may be insufficient due to a premature loss of contact resistance. Thus, correct initiation of the force relative to the current is essential. Dual force schedules are frequently used; therefore, a low-inertia welding head can be beneficial because it allows a more effective buildup of force to its final level.

Because of the higher secondary currents required when welding aluminum alloys, direct current secondary rectified or frequency converter welding machines are favored, particularly if high-quality welds are required (such as in aircraft applications).

Resistance Spot Welding

Neville T. Williams, British Steel

References

1. *RESISTANCE WELDING MANUAL*, 4TH ED., RESISTANCE WELDER MANUFACTURERS ASSOCIATION, 1989
2. *RECOMMENDED PRACTICES FOR AUTOMOTIVE WELD QUALITY--RESISTANCE SPOT WELDING*, ANSI/AWS D8.7-88, AMERICAN WELDING SOCIETY, 1988

Projection Welding

Jerry E. Gould, Edison Welding Institute

Introduction

PROJECTION WELDING (PW) is a variation of resistance welding in which current flow is concentrated at the point of contact with a local geometric extension of one (or both) of the parts being welded. These extensions, or projections, are used to concentrate heat generation at the point of contact. The process typically uses lower currents, lower forces, and shorter welding times than does a similar application without the projections. Projection welding is often used in the most difficult resistance-welding applications because a number of welds can be made at one time, which speeds up the manufacturing process.

PW applications are generally categorized as being either embossed-projection welding or solid-projection welding. These variations are shown in Fig. 1 and 2. Embossed-projection welding is generally a sheet-to-sheet joining process, in which a projection is stamped onto one of the sheets to be joined. Then, resistance welding is conducted on a stack of sheets. Heat initially concentrates at the contact point and in the walls of the projection during resistance welding. Early in the process, the projection itself collapses back into the original sheet. However, the initial heating raises the local resistivity of the joint area, allowing continued resistance heating at this location. Weld development then proceeds in the conventional manner, by forming a fused weld nugget.

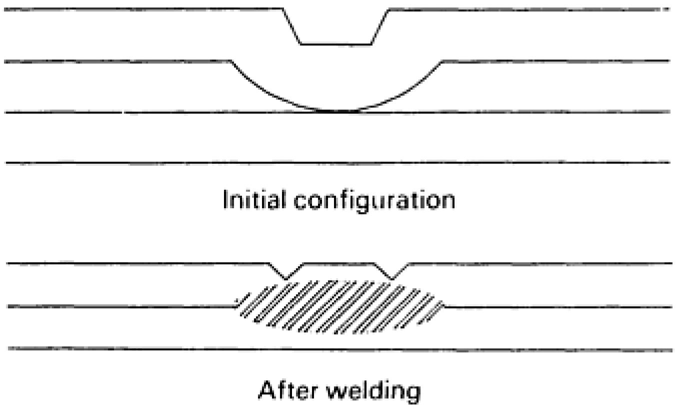

FIG. 1 TYPICAL STACKED CONFIGURATION FOR EMBOSSED-PROJECTION WELDING OF SHEET

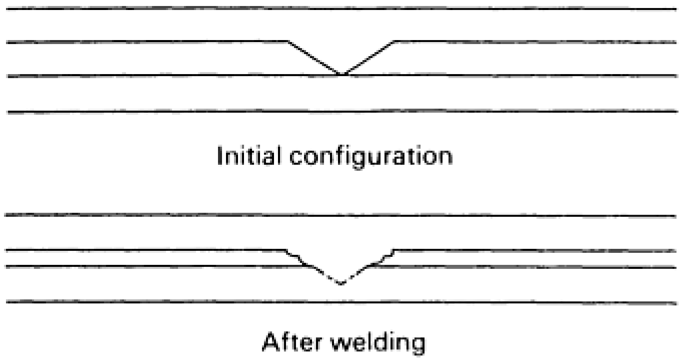

FIG. 2 TYPICAL CONFIGURATION FOR SOLID-PROJECTION WELDING

Solid-projection welding requires that the projection be forged onto one of the two components. Then, during resistance welding, the contact point and the projection itself experience preferential heating. In this case, the projection cannot simply collapse, as it does in embossed-projection welding. Rather, the projection collapses by penetrating the opposing material and by extrusion to the periphery. When compared with embossed-projection welding, the resulting joints are solid-state, rather than fusion, welds. The actual joints are caused by a combination of material forging and diffusion bonding, much like they are in resistance butt and flash butt welding.

Projection Welding

Jerry E. Gould, Edison Welding Institute

Applications

Examples of projection welding are shown in Fig. 3. These applications, which range from sheet-to-sheet joints, to cross-wire welds, to annular attachments, to nut welds, to weld screws, include both embossed- and solid-projection types of welding. Specific welding examples, as well as material effects, are described below.

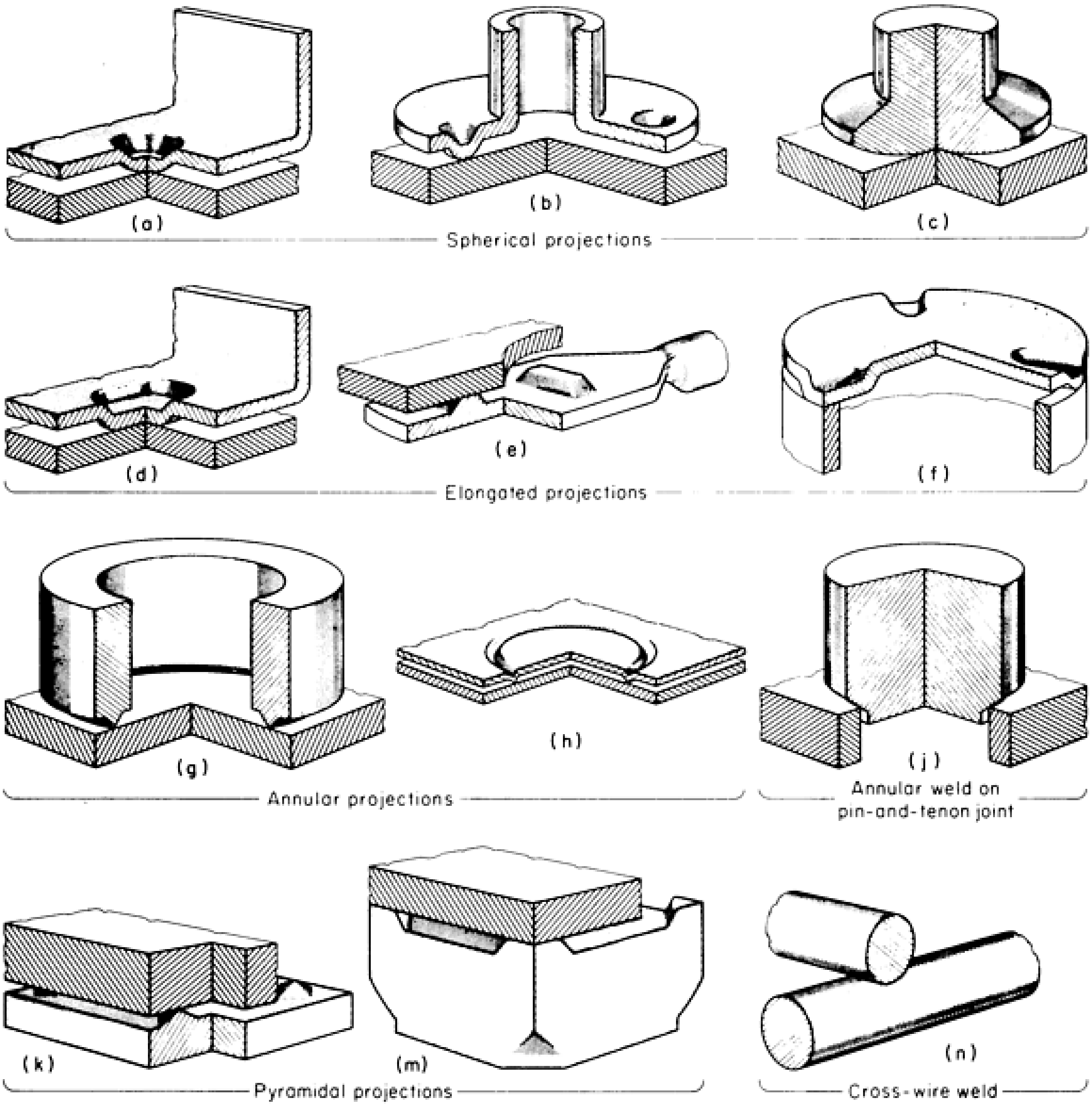

FIG. 3 DIFFERENT PROJECTION-WELDING CONFIGURATIONS

Embossed Projections

Typical embossed-projection weld configurations are shown in Fig. 3(a), (b), (d), and (h). The embossed-projection technique is primarily used for sheet-to-sheet applications. It does involve the additional expense of stamping the projections and ensuring that the electrodes are properly located over the projections. However, the ability of the projection to create highly localized joints allows this technique to extend the applications of resistance welding. Most notably, projection welding widely increases the allowable thickness mismatch associated with resistance welding.

Generally, resistance welding is considered adaptable to thickness variations of less than 3:1. However, with the addition of the projection, there is no limit to the difference in joinable thicknesses. Similarly, embossed-projection welding is highly adaptable to the joining of multiple-sheet stacks. The approach is similar in that the projections are used to localize heat at critical interfaces in the overall joint.

Another more recent use of embossed-projection welding is in applications where external marking must be minimized. This is a concern when a primer has been applied to one side of one of the sheets being joined. Because it is very effective at localizing heat at the interface, and because it uses shorter welding times, the PW process is more adiabatic and there is less residual heat that can damage outside surfaces. However, surface marking still does occur.

Solid Projections

A considerably wider range of solid-projection welding processes is commonly used in production applications. Annular projection welding, like embossed-projection welding, is commonly used to provide a highly localized joint and to minimize thermal damage to other parts of the structure. Common variations of solid-projection welding are described below.

Annular Projection Welds. One of the most common applications of solid-projection welding is to attach either tubular components or members with circular bases to flat or curved substrates. This is commonly accomplished by annular projection welding, in which the projection is machined onto the circular base or tubular section. Then, resistance welding is conducted to set the projection into the substrate. A typical weld geometry is shown in Fig. 3(g). Annular projection welding generally provides high-integrity joints and can be used for leak-tight applications.

Cross-wire welding is a variation of solid-projection welding in which the projection is formed by the contact point of two crossing wires. The geometry for cross-wire welding is shown in Fig. 3(n). Upon resistance welding, heat is maximized at the location of the wire-wire contact, and the parts are subsequently forged together. Depending on the application, either highly localized joints (minimal forging) or heavily forged joints can be made. In some applications, cross-wire welding is conducted at rates up to 30 welds/s.

Weld nuts represent an application of solid-projection welding in which the projections themselves occur as actual extensions of material from the nut. Typically, weld nuts contain 3, 4, or 6 of these projections, or feet, spaced around the nut periphery. During welding, all projections are attached simultaneously. Examples of weld nut geometries are shown in Fig. 3(k) and 3(m).

Edge-to-sheet welds are typically a cross between embossed- and solid-projection welding. They are used to attach the end of a sheet to the flat face of an opposing sheet (Fig. 3f). The projections for welding are generally stamped into the face of the attaching sheet. During welding, the projection does collapse locally back into the base sheet. However, bonding is strictly solid state, similar to that of other solid-projection welding processes.

Material Effects

Solid-projection welds are essentially strain-assisted diffusion bonds. Because the projections typically collapse at very high temperatures (generally, within several hundred °C of the melting point), diffusion bonding can occur within the very limited available time (usually, less than 1 s).

Not surprisingly, some of the material-related factors that affect diffusion bonding also affect solid-projection welding. The most notable of these is the ability of the metal to dissolve its own oxide. As a result, materials that do not energetically favor the solid solubility of oxygen, rather than the formation of the oxide at aluminum forging temperatures, will be relatively difficult to projection weld.

The strength-temperature relationship also affects projection weldability. Materials that maintain their strengths at relatively high temperatures permit substantial heating to occur before projection collapse. This heat then becomes available to promote diffusion after projection collapse. Premature collapse results in lower temperatures in which diffusion can occur and in reduced current density, which prohibits further resistance heating.

Bulk resistivity also plays a role in projection welding, but to a lesser degree. Increased bulk resistivity can reduce the effectiveness of the projection as a current concentrator. With increasing bulk resistivity, the tendency is for delocalized heating and general, rather than local, collapse of the projection. As a result, high-resistivity materials are more difficult to projection weld.

Mild steels and low-alloy, nickel-base alloys are ideal materials for projection welding, because they readily dissolve their own oxides and have adequate strength-temperature and resistivity properties. Stainless steels and higher-alloy nickel-base materials become slightly more difficult to weld, because of the formation of more-stable chromium and aluminum oxides, increased high-temperature properties, and higher resistivities.

Projection welding is commonly applied to copper and copper-base alloys. In many applications, projections are virtually required for resistance welding, because of the high conductivity of these materials.

On the other hand, aluminum and aluminum-base alloys are very difficult to projection weld. The aluminum oxide is so tenacious that solid-projection welding, in particular, is nearly impossible. In addition, it is very difficult to localize heat, because aluminum alloys soften at such low temperatures.

Conventional titanium alloys are also relatively difficult to projection weld. Although titanium readily dissolves its own oxide, its high resistivity and low forging temperature generally cause premature collapse of the projection.

Projection Welding

Jerry E. Gould, Edison Welding Institute

Equipment

Resistance projection welding generally utilizes slightly modified resistance-welding equipment. Because the nature of projection welding is to direct current flow to the desired weld locations, there is considerably more flexibility in tooling. A specific example is spot welding that requires electrodes of specific sizes and shapes in order to localize current for welding.

Projection welding is typically done with large, flat electrodes. In most applications, tooling is simply shaped to match the contour of the part in the contacting location. Given the current-locating characteristic of projections, multiple projection welds can be made simultaneously.

Projection welding does require high compliance or the rapid response of the loading system. This is necessary because projection welding requires the collapse of a projection, which, in turn, requires some motion of the welding head. If the follow-up of the welding head is insufficient, then a local loss of force will occur, potentially causing catastrophic overheating of the contacting surfaces. Therefore, PW systems typically utilize "fast follow-up" heads (Fig. 4), which consist of relatively low inertia (mass) head components and mechanisms to maintain a nominally constant force. The head shown in Fig. 4 uses a relatively low volume diaphragm assembly and a spring to maintain this force on the components. Most systems, however, use either these low-volume diaphragm assemblies or springs to accomplish fast follow-up.

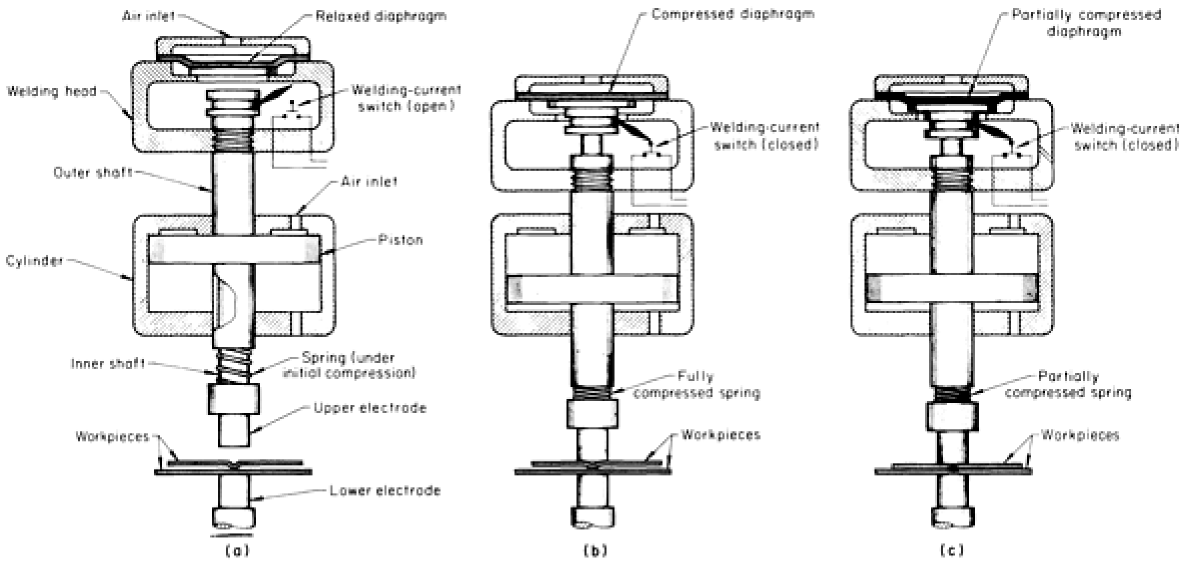
Air inlet
Relaxed diaphragm
Welding head
Welding-current switch (open)
Outer shaft
Air inlet
Piston
Cylinder
Inner shaft
Spring (under initial compression)
Upper electrode
Workpieces
Lower electrode
(a)
Compressed diaphragm
Welding-current switch (closed)
Fully compressed spring
Workpieces
(b)
Partially compressed diaphragm
Welding-current switch (closed)
Partially compressed spring
Workpieces
(c)

FIG. 4 TYPICAL FAST FOLLOW-UP (LOW INERTIA) HEAD FOR PROJECTION WELDING. (A) IN OPEN POSITION. (B) IN POSITION FOR SQUEEZING AND HEATING THE PROJECTION. (C) AT INSTANT OF PROJECTION COLLAPSE AND START OF NUGGET FORMATION

Projection Welding

Jerry E. Gould, Edison Welding Institute

Personnel

Like all resistance welding processes, PW processes utilize basically automated equipment, which means that the expertise level to operate the equipment is relatively low. Generally, the skill levels and the health and safety issues are identical to those of conventional resistance spot welding. An area in which expertise is beneficial to hardware function and on-stream weld quality is operator attention to detail. Resistance welding is subject to variations in setup conditions, and projection welding may be extraordinarily sensitive to setup variations in situations where the placement of projections is critical.

With limited expertise, operators or supervisory staff can visually inspect the characteristics of the process and the welds themselves to identify improper setups, as well as maintenance needs. These quality assurance efforts, which fall under the auspices of total quality management or total quality joining, can be established with a minimum of on-line training and experience.

Projection Welding

Jerry E. Gould, Edison Welding Institute

Process Requirements

The types of hardware required for projection welding are analogous to those required for other resistance-welding processes.

The embossed-projection welding of heavy-gage, intermediate-gage, and thin-gage sheet, as well as joints of dissimilar thicknesses, is described below.

Heavy-Gage Sheet Steels. The projection welding of heavy-gage steels, which involves an embossed-projection welding process, has many of the characteristics of solid-projection welding particularly in its early stages, because of the thickness of the sheet.

In addition to defining the geometry of the projection, projection designs should provide an annular relief for projection material that is forged to the side during welding. Projection and die geometries for steels that range from 3.12 to 6.22 mm (0.123 to 0.245 in.) are shown in Table 1. The process requirements for forming these welds are given in Table 2. All welding schedules described in this article are single-pulse welding schedules.

TABLE 1 PROJECTION AND DIE GEOMETRIES FOR WELDING A RANGE OF HEAVY-GAGE STEELS

Data are for tool steel hardened to 50-52 HRC.

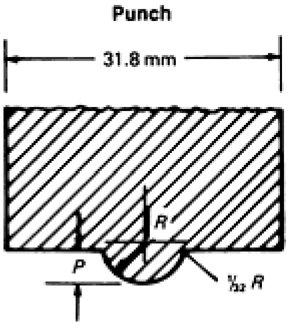

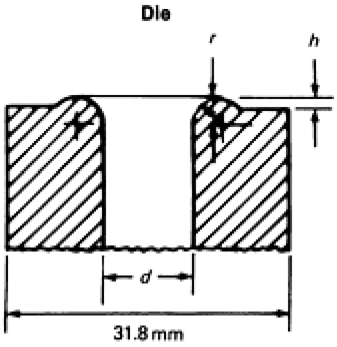

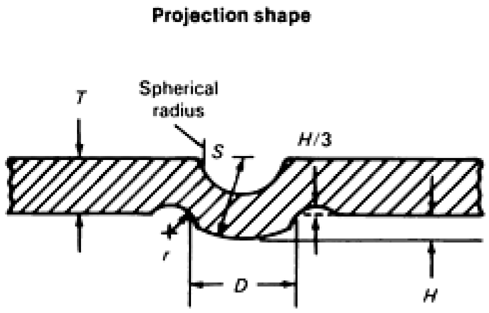

USS GAGE NO.	**MATERIAL THICKNESS (*T*)**		**PROJECTION**						**PUNCH**				**DIE**					
			HEIGHT (*H*), ±2%		**DIAMETER (*D*), ±5%**		**RADIUS (*S*)**		**RADIUS (*R*), ±0.1 MM (±0.005 IN.)**		**HEIGHT (*P*), ±2%**		**HOLE DIAMETER (*D*), ±0.1 MM (±0.005 IN.)**		**RECESS RADIUS (*R*), *R* = *S*/3**		**RECESS HEIGHT (*H*), *H* = *H*/3**	
	mm	**in.**	**mm**	**in.**	**mm**	**in.**	**mm**	**in.**	**mm**	**in.**	**mm**	**in.**	**mm**	**in.**	**mm**	**in.**	**mm**	**in.**
11	3.12	0.123	1.47	0.058	6.86	0.270	4.98	0.196	2.39	0.094	1.91	0.075	5.61	0.221	1.65	0.065	0.483	0.019
10	3.42	0.135	1.57	0.062	7.62	0.300	5.46	0.215	2.77	0.109	2.06	0.081	6.35	0.250	1.83	0.072	0.508	0.020
9	3.88	0.153	1.63	0.064	8.38	0.330	5.97	0.235	3.18	0.125	2.16	0.085	6.86	0.270	1.98	0.078	0.533	0.021
8	4.17	0.164	1.73	0.068	9.14	0.360	6.30	0.248	3.58	0.141	2.31	0.091	7.54	0.297	2.11	0.083	0.584	0.023
7	4.54	0.179	2.03	0.080	9.90	0.390	6.96	0.274	3.96	0.156	2.64	0.104	8.33	0.328	2.31	0.091	0.686	0.027
6	4.95	0.195	2.13	0.084	10.4	0.410	7.26	0.286	3.96	0.156	2.82	0.111	8.59	0.338	2.41	0.095	0.711	0.028
5	5.33	0.210	2.28	0.090	11.2	0.440	7.75	0.305	4.75	0.187	3.05	0.120	9.09	0.358	2.57	0.101	0.762	0.030
4	5.72	0.225	2.54	0.100	11.9	0.470	8.26	0.325	4.75	0.187	3.35	0.132	9.35	0.368	2.74	0.108	0.838	0.033
3	6.22	0.245	2.84	0.112	13.5	0.530	9.27	0.365	4.75	0.187	3.76	0.146	10.3	0.406	3.07	0.121	0.940	0.037

Source: Ref 1

TABLE 2 PROCESS REQUIREMENTS FOR PROJECTION WELDING OF A RANGE OF HEAVY-GAGE LOW-CARBON STEELS

USS GAGE NO.	THICKNESS (A)(B)		PROJECTION SIZE				MINIMUM				ELECTRODE FORCE WELD		ELECTRODE FORCE FORGE		UPSLOPE TIME(D), CYCLES	WELD-TIME(D), CYCLES	WELDING CURRENT(C), A	TENSILE-SHEAR STRENGTH(F)	
			DIAMETER		HEIGHT		MINIMUM SPACING, CENTERLINE TO CENTERLINE		MINIMUM CONTACT OVERLAP										
	mm	in.	mm	in.	mm	in.	mm	in.	mm	in.	mn	lbf	mn	lbf				mpa	ksi
SCHEDULE A: WELDING NORMAL-SIZED WELDS																			
9	3.89	0.153	8.38	0.330	1.57	0.062	44.5	1.75	22.9	0.9	8.9	2000	17.8	4000	15	60	15,400	52	7.5
8	4.17	0.164	8.89	0.350	1.73	0.068	45.7	1.80	24.1	0.95	10.2	2300	20.5	4600	15	70	16,100	56	8.1
7	4.55	0.179	9.91	0.390	2.03	0.080	48.3	1.90	25.4	1.0	11.7	2630	23.4	5260	20	82	17,400	66	9.5
6	4.95	0.195	1.04	0.410	2.13	0.084	50.8	2.00	26.7	1.05	13.0	2930	26.1	5860	20	98	18,800	78	11.3
5	5.33	0.210	1.12	0.440	2.34	0.092	53.3	2.10	29.2	1.15	14.2	3180	28.3	6360	25	112	20,200	86	12.5
4	5.72	0.225	1.19	0.470	2.54	0.100	33.0	1.30	30.5	1.20	16.1	3610	32.1	7220	25	126	21,500	103	15.0
3	6.22	0.245	1.35	0.530	2.84	0.112	63.5	2.50	33.0	1.30	17.4	3900	34.7	7800	30	145	23,300	119	17.3
SCHEDULE B: WELDING SMALL-SIZED WELDS																			
9	3.89	0.153	6.86	0.270	1.47	0.058	40.6	1.60	19.1	0.75	6.2	1400	12.5	2800	15	60	11,100	35	5.1
8	4.17	0.164	7.37	0.290	1.57	0.062	41.9	1.65	20.3	0.80	6.3	1425	12.7	2850	15	70	11,800	38	5.5
7	4.55	0.179	7.87	0.310	1.70	0.067	43.2	1.70	21.6	0.85	6.7	1500	13.4	3000	20	82	12,800	45	6.5
6	4.95	0.195	8.38	0.330	1.83	0.072	44.5	1.75	22.9	0.90	7.1	1600	14.2	3200	20	98	13,900	53	7.7
5	5.33	0.210	8.89	0.350	1.96	0.077	45.7	1.80	24.1	0.95	7.7	1730	15.4	3460	25	112	14,900	59	8.5
4	5.72	0.225	9.40	0.370	2.08	0.082	48.3	1.90	25.4	1.00	8.3	1870	16.6	3740	25	126	16,000	72	10.4
3	6.22	0.245	9.91	0.390	2.24	0.088	53.3	2.10	27.9	1.10	9.3	2100	18.7	4200	30	145	17,300	83	12.0

Source: Ref 1

(A) LOW-CARBON STEEL, SAE 1005-1010, 290-380 MPA (42-55 KSI) ULTIMATE TENSILE STRENGTH.

(B) SURFACE OF STEEL MAY BE OILED LIGHTLY BUT FREE FROM GREASE, SCALE, AND DIRT.

(C) ON SINGLE FORCE WELDS USE ONLY WELD FORCE AS ELECTRODE FORCE. ELECTRODE FORCE CONTAINS NO FACTOR TO FURTHER FORM POORLY MADE PARTS.

(D) BASED ON 60 HZ.

(E) STARTING VALUES SHOWN ARE BASED ON EXPERIENCE OF RWMA MEMBER COMPANIES.

(F) TENSILE-SHEAR STRENGTH PER PROTECTION DEPENDS ON THE JOINT DESIGN.

Slope control is recommended to prevent preflashing of the projection upon initiation of the welding current. Forge forces are also recommended, because weld porosity is often a concern with heavy-gage steel. Electrode or die wear may also be a concern, as it is in the spot welding of heavy-section steels. In such applications, pulsation welding schedules might be recommended.

Intermediate-Gage Sheet Steels. The projection welding of steels ranging in gage from 0.56 to 3.43 mm (0.022 to 0.135 in.), using single-point projections with single-impulse welding schedules, is well established. Projection-stamping die designs are given in Table 3, whereas process requirements are given in Table 4. Lower forces and shorter welding times are required, when compared with the conventional spot welding of these gages.

TABLE 3 PROJECTION AND DIE GEOMETRIES FOR WELDING A RANGE OF INTERMEDIATE-GAGE STEELS TO MAKE SPHERICAL PROJECTIONS

Data are for tool steel hardened to 50-52 HRC.

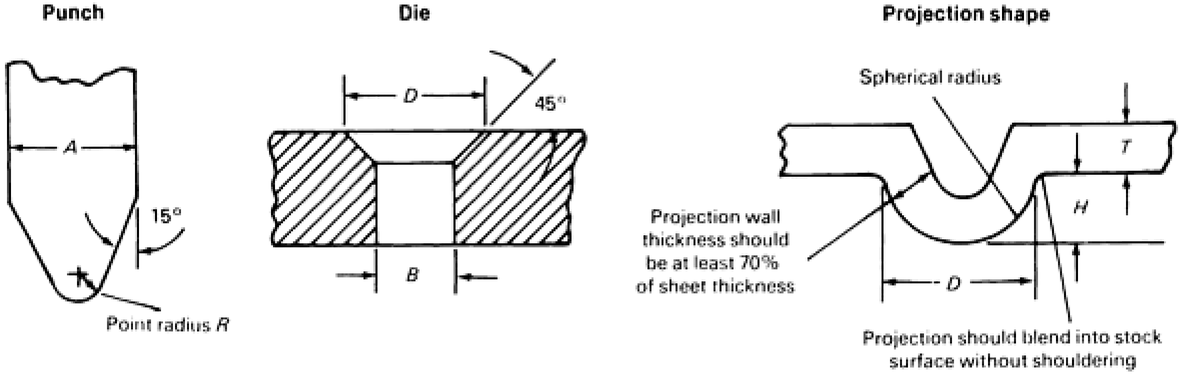

Punch

Die

Projection shape

A
15°
Point radius R
D
45°
B
Spherical radius
T
H
D
Projection wall thickness should be at least 70% of sheet thickness
Projection should blend into stock surface without shouldering

USS GAGE NO.	MATERIAL THICKNESS (*T*)		PROJECTION				PUNCH				DIE			
			HEIGHT (*H*), ±2%		DIAMETER (*D*), ±5%		DIAMETER (A)		POINT RADIUS (*R*), ±0.05 MM (±0.002 IN.)		HOLE DIAMETER (*B*), ±0.1 MM (±0.005 IN.)		CHAMBER DIAMETER (*D*)	
	mm	in.	mm	in.	mm	in.	mm	in.	mm	in.	mm	in.	mm	in.
25-21	0.56-0.86	0.022-0.034	0.64	0.025	2.3	0.090	9.53	0.375	0.79	0.031	1.93	0.076	2.29	0.090
20-19	0.91-1.1	0.036-0.043	0.89	0.035	2.8	0.110	9.53	0.375	1.19	0.047	2.26	0.089	2.79	0.110
18-17	1.2-1.4	0.049-0.054	0.97	0.038	3.6	0.140	9.53	0.375	1.19	0.047	2.64	0.104	3.56	0.140
16-15	1.5-1.7	0.061-0.067	1.1	0.042	3.8	0.150	9.53	0.375	1.57	0.062	3.05	0.120	3.81	0.150
14	1.9	0.077	1.2	0.048	4.6	0.180	9.53	0.375	1.57	0.062	3.66	0.144	4.57	0.180
13	2.3	0.092	1.3	0.050	5.3	0.210	12.7	0.500	1.98	0.078	4.37	0.172	5.33	0.210
12	2.72	0.107	1.4	0.055	6.1	0.240	12.7	0.500	1.98	0.078	4.98	0.196	6.10	0.240
11	3.12	0.123	1.5	0.058	6.9	0.270	12.7	0.500	2.39	0.094	5.61	0.221	6.86	0.270
10	3.43	0.135	1.6	0.062	7.6	0.300	12.7	0.500	2.77	0.109	6.35	0.250	7.62	0.300

Source: Ref 1

TABLE 4 PROCESS REQUIREMENTS FOR PROJECTION WELDING OF A RANGE OF INTERMEDIATE-GAGE LOW-CARBON STEELS

Data are for SAE 1005-1010 steels with ultimate tensile strengths ranging from 290-380 MPa (42-55 ksi). Surface of steel may be oiled lightly but must be free from grease, scale, and dirt.

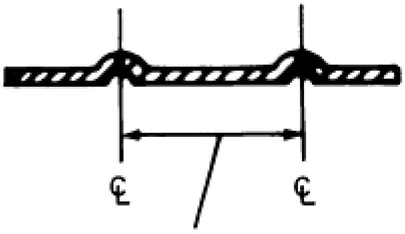

Spacing between projections

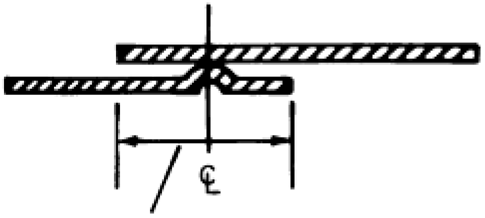

Contacting overlap

CONFIGURATION(A)	USS GAGE NO.	MATERIAL THICKNESS		PROJECTION SIZE				MINIMUM SPACING CENTERLINE TO CENTERLINE		MINIMUM CONTACT OVERLAP	
				DIAMETER		HEIGHT					
		mm	**in.**	**mm**	**in.**	**mm**	**in.**	**mm**	**in.**	**mm**	**in.**
A	25	0.559	0.022	2.29	0.090	0.635	0.025	9.65	0.38	6.35	0.25
B	23	0.711	0.028	2.29	0.090	0.635	0.025	9.65	0.38	6.35	0.25
C	21	0.864	0.034	2.79	0.110	0.889	0.035	12.7	0.50	9.65	0.38
D	19	1.09	0.043	2.79	0.110	0.889	0.035	12.7	0.50	9.65	0.38
E	18	1.24	0.049	3.56	0.140	0.965	0.038	19.1	0.75	12.7	0.50
F	16	1.55	0.061	3.81	0.150	1.07	0.042	19.1	0.75	12.7	0.50
G	14	1.96	0.077	4.57	0.180	1.22	0.048	22.4	0.88	12.7	0.50
H	13	2.34	0.092	5.33	0.210	1.27	0.050	26.9	1.06	15.7	0.62
I	12	2.72	0.107	6.10	0.240	1.40	0.055	31.8	1.25	19.1	0.75
J	11	3.12	0.123	6.86	0.270	1.47	0.058	38.1	1.50	20.6	0.81
K	10	3.43	0.135	7.62	0.300	1.57	0.062	41.4	1.63	22.4	0.88

CONFIGURATION(A)	WELD TIME(B), CYCLES	ELECTRODE FORCE(C)		WELDING CURRENT(D), A	TENSILE-SHEAR STRENGTH(E)	
		kn	**lbf**		**mpa**	**ksi**
WELDING SCHEDULE A (FOR SINGLE PROJECTION)(F)						
A	3	0.67	150	4,400	2.6	0.37
B	3	0.87	195	5,500	3.4	0.50
C	3	1.1	240	6,600	4.8	0.70
D	5	1.5	330	8,000	7.3	1.06
E	8	1.8	400	8,800	9.0	1.30
F	10	2.4	550	10,300	12.4	1.80
G	14	3.6	800	11,850	16.7	2.43
H	16	4.5	1020	13,150	22.4	3.25
I	19	5.6	1250	14,100	26.5	3.85
J	22	6.7	1500	14,850	33.1	4.80
K	24	7.3	1650	15,300	37.9	5.50
WELDING SCHEDULE B (FOR 1-3 PROJECTIONS), EACH PROJECTION(G)						
A	6	0.67	150	3,850	2.3	0.33
B	6	0.67	150	4,450	3.0	0.43
C	6	0.67	150	5,100	3.7	0.53
D	10	0.93	210	6,000	6.1	0.88
E	16	1.2	270	6,500	7.6	1.10
F	20	1.6	365	7,650	11.0	1.58
G	28	2.4	530	8,850	14.8	2.15
H	32	3.0	680	9,750	19.3	2.80
I	38	3.7	830	10,600	23.8	3.45
J	45	4.5	1000	11,300	29.0	4.20
K	48	4.9	1100	11,850	33.4	4.85
WELDING SCHEDULE C (FOR [GES]3 PROJECTIONS), EACH PROJECTION(H)						

B	8	0.45	100	3300	2.3	0.34
C	11	0.56	125	3800	2.9	0.43
D	15	0.71	160	4300	5.0	0.72
E	19	0.98	220	4600	6.1	0.88
F	25	1.5	330	5400	8.5	1.23
G	34	2.1	470	6400	12.1	1.75
H	42	2.7	610	7200	16.0	2.33
I	50	3.3	740	8300	20.0	2.90
J	60	4.0	900	9200	24.8	3.60
K	66	4.5	1000	9900	29.3	4.25

(A) CONFIGURATION DESIGNATIONS ARE FOR THE PURPOSES OF THIS TABLE ONLY.
(B) BASED ON 60 HZ.
(C) ELECTRODE FORCE CONTAINS NO FACTOR TO FURTHER FORM POORLY MADE PARTS.
(D) STARTING VALUES SHOWN ARE BASED ON EXPERIENCE OF RWMA MEMBER COMPANIES.
(E) TENSILE-SHEAR STRENGTH PER PROJECTION DEPENDS ON THE JOINT DESIGN.
(F) SCHEDULE A IS USABLE FOR WELDING MORE THAN ONE PROJECTION IF CURRENT IS DECREASED, BUT EXCESSIVE WELD EXPULSION MAY RESULT AND POWER DEMAND WILL BE GREATER THAN THAT FOR SCHEDULE B OR C.
(G) SCHEDULE B IS USABLE FOR WELDING MORE THAN THREE PROJECTIONS, BUT SOME WELD EXPULSION MAY RESULT, AND POWER DEMAND WILL BE GREATER THAN THAT FOR SCHEDULE C.
(H) SCHEDULE C IS USABLE FOR WELDING LESS THAN THREE PROJECTIONS WITH WELD CURRENT INCREASED APPROXIMATELY 15% AND POSSIBLE OBJECTIONABLE FINAL SHEET SEPARATION.

Thin-Gage Sheet. The projection welding of thin-gage steel sheet (<0.50 mm, or 0.020 in.) differs significantly from the projection welding of intermediate-gage sheet in that single-point projections become mechanically unstable. It also becomes increasingly difficult to use single-point projections to make adequate-sized welds. Therefore, annular projections are typically recommended for thin-gage steel. Table 5 provides the annular projection weld geometry, as well as the required process conditions.

TABLE 5 PROJECTION DESIGNS AND PROCESS REQUIREMENTS FOR ANNULAR PROJECTION WELDING OF THIN-GAGE LOW-CARBON STEEL

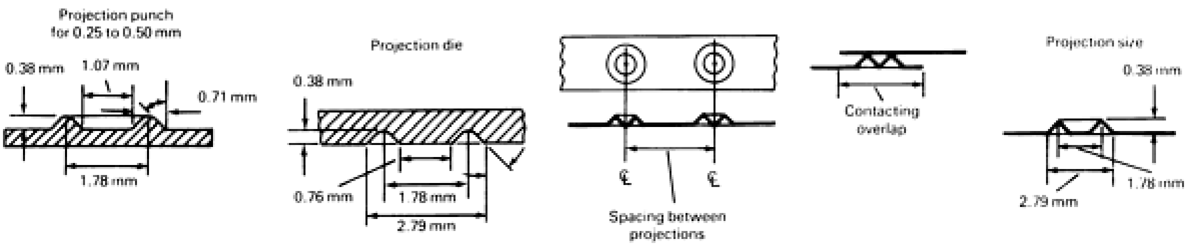

USS GAGE NO.	THICKNESS(A)(B)		MINIMUM SPACING CENTERLINE TO CENTERLINE		MINIMUM CONTACT OVERLAP		WELD TIME(C), CYCLES	ELECTRODE FORCE(D)		WELDING CURRENT(E), A	TENSILE-SHEAR STRENGTH EACH PROJECTION(F)			
											ONE		TWO OR MORE	
	mm	in.	mm	in.	mm	in.		n	lbf		mpa	ksi	mpa	ksi

31	0.279	0.011	7.87	0.31	6.35	0.25	6	490	110	5200	1.3	0.19	1.0	0.14
26	0.483	0.019	7.87	0.31	6.35	0.25	6	1000	225	5400	2.8	0.40	1.9	0.28

Source: Ref 1

(A) SAE 1010, LOW-CARBON STEEL, 290-380 MPA (42-55 KSI) ULTIMATE TENSILE STRENGTH.
(B) SURFACE OF STEEL MAY BE OILED LIGHTLY, BUT MUST BE FREE FROM GREASE, SCALE, AND DIRT.
(C) ELECTRODE FORCE CONTAINS NO FACTOR TO FURTHER FORM POORLY MADE PARTS.
(D) BASED ON 60 HZ.
(E) STARTING VALUES SHOWN ARE BASED ON EXPERIENCE OF RWMA MEMBER COMPANIES.
(F) APPROXIMATE STRENGTH PER PROJECTION DEPENDS ON JOINT DESIGN.

Dissimilar Thickness Joints. As previously noted, the PW process is ideally suited for the joining of sheets of widely disparate thickness. When fabricating such joints, the best practice is to place the projection on the thicker sheet, where possible, using the projection design that is appropriate for the thinner sheet. This is done because the projection has the effect of concentrating the heat at the contact surface, regardless of the thicknesses of the sheets being joined. As a result, selecting conditions based on the thinner sheet has the effect of sizing the weld for that (attached) material.

The solid projections that are described below include nut, cross-wire, and annular projection configurations.

Nut Welding. Specific criteria for the design of weld nut projections, as well as appropriate processing conditions, are not well established. However, there is a general consensus that the ideal number of projections on a weld nut is three, because it is assumed that all three will contact the opposing surface when the electrode force is applied, if the projections are of the same height and if the electrodes are aligned properly.

The actual design of the welding projections is not well established, although 90° included angle square pyramidal and hemispherical projections are common. Such projections are nominally scaled according to the opposing sheet. Designs for embossed projections of the appropriate sheet size can be used as a guideline.

Cross-Wire Welding. The process conditions for the cross-wire welding of a range of wire thicknesses are given in Table 6. The conditions are usually specified in terms of requirements to achieve a certain level of compression set-down for the wires. Set-down is defined in Fig. 5. In general, joint strength increases as the degree of set-down increases. The defined process conditions are for mild steels. Higher-strength materials, such as stainless steels or nickel-base alloys, will require higher forces and longer welding times to promote wire collapse and diffusion across the bond line.

TABLE 6 PROCESS REQUIREMENTS FOR CROSS-WIRE WELDING OF HOT- AND COLD-DRAWN STEEL WIRES IN A RANGE OF THICKNESSES

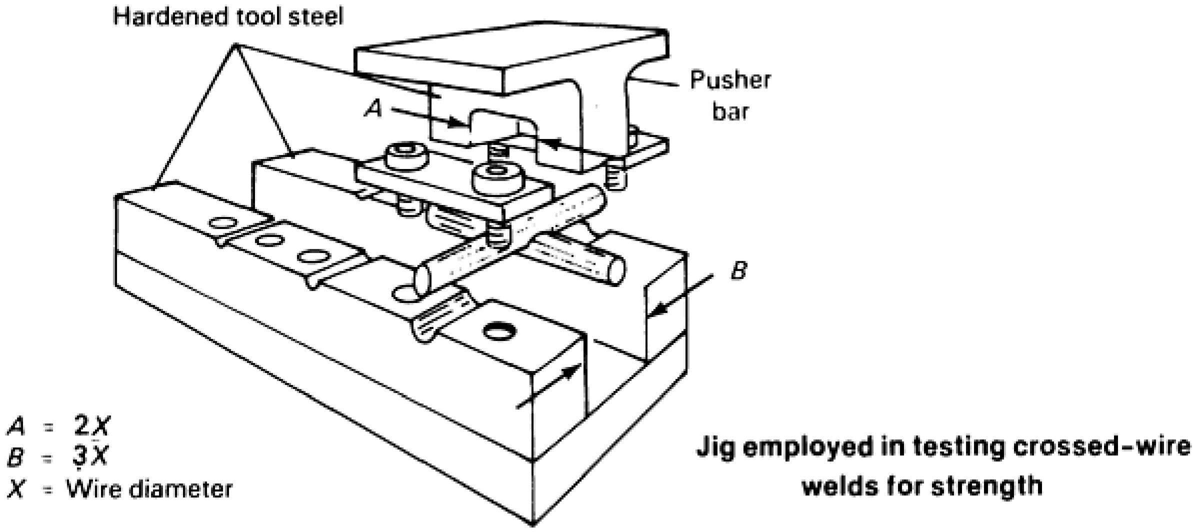

Jig employed in testing crossed-wire welds for strength

COLD-DRAWN WIRE								HOT-DRAWN WIRE						
WIRE DIAMETER		WELD TIME(A), CYCLES	WELD FORCE		WELDING CURRENT(B), A	WELD STRENGTH		WELD TIME(A), CYCLES	WELD FORCE		WELDING CURRENT(B), A	WELD STRENGTH		
mm	in.		kn	lbf		mpa	ksi		kn	lbf		mpa	ksi	
15% SETDOWN														
0.63	$\frac{1}{16}$	5	0.45	100	600	3.1	0.45	5	0.45	100	600	2.4	0.35	
3.18	$\frac{1}{8}$	10	0.56	125	1,800	6.8	0.98	10	0.56	125	1,850	5.2	0.75	
4.76	$\frac{3}{16}$	17	1.60	360	3,300	13.8	2.0	17	1.60	360	3,500	10.3	1.5	
6.35	$\frac{1}{4}$	23	2.58	580	4,500	25.5	3.7	23	2.58	580	4,900	19.3	2.8	
7.94	$\frac{5}{16}$	30	3.67	825	6,200	35.2	5.1	30	3.67	825	6,600	31.7	4.6	
9.53	$\frac{3}{8}$	40	4.90	1,100	7,400	46.2	6.7	40	4.90	1,100	7,700	42.7	6.2	
11.1	$\frac{7}{16}$	50	6.22	1,400	9,300	66.2	9.6	50	6.22	1,400	10,000	60.7	8.8	
12.7	$\frac{1}{2}$	60	7.57	1,700	10,300	84.1	12.2	60	7.57	1,700	11,000	79.3	11.5	
30% SETDOWN														
0.63	$\frac{1}{16}$	5	0.67	150	800	3.4	0.500	5	0.67	150	800	2.8	0.40	
3.18	$\frac{1}{8}$	10	1.16	260	2,650	7.6	1.1	10	1.16	260	2,770	5.9	0.85	
4.76	$\frac{3}{16}$	17	2.67	600	5,000	16.5	2.4	17	2.67	600	5,100	11.7	1.7	
6.35	$\frac{1}{4}$	23	3.78	850	6,700	29.0	4.2	23	3.78	850	7,100	20.7	3.0	
7.94	$\frac{5}{16}$	30	6.45	1,450	9,300	42.1	6.1	30	6.45	1,450	9,600	34.5	5.0	
9.53	$\frac{3}{8}$	40	9.17	2,060	11,300	57.9	8.4	40	9.17	2,060	11,800	46.9	6.8	
11.1	$\frac{7}{16}$	50	12.9	2,900	13,800	77.9	11.3	50	12.9	2,900	14,000	66.2	9.6	

12.7	$\frac{1}{2}$	60	15.1	3,400	15,800	93.8	13.6	60	15.1	3,400	16,500	85.5	12.4
50% SETDOWN													
0.63	$\frac{1}{16}$	5	0.89	200	1,000	3.8	0.55	5	0.89	200	1,000	3.1	0.45
3.18	$\frac{1}{8}$	10	1.56	350	3,400	9.0	1.3	10	1.56	350	3,500	6.2	0.90
4.76	$\frac{3}{16}$	17	3.34	750	6,000	17.2	2.5	17	3.34	750	6,300	12.4	1.8
6.35	$\frac{1}{4}$	23	5.52	1,240	8,600	30.3	4.4	23	5.52	1,240	9,000	21.4	3.1
7.94	$\frac{5}{16}$	30	8.90	2,000	11,400	44.8	6.5	30	8.90	2,000	12,000	36.5	5.3
9.53	$\frac{3}{8}$	40	13.4	3,000	14,400	60.7	8.8	40	13.4	3,000	14,900	49.6	7.2
11.1	$\frac{7}{16}$	50	19.8	4,450	17,400	82.1	11.9	50	19.8	4,450	18,000	70.3	10.2
12.7	$\frac{1}{2}$	60	23.6	5,300	21,000	100.7	14.6	60	23.6	5,300	22,000	89.6	13.0

Source: Ref 1

(A) 60 HZ.

(B) STARTING VALUES SHOWN ARE BASED ON EXPERIENCE OF RWMA MEMBER COMPANIES.

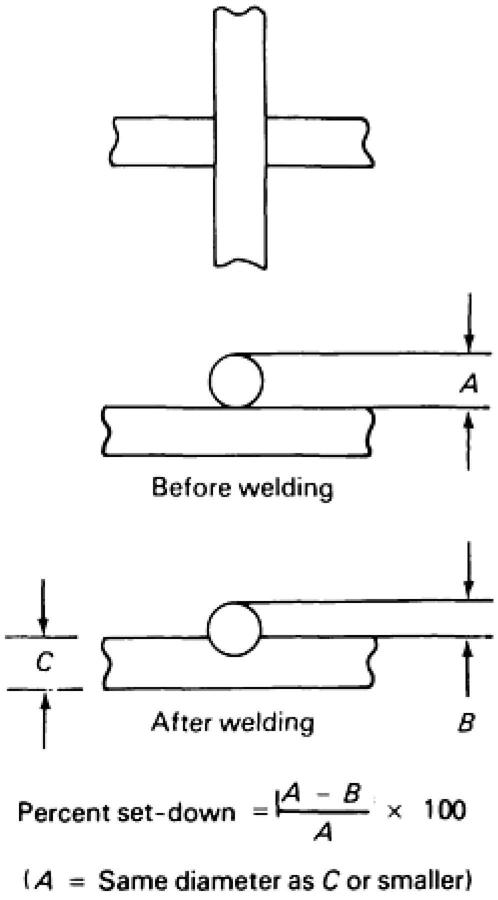
A
Before welding
C
After welding
B
Percent set-down = (A − B)/A × 100
(A = Same diameter as C or smaller)

FIG. 5 DEFINITION OF HOW SET-DOWN IS ESTIMATED ON CROSS-WIRE WELDS. SOURCE: REF 1

Annular Projection Welding. Neither specific projection designs nor process conditions for annular projection welding are well established. There is a general consensus on the use of 90 ° included angle triangular cross section projections in sizes that are scaled to the thickness of the opposing sheet. Typical practice includes approximately 4.5 to 8.9 kN (1000 to 2000 lbf) of force per linear inch (or 25 mm) of projection. Weld times for annular projections for diameters ranging from 6.4 to 25 mm (0.25 to 1.0 in.) usually range from a single cycle to about 10 cycles. Typically, longer welding times result in more-robust processing conditions. However, maximum welding times are often limited by concerns over peripheral thermal damage.

The best annular projection design will recess the base diameter of the projection from the outside diameter of the part. Reasons for this are shown in Fig. 6. Projection designs that extend to the outside diameter of the part have the potential for unstable projection collapse, which can result in the formation of an incipient notch at the outside diameter of the part and an increase in the incidence of weld-related failures.

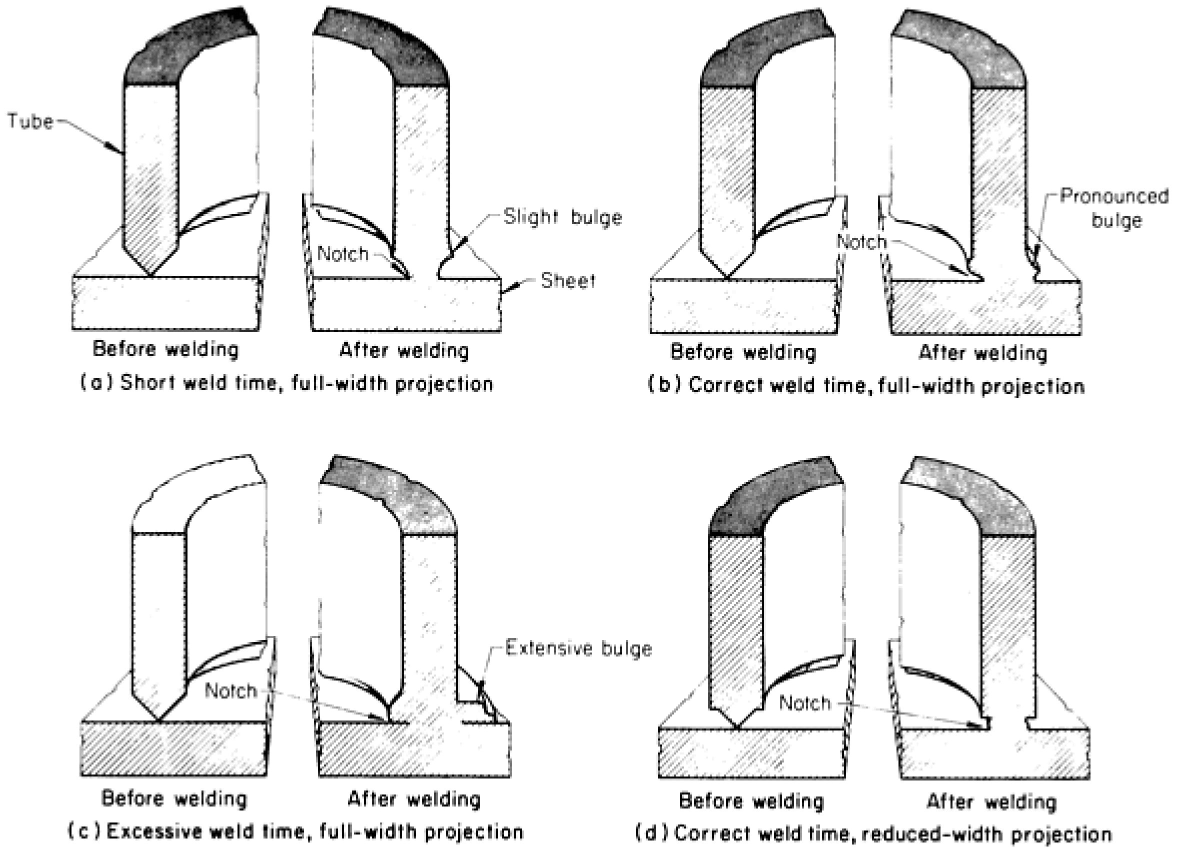
Tube
Slight bulge
Notch
Sheet
Before welding
After welding
(a) Short weld time, full-width projection
Pronounced bulge
Notch
Before welding
After welding
(b) Correct weld time, full-width projection
Extensive bulge
Notch
Before welding
After welding
(c) Excessive weld time, full-width projection
Notch
Before welding
After welding
(d) Correct weld time, reduced-width projection

FIG. 6 CHARACTERISTICS OF PROJECTION COLLAPSE DURING ANNULAR PROJECTION WELDING WITH DIFFERENT BASE PROJECTION WIDTHS. (A) SHORT WELD TIME, FULL-WIDTH PROJECTION. (B) CORRECT WELD TIME, FULL-WIDTH PROJECTION. (C) EXCESSIVE WELD TIME, FULL-WIDTH PROJECTION. (D) CORRECT WELD TIME, REDUCED-WIDTH PROJECTION

Larger-diameter annular projection welds do cause a problem in attaining the proper heat balance around the periphery of the projection. Typically, the side of the projection that is toward the throat of the machine will heat preferentially for two reasons. First, the shortest conductive (and highest current flow) path is along the inside of the throat, including the inboard side of the projection. Second, there is an inductive "skin" effect associated with alternating-current machines, which tends to concentrate current along the inside of the welder throat. Intermediate-diameter projections have been successfully welded using direct current, because it minimizes the skin effect and appears to allow better heat balance.

Larger-diameter projections may require dual-throat machines, which have dual transformers located on opposite sides of the workpiece. Each has its own "throat." The effect is to balance the resistive/inductive effects, allowing the welding of very large diameter projections.

Reference cited in this section

1. D. BENETEAU, ED., *RESISTANCE WELDING MANUAL,* 4TH ED., G.H. BUCHANAN CO., 1989

Projection Welding

Jerry E. Gould, Edison Welding Institute

Reference

1. D. BENETEAU, ED., *RESISTANCE WELDING MANUAL,* 4TH ED., G.H. BUCHANAN CO., 1989

Resistance Seam Welding

Michael J. Karagoulis, General Motors Corporation

Introduction

RESISTANCE SEAM WELDING (RSEW) is a process in which heat generated by resistance to the flow of electric current in the work metal is combined with pressure to produce a welded seam. The resulting seam consists of a series of spot welds. The specific seam weld process can be classified as:

- *ROLL SPOT WELDING* (RELATIVELY LARGE UNWELDED GAPS BETWEEN NUGGETS)
- *REINFORCED ROLL SPOT WELDING* (SMALL GAPS BETWEEN NUGGETS)
- *LEAK-TIGHT SEAM WELDING* (NUGGET OVERLAP)

Two rotating circular electrode wheels are often used to apply current, force, and cooling to the work metal. A variety of workpiece/wheel configurations are possible (Fig. 1). When two electrode wheels are used, one or both wheels are driven, either by a direct drive of the wheel axles or by a knurl drive that contacts the peripheral surface of the electrode wheel (Fig. 2). For some applications, the electrode wheels idle while the workpiece is driven.

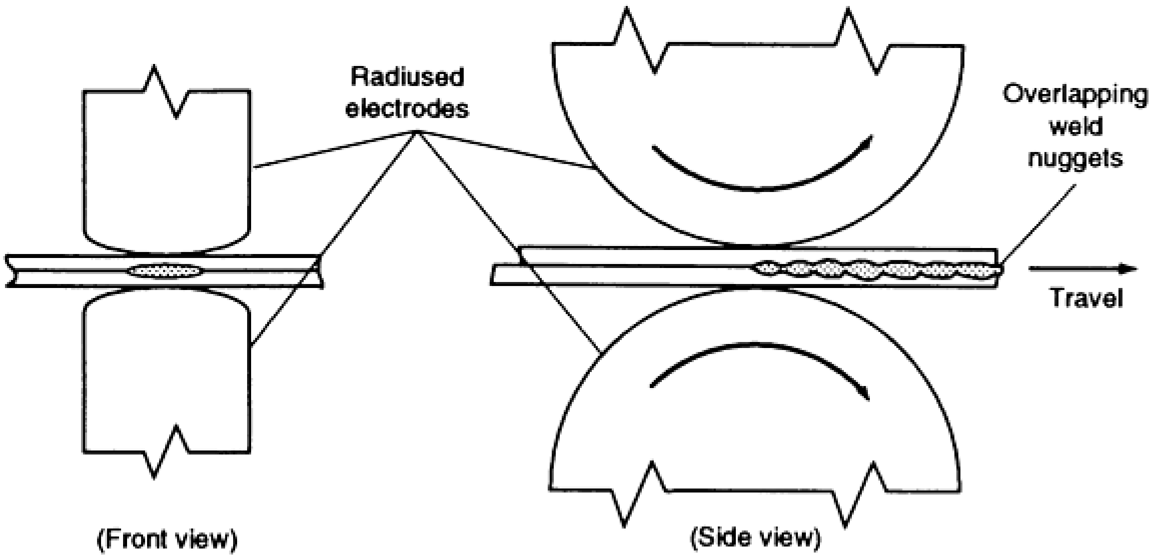
Radiused electrodes
Overlapping weld nuggets
Travel
(Front view)
(Side view)

(a) Lap seam weld

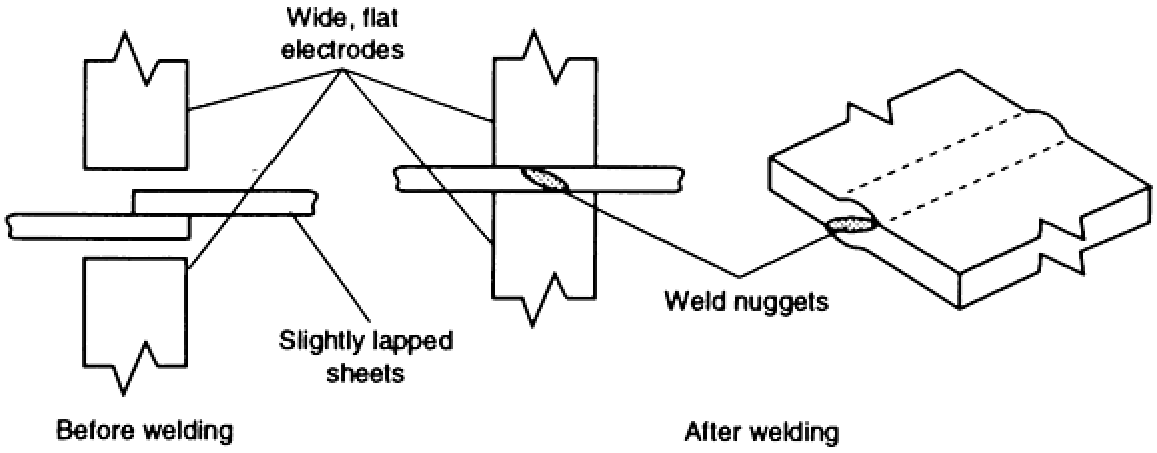
Wide, flat electrodes
Weld nuggets
Slightly lapped sheets
Before welding
After welding

(b) Mash seam weld (front view)

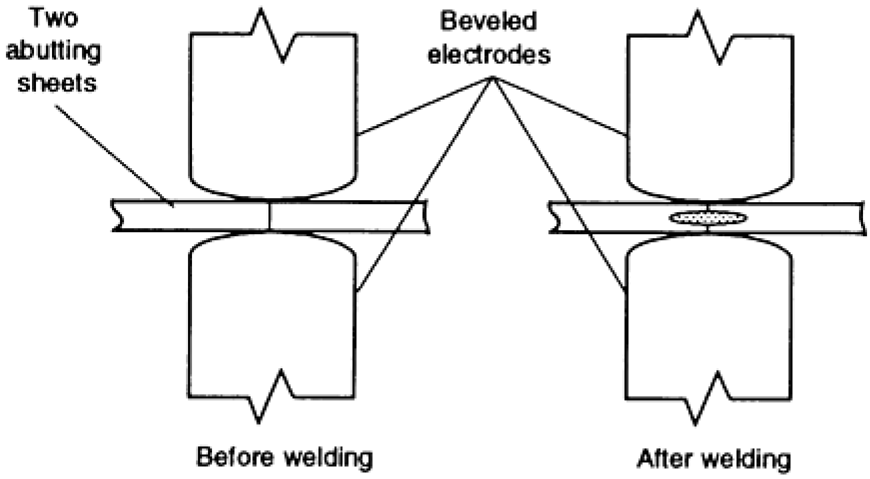
Two abutting sheets
Beveled electrodes
Before welding
After welding

(c) Butt seam weld (front view)

FIG. 1 BASIC CONFIGURATION OF JOINTS AND RESULTANT WELDS FORMED IN RESISTANCE SEAM WELDING. ADAPTED FROM *WELDING HANDBOOK*, 8TH ED., VOL 2, AMERICAN WELDING SOCIETY, 1991, P 554

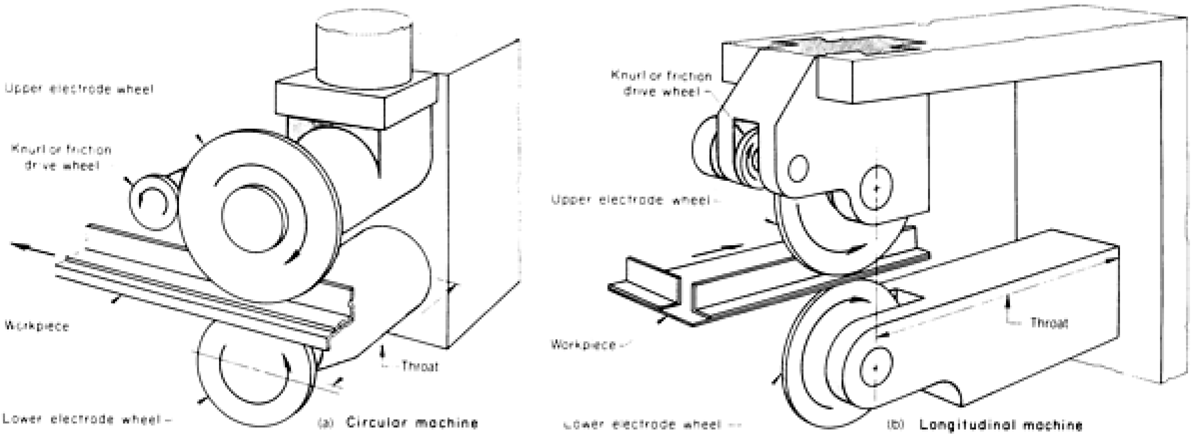
Upper electrode wheel
Knurl or friction drive wheel
Workpiece
Throat
Lower electrode wheel
Knurl or friction drive wheel
Upper electrode wheel
Throat
Workpiece
Lower electrode wheel

(a) Circular machine

(b) Longitudinal machine

FIG. 2 POSITION OF ELECTRODE WHEELS RELATIVE TO THROAT ON RESISTANCE SEAM WELDING MACHINES

Other electrode systems for seam welding may use one wheel and one flat bar-type electrode (Fig. 3), or a wire-feed system in which a copper wire is fed into a groove on the wheel. This wire-feed process continuously provides a new copper surface for contact with the work metal (Fig. 4).

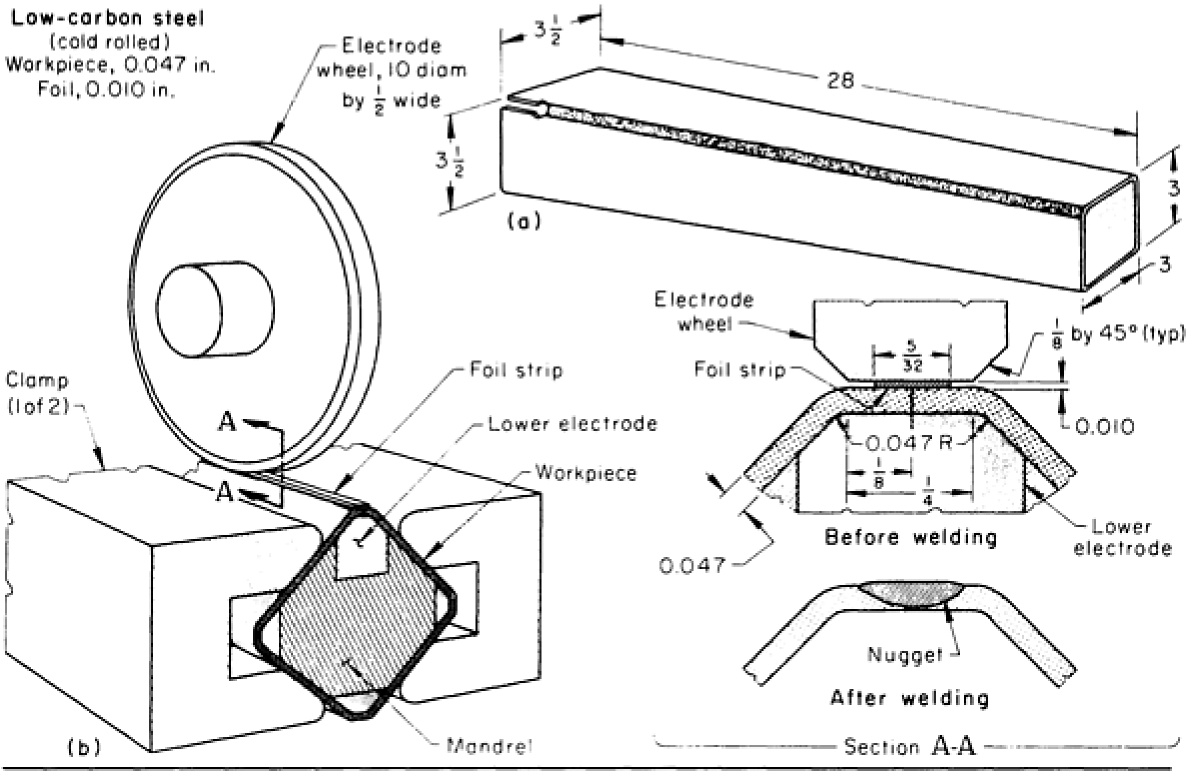
Low-carbon steel
(cold rolled)
Workpiece, 0.047 in.
Foil, 0.010 in.
Electrode wheel, 10 diam by 1/2 wide
3 1/2
28
3 1/2
3
3
(a)
Electrode wheel
1/8 by 45° (typ)
Foil strip
5/32
0.010
0.047 R
1/8
1/4
Before welding
Lower electrode
0.047
Nugget
After welding
Section A-A
Clamp (1 of 2)
A
A
Foil strip
Lower electrode
Workpiece
Mandrel
(b)

EQUIPMENT SPECIFICATIONS AND WELDING PARAMETERS	
POWER SUPPLY	440 V, SINGLE PHASE
WELDING MACHINE	AUTOMATIC, LONGITUDINAL, WITH MOVABLE CARRIAGE FOR UPPER ELECTRODE
RATING AT 50% DUTY CYCLE	300 KVA

HEAT CONTROL	PHASE SHIFT
ELECTRODE-RAM OPERATION	AIR CYLINDER
ELECTRODE FORCE, MAX	5.4 KN (1200 lbf)
UPPER ELECTRODE	13 MM ($\frac{1}{2}$ in.) WIDE BY 255 MM (10 IN.) DIAM; DOUBLE BEVEL; 6.4 MM ($\frac{1}{4}$ in.) WIDE FIAT FACE
LOWER ELECTRODE	710 mm (28 in.) LONG BAR; DOUBLE BEVEL; 6.4 MM ($\frac{1}{4}$ in.) WIDE FLAT FACE (SEE DRAWING)
ELECTRODE MATERIAL	RWMA CLASS 2
WELDING CURRENT	CONTINUOUS
WELDING SPEED	6 M/MIN (20 FT/MIN)
PRODUCTION RATE	220 LEGS PER HOUR

FIG. 3 SETUP FOR FOIL BUTT SEAM WELDING OF A TABLE LEG MADE OF LOW-CARBON STEEL. DIMENSIONS GIVEN IN INCHES

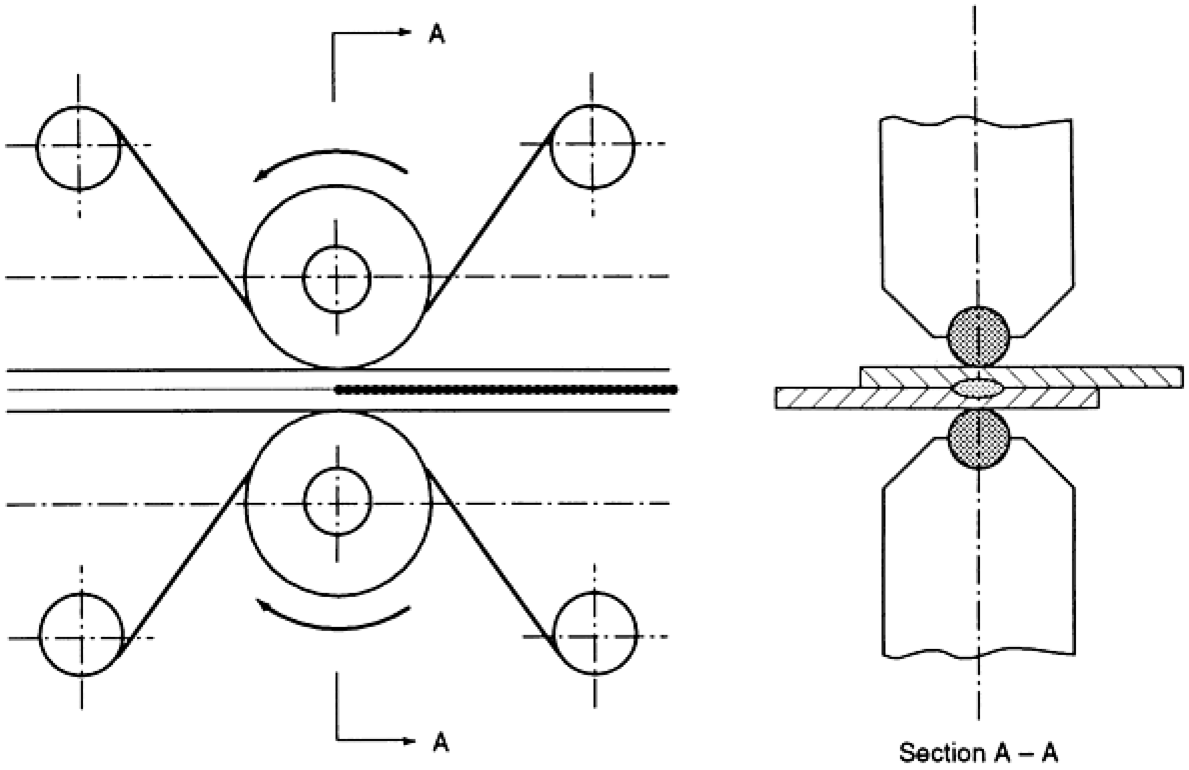

FIG. 4 TYPICAL SETUP FOR AN ELECTRODE WIRE SEAM WELDING PROCESS SHOWING POSITIONS OF WIRE RELATIVE TO GROOVED WHEEL

In seam welding, a series of spot welds is made without retracting the electrode wheels or releasing the electrode force between spots. The electrode wheels may rotate either continuously or intermittently. Weld speed, current magnitude, current waveform, cooling system, and the electrode parameters of force, shape, and diameter are all related and must be carefully selected to optimize the process and produce the highest quality weld.

The principles described in the article "Resistance Spot Welding" in this Volume are also generally applicable to seam welding.

Resistance Seam Welding

Michael J. Karagoulis, General Motors Corporation

RSEW Applications

Lap Seam Weld. A joint in which the workpieces are overlapped sufficiently to prevent the sheet edges from becoming part of the weld is classified as a lap seam weld (Fig. 1a). Lap seams are popular in automotive applications, such as automotive fuel tanks, catalytic converters, mufflers, and roof joints, as well as in nonautomotive applications, such as furnace heat exchangers, water tanks, and certain types of can making (Fig. 5). Lap seam welding of multiple stackups and dissimilar thicknesses is also possible.

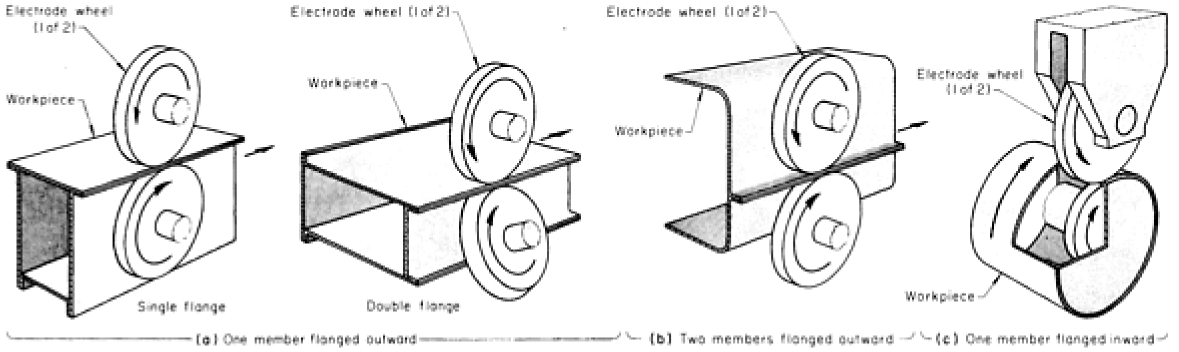

FIG. 5 ARRANGEMENT OF ELECTRODE WHEELS RELATIVE TO WORKPIECE FOR SELECTED FLANGE-JOINT LAP SEAM WELDS

Mash Seam Weld. A joint in which two sheets are overlapped by only one to two times the sheet thickness is classified as a mash seam weld (Fig. 1b). The weld area is forged or mashed down during welding to an overthickness 5 to 25% greater than a single sheet thickness. It is necessary to have some amount of overthickness so the current path can be controlled, because the electrode wheels in mash seam welding are wide and flat-edged. Surface appearance and thickness can be improved through the use of hard steel planishing wheels to cold work the joint before or after welding. Because of the excessive hardening induced by cold work, postweld heat treatment may be necessary (Ref 1).

Typical applications include drums, buckets, vacuum-jacketed bottles, aerosol cans, water tanks, and steel mill coil joining. Sheets of dissimilar thicknesses and/or coatings have been successfully mash seam welded. This development has created a whole new set of applications in tailored blank manufacturing, primarily for automotive use (Ref 2).

Butt Seam Weld. A joint in which two abutting edges are welded is classified as a butt seam joint (Fig. 1c and 3). The thickness of the weld should be approximately the same, or slightly less than, the sheet thickness. Butt seam welding is typically reserved for applications in which other butt welding processes cannot be used (for example, for tube welding and for sheet metal in railroad cars).

References cited in this section

1. K.M. PICKETT, M.F. SHI, K.K. BHATT, AND J. SELVARAJ, STRUCTURE PROPERTY RELATIONSHIPS OF RESISTANCE MASH SEAM WELDED TAILORED BLANKS WITH POST HEAT TREATMENT, *24TH CONF. INT. METALLOGRAPHIC SOC.* (DENVER), AUG 1992

2. Y. ADONYI, J. MILIAN, AND L. ROUDABUSH, WELDED STEEL SHEET PERFORMANCE FOR TAILORED BLANKS, *PROC. SHEET METAL WELDING CONF. V*, PAPER C1, AMERICAN WELDING SOCIETY, OCT 1992

Resistance Seam Welding

Michael J. Karagoulis, General Motors Corporation

Advantages and Limitations

Advantages of seam welding, compared to resistance spot welding, projection welding, and laser welding, are:

- GAS-TIGHT OR LIQUID-TIGHT JOINTS CAN BE PRODUCED (NOT POSSIBLE WITH SPOT WELDING OR PROJECTION WELDING).
- SEAM WIDTH MAY BE LESS THAN THE DIAMETER OF SPOT WELDS, BECAUSE THE ELECTRODE CONTOUR CAN BE CONTINUOUSLY DRESSED AND IS THEREFORE OF A STABLE SHAPE.
- HIGH-SPEED WELDING (ESPECIALLY ON THIN STOCK) IS POSSIBLE.
- TOOLING COST IS GENERALLY FAVORABLE PER INCH OF FLANGE WELDED.
- COATED STEELS ARE GENERALLY MORE WELDABLE USING SEAM WELDING THAN SPOT WELDING, BECAUSE COATING RESIDUE CAN BE CONTINUOUSLY REMOVED FROM THE ELECTRODE WHEELS IF SPECIAL PROVISIONS ARE MADE.
- COATED STEELS ARE GENERALLY MORE WELDABLE USING SEAM WELDING THAN LASER WELDING, BECAUSE COATING VOLATILITY IS MINIMIZED BY THE INTENSE PRESSURE FIELD IN THE WELD ZONE.
- RESISTANCE SEAM WELDING IS NOT PARTICULARLY FITUP-SENSITIVE COMPARED TO LASER WELDING.
- THE HARDNESS OF RESISTANCE SEAM WELDS MADE WITH AIR COOLING IS LESS THAN THAT OF LASER WELDS (120 HV VERSUS 250 HV, RESPECTIVELY, FOR DRAWING-QUALITY BARE STEEL, AND 170 HV VERSUS 300 HV, RESPECTIVELY, FOR ORGANIC COATED DRAWING-QUALITY STEEL) (REF 3).

Limitations of seam welding, apart from those it shares with spot welding and projection welding, are:

- WELDS MUST ORDINARILY PROCEED IN A SINGLE PLANE OR ON A UNIFORMLY CURVED SURFACE. FOR SPECIAL CASES, MULTIPLANE WELDING HAS BEEN ACHIEVED USING A TILTING WELDING HEAD. THE RADIUS OF CURVATURE CONNECTING THE PLANES MUST BE SUFFICIENTLY LARGE.
- OBSTRUCTIONS ALONG THE PATH OF THE ELECTRODE WHEEL MUST BE AVOIDED OR COMPENSATED FOR IN THE DESIGN OF THE WHEEL.
- MATERIAL HANDLING MUST NOT INDUCE EXTRANEOUS FORCES INTO THE FRAGILE, MOLTEN WELD ZONE DURING WELDING.
- THE LENGTH OF SEAMS MADE IN A LONGITUDINAL SEAM WELDING MACHINE IS LIMITED BY THE THROAT DEPTH OF THE MACHINE (FIG. 2B). CURRENT SHUNTING THROUGH THE WELD, OR A CHANGE IN ELECTRICAL IMPEDANCE CAUSED BY THE MOVEMENT OF THE PRODUCT INSIDE THE THROAT OF THE MACHINE, CAN REQUIRE CURRENT COMPENSATION AS WELDING PROCEEDS.
- COMPONENTS USING MULTIPLE CROSSING SEAM WELDS CAN BE QUALITY-SENSITIVE AT THE WELD INTERSECTIONS.
- EXTERNAL WATER COOLING OF THE ELECTRODES AND THE WELD ZONE MAY BE REQUIRED FOR HIGH-SPEED WELDING. (UNDER HIGH-SPEED CONDITIONS THE WELD

NUGGETS ARE STILL MOLTEN AS THEY LEAVE THE PRESSURE FIELD OF THE WHEELS). EXTERNAL COOLING MAY ADD TOOLING COST FOR WATER CONTAINMENT AND WATER REMOVAL FROM THE PARTS AFTER WELDING.

- MATERIALS REQUIRING POSTHEAT AND TEMPER SHOULD NOT BE SEAM WELDED WITHOUT SPECIAL CONSIDERATIONS.
- CHROMATES AND INSULATING LACQUER COATINGS ARE NOT RESISTANCE SEAM WELDABLE.

Reference cited in this section

3. VDI REPORT 818, VEREIN DEUTSCHER INGENIEURE, 1990, P 256 (IN GERMAN)

Resistance Seam Welding

Michael J. Karagoulis, General Motors Corporation

Fundamentals of Lap Seam Welding

Empirical seam welding parameters have been developed over the years and have proven useful under many conditions (Ref 4, 5). However, this discussion is devoted to an alternative approach to process optimization, based on principles of heat transfer in the weld. Hopefully, this approach will shed some light on difficult applications in which traditional methods may fall short of what is necessary to make quality welds.

Weld Time, Speed, and Current Pulsation

Certain concepts from resistance spot welding begin to explain resistance seam welding, because seam welding generally consists of a series of overlapping spot welds.

Spot Weld Analogy. Empirically, the optimum time for the spot welding of low-carbon steel sheet is approximately (Ref 6):

$$T = 0.0292\ S^2 \qquad \textbf{(EQ 1)}$$

where S is the thickness of the total stackup of sheets (in millimeters) and t is the optimum weld time (in seconds).

Applying current for this length of time maximizes the current range from minimum weld diameter through the onset of either cracking, overindentation, or electrode sticking. If weld time is less than this, there will be insufficient time for proper heat flow to occur inside the weld. As such, the weld will begin to exhibit adiabatic character ("thermal runaway"), nugget size instability, and surface expulsion. By the same measure, if weld time is too long, the extra cycles simply heat up the electrodes. Hot electrodes are less capable of keeping the surface of the weld cool to avoid electrode sticking and metallurgical discontinuities on the weld surface. In addition, excessive weld time will also compromise electrode life, waste energy, and could even slow down the production rate of an entire manufacturing line. Thus in any case, for spot welding there seems to be a proper rate of welding, based on material thickness and thermal diffusivity (in the thickness direction).

Application to Seam Welding. For seam welding, the same principle of optimum weld time applies, but weld time refers to how long a given point on the weld will be in contact with the wheel as it passes through the welder. In practice, travel speed and electrode footprint length, *L*, should be controlled so that the resulting weld time is matched with the 0.0292 S^2 rule (see Eq 1, Eq 5, and Fig. 6). The following equations may be used (forward or backward) to fix some variables and calculate the others that result from these assumptions.

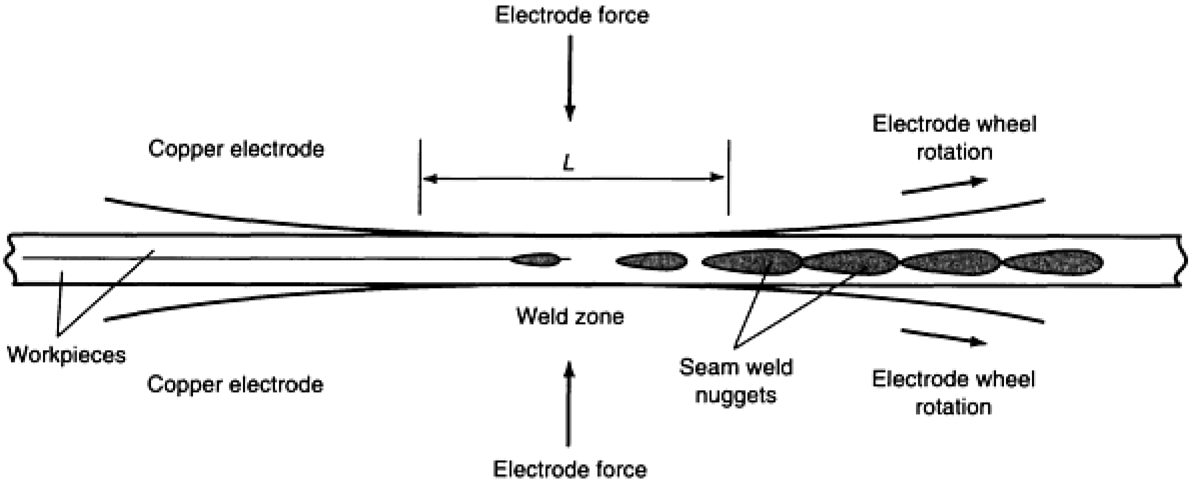
Electrode force
Copper electrode
L
Electrode wheel rotation
Workpieces
Weld zone
Seam weld nuggets
Copper electrode
Electrode wheel rotation
Electrode force

FIG. 6 SCHEMATIC SHOWING LENGTHWISE SECTIONAL VIEW OF PROPER NUGGET DEVELOPMENT IN THE WELD ZONE OF A PULSED LAP SEAM WELD. THE LENGTH OF THE CONTACT FOOTPRINT, *L,* IS SUFFICIENT TO CONTAIN THREE NUGGETS UNDER THE WHEEL (THAT IS, *N* = 3).

For lap seams of low-carbon sheets, Eq 1 becomes:

$$T = N/F = 0.0292\ S^2 \qquad \textbf{(EQ 2)}$$

where *n* is the number of nuggets under development inside the weld zone footprint at any one time (Fig. 6) and *f* is the impulse frequency of the applied current (on-time and off-time).

For process control and leak-tight welds, a good rule for the nugget spacing is:

$$1.7\ S \leq D \leq 2.5\ S \qquad \textbf{(EQ 3)}$$

where *d* is the nugget spacing (in mm).

Welding speed, *v* (in mm/s), is therefore:

$$V = D \cdot F \qquad \textbf{(EQ 4)}$$

However, to predict the optimum welding speed, one must first know *L*, the length of the contact footprint (in mm):

$$L = D \cdot N \qquad \textbf{(EQ 5)}$$

One can calculate the required *L* from *d* and *n,* and then determine the wheel diameter, force, and current needed to achieve *L.*

Substituting Eq 5 and Eq 3 into Eq 4:

$$V = L \cdot F/N \qquad \textbf{(EQ 6)}$$

and

$$L_{\mathrm{MAX}} = 2.5\, N \cdot S \qquad \textbf{(EQ 7)}$$

Rearranging Eq 2:

$$N = 0.0292\, F \cdot S^2 \qquad \textbf{(EQ 8)}$$

and:

$$V_{\mathrm{MAX}} = L_{\mathrm{MAX}} \cdot F/N \qquad \textbf{(EQ 9)}$$

to obtain:

$$V_{\mathrm{MAX}} = 85.6\, N/S \qquad \textbf{(EQ 10)}$$

It is important to note that the current range of seam welding is widened by the judicious use of impulsed current to form distinct nuggets. Impulse profiles of 25 to 33% off-times are commonly used. Impulses contain off-time for the electrodes to draw heat away from the weld surface. Therefore, the occurrence of surface discontinuities is lessened. In addition, off-time allows heat to redistribute itself within the weld, away from local hot spots. Hot spots are places within the weld where thermal spikes may be occurring due to either a relatively high local resistance, a nonuniform current density, a multiple stackup, or a stackup of unequal sheets. Hot spots are especially common along the weld interface when welding either thick plates or prepainted sheets. Hot spots on the outside surface of a weld occur naturally with unequal sheet thicknesses (on the thin side), or when welding is faster than permitted by Eq 2. On the other hand, hot spots can also occur within the sheet bulk, if due to erratic time distribution of current during on-time. Bulk hot spots begin to occur when welding near the maximum speed for a material, because current is actually not applied uniformly throughout the on-time, and time is required for the resulting nonuniform heat pattern to redistribute itself (Fig. 7). The ideal impulse shape under high-speed conditions would seem to be a square wave (Fig. 8a). Because typical industrial power sources can only approximate this shape (Fig. 8b), it is worthwhile to accurately record the waveform of the actual secondary current in order to monitor its shape and stability during welding.

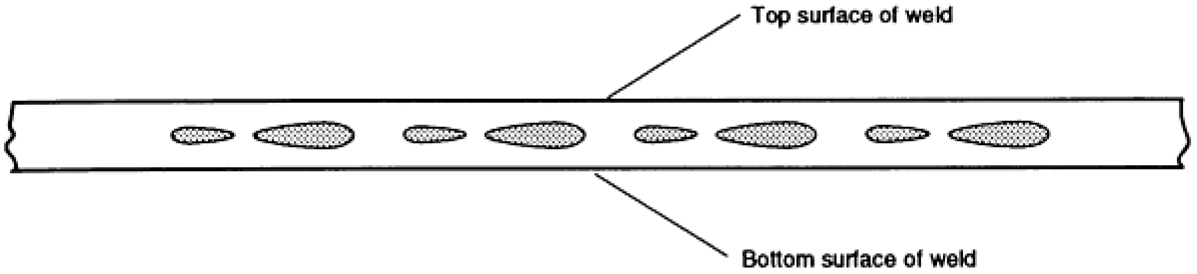

FIG. 7 SCHEMATIC SHOWING LENGTHWISE SECTIONAL VIEW OF THE WELD CENTERLINE OF A LAP SEAM JOINT WITH IRREGULAR NUGGET SHAPES. THE LACK OF UNIFORMITY IN THE SHAPE OF THE NUGGETS IS ATTRIBUTABLE TO AN IRREGULAR PULSE SHAPE THAT DOES NOT PROVIDE SUFFICIENT TIME FOR HEAT FLOW TO BLEND TOGETHER THE DOUBLETS.

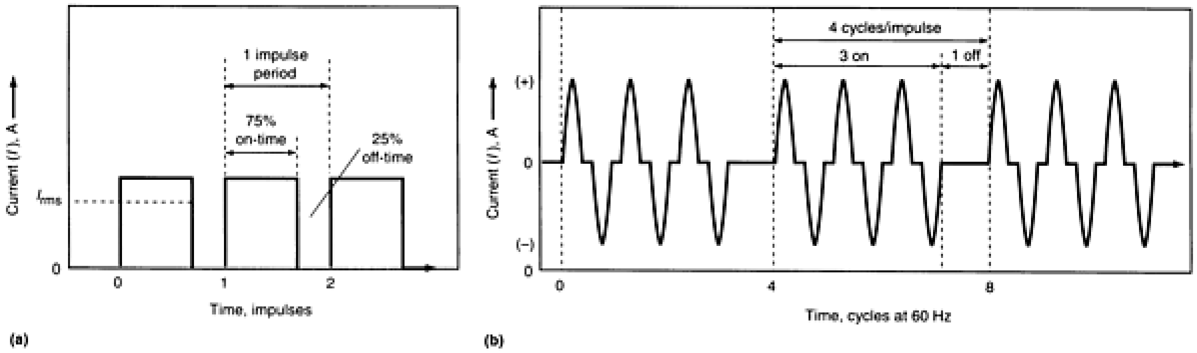

FIG. 8 WELD CURRENT PROFILES FOR SEAM WELDING. (A) IDEAL IMPULSE SHAPE. ON-TIME IS 75% OF IMPULSE PERIOD WHILE OFF-TIME IS 25% OF IMPULSE PERIOD. (B) ACTUAL IMPULSE SHAPE IS TYPICALLY COMPLEX AND NON-IDEAL. IN THIS SPECIFIC CASE, IMPULSES ARE FROM AN AC SINGLE-PHASE (3 ON/1 OFF) CURRENT SOURCE.

Current range is greater in impulse mode than in nonimpulse (continuous current) mode. The extra range seems to provide more robust penetration under difficult welding conditions, such as when intersecting other welds or turning corners with the seam.

One of the disadvantages of seam welding relative to spot welding is that solidification of the weld nugget occurs after the weld metal leaves the pressure field of the electrode wheels. Thus, if there is a large separating force on the weld before it freezes, it can come apart as it leaves the wheels. This postweld separation has been called "unzipping" or "alligatoring." Impulses effectively cure this problem because the off-times create a modulating penetration profile (Fig. 6), providing cool tack joints between the molten spots.

It is desirable from a process control standpoint to select an impulse schedule that will produce nuggets of nearly equal length and width, in conjunction with the other parameters used (speed, wheel diameter, wheel width, and power source).

In Table 1, theoretical parameters from Eq 2, 3, 4, 5, 6, 7, 8, 9, and 10 have been calculated for several commonly available impulse frequencies. The results can be used to predict optimum welding parameters from these basic principles.

TABLE 1 THEORETICAL PARAMETERS FOR LAP SEAM WELDING OF TWO EQUAL SHEETS OF MILD STEEL (PER EQ 2, 3, 4, 5, 6, 7, 8, 9, 10)

IMPULSED CURRENT (I)			**UNCOATED(A), ONE SPOT UNDER WHEEL (N = 1)**						**COATED OR UNCOATED(A)**											
									TWO SPOTS UNDER WHEEL (N = 2)						**THREE SPOTS UNDER WHEEL (N = 3)**					
			STACKUP (S_{MIN})		**WELD ZONE LENGTH (L)**		**TRAVEL SPEED (V)**		**STACKUP (S_{MIN})**		**WELD ZONE LENGTH (L)**		**TRAVEL SPEED (V)**		**STACKUP (S_{MIN})**		**WELD ZONE LENGTH (L)**		**TRAVEL SPEED (V)**	
Type	**On-time/off-time**	**Impulse frequency (f), hz**	**mm**	**in.**	**mm**	**in.**	**mm/s**	**in./min**	**mm**	**in.**	**mm**	**in.**	**mm/s**	**in./min**	**mm**	**in.**	**mm**	**in.**	**mm/s**	**in./min**
AC	(B)	120	0.53	0.021	1.34	0.053	160	370	0.76	0.030	3.78	0.149	227	535	0.93	0.037	6.94	0.273	278	656
AC	0.33/0.33	90	0.62	0.024	1.54	0.061	139	328	0.87	0.034	4.36	0.172	196	464	1.07	0.042	8.01	0.315	240	568
DC	0.67/0.33	60	0.76	0.030	1.89	0.074	113	268	1.07	0.042	5.34	0.210	160	379	1.31	0.052	9.81	0.386	196	464
AC	1/0.33	45	0.87	0.034	2.18	0.086	98	232	1.23	0.048	6.17	0.243	139	328	1.51	0.059	11.3	0.445	170	402
AC	1/0.67	36	0.98	0.039	2.44	0.096	88	207	1.38	0.054	6.90	0.272	124	293	1.69	0.067	12.7	0.500	152	359
AC	1.33/0.67	30	1.07	0.042	2.67	0.105	80	189	1.51	0.059	7.56	0.298	113	268	1.85	0.073	13.9	0.547	139	328
DC	1/1	30	1.07	0.042	2.67	0.105	80	189	1.51	0.059	7.56	0.298	113	268	1.85	0.073	13.9	0.547	139	328
AC	2/1	20	1.31	0.052	3.27	0.129	65	155	1.85	0.073	9.25	0.364	93	219	2.27	0.089	17.0	0.669	113	268
AC	3/1	15	1.51	0.059	3.78	0.149	57	134	2.14	0.084	10.7	0.421	80	189	2.62	0.103	19.6	0.748	98	232
AC	3/2	12	1.69	0.067	4.22	0.166	51	120	2.39	0.094	11.9	0.468	72	169	2.93	0.115	21.9	0.862	88	207
AC	4/2	10	1.85	0.073	4.63	0.182	46	109	2.62	0.103	13.1	0.516	65	155	3.21	0.126	24.0	0.945	80	189
AC	5/2	8.57	2.00	0.079	5.00	0.197	43	101	2.83	0.111	14.1	0.555	61	143	3.46	0.136	26.0	1.024	74	175
AC	6/2	7.50	2.14	0.084	5.34	0.210	40	95	3.02	0.119	15.1	0.594	57	134	3.70	0.147	27.8	1.094	69	164
AC	6/3	6.67	2.2	0.08	5.6	0.22	38	89	3.2	0.12	16.	0.63	53	126	3.9	0.15	29.	1.15	65	155

			7	9	7	3			0	6	0	0			2	4	4	7		
AC	7/3	6	2.39	0.094	5.97	0.235	36	85	3.38	0.133	16.9	0.665	51	120	4.14	0.163	31.0	1.220	62	147

(A) *N*, NUMBER OF NUGGET SPOTS UNDER THE ELECTRODE WHEEL; *S*, STACKUP, SUM OF BOTH SHEET THICKNESSES; *L*, CONTACT LENGTH OF ELECTRODE ON WORKPIECE.

(B) 60 HZ CONTINUOUS

The optimum weld speed may be identified experimentally as the speed with the widest current range for making good welds. Good welds are nominally free of typical metallurgical discontinuities (for example, surface cracks, gross lack of fusion at the weld interface, or surface melting eruptions).

When welding dissimilar thicknesses, on-time must not be greater than the optimum weld time for the thinnest outer sheet (*TOS*). Mathematically:

$$T = 4(0.0292)\ (TOS)^2 \qquad \textbf{(EQ 11)}$$

where *TOS* is in millimeters.

One way to satisfy the weld time needs of both total stackup, *S* (Eq 1) and *TOS* for thick-thin welding is to raise *n*, the number of nuggets in the weld zone, to 2 or 3. The goal here is to weld for the proper total time dictated by Eq 1, but to do it with pulses of on-time no greater than dictated by Eq 11. To increase *n*, one must carefully consider the electrode wheels as well as the impulse schedule. Even when taking advantage of pulsation, it may become impossible to seam weld extremely dissimilar thicknesses without damaging the thinnest outer sheet. This difficulty should become pronounced for thickness ratios above about 3 to 1.

Wheel Geometry, Weld Force, and Wheel Maintenance

The primary role of wheel geometry and force in seam welding is to control the dimensions of the weld zone or the size of the footprint that the wheel makes on the workpiece while welding.

The length of the footprint, *L*, is an important factor in the weld time equation (see the section "Weld Time, Speed, and Current Pulsation" in this article). The variable *L* is primarily controlled by wheel diameter, force, and current. (As either force or current is increased, the wheel will sink deeper into the work and *L* will increase.) It is worthwhile to measure or estimate *L*. This has been done experimentally using special voltage probes on the electrodes and by an optical silhouette method (Ref 7). A less accurate (but simpler) method is to read the heat pattern at the beginning and end of the weld, in order to estimate *n*, the number of nuggets under the wheel at any one time. For bare steel, *n* can be as low as 1. However, the weldability of coated materials and dissimilar thicknesses improves when *n* is as high as 2 or 3 (see Fig. 6). Keep in mind that the contact length will decrease somewhat as the wheel wears to a smaller diameter. (Normally this change should not be of concern, except possibly when welding beyond the theoretical top speed for a material combination.)

Footprint width is best matched with impulse length, *d*, to produce nearly round nuggets. Width, however, is of secondary importance to impulse length.

A transverse radius on the wheel generally improves the welding range, because under high current conditions it automatically reduces current density by sinking further into the work.

Under given electrode conditions, increasing weld force improves the welding range for speed and helps control cracking.

It is generally true that the current range improves with higher wheel pressure, until longitudinal splits begin to occur at seam starts or stops. On the other hand, at low force, transverse surface cracks may result at the upper end of the current range, especially on coated steels. A good role of thumb is to weld at 80% of the seam-splitting force, or less. In addition, it may be helpful to de-energize the weld force except while welding, so as not to cause undue wheel wear.

For the highest quality joints it may be necessary to lightly machine the wheels as they are used, in order to maintain a constant edge profile. The types of wheel dressing tools are:

- *SCRAPERS* TO REMOVE CONTAMINATION WHEN WELDING COATED MATERIALS
- *WIDTH CUTTERS* TO REMOVE MUSHROOMED COPPER AND CONTROL WHEEL WIDTH
- *KNURLERS* (FIG. 2), TO MAINTAIN THE TRANSVERSE PROFILE ON THE WHEEL AND DRIVE THE WHEELS AS NEEDED

If wheel maintenance on the machine is not feasible, the wheels can be removed and redressed off-line. Alternatively, the wire-electrode seam welding process provides a constant wheel condition and thus eliminates the need for continuous wheel dressing (Fig. 4).

Cooling the Weld

The weld cooling system is very important for process control of seam welding. Its primary role is to minimize thermal damage to the outer surfaces of the weld while the weld interface is being heated above the melting point. The key elements in the success of the weld cooling system are:

- *ELECTRODE MATERIAL* (THERMAL OR ELECTRICAL CONDUCTIVITY)
- *ELECTRODE TRANSVERSE SHAPE* (RADIUS PROFILE IS BEST)
- *ELECTRODE CLEANLINESS* (SHOULD BE DRESSED FREE FROM PICKUP OF NONCONDUCTING COATINGS)
- *ELECTRODE COOLING SYSTEM*
- *COOLING WATER* APPLIED DIRECTLY TO THE WELD (IF APPLICABLE)

Cooling water for seam welding wheels can be either internal or external. Internal cooling systems cool the wheel but not the weld. The key advantage of internal cooling is that the process is dry; the workpiece, the machine, and the operator stay dry. The cost of installing floor drains and water containment facilities is avoided. Neither is it necessary to remove water from inside the workpiece (as in the case of automotive fuel tanks). However, internally cooled wheels may cost more due to additional machining costs. Thus, operating costs for internal cooling may be higher than for external cooling. Some internal systems do not directly cool the wheel, but only the wheel shaft.

On the other hand, with external cooling, spray nozzles are aimed with some precision at the workpiece-wheel separation point. The benefits of external cooling include:

- THERMAL DAMAGE TO THE WELD SURFACE AND ANY COATINGS IS REDUCED, SO THAT WELDABILITY IS IMPROVED.
- WELD DISTORTION IS REDUCED.
- FUME EMISSION IS REDUCED.
- WHEEL CLEANLINESS IMPROVES, BECAUSE WATER IS DIRECTLY APPLIED TO THE HOTTEST POINT ON THE WHEEL, PRODUCING A QUENCH/CLEANING ACTION ON THE WHEEL SURFACE.
- WELD HARDNESS WILL INCREASE, ADDING STRENGTH TO THE WELD UNDER CERTAIN LOADING CONDITIONS SUCH AS CRUSH LOADING. (*CAUTION: SOME QUENCH-SENSITIVE MATERIALS MAY NOT ALLOW EXTERNAL COOLING.*)

In welding bare steel, rust-preventive water solutions are sometimes used, such as 5% borax ($Na_2B_4O_7 \cdot 10H_2O$).

Metals Welded

Most low-carbon, high-carbon, low-alloy, and stainless steels, and many coated steels above approximately 0.15% tend to form areas of hard martensite upon cooling. In critical applications, the welds may require postweld tempering to reduce the hardness and brittleness. In some cases, this can be done in the welding machine. Tempering can also be done in a furnace or by induction heating.

Aluminum and aluminum alloys can be lap seam welded but not mash seam welded, due to their narrow plastic range. Nickel and nickel alloys can also be seam welded, but seam welding is not recommended for copper and high-copper alloys. Compatible combinations of dissimilar metals and alloys also can be seam welded. For additional information on the seam welding of stainless steel, aluminum alloys, and copper alloys, see the article "Procedure Development and Practice Considerations for Resistance Welding" in this Volume.

References cited in this section

4. "RECOMMENDED PRACTICES FOR RESISTANCE WELDING," AWS C1.1, AMERICAN WELDING SOCIETY
5. *RESISTANCE WELDING MANUAL,* 4TH ED., RESISTANCE WELDER MANUFACTURERS ASSOCIATION, 1989
6. T. OKUDA, SPOT WELDING OF THICK PLATES, PART I: THE LAW OF THERMAL SIMILARITY, *YOSETSU GIJUTSI,* VOL 21 (NO. 9), 1973, P 85-88 (IN JAPANESE)
7. T. YAMAMOTO AND T. OKUDA, FUNDAMENTAL STUDY ON WELD FORMATION PHENOMENA OF HIGH SPEED LAP SEAM WELDING, 3RD REPORT: VOLTAGE ACROSS ELECTRODES AND CURRENT FLOW AREA OF PLATE SURFACE, *YOSETSU GAKKAISHI,* VOL 46 (NO. 10), 1977, P 748-753 (IN JAPANESE)

Resistance Seam Welding

Michael J. Karagoulis, General Motors Corporation

Types of Seam Welds

Several types of resistance seam welds can be made:

- *LAP SEAM WELDS* JOINING OVERLAPPING FLAT SHEETS (FIG. 1A)
- *FLANGE-JOINT LAP SEAM WELDS* WITH AT LEAST ONE FLANGE OVERLAPPING THE MATING PIECE (FIG. 5A)
- *FOIL LAP SEAM WELDS,* IN WHICH AN ALLOY METAL FOIL IS PLACED AT THE WELD INTERFACE
- *MASH SEAM WELDS* WITH WORK METAL COMPRESSED AT THE JOINT TO REDUCE JOINT THICKNESS (FIG. 1B)
- *BUTT SEAM WELDS* (FIG. 1C AND 3)

In addition, both low-frequency and high-frequency induction and resistance welding are used for butt welding strip stock into tubes and other shapes.

Lap Seam Welds

The most common type of seam weld is a simple lap seam in which the pieces to be welded are lapped sufficiently to prevent the sheet edges from becoming part of the weld (Fig. 1a). Applications include sealing of cans and tanks in which liquid-tight or gas-tight joints are needed.

Flange-joint lap seam welds are used in joining assemblies having one straight member and one outward-flanged member at the joint to be welded (for example, the duct in Fig. 5a) or assemblies with two outward-ranged members (for example, the automotive gasoline tanks in Fig. 5b). Common applications include containers with out-turned bottoms or tops, and ducts or structural parts with out-turned sides. Unless the workpiece length is less than the usable throat depth of a longitudinal machine, seam welds of this type are made on a circular machine.

Often, container ends are dished for added strength, and one or both electrode wheels must be mounted at an angle to clear the workpiece. Wheel diameter limitations may also necessitate setting a wheel at an angle (Fig. 5a, right view).

Flange-joint seam welds can also be used for welding assemblies in which the flanges face inward at the joint (Fig. 5c). However, to reduce overhang of the arm supporting the lower electrode or to reduce the required throat depth of the machine, the length of the workpiece should be kept to a minimum. If the lower electrode support is small, too much overhang can result in excessive deflection of the support and thus cause inconsistent weld quality because of force variation.

Mash Seam Welds

Mash seam welds are produced by overlapping two sheets by an amount ranging from one to two times the sheet thickness and applying electrode force and current. The resulting over-thickness is 5 to 25%.

Usually wide, flat edges are used on the electrode wheels. The joint itself acts as a current localizer, as in projection welding, because the overlap is thicker than either sheet by itself. Because the wheels are wide and flat-edged, the transverse position of the joint on the wheels can be oscillated as needed to prevent a groove from forming on the wheels and to increase the period between wheel dressings.

Careful selection of an impulse weld schedule will result in higher penetration and greater joint strength, although the "bumpy" appearance of a weld made with pulsating current may not meet product requirements. Cosmetic mash seam welds can be made with nonpulsed (continuous) current, or by pulsing according to the minimum nugget spacing afforded by Eq 3 (for mash seams use S = thickness after welding).

Because sheet overlap is a critical parameter, workpieces must be rigidly and accurately clamped or tack welded to prevent lateral motion during welding. Clamp distance is highly important and is related to stock thickness. On continuously fed, moving-part setups and thin metal processes (where clamp distance would be too small), clamps may not be feasible. In these cases, a Z-bar and nosepiece tooling arrangement is used to help position the edges to be welded.

In applications in which the mashed surface must be as flat and burr-free as possible after welding to facilitate porcelain enameling and to present a good appearance, a bar-type electrode is used against the surface to be enameled, and an electrode wheel is used against the other surface. The bar-type electrode may require periodic dressing to avoid excessive grooving and subsequent loss of current density. When a premium finish appearance is required, the weld can be ground or roll planished to remove surface imperfections.

In steel coil joining, it is sometimes desirable for the final thickness of the mash seam weld to be similar to the single sheet thickness. To minimize the thickness gain from welding, pre- and/or postweld planishing rolls are used. These rolls are hard steel wheels, typically 100 mm (4 in.) in diameter and 20 mm (0.75 in.) in width. Planishing rolls are typically loaded to 22 kN (5000 lbf) to cold work the weld. In this way the overthickness may be held to as little as 5%. Also, the overlap may be retracted slightly after the preweld planish in order to further limit over-thickness.

The highest quality mash seam welds are made in low-carbon steels and stainless steels. The process is widely used in the manufacture of refrigerators, stoves, laundry equipment, and other products that receive porcelain enamel coatings. Mash seams are gaining popularity in automotive applications for creating tailored blanks that reduce weight and material loss in certain large stampings (for example, body sides and doors), reduce manufacturing complexity, and reduce the number of components and welds required to make finished automotive subassemblies.

Because of their narrow plastic range, nonferrous metals generally cannot be mash seam welded.

Butt Seam Welds

Butt seam welds can be classified as either autogenous or exogenous.

Autogenous (nonfilled) butt seam welds are made by a special seam welding technique in which the two abutting edges are heated and lightly forged together. There is a slight depression where the electrode wheels compress the plastic metal. This type of weld usually has low strength and a reduced cross-sectional area. Resistance tube welding may be a better alternative to this process.

Exogenous (Filled) Butt Seam Welds. The strength of a butt seam weld can be increased by adding filler metal to the outer surface of the weld, against the electrode. Filler metal, in the form of a foil or a shaped wire, is introduced on one or both surfaces adjacent to the joint as the workpiece is passed between conventional seam welding electrodes (Fig. 3). This technique produces a smooth, nonoverlapping seam that is high in strength with good appearance. Highest quality exogenous butt seam welds show nugget penetration to the filler-workpiece interface, complete filler bonding, and smooth, regular surface contours. The only edge preparation required is to ensure that the sheared edges are clean and straight without excessive burr or gap when they are butted together. The filler acts as a bridge to:

- DISTRIBUTE THE WELDING CURRENT MORE EVENLY TO BOTH SHEETS
- CONCENTRATE THE CURRENT IN THE JOINT
- HELP CONTAIN THE MOLTEN NUGGET AS IT GROWS AND THEN COOLS
- PROVIDE METAL FOR A SLIGHTLY RAISED BEAD, WHICH WILL ADD STRENGTH

The variables having the greatest effect on the tensile strength of filled butt seam welds are current, speed, size of the gap between the edges of the butted workpieces, and work metal thickness. Surface grinding will improve cosmetic appearance but also may reduce weld strength.

For the commonly used thicknesses of low-carbon steel, welding speeds similar to those used in lap seam welding are permissible, providing that the gap is no wider than 0.38 mm (0.015 in.). Either pulsed or continuous current can be used.

Resistance Seam Welding

Michael J. Karagoulis, General Motors Corporation

Processing Equipment

A seam welding machine is similar in construction to a spot welding machine, except that one or two electrode wheels are substituted for the spot welding electrodes.

Seam Welding Machines

In general, seam welding is done in a press-type resistance welding machine. Most seam welding machines are powered by alternating current (ac), either three-phase or single phase; some are designed for use with direct current (dc). Others superimpose a high frequency (16.5 kHz) on the basic power line. Stored-energy seam welding machines have been built, but this type finds little use. Direct current frequency converter power supplies are used to tailor custom pulsation schedules.

Equipment for seam welding must be capable of:

- DELIVERING LOW-VOLTAGE, HIGH-AMPERAGE CURRENT
- SUPPORTING THE ELECTRODES AND WORKPIECE AND APPLYING THE ELECTRODE FORCE
- MOVING THE WORKPIECE AND/OR DRIVING THE ELECTRODES
- REGULATING, TIMING, AND SEQUENCING THE APPLICATION OF THE WELDING CURRENT, FORCE, AND COOLING WATER (IF APPLIED)
- CAREFUL PART HANDLING, SUCH THAT EXTRANEOUS FORCES ARE NOT IMPOSED DURING WELDING

Electrode Force and Support. The upper electrode wheel is mounted to, and insulated from, the operating head. The head, which is actuated by a direct-acting air or hydraulic cylinder, applies the electrode force. The lower electrode is either a wheel, a platen, or a mandrel and is mounted on a supporting arm, table, or knee. This lower support can be made adjustable for applications in which the work metals must be maintained at a constant level above the floor.

Electrode or Workpiece Drives. Workpieces are moved by rotating the electrodes with knurl or friction drive, by direct drive of the wheel shaft, by a driven workpiece and idling wheels, or by clamping the workpieces to a bar electrode and moving the bar electrode (Fig. 3).

Weld travel tends to be continuous on thin steel sheets (less than 1.7 mm, or 0.067 in.), whereas it is intermittent on thick sheets (greater than 3.0 mm, or 0.118 in.). The need for intermittent weld travel arises from the longer weld time requirements of thick steel (Eq 1).

Knurl Drive versus Friction Drive. In a knurl drive or friction drive, one or both of the electrodes are rotated by a knurl or friction wheel on the peripheral surface of the electrode wheel, to provide nearly constant linear speed regardless of the diameter of the electrode wheel. The knurl drive, in which ridges or beads on the drive wheel aid in turning the electrode, minimizes slippage and is more positive than the friction drive, in which the turning surface of the drive wheel is smooth. The knurled wheel also removes some of the pickup from the electrode wheels. A knurl drive is used when welding coated metals, such as galvanized steel and terneplate, in any application in which the electrodes are likely to pick up material from the work metal, and for scaled stock that cannot be precleaned. A knurl drive usually is not employed if good cosmetic appearance of the weld is important. Knurl or friction drives can be used only with electrode wheels large enough in diameter to allow clearance.

Gear drive is used if small-diameter electrode wheels cannot be driven by the use of knurled or friction wheels because of interference, or if the application cannot tolerate an electrode wheel that has been roughened by a knurled drive roll. With gear drive, in which the mounting shaft of the electrode wheel is driven, the speed of welding decreases as the wheel wears down.

In a standard seam welding machine, interference may exist at some minimum distance between electrode wheel centers. If one wheel must be small to clear the workpiece, the other must be correspondingly large. If the ratio of the electrode wheel diameters is greater than 2 to 1, the smaller wheel should be driven, to minimize slippage. Welding speed can be kept constant in spite of wheel wear by a variable-speed gear-driven mechanism.

Movable Carriages. When a mandrel or a platen is used as the lower electrode (Fig. 3), the operating head carrying the upper electrode may be mounted in a carriage. The carriage is moved by an air or hydraulic cylinder, or by a motor-driven screw, and thus the upper electrode is passed over both the workpiece and the lower electrode. The upper electrode is free-wheeling, but it may be equipped with an idling knurl for dressing. Sometimes it is necessary to use a high-slip drive to assist on thin materials, to prevent wrinkling. In some machines, the workpiece is clamped to a bar-type lower electrode and moved under an idling fixed-position upper electrode wheel.

Types of Machines. There are four basic types of resistance seam weld machines: circular, longitudinal, universal, and portable.

Circular. The axis of rotation of the electrode wheels is at right angles to the front of the machine (Fig. 2a). This type of machine is used for circular work, such as for welding the heads on containers and for flat work requiring long seams. Head pivot capability for welding on an inclined plane is optional.

Longitudinal. The axis of rotation of the electrode wheels is parallel to the front of the machine (Fig. 2b), and throat depth is typically 305 to 915 mm (12 to 36 in.). This type of machine is used for welding short longitudinal seams in containers, for attaching pieces to containers, for coil joining, and for similar work. Machines with traveling heads or traveling electrodes, in which a mandrel or a platen is used for the lower electrode, are normally of the longitudinal type.

Universal. A swivel-type head and interchangeable lower arms allow the axis of rotation of the electrode wheels to be set either at right angles or parallel to the front of the machine.

Portable. The workpiece is clamped in a fixture, and a portable welding head is moved over the seam, either manually or by using a robot. This type of machine is used when the workpiece is too bulky to be handled by a regular machine or when it is more efficient to move the welding head rather than the workpiece. This type of machine is sometimes used for joining the roof and body sides on automobiles. The portable welding head is moved by motor-driven wheels, and an air cylinder mechanism provides the electrode force.

Common power sources used for seam welding are similar to those used for spot welding. They usually consist of a welding transformer, with a tap switch in the primary circuit, driven by a silicon-controlled rectifier switching system and a secondary circuit with electrodes for transferring the welding current to the work metal.

Controls for seam welding are generally similar to those for spot welding, except for differences related to the relative motion of the work and the electrodes. Most modern weld timers are capable of generating impulse weld schedules for seam welding. (Higher impulse frequencies for welding thin stock can be obtained using special high-frequency systems, although complete machine redesign may be necessary due to transformer and throat area limitations that are directed by the frequency.) Tap switches and a phase-shift heat control are used as in spot welding for changing the transformer output current. In theory, direct feedback current regulation would be a desirable feature in seam welding, and some work has been done to develop and improve this. Alternating current is inherently more stable in secondary root-mean-square (rms) current than dc, because in ac there is essentially no current memory from one half-cycle to the next, or from one impulse to the next.

Selection of Electrodes

Alloy type and wheel configuration are the primary factors that determine electrode selection.

Class 1 Copper. Electrode wheels made of Resistance Welder Manufacturers Association (RWMA) class 1 copper (RWMA No. 1.16200) have been used for seam welding of aluminum and magnesium alloys, galvanized steel, and tin-plated steel. The minimum electrical conductivity of class 1 copper is 80% International Annealed Copper Standard (IACS). Recently, this material has been replaced by hot-forged and heat-treated RWMA class 2 copper in the overaged condition (RWMA No. 1.18200). With adequate cooling and proper parameters, the high electrical and thermal conductivities of this material keep the electrode-workpiece interface at a temperature below the point where metallurgical discontinuities are formed near the weld surface.

Class 2 Copper. Seam welding of low-carbon and low-alloy steel is usually done with electrodes made of hot-forged and heat-treated RWMA class 2 copper (RWMA No. 2.18200). Minimum conductivity is 75% IACS. Minimum hardness is typically 65 HRB. Class 2 also works well with all types of coated steel, but the metallurgical optimization of weld parameters and provisions for wheel maintenance are critical to the success of welding coated steels. For longer electrode life, premium class 2 wheels are available in the 20% coldworked form, where a minimum hardness of 75 HRB may be specified. Coldworked wheels show no appreciable loss of conductivity over the forged class 2 material.

Class 20 copper (RWMA No. 20.15760) is also used for electrode wheels, with properties similar to those of class 2 premium wheels. Class 20 wheels are made using a powder metallurgy process that produces a pure copper product, dispersion-strengthened with alumina. The alumina adds hot strength and tends to reduce pickup during welding of coated steels.

Class 3 copper (RWMA No. 3.17510) is sometimes used for seam welding materials with lower electrical conductivities, such as stainless steel, Nichrome, and Monel alloys. However, class 3 is used with special ventilation only because of the health hazard of atmospheric beryllium when welding or machining with class 3 material. (Beryllium-free class 3 alloy RWMA 3.18000 is also available.) Class 3 wheels have lower electrical conductivity (45% IACS), so they tend to run hot.

Sizes and Shapes. Electrode wheels range in diameter from 50 to 610 mm (2 to 24 in.). Popular sizes range from 100 to 305 mm (4 to 12 in.), with widths from 6 to 19 mm ($\frac{1}{4}$ to $\frac{3}{4}$ in.). Edge contours may be flat, radiused, beveled flat, or angled. In lap seam welding, weld quality seems to be easiest to control with a radius edge, as discussed above. The need for nonradius contours is usually a matter of cosmetic appearance, or due to a tool access limitation. Mash seam welders very often use a flat-edge wheel, as in Fig. 1(b).

Resistance Seam Welding

Michael J. Karagoulis, General Motors Corporation

Weld Quality and Process Control

Welding Parameters. It is best to design the entire process around proven weld parameters that have been metallurgically validated on workpieces very similar to actual production parts. Sensitive parts of the weld typically are starts, stops, corners, adjacent shunt paths, intersecting seams, and areas of poor metal fit. On coated metal, parameters should be flexible enough to handle normal coating variations.

To select starting weld parameters, use the guidelines presented in the section "Fundamentals of Lap Seam Welding" in this article and Table 1. For process design, the givens are usually material thickness, available pulsation capability, and wheel diameter. From these three independent variables will spring all of the dependent variables (for example, speed, force, current, and so on) from Eq 2, 3, 4, 5, 6, 7, 8, 9, 10, and 11. At this point, estimates of S and L would be helpful in order to begin the calculations. Use working values that are not too near the extreme limits of either the material, the weld tool, or the plant facilities. To make leak-tight seams, weld spots should have approximately equal length and width. Ideally, the same welding parameters should be used for the entire length of the weld.

Verify the quality of resulting welds using chisel tests, metallographic sections, pressure tests, or fatigue tests, depending on product requirements. Record target values for speed, force, pulsation, and water cooling (if applicable). Record tap setting and lower and upper rms secondary current limits for acceptable weld quality. Record the actual secondary current waveform to see that it is acceptable (as in Fig. 8) and consistent throughout the weld. Process changes should be made only on the basis of data.

Establish production procedures for maintenance of electrode wheels and cooling sprays (if applicable). Check process parameters against required values frequently, and make adjustments and repairs as needed.

Nondestructive testing (for example, resistivity, ultrasonic testing, dye penetrant, magnetic particle, or eddy current) may be useful in some low-volume testing applications. However, such techniques should come into use only after the optimum parameters are verified metallographically and the process is well understood. In many instances, much can be observed from the external appearance of a lap seam weld, provided the weld surface is not planished or scarfed after welding.

Coated Steels. For a process that has been properly set up and optimized as described in the section "Welding Parameters" in this article, the difference in parameters between coated and uncoated steel lies mainly in weld current. Steels with low-melting metallic coatings require more current than uncoated steels. Organic coated steels require about the same current as uncoated steels (although the current range for making good welds will be smaller for the organic coated steels). Some coatings (for example, zinc and aluminum) are particularly active at the electrode-workpiece interface. For these surface-active coatings, speed, force, and pulsation are selected to keep the surface temperature of the weld below the point where metallurgical discontinuities are developed. Diligent maintenance of the wheel condition and the cooling system cannot be overstated for sensitive applications.

Michael J. Karagoulis, General Motors Corporation

References

1. K.M. PICKETT, M.F. SHI, K.K. BHATT, AND J. SELVARAJ, STRUCTURE PROPERTY RELATIONSHIPS OF RESISTANCE MASH SEAM WELDED TAILORED BLANKS WITH POST HEAT TREATMENT, *24TH CONF. INT. METALLOGRAPHIC SOC.* (DENVER), AUG 1992
2. Y. ADONYI, J. MILIAN, AND L. ROUDABUSH, WELDED STEEL SHEET PERFORMANCE FOR TAILORED BLANKS, *PROC. SHEET METAL WELDING CONF. V*, PAPER C1, AMERICAN WELDING SOCIETY, OCT 1992
3. VDI REPORT 818, VEREIN DEUTSCHER INGENIEURE, 1990, P 256 (IN GERMAN)
4. "RECOMMENDED PRACTICES FOR RESISTANCE WELDING," AWS C1.1, AMERICAN WELDING SOCIETY
5. *RESISTANCE WELDING MANUAL*, 4TH ED., RESISTANCE WELDER MANUFACTURERS ASSOCIATION, 1989
6. T. OKUDA, SPOT WELDING OF THICK PLATES, PART I: THE LAW OF THERMAL SIMILARITY, *YOSETSU GIJUTSI*, VOL 21 (NO. 9), 1973, P 85-88 (IN JAPANESE)
7. T. YAMAMOTO AND T. OKUDA, FUNDAMENTAL STUDY ON WELD FORMATION PHENOMENA OF HIGH SPEED LAP SEAM WELDING, 3RD REPORT: VOLTAGE ACROSS ELECTRODES AND

CURRENT FLOW AREA OF PLATE SURFACE, *YOSETSU GAKKAISHI,* VOL 46 (NO. 10), 1977, P 748-753 (IN JAPANESE)

Resistance Seam Welding

Michael J. Karagoulis, General Motors Corporation

Selected References

- Y. ADONYI, RESISTANCE SEAM WELDING OF ENAMELING QUALITY STEEL SHEET, *PROC. PORCELAIN ENAMEL INSTITUTE TECHNICAL FORUM,* VOL 52, 1990, P 729-748
- M.J. KARAGOULIS AND M.A. PUSATERI, SEAM WELDING AND PRE-PAINTED FUEL TANKS AT GENERAL MOTORS, *PROC. SHEET METAL WELDING CONF. II,* AMERICAN WELDING SOCIETY, OCT 1986
- C.A. LEFFEL, "AN INVESTIGATION INTO THE RESISTANCE SEAM WELDABILITY OF COATED STEELS," MASTER'S THESIS, THE OHIO STATE UNIVERSITY, 1990
- A.A. MAZAEV AND G.V. NAZAROV, EFFECT OF THE FORM OF THE WELDING PULSE ON HEATING OF THE WELD ZONE IN ONE-SIDE RESISTANCE SPOT AND SEAM WELDING, *FIZ. KHIM. OBRAB. MATER.,* VOL 17 (NO. 5), 1983, P 125-131 (UDC 621.791.762.5 IS IN ENGLISH)
- A.W. SCHUELER, RESISTANCE SEAM WELDING, *WELD. RES. COUNC.,* BULLETIN 225, APRIL 1977
- SOUDRONIC LTD LITERATURE, 465 NORTH STATE RD., BRIARCLIFF MANOR, NY 10510, USA
- K. TORII, Y. KADANO, AND Y. TAMURA, A STUDY ON THE DEFECTS IN SEAM WELDS OF METAL COATED STEEL SHEETS, *TRANS. JPN. WELD. SOC.,* VOL 2 (NO. 1), 1971, P 77-85
- S.A. WESTGATE, THE HIGH SPEED RESISTANCE SEAM WELDING OF NOMINALLY 1.2 MM LOW CARBON STEEL SHEETS, *BR. WELD. INST.,* DOC 240/1984, 1984
- N.T. WILLIAMS, D.E. THOMAS, AND K. WOOD, HIGH SPEED SEAM WELDING OF TINPLATE CANS, PARTS I AND III, *MET. CONSTR.,* APRIL 1977, P 157-160, MAY 1977, P 406-408

Flash Welding

H.J. Latimer and R.B. Matteson, Taylor-Winfield Corporation

Introduction

FLASH WELDING (FW) is a resistance welding process in which a butt joint weld is produced by a flashing action and by the application of pressure. In basic terms, it is a melting and a forging process. The process is capable of producing welded joints with strengths equal to those of the parent materials. Figure 1 is a schematic representation of a typical flash welding operation.

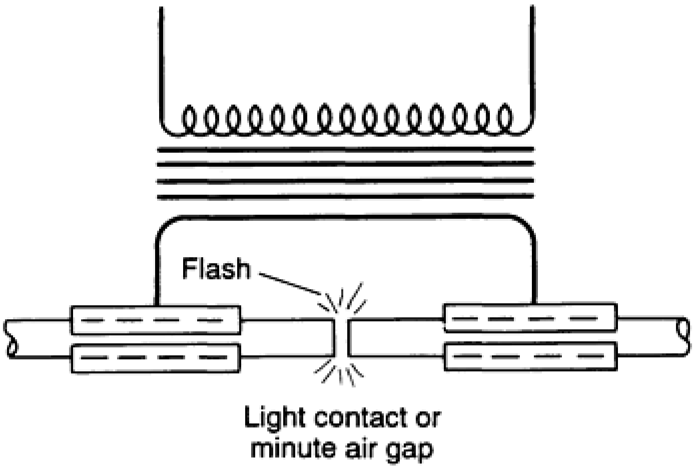

FIG. 1 TYPICAL FLASH-WELDING SETUP SHOWING AIR GAP PRESENT BETWEEN ENDS OF TWO BARS

Flash welding is used to join metallic parts that have similar cross sections, in terms of size and shape. The process lends itself to joining nearly all grades of steel, aluminum, brass, and copper parts, in addition to selected dissimilar materials.

Flash Welding

H.J. Latimer and R.B. Matteson, Taylor-Winfield Corporation

Process Applications

Materials with cross sections ranging from 65 to 13,000 mm^2 (0.1 to 20 in.^{2}) are commonly flash welded. The process is also used to manufacture components of solid and tubular shape, in addition to strips and complex configurations. Flash welding is commonly used to weld rings that range in diameter from a few inches (Fig. 2) to as much as 4.6 m (15 ft) (Fig. 3). The aircraft engine industry uses flash welded rings made from heat-resistant materials in accordance with military specifications for jet aircraft.

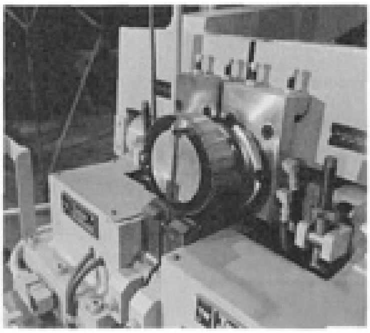

FIG. 2 FLASH-WELDING MACHINE USED TO FABRICATE A COILED STEEL STRIP INTO A SMALL-DIAMETER RING WITH A FLANGE

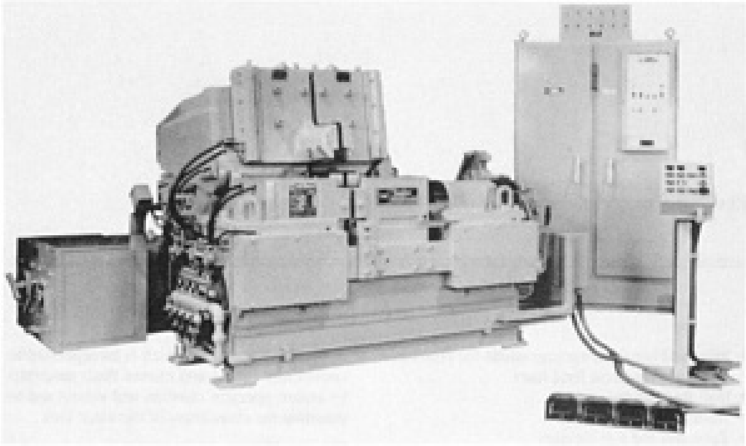

FIG. 3 FLASH-WELDING MACHINE WITH CAPACITY TO WELD RINGS UP TO 4.6 M (15 FT) IN DIAMETER

Other flash-welding applications include:

- CHAIN LINKS
- TRANSMISSION BANDS
- AUTOMOTIVE FLYWHEEL RING GEARS
- STRIPS THAT ARE JOINED FOR CONTINUOUS PROCESSING LINES
- WIRE AND BAR DRAWING OPERATIONS FOR CONTINUOUS STAMPING PRESS FEED LINES
- ROLL FORM LINES
- AIRCRAFT LANDING GEAR
- TUBE AND ROD EYE CLEVISES
- BAND-SAW BLADES (FIG. 4)
- DRILL EXTENSIONS USING TWO DIFFERENT MATERIALS
- MITER JOINTS FOR AUTOMOTIVE AND HOME WINDOW FRAMES
- CRANKSHAFT COUNTER WEIGHTS
- ALUMINUM-TO-COPPER ELECTRICAL TRANSITION FOR POWER TRANSMISSION
- ANODES USED IN ALUMINUM SMELTERS

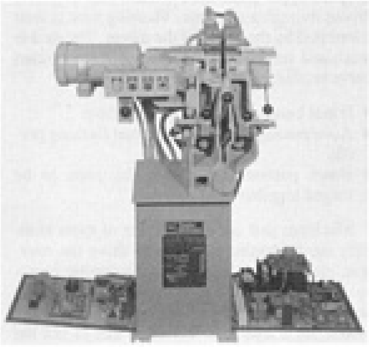

FIG. 4 FLASH-WELDING MACHINE INSTALLATION EQUIPPED TO WELD BAND-SAW BLADES

Flash Welding

H.J. Latimer and R.B. Matteson, Taylor-Winfield Corporation

Processing Equipment

A typical flash-welding machine consists of seven major components:

- A MACHINE BED THAT HAS A FIXED PLATEN ATTACHED TO IT, AS WELL AS A SET OF ELECTRICALLY INSULATED WAYS THAT SUPPORT A MOVABLE PLATEN
- A MOVABLE PLATEN, WHICH IS GENERALLY MOUNTED ON THE ELECTRICALLY INSULATED WAYS THAT ARE ATTACHED TO THE MACHINE BED
- TWO CLAMPING ASSEMBLIES, ONE OF WHICH IS RIGIDLY ATTACHED TO EACH PLATEN TO BOTH ALIGN AND HOLD THE PARTS TO BE WELDED IN PLACE
- EQUIPMENT FOR CONTROLLING THE MOTION OF THE MOVABLE PLATEN
- A WELDING TRANSFORMER WITH ADJUSTABLE TAPS
- CONTROLS AND MONITORS FOR MACHINE FUNCTIONS
- A VENT/FLASH CATCHER, WHICH IS INCORPORATED TO COLLECT ANY FUMES AND EXCESS FLASH MATERIALS TO ENSURE OPERATOR COMFORT AND SAFETY AND TO MAINTAIN THE CLEANLINESS OF THE WORK AREA

The movable platen of a flash welder can be driven by a number of different methods. The simplest type often uses a cam that is motor driven through a gear box. Flashing time is then controlled by the speed of the motor. The cam is machined to produce the desired flashing cam curve profile, including the:

- INITIAL BURN-OFF REGION OF LINEAR FLASH
- ACCELERATING CURVE FOR THE ACTUAL FLASHING PROFILE
- UPSET PORTION THAT CAUSES THE PARTS TO BE FORGED TOGETHER AFTER FLASHING

Machines that are either larger or more complex use hydraulic cylinders to drive the movable platen and to incorporate hydraulic servo-systems. Although it is possible to use a servovalve to directly control the position of the platen via a hydraulic cylinder, this is not the preferred method, because contamination by the oil is extremely difficult to prevent. The better approach is to use a hydraulic servoactuator, in which the servovalve operates on its own refined oil system and the oil that circulates through the platen cylinder is derived from another hydraulic source. This approach has been shown to be more reliable and more accurate, and to vastly improve servovalve life.

Controls and Auxiliary Equipment. Electrical controls on flash-welding equipment are designed to sequence the machine, control the weld current, and precisely control the platen position during flashing and upsetting. These controls often comprise programmable controllers, welding current controllers, and either motor or servovalve controllers. Silicon-controlled rectifier (SCR) contactors (solid-state switches) are used as the devices to control the flow of welding current from the power lines to the welding transformer. Older machines used ignitrons for this purpose, but SCRs have fully replaced them in recent years. Preheat and postheat cycles are often required procedures in flash welding operations; timers with phase-shift heat control are used to accomplish these important functions.

Power Sources. The most common source of power for flash welders is single-phase alternating current (ac) power. Flash welding transformers are KVA rated at 50% duty cycle.

Conversions From ac to dc Power. If an extremely large demand current is required and an adequate single-phase ac source is not available, then ac to dc conversion is employed using a three-phase transformer/rectifier. The most frequently used rectifier systems are the three-phase half-wave and three-phase full-wave types. In actual practice, no discernible differences in the weld produced by these two rectifier systems can be detected.

State-of-the Art Welding Unit Components. Devices that monitor and control selected process parameters are commonly incorporated into the latest flash welding units.

A linear ramp preflashing approach rate is usually built into the flashing cam and minimizes any tendency of the weld to freeze at the beginning of flashing. The term "flashing cam" is used regardless of whether the cam is a motor-driven machined cam or if the position curve is generated electronically.

Natural Log Flashing Curves. Numerous technical articles have been published on what constitutes the best flashing curve. Although engineers are divided on this issue, it has been shown that the use of natural log curves results in significantly better welds than are produced using any other curve. The natural log system conforms to the natural heat-transfer characteristics of the metal being welded, thereby creating a flashing progression that parallels the development of heat at the interface with a natural log gradient toward the clamp.

Ramp Turn-On of SCRs. Not to be confused with the linear approach ramp, this control device slowly and progressively turns on the SCRs at the beginning of the flashing cycle. Because the magnetic condition of a weld transformer may be unknown at the end of a previous weld, the ramp turn-on device is used to restore the transformer to a known condition. This greatly extends the life of the transformer itself, in addition to the SCR contactor, because it avoids the high inrush of current that would have occurred when the next weld was initiated.

No Phase Shift During Flashing. This control feature ensures that full voltage will be applied to the dies whenever flashing conditions are present, except during the ramp turn-on interval. A phase shift can reduce the average voltage, but does not reduce the peak voltage. During the phase-delay period of the voltage waveform, the platen continues to travel forward until a peak voltage is suddenly applied, causing violent flashing and deep cratering. Although the pyrotechnics of a phase-shift flashing action can be quite impressive, the resulting weld will be a disaster. The best flashing action is relatively subdued and produces flash particles of significantly lower intensity and size.

The oxygen depletion system used in this process is a patented method of combusting the oxygen that would otherwise exist in the weld interface region. In actual use, the oxygen depletion system does more than remove oxygen. By heating the gases in the interface region, the ionization voltage is reduced and higher flashing currents can be sustained at relatively low flashing voltages.

Upset Welding

W.R. Kanne, Jr., Westinghouse Savannah River Company

Introduction

UPSET WELDING (UW) is a resistance welding process utilizing both heat and deformation to form a weld. The heat is produced by resistance to the flow of electrical current at the interface of the abutting surfaces to be joined. The deformation results from force on the joint in combination with softening from the electrical resistance heat. Upset welding typically results in solid-state welds (no melting at the joint). The deformation at the weld joint provides intimate contact between clean adjoining surfaces, allowing formation of strong metallurgical bonds. If any melting does occur during upset welding, the molten metal is typically extruded out of the weld joint area.

A wide variety of shapes and materials can be joined using upset welding in either a single-pulse or continuous mode. Wire, bar, strip, and tubing can be joined end to end with a single pulse of welding current. Seams on pipe or tubing can be joined using continuous upset welding by feeding a coiled strip into a set of forming rolls, resistance heating the edges with wheel electrodes, and applying a force to upset the edges together.

Equipment for single-pulse upset welding is relatively simple. It consists of a pneumatic or hydraulic system for force application, transformers or a bank of diodes as a source of electrical current, and a standard resistance-welding controller. A data acquisition system usually is employed to record the force, current, voltage, and motion of the weld head during welding. Equivalent welds have been made using both alternating and direct current.

Upset welds have similar characteristics to inertia friction welds, which are also solid-state welds. The amount of deformation is usually less for upset welds, and the deformation can be more precisely controlled using upset welding. For example, a pipe butt weld made using inertia friction welding will have a large upset on both the inside and outside, whereas an upset weld can be controlled, through joint design and welding parameters, to have essentially no internal upset (Fig. 1).

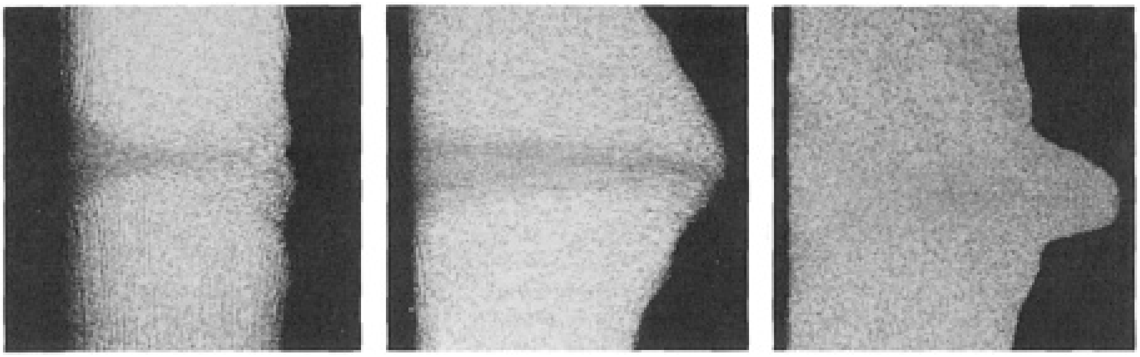

FIG. 1 UPSET WELD CONFIGURATIONS IN BUTT WELDS OF PIPE WITH AN OUTSIDE DIAMETER OF 25 MM (1 IN.) AND A WALL OF 3 MM (0.1 IN.). UPSET HAS BEEN MACHINED FROM THE OUTSIDE SURFACE (ON LEFT). THE INTERNAL UPSET CONFIGURATION WAS CONTROLLED BY ADJUSTING WELD PARAMETERS AND JOINT DESIGN. (A) MINIMUM UPSET. (B) LARGE UPSET. (C) NARROW UPSET AT LOW FORCE. COURTESY OF WESTINGHOUSE SAVANNAH RIVER COMPANY

Upset Welding

W.R. Kanne, Jr., Westinghouse Savannah River Company

Advantages

Solid-state upset welding has advantages compared to typical fusion welding processes. These advantages are due to the simplicity of the welding process and the resulting solid-state weld microstructure. These advantages include:

- *SPEED:* THE UPSET WELDING PROCESS IS FAST (USUALLY LESS THAN 1 S).
- *EASE OF CONTROL:* THE PROCESS HAS ONLY THREE PRIMARY VARIABLES (CURRENT, FORCE, AND TIME).
- *FEWER DEFECTS:* TYPICAL FUSION WELDING DEFECTS SUCH AS POROSITY, MISSED JOINTS, INCOMPLETE FUSION, SPATTER, AND SOLIDIFICATION CRACKING DO NOT OCCUR IN UPSET WELDS.
- *ENHANCED WELD PROPERTIES:* THE METALLURGICAL PROPERTIES OF THE WELD METAL AND HEAT-AFFECTED ZONE IN UPSET WELDS ARE THOSE OF HOT-WORKED MATERIAL. IN OTHER WORDS, THE STRENGTH OF THE WELD ZONE IS NOT REDUCED TO THAT OF AN ANNEALED STRUCTURE, AS IN FUSION WELDS.
- *SIMPLICITY OF EQUIPMENT:* UPSET WELDING EQUIPMENT IS NOT COMPLEX, INVOLVES NO ROTATING PARTS, AND REQUIRES MINIMAL MAINTENANCE.
- *LESS-STRICT COMPOSITION REQUIREMENTS:* MINOR ALLOYING ELEMENTS DO NOT AFFECT UPSET WELD QUALITY, AND THUS ATTENTION TO ALLOY COMPOSITION FOR WELDABILITY CAN BE ELIMINATED.
- *ABILITY TO JOIN DIFFICULT-TO-WELD MATERIALS:* ALLOYS NORMALLY CONSIDERED UNWELDABLE CAN BE JOINED BY UPSET WELDING. FOR EXAMPLE, A VARIETY OF STAINLESS STEELS (INCLUDING A-286), SUPERALLOYS (INCLUDING TD NICKEL), REFRACTORY METALS (INCLUDING TUNGSTEN), GRADE 2 TITANIUM, AND ALUMINUM ALLOYS (INCLUDING 2024) HAVE BEEN SUCCESSFULLY UPSET WELDED.

The effect of welding conditions, other than the basic parameters of force, current, and time, is generally minimal. Surface roughness usually has little effect, because it is overshadowed by the deformation that takes place during welding. Surface cleanliness and welding atmosphere may or may not be important depending on the amount of deformation that occurs to break up any oxide present before, or formed during, welding.

Upset Welding

W.R. Kanne, Jr., Westinghouse Savannah River Company

Limitations

A perceived limitation of the UW process is the lack of a good nondestructive testing technique for determining the quality of solid-state bonds. Normally, control of process variables is sufficient to ensure quality. Nondestructive methods, such as ultrasonics, are being applied to solid-state welds for determination of bond quality. For upset welding of large parts, the peak power drawn from the electrical supply may be a disadvantage.

Upset Welding

W.R. Kanne, Jr., Westinghouse Savannah River Company

Types of Upset Welds

Fixturing of the parts to be joined is usually necessary to achieve alignment of the parts, carry the electrical current to the joint area, and confine deformation to the joint area. In some cases the increased heat at the joint--which is caused by the higher electrical resistivity at the joint--is sufficient to limit deformation to the weld joint area. Electrodes are made from class 2 copper, which may be faced with a harder material such as copper-infiltrated tungsten carbide. Fixturing used to join two cylindrical parts consists of a pair of electrodes confined in an alignment sleeve (Fig. 2). The metallurgical structure of the weld area is that of a hot-worked material with a good diffusion bond across the weld interface, as can be seen in Fig. 1. This microstructure results in higher strength in the weld area than can be achieved for fusion welds (Fig. 3).

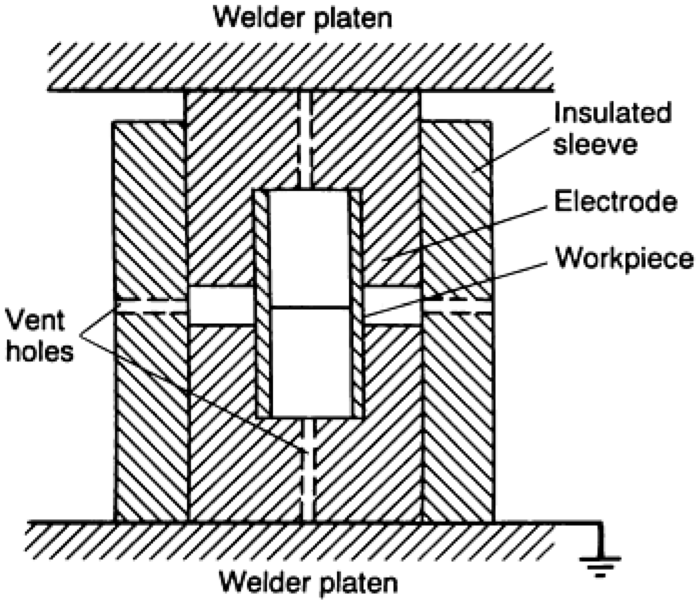

FIG. 2 FIXTURING ARRANGEMENT FOR UPSET WELDING OF CYLINDRICAL PARTS

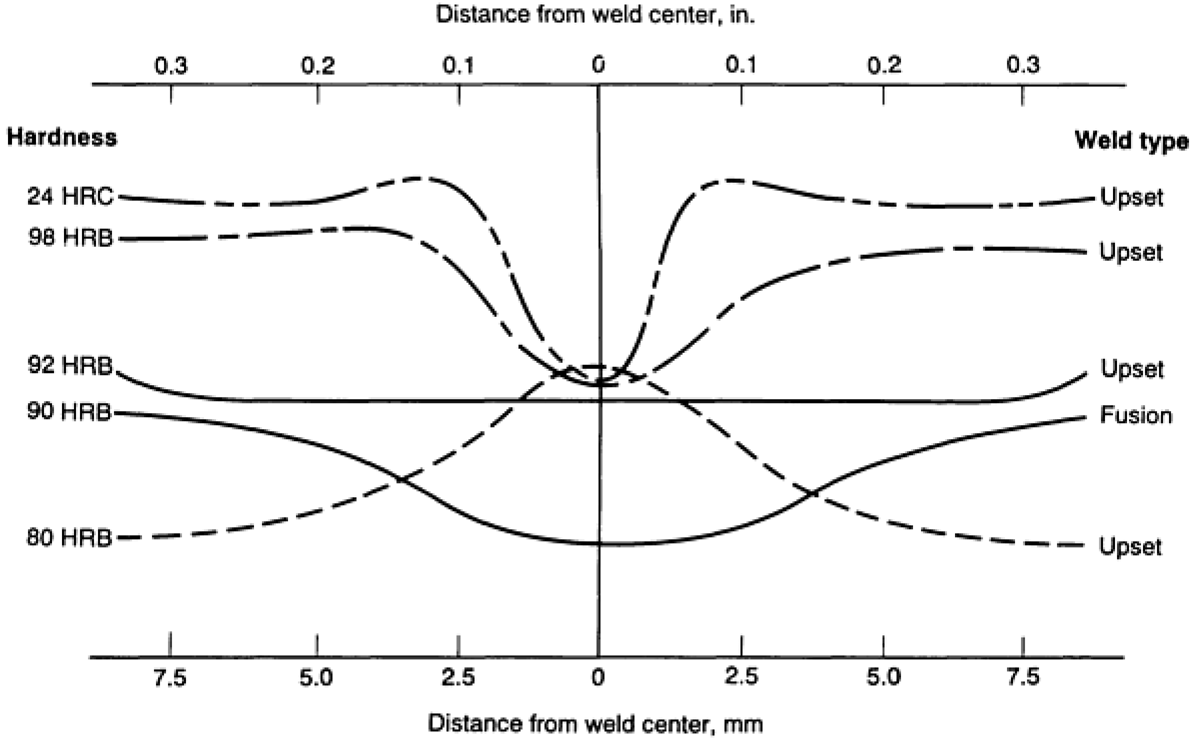

FIG. 3 HARDNESS ACROSS UPSET WELDS AND IN MATERIALS OF VARIOUS STRENGTHS COMPARED TO FUSION WELD HARDNESS. NOTE THAT THE WELD AREA HARDNESS (AND THEREFORE STRENGTH) IS GREATER FOR THE UPSET WELDS THAN FOR THE FUSION WELD.

Plug welds are a type of upset weld that requires little fixturing. Plug welds up to 125 mm (5 in.) in diameter are used to seal the fill holes of canisters employed in the long-term disposal of high-activity nuclear waste. The canisters, which are made from 304L stainless steel, are 610 mm (24 in.) in diameter and 3 m (118 in.) in height. The plug configuration is shown in Fig. 4(a). The considerable upsetting that occurs during the welding process results in a solid-state weld (Fig. 4b). Typically, these welds are made using pressure of 300 kN (75,000 lbf) with 230,000 A current applied for 1.5 s. The resulting welds have the high reliability required for long-term containment of radioactive material.

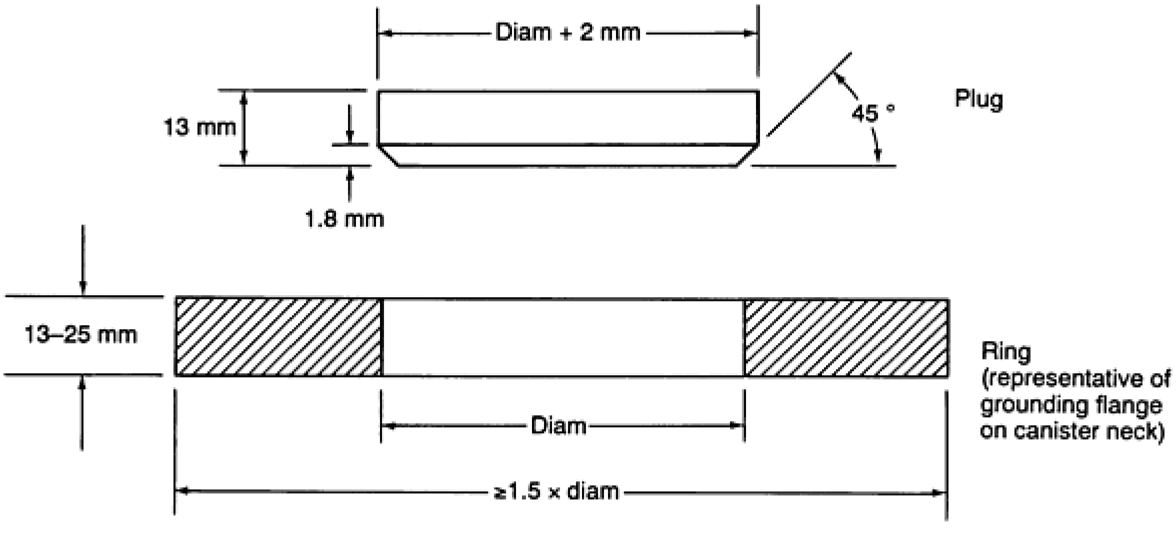

FIG. 4 UPSET PLUG WELD FOR CANISTER CLOSURE. (A) DESIGN OF WELD JOINT. (B) METALLOGRAPHIC SECTION SHOWING SOLID-STATE WELD

Alternative Energy Supply

A homopolar generator is an alternative method for supplying the electrical current for upset welding. Homopolar generators store energy in a large flywheel and then quickly convert the stored energy to electrical energy. The demand on utility power lines is thus spread over a considerable time period, yet the high peak power required for upset welding large parts can be met over a period of about 1 s. This low demand on the electrical grid is the principal advantage of the homopolar generator, whereas lack of good control of the current waveform and the complexity of the equipment are drawbacks. The application of homopolar generators to resistance upset welding was developed in the late 1970s and has since been applied to welding of pipes up to 168 mm (6.6 in.) in diameter.

Selected References

- T.A. AANSTOOS AND J.M. WELDON, HOMOPOLAR PULSE UPSET WELDING OF APT X-60 HIGH STRENGTH LINE PIPE, *WELD. J.*, JULY 1984, P 23-28
- B.J. EBERHARD AND J.W. KELKER, JR., HIGH CURRENT RESISTANCE WELDING OF NUCLEAR WASTE CONTAINERS, *WELD. J.*, JUNE 1982, P 15-19
- W.R. KANNE, JR., SOLID-STATE RESISTANCE WELDING OF CYLINDERS AND SPHERES, *WELD. J.*, MAY 1986, P 33-38
- *WELDING HANDBOOK*, 8TH ED., VOL 2, AWS, P 598-602

High-Frequency Welding

Warren Smith and Julian Roberts, Thermatool Corporation

Introduction

HIGH-FREQUENCY WELDING is a welding process in which the heat source used to melt the joining surfaces is obtained from high-frequency (HF) alternating current (ac) resistance heating. High-frequency current has certain characteristics that make it useful for welding. Unlike direct current (dc) or low-frequency alternating current, HF current tends to flow at high densities along surfaces (skin effect) and seeks adjacent parallel surfaces for its return path (proximity effect). This means that the heating and subsequent melting can be efficiently concentrated and focused to the surfaces where it is needed.

In nearly all HF welding processes, the joining surfaces are mechanically squeezed together (upset) after they are melted. This procedure helps produce a high-quality weld by squeezing out residual oxides and molten metal, which are detrimental to weld integrity. The removal of molten metal from the weld zone is beneficial because it eliminates the potentially harmful effects associated with cast structures, such as lower fracture toughness and poorer corrosion resistance. The upset also hot-works the bond plane and subsurface material, which can result in enhanced properties.

The tubemaking incorporates the simplest method of controlling HF currents, that is, the generation of an edge "V." High-frequency currents can be supplied to the welding process or workpiece by using either an induction coil (known as high-frequency induction welding, HFIW) or electrical contacts (known as high-frequency resistance welding, HFRW). The edge "V" is a series component of the circuit (Fig. 1). Because of proximity and skin effects, current flow is concentrated across the full face width of the strip edges, resulting in controlled surface melting and subsequent welding.

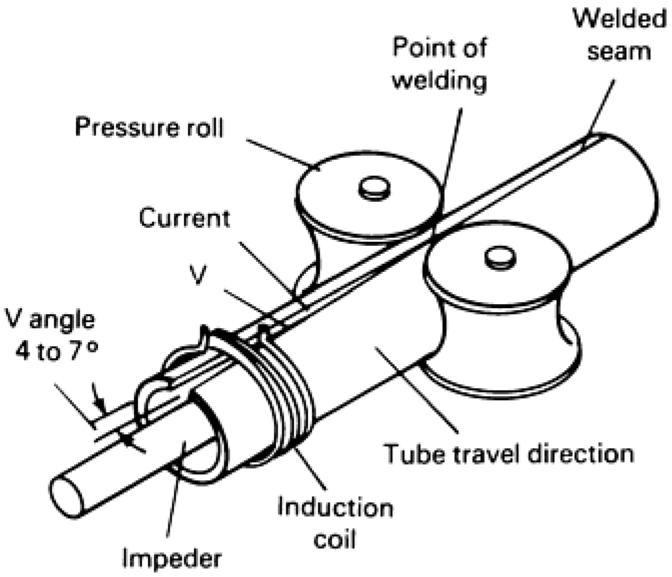

FIG. 1 JOINING A TUBE SEAM BY HIGH-FREQUENCY INDUCTION WELDING

In this process, a coil of strip material is first formed into a longitudinal hollow on a forming mill. The two edges of the strip are then HF welded and squeezed together to form a continuous welded seam.

Materials that can be successfully HF welded into tube and pipe include carbon steels, stainless steels, aluminum, copper, brass, and titanium. Exceptions are materials that are unstable at welding temperatures, have a negligible hot-work capability, or experience property deterioration that cannot be subsequently recovered. Where necessary, a gas shield can be provided for reactive metals. Tube or pipe sizes fabricated by HF welding range in size from 13 to 1220 mm ($\frac{1}{2}$ to 48 in.) in diameter.

The advantages of HF welding are that it is well suited for high-speed welding and can weld a large range of product sizes and materials. Weld quality is not particularly sensitive to the presence of air, and special atmospheres are usually not needed. Weld quality is relatively (but not completely) tolerant of surface oxides and contamination.

The disadvantages of HF welding are that it is not well suited for low welding speeds and small-scale operations where welding is done by hand. HF welding must be done continuously; continuous welds cannot be made in stop/start operations because a discontinuity in the weld will usually occur.

High-Frequency Welding

Warren Smith and Julian Roberts, Thermatool Corporation

Applications

High-frequency welding is most suitable in applications that involve the continuous edge, or butt, joining of metals. The largest single use of HF welding is in the manufacture of tube and pipe. High-frequency welding is useful in the manufacture of certain types of heat-exchanger tube, where the edge of a rectangular strip is continuously welded onto the outside diameter of a tube to form a cooling fin. The fin can be welded around the tube in a spiral configuration (Fig. 2) or it can be straight and parallel to the pipe axis. The greatest volume of heat exchanger tube is made with low-carbon steel for both the tube and fins. However, other alloy combinations are growing in use, such as low-carbon steel fins on both stainless steel and chromium/molybdenum alloy steel tubes, stainless steel fins on low-carbon steel and stainless steel tubes, and aluminum fins on cupronickel tube.

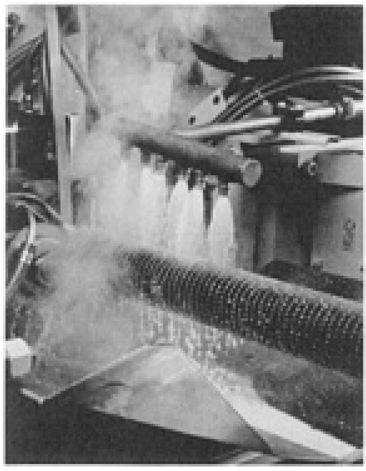

FIG. 2 MANUFACTURING A SPIRAL FIN TUBE USING HIGH-FREQUENCY WELDING

High-frequency welding can also be used for the production of structural shapes, such as T-sections and I- and H-beams. The specific shape is constructed by welding rectangular strips together to produce the desired configuration. A heating "V" is established as the strip materials are brought together.

The end, or butt, welding of metal strip is another application in which HF welding can be used. Single lengths of strip are welded to form a hoop that is subsequently formed into products such as wheel rims. High-frequency butt welding is also used to make long continuous lengths of strip by butt welding together a series of shorter lengths. This is done for certain manufacturing processes that require strip material longer than that which is commercially available.

High-Frequency Welding

Warren Smith and Julian Roberts, Thermatool Corporation

Equipment

Units for producing power for HF welding include vacuum tube oscillators and solid-state inverters. Depending on the application, frequency requirements can vary from 100 to 700 kHz, whereas power requirements vary from 30 to 1000 kW. Solid-state inverters produce frequencies ranging from 100 to 400 kHz. Vacuum-tube oscillators are available in frequencies up to 700 kHz. Because of their improved efficiency, solid-state units are replacing vacuum-tube units when frequency requirements permit.

Water-cooled copper conductors are used to carry the power to the coil or contacts. These conductors must be of minimum length and closely spaced to minimize impedance losses. The use of induction coils or electrical contacts to deliver HF power to the workpiece has specific advantages and disadvantages.

Induction coils are the least-efficient vehicle. They can only be used in applications where the geometry of the workpiece can act as a closed circuit to pick up the induced power (for example, the welding of tube). The advantages of an induction coil are that it is generally more convenient to use and requires less maintenance than contacts. Because the coils do not contact the workpiece, they leave no marks or scratches. Induction coils are usually made from a water-cooled copper conductor that has 1 to 3 turns.

The delivery of HF power through electrical contacts is more efficient than that through the use of coils because the power can be more directly and efficiently distributed to the weld area. The disadvantages of contacts are that they wear out, require periodic replacement, and can scratch, mar, or leave deposits on the surface of a workpiece, which can be objectionable in some cases.

In the production of tube and pipe, induction coils are usually used to make the smaller diameters (<305 mm, or 12 in.) because their advantages outweigh the loss of efficiency. Contacts are generally used for larger diameters because of their greater efficiency. Contacts are also used in structural shapes and fin heat-exchanger tubes, as well as for end welding, because induction coils cannot induce the current densities necessary for welding.

Impeders are used to improve welding efficiency in tube and pipe production. Impeders that consist of one or more water-cooled cores of ferrite material are placed inside the tube parallel to and under the weld area. The reason for this is the existence of a power-dissipating circuit loop consisting of the inside surface of the tube being welded. The impeder increases the inductive impedance of this path, which bypasses the weld zone.

Inert gases are not needed for HF welding unless certain reactive metals are used. Argon is sometimes employed for the welding of stainless steel, and it is usually necessary for the welding of titanium. Although cleanliness of weld surfaces is not particularly necessary for easily welded materials, a lack of cleanliness is often the cause of problems in more-sensitive metals.

High-Frequency Welding

Warren Smith and Julian Roberts, Thermatool Corporation

Personnel

A considerable amount of on-the-job training is required for new operators of HF welding equipment, as would be the case in any other mill process. High-frequency welding is a high-speed and, usually, high-volume process that often operates in conjunction with precision forming and handling equipment. The operating speed of tube mills ranges from 30 to 240 m/min (100 to 800 ft/min). A tube must be precisely formed prior to welding. An operator, therefore, must be familiar with both the operation of the welding equipment and the associated mechanical equipment for the forming or handling of the metal. Troubleshooting and repair of the welding equipment generally require specially trained personnel and are often conducted by the manufacturer. Improved and simplified systems for speed and/or power and weld temperature control are reducing the operator dependence of the process.

High-Frequency Welding

Warren Smith and Julian Roberts, Thermatool Corporation

Safety

Serious consideration must be given to the health and safety of welding operators, maintenance personnel, and other workers in the area of the welding operations. Good engineering practice must be followed in the design, construction, installation, operation, and maintenance of equipment, controls, power supplies, and tooling in order to conform to federal, state, and local safety regulations, as well as those of the company using the equipment.

Because HF welding power sources are electrical devices, they require all the usual precautions in handling and repairing such equipment. Voltages, which range from 400 to 25,000 V, can be either low or high frequency and may be lethal. To prevent injury, proper care and safety precautions should be taken while working on high-frequency systems and generators. Modern units are equipped with safety interlocks on access doors, as well as automatic safety grounding devices that prevent unsafe operation of the equipment. The equipment should not be operated when panels or high-voltage covers have been removed or when interlocks and grounding devices have been blocked.

High-frequency, high-voltage leads must be encased in grounded metal ductwork to ensure safety and to minimize electromagnetic radiation. Induction coils and contact systems should always be properly grounded for operator protection. Grounding lines should be kept short to minimize inductive reactance. Injuries from high-frequency power, especially at the upper range of welding frequencies, tend to produce severe, local, surface tissue damage. Care should be taken that the magnetic field from the output system, particularly the output transformer, does not heat adjacent metallic sections by induction.

Metal or flux fumes should always be vented and the correct safety glasses should be worn in the presence of liquid metal or high-intensity visible radiation.

Warren Smith and Julian Roberts, Thermatool Corporation

Inspection and Quality Control

Numerous nondestructive and destructive tests can be used to determine weld quality. For example, weld quality can be determined nondestructively by eddy current testing (ASTM E 309 and ASTM A 426), ultrasonic testing (ASTM E 213 and ASTM A 273), flux leakage examination (ASTM E 570), and hydrostatic pressure testing (ASTM 450).

A number of destructive tests that utilize small samples of the product are identified in ASTM A 450. These tests include the flattening test (crush test), reverse flattening test, and the flare test.

Metallographic examination and microhardness testing can also be used to determine weld quality and the character of the adjacent heat-affected zone.

Introduction

ELECTRON-BEAM WELDING (EBW) is a high-energy density fusion process that is accomplished by bombarding the joint to be welded with an intense (strongly focused) beam of electrons that have been accelerated up to velocities 0.3 to 0.7 times the speed of light at 25 to 200 kV, respectively. The instantaneous conversion of the kinetic energy of these electrons into thermal energy as they impact and penetrate into the workpiece on which they are impinging causes the weld-seam interface surfaces to melt and produces the weld-joint coalescence desired. Electron-beam welding is used to weld any metal that can be arc welded: weld quality in most metals is equal to or superior to that produced by gas-tungsten arc welding (GTAW).

Because the total kinetic energy of the electrons can be concentrated onto a small area on the workpiece, power densities as high as 10^8 W/cm^2 (10^7 W/in.^{2}) can be obtained (Ref 1). That is higher than is possible with any other known continuous beam, including laser beams. The high-power density plus the extremely small intrinsic penetration of electrons in a solid workpiece results in almost instantaneous local melting and vaporization of the workpiece material. That characteristic distinguishes EBW from other welding methods in which the rate of melting is limited by thermal conduction.

Acknowledgements

This article was adapted from "Electron Beam Welding" in the 9th Edition Volume 6 *Metals Handbook.* Special thanks are due to the committee that prepared the original article.

Reference

1. H.B. CARY, *MODERN WELDING TECHNOLOGY,* 2ND ED., PRENTICE HALL, 1989, P 256

Principles of Operation

Figure 1 shows an electron-beam weld being performed at a weld chamber pressure of approximately 3 Pa (2 × 10^{-2} torr). The "beam" shown in Fig. 1 is the visible glow that results from the residual (ambient) gas molecules that are excited by the electrons in the actual electron beam.

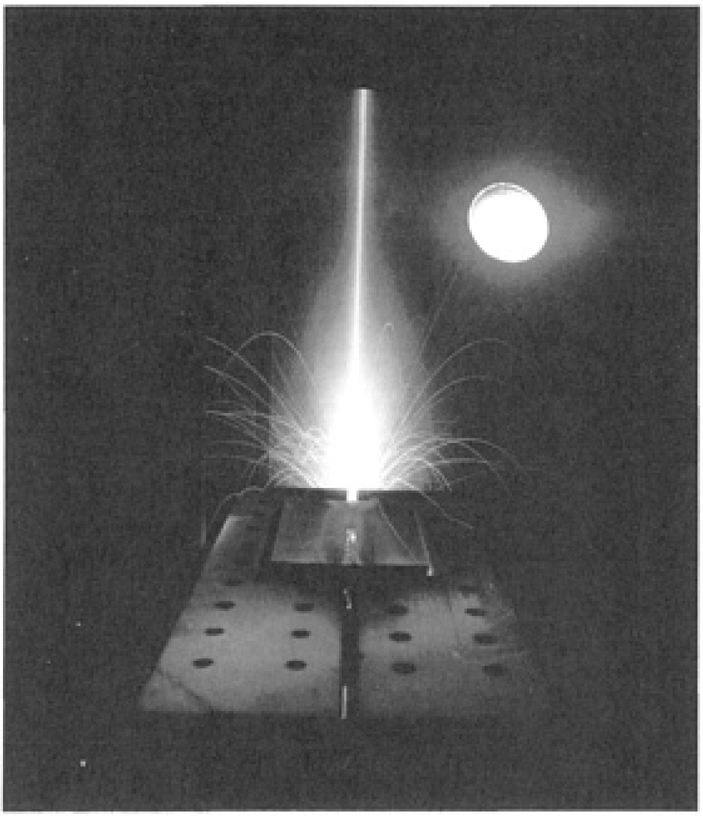

FIG. 1 WELD GENERATED BY AN ELECTRON BEAM OPERATING IN AN ENCLOSED CHAMBER MAINTAINED AT A PRESSURE OF APPROXIMATELY 3 PA (2 × 10^{-2} TORR)

Basically, the electron beam is formed (under high-vacuum conditions) by employing a triodestyle electron gun consisting of a cathode, a heated source (emitter) of electrons that is maintained at some high negative potential; a grid cup, a specially shaped electrode that can be negatively biased with respect to the hot cathode emitter (filament); and an anode, a ground potential electrode through which the electron flow passes in the form of a collimated beam. The hot cathode emitter (filament) is made from a high-emission material, such as tungsten or tantalum. This emitter material, usually available in wire, ribbon, or sheet form, is fabricated into the desired shape for being either directly or indirectly heated to the required emitting temperature of about 2500 °C (4500 °F).

Electrons emitted from the surface of the filament are accelerated to a high velocity and shaped into a collimated beam by the electrostatic field geometry generated from the cathode/grid/anode configuration employed, thus producing a steady stream of electrons that flows through an aperture in the ground plane anode. By varying the negative potential difference between the grid and cathode, this flow of electrons can be altered easily (i.e., gated "on/off" or ramped up/down to different levels) in a precisely controlled manner.

Diode-style electron guns are also employed, but not to the extent that triode-style electron guns are. In a diode gun, the specially shaped electrode (grid cup) is maintained at the same voltage as the emitter, thus making the diode gun a two-element (cathode and anode) device. With this design, the flow of electrons from a diode gun cannot be adjusted by simply varying a grid voltage, as is done with triode guns, and beam current adjustments are usually accomplished by varying the operating temperature of the cathode emitter instead.

Once the electrons exit the anode, they receive the maximum energy input allowable from the operating voltage being applied to the gun. Electrons then pass down through the electron beam column assembly and into the field of an electromagnetic focusing coil (a magnetic lens). This focusing lens reduces the diameter of the electron beam, as it continues in its passage, and focuses the stream of electrons down to a much smaller beam cross section in the plane of the workpiece. This reduction in beam diameter increases the energy density, producing a very small, high-intensity beam spot at the workpiece. In addition, an electromagnetic deflection coil (positioned below the magnetic lens) can be employed to "bend" the beam, thus providing the flexibility to move the focused beam spot. Figure 2 illustrates the main elements of the electron beam welding head. As described above and shown in Fig. 2, electrons are emitted from the cathode and are accelerated to high speed by the voltage between cathode and anode.

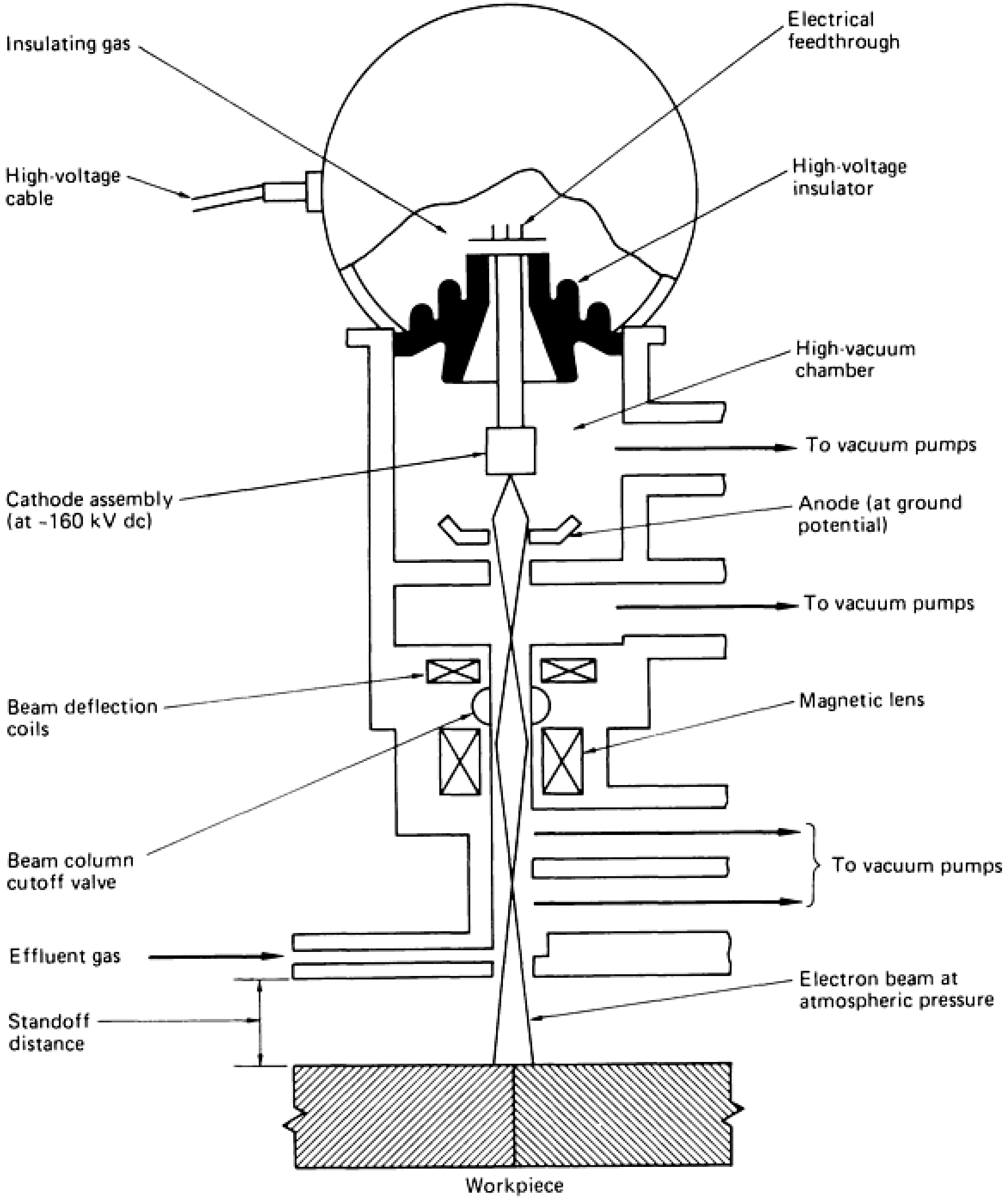
Insulating gas
Electrical feedthrough
High-voltage cable
High-voltage insulator
High-vacuum chamber
To vacuum pumps
Cathode assembly (at -160 kV dc)
Anode (at ground potential)
To vacuum pumps
Beam deflection coils
Magnetic lens
Beam column cutoff valve
To vacuum pumps
Effluent gas
Standoff distance
Electron beam at atmospheric pressure
Workpiece

FIG. 2 SCHEMATIC SHOWING PRIMARY COMPONENTS OF AN ELECTRON-BEAM WELDING HEAD

The "gun" portion of an electron gun/column assembly generally is isolated from the welding chamber through the use of valves when desired. The gun may be maintained in a vacuum on the order of 13 mPa (1 × 10^{-4} torr) when the welding chamber is vented to atmosphere (for access reasons). This level of vacuum in the gun region is needed to maintain gun component cleanliness, prevent filament oxidation, and impede gun arcing (high-pressure short circuiting between electrodes at different voltages). During welding, this same degree of vacuum is required in both the gun column and welding chamber areas to minimize scattering of beam electrons by excitation collisions with residual air molecules as they traverse the distance from the gun to the workpiece. This type of interaction tends to produce a broader beam spot and a resulting decrease in energy density. Generally, electron guns are operated with applied voltages that vary from 30 to 200 kV, and they employ beam currents that range from 0.5 to 1500 mA. Electron-beam welding equipment with power levels up to 30 kW is common, and several units with power levels of up to 200 kW are commercially available.

Typically, high-vacuum EBW beams can be focused down to spot sizes in the range of 0.25 to 1.3 mm (0.010 to 0.050 in.) in diameter, with a power density of about 10^7 W/cm^2 (10^6 W/in.^{2}). This high level of beam spot intensity generates temperatures of approximately 14,000 °C (25,000 °F) and is sufficient to vaporize almost any material, forming a vapor hole that penetrates deep into the workpiece. When this vapor hole is advanced along a weld joint, the weld is produced by three effects that occur simultaneously: (1) the material at the leading edge of the vapor hole melts; (2) this molten material flows around the sides of the vapor hole to the trailing edge; and (3) this continuous flow of molten material fills in the trailing edge of the advancing vapor hole and solidifies as the vapor hole moves forward to produce a continuous weld.

Originally, EBW generally was performed only under high-vacuum ($\leq$13 mPa, or 1 × 10^{-4} torr) conditions; because an ambient vacuum environment was required to generate the beam, welding the part within the same clean atmosphere was considered beneficial. However, as the demand for greater part production increased, it was found that the weld chamber vacuum level need not be as high as that needed for the gun region; ultimately, the need for any type of vacuum surrounding the workpiece was totally eliminated for some applications. Currently, three distinct modes of EBW are employed:

- HIGH-VACUUM (EBW-HV), WHERE THE WORKPIECE IS IN AN AMBIENT PRESSURE RANGING FROM 0.13 TO 0.30 MPA (10^{-6} TO 10^{-3} TORR)
- MEDIUM-VACUUM (EBW-MV), WHERE THE WORKPIECE MAY BE IN A "SOFT" OR "PARTIAL" VACUUM RANGING FROM 0.13 TO 3300 PA (10^{-3} TO 25 TORR)
- NONVACUUM (EBW-NV), WHICH IS ALSO REFERRED TO AS ATMOSPHERIC EBW, WHERE THE WORKPIECE IS AT ATMOSPHERIC PRESSURE IN AIR OR PROTECTIVE GAS

In all EBW applications, the electron-beam gun region is maintained at a pressure of 13 mPa (10^{-4} torr) or lower.

Advantages

One of the prime advantages of EBW is the ability to make welds that are deeper and narrower than arc welds, with a total heat input that is much lower than that required in arc welding. This ability to achieve a high weld depth-to-width ratio eliminates the need for multiple-pass welds, as is required in arc welding. The lower heat input results in a narrow workpiece heat-affected zone (HAZ) and noticeably fewer thermal effects on the workpiece.

In EBW, a high-purity vacuum environment can be used for welding, which results in freedom from impurities such as oxides and nitrides. The ability to employ higher weld speeds, due to the high melting rates associated with the concentrated heat source, reduces the time required to accomplish welding, thereby resulting in an increased productivity and higher energy efficiency for the process. Total energy conversion efficiency of EBW is approximately 65%, which is slightly higher than so-called conventional welding processes and much higher than other types of high-energy-density welding processes, such as laser-beam welding (EBW). Additional information is available in the article "Laser-Beam Welding" in this Section.

These characteristics (1) minimize distortion and shrinkage during welding; (2) facilitate welding of most hardened or work-strengthened metals, frequently without significant deterioration of mechanical properties in the weld joint; (3) facilitate welding in close proximity to heat-sensitive components or attachments; (4) allow hermetic seal welding of evacuated enclosures, while retaining a vacuum inside the component; and (5) permit welding of refractory metals, reactive metals, and combinations of many dissimilar metals that are not joinable by arc welding processes. The ability to project the electron beam a distance of over 510 mm (20 in.) under high-vacuum conditions, as well as the low end of medium-vacuum conditions, allows otherwise inaccessible welds to be completed.

Limitations

Equipment costs for EBW generally are higher than those for conventional welding processes. However, when compared to other types of high-energy density welding (such as LBW), production costs are not as high. The cost of joint preparation and tooling is more than that encountered in arc welding processes, because the relatively small electron beam spot size that is used requires precise joint gap and position.

The available vacuum chamber capacities are limited; workpiece size is limited, to some degree, by the size of the vacuum chamber employed. Consequently, the production rate (as well as unit cost) is affected by the need to pump down the chamber for each production load. Because the electron beam is deflected by magnetic fields, nonmagnetic or degaussed metals must be used for tooling and fixturing that are near the beam path.

Although most of the above advantages and disadvantages generally are applicable to all modes of EBW, several do not specifically apply to EBW-NV. Nonvacuum EBW does not offer the advantage of a high-purity environment (unless some form of inert-gas shielding is provided), and it is not subject to vacuum chamber limitations. Because welding is not done within the confines of a vacuum environment, the maximum practical "standoff," the working distance between the bottom of the electron beam column and the top of the workpiece, currently used on EBW-NV systems is limited to approximately 35 mm ($1\frac{3}{8}$ in.). The maximum attainable penetration is currently limited to about 25 to 32 mm (1 to 1.25 in.) for 60-kW, 165 kV machines. These limitations, however, are offset by higher production rates and a lower rate of vacuum-related equipment problems that are typical of EBW-NV systems.

Process Control

Basic variables employed for controlling the results of an electron-beam weld include accelerating (applied gun) voltage, beam current, welding (beam spot travel) speed, focusing current, and standoff (gun column assembly to workpiece) distance. The final beam spot size that is produced in the plane of the workpiece is determined by:

- CHARACTERISTICS OF THE ELECTRON BEAM GUN/COLUMN ASSEMBLY (GUN AND ELECTRON OPTICS)
- FOCUSING CURRENT, WHICH CONTROLS THE FOCAL LENGTH OF THE LENS AND RESULTING BEAM FOCUS LOCATION
- GUN COLUMN ASSEMBLY-TO-WORKPIECE DISTANCE
- ACCELERATING VOLTAGE
- BEAM CURRENT

Each of these variables, separately and jointly, affects final beam spot size.

Increasing the accelerating voltage or beam current increases the depth of penetration; the product of these two variables--the beam power--determines the amount of metal melted for a given exposure time. Increasing welding speed--relative travel motion between beam spot and workpiece--without changing any other process variable, reduces depth of penetration and correspondingly reduces weld width. Changing any of the other basic control variables to increase beam spot size, thereby lowering the beam spot intensity (e.g., reducing the power density), reduces depth of penetration and increases weld width, if welding speed is left unchanged. Beam deflection can be used to change the impact angle of the beam (slightly) or to produce controlled patterns of beam oscillation to create greater beam spot size or other special effects. The beam can also be pulsed to vary the average amount of power inputted per unit of time.

With EBW-NV, the accelerating voltage and focus current are normally preset and held at a fixed value. Beam current may be preset for any application and then turned "on/off" (or programmed to vary in predetermined fashion) with workpiece travel. Workpiece speed and, in certain instances, standoff distance may also be preset to a desired value or varied with workpiece travel. On many of the latest computerized numerical control (CNC) types of EBW-NV systems, beam current, standoff distance, and welding speed are varied simultaneously, as required, throughout the length of the weld. Because of the limited standoff distance involved with EBW-NV, beam deflection normally is not feasible, and beam oscillatory motion generally is not employed.

Operation Sequence and Preparation

Tooling and welding procedures for each EBW application are developed first on experimental workpieces. Details of the welding sequence vary somewhat, depending on equipment differences and application requirements. A typical sequence of operation for welding in vacuum is to:

- ASSEMBLE AND PREPARE WORK AND FIXTURES FOR WELDING. THIS INCLUDES CLEANING AND MAY INCLUDE DEMAGNETIZING, PREHEATING, AND TACK WELDING.
- LOAD FIXTURED WORK ONTO WORKTABLE OR WORK-HOLDING MECHANISM IN WELDING CHAMBER.
- START CHAMBER PUMPDOWN.
- AFTER CHAMBER PRESSURE HAS BEEN REDUCED TO 0.013 TO 13 PA (10^{-4} TO 10^{-1} TORR), FOCUS ON A TARGET BLOCK AND SET BEAM PARAMETERS.
- ALIGN JOINT TO THE BEAM POSITION, USING VERY LOW POWER BEAM SPOT.
- BEGIN WELDING; THIS USUALLY IS PERFORMED AUTOMATICALLY, BUT CAN BE PERFORMED MANUALLY.
- TERMINATE THE WELDING CYCLE.
- ALLOW WORK TO COOL SUFFICIENTLY IF MADE OF REACTIVE MATERIAL, THEN ADMIT AIR TO THE CHAMBER AND REMOVE FIXTURED WORK.

For nonvacuum welding, the electron gun/column assembly is maintained at established pressure levels; work-handling and welding operations for EBW-NV usually are mechanized for high-speed production, and work and fixturing are prepared before a production run is begun, or they are completed as part of the production line operation. Beam parameters, alignment of joint with beam position, and transfer and movement of work for welding are established before a production run. Alignment of the joint with beam position is slightly less critical than for vacuum electron-beam welding, because beam spot size and weld width are slightly larger in nonvacuum welding.

Joint Preparation. A joint for EBW ordinarily has close fitted, abutting, square-groove faces, and filler metal usually is not used. Filler metal can be used if desired, as in EBW of dissimilar metals and alloys. Generally, the faces of the joint are machined to a surface roughness of 3.20 μm (125 μin.) or less.

Surface finish on the weld groove faces may not be critical, depending on part and joint design and the requirements for weld properties. In studies on butt welding of 50 mm (2 in.) thick aluminum alloy 2219 and Ti-6A1-4V specimens that had groove faces with surface roughness values of 1.60, 3.20, 6.3, 12.5, and 25.0 μm (63, 125, 250, 500, and 1000 μin.), respectively, surface finish had no effect on weld quality, as determined by visual examination and x-ray radiography; all welds were sound. Edge surface finish is much less critical on broad welds than on narrow welds. Edge roughness is not critical on lap joints in thin metal, as long as burrs do not separate the surfaces.

Joint Fit-Up. A butt joint is not open, as in arc welding, but has closely fitted, nearly parallel surfaces to enable the narrow electron beam to fuse base metal on both sides of the joint. The members of a joint to be melt-through welded also are closely fitted. Fit-up tolerance depends on work-metal thickness and joint design, but is usually 0.13 mm (0.005 in.) or less. Joint gap is usually smaller for thin work metal and unbacked joints and may be only 0.05 mm (0.002 in.) maximum. Interference fits may be used where shrinkage can cause cracking, as in circular joints on hardenable metals.

Joint gaps of about 0.08 mm (0.003 in.) maximum commonly are used for making narrow vacuum welds, whereas joint gaps of up to 0.8 mm (0.030 in.) are not uncommon in EBW-NV. In making deep welds, poor fit-up or too large a joint gap can cause excessive shrinkage, underfill, undercut, voids, cold shuts, and missed joints. In most metals, joint gap should not exceed 0.25 mm (0.010 in.) for narrow welds deeper than about 50 mm ($\frac{1}{2}$ in.), although sound welds have been obtained using joint gaps of 0.5 mm (0.020 in.) by increasing the weld width to about 6.4 mm (0.25 in.).

Cleaning. Workpiece surfaces must be properly cleaned for in-vacuum EBW. Inadequate surface cleaning of the weld metal can cause weld flaws and a deterioration of mechanical properties. Inadequate cleaning of weld surfaces also adversely affects pumpdown time and gun operational stability, as well as contributing to the rapid degradation of the oil used in the vacuum pumps.

In EBW-NV, cleanliness of workpiece surfaces is important, but not as critical as with vacuum EBW. Although the effect of workpiece cleanliness on pumpdown time and gun stability is reduced in nonvacuum welding, it still affects final weld quality.

Wire brushing generally is not recommended, because contaminants may become embedded in the metal surface. Acetone is preferred for cleaning electron gun components and workpiece parts. If workpieces are cleaned in chlorine or other halogen-containing compounds, residue from these compounds must be removed by another cleaning method (usually thorough washing in acetone) before welding.

Fixturing Methods. Methods used for fixturing workpieces in EBW are similar to those used for GTAW of precision parts without the use of filler metal, except that clamping force is usually lower and all fixturing and tooling materials should be made of nonmagnetic materials.

Because total heat input to the weld is much less in EBW than for arc welding, and because the heat is highly localized, heavy fixturing, massive heat sinks, or water cooling is not needed. C-clamps are sufficient for many parts. In some applications, clamping may be supplemented or replaced by small tack welds or by a shallow weld pass (sealing pass) over the joint; the penetration weld is completed later at full power.

For EBW-NV, general-purpose welding positioners are satisfactory. Locating and aligning mechanisms are of simpler design than for welding in a vacuum; accessibility and beam widths for nonvacuum welding do not require as much accuracy in tracking of the joint. Maintenance of work-handling equipment for EBW-NV is simplified because of its out-of-chamber location.

Demagnetization. Workpieces and fixtures made of magnetic materials should be demagnetized before welding. Residual magnetism may result from magnetic-particle testing, magnetic chucks, or electrochemical machining. Even a small amount of residual magnetism can cause beam deflection. Workpieces are usually demagnetized by placing them in a 60 cycle inductive field and then slowly removing them. Equipment used for magnetic-particle testing may be used. Before welding, workpieces should be checked with a gauss meter. Acceptable gauss-meter readings vary from 5×10^{-5} T ($\frac{1}{2}$ G) for very narrow (highly critical) welds up to as much as 2×10^{-4} to 4×10^{-4} T (2 to 4 G) for relatively wide welds.

Pumpdown. The time required to pump the work chamber of an in-vacuum EBW unit down to the desired ambient pressure depends on the chamber size used, type of pump employed, and level of vacuum required. Electron-beam welding equipment currently employs either computer or program logic controls, normally providing automatic vacuum sequencing.

For EBW-HV, a pressure of 13 mPa (10^{-4} torr) or less is produced by a mechanical pump (either a simple piston, a rotary vane type, or a combination pump-and-blower package), operating in conjunction with an oil-diffusion pump. The size of these pumps depends mainly on the size of the weld chamber used, as well as on the final operating vacuum level and total pump-down time needed.

Generally, a 1.1 m^3 (40 ft^3) chamber employing a 250 mm (10 in.), 250 m^3/min (8820 ft^3/min) diffusion pump and a 17 m^3/min (600 ft^3/min) mechanical pump package can be expected to reach a welding pressure of less than 40 mPa (3×10^{-4} torr) in approximately 4 min, while an 11 m^3 (400 ft^3) chamber employing a 510 mm (20 in.), 1070 m^3/min (37,800 ft^3/min) diffusion pump and a 37 m^3/min (1300 ft^3/min) mechanical pump package can be expected to pump down to the same pressure in approximately 12 min.

If desired, the pumpdown time of the 11 m^3 (400 ft^3) chamber could be reduced to under 6 min by increasing the size of the pumping system to an 815 mm (32 in.) 3090 m^3/min (109,200 ft^3/min) diffusion pump and a 77.9 m^3/min (2750 ft^3/min) mechanical pump package; however, the higher cost associated with this increased pumping capability would have to be evaluated against the financial benefit of the saving in pumpdown time to determine whether it would be economically feasible. In the examples above, weld chambers are assumed to be clean, dry, and empty. The cleanliness of the work chamber, the amount of water vapor (humidity) in the ambient air, and the workpiece assembly (material and shape of both weldment and fixturing) thus affect the actual pumpdown time obtained in any production application.

For partial-vacuum EBW, where the pressure level required in the weld chamber is approximately 13 Pa (10^{-1} torr) and a pressure level of 13 mPa (10^{-4} torr) is provided only in the electron-beam gun/column regions, diffusion pumps are not needed for the weld chamber, and weld chamber pumping is accomplished strictly by mechanical pumping. For high-production EBW of small parts in a partial vacuum, pumpdown lime ranges from about 20 s for relatively large-volume (general-purpose) partial-vacuum weld chambers of about 0.14 m^3 (5 ft^3) down to less than 5 s for fairly small-volume chambers of 0.014 m^3 (0.5 ft^3), with the use of a mechanical pump rated at 37 m^3/min (1300 ft^3/min).

Preheat and Postheat. Most commonly welded metals can be processed with EBW methods, even in thick sections without preheating, because of the extremely narrow width of the HAZ. Hardenable and difficult-to-weld metals may need to be preheated, especially for thick sections and in applications when the weld is restrained.

High-strength alloy steels and tool steels thicker than about 9.5 mm ($\frac{3}{8}$ in.) ordinarily must be preheated before EBW to prevent cracking. Deep circular welds, especially partial-penetration welds, in thick sections of carbon steel containing more than about 0.35% C usually require preheating. However, welds subject to less restraint, such as circumferential welds on cylindrical shapes, can be made on 13 mm ($\frac{1}{2}$ in.) thick 0.50% C steel without preheating.

Preheating, when required, is usually done before the work is placed in the work chamber. Selection of heating method depends on the size and shape of the work and the preheat temperature; a combination of methods can be used. Torch and furnace heating are widely used; induction and infrared-radiation heating are also used. On small parts or where distortion from localized heating is not a problem and where increased cycle time can be tolerated, heating is sometimes done with a defocused electron beam; this method can also be used to supplement other methods of heating.

When postheating is used on EBW parts, it is accomplished by conventional means (i.e., stress relieving or tempering) after removing the work from the welding chamber.

Operating Conditions. Starting and stopping the weld usually require special consideration to avoid uniformity of the weld at these points and possible melt-through and loss of metal.

One technique that is used to avoid these difficulties is to start the weld at full beam power on a starting tab of the work metal that is tightly fitted against one end of the joint and to conclude the weld on a runoff tab at the other end of the joint. The use of starting and runoff tabs prevents underfill at the ends of the joint, which is caused by the introduction or exit of the beam. Workpieces can also be made oversize to provide extra material for starting and stopping. The tabs or the extra material can be machined off after welding. Starting and runoff tabs are used mainly in low-production operations.

Another technique is to start and stop the weld on the work, raising the current gradually (upslope) at the beginning of the weld and reducing it gradually (downslope) at the end of the weld. Upslope and downslope power may be used at controlled rates and time intervals, as established for a specific application, and can be programmed into the welding procedure. The use of upslope and downslope is of special value where the weld is a closed path, as in welding circular and circumferential joints.

In many applications of closed-path welds, the weld can be started at full power, with downslope at the end providing a sufficiently gradual termination and a suitable distance of overlap. Downslope of current after overlapping the beginning of the weld on circular welds in thick sections of high-hardenability steels is critical to avoid porosity and cracking. Other parameters may also be adjusted to avoid these defects. Upslope is usually rapid; downslope is from a few degrees to a major portion of a revolution.

A "cosmetic pass" is made when needed to smooth or flatten the crown of a weld that is irregular or too high. Such a pass is used to correct undercut or underfill. The beam is usually defocused or reduced in power, or both, in making a cosmetic pass. Filler metal may also be added in making a cosmetic pass intended to correct undercut or underfill.

Weld Geometry

The shape of the parts to be welded and the corresponding joint designs are critical to the successful application of EBW in vacuum or at atmospheric pressures (nonvacuum). While minimum heat input and low thermal distortions are important advantages of EBW, the molten metal still shrinks as it solidifies. Shrinkage stresses may lead to microcracks if parts, due to design restrictions, are unable to shrink at corresponding rates and the joint volume is completely constrained. Additional information is available in the article "Procedure Development and Practice Considerations for Electron-Beam Welding" in this Volume.

Joint Design

Butt, corner, T-, lap, and edge joints can be made by the EBW process using square-groove or seam welds. Fillet welds, which are difficult to make with vacuum EBW, are readily made using EBW-NV. Square-groove welds require fixturing to maintain fit-up and alignment of the joint. They can, however, be self-aligning if a rabbeted joint design is used. Self-alignment is particularly important in batch loading the vacuum chamber for efficient work manipulation. The weld-metal area can be increased using a scarf joint, but fit-up and alignment of the joint are more difficult than with a square-groove weld. Most joints are designed to be welded in a single pass with full penetration or penetration to a specified depth.

Electron-Beam Welding Machines

This brief review is aimed at introducing the potential user to some of the EBW equipment currently available. Further advances in the control and programming of these systems can be expected because of the likelihood of uninterrupted progress in the electronics and computer industries.

As discussed earlier in this article, the EBW process is accomplished at three pressure-dependent lines, referred to as the three modes of EBW. The original mode is the high or "hard" vacuum mode in which welding is carried out in the pressure range of 0.13 to 130 mPa (10^{-6} to 10^{-3} torr). In the second mode, medium vacuum, the pressure ranges from 0.13 to 3300 Pa (10^{-3} to 25 torr). The term "medium vacuum" includes the range of pressure (0.13 to 130 Pa, or 10^{-3} to 1 torr) referred to as the "soft" or partial vacuum. The third mode is called non-vacuum or atmospheric, with welding carried out at atmospheric pressure.

All three modes employ an electron-beam gun/column, a power supply with controls, one or more vacuum pumping systems, and work-handling equipment. Although in nonvacuum welding the workpiece is not placed in an evacuated work chamber, the electron beam gun/column must be in a vacuum environment. The electron gun in all three modes is held at a pressure of 13 mPa (10^{-4} torr) or less; otherwise, the high voltage required for the acceleration of the electrons could not be sustained.

Electron-beam welding equipment comes in two basic designs: (1) the low-voltage system, which uses accelerating voltages in the 15 to 60 kV range; and (2) the high-voltage system, with accelerating voltages in the 100 to 200 kV range. Beam powers up to 100 kW are available with both high-voltage and low-voltage equipment.

The lower-voltage machines operate at a higher current. Typically, 30 to 60 kV machines operate at a 500 mA beam current. The high-voltage machines of 150 kV operate at 40 mA. The higher-voltage machines produce a greater depth-to-width ratio of the weld nugget. This could be the difference between a 12:1 depth-to-width ratio versus a 25:1 depth-to-width ratio. The higher-voltage machines can utilize a longer standoff distance than can low-voltage machines; however, low-voltage machines are simpler in construction and less maintenance is required (Ref 1).

Low-voltage equipment is more suitable for high- and medium-vacuum operations, while nonvacuum welding is carried out at higher voltages (130 to 175 kV minimum). All three modes, however, are operational with high-voltage equipment. The nonvacuum, high-voltage systems generally are used for welding materials less than 25 mm (1 in.) thick. With low-voltage systems, the gun may be fixed in position on the chamber or may be mobile inside the chamber. With high-voltage equipment, the gun is generally fixed in position on the chamber. Figure 3 shows a typical floor plan of a high-voltage, high-vacuum welding facility with a welding chamber size of 1725 by 1725 by 1980 mm (68 by 68 by 78 in.).

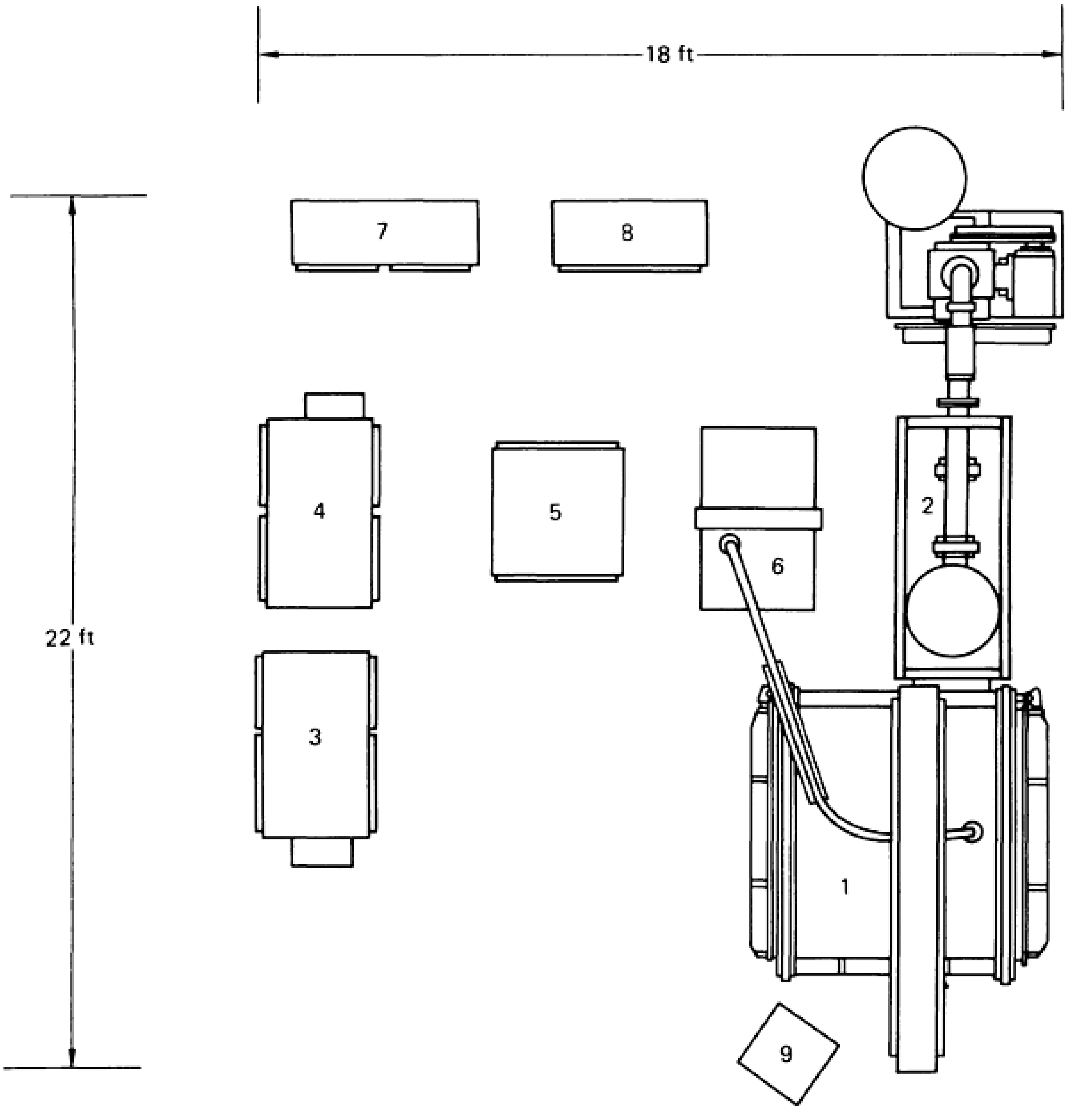

FIG. 3 OVERHEAD VIEW SHOWING TYPICAL FLOOR LAYOUT OF HIGH-VOLTAGE, HIGH-VACUUM EBW SYSTEM. (1) VACUUM CHAMBER AND MANIPULATING MECHANISMS. (2) VACUUM PUMPING SYSTEM. (3) COMPUTER CONTROL CABINET. (4) SERVO CABINET. (5) AUTOMATIC VOLTAGE REGULATOR. (6) HIGH-VOLTAGE POWER SUPPLY. (7) POWER DISTRIBUTION CABINET. (8) PUMPING CONTROL CABINET. (9) OPERATOR CONSOLE

Figure 4 shows a standard high-vacuum machine with a chamber size of 2845 by 1525 by 1830 mm (112 by 60 by 72 in.). This type of unit is available with power ratings of 7.5 kW (150 kV, 50 mA) and 25 kW (150 kV, 267 mA). The gun/column in this unit is fixed, and the work travels under the gun. Work-handling equipment makes processing of workloads of up to 1400 kg (3000 lb) at welding speeds of 3300 mm/min (130 in./min) possible. This system can be obtained with various degrees of automation; most utilize CNC operations. Figure 5 shows a high-vacuum unit with a chamber size of 3505 by 2845 by 2690 mm (138 by 112 by 106 in.). This installation comes with a movable gun. Low-voltage, high-vacuum systems with movable guns are available with power ratings of 7.5 kW (60 kV, 125 mA) 15 kW (60 kV, 250 mA), 30 kW (60 kV, 125 mA), and 42 kW (60 kV, 700 mA).

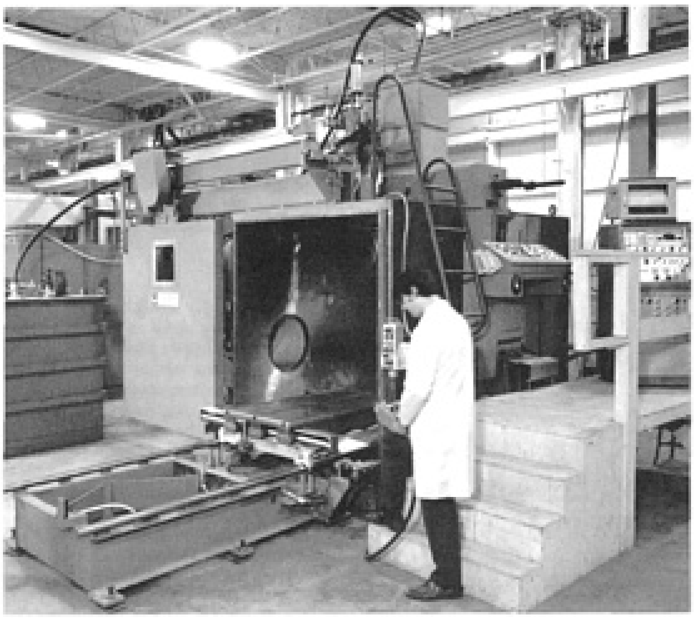

FIG. 4 TYPICAL EBW-HV UNIT WITH A 2845 × 1525 × 1830 MM (112 × 60 × 72 IN.) CHAMBER. COURTESY OF LEYBOLD HERAEUS VACUUM SYSTEMS INC.

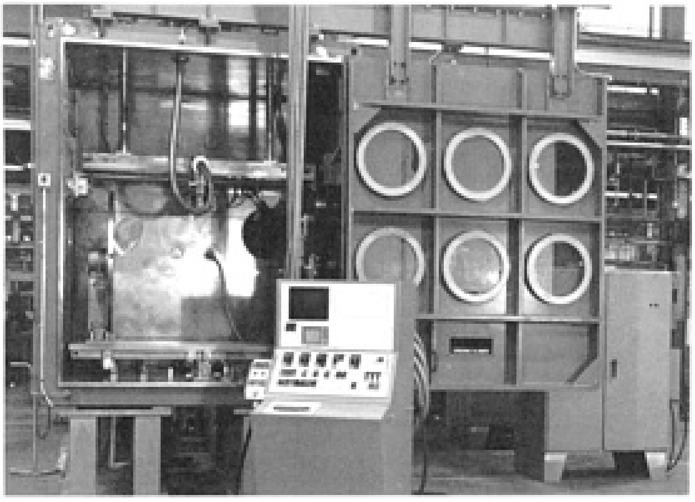

FIG. 5 STANDARD LOW-VOLTAGE EBW-HV UNIT CONSISTING OF A 3505 × 2845 × 2690 MM (138 × 112 × 106 IN.) CHAMBER. COURTESY OF SCIAKY BROS., INC.

Medium-vacuum machines are frequently special-purpose units tooled for particular assemblies. Figure 6 shows a dual-system partial-vacuum welding installation for joining ring gear and counterweight onto flywheels. Dual medium-vacuum systems of this type, which are used for high-production applications, can be readily modified to produce alternate parts by simple changes in the work-handling components and weld programming. As with the system shown in Fig. 4, a wide range of automation and computer control is available.

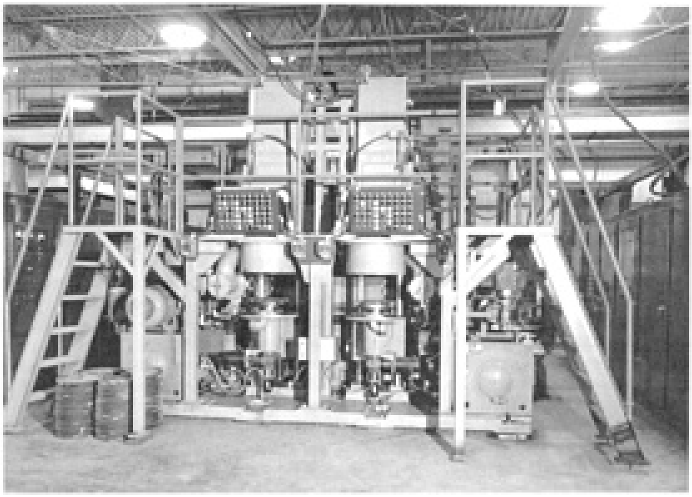

FIG. 6 TWIN EBW-MV INSTALLATION DESIGNED TO WELD A RING GEAR ONTO A FLYWHEEL. COURTESY OF LEYBOLD-HERAEUS VACUUM SYSTEMS INC.

The medium-vacuum mode of EBW was initially employed to increase the production capacity of the process. This mode of application typically involves the use of a chamber custom designed to provide a minimum of weld-zone volume that has to be evacuated during each part cycle (that is, a chamber envelope that is not much bigger in size than the particular workpiece to be welded plus any fixturing it may require). Because only a small volume needs to be evacuated, coupled with the fact that medium-vacuum EBW is normally performed in an ambient vacuum level of nominally 6.5 Pa (5×10^{-2} torr)--which is several orders of magnitude higher in pressure than high-vacuum EBW--the pumpdown times necessary are much less than those required on a high-vacuum EBW system (that is, times of <10 s compared to >3 min). Thus, in order to take advantage of this significant reduction in the time required for evacuating the weld zone, various methods were developed to rapidly transport parts into or out of the weld area on EBW-MV systems. Figure 7 shows one of the conventional methods used for this purpose, employing a dial table arrangement for repeatedly indexing a set of work stations through the weld area. Each individual work station contains whatever fixturing might be required to hold the part. The parts can usually be loaded/unloaded into and out of these work stations either manually or automatically. This particular method for achieving rapid part transfer is capable of being employed both with the drop-bottom (discretely pumped) EBW-MV unit and the sliding-seal (continuously pumped) EBW-MV unit.

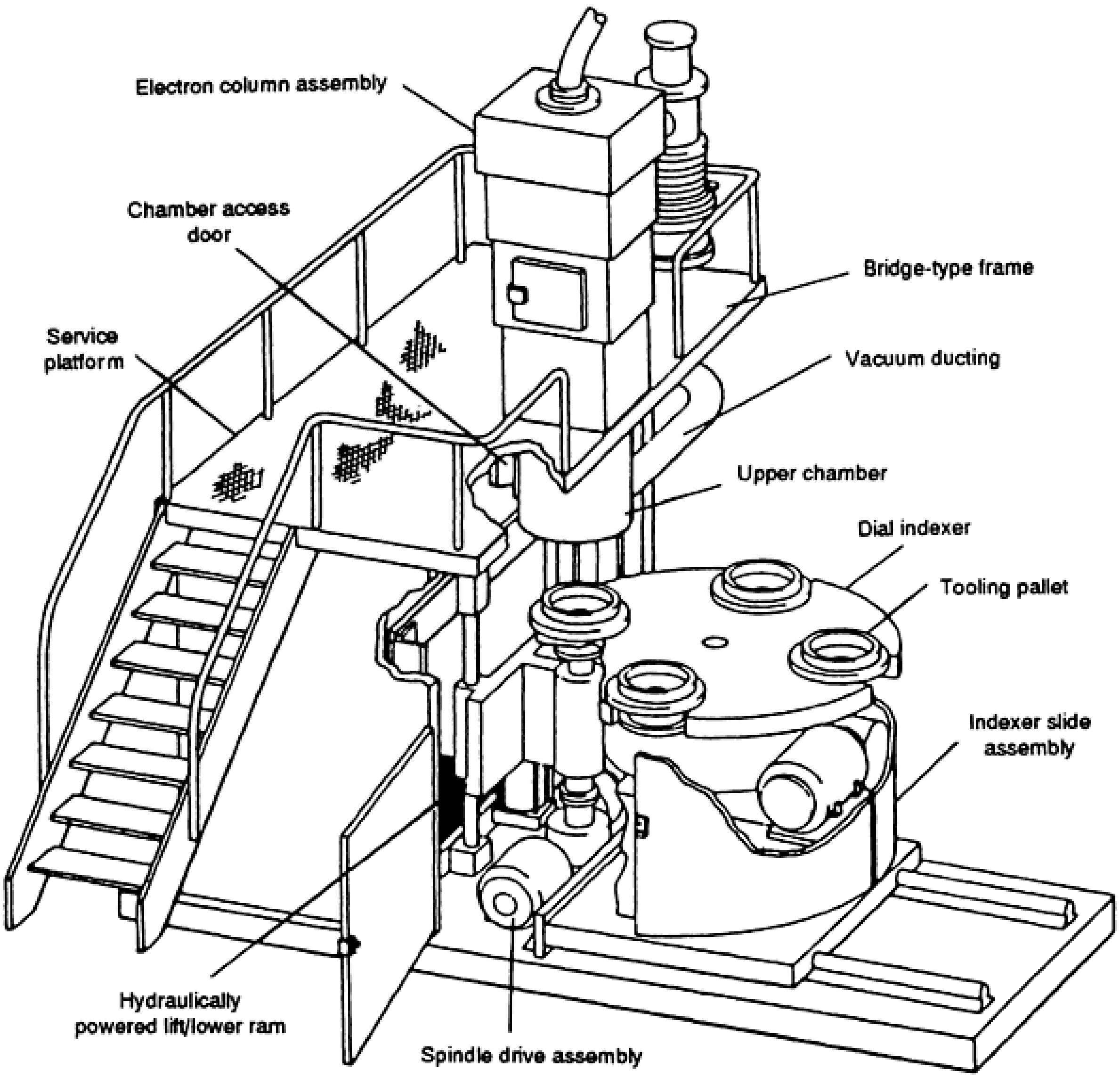
Electron column assembly
Chamber access door
Service platform
Bridge-type frame
Vacuum ducting
Upper chamber
Dial indexer
Tooling pallet
Indexer slide assembly
Hydraulically powered lift/lower ram
Spindle drive assembly

FIG. 7 SCHEMATIC SHOWING A CUTAWAY VIEW OF A DIAL TABLE DEVICE INCORPORATED INTO AN EBW-(MV/PV) SYSTEM TO PROVIDE INDEXING OF WORK STATION INTO WELDING AREA

Figure 8 shows a custom partial-vacuum system, typical of the variety used for high-production automotive parts manufacture. Installations of this type currently are available with power ratings of 100 kW (100 kV, 1 A) for welding heavy cross sections from 50 to 205 mm (2 to 8 in.) thick.

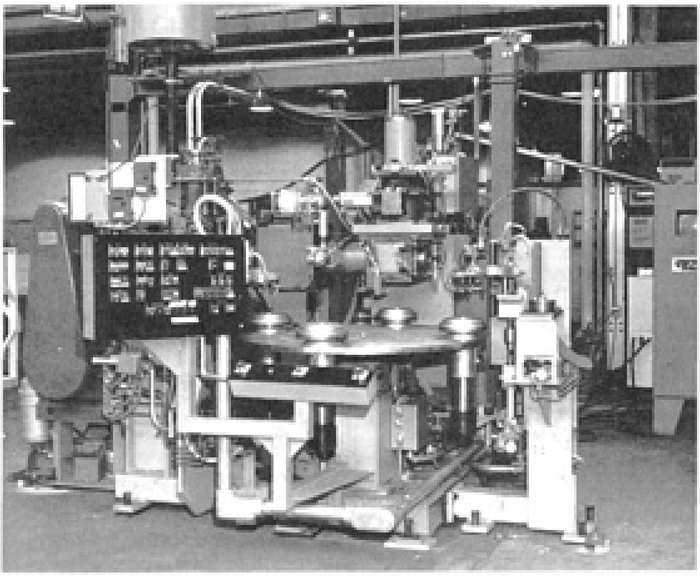

FIG. 8 CUSTOM-DESIGNED EBW-MW SYSTEM USED TO PRODUCE AUTOMOTIVE PARTS. COURTESY OF SCIAKY BROS., INC.

Some smaller units are also available that feature dual-mode operation; they operate in either the high- or medium-vacuum mode, depending on the application. Figure 9 shows a dual-mode portable machine rated at 6 kW (60 kV, 100 mA) with fully automatic and preset pumping and welding cycles. The small weld chamber can be pumped down to 65 mPa (5×10^{-4} torr) in 10 s.

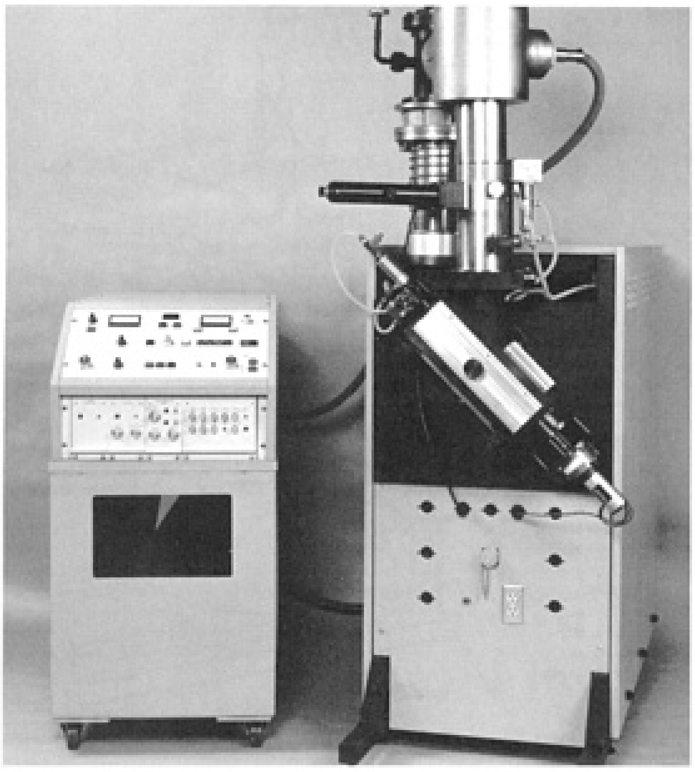

FIG. 9 STANDARD DUAL-MODE PORTABLE EBW SYSTEM. COURTESY OF EBTEC CORP.

The system shown in Fig. 10 is an example of a nonvacuum unit featuring a standard dial index table. Parts are loaded into a fixture nest in the table and are indexed through a small, radiation-tight enclosure in which parts are welded at 100 kPa (1 atm). This nonvacuum unit has power ratings of 17.5 kW (174 kV, 100 mA) or 35 kW (175 kV, 200 mA) and can be used either as a standard system or as a custom-designed special task system.

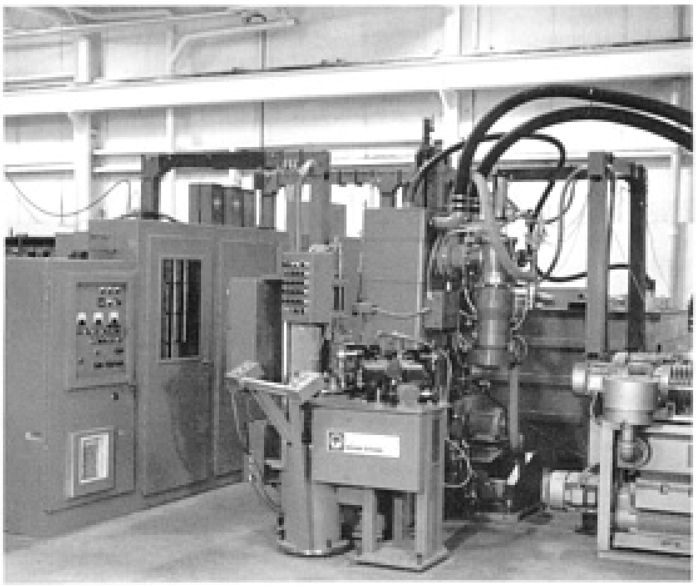

FIG. 10 STANDARD DIAL-INDEX TABLE-TYPE EBW-NV SYSTEM. COURTESY OF LEYBOLD-HERAEUS VACUUM SYSTEMS INC.

Computer control systems available in many EBW units. Control of welding operations can be expanded to include joint locators, five-axis programmable contouring, and several digitizing provisions that can be actuated on the basis of information generated by the seam locator. Many EBW systems can also be adapted for electron-beam heat-treating applications.

Reference cited in this section

1. H.B. CARY, *MODERN WELDING TECHNOLOGY,* 2ND ED., PRENTICE HALL, 1989, P 256

Safety

Protection must be provided by equipment design and arrangement, and by safety precautions in EBW and related operations, against the usual hazards of welding and the special hazards of exposure to (1) the high voltages involved in generating the electron beam, (2) the beam itself (direct viewing of intense radiation emitted by molten weld metal can be harmful to eyesight and thus the beam operation should be viewed only through a filter lens commonly used for arc welding), and (3) radiation of x-rays produced by impingement of the beam on the work or other materials. From a safety standpoint, an accelerating voltage of less than 20 kV produces soft x-rays, while an accelerating voltage of over 20 kV produces hard x-rays (Ref 1). Shielding is more demanding against radiation as the acceleration voltage increases. Suitable precautionary measures are described in AWS F2.1-78, "Recommended Safe Practices for Electron Beam Welding and Cutting," and in AWS/ANSI Z49.1, "Safety in Welding and Cutting."

Reference cited in this section

1. H.B. CARY, *MODERN WELDING TECHNOLOGY,* 2ND ED., PRENTICE HALL, 1989, P 256

Electron-Beam Welding

References

1. H.B. CARY, *MODERN WELDING TECHNOLOGY,* 2ND ED., PRENTICE HALL, 1989, P 256
2. J.F. LOWRY, J.H. FINK, AND B.W. SCHUMACHER, A MAJOR ADVANCE IN HIGH-POWER ELECTRON BEAM WELDING IN AIR, *J. APPL. PHYS.*, VOL 47, 1976, P 95-106

Laser-Beam Welding

Jyoti Mazumder, University of Illinois at Urbana-Champaign

Introduction

LASER-BEAM WELDING (LBW) uses a moving high-density (10^5 to 10^7 W/cm^2, or 6×10^5 to 6×10^7 W/in.^{2}) coherent optical energy source called a laser as the source of heat. "Laser" is an acronym for "light amplification by stimulated emission of radiation." The coherent nature of the laser beam allows it to be focused to a small spot, leading to high energy densities.

Lasers have been promoted as potentially useful welding tools for a variety of applications. Until the 1970s, however, laser welding had been restricted to relatively thin materials and low speeds because of the limited continuous power available. By 1965, a variety of laser systems had been developed for making microwelds in electronic circuit boards, inside vacuum tubes, and in other specialized applications where conventional technology was unable to provide reliable joining. The availability of high-power continuous-wave (CW) carbon dioxide (CO_2) and neodymium-doped yttrium aluminum garnet (Nd:YAG) lasers and the limitations of current welding technology have promoted interest in deep-penetration welding in the past 20 years using these devices.

The ability of the laser to generate a power density greater than 10^6 W/cm^2 (6×10^6 W/in.^{2}) is a primary factor in establishing its potential for welding (Table 1). Numerous experiments have shown that the laser permits precision (that is, high quality) weld joints rivaled only by those made with an electron beam.

TABLE 1 ENERGY CONSUMPTION AND EFFICIENCY OF LBW RELATIVE TO OTHER SELECTED WELDING PROCESSES

WELDING PROCESS	INTENSITY OF ENERGY SOURCE		JOINING EFFICIENCY, MM^2/KJ	FUSION ZONE PROFILE
	w/cm^2	w/in.^{3}		
OXYACETYLENE (OAW)	10^2-10^3	6×10^2-6×10^3	0.2-0.5	SHALLOW FOR SINGLE PASS
ARC WELDING	5×10^2-10^4	3×10^3-6×10^4	0.8-2$^{(A)}$ 2-3$^{(B)}$ 4-10$^{(C)}$	SHALLOW FOR SINGLE PASS
PLASMA ARC (PAW)	10^3-10^6	6×10^3-6×10^6	5-10	SHALLOW AT LOW-ENERGY END DEEP PENETRATION AT HIGH-ENERGY END
LASER BEAM	10^5-10^7	6×10^5-6×10^7	15-25	SHALLOW AT LOW-ENERGY DENSITY RANGE DEEP PENETRATION AT HIGH-ENERGY DENSITY RANGE
ELECTRON BEAM	10^5-10^8	6×10^5-6	20-30	DEEP PENETRATION

		$\times 10^{8}$		

(A) GAS-TUNGSTEN ARC WELDING (GTAW).
(B) GAS-METAL ARC WELDING (GMAW).
(C) SUBMERGED ARC WELDING (SAW)

Laser-Beam Welding

Jyoti Mazumder, University of Illinois at Urbana-Champaign

Advantages and Limitations of LBW

Laser welding offers the following advantages (Ref 1, 2):

- LIGHT IS INERTIALESS (HENCE, HIGH PROCESSING SPEEDS WITH VERY RAPID STOPPING AND STARTING BECOME POSSIBLE).
- FOCUSED LASER LIGHT PROVIDES HIGH ENERGY DENSITY.
- LASER WELDING CAN BE USED AT ROOM ATMOSPHERE.
- DIFFICULT-TO-WELD MATERIALS (FOR EXAMPLE, TITANIUM, QUARTZ, ETC.) CAN BE JOINED.
- WORKPIECES DO NOT NEED TO BE RIGIDLY HELD.
- NO ELECTRODE OR FILLER MATERIALS ARE REQUIRED.
- NARROW WELDS CAN BE MADE.
- PRECISE WELDS (RELATIVE TO POSITION, DIAMETER, AND PENETRATION) CAN BE OBTAINED.
- WELDS WITH LITTLE OR NO CONTAMINATION CAN BE PRODUCED.
- THE HEAT-AFFECTED ZONE (HAZ) ADJACENT TO THE WELD IS VERY NARROW.
- INTRICATE SHAPES CAN BE CUT OR WELDED AT HIGH SPEED USING AUTOMATICALLY CONTROLLED LIGHT DEFLECTION TECHNIQUES.
- THE LASER BEAM CAN ALSO BE TIME SHARED.

Precise part fit-up and alignment are much more critical in laser welding than in ordinary arc welding. The typical focal spot diameter for a laser beam ranges from 100 to 1000 μm (0.004 to 0.040 in.). In addition, capital cost for laser welding devices is almost 10 times more expensive than comparable power arc welding systems. On the other hand, laser welding can provide much higher throughput relative to conventional arc welding. When the capital cost of laser-beam welding is compared to electron-beam welding (EBW), laser-beam welding becomes the more cost-effective of the two processes because no vacuum enclosure is necessary for laser-beam welding.

The penetration depth obtained in laser welding is less than that observed in electron-beam welding. The maximum thickness of type 304 stainless steel plate that can be welded using a 77 kW (105 hp) CO_2 laser is 50 mm (2 in.) (Ref 3), whereas EBW can produce welds in type 304 stainless steel up to several inches in thickness. However, the penetration depth of EBW extends only a relatively short distance under atmospheric pressure. Welding under a vacuum is required to obtain optimum efficiency in EBW. A laser beam, however, can be transmitted an appreciable distance through the atmosphere without serious attenuation or optical degradation because of its coherent nature. Laser-beam welding thus offers an easily maneuverable, chemically clean, high intensity, atmospheric welding process with narrow HAZ and subsequent low distortion. Peak penetration, $p_{\max}$, for LBW is defined by:

$$P_{\mathrm{MAX}} \propto P^{0.7} \qquad \textbf{(EQ 1)}$$

where P is the power (in watts). In terms of weld width (w) and depth (d), both conduction-mode welding (w/d [ges] 1) and deep-penetration welding ($w/d < 1$) can be obtained with lasers.

References cited in this section

1. I.J. SPALDING, LASER SYSTEM DEVELOPMENTS, *PHYS. BULL.*, JULY 1971, P 402
2. Y.S. ARATA AND J. MIYAMOTO, *TECHNOCRAT,* VOL 11 (NO. 5), 1978, P 33
3. C.M. BANAS, "LASER WELDING TO 100 KW," REPORT NO. R76-912260-2, UNITED TECHNOLOGIES, FEB 1977

Laser-Beam Welding

Jyoti Mazumder, University of Illinois at Urbana-Champaign

Process Applications

A large variety of metals and alloys have been welded with pulsed or CW lasers. Until the first reported demonstration of deep-penetration laser-beam welding by Brown and Banas (Ref 4), most laser welds were made with pulsed lasers.

The localized heating obtained with laser sources was soon realized to be an important advantage. Anderson and Jackson (Ref 5) reported an interesting comparison between heating effects produced with a conventional arc source and those occurring with a pulsed laser for an iron substrate. They show that not only is the laser-produced HAZ small, but also that the laser source is more efficient because it requires only 37.8 J/cm^2 (0.229 Btu/in.^{2}) to produce melting to the required depth, while the arc source must deliver 246 J/cm^2 (1.49 Btu/in.^{2}) to the workpiece. Localized heating makes the laser ideal for welding electronic components on a printed circuit board (PCB) where high average temperatures in even small volumes surrounding the weld region cannot be tolerated (Ref 6).

Microwelding with Pulsed Lasers. Microwelding with pulsed ruby lasers and pulsed Nd:YAG lasers is credited with some of the earliest successful applications of laser welding, including laser welding of thermocouple gages in the Apollo lunar probes in the late 1960s (Ref 7). Table 2 shows the various lasers available for welding and the maximum thicknesses that can be welded with these lasers.

TABLE 2 PARAMETERS FOR SELECTED PULSED AND CONTINUOUS WAVE LASERS USED FOR LBW APPLICATIONS

LASER	PULSE LENGTH, MS	PULSE ENERGY, J	PEAK POWER, KW	MAXIMUM WELD THICKNESS$^{(A)}$		WELDING SPEED	
				mm	in.	mm/s	in./min
PULSED							
RUBY	3-10	20-50	1-5	0.13-0.50	0.005-0.020	1.2	3.0
ND:GLASS	3-10	20-50	1-5	0.13-0.50	0.005-0.020	0.63	1.5
ND:YAG	3-10	10-100	1-10	0.13-0.60	0.005-0.025	2.1	5.0
CO$_2$	5-20	0.1-10	1-5	0.13	0.005	1.2	3.0
CONTINUOUS WAVE							
ND:YAG	. . .	. . .	1.8	5.56	0.022	5.8	14.0
CO$_2$ (DC EXCITED)	. . .	. . .	1	0.60	0.025	12.7	30.0
CO$_2$ GAS DYNAMIC	. . .	. . .	20	19.0	0.750	21.2	50.0
CO$_2$ GAS DYNAMIC	. . .	. . .	77	50.8	2.00	26.7	63.0

CO_2 (RF EXCITED)(B)	. . .	. . .	5	10.0	0.4	10.8	25.6

Source: Ref 3, 8, 9, 10

(A) DATA ARE FOR TYPE 304 STAINLESS STEEL.

(B) DATA PROVIDED BY TRUMPF INDUSTRIAL LASER.

Microwelding with pulsed solid-state lasers is now used to:

- SEAL BATTERIES FOR DIGITAL WATCHES AND PACEMAKERS
- SEAL INK CARTRIDGES FOR FOUNTAIN PENS
- WELD WIRES TO TERMINALS ON SHUNT PLATES
- WELD TELEPHONE SWITCHING DEVICES
- ATTACH BATTERY TERMINALS IN DISK CAMERAS
- JOIN LAMP TERMINAL FEED-THROUGHS
- ASSEMBLE GYROSCOPE BEAMING DEVICES
- CONSTRUCT TANTALUM DIAPHRAGMS
- WELD REPAIR INCONEL 718 CASTINGS

Laser microwelding has found its way into almost all types of manufacturing industries including aerospace, automotive, electronics, and consumer durables (Ref 11). Figure 1 shows the approximate penetration possible for high-quality welds using Nd:YAG lasers. It should be noted that for the same average power, pulsed Nd:YAG lasers provide much higher penetration levels than continuous-wave Nd:YAG lasers.

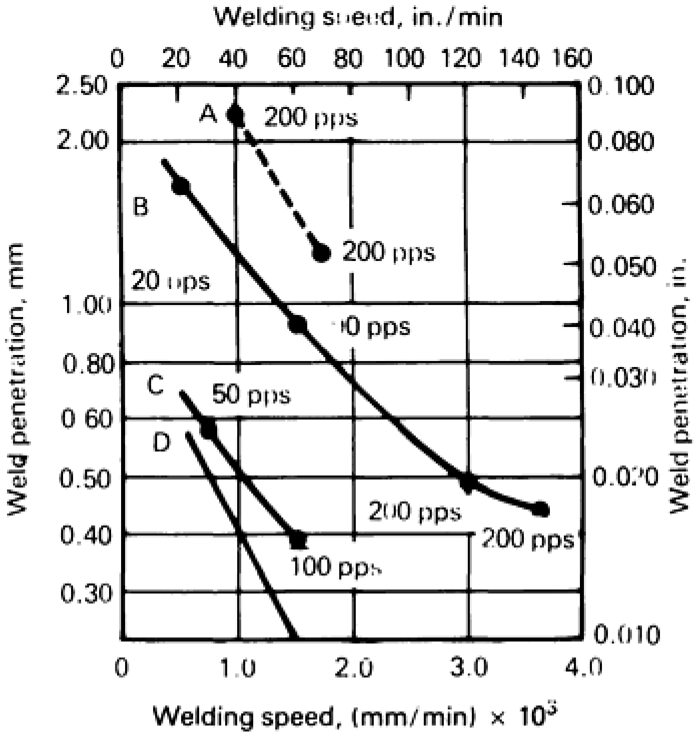

CURVE	WELDING MODE	AVERAGE OUTPUT, W	
		PULSED	CONTINUOUS WAVE

A	PENETRATION	400	. . .
B	CONDUCTION	400	. . .
C	CONDUCTION	150	. . .
D	CONDUCTION	. . .	400

FIG. 1 PLOT OF WELD PENETRATION VERSUS WELDING SPEED AS A FUNCTION OF WELDING MODE, MODE OF OPERATION, AND AVERAGE OUTPUT FOR ND:YAG LASERS AT A WAVELENGTH OF 1.06 μM (41.7 μIN.). SOURCE: REF 12

Penetration Welding. The scope of technically and commercially feasible LBW applications has increased greatly since the demonstration of penetration welding using a multikilowatt CW CO_2 laser by Brown and Banas (Ref 4). Numerous experiments have shown that the laser produces precision weld joints of high quality in ferrous alloys, nickel alloys, and titanium alloys that is rivaled only by those made by EBW. Data describing the plate thickness of different laser-welded materials are given in Table 3.

TABLE 3 PARAMETERS FOR HIGH-POWER LBW RELATIVE TO ALLOY TYPE FOR BEAD-ON-PLATE JOINTS

MATERIAL	TYPE OF JOINT	THICKNESS		LASER POWER, KW
		mm	in.	
SHIP STEEL, GRADES A, B, C	BUTT	28.6	1.125	12.8
	BUTT	25.4	1.0	12.0
	BUTT	19.0	0.75	12.0
	BUTT	15.9	0.625	12.0
	BUTT	12.7	0.50	12.0
	BUTT	9.5	0.375	10.8
	TEE	9.5-12.7	0.375 TO 0.5	11.9
LOW-ALLOY CARBON STEEL	TEE	9.5-12.7	0.375 TO 0.5	7.5
AISI 1010	BEAM PLATE	4	0.157	1.8(A)
AISI 4130	BUTT	15.2	0.60	14
LOW-ALLOY HIGH-STRENGTH STEEL, 300M	BUTT	19.0	0.75	14
ARCTIC PIPELINE STEEL X-80	BUTT	13.2	0.52	12
D-6AC STEEL	BUTT	6.4	0.25	15
HY-80	. . .	12.5	0.49	10.6
HY-130 STEEL	BUTT	6.4	0.25	5.5
HY-180 STEEL (HP9-4-20)	BUTT	1.6	0.062	5.5
	. . .	1.6	0.062	10.5
	. . .	16.3	0.64	5.5
NICKEL-BASE ALLOY, INCONEL 718	BUTT	14.5	0.57	14
STAINLESS STEEL				
AISI TYPE 304	. . .	6	0.24	17
	. . .	12.5	0.49	17
	. . .	17	0.67	17
	. . .	16.5	0.65	17
	. . .	3.8	0.15	11.5
	. . .	5.6	0.22	11.5(B)

				1.8(C)
	. . .	8.9	0.35	8
	. . .	20.3	0.8	20
	. . .	12.7	0.5	20
	. . .	50.8	2.0	77
AISI TYPE 321	BUTT	14.5	0.57	14
ALUMINUM ALLOY:				
2219	BUTT	12.7	0.50	13
2219	BUTT	6.4	0.25	5
2219	BUTT	6.4	0.25	16
5083	BUTT	6.4	0.25	7
5456	BUTT	3.2	0.125	5.5
5456	BUTT	12.5	0.187	8
5456	BUTT	12.5	0.49	. . .
TITANIUM ALLOY, TI-6AL-4V	BUTT	15.2	0.60	13.5
	. . .	3.2, 6.4	0.125, 0.25	5.5
	. . .	3	0.12	4.7
	. . .	12.5	0.49	11

Source: Ref 10, 13

(A) ND:YAG LASER.

(B) CO_2 LASER.

(C) CONTINUOUS WAVE ND:YAG LASER.

Aluminum and its alloys have recently generated lots of interest in the laser welding community because of their potential application in the auto industry for their weight reduction properties (Ref 14, 15, 16, 17, 18, 19). During the 1970s and 1980s, aluminum alloys were reported to be rather difficult to weld due to the high initial surface reflections of 10.6 μm (417 μin.) radiation emitted by CO_2 lasers (Ref 20, 21, 22, 23, 24, 25, 26, 27). Surface coating often reduces this problem. Low melting point and relevant fluidity sometimes lead to "drop through" problems, and thus energy input is a critical parameter for laser welding. Another critical factor involving laser welding is the vaporization of alloying elements such as magnesium. Suppression of the plasma with a properly designed shielded gas nozzle can be used to avoid the loss of alloying elements (Ref 25). The other alternative is to use filler metal (Ref 17). Although the welding of aluminum is difficult with LBW, it has already been demonstrated that, with proper precautions, welds having tensile strength equivalent to the parent materials can be obtained.

Steels. A great deal of laser welding research has been conducted on ferrous alloys. Most of the initial parametric studies were carried out on stainless steel (Ref 4, 27, 28). Stainless steel was investigated (Ref 28, 29) because of its importance in the power plant and chemical industries. Laser welding of rimmed steel sheet was evaluated for automotive applications by Baardsen *et al.* (Ref 30). High-speed LBW of tin plate and tin-free steel for container industries was reported by Mazumder and Steen (Ref 31). Results in HY-alloy steels were evaluated and reported by Banas (Ref 32) and by Metzbower and Moon (Ref 21). Laser-beam welding of modified 4340 alloy was reported by Seaman and Hella (Ref 33). Several other ferrous alloys that have been studied for the applicability of laser welding include X-80 Arctic pipeline steel (Ref 34), tanker construction steels (Ref 35), D-6ac low-alloy ultrahigh-strength steel (Ref 36), and high-strength low-alloy grade steels (Ref 36).

One of the earliest applications of LBW was for the welding of planetary gear assemblies for the automotive industry. Welding was carried out at 10 m/min (400 in./min) using a 5 kW (7 hp) laser. Subsequently, LBW was applied to produce automobile powertrain components. A most recent mass production application is in the production of the "body-in-white" tailored blank shown in Fig. 2. An LBW process developed by Fraunhofer Institute at Aachen that uses a polarized beam was applied by Krupps Steel to produce steel pipe (Ref 38).

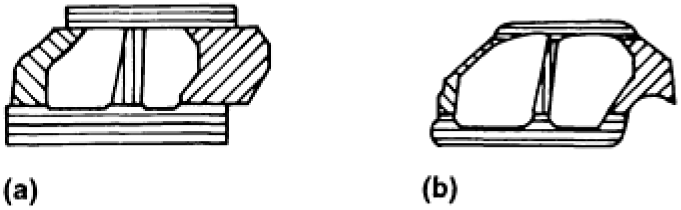

FIG. 2 SEQUENCE OF OPERATIONS REQUIRED TO PRODUCE "BODY-IN-WHITE" TAILORED BLANK AUTOMOTIVE BODIES USING LASER-BEAM WELDING. (A) ARRANGEMENT AND WELDING OF DIVIDED-TYPE BODY PANELS BEFORE FORMING. (B) TAILORED BLANK OBTAINED AFTER STRUCTURAL COMPONENTS HAVE BEEN FORMED IN A PRESS. SOURCE: REF 37

Titanium and Titanium Alloys. Until recently, the EBW technique has been the most popular method used to weld Ti-6Al-4V, which is the most widely used titanium material in the aerospace industry due to its remarkable strength-to-weight ratio. However, the deep penetration of EBW can be obtained only up to a short distance under nonvacuum conditions. For optimum efficiency, electron beam welding must be carried out in an evacuated chamber (Ref 24). In contrast, CO_2 laser beams can be transmitted for appreciable distances through the atmosphere without serious attenuation or optical degradation.

The application of the laser technique to a metal such as a titanium alloy, which is normally difficult to weld, is not only of direct interest to the aerospace and chemical industries, but also is of interest to those who need to weld chemically sensitive metals with complex temperature-dependent structures. The need for better joining methods for titanium and its alloys resulted in several investigations of LBW techniques over various power ranges (maximum 16 kW, or 21 hp). A detailed review is available in Ref 13. Welding speeds in excess of 15 m/min (49 ft/min) were used to weld 1 mm (0.04 in.) thick Ti-6Al-4V with a 4.7 kW (6.3 hp) CO_2 laser, and materials up to 12.5 mm (0.492 in.) thick were joined using an 11 kW (15 hp) CO_2 laser.

References cited in this section

3. C.M. BANAS, "LASER WELDING TO 100 KW," REPORT NO. R76-912260-2, UNITED TECHNOLOGIES, FEB 1977
4. C.O. BROWN AND C.M. BANAS, "DEEP PENETRATION LASER WELDING," AWS ANNUAL MEETING (SAN FRANCISCO), APRIL 1971
5. J.E. ANDERSON AND J.E. JACKSON, *WELD. J.,* VOL 44, 1965, P 1018
6. W.W. DULEY, *CO_2 LASERS: EFFECTS AND APPLICATIONS,* ACADEMIC PRESS, 1976
7. A.J. MOOREHEAD AND P.N. TURNER, *WELD. J.,* VOL 49, 1970, P 15
8. M.A. ACHAREKAR, LASERS, *WELD. ENG.,* DEC 1974, P 9-11
9. T. WEBER, HOBART LASER PRODUCTS, INC., LIVERMORE, CA, 1992, PRIVATE COMMUNICATION
10. M.M. SCHWARTZ, *METAL JOINING MANUAL,* MCGRAW-HILL, 1979
11. S.R. BOLIN, ND-YAG LASER APPLICATION SURVEY, CHAPTER 8, *LASER MATERIALS PROCESSING,* M. BASS, ED., NORTH HOLLAND PUBLISHING CO., AMSTERDAM, 1983
12. H.L. MARSHALL, *PROC. OF THE SPIE,* VOL 86, 1976
13. J. MAZUMDER, LASER WELDING, CHAPTER 3, *LASER MATERIALS PROCESSING,* M. BASS, ED., NORTH HOLLAND PUBLISHING CO., AMSTERDAM, 1983
14. L. RAPP, C. GLUMANN, F. DOUSINGER, AND H. HÜGEL, *LASER TREATMENT OF MATERIALS,* B.L. MORDIKE, ED., DGM INFORMATIONSGESELLSCHAFT MBH, GERMANY, 1992, P 99-104
15. A.S. BRANDSON AND T. ENDERS, *LASER TREATMENT OF MATERIALS,* B.L. MORDIKE, ED., DGM INFORMATIONSGESELLSCHAFT MBH, GERMANY, 1992, P 117-124

16. H. SAKAMOTO, K. SHIBATA, AND F. DAUSINGER, *LASER TREATMENT OF MATERIALS,* B.L. MORDIKE, ED., DGM INFORMATIONSGESELLSCHAFT MBH, GERMANY, 1992, P 125-130
17. J. BERKMANN, S.K. BEHLER, AND E. BEYER, *LASER TREATMENT OF MATERIALS,* B.L. MORDIKE, ED., DGM INFORMATIONSGESELLSCHAFT MBH, GERMANY, 1992, P 151-156
18. A. LANG AND H.W. BERGMAN, *LASER TREATMENT OF MATERIALS,* B.L. MORDIKE, ED., DGM INFORMATIONSGESELLSCHAFT MBH, GERMANY, 1992, P 163-170
19. L.C. MALLORY, R.F. ORR, AND W. WELLS, *LASER MATERIALS PROCESSING III,* J. MAZUMDER AND K. MUKHERJEE, ED., MINERALS, METALS, AND MATERIALS SOCIETY, 1989, P 123-134
20. R.C. CRAFER, *WELD. INST. RES. BULL.,* APRIL 1976, P 17
21. E.A. METZBOWER AND D.W. MOON, MECHANICAL PROPERTIES, FRACTURE TOUGHNESS, AND MICROSTRUCTURES OF LASER WELDS OF HIGH STRENGTH ALLOYS, *PROC. CONF. ON APPLICATIONS OF LASERS IN MATERIALS PROCESSING* (WASHINGTON, DC), 18-20 APRIL 1979
22. D.B. SNOW AND E.M. BREINAN, "EVALUATION OF BASIC WELDING CAPABILITIES," UNITED TECHNOLOGIES RESEARCH CENTER, JULY 1978
23. E.M. BREINAN, C.M. BANAS, AND M.A. GREENFIELD, LASER WELDING--THE PRESENT STATE OF THE ART, DOC IV-181-75, *11TH ANNUAL MEETING OF THE WELDING INSTITUTE* (TEL AVIV), 6-12 JULY 1975, P 1-53
24. M.M. SCHWARTZ, *METAL JOINING MANUAL,* MCGRAW-HILL, 1979
25. A. BLAKE AND J. MAZUMDER, *ASME J. ENG. IND.,* VOL 107 (NO. 3), AUG 1985, P 275-280
26. D.W. MOON AND E. METZBOWER, *WELD. J.,* 1983, P 535-585
27. E.V. LOCKE AND R.A. HELLA, *IEEE J. QUANTUM ELECTRONICS,* VOL QE-10 (NO. 2), FEB 1974, P 179-185
28. R.C. CRAFER, ADVANCES IN WELDING PROCESSES, PAPER NO. 46, *PROC. 4TH LNT. CONF.* (HARROGATE, YORKS, ENGLAND), 9-11 MAY 1978, P 267-278
29. R.A. WILLGOSS, J.H.P.C. MEGAW, AND J.N. CLARK, *WELD. MET. FABR.,* MARCH 1979, P 117-126
30. E.L. BEARDSON, D.J. SCHMATZ, AND R.E. BISARO, *WELD. J.,* VOL 52, APRIL 1973, P 227-229
31. J. MAZUMDER AND W.M. STEEN, LASER WELDING OF STEELS USED IN CAN MAKING, *WELD. J.,* VOL 60 (NO. 6), JUNE 1981, P 19-25
32. C.M. BANAS, "LASER WELDING DEVELOPMENTS," PROC. CEGB INT. CONF. ON WELDING RES. RELATED TO POWER PLANTS (SOUTHAMPTON, ENGLAND), 17-21 SEPT 1972
33. F.D. SEAMAN AND R.A. HELLA, "ESTABLISHMENT OF A CONTINUOUS WAVE LASER WELDING PROCESS," IR-809-3 (1-10), REPORT IN CONFERENCE PROCEEDINGS AFML CONTRACT F336 15-73-C5004, OCT 1976
34. E.M. BREINAN AND C.M. BANAS, PRELIMINARY EVALUATION OF LASER WELDING OF X-80 ARCTIC PIPELINE STEEL, *WRC BULL.,* DEC 1971, P 201
35. C.M. BANAS AND G.T. PETERS, "STUDY OF THE FEASIBILITY OF LASER WELDING IN MERCHANT SHIP CONSTRUCTION," CONTRACT NO. 2-36214, U.S. DEPT. OF COMMERCE, FINAL REPORT TO BETHLEHEM STEEL CORP., AUG 1974
36. M. YESSIK AND D.J. SCHMATZ, "LASER PROCESSING IN THE AUTOMOTIVE INDUSTRY," PAPER MR74-962, SME, 1974
37. F.A. DIPIETRO, *LASER SYSTEM APPLICATION IN INDUSTRY,* ATA-ASSOCIAZIONE TECNIER DELL' AUTOMOBILE, TORINO, ITALY, 1990, P 103-120
38. E. BEYER, FRAUNHOFER INSTITUTE, AACHEN, GERMANY, PRIVATE COMMUNICATION, 1992

Laser-Beam Welding

Jyoti Mazumder, University of Illinois at Urbana-Champaign

Fundamentals of Laser Welding

Lasers are capable of both conduction-mode welding and deep-penetration welding.

Conduction-Mode Welding. Momentum transfer or convection dominates conduction-mode welding. High-power density (10^5 to 10^7 W/cm^2, or 6×10^5 to 6×10^7 W/in.^{2}) in laser welding produces a temperature gradient of the order of 10^6 K/cm (5×10^6 °F/in.) and this in turn leads to surface-tension-driven thermocapillary flow (Marangoni convection) with surface velocities of the order of 1 m/s (Ref 39). Convection is the single most important factor affecting the geometry of the laser melt pool (that is, pool shape, aspect ratio, and surface ripples) and can result in defects such as variable penetration, porosity, and lack of fusion. Convection is also responsible for mixing and therefore affects the composition of the melt pool during laser welds because pool configuration in conduction-mode laser welding is a function of the Prandtl number (kinematic viscosity/molecular diffusivity), Pr_m, of the materials (Ref 40). In materials with low Prandtl numbers (for example, aluminum, with $Pr_m = 0.02$), the pool shape is more spherical and is dominated by conduction heat transfer when compared to high Prandtl number material (for example, steel, with $Pr_m = 0.1$), where pool shape is shallow and wide because it is dominated by the surface-tension-driven flow (Ref 40). Free surface deformation leading to defects such as undercuts is also influenced by convection. A small amount of surface reactant elements (for example, sulfur) can change the convection direction. For most metals, thermocapillary flow drives the hot metal to the cold side (Fig. 3), but the addition of sulfur changes the sign of the temperature dependence function of the surface tension and drives the liquid metal to the center of the pool.

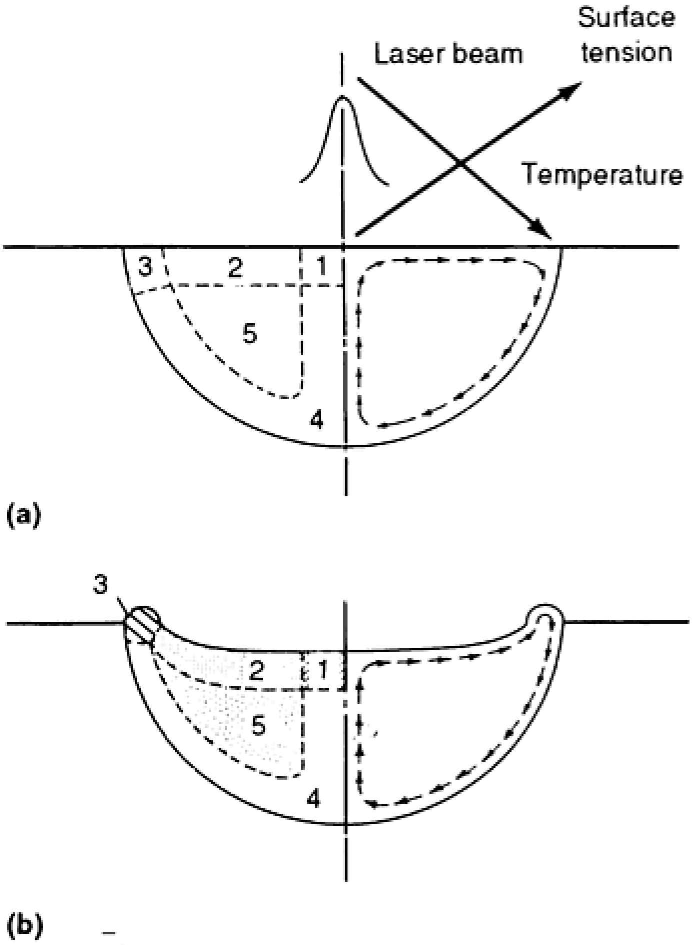

FIG. 3 SCHEMATIC SHOWING EFFECT OF CONVECTION ON LBW MELT POOL CONFIGURATION. (A) SPHERICAL SHAPE WITH FLAT SURFACE TYPICAL OF LOW PR_M MATERIALS. (B) SHALLOW AND UNDERCUT FREE SURFACE CHARACTERISTICS OF HIGH PR_M MATERIALS. NUMBERS IN FIGURE IDENTIFY SPECIFIC REGIONS: (1) STAGNATION FLOW REGION; (2) FREE SURFACE BOUNDARY LAYER REGION; (3) COOLED CORNER REGION; (4) SOLID-LIQUID INTERFACE BOUNDARY LAYER REGION; AND (5) ISOTHERMAL INVISCID CORE. SOURCE: REF 39

Deep-Penetration-Mode Welding. The mechanism of deep-penetration welding by a laser beam is very similar to that encountered with an electron beam (that is, energy transfer is via "keyhole" formation) (Ref 41, 42, 43). This keyhole may be produced when a beam of sufficiently high power density causes vaporization of the substrate and the pressure produced by the vapor in the crater causes displacement of the molten metal upward along the walls of the hole. This hole acts as a blackbody, aiding the absorption of the laser beam as well as distributing the heat deep in the material. The energy in most conventional welding processes is deposited at the surface of the workpiece and brought into the interior by conduction.

The conditions of energy and material flow during beam welding were investigated theoretically by Klemens (Ref 44). According to Klemens, the keyhole or cavity is formed only if the beam has sufficient power density. The keyhole is filled with gas or vapor created by continuous vaporization of wall material by the beam. This cavity is surrounded by liquid that, in turn, is surrounded by solid.

The flow of the liquid and the surface tension tends to obliterate the cavity while the vapor, which is continuously generated, tends to maintain the cavity. There is a continuous flow of material out of the cavity at the point where the beam enters. For a moving beam, this "keyhole" achieves a steady state (that is, the cavity and the beam associated with the molten zone move forward at the speed set by the advance of the beam). The material lost by vaporization shows up as a depression in the solidified melt as porosity, as an inward deformation of the workpiece, or possibly as a combination of these effects. The requirement that sufficient vapor be produced to maintain a steady state leads to a minimum advance speed for a steady state. While the cavity moves through the solid and liquid material at a speed determined by the motion of the beam, material must be moved continuously from the region ahead of the cavity to the region behind it. This is confirmed by the experimental work of Sickman and Morijn (Ref 45, 46) and Arata *et al.* (Ref 47).

An interesting experiment on the mechanism of deep-penetration welding with a CW CO_2 laser has been reported by Sickman and Morijn (Ref 45, 46). It was conducted on transparent fused quartz so that the formation of the weld as a function of time could be photographed. The tip of the hole appears to bend toward the direction in which the workpiece is moving. This process is caused by the reflection of laser light from the leading edge of the hole. Material evaporated at this surface is effectively trapped by the cooler trailing edge. Thus, material is transported across the laser beam from the hot leading edge to the cooler trailing edge without significant ejection of material back out toward the beam. For welding speeds in the range of 10 to 45 mm/min (0.4 to 1.8 in./min), the depth of penetration was linearly related to welding speed. As expected, the minimum penetration was obtained at maximum welding speed.

Fluid flow during penetration welding was also studied by Arata *et al.* (Ref 47), using laser beam irradiation of low-viscosity glass at high temperature. High-speed photography of the phenomenon, taken at 8000 frames per second (fps), clearly shows melt flow and the motion of the cavity. Molten fluid, which forms at the front wall of the cavity, accelerates at an angle along the wall as it is driven by the forces of the turbulence generated by laser vaporization. In this process, a large vortex is formed behind the cavity near the weld surface. This vortex is considered to be the cause of the so-called "wine-glass" beads that are produced by the process.

The transport of material is mainly due to flow in the liquid. However, part of the material is transported in the vapor phase, and this vapor transport generates the excess pressure that drives the liquid flow.

Laser Welding Parameters. The major independent process variables for laser welding include:

- INCIDENT LASER BEAM POWER
- INCIDENT LASER BEAM DIAMETER
- ABSORPTIVITY
- TRAVERSE SPEED OF THE LASER BEAM ACROSS THE SUBSTRATE SURFACE

Parameters such as weld design, shielding gas, gap size for butt welds, and depth of focus with respect to the substrate also play important roles. These parameters are discussed in the article "Procedure Development and Practice Considerations for Laser-Beam Welding" in this Volume.

The dependent variables are considered to be:

- DEPTH OF PENETRATION

The effects of some of the important variables are briefly discussed below.

Laser-Beam Power. The depth of penetration with laser welding is directly related to the power density of the laser beam and is a function of incident beam power and beam diameter. For a constant beam diameter, penetration typically increases as the beam power is increased. Locke *et al.* (Ref 27) and Baardsen *et al.* (Ref 30) report that penetration increases almost linearly with incident laser power. It is generally observed that for LBW of a particular thickness, a minimum threshold power is required.

Laser-Beam Diameter. This parameter is one of the most important because it determines the power density. However, it is very difficult to measure for high-power laser beams. This difficulty is due partly to the nature of the beam diameter and partly to the definition of what is to be measured. A Gaussian beam diameter, d_G, can be defined as the diameter where the power has dropped to $1/e^2$ or $1/e$ of the central value:

$$\frac{1}{e^2} < P_{dG} < \frac{1}{e}$$ **(EQ 2)**

The beam diameter defined on the basis of $1/e^2$ of the central value contains more than 80% of the total power, whereas the power contained for $1/e$ beam definition is slightly over 60% (Ref 49). Therefore, the $1/e^2$ beam diameter is recommended.

Many techniques have been used to measure the beam diameter, but most are unsatisfactory in some respect. Single isotherm contouring techniques (for example, charting paper and drilling acrylic or metal plates) suffer because the particular isotherm they plot depends both on power and on exposure time. These techniques are also highly unlikely to coincide with either the $1/e$ or $1/e^2$ diameters. Multiple isotherm contouring techniques overcome these difficulties but are tedious to interpret. However, several commercial systems are now available (see the article "Procedure Development and Practice Considerations for Laser-Beam Welding" in this Volume).

Absorptivity. The efficiency of LBW depends on the absorption of light energy by the workpiece. Any heat transfer calculation for laser processing is based on the energy absorbed by the workpiece.

The infrared absorption of metals largely depends on the conductive absorption by free electrons. Therefore, absorptivity is a function of the electrical resistivity of the substrate material. Arata and Miyamoto (Ref 50) and McCay *et al.* (Ref 51) measured the absorptivity of polished surfaces of various materials and concluded that absorptivity is proportional to the square root of the electrical resistivity. This agrees closely with:

$$A = 112.2\sqrt{r_r}$$ **(EQ 3)**

where A is the absorptivity and ρ_r is the electrical resistivity. A temperature-dependent relationship between the electrical resistivity and emissivity of the metal was derived by Bramson (Ref 52). Bramson's formula can be used for theoretical calculation of the absorptivity from the electrical resistivity because absorptivity equals emissivity. However, Eq 4 will be valid only for metals without surface oxide layers that are heated in a vacuum. The presence of an oxide layer will increase the absorptivity. The relationship between the emissivity and the electrical resistivity of a substrate for the perpendicular incidence of radiation of long wavelength derived by Bramson is:

$$e_l(T) = 0.365\left[\frac{r_r(T)}{l}\right]^{1/2} - 0.667\left[\frac{r_r(T)}{l}\right] + 0.006\left[\frac{r_r(T)}{l}\right]^{3/2}$$ **(EQ 4)**

where ρ_r (T) is the electrical resistivity at absolute temperature, T, expressed in Ohm-centimeter ($\Omega \cdot$ cm), and $\varepsilon\lambda$ (T) is the emissivity of the substrate at temperature T (in °C) for radiation having a wavelength, λ.

The estimated absorptivity for Ti-6Al-4V at 300 °C (572 °F) using Bramson's formula is approximately 15%. Experimental data published by Arata and Miyamoto (Ref 50) and Bramson (Ref 52) indicate that the absorptivities of aluminum, silver, and copper are between 2 to 3% and those of stainless steel (type 304), iron, and zirconium are below 15% even in the molten state. Reflection losses of this magnitude are significant. Therefore, when a sheet metal product form is welded, measures must be taken to avoid reflection losses. Applying an absorbent powder to the surface or forming an anodized film on the surface are two techniques that are considered to be very effective (Ref 51).

Absorptivity can also be increased by the use of reactive gases. Jørgensen (Ref 53) reported that the addition of 10% O_2 to an argon shielding gas gives an increase of up to 100% in welding depth. Jørgensen also found that gas flow had no significant effect on weld depth. However, an increase in depth was associated with a decrease in reflectivity, which was obtained by the addition of a small amount of oxygen.

Although metals are poor absorbers of infrared energy at room temperature, above a certain threshold (approximately 10^6 to 10^7 W/cm^2, or 6×10^6 to 6×10^7 W/in.^{2}) energy transfer via the keyhole leads to much higher effective absorptivity. Once a keyhole has formed, absorptivity increases rapidly. Multiple internal reflections inside a keyhole are responsible for the deep-penetration welding by the laser despite the large convergence angles for laser beams relative to electron beams (Ref 51). However, the threshold energy required for keyhole formation in laser welding is higher than that required for an electron beam (1.5×10^5 W/cm^2, or 9.5×10^5 W/in.^{2}) because of the poor absorptivity. Nevertheless, energy transfer by this keyhole mechanism permits the laser to provide efficient welding of even highly reflective materials (for example, aluminum) (Ref 25).

Traverse Speed. The correlation of penetration depth relative to welding speed with both LBW and EBW processes was studied by Duley (Ref 41) and Locke and Hella (Ref 27). The penetration in the laser weld is consistently less than that obtained with an electron beam, but the relative difference between the two penetration depths diminishes as the welding speed is increased. However, Duley (Ref 41) found this somewhat surprising because as pointed out by Baardsen *et al.* (Ref 48), the time to form a void or keyhole depends on the illumination time for a particular area on the surface of the workpiece as the welding speed is increased. When this occurs, the average power dissipated in the sheet is expected to drop because the keyhole is no longer a completely effective trap for the incident laser radiation. For an electron beam, absorptivity of the material is independent of the shape and extent of the keyhole. Hence, the total power dissipated in the workpiece is less dependent on the welding speed. However, Crafer (Ref 54) reports that the keyhole penetration threshold for laser beams or electron beams of radius 0.1 mm (0.004 in.) incident on a steel surface of thermal diffusivity 10 mm^2/s (0.4 ft^2/h) is achieved in 1 ms. This could be regarded as instantaneous. Again, for very low welding speeds, the penetration depth of laser welds becomes significantly less than that attainable with the electron beam. According to Locke *et al.* (Ref 28), this can be attributed to the formation of a plasma cloud, which attenuates the incident beam.

References cited in this section

25. A. BLAKE AND J. MAZUMDER, *ASME J. ENG. IND.,* VOL 107 (NO. 3), AUG 1985, P 275-280
27. E.V. LOCKE AND R.A. HELLA, *IEEE J. QUANTUM ELECTRONICS,* VOL QE-10 (NO. 2), FEB 1974, P 179-185
28. R.C. CRAFER, ADVANCES IN WELDING PROCESSES, PAPER NO. 46, *PROC. 4TH LNT. CONF.* (HARROGATE, YORKS, ENGLAND), 9-11 MAY 1978, P 267-278
30. E.L. BEARDSON, D.J. SCHMATZ, AND R.E. BISARO, *WELD. J.,* VOL 52, APRIL 1973, P 227-229
39. J. MAZUMDER, *OPT. ENG.,* VOL 30 (NO. 8), 1991, P 1208-1219
40. C.L. CHAN, J. MAZUMDER, AND M.M. CHAN, *METALL. TRANS.* A, VOL 15, 1982, P 2175
41. W.W. DULEY, *CO_2 LASERS: EFFECTS AND APPLICATIONS,* ACADEMIC PRESS, 1976, P 241
42. A.M. MALEKA, ED., *ELECTRON BEAM WELDING PRINCIPLES AND PRACTICE,* WELDING INSTITUTE, MCGRAW-HILL, 1971, P 95-96
43. D.T. SWIFTHOOK AND E.E.F. GICK, *WELD. J.,* 1973, P 492S-499S
44. P.G. KLEMENS, *J. APPL. PHYSICS,* VOL 47, 1976, P 2165-2174
45. J.G. SICKMAN AND R. MORIJN, *PHILLIPS RES. REP.,* VOL 23, 1968, P 376
46. J.G. SICKMAN AND R. MORIJN, *PHILLIPS RES. REP.,* VOL 23, 1968, P 375
47. Y. ARATA, H. MARUO, I. MIYAMOTO, AND Y. INOUE, "DYNAMIC BEHAVIOR OF LASER

WELDING," IIW DOC IV/222/77, INTERNATIONAL INSTITUTE OF WELDING, 1977
48. E.L. BAARDSEN, D.J. SCHMATZ, AND R.E. BISARO, *WELD. J.,* VOL 52, APRIL 1973, P 227-229
49. J.E. HARRY, *INDUSTRIAL APPLICATION OF LASERS,* MCGRAW-HILL, 1974
50. Y. ARATA AND I. MIYAMOTO, *LASER FOCUS,* VOL 3, 1977
51. M.H. MCCAY, T.D. MCCAY, A. SEDGHINA-SAB, AND D.R. KEEFER, *LASER MATERIALS PROCESSING III,* J. MAZUMDER AND K. MUKHERJEE, ED., MINERALS, METALS, AND MATERIALS SOCIETY, 1989
52. M.A. BRAMSON, *INFRARED RADIATION: A HANDBOOK FOR APPLICATION,* PLENUM PRESS, 1968
53. M. JØRGENSEN, *MET. CONSTR.,* VOL 12 (NO. 2), FEB 1980, P 88
54. R.C. CRAFER, *WELD. INST. RES. BULL.,* VOL 17, FEB 1976

Laser-Beam Welding

Jyoti Mazumder, University of Illinois at Urbana-Champaign

Processing Equipment

Solid-State Optically Pumped Lasers. The majority of pulsed ruby and Nd:YAG solid-state lasers are used for microwelding applications. Continuous wave CO_2 (commercially available up to 25 kW, or 34 hp) and Nd:YAG (commercially available up to 2.4 kW, or 3.2 hp) versions are primarily used for deep-penetration welding. Table 2 gives the capabilities of lasers used for welding. Typical specifications of commercially available Nd:YAG lasers feature:

- UP TO 400 TO 600 W (0.5 TO 0.8 HP) AVERAGE POWER WITH PEAK POWER OF UP TO 8 TO 10 KW (11 TO 13 HP)
- PULSING WIDTH OF 250 MS
- PULSING RATE OF 400 PULSES PER SECOND (PPS)

Nd:YAG rods used as the active medium are pumped by a xenon flash lamp arranged in an elliptical cavity (Fig. 4). Typically, the cavity is coated with gold to maximize reflection of the flash lamp and subsequently to maximize the absorption in the rod. Recently, diode lasers have been explored as the pumping source to replace the flash lamp as the energy source. This improves the pumping efficiency and minimizes the thermal distortion of the medium.

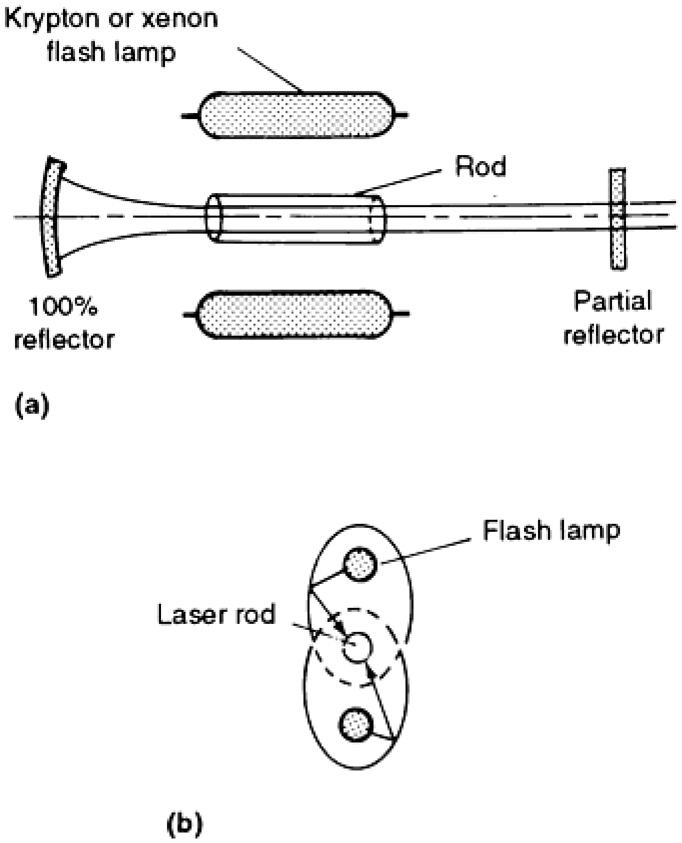

FIG. 4 TYPICAL CONSTRUCTION OF A PULSED OPTICALLY PUMPED ND: YAG SOLID-STATE LASER. (A) TOP VIEW. (B) END VIEW, SHOWING ELLIPTICAL CROSS SECTION OF REFLECTOR CAVITY HAVING THE ROD AT ONE FOCUS OF THE ELLIPSE AND THE FLASH LAMP AT THE OTHER FOCUS

Another new development in the Nd:YAG laser field is the face-pumped Nd:YAG laser (Fig. 5). The crystal slab provides much better beam quality because the temperature fluctuation across the active medium is much less for the slab configuration than the rod configuration. The "rod" geometry is at a disadvantage because only the surface of the rod is used as the active medium to produce a temperature gradient between the center and the surface. In the slab geometry, however, almost the entire crystal serves as the active medium, thus minimizing temperature gradient, which in turn reduces the thermal lensing effect. Although these solidstate optical systems have been used internally by General Electric for a long time, only recently, after the patent expired, did commercial systems become available. Beam generation in most rod systems is pulsed, whereas in most slab systems it is continuous wave.

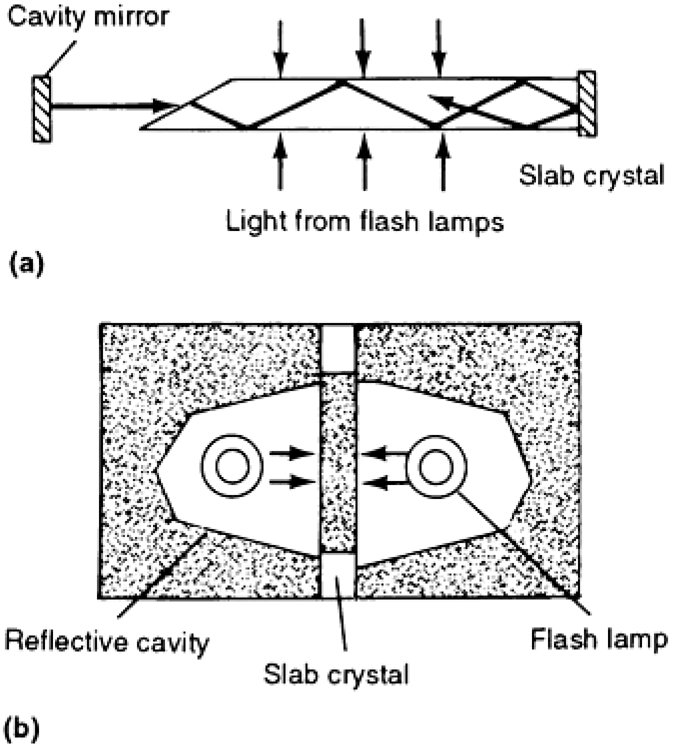

FIG. 5 KEY COMPONENTS OF A FACE-PUMPED ND:YAG SOLID-STATE LASER INCORPORATING A SLAB CRYSTAL. (A) FRONT VIEW. (B) END VIEW (CROSS SECTION)

The emergence of commercial multikilowatt (up to 2.4 kW, or 3.2 hp) CW Nd:YAG lasers is a recent development that has generated considerable interest with the laser industry because the energy can be transmitted through fiber optics. The maneuverability of the beam provided by fiber optics opens many complicated manufacturing operations previously closed to LBW processing.

Gas Lasers. Continuous wave CO_2 lasers are generally chosen for penetration welding applications. Figure 6 shows different types of pumping designs for CO_2 lasers. The primary components of a fast-axial-flow (FAF) CO_2 laser are shown in Fig. 7. Table 4 gives the four major types of CO_2 lasers commercially available.

TABLE 4 TYPICAL OUTPUT POWER OF COMMERCIALLY AVAILABLE CW CO_2 LASERS

DESIGNATION		**ELECTRIC EXCITATION DISCHARGE$^{(A)}$**	**OUTPUT POWER PER CAVITY LENGTH**	
FLOW CONFIGURATION	**TYPE**		**w/m**	**w/ft**
AXIAL	SLOW	DC	60	18
	FAST	DC	600	180
		RF	750	230
TRANSVERSE	. . .	. . .	7000	2100

(A) DC. DIRECT CURRENT; RF, RADIO FREQUENCY

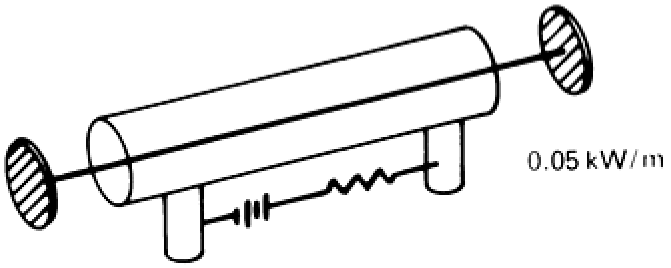

Discharge tube

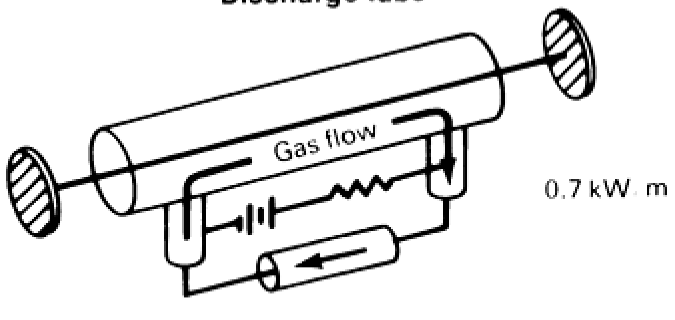
Gas flow
0.7 kW/m

Axial flow

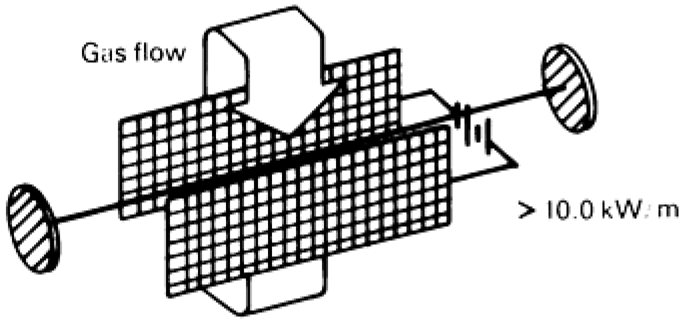
Gas flow
> 10.0 kW/m

Transverse flow

FIG. 6 TYPES OF ELECTRIC EXCITATION DISCHARGE GAS LASERS AND THE PRACTICAL LIMITS OF THE OUTPUT POWER PER CAVITY LENGTH. COURTESY OF TRUMPF LASER

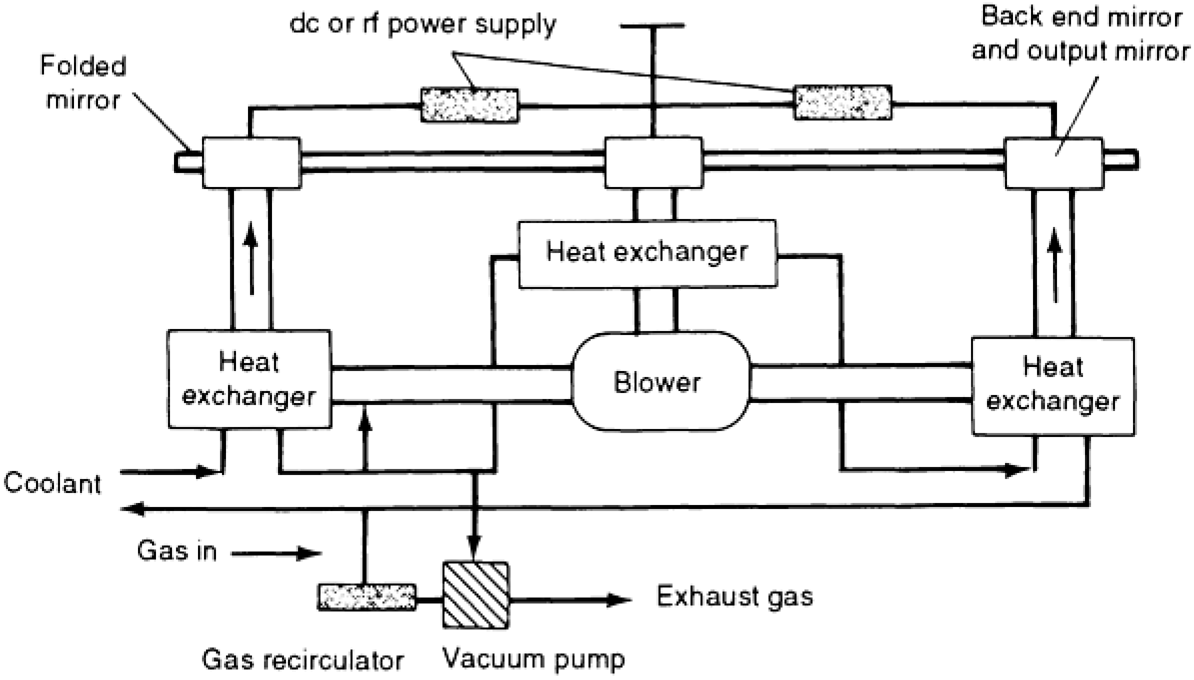

FIG. 7 FLOWCHART SHOWING HEAT EXTRACTION DEVICES USED TO COOL A FAST AXIAL-FLOW LASER. SOURCE: REF 55

Robotics. Robotic equipment is now available that can be integrated with lasers to provide beam delivery to the workpiece. In addition, almost any multiaxis or single-axis workstation can be integrated with lasers to steer the beam to the workpiece.

Ergonomics. Proper use of the laser as a machine tool requires a basic knowledge of optics. Knowledge of optical properties such as beam spatial distribution, depth of focus, and wavelength is critical to LBW processing. Textbooks should be consulted to make operators familiar with the fundamentals of laser physics and optics.

Reference cited in this section

55. W.M. STEEN, *LASER MATERIALS PROCESSING,* SPRINGER-VERLAG, 1991

Laser-Beam Welding

Jyoti Mazumder, University of Illinois at Urbana-Champaign

Health and Safety

Laser-beam welding hazards differ substantively from hazards encountered in other welding techniques. The hazards are not readily apparent, and inexperienced personnel may suffer permanent injury before the existence of hazardous conditions is recognized. For this reason, the American National Standards Institute specification ANSI Z136.1, "Safe Use of Lasers" (latest edition), requires that each facility using lasers designates an individual as "laser safety officer." This individual should be familiar with laser safety and ANSI Z136.1. The officer should monitor the use of lasers to ensure adherence to safe laser practice and the ANSI requirements. Because of the complexity of laser safety, this approach is strongly endorsed. The following brief review of laser safety by itself is not sufficient to ensure personnel safety.

Electrical Hazards. All lasers used for welding employ high voltages capable of inflicting lethal electric shocks. Therefore, maintenance should be performed by personnel familiar with high-voltage safety procedures.

Power supplies for high-power lasers contain capacitors capable of lethal shock even after initial discharge due to a phenomenon known as charge buildup. To facilitate safe maintenance access, the following provisions should be included:

- AN AUTOMATIC DISCHARGE AND GROUNDING CIRCUIT THAT IS ACTUATED WHEN THE LASER IS TURNED OFF
- DISCHARGE AND GROUNDING INTERLOCKS ON ALL ACCESS PANELS
- GROUNDING RODS FOR MANUAL VERIFICATION OF COMPLETE DISCHARGE. SAFETY GLASSES SHOULD BE USED, BECAUSE EXPLOSION-LIKE DISCHARGES ARE POSSIBLE ON PARTLY CHARGED CAPACITORS
- GROUNDING STRAPS TO SHORT OUT CAPACITORS TO PREVENT CHARGE BUILD-UP

In addition, all capacitors should be discharged and grounded before any work is performed on or near high-voltage components.

Control of laser performance usually requires switching of capacitors. Insulated switches that do not expose personnel to electric conductors are preferred. If switching requires work on bus bars, the procedure listed above should be followed, and the use of insulated tools should be considered. Cooling water leaks are not acceptable, especially when electrical and cooling lines share the same umbilical.

Eye Hazards. Any laser beam capable of welding metals is also capable of causing serious eye damage. Personnel exposure to the beam and any specularly reflected beam must therefore be prevented at all times. Certain lasers, however, are also capable of producing diffuse reflections that can cause permanent eye damage. Hence, viewing of the impact area of the laser beam or reflected beams also must be prevented. The preferred method for this is by complete enclosure.

Enclosures can range from a simple sleeve between the laser optic and the part to be welded to fully automatic operations in enclosed rooms. The following are general guidelines:

- THE ENCLOSURE MUST BE OPAQUE TO THE LASER WAVELENGTH. METAL ENCLOSURES ARE GENERALLY SUITABLE, BUT PLASTICS ALSO ARE A POSSIBLE SELECTION. INFRARED LASERS, FOR EXAMPLE, CAN BE ENCLOSED IN CLEAR POLY METHYLMETHACRYLATE. THE REQUIRED THICKNESS CAN BE CALCULATED USING ANSI Z136.1 AND THE TRANSMITTANCE PROPERTIES OF THE MATERIAL.
- THE ENCLOSURE MUST BE INTERLOCKED TO PREVENT FIRING OF THE LASER BEAM WHEN PERSONNEL COULD BE EXPOSED. WHEN THE PIECE TO BE WELDED IS A PART OF THE ENCLOSURE, THE INTERLOCKS SHOULD ALSO PREVENT FIRING, UNLESS THE PIECE IS IN PLACE. IN THE CASE OF PULSED LASERS, BREAKING OF THE INTERLOCKS SHOULD ALSO DISCHARGE THE STORED ENERGY INTO A DUMMY LOAD.
- SIGNS ARE REQUIRED AT ACCESS POINTS TO THE ENCLOSURES; SEE ANSI Z136.1.
- VIEWING OF THE WELD AREA CAN BE ACCOMPLISHED IN SEVERAL WAYS. MOST COMMON ARE VIEWING PORTS WITH FILTERS AND TELEVISION MONITORS. FOR MICROWELDING, MICROSCOPES SHOULD HAVE FILTERED VIEWING OPTICS OR FLIP MIRRORS THAT PERMIT EITHER WELDING OR VIEWING.
- ALIGNMENT OF LASER WELDING SYSTEMS SHOULD BE ACCOMPLISHED USING LOW-POWER LASERS.

Laser welding can also be performed with personnel in attendance. When done in this manner, the hazard must be evaluated for each welding process using the ANSI standard. General requirements include:

- THE LASER WELDING AREA MUST BE COMPLETELY ENCLOSED AND ACCESS MUST BE

RESTRICTED.

- IN MOST CASES, LASER EYE PROTECTION IS REQUIRED FOR ALL PERSONNEL IN THE AREA. THE OPTICAL DENSITY OF THE EYE PROTECTION MUST BE CALCULATED TO REDUCE THE POTENTIAL EYE EXPOSURE TO LESS THAN THE MAXIMUM PERMISSIBLE EXPOSURE (MPE) LEVEL.
- THE WELDING BEAM MUST BE CAREFULLY CONTROLLED AND SHOULD STILL BE ENCLOSED TO THE GREATEST EXTENT POSSIBLE. EYE PROTECTION IS LIKELY TO FAIL ON EXPOSURE TO THE PRIMARY BEAM.
- COMPLIANCE WITH PERSONNEL TRAINING REQUIREMENTS IS EXCEEDINGLY IMPORTANT.

Skin Exposure. Skin exposure to the primary beam obviously can result in thermal burns and must be prevented by complete enclosure and operator training. Even in attended operations where the operator is permitted to view the welding process, provision of partial enclosures to prevent personnel from placing any part of the body into the beam path is desirable.

The ANSI standard also prescribes MPE values for skin exposure. In the visible and near infrared regions, these are much greater than the MPE values for the eyes; hence, a problem from excessive exposure due to diffuse reflections at these wavelengths rarely occurs. Excessive skin exposures in the ultraviolet and far infrared regions are possible, however. Typically, ultraviolet exposures can be controlled with clothing of tightly woven material and barrier creams applied to exposed skin. Harmful levels of ultraviolet light can also be generated by flash lamps; hence, covers should be kept in place.

Chemical Hazards. Laser welding generates metal fumes similar to other welding processes, and the hazard is largely dependent on the composition of the welded metals. Ventilation is required to meet OSHA standards and American Conference of Governmental and Industrial Hygienists (ACGIH) threshold limit values. In high-power welding applications, fumes can be generated in sufficient quantities to make local exhaust ventilation, in addition to general room ventilation, both necessary and economical.

Harmful fumes or vapors can also be generated when laser energy is deposited in unwanted materials, such as breakdown of plastic enclosure materials due to laser exposure and failure of exotic lens materials due to thermal runaway. These conditions are best controlled by careful material selection and monitoring of welder performance. Finally, laser optics cleaning agents may be toxic and flammable and should be handled accordingly.

Training, Medical Examinations, and Documentation. ANSI Z136.1 requires that training in the potential hazards and control measures be provided to operators, engineers, technicians, and maintenance and service personnel. Special training on subjects such as potential hazards (including biological effects), control measures, and applicable standards is required for the laser safety officer. A model safety and training program is outlined in Appendix D of the standard.

Routine medical surveillance of laser users is no longer mandated by the standard. However, an employer may wish to provide such for medical/legal reasons (that is, to document what eye damage existed prior to commencement of laser work and that no additional damage has occurred).

Good documentation includes:

- LISTING OF LASERS, DURATION OF USE
- RESULTS OF HAZARD SURVEYS AND CALCULATIONS OF ACCESSIBLE LASER RADIATION
- INTERLOCK TESTS
- LASER USE PROCEDURES
- EMPLOYEES WORKING IN LASER AREA
- DATE AND EXTENT OF EMPLOYEE TRAINING
- DATES AND RESULTS OF MEDICAL EXAMINATIONS (IF ANY)
- TRAINING AND QUALIFICATIONS OF LASER SAFETY OFFICER

Laser-Beam Welding

Jyoti Mazumder, University of Illinois at Urbana-Champaign

References

1. I.J. SPALDING, LASER SYSTEM DEVELOPMENTS, *PHYS. BULL.*, JULY 1971, P 402
2. Y.S. ARATA AND J. MIYAMOTO, *TECHNOCRAT,* VOL 11 (NO. 5), 1978, P 33
3. C.M. BANAS, "LASER WELDING TO 100 KW," REPORT NO. R76-912260-2, UNITED TECHNOLOGIES, FEB 1977
4. C.O. BROWN AND C.M. BANAS, "DEEP PENETRATION LASER WELDING," AWS ANNUAL MEETING (SAN FRANCISCO), APRIL 1971
5. J.E. ANDERSON AND J.E. JACKSON, *WELD. J.,* VOL 44, 1965, P 1018
6. W.W. DULEY, *CO_2 LASERS: EFFECTS AND APPLICATIONS,* ACADEMIC PRESS, 1976
7. A.J. MOOREHEAD AND P.N. TURNER, *WELD. J.,* VOL 49, 1970, P 15
8. M.A. ACHAREKAR, LASERS, *WELD. ENG.,* DEC 1974, P 9-11
9. T. WEBER, HOBART LASER PRODUCTS, INC., LIVERMORE, CA, 1992, PRIVATE COMMUNICATION
10. M.M. SCHWARTZ, *METAL JOINING MANUAL,* MCGRAW-HILL, 1979
11. S.R. BOLIN, ND-YAG LASER APPLICATION SURVEY, CHAPTER 8, *LASER MATERIALS PROCESSING,* M. BASS, ED., NORTH HOLLAND PUBLISHING CO., AMSTERDAM, 1983
12. H.L. MARSHALL, *PROC. OF THE SPIE,* VOL 86, 1976
13. J. MAZUMDER, LASER WELDING, CHAPTER 3, *LASER MATERIALS PROCESSING,* M. BASS, ED., NORTH HOLLAND PUBLISHING CO., AMSTERDAM, 1983
14. L. RAPP, C. GLUMANN, F. DOUSINGER, AND H. HÜGEL, *LASER TREATMENT OF MATERIALS,* B.L. MORDIKE, ED., DGM INFORMATIONSGESELLSCHAFT MBH, GERMANY, 1992, P 99-104
15. A.S. BRANDSON AND T. ENDERS, *LASER TREATMENT OF MATERIALS,* B.L. MORDIKE, ED., DGM INFORMATIONSGESELLSCHAFT MBH, GERMANY, 1992, P 117-124
16. H. SAKAMOTO, K. SHIBATA, AND F. DAUSINGER, *LASER TREATMENT OF MATERIALS,* B.L. MORDIKE, ED., DGM INFORMATIONSGESELLSCHAFT MBH, GERMANY, 1992, P 125-130
17. J. BERKMANN, S.K. BEHLER, AND E. BEYER, *LASER TREATMENT OF MATERIALS,* B.L. MORDIKE, ED., DGM INFORMATIONSGESELLSCHAFT MBH, GERMANY, 1992, P 151-156
18. A. LANG AND H.W. BERGMAN, *LASER TREATMENT OF MATERIALS,* B.L. MORDIKE, ED., DGM INFORMATIONSGESELLSCHAFT MBH, GERMANY, 1992, P 163-170
19. L.C. MALLORY, R.F. ORR, AND W. WELLS, *LASER MATERIALS PROCESSING III,* J. MAZUMDER AND K. MUKHERJEE, ED., MINERALS, METALS, AND MATERIALS SOCIETY, 1989, P 123-134
20. R.C. CRAFER, *WELD. INST. RES. BULL.,* APRIL 1976, P 17
21. E.A. METZBOWER AND D.W. MOON, MECHANICAL PROPERTIES, FRACTURE TOUGHNESS, AND MICROSTRUCTURES OF LASER WELDS OF HIGH STRENGTH ALLOYS, *PROC. CONF. ON APPLICATIONS OF LASERS IN MATERIALS PROCESSING* (WASHINGTON, DC), 18-20 APRIL 1979
22. D.B. SNOW AND E.M. BREINAN, "EVALUATION OF BASIC WELDING CAPABILITIES," UNITED TECHNOLOGIES RESEARCH CENTER, JULY 1978
23. E.M. BREINAN, C.M. BANAS, AND M.A. GREENFIELD, LASER WELDING--THE PRESENT STATE OF THE ART, DOC IV-181-75, *11TH ANNUAL MEETING OF THE WELDING INSTITUTE* (TEL AVIV), 6-12 JULY 1975, P 1-53
24. M.M. SCHWARTZ, *METAL JOINING MANUAL,* MCGRAW-HILL, 1979
25. A. BLAKE AND J. MAZUMDER, *ASME J. ENG. IND.,* VOL 107 (NO. 3), AUG 1985, P 275-280

26. D.W. MOON AND E. METZBOWER, *WELD. J.,* 1983, P 535-585
27. E.V. LOCKE AND R.A. HELLA, *IEEE J. QUANTUM ELECTRONICS*, VOL QE-10 (NO. 2), FEB 1974, P 179-185
28. R.C. CRAFER, ADVANCES IN WELDING PROCESSES, PAPER NO. 46, *PROC. 4TH LNT. CONF.* (HARROGATE, YORKS, ENGLAND), 9-11 MAY 1978, P 267-278
29. R.A. WILLGOSS, J.H.P.C. MEGAW, AND J.N. CLARK, *WELD. MET. FABR.,* MARCH 1979, P 117-126
30. E.L. BEARDSON, D.J. SCHMATZ, AND R.E. BISARO, *WELD. J.,* VOL 52, APRIL 1973, P 227-229
31. J. MAZUMDER AND W.M. STEEN, LASER WELDING OF STEELS USED IN CAN MAKING, *WELD. J.,* VOL 60 (NO. 6), JUNE 1981, P 19-25
32. C.M. BANAS, "LASER WELDING DEVELOPMENTS," PROC. CEGB INT. CONF. ON WELDING RES. RELATED TO POWER PLANTS (SOUTHAMPTON, ENGLAND), 17-21 SEPT 1972
33. F.D. SEAMAN AND R.A. HELLA, "ESTABLISHMENT OF A CONTINUOUS WAVE LASER WELDING PROCESS," IR-809-3 (1-10), REPORT IN CONFERENCE PROCEEDINGS AFML CONTRACT F336 15-73-C5004, OCT 1976
34. E.M. BREINAN AND C.M. BANAS, PRELIMINARY EVALUATION OF LASER WELDING OF X-80 ARCTIC PIPELINE STEEL, *WRC BULL.,* DEC 1971, P 201
35. C.M. BANAS AND G.T. PETERS, "STUDY OF THE FEASIBILITY OF LASER WELDING IN MERCHANT SHIP CONSTRUCTION," CONTRACT NO. 2-36214, U.S. DEPT. OF COMMERCE, FINAL REPORT TO BETHLEHEM STEEL CORP., AUG 1974
36. M. YESSIK AND D.J. SCHMATZ, "LASER PROCESSING IN THE AUTOMOTIVE INDUSTRY," PAPER MR74-962, SME, 1974
37. F.A. DIPIETRO, *LASER SYSTEM APPLICATION IN INDUSTRY,* ATA-ASSOCIAZIONE TECNIER DELL' AUTOMOBILE, TORINO, ITALY, 1990, P 103-120
38. E. BEYER, FRAUNHOFER INSTITUTE, AACHEN, GERMANY, PRIVATE COMMUNICATION, 1992
39. J. MAZUMDER, *OPT. ENG.,* VOL 30 (NO. 8), 1991, P 1208-1219
40. C.L. CHAN, J. MAZUMDER, AND M.M. CHAN, *METALL. TRANS.* A, VOL 15, 1982, P 2175
41. W.W. DULEY, *CO_2 LASERS: EFFECTS AND APPLICATIONS,* ACADEMIC PRESS, 1976, P 241
42. A.M. MALEKA, ED., *ELECTRON BEAM WELDING PRINCIPLES AND PRACTICE,* WELDING INSTITUTE, MCGRAW-HILL, 1971, P 95-96
43. D.T. SWIFTHOOK AND E.E.F. GICK, *WELD. J.,* 1973, P 492S-499S
44. P.G. KLEMENS, *J. APPL. PHYSICS,* VOL 47, 1976, P 2165-2174
45. J.G. SICKMAN AND R. MORIJN, *PHILLIPS RES. REP.,* VOL 23, 1968, P 376
46. J.G. SICKMAN AND R. MORIJN, *PHILLIPS RES. REP.,* VOL 23, 1968, P 375
47. Y. ARATA, H. MARUO, I. MIYAMOTO, AND Y. INOUE, "DYNAMIC BEHAVIOR OF LASER WELDING," IIW DOC IV/222/77, INTERNATIONAL INSTITUTE OF WELDING, 1977
48. E.L. BAARDSEN, D.J. SCHMATZ, AND R.E. BISARO, *WELD. J.,* VOL 52, APRIL 1973, P 227-229
49. J.E. HARRY, *INDUSTRIAL APPLICATION OF LASERS,* MCGRAW-HILL, 1974
50. Y. ARATA AND I. MIYAMOTO, *LASER FOCUS,* VOL 3, 1977
51. M.H. MCCAY, T.D. MCCAY, A. SEDGHINA-SAB, AND D.R. KEEFER, *LASER MATERIALS PROCESSING III,* J. MAZUMDER AND K. MUKHERJEE, ED., MINERALS, METALS, AND MATERIALS SOCIETY, 1989
52. M.A. BRAMSON, *INFRARED RADIATION: A HANDBOOK FOR APPLICATION,* PLENUM PRESS, 1968
53. M. JØRGENSEN, *MET. CONSTR.,* VOL 12 (NO. 2), FEB 1980, P 88
54. R.C. CRAFER, *WELD. INST. RES. BULL.,* VOL 17, FEB 1976
55. W.M. STEEN, *LASER MATERIALS PROCESSING,* SPRINGER-VERLAG, 1991

Laser-Beam Welding

Jyoti Mazumder, University of Illinois at Urbana-Champaign

Selected References

- M. BORN AND E. WOLF, *PRINCIPLES OF OPTICS*, PERGAMON PRESS, 1987
- O. SVELTO, *PRINCIPLES OF LASERS*, PLENUM PRESS, 1982
- J. VERDEYEN, *LASER ELECTRONICS*, PRENTICE-HALL, 1989

Electroslag and Electrogas Welding*

S. Liu, Colorado School of Mines; S.D. Brandi, Escola de Politecnica da USP; R.D. Thomas, Jr., R.D. Thomas and Company

Introduction

ELECTROSLAG WELDING AND ELECTROGAS WELDING are two related procedures that are presently used to weld thick-section materials in the vertical or near-vertical position between retaining shoes. Primarily applied for joining steels of thicknesses over 50 mm (2 in.), electroslag welding (ESW) involves high energy input relative to other welding processes, resulting in generally inferior mechanical properties, specifically lower toughness of the heat-affected zone (HAZ). However, the high deposition rate and relatively low cost of the process make it attractive for heavy structural fabrication. The as-welded properties of electrogas welding (EGW), usually applied to steels under 50 mm (2 in.), are generally superior to those of electroslag welds, and the process is commonly applied to the field erection of storage vessels and other less critical structures.

Electroslag welding is a vertical welding process producing coalescence with molten slag which melts the filler metal and the surface of the work to be welded. Confined by cooling shoes, the molten weld pool is shielded by the molten slag, which moves along the full cross section of the joint as welding progresses. The conductive slag is maintained in a molten condition by its resistance to electric current passing between the electrode and the work (Ref 1). ESW can be considered a progressive melting and casting process in which the heat of a bath of molten flux is used to melt the filler metal and the edges of the plates to be welded. Electric arc occurs only at the beginning of the process, and once a molten bath is achieved, the arc is extinguished. During the process, flux is added periodically or continuously to maintain an adequate slag covering over the pool of molten metal. Two or more retaining shoes hold the molten metal in place until it has solidified. In normal operation with a constant potential power source, the electrode melts off while dipping only partly through the flux bath and gathers in the molten metal puddle. In the case of low-carbon steel, the temperature of the bath is reported to be in the vicinity of 1925 °C (3500 °F), while the surface temperature is approximately 1650 °C (3000 °F) (Ref 2). The major process variables are welding current and voltage. Welding current is directly responsible for the electrode melt rate, while voltage influences the base metal penetration and weld bead width. Both variables are sensitive to the physical properties of the welding flux, such as electrical resistivity and fluidity.

Electrogas welding is a method of gas-metal arc welding (if a solid wire is used) or flux-cored arc welding (if a tubular wire is used), wherein an external gas is supplied to shield the arc and molding shoes are used to confine the molten weld metal for vertical position welding (Ref 2). Electrogas welding may or may not use an added flux. In the solid wire process, CO_2 shielding gas is commonly used and no flux is added. With the flux-cored process, the core ingredients provide a small amount of flux to form a thin deposit of slag between the weld and the shoes. Self-shielding electrodes eliminate the need for external shielding gas. A major difference between ESW and EGW is that the former relies on slag conduction to carry the welding current and the latter uses arc conduction. Despite the differences, similarities between ESW and EGW in terms of equipment, joint preparation, and welding procedures are such that they can be grouped into one category and described as allied processes.

References

1. ELECTROSLAG AND ELECTROGAS WELDING, *AWS WELDING HANDBOOK*, VOL 2 (NO. 7), AWS, 1978, P 226-260
2. H.C. CAMPBELL, ELECTROSLAG, ELECTROGAS, AND RELATED WELDING PROCESSES, *WRC BULL.*, VOL 154, 1970

Note

* ADAPTED FROM R.D. THOMAS, JR., AND S. LIU, INTERPRETIVE REPORT ON ELECTROSLAG, ELECTROGAS, AND RELATED PROCESSES, *WELDING RESEARCH COUNCIL BULLETIN*, NO. 338, NOVEMBER 1988. USED WITH PERMISSION OF THE WELDING RESEARCH COUNCIL.

Electroslag and Electrogas Welding*

S. Liu, Colorado School of Mines; S.D. Brandi, Escola de Politecnica da USP; R.D. Thomas, Jr., R.D. Thomas and Company

Fundamentals of the Electroslag Process

Heat Flow Conditions. Electroslag welding is quite similar to *in situ* casting, with large volumes of molten metal and high heat content. When compared with other arc welding processes, electroslag welds have a long thermal cycle with very slow cooling rate. Electroslag welding generally consumes hundreds of kilo-joules per inch, as compared to 10 to 40 kJ/mm (250 to 1000 kJ/in.) found in most arc welding. Figure 1 shows a typical thermal cycle of an electroslag weld compared with that of an arc welded weld. As a consequence of the thermal experience, weld metal solidification is extremely slow, resulting in a coarse primary solidification structure. The heat absorbed into the base metal also creates an extremely large HAZ (Fig. 2).

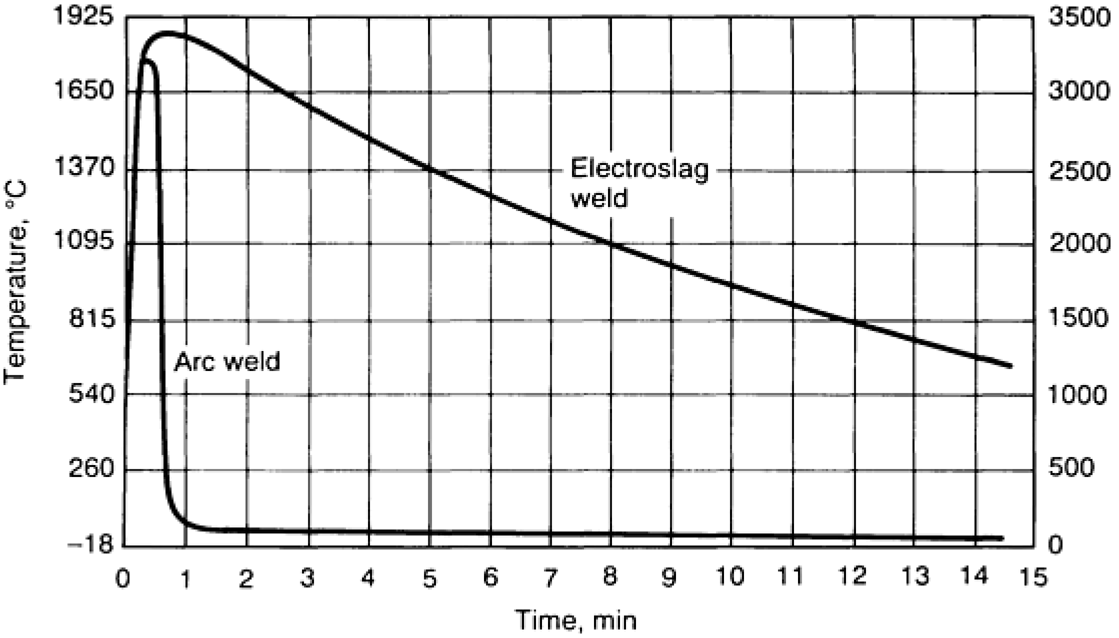

FIG. 1 TYPICAL THERMAL CYCLE OF AN ELECTROSLAG WELD RELATIVE TO THAT OF AN ARC WELDED WELD

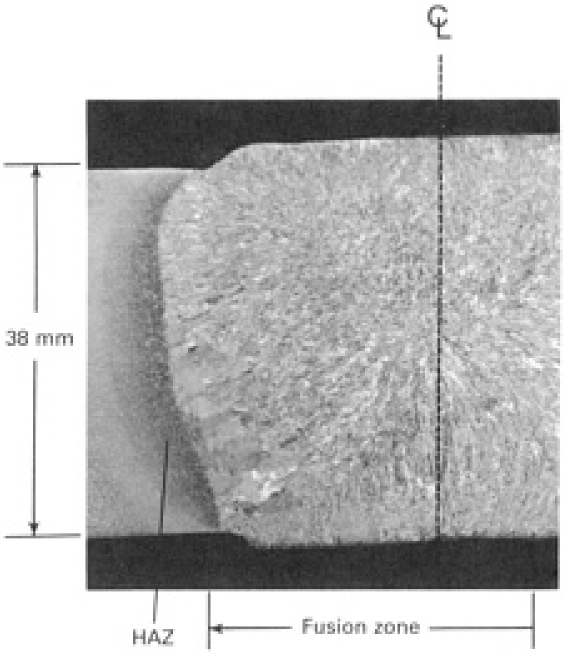
℄
38 mm
HAZ
Fusion zone

FIG. 2 MACROGRAPH SHOWING FUSION ZONE AND HEAT-AFFECTED ZONE IN AN ELECTROSLAG WELDMENT

A heat balance diagram of a typical electroslag weld (Fig. 3) illustrates that approximately 60% of the heat is absorbed by the workpart, close to 25% of the total heat is expended in the melting of the electrode, and around 10% of the heat is used to superheat the molten metal (Ref 3). The amount of heat extracted by the cooling shoes varies, depending on the thickness of the plate and on the welding conditions. In the welding of steel plates 90 mm (3 $\frac{1}{2}$ in.) thick, less than 10% of the heat of the molten slag and metal pool is transferred to the cooling shoes. In thinner plates, however, the cooling shoes play a more significant role in the heat balance.

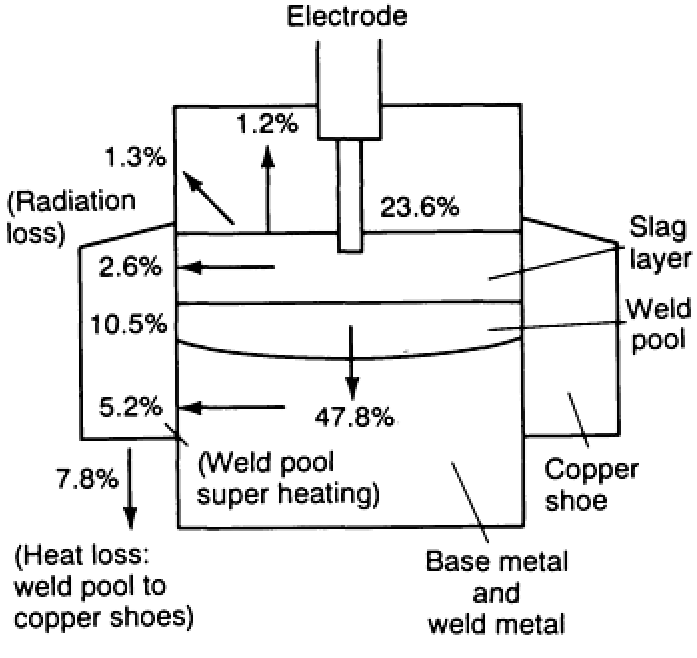
Electrode
1.2%
1.3%
(Radiation loss)
23.6%
Slag layer
2.6%
Weld pool
10.5%
5.2%
47.8%
7.8%
(Weld pool super heating)
Copper shoe
(Heat loss: weld pool to copper shoes)
Base metal and weld metal

FIG. 3 HEAT BALANCE DIAGRAM OF A TYPICAL ELECTROSLAG WELDMENT. SOURCE: REF 3

Mathematical models are used to estimate the three-dimensional temperature field in the slag, metal pool, and base metal regions in an electroslag weldment, and to predict HAZ size and grain growth in the HAZ. Figure 4 shows an example of the calculated temperature distribution for a base plate 25.4 mm (1 in.) thick, 470 mm (18.5 in.) long, and 610 mm (24 in.) wide, at the time that corresponds to half the total time required for completion of the weld (Ref 4). The maximum temperatures reached in the slag and weld pool were 2230 K (3555 °F) and 1900 K (2960 °F), respectively. The temperature distribution in a weldment along planes parallel and normal to the parent plate surface can also be represented in the form of isometric temperature plots (Fig. 5).

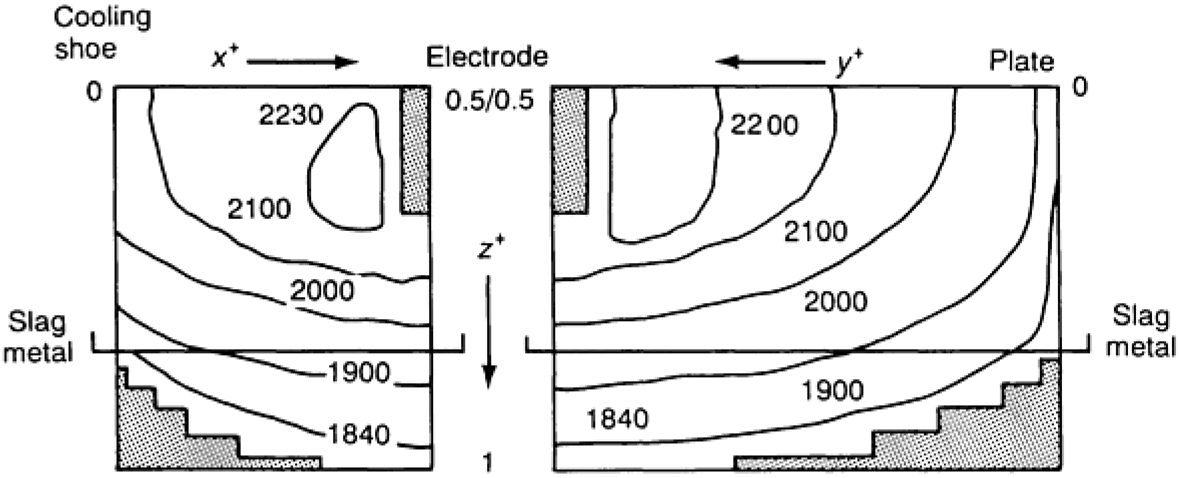

FIG. 4 CALCULATED TEMPERATURE DISTRIBUTION OF AN ELECTROSLAG WELDMENT SHOWING THREE-DIMENSIONAL TEMPERATURE FIELD IN THE SLAG, METAL POOL, AND BASE METAL. ALL TEMPERATURES SHOWN ARE IN KELVIN. SOURCE: REF 4

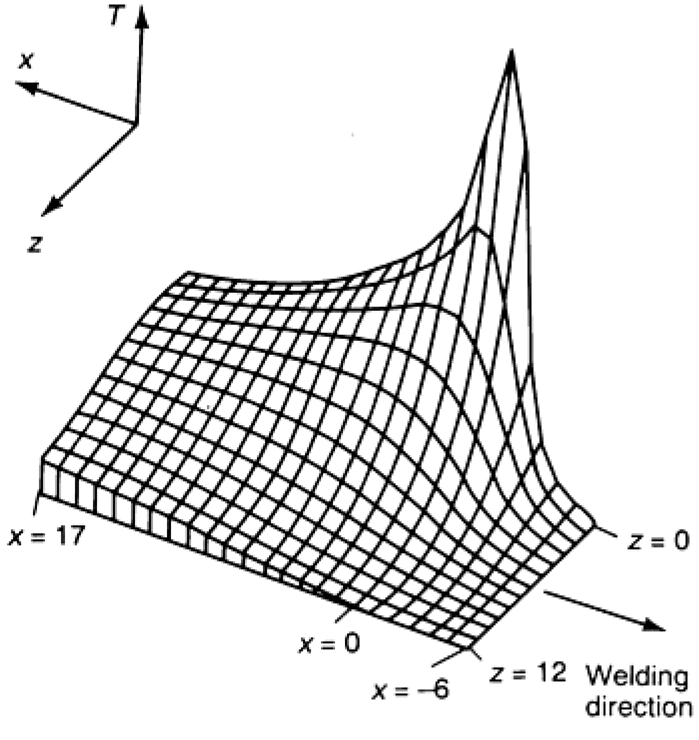

FIG. 5 ISOMETRIC THREE-DIMENSIONAL TEMPERATURE DISTRIBUTION IN AN ELECTROSLAG WELDMENT ALLOWING IMPROVED VISUALIZATION OF THE TEMPERATURE PROFILE. SOURCE: REF 5

The energy balance in the slag phase is summarized in Table 1. A major portion of the heat generated in the slag layer was transferred to the cooling shoes (36.1%). Approximately 22 and 15% of the heat was used in base metal heating and electrode melting, respectively. The heat-generating patterns are highly sensitive to the geometric location of the electrode in the slag. With careful control of the process variables (such as gap distance and convective flow suppression) by external application of electromagnetic field, reductions of two or three times the usual heat input could be practical.

TABLE 1 ENERGY BALANCE IN THE SLAG PHASE OF ELECTROSLAG WELDING

PARAMETER[(A)]	ENERGY DISSIPATED, %
HEAT TRANSFER FROM SLAG:	
TO COPPER SHOE	36.10
TO BASE PLATES	20.22
TO LIQUID METAL	8.71
HEAT USED TO HEAT THE ELECTRODE TO ITS MELTING POINT	15.26
HEAT REQUIRED:	
TO MELT THE ELECTRODE	4.07
TO HEAT FLUX COOLING ON THE CONSUMABLE GUIDE	1.84
HEAT CARRIED BY LIQUID METAL DROPS	2.04
RADIATION LOSS FROM THE SLAG SURFACE	1.75
ELECTROCHEMICAL LOSS	10.00

Source: Ref 5

(A) ACTUAL WIDTH OF THE SLAG POOL 0.044 MM (0.0017 IN.); CURRENT 534 A; OTHER CONDITIONS ARE THE SAME AS THOSE PRESENTED IN TABLE 2.

To distribute the energy uniformly across the thickness of the weldment, empirical rules have been developed (Table 2) to determine the preferred number of electrodes, the wire spacing, and the traverse (oscillation) distance (Fig. 6). Oscillation speeds depend on the plate thickness, usually allowing a traverse time of 3 to 5 s; a dwell time at the end of each traverse ensures adequate penetration at the plate edges.

TABLE 2 EFFECT OF JOINT GEOMETRY ON MULTIPLE-WIRE BUTT ELECTROSLAG WELDING

STEEL PLATE THICKNESS		NUMBER OF WIRES	OSCILLATION	SQUARE BUTT GAP OPENING	
mm	in.			mm	in.
19-75	$\frac{3}{4}$ -3	1	NO	22-32	$\frac{7}{8}$ -1 $\frac{1}{4}$
25-130	1-5	1	YES	32-35	1 $\frac{1}{4}$ -1 $\frac{3}{8}$
75-150	3-6	2	NO	32-35	1 $\frac{1}{4}$ -1 $\frac{3}{8}$
75-280	3-11	2	YES	32-35	1 $\frac{1}{4}$ -1 $\frac{3}{8}$
130-535	5-21	3	YES	35	1 $\frac{3}{8}$

Source: Ref 6

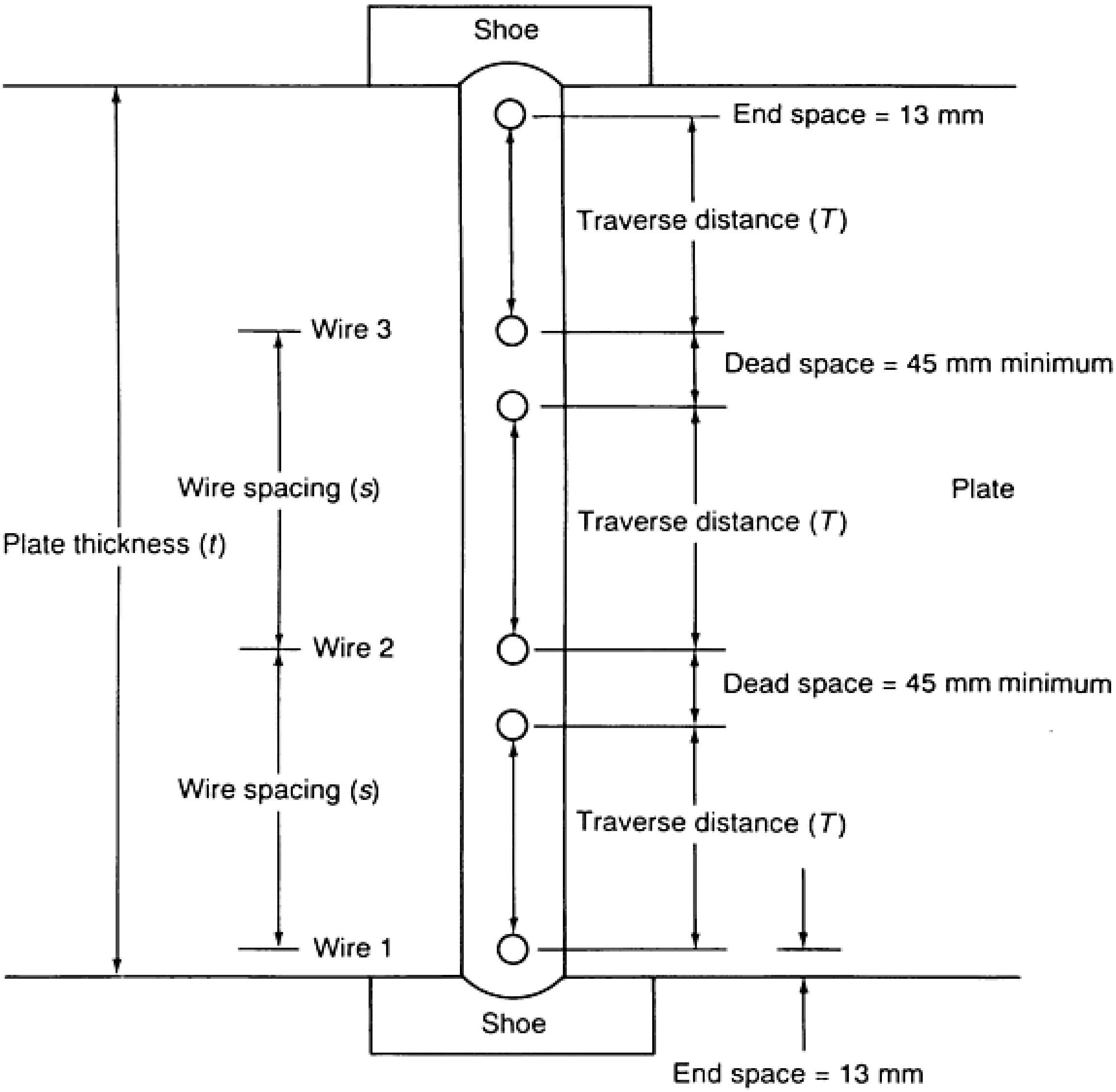
Shoe
End space = 13 mm
Traverse distance (T)
Wire 3
Dead space = 45 mm minimum
Wire spacing (s)
Plate
Traverse distance (T)
Plate thickness (t)
Wire 2
Dead space = 45 mm minimum
Wire spacing (s)
Traverse distance (T)
Wire 1
Shoe
End space = 13 mm

FIG. 6 WIRE SPACING DIAGRAM FOR MULTIPLE-ELECTRODE ELECTROSLAG WELDING INCORPORATING OSCILLATION FOR BETTER ENERGY DISTRIBUTION ACROSS THE THICKNESS OF THE WELDMENT. SOURCE: REF 6

Metal Transfer and Weld Pool Morphology. The droplet transfer rate and the length of time each droplet is in contact with the slag layer profoundly affect the chemical composition and the metallurgical properties of the weld pool.

Droplet Formation. In steady-state operation, the molten filler metal transfers into the weld pool in a globular mode. Current and voltage oscillograms indicate that despite the droplet size decrease observed with increasing voltage, welding current has no significant effect on the size of the droplets. Rather, the length of time that a droplet is in contact with the slag layer decreases considerably with increasing current. On the other hand, the length of time a droplet is in contact with the layer of slag increases with voltage, because the electrode tip is further away from the molten metal pool. The extent of interaction between a metal droplet and the slag layer determines the chemical composition of the weld pool.

Weld Pool Morphology. In an electroslag weldment, solidification begins at the fusion line, surfaces adjacent to the retaining shoes, and progresses toward the center of the weld. Because the process is continuous in a vertical or near vertical-up position, solidification also progresses from the bottom toward the top part of the joint. The angles at which the columnar grains meet at the center of the weld depend on the shape of the weld pool, which can be described by the weld pool form factor. Form factor is defined as the ratio between the width, *W*, and the maximum depth, *h*, of the pool (Fig. 7). Welds having a high form factor (>2.0) will have grains meeting at an acute angle at the center line, while welds with a low form factor (<1.0) will solidify with grains meeting at an obtuse angle. Low form factor is highly undesirable because of the potential accumulation of residual elements at the center of the weld joint. High welding current usually results in a low form factor, while low welding current usually results in a high form factor and shallow metal pool. High voltage promotes shallow pools and low voltage. Figure 7 shows two examples of different pool morphologies and form factors as a function of the welding current and voltage.

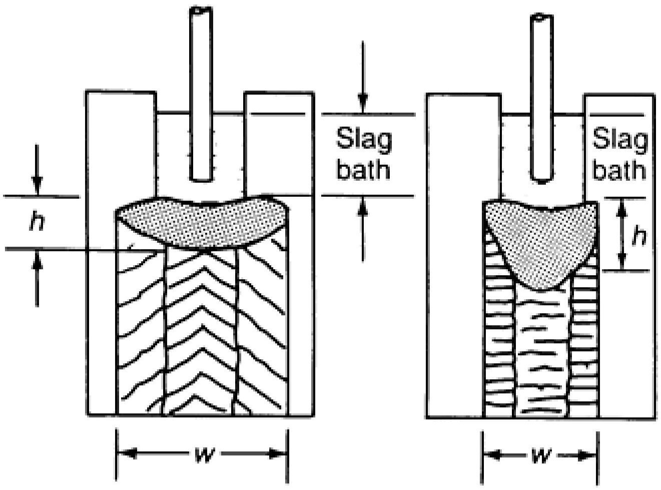

EFFECT OF WELDING PARAMETERS ON WELD METAL POOL SHAPE

PARAMETER	EFFECT OF PARAMETER INCREASE ON WELD-METAL POOL DIMENSION(A)		
	POOL WIDTH (*W*)	POOL DEPTH (*H*)	FORM FACTOR (*F* = *W*/*H*)
CURRENT OR WIRE-FEED SPEED:			
LOW VALUES	A	A	B
HIGH VALUES	D	A	D
VOLTAGE	A	B	A
SLAG POOL DEPTH	D	B	D
ROOT OPENING	A	C	A

FIG. 7 EFFECT OF INCREASE IN WELDING PARAMETERS ON WELD POOL FORM FACTOR. (A) OPTIMUM WELD POOL DIMENSIONS (SHALLOW WELD POOL, HIGH FORM FACTOR, ACUTE ANGLE BETWEEN GRAINS). (B) UNDESIRABLE WELD POOL DIMENSIONS (DEEP WELD POOL, LOW FORM FACTOR, OBTUSE ANGLE BETWEEN GRAINS). SOURCE: REF 6

An additional factor controlling the weld puddle morphology is the welding flux conductivity. Actually, the bulk of the electrical energy is converted into thermal energy in a thin layer of the slag contiguous to the electrode tip, which acts as the heat source in ESW (Ref 3). High conductivity fluxes generate less heat and result in less base metal penetration and shallower weld pools.

Out-of-Position (Nonvertical) ESW. Electroslag welding has been successfully done in joints that are inclined from the vertical position. The stability of the process and the quality of the weld joint are both functions of the angle of inclination. Weld morphology, including penetration, HAZ, and weld metal microstructure, is sensitive to the welding parameters. Preferential radiative heating of the upper weld surface is considered the cause of asymmetric penetration and HAZ geometry (Ref 7).

Weld Pool Penetration and Magnetic Field Coupling. Magnetic field is not generally applied in ESW. For electroslag overlays, however, some use of magnetic field to promote fluid flow and to drive the molten slag layer from the hotter region to the colder region to eliminate undercutting has been reported (Ref 8). However, the effect of magnetic stirring, which induced convective fluid motion in the weld pool and disrupted the solidification and growth of the columnar dendrites, can be seen in Fig. 8 (Ref 9, 10). Mechanical vibration on electroslag weld pools generally does not cause significant microstructural modification. However, some refinement of the as-deposited grains can result when the slag-molten metal interface is disturbed. The use of a quartz-shielded electrode guide tube as an extended stirrer has been reported to form a deeper weld pool, with the form factor decreasing from 3 to 1.5 (Ref 11, 12). Other techniques for refining weld metal microstructures include ultrasonic cavitation and the use of chemical inoculators.

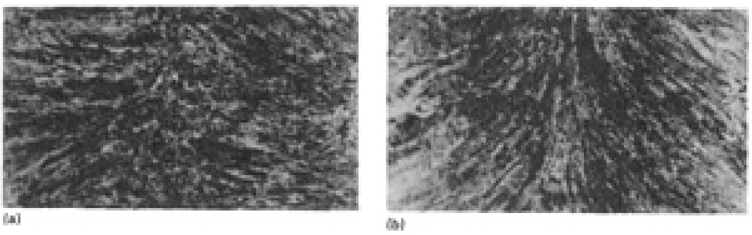

FIG. 8 MACROGRAPH SHOWING EFFECT OF SUPERIMPOSED MAGNETIC FIELD ON ELECTROSLAG WELDMENTS. (A) FIELD PRESENT. (B) FIELD ABSENT. BOTH 1.6×. SOURCE: REF 9

References cited in this section

3. B. PATON, *ELECTROSLAG WELDING,* AWS, 1962
4. T. DEBROY, J. SZEKELY, AND T. EAGER, HEAT GENERATION PATTERNS AND TEMPERATURE PROFILES IN ELECTROSLAG WELDING, *METALL. TRANS. B,* VOL 11, 1980, P 593-605
5. R.H. FROST, J.E. JONES, AND D.L. OLSON, ELECTROSLAG WELDING OF PRESSURE VESSEL STEELS, *ELECTROSLAG WELDING FOR MARINE APPLICATIONS,* PAPER 2, U.S. NAVY ACADEMY, MARCH 1985
6. "ELECTROSLAG ELECTROGAS TIPS & TECHNIQUES," BULLETIN OF ELECTROTHERM CORP., 1973
7. J.E. JONES, D.L. OLSON, AND G.P. MARTINAS, METALLURGICAL AND THERMAL CHARACTERISTICS OF NON-VERTICAL ELECTROSLAG WELDING, *WELD. J.,* VOL 59 (NO. 9), 1980, P 245S-254S
8. S. NAKANO, N. NISHIYAMA, T. HIRO, AND J. TSUBOI, MAGLAY PROCESS--ELECTROMAGNETIC CONTROLLED OVERLAY WELDING PROCESS WITH ESW, *KAWASAKI STEEL TECH. REP.,* VOL 2, 1981, P 31-42
9. S. LIU AND C.T. SU, GRAIN REFINEMENT IN ELECTROSLAG WELDMENTS BY METAL POWDER ADDITION, *WELD. J.,* VOL 68 (NO. 4), 1989, P 132S-144S
10. C.T. SU, "HIGH SPEED ELECTROSLAG WELDING," M.S. THESIS, PENNSYLVANIA STATE UNIV., 1987
11. S. VENKATARAMAN, J.H. DEVLETIAN, W.E. WOOD, AND D.G. ATTERIDGE, GRAIN REFINEMENT DEPENDENCE ON SOLIDIFICATION AND SOLID STATE REACTIONS IN ELECTROSLAG WELDS, *GRAIN REFINEMENT IN CASTINGS AND WELDS,* AIME, 1982, P 275-288
12. D.G. ATTERIDGE, S. VENKATARAMAN, AND W.E. WOOD, "IMPROVING THE RELIABILITY AND INTEGRITY OF CONSUMABLE GUIDE ELECTROSLAG WELDMENTS IN BRIDGE STRUCTURES," RESEARCH REPORT, OREGON GRADUATE CENTER, 1982

Electroslag and Electrogas Welding*

S. Liu, Colorado School of Mines; S.D. Brandi, Escola de Politecnica da USP; R.D. Thomas, Jr., R.D. Thomas and Company

Constitutive Equations for Welding Current, Voltage, and Travel Rate

The power input per unit length of a weld is the primary variable that influences the degree of base metal melting, the amount of filler metal deposition, and the thermal history of the weldment. In ESW, current and travel speed are not independent variables, as in the other arc welding processes. For a constant voltage power supply, the current and power are both functions of the resistance of the slag pool. Hence, they are functions of the electrode feed rate, the mechanics of electrode melting, and the nature of the electrical and thermal transport at the electrodeslag interface. To control the welding process, a constitutive equation has been proposed (Ref 13):

$$I = B\, W^{1/2} V^{1/3} \qquad \textbf{(EQ 1)}$$

where I is the welding current, W is the electrode feed rate, V is the welding voltage, and B is a proportionality constant. Equation 1 expresses the functional relationship between the various process variables and has the primary objective of selecting a voltage and current range that provides sufficient energy input for base plate penetration without excessive overheating of the HAZ. The relation can be empirically determined from Fig. 9. The total energy input per unit area of weld surface, E, can also be determined:

$$E = B \frac{L_g}{2p\, r^2 (L+g)} . \frac{V^{5/3}}{I} \qquad \textbf{(EQ 2)}$$

where L is the width of the base plate, g is the root gap opening, and r is the radius of the welding electrode. Note also that this equation takes into consideration the electrode-to-joint geometry, which is related to the electrical and thermal transport in the process.

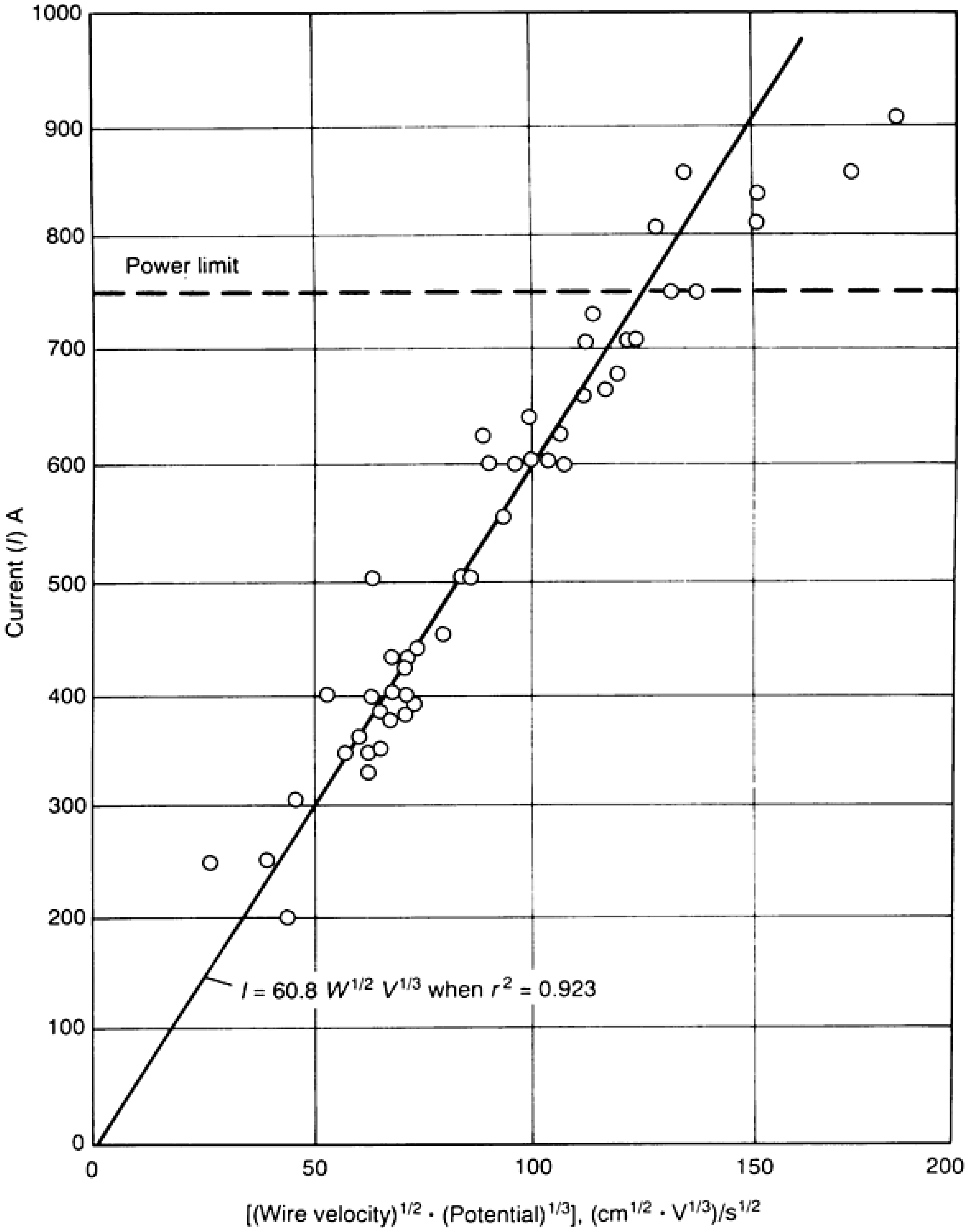

FIG. 9 FUNCTIONAL RELATION AMONG WELDING CURRENT (*I*), POTENTIAL (*V*), AND WIRE-FEED RATE (*W*) FOR ELECTROSLAG WELDING OF 2 $\frac{1}{4}$ CR-1MO STEEL. LINE INDICATES VALUE OF $I = BW^{1/2}V^{1/3}$ FOR B = 60.8 AND R = 0.923. SOURCE: REF 14

Unlike arc welding processes, in which an increase in current results in greater energy input and deeper penetration, the reverse occurs in ESW. An increase in current requires an increase in electrode burn-off rate, which increases the welding speed and decreases the energy input per unit area of weld. An excessive increase in the current at constant voltage, however, can shift the process from penetration to nonpenetration of the base plate. The process control boundaries for ESW can be easily developed (Fig. 10). Boundaries A through D enclose an operating space that exhibits good process stability and base metal melting. Outside the indicated boundary, the process often appears to operate well but with inadequate base metal penetration (Ref 13).

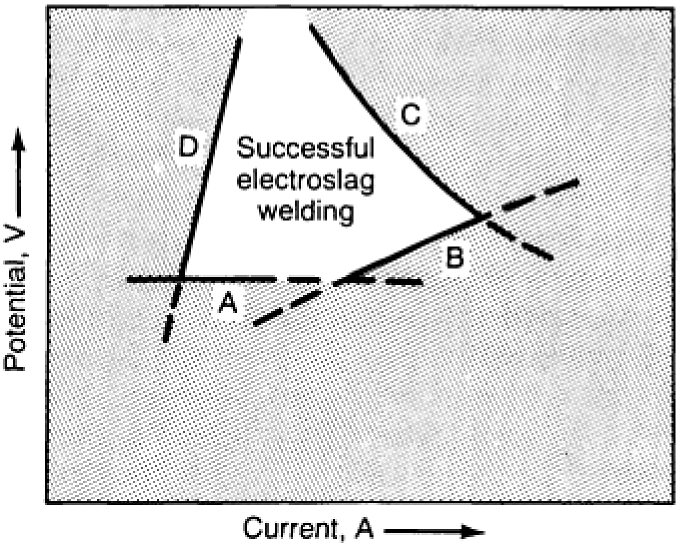

FIG. 10 OPERATING PARAMETER WINDOW FOR ELECTROSLAG WELDING. BOUNDARY A REPRESENTS THE VOLTAGE THRESHOLD FOR PARENT PLATE FUSION AT LOW-POWER INPUTS. BOUNDARY B REPRESENTS THE CONSTITUTIVE EQUATION FOR ADEQUATE PENETRATION AT HIGH POWER LEVELS. BOUNDARY C REPRESENTS THE MAXIMUM POWER OUTPUT OF THE WELDING POWER SUPPLY. BOUNDARY D REPRESENTS THE LIMIT OF ELECTRODE FEED RATE AT WHICH THE WIRE ELECTRODE MELTS BY OHMIC HEATING. SOURCE: REF 14

A change in electrode geometry and nature (cored) may result in a different constitutive equation (Ref 9, 15):

$$I = B' W^{0.35} V^{1.63} \qquad \textbf{(EQ 3)}$$

where B' is the proportionality constant. The total energy input, E, is:

$$E = B' \frac{L_g - 2w_g t_g}{[L + (g + 2t_g)][w_c t_c d_c]} \cdot \frac{V^4}{I^2} \qquad \textbf{(EQ 4)}$$

where w_g is width of the consumable plate guide; t_g is the thickness of the consumable plate guide; w_e is the width of the consumable electrode; and t_e is the thickness of the consumable electrode.

References cited in this section

9. S. LIU AND C.T. SU, GRAIN REFINEMENT IN ELECTROSLAG WELDMENTS BY METAL POWDER ADDITION, *WELD. J.,* VOL 68 (NO. 4), 1989, P 132S-144S
13. R.H. FROST, G.R. EDWARDS, AND M.D. RHEINLANDER, A CONSTITUTIVE EQUATION FOR THE CRITICAL ENERGY INPUT DURING ELECTROSLAG WELDING, *WELD. J.,* VOL 60 (NO. 1), 1981, P 12-62
14. S. LIU AND C.T. SU, PERFORMANCE EVALUATION OF A METAL POWDER CORED STRIP ELECTRODE IN HIGH SPEED ELECTROSLAG WELDING, *ADVANCES IN WELDING SCIENCE AND TECHNOLOGY,* ASM INTERNATIONAL, 1986, P 401-412
15. R.D. THOMAS, JR., CORROSION RESISTANT WELD OVERLAYS BY THE DUAL STRIP PROCESS, *BRIT. WELD. J.,* MAY 1966

Electroslag and Electrogas Welding*

S. Liu, Colorado School of Mines; S.D. Brandi, Escola de Politecnica da USP; R.D. Thomas, Jr., R.D. Thomas and Company

ESW Consumables

ESW Fluxes. Similar to the fluxes used in other welding processes, ESW fluxes are formulated to refine the weld metal, to coat the surface of the retaining shoes and the completed weld metal, and to protect the molten metal from oxidation. However, several physical and chemical properties distinguish ESW fluxes from the others used in arc welding. Electroslag welding fluxes are invariably fused, rather than agglomerated. Compared with the arc welding fluxes, fluxes for ESW are higher in resistivity, because the arc is extinguished soon after the process becomes stable. Sometimes, an agglomerated starting flux with high conductivity is used to initiate the process and form the weld pool. After that, a running flux of high resistivity is added to generate heat for melting the filler metal and to maintain steady welding operation. In fact, for high-current welding the use of starting flux is often omitted.

A typical running flux for low-carbon steels has the following composition (Ref 11):

CONSTITUENT	CONTENT, WT%
SIO_2	25
MNO	10
CAF_2	15
AL_2O_3	25
CAO	15
MGO	10

Electrical Resistivity. A flux of high resistance (or low conductivity) will draw less current, resulting in a colder weld pool and lower base metal penetration. It will also allow the wire to drive deeper into the pool. On the other hand, a flux of low resistance may draw excessive current, raising the temperature of the bath until the process stabilizes with a shorter electrode extension. However, if resistance is too low, then arcing may occur between the electrode and the slag bath surface, especially at higher voltages. This condition is aggravated in fluxes that show steeply increasing conductivity with rising temperature. In terms of process conditions, lower slag resistivity may result in operation at lower voltage.

Fluidity. In addition to its heat-generating ability, the molten slag must also have sufficient fluidity to cause rapid convection and good circulation, needed to distribute heat throughout the weld joint. Fluidity of a slag depends mainly on its chemical composition and operating temperature. The melting point of a flux must be below that of the base metal for weld pool refining, and its boiling temperature must be higher than the operating temperature to avoid loss by vaporization. Any preferential loss (of one or more ingredients) will alter the bath composition, which may greatly change the slag fluidity and electrical conductivity, resulting in improper shielding. If the composition of the fluxes is altered during operation, the change in energy imparted to the slag may cause an increase in temperature and in conductivity during welding of long seams, which may give rise to arcing on the top surface of the slag bath.

Low fluidity will tend to trap slag inclusions in the weld metal; excessive fluidity will cause leakage through the small space between the work and the retaining shoes. For joining thin plates, higher fluidity fluxes are desirable to obtain good circulation in the relatively small bath.

Metallurgical Compatibility. Finally, the slag should be metallurgically compatible with the alloy being welded. For steel welding, the fluxes are generally mixed oxides of silicon, manganese, titanium, calcium, magnesium, and aluminum (Ref 1). These oxide components play an important role in the shielding and refining of the weld pool. Calcium fluoride (CaF_2) is added to basic oxides or silicate systems to achieve the proper resistivity and fluidity. Increasing CaF_2 decreases viscosity, melting point, and resistivity. Additions of TiO_2 also decrease resistivity, while Al_2O_3 increases it. However, TiO_2 also increases slag viscosity (Ref 2, 3). Special applications, such as inclusion control or sulfur removal, may require the addition of rare-earth compounds (Ref 16).

Slag detachability after weld metal solidification is not a major problem in ESW. The addition of large quantities of TiO_2, however, will often result in difficult slag removal. Fluoride additions generally improve slag detachability.

ESW Electrodes. Various ESW techniques are available, depending on the type of electrodes and the feeding mechanism. The electrodes can be solid wires, tubular flux-cored wires, large-section solid electrodes, and large-section cored electrodes. The guides or nozzles can be consumable or nonconsumable. The conventional method uses nonconsumable guides (also known as snorkels), which are maintained approximately 50 to 75 mm (2.0 to 3.0 in.) above the molten flux. In this case, a mobile feeding head is used and is raised vertically to match the weld pool travel speed. If the electrode feeding mechanism is stationary, then the nozzle will be "consumed," becoming part of the weld metal when it is reached by the slag pool. Accordingly, the materials used in consumable guides generally match the chemical composition of the electrode or base metal. A consumable guide can either be a thin-wall tube or an assembly of plates or rods with conduits to feed the electrode wire. Fin-shape guide tubes are prepared by welding webs to circular-section guide tubes to obtain more uniform weld pool heating (Ref 1, 3, 6). Among the different types of electrodes, only the wire type and the large-section cored electrodes require the use of nozzles for delivering the electrode into the root gap. Bare guides sometimes need insulation along their lateral surfaces if they cannot be precisely aligned within the weld joint. One solution is to insert insulating plugs at critical locations in the assembly. Flux-coated consumable nozzles are available to minimize the problem of insulating the electrode assembly. Figure 11 shows several shapes of electrodes and guides used in ESW.

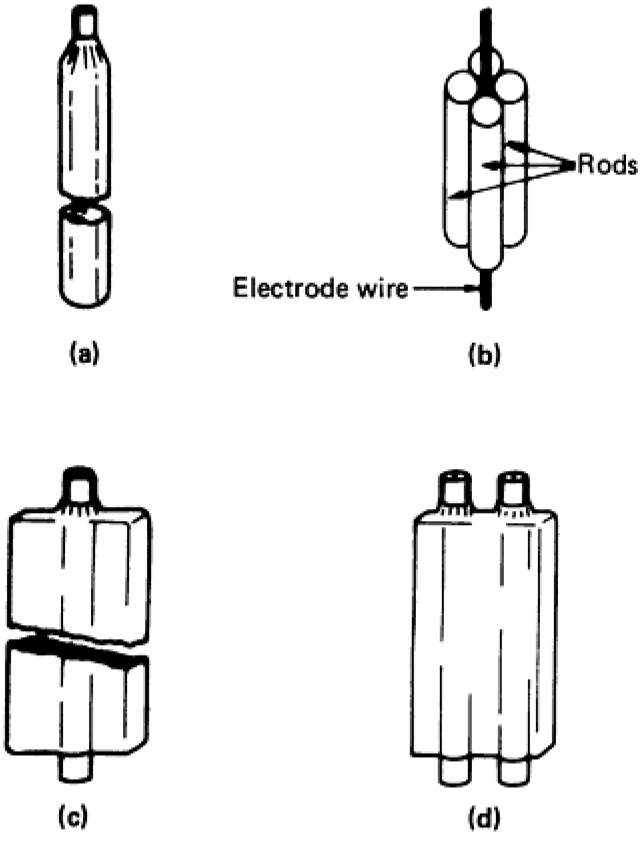

FIG. 11 SELECTED ELECTRODES AND GUIDES (NOZZLES) USED IN ELECTROSLAG WELDING. (A) SINGLE FLUX-COVERED TUBE. (B) CLUSTER OF RODS TAPED TOGETHER. (C) FLUX-COVERED WING NOZZLE. (D) FLUX-COVERED WING OR WEB NOZZLE WITH TWO TUBES

References cited in this section

1. ELECTROSLAG AND ELECTROGAS WELDING, *AWS WELDING HANDBOOK,* VOL 2 (NO. 7), AWS, 1978, P 226-260
2. H.C. CAMPBELL, ELECTROSLAG, ELECTROGAS, AND RELATED WELDING PROCESSES, *WRC*

BULL., VOL 154, 1970

3. B. PATON, *ELECTROSLAG WELDING,* AWS, 1962
6. "ELECTROSLAG ELECTROGAS TIPS & TECHNIQUES," BULLETIN OF ELECTROTHERM CORP., 1973
11. S. VENKATARAMAN, J.H. DEVLETIAN, W.E. WOOD, AND D.G. ATTERIDGE, GRAIN REFINEMENT DEPENDENCE ON SOLIDIFICATION AND SOLID STATE REACTIONS IN ELECTROSLAG WELDS, *GRAIN REFINEMENT IN CASTINGS AND WELDS,* AIME, 1982, P 275-288
16. A. MITCHELL AND G. BEYNON, ELECTROCHEMICAL REACTIONS IN THE ELECTROSLAG WELDING OF THICK WALLED STRUCTURES, *AVTOM. SVARKA,* VOL 6, 1965, P 32-37

Electroslag and Electrogas Welding[*]

S. Liu, Colorado School of Mines; S.D. Brandi, Escola de Politecnica da USP; R.D. Thomas, Jr., R.D. Thomas and Company

ESW Consumables

ESW Fluxes. Similar to the fluxes used in other welding processes, ESW fluxes are formulated to refine the weld metal, to coat the surface of the retaining shoes and the completed weld metal, and to protect the molten metal from oxidation. However, several physical and chemical properties distinguish ESW fluxes from the others used in arc welding. Electroslag welding fluxes are invariably fused, rather than agglomerated. Compared with the arc welding fluxes, fluxes for ESW are higher in resistivity, because the arc is extinguished soon after the process becomes stable. Sometimes, an agglomerated starting flux with high conductivity is used to initiate the process and form the weld pool. After that, a running flux of high resistivity is added to generate heat for melting the filler metal and to maintain steady welding operation. In fact, for high-current welding the use of starting flux is often omitted.

A typical running flux for low-carbon steels has the following composition (Ref 11):

CONSTITUENT	CONTENT, WT%
SIO_2	25
MNO	10
CAF_2	15
AL_2O_3	25
CAO	15
MGO	10

Electrical Resistivity. A flux of high resistance (or low conductivity) will draw less current, resulting in a colder weld pool and lower base metal penetration. It will also allow the wire to drive deeper into the pool. On the other hand, a flux of low resistance may draw excessive current, raising the temperature of the bath until the process stabilizes with a shorter electrode extension. However, if resistance is too low, then arcing may occur between the electrode and the slag bath surface, especially at higher voltages. This condition is aggravated in fluxes that show steeply increasing conductivity with rising temperature. In terms of process conditions, lower slag resistivity may result in operation at lower voltage.

Fluidity. In addition to its heat-generating ability, the molten slag must also have sufficient fluidity to cause rapid convection and good circulation, needed to distribute heat throughout the weld joint. Fluidity of a slag depends mainly on its chemical composition and operating temperature. The melting point of a flux must be below that of the base metal for weld pool refining, and its boiling temperature must be higher than the operating temperature to avoid loss by vaporization. Any preferential loss (of one or more ingredients) will alter the bath composition, which may greatly change the slag fluidity and electrical conductivity, resulting in improper shielding. If the composition of the fluxes is altered during operation, the change in energy imparted to the slag may cause an increase in temperature and in conductivity during welding of long seams, which may give rise to arcing on the top surface of the slag bath.

Low fluidity will tend to trap slag inclusions in the weld metal; excessive fluidity will cause leakage through the small space between the work and the retaining shoes. For joining thin plates, higher fluidity fluxes are desirable to obtain good circulation in the relatively small bath.

Metallurgical Compatibility. Finally, the slag should be metallurgically compatible with the alloy being welded. For steel welding, the fluxes are generally mixed oxides of silicon, manganese, titanium, calcium, magnesium, and aluminum (Ref 1). These oxide components play an important role in the shielding and refining of the weld pool. Calcium fluoride (CaF_2) is added to basic oxides or silicate systems to achieve the proper resistivity and fluidity. Increasing CaF_2 decreases viscosity, melting point, and resistivity. Additions of TiO_2 also decrease resistivity, while Al_2O_3 increases it. However, TiO_2 also increases slag viscosity (Ref 2, 3). Special applications, such as inclusion control or sulfur removal, may require the addition of rare-earth compounds (Ref 16).

Slag detachability after weld metal solidification is not a major problem in ESW. The addition of large quantities of TiO_2, however, will often result in difficult slag removal. Fluoride additions generally improve slag detachability.

ESW Electrodes. Various ESW techniques are available, depending on the type of electrodes and the feeding mechanism. The electrodes can be solid wires, tubular flux-cored wires, large-section solid electrodes, and large-section cored electrodes. The guides or nozzles can be consumable or nonconsumable. The conventional method uses nonconsumable guides (also known as snorkels), which are maintained approximately 50 to 75 mm (2.0 to 3.0 in.) above the molten flux. In this case, a mobile feeding head is used and is raised vertically to match the weld pool travel speed. If the electrode feeding mechanism is stationary, then the nozzle will be "consumed," becoming part of the weld metal when it is reached by the slag pool. Accordingly, the materials used in consumable guides generally match the chemical composition of the electrode or base metal. A consumable guide can either be a thin-wall tube or an assembly of plates or rods with conduits to feed the electrode wire. Fin-shape guide tubes are prepared by welding webs to circular-section guide tubes to obtain more uniform weld pool heating (Ref 1, 3, 6). Among the different types of electrodes, only the wire type and the large-section cored electrodes require the use of nozzles for delivering the electrode into the root gap. Bare guides sometimes need insulation along their lateral surfaces if they cannot be precisely aligned within the weld joint. One solution is to insert insulating plugs at critical locations in the assembly. Flux-coated consumable nozzles are available to minimize the problem of insulating the electrode assembly. Figure 11 shows several shapes of electrodes and guides used in ESW.

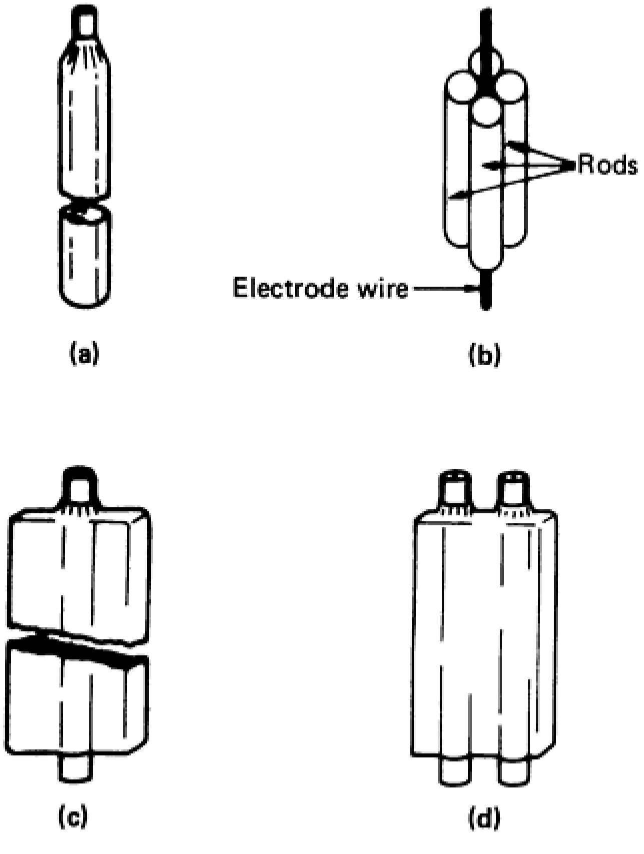

FIG. 11 SELECTED ELECTRODES AND GUIDES (NOZZLES) USED IN ELECTROSLAG WELDING. (A) SINGLE FLUX-COVERED TUBE. (B) CLUSTER OF RODS TAPED TOGETHER. (C) FLUX-COVERED WING NOZZLE. (D) FLUX-COVERED WING OR WEB NOZZLE WITH TWO TUBES

References cited in this section

1. ELECTROSLAG AND ELECTROGAS WELDING, *AWS WELDING HANDBOOK,* VOL 2 (NO. 7), AWS, 1978, P 226-260
2. H.C. CAMPBELL, ELECTROSLAG, ELECTROGAS, AND RELATED WELDING PROCESSES, *WRC BULL.,* VOL 154, 1970
3. B. PATON, *ELECTROSLAG WELDING,* AWS, 1962
6. "ELECTROSLAG ELECTROGAS TIPS & TECHNIQUES," BULLETIN OF ELECTROTHERM CORP., 1973
11. S. VENKATARAMAN, J.H. DEVLETIAN, W.E. WOOD, AND D.G. ATTERIDGE, GRAIN REFINEMENT DEPENDENCE ON SOLIDIFICATION AND SOLID STATE REACTIONS IN ELECTROSLAG WELDS, *GRAIN REFINEMENT IN CASTINGS AND WELDS,* AIME, 1982, P 275-288
16. A. MITCHELL AND G. BEYNON, ELECTROCHEMICAL REACTIONS IN THE ELECTROSLAG WELDING OF THICK WALLED STRUCTURES, *AVTOM. SVARKA,* VOL 6, 1965, P 32-37

Electroslag and Electrogas Welding*

S. Liu, Colorado School of Mines; S.D. Brandi, Escola de Politecnica da USP; R.D. Thomas, Jr., R.D. Thomas and Company

Metallurgical and Chemical Reactions

Fusion Zone Compositional Effects. In ESW of carbon and low-alloy steels, the major reactions involve manganese, silicon, carbon, and impurities such as oxygen, sulfur, and phosphorus. In the welding of high-alloy steels, reactions involving chromium, titanium, aluminum, and other elements are important. The extent and control of reactions are determined mainly by the concentrations of the elements in the liquid metal and by the concentrations of their oxides in the slag. The temperature in the reaction zone, the metal droplet-slag surface area, and the duration of contact between the molten metal and slag are other factors that affect the final weld metal composition. The concentration of elements in the metal and slag is determined by the electrode wire and flux compositions and by the renewal of the slag pool (that is, the frequency with which the make-up flux is added to the pool).

Thermochemistry. In the case of slags containing high MnO and SiO_2, the reactions between the metal and slag can be described:

$$[FE] + (MNO) \rightarrow [MN] + (FEO) \qquad \textbf{(EQ 5)}$$

$$2[FE] + (SIO_2) \rightarrow [SI] + 2(FEO) \qquad \textbf{(EQ 6)}$$

In Eq 5 and 6, [*M*] and (*N*O) represent the activities of the metal *M* in the weld pool and of the oxide of element *N* in the slag layer. However, for all practical purposes, concentrations of the metallic element and oxide can be used instead of their activities in thermodynamic calculations. When electrodes containing high silicon are used, the reduction of manganese oxide by silicon must also be taken into account:

$$[SI] + 2(MNO) \rightarrow 2[MN] + (SIO_2) \qquad \textbf{(EQ 7)}$$

The lower the temperature, the more intense will be the reaction between manganese oxide and silicon. Changes in welding conditions, such as power input and temperature distribution, will have a definite effect on the recovery of alloying elements.

Like silicon, aluminum is often employed as a deoxidizer for steels and, to a lesser degree, as a microalloying element. The deoxidation reaction is:

$$2[AL] + 3(FEO) \rightarrow 3[FE] + (AL_2O_3) \qquad \textbf{(EQ 8)}$$

The reaction of aluminum with nitrogen is also important because it has a strong affinity for dissolved nitrogen in the weld pool and because aluminum nitride is responsible for the pinning of grain boundaries and grain growth control.

Titanium is widely used to alloy and modify steels. In welding, it is oxidized by the slag:

$$[TI] + 2(FEO) \rightarrow 2[FE] + (TIO_2) \qquad \textbf{(EQ 9)}$$

Similar to aluminum, titanium also reacts with carbon, nitrogen, and oxygen to form TiN, TiC, Ti(C,N), TiO, and TiO_2 particles that may remain trapped in the weld metal, moderating excessive grain growth.

The oxidation of carbon may take place as a result of carbon reacting with dissolved oxygen and metal oxides in the liquid metal:

$$[C] + [O] \rightarrow CO \qquad \textbf{(EQ 10)}$$

$$[C] + (MnO) \rightarrow CO + [Mn] \qquad \textbf{(EQ 11)}$$

$$2[C] + (SiO_2) \rightarrow 2CO + [Si] \qquad \textbf{(EQ 12)}$$

$$[C] + (FeO) \rightarrow CO + [Fe] \qquad \textbf{(EQ 13)}$$

Other minor constituents, such as vanadium and niobium, are oxidized by the slag at the expense of FeO to form a series of oxides, such as VO, V_2O_5, NbO, and Nb_2O_5.

When ESW is used to surface machine parts, metal-cored tubular wire electrodes are often used to provide high contents of alloy elements, such as chromium, molybdenum, vanadium, or titanium, in the deposited metal. When silicate fluxes are used, the oxidation of chromium takes place according to the following main reactions:

$$[Cr] + (MnO) = (CrO) + [Mn] \qquad \textbf{(EQ 14)}$$

$$[CR] + (FeO) \rightarrow (CrO) + [Fe] \qquad \textbf{(EQ 15)}$$

The welding consumables (fluxes and welding electrodes) and base metal are the major contributors of tramp elements, such as sulfur and phosphorus, to the weld pool. In addition, the temperature and welding conditions in ESW are not optimal for dephosphorization. As a result, increased migration of phosphorus from the slag to the weld pool occurs not only at the initiation of the weld but also at points where the electroslag process is resumed after an interruption. Sulfur in carbon steel welds is considerably more dangerous than phosphorus because it is the main cause of hot cracking. More recently, desulfurization slags with approximately 60 to 85 wt% CaF_2, and 15 to 40 wt% CaO have been adopted in ESW. Substantial desulfurization can also be achieved when carbon is added to a lime-fluorspar system.

There is also indication that in ESW the slag bath receives oxygen continuously from the surrounding atmosphere. With higher oxygen potential, the slag will also tend to oxidize the weld metal. Therefore, some modified electroslag processes include an inert gas cover on top of the slag layer to ensure a more deoxidized weld pool.

Electrochemistry. In addition to the thermochemistry of the process, electrochemistry plays an important role in ESW (Ref 17, 18). Because the anode and cathode reactions all occur in one pool, the weight of electrode melted is approximately equal to the weight recovered in the weld pool. Current efficiency can then be expressed as a function of electrode melted and element concentration change. Anodic current efficiency data show that the bulk of the current at the anode is carried by the oxidation reactions of iron, chromium, manganese, and silicon (Ref 19), which also indicates that electrochemical reactions account for a large fraction of the charges transferred. The major charge-carrying reactions at the cathode are the reductions of iron, aluminum, chromium, manganese, and silicon cations from the flux. Furthermore, the current efficiencies for electrochemical reactions involving chromium, manganese, silicon, molybdenum, oxygen, and aluminum increase with increasing current density. The concentration changes are directly proportional to the welding current and to the current efficiency for the reactions, and they are inversely proportional to the electrode melt rate.

Weld Metal Inclusions. Due to the almost "equilibrium" operating conditions of ESW as compared with other arc welding processes, inclusion separation is expected to be more complete. The larger volume of molten metal, less turbulent weld pool, slower cooling, and slower solidification rate in ESW contribute to a cleaner weld metal with relatively few submicron-size inclusions.

Solidification Structure. The morphology of steel electroslag weld deposits was first characterized and summarized by Paton (Fig. 12). Yu, Ann, Devletian, and Wood (Ref 20) further characterized the solidification substructure of low-carbon structural steel electroslag weldments and related them to the weld pool form factor:

- *GROUP 1:* TRANSITION FROM CELLULAR TO COLUMNAR DENDRITES OCCURS AT A SHORT DISTANCE FROM THE FUSION LINE. WELDS FROM THIS GROUP SHOW SHALLOW MOLTEN METAL POOLS WITH HIGH FORM FACTOR.
- *GROUP 2:* CELLULAR TO COLUMNAR DENDRITE TRANSITION OCCURS FURTHER AWAY FROM THE FUSION LINE, WITH A FINE COLUMNAR GRAIN ZONE AT THE CENTER OF THE WELD.
- *GROUP 3:* CHARACTERIZED BY A SHALLOW ANGLE OF INCLINATION (APPROXIMATELY 10°) BETWEEN THE COLUMNAR GRAINS AND THE FUSION LINE. THE TRANSITION FROM DENDRITIC TO COLUMNAR DENDRITES OCCURS CLOSE TO THE CENTER LINE OF THE WELD. THE GRAINS MEET AT THE CENTER AT AN OBTUSE ANGLE.
- *GROUP 4:* DENDRITIC TO COLUMNAR DENDRITE TRANSITION OCCURS CLOSE TO THE CENTER LINE, WITH EQUIAXED GRAINS AT THE CENTER OF THE WELD.

The deep pools and low form factor generally found in group 3 and group 4 welds are also responsible for the high sensitivity to centerline cracking and radial hot cracking of these welds.

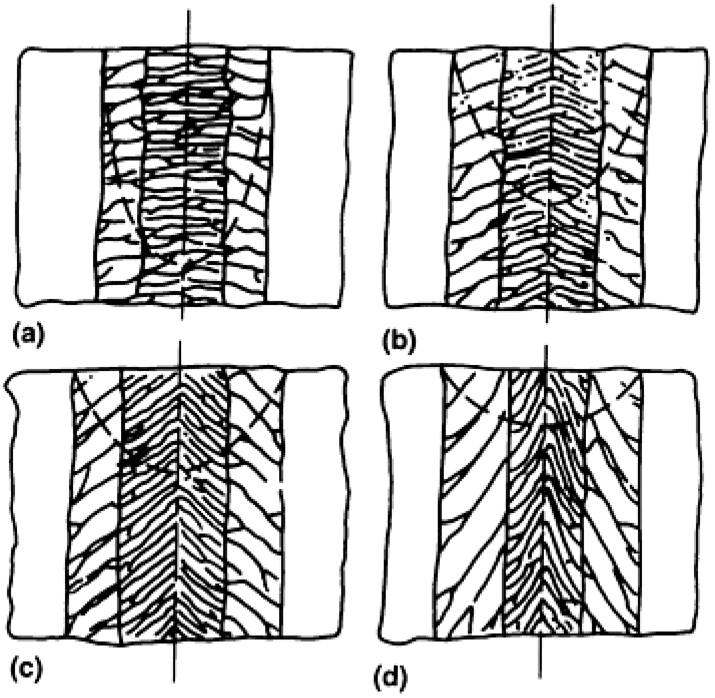

FIG. 12 ELECTROSLAG WELD METAL SOLIDIFICATION STRUCTURE ACCORDING TO THE VARIATION OF THE ORIENTATION AND THE THICKNESS OF THE COLUMNAR GRAINS ZONE. (A) GROUP 1. (B) GROUP 2. (C) GROUP 3. (D) GROUP 4. SEE TEXT FOR DETAILS. SOURCE: REF 3

Solid-State Transformations. The cooling rate of an electroslag weld metal is so much slower than those of other welding processes that reasonably coarse microstructures are found. During the austenite-to-ferrite transformation, primary ferrite and grain boundary ferrite are the principal constituents, along with carbides, mostly in the form of pearlite. In low-alloy steels, acicular ferrite, martensite, bainite, and retained austenite can be found only in those welds that have considerable amounts of alloying ingredients.

References cited in this section

3. B. PATON, *ELECTROSLAG WELDING,* AWS, 1962
17. A. MITCHELL AND G. BEYNON, ELECTROCHEMICAL REACTIONS IN THE ELECTROSLAG PROCESS, CH. 2, *BUR. MINES BULL.,* 1976, P 669
18. M.E. PEOVER, ELECTROSLAG REMELTING, A REVIEW OF ELECTRICAL AND ELECTROCHEMICAL ASPECTS, *J. INST. MET.,* VOL 100, 1972, P 97
19. R.H. FROST, D.L. OLSON, AND G.R. EDWARDS, "THE INFLUENCE OF ELECTROCHEMICAL REACTIONS ON THE CHEMISTRY OF THE ELECTROSLAG WELDING PROCESS, *MODELING OF CASTING AND WELDING PROCESSES II,* AIME, 1983, P 279-294
20. D. YU, H. ANN, J.H. DEVLETIAN, AND W.E. WOOD, SOLIDIFICATION STUDY OF NARROW-GAP ELECTROSLAG WELDING, *WELDING RESEARCH: THE STATE OF THE ART,* E. NIPPES AND D. BALL, ED., ASM, 1985, P 21-32

Electroslag and Electrogas Welding[*]

Electroslag Process Development

High-Productivity Electroslag Processes. Responding to the technological challenge of the need for a large volume of deposited metal and low heat input, many high-speed ESW techniques have been developed (Ref 9, 10, 11, 12, 15, 21, 23, 24, 25, 26). There have been essentially two approaches:

- NARROWING THE GAP TO DECREASE THE VOLUME OF METAL DEPOSIT (THEREBY SHORTENING THE TIME NEEDED TO FORM THE WELD JOINT)
- INCREASING THE WELDING SPEED

With either approach, the productivity is increased and the heat can be distributed throughout a longer section of the weld.

With respect to the narrow-gap approach, the extent of gap cross-sectional area reductions is limited by the stability of the process and the size of the electrode and guide tube. Too narrow a gap may cause arcing between the electrode and the base metal. Solid flux blocks may be inserted at regular intervals to avoid arcing between a solid strip electrode in a thin-channel consumable guide and the base metal. Gaps as small as 12 mm ($\frac{1}{2}$ in.) have been reported (Ref 22). Narrow-gap welding, however, can result in higher base metal dilution.

The use of powdered filler metal has been reported to increase the deposition rate in ESW (to over 24 kg/h, or 53 lb/h) with reduction in the specific heat input and to refine the weld metal and HAZ microstructure. The supply of metal powder (ferrous, mainly) to the weld metal can be obtained by a flow of argon gas (Fig. 13). Welding speed as high as 230 mm (9 in.) per minute has been reported. Notch impact testing results measured with samples taken 1 to 2 mm (0.04 to 0.08 in.) from the fusion line showed good Charpy impact toughness.

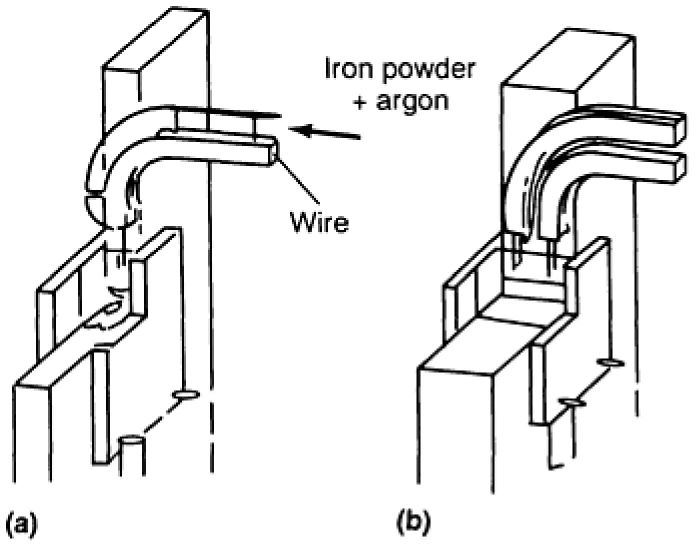

FIG. 13 EQUIPMENT SETUP FOR SQUARE-GROOVED HIGH-SPEED WELDING WITH METAL POWDER ADDITION. (A) PLATE THICKNESS OF 10 TO 40 MM ($\frac{3}{8}$ TO 1 $\frac{9}{16}$ IN.). (B) PLATE THICKNESS OF 40 TO 100 MM (1 $\frac{9}{16}$ TO 4 IN.). SOURCE: REF 25

Metal powder cored strip electrode for consumable guide ESW has also been used (Ref 9, 17). The average weld deposition rate is approximately two to three times higher than the solid wire process, over 38 kg/h (85 lb/h). The specific heat input of these welds was lower, at 0.8 kJ/mm^2 (520 kJ/in.^{2}). Rather than add metal powder to increase the deposition rate and absorb the excess thermal energy in the molten metal bath, a separate filler wire can be used (Ref 27). This is shown in Fig. 14. This setup resembles the feeding of a cold wire into plasma arc welding or gas-tungsten arc welding processes. Not only is the productivity of the process increased, but the HAZ width is also reduced.

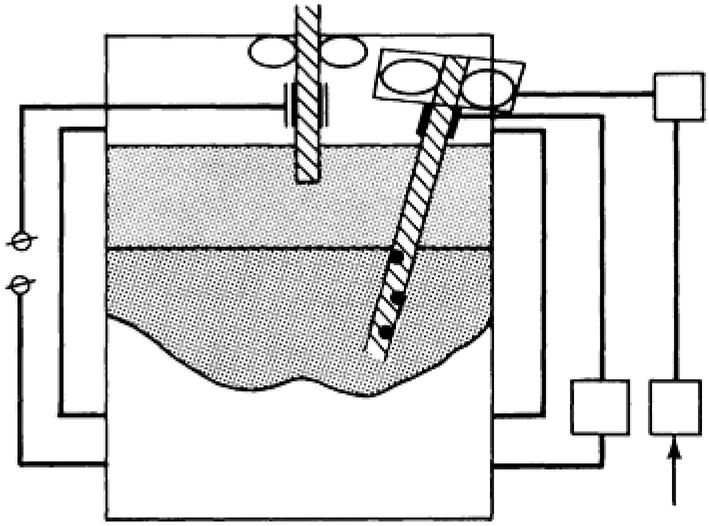

FIG. 14 SCHEMATIC OF ELECTROSLAG WELDING PROCESS USING SEPARATE FILLER WIRE TO INCREASE DEPOSITION RATE AND ABSORB EXCESS THERMAL ENERGY IN MOLTEN METAL BATH. SOURCE: REF 28

Electroslag Surfacing. Surfacing by means of weld overlays has a long history. Electroslag surfacing can be done in either the vertical or flat position. Where the area to be surfaced is small relative to the base metal, vertical electroslag surfacing is usually used. Adapting the vertical orientation of ESW to produce composite metallic structures involves replacing one of the parent metal plates with a water-cooled mold. An interesting application is tool steel ingots that are given a soft steel electroslag welded deposit on each end to facilitate the rolling operation.

For large areas, such as plates for tube sheets or cylindrical pressure vessels, electroslag surfacing is conducted in the flat position. Excellent process stability is one of the characteristics of the process. Even with electrodes as wide as 150 mm (6 in.), the quality of the surface weld is still comparable to that of welds produced by other processes. To minimize the undercutting problem and improve bead smoothness, Kawasaki Steel (Ref 6) developed the Maglay process, which utilizes an external magnetic field to control the flow of molten slag and metal. Dilution can also be controlled, to less than 10%. Forsberg (Ref 28) reports the development of a similar process, which uses strips of 60 to 90 mm (2.4 to 3.5 in.) width and emphasizes the higher deposition rate and lower dilution ratio.

References cited in this section

6. "ELECTROSLAG ELECTROGAS TIPS & TECHNIQUES," BULLETIN OF ELECTROTHERM CORP., 1973
9. S. LIU AND C.T. SU, GRAIN REFINEMENT IN ELECTROSLAG WELDMENTS BY METAL POWDER ADDITION, *WELD. J.,* VOL 68 (NO. 4), 1989, P 132S-144S
10. C.T. SU, "HIGH SPEED ELECTROSLAG WELDING," M.S. THESIS, PENNSYLVANIA STATE UNIV., 1987
11. S. VENKATARAMAN, J.H. DEVLETIAN, W.E. WOOD, AND D.G. ATTERIDGE, GRAIN REFINEMENT DEPENDENCE ON SOLIDIFICATION AND SOLID STATE REACTIONS IN ELECTROSLAG WELDS, *GRAIN REFINEMENT IN CASTINGS AND WELDS,* AIME, 1982, P 275-288
12. D.G. ATTERIDGE, S. VENKATARAMAN, AND W.E. WOOD, "IMPROVING THE RELIABILITY AND INTEGRITY OF CONSUMABLE GUIDE ELECTROSLAG WELDMENTS IN BRIDGE

STRUCTURES," RESEARCH REPORT, OREGON GRADUATE CENTER, 1982

15. R.D. THOMAS, JR., CORROSION RESISTANT WELD OVERLAYS BY THE DUAL STRIP PROCESS, *BRIT. WELD. J.,* MAY 1966
17. A. MITCHELL AND G. BEYNON, ELECTROCHEMICAL REACTIONS IN THE ELECTROSLAG PROCESS, CH. 2, *BUR. MINES BULL.,* 1976, P 669
21. V.I. AVRAMENKO, B.F. LEBEDEV, AND V.I. BOZHKO, SOME WAYS OF INCREASING THE PRODUCTIVITY OF ELECTROSLAG WELDING, *SVAR. PROIZVOD.,* VOL 10, 1973, P 16
22. K. WATANABE, I. SEJIMA, S. KOKURA, G. TAKI, AND H. MIYAKE, PROBLEMS AND IMPROVEMENT OF LARGE HEAT INPUT ELECTROSLAG WELDING, *J. JPN. WELD. SOC.,* 1975, P 519
23. S.A. SMIRNOV AND L.A. EFIMENKO, SPECIAL STRUCTURAL FEATURES AND MECHANICAL PROPERTIES OF ELECTROSLAG WELDED JOINTS MADE USING POWDERED FILLER METAL, *SVT. SVARKA,* VOL 9, 1973, P 46
24. F. EICHHORN AND J. REMMEL, SITUATION OF RESEARCH IN ELECTROSLAG WELDING--A TENDENCY OF FURTHER DEVELOPMENT, *IND. WELD. J.,* VOL 4, 1983, P 37
25. F. EICHHORN, J. REMMEL, AND B. WUBBELS, HIGH SPEED ELECTROSLAG WELDING, *WELD. J.,* VOL 62 (NO. 1), 1984, P 37
26. F. EICHHORN AND J. REMMEL, EFFICIENT FILLET WELDING IN THE VERTICAL WELDING POSITION WITH ELECTROGAS AND ELECTROSLAG WELDING METHODS, DOC. XII-908-85, IIW, 1985
27. B.F. YAKUSHIN *ET AL.,* IMPROVING THE CAPACITY OF ELECTROSLAG WELDED JOINTS FOR RESISTING HOT CRACKING, *WELD. RES. ABROAD,* JUNE/JULY 1984 (TRANSLATION FROM RUSSIAN BY AUTOMATIC WELDING, OCT 1982)
28. S. FORSBERG, RESISTANCE ELECTROSLAG (RES) SURFACING, *WELD. J.,* VOL 63 (NO. 8), 1985, P 41-48

Electroslag and Electrogas Welding*

S. Liu, Colorado School of Mines; S.D. Brandi, Escola de Politecnica da USP; R.D. Thomas, Jr., R.D. Thomas and Company

Electrogas Welding

Electrogas welding (EGW) can be classified as a gas-metal arc welding process. It was introduced in the early 1960s to perform single-pass welding of relatively thin plates in the vertical position. As shown in Fig. 15, the process is very similar to ESW, with the exception of the presence of an arc and the weld pool shielding mechanism (Ref 6). In most cases, shielding gas is used together with a bare or flux-cored wire to provide a shielding gas on top of the molten pool of slag and weld metal. Self-shielded flux-cored wires are often used without a separate gas shield. The production rate can be as high as 34 kg/h (75 lb/h), depending on the joint configuration and plate size. Special flux-cored wires that contain fewer slagging ingredients must be formulated for EGW, because slag buildup on the molten weld metal surface can affect arc stability and cause slag entrapment. The thicker the plate being welded, the less flux needed. When bare wire is used, small amounts of granular flux are usually added at the start. The slag needed for protection is produced from the deoxidizers in the electrode wire. For steel welding, the most common shielding gas is carbon dioxide. Additions of argon (75 to 80 vol%) to carbon dioxide seem to improve the arc stability and weld metal properties. Flow rates are usually of the order of 14 to 19 L/min (30 to 40 ft^3/h).

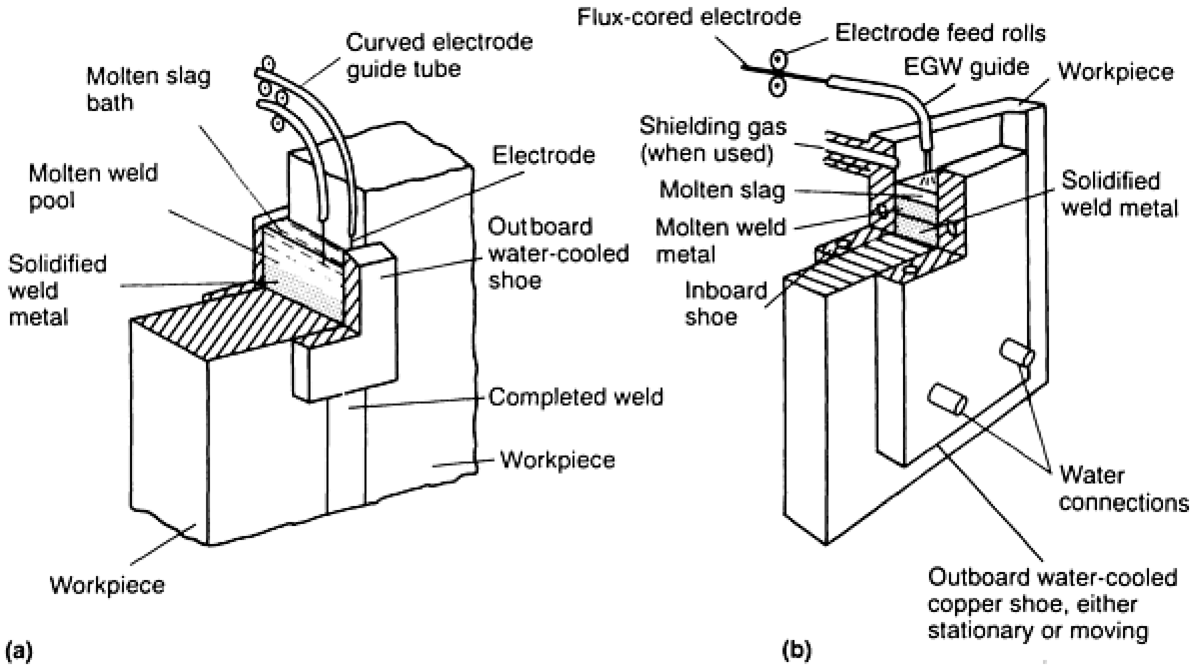
Curved electrode guide tube
Molten slag bath
Molten weld pool
Solidified weld metal
Electrode
Outboard water-cooled shoe
Completed weld
Workpiece
Workpiece
(a)
Flux-cored electrode
Electrode feed rolls
EGW guide
Workpiece
Shielding gas (when used)
Molten slag
Molten weld metal
Inboard shoe
Solidified weld metal
Water connections
Outboard water-cooled copper shoe, either stationary or moving
(b)

FIG. 15 SCHEMATICS COMPARING PRIMARY COMPONENTS OF TWO VERTICAL WELDING PROCESSES IN WHICH MOLTEN WELD POOLS ARE CONFINED BY COOLING SHOES. (A) ELECTROSLAG WELDING. (B) ELECTROGAS WELDING

Heavy aluminum weldments are produced by EGW using argon-helium gas shielding with flow rates of 28 to 38 L/min (60 to 80 ft^3/h).

In addition to its higher deposition rate, EGW presents the advantages of lower heat input, more refined microstructure, and improved HAZ properties. Consequently, postweld heat treatment is not mandatory in many EGW applications. The process also offers easy restarting if the weld is interrupted (complicated in ESW). However, as the thickness of a plate increases, the shielding gas often fails to cover the entire joint area, leading to unacceptable porosity. Thick plates favor the choice of ESW because electroslag welds exhibit fewer inclusions and defects.

Multipass ESW and EGW. The metallurgical problems encountered in single-pass electroslag and electrogas processes can be largely overcome by multipass welding. An example is a thick steel square-butt electroslag weld joint with a fixed copper chill bar or water-cooled copper tube inserted at midwall, and a weld being produced with a moving shoe on one side. The copper chill is then removed, and the opposite side is welded. In practice, however, several problems may appear. One is related to the removal of the copper chill, gripped by the shrinkage that occurs during the first weld pass. Another is the problem of ensuring penetration by the second pass into the first pass. Unless the plate edges are machined and carefully fitted up, slag leakage may also occur.

Reference 29 reports the use of a copper chill bar that penetrated to the root of one side of a double V-joint preparation while EGW proceeded from the opposite side. An alternative method is to manually deposit the root pass, to avoid the problem of close fit-up tolerances needed for complete penetration. Where welding can only be approached from one side, conventional backing bars are used, and the copper shoe protrudes into the joint to mold the initial pass, followed by the surface pass deposited in the conventional manner. This technique is satisfactory for EGW, but slag retention in the cavity formed by the protruding shoe and the base plates is a challenge in the electroslag process.

The consumable nozzle electroslag process can be used to make a multipass electroslag weld with a pass from each side. A tightly fitted spacer bar is placed at midwall. The electroslag weld pass from the first side partially penetrates the spacer bar, and the second pass from the opposite side consumes the remaining spacer and penetrates the first pass.

These process modifications do not solve the problem of the wide HAZ on the surface of the plates being welded, because toughness and fatigue properties at or near the outer surface of a weld are essential for the integrity and performance of the structure.

References cited in this section

6. "ELECTROSLAG ELECTROGAS TIPS & TECHNIQUES," BULLETIN OF ELECTROTHERM CORP., 1973
29. P.C. ARNOLD AND D.C. BERTOSSA, MULTIPLE PASS AUTOMATIC VERTICAL WELDING, *WELD. J.,* VOL 45 (NO. 8), 1966, P 651-660

Electroslag and Electrogas Welding*

S. Liu, Colorado School of Mines; S.D. Brandi, Escola de Politecnica da USP; R.D. Thomas, Jr., R.D. Thomas and Company

---

Electroslag and Electrogas Process Applications

Carbon and Low-Alloy Steels. The joining of heavy-section steels is the most common application of both ESW and EGW. As previously stated, ESW is commonly done on plates of thickness 50 mm (2 in.) and greater. In fact, economy is greatly increased if the section thickness is greater than 100 mm (4 in.). The range of application of EGW is between 9.5 mm ($\frac{3}{8}$ in.) and 75 mm (3 in.). When compared with arc welding processes, minimum distortion, vertical position, and minimum joint penetration are the major advantages. Typical applications include fabrication of pressure vessels, nuclear components, power generation equipment, rolling mills, heavy presses, bridges, ships, and oil drilling rigs (Ref 30). Other structural applications reported are blast furnace shells (carbon steel) and a wind tunnel structure (HY-100 steel) (Ref 31, 32).

Structural Steels. A considerable number of highway bridges were fabricated using ESW prior to its ban in bridge construction. Some failures have been attributed to defects such as hydrogen cracking, lack of fusion, low HAZ fatigue properties, and so on. However, properly made welds were found to meet the radiographic standards required for impact and fatigue loading of structural members (Ref 33). In the case of low-alloy steels for structural applications, austenitizing postweld heat treatments have been considered necessary for electroslag welds. However, when proper attention is given to consumable (flux and electrode) selection, a stress-relieving heat treatment is generally sufficient to obtain good impact properties (Ref 34).

Shipbuilding. Electroslag and electrogas welding can be used for the welding of components and for on-ship welding. One application is the propeller shaft bracket assembly for natural gas and oil tankers. Stem frame parts, heavy rudderstock plates, longitudinal and vertical hull stiffeners, and shaft struts are often electroslag welded (Ref 35).

Pressure Vessels. Thick-wall pressure vessels used in the chemical, power generating, petroleum, and marine industries are manufactured using ESW. Plates are rolled to form the shell of the pressure vessel, and the longitudinal seam created is then welded. Lifting lugs on the vessels, nozzles, and branch pipes are also electroslag welded to thick-walled vessels. Due to the high heat input, the welding practice must be critically controlled. Where code requirements are specified, a normalizing heat treatment is also performed.

Penetrating members, such as nozzles and pipe fittings, are often costly to weld into thick-wall pressure vessels by conventional processes. A unique method employing both welding and casting techniques has been proposed by Norcross (Ref 36). A cylindrical water-cooled mold is tightly affixed to the outside surface of the vessel, and weld metal is cast in place. The hole of the nozzle is then machined to provide an opening into the vessel, at the end of which pipes or other attachments can be easily welded (Fig. 16). Even large-diameter penetrating members can be made by this method, by casting hollow cylinders after first tapering the vessel wall to allow the starting zone of the electroslag weld to be machined away (Ref 38, 39).

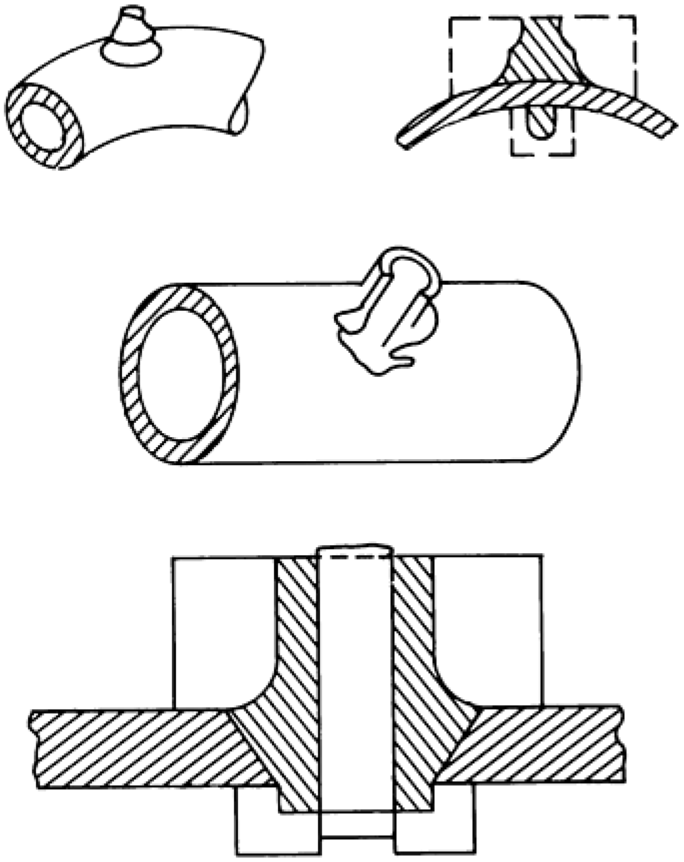

FIG. 16 APPLICATION OF ELECTROSLAG WELDING TO INCORPORATE PENETRATING MEMBERS SUCH AS NOZZLES ONTO THICK-WALL PRESSURE VESSELS. SOURCE: REF 39

Heavy Machinery. Large presses and machine tools are often manufactured from plates that are larger than the mills can produce, so ESW is used to join large plates together. Motor frames, press frames, gear blanks, turbine rings, crane rails, and crusher bodies are some of the common applications. In the transmission pipeline industries, ball valves and stop-gate valves are often electroslag welded.

Dissimilar Metal Joining. Electroslag remelting has been used to develop transition pieces for heavy-wall steam piping where low-alloy chromium-molybdenum steel pipe is to be joined to austenitic stainless steel pipe. Two round ingots, one of the ferritic alloy and one of the austenitic stainless steel, are welded together to form the electrode for electroslag remelting. As the electrode is melted in the molten slag bath, the composition gradually changes from one to the other, depositing a graded alloy billet which is later pierced and formed into a transition piece (Ref 37, 39, 40).

Joining and Repairing of Castings. Electroslag welding is also used to weld cast components. Instead of producing a large casting, many small castings of lower cost and simpler fabrication are produced and welded together. One such application is described by Brosholen, Skaug, and Vesser (Ref 41) to construct a cast steel propeller shaft bracket. Cast iron has been electroslag welded using a cored wire containing either white or gray cast iron powders and a graphite nozzle (Ref 42). Defective thick sections of cast steels can be repaired by ESW, by drilling a series of connecting holes into the defect zone. A close-fitting water-cooled copper mold is inserted in one of the drilled holes while welding proceeds in the adjacent hole. On completion of that weld, the mold is moved to the next hole before the molten slag has cooled, which allows ESW to start in the vacated hole (Ref 43, 44).

Tool and Die Surfacing and Welding. The slow thermal cycles involved in ESW are favorable for depositing hardenable alloys. For large forging hammers and forming dies, hard wear-resisting alloys can be spilled to the surface of a carbon steel backing using multiple electrodes and current pulsing techniques (Ref 45, 46, 47).

Stainless Steel and Nickel-Base Alloys. Two halves of large 38 tonne (42 ton) cast stainless steel pumps for nuclear electric power facilities have been successfully joined by circumferential welding of their equatorial surfaces and have been qualified to rigid nuclear code requirements. As-welded electroslag welds in austenitic stainless steels were found to be insensitive to knife-line attack after sensitization for 1 to 5000 h at temperatures ranging from 450 to 750 °C (842 to 1382 °F) in tests (Ref 48). It has also been demonstrated that ESW is suitable for welding 25 mm (1 in.) thick alloy 600 plates and 25 mm (1 in.) and 114 mm ($4\frac{1}{2}$ in.) thick plates of alloy 800. Matching filler metals and Inconel filler-metal 82 produced sound welds with excellent elevated-temperature strength and stress-rupture results. Because stainless steels and nickel-base alloys do not undergo allotropic transformations, electroslag welds do not require a high-temperature postweld heat treatment.

Aluminum. Both ESW and EGW have been used to join thick sections of aluminum. One such application is to join 240 mm (9.5 in.) thick electrical conductors (busbars). Instead of the typical copper shoes, graphite shoes are used to obtain satisfactory weld surfaces and edge penetration (Ref 49). Gagan *et al.* (Ref 50) reported erratic penetration encountered in ESW of thick-section aluminum due to the magnetic fields generated by the welding process, which required special shielding screens around the weld zone. The high conductivity of aluminum and the prompt formation of a refractive oxide are also of concern. High-fluoride-content fluxes and an inert gas shield in EGW typically produce good results in welding thick-section aluminum. For ESW of aluminum containing 4.5 wt% Mg, the following flux has been suggested (Ref 51):

CONSTITUENT	CONTENT, WT%
MGF$_2$	1-30
MGCL$_2$	5-60
LIF	5-60
KCL	5-6

As welding progresses, additions are made to the flux containing 30 wt% each of KCl, LiF, and MgCl$_2$, and 10 wt% MgF$_2$. This flux is reported to work well for ESW of aluminum with copper shoes (Ref 51).

Titanium. The reactivity of titanium requires that residual elements such as hydrogen, oxygen, and nitrogen be minimized in ESW. A slag bath high in halides and containing virtually no oxides, and an inert gas (argon) shield over the bath, are required. Plate electrodes seem to produce better results (Ref 3, 52). Commonly, fluxes are halide compounds based on CaF$_2$ with ram-earth additions to produce welds low in oxygen, nitrogen, and hydrogen. Devletian (Ref 53) reported that because the high resistivity of titanium promotes rapid ohmic heating of the titanium consumables, stable welding operation generally requires large-diameter electrodes for nonconsumable guide welding, and nozzles or guide plates for consumable guide welding.

References cited in this section

3. B. PATON, *ELECTROSLAG WELDING,* AWS, 1962
30. D.J. ELLIS AND A.F. GIFFORD, APPLICATION OF ELECTRO-SLAG AND CONSUMABLE GUIDE WELDING--PARTS 1-6, *MET. FABR.*, APRIL-NOV 1973
31. H. PASS, THE ELECTRO-SLAG WELDING OF A BLAST FURNACE HEARTH JACKET, *WELD. MET. FABR.,* JAN/FEB 1976
32. ELECTROSLAG WELDING HELPS TO RESTORE WIND TUNNEL, *MET. PROGR.*, MARCH 1976
33. J.S. NOEL AND A.A. TOPRAC, STATIC, FATIGUE, AND IMPACT STRENGTH OF ELECTROSLAG WELDMENTS, RESEARCH REPORT 157-1F, RESEARCH PROJECT 3-5-71-157, CENTER FOR HIGHWAY RESEARCH, UNIV. OF TEXAS, 1971
34. A.M. MAKARA *ET AL.,* ELECTROSLAG WELDING OF STRUCTURAL STEELS WITHOUT NORMALIZING, *AVTOM. SVARKA,* VOL 7, 1974, P 511-514
35. B. HOWSER, APPLICATION AND TRENDS IN ELECTROSLAG WELDING IN THE UNITED STATES, *ELECTROSLAG WELDING FOR MARINE APPLICATIONS*, U.S. NAVAL ACADEMY,

MARCH 1985, P 4-1
36. J.E. NORCROSS, ELECTROSLAG DEVELOPMENTS, *FWP J.,* VOL 18 (NO. 2), 1978, P 41-58
37. D. YAPP AND A.P. BENNETT, DEVELOPMENT OF ELECTROSLAG-MELTED GRADED TRANSITION JOINTS, *WELDING RESEARCH RELATED TO POWER PLANT,* PAPER 34, CENTRAL ELECTRICITY GENERATING BOARD, SOUTHAMPTON, ENGLAND, SEPT 1972, P 464-481
38. J.E. NORCROSS, ELECTROSLAG BOSS AND PROCESS, U.S. PATENT 4,130,931, DEC 1978
39. ELECTROSLAG CASTING AND WELDING, BRITISH PATENT 1,390,674, APRIL 1975
40. A.P. BENNETT, AN ELECTRICAL ANALOGUE OF AN ELECTROSLAG REMELTING OPERATION USED AS A GUIDE TO IMPROVED CONTROL OF THE PROCESS, *MET. MATER.,* APRIL 1972, P 146-149
41. A. BROSHOLEN, E. SKAUD, AND J.J. VESSER, ELECTROSLAG WELDING OF LARGE CASTINGS FOR SHIP CONSTRUCTION, *WELD. J.,* VOL 56 (NO. 8), 1977, P 26-30
42. Y. ISHII, ELECTROSLAG WELDING OF CAST IRON, *TRANS. JAP. WELD. SOC.,* NO. 2, SEPT 1970, P 241-252
43. V.F. YAKOVLEV, P.I. KOVALKIN, N.I. EVDOKIMOV, M.G. KOZULIN, AND I.I. SUSHCHUCK-SLYUSARENKO, ELECTROSLAG WELDING OF STEEL CASTING DEFECTS, *AVTOM. SVARKA,* VOL 2, 1970, P 72
44. V.G. SVYNYUK, ELECTROSLAG WELDING FOR RECTIFYING DEFECTS IN CAST CENTRIFUGAL PUMP WHEELS, *AUTOM. WELD.,* VOL 28 (NO. 7), 1975, P 57-58
45. V.T. ARSENKIN, V.G. RADCHENKO, AND D.M. LIKHOHSHERSTOV, THE ELECTROSLAG DEPOSITION OF DIE STEEL ON CARBON STEEL, *AVTOM. SVARKA,* VOL 3, 1976, P 46-49
46. M.P. IVANAN, *ET AL.,* ELECTROSLAG HARDFACING OF BEATERS OF HAMMER CRUSHERS, *SVAR. PROIZVOD.,* VOL 7, 1974, P 48
47. Y.I. GORBACHEV, E.A. KOVALEVSKII, AND A.V. MATELEV, ELECTROSLAG WELDING OF DIE STEEL, *SVAR. PROIZVOD.,* VOL 11, 1984, P 12-13
48. N. LEHKA, PROBLEMS ASSOCIATED WITH THE CORROSION OF THE PROPERTIES OF ELECTROSLAG WELDED JOINTS IN AUSTENITIC STEELS OF INCREASED STRENGTH, *ZVRACSKE SPRAVY/WELDING NEWS,* VOL 26 (NO. 3), 1976, P 61-68
49. P.M. BARTLE, DEVELOPMENT OF AN ELECTROGAS TECHNIQUE FOR WELDING THICK ALUMINUM, *WELD. INST. RES. BULL.,* VOL 12 (NO. 3), 1971, P 67-71
50. Y.G. GAGEN, *ET AL.,* ELECTROSLAG WELDING OF ALUMINUM BUS-BARS LOCATED IN THE ZONE OF A STRONG MAGNETIC FIELD, *SVAR. PROIZVOD.,* VOL 12, 1972, P 19-21
51. ELECTROSLAG OR SUBMERGED ARC WELDING OF ALUMINIUM OR ITS ALLOYS OR ELECTROSLAG REMELTING ALUMINUM, BRITISH PATENT 1,544,248, APRIL 1979
52. V.Y. MALIN, ELECTROSLAG WELDING OF TITANIUM AND ITS ALLOYS, *WELD. J.,* VOL 64, FEB 1985, P 42-49
53. J.H. DEVLETIAN, ELECTROSLAG WELDING OF NON-FERROUS METALS--A REVIEW, *ELECTROSLAG WELDING FOR MARINE APPLICATIONS,* U.S. NAVAL ACADEMY, 1985, P 3-1 TO 3-32

Electroslag and Electrogas Welding[*]

S. Liu, Colorado School of Mines; S.D. Brandi, Escola de Politecnica da USP; R.D. Thomas, Jr., R.D. Thomas and Company

---

Problems and Quality Control

Fusion Zone. As discussed earlier in the section "Solidification Structure" of this article, both cellular and columnar dendrites are observed in the solidification structure of an electroslag weldment. Particularly in the cases of welds with low form factor, in which the transition from cellular to columnar dendrites occurs close to the centerline of the weld and the grains meet at an obtuse angle, centerline cracking and radial hot cracking are more frequently observed. These defects can be attributed to the combined effects of temperature gradient, solidification rate, degree of restraint in fit-up of the weld, high welding speed, and low form factor. In general, a long, straight-sided columnar grain structure (the result of high welding speed) tends to be weaker under load than the more equiaxed and finer grain structure of a slow-speed weld. At the same time, the cellular structure may be coarser and higher in segregation for the case of low welding speed. Figure 17 shows a proposed mechanism of solidification cracking.

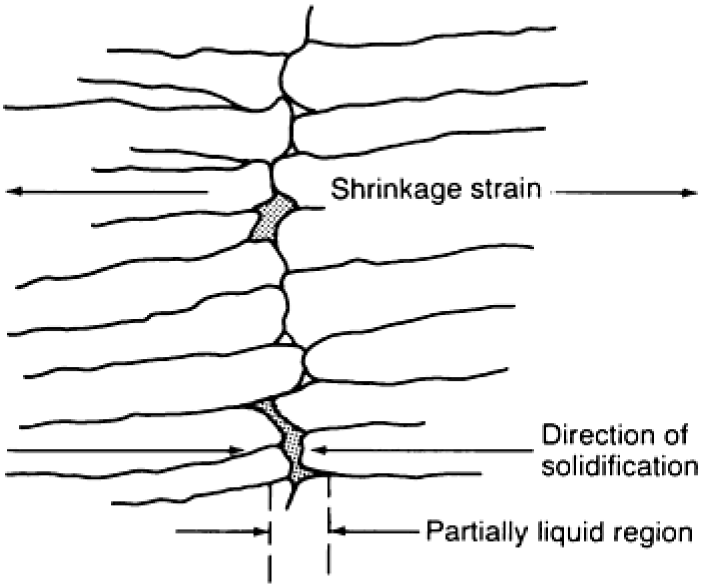

FIG. 17 SCHEMATIC INDICATING PROJECTED MECHANISM OF SOLIDIFICATION CRACKING IN AN ELECTROSLAG WELDMENT. SOURCE: REF 54, 55

Partially Melted Zone. Liquation cracking is associated with the melting of heavily segregated grain boundaries near the fusion line region. Carbon and manganese are the most common alloying elements involved in the formation of compounds such as (Mn,Fe)S that lower the melting temperature of the grain boundaries. Phosphorus, nitrogen, and boron are some of the other embrittling agents found in ferrous alloys. During cooling, residual stresses may build up and rupture these weakened boundaries. Under the slow heating and cooling conditions experienced by an electroslag weld, the susceptibility of a ferrous alloy to liquation cracking is increased. It has been reported that liquation cracks can be eliminated in heavy electroslag welds by microalloying the steels with cerium, titanium, and other elements to combine with sulfur and form small high-melting-point inclusions (Ref 56).

Temper Embrittlement. Thick-section $2\frac{1}{4}$ Cr-1Mo steel plates used in pressure vessels for petroleum and chemical service at elevated temperatures are susceptible to temper embrittlement in service. The problem has been attributed to the presence of residual elements such as phosphorus and antimony. Bruscato (Ref 55) established an *X* factor to describe the effect of the residual elements:

$$X = \frac{10P + Sb + 4Sn + As}{100} \qquad \textbf{(EQ 16)}$$

where *X* is in parts per million. Acceptable weld metal ductility can be obtained if the manganese and silicon contents and the *X* factor are low. Reduction of these elements can be achieved using basic fluxes. Another indicator, the Watanabe number, *J*, shows that welds and base metal are not susceptible to temper embrittlement if *J* is less than 200 (Ref 57):

$$J = (\mathrm{SI} + \mathrm{MN})\,(\mathrm{P} + \mathrm{SN}) \times 10^4 \qquad \textbf{(EQ 17)}$$

Hydrogen Cracking. Generally speaking, ESW is performed under perfect slag shielding conditions, and the rate of cooling of the weld metal is low. Therefore, the occurrence of hydrogen-induced cracking is minimal. However, under circumstances that require the use of a moist argillaceous material to prevent the slag from leaking out of the joint region, an atmosphere high in water vapor is present. Microcracks, and sometimes blowholes, may result in the weld metal. Grain boundary separation, observed in some electroslag welds, can be completely eliminated by postweld heat treating at 300 °C (570 °F) immediately after welding, which indicates that by allowing hydrogen to diffuse out of the specimen, the integrity of the welds can be maintained. This effect is further evidenced by the higher cracking incidence at the bottom part of an electroslag weld. In addition, diffusible hydrogen content is high at a region close to the weld start, decreasing to a relatively constant value at short distances from the starting point. By increasing the depth of the pool, the cooling rate is reduced, and hydrogen escapes by diffusion before it causes damage (Ref 58).

Weld Distortion. When compared with welds produced using other processes, electroslag slag welds do not present significant distortion problems. Measurements show that both transverse and angular distortion are present and that the gap distance changes as welding progresses. However, the distortion is only of the order of 1 to 2%. Angular distortion occurs because of the rapid cooling of the welded portion of the joint, which tends to draw the parts together, reducing the gap distance. Material ahead of the weld is separated by a gap, so heating has no effect. Material in the weld is either liquid or so hot that it is soft. Neither can support much of a load. As the weld progresses, the very bottom of the weld (coldest metal) resists motion. A contraction of 3 mm ($\frac{1}{8}$ in.) at the top of a weld 1.3 m (4.3 ft) long is normal. This separation must be allowed during joint setup to avoid jamming the guides and disrupting the normal progress of welding. Parts to be joined must be fitted up so that the gap is wider at the top of the joint. For improved dimensional accuracy, welding must be done at the maximum travel speed attainable.

Postweld Heat Treatment. As discussed in the previous section "Solidification Structure" of this article, most of the as-welded electroslag weldments have very coarse as-cast structure in the weld metal and coarse grains in the HAZ. As a result, such electroslag weldments may not qualify for many critical applications. For example, in the case of ferritic steels, normalizing is generally required to refine the weld metal and HAZ structure. In the case of electroslag welded nuclear transport flasks, normalizing is frequently followed by tempering (Ref 59). In some low-alloy steels, subcritical postweld heat treatments (for example, stress relief) may not be useful because they can be either detrimental or harmful to mechanical properties, particularly notch toughness. As an example, stress relief $2\frac{1}{4}$ Cr-1Mo steel weldments at 760 °C (1400 °F) significantly increased the fusion zone and HAZ toughness, but they coarsened the carbides in the base metal, reducing its toughness (Ref 60). Quenched and tempered material, when joined by ESW, must be heat-treated after welding to obtain adequate mechanical strength in the weld and HAZ. Kapadia (Ref 61) examined the need for stress-relieving components that will be exposed to fatigue loading and determined that in butt welds in the as-welded condition the outside surfaces cool and contract before the center, and thus compressive residual stresses are present, which improves fatigue life. Stress-relief heat treatments actually reduce this beneficial effect.

References cited in this section

54. K. EASTERLING, *INTRODUCTION TO THE PHYSICAL METALLURGY OF WELDING*, BUTTERWORTHS, 1983
55. R. BRUSCATO, TEMPER EMBRITTLEMENT AND CREEP EMBRITTLEMENT OF 2-1/4CR 1MO SHIELDED METAL ARC WELD DEPOSITS, *WELD. J.*, VOL 9 (NO. 4), 1970, P 148S-156S
56. A.M. MAKARA, Y.Y. KOKALEV, AND I.V. NOVIKOV, LIQUIDATION CRACKS IN THE HAZ IN STRUCTURAL STEELS DURING ELECTROSLAG WELDING, *AVTOM. SVARKA*, VOL 5, 1972, P 1-5
57. Y. MURAKAMI, T. NAMURA, AND J. WATANABE, HEAVY SECTION 2-1/4CR-1MO STEEL FOR HYDROGENATION REACTORS, STP 755, ASTM, 1982, P 383-417
58. T. KUNIHIRO AND H. NAKAJIMA, MICRO-CRACKING IN CONSUMABLE-NOZZLE ELECTROSLAG WELD METAL, *SIGNIFICANCE OF DEFECTS IN WELDED STRUCTURES*, UNIV. TOKYO PRESS, 1973, P 105-109
59. S.S. TULIANI, K.S. PROBERT, AND A.H. BRISCOE, FABRICATION OF ELECTROSLAG WELDED MAGNOX FUEL TRANSPORT FLASKS, *WELDING AND FABRICATION IN THE NUCLEAR INDUSTRY*, 1979, P 327-333

60. O. SERRANO, G.E. EDWARDS, AND R.H. FROST, A COMPARISON OF AS-WELDED AND STRESS-RELIEVED 2-1/4CR-1MO STEEL ELECTROSLAG WELDMENTS: MICROSTRUCTURAL AND PROPERTIES, *THE APPLICATION OF 2-1/4CR-1MO STEEL FOR THICK WALL PRESSURE VESSELS,* ASTM, 1980
61. B.M. KAPADIA, INFLUENCE OF RESIDUAL STRESS ON FATIGUE CRACK PROPAGATION IN ELECTROSLAG WELDS, *FATIGUE TESTING OF WELDMENTS,* STP 648, ASTM, 1977, P 244-260

Electroslag and Electrogas Welding*

S. Liu, Colorado School of Mines; S.D. Brandi, Escola de Politecnica da USP; R.D. Thomas, Jr., R.D. Thomas and Company

---

References

1. ELECTROSLAG AND ELECTROGAS WELDING, *AWS WELDING HANDBOOK,* VOL 2 (NO. 7), AWS, 1978, P 226-260
2. H.C. CAMPBELL, ELECTROSLAG, ELECTROGAS, AND RELATED WELDING PROCESSES, *WRC BULL.,* VOL 154, 1970
3. B. PATON, *ELECTROSLAG WELDING,* AWS, 1962
4. T. DEBROY, J. SZEKELY, AND T. EAGER, HEAT GENERATION PATTERNS AND TEMPERATURE PROFILES IN ELECTROSLAG WELDING, *METALL. TRANS. B,* VOL 11, 1980, P 593-605
5. R.H. FROST, J.E. JONES, AND D.L. OLSON, ELECTROSLAG WELDING OF PRESSURE VESSEL STEELS, *ELECTROSLAG WELDING FOR MARINE APPLICATIONS,* PAPER 2, U.S. NAVY ACADEMY, MARCH 1985
6. "ELECTROSLAG ELECTROGAS TIPS & TECHNIQUES," BULLETIN OF ELECTROTHERM CORP., 1973
7. J.E. JONES, D.L. OLSON, AND G.P. MARTINAS, METALLURGICAL AND THERMAL CHARACTERISTICS OF NON-VERTICAL ELECTROSLAG WELDING, *WELD. J.,* VOL 59 (NO. 9), 1980, P 245S-254S
8. S. NAKANO, N. NISHIYAMA, T. HIRO, AND J. TSUBOI, MAGLAY PROCESS--ELECTROMAGNETIC CONTROLLED OVERLAY WELDING PROCESS WITH ESW, *KAWASAKI STEEL TECH. REP.,* VOL 2, 1981, P 31-42
9. S. LIU AND C.T. SU, GRAIN REFINEMENT IN ELECTROSLAG WELDMENTS BY METAL POWDER ADDITION, *WELD. J.,* VOL 68 (NO. 4), 1989, P 132S-144S
10. C.T. SU, "HIGH SPEED ELECTROSLAG WELDING," M.S. THESIS, PENNSYLVANIA STATE UNIV., 1987
11. S. VENKATARAMAN, J.H. DEVLETIAN, W.E. WOOD, AND D.G. ATTERIDGE, GRAIN REFINEMENT DEPENDENCE ON SOLIDIFICATION AND SOLID STATE REACTIONS IN ELECTROSLAG WELDS, *GRAIN REFINEMENT IN CASTINGS AND WELDS,* AIME, 1982, P 275-288
12. D.G. ATTERIDGE, S. VENKATARAMAN, AND W.E. WOOD, "IMPROVING THE RELIABILITY AND INTEGRITY OF CONSUMABLE GUIDE ELECTROSLAG WELDMENTS IN BRIDGE STRUCTURES," RESEARCH REPORT, OREGON GRADUATE CENTER, 1982
13. R.H. FROST, G.R. EDWARDS, AND M.D. RHEINLANDER, A CONSTITUTIVE EQUATION FOR THE CRITICAL ENERGY INPUT DURING ELECTROSLAG WELDING, *WELD. J.,* VOL 60 (NO. 1), 1981, P 12-62
14. S. LIU AND C.T. SU, PERFORMANCE EVALUATION OF A METAL POWDER CORED STRIP ELECTRODE IN HIGH SPEED ELECTROSLAG WELDING, *ADVANCES IN WELDING SCIENCE AND TECHNOLOGY,* ASM INTERNATIONAL, 1986, P 401-412
15. R.D. THOMAS, JR., CORROSION RESISTANT WELD OVERLAYS BY THE DUAL STRIP

PROCESS, *BRIT. WELD. J.,* MAY 1966

16. A. MITCHELL AND G. BEYNON, ELECTROCHEMICAL REACTIONS IN THE ELECTROSLAG WELDING OF THICK WALLED STRUCTURES, *AVTOM. SVARKA,* VOL 6, 1965, P 32-37
17. A. MITCHELL AND G. BEYNON, ELECTROCHEMICAL REACTIONS IN THE ELECTROSLAG PROCESS, CH. 2, *BUR. MINES BULL.,* 1976, P 669
18. M.E. PEOVER, ELECTROSLAG REMELTING, A REVIEW OF ELECTRICAL AND ELECTROCHEMICAL ASPECTS, *J. INST. MET.,* VOL 100, 1972, P 97
19. R.H. FROST, D.L. OLSON, AND G.R. EDWARDS, "THE INFLUENCE OF ELECTROCHEMICAL REACTIONS ON THE CHEMISTRY OF THE ELECTROSLAG WELDING PROCESS, *MODELING OF CASTING AND WELDING PROCESSES II,* AIME, 1983, P 279-294
20. D. YU, H. ANN, J.H. DEVLETIAN, AND W.E. WOOD, SOLIDIFICATION STUDY OF NARROW-GAP ELECTROSLAG WELDING, *WELDING RESEARCH: THE STATE OF THE ART,* E. NIPPES AND D. BALL, ED., ASM, 1985, P 21-32
21. V.I. AVRAMENKO, B.F. LEBEDEV, AND V.I. BOZHKO, SOME WAYS OF INCREASING THE PRODUCTIVITY OF ELECTROSLAG WELDING, *SVAR. PROIZVOD.,* VOL 10, 1973, P 16
22. K. WATANABE, I. SEJIMA, S. KOKURA, G. TAKI, AND H. MIYAKE, PROBLEMS AND IMPROVEMENT OF LARGE HEAT INPUT ELECTROSLAG WELDING, *J. JPN. WELD. SOC.,* 1975, P 519
23. S.A. SMIRNOV AND L.A. EFIMENKO, SPECIAL STRUCTURAL FEATURES AND MECHANICAL PROPERTIES OF ELECTROSLAG WELDED JOINTS MADE USING POWDERED FILLER METAL, *SVT. SVARKA,* VOL 9, 1973, P 46
24. F. EICHHORN AND J. REMMEL, SITUATION OF RESEARCH IN ELECTROSLAG WELDING--A TENDENCY OF FURTHER DEVELOPMENT, *IND. WELD. J.,* VOL 4, 1983, P 37
25. F. EICHHORN, J. REMMEL, AND B. WUBBELS, HIGH SPEED ELECTROSLAG WELDING, *WELD. J.,* VOL 62 (NO. 1), 1984, P 37
26. F. EICHHORN AND J. REMMEL, EFFICIENT FILLET WELDING IN THE VERTICAL WELDING POSITION WITH ELECTROGAS AND ELECTROSLAG WELDING METHODS, DOC. XII-908-85, IIW, 1985
27. B.F. YAKUSHIN *ET AL.,* IMPROVING THE CAPACITY OF ELECTROSLAG WELDED JOINTS FOR RESISTING HOT CRACKING, *WELD. RES. ABROAD,* JUNE/JULY 1984 (TRANSLATION FROM RUSSIAN BY AUTOMATIC WELDING, OCT 1982)
28. S. FORSBERG, RESISTANCE ELECTROSLAG (RES) SURFACING, *WELD. J.,* VOL 63 (NO. 8), 1985, P 41-48
29. P.C. ARNOLD AND D.C. BERTOSSA, MULTIPLE PASS AUTOMATIC VERTICAL WELDING, *WELD. J.,* VOL 45 (NO. 8), 1966, P 651-660
30. D.J. ELLIS AND A.F. GIFFORD, APPLICATION OF ELECTRO-SLAG AND CONSUMABLE GUIDE WELDING--PARTS 1-6, *MET. FABR.,* APRIL-NOV 1973
31. H. PASS, THE ELECTRO-SLAG WELDING OF A BLAST FURNACE HEARTH JACKET, *WELD. MET. FABR.,* JAN/FEB 1976
32. ELECTROSLAG WELDING HELPS TO RESTORE WIND TUNNEL, *MET. PROGR.,* MARCH 1976
33. J.S. NOEL AND A.A. TOPRAC, STATIC, FATIGUE, AND IMPACT STRENGTH OF ELECTROSLAG WELDMENTS, RESEARCH REPORT 157-1F, RESEARCH PROJECT 3-5-71-157, CENTER FOR HIGHWAY RESEARCH, UNIV. OF TEXAS, 1971
34. A.M. MAKARA *ET AL.,* ELECTROSLAG WELDING OF STRUCTURAL STEELS WITHOUT NORMALIZING, *AVTOM. SVARKA,* VOL 7, 1974, P 511-514
35. B. HOWSER, APPLICATION AND TRENDS IN ELECTROSLAG WELDING IN THE UNITED STATES, *ELECTROSLAG WELDING FOR MARINE APPLICATIONS,* U.S. NAVAL ACADEMY, MARCH 1985, P 4-1
36. J.E. NORCROSS, ELECTROSLAG DEVELOPMENTS, *FWP J.,* VOL 18 (NO. 2), 1978, P 41-58

37. D. YAPP AND A.P. BENNETT, DEVELOPMENT OF ELECTROSLAG-MELTED GRADED TRANSITION JOINTS, *WELDING RESEARCH RELATED TO POWER PLANT,* PAPER 34, CENTRAL ELECTRICITY GENERATING BOARD, SOUTHAMPTON, ENGLAND, SEPT 1972, P 464-481
38. J.E. NORCROSS, ELECTROSLAG BOSS AND PROCESS, U.S. PATENT 4,130,931, DEC 1978
39. ELECTROSLAG CASTING AND WELDING, BRITISH PATENT 1,390,674, APRIL 1975
40. A.P. BENNETT, AN ELECTRICAL ANALOGUE OF AN ELECTROSLAG REMELTING OPERATION USED AS A GUIDE TO IMPROVED CONTROL OF THE PROCESS, *MET. MATER.,* APRIL 1972, P 146-149
41. A. BROSHOLEN, E. SKAUD, AND J.J. VESSER, ELECTROSLAG WELDING OF LARGE CASTINGS FOR SHIP CONSTRUCTION, *WELD. J.,* VOL 56 (NO. 8), 1977, P 26-30
42. Y. ISHII, ELECTROSLAG WELDING OF CAST IRON, *TRANS. JAP. WELD. SOC.,* NO. 2, SEPT 1970, P 241-252
43. V.F. YAKOVLEV, P.I. KOVALKIN, N.I. EVDOKIMOV, M.G. KOZULIN, AND I.I. SUSHCHUCK-SLYUSARENKO, ELECTROSLAG WELDING OF STEEL CASTING DEFECTS, *AVTOM. SVARKA,* VOL 2, 1970, P 72
44. V.G. SVYNYUK, ELECTROSLAG WELDING FOR RECTIFYING DEFECTS IN CAST CENTRIFUGAL PUMP WHEELS, *AUTOM. WELD.,* VOL 28 (NO. 7), 1975, P 57-58
45. V.T. ARSENKIN, V.G. RADCHENKO, AND D.M. LIKHOHSHERSTOV, THE ELECTROSLAG DEPOSITION OF DIE STEEL ON CARBON STEEL, *AVTOM. SVARKA,* VOL 3, 1976, P 46-49
46. M.P. IVANAN, *ET AL.,* ELECTROSLAG HARDFACING OF BEATERS OF HAMMER CRUSHERS, *SVAR. PROIZVOD.,* VOL 7, 1974, P 48
47. Y.I. GORBACHEV, E.A. KOVALEVSKII, AND A.V. MATELEV, ELECTROSLAG WELDING OF DIE STEEL, *SVAR. PROIZVOD.,* VOL 11, 1984, P 12-13
48. N. LEHKA, PROBLEMS ASSOCIATED WITH THE CORROSION OF THE PROPERTIES OF ELECTROSLAG WELDED JOINTS IN AUSTENITIC STEELS OF INCREASED STRENGTH, *ZVRACSKE SPRAVY/WELDING NEWS,* VOL 26 (NO. 3), 1976, P 61-68
49. P.M. BARTLE, DEVELOPMENT OF AN ELECTROGAS TECHNIQUE FOR WELDING THICK ALUMINUM, *WELD. INST. RES. BULL.,* VOL 12 (NO. 3), 1971, P 67-71
50. Y.G. GAGEN, *ET AL.,* ELECTROSLAG WELDING OF ALUMINUM BUS-BARS LOCATED IN THE ZONE OF A STRONG MAGNETIC FIELD, *SVAR. PROIZVOD.,* VOL 12, 1972, P 19-21
51. ELECTROSLAG OR SUBMERGED ARC WELDING OF ALUMINIUM OR ITS ALLOYS OR ELECTROSLAG REMELTING ALUMINUM, BRITISH PATENT 1,544,248, APRIL 1979
52. V.Y. MALIN, ELECTROSLAG WELDING OF TITANIUM AND ITS ALLOYS, *WELD. J.,* VOL 64, FEB 1985, P 42-49
53. J.H. DEVLETIAN, ELECTROSLAG WELDING OF NON-FERROUS METALS--A REVIEW, *ELECTROSLAG WELDING FOR MARINE APPLICATIONS,* U.S. NAVAL ACADEMY, 1985, P 3-1 TO 3-32
54. K. EASTERLING, *INTRODUCTION TO THE PHYSICAL METALLURGY OF WELDING,* BUTTERWORTHS, 1983
55. R. BRUSCATO, TEMPER EMBRITTLEMENT AND CREEP EMBRITTLEMENT OF 2-1/4CR 1MO SHIELDED METAL ARC WELD DEPOSITS, *WELD. J.,* VOL 9 (NO. 4), 1970, P 148S-156S
56. A.M. MAKARA, Y.Y. KOKALEV, AND I.V. NOVIKOV, LIQUIDATION CRACKS IN THE HAZ IN STRUCTURAL STEELS DURING ELECTROSLAG WELDING, *AVTOM. SVARKA,* VOL 5, 1972, P 1-5
57. Y. MURAKAMI, T. NAMURA, AND J. WATANABE, HEAVY SECTION 2-1/4CR-1MO STEEL FOR HYDROGENATION REACTORS, STP 755, ASTM, 1982, P 383-417
58. T. KUNIHIRO AND H. NAKAJIMA, MICRO-CRACKING IN CONSUMABLE-NOZZLE ELECTROSLAG WELD METAL, *SIGNIFICANCE OF DEFECTS IN WELDED STRUCTURES,* UNIV. TOKYO PRESS, 1973, P 105-109
59. S.S. TULIANI, K.S. PROBERT, AND A.H. BRISCOE, FABRICATION OF ELECTROSLAG WELDED

MAGNOX FUEL TRANSPORT FLASKS, *WELDING AND FABRICATION IN THE NUCLEAR INDUSTRY*, 1979, P 327-333

60. O. SERRANO, G.E. EDWARDS, AND R.H. FROST, A COMPARISON OF AS-WELDED AND STRESS-RELIEVED 2-1/4CR-1MO STEEL ELECTROSLAG WELDMENTS: MICROSTRUCTURAL AND PROPERTIES, *THE APPLICATION OF 2-1/4CR-1MO STEEL FOR THICK WALL PRESSURE VESSELS,* ASTM, 1980
61. B.M. KAPADIA, INFLUENCE OF RESIDUAL STRESS ON FATIGUE CRACK PROPAGATION IN ELECTROSLAG WELDS, *FATIGUE TESTING OF WELDMENTS,* STP 648, ASTM, 1977, P 244-260

Oxyfuel Gas Welding

Revised by William Ballis, Columbia Gas of Ohio

Introduction

OXYFUEL GAS WELDING (OFW) is a manual process in which the metal surfaces to be joined are melted progressively by heat from a gas flame, with or without filler metal, and are caused to flow together and solidify without the application of pressure to the parts being joined. The most important source of heat for OFW is the oxyacetylene welding (OAW) torch.

The simplest and most frequently used OFW system consists of compressed gas cylinders, gas pressure regulators, hoses, and a welding torch. Oxygen and fuel are stored in separate cylinders. The gas regulator attached to each cylinder, whether fuel gas or oxygen, controls the pressure at which the gas flows to the welding torch. At the torch, the gas passes through an inlet control valve and into the torch body, through a tube or tubes within the handle, through the torch head, and into the mixing chamber of the welding nozzle or other device attached to the welding torch. The mixed gases then pass through the welding tip and produce the flame at the exit end of the tip. This equipment can be mounted on and operated from a cylinder cart, or it can be a stationary installation. Filler metal, when needed, is provided by a welding rod that is melted progressively along with the surfaces to be joined.

Acknowledgements

This article was revised from "Oxyfuel Gas Welding Processes and Their Application to Steel" in Volume 6 of the 9th Edition of *Metals Handbook.* Special thanks are due to the authors of that article: Gene Meyer, Victor Equipment; Joseph E. McQuillen, Air Products and Chemicals, Inc.; and Clarence E. Vaughn, Chicago Bridge & Iron Company.

Oxyfuel Gas Welding

Revised by William Ballis, Columbia Gas of Ohio

Capabilities, Advantages, and Limitations

In OFW, the welder has considerable control over the temperature of the metal in the weld zone. When the rate of heat input from the flame is properly coordinated with the speed of welding, the size, viscosity, and surface tension of the welding pool can be controlled, permitting the pressure of the flame to be used to aid in positioning and shaping the weld. The welder has control over filler-metal deposition rates because the sources of heat and filler metal are separate. Heat can be applied preferentially to the base metal or the filler metal without removing either from the flame envelope. With these capabilities, OFW can be used for joining thin sheet metal, thin-walled tube, small pipe, and assemblies with poor fit-up, as well as for smoothing or repairing rough arc welds.

The equipment is versatile, low-cost, self-sufficient, and usually portable. It can be used for preheating, postheating, welding, braze welding, and torch brazing, and it is readily converted into oxygen cutting. The process can be adapted to short production runs, field work, repairs, and alterations.

Metals That Can Be Oxyfuel Gas Welded. Most ferrous and nonferrous metals can be oxyfuel gas welded. Oxyacetylene supplies the heat intensity and flame atmosphere necessary for welding carbon steel, cast iron, and other ferrous, copper, and nickel alloys (Table 1). Aluminum and zinc alloys can also be welded by the oxyacetylene process. Oxyfuel gas welding of steel is done almost exclusively with an oxyacetylene flame. Hydrogen, natural gas, propane, and several proprietary gases are used as fuel gases in welding metals with lower melting temperatures, such as aluminum, magnesium, zinc, lead, and some precious metals. Metals unsuited to OFW are the refractory metals, such as niobium, tantalum, molybdenum, and tungsten, as well as the reactive metals, such as titanium and zirconium.

TABLE 1 FERROUS AND NONFERROUS METALS THAT CAN BE WELDED BY OXYACETYLENE WELDING

Base metal	Filler-metal type	Flame type	Flux type
Aluminums	(a)	Slightly reducing	Aluminum
Brasses	Navy brass	Slightly oxidizing	Borax
Bronzes	Copper tin	Slightly oxidizing	Borax
Copper	Copper	Neutral	(b)
Copper nickel	Copper nickel	Reducing	(b)
Inconel	(a)	Slightly reducing	Fluoride
Iron, cast	Cast iron	Neutral	Borax
Iron, wrought	Steel	Neutral	(b)
Lead	Lead	Slightly reducing	(b)
Monel	(a)	Slightly reducing	Monel
Nickel	Nickel	Slightly reducing	(b)
Nickel silver	Nickel silver	Reducing	(b)
Low-alloy steel	Steel	Slightly reducing	(b)
High-carbon steel	Steel	Reducing	(b)
Low-carbon steel	Steel	Neutral	(b)
Medium-carbon steel	Steel	Slightly reducing	(b)
Stainless steel	(a)	Slightly reducing	Stainless steel

Source: Ref 1

(A) MATCH BASE METAL.

(B) NO FLUX REQUIRED.

Fluxes. Except for lead, zinc, and some precious metals, OFW of nonferrous metals, cast irons, and stainless steels generally requires a flux. In welding carbon steel, the gas flame shields the weld adequately, and no flux is required. Adjustment for correct flame atmosphere is important, but the absence of flux results in one less variable to control.

Major Applications. Oxyfuel gas welding can be used to join thin carbon steel sheet and carbon steel tube and pipe. The advantages of OFW include the ability to control heat input, bridge large gaps, avoid melt-through, and clearly view the weld pool. Carbon steel sheet, formed in a variety of shapes, can often be welded more economically by OFW than by other processes. Oxyfuel gas welding is capable of joining small-diameter carbon steel pipe (up to about 75 mm, or 3 in., in diameter) with resulting weld quality equal to competitive processes and often with greater economy. Pipe with wall thickness up to 4.8 mm ($\frac{3}{16}$ in.) can be welded in a single pass.

Welder Skill. Oxyfuel gas welding requires skill in manipulating the welding rod and the torch flame. In depositing a weld, the welder uses both hands to melt base metal and filler metal, control the weld pool, and obtain progressive solidification of weld metal in the correct bead shape.

Reference cited in this section

1. H.B. CARY, *MODERN WELDING TECHNOLOGY,* PRENTICE-HALL, 1989, P 211

Oxyfuel Gas Welding

Revised by William Ballis, Columbia Gas of Ohio

Gases

Oxygen and acetylene are the principal gases used in OFW. Oxygen supports combustion of the fuel gases. Acetylene supplies both the heat intensity and the atmosphere needed to weld steel. Hydrogen, natural gas, propane, and proprietary gases are used only to a limited extent in oxyfuel gas welding or brazing of metals with a low melting temperature.

Oxygen. Only by burning selected fuel gases with high-purity oxygen in a high-velocity flame can the high heat transfer intensity required in OFW be obtained. Oxygen is supplied for oxyfuel gas welding and cutting at a purity of 99.5% and higher, because small percentages of contaminants have a noticeable effect on combustion efficiency.

When the consumption requirement is relatively small, the oxygen is supplied and stored as a compressed gas in a standard steel cylinder under an initial pressure of up to 180 MPa (26 ksi). The most frequently used cylinder (Fig. 1) has a capacity of 6.91 m^3 (244 scf).* The gas is distributed for use under reduced pressure. When consumption of oxygen is somewhat greater, banks of cylinders are joined through a manifold to permanent pipeline systems that terminate at various stations of use.

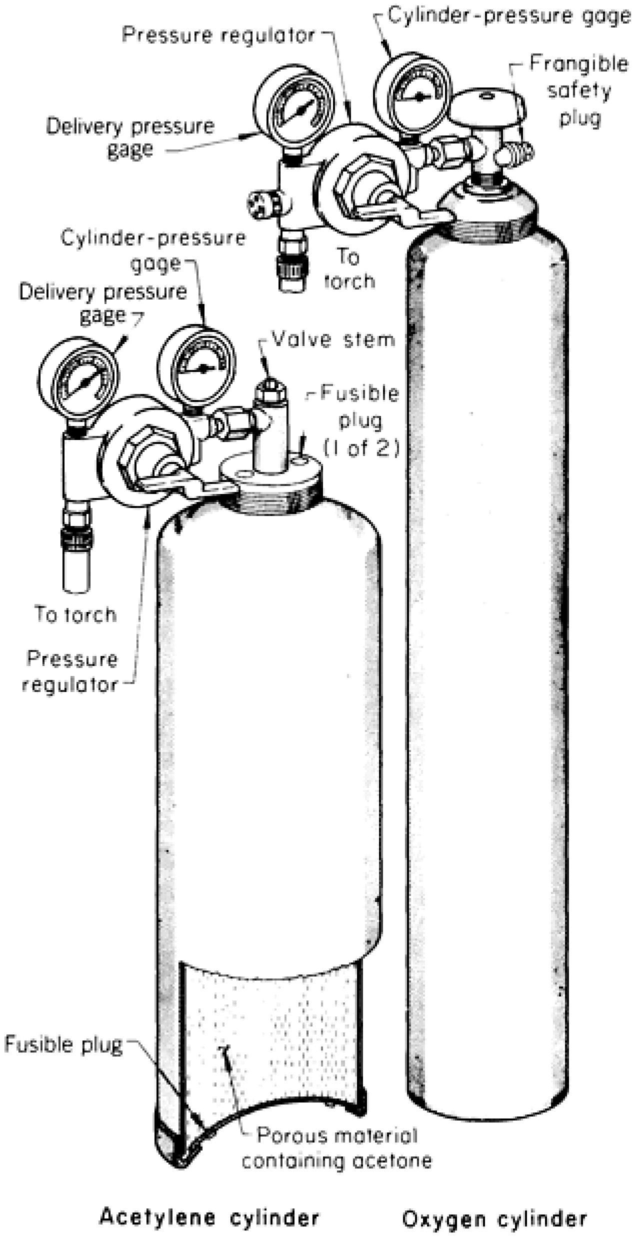
Cylinder-pressure gage
Pressure regulator
Frangible safety plug
Delivery pressure gage
Cylinder-pressure gage
To torch
Delivery pressure gage
Valve stem
Fusible plug (1 of 2)
To torch
Pressure regulator
Fusible plug
Porous material containing acetone
Acetylene cylinder
Oxygen cylinder

FIG. 1 GAS CYLINDERS AND REGULATORS USED IN OXYFUEL GAS WELDING. THE ACETYLENE CYLINDER SHOWN IS 1029 MM (40.5 IN.) HIGH, 314 MM (12.35 IN.) IN DIAMETER, AND HAS A WALL THICKNESS OF 4.4 MM (0.175 IN.). THE OXYGEN CYLINDER IS 1295 MM (51 IN.) HIGH, 229 MM (9 IN.) IN DIAMETER, AND HAS A WALL THICKNESS OF 6.3 MM (0.25 IN.).

When oxygen consumption exceeds approximately 6.91 m^3 (244 scf) cylinders per week, it may be more economical to obtain and store oxygen in liquid form. Liquid oxygen can be supplied in portable cryogenic cylinders of 130 m^3 (4500 scf) capacity, or it can be delivered in bulk to an on-site cryogenic storage tank. The gas is then piped to points of use. The distribution and use of oxygen are covered by laws and safety regulations designed to help prevent injury to persons and damage to property.

Acetylene is a hydrocarbon gas with the chemical formula C_2H_2. When under pressure of 203 kPa (29.4 psi) and above, acetylene is unstable, and a slight shock can cause it to explode, even in the absence of oxygen or air. Safety rules for the use of acetylene and the handling of acetylene equipment are extremely important.

Acetylene should not be used at pressure greater than 105 kPa (15 psi). Acetylene generators for on-site gas production are constructed so that the gas is not given off at pressures much greater than 105 kPa (15 psi). Commercially supplied portable cylinders are specially constructed (Fig. 1) to store acetylene under high pressure. By dissolving acetylene in liquid acetone, a cylinder such as that shown in Fig. 1 can be used to store about 7.79 m^3 (275 scf) of acetylene under a pressure of 1.7 MPa (250 psi). This pressure must be reduced to 105 kPa (15 psi) or less by the regulator (Fig. 1) before the gas enters the hose or distribution line.

Acetylene cylinders must not be subjected to sudden shock and should be stored well away from any source of heat or sparks. The cylinders must be stored in an upright position to keep the acetone from escaping during use. Under normal sustained use, withdrawal rate from an acetylene cylinder should not exceed one-seventh of the cylinder capacity per hour.

Hydrogen is used chiefly for welding lower-melting-temperature metals, such as aluminum, magnesium, and lead. It cannot be used to weld common thicknesses of steel sheet, because it results in a flame temperature that is too low to produce good fusion. Hydrogen has, however, been used in welding thin sheet, where its lower combustion intensity (about 60% of that of acetylene) can be an advantage. It is used for brazing and, to some extent, for braze welding. Hydrogen is available in compressed-gas cylinders of various sizes.

Natural gas, propane, and proprietary gases can be used with oxygen to weld some lower-melting-temperature metals. Special welding rods may be required depending on the gas, base metal, and quality of metal desired. These gas mixtures cannot usually be applied to the welding of steel because when they are burned at temperatures high enough for welding, their flame atmospheres are excessively oxidizing. When ratios of oxygen to fuel gas are reduced to a carburizing condition, flame temperatures are too low. Therefore, these gases are usually limited to brazing, braze welding, heating, and related processes.

Note cited in this section

* * A STANDARD CUBIC FOOT (SCF) OF GAS IS DEFINED AS EQUIVALENT TO 0.028 M^3 (1 FT^3) OF GAS AT 20 °C (70 °F) AND 100 KPA (1 ATM, OR 14.7 PSI) PRESSURE. THIS DEFINITION, WHICH IS USED IN THE GAS INDUSTRY AND SOME ENGINEERING PRACTICE, DIFFERS FROM THE STANDARD TEMPERATURE AND PRESSURE OF 0 °C (32 °F) AND 100 KPA (1 ATM) USED IN SCIENTIFIC WORK.

Oxyfuel Gas Welding

Revised by William Ballis, Columbia Gas of Ohio

Equipment

The principal function of OFW equipment is to supply the oxyfuel gas mixture to the welding tip at the correct rate of flow, exit velocity, and mixture ratio. The rate of gas flow affects the quantity of metal melted; the pressure and velocity affect the manipulation of the weld pool and the rate of heating; and the ratio of oxygen to fuel gas determines the flame temperature and the atmosphere, which must be chemically suited to the metal being welded. Important elements in an OFW system include gas storage facilities, pressure regulators, hoses, torches, related safety devices, and accessories.

Gas Storage Equipment. Compressed-gas cylinders and storage tanks are used as on-site supply sources for gases. They are constructed under regulations of the Department of Transportation and are regulated to some extent by federal, state, and local laws. Cylinders are designed for specific gases and are not generally interchangeable. Sizes and threading of cylinder connections for oxygen, for example, differ from those for acetylene, hydrogen, and other gases. Only the appropriate fittings can be used for delivery of compressed gas. Users should not tamper with valves or safety devices on cylinders.

When portable equipment is not required and gas consumption is large enough, permanent installations are constructed. Gases can be supplied from a variety of bulk-storage vessels, manifolded gas cylinders, or gas generators. The gases are then separately piped, at suitable pressure, to terminal stations, where they are drawn off through station regulators and used as in a portable setup (Ref 2).

Pressure regulators reduce the supply gas pressure to a desired delivery pressure. They are designed for specific gases and are not generally interchangeable. In Fig. 1, oxygen and acetylene pressure regulators used in OFW are shown attached to their respective cylinders and hoses. In operation, gas enters the inlet side of the regulator at cylinder pressure and emerges from the outlet side at the desired delivery pressure.

Regulators are made for various ranges of inlet and outlet pressure. They can be adjusted within their delivery pressure range by turning an adjusting screw. Although some regulators do not have gages and are preset to deliver at a specific and constant pressure, most are equipped with two pressure gages. The one on the outlet side permits the operator to read the adjusted delivery pressure; the one on the inlet side indicates the pressure in the cylinder.

Two basic types of regulators are used in OFW: single-stage and two-stage. Both are available with either direct or inverse actuation of the valve mechanism. Regulators with direct actuation are known also as positive or nozzle regulators; those with inverse actuation, as negative or stem regulators. All single-stage regulators reduce cylinder pressure in one step. Pipeline regulators, made for lower pressures, must not be used on cylinders (the higher cylinder pressure can blow apart a pipeline regulator).

When adjusted for a desired delivery pressure, regulators continue to deliver gas at the pressure shown on the outlet gage within a fairly narrow range of deviation, as the cylinder pressure drops. The amount and direction of the deviations depend on the design of the regulator. The variation in delivery pressure as cylinder pressure decreases is shown in Fig. 2 for three types of regulators. Both types of single-stage regulators may require occasional adjustment as cylinder pressure drops. The more costly two-stage regulator supplies a more constant working pressure until cylinder pressure nearly equals delivery pressure. On full oxygen cylinders exposed to low outdoor temperatures in winter, two-stage regulators are more reliable than single-stage regulators. The temperature drop of expanding gas is less severe because pressure is reduced in two steps. Less chance exists for ice to form and clog regulator passages if water vapor is present.

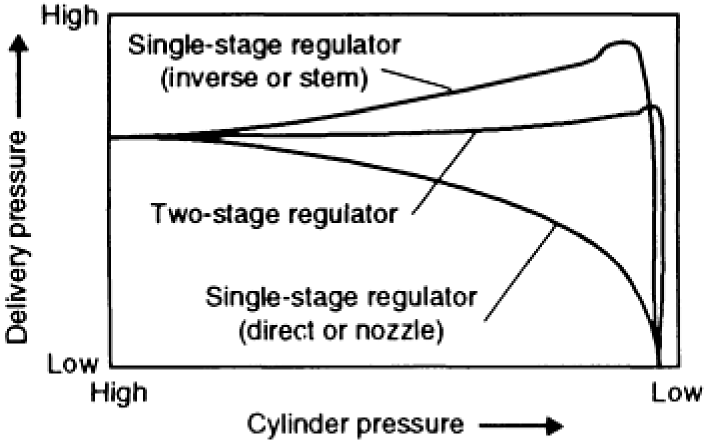

FIG. 2 DELIVERY PRESSURE FOR SELECTED TYPES OF PRESSURE REGULATORS

Hoses. Flexible hoses permit gas cylinders and regulators to be kept at a safe distance from the working area and also allow the welder freedom of movement. Hoses usually range from 3.2 to 13 mm ($\frac{1}{8}$ to $\frac{1}{2}$ in.) in inside diameter, with larger sizes available for special applications, and are usually available in 7.6 m (25 ft) lengths. If hoses are longer than 7.6 m (25 ft), either the next larger size hose or a length of larger diameter hose with a short length of the usual size hose should be used to connect to the welding torch. This facilitates ease of movement. To ensure that oxygen and fuel gas hoses are used with the proper gas and correct fittings, oxygen hoses are generally green, with right-hand threaded fittings; acetylene and other fuel gas hoses are generally red, with left-hand threaded fittings. Left-hand threaded fittings are grooved on the outside so they can be readily identified.

Welding torches control the operating characteristics of the welding flame and enable the flame to be manipulated during welding. The choice of torch size and style depends on the work to be performed. Aircraft welding torches, for example, are small and light to permit ease of handling. Most torch styles permit one of several sizes of welding tips or a cutting attachment to be added.

The general construction of an OFW torch is shown schematically in Fig. 3. The principal operating parts are inlet valves, rear body, handle, and head. A mixing chamber with a welding tip, heating nozzle, or cutting attachment is added to the head. By unscrewing the sleeve nut (Fig. 3), the welding tip and mixing-chamber assembly can be removed and replaced by units with different capacity.

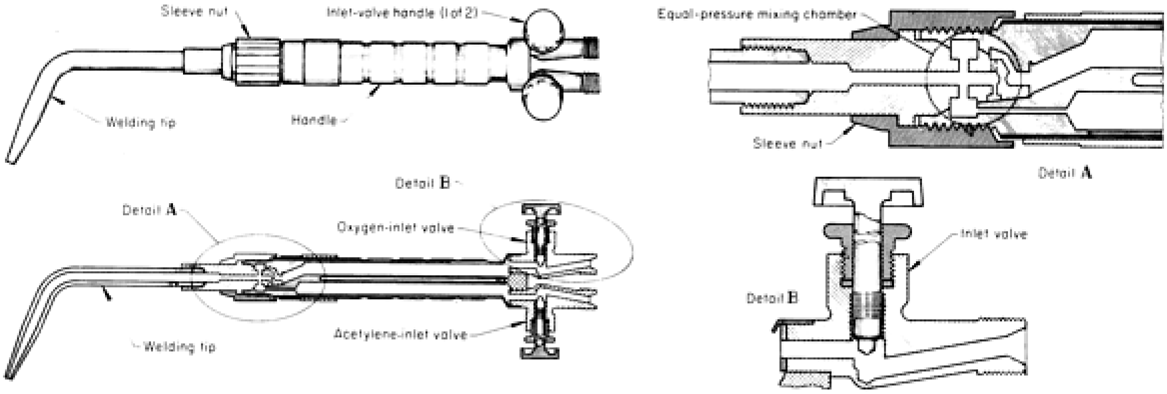
Sleeve nut
Inlet-valve handle (1 of 2)
Welding tip
Handle
Detail B
Detail A
Oxygen-inlet valve
Acetylene-inlet valve
Welding tip
Equal-pressure mixing chamber
Sleeve nut
Detail A
Inlet valve
Detail B

FIG. 3 SCHEMATIC SHOWING CROSS-SECTIONAL VIEWS OF GAS PASSAGES IN A TYPICAL OXYFUEL GAS WELDING TORCH

Torch inlet valves provide the welder with two important controls. First, the gas pressure, velocity, and flow rate can be adjusted within the limits set by the pressure regulators and the practical requirements of the welding flame. Second, the ratio of oxygen to fuel gas can be varied. The ability of the oxyfuel flame to produce the combustion intensity needed for welding, while providing a suitable protective atmosphere for the weld metal, is largely a property of the fuel gas, but correct torch settings must be used.

Mixing Chambers. In Fig. 3, the mixing chamber is shown as part of an assembly with the welding tip. Two general mixing chambers are available: equal-pressure (also called positive-pressure and medium-pressure) and injector. In an equal-pressure mixing chamber, gases are at approximately the same pressure and are mixed by directing the fuel gas into the oxygen stream. Detail A in Fig. 3 shows an equal-pressure mixing chamber assembled in position. Figure 4 is a spiral equal-pressure mixer.

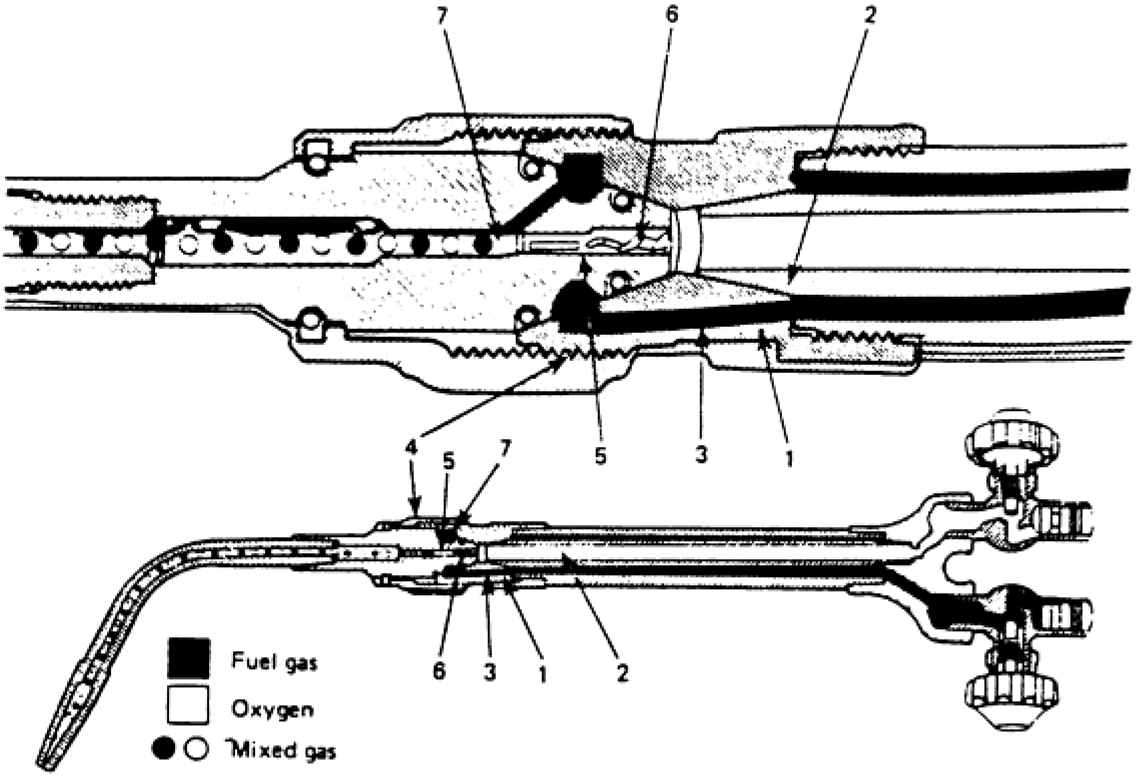

FIG. 4 SCHEMATIC SHOWING CROSS-SECTIONAL VIEW OF A SPIRAL EQUAL-PRESSURE MIXER. (1) WELDING TORCH HEAD. (2) OXYGEN TUBE FROM TORCH HEAD. (3) ACETYLENE (FUEL GAS) PASSAGES FROM TORCH HEAD. (4) NOZZLE NUT. (5) WELDING NOZZLE CONE END. (6) SPIRAL IN WELDING NOZZLE. (7) MIXER ORIFICE AND MIXING CHAMBER

In the injector mixing chamber, low-pressure fuel gas is aspirated by directing it into a high-velocity stream of oxygen. A nozzle system based on the flow principles of the venturi tube is used. Injector torches are useful when fuel gases are supplied at pressures too low to produce a flame of adequate combustion intensity. Supplying oxygen at desired pressure is not usually difficult. By varying the design of the injector nozzle, different degrees of aspiration can be obtained, and designs for injector mixing chambers differ considerably in detail.

Welding tips are replaceable nozzles that control gas flow through the diameter of the exit orifice. Tips of various orifice diameter are usually available for any welding torch. Orifice diameters are identified by drill-size number, decimal size in inches, or manufacturer's code number. Because code numbers of different manufacturers do not necessarily correspond, drill sizes or decimals are needed to compare the orifice size of different makes of tips. The performance of tips of equal size at equal pressure-regulator settings may differ, however, because of differences in torch and mixing-chamber designs.

Small-diameter tips produce small flames for welding thin sections; large-diameter tips are required for heavier work. Welding tips are made with a smooth bore at the exit end to ensure laminar flow and a uniform flame. The influence of bore shape on the shape of the flame is shown in Fig. 5. When foreign matter, such as carbon, dirt, or weld spatter, enters the welding-tip orifice, it must be carefully removed. Specially designed dressing tools (known as tip cleaners) are available for this purpose.

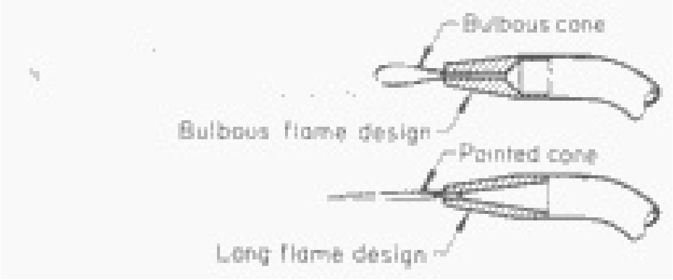

FIG. 5 VARIATION IN THE SHAPE OF THE INNER CONE OF AN OXYACETYLENE WELDING FLAME PRODUCED BY WELDING TIPS WITH TWO DIFFERENT BORE SHAPES. VARYING THE TRANSITION TAPER PRODUCES CONE SHAPES INTERMEDIATE BETWEEN THOSE SHOWN.

Accessories essential to OFW include a friction lighter for igniting the torch; welder's goggles, gloves, and protective clothing; and related safety devices.

Welder's goggles are covered by ANSI standard Z49.1 (Ref 3), which suggests the following lens shade numbers for use in OFW of steel:

STEEL THICKNESS		SHADE NUMBER
MM	IN.	
≤ 3.2	$\leq \frac{1}{8}$	4 OR 5
3.2-13	$\frac{1}{8} - \frac{1}{2}$	5 OR 6
>13	$> \frac{1}{2}$	6 TO 8

For general OFW, goggles must have side shields.

References cited in this section

2. "OXYGEN-FUEL GAS SYSTEMS FOR WELDING AND CUTTING," BULLETIN NO. 51, NATIONAL FIRE PROTECTION ASSOCIATION, 1977
3. "SAFETY IN WELDING AND CUTTING," ANSI/AWS ZA9.1, AMERICAN WELDING SOCIETY, 1983

Oxyfuel Gas Welding

Revised by William Ballis, Columbia Gas of Ohio

Selection of Tip-Orifice Size and Gas Pressure

Torch manufacturers provide charts that give tip-orifice sizes to be used for welding different thicknesses of metal. Table 2 presents the data from a chart showing these relationships. Tip size number and drill size are not precise means of comparing tips because significant differences exist between tips of the same dimensions when operated under positive pressure and when operated under injector pressure. The only valid basis of comparison is the volume of fuel passing through the tip in a unit of time. Most tips operate within a range, usually recorded as cubic feet of fuel per hour. The manufacturer's recommendations should always be followed for optimum performance and safety.

TABLE 2 TYPICAL TIP-ORIFICE SIZES AND ACETYLENE CONSUMPTION FOR OXYFUEL GAS WELDING OF 0.25 TO 6.4 MM (0.010 TO $\frac{1}{4}$ IN.) THICK STEEL

THICKNESS OF STEEL		**TIP-ORIFICE SIZE**			ACETYLENE CONSUMPTION	
		DIAMETER		**DRILL NO.**		
MM	**IN.**	**MM**	**IN.**		**M^3/H**	FT^3/H
0.25	0.010	0.571	0.0225	74	≤ 0.028	≤ **1**
0.41	0.016	0.711	0.0280	70	≤ 0.028	≤ **1**
0.48	0.019	0.711	0.0280	70	≤ 0.028	≤ **1**
0.8	$\frac{1}{32}$	0.889	0.0350	65	0.014-0.057	$\frac{1}{2}$ **-2**
1.6	$\frac{1}{16}$	1.12	0.0465	56	0.028-0.11	**1-4**
2.4	$\frac{3}{32}$	1.12-1.40	0.0465-0.0550	56-54	0.11-0.17	**4-6**
3.2	$\frac{1}{8}$	1.40-1.51	0.0550-0.0595	54-53	0.17-0.28	**6-10**
4.8	$\frac{3}{16}$	1.51-1.78	0.0595-0.0700	53-50	0.28-0.48	**10-17**
6.4	$\frac{1}{4}$	**1.78-2.06**	**0.0700-0.0810**	**50-46**	**0.48-0.85**	**17-30**

Oxyfuel Gas Welding

Revised by William Ballis, Columbia Gas of Ohio

Flame Adjustment

Different welding atmospheres and flame temperatures can be produced by varying the relative amounts of oxygen and fuel gas in the gas flowing to the tip of the torch. Usually, a welder makes the appropriate adjustments in gas flow based on the appearance of the flame. This is not true for oxyhydrogen welding, however. The sequence for setting up a positive-pressure welding outfit is:

- CHECK ALL PARTS OF THE APPARATUS, MAKING SURE THEY ARE FREE OF DIRT, OIL, OR GREASE AND IN PROPER WORKING CONDITION.
- CRACK EACH CYLINDER VALVE INDIVIDUALLY TO BLOW OUT FOREIGN MATTER. MAKE SURE VENTED GASES ARE SAFELY DISPERSED. WIPE OUT THE CYLINDER-VALVE OUTLET WITH A CLEAN, LINT-FREE CLOTH.
- ATTACH THE OXYGEN REGULATOR TO THE OXYGEN CYLINDER OR MANIFOLD.
- ATTACH THE ACETYLENE (OR FUEL GAS) REGULATOR TO THE FUEL CYLINDER OR SOURCE.
- CONNECT THE WELDING HOSE TO THE REGULATORS AND THE WELDING TORCH.
- OPEN THE CYLINDER VALVE SLOWLY AND CAREFULLY. THE OPERATOR SHOULD NEVER STAND IN FRONT OF THE REGULATOR WHEN OPENING THE CYLINDER VALVE.
- PURGE THE OXYGEN LINE WHILE THE ACETYLENE LINE IS CLOSED AND THE ACETYLENE LINE WHILE THE OXYGEN LINE IS CLOSED. VENT GASES SAFELY.
- SET THE OXYGEN AND FUEL GAS REGULATORS TO THE RECOMMENDED WORKING

PRESSURE WITH APPROPRIATE TORCH VALVE OPEN.

- OPEN THE ACETYLENE (OR FUEL GAS) INLET VALVE AND LIGHT THE WELDING TORCH, USING A SPARK LIGHTER.
- OPEN THE OXYGEN INLET VALVE AND ADJUST THE FLAME, USING BOTH INLET VALVES.

For injector equipment, the sequence and method of setting up differ because high-pressure oxygen aspirates low-pressure acetylene into the torch. The differences among brands make it essential to always follow the manufacturer's directions for setting up.

Oxyfuel Gas Welding

Revised by William Ballis, Columbia Gas of Ohio

Oxyacetylene Combustion

As the acetylene/oxygen mixture burns from the tip of the welding torch, it displays several clearly recognizable zones of combustion. The overall chemical equation for the complete combustion of acetylene is:

$$2C_2H_2 + 5O_2 \rightarrow 4CO_2 + 2H_2O \qquad \textbf{(EQ 1)}$$

Combustion takes place in two stages. The first stage:

$$2C_2H_2 + 2O_2 \rightarrow 4CO + 2\,H_2 \qquad \textbf{(EQ 2)}$$

uses the oxygen supplied from the cylinder and available in the oxyacetylene mixture. The reaction can be seen as the small inner cone of the flame. The highest temperature is at the point of this cone. The second stage:

$$4CO + 2H_2 + 3O_2 \rightarrow 4CO_2 + 2H_2O \qquad \textbf{(EQ 3)}$$

uses the oxygen supplied from the air surrounding the flame. This combustion zone constitutes the outer envelope of the flame.

About two-fifths of the oxygen necessary for the complete combustion of acetylene comes from the oxygen cylinder; the remainder comes from the air. Because of the need for supplemental oxygen from the atmosphere, the acetylene/oxygen flame cannot be used inside tubes of structures subject to oxygen depletion from OFW. By varying the relative amounts of acetylene and oxygen in the gas mixture in the torch, a welder can produce different flame atmospheres and temperatures.

The second equation shows that in the first stage, when equal amounts of oxygen and acetylene are burning, neither excess acetylene nor excess oxygen is present at the high-temperature tip of the inner cone. For this reason, the flame is called neutral, and the gas mixture is often described as an acetylene-to-oxygen ratio of 1 to 1, or an equal ratio. The condition of the flame is important because the inner cone is held close to, but not touching, the work metal. If excess oxygen is present, the molten metal foams and sparks, and brittle oxides may form in the weld metal. If acetylene is present in sufficient excess, indicated by an acetylene feather greater than about one-half the length of the inner cone, carbon can enter the metal. Some of the carbon forms carbides and some burns, causing gas and porosity in the solidified weld metal. In austenitic stainless steels, carbides formed in this manner aggravate susceptibility to intergranular corrosion. In welding steel, the flame should be as nearly neutral as possible, but a perfectly neutral flame is difficult to recognize, and a flame containing a slight excess of acetylene is often used.

A neutral flame is obtained by observing the size and color of the combustion zones in the flame as the oxygen-to-acetylene ratio is changed by adjusting the torch inlet valves. Figure 6 shows five typical flame conditions that appear as oxygen flow is increased from zero to excess oxygen, as well as a separated flame condition that results from excessive gas pressure.

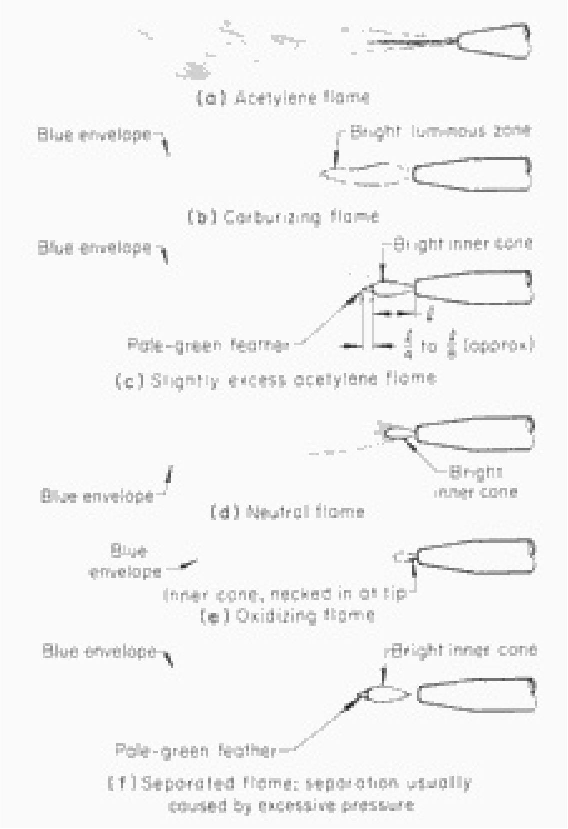
(a) Acetylene flame
Blue envelope
Bright luminous zone
(b) Carburizing flame
Blue envelope
Bright inner cone
ℓ
Pale-green feather
$\frac{\ell}{4}$ to $\frac{\ell}{8}$ (approx)
(c) Slightly excess acetylene flame
Bright inner cone
Blue envelope
(d) Neutral flame
Blue envelope
Inner cone, necked in at tip
(e) Oxidizing flame
Blue envelope
Bright inner cone
Pale-green feather
(f) Separated flame; separation usually caused by excessive pressure

FIG. 6 FLAME CONDITIONS OBTAINED AS OXYGEN FLOW RATE INCREASES FROM ZERO TO AN EXCESS OF OXYGEN IN AN OXYACETYLENE WELDING TORCH

Acetylene Flame. When acetylene alone is burned in air, it produces a flame that varies in color from yellow near the torch tip to orange-red at the outer extremity. If soot particles are present at the end of the flame, the volume of gas should be increased until the flame burns clear (Fig. 6a).

Carburizing Flame. As the oxygen valve in the torch is progressively opened and the ratio of oxygen to acetylene increases, the flame becomes generally luminous. Then, the luminous portion contracts toward the welding tip, forming a distinct bright zone within a blue outer envelope (Fig. 6b). This is a carburizing flame because it has a large excess of acetylene; it is sometimes described as a soft flame because it has very little force. It has a relatively low temperature and is used in silver brazing and soldering, as well as in the welding of lead.

Reducing Flame. As more oxygen is introduced, the bright zone of the flame contracts farther and is seen to consist of two parts: a bright inner cone and a pale-green streamer, or feather, trailing off its end into the blue envelope. The streamer or feather is caused by a slight excess of acetylene. It disappears as the oxygen-to-acetylene ratio approaches 1 to 1. In Fig. 6(c), the feather is shown adjusted to about one-quarter the length of the inner cone. For welding steel, the length of the feather should be about one-eighth to one-quarter, but never more than one-half, the length of the inner cone. The flame is properly described as a slightly excess acetylene or reducing flame. It should not be called a carburizing flame because it does not carburize the metal, but it does ensure the absence of the oxidizing condition. It is frequently used for welding with low-alloy steel rods. The flame temperature at the tip of the inner cone is about 2930 to 3040 °C (5300 to 5500 °F).

Neutral Flame. In the neutral flame shown in Fig. 6(d), the oxygen-to-acetylene ratio is 1 to 1 (more accurately, 1.1 to 1), and the temperature at the tip of the inner cone is probably above 3040 °C (5500 °F) (Fig. 7). As pointed out earlier, making the precise adjustment of the inlet valves that results in a neutral flame is difficult, particularly in sunlight, because the oxidizing flame is of similar appearance. The neutral flame is ideal for the welding of steel, and it is also needed when the presence of carbon must be strictly avoided. When the oxidizing condition is unacceptable, as in welding stainless steel, the use of a neutral flame is essential for good results.

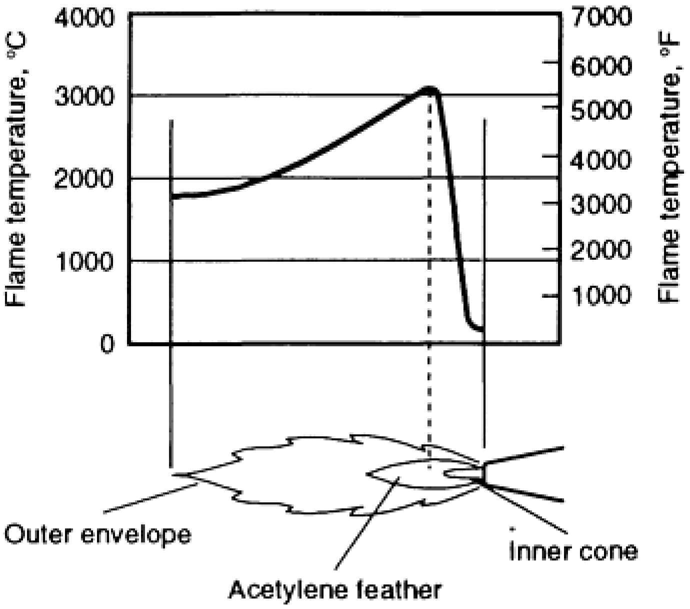

FIG. 7 FLAME TEMPERATURE AS A FUNCTION OF RELATIVE DISTANCE FROM THE TORCH TIP FOR A NEUTRAL (1:1 RATIO OF ACETYLENE TO OXYGEN) OXYACETYLENE FLAME. SOURCE: REF 4

Oxidizing Flame. An oxidizing flame is shown in Fig. 6(e). The adjustment of the inlet valves for this flame is difficult because it cannot be made on the basis of luminosity. When the flame tends to neck in at the juncture of the inner cone and the tip of the torch, this best indicates an oxidizing condition. When the flame is adjusted to be extremely oxidizing, it may also produce a hissing sound. An oxidizing flame should never be used in welding steel. It is used only in welding copper and certain copper-base alloys, and the flame should be sufficiently rich in oxygen to ensure that a film of oxide slag forms over the weld to provide shielding for the weld pool. With oxygen-to-acetylene ratios of about $1\frac{3}{4}$ to 1, flame temperature can approach 3315 °C (6000 °F).

Separated Flame. When gas pressure is too high for the size of the tip used, the flame can separate from the tip (Fig. 6f) and may even blow out. This is an unusable flame condition and must be avoided. Another cause of this condition is a clogged tip orifice.

Shape of Flame Cone. The design of the welding tip largely determines the shape of the inner cone of the oxyacetylene flame (Fig. 5); however, as the amount of oxygen used in the gas mixture is increased, the shape of the cone becomes more pointed. Both bulbous and pointed cones produce sound welds. The flame with the bulbous inner cone is preferred for welding in deep grooves in heavy sections. The flame with the long, pointed inner cone is softer and is generally preferred for welding thin sheet and aircraft assemblies.

Reference cited in this section

4. H.B. CARY, *MODERN WELDING TECHNOLOGY,* PRENTICE-HALL, 1989, P 210

Revised by William Ballis, Columbia Gas of Ohio

Oxyhydrogen Combustion

Complete combustion of hydrogen requires an oxygen-to-hydrogen ratio of 1 to 2, as can be seen from the following equation:

$$2H_2 + O_2 \rightarrow 2H_2O \qquad \textbf{(EQ 4)}$$

This gas mixture produces a strongly oxidizing flame having a temperature of about 2760 °C (5000 °F).

It is impossible to obtain a neutral oxyhydrogen flame by the visual methods of flame adjustment described for the oxyacetylene flame. The oxyhydrogen flame itself is scarcely visible, and no combustion zones, typical of the hydrocarbon gases, can be determined. To avoid an oxidizing flame, the pressure regulators must be set to ensure an excess of hydrogen. The flame is then reducing, but not carburizing. It has no carbon, and the temperature is several hundred degrees lower than that of the neutral flame. Metering flow regulators permit establishing the desired ratio of hydrogen to oxygen, usually 4 to 1. The oxyhydrogen flame is useful for welding and brazing aluminum alloys and lead.

Oxyfuel Gas Welding

Revised by William Ballis, Columbia Gas of Ohio

Combustion of Natural Gas and Propane

Complete combustion of natural gas (methane) and propane is shown, respectively, by the following equations:

$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O \qquad \textbf{(EQ 5)}$$

$$C_3H_8 + 5O_2 \rightarrow 3CO_2 + 4H_2O \qquad \textbf{(EQ 6)}$$

When the flame temperature is high enough to weld steel, the flame atmosphere is excessively oxidizing, but when the ratio of oxygen to fuel gas is decreased to produce a carburizing condition, flame temperature is too low for welding steel.

Oxyfuel Gas Welding

Revised by William Ballis, Columbia Gas of Ohio

Welding Rods

Filler metal for OFW of low-carbon steel is available in the form of cold-drawn steel rods 915 mm (36 in.) long and 1.6 to 4.0 mm ($\frac{1}{16}$ to $\frac{5}{32}$ in.) in diameter. Welding rods for OFW of other metals are supplied in various lengths, depending on whether they are wrought or cast.

Specifications. Steel welding rods have been standardized in the American Welding Society (AWS) specification A5.2, "Iron and Steel Gas Welding Rods." This specification shows three classifications of welding rods based on the minimum tensile strength of all-weld-metal and transverse weld test specimens:

AWS	MINIMUM TENSILE	MINIMUM

CLASSIFICATION	**STRENGTH**[(A)]		ELONGATION IN 4*D*[(B)], %
	MPA	**KSI**	
RG65	460	67	**16**
RG60	415	60	**20**
RG45	**310**	**45**	**. . .**

(A) ALL-WELD-METAL TEST SPECIMENS, 6.40 ± 0.13 MM (0.252 ± 0.005 IN.) IN DIAMETER.

(B) ELONGATION IN APPROXIMATELY 25 MM (1 IN.)

The mechanical properties specified for the RG65, RG60, and RG45 rods are obtained by welding in a neutral to slightly excess acetylene atmosphere.

The specification covers the chemical composition of the welding rods only to the extent of limiting sulfur and phosphorus to 0.040% maximum, and aluminum, if present, to 0.02% maximum. Therefore, welding rods of different manufacturers may vary appreciably in chemical composition.

Rods for OFW of steel have no flux covering. In the absence of flux coverings, weld-metal properties depend on chemical composition of the welding rod, control of the welding atmosphere, and techniques used to provide for mixing of base metal and filler metal.

Weld-Metal Strengthening. The ability to control the properties of the weld by mixing base metal and filler metal in the weld pool means that the choice of welding rod can influence weld strength to a considerable extent. Fully reinforced welds in thin-walled tubes of 4130 welded with RG45 rod consistently showed tensile strengths of 620 to 690 MPa (90 to 100 ksi). When the welds were made with RG60 rods, the strengths increased to 690 to 860 MPa (100 to 125 ksi), and with RG65 rods, strengths as high as 1000 MPa (145 ksi) were attained when the joint was heat treated after welding.

Class RG65 welding rods have a low-alloy steel composition and are used for OFW of carbon and low-alloy steels that have strengths of 450 to 515 MPa (65 to 75 ksi). They are used on sheet, plate, tube, and pipe. These rods give the highest strengths in welding 4130, 4340, and 8630 alloy steels when base metal and filler metal are mixed properly. The end use has a marked effect on selection of filler metal. If the base metal was selected to meet a specific corrosion- or heat-resistant application, the filler metal should be of similar composition. However, if a room-temperature mechanical property is the primary requirement, the strength and ductility of the filler metal should be the basis of selection.

Class RG60 welding rods are probably used most widely. They are generally made of low-alloy steel and are preferred for OFW of carbon and low-alloy steels in the tensile-strength range of 345 to 450 MPa (50 to 65 ksi). Class RG60 rods are most commonly used for welding carbon steel pipes for power plants.

Class RG45 welding rods have a simple low-carbon steel composition. These rods can be used for OFW of carbon and low-alloy steels.

Oxyfuel Gas Welding

Revised by William Ballis, Columbia Gas of Ohio

Welding Techniques

Most OFW is done with the one-pass forehand technique, particularly on thinner materials. A two-pass weld, using the backhand technique for the first pass and the forehand technique for the second pass, is preferred for thicker material (maximum thickness of 4.8 mm, or $\frac{3}{16}$ in.).

Forehand Welding. In forehand welding (Fig. 8), the flame is pointed away from the completed weld in the direction of welding. The torch is held at about 90° to the workpiece. The welding rod is held at an angle of about 30° to the workpiece, and the flame is between the tip of the rod and the weld. The inner cone of the flame is held close to the work without actually touching it. The torch is moved from side to side or oscillated so that the flame heats the welding rod, the edges to be welded, and the weld metal ahead of the flame. The rod may be oscillated in a direction counter to the side-to-side movement of the torch. Both the welding rod-tip and the weld pool should be kept under the shielding influence of the flame. The forehand technique utilizes a zero work angle for both the torch and the welding rod.

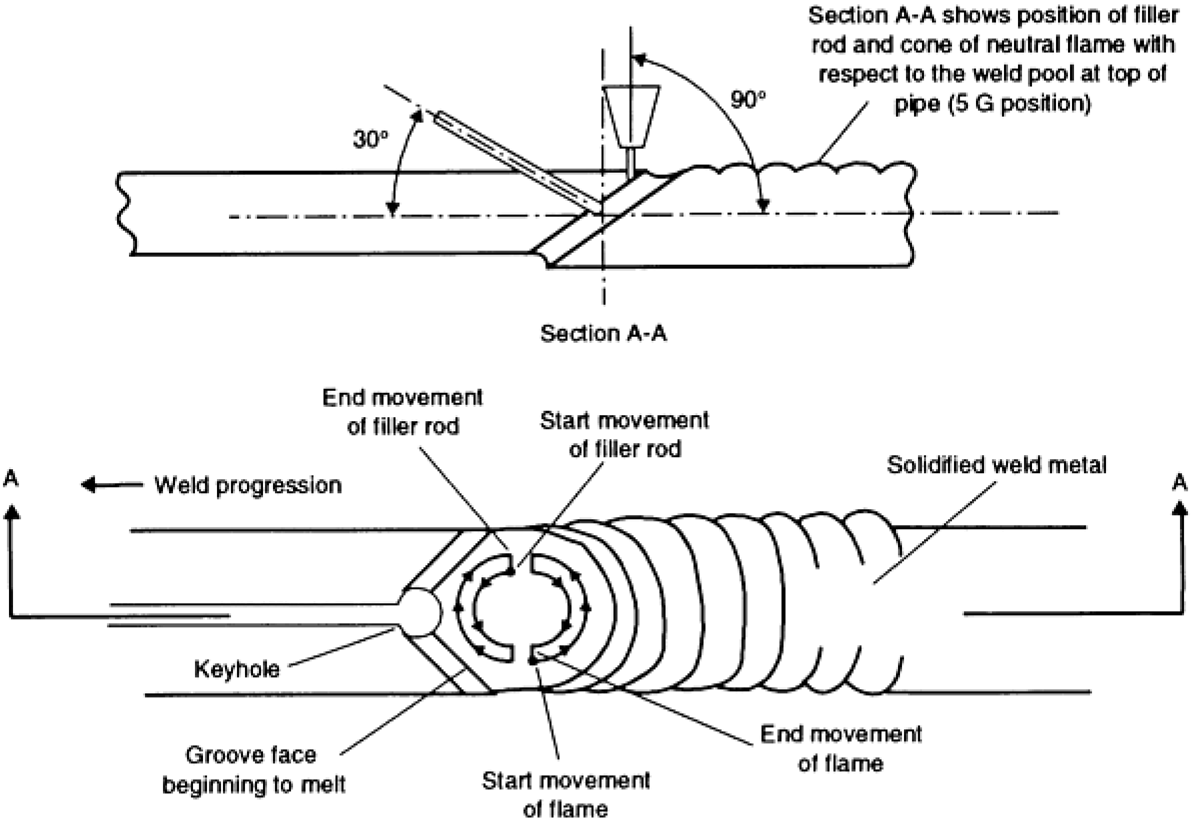
Section A-A shows position of filler rod and cone of neutral flame with respect to the weld pool at top of pipe (5 G position)
30°
90°
Section A-A
End movement of filler rod
Start movement of filler rod
A
Weld progression
Solidified weld metal
A
Keyhole
Groove face beginning to melt
Start movement of flame
End movement of flame

FIG. 8 ORIENTATION AND MOTION OF TORCH RELATIVE TO WORKPIECE IN ONE-PASS FOREHAND OXYACETYLENE WELDING. THE CONE OF FLAME SHOULD BE KEPT CLOSE TO, BUT SHOULD NEVER TOUCH, THE WELD POOL OR PIPE GROOVE FACE. SOURCE: REF 5

Backhand Welding. In backhand welding (Fig. 9), the flame is pointed toward the completed weld. The tip of the welding rod is held between the flame and the weld at an angle of about 30° to the workpiece, and the torch is held at an angle of about 30° to the workpiece. Rod and torch are usually oscillated as shown in Fig. 9. Because of less side-to-side movement of the flame and less melting of the joint edges, a backhand weld is more likely to retain the properties of filler metal unaltered by the base metal. The weld pool receives less agitation, and the weld metal is better protected by the flame.

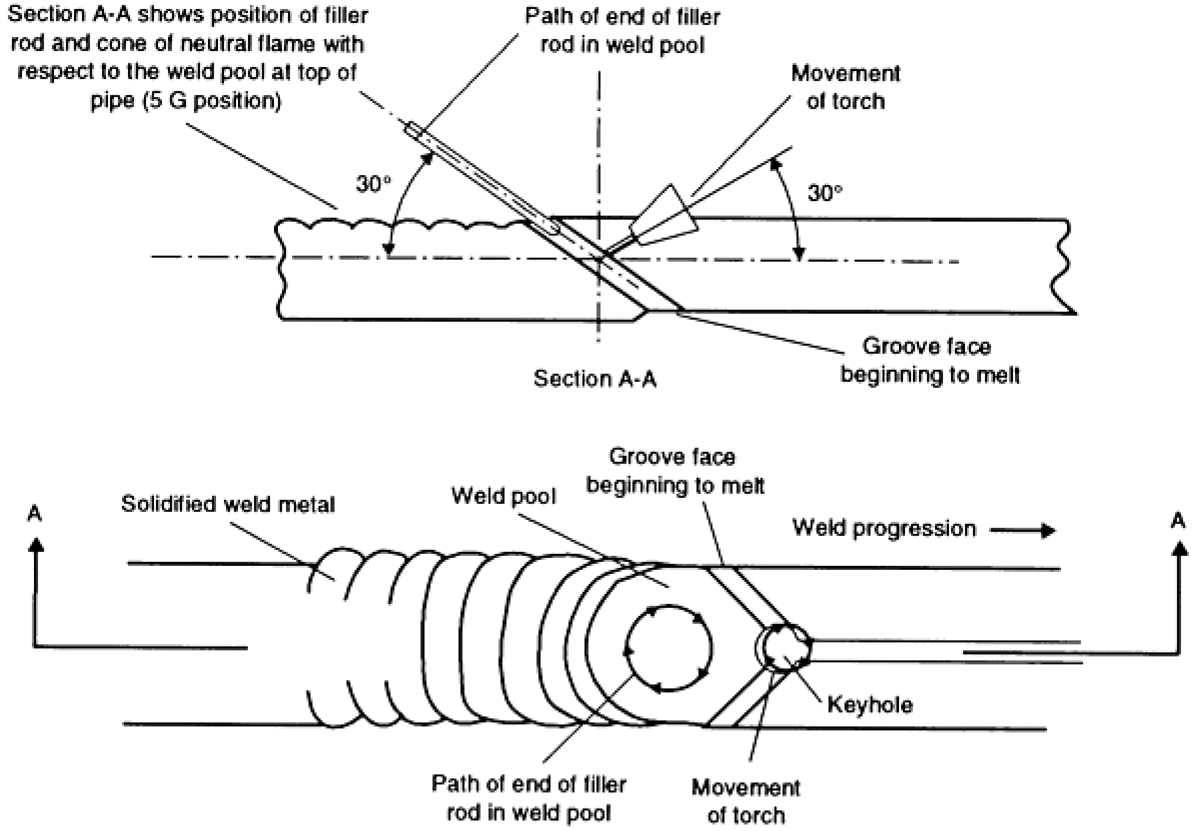
Section A-A shows position of filler rod and cone of neutral flame with respect to the weld pool at top of pipe (5 G position)
Path of end of filler rod in weld pool
Movement of torch
30°
30°
Section A-A
Groove face beginning to melt
A
Solidified weld metal
Weld pool
Groove face beginning to melt
Weld progression
A
Keyhole
Path of end of filler rod in weld pool
Movement of torch

FIG. 9 ORIENTATION AND MOTION OF TORCH RELATIVE TO WORKPIECE IN ONE-PASS BACKHAND OXYACETYLENE WELDING. THE CONE OF FLAME SHOULD BE KEPT CLOSE TO, BUT SHOULD NEVER TOUCH, THE WELD POOL OR PIPE GROOVE FACE. SOURCE: REF 5

Backhand welding is often preferred for joints in the 2G fixed position. Smaller grooves can be used with this method because the flame does not need to be moved around the rod to melt the edges of the groove. Proponents of this technique believe that it saves time, filler metal, and gas. When working with a weld fillet, the flame should be directed primarily at the heavier and continuous section of base metal.

Reference cited in this section

5. "RECOMMENDED PRACTICES AND PROCEDURES FOR WELDING LOW CARBON STEEL PIPE," ANSI/AWS D10.12, AMERICAN WELDING SOCIETY, 1989

Oxyfuel Gas Welding

Revised by William Ballis, Columbia Gas of Ohio

Joint Design and Edge Preparation

Joints used in OFW are butt, lap, edge, T-, and comer joints. Either fillet or groove welds are used, depending on workpiece and strength requirements.

Sheet. Five joints used for single-pass OFW of low-carbon steel sheet are shown in Fig. 10. For OFW, beveling of joint edges is not needed in sheet up to 3.2 mm ($\frac{1}{8}$ in.) thick, and if proper shearing practice has been followed in trimming sheets to size, no special edge preparation is required. However, edges must be free of rust, dirt, oil, and grease.

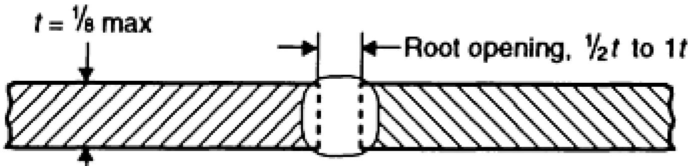

Butt joint
(square-groove weld)

(a)

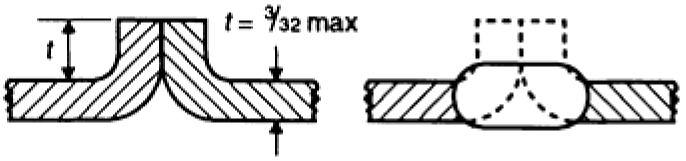

Before welding After welding

Short-flanged-edge butt joint
(melt-through weld)

(b)

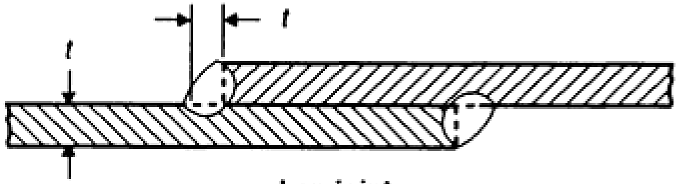

Lap joint
(double fillet weld)

(c)

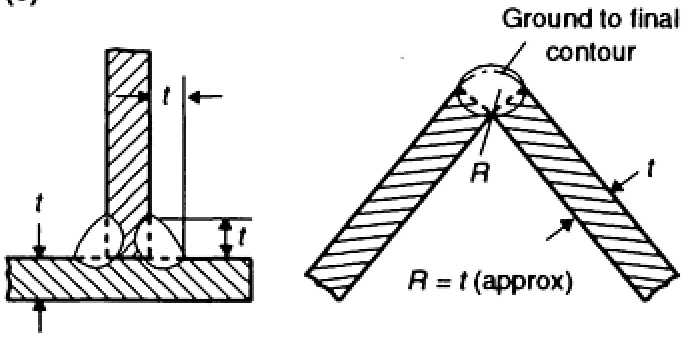

T-joint
(double fillet weld)

Corner-edge joint
(V-groove weld)

(d) (e)

FIG. 10 TYPES OF JOINTS AND RESULTING SINGLE-PASS WELDS OBTAINED WITH OXYFUEL GAS WELDING OF THIN SHEET MATERIAL. MEASUREMENTS IN INCHES

Square-groove butt joints (Fig. 10a) can be single-pass welded from one side in sheet up to approximately 3.2 mm ($\frac{1}{8}$ in.) thick. Complete-penetration welds, properly made, develop the same or greater strength in low-carbon steel as the members joined.

Short-Flanged-Edge Butt Joints. For making butt joints in sheet up to about 2.4 mm ($\frac{3}{32}$ in.) thick, a short flanged edge (Fig. 10b) is beneficial. The weld is called a melt-through weld, because the flanges are melted down to form the butt joint. Filler metal is not required. The oscillating technique for melting down the flanges is shown in Fig. 11. The flanges keep the sheet in flat alignment.

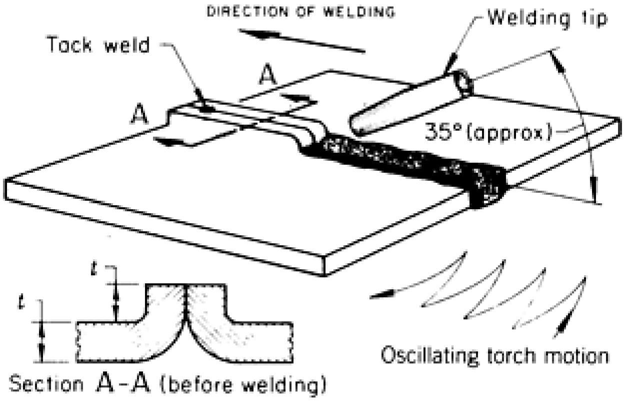

FIG. 11 OXYFUEL GAS WELDING PRACTICE USED TO PRODUCE A SHORT-FLANGED-EDGE BUTT JOINT IN THIN SHEET MATERIAL (≤ 2.4 MM, OR $\frac{3}{32}$ IN., THICK)

Although the sheet has to undergo a special flanging operation, this joint is practical for long production runs when filler-metal deposition and melt-through are difficult to control. High welding speeds can be attained. Sheet thicknesses are limited to approximately 0.8 mm ($\frac{1}{32}$ in.) because short flanges are difficult to bend in heavier material. In other flanged-edge joints, the weld is deposited on the edges, and the flanges retain their structural identity.

Without support from adjacent sheet, long sections of thin material cannot be welded without buckling and distortion. The amount of distortion resulting from expansion during welding and contraction during cooling of the welded sheet-metal section or structure is proportional to the amount of heat applied to the metal as well as to its inherent stiffness and rigidity. Jigs and fixtures are essential to counteract the adverse effects of welding heat. Figure 12(a) shows the elements of a jig used for making butt or flange welds in thin sheet.

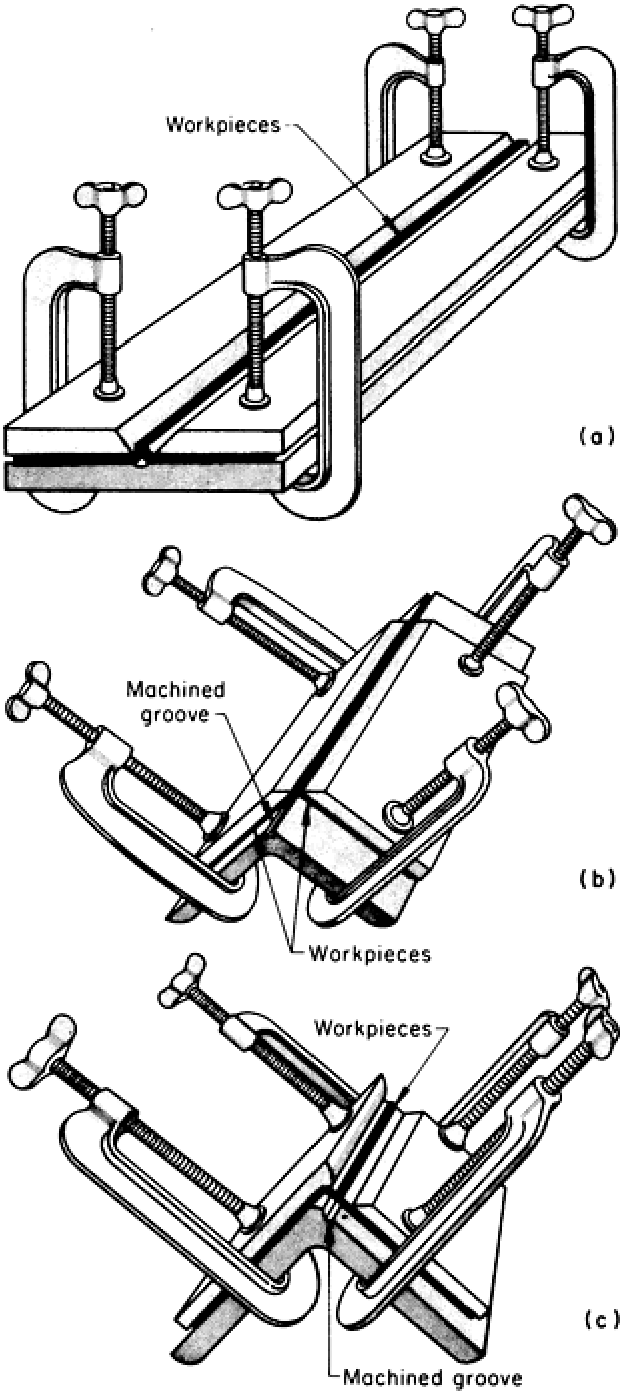

FIG. 12 APPLICATION OF JIGS TO MAINTAIN FLATNESS AND TO MINIMIZE DISTORTION WHEN WELDING THIN SHEET MATERIAL

Lap and T-joints, shown in Fig. 10(c) and 10(d), are single or double fillet welded, depending on strength requirements. A single fillet weld should not be used if the joint is likely to open under service loads. The amount of overlap in lap joints varies with design requirements and sheet thickness. For double-welded lap joints, a minimum overlap of approximately 25 mm (1 in.) is usually adequate for sheet.

The size of fillet weld, with size defined as the length of the shortest side of the fillet, is generally equal to the thickness of the sheet, because this size weld is easily produced in a single pass and has a good appearance.

Both lap and T-joints are somewhat more difficult to weld in thin sections than in thick sections. Danger of melt-through exists when a welder attempts to penetrate to the root of the joint.

Corner-edge joints (Fig. 10e) are used in thin material, mainly because of their simplicity and low cost. In addition, skillful OFW can produce a weld of good external appearance in a single pass.

To improve internal appearance and to provide added strength, an internal pass or removable ceramic backing is needed. Square comer joints in thin material can be made with good appearance and strength and with little or no distortion by using a jig as shown in Fig. 12(b). A recess is machined or ground in the corner of the backing plate to provide for root reinforcement of the V-groove weld.

When parts with rounded corner joints are designed in thin sheet, corners should be bent rather than joined. Joints next to corners can be made by butt welding in a jig constructed as shown in Fig. 12(c). Note the use of a recess in the backing member to provide for root reinforcement of the weld.

Plate. Butt, lap, T-, and corner joints are oxyfuel gas welded in steel plate up to 4.8 mm ($\frac{3}{16}$ in.) thick.

Butt joints welded from one side only require beveled edges in thicknesses 3.2 mm ($\frac{1}{8}$ in.) and over.

Fillet Welds. When compared to groove welds, fillet welds have advantages and limitations. No edge preparation such as chamfering or beveling of joint edges is required. If the weld serves mainly to hold structural members together, or if forces transmitted by the weld are low, fillet welds can be relatively small and economical.

Groove welds are used in butt joints and can be used in lap, T-, and corner joints as an alternative to fillet welds. In low-carbon steel, properly made complete-penetration groove welds develop strength equal to the strength of the members joined.

Typical grooves used for OFW of butt, T-, and corner joints in steel plate with thicknesses of 4.8 mm ($\frac{3}{16}$ in.) or less are shown in Fig. 13. Grooves for OFW are the same as those for arc welding, except that the root opening is larger.

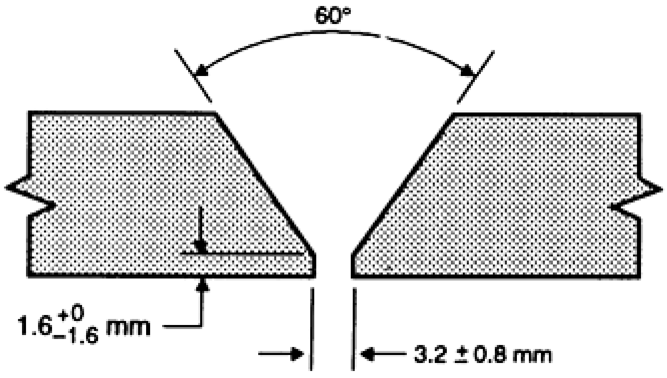

FIG. 13 USE OF GROOVES TO FACILITATE OXYFUEL GAS WELDING OF ≤ 4.8 MM (≤ $\frac{3}{16}$ IN.) THICK PLATES

Pipe. All OFW of low-carbon steel pipe is done with the oxyacetylene process. The techniques used are the one-pass forehand technique, the one-pass backhand technique, or the two-pass technique using a backhand first pass with a forehand second pass. Most production oxyacetylene welding (OAW) in the horizontal fixed (5G) or rolled position (1G) is done using the two-pass technique. When the pipe is fixed in the vertical position (2G), most welders use the one-pass backhand technique. Oxyacetylene welding is cost effective for pipe with diameters of 150 mm (6 in.) and less and with wall thicknesses of 4.8 mm ($\frac{3}{16}$ in.) or less.

Most OAW of low-carbon steel pipe is done using a neutral flame and a work angle of 0°. The travel angles for the torch and welding rod vary with position (Ref 3).

Full-Penetration Welds. If a full penetration weld is desired, a keyhole must be maintained at all times. The torch and rod movements are shown in Fig. 8 for the one-pass forehand method and in Fig. 9 for the one-pass backhand method.

Tack welds should be small and have the general appearance shown in Fig. 14. Both ends of the tack should have a keyhole. The start end of the tack is reheated to form the start keyhole in order to obtain completed fusion when the first pass is connected with the tack weld.

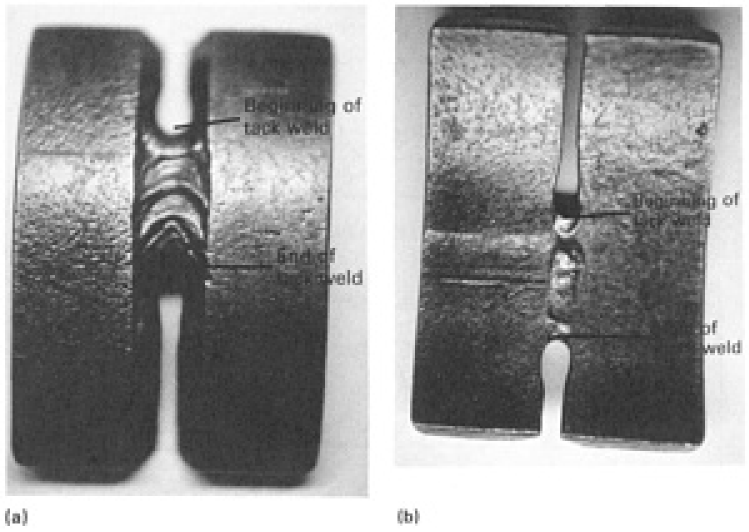
Beginning of tack weld
End of tack weld
Beginning of tack weld
of weld
(a)
(b)

FIG. 14 TACK WELD IN API 5L STEEL PIPE. PIPE DIAMETER, 150 MM (6 IN.) WALL THICKNESS, 7 MM (0.28 IN.). (A) OUTSIDE DIAMETER VIEW. (B) INSIDE DIAMETER VIEW

The 1G, 2G, and 5G positions for welding pipe are shown in Fig. 15.

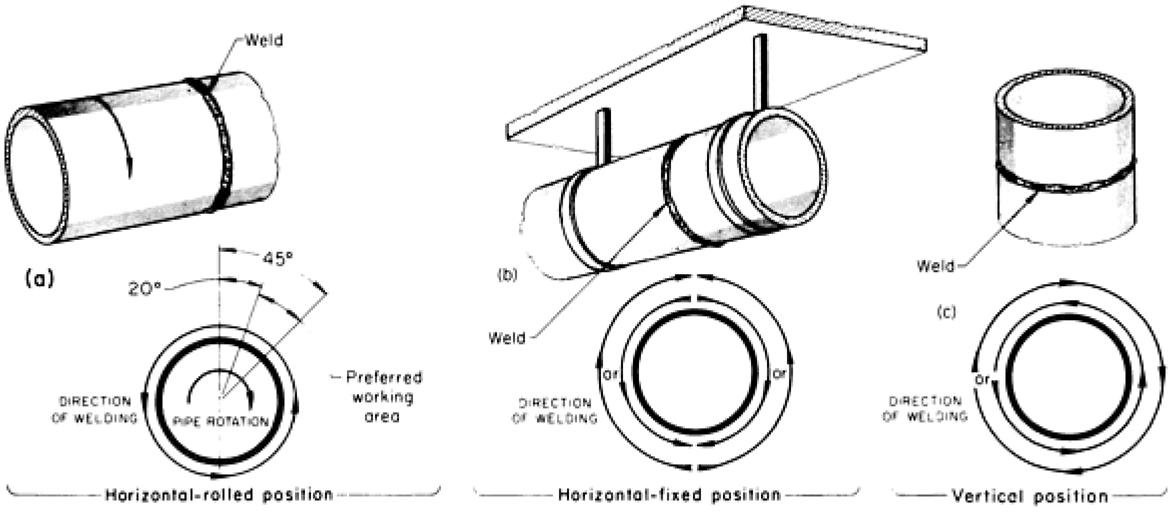
Weld
(a)
45°
20°
Preferred working area
DIRECTION OF WELDING
PIPE ROTATION
Horizontal-rolled position
(b)
Weld
or
or
DIRECTION OF WELDING
Horizontal-fixed position
Weld
(c)
or
DIRECTION OF WELDING
Vertical position

FIG. 15 TYPICAL WELDING POSITIONS FOR OXYFUEL GAS WELDING OF PIPE

Applications. The oxyacetylene process can be used to join low-carbon steel pipe having a maximum diameter of 150 mm (6 in.) and a maximum wall thickness of 4.8 mm ($\frac{3}{16}$ in.). Because of the slow travel speed and close adjustment of welding heat that characterize OAW, small-diameter steel pipe can be oxyacetylene welded faster and with better quality than can be obtained by using most other welding processes.

One of the reasons that the oxyacetylene process is seldom used today is that it is becoming a lost art. There are very few welders that are knowledgeable in the techniques used to oxyacetylene weld steel pipe. Good sources of information on oxyacetylene welding practices are manufacturers of oxyacetylene torch equipment, the American Welding Society, and welding textbooks.

Thin Sheet. Techniques used in OFW of thin sheet depend on joint design and welding position, as well as on the style used by the welder. Although forehand and backhand techniques can be used for OFW of thin sheet, forehand welding is often preferred because the flame points away from the completed weld. This results in less heat buildup and less risk of distortion and melt-through. Proponents of the forehand technique also believe that thin sections can be forehand welded faster using flat or horizontal positions because the flame forces the weld metal to flow in the direction of welding. Figure 11 shows forehand welding of a short-flanged-edge butt joint.

Assemblies of steel sheet less than 3.2 mm ($\frac{1}{8}$ in.) thick are often made by OFW. Because of the relatively long welding time involved, OFW does not permit critical control over total heat input, and distortion can occur when long, continuous welds are made in large, flat thin sheets. For this reason, OFW of thin sections should be done on small parts, where welds can be short, or on sections that have been strengthened by forming. Welding fixtures are often used to prevent distortion and to maintain alignment of joint edges.

The problems arising from difficulties in heat control in joining thin sections by OFW are sometimes solved by using other metal-joining processes:

- SHORT CIRCUITING GAS METAL ARC WELDING
- PULSED ARC WELDING
- GAS TUNGSTEN ARC WELDING
- PLASMA ARC WELDING
- ELECTRON-BEAM WELDING
- ULTRASONIC WELDING
- RESISTANCE WELDING
- BRAZING
- BRAZE WELDING

Products made by OFW of thin sheet use materials ranging in thickness from 0.607 mm (0.0239 in., or 24 gage) to 3.038 mm (0.1196 in., or 11 gage). Typical items include truck bodies, furniture, office equipment, utensils, enclosures, and refrigeration equipment.

Reference cited in this section

3. "SAFETY IN WELDING AND CUTTING," ANSI/AWS ZA9.1, AMERICAN WELDING SOCIETY, 1983

Oxyfuel Gas Welding

Revised by William Ballis, Columbia Gas of Ohio

Repairs and Alterations

Oxyfuel gas welding is frequently used for repairs and alterations because the equipment is portable, welding can be done in all positions, and acetylene and oxygen are readily available. Also, gas cutting, braze welding, brazing, or flame heating is often needed, and the same personnel, equipment, and gases required for these operations can be used for OFW.

Repair Welding. Edge preparation for repair welding generally involves cutting a V-groove to expose clean metal down to the root of the joint. The groove must be large enough to allow room for manipulation of the torch and welding rod.

On many repair welding jobs, sections that must be rebuilt to specific contours or dimensions require backing to support the weld metal. Carbon paste, which can be molded, or carbon plate, which can be easily machined, can be used for backing (Fig. 16). To avoid carbon pickup, copper or a thick steel plate can be used. Welding of the backing to the part should be prevented.

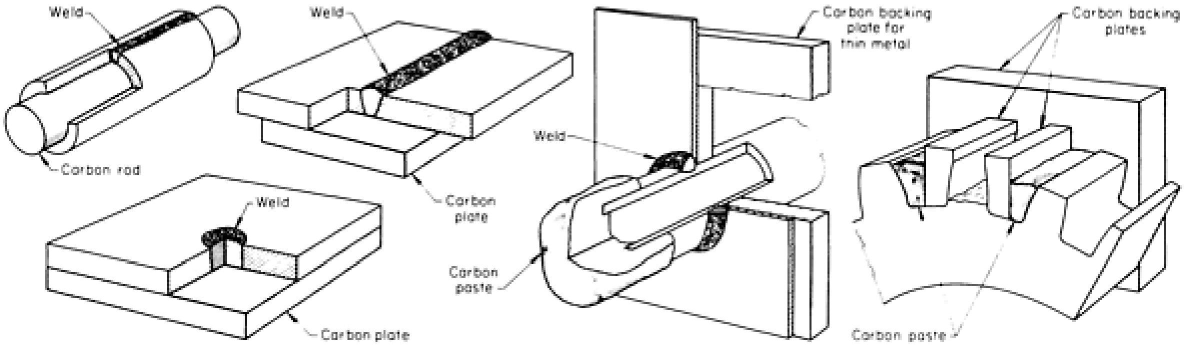

FIG. 16 USE OF CARBON BACKING TO PROVIDE SUPPORT WHEN REPAIR WELDING SURFACES AND CONTOURS

In weld repairing of parts that have cracked, the crack should be fully dressed out. Metal can be removed by chipping, grinding, machining, arc air gouging, or oxyfuel gouging. For OFW, oxyfuel gouging is convenient, because a cutting torch is generally available and only a special gouging tip is needed.

When using oxyfuel gouging for dressing cracks, the crack remains visible until it is completely removed. Grinding may result in a fine crack being smeared over, and chipping is also likely to cover the crack. The disadvantage in using oxyfuel gouging is that the heat input may cause the crack to propagate through differential expansion in the workpiece. The risk is especially great if some of the stresses that caused the crack are still present. Drilling a 6.4 to 13 mm ($\frac{1}{4}$ to $\frac{1}{2}$ in.) hole at the end of the crack may prevent the crack from propagating. Cleanliness of the joint is also a prerequisite for OFW. Rust, scale, dirt, grease, oil, paint, and slag must be removed from the joint areas. The soundness of the weld metal depends in large part on the care used in cleaning. To ensure complete crack removal, liquid-penetrant or magnetic-particle testing can be used.

Workpiece Alterations. When OFW is used to correct production parts that would be rejected because of a small error or defect in manufacture, it must be determined if the salvaged part, after welding, is as good as a new part for the required service. Some fabricating codes identify the defects that may be rectified by welding and define the standards for repair. Rectification may include correction of defects, building up of undersize parts, and filling in of drilled holes.

Preheating and Postheating

Preweld and postweld heat treatments are not usually required to reduce the hardness of low-carbon steel, but either or both treatments are beneficial and can be effective in avoiding or reducing distortion. Most steels that are oxyacetylene gas welded fall within the low-carbon range.

Oxyacetylene gas welding distributes a large amount of heat over a wide area. This results in relatively slow cooling rates and relatively low stress gradients, which, in turn, reduce the degree of hardening and the magnitude of residual stresses that are usually associated with welding heat cycles.

Example 1. Use of Oxyacetylene Welding to Eliminate Postweld Tempering. Aircraft landing gear side stay assemblies made of 1330 steel (Fig. 17) were joined in the form of a trapezoid. The assembly was originally welded by the shielded metal arc welding process; E7016 electrodes were used. Several assemblies cracked in a tube adjacent to a weld after a short time in service. Examination revealed high hardness (50 HRC) in the heat-affected zone (HAZ) was caused by the chilling effect of the large mass of metal in the end fittings. A postweld tempering treatment at 650 °C (1200 °F) to reduce the high hardness could not be used because the tubes became distorted.

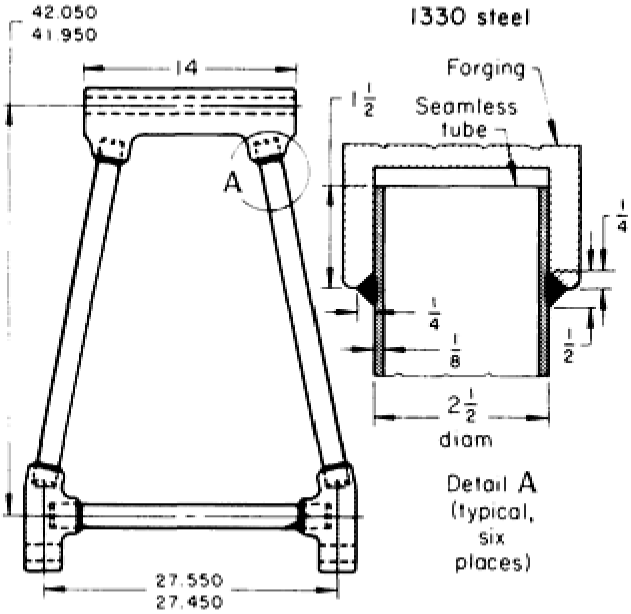

FIG. 17 OXYACETYLENE WELDING OF AIRCRAFT LANDING GEAR SUPPORT. USE OF OAW REDUCED HAZ HARDNESS AND ELIMINATED NEED FOR POSTWELD TEMPERING. MEASUREMENTS SHOWN IN SCHEMATIC ARE IN INCHES.

Local stress relief of the welded joint with a gas torch was successful but laborious, and it was finally decided to oxyacetylene gas weld the assembly. Hardness tests in the area of the welded joint showed that the maximum hardness in the HAZ was 285 HB (29 HRC). This was only moderately higher than the hardness of the end fittings and tubes, which had a hardness of 223 HB (98 HRB). The improvement was believed to result from the wider area heated during OAW, which produced a lower temperature gradient.

Bridging Gaps in Poor Fit-Ups

The ability to control the flow of molten weld metal in OFW provides a high degree of versatility in bridging large gaps caused by a poor fitup. By manipulating the gas torch, weld-metal temperature can be held to the minimum, and the pressure of the gas flame can be used to help support the weld pool in any position, including overhead.

Safety

The gases used in OFW can form explosive mixtures if improperly handled, and burning of these gases must be controlled within recommended limits for the safety of the welder and surrounding property. Workers must be familiar with safe welding practices and should be aware that any deviation from recognized safety standards can result in damage to apparatus or in personal injury. Workers in charge of welding operations should be instructed and judged competent by their employers.

Rules and instructions covering safety, operation, and maintenance procedures are readily available. Federal, state, and local governments, in addition to equipment manufacturers, gas suppliers, and regulatory agencies, have cooperated in publishing documents on safe welding practices.

Additional safety guidelines and reference sources for further information are provided in the article "Safe Practices" in this Volume.

References

1. H.B. CARY, *MODERN WELDING TECHNOLOGY,* PRENTICE-HALL, 1989, P 211
2. "OXYGEN-FUEL GAS SYSTEMS FOR WELDING AND CUTTING," BULLETIN NO. 51, NATIONAL

FIRE PROTECTION ASSOCIATION, 1977
3. "SAFETY IN WELDING AND CUTTING," ANSI/AWS ZA9.1, AMERICAN WELDING SOCIETY, 1983
4. H.B. CARY, *MODERN WELDING TECHNOLOGY,* PRENTICE-HALL, 1989, P 210
5. "RECOMMENDED PRACTICES AND PROCEDURES FOR WELDING LOW CARBON STEEL PIPE," ANSI/AWS D10.12, AMERICAN WELDING SOCIETY, 1989

Thermite Welding

Bernard Schwartz, Norfolk Southern Railway Company

Introduction

THERMITE WELDING (TW) is a fusion welding process in which two metals become bonded after being heated by superheated metal that has experienced an aluminothermic reaction. The liquid metal that results from the reaction between a metal oxide and aluminum acts as the filler metal. This exothermic process was discovered in 1898, in Germany, by Dr. Hans Goldschmidt of Goldschmidt AG.

In the United States, this process was used in the early 1920s to join trolley car track. In 1933, thermite welding was used on the Delaware and Hudson Railroad at Albany, NY to produce long rail sections. In 1938, a researcher at the Electric Railway Improvement Company invented a copper-base aluminothermic process for welding copper conductors to steel rails. Currently, thermite welding is widely used in the field welding of track, where its portability and versatility are strong assets.

It should be noted that the standard term for this process is often listed as "thermit" welding. However, Thermit is a registered trademark of Th. Goldschmidt AG of Essen, Germany. Therefore, this Volume uses the generic term, "thermite."

Thermite Welding

Bernard Schwartz, Norfolk Southern Railway Company

Thermite Welding Principles

The aluminothermic reaction that occurs in thermite welding follows the general formula:

$$\textit{metal}\text{ oxide} + \underset{\text{(reducing agent)}}{\text{Aluminum}} \rightarrow \underset{\text{oxide+metal+heat}}{\text{Aluminum}}$$

whereas the typical reaction that occurs in the thermite welding of rails is:

$$FE_2O_3 + 2AL \rightarrow 2FE + AL_2O_3 \text{ (HEAT } \rightarrow 850 \text{ KJ)}$$

and the typical reaction that occurs in the welding of copper conductors to steel rails is:

$$3CU_2O + 2AL \rightarrow 6CU + AL_2O_3 \text{ (HEAT } \rightarrow 1060 \text{ KJ)}$$

This exothermic reaction is extremely violent if only the metal oxide and aluminum reducing agent are used. Pellets of ferroalloy are added to cool this reaction from a typical temperature of 3090 °C (5600 °F) to 2480 °C (4500 °F). These additions also are used to produce the desired chemistry. The amount of added alloy is very critical, because larger amounts will cool the reaction to temperatures below 2040 °C (3700 °F), at which point the slag/metal separation could be incomplete.

Thermite Welding

Bernard Schwartz, Norfolk Southern Railway Company

Thermite Welding Metallurgy

Metallurgical structures that are present in thermite welds depend on the chemical composition of the weld metal and on the cooling rate of the joint after pouring is completed.

Figure 1 shows a typical macrostructure of a carbon steel rail thermite weld and the microstructure of the fusion zone, the fusion line, the end of the heat-affected zone (HAZ), and the unaffected rail. Typically, the weld is 100% pearlite with varying degrees of coarseness. This is due to the slow cooling rate that completes transformation before reaching the martensite start temperature and the formation of untempered martensite.

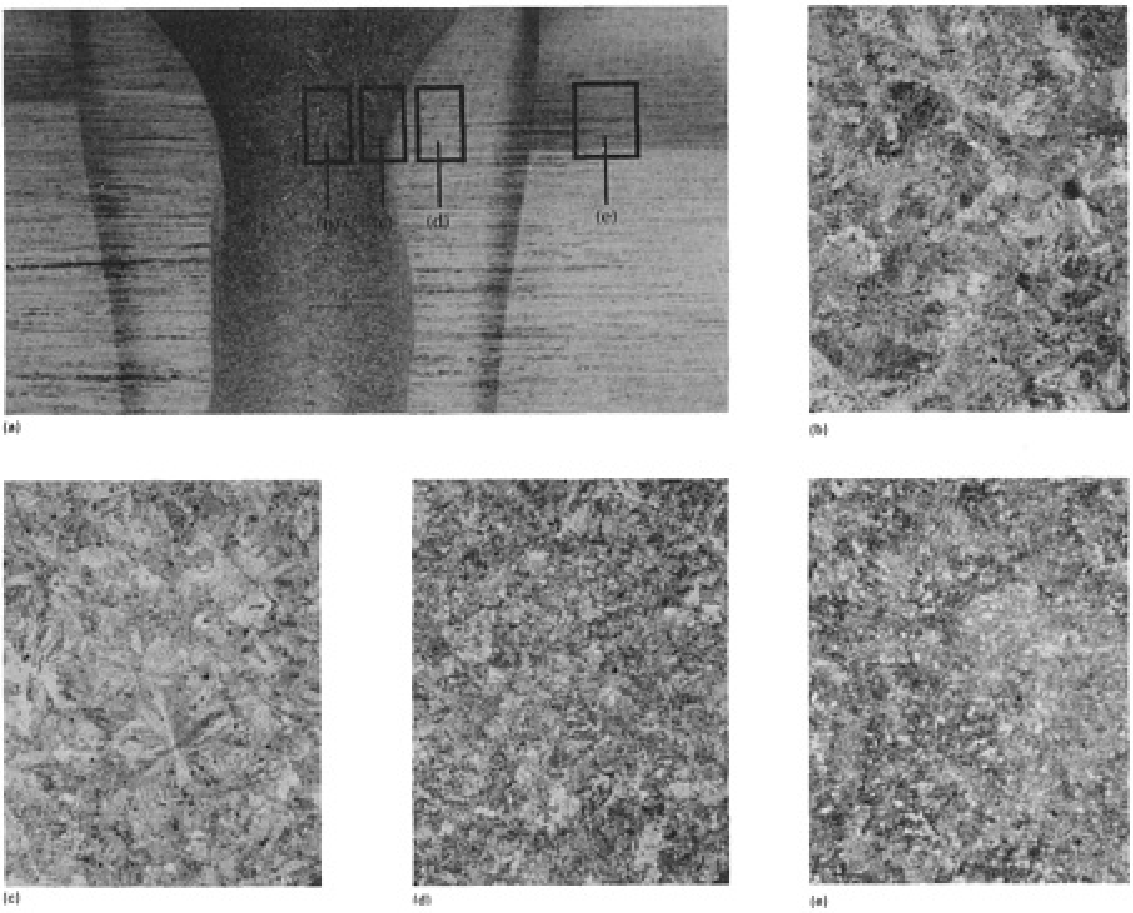

FIG. 1 CARBON STEEL RAIL THERMITE WELD. (A) MACROSTRUCTURE. (B) WELD MATERIAL. 65×. (C) FUSION LINE AREA. 65×. (D) HEAT-AFFECTED ZONE. 65×. (E) UNAFFECTED RAIL AREA. 65×

Thermite Welding

Bernard Schwartz, Norfolk Southern Railway Company

Applications for Thermite Welding

Rail welding is the most widely used application of the thermite welding process. The use of that process, as well as the electric flash butt welding process, has eliminated joint bars (mechanical fasteners) and greatly lessened track maintenance. Long sections of continuous welded rail (CWR) (typically, 440 m, or 1440 ft) that are welded by the flash butt welding process are transported to the track site and set into place. Then, the CWR sections are joined by thermite welding. A short preheating process is typically used in the thermite welding of rail.

The thermite process is also utilized to replace rail defects, to weld insulated joint assemblies into track, and to make electrical bonding connections using copper conductors between rails that will not be welded together.

Process Description. The thermite welding of rails, including a short preheat, involves the following basic operations:

- CUT OR SPACE THE RAILS TO BE WELDED SO THAT THE GAP SPECIFIED BY THE MANUFACTURER IS ACHIEVED
- REMOVE SLAG, MOISTURE, FOREIGN MATTER, AND SUCH FROM THE AREA TO BE WELDED
- ALIGN THE RAIL ENDS, BOTH VERTICALLY AND HORIZONTALLY
- CENTER THE THERMITE MOLDS OVER THE RAIL GAP AND SEAL THE GAPS WITH LUTING MATERIAL
- SITUATE THE PREHEATING EQUIPMENT, MOLD CLAMPS, AND THERMITE CRUCIBLE IN POSITION ABOVE THE JOINT GAP AND THE MOLD OPENING (FIG. 2)
- ADJUST THE PREHEATING EQUIPMENT TO ACHIEVE THE MOST UNIFORM HEATING OF THE RAIL ENDS FROM ONE SIDE TO THE OTHER AND FROM TOP TO BOTTOM
- IGNITE THE THERMITE CHARGE AND ALLOW IT TO TAP INTO THE MOLD, THE TIMING OF WHICH IS USUALLY CONTROLLED BY A SELF-TAPPING THIMBLE. THE LIQUID METAL FLOWS ONTO A DIVERTER PLUG TO REDIRECT THE FLOW OF LIQUID METAL INTO THE SIDE SPRUES SO THAT THE CASTING IS PRODUCED FROM THE BOTTOM UPWARD
- ALLOW THE PROPER SOLIDIFICATION TIME, AS RECOMMENDED BY THE MANUFACTURER
- REMOVE THE MOLD MATERIAL AND FINISH-GRIND THE WELD TO RAILROAD SPECIFICATIONS
- CONDUCT A FINAL INSPECTION OF THE WELD, BOTH VISUALLY AND WITH THE AID OF ULTRASONIC EQUIPMENT

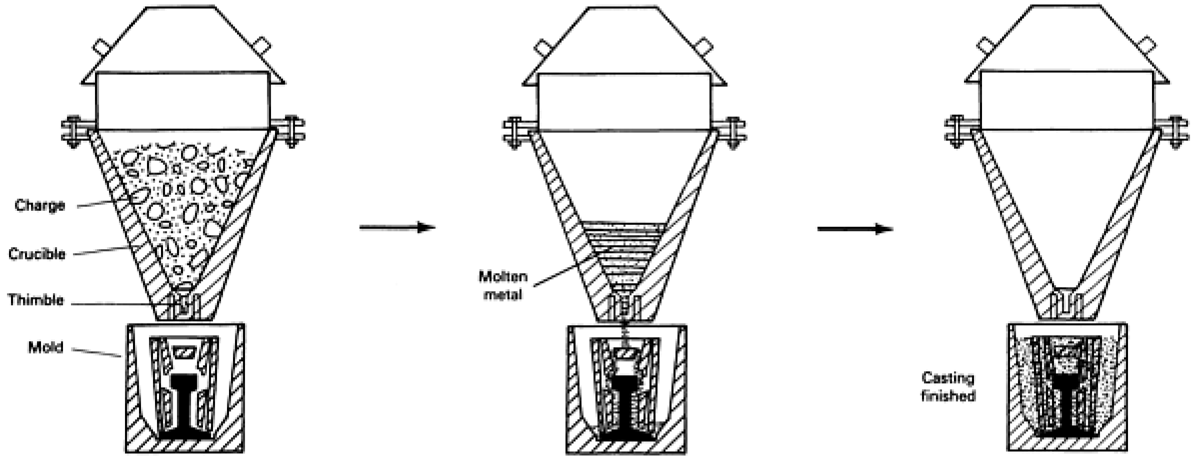

FIG. 2 TYPICAL CRUCIBLE-MOLD SETUP FOR RAIL THERMITE WELDING

Electrical Connections. The thermite welding process also is used to weld electrical conducting joints, particularly to provide electrical continuity for railroad signal systems. A copper oxide powder is reduced by aluminum to form a metallurgical bond between the steel rail and the copper conductor. The resulting joint has an excellent current-carrying capacity and the ability to withstand corrosion, when compared with a mechanical bond.

Repair Welding. Large-diameter rolls, shafts, ingot molds, and heavy mill housings can be successfully repaired by using the thermite welding process. The metallurgical aspects of the process, when used to make repair welds, are similar to those of rail welding, except for the large volume of the weld. The primary difference is that a customized mold, which is usually manufactured on-site, is used. A wax pattern is built around the weld cavity, and a collar is shaped in wax to overlap the base metal. Sand is hand rammed around the pattern within a mold box. Then, sufficient gates and risers are incorporated into the system.

New Railroad Applications. Recent developments include:

- HIGHER-STRENGTH THERMITE MIXTURES THAT MORE CLOSELY MATCH THE HARDNESS OF CURRENT HIGH-STRENGTH RAILS (RAIL HARDNESS HAS STEADILY INCREASED TO 300 HB, MINIMUM, FOR STANDARD CARBON RAILS AND TO 360 HB, MINIMUM, FOR PREMIUM RAILS)
- ONE-USE CRUCIBLES THAT REDUCE OR ELIMINATE WELD-METAL CONTAMINATION BY THE RESIDUAL ELEMENTS AND IMPURITIES THAT REMAIN IN CONVENTIONAL REUSABLE CRUCIBLES
- POSTWELD HEAT TREATMENT THAT REFINES THE GRAINS IN OLDER WELDS AND IMPROVES THE STRUCTURAL PROPERTIES OF THE THERMITE WELD TO THE LEVEL OF THOSE OF ELECTRIC FLASH BUTT WELDS
- A HEAD WASH REPAIR WELD THAT CAN REPAIR DISCONTINUITIES ON THE RUNNING SURFACE OF THE RAIL HEAD. (IN THE HEAD WASH REPAIR, RAIL SURFACE DEFECTS OR DISCONTINUITIES ARE TORCH WASHED AWAY DOWN TO THE SOUND METAL. THIS OCCURS OVER A VERY SHORT TIME SPAN. THE TORCH WASH AREA IS CLEARED, A THERMITE MOLD IS APPLIED, AND THE AREA IS PREHEATED AND THEN THERMITE WELDED TO PRODUCE A COMPLETE RAIL HEAD OF SOUND METAL.)
- WIDE-GAP REPAIR WELDS (50 AND 75 MM, OR 2 AND 3 IN.) THAT CAN BE UTILIZED TO WELD WIDER GAPS AND TO ELIMINATE SPECIFIC RAIL DEFECTS, WITHOUT REQUIRING THE INSTALLATION OF A PLUG RAIL
- NEW MOLD CONFIGURATIONS (REPOSITIONED RISERS, METAL GUSSETS, AND SO ON) THAT MAY INCREASE THE MECHANICAL PROPERTIES OF RAIL THERMITE WELDS
- A ZIRCON-BASE MOLD WASH COATING FOR THERMITE WELD MOLDS THAT IMPROVES THE SURFACE CONDITION OF RAIL THERMITE WELDS, THEREBY IMPROVING FATIGUE LIFE

Thermite Welding

Bernard Schwartz, Norfolk Southern Railway Company

Thermite Welding Safety

As already noted, the thermite welding process produces a violent reaction and a large release of heat. The presence of moisture in the charge, crucible, molds, or tooling should be avoided in order to prevent the rapid release of steam and the resulting ejection of molten metal from the crucible. The thermite charge and molds must be stored in a dry location. Tooling should be preheated prior to use, and the surrounding work area should be dry and free of combustible materials.

Appropriate protective clothing, safety boots, gloves, safety glasses, or a full face shield with filter lenses should be worn by all personnel in the immediate vicinity.

Thermite Welding

Bernard Schwartz, Norfolk Southern Railway Company

Selected References

- *METALS HANDBOOK,* 9TH ED., VOL 6, *WELDING, BRAZING, AND SOLDERING,* AMERICAN SOCIETY FOR METALS, 1983, P 692-704
- RAIL, CHAPTER 4, *AREA MANUAL,* AMERICAN RAILWAY ENGINEERING ASSOCIATION, 1980, P 4-2-6.7 TO 4-2-7
- *WELDING HANDBOOK,* 8TH ED., VOL 2, *WELDING PROCESSES,* AMERICAN WELDING SOCIETY, 1991, P 892-900

High-Temperature Solid-State Welding

W. Lehrheuer, Forschungszentrum Jülich GmbH

Introduction

HIGH-TEMPERATURE SOLID-STATE WELDING is used to join members of similar or dissimilar materials by compressing and heating the surfaces to be joined in a suitable atmosphere and by maintaining pressure and temperature until the parts are joined.

The processes that occur on the surfaces to be joined can be categorized by three stages.

Primary Bonding. In this first stage, the surface layers of the pieces being welded are prepared to accept the initial bond. The deformation of rough points will crack brittle surface layers and, thus, will achieve limited local bonding of metallic surfaces, leading to reactions between the atoms of the pieces being welded.

Bond-Surface Extension. In this second stage, the bond surface is enlarged by virtue of plastic flow processes, which are determined by surface loading, and the mechanical properties of the pieces being welded. An expansion of the weld surface area occurs because of diffusion along the surfaces being bonded. The metallic bond becomes improved as a result of the removal of surface films and impurities by dissolution and diffusion into the base material. The diffusion process actually reorients surface atoms, thereby providing a configuration that is more suitable for bonding. Finally, the reaction processes that occur between the atoms (and their electrons) of the enlarged bond surfaces of the pieces being welded result in good adhesion between the components.

Elimination of the Original Joining Surface. In this third stage, the joined zone experiences recrystallization, grain growth, and diffusion that are beyond the characteristics of the original weld interface. The formation of a largely homogeneous weld with normal metallic cohesion is achieved by reactions that occur between the materials being joined.

Because diffusion processes have a significant role in this type of welding, the operation is frequently called "diffusion welding" or "diffusion bonding."

High-Temperature Solid-State Welding

W. Lehrheuer, Forschungszentrum Jülich GmbH

Process Advantages and Disadvantages

The major advantages of the process are that:

- IT ALLOWS THE PREPARATION OF STRESS-RELIEVED AND, IN MANY CASES, PRIMARILY HOMOGENEOUS 0JOINTS ON VERY DIFFERENT MATERIALS.
- IT ALLOWS THE WELDING OF MATERIAL COMBINATIONS THAT CANNOT BE JOINED BY FUSION WELDING. THIS APPLIES TO MATERIALS THAT FORM BRITTLE INTERMETALLIC PHASES AND TO METALS THAT ARE JOINED TO NONMETALS.
- IT ENABLES THE JOINING OF COMPONENTS WITH VERY DIFFERENT JOINING SURFACE GEOMETRIES THAT ARE NOT SUBJECT TO ANY PARTICULAR REQUIREMENTS. HOWEVER, THE PRESSURE APPLICATION METHOD CAN BE A LIMITING FACTOR. MATERIALS ARE JOINED IN THE SOLID STATE. METALLURGICAL CONDITIONS ARE GENERALLY FAVORABLE AND EASY TO INFLUENCE, BECAUSE THERE IS LESS SHRINKAGE AND FEWER STRESSES.
- COMPONENT DEFORMATION IS KEPT WITHIN RELATIVELY CLOSE LIMITS, THEREBY OMITTING THE NEED FOR REWORKING, IN MANY CASES.
- THE PROCESS MAY BE ALMOST FULLY AUTOMATED. THEREFORE, HIGHLY QUALIFIED TECHNICIANS ARE NOT REQUIRED FOR PRODUCTION TASKS.

The disadvantages of the process are:

- HIGH EQUIPMENT COST
- HIGH DEMANDS IN TERMS OF SURFACE CLEANLINESS AND PRECISION, AS WELL AS WELDING ATMOSPHERE
- CONSIDERABLE TIME REQUIREMENTS FOR EXECUTING THE JOINT, EXCEEDING BY FAR THE REQUIREMENTS OF MOST OTHER PROCESSES
- INCREASED CAPITAL COSTS WITH INCREASING COMPONENT SIZE, BECAUSE A WELDING CHAMBER BECOMES NECESSARY
- VERIFICATION OF PROPER JOINT EXECUTION BY NON-DESTRUCTIVE TESTING IS SIGNIFICANTLY IMPAIRED, IN MANY CASES

Although these disadvantages and process features prevent large-scale technical applications that replace welding, the possibilities for using this process to solve special problems are impressive.

High-Temperature Solid-State Welding

W. Lehrheuer, Forschungszentrum Jülich GmbH

Equipment

Mass production is not possible when this process is used. The type of equipment, process temperatures, atmospheres, materials, component sizes, and numbers of workpieces can differ in each application and can result in entirely different systems and devices. The process is frequently conducted in facilities designed for other purposes, such as hot presses, hot isostatic pressing facilities, and vacuum or inert-atmosphere furnaces (with dead-weight loading). The configuration shown in Fig. 1 represents a typical equipment design.

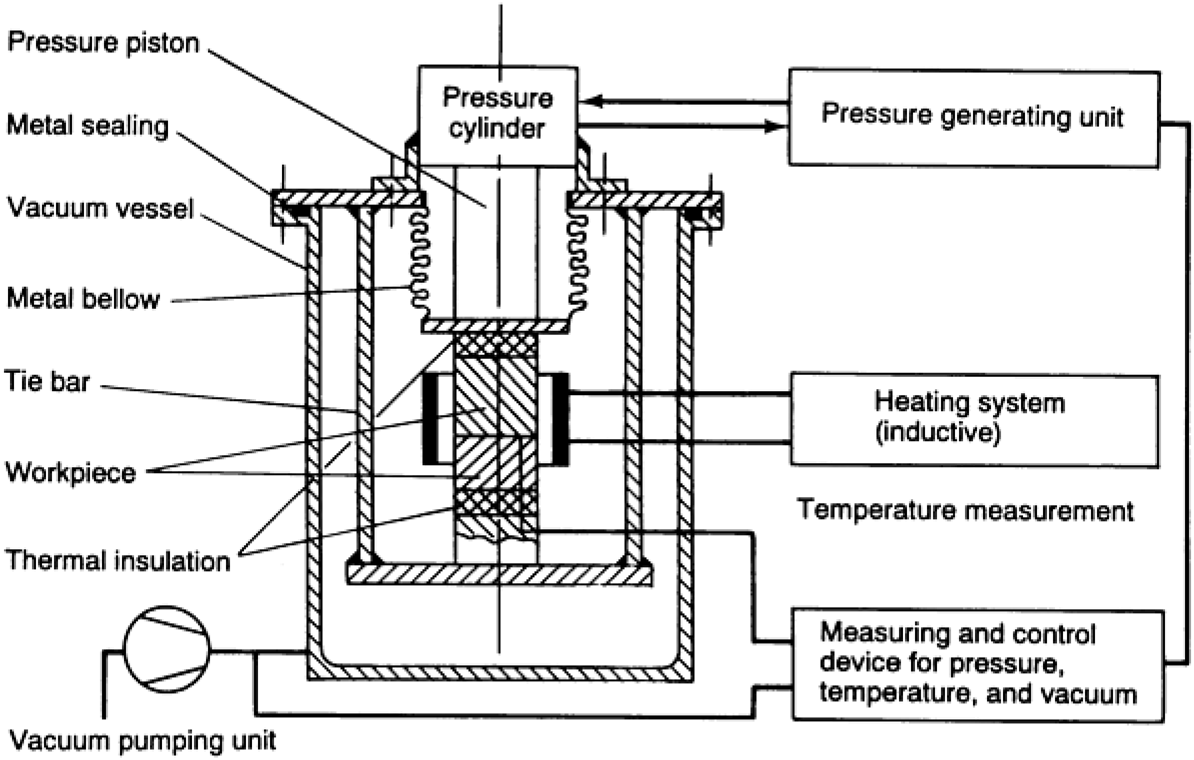
Pressure piston
Metal sealing
Vacuum vessel
Metal bellow
Tie bar
Workpiece
Thermal insulation
Vacuum pumping unit
Pressure cylinder
Pressure generating unit
Heating system (inductive)
Temperature measurement
Measuring and control device for pressure, temperature, and vacuum

FIG. 1 PRINCIPLE OF OPERATION OF TYPICAL HIGH-TEMPERATURE SOLID-STATE WELDING EQUIPMENT

High-Temperature Solid-State Welding

W. Lehrheuer, Forschungszentrum Jülich GmbH

Procedures for High-Temperature Solid-State Welding

The procedures used to make this type of weld can be described in simple terms. Clean surfaces are pressed together in an atmosphere that is suited for the materials being joined and then heated to welding temperatures. The temperature is maintained until welding is completed. The component is then cooled.

The most important process parameters are temperature, pressure, welding time, and welding atmosphere. Figure 2 shows the dependence of the strength of a diffusion weld on temperature, pressure, and time. It should be noted that these three parameters are not independent of one another. Within certain limits, higher temperature can shorten the time required for welding and reduce the pressure, and vice versa.

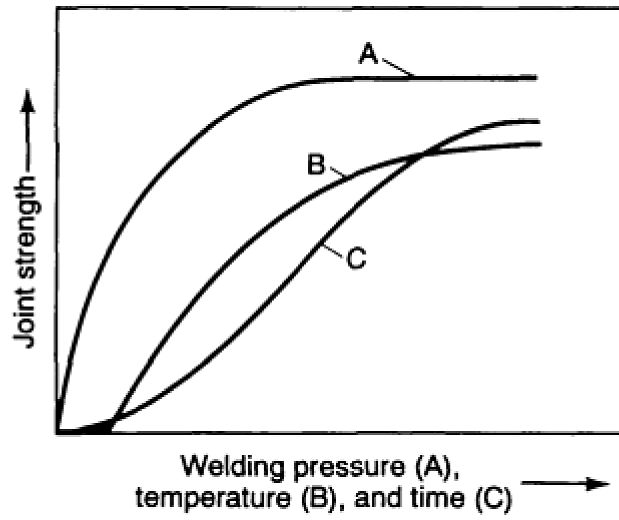

FIG. 2 INFLUENCE OF WELDING PRESSURE, TEMPERATURE, AND TIME ON JOINT STRENGTH

Temperature is the most important welding parameter. The requisite temperature depends on the material or the combination of materials. It must be sufficiently high to ensure plastic flow and intimate contact of the surfaces at the selected surface pressure, as well as adequate diffusion within a reasonable time. As a general rule, a welding temperature ranging from 0.5 to 0.7 T_s is advisable, where T_s is the melting temperature, in K, of the lower-melting-point material or component.

Pressure is required in order to achieve a large contact area at the welding temperature. The pressure level depends on the mechanical properties of the materials being joined. Its upper limit is determined by the admissible degree of deformation. A pressure increase beyond a certain level does not improve strength any further. As a general rule, a surface pressure that ranges from 1 to 20 MPa (0.15 to 2.9 ksi) can be considered adequate for the execution of high-temperature solid-state welds. The deformation at the joint interface, as a result of surface pressure, also has a positive influence on the joint. Surface impurity layers, such as oxides, are cracked, surface roughness is removed, and the atoms in the joint surfaces are brought close together. Plastic deformation facilitates the approach of the metallic atoms by fragmentation and dispersion of surface oxide and impurity barriers. Therefore, it is often better to match a surface of elevated roughness by deformation in the surface region than it is to bond, for example, highly polished surfaces with minor deformation. A precondition, however, is good surface cleaning and the removal of impurity surface layers prior to welding.

The welding time required for the preparation of a proper joint can range from a few minutes to several hours, depending on the material or material combination, as well as on the selected welding temperature.

The welding atmosphere used in most welding systems is a vacuum. In most cases, a vacuum of 0.01 Pa (1.4×10^{-6} psi) is sufficient. The use of high-purity shielding gases, such as argon or helium, is possible for many materials, but generally does not present any technical or economic advantages. Materials that are relatively insensitive to oxygen, such as structural steels, have been successfully welded in air. Welding has also been performed in liquids, such as molten salt.

Prior to welding, it is important to clean the surfaces to be joined and to remove any surface layers, especially oxide films. Although it is not possible to prevent the formation of new layers in air, these newly formed layers are very thin and hardly impair the welding process. Higher welding temperatures and greater deformation on the surfaces reduce the influence of such layers. Surface cleaning can be accomplished by chemical or mechanical methods. Proven techniques are degreasing with a suitable agent or removing the layers with wire brushes immediately prior to welding.

The most metallurgically simple type of weld results when similar materials are joined without using an interlayer material. In this case, the properties of only one material determine the process parameters. Two materials are involved in the rarely practiced joining of similar materials using an interlayer of softer material, which may facilitate surface matching or improve diffusion. An interlayer is more frequently used in the joining of dissimilar materials. When interlayers are used, the welding process involves a metallurgically complicated multicomponent system, even in the case of pure metals. The interlayers primarily serve to bridge different melting points and different coefficients of thermal expansion, improve the diffusion properties, and form favorable structures in the joining zones. Practically any metal combination can he joined using suitable interlayers. Even extreme combinations, such as molybdenum and lead or tungsten and lead, can be processed by selecting several suitable interlayers, although several welding operations at different temperatures would be required.

In addition to welding metals to each other, diffusion welding permits the joining of metals to nonmetals. However, the problems encountered in this case are even more complex, because the chemical and physical properties of the components being welded differ significantly from each other. Nevertheless, welds between metals and glasses, as well as between metals and ceramics, have been successfully produced. The main problems arise from differences in the coefficients of thermal expansion and in the brittleness of the nonmetallic members.

High-Temperature Solid-State Welding

W. Lehrheuer, Forschungszentrum Jülich GmbH

Practical Applications

The high-temperature solid-state welding process has been primarily used on an industrial scale in the fields of aerospace and nuclear engineering. A typical application is the fabrication of complicated hollow structures. Figure 3 shows a cross section of a rotating-anode x-ray tube. The body consists of austenitic stainless steel, whereas the outer welded ring is made of aluminum. Both parts form several cooling channels.

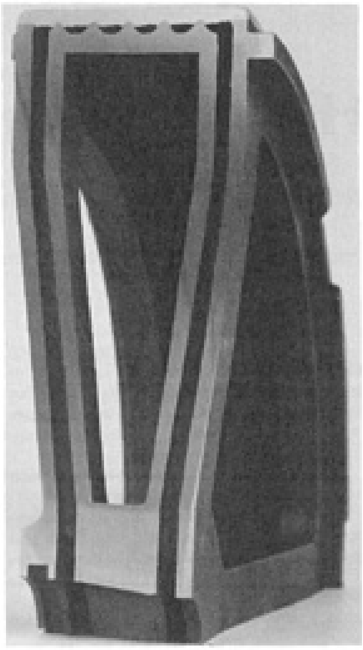

FIG. 3 ROTATING X-RAY ANODE WITH OUTER ALUMINUM RING DIFFUSION WELDED TO STAINLESS STEEL BODY

Another important application is the joining of dissimilar materials. Of particular significance to nuclear engineering applications are joints made of aluminum and stainless steel. Figure 4 shows welds of these materials in different joint geometries. An economical use of this process for mass production frequently is not possible, because of the long welding times and expensive equipment involved. Therefore, applications will probably continue to be limited to special complicated geometries, expensive materials and dissimilar material combinations. In recent years, the combined use of this process with a heat-treatment step has been increasing. Economic efficiency is much easier to achieve with such a process combination. The process is undoubtedly of great significance for joining very dissimilar materials, including metals and ceramics, and the newly developed high-temperature materials, anisotropic materials, and fiber-reinforced materials.

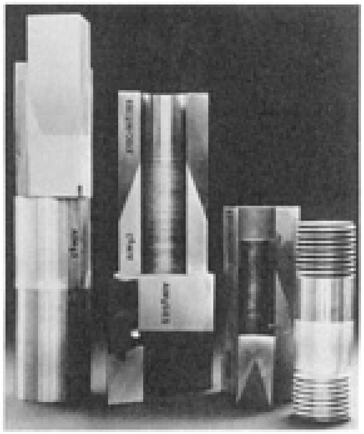

FIG. 4 ALUMINUM-STAINLESS STEEL JOINTS

High-Temperature Solid-State Welding

W. Lehrheuer, Forschungszentrum Jülich GmbH

Selected References

- V.A. BACHIN AND K.E. TSIOLKOVSKY, "DIFFUSION WELDING METALS VERSUS NON-METALS," DVS REPORT 148, DVS-VERLAG, DÜSSELDORF, 1992
- W.D. LUDEMAN, "A FUNDAMENTAL STUDY OF THE PRESSURE WELDING OF DISSIMILAR METALS THROUGH OXIDE LAYERS," UCRL-50744, LAWRENCE RADIATION LABORATORY, UNIVERSITY OF CALIFORNIA, OCT 1969
- N.F. KAZAKOV, VACUUM DIFFUSION WELDING OF METALS AND NON-METALS, *WELD. PROD.,* VOL 11 (NO. 63), 1967, P 83-86
- Y.N. KOPYLOV AND N.F. KAZAKOV, DIFFUSION WELDING OF QUARTZ GLASS TO METALS, *AUT. WELD.,* VOL 6 (NO. 73), 1968, P 77-78
- A.Y. MAKARKIN, M.A. PAVLOVA, AND I.I. METELKIN, WELDING TOGETHER CERAMIC AND METALS, *WELD. PROD.,* VOL 6, 1967, P 17-22
- M.M. SCHWARTZ, "DIFFUSION WELDING," MARTIN MARIETTA CORPORATION, MARCH 1970
- J.K. SOLBERG AND L.K. RUDSRUD, "DIFFUSION BONDING OF 253 MA AUSTENITIC STAINLESS STEEL," DVS REPORT 148, DVS-VERLAG, DÜSSELDORF, 1992
- V.T. SLEPUKHA, FORMATION OF SURFACES FREE FROM OXIDE FILMS IN VACUUM DIFFUSION WELDING, *SVAR. PROIZ.,* VOL 5, 1965, P 38-40
- R.F. TYLECOTE, *THE SOLID-PHASE WELDING OF METALS,* ST. MARTINS PRESS, 1968

Low-Temperature Solid-State Welding

R.S. Rosen, Lawrence Livermore National Laboratory

Introduction

SOLID-STATE WELDING is the joining of base metals (metals, alloys, ceramics, etc.) by chemical bond formation without heating the base metals above their respective melting points. The advantages of solid-state welding processes include the ability to join parts to a finished size or shape without additional machining, and the ability to join multiple interfaces during a single welding cycle. The disadvantages of solid-state welding processes compared with conventional fusion welding or brazing include generally longer welding cycles, and more restrictive size limitations for the parts to be joined. In this article, low-temperature solid-state welding processes are described in relation to the interlayer fabrication method, welding method, and welding parameters. These findings are discussed relative to the mechanical properties of the joints.

Low-Temperature Solid-State Welding

R.S. Rosen, Lawrence Livermore National Laboratory

Applications

The utilization of soft-metal interlayers for solid-state welding may be an appropriate joining method when brittle intermetallic formation, differential thermal contraction during cooling, or oxide dissociation and dissolution temperatures preclude conventional methods such as fusion welding, brazing, and direct solid-state welding of the base metals. An interlayer can be in the form of a foil, or it can be applied to one or both of the base-metal surfaces by various coating methods (e.g., electrodeposition, plasma spraying, or vapor-deposition methods). Methods utilizing coated interlayers require a two-step joining procedure (coating and welding) as opposed to joints fabricated with foil interlayers, which may only require a single solid-state welding step. However, methods utilizing coated interlayers have the advantage over foil interlayers in not requiring solid-state welding at the interlayer/base-metal interface. Solid-state welding of some interlayer/base-metal combinations may require high temperatures and/or pressures, due to the high dissociation temperatures of the surface oxides, which may preclude or restrict the use of foil interlayers.

Interlayer Materials. Silver is the predominant material utilized as a low-temperature interlayer because high-strength joints can be fabricated at relatively low temperatures (473 to 673 K, or 392 to 752 °F) and pressures (100 to 200 MPa, or 15 to 30 ksi) due to the low dissociation temperature (<460 K, or 370 °F) of silver oxide. Other interlayer materials utilized--such as copper (Ref 1), aluminum (Ref 2), and aluminum-silicon alloys (Ref 3)--require higher welding temperatures (873 to 1273 K, or 1110 to 1832 °F) to achieve joint strengths comparable to those of silver. Gold, thought to be a promising interlayer material for low-temperature applications due to the absence of a surface oxide layer at room temperature, has not been utilized as much. This may be due to its propensity to form brittle intermetallics at relatively low temperatures with a number of base metals (Ref 4). Low-temperature (616 K, or 525 °F) solid-state welding of dissimilar (electroplated) interlayer materials, such as gold-to-silver and gold-to-copper, has been achieved at a pressure of 50 MPa (7 ksi) for 3 h using various base metals (Ref 5).

Joint Geometry. Soft interlayer welds between stronger base materials may have ultimate tensile strengths much higher than those of the bulk interlayer materials. For example, the tensile strength of silver interlayer solid-state welds between maraging steel base metals has been reported to be nearly 800 MPa (115 ksi) (Ref 6), while the ultimate tensile strength of the bulk silver is only about 250 MPa (35 ksi) (Ref 7). The high tensile strength of the joint is due to the mechanical constraint provided by the stronger base metal, which restricts transverse contraction of the interlayer. The constraint establishes radial (triaxial) stresses that inhibit the development of shear stresses within the joint, thereby decreasing the effective stress in the interlayer. This reduction in stress inhibits the plastic flow that leads to the ductile fracture. The base metals provide relatively little mechanical constraint of the interlayer during shear loading because the shear stress is parallel to the plane of the interlayer. Therefore, soft interlayer welds between stronger base metals are capable of higher loads in tension (loading perpendicular to the plane of the interlayer, such as in a butt joint) than in shear (loading parallel to the plane of the interlayer, such as in a lap joint).

In general, joint strengths of soft interlayer solid-state welds increase with decreasing interlayer thickness and increasing welding temperature and pressure. However, increasing welding times do not necessarily result in increased joint strength. For example, resultant joint strengths of silver solid-state welds have been reported to be independent of welding times from 1 to 240 min (Ref 4). In addition, decreasing interlayer thickness-to-diameter ratios to less than 0.025 does not result in higher joint strengths and may actually result in a lowering of tensile strength (Ref 8, 9).

References cited in this section

1. A. URENA, J.M.G.D. SALAZAR, AND J. QUINONES, DIFFUSION BONDING OF ALUMINA TO STEEL USING A SOFT COPPER INTERLAYER, *J. MATER. SCI.,* VOL 27, 1992, P 599-606
2. M.G. NICHOLAS AND R.M. CRISPIN, DIFFUSION BONDING STAINLESS STEEL TO ALUMINA USING ALUMINUM INTERLAYERS, *J. MATER. SCI.,* VOL 17, 1982, P 3347-3360
3. T. YAMADA, K. YOKOI, AND A. KOHNO, EFFECT OF RESIDUAL STRESS ON THE STRENGTH OF ALUMINA-STEEL JOINT WITH AL-SI INTERLAYER, *J. MATER. SCI.,* VOL 25, 1990, P 2188-2192
4. J.L. KNOWLES AND T.H. HAZLETT, HIGH-STRENGTH LOW-TEMPERATURE BONDING OF BERYLLIUM AND OTHER METALS, *WELD. J.,* VOL 49 (NO. 7). 1970, P 301S-310S
5. J.T. NIEMANN, R.P. SOPHER, AND P.J. RIEPPEL, DIFFUSION BONDING BELOW 1000 °F, *WELD. J.,* VOL 37, 1958, P 337S-342S
6. M.E. KASSNER, R.S. ROSEN, AND G.A. HENSHALL, DELAYED MECHANICAL FAILURE OF SILVER-INTERLAYER DIFFUSION BONDS, *METALL. TRANS. A,* VOL 21, 1990, P 3085-3100
7. M.E. KASSNER, THE RATE DEPENDENCE AND MICROSTRUCTURE OF HIGH-PURITY SILVER DEFORMED TO LARGE STRAINS BETWEEN 0.16 AND 0.30 TM, *METALL. TRANS. A,* VOL 20, 1989, P 2001-2010
8. R.S. ROSEN AND M.E. KASSNER, DIFFUSION WELDING OF SILVER INTERLAYERS COATED ONTO BASE METALS BY PLANAR-MAGNETRON SPUTTERING, *J. VAC. SCI. TECHNOL. A,* VOL 8 (NO. 1), 1990, P 19-29
9. R.A. MUSIN, V.A. ANTSIVEROV, Y.A. BELIKOV, Y.V. LYAMIN, AND A.N. SOLOKOV, THE EFFECTS ON THE STRENGTH OF THE JOINT OF THE THICKNESS OF THE SOFT INTERLAYER IN DIFFUSION WELDING, *AUTO. WELD.,* VOL 32, 1979, P 38-40

Low-Temperature Solid-State Welding

R.S. Rosen, Lawrence Livermore National Laboratory

Surface Preparation

Surface Cleaning. Prior to coating and/or welding, machined base-metal surfaces are usually degreased by a multiple-step procedure. The following method (Ref 8) has been used to clean a variety of base metals (for example, type 304 stainless steel, maraging steel, uranium, beryllium, and aluminum) before they undergo subsequent ion-sputter etching and silver interlayer deposition in vacuum:

1. WASH WITH ABRASIVE DETERGENT FOLLOWED BY A HOT-WATER RINSE
2. RINSE IN DEIONIZED WATER
3. ULTRASONICALLY CLEAN IN ETHYL ALCOHOL
4. BLOW DRY USING INERT GAS

For other fabrication methods, such as electrodeposited interlayers or foils, the base metal surfaces are degreased prior to plating or joining to remove organic surface films that would otherwise degrade the metallic bonding at the interlayer/base-metal interfaces.

Surface Finish. Joint strength generally increases with decreasing base-metal surface roughness and flatness. This is true for foil interlayers and for coated interlayers, where the surface roughness of the coating generally replicates that of the underlying base-metal surface. Table 1 shows the increases in tensile strengths of silver interlayer welds utilizing lapped base-metal surfaces (0.15 μm, or 6 μin., flatness, 0.03 μm, or 1.2 μin., arithmetic average surface roughness, R_a) over those of machined surfaces (2 μm, or 80 μin., flatness, 0.1 μm, or 4 μin., R_a) (Ref 8, 10). The increase in strength achieved through use of smoother base-metal surfaces is generally attributed to more complete welding from the increased contact area of the laying surfaces.

TABLE 1 EFFECT OF BASE-METAL SURFACE FINISH ON THE TENSILE STRENGTH OF JOINTS SOLID-STATE WELDED USING SILVER INTERLAYERS

BASE METAL	**SURFACE FINISH (TO BE COATED)**	**TENSILE STRENGTH**		**PLASTIC STRAIN IN BASE METAL AT FRACTURE, %**	INTERLAYER THICKNESS, $T/D \times 100$
		MPA	**KSI**		
MARAGING STEEL	MACHINED	656	95	0	**2.4**
MARAGING STEEL(A)	LAPPED	758	110	0	**2.4**
TYPE 304 STAINLESS STEEL(A)	MACHINED	400	58	2.2	**2.4**
TYPE 304 STAINLESS STEEL(A)	LAPPED	414	60	2.8	**2.4**
BERYLLIUM(B)	MACHINED	235	34	~0.2	**0.1**
BERYLLIUM(B)	**LAPPED**	**280**	**41**	**~1**	**0.1**

(A) SOURCE: REF 8.

(B) SOURCE: REF 9

References cited in this section

8. R.S. ROSEN AND M.E. KASSNER, DIFFUSION WELDING OF SILVER INTERLAYERS COATED ONTO BASE METALS BY PLANAR-MAGNETRON SPUTTERING, *J. VAC. SCI. TECHNOL. A,* VOL 8 (NO. 1), 1990, P 19-29
9. R.A. MUSIN, V.A. ANTSIVEROV, Y.A. BELIKOV, Y.V. LYAMIN, AND A.N. SOLOKOV, THE EFFECTS ON THE STRENGTH OF THE JOINT OF THE THICKNESS OF THE SOFT INTERLAYER IN DIFFUSION WELDING, *AUTO. WELD.,* VOL 32, 1979, P 38-40
10. E.R. NAIMON, R.G. KURZ, D. VIGIL, AND L. WILLIAMS, "SILVER FILMS FOR SOLID STATE BONDING," REPORT RFP-312S, ROCKWELL INTERNATIONAL, 1981

Low-Temperature Solid-State Welding

R.S. Rosen, Lawrence Livermore National Laboratory

Interlayer Fabrication Method

Vacuum Coated Interlayers. Specimens to be coated are usually mounted in a fixture (which may be water-cooled) inside the vacuum chamber, with the base-metal surfaces facing the evaporation or sputtering source. The chamber is evacuated using a roughing pump is conjunction with an ultrahigh vacuum turbomolecular or cryogenic pump. After achieving the desired chamber base pressure (typically 10 μPa, or 0.075 μtorr), argon gas is then introduced into the vacuum chamber and maintained at a pressure of about 5 Pa (38 mtorr) for sputtering. A glow-discharge power supply is then used to initiate a plasma at 1 to 2 kV with reference to the base-metal specimens. Depending on the base metals utilized, specimens are usually sputter-etched for 5 to 45 min, resulting in 0.1 to 1 μm (4 to 40 μin.) etched from the surfaces. Figure 1 shows a schematic of a typical vacuum chamber and the two-stage etching and sputter-coated procedure. The shutter is shown placed between the target surface during the etching phase. The coating phase is accomplished (after evacuating the argon gas from the chamber) by evaporation of the interlayer material, or by using a high-rate sputter-target source such as a planar magnetron. Deposition of interlayer metal onto the specimens is usually started after reducing the chamber pressure to 1 t 2 Pa (8 to 15 mtorr). Typical deposition rates at the specimen surfaces are 5 to 20 nm/s (50 to 200 Å/s). Some interlayer/base-metal combinations are capable of achieving high interfacial strengths without having to sputter-etch the native oxide layer prior to deposition. Adhesion of the deposited material is enhanced by preheating the base metals prior to deposition. For the case of beryllium base metal, high-strength joints result when silver is vacuum coated onto the beryllium surfaces at base-metal temperatures exceeding 773 K (932 °F).

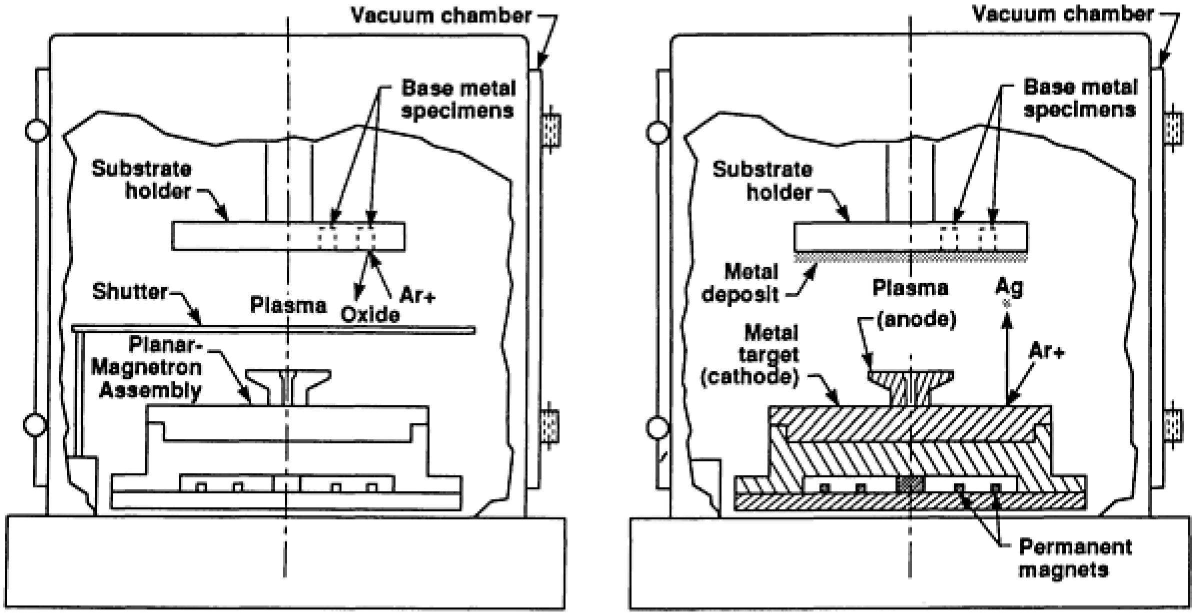
Vacuum chamber
Base metal specimens
Substrate holder
Shutter
Plasma
Oxide
Ar+
Planar-Magnetron Assembly
Vacuum chamber
Base metal specimens
Substrate holder
Metal deposit
Plasma (anode)
Ag
Metal target (cathode)
Ar+
Permanent magnets

FIG. 1 SCHEMATIC OF VACUUM COATING CHAMBER AND ASSOCIATED HARDWARE SHOWING (A) ION-ETCHING AND (B) INTERLAYER METAL DEPOSITION PHASES. AN EVAPORATION SOURCE MAY BE USED INSTEAD OF THE SPUTTERING SOURCE TO DEPOSIT THE INTERLAYER MATERIAL.

Electrodeposited Interlayers. Prior to plating of the interlayer metal onto the base-metal surfaces, the specimens are generally masked to restrict the deposition to the surfaces to be joined. The specimens are usually degreased and then chemically or electrochemically etched in an appropriate solution to remove surface oxide layers. The initial plated layer (~1 to 5 μm, or 40 to 200 μin.) may be deposited at higher current densities (>200 A/m^2, or 20 A/ft^2) to improve the adhesive strength, followed by plating of the remaining interlayer thickness (~10 to 100 μm, or 400 to 4000 μin.) at lower current densities (~40 A/m^2, or 4 A/ft^2). It should be noted that electrodeposited interlayer metals are usually not as pure as vapor-deposited metals due to the presence of impurities in the plating solutions.

Foil Interlayers. Use of foils requires the same cleaning considerations for surface oxides and contaminants as that of deposited interlayers. The base metal as well as the foil material may be etched to remove the oxide layer prior to welding. Foils are usually cut to the desired shape prior to placing between the base metal surfaces to be joined.

Low-Temperature Solid-State Welding

R.S. Rosen, Lawrence Livermore National Laboratory

Welding Methods

The mechanism of solid-state welding is believed to consist of:

- PLASTIC DEFORMATION OF THE CONTACTING SURFACE ASPERITIES DURING INITIAL APPLICATION OF THE LOAD
- OXIDE DISSOLUTION AND DIFFUSION AWAY FROM THE FORMING INTERFACE
- TIME-DEPENDENT PLASTICITY (CREEP) AND/OR DIFFUSION OF METALS TO THE

In addition to these processes, recrystallization and grain growth of the (formed) interface may occur.

Uniaxial Compression. Solid-state welding by uniaxial compression is usually accomplished in a vacuum furnace; for the case of silver at temperatures above about 473 K (392 °F), welding may be accomplished directly in air. Pairs of specimens to be joined are usually loaded with a hydraulic ram. If a vacuum furnace is used, the chamber is evacuated to a gas pressure below about 100 μPa (0.75 μtorr) prior to applying the peak welding pressure (~70 to 210 MPa, or 10 to 30 ksi) to the coated surfaces in contact. The specimen temperature is then elevated to peak temperature (~473 to 873 K, or 392 to 1110 °F) and maintained for 10 to 240 min. The furnace is usually cooled to below 323 K (122 °F) before the pressure is relieved.

Figure 2 shows a schematic of a typical vacuum furnace used for brazing. The constraining fixture shown surrounding the specimen is used to prevent excessive deformation from base-metal yielding during applied loading. This is often the case when dissimilar metals such as aluminum and stainless steels are joined at temperatures and pressures exceeding 473 K (392 °F) and 100 MPa (15 ksi), respectively, using silver as the interlayer material.

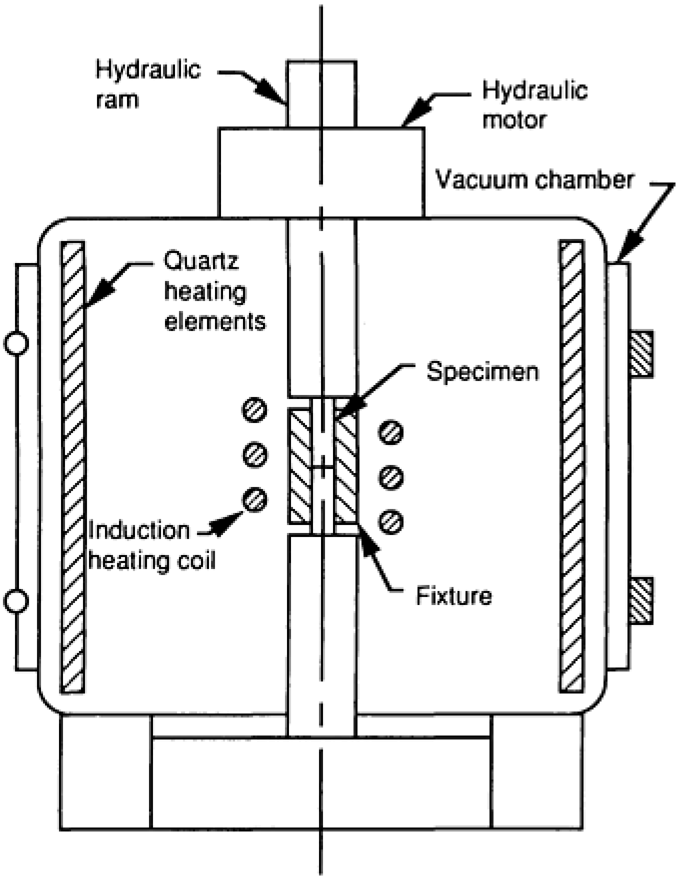

FIG. 2 SCHEMATIC OF A TYPICAL VACUUM FURNACE USED FOR SOLID-STATE WELDING BY UNIAXIAL COMPRESSION. HEATING WOULD BE BY EITHER AN INDUCTION COIL OR RADIANT ELEMENTS AS SHOWN. THE CONSTRAINING FIXTURE SURROUNDING THE SPECIMEN IS USED TO PREVENT EXCESSIVE DEFORMATION FROM BASE-METAL YIELDING DURING APPLIED LOADING.

Isostatic Pressure. Solid-state welding by hot isostatic pressing is accomplished in an autoclave. Parts to be joined are placed with the coated surfaces in contact and then encapsulated with a thin (0.25 to 1 mm, or 0.01 to 0.04 in.) stainless steel membrane, or "can." The cans are sealed using a suitable fusion welding method such as electron-beam or gas-tungsten arc welding. The sealed assemblies are usually evacuated through a small tube welded on one end of the can using a vacuum pump. The evacuated assemblies are placed in an autoclave and isostatically compressed with gas or liquid to a pressure of about 100 MPa (15 ksi). The temperature is then raised to the peak value of 473 to 873 K (392 to 1110 °F) while the gas pressure is increased to the desired peak value (~70 to 350 MPa, or 10 to 50 ksi). Peak temperature and pressure are maintained for 1 to 2 h. The autoclave is usually cooled to below 373 K (212 °F) before the gas is vented.

Final Machining of Welded Parts

After welding, the parts may be finish machined if desired. For the case of dissimilar base metals where one material is considerably stiffer than the other, some allowances may have to be made for single-point cutting-type operations where the tool passes over the joint. For example, machining by lathe turning has resulted in a step remaining at the joint between the base metals due to the differences in elasticity of the base metals. Additionally, lathe cutting with a single-point tool may result in a smearing of the harder base metals across surfaces of the softer interlayer material. These problems can be avoided by utilizing surface grinding as the final machining operation.

References

1. A. URENA, J.M.G.D. SALAZAR, AND J. QUINONES, DIFFUSION BONDING OF ALUMINA TO STEEL USING A SOFT COPPER INTERLAYER, *J. MATER. SCI.,* VOL 27, 1992, P 599-606
2. M.G. NICHOLAS AND R.M. CRISPIN, DIFFUSION BONDING STAINLESS STEEL TO ALUMINA USING ALUMINUM INTERLAYERS, *J. MATER. SCI.,* VOL 17, 1982, P 3347-3360
3. T. YAMADA, K. YOKOI, AND A. KOHNO, EFFECT OF RESIDUAL STRESS ON THE STRENGTH OF ALUMINA-STEEL JOINT WITH AL-SI INTERLAYER, *J. MATER. SCI.,* VOL 25, 1990, P 2188-2192
4. J.L. KNOWLES AND T.H. HAZLETT, HIGH-STRENGTH LOW-TEMPERATURE BONDING OF BERYLLIUM AND OTHER METALS, *WELD. J.,* VOL 49 (NO. 7). 1970, P 301S-310S
5. J.T. NIEMANN, R.P. SOPHER, AND P.J. RIEPPEL, DIFFUSION BONDING BELOW 1000 °F, *WELD. J.,* VOL 37, 1958, P 337S-342S
6. M.E. KASSNER, R.S. ROSEN, AND G.A. HENSHALL, DELAYED MECHANICAL FAILURE OF SILVER-INTERLAYER DIFFUSION BONDS, *METALL. TRANS. A,* VOL 21, 1990, P 3085-3100
7. M.E. KASSNER, THE RATE DEPENDENCE AND MICROSTRUCTURE OF HIGH-PURITY SILVER DEFORMED TO LARGE STRAINS BETWEEN 0.16 AND 0.30 TM, *METALL. TRANS. A,* VOL 20, 1989, P 2001-2010
8. R.S. ROSEN AND M.E. KASSNER, DIFFUSION WELDING OF SILVER INTERLAYERS COATED ONTO BASE METALS BY PLANAR-MAGNETRON SPUTTERING, *J. VAC. SCI. TECHNOL. A,* VOL 8 (NO. 1), 1990, P 19-29
9. R.A. MUSIN, V.A. ANTSIVEROV, Y.A. BELIKOV, Y.V. LYAMIN, AND A.N. SOLOKOV, THE EFFECTS ON THE STRENGTH OF THE JOINT OF THE THICKNESS OF THE SOFT INTERLAYER IN DIFFUSION WELDING, *AUTO. WELD.,* VOL 32, 1979, P 38-40
10. E.R. NAIMON, R.G. KURZ, D. VIGIL, AND L. WILLIAMS, "SILVER FILMS FOR SOLID STATE BONDING," REPORT RFP-312S, ROCKWELL INTERNATIONAL, 1981

Explosion Welding

John G. Banker and Edward G. Reineke, Explosive Fabricators, Inc.

Introduction

EXPLOSION WELDING (EXW) is a solid-state metal-joining process that uses explosive force to create an electron-sharing metallurgical bond between two metal components. Although the explosive detonation generates considerable heat, there is not time for heat transfer to the component metals; therefore, there is no appreciable temperature increase in the metals. The physics of the process are discussed in the article "Fundamentals of Explosion Welding" in this Volume.

John G. Banker and Edward G. Reineke, Explosive Fabricators, Inc.

EXW Process Attributes

Metallurgical Attributes. Because of the absence of heating, EXW products do not exhibit many of the metallurgical characteristics of fusion-welded, brazed, or hot-rolled/forged products. Unlike those processes, in EXW (Ref 1, 2, 3, 4):

- THE COMPONENT METALS REMAIN IN THEIR WROUGHT STATES; CONTINUOUS CASE STRUCTURES ARE NOT CREATED.
- THE MICROSTRUCTURES, MECHANICAL PROPERTIES, AND CORROSION PROPERTIES OF THE WROUGHT PARENT COMPONENTS ARE NOT ALTERED FROM APPLICABLE PREBONDING SPECIFICATION REQUIREMENTS.
- THERE ARE NO HEAT-AFFECTED ZONES.
- THERE ARE NO CONTINUOUS-MELT BANDS EXHIBITING MIXED CHEMISTRY.
- THERE IS ALMOST NO DIFFUSION OF ALLOYING ELEMENTS BETWEEN COMPONENTS.

Metals Combinations. Explosion welding is an effective joining method for virtually any combination of metals. The only metallurgical limitation is sufficient ductility and fracture toughness to undergo the rapid deformation of the process without fracture. Generally accepted limits are 10% and 30 J (22 ft · lbf) minimum, respectively.

Explosion welding is suitable for joining metals of the same type--for example, steel to steel--as well as metals with substantially different densities, melting points, and/or yield strengths--for example, tantalum to alloy titanium. The process is commonly used to join corrosion-resistant alloys to carbon or alloy steels.

Size Limitations. Explosion welding is appropriate for use over a broad range of sizes. It has been used for joining electronic components of widths less than 0.5 mm (0.020 in.) and for cladding plates up to 5 × 13 m (16 × 40 ft). Cladding metal thicknesses can range from 0.025 to 100 mm (0.001 to 4 in.), and base metal thickness can range from 0.025 mm (0.001 in.) to over 1 m (40 in.). The size limits are generally mandated by component metal manufacturability and transport requirements.

Configuration Limitations. Explosion welding is limited to the bonding of flat surfaces or coaxial cylindrical surfaces. It is ideally suited for bonding large, flat areas, as in clad plate manufacture.

Bond Zone Morphology. Explosion welding is traditionally known for its wavy bond zone morphology. The physics of the wave formation process and its relation to process variables are discussed in the article "Fundamentals of Explosion Welding" in this Volume. The EXW bond zone morphology is flat at the low end of the bonding velocity range and transitions to wavy as the velocity is increased. Through proper selection of EXW parameters, the bond zone morphology can be tailored to satisfy product performance requirements. Although a low-velocity wavy bond zone is preferred for most needs (Fig. 1), certain applications are better served by flat or turbulent-wavy interfaces (Ref 5).

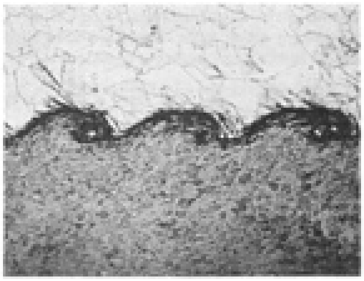

FIG. 1 BOND ZONE PATTERN TYPICAL OF EXPLOSION CLAD METALS. MATERIALS ARE TYPE 304L STAINLESS STEEL AND MEDIUM-CARBON STEEL. 20×

Interlayers. The physical and mechanical performance requirements of the EXW bond zone are determined by the product application. With some combinations of metals, the required bond zone performance characteristics are best achieved with an interlayer between the two primary metals. For example, a titanium interlayer can be used to increase the performance temperature of an aluminum/copper weld, and a silver interlayer can be used to improve the leak tightness of an aluminum/stainless steel weld.

References cited in this section

1. R.E. SAVIDGE, *MECHANICAL PROPERTIES OF EXPLOSION CLAD BONDS OF HIGH NICKEL ALLOYS ON LOW ALLOY STEELS AND THE EFFECT OF EXPLOSION CLADDING ON THE FRACTURE PROPERTIES OF THE STEEL,* THIRD INTERNATIONAL CONFERENCE ON HIGH ENERGY RATE FABRICATION, 1971, P 3.3.1
2. W. LUCAS, J.D. WILLIAMS, AND B. CROSSLAND, *SOME METALLURGICAL OBSERVATIONS ON EXPLOSION WELDING,* SECOND INTERNATIONAL CONFERENCE OF THE CENTRE FOR HIGH ENERGY FORMING, 1969, P 8.1.2
3. S.H. CARPENTER AND M. NAGARKAR, *THE EFFECTS OF EXPLOSIVE WELDING ON THE KINETICS OF METALLURGICAL REACTIONS,* THIRD INTERNATIONAL CONFERENCE OF THE CENTRE FOR HIGH ENERGY FORMING, 1971, P 3.4.1
4. L.F. TRUEB, MICROSTRUCTURAL EFFECTS OF HEAT TREATMENT ON THE BOND INTERFACE OF EXPLOSIVELY WELDED METALS, *METALL. TRANS.,* VOL 2, JAN 1971, P 145
5. A. SZECKET, O.T. INAL, AND J. ROCCO, EXPLOSIVE WELDING OF ALUMINUM TO STEEL; A WAVY VERSUS A STRAIGHT INTERFACE, *HIGH ENERGY RATE FABRICATION,* 1984, P 153

Explosion Welding

John G. Banker and Edward G. Reineke, Explosive Fabricators, Inc.

EXW Products and Applications

Common Industrial Applications. Because of the unique safety and noise-vibration considerations, EXW is generally performed in relatively isolated facilities by companies specializing in explosives operations. The major products are large clad plates, which are used in pressure vessel, heat exchanger, and electrochemical process applications (Fig. 2). The dissimilar-metal welding features of EXW are made available to a broader industrial base through the concept of welding transition joints (Fig. 3).

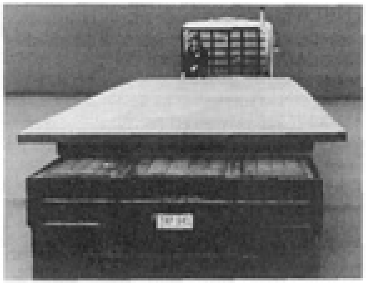

FIG. 2 TITANIUM/STEEL EXPLOSION-BONDED CLAD PLATE

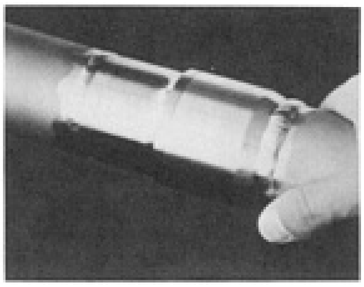

FIG. 3 ALUMINUM-TO-STAINLESS WELD ACCOMPLISHED WITH AN EXPLOSION-BONDED TUBULAR TRANSITION JOINT. ALUMINUM AND STAINLESS STEEL ARE WELDED TO THE RESPECTIVE ENDS USING CONVENTIONAL FUSION-WELDING PROCESSES.

Transition joint components are produced by explosion welders at their production facilities and then provided in the form of blocks, strips, or tubular couplings to equipment fabricators for use in the conventional welding of dissimilar metals. Common applications are aluminum/steel (Ref 6, 7), aluminum/titanium, and aluminum/copper-nickel for shipboard construction (Fig. 4); aluminum/copper and aluminum/steel for electrical contacts and buss systems; aluminum/stainless steel for cryogenic pipe couplings; titanium/stainless steel for aerospace and aircraft applications (Fig. 5); and Kovar/copper and Kovar/aluminum for electronic packaging (Ref 8).

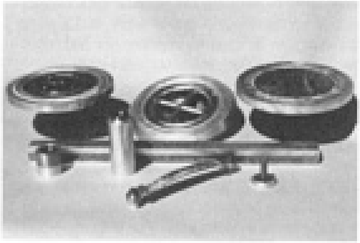

FIG. 4 TRANSITION JOINT MATERIALS FOR JOINING DISSIMILAR METALS IN SHIPBOARD AND MARINE EQUIPMENT CONSTRUCTION

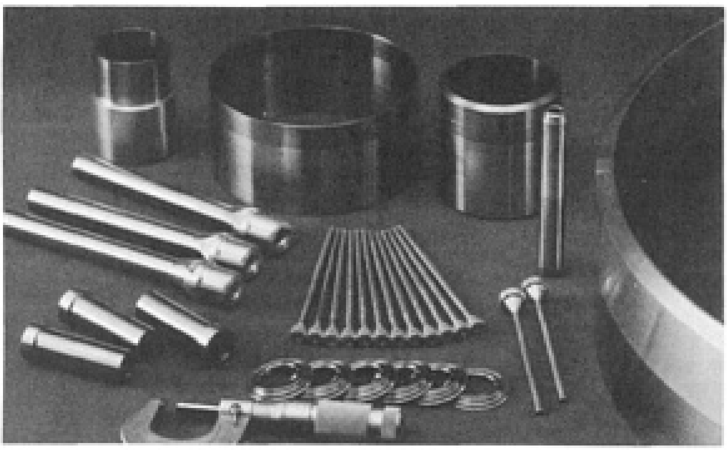

FIG. 5 TUBULAR WELDING TRANSITION JOINTS FOR HERMETIC WELDS BETWEEN TITANIUM, ALUMINUM, AND ZIRCONIUM AND STAINLESS STEEL

Shop Welding Applications. Small, specialty EXW devices have been developed for a few unique shop applications. Explosion welding is used in heat exchanger shops for joining tubes to tubesheets and tube plugging. Explosion spot welding devices have also been used in limited commercial applications (Ref 9).

References cited in this section

6. C.R. MCKENNEY AND J.G. BANKER, *EXPLOSION-BONDED METALS FOR MARINE STRUCTURAL APPLICATIONS,* THE SOCIETY OF NAVAL ARCHITECTS AND MARINE ENGINEERS, NOV 1970, P 285-292
7. E. GAINES AND J. BANKER, *SHIPYARD ALUMINUM/STEEL WELDED TRANSITION JOINTS,* 1991 SHIP PRODUCTION SYMPOSIUM, SEPT 1991
8. H. MANSELL, HYBRID METAL PACKAGES BY EXPLOSION BONDING, *HYBRID CIRCUIT TECHNOL.,* SEPT 1990, P 67
9. L.G. LAZARI, EXPLOSIVE WELDING AND ITS PRACTICAL APPLICATIONS, *WELD. REV.,* MAY 1988, P 74

John G. Banker and Edward G. Reineke, Explosive Fabricators, Inc.

EXW Manufacturing Process and Practice

Safety and Regulations. Safety is of critical importance in the practice of EXW. Guidelines and regulations for the safe handling and transport of explosives are provided by the U.S. Bureau of Alcohol, Tobacco, and Fire Arms (Ref 10), the U.S. Department of Transportation in its Federal Motor Carrier Safety Regulations (Ref 11), and the Institute of Makers of Explosives Safety Library Publications (Ref 12). Licensing and regulations for the purchase, transport, manufacture, and use of explosives may vary significantly among states and municipalities. Depending on the types of explosives used, the detonation by-products may be subject to environmental regulation. Potential practitioners of explosion-bonding operations should understand the safety standards and regulations before proceeding (Ref 13).

Noise and Vibration Abatement. The sound and ground vibration generated during EXW can be considerable. Common methods for noise abatement include performing the operation in a highly remote geographic location, in a mine or bunker, or in a specially designed blast chamber or vacuum chamber. Safety considerations vary among these options. Precautions must be taken to ensure that personnel are protected from shrapnel and facility deterioration.

Process Geometry. Two basic geometric configurations of the EXW process are commonly used: angle bonding and parallel-plate bonding. Angle bonding is normally used for bonding sheet components and tubes, where the required bond width does not exceed 20 times the flyer plate thickness. The more commonly used parallel-plate geometry (Fig. 6) is applicable for welding larger flat areas and concentric cylinders (Ref 14).

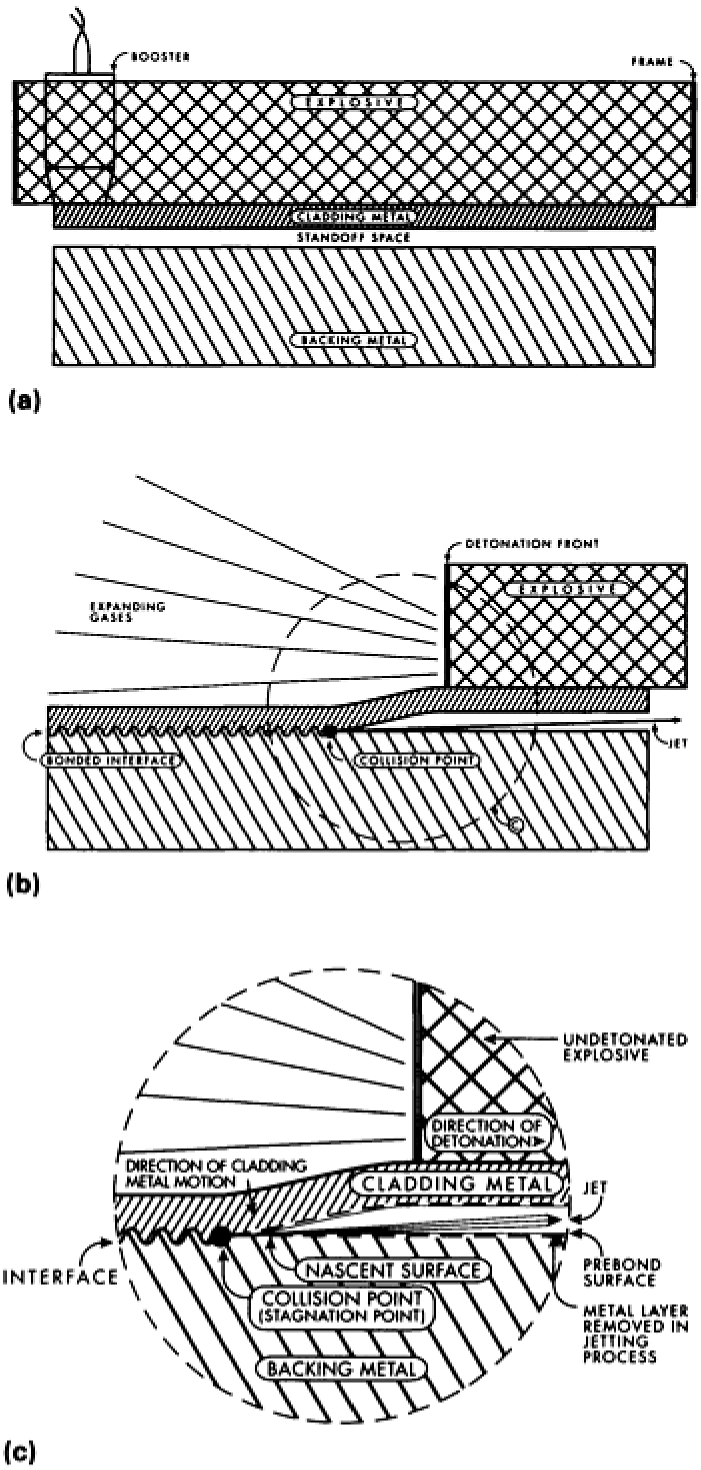
BOOSTER
FRAME
EXPLOSIVE
CLADDING METAL
STANDOFF SPACE
BACKING METAL
(a)
DETONATION FRONT
EXPLOSIVE
EXPANDING GASES
JET
BONDED INTERFACE
COLLISION POINT
(b)
UNDETONATED EXPLOSIVE
DIRECTION OF DETONATION
DIRECTION OF CLADDING METAL MOTION
CLADDING METAL
JET
INTERFACE
NASCENT SURFACE
PREBOND SURFACE
COLLISION POINT (STAGNATION POINT)
METAL LAYER REMOVED IN JETTING PROCESS
BACKING METAL
(c)

FIG. 6 PARALLEL-PLATE EXPLOSION WELDING PROCESS. (A) EXPLOSION-CLADDING ASSEMBLY BEFORE DETONATION. (B) EXPLOSION-CLADDING ASSEMBLY DURING DETONATION. (C) CLOSE-UP OF (B) SHOWING MECHANISM FOR JETTING AWAY THE SURFACE LAYER FROM THE PARENT LAYER

Basic parameters of EXW are flyer plate velocity, collision point velocity, and collision angle. The process variables that affect these parameters are explosive detonation rate, explosive mass, flyer plate mass, standoff distance, and/or preset angle. These process variables are normally controlled to provide:

- A FLYER PLATE VELOCITY OF 250 TO 500 M/S (820 TO 1600 FT/S)
- A COLLISION POINT VELOCITY OF 1500 TO 3500 M/S (4900 TO 11,000 FT/S) BUT LESS THAN 120% OF THE SONIC VELOCITY OF THE COMPONENT METALS
- A COLLISION ANGLE OF 5 TO 20°

Selecting the specific parameters is discussed in the Section "Procedure Development and Practice Considerations for Solid-State Welding" in this Volume.

Sizing Considerations. It is common practice to provision materials that are wider, longer, and thicker than required in the finished product to compensate for thinning and edge effects. In most situations, metals cannot be reliably bonded fully to the edges. Typically, edge nonbond allowances are four to eight times flyer plate thickness, depending on the combination of metals. Additionally, an area of nonbond is typical near the explosive initiator.

The thickness of the component metals is reduced during surface preparation and bonding. Surface conditioning removes up to 0.75 mm (0.030 in.) from the thickness. During the bonding operation, the force vector parallel to the detonation causes a 1 to 2% increase in plate lateral dimensions; this is accomplished by a corresponding reduction in thickness. The jetting of the metal surfaces during the bonding process causes only a thin layer to be removed, usually less than 0.05 mm (0.002 in.).

References cited in this section

10. *ATF--EXPLOSIVES LAW AND REGULATIONS,* BUREAU OF ALCOHOL, TOBACCO, AND FIREARMS, DEPARTMENT OF THE TREASURY, JUNE 1990
11. *FEDERAL MOTOR CARRIER SAFETY REGULATIONS,* U.S. DEPARTMENT OF TRANSPORTATION; FEDERAL HIGHWAY ADMINISTRATION, FEB 1991
12. *SUGGESTED CODE OF REGULATIONS FOR THE MANUFACTURE, TRANSPORTATION, STORAGE, SALES, POSSESSION AND USE OF EXPLOSIVE MATERIALS,* INSTITUTE OF MAKERS OF EXPLOSIVES, JAN 1985
13. SALES DEVELOPMENT SECTION OF THE EXPLOSIVES DEPARTMENT OF E.I. DU PONT DE NEMOURS & CO., INC., *BLASTERS' HANDBOOK*
14. W. SHARP, THE EXPLOSION CLADDING PROCESS, *ENG. DES.,* SEPT/OCT 1979

Typical EXW Process Sequence (Parallel-Plate Bonding)

Surface Preparation. The faying surfaces of the components are ground to achieve a smooth, pitfree finish of 1 to 3 μm (50 to 100 μin.).

Assembly. The plates are fixtured parallel while separated by a standoff gap. Methods for determining the standoff gap are presented in the article "Introduction to Solid-State Welding" and the Section "Procedure Development and Practice Considerations for Solid-State Welding" in this Volume. Small standoff devices are required to bear the weight of the flyer plate and the explosive. Edge standoff devices are easily ejected from the system during bonding; therefore, material of construction is rarely critical. Internal standoff devices must be of sufficiently low mass to be removed by the jet, or must be located in trim regions. Commonly used internal standoff materials are thin metal ribbons and foam blocks.

Explosive Loading. The explosive is placed on top of the flyers metal surface. The most commonly used explosives are granular or liquid and are typically contained within a simple box structure fabricated around the perimeter of the flyer plate. Initiating explosives are positioned in a suitable location on the plate surface.

The required detonation range of EXW is below that of most commercially manufactured explosives. Consequently, most practitioners of EXW use proprietary explosive blends. Appropriate detonation characteristics can be achieved by specialized blending of commercial explosives such as Amatol, Dynamite, or NCN.

Bonding. The primer is electrically initiated, causing a progressive detonation of the explosive, which creates the actions shown in Fig. 6 and welds the plates together.

Flattening. The energy of bonding typically creates sufficient deformation that flattening or straightening is required prior to further processing. Flattening is performed with equipment of the same design used in plate and sheet manufacture.

Testing and inspection are performed as mandated by the product specifications. Ultrasonic inspection, using standard procedures such as ASTM A 578, is the most commonly accepted method for inspecting the bond continuity. Bond shear strength is normally measured using the test defined in ASTM A 265. The tensile strength of the bond is normally measured using a Ram Tensile Procedure; the standard design is presented in MIL-J-24445.

Preparation of the Required Product. Because of the process edge effects, it is common practice to cut the product from a larger plate after bonding. The process options vary with the metal combinations and include oxyfuel gas cutting, plasma cutting, waterjet cutting, sawing, and machining.

Heat Treatment. Most metal systems do not require postbond heat treatment; however, some combinations provide superior properties if they are given a stress relief after bonding. For example, titanium/steel and zirconium/steel exhibit improved bond toughness after stress relief, and ferritic stainless steels exhibit an improvement in ductile-to-brittle transition temperature after stress relief.

Explosion Welding

John G. Banker and Edward G. Reineke, Explosive Fabricators, Inc.

Specifications

There are no universal ASTM, ASME, or military standard specifications for EXW products. Specific combinations of metals are covered in ASTM A 263, A 264, A 265, and B 432 (and their ASME equivalents) and in MIL-J-24445. The International Explosion Metalworking Association (IEMA) has established IEMA Guide Note 100, which provides recommended specification content.

Explosion Welding

John G. Banker and Edward G. Reineke, Explosive Fabricators, Inc.

Process Variations

The preceding discussion describes the process for welding a two-component flat plate product. Essentially the same approach can be used for producing multilayer products or cylindrical shapes:

- MULTIPLE LAYERS CAN BE BONDED SEQUENTIALLY IN A SINGLE BONDING OPERATION; AS MANY AS 50 INDIVIDUAL THIN LAYERS HAVE BEEN IN A SINGLE ASSEMBLY.
- TWO-SIDED CLAD PLATES, WITH A HEAVY CORE METAL, CAN BE PRODUCED BY

SIMULTANEOUSLY BONDING BOTH SIDES USING A VERTICAL ASSEMBLY.

- TWO OR MORE CYLINDERS CAN BE BONDED CONCENTRICALLY. THE EXPLOSIVE IS PLACED EITHER INSIDE OF THE CYLINDER OR OUTSIDE OF THE OUTER CYLINDER, DEPENDING ON DIAMETER AND WALL THICKNESS.

Explosion Welding

John G. Banker and Edward G. Reineke, Explosive Fabricators, Inc.

References

1. R.E. SAVIDGE, *MECHANICAL PROPERTIES OF EXPLOSION CLAD BONDS OF HIGH NICKEL ALLOYS ON LOW ALLOY STEELS AND THE EFFECT OF EXPLOSION CLADDING ON THE FRACTURE PROPERTIES OF THE STEEL,* THIRD INTERNATIONAL CONFERENCE ON HIGH ENERGY RATE FABRICATION, 1971, P 3.3.1
2. W. LUCAS, J.D. WILLIAMS, AND B. CROSSLAND, *SOME METALLURGICAL OBSERVATIONS ON EXPLOSION WELDING,* SECOND INTERNATIONAL CONFERENCE OF THE CENTRE FOR HIGH ENERGY FORMING, 1969, P 8.1.2
3. S.H. CARPENTER AND M. NAGARKAR, *THE EFFECTS OF EXPLOSIVE WELDING ON THE KINETICS OF METALLURGICAL REACTIONS,* THIRD INTERNATIONAL CONFERENCE OF THE CENTRE FOR HIGH ENERGY FORMING, 1971, P 3.4.1
4. L.F. TRUEB, MICROSTRUCTURAL EFFECTS OF HEAT TREATMENT ON THE BOND INTERFACE OF EXPLOSIVELY WELDED METALS, *METALL. TRANS.*, VOL 2, JAN 1971, P 145
5. A. SZECKET, O.T. INAL, AND J. ROCCO, EXPLOSIVE WELDING OF ALUMINUM TO STEEL; A WAVY VERSUS A STRAIGHT INTERFACE, *HIGH ENERGY RATE FABRICATION*, 1984, P 153
6. C.R. MCKENNEY AND J.G. BANKER, *EXPLOSION-BONDED METALS FOR MARINE STRUCTURAL APPLICATIONS,* THE SOCIETY OF NAVAL ARCHITECTS AND MARINE ENGINEERS, NOV 1970, P 285-292
7. E. GAINES AND J. BANKER, *SHIPYARD ALUMINUM/STEEL WELDED TRANSITION JOINTS,* 1991 SHIP PRODUCTION SYMPOSIUM, SEPT 1991
8. H. MANSELL, HYBRID METAL PACKAGES BY EXPLOSION BONDING, *HYBRID CIRCUIT TECHNOL.,* SEPT 1990, P 67
9. L.G. LAZARI, EXPLOSIVE WELDING AND ITS PRACTICAL APPLICATIONS, *WELD. REV.*, MAY 1988, P 74
10. *ATF--EXPLOSIVES LAW AND REGULATIONS*, BUREAU OF ALCOHOL, TOBACCO, AND FIREARMS, DEPARTMENT OF THE TREASURY, JUNE 1990
11. *FEDERAL MOTOR CARRIER SAFETY REGULATIONS*, U.S. DEPARTMENT OF TRANSPORTATION; FEDERAL HIGHWAY ADMINISTRATION, FEB 1991
12. *SUGGESTED CODE OF REGULATIONS FOR THE MANUFACTURE, TRANSPORTATION, STORAGE, SALES, POSSESSION AND USE OF EXPLOSIVE MATERIALS,* INSTITUTE OF MAKERS OF EXPLOSIVES, JAN 1985
13. SALES DEVELOPMENT SECTION OF THE EXPLOSIVES DEPARTMENT OF E.I. DU PONT DE NEMOURS & CO., INC., *BLASTERS' HANDBOOK*
14. W. SHARP, THE EXPLOSION CLADDING PROCESS, *ENG. DES.*, SEPT/OCT 1979

Forge Welding

FORGE WELDING (FOW) is a solid-state process in which the workpieces are heated to the welding temperature and then applied with blows sufficient to cause permanent deformation at the faying surfaces. It is most commonly applied to the butt welding of steels. As contrasted with hot pressure (thermocompression) welding of ductile face-centered cubic (fcc) metals, which is normally performed at temperatures of less than one-half the melting temperature (T_m), forge welding is typically conducted at temperatures in the 0.8 to 0.9 T_m range. The forge welding temperature is generally selected to be as high as possible with due consideration to avoiding such metallurgical problems as hot shortness, embrittlement, sensitization, and excessive grain coarsening. This implies an understanding of the unique metallurgical problems of the alloy to be welded.

Forge welding requires the application of pressure by means of either a hammer (hammer welding), rolls (roll welding), or dies (die welding). Joint configurations differ depending on whether the joints are to be produced manually or using automatic equipment. Typical joint designs used in manual forge welding operations are shown in Fig. 1. The joint surfaces in Fig. 1 are slightly rounded or crowned to ensure that the centerline region of the components joined will be welded first to force any contaminants (for example, slag, dirt, or oxide) present on the surfaces out of the joint. Typical joint configurations used for automatic forge welding operations are shown in Fig. 2 (Ref 1).

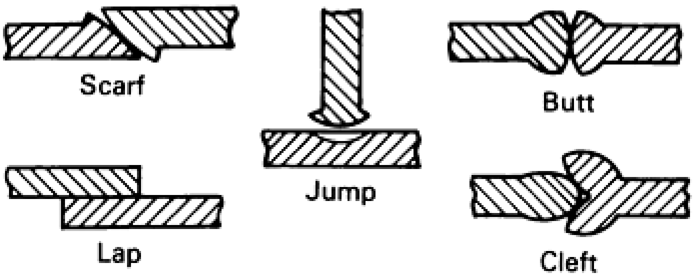

FIG. 1 TYPICAL JOINT CONFIGURATIONS USED FOR MANUAL FORGE WELDING APPLICATIONS. SOURCE: REF 1

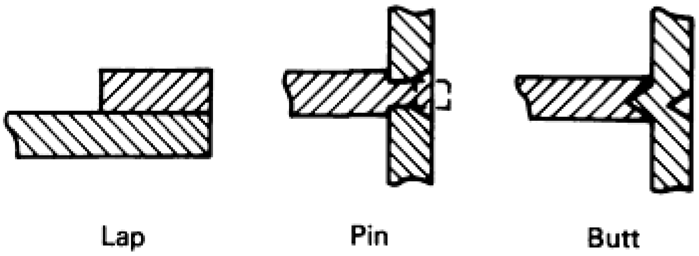

FIG. 2 RECOMMENDED JOINT CONFIGURATIONS USED IN AUTOMATIC FORGE WELDING APPLICATIONS. SOURCE: REF 1

Hydraulic presses are typically employed to apply pressure. Presses are often highly automated, featuring microprocessor control of pressure and temperature cycles. Heat is applied locally to the joint area by multiple-tip oxyacetylene torches, resistance heating, or induction heating. Often the oxyacetylene torches are oscillated to ensure uniformity of heating. In a closely related process, magnetically induced arc butt welding, the surfaces to be welded are heated by a rapidly rotating arc plasma. Generally, the process is conducted in the open air, with oxygen partially occluded from the joint area by the initial contact of the faying surfaces. When employing oxyacetylene torches, a slightly reducing flame affords some atmospheric protection. Vacuum, inert, and reducing atmospheres have been used.

The normal welding sequence is to (1) apply sufficient pressure to firmly seat the faying surfaces against one another, (2) heat the joint to welding temperature, and (3) rapidly apply additional pressure to upset the weld zone. Typical weld durations are 1 to 2 min. A less common procedure is to initially apply high pressure and permit deformation to occur during the heating cycle. Most forge welding employs sufficient pressure to upset the surface until the increase in the surface area is 125% or more. However, such high deformation tend to cause flow lines to bend toward the surface during upsetting. Consequently, alloys that contain significant stringers and inclusions may exhibit poor impact or fatigue properties when welded with high amounts of upset. This effect may be minimized by reducing the upset, which normally requires increasing welding temperature and/or time to ensure complete elimination of voids and surface oxide.

Forge welding is most commonly applied to carbon and low-alloy steels, with typical welding temperatures of about 1125 °C (2060 °F). Low-carbon steels can be used in the as-welded condition, but medium-carbon steels and low-alloy steels normally are given full heat treatments following welding. In those cases where full heat treatment is impractical, but hardening due to rapid cooling has occurred, induction heating may be used to temper the weld zone. Other metals welded by forge welding include high-alloy steels, nickel-base alloys, cobalt-base alloys, aluminum alloys, titanium alloys, and tungsten. Applications of this process include welding rods, bars, tubes, rails, aircraft landing gear, chains, and cans. The forge welding process is competitive with flash welding (see the article "Flash Welding" in this Volume) and friction welding (see the article "Friction Welding" in this Volume).

Flux must be added when forge welding certain meals to prevent the formation of oxide scale. The flux and the oxides present combine to form a protective coating on the heated surfaces of the metal that prevents the formation of additional oxide and lowers the melting point of the existing oxide.

Silica sand and borax are two fluxes commonly used on steels. Silica sand can be used as a flux in the forge welding of low-carbon steel. The oxides of very low-carbon steels (ingot irons) and wrought irons do not require fluxes because their oxides have low melting points. Borax, sprinkled on the workpiece while it is in the process of being heated, is commonly used in the forge welding of high-carbon steels because of its low fusion point (Ref 1).

Weld durations for resistance welding are very short (seconds) compared to those for forge welding where gas torches are employed. This is because heat is generated internally in resistance welding, but externally with gas torches. Terms that are sometimes used synonymously with forge welding include pressure welding, upset welding, and solid-state resistance welding.

Reference

1. R.L. O'BRIEN, ED., *WELDING HANDBOOK,* VOL 2, 8TH ED., AWS, 1991, P 917-918

Reference

1. R.L. O'BRIEN, ED., *WELDING HANDBOOK,* VOL 2, 8TH ED., AWS, 1991, P 917-918

Cold Welding

Karl Thomas, Technische Universität Braunschweig; Manfred Petri, Gerhard Petri GmbH + Co. KG

Introduction

COLD WELDING (also known as cold pressure welding) in normal ambient conditions involves the destruction of the surface (that is, oxide) layers of the metallic materials in the weld area. This exposes areas of clean metal surface on the two components to be welded, which must be brought into contact with each other to generate the interatomic forces needed to form a weld. Examinations of butt-welded joints of aluminum using transmission electron microscopy (TEM) have shown that the common deformation of the material in the weld zone leads to the destruction of the surface (that is, oxide and contaminant). TEM showed these surface layers in the form of numerous minute particles in the joint. The largest of these particles had a diameter of 12 nm (120 Å).

The actual interface between two dissimilar materials is a phase boundary, the same as a grain boundary between similar materials. Thermally induced processes such as recrystallization or diffusion are not required for welding to take place (Ref 1, 2). This makes cold-formable materials with a brittle cover layer especially suitable for cold pressure welding. The technique is frequently used with nonferrous metals (primarily copper and aluminum) (Fig. 1).

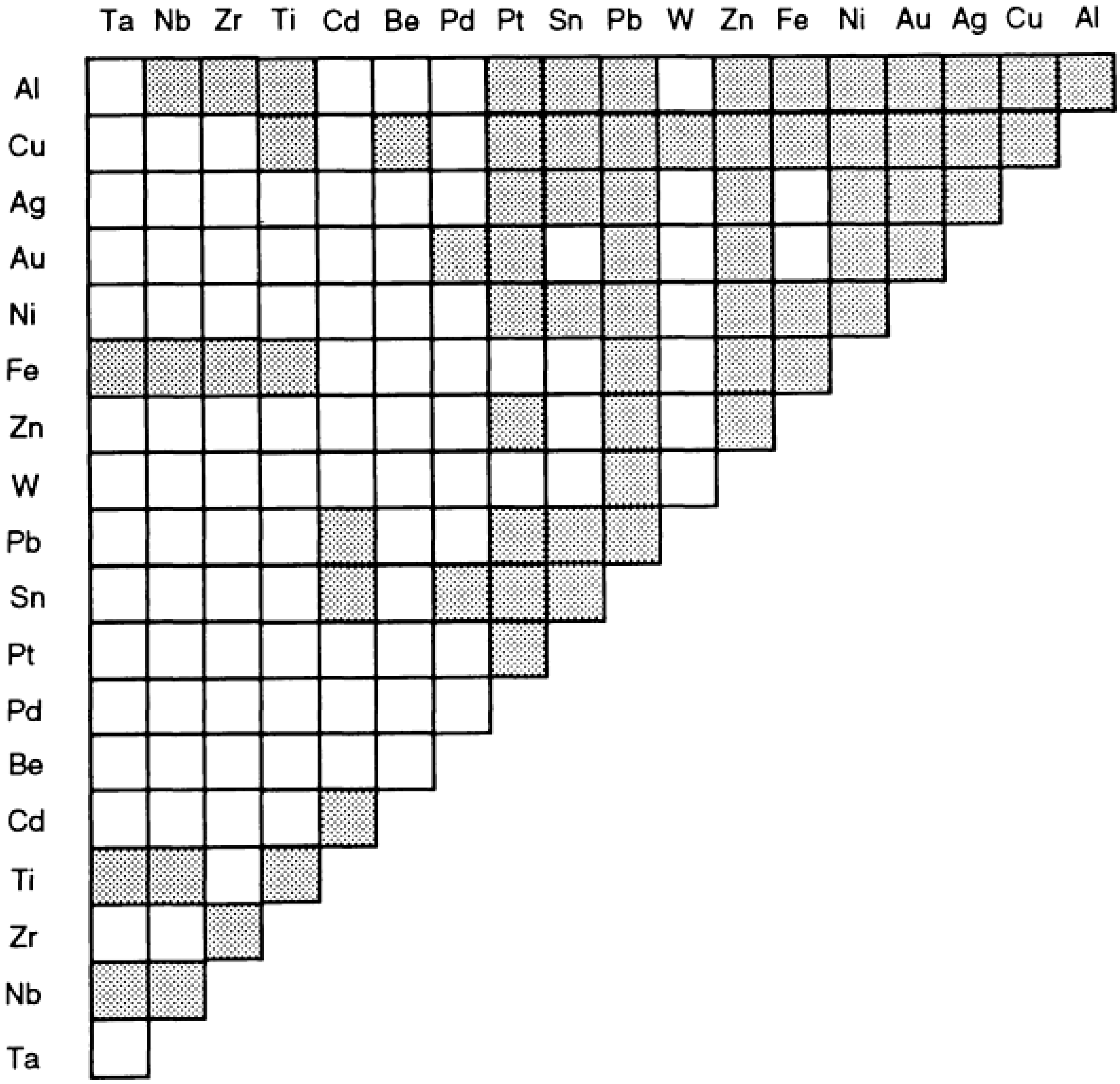

	Ta	Nb	Zr	Ti	Cd	Be	Pd	Pt	Sn	Pb	W	Zn	Fe	Ni	Au	Ag	Cu	Al
Al	□	▨	▨	▨	□	□	□	▨	▨	▨	□	▨	▨	▨	▨	▨	▨	▨
Cu	□	□	□	▨	□	▨	□	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	
Ag	□	□	□	□	□	□	□	▨	▨	▨	□	▨	□	▨	▨	▨		
Au	□	□	□	□	□	□	▨	▨	□	▨	□	▨	□	▨	▨			
Ni	□	□	□	□	□	□	□	▨	▨	▨	□	▨	▨	▨				
Fe	▨	▨	▨	▨	□	□	□	□	□	▨	□	▨	▨					
Zn	□	□	□	□	□	□	□	▨	□	▨	□	▨						
W	□	□	□	□	□	□	□	□	□	▨	□							
Pb	□	□	□	□	▨	□	□	▨	▨	▨								
Sn	□	□	□	□	▨	□	▨	▨	▨									
Pt	□	□	□	□	□	□	□	▨										
Pd	□	□	□	□	□	□	□											
Be	□	□	□	□	□	□												
Cd	□	□	□	□	▨													
Ti	▨	▨	□	▨														
Zr	□	□	▨															
Nb	▨	▨																
Ta	□																	

FIG. 1 CHART SHOWING COMBINATIONS OF METALS THAT CAN BE SUCCESSFULLY COLD WELDED (INDICATED BY SHADED BOX). SOURCE: ADAPTED FROM REF 3

In general, welds are examined by destructive techniques. As with other pressure welded joints, nondestructive testing methods fail to generate reliable results. It is expedient, therefore, to prepare the parts to be joined with great care and to control the parameters of the welding process as closely as the equipment tolerance specifications allow.

References

1. H. KREYE AND K. THOMAS, ELECTRON-MICROSCOPICAL TESTS AND THE BONDING MECHANISM OF COLD PRESSURE WELDING, *WELD. CUTTING,* VOL 29, 1977, P 249-252
2. H. PRIES, TESTS ON THE BONDING MECHANISM OF WELDING METALS TO NON-METALS IN ULTRA-HIGH VACUUM, *VDI PROGRESS REPORTS,* SERIES 5, *PRIMARY AND MANUFACTURING*

MATERIALS, NO. 89, VDI, 1985.

3. N. BAY, COLD WELDING: CHARACTERISTICS, BONDING MECHANISMS, BOND STRENGTH (PART 1), *MET. CONSTR.,* VOL 18 (NO. 8), JUNE 1986, P 369-372

Cold Welding

Karl Thomas, Technische Universität Braunschweig; Manfred Petri, Gerhard Petri GmbH + Co. KG

Variations of Cold Pressure Welding

Because welding lends itself to numerous techniques of forming, there are many variants of the cold pressure welding process. Cold pressure welding can be accomplished by deforming in a lap or butt configuration, drawing, extrusion, and rolling. The following discussion deals with only the first three variants.

Cold Pressure Lap Welding. The geometry of the cold pressure lap welding method (Fig. 2) is similar to the geometry of resistance spot welding. Two sheets of metal are joined in such a way that the direction of material flow is perpendicular to the direction of pressure. After a material-dependent reduction of the sheet thickness, the joint is formed. A high-quality joint will not be formed unless all superfluous oil has been removed and the contact surfaces have been mechanically cleaned (for example, by scratch-brushing with a rotating steel brush). Cleaning should be performed immediately before welding. Depending on the material, sheet thicknesses between 0.1 to 15 mm (0.004 to $\frac{19}{32}$ in.) can be welded.

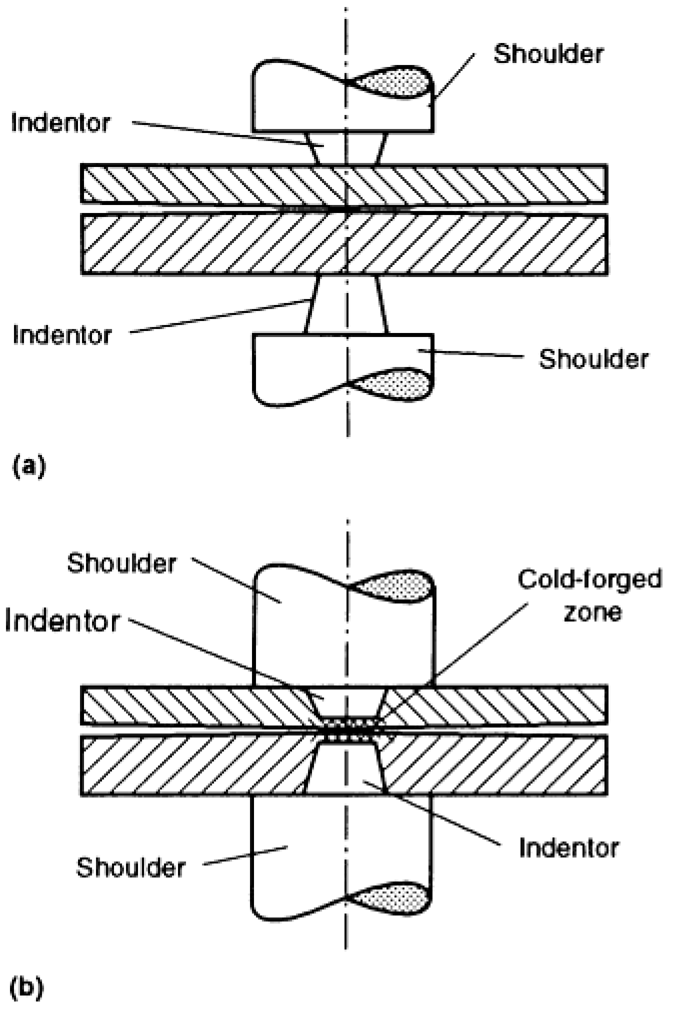

FIG. 2 SCHEMATIC SHOWING TWO DIES BEING USED FOR COLD PRESSURE LAP WELDING OF TWO METALLIC SHEETS. (A) DIE POSITION BEFORE WELDING. (B) DIE POSITION AFTER WELDING

The welds are produced by annular, point, or line-shaped pressure dies in mechanical or hydraulic presses or pneumatic clamps. The reduction in thickness produced by the dies must be limited by a solid stop or deformation measurement method. The bending distortion of the sheets is small when several spots or adjacent narrow areas are welded at the same time. Dies that generate a ring-shaped area of pressure produce welds superior to dies that generate a circular area of pressure. For line welds, dies with relatively long and narrow pressure areas should be used. The width of the pressure area and its distance to an adjacent pressure area will be greater with increasing thickness of the plate. For example, coils of aluminum and copper for electric machines were welded to each other and to connecting pieces. The joints consisted of two aluminum parts having cross-sections measuring 7 by 1.6 mm (0.3 by 0.063 in.) and welds between aluminum measuring 7 by 1.6 mm (0.3 by 0.063 in.) and copper measuring 7 by 1 mm (0.3 by 0.039 in.) cross section (Ref 4). It is possible to produce cold pressure lap welds with very large cross sections if the proper tooling and required pressure forces are available.

Cold pressure butt welding implies that the ends of two bars of similar or dissimilar metals are jointly upset. The two bars are placed in a device or machine having suitable clamping shoes (Fig. 3). The compressive force applied along the axis of the bars causes an expansion of the contact surfaces and forms a bulb. Both bar ends will weld together when the contact surfaces have reached a certain size, which depends on both the material and the condition of the contact surface. The strength of the joint increases as the contact surface expands until it reaches the strength of the softer of the two materials used. The strength of the softer material is thus enhanced by cold work hardening.

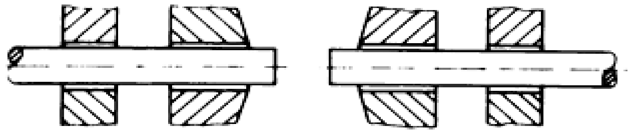

Bars clamped in dies

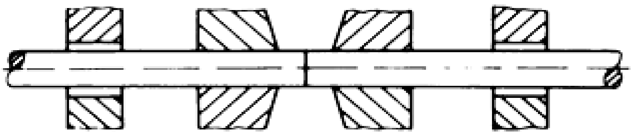

Beginning of first upsetting operation

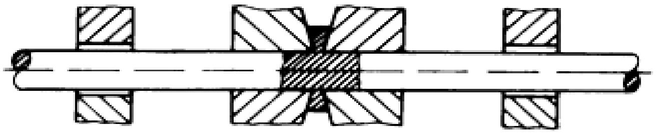

End of first upsetting operation

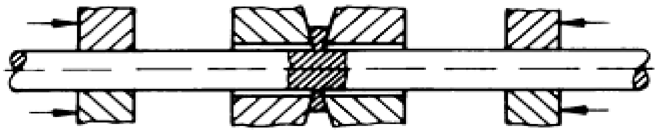

Initial loading by auxiliary clamp jaws along the bar; removal of main clamp jaws

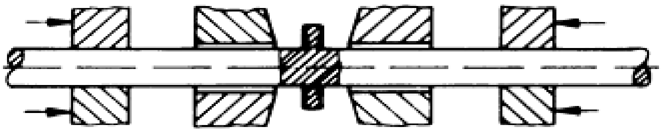

Movement of main clamp jaws

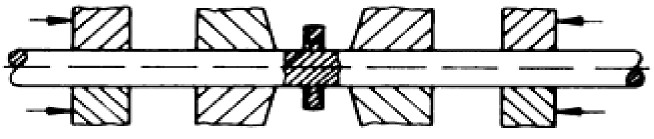

Main clamp jaws grip ends of each bar

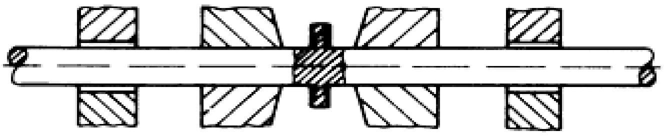

Beginning of second upsetting sequence

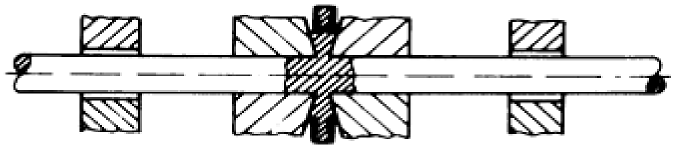

End of second upsetting sequence

FIG. 3 COLD PRESSURE BUTT WELDING OF TWO BARS USING MULTIPLE-STEP UPSETTING METHOD

Cold pressure butt welding may require one or more steps of upsetting. The parameter characterizing the weld is the surface extension parameter, V_o:

$$Vo = \left(\frac{A_s}{A_o}\right) - 1 \qquad \textbf{(EQ 1)}$$

where A_s is the size of the contact surface after welding and A_o is the original surface area. The extension of the surface area can be particularly large when several steps of upsetting are used on materials that are easily cold-formed. When single-step upsetting and a rather simple upsetting device are used, the material ends that project from the clamping shoes may buckle. This limits the surface expansion that can be obtained.

When single-step upsetting is applied on machines made for this purpose and in multiple-step upsetting, buckling is not a consideration, because only very short rod ends will project from the clamping shoes. Because the clamping shoes with an initial loading are transposed in multiple-step upsetting, no cracks can form in the weld during transposition (Fig. 3). No gases from the ambient atmosphere can contact the exposed surface and, hence, no new oxides form.

Many devices have been developed for cold pressure butt welding. There are hand-held tongs for wires in diameters up to 1 mm (0.04 in.) and even an oil-hydraulically powered welding machine with a programmable device to interface with a multiple-step upsetting system. The largest known system applies upsetting forces as high as 1.2 MN (135 tonf), which permits the joining of aluminum up to 1500 mm^2 (2.3 $in.^2$) in cross-sectional area and of copper up to 1000 mm^2 (1.55 $in.^2$) in cross-sectional area (Ref 4). A system capable of butt welding of plastic-coated aluminum ribbons of 0.1 to 0.2 mm (0.004 to 0.008 in.) thickness and up to 50 mm (2 in.) wide has been constructed. This system can complete as many as 99 upsetting steps to produce a weld (Ref 5).

Figure 4 is a schematic of the mechanical part of a servohydraulic, computer-aided cold pressure welding machine. The base, which consists of a base plate and two vertical side plates, houses two horizontally arranged structural members. These structural members support two traverses, one of which can be moved by two hydraulic actuators that operate in unison. The second traverse is held by several binder plugs that have to be loosened only for changeover to a new welding job. Bolted to the traverse are two dies that support the blanks to be welded. The dies, in turn, are attached to two fixtures to which the piston rods of the hydraulic actuators are bolted. Each hydraulic actuator has a sensor to monitor the inductive force acting on the piston rod and to record the inductive displacement. Software is available to produce butt-welded joints by two methods. In cold pressure butt welding with continuous thrust, the piston rods of the 630 kN cylinders are displaced by the same amount. The 250 kN cylinders maintain their positions as set by the computer program. They determine the thickness of the upset. Three bladelike cutters are arranged on the spot faces on the dies. They produce notches in the forming flash that cause the flash to rupture under circumferential tension. While weld material is flowing, the flash is automatically sheared off. The surface extensions obtained are very large. Because of the high tool load, this method should be used only with materials of low tensile strength and moderate work-hardening properties.

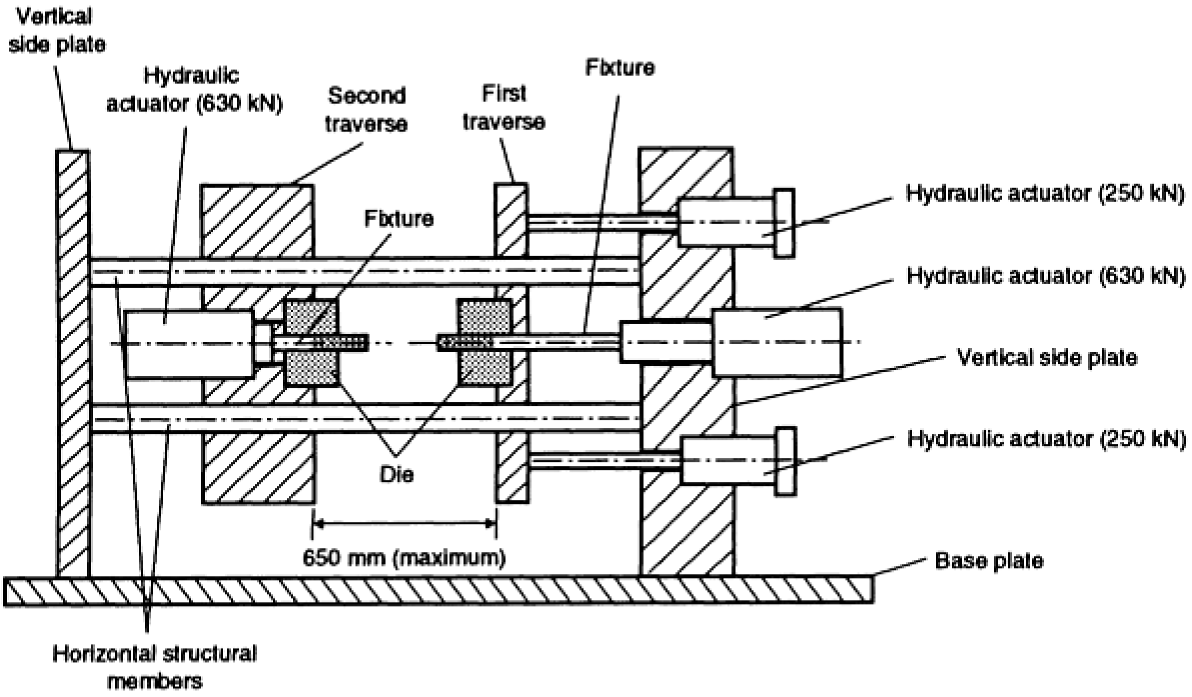
Vertical side plate
Hydraulic actuator (630 kN)
Second traverse
First traverse
Fixture
Fixture
Hydraulic actuator (250 kN)
Hydraulic actuator (630 kN)
Vertical side plate
Hydraulic actuator (250 kN)
Die
650 mm (maximum)
Base plate
Horizontal structural members

FIG. 4 SCHEMATIC SHOWING SERVOHYDRAULIC, COMPUTER-AIDED COLD PRESSURE WELDING MACHINE CONSISTING OF TWO 250 KN (28 TONF) ACTUATORS AND TWO 630 KN (71 TONF) ACTUATORS

In the second method, multiple-step upsetting, the computer controls a sequence of operations that corresponds to the schematic in Fig. 3. Large surface extensions are obtained with this method, but the tool load is lower than in the continuous thrust method. Figure 5 illustrates two copper-aluminum joints produced by the two methods in which the upset metal was not removed automatically (Ref 6).

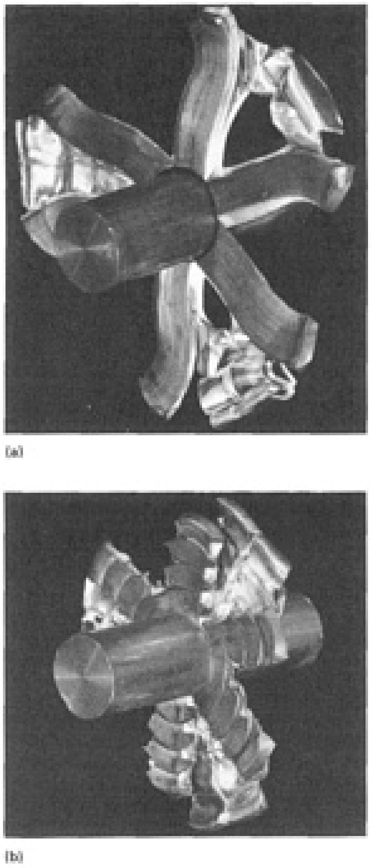

FIG. 5 COPPER-ALUMINUM JOINT OBTAINED BY BUTT WELDING OF TWO 20 MM ($\frac{25}{32}$ IN.) COMPONENTS USING TWO DIFFERENT TECHNIQUES. (A) CONTINUOUS THRUST METHOD. (B) MULTIPLE-STEP UPSETTING METHOD

An example of multiple-step upsetting is the production of interlocking connections of copper and aluminum components for installation in national power grids. The connecting pieces are welded by the single-step upsetting method from 12 to 45 mm ($\frac{15}{32}$ to $1\frac{3}{4}$ in.) diameter blanks. Before welding, the contact surfaces on the blanks are machined with a dry milling tool. The milling tool is periodically cleaned with a solvent. In a manual operation, the blanks are inserted in the upsetting die of a power press without touching the contact surfaces. They are welded together and ejected, and the flash is immediately removed by subsequent machining. Before the mechanical operation is finished, a connecting plate is attached to the copper side of the component by deformation and a hole with a cross-section formed by backward extrusion is made in the aluminum. By deforming the copper closely adjacent to the seam, the quality of the joint can be safely assessed because welds of insufficient strength would rupture along the seam (Ref 7).

Cold Pressure Welding in Drawing Processes. By common deformation in the drawing tool of a drawing bench, multilayer tubes or solid multilayer (clad) bars can be made that have numerous inside and outside layers that consist of different materials (Fig. 6). In this variant also, the contact surfaces must be thoroughly cleaned before welding. This need for cleanliness limits the initial tube length to 800 mm (32 in.). High-intensity compressive stresses exist between the commonly deformed contact surfaces during the drawing operation. The extension of the contact surfaces along the longitudinal axis of the bar is the essential characteristic of the weld. Shear strength values of about 50 MPa (7 ksi) have been reported for copper-aluminum joints after a surface extension of 300% (Ref 8).

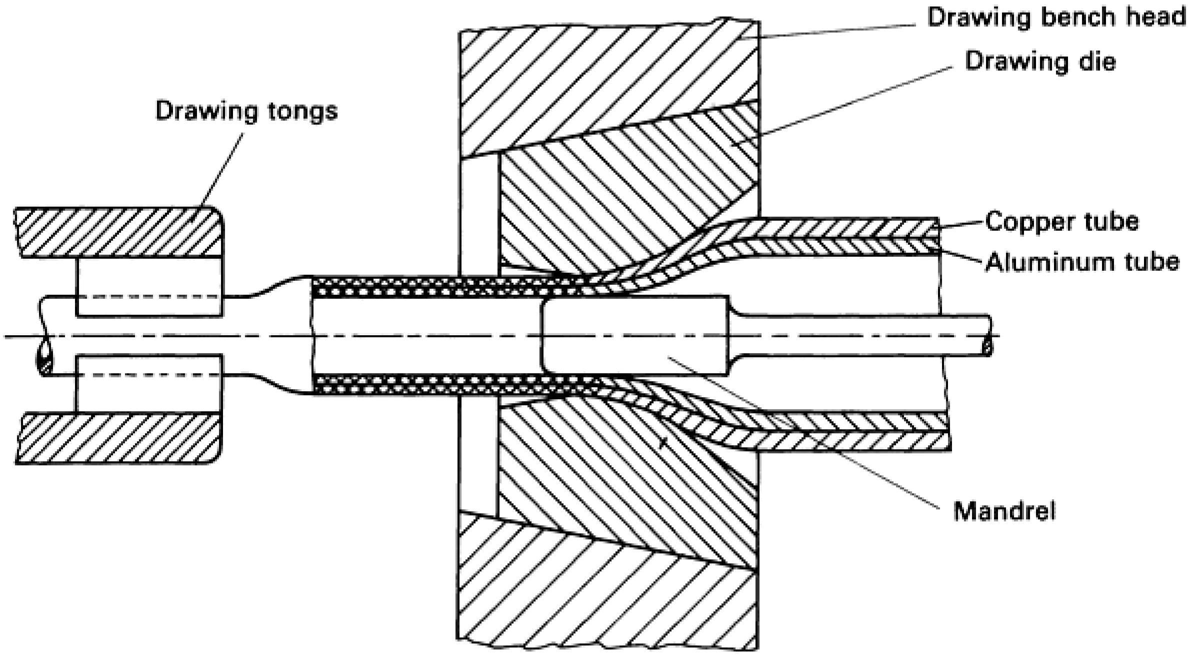

FIG. 6 APPLICATION OF DRAWING TO COLD PRESSURE WELD MULTILAYER TUBES INCORPORATING COPPER ON THE OUTSIDE DIAMETER AND ALUMINUM ON THE INSIDE DIAMETER

References cited in this section

4. A. STEFANESCU AND T. ZRNA, SVAROVÁNÍ TLAKEM ZA STUDENA, SNTL--NAKLADATATELSTVÍ TECHNICKÉ LITERATURY, 1975
5. J. RUGE, SOLID STATE WELDING, COLD PRESSURE WELDING, DIFFUSION WELDING, SPIN WELDING, *DVS REPORTS,* NO. 65, 1980, P 41-46
6. K. THOMAS, APPLICATION OF PROCESS CONTROL COMPUTERS IN SPIN WELDING AND COLD PRESSURE WELDING, *DVS REPORTS,* NO. 120, 1989, P 167-185
7. J. RUGE, H. PRIES, AND K. THOMAS, SURVEY ON THE ACTUAL STAGE AND DEVELOPMENT TRENDS IN COLD PRESSURE WELDING, *DVS REPORTS,* NO. 139, 1991, P 25-32
8. J. RUGE, *MANUAL OF WELDING TECHNICS,* VOL II, 2ND ED., SPRINGER-VERLAG, 1980

Cold Welding

Karl Thomas, Technische Universität Braunschweig; Manfred Petri, Gerhard Petri GmbH + Co. KG

References

1. H. KREYE AND K. THOMAS, ELECTRON-MICROSCOPICAL TESTS AND THE BONDING MECHANISM OF COLD PRESSURE WELDING, *WELD. CUTTING,* VOL 29, 1977, P 249-252
2. H. PRIES, TESTS ON THE BONDING MECHANISM OF WELDING METALS TO NON-METALS IN ULTRA-HIGH VACUUM, *VDI PROGRESS REPORTS,* SERIES 5, *PRIMARY AND MANUFACTURING MATERIALS,* NO. 89, VDI, 1985.
3. N. BAY, COLD WELDING: CHARACTERISTICS, BONDING MECHANISMS, BOND STRENGTH (PART 1), *MET. CONSTR.,* VOL 18 (NO. 8), JUNE 1986, P 369-372
4. A. STEFANESCU AND T. ZRNA, SVAROVÁNÍ TLAKEM ZA STUDENA, SNTL--

NAKLADATATELSTVÍ TECHNICKÉ LITERATURY, 1975

5. J. RUGE, SOLID STATE WELDING, COLD PRESSURE WELDING, DIFFUSION WELDING, SPIN WELDING, *DVS REPORTS,* NO. 65, 1980, P 41-46
6. K. THOMAS, APPLICATION OF PROCESS CONTROL COMPUTERS IN SPIN WELDING AND COLD PRESSURE WELDING, *DVS REPORTS,* NO. 120, 1989, P 167-185
7. J. RUGE, H. PRIES, AND K. THOMAS, SURVEY ON THE ACTUAL STAGE AND DEVELOPMENT TRENDS IN COLD PRESSURE WELDING, *DVS REPORTS,* NO. 139, 1991, P 25-32
8. J. RUGE, *MANUAL OF WELDING TECHNICS,* VOL II, 2ND ED., SPRINGER-VERLAG, 1980

Coextrusion Welding

Introduction

COEXTRUSION WELDING (CEW) is a solid-state process that produces a weld by heating two or more workpieces to the welding temperature and forcing them through an extrusion die. The process typically is conducted at elevated temperatures not only to improve welding but also to lower extrusion pressures. Some cold coextrusion welding of aluminum and copper has been performed. For hot coextrusion, the parts to be welded are often assembled in a can or retort that is designed with the appropriate leading taper and wall thickness to promote initiation of extrusion. For reactive metals, such as zirconium, titanium, and tantalum, the retort may be evacuated and sealed. Both forward and back coextrusion have been employed, but forward coextrusion is the usual mode. A principal advantage of coextrusion welding is that the high isostatic pressures associated with the process are favorable to the deformation welding of low-ductility alloys.

In a similar process, extrusion welding has been used to butt weld tubes. The ends of the tubes are prepared for extrusion by beveling at a 45 to 60° angle to produce an overlapping joint (Fig. 1). The leading tube contains the female portion of the beveled joint and is the stronger of the two metals in dissimilar metal joints. Extrusion press die angles of 30 to 35° are common. An advantage of extrusion welding over other methods of deformation butt welding of tubes is that there is no flash or upset to remove following extrusion.

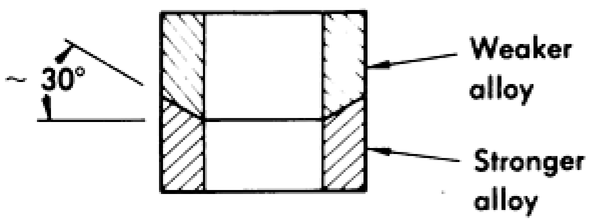

Before extrusion

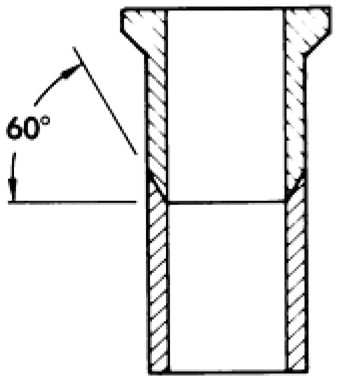

After extrusion

FIG. 1 DESIGN FOR BUTT WELDING OF TUBES BY EXTRUSION

Coextrusion Welding

Common Metals Welded

The most common metals welded by the coextrusion process include low-carbon steel, aluminum, aluminum alloys, copper, and copper alloys. Additional applicable materials include nickel, nickel-base alloys, zirconium, titanium, tantalum, and niobium.

Roll Welding

Introduction

ROLL WELDING (ROW) is a process in which two or more sheets or plates are stacked together and then passed through rolls until sufficient deformation has occurred to produce solid-state welds.

Roll Welding

Process Description

Two modes of roll welding are common. In the first, the parts to be welded are merely stacked and passed through the rolls. The second method, usually termed pack rolling, involves sealing the parts to be rolled in a pack or sheath and then roll welding the pack assembly.

The first method is more generally employed in the cold welding of ductile metals and alloys. Sometimes the stack to be welded is first tack welded at several locations to ensure alignment during rolling. Also, when using this method, the deformation during the first rolling pass must exceed the threshold for welding (typically greater than 60% for cold rolling) to keep the parts together. The required first pass reduction can be reduced by hot rolling, if the metals to be rolled can tolerate preheating without excessive oxidation. Once the first pass has been accomplished, the reduction per pass can be decreased, as is often desirable because roll-separating forces increase as the parts to be rolled become thinner. However, the nonuniform stress distribution that builds up during a sequence of very light passes can cause the weld to open up or "alligator." Therefore, the reduction for subsequent passes is generally a compromise between applying excessive separating forces and "alligatoring."

In pack roll welding, the parts to be welded are completely enclosed in a pack that is sealed (typically by fusion welding) and often evacuated to provide a vacuum atmosphere. This may be accomplished by a frame that surrounds the parts to be welded, which is sandwiched by two lids, or may simply consist of two covers formed to encapsulate the parts to be welded (Fig. 1). Semikilled or killed low-carbon steel is a common material for the pack, but is not suitable for all alloy and temperature combinations. Although the preparation costs of pack roll welding are significant, the process has the advantages of (1) providing atmospheric protection, which may be particularly important for reactive alloys such as those of titanium, zirconium, niobium, and tantalum; and (2) permitting welding of complex assemblies involving several layers of parts. A significant limitation of the process is that packs become difficult to process when their length exceeds several feet.

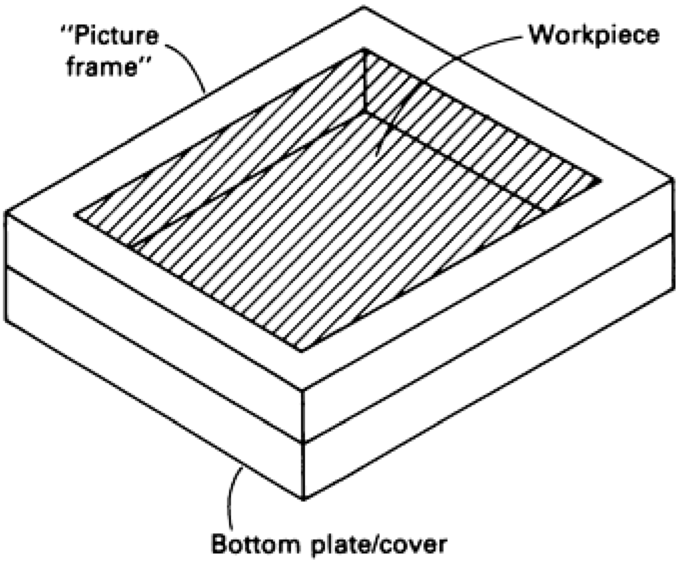

FIG. 1 PACK USED FOR ROLL WELDING. THE TOP COVER, WHICH WOULD BE WELDED IN PLACE BEFORE ROLLING, IS NOT SHOWN.

Titanium and titanium alloys are considerably more difficult to pack roll than are either steels or copper and aluminum alloys because most titanium alloys have very narrow working temperature ranges. To overcome this problem, titanium alloys that are pack rolled are sometimes encased in a steel envelope, or can. The can is evacuated to minimize oxidation of the work metal, and also serves to minimize heat loss to the relatively cold rolls upon deformation. The narrow working temperature range of titanium alloys makes the rolling of these materials labor intensive. Rolling passes are done by hand on relatively small pieces, and many intermediate reheating steps are required. Some β titanium alloys, however, are being continuously rolled on hot and cold strip mills.

Nonpack rolling, of course, can employ continuous strip rolling. A seldom used hybrid version of roll welding is to seal the edges of the parts to be welded by fusion welding prior to rolling. This method does not provide atmospheric protection to the outer surfaces, as in pack roll welding, but does occlude air from the more critical inner surfaces.

In roll welding, successful welding depends not only on the total amount of deformation but also on the pressure during rolling. A high isostatic pressure component during rolling is promoted by large reductions per pass, large roll diameters, and rough roll surfaces. A nonuniform distribution of normal stresses, referred to as pressure hill effects, is caused by frictional and mechanical restraint and results in lower pressure near the outer edges of the parts being rolled. Therefore, rolling conditions must be such to ensure sufficient pressure near the edges to avoid weak welds in these regions.

Copper, aluminum, and their alloys are often cold roll welded. Hot rolling is applicable to a wide spectrum of alloys. Typically, hot rolling is performed above recrystallization temperatures. In practice, weldability is often improved by raising the temperature to the region where oxygen mobility becomes significant to promote oxide assimilation by diffusion. Also, welded assemblies are often postheat treated to cause recrystallization (in cold rolling), oxygen diffusion, and/or sintering. In welding dissimilar metals, selection of temperatures for rolling and postheating must take into consideration the possibility of intermetallic compound formation, eutectic melting, and/or porosity due to Kirkendall diffusion. The most common metals that are roll welded are low-carbon steels, aluminum and aluminum alloys, copper and copper alloys, and nickel. Low-alloy steels, high-alloy steels, nickel-base alloys, and titanium and titanium alloys have also been roll welded.

Applications

One of the more important applications of roll welding is the fabrication of heat exchangers. A patented roll welding process is described in the section "Roll Welding Heat Exchangers" in this article. This process is used to fabricate flat solar collectors. In another method of heat-exchanger fabrication, three plates are stacked, with the middle plate containing the flow channels. The stacked plates are initially held together by tack fusion welds or resistance spot welds. This approach has been employed for both aluminum and steel heat exchangers. A second important application of this process is in the cladding of sheet metal products.

Roll Welded Heat Exchangers. A principal application of roll welding is the production of heat-exchanger panels with flow tubes as an integral part of the composite. Millions of refrigerator evaporator plates have been formed by this process. A cross section of a similar application, solar collector panels, is illustrated in Fig. 2. Roll welded heat-exchanger panels are most often made of aluminum, although copper and stainless steel panels have also been produced commercially.

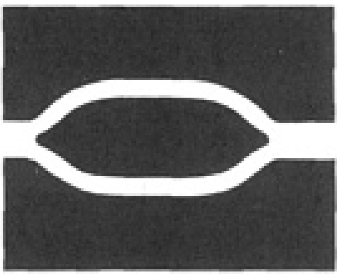

FIG. 2 CROSS SECTION OF AN ALUMINUM SOLAR COLLECTOR PANEL PRODUCED BY THE ROLL WELDING PROCESS

The first major step in the production of roll welded heat exchangers is to thoroughly clean the surfaces of the sheets to be joined by combinations of solvent cleaning, chemical cleaning, and wire brushing. Next, a stop-weld pattern is reproduced on one of the sheets that will later form the passageways in the heat exchanger. This is accomplished by transferring ink onto the surface by silkscreen processing. These stop-off inks contain materials such as graphite or titanium oxide.

The prepared plates are placed together and temporarily joined by spot welding to form a sandwich that is then roll welded, with at least partial welding occurring during the first rolling pass. The sandwich sometimes is heated prior to this first rolling pass. Whether hot rolling or cold rolling is employed, the major reduction takes place in this first pass (50 to 80%). Subsequently, cold rolling may be employed to reduce the panel to final gage. Also, panels are generally given a postweld heat treatment. This heat treatment results in an annealed part and, particularly in the case of cold rolled panels, may improve the weld quality through recrystallization and diffusion. Finally, the passageways are formed by inflating them with air pressure while the sandwich is held between platens. The panel is then ready for trimming, additional forming, and painting, as required for the application.

The process has been used to produce heat exchangers for refrigerators, solar panels, poultry incubators, electronic equipment temperature controllers, industrial heat exchangers, and special applications such as thermal control of spacecraft. Modifications of the process permit as many as six sheets to be welded at one time and inclusion of passageways oriented along different principal directions at the different layers. Also, it is possible to produce panels where the passageways are expanded on one side only.

Cladding of Metals by Strip Roll Welding. The common early application of strip roll welding was the fabrication of bimetallic strips for thermostats. This remains an important application, with thermostats taking many complex forms and finding use in furnaces, color televisions, cars, and numerous industrial controls.

A key cladding application is the production of coins by the U.S. Mint. Silver shortages in the 1960s and 1970s resulted in the introduction of coins made from new materials. These coins required a unique set of properties for acceptance by the general public and use in automatic vending machines. Copper clad with cupro-nickel was found to meet these requirements (Fig. 3). The cladding process is relatively simple, although stringent in its requirement for surface preparation. Cladded strips are produced by continuous rolling with the surfaces prepared just prior to rolling by processes such as wire brushing. Welding typically is accomplished in a single rolling pass. Subsequent heat treatments may be employed to improve the weld quality by processes such as sintering, diffusion, and recrystallization.

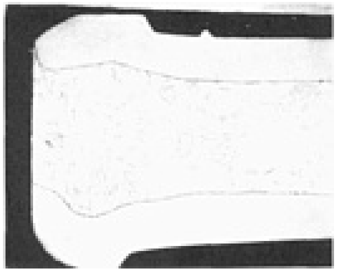

FIG. 3 METALLOGRAPHIC CROSS SECTION OF THE EDGE OF A CUPRONICKEL-CLAD COPPER COIN

The starting material for the U.S. quarter is two outer layers of 75Cu-25Ni, each 1.2 mm (0.048 in.) thick, with an inner layer of pure copper that is 5.1 mm (0.20 in.) thick. To obtain good bond strength, the surfaces are chemically cleaned and wire brushed. The strips are then roll welded to a combined thickness of 2.29 mm (0.090 in.) (Fig. 4). A second rolling operation reduces the final thickness 1.36 mm (0.0535 in.). The strips thus undergo a total reduction of 82% (Ref 1).

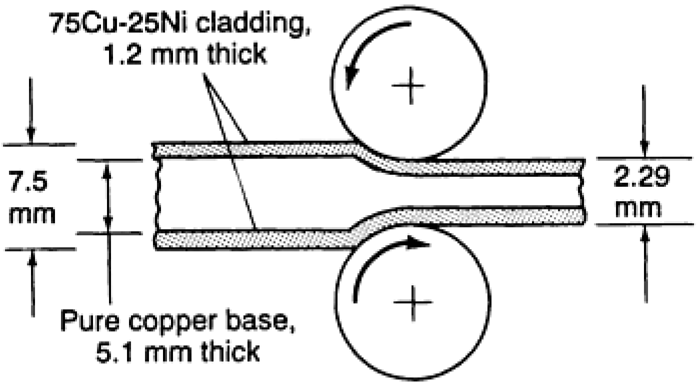

FIG. 4 SCHEMATIC SHOWING 70% REDUCTION IN CROSS SECTION OF A U.S. MINT QUARTER COIN (COMPOSED OF AN INSIDE LAYER OF PURE COPPER SANDWICHED BETWEEN TWO THIN LAYERS OF CUPRONICKEL MATERIAL) AS IT UNDERGOES AN INITIAL ROLL WELDING OPERATION. A SECOND ROLLING OPERATION FURTHER REDUCES THE MATERIAL TO A FINAL THICKNESS OF 1.36 MM (0.0535 IN.) FOR A TOTAL REDUCTION OF 82%. SOURCE: REF 1

Similar processing is applied to many new products, including electrical contacts, conductive springs, automotive windshield wiper sockets, automotive trim, clad cookware, wrap for underground cable, and electromagnetic shielding (Table 1).

TABLE 1 TYPICAL PROPERTIES OF COMMON ROLL WELDED CLAD LAMINATES

MATERIALS SYSTEM	**COMPOSITE RATIO, %**	**THICKNESS**		**WIDTH**		**TENSILE STRENGTH**		**YIELD STRENGTH**		**ELONGATION, %**	APPLICATIONS
		MM	**IN.**	**MM**	**IN.**	**MPA**	**KSI**	**MPA**	**KSI**		
COPPER 10300/LOW-CARBON STEEL	50:50	0.25-2.54	0.010-0.100	12.7-50	0.5-2	260(A)	38(A)	215(A)	31(A)	40(A)	**TYPICALLY USED FOR COMMUTATORS IN ELECTRIC MOTORS, REPLACING COPPER 10400. THE LOWER CONDUCTIVITY OF STEEL CAUSES LESS HEAT LOSS DURING WELDING, RESULTING IN FASTER, BETTER QUALITY WELDS. THE ADDED STRENGTH OF STEEL ALLOWS THICKNESS REDUCTION.**
DEOXIDIZED COPPER/CARBON STEEL/DEOXIDIZED COPPER	5:90:5, 10:80:10, 15:70:15	0.25-2.54	0.010-0.100	≤ 610	≤ 24	420(B)	61(B)	250(B)	36(B)	35(B)	**USED IN HEAT EXCHANGERS, REPLACING BRAZING SHIM MATERIAL. ONE-PIECE MANUFACTURING ELIMINATES SEPARATE SHIMS AND REDUCES ASSEMBLY COST. BECAUSE NO FERROUS SURFACES ARE EXPOSED, NO FLUX IS NEEDED.**
TYPE 434 STAINLESS/5052 ALUMINUM	40:60	0.56-0.76	0.022-0.030	≤ 610	≤ 24	395	57	360	52	12	**WIDELY USED FOR AUTOMOTIVE BODY MOLDINGS, DRIP RAILS, ROCKER PANELS, AND OTHER TRIM COMPONENTS, OFTEN REPLACING SOLID STAINLESS STEEL OR ALUMINUM. STAINLESS STEEL PROVIDES BRIGHT APPEARANCE; THE HIDDEN ALUMINUM BASE PROVIDES CATHODIC PROTECTION, CORRODING SACRIFICIALLY TO THE BODY STEEL.**
COPPER 11000/1100 ALUMINUM	50:50	0.51-2.54	0.020-0.100	50-510	2-20	180(C)	26(C)	12(C)	1.7	35(C)	**DEVELOPED SPECIFICALLY FOR TRANSITION JOINTS BETWEEN COPPER AND ALUMINUM IN REFRIGERATION EQUIPMENT. USUALLY USED IN THE FORM OF DEEP-DRAWN TUBING.**
C1008 STEEL/TYPE 347 STAINLESS STEEL/ C1008 STEEL	43:10:45	0.36	0.014	305	12	393	57	195	28	35	**USED IN HYDRAULIC TUBING IN VEHICLES, REPLACING TERNE-COATED CARBON STEEL TUBING. THE OUTER LAYER OF CARBON STEEL CATHODICALLY PROTECTS THE STAINLESS CORE OF THE TUBE, EXTENDING ITS LIFE SIGNIFICANTLY.**

NICKEL 201/TYPE 304 STAINLESS STEEL/NICKEL 201	7.5:85:7.5	0.20-2.41	0.008-0.095	25-64	1-2.5	. . .	. . .	310	45	40	**USED IN FORMED CANS FOR TRANSISTOR AND BUTTON CELL BATTERIES, REPLACING SOLID NICKEL AT A LOWER COST.**
COPPER 10300/TYPE 430 STAINLESS STEEL/ COPPER 10300	17:66:17, 20:60:20, 33:34:33	0.10-0.15	0.004-0.006	12.7-150	0.5-6	415(D)	60(D)	275	40	20(D)	**REPLACES HEAVIER GAPES OF COPPER AND BRONZE IN BURIED COMMUNICATIONS CABLE. THE STAINLESS STEEL PROVIDES RESISTANCE TO GNAWING BY RODENTS, WHICH IS A SERIOUS PROBLEM IN UNDERGROUND INSTALLATIONS.**
PHOSPHOR BRONZE 51000/CARBON STEEL/COPPER	**10:85:5**	**0.38-1.52**	**0.015-0.060**	**≤ 610**	**≤ 24**	**455(E)**	**66(E)**	**435**	**63**	**16(E)**	**USED IN AUTOMOTIVE APPLICATIONS, SUCH AS WINDSHIELD-WIPER BEARINGS. THE BRONZE PROVIDES THE BEARING QUALITIES, THE STEEL PROVIDES STRENGTH, AND THE COPPER AIDS TOOLING LIFE AND PROVIDES CORROSION RESISTANCE IN SERVICE.**

Source: Ref 2

(A)	IN EIGHTH-HARD TEMPER.
(B)	10:80:10 MATERIAL.
(C)	IN THREE-QUARTER-HARD TEMPER.
(D)	20:60:20 MATERIAL.
(E)	IN QUARTER-HARD TEMPER.

Essentially all combinations of ductile metals and alloys can be clad by roll welding, although the oxides of some metals may make welding difficult and other combinations may suffer from intermetallic compound formation on heating. Most engineering metals and alloys can be processed into clad-laminate form, but some cannot (Table 2). One characteristic that makes some metals difficult or impossible to combine by deformation is low ductility. Such metals cannot be co-reduced to the extent necessary to achieve a metallurgical bond. Another characteristic that can cause problems is the tendency of some metals to form a tenacious oxide film that inhibits weld formation. Finally, some metals are thermally unstable; they form brittle intermetallic compounds when heated above a certain temperature. Those metals that are readily processed into clad laminates have opposite characteristics. They are ductile and easily cleaned of oxides, and they form no intermetallic compounds. In most cases, when a metal is named, alloys of that metal also apply.

TABLE 2 RELATIVE WELDABILITY OF SELECTED DISSIMILAR METALS AND ALLOYS ROLL WELDED INTO CLAD-LAMINATE FORM

BASE METAL NO. 1	BASE METAL NO. 2																		
	Ag	**Al**	**Al alloys**	**Au**	**Carbon Steel**	**Co**	**Cu**	**Mn**	**Mn-Ni**	**Nb**	**Ni**	**Pt**	**Stainless Steel**	**Steel**	**Sn**	**Ta**	**Ti**	**U**	Zr
AG	. . .	. . .	. . .	. . .	. . .	. . .	A	B	. . .	. . .	. . .	. . .	. . .	B	. . .	. . .	. . .	. . .	**. . .**
AL	. . .	. . .	A	C	B	C	B	. . .	. . .	. . .	B	. . .	B	. . .	. . .	. . .	. . .	. . .	**C**
ALFESIL	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	**D**
BE	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	**D**
CARBON STEEL	. . .	B	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	B	. . .	. . .	. . .	B	. . .	**. . .**
CU	A	B	. . .	A	. . .	. . .	. . .	B	B	B	A	B	B	A	A	B	B	. . .	**B**
MN	B	. . .	. . .	. . .	. . .	. . .	B	. . .	. . .	. . .	A	B	. . .	. . .	. . .	. . .	. . .	. . .	**. . .**
NI	. . .	B	. . .	A	. . .	. . .	A	. . .	. . .	. . .	. . .	B	. . .	B	A	. . .	. . .	. . .	**. . .**
NB	. . .	. . .	. . .	. . .	. . .	. . .	B	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	B	. . .	. . .	**. . .**
STAINLESS STEEL	. . .	B	. . .	. . .	B	. . .	B	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	B	. . .	**B**
STEEL	B	. . .	. . .	. . .	. . .	. . .	A	. . .	. . .	. . .	B	. . .	. . .	. . .	. . .	. . .	. . .	. . .	**B**
U	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**B**

A, easy to weld; B, difficult but possible to weld; C, impractical to weld; D, impossible to weld.

Source: Ref 2

References cited in this section

1. S. KALPAKJIAN, *MANUFACTURING PROCESSES FOR ENGINEERING MATERIALS,* ADDISON-WESLEY, 1984, P 703-704
2. R.G. DELAGI, "DESIGNING WITH CLAD METALS," METALLURGICAL MATERIALS DIVISION OF TEXAS INSTRUMENTS, INC.

Roll Welding

References

1. S. KALPAKJIAN, *MANUFACTURING PROCESSES FOR ENGINEERING MATERIALS*, ADDISON-WESLEY, 1984, P 703-704
2. R.G. DELAGI, "DESIGNING WITH CLAD METALS," METALLURGICAL MATERIALS DIVISION OF TEXAS INSTRUMENTS, INC.

Friction Welding

Schillings Tsang, EG&G Rocky Flats

Introduction

FRICTION WELDING (FRW), in its simplest form, involves two axially aligned parts. While one part is rotated, the other stationary part is advanced to make pressure contact. Axial force then increases to generate the frictional heat necessary for welding at the abutting surfaces in order to form a solid-state joint. Friction welding can be divided into two major process variations, depending on the manner by which rotational energy is converted into frictional heat. The first process, direct-drive, or continuous-drive, FRW, has been used commercially since the 1940s. It requires constant energy from a source for any desired duration. The second process, inertia-drive FRW, which was developed in the early 1960s, uses the kinetic energy stored in a rotating flywheel.

Friction Welding

Schillings Tsang, EG&G Rocky Flats

Direct-Drive Friction Welding

Figure 1 shows the layout of a direct-drive FRW system. The spindle is first driven to a predetermined constant speed, and the two parts are brought together under a preset axial force. Both rotation and force are maintained for a specific period determined either by a time or a distance, so that the frictional heat will raise temperatures at the abutting surfaces enough to render the material plastic and suitable for welding. (However, direct-drive friction welds are almost never made using a single level of axial load. The vast majority of welds are made using a minimum of two axial force levels. The second axial load is basically added to the beginning of the weld cycle to yield a preheating phase. In fact, direct-drive friction welds using three axial loads are more commonly applied than those using a single load.) The spindle is then disengaged from the driving unit, and a brake is applied to bring the spindle to rest. At the same time, axial force either remains unchanged or is raised to complete the weld.

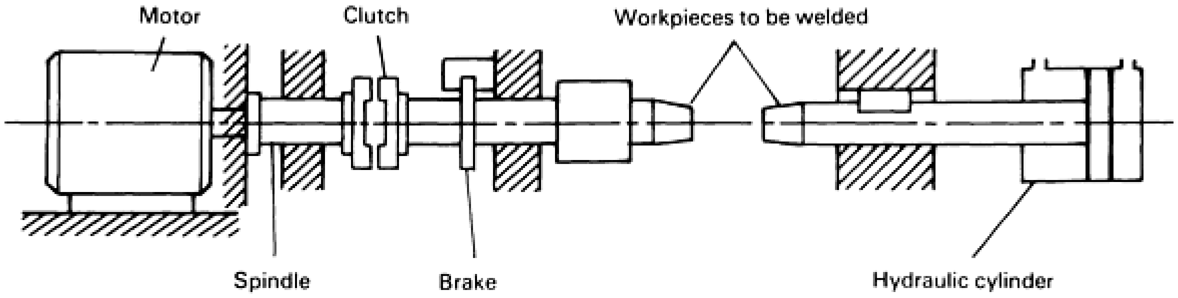
Motor
Clutch
Workpieces to be welded
Spindle
Brake
Hydraulic cylinder

FIG. 1 SCHEMATIC SHOWING RELATION OF WORKPIECES TO KEY COMPONENTS OF A DIRECT-DRIVE FRW SYSTEM. SOURCE: REF 1

Reference cited in this section

1. INERTIA WELDING: SIMPLE IN PRINCIPLE AND APPLICATION, *WELD. MET. FABRIC.*, VOL 47 (NO. 8), 1979, P 585

Friction Welding

Schillings Tsang, EG&G Rocky Flats

Inertia-Drive Friction Welding

In inertia-drive FRW (Fig. 2), a flywheel and the rotating part are mounted in a spindle, which is driven to the desired speed. The drive source is then disengaged, and the two parts make contact under a preset axial force. The free-rotating flywheel decelerates under either the same applied force or later under a larger force. Meanwhile, kinetic energy stored in the flywheel and spindle is converted to frictional heat at the abutting surfaces. The weld is complete when the flywheel comes to a stop. A subsequent higher forging force may be used after the flywheel has stopped.

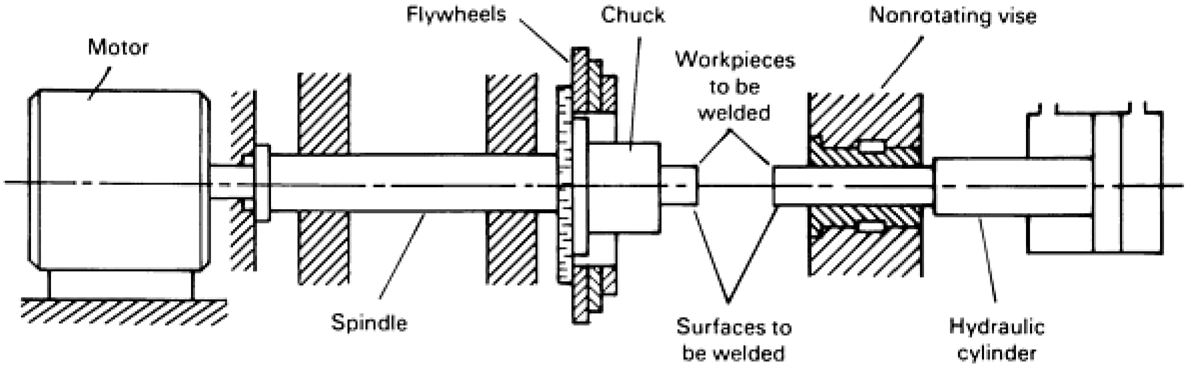
Flywheels
Chuck
Nonrotating vise
Motor
Workpieces to be welded
Spindle
Surfaces to be welded
Hydraulic cylinder

FIG. 2 SCHEMATIC SHOWING RELATION OF WORKPIECES TO KEY COMPONENTS OF AN INERTIA-DRIVE FRW SYSTEM. SOURCE: REF 1

Reference cited in this section

1. INERTIA WELDING: SIMPLE IN PRINCIPLE AND APPLICATION, *WELD. MET. FABRIC.*, VOL 47

(NO. 8), 1979, P 585

Welding Parameters

Direct-Drive FRW Variables. Three parameters control the character of a weld in direct-drive FRW: rotational speed, duration of rotation, and axial force. During welding, there is not only an axial shortening of the part length, often called axial displacement or upset, but also a resisting torque of friction to rotation, which also undergoes change. Figure 3 shows the change in various events occurring throughout the whole process. Based on the shape of the friction torque curve, it is convenient to divide direct-drive FRW into three phases:

- *PHASE 1:* INITIAL FRICTION (THAT IS, BREAK-IN OR FIRST FRICTION) PHASE
- *PHASE 2:* HEATING (THAT IS, FRICTION) PHASE
- *PHASE 3:* FORGING (THAT IS, UPSETTING) PHASE

In phase 1, the torque rises rapidly after the start of the process. It then peaks and drops before leveling off when phase 1 ends. The rapid rise and gradual fall of torque is associated with the interlocking and breaking of asperities and subsequent softening of the material at the faying surfaces by frictional heating.

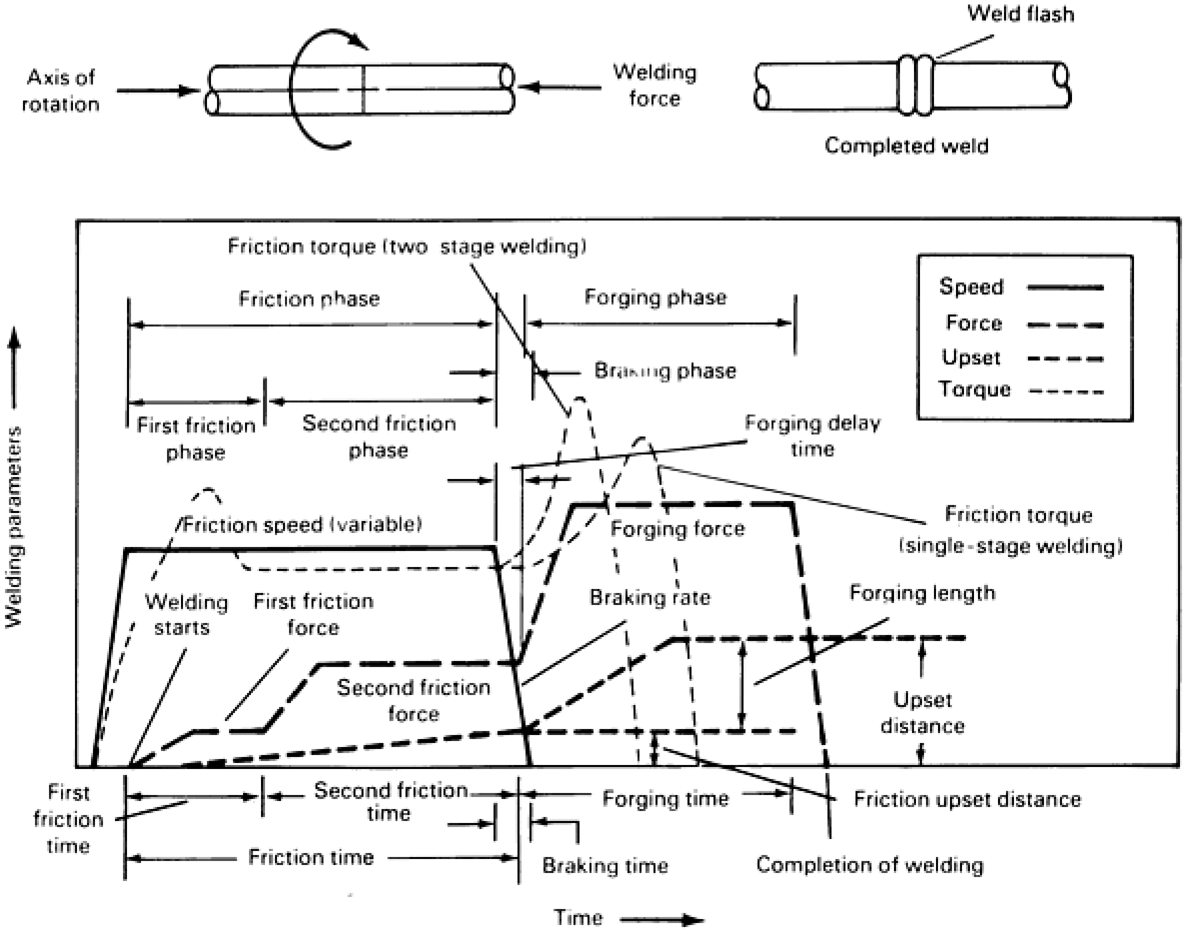

FIG. 3 PLOT OF WELDING PARAMETERS VERSUS TIME FOR A DIRECT-DRIVE FRW SYSTEM. COURTESY OF D.L.

Friction torque remains somewhat constant in phase 2, indicating that the process reaches a balance of effects between strain hardening and thermal softening. Both the faying surfaces and material immediately behind are by then sufficiently heated to permit the two parts to be forged together.

Forging takes place in phase 3, which starts at the time of declutching and braking. Spindle rotation is immediately retarded, and the deceleration depends on braking time. Because the brake force itself can be set, braking time can also be a variable. Deceleration varies in the welding of different materials. However, the torque may not rise at all but drops abruptly at the onset of phase 3 if braking is sudden. This brings the rotating spindle to rest almost instantly, and the larger force is applied after spindle rotation stops. In this case, there is no torsional forge, and forging is brought about by upsetting.

Axial force in phase 3 is usually increased to effect forging. The friction torque again rises after the onset of this phase, reaching another peak before sharply falling off to zero. This peak varies with deceleration and applied axial force. Under some circumstances, this final peak can be omitted by delaying the onset of the forging force.

This is done only at or after the stoppage of spindle rotation by braking. In this case, forging is carried out by upsetting without torsional forging. Although two-stage welding is applied more frequently in direct-drive friction welding, one-stage welding is sometimes used, especially in research work (Ref 2).

When the axial force remains unchanged, slower deceleration (longer braking time by reduced braking) leads to higher peak. If the axial force is increased in phase 3, braking time is shortened, but the peak still rises because of the larger applied force. Rising friction torque in phase 3, characteristic of FRW, contributes to torsional forge, which is more effective than conventional forge by upsetting. Phase 3 ends shortly after spindle rotation stops. This period of forging or upsetting time can also be considered a fourth welding parameter.

Inertia-Drive FRW Variables. There are three welding parameters in inertia-drive FRW: flywheel mass (expressed by moment of inertia), rotational speed, and axial force. The events described for direct-drive FRW also occur in inertia-drive FRW. Except for rotational speed, the curves shown in Fig. 4 are similar to those shown in Fig. 3. The process can also be divided into three phases based on the shape of the friction-torque curve. Unlike direct-drive FRW, duration of rotation and forging time in phase 3 are not predetermined but controlled by the three welding parameters.

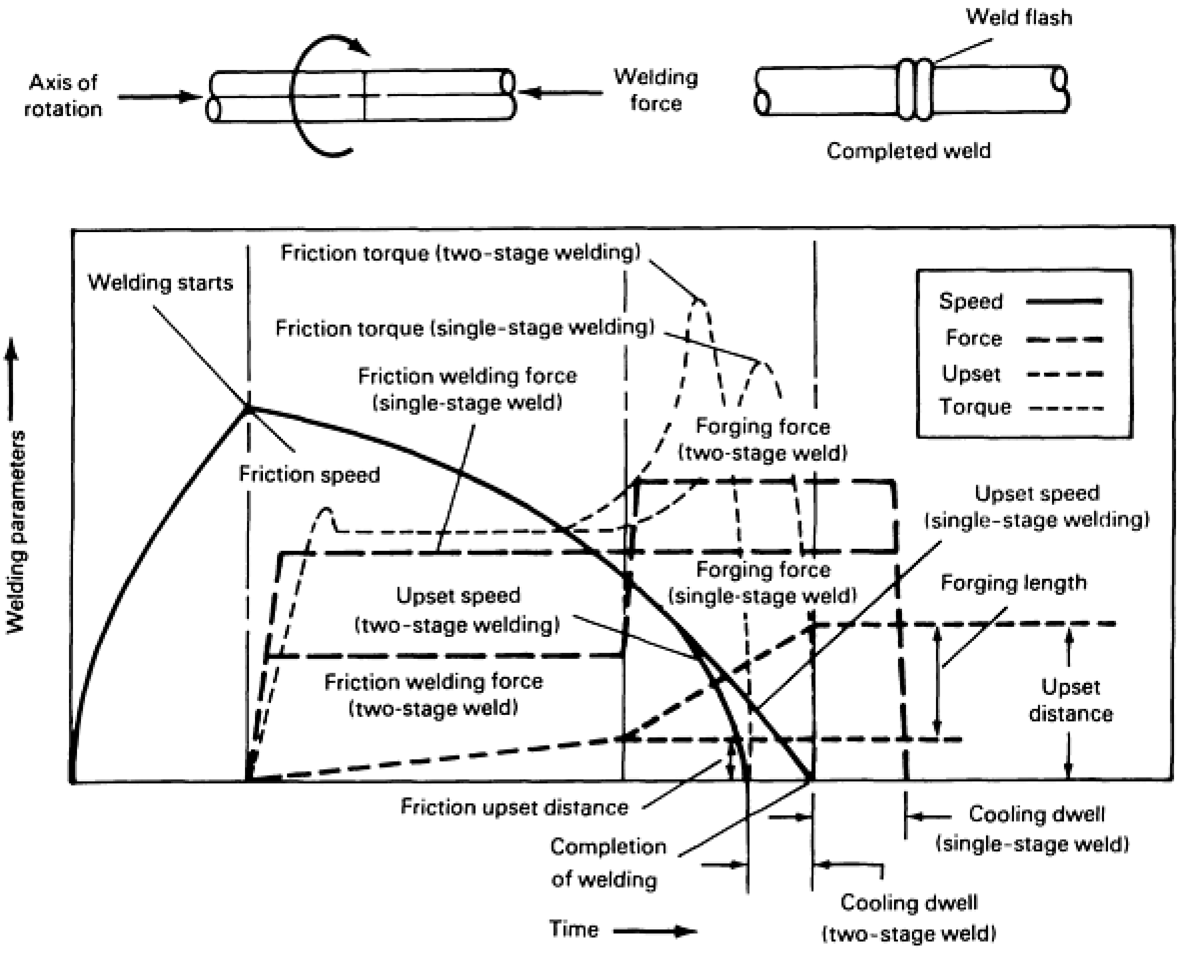

FIG. 4 PLOT OF WELDING PARAMETERS VERSUS TIME FOR AN INERTIA-DRIVE FRW SYSTEM. COURTESY OF D.L. KURUZAR, MANUFACTURING TECHNOLOGY, INC.

If axial force differs in phases 2 and 3, the process is called two-stage welding, and the forces, called heating force and forging force, respectively, are both included as a welding parameter of axial force. On the other hand, when the force remains constant throughout the process, it is called one-stage welding. The second friction-torque peak in two-stage welding is generally higher than that in one-stage welding because of the larger axial force applied in phase 3. As stated in the section "Direct-Drive FRW Variables" in this article, the peak in one-stage welding depends on braking time: It rises with longer braking time.

In two-stage welding, the larger axial forging force is applied when the flywheel rotational speed decreases to a fraction of the initial value--for example, 0.1 to 0.5. Friction torque again rises to a second peak, which as in direct-drive FRW, is higher than in one-stage welding. This phase is prolonged beyond cessation of the flywheel rotation. This extra time, called dwell time, is used for cooling of the joint to avoid debonding by the release of stored strain energy. Although dwell times can range from 0 to 30 s, 1 to 2 s is typical.

The kinetic energy used for inertia-drive FRW depends on the product of the flywheel moment of inertia and the square of rotational speed. The frictional heat generated at the faying surfaces, however, is determined by the rate of change of this energy (or the energy delivery rate), which is influenced by the moment of inertia of the flywheel and the axial force. For a given kinetic energy input and a given axial force, a smaller flywheel (with smaller moment of inertia) delivers the energy at the faying surfaces faster than a larger flywheel (with larger moment of inertia). For a given energy input using the same flywheel, a larger axial force increases flywheel deceleration rates, intensifies heat generation, and raises temperatures at the faying surfaces. A rapid energy input and a high rate of heat generation, typical of inertia-drive FRW, result in very short weld cycles. There is little time for heat to dissipate in the axial direction, so the heat-affected zone (HAZ) remains narrow.

Direct-Drive Versus Inertia-Drive FRW. Lower rotational speed and smaller axial force are generally used in direct-drive FRW. Direct-drive FRW also offers greater latitude in process variables, because heating time in phases 1 and 2 is preset as duration of rotation or upset distance, which is a welding parameter, and because forging phase can be controlled by braking time, axial force, and forging time. Nevertheless, both processes do produce welds of equal quality when proper welding parameters are implemented. Torsional forging is required when forging by upsetting with inertia-drive friction welding; it is not required when forging by upsetting with direct-drive friction welding. A comparison of the two processes is presented in Table 1.

TABLE 1 PROCESS PARAMETERS OF DIRECT-DRIVE FRW SYSTEMS RELATIVE TO INERTIA-DRIVE FRW SYSTEMS

It is assumed that similar or dissimilar metals of the same size are being joined.

PROCESS PARAMETER	**DIRECT-DRIVE FRW**	INERTIA-DRIVE FRW
WELDING PARAMETERS	ROTATIONAL SPEED	**FLYWHEEL MOMENT OF INERTIA**
	DURATION OF ROTATION OR FRICTION BURNOFF DISTANCE	**ROTATIONAL SPEED**
	AXIAL FORCE	**AXIAL FORCE**
CONVERSION OF ENERGY TO FRICTIONAL HEAT	CONSTANT ENERGY IN A PRESET DURATION OF ROTATION	**FIXED ENERGY IN FLYWHEEL; DURATION OF ROTATION DETERMINED BY WELDING PARAMETERS**
ENERGY INPUT	LOW	**HIGH**
HEAT GENERATION RATE	LOW	**HIGH**
CYCLE TIME	SIMILAR	**SIMILAR**
HAZ WIDTH	WIDE	**NARROW**
MACHINE SYSTEM SPINDLE RIGIDITY	**LESS RIGIDITY**	**GREATER RIGIDITY AND MORE POWERFUL CLAMPING FORCES REQUIRED**

Reference cited in this section

2. C.R.G. ELLIS, CONTINUOUS DRIVE FRICTION WELDING OF MILD STEEL, *WELD. J.*, VOL 51 (NO. 4), 1972, P 183-185

Friction Welding

Schillings Tsang, EG&G Rocky Flats

Weld Quality and Product Evaluation

At present, there is no satisfactory nondestructive method for detecting a friction weld of poor strength. (X-rays are used to identify defects that can cause welds of poor quality.) Weld quality control and product evaluation must rely on the control of the process and the machine that is used to produce the welds.

In direct-drive FRW, both rotational speed and axial force affect axial shortening or upset in addition to the mechanical properties of the weld produced. There can be a correlation between the upset and the quality of the weld. Further, upset in phase 2 increases almost linearly with time. The upset of this phase, often called friction burn off, and its rate, the friction burn off rate, have been believed to control weld quality. Indeed, some researchers use burn off instead of duration of rotation as a welding parameter. In inertia-drive FRW, because duration of rotation is not preset but controlled by the three welding parameters, the total upset, instead of friction burn off, is supposed to indicate weld quality. This presumes that the same amount of kineticenergy and axial force will produce the same amount of upset at the end of the process (given the same materials and cross sections).

The welding parameters of either direct-drive or inertia-drive FRW are set using a particular machine. So once this machine is properly calibrated, the process becomes almost entirely machine-dependent. Regardless of the process used to evaluate the product of friction welding under a given set of parameters, once variations in machine components and dimensional tolerances for a particular weld are determined, an acceptable tolerance band of the total upset can be established as a standard. A monitor will compare the upset of each weld relative to this reference. The machine will automatically stop to indicate a defective weld when its upset falls outside the proven band of the standard.

Advantages and Limitations

Advantages of FRW. Friction welding has a number of advantages over other joining processes (such as fusion welding, brazing, and diffusion bonding):

- SPECIAL ATTENTION TO SURFACE CLEANLINESS IS NOT NECESSARY, BECAUSE FRW TENDS TO DISRUPT, DISPLACE, AND FINALLY REMOVE SURFACE FILMS IN WELD FLASH (THIS DOES NOT APPLY TO HEAVY MILL OR HEAT-TREAT SCALES, WHICH CAN INTERFERE WITH FRICTIONAL HEATING).
- FILLER METAL, FLUX, AND SHIELDING GAS ARE NOT REQUIRED. UNLIKE FUSION WELDING, FRICTION WELDING IS NOT HAZARDOUS TO OPERATOR HEALTH AND IS SAFER, BECAUSE THERE IS NO METAL SPATTER, RADIATION, FUME, OR ELECTRIC HAZARD INVOLVING HIGH VOLTAGE, ARCS, AND SPARKS.
- DEFECTS ASSOCIATED WITH MELTING-SOLIDIFICATION PHENOMENA ARE NOT PRESENT IN FRW, BECAUSE IT IS A SOLID-STATE PROCESS.
- IT IS POSSIBLE TO MAKE TRANSITION JOINTS OF DISSIMILAR METALS THAT ARE DIFFICULT OR EVEN IMPOSSIBLE TO WELD BY OTHER PROCESSES (FOR EXAMPLE, REFRACTORY AND EXOTIC METALS).
- LOWER LABOR COSTS, SIMPLE PART DESIGN, SIMPLER WELD FIXTURES, LOWER ENERGY REQUIREMENTS, AND A SHORT WELD CYCLE (A MATTER OF SECONDS) MAKE THE PROCESS COST EFFECTIVE FOR PRODUCING COMPONENTS NORMALLY MADE BY OTHER PROCESSES.

Limitations of FRW include the following:

- THE WELDING AREA OF AT LEAST ONE PART MUST BE ROTATIONALLY SYMMETRICAL, SO THAT THE PART CAN ROTATE ABOUT THE AXIS OF THE WELDING PLANE. TYPICAL PART GEOMETRIES THAT CAN BE FRICTION WELDED ARE: BAR TO BAR, BAR TO TUBE, BAR TO PLATE, TUBE TO TUBE, AND TUBE TO PLATE.
- THIS PROCESS IS NORMALLY LIMITED TO MAKING FLAT AND ANGULAR (OR CONICAL) BUTT JOINTS.
- THE MATERIAL OF AT LEAST ONE COMPONENT MUST BE PLASTICALLY DEFORMABLE UNDER THE GIVEN WELDING CONDITIONS. FOR EXAMPLE, ALUMINA CAN BE JOINED TO ALUMINUM, BUT ALUMINA CANNOT BE JOINED TO ALUMINA. ANOTHER EXAMPLE IS THE JOINING OF POLYMERS (THAT IS, PLASTICS). THERMOPLASTICS, UNLIKE THERMOSETTING PLASTICS, CAN BE FRICTION WELDED BECAUSE THERMOPLASTICS ARE SOFTENED BY FRICTIONAL HEAT BUT THERMOSETTING PLASTICS DECOMPOSE.

Process Applications

The applications of FRW fall into two categories:

- *BATCH AND JOBBING WORK:* USED WHEN A RELATIVELY LOW VOLUME OF PARTS IS REQUIRED, ESPECIALLY WHEN EXPENSIVE DIES MUST BE USED BUT CANNOT BE JUSTIFIED
- *MASS PRODUCTION:* APPLICATIONS THAT SPAN A WIDE RANGE OF INDUSTRIES (AGRICULTURAL, AIRCRAFT ENGINE, AUTOMOTIVE, EARTHMOVING, ELECTRICAL, PETROLEUM, AND SO ON) PRODUCING COMPONENTS RANGING FROM SIMPLE BUTT JOINTS OF DRIVE SHAFTS AND OIL DRILLING PIPES TO COMPLICATED OR CRITICAL AIRCRAFT ENGINE COMPONENTS

Friction Welding

Schillings Tsang, EG&G Rocky Flats

References

1. INERTIA WELDING: SIMPLE IN PRINCIPLE AND APPLICATION, *WELD. MET. FABRIC.*, VOL 47 (NO. 8), 1979, P 585
2. C.R.G. ELLIS, CONTINUOUS DRIVE FRICTION WELDING OF MILD STEEL, *WELD. J.*, VOL 51 (NO. 4), 1972, P 183-185

Radial Friction Welding

E.D. Nicholas, The Welding Institute

Introduction

THE CONVENTIONAL FRICTION-WELDING PROCEDURE of rotating one component against another while an axial thrust is being applied is well known. Autogenous welds in the solid state are made in one operation. The procedure is an accepted production joining method because it provides excellent weld quality and good productivity, it allows the use of unskilled labor, and it enables a wide range of similar and dissimilar metal combinations to be joined. However, continuous-drive and stored-energy (inertia) friction-welding systems have not been as successfully exploited for the joining of long, hollow sections, where restrictions in the bore cannot be tolerated.

The difficulties involved in rotating an intermediate length of hollow section, and in applying uniform and consistent welding forces to both weld interfaces, represent some of the problems that face the machine designer. Furthermore, because two internal upset metal collars are formed, additional machining of the bore is necessary in order to avoid restriction. In 1975, a new approach, called radial friction welding (Ref 1), attempted to provide a means for overcoming these limitations.

Method. The radial friction-welding process adopts the principle of rotating and compressing a solid ring around two stationary pipe ends (Fig. 1a). The pipes to be welded are bevelled to provide a "V" groove when they are butted together. They are then securely clamped to prevent axial and rotational movement. A solid, internally bevelled ring of compatible material with a bevelled angle that is less than that of the pipes is positioned around the pipe ends.

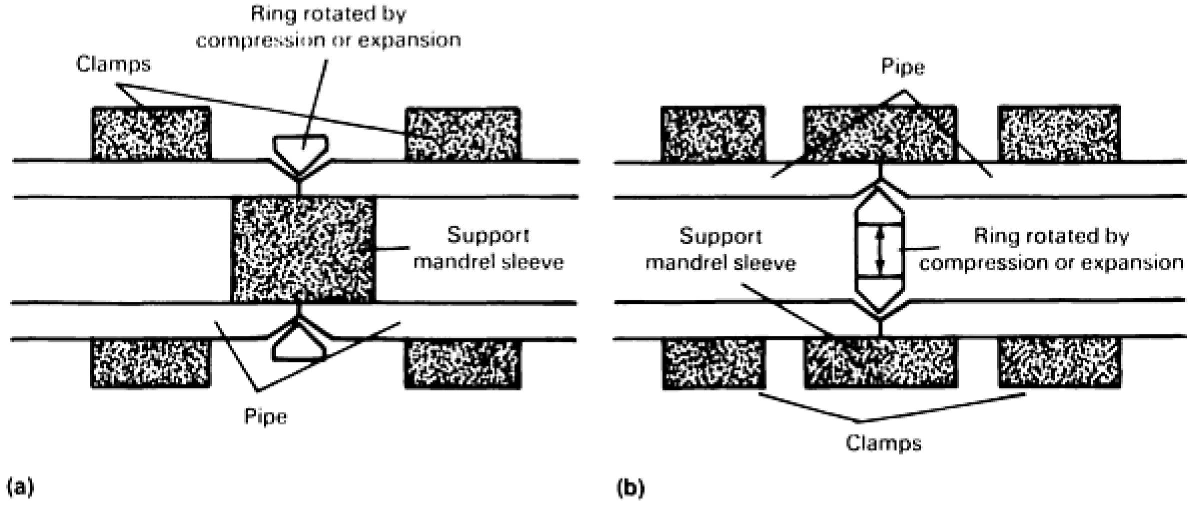
Ring rotated by compression or expansion
Clamps
Support mandrel sleeve
Pipe
(a)
Pipe
Support mandrel sleeve
Ring rotated by compression or expansion
Clamps
(b)

FIG. 1 RADIAL FRICTION WELDING. (A) USING COMPRESSION. (B) USING EXPANSION

The "V" groove for the 108 mm ($4\frac{1}{4}$ in.) outside diameter by 9.5 mm ($\frac{3}{8}$ in.) wall thickness pipe is typically 100°, while the internal bevel of the ring is 80°. It is important for the latter angle to be less than that of the groove in order to limit the initial peak torque and also to promote metal flow from the bottom of the groove in an outward direction toward the outside diameter. Optimum angles for selected sizes of tubes and pipes are determined experimentally.

A support mandrel is located in the bore at the weld position, in order to prevent both the flow of metal into the bore and the collapse of the pipe ends. The ring is then rotated and subjected to radial compressive loading in order to obtain the frictional interaction between the rubbing surfaces, which, in turn, will generate the thermomechanical conditions necessary for weld formation. After a predetermined heating duration, the ring rotation is terminated. The level of compressive load is then either maintained or increased to consolidate the bond. Alternatively, it is also feasible to change the direction of deformation (Fig. 1b). Thus, by providing an expansion load to the ring, the ring can be deformed radially outwards to make the weld.

The included, or bevelled, angle of the solid-compression ring (Fig. 1a) or the expansion ring (Fig. 1b) can have a pronounced effect on weld integrity. The joining of the abutted area is effected and promoted by metal flow from the bore to the outer surface of the pipe. There can be little or no penetration at the pipe root. If the included angle is too narrow, the result will be a weak bond area at the weld root, leading to poor weld strength and a short fatigue life.

Process Development. It was first necessary to prove that the concept of both rotating and compressing/expanding solid rings was feasible. Consequently, conventional friction-welding machines that could provide both rotary motion and axial force were used in the initial development trials. Various bolt-on units were designed, built, and then fitted to these machines.

The results of these initial experiments are summarized in Fig. 2. Although the units were relatively simple, the concept was able to produce sound, solid-state welds with a variety of material combinations. One important outcome of this development effort was that nickel-base alloys were identified as being necessary for the internal mandrel, because they provided good resistance to radial compression loads and wear at the elevated temperatures that were attained during welding. Additional information is available in the section "Prototype/Production Machine" in this article.

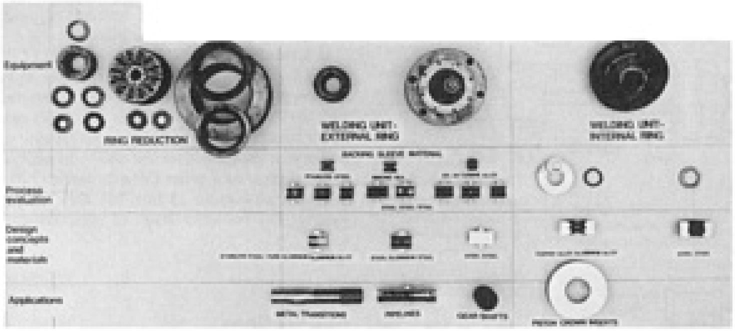
Equipment
RING REDUCTION
WELDING UNIT-
EXTERNAL RING
WELDING UNIT-
INTERNAL RING
BACKING SLEEVE MATERIAL
Process
evaluation
Design
concepts
and
materials
Applications
METAL TRANSITIONS
PIPELINES
GEAR SHAFTS
PISTON CROWN INSERTS

FIG. 2 PROCESS EVOLUTION OF RADIAL FRICTION WELDING

Reference

1. "FRICTION WELDING METHODS AND APPARATUS," U.K. PATENT 1,505,832, 30 MAR 1978

Radial Friction Welding

E.D. Nicholas, The Welding Institute

Equipment

When a prototype radial friction welder was built in the mid-1970s, the primary emphasis was directed toward the application of pipe joining. From the early experimental trials, it was recognized that the shortcomings of using this process in production would be:

- THE NEED FOR AXIAL MOVEMENT OF THE RING THROUGH A TAPER HOUSING
- THE ADDITIONAL MACHINING OF DRIVE GROOVES IN THE RING
- THE RELATIVELY COMPLEX RING GEOMETRY, WHICH WOULD INCLUDE A TAPER SURFACE ON ITS PERIPHERY

Consequently, the idea of using circumferentially located individual jaws (similar to multijaw collet chucks) to provide radial compression and rotation of simple rings was considered. An evaluation program using the arrangement (Ref 2) proved to be successful and, thus, provided the impetus to design and construct a prototype machine based on the multijaw rotation/compression arrangement.

Research Machine. A unique friction-welding facility was developed in order to join pipe of maximum dimensions, that is, 51 mm (2 in.) in outside diameter (OD), by 6 mm (0.24 in.) in wall thickness. The resulting machine is shown in Fig. 3 and its major features are described below.

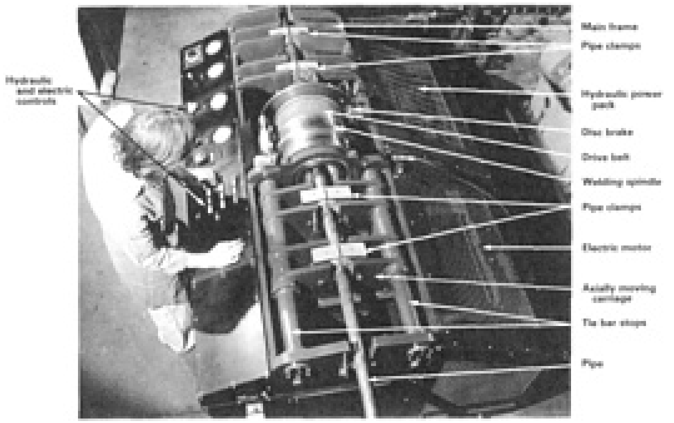

FIG. 3 RADIAL FRICTION-WELDING MACHINE WITH A 51 MM (2 IN.) DIAMETER

Welding Spindle. It was imperative that the ring rotation/compression unit, or welding spindle, be capable of exerting a grip on the ring that was adequate for the transmission of welding power during rotation and of delivering a radial compressive force to reduce the ring into the weld groove. The welding spindle was supposed to incorporate a multijaw arrangement. The heavy steel body of the welding spindle, which needed to contain the radial compressive forces, was supported at each end by simple bearings, and was rotated through a "V"-drive pulley on its outside surface. The spindle was fixed in the machine frame using two simple yoke clamps. A hydraulic caliper disk brake unit was incorporated, when necessary, to assist ring deceleration.

An electric motor rated at 56 kW (75 hp) provided rotation to the welding spindle via a clutch and transmission. The latter enables the setting of rotation speeds ranging from 700 to 2000 rev/min, after a suitable selection of pulley ratios. The clutch and final belt drive were located in line with the welding spindle. The inside of the welding spindle housed an annular piston, which moved in an axial direction, when actuated by hydraulic means, to operate the 12 ring compression jaws via wedges. A maximum axial load of 350 kN (7.8×10^4 lbf) could be generated when oil was supplied to the piston through a rotary oil distributor, at a pressure of 3.4 MPa (0.5 ksi).

The welding spindle had a 340 mm (13.4 in.) OD and was 595 mm (23.4 in.) in length. Consequently, the pipe that passed through it had to be supported close to the weld area by a manually operated collet assembly at the center of the welding spindle.

Pipe Clamping. In order to react to both the axial and torsional forces generated on the pipes, four clamps (two for each pipe) were housed in the top bed of the machine (Fig. 3). The clamps opened and closed by the vertical movement of hydraulic rams that forced wedges between the pivoted jaws. These rams are capable of exerting a maximum thrust of 200 kN (4.5×10^4 lbf).

The machine framework was manufactured as two main assemblies. The machine bed was a heavy-section box-type fabrication that housed the welding spindle and pipe clamps. In contrast, the support framework was of a lighter construction in order to accommodate the motor drive, transmission, and other auxiliary equipment.

Hydraulic Services. The hydraulic circuit had a conventional layout, in that all services were supplied from a dual pump. A gear pump that operated at a pressure of 13.8 MPa (2 ksi) supplied the clutch, bore support mandrel, and moving carriage, whereas a high-volume vane pump that was set to provide a maximum pressure of 3.4 MPa (0.5 ksi) fed only the welding head through a flow-control valve, in order to operate the axially moving annular piston. The latter pump was used to developed the radial compressive forces on the ring. The flow-control valve could be set to provide different rates of piston movement and, thus, varying radial displacement rates to the ring.

Electrical Services. The electrical circuit utilized standard machine controls with relays and timers to sequence the operation of the initial ring preload pressure, the weld heating cycle, the disk brake, the application of forge force, the compression jaws, the mandrel, and the retraction of the moving carriage.

Internal Support Mandrel. The ideal arrangement for a support mandrel should include materials that possess good wear resistance at elevated temperatures and the capability to expand and contract from the pipe bore. The ability of the mandrel to expand proved to be useful to both round and size the pipe ends prior to welding. However, problems were anticipated during the development effort because of the limited working area available for the 50 mm (2 in.) OD pipe (that is, the 38 mm, or 1.5 in., bore), and because of the requirement to eliminate longitudinal gaps in the mandrel.

An expanding mandrel that closed 1 mm (0.039 in.) below the minimum bore diameter was devised. It was operated by a hydraulic cylinder through a draw bar. The mandrel consisted of six segments machined from a nickel-base superalloy (Nimocast PK24). Earlier trials had proved the suitability of this material. Three of the segments were used as wedges, and all segments were operated by the tapered cone on the draw bar, with the wedge segments being removed by the extraction cone first, upon release of the mandrel.

The operating sequence to produce a weld was determined to consist of the steps described below. First, both pipes and ring were machined to provide the necessary weld groove and bevel, respectively. Then, the 16 mm (0.630 in.) wide ring was located in the reduction jaws and held under a light preload pressure.

Next, one pipe was positioned in the welding spindle and clamped, while the collet-steadying device was locked. (The collet-steadying device is of tubular construction and located around that portion of the pipe positioned inside the radial compression. When activated, it provides additional support to the pipe.) Then, the second pipe was fitted with the bore support mandrel, loaded into the machine, and clamped so that it butted against the other pipe.

After setting the rotation speed, as well as the friction welding force timing device that controlled heating duration and the flow-control valve that provided a fixed rate of piston movement (compression rate), the weld sequence was initiated. Upon completion of the weld cycle, the reduction jaws were retracted and the mandrel was sequenced to be extracted after 5 s. To remove the welded pipe, the collet-steadying device was released, along with both pairs of clamps.

Subsequent trials with this research machine demonstrated that the radial friction-welding process could produce welds of acceptable mechanical and metallurgical properties (Ref 3), and that the process could operate in the field, either on land or offshore.

Prototype/Production Machine. During the 1980s, a Norwegian company continued with the development of the process. Its efforts culminated in the design, manufacture, and subsequent commissioning of a prototype/production machine that could weld pipes ranging in size from 89 to 170 mm (3.5 to 6.625 in.) (Ref 4). This machine, which was constructed by another firm based in Oslo Norway, was designed on a twin tie-bar arrangement, upon which the welding head and pipe clamps were mounted (Fig. 4). The welding head, which consists of the drive spindle for rotating the ring, is driven by a 370 kW (500 hp) direct-current motor. The pipe clamps operate hydraulically, as does the ring compression. Additional pipe support and alignment are provided by collets located near and inside the welding head. Although it is not yet fitted as such, the machine has space for a machining station to allow the removal of excess ring/flash metal soon after welding.

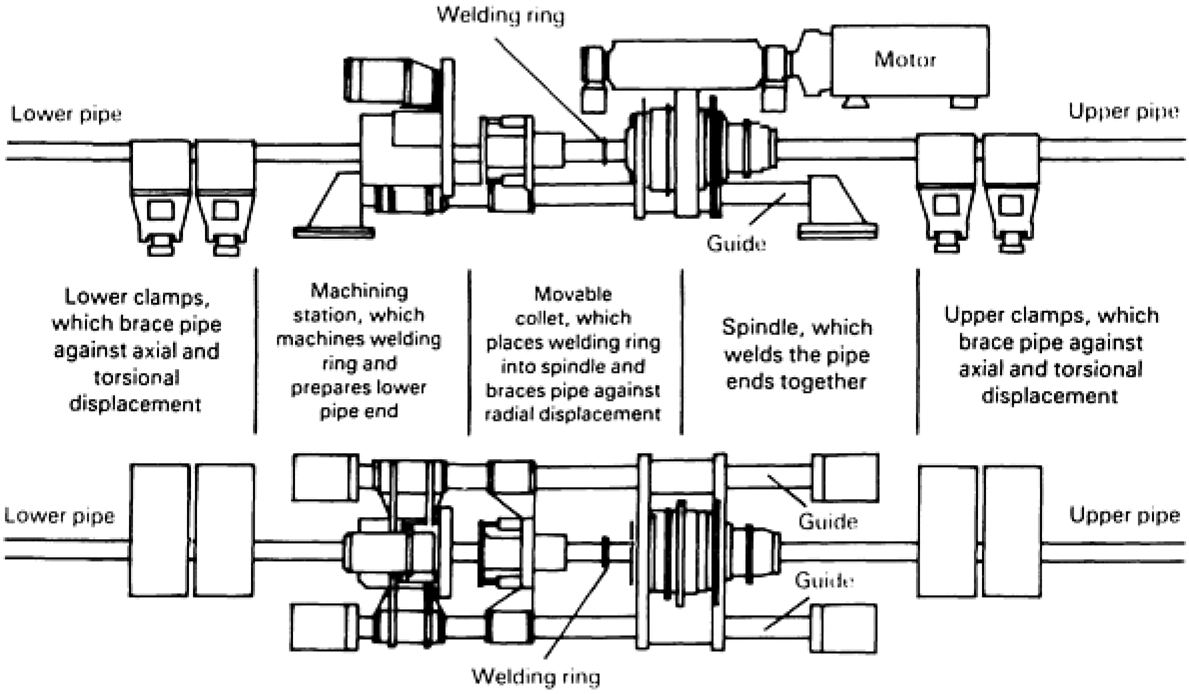
Welding ring
Motor
Lower pipe
Upper pipe
Guide
Lower clamps, which brace pipe against axial and torsional displacement
Machining station, which machines welding ring and prepares lower pipe end
Movable collet, which places welding ring into spindle and braces pipe against radial displacement
Spindle, which welds the pipe ends together
Upper clamps, which brace pipe against axial and torsional displacement
Lower pipe
Guide
Upper pipe
Guide
Welding ring

FIG. 4 SCHEMATIC OF THE PROTOTYPE RADIAL FRICTION-WELDING MACHINE

In a parallel effort to the development of this machine, special compression rigs for radial friction welding were manufactured to interface with 1000 to 1800 kN (110 to 200 tonf) axial thrust friction welders. These units provide experimental facilities with the ability to evaluate the process over a wider OD size range (100 to 273 mm, or 4 to 10.75 in.), not only for pipe joining, but for other areas of application, as well.

References cited in this section

2. E.D. NICHOLAS AND R.H. LILLY, RADIAL FRICTION WELDING, *PROC. CONF. ADVANCES IN WELDING PROCESSES* (HARROGATE, U.K.), 1978, P 48
3. E.D. NICHOLAS, RADIAL FRICTION WELDING, *WELD. J.*, VOL 62 (NO. 7), 1983, P 17
4. S.B. DUNKERTON, A. JOHANSEN, AND S. FRICH, RADIAL FRICTION WELDING FOR OFFSHORE PIPELINES, *WELD. J.*, VOL 66 (NO. 7), 1987, P 40

Radial Friction Welding

E.D. Nicholas, The Welding Institute

Applications

The applications that would be suitable for radial friction welding are:

- ATTACHMENT OF COLLARS TO SHAFTS
- WELDING OF ROTATING DRIVING BANDS TO ARTILLERY SHELLS
- ATTACHMENT OF WEAR/SUPPORT RINGS ONTO CYLINDRICAL BODIES
- PIPE JOINING AT BOTH LAND AND OFFSHORE SITES
- FABRICATION OF SHAFTS OR PIPEWORK SYSTEMS WITH CORRECT ALIGNMENT OF

PARTS
- REPAIR OF PIPELINES, BOTH ABOVE WATER AND UNDERWATER
- MANUFACTURE OF BIMETAL TRANSITIONS USING A COMPATIBLE INTERMEDIATE MATERIAL
- INSERTION OF NONMAGNETIC CENTER PIECES
- INSERTION OF BETTER HEAT- OR WEAR-RESISTANT MATERIALS INTO DIESEL PISTON COMBUSTION BOWLS

Radial Friction Welding

E.D. Nicholas, The Welding Institute

References

1. "FRICTION WELDING METHODS AND APPARATUS," U.K. PATENT 1,505,832, 30 MAR 1978
2. E.D. NICHOLAS AND R.H. LILLY, RADIAL FRICTION WELDING, *PROC. CONF. ADVANCES IN WELDING PROCESSES* (HARROGATE, U.K.), 1978, P 48
3. E.D. NICHOLAS, RADIAL FRICTION WELDING, *WELD. J.*, VOL 62 (NO. 7), 1983, P 17
4. S.B. DUNKERTON, A. JOHANSEN, AND S. FRICH, RADIAL FRICTION WELDING FOR OFFSHORE PIPELINES, *WELD. J.*, VOL 66 (NO. 7), 1987, P 40

Friction Surfacing

E.D. Nicholas, The Welding Institute

Introduction

FRICTION WELDING, a solid-state (non-melting) joining process, relies on the presence of relative motion between the parts while they are being pressed together under an applied axial force to generate the thermomechanical conditions for welding. The process, in its simplest form, makes use of rotary motion. A modification of conventional friction welding makes it possible to deposit layers of metal onto a substrate, a metal-transfer method that is called "friction surfacing."

Although surfacing by friction can be regarded as novel, the original concept was reported, and a patent application filed, in 1941 (Ref 1). The technology was dormant until the late 1950s, when the process was rediscovered by Russian workers (Ref 2). A review of the literature has identified a wide variety of applications that utilized the process at that time:

- SHAFT RECLAMATION (REF 3)
- HARDFACING OF AGRICULTURAL PARTS (REF 4)
- RECONDITIONING BUSHINGS OF SLIDING FRICTION BEARINGS AND ROLL TRUNNIONS (REF 5, 6, 7, 8, 9)
- HARDFACING (REF 10) AND IMPROVING WEAR RESISTANCE IN GAS AND OIL EQUIPMENT (REF 11)

References

1. H. KLOPSTOCK AND A.R. NEELANDS, "AN IMPROVED METHOD OF JOINING AND WELDING METALS," BRITISH PATENT SPECIFICATION 572789, APPLICATION DATE, OCT 1941
2. E. BISHOP, FRICTION WELDING IN THE SOVIET UNION, *WELD. MET. FABR.,* OCT 1960, P 408-410
3. KH.A. TYAYAR, FRICTION WELDING IN THE RECONDITIONING OF WORN COMPONENTS, *SVAR. PROIZVOD.,* VOL 1 (NO. 10), 1959, P 3-24
4. R.I. ZAKSON AND F.G. TURUKIN, FRICTION WELDING AND HARDFACING OF AGRICULTURAL MACHINE PARTS, *AVESTA SVARKA,* VOL 3, 1965, P 48-50
5. V.YA. KERSHENBAUM AND B.A. AVERBUKH, SPECIAL FEATURES IN THE FRICTION DEPOSITION OF BRONZE ON STEEL, *AVESTA SVARKA,* VOL 3, 1964, P 19-22
6. G.K. SCHREIBER, V.YA. KERSHENBAUM, AND S.G. TKACHENKO, INFLUENCE OF FRICTION SURFACING CONDITIONS ON TRANSITION LAYER PROPERTIES IN CLAD METALS, *SVAR, PROIZVOD.,* VOL 9, 1971, P 32-33
7. V.YA. KERSHENBAUM *ET AL.,* PERFORMANCE OF STEEL BRONZE BIMETAL PRODUCED BY FRICTION SURFACING, *SVAR. PROIZVOD.,* VOL 7, 1972, P 29-30
8. V.YA KERSHENBAUM AND B.A. AVERBUKH, OPTIMUM DIMENSIONS OF FRICTION SURFACED COMPONENTS, *AVESTA SVARKA,* VOL 5, 1972, P 64-66
9. B.A. AVERBUKH AND A.L. VELINSKII, THERMAL EFFECTS IN FRICTION SURFACING, *AVESTA SVARKA,* VOL 11, 1975, P 29-31
10. E.I. PLUZHIKOVA, E.L. DEMINA, V.YA. KERSHENBAUM, AND G.K. SCHREIBER, FRICTION FACING OF CAST IRON ON STEEL, TRANSLATED FROM *METALLOVED. TERM. OBRAB. MET.,* VOL 11, 1971, P 46-47
11. M.V. NOGIK *ET AL.,* FRICTION HARDFACING STEEL WITH STELLITE V3K, *SVAR. PROIZVOD.,* VOL 8, 1970, P 16-17

Friction Surfacing

E.D. Nicholas, The Welding Institute

Process

In the friction-surfacing process, a rotating consumable is brought into contact with a moving substrate, which results in a deposited layer on the substrate (Fig. 1). First, the consumable rod is rotated to the desired speed. Next, it is brought into contact, under an axial force, with the stationary substrate. An initial rubbing period is required to allow the generation of a plasticized layer in the consumable. Several seconds are required before this condition is achieved. Then, the substrate is traversed across the consumable, leaving a layer of deposited material. The deposit is characterized by a fine hot-worked microstructure, zero dilution, and a very strong bond with the substrate.

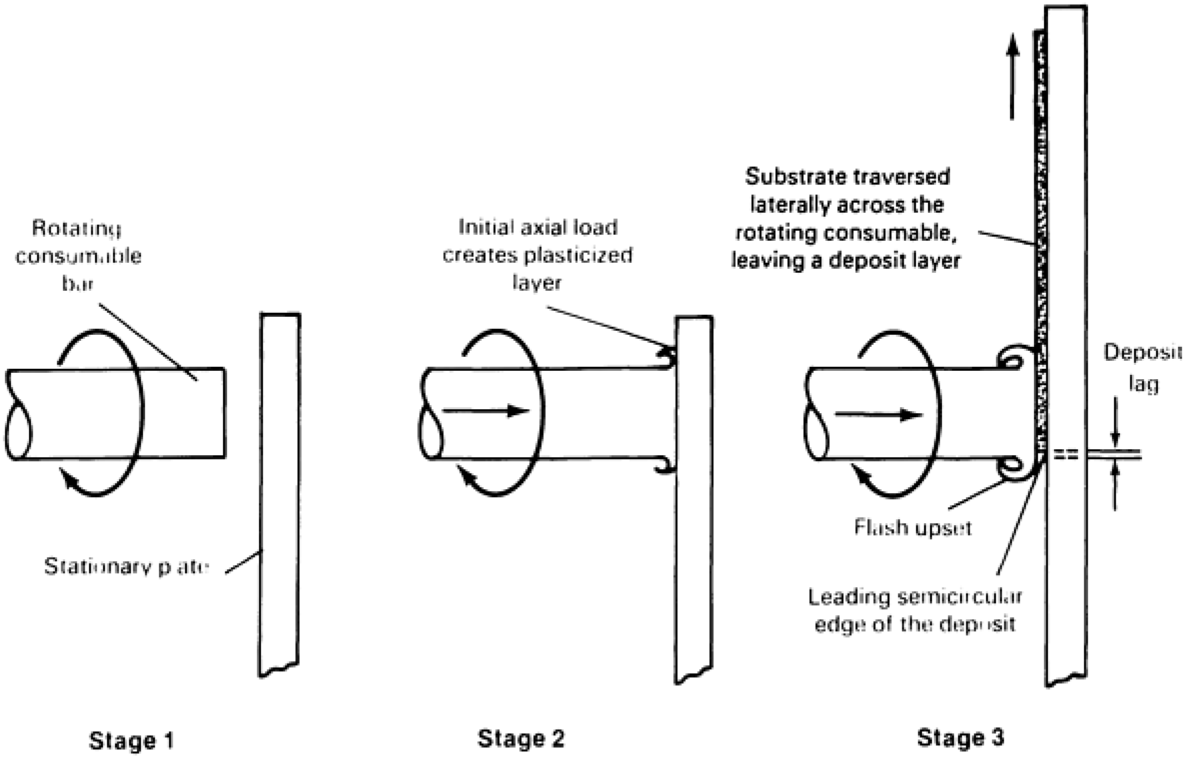
Rotating consumable bar
Stationary plate
Initial axial load creates plasticized layer
Substrate traversed laterally across the rotating consumable, leaving a deposit layer
Deposit lag
Flash upset
Leading semicircular edge of the deposit
Stage 1
Stage 2
Stage 3

FIG. 1 BASIC TECHNIQUE FOR FRICTION SURFACING

The process utilizes equipment that is similar to a machine tool, for which the surfacing conditions are preset. Therefore, the process can provide good reproducibility, and is not dependent on operator skill.

The width and thickness of the deposited layer is influenced by the primary surfacing variables of rotational speed, axial force, substrate traverse rate, consumable diameter, and type of material. Generally, the fully bonded width of the deposit is approximately 1 to 3 mm (0.04 to 0.12 in.) less than the diameter of the consumable, with the extreme angles of the deposit being characterized by lack of bond. It is thought that this lack of bond is associated with nonuniform pressure distribution across the consumable; lack of restraint to the highly plasticized flowing material; and the reduced time in action seen at the outer edges as compared to the center of the consumable.

The deposit thickness in a single pass is considerably influenced by the type of consumable material. Materials that have good resistance to plastic deformation at elevated temperatures, such as the nickel-base alloys and the Stellite alloys, will only yield thicknesses of the order of 0.5 to 1 mm (0.02 to 0.04 in.). When austenitic stainless steel and carbon steels are used, the thickness increases to approximately 3 mm (0.12 in.). For materials with low strength at elevated temperature, such as aluminum and its alloys, thicknesses can approach 5 to 6 mm (0.20 to 0.24 in.).

Representative variables associated with friction surfacing conditions are shown in Table 1. A typical deposit of austenitic stainless steel from a 25 mm (1 in.) diameter rod that has been laid down on mild steel by the friction surfacing process is shown in Fig. 2(a). The layer is characterized by a uniformly rippled surface of consistent width and thickness. A transverse section reveals lack of bond at the edges and sound bond formation between the deposit and substrate, above the narrow, crescent-shaped, heat-affected zone (HAZ) in the latter (Fig. 2b). Figure 2(c) shows the microstructural features at the bond interface. It is important to remember that a uniform and smooth mushroom-shaped upset is continuously formed on the consumable rod during the surfacing operation.

TABLE 1 FRICTION-SURFACING PARAMETERS FOR SELECTED MATERIAL COMBINATIONS

COMBINATION	**CONSUMABLE DIAMETER**		**ROTATION SPEED, REV/MIN**	**APPLIED FORCE**		**TRAVERSE RATE**	
	mm	**in.**		**kn**	**10^3 lbf**	**mm/s**	**in./s**
AUSTENITIC STAINLESS STEEL TO MILD STEEL	25	1.0	550	50	11.3	5	0.20
STELLITE 6 TO AUSTENITIC STAINLESS STEEL	20	0.8	330	50	11.3	2.5	0.10
AL-4CU TO AL-4CU	25	1.0	780	17	3.8	4	0.16

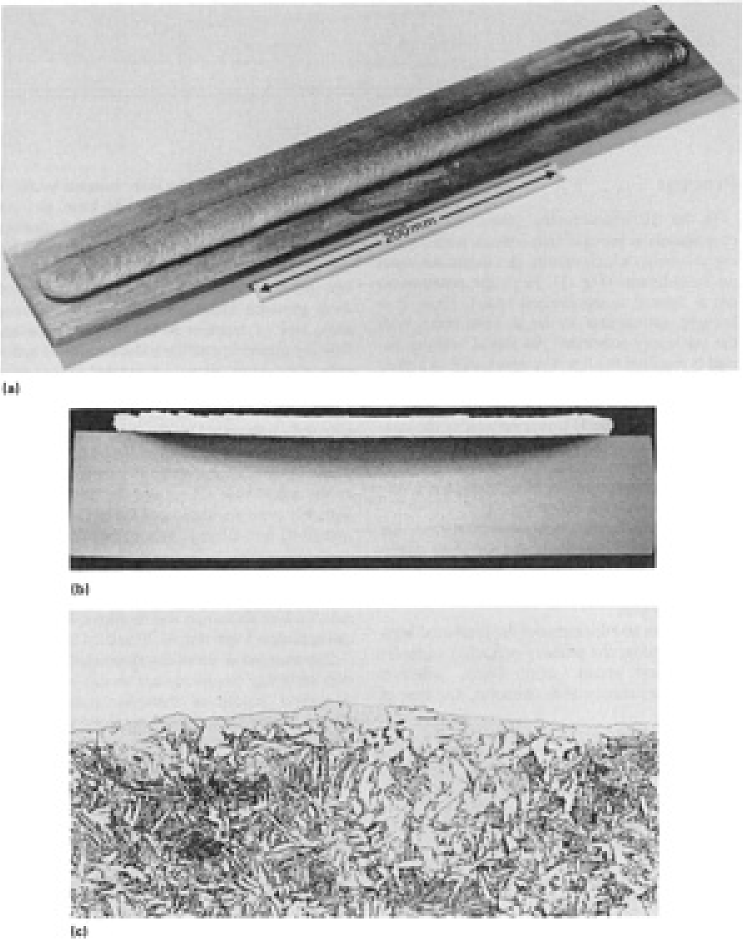

FIG. 2 THREE VIEWS OF 500 MM (20 IN.) FRICTION-SURFACED DEPOSIT OF AUSTENITIC STAINLESS STEEL ON MILD STEEL. (A) VIEW SHOWING ENTIRE SECTION. (B) TRANSVERSE SECTION OF DEPOSIT. (C) MICROSTRUCTURE AT THE DEPOSIT SUBSTRATE WELD INTERFACE. 550×

If it is necessary to clad larger surface areas, then multiple side-by-side layers can be laid down. If thicker deposits are needed, then layers one on top of another can be used. If both approaches are called for, then interpass machining may be required to ensure sound adhesion properties.

Friction Surfacing

E.D. Nicholas, The Welding Institute

Equipment

The earliest feasibility studies were conducted on a modified, continuous-drive, friction-welding machine (Ref 12). However, further practical investigations necessitated an increased surfacing capacity. Therefore, a milling machine (Ref 13) designed for the heaviest class of milling was used for longer deposition runs and larger-diameter components.

Ample power for the steady rotation of 64 spindle speeds between 12 and 540 rev/min is provided by the two-speed reversing motor, which develops 22 kW at 1430 rev/min and 15 kW at 960 rev/min. The robust machine frame minimizes deflection during friction surfacing, and the table, which moves longitudinally and transversely, also is designed to withstand large loads. The table is driven by a 2.2 kW variable-speed motor for 1575 mm (65 in.) of longitudinal travel and 724 mm (30 in.) of transverse travel at a lateral rate between 0.5 and 17 mm/s (0.02 and 0.68 in./s).

A rotary manipulator is used for circumferential deposition onto curved surfaces that are up to 480 mm (20 in.) in diameter. This manipulator is designed to handle the desired radial compressive loads while providing the required peripheral velocity. Similarly, by using a 610 mm (24 in.) diameter rotary table, annular deposits can be applied to the contact face of disks and flanges. Modifications have included fitting a direct-current variable-speed 2.2 kW motor to allow the feed rate of the consumable bar to be varied independently of the table traverse. Changes include a control system that allows consumable feed rates to be changed during the process, in these ways:

- TOUCH DOWN (SLOW FEED RATE, WITH NO LATERAL MOVEMENT)
- RAMP-UP (GRADUAL INCREASE IN THE FEED RATE TO THE SET VALUE)
- CONSTANT FEED RATE DURING LATERAL MOVEMENT
- PAUSE (IN LATERAL MOVEMENT) AT END OF RUN
- WITHDRAWAL OF THE CONSUMABLE (WHILE STILL ROTATING)
- STOP

The ability to use converted machine tools offers a low-cost option in the technology of friction surfacing. A certain manufacturer has recently designed and built a machine dedicated to this surfacing method (Fig. 3). The machine, which is particularly suited for longitudinal deposition, can:

- PUT DOWN RUNS OF 3 M (10 FT)
- USE CONSUMABLES UP TO 70 MM (3 IN.) IN DIAMETER
- PROVIDE A RANGE OF ROTATION SPEEDS FROM 250 TO 1800 REV/MIN
- DEVELOP AXIAL FORCES UP TO A MAXIMUM OF 180 KN (40.5 × 10^3 LBF)

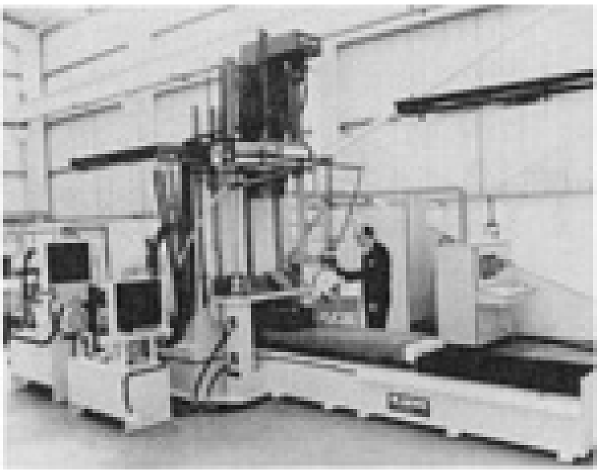

FIG. 3 COMMERCIAL FRICTION-SURFACING MACHINE. COURTESY OF BLACKS EQUIPMENT LTD.

References cited in this section

12. W.M. THOMAS, E.D. NICHOLAS, AND S.B. DUNKERTON, "FEASIBILITY STUDIES INTO SURFACING BY FRICTION WELDING," TWI RESEARCH REPORT 236, THE WELDING INSTITUTE, APRIL 1984
13. W.M. THOMAS, SOLID PHASE CLADDING BY FRICTION SURFACING, *PROC. TWI INTERNATIONAL SYMPOSIUM 18,* APRIL 1988, P 18

Friction Surfacing

E.D. Nicholas, The Welding Institute

Applications

The available methods, in terms of surfacing and surface-treatment technology, are far from complete. The development of friction surfacing provides an alternative (in some cases, complementary) technique. The process may increasingly accompany other solid-state, large-overlay techniques, especially with respect to localized repair and the reclamation of explosively clad and roll-bonded composite materials.

Hardfacing applications on cutting edges and agricultural tools have been reported to be successful (Ref 11). Other workers (Ref 14, 15) suggest that thin-clad sheet can be produced from heavily cold-rolled friction-surfaced deposits. They believe that another promising application is the hermetic sealing of containers with heat-sensitive contents. Friction surfacing also will undoubtedly be applied to the reclamation of worn parts. The reconditioning of worn shafts, in which good-quality deposits were achieved, has been described (Ref 13).

Initially, potential applications for this solidstate technique are in the cladding of localized areas. Typical examples include the anticorrosion surfacing of slide valve plates, the surfacing of the annular contact face of composite pipe flanges, the inlay cladding by suitable strategic materials in positions that suit bearings and seal-contact areas on shafts, and the cladding of the exposed regions of shafts that are used in pernicious environments.

Various substrate geometries that have interesting potential for friction cladding are shown in Fig. 4. Shell banding may be another application, where suitably soft material is friction surfaced to the bodies of artillery shells (Ref 16). Other possibilities include the manufacture of specialized wear tiles (for sinter plant and quarrying equipment) and the in-situ reclamation of worn railway points. Turbine blades, guillotine and shear blades, disk brakes, machine-tipped tools, and press tool dies are also considered to be potential friction-surfacing applications.

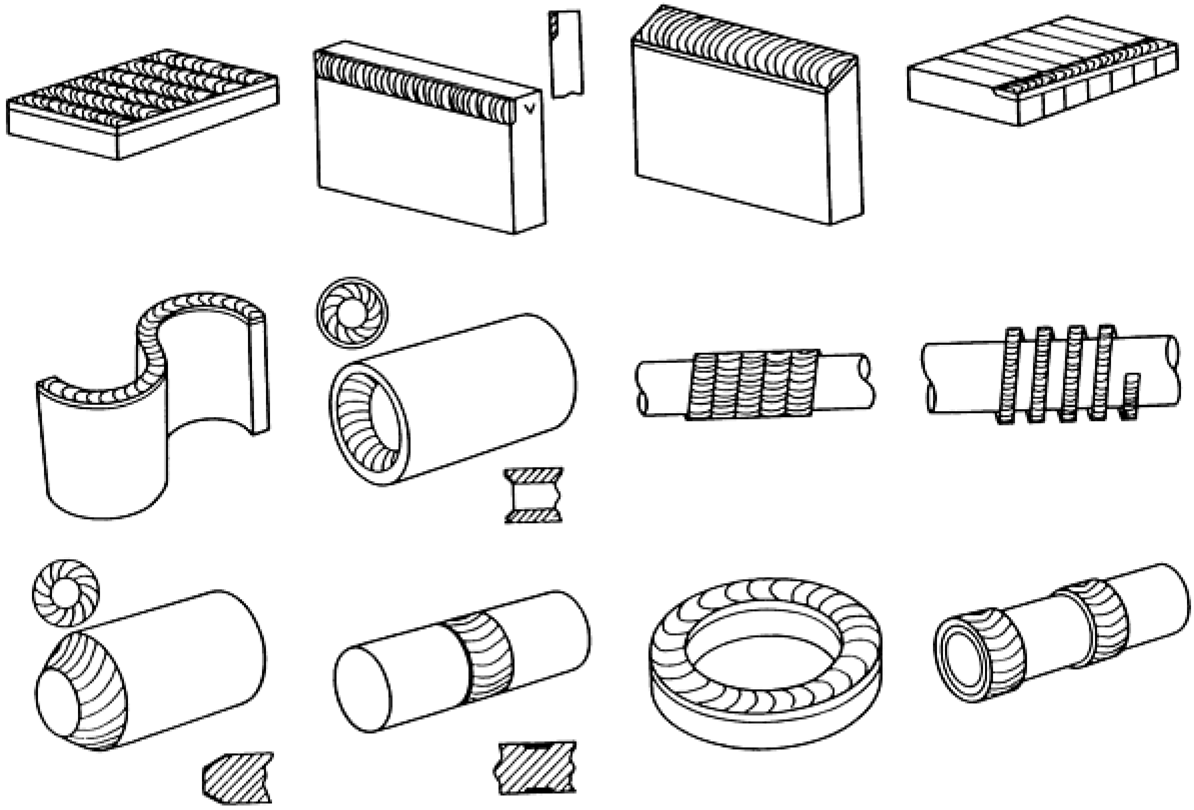

FIG. 4 GEOMETRIC ARRANGEMENTS FOR FRICTION SURFACING

Monolithic Structures. Friction surfacing has now been applied to a wide range of monolithic deposit/substrate combinations. There has been considerable success in producing high-strength bonding, fine-grained forged microstructures, and full homogeneity. A lack of porosity has also been attained, as have desirable surface properties, such as corrosion and wear characteristics that either equal or surpass the bar stock from which the deposit originated. Combinations include:

- INCO 625 TO AUSTENITIC STAINLESS STEEL
- AUSTENITIC STAINLESS STEEL TO CARBON STEEL
- STELLITE 6 AND 12 TO AUSTENITIC STAINLESS STEEL
- HASTELLOY TO AUSTENITIC STAINLESS STEEL
- TI-6AL-4V ALLOY TO TI-6AL-4V ALLOY
- AL-4CU ALLOY TO AL-4CU ALLOY
- HARDFACING STEELS, SUCH AS D2 AND H13, TO MEDIUM-CARBON STEELS

These and other applications are evidence that the process is coming of age and should be given serious consideration in the high-technology industries.

Metal-matrix composites. (MMC) offer considerable potential in high-temperature and wear-resistant applications. During 1988, friction surfacing with a MMC was accomplished using a consumable rod (matrix) that had a predrilled hole filled with alumina powder. Success was achieved, particularly when the matrix metal was 6063 aluminum alloy. It was equally possible to deposit a ferrous-based MMC (5 to 7 vol% of alumina in 4130 steel) from a proprietary MMC consumable rod produced by the Osprey process. The MMC deposited most easily onto a carbon steel, providing a hard, fine-grained martensitic matrix with uniformly distributed particulates of alumina. Further work was done to increase the volume fraction and distribution of the particulate in the matrix for the inclusion of zirconia and alumina by introducing changes in hole geometry and position in the consumable rod (Ref 17). An encouraging volume fraction of approximately 14% was obtained, although the particulate distribution needed to be improved.

References cited in this section

11. M.V. NOGIK *ET AL.*, FRICTION HARDFACING STEEL WITH STELLITE V3K, *SVAR. PROIZVOD.*, VOL 8, 1970, P 16-17
13. W.M. THOMAS, SOLID PHASE CLADDING BY FRICTION SURFACING, *PROC. TWI INTERNATIONAL SYMPOSIUM 18,* APRIL 1988, P 18
14. G.M. BEDFORD AND P.J. RICHARDS, "RECENT DEVELOPMENTS OF FRICTION COATING REPAIR AND RECLAMATION," THE ROYAL SOCIETY (LONDON, U.K.), 24-25 SEPT 1984
15. G.M. BEDFORD AND P.J. RICHARDS, "THE ABSENCE OF DILUTION IN FRICTION SURFACING AND LATERAL FRICTION WELDING," PAPER NO. 70, THE 1ST INTERNATIONAL CONFERENCE ON SURFACE ENGINEERING (BRIGHTON, U.K.), 25-28 JUNE 1985
16. E.D. NICHOLAS AND W.M. THOMAS, METAL DEPOSITION BY FRICTION WELDING, *WELD J.*, VOL 65 (NO. 8), AUG 1986, P 17-27
17. P.L. THREADGILL AND W.M. THOMAS, MANUFACTURE OF METAL MATRIX COMPOSITE CLAD LAYERS DURING FRICTION SURFACING: PRELIMINARY STUDIES, *PROC. EUROJOIN 1 CONF.* (STRASBOURG), NOV 1992, P 433-440

Friction Surfacing

E.D. Nicholas, The Welding Institute

References

1. H. KLOPSTOCK AND A.R. NEELANDS, "AN IMPROVED METHOD OF JOINING AND WELDING METALS," BRITISH PATENT SPECIFICATION 572789, APPLICATION DATE, OCT 1941
2. E. BISHOP, FRICTION WELDING IN THE SOVIET UNION, *WELD. MET. FABR.,* OCT 1960, P 408-410
3. KH.A. TYAYAR, FRICTION WELDING IN THE RECONDITIONING OF WORN COMPONENTS, *SVAR. PROIZVOD.,* VOL 1 (NO. 10), 1959, P 3-24
4. R.I. ZAKSON AND F.G. TURUKIN, FRICTION WELDING AND HARDFACING OF AGRICULTURAL MACHINE PARTS, *AVESTA SVARKA,* VOL 3, 1965, P 48-50
5. V.YA. KERSHENBAUM AND B.A. AVERBUKH, SPECIAL FEATURES IN THE FRICTION DEPOSITION OF BRONZE ON STEEL, *AVESTA SVARKA,* VOL 3, 1964, P 19-22
6. G.K. SCHREIBER, V.YA. KERSHENBAUM, AND S.G. TKACHENKO, INFLUENCE OF FRICTION SURFACING CONDITIONS ON TRANSITION LAYER PROPERTIES IN CLAD METALS, *SVAR, PROIZVOD.,* VOL 9, 1971, P 32-33
7. V.YA. KERSHENBAUM *ET AL.,* PERFORMANCE OF STEEL BRONZE BIMETAL PRODUCED BY FRICTION SURFACING, *SVAR. PROIZVOD.,* VOL 7, 1972, P 29-30
8. V.YA KERSHENBAUM AND B.A. AVERBUKH, OPTIMUM DIMENSIONS OF FRICTION SURFACED COMPONENTS, *AVESTA SVARKA,* VOL 5, 1972, P 64-66
9. B.A. AVERBUKH AND A.L. VELINSKII, THERMAL EFFECTS IN FRICTION SURFACING, *AVESTA SVARKA,* VOL 11, 1975, P 29-31
10. E.I. PLUZHIKOVA, E.L. DEMINA, V.YA. KERSHENBAUM, AND G.K. SCHREIBER, FRICTION FACING OF CAST IRON ON STEEL, TRANSLATED FROM *METALLOVED. TERM. OBRAB. MET.*, VOL 11, 1971, P 46-47
11. M.V. NOGIK *ET AL.*, FRICTION HARDFACING STEEL WITH STELLITE V3K, *SVAR. PROIZVOD.*, VOL 8, 1970, P 16-17
12. W.M. THOMAS, E.D. NICHOLAS, AND S.B. DUNKERTON, "FEASIBILITY STUDIES INTO SURFACING BY FRICTION WELDING," TWI RESEARCH REPORT 236, THE WELDING INSTITUTE, APRIL 1984

13. W.M. THOMAS, SOLID PHASE CLADDING BY FRICTION SURFACING, *PROC. TWI INTERNATIONAL SYMPOSIUM 18,* APRIL 1988, P 18
14. G.M. BEDFORD AND P.J. RICHARDS, "RECENT DEVELOPMENTS OF FRICTION COATING REPAIR AND RECLAMATION," THE ROYAL SOCIETY (LONDON, U.K.), 24-25 SEPT 1984
15. G.M. BEDFORD AND P.J. RICHARDS, "THE ABSENCE OF DILUTION IN FRICTION SURFACING AND LATERAL FRICTION WELDING," PAPER NO. 70, THE 1ST INTERNATIONAL CONFERENCE ON SURFACE ENGINEERING (BRIGHTON, U.K.), 25-28 JUNE 1985
16. E.D. NICHOLAS AND W.M. THOMAS, METAL DEPOSITION BY FRICTION WELDING, *WELD J.,* VOL 65 (NO. 8), AUG 1986, P 17-27
17. P.L. THREADGILL AND W.M. THOMAS, MANUFACTURE OF METAL MATRIX COMPOSITE CLAD LAYERS DURING FRICTION SURFACING: PRELIMINARY STUDIES, *PROC. EUROJOIN 1 CONF.* (STRASBOURG), NOV 1992, P 433-440

Ultrasonic Welding

Janet Devine, Sonobond Ultrasonics

Introduction

ULTRASONIC WELDING (USW) is a quasi-solid-state process that produces a weld by introducing high-frequency vibration to the weldment as it is held under moderately high clamping forces. The weld is produced without significant melting of the base materials.

In some respects, ultrasonic welding is an infant process that still awaits thorough exploration. A greater understanding is needed of the processes that occur at the bond interface. Specifically, the interaction of the process parameters, as well as their role in bond development, needs to be better understood.

The advantages of ultrasonic welding are that it:

- PERMITS JOINING OF THIN MATERIALS TO THICK MATERIALS
- PERMITS DISSIMILAR METAL JOINTS
- PROVIDES JOINTS WITH GOOD THERMAL AND ELECTRICAL CONDUCTIVITY
- JOINS METALS WITHOUT THE HEAT OF FUSION
- PROVIDES EFFICIENT ENERGY USE
- TYPICALLY REQUIRES NO FILLER MATERIAL, FLUX, OR SPECIAL ATMOSPHERE
- TYPICALLY REQUIRES NO SPECIAL CLEANING PROCESSES
- WELDS THROUGH MOST OXIDES

Applications. Commercially successful applications generally have certain characteristics. First, joints must be lap joints, not butt joints. Second, thin sections are required adjacent to the welding tip. Third, better results are obtained with nonferrous alloys.

Production applications include electrical wire harnesses for the appliance and automotive industry; buss bars; fuses; circuit breakers; contacts; ignition modules; starter motors; aluminum and copper foil; battery foils; capacitors; encapsulation of explosives; microelectronic wires; heating, ventilation, and air conditioning (HVAC) tubing; and many others. Military applications in the aircraft industry are described in MIL-STD 1947, issued May 15, 1985.

Ultrasonic Welding

Janet Devine, Sonobond Ultrasonics

Equipment and Process Parameters

An ultrasonic welding system requires a power supply that converts line power to the high frequency and high voltage needed by the transducer. The transducer transforms high-frequency electrical energy to vibratory energy and is incorporated into the welding head, which also provides the means (that is, either pneumatic, hydraulic, or mechanical) to clamp the workpieces. The transducer assembly also incorporates components or waveguides to transmit the energy to the desired weld area.

Process Mechanism. Ultrasonic welding products a weld by oscillating shear forces at the interface between the two metals being joined while they are held together under moderate static clamping force. The resulting internal stresses result in elastoplastic deformation at the interface.

Highly localized interfacial slip at the interface tends to break up oxides and surface films, permitting metal-to-metal contact at many points. As continued oscillation breaks down the points and the contact area grows, diffusion occurs across the interface to produce a structure similar to that of a diffusion weld.

Ultrasonic welding produces a localized temperature rise from the combined effects of elastic hysteresis, interfacial slip, and plastic deformation. The welding process is completed without having fully melted metal at the interface when the correct combination of force, power, and time parameters are used.

Interface temperature rise is greater for metals with low thermal conductivity (for example, steel) than it is for metals of high conductivity (for example, aluminum or copper). Ultrasonic welding of such high-conductivity materials consumes substantially less energy than does resistance welding.

In the case of alloys that have a broad melting temperature range, it is likely that as the low end of the range is reached, a slushing condition that facilitates plasticity in the weld interface is produced.

Process Variations and Limitations. Variations of the USW process produce different weld geometries. There are spot, line, continuous seam, and ring welding machines. Two other versions of spot welding machines are used to join microelectronic components.

Spot welds can be circular, elliptical, or rectangular, and solid or ringlike in geometry. They are formed when the material is clamped between a shaped tip (sometimes called a sonotrode) and an anvil (Fig. 1). The tip vibrates as ultrasonic energy is momentarily introduced in a plane parallel to the interface and perpendicular to the clamping force. Although weld time varies according to the thickness and composition of the material to be joined and the power of the welding machine, most spot welds can be produced in less than 1.5 s (Fig. 2).

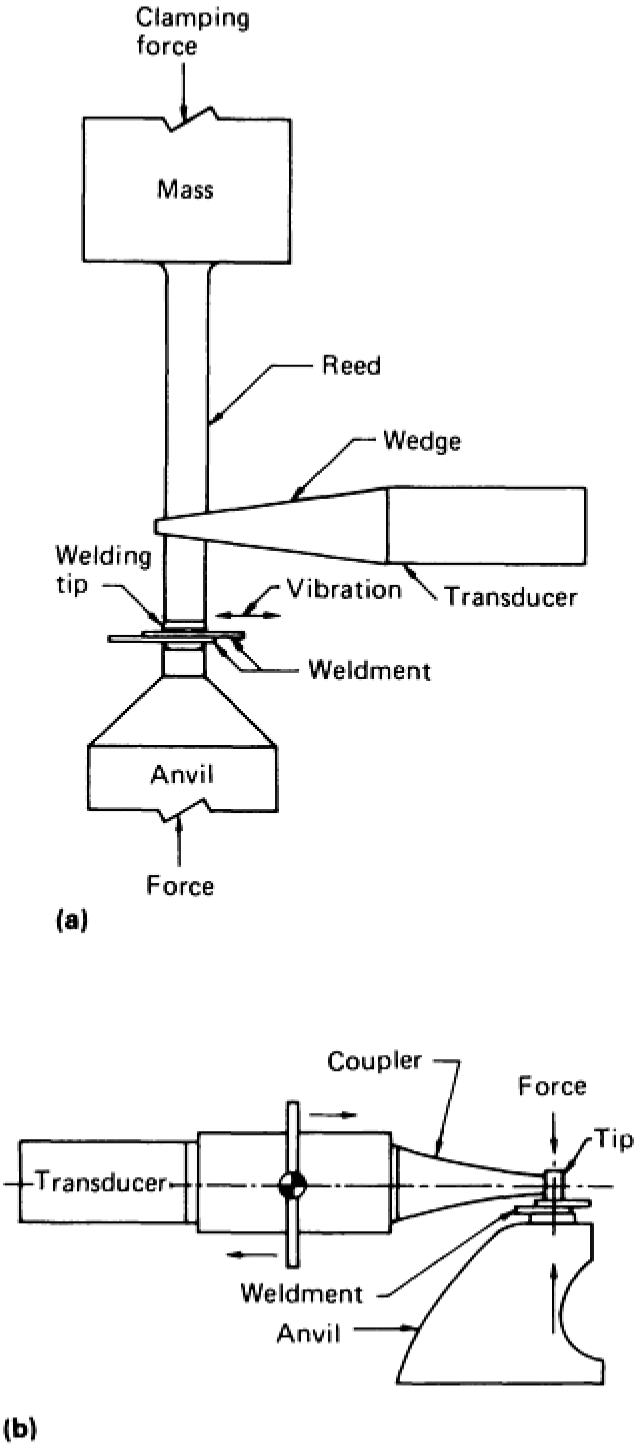
Clamping force
Mass
Reed
Wedge
Welding tip
Vibration
Transducer
Weldment
Anvil
Force
(a)
Coupler
Force
Tip
Transducer
Weldment
Anvil
(b)

FIG. 1 TWO VERSIONS OF AN ULTRASONIC WELDING SYSTEM USED FOR SPOT WELDING APPLICATIONS. (A) WEDGE-REED SYSTEM. (B) LATERAL DRIVE SYSTEM

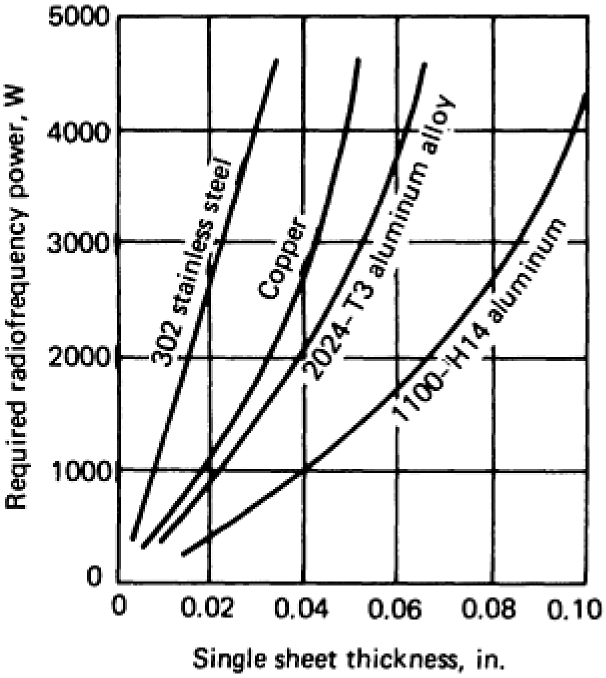

FIG. 2 PLOT OF WELDING MACHINE POWER CONSUMPTION VERSUS MATERIAL THICKNESS FOR SELECTED METALS JOINED BY ULTRASONIC WELDING

Welds can be made adjacent to or overlapping previous welds to form a continuous welded joint. A product clamp may be necessary to prevent the dispersion of ultrasonic energy into adjacent areas of the workpiece. The product clamp is usually concentric with the welding tip and has a slightly larger diameter than the tip.

Line welds are produced by a variation of spot welding in which the weld geometry is elongated by using a linear sonotrode tip and anvil. Custom multiple transducer heads have been used to produce line welds that are several inches long, but most commercially available equipment is limited to lines of 38 mm (1.5 in.) or less in length. Adjacent welds can produce a longer line. Typically, the longer welds are attainable only in thin materials (that is, less than 0.25 mm, or 0.010 in., thick). Single line welds up to 50 mm (2 in.) long have been made to join expanded nickel foil to solid foil for a lithium battery application.

Line welding is also used to seal copper tubes in HVAC applications (Fig. 3). This technique can replace seals normally produced by crimping and brazing.

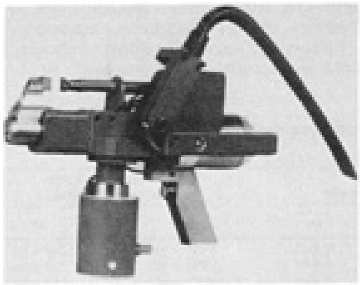

FIG. 3 TYPICAL ULTRASONIC LINE WELDER USED FOR METAL TUBE CLOSURE APPLICATIONS

Continuous seam welds are produced when a disk-shaped ultrasonically vibrating roller is rotated and traversed over a workpiece that is supported on a fixed anvil. Typical uses include joining foil ends in aluminum and copper foil mills. Commercial equipment is available to weld sheet thicknesses up to about 0.15 mm (0.006 in.). High-frequency systems (typically, 50 kHz) permit excellent welds in even the thinnest of foils, such as 0.0043 mm (0.00017 in.), without tearing or puckering. This technique is also used to join 0.038 mm (0.0015 in.) aluminum interconnects to foil in photovoltaic panels.

Ring Welds. A circular tip used on a spot welder can be used to form a ring weld. Systems designed especially for ring welding often use the torsional or circular motion of an annular-shaped tip instead of a forward and backward motion. Such a system utilizes two transducers, one of each side of a hollow reed. Each transducer produces motion 180° out of phase with the other, thus causing a torsional motion at the interface of the weldment.

Ring welds with diameters up to about 50 mm (2 in.) and an annular weld track of about 1.25 mm (0.050 in.) have been produced in thin aluminum or copper foils. Typical applications include the encapsulation of liquid and powder propellant or explosive materials by welding a thin foil cover on a container. Foils are usually 0.2 mm (0.008 in.) or less. The weld process does not produce much heat, which makes it suitable for use with heat-sensitive materials. Many small, high-frequency (28 kHz) systems are in operation to hermetically seal small explosive initiators or fuses for armaments.

Microelectronic Welds. fine wire bonding represents the earliest widely used USW application and still accounts for a large volume of industrial activity. Millions of wire bonds are performed daily. Figure 4 shows an ultrasonic wire welding apparatus.

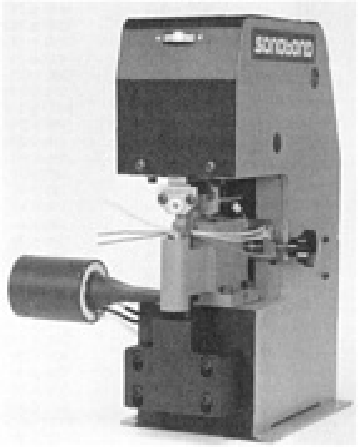

FIG. 4 ULTRASONIC WIRE WELDER USED TO BOND WIRES FOR MICROELECTRONIC APPLICATIONS

Wire diameters range from less than 0.025 to 0.5 mm (0.001 to 0.020 in.), and the highest volume occurs in the 0.025 to 0.050 mm (0.001 to 0.002 in.) diameter range. Vibratory action at high frequency (typically, 60 kHz) removes surface contaminants, induces material flow, and permits a solid-state weld between the wire and either the metallized bond pad or the leads on the semiconductor package.

A combination of ultrasonic and thermocompression bonding techniques, known as thermosonic bonding, is now a popular wire bonding method. The technique involves ultrasonic welding with heated substrates, typically with interface temperatures that range from 100 to 200 °C (210 to 390 °F).

Ultrasonic Welding

Janet Devine, Sonobond Ultrasonics

Personnel

Operators of the USW process do not need elaborate or extensive training. Once the process parameters have been determined by a process engineer, the operator is required only to load the parts into a nest of supporting/locating anvil assembly, to press the palm buttons or other starting device, and then to unload the finished part. In-process monitoring can be performed either electronically, using a microprocessor, or visually by the operator, which usually requires the operator to look at a power indication gage. The process engineer needs to work closely with the equipment manufacturer for guidance regarding tooling design and the selection of process parameters.

No arc, spark, or molten filler material is associated with ultrasonic welding, and no electrical current passes through the weldment. Because the welder is usually configured in a press form with moderately high forces, the normal precautions need to be observed to ensure operator safety, that is, the use of anti-tie-down palm buttons or a similar provision to protect the hands of the operator. Another common consideration involves the high-frequency noise level produced by the ultrasonic vibration. In some instances, especially with the higher power level equipment, the noise exceeds the Occupational Safety and Health Administration (OSHA) approved levels and sound-deadening barriers or enclosures must be provided. Even where the noise level meets the approved requirements, the operator may find it necessary to wear ear protection.

Ultrasonic Welding

Janet Devine, Sonobond Ultrasonics

USW Procedures

The USW process requires the overlapping of the materials to be welded. Generally, the materials need only be presented to the welder in proper orientation. Correct orientation is usually achieved by using a nest or anvil fixture, which supports the parts while they are being welded. When joining stranded wires to other solid or stranded wires or to a terminal, a "gathering" fixture must be used to pull the wires together and to exert a slight pressure while welding to prevent the wires from escaping from the intended weld area. This type of fixture is usually supplied by the manufacturer with machines intended for use with wires and can be adjusted to accommodate a wide range of sizes and combinations.

Special considerations described below include the condition of the surface, the use of an interlayer, and the control of resonance.

Surface Condition. Most of the readily weldable materials, such as aluminum, copper, or brass, can be welded as received from the mill or must be degreased with a common solvent or detergent to remove surface lubricants. Oxide coatings will disperse during the process, unless they are very thick. Heavy surface scale should be removed by mechanical abrading or chemical etching before welding. The time lapse between cleaning and welding is generally not critical, unless the atmosphere is corrosive. Some types of coatings and insulations (for example, low-temperature magnet-wire coating) may be penetrated during the welding process, whereas other types must be mechanically removed. Fairly consistent surface cleanliness and quality must be maintained to ensure uniform weld quality.

Use of an Interlayer. A useful technique for improving the weld quality of some weldments involves placing a thin foil, usually aluminum or copper, between the metals to be bonded. This is particularly useful when materials of varying hardness are to be bonded. This interlayer is sometime more convenient and cost effective than plating the materials with a more weldable material (for example, copper or gold).

In a technique known as weld bonding, a layer of adhesive is placed between the panels to be ultrasonically welded. This technique not only provides a watertight seal, but also increases the weld strength beyond that obtained by either adhesive bonding or ultrasonic welding alone. Either a paste adhesive or a fabric-supported adhesive can be used in the ultrasonic bonding process.

Control of Resonance. Complex workpieces, especially those with multiple parts or thin wall sections, may be induced to vibrate by the ultrasonic welding system, which can produce fractures in the workpiece itself and in previously made welds or can cause inconsistent weld quality. This resonance can be either eliminated or minimized by applying pressure to the vibrating section. For example, in the ultrasonic welding of aluminum foil layers to the studs of capacitor caps, the stud must be clamped tightly or else the vibration will not only prevent the formation of a good joint, but can even melt the plastic cap that surrounds the stud.

Another option is to clamp the vibrating part to a comparatively large fixture or anvil. Significant pressure is required, and some machines come equipped with a product clamp for this purpose. Because resonance in the tooling can also occur, fixtures should be rugged and should not contain multiple small devices, such as springs or pins. It is best to avoid using light materials (for example, aluminum) for part fixtures, anvils, or supports. Steel is the preferred material for these components.

Tooling, Tips, and Anvils. The welding tip (or sonotrode) that contacts the weldment is usually made of the high-quality heat-treated tool steel. A precision fit between the tip and the waveguides is necessary to ensure efficient transmission of the vibratory energy. A locking (Morse) taper is frequently used, and the fit should cover 75% of the contacting surface area between the tip and its matching receptacle. In lower-power systems, the tip and the waveguide (horn) can be integral and can sometimes have several surfaces for welding. Rotation of the horn provides a new welding surface. A welding tip with a taper lock fit is less expensive to replace and easier to resurface, when necessary, than an integral horn-tip combination.

Certain alloys, especially the very soft aluminum alloys, may stick to the tip when welded. A mechanical stripper may be needed to pull the part free. Sometimes, a low-power, ultrasonic pulse may be sufficient to remove the stuck part from the tip. If a nugget remains on the tip, a weld pulse, with the tip clamped against a thick piece of brass, can remove the sticking nugget easier than mechanical abrasion methods.

Tips composed of exotic alloys have been used to prevent particularly tenacious sticking conditions, but with limited success. A steel shim with an oxidized surface was found to be particularly effective in preventing both sticking and deformation when bonding high-strength aluminum and titanium alloys, which are extensively used by the aircraft industry.

The anvil tip is subject to the same problems of wear and tip sticking that plague the sonotrode tip. The same high-strength heat-treated tool steel (typically, M2 heat treated to 58 to 60 HRC hardness) is recommended.

Welding tip and anvil tip surfaces with serrated or crosshatched patterns are useful in preventing slip between the tip and the weldment. Slip can result in a weld location between the metal and the tools, instead of at the required interface. A typical crosshatched pattern would be 0.5 mm (0.02 in.) peak-to-peak and about 0.2 mm (0.008 in.) deep.

Special Atmospheres. Although ultrasonic welding does not require a special atmosphere, it may be applicable under certain conditions. For example, use of an inert gas can reduce or prevent oxidation when bonding a metal such as lithium. Ultrasonic welding is not adversely affected by the presence of an inert atmosphere.

Weld quality is described below in terms of influencing factors, surface appearance and deformation, and metallographic examination.

Influencing Factors. The quality of ultrasonic welds is affected by these parameters:

- COMPOSITION AND GEOMETRY OF THE WELDMENT
- HARDNESS OF THE WORKPIECE
- CLEANLINESS OF THE WELDMENT
- SELECTION OF WELDING CONDITIONS, SUCH AS POWER, CLAMPING FORCE, AND WELD TIME
- CAPACITY OF THE TOOLING TO PROPERLY SUPPORT AND CLAMP THE PARTS TO PREVENT UNWANTED VIBRATION

Surfaces to be welded should be reasonably flat the parallel. This is especially critical for ring welding where a high degree of hermeticity is required. Some materials may be weldable in the as-received condition. However, a change in lubricants or other surface condition can require an adjustment in machine settings to maintain quality. Therefore, it is sometimes advisable to degrease or to abrade surfaces before welding to maintain a certain level of consistency.

Surface Appearance and Deformation. Depending on the material and the tip geometry, the surface of an ultrasonic weld can leave a slight scuff mark or a significant depression. This thickness deformation is more visible in soft, ductile materials, such as soft aluminum. The actual weld interface is usually smaller than the surface impression. Harder materials generally have a shallower and smaller surface depression than soft, ductile materials.

A tip surface that has serrations or a cross-hatched pattern will replicate this pattern in the surface of the weldment. A spherical radius on the tip will generally produce a deeper, bowl-shaped depression than a flat tip of the same diameter.

Stranded or braided wires can be welded to form a solid cross section, if required. Slightly lower power, time, or force can give a compressed, but not solid, cross section.

Judicious radiusing and angling of tools is recommended to avoid sharp transitions in areas that can lead to early failure of an assembly.

The metallographic examination of ultrasonic welds in a wide variety of metals reveals phenomena that occur in the microstructure, such as surface film and oxide disruption, plastic flow and extrusion, recrystallization, phase transformation, and diffusion. Photomicrographs of selected ultrasonically welded materials are shown in the Section "Procedure Development and Practice Considerations for Solid-State Welding" in this Volume.

A heat-affected zone is significant in certain alloys, such as aluminum and nickel. Phase transformation, recrystallization, and precipitation may occur.

Diffusion across the interface is usually shallow, because of the relatively short weld times required, although significant penetration across the interface can take place. Alloying may occur then welding certain dissimilar metals, and the possibility of galvanic corrosion should be considered.

Janet Devine, Sonobond Ultrasonics

Mechanical Properties

Tensile shear tests conducted on single-weld lap joints indicated that failure was caused by a fracture in the base metal itself or by a tear out of the weld button, rather than shear in the actual weld. Typical spot weld strengths in a variety of metals are summarized in Tables 1 and 2. These data were published as the result of a government-sponsored research program undertaken in the 1970s. With more modern equipment and practices, improved results can be obtained.

TABLE 1 TYPICAL SHEAR STRENGTHS OF ULTRASONIC SPOT-TYPE WELDS IN SELECTED PURE METALS AND ALLOYS

METAL	ALLOY OR PRODUCT TYPE	SHEET THICKNESS		MEAN STRENGTH WITH 90% CONFIDENCE INTERVAL[(A)]	
		mm	in.	n	lbf
ALUMINUM	X2020-T6	1.02	0.040	5520 ± 220	1240 ± 50
COPPER	ELECTROLYTIC	1.02	0.040	3470 ± 90	780 ± 20
		1.14	0.045	3780 ± 90	850 ± 20
BRASS	70-30 (HARD)	0.74	0.029	2400 ± 135	540 ± 30
STEEL	1020	0.64	0.025	2220 ± 90	500 ± 20
	AISI TYPE 301	0.20	0.008	2220 ± 45	500 ± 10
		0.36	0.014	3910 ± 180	880 ± 40
		0.41	0.016	6000 ± 710	1350 ± 160
	AISI TYPE 302	0.33	0.013	1820 ± 90	410 ± 20
		0.46	0.018	2800 ± 135	630 ± 30
		0.64	0.025	4630 ± 310	1040 ± 70

		0.81	0.032	5780 ± 900	1300 ± 200
	AISI TYPE 316	0.30	0.012	1020 ± 180	230 ± 40
		0.46	0.018	1960 ± 220	440 ± 50
	17-7 PH	0.20	0.008	1420 ± 90	320 ± 20
		0.51	0.020	440 ± 440	990 ± 100
		1.02	0.040	6230 ± 440	1400 ± 100
	PH 15-7 MO	0.20	0.008	1200 ± 45	270 ± 10
	AM350	0.20	0.008	1380 ± 90	310 ± 20
	AM335	0.20	0.008	1690 ± 180	380 ± 40
	A-286	0.38	0.015	3020 ± 310	680 ± 70
MOLYBDENUM	ARC-CAST	0.38	0.015	1070 ± 180	240 ± 40
		0.51	0.020	1470 ± 180	330 ± 40
		0.64	0.025	1600 ± 180	360 ± 40
	SINTERED	0.38	0.015	1510 ± 310	340 ± 70
		0.51	0.020	1690 ± 335	380 ± 80
		0.64	0.025	1910 ± 220	430 ± 50
TANTALUM	. . .	0.25	0.010	1110 ± 135	250 ± 30
TITANIUM ALLOY	TI-5AL-2.5SN	0.71	0.028	8670 ± 535	1950 ± 120
	TI-8MN	0.81	0.032	7690 ± 890	1730 ± 200
	TI-6AL-4V	1.02	0.040	10005 ± 800	2260 ± 180
ZIRCALOY-2	. . .	0.51	0.020	2760 ± 90	620 ± 20
NICKEL ALLOY	J-1500	0.33	0.013	3690 ± 735	830 ± 30
	INCONEL	0.25	0.010	755 ± 310	170 ± 70
		0.51	0.020	980 ± 445	220 ± 100
	INCONEL X	0.81	0.032	6760 ± 445	1520 ± 100
	K-MONEL	0.51	0.020	3025 ± 355	680 ± 80
		0.81	0.032	4000 ± 645	900 ± 60

(A) 90% CONFIDENCE INTERVAL INDICATES STATISTICAL PROBABILITY THAT STRENGTHS OF 90% OF THE WELDS WILL FALL WITHIN THE RANGE INDICATED.

TABLE 2 TENSION-TO-SHEAR RATIOS FOR SELECTED ULTRASONICALLY WELDED MATERIALS

METAL	TYPE	CONDITION	THICKNESS		CROSS-TENSION STRENGTH		TENSILE-SHEAR STRENGTH		TENSION-TO-SHEAR RATIO
			mm	**in.**	**n**	**lbf**	**n**	**lbf**	
ALUMINUM	2014-T6	. . .	0.81	0.032	1025	230	3115	700	0.33
			1.02	0.040	1380	310	3290	740	0.42
			1.27	0.050	1245	280	4005	900	0.31
	2024-T3	BARE	1.02	0.040	1070	240	3780	850	0.28
	2024-T3	ALUMINUM CLAD	1.27	0.050	1200	270	4715	1090	0.25
	7075-T6	ALUMINUM CLAD	1.27	0.050	1110	250	4890	1100	0.23
TITANIUM	TI-6AL-4V	. . .	0.64	0.025	1155	260	5425	1220	0.21
	TI-8MN	. . .	0.81	0.032	1955	440	7785	1750	0.25
MOLYBDENUM	SINTERED	. . .	0.13-0.25	0.005-0.010	110	25	425	95	0.26
TANTALUM	SINTERED	. . .	0.13-0.25	0.005-0.010	135	30	380	85	0.35
STAINLESS STEEL	17-7 PH	ANNEALED	0.51	0.020	845	190	2400	540	0.35
		HEAT-	0.30	0.012	490	110	1600	360	0.31

		TREATED							
		HEAT-TREATED AND AGED	0.13	0.005	290	65	690	155	0.42

Consistency is generally good, with one standard deviation, σ, typically less than 5% of the average strength value. Consistency of results in the manufacturing environment will follow from control of vital parameters (for example, part temperature, tooling temperature, machine and tooling stability, and surface condition of the parts).

Welds tested after thermal cycling, exposure to salt baths, and other corrosive environments maintain a relatively high tensile strength level.

Fatigue strength of ultrasonically welded metals often exceeds that of fusion-welded metals, because the ultrasonic bonding does not leave the cast button structure that is typical of melted and resolidified metal.

For relatively thin foils, the ductility of the ultrasonic bond permits the reforming of parts after welding without cracking the welded joint.

Ultrasonic Welding

Janet Devine, Sonobond Ultrasonics

Selected References

- J. DEVINE, ULTRASONIC BONDING FOR MOTOR MANUFACTURE, *PROC. ICWA CONF.* (BOSTON), 1979, P 1-5
- J. DEVINE, ULTRASONIC WELDING, *WELDING HANDBOOK,* 8TH ED., VOL 2, CHAPTER 25, 1991, P 784-812
- J. DEVINE, JOINING ELECTRICAL CONTACTS?, *WELD. DES. FABR.,* VOL 53 (NO. 3), 1980, P 112-115
- J. DEVINE, ULTRASONIC WELDING HELPS LIGHTEN AIRCRAFT, *WELD. DES. FABR.,* VOL 51 (NO. 8), 1978, P 74-76
- J. DEVINE AND R.G. VOLLMER, ULTRASONIC BONDING ARRIVES, *U.S. ARMY MAN. TECH. J.,* VOL 3 (NO. 1), 1978, P 11-14
- J. DEVINE, G.K. DINGLE, AND R.G. VOLLMER, ULTRASONIC BONDING, PANACEA OR PIE IN THE SKY, *PROC. ULTRASONIC INDUSTRIES ASSOCIATION* (NEW YORK), 1977, P 1-15
- T.J. KELLY, ULTRASONIC WELDING OF CUNI TO STEEL, *WELD. J.,* VOL 60 (NO. 4), 1981, P 29-31
- F.R. MEYER, ULTRASONIC WELDING PROCESS FOR DETONATABLE MATERIALS, *NATL. DEF.,* VOL 70 (NO. 334), 1976, P 291-293
- F.R. MEYER, ULTRASONICS PRODUCES STRONG OXIDE-FREE WELDS, *ASS. ENG.,* VOL 20 (NO. 5), 1977, P 26-29
- T. RENSHAW, J. CURATOLA, AND A. SARRANTONIO, DEVELOPMENTS IN ULTRASONIC WELDING FOR AIRCRAFT, *PROC. 11TH NATL. SAMPE TECHNICAL CONF.* (BOSTON). 1979, P 681-693
- T. RENSHAW AND A. SARRANTONIO, "PROPERTIES OF LARGE MULTISPOT ULTRASONICALLY WELDED JOINTS," AIAA STRUCTURES AND MATERIALS CONFERENCE (SEATTLE), 12-14 MAY 1980
- T. RENSHAW, K. WONGWIWAT, AND A. SARRANTONIO, A COMPARISON OF PROPERTIES OF SINGLE OVERLAP TENSION JOINTS PREPARED BY ULTRASONIC WELDING AND OTHER MEANS, *PROC. AIAA/ASME/ASCE 23RD CONF.* (NEW ORLEANS), 1983, P 1-8

- T. RENSHAW, AIRCRAFT SERVICE TESTING OF ULTRASONICALLY WELDED PANELS, *PROC. 16TH NATIONAL SAMPE TECHNICAL CONF.* (ALBUQUERQUE), 1984
- "ULTRASONIC WELDING OF ALUMINUM AND ALUMINUM ALLOY MATERIALS," MIL-STD 1947, 15 MAY 1985
- WELDING ALUMINUM THEORY AND PRACTICE, *ULTRASONIC WELDING,* THE ALUMINUM ASSOCIATION, 1989, P 14.3, 14.4

Torch Brazing

Charles E. Fuerstenau, Lucas-Milhaupt, Inc.

Introduction

TORCH BRAZING (TB) utilizes a fuel gas flame as the heat source for the brazing process. The fuel gas is mixed with either air or oxygen to produce a flame, which is applied to the workpiece until the assembly reaches the proper brazing temperature. Then, preplaced filler metal will be melted or hand-fed wire can be introduced.

Torch Brazing

Charles E. Fuerstenau, Lucas-Milhaupt, Inc.

Advantages and Limitations

Torch brazing is used with various base metals and on many different sizes of assemblies. The process offers many advantages, including:

- FLEXIBILITY, IN THAT ONE TORCH WITH MULTIPLE TIPS CAN BE USED TO BRAZE A VARIETY OF ASSEMBLIES
- LOW CAPITAL EQUIPMENT COST (MANUAL TORCH BRAZING)
- ENTIRE ASSEMBLY DOES NOT HAVE TO BE HEATED; SMALL JOINTS ON LARGE ASSEMBLIES CAN BE HEATED LOCALLY
- AUTOMATION IS POSSIBLE IN MANY CASES
- MOST BASE METALS AND COMBINATIONS OF BASE METALS CAN BE TORCH BRAZED IF A SUITABLE FLUX IS AVAILABLE

Although the process provides versatile, low-cost heating for brazing, its limitations include:

- OXIDATION/DISCOLORATION CAN OCCUR ON SURFACES OF THE ASSEMBLY NOT COVERED WITH FLUX, BECAUSE PROCESS IS CONDUCTED IN AIR
- FLUX RESIDUES NEED TO BE REMOVED AFTER BRAZING
- HIGHLY REACTIVE MATERIALS, SUCH AS TITANIUM AND ZIRCONIUM, CANNOT BE TORCH BRAZED, BECAUSE NO FLUX IS AVAILABLE
- LARGE ASSEMBLIES CAN BE DIFFICULT TO HEAT, BECAUSE OF THE LOCALIZED NATURE OF FLAME HEATING

Torch Brazing

Charles E. Fuerstenau, Lucas-Milhaupt, Inc.

Applications

Torch brazing is commonly used on copper, brass, and other copper alloys, as well as steel, stainless steel, aluminum, carbides, and various heat-resistant materials. Most combinations of these materials can also be torch brazed. It is necessary to use flux with these materials, except when a phosphorus brazing alloy is used to braze pure copper parts. In this case, the phosphorus acts as the flux. The low-temperature silver-base and silver/copper/phosphorus filler metals are commonly used with torch brazing. Various other copper-base and gold-base filler metals can also be used with this process.

Torch brazing is often used to join copper and Bundy steel tube assemblies for the heating, air conditioning, and refrigeration industries. The process is also commonly used when brazing heat exchangers, bicycles, furniture, carbide tools, plumbing components, automotive subassemblies, medical instruments, and many other workpiece types. A wide range of components can be torch brazed, including small joints for jewelry parts, large-diameter (75 mm, or 5 in.) tubes, and fitting joints. The process provides strong, leak-tight joints on a wide variety of base materials.

Torch Brazing

Charles E. Fuerstenau, Lucas-Milhaupt, Inc.

Equipment

The equipment for torch brazing has several components. Single torches are typically used for hand brazing, whereas multiple torches can be used in an automated system. The gas-oxygen is mixed in the torch body and is adjusted using the needle valves on the torch. Gas-air combinations can be mixed at the torch or, alternatively, a central mixing system can be used to supply many torches, particularly in automated applications.

The typical fuel gases used in torch brazing are acetylene, propane, and methane (natural gas). Various flame temperatures and heat contents are given in Table 1.

TABLE 1 COMMON FUEL GASES USED IN TORCH BRAZING

FUEL GAS	FORMULA	OXYGEN-TO-FUEL GAS RATIO$^{(A)}$	FLAME TEMPERATURE FOR OXYGEN		HEAT CONTENT	
			°c	°f	mj/m^3	btu/ft^3
ACETYLENE	C_2H_2	2.5	3087	5589	55	1470
PROPANE	C_3H_8	5.0	2526	4579	104	2498
METHANE	CH_4	2.0	2538	4600	37	1000

(A) THE VOLUME UNITS OF OXYGEN REQUIRED TO COMPLETELY BURN A UNIT VOLUME OF FUEL GAS; A PORTION OF THE OXYGEN IS OBTAINED FROM THE ATMOSPHERE.

An automated brazing system that can braze up to 40,000 heat-exchanger return bend joints per hour is shown in Fig. 1. Multiple torches are used to provide the necessary heat.

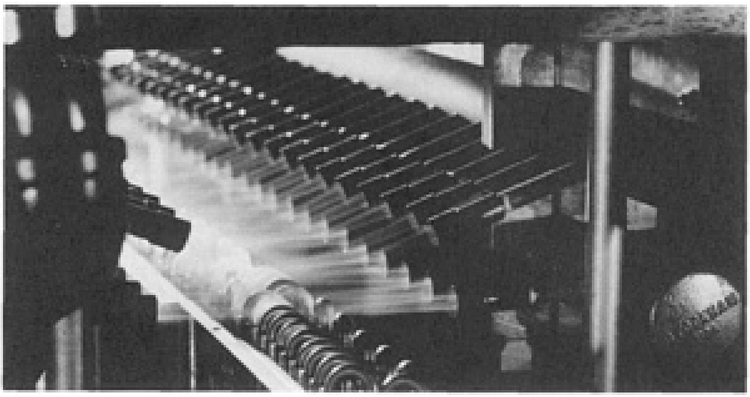

FIG. 1 AUTOMATED TORCH BRAZING SYSTEM

Torch brazing gases are normally supplied from bulk sources. An individual gas cylinder may supply one torch, whereas large bulk tanks are used to supply many torches or an automated system utilizing many burners. In all torch brazing systems, regulators are used to safely control the gas distribution. Individual regulators are used on the gas and oxygen lines. Standard safety precautions should be utilized when handling the equipment, and operators should be thoroughly trained in safety practices related to the use of compressed and combustible gases.

Techniques

The key to torch brazing, as in all methods of heating, is to heat evenly the components to be brazed. This may require directing the torch primarily on a large, rather than small, component. Because the filler metal will flow to the hottest portion of an assembly, the flame should be applied in such a way as to flow the filler metal in the required direction.

When preplaced (preform or paste) filler metal is used in torch brazing, the operator need only know where and how much heat to apply. When the operator is also responsible for the filler metal, he must not apply it until the assembly has reached brazing temperatures. This may require practice and experience. Torch brazing is a skill that can be attained relatively quickly with the proper training.

Selected References

- *BRAZING HANDBOOK,* 4TH ED., AMERICAN WELDING SOCIETY, 1991
- *THE BRAZING BOOK,* HANDY & HARMAN/LUCAS-MILHAUPT, INC., 1991

Furnace Brazing

Toshi Oyama and Howard Mizuhara, WESGO, Inc.

Introduction

FURNACE BRAZING is a mass production process for joining the components of small assemblies with a metallurgical bond, using a nonferrous filler metal as the bonding material and a furnace as the heat source (Fig. 1). Furnace brazing technology was initiated in the 1920s and was first used commercially circa 1930, primarily to provide a brazing process that did not require a chemical flux, thereby eliminating the flux entrapment problem. Currently, furnace brazing is widely applied in a variety of industries. The essentially automatic nature of the process and its use of unskilled labor account for its popularity.

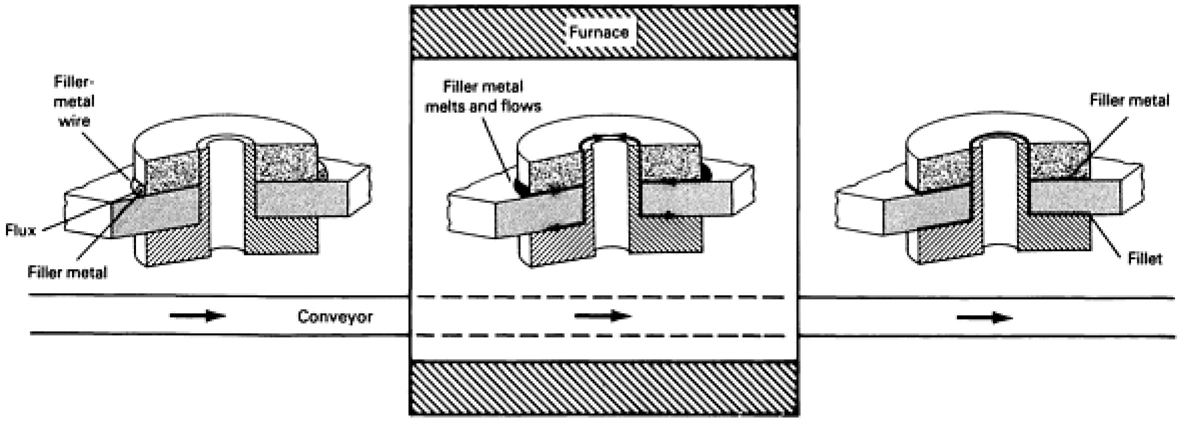

FIG. 1 CROSS-SECTIONAL VIEW OF THREE-PART ASSEMBLY AS IT PROCEEDS THROUGH FURNACE BRAZING OPERATION. FILLER-METAL WIRE (FILLER METAL ENCIRCLED BY FLUX) GRADUALLY MELTS AND FLOWS INTO VOIDS BETWEEN THE THREE SECTIONS AND THEN COOLS TO FORM A SINGLE-UNIT WORKPIECE AS IT IS CONVEYED THROUGH THE FURNACE.

In the furnace brazing process, cleaned parts and brazing filler metal are assembled, placed in a furnace, and heated to the brazing temperature. Furnace brazing is only practical if the filler metal can be placed on the joint before brazing and retained in position during brazing. Heating rate, brazing temperature and time, and cooling rate are controlled. Furthermore, in most cases, the brazing is carried out in a controlled atmosphere, which may be either reducing, inert, or vacuum. The advantages and limitations of furnace brazing are identified below.

Process Advantages. There are at least five advantages to using the furnace brazing process:

- IT IS A METHOD IN WHICH MANY VARIABLES CAN BE EASILY CONTROLLED TO ENSURE REPEATABILITY OF THE PROCESS AND GUARANTEE A HIGH-QUALITY JOINT. BRAZING TEMPERATURE AND PROCESS DURATION, AS WELL AS HEATING AND COOLING RATES, CAN BE CONTROLLED AND MONITORED. IN ADDITION, THE BRAZING ATMOSPHERE CAN BE EASILY REGULATED.
- WHEN ALL THE BRAZING PROCEDURES ARE WELL ESTABLISHED, RELATIVELY UNSKILLED OPERATORS CAN CARRY OUT EVERYDAY MANUFACTURING OPERATIONS.
- THE ABSENCE OF FLUX ENTRAPMENT ALLOWS MORE FLEXIBILITY IN JOINT DESIGN. THE POSTBRAZE CLEANING OF THE BRAZED PART AND THE FURNACE ARE UNNECESSARY.
- BECAUSE A WHOLE ASSEMBLY IS HEATED, DISTORTION OF THE PARTS CAN BE MINIMIZED OR ELIMINATED. IN SOME CASES, HEAT TREATMENT OF THE PART CAN BE INCORPORATED INTO THE BRAZING CYCLE.
- MORE THAN ONE JOINT PER WORKPIECE CAN BE BRAZED IN A BRAZING CYCLE. SEVERAL DIFFERENT ASSEMBLIES REQUIRING THE SAME BRAZING CONDITIONS CAN ALSO BE BRAZED SIMULTANEOUSLY.

Process limitations include:

- THE INITIAL INVESTMENT IN FURNACE BRAZING EQUIPMENT IS RELATIVELY HIGH. FURTHERMORE, THE MAINTENANCE OF THE EQUIPMENT IS CRITICAL AND CAN BE MORE EXPENSIVE THAN OTHER BRAZING TECHNIQUES. THE COST OF A SPECIAL ATMOSPHERE MUST ALSO BE CONSIDERED.
- BECAUSE THE WHOLE ASSEMBLY IS HEATED, THE COST OF HEATING EXCEEDS THAT OF OTHER BRAZING OPERATIONS. IN ADDITION, MICROSTRUCTURAL CHANGES IN THE BASE MATERIAL CAN OCCUR, DEPENDING ON THE MATERIAL BEING BRAZED AND THE BRAZING TEMPERATURE.
- BECAUSE BRAZING IS PERFORMED INSIDE THE FURNACE, JOINT DESIGN IS IMPORTANT. THE FIXTURES FOR HOLDING A BRAZED ASSEMBLY REQUIRE EXTRA HEAT MASS, WHICH INCREASES THE MANUFACTURING COST. PROPER JOINT DESIGN CAN MINIMIZE OR EVEN ELIMINATE FIXTURE COMPLEXITY.

Furnace Brazing

Toshi Oyama and Howard Mizuhara, WESGO, Inc.

Furnaces and Control Instrumentation

Brazing furnaces can be either oxyfuel fire or electrically heated. Whether they furnish heat directly or indirectly, all furnaces must provide a uniform work-load temperature. For example, the brazing of high-temperature nickel-base alloys at 1095 °C (2000 °F) and higher generally requires a hot-zone uniformity of ±8 °C (±15 °F). When brazing aluminum alloys, the temperature uniformity must be ±3 °C (±5 °F). Thus it is vital that each work load be closely monitored with a minimum of two thermocouples attached to or embedded in the parts to be brazed.

There are basically three types of furnaces: continuous, semicontinuous, and batch. Proper furnace selection primarily depends on the application. An example of the continuous-type furnace is a belt-conveyor furnace (Fig. 2). The brazing conditions are controlled by the speed of the conveyor and the set temperature. Because initial start-up involves several days to heat and to prepare the furnace for operation, these parameters are constantly maintained. The furnaces are seldom shut down and, when idling, are usually run at a temperature that is approximately several hundred degrees below the operating temperature. Therefore, the cost of furnace operation is very high if the equipment is not fully utilized. The furnace is used for mass production applications (that is, production rates of a few thousand parts per hour). Typically, a reducing gas is introduced into the brazing chamber. When the furnace is fully utilized, the cost per unit part becomes cost effective.

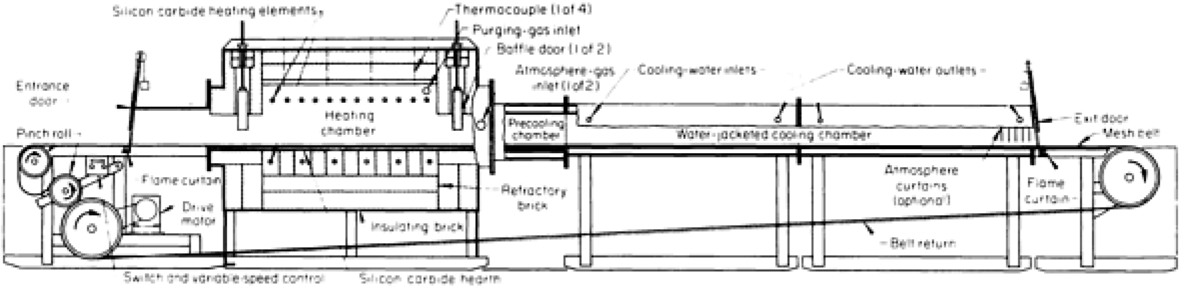
Silicon carbide heating elements
Thermocouple (1 of 4)
Purging-gas inlet
Baffle door (1 of 2)
Atmosphere-gas inlet (1 of 2)
Cooling-water inlets
Cooling-water outlets
Entrance door
Heating chamber
Precooling chamber
Water-jacketed cooling chamber
Exit door
Mesh belt
Pinch roll
Flame curtain
Drive motor
Refractory brick
Insulating brick
Atmosphere curtains (optional)
Flame curtain
Belt return
Switch and variable-speed control
Silicon carbide hearth

FIG. 2 MESH-BELT CONVEYOR BRAZING FURNACE THAT UTILIZES A WATER-JACKETED COOLING CHAMBER

Retort (or bell-type) furnaces are often used as semicontinuous furnaces (Fig. 3). In this type of furnace, brazing assemblies are encapsulated in the retort and a special gas is introduced. Then, the retort is placed in a furnace for brazing.

In the semicontinuous operation, two or more retorts are used in the production process. While one retort is in a brazing cycle, another retort is cooled, unloaded, and reloaded. Such furnace systems are typically equipped with microprocessors to control the furnace cycle.

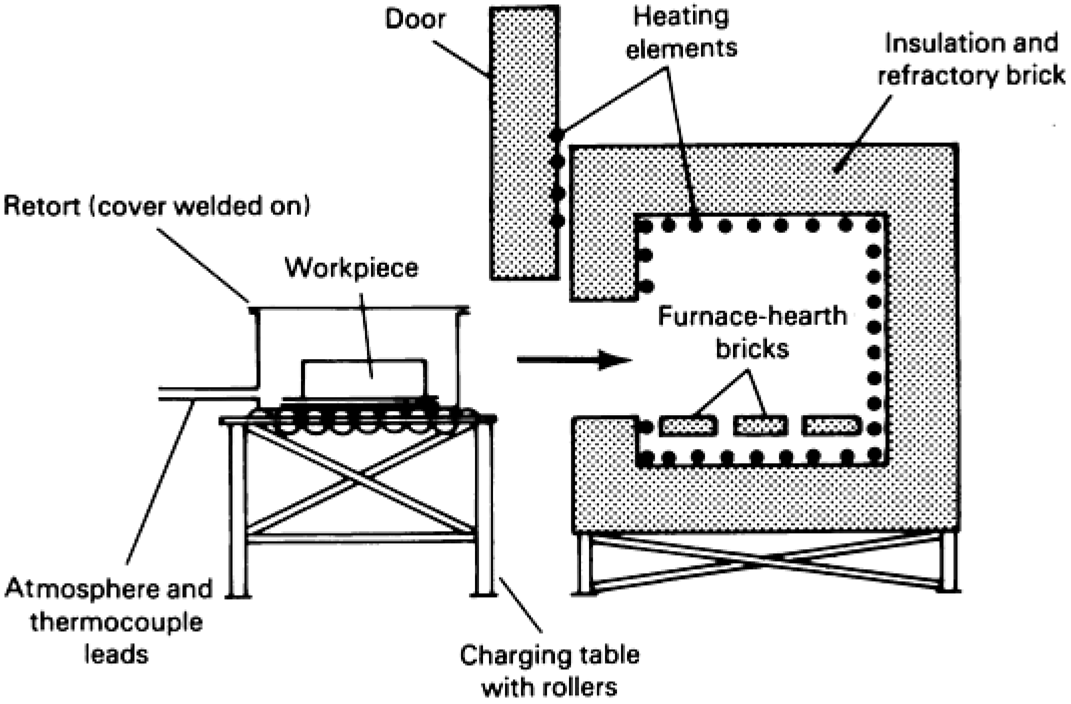
Door
Heating elements
Insulation and refractory brick
Retort (cover welded on)
Workpiece
Furnace-hearth bricks
Atmosphere and thermocouple leads
Charging table with rollers

FIG. 3 SEALED RETORT ABOUT TO BE LOADED INTO ELECTRIC-FIRED BOX FURNACE IN ORDER TO BRAZE A WORKPIECE IN A CONTROLLED ATMOSPHERE

The vacuum furnace is typically used for batch processing (Fig. 4), although a semicontinuous vacuum furnace version is also available. The vacuum brazing furnace is typically interfaced with a microprocessor (Fig. 5). The programmable microprocessor controls all of the brazing parameters: heating rate, brazing temperature and time, cooling rate, and back-filled gas. Because several different brazing cycles can be stored in the microprocessor memory, the vacuum furnaces are quite versatile. Furthermore, an alarm system can be installed in the microprocessor to detect any operational abnormality that has the potential to terminate the brazing operation.

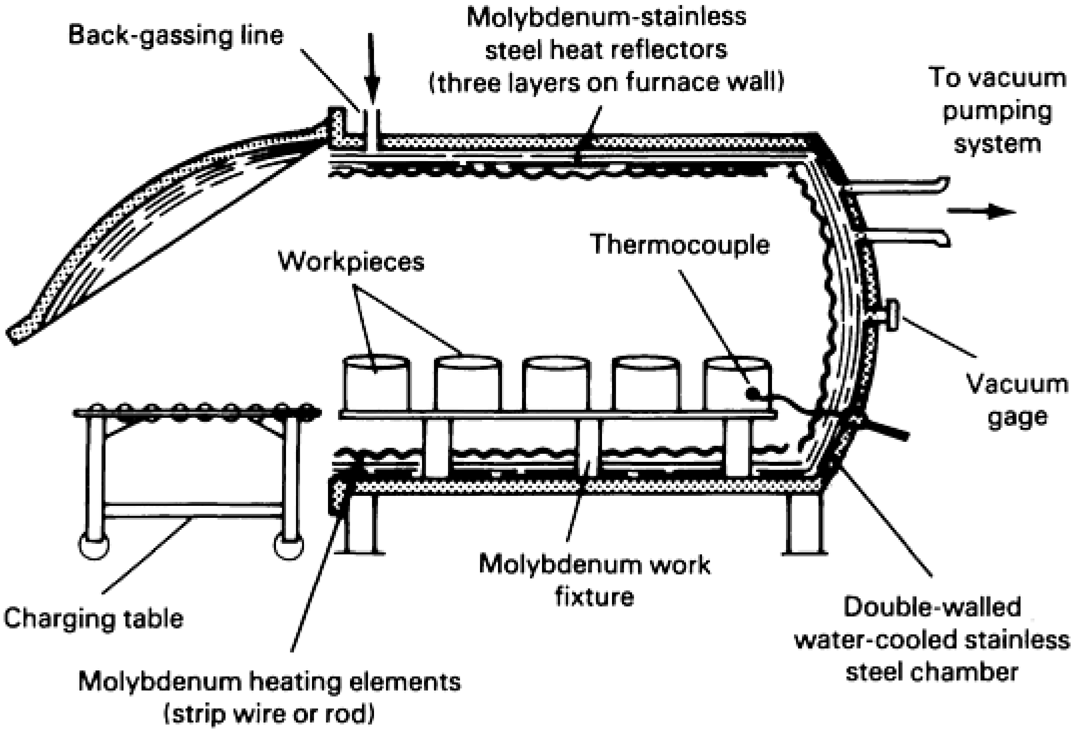
Back-gassing line
Molybdenum-stainless steel heat reflectors (three layers on furnace wall)
To vacuum pumping system
Workpieces
Thermocouple
Vacuum gage
Charging table
Molybdenum work fixture
Double-walled water-cooled stainless steel chamber
Molybdenum heating elements (strip wire or rod)

FIG. 4 TYPICAL CONSTRUCTION OF HORIZONTAL SIDE-LOADED COLD-WALL VACUUM FURNACE, WHICH IS IDEAL FOR BRAZING OF SMALL ASSEMBLIES THAT CAN BE PLACED IN STACKED BASKETS OR ON TIERS OF WORK GRIDS

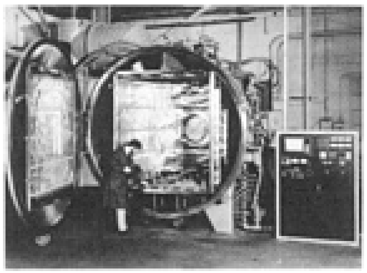

FIG. 5 BRAZING FURNACE INSTALLATION SHOWING RELATIVE SIZE OF BRAZING CHAMBER OF THE UNIT TO THE MICROPROCESSOR-BASED CONTROL MODULE. COURTESY OF CENTORR/VACUUM INDUSTRIES

Furnace Brazing

Toshi Oyama and Howard Mizuhara, WESGO, Inc.

Industrial Applications

The furnace brazing process is employed in many industries because of its high quality and reproducibility. The vacuum device, jet engine, and automotive industries represent three examples. In the vacuum device industry, a vacuum tube with an envelope of alumina ceramic is used, the nature of which requires that the whole assembly be heated. The control of heating and cooling rates is very important to the prevention of alumina ceramic cracking.

Other applications of furnace brazing can be found in the jet engine industry, where the quality of the brazed joint and the dimensional control of the brazed assembly are critical. Vacuum furnace brazing is the most common system used in the production of jet engine components.

The automotive industry makes extensive use of the continuous belt furnace because it is easily automated and is a cost-effective, high-production joining process for components made of carbon steel, stainless steel, and aluminum.

Furnace Brazing

Toshi Oyama and Howard Mizuhara, WESGO, Inc.

Health and Safety Guidelines

Although everyday production is relatively simple and trouble-free, the furnace operator should be instructed on the proper use of the equipment, the hazards of the gases and atmospheres used, and the importance of proper operating procedures. An alarm system should be installed to monitor furnace parameters and to alert operators of any potential equipment breakdowns or hazardous working conditions. Routine maintenance of the furnace and auxiliary equipment is vital if accidents are to be prevented.

Furnace Brazing

Toshi Oyama and Howard Mizuhara, WESGO, Inc.

Selected References

- M.M. SCHWARTZ, *BRAZING,* ASM INTERNATIONAL, 1987
- *BRAZING HANDBOOK,* AWS, 1991
- *MANUFACTURING TECHNIQUES IN BRAZING,* YOUSETSU SHINBUN SHA, 1982 (IN JAPANESE)
- "SAFETY IN WELDING AND CUTTING," STANDARD ZA9.1, ANSI/AWS

Induction Brazing

J.F. Libsch, Lepel Corporation

Introduction

INDUCTION BRAZING (IB) is a process that joins metal components to form an assembly by selectively heating the joint area to the brazing temperature with electrical energy transmitted from an inductor that has been energized by a heating generator. It is classified as a brazing process because joints are made at elevated temperatures (540 °C, or 1000 °F, or above), using copper alloys that often contain silver (Ref 1).

Heating results from resistance to the flow of induced current (I^2R losses) in the assembly components. It occurs when the electrically conductive components are placed in the electromagnetic field created by rapidly alternating current flow in the inductor (Fig. 1). With proper coil design, heating occurs rapidly at the joint area, providing localized brazing temperatures. Response to the electromagnetic field depends on the frequency of the alternating current and the nature of the materials being heated, as well as on the shape of the component parts and coil design. Nonmagnetic materials with low electrical resistance such as copper and brass take longer to heat than magnetic, higher resistivity, ferrous materials such as steel.

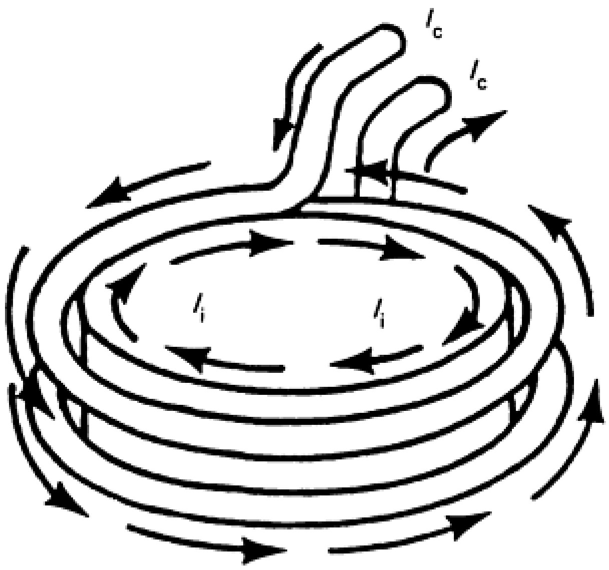

FIG. 1 RELATION OF COIL CURRENT, I_C, TO INDUCED CURRENT IN THE WORKPIECE, I_I, GENERATED BY ELECTROMAGNETIC FIELD IN THE INDUCTION COIL

While a strong electromagnetic field and high power density (2×10^{-3} kW/mm^2, or 1 kW/in.^{2}) can create highly localized temperatures, most brazing applications use relatively low power densities of 1×10^{-3} kW/mm^2 (0.5 kW/in.^{2}) to ensure that the joint area comes to brazing temperature uniformly. Heat conduction plays an important role.

Because a multitude of joint configurations are encountered, proper inductor or coil design is essential to ensure uniform joint temperature and to permit handling of components onto the coil for processing. Some typical coils that are fabricated from copper tubing and that use water cooling are shown in Fig. 2. Inductors made from solid copper bar for single- or multiple-joint heating and that employ both internal and external water cooling are used for some applications.

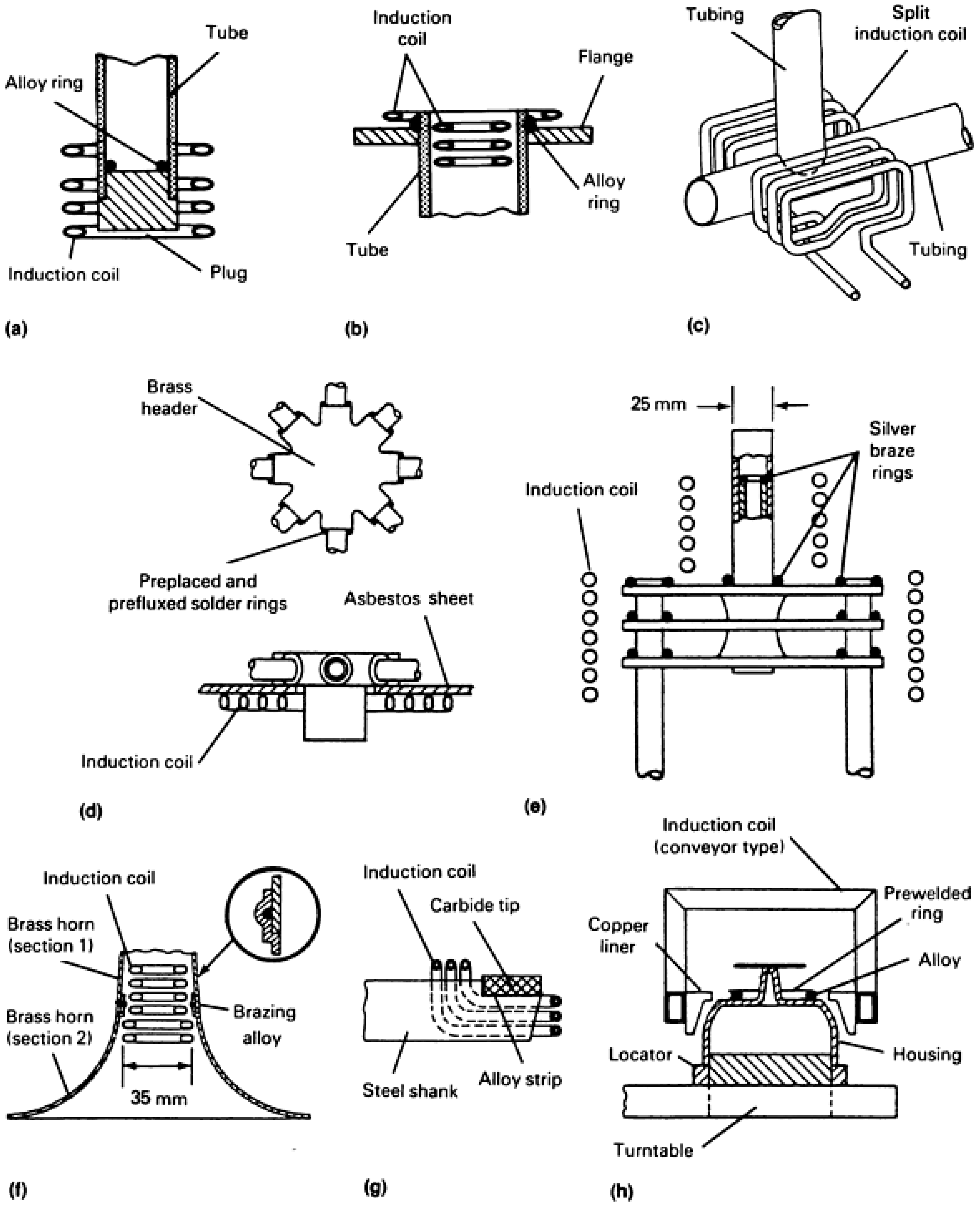
Tube
Alloy ring
Induction coil
Plug
(a)
Induction coil
Flange
Alloy ring
Tube
(b)
Tubing
Split induction coil
Tubing
(c)
Brass header
Preplaced and prefluxed solder rings
Asbestos sheet
Induction coil
(d)
25 mm
Silver braze rings
Induction coil
(e)
Induction coil
Brass horn (section 1)
Brass horn (section 2)
Brazing alloy
35 mm
(f)
Induction coil
Carbide tip
Steel shank
Alloy strip
(g)
Induction coil (conveyor type)
Prewelded ring
Copper liner
Alloy
Locator
Housing
Turntable
(h)

FIG. 2 TYPICAL COIL AND JOINT CONFIGURATIONS USED IN INDUCTION BRAZING. (A) SOLENOID COIL FOR PLUG-TO-TUBE JOINT (NOTE LOCATION OF BRAZING ALLOY RING). (B) INTERNAL-EXTERNAL COIL FOR FLANGE-TO-TUBE JOINT (FLANGE CHAMFERED TO ASSIST PREPLACED ALLOY RING. (C) SPLIT SOLENOID COIL FOR TUBE-TO-TUBE T-JOINT (BRAZE RING PREPLACED INTERNALLY TO PROVIDE UNIFORM FILLET). (D) PANCAKE OR PIE WOUND COIL FOR HEATING BRASS HEADER TO PERMIT SIMULTANEOUS BRAZING OF EIGHT COPPER TUBES TO HEADER. (E) EXTERNAL COILS FOR SIMULTANEOUS PRODUCTION OF A NUMBER OF BRAZED JOINTS. (F) FORMED INTERNAL COIL TO JOIN TWO SECTIONS OF MUSICAL HORN (SPECIAL JOINT DESIGN FOR PREPLACEMENT OF BRAZING ALLOY MAINTAINS SMOOTH JOINT ON OUTSIDE OF INSTRUMENT. (G) OPEN-END COIL FOR BRAZING CARBIDE TIPS TO SHANKS (ALLOY STRIP PREPLACED AS SHOWN). (H) CONVEYOR TYPE COIL FOR CONTINUOUSLY BRAZING FUSE ASSEMBLIES ON A ROTARY FIXTURE

Reference

1. "SPECIFICATION FOR FILLER METALS FOR BRAZING," A.58-89, AMERICAN WELDING SOCIETY, 1989

Induction Brazing

J.F. Libsch, Lepel Corporation

Process Applications

Induction brazing is used where relatively high-strength heat-resistant joints are required between components parts. The strength of the joint depends on the joint design, the brazing alloy, the strength of the components being joined, and the clearance between the components. A clearance of 0.038 to 0.050 mm (0.0015 to 0.002 in.) is recommended. Induction brazing has been used extensively in the production of consumer and industrial products, structural assemblies, electric and electronic products, mining equipment, machine and hand tools, military and ordnance equipment, and aerospace components.

Advantages. Induction brazing features rapid heating that produces joints in seconds because each joint area is exposed to the electromagnetic field of the specific coil involved. Uniform and reliable joints can be produced piece after piece with smooth joint fillets, eliminating the need for finishing operations. Because a suitable induction coil can provide heat simultaneously at several locations in an assembly, a number of joints can be made in a single heating operation. Induction brazing provides control of the heat pattern and rapid heating, which serve to localize the heated zone, thereby preventing or minimizing deterioration of properties in previously heat-treated or cold-worked materials, and, if desired, permitting sequential brazing of nearby joints using alloys with decreasing melting temperatures.

Induction brazing uses predetermined amounts of brazing alloys in the form of rings, strip, or powder; thus, the possibility of excessive material use is avoided, lowering joining alloy costs. Rapid heating minimizes oxidation and discoloration in most instances, even without the use of a controlled atmosphere.

Limitations of induction brazing include:

- DIFFICULTY IN HEATING OF COMPLEX ASSEMBLIES
- NEED FOR RELATIVELY UNIFORM CLEARANCE BETWEEN COMPONENTS
- INITIAL COST OF EQUIPMENT
- NEED FOR SPECIALIZED KNOWLEDGE FOR SYSTEM DESIGN

For complex assemblies involving several brazed joints and special geometry, induction brazing fixturing may be difficult, and furnace brazing may be a more practical process choice.

Induction Brazing

J.F. Libsch, Lepel Corporation

Brazing Equipment and Fixturing

Major changes have been made in the design of generators and temperature control equipment in recent years. Most significant perhaps has been the development of solid-state generators to complement or replace motor generators and vacuum tube generators. Solid-state generators of intermediate power levels that operate at frequencies of 50 to 200 kHz or 50 to 350 kHz have become available, providing an alternate to tube generators that operate at nominal frequencies of 450 kHz. Improvements have been made in generator control design, permitting precise and instantaneous response to changes transmitted by a suitable sensor element.

The proper size generator required for a given operation is best determined by an actual trial in the laboratory. However, a preliminary estimate can be made by considering the power absorbed by the pieces to be joined, the amount of radiation lost from the heated piece, and the required production rate. (The accuracy of this estimate depends on good coil design and electrical matching between the generator and the work load.) The curves in Fig. 3 represent the approximate heat content or heat energy required to heat 1 kg (2.2 lb) of material to the temperatures indicated from 20 °C (68 °F); these curves reflect the variation in specific heat with temperature for each metal.

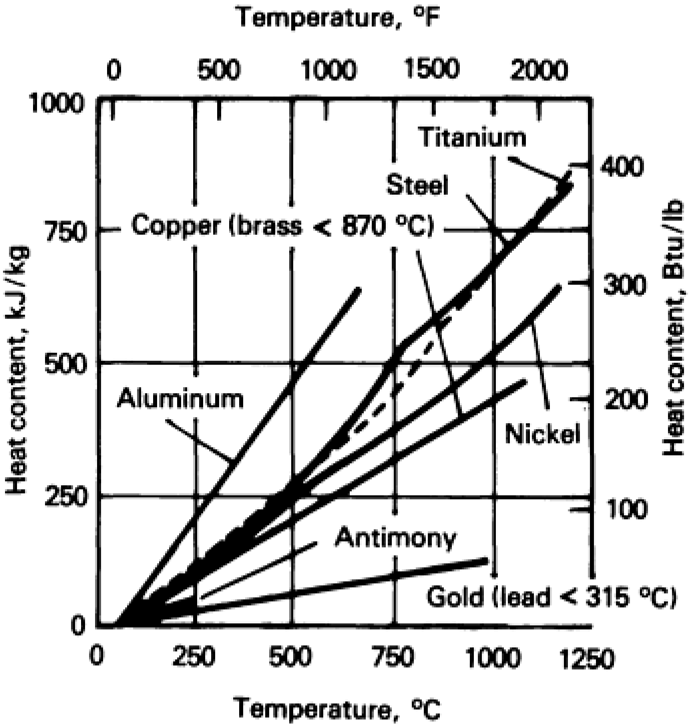

FIG. 3 PLOT OF HEAT CONTENT (MASS BASIS) VERSUS TEMPERATURE FOR SELECTED METALS AND ALLOYS. SOURCE: REF 3

Precise determination of the weight being heated in most brazing applications is difficult because heat is conducted away from the joint area, thus creating a temperature gradient for some distance beyond the induction-heated joint and adding to the amount of energy required. Furthermore, Fig. 3 does not account for the energy required to compensate for radiation losses. Fortunately, the temperatures most often used in induction brazing are relatively low (620 to 815 °C, or 1150 to 1500 °F); thus, radiation losses are small and can generally be neglected in the preliminary estimate. While rapid heating rates using high power may increase production, they are a frequent source of difficulty in induction brazing; very rapid heating exaggerates the effects of electrical resistivity and thermal conduction. Slower heating rates minimize the effects of thermal conduction when dissimilar metal joints, such as steel to copper, are involved. In some situations, slower heating rates can help compensate for considerable differences in the mass of the components being brazed.

The most common frequencies currently used for brazing applications are 10 kHz, 25 to 50 kHz, 50 to 200 kHz, and 50 to 350 kHz with solid-state generators, and 250 to 450 kHz with tube equipment. The lower frequencies heat more deeply and may provide more uniform heating of the matching surfaces if one of the components to be brazed has a large mass. Higher operating frequencies, such as 450 kHz, may be suitable for brazing of nonferrous metals, or for brazing of steel to a nonferrous metal or to a nonmagnetic (austenitic) steel. High frequency may also provide more efficient heating of parts; in fact, tube equipment operating at 2.5 to 8 MHz may be necessary for very thin components. As a general. rule, cost in dollars per kilowatt is lower with lower frequency generators.

Figure 4 shows a system for induction brazing components for refrigeration and air conditioning using a 7.5 kW solid-state unit operating at a frequency of 100 kHz.

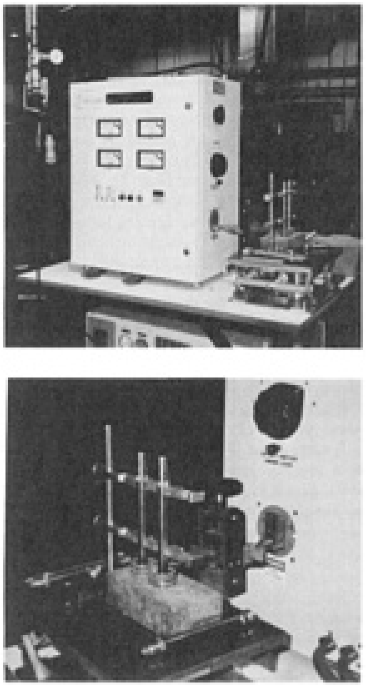

FIG. 4 INDUCTION BRAZING INSTALLATION THAT USES A 7.5 KW SOLID-STATE INDUCTION GENERATOR OPERATING AT 100 KHZ TO JOIN BRASS AND STAINLESS STEEL COMPONENTS FOR HEAVY-DUTY VEHICLE REFRIGERATION AND AIR CONDITIONING SYSTEMS. COURTESY OF RED DOT CORPORATION AND LEPEL CORPORATION

An inert or reducing atmosphere surrounding the brazed joint area is desirable in some applications. Appropriate fixturing and/or bell-jar arrangements (Ref 4) are used in such applications.

References cited in this section

3. C.A. TUDBURY, *BASICS OF INDUCTION HEATING,* VOL 1
4. P.F. GERBOSI AND J.F. LIBSCH, CONTROLLED-ATMOSPHERE BRAZING WITH INDUCTION HEATING, *WELD. J.,* OCT 1989, P 32-37

Induction Brazing

J.F. Libsch, Lepel Corporation

Joint Design

Figure 2 illustrates several basic joint designs for induction brazing as well as typical coils and some frequent applications (Ref 5). The schematics show the placement of brazing alloy preforms and coil-part arrangements.

Joint Parameters (Ref 2). The following joint parameters for parts to be brazed by induction heating require special attention:

- HEATING PATTERN
- METHOD OF PREPLACING THE JOINING ALLOY
- CLEARANCES BETWEEN MATING PARTS

- THERMAL CONDUCTIVITY AND EXPANSION CHARACTERISTICS OF THE MATERIALS TO BE JOINED

Effect of Thermal Expansion on Stress (Ref 2). Stress concentration and high residual stresses in the joint become critical factors when the joint members are stronger than the brazing filler metal, or when differential contraction takes place with dissimilar materials. The marked difference in the thermal expansion between tungsten carbide (5.9×10^{-6}/K, or 3.3×10^{-6}/°F) and carbon steel (12.1×10^{-6}/K, or 6.7×10^{-6}/°F) may be particularly troublesome when the carbide is placed in tension. This frequently results in cracked carbide. A "sandwich" braze utilizing a clad brazing strip (that is, copper clad with silver brazing filler metal on both sides) can overcome this problem. In a sandwich braze, stresses are minimized as a result of plastic deformation in the low-yield-strength copper layer.

References cited in this section

2. *BRAZING HANDBOOK*, 4TH ED., AMERICAN WELDING SOCIETY, 1989, P 214-218
5. J.F. LIBSCH AND P. CAPOLONGO, HIGH FREQUENCY INDUCTION BRAZING AND SOLDERING, *LEPEL REV.*, VOL 1 (NO. 5)

Human Factors Engineering Ergonomics

In induction brazing, the heating function is essentially independent of the operator once the generator, coil, and heating program have been set up. The amount of handling of the component parts, the brazing alloy preforms, and flux depend upon the degree to which the application is automated. In fully automated systems, the operator oversees the operations and ensures the supply of the various joint elements from feed mechanisms, including mechanisms that provide automatic placement of brazing alloy preforms and fluxing. In semiautomatic systems, the operator loads the component parts on a fixture, preplaces the preform, and fluxes the joint region at a workstation on a continuously moving conveyor, rotary, or indexing device. In these systems, the operator serves primarily as an unskilled machine operator. Specialized complex assemblies (for example, electron tubes) that are brazed in controlled atmospheres may require skilled personnel with specialized training in induction heating processing.

Good joint design provides for simple inspection of the assembly, which involves simply noting alloy flow from the joint area to the edges of the joint. The presence of smooth continuous fillets also indicates a successful joint. Complex inspection methods such as radiography or ultrasonic inspection may be specified in critical applications.

References

1. "SPECIFICATION FOR FILLER METALS FOR BRAZING," A.58-89, AMERICAN WELDING SOCIETY, 1989
2. *BRAZING HANDBOOK*, 4TH ED., AMERICAN WELDING SOCIETY, 1989, P 214-218
3. C.A. TUDBURY, *BASICS OF INDUCTION HEATING*, VOL 1
4. P.F. GERBOSI AND J.F. LIBSCH, CONTROLLED-ATMOSPHERE BRAZING WITH INDUCTION

HEATING, *WELD. J.*, OCT 1989, P 32-37
5. J.F. LIBSCH AND P. CAPOLONGO, HIGH FREQUENCY INDUCTION BRAZING AND SOLDERING, *LEPEL REV.*, VOL 1 (NO. 5)

Induction Brazing

J.F. Libsch, Lepel Corporation

Selected References

- M.M. SCHWARTZ, *BRAZING,* ASM INTERNATIONAL, 1987

Dip Brazing

Daryl D. Peter, Daryl D. Peter and Associates

Introduction

DIP BRAZING (DB) is one of the oldest brazing processes. The materials to be joined are immersed in a hot liquid, which is either a molten flux or a molten filler metal that usually contains a layer of flux to prevent oxidation. Dip brazing is primarily used because of its ability to uniformly heat parts that have a wide range of thicknesses. The process is simple when compared with modem, complex brazing equipment and processes. The drawback of the process is that it coats everything with flux or braze alloy; flux can be corrosive to fixtures and surrounding metal structures and equipment. Even when compared with more modern heating methods, the speed of immersion heating is hard to match and dip brazing, therefore, continues to be a common process.

Dip Brazing

Daryl D. Peter, Daryl D. Peter and Associates

Furnace Construction

Dip brazing systems can be described by the type of pot vessel, heating method, and heating media that is used. Vessel construction can be ceramic, fired brick, or, possibly, cast (Fig. 1). A metal tank of a suitable alloy can also be used. Heating is facilitated by either gas or electric current. Flux baths favor electric heating, because of their higher resistivity when molten.

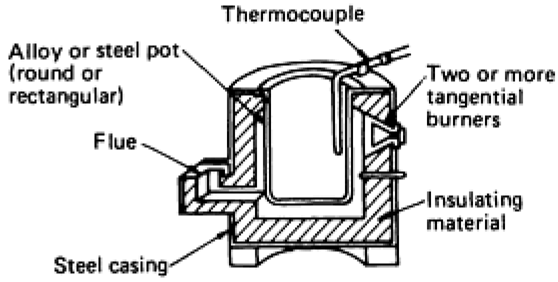
Thermocouple
Alloy or steel pot (round or rectangular)
Two or more tangential burners
Flue
Insulating material
Steel casing

(a) Gas fired or oil fired

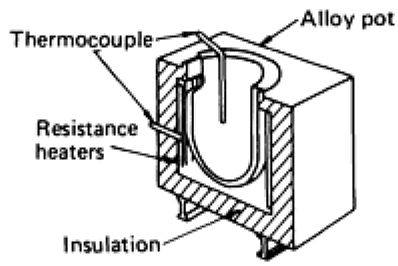
Alloy pot
Thermocouple
Resistance heaters
Insulation

(b) Resistance heated

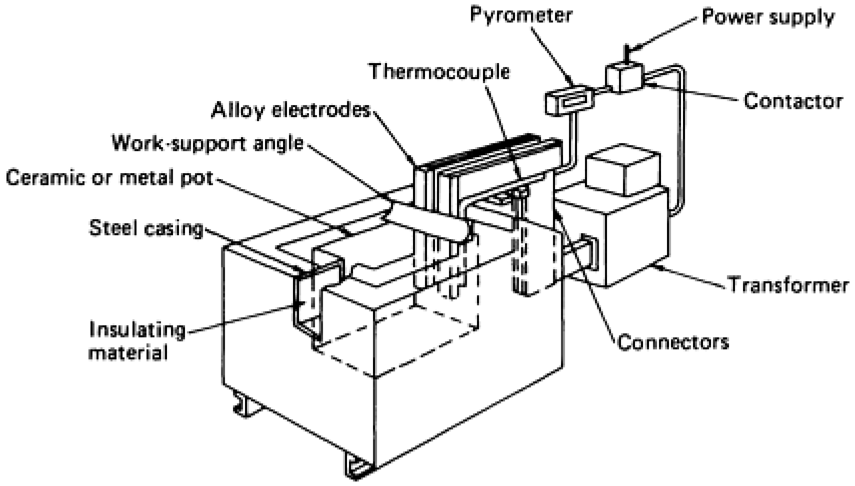
Pyrometer
Power supply
Thermocouple
Contactor
Alloy electrodes
Work-support angle
Ceramic or metal pot
Steel casing
Transformer
Insulating material
Connectors

(c) Immersed electrodes

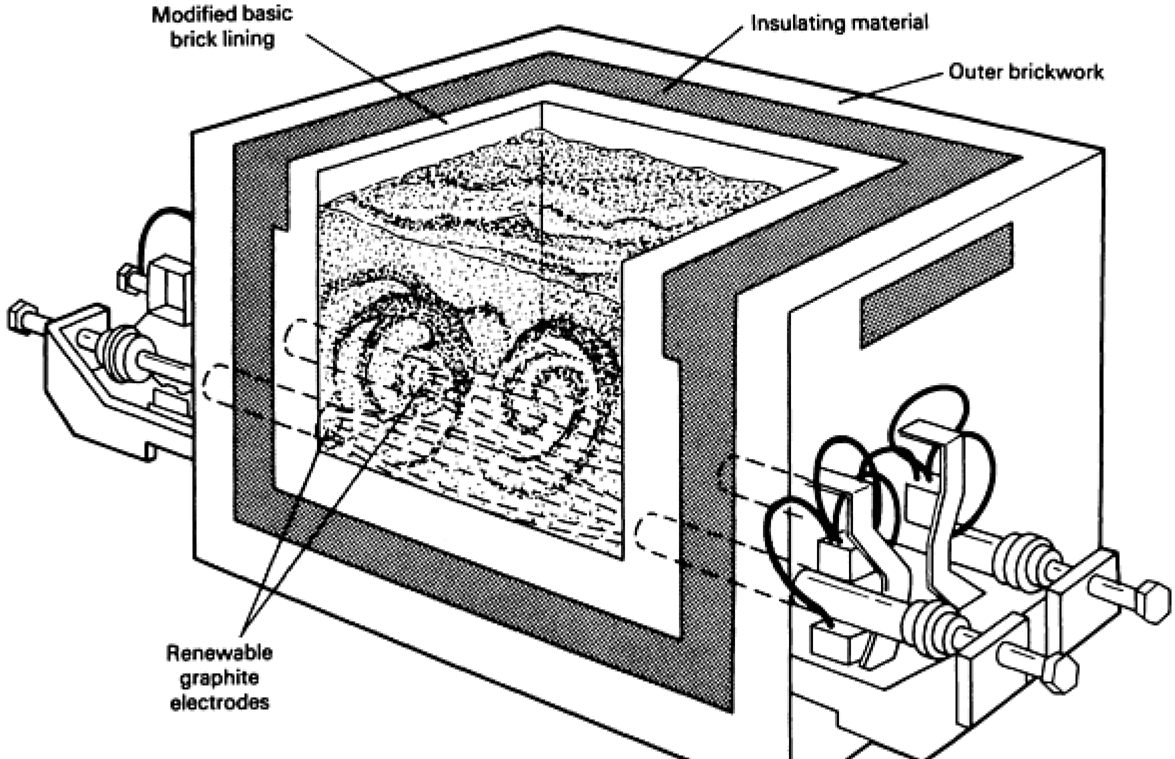
Modified basic brick lining
Insulating material
Outer brickwork
Renewable graphite electrodes

(d) Electrically heated (submerged electrodes)

FIG. 1 PRINCIPAL TYPES OF FURNACES USED FOR MOLTEN-SALT-BATH DIP-BRAZING APPLICATIONS. (A) AND (B) EXTERNALLY HEATED; (C) AND (D) INTERNALLY HEATED

Figure 2 shows a typical furnace used for flux brazing applications. Electricity is conducted into the bath by either carbon or nickel electrodes that penetrate the furnace wall. Ultrasonic generators can also be incorporated into the molten metal processes to provide a scrubbing action.

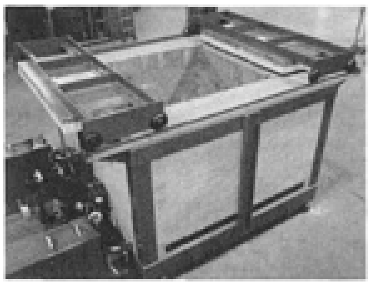

FIG. 2 SUBMERGED-ELECTRODE FURNACE INSTALLATION INCORPORATING A TOP-LOADING DOOR MECHANISM WITH RAIL TRAVERSE

Dip Brazing

Daryl D. Peter, Daryl D. Peter and Associates

Process Details

The process for brazing in molten flux is straightforward. To prepare for brazing, the operator must perform the following steps:

- CLEAN THE WORKPIECE
- MAINTAIN FLUX ACTIVITY
- ASSEMBLE THE PARTS AND POSITION THE FILLER METAL
- FIXTURE TO MAINTAIN CONTACT WITH PARTS
- HEAT UNIFORMLY TO MINIMIZE DISTORTION

The fixtured parts are then placed into a preheated furnace and heated to a temperature that is both above the liquidus of the flux in the immersion bath and below any temperature detrimental to the fixtured parts. Next, the parts are transferred to the bath and immersed for a period of time that is sufficient to reach brazing temperatures. They are then removed and allowed to drain and cool enough to set the brazing alloy, before being transferred to a cooling area. When the parts have cooled to a temperature at which distortion is no longer a factor, they are then quenched in hot water. The flux is then removed and the brazements are cleaned.

The process technician must work closely with the designer to ensure that the part design will not trap flux or air during brazing. In addition, the process technician is typically responsible for the choice of materials used in fixturing (ensuring that they are compatible with the material in the bath, immersion angles to prevent trapped air, and speed through the liquid to control erosion). Fluxes can be very damaging to most metals. Stainless steel that is satisfactory for torch or induction brazing applications may not be suitable for dip brazing. High-nickel content alloys are therefore more commonly used, because of their superior corrosion-resistant properties.

Dip Brazing

Daryl D. Peter, Daryl D. Peter and Associates

Equipment Maintenance

Preventive maintenance is critical to the process and can be divided into three categories:

- TEMPERATURE CONTROL
- CONTROL OF THE LIQUID
- MAINTENANCE OF THE VESSEL

A daily satisfactory braze sample tends to verify that all is well. Once every 30 days, an inspection of the operating electronics for accuracy and a survey of the bath temperature are recommended. The composition of the bath tends to change slowly. However, the time interval between tests will be determined by how the bath is used and by the nature of the bath. This test frequency interval can range from one week to one month in duration.

Care of the vessel is usually the most neglected part of the preventive maintenance system. Besides looking unkempt, it is totally inhospitable, because of the heated liquid. As the repository of the molten material, the vessel also holds some form of solid residues (for example, oxides, salts, and organics) from all items that have been dipped since its last thorough cleaning. It is also likely to contain parts of fixtures, parts that fell off the rack, dripped braze alloy, and other substances.

For the most efficient performance, it is necessary to clean the pot daily. The material (that is, sludge) that collects on the vessel bottom stays soft for a short time and should be removed on a regular basis. Chipping and hammering to remove hardened sludge material on cemented fire brick can shorten the life of the pot. Common techniques for removing sludge before hardening include scraping and scooping. Scooping tools should be ventilated to allow the liquid to drain back into the pot. Disposal of the sludge should conform to environmental regulations. Sludge buildup causes problems with both bath chemistry and heating. If allowed to remain, the buildup will eventually either produce poor-quality brazements or cause damage to the pot.

Dip Brazing

Daryl D. Peter, Daryl D. Peter and Associates

Applications

The use of dip brazing in a molten filler metal bath is limited to the brazing of wire, small components, and aluminum return bends, whereas dip brazing in a molten-salt (flux) bath is used extensively for the brazing of aluminum, copper, and ferrous alloys (Table 1).

TABLE 1 PARAMETERS FOR MOLTEN-SALT DIP BRAZING OF SELECTED ALLOY ASSEMBLIES

WORKPIECE ALLOY(S)	TEMPERATURE		FILTER METAL	SALT
	°C	°F		
ALUMINUM	540-615	1000-1140	(A)	FLUORIDE-CHLORIDE-BASE
COPPER	815-870	1500-1600	BCUP(B)	CHLORIDE-BASE
FERROUS	915-955	1675-	RBCUZN	CHLORIDE-BASE +

		1750		BORAX[(C)]
	980-1205	1800-2200	BNI	CYANIDE-BASE
	1120	2050	BCU	CHLORIDE-BASE
FERROUS AND NONFERROUS	675-1010	1250-1850	BAG[(D)]	CHLORIDE-BASE

Source: Ref 1

(A) SALT ACTS AS BOTH FILLER METAL AND HEATING MEDIUM.
(B) SELF-FLUXING.
(C) NO DEZINCIFICATION TAKES PLACE.
(D) SEPARATE FLUX REQUIRED.

Dip-brazed parts are ideally suited for limited as well as high-production runs. The method of applying the process can be either manual or automatic. The process is used to braze small to very large assemblies with multiple or hidden joints. It is appropriate for all metals that can be brazed and is particularly suited for aluminum and other alloys that have melting points very close to the brazing temperature. The brazing operation can also simultaneously serve as a source for heat treating, carburizing, or cyaniding.

Reference cited in this section

1. *BRAZING HANDBOOK*, 4TH ED., AWS, 1991, P 223

Dip Brazing

Daryl D. Peter, Daryl D. Peter and Associates

Carbon and Low-Alloy Steels

The types of salts used in the dip brazing of carbon and low-alloy steels are neutral chloride salts, neutral chloride salts plus a fluxing agent such as borax or cryolite, and carburizing and cyaniding salts, which are also fluxing types of salts. Specific types and compositions of brazing salts and temperatures used for brazing of carbon and low-alloy steels with various filler metals are given in Table 2.

TABLE 2 TYPICAL SALTS USED FOR MOLTEN-SALT DIP BRAZING OF CARBON AND LOW-ALLOY STEELS WITH SELECTED FILLER METALS

FILLER METAL	**TYPE OF SALT**	**NOMINAL COMPOSITION, %**	**BRAZING TEMPERATURE RANGE**[(A)]	
			°C	**°F**
BAG-1 THROUGH BAG-8, AND BAG-18	NEUTRAL	55 $BACL_2$, 25 NACL, 20 KCL	620-870	1150-1600
	CYANIDING-FLUXING	20-30 NA_2CO_3, 20-30 KCL, 30-40 NACN	650-870	1200-1600
	NEUTRAL	50 NACL, 50 KCL	730-870	1350-1600
RBCUZN-A	NEUTRAL	80 $BACL_2$, 20 NACL	915-940	1675-1725

	FLUXING	79 $BACL_2$, 20 NACL, 1 BORAX	915-940	1675-1725
	CARBURIZING-FLUXING (WATER SOLUBLE)	30 NACL, 30 KCL, 20 CARBONATE, 15-20 NACN, ACTIVATOR (PROPRIETARY)	915-940	1675-1725
	CARBURIZING AND SELF-FLUXING	50 CARBONATE, 50 CHLORIDE WITH GRAPHITE ADDITION$^{(B)}$	815-925	1500-1700
RBCUZN-D	NEUTRAL	90 $BACL_2$, 10 NACL	1040-1050	1900-1925
BCU-1 AND 1A	NEUTRAL	95 $BACL_2$, 5 NACL	1095-1150	2000-2100
	NEUTRAL	100 $BACL_2$	1095-1150	2000-2100

(A) TEMPERATURES SHOWN ARE THOSE OF THE SALT BATH.

(B) USED WITH MECHANICAL AGITATION

Neutral salts, so called because normally they do not add or subtract anything from the surface of the steel being treated, protect the surface from attack by oxygen in the air. Oxide on the workpiece, however, cannot be reduced by the salt, and a flux must generally be provided.

The neutral salts are mildly oxidizing to steel when they are used at recommended austenitizing temperatures. The oxides produced by heating steel in molten salt are largely soluble; hence, the steel is scale-free after heating. The accumulation of oxide in the molten salt, however, progressively makes the salt more decarburizing, and for this reason baths may require periodic rectification.

Flux that is applied to the surface of the assembly and dried before the assembly is immersed in the neutral salt will be quickly dissipated by dissolving in the salt or escaping from the surface of the bath as a volatilized salt or gas. For this reason, there is generally no difficulty in removing flux from an assembly that has been brazed in a salt bath.

Fluxing agents, such as borax and cryolite, are added to neutral chloride salts to produce a fluxing environment in the bath. When these fluxing agents are used with silver alloy or copper-zinc filler metals, periodic flux additions are required to maintain the fluxing potential of the bath. Above 650 °C (1200 °F), the fluxing potential can decrease rapidly because of oxidation from contact with air or the parts being brazed; therefore, the fluxing agent must be replenished more frequently.

Carburizing and cyaniding salts provide their own fluxing action. In addition, they supply carbon or carbon and nitrogen to the surface of the steel assembly as it is being brazed. Although silver brazing alloys have been used successfully, RBCuZn-A filler metal is generally preferred. A case depth of up to 0.30 mm (0.012 in.) can be obtained without adversely affecting the quality of a joint brazed with this copper-zinc filler metal.

Fluxes. An adequate fluxing environment is needed to ensure good flow and penetration of the brazing alloy in molten-salt-bath dip brazing. When brazing is done in a neutral flux chloride salt bath, an active flux is usually applied to the assemblies before brazing. Generally, the application of additional flux to the assembly isnot necessary when using a cyanide bath or another active fluxing bath.

Flux can be applied by brushing, dipping, or spraying the parts to be brazed before, during, or after assembly. After flux application, if any moisture is present, the assemblies must be preheated to dry them before immersion in the salt bath. Typical fluxes employed for prefluxing carbon steels and low-alloy steels that are to be brazed in a salt bath are American Welding Society (AWS) types 3A and 3B.

A temporary insulating cocoon of frozen salt forms instantly around a piece of cold metal when it is immersed in a molten-salt bath, and externally located filler metal can reach the melting temperature range before the workpiece has reached a temperature high enough for proper wetting to take place. When this occurs, the molten filler metal flows away from the joint with no brazing. This can be avoided either by preheating the assembly to a temperature above the melting range of the flux or by relocating the filler metal inside the joint in grooves, recesses, or drilled holes.

Filler Metals. The brazing filler metals shown in Table are the most widely used for molten-salt-bath dip brazing of carbon and low-alloy steels. Although silver alloys BAg-13 and 13A (not shown in Table 2) can be used for brazing in a molten-salt bath, they have been supplanted in most applications by copper-zinc alloys, which are less costly and have similar brazing temperature ranges. The rapid heating rate and nonoxidizing environment in a salt bath minimize dezincification of copper-zinc alloys, thereby facilitating the use of these alloys.

The brazing filler metal must be in contact with the joint. Filler metals in the form of wire, strip, powder, paste, and cladding are available. Special preform rings or other shapes can be obtained for specific joint requirements.

For additional information, see the article "Brazing of Carbon Steel and Cast Iron" in this Volume.

Stainless Steels

The brazing of stainless steel by immersing all or a portion of the assembly in molten salt offers essentially the same advantages that would apply to brazing similar assemblies made of carbon steel. The same limitations are also applicable. See the article "Brazing of Stainless Steels" in this Volume for more information.

Cast Irons

One of the more-common applications for the dip brazing of cast iron is the joining of steel tubing to headers, special fittings, and stanchions. Various mechanical assemblies are brazed to cast iron because different metal properties are needed in different parts of the assembly or, more often, because the fitting portion of such an assembly can be made most economically from cast iron.

Assemblies containing malleable iron castings are water quenched to remove the brazing salt--a practice that is common with steel assemblies. To avoid warpage and cracking, however, brazed assemblies containing gray and ductile iron castings should never be water quenched from the brazing temperature. Instead, the assemblies should be cooled slowly to about 150 °C (300 °F) and then immersed in water to remove the salt.

For additional information, see the article "Brazing of Carbon Steel and Cast Iron" in this Volume.

Aluminum Alloys

The best method of heating and fluxing aluminum joints simultaneously is to immerse the entire assembly in a bath of molten flux. Because of the low specific heat of flux, assemblies usually are heated to about 540 °C (1000 °F) prior to flux immersion. Dip brazing has been used successfully in the manufacture of complex, multiple-joint heat exchangers.

Immersing the entire assembly into molten flux has many advantages. Heat is applied to all parts simultaneously and uniformly, and air is replaced by a buoyant and surface-active environment, promoting brazing filler metal flow. In addition, the uniform temperature permits production assembly of parts with dimensional tolerances as low as ±0.05 mm (±0.002 in.) or even less.

Heat-transfer units assembled from alternate corrugated and flat aluminum brazing sheets or from various crimped and formed pieces are examples of the type of work that dip brazing can handle advantageously. Units weighing up to 9 Mg (20,000 lb) have been joined by dip brazing. Certain designs have to withstand a service pressure of 8 MPa (1150 psi). Brazing sheet is essential to this type of work, reducing assembly and brazing costs. The rapid and even heating and flux buoyancy minimize distortion.

For assemblies designed with components in close proximity, flux removal can be tedious and expensive. For instance, when components such as those of a heat-exchanger matrix are spaced closer than 3.2 mm ($\frac{1}{8}$ in.), the flux holds to the surfaces by surface tension and capillary action; it will not drain from the components freely. This is not as great a problem with spacings greater than 3.2 mm ($\frac{1}{8}$ in.) between components of normal length; wider spacing of long components is desirable.

For additional information, see the article "Brazing of Aluminum Alloys" in this Volume.

Daryl D. Peter, Daryl D. Peter and Associates

Safety Precautions

When dip-brazing furnaces contain flux, corrosion of the surroundings is a problem. Ventilation is critical, and protection of all metal exposed to the fumes is mandatory. Scrubbers are often required for the air removed from above the liquid.

Safety equipment, such as full-face protection, heat suits, heat shields, and gloves, should be required when working around molten flux or metal. Clothing of synthetic fibers should be prohibited. In addition, nothing wet or even damp should be immersed, because the resulting burst of steam may blow the molten material over anyone standing near the flux furnace installation.

Daryl D. Peter, Daryl D. Peter and Associates

Reference

1. *BRAZING HANDBOOK*, 4TH ED., AWS, 1991, P 223

Resistance Brazing

Allen Cedilote, Industrial Testing Laboratory Services

Introduction

RESISTANCE BRAZING (RB) is a resistance joining process. The workpieces are heated locally, and the filler metal that is preplaced between the workpieces is melted by the heat generated from resistance to the flow of electric current through the electrodes and the work. In the usual application of RB, the heating current is passed through the joint itself. Equipment is the same as that used for resistance welding, and the pressure needed for establishing electrical contact across the joint is ordinarily applied through the electrodes (Fig. 1). The electrode pressure is also the usual means for providing the tight fit needed for capillary behavior in the joint. The heat for resistance brazing can be generated mainly in the workpieces themselves, in the electrodes, or in both, depending on the electrical resistivity and dimensions.

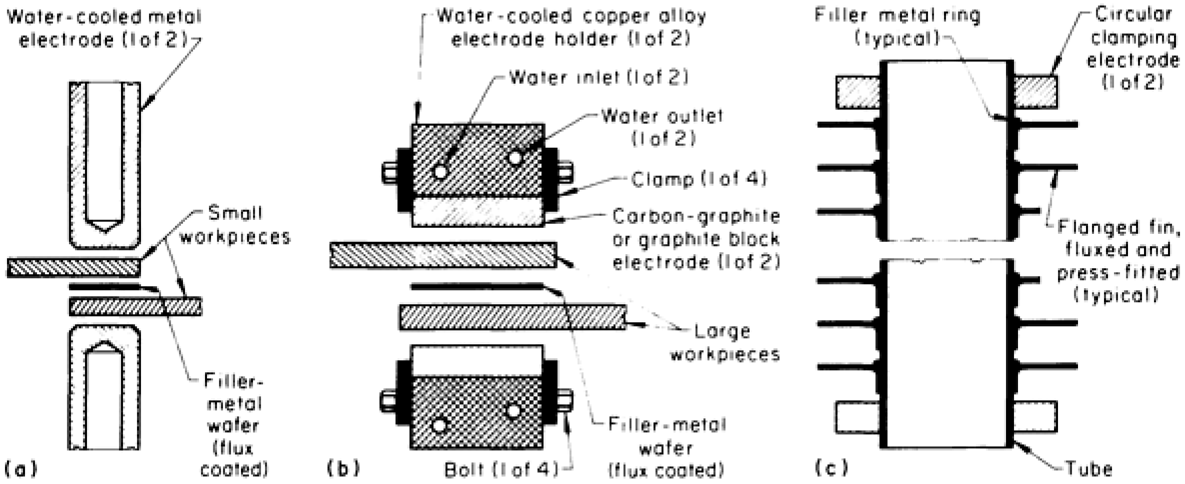
Water-cooled metal electrode (1 of 2)
Small workpieces
Filler-metal wafer (flux coated)
(a)
Water-cooled copper alloy electrode holder (1 of 2)
Water inlet (1 of 2)
Water outlet (1 of 2)
Clamp (1 of 4)
Carbon-graphite or graphite block electrode (1 of 2)
Large workpieces
Filler-metal wafer (flux coated)
Bolt (1 of 4)
(b)
Filler metal ring (typical)
Circular clamping electrode (1 of 2)
Flanged fin, fluxed and press-fitted (typical)
Tube
(c)

FIG. 1 TYPICAL SETUPS FOR RESISTANCE BRAZING. (A) FOR SMALL PARTS OR SMALL FLAT PORTIONS OF LARGER COMPONENTS, USING OPPOSED WATER-COOLED METAL ELECTRODES OF THE CONVENTIONAL RESISTANCE WELDING TYPE. (B) FOR LARGE FLAT PARTS, TYPICALLY OF A HIGHLY CONDUCTIVE METAL (FOR EXAMPLE, COPPER), USING OPPOSED CARBON BLOCK ELECTRODES ATTACHED TO WATER-COOLED COPPER ALLOY ELECTRODE HOLDERS. (C) FOR JOINING FLANGED FINS TO A TUBE, USING CIRCULAR CLAMPING ELECTRODES

Resistance brazing is used for many applications involving small workpieces, for small joints that are part of very large equipment, or for low-volume production runs because it provides:

- VERY RAPID AND HIGHLY LOCALIZED HEATING
- FLAMELESS, CLEAN OPERATION
- PRECISE BRAZING THAT IS EASY TO CONTROL
- HIGH-QUALITY JOINTS WITHOUT EXTENSIVE OPERATOR TRAINING
- EQUIPMENT COST SAVINGS (ONE TRANSFORMER AND A PAIR OF ELECTRODES CAN BE USED FOR A WIDE VARIETY OF JOINT DESIGNS)
- EASY PORTABILITY IN SOME OF ITS EQUIPMENT (FIG. 2)

However, RB is impractical for:

- JOINT AREAS THAT EXCEED APPROXIMATELY 1300 MM^2 (2 IN.^{2})
- WORKPIECES THAT ARE LARGE OR HEAVY
- WORKPIECES THAT HAVE VARIABLE THICKNESS OR COMPLEX CONTOURS

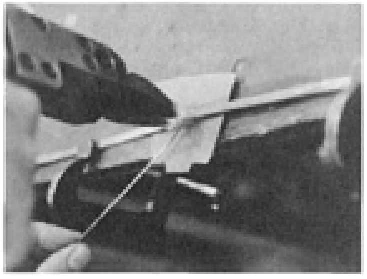

FIG. 2 PORTABLE ELECTRICAL RB ELECTRODES USED TO JOIN A WIRE TO A PLATE. COURTESY OF ROBIN

GROSS GOURLEY, WESTINGHOUSE ELECTRIC CORPORATION

Resistance Brazing

Allen Cedilote, Industrial Testing Laboratory Services

Process Parameters

Workpiece material, workpiece size and shape, and joint design are the major factors that must be considered in developing a detailed RB procedure. The workpiece material is of primary importance because it has the greatest influence on the choice of electrode material and on the type of filler metal selected (see the Sections "Brazeability and Solderability of Engineering Materials" and "Procedure Development and Practical Considerations for Brazing and Soldering" in this Volume).

High-resistance workpieces should be brazed with low-resistance electrodes; conversely, high-conductivity (that is, low-resistance) workpieces must be brazed with low-conductivity electrodes. If at least one workpiece is made of a high-resistance material and if that workpiece is of sufficient mass, then the electrodes can be a high-conductivity material, such as a chromium-copper alloy (UNS C18200). When such an RB electrical circuit (which includes not only the transformer and electrodes but also at least one workpiece) is energized, heat will be generated within the high resistance circuit component of the circuit (that is, the workpiece). Conversely, if all workpieces have low electrical resistance, then the brazing current flowing through them will not generate enough heat to melt the filler metal. In this case, the electrodes must supply the necessary heat by virtue of their electrical resistance. Consequently, in resistance brazing high-conductivity workpieces, graphite (and in some cases, tungsten or molybdenum) electrodes are used. The heat generated in the electrodes flows into the workpieces by conduction or convection through the electrode-to-workpiece contact areas.

Assuming that the brazing procedure is fundamentally correct and capable of yielding properly bonded workpieces, there are a number of factors that strongly affect the quality of an RB joint. In descending order of priority, the factors that contribute to high quality in an RB joint are:

- CLEANLINESS OF THE WORKPIECES AND FILLER METAL SURFACES
- PROPER FIT BETWEEN THE WORKPIECES
- UNIFORMITY OF HEATING TO BRAZING TEMPERATURE
- MAINTAINING UNIFORM TEMPERATURE DURING BRAZING
- AVOIDING OVERHEATING
- MAINTAINING THE BRAZING TEMPERATURE FOR NO LONGER THAN SPECIFIED
- KEEPING THE PROPER LEVEL OF PRESSURE ON THE WORKPIECES DURING BRAZING

Each of these vital factors is influenced by the joint/workpiece design and the brazer's skill, experience, and ability to focus on the dynamic details of the process.

Workpiece and braze filler metal cleanliness is an absolute requirement for all types of brazing. In particular, cleanliness is of utmost importance in RB because in torch, furnace, and induction brazing the gaseous atmosphere around the workpieces can be controlled relatively easily to decompose surface oxides. Almost all common contaminants (grease, oil, perspiration, dust, abrasive particles, and oxides) are effective barriers to braze metal flow. At brazing temperature, most of these contaminants either vaporize or outgas, and because gases expand when heated, they can prevent bonding over a large area when confined within the braze joint.

Regardless of the electrical conductivity of the electrodes or workpieces, it is important to position and maintain the electrodes so as to sustain maximum electrode-to-workpiece contact through the entire brazing cycle. In addition, although it is preferable for the electrode-to-workpiece contact area to equal that of the braze joint in order to achieve even heating, the contact area should be no less than half the area of the braze joint.

For high-conductivity electrodes, maximum contact area is an optimum condition because:

- ELECTRODE-TO-WORKPIECE INTERFACE RESISTANCE IS MINIMIZED, THEREBY PROLONGING ELECTRODE LIFE AND ELIMINATING LOCALIZED HOTSPOTS ON THE WORKPIECE
- CURRENT FLOW INTO THE WORKPIECE WILL BE MORE EVENLY DISTRIBUTED, THUS HEATING THE WORKPIECE MORE UNIFORMLY

For low-conductivity electrodes, maximum contact area is even more critical because the workpieces are heated only by conduction from the electrode surfaces.

Because molten braze filler metal tends to flow to the hottest region in the joint, uniform heating is vitally important in achieving a uniform brazed joint. Therefore, it is necessary that the shape of the electrodes conform closely to the contour of the workpieces.

Pressure on the workpieces must be controlled carefully in order to achieve a properly bonded joint. Thermal expansion (and, commonly, differential thermal expansion in dissimilar joints) of the workpieces and electrodes must be allowed for. On one hand, too much pressure can squeeze molten filler metal out of the joint. Excessive pressure can cause the formation of a complete fillet around the joint perimeter while little or no bonding exists within. On the other hand, insufficient pressure can allow an unacceptably thick layer of filler metal to remain in the joint, which can result in a low-strength brazement that could fail prematurely in service.

Resistance Brazing

Allen Cedilote, Industrial Testing Laboratory Services

Equipment and System Selection

Resistance brazing may be classified as either manual or automatic, with manual resistance brazing being by far the most widely used.

In both manual and automatic RB, the kinds and models of transformers that provide the power are the same as those used for resistance welding. In many cases that involve high-resistance workpieces, the electrodes are also the same as those used in resistance welding applications. Welding transformers for RB can be purchased with automatic current/voltage controls, timers, electrode motion and pressure controls, and even automatic parts-handling devices, presuming that the production run is sufficient to warrant the added investment and a complexity of process optimization.

For manual RB, electrodes are usually incorporated into tongs made of chromium-copper, silver-tungsten, or copper-tungsten (for high-resistivity workpieces), and graphite, molybdenum, or tungsten (for low-resistivity workpieces). Graphite is the most common material for high-resistance electrodes because of its low cost and easy machinability. In addition, because it is not wetted by molten braze filler metals, there is no danger that graphite electrodes will be brazed to the workpieces. However, hot graphite is attacked by atmospheric oxygen during the brazing process. Thus, graphite electrodes must be inspected often and dressed as needed in order to maintain good workpiece contact. Special oxidation-resistant grades are available from graphite suppliers.

There is no standard design for manual RB tongs. Most RB organizations design and fabricate their own. As a result, there is an enormous variety of manual RB tongs and electrodes. However, the various types do contain some common elements; for example:

- WATER-COOLED, QUICK-RELEASE ELECTRODE CLAMPS (USUALLY MADE OF COPPER OR COPPER-CHROMIUM)
- WATER-COOLED CABLES FROM THE TRANSFORMER TO THE TONGS
- NONCONDUCTIVE HANDLES
- INSULATION OF THE ELECTRODES FROM EACH OTHER
- GROUND-FAULT INTERRUPTER PROTECTION

- INSULATING SHIELDS TO PREVENT INADVERTENT CONTACT WITH HOT OR ELECTRICALLY ENERGIZED AREAS

Resistance Brazing

Allen Cedilote, Industrial Testing Laboratory Services

Metal Electrodes

Selection of electrode material for RB depends on the electrical conductivity of the workpieces to be joined, the design and dimensions of the joint, appearance requirements for the brazed product, susceptibility of the workpieces to damage or marking, pressure needed on the work, production quantity of brazed joints, and cost of reconditioning or replacing an electrode.

The materials used most frequently in RB are RWMA class 2 (chromium-copper), RWMA class 14 (molybdenum), and various grades of carbon-graphite and graphite. Other standard and special electrode materials are sometimes used for special applications.

Class 2 (chromium-copper), a general-purpose resistance welding electrode material with a minimum electrical conductivity of about 75% IACS, is usually used to make electrode holders and shanks of faced electrodes for RB.

Refractory metal electrodes include class 14 (molybdenum), which is intermediate in mechanical and electrical properties in that classification, having a nominal electrical conductivity of 30% IACS. This material and RWMA class 13 (tungsten), because of resistance to high temperatures and nonsticking characteristics, are the only common electrode materials that have long life in resistance brazing of copper and other highly conductive nonferrous metals (i.e., aluminum and precious metals). They are about equal in electrical conductivity.

Class 14 (molybdenum) is generally preferred because class 13 (tungsten) is harder to machine and is likely to develop radial cracks in service. For reasons of economy, common practice is to use facings, buttons, or inserts of the refractory metal instead of making the entire electrode of the refractory metal.

Special electrode materials are sometimes needed to meet the unusual requirements of specific RB applications. An electrode made of 60% Pt and 40% Rh provides long life in the high-speed mass-production RB of insulated copper wires to copper terminal pads clad with BCuP-5. This electrode can withstand the heating used to melt the polyurethane insulation on the copper wire, the contact with molten BCuP-5 filler metal, and the heating in air used for burning off insulation residues after each brazed connection is made. This electrode material can be used at temperatures approaching 925 °C (1700 °F).

Carbon Electrodes

Ordinarily, two general types of carbon electrodes are used in RB: carbon-graphite and electrographite (artificial graphite). These electrode materials are made by simultaneously heating and blending the finely divided raw materials with coal tar pitch, which serves as a binder.

The stock from which RB electrodes are made comprises a few of the many grades of carbon stock that are manufactured primarily for use in melting ferrous and nonferrous metals. Composition and properties of the commercial carbon electrode materials vary; no generally accepted industry standards and terminology exist. The properties of five grades that are generally typical of the materials used in carbon electrodes for resistance brazing are given in Table 1.

TABLE 1 PROPERTIES OF FIVE CARBON ELECTRODE MATERIALS USED IN RESISTANCE BRAZING

GRADE	ELECTRODE MATERIAL	ELECTRICAL RESISTIVITY		SCLEROSCOPE HARDNESS	FLEXURAL STRENGTH (MINIMUM)		APPARENT DENSITY	
		μΩ· mm	Ω· in.		mpa	ksi	g/cm^3	$lb/in.^3$
1	CARBON-GRAPHITE, HARD(A)	20	0.00080	70	24	3.5	1.74	0.0629
2	CARBON-GRAPHITE, HARD, OXIDATION RESISTANT(B)	20	0.00080	70	24	3.5	1.75	0.0632
3	CARBON-GRAPHITE, SOFT	19	0.00075	40	16.5	2.4	1.57	0.0567
4	ELECTROGRAPHITE(A)	11	0.00042	50	17.2	2.5	1.73	0.0625
5	ELECTROGRAPHITE, OXIDATION RESISTANT(B)	11	0.00042	50	17.2	2.5	1.75	0.0632

(A) THIS TYPE OF CARBON ELECTRODE MATERIAL IS ALSO FREQUENTLY USED IN AIR CARBON ARC CUTTING.

(B) SIMILAR TO THE ELECTRODE MATERIAL LISTED IMMEDIATELY ABOVE, BUT IMPREGNATED WITH A SMALL PERCENTAGE OF AN OXIDATION RETARDANT, USUALLY AN INORGANIC COMPOUND CONTAINING BORON OR PHOSPHORUS, FOR LONGER LIFE

Carbon-Graphite. The first three types of electrode materials listed in Table 1 are called carbon-graphite. They are made from mixtures of finely ground petroleum coke (carbon) and natural or artificial (electro) graphite with coal tar pitch. These electrode materials are extruded in suitable shapes and heated in an electric furnace at about 815 °C (1500 °F) to develop the desired properties shown in Table 1. The heating converts about half of the pitch to carbon; the remainder is driven off as gases.

The final properties can be varied from those shown by changing the proportions of the raw materials, the particle size, and the time and temperature of heating. Increasing the ratio of graphite to carbon in the mixture, prolonging the heating time, or increasing the temperature lowers resistivity, hardness, and strength. Density is influenced chiefly by particle size.

Oxidation resistance for some grades of carbon-graphite is improved, for longer electrode life, by impregnating the cured material with a small percentage of an oxidation retardant, usually an inorganic compound that contains boron or phosphorus. In Table 1, oxidation-resistant grades 2 and 5 are the same as grades 1 and 4, respectively, except for the presence of an additive of this type.

Electrographite Electrode Material. Carbon-graphite grades and electrographite materials are produced in the same way except that the extruded mixture of raw materials is heated in "beehive" furnaces at about 2480 to 2705 °C (4500 to 4900 °F) for approximately four weeks, thus converting the mixture to graphite (called artificial graphite or electrographite). Properties of a representative material of this type and of an oxidation-resistant or long-life impregnated form of the same material are given in Table 1. Resistivity, hardness, and flexural strength of electrographite electrode materials are substantially lower than for the hard carbon-graphite electrode materials.

Selection of Carbon Electrode Material. The general types of carbon electrode material shown in Table 1 (carbon, graphite, and electrographite) have almost completely replaced the formerly used straight-carbon types, which are made by heating finely ground petroleum coke and a binder at about 815 °C (1500 °F). The straight-carbon grades typically have an electrical resistivity of about 51.2 μΩ· mm (0.0020 Ω· in.) and a Scleroscope hardness of about 100. Machining of these materials is much more difficult and costly than that of carbon-graphite and electrographite.

The hard carbon-graphite types are general-purpose materials and are preferred for most RB with carbon electrodes because brazing temperatures can be reached with less current than when the less-resistive electrographite types are used. Electrographite is preferred for RB of metals that have a high surface resistance, particularly when one of the metals being joined is steel or another iron-base alloy. Soft carbon-graphite material combines high heating capacity with a low tendency to produce local hot spots on the workpieces, but has comparatively low wear resistance. The first and fourth grades in Table l are the same types of material ordinarily used in air-carbon arc cutting.

In RB with carbon electrodes, nonuniform current flow and resultant local overheating of the electrodes can shorten electrode life excessively. To prevent this, carbon electrodes must be tightly fitted into matching tapered adapters or clamped securely to the electrode holder making contact with the holder over as large an area as possible. Provision also must be made for adequate flow of cooling water in the electrode holder. In addition, some carbon brazing electrodes are electroplated or sprayed with a copper coating about 0.05 mm (0.002 in.) in thickness to reduce contact resistance against the electrode holder and to minimize internal temperature buildup in the carbon electrode.

Design of Electrodes

Commercially available resistance seam welding electrodes are used where the design and dimensions of workpiece and joint permit. The electrode tip is machined, where necessary, to provide a tip shape and face dimensions suitable for the work. Carbon electrodes, depending on the size and design of the work, are in the shape of either standard metallic electrodes or flat blocks that may be several inches in length and width (size is usually limited by ability to obtain uniform heating and water cooling or by the current capacity of the power source).

When a high-resistance or long-wearing electrode is required, it is often used in the form of facings, inserts, or buttons that are attached to electrode shanks or bodies made of a less-expensive material. Electrode design is developed to work with workpiece and joint design in eliminating the need for holding and locating fixtures or clamps wherever possible and in permitting rapid and easy loading and unloading of workpieces, filler metal, and flux. Provision for water cooling of electrode holders and electrodes generally is the same as in resistance welding.

Filler Metals

Of the large number of filler metals available, only a few are used extensively in RB (Table 2). Selection of filler metal for RB is similar to that for other brazing processes and is discussed in the Section "Procedure Development and Practice Considerations for Brazing and Soldering" in this Volume.

TABLE 2 COMPOSITION AND THERMAL PROPERTIES OF SELECTED FILLER METALS USED FOR RESISTANCE BRAZING OPERATIONS

AWS CLASS	UNS NUMBER	COMPOSITION, WT%						LIQUIDUS		SOLIDUS	
		Ag	Cu	Zn	Cd	P	Sn	°C	°F	°C	°F
BAG-1	P07450	45	15	16	24	. . .	. . .	620	1145	605	1125
BAG-1A	P07500	50	15.5	16.5	18	. . .	. . .	635	1175	625	1160
BAG-2	P07350	35	26	21	18	. . .	. . .	700	1295	605	1125
BAG-7	P07563	56	22	17	. . .	. . .	5	650	1205	620	1145
BAG-8	P07720	72	28	. . .	. . .		. . .	780	1435	780	1435
BAG-18	P07600	60	30	. . .	. . .	. . .	10	720	1325	600	1115
BCUP-1	C55180	. . .	95	. . .	. . .	5	. . .	925	1695	710	1310
BCUP-2	C55181	. . .	97	. . .	. . .	3	. . .	795	1460	710	1310
BCUP-5	C55284	15	80	. . .	. . .	5	. . .	800	1475	645	1190

Source: Ref 1

More attention is given in RB to selecting compatible filler metals having the lowest brazing temperature, because in RB, the maximum local temperature reached by the work must be kept as low as possible, while providing uniform heating of the faying surfaces and the filler metal. Fluidity of the filler metal is not critical in most RB, because the filler metal is usually preplaced and the bond area is relatively large. The general types of filler metal usually selected for resistance brazing various classes of work metals are:

WORKPIECE METAL	**FILLER METAL ALLOYS**
STEEL, STAINLESS STEEL, HEAT-RESISTANT ALLOYS, COPPER, COPPER ALLOYS, NICKEL ALLOYS	SILVER (BAG TYPE)
ALUMINUM ALLOYS	AL-SI
COPPER AND COPPER ALLOYS	CU-P

These types of filler metal all have relatively low brazing temperatures (Table 3).

TABLE 3 SOLIDUS, LIQUIDUS, AND BRAZING TEMPERATURES FOR SELECTED RESISTANCE BRAZING FILLER METALS

FILLER METAL	**SOLIDUS**		**LIQUIDUS**		**BRAZING TEMPERATURE**	
	°C	**°F**	**°C**	**°F**	**°C**	**°F**
BCUP-2	710	1310	795	1460	730-845	1350-1550
BCUP-5	645	1190	800	1475	705-815	1300-1500
BAG-1A	625	1160	635	1175	635-760	1175-1400

Silver Alloys. Of the silver alloy (BAg type) group of filler metals, the two most often used are BAg-1 and BAg-1A, which are free-flowing alloys that permit the use of low brazing temperatures. In addition, their narrow melting ranges (-7 and -9 °C, or 20 and 15 °F, respectively) prevent liquation, making them insensitive to variations in the rate of heating or cooling. Their narrow melting ranges are also advantageous in step brazing. For making corrosion-resistantbrazed joints in stainless steel, BAg-3 and BAg-18 are preferred.

Aluminum-Silicon Alloys. Aluminum-silicon alloys (BAlSi type) are used in the RB of aluminum alloys. These low-melting alloys are available in sheet or wire form; BAlSi-2 and BAlSi-5 are also available as cladding on aluminum brazing sheet. Temperature must be controlled with special care in brazing of the lower melting brazeable aluminum alloys, such as wrought alloy 6151 (UNS A96151) and cast alloys 443.0 (UNS A04430) and 356.0 (UNS A03560), because the melting range of these filler metals approaches closely or overlaps the melting range of the work metal.

Copper-Phosphorus Alloys. Filler metals of copper-phosphorus alloys (BCuP type) are widely used for resistance brazing of copper and copper alloys. They are low-cost, general-purpose filler metals and have the special advantage of being self-fluxing on copper (by virtue of their phosphorus content), although flux is ordinarily needed when using them on copper alloys because phosphorus does not reduce metallic oxides other than those of copper. BCuP-5 is by far the most frequently used of the group.

In some applications on copper or copper alloys, the wider melting range, higher working temperature, and lower ductility of the BCuP filler metals make them less suitable than the BAg types. For example, in applications requiring higher ductility and corrosion resistance, silver brazing alloys are sometimes preferred to the BCuP filler metals for brazing of copper, despite the resulting need for fluxing. The difficulty in restricting flow of BCuP filler metals to joint surfaces may be a disadvantage.

Reference cited in this section

1. *BRAZING HANDBOOK,* 4TH ED., AMERICAN WELDING SOCIETY, 1991, P 239

Allen Cedilote, Industrial Testing Laboratory Services

Reference

1. *BRAZING HANDBOOK,* 4TH ED., AMERICAN WELDING SOCIETY, 1991, P 239

Diffusion Brazing

W. Daniel Kay, Wall Colmonoy Corporation

Introduction

DIFFUSION BRAZING (DFB) is a process that coalesces, or joins, metals by heating them to a suitable brazing temperature at which either a preplaced filler metal will melt and flow by capillary attraction or a liquid phase will form *in situ* between one faying surface and another. In either case, the filler metal diffuses into the base metal until the physical and mechanical properties of the joint become almost identical to those of the base metal. Pressure may or may not be applied to accomplish this.

"Diffusion bonding," a term that can refer to either diffusion brazing or diffusion welding, is now considered to be a nonstandard term. Where diffusion brazing is required, it should be clearly specified, However, in the aerospace industry, diffusion brazing is still commonly referred to by such tradenames as Activated Diffusion Bonding (ADB), Activated Diffusion Healing (ADH), and Transient Liquid-Phase Bonding (TLP). In each case, the process is actually diffusion brazing.

Two critical aspects of DFB are:

- A *LIQUID* FILLER METAL MUST BE FORMED AND BECOME ACTIVE IN THE JOINT AREA.
- EXTENSIVE DIFFUSION OF FILLER METAL ELEMENTS INTO THE BASE METAL MUST OCCUR.

This diffusion process often results in the total loss of identity of the original brazed joint (Fig. 1).

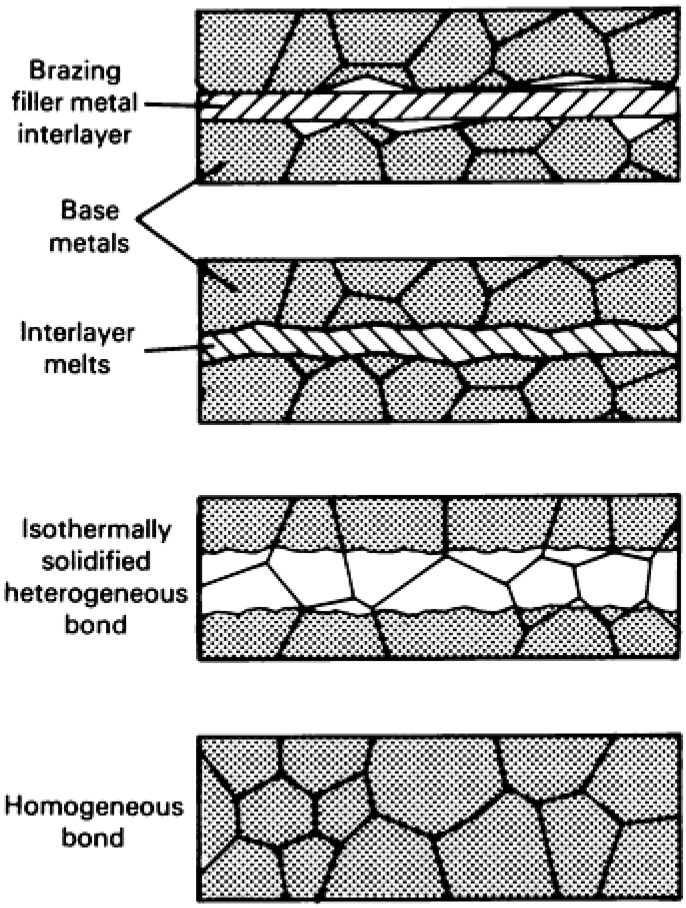
Brazing filler metal interlayer
Base metals
Interlayer melts
Isothermally solidified heterogeneous bond
Homogeneous bond

FIG. 1 DIFFUSION PROCESS RESULTING IN LOSS OF IDENTITY OF ORIGINAL BRAZED JOINT

Diffusion Brazing

W. Daniel Kay, Wall Colmonoy Corporation

Diffusion Brazing Processes

The most advanced diffusion brazing processes are used in the aerospace industry, particularly for brazements involving titanium, nickel, cobalt, and aluminum alloys. However, many similar and dissimilar metal combinations can be diffusion brazed, because diffusion is involved in all brazing processes, to a lesser or greater extent.

Although diffusion brazing is commonly performed in furnaces specifically set aside for brazing or heat treating, it is also possible to initially braze two components together by induction heating or by torch brazing, and then place the brazement into a furnace for the extended diffusion cycle. This cycle can range from 30 min to 80 h, or even longer.

As an example of a typical DFB process, a round bar of 304L stainless steel was diffusion brazed (end-to-end) to a round bar of pure copper in a furnace at 930 °C (1700 °F) for 60 min at temperature, using pure dry hydrogen as the atmosphere. The laying surface of the stainless steel was electrolytically plated with about 0.13 mm (0.005 in.) of pure silver. As this assembly was heated (without external pressure), the silver and copper diffused together to form a liquid eutectic filler metal alloy (72Ag-28Cu) in the butt joint at 780 °C (1435 °F), which metallurgically bonded the stainless steel and copper bars together permanently. When the brazement was subsequently heated in another furnace at 1040 °C (1900 °F), with a weight attached to the copper side of the part (in an effort to remelt the braze joint and take the brazement apart), it was found that the copper elongated until the pure copper portion of the part, rather than the brazed joint, failed. Diffusion had been complete enough that the remelt temperature of the brazement approached that of pure copper.

This example shows that, once the liquid has been distributed throughout the joint area, the assembly should be held at the brazing temperature long enough for one or more of the filler metal elements to adequately diffuse into the base metals (and vice versa). In this way, the mechanical properties of the brazed joint area become essentially identical to those of either or both of the base metals.

The amount of diffusion that occurs will be a function of:

- THE BRAZING TEMPERATURE
- THE LENGTH OF TIME THE PART IS HELD AT TEMPERATURE
- THE QUANTITY OF FILLER METAL AVAILABLE FOR DIFFUSION
- THE MUTUAL SOLUBILITY OF FILLER METAL AND THE BASE METALS

A fully diffused joint will lose all joint identity because of base-metal grain growth across the entire joint (Fig. 2).

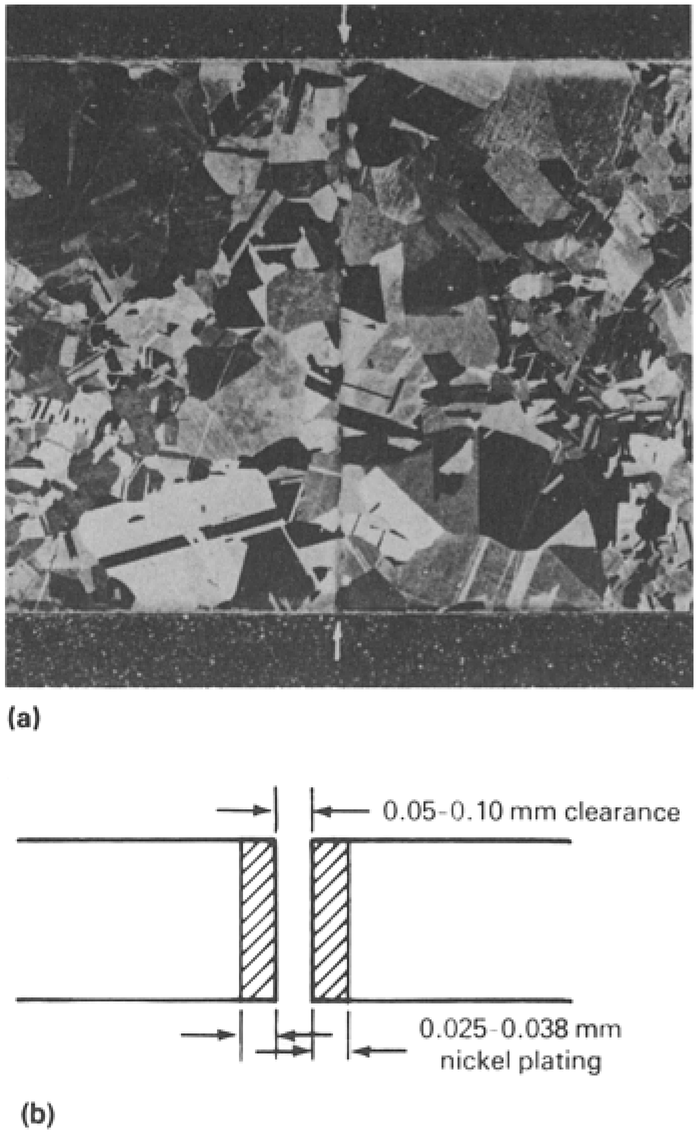

FIG. 2 DFB JOINT ILLUSTRATING HOW THE BRAZING FILLER METAL AND NICKEL PLATE DIFFUSED INTO THE BASE METAL, ALMOST OBLITERATING THE JOINT. SPECIMEN: NICKEL-PLATED NIMONIC 80A 6.4 MM (0.252 IN.) TENSILE TEST BAR MACHINED FROM RECTANGULAR BRAZED BLOCKS. BRAZING PROCEDURE: 30 MIN AT 1175 °C (2150 °F); NICROBRAZ 125 BRAZING FILLER METAL; AGED AFTER BRAZING. (A) MACROSTRUCTURE OF JOINT. 8×. (B) SCHEMATIC OF JOINT

When a brazement is to be made using either a preplaced sheet of filler metal or a layer of filler metal powder, it is wise to apply pressure to the assembly so that, when filler metal becomes molten, the gap between the two base metals is reduced. As an example of this process, a cobalt-base alloy was diffusion brazed at 1120 °C (2050 °F) in a vacuum furnace using a nickel-chromium-boron filler material in sheet form. The assembly was held at temperature for 8 h. Even though the solidus of the filler metal was only 970 °C (1780 °F), the braze joint completely lost its identity through diffusion, and the resulting remelt temperature was above 1370 °C (2500 °F). In this case, pressure was applied using simple weights on top of the part. Because the filler metal contained boron as a melting-point depressant, the diffusion of the boron out of the filler metal into the base metal caused the remelt temperature of the filler metal to rise. Because boron is also a very small atom, it very readily diffuses (as an interstitial atom) away from the joint as the temperature of the base metal rises.

Diffusion brazing is very advantageous for the effective joining of aerospace components that are intended for elevated-temperature service, such as jet engine hot-section parts. It is critical that they be able to perform well and safely at operating temperatures from 1040 to 1150 °C (1900 to 2100 °F). These objectives can be readily achieved when diffusion brazing is properly understood and used.

Diffusion Brazing

W. Daniel Kay, Wall Colmonoy Corporation

Selected References

- I.M. BARTA, LOW TEMPERATURE DIFFUSION BONDING OF ALUMINUM ALLOYS, *WELD. J.,* VOL 43 (NO. 6), JUNE 1964, P 241S-247S
- P.M. BARTLE, DIFFUSION BONDING: A LOOK AT THE FUTURE, *WELD. J.,* VOL 54 (NO. 11), NOV 1975, P 799-804
- G.F. BLANK, A PRACTICAL GUIDE TO DIFFUSION BONDING, *MATER. DES. ENG.,* OCT 1966, P 76-79
- C.L. CLINE, AN ANALYTICAL AND EXPERIMENTAL STUDY OF DIFFUSION BONDING, *WELD. J.,* VOL 45 (NO. 11), NOV 1966, P 481S-489S
- W.A. DEMO AND S.J. FERRIGNO, BRAZING METHOD HELPS REPAIR AIRCRAFT GAS-TURBINE NOZZLES, *ADV. MATER. PROCESS.,* MARCH 1992, P 43-45
- D.S. DUVALL, W.A. OWCZARSKI, AND D.F. PANIONIS, TLP BONDING: A NEW METHOD FOR JOINING HEAT RESISTANT ALLOYS, *WELD. J.,* VOL 53 (NO. 4), APRIL 1974, P 203-214
- A.H. FREEDMAN, BASIC PROPERTIES OF THIN-FILMED DIFFUSION BRAZED JOINTS IN TI-6AL-4V, *WELD. J.,* VOL 50 (NO. 8), AUG 1971, P 343S-356S
- G.S. HOPPIN III AND T.F. BERRY, ACTIVATED DIFFUSION BONDING, *WELD. J.,* VOL 49 (NO.11), NOV 1970, P 505S-509S
- S. LIU, D.L. OLSON, G.P. MARTINS, AND G.R. EDWARDS, THE USE OF COATING AND INTERLAYER TECHNOLOGY IN BRAZING, *WELD. J.,* VOL 70 (NO. 8), AUG 1991, P 207S-215S
- J.T. NIEMANN, R.P. SOPHER, AND P.J. RIEPPEL, DIFFUSION BONDING BELOW 1000 °F, *WELD. J.,* VOL 37 (NO. 8), AUG 1958, P 337S-342S
- R.L. PEASLEE, DIFFUSION BRAZING, *WELD. J.,* VOL 55 (NO. 8), AUG 1976, P 695-696
- J.E. RAMIREZ AND S. LIU, DIFFUSION BRAZING IN THE NICKEL-BORON SYSTEM, *WELD. J.,* VOL 71 (NO. 10), OCT 1992, P 365S-375S
- M.M. SCHWARTZ, DIFFUSION BRAZING TITANIUM SANDWICH STRUCTURES, *WELD. J.,* VOL 57 (NO.9), SEPT 1978, P 35-38
- R.F. SEKERKA, ON THE MODELING OF DIFFUSION BONDING, *PROC. PHYSICAL METALLURGY,* TMS-AIME, 1980
- R.R. WELLS, MICROSTRUCTURAL CONTROL OF THIN-FILM DIFFUSION-BRAZED TITANIUM,

WELD. J., VOL 55 (NO. 1), JAN 1976, P 20S-28S

Exothermic Brazing

Donald W. Bucholz, IBM Federal Systems Corporation

Introduction

EXOTHERMIC BRAZING (EXB) is a process that utilizes the heat produced in a solid-state chemical reaction to melt a conventional filler metal or to produce molten filler metal as a product of the reaction.

When used to melt a conventional filler metal, exothermic brazing is similar to thermite welding (see the article "Thermite Welding" in this Volume) in that the heat required is derived from the chemical reaction, but it is different from thermite welding because only the heat is used. The brazing filler metal is not a component of the exothermic heat source and therefore does not include contaminants from the reaction.

When the EXB process is used to produce molten filler metal as a product of the reaction, an exothermic reaction occurs between two or more melting-point-depressant elements, which react to form a compound with a higher melting point. The result is an intermetallic compound that forms a heterogeneous filler metal. During this type of reaction, surface oxides are either displaced or disintegrated, which can eliminate the need for protective atmospheres or fluxes (Ref 1).

Reference

1. N. BREDSZ *ET AL.,* U.S. PATENT 3,415,697, 10 DEC 1968

Exothermic Brazing

Donald W. Bucholz, IBM Federal Systems Corporation

Process Advantages and Disadvantages

Exothermic brazing is most frequently used to make brazed joints in remote locations, inaccessible areas, physically complex structures, or situations that require the application of localized heat for a brief period of time. The basic advantages of this process are:

- RAPID CYCLE TIME (20 TO 30 S)
- LOCALIZED APPLICATION OF HEAT
- SIMPLICITY AND PORTABILITY OF EQUIPMENT
- PREPLACEMENT OF FILLER METAL

Some disadvantages of the EXB process are:

- DIFFERING JOINT DESIGNS REQUIRE EXPERIMENTATION PRIOR TO ACTUAL PRODUCTION IN ORDER TO EFFECTIVELY BRAZE EXOTHERMICALLY
- SAFETY ISSUES THAT CONCERN HANDLING AND APPLICATION MUST BE THOROUGHLY EVALUATED WHEN USING IGNITABLE, HIGH-HEAT-PRODUCING MATERIALS
- FILLER METAL SELECTION IS LIMITED BY THE THERMODYNAMIC BEHAVIOR OF THE EXOTHERMIC MATERIAL

Exothermic Brazing

Donald W. Bucholz, IBM Federal Systems Corporation

Types of Compounds

Exothermic compounds that generate enough heat to melt conventional filler metals have been developed for iron alloys, copper alloys, and refractory metals. Aluminum compounds that react and produce a heterogeneous filler metal can be used to join aluminum alloys.

The aerospace industry has utilized exothermic brazing to replace hydraulic fittings in corrosion-resistant steel tubing (AM-350 CRT). An example of this application using a commercial filler metal and an exothermic compound to braze a 25.4 mm (1 in.) outside diameter stainless steel tube with a 1 mm (0.04 in.) wall thickness has shown that a satisfactory braze joint can be made (Ref 2, 3). The exothermic compound in that example consisted of 30 g (1 oz) of 19.90% Mg, 51.84% NiO, 18.24% MnO, 4.24% Al, and 5.36% MnO_2.

Refractory metals exothermically brazed at approximately 1760 °C (3200 °F) with conventional filler metals, using either 24.9% B and 75.1% V_2O_5 or 19.8% B, 60.2% V_2O_5, 16.5% Ta_2O_5, and 6.8% Al as the exothermic compound, have shown minimal base metal recrystallization and braze filler metal/base metal interdiffusion (Ref 4).

Aluminum alloys have been brazed utilizing an exothermic compound consisting of powdered mixtures of: 3.5 g (0.12 oz) of 50% Al and 50% Mg; 10.0 g (0.35 oz) of 88% Al and 12% Si; and 13.5 g (0.48 oz) Al. The resulting intermetallic compound that forms the filler metal is composed of 27 g (0.95 oz) of 89.1% Al, 10.2% Mg_2Si, and 0.7% Si. More than a dozen of these types of exothermic brazing filler metals used with aluminum alloys are described in Ref 1 and 5.

References cited in this section

1. N. BREDSZ *ET AL.*, U.S. PATENT 3,415,697, 10 DEC 1968
2. R.A. LONG, EXOTHERMICALLY BRAZED HYDRAULIC FITTINGS FOR AIRCRAFT, *PROC. AWS NATIONAL FALL MEETING*, AWS, OCT 1966, P 28-38
3. R.A. LONG *ET AL.*, U.S. PATENT 3,308,532, 14 MARCH 1967
4. R.A. LONG, EXOTHERMIC BRAZING OF REFRACTORY METALS, *PROC. METALL. SOC. CONF.*, AIME, 1966, P 249-258,
5. W.V. KNOPP *ET AL.*, U.S. PATENT 3,899,306, 12 AUG 1975

Exothermic Brazing

Donald W. Bucholz, IBM Federal Systems Corporation

Equipment

The type of equipment used for exothermic brazing is shown in Fig. 1. The principal components are the joint itself, the brazing filler metal (if it is not a component of the exothermic compound), the exothermic compound, the igniter, the tubing sleeve insulation, and the container. Variations that utilize different joint designs, shielding gases, and external sources for preheating can also be used.

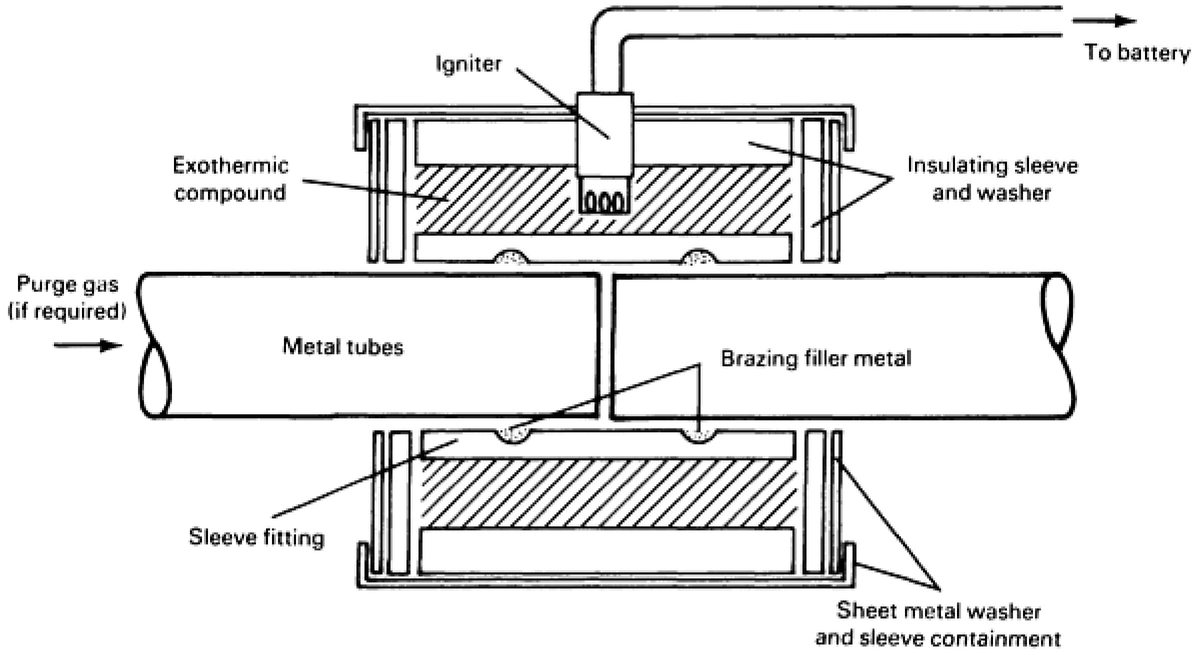

FIG. 1 TYPICAL ARRANGEMENT FOR EXOTHERMIC BRAZING OF TUBE. SOURCE: REF 2

A moderate amount of technical skill and knowledge of the fundamental principles, as well as empirical experience for the particular joining situation, are needed in order to make high-quality brazed joints. In addition to typical joint inspection methods and braze filler metal quality control requirements, exothermic brazing also can require lot control for exothermic compounds to ensure uniformity of performance.

Reference cited in this section

2. R.A. LONG, EXOTHERMICALLY BRAZED HYDRAULIC FITTINGS FOR AIRCRAFT, *PROC. AWS NATIONAL FALL MEETING,* AWS, OCT 1966, P 28-38

Exothermic Brazing

Donald W. Bucholz, IBM Federal Systems Corporation

Procedure

When performing exothermic brazing, the standard steps that are followed in any other brazing process apply, such as joint design, alloy selection, and surface preparation. However, additional parameters must be considered: selection of exothermic compound, required quantity of exothermic compound, and brazing assembly configuration.

The typical effects of joint mass on the heat-up time and the mass of exothermic material that is required to achieve a specific temperature are shown in Fig. 2 and 3, respectively. The exothermic material is usually placed around or over the joint to be made, and an electric igniter is embedded in the exothermic compound. The assembly is typically insulated with ceramic or ceramic fiber and then covered and held in place with metal sheet or foil. The electrical leads of the igniter should be connected to a low-voltage battery. Once the assembly has been prepared, the battery and igniter circuit is closed, and the process runs to completion.

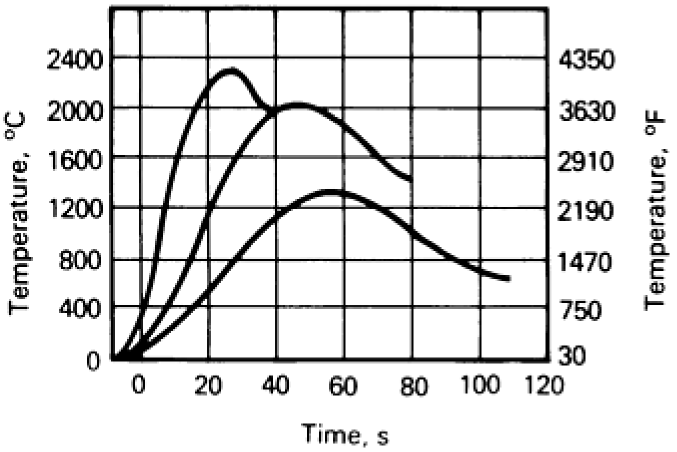

FIG. 2 HEATING VERSUS TIME CHARACTERISTICS FOR EXOTHERMIC BRAZING. LARGER MASSES REQUIRE LONGER TIMES TO HEAT UP AND ACHIEVE LOWER MAXIMUM TEMPERATURES FOR A FIXED AMOUNT OF EXOTHERMIC COMPOUND. SOURCE: REF 2, 6

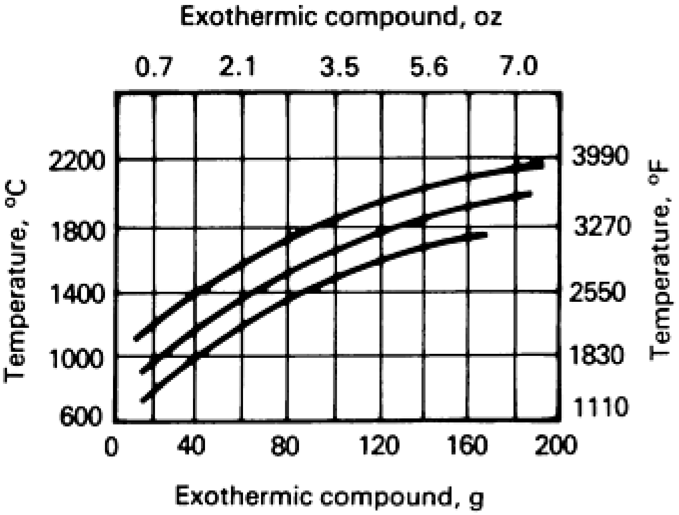

FIG. 3 MAXIMUM TEMPERATURE VERSUS MASS OF EXOTHERMIC COMPOUND (CHARACTERISTIC CURVE). A PARTICULAR EXOTHERMIC COMPOUND WILL PRODUCE A SINGLE CHARACTERISTIC CURVE FOR A GIVEN JOINT MASS AND CONFIGURATION. SOURCE: REF 6

References cited in this section

2. R.A. LONG, EXOTHERMICALLY BRAZED HYDRAULIC FITTINGS FOR AIRCRAFT, *PROC. AWS NATIONAL FALL MEETING,* AWS, OCT 1966, P 28-38
6. I.P. CHEKUNOV *ET AL.*, EXOTHERMAL BRAZING OF STEEL PIPING WITH HIGH TEMPERATURE BRAZING ALLOYS, *WELD. PROD.* (ENGLISH TRANSLATION OF *SVAR. PROIZVOD.*), NOV 1972, P 39-42

Exothermic Brazing

Donald W. Bucholz, IBM Federal Systems Corporation

References

1. N. BREDSZ *ET AL.,* U.S. PATENT 3,415,697, 10 DEC 1968
2. R.A. LONG, EXOTHERMICALLY BRAZED HYDRAULIC FITTINGS FOR AIRCRAFT, *PROC. AWS NATIONAL FALL MEETING,* AWS, OCT 1966, P 28-38
3. R.A. LONG *ET AL.,* U.S. PATENT 3,308,532, 14 MARCH 1967
4. R.A. LONG, EXOTHERMIC BRAZING OF REFRACTORY METALS, *PROC. METALL. SOC. CONF.,* AIME, 1966, P 249-258,
5. W.V. KNOPP *ET AL.,* U.S. PATENT 3,899,306, 12 AUG 1975
6. I.P. CHEKUNOV *ET AL.,* EXOTHERMAL BRAZING OF STEEL PIPING WITH HIGH TEMPERATURE BRAZING ALLOYS, *WELD. PROD.* (ENGLISH TRANSLATION OF *SVAR. PROIZVOD.*), NOV 1972, P 39-42

Exothermic Brazing

Donald W. Bucholz, IBM Federal Systems Corporation

Selected References

- R. BECK *ET AL.,* "ASSESSMENT OF THE RESULTS OF THE SKYLAB SPACE PROCESSING EXPERIMENTS WITH RESPECT TO THEIR SCIENTIFIC-TECHNICAL SIGNIFICANCE," FINAL REPORT NO. ESA-CR(P)-623, EUROPEAN SPACE RESEARCH ORGANISATION, OCT 1974
- "EXOTHERMIC BRAZING SPECIMENS", FINAL REPORT NO. NASA-CR-103169, WHITAKER CORPORATION, SAN DIEGO RESEARCH AND DEVELOPMENT DIVISION, DEC 1970
- R.W. HEINE *ET AL.,* "FLIGHT/GROUND SAMPLE COMPARISON RELATING TO FLIGHT EXPERIMENT M552, EXOTHERMIC BRAZING," REPORT NO. NASA-CR-12059, UNIVERSITY OF WISCONSIN, DEPARTMENT OF METALLURGY AND MINERAL ENGINEERING, PREPARED FOR GEORGE C. MARSHALL SPACE FLIGHT CENTER, DEC 1973
- R.A. LONG, U.S. PATENT 3,358,356, 19 DEC 1967
- R.A. LONG, U.S. PATENT 3,287,190, 22 NOV 1966
- H.E. PATTEE AND R.E. MONROE, "PHASE B REPORT ON MATERIALS PROCESSING IN SPACE M512," REPORT NO. NASA-CR-120266, BATTELLE COLUMBUS LABS, PREPARED FOR GEORGE C. MARSHALL SPACE FLIGHT CENTER, JULY 1973
- H.E. PATTEE AND R.E. MONROE, "CHARACTERIZATION OF EXOTHERMIC BRAZING COMPONENTS, SKYLAB EXPERIMENT M552," REPORT NO. NASA-CR-120518, BATTELLE COLUMBUS LABS, PREPARED FOR GEORGE C. MARSHALL SPACE FLIGHT CENTER, DEC 1973

Exothermic Brazing

Donald W. Bucholz, IBM Federal Systems Corporation

Selected References

- R. BECK *ET AL.,* "ASSESSMENT OF THE RESULTS OF THE SKYLAB SPACE PROCESSING EXPERIMENTS WITH RESPECT TO THEIR SCIENTIFIC-TECHNICAL SIGNIFICANCE," FINAL REPORT NO. ESA-CR(P)-623, EUROPEAN SPACE RESEARCH ORGANISATION, OCT 1974
- "EXOTHERMIC BRAZING SPECIMENS", FINAL REPORT NO. NASA-CR-103169, WHITAKER

CORPORATION, SAN DIEGO RESEARCH AND DEVELOPMENT DIVISION, DEC 1970

- R.W. HEINE *ET AL.*, "FLIGHT/GROUND SAMPLE COMPARISON RELATING TO FLIGHT EXPERIMENT M552, EXOTHERMIC BRAZING," REPORT NO. NASA-CR-12059, UNIVERSITY OF WISCONSIN, DEPARTMENT OF METALLURGY AND MINERAL ENGINEERING, PREPARED FOR GEORGE C. MARSHALL SPACE FLIGHT CENTER, DEC 1973
- R.A. LONG, U.S. PATENT 3,358,356, 19 DEC 1967
- R.A. LONG, U.S. PATENT 3,287,190, 22 NOV 1966
- H.E. PATTEE AND R.E. MONROE, "PHASE B REPORT ON MATERIALS PROCESSING IN SPACE M512," REPORT NO. NASA-CR-120266, BATTELLE COLUMBUS LABS, PREPARED FOR GEORGE C. MARSHALL SPACE FLIGHT CENTER, JULY 1973
- H.E. PATTEE AND R.E. MONROE, "CHARACTERIZATION OF EXOTHERMIC BRAZING COMPONENTS, SKYLAB EXPERIMENT M552," REPORT NO. NASA-CR-120518, BATTELLE COLUMBUS LABS, PREPARED FOR GEORGE C. MARSHALL SPACE FLIGHT CENTER, DEC 1973

Brazing With Clad Brazing Materials

Sunil Jha, Michael Karavolis, Kevin Dunn, and James Forster, Texas Instruments, Inc.

Introduction

BRAZING is the preferred method for manufacturing complex assemblies such as heat exchangers, which contain a large number of joints (Ref 1). In the conventional plate-and-fin type of heat exchanger, the fins are brazed to the plates. Often, the brazing filler metal, which can be in the form of a foil, a paste, or a powder that has been consolidated into a sheet using a polymer carrier, is inserted between the plate and fin, and a multiple stack of plate, brazing foil, and fin is fixtured for brazing in a furnace. This complex assembly requires the precise fit of multiple components and the use of a flux, which allows the filler metal to "wet" the surfaces to be joined. To reduce part count and thereby simplify the assembly operation, the brazing filler metal can be clad to the plate material by a roll bonding technique.

This article focuses on clad brazing material, which is defined as any base material or alloy that is clad with an appropriate lower-melting-point brazing filler metal. The most common example is the aluminum-braze clad sheet that is used in the fabrication of aluminum heat exchangers (Ref 2). In these sheets, the aluminum brazing filler alloy (BAlSi series) is clad to the aluminum-base alloy by hot roll bonding. The aluminum-clad brazing sheets can be brazed in a variety of atmospheres, and are used in numerous applications.

Another commercial example of a clad brazing material is copper-clad stainless steel sheet, which is used in the fabrication of heat exchangers for automotive, electronic, and other applications. The copper layer acts as the brazing filler material. The clad brazing material can be formed by drawing, bending, stamping, or other metalworking processes. The formed parts can be assembled and brazed without using any additional filler metal in the joint. Clad brazing materials are frequently used as one member of an assembly, in which the mating piece is made from a brazeable alloy, so that the clad filler metal flows by capillarity to the joint area.

On a more limited scale, aluminum-clad titanium strips have been used to fabricate titanium heat exchangers for electronics applications. In addition, silver brazing alloy (BAg3) that is clad on copper is often used in the brazing of carbide tools (Ref 3). The copper core remains in the joint area and acts as a ductile member that plastically deforms during cooling of the brazed joint.

References

1. T. PROFUGHI, BRAZING: THE SOLUTION FOR MANY COMPLEX ASSEMBLY PROBLEMS, *HEAT TREAT.*, OCT 1990, P 16-17
2. *BRAZING MANUAL*, 3RD ED., AWS, 1984, P 150

3. *BRAZING MANUAL*, 3RD ED., AWS, 1984, P 236

Brazing With Clad Brazing Materials

Sunil Jha, Michael Karavolis, Kevin Dunn, and James Forster, Texas Instruments, Inc.

Cladding Materials

Clad brazing materials are produced as strips, using the roll bonding technique. The strips comprise a base metal that is clad with a brazing filler metal on either one or both sides. These products are used primarily in high-volume manufacturing operations, such as the production of heat exchangers, brazed bellows, and honeycomb structures. The use of a self-brazing sheet reduces the total part count, simplifies the assembly operation (because the brazing filler metal is always present on the core material), and reduces assembly time and, therefore, cost. In addition, there is no need for the application of flux, nor for its subsequent removal. This not only saves the initial purchase cost of the flux, but the waste-management cost that is associated with the disposal of the spent material.

Aluminum-braze clad products, copper-clad steel, and copper-clad stainless steel are the most commonly used clad brazing materials. Table 1 lists the typical clad brazing strip products that are used in the fabrication of compact heat exchangers for the automotive, truck, and aircraft industries.

TABLE 1 TYPICAL CLAD BRAZING STRIP PRODUCTS

MATERIAL	CLAD LAYER RATIOS	THICKNESS		TEMPER	TENSILE STRENGTH		0.2% YIELD STRENGTH		ELONGATION, %
		mm	in.		mpa	ksi	mpa	ksi	
C12200 COPPER CLAD TO 409 STAINLESS STEEL	10/80/10 5/80/15	0.51-0.76	0.020-0.030	ANNEALED	400	58	230	33	37
C12200 COPPER CLAD TO 304 L STAINLESS STEEL	13.5/86.5	0.51-0.30	0.020-0.012	ANNEALED	635	92	290	42	56
	10/80/10			ANNEALED	620	90	275	40	55
C12200 COPPER CLAD TO 1008 STEEL	10/80/10	0.38	0.015	NO. 4 TEMPER	380	55	290	42	35
C52400 PHOSPHOROUS BRONZE CLAD TO C10200	10/80/10	0.51	0.020	ANNEALED	275	40	97	14	48

The thickness of the brazing filler metal used in a clad brazing material can range from 0.050 to 2.54 mm (0.002 to 0.100 in.) and will represent 5 to 20% of the total strip thickness. This range of strip thickness covers the majority of applications in the heat-exchanger industry. Figure 1 shows a typical clad brazing strip of copper-clad stainless steel.

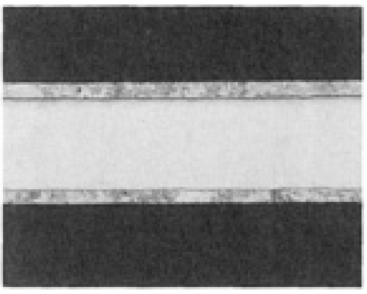

FIG. 1 PHOTOMICROGRAPH OF TYPICAL CLAD BRAZING MATERIAL, C12200 COPPER CLAD TO 304 L STAINLESS STEEL

Figure 2 compares the steps in using brazing preforms to fabricate a brazed assembly with the steps involved in using clad brazing materials. The latter is shown to be more efficient.

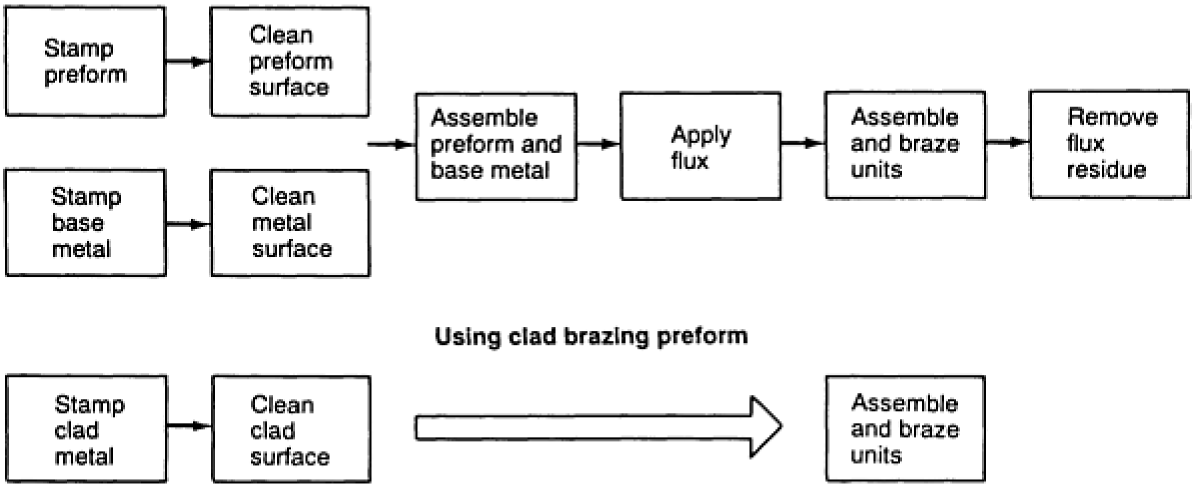
Stamp preform
Clean preform surface
Stamp base metal
Clean metal surface
Assemble preform and base metal
Apply flux
Assemble and braze units
Remove flux residue
Using clad brazing preform
Stamp clad metal
Clean clad surface
Assemble and braze units

FIG. 2 COMPARISON OF BRAZING PROCESS USING BRAZING PREFORM VERSUS PROCESS USING CLAD BRAZING SHEET

Figure 3 depicts an automotive transmission fluid cooler that was assembled using clad brazing materials. A turbulator is brazed to a copper-clad stainless steel base and cover. The base and cover are formed from a stainless steel strip containing copper braze on one side. After brazing, the dimensional changes in this part are minimal, which is important when making a hermetically sealed heat exchanger.

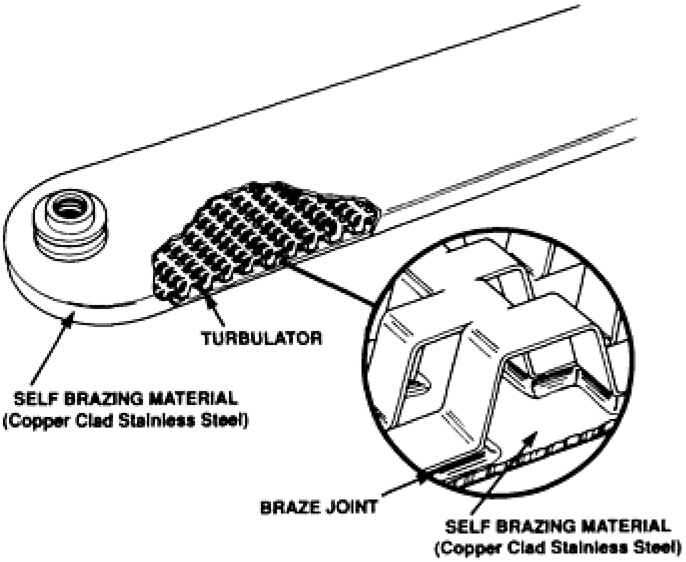

FIG. 3 HEAT EXCHANGER FABRICATED USING CLAD BRAZING ("SELF-BRAZING") MATERIALS

To fabricate clad brazing materials, the brazing filler metal must be ductile enough to be roll bonded to the base metal. The roll bonding technique is described in the article "Roll Welding" in this Volume and in Ref 4 and 5. A nickel brazing alloy that contains high levels of boron and silicon as melting-point depressants, such as the BNi-W series, is not available in thin-strip form, because these materials are brittle. Therefore, it cannot be roll bonded or processed by conventional metalworking methods.

The advantages of using clad brazing materials are that they:

- REDUCE THE TOTAL NUMBER OF PARTS THAT NEED TO BE ASSEMBLED
- ELIMINATE THE STAMPING AND ASSEMBLY COSTS ASSOCIATED WITH THE PRODUCTION OF A BRAZING FILLER METAL PREFORM
- ELIMINATE THE USE OF FLUXES AND REDUCING OR SURFACE-ACTIVATING ATMOSPHERES, BECAUSE ROLL BONDING ALLOWS THE BRAZING FILLER METAL TO METALLURGICALLY BOND TO THE SURFACES TO BE JOINED
- IMPROVE THE MANUFACTURING YIELD, BECAUSE PRECISE POSITIONING OF THE FILLER METAL IN THE JOINT AREA IS POSSIBLE
- ENHANCE DESIGN FLEXIBILITY, BECAUSE THE BRAZING FILLER METAL IS INTEGRAL WITH THE STRUCTURAL COMPONENTS THROUGH ALL MANUFACTURING STEPS, WHICH PERMITS THE DESIGN OF COMPACT, MORE COMPLEX, AND COST-EFFECTIVE ASSEMBLIES
- IMPROVE PROCESS REPEATABILITY, BECAUSE OF THE UNIFORMITY OF BRAZING FILLER METAL AND TIGHTER DESIGN TOLERANCES
- ELIMINATE HANDLING PROBLEMS WITH THIN AND WEAK FILLER METAL SHIMS, FOILS, OR PASTES

References cited in this section

4. G. DURST, A NEW DEVELOPMENT IN METAL CLADDING, *J. MET.*, MARCH 1956
5. R.G. DELAGI, DESIGNING WITH CLAD METALS, *MACH. DES.*, NOV 1980

Sunil Jha, Michael Karavolis, Kevin Dunn, and James Forster, Texas Instruments, Inc.

Brazing With Clad Brazing Materials

The design and manufacturing considerations associated with obtaining a good joint when using a clad brazing material are similar to those associated with producing joints when using conventional brazing techniques (Ref 6). Because the brazing filler metal is metallurgically bonded to the base metal, it acts as a protective (oxide-inhibiting) layer. This somewhat reduces the requirement for cleaning and fixturing before brazing, because the total part count is reduced. Similarly, the use of clad brazing materials often imposes a less-stringent requirement on the furnace atmosphere.

In the case of copper-clad stainless steel, the copper protects the surface of the stainless steel during heating. At the brazing temperature, it readily melts and flows to form the braze joint. Therefore, self-brazing materials often eliminate the need for pretreatments and fluxes to ensure adequate flow and wetting.

Clad brazing materials should not be used solely as a substitute material or as a joining method in an existing design that is based on conventional assembly techniques and placement of the filler metal. The greatest benefits of clad brazing materials are obtained when their use is comprehended in the basic design of the assembly and in the manufacturing process. The decision to use clad brazing material in the design stage of a heat exchanger allows greater design flexibility, tighter tolerances, and a more-robust process, because the design engineer does not need to allow spaces for wires or shims, or to specify the application method and handling procedure for the brazing filler metal. During the high-volume production of brazed assemblies that utilize clad brazing materials, the same amount of brazing filler metal is located at any given location every time.

Reference cited in this section

6. *THE BRAZING BOOK,* HANDY AND HARMAN, 1985

Brazing With Clad Brazing Materials

Sunil Jha, Michael Karavolis, Kevin Dunn, and James Forster, Texas Instruments, Inc.

References

1. T. PROFUGHI, BRAZING: THE SOLUTION FOR MANY COMPLEX ASSEMBLY PROBLEMS, *HEAT TREAT.,* OCT 1990, P 16-17
2. *BRAZING MANUAL,* 3RD ED., AWS, 1984, P 150
3. *BRAZING MANUAL,* 3RD ED., AWS, 1984, P 236
4. G. DURST, A NEW DEVELOPMENT IN METAL CLADDING, *J. MET.,* MARCH 1956
5. R.G. DELAGI, DESIGNING WITH CLAD METALS, *MACH. DES.,* NOV 1980
6. *THE BRAZING BOOK,* HANDY AND HARMAN, 1985

Iron Soldering

Mark Cowell, Metcal, Inc.

Equipment

Hand soldering equipment can be broadly classified by its temperature control method. There are constant-voltage, variable temperature, and tip-temperature-controlled soldering iron types, each of which is described below.

Constant-voltage soldering irons are temperature-limited, but do not incorporate a temperature sensor. Temperature calibration of the tip is not possible, because these irons depend on the increasing resistance of the heating element to limit tip temperature. Line voltage fluctuations and ambient temperatures greatly affect the temperature of the soldering tip. A constant-voltage soldering iron is the least-expensive hand soldering tool. However, this type of soldering iron requires highly skilled operators to create reliable joints.

Variable-temperature soldering irons incorporate a temperature control device in the soldering handle. A control knob on the power source provides a means for the user to vary the tip temperature. When the soldering tip achieves the prescribed temperature, current to the heater is turned off. The tip cools to the lower temperature setpoint of the controller, and current to the heater is turned back on. Tip temperature calibration is possible with variable-temperature soldering irons. This type of iron can be adjusted so that its temperature is balanced with the thermal requirements of the joint. However, one of its drawbacks is that an unskilled or inattentive operator may leave the temperature at the highest setting, in an attempt to solder more quickly, damaging heat-sensitive components or printed circuit boards. Damage can also occur when the heat is too low and an operator stays on the termination too long while adding pressure.

Temperature-controlled irons incorporate a temperature sensor in the soldering tip. These soldering irons operate at a manufacturer's preselected temperature, and tip temperature calibration is not required. Tips are selected for their shape, as well as their operating temperatures. At the preselected temperature, either the resistance of the heating element changes or the current supplied to the heating element is turned off. The resulting reduction in heat generation stabilizes the tip temperature at its setpoint. Temperature-controlled soldering tips help ensure that the proper tip is used in a production environment, thus eliminating costly damage to printed circuit boards and components. This tool requires the least amount of operator training.

Iron Soldering

Mark Cowell, Metcal, Inc.

Soldering Iron Selection

When selecting a soldering iron, potential damage to the printed circuit board and its components must be considered. Damage can result from excessive joint temperatures or electrical overstress. Thermal damage can result from an iron that is not temperature controlled or is operated at too high a temperature, whereas electrical overstress results from the flow of electrical energy (leakage or electrostatic) when the iron is placed in contact with the component, which damages or weakens an electrical component. Military soldering standard MIL-STD-2000A and industry soldering standard IPC-S-815B cover the minimum soldering requirements in detail.

Figure 2 illustrates a typical soldering iron tip and joint thermal profile for repetitive joints. When an iron and tip have been properly selected, the tip temperature will decrease when the tip is placed in contact with the joint and solder is applied. When the temperature of the joint is from 28 to 55 °C (50 to 100 °F) greater than the solder liquidus temperature, the tip is removed from the joint. The tip temperature recovers (increases) to its setpoint as the solder joint cools.

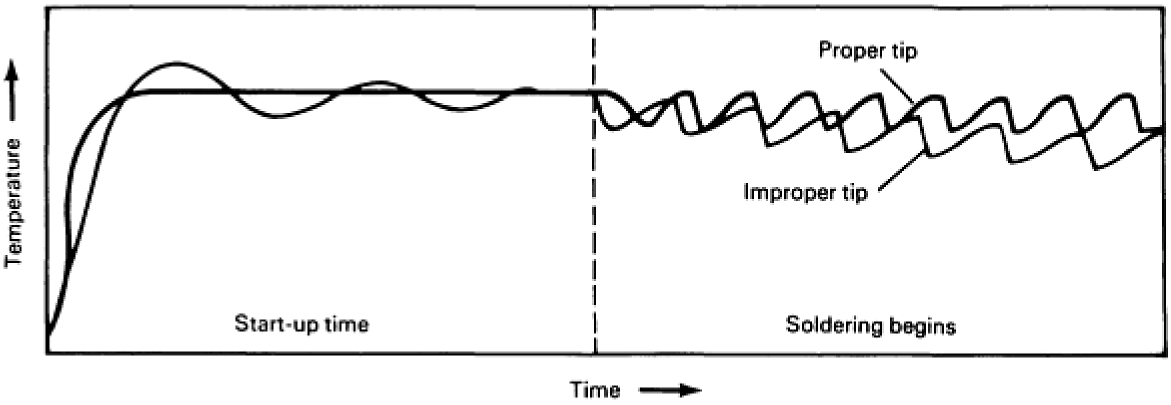

FIG. 2 TYPICAL SOLDERING IRON TIP AND JOINT THERMAL PROFILE

When a soldering iron is improperly selected, each successive joint connection decreases the working temperature of the tip. After several solder joints have been completed, the tool may be unable to form reliable joints. A properly selected soldering iron will stabilize the tip at a temperature that does not damage components, while providing a working temperature sufficient for soldering successive joints reliably.

Iron Soldering

Mark Cowell, Metcal, Inc.

Selected References

- M. FREDERICKSON, *CONDUCTIVE POINT-TO-POINT SOLDERING,* ELECTRONICS MANUFACTURING PRODUCTIVITY FACILITY (NAVIRSA-DET.), 1989
- L. LAMBERT, *SOLDERING FOR ELECTRONIC ASSEMBLIES,* MARCEL DEKKER, 1988
- H. MANKO, *SOLDERS AND SOLDERING,* MCGRAWHILL, 1964, P 225-250
- R. SKIPP, *SOLDERING HANDBOOK,* BSP PROFESSIONAL BOOKS, 1988, P 28-47

Torch Soldering

Charles E. Fuerstenau, Lucas-Milhaupt, Inc.

Introduction

TORCH SOLDERING (TS) utilizes a fuel gas flame as the heat source in the soldering process. The fuel gas is mixed with either air or oxygen to produce the flame, which is applied to the materials to be soldered until the assembly reaches the proper soldering temperature. Solder filler metal, which melts at temperatures below 450 °C (840 °F), is added to the assembly to bond it. Successful torch soldering is accomplished when parts are clean and fit together closely, and when oxides are not excessive.

Torch soldering is used in numerous industries that utilize a variety of base metals. It is most often used to produce a leak-tight assembly with some degree of mechanical strength. This process does have several advantages when compared with other metal-joining methods.

These advantages include:

- RELATIVELY LOW TEMPERATURES, WHICH RESULT IN MINIMAL DISTORTION AND HEAT DISCOLORATION OF THE SOLDERED ASSEMBLY
- FLEXIBILITY, SUCH THAT ONE TORCH CAN BE USED TO SOLDER A VARIETY OF ASSEMBLIES
- RELATIVELY INEXPENSIVE CAPITAL EQUIPMENT COST
- ABILITY TO SOLDER MOST BASE METALS AND BASEMETAL COMBINATIONS WITH AN APPROPRIATE FLUX
- LOCAL HEATING OF SMALL JOINTS ON LARGE ASSEMBLIES, WHICH PRECLUDES THE NEED TO HEAT THE ENTIRE ASSEMBLY
- PROCESS AUTOMATION, IN MANY CASES

The limitations of the process are that:

- A FLUX IS NEEDED BECAUSE THE PROCESS IS ACCOMPLISHED IN AIR, AND THIS FLUX OFTEN MUST BE REMOVED IN ORDER TO PREVENT PART CORROSION
- HIGHLY REACTIVE MATERIALS, SUCH AS TITANIUM AND MANY REFRACTORY METALS, CANNOT BE SOLDERED, BECAUSE NO SUITABLE FLUXES ARE AVAILABLE
- SOLDERED JOINTS DO NOT NORMALLY EXCEED THE STRENGTH OF THE PARTS BEING SOLDERED, THAT IS, THE PROCESS PRODUCES RELATIVELY WEAK JOINTS, WHEN COMPARED WITH BRAZED AND WELDED ASSEMBLIES

Torch Soldering

Charles E. Fuerstenau, Lucas-Milhaupt, Inc.

Applications

Materials. Torch soldering is used extensively on copper, brass, and other copper alloys. Steel, stainless steel, aluminum, gold, and other metals, as well as many combinations of these, also can be soldered by this process.

Industries. The plumbing industry utilizes this process extensively on residential copper tubing for water lines, primarily because of its flexibility and ease of use. Soldered copper water lines provide leak-tight joints with the requisite strength. Torch heating is also used for liquid and gas conduit in industrial processing applications.

This process is also used by many other industries in assembly processes, such as the soldering of jewelry, aluminum heat exchangers, radiators, and pressure-sensing devices. Tube-to-fitting types of assemblies also are commonly torch soldered, as are butt and lap joints.

Torch Soldering

Charles E. Fuerstenau, Lucas-Milhaupt, Inc.

Equipment

Torch. Typically, the equipment required for torch soldering is simple. Single torches are often used for hand soldering, whereas multiple torches can be used together on automated systems.

Gases. Fuel gas and air systems are frequently used for torch soldering because of the low heat requirements. The plumbing industry often utilizes a simple hand-held propane/air torch that is attached to the top of a small, rechargeable cylinder. Surrounding air is mixed with the propane from the cylinder in the appropriate ratio.

Gas/oxygen systems can also be used for torch soldering, although they produce more heat than gas/air systems. This amount of heat is not normally required for torch soldering.

A reducing flame--that is, a flame where excess fuel is present and surrounding air is used to complete combustion--is typically used when soldering. A reducing flame helps keep the assembly clean by reducing surface oxides.

Filler Metals. Many different filler metals are used in the torch soldering process. Tin-lead, tin-silver, and tin-antimony filler metals are commonly used on copper alloys, steel, and stainless steels. Zinc-aluminum is used exclusively on aluminum-base metals. The solders that are commonly used as filler metals are shown in Table 1.

TABLE 1 COMMON SOLDERS FOR TORCH SOLDERING

COMPOSITION	SOLIDUS TEMPERATURE		LIQUIDUS TEMPERATURE		COMMENTS
	°C	°F	°C	°F	
63SN-37PB	183	361	183	361	EUTECTIC COMPOSITION; MOST FLUID OF THE SNPB ALLOYS
40SN-60PB	183	361	238	460	LONG MELTING RANGE; GOOD FOR FILLING GAPS
96.5SN-3.5AG	221	430	221	430	FLUID SOLDER; STRONGEST OF LOW-TEMPERATURE SOLDERS
95SN-5SB	233	452	240	464	FOR COPPER- AND IRON-BASE ALLOYS; NOT FOR USE ON ZINC-BEARING ALLOYS, SUCH AS BRASS
98ZN-2AL	376	710	385	725	FOR SOLDERING OF ALUMINUM; USED EXTENSIVELY ON HEAT EXCHANGERS

Basic Heating Techniques

The key to torch soldering, as in all methods of heating, is to equally heat the components that need to be soldered. This may require directing the torch to a larger mass component of an assembly. Because the filler metal will flow to the hottest portion of an assembly, the flame should be applied so that the filler metal will flow in the intended direction. Also, movement of the torch is required to avoid overheating or burning the assembly.

Figure 1 shows the appropriate wire and torch placement for the soldering of tubing. Proper heating will result in the flow of filler metal through the joint, resulting in a fillet both inside and outside the joint. This ensures that a leak-tight joint has been produced.

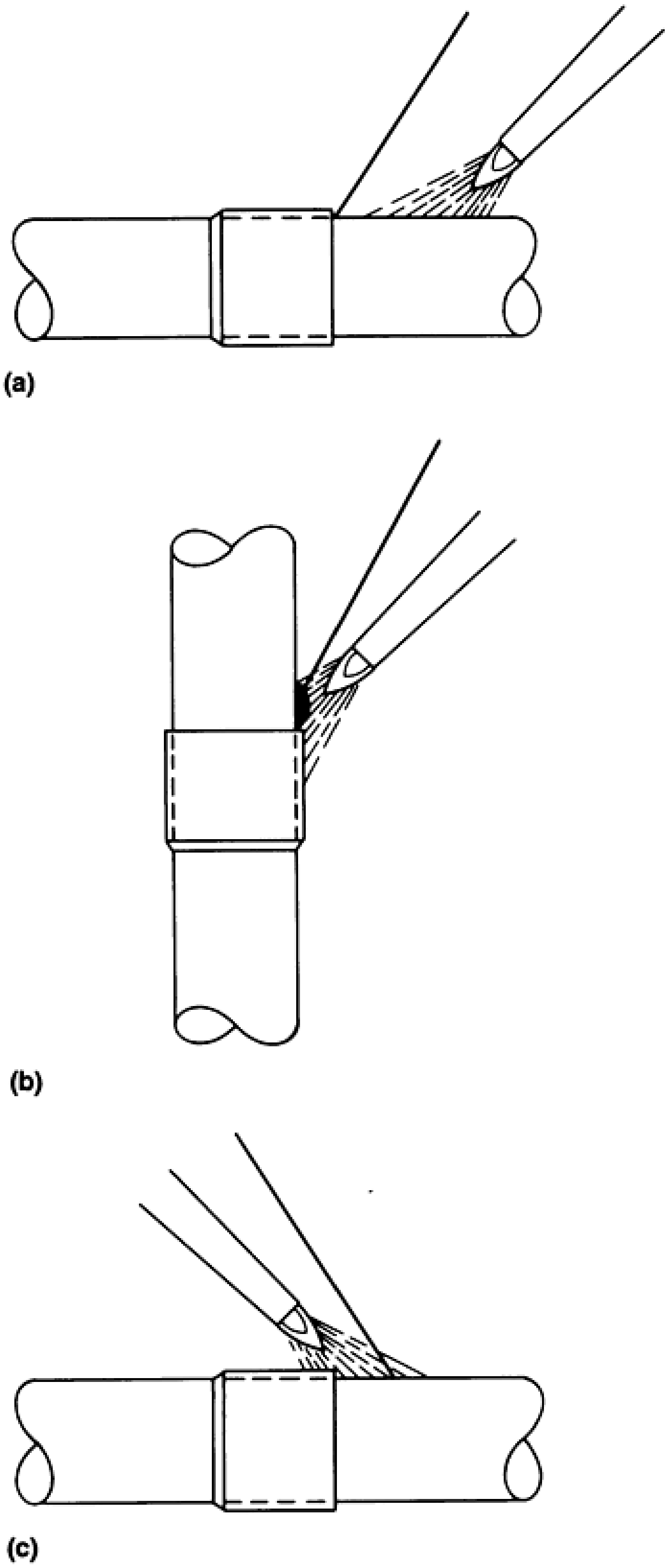

FIG. 1 EXAMPLES OF PLACEMENT OF FILLER-METAL WIRE AND FLAME. (A) AND (B) PROPER PLACEMENT. (C) POOR PLACEMENT

Torch soldering techniques are easy to learn when proper training and experience are provided.

Torch Soldering

Charles E. Fuerstenau, Lucas-Milhaupt, Inc.

Selected References

- *FILLER METAL SELECTION GUIDE,* HANDY AND HARMAN/LUCAS-MILHAUPT, 1992
- *THE BRAZING BOOK,* HANDY AND HARMAN/LUCAS-MILHAUPT, 1991

Furnace and Infrared Soldering

Phil Zarrow, Synergistek Associates

Introduction

FURNACE SOLDERING (FS) encompasses a group of reflow soldering techniques in which the parts to be joined and preplaced filler metal are put in a furnace and then heated to the soldering temperature. Five reflow technologies are currently certified for use in surface-mount technology (SMT) applications:

- TYPE A: VAPOR PHASE
- TYPE B: AREA CONDUCTION (THAT IS, LINEAR CONDUCTION)
- TYPE C: HOT BAR
- TYPE D: CONVECTION AND CONVECTION/INFRARED (IR)
- TYPE E: LASER

Of these five methods, three are considered to be mass reflow techniques (Types A, B, and D), because all of the solderable interconnections on the surface of a printed wiring board (PWB) assembly are brought through the reflow heating cycle simultaneously.

Vapor-phase reflow, or condensation heating, had early prominence as an SMT reflow soldering method. A fluid that has a boiling point higher than the solder melting point is heated and the vapor is contained in a chamber, through which the PWB assembly is passed. Heat transfer is achieved through the heat of vaporization released by the fluid as it condenses on the PWB surface. The fluids are compatible with most substrate and device packaging materials. The process is relatively insusceptible to variances in component distribution and assembly densities. Additional information can be found in the article "Vapor-Phase Soldering" in this Volume.

Area conduction comprises multiple "hot-plates" with a substrate conveyance system. This method can also be combined with other energy-transfer techniques, including convection and infrared, when appropriate for specific applications.

Infrared heating technology has also been very successfully adapted to SMT reflow soldering. Area IR emitters, that is, radiant panels, and lamp radiant emitters that are arranged in arrays with reflectors are incorporated in a conveyorized oven to provide the source of heat. By using mid- to far-infrared emissions and presenting a large heated surface area to heat the resident atmosphere in the oven, the dual heat transfer technologies (that is, IR heat provided by the direct lamp heat and conduction-generated heat provided by the indirect heating of the atmosphere) can achieve a very controlled rate of heating, which is ideal for SMT reflow soldering.

IR reflow has continued to evolve. Emphasis has been placed on enhancing heat transfer through forced convection (class 3), also known as convection-dominant reflow soldering (Fig. 1 and 2). Convection evens out the "hot" and "cold" spots caused by different surface adsorptivity. The intent is to approach the beneficial near-equilibrium aspects of vapor-phase reflow, but with the multizone profile control that has been traditionally offered by the convection/IR method. This objective is driven by the more-complex soldering applications, in terms of assembly mass and surface geometries.

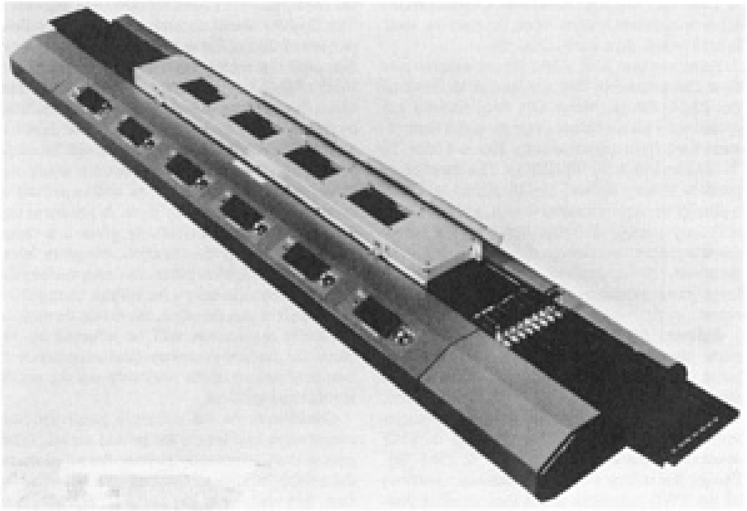

FIG. 1 CONDUCTION REFLOW SOLDERING OVEN WITH FORCED CONVECTION ENHANCEMENT. COURTESY OF SIKAMA INTERNATIONAL

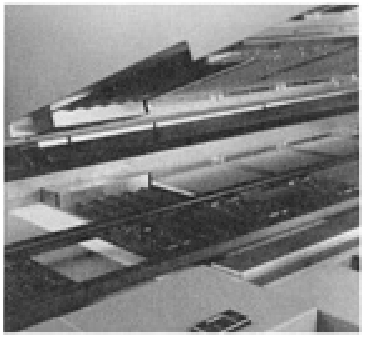

FIG. 2 FORCED CONVECTION REFLOW SOLDERING OVEN USED TO SOLDER PWB COMPONENTS. COURTESY OF VITRONICS CORPORATION

Furnace and Infrared Soldering

Phil Zarrow, Synergistek Associates

Reflow Profile

The heating cycle of the PWB assembly should be precisely controlled. The reflow profile is composed of four regions (Fig. 3):

- PREHEAT (PREBAKE)
- PREFLOW (SOAK)
- REFLOW
- COOLDOWN

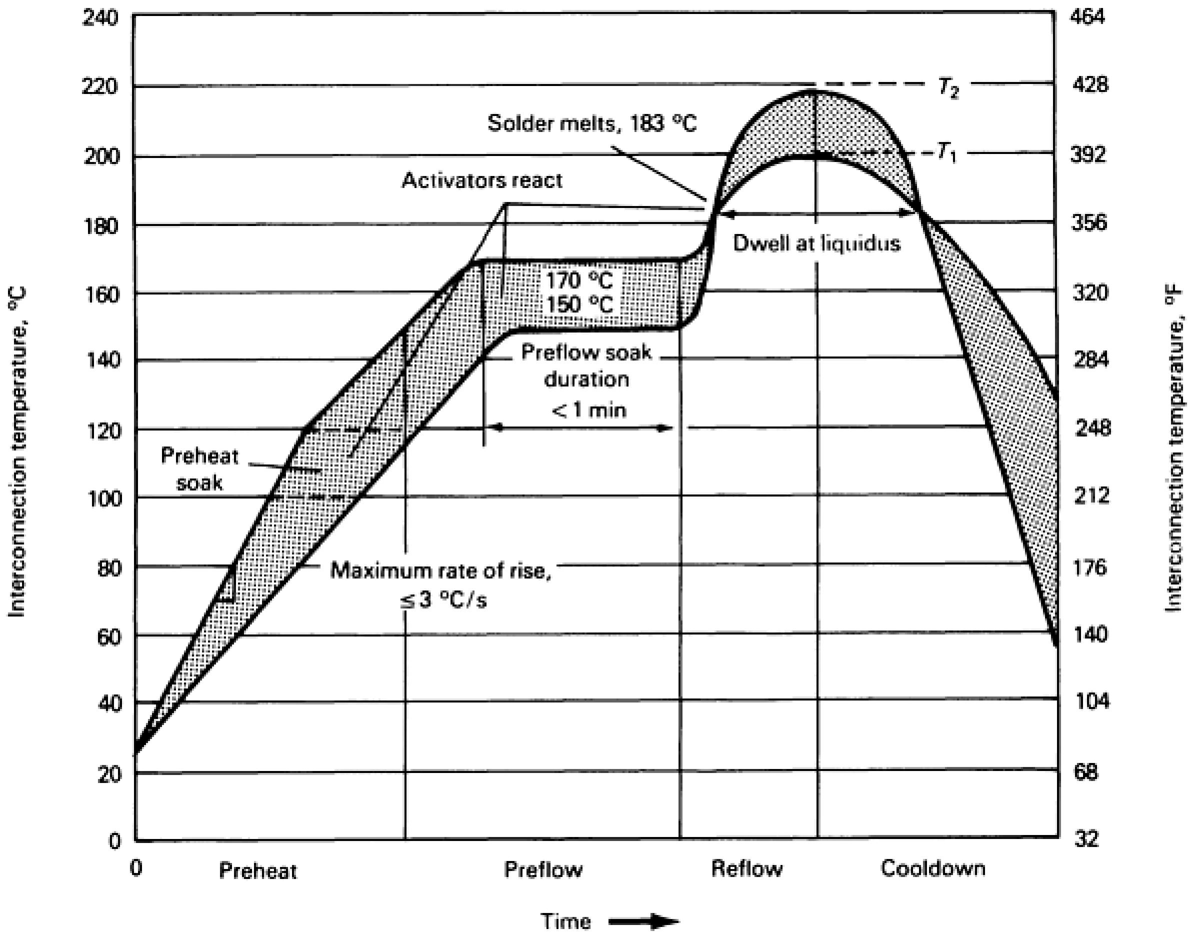

FIG. 3 PLOT OF INTERCONNECTION TEMPERATURE VERSUS ELAPSED TIME SHOWING FOUR REGIONS THAT CONSTITUTE REFLOW PROFILE. TOTAL ELAPSED TIME FROM PREHEAT TO COOLDOWN VARIES FROM 3 TO 6 MIN.

Preheat. As the PWB assembly enters the oven tunnel (at ambient temperature) and energy is introduced to the assembly, the heat transfer process begins. The interconnection temperature is brought up to approximately 100 to 120 °C (212 to 250 °F) to reduce the risk of thermal shock to the components and PWB delamination. The control of the rate of heating is most critical in the preheat portion of the profile. Many component manufacturers, particularly suppliers of ceramic capacitors, recommend a preheat rate of heating that is less than 4 °C/s (7 °F/s). The rate of heating should be based on the application, and the component complement should be considered.

Throughout the preheat stage, activators commence "scrubbing" and some solvents begin to evaporate. In certain solder paste formulations, some of the flux activators may begin oxide reduction at this stage.

Preflow. Before it attains reflow temperatures, the assembly is allowed to "soak" for a period of time. Those PWB assemblies that have polymer-base substrates require time to allow the transfer of heat throughout the assembly, in order to reduce thermal gradients. The remainder of the noncontributing solvents contained in the solder paste are driven off. As the temperature of the solderable interconnections are brought to just below the solder melting point, flux activators react with the surface to clean component and pad interfacing surfaces.

The temperature range and duration of this stage varies among flux types and the solder paste formulations that different manufacturers produce. Generally, rosin mildly activated (RMA) fluxes become fully activated in a preflow soak at temperatures between 150 and 170 °C (300 and 340 °F). Some RMA formulations achieve optimum results when the preflow soak is held to less than 1 rain (Fig. 3).

Some organic acid (OA) fluxes exhibit preflow characteristics that are similar to those of the RMA fluxes. Many OA formulations are optimized with a gradual slope in which temperature rises from approximately 100 to 170 ± 15 °C (212 to 340 ± 30 °F) (Ref 1). The duration of preflow is very critical to OA fluxes and will typically be approximately 1 min, although this will vary among different formulations (some manufacturers recommend minimizing the duration). If the preflow time interval is too long, paste oxidation and flux degradation can occur.

Reflow. As the temperature of the solder paste on the interconnect passes the melting point of the solder alloy and enters a molten state, the reflow portion of the process begins. With 63Sn-37Pb, a eutectic solder that represents the most commonly used alloy in SMT assembly, reflow occurs at 183 °C (361 °F). During the reflow stage, all solderable portions of the PWB assembly reach their desired peak temperatures. Most solder manufacturers recommend bringing the interconnect temperature to approximately 15 to 30 °C (27 to 54 °F) above the alloy melting point to ensure complete melting, as well as good solder flow, and to aid fillet formation. This is designated T_1 in Fig. 3. The supplier should be consulted to provide the actual temperature for a particular formulation.

The duration of time at which a solderable interconnect resides above the melting point of the solder paste is called the "dwell at liquidus." The liquidus dwell extends through the reflow portion of the profile and into the cooldown portion until the solder reaches solidus. This is the most critical portion of the reflow profile. Because these are the highest temperatures endured by the assembly during the process, the duration at these elevated temperatures should be minimized. Intermetallic growth occurs while the solder is in the molten state, as well as a result of diffusion over a period of time. A sustained liquidus dwell period essentially gives it a "head start." The longer the duration, the more intermetallic growth takes place. Because cooling the interconnection to below the solidus temperature is also part of this duration, the actual dwell for a particular application will be affected by the mass and surface geometry (and expediency of heat dissipation) of the assembly and the resulting thermal gradient.

Cooldown. As the assembly passes its peak temperature and leaves the heated tunnel, it begins to cool, eventually passing the point where the solder solidifies. Cooling can be aided by fans, blowers, inert gas curtains, refrigeration, normal exposure to ambient air, or a combination of several of these mechanisms. To avoid thermal shock to components, it is currently recommended that the rate of cooling not exceed -4 °C/s (-7 °F/s). However, rapid cooling does promote a finer grain growth in the joint and contributes to a stronger interconnect.

The actual time and temperature duration range for each zone not only varies among the different types of fluxes used in SMT soldering, but also among similar flux designation formulations of different manufacturers. The user should consult his supplier to obtain the actual time and temperature durations that are recommended for the specific solder paste being used. It is important to note that most fluxes have a fairly substantial "envelope" in terms of temperatures at the different stages. In some cases, however, actual duration can be more stringent.

Reference cited in this section

1. J.S. HWANG, *SOLDER PASTE IN ELECTRONIC PACKAGING*, VAN NOSTRAND REINHOLD, 1989, P 38

Furnace and Infrared Soldering

Phil Zarrow, Synergistek Associates

Reflow Schedule

A reflow specification should be required for any facility that uses mass reflow techniques. This procedure should mandate the temperatures that the PWB assembly will be exposed to at the peak of the reflow cycle. The full peak liquidus temperature, T_1, is the minimum peak temperature that any solder and interconnection must be exposed to for proper joint formation and is a parameter of the reflow schedule. It is affected by factors such as solder paste alloy, metal content, particle size, flux, surface tension, and fluidity of the solder. As noted earlier, this value will vary among different formulations and should be specified by the solder paste manufacturer (see the section "Reflow Profile" in this article).

The most vulnerable component (MVC) maximum peak temperature is represented by the value T_2. It identifies the component that has the lowest thermal threshold and is most likely to be thermally damaged. The MVC might be an inserted component with a low material melting point (that is, an electrolytic capacitor), a plastic connector, a switch, a label, or some other heat-sensitive part or material. The lowest common denominator within the assembly will determine the highest temperature to which any part should be exposed. A buffer of -5 °C (-9 °F) should be taken from that threshold to establish the MVC temperature used in the specification.

In some applications, the MVC temperature might be determined by the internal heat sensitivity of an active device. Hence, the internal temperature tolerances should also be taken into consideration.

Processing in Inert Atmosphere

Reflow soldering in an inert atmosphere can, in some instances, be beneficial to improved wettability. However, implementing nitrogen in the reflow oven is being driven by two specific factors: "no-clean" fluxes and the bare copper assembly process.

No-clean (or low-residue) fluxes represent one solution to the search for alternatives to cleaning PWB assemblies with environmentally damaging chlorofluorocarbon (CFC) solvents. The result is the development of solder formulations that do not require cleaning.

A number of factors will ultimately affect feasibility for a given application. Besides external cosmetic appearance requirements, both solder joint quality and testability (in light of remaining residue) will be dictated by other user standards and requirements. Although some low-residue solder paste formulations that work well in ambient (oxygen-laden) atmosphere are available, many very low residue no-clean solder pastes require that reflow take place in an inert (N_2) atmosphere. Some no-clean solder paste formulations have been successfully tested at approximately 300 ppm of O_2, but the higher the concentration of inert atmosphere, the wider the process window. Optimum results, in terms of very low residue, have been attained with nitrogen atmosphere of less than 20 ppm O_2.

The bare copper assembly process has been prompted by the need for coplanarity on the PWB substrate when placing fine- pitch devices. Hot-air leveling of solder during PWB fabrication has not attained adequate levels of coplanarity when components of 0.51 mm (0.020 in.) pitch or finer are being attached. In the bare copper assembly process, the PWB is fabricated in the usual manner, except that the tin-lead solder coating is omitted. Instead, the substrate is coated with a sealant (that is, inhibitor) that prevents the copper from oxidizing. In assembly, solder paste deposition (screened or stencilled) and component placement are accomplished "as usual." During reflow, the sealant vaporizes between 100 and 120 °C (212 and 250 °F) (during the preheat stage), which exposes the copper to the solder paste. Because the copper is exposed, the furnace must contain an inert atmosphere to prevent excessive oxidation of the lands. Just how oxygen-free the oven atmosphere needs to be varies among users. Some obtain successful results in atmospheres composed of approximately 500 to 700 ppm O_2. The most stringent specification requires a nitrogen atmosphere with less than 50 ppm O_2 at any point in the process where the PWB assembly becomes elevated above 150 °C (300 °F).

Reference

1. J.S. HWANG, *SOLDER PASTE IN ELECTRONIC PACKAGING*, VAN NOSTRAND REINHOLD, 1989, P 38

Phil Zarrow, Synergistek Associates

Selected References

- C.L. HUTCHINS, SOLDERING SURFACE MOUNT ASSEMBLIES, *ELECTRON. PACK. PROD.*, AUG 1992
- N. SOCOLOWSKI, SOLDER PASTE-KEY PARAMETERS TO FINE PITCH COMPONENT ASSEMBLIES, *NEPCON WEST PROC.*, MARCH 1990
- P. ZARROW, MAJOR FACTORS FOR SUCCESS IN SMT REFLOW SOLDERING, *ELECTRON. PACK. PROD.*, JUNE 1992
- P. ZARROW, "SMT REFLOW SOLDERING WORK-BOOK," SYNERGISTEK ASSOCIATES, 1992

Dip Soldering

Roy E. Beal, Amalgamated Technologies, Inc.

Introduction

DIP SOLDERING (DS) is accomplished by submerging parts to be joined into a molten solder bath. The molten bath can be any suitable filler metal, but the selection is usually confined to the lower melting point elements. The most common dip soldering operations use zinc-aluminum and tin-lead solders.

The molten bath can be heated by electricity or gas. The bath container is made from ceramic materials or a metal that is nonreactive to the filler metal used for dipping. The dip baths, which can range in size from very small to large, are used in a wide range of industries.

The process of dip soldering is simple to carry out. The joint areas are cleaned and then coated with an appropriate flux. The parts are then lowered into the molten filler metal and joining takes place by capillary attraction into the joints. The solder does not wet surfaces that are not coated with flux. Jigs and fixtures are normally essential to hold components at proper clearances for joining purposes and to maintain overall dimensions of the finished product. They should not be wetted by the solder. Furthermore, any substrates that must be submerged should be able to tolerate the solder temperature.

The molten bath supplies both the heat for raising the temperature of the part and the filler metal to make the joints. It is therefore important that the dip soldering pot capacity be large enough to allow heating of the part and joining without significantly reducing the temperature of the molten filler metal. For some applications, soldering pots are designed for specific production rates.

An advantage of the dip soldering process is the wide range of part sizes, shapes, and thicknesses that can be accommodated by the same apparatus and still result in satisfactory joints. Larger parts require a preheating operation in order to reduce thermal shock when the part enters the dip pot. The preheat also allows quicker attainment of equilibrium between the molten metal in the pot and the part being joined. Atmosphere protection of the dip soldering pot is necessary when the filler metal has oxidation tendencies. Surface oxides on the workpieces must be removed prior to dip soldering, because they can interfere with a proper joining operation.

Dip Soldering

Roy E. Beal, Amalgamated Technologies, Inc.

Applications

Dip soldering is used in job shops, as well as high-rate production situations. Therefore, the parts joined by dip soldering vary from individually designed products to mass-produced items. Dip soldering is ideal for the production of prototype engineered products, which are often made for test purposes. The process also can be used to join limited-production items for the automotive, aircraft, consumer, electronic, and other industries. High-production products are typically heat exchangers for automotive use, high-volume telecommunications equipment, and appliances.

Joints produced by dip soldering techniques are usually sleeve, lap, or T-joints that fill by capillary attraction. The parts to be joined are assembled with gaps appropriate to the particular filler metal surface tension and viscosity characteristics. The joint/fixture design must accommodate thermal expansion that is due to heating the entire substrate or part. Joint gaps also depend on the reactions of the molten filler metal with the material to be joined. For example, aluminum filler metals (or those with zinc, tin-zinc, or zinc-aluminum alloys) react more quickly with aluminum than a tin-lead solder will with copper and therefore require larger gaps for joining.

Roy E. Beal, Amalgamated Technologies, Inc.

Equipment

Dip soldering equipment is very basic, comprising a bath made of metal or ceramic to hold the molten filler metal and a means to heat and control the molten filler metal temperature. In the radiator industry, job shops use gas heating with their pots for the sake of economy. Much equipment is made to operate electrically, either because of space considerations or because the pots are used in conjunction with other equipment on a production line that might be affected by a gas heating system. Some systems include a metal pump that raises the level of the liquid metal in areas smaller than the overall pot in order to ensure that no surface oxides are present during the dip operation. This type of equipment is used in automotive radiator manufacture. A typical installation is illustrated in Fig. 1.

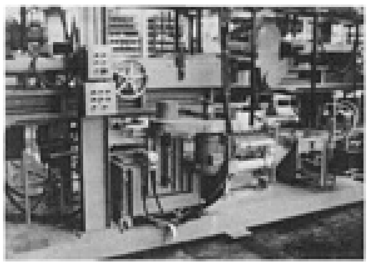

FIG. 1 DIP SOLDERING POT IN A RADIATOR PRODUCTION LINE

Roy E. Beal, Amalgamated Technologies, Inc.

Personnel

Solder pots are generally run by production personnel who should be trained to handle and care for equipment that contains molten metals. A substrate that has been raised to the soldering temperature represents a high-temperature hazard. Parts must be dry upon entering the dip solder pot to avoid splashes or explosions. Suitable protective wear (with guards, where possible) is necessary. Good ventilation is important, because of flux and, possibly, metal fumes. Generally, operators of tin-lead solder dip pots are now checked for lead content in the blood, because of possible health hazards.

Resistance Soldering

Karl Lazar

Introduction

RESISTANCE SOLDERING (RS) is a soldering process in which the heat needed to melt the solder is developed by the resistance of the material when a large electrical current is supplied. Resistance soldering can be applied to electrically conductive materials that allow the passage of electric current. The process can be used for selective spot soldering of small components, for the soldering of closely placed parts on an assembly, or for heat restriction when necessary. It is similar in many ways to resistance brazing (see the article "Resistance Brazing" in this Volume).

When the current is applied, rapid local heating occurs, melting the solder. This molten solder wets the surface. As a result, the resistance in the material falls and the current increases, tripping a control. The heat rapidly dissipates into the surrounding area, and the solder quickly solidifies.

Resistance Soldering

Karl Lazar

Process Applications

The RS process can be used in all soldering operations and with all solderable metals. The only limitations are the thickness and the design of the parts to be soldered. Resistance soldering is used to join steels (for example, carbon, low alloy, and stainless) and nonferrous alloys (for example, aluminum and aluminum alloys, nickel and nickel alloys, and copper and copper alloys) up to 3.2 mm ($\frac{1}{8}$ in.) in thickness (Ref 1).

The RS process is suitable where:

- AN OPEN FLAME CONSTITUTES A POTENTIAL HAZARD IN THE WORKPLACE.
- THE HEAT NEEDED FOR SOLDERING MUST BE CONFINED TO A SPECIFIC AREA.
- THE OXIDATION ATTRIBUTED TO HEATING OF THE SOLDER MUST BE MINIMIZED.
- THE COMPONENTS TO BE JOINED ARE INACCESSIBLE WITH A CONVENTIONAL SOLDERING IRON.
- THE BEAD PRODUCED CAN BE ADAPTED TO AUTOMATED PROCESSING/MASS PRODUCTION METHODS.

Reference cited in this section

1. H.E. BOYER AND T.L. GALL, ED., *METALS HANDBOOK DESK EDITION*, AMERICAN SOCIETY FOR METALS, 1985, P 30-8 TO 30-9

Resistance Soldering

Karl Lazar

Equipment

Resistance soldering equipment typically consists of two electrodes that have been modified for use with a specific part configuration. These electrodes are made of a high-resistance electrical conductor (for example, carbon, graphite, or tungsten). The geometry of the electrode can assume any one or a combination of the following forms (Fig. 1):

- PENCIL
- TWIN-PENCIL
- ROLLER
- CLAMP
- PLIER-TYPE

The electrodes are connected to the transformer by oversized current-carrying conductors. The heart of the system is a heavy-duty, variable, step-down transformer designed to develop an output of 2 to 25 V_{ac} and a current of 5 to 500 A (Fig. 2). A footswitch, a timer, or an adjustable-resistance bridging device across the electrodes can be used to provide a method for the parts to reach the desired solder temperature and shutdown. This allows the solder to solidify and the parts to return to ambient temperature in a 4 to 6 s cycle.

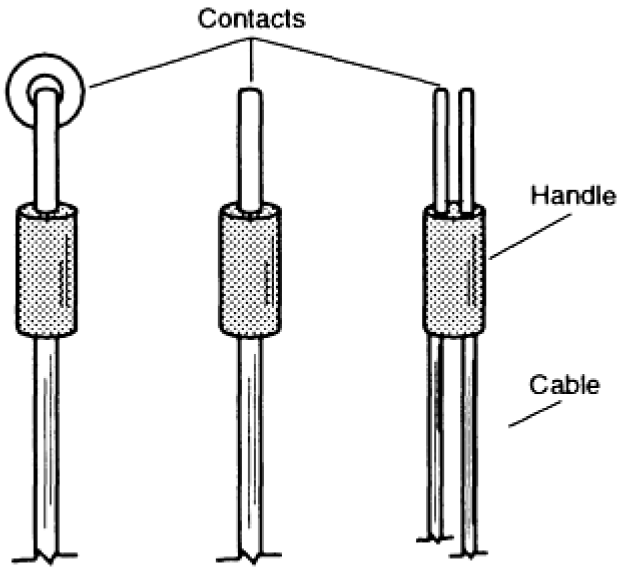
Contacts
Handle
Cable

Roller electrode **Pencil electrode** **Twin-pencil electrode**

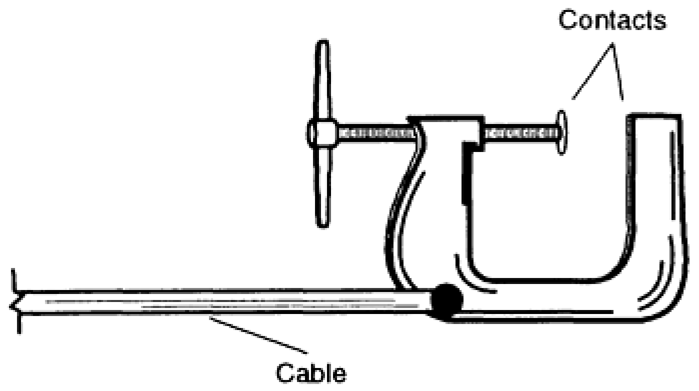
Contacts
Cable

Clamp electrode

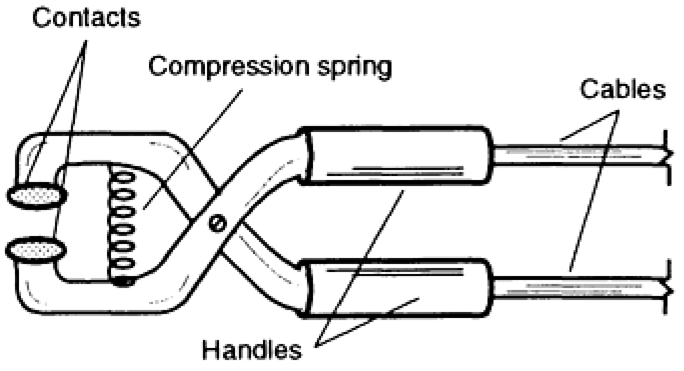
Contacts
Compression spring
Cables
Handles

Plier-type electrode

FIG. 1 TYPICAL ELECTRODE CONFIGURATIONS USED IN RESISTANCE SOLDERING APPLICATIONS

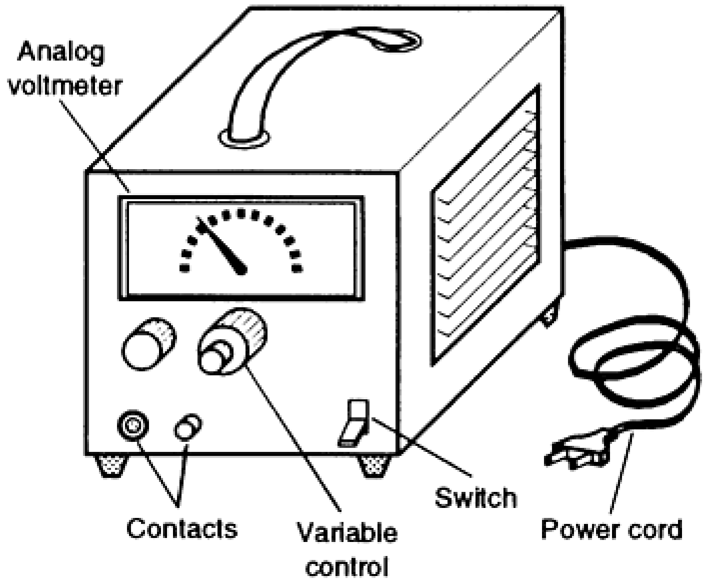

FIG. 2 TYPICAL VARIABLE STEP-DOWN TRANSFORMER WITH 2 TO 25 V OUTPUT (AT 5 TO 500 A CURRENT) USED IN RESISTANCE SOLDERING PROCESSING

Personnel Training

Soldering personnel with a strong basic training will need some additional special training to operate an RS system. The operator's function is to load and unload the parts and to maintain or replace the electrode heads when required. General soldering health and safety rules should be observed to avoid electrical shock and the potential for overheating solder to cause spatter. Typically, the engineer who originally designs the electrode heads is responsible for the equipment's initial start-up.

Resistance Soldering Practice

The RS process requires preassembly of the workpiece components. Because the workpiece and the solder are integral parts of the current loop, care must be taken to ensure that the current path is not blocked by any foreign or nonmetallic substance. Hence, the preassembled workpiece requires a clean, deburred surface and the proper solder alloys (preform or in paste form). The use of wire solder and flux is not recommended because of the rapidity of heating and the potential hazard of electrical shock.

The preassembled workpieces are positioned in a grounded jig or clamp, and the movable electrode is brought in contact with the workpiece to complete the circuit. When the power is turned on, the operator waits for signs that the solder is starting to flow and then immediately turns the power off. When using automatic soldering equipment, the duration of the current flow is monitored by sensors that determine the length of the "on" cycle. As the workpiece cools, the solder solidifies, and the workpiece is removed from the jig. The next assembly can then be placed in the jig. A properly designed movable electrode head allows multiple solder connections to be produced in a single operation.

Resistance Soldering

Karl Lazar

Reference

1. H.E. BOYER AND T.L. GALL, ED., *METALS HANDBOOK DESK EDITION,* AMERICAN SOCIETY FOR METALS, 1985, P 30-8 TO 30-9

Laser Soldering

Riccardo Vanzetti, Vanzetti Systems, Inc.

Introduction

INDUSTRIAL LASERS are able to deliver large amounts of heat with great precision and without contact, making them ideal for applications that have either a destructive nature, such as cutting or drilling, or a constructive nature, such as soldering or annealing.

The use of focused energy generated by either a CO_2 or a neodymium-doped yttrium-aluminum-garnet (Nd:YAG) laser continues to gain acceptance as a process for a number of operations requiring high-temperature injection with speed and precision. For example, in the electronics field, component soldering can be carried out one joint at a time with highly reliable results. Laser soldering uses the well-focused, highly controlled beam to deliver energy to a desired location for a precisely measured length of time.

Advantages of laser soldering are that it is a noncontact procedure; it avoids thermal stress by localizing heat input and confining it to the solder joint area; it reduces intermetallic compound formation, due to rapid joint formation which results in a more ductile joint; and it involves a fine grain size (due to rapid cooling), resulting in better fatigue properties. Disadvantages include low output rates, high power requirements, and high capital requirements.

This discussion focuses on two types of laser soldering operations, blind laser soldering and intelligent laser soldering, and the conditions under which their application is suitable. Key attributes of each system, relative to other types of soldering operations, are also identified.

Laser Soldering

Riccardo Vanzetti, Vanzetti Systems, Inc.

Blind Laser Soldering

The early configurations of production lasers were of the "blind" (no feedback) type. The system included a laser programmed for power, as well as length of heat injection time. The focal spot of the laser beam was brought to impinge on the target by moving either the target itself or the laser beam, or a combination of the two.

Among the vast array of currently available lasers, the preferred systems for industrial soldering applications are the CO_2 laser, which emits at a wavelength of 10.6 μm, and the Nd:YAG laser, which emits at a wavelength of 1.06 μm. The total energy emitted by these lasers (or by any laser, in fact) is in part absorbed by the target, in part reflected by the target, and the balance causes an instantaneous emission of highly energized atomic and subatomic particles from the target's uppermost surface layer. The ratio of these three parts is a function of laser wavelength and of surface emissivity of the target. The latter, in turn, is a function of chemical composition, crystal structure, and degree of smoothness.

For all practical purposes, certain solder alloys are known to reflect about 74% of the energy from a CO_2 laser, but just 21% of the energy from a Nd:YAG laser. These figures are approximate, because the presence of fluxes and activators reduces the reflection factor. However, the Nd:YAG laser, when used for soldering, is more efficient.

In a blind system, the coordinates of the target, the power of the laser, the diameter of the beam at its focal point, the duration of the incident heat pulse, and other parameters must be preprogrammed and cannot be changed during the operation. Without a feedback mechanism, the quality of the work done by the laser can only be assessed after the operation has been completed.

This approach is adequate for highly repeatable production situations in which the targets are nearly identical to one another. Once the optimum settings for all variables have been established, uniformly good results can be expected. Care must be taken to avoid any process variations, such as changes in surface emissivity of the target or its contamination by a host of other factors, including solderability degradation from oxidation and handling.

However, blind laser soldering is less suitable in electronics applications, where components and the printed circuit pose a serious problem, because no two joints are identical. Although they might visually appear to be identical, their heat requirements can vary substantially. This is due to hidden differences in their heat-sinking efficiency, which is a function of the thermal mass of the subsurface elements connected to the joint. Especially critical is the diameter of the laser beam at the focal spot, because it must cover proportional areas of the elements to be joined, plus some of the solder material. This is why blind laser soldering seldom finds practical applications in the field of electronics. The most difficult applications involve fine pitch technology (FPT) and tape automated bonding (TAB), where the very small dimensions of the joints make the use of conventional soldering systems critical.

Key attributes of blind laser soldering are:

- *REPEATABILITY:* GOOD FOR LARGE PRODUCTION RUNS OF IDENTICAL PARTS, BUT POOR FOR PROCESSES THAT INVOLVE A NUMBER OF NONCONTROLLABLE ITEMS
- *SPEED:* FAST, WHEN COMPARED WITH ANY SOLDERING APPROACH IN WHICH JOINTS ARE MADE ONE AT A TIME
- *QUALITY:* GOOD, WHEN OPTIMIZED FOR LARGE PRODUCTION RUNS OF IDENTICAL CONDUCTORS AND PARTS
- *COST:* EXPENSIVE, WHEN COMPARED WITH MOST OTHER SOLDERING TECHNOLOGIES
- *SAFETY:* CRITICAL, BECAUSE SERIOUS BURNS AND EYE DAMAGE CAN RESULT IF SPECIAL PRECAUTIONS ARE NOT TAKEN
- *FLEXIBILITY:* POOR, BECAUSE ANY DRIFT OF THE MANY PROCESS VARIABLES WILL CAUSE DEFECTS, UNLESS THE PROGRAM IS MODIFIED ACCORDINGLY

Laser Soldering

Riccardo Vanzetti, Vanzetti Systems, Inc.

Intelligent Laser Soldering

Because solder joints can differ from one another in terms of thermal mass, emissivity, contamination, presence of internal thermal barriers, and heat-sinking efficiency, an infrared (IR) detector was added to create an "intelligent" laser soldering system. This system enables the precise control of heat input to the joint, because the IR "eye," constantly focused on the joint being heated, monitors its thermal radiation, and turns off the laser beam right after lead and pad substrates have been thermally soaked by the molten solder. The resulting solder joints are neither overexposed nor underexposed.

Figure 1, which depicts an intelligent system, shows two lasers with coaxial beams: a He-Ne 0.5 mW visible beam and a Nd:YAG 50 W "heating" beam emitting at a 1.06 μm wavelength. After traveling through an optical fiber, these two beams are focused on the solder joint to be formed. At the same time, the IR (thermal) radiation emitted by the heated solder joint is focused on the IR detector, which does not register the wavelength of the laser beams and therefore is not disturbed by their reflections. The visible image of the solder joint is picked up by a closed-circuit television camera and displayed on a television screen. These three optical signals are separated by dichroic mirrors in the optical head while the solder joints are brought, one at a time, under the focal spot of the system. The preprogrammed sequence is based on the *x* and *y* coordinates of each joint.

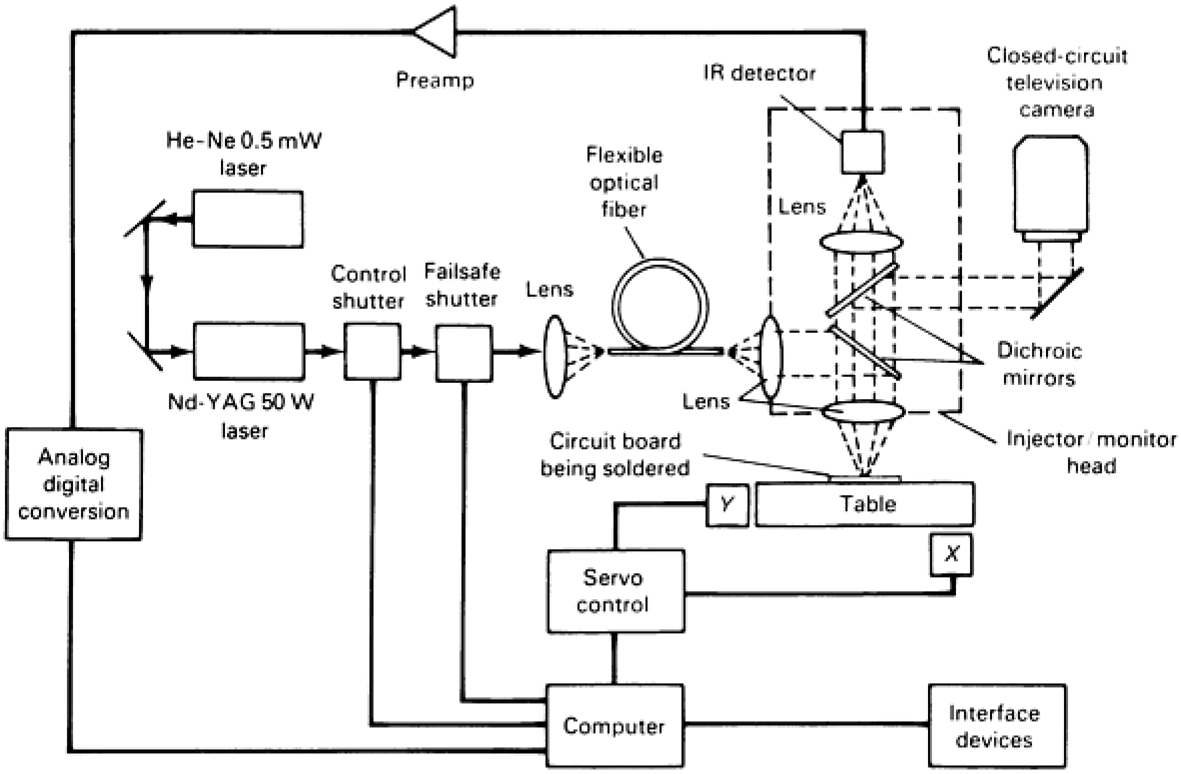
Preamp
IR detector
Closed-circuit television camera
He-Ne 0.5 mW laser
Flexible optical fiber
Lens
Control shutter
Failsafe shutter
Lens
Dichroic mirrors
Nd-YAG 50 W laser
Lens
Injector/monitor head
Circuit board being soldered
Analog digital conversion
Y
Table
X
Servo control
Computer
Interface devices

FIG. 1 SCHEMATIC OF INTELLIGENT LASER SOLDERING SYSTEM

The output of the IR detector is an analog signal that depicts, in real time, the radiation history of a spot on the surface of the solder joint during fabrication. This signal, called the "infrared signature" of the joint, is unique for every joint. The amplitude of the signal records the thermal radiation of the joint as it first rises during the heat injection by the laser and then as it decays after the laser is shut off. After the IR signature is turned into digital form, it is forwarded to the computer, which watches for the dip in the amplitude of the signal radiated by the joint, which pinpoints the transition from a solid to a liquid phase. At this point, the computer decreases the power of the laser for several milliseconds (the "dwell" phase) to allow adequate thermal soaking of the substrate. The laser is then turned off to allow the joint to quickly cool down to ambient temperature.

Figure 2 shows the theoretical signature of a typical good joint. The top curve is the radiation signal, as seen by the detector, and the lower curve is the temperature of the joint at each corresponding point in time. The sharp peak of the upper curve signals, with unequivocable precision, the instant when melting begins, so that the computer can initiate the laser control sequence, thus making it impossible to either overheat or underheat the joint. The resulting joint is free from defects such as excess intermetallics, nonreflowed solder, voids, cracks, or substrate damage.

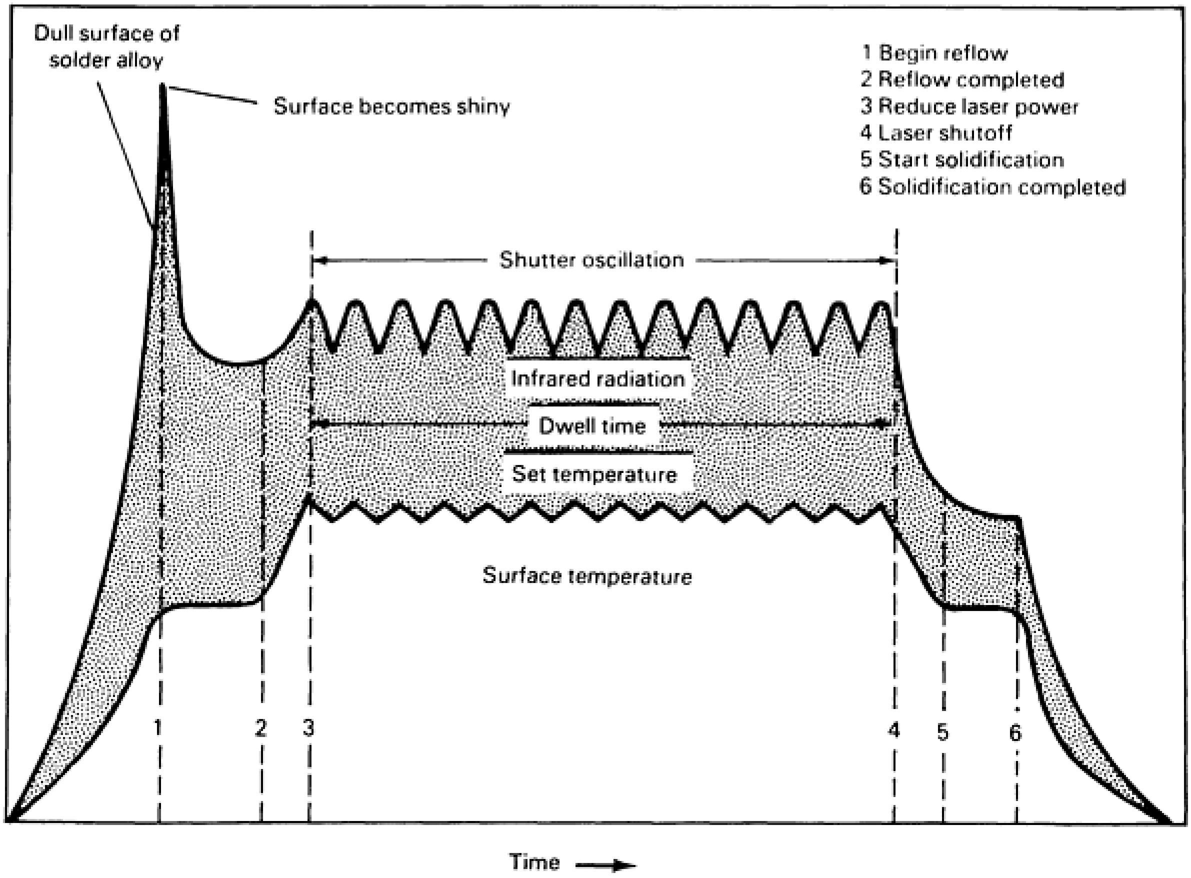

FIG. 2 THEORETICAL SIGNATURE OF A TYPICAL GOOD SOLDER JOINT

Key attributes of intelligent laser soldering are:

- *REPEATABILITY:* GOOD FOR ANY TARGET
- *SPEED:* FAST, WHEN COMPARED WITH ANY OTHER SOLDERING AND INSPECTION SYSTEM, BECAUSE THE IR SIGNATURE IS THE INSPECTION CERTIFICATE, AND IS AVAILABLE SIMULTANEOUSLY WITH THE COMPLETION OF THE JOINT
- *QUALITY:* GOOD FOR ANY TARGET
- *COST:* INEXPENSIVE, WHEN COMPARED WITH THE COST OF THE TWO SYSTEMS IT REPLACES, THAT IS, SOLDERING AND INSPECTION
- *SAFETY:* CRITICAL, BECAUSE SERIOUS BURNS AND EYE DAMAGE CAN RESULT UNLESS SPECIAL PRECAUTIONS ARE USED
- *FLEXIBILITY:* EXCELLENT, BECAUSE NO MATTER HOW DIFFERENT THE JOINTS ARE FROM ONE ANOTHER, EACH RECEIVES A CUSTOM DOSE OF THE LASER HEAT NECESSARY FOR OPTIMUM RESULTS

Hot Gas Soldering

Michael D. Frederickson, Electronics Manufacturing Productivity Facility

Introduction

HOT GAS SOLDERING is a process that is commonly used in applications where the workpiece thermal mass is small and the melting temperature of the solder is relatively low. The efficiency of heating large-mass substrates with hot gas is low, when compared with radiative or conductive methods of heating. The electronics industry utilizes hot gas soldering to reflow or melt solder in localized areas on circuit assemblies. A common application is the reworking or removal of electronic components.

Hot Gas Soldering

Michael D. Frederickson, Electronics Manufacturing Productivity Facility

Process Characteristics

Characteristics of hot gas heating that are critical to its effectiveness in soldering are:

- FOCUS OF GAS FLOW
- GAS FLOW RATES (VELOCITY AND VOLUME)
- TEMPERATURE CONTROL
- GAS MEDIA OR CONSTITUENT

The focus of gas flow defines the area of localized heating, determines the efficiency rate of heating a defined area, and controls the excessive heating of adjacent areas. In most electronic applications, because the areas that need to reach the solder melting temperature are relatively small, confining thermal energy to a small area is critical. The focus of gas flow is usually controlled via nozzle design or the use of baffle plates. The nozzle collimates the gas stream to optimize the gas impingement across one or more solder interconnects, whereas the baffle plates attempt to block the gas from impinging on adjacent areas or on a component body directly. Common nozzle materials are stainless steel or quartz glass.

Gas flow rates primarily determine heat transfer efficiency (that is, the temperature and heating rate of the workpiece) and may affect the total area being heated. Insufficient gas flow will reduce the efficiency of thermal energy transfer, whereas excessive gas flow will disperse heat across the assembly, which may result in the overheating of adjacent areas. On some hot gas soldering systems, the flow rates are adjustable in both volume and velocity and can be used to establish the most appropriate heating profile for soldering. The selection of optimum flow rates usually depends on the gas temperature selected for the soldering process and on the size of the openings in the nozzle, as well as the configuration of the workpiece or circuit board.

The gas temperature also affects the heat transfer efficiency and ultimately the time required to form the solder interconnect. Typically, the gas is heated by passing it through a set of resistive heating elements. The usual temperature of interest for soldering process control is the temperature at the exit aperture of the nozzle. Because air is a poor heat conductor, the gas temperature will fall off rapidly as the distance from the nozzle to the solder interconnect increases. In electronics applications, the component body tends to heat more rapidly than the actual interconnect; therefore, excessive nozzle-exit temperatures with a high flow rate may thermally damage components or circuit boards.

Typical gas media used in hot gas soldering are air, nitrogen, and a nitrogen-hydrogen mix ($75N_2$-$25H_2$). Air is the most common gas used because of its ready availability and low cost. Nitrogen is sometimes used to minimize the oxidation of the component and board metallizations and to potentially improve solder wetting performance and interconnect reliability. In some cases, a nitrogen-hydrogen mixture is used to capitalize on its lower density relative to air. The lighter gases allow for the same heat transfer performance and inert nature, while reducing the force of the gas against the area being soldered. This reduction in the force of the gas decreases the dispersive effects upon impingement. In addition, some no-clean flux materials perform better in an inert atmosphere, such as nitrogen, during the soldering process.

Hot Gas Soldering

Michael D. Frederickson, Electronics Manufacturing Productivity Facility

Process Parameters

Heating profiles are established in hot gas soldering by using a combination of gas flow rates, gas temperatures, and time. Some systems use lower flow rates and higher temperatures, whereas others use higher flow rates and lower temperatures. The actual selection of flow rates and temperatures depends on the physical and dimensional characteristics of the item being soldered, as well as component mounting density. Figure 1 illustrates a thermal profile of a component being soldered and the temperature across adjacent components.

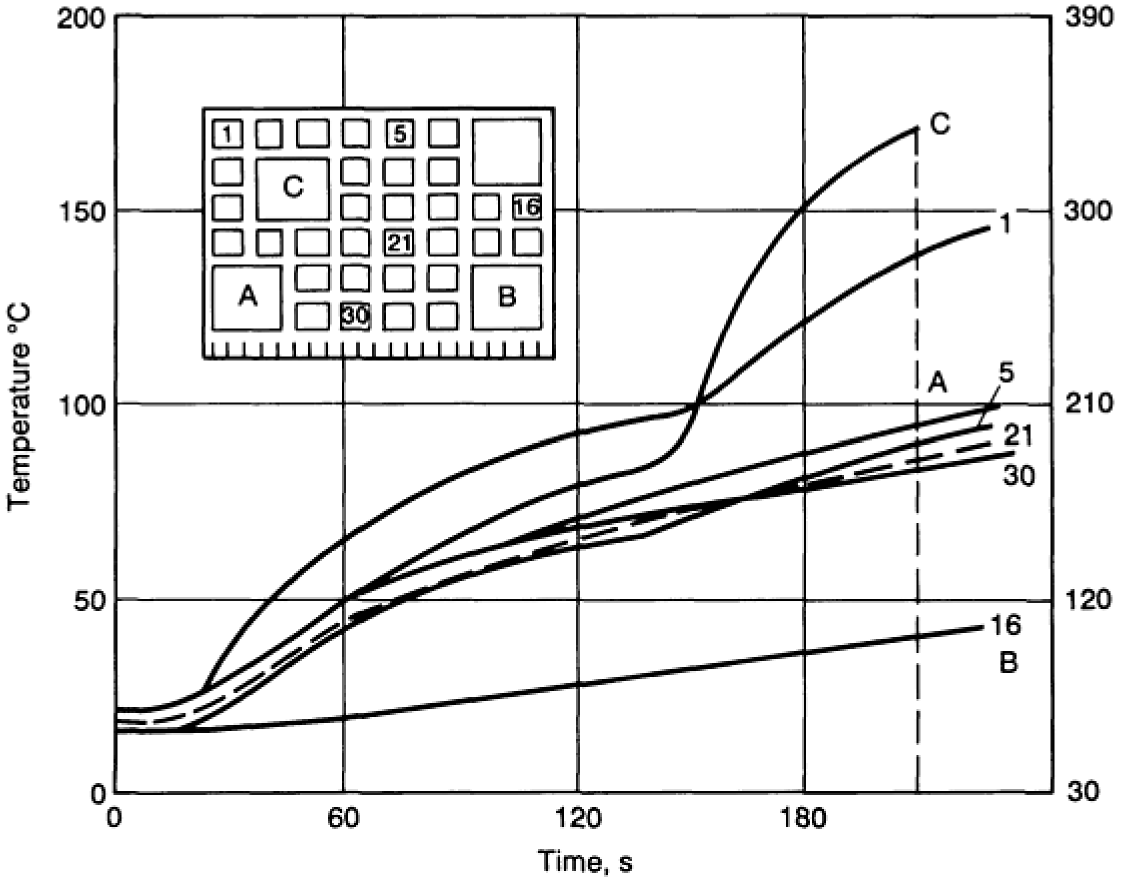

FIG. 1 TIME-TEMPERATURE RELATIONSHIP OF COMPONENT BEING SOLDERED (LINE C) AND ADJACENT COMPONENTS (LINES A AND B). LOCATIONS OF SPECIFIC SENSOR COMPONENTS ARE SHOWN IN INSET DIAGRAM.

Reliability concerns do exist when hot gas soldering techniques are used in electronics assembly. There is a potential for solder joint fracture when melting or reflowing a portion of a multi-leaded component. Fractures in solder joints occur when the joints are thermally cycled in use because of the induced stress caused by partially reflowing the component. If the hot gas process excessively heats adjacent components, causing a percentage of the interconnects on a component to reflow, then a residual stress may be established, which can significantly reduce the overall fatigue life of the solder interconnects. Therefore, gas flow rates and focus are critical to the overall reliability of the solder interconnects. The use of properly maintained equipment and trained operators should minimize these concerns.

Processing concerns also exist when using hot gas soldering in electronics assembly. Because this technique has inefficiencies relative to conductive and radiative methods, it is difficult to reflow solder joints that have large heat sinks. Therefore, an additional heating process may be needed to reduce the total hot gas heat energy required to reflow the solder interconnects. There are four methods of preheating the assemblies: oven bake, diffused hot gas, radiant, and conduction (edge rail or hot plate) heating. All methods are effective. However, the working time of assemblies preheated in an oven is very short. The selection of the preheating technique depends on the physical characteristics of the assembly being soldered and on the process equipment.

Hot Gas Soldering

Michael D. Frederickson, Electronics Manufacturing Productivity Facility

---

Selected References

- C. LEA, *A SCIENTIFIC GUIDE TO SURFACE MOUNT TECHNOLOGY,* ELECTROCHEMICAL PUBLICATIONS LIMITED, 1988
- "SURFACE MOUNT REPAIR STUDY," TECHNICAL REPORT TP60-07073, NAVAL SURFACE WARFARE CENTER, 1991
- R.S. WASSINK, *SOLDERING IN ELECTRONICS,* ELECTROCHEMICAL PUBLICATIONS LIMITED, 1989

Induction Soldering

Paul T. Vianco, Sandia National Laboratories

Introduction

INDUCTION OR (RADIO FREQUENCY) HEATING is a versatile means of providing heat to the joint area. Heating is caused by electrical resistance to eddy currents induced in the workpiece. These currents are induced by the rapidly changing magnetic field generated by a coil supplied with an alternating current. The eddy currents are generated at the surface of the workpiece (skin effect) and diminish toward the interior. The depth of direct heating by eddy currents depends on the electromagnetic properties of the workpiece material, the frequency of the alternating current (higher frequency results in shallower penetration depth), and the power level of the magnetic field (i.e., the current in the coil).

The advantages of induction heating include the ability to supply heat uniformly over the entire joint area while maintaining a localized temperature rise so heat sensitive materials or devices neighboring the joint are not damaged. Induction heating is a compromise between global heating processes such as furnaces, which warm the entire part, and more concentrated heat sources, such as torches or soldering irons, which warm only a very localized area. A second advantage of induction heating is that the temperature rise can be very rapid, thereby limiting heat loss into the workpiece, oxidation of the substrate, and thermal degradation of the flux. Limitations that must be considered with the application of induction heating are the cost of the equipment and the need for specially trained personnel. These personnel are needed to ensure proper equipment setup so that efficient transfer (coupling) of energy between the coil and the workpiece take place.

In the induction process, heat energy is not delivered to the workpiece; instead, it is generated within the material by electromagnetic field coupling between the workpiece and the alternating current in the coil (i.e., the workpiece and the coil form an inductance-capacitance-resistance circuit with the power supply).

Acknowledgement

This work was performed at Sandia National Laboratories and supported by the U.S. Department of Energy under contract DE-AC04-76DP00789.

Induction Soldering

Paul T. Vianco, Sandia National Laboratories

Coupling Efficiency

The coupling efficiency is optimized over a narrow range of operating parameters (frequency, coil turns, coil proximity, etc.) per the given workpiece parameters (electromagnetic properties and part geometry). The first consideration regarding the use of induction heating in a soldering operation is the electrical and magnetic properties of the materials to be joined. First, the energy coupling is more efficient in ferromagnetic materials, that is, materials with large magnetic permeability values (e.g., iron, low-carbon steels, nickel, etc.). Permeability being the same, materials with higher electrical resistivity heat more efficiently than low-resistivity materials. Therefore, heating efficiency is optimized by joints having metals with large permeability and high electrical resistance as opposed to low-permeability, low-resistance metals (e.g., copper, brass, or aluminum). Figure 1 illustrates the energy input required to bring several ferrous and nonferrous materials to 188 °C (370 °F) (Ref 1). These differences become especially critical when joining dissimilar materials, where one part may heat up more quickly than the other.

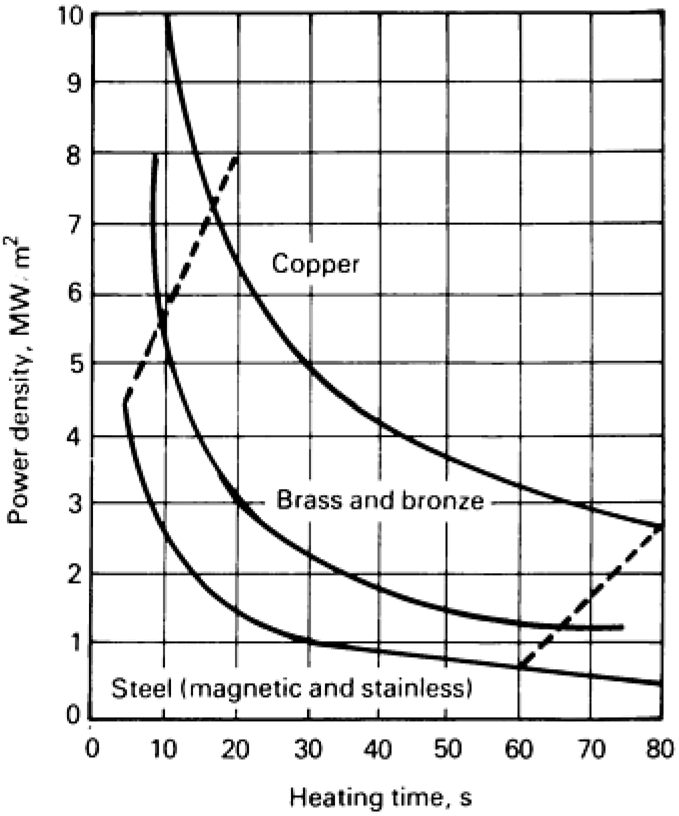

FIG. 1 POWER INPUTS AND HEATING TIMES FOR SELECTED FERROUS AND NONFERROUS MATERIALS. SOURCE: REF 1

Resistivity Effects. The coupling efficiency changes as the material warms up due to the temperature dependence of the electrical and/or magnetic properties. Resistivity increases with temperature for most metals and is the most important concern in nonmagnetic materials. Table 1 shows the resistivity changes in materials over the typical range of soldering temperatures. Unfortunately, changes to the overall coupling efficiency (and therefore heating rate) caused by resistance changes are not easily predicted because part geometry and penetration depth of the electromagnetic field have a synergistic effect.

TABLE 1 RELATIONSHIP OF TEMPERATURE TO ELECTRICAL RESISTIVITY, THE PRODUCT OF SPECIFIC HEAT AND MATERIAL DENSITY, AND THERMAL CONDUCTIVITY FOR SELECTED MATERIALS

TEMPERATURE		ELECTRICAL RESISTIVITY (ρ), 10^{-6} Ω· M	PRODUCT OF SPECIFIC HEAT AND MATERIAL DENSITY ($C\gamma$), 10^6 W · S/M^3 · K	THERMAL CONDUCTIVITY (K), W/M · K
°C	°F			
MILD STEEL (0.23% C)				
20	68	0.160	3.65	52.0
100	212	0.220	3.85	51.0
200	392	0.290	4.10	49.0
300	572	0.380	4.40	46.0
400	752	0.483	4.77	43.0
500	932	0.610	5.19	39.3
STAINLESS STEEL (19.11% CR, 8.14% NI, 0.60% W)				
20	68	0.695	4.04	15.9
100	212	0.776	4.15	16.3
200	392	0.850	4.24	17.2
300	572	0.915	4.36	18.0
400	752	0.976	4.51	19.7
500	932	1.030	4.83	21.4
COPPER				
20	68	0.017	3.39	395
100	212	0.022	3.48	387
200	392	0.033	3.57	380
300	572	0.037	3.65	373
400	752	0.044	3.72	366
500	932	0.052	3.77	360
ALUMINUM				
20	68	0.027	2.52	211
100	212	0.0364	2.59	219
200	392	0.0478	2.65	224
300	572	0.0599	2.71	223
400	752	0.073	2.78	216
500	932	0.087	2.84	209

Source: Ref 1

Curie Temperature Effects. In magnetic materials, a significant reduction in coupling efficiency occurs when the material passes through its Curie temperature upon heating (i.e., changes from ferromagnetic to paramagnetic). Curie points for iron and nickel are 770 °C (1418 °F) and 358 °C (676 °F), respectively (Ref 2). Clearly, only nickel would be of concern at soldering temperatures. The magnitude of the coupling loss depends strongly on the geometry of the part and the coil configuration. Lost coupling efficiency due to resistance or magnetic property variations can be regained by changes to the power supply frequency during the soldering procedure. However, in the case of the Curie transition, the exceptional heating rate of ferromagnetic materials can never be wholly recovered.

Temperature rise in the workpiece is also affected by the thermal mass of the part(s) as reflected by thermal conductivity, specific heat, and size of the part. Heat loss from the joint area due to thermal conductivity as well as excessive part size can lengthen heating times, making it difficult to uniformly heat joints containing dissimilar metals. Thermal mass is often expressed quantitatively by the product of the specific heat (c) and the material density (γ). Values of $c\ \gamma$ are given in Table 1.

References cited in this section

1. P. SIMPSON, *INDUCTION HEATING*, MCGRAW-HILL, 1960
2. F. LOBKOWICZ AND A. MELISSINOS, *PHYSICS FOR SCIENTISTS AND ENGINEERS*, SAUNDERS CO., 1975, P 355

Induction Soldering

Paul T. Vianco, Sandia National Laboratories

Workpiece Geometry

Induction heating is more efficient and more easily practiced on workpieces with simple, axis-symmetric geometries because they, in turn, require simple coil geometries. Electromagnetic energy is more easily coupled into the part due to closer proximity to the workpiece and the generation of a more uniform magnetic field. Uniformity, as well as field strength, improves as the number of turns in the coil is increased. A larger coil also provides a more uniform field but at the expense of delocalizing the heating area. Some sample coil configurations for typical part geometries are shown in Fig. 2. Finally, uniform heating of the workpiece can be achieved in relatively nonuniform magnetic fields by rotating the part within the coil.

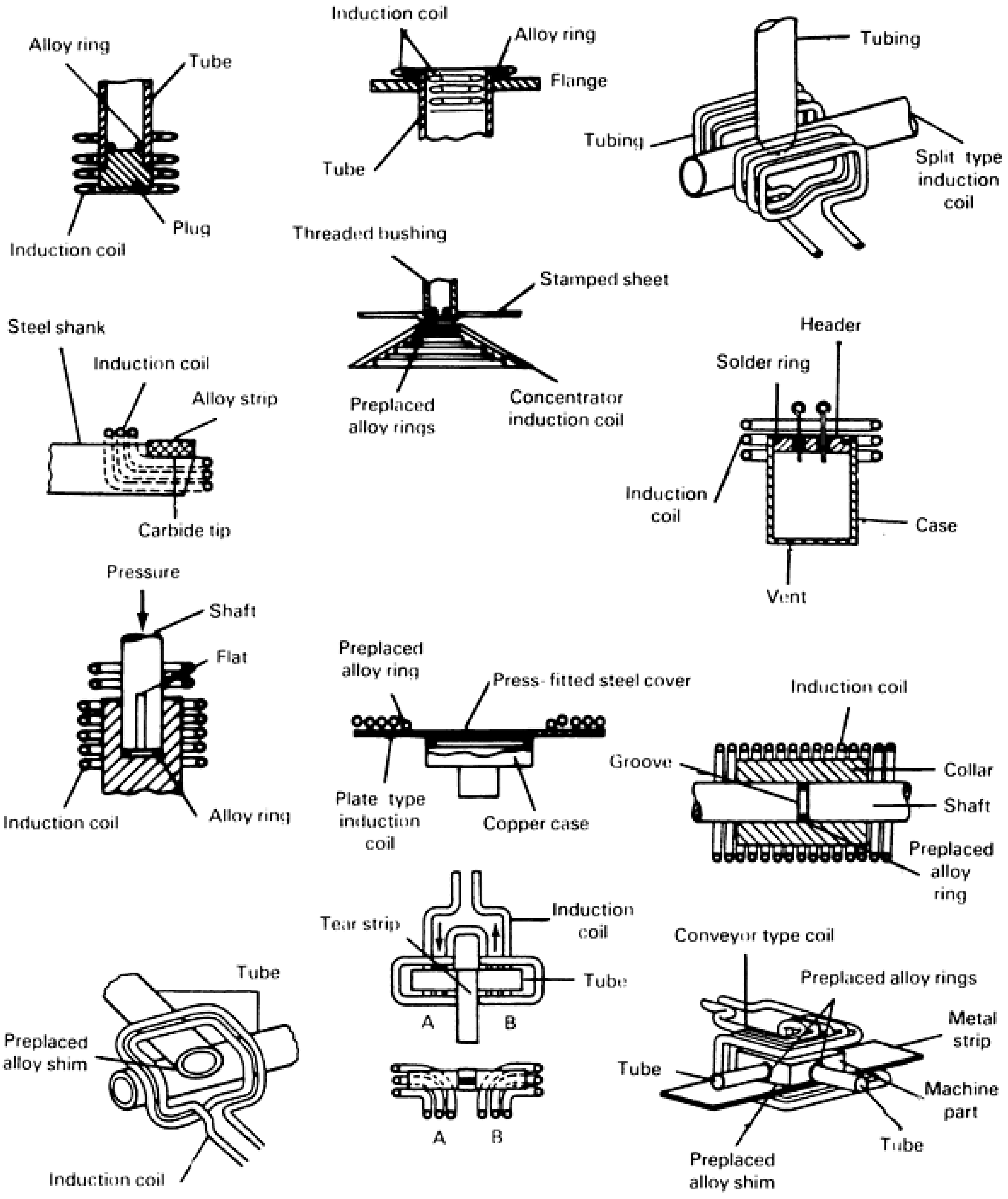
Alloy ring
Tube
Plug
Induction coil
Induction coil
Alloy ring
Flange
Tube
Tubing
Tubing
Split type induction coil
Threaded bushing
Stamped sheet
Preplaced alloy rings
Concentrator induction coil
Steel shank
Induction coil
Alloy strip
Carbide tip
Header
Solder ring
Induction coil
Case
Vent
Pressure
Shaft
Flat
Induction coil
Alloy ring
Preplaced alloy ring
Press- fitted steel cover
Plate type induction coil
Copper case
Induction coil
Groove
Collar
Shaft
Preplaced alloy ring
Tube
Preplaced alloy shim
Induction coil
Tear strip
Induction coil
Tube
A
B
A
B
Conveyor type coil
Preplaced alloy rings
Metal strip
Tube
Machine part
Tube
Preplaced alloy shim

FIG. 2 COIL CONFIGURATIONS FOR TYPICAL JOINT SHAPES. SOURCE: REF 1

Use of induction heating becomes more limited as the workpiece geometry becomes complex. First, the placement of the coils about or near the joint area becomes more difficult. Secondly, corners and protrusive shapes in the magnetic field may develop hot spots caused by proximity to the coil or low thermal mass, thereby possibly damaging the material at those locations. To accommodate complex part configurations, coil shapes can be altered to vary the local magnetic field strength and hence change the localized heating of the part. For example, the overheating of a corner or protrusion can be reduced by locally decreasing the number of coil turns in that area or by positioning the coil farther from the workpiece. The opposite actions apply to part configurations with large thermal mass that require additional power input to prevent heat-sinking effects.

In addition to surrounding the part with the coil to maximize coupling, coils can be used as proximity sources for localized heating. In this technique, a coil (typically smaller than the part) is located next to the joint area to be heated. As expected, the limited magnetic field intensity causes this technique to have very low coupling. The heating area can be extended by manually scanning the coil over a larger area (such as is done in torch heating).

Reference cited in this section

1. P. SIMPSON, *INDUCTION HEATING*, MCGRAW-HILL, 1960

Paul T. Vianco, Sandia National Laboratories

Preplaced Solder

Soldering by induction heating necessarily implies joints designed for the use of solder preforms or paste; these items are introduced into the joint area prior to applying heat. The amount of solder is much less than the mass of the parts, so the effect of the preform on the induction heating process is typically negligible. Space restraints, as well as the danger of electrical shock and severe burns, prohibit manual interaction with the workpiece during heating. Therefore, certain practices associated with the use of preplaced solder must be followed:

- AN ADEQUATE SUPPLY OF SOLDER ALLOY MUST BE PRESENT TO COMPLETELY FILL THE GAP.
- PROVISION MUST BE MADE FOR SOLDER FLOW THROUGH THE JOINT AND THE ESCAPE OF FLUX VOLATILES FROM THE JOINT.
- DESIGN CONSIDERATIONS MUST ALLOW FOR AN INCREASED PROPENSITY OF VOID FORMATION DUE TO LIMITED CAPILLARY FILL WITH PREPLACED SOLDER.
- FIXTURING THAT SECURES THE WORKPIECE DURING SOLDERING MUST MAINTAIN SUBSTRATE ALIGNMENT AND THE JOINT GAP SPECIFIED FOR THE FINAL PRODUCT.

Wetting of the parts can be enhanced by hot solder dipping or by electroplating a coating of tin-lead solder on the joint surfaces prior to final assembly. Although typically performed in air, induction soldering can also be practiced in inert atmospheres or in vacuum (e.g., localized cover gas blankets or more sophisticated equipment such as glove boxes and large chambers). Induction heating is well suited for robotic automation, including bench-top systems as well as larger-scale factory production lines. Monitoring the workpiece temperature by optical pyrometry can provide a feedback signal for real-time control of the induction power source, thereby improving reproducibility of the heating schedule parameters.

Paul T. Vianco, Sandia National Laboratories

Setup Parameters

The heating requirements of some simple workpiece geometries (and accompanying simple coil geometries) have been calculated, and performance parameters have been tabulated in charts (Ref 1). However, maximizing the energy coupled into a particular workpiece is more often the result of trial-and-error experiments with coil configurations, power supply frequency, and voltage. Nonoptimum coupling causes greater current to be drawn from the power supply to achieve the desired heating of the part, possibly exceeding the power capacity of the generator. Joints between dissimilar metals further lessen optimum coupling and diminish efficiency.

Although induction heating may require some trial-and-error efforts to optimize the process for a given product, some preliminary calculations can provide approximate design criteria from which to begin process development.

The first parameter required to set up an induction system is the power required to heat the workpiece(s). For localized heating, heat is lost to the remaining mass of the part as well as through radiation, conduction, and convection with the atmosphere. For a first-order approximation, ambient heat loss can be neglected, particularly with the relatively low temperatures necessary for soldering processes. However, conduction into the solid away from the joint area can be significant; moreover, the calculations to quantify the effect may be very complex. An estimate of the power, ***P,*** required to heat a section of material can be determined from:

$$P = \frac{g\,c\Delta TV}{t} \qquad \textbf{(EQ 1)}$$

where V is the volume to be heated, c is the specific heat of the material, ΔT is the temperature rise, γ is the density of the material, and t is the time for the temperature increase to take place. Of course, the power rating of the induction heating equipment must be higher than that calculated in Eq 1 to account for coupling inefficiencies and the heat loss mechanisms noted above. The overall efficiency of induction heating nonferrous metals is 50 to 75%. Values are approximately 75% for magnetic metals that are below the Curie temperature (i.e., in the ferromagnetic state).

Selection of a power frequency depends on the type of soldering material and the size of the workpiece. Table 2 gives typical frequencies for the through-heating of nonferrous metals. The representative geometry is that of round bar. A similar range of values can also be used for magnetic materials below the Curie temperature. Heating the material above the Curie point requires a frequency increase to maintain reasonable power coupling, particularly as the part size increases (power input to the workpiece is proportional to the square root of the frequency). Frequency values at temperatures exceeding the Curie point are 3 to 10 kHz for part sizes of 5 to 40 mm (0.20 to 1.6 in.) diameter and 1 to 3 kHz for diameters of 40 to 150 mm (1.6 to 5.9 in.).

TABLE 2 FREQUENCIES FOR THE THROUGH-HEATING OF NONFERROUS METAL ROUND BARS

BAR DIAMETER		**FREQUENCY, KHZ**
mm	**in.**	
0-12	0-$\frac{1}{2}$	450
12-25	$\frac{1}{2}$-1	10 AND 3
25-75	1-3	1
>75	>3	50

A parameter of particular interest to induction heating is the penetration depth of heating (also called the skin depth), and this is denoted by the symbol δ. The skin depth is defined as the position in the solid away from the surface at which the magnitude of the magnetic field falls to approximately 37% of its value at the part surface. The skin depth is given by:

$$d = \frac{r}{pm\,f} \qquad \textbf{(EQ 2)}$$

where ρ is the resistivity of the part, $\rho$is the absolute magnetic permeability, and f is the frequency. The absolute permeability is equal to the product $\mu_r\mu_0$, where μ_r is the relative permeability of the material (e.g., pure iron = 18,000 and nonmagnetic copper = 1) and μ_0 is the permeability of "free space," which equals 1.257×10^{-6} V · s/m · A (a universal constant). Skin depth increases as the material resistivity increases and as the frequency decreases. For magnetic materials at temperatures below the Curie point, the value of μ_r is not a constant; it is dependent on the magnetic field intensity, H_0. For example, the value of μ_r for a low-carbon 1044 steel is $734{,}000/H_0^{0.92}$.

Reference cited in this section

1. P. SIMPSON, *INDUCTION HEATING,* MCGRAW-HILL, 1960

Reference cited in this section

1. P. SIMPSON, *INDUCTION HEATING,* MCGRAW-HILL, 1960

Induction Soldering

Paul T. Vianco, Sandia National Laboratories

Safety Concerns

The safety hazards accompanying the use of induction heating include electrical shock from contact with the coils as well as burns caused by the heated workpiece. A third hazard is the danger of steam explosion caused by arcing between the coil and the part. Typically, the heating coils are hollow copper tubing; they are cooled by the internal flow of water or a water-base coolant. The vapor cloud accompanying the spatter of flux during soldering can cause an electric arc that may puncture the copper tubing if the induction coil is close to the workpiece. Arcing can also be caused by movement of the coil windings through forces generated by the magnetic field. Water contacting the hot part produces an immediate, violent steam explosion. To prevent this occurrence, adequate clearance should be maintained between the coil and the part (typical minimum spacing is 6.4 to 13 mm, or 0.25 to 0.50 in.). Moreover, trial heating runs, in which the current is slowly increased to assess coil movement and part-to-coil shorting, should be performed prior to product assembly.

Induction Soldering

Paul T. Vianco, Sandia National Laboratories

References

1. P. SIMPSON, *INDUCTION HEATING,* MCGRAW-HILL, 1960
2. F. LOBKOWICZ AND A. MELISSINOS, *PHYSICS FOR SCIENTISTS AND ENGINEERS,* SAUNDERS CO., 1975, P 355

Wave Soldering

Paul T. Vianco, Sandia National Laboratories

Introduction

WAVE SOLDERING is one of the primary techniques for mass assembly of printed wiring boards involving through holes, surface mount devices, or a combination of these two technologies. A schematic of the wave soldering process is shown in Fig. 1. A solder "fountain" or "wave" is created by a pump located at the bottom of the solder pot; suitable baffles are mounted in the pot to direct the flow of solder into the desired configuration. The printed wiring board is placed onto a conveyor, which brings it into contact with the wave surface. Note from Fig. 1 that the circuit boards travel along the surface of the solder; molten alloy does not flow on top of the board. As the printed circuit board passes on the wave, the solder wets the surface-mount package leads, terminations, and exposed metal surfaces in the circuit board, and also fills plated through holes. This technique can produce several thousand solder joints in a matter of minutes.

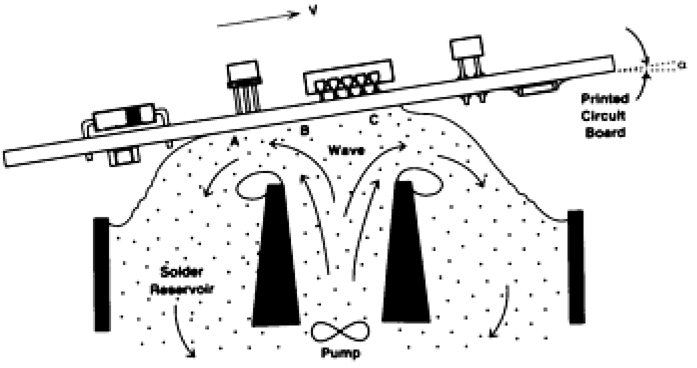

FIG. 1 SCHEMATIC OF THE WAVE SOLDERING PROCESS. THE THREE IMPORTANT PROCESS CONTROL REGIONS ARE THE ENTRY (A), THE INTERIOR (B), AND THE PEEL-BACK REGION (C).

The implementation of wave soldering for surface-mount technology requires that the devices be glued to the substrate prior to wave soldering. Moreover, the entire surface-mount package must be able to withstand contact with the flux and temperatures of the molten solder without a loss of reliability. Small-outline transistors and discrete components (chip resistors, capacitors, etc.) are readily attached by wave soldering; larger packages (e.g., leaded or leadless chip carriers) may be damaged by exposure to the harsh environments (flux and molten solder) as well as be more prone to solder defects (Ref 1).

The wave soldering technique encompasses a sequence of processes, all of which are typically contained in the same apparatus. First, the substrate receives a coating of flux. Flux application methods include:

- A WAVE TECHNIQUE SIMILAR TO THAT SHOWN IN FIG. 1
- COATING THE CIRCUIT BOARD BY A FLUX FOAM CREATED BY PASSING AIR OR NITROGEN THROUGH A FLUX BATH TO GENERATE THE FOAM (OR FROTH) ON THE BATH SURFACE
- DIRECTLY SPRAYING THE FLUX ONTO THE BOARD SURFACE

After flux is applied, the substrate is passed through a preheating stage. Warming the board promotes activation of the flux, accelerates the evaporation of volatiles from the flux (which can cause voids and solder balls by spattering upon contact with the hot solder bath), and reduces thermal shock to the substrate and devices when it passes onto the solder wave. Then, the circuit board contacts the solder wave for the formation of the joints. After passing the wave, the board cools through natural heat loss or, more quickly, by the use of forced air.

Acknowledgement

This work was performed at Sandia National Laboratories and supported by the U.S. Department of Energy under contract DE-AC04-76DP00789.

Reference

1. R. PRASAD, *SURFACE MOUNT TECHNOLOGY: PRINCIPLES AND PRACTICES*, VAN NOSTRAND-RHEINHOLD, 1989, P 426-428

Wave Soldering

Paul T. Vianco, Sandia National Laboratories

Design Considerations and Process Parameters

The successful assembly of a printed circuit board by the wave soldering technique relies on the proper process parameters and the layout of the board for the respective technology (through-hole, surface-mount, or mixed). The hole or via size (diameter) must permit adequate capillary filling by the molten solder and form the joint or to interconnect within the short time period (5 to 7 s) that the board contacts the wave. Holes with small diameters (<0.076 mm, or 0.003 in.) and long length joints with very thin gaps restrict the elimination of flux volatiles, causing voids and insufficient solder quantity. Excessively large holes (>1.0 mm, or 0.039 in.) and gaps (>0.40 mm, or 0.016 in.) decrease the capillary driving force of the solder to fill the volume and create the joint; in some cases, the solder simply drains from the hole under its own weight.

In addition to insufficient solder in holes and vias, other defects in wave-soldered products include bridges, icicles, and skips. Bridges are defects in which a quantity of solder forms a short across two conductors (leads, lands, or traces). A piece of solder that does not short two conductors, but rather hangs from one surface, is an icicle. Skips refer to the failure of solder in the wave to reach the metal surface and form the joint.

Design Considerations

Design considerations, such as the orientation of through-hole leads and surface-mount devices with respect to the direction of travel past the wave, as well as the proximity of packages to 0one another, can dramatically affect yields (Ref 1, 2). Figure 2 is a diagram showing the preferred orientation of through-hole and surface-mount packages.

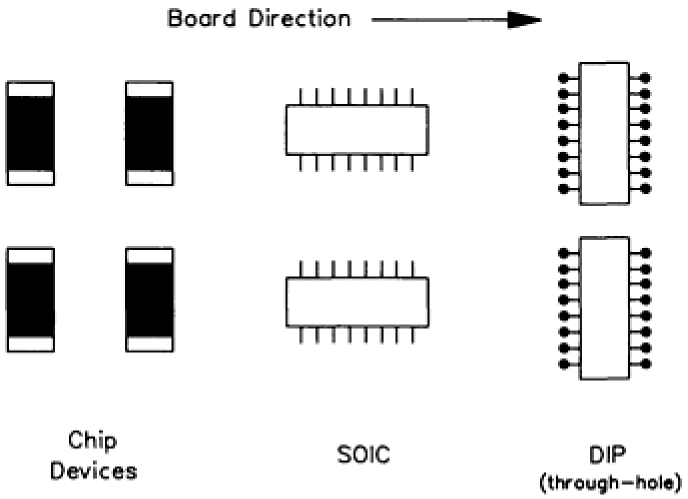

FIG. 2 SCHEMATIC OF PREFERRED DEVICE ORIENTATION FOR WAVE-SOLDERED ASSEMBLIES. SOIC, SMALL-OUT-LINE INTEGRATED CIRCUIT (SURFACE-MOUNT); DIP, DUAL-INLINE PACKAGE (THROUGH-HOLE)

Through-Hole Technology. In the case of through-hole assemblies, the leads on the side of a through-hole 16-pin dual-in-line package (DIP) should be oriented perpendicular to the travel direction of the conveyor to prevent bridging between them. Moreover, the bridging of leaded packages can be limited by maintaining the lead extension beneath the circuit board at less than 1.0 mm (0.039 in.). Leads that are too close to each other also increase the chances of solder bridges. Solder skips are less frequently observed on through-hole products (assuming adequate wettability of the metal surfaces) because the device package body is on the top surface, and therefore does not shadow joints from the wave.

Surface-Mount Technology. The circumstances of solder defect function are different for surface-mount technology because the entire package passes through the wave. Solder joint skips are more common because the package body can shadow nearby leads or terminations from contact with the wave, causing an inadequate supply of solder at the joint. Therefore, the joint areas of the package should be fully exposed to the wave, as suggested by Fig. 2. Packages with leads or terminations on all four sides experience an increased propensity for defects on the leads in the nonpreferred orientation. Also, the greater density of devices on surface-mount circuit boards increases the frequency of solder skips due to the closer proximity of the packages. Finally, the finer lead pitch used on surface-mount leaded and leadless (chip carrier) packages limits or prohibits their assembly on boards by wave soldering.

Process Parameters

The second factor critical to successful implantation of wave soldering of printed circuit boards is the process parameters. Process parameters can be divided into three groups: the fluxing operation, the solder wave properties, and the process schedule.

The fluxing operation not only includes the particular parameters (wave or foam height, spraying pressure, etc.) needed to ensure adequate flux supply to the board for solder wetting, but also the condition of the flux. In-line fluxing techniques (wave or foam) frequently use the same quantity of flux to process a large number of boards. Deterioration of the flux due to board contaminants, as well as loss of the flux vehicle, must be monitored to minimize defects in the final product. Specific gravity is used to indicate flux condition.

Low-solids fluxes are gaining popularity as a "no-clean" alternative to rosin-base fluxes, including their use in wave soldering. The low-solids contents (typically less than 5%) make these fluxes difficult to foam, so spray or wave applications are preferred (Ref 3). Tighter control of the flux chemistry, reduced solder pot temperatures (by 10 to 15 °C, or 18 to 27 °F), shorter solder dwell times (by 1 to 2 s), and the use of inert or reducing atmospheres (Ref 4) can reduce the number of defects.

Solder Wave Properties. The second process factor is the solder wave properties. The simple wave geometry (also called the T-wave) is shown in Fig. 1. Three regions of the wave are identified. The entry region (on the left with the board moving to the right) should be sufficiently turbulent to supply solder to all of the exposed metal surfaces and should permit the escape of flux volatiles that may cause voids in the joint. In the case of surface-mount assemblies, fluid motion must be sufficiently active to force solder close to the packages for adequate wetting of terminations and leads, thereby avoiding solder skips. The interior region of the wave acts as the heat source and solder supply to promote wetting of the substrate features. The exit or peelback region is the point at which the board leaves the wave and is critical in preventing solder bridge and icicle defects on the circuit board. For example, defects are reduced by having a smooth wave that separates very slowly from the board. This effect is achieved by matching the solder flow velocity with the speed of the circuit board (Ref 5) as well as by making adjustments to the wave configuration. Hot air knives positioned after the wave have been used to remove solder bridges and icicles prior to solidification of the solder.

Figure 3 shows several equipment modifications used to produce different wave geometries. The extended T-wave and lambda-wave modifications to the standard T-wave cause a smooth surface at the exit region to limit bridges, icicles, and the formation of excessive dross. The dual-wave configuration provides separate turbulent and smooth (laminar) waves to effect the desired solder joint qualities. Note that these modifications generally require that the circuit board laminate and devices withstand a longer period of contact with the molten solder.

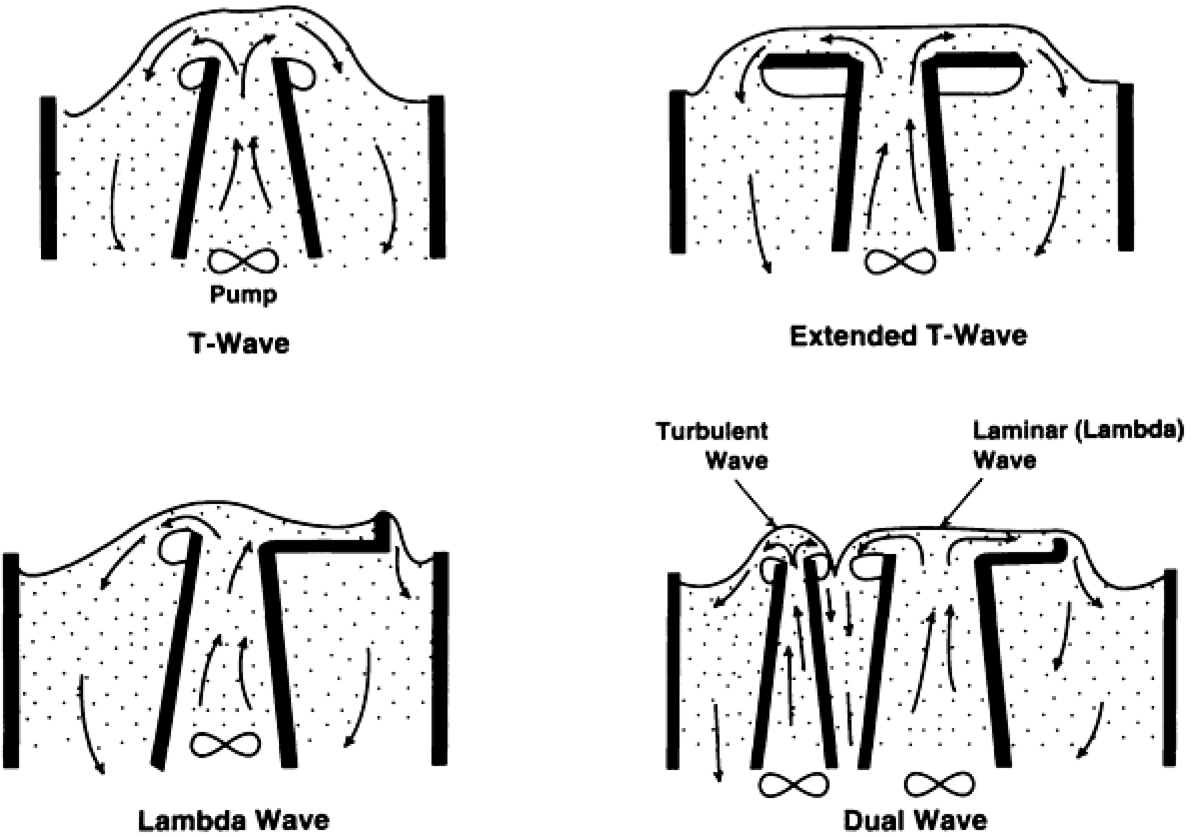

FIG. 3 EQUIPMENT MODIFICATIONS USED TO PRODUCE DIFFERENT WAVE GEOMETRIES. THE NOMENCLATURE DESCRIBES THE PROFILE GEOMETRY OF THE WAVE.

Dross formation tends to be high in wave soldering due to the large surface areas of solder exposed to the atmosphere (which is further aggravated by the turbulent flow of the solder in the pot). Dross particles can become embedded in the solder joints as well as disturb the peel back of the solder upon exiting the board. The use of inert atmospheres, reduced solder temperatures, and oils to cover the solder surface limit the formation of dross.

Solder temperatures are typically higher than those used in reflow processes. The higher temperatures are required to provide sufficient heat transfer to the substrate for adequate wetting as well as to reduce the solder surface tension to limit defects at the peel-back region. Solder pot temperatures are typically 50 to 80 °C (90 to 145 °F) above the liquidus temperature of the alloy; higher values are required for thicker substrates (e.g., multilayer or controlled-expansion substrates). The potential for damage to the laminate, including warpage and distortion as well as delamination of the conductor, increases with higher solder temperatures. Continuous processing with wave soldering equipment can cause a build-up of contaminants in the solder pot (e.g., nickel, copper, and gold from circuit board and lead finishes). See the article "Soldering in Electronic Applications" in this Volume for guidelines on acceptability limits for solder pot contamination.

In addition to the traditional tin-lead alloys, wave soldering has been adapted to other solders, including the eutectic tin-bismuth solder and the tin-lead-bismuth alloys (Ref 6).

Process Schedule. The third factor in the wave soldering technique is the process schedule, which includes the preheat zone temperature profile, the solder wave temperature, the conveyor tilt angle (α in Fig. 1, which affects the exit geometry of the board), and the conveyor speed. The conveyor speed determines the dwell time of the circuit board in the preheat zone and solder wave. Although specific settings depend on the type of equipment and the product being assembled, some general guidelines can be applied. First, the preheat stage should bring the board temperature to the range of 100 to 150 °C (210 to 300 °F). These temperatures provide suitable activation of the flux and lessen thermal shock to the circuit board (and devices) upon contact with the solder wave. The dwell time in the molten solder wave should be limited to 3 to 5 s. Dwell times at the molten solder temperature can reach 5 to 10 s for dual-wave systems. Time-temperature exposure limits for soldering processes are typically specified for devices and the laminate; manufacturer recommendations should always be followed. Also, dwell times should be minimized to limit the leaching of conductor or lead finishes (e.g., copper, nickel, or gold) into the solder.

References cited in this section

1. R. PRASAD, *SURFACE MOUNT TECHNOLOGY: PRINCIPLES AND PRACTICES,* VAN NOSTRAND-RHEINHOLD, 1989, P 426-428
2. R. GENGLER AND J. HABIB, MINIMIZING DEFECTS IN THE MASS SOLDERING OF PRINTED WIRING ASSEMBLIES, *WESTERN ELECTRIC ENGINEER,* 1983
3. D. KOCKA, NO-CLEAN FLUXES ARE A VIABLE ALTERNATIVE TO CFC CLEANING,*ELECT. PACK. PROD.,* JUNE 1990, P 95
4. P. FODOR AND P. LENSCH, COVER GAS SOLDERING LEAVES NOTHING TO CLEAN OFF PCB ASSEMBLY, *ELECT. PACK. PROD.,* APRIL 1990, P 64
5. R. BOTHAM, C. LOWELL, AND J. STERRITT, WAVE SOLDERING MIXED TECHNOLOGY BOARDS, *ELECT. PACK. PROD.,* NOV 1990, P 28
6. M. NYLEN, K. JOSEFSON, AND H. STEEN, USE OF TIN-LEAD-BISMUTH ALLOY AS A SUBSTITUTE FOR EUTECTIC TIN-LEAD IN WAVE SOLDERING, *BRAZING AND SOLDERING,* SPRING 1988, P 38

Wave Soldering

Paul T. Vianco, Sandia National Laboratories

Defects

Certain defect trends can be identified with particular process conditions (Ref 7). For example, voids and blowholes are more frequent at fast conveyor speeds and higher solder temperatures. These process conditions generate an excessively fast heating rate, which causes the rapid volatilization of the flux components and subsequent void formation. Rapid conveyor speed can be responsible for several other defects:

- AN INCREASED PROPENSITY FOR POOR FILLING OF HOLES AND VIAS CAUSED BY INSUFFICIENT DWELL TIME IN THE SOLDER WAVE. AS A RESULT, THE WETTING KINETICS MAY NOT FILL THE VOLUME (THIS IS FURTHER EXAGGERATED BY LOW SOLDER TEMPERATURES)
- THE PRESENCE OF SOLDER BRIDGES AND ICICLES DUE TO POOR MATCHING OF THE SUBSTRATE SPEED AND FLOW OF THE SOLDER WAVE AT THE PEEL-BACK REGION UPON BOARD EXIT
- CRACKING OF SURFACE-MOUNT CERAMIC CHIP CAPACITORS CAUSED BY THERMAL SHOCK

High solder pot temperatures and slow conveyor speeds (i.e., long dwell times) can create difficult-to-remove flux residues. Also, excessive warpage or distortion of the board results from inadequate support of the substrate by the conveyor system; this is particularly acute with thinner laminates, higher soldering temperatures, or longer dwell times (e.g., dual-wave systems). Thicker circuit boards (backplanes) may also warp due to the large thermal gradients through the thickness of the laminate due to inadequate preheating of the board.

Clearly, the various process parameters can interact to produce certain defects. Therefore, optimization experiments and statistical process control are required to obtain and maintain acceptable product yields. The qualitative guidelines cited above provide a first measure of correlating defects to process conditions for subsequent improvement to the assembly procedure.

Reference cited in this section

7. C. LEA, *A SCIENTIFIC GUIDE TO SURFACE MOUNT TECHNOLOGY,* ELECTROCHEMICAL PUBLISHERS, LTD., 1988, P 148-153

Wave Soldering

Paul T. Vianco, Sandia National Laboratories

References

1. R. PRASAD, *SURFACE MOUNT TECHNOLOGY: PRINCIPLES AND PRACTICES,* VAN NOSTRAND-RHEINHOLD, 1989, P 426-428
2. R. GENGLER AND J. HABIB, MINIMIZING DEFECTS IN THE MASS SOLDERING OF PRINTED WIRING ASSEMBLIES, *WESTERN ELECTRIC ENGINEER,* 1983
3. D. KOCKA, NO-CLEAN FLUXES ARE A VIABLE ALTERNATIVE TO CFC CLEANING,*ELECT. PACK. PROD.,* JUNE 1990, P 95
4. P. FODOR AND P. LENSCH, COVER GAS SOLDERING LEAVES NOTHING TO CLEAN OFF PCB ASSEMBLY, *ELECT. PACK. PROD.,* APRIL 1990, P 64
5. R. BOTHAM, C. LOWELL, AND J. STERRITT, WAVE SOLDERING MIXED TECHNOLOGY BOARDS, *ELECT. PACK. PROD.,* NOV 1990, P 28
6. M. NYLEN, K. JOSEFSON, AND H. STEEN, USE OF TIN-LEAD-BISMUTH ALLOY AS A SUBSTITUTE FOR EUTECTIC TIN-LEAD IN WAVE SOLDERING, *BRAZING AND SOLDERING,* SPRING 1988, P 38
7. C. LEA, *A SCIENTIFIC GUIDE TO SURFACE MOUNT TECHNOLOGY,* ELECTROCHEMICAL PUBLISHERS, LTD., 1988, P 148-153

Vapor-Phase Soldering

Dale L. Linman, Centech Corporation

Introduction

VAPOR-PHASE SOLDERING is a process of condensation heating, in which a product prepared for soldering is passed through or into a layer of saturated vapor. The vapor condenses on the relatively cool part, which transfers the latent heat of vaporization to the part and heats it rapidly and uniformly. The process reaches thermal equilibrium in a few seconds at a temperature that is only limited by the boiling point of the fluid used in the process. Typically, newer equipment utilizes a single perfluorocarbon fluid vapor layer, which does not contain chlorine or bromine and is environmentally safe. This fluid is also chemically inert, thermally stable, and has very low solvent action.

Although the process is used in a variety of heating applications, such as curing epoxies and polymers or stress relieving various platings, it is primarily useful when soldering surface-mounted components to various substrate materials. Because of the precise temperature control that this process offers (by virtue of a fixed boiling point), multilevel soldering is possible using up to three solder compositions with three different melting temperatures. Fluid temperatures that range from 100 to 265 °C (212 to 510 °F) are available. The most common fluid used boils at 215 °C (420 °F) and is used for 63-37 type solder, which melts at 183 °C (360 °F). A 30 °C (55 °F) differential is typical, to enable the heating process to proceed rapidly. Temperature precision also prevents the assembly from overheating.

Vapor-Phase Soldering

Typical Materials

The typical materials that are joined by this process are solder-coated substrates made from various ceramic materials and G10 or FR4 epoxy laminates mated with surface-mounted components. Solder paste is typically applied to the substrate by screening, stenciling, or using *X-Y* dot placement. Components are then placed in the correct position in the solder paste, which temporarily holds them until they are soldered. The assembly is then passed through the vapor-phase setup, where the solder is melted, joining the components to the substrate. Typical surface-mount lead density is 1.25 mm (0.050 in.) on centers. Fine-pitch components as small as 0.38 mm (0.015 in.) on centers can also be processed. There is no inherent lead-pitch limitation in the vapor-phase process, because it is essentially an oxygen-free soldering process. Because of its uniformity, rapid heating, precise temperature control, and oxygen-free environment, the process is preferred for high-value assemblies. Also, because of the thermal capacity, ease of profiling, and forgiving nature of the process, it is also used when high throughput is required from limited floor space or when a large number of different types of boards need to be run in a given time period.

Equipment is normally configured for either batch or in-line operations (Fig. 1 and 2). The majority of machines available are of single-vapor design (Fig. 3). Earlier models had vertical access and used less-expensive fluid as a cover layer. These machines also lacked a preheating capability, which was found to be necessary in order to reduce the temperature differential at the point of solder melting. This measure prevents wicking and other undesirable characteristics that can develop when one part of a solder joint heats faster than the other. Both a preheating capability, typically provided by infrared panels, and a cooldown stage have been incorporated in newer models. The cooldown stage brings the solder joint through the liquidus phase more quickly (Fig. 4). This procedure reduces the time that the joint can form undesirable intermetallic layers and simultaneously improves the grain structure of the joint, thus improving its capability to withstand thermal cycling.

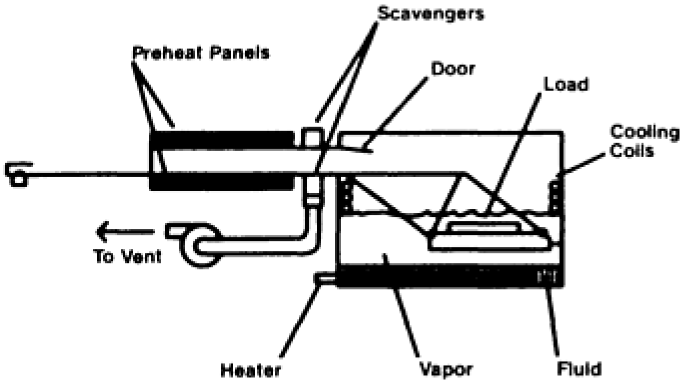

FIG. 1 PELLETIZED VAPOR-PHASE BATCH SYSTEM

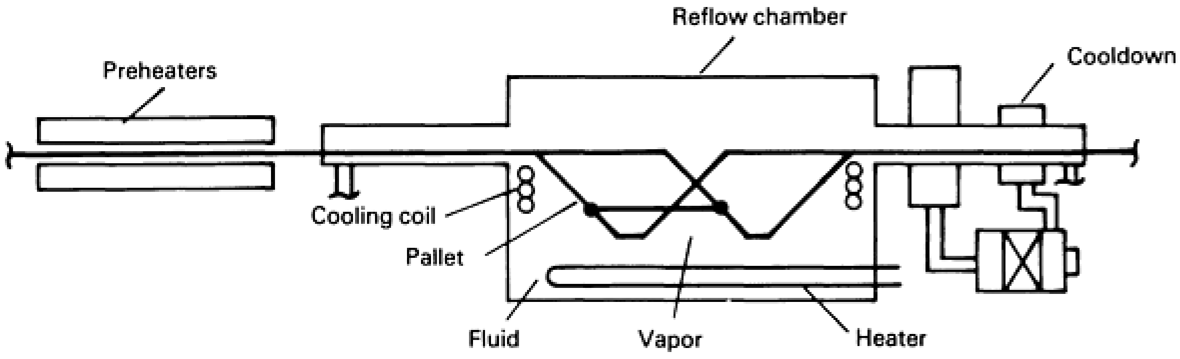

FIG. 2 THIRD-GENERATION VAPOR-PHASE REFLOW SYSTEM

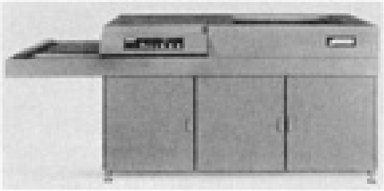

FIG. 3 SINGLE-VAPOR BATCH-TYPE EQUIPMENT

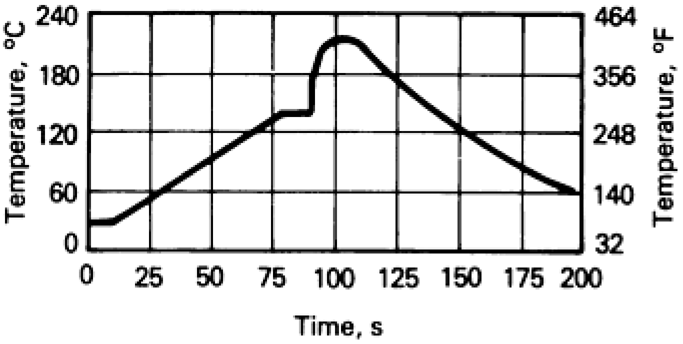

FIG. 4 THERMAL PROFILE

The stability of the vapor-phase process and its ease of profiling means that machine operators require little training. If the overall process is established correctly, then there are few elements that require adjustments. A smooth, vibration-free conveying mechanism enables defect-free soldering and assembly.

The proper setup procedure is to attach a thermocouple to a strategic location on the printed board, such as a sensitive component or a solder joint. Then, the board is processed to reach a preheat temperature of between 125 and 150 °C (255 and 300 °F) in 50 to 60 s. A desirable heating rate is approximately a 2 °C/s (4 °F/s) rise to within 100 °C (180 °F) of the reflow temperature. Major parameters are dwell time (batch machines), conveyor speed (in-line machines), and preheat panel temperature. Reflow time is established by observing the solder melt through the window in the chamber. In the batch machine, the dwell times can be set from those obtained and the process replicated. With in-line machines, the reflow time is established by the conveyor speed and the preheat temperature is adjusted with the panel temperature. Similar board assemblies can be processed using very similar profiles, which minimizes lost machine time due to the complex repetitive profiling that is common with other reflow methods.

Because vapor-phase soldering depends on the series of processes that precede it, every parameter leading to the actual soldering process must be carefully controlled to ensure a high yield. For example, thickness, viscosity, and location accuracy are just some of the parameters involved in solder paste application. Pad design and board layout are equally important to the process, as is the accurate placement of components. With proper controls, it becomes possible to solder with defect rates under 50 ppm using the vapor-phase process. Information on guidelines for parameter control can be obtained by contacting the Institute for Interconnecting and Packaging Electronic Circuits.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Introduction

THE SELECTION of materials for welded construction applications involves a number of considerations, including design codes and specifications where they exist. Mobile structures, such as automobiles and aerospace vehicles, have quite different materials requirements for weight, durability, and safety than stationary structures, such as buildings and bridges, which are built to last for many years. In every design situation, economics--choosing the correct material for the life cycle of the part and its cost of fabrication--are of great importance. Design codes or experience frequently offers an adequate basis for material selection; for new or specialized applications, however, the engineer encounters problems of an unusual nature and thus must rely on basic properties of the material, such as strength, corrosion or erosion resistance, ductility, and toughness. The properties of the various metallurgical structures associated with the thermal cycles encountered in the welding operation must also be included in the design process. The various subsections in this article offer guidance for material selection applications involving bridges and buildings, pressure vessels and piping, shipbuilding and offshore structures, aerospace systems, machinery and equipment, automobiles, railroad systems, and sheet metal. Material properties and welding processes that may be significant in meeting design goals are described.

Note

* R. DAVID THOMAS, JR., *CHAIRMAN*, R.D. THOMAS & COMPANY; BRUNO L. ALIA; WILLIAM R. APBLETT. AMET ENGINEERING; ROBERT G. BARTIFAY, ALUMINUM COMPANY OF AMERICA; STEPHEN A. COUGHLIN, ACF INDUSTRIES, INC.; GREGORY MELEKIAN, GENERAL MOTORS CORPORATION; ANTHONY R. MELLINI, ST., MELLINI AND ASSOCIATES, INC.; LARRY PERKINS, WRIGHT LABORATORY; JAMES E. ROTH, JAMES E. ROTH, INC.; WILLIAM J. RUPRECHT, GENERAL ELECTRIC COMPANY; BERNARD E. SCHALTENBRAND, ALUMINUM COMPANY OF AMERICA; ROBERT E. SOMERS, SOMERS CONSULTANTS; ROGER K. STEELE, AAR TECHNICAL CENTER

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Overview

Principles of Material Selection

Structural Integrity. Before a structure is initially placed in service, the properties of its materials of construction must be adequate to ensure that the structure will perform as intended. Its components must be strong enough to bear the loads that will be imposed. It must be capable of meeting various inspection and qualification tests required by the end user to measure base material and weld soundness, ductility, toughness, and strength.

Length of Service. The service environment is often the largest influence on material selection. Materials of construction rarely last indefinitely. One of the principal selection considerations is the intended length of service for the structure. The designer needs to choose materials that will meet the conditions imposed initially and that can adapt to possible changes in conditions during the life of the structure.

The length of service is often dictated by obsolescence. Military hardware fabricated during wartime is an example of structures with relatively high obsolescence and short service life. Rocket cases for the propulsion of space vehicles need to last one launching. At the other end of the spectrum, hydroelectric projects have expected lifetimes of a century or longer.

Materials or components rarely reach the end of their useful lives at the same time. Although a hydroelectric project is designed to last 100 years, some components may reach the end of their expected lives in less time, either by obsolescence due to improved technology or by unavoidable wear. At the outset of the material selection process, the length-of-service criterion must take into consideration the following: inspectability for potential failure, ease of maintenance of areas subject to deterioration, and ease of replacement of obsolete or life-exhausted components.

Joining Process. The suitability of the materials for the fabrication process to be employed must be considered early in the selection process. Weldability is the property of a material that dictates its ease of joining. The term most often applies to various types of steels joined by arc welding processes. This property often determines whether preheat is needed for successful welding and whether some type of postweld heat treatment is required to restore properties degraded by the welding operation.

Although weldability usually relates to the joining of relatively thick plates, the choice of materials for sheet metal applications must also take the welding process into account. The thinner the metal, the more difficult it is to arc weld. Brazing, soldering, or resistance welding is often used to join thin sheet materials. Resistance welding is frequently employed in the assembly of automobile bodies, for example; however, this process may not perform as reliably on coated steels as it does on uncoated steels.

Design Codes. Material selection is often dictated by the codes by which the design is governed, but there is still frequently a choice of materials. For example, pressure vessel codes offer a choice of steels, depending on the service temperature. Structural codes for buildings and bridges allow the use of both plain carbon and low-alloy steels. However, the codes tend to narrow the choice of materials.

Economics. An overriding selection criterion is often cost. In the broadest sense, this can be determined by the price of the material. However, a designer must assess not only the unit cost of the selected material, but also the costs imposed by fabrication. Compared with unalloyed steel, a low-alloy steel may allow a structure to be built using 25% less steel at a nominal 15% increase in unit price, but the cost of fabrication, such as the need for preheat and postweld heat treatment, may outweigh the material saving. A more costly steel that does not require painting or other protection from atmospheric corrosion may be a more economical choice, if not for initial construction, then for the life of the structure.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Service Conditions

The service environment is a principal determinant in the selection of materials for welded construction applications. The physical, mechanical, and corrosion properties of a material must be considered (Table 1).

TABLE 1 PRINCIPAL ATTRIBUTES OF A MATERIAL

PHYSICAL PROPERTIES

- DENSITY
- MELTING POINT
- THERMAL EXPANSION
- ELECTRICAL CONDUCTIVITY
- FERROMAGNETISM

MECHANICAL PROPERTIES

- TENSILE STRENGTH
- IMPACT STRENGTH
- FATIGUE STRENGTH
- CREEP STRENGTH
- DUCTILITY
- HARDNESS
- FRACTURE TOUGHNESS
- ELASTIC LIMIT

CORROSION PROPERTIES

- GENERAL CORROSION
- PITTING ATTACK
- STRESS CORROSION
- EROSION
- CAVITATION
- OXIDATION

Ambient Temperatures and Pressures. Most welded structures are designed to work in normal atmospheric conditions. The properties of common materials are well known for such conditions, and ample data are available to assist the designer in material selection.

Elevated Temperatures. The chemical industry has many processes that operate at temperatures of 1000 °C (1830 °F) and higher. Fossil fuel electric power generation makes use of steam temperatures from 500 to 600 °C (930 to 1110 °F). At such temperatures, most steels will exhibit creep--that is, a gradual deformation in the direction of the principal stresses--and thus the creep strength of a material becomes the controlling property. Creep strength determines the unit stress allowed to minimize the rate of creep and hence the permanent distortion and ultimate failure of a component.

Materials operating at high temperatures are frequently exposed to corrosive, erosive, or oxidizing environments, which may determine the life span of a structure. The resistance of proposed materials of construction to these types of environments should be examined during the selection process.

In processes where heat transfer occurs, the thermal conductivity of a material becomes an important consideration. Allowance must be made in the design for the expansion and contraction of components during heating and cooling--phenomena related to the coefficient of thermal expansion of a material.

Low Temperatures. Most ferritic metals become brittle at low temperatures, losing their ability to deform plastically to adjust to localized stresses and to accept a sudden shock. Thus, materials for use at low temperatures must have good toughness, a property usually measured by the Charpy V-notch impact test. For low-temperature applications, as for elevated-temperature service, the physical properties of thermal conductivity and coefficient of expansion become important.

High Pressures. Structures intended to contain high pressures are usually weldments and are designed according to American Society of Mechanical Engineers (ASME) codes for pressure vessels and piping. Strength properties are of primary importance. The codes specify the unit stress allowed for pressure-containing components. In cylindrical vessels, the maximum stress is in the tangential (i.e., hoop) direction, and the axial stress is half the hoop stress. In its simplest form, hoop stress is determined by:

$$S = \frac{PD}{2T}$$

where S is unit stress, P is pressure, D is the diameter of the cylinder, and T is wall thickness.

The tensile strength of a material determines the allowable stress and is specified in the ASME codes. The codes for steels allow a stress in excess of two-thirds of the minimum yield strength but less than 90% of the minimum yield strength. The minimum wall thickness, T_m, is then determined by:

$$T_m = \frac{PD}{2S + P_y} + A$$

where the extra thickness allowance, A, provides for anticipated corrosion or wear, and the coefficient y is specified by the code and varies with the temperature of operation and the wall thickness relative to the diameter.

Vacuum Conditions. Materials that are selected for very-low-pressure applications are primarily concerned with the tightness of the joints to prevent leakage from the exterior higher (or ambient) pressures into the vacuum chamber. For welded containers, the properties of the weld metal are critical to ensure lack of defects, such as porosity, cracks, and microfissures.

Vacuum conditions exist in outer space. In such applications, the vacuum condition is on the exterior of the vessel, and design for pressure containment is similar to that described for high pressures. Material selection also requires consideration of dynamic loads experienced during launches (hence, toughness), density for light weight (as in aircraft and aerospace applications), and corrosive and erosive effects, especially for reentry into the atmosphere.

Underwater Applications. A wide variety of material properties are involved in the application of weldments under water. Corrosion resistance is frequently a major concern, especially in seawater. Ships and offshore drilling platforms are notable examples.

For ships and for piping operating in biologically active waters, microbiologic organisms pose an unusual problem in material selection. Barnacles that attach themselves to the hulls of ships greatly increase the resistance to propulsion. The use of copper cladding or paints containing copper-bearing compounds is one means of discouraging barnacle formation. In turbid waters, piping is sometimes attacked by microorganisms that form surface pits, causing cracks in areas of local tension stresses. Because high tensile stresses frequently occur in or adjacent to welds, failures of weldments in these conditions can be attributed to these organisms, even in materials that otherwise are considered corrosion resistant to such environments.

Corrosion, Erosion, and Oxidation. Corrosion generally refers to surface attack by aqueous solutions or organic chemicals in either the liquid or vapor phase. Erosion implies the mechanical removal of metal from the surface, sometimes by high-pressure impingement of a fluid, in which case it is termed cavitation. Oxidation commonly occurs in metals at elevated temperatures and results in the removal of metal in the form of oxides.

Data on the resistance of materials to deterioration in many of these environments are frequently available from handbooks and material suppliers. In the case of weldments, the performance depends largely on experience gained in a variety of applications. The weld metal has a structure distinctly different from that of the parent metal, and sometimes a different composition as well. An electrochemical potential may exist between the weld and parent metal structures, causing attack on one or the other (frequently the weld metal because of its lower surface area).

Corrosive attack in weldments often is attributed to stress-corrosion cracking (SCC). In the presence of chloride ions (even at levels of parts per million), high tensile stress zones in or adjacent to welds can cause failure in materials whose corrosion life in the absence of stress is more than adequate for a particular application.

The welds in components carrying fluids at high velocity may erode during service. Under such conditions, hardness is the critical property. Localized areas of components, such as valve seats and turbine blades, may be protected from erosion by applying a hard weld metal, known as hard surfacing. For more information, see the article "Hardfacing, Weld Cladding, and Dissimilar Metal Joining" in this Volume.

In high-temperature service, allowance must be made for the thinning of metal sections due to progressive oxidation. In some applications the initial oxide layer protects the underlying metal, and progress is arrested. Adjacent to welds, oxidation can form preferentially in the highly stressed zones adjacent to welds and progress through the wall, causing weakness and possible failure.

Fatigue. In weldments subjected to fluctuating stresses, the fatigue strength of a metal may be the primary consideration in material selection and design loads. In the absence of flaws, the fatigue strength of most metals used for dynamically loaded structures is considerably higher than that of the welds. Designs for such structures must take into consideration both service experience and laboratory experiments dealing with various types of welds.

Weld metal cracks (even fine microfissures), lack of penetration or lack of fusion at the interface, and geometric discontinuities on the surface (such as overlaps and undercuts) are the major causes of poor performance in fatigue tests of welds. In the absence of such flaws, the fatigue properties of the weld metal itself or of the weld heat-affected zone (HAZ) rarely need to be considered, even though these zones may be inferior compared with the wrought parent metals. Design allowances for fatigue of weldments primarily take into account potentially damaging flaws.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Effect of Weld Thermal Cycles

Thermal Excursions During Welding. The thermal cycles encountered during the welding operation must be considered when selecting material for weldments. To make a weld requires that temperatures in the fusion zone reach above the melting point of the metal. Zones adjacent to the fusion region reach varying peak temperatures, depending on their distance from the fusion zone. Each of these thermal cycles causes metallurgical changes that affect the properties of the metal.

In the weld metal itself, the dendritic structure formed during solidification in a single thermal cycle is similar to that of a casting. In multipass welding, which is common when arc welding thicknesses greater than 6.4 mm ($\frac{1}{4}$ in.), the weld metal will undergo several thermal cycles, a process that refines the dendritic structure and alters the weld properties.

Similarly, in the HAZs immediately adjacent to the weld, one or more heat cycles will be experienced, causing changes in properties. When selecting materials for welded construction, the properties of these HAZs must be considered.

Cooling from the peak temperatures causes residual stresses, most often of a magnitude matching the yield strength of the material. Residual stresses may add to the service stresses. In cases where residual stresses may influence the integrity of the structure during its life expectancy, the designer may specify a postweld heat treatment.

Postweld Heat Treatments. Many weldments are placed in service in the as-welded condition. Under certain circumstances, postweld heat treatments may be required, thereby changing the properties of the selected metal.

Stress relieving is a thermal cycle that takes a weldment up to a high temperature where its yield strength is substantially reduced, allowing localized plastic deformation of the highly stressed regions. In ferritic steels, most stresses are relieved by heating to a temperature of less than 720 °C (1330 °F) for carbon steels (note that alloy steel values are higher). This is below the temperature transformation of ferrite to austenite, known as the critical temperature. This softens the structures that have been hardened by fast cooling from above the austenite-forming temperatures. Such subcritical heat treatments are said to "temper" the steel.

To cause the structures to become more similar to that of the parent metal, a supercritical heat treatment is sometimes specified. The properties of such welds are then likely to be comparable to those of the parent metal, provided that the compositions are similar.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Bridges and Buildings

This section deals with large welded structures built of plain carbon steels--the tonnage items in the construction industry. These structures generally conform to established codes.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Material Selection

The four code-writing organizations listed in Table 2 specify the chemical compositions, tensile properties, and other pertinent characteristics of steels used for large welded structures. Most employ the steel specifications of ASTM.

TABLE 2 CODES FOR STEEL STRUCTURES

ORGANIZATION	TITLE	SPECIFICATION NO.
AMERICAN WELDING SOCIETY (AWS)	"STRUCTURAL WELDING CODE--STEEL"	D1.1
	"BRIDGE WELDING CODE"	D1.5
AMERICAN INSTITUTE OF STEEL CONSTRUCTION (AISC)	"DESIGN, FABRICATION, AND ERECTION OF STRUCTURAL STEEL FOR BUILDINGS"	. . .
AMERICAN ASSOCIATION OF STATE HIGHWAY AND TRANSPORTATION OFFICIALS (AASHTO)	"STANDARD SPECIFICATION FOR WELDING STRUCTURAL STEEL HIGHWAY BRIDGES"	. . .
AMERICAN RAILWAY ENGINEERS ASSOCIATION (AREA)	"MANUAL FOR RAILWAY ENGINEERING"	. . .

The "Structural Welding Code--Steel" (AWS D1.1) considers three classes of structures:

- STATICALLY LOADED (BUILDINGS)
- DYNAMICALLY LOADED (BRIDGES)
- TUBULAR (OFFSHORE AND ARCHITECTURAL)

Fifteen ASTM specifications are applicable to buildings, 13 to bridges, and 34 to tubular structures, including four of the American Petroleum Institute (API) and two of the American Bureau of Shipping (ABS). The six most commonly used ASTM specifications, applicable to all three classes, are:

NO.	TITLE
A 36	"SPECIFICATION FOR STRUCTURAL STEEL"
A 500	"SPECIFICATION FOR COLD-FORMED WELDED AND SEAMLESS CARBON STEEL STRUCTURAL TUBING"
A 501	"SPECIFICATION FOR HOT-FORMED WELDED AND SEAMLESS CARBON STEEL STRUCTURAL TUBING"
A 514	"SPECIFICATION FOR HIGH-YIELD STRENGTH, QUENCHED AND TEMPERED ALLOY STEEL PLATE, SUITABLE FOR WELDING"
A 572	"SPECIFICATION FOR HIGH-STRENGTH LOW-ALLOY COLUMBIUM-VANADIUM STEELS OF STRUCTURAL QUALITY"

A 588	"SPECIFICATION FOR HIGH-STRENGTH LOW-ALLOY STRUCTURAL STEEL WITH 50 KSI (345 MPA) MINIMUM YIELD POINT TO 4 IN. THICK"
A 852	"STANDARD SPECIFICATION FOR QUENCHED AND TEMPERED LOW-ALLOY STRUCTURAL STEEL PLATE WITH 70 KSI YIELD STRENGTH TO 4 IN. THICK"

These specifications vary considerably in properties important to notch sensitivity (toughness) and weldability. The summary in Table 3 provides a guide as to what to look for in these and other specifications referred to in the structural codes. The supplementary requirements in some of these specifications are useful to the designer and fabricator when evaluating the need for additional acceptance tests for unusual applications.

TABLE 3 PROPERTIES RELATING TO SELECTION FOR SERVICE OF ASTM STEELS APPROVED IN AWS D1.1-92

Includes only the seven steels approved for all three AWS structural classes: (1) statically loaded, (2) dynamically loaded, and (3) tubular designs.

ASTM NO.	**MAXIMUM CARBON CONTENT, %**	**TENSILE TEST REQUIREMENTS(A)**	**DEOXIDATION**	**SUPPLEMENTARY REQUIREMENTS**
A 36	0.25-0.29 (FOR PLATE THICKER THAN 4 IN., OR 100 MM)	400-550 MPA (55-80 KSI); 250 MPA (36 KSI) MIN YP	NO RIMMED OR CAPPED STEEL FOR PLATE THICKER THAN 13 MM ($\frac{1}{2}$ IN.)	CHARPY V-NOTCH TESTING; SILICON KILLED, FINE-GRAIN PRACTICE
A 500	0.23-0.30	MIN TENSILE, THREE GRADES; 230, 290, 320 MPA (33, 42, 46 KSI) MIN YP	. . .	. . .
A 501	0.26-0.30	400 MPA (58 KSI) MIN TENSILE; 250 MPA (36 KSI) MIN YP	. . .	. . .
A 514	0.10-0.21 (15 PROPRIETARY COMPOSITIONS)	760-900 MPA (110-130 KSI) TO 65 MM (2.5 IN.) THICK; 690 MPA (100 KSI) MIN YP	FULLY KILLED, FINE-GRAIN PRACTICE	CLOSER THAN STANDARD FLATNESS
A 516	0.18-0.27	PLATES THICKER THAN 38 MM (1.5 IN.) NORMALIZED	KILLED, FINE-GRAIN PRACTICE	VACUUM TREATMENT; CHARPY V-NOTCH TEST; DROP-WEIGHT TEST(B)
A 572	0.21-0.26	MIN TENSILE, FOUR GRADES; 290-450 MPA (42-65 KSI) MIN YP	. . .	MAX TENSILE
A 588	0.15-0.20 (WEATHERING STEELS)	MIN TENSILE	FINE-GRAIN PRACTICE	MAX TENSILE; CHARPY V-NOTCH TEST; DROP-WEIGHT TEST(B); ULTRASONIC TEST

(A) YP, YIELD POINT.

(B) SEE ASTM E 208.

In general, the lower the carbon content and the finer the grain size of a steel, the better its notch toughness and weldability. The methods for achieving superior properties add to the unit price of the steel, so a designer striving for a low-cost structure tries to reduce steel cost, which is usually the most expensive single item in a job. This is especially true of static structures.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Environment

Low Temperatures. In the design of bridges, certain components merit special-toughness steel. Such components, termed fracture critical tension members (FCMs), would allow the bridge to collapse should they fail. The steel selection criteria are based on the minimum Charpy V-notch energy at a given temperature. For bridges financed in part by the Federal Highway Administration (FHA), the minimum winter temperature is the criterion used for the required Charpy V-notch energy at that temperature. A bridge located in the northern United States requires FCMs made with steel that exhibits the required minimum Charpy V-notch energy at a lower temperature than the same bridge constructed in the southern U.S.

The same FCM concept also applies for railroad bridges constructed under the codes established by AREA. The Charpy V-notch requirements for the three temperature zones defined by AREA codes are shown in Table 4. The required temperatures for measuring the Charpy V-notch energy are 21 °C (70 °F) for zone 1, 4 °C (40 °F) for zone 2, and -23 °C (-10 °F) for zone 3. Higher minimum Charpy V-notch values are required for weathering steels in heavier thicknesses. These values have been selected on the basis of extensive studies of steel structures, especially those of the Liberty ship failures during World War II.

TABLE 4 CHARPY IMPACT TEST REQUIREMENTS FOR WELDED STRUCTURAL STEEL: FRACTURE CRITICAL MEMBERS

ASTM	THICKNESS	MINIMUM AVERAGE ENERGY AND TEST TEMPERATURE		
		ZONE 1	ZONE 2	ZONE 3
A 36 OR A 709, GRADE 36	≤38 MM (≤1.5 IN.)	34 J (25 FT · LBF) AT 21 °C (70 °F)	34 J (25 FT · LBF) AT -12 °C (10 °F)	34 J (25 FT · LBF) AT 4 °C (40 °F)
	>38 TO ≤100 MM (>1.5 TO ≤4 IN.)	34 J (25 FT · LBF) AT -1 °C (30 °F)	34 J (25 FT · LBF) AT -7 °C (20 °F)	34 J (25 FT · LBF) AT -23 °C (-10 °F)
A 572, GRADE 50	≤38 MM (≤1.5 IN.)	34 J (25 FT · LBF) AT 21 °C (70 °F)	34 J (25 FT · LBF) AT -12 °C (10 °F)	34 J (25 FT · LBF) AT 4 °C (40 °F)
A 709, GRADE 50	>38 TO ≤50 MM (> 1.5 TO ≤2 IN.)	34 J (25 FT · LBF) AT -1 °C (30 °F)	34 J (25 FT · LBF) AT -7 °C (20 °F)	34 J (25 FT · LBF) AT -23 °C (-10 °F)
A 588 A 709, GRADE 50 W	>38 TO ≤100 MM (>1.5 TO ≤4 IN.)	41 J (30 FT · LBF) AT -1 °C (30 °F)	41 J (30 FT · LBF) AT -7 °C (20 °F)	41 J (30 FT · LBF) AT -23 °C (-10 °F)
MINIMUM SERVICE TEMPERATURE		-18 °C (0 °F)	-34 °C (-30 °F)	-51 °C (-60 °F)

Source: Adapted from "Manual for Railway Engineering," Table 1.14.7. American Railway Engineers Association

Earthquakes. Structures for locations subject to earthquakes should be designed to withstand the impact of such events. This is a constantly evolving technology. For up-to-date information, contact the Earthquake Engineering Research Center (EERC) at the University of California at Berkeley.

Atmospheric Corrosion. Steel structures that are to remain exposed to the atmosphere without painting or other surface protection can be constructed of a steel that forms an adherent oxide. Such steel is known as weathering steel and is covered by ASTM specification A 588. Corrosion resistance is provided by additions of approximately 0.35% Cu in most grades and/or small amounts of chromium or nickel in other grades. Weathering steels are suitable for resisting corrosion in air, but are not significantly better than other structural steels under water. In highway bridges where salt is used to control icing conditions, progressive corrosion is encountered, especially at expansion joints. A matching filler metal should be used to weld these steels, such as those with the designation "W" in the AWS specifications (e.g., E7018W in A5.5 or EW in A5.23).

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Fitness for Service

No structure is entirely free from flaws. Fitness for service is a term applied to an evaluation of the flaws found on inspection. The codes define the allowance for defects considered acceptable for the intended service.

Workmanship. Prior to assembly, the base metals are inspected for evidence of surface flaws that could impair the integrity of the structure. Before welding, the prepared edges are examined for discontinuities longer than 25 mm (1 in.). Such flaws must be removed and weld repaired before final assembly.

Distortion and shrinkage result from the welding operation. Permissible variations in the dimensions, straightness, and camber of structural members are defined in the codes. Peening of intermediate layers of multi-pass welds can be used to control distortion, but should be avoided on the weld surfaces.

Acceptable weld profiles are also defined as part of the workmanship standards of the codes for buildings and bridges. Unacceptable profiles, cracks, undercuts, overlaps, and other deficiencies are generally repaired by grinding the objectionable flaw and, if necessary, rewelding.

Inspection. Visual inspection is relied on for assessing workmanship criteria. In the construction of large structures, the general contract defines the extent of nondestructive testing of welds that may be required. Methods of nondestructive testing include radiography, ultrasonic testing, magnetic-particle inspection, and dye-penetrant inspection. For more information, see the article "Inspection of Welded Joints" in this Volume.

Radiography and ultrasonic testing allow examination of the full cross section of a weld. Flaws larger than 1% of the thickness, including cracks, lack of fusion, inclusions, and porosity can be detected by these techniques. X-ray analysis will detect only cracks substantially parallel to the impinging x-ray beam. Ultrasonic pulses can be introduced at an angle of 30 or 45° from the perpendicular to the plate surface, which permits a more complete inspection coverage of a weld.

Magnetic-particle and dye-penetrant methods are limited to the detection of defects on or near the surface. Dye penetrants can detect surface flaws only, whereas magnetic-particle testing can detect flaws slightly below the surface.

Flaw Assessment. The integrity and performance of welded structures depend heavily on the assessment of the significance of flaws identified during inspection. The discipline of fracture mechanics allows judgments to be made as to the likelihood of a flaw to cause a failure.

For static structures, the acceptance of a flaw of a given size that is found during inspection makes use of the material property evaluated by the crack-opening displacement (COD) test. Acceptance is based on the principle illustrated in Fig. 1. An example of the application of the COD test to flaw assessment is shown in Fig. 2.

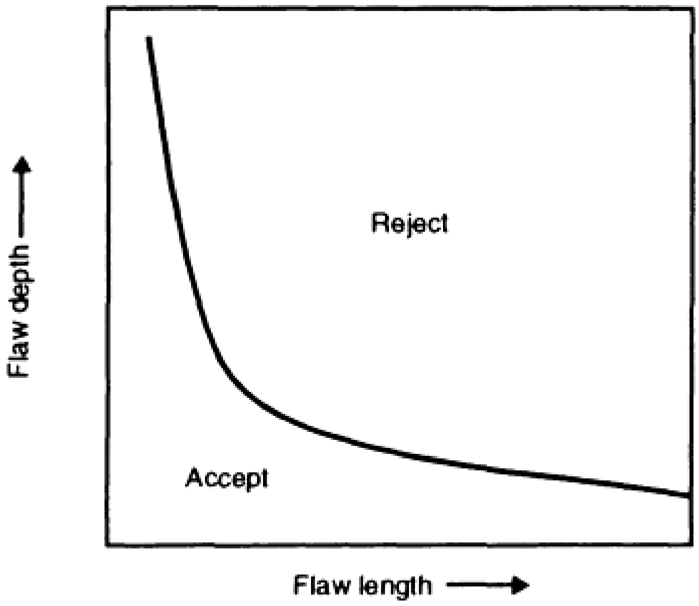

FIG. 1 SCHEMATIC OF AN ALLOWABLE FLAW-SIZE CURVE

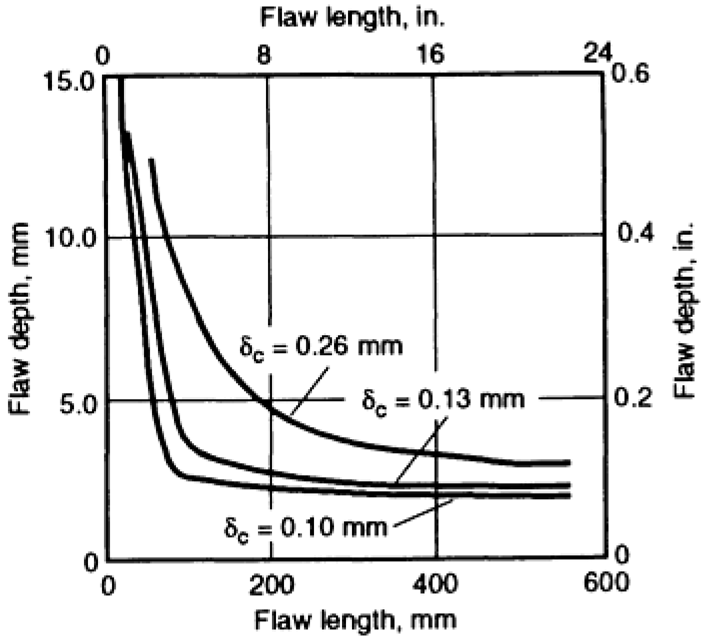

FIG. 2 ALLOWABLE FLAW SIZES DETERMINED FOR GIRTH WELDS IN PIPELINES CONSTRUCTED FROM API X70 STEEL, ASSUMING A COD VARYING FROM 0.1 TO 0.25 MM (0.004 TO 0.010 IN.), A FLOW STRESS OF 520 MPA (75 SKI), AND AN APPLIED STRESS OF 435 MPA (63 KSI). SOURCE: REF 1

Tolerance for Flaws. The inherent material properties of steels and welds that allow them to withstand the existence of flaws are measured by tensile testing (ductility), Charpy impact testing (toughness), and COD testing. For welds in FCMs, values of these measurements should be well above the minimum specification requirements. These tests are conducted on the welded joints prepared according to the specified qualification test procedures. An evaluation of ductility, toughness, and crack-opening displacement should be obtained on the weld metal, the HAZ, and the unaffected base metal.

Reference cited in this section

1. R.P. REED, M.B. KASEN, H.I. MCHENRY, C.M. FORTUNKO, AND D.T. REED, *FITNESS FOR SERVICE CRITERIA FOR PIPELINE GIRTH WELD QUALITY,* BULL. 296, WELDING RESEARCH COUNCIL, JULY 1984

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Pressure Vessels and Piping

Codes and Specifications

The primary users of pressure vessels and piping are the chemical, petroleum, and electric power industries. While there are numerous codes, standards, and specifications that govern the selection of materials for pressure vessels and piping, the manufacture of boilers and pressure vessels in the United States falls under the general specification issued by ASME titled "Boiler and Pressure Vessel Code." This code is published in several parts. The rules governing the design of boilers are included in section I, "Power Boilers." The design rules for pressure vessels are included in section VIII, divisions 1 and 2, "Rules for Construction of Pressure Vessels." The materials requirements are included in section II, "Materials Specifications":

- PART A: FERROUS
- PART B: NONFERROUS
- PART C: WELDING RODS, ELECTRODES, AND FILLER METALS
- PART D: PROPERTIES

The ferrous (SA) and nonferrous (SB) ASME materials specifications were originally developed by ASTM and in most cases are identical. Thus, ASME specifications designated SA 204 and SB 467 may also be found as ASTM specifications A 204 and B 467, respectively.

In the United States, the AWS filler metal specifications are the primary guidelines used by the pressure vessel and piping industry. Here again, the ASME specifications (part C) are identical. The ASME specifications carry the initial designation SFA to correspond to the AWS specification designations starting with A; for example, SFA 5.5 is identical to AWS A 5.5. The AWS specifications are written to provide the specific chemical analysis of the filler material and the mechanical properties (minimum tensile and yield strengths) of the deposited weld metal.

Standards for pressure piping are covered by the American National Standards Institute (ANSI) document B31.1, "Code for Pressure Piping." This standard also applies to industrial pipelines and piping. Material selection is primarily from the ASTM and ASME materials specifications, although the API does provide a number of pipe specifications for material with yield strengths ranging from 170 to 760 MPa (25 to 110 ksi) and tensile strengths ranging from 275 to 860 MPa (40 to 125 ksi). The intended service temperatures, corrosion resistance, and fabricability are the primary considerations in the selection of materials for pressure vessels and piping.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Environment

Ordinary-Temperature Service (-30 to 345 °C, or -20 to 650 °F). The ASME, recognizing the fact that the ultimate strength of steels remains relatively constant over the temperature range from -30 to 345 °C (-20 to 650 °F), elected to set a safe maximum design stress of 25% of the ultimate tensile strength at temperatures over this temperature range. Consequently, the plain carbon steels are the most economical, with the popular grades being SA 515 ("Pressure

Vessel Plates, Carbon Steel, for Intermediate- and Higher-Temperature Service") and SA 516 ("Pressure Vessel Plates, Carbon Steel, for Moderate- and Lower-Temperature Service"). SA 516, a fine-grain, silicon-aluminum killed steel, is often preferred to SA 515, a coarse-grain, silicon killed steel, because of its better notch toughness characteristics.

A steam drum for an electric utility boiler fabricated from SA 515 grade 70 material is shown in Fig. 3. Where reduction in weight is important and operating pressures are high, the higher strength weldable steels are used, such as SA 302 ("Pressure Vessel Plates, Alloy Steel Manganese-Molybdenum and Manganese-Molybdenum-Nickel"). The quenched and tempered version of this steel, SA 533 (grade B, class 1), is used extensively in nuclear reactor vessels. Pertinent data for these materials are listed in Table 5. The specifications listed in Table 5 (and also Tables 6 and 7, to be discussed later) are for plate material. Similar ASME specifications exist for forgings, pipe, tube, and cast material.

TABLE 5 PRESSURE VESSEL STEELS FOR ORDINARY-TEMPERATURE SERVICE

ASME NO.	**ALLOY AND NOMINAL COMPOSITION**	**MINIMUM SPECIFIED TENSILE STRENGTH**		**ALLOWABLE DESIGN STRESS**	
		mpa	**ksi**	**mpa**	**ksi**
SA 515	PLAIN-CARBON; COARSE-GRAIN PRACTICE				
SA 516	PLAIN-CARBON; FINE-GRAIN PRACTICE; IMPROVED TOUGHNESS				
GRADE 55	0.26% C MAX; 0.60-0.90% MN	380	55	95	13.75
GRADE 60	0.27% C MAX; 0.60-0.90% MN	415	60	103	15.00
GRADE 65	0.29% C MAX; 0.85-1.20% MN	450	65	112	16.25
GRADE 70	0.31% C MAX; 0.85-1.20% MN	485	70	121	17.50
SA 302					
SA 302 GRADE A	MN-MO	515	75	130	18.8
GRADE B	MN-MO	550	80	138	20
GRADE C	MN-MO-NI	550	80	138	20
GRADE D	MN-MO-NI	550	80	138	20
SA 202					
GRADE A	CR-MN-SI	515	75	130	18.8
GRADE B	CR-MN-SI	585	85	147	21.3
SA 537					
CLASS 1	C-MN-SI; NORMALIZED; 63.5 MM (2.5 IN.) MAX THICKNESS	485	70	121	17.5
CLASS 2	C-MN-SI: QUENCHED AND TEMPERED; 63.5 MM (2.5 IN.) MAX THICKNESS	550	80	138	20

TABLE 6 PRESSURE VESSEL STEELS FOR LOW-TEMPERATURE SERVICE

ASME NO.	**ALLOY AND NOMINAL COMPOSITION**	**MINIMUM SPECIFIED TENSILE STRENGTH**		**ALLOWABLE DESIGN STRESS**		**LOWEST TEST TEMPERATURE FOR CHARPY V-NOTCH TEST**	
		MPa	**ksi**	**MPa**	**ksi**	**°C**	**°F**
SA 516	PLAIN-CARBON; FINE-GRAIN						

	PRACTICE						
GRADE 60	0.27% C MAX; 0.60-0.90% MN	415	60	103	15.0	-46	-50
GRADE 65	0.29% C MAX; 0.85-1.20% MN	450	65	112	16.25	-46	-50
GRADE 70	0.31% C MAX; 0.85-1.20% MN	485	70	121	17.5	-46	-50
SA 537	FINE-GRAIN PRACTICE						
CLASS 1	NORMALIZED; 0.24% C MAX; 0.70-1.35% MN	485	70	. . .	. . .	-62	-80
CLASS 2	QUENCHED AND TEMPERED; 0.24% C MAX; 0.70-1.35 MN	550	80	. . .	. . .	-60	-76
SA 203							
GRADE A	2.4% NI	450	65	112	16.25	-60	-75
GRADE B	2.4% NI	485	70	121	17.5	-60	-75
GRADE D	3.5% NI	450	65	112	16.25	-100	-150
GRADE E	3.5% NI	485	70	121	17.5	-100	-150
SA 645	5.0% NI	655	95	164	23.75	-170	-275
SA 553							
TYPE II	8.0% NI	690	100	172	25.0	-170	-175
TYPE I	9.0% NI	690	100	172	25.0	-195	-320

TABLE 7 PRESSURE VESSEL STEELS FOR HIGH-TEMPERATURE SERVICE

ASME NO.	ALLOY AND NOMINAL COMPOSITION	NORMAL TEMPERATURE RANGE OF USAGE		ALLOWABLE STRESS AT LOW END OF USAGE TEMPERATURE		ALLOWABLE STRESS AT HIGH END OF USAGE TEMPERATURE	
		°C	**°F**	**Mpa**	**Ksi**	**Mpa**	**Ksi**
SA 204, GRADE C	0.28% C MAX; 0.45-0.60% MO	430-510	800-950	130	18.8	57	8.2
SA 302, GRADE B	0.25% C MAX; 1.15-1.50% MN; 0.45-0.60% MO	430-510	800-950	130	18.8	57	8.2
SA 387, GRADE 12							
CLASS 1	1.0CR-0.5MO	455-565	850-1050	92	13.4	30	43
CLASS 2	1.0CR-0.5MO	345-480	650-900	112	16.3	104	15.1
SA 387, GRADE 112							
CLASS 1	1.25CR-0.5MO	455-565	850-1050	101	14.6	32	4.6
CLASS 2	1.25CR-0.5MO	345-480	650-900	130	18.8	110	15.9
SA 387, GRADE 22							
CLASS 1	2.25CR-1.0MO	455-	850-	99	14.4	40	5.8

		595	1100				
CLASS 2	2.25CR-1.0MO	370-480	700-900	119	17.2	109	15.8
SA 387, GRADE 5							
CLASS 1	5.0CR-0.5MO	480-620	900-1150	83	12.1	29	4.2
CLASS 2	5.0CR-0.5MO						
SA 387, GRADE 9	9.0CR-1.0MO	510-595	950-1100	73	10.6	22	3.3
SA 387, GRADE 91	9.0CR-1.0MO + NI, V, NB, N, AL	540-650	1000-1200	99	14.3	28	4.3
SA 240	AUSTENITIC STAINLESS STEELS						
GRADE 304H	18CR-8NI	595-815	1100-1500	61	8.9	9.7	1.4
GRADE 316H	16CR-12NI-2MO	595-815	1100-1500	71	10.3	8.3	1.2
GRADE 321H	18CR-10NI-TI	595-815	1100-1500	48	6.9	2	0.3
GRADE 347H	18CR-10NI-NB	595-815	1100-1500	90	13.0	9	1.3

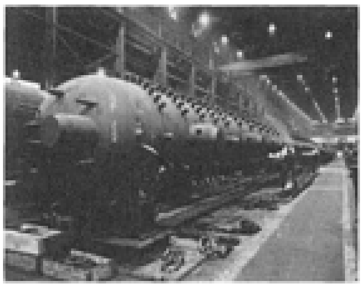

FIG. 3 STEAM DRUM FOR AN SA 515 STEEL ELECTRIC UTILITY BOILER

Low-Temperature Service (-30 to -195 °C, or -20 to -320 °F). The commonly measured mechanical properties of steels, such as ultimate strength, yield strength, and elongation, do not always ensure freedom from failure at low temperatures, even when the design load is not exceeded. To ensure safe performance, the steel must be resistant to the initiation and propagation of a crack under all service conditions. Resistance to brittle crack initiation and propagation is commonly referred to as notch toughness or notch ductility. Factors that adversely affect notch toughness include cold working, strain aging, large grain size, increasing section thickness, and increasing the compositional contents of carbon, sulfur, phosphorus, nitrogen, and silicon. Beneficial factors include fine grain size, deoxidation, heat treatment, increasing the manganese and nickel contents, and decreasing the sulfur content. The notch toughness of a steel is also dependent on fabrication procedures, workmanship, and design details. Even the best design will contain some notch effects, and there will also be some imperfections in the highest level of workmanship. In the presence of these notches, the breaking strength of a pressure vessel and, more importantly, its failure mode (ductile or brittle) depend largely on the capacity of the steel at the root of some critical notch to yield plastically as the stress increases.

In thick sections, plain carbon steels produced according to fine-grain practice and normalized or quenched and tempered are used for service to -45 °C (-50 °F). Low-carbon high-nickel (e.g., 9% Ni) alloy steels are used for service down to -195 °C (-320 °F). Ferritic steels covered by ASME specifications are shown in Table 6. Austenitic chromium-nickel steels, aluminum, and special copper and aluminum-base alloys have been found to be particularly suitable for applications close to absolute zero. Because austenitic steels have a face-centered cubic (fcc) crystal structure, they retain their toughness to very low temperatures.

High-Temperature Service (345 to 815 °C, or 650 to 1500 °F). In the design of pressure vessels and piping, engineers and designers are confronted with the problem of selecting materials for a wide range of high-temperature service conditions. The chromium-molybdenum ferritic steels and austenitic stainless steels are generally used for design temperatures above 425 °C (800 °F). In addition to service temperature, corrosion resistance, and fabricability, the following conditions should be considered in high-temperature applications:

- POSSIBLE MAXIMUM TEMPERATURE
- TYPE AND SIZE OF LOAD
- EXPECTED LIFE OF THE STRUCTURE
- COST

Service experience and laboratory test data have established the normal temperature range of usage shown in Table 7 for the materials commonly used for high-temperature service. The allowable stresses in the ASME code, section VIII, division 1, for the low and high ends of the usage temperature ranges are also given in Table 7.

The following material properties must be taken into account in addition to the aforementioned service conditions:

- CREEP STRENGTH
- STRESS-RUPTURE LIFE
- DUCTILITY
- SHORT-TERM TENSILE PROPERTIES
- SURFACE STABILITY
- STRUCTURAL STABILITY
- THERMAL CONDUCTIVITY
- THERMAL FATIGUE
- THERMAL EXPANSION
- WELDABILITY
- HOT AND COLD WORKABILITY

All of these factors are important in arriving at the most suitable alloy for a given application.

Corrosion. Because of the many chemical compounds that may be encountered in pressure vessel and piping service, a discussion of corrosion is outside the scope of this section. However, there are two areas of concern in both the petrochemical field and the electric power industry: the effect of hydrogen absorption on materials properties and the possible metal loss due to steam/air oxidation.

High-Pressure, High-Temperature Hydrogen Service. Pressure vessels and piping exposed to hot high-pressure hydrogen may embrittle, crack, and ultimately fail without visible thinning of the metal or a superficially observable change in appearance of the exposed surfaces. A tool known as the Nelson diagram (Ref 3) is generally used to predict the behavior of alloy material in high-pressure, high-temperature hydrogen (Fig. 4). The Nelson diagram is based on actual service experience and indicates the temperature-pressure limits below which a material has not failed to date. Allowance must be made for temperature and pressure excursions that may occur over the design life of the component. Hydrogen can affect the mechanical properties of steel in two ways. Absorbed hydrogen causes a reduction in yield strength and impairs material toughness. More advanced attack, as manifested by fissuring and decarburization, is reflected in severely reduced tensile strength, ductility, and toughness.

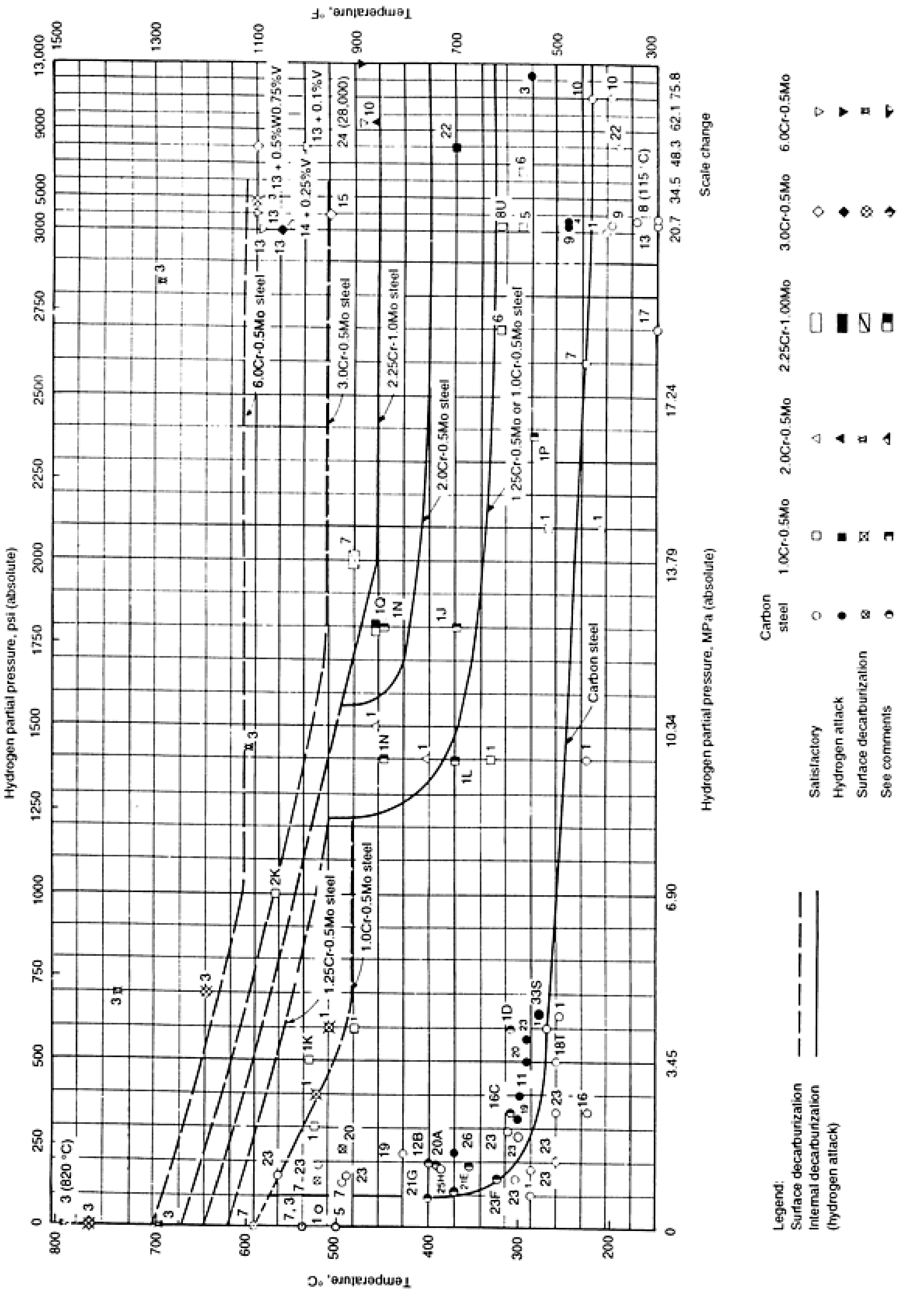

FIG. 4 NELSON DIAGRAM. 1, THE LIMITS DESCRIBED BY THESE CURVES ARE BASED ON ACTUAL SERVICE EXPERIENCE AND ON INFORMATION GATHERED BY API (REF 3). 2, AUTENITIC STAINLESS STEELS ARE GENERALLY NOT DECARBURIZED IN HYDROGEN AT ANY TEMPERATURE OR HYDROGEN PRESSURE. 3, THE LIMITS DESCRIBED BY THESE CURVES ARE BASED ON EXPERIENCE WITH CAST STEEL AS WELL AS ANNEALED AND NORMALIZED STEELS AT STRESS LEVELS DEFINED BY THE ASME CODE, SECTION VIII, DIVISION I. LINES REPRESENT THE UPPER LIMIT OF CONDITIONS FOR ACCEPTABLE USE OF THE ALLOY STEEL. SOURCE:

REF 3

The recent trend in low-alloy steels for hydrogen service has been in the direction of steels of the chromium-molybdenum-vanadium type, such as 3Cr-1Mo-0.25V or 2.25Cr-1Mo-0.25V. For the most severe corrosive conditions, particularly those involving hydrogen sulfide (H_2S) in addition to hydrogen, austenitic steels of the 18Cr-8Ni type are used.

The National Association of Corrosion Engineers (NACE) standard MR-01-75 is the accepted reference in terms of materials and fabrication requirements for H_2S service. This standard specifies the materials and fabrication that are considered to provide satisfactory resistance to H_2S attack. Carbon steels for use in wet H_2S service are usually evaluated in accordance with NACE TM-01-77 (test solution) and TM-02084 (test method).

The relatively high cost of the austenitic steels frequently dictates the use of carbon or low-alloy steel vessels with applied corrosion-resistant cladding. In general, three methods of attaching the cladding to the backing plate are utilized: integral cladding, strip lining, and weld overlay cladding. Integrally clad plate is fabricated by hot rolling in a steel mill or by explosive cladding. Strip lining is applied by using long, narrow alloy sheet attached to the vessel by a continuous weld around the edges. Weld overlay cladding produces a continuous bonded layer of corrosion-resistant alloy on the base metal. Weld overlay cladding can be applied on irregular shapes, such as nozzles, thus avoiding dissimilar-alloy welds in the main pressure boundary.

Oxidation in Steam Service. In choosing a steel for service at elevated temperature, consideration must be given not only to its resistance to applied stress at temperature but also to its resistance to scaling, both of which decrease rapidly with increasing temperature above about 600 °C (1110 °F). Chromium, alone or in combination with silicon or molybdenum, contributes to increased resistance. For plain carbon steels, Fe_3O_4 is the only oxidation product and is stable until a temperature of 570 °C (1060 °F) is reached. Above 570 °C (1060 °F), FeO is the stable oxide phase and contributes to more rapid metal loss. The addition of chromium raises the temperature at which the Fe_3O_4/FeO transition occurs (see Table 8). Metal loss due to oxidation is of relatively little importance in heavy-wall pressure vessels; however, it deserves consideration in thinner-wall piping and in thinner low-pressure vessels.

TABLE 8 FE_3O_4/FEO TRANSITION TEMPERATURES

ALLOY	TEMPERATURE	
	°C	°F
0.5CR-0.5MO	590	1095
1CR-0.5MO	600	1110
1.25CR-0.5MO	605	1120
2.25CR-1MO	615	1140
5CR-0.5MO	625	1155

Other Factors Causing Degradation of Materials (Ref 4). Specific service conditions may degrade materials to an extent that requires replacement or retirement of components in the system. One of the more common failure modes in pressure vessels is SCC, which can be characterized as either intergranular (IGSCC) or transgranular (TGSCC), depending on the environment. Hydrogen-induced cracking (HIC) is another failure mechanism, again influenced by the localized stress condition.

Low-alloy steels in service at temperatures above 450 °C (850 °F) will suffer a loss in ductility due to temper embrittlement unless the composition is rigidly controlled. Phosphorus, tin, arsenic, and antimony are the principal elements that contribute to temper embrittlement.

The ductility of steels operating in the creep temperature range (generally above 480 °C, or 900 °F) suffers from a condition known as creep embrittlement. The loss in ductility is caused by the formation of small cavities in grain boundaries. In the presence of relatively low stress, the number of cavities grow and eventually link together to form microcracks, thus preventing the plastic deformation that is characteristic of ductile behavior. Sulfur is one of the elements that contribute to cavity formation and loss in ductility in creep service.

References cited in this section

3. *STEELS FOR HYDROGEN SERVICE AT ELEVATED TEMPERATURES AND PRESSURES IN PETROLEUM REFINERIES AND PETROCHEMICAL PLANTS,* 4TH ED., API PUBL. 941, AMERICAN PETROLEUM INSTITUTE, APRIL 1990
4. S. YUKAWA, *GUIDELINE FOR PRESSURE VESSEL SAFETY ASSESSMENT,* NIST SPEC. PUBL. 780, NATIONAL INSTITUTE FOR STANDARDS AND TECHNOLOGY, P 14-22

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Fabricability

Fabrication operations can affect the performance of pressure vessels and piping in several ways: (1) cold forming operations may lead to strain aging, with a loss in toughness and ductility; (2) hot forming, unless carefully controlled, may produce undesirable changes in microstructure; (3) welding operations may produce an HAZ with reduced toughness and high hardness, and may also produce high tensile residual stresses that could lead to stress-rupture cracking (reheat cracking) during postweld heat treatment of susceptible materials.

Welding Process Selection. Considerations which govern the selection of a welding process are those that define the end use of a vessel, such as toughness, tensile properties, creep properties, and corrosion. There are also practical considerations, such as weldment size, equipment limitations, welding position, and access to the weldment area.

Two welding processes are primarily used: shielded metal arc welding (SMAW) and submerged arc welding (SAW). Other processes, such as electroslag welding, gas-tungsten arc welding, and gas-metal arc welding, are also used, but historically SMAW (see the article "Shielded Metal Arc Welding" in this Volume) and SAW (see the article "Submerged Arc Welding" in this Volume) have been dominant in vessel manufacture. A characteristic of both the SMAW and SAW processes is that they rely on the use of a flux to produce a slag over the molten weld pool, protecting it from atmospheric contamination. These fluxes vary significantly in formulation and operating characteristics and can modify the chemical composition, inclusion content, and microstructure of the final weld metal.

Weldability. Steels for pressure vessel fabrications are often classified as weldable based on composition, thickness, and need for preheat (see the article "Weldability Testing" in this Volume). Carbon and low-alloy steels are the most frequently used materials for pressure vessels; consequently, their weldability has received the greatest attention.

Preheating is utilized to prevent cracking, to permit diffusion of hydrogen from the weld and HAZ during and following welding, to reduce distortion, and to prevent loss of toughness and ductility. Preheating is also used to ensure that all traces of moisture have been removed from the surface of the material. Of the various schemes proposed for predicting preheating requirements, most are based on the carbon equivalent (C_{eq}). A variety of equations have been developed by various researchers for determining the C_{eq}, one of which has demonstrated that prevention of cracking due to hydrogen can be correlated with chemical composition in terms of a carbon equivalent formula of the form:

$$C_{eq} = C + \frac{Mn}{6} + \frac{Si}{24} + \frac{Ni}{40} + \frac{Cr}{5} + \frac{Mo}{4}$$

This formula was developed by Winn (Ref 5). A plot of carbon equivalents for a series of steel compositions used by Winn against preheat temperature is shown in Fig. 5. A linear relationship exists between the steel chemical composition, expressed by the carbon equivalent formula, and the minimum preheat temperature to avoid HAZ cracking.

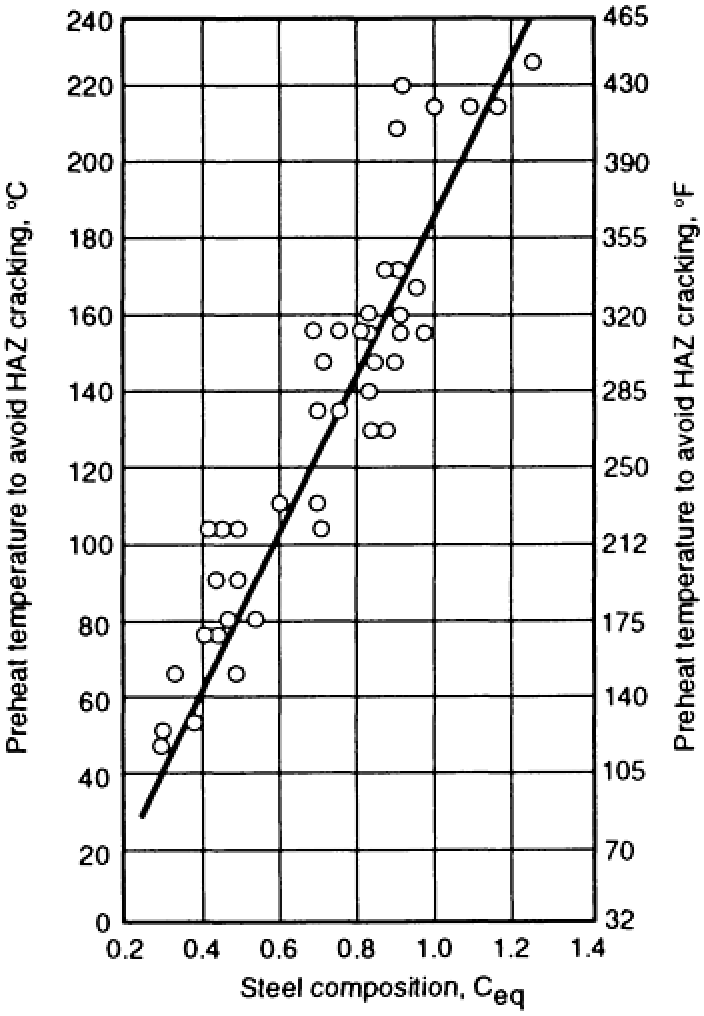

FIG. 5 CARBON EQUIVALENT VERSUS MINIMUM PREHEAT TEMPERATURE. THE BEST-FIT LINE SHOWN MAY BE REPRESENTED APPROXIMATELY BY: $T = 210(C_{EQ}) - 25$, WHERE T IS THE PREHEAT TEMPERATURE (°C).

The formula used by the International Institute of Welding (IIW) for judging the risk of underbead cracking and the resultant need for preheat is as follows:

$$C_{eq} = C + \frac{Mn}{6} + \frac{Cr + Mo + V}{5} + \frac{Ni + Cu}{15}$$

Regardless of the formula used, the C_{eq} values must be considered as approximations, because all factors affecting hardenability (e.g., section thickness, mass, restraint, grain size, and environmental effects) are not included.

Weld Zones. Based on work published by Savage, Nippes, and Szekeres (Ref 6), a typical weld consists of the composite zone, the unmixed zone, the weld interface, the partially melted zone, the HAZ, and the unaffected base metal (Fig. 6). All of these zones possess different mechanical properties, which are determined by the chemical compositions of the weld metal and the base metal. The preheat used, the weld energy input, and whether or not postweld heat treatment is specified also affect the resultant mechanical properties of the weldment.

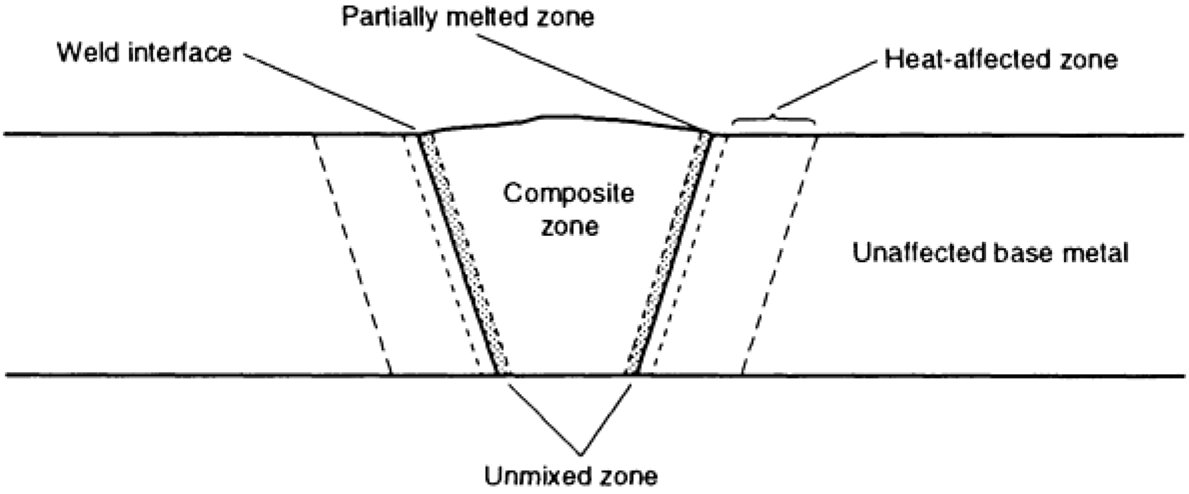

FIG. 6 TYPICAL ZONES ASSOCIATED WITH FUSION WELDS. SOURCE: REF 6

Postweld Heat Treatment. The circumstances under which the postweld heat treatment of pressure vessels is required are set forth in the ASME pressure vessel codes. They are determined primarily by the steel composition, thickness, specific fabrication operations, and intended service.

The principal beneficial effects of postweld heat treatment are the relief of residual stresses and the tempering action on the microstructures produced by the welding thermal cycles. Undesirable metallurgical phenomena may also be produced, such as temper embrittlement or cracking due to stress-rupture failures at grain boundaries.

References cited in this section

5. W.H. WINN, WELDABILITY OF LOW ALLOY STEELS, *BR. WELD. J.,* VOL 11 (NO. 8), AUG 1964
6. W.F. SAVAGE, E.F. NIPPES, AND E.S. SZEKERES, A STUDY OF WELD INTERFACE PHENOMENA IN A LOW ALLOY STEEL, *WELD. J.,* SEPT 1976, P 260S-268S

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions[*]

Mechanical Properties of Weldments

Design Considerations. Some designers have tacitly assumed that a pressure vessel behaves as if it were manufactured from an isotropic material with the same properties throughout the vessel. Because of the presence of welded seams and welded attachments, however, no actual vessel fulfills this simple concept. Welds produce regions where the mechanical properties are considerably different from those of the base material. The stress-rupture life and the ductility at rupture may also be significantly different.

Weld Metal Strength. There is a tendency for weld metal to overmatch the parent metal in strength at ambient and low temperatures. While a slight overmatch is not harmful, excessive hardness in welds can cause greater susceptibility to weld cracking.

Creep of Welds in High-Temperature Service. Recent developments suggest that welded joints in some chromium-molybdenum steels are a potential source of weakness in creep. Vessels that have performed satisfactorily in service thus far may have safety factors lower than those suggested by present design rules. This arises from the mismatch of creep properties in the different weld zones, especially in thick sections. In recognition of this fact, section III of the ASME code, "Nuclear Power Plant Components," has assigned stress-reduction factors to weldments operating in the creep range of temperatures.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Shipbuilding and Offshore Structures

The materials used for ships and offshore structures encounter a broad range of service conditions and must be capable of withstanding the rigors of weather and sea. With the advent of sophisticated communications systems, ships can often avoid heavy weather, whereas site-dependent offshore structures must be built to withstand any service condition that arises.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Ships

Steel ship structures are comprised of shell plating and shapes joined by welding. The structures vary considerably among various ship types, including small boats 10 to 25 m (30 to 80 ft) long to supertankers with lengths ranging from 100 to 400 m (330 to 1300 ft) or more. In between are container ships, liquefied natural gas (LNG) and liquefied petroleum gas (LPG) carriers, passenger ships, ore carriers, and others. Requirements for ordinary- and higher-strength steels according to the rules of the American Bureau of Shipping Rules (ABS) are given in Tables 9 and 10, respectively.

TABLE 9 ABS REQUIREMENTS FOR ORDINARY-STRENGTH HULL STRUCTURAL STEEL (≤50 MM, OR 2 IN., THICK)

GRADE	DEOXIDATION	CHEMICAL COMPOSITION(A), %				
		C	Mn	P	S	Si
A	SEMIKILLED OR KILLED FOR MATERIAL THICKER THAN 12.5 MM (0.5 IN.)	0.23 MAX(B)	2.5 × C MIN FOR PLATE THICKER THAN 12.5 MM (0.5 IN.)	0.04 MAX	0.04 MAX	. . .
B	SEMIKILLED OR KILLED	0.21 MAX	0.08-1.10; 0.60 MIN FOR FULLY KILLED OR COLD FLANGING	0.04 MAX	0.04 MAX	0.35 MAX
D	FULLY KILLED; FINE-GRAIN PRACTICE(C)	0.21 MAX	0.70-1.35; 0.60 MIN FOR THICKNESS ≤25 MM (1 IN.)	0.04 MAX	0.04 MAX	0.10-0.35
E	FULLY KILLED; FINE-GRAIN PRACTICE	0.18 MAX	0.70-1.35	0.04 MAX	0.04 MAX	0.10-0.35

TENSILE STRENGTH		YIELD POINT (MIN)		ELONGATION (MIN)
Mpa	Ksi	Mpa	Ksi	
FOR ALL GRADES:		FOR ALL GRADES:		FOR ALL GRADES:
400-490	58-71	235	34	21% IN 200 MM (8 IN.) OR 24% IN 50 MM (2 IN.)
FOR GRADE A SHAPES		FOR GRADE A THICKER THAN		FOR COLD FLANGING QUALITY:

AND BARS:		25 MM (1 IN.)		
400-550	58-80	220	32	23% MIN IN 200 MM (8 IN.)
FOR COLD FLANGING QUALITY:		FOR COLD FLANGING:		
380-450	55-65	205	30	

GRADE	**CHARPY V-NOTCH IMPACT TEMPERATURE**		**ENERGY AVERAGE (MIN), J (FT · LBF)**		**HEAT TREATMENT**	**MARKING**
	°C	**°F**	**LONGITUDINAL SPECIMENS**	**TRANSVERSE SPECIMENS**		
A	. . .	. . .	. . .	. . .	. . .	AB/A
B	0	32	27 (20)	20 (14)	. . .	AB/B
D	-10	14	27 (20)(D)	20 (14)(D)	NORMALIZED FOR MATERIAL THICKER THAN 35 MM (1.4 IN.)(E)	AB/D(F)
E	-40	-40	27 (20)	20 (14)	NORMALIZED(G)	AB/E

(A) FOR ALL GRADES EXCLUSIVE OF GRADE A SHAPES AND BARS, THE SUM OF CARBON CONTENT AND ONE-SIXTH OF THE MANGANESE CONTENT IS NOT TO EXCEED 0.40%. THE UPPER LIMIT OF MANGANESE MAY BE EXCEEDED UP TO A MAXIMUM OF 1.65%, PROVIDED THAT THIS CONDITION IS SATISFIED. THE CONTENTS OF NICKEL, CHROMIUM, MOLYBDENUM, AND COPPER ARE TO BE DETERMINED AND REPORTED, AND WHEN THE AMOUNT PRESENT IS LESS THAN 0.02%, THESE ELEMENTS MAY BE REPORTED AS 0.02%.

(B) A MAXIMUM CARBON CONTENT OF 0.26% IS ACCEPTABLE FOR GRADE A PLATES ≤12.5 MM (0.5 IN.) AND ALL THICKNESSES OF GRADE A SHAPES AND BARS.

(C) GRADE D MAY BE FURNISHED SEMIKILLED IN THICKNESSES UP TO 35 MM (1.4 IN.), PROVIDED THAT STEEL >25 MM (1 IN.) THICK IS NORMALIZED. IN THIS CASE THE REQUIREMENTS RELATIVE TO MINIMUM SILICON AND ALUMINUM CONTENTS AND FINE-GRAIN PRACTICE DO NOT APPLY.

(D) IMPACT TESTS ARE NOT REQUIRED FOR NORMALIZED GRADE D STEEL WHEN FURNISHED FULLY KILLED, FINE-GRAIN PRACTICE.

(E) CONTROLLED ROLLING OR THERMOMECHANICAL CONTROLLED ROLLING OF GRADE D STEEL MAY BE CONSIDERED AS A SUBSTITUTE FOR NORMALIZING.

(F) GRADE D HULL STEEL THAT IS NORMALIZED, THERMOMECHANICAL CONTROLLED ROLLED, OR CONTROLLED ROLLED IN ACCORDANCE WITH (E) IS TO BE MARKED AB/DN.

(G) CONTROLLED ROLLING OR THERMOMECHANICAL CONTROLLED ROLLING OF GRADE E SHAPES AND THERMOMECHANICAL CONTROLLED ROLLING OF GRADE E PLATES MAY BE CONSIDERED AS A SUBSTITUTE FOR NORMALIZING.

TABLE 10 ABS REQUIREMENTS FOR HIGHER-STRENGTH HULL STRUCTURAL STEEL (≤50 MM, OR 2 IN., THICK)

GRADE(A)	**DEOXIDATION**	**TENSILE STRENGTH**		**YIELD POINT (MIN)**		**ELONGATION (MIN)**	**CHARPY V-NOTCH IMPACT TEMPERATURE**		**ENERGY AVERAGE (MIN), J (FT · LBF)**		**HEAT TREATMENT**	**MARKING**
		Mpa	**Ksi**	**Mpa**	**Ksi**		**°C**	**°F**	**LONGITUDINAL SPECIMENS**	**TRANSVERSE SPECIMENS**		
AH32	SEMIKILLED	470-	68-	315	45.5		0	32	34 (25)(B)	24 (17)(B)	. . .	AB/AH32

	OR KILLED	585	85									
AH36		490-620	71-90	355	51							AB/AH36
DH32	KILLED; FINE-GRAIN PRACTICE(C)	470-585	68-85	315	45.5	FOR ALL GRADES: 19% IN 200 MM (8 IN.) OR 22% IN 50 MM (2 IN.)	-20	-4	34 25(B)	24 (17)(B)	NORMALIZED(D) FOR MATERIAL >12.5 MM (0.5 IN.) THICK(E)	AB/DH32(F)
DH36		490-620	71-90	355	51							AB/DH36(F)
EH32	KILLED; FINE-GRAIN PRACTICE (C)	470-585	68-85	315	45.5		-40	-40	34 (25)	24 (17)	NORMALIZED(G)	AB/EH32
EH36		490-620	71-90	355	51							AB/EH36

(A) THE NUMBERS FOLLOWING THE GRADE DESIGNATION INDICATE THE YIELD POINT OR YIELD STRENGTH TO WHICH THE STEEL IS ORDERED AND PRODUCED IN KGF/MM^2. THE CHEMICAL COMPOSITION FOR ALL GRADES IS: 0.18% MAX C; 0.90-1.60% MN (GRADE AH THAT IS ≤12.5 MM, OR 0.5 IN., THICK MAY HAVE A MINIMUM MANGANESE CONTENT OF 0.70%); 0.04% MAX P; 0.04.% MAX S; 0.10-0.50% SI (GRADE AH PLATE THAT IS ≤12.5 MM, OR 0.5 IN., THICK AND ALL THICKNESSES OF AH SHAPES MAY BE SEMIKILLED, IN WHICH CASE THE 0.10% MIN SI DOES NOT APPLY; UNLESS OTHERWISE SPECIALLY APPROVED, AH PLATE >12.5 MM, OR 0.5 IN., THICK ARE TO BE KILLED WITH 0.10-0.50% SI); 0.40% MAX NI; 0.25% MAX CR; 0.08% MAX MO: 0.35% MAX CU; 0.05% MAX NB; 0.10 MAX V. NICKEL, CHROMIUM, AND MOLYBDENUM MAY BE REPORTED AS 0.02% WHEN THE AMOUNT PRESENT IS LESS THAN 0.02%. NIOBIUM AND VANADIUM NEED NOT BE REPORTED ON THE MILL SHEET UNLESS INTENTIONALLY ADDED.

(B) IMPACT TESTS ARE NOT REQUIRED FOR AH THAT IS ≤12.5 MM (0.5 IN.) THICK OR FOR ALUMINUM TREATED GRADE AH THAT IS ≤35 MM (1.4 IN.) THICK, NOR FOR FULLY KILLED FINE-GRAIN NORMALIZED GRADES AH OR DH THAT IS ≤50 MM (2 IN.) THICK.

(C) GRADES DH AND EH ARE TO CONTAIN AT LEAST ONE OF THE GRAIN-REFINING ELEMENTS IN SUFFICIENT AMOUNT TO MEET THE FINE-GRAIN PRACTICE REQUIREMENT.

(D) CONTROLLED ROLLING OR THERMOMECHANICAL CONTROLLED ROLLING OF GRADE D STEEL MAY BE CONSIDERED AS A SUBSTITUTE FOR NORMALIZING.

(E) CONTROLLED ROLLING OR THERMOMECHANICAL CONTROLLED ROLLING OF GRADE EH SHAPES AND THERMOMECHANICAL CONTROLLED ROLLING OF GRADE EH PLATES MAY BE CONSIDERED AS A SUBSTITUTE FOR NORMALIZING.

(F) THE MARKING AB/DHN IS TO BE USED TO DENOTE GRADE DH PLATES WHICH HAVE EITHER BEEN NORMALIZED, THERMOMECHANICAL CONTROLLED ROLLED, OR CONTROL ROLLED IN ACCORDANCE WITH APPROVED PROCEDURE.

(G) FOR DH TREATED WITH NIOBIUM OR VANADIUM, OR WHEN USED IN COMBINATION WITH ALUMINUM OR EACH OTHER. DH TREATED ONLY WITH ALUMINUM IS TO BE

NORMALIZED WHEN >25 MM (1 IN.) THICK.

Tankers and Container Ships. Large ocean-going tankers have been built with both single and double hulls, with the latter type becoming more popular due to environmental considerations and the belief that oil spills are more readily prevented by double-hull construction. However, most tankers in service today have single hulls. In the design of typical ships, 0 °C (32 °F) is normally assumed to be the lowest service temperature.

Material selection for single-hull tankers for a particular steel strength level is based on stress level, location, thickness, and toughness. A typical cross section of the shell and longitudinal bulkheads for a large tanker is shown in Fig. 7. Plating for various parts of the ship can be categorized as secondary, intermediary, and special application. The four corners of a ship and the deck and bottom plating in the way of the longitudinal bulkheads are usually highly stressed and are made from "special" steels with improved toughness properties. The side shell and areas close to the neutral axis of the ship are the least stressed, and thus the toughness of the steels employed is of "secondary" importance. The remaining deck and bottom of a tanker are moderately stressed and are constructed of intermediate-toughness steels.

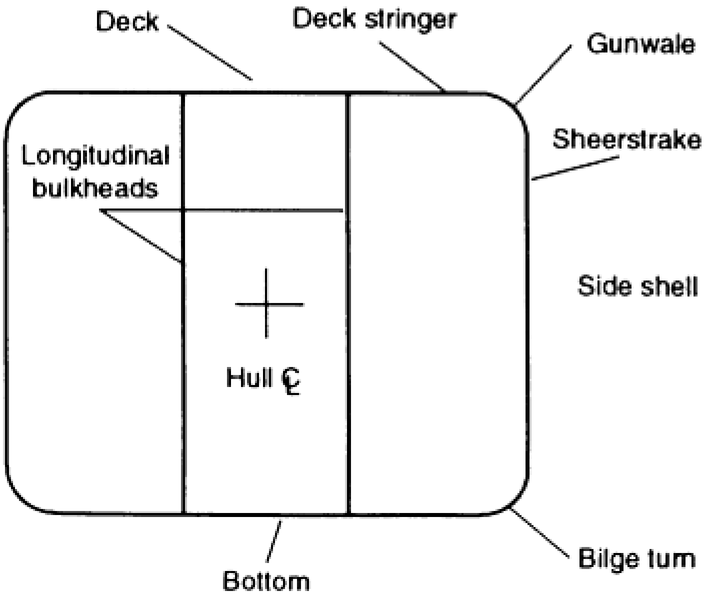

FIG. 7 MIDSHIP CROSS SECTION OF A LARGE TANKER, INDICATING SHELL PLATING AND LONGITUDINAL BULKHEADS.

The design considerations for tankers can also apply to container ships. However, because of the large cutouts on the deck that allow access for loading and unloading, the most critical structural area of a container ship is the box girder that connects the side shell to the deck. Major stresses are transmitted through this member, and thus the steels selected for these areas must have improved notch toughness.

Liquefied Natural Gas/Liquefied Petroleum Gas Ships. The minimum service temperature for most ships is approximately 0 °C (32 °F), and hull steels are selected and applied on this basis. Where temperatures below 0 °C (32 °F) are expected, application adjustments or selection of materials with superior low-temperature notch toughness must be considered. In the marine industry, low-temperature service is a significant factor in the design of ships carrying liquefied gases in bulk and of structures operating for prolonged periods in cold weather. Liquefied gases--both petroleum and natural--are transported in bulk by ship at temperatures ranging from -50 to -165 °C (-58 to -265 °F).

Steels for ship hulls are brittle at cryogenic temperatures. Thus, for safety reasons, it is essential to prevent the liquefied gas from contacting the steel hull and to minimize excessive conduction cooling of the hull. The primary requisite for the cargo tank material is high notch toughness at low temperature. Also, because shipboard tanks are subjected to dynamic loads resulting from ship motions, fatigue is another consideration.

Fracture mechanics concepts, which require knowledge of fracture toughness and fatigue crack-growth rate properties at low temperature, are frequently used in design. One design concept applied is "leak before failure," combined with a small-leak protection system. Utilizing this concept, a ship will be able to journey to an unloading and repair site before a detected flaw grows to critical size. Other systems are provided with large secondary containment barriers that can contain major leaks while permitting the ship to proceed to an unloading facility. All designs incorporate insulation to prevent excessive conduction cooling of the hull structure. For LNG containment systems, designed for cargo temperatures of -65 °C (-265 °F), external insulations of balsa, pearlite, polyurethane foam, and polyvinyl chloride foam have been used. For higher-temperature liquefied gases, such as LPG transported at a temperature of about -50 °C (-60 °F), some tanks have been internally insulated, which permits the use of less expensive tank structure materials than would be necessary with external insulation.

The worldwide adoption of International Maritime Organization (IMO) requirements for the carriage of liquefied gases by sea has unified material requirements for ships carrying cryogenic cargos in bulk. For the primary containment of such cryogenic cargos as LNG (methane) at a temperature of -165 °C (-265 °F), the most commonly selected materials have hexagonal close-packed (hcp) crystal structures; these include alloys such as 5083 aluminum, austenitic stainless steels, and 36% Ni steels, selected because of their superior low-temperature toughness properties. Nickel-bearing body-centered cubic (bcc) alloys, such as 9% Ni steels, have also been used, although less extensively.

The most important characteristic of the low-temperature mechanical behavior of ferritic steels is the ductile-to-brittle fracture transition. The transition occurs over a narrow temperature range for sharply notched specimens, and at temperatures below the transition, failure occurs by the brittle cleavage mode. This low-temperature embrittlement severely restricts the use of ferritic materials in cryogenic systems. Structural applications of ferritic materials are generally limited to environments where the service temperature remains above the ductile-to-brittle transition. The ferritic steels that are used for low-temperature applications usually contain nickel to improve low-temperature toughness. At the lower end of the cargo temperature range, austenitic alloys and 5083 aluminum are commonly used.

Table 11 indicates the service temperature specified in the ABS rules for each material. Austenitic stainless steel and 36% Ni steel are used for nonpressurized membrane tank designs. These alloys are provided in thin 0.5 to 1.5 mm (0.02 to 0.06 in.) sheets, usually corrugated to provide stiffness and allow for thermal expansion.

TABLE 11 MATERIALS FOR CRYOGENIC APPLICATIONS

MATERIAL	**ASTM NO.**	**SERVICE TEMPERATURE**	
		°C	**°F**
2% NI STEEL	A 203, GRADE A	-62	-80
	A 203, GRADE B	-59	-75
3% NI STEEL	A 203, GRADE D	-90	-130
	A 203, GRADE E	-79	-1L0
5% NI STEEL	A 645	-105(A)	-155(A)
9% NI STEEL	A 353	-196	-320
9% NI STEEL	A 553	-196	-320
AUSTENITIC STAINLESS STEEL	A 240	-196	-320
36% NI ALLOY	A 658	-196(B)	320(B)
ALUMINUM ALLOY	E 209, ALLOY 5083	-196	-320

(A) 5% NI STEEL MAY BE USED DOWN TO -165 °C (-265 °F) PROVIDED THAT IMPACT TESTS ARE CONDUCTED AT -196 °C (-320 °F).

(B) CHEMISTRY WILL BE SPECIALLY CONSIDERED FOR LOWERING THE COEFFICIENT OF EXPANSION.

For spherical pressurized tanks, 5083 aluminum alloy at a thickness of 40 to 60 mm (1.6 to 2.4 in.) is used to form tanks up to 35 m (115 ft) in diameter. The aluminum tank is attached to the steel hull by means of an explosion-bonded aluminum-to-steel plate; the tank connects to the aluminum side, and the steel side connects to the hull. This type of explosion-bonded plate is also used for connecting aluminum superstructures to steel decks.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Offshore Structures

Several important characteristics and operations of many offshore drilling units dictate the use of materials different from those used for ships, or restrictions on the application of similar materials. Much offshore exploration by mobile drilling or fixed units is carried out in areas where ambient air temperatures remain below 0 °C (32 °F) for long periods, such as in the Arctic. Toughness for these applications must be considered on the basis of the lowest expected steel temperature, because ferritic steels undergo a ductile-to-brittle fracture transition. The lowest steel temperature expected in service is estimated from meteorological data representing the coldest region in which the structure is designed to operate.

Toughness Versus Structural Application. In general, toughness testing of structural steel for offshore structures is conducted at a temperature below the minimum expected service temperature, the extent of the temperature increment being based on certain design criteria, such as anticipated stresses, stress concentrations, and redundancy of the structural member. For example, the ABS defines three structural categories to which different levels of toughness are assigned (see Table 12). The three categories are assigned to structural components based on consideration of the following: consequence of failure; redundancy of component; service experience with similar design; stress level, including stress concentration and restraint; fatigue; loading rate; and capability of inspection and repair in service.

TABLE 12 TEST TEMPERATURE FOR STEELS WITH A YIELD STRENGTH OF LESS THAN 415 MPA (60 KSI)

STRUCTURAL CATEGORY	DESCRIPTION	CHARPY V-NOTCH TEST TEMPERATURE
SECONDARY	LEAST CRITICAL: REDUNDANT	AT SERVICE TEMPERATURE
PRIMARY	MAIN STRUCTURAL MEMBERS THAT SUSTAIN PRIMARY TENSILE STRESS AND WHOSE FAILURE WOULD JEOPARDIZE THE SAFETY OF THE STRUCTURE	10 °C (18 °F) BELOW SERVICE TEMPERATURE
SPECIAL	MOST CRITICAL: FRACTURE CRITICAL MEMBERS THAT COULD EXPERIENCE RAPID LOADING AT POINTS OF STRESS CONCENTRATION	30 °C (55 °F) BELOW SERVICE TEMPERATURE

Absorbed energy obtained by a Charpy V-notch test is the most commonly used criterion for grading steels according to toughness. This criterion cannot be related directly to design, but is used as an expression of a toughness quality level.

Structural steel elements for offshore structures are often fabricated from cold-formed sections. As most structural steels are strain-age sensitive, it is important to evaluate the extent to which a given steel loses toughness after being cold strained. Outer fiber strains of less than 3% usually do not reduce toughness significantly. In general, at above 3% strain the material is either heat treated to restore the toughness or tested to determine that the poststrain toughness has not degraded below minimum acceptable values.

Quenched and Tempered Steels. Certain offshore drilling unit designs, especially the self-elevating type, make weight savings in the upper structure and the legs an important consideration in order to reduce the size and power requirements for lifting equipment and to aid stability. Therefore, high-strength quenched and tempered martensitic steels with yield strengths of approximately 700 MPa (100 ksi) find wider application in offshore structures than in ships. Areas subject to concentrated loads, such as openings for leg structures, cantilever overhangs for drilling booms, helicopter decks, and crane pedestals, are other locations on offshore drilling structures in which high-strength steels are commonly used.

Welding Process Selection

Welding has exclusively replaced riveting for joining structural steels in the fabrication of ships and offshore structures. In shipbuilding, welding permits large subassemblies to be built quickly and efficiently; these subassemblies are then welded together to form ship hulls.

Manual Welding. Shielded metal arc welding with stick electrodes, semiautomatic gas metal arc welding (GMAW), and flux-cored arc welding (FCAW) are all used in shipbuilding. These processes offer a high degree of versatility because of the portability of equipment and are commonly used for short weld runs and for erection welds.

Gravity-feed welding is used in the assembly of ships and offshore structures. It is carried out using contact or drag-type electrodes with tripod-type holding mechanisms, which allow welding to be done semiautomatically. An operator places the electrode in the holder and positions the holding mechanism, and welding progresses unrestricted until the electrode is consumed. Long electrodes enable welds more than 750 mm (30 in.) in length to be made with one electrode. Specific electrode diameters with proper current and voltage settings control the size of the deposited weld. One operator can handle two to six gravity-type devices at the same time.

This process has been largely employed for attaching stiffeners to plating by fillet welding. It is particularly attractive when the weld lengths are less than 3 m (10 ft), in which case the use of a fully automatic process may not be justified due to setup time, portability, and other considerations. Applications include internal stiffening of drilling-rig support columns and braces, as well as stiffening of decks.

Mechanized Welding. Submerged arc welding is used for making long, continuous butt welds to join plate panels or for joining stiffeners to plating by fillet welding. Modern shipyards that build large ships, especially tankers, where a considerable amount of flat panel work is required, have invested heavily in panel lines to further mechanize or automate the SAW process. Such panel lines are capable of making panel subassemblies up to about 20 m^2 (200 ft^2) in size.

One-side welding using SAW was introduced and perfected in Japan during the 1960s and early 1970s in connection with a massive tanker-building program for very large and ultralarge crude-oil carriers. In one-side welding, weld backing on the underside is provided by copper bars, ceramic forms, solid flux systems, or powdered flux held in place by various troughing systems.

Electroslag welding (ESW) and electrogas welding (EGW) have been used to make relatively long (more than 10 m, or 30 ft) continuous vertical erection welds in the side shells of large ships. These processes offer higher deposition and production rates than more conventional welding processes. The minimum plating thickness is about 19 to 25 mm ($\frac{3}{4}$ to 1 in.).

Consumable-nozzle ESW, a variation of the basic electroslag process, is used for making vertical butt welds in deck and bottom longitudinals and in heavy face plates of girders.

Oxyfuel Gas (OFG) and Plasma Arc Cutting. Oxyfuel gas cutting is used predominantly for cutting panels and stiffeners and for handfitting miscellaneous members. Plasma arc cutting has been used for mechanized cutting of plating up to about 12 to 20 mm (0.5 to 0.8 in.) thick. It offers the advantage of faster cutting speeds and less heat distortion as compared with gas cutting for thin plating. Because the significant tonnages of plating and shaping that must be cut or beveled for welding large ships, a high degree of mechanization has been introduced. Numerical control cutting permits panels up to about 2.5 by 15 m (8 by 50 ft) to be completely cut to size with appropriate cutouts, as well as edge beveling for welding to make large subassemblies.

Weld Considerations

The materials used for ship construction are generally ordinary-strength or higher-strength steels with a maximum yield strength of 355 MPa (52 ksi) (see Tables 9 and 10). For offshore drilling units and platforms, a variety of high-strength quenched and tempered steels with yield strengths up to 700 MPa (100 ksi) are also used. Weld filler metals should approximate the tensile and toughness properties of the base material; where materials of two different strengths or toughnesses are to be joined, the lower-strength or lower-toughness weld metal can be used.

Weld metal toughness is an important parameter in shipbuilding. Quantitative measurements of toughness are based on Charpy V-notch properties, and values comparable to those of the base materials are specified for weld metal. For offshore structures such as fixed platforms or drilling units, which must stay on station and encounter severe storms, winds, waves, and temperature fluctuations, toughness is also an important consideration. Materials and weld metals with yield strengths of up to 700 MPa (100 ksi) are commonly offered for tubular leg chord sections, braces, and rack material. The racks that form part of the leg chord are often 100 to 150 mm (4 to 6 in.) thick.

Hydrogen Cracking. When joining thick high-strength quenched and tempered steels, appropriate preheat and interpass temperatures are necessary to avoid HIC (see the article "Cracking Phenomena Associated with Welding" in this Volume). This condition is generally more prevalent in HSLA steels than in ordinary-strength steels, and its occurrence is generally proportional to the hardenability and tensile strength of the base material. Hydrogen cracks are usually found in the HAZ and can occur in both fillet- and butt-welded connections. Hydrogen cracking is often referred to as cold or delayed cracking because of the tendency for these cracks to occur and propagate after the weld has cooled to ambient temperatures. It is not uncommon for such cracks to be revealed hours or even days after welding, depending on the restraint conditions within the structure. For this reason, it is common practice to postpone the final nondestructive testing until several days after completing a weld.

Hydrogen sources in and around the arc atmosphere include moisture in the electrode coating, flux, gas, and base material and the presence of hydrocarbons, such as grease, paint, or oil. Moisture requirements for electrode coatings, along with storage and conditioning recommendations, are included in specifications and become more severe as strength levels increase. In SAW, where the flux is a common source of moisture, storage should be controlled to avoid moisture contamination from the atmosphere. Inadequate protection from hydrogen must be provided, especially in final erection or field welding applications, where hydrogen sources are more prevalent and where accessibility to the welded joint is less than ideal.

Lamellar tearing is the separation of steel plate underneath a weld in a plane generally parallel to the plate surface. It usually results from high through-thickness strains induced by weld shrinkage. These strains cause nonmetallic inclusions present in the steel to decohere. The resulting tears or cracks are planar in appear- ance, with some steplike characteristics. Welded connections joining two or more tubular sections are common in offshore structures. On several occasions, lamellar tearing has been observed in diaphragm plates used between two tubulars to simplify construction. Improvement in through-thickness properties is possible by using steels with fewer impurities or nonmetallic inclusions, lowering the permitted sulfur contents (typically 0.010% maximum), and treating the molten steel with rare earth elements, which combine with sulfur to form refractory sulfides that resist deformation at hot rolling temperatures. Another approach to eliminate lamellar tearing, and to simplify fabrication, is to use steel castings for the node connections. Centrifugally cast straight tubulars have also been proposed for offshore structural applications.

Lamellar tearing has been reported in sheer-strake-to-deck connections in ships. This problem can be eliminated by the use of gunwale connections. Lamellar tearing or cracking at welded connections is best avoided by the selection of a design that reduces the tendency for stressing through the thickness of the structural members.

In the construction of offshore drilling units and platforms, lamellar tearing presents a more acute problem because of the complexity of welded joints and connections. Sometimes joints may be unavoidably susceptible to lamellar tearing; in such cases, carefully planned welding sequences should be employed. It is common practice to use other than full-penetration welds; for instance, double-bevel partial-penetration T-welds are often used. Steel exhibiting the equivalent of 20% (minimum) reduction of area in a through-thickness tensile specimen (as per ASTM A 770) is considered to be less susceptible to lamellar tearing. Commonly used structural steels, including ABS hull steels, that meet the A 770 specification are available from some producers. Such steels have additional restrictions on chemistry and are produced using special melting practices.

Undermatching Weld Metal Strength. In some cases, weld metal that undermatches the base plates may be used in the fabrication and repair of high-strength steels, primarily for fillet welding and local repairs. The use of the lower-strength weld metal minimizes cracking tendency by providing a weld with reduced residual stresses and somewhat higher ductility. Such lower-strength filler metals are permitted when the strength level of the particular weld fulfills design requirements. When lower-strength covered electrodes are used to weld higher-strength materials, special precautions associated with welding the higher-strength materials are still required to prevent hydrogen cracking. When E80*xx* or E90*xx* electrodes are used to weld 700 MPa (100 ksi) yield strength material, such as ASTM A 514, the electrodes should be subjected to appropriate baking to reduce electrode-covering moisture to 0.15% maximum. This precaution should also be taken in the more common case where E7018-type electrodes are used to join high-strength steels (e.g., 700 MPa, or 100 ksi, yield strength) to lower-strength steels.

To summarize, in the following circumstances, the use of undermatching electrodes should be considered:

- *FILLET WELDING:* USE OF E9018 OR EVEN E8018 ELECTRODES FOR FILLET WELDING OF STEELS WITH A YIELD STRENGTH OF 700 MPA (100 KSI) IS A CONVENIENT WAY OF MINIMIZING CRACKING TENDENCY IN THE WELD METAL AND HAZ. ELECTRODES SHOULD BE BAKED TO A 0.15% MAXIMUM MOISTURE CONTENT.
- *REPAIR WELDING:* THE LOWER STRENGTH AND LOWER RESIDUAL STRESS ASSOCIATED WITH THE E9018 ELECTRODE MAKE IT THE ELECTRODE OF CHOICE FOR ISOLATED REPAIRS IN BUTT OR FILLET WELDS IN THE HIGHER-STRENGTH STEELS (700 MPA, OR 100 KSI, YIELD STRENGTH), PROVIDED THAT THE OVERALL STRENGTH OF THE STRUCTURE IS MAINTAINED TO MEET DESIGN REQUIREMENTS.

Aerospace

Aerospace systems push materials to the limits of their physical and mechanical properties. Therefore, it is important to optimize designs, which often involves consideration of nonmechanical fastening methods. Welding, brazing, soldering, and/or adhesive bonding are the only options in many aerospace structures. Whether the driving force is weight, leakproof seals, continuous load paths, or any number of other requirements, nonmechanical joining of materials is essential in the manufacture of aerospace systems. The list of materials used in the construction of such systems is quite large. This section will present an overview of mechanical and physical property considerations as they relate to joining.

Material Selection Criteria

Properties. Modern aerospace systems employ metals, ceramics, composites, and polymers. Important mechanical and physical properties for aerospace applications are listed in Table 13 (Ref 7).

TABLE 13 MATERIAL PROPERTIES OF IMPORTANCE IN AEROSPACE APPLICATIONS

PHYSICAL PROPERTIES

CRYSTAL STRUCTURE
DENSITY
MELTING POINT
VAPOR PRESSURE
VISCOSITY
POROSITY
PERMEABILITY
REFLECTIVITY
TRANSPARENCY
OTHER OPTICAL PROPERTIES
DIMENSIONAL STABILITY
ELECTRICAL/MAGNETIC PROPERTIES
CONDUCTIVITY
DIELECTRIC CONSTANT
COERCIVE FORCE
HYSTERESIS
SUSCEPTIBILITY
PERMEABILITY
REMANENCE
MECHANICAL PROPERTIES
HARDNESS
MODULUS OF ELASTICITY TENSION
COMPRESSION
POISSON'S RATIO
STRESS-STRAIN CURVE
YIELD STRENGTH TENSION
COMPRESSION
SHEAR
ULTIMATE STRENGTH TENSION
SHEAR
BEARING
FATIGUE PROPERTIES SMOOTH
NOTCHED
CORROSION FATIGUE
ROLLING CONTACT
FRETTING
CHARPY TRANSITION TEMPERATURE
FRACTURE TOUGHNESS (K_{IC})
HIGH-TEMPERATURE PROPERTIES CREEP
STRESS RUPTURE
DAMPING PROPERTIES
WEAR PROPERTIES GALLING
ABRASION
EROSION
CAVITATION

SPALLING
BALLISTIC IMPACT
THERMAL PROPERTIES
CONDUCTIVITY
SPECIFIC HEAT
COEFFICIENT OF EXPANSION
EMISSIVITY
ABSORPTIVITY
ABLATION RATE
FIRE RESISTANCE
MAXIMUM/MINIMUM OPERATING TEMPERATURE
CHEMICAL PROPERTIES
POSITION IN ELECTROMOTIVE SERIES
CORROSION AND DEGRADATION ATMOSPHERIC
SALT WATER
ACIDS
HOT GASES
ULTRAVIOLET
OXIDATION
THERMAL STABILITY
BIOLOGICAL STABILITY
STRESS CORROSION
HYDROGEN EMBRITTLEMENT
HYDRAULIC PERMEABILITY
FABRICATION PROPERTIES
CASTABILITY
HEAT TREATABILITY
HARDENABILITY
FORMABILITY
MACHINABILITY
WELDABILITY

Source: Ref 7

Applications. The service requirements for a particular application will order the material properties. As shown in Fig. 8, material selection is based on a number of interrelated factors, including material properties, manufacturability, repairability, and availability (Ref 8). Because compromise is required to ensure optimal material selection, both the service requirements and the materials characterization must be thorough and complete. Design also plays an important role. For example, if a material is selected for a design using mechanical fastening, and nonmechanical joining is substituted as the manufacturing technique, the material and the design must be reevaluated. A design requiring mechanical fastening generally is not optimal for nonmechanical joining.

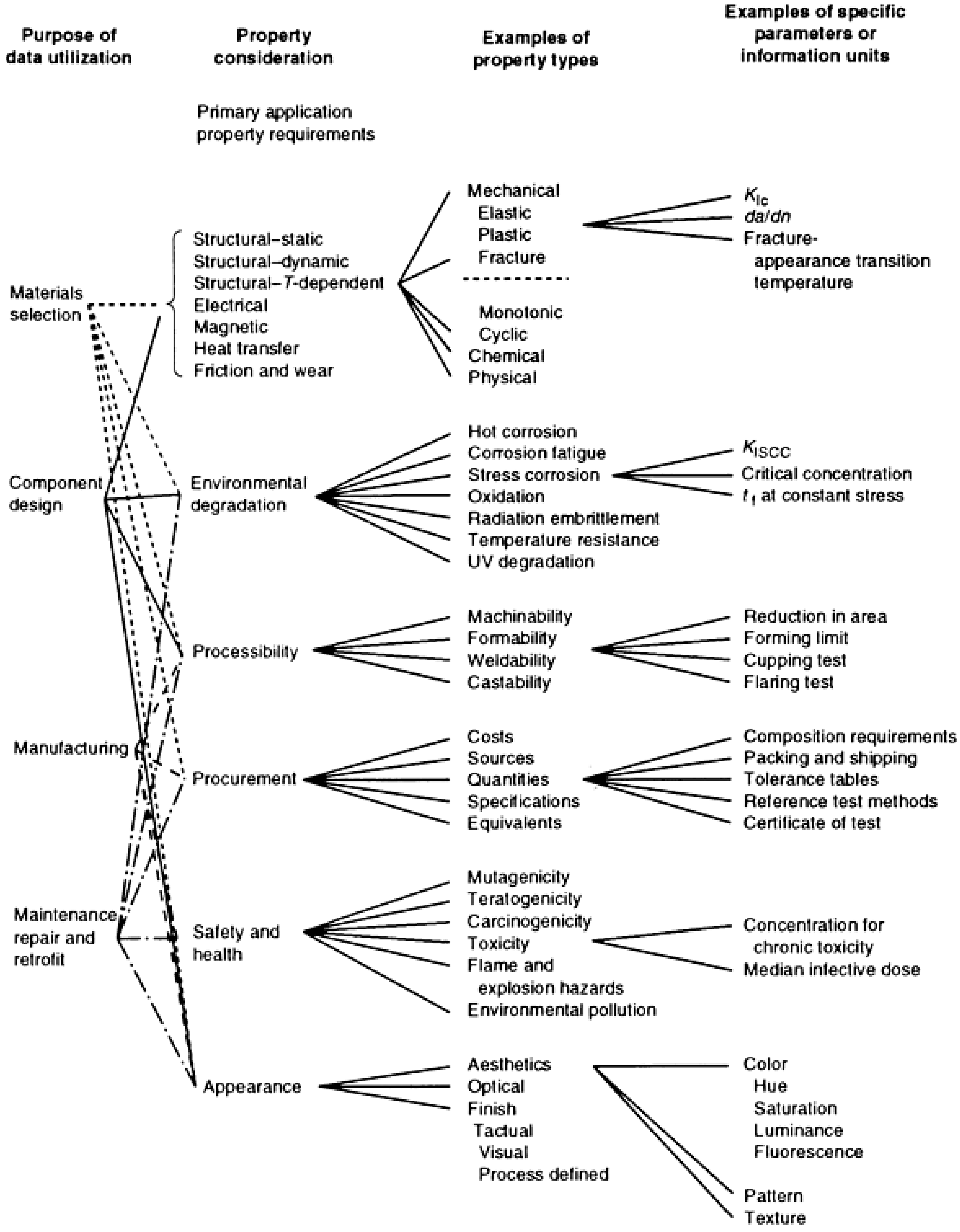
Purpose of data utilization
Property consideration
Examples of property types
Examples of specific parameters or information units
Primary application property requirements
Materials selection
Component design
Manufacturing
Maintenance repair and retrofit
Structural–static
Structural–dynamic
Structural–T-dependent
Electrical
Magnetic
Heat transfer
Friction and wear
Mechanical
Elastic
Plastic
Fracture
Monotonic
Cyclic
Chemical
Physical
K_{Ic}
da/dn
Fracture-appearance transition temperature
Environmental degradation
Hot corrosion
Corrosion fatigue
Stress corrosion
Oxidation
Radiation embrittlement
Temperature resistance
UV degradation
K_{ISCC}
Critical concentration
t_f at constant stress
Processibility
Machinability
Formability
Weldability
Castability
Reduction in area
Forming limit
Cupping test
Flaring test
Procurement
Costs
Sources
Quantities
Specifications
Equivalents
Composition requirements
Packing and shipping
Tolerance tables
Reference test methods
Certificate of test
Safety and health
Mutagenicity
Teratogenicity
Carcinogenicity
Toxicity
Flame and explosion hazards
Environmental pollution
Concentration for chronic toxicity
Median infective dose
Appearance
Aesthetics
Optical
Finish
Tactual
Visual
Process defined
Color
Hue
Saturation
Luminance
Fluorescence
Pattern
Texture

FIG. 8 INTERRELATIONSHIPS AMONG AEROSPACE DESIGN, MANUFACTURE, MAINTENANCE AND REPAIR, AND MATERIAL SELECTION. SOURCE: REF 8

Thermal Stability. A clear understanding of the material properties of importance for a given application is crucial to optimal material selection. For example, thermal stability, both metallurgical and physical, is a very important factor in elevated-temperature applications. Over time, microstructural thermal instability may alter mechanical properties, such as toughness, creep, or crack growth. The thermal stability of a material may be affected by a thermal joining process. Precipitation of unstable phases in the fusion zones of structural alloys, such as advanced titanium alloys, can be difficult to evaluate. Also, the tensile strength of high-strength aluminum alloys can be adversely affected during high-temperature curing of adhesives if the aging temperature is exceeded for a sufficient length of time (Ref 9). Thermal stability can be maintained by ensuring proper material chemistry, heat treatment, joining process, joining consumable, and/or postjoining thermal treatments. Thermal instability may manifest itself as creep, cracking, buckling, and/or a reduction in corrosion resistance (Ref 4, 10, 11, 12). The corrosion properties of stainless steels may be severely impaired by elevated-temperature exposure, such as welding and brazing. The degradation referred to as sensitization can also reduce fatigue performance. Thermal stability has been used here as an example. A similar discussion could be presented for each property listed in Table 13.

Weight is an overriding concern in aerospace applications because of its adverse effect on performance and operational costs. Many decisions are based on the specific property of a material, such as those shown in Fig. 9. Specific properties are defined as the mechanical or physical property of the material divided by its density:

$$\frac{\textit{Material}\text{ property}}{\text{Material density}} = \textit{specific}\text{ property}$$

When using specific properties, one must be careful to fully characterize anisotropic materials. When considering a material for a nonmechanical joining application, an understanding of the effect of joining on these properties is essential. Joinability often is ignored during the development of aerospace materials (and materials in general) and can considerably inhibit their versatility during construction. Because compromise is the standard when selecting materials for a specific application, the effect of joining on properties must be considered.

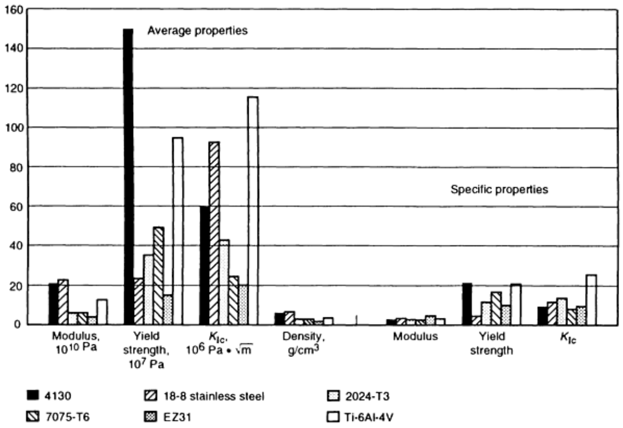

FIG. 9 AVERAGE VERSUS SPECIFIC PROPERTIES

Specifications. Aerospace materials are most often purchased to a military (MIL) or Aerospace Materials Specification (AMS). Some materials, such as A286, are available in standard and weldable grades. The weldable grade is formulated to reduce the tendency for microcracking. Filler metal specifications are provided by AWS or by the Society of Automotive Engineers (SAE) as Aerospace Materials Specifications.

Economics. Cost is another important material consideration. The cost of the raw materials is usually a small percentage of the cost of the end product in applications such as aerospace airframes and engine components, structures, and assemblies. The costs of some common materials as a function of volume production are presented in Fig. 10. However, material selection can have a significant cost impact in the areas of manufacturability and life-cycle costs (repairability and maintainability). The designation of materials must be as specific as possible to ensure maintenance of quality for joining applications (Ref 10, 11, 13). If multiple welds are to be performed on a given structure, the requirement for intermediate heat treatments or other processing steps must be considered. In large structures where full heat treatments are difficult because of distortion, the reduction in properties may require increased joint thickness, thereby increasing weight and complicating the manufacturing process.

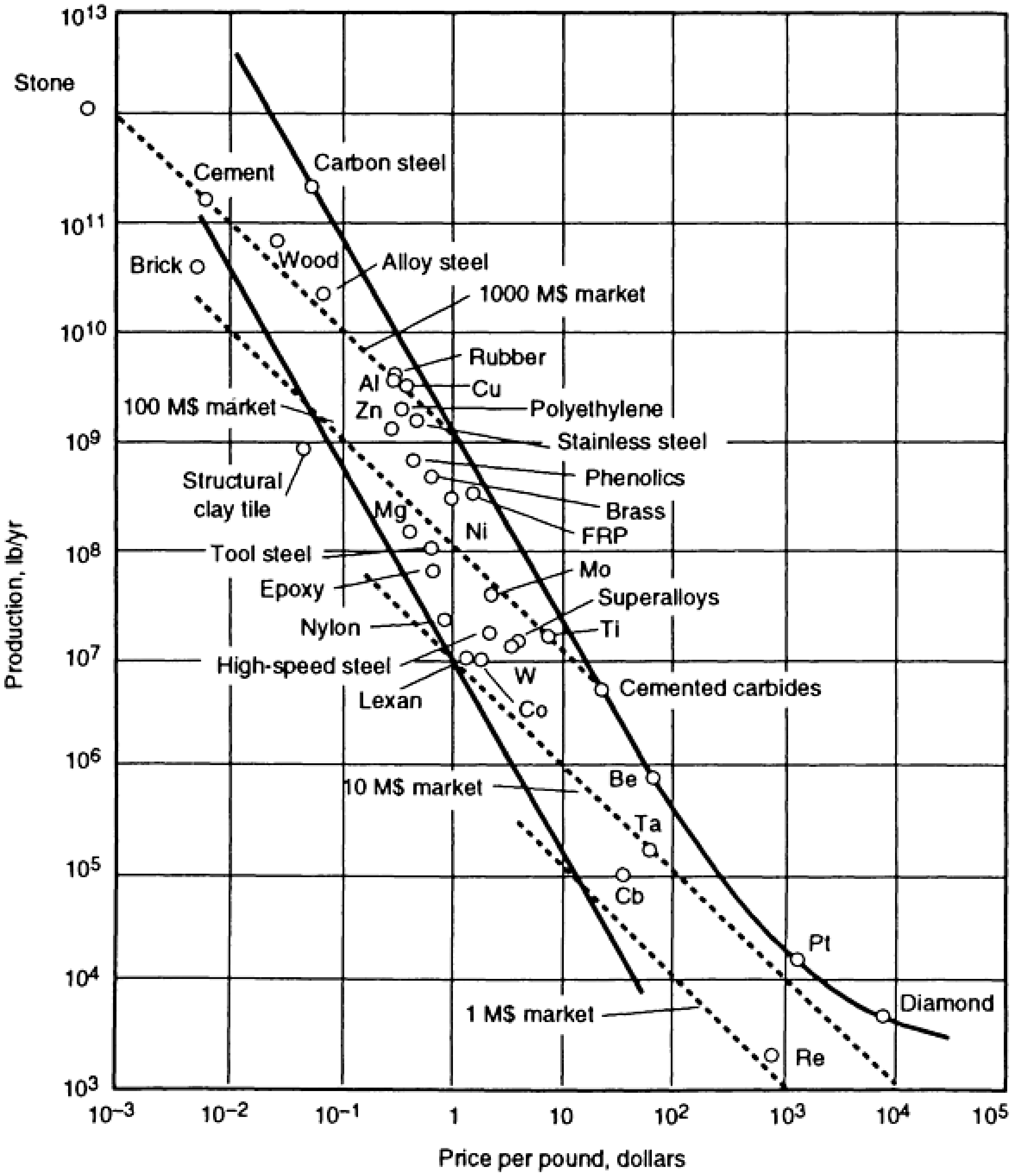

FIG. 10 COST VERSUS PRODUCTION FOR SEVERAL COMMON AEROSPACE MATERIALS. SOURCE: REF 8

Joining process selection and practice are influenced by material selection. For example, titanium requires vacuum or complete inert shielding (chambers with glove ports) to produce high-quality joints when elevated-temperature joining processes are used, which increases production costs in comparison with some other materials.

Finally, the environmental ramifications of joining must be considered when selecting a material. Toxic materials, such as beryllium, are sometimes used in aerospace structures because of their very high specific properties. However, when joining these materials, especially by welding, health and safety issues may drive manufacturing costs very high.

References cited in this section

4. S. YUKAWA, *GUIDELINE FOR PRESSURE VESSEL SAFETY ASSESSMENT,* NIST SPEC. PUBL. 780, NATIONAL INSTITUTE FOR STANDARDS AND TECHNOLOGY, P 14-22
7. G.E. DIETER, *ENGINEERING DESIGN,* MCGRAW-HILL, 1983
8. B.D. TAPLEY, *ESCHBACK'S HANDBOOK OF ENGINEERING FUNDAMENTALS,* JOHN WILEY & SONS, 1990
9. *AEROSPACE STRUCTURAL MATERIALS HANDBOOK,* VOL III, 1984
10. SIMS AND HAGEL, *THE SUPERALLOYS,* JOHN WILEY & SONS, 1972
11. CASTRO AND DECADENT, *WELDING METALLURGY OF STAINLESS AND HEAT RESISTING STEELS,* CAMBRIDGE PRESS, 1968
12. EVANS AND WILSHIRE, *CREEP IN METALS AND ALLOYS,* DOTESIOS PRINTERS, U.K., 1985
13. *AEROSPACE STRUCTURAL MATERIALS HANDBOOK,* VOL IV, 1984

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions[*]

Joinability

Material Characteristics Related to Joining. Materials for aerospace use can be divided into two major categories: structural and engine, the latter being materials for elevated temperatures. There is considerable overlap, but it is instructive to examine materials from these points of view. Although the practice of welding of critical structural members in modern aircraft is small, but growing, the use of welding in missiles and space hardware has always been extensive. The joinability of aerospace materials takes on various definitions. Joint efficiency, cracking resistance, welding characteristics (such as flow), maintenance of corrosion resistance, and long-term metallurgical stability are all measures of joinability. Design, process, and consumable selection affect the joining characteristics of materials.

Dissimilar-materials joining muddies these definitions even further. For example, the welding of titanium to stainless steel is not recommended, but brazing these two materials together is possible. The usefulness of the hybrid braze would be determined by the service requirements. Joinability is often determined by standardized tests or by the ability of a material to meet certain mechanical or inspection requirements.

When evaluating joinability, it is important to consider the range of conditions to be experienced n production. For example, most material specifications allow a range of compositions. Because joinability may vary considerably as a function of chemistry, addressing these concerns prior to production will avoid costly rework.

The joinability of a material is influenced by the condition of the material when joined. For example, nickel-base superalloys should always be stress relieved prior to brazing to prevent stress cracking (Ref 14). Heat-treatable aluminum alloys are sometimes welded in the T4 condition and then postweld aged to the T6 condition to achieve more uniform properties in the weldment. Alpha-beta titanium alloys are sensitive to preweld heat treatment.

The thermal history can become very important during repair operations. If a material has been exposed to elevated temperatures or high loading in service, its joinability may be altered. In such instances, specialized testing may be required to ensure adequate joinability for repair operation.

The joinability of a material may be contingent on the postjoining operations, such as heat treatments. Many as-joined materials do not have the properties required for a particular application, but can be brought to an acceptable condition by a postjoining process. Postjoining thermal treatments can consist of full heat treatments (solution, intermediate, age), stress relief, or stabilization treatments (age). These can be localized or can be applied to the entire part. Other postjoining treatments include shot peening, machining, vibration stress relief, and sizing.

Base Materials. When characterizing the joinability of a material, it is important to consider the effect of joining on the base material as well as on the nexus (the actual joint; i.e., the fusion zone in welding). Nonthermal joining operations, such as room-temperature-cured adhesive bonding, may have no effect on the base material. In contrast, thermal processes, such as brazing, soldering, and welding, can radically change material properties, especially near the nexus. Characterization of the affected base material can be difficult because of its size and makeup. For example, the HAZ in an electron beam weld in thick-section titanium may be as little as half the fusion-zone width. If the weld is on the order of 2.5 mm (0.10 in.) in width, the HAZ is only 1.25 mm (0.05 in.) in width. Locating a notch in the HAZ to evaluate fatigue or fracture toughness can be challenging. Moreover, if the HAZ is small enough, its effect on the service performance of the joint may be negligible, which signals a caution when service life is evaluated by using "simulated" HAZ microstructure. Joints of the same configuration as the production parts should be used whenever possible.

Filler Materials. The selection of filler materials is a critical factor that sometimes receives less than proper attention when joining aerospace materials (Ref 15). Filler materials are most often selected on the basis of chemistry match. However, other considerations sometimes take precedence, such as ferrite control in the welding of stainless steels, distortion/residual stress control for high-strength steels, optimization of elevated-temperature properties for nickel-base superalloys, and strength optimization in aluminum welding or optimization of environmental resistance in brazing and adhesive bonding.

Selection of filler materials for dissimilar-material joining requires careful consideration. Some brazing alloys contain melting-point depressants to allow brazing at temperatures lower than the operating environment (diffusion brazing). The effect of the formation of detrimental phases in the base materials is a factor when these alloys are used. Special corrosion or wear requirements will sometimes direct the use of a particular joining consumable.

Joining Techniques Other Than Fusion Welding. Process is another consideration in the joinability of aerospace materials. Alloys that are nonfusion weldable may be readily joined by other methods, such as friction, explosion, diffusion, brazing, or soldering.

Service Considerations. The joinability of materials is influenced by service requirements. A titanium weldment used for a missile fuselage or a wing support must have high strength, low fatigue crack-growth rates, and high toughness. Good corrosion resistance and thermal stability are also desirable. The properties of the weldments made from such material must reflect the base material requirements. On the other hand, an Inconel 626 weld used to build up a turbine blade tip may not have the same requirements as the turbine blade. Ease of welding and adequate elevated-temperature oxidation resistance are overriding concerns in this case. Because the tip buildup is not "structural" in nature and the consequences of failure not as severe as for the wing support, weldability is based on different criteria. Therefore, the joinability of a given material can vary depending on the end service requirements.

References cited in this section

14. *BRAZING HANDBOOK,* AMERICAN WELDING SOCIETY, 1976
15. D. HARVEY, FILLER METAL SELECTIONS FOR CRITICAL AEROSPACE FABRICATION, *WELD. DESIGN FABR.,* VOL 60 (NO. 2), MARCH 1991, P 765-80

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Damage Tolerance

Existence of Flaws. Damage tolerance is a design methodology often used for aerospace systems. The basic premise is that all structures contain inherent flaws. The effect of these flaws on flight safety depends on their geometry, growth rate at the service conditions, and critical size, and on the inspectability of the structure and the consequence of failure (Ref 16).

In the design of damage-tolerant structures, it is imperative that the relevant material properties, such as fatigue crack-growth rates, fracture toughness, and SCC, be well characterized. When applying damage-tolerant methodologies to structures joined by other methods than mechanical, unique considerations include:

- RESIDUAL STRESS EFFECTS
- PROPERTY GRADIENTS ACROSS THE JOINT
- INHERENT FLAWS, SUCH AS MICROFISSURES AND POROSITY
- LONG-TERM STABILITY
- CORROSION
- STRESS CONCENTRATIONS

These concerns are not well characterized for many of the materials and joining processes used in aerospace.

Finally, because the properties of the base metal may be altered nonuniformly during thermal joining, specialized testing may be required. For example, the crack-growth rate may be improved in the fusion zone of a weld, but the fracture toughness and SCC resistance may be reduced in the HAZ. All of these factors are important to the damage-tolerance analysis, and thus full characterization is required.

Inspection Criteria. For complex structures, the limiting factor may well be the inspectability of the joint. As a result, many welding specifications used for aerospace production, such as MIL-STD-2219 and AMS 2680, include defect allowable tables. When contractually requiring one of these standards, it is important to review the defect allowables to ensure that they are realistic for the application. Many unnecessary repairs are performed on welded structures because of improper defect criteria. As a result of the additional thermal cycles, the repairs may be more detrimental than the flaw in many materials. For many joining processes and materials combinations used in aerospace, the information necessary to perform a damage-tolerance analysis is not available; therefore, the defect criteria in the welding specifications are used, as this procedure is less expensive than developing the required data. Also, the effect of residual stress on fracture toughness and fatigue crack growth is difficult to evaluate with current damage-tolerance methodologies. Much work needs to be accomplished in this area.

Nondestructive examination (NDE) concerns require consideration of both the process and design limitations. Many times joints are designed without adequate attention to inspection requirements, often making effective NDE impossible due to inaccessibility.

Reference cited in this section

16. *USAF DAMAGE TOLERANT DESIGN HANDBOOK: GUIDELINES FOR ANALYSIS AND DESIGN OF DAMAGE TOLERANT AIRCRAFT,* U.S. AIR FORCE, 1984

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions[*]

Advanced Materials

Some materials used for aerospace systems present special problems in relation to joining. Among these materials are composites and intermetallics. Composite materials, including metal matrix and organic matrix, are used because of their increased stiffness, damping properties, and high specific properties. Composites present challenges when joined by any method, and much work is currently underway to establish their joinability. Maintenance of properties when using nonmechanical joining is difficult. Metal-matrix composites have been successfully joined by brazing, diffusion, and high-energy density welding. During the development of these materials, joinability was not a high priority and is thus limited.

The bond integrity between the fibers and the matrix of a composite greatly affects material properties. Special coatings are applied to the fibers to ensure adequate bonding. The stability of these coatings to a large degree determines joinability when using elevated-temperature processes. Coating schemes that optimize joinability should be addressed during materials development. A further challenge involves the attainment of material properties across the joint.

Intermetallics, such as Ti_3Al_2, offer very promising elevated-temperature properties. Limited studies to date concerning the fusion weldability of these materials indicate that tight control of base-metal chemistry, accurate regulation of heat input and cooling rate, and joint design to minimize residual stress are required to produce sound joints (Ref 17). Brazing offers some advantages, but may be limited by operating temperature unless suitable diffusion-bonding fillers can be developed. The joinability of these types of emerging materials will to a large degree determine their long-term usefulness in aerospace systems.

Reference cited in this section

17. PATTERSON *ET AL.* TITANIUM ALUMINIDE: ELECTRON BEAM WELDABILITY," *WELD. J.,* JAN 1990, P 39S-44S

Machinery and Equipment

Welding plays an important role in the wide variety of structures broadly characterized as machinery and equipment. This section will discuss three categories: stationary equipment, such as presses, rolling mills, and machine tools; mobile equipment, such as cranes, earthmoving, and other construction equipment; and rotating components of machines. In all of these categories, the materials requirements for serviceability are distinctly different for the structural machine support parts and for those involving the machine-moving and working parts.

Applications

Stationary Equipment. The working parts of machine tools require stability. Massive cast iron or cast steel bases are used to provide a damping capacity, thus ensuring that cutting tools do not vibrate or chatter. In applications that use weldments, thick steel components are specified.

The principal design requirement for metalworking presses is adequate strength to counter the forces imposed during press operation. Weldments of various sections using both wrought and cast carbon steels offer adequate strength and rigidity. The elastic deflections of the components created by work forces affect the design, which must take into account the cyclical nature of the stress and hence the fatigue properties of the materials and welded joints.

Mobile Equipment. In the design of components for earthmoving and other construction equipment, static and dynamic loads are important factors, as are erosion, wear, weight, and ease of maintenance. The materials for each working part of such equipment must be considered separately.

As an example, booms used in mobile industrial cranes must be light in weight, yet rigid under load. Wear resistance is another important factor in this application.

Rotating Elements. Materials specified for the rotating parts of pumps, fans, compressors, and similar machinery have a wide variety of properties. For large rotating components, weldments are commonly required, sometimes employing different metals for improved performance in specific environments. Erosion, corrosion, and cavitation are often the mechanisms that limit the lives of such machine elements. Critical surfaces are frequently weld surfaced with corrosion-resistant materials.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Design Considerations

Welds and members of weldments are sized to withstand the working stress and stress range induced by static and fluctuating loads. The magnitude of the fluctuation between maximum and minimum working stress is termed the working stress range. The magnitude of the static and fluctuating stresses is influenced by such factors as dead load, live load, type of operating mechanism, machine geometry, component size, shock or impact, accidental overloads, corrosion, wear misalignment, stress concentration, and weld joint design. The allowable stress and stress range must be greater than the expected working stress and stress range. The relationships among the various stresses, stress ranges, strengths, and loads are shown graphically in Fig. 11.

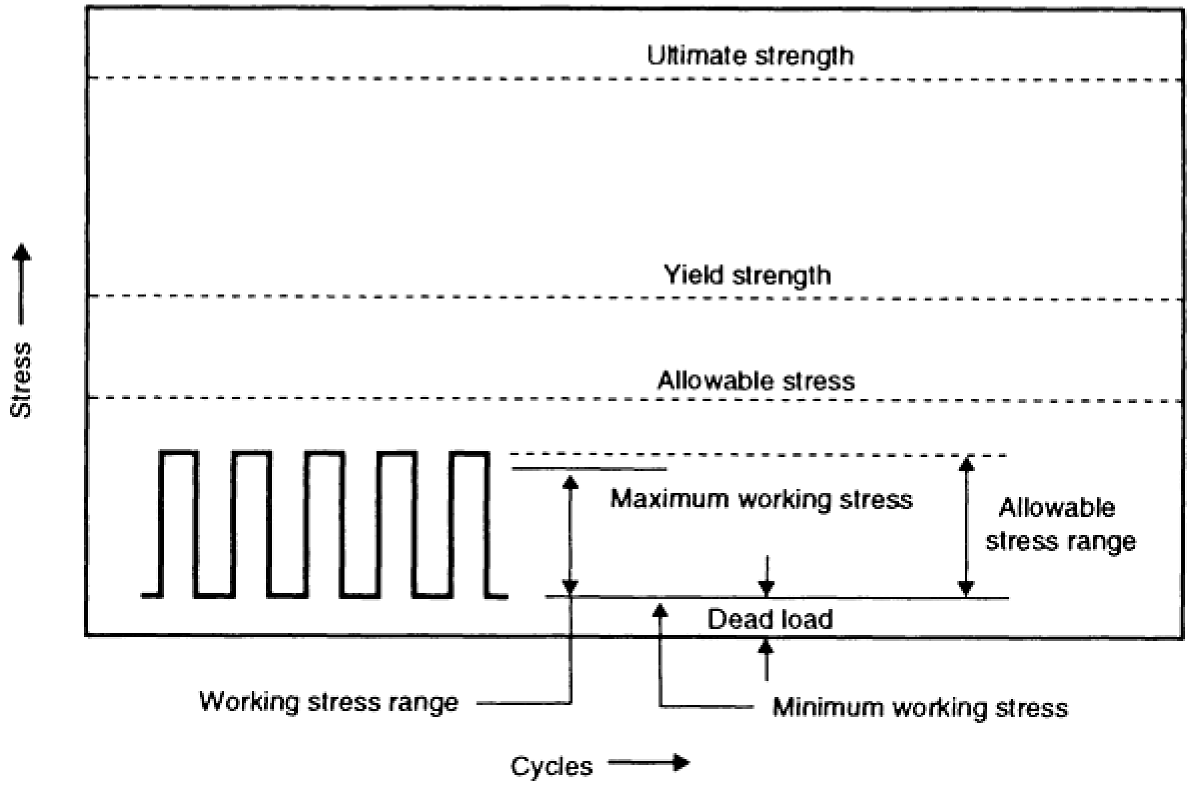

FIG. 11 STRESS RELATIONSHIPS RESULTING FROM STATIC AND FLUCTUATING LOADS

Allowable stresses in weld and base metals are usually specified by an applicable code or standard based on a history of satisfactory operating performance. In the absence of an industry standard, as in the case of most machinery and equipment, the determination of an allowable stress value is left to the discretion of the manufacturer or designer. In such case, the allowable stress (SA) is calculated by dividing the ultimate tensile strength (UTS) or yield strength (YS) by an appropriate factor of safety (FS) selected by the designer:

$$SA = \frac{UTS}{\mathrm{FS}} \text{ or } \frac{\mathrm{YS}}{\mathrm{FS}}$$

The selection of an appropriate factor of safety to prevent failure from unpredictable causes is empirical and depends on industry data and the experience and judgment of the designer. Table 14 provides guidance in selecting a factor of safety, based on the ultimate tensile strength of the material. The assigned factors in Table 14 may be increased as warranted by safety, cost, plant utilization, and engineering judgment. Also, the assigned factors may be reduced if detailed analysis increases confidence in design factors, such as loading, material properties, and fabrication controls.

TABLE 14 FACTOR OF SAFETY SELECTION CRITERIA

CONDITION	**ASSIGNED FACTOR OF SAFETY$^{(A)}$**
EXCEPTIONALLY RELIABLE MATERIALS USED UNDER CONTROLLABLE CONDITIONS AND SUBJECTED TO LOADS AND STRESSES THAT CAN BE DETERMINED WITH CERTAINTY. USED ALMOST INVARIABLY WHERE LOW WEIGHT IS A PARTICULARLY IMPORTANT CONSIDERATION	2.5-3
WELL-KNOWN MATERIALS, UNDER REASONABLY CONSTANT ENVIRONMENTAL CONDITIONS, SUBJECTED TO LOADS AND STRESSES THAT CAN BE DETERMINED	4-5
LESS TRIED OR BRITTLE MATERIALS UNDER AVERAGE CONDITIONS OF ENVIRONMENT, LOAD, AND STRESS	5-6
KNOWN MATERIALS THAT ARE TO BE USED IN UNCERTAIN ENVIRONMENTS OR SUBJECTED TO UNCERTAIN STRESSES	6-8

(A) FOR REPEATED LOADS, THE FACTORS SHOULD BE DOUBLED. FOR IMPACT LOADS. THE FACTORS SHOULD BE INCREASED AS WARRANTED.

When a design is analyzed using a mathematical model, it must be understood that subsequent calculations only approximate the working stress. With all models, various assumptions, limitations, and simplifications used to fit equations to the component model influence the accuracy of the calculations. (This is true even when using finite-element methods.) The calculated stress values that result from applying "good" theory to a "good" model are, at best, ballpark figures that give the design engineer a rational basis for making design decisions. For the most part, the allowable stress ranges specified in current standards for welded joints subjected to fatigue loading are based on testing of representative full-size welded joints in actual or mockup structures.

Static Loads. Nearly every machine is subjected to static and fluctuating loads; few useful machine structures have only static loads applied. Structures subjected to slowly applied loads, few cycles (less than 20,000) of loading, or nearly steady loads are considered statically loaded and are designed using rules that apply to static structures. These rules assume that localized areas of high stress due to stress concentrations are relieved by yielding of the highly stressed material to redistribute internal stresses.

Steels commonly used for welded construction of machinery and equipment have yield strengths of less than 700 MPa (100 ksi) and are listed in Table 15. Class I materials have the highest weldability, and class V have the lowest. Carbon and carbon-manganese steels--those class I and class II steels designated by the American Iron and Steel Institute (AISI), SAE, and ASTM--are the most commonly used materials for welded machinery and equipment. These steels do not require preheat unless the thickness is greater than 150 mm ($1\frac{1}{2}$ in.). High-strength low-alloy steels (classes III, IV, and V) allow the thickness and hence the weight of a machine structure subjected to high static loads to be reduced.

TABLE 15 WELDABILITY CLASSES OF STEELS USED IN MACHINERY AND EQUIPMENT

STEEL	**TYPICAL MINIMUM YIELD**	**CARBON EQUIVALENT (C_{EQ})**	**SPECIFICATION**

	STRENGTH		(MAX)(A)			
	MPA	KSI		AISI-SAE	ASTM	API
STRUCTURAL CARBON						
CLASS I	200-400	30-60	0.38	1005, 1006, 1008, 1010, 1012, 1015, 1016, 1017, 1018, 1020, 1021	. . .	. . .
CLASS II	240-380	35-55	0.48	. . .	A 36, A 38 (GRADE Y35), A 53 (GRADE B), A 106, A 131, A 139, A 374, A 500, A 501, A 516, A 524, A 529, A 570 (GRADES D, E), A 573	52 (GRADE B)
HSLA						
CLASS III	275-380	40-55	0.63	. . .	A 242, A 441, A 537 (GRADES A, E), A 572 (GRADES 42, 45, 50), A 588, A 618	5LX (GRADE 42)
CLASS IV	410-450	60-65	NOT SPECIFIED	. . .	A 572 (GRADES 60, 65)	. . .
QUENCHED AND TEMPERED						
CLASS V	620-700	90-100	0.74	. . .	A 514, A 517	. . .

(A) CARBON EQUIVALENT IS A MEASURE OF WELDABILITY. FOR UNALLOYED (STRUCTURAL) STEELS. C_{EQ} = %C + (MN/4) + (SI/4). FOR LOW-ALLOY STEELS, C_{EQ} = %C + (MN/6) + (NI/20) + (CR/10) + (MO/40) + (V/10).

The AISI and SAE carbon-manganese steels have carbon contents of less than 0.25% and manganese contents of less than 0.90% and are supplied to meet composition specifications that do not include mechanical properties. Unless verified by mechanical tests, an ultimate tensile strength of 380 MPa (55 ksi) and a yield strength of 190 MPa (27.5 ksi) should be used for design purposes. These steels have the highest weldability rating shown in the machinery welding specifications that classify steels according to weldability.

ASTM material specifications include mechanical property requirements that demand higher strength than the AISI and SAE carbon-manganese steels, and thus offer the opportunity to reduce thickness and weight for statically loaded conditions. The weldability of these higher-strength steels is lower, and they usually require higher preheat and low-hydrogen welding processes.

Allowable stress for statically loaded steel base metals can be determined from the equation given above after selecting a factor of safety. Allowable stress levels for statically loaded weld metal are presented in Table 16.

TABLE 16 ALLOWABLE STRESSES IN WELD METAL

TYPE OF WELD	STRESS IN WELD	ALLOWABLE STRESS	REQUIRED WELD STRENGTH LEVEL(A)
GROOVE WELDS WITH COMPLETE JOINT PENETRATION	TENSION NORMAL TO EFFECTIVE AREA	SAME AS BASE METAL	MATCHING WELD METAL SHALL BE USED.
	COMPRESSION NORMAL TO	SAME AS BASE METAL	WELD METAL WITH A STRENGTH LEVEL EQUAL

	EFFECTIVE AREA		TO OR ONE CLASSIFICATION (69 MPA, OR 10 KSI) LESS THAN MATCHING WELD METAL MAY BE USED.
	TENSION OR COMPRESSION PARALLEL TO WELD AXIS	SAME AS BASE METAL	WELD METAL WITH A STRENGTH LEVEL EQUAL TO OR LESS THAN MATCHING WELD METAL MAY BE USED.
	SHEAR ON EFFECTIVE AREA	0.27 × NOMINAL TENSILE STRENGTH OF WELD METAL, EXCEPT SHEAR STRESS ON BASE METAL SHALL NOT EXCEED 0.40 × YIELD STRENGTH OF BASE METAL.	
GROOVE WELDS WITH PARTIAL JOINT PENETRATION	COMPRESSION NORMAL TO EFFECTIVE AREA: JOINT NOT DESIGNED TO BEAR	0.45 × NOMINAL TENSILE STRENGTH OF WELD METAL, EXCEPT STRESS ON BASE METAL SHALL NOT EXCEED 0.55 × YIELD STRENGTH OF BASE METAL.	WELD METAL WITH A STRENGTH LEVEL EQUAL TO OR LESS THAN MATCHING WELD METAL MAY BE USED.
	COMPRESSION NORMAL TO EFFECTIVE AREA; JOINT DESIGNED TO BEAR	SAME AS BASE METAL	
	TENSION OR COMPRESSION PARALLEL TO WELD AXIS(B)	SAME AS BASE METAL	
	SHEAR PARALLEL TO WELD AXIS	0.27 × NOMINAL TENSILE STRENGTH OF WELD METAL. EXCEPT SHEAR STRESS ON BASE METAL SHALL NOT EXCEED 0.36 × YIELD STRENGTH OF BASE METAL.	
	TENSION NORMAL TO EFFECTIVE AREA	0.27 × NOMINAL TENSILE STRENGTH OF WELD METAL, EXCEPT SHEAR STRESS ON BASE METAL SHALL NOT EXCEED 0.55 × YIELD STRENGTH OF BASE METAL.	
FILLET WELDS	SHEAR ON EFFECTIVE AREA	0.27 × NOMINAL TENSILE STRENGTH OF WELD METAL	WELD METAL WITH A STRENGTH LEVEL EQUAL TO OR LESS THAN

	TENSION OR COMPRESSION PARALLEL TO WELD AXIS(B)	SAME AS BASE METAL	MATCHING WELD METAL MAY BE USED
PLUG AND SLOT WELDS	SHEAR PARALLEL TO FAYING SURFACES (ON EFFECTIVE AREA)	0.27 × NOMINAL TENSILE STRENGTH OF WELD METAL, EXCEPT SHEAR STRESS ON BASE METAL SHALL NOT EXCEED 0.36 × YIELD STRENGTH OF BASE METAL.	WELD METAL WITH A STRENGTH LEVEL EQUAL TO OR LESS THAN MATCHING WELD METAL MAY BE USED.

(A) FOR MATCHING WELD METAL, SEE TABLE 20.

(B) FILLET WELDS AND GROOVE WELDS WITH PARTIAL JOINT PENETRATION JOINING THE COMPONENT ELEMENTS OF BUILT-UP MEMBERS, SUCH AS FLANGE-TO-WEB CONNECTIONS, MAY BE DESIGNED WITHOUT REGARD TO THE TENSILE OR COMPRESSIVE STRESS IN THESE ELEMENTS PARALLEL TO THE AXIS OF THE WELDS.

Fluctuating Loads. Fatigue is the process of progressive localized damage that may result in cracks or complete fracture after a sufficient number of working stress fluctuations. Fatigue life is the number of stress fluctuations sustained to failure.

The fatigue design of welded joints is different from the "modified Goodman diagram" method used for plain (unwelded) machine and structural members. The fatigue life of plain base metal is a function of its tensile strength. Welded joints introduce stress concentrations in a structure, which reduces fatigue life. Examples of stress concentrations include weld ripple, undercut, crowned reinforcement, lack of penetration, slag inclusions, microcracks, and craters. The fatigue life of a steel welded joint is independent of base metal tensile strength. An ASTM A 36 steel weld joint has the same fatigue strength as a joint made from ASTM A 514 steel. The fatigue life of a steel weld joint is indirectly proportional to working stress range; that is, as the magnitude of working stress range increases, fatigue life decreases (see Table 17). The fatigue life of a steel weld joint is a function of the weld joint category, which in turn is a function of joint geometry, type of weld, direction of load application, and weld quality assurance (see Table 18 and Fig. 12).

TABLE 17 ALLOWABLE STRESS RANGES

STRESS CATEGORY (SEE TABLE 18)	NOMINAL NUMBER OF LOADING CYCLES							
	20,000-100,000		100,000-500,000		500,000-2,000,000		>2,000,000	
	MPA	KSI	MPA	KSI	MPA	KSI	MPA	KSI
A	434	63	255	37	165	24	165	24
B	338	49	200	29	124	18	110	16
B'	269	39	159	23	103	15	83	12
C	241	35	145	21	90	13	69(A)	10(A)
D	193	28	110	16	69	10	48	7
E	152	22	90	13	55	8	31	4.5
E'	110	16	63	9.2	40	5.8	18	2.6
F	103	15	83	12	62	9	55	8

(A) FLEXURAL STRESS RANGE OF 83 MPA (12 KSI) PERMITTED AT TOE OF STIFFENER WELDS OR FLANGES

TABLE 18 STRESS CATEGORY CLASSIFICATIONS

GENERAL CONDITION	SITUATION	TYPES OF STRESS[A]	STRESS CATEGORY (SEE TABLE 17)	ILLUSTRATIVE EXAMPLE NOS. (SEE FIG. 12)[B]
PLAIN MATERIAL	BASE METAL WITH ROLLED OR CLEANED SURFACE. FLAME-CUT EDGES WITH ANSI SMOOTHNESS OF 1000 OR LESS	T OR REV	A	1, 2
BUILT-UP MEMBERS	BASE METAL IN MEMBERS WITHOUT ATTACHMENTS, BUILT-UP PLATES, OR SHAPES CONNECTED BY CONTINUOUS FULL-PENETRATION GROOVE WELDS OR BY CONTINUOUS FILLET WELDS PARALLEL TO THE DIRECTION OF APPLIED STRESS	T OR REV	B	3-6
	BASE METAL IN MEMBERS WITHOUT ATTACHMENTS, BUILD-UP PLATES, OR SHAPES CONNECTED BY FULL-PENETRATION GROOVE WELDS WITH BACKING BARS NOT REMOVED, OR BY PARTIAL-PENETRATION GROOVE WELDS PARALLEL TO THE DIRECTION OF APPLIED STRESS	T OR REV	B'	3-6
	BASE METAL AT TOE WELDS ON GIRDER WEBS OR FLANGES ADJACENT TO WELDED TRANSVERSE STIFFENERS	T OR REV	C	7
	BASE METAL AT ENDS OF PARTIAL-LENGTH WELDED COVER PLATES NARROWER THAN THE FLANGE, HAVING SQUARE OR TAPERED ENDS, WITH OR WITHOUT WELDS ACROSS THE ENDS OR WIDER THAN FLANGE WITH WELDS ACROSS THE ENDS			
	FLANGE THICKNESS ≤20 MM (0.8 IN.)	T OR REV	E	5
	FLANGE THICKNESS >20 MM (0.8 IN.)	T OR REV	E'	5
	BASE METAL AT END OF PARTIAL-LENGTH WELDED COVER PLATES WIDER THAN THE FLANGE WITHOUT WELDS ACROSS THE ENDS		E'	5
GROOVE WELDS	BASE METAL AND WELD METAL AT FULL-PENETRATION GROOVE WELDED SPLICES OF PARTS OF SIMILAR CROSS SECTION GROUND FLUSH, WITH GRINDING IN THE DIRECTION OF APPLIED STRESS AND WITH WELD SOUNDNESS ESTABLISHED BY RADIOGRAPHIC OR ULTRASONIC INSPECTION IN	T OR REV	B	10, 11

	ACCORDANCE WITH THE REQUIREMENTS OF 9.25.2 OR 9.25.3 OF AWS D1. I			
	BASE METAL AND WELD METAL AT FULL-PENETRATION GROOVE WELDED SPLICES AT TRANSITIONS IN WIDTH OR THICKNESS; WITH WELDS GROUND TO PROVIDE SLOPES NO STEEPER THAN 1 TO $2\frac{1}{2}$, WITH GRINDING IN THE DIRECTION OF APPLIED STRESS AND WITH WELD SOUNDNESS ESTABLISHED BY RADIOGRAPHIC OR ULTRASONIC INSPECTION IN ACCORDANCE WITH THE REQUIREMENTS OF 9.25.2 OR 9.25.3 OF AWS D1. 1			
	A 514 BASE METAL	T OR REV	B'	12, 13
	OTHER BASE METALS	T OR REV	B	12, 13
	BASE METAL AND WELD METAL AT FULL-PENETRATION GROOVE WELDED SPLICES, WITH OR WITHOUT TRANSITIONS HAVING SLOPES NO GREATER THAN 1 TO $2\frac{1}{2}$ WHEN REINFORCEMENT IS NOT REMOVED BUT WELD SOUNDNESS IS ESTABLISHED BY RADIOGRAPHIC OR ULTRASONIC INSPECTION IN ACCORDANCE WITH THE REQUIREMENTS OF 9.25.2 OR 9.25.3 OF AWS D1.1	T OR REV	C	10-13
PARTIAL-PENETRATION GROOVE WELDS	WELD METAL OF PARTIAL-PENETRATION TRANSVERSE GROOVE WELDS, BASED ON EFFECTIVE THROAT AREA OF THE WELD OR WELDS	T OR REV	$F^{(C)}$	16
FILLET-WELDED CONNECTIONS	BASE METAL AT INTERMITTENT FILLET WELDS	T OR REV	E	. . .
	BASE METAL AT JUNCTION OF AXIALLY LOADED MEMBERS WITH FILLET-WELDED END CONNECTIONS. WELDS SHALL BE DISPOSED ABOUT THE AXIS OF THE MEMBER SO AS TO BALANCE WELD STRESSES.			
	$B \leq 25$ MM (1 IN.)	T OR REV	E	17, 18
	$B > 25$ MM (1 IN.)	T OR REV	E'	17, 18
	BASE METAL AT MEMBERS			

	CONNECTED WITH TRANSVERSE FILLET WELDS			
	B ≤13 MM (0.5 IN.)	T OR REV	C(C)	20, 21
	B > 13 MM (0.5 IN.)			
FILLET WELDS	WELD METAL OF CONTINUOUS OR INTERMITTENT LONGITUDINAL OR TRANSVERSE FILLET WELDS	S	F(C)	15, 17, 18, 20, 21
PLUG OR SLOT WELDS	BASE METAL AT PLUG OR SLOT WELDS	T OR REV	E	27
	SHEAR ON PLUG OR SLOT WELDS	S	F	27
MECHANICALLY FASTENED CONNECTIONS	BASE METAL AT GROSS SECTION OF HIGH-STRENGTH BOLTED SLIP-CRITICAL CONNECTIONS, EXCEPT AXIALLY LOADED JOINTS THAT INDUCE OUT-OF-PLANE BENDING IN CONNECTED MATERIAL	T OR REV	B	8
	BASE METAL AT NET SECTION OF OTHER MECHANICALLY FASTENED JOINTS	T OR REV	D	8, 9
	BASE METAL AT NET SECTION OF FULLY TENSIONED HIGH-STRENGTH, BOLTED-BEARING CONNECTIONS	T OR REV	B	8, 9
ATTACHMENTS	BASE METAL AT DETAILS ATTACHED BY FULL-PENETRATION GROOVE WELDS SUBJECT TO LONGITUDINAL AND/OR TRANSVERSE LOADING WHEN THE DETAIL EMBODIES A TRANSITION RADIUS, *R*, WITH THE WELD TERMINATION GROUND SMOOTH AND FOR TRANSVERSE LOADING, THE WELD SOUNDNESS ESTABLISHED BY RADIOGRAPHIC OR ULTRASONIC INSPECTION IN ACCORDANCE WITH 9.25.2 OR 9.25.3 OF AWS D1.1			
	LONGITUDINAL LOADING			
	R > 600 MM (24 IN.)	T OR REV	B	14
	600 MM (24 IN.) > *R* > 150 MM (6 IN.)	T OR REV	C	14
	150 MM (6 IN.) > *R* > 50 MM (2 IN.)	T OR REV	D	14
	50 MM (2 IN.) > *R*	T OR REV	E	14
	DETAIL BASE METAL FOR TRANSVERSE LOADING: EQUAL THICKNESS AND REINFORCEMENT REMOVED			
	R > 600 MM (24 IN.)	T OR REV	B	14

	600 MM (24 IN.) > *R* > 150 MM (6 IN.)	T OR REV	C	14
	150 MM (6 IN.) > *R* > 50 MM (2 IN.)	T OR REV	D	14
	50 MM (2 IN.) > *R*	T OR REV	E	14, 15
	DETAIL BASE METAL FOR TRANSVERSE LOADING; EQUAL THICKNESS AND REINFORCEMENT NOT REMOVED			
	R > 600 MM (24 IN.)	T OR REV	C	14
	600 MM (24 IN.) > *R* > 150 MM (6 IN.)	T OR REV	C	14
	150 MM (6 IN.) > *R* > 50 MM (2 IN.)	T OR REV	D	14
	50 MM (2 IN.) > *R*	T OR REV	E	14, 15
	DETAIL BASE METAL FOR TRANSVERSE LOADING: UNEQUAL THICKNESS AND REINFORCEMENT REMOVED			
	R > 50 MM (2 IN.)	T OR REV	D	14
	50 MM (2 IN.) > *R*		E	14, 15
	DETAIL BASE METAL FOR TRANSVERSE LOADING; UNEQUAL THICKNESS AND REINFORCEMENT NOT REMOVED			
	ALL *R*	T OR REV	E	14, 15
	DETAIL BASE METAL FOR TRANSVERSE LOADING			
	R > 150 MM (6 IN.)	T OR REV	C	19
	150 MM (6 IN.) > *R* > 50 MM (2 IN.)	T OR REV	D	19
	50 MM (2 IN.) > *R*	T OR REV	E	19
	BASE METAL AT DETAIL ATTACHED BY FULL-PENETRATION GROOVE WELDS SUBJECT TO LONGITUDINAL LOADING			
	50 MM (2 IN.) < *A* < 12*B* OR 100 MM (4 IN.)	T OR REV	D	15
	A > 12*B* OR 100 MM (4 IN.) WHEN *B* ≤25 MM (1 IN.)	T OR REV	E	15
	A > 12*B* OR 100 MM (4 IN.) WHEN *B* > 25 MM (1 IN.)	T OR REV	E'	15
	BASE METAL AT DETAIL ATTACHED BY FILLET WELDS OR PARTIAL-PENETRATION GROOVE WELDS			

	SUBJECT TO LONGITUDINAL LOADING			
	A < 50 MM (2 IN.)	T OR REV	C	15, 23-26
	50 MM (2 IN.) < *A* < 12*B* OR 100 MM (4 IN.)	T OR REV	D	15, 23, 24, 26
	A > 12*B* OR 100 MM (4 IN.) WHEN *B* ≤25 MM (1 IN.)	T OR REV	E	15, 23, 24, 26
	A > 12*B* OR 100 MM (4 IN.), WHEN *B* > 25 MM (1 IN.)	T OR REV	E'	15, 23, 24, 26
	BASE METAL ATTACHED BY FILLET WELDS OR PARTIAL-PENETRATION GROOVE WELDS SUBJECTED TO LONGITUDINAL LOADING WHEN THE WELD TERMINATION EMBODIES A TRANSITION RADIUS WITH THE WELD TERMINATION GROUND SMOOTH			
	R > 50 MM (2 IN.)	T OR REV	D	19
	R ≤50 MM (2 IN.)	T OR REV	E	19
	FILLET-WELDED ATTACHMENTS WHERE THE WELD TERMINATION EMBODIES A TRANSITION RADIUS, WELD TERMINATION GROUND SMOOTH, AND MAIN MATERIAL SUBJECT TO LONGITUDINAL LOADING			
	DETAIL BASE METAL FOR TRANSVERSE LOADING:			
	R > 50 MM (2 IN.)	T OR REV	D	19
	R < 50 MM (2 IN.)	T OR REV	E	19
	BASE METAL AT STUD-TYPE SHEAR CONNECTOR ATTACHED BY FILLET WELD OR AUTOMATIC END WELD	T OR REV	C	22
	SHEAR STRESS ON NOMINAL AREA OF STUD-TYPE SHEAR CONNECTORS	S	F	. . .

(A) T, RANGE IN TENSILE STRESS ONLY; REV, RANGE INVOLVING REVERSAL OF TENSILE OR COMPRESSIVE STRESS; S, RANGE IN SHEAR, INCLUDING SHEAR-STRESS REVERSAL.

(B) THESE EXAMPLES ARE PROVIDED AS GUIDELINES AND ARE NOT INTENDED TO EXCLUDE OTHER REASONABLY SIMILAR SITUATIONS.

(C) ALLOWABLE FATIGUE STRESS RANGE FOR TRANSVERSE PARTIAL-PENETRATION AND TRANSVERSE FILLET WELDS IS A FUNCTION OF THE EFFECTIVE THROAT, DEPTH OF PENETRATION, AND PLATE THICKNESS.

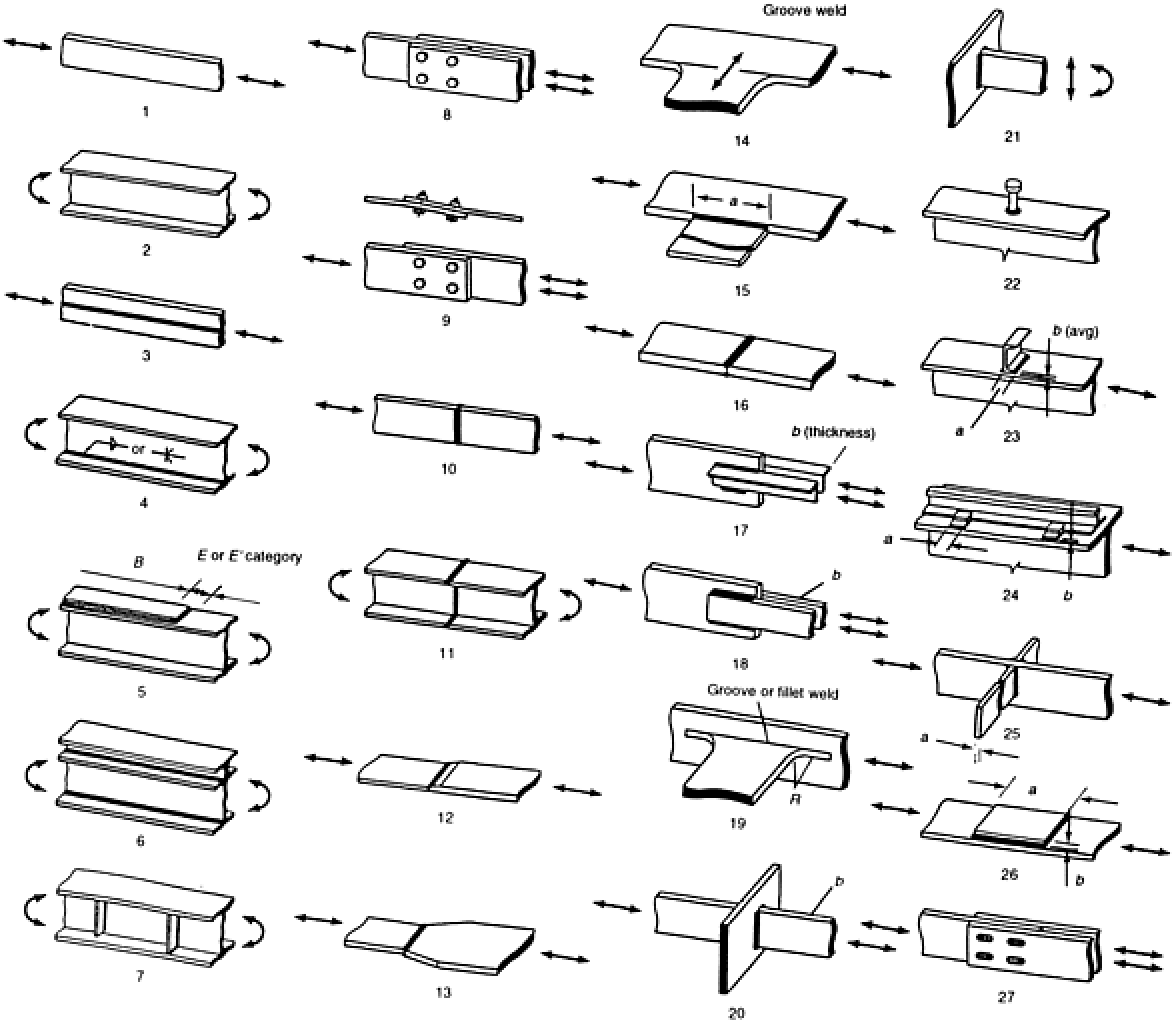

FIG. 12 ILLUSTRATIVE EXAMPLES OF THE STRESS CATEGORIES IN TABLE 18

The primary design requirement for some machine members can be rigidity. Such members, even when subjected to both static and fluctuating loads, are treated as statically loaded structures. Deflection under load is very small and the working stress is only 14 to 28 MPa (2 to 4 ksi). In this case, the weld need not be sized to carry the full load capacity of the member.

Loads are considered cyclic if the stress range exceeds 10% of the allowable stress range or if the total life cycles exceed 20,000. Care must be taken when selecting higher-strength materials, because the allowable stress range of welded joints is independent of material strength. The maximum allowable stress is higher for higher-strength steels, but the allowable stress range is the same. The use of higher-strength steels in cyclic load applications is beneficial only when the part is not welded or when the part is in a tensile prestress condition.

The materials listed in Table 15 are commonly used steels that are considered prequalified. Other steels, such as AISI 4340 and 4140, are also commonly used in machinery components, but are not prequalified. Weld procedure qualifications must be developed as required by the appropriate machinery weld specifications. Preheats for these high-hardenable steels must be above the martensitic start (M_S) temperature of 290 °C (550 °F) for AISI 4340 and 4140 to allow the formation of the more ductile bainitic structure and thus avoid brittleness and HAZ cracks. A low-hydrogen welding process is mandatory. The postheat temperature must be no higher than 28 °C (50 °F) below the hardening temper temperature incurred when the part was prehardened.

A table of welding practices applicable to most steels for machinery and equipment is available in Ref 18. Related AWS specifications are listed in Table 19.

TABLE 19 AWS SPECIFICATIONS RELATING TO MACHINERY AND EQUIPMENT

AWS NO.	TITLE
D14.1	"WELDING INDUSTRIAL AND MILL CRANES AND OTHER MATERIAL HANDLING EQUIPMENT"
D14.2	"METAL CUTTING MACHINE TOOL WELDMENTS"
D14.3	"WELDING EARTHMOVING AND CONSTRUCTION EQUIPMENT"
D14.4	"CLASSIFICATION AND APPLICATION OF WELDED JOINTS FOR MACHINERY AND EQUIPMENT"
D14.5	"WELDING OF PRESSES AND PRESS COMPONENTS"
D14.6	"WELDING ROTATING ELEMENTS OF EQUIPMENT"

Service Life. Material selection for machine tools, presses, and heavy rolling and forging equipment is usually governed by an indefinite life expectancy. The working parts (e.g., tools, hammers, forming or punching dies, and mill rolls) and the moving parts (e.g., motors, gears, and clutches) are subject to wear and may need to be either replaced or rebuilt. The basic equipment, however, should be constructed from materials that will last until the machine is retired due to obsolescence.

Obsolescence is also a design consideration in material selection for mobile equipment. Because of weight restrictions, however, material election must make use of the maximum load-carrying capacity of the principal components. The life of wear surfaces, especially in earth-moving equipment, is often enhanced by careful selection of weld surfacing materials.

Maintenance and repair of components must also factor into the material selection process. If it is known that parts will need replacement, access for this type of maintenance should be part of the design. If welds must be removed to provide access, necessitating rewelding, the weldability of the steel used for that part is important. A design that seems perfectly feasible during the initial fabrication of a component may present difficult, or even insoluble, problems during disassembly and reassembly for replacement or repair of a worn part.

Safety. Construction equipment is used in a variety of applications, some of which may endanger the operator or other workers at the site. The designer is responsible for ensuring that the material selected is appropriate for the intended use, making allowances for improper uses that cause stress higher than the rated load stress. Material selection should therefore be based on conservative estimates for components whose failure might jeopardize operators and others.

Economics. Machinery and equipment involve the same economic principles with regard to material choice as other types of welded structures. In most applications the unit cost of the material is a small percentage of the overall cost of the fabricated component or machine. Too often the driving force for low initial cost fails to consider the life-cycle cost of a machine. The cost of maintenance, repair, or replacement of components subject to failure or wear can in many cases justify the selection of more costly materials with longer life cycles. Pumps that handle corrosive liquids and dredges that handle erosive slurries are useful examples. Impellers and casings made of materials resistant to corrosion and erosion can add enormously to the utility of a pump over its anticipated life in comparison with materials that may require frequent replacement or rebuilding. In some cases, maintaining the efficiency of a machine component during its life by selecting materials with greater resistance to the environment will offset its cost. Life-cycle cost assessments should be applied to the choice of materials for machinery and equipment during the design phase.

Reference cited in this section

18. R.D. STOUT, *WELDABILITY OF STEELS,* 4TH ED., WELDING RESEARCH COUNCIL, 1987, P 338-437

Material Properties of Weldments

Assessment of the performance of welded components must include the ability of the various metallurgical structures to perform the tasks specified by the designer. These structures are broadly defined as the fusion zone and the heat-affected zone. The thermal excursions that occur during welding invariably result in residual stresses, which in some instances may add considerably to the service stresses encountered by a weldment. The effect of welds on the service-ability of weldments for machinery and equipment may influence material selection.

Heat-Affected Zone. In unalloyed steels, the properties of a region that has experienced one or more thermal excursions during welding are quite similar to those of the unaffected base metal. In low-alloy steels, especially those in weldability classes IV and V (Table 15), the HAZ becomes considerably stronger, harder, and less ductile. This impairs the ability of the metal to adjust to the strains imposed by static loads; moreover, the fatigue properties are degraded, lowering resistance to cyclic loads. When low-alloy steels are used for critical weldments in machinery, postweld heat treatments are employed to overcome most of these adverse effects. Preheat is necessary to permit these hardenable steels to overcome the effect of hydrogen, which can cause cracks in the HAZ during arc welding operations.

Fusion Zone. In an arc weld, the fusion zone comprises the amount of base metal that has been melted in addition to the filler metal that is fused by the arc. This is a cast structure distinctly different from the base material--which in wrought materials is usually a polycrystalline structure of more or less uniform size. The welding filler metals normally used for unalloyed carbon steels produce welds whose yield strength is generally higher in the as-deposited condition than that of the base metal. In most machinery applications, this higher strength is considered advantageous. Failures in static loading will therefore depend on the strength of the base metal, which the designer has chosen based on its known properties. Weld failures in statically loaded full-strength welds are thus unlikely, assuming that the welds meet specified workmanship standards.

Low-alloy steel welds of matching weld-metal composition are considerably stronger than the base metal. In these applications, it is often preferable to choose a filler metal of lower alloy content (especially carbon) to match, or even undermatch, the strength of the base metal. As long as the yield strength of the weld metal is not less than that of the base metal, static loads will not cause significant plastic deformation in the welds. Because filler metals are traditionally classified on the basis of the ultimate tensile strength of the welds that they produce, it is possible to use filler metal that is one classification lower in ultimate strength than the base metal. Generally, the yield strength of the lower-strength weld metal will more nearly match that of the base metal. Table 20 lists matching filler metal specifications.

TABLE 20 MATCHING FILLER METAL SPECIFICATIONS

BASE METAL(B)	**WELDING PROCESS**(A)			
	SMAW	**SAW**	**GMAW**	**FCAW**
CLASSES I AND II	AWS A5.1 OR A5.5 E60*XX* OR E70*XX*	AWS A5.17 F6*X* OR F7*X*-E*XXXX*	AWS A5.18 E70S-*X*	AWS A5.20 E60T-*X* OR E70T-1 (EXCEPT E*XX*T-2 AND E*XX*T-3)(C)
CLASS III	AWS A5.1 OR A5.5 E70*XX*(D)	AWS A5.17 F7*X*-E*XXXX*	AWS A5.18 E70S-*X*	AWS A5.20 E70T-*X* (EXCEPT E70T-2 AND E70-T3)(C)
CLASS IV(E)	AWS A5.5 E80*XX*(D)	AWS A5.23 F8*X*-E*XXX*	GRADE E80S(F)	GRADE E80T(F)
CLASS V	AWS A5.5 E110*XX*(D)	AWS A5.23 F11*X*-E*XXX*	GRADE E110S(F)	GRADE E110T(F)

Note: The use of the same type of filler metal having the next higher mechanical properties as listed in AWS specification is permitted.

(A) A FILLER METAL OF A LOWER STRENGTH MAY BE USED WHERE THE DESIGN ENGINEER HAS STIPULATED LESS THAN 100% JOINT EFFICIENCY.

(B) IN JOINTS INVOLVING BASE METALS OF DIFFERENT YIELD POINTS OR STRENGTHS, FILLER METALS APPLICABLE TO THE LOWER-STRENGTH BASE METAL MAY BE USED.

(C) IF TYPE 2 FILLER METALS ARE USED ON SUCCESSFULLY TESTED PROTOTYPE VEHICLES, THIS QUALIFIES THE USE OF TYPE 2 FILLER METALS FOR PRODUCTION.

(D) LOW-HYDROGEN CLASSIFICATION.

(E) WHEN WELDS ARE TO BE THERMALLY STRESS RELIEVED, THE DEPOSITED WELD METAL SHALL NOT EXCEED 0.05% V.

(F) FOR WELD METAL HAVING A MINIMUM SPECIFIED YIELD STRENGTH GREATER THAN 415 MPA (60 KSI), THE USER SHALL DEMONSTRATE THAT EACH COMBINATION OF ELECTRODE AND SHIELDING PROPOSED FOR USE WILL PRODUCE LOW-ALLOY WELD METAL HAVING THE MECHANICAL PROPERTIES LISTED. THE MECHANICAL PROPERTIES SHALL BE DETERMINED FROM A MULTIPLE-PASS WELD MADE IN ACCORDANCE WITH THE TEST REQUIREMENTS OF THE LATEST EDITION OF AWS A5.18 ("SPECIFICATION FOR MILD STEEL ELECTRODES FOR GAS METAL ARC WELDING") OR A5.20 ("SPECIFICATION FOR MILD STEEL ELECTRODES FOR FLUX CORED ARC WELDING"), AS APPLICABLE. WHEN AN APPLICABLE AWS FILLER METAL SPECIFICATION IS ISSUED, IT WILL TAKE PRECEDENCE OVER ALL OTHERS, AND TESTING BY THE USER WILL NOT BE REQUIRED. THE MECHANICAL PROPERTY TESTS REQUIRED FOR GRADES E110S. E100T, AND E110T SHALL BE MADE USING ASTM A 514 BASE METAL. WHEN REQUESTED, THE ELECTRODE MANUFACTURER SHALL FURNISH CERTIFICATION THAT THE ELECTRODE WILL MEET THE REQUIREMENTS OF CLASSIFICATION OR GRADE.

As-Welded Versus Postweld Heat-Treated Welds. Most welds used for machinery and equipment, including HSLA steels, are placed in service in the as-welded condition. Postweld heat treatment is usually specified to reduce residual stress and to attain dimensional stability when the weldment is to be machined after welding and the machining dimensional tolerances are small. Postweld heat treatment is sometimes specified to improve corrosion resistance in the welded area. The allowable stresses in specifications for components to be postweld heat treated are based on experience or on laboratory tests of the various types of weld joints.

High residual stresses remain in welds that have not undergone postweld heat treatment. In some cases, service loads added to the residual stresses in the weld area will cause sufficient distortion to present problems of dimensional misalignment. A stress-relief thermal treatment may be needed to overcome problems caused by residual stresses. Additional information is available in the article "Residual Stresses and Distortion" in this Volume.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Automobiles

Since its inception nearly a century ago, the automobile industry has relied heavily on joining technologies. Improvements in those technologies over the years have enabled improvements in automotive design--a constantly changing field driven by the ever-increasing demands of customers and governmental agencies. Customers want vehicles to meet criteria ranging from performance and function to appearance and styling; governmental agencies dictate that vehicles meet stringent safety and environmental standards. Producing automotive designs under such demands and constraints can often create difficulties in joining operations.

Service Conditions. The automotive design engineer must consider numerous types of service conditions. Structurally, a vehicle must be capable of withstanding on- and/or offroad service, must be crashworthy in order to meet Federal Motor Vehicle Safety Standards (FMVSS), and yet must possess an economy in terms of overall weight to meet Corporate Average Fuel Economy (CAFE) regulations. A vehicle must be able to withstand diverse environmental conditions, ranging from the external environment (weather, electrolyte exposure, acid rain, etc.) to the internal environment (engine heat, fuels, lubricants, sealers, etc.), and must be watertight and airtight as well. Finally, all of this must be accomplished within the boundaries of cost-effective and manufacturable processes, including joining.

Welding Processes. The automobile industry utilizes a wide array of welding processes, materials, coatings, and designs. Given the many possible combinations of these factors, difficulties in the manufacturing process are inevitable. However, many problems can be avoided by exercising care in the initial design phase. By taking into account structural requirements (strength, crashworthiness, and fatigue resistance) as well as appearance requirements (corrosion resistance and finish), the design engineer can determine the most effective combination of materials and joining processes to fit the intended design and maximize performance.

Joining processes commonly used in the automotive industry include, but are not limited to, the following:

- RESISTANCE SPOT WELDING
- GAS-METAL ARC WELDING
- GAS-TUNGSTEN ARC WELDING
- FLUX-CORED ARC WELDING
- RESISTANCE PROJECTION WELDING
- STUD ARC WELDING
- LASER BEAM WELDING
- RESISTANCE SEAM WELDING
- BRAZING
- SOLDERING
- PLASTIC WELDING

These processes are applied and controlled either by the automotive industry itself or by suppliers within the industry. In either case, each joining technique should be governed by a reliable set of standards to ensure that the technique is properly defined and applied. The sections that follow will describe how the more commonly used joining techniques listed above are affected by the various material and mechanical factors encountered in the automotive industry.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Resistance Spot Welding

Resistance spot welding is the primary joining process used in automobile production. Its widespread use can be attributed to its many advantages. Specifically, the method is easy, cost effective, and, when properly applied, can produce structures with greater integrity than the sum of the individual welds. Therefore, RSW lends itself to the extremely high production volumes common in the automotive industry. Despite the obvious advantages of the spot welding process, adequate results cannot be achieved without special consideration of a number of factors that can ultimately affect the welded structure, including raw material coatings, processing fluids, raw material chemistry, joint design, joint sealers, and joint adhesives.

Steel Composition. Raw material chemistry is of primary importance in the design of a spot-welded structure. Most automotive applications use SAE 1008-1010 steel. This material has a low carbon equivalent, which makes it readily formable and weldable and relatively inexpensive and abundantly produced. However, low-carbon material is not used exclusively. Automobile manufacturers must reduce weight while retaining structural integrity. This has resulted in increasing use of high-strength steels, such as bake-hardenable, dual-phase, precipitation-hardenable, and martensitic steels. These materials are inherently more difficult to weld because of the residual elements added for solid-solution strengthening, but weldable grades are available. As a rule, welding of these materials requires increased heat input compared with low-carbon steel material.

Steel Thickness. In the design of a spot-welded joint, a critical factor is the combination of materials to be joined, along with their thicknesses. As the number of sheets to be welded increases, the difficulty of the welding operation also increases. Similarly, welding becomes more difficult as the ratio of material thicknesses between sheets becomes overly large. Flange size is another important factor. Flanges must be large enough to accommodate weld equipment constraints and required weld sizes. Too often, designers fail to consider these important yet simple manufacturing requirements.

Coated Steels. Due to consumer demand, the automobile industry has been striving in recent years to produce vehicles with greater corrosion resistance. For the most part, this demand has been met by the use of steels coated by galvanizing, aluminizing, prepainting, or laminating. Such coatings greatly enhance the corrosion resistance of the base material, but reduce its weldability. In general, spot welding of coated materials generally requires an increase in weld current and hold time. Greater electrode wear is also exhibited, which can ultimately affect the resultant welds.

Joint Sealers. Passenger comfort is always an overriding concern in the design of a vehicle; therefore, much attention is paid to the reduction of wind noise, fume intrusion, and water leakage. Body sealers, particularly weld-through sealers, are used to control these factors. Welding problems associated with the use of the sealers include reduced conductivity at the faying surface and increased fuming as the sealer is burned away at the weld. Solutions to these problems include increasing heat input to ensure proper welding and providing for adequate ventilation to remove fumes generated (although minimal) by the vaporized sealer.

Joint adhesives are used to enhance joint integrity. They are commonly used in conjunction with the spot welding process to improve joint shear strength. Similar to joint sealers, adhesives can cause problems in welding, which can be overcome by the methods described above.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Gas Metal Arc Welding

The GMAW process is widely used in the automotive industry to weld frames, brackets, body panels, and many other structural components. This method is easily applied (automated and manually) and lends itself well to high productivity at a relatively low cost; in addition, it allows most metals to be joined to themselves. Disadvantages include the fact that GMAW is sensitive to material chemistry, surface coatings, and joint design. When these factors are not considered in the initial design, weld quality problems, such as cracks, porosity, lack of fusion, and burnthrough, often result, which may affect the structural integrity of an assembly.

Steel Properties. In the automotive industry, a wide variety of steel chemistries are used to meet strength and durability requirements, ranging from low-strength plain carbon steels to high-strength alloys. To avoid GMAW problems associated with chemistry, the designer must take into account the compatibility of the materials to be joined as well as their individual chemistries. Both plain carbon and alloy steels with carbon equivalents of approximately 0.30 are generally considered to be weldable. Once the material chemistries have been established, the appropriate solid filler material must be chosen to produce a weld with characteristics similar to those of the base materials being joined. By taking these initial design steps, problems such as underbead cracking and weld-metal embrittlement can be avoided.

Surface Condition. The GMAW process is sensitive to surface impurities on the base material, such as corrosion, dirt, and oil. Surface coatings can also adversely affect the resultant weld. When improperly prepared surfaces are welded by this method, quality problems (typically porosity and cracking) are often observed. The solution involves simply removing the surface contaminants. Surface coatings are not dealt with as easily, for either the coating must be selectively removed from the weld joint or the welding parameters must be adjusted to reduce the effects of the off-gassing material. This is sometimes done through a trial-and-error process, but can be refined to produce sound welds. In addition, if the coatings are removed in the weld area, provisions must be made to reapply corrosion protection where necessary.

Joint Preparation. Joint designs for GMAW are quite important in the automotive industry. Joints must be designed in such a way as to minimize the amount of joint preparation necessary, yet produce a weld that will serve the intended purpose. Joint preparation is a time-consuming process, which is not desirable for high-volume vehicle production.

Therefore, naturally occurring joints that result from strategic part placement are preferred to prepared joints, such as groove welds, which often require extra processing steps.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Gas Tungsten Arc Welding and Flux-Cored Arc Welding

The GTAW process is well suited to handle thin-section welding with or without filler materials. The FCAW process can handle the same types of welding conditions as the GMAW process, despite the filler material differences (flux-filled wires versus solid wires, respectively). In terms of automotive applications, both processes involve considerations similar to those listed for the GMAW process.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Fasteners: Resistance Projection Welding and Stud Welding

The intricate assembly of an automobile requires that parts be securely fastened, yet the ease of manufacturing must not be compromised. In order to accomplish this task, projection-welded screws, nuts, and studs are often used. These fasteners must possess welds that can withstand the rigors of a production environment as well as endure the torquing operations in assembly. Types of service conditions that can affect the welding and subsequent use of these parts include:

- BASE METAL/FASTENER CHEMISTRY
- BASE METAL/FASTENER STRENGTH CHARACTERISTICS
- BASE METAL/FASTENER COATINGS
- BASE METAL CLEANLINESS PRIOR TO WELDING

Materials Properties. Selection of the correct base metal and fastener chemistries is extremely important to ensure compatibility at the weld interface and HAZs. In order to avoid exotic weld schedules, materials with carbon equivalents of approximately 0.30 are ordinarily used, as they are readily weldable. It should be noted that these types of welds result in a similar type of weld metal/HAZ transition as other joining processes; this fact must be recognized in the design of such weldments.

Base-metal strength and fastener strength characteristics can affect the ultimate mechanical joint integrity. Designs must reflect conditions where the base-metal strength is compatible with that of the fastener; otherwise, premature failures or insufficient mechanical joints may result. The different strength characteristics may be a result of chemistry, heat treatment, cold working, or simply section thickness.

Surface Condition. Projection-welded fasteners may also be affected by coatings present on the base material or the fastener. As described previously, the automotive industry utilizes different corrosion-resistant coatings on many areas of vehicles. Such coatings can affect projection or stud welds in several ways. If the base material is zinc coated or the fastener is coated with a corrosion-resistant plating, the arc characteristics of the welding process are much different than for bare materials. In either case, provisions must be made in welding schedules. Lastly, use of an antispatter/antifriction coating, such as Teflon, can also affect the welding process. Such coatings are being used increasingly to prevent production problems associated with weld spatter and expulsions generated by other processes.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Laser Beam Welding

Laser beam welding is an emerging technology that has shown promise for use in the automotive industry. Basically, the technique utilizes a high-power laser beam capable of penetrating deep into the weld joint. Laser welding offers high speed, precision, and flexibility to automotive applications. The resultant welds exhibit very small HAZs and are easily formed to final part configuration without weld degradation.

The diverse capabilities of the laser welding technique enable the user to produce parts with properties unattainable with common methods, such as resistance spot welding or GMAW. Laser welding is most often used to make lap and butt joints. It allows blanks to be fabricated through the welding of different-gage materials as well as materials with dissimilar coatings. After the welding operation, the parts can be formed to their final shape. This allows heavier-gage materials to be used in place of reinforced areas, thereby eliminating parts. Similarly, coated parts can be strategically placed in corrosion-prone areas, which eliminates the need to coat the entire assembly.

Despite the obvious advantages of laser welding, application of the technique requires care, especially with the types of materials used in automotive applications. Coated steels complicate the laser welding process, because the off-gassing of the lower-boiling-point material interferes with the maintenance of the weld keyhole. Once this keyhole is broken, the weld must be restarted, resulting in inferior quality. Joint edges and part alignment/fit-up are also common production problems when laser welding. They affect the continuity of the welding process by disrupting the laser beam or by deflecting the beam from the joint.

Other Joining Processes

Many other joining techniques are commonly used in the automotive industry, but to a lesser extent than those discussed above. Some of these processes include resistance seam welding, soldering, brazing, and plastic welding, each of which requires due care in its application to ensure the production of sound joints.

Railroad Equipment

The transporting of freight and passengers by railroads in the United States involves over 200,000 miles of track, 1,200,000 freight cars, 19,000 locomotives, and thousands of passenger vehicles (Ref 19). Welding is the major process used in the fabrication of railroad facilities and structures, as well as in their repair and maintenance. Practically all known welding processes are being used or have been applied to components used by the industry. This section will deal with freight cars, locomotives, and track.

Reference cited in this section

19. *RAILROAD FACTS,* ASSOCIATION OF AMERICAN RAILROADS, 1990

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Freight Cars

Freight car requirements vary considerably, depending on the type of load being carried. A variety of products are transported by rail, including various liquid and solid commodities. Some commodities must be kept at low temperatures and some at high pressures, and some are classified as hazardous. Public safety is always a key issue.

Many freight cars are in "interchange service," which means that they may be handled by any railroad. It also means that repair work may be done by a variety of shops throughout North America. Ease of repair welding is therefore a critical consideration in material selection.

It is not unusual to expect a freight car to last 40 years. Cars for interchange are designed for required service conditions, which for high-utilization cars may be up to 3 million miles of service. Freight cars must be able to sustain 4.5 MN (10^6 lbf) of squeeze and single-end impacts of 5.6 MN (1.25×10^6 lbf). Designs must include buff and draft forces of up to 3.5 MN (630,000 lbf) without exceeding the yield strength of the materials.

Freight cars are subject to a variety of cyclic stresses due to train action, track irregularities, and track hunting. Most failures are caused by fatigue crack propagation, with a lesser number caused by a single overspeed impact.

Types of freight cars in common use include tank, flat, gondola, box, hopper, and covered hopper cars. Of these, only tank and covered hopper cars have seen a net increase in number over the last 21 years (Ref 20). The largest increase has been in the use of covered hopper cars, which handle a variety of dry commodities that must be kept clean.

Tank cars consist of a tank, which is basically a pressure vessel, and an underframe, which provides connection to other cars and to the "trucks." Underframes may be continuous or may consist of attachments to the ends of the tank, with the tank being a structural member. Tank cars and covered hopper cars are often lined with epoxy, rubber, or other protective materials to prevent car body corrosion and/or contamination of the commodity being carried.

Specifications. Freight car construction is regulated by the Association of American Railroads (AAR). Tank car welding is covered by the AAR "Tank Car Specification." For other types of freight car welding, the AAR references the AWS "Railroad Welding Specification" (D15.1).

Design Considerations and Materials of Construction. With the type of welded construction employed on freight cars, the use of higher-strength materials does not improve fatigue resistance. Where fatigue is a significant factor, structural designs incorporating welding are based on carbon steel and HSLA steels with moderate strengths (usually 345 MPa, or 50 ksi, yield strength or less). For applications requiring high static strength and moderate fatigue resistance, quenched and tempered steels with yield strengths of 690 MPa (100 ksi) and high-strength controlled rolled steels with yield strengths of 550 MPa (80 ksi) have been used.

The most common steels used in the production of modern tank cars are ASTM A 516 grade 70 and AAR TC128 grade B (similar to ASTM A 612 grade B). For all pressure tank cars and for some nonpressure cars, these steels are used in the normalized condition. For general-purpose nonpressure cars, hot-rolled steel is used.

Certain components are exposed to severe wear conditions, including "truck" components as well as couplers and related wear plates. Wear materials used include plain high-carbon steel, quenched and tempered alloy steel, and austenitic manganese steel, which are attached by welding or by mechanical fasteners.

Some commodities warrant the use of stainless steel tanks or covered hopper cars, which are generally fabricated of AISI type 304 or 316 stainless steel. However, this constitutes a small percentage of the freight car business.

The empty car weight, or tare weight, is an important factor in determining the hauling capacity of a car. The tare weight is deducted from the total allowed weight (i.e., gross rail load) to determine the weight of commodity that can be loaded (i.e., load capacity). The energy used to haul the tare weight around is wasted fuel. For this reason, the use of aluminum car bodies has recently become more popular, especially in coal-hauling service. The corrosion resistance of aluminum alloys is an added benefit. In some designs, the aluminum is mechanically fastened to the welded steel underframe, and little or no welding of the aluminum is required.

Freight cars generally consist of six components: tracks, brakes, underframes, draft components, safety appliances, and bodies. Figure 13 shows a side view of a general-purpose boxcar. A high percentage of the welding operation is performed in the underframes and the car bodies. Underframe components typically are from 6.4 to 25 mm ($\frac{1}{4}$ to 1 in.) thick, but may be thicker for heavy-duty cars. Bodies vary from 3.2 to 7.8 mm ($\frac{1}{8}$ to $\frac{5}{16}$ in.) in thickness. Tank car tanks range from 11 mm ($\frac{7}{16}$ in.) to slightly more than 25 mm (1 in.) in thickness.

Dimensions

Nominal capacity: 71,000 kg (70 tons)	Inside width: 2.9 m (9.5 ft)
Cubic capacity: 145 m^3 (5090 ft^3)	Side door opening: 3 m (10 ft)
Estimated light weight: 27,000 kg (59,600 lb)	Clear height door opening: 3 m (9.9 ft)
Inside length: 15 m (50.5 ft)	Inside height: 3.2 m (10.6 ft)

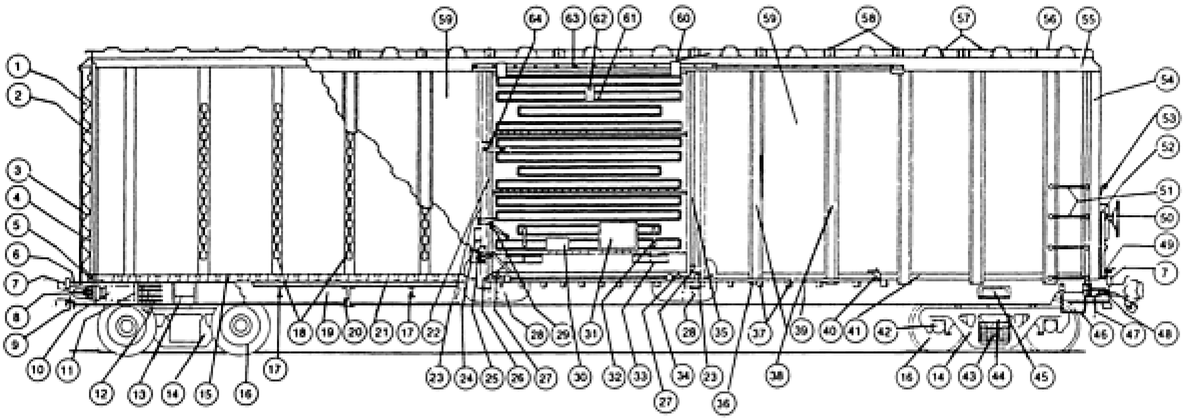

FIG. 13 SIDE VIEW OF A GENERAL-PURPOSE BOXCAR. 1, END SHEET (TOP); 2, END LINING (TOP); 3, END SHEET (BOTTOM); 4, END LINING (BOTTOM); 5, END SILL; 6, STRIKER FACE RIB; 7, STRIKER FACE; 8, KEY SLOT LINER AND FRONT DRAFT LUG; 9, COUPLER CARRIER; 10, DRAFT GEAR; 11, FRONT DRAFT GEAR STOP; 12, REAR DRAFT LUG RIBS; 13, COMBINATION CENTER FILLER/CENTER PLATE; 14, TRUCK FRAME; 15, NAILABLE STEEL FLOORING; 16, WHEEL; 17, FLOOR BEAM; 18, LADING STRAP ANCHOR; 19, CENTER SILL; 20, CROSSBEARER; 21, FLOOR STRINGER; 22, DOOR FRONT STOP; 23, DOOR POST; 24, DOOR LOCK BRACKET; 25, DOOR LOCK; 26, DOOR LOCK REINFORCING GUSSET; 27, DOOR ROLLERS; 28, DOOR POST AND SIDE SILL GUSSET; 29, STARTER AND CLOSURE LEVER; 30, ROUTING BOARD; 31, PLACARD BOARD; 32, DOOR HANDLE; 33, FORKLIFT PUSHER BLOCK; 34, DOOR CATCH; 35, SPARK STRIP; 36, SIDE SILL REINFORCEMENT; 37, DOOR TRACK BRACKET; 38, SIDE POST; 39, DOOR TRACK (BOTTOM); 40, DOOR CATCH ASSEMBLY; 41, SIDE SILL; 42, AXLE; 43, TRUCK SPRINGS; 44, TRUCK BOLSTER; 45, JACKING PAD; 46, SILL STEP; 47, UNCOUPLING ROD; 48, KEY SLOT; 49, CROSSOVER STEP; 50, HAND BRAKE WHEEL; 51, LADDER GRAB; 52, LADDER STILE; 53, GRAB; 54, CORNER POST; 55, SIDE PLATE; 56, ROOF SHEET (END); 57, ROOF SHEET (INTERMEDIATE); 58, SEAM CAP; 59, SIDE SHEET (PLAIN); 60, DOOR SAFETY HANGER; 61, SLIDING DOOR; 62, LIFTING LUG; 63, DOOR TOP RETAINER; 64, ANTIPILFERAGE LOCK. COPYRIGHT SIMMONS-BOARDMAN PUBLISHING COMPANY/THE RAILWAY EDUCATIONAL BUREAU. REPRODUCED WITH PERMISSION

Welding Processes. The most common welding processes used in the manufacture of freight cars are FCAW, GMAW, and SAW. SMAW has been largely displaced by semiautomatic, mechanized, or automatic welding procedures. Seams in tank car tanks are produced primarily by SAW. Except for carbon and low-alloy steel tank car tanks, which are stress relieved, most freight car components are used in the as-welded condition.

Reference cited in this section

20. FREIGHT CARS 1992: THE QUALITY PROCESS AT WORK, *PROGRESS. RAILROAD.*, APRIL 1992

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Locomotives

Modern locomotives are diesel-electric, ranging from 0.75 to 3.3 MW (1000 to 4500 hp) in power and weighing as much as 181,000 kg (400,000 lb). The life of a locomotive is expected to be 20 years and with rebuild can be extended to as long as 40 years. Locomotives are generally owned by a railroad and remain within its control. Maintenance is normally done by the owners, but major overhauls and rebuilds are performed by remanufacturing specialists.

Design Considerations and Materials of Construction. The modern diesel-electric locomotive consists of welded components fabricated using the full range of welding processes. The basic components are the underframe, cab structures, engine, generator, tracks, traction motors, electrical cabinets, fuel tank, air compressor, and brake equipment. Components are fabricated from rolled plate, sheet metal, forgings, castings, and various shapes.

Locomotives must be able to withstand forces similar to those experienced by freight cars, including 4.5 MN (10^6 lbf) of squeeze. Buff and draft forces are transmitted through a self-supporting rigid underframe consisting of a fabricated draft gear pocket, "T" or "I" sills, and a bottom plate.

Materials used for fabrication of the main components of a locomotive are typically selected on the basis of good weldability in order to facilitate original construction and subsequent maintenance and rebuild. The underframe, fuel tank, traction motor frame, generator frame, and other structural steel components are fabricated from low-carbon steel containing less than 0.26% C, with and without added alloying. Typical tensile and yield strengths are 415 and 275 MPa (60 and 40 ksi), respectively. Weld filler metals used for these components include E70*xx*, ER70S-X, and E70T-X-type electrodes. Although higher-strength base materials do not improve high-cycle fatigue life, quenched and tempered steels with a tensile strength of 690 MPa (100 ksi) and normalized carbon-manganese steels with a tensile strength of 550 MPa (80 ksi) may be used when high static strength or weight reduction is an issue.

Diesel locomotive track frames are made of cast steel. Except for the application of wear plates, welding on cast steel truck frames is accomplished using E70-type electrodes. Wear plates made of medium-carbon, high-carbon, or austenitic manganese steels are welded using austenitic stainless steel electrodes.

With the exception of certain critical subassemblies (bolsters, center plates, and traction motor frames), which are stress relieved, most diesel locomotive structural fabrications are used in the as-welded condition.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Track

Rails are designed for strength and wear resistance. Locomotives and freight cars weighing up to 181,000 kg (400,000 lb) and riding on steel wheels create high bearing stresses, requiring very-high-strength rail material. Bearing stresses may be up to 2750 MPa (400 ksi).

Rail life averages 70 years. Replacement of rail is extremely expensive. Maintenance of high wear areas, such as switch points and rail ends, requires that repairs be made economically in the field. The desire for long, continuous rails necessitates the use of materials that allow butt welds to be made in the rail that have properties matching the unwelded rail.

Design Considerations and Materials of Construction. There are two categories of rail materials. The largest category by far is T-rail, typically of near-eutectoid composition (~0.8% C), which makes up the thousands of miles of railroad track. The much smaller category is austenitic manganese steel, which is used extensively in cast frogs. The welding practices for each type of material are very different. This section will address the metallurgical reasons for these differences. Welding is applied to T-rail for joining and repair purposes, and is applied to austenitic manganese steel solely for repair purposes.

The controlling feature for welding of near-eutectoid T-rail is hardenability--that is, its tendency to form transformation products such as martensite and bainite, rather than the preferred pearlite. Untempered martensite is unacceptable because of its possible embrittling effect on rail steel; near-eutectoid composition bainite and tempered martensite are avoided because of their poor wear resistance compared with pearlite of the same hardness.

Historically, rail steels have had carbon contents near 0.7 to 0.75%, with manganese contents of 0.85 to 0.95% and silicon levels of up to 0.5% maximum. The as-hot-rolled hardness generally has been near 240 to 270 HB. More modern conventional carbon rail steel has a slightly higher maximum manganese content (1.25%) and allows up to 0.25% Cr to be added to ensure as-hot-rolled hardness near 300 HB.

Metallurgical Characterization. The time-temperature-transformation (TTT) characteristics of steel control the type of transformation product that results from the welding process. A typical continuous cooling curve diagram for the older grades of conventional carbon rail steel is illustrated at the left in Fig. 14. In an effort to ensure that the transformation product resulting from welding is pearlitic, the maximum cooling rate through the transformation region generally must not exceed approximately 4 °C/s (7 °F/s). For rates up to this maximum level, transformation will take place entirely between the pearlite start and finish boundaries. However, should the cooling rates exceed approximately 4 °C/s (7 °F/s), it is possible that the cooling path will miss the nose of the pearlite finish boundary such that untransformed austenite will be available for transformation to bainite or untempered martensite at temperatures below 540 °C (1000 °F).

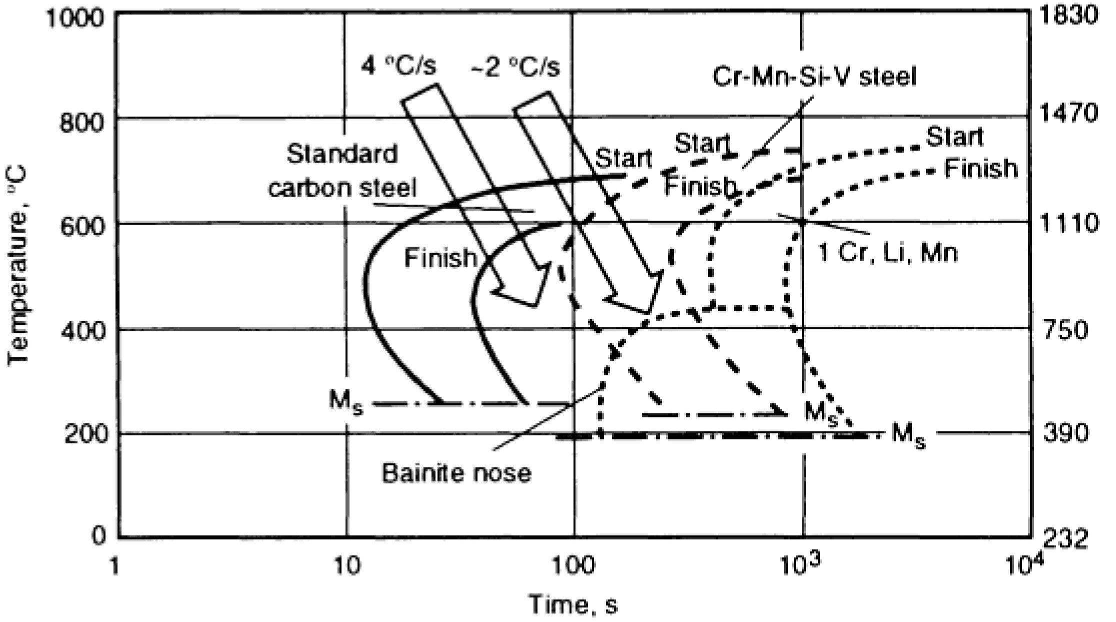

FIG. 14 TTT CHARACTERISTICS OF RAIL STEELS

Alloying additions are sometimes made to near-eutectoid rail steels in an effort to produce pearlite having a smaller spacing between the Fe_3C platelets during cooling following hot rolling; this yields superior hardness and greater wear resistance. The alloying additions most commonly used are the carbide formers: chromium, molybdenum, and vanadium. If added in sufficient amounts (1% Cr; 0.6 to 0.8% Cr and 0.20% Mo; or 1% Cr and 0.08% V), these types of alloy additions tend to increase hardenability significantly and in so doing shift the diagram to longer times. In fact, a specific bainite nose can be obtained (see Fig. 14).

Welding Thermal Cycles. One of the limitations of the alloy approach to strengthening is that a welding practice resulting in a cooling rate near 2 °C/s (3.6 °F/s) (which would be perfectly satisfactory for conventional carbon steel rail) would now lead to the formation of undesirable bainite or untempered martensite. Welding practices must be modified to slow the cooling rate to less than 2 °C/s (3.6 °F/s).

By and large, most rail welding is done to join rails into strings up to 440 m (1440 ft) in length. The most common practice for that purpose is flash butt welding utilizing either ac or dc current. This practice generally involves the use of a number of preheating cycles to heat the rail ends and control the subsequent cooling rate. When the rail ends have been sufficiently preheated, a flashing phase is initiated that removes oxidized metal and establishes a metal ion plasma in which forge welding of the rail ends can be accomplished. The forging operation ideally expels all liquid metal from the bond interface, such that the weld is in reality a solid-state bond.

Postweld Heating and Cooling Cycles. For significantly alloyed rail steels, adjustment of the preheat practice may not be sufficient to limit the final cooling rate. To this end, postheat cycles are applied after the forging cycle. The added expense of postheating has caused railroads to prefer the use of heat-treated carbon steel rails--both those achieving full section hardening and those in which only the head is hardened. These rails generally require no modification of the electric flash butt welding practice. However, a consequence of welding most heat-treated rails (except those classified as slack quenched) is the loss of the heat-treat strengthening in the weldment. This loss can lead to severe batter of the weldment under heavy wheel loads. Thus, some railroads have introduced a postweld air quench to refine the pearlite in weldments.

An alternative approach to air quenching is the use of head-hardened rails having their composition adjusted slightly to achieve slack quenching during the normal cooling after flash butt welding. Typically, such modified compositions will include increased silicon (up to 0.8%) and chromium (up to 0.5%) levels. After normal cooling, the hardness of the weldment will have returned to about the same level (~350 HB) as that of the head-hardened region of the rail. Although this approach works very well, the cost of the rail may be notably higher than that of conventional head-hardened rail. Neither air quenching nor slack quenching prevents the softening that occurs in the outer boundaries of the HAZs.

Other Welding Processes. Rails are also joined together in track by thermite welding--a metal casting process based on the reaction between aluminum powder and iron oxide (see the article "Thermite Welding" in this Volume). The resulting cast structure has low ductility and toughness compared with the hot-worked structure of a flash butt weld. The cooling rates typical of thermite welding are low in comparison with those of the flash butt process, so there generally is no need to alter standard welding practices for use with alloy rail.

Welding also is applied to near-eutectoid rail steels to repair battered and chipped rail ends, wheel burns, and worn and chipped switch points and frogs. Both arc welding and oxyfuel gas welding processes are possible, although arc welding is preferred. Either coated electrodes or flux-cored wire can be used. Electrodes may contain high nickel and manganese contents, and deposits are generally austenitic. Both longitudinal and transverse lay-down patterns have been employed.

Austenitic Manganese Steels. The nominal chemical composition range for austenitic manganese steel is 10 to 14% Mn and 1.0 to 1.4% C. The presence of such high levels of manganese stabilizes the high-temperature fcc austenite phase so that, with rapid cooling, the austenite is retained to room temperatures. The austenite has a relatively low yield strength that is not strongly influenced by carbon or manganese levels. However, its ultimate tensile strength, which is influenced by carbon or manganese levels, can be quite high in comparison with yield strength. This wide separation imbues the steel with a much higher work-hardening capacity than pearlitic rail steels. It is this high work-hardening capacity that makes manganese steels effective in resisting impact loads such as those occurring at railroad frogs. Although the high manganese content can delay the transformation of the high-temperature austenite phase to the room-temperature equilibrium phases (ferrite and carbides), it cannot completely suppress the transformation. If temperatures are elevated to 315 °C (600 °F) for very long periods or to temperatures 480 to 540 °C (900 to 1000 °F) for much shorter periods (a few minutes), the steel can become embrittled. The mechanism of embrittlement is the precipitation of carbides, usually along austenite grain boundaries. For this reason, repair welding must cease when the casting temperature reaches 315 to 370 °C (600 to 700 °F), and the parts must be allowed to cool to a more suitable range.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Sheet Metals

This section deals with a wide variety of sheet metal fabrications, excluding those discussed in previous sections of this article. Ductwork of many types, small and large appliances, containers, pharmaceutical and food processing vessels, and industrial and domestic kitchen equipment are examples of sheet metal applications.

Service conditions are the principal consideration in material selection for such applications. The material attributes listed in Table 1 are a useful starting point. If an application requires cold forming or deep drawing, the ductility of the material will be of primary importance. Corrosion properties are paramount in applications where the weldment will not be protected by some type of coating following fabrication. Mechanical properties figure prominently in designs for structural applications. Density becomes important in many applications for products where shipping weight is a factor. Fracture toughness, conductivity, and magnetic properties are rarely significant in sheet metal fabrications.

Weldability is extremely important in the material selection process, particularly when an application involves very thin sheet materials. The choice of welding process will be determined by the ability to weld thin materials without burnthrough. When dealing with sheet metals, the ability to make sound welds without burnthrough is often synonymous with weldability.

The weldability of sheet metals is vitally affected by the welding process selected, the manner of welding (manual, semiautomatic, or automatic), and the number of parts to be fabricated in a given manufacturing operation. This section will deal with material selection criteria as determined by the joining process to be employed.

Welding process selection is greatly influenced by the need to control distortion. Thinner material is inherently more difficult to weld because of the thermal cycles encountered in welded joints. Distortion is minimized by rigid fixturing and by use of high-energy rapid-travel-rate welding processes. Brazing or soldering allows lower peak temperatures to be used for joining, which is advantageous in distortion control. Materials with low heat conductivity, such as stainless steels, aggravate distortion problems. Warpage of materials during joining is one of the principal problems encountered in the fabrication of sheet metal by any of the joining processes that involve fusion of metal.

Joining Processes

Resistance Welding. Spot, seam, and projection welding are commonly used in sheet metal fabrications. The most commonly used material joined by these processes is low-carbon steel, although low-alloy steels, stainless steels, nickel, and copper-base alloys are also readily joined.

When an unalloyed carbon steel is suitable for an application, the carbon content is generally limited to 0.20% maximum. For products that are to be enameled after welding, the carbon content is usually kept below 0.04% to avoid blistering of the enamel. The designer can then take advantage of the enamel thickness in assessing the strength properties of the part, which will offset the relatively low strength of the very-low-carbon steel.

Spot welding and projection welding are advantageous for control of distortion, but are unsatisfactory for containment of fluids or for elimination of crevices. Seam welding can be suitable for fluid containment, but distortion is a greater problem and crevices continue to be present. As described in the section titled "Automobiles" in this article, joint sealants and adhesives can be used with the resistance welding processes, offering the possibility of fluid containment and crevice elimination.

Arc Welding. Developments over the past 20 years have greatly increased the application of arc welding in the fabrication of sheet metals. All of the arc welding processes are employed, although some are more suitable than others.

The simplest arc welding process is SMAW. For nonrepetitive parts, it is the preferred process and is suitable for thicknesses of most steels down to about 2 mm (0.080 in.). Covered electrodes down to 1.6 mm ($\frac{1}{16}$ in.) in diameter are available in most types of low-carbon, low-alloy, and stainless steels. Burnthrough is a problem with the thinner sheet metals, more so with stainless steel than with carbon steels because of its low thermal conductivity. The SMAW process is commonly used in job shops and for repairs. Fixturing for prevention of distortion and warpage is needed in cases where large, flat surfaces must be kept flat.

Manual welding with the gas-shielded arc welding processes (GMAW, FCAW, and GTAW) offers improved control of the arc energy, especially when using more sophisticated power sources with pulsed current controls. However, travel rate in manual welding is generally slow, so the distortion problem is similar to that encountered with SMAW. Automation of the process offers increased welding speed, use of fixtures with heat sinks for more rapid cooling, and higher productivity and quality.

High-energy-beam welding processes, such as electron and laser beam, offer even higher welding speed and lower distortion. Materials sensitive to atmospheric effects during welding, such as titanium and zirconium, are especially suited to electron beam welding in vacuum chambers or with an inert gas trailing shield. Laser welding is being increasingly used for unalloyed carbon steel sheet fabrication, because the process requires no protection from the atmosphere during welding and because relatively low-power laser equipment of nominal cost is adaptable for thin metal sections. Fully automatic welding systems are generally required for high-energy-beam welding processes.

Brazing and soldering are suitable joining methods for many sheet metal materials. The metal surfaces must be compatible with the liquid-phase joining material. For successful joining, both brazing and soldering depend on capillary action between the mating surfaces to cause flow of the liquid joining material. Oxides on the metal surface must be mechanically or chemically removed in many instances for the liquid metal to flow freely. Fluxes are frequently used to assist this flow process.

These joining processes are easily adaptable to manual low-volume fabrication using oxyfuel torches for the heat source. Job shops customarily use these methods for fabrication or repair. The lower joining temperature is a distinct advantage in that fusion of the base metals does not take place, minimizing distortion.

Soldering is commonly used to join galvanized steel and tin plate. The surface zinc or tin becomes alloyed with the solder during the fusion process and assists in establishing an adherent bond with the underlying steel. The coated surface adjacent to the joint is unaffected and is thus resistant to the service environment.

Brazing requires much higher temperatures and offers many more types of filler metals than soldering. A filler metal must be chosen not only for its ability to flow at the brazing temperature, but also for its compatibility with the service environment. The filler metal is distinctly different from the base metal and may set up electrochemical cells, causing localized corrosion adjacent to the joint.

Brazing and soldering are superior to resistance welding for producing crevice-free joints, but are usually less satisfactory than the arc welding processes for joints that require mechanical and corrosion-resistance properties comparable to those of the base metals.

Nonfusion Joining Processes. Bonding of metal surfaces can be accomplished in many metal systems without fusion. Diffusion bonding, ultrasonic welding, and similar techniques can be used to join very thin metals. These processes are particularly appropriate for the precious metals--for example, in making contacts to foils in electronic circuitry. They are economically unsuitable for continuous seams in liquid or gas containers.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Materials and Properties

Steels. Low-carbon steels are characterized by their considerable ductility at ambient temperatures, which becomes even higher at elevated temperatures. During the welding thermal cycle, plastic deformation occurs in the HAZ. Unless restrained by clamping devices or by the inherent shape of the piece, unwanted distortion in the weld area will result. The thinner the steel, the greater the distortion problem.

Fusion welds have greater strength and lower ductility than the base metal. If a sheet metal is to be cold formed after joining, this strength mismatch may need to be addressed in the design. Although the ductility of the weld is less than that of the base metal, it is usually adequate for most cold-forming operations. Distortion that forms during the welding operation can sometimes be ignored if the component is subsequently cold formed.

In corrosive or high-temperature environments, carbon steel weldments will perform in a manner generally comparable to that of the unaffected base metal. The cast structure of the weld is often less resistant to some environments than the wrought structure of the unaffected steel. Also, the welding process usually results in fine oxide inclusions in the weld, which are often less resistant to corrosion.

The higher strength of low-alloy steels does not offer the same advantages to sheet metal welding applications that it does to structural steel applications. Thinner sections might be possible by using the higher-strength steels, but distortion and weldability considerations usually make such choices uneconomical unless weight savings are critical.

Table 21 lists the principal ASTM specifications for weldable carbon and low-alloy sheet steels.

TABLE 21 PRINCIPAL ASTM SPECIFICATIONS FOR WELDABLE CARBON AND LOW-ALLOY SHEET STEELS

TYPE OF STEEL	PURPOSE	ASTM NO.	
		COLD ROLLED	HOT ROLLED
LOW CARBON	COMMERCIAL QUALITY	A 366	A 569
		A 794	A 659
	STRUCTURAL	A 611	A 570
	ENAMELING	A 424	A 424
	DEEP DRAWING	A 619	A 621
		A 620	A 622
CARBON-MANGANESE	ATMOSPHERIC CORROSION	A 606	A 606
	PRESSURE VESSELS	. . .	A 414
LOW ALLOY	HIGH STRENGTH	. . .	A 715
	NIOBIUM OR VANADIUM ADDED	A 607	A 607

Aluminum and aluminum alloy specifications for sheet metal are included in ASTM B 209. The low density of aluminum when compared with most metals makes aluminum sheet metal desirable for fabrications where weight savings are important. It is also useful for its adherent oxide, which makes it resistant to many environments at ambient temperatures.

The refractory oxide on the surface of aluminum becomes increasingly important as the thickness of the sheet is reduced. Gas-shielded arc welding processes, especially those using ac power, are essential to break down this adherent refractory aluminum oxide in order to make successful welds. Manual welding with covered electrodes is employed occasionally; oxyfuel welding is not suitable. Brazing and soldering are possible only with chemically active fluxes that dissolve the oxide layer.

The high thermal conductivity of aluminum can cause distortion. Gas-shielded arc welding processes overcome this problem by concentrating the arc energy sufficiently to cause fusion to occur. Electron beam welding is also suitable.

Aluminum alloys offer higher strength, which is especially important in applications where weight saving is a factor. Furthermore, their thermal conductivity is lower than that of pure aluminum, making fusion welding easier.

Distortion control is usually provided by rigid fixturing and backups. The metallic backup of the fixtures offers the advantage of heat sinks and backside weld contour control. Adding an inert backing gas often prevents oxidation of the molten weld pool.

Stainless Steel. The corrosion resistance of stainless steel is provided by chromium in an amount of at least 12%. When the content of nickel is greater than 8%, the austenitic form of iron becomes stable at ambient temperatures. The most frequently employed stainless steels for sheet metal fabrication contain 18% Cr and 8% Ni and are known as austenitic stainless steels. ASTM specifications normally used in stainless steel sheet metal applications are listed in Table 22.

TABLE 22 PRINCIPAL ASTM SPECIFICATIONS FOR WELDABLE STAINLESS STEEL SHEET

TYPE OF STEEL	PURPOSE	ASTM NO.
AUSTENITIC CHROMIUM-NICKEL	GENERAL REQUIREMENTS	A 480
	GENERAL PURPOSES	A 167
		A 240
	ARCHITECTURAL	A 666
	ROLL CLAD	A 264
DUPLEX CHROMIUM-NICKEL	GENERAL PURPOSES	A 240
FERRITIC CHROMIUM	GENERAL REQUIREMENTS	A 480
	GENERAL PURPOSES	A 240
		A 276
	ROLL CLAD	A 263

The austenitic stainless steels are extremely ductile and easily welded. The weld strengths, both yield and ultimate tensile, are generally higher than those of the base metal, but ductility remains high, permitting subsequent cold forming or straightening if necessary.

The thermal conductivity of stainless steels is about 30% less than that of carbon steels. The thermal cycles during welding cause plastic deformation to occur in the widely heated zones adjacent to the weld. Moreover, the coefficient of thermal expansion for stainless steels is about 30% greater than that of carbon steels. These two properties account for the even greater distortion problems encountered in the austenitic stainless steels compared with carbon steels. Therefore, rigid fixturing and heat sinks are essential when welding these alloys.

Stainless steels are principally used for their resistance to aqueous corrosion and for their strength and oxidation resistance at elevated temperatures. With regard to certain corrosive media, it is important to recognize that the thermal cycles during welding cause carbide precipitation to occur in the HAZ, with severe loss in corrosion resistance. This can be overcome by the use of low-carbon types (0.03% maximum) or those stabilized with titanium or niobium. The stabilized stainless steels are also stronger at elevated temperatures. The fluidity of the very-low-carbon stainless steels often affects the attainment of the desired weld profiles. This is particularly troublesome in autogenous GTAW when sulfur is unusually low (<0.010%).

Because many applications require avoidance of crevices at weld joints, arc welding is the preferred joining method. Shielding from the atmosphere during welding is essential, and while manual arc welding with covered electrodes is suitable for many applications, gas-shielded processes, especially those using inert gases, are useful in minimizing the oxidation of chromium and other essential alloying ingredients.

Butt joints in sheet metals are welded in a single pass. If the backside of the weld will be exposed in the finished weldment, it should have a suitable contour. Purging of the backside of the joint with an inert or inactive gas, usually argon or helium, but sometimes nitrogen, is useful in producing the desired finish. Gas purging is often built into the distortion-control fixturing.

Austenitic stainless steel welds are almost invariably put into service in the as-deposited condition. The residual stresses associated with the welding operation can give rise to SCC in certain service environments. This situation can be substantially overcome by selecting one of the duplex stainless steels, which have structures of roughly half austenite and half ferrite. These steels contain less nickel than the austenitic stainless steels and provide corrosion resistance by the addition of molybdenum. Filler metals with matching compositions are available for the duplex stainless steels; more commonly, welds are made using an austenitic stainless steel filler metal, provided that the type is suitable for the service environment.

The ferritic stainless steels that contain little or no nickel are less suitable for sheet metal fabrications. Although they do not present the problems of the austenitic stainless steels in terms of low thermal conductivity and high coefficient of expansion, their metallurgical weldability is troublesome, forming constituents that have low ductility in the HAZs. Postweld heat treatments are required to restore ductility, which in practice complicate the distortion problem.

Other Metals and Alloys. ASTM specifications for weldable nonferrous sheet metals are listed in Table 23. Like the stainless steels, nickel-base alloys are commonly used in sheet metal weldments. The welding processes and controls are similar to those described for the austenitic stainless steels, although corrosion impairment due to carbide precipitation is not a concern. Unalloyed nickel, nickel-chromium alloys, and nickel-copper alloys are readily weldable, although their physical properties can affect distortion and burnthrough.

TABLE 23 PRINCIPAL ASTM SPECIFICATIONS FOR WELDABLE NONFERROUS SHEET METALS

TYPE	PRINCIPAL ALLOYING ELEMENTS	ASTM NO.
NICKEL ALLOYS	NONE	B 162
	CU	B 127
	CR-FC	B 168
	MO	B 333
	FE-CR	B 409
	FE-CR-MO-CU	B 424
	MO-CR-FE	B 434
	CR-MO-FE	B 435
	CR-MO-NB	B 443
	CR-FE-MO-CU-NB	B 463
	FE-CR-SI	B 536
	MO-CR	B 575
	CR-FE-MO-CU	B 582
	FE-CR-MO-NB	B 599
	FE-CR-MO	B 620
	FE-CR-MO-CU	B 625
COPPER ALLOYS	NONE	B 152
	SI	B 96, B 97
	SN	B 103
	AL	B 169
	NI	B 402, B 422
TITANIUM AND ITS ALLOYS	. . .	B 265
ZIRCONIUM AND ITS ALLOYS	. . .	B 352, B 551
NIOBIUM AND ITS ALLOYS	. . .	B 393
MOLYBDENUM AND ITS ALLOYS	. . .	B 386

Copper and its alloys in sheet metal form are somewhat more difficult to weld. Unalloyed copper has a very high thermal conductivity. The tin-alloyed bronzes are much more easily welded, whereas suitable fusion welds are virtually impossible in the zinc-alloyed brasses. Copper-nickel alloys (both 10% and 30% Ni), on the other hand, are easily welded with many of the processes and controls cited for stainless steels.

Titanium is used in sheet metal fabrications for its superior corrosion-resistance properties in many environments. Gas shielding or vacuum environments are needed during the welding operation to protect against contamination of the weld zone by oxygen and nitrogen from the air. Gas-shielded arc welding processes are almost invariably used, because smooth contour welds free from crevices are usually demanded. These can be fabricated without resorting to vacuum chambers, provided that both sides of the weld at the welding arc and in the trailing solidifying weld regions are protected from the atmosphere.

Zirconium, niobium, and molybdenum and their alloys are occasionally produced in sheet metal form and are welded using the gas-shielded processes.

High-energy-beam welding processes are also adaptable to exotic metals and alloys. Electron beam welding in vacuum chambers is especially suited to joining these metals, the unit cost of which is high. The added equipment and fixturing costs can be justified by the reliable welds that are produced.

Coated Metals. Among the most commonly used coated sheet metals are galvanized steel and tin plate. As described above, soldering is a useful technique for joining these metals. Arc welding is possible, but it destroys the coating adjacent to the weld, thus negating the corrosion resistance for which the coated steel was specified. If arc welding is used on galvanized steel, good ventilation must be provided to protect the operator from zinc fumes.

In applications where the corrosion resistance offered by austenitic stainless steels is needed on only one side of a vessel, roll-clad steels are often an economical choice, especially for heavier-gage thicknesses. Distortion during welding is much less of a problem than with solid stainless steel because of the physical properties of the steel backing and its greater thickness. The welding of clad steels in thin sections is usually accomplished in a single pass using a higher-alloyed filler metal (e.g., type 309 for joining type 304 clad steel) to offset the effect of dilution with the unalloyed steel backing.

Painted sheet steels are commonly encountered in repair welding applications. Paints or other surface contaminants should be removed before welding to ensure sound joints.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Surface Finishing

Protective Coatings. Sheet metal weldments fabricated from unalloyed carbon steel are often more vulnerable to failure by corrosion or other environmental deterioration because of their relatively thin sections. Moreover, the welds themselves are frequently anodic to the base metal, causing corrosion failure in those regions. Suitable protective coatings, such as enamels, are often mandated for sheet steel weldments.

Postweld Finishing. Sheet metal fabrications frequently remain in the as-deposited condition, with only slag removal considered essential. This places a premium on smooth contours and good workmanship standards for meeting appearance criteria.

Appearance demands sometimes require the removal of all evidence of the weld. Grinding of the weld surface to blend with the base metal is often performed for appearance reasons or to remove notches for fatigue service applications. The amount of grinding can be substantially reduced if the welding process is intelligently selected and if the controls are carefully set forth in the fabrication specifications. Sheet metal fabrications in which the welds are to be ground smooth are more economically produced when the amount of weld reinforcement is kept to a minimum by appropriate welding process selection and controls.

Stainless steel sheet fabrications often require polishing to match the finish of the base metal. Polishing of the welds can be done either on the as-deposited finish or, if necessary, on the ground and blended finish.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Fabrication Codes and Specifications

Most of the requirements for sheet metal fabrications are set forth in user specifications or contractual documents. Aside from those considered in the preceding sections, few industry-wide codes are available. The American Welding Society publishes a specification titled "Structural Welding Code--Sheet Steel" (AWS D9.1) that covers the welding of low-carbon hot-rolled and cold-rolled sheet and strip steel with and without zinc coatings (galvanized steel). Qualification of welding procedures and welding operators is set forth in this specification.

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

References

1. R.P. REED, M.B. KASEN, H.I. MCHENRY, C.M. FORTUNKO, AND D.T. REED, *FITNESS FOR SERVICE CRITERIA FOR PIPELINE GIRTH WELD QUALITY,* BULL. 296, WELDING RESEARCH COUNCIL, JULY 1984
2. "BOILER AND PRESSURE VESSEL CODE," SECTION II, AMERICAN SOCIETY OF MECHANICAL ENGINEERS
3. *STEELS FOR HYDROGEN SERVICE AT ELEVATED TEMPERATURES AND PRESSURES IN PETROLEUM REFINERIES AND PETROCHEMICAL PLANTS,* 4TH ED., API PUBL. 941, AMERICAN PETROLEUM INSTITUTE, APRIL 1990
4. S. YUKAWA, *GUIDELINE FOR PRESSURE VESSEL SAFETY ASSESSMENT,* NIST SPEC. PUBL. 780, NATIONAL INSTITUTE FOR STANDARDS AND TECHNOLOGY, P 14-22
5. W.H. WINN, WELDABILITY OF LOW ALLOY STEELS, *BR. WELD. J.,* VOL 11 (NO. 8), AUG 1964
6. W.F. SAVAGE, E.F. NIPPES, AND E.S. SZEKERES, A STUDY OF WELD INTERFACE PHENOMENA IN A LOW ALLOY STEEL, *WELD. J.,* SEPT 1976, P 260S-268S
7. G.E. DIETER, *ENGINEERING DESIGN,* MCGRAW-HILL, 1983
8. B.D. TAPLEY, *ESCHBACK'S HANDBOOK OF ENGINEERING FUNDAMENTALS,* JOHN WILEY & SONS, 1990
9. *AEROSPACE STRUCTURAL MATERIALS HANDBOOK,* VOL III, 1984
10. SIMS AND HAGEL, *THE SUPERALLOYS,* JOHN WILEY & SONS, 1972
11. CASTRO AND DECADENT, *WELDING METALLURGY OF STAINLESS AND HEAT RESISTING STEELS,* CAMBRIDGE PRESS, 1968
12. EVANS AND WILSHIRE, *CREEP IN METALS AND ALLOYS,* DOTESIOS PRINTERS, U.K., 1985
13. *AEROSPACE STRUCTURAL MATERIALS HANDBOOK,* VOL IV, 1984
14. *BRAZING HANDBOOK,* AMERICAN WELDING SOCIETY, 1976
15. D. HARVEY, FILLER METAL SELECTIONS FOR CRITICAL AEROSPACE FABRICATION, *WELD. DESIGN FABR.,* VOL 60 (NO. 2), MARCH 1991, P 765-80
16. *USAF DAMAGE TOLERANT DESIGN HANDBOOK: GUIDELINES FOR ANALYSIS AND DESIGN OF DAMAGE TOLERANT AIRCRAFT,* U.S. AIR FORCE, 1984

17. PATTERSON *ET AL.* TITANIUM ALUMINIDE: ELECTRON BEAM WELDABILITY," *WELD. J.,* JAN 1990, P 39S-44S
18. R.D. STOUT, *WELDABILITY OF STEELS,* 4TH ED., WELDING RESEARCH COUNCIL, 1987, P 338-437
19. *RAILROAD FACTS,* ASSOCIATION OF AMERICAN RAILROADS, 1990
20. FREIGHT CARS 1992: THE QUALITY PROCESS AT WORK, *PROGRESS. RAILROAD.,* APRIL 1992

Material Requirements for Service Conditions

ASM Committee on Material Requirements for Service Conditions*

Selected References

- "RECOMMENDED PRACTICE FOR RESISTANCE WELDING," AWS C1.1-66, AMERICAN WELDING SOCIETY, 1966
- "RECOMMENDED PRACTICE FOR RESISTANCE WELDING COATED LOW CARBON STEELS," AWS C1.3-70, AMERICAN WELDING SOCIETY, 1970
- "AUTOMOTIVE RESISTANCE SPOTWELDING," ENGINEERING STANDARD GM4488M, GENERAL MOTORS CORP., AUG 1990
- G.E. LINNERT, *WELDING METALLURGY,* VOL II, AMERICAN WELDING SOCIETY, 1967
- NATIONAL RESEARCH COUNCIL, *MATERIALS SCIENCE AND ENGINEERING FOR THE 1990'S: MAINTAINING COMPETITIVENESS IN THE AGE OF MATERIALS,* NATIONAL ACADEMY PRESS, 1989
- K. MASUBUCHI, *ANALYSIS OF WELDED STRUCTURES,* PERGAMON PRESS, 1980
- RUSSELL *ET AL.,* "CONSIDERATIONS OF METAL JOINING PROCESSES FOR SPACE FABRICATION, CONSTRUCTION AND REPAIR," PAPER PRESENTED AT 23RD INT. SAMPE TECH. CONF., 21-24 OCT 1991

Introduction to the Selection of Carbon and Low-Alloy Steels

Bruce R. Somers, Lehigh University

Introduction

CARBON AND LOW-ALLOY STEELS are welded more frequently than any other materials because of their widespread applications and good weldability. This versatility is principally due to the metallurgical characteristics of the iron-base system. The characteristics include the ability to undergo allotropic (that is, microstructural) transformations that allow the opportunity for hardening and strengthening through martensitic and bainitic transformations or precipitation mechanisms in addition to the ability to be readily alloyed with a wide variety of elements.

This Section of the Handbook will provide the reader with information about the suitability of these materials for particular types of assemblies. This Section will also provide in-depth metallurgical information about the response of these materials to welding conditions and microstructural evolution in the weld heat-affected zone (HAZ).

Introduction to the Selection of Carbon and Low-Alloy Steels

Bruce R. Somers, Lehigh University

Weldability

The articles in this Section focus on the weldability of carbon and low-alloy steels. The user should consider weldability as being divided into two general classes: (1) fabrication weldability and (2) service weldability.

Fabrication weldability addresses the question, "Can one join these materials by welding without introducing detrimental discontinuities?" The acceptability of these discontinuities depends on the application requirements for the particular weldment. The fabrication weldability of a steel may be adequate for a noncritical application. However, this same steel may not be recommended for a critical application, or special precautions, such as preheat, may be required when welding. Fabrication weldability deals primarily with discontinuities such as hydrogen-assisted cold cracks, hot cracks, reheat cracks, lamellar tearing, and porosity.

Service weldability addresses the question, "Will the finished weldment have properties adequate to serve the intended function?" An important aspect of service weldability is the comparison of HAZ properties with those of the unaffected base metal. Again, the acceptability of the service weldability depends on the intended application. The service weldability of a particular steel may be acceptable for an application where corrosion is of prime importance and toughness is secondary. However, the same steel may be unacceptable for an application where toughness is most important.

Service weldability involves the effect of the welding thermal cycle on the properties in the HAZ. Service weldability often determines the range of heat inputs allowable for a particular steel. Low heat inputs may introduce unacceptable low-toughness microstructures, as well as fabrication-weldability problems associated with cold cracking. High heat inputs can introduce coarse microstructures with both low toughness and low strength. The heat input alone does not control the resulting microstructure and HAZ properties, but the induced thermal cycle controls the microstructure and properties. Therefore, both heat input and thickness should be considered.

Introduction to the Selection of Carbon and Low-Alloy Steels

Bruce R. Somers, Lehigh University

Classification of Steels

The carbon and low-alloy steels discussed in this Section cover a wide variety of compositions and properties. Steels are most often classified according to their carbon and/or alloy content. The various classifications have become known under assorted designations, such as plain-carbon steel, carbon-manganese steel, medium-carbon steel, low-alloy steel, high-strength low-alloy steel, and microalloyed steel. Recently, a new classification of steel has introduced the steel-processing technique as a categorization factor. These steels, known by various designations, are most often described as thermal-mechanical-controlled processing (TMCP) steels. The boundaries between all the above classes are often diffuse, they frequently overlap, and they are sometimes arbitrary (Table 1).

TABLE 1 CARBON STEELS CLASSIFIED ACCORDING TO CARBON CONTENT

REF	CARBON, %			
	LOW-CARBON STEEL	MILD STEEL	MEDIUM-CARBON STEEL	HIGH-CARBON STEEL
1	≤0.15	0.15-0.30	0.30-0.50	0.50-1.00
2	≤0.15	0.15-0.35	0.35-0.60	0.60-1.0
3	<0.30	. . .	0.30-0.60	>0.6
4	≤0.30	≤0.25	0.30-0.60	0.60-1.00

The steels covered in this Section have been grouped into six general classes according to composition or processing:

- LOW-CARBON STEELS
- HIGH-STRENGTH LOW-ALLOY STEELS
- QUENCHED-AND-TEMPERED STEELS

- HEAT-TREATABLE LOW-ALLOY STEELS
- THERMAL-MECHANICAL-CONTROLLED PROCESSING STEELS
- CHROMIUM-MOLYBDENUM STEELS

There will be some overlap between these classifications because fabricators often use some steels in more than one heat-treated condition. Typical steel compositions in these classifications are given in Table 2.

TABLE 2 COMPOSITION AND CARBON EQUIVALENT OF SELECTED CARBON AND LOW-ALLOY STEELS

STEEL DESIGNATIONS			COMPOSITION, WT%										CARBON EQUIVALENT, %
AISI	ASTM	MIL-STD	C	Mn	P	S	Si	Ni	Cr	Mo	V	Nb	
LOW-CARBON STEELS													
1010, 1020, 1030	A 36	. . .	0.10-0.30	0.50-1.0	<0.04	<0.05	0.10-0.25	. . .	. . .	. . .	. . .	. . .	0.20-0.50
HIGH-STRENGTH LOW-ALLOY STEELS													
. . .	A 242, A 572, A 588	. . .	0.10-0.25	0.50-1.5	<0.04	<0.05	0.20-0.35	0-0.75	0-0.75	0-0.25	0-0.05	0-0.04	0.25-0.75
QUENCHED-AND-TEMPERED STEELS													
. . .	A 508, A 517	S-16216 (HY-80)	0.10-0.30	0.20-1.5	<0.04	<0.05	0.20-0.35	0-3.4	0-1.5	0-0.5	0-0.05	0-0.04	0.35-1.10
HIGH-TREATABLE LOW-ALLOY STEELS													
4140, 4340	. . .	. . .	0.30-0.50	0.50-1.0	<0.04	<0.05	0.15-0.30	0-3.0	0.50-1.0	0.15-0.25	. . .	. . .	0.55-1.20
THERMAL-MECHANICAL-CONTROLLED PROCESSING STEELS													
. . .	A 841		0.04-0.15	0.70-1.5	<0.03	<0.03	0.20-0.35	. . .	0-0.25	0-0.05	0-0.05	0-0.03	0.20-0.50
CHROMIUM-MOLYBDENUM STEELS													
. . .	A 217, A 387	. . .	0.10-0.20	0.50-0.70	<0.035	<0.04	0.15-0.50	. . .	0.50-9.0	0.50-1.0	. . .	. . .	0.38-2.0

Low-carbon steels contain up to approximately 0.30 wt% C and up to about 1.65 wt% Mn. Most as-rolled steel used for welded applications consists of low-carbon steel. This group encompasses steels that can have wide variations in weldability. For example, it is possible to weld low-carbon steels with less than 0.15% C by most welding processes. It is also possible to weld low-carbon steels containing 0.15 to 0.30% C (sometimes termed mild steels) in thicknesses up to

25 mm (1 in.). However, thicker sections of these steels may require additional measures in order to produce successful welds.

High-strength low-alloy (HSLA) steels are designed to provide better mechanical properties than those of conventional carbon steels. These steels generally have yield strengths of 290 to 550 MPa (42 to 80 ksi). They are also generally of the carbon-manganese type, with very small additions of niobium and vanadium to ensure both grain refinement and precipitation hardening. The term microalloyed steels is often used in reference to these materials. Fabricators usually weld these steels in the as-rolled or the normalized condition. The weldability of most HSLA steels is similar to that of mild steel.

Recently, a new family of steels has been appearing in the literature under the designation HSLA steels. These are the low-carbon, copper-bearing age-hardening steels, also frequently referred to by their ASTM designations, A 710 or A 736 (Table 3); or by their military specification grade designation, HSLA 80 or HSLA 100 (MIL-S-24645). These steels are not truly low-alloy, because the copper, nickel, and chromium contents are all usually near 1%. However, the weldability of these materials is very good, mainly because of their low carbon contents (less than 0.06% C). These steels are usually used in the quenched-and-aged condition. Because of these two conditions, it is probably more appropriate to group these materials with the thermal-mechanical-controlled processing steels.

TABLE 3 HEAT ANALYSIS COMPOSITIONS OF SELECTED LOW-CARBON COPPER-BEARING AGE-HARDENING STEELS

SPECIFICATION	GRADE	COMPOSITION, %									
		C (max)	Mn	P (max)	S (max)	Si (max)	Cr	Ni	Mo	Cu	Nb (min)
A710/A710M	A	0.07	0.40-0.70	0.025	0.025	0.40	0.60-0.90	0.70-1.00	0.15-0.25	1.00-1.30	0.02
	B	0.06	0.40-0.65	0.025	0.025	0.15-0.40	. . .	1.20-1.50	. . .	1.00-1.30	0.02
	C	0.07	1.30-1.65	0.025	0.025	0.40	. . .	0.70-1.00	0.15-0.25	1.00-1.30	0.02
A736/A736M	A	0.07	0.40-0.70	0.025	0.025	0.40	0.60-0.90	0.70-1.00	0.15-0.25	1.00-1.30	0.02
	C	0.07	1.30-1.65	0.025	0.025	0.40	. . .	0.70-1.00	0.15-0.25	1.00-1.30	0.02

Source: Ref 5

Quenched-and-tempered steels are heat treated to provide yield strengths of 345 to 1035 MPa (50 to 150 ksi). ASTM A 514 and A 517 are well-known examples of this category of steel. Other examples include the HY-80, HY-100, and HY-130 nickel-chromium-molybdenum steels covered by military specifications. Weldments fabricated from these materials generally do not need further heat treatment except for a postweld heat treatment (stress relief) in some special applications. For example, HY-100 and HY-130 require controls on the filler-metal carbon content, preheat/interpass minimums and maximums, heat input, and postweld soaks to achieve the desired military mechanical properties.

Advantages of HSLA Steels Over Selected Quenched-and-Tempered Steels. In certain situations, the U.S. Navy substitutes HSLA-80 for HY-80, and HSLA-100 for HY-100. One significant advantage of this substitution is the reduced welding preheat required for the copper age-hardening HSLA versus the HY steels. It should also be noted that HSLA-100 steel does not weld like a mild steel.

Heat-treatable low-alloy (HTLA) steels are usually reaustenitized, then quenched and tempered after welding. These are relatively hardenable steels that in their quenched-and-tempered condition develop yield strengths well above 965 MPa (140 ksi). Weld metals usually cannot develop acceptable combinations of strength and toughness at these levels in the as-welded or stress-relieved condition. Therefore, it is necessary to reaustenitize and then quench and temper the entire weldment after welding. These steels are often referred to by their AISI designations (for example, AISI 4140 and AISI 4340).

Thermal-mechanical-controlled processing (TMCP) steels are a relatively recent addition to the family of steels. These steels are generally produced with a combination of controlled rolling followed by accelerated cooling or in-line direct quenching. This processing allows the steelmaker to develop a combination of high strength and high toughness while maintaining good weldability. The weldability is good because the alloy content of these steels can be kept very lean, with carbon contents generally below 0.06 wt%. Yield strength levels as high as 700 MPa (100 ksi) and above are possible with these steels. Generally, they can be welded without preheat. However, at the high strength levels, preheat may be required in order to prevent cracking in the weld metal.

Chromium-molybdenum steels are extensively used for elevated-temperature applications in the power and petroleum refinery industries. The chromium content of these steels varies from 0.5 to 9%, and the molybdenum content from 0.5 to 1.0%. Typical compositional ranges for chromium-molybdenum steels are given in Table 2. They are usually supplied in the normalized-and-tempered or the quenched-and-tempered condition.

Because these materials have reasonable hardenability, adequate precautions must be taken to avoid hydrogen-assisted cold cracking. The service application frequently imposes additional requirements on the welding of these steels. For example, in the power industry, these materials are required for their creep resistance, and the weld metal and HAZ must provide adequate creep properties. The corrosion environment in refineries requires that the maximum HAZ hardness be limited to avoid corrosion cracking.

References cited in this section

1. *WELDING HANDBOOK*, VOL 4, 7TH ED., AWS, 1982, P 9
2. G. LINNERT, *WELDING METALLURGY*, VOL 1, 3RD ED., P 45
3. S. KALPAKJIAN, *MANUFACTURING PROCESSES FOR ENGINEERING MATERIALS*, ADDISON-WESLEY, 1984, P 146
4. *PROPERTIES AND SELECTION: IRONS, STEELS, AND HIGH-PERFORMANCE ALLOYS*, VOL 1, *ASM HANDBOOK*, ASM INTERNATIONAL, 1990, P 147-148, 390
5. *1989 ANNUAL BOOK OF ASTM STANDARDS*, VOL 1.04, *STEEL*: *STRUCTURAL*, *REINFORCING*, *PRESSURE VESSEL*, RAILWAY ASTM, 1989, P 494, 523

Introduction to the Selection of Carbon and Low-Alloy Steels

Bruce R. Somers, Lehigh University

Relative Susceptibility of Steels to Hydrogen-Assisted Cold Cracking

Figure 1 illustrates where these classes of steel are located on a Graville diagram. Graville has suggested that the susceptibility to hydrogen-assisted cold cracking can be evaluated by calculating the carbon equivalent (CE) and comparing it to the carbon content (Fig. 1). Zone I steels have low carbon and low hardenability and are not very susceptible to cracking. Zone III steels have both high carbon and high hardenability, and all welding conditions will produce crack-sensitive microstructures. Therefore, to avoid hydrogen-assisted cold cracking in Zone III steels, the user must apply low-hydrogen procedures, including preheat and postweld heat treatment. Zone II steels have higher carbon levels with lower hardenability. Thus, it is possible to avoid crack-sensitive microstructures by restricting HAZ cooling rates. This can be accomplished through control of heat input and, to a minor extent, with preheat.

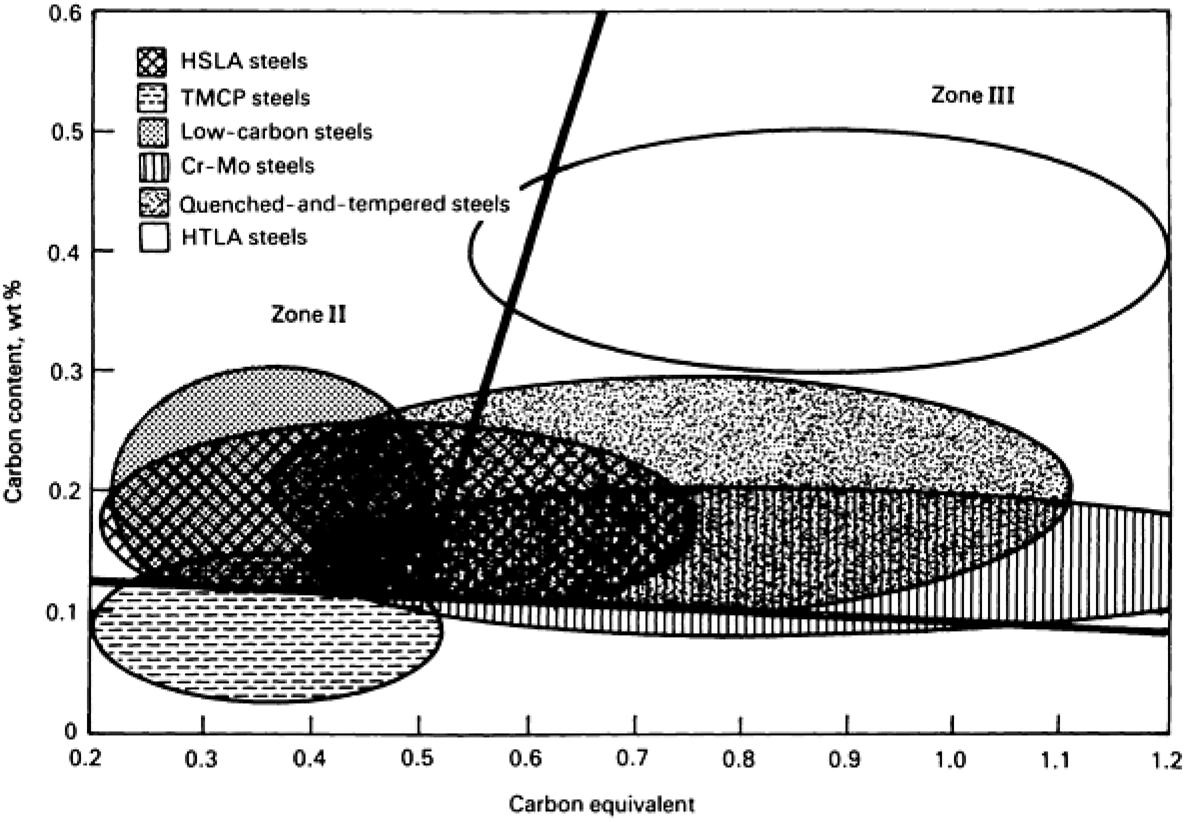

FIG. 1 GRAVILLE DIAGRAM SHOWING SUSCEPTIBILITY OF STEELS TO HYDROGEN-ASSISTED COLD CRACKING RELATIVE TO CARBON CONTENT AND CARBON EQUIVALENT (CE), WHERE CE = %C + (%MN + %SI)/6 + (%NI + %CU)/15 + (%CR + %MO + %V)/5. SUSCEPTIBILITY TO COLD CRACKING PROGRESSIVELY INCREASES AS STEELS MIGRATE FROM ZONES I TO II, TO III. SEE TEXT FOR EXPLANATION. SOURCE: REF 6

Figure 1 also shows that heat-treatable alloy steels, squarely in Zone III, require special considerations for welding. Chromium-molybdenum and quenched-and-tempered steels also require attention, as will some HSLA steels. As mentioned in the section "Low-Carbon Steels" of this article, low-carbon steels are readily welded except in thick sections, for which some precautions may be necessary. The TMCP steels have been specifically developed to lie in Zone I, and so their weldability is excellent. Although Fig. 1 represents only one aspect of weldability and there are many other considerations, the desired preference with regard to hydrogen-assisted cold cracking is to use steels that push the compositional envelope toward the lower-left corner of the Graville diagram.

Reference cited in this section

6. B.A. GRAVILLE, COLD CRACKING IN WELDS IN HSLA STEELS, *WELDING OF HSLA (MICROALLOYED) STRUCTURAL STEELS*, PROC. INT. CONF., AMERICAN SOCIETY FOR METALS, NOV 1976

Introduction to the Selection of Carbon and Low-Alloy Steels

Bruce R. Somers, Lehigh University

References

1. *WELDING HANDBOOK*, VOL 4, 7TH ED., AWS, 1982, P 9
2. G. LINNERT, *WELDING METALLURGY*, VOL 1, 3RD ED., P 45
3. S. KALPAKJIAN, *MANUFACTURING PROCESSES FOR ENGINEERING MATERIALS*, ADDISON-WESLEY, 1984, P 146
4. *PROPERTIES AND SELECTION: IRONS, STEELS, AND HIGH-PERFORMANCE ALLOYS*, VOL 1, *ASM HANDBOOK*, ASM INTERNATIONAL, 1990, P 147-148, 390
5. *1989 ANNUAL BOOK OF ASTM STANDARDS*, VOL 1.04, *STEEL*: *STRUCTURAL*, *REINFORCING*, *PRESSURE VESSEL*, RAILWAY ASTM, 1989, P 494, 523
6. B.A. GRAVILLE, COLD CRACKING IN WELDS IN HSLA STEELS, *WELDING OF HSLA (MICROALLOYED) STRUCTURAL STEELS*, PROC. INT. CONF., AMERICAN SOCIETY FOR METALS, NOV 1976

Influence of Welding on Steel Weldment Soundness

A. Lesnewich

Introduction

ARC WELDING is a process by which metals are joined by coalescence. In most cases, the process uses a compatible filler metal. Before a well-bonded joint can be produced, the joint surfaces must be heated above their melting temperatures in order to completely fuse with the weld metal. Although the metallurgical reactions that involve melting, solidification, and solid-state transformations may not be unusual, the temperatures and cooling rates observed are severe. Active gases also are present and can dissolve in the fused metal. Fluxes are introduced to alloy with and protect the weld metal.

Generally, joints are rigid and will restrain dimensional changes caused by shrinkage and solid-state transformations, producing residual stresses of yield-strength magnitude. Because the metallurgical changes do not occur under equilibrium conditions, and because the stresses are high, many of the reactions can take place in either or both the weld metal and the heat-affected zone (HAZ) of the plate and can produce defects that impair their soundness.

The intent of this article is to point out those defects. However, because of the tremendous variability of the welding processes, it is impossible to provide much detail about the exact mechanisms involved or the corrections that can be made. Furthermore, many corrective measures are self-evident once most defects are explained. One problem, which relates to hydrogen, is not so simple. Because this problem is becoming more relevant as more high-strength, low-alloy (HSLA) steels are being welded, the subject of hydrogen-induced cracking will be emphasized.

Common Defects Associated with Arc Welds

Porosity is caused by the entrapment of small pockets of gas, particularly nitrogen and hydrogen, which typically have a higher solubility in molten, rather than solid, iron. During solidification, gases attempt to leave the weld metal. However, because of high solidification rates, some gas may become trapped. This entrapment depends both on the rate of gas dissolution and on the rate of weld-metal solidification. If the dissolution rate is high, gas bubbles have a chance to develop and escape before the metal solidifies. If the rate is very low the gas may remain in solution, which avoids porosity but allows other problems such as hydrogen-induced cracking or poor toughness. At intermediate rates, the gas can nucleate and, depending on the amount of gas dissolved in the weld metal and the weld-solidification rates, develop bubbles that become trapped. A very severe form of porosity, called worm holes, occurs when the rates of gas evolution and solidification are similar, causing elongated gas pockets to develop instead of essentially spherical bubbles.

Among the possible sources of hydrogen are moisture in fluxes, hydrocarbons in either wiredrawing lubricants or surface contaminants in the joint to be welded, and water leaks in gas-metal arc welding (GMAW) equipment. Nitrogen is gleaned from air that enters the arc regions as a result of poor shielding of the arc. With GMAW, this can happen when the gas flow rates are either so low that crossdraf ts displace the shielding or so high that the surrounding atmosphere is aspirated into the shielding gas. With the shielded metal arc welding (SMAW) process, this can happen because welders are not sufficiently skilled or use improper procedures that cause the arc length to be excessive.

Incomplete fusion can take a number of forms, such as inadequate joint penetration (Fig. 1), absence of root fusion, or lack of side-wall fusion (Fig. 2). These defects can be caused by: inadequate energy input to the weld, particularly insufficient current; excessive travel speed, which allows weld metal to flow ahead of the arc; or improper electrode angles or work positions.

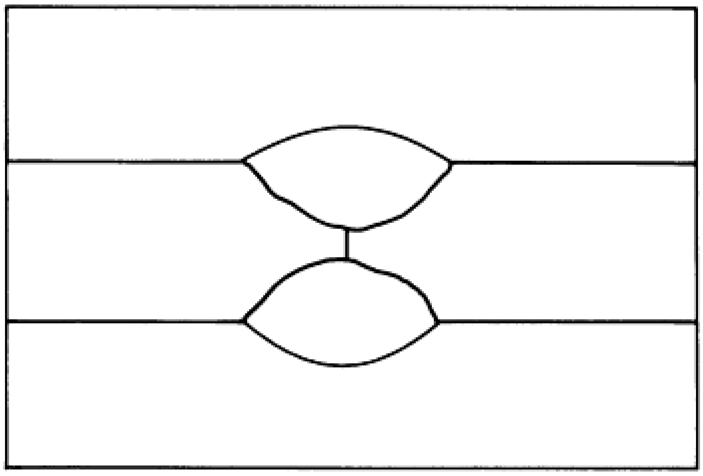

FIG. 1 INCOMPLETE PENETRATION OF WELD INTO BUTT JOINT

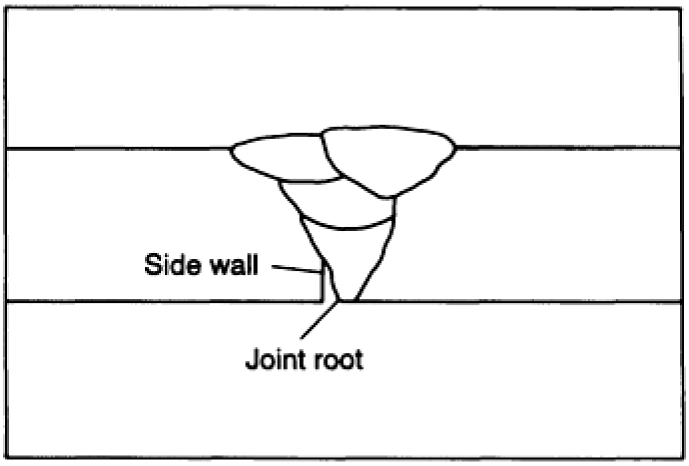

FIG. 2 INCOMPLETE FUSION OF WELD METAL TO SIDE WALL AND TO JOINT ROOT

Problems with joint penetration and root fusion commonly are due to the use of a joint design that is not proper for the welding process being used or to a disregard for the procedures that are intended to provide adequate penetration of the arc. In most cases, this would mean that the welding current is too low. However, in the case of the gas-shielded welding processes, it could mean that the wrong shield gas is being used. For example, with argon-rich gas mixtures, the penetration pattern is relatively shallow, with the exception of a fairly deep central "finger," which is shown in Fig. 3. Unfortunately, this finger generally is not positioned centrally and, therefore, cannot be relied upon. However, shield gas mixtures that are rich in helium or carbon dioxide are capable of a more uniform and deeper useful penetration pattern.

Poor root fusion that occurs when welding from one side requires either a modification in the joint design to allow better penetration or a change to welding from both sides of the plate.

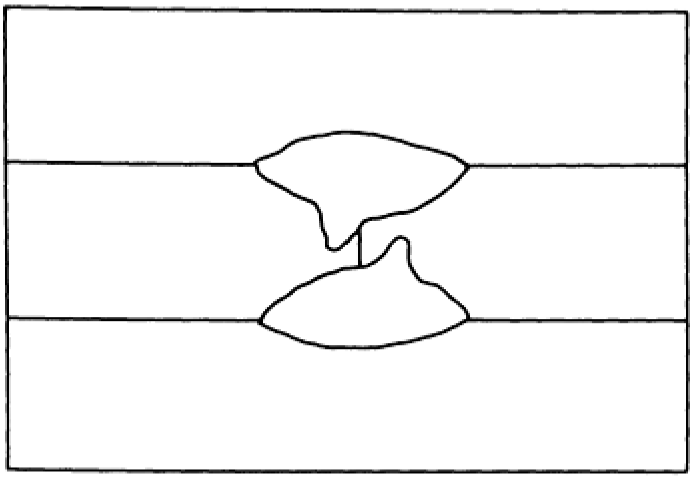

FIG. 3 OFF-CENTER PENETRATION OF CENTRAL CORE OF WELD USING ARGON SHIELD

In most cases, the lack of side-wall fusion between the weld metal and the joint occurs when proper procedures or manipulative techniques are not being used by welders. With the GMAW process, it can be due to using inappropriate variations, such as short-circuiting transfer, when welding heavy sections. Short-circuiting transfer is effective only at low energy levels, which makes it very suitable for welding sheet or thin plate in all positions. This is because the process is designed to provide little penetration and to quickly freeze weld metal. For that reason, the weld metal may not fuse the side walls of joints from which heat is extracted rapidly, that is, those thicker than 6 mm ($\frac{1}{4}$ in.). Both a spray arc with argon and a buried arc with carbon dioxide shielding will deposit welds that are too massive and fluid to be supported in the vertical or overhead positions. However, these processes are very effective for making welds in the flat or horizontal positions. On the other hand, the pulsed-arc variation with argon-rich shielding is very effective in all positions, offering both adequate penetration and control of the weld pool to prevent defects caused by poor side-wall fusion.

Hot cracks, also called centerline or solidification cracks (Fig. 4), are caused by the rejection of low-melting constituents along the centerline of restrained welds. They are called hot cracks because they develop immediately after welds are completed and, sometimes, while the welds are being made. If the welds are broken to expose these cracks, they are found to be blued, or heat tinted. These defects, which are often caused by sulfur and phosphorus, are more likely to occur in higher-carbon alloys. Nearly always, the base plate is their source. The susceptibility to cracking, based on weld composition, has been compared with empirical equations, such as:

$$\begin{aligned} UCS &= 230\ \mathrm{C} + 190\ \mathrm{S} + 75\ \mathrm{P} + 45\ \mathrm{NB} \\ &\quad - 12.3\ \mathrm{SI} - 5.4\ \mathrm{MN} - 1 \end{aligned} \qquad \textbf{(EQ 1)}$$

where the elemental values are given in percent. If the *UCS* value is less than 10, then the susceptibility to cracking is low, whereas a value higher than 30 means that this susceptibility is high, and a value between 10 and 30 means that the welding procedures become controlling.

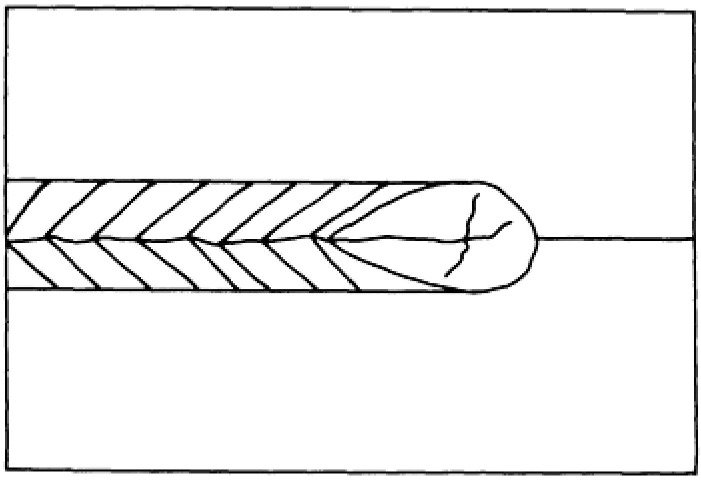

FIG. 4 HOT CRACK AND CRATER CRACKS IN WELD BEAD

Such defects are more likely to occur with welding processes or procedures that produce high dilution (that is, deep penetration). Another factor that contributes to centerline cracking is a sharp tear-drop profile of the weld crater, which is characteristic of high welding speeds. In these situations, the weld crater often develops shrinkage cracks, called crater cracks. Both the teardrop crater and deep penetration are produced with the submerged arc welding (SAW) process and the GMAW process using carbon dioxide shielding. The problem also may occur in fillet welds that are very concave, because their cross section might not be sufficient to tolerate the transverse stresses that are due to weld shrinkage.

In most cases, the problem can be prevented by keeping the combined phosphorus and sulfur levels below 0.06%. However, when welding highly restrained joints using high-strength steels, a combined level below 0.03% might be necessary. When the steels to be welded contain excessive amounts of sulfur or phosphorus, hot cracks can be avoided by:

- USING WELDING PROCESSES OR PROCEDURES THAT ARE NOT DEEPLY PENETRATING
- SELECTING TRAVEL SPEEDS THAT ARE SUFFICIENTLY SLOW TO PREVENT THE FORMATION OF TEAR-DROP CRATERS
- PROVIDING CONVEX BEAD PROFILES
- FILLING THE CRATERS AT THE END OF EACH BEAD

Lamellar tearing occurs in the base plate when stressed through its thickness and generally is found just below the HAZ (Fig. 5). It is associated with banded steels that contain thin layers of inclusions located beneath the plate surfaces. If the dirty steels must be used, the problem can be prevented by changing the joint design to minimize strain through the thickness of the plate at the weld.

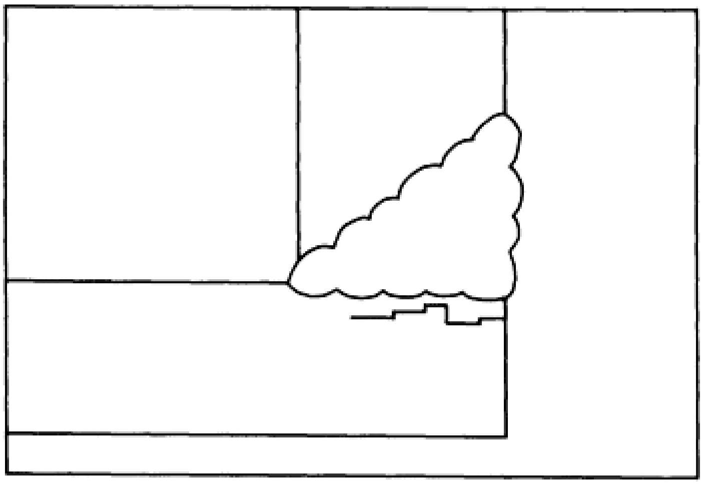

FIG. 5 LAMELLAR TEARING

An undercut is an irregular gouge that is generally found in the upper toe of a horizontal fillet weld. The base plate in that section of the weld shown on the left side of Fig. 6 had been melted by the arc, but not refilled by the weld metal. Most often, this defect is caused by improperly selected welding conditions such as the electrode angle, travel speed, and welding current. It is more likely to occur when attempting to make fillet welds with legs that are more than 8 mm (0.3 in.) in length. With the GMAW process, it also can occur when using an argon shield containing less than 2% oxygen. Undercut also can be found in welds made in the vertical position, where it is generally attributed to excessive weaving.

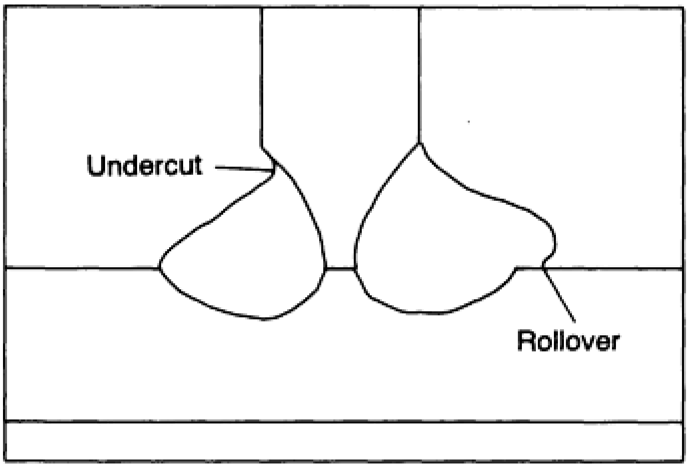

FIG. 6 UNDERCUT IN VERTICAL LEG OF FILLET WELD AND ROLLER IN HORIZONTAL FILLET WELD

A rollover, also called an overlap (Fig. 6), generally is associated with fillet welds and can be found when either the welding current is too low to properly fuse the base plate or the travel speed is too low to accept the amount of metal being deposited. Poor manipulation of the electrode during the SMAW process also can be a factor.

Inclusions are produced by slag that is trapped between weld passes. They originate as pieces of unfused fluxes that may be trapped in a joint, or as slag that is allowed to flow ahead of the arc and is covered by the weld, or as solidified slag that has not been removed between weld passes, or as heavy mill scale that has not been removed from a joint prior to welding. The problem is most common with the SMAW process because it can be aggravated by poor manipulative techniques on the part of the welder. The presence of inclusions can be anticipated when welding over highly crowned or rough welds because their edges are difficult to clean between passes or to penetrate during welding.

Prevention is possible by:

- TRAINING WELDERS TO DEPOSIT WELDS THAT HAVE PROPERLY FLAT PROFILES
- POSITIONING WELDS TO ALLOW HIGHER ENERGY AND MORE FLUID DEPOSITS TO BE MADE
- PREVENTING THE DEVELOPMENT OF RUST BETWEEN PASSES
- ENSURING THAT THE WELDS ARE PROPERLY CONDITIONED BETWEEN PASSES BY EITHER CLEANING OR GRINDING

Influence of Welding on Steel Weldment Soundness

A. Lesnewich

Hydrogen-Induced Cracking

Hydrogen-induced cracking (HIC) is a phenomenon primarily associated with welds in low-alloy steels. Figure 7 depicts a typical crack, which was taken from the HAZ of a weld. The factors that contribute to HIC are the presence of hydrogen, high tensile stress, susceptible microstructures, temperatures between approximately 200 and -100 °C (400 and -150 °F), and time. At lower strength levels (approximately 485 MPa, or 70 ksi) HIC is most often observed as longitudinal cracks in the HAZ of the base metal, often called underbead cracking. At higher strength levels (approximately 830 MPa, or 120 ksi, and higher), transverse cracks can occur in the weld metal as well.

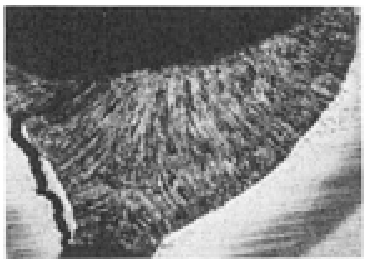

FIG. 7 PHOTOMICROGRAPH OF HYDROGEN-INDUCED CRACK IN HAZ OF WELD. 5×

The commonly used expression "hydrogen embrittlement" suggests that hydrogen impairs the toughness of welds, but that designation is a misnomer. Impact tests on material removed from the area between cracks show that the material exhibits levels of toughness equivalent to welds made in the absence of hydrogen and, of course, cracks. However, tensile ductility can be reduced because HIC can occur while the tensile test is in progress, which reduces the cross-sectional area of the test specimens. The resulting defects in the fractured surfaces are called "fisheyes."

Cold cracking is another expression that has been used to differentiate these cracks from the hot cracks that are found in weld metal and are produced by low-melting constituents that segregate during solidification. Delayed cracking is another term that has been used. It is descriptive because HIC might not occur for days or weeks. When HIC is anticipated, welds often are not radiographed for a week or more in order to allow the cracks to develop.

Mechanism. Hydrogen is a ubiquitous contaminant in all arc welding processes. It exists in the water that is unavoidable in fluxes, in the organic lubricants on the surfaces of filler wires, in the debris that collects in weld joints, and in the moisture in air that can be aspirated into the arc stream. Hydrogen has a much higher solubility in molten iron than it does in solid iron, and its solubility decreases with temperature in the solid state as well. The hydrogen solubility in iron as a function of temperature is shown in Fig. 8. The solubility above the liquidus at 1500 °C (2730 °F) is about 30 ppm by weight, but about 8 ppm in the solid state. At 400 °C (750 °F), its solubility drops to less than 1 ppm. The rates of weld-metal solidification are very high and, as a result, the hydrogen that has dissolved in the fused weld metal is retained.

Although that which does escape as gas often is trapped in the form of small bubbles or weld-metal porosity, a substantial amount does remain in the solidified weld metal as supersaturated hydrogen. The residue might appear to be insignificant, but it should be recognized that as little as 1 ppm of hydrogen can cause cracking problems in high-strength steels.

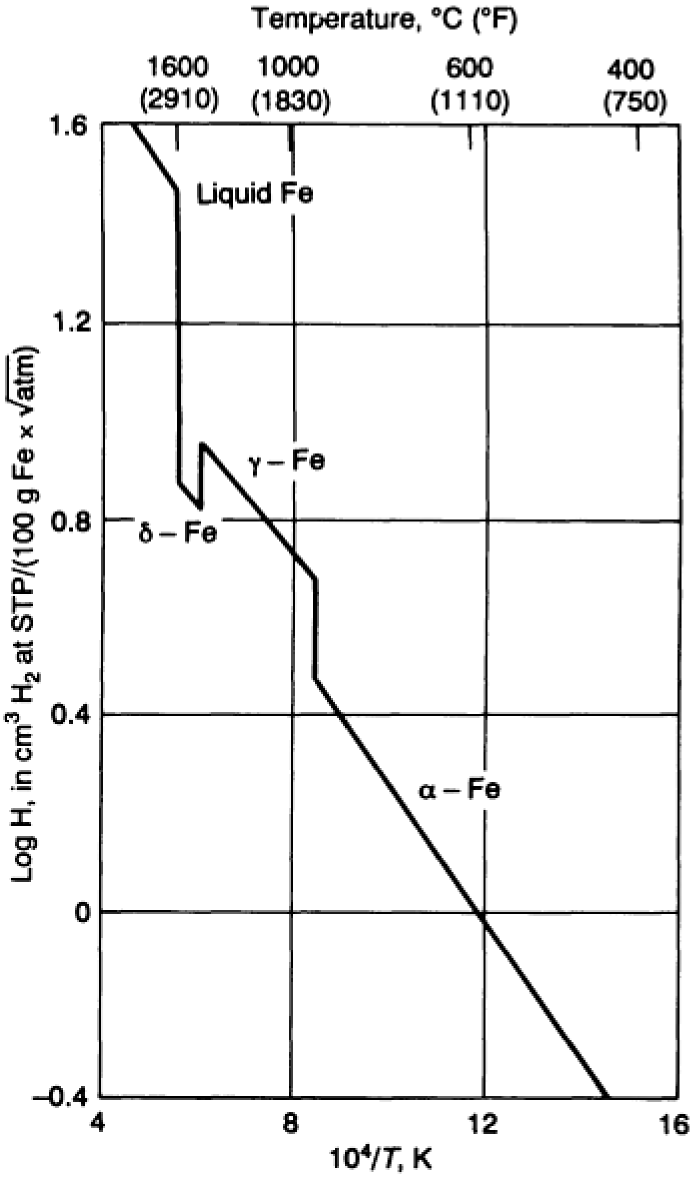

FIG. 8 EFFECT OF TEMPERATURE ON THE SOLUBILITY OF HYDROGEN IN IRON

During the cooling interval, the atomic hydrogen diffuses rapidly, with some going into the weld HAZ, some escaping to the air, and the rest remaining within the weld metal. Given the right conditions, these highly mobile atoms seek rifts and discontinuities in the metal lattice and concentrate at those points. In concert with the residual stresses in the lattice that are due to external restraint and volume changes caused by solidification and solid-state transformation, the hydrogen enlarges the discontinuities to form microcracks. The localized stresses are relieved abruptly as the atoms penetrate the fissures and become trapped as molecules. The resulting microcracks, which have sharp tips, also are associated with high stress concentrations at which additional atoms collect. Those stresses build up until they, too, are relieved as the crack is extended. This process of stress build up and relief by cracking continues until:

- THE CROSS-SECTIONAL AREA IS REDUCED SUFFICIENTLY TO CAUSE FAILURE.
- HYDROGEN ESCAPES IN A SUFFICIENT AMOUNT TO LOWER ITS CONCENTRATION BELOW THAT WHICH IS NEEDED FOR CRACKING TO PROCEED.

- UNDERBEAD CRACKS HAVE REDUCED THE RESIDUAL STRESSES IN THE WELD BELOW THAT LEVEL NEEDED FOR CRACKING TO PROCEED.

HIC does not occur spontaneously, but as discrete steps. The stepwise progression can be observed acoustically. In small specimens, its progression also can be monitored by measuring changes in resistance, as shown in Fig. 9. The lower portion of this illustration depicts the changes in resistance that occur after the process of HIC begins, as well as the way in which HIC progresses a step at a time until failure. The upper portion of Fig. 9 illustrates the sensitivity of HIC to the level of external stress. Failure occurs quickly when the stress on a specimen exceeds its tensile strength, whether or not hydrogen is present. However, when enough hydrogen is present, damage that is caused by HIC can be initiated with stresses well below the tensile strength. Given enough hydrogen and time, HIC can cause failure. Note that the time needed for initiating the cracks and leading to failure increases as the stress is reduced.

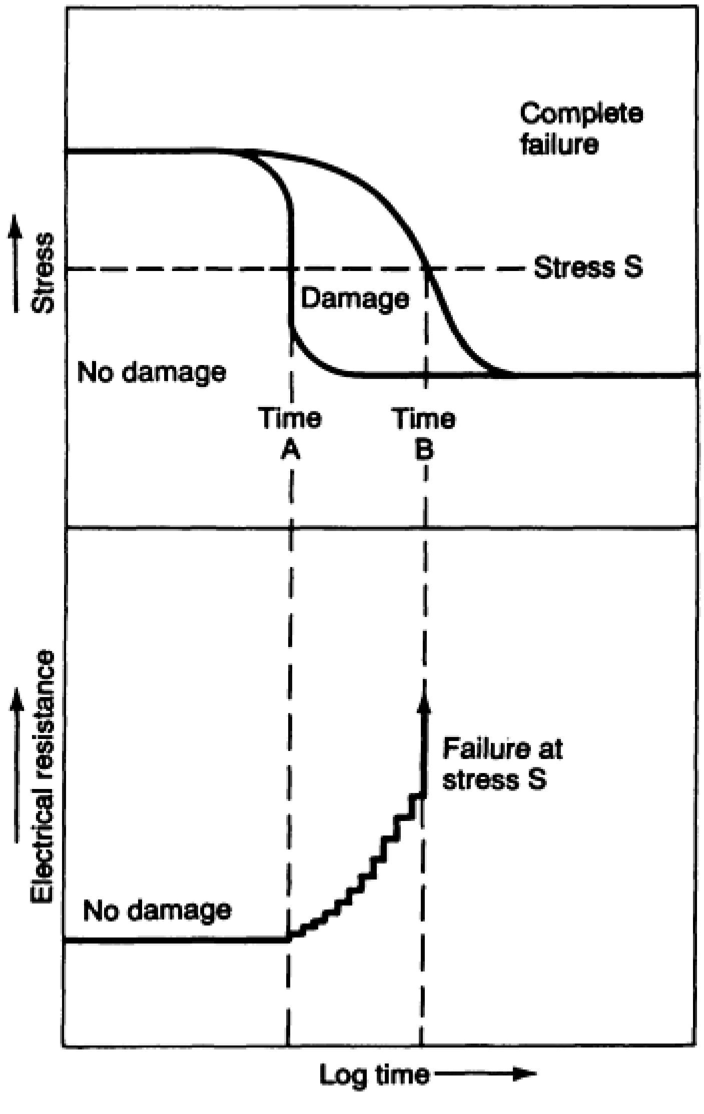

FIG. 9 DAMAGE CAUSED BY HIC, AS AFFECTED BY APPLIED STRESS. IN SMALL SPECIMENS, ELECTRICAL RESISTANCE INCREASES AS HYDROGEN-INDUCED CRACKS GROW.

Most important is the observation that HIC does not occur below a critical stress. In addition to the applied stress, the amount of hydrogen dissolved in the steel also plays an important role. With increasing hydrogen, less stress is needed to initiate HIC, and the times required for its initiation are reduced, as well. The interaction of these two variables, stress and hydrogen, on HIC is shown in Fig. 10. Note that both the time to initiate HIC and the critical stress below which failure will not occur are inversely proportional to the amount of hydrogen present in the steel.

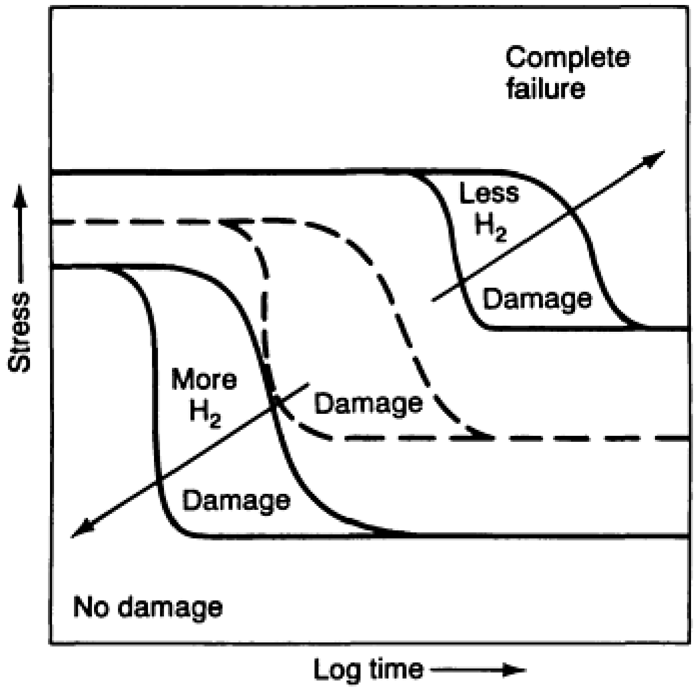

FIG. 10 EFFECT OF HYDROGEN AND STRESS ON DAMAGE CAUSED BY HIC. AS HYDROGEN INCREASES, LESS STRESS IS NEEDED TO CAUSE DAMAGE BY HYDROGEN-INDUCED CRACKS.

The third variable that affects HIC is the microstructure of the steel (either the weld metal or the HAZ). Twinned martensites, which occur in higher-carbon-content steels (above approximately 0.3% C), are the most troublesome, although the problem can occur with all acicular microstructures, including the bainites. This hypothesis might be flawed because the acicular microstructures are typical of those associated with high-strength steels, and higher stresses, in themselves, are an aggravating factor with HIC. However, as shown in Fig. 11, a steel with a relatively tolerant microstructure can exhibit a higher critical stress than a stronger steel with a sensitive microstructure. Note that the stronger steel is more sensitive to hydrogen with regard to both an earlier initiation time for HIC and a lower critical stress. Such behavioral differences have been observed between high-strength martensitic and weaker bainitic steels.

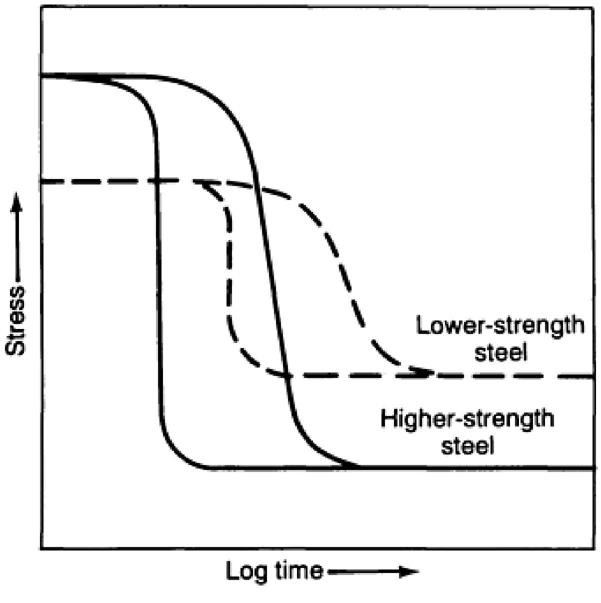

FIG. 11 EFFECT OF STEEL STRENGTH ON SUSCEPTIBILITY TO HIC DAMAGE. STRONGER STEELS CAN BE MORE SENSITIVE TO HYDROGEN, RESULTING IN LOWER CRITICAL STRESSES THAN THOSE EXHIBITED BY WEAKER

STEELS, WHICH ARE MORE TOLERANT.

Inclusions also are important. The toughness of HSLA steels is impaired by contaminants, particularly when in the form of inclusions. However, because inclusions can serve as sinks for hydrogen atoms, they also can have a beneficial effect. For this reason, some very high-purity steels have been shown to be surprisingly sensitive to HIC.

One should not conclude that welds need to be stressed externally in order to develop HIC. The differential shrinkage associated with fusion welding always produces residual stresses in weldments, and, with rare exceptions, those stresses are at least equivalent to the yield strength of the weakest component of the joint. Because most weld metals are stronger than the base metal, the residual stresses are close to the yield strength of the base plate. Often, it is possible to minimize the development of HIC in critical structures by selecting a weaker, or undermatching, weld metal to keep the residual stresses as low as possible. With some applications, such as those involving fatigue, a weaker but sound structure could be more suitable than one that contains HIC. However, given a sensitive microstructure and enough hydrogen, the critical stress can be very low, that is, substantially less than typical residual stresses. Therefore, if HIC is a problem, it will develop, in most cases, before a welded structure leaves the fabrication shop.

Another important observation that needs to be noted is that the mechanism of hydrogen-induced cracking is affected by temperature. The chances for HIC are minimal when temperatures exceed about 205 °C (400 °F). At higher temperatures, the hydrogen diffusion rates are too high to allow the atoms to concentrate at lattice defects or other sharp discontinuities in the weld. Because the mobility of hydrogen is essentially nil, HIC is unlikely to occur in welds cooled below -130 °C (-200 °F). The relevance of the different behaviors at temperatures above 205 °C (400 °F) will be discussed later.

Control of Hydrogen-Induced Cracks. In reviewing the metallurgical requirements for HIC, it is evident that a number of approaches might be taken to prevent its occurrence. They include reducing the residual stresses associated with a weldment; either avoiding acicular microstructures in the weld metal and HAZ or at least selecting those that are bainitic, rather than martensitic; reducing the amount of hydrogen dissolved in the weld metal during the welding operation; or allowing hydrogen to escape before it can cause damage. The most appropriate of these options will depend on the size of the component to be welded, the mechanical properties required, the service anticipated, the welding process to be used, and any cost constraints. In most cases, compromises have to be made and a combination of these approaches probably will prove to be most cost effective.

As mentioned earlier, the residual stresses in welds are generally equivalent to the yield strength of the weakest material in the joint. In joint configurations that introduce high triaxial stresses, the residual stresses can be significantly higher than the yield strengths. Although designers rarely use weaker materials just to reduce residual stresses, they should recognize that HIC can have a significant effect on the fatigue life of a structure. To accommodate weaker steels, a more-acceptable compromise might be to redesign the weldment to incorporate thicker sections. However, other approaches can be taken to gain full advantage of the strengths available in low-alloy steels without incurring HIC.

For the reasons mentioned earlier, the possibility of changing the microstructure of the weld metal or the HAZ is remote, unless given the option of selecting different alloys. In that case, the alloy that is most forgiving of HIC should be chosen.

Another method for reducing the residual stresses in welds is by using a postweld heat treatment at temperatures below the critical temperature. Because steels are weaker at higher temperatures, a significant reduction of residual stresses is possible by heating welds to a temperature at which plastic yielding can occur. For a steel with a tempered martensite structure, the most suitable choice for this heat treatment would be at or just below its original tempering temperature, which is generally close to 620 °C (1150 °F). This treatment is called stress relief annealing (SRA). For this treatment to be effective, the weldment must be placed in a suitably large furnace before its temperature drops below 205 °C (400 °F) and then, to prevent difficulties associated with distortion, heated and cooled slowly. Considering the temperatures and times needed for the SRA treatment, it is obvious that all of the diffusible hydrogen in the weld will escape. However, unless the stresses in a weld must be relieved for reasons other than the avoidance of HIC, SRA can prove to be a very expensive option.

Postheating does have a place in the scheme of preventing HIC. It is not necessary to reheat weldments to temperatures that are much higher than 205 °C (400 °F) in order to accelerate the escape of hydrogen and still avoid the temperature range within which HIC is likely. Such thermal treatments are excellent candidates for welded components that are small enough to be preheated in a furnace prior to welding and returned to the furnace immediately after welding for a period of time that allows all of the hydrogen to escape. This approach is particularly important for very high strength alloys, such as SAE 4340, that are very sensitive to cracking problems associated with hydrogen.

Similar results are possible by slowing the rates at which welds are allowed to cool after welding. This provides more time for hydrogen to escape before temperatures drop below 205 °C (400 °F). Retarding the cooling rate also allows the transformation of austenite to softer microstructures that are less sensitive to HIC.

The cooling rates of arc welds are affected primarily by three factors: the temperature of the joint before welding begins, the arc energy input during welding, and the joint thickness. The initial temperature can be that of the shop in which the steel had been stored, or the temperature to which the weldment had been heated as the result of a previous weld by external methods (the interpass temperature), or the temperature to which the joint had been heated (the preheat temperature). As preheat temperatures are increased, the cooling rates decrease. The arc energy input is defined by the electrical energy dissipated by the arc and the speed at which the arc is moved along the joint. Higher arc energy inputs retard the cooling rate.

The joint thickness also affects cooling rates because much of the heat entering the joint is extracted by conduction into the body of the weldmerit. Conduction is at a maximum with three-dimensional cooling. This occurs when the joint is thicker than about 25 mm (1 in.). Conduction is less effective in thinner sections, that is, the weld cooling rate is inversely proportional to thickness. Although the cooling rates of thin sections are also influenced by radiation and convection, the effect is much less pronounced than that of conduction.

The variables described above can be incorporated into a single equation that allows calculations to be made of the rate at which welds cool at a specific temperature:

$$CR_T = C\frac{(T - T_0)^2}{E} \qquad \textbf{(EQ 2)}$$

where CR_T is the cooling rate at temperature T, C is a constant of proportionality (including an adjustment for plate thickness, if it is thinner than 25 mm, or 1 in.), T_o is the preheat or interpass temperature, E is the arc energy input, which is calculated as:

$$E = \frac{VI}{S} \qquad \textbf{(EQ 3)}$$

where V is the arc voltage, I is the welding current, and S is the arc travel speed. By combining the above two equations, a general expression for cooling rate is obtained:

$$CR_T = C\frac{(T - T_0)^2 S}{VI} \qquad \textbf{(EQ 4)}$$

This equation was developed for the purpose of predicting weld and HAZ microstructures in conjunction with continuous-cooling transformation diagrams. These diagrams allow the determination of the cooling rates above which strong martensites or bainites are ensured or below which they can be avoided. The same equation can be used to calculate the cooling rates at temperatures critical to the evolution of hydrogen and the avoidance of HIC.

The adjustment of welding procedures is accomplished by varying the current or the travel speed. Voltage is a strongly dependent variable that is determined by the welding process; the characteristics of the electrodes, fluxes, or shielding gases; and the current. It should not be viewed as a variable with which to control weld cooling rates.

The other method of retarding cooling rates, which is probably the most common method, is to control the preheat or interpass temperature of the joint prior to welding. Relatively small changes in those temperatures can exert strong effects on cooling rates at temperatures around 205 °C (400 °F), which is critical with regard to the onset of HIC. For example, by increasing the preheat temperature from 20 to 95 °C (70 to 200 °F), the cooling rate at 205 °C (400 °F) is reduced by approximately one third. By preheating to 150 °C (300 °F), the cooling rate is reduced by a factor of approximately ten, which is a very significant amount when fabricating high-strength steels that have little tolerance to HIC.

Unfortunately, preheating is expensive, it can affect the weld microstructures, and it can make working conditions unbearable for welders. Nevertheless, as Fig. 12 shows, a preheat is important for reducing HIC. The data plotted in Fig. 12 clearly show how a preheat affects the lower critical stress in the HAZ of HY-80 steel when welded with an E11018-M covered electrode. The ultimate tensile stress of this heat of HY-80 is approximately 760 MPa (110 ksi). However, with a 25 °C (78 °F) preheat, that is, room temperature, failure was caused by HIC in less than 10 min, at a stress level of only 480 MPa (70 ksi). The lower critical stress below which failure did not occur was only 415 MPa (60 ksi). By preheating to a temperature of 120 °C (250 °F), the critical stress was increased to 620 MPa (90 ksi), which is approximately the yield strength of the HY-80 steel, but still considered unsafe. To avoid HIC entirely, under the conditions used to produce the welds shown, the preheat temperature would need to exceed 150 °C (300 °F).

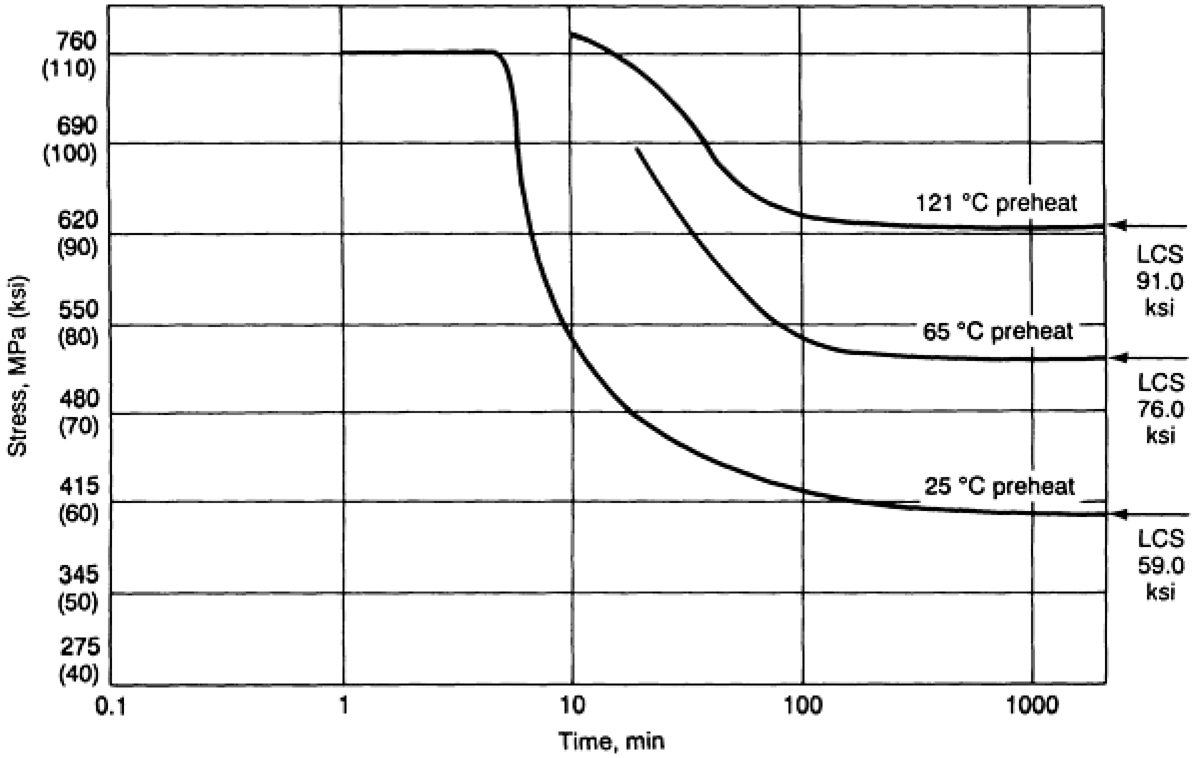

FIG. 12 EFFECT OF PREHEAT ON HIC SUSCEPTIBILITY. PREHEAT INCREASES CRITICAL STRESS NEEDED TO CAUSE HIC. LCS, LOWER CRITICAL STRESS

A number of approaches can be used to select the most appropriate temperature for preheating a steel to avoid HIC. Some approaches rely on empirically derived tables that list the steels and recommended welding procedures, including those for preheat and postheat. Another relates cracking tendencies quantitatively to the hardenability of the alloy steel, calculating it on the basis of the carbon equivalence. One such formula for carbon equivalence is given as:

$$CE = C + \frac{Mn}{6} + \frac{Si}{24} + \frac{Ni}{40} + \frac{Cr}{5} + \frac{Mo}{4} \quad \textbf{(EQ 5)}$$

For applications that involve welds made with covered electrodes, the recommended preheat temperatures for steels having different carbon equivalencies are plotted in Fig. 13. Although considerable scatter is seen, the overall trend demonstrates a linear relationship between the carbon equivalence and the preheat temperature. For a quick approximation of the required preheat, the following relationship can be used:

$$T_O = 200\ CE \quad \textbf{(EQ 6)}$$

where T_o is in °C.

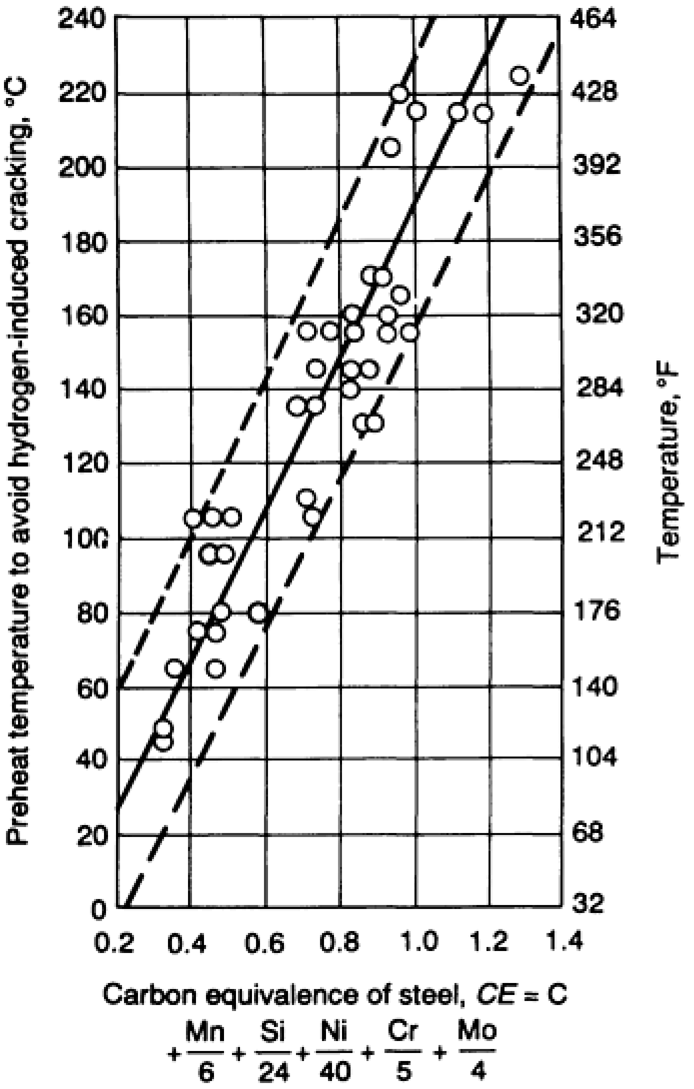

FIG. 13 EFFECT OF CARBON EQUIVALENT ON PREHEAT REQUIREMENTS. AMOUNT OF PREHEAT NEEDED TO PREVENT HIC IS PROPORTIONAL TO THE CARBON EQUIVALENT OF THE STEEL BEING WELDED

To include the scatter band that incorporates all of the data points, a more-precise interaction between the carbon equivalence and the preheat temperature can be shown as:

$$T_O = 210\ CE + 15/-45 \qquad \textbf{(EQ 7)}$$

The scatter band of 60 °C (108 °F) is quite large, which suggests that the upper portion be used for selecting suitable preheat temperatures with which to avoid potential problems. However, if metallurgical softening needs to be avoided, then the most appropriate course of action would be to rely on laboratory trials for determining the minimum effective level of preheat. Of course, such a determination would require that the energy input, the thickness of the joint, and the welding process also be considered.

Measurement of Hydrogen. Direct measurements of hydrogen in weld metal are difficult. Unless great care is taken to stop its escape from a weld before an analysis can be made, the amount found will not be representative of that which might have caused a crack to develop. This means that specimens must be designed to be analyzed quickly or supercooled in liquid nitrogen to stop the diffusion of hydrogen while awaiting analysis. The procedure recommended by the American Welding Society (AWS A4.386) measures the volume of hydrogen gas that escapes from a test weld approximately 75 mm (3 in.) in length. It is collected in either a eudiometer tube (in a mercury or glycerine bath) or in the isolation chamber of a gas chromatograph.

Indirect methods also have been used by measuring the sources of the hydrogen. For wires used in the GMAW and SAW processes, this can be done by measuring the hydrocarbons on their surface. Mass spectrometry can be used for the analysis. For the SMAW and SAW processes, the moisture adsorbed in the fluxes can be determined. Often, this is done by measuring weight loss after drying at high temperatures of approximately 400 to 425 °C (750 to 800 °F). The problem with indirect measurements is that the efficiency of transfer of the hydrogen to the weld from the wires or fluxes is difficult to predict. It can be very dependent on the welding procedures. Therefore, empirical results are used to relate the amount of hydrogen present in the welding materials to the HIC in the weldment. For this reason, a comparison among processes becomes very difficult. However, even the measurements of gas evolution can be faulted, because only the diffusible hydrogen is measured. Some remains in solution and some is trapped within weld defects or inclusions.

Importance of Welding Processes. The arc welding processes require a source of filler metal and methods for protecting and controlling the arc and the deposited metal. In almost all cases, the filler metals are provided in the form of rods, continuous wires, or continuous tubes. The surfaces of all of these materials are contaminated with residues of hydrogen-rich drawing lubricants. In the GMAW process, a shield gas is used for protection. For cored wires, a combination of shield gases and fluxes are used. The submerged-arc and covered-electrode methods involve only fluxes. All of the fluxes are sources of chemically combined or adsorbed water. The quantity of hydrogen dissolved in weld metal can vary, not only between but within processes. Some indication of the variability to be expected is shown in Table 1.

TABLE 1 EFFECT OF WELDING PROCESSES AND ELECTRODES ON HYDROGEN LEVELS IN WELDS

TYPE OF ELECTRODE	DIFFUSIBLE H_2, ML/100 G
GMAW SOLID WIRE	0-10
BASIC SMAW ELECTRODE	3-20
ACID SMAW ELECTRODE	20-40
SAW SOLID WIRE AND FLUX	4-20
GMAW CORED WIRE	3-30

Of all of the arc welding processes using consumable electrodes, the GMAW process is associated with the lowest hydrogen levels, the primary source being residual drawing lubricants on the wire surface. Totally dry wire is unacceptable, because it is difficult to feed. The amount of residual lubricant generally is not a problem with steels having yield strengths below 520 MPa (75 ksi). However, as the yield strengths approach 620 MPa (90 ksi), the residual lubricants become a potentially important factor if HIC is to be avoided, unless relatively high preheat temperatures can be used. When the yield strengths exceed 830 MPa (120 ksi), the residual lubricants must be kept as low as possible.

The importance of such residues is shown in Fig. 14, which summarizes the effects of hydrogen on HIC in welds that have a yield strength of 930 MPa (135 ksi) and shows how it can be minimized by controlling cooling rates. In this case, the cooling rates indicated were determined at 540 °C (1000 °F), a temperature close to that at which the weld metal transforms from austenite to martensite. The closed circles indicate the presence of HIC in the welds, whereas the open circles indicate sound welds. At the relatively rapid cooling rates of 28 °C/s (50 °F/s), 4 ppm of hydrogen on the wire surface is shown to have caused HIC. To be securely free of HIC, the hydrogen had to be maintained at levels below 3 ppm. By adjusting the welding procedures, preheat temperatures, or both, in order to retard the cooling rate at 540 °C (1000 °F) to less than 20 °C/s (35 °F/s), the tolerance for hydrogen on the wire was increased to 5 ppm.

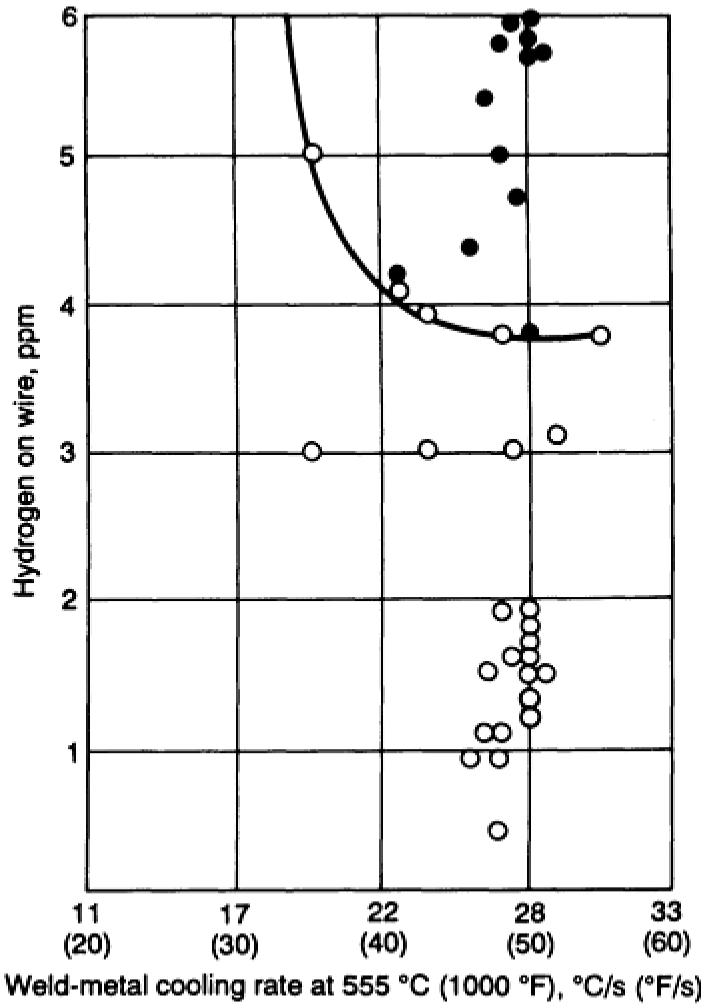

FIG. 14 INFLUENCE OF HYDROGEN PRESENT ON A WIRE SURFACE AND WELD-METAL COOLING RATE ON HIC. DATA ARE BASED ON GMAW WELD METAL WITH YIELD STRENGTH OF 930 MPA (135 KSI).

As mentioned earlier in the discussion of hydrogen measurements, it is difficult to predict the amount of hydrogen that will transfer to a weld from surface contaminants that are decomposed in the arc (or before reaching the arc), particularly when the levels are measured in single-digit parts per million. These levels are so low as to preclude the use of gas-evolution techniques to measure the hydrogen.

The higher tolerance for wire surface contaminants at lower cooling rates may be due as much to softer microstructures as it is to the escape of hydrogen. The effect of cooling rate on the tensile yield strength of these alloys is illustrated in Fig. 15. To retain high strengths, the higher cooling rates are essential. In this example, a very abrupt drop in strength is shown as the cooling rate drops below 10 °C/s (20 °F/s). Obviously, to obtain the strongest possible welds without encountering HIC, it is necessary to minimize the presence of any contaminants that contain hydrogen.

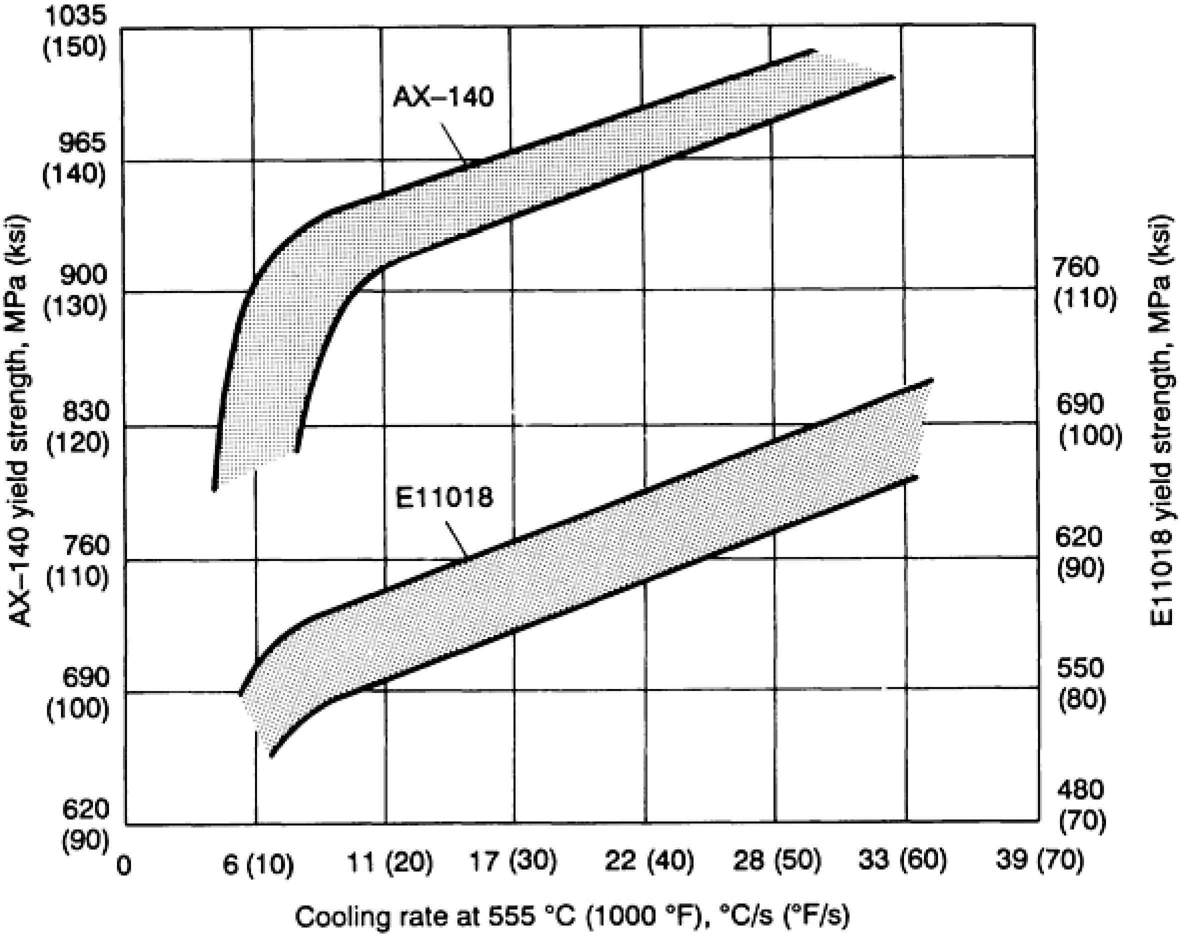

FIG. 15 EFFECT OF COOLING RATE ON STRENGTH OF HSLA STEEL WELD METAL. TENSILE YIELD STRENGTH OF HSLA STEEL WELD METAL IS AFFECTED BY THE COOLING RATE AT WHICH IT TRANSFORMS. DATA ARE BASED ON AX-140 GMAW AND E11018 SMAW FILLER METALS.

The attainment of such low levels of hydrogen is not possible with any of the other arc welding processes, because they require fluxes instead of shield gases for protection. Fluxes can absorb water. The data in Fig. 16 summarize the importance of moisture in a submerged arc flux on the cracking sensitivity of a Soviet weld metal that has a yield strength of 830 MPa (120 ksi). It demonstrates that diffusible hydrogen levels as low as 7 mL/100 g can drop the critical strength to 105 MPa (15 ksi). (Note that a hydrogen content of 1 ppm is equivalent to 1.11 mL/100 g). Even baking the flux to reduce the weld-diffusible hydrogen below 2 mL/100 g did not eliminate HIC. The critical stress remained below 415 MPa (60 ksi). It is evident that the welding conditions used for these submerged arc welds were not acceptable. Either this steel is unusually sensitive to hydrogen or the flux used is not capable of being dried sufficiently to reduce hydrogen contamination.

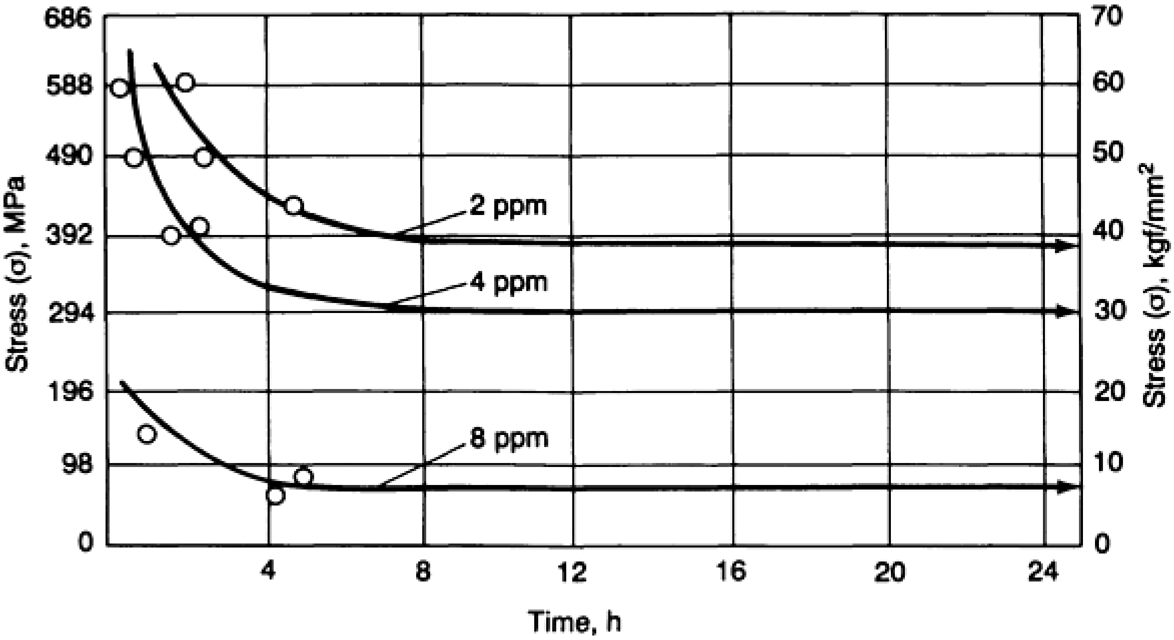

FIG. 16 EFFECT OF HYDROGEN ASSOCIATED WITH SUBMERGED ARC WELDS ON THE LOWER CRITICAL STRESS AND THE TIME NEEDED FOR HIC. FLUX AS RECEIVED, 4 PPM; FLUX BAKED, 2 PPM; FLUX MOISTENED, 8 PPM. DATA ARE BASED ON 30KH2N2M SOVIET STEEL WITH A YIELD STRENGTH OF 930 MPA (135 KSI).

Similar HIC problems are encountered in the SMAW process when weld strengths exceed 480 MPa (70 ksi). For this reason, low-hydrogen electrodes were developed specifically to minimize, if not prevent, these problems. Low-hydrogen electrode coatings are formulated without any organic materials, such as the cellulose that characterizes the EXX10 classes. These low-hydrogen coatings are baked at temperatures exceeding 425 °C (800 °F) to reduce residual moisture to a level of approximately 0.1 wt%. This is nearly the lowest practicable level, because the absence of moisture in a coating tends to make it brittle. The effect of baking on the residual moisture during initial manufacture is shown in Fig. 17. Even with careful control of formulations and baking, the moisture level of covered-electrode coatings cannot be reduced to levels sufficiently low to prevent HIC in steels having yield strengths higher than 830 MPa (120 ksi).

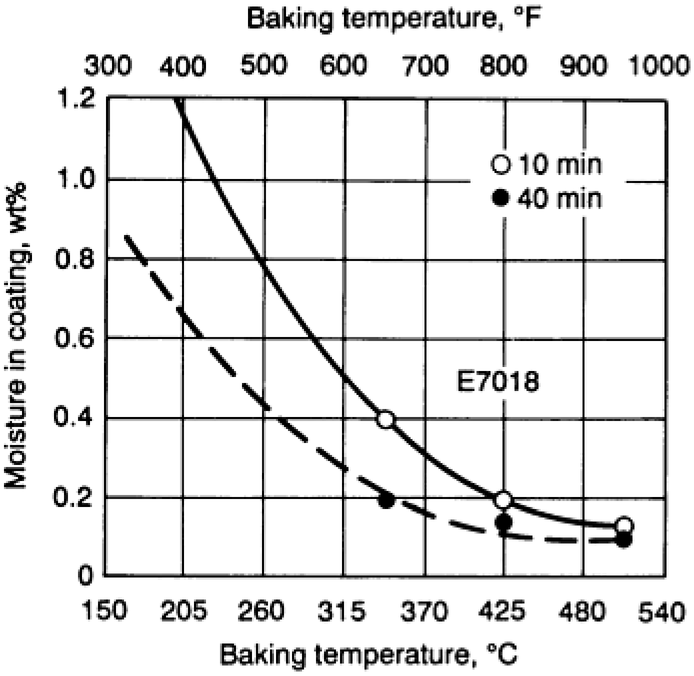

FIG. 17 EFFECT OF BAKING TIME AND TEMPERATURE ON MOISTURE IN E-7018 COVERED ELECTRODES

The moisture in low-hydrogen electrodes generally is specified as 0.2% max. This moisture level is what is expected to be found in coatings of commercial low-hydrogen electrodes, immediately after being removed from hermetically sealed containers. However, if exposed to humid, warm air, these electrode coatings will reabsorb moisture, as shown in Fig. 18. The rate of moisture pickup depends on the constituents in the coating. In some cases, reabsorbed moisture can reach levels exceeding 1%. For this reason, electrodes must be stored in heated ovens on hot and humid days and exposed to shop environments for only short times.

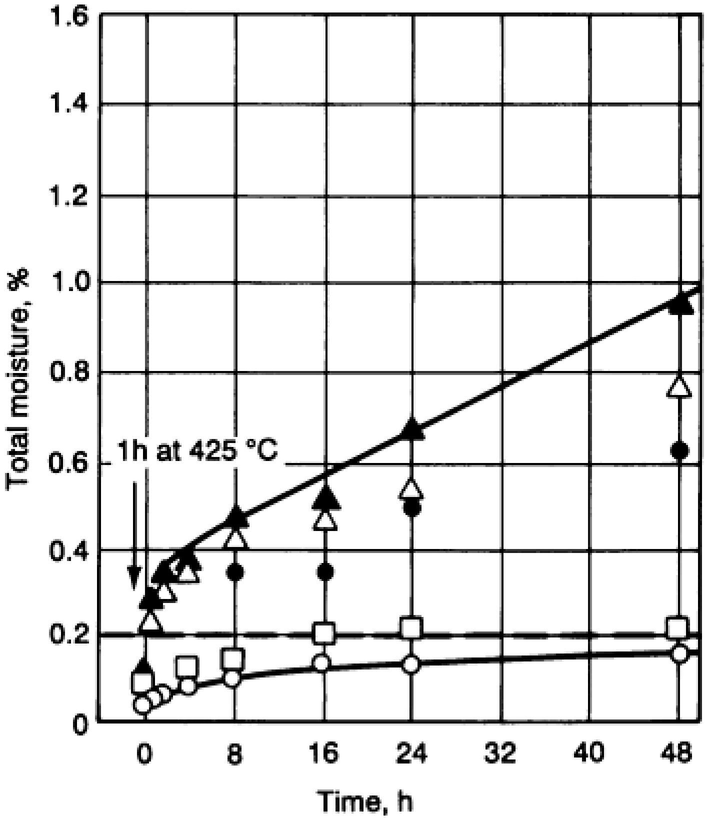

FIG. 18 MOISTURE ABSORPTION IN COVERED-ELECTRODE COATINGS WHEN EXPOSED TO MOIST, HUMID ENVIRONMENTS. THE AMOUNT ABSORBED DEPENDS ON THE ELECTRODE AND INCREASES WITH TIME. TEMPERATURE, 21 ± 3 °C (70 ± 5 °F); RELATIVE HUMIDITY, 65 ± 5%

Moisture-resistant coatings have been developed to counter the reabsorption problem. Their rates of reabsorption are typical of the lower curve shown in Fig. 13. Although they are quite safe when exposed to the relatively cool and moderately humid environment indicated, extra precautions are essential when welding in tropical conditions.

It is possible to salvage electrodes that have become "wet" by rebaking them at temperatures that approach those used during their manufacture. Typical results are shown in Fig. 19. Drying times of about 1 h are needed to recondition typical electrodes at approximately 400 to 425 °C (750 to 800 °F). Although rebaking can salvage electrodes that are inadvertently exposed to moist conditions, it should not be repeated. Covered electrodes are alloyed with metal powders that can be oxidized during rebaking operations. Therefore, the resulting alloys are leaner and weaker, as seen in Table 2.

TABLE 2 EFFECT OF REBAKING ON COMPOSITION AND STRENGTH

NUMBER OF	COMPOSITION, WT%	YIELD STRENGTH

BAKE CYCLES	C	MN	SI	MPA	KSI
1	0.069	1.54	0.39	723.3	104.9
2	0.064	1.46	0.37	714.3	103.6
3	0.066	1.45	0.37	688.8	99.9

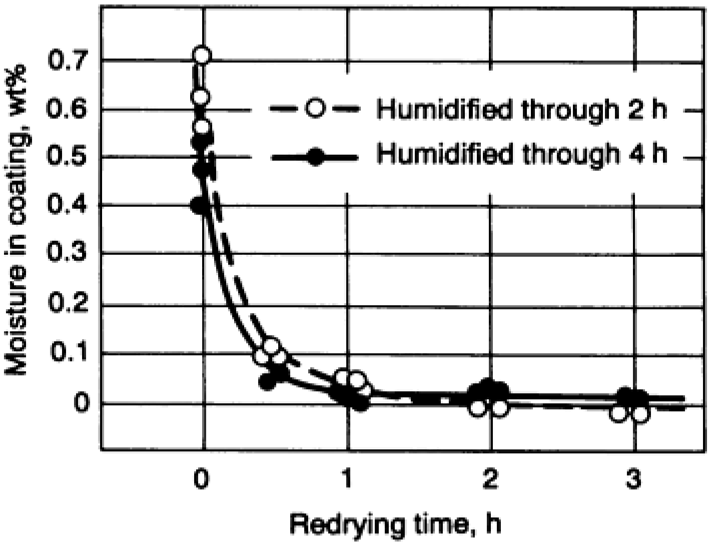

FIG. 19 EFFECT OF DRYING TIME ON MOISTURE ABSORBED BY COVERED ELECTRODES AFTER EXPOSURE TO MOIST ENVIRONMENTS. E11016 LOW-HYDROGEN TYPE ELECTRODE, 4 MM (0.16 IN.); REDRYING TEMPERATURE, 400 °C (750 °F)

Rebaking caused a loss in both manganese and silicon in the weld metal, resulting in a drop in the weld yield strength. These data were obtained after very controlled rebaking. Unfortunately, the same care is not always taken in shop environments. Significantly greater losses in the manganese and silicon contents, as well as mechanical properties, can be expected.

Influence of Welding on Steel Weldment Properties

C.C. Chen and Abe Pollack, Microalloying International, Inc.

Introduction

THE FUNDAMENTAL AND THE SPECIFIC FACTORS that control the properties of steel weldments in both the weld metal and heat-affected zone (HAZ) of carbon and low-alloy steels are described in this article. The influence of welding processes, welding consumables, and welding parameters on the weldment properties also will be discussed. The service properties of weldments in corrosive environments and subjected to cyclic loading will be considered, as well.

Heat-Affected Zone

In terms of materials selection, the service characteristics of the HAZ should be emphasized more than those of the weld metal. This is because the metallurgical and mechanical properties of the HAZ are directly related to the selected base metal, even though they can be modified by welding parameters and postweld heat treatment (PWHT). It is also true that any metallurgical and/or weldability problems associated with the HAZ characteristics will be more difficult to handle than those associated with the weld metal. Welding problems that occur in the weld metal can often be overcome by changing the welding electrode and/or other welding consumables. In contrast, problems with the HAZ sometimes can only be resolved by changing the base metal, which is often an excessively costly measure, or by changing the heat input.

Various empirical carbon equivalents have been developed and utilized to evaluate the weldability (hydrogen-induced cold-cracking tendency) of the base materials. In the United States and European countries, the International Institute of Welding (IIW) equation is most common:

$$CE(IIW) = C + \frac{Mn}{6} + \frac{Cu + Ni}{15} + \frac{Cr + Mo + V}{5} \qquad \textbf{(EQ 1)}$$

In Japan, the Ito-Bessyo (Ref 1) composition-characterizing parameter, P_{cm}, is more widely used. This carbon equivalent is considered to allow a more realistic assessment of the weldability of low-carbon steels, when compared with the IIW equation:

$$P_{cm} = C + \frac{Si}{30} + \frac{Mn + Cu + Cr}{20} + \frac{Ni}{60} + \frac{Mo}{15} + \frac{V}{10} + 5B \qquad \textbf{(EQ 2)}$$

Yurioka *et al.* (Ref 2) have also developed another carbon-equivalent number, CEN, which incorporates both the IIW carbon equivalent and the P_{cm}. The CEN is similar to the P_{cm} for steels with carbon levels that are less than 0.17%, and it follows the IIW formula at high carbon levels.

$$CEN = C + A(C)\left(\frac{Si}{24} + \frac{Mn}{6} + \frac{Cu}{15} + \frac{Ni}{20} + \frac{Cr + Mo + Nb + V}{5} + 5B\right) \qquad \textbf{(EQ 3)}$$

where A(C) equals 0.75 plus 0.25 tan*h* [20(C - 0.12)].

Although these carbon-equivalent expressions were initially developed to characterize the hydrogen cracking tendency for steel plates, they also have been utilized to assess the hardenability of steels based on their chemistry. Figure 1 illustrates the correlation between carbon equivalent, P_{cm}, and the martensite volume in both the coarse-grained region and the grain-refined region (Ref 3). As the P_{cm} increases, the martensite volume and hardness in both regions increase.

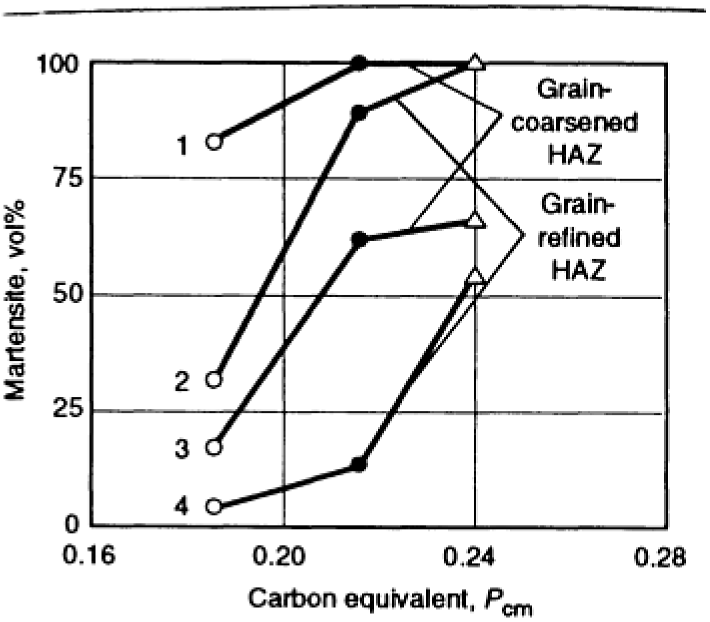

FIG. 1 RELATIONSHIP BETWEEN HAZ VOLUME FRACTION OF MARTENSITE AND THE P_{CM} CARBON EQUIVALENT

OF THERMALLY CYCLED SPECIMENS. FOUR THERMAL PROGRAMS ARE INCLUDED: (1) PEAK TEMPERATURE (T_P) ≈1350 °C (2460 °F), COOLING TIME FROM 800 TO 500 °C ($\Delta T_{8/5}$) = 3 S; (2) T_P ≈1000 °C (1830 °F), $\Delta T_{8/5}$ = 3 S; (3) T_P ≈1350 °C (2460 °F), $\Delta T_{8/5}$ = 13 S; (4) T_P ≈1000 °C (1830 °F), $\Delta T_{8/5}$ = 13 S. SOURCE: REF 3

Various regions in the HAZ of ferritic steels have been defined and characterized using a specific peak temperature of the HAZ weld thermal cycle in conjunction with an iron-carbon equilibrium phase diagram (Ref 4). These regions are the coarse-grained, grain-refined, intercritical, subcritical, and partly molten regions. However, the metallurgical behavior in the HAZ that undergoes rapid heating and cooling is a nonequilibrium process. Consequently, a continuous cooling transformation (CCT) diagram is more appropriately used to predict the microstructures in the HAZ.

The cooling rate is closely related to plate thickness, joint geometry, and welding heat input. The cooling rate that is often most critical in welding applications is that which falls between 800 and 500 °C (1470 to 930 °F). As the welding heat input is increased or the plate thickness is reduced, the cooling rate between these temperatures is reduced. (The effects of welding parameters on the HAZ cooling rate will be discussed in detail later in this article.)

The effects of major alloying elements in carbon and low-alloy steels on HAZ microstructure and toughness is summarized below (Ref 3).

Carbon has a tremendously important role in the overall hardness of welded steels. This can be clearly seen by its role in the carbon-equivalent and P_{cm}, equations. An increase in carbon level promotes the formation of lower-temperature transformation products, such as bainite and martensite, and leads to a significant reduction in HAZ cleavage resistance.

Manganese. In addition to solid-solution hardening, manganese can lower the transformation temperature of austenite to ferrite while providing strengthening effects by grain refinement. However, the level of manganese should be limited to minimize solidification segregation and microstructural banding.

Chromium is a solid-solution strengthener and carbide former. It increases the hardenability of steels and improves oxidation and corrosion resistance. The addition of chromium can be favorable in cases where the precipitation of chromium carbides suppresses the formation of ferrite side plates through a pinning effect.

Nickel is considered to have a beneficial effect on steel transformation, similar to that of manganese, by lowering the austenite transformation temperature. The addition of nickel also can improve toughness, as well as provide a solid-solution hardening effect.

Vanadium and niobium are added in small quantities in low-alloy steels to obtain the desired mechanical properties. By forming V(C,N) and Nb(C,N), the vanadium and niobium can retard the recrystallization and grain growth of austenite during rolling and normalizing. During fusion welding, especially with high heat input, V(C,N) and Nb(C,N) dissolve in the HAZ with peak temperatures greater than 1100 °C (2012 °F) and reprecipitate during slower cooling. The reprecipitation of V(C,N) and Nb(C,N) impairs the toughness of the HAZ.

Titanium, in the form of stable titanium nitride, has been found to prevent coarsening of the prior austenite grain in regions adjacent to the weld fusion boundary.

References cited in this section

1. Y. ITO AND K. BESSYO, "WELDABILITY OF HIGH STRENGTH STEELS RELATED TO HEAT AFFECTED ZONE CRACKING," DOCUMENT IX-567-68, IIW, 1968
2. N. YURIOKA, H. SUZUKI, AND S. OHSHITA, *WELD. J.*, VOL 62, 1983, P 1475-1535
3. O. GRONG AND O.M. AKSELSEN, HAZ TOUGHNESS OF MICROALLOYED STEELS FOR OFFSHORE, *MET. CONSTR.*, SEPT 1986, P 557
4. K.E. EASTERLIN, *INTRODUCTION TO THE PHYSICAL METALLURGY OF WELDING*, BUTTERWORTHS, 1983

Weld Metal

The use of a filler metal is very common in many fusion-welding processes. When selecting the proper filler metal/electrode, the primary considerations are whether or not the weld metal can be produced defect-free and whether or not the weld metals are compatible with the base metal and can provide satisfactory properties. These characteristics are determined by the:

- CHEMISTRY OF THE ELECTRODES
- DILUTION OF THE BASE METAL
- FLUX SYSTEM OR SHIELDING GAS
- WELD-POOL SOLIDIFICATION AND SUBSEQUENT COOLING AND TRANSFORMATION

The selection of the proper filler metal is not based on matching the chemistry of the base metal. Rather, it is based on matching the weld metal and base metal service properties. Using a filler metal with a chemistry identical to that of the base metal may not provide the desired results, because the microstructures of the weld metal are entirely different from those of the base metal. For most carbon and low-alloy steels, the solidification and rapid cooling rate involved in fusion welding will result in a weld metal that has higher strength and lower toughness properties than the base metal when they are of the same chemistry. Consequently, the filler metal often contains a lower carbon level than the base metal. The strength of the weld metal is not improved by increasing the carbon content, but by adding the alloying elements that provide solid-solution or precipitation strengthening and modification of the microstructures (Ref 5).

The weld-metal microstructure in carbon and low-alloy steels contains a variety of constituents, ranging from blocky ferrite, to acicular ferrite, to bainite, to martensite. The effects of alloying elements and cooling on the weld-metal microstructure can be expressed in a schematic CCT diagram (Fig. 2), whereas the effects of alloying elements on the hardness and notch toughness of weld metal are illustrated in Fig. 3.

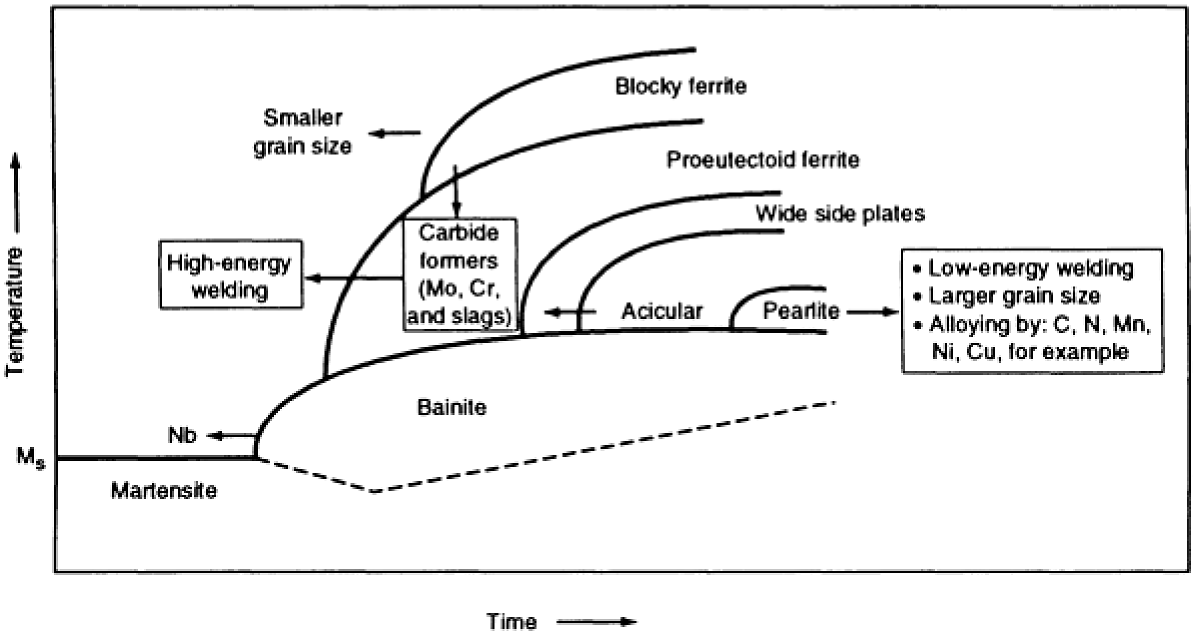
Blocky ferrite
Smaller grain size
Proeutectoid ferrite
Wide side plates
Carbide formers (Mo, Cr, and slags)
High-energy welding
Acicular
Pearlite
• Low-energy welding
• Larger grain size
• Alloying by: C, N, Mn, Ni, Cu, for example
Temperature
Bainite
Nb
M_s
Martensite
Time

FIG. 2 SCHEMATIC CCT DIAGRAM FOR STEEL WELD METAL, SUMMARIZING THE POSSIBLE EFFECT OF MICROSTRUCTURE AND ALLOYING ON THE TRANSFORMATION PRODUCT FOR A GIVEN WELD COOLING TIME. SOURCE: REF 4

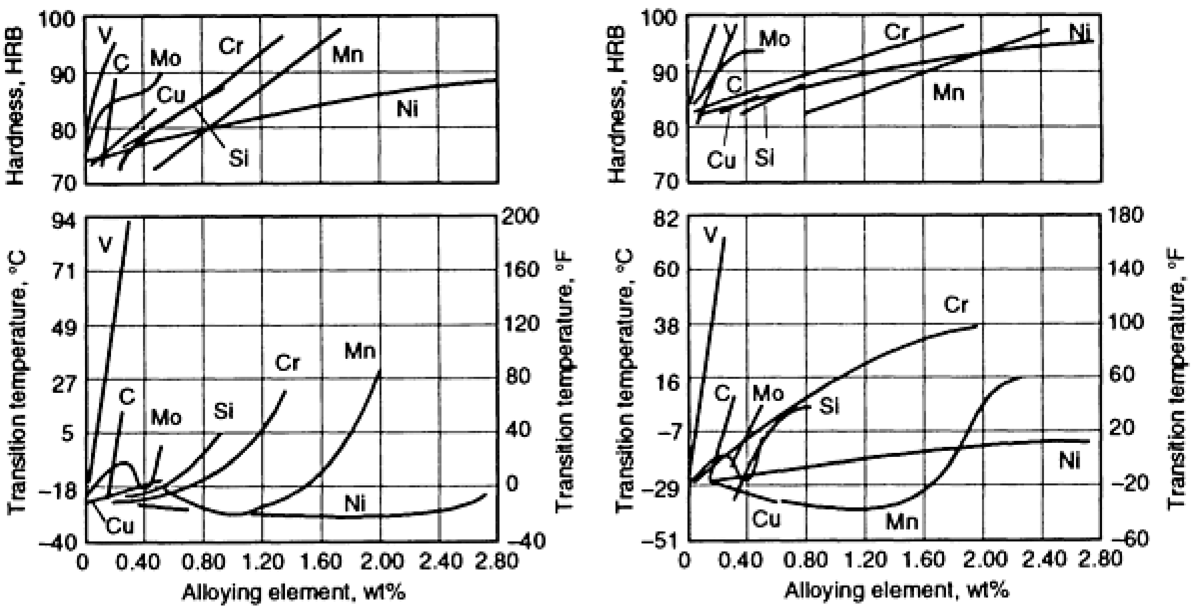

FIG. 3 EFFECTS OF ALLOY ADDITIONS ON HARDNESS AND NOTCH-TOUGHNESS OF WELD METAL. TRANSITION TEMPERATURE MEASURED AT 20 J (15 FT · LBF). (A) SUBMERGED ARC WELD METAL. (B) GAS-METAL ARC WELD METAL. SOURCE: REF 6

Dilution. The melting of base metal and the subsequent mixing with filler metal causes the final chemistry of the weld deposit to be between that of the base and filler metals. When a filler metal has the same chemistry as the base metal, the final weld-metal chemistry remains the same, theoretically. Depending on the amount of base metal that is melted and the amount of filler metal that is added, the final chemistry of the weld deposit can be approximated using the dilution equation:

$$Dilution, \% = \left(\frac{Weight \text{ of parent metal melted}}{\text{Total weight of used metal}} \right) 100 \qquad \textbf{(EQ 4)}$$

Figure 4 shows how factors such as joint configuration and edge preparation influence dilution. Welding processes and parameters that increase penetration can also increase dilution (Ref 7).

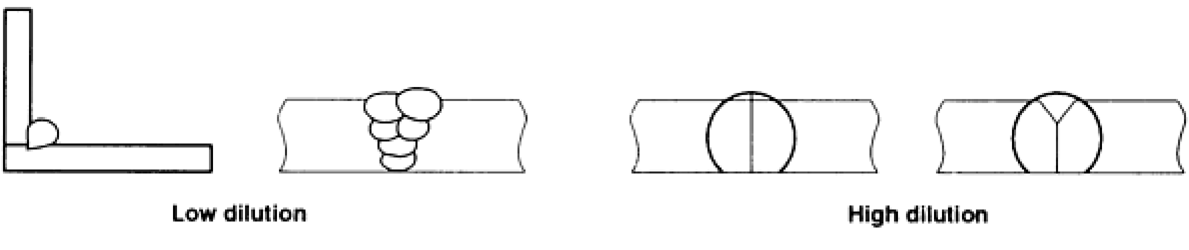

FIG. 4 DILUTION OF WELD DEPOSIT. SOURCE: REF 7

In a single-pass weld, certain element contents in the weld metal can be approximated using this formula: Specific element in weld metal equals (that element in the base metal multiplied by the dilution) plus (that element in the filler metal multiplied by [1 - dilution]).

In certain cases, greater dilution is favorable, considering the less-drastic change in mechanical and other properties across the fusion boundary. However, in applications involving weld cladding or weld surfacing, it is more desirable to minimize the dilution. This is because filler metal contains more-expensive alloying elements tailored to impart specific weld-metal properties. When there are concerns that the base metal may introduce undesired impurities and/or alloying elements, then welding practices that lower the dilution are recommended. This also includes considerations for phosphorus and sulfur and for higher carbon and silicon contents, which adversely affect the properties of weld metal.

References cited in this section

4. K.E. EASTERLIN, *INTRODUCTION TO THE PHYSICAL METALLURGY OF WELDING,* BUTTERWORTHS, 1983
5. C.F. CHOU, *THEORY AND TECHNOLOGY OF METAL WELDING,* VOL 2, MECHANICAL INDUSTRY PUBLISHER, 1981 (IN CHINESE)
6. R.D. STOUT AND W.D. DOTY, *WELDABILITY OF STEELS,* 3RD ED., S. EPSTEIN AND R.E. SOMERS, ED., WELDING RESEARCH COUNCIL, NEW YORK, 1978
7. J.F. LANCASTER, *METALLURGY OF WELDING,* GEORGE ALLEN & UNWIN, 1980

Influence of Welding on Steel Weldment Properties

C.C. Chen and Abe Pollack, Microalloying International, Inc.

Single-Pass versus Multipass Welding

In terms of mechanical properties, especially toughness, multipass weld metal is usually better than single-pass weld metal for these reasons:

- THE REHEAT THERMAL CYCLE OF EACH SUBSEQUENT WELDING PASS NORMALIZES AND REFINES PORTIONS OF THE MICROSTRUCTURE IN THE PREVIOUS WELD METAL.
- SUBSEQUENT WELD RUNS TEMPER THE PREVIOUS WELD METAL AND REDUCE RESIDUAL STRESSES FROM PREVIOUS RUNS.
- THE TOTAL ENERGY INPUT PER PASS IS REDUCED, WHICH IS HELPFUL FOR LIMITING THE AMOUNT OF GRAIN GROWTH.
- PREVIOUS WELD PASSES PROVIDE A CERTAIN PRE-HEAT THAT SLOWS DOWN THE COOLING RATE OF THE SUBSEQUENT PASS.

Figure 5 shows that the weld-metal ductility (cross-sectional area) increases as the total grain-refined areas are increased (Ref 8).

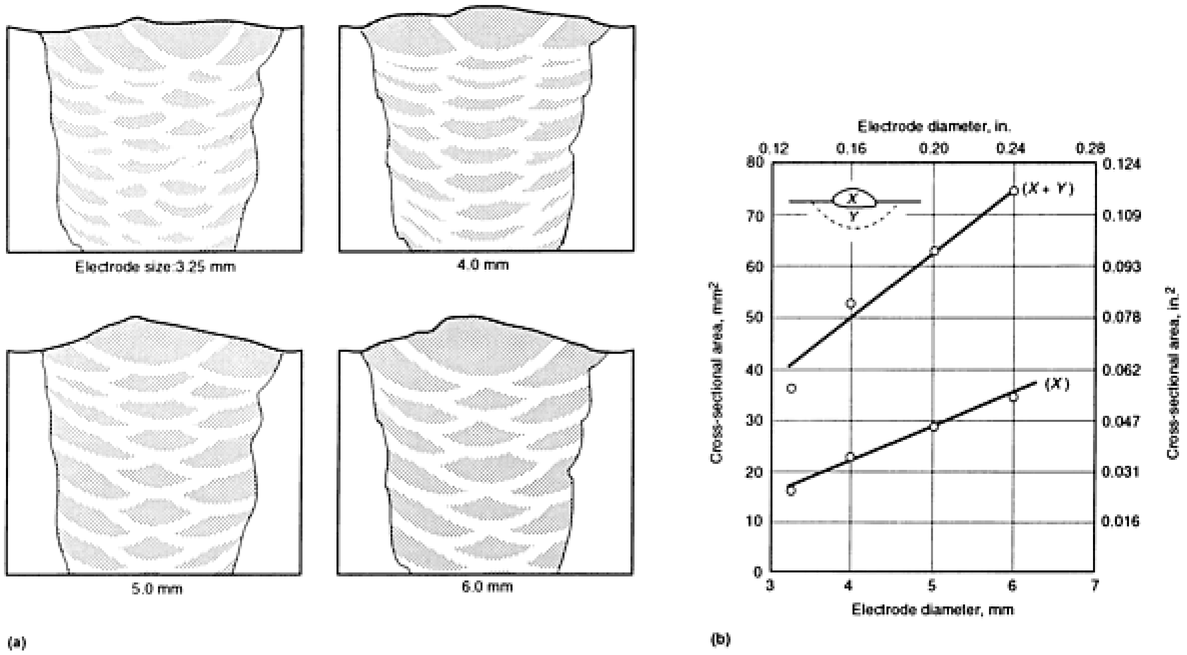

FIG. 5 EFFECT OF ELECTRODE SIZE ON AMOUNT OF RECRYSTALLIZED WELD METAL IN MULTIRUN WELDING. (A) CROSS SECTIONS AS A FUNCTION OF ELECTRODE DIAMETER; WHITE AREAS REPRESENT RECRYSTALLIZED WELD METAL. (B) GRAPHICAL REPRESENTATION. SOURCE: BASED ON REF 8

In contrast, when an active flux is used in submerged arc welding, the number of weld passes to be applied in a joint is limited. This is because there is a tendency for alloying elements (such as silicon and manganese) to build up during multipass welding, which impairs the fracture toughness values in the final weld pass.

Reference cited in this section

8. G.M. EVANS, FACTOR AFFECTING THE MICROSTRUCTURES AND PROPERTIES OF C-MN ALL-WELD METAL DEPOSITS, *WELD. REV.,* VOL 4, MAY 1982

Influence of Welding on Steel Weldment Properties

C.C. Chen and Abe Pollack, Microalloying International, Inc.

Steel Types and Weldability

The high-strength, low-alloy (HSLA) steels were designed to have improved mechanical properties and weldability, when compared with conventional carbon steels. These improvements are obtained through the addition of small amounts of alloying elements, such as niobium, vanadium, titanium, nitrogen, and carbon, which strengthen the ferrite, increase the hardenability, and control the grain size. Most of these steels are provided in the as-rolled, normalized condition, and some are provided in the precipitation-treated condition, as well. The finer grain size and greater amount of pearlite account for the greater strength and toughness of this group of steels. In addition, the lower carbon content (0.22% max) leads to good weldability. Figure 6 shows the CCT diagram for a titanium-boron HSLA steel (Ref 9).

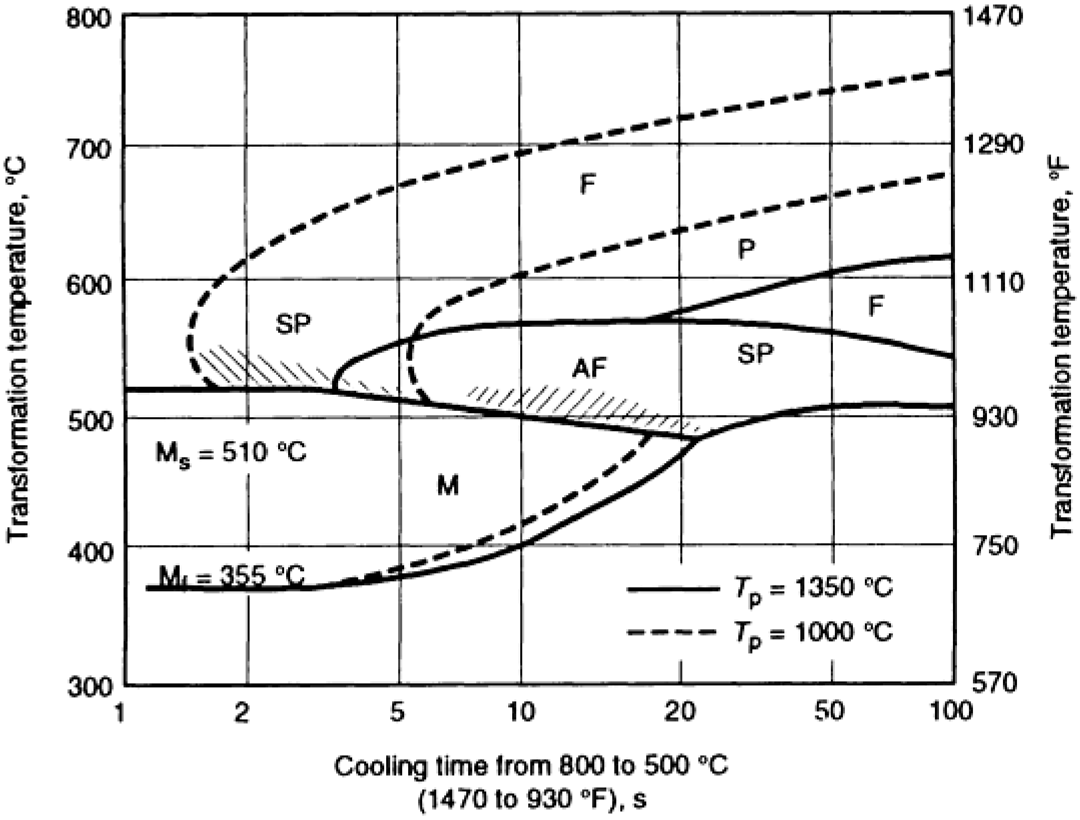

FIG. 6 CCT DIAGRAM FOR A TITANIUM-BORON MICROALLOYED STEEL. T_P, PEAK TEMPERATURE. SOURCE: REF 9

Generally, the weldability of HSLA steels is similar to that of mild steels (Ref 10). Although HSLA steels have good weldability, the consideration of preheating and controlling hydrogen in the welding process is still necessary to ensure a successful weldment. The embrittlement in the coarse-grained HAZ is one of the concerns during the welding of HSLA steels. This embrittlement is attributed to the slow cooling rate leading to the formation of undesirable microstructures, such as coarse ferrite side plates (Widmanstätten ferrite and upper bainite) and grain-boundary ferrite (Ref 3), especially at high-energy input levels.

For HSLA steels that contain niobium and vanadium, the weld metal and the HAZ will exhibit low notch toughness at high heat inputs, such as those used in the electroslag and submerged arc welding processes. This is because the high heat input increases the extent of coarse-grained HAZ. The associated low cooling rate in the HAZ reduces the scope for grain refinement and increases the likelihood of the precipitation of niobium and vanadium carbides, leading to embrittlement of the coarse-grained HAZ. The high heat input and the resulting greater dilution also brings the precipitation-hardening problem of niobium and vanadium carbides to the weld metal.

The use of steels with small additions of titanium (up to 0.04%) and nitrogen has been found to effectively minimize the grain growth in the coarse-grained HAZ (Ref 11). This is due to the greater stability of titanium nitride at the peak temperature experienced in coarse-grained regions (>1100 °C, or 2012 °F), where niobium and vanadium carbides are turned back into solution.

Low-carbon steels comprise two classes of steels, namely, low-carbon steels with a carbon content of less than 0.15% and mild steels with a carbon content between 0.15 and 0.30%.

For these low-carbon steels in the annealed condition, the major microstructure is relatively soft ferrite with scattered small carbide particles, whereas low-carbon steel in the hot-rolled or normalized condition has islands of pearlite. During fusion welding, the HAZ will undergo transformation from ferrite to austenite upon heating, and from austenite back to ferrite during cooling. At this carbon level, the hardening involved in the HAZ is approximately 10 HRC or less. However, when the weldment undergoes a cold-forming operation, it is recommended that the carbon content be limited to 0.06%. The HAZ of steels with higher carbon contents may not have good ductility to accommodate the strain involved in cold forming (Ref 10).

When low-carbon steels are furnished as rimmed steels, their central core regions tend to contain concentrations of elements such as carbon monoxide and carbon dioxide, as well as elements such as sulfur and phosphorus, as a result of segregation during the rimming action. In addition, the gaseous species that result from the reaction between dissolved oxygen and carbon are also trapped in the central core region. During the welding operation, certain amounts of base metal will melt and mix with filler metal because of dilution. The high-temperature, molten weld pool provides conditions that allow the incomplete oxygen-carbon reaction to resume and trapped carbon monoxide and carbon dioxide to evolve. When the weld-pool solidification rate is too rapid for gaseous species to escape, their entrapment will lead to pores in the weld metal. Consequently, sufficient weld-pool deoxidation is the practical way of avoiding porosity problems when welding rimmed steels (Ref 10).

When low-carbon steels are furnished as killed steels, weld-metal porosity is no longer a problem during fusion welding. Instead, a viscous refractory slag sometimes forms on the weld-pool surface of the killed steels, which makes the flow of molten metal sluggish and weld-pool manipulation difficult. The slag becomes viscous when deoxidizers, such as the aluminum, titanium, and zirconium used in the steel-ingot killing practice, form oxides with relatively high melting temperatures. This problem with slag is more intensified for oxyacetylene welding, because of the low temperatures involved in that process. The common remedy is to select a filler metal that contains a sufficient amount of manganese and silicon to reduce the melting point and improve the flow of the slag (Ref 10).

Mild steels are also considered to be very weldable, and cold cracking is usually not a problem when the carbon content is less than 0.2% and the manganese content is less than 1.0%. Preheat, interpass temperature control, PWHT, or special welding procedures are usually not necessary when welding plate with thicknesses of 25 mm (1 in.) or less. The carbon content is usually higher in thicker plates in order to provide the required strength. In addition, the related HAZ cooling rate during the fusion welding of a thicker plate is higher. Consequently, when welding plate that is thicker than 25 mm (1 in.), or when the carbon and manganese contents are beyond 0.3 and 1.4%, respectively, proper caution and the use of low-hydrogen electrodes/fluxes is required to prevent hydrogen cracking (Ref 10).

A useful guideline in selecting carbon-manganese and microalloyed carbon-manganese steels with optimum weldability and toughness, as established by Dolby (Ref 4, 12), is provided below.

For carbon-manganese and microalloyed carbon-manganese steels with more than 0.1% C, use steel base materials with both a low-carbon content and a low-carbon equivalent, but high toughness properties. Select steels that are aluminum-treated to restrict the width of the coarse-grained region in the HAZ and minimize the free-nitrogen level (aluminum-nitride precipitates are relatively stable). The selected steels should be clean. The aluminum-treated or vacuum degassed steels will have lower sulfur levels. The calcium treatment for sulfide shape control will improve the mechanical properties in the through-thickness direction and reduce the lamellar tearing associated with liquidation cracking problems.

When using welding processes with high-energy input for the same base-metal carbon-manganese compositions, the addition of niobium and vanadium will lower HAZ toughness, because of the niobium and vanadium precipitations that occur during cooling and reheating.

Microalloyed steels with less than 0.1% C should have higher toughness and a reduced tendency for hydrogen cracking in the HAZ. Clean aluminum-treated steels with high toughness values should be selected for the reasons described above for carbon-manganese steels. The same precaution regarding welding processes with high-energy inputs, as mentioned above, also applies.

Low-alloy steels that have a low-carbon level should be selected. A high cleavage resistance can be obtained when the hardenability of the steels is sufficient to develop low-carbon martensite for the welding conditions used. Caution should be used in the PWHT of steels that have significant amounts of secondary precipitation-hardening alloying elements, such as vanadium and niobium. Care also should be given to the PWHT of steels that contain high amounts of residual elements such as antimony, phosphorus, arsenic, and tin.

Clean, aluminum-treated steels with high toughness values should be selected for the reasons described above for carbon-manganese steels. The same precaution regarding welding processes with high-energy inputs, as mentioned above, also applies.

Thermomechanically controlled process (TMCP) steels are recently developed steels that exhibit significantly improved strength and toughness properties and weldability. The new technology of using accelerated cooling makes it possible to achieve the same strength level as conventional controlled rolled steels at a lower carbon content (<0.06%). Instead of strengthening by increasing the carbon content, TMCP steels derive their strength and toughness from the very fine ferrite and second-phase microstructure (finely dispersed pearlite or bainite) that occur during the accelerated cooling of the manufacturing process. As a result of the reduction in carbon content and carbon equivalent, these steels significantly improve their resistance to hydrogen-induced cold cracking. Therefore, the concern with preheat, interpass temperature control, and PWHT during welding fabrication is not critical.

However, HAZ softening, especially at high heat inputs, is a concern, because the favorable microstructure of the TMCP steel is reverted during the slow cooling in the HAZ at high heat input levels. Figure 7 shows the hardness profiles in the HAZ of a TMCP steel welded using various heat inputs (Ref 13).

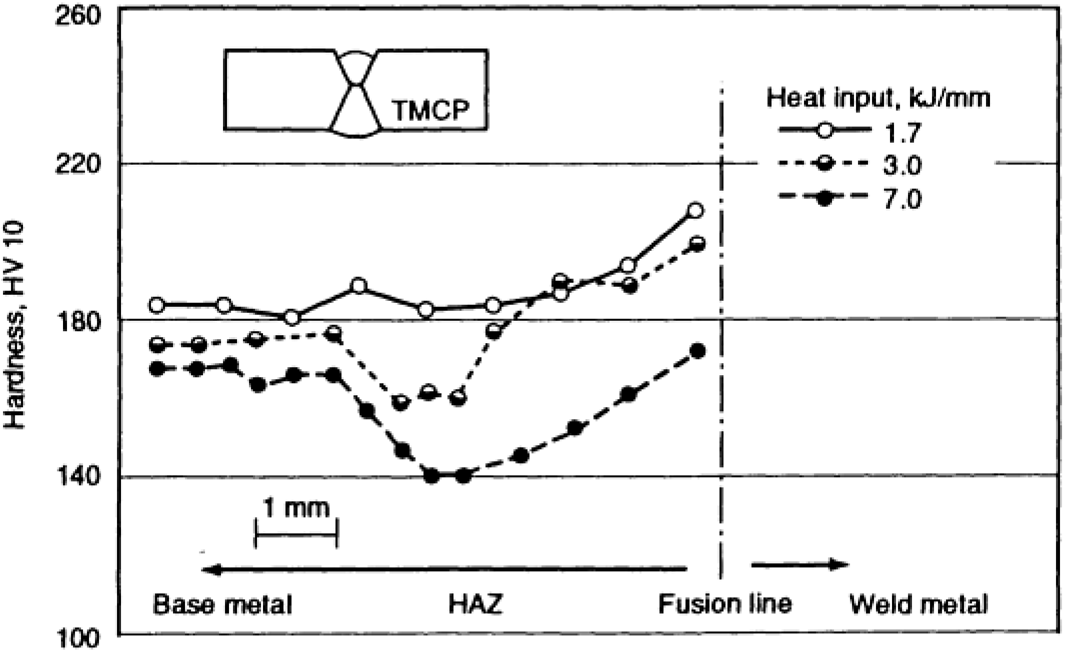

FIG. 7 VARIATION OF HARDNESS PROFILES IN HAZ OF TMCP STEEL WELDED WITH HEAT INPUT OF 1.7, 3.0, AND 7.0 KJ/MM (43, 76, AND 178 KJ/IN.). CARBON, 0.06%; SILICON, 0.14%; MANGANESE, 1.33%; PHOSPHORUS, 0.010%; SULFUR, 0.001%; COPPER, 0.31%; NICKEL, 0.31%; CHROMIUM, 0.05%; NIOBIUM, 0.15%; AND ALUMINUM, 0.034%. SOURCE: REF 13

Chromium-molybdenum steels are relatively inexpensive materials, when compared with high-alloy steels used in the power and petrochemical industries, because of their oxidation resistance, elevated-temperature creep resistance, and sulfide corrosion resistance. The chromium content provides oxidation corrosion resistance, whereas the molybdenum content increases elevated-temperature strength. The presence of finely distributed stable carbide enhances the creep resistance. Because of the alloying element content and approximate 0.15% C level, these steels are air hardenable. Most of these steels are supplied in normalized and tempered or quenched and tempered condition.

The weldability of chromium-molybdenum steels is very similar to that of quenched and tempered and hardenable low-alloy steels. The major problem in the HAZ is cracking in the hardened coarse-grained region, as well as HAZ softening between Ac_1 and Ac_3. Reheat cracking during PWHT and long-term exposure in elevated-temperature service conditions also can cause severe problems. The appropriate preheat and interpass temperature should be selected, and low-hydrogen practice should be used. The recommended practice for welding these steels is provided in ANSI/AWS D 10.8-86.

Although a PWHT is not required for chromium-molybdenum steels with lower chromium contents and thinner gages, it is often conducted immediately after welding as part of the welding procedure. The PWHT of a chromium-molybdenum weldment is also referred to as a stress-relief heat treatment. Such a heat treatment is designed to relieve the residual stresses by a creep-relaxation process and to temper the as-welded hardened microstructures to improve the fracture toughness of the HAZ and the weld metal. Additional beneficial effects of a PWHT are to allow for more dissipation of hydrogen in the weld region and to further reduce the likelihood of hydrogen-induced cold cracking.

Although a stress-relief treatment or PWHT is often used in chromium-molybdenum steel weldments, this type of steel is susceptible to either stress-relief cracking in the HAZ during such a treatment or reheat cracking during postweld reheating processes, including short-term exposure to elevated-temperature service environments. The temperature range for the occurrence of stress-relief/reheat cracking is between 500 and 700 °C (932 and 1292 °F). Although it is not clear, the cracking mechanism is believed to be related to the strengthening of the grain interior as a result of precipitation. Cracking occurs during reheating when the relatively soft grain boundaries that are free of precipitates cannot accommodate the plastic deformation during creep relaxation. A parameter known as the P_{SR} has been developed (Ref 14) to correlate the susceptibility of chromium-molybdenum steels to reheat cracking when the steels contain $\leq 1.5\%$ Cr, $\leq 1.0\%$ Cu, $\leq 2.0\%$ Mo, and $\leq 0.15\%$ V, Nb, and Ti:

$$P_{SR} = \mathrm{CR} + \mathrm{CU} + 2\mathrm{MO} + 10\mathrm{V} + 7\mathrm{NB} + 5\mathrm{TI} - 2 \qquad \textbf{(EQ 5)}$$

When P_{SR} is less than zero, the material is considered sensitive to reheat cracking. However, when there is more than 2% Cr, the tendency for cracking is eliminated.

Another parameter that takes into account the tramp elements that lead to grain-boundary embrittlement and increase the tendency for reheat cracking is the metal composition factor (MCF) (Ref 15):

$$\mathrm{MCF} = \mathrm{SI} + 2\mathrm{CU} + 2\mathrm{P} + 10\mathrm{AS} + 15\mathrm{SN} + 20\mathrm{SB} \qquad \textbf{(EQ 6)}$$

A greater MCF also increases the susceptibility of chromium-molybdenum steels to reheat cracking.

The common welding practices that are used to reduce the likelihood for reheat cracking can be summarized as (Ref 5):

- REDUCE THE STRESS RISERS IN WELDMENT DESIGN.
- USE WELDING PRACTICES THAT MINIMIZE RESIDUAL STRESS, SUCH AS INCREASING THE PREHEAT TEMPERATURE AND REDUCING RESTRAINT.
- WHEN APPLICABLE AND NECESSARY, REDUCE WELD-METAL STRENGTH TO ACCOMMODATE THE PLASTIC DEFORMATION IN THE WELD METAL, AND REDUCE THE STRESS CONCENTRATION IN THE COARSE-GRAINED REGION.

Quenched and tempered steels (Q&T) steels are normally furnished in heat-treated conditions that involve austenitizing and/or quenching and tempering to obtain high-strength properties. The hardenability of Q&T steels is such that the HAZ consists of microstructures of low-carbon martensite and bainite. This type of as-welded HAZ microstructure has desirable mechanical properties that are close to those of the base metal. Consequently, these materials generally do not require a PWHT or stress-relief treatment, except in certain special situations.

Unlike other hardenable low-alloy steels in which high-energy input has to be used to avoid the formation of martensite in the HAZ, the use of Q&T steels requires that welding conditions include a cooling rate in the HAZ that is rapid enough to ensure the reformation of martensite and bainite microstructures. This is necessary because a HAZ cooling rate is too slow will cause the austenitized HAZ to transform into ferrite and a mixture of bainite and martensite. When proetectoid ferrite is undergoing transformation from austenite, the untransformed austenite will become enriched in carbon and will subsequently transform into hard and brittle bainite and martensite. This mixed microstructure of ferrite, bainite, and martensite leads to embrittlement of the coarse-grained HAZ. The slower the cooling rate, the greater the extent of embrittlement in the HAZ. When welding less-hardenable or thinner material, an even faster critical cooling rate (less heat input) is required to avoid transformation of brittle mixed microstructures (Ref 10).

Another concern with welding Q&T steels is the strict maintenance of low-hydrogen welding practice in order to prevent underbead cold cracking. A preheat is one of the most effective ways of reducing the tendency for cold cracking. However, a preheat also significantly reduces the cooling rate in the HAZ. Consequently, it should be applied in such a manner that a satisfactorily fast cooling rate can be achieved in the HAZ.

When welding Q&T steels with a higher strength level, the tendency for hydrogen cracking to occur in the weld metal increases. Consequently, the allowable moisture content in the electrode coating and flux, as well as the handling of these materials, become much more critical. For Q&T steels with a yield strength level under 483 MPa (70 ksi), a 0.4 wt% moisture content is the allowable limit in covered electrode, whereas for steels with a strength level above 690 MPa (100 ksi), the limit is 0.1 wt% (Ref 10). The use of a temper bead technique, as shown in Fig. 8, also can help avoid undesirable regions of high hardness and low toughness in the HAZ.

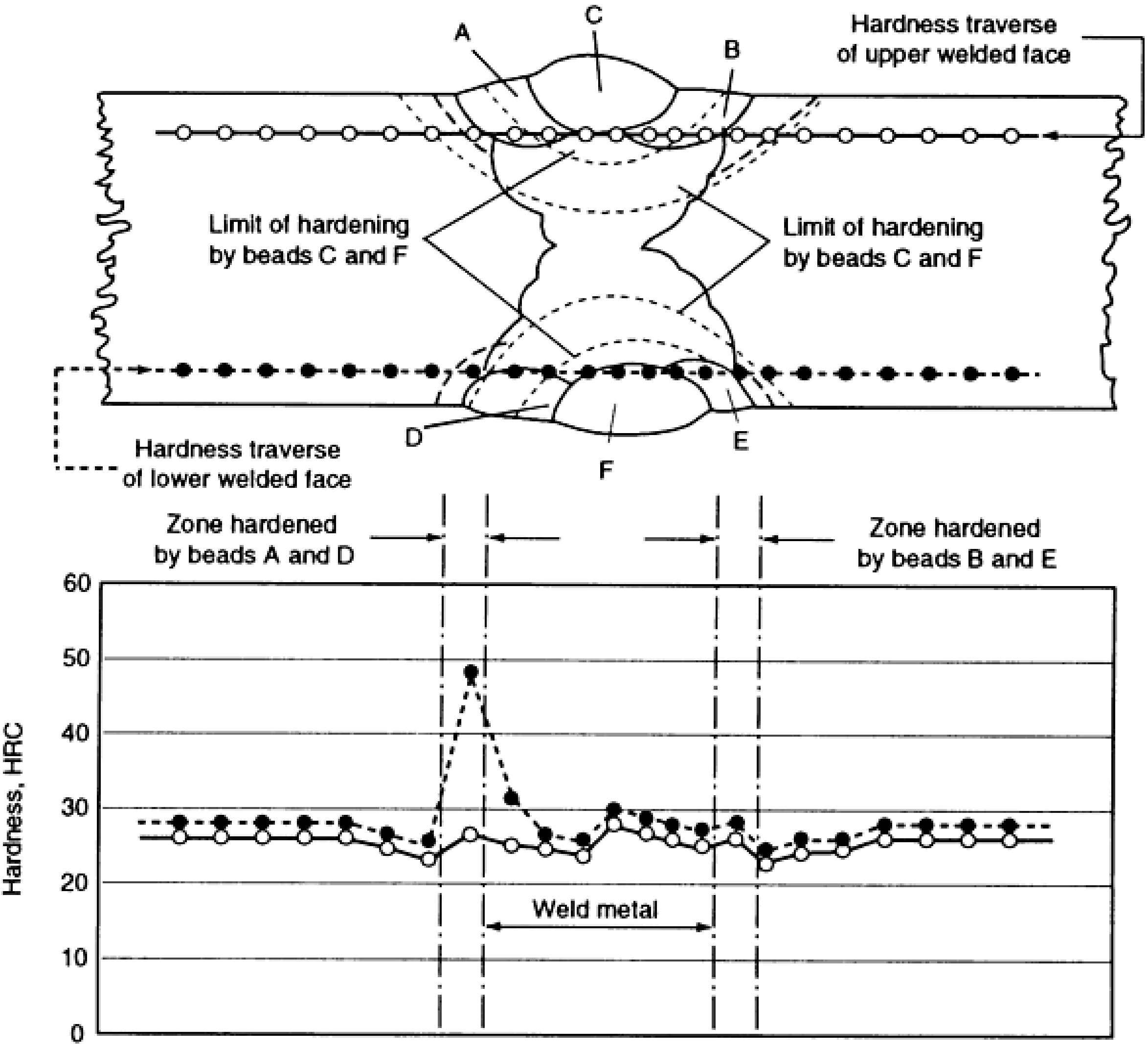

FIG. 8 TEMPERING BEAD TECHNIQUE FOR MULTIPASS WELDING OF A BUTT JOINT IN A Q&T ALLOY STEEL. HARDNESS VALUES FOUND BY UPPER AND LOWER TRAVERSES. DISTRIBUTION OF HARDNESS IS ALMOST IDENTICAL, EXCEPT FOR UNTEMPERED ZONE OF BEAD D, AT LEFT. SOURCE: REF 16

Multipass welding is a good practice when welding Q&T steels. In addition to the refining and tempering effects on the weld metal of the previous weld pass, the smaller heat input of multipass welding, compared to that of single-pass welding, can help achieve a fast cooling rate. When hydrogen cracking in weld metal is a concern, it is recommended that thin-layer multipass welding be used, along with soaking at the interpass temperature for a predetermined time before depositing the subsequent pass (Ref 17). Although this practice helps dissipate the hydrogen in the weld metal in each weld pass, productivity is decreased.

In Q&T steels, there is a softening region in the HAZ that is caused by a HAZ thermal cycle with a peak temperature between A_{c1} and A_{c3} (intercritical region). At temperatures below A_{c3} during heating, the carbides do not completely dissolve in the austenite. Therefore, the carbon level in the austenite is a lower concentration. During cooling, the unsaturated austenite transforms at high temperature and results in a microstructure with lower strength.

References cited in this section

3. O. GRONG AND O.M. AKSELSEN, HAZ TOUGHNESS OF MICROALLOYED STEELS FOR OFFSHORE, *MET. CONSTR.,* SEPT 1986, P 557
4. K.E. EASTERLIN, *INTRODUCTION TO THE PHYSICAL METALLURGY OF WELDING,* BUTTERWORTHS, 1983
5. C.F. CHOU, *THEORY AND TECHNOLOGY OF METAL WELDING,* VOL 2, MECHANICAL INDUSTRY PUBLISHER, 1981 (IN CHINESE)
9. O.M. AKSELSEN, "MICROSTRUCTURE AND TOUGHNESS IN HAZ OF BORON CONTAINING STEELS," REPORT STF A84065, THE WELDING CENTER (SINTEF), TRONDHEIM, NORWAY, 1984
10. W.H. KEARNS, *WELDING HANDBOOK,* VOL 4, AWS, 1984
11. M. GRAY, WELDABILITY OF NIOBIUM CONTAINING HIGH STRENGTH LOW ALLOY STEEL, *WELD. RES. COUNC. BULL.,* NO. 213, FEB 1976
12. R.E. DOLBY, "FACTORS CONTROLLING WELD TOUGHNESS--THE PRESENT POSITION," REPORT NO. 14, THE WELDING INSTITUTE, 1976
13. H.J. KIM AND J.G. YOUN, CHARACTERISTICS OF TMCP STEEL AND ITS SOFTENING, *WELDING METALLURGY OF STRUCTURAL STEEL,* J.Y. KOO, ED., AIME, 1987
14. Y. ITO AND M. NAKANISH, STUDY ON STRESS RELIEF CRACKING IN WELDING LOW ALLOY STEELS, *SUMITOMO SEARCH,* NO. 7, MAY 1972, P 27-36
15. C.J. MCMAHON, JR., R.J. DOBBS, AND D.H. GENTUER, STRESS RELIEF CRACKING IN MN, MO, NI, AND MN-MO-NI-CR PRESSURE VESSEL STEELS, *MATER. SCI. ENG.,* VOL 37, 1979, P 179-186
16. G.E. LINERT, *WELDING METALLURGY,* AWS
17. D.N. SHACKLETON, *WELDING HY-100 AND HY-130 STEELS,* THE WELDING INSTITUTE, 1973

Influence of Welding on Steel Weldment Properties

C.C. Chen and Abe Pollack, Microalloying International, Inc.

Weldment Considerations

Corrosion. Carbon and low-alloy steels used for structural applications are not commonly utilized in severe corrosion environments. However, they are used in moderate corrosive service conditions, such as those in oil refinery facilities and sour gas/oil pipelines. The presence of a weld often leads to a reduction in corrosion resistance because of these circumstances (Ref 6):

- COMPOSITIONAL VARIATIONS IN THE BASE METAL, HAZ, AND WELD METAL THAT RESULT IN A CONDITION FAVORING GALVANIC CORROSION
- PRESENCE OF WELDING RESIDUAL STRESSES THAT LEAD TO STRESS-CORROSION CRACKING
- PRESENCE OF WELD DISCONTINUITIES, SUCH AS SURFACE FLAWS, WHICH ACT AS PREFERENTIAL SITES FOR LOCAL CORROSION ATTACK

In a corrosive environment, the prevention of hydrogen-induced cracking (HIC) and sulfide stress-corrosion cracking (SCC) is of great importance.

Hydrogen-induced cracking, which has been observed in both high- and low-strength steels, even under nonstressed conditions, occurs primarily in the low-strength steels that are exposed to a hydrogen-containing environment. Because of its rapid cooling and solidification, weld metal forms a structure of rendrites and has oxide inclusions dispersed in the form of fine globules. It has been confirmed that weld metals, even when used without a filler metal of special chemistry, do not develop HIC up to a maximum hardness of 280 HV. In comparison, HIC has been observed primarily in the base metal and HAZ (Ref 18, 19).

Stress-corrosion cracking is defined as a cracking phenomenon that occurs under sulfide-corrosive conditions when steel materials are subjected to stress. Steels that are to be used in environments containing hydrogen sulfide can be selected per Ref 20. According to this standard, a value of 22 HRC (248 Hv) is recommended as the hardness limit for steel line pipe used in sour-gas service.

In the case of a weldment, the HAZ generally becomes harder than the base metal, and, consequently, is more susceptible to SCC. In addition to the more susceptible microstructure, other factors that contribute adversely to SCC are weld residual stresses and the presence of the entrant angle of the weld cap as a stress raiser in the welded region.

When the welded joints are quenched and tempered or normalized, the HAZ disappears and the tendency for the previous HAZ to develop SCC is altered (Ref 18, 19). The SCC susceptibility is greater in the as-welded condition than in the Q&T (PWHT) condition.

Fatigue Strength of Weld Joints. Fatigue cracks in a welded structure under cyclic loading are often observed to develop at weld toe regions. This is because the weld toes are regions of stress concentration. The toes are the point at which the base metal and weld metal meet and where changes are greatest in both sections. More importantly, slag intrusions that result from a welding operation are often present at the weld toe regions.

Various methods have been utilized to improve the fatigue strength of welded structures by reducing the stress concentration in the regions of weld toes and weld reinforcements. These methods are (Ref 21, 22):

- REMOVING THE WELD REINFORCEMENT ON BOTH SIDES OF A BUTT JOINT BY MACHINING THE WELD TO THE LEVEL OF THE BASE METAL
- INCREASING THE RADIUS AND ENTRANT ANGLE OF WELD TOES BY MECHANICALLY GRINDING THE INTERSECTING REGIONS BETWEEN THE WELD AND THE BASE METAL
- CHANGING THE STATE OF THE SURFACE-WELD RESIDUAL STRESS BY THE MECHANICAL COLD WORKING OF THE WELD SURFACE AND THE BASE METAL AT THE WELD TOE REGIONS
- COATING AND/OR PAINTING THE WELDS AND BASE METAL TO PREVENT CORROSION IN THE REGION OF STRESS CONCENTRATION
- INCREASING THE TOE RADIUS AND WELD ENTRANT ANGLE BY ADDING A WELD BEAD ON BOTH SIDES OF THE REINFORCEMENTS USING A FILLER METAL WITH HIGH FLUIDITY
- USING WELDING CONDITIONS THAT WILL RESULT IN A GREATER WELD TOE RADIUS AND ENTRANT ANGLE
- REMELTING THE SURFACE IN THE WELD TOE REGIONS FOR FLATTENING AND SMOOTHING WELD PROFILES USING THE GAS-TUNGSTEN ARC WELDING PROCESS

References cited in this section

6. R.D. STOUT AND W.D. DOTY, *WELDABILITY OF STEELS,* 3RD ED., S. EPSTEIN AND R.E. SOMERS, ED., WELDING RESEARCH COUNCIL, NEW YORK, 1978
18. T. TAIRA *ET AL.,* "SULFIDE CORROSION CRACKING OF GIRTH-WELDED JOINT OF LINE PIPE FOR SOUR GAS SERVICE," CORROSION '79, T-1F SYMPOSIUM (ATLANTA, GA), NACE, MARCH 1979
19. M. ITOH *ET AL.,* "HYDROGEN SULFIDE CORROSION CRACKING OF LARGE DIAMETER LINE PIPE UP TO GRADE X-65," MIDDLE EAST CORROSION CONFERENCE (BAHRAIN), NACE, APRIL 1979
20. "SULFIDE STRESS CRACKING RESISTANT METALLIC MATERIALS FOR OIL FIELD

EQUIPMENT," STANDARD MATERIAL REQUIREMENTS DOCUMENT MR0175-91, NACE
21. Z. ZACZEK, IMPROVEMENT IN THE FATIGUE STRENGTH OF BUTT WELDED JOINTS BY TIG REMELTING OF WELDING REINFORCEMENTS, *MET. CONSTR.*, JULY 1986
22. G.S. BOOTH, IMPROVING THE FATIGUE STRENGTH OF WELDED JOINT BY GRINDING--TECHNIQUES AND BENEFITS, *MET. CONSTR.*, JULY 1986

Influence of Welding on Steel Weldment Properties

C.C. Chen and Abe Pollack, Microalloying International, Inc.

Effect of Welding Procedure on Steel-Weldment Properties

The American Welding Society defines a welding procedure as the detailed methods and practices that include all joint welding procedures involved in the production of a weldment (Ref 23). In a comprehensive review of the relationship between procedures and weldment properties, the conclusions described below were made (Ref 23).

First, for any welding process/steel combination, welding procedures have a predominant role, along with the electrode (if a filler metal is used), in determining the quality and mechanical properties of the weldment. However, these two variables (welding procedures and mechanical properties) do not have a direct cause-effect relationship.

Second, the mechanical properties of any metal are attributable only to its soundness, microstructure, and chemical composition, and to whether or not it is base plate, weld metal, vapor-deposited metal, electrodeposited metal, or any other type. The response of a metal to mechanical forces depends on its current state and not on the manner in which this state was created.

Third, metallurgical research has established relationships between mechanical properties and microstructure, such as the inverse relationship between grain size and strength, the difference in strength between pearlite and martensite, the embrittling effect of grain-boundary films, the deleterious effect of inclusions, and others. Conversely, there is no direct relationship between the current setting of a welding machine and the final properties of the weldment. Rather, the true cause and effect relationship is that:

- INCREASED CURRENT INTRODUCES MORE HEAT INTO THE ALLOY
- INCREASED HEAT AFFECTS THE STRUCTURE OF BOTH THE WELD METAL AND THE HAZ IN A PARTICULAR MANNER, DEPENDING ON THE ALLOY CHEMISTRY, PHASES THAT ARE PRESENT, AND OTHER FACTORS
- DIFFERENCES IN STRUCTURE PRODUCE DIFFERENT MECHANICAL PROPERTIES

Described below are the specific welding procedure factors that affect weldment properties.

Preheat Temperature

Preheating is the application of heat to the base metal immediately before welding, brazing, soldering, or cutting (Ref 24). The preheat temperature depends on many factors, such as the composition and mass of the base metal, the ambient temperature, and the welding procedure.

Preheating is used to:

- REDUCE SHRINKAGE STRESSES IN THE WELD AND ADJACENT BASE METAL, WHICH IS ESPECIALLY IMPORTANT WITH HIGHLY RESTRAINED JOINTS
- PROVIDE A SLOWER RATE OF COOLING THROUGH THE CRITICAL TEMPERATURE RANGE (APPROXIMATELY 879 TO 720 °C, OR 1600 TO 1330 °F), WHICH PREVENTS EXCESSIVE HARDENING AND LOWERS THE DUCTILITY OF BOTH THE WELD AND HEAT-

AFFECTED AREA OF THE BASE PLATE

- PROVIDE A SLOWER RATE OF COOLING DOWN TO 205 °C (400 °F), ALLOWING MORE TIME FOR ANY HYDROGEN THAT IS PRESENT TO DIFFUSE AWAY FROM THE WELD AND ADJACENT PLATE IN ORDER TO AVOID UNDERBEAD CRACKING

Gas torches, gas burners, heat-treating furnaces, electric-resistance heaters, low-frequency induction heating, and temporary furnaces are some of the preheating methods that are used. The selection of the method depends on several factors, such as the preheat temperature, the length of preheating time, the size and shape of the parts, and whether it is a one-of-a-kind or a continuous production type of operation.

For critical applications, the preheat temperature must be precisely controlled. In these cases, controllable heating systems are used, and thermocouples are attached directly to the part being heated. The thermocouple measures the exact temperature of the part and provides a signal to a controller, which regulates the fuel or electrical power required for heating. By accurately regulating the fuel or power, the temperature of the part being heated can be held to close limits. Most code work requires precise heat-temperature control. Preheat and interpass temperature requirements are described in the article "Welding of Carbon Steels" in this Volume and in ANSI/AWS D1.1-92.

The effect of preheat is illustrated by the data in Table 1. The analysis indicates that the extent to which the ductility and fracture-transition temperature are affected depends on the chemical composition of the steel. Changes in preheat temperature for the steel with 0.24% C were effective only above 120 °C (250 °F), whereas for the steel with 0.18% C, the changes were effective over the entire range from -12 to 260 °C (10 to 500 °F).

TABLE 1 EFFECT OF PREHEAT AND STRESS-RELIEF TREATMENT ON TRANSITION TEMPERATURE OF WELDED STEELS

CONDITION OF PLATE	**19 MM ($\frac{3}{4}$ IN.) 0.18% C, SEMIKILLED STEEL**				**19 MM ($\frac{3}{4}$ IN.) 0.24% C, SEMIKILLED STEEL**			
	DUCTILITY-TRANSITION TEMPERATURE		**FRACTURE-TRANSITION TEMPERATURE**		**DUCTILITY-TRANSITION TEMPERATURE**		**FRACTURE-TRANSITION TEMPERATURE**	
	°C	**°F**	**°C**	**°F**	**°C**	**°F**	**°C**	**°F**
AS-ROLLED PLATE	-79	-110	-1	30	-40	-40	66	150
WELDED AT 5 °C (10 °F)	-8	18	10	50	14	58	60	140
WELDED AT 21 °C (70 °F)	-18	0	-9	15	17	62	68	155
WELDED AT 65 °C (150 °F)	-42	-44	-9	15	19	67	66	150
WELDED AT 120 °C (250 °F)	-57	-70	-18	0	18	65	54	130
WELDED AT 250 °C (400 °F)	-36	-32	-18	0	-1	30	49	120
WELDED AT 260 °C (500 °F)	-50	-58	-46	-50	-18	0	82	180
WELDED AT 21 °C (70 °F) AND STRESS RELIEVED AT 620 °C (1150 °F)	-51	-60	-18	0	-18	0	60	140
WELDED AT 38 °C (100 °F) AND STRESS RELIEVED AT 620 °C (1150 °F)	-58	-73	-7	+20	. . .	. . .	. . .	. . .

Note: Ductility-transition temperature selected at 1% lateral contraction; fracture-transition temperature selected at 50% shear.

Source: Ref 6

Interpass Temperature

The interpass temperature, which is involved in multipass welds, is represented by the minimum and maximum temperatures of the deposited weld metal and adjacent base metal before the next pass is started.

Usually, a steel that requires preheating to a specified temperature also must be kept at this temperature between weld passes. With many weldments, the heat input during welding is adequate to maintain the interpass temperature. On a massive weldment, it is not likely that the heat input of the welding process will be sufficient to maintain the required interpass temperature. If this is the case, then torch heating between passes may be required.

Once an assembly has been preheated and the welding begun, it is desirable to finish welding as soon as possible, in order to avoid the need for interpass heating.

Because the purpose of preheating is to reduce the cooling rate, it logically follows that the same slow cooling should be accorded all passes. This can only be accomplished by maintaining an interpass temperature that is at least equal to the preheat temperature. If this is not done, then each individual bead will be subjected to the same high quench rate as the first bead of a nonpreheated assembly.

The required minimum temperatures, based on specific steel, welding process, and plate thickness, should be heeded. When heat buildup becomes excessive, the weldment must be allowed to cool, but not below the minimum interpass temperature. The temperature of the welding area must be maintained within minimum and maximum interpass temperatures.

Postweld Heat Treatment

A PWHT is generally considered necessary for welds in thicker-section steels, in order to reduce the high as-welded residual stress levels and improve the toughness and defect tolerance of the joint. Many fabrication codes offer guidance on the duration and temperature of the PWHT, although some discrepancies between various codes exist. In assessing the time and temperature required to provide a suitable PWHT, it is necessary to know how such parameters will respond to various heat-treatment schedules.

In all heat treatments, with the possible exception of fusion, the heating rate and time can be specified. The maximum temperature is related to the composition of the steel, the holding time (at the maximum temperature) is related to the material thickness, and the cooling rate is related to the particular treatment and to the code. The rate of heating usually ranges from 149 to 177 °C (300 to 350 °F) per hour. The holding temperature is usually 2.4 min/mm (1 h/in.) of maximum thickness in order to provide uniform heating throughout. The cooling rate also ranges from 149 to 177 °C (300 to 350 °F) per hour, down to a specific temperature. In some cases, the cooling rate can be increased when the part has cooled to a temperature from 260 to 316 °C (500 to 600 °F). These rates of heating, holding, and cooling are usually part of the specification and must be followed explicitly.

Example 1: Effects of PWHT on the HAZ Properties of Welds in a Carbon-Manganese Grade Steel.

An investigation (Ref 25) of the effects of PWHT has reported the variation in HAZ properties (strength, hardness, and toughness) with different PWHT parameters. The tests were carried out on a 50 mm (2 in.) BS1501-225-490B-LT50 steel manufactured by the British Steel Corporation. The steel was a carbon-manganese grade that had been silicon killed and treated with aluminum and niobium. This grade was chosen because it is increasingly being used in pressure-vessel applications and is also similar to BS4360 grade 50E, a steel that is used in many off-shore fabrications and other structures. The submerged arc welding process was used with a heat input of approximately 3.4 kJ/mm (89 kJ/in.) of the fill passes. In addition, the effects of multiple PWHT cycles, which are often encountered in pressure-vessel fabrication or in cases where repairs have been made, have been investigated in the same manner.

Longitudinal tensile tests of the HAZ were conducted, and the data are plotted in Fig. 9(a). No significant effect that was due to multiple stress-relief cycles was noted in the HAZ regions. The values observed exceeded the plate minimum requirements. Adequate ductility and elongation values were observed at 600 °C (1112 °F). Figure 9(b) indicates no significant effect of PWHT time on the HAZ yield-stress tensile strength or ductility at this temperature. The data plotted in Fig. 9(c) indicated that only a minor reduction in yield stress and tensile strength were observed with increasing PWHT temperature. No significant effects on ductility were observed.

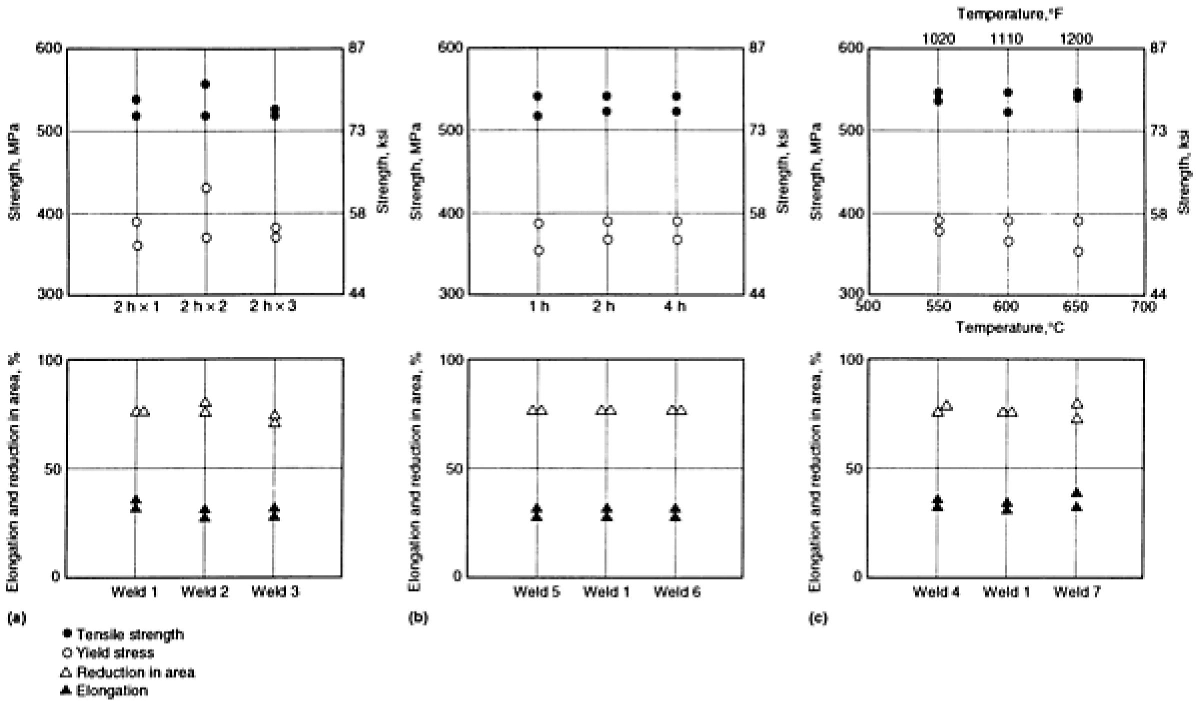

FIG. 9 EFFECT OF PWHT ON THE HAZ TENSILE PROPERTIES OF A CARBON-MANGANESE GRADE STEEL. (A) EFFECT OF MULTIPLE PWHT AT 600 °C (1112 °F), (B) EFFECT OF PWHT TIME. (C) EFFECT OF PWHT TEMPERATURE. SOURCE: REF 25

Charpy Tests. Five testing temperatures were chosen and three specimens were tested at each of the temperatures, which were -80 °C (-112 °F), -60 °C (-76 °F), -40 °C (-40 °F), -20 °C (-4 °F), and 0 °C (32 °F). For the welds, one specimen was taken from the first side of the HAZ, one from the second side of the HAZ, and one from the HAZ at the plate midthickness, which was associated with second-side welds. The data in Fig. 10 show that multiple PWHT cycles had only a marginal effect, in that a slight increase in the number of low values at low temperatures was observed at two and three cycles, particularly from specimens machined from the plate midthickness.

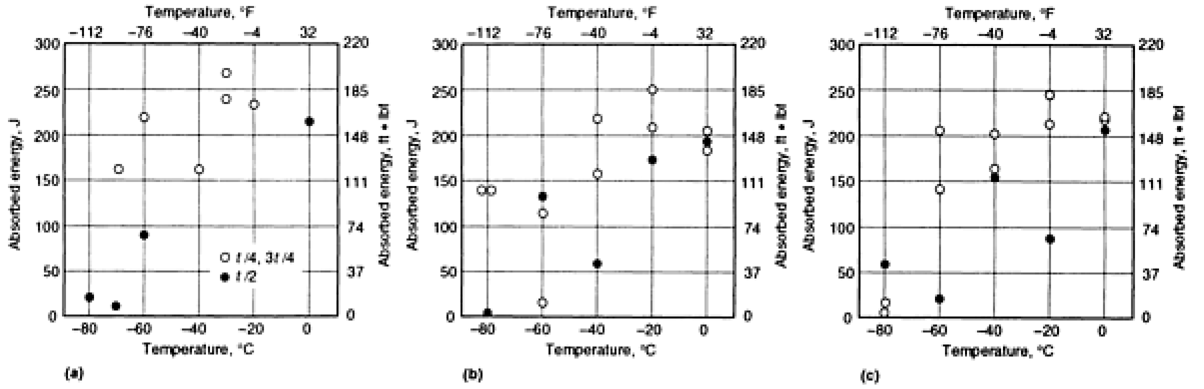

FIG. 10 EFFECT OF MULTIPLE PWHT CYCLES ON THE HAZ CHARPY V-NOTCH PROPERTIES OF A CARBON-MANGANESE GRADE STEEL. (A) WELD 1; PWHT, 1 × 2H. (B) WELD 2; PWHT, 2 × 2 H. (C) WELD 3; PWHT, 3 × 2 H. SOURCE:REF 25

Although extended PWHT has not led to an appreciable variation in HAZ Charpy properties from the ($\frac{1}{4}$)t and ($\frac{3}{4}$)t locations, a pronounced drop in absorbed energy levels has been observed after 4 h in specimens machined from the plate midthickness. This is clearly shown in Fig. 11, from which the toughness at the plate midthickness appears to fall off with time, although this does not happen at the $(\frac{1}{4})t$ locations.

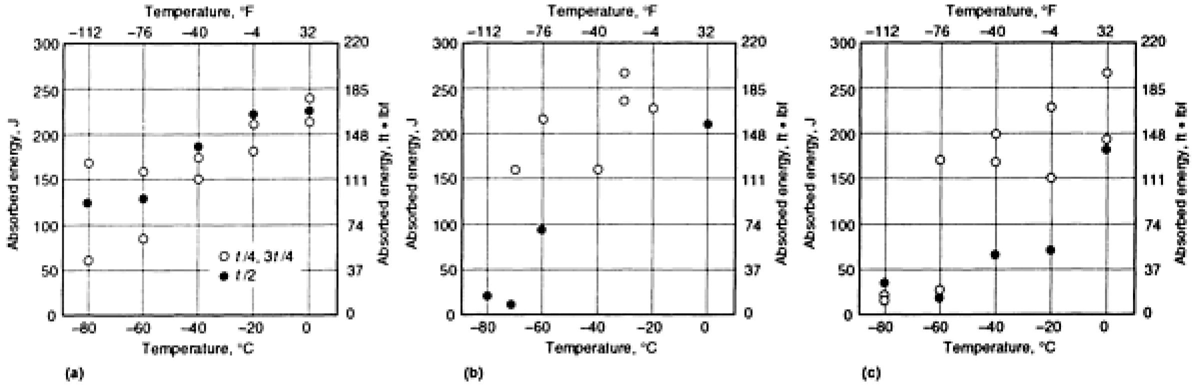

FIG. 11 EFFECT OF PWHT TIME AT 600 °C (1112 °F) ON THE HAZ CHARPY V-NOTCH PROPERTIES OF A CARBON-MANGANESE GRADE STEEL. (A) WELD 5; PWHT TIME, 1 H. (B) WELD 1; PWHT TIME, 2 H. (C) WELD 6; PWHT TIME, 4 H. SOURCE:REF 25

The effects of increasing the PWHT temperature are shown in Fig. 12, from which it is clear that no pronounced effect of temperature on toughness exists, except at the plate midthickness. Here, there seems to be a trend of increasing transition temperature with increasing PWHT temperature.

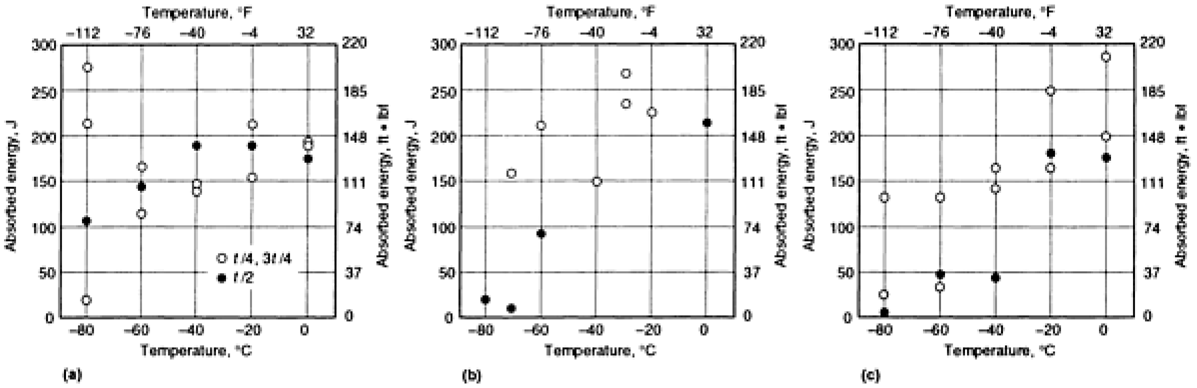

FIG. 12 EFFECT OF PWHT TEMPERATURE ON THE HAZ CHARPY V-NOTCH PROPERTIES OF A CARBON-MANGANESE GRADE STEEL. (A) WELD 4, 550 °C (1020 °F). (B) WELD 1, 600 °C (1110 °F). (C) WELD 7, 650 °C (1200 °F). SOURCE:REF 25

Hardness Survey. Hardness traverses were conducted in the macrosection from each weld after PWHT, reaching from the parent plate that was unaffected by the welding processes into the weld metal. Similar traverses were made on welds 1 and 2 in the as-welded condition. On each macrosection, three traverses were made at these locations: 3 mm (0.12 in.) below the surface of the first side; at the plate midthickness; and 3 mm (0.12 in.) below the surface of the second side. In all cases, a 98 N (10 kgf) load was used on a Vickers hardness machine.

The as-welded hardness data from welds 1 and 2 are shown in Fig. 13. Plate values were approximately 170 to 180 HV 10, and weld-metal values were 200 to 220 HV 10. Peak HAZ hardness was below 240 HV 10.

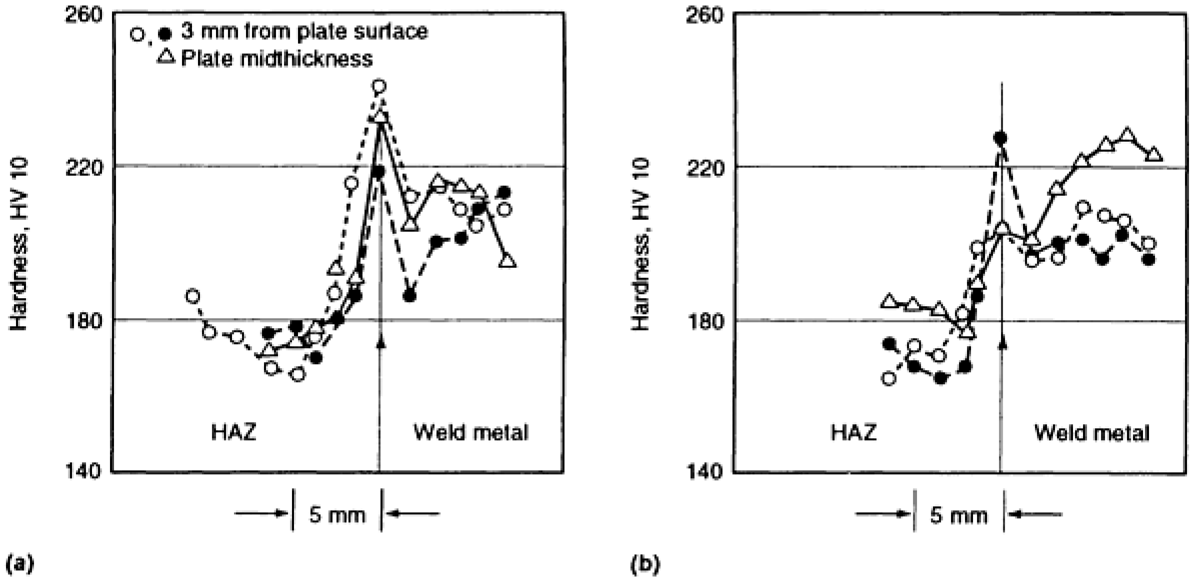

FIG. 13 AS-WELDED HARDNESS DATA FOR TWO WELDS IN A CARBON-MANGANESE GRADE STEEL. (A) WELD 1. (B) WELD 2. SOURCE: REF 25

The effect of multiple PWHT cycles is shown in Fig. 14. Multiple PWHTs had little influence on HAZ or plate hardness, where peak HAZ values of less than 240 HV were obtained and plate values were approximately 140 to 160 HV 10. In all welds, the hardness at the plate midthickness was slightly higher than it was at the surfaces, because of the presence of a segregated band, as is frequently observed in continuously cast steel.

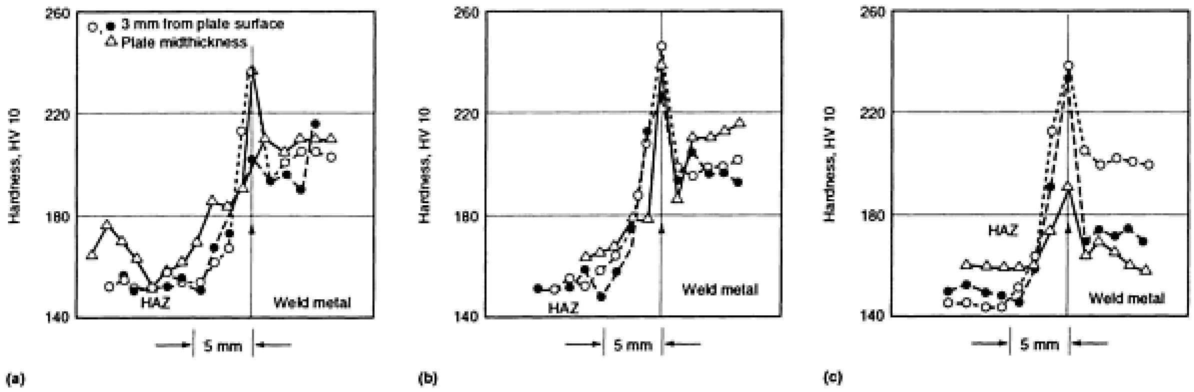

FIG. 14 EFFECT OF MULTIPLE PWHT CYCLES ON THE HARDNESS OF WELDS IN A CARBON-MANGANESE GRADE STEEL. (A) WELD 1; PWHT, 1 × 2 H. (B) WELD 2; PWHT, 2 × 2 H. (C) WELD 3; PWHT, 3 × 2 H. SOURCE: REF 25

Weld-metal hardness shows a more noticeable effect when exposed to multiple stress-relief cycles. Welds 1 and 2 had a fairly uniform hardness of approximately 200 HV, but after the third PWHT cycle, the weld-metal hardness levels dropped to 160 to 220 HV, with the lowest values appearing at the weld midthickness.

The values in Fig. 15 show only a slight reduction in HAZ hardness as PWHT time increases. Plate values are unaffected. The weld metal behaves in a nonsystematic manner. Values between approximately 190 and 210 HV 10 were observed after 1 h, increasing to approximately 200 to 220 HV 10 after 2 h, and decreasing to approximately 160 to 195 HV 10 after 4 h.

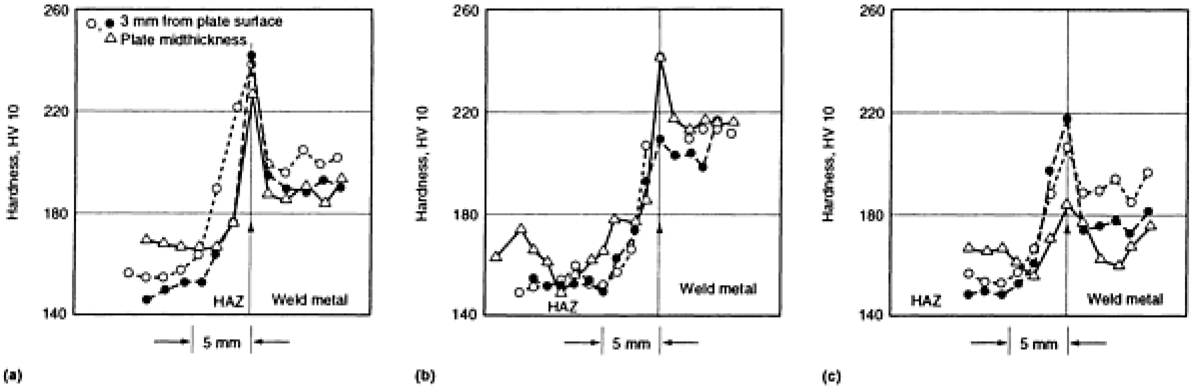

FIG. 15 EFFECT OF PWHT TIME AT 600 °C (1112 °F) ON THE HARDNESS OF WELDS IN A CARBON-MANGANESE GRADE STEEL. (A)WELD 5, 1 H. (B)WELD 1, 2 H. (C) WELD 6, 4 H. SOURCE: REF 25

Figure 16 shows that plate and HAZ hardness levels appear to be fairly insensitive to PWHT temperature. No obvious trends in the variation of hardness with temperature are apparent. However, a slight drop in plate hardness has been observed. Weld-metal hardness appears to decline after a 650 °C (1202 °F) PWHT, where values of approximately 180 to 190 HV 10 have been recorded. This compared with values of approximately 190 to 210 HV 10, which were recorded at 550 and 600 °C (1022 and 1112 °F).

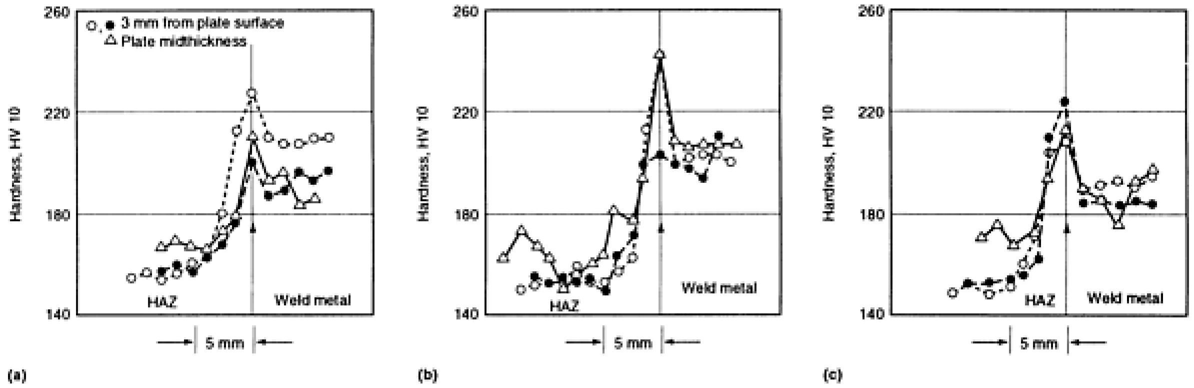

FIG. 16 EFFECT OF PWHT TEMPERATURE ON THE HARDNESS OF WELDS IN A CARBON-MANGANESE GRADE STEEL. (A) WELD 4, 550 °C (1020 °F). (B) WELD 1, 600 °C (1110 °F). (C) WELD 7, 650 °C (1200 °F). SOURCE: REF 25

It was concluded that increasing the PWHT time at 600 °C (1112 °F), whether by prolonging single treatments or by using multiple treatments, had little effect on the mechanized properties. In addition, increasing the PWHT temperature from 550 to 650 °C (1022 to 1202 °F) resulted in no significant drop in HAZ strength or hardness. However, evidence for a drop in the HAZ Charpy V notch-toughness at the plate midthickness with increasing temperature was found.

Heat Input

The welding process and procedure both influence the energy input used to made a weld. The greater the energy input, the slower the cooling rate. Heat input is a function of welding current, arc voltage, and travel speed. To increase the heat input, either the welding current should be increased or the travel speed should be reduced. Welding current is related to the process and the electrode size. Heat input is calculated by using the following formula:

$$HE = EI\frac{60}{S} \quad \textbf{(EQ 7)}$$

where *HE* is the energy input in joules per linear measure of weld, *E* is the arc voltage in volts, *I* is the welding current in amperes, and *S* is the travel speed (linear measure) per minute.

By increasing the amperage or voltage, heat input increases, but by increasing the travel speed, heat input decreases. The voltage has a minor effect, because it varies only slightly, when compared with the other factors. The general higher input reduces the cooling rate. This must be used with caution, because with Q&T steels, too high a heat input will tend to soften the HAZ, and its strength level will be reduced.

In relatively low hardenability steels, it is possible to produce an unhardened HAZ by increasing the heat input. In higher-hardenability steels, the tendency toward cracking and the maximum hardness will be reduced by a slower cooling rate.

There are limits to the amount of heat input that can be used. In this case, preheating is used in order to reduce cooling rates.

Effect of Welding Process. Each welding process has a different thermal cycle. Figure 17 shows the time-temperature relationship of base metal taken immediately adjacent to welds made by two welding processes (Ref 26). The rate of heat rise, the maximum temperature, the time at high temperature, and the rate at which the metal cools are quite different for the shielded metal arc and the electroslag welding processes. Those processes with the highest concentration of heat will cause the temperature to both rise and fall much more rapidly. The curve shown for the shielded metal arc weld rises almost instantaneously, and the cooling rate of the base metal is a very steep slope, which indicates quick cooling. The curve for electroslag welding rises more slowly, is held at a high temperature for a fairly long time, and then decreases slowly. The temperature changes that occur during an arc welding operation are much quicker and more abrupt than for most metallurgical processes. The metallurgical reactions from the heat of welding do not follow the normal heat-treating relationships. The temperature changes with electroslag welding are more similar to those encountered in foundry metallurgy.

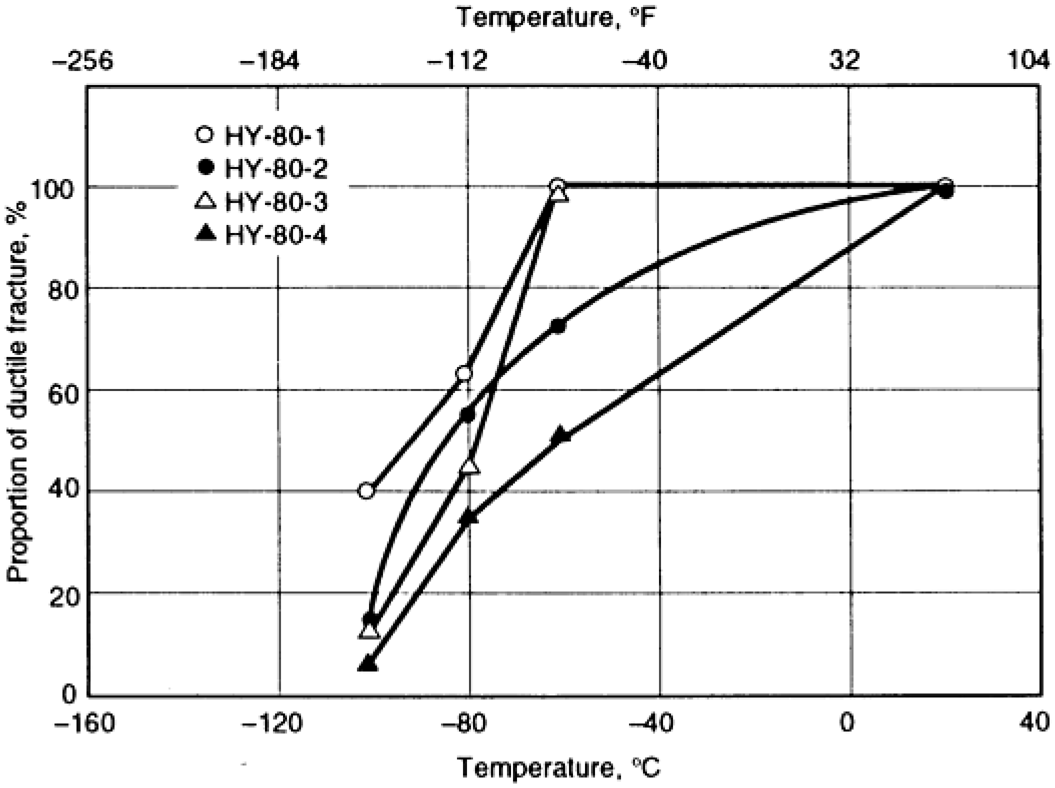

FIG. 17 PERCENT DUCTILE FRACTURE VERSUS TEST TEMPERATURE FOR THE SUBMERGED ARC BEND-IN-

Example 2: Effect of Heat Input on Properties of HY-80 and HSLA-80 Weld Metals.

Because the low-alloy Q&T steels are already in a heat-treated condition, any heating beyond a certain temperature will destroy the properties developed in them by the manufacturing process. A study (Ref 27) was conducted to assess the influence of composition and energy input on the structure and properties of single-pass submerged arc bead-in-groove welds produced on HY-80 and HSLA-80 steels.

Table 2 and 3 gives the chemical composition of the high-strength steels and the welding electrodes used in the study; Table 3 lists the welding conditions. The single-pass welds were produced at energy input levels of 1, 2, 3, and 4 kJ/mm (25, 51, 76, and 101 kJ/in.). The size of the machined grooves was increased to accommodate the increase in energy input, while maintaining the same level of dilution (50%) from the base plate.

TABLE 2 CHEMICAL COMPOSITIONS OF HIGH-STRENGTH STEELS AND WELDING ELECTRODES FROM EXAMPLE 2

MATERIAL	COMPOSITION, WT%									
	C	Mn	Si	S	P	Ni	Cr	Mo	Cu	Nb
HY-80	0.17	0.30	0.19	0.014	0.007	2.59	1.53	0.42	0.03	0.005
HSLA-80	0.06	0.50	0.27	0.004	0.007	0.92	0.66	0.25	1.02	0.044
100S-1	0.07	1.57	0.52	0.004	0.004	1.86	0.03	0.28	0.06	
HY-80^(A)	0.12-0.18	0.10-0.40	0.15-0.35	0.002-0.020	0.020	2.00-3.25	1.00-1.80	0.20-0.60	0.25	
HSLA-80^(B)	0.06	0.40-0.70	0.40	0.006	0.020	0.70-1.00	0.60-0.90	0.15-0.25	1.00-1.30	0.02-0.06
100S-1^(C)	0.08	1.25-1.80	0.20-0.55	0.01	0.01	1.40-2.10	0.30	0.25-0.55	0.30	

Source: Ref 27

(A) MILITARY SPECIFICATION MLL-S-16216J.

(B) MILITARY SPECIFICATION MIL-S-24645A.

(C) MILITARY SPECIFICATIONS MIL-E-23765/2C.

TABLE 3 WELDING CONDITIONS FOR THE SUBMERGED ARC BEAD-IN-GROOVE WELDS FROM EXAMPLE 2

ENERGY INPUT		CURRENT, A	TRAVEL SPEED	
kj/mm	kj/in.		mm/s	in./s
1	25	400	12.5	0.50
2	50	450	7.8	0.31
3	75	500	5.9	0.24
4	100	570	5.1	0.20

Note: Voltage, 35 V; electrode extension, 35 mm; and preheat temperature, 93 °C (199 °F).

Source: Ref 27

Effect on HY-80 Tensile Properties. The weld-metal tensile results noted in Table 4 exhibited yield strengths that were greater than the targeted level of 56 MPa (81.9 ksi) with HY-80-1 weld deposited at 1 kJ/mm (25 k J/in.), which displayed very high yield strength and lower ductility. Upon increasing the energy input from 1 to 4 kJ/mm (25 to 101 kJ/in.), there was a pronounced decrease in yield strength (209 MPa, or 30 ksi) for the HY-80 welds.

TABLE 4 TENSILE PROPERTIES OF SUBMERGED ARC BEAD-IN-GROOVE WELDS DEPOSITED IN HY-80 STEEL FROM EXAMPLE 2

WELD NO.	ENERGY INPUT		YIELD STRENGTH		ULTIMATE STRENGTH		ELONGATION, %	REDUCTION IN AREA, %
	KJ/MM	KJ/IN.	MPA	KSI	MPA	KSI		
HY-80-1	1	25	875	127	1124	163	18	54
HY-80-2	2	50	745	108	930	135	24	59
HY-80-3	3	75	680	99	858	124	25	64
HY-80-4	4	100	666	97	867	126	25	63

Source: Ref 27

Effect on HY-80 Charpy Properties. Charpy impact transition curves for HY-80 welds based on 5 × 10 mm (0.2 × 0.4 in.) subsized specimens and plots of the percentage of ductile fracture versus test temperature are shown in Fig. 12. The investigators noted that, for HY-80 welds, the lowest energy input weld displayed the lowest toughness values, despite ductile behavior at temperatures above -80 °C (-112 °F). Figure 12 shows the variation in proportion of ductile fracture for the series of HY-80 welds. For 2 and 4 kJ/mm (5 and 10 kJ/in.) welds, ductile fracture occurred at temperatures above -50 °C (-58 °F), with the transition to brittle cleavage fracture occurring below -50 °C (-58 °F). The 1 and 3 kJ/mm (25 and 75 kJ/in.) welds displayed fully ductile fracture at temperatures above -80 °C (-112 °F).

Effect on HSLA-80 Tensile Properties. Tensile property results (Table 5) show that the HSLA-80 welds displayed lower strengths and better ductilities than the HY-80 welds. A slight decrease (45 MPa, or 6.5 ksi) was observed for the HSLA-80 welds, compared with the pronounced decreased in yield strength (209 MPa, or 30 ksi) for the HY-80 welds.

TABLE 5 PROPERTIES OF SUBMERGED ARC BEAD-IN-GROOVE WELDS DEPOSITED IN HSLA-80 STEEL FROM EXAMPLE 2

WELD NO.	ENERGY INPUT		YIELD STRENGTH		ULTIMATE STRENGTH		ELONGATION, %	REDUCTION IN AREA, %
	kj/mm	kj/in.	MPa	ksi	MPa	ksi		
HSLA-80-1	1	25	640	93	806	117	25	64
HSLA-80-2	2	50	618	90	756	110	28	66
HSLA-80-3	3	75	607	88	720	105	27	65
HSLA-80-4	4	100	595	86	710	103	28	65

Source: Ref 27

Effect on HSLA-80 Charpy Properties. In the case of the HSLA-80 welds, only slight variations in notch toughness were observed, as plotted in Fig. 18. The weld metals deposited at energy inputs of ≤3 kJ/mm (75 kJ/in.) displayed ductile behavior down to -80 °C (-112 °F). The transition to cleavage fracture occurred at temperatures between -80 and -150 °C (-112 and -238 °F), as shown in Fig. 18. The superior toughness of the HSLA-80 weld metals was generally attributed to the combined effect of the acicular ferrite microstructure and the lower yield strength.

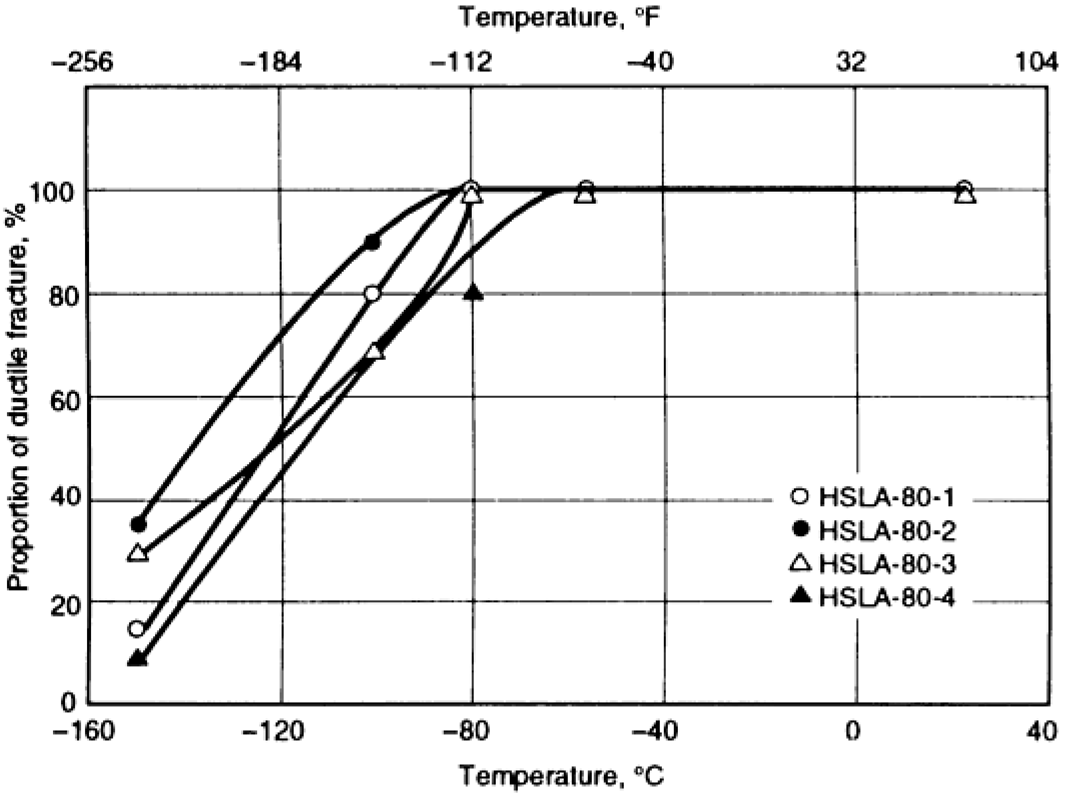

FIG. 18 PERCENT DUCTILE FRACTURE VERSUS TEST TEMPERATURE FOR THE SUBMERGED ARC BEAD-IN-GROOVE WELDS DEPOSITED IN HSLA STEEL. SOURCE: REF 27

With regard to the effect of heat input on weldment properties, the investigators concluded that:

- A PRONOUNCED DECREASE IN WELD-METAL YIELD STRENGTH AND REDUCTION IN HARDNESS WERE ALSO OBSERVED FOR THE HY-80 WELDS WITH INCREASING ENERGY INPUT. POOR LOW-TEMPERATURE NOTCH TOUGHNESS WAS DISPLAYED BY BOTH THE LOW- AND HIGH-ENERGY INPUT WELDS, WHEREAS CONSIDERABLE IMPROVEMENTS OCCURRED FOR THE 2 AND 3 KJ/MM (50 AND 75 K J/IN.) WELDS.
- THE NOTCH TOUGHNESS WAS RELATED TO THE PACKET SIZE AND THE RELATIVE STRENGTH OF THE MICROSTRUCTURES. THE EXCESSIVELY HIGH YIELD STRENGTH OF THE HY-80-1 WELD RESULTED IN LOW RESISTANCE TO BOTH DUCTILE AND CLEAVAGE FRACTURE, DESPITE FINE CLEAVAGE FACETS, WHICH CORRESPONDED CLOSELY TO THE MARTENSITIC PACKET SIZES (≤ 5 μM, OR 0.2 MIL). AT A HIGH-ENERGY INPUT, POOR TOUGHNESS OCCURRED, BECAUSE OF AN INCREASE IN THE PACKET SIZE FOR THE FULLY BAINITIC MICROSTRUCTURE.
- SUPERIOR NOTCH TOUGHNESS PROPERTIES WERE OBSERVED TO BE INDEPENDENT OF THE ENERGY INPUT FOR THE HSLA-80 WELDS AS A RESULT OF THE TRANSFORMATION TO A HIGH PROPORTION OF ACICULAR FERRITE WITH THE LOWER YIELD STRENGTHS.

References cited in this section

6. R.D. STOUT AND W.D. DOTY, *WELDABILITY OF STEELS,* 3RD ED., S. EPSTEIN AND R.E. SOMERS, ED., WELDING RESEARCH COUNCIL, NEW YORK, 1978
23. "STANDARD WELDING TERMS AND DEFINITIONS," ANSI/AWS A3.0-85, AWS
24. H.B. CARY, *MODERN WELDING TECHNOLOGY,* PRENTICE HALL, 1979

25. P.L. THREADGITT AND R.H. LEGGOTT, EFFECTS OF POSTWELD HEAT TREATMENT ON MECHANICAL PROPERTIES AND RESIDUAL STRESS LEVELS OF SUBMERGED ARC WELDING IN A C-MN-NB-AL STEEL, NOV 1984
26. I. HRIVNAK, THE THEORY OF MILD STEEL AND MICRO ALLOY STEEL WELDABILITY, 1969
27. J.A. GIANETOO, N.S. SMITH, J.T. MCGRATH, AND J.T. BOWKER, EFFECT OF COMPOSITION AND ENERGY INPUT ON STRUCTURAL AND PROPERTIES OF HIGH-STRENGTH WELD METALS, *WELD. J. RES. SUPP.*, NOV 1992

Influence of Welding on Steel Weldment Properties

C.C. Chen and Abe Pollack, Microalloying International, Inc.

References

1. Y. ITO AND K. BESSYO, "WELDABILITY OF HIGH STRENGTH STEELS RELATED TO HEAT AFFECTED ZONE CRACKING," DOCUMENT IX-567-68, IIW, 1968
2. N. YURIOKA, H. SUZUKI, AND S. OHSHITA, *WELD. J.,* VOL 62, 1983, P 1475-1535
3. O. GRONG AND O.M. AKSELSEN, HAZ TOUGHNESS OF MICROALLOYED STEELS FOR OFFSHORE, *MET. CONSTR.,* SEPT 1986, P 557
4. K.E. EASTERLIN, *INTRODUCTION TO THE PHYSICAL METALLURGY OF WELDING,* BUTTERWORTHS, 1983
5. C.F. CHOU, *THEORY AND TECHNOLOGY OF METAL WELDING,* VOL 2, MECHANICAL INDUSTRY PUBLISHER, 1981 (IN CHINESE)
6. R.D. STOUT AND W.D. DOTY, *WELDABILITY OF STEELS,* 3RD ED., S. EPSTEIN AND R.E. SOMERS, ED., WELDING RESEARCH COUNCIL, NEW YORK, 1978
7. J.F. LANCASTER, *METALLURGY OF WELDING,* GEORGE ALLEN & UNWIN, 1980
8. G.M. EVANS, FACTOR AFFECTING THE MICROSTRUCTURES AND PROPERTIES OF C-MN ALL-WELD METAL DEPOSITS, *WELD. REV.,* VOL 4, MAY 1982
9. O.M. AKSELSEN, "MICROSTRUCTURE AND TOUGHNESS IN HAZ OF BORON CONTAINING STEELS," REPORT STF A84065, THE WELDING CENTER (SINTEF), TRONDHEIM, NORWAY, 1984
10. W.H. KEARNS, *WELDING HANDBOOK,* VOL 4, AWS, 1984
11. M. GRAY, WELDABILITY OF NIOBIUM CONTAINING HIGH STRENGTH LOW ALLOY STEEL, *WELD. RES. COUNC. BULL.,* NO. 213, FEB 1976
12. R.E. DOLBY, "FACTORS CONTROLLING WELD TOUGHNESS--THE PRESENT POSITION," REPORT NO. 14, THE WELDING INSTITUTE, 1976
13. H.J. KIM AND J.G. YOUN, CHARACTERISTICS OF TMCP STEEL AND ITS SOFTENING, *WELDING METALLURGY OF STRUCTURAL STEEL,* J.Y. KOO, ED., AIME, 1987
14. Y. ITO AND M. NAKANISH, STUDY ON STRESS RELIEF CRACKING IN WELDING LOW ALLOY STEELS, *SUMITOMO SEARCH,* NO. 7, MAY 1972, P 27-36
15. C.J. MCMAHON, JR., R.J. DOBBS, AND D.H. GENTUER, STRESS RELIEF CRACKING IN MN, MO, NI, AND MN-MO-NI-CR PRESSURE VESSEL STEELS, *MATER. SCI. ENG.,* VOL 37, 1979, P 179-186
16. G.E. LINERT, *WELDING METALLURGY,* AWS
17. D.N. SHACKLETON, *WELDING HY-100 AND HY-130 STEELS,* THE WELDING INSTITUTE, 1973
18. T. TAIRA *ET AL.,* "SULFIDE CORROSION CRACKING OF GIRTH-WELDED JOINT OF LINE PIPE FOR SOUR GAS SERVICE," CORROSION '79, T-1F SYMPOSIUM (ATLANTA, GA), NACE, MARCH 1979
19. M. ITOH *ET AL.,* "HYDROGEN SULFIDE CORROSION CRACKING OF LARGE DIAMETER LINE

PIPE UP TO GRADE X-65," MIDDLE EAST CORROSION CONFERENCE (BAHRAIN), NACE, APRIL 1979

20. "SULFIDE STRESS CRACKING RESISTANT METALLIC MATERIALS FOR OIL FIELD EQUIPMENT," STANDARD MATERIAL REQUIREMENTS DOCUMENT MR0175-91, NACE
21. Z. ZACZEK, IMPROVEMENT IN THE FATIGUE STRENGTH OF BUTT WELDED JOINTS BY TIG REMELTING OF WELDING REINFORCEMENTS, *MET. CONSTR.,* JULY 1986
22. G.S. BOOTH, IMPROVING THE FATIGUE STRENGTH OF WELDED JOINT BY GRINDING--TECHNIQUES AND BENEFITS, *MET. CONSTR.,* JULY 1986
23. "STANDARD WELDING TERMS AND DEFINITIONS," ANSI/AWS A3.0-85, AWS
24. H.B. CARY, *MODERN WELDING TECHNOLOGY,* PRENTICE HALL, 1979
25. P.L. THREADGITT AND R.H. LEGGOTT, EFFECTS OF POSTWELD HEAT TREATMENT ON MECHANICAL PROPERTIES AND RESIDUAL STRESS LEVELS OF SUBMERGED ARC WELDING IN A C-MN-NB-AL STEEL, NOV 1984
26. I. HRIVNAK, THE THEORY OF MILD STEEL AND MICRO ALLOY STEEL WELDABILITY, 1969
27. J.A. GIANETOO, N.S. SMITH, J.T. MCGRATH, AND J.T. BOWKER, EFFECT OF COMPOSITION AND ENERGY INPUT ON STRUCTURAL AND PROPERTIES OF HIGH-STRENGTH WELD METALS, *WELD. J. RES. SUPP.,* NOV 1992

Introduction to the Selection of Stainless Steels

John C. Lippold, Edison Welding Institute

Introduction

STAINLESS STEELS are an important class of engineering alloys used in both wrought and cast form for a wide range of applications and in many environments. Stainless steels are used extensively in the power generation, pulp and paper, and chemical processing industries, but are also chosen for use in many everyday household and commercial products. The widespread use of stainless steels and their importance in critical industrial technologies has led to considerable investigation of the weldability and service integrity of these steels. The purpose of the articles in this Section of the Handbook is to aid the reader in the selection of stainless steels based on these criteria.

Introduction to the Selection of Stainless Steels

John C. Lippold, Edison Welding Institute

Selection of Stainless Steels

Fundamentally, stainless steels are based on the iron-chromium, iron-chromium-carbon, and iron-chromium-nickel systems, but may contain a number of other alloying additions that alter their microstructures and/or properties. The "stainless" nature of these steels arises primarily from the addition of chromium in quantities greater than 12 wt%. This level of chromium ensures that a continuous layer of protective chromium-rich oxide forms on the surface. In practice, however, stainless steels may contain as little as 9 wt% Cr and be subject to general corrosion ("rusting") at ambient temperatures. Few stainless steels contain more than 30 wt% Cr or less than 50 wt% Fe.

Historically, stainless steels have been classified by microstructure and are described as ferritic, martensitic, austenitic, or duplex (austenitic and ferritic). In addition, a number of precipitation-hardenable (PH) martensitic and austenitic stainless steels exist and are normally classified separately as PH stainless steels. Cast versions of all of the microstructural variants exist, but their welding behavior is often unique relative to the corresponding wrought alloys. The articles in this Section are organized according to the categories of stainless steels described above.

In order to control microstructure and properties, a number of alloying elements are added to the basic iron-chromium, iron-chromium-carbon, and iron-chromium-nickel systems; these alloying elements include manganese, silicon, molybdenum, niobium, titanium, and nitrogen. In order to broadly describe the effect of composition on microstructure in a wide range of stainless steels, the concept of chromium and nickel equivalents was developed to normalize the effect of these alloying additions on microstructural evolution, relative to chromium and nickel. Plotting the chromium and nickel equivalents on opposing axes provides a graphic depiction of the relationship between composition and microstructure for stainless steel welds. The Schaeffler diagram (Fig. 1) has become known as the "roadmap" of stainless steels. The compositional ranges of the ferritic, martensitic, austenitic, and duplex alloys have been superimposed on this diagram.

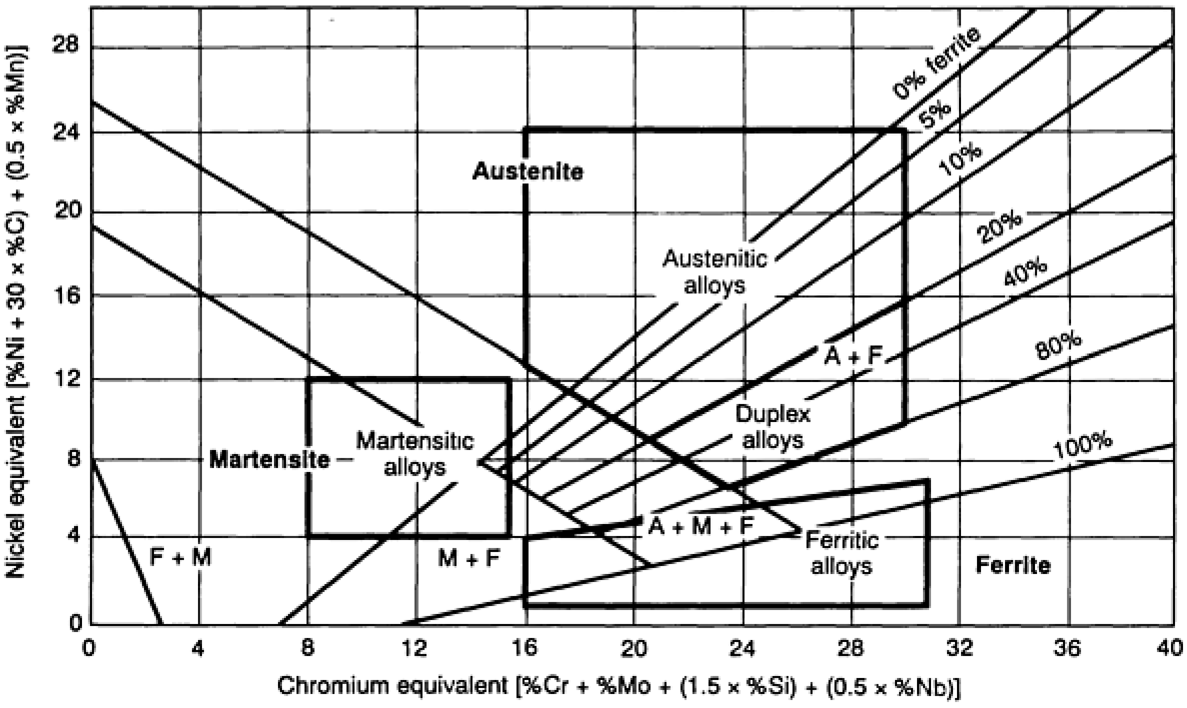

FIG. 1 SCHAEFFLER CONSTITUTION DIAGRAM FOR STAINLESS STEELS

The following articles describe the metallurgical aspects of welded stainless steels to be considered when selecting these materials for particular engineering applications and service conditions. Each article describes the microstructural evolution of the weld metal and the heat-affected zone, susceptibility to defect formation during welding, mechanical and corrosion properties, and weld process tolerance. In addition, each article lists the compositions and base metal properties for commercially available alloys. Information regarding filler metal selection and detailed process applications related to the joining of stainless steels can be found in the articles "Welding of Stainless Steels," "Brazing of Stainless Steels," and "General Soldering" in this Volume.

Selection of Wrought Martensitic Stainless Steels

J.R. Davis, Davis & Associates

Introduction

MARTENSITIC STAINLESS STEELS are essentially alloys of chromium and carbon that possess a body-centered cubic (bcc) or body-centered tetragonal (bct) crystal (martensitic) structure in the hardened condition. They are ferromagnetic and hardenable by heat treatments. They generally resist relatively mild corrosive environments.

The chromium content of these materials generally ranges from 11.5 to 18 wt%, and their carbon content can be as high as 1.2 wt%. The chromium and carbon contents are balanced to ensure a martensitic structure after hardening. Tables 1 and 2 list the chemical compositions for standard and nonstandard grades, respectively.

TABLE 1 CHEMICAL COMPOSITIONS OF STANDARD WROUGHT MARTENSITIC STAINLESS STEELS

TYPE	UNS NO.	COMPOSITION, %[(A)]							
		C	Mn	Si	Cr	Ni	P	S	Other
403	S40300	0.15	1.00	0.50	11.5-	. . .	0.04	0.03	. . .

					13.0				
410	S41000	0.15	1.00	1.00	11.5-13.5	. . .	0.04	0.03	. . .
414	S41400	0.15	1.00	1.00	11.5-13.5	1.25-2.50	0.04	0.03	. . .
416	S41600	0.15	1.25	1.00	12.0-14.0	. . .	0.06	0.15 MIN	0.6 MO$^{(B)}$
416SE	S41623	0.15	1.25	1.00	12.0-14.0	. . .	0.06	0.06	0.15 MIN SE
420	S42000	0.15 MIN	1.00	1.00	12.0-14.0	. . .	0.04	0.03	. . .
420F	S42020	0.15 MIN	1.25	1.00	12.0-14.0	. . .	0.06	0.15 MIN	0.6 MO$^{(B)}$
422	S42200	0.20-0.25	1.00	0.75	11.5-13.5	0.5-1.0	0.04	0.03	0.75-1.25 MO; 0.75-1.25 W; 0.15-0.3 V
431	S43100	0.20	1.00	1.00	15.0-17.0	1.25-2.50	0.04	0.03	. . .
440A	S44002	0.60-0.75	1.00	1.00	16.0-18.0	. . .	0.04	0.03	0.75 MO
440B	S44003	0.75-0.95	1.00	1.00	16.0-18.0	. . .	0.04	0.03	0.75 MO
440C	S44004	0.95-1.20	1.00	1.00	16.0-18.0	. . .	0.04	0.03	0.75 MO

(A) SINGLE VALUES ARE MAXIMUM VALUES UNLESS OTHERWISE INDICATED.

(B) OPTIONAL

TABLE 2 CHEMICAL COMPOSITIONS OF NONSTANDARD WROUGHT MARTENSITIC STAINLESS STEELS

DESIGNATION$^{(A)}$	**UNS NO.**	**COMPOSITION, %$^{(B)}$**							
		C	**Mn**	**Si**	**Cr**	**Ni**	**P**	**S**	**Other**
TYPE 410S	S41008	0.08	1.00	1.00	11.5-13.5	0.60	0.040	0.030	. . .
TYPE 410 CB (XM-30)	S41040	0.15	1.00	1.00	11.5-13.5	. . .	0.040	0.030	0.05-0.20 NB
HT9	DIN 1.4935$^{(C)}$	0.17-0.23	0.30-0.80	0.10-0.50	11.0-12.5	0.30-0.80	0.035	0.035	0.80-1.20 MO; 0.25-0.35 V; 0.4-0.6 W
416 PLUX X (XM-6)	S41610	0.15	1.5-2.5	1.00	12.0-14.0	. . .	0.060	0.15 MIN	0.6 MO
TYPE 418 (GREEK ASCOLLOY)	S41800	0.15-0.20	0.50	0.50	12.0-14.0	1.8-2.2	0.040	0.030	2.5-3.5 W
TRIMRITE	S42010	0.15-0.30	1.00	1.00	13.5-15.0	0.25-1.00	0.040	0.030	0.40-1.00 MO
TYPE 429 F SE	S42023	0.3-0.4	1.25	1.00	12.0-14.0	. . .	0.060	0.060	0.15 MIN SE; 0.6 ZR; 0.6 CU
LAPELLOY	S42300	0.27-0.32	0.95-1.35	0.50	11.0-12.0	0.50	0.025	0.025	2.5-3.0 MO; 0.2-0.3 V
TYPE 440 F	S44020	0.95-1.20	1.25	1.00	16.0-18.0	0.75	0.040	0.10-0.35	0.08 N

TYPE 440 F SE	S44023	0.95-1.20	1.25	1.00	16.0-18.0	0.75	0.040	0.030	0.15 MIN SE; 0.60 MO

(A) XM DESIGNATIONS IN THIS COLUMN ARE ASTM DESIGNATIONS FOR THE LISTED ALLOY.

(B) SINGLE VALUES ARE MAXIMUM VALUES UNLESS OTHERWISE INDICATED.

(C) GERMAN (DIN) SPECIFICATION

The most commonly used alloy within this stainless steel family is type 410, which contains about 12 wt% Cr and 0.1 wt% C to provide strength. The carbon level and, consequently, strength, increase in the 420, 440A, 440B, and 440C alloy series. The latter three alloys, in particular, have an increased chromium level in order to maintain corrosion resistance.

Molybdenum can be added to improve mechanical properties or corrosion resistance, as it is in type 422 stainless steel. Nickel can be added for the same reasons, as it is in types 414 and 431. When higher chromium levels are used to improve corrosion resistance, nickel also serves to maintain the desired microstructure and to prevent excessive free ferrite. The limitations on the alloy content required to maintain the desired fully martensitic structure restrict the obtainable corrosion resistance to moderate levels.

In the annealed condition, martensitic stainless steels have a tensile yield strength of approximately 275 MPa (40 ksi) and can be moderately hardened by cold working. However, martensitic alloys are typically heat treated by both hardening and tempering in order to yield strength levels up to 1900 MPa (275 ksi), depending on carbon level, primarily. These alloys have good ductility and toughness properties, which decrease as strength increases. Depending on the heat treatment, hardness values range from approximately 150 HB (80 HRB) for materials in the annealed condition to levels greater than 600 HB (58 HRC) for fully hardened materials (Fig. 1). Basic room-temperature properties of the martensitic types are given in Table 3. Typical physical properties of annealed martensitic stainless steels are listed in Table 4.

TABLE 3 MINIMUM MECHANICAL PROPERTIES OF MARTENSITIC STAINLESS STEELS

PRODUCT FORM(A)	CONDITION	TENSILE STRENGTH		0.2% YIELD STRENGTH		ELONGATION, %	REDUCTION IN AREA, %	ROCKWELL HARDNESS	ASTM SPECIFICATION
		MPA	KSI	MPA	KSI				
TYPE 403 (UNS S40300)									
B, F	ANNEALED, HOT FINISHED	485	70	275	40	20	45	. . .	A 276, A 473, A 479
B	ANNEALED, COLD FINISHED	485	70	275	40	16	45	. . .	A 276
B	INTERMEDIATE TEMPER, HOT FINISHED	690	100	550	80	15	45	. . .	A 276
B	INTERMEDIATE TEMPER, COLD FINISHED	690	100	550	80	12	40	. . .	A 276
B	HARD TEMPER, HOT OR COLD FINISHED	825	120	620	90	12	40	. . .	A 276
W	ANNEALED	485	70	275	40	20	45	. . .	A 580
W	ANNEALED, COLD FINISHED	485	70	275	40	16	45	. . .	A 580
W	INTERMEDIATE TEMPER, COLD FINISHED	690	100	550	80	12	40	. . .	A 580
W	HARD TEMPER, COLD FINISHED	825	120	620	90	12	40	. . .	A 580
P, SH, ST	ANNEALED	485	70	205	30	25(B)	. . .	88 HRB MAX	A 176
TYPE 410 (UNS S41000)									
B, F	ANNEALED, HOT FINISHED	485	70	275	40	20	45	. . .	A 276, A 473, A 479
B	ANNEALED, COLD FINISHED	485	70	275	40	16	45	. . .	A 276
B	INTERMEDIATE TEMPER, HOT FINISHED	690	100	550	80	15	45	. . .	A 276
B	INTERMEDIATE TEMPER, COLD FINISHED	690	100	550	80	12	40	. . .	A 276
B	HARD TEMPER, HOT OR COLD FINISHED	825	120	620	90	12	40	. . .	A 276
W	ANNEALED	485	70	275	40	20	45	. . .	A 580
W	ANNEALED, COLD FINISHED	485	70	275	40	16	45	. . .	A 580
W	INTERMEDIATE TEMPER, COLD FINISHED	690	100	550	80	12	40	. . .	A 580

W	HARD TEMPER, COLD FINISHED	825	120	620	90	12	40	. . .	A 580
P, SH, ST	ANNEALED	450	65	205	30	22(B)	. . .	95 HRB MAX	A 176
P, SH, ST	ANNEALED	450	65	205	30	20	. . .	95 HRB MAX	A 240
TYPE 410S (UNS S41008)									
F	ANNEALED	450	65	240	35	22	45	. . .	A 473
P, SH, ST	ANNEALED	415	60	205	30	22(B)	. . .	88 HRB MAX	A 176, A 240
TYPE 410CB (UNS S41040)									
B	ANNEALED, HOT FINISHED	485	70	275	40	13	45	. . .	A 276, A 479
B	ANNEALED, COLD FINISHED	485	70	275	40	12	35	. . .	A 276, A 479
B	INTERMEDIATE TEMPER, HOT FINISHED	860	125	690	100	13	45	. . .	A 276, A 479
B	INTERMEDIATE TEMPER, COLD FINISHED	860	125	690	100	12	35	. . .	A 276, A 479
E-4 (UNS S41050)									
P, SH, ST	ANNEALED	415	60	205	30	22	. . .	88 HRB MAX	A 276, A 240
TYPE 414 (UNS S41400)									
B	INTERMEDIATE TEMPER, COLD OR HOT FINISHED	795	115	620	90	15	45	. . .	A 276, A 479
W	ANNEALED, COLD FINISHED	1030 MAX	150 MAX	. . .	. . .	. . .	. . .	. . .	A 580
CA6NM (UNS S41500)									
P, SH, ST	TEMPERED	795	115	620	90	15	. . .	32 HRC MAX	A 176, A 240
B, F	TEMPERED	795	115	620	90	15	45	. . .	A 276, A 473, A 479
TYPES 416 (UNS S41600) AND 416SE (UNS S41623)									
F	ANNEALED	485	70	275	40	20	45	. . .	A 473
W	ANNEALED	585-860	85-125	. . .	. . .	. . .	. . .	. . .	A 581
W	INTERMEDIATE TEMPER	795-1000	115-145	. . .	. . .	. . .	. . .	. . .	A 581
W	HARD TEMPER	965-1210	140-175	. . .	. . .	. . .	. . .	. . .	A 581

TYPE 416 PLUS X (UNS S41610)									
W	ANNEALED	585-860	85-125	. . .	. . .	. . .	. . .	. . .	A 581
W	INTERMEDIATE TEMPER	795-1000	115-145	. . .	. . .	. . .	. . .	. . .	A 581
W	HARD TEMPER	965-1210	140-175	. . .	. . .	. . .	. . .	. . .	A 581
TYPE 418 (UNS S41800)									
B, F	TEMPERED AT 620 °C (1150 °F)	965	140	760	110	15	45	. . .	A 565
TYPE 420 (UNS S42000)									
B	TEMPERED AT 204 °C (400 °F)	1720	250	1480(C)	215(C)	8(C)	25(C)	52 HRC(C)	. . .
W	ANNEALED, COLD FINISHED	860 MAX	125 MAX	. . .	. . .	. . .	. . .	. . .	A 580
P, SH, ST	ANNEALED	690	100	. . .	. . .	15	. . .	96 HRB MAX	A 176
TRIMRITE (UNS S42010)									
W	ANNEALED	690 MAX	100 MAX	. . .	. . .	. . .	. . .	. . .	A 493
W	LIGHTLY DRAFTED	725 MAX	105 MAX	. . .	. . .	. . .	. . .	. . .	A 493
TYPE 422 (UNS S42200)									
B, F	TEMPERED AT 675 °C (1250 °F)	825	120	585	85	17(D)	35	. . .	A 565
B, F	TEMPERED AT 620 °C (1150 °F)	965	140	760	110	13	30	. . .	A 565
LAPELLOY (UNS S42300)									
B, F	TEMPERED AT 620 °C (1150 °F)	965	140	760	110	8	20	. . .	A 565
TYPE 431 (UNS S43100)									
F	INTERMEDIATE TEMPER	795	115	620	90	15	. . .	. . .	A 473
F	HARD TEMPER	1210	175	930	135	13	. . .	. . .	A 473
W	ANNEALED, COLD FINISHED	965 MAX	140 MAX	. . .	. . .	. . .	. . .	. . .	A 580
W	ANNEALED	760	110	. . .	. . .	. . .	. . .	. . .	A 493
W	LIGHTLY DRAFTED	795	115	. . .	. . .	. . .	. . .	. . .	A 493
TYPE 440A (UNS S44002)									
B	ANNEALED	725(C)	105(C)	415(C)	60(C)	20(C)	. . .	95 HRB(C)	. . .

B	TEMPERED AT 135 °C (600 °F)	1790(C)	260(C)	1650(C)	240(C)	5(C)	20(C)	51 HRC(C)	. . .
W	ANNEALED, COLD FINISHED	965 MAX	140 MAX	. . .	. . .	. . .	. . .	. . .	A 580
TYPE 440B (UNS S44003)									
B	ANNEALED	740(C)	107(C)	425(C)	62(C)	18(C)	. . .	96 HRB(C)	. . .
B	TEMPERED AT 315 °C (600 °F)	1930(C)	280(C)	1860(C)	270(C)	3(C)	15(C)	55 HRC(C)	. . .
W	ANNEALED, COLD FINISHED	965 MAX	140 MAX	. . .	. . .	. . .	. . .	. . .	A 580
TYPE 440C (UNS S44004)									
B	ANNEALED	760(C)	110(C)	450(C)	65(C)	14(C)	. . .	97 HRB(C)	. . .
B	TEMPERED AT 315 °C (600 °F)	1970(C)	285(C)	1900(C)	275(C)	2(C)	10(C)	57 HRC(C)	. . .
W	ANNEALED, COLD FINISHED	965 MAX	140 MAX	. . .	. . .	. . .	. . .	. . .	A 580

(A) B, BAR; F, FORGINGS; P, PLATE; SH, SHEET; ST, STRIP; W, WIRE.

(B) 20% ELONGATION FOR 1.3 MM (0.050 IN.) AND UNDER IN THICKNESS.

(C) TYPICAL VALUES.

(D) MINIMUM ELONGATION OF 15% FOR FORGINGS

TABLE 4 TYPICAL PHYSICAL PROPERTIES OF WROUGHT MARTENSITIC STAINLESS STEELS IN THE ANNEALED CONDITION

TYPE	UNS NO.	DENSITY, G/CM^3	ELASTIC MODULUS		MEAN CTE FROM 0 °C (32 °F) TO:			THERMAL CONDUCTIVITY				SPECIFIC HEAT(A)		ELECTRICAL RESISTIVITY, NΩ · M	MAGNETIC PERMEABILITY(B)	MELTING RANGE	
								AT 100 °C (212 °F)		AT 500 °C (932 °F)							
			GPA	10^6 PSI	100 °C (212 °F), 10^{-6}/K	315 °C (600 °F), 10^{-6}/K	538 °C (1000 °F), 10^{-6}/K	W/M · K	BTU/FT · H · °F	W/M · K	BTU/FT · H · °F	J/KG · K	BTU/LB · H · °F			°C	°F
410	S41000	7.8	200	29.0	9.9	11.4	11.6	24.9	14.4	28.7	16.6	460	0.11	570	700-1000	1480-1530	2700-2790
414	S41400	7.8	200	29.0	10.4	11.0	12.1	24.9	14.4	28.7	16.6	460	0.11	700	. . .	1425-1480	2600-2700
416	S41600	7.8	200	29.0	9.9	11.0	11.6	24.9	14.4	28.7	16.6	460	0.11	570	700-1000	1480-1530	2700-2790
420	S42000	7.8	200	29.0	10.3	10.8	11.7	24.9	14.4	. . .	. . .	460	0.11	550	. . .	1450-1510	2650-2750
422	S42200	7.8	. . .	. . .	11.2	11.4	11.9	23.9	13.8	27.3	15.8	460	0.11	. . .	. . .	1470-	2675-

																1480	2700
431	S43100	7.8	200	29.0	10.2	12.1	. . .	20.2	11.7	. . .	. . .	460	0.11	720	. . .	. . .	. . .
440A	S44002	7.8	200	29.0	10.2	. . .	. . .	24.2	14.0	. . .	. . .	460	0.11	600	. . .	1370-1480	2500-2700
440C	S44004	7.8	200	29.0	10.2	. . .	. . .	24.2	14.0	. . .	. . .	460	0.11	600	. . .	1370-1480	2500-2700

CTE, coefficient of thermal expansion.

(A) AT 0 TO 100 °C (932 TO 212 °F).

(B) APPROXIMATE VALUES

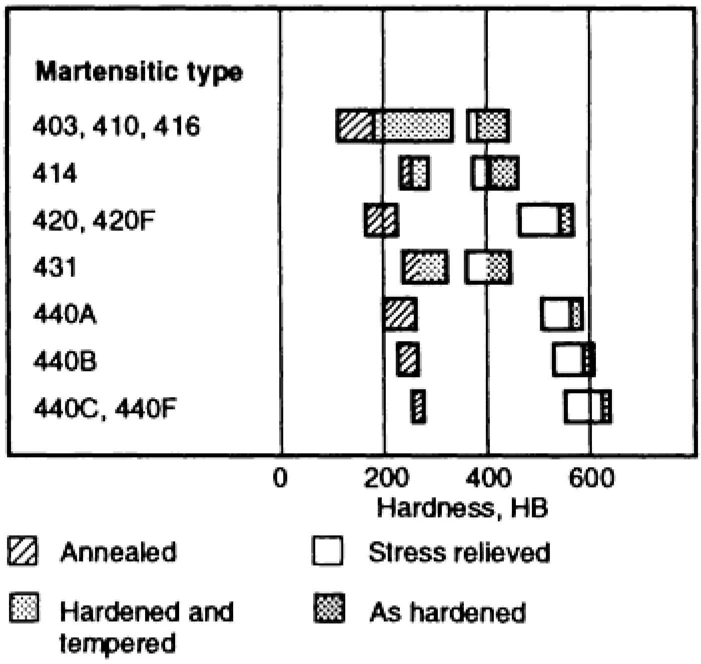

FIG. 1 RANGE OF HARDNESS IN VARIOUS MARTENSITIC STAINLESS STEELS

Martensitic stainless steels are specified when the application requires good tensile strength, creep, and fatigue strength properties, in combination with moderate corrosion resistance and heat resistance up to approximately 650 °C (1200 °F). In the United States, low- and medium-carbon martensitic steels (for example, type 410 and modified versions of this alloy) have been used primarily in steam turbines, jet engines, and gas turbines. Martensitic stainless steels are also used in petroleum and petrochemical equipment. Other applications for higher-carbon-level grades (type 440 grades) include cutlery, surgical and dental instruments, scissors, springs, valves, gears, shafts, cams, and ball bearings.

In Europe, these alloys have been used in similar applications. In addition, certain alloys, particularly 12Cr-1Mo-0.3V (HT9), have been widely used in elevated-temperature pressure-containment applications, including steam piping and steam generator reheater and superheater tubing used in fossil fuel power plants. The 12Cr-1Mo-0.3V alloy also has been investigated as a candidate material for the first wall of fusion reactors in both Europe and the United States (Ref 1, 2, 3, 4).

References

1. T. LECHTENBERG, WELDING AND FABRICATION CONSIDERATIONS FOR A 12CR-1MO MARTENSITIC STEEL FOR FUSION ENERGY SYSTEMS, *FERRITIC STEELS FOR HIGH-TEMPERATURE APPLICATIONS,* AMERICAN SOCIETY FOR METALS, 1983, P 163-177
2. J.C. LIPPOLD, TRANSFORMATION AND TEMPERING BEHAVIOR OF 12CR-1MO-0.3V MARTENSITIC STAINLESS STEEL WELDMENTS, *J. NUCL. MATER.,* VOL 103 AND 104, 1981, P 1127-1132
3. J.C. LIPPOLD, "MICROSTRUCTURAL CHARACTERIZATION OF AUTOGENOUS GTA WELDS IN A 12CR-1MO-0.3V STEEL," DOCUMENT NO. SAND80-8236, SANDIA NATIONAL LABORATORIES, OCT 1980
4. J.C. LIPPOLD, THE EFFECT OF POSTWELD HEAT TREATMENT ON THE MICROSTRUCTURE AND PROPERTIES OF THE HEAT-AFFECTED ZONE IN 12CR-1MO-0.3V (HT9) WELDMENTS, *PROC. TOPICAL CONFERENCE ON FERRITIC ALLOYS FOR USE IN NUCLEAR ENERGY TECHNOLOGIES,* TMS, 1984, P 497-506

J.R. Davis, Davis & Associates

General Welding Characteristics

As mentioned above, the compositions of martensitic stainless steels are specifically formulated to render them amenable to a quench-and-temper heat treatment in order to produce high levels of strength and hardness. The response of martensitic stainless steels to the heattreat process is essentially the same as that of plain carbon or low-alloy steels, where maximum strength and hardness primarily depend on carbon content. Although the chromium level of approximately 11 to 18 wt% in martensitic grades is the same as in ferritic stainless steels, the higher carbon content of the martensitic grades results in a complete transformation from delta (δ) ferrite to austenite at high temperature (~980 °C, or 1800 °F), followed by a subsequent change to the hard martensite phase upon rapid cooling. The thermal cycle of heating and rapid cooling, which occurs within the confined heat-affected zone (HAZ) during welding, is equivalent to a quenching cycle.

The high-carbon martensitic structure that is produced is extremely brittle in the untempered condition. Cracking can occur when the heated weldment and the surrounding martensitic HAZ are unable to contract to the same degree and at the same rate as the weld metal. Preheat and postweld cooling rates are important vehicles for controlling these shrinkage stresses and will be further discussed later in this article. Because of their response to welding thermal cycles, martensitic stainless steels are considered to be the most difficult of the five stainless steel families to weld.

The extent to which HAZ hardening is present depends on the carbon content of the base metal. Higher carbon contents (>0.15%) will produce greater hardness and, therefore, an increased susceptibility to cracking. Thus, type 440A, B, and C stainless steels are not usually considered for applications that require welding, and filler metals are not readily available in type 440 compositions.

In addition to the problems that result from localized stresses associated with the volume change upon martensitic transformation, the risk of cracking will increase when hydrogen from various sources is present in the weld metal. Therefore, the use of low-hydrogen consumables is mandatory. Hydrogen-induced cold cracking in martensitic stainless steel weldments is also discussed in more detail later in this article.

J.R. Davis, Davis & Associates

Weld Microstructure (Ref 2, 3, 4)

Detailed studies have been made on the transformation, tempering behavior, and resulting microstructures of martensitic stainless steel weldments. These studies were spurred by an interest in using 12Cr-1Mo-0.3V (HT9) martensitic stainless steel as a candidate material for the first wall/blanket of fusion-containment devices (Tokamaks), because of its improved resistance to swelling, low coefficient of thermal expansion (CTE), and high thermal conductivity relative to austenitic stainless steel (Ref 2).

During the welding of martensitic stainless steels, the fusion zone and that portion of the HAZ that is heated to a temperature higher than the austenitization temperature is transformed to a brittle, untempered martensite upon cooling to room temperature (Ref 4). Because the ductility, or fracture toughness, of the untempered martensite is inferior to that of the surrounding quenched-and-tempered microstructure, the material is unfit for service in the as-welded condition. Consequently, a postweld heat treatment (PWHT) must be employed to reduce the hardness and increase the toughness of the weld region. Recommended PWHT practices for martensitic stainless steels, including HT9, are described in the section "Specific Welding Recommendations" in this article.

The as-welded properties of the weld fusion zone and its response to PWHT can often be controlled by the appropriate choice of weld filler metal (Ref 4). However, the properties of the HAZ can be controlled only by the judicious selection of the welding process and process parameters, and by the use of a PWHT. As a result, the utilization of martensitic stainless steels depends, to a large extent, on the ability of the user to understand and control the properties of the weld HAZ.

An extensive review of the microstructure and properties of the HAZ in martensitic HT9 weldments is provided by Lippold (Ref 2, 3, 4). In his studies, a variety of autogenous bead-on-plate gas-tungsten arc welds were evaluated using both light optical microscopy and microhardness techniques. Prior to welding, the base-metal plates were austenitized for 30 min at 1040 °C (1900 °F), followed by an air cool. Tempering was performed at 760 °C (1400 °F) for 1 h. The base-metal microstructure consisted of a mixture of tempered lath martensite and alloy carbides, as shown in Fig. 2.

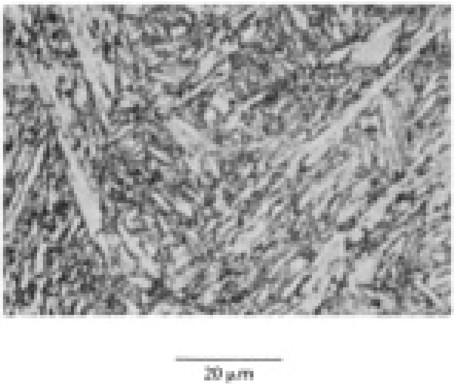

FIG. 2 QUENCHED-AND-TEMPERED BASE METAL (HT9) MICROSTRUCTURE, 22 TO 25 HRC. SOURCE: REF 2

The welding parameters used in these studies are listed in Table 5. The welding current was varied in the range from 100 to 225 A in order to produce weld heat inputs ranging from 0.71 to 2.13 kJ/mm (18 to 54 kJ/in.). The plates were clamped in a copper fixture during welding in order to stimulate the structural restraint experienced by actual components during fabrication.

TABLE 5 PARAMETERS FOR GAS-TUNGSTEN ARC WELDING OF HT9 MARTENSITIC STAINLESS STEEL

VOLTAGE, V	12-16
ARC CURRENT, A	100-225
TRAVEL SPEED, MM/S (IN./MIN)	1.69 (4)
ELECTRODE/WORK DISTANCE, MM (IN.)	2.38 (0.094)
ELECTRODE DIAMETER, MM (IN.)	3.18 (0.125)
ELECTRODE MATERIAL	W-2THO$_2$
SHIELDING GAS (ARGON), MM^3/S (FT^3/H)	365 (30)
PREHEAT	NONE

Source: Ref 2, 3

As shown in Fig. 3, the weld area consisted of the base metal (described above and shown in Fig. 2), the HAZ, and the fusion zone. Prior to tempering, the fusion zone microstructure comprised a mixture of untempered martensite and δ ferrite, as shown in Fig. 4.

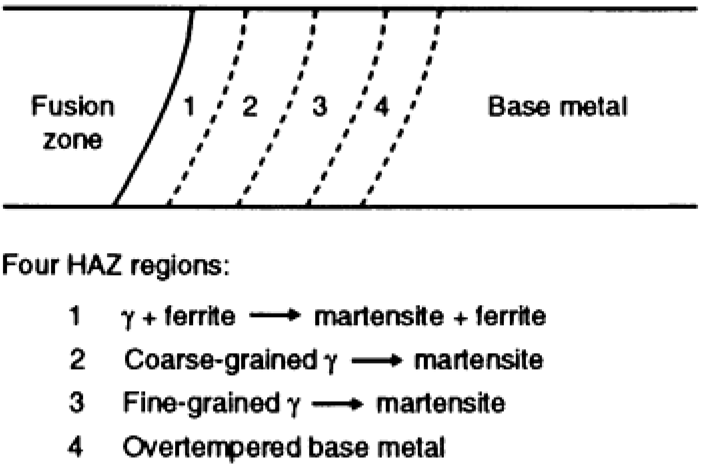
Fusion zone
1
2
3
4
Base metal
Four HAZ regions:
1 γ + ferrite ⟶ martensite + ferrite
2 Coarse-grained γ ⟶ martensite
3 Fine-grained γ ⟶ martensite
4 Overtempered base metal

FIG. 3 FOUR HAZ REGIONS OBSERVED IN HT9 WELDMENTS. γ REPRESENTS AUSTENITE. SOURCE: REF 3

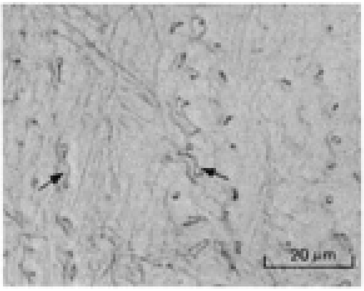

FIG. 4 AS-WELDED GAS-TUNGSTEN ARC FUSION ZONE MICROSTRUCTURE IN HT9 WELDMENTS. ARROWS INDICATE METASTABLE δFERRITE, WHICH CONSTITUTES APPROXIMATELY 2 TO 3 VOL% OF THE STRUCTURE. 1000×. SOURCE: REF 2

Four distinct regions were identified within the HAZ. These regions can be related to the phase fields on the iron-chromium-carbon equilibrium diagram shown in Fig. 5. The microstructures associated with these four regions are (Ref 4):

- *REGION 1,* WHICH IS A TWO-PHASE REGION ADJACENT TO THE FUSION LINE THAT CONSISTS OF UNTEMPERED MARTENSITE AND FERRITE. THE FERRITE FORMS ALONG PRIOR AUSTENITE GRAIN BOUNDARIES AS THIS REGION IS HEATED INTO THE AUSTENITE-PLUS-FERRITE PHASE FIELD (DESIGNATED 1 IN FIG. 3 AND 5).
- *REGION 2,* WHICH CONSISTS ENTIRELY OF UNTEMPERED MARTENSITE. AS THIS PORTION OF THE HAZ IS HEATED INTO THE HIGH-TEMPERATURE REGIME OF THE AUSTENITE PHASE FIELD, ALLOY CARBIDES FROM THE ORIGINAL BASE-METAL MICROSTRUCTURE COMPLETELY DISSOLVE, RESULTING IN A CARBIDE-FREE UNTEMPERED MARTENSITE THAT EXHIBITS A LARGE PRIOR AUSTENITE GRAIN.
- *REGION 3,* WHICH CONTAINS UNTEMPERED MARTENSITE INTERSPERSED WITH UNDISSOLVED ALLOY CARBIDES. BECAUSE THIS REGION EXPERIENCES A TEMPERATURE EXCURSION INTO THE LOWER REGIME OF THE AUSTENITE PHASE FIELD, CARBIDE DISSOLUTION IS INCOMPLETE AND THE PRIOR AUSTENITE GRAIN SIZE IS SMALLER RELATIVE TO REGION 2.
- *REGION 4,* WHICH CONSISTS OF TEMPERED MARTENSITE AND OVERAGED CARBIDES. IT APPEARS TO BE VIRTUALLY IDENTICAL TO THE BASE-METAL MICROSTRUCTURE.

Representative microstructures of these four regions are shown in Fig. 6.

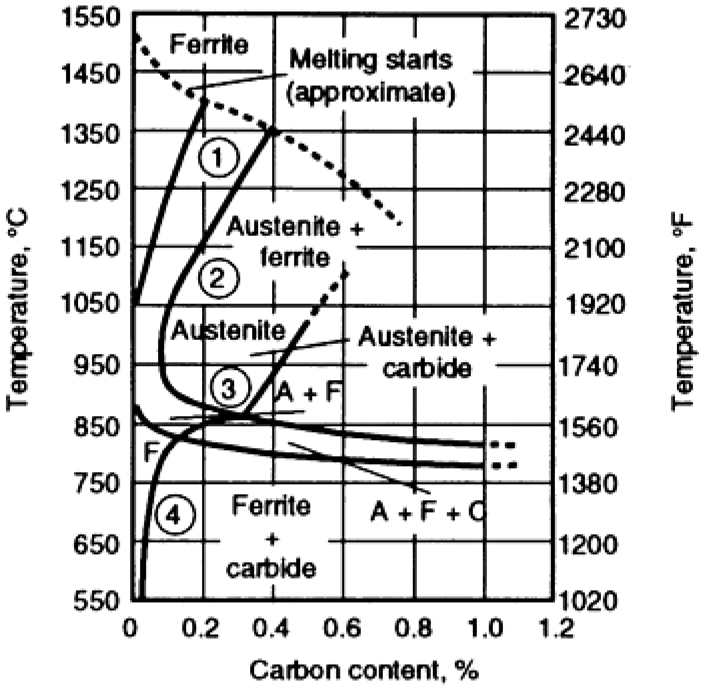

FIG. 5 IRON-CHROMIUM-CARBON PSEUDO-BINARY PHASE DIAGRAM FOR 12 WT% CR STEEL. CIRCLED NUMBERS REPRESENT THE FOUR HAZ REGIONS SHOWN IN FIG. 3 AND 6. SOURCE: ADAPTED FROM REF 5

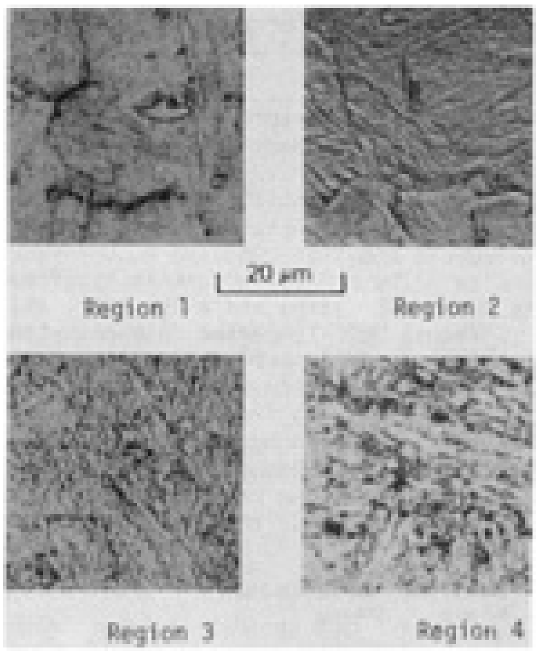

FIG. 6 GAS-TUNGSTEN ARC WELDED HAZ MICROSTRUCTURES IN HT9 WELDMENTS. SOURCE: REF 2

The microstructures in the HAZ can also be differentiated by means of a microhardness traverse, as shown in Fig. 7. The drop in hardness in Region 1 results from the presence of the softer ferrite phase situated at prior austenite grain boundaries within the untempered martensite matrix. The highest hardness in the HAZ occurs in the region where carbide dissolution is nearly complete. Because the hardness of the martensite is almost entirely a function of the carbon content of the prior austentite, Region 2 exhibits the peak HAZ hardness. Within Region 3, the carbon content of the austenite is less, because of the incomplete dissolution of the carbides at lower austenitization temperatures. Therefore, the resultant microstructure, which consists of untempered martensite and residual carbides, is softer than the adjacent untempered microstructure (Region 2). The hardness decrease in Region 4 results from overtempering the base-metal microstructure at a temperature just below Ac_1 (the temperature at which austenite begins to form during heating), which is approximately 840 °C (1545 °F).

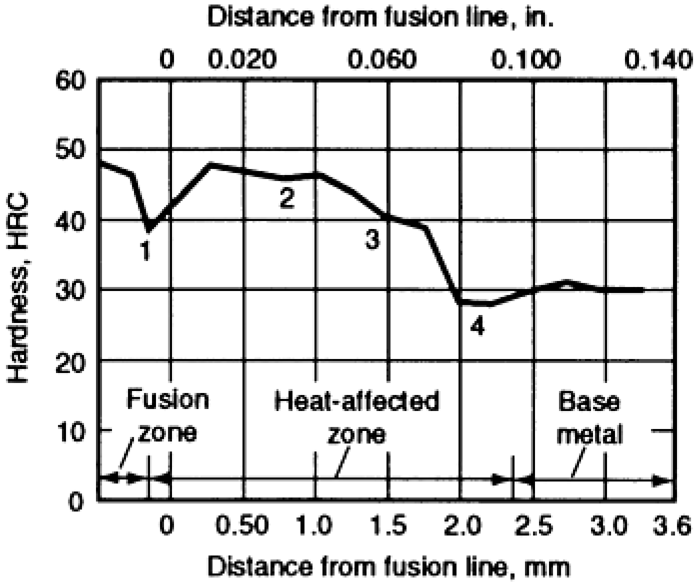

FIG. 7 MICROHARDNESS TRAVERSE ACROSS A GAS-TUNGSTEN ARC WELD IN HT9 STAINLESS STEEL (NO PREHEAT OR POSTHEAT TREATMENT). SOURCE: REF 3

Figures 8 and 9 demonstrate the effect of PWHT on the hardness and toughness, respectively, of the four HAZ regions described above. As shown in these figures, a PWHT of 760 °C (1400 °F) for 1 h produced superior results.

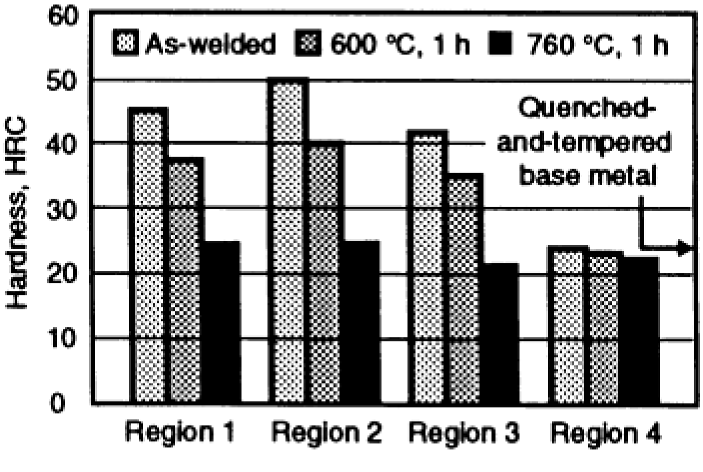

FIG. 8 HARDNESS AS A FUNCTION OF PWHT TEMPERATURE IN THE FOUR HAZ REGIONS OF HT9 WELDMENTS. SOURCE: REF 4

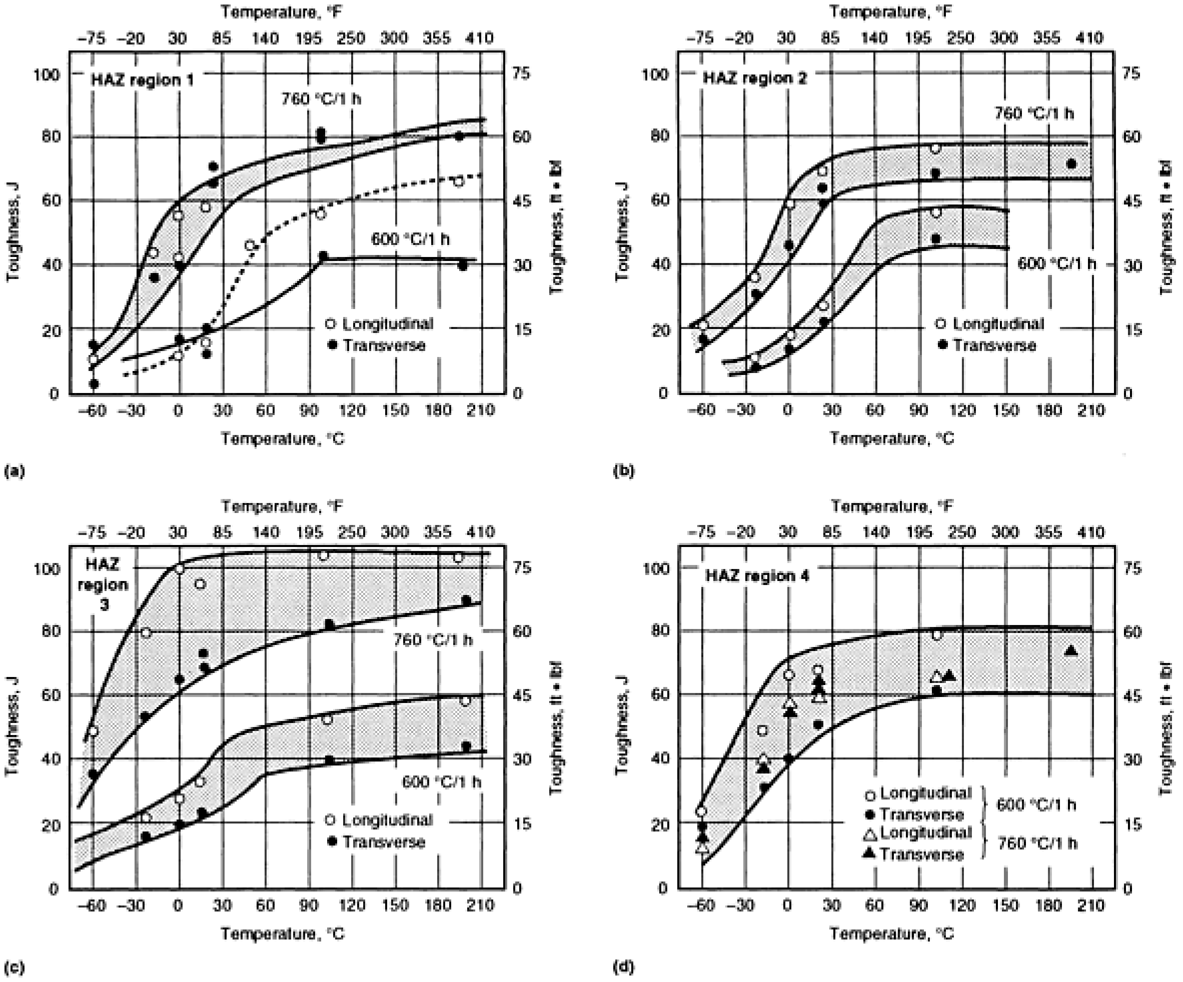

FIG. 9 EFFECT OF ORIENTATION AND PWHT ON THE TOUGHNESS OF HT9 WELDMENTS. (A) REGION 1. (B) REGION 2. (C) REGION 3. (D) REGION 4. SEE ALSO FIG. 3 AND 6. SOURCE: REF 4

Similar studies on HT9 have been carried out by Fenn and Jordan, with particular attention being given to δ ferrite formation (Ref 6). Maintaining a δ ferrite content that is as low as possible (< 10 vol%) is important, because a homogeneous martensitic structure that is free of δ ferrite islands results in a tougher weld. The investigations of these workers covered a variety of welding processes and a wide range of welding heat inputs, and involved real and simulated HAZ material.

References cited in this section

2. J.C. LIPPOLD, TRANSFORMATION AND TEMPERING BEHAVIOR OF 12CR-1MO-0.3V MARTENSITIC STAINLESS STEEL WELDMENTS, *J. NUCL. MATER.,* VOL 103 AND 104, 1981, P 1127-1132
3. J.C. LIPPOLD, "MICROSTRUCTURAL CHARACTERIZATION OF AUTOGENOUS GTA WELDS IN A 12CR-1MO-0.3V STEEL," DOCUMENT NO. SAND80-8236, SANDIA NATIONAL LABORATORIES, OCT 1980
4. J.C. LIPPOLD, THE EFFECT OF POSTWELD HEAT TREATMENT ON THE MICROSTRUCTURE AND PROPERTIES OF THE HEAT-AFFECTED ZONE IN 12CR-1MO-0.3V (HT9) WELDMENTS, *PROC. TOPICAL CONFERENCE ON FERRITIC ALLOYS FOR USE IN NUCLEAR ENERGY TECHNOLOGIES,* TMS, 1984, P 497-506
5. J.Z. BRIGGS AND T.D. PARKER, *THE SUPER 12% CR STEELS,* CLIMAX MOLYBDENUM

COMPANY, 1965
6. R. FENN AND M.F. JORDAN, MICROSTRUCTURE OF WELD HEAT-AFFECTED ZONE IN 12CR-1MO STEEL, *MET. TECHNOL.*, VOL 9 (NO. 8), 1982, P 327-337

Selection of Wrought Martensitic Stainless Steels

J.R. Davis, Davis & Associates

Weldability

Martensitic stainless steels can be welded in the annealed, hardened, and hardened-and-tempered conditions. Regardless of the prior condition of the steel, welding produces a hardened martensitic zone adjacent to the weld. In other words, the high-temperature HAZ will be in the "as-quenched" condition after welding, regardless of the prior condition of the material. In addition, the HAZ hardness is very much independent of the cooling rate over the temperature range experienced in common arc welding practices. This is evident from the isothermal transformation diagram shown in Fig. 10. Because such high hardness values render the material prone to cracking during fabrication, the selection of appropriate preheating levels and welding procedures is critical to the success of the welding process.

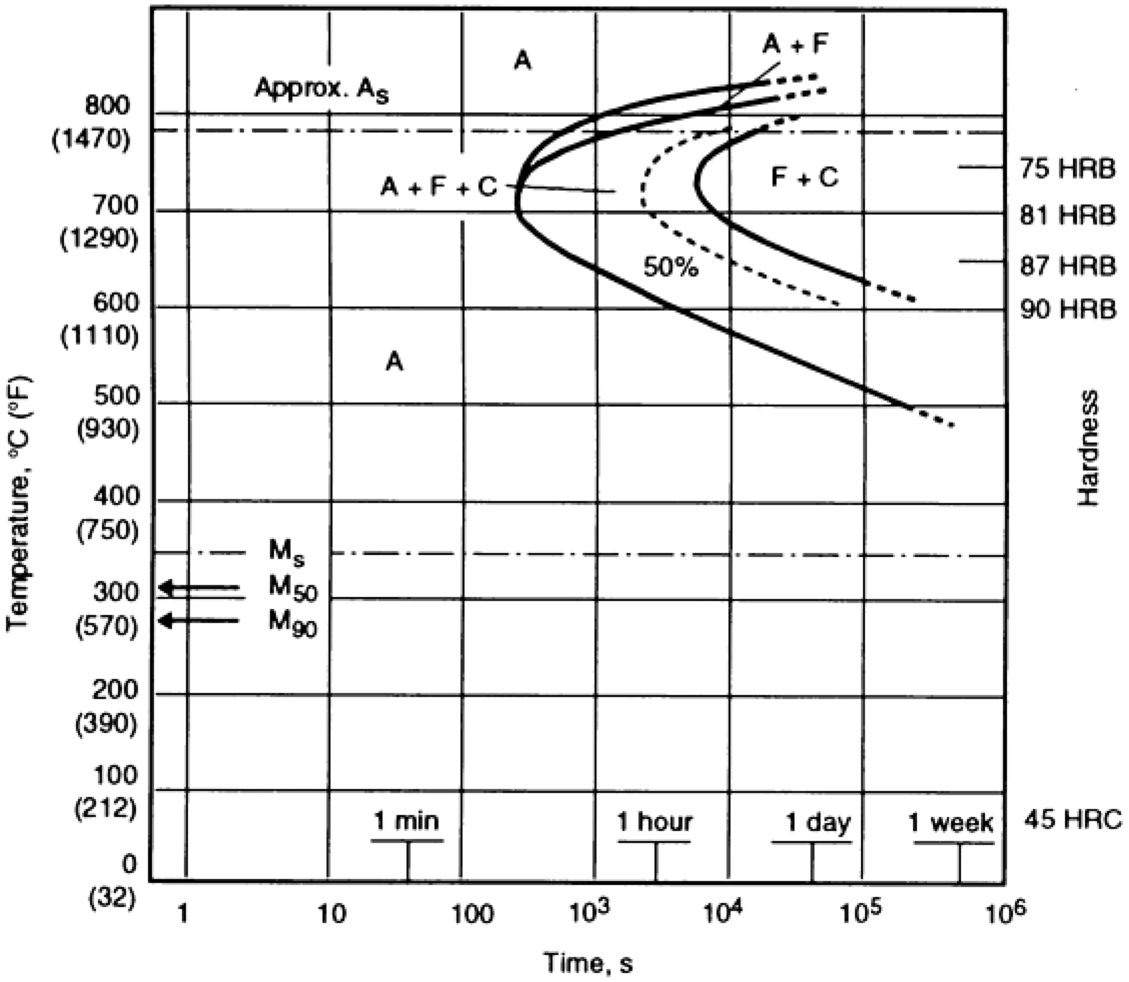

FIG. 10 ISOTHERMAL TRANSFORMATION DIAGRAM FOR TYPE 410 STAINLESS STEEL

Hydrogen-Induced Cold Cracking. Weld-area cracking in martensitic stainless steels is primarily due to the presence of hydrogen in the hardened structure (Ref 7). Hydrogen-induced cold cracks form in welds when the weldment is at or near room temperature (typically, less than 150 °C, or 300 °F). Cracking may occur almost immediately or hours after cooling. In steel, cold cracks depend on the presence of a tensile stress, a susceptible microstructure, and sufficient hydrogen in the weld. Elimination of one or more of these factors greatly reduces crack susceptibility. The stress may arise from restraint by other components of a weldment or from the simple thermal stresses created by welding in a butt, groove, or "T" joint.

The susceptibility of the microstructure to cold cracking relates to the solubility of hydrogen and the possibility of supersaturation. Austenite, in which hydrogen is highly soluble, is least susceptible to cold cracking, and martensite, in which the solubility is lower, is most susceptible, because the rapid cooling necessary for the austenite-to-martensite transformation traps the hydrogen in a state of supersaturation in the martensite. Because the solubility of hydrogen in bcc martensite is low and diffusivity is high, austenitic filler metals are often selected to weld martensitic stainless steels. The discussion on filler metal selection later in this article provides more details.

The presence of hydrogen in a weld is generally due to moisture that is introduced in the shielding gas (or the electrode coating or flux), dissociated by the arc to form elemental hydrogen, and dissolved by the molten weld pool and by the adjacent region in the HAZ. In the supersaturated state, the hydrogen diffuses to regions of high stress, where it can initiate a crack. Continued diffusion of the hydrogen to the region of stress concentration at the crack tip extends the crack. This behavior means that hydrogen-induced cold cracking is time dependent--that is, time is needed for hydrogen diffusion--and the appearance of detectable cracks can be delayed until long after the weld has passed inspection. Additional information on the mechanisms of hydrogen cracking of weldments can be found in Ref 8 and 9.

For most commercial steels, the avoidance of hydrogen-induced cold cracking is based on the control of hydrogen (Ref 7). The sources of hydrogen (water, oils, greases, waxes, and rust that contains hydrogen or hydrates) should be eliminated. The hydrogen potential also can be minimized by using low-hydrogen, inert-gas welding processes, such as gas-tungsten arc welding (GTAW) or gas-metal arc welding (GMAW), or by paying stringent attention to consumable drying and baking for flux-shielded welding processes, such as shielded metal arc welding (SMAW) or flux-cored arc welding (FCAW). In addition, preheating must be applied to slow the rate of cooling. This allows more time for hydrogen to diffuse away from the weld area during cooling within the austenite range and, especially, following transformation to martensite.

Preheating is generally carried out in the temperature range from 200 to 300 °C (400 to 600 °F). In multipass welds, interpass temperatures must be maintained at the same level, and it is frequently beneficial to hold this temperature for some time after arc extinction to permit further hydrogen diffusion out of the joint. Postweld tempering is carried out to improve weld-area toughness. For higher-carbon steels (>0.2 wt% C), this process should be applied as soon as possible after welding to avoid the possibility of hydrogen-induced cracking from atmospheric corrosion. The weld must be cooled to a sufficiently low temperature to induce the austenite-to-martensite transformation (in principle, below the martensite completion point, M_f, as described below) prior to PWHT.

Material Composition and Selection of Preheat Temperature (Ref 10). The temperature at which austenite transforms to martensite upon cooling in martensitic stainless steels depends on material composition, particularly the carbon, manganese, nickel, chromium, and molybdenum contents. The martensitic start temperature, M_s, for such steels can be calculated using the following equation developed by The Welding Institute (Ref 11):

$$M_S = 540 - (497 \times \%C + 6.3 \times \%MN + 36.3 \times \%NI + 10.8 \times \%CR + 46.6 \times \%MO)\ ^\circ C \quad \textbf{(EQ 1)}$$

It should be noted that all of the elements listed above lower the M_s temperature.

Using Eq 1 for a type 410 stainless steel of the composition listed in Table 1, the M_s temperature is approximately 350 °C (660 °F). The temperature at which the transformation to martensite is complete, denoted by M_f, is approximately 100 °C (180 °F) below the M_s, that is, 256 °C (493 °F) for type 410. The preheat level should be below the M_f wherever possible, for example, 250 °C (480 °F) for type 410. However, if the M_f is less than about 150 °C (300 °F), as is the case with high-carbon (>0.2 wt% C) steels, it is not advisable to select a lower preheat level, because the embrittling effect of hydrogen can be significant at this temperature.

In most cases, the reason for choosing a preheat temperature below the M_f is to ensure that maximum transformation to martensite occurs. If this has not happened, then subsequent PWHT immediately after welding may be ineffective for two reasons. First, the partially austenitic structure will retain hydrogen, which can lead to hydrogen cracking problems upon eventual cooling to room temperature, where untempered martensite with a high hydrogen concentration could form. Second, an isothermal transformation of the weld can take place during heat treatment, and the martensite that forms upon cooling to room temperature will be untempered and hard.

The major compositional factor influencing hydrogen-induced cold cracking is the material carbon content, because it determines the hardness of the transformed structure and its sensitivity to hydrogen embrittlement (Ref 7). Below levels of approximately 0.1 wt% C, and at least in thin sections (<5 mm, or 0.2 in.) it may be possible to weld without a preheat, if hydrogen levels are low. For materials with high carbon levels (above 0.2 wt% C), both a preheat and a PWHT are essential. Such materials are particularly difficult to weld, because the M_f temperature may be well below 150 °C (300 °F). In such cases, alloys are frequently welded with a preheat of 250 °C (480 °F), but are allowed to cool to an intermediate temperature of approximately 150 °C (300 °F) prior to a hydrogen release treatment or a full PWHT or both. For both low- and high-carbon martensitic stainless steels, it should be stressed that the weld should not be allowed to cool to room temperature after welding before either a full PWHT or a hydrogen release treatment is conducted.

Conventional martensitic stainless steels containing 12 wt% Cr and 0.15 wt% C have been successfully welded for many years, but a number of developments have reduced the risk of hydrogen-induced cold cracking. In terms of composition, this is best achieved by lowering the carbon content, except that this also reduces the hardenability of the material. Consequently, alloys that contain nickel have been produced in order to attain higher hardenability. The most common of these grades, particularly in Europe, is the 13Cr-4Ni-0.05C steels (Ref 12). Although the additional alloying reduces the M_s temperature, HAZ hydrogen-induced cold cracking resistance is substantially improved.

References cited in this section

7. R.A. WALKER AND T.G. GOOCH, ARC WELDABILITY OF STAINLESS STEELS, *MET. MATER.*, VOL 2 (NO. 1), 1986, P 18-24
8. J.H. DEVLETIAN AND W.E. WOOD, PRINCIPLES OF JOINING METALLURGY, *METALS HANDBOOK*, 9TH ED., VOL 6, AMERICAN SOCIETY FOR METALS, 1983, P 21-49
9. R.D. STOUT *ET AL.*, *WELDABILITY OF STEELS*, 3RD ED., WELDING RESEARCH COUNCIL, 1978
10. D.N. NOBLE, "REVIEW OF PRACTICES FOR WELDING TYPE 410 STAINLESS STEEL," EDISON WELDING INSTITUTE
11. T.G. GOOCH, WELDING MARTENSITIC STAINLESS STEELS, *THE WELDING INSTITUTE RESEARCH BULLETIN*, VOL 18, DEC 1977
12. E. FOLKHARD, WELDING METALLURGY OF LOW CARBON CHROMIUM-NICKEL MARTENSITIC STAINLESS STEELS, *WELDING METALLURGY OF STAINLESS STEELS*, SPRINGER-VERLAG, 1984, P 179-184

Selection of Wrought Martensitic Stainless Steels

J.R. Davis, Davis & Associates

Heat Treatment

Preheating is one of the most effective means of avoiding weld cracking in martensitic stainless steels. The preheating temperature usually ranges from 200 to 300 °C (400 to 600 °F). As discussed previously, the material composition (particularly the carbon content) affects the preheat temperature. The use of the M_s equation (Eq 1) is helpful in determining an acceptable preheat temperature.

Another consideration that influences preheating requirements is material thickness. As described above, the restraint imposed by material thickness is also one of the factors controlling hydrogen-induced cold cracking in any welded joint (Ref 10). If the weld metal or HAZ has been embrittled by the presence of hydrogen, then the level of restraint in the joint may determine whether cracking occurs. The general rule for joint thickness is that the thicker the material, the greater the restraint, that is, the more material that is present, the greater the force opposing the dimensional changes in the weld area that is brought about through thermal and transformation stresses. In a repair weld situation that involves a groove, the restraint will be very high and a conservative approach is recommended. However, for a simple butt joint involving very thin material, for example, 3 mm (0.12 in.) or less, a preheat is generally considered unnecessary. This is particularly true with low-carbon (<0.10 wt% C) base metals. However, a preheating step is typically carried out.

Postweld Heat Treatment. The functions of a PWHT are to temper the martensite in the weld metal and HAZ, in order to reduce the hardness and increase the toughness, and to decrease residual stresses associated with welding. Postweld heat treatments used for martensitic stainless steels usually take one of two forms. The most common form involves tempering the weldment by heating it below the austenite-start temperature, A_s (Fig. 10). This is generally carried out at temperatures ranging from 650 to 750 °C (1200 to 1400 °F). The weldment should be held at temperature for a minimum of 1 h per inch of weld thickness (Ref 10). Another form of PWHT involves heating the weldment above the austenite-finish temperature, A_f, to austenitize the entire mass, and then cooling it back to just above room temperature, followed by a second heat treatment at a temperature below the A_s, in order to temper the metal to the desired properties. The latter PWHT is utilized only when matching filler metals are used and maximum joint toughness must be achieved.

In addition, a low-temperature PWHT is sometimes carried out to facilitate the diffusion of hydrogen from the joint and decrease the likelihood of hydrogen-induced cold cracking (Ref 10). This type of PWHT is carried out at, or slightly above, the preheat temperature range of, for example, 300 to 400 °C (550 to 750 °F). The 450 to 500 °C (840 to 930 °F) range should be avoided because it can compromise weld toughness.

Except for very small weldments or very-low-carbon base metals, martensitic stainless steels are not typically used in the as-welded condition. If a weldment of martensitic stainless steel must be used in the as-welded condition, then it is recommended that both autogenous welds and welds with matching filler metal be avoided.

Correlations of Heat Treatment and Carbon Content. The following correlations of preheating and PWHT practice with carbon contents and welding characteristics of martensitic stainless steels can be used:

- *CARBON CONTENT BELOW 0.10%:* NEITHER PREHEATING NOR PWHT IS REQUIRED EXCEPT FOR VERY HEAVY SECTIONS; MARTENSITIC STAINLESS STEELS WITH CARBON CONTENTS THIS LOW ARE NOT STANDARD
- *CARBON CONTENT 0.10 TO 0.20%:* PREHEAT TO 260 °C (500 °F); WELD AT THIS TEMPERATURE; SLOW COOL BELOW 65 °C (150 °F); TEMPER
- *CARBON CONTENT 0.20 TO 0.50%:* PREHEAT TO 260 °C (500 °F); WELD AT THIS TEMPERATURE; SLOW COOL BELOW 65 °C (150 °F); AUSTENITIZE AND TEMPER
- *CARBON CONTENT OVER 0.50%:* PREHEAT TO 260 °C (500 °F); WELD WITH HIGH HEAT INPUT; ANNEAL WITHOUT LETTING THE WELDMENT COOL BELOW 260 °C (500 °F); AUSTENITIZE AND TEMPER

Reference cited in this section

10. D.N. NOBLE, "REVIEW OF PRACTICES FOR WELDING TYPE 410 STAINLESS STEEL," EDISON WELDING INSTITUTE

Filler-Metal Selection

On the basis of hydrogen-cracking risk, the use of austenitic filler metals is preferable to the use of martensitic consumables of "matching" composition, because the solubility of hydrogen in a bcc structure (martensite) is low and diffusivity is high, as noted earlier. This contrasts with the situation in fcc materials, such as austenite, where the solubility is high and diffusivity is low (Ref 10). It follows that if a matching martensitic filler metal is employed, then the hydrogen introduced via the welding consumable (which occurs in every case, regardless of cleanliness) will diffuse into the adjacent HAZ, where it has the opportunity to embrittle the steel. However, if an austenitic filler metal is used, then the majority of hydrogen introduced through welding will remain in the austenitic weld metal, where its solubility is high. A high hydrogen concentration in the austenitic weld metal is unlikely to cause weld-metal hydrogen-induced cracking, because the embrittling effect of hydrogen is considerably less in a predominantly austenitic weld structure than it is in a martensitic one. Suitable austenitic filler metals used for welding martensitic stainless steels include AWS E/ER308, 309, and 310, or a duplex austenitic-ferritic filler metal, such as 312.

Matching filler metals are used when the strength of the weld metal must be similar to that of the base metal or when maximum weld toughness is specified. The availability of matching filler metals is a problem for some martensitic stainless steels. Only AWS E410 and 410NiMo types are available in covered-electrode form. These compositions, as well as AWS ER420, are available as solid wires. Chemical compositions of martensitic stainless steel filler metals are given in Table 6.

TABLE 6 ALL-WELD-METAL CHEMICAL COMPOSITIONS FOR MARTENSITIC STAINLESS STEEL FILLER METALS

AWS CLASSIFICATION	COMPOSITION, %								
	C	**Cr**	**Ni**	**Mo**	**Mn**	**Si**	**P**	**S**	**Cu**
SMAW (PER AWS 5.4)									
E410	0.12	11.0-13.5	0.60	0.75	1.0	0.90	0.04	0.03	0.75
E410NIMO	0.06	11.0-12.5	4.0-5.0	0.40-7.0	1.0	0.90	0.04	0.03	0.75
GMAW, GTAW, AND SAW (PER AWS 5.9)									
ER410	0.12	11.5-13.5	0.6	0.75	0.6	0.50	0.03	0.03	0.75
ER410NIMO	0.06	11.0-12.5	4.0-5.0	0.4-0.7	0.6	0.50	0.03	0.03	0.75
ER420	0.15-0.40	12.0-14.0	0.6	0.75	0.6	0.50	0.03	0.03	0.75

Electrodes and welding rods that are suitable for use as filler metals in the welding of martensitic stainless steels are listed in Table 7. More detailed information on filler-metal selection, including case histories that review the steps necessary during the filler-metal selection process, can be found in the article "Welding of Stainless Steels" in the Section "Consumable Selection, Procedure Development, and Practice Considerations for Welding" in this Volume.

TABLE 7 FILLER METALS FOR USE IN ARC WELDING OF SELECTED MARTENSITIC STAINLESS STEELS

TYPE OF STEEL	ELECTRODE OR WELDING ROD(A)
403	410, 410NIMO, 309
410	410, 308, 309
414	410NIMO, 410, 309
416	312
420	420, 308, 309
431	410, 310
440A	312

Source: Carpenter Technology Corporation, Stoody Company, and *Metals Handbook,* 9th ed., Vol 6

(A) PREFIX E OR ER OMITTED.

Reference cited in this section

10. D.N. NOBLE, "REVIEW OF PRACTICES FOR WELDING TYPE 410 STAINLESS STEEL," EDISON WELDING INSTITUTE

Selection of Wrought Martensitic Stainless Steels

J.R. Davis, Davis & Associates

Specific Welding Recommendations

This section provides specific arc welding procedural recommendations for the more commonly welded martensitic stainless steels. The free-machining grades (types 416, 416Se, 420F, 440F, and 440FSe) are not included, because these enhanced-machining alloys are not recommended for welding. Weld data related to the welding of martensitic stainless steels are available from both alloy producers and manufacturers of welding equipment and consumables. For example, some of the data listed below were obtained from Carpenter Technology Corporation and from Sandvik. Such companies should be consulted prior to welding martensitic stainless steels.

Type 403 can be satisfactorily welded using AWS E/ER309 austenitic filler metal. If a martensitic consumable must be used, then either AWS E/ER410 or 410NiMo are acceptable. Parts should be preheated to at least 200 to 250 °C (400 to 450 °F) before welding to prevent cracking. Following welding, a PWHT of 650 to 750 °C (1200 to 1400 °F) should be conducted.

Because of the considerable cracking that can occur when welding thick sections (greater than 25 mm, or 1 in.) of type 403, a program designed to select the optimal filler metal to produce crack-free welds was conducted by Westinghouse Electric Corporation (Ref 13). This study compared the mechanical properties of AWS ER410 and 410NiMo gas-metal arc welds in 25 mm (1 in.) thick type 403 stainless steel (normalized and tempered to 585 MPa, or 85 ksi; 0.2% yield strength). Table 8 lists the welding parameters that were used in this study.

TABLE 8 PARAMETERS FOR GMAW OF TYPE 403 MARTENSITIC STAINLESS STEEL

	ELECTRODE	
	ER410NIMO	**ER410**
VOLTAGE, V	30.5	28.0
ARC CURRENT, A	460	420
TRAVEL SPEED, MM/S (IN./MIN)	3.8 (9)	5.0 (12)
ELECTRODE DIAMETER, MM (IN.)	2.38 ($\frac{3}{32}$)	2.38 ($\frac{3}{32}$)
SHIELDING GAS (ARGON), MM^3/S (FT^3/H)	600 (50)	600 (50)
PREHEAT, °C (°F)	NONE	200 TO 250 (400 TO 450)

Source: Ref 13

Weldments were tested using U-bend, tensile, stress-rupture, and Charpy V-notch tests. Results indicated that AWS ER410NiMo weldments exhibited superior tensile and impact properties (Fig. 11 and 12), but had lower stress-rupture values (Fig. 13).

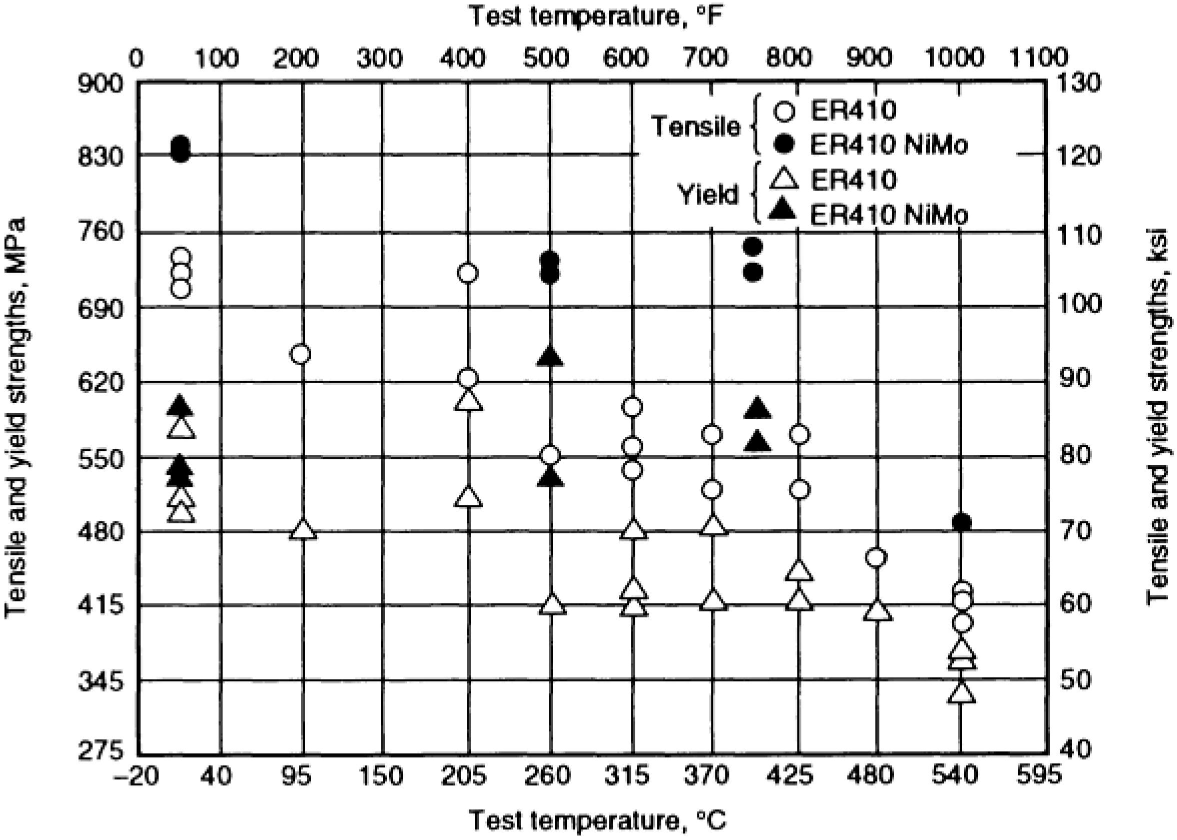

FIG. 11 STRENGTH VERSUS TEST TEMPERATURE FOR TWO MARTENSITIC STAINLESS FILLER METALS. SOURCE: REF 13

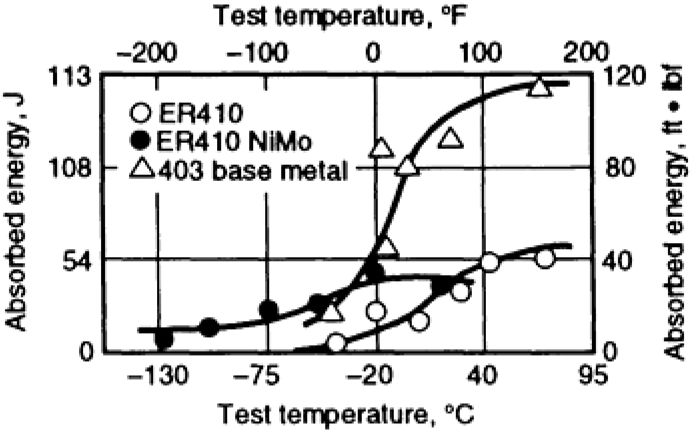

FIG. 12 CHARPY V-NOTCH VALUES VERSUS TEST TEMPERATURE FOR TWO MARTENSITIC STAINLESS STEEL FILLER METALS. SOURCE: REF 13

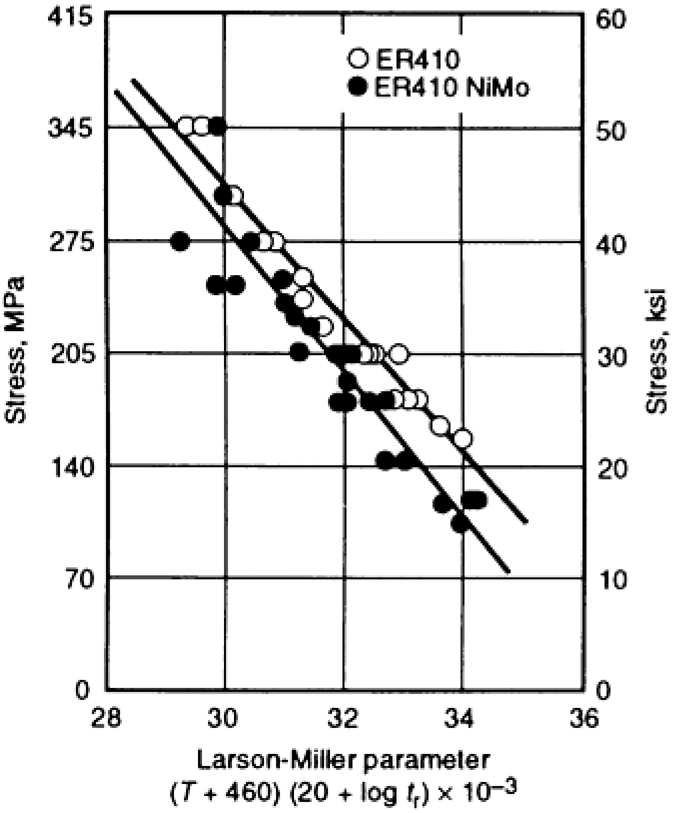

FIG. 13 RUPTURE STRESS AS A FUNCTION OF THE LARSON-MILLER PARAMETER FOR TWO MARTENSITIC STAINLESS STEEL FILLER METALS. SOURCE: REF 13

Type 410. The following recommendations for the arc welding of type 410 stainless steel are taken from Ref 10:

- PREHEAT TO 250 °C (480 °F) AND MAINTAIN THIS TEMPERATURE DURING WELDING. USE AN AUSTENITIC CONSUMABLE SUCH AS AWS E/ER308 OR 309. IF A STRONGER WELD IS DESIRED, AWS E/ER410 SHOULD BE USED.
- AFTER WELDING, ALLOW THE JOINT TO COOL TO THE PREHEAT TEMPERATURE, THAT IS, 250 °C (480 °F).
- BEFORE ALLOWING THE JOINT TO COOL TO ROOM TEMPERATURE, CONDUCT A TEMPERING OPERATION AT 720 °C (1330 °F) AND ALLOW APPROXIMATELY 1 H PER INCH OF THICKNESS, WITH A MINIMUM OF 1 H.
- COOL TO ROOM TEMPERATURE.

Type 414 can be welded using a filler metal of similar composition (AWS E/ER410 or 410NiMo) or an austenitic consumable (AWS E/ER309). Cracking is best prevented by preheating the part to at least 200 °C (400 °F) and maintaining that as a minimum interpass temperature. In addition, a PWHT that is in the range from 650 to 700 °C (1200 to 1300 °F) should be conducted.

Type 420. Satisfactory welds have been obtained in type 420 by preheating to a minimum of 200 °C (400 °F) before welding, followed by a PWHT of 675 to 750 °C (1250 to 1400 °F) (Ref 14). If a filler metal is required, then AWS ER420 should be considered when the mechanical properties of the weld are important. If the mechanical properties of the weld do not have to match those of the base metal, then AWS E/ER309 should be considered.

HT9 (12Cr-1Mo-0.3V) martensitic stainless steel can be welded using shielded-metal arc, gas-tungsten arc, or gas-metal arc welding processes (Ref 1). The joint should be preheated to a temperature between 200 and 400 °C (400 and 600 °F). In the case of multipass welds, an interpass temperature of 250 °C (480 °F) is specified to prevent the transformation to untempered martensite that occurs upon cooling to room temperature. Finally, a PWHT in the temperature range from 740 to 780 °C (1365 to 1470 °F), for a period ranging from 30 min to 2 h, is required immediately following welding. If a filler metal is used, then consumables of matching composition should be selected.

Type 431 can be welded satisfactorily with most of the arc welding processes. If a filler metal is required, then one of a similar composition, such as AWS E/ER410, should be considered. An austenitic filler metal, such as AWS E/ER310, can also be used if the mechanical properties of the weld and the response to heat treatment are not required to be similar to those of the base metal. To prevent cracking of the weldment, it is necessary to preheat the base metal to a temperature between 200 and 300 °C (400 and 600 °F) and maintain a minimum interpass temperature of 200 °C (400 °F). After welding, the weldment should receive a PWHT at 650 °C (1200 °F).

Types 440A, B, and C. Because of their high carbon contents (0.6 to 1.2 wt% C, as listed in Table 1), these alloys are not generally considered to be weldable. However, successful welds using austenitic filler metals (AWS E312) have been made in type 440A stainless steel, which has the lowest carbon content of this group (0.60 to 0.75 wt% C). A case history outlining the proper gas-metal arc welding of type 440A stainless steel is given in the article "Welding of Stainless Steels" in this Volume.

References cited in this section

1. T. LECHTENBERG, WELDING AND FABRICATION CONSIDERATIONS FOR A 12CR-1MO MARTENSITIC STEEL FOR FUSION ENERGY SYSTEMS, *FERRITIC STEELS FOR HIGH-TEMPERATURE APPLICATIONS,* AMERICAN SOCIETY FOR METALS, 1983, P 163-177
10. D.N. NOBLE, "REVIEW OF PRACTICES FOR WELDING TYPE 410 STAINLESS STEEL," EDISON WELDING INSTITUTE
13. R.E. CLARK AND L.E. WAGNER, COMPARISON OF ER410 AND ER410NIMO GAS METAL ARC WELD METAL, *CAST METALS FOR STRUCTURAL AND PRESSURE CONTAINMENT APPLICATIONS,* AMERICAN SOCIETY OF MECHANICAL ENGINEERS, 1979, P 153-169
14. *WELDING HANDBOOK,* 7TH ED., VOL 4, AMERICAN WELDING SOCIETY, 1982, P 87-93

Non-Arc Welding Processes

Although arc welding processes (particularly GMAW, GTAW, and SMAW) are commonly used to join martensitic stainless steels, a variety of other joining methods also can be employed, such as laser- and electron-beam welding, resistance welding, flash welding, and friction welding. However, regardless of the method selected, a hard (untempered) weld microstructure results from the welding process, and the same PWHTs as those discussed above must be carried out.

Laser-Beam Welding. Although the lower heat input that is typical of laser welds results in faster cooling rates and steeper temperature gradients in the weld HAZ, the problems of high hardness and low toughness in the weld area persist. Lippold has studied the microstructure of laser-beam welded martensitic stainless steels (Ref 2). The microstructure along the fusion boundary consists almost entirely of untempered martensite. The HAZ consists of two distinct microstructural regions. The first region immediately adjacent to the fusion line consists entirely of untempered martensite and extends less than 0.25 mm (0.01 in.) from the fusion line. The microstructure that is more remote from the fusion line consists of untempered martensite and carbides. The as-welded hardness of both the fusion zone and HAZ range from 52 to 56 HRC. As a result, a postweld tempering treatment is necessary. Lippold has shown that the tempering response of martensitic stainless steel laser welds is similar to that of arc welded specimens. As shown in Fig. 14, tempering for 1 h at 800 °C (1470 °F) reduced the as-welded hardness of laser welds.

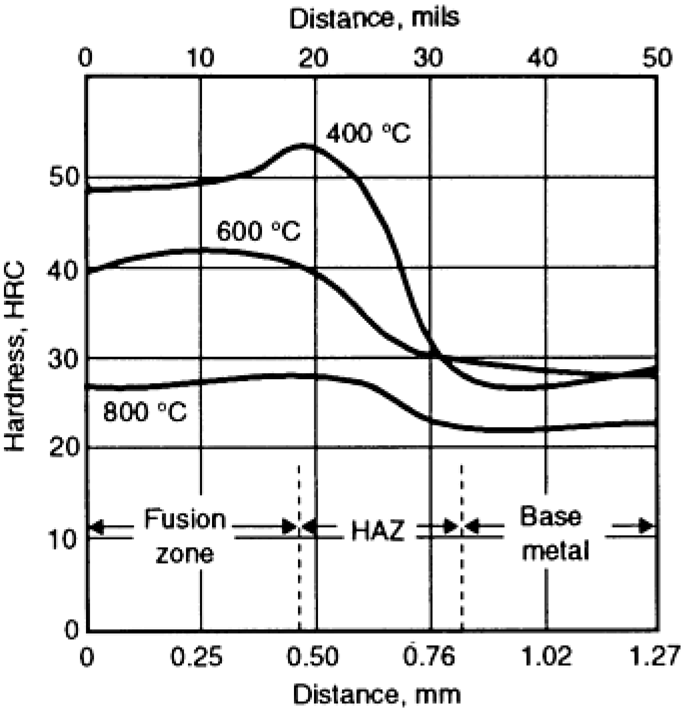

FIG. 14 TEMPERING BEHAVIOR OF HT9 LASER WELDS. SOURCE: REF 2

Lippold has also studied the effects of welding parameters on the quality of laser welds in HT9 martensitic steels (Ref 15). Laser welds were made in 6.35 mm (0.25 in.) thick HT9 plate using a variety of travel speed/focal length combinations at a constant power level of 6 kW. Welds made using low travel speed (1.27 mm/s, or 30 in./min), high heat input (0.47 kJ/mm, or 12 kJ/in.), and sharp focus produced sound welds. Welds performed at sharp focus with travel speeds ranging from 1.69 to 4.23 mm/s (40 to 100 in./min) and heat inputs decreasing from 0.35 to 0.14 kJ/mm (12 to 3.6 kJ/in.) exhibited scattered porosity and occasional centerline cracking that occurred during the final stages of solidification. Defocusing the laser beam relative to the plate resulted in welds containing severe porosity and centerline cracks. Microstructural and microhardness evaluation of the welds indicated that the region of highest hardness occurred in the HAZ immediately adjacent to the fusion line.

Laser-beam welds can be autogenous, or, alternatively, filler metals can be used. Martensitic stainless steels can be joined to each other or to dissimilar materials. For example, successful welds have been made between type 403 stainless steel and Inconel 600 using AWS A5.14 Class ENiCr-3 (Ni-20Cr-2.5Mn-2.25Nb) filler wire (Ref 16).

Electron-beam (EB) welding has been successfully conducted on modified 12 wt% Cr steels used for high-pressure, high-temperature components in power plants and in the chemical industry. In one study, EB welds were made to join 21 mm (0.83 in.) thick Fe-12Cr-1Mo plates (Ref 17). After tempering at 740 °C (1365 °F) for 2 h, the welds were investigated by Charpy V-notch and tensile tests at temperatures up to 550 °C (1020 °F). Impact properties of the EB welded joint were superior to those of submerged arc welded joints in the same alloy (Fig. 15) and exceeded those of the base metal. As is the case with laser-beam welds, the HAZ of the EB welds consisted of untempered martensite. A microhardness traverse across the weld revealed that the hardness in the fusion zone and HAZ region ranged from 300 to nearly 600 HV (30 to 55 HRC) (Fig. 16). Tempering subsequently lowered the hardness to below 300 HV.

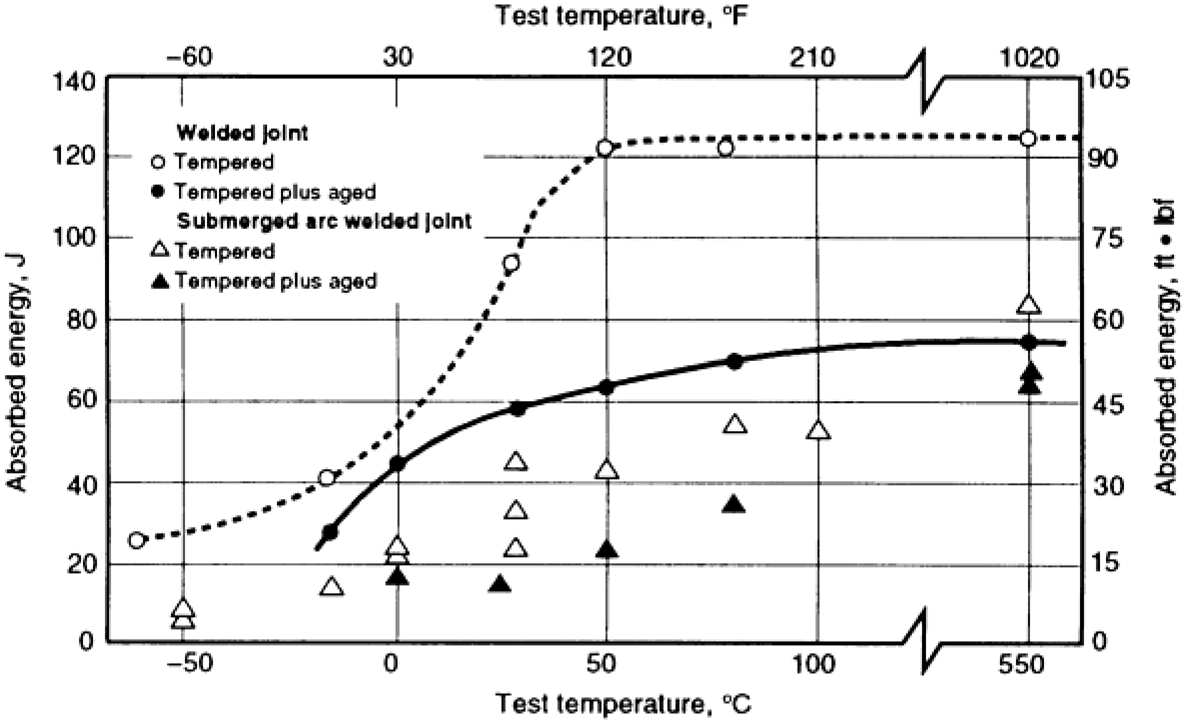

FIG. 15 CHARPY V-NOTCH DATA FOR WELD JOINTS OF EB WELDED ALLOY FE-0.2C-12CR-1MO PLATES, COMPARED WITH IMPACT DATA FOR SUBMERGED ARC WELDED JOINT USING A FILLER METAL SIMILAR TO FE-0.2C-12CR-1MO. CONDITIONS: TEMPERED AT 740 °C (1365 °F) AND TEMPERED PLUS AGED AT 550 °C (1020 °F) FOR 5000 H. SPECIMEN ORIENTATION: TRANSVERSE TO WELD. SOURCE: REF 17

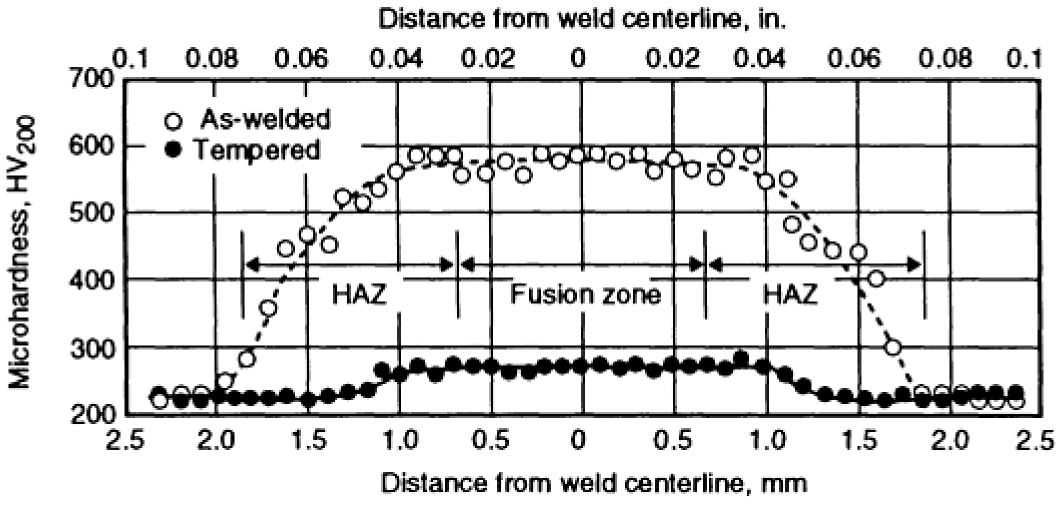

FIG. 16 MICROHARDNESS TRAVERSE ACROSS THE WELD IN TRANSVERSE CROSS SECTIONS OF EB WELDED ALLOY FE-0.2C-12CR-1MO PLATE AT THE CENTER OF PLATE THICKNESS; AS-WELDED AND CONDITIONED AND TEMPERED FOR 2 H AT 740 °C (1365 °F). SOURCE: REF 17

Friction welding can be used to join martensitic stainless steels to themselves, as well as to other steels. As is the case with other welding processes, changes in hardness at and near the weld interface can be expected (Ref 18). Figure 17 shows changes in hardness of a welded joint made of JIS S45C (0.45 wt% C) carbon steel and type 440C stainless steel.

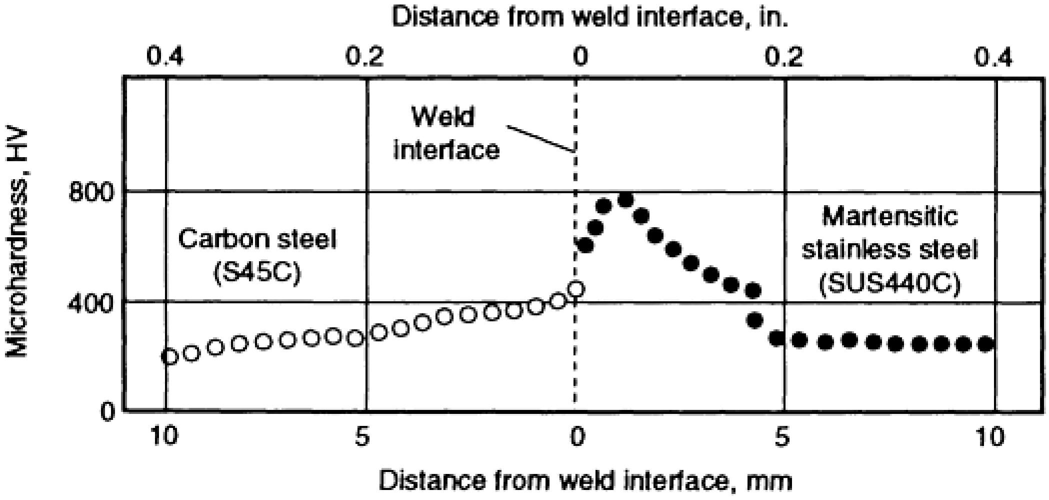

FIG. 17 CHANGES IN HARDNESS AT AND NEAR A FRICTION WELD INTERFACE. SOURCE: REF 18

Resistance Welding. Martensitic stainless steels can be spot welded in the annealed, hardened, or quenched-and-tempered conditions. Regardless of the initial base-metal hardness and the welding schedule, the HAZ adjacent to the weld nugget quenches to martensite. The hardness of the weld nugget and HAZ primarily depends on the carbon content of the steel, although it can be controlled somewhat with preheat, postheat, and tempering during the spot-welding cycle. The likelihood of cracking in the HAZ increases with the carbon content of the steel.

Flash Welding. As is the case with resistance spot welding, a hard HAZ is formed during the flash welding of martensitic stainless steels. This hard HAZ can be softened somewhat by a tempering cycle in the welding machine. A PWHT will produce welds with properties approximately equal to those of the base metal (Ref 19).

References cited in this section

2. J.C. LIPPOLD, TRANSFORMATION AND TEMPERING BEHAVIOR OF 12CR-1MO-0.3V MARTENSITIC STAINLESS STEEL WELDMENTS, *J. NUCL. MATER.,* VOL 103 AND 104, 1981, P 1127-1132
15. J.C. LIPPOLD, ANALYSIS OF WELDS IN HT9, *ALLOY DEVELOPMENT FOR IRRADIATION PERFORMANCE,* DOCUMENT DOE/ER 0045-4, U.S. DEPARTMENT OF ENERGY, 1981, P 126-139
16. G.J. BRUCK, DILUTION CONTROL IN HORIZONTAL LASER BEAM WELDING OF DISSIMILAR METALS, *THE CHANGING FRONTIERS OF LASER MATERIALS PROCESSING,* IFS PUBLICATIONS, 1987, P 149-159
17. K. KUSSMAUL *ET AL.,* MECHANICAL PROPERTIES AND STRUCTURE OF ELECTRON BEAM WELDS IN ALLOY FE-0.2%C-12%CR-1%MO, *WELD J.,* VOL 63 (NO. 9), SEPT 1984, P 267S-272S
18. A. ISHIBASHI *ET AL.,* STUDIES ON FRICTION WELDING OF CARBON AND ALLOY STEELS (3RD REPORT: ADEQUATE WELDING CONDITIONS FOR HIGH ALLOY STEELS AND DISTRIBUTIONS OF ALLOY ELEMENTS NEAR WELD INTERFACE), PAPER NO. 216, *BULL. JPN. SOC. MECH. ENG.,* VOL 26, JUNE 1983, P 1080-1087
19. L.W. WELLER AND R.H. FOXALL, FLASH WELDING, *METALS HANDBOOK,* 9TH ED., VOL 6, AMERICAN SOCIETY FOR METALS, 1984, P 557-580

Selection of Wrought Martensitic Stainless Steels

J.R. Davis, Davis & Associates

References

1. T. LECHTENBERG, WELDING AND FABRICATION CONSIDERATIONS FOR A 12CR-1MO MARTENSITIC STEEL FOR FUSION ENERGY SYSTEMS, *FERRITIC STEELS FOR HIGH-TEMPERATURE APPLICATIONS*, AMERICAN SOCIETY FOR METALS, 1983, P 163-177
2. J.C. LIPPOLD, TRANSFORMATION AND TEMPERING BEHAVIOR OF 12CR-1MO-0.3V MARTENSITIC STAINLESS STEEL WELDMENTS, *J. NUCL. MATER.*, VOL 103 AND 104, 1981, P 1127-1132
3. J.C. LIPPOLD, "MICROSTRUCTURAL CHARACTERIZATION OF AUTOGENOUS GTA WELDS IN A 12CR-1MO-0.3V STEEL," DOCUMENT NO. SAND80-8236, SANDIA NATIONAL LABORATORIES, OCT 1980
4. J.C. LIPPOLD, THE EFFECT OF POSTWELD HEAT TREATMENT ON THE MICROSTRUCTURE AND PROPERTIES OF THE HEAT-AFFECTED ZONE IN 12CR-1MO-0.3V (HT9) WELDMENTS, *PROC. TOPICAL CONFERENCE ON FERRITIC ALLOYS FOR USE IN NUCLEAR ENERGY TECHNOLOGIES,* TMS, 1984, P 497-506
5. J.Z. BRIGGS AND T.D. PARKER, *THE SUPER 12% CR STEELS*, CLIMAX MOLYBDENUM COMPANY, 1965
6. R. FENN AND M.F. JORDAN, MICROSTRUCTURE OF WELD HEAT-AFFECTED ZONE IN 12CR-1MO STEEL, *MET. TECHNOL.*, VOL 9 (NO. 8), 1982, P 327-337
7. R.A. WALKER AND T.G. GOOCH, ARC WELDABILITY OF STAINLESS STEELS, *MET. MATER.*, VOL 2 (NO. 1), 1986, P 18-24
8. J.H. DEVLETIAN AND W.E. WOOD, PRINCIPLES OF JOINING METALLURGY, *METALS HANDBOOK*, 9TH ED., VOL 6, AMERICAN SOCIETY FOR METALS, 1983, P 21-49
9. R.D. STOUT *ET AL.*, *WELDABILITY OF STEELS,* 3RD ED., WELDING RESEARCH COUNCIL, 1978
10. D.N. NOBLE, "REVIEW OF PRACTICES FOR WELDING TYPE 410 STAINLESS STEEL," EDISON WELDING INSTITUTE
11. T.G. GOOCH, WELDING MARTENSITIC STAINLESS STEELS, *THE WELDING INSTITUTE RESEARCH BULLETIN,* VOL 18, DEC 1977
12. E. FOLKHARD, WELDING METALLURGY OF LOW CARBON CHROMIUM-NICKEL MARTENSITIC STAINLESS STEELS, *WELDING METALLURGY OF STAINLESS STEELS,* SPRINGER-VERLAG, 1984, P 179-184
13. R.E. CLARK AND L.E. WAGNER, COMPARISON OF ER410 AND ER410NIMO GAS METAL ARC WELD METAL, *CAST METALS FOR STRUCTURAL AND PRESSURE CONTAINMENT APPLICATIONS,* AMERICAN SOCIETY OF MECHANICAL ENGINEERS, 1979, P 153-169
14. *WELDING HANDBOOK,* 7TH ED., VOL 4, AMERICAN WELDING SOCIETY, 1982, P 87-93
15. J.C. LIPPOLD, ANALYSIS OF WELDS IN HT9, *ALLOY DEVELOPMENT FOR IRRADIATION PERFORMANCE,* DOCUMENT DOE/ER 0045-4, U.S. DEPARTMENT OF ENERGY, 1981, P 126-139
16. G.J. BRUCK, DILUTION CONTROL IN HORIZONTAL LASER BEAM WELDING OF DISSIMILAR METALS, *THE CHANGING FRONTIERS OF LASER MATERIALS PROCESSING,* IFS PUBLICATIONS, 1987, P 149-159
17. K. KUSSMAUL *ET AL.*, MECHANICAL PROPERTIES AND STRUCTURE OF ELECTRON BEAM WELDS IN ALLOY FE-0.2%C-12%CR-1%MO, *WELD J.*, VOL 63 (NO. 9), SEPT 1984, P 267S-272S
18. A. ISHIBASHI *ET AL.*, STUDIES ON FRICTION WELDING OF CARBON AND ALLOY STEELS (3RD REPORT: ADEQUATE WELDING CONDITIONS FOR HIGH ALLOY STEELS AND DISTRIBUTIONS OF ALLOY ELEMENTS NEAR WELD INTERFACE), PAPER NO. 216, *BULL. JPN.*

SOC. MECH. ENG., VOL 26, JUNE 1983, P 1080-1087

19. L.W. WELLER AND R.H. FOXALL, FLASH WELDING, *METALS HANDBOOK*, 9TH ED., VOL 6, AMERICAN SOCIETY FOR METALS, 1984, P 557-580

Selection of Wrought Ferritic Stainless Steels

K.F. Krysiak, Hercules, Inc.; J.F. Grubb, Allegheny Ludlum Corporation; B. Pollard; R.D. Campbell, Joining Services, Inc.

Introduction

FERRITIC STAINLESS STEELS comprise approximately one-half of the SAE-AISI-type 400-series stainless steels. These steels contain from 10.5 to 30% Cr along with other alloying elements, notably molybdenum. Ferritic stainless steels are noted for their excellent stress-corrosion cracking (SCC) resistance and good resistance to pitting and crevice corrosion in chloride environments.

Following a discussion on how ferritic stainless steels can be classified, this article will review the metallurgical characteristics of various ferritic grades as well as the factors that influence their weldability. Emphasis will be placed on arc welding processes. Additional information on ferritic stainless steels can be found in the companion article "Welding of Stainless Steels" in this Volume.

Selection of Wrought Ferritic Stainless Steels

K.F. Krysiak, Hercules, Inc.; J.F. Grubb, Allegheny Ludlum Corporation; B. Pollard; R.D. Campbell, Joining Services, Inc.

Classification of Ferritic Stainless Steels

The ferritic stainless steels can be classified into three groups. These include: Group I alloys, which are the standard ferritic stainless steels; Group II alloys, which are modified versions of the standard alloys; and Group III alloys, which contain very low interstitial element (carbon, nitrogen, and oxygen) contents or stabilizing elements for improved corrosion resistance and ductility in the as-welded condition.

Group I Alloys. Ferritic stainless steels of the 400-series Group I variety have been available for many years and are used primarily for their resistance to corrosion and scaling at elevated temperatures. Table 1 lists chemical compositions of standard-grade Group I ferritic stainless steels. While these alloys have useful properties in the wrought condition, welding is known to reduce toughness and ductility and corrosion resistance because of grain coarsening and formation of martensite (Ref 1). Welding these commercial grades usually requires preheat and postweld heat treatment. For these reasons, the application of the 400-series Group I ferritic stainless steels is not as extensive as it might otherwise be, compared with the 300-series austenitic stainless steels, which enjoy ease of weldability.

TABLE 1 NOMINAL CHEMICAL COMPOSITION OF REPRESENTATIVE GROUP I STANDARD-GRADE 400-SERIES FERRITIC STAINLESS STEELS

UNS NO.	TYPE	COMPOSITION$^{(A)}$, WT%			
		C	Cr	Mo	OTHER
S42900	429	0.12	14.0-16.0	. . .	. . .
S43000	430	0.12	16.0-18.0	. . .	. . .
S43020	430F	0.12	16.0-18.0	0.6	0.06 P; 0.15 MIN S
S43023	430FSE	0.12	16.0-18.0	. . .	0.15 MIN SE
S43400	434	0.12	16.0-18.0	0.75-1.25	. . .

S43600	436	0.12	16.0-18.0	0.75-1.25	NB + TA = 5 × %C MIN
S44200	442	0.20	18.0-23.0	. . .	. . .
S44600	446	0.20	23.0-27.0	. . .	. . .

(A) SINGLE VALUES ARE MAXIMUM UNLESS OTHERWISE INDICATED

Group II Alloys. To overcome some of the difficulties and to improve weldability, several of the standard grade ferritics have been modified. These Group II ferritic stainless steels which contain lower levels of chromium and carbon, along with additions of ferrite stabilizers, are listed in Table 2. Applications of these alloys involve exposure to high temperatures, such as for quenching racks and annealing boxes, and also for tanks for agricultural sprays, transformer cases, and automotive applications.

TABLE 2 CHEMICAL COMPOSITIONS OF GROUP II FERRITIC STAINLESS STEELS

UNS NO.	ALLOY DESIGNATION	COMPOSITION(A), WT%				
		C	Cr	Mo	Ni	Other
S40500	405	0.08	11.5-14.5			0.10-0.30 AL
S40900	409	0.08	10.5-11.75	. . .	0.5	TI = 6 × C MIN TO 0.75 MAX
. . .	409CB	0.02(B)	12.5(B)	. . .	0.2(B)	0.4 NB(B)
S44100	441	0.02(B)	18.0(B)	. . .	0.3(B)	0.7 NB(B), 0.3 TI(B)
. . .	AL433	0.02(B)	19.0(B)	. . .	0.3(B)	0.4 NB(B), 0.5 SI(B), 0.4 CU(B)
. . .	AL446	0.01(B)	11.5(B)	. . .	0.2(B)	0.2 NB(B), 0.1 TI(B)
. . .	AL468	0.01(B)	18.2(B)	. . .	0.2(B)	0.2 NB(B), 0.1 TI(B)
. . .	YUS436S	0.01(B)	17.4(B)	1.2(B)	. . .	0.2 TI(B)
S43035	439	0.07	17.00-19.00	. . .	0.5	TI = 0.20 + 4 (C + N) MIN TO 1.0 MAX
. . .	12SR	0.2	12.0	. . .	. . .	1.2 AL: 0.3 TI
. . .	18SR	0.04	18.0	. . .	. . .	2.0 AL; 0.4 TI
K41970	406	0.06	12.0-14.0	. . .	0.5	2.75-4.25 AL; 0.6 TI

(A) SINGLE VALUES ARE MAXIMUM UNLESS OTHERWISE INDICATED.

(B) TYPICAL VALUE

For example, type 405, containing nominally 12% Cr, is made with lower carbon and a small aluminum addition of 0.20% to restrict the formation of austenite at high temperature so that hardening is reduced during welding. For maximum ductility and corrosion resistance, however, postweld annealing is necessary (Ref 2). Recommendations for welding include either a 430- or a 309-type filler metal, the latter being used when increased weld ductility is desired.

Type 409 is one of the most widely used ferritic stainless steels. This alloy, which has a titanium addition, is used extensively in automotive exhaust systems and is often welded using the resistance welding processes.

Group III Alloys. In the late 1960s and early 1970s, researchers recognized that highly alloyed (higher chromium and molybdenum contents) ferritic stainless steels possessed a desirable combination of resistance to general corrosion, pitting, and SCC (Ref 3, 4, 5, 6, 7, 8, 9, 10). These properties made them attractive alternatives to the 300-series austenitic stainless steels that are commonly plagued by failure as a result of chloride SCC. It was reasoned that by controlling the interstitial element content (carbon, C, nitrogen, N, and oxygen, O) of these new ferritic alloys, either by ultrahigh purity processing methods or by stabilization, the formation of martensite could be eliminated, as well as the need for preheat and postweld heat treatment, so that welds would be corrosion resistant, tough, and ductile in the as-welded condition. To achieve these results, electron-beam vacuum refining, vacuum and argon-oxygen decarburization (AOD), and vacuum induction melting processes were used. From this beginning, two basic ferritic stainless alloy systems evolved.

Ultrahigh-Purity Group III Alloys. Ferritic alloys that have an interstitial element (C + N) content less than 150 ppm are termed ultrahigh purity. Examples of these alloys, shown in Table 3, are available commercially. These alloys are produced by either electron beam vacuum refining, vacuum induction melting (VIM), or vacuum oxygen decarburization

(VOD). Plate is available up to 13 mm ($\frac{1}{2}$ in.) in thickness. Purity is especially important with the higher-chromium alloys because of the effect on ductility, toughness, and weldability. Although not mentioned in the chemistry specification for these alloys, oxygen and hydrogen are particularly harmful, and their levels must also be carefully restricted. A modest amount of niobium (columbium) is added to some of these alloys to preserve corrosion resistance in the welded condition.

TABLE 3 NOMINAL CHEMICAL COMPOSITIONS OF GROUP III ULTRAHIGH-PURITY FERRITIC STAINLESS STEELS

UNS NO.	ALLOY DESIGNATION	COMPOSITION(A), WT%						
		C	N	Cr	Mo	Ni	Nb	Other
S44726	E-BRITE 26-1 (XM-27)	0.010	0.015	25-27	0.75-1.5	0.30	0.05-0.20	0.4 MN
S44800	AL 29-4-2	0.010	0.020	28-30	3.5-4.2	2.0-2.5	. . .	. . .
S44700	AL 29-4	0.010	0.020	28-30	3.5-4.2	0.15	. . .	0.3 MN
. . .	SHOMAC 30-2	0.003(B)	0.007(B)	30(B)	2(B)	0.2(B)	. . .	0.3 MN
S44400	YUS 190L	0.004(B)	0.0085(B)	18(B)	2(B)	0.4(B)	. . .	. . .

(A) SINGLE VALUES ARE MAXIMUM UNLESS OTHERWISE STATED.

(B) TYPICAL VALUE

Intermediate-Purity Group III Alloys. When the (C + N) interstitial element content exceeds 150 ppm (but is less than 800 ppm), ferritic alloys are considered to be of intermediate purity. Alloys listed in Table 4 are available commercially and are produced typically by the AOD process. Because these steels have higher carbon and nitrogen contents, their ductility and toughness are inferior to those of the ultrahigh-purity alloys. For this reason, metal thickness is limited to a maximum of approximately 3 mm ($\frac{1}{8}$ in.) where toughness is adequate. In order to maintain acceptable corrosion resistance in the as-welded condition, titanium and/or niobium are added to tie up the carbon and nitrogen that otherwise would combine with the chromium to form chromium carbides and nitrides, which can lead to preferential intergranular attack in many environments. This form of attack is most readily observed in the weld heat-affected zone (HAZ). Although niobium and titanium are both stabilizers, they are not equivalent. Autogenous welds in titanium-containing alloys typically exhibit an equiaxed grain region near the weld centerline. Alloys stabilized only with niobium typically show a columnar weld structure with a sharply defined centerline and a propensity for weld cracking. To avoid this, one must: (1) keep the carbon content below 0.01%, (2) keep the niobium content below 0.1%, and (3) add titanium in amounts greater than 0.1%.

TABLE 4 NOMINAL CHEMICAL COMPOSITIONS OF GROUP III INTERMEDIATE-PURITY FERRITIC STAINLESS STEELS

UNS NO.	ALLOY DESIGNATION	COMPOSITION(A), WT%					
		C	N	Cr	Mo	Ni	Ti
S44626	26-1 TI	0.02(B)	0.025(B)	26(B)	1(B)	0.25(B)	0.5(B)
S44400	AISI 444	0.02(B)	0.02(B)	18(B)	2(B)	0.4(B)	0.5(B)
S44660	SEA-CURE	0.025	0.035	25-27	2.5-3.5	1.5-3.5	[0.20 + 4 (C + N)] ≤ (NB + TI) ≤0.80
S44635	NU MONIT	0.025	0.035	24.5-26	3.5-4.5	3.5-4.5	[0.20 + 4 (C + N)] ≤ (NB + TI) ≤0.80
S44735	AL 29-4C	0.030	0.045	28-30	3.6-4.2	1.0	6 (C + N) ≤ (NB + TI) ≤1.0

(A) SINGLE VALUES ARE MAXIMUM VALUES UNLESS OTHERWISE STATED.

(B) TYPICAL VALUE

Although it was claimed that the ultrahigh-purity alloys were costly to produce, many users found that the cost differential of similar alloys produced by the AOD melting practice was not economical enough to warrant purchase of these alloys in preference to similar alloys of ultrahigh purity, especially considering the inferior ductility and toughness of welded AOD-refined materials.

References cited in this section

1. *METALS HANDBOOK,* VOL 6, 8TH ED., *WELDING AND BRAZING,* AMERICAN SOCIETY FOR METALS, 1971, P 248
2. WORKING DATA, CARPENTER STAINLESS STEELS, 1971
3. W.A. MATEJKA AND R.J. KNOTH, E-BRITE 26-1, THE HIGH-PURITY APPROACH TO HIGH CHROMIUM FERRITIC STAINLESS STEELS, *ASM SYMPOSIUM,* BAL HARBOUR, FLORIDA, ASTM, 6 DEC 1973
4. R.N. WRIGHT, MECHANICAL BEHAVIOR AND WELDABILITY OF HIGH-CHROMIUM FERRITIC STAINLESS STEELS AS A FUNCTION OF PURITY, *WELD. J.,* OCT 1971, P 434S-440S
5. I.A. FRANSON, MECHANICAL PROPERTIES OF HIGH-PURITY FE-26CR-1MO FERRITIC STAINLESS STEEL, *METALL. TRANS.,* VOL 5, NOV 1974, P 2257-2264
6. J.J. DEMO, WELDABLE AND CORROSION-RESISTANT FERRITIC STAINLESS STEELS, *METALL. TRANS.,* VOL 5, NO. 11, NOV 1974, P 2253-2256
7. K.E. PINNOW, J.P. BRASSANELLI, AND A. MOSKOWITZ, A NEW HIGH-CHROMIUM FERRITIC STAINLESS STEEL, *MET. ENG. QUART.,* AUG 1975, P 32-41
8. R.F. STEIGERWALD, LOW INTERSTITIAL FE-CR-MO FERRITIC STAINLESS, *SOVIET-AMERICAN SYMPOSIUM ON NEW DEVELOPMENTS IN THE FIELD OF MOLYBDENUM-ALLOYED CAST IRON AND STEEL,* MOSCOW, JAN 1973
9. A.P. BOND, *TRANS. TMS-AIME,* VOL 245, 1969, P 2127
10. M.A. STREICHER, DEVELOPMENT OF PITTING-RESISTANT FE-CR-MO ALLOYS, *CORROSION,* VOL 30 (NO. 3), 1974, P 77-91

Selection of Wrought Ferritic Stainless Steels

K.F. Krysiak, Hercules, Inc.; J.F. Grubb, Allegheny Ludlum Corporation; B. Pollard; R.D. Campbell, Joining Services, Inc.

Metallurgical Characteristics

All of the ferritic stainless steels ideally possess the body-centered cubic (bcc) crystal structure known as ferrite at all temperatures below their melting temperatures (Fig. 1). Many of these alloys are subject to the precipitation of undesirable intermetallic phases when exposed to certain temperature ranges. The higher-chromium alloys can be embrittled by precipitation of the tetragonal sigma (σ) phase, which is based on the compound FeCr. Sigma phase has a lower temperature limit of formation (about 440 °C, or 825 °F). The σ solvus for the E-Brite alloy (see Table 3) is relatively low, probably less than 675 °C (1250 °F), and σ formation at low temperatures is sluggish. As a result, σ precipitation is rarely an issue for the E-Brite alloy during welding. The presence of increased chromium and molybdenum contents in other ultrahigh-purity ferritic stainless steels raises their σ solvus temperatures to approximately 1000 °C (1830 °F). Molybdenum also promotes formation of the complex cubic chi (χ) phase, which has a nominal composition of $Fe_{36}Cr_{12}Mo_{10}$. Sigma and χ phases are also stabilized by titanium. In the alloys other than E-Brite, these intermetallic phases form most rapidly at about 850 °C (1560 °F) and these alloys must consequently be cooled rapidly after annealing. Service exposure of welded ferritic alloys containing high levels of chromium and molybdenum should avoid these temperature ranges.

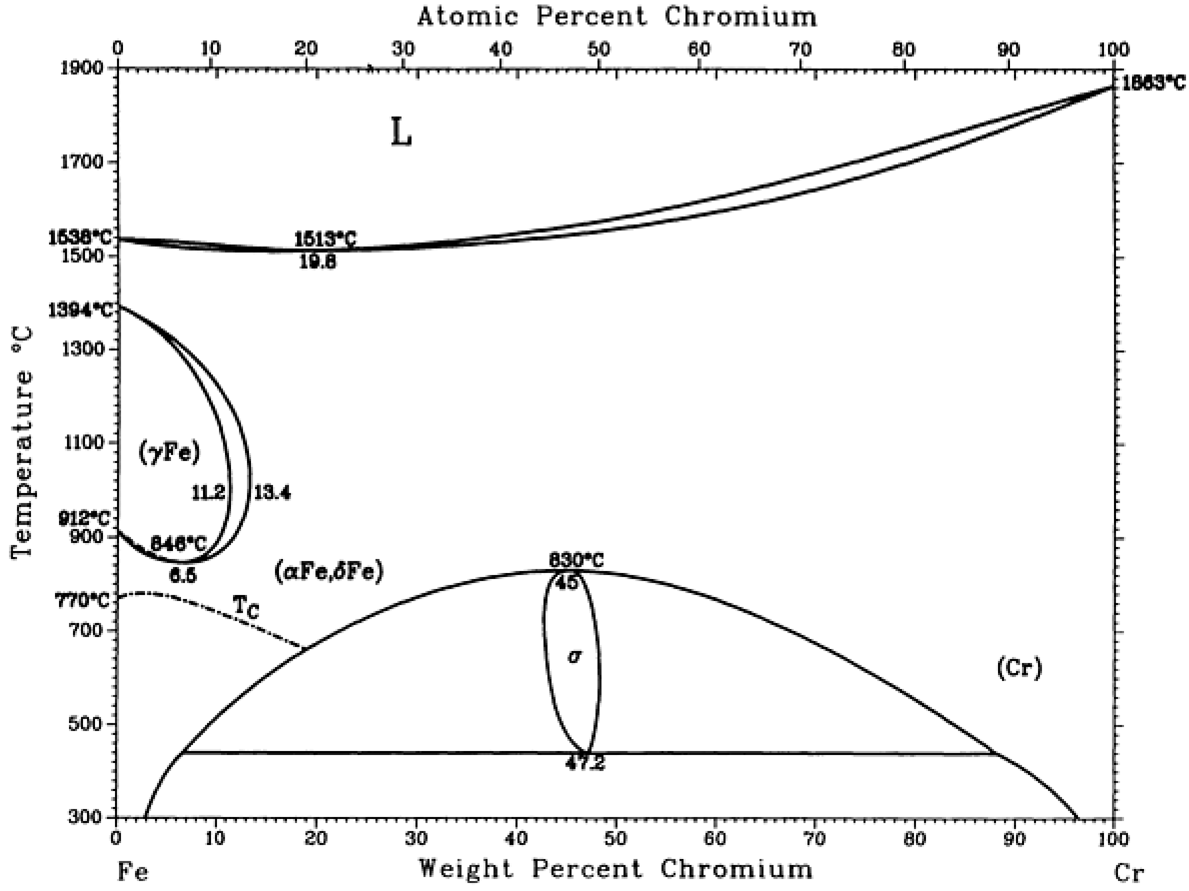

FIG. 1 THE IRON-CHROMIUM PHASE DIAGRAM. SOURCE: ASM/NIST PHASE DIAGRAM PROJECT

All of the ferritic stainless steels with chromium contents higher than 12% are also subject to embrittlement when exposed in the 370 to 550 °C (700 to 1200 °F) range. This embrittlement is most rapid at about 475 °C (885 °F) and is usually called "885 °F (475 °C) embrittlement." The rate of 475 °C (885 °F) embrittlement increases with increasing chromium plus molybdenum contents. It is generally agreed that the severe embrittlement which occurs upon long-term exposure at these temperatures is due to the decomposition of the iron-chromium ferrite phase into a mixture of iron-rich alpha (α) and chromium-rich alpha-prime (α') phases. This embrittlement is thus often called "alpha-prime embrittlement." Additional reactions such as chromium carbide and nitride precipitation may play a significant role in the more rapid, early stages of 475 °C (885 °F) embrittlement. Short-term heat treatments at 620 to 650 °C (1120 to 1210 °F) can often alleviate 885 °F embrittlement, and 885 °F embrittlement can always be cured by a full anneal heat treatment followed by rapid cooling.

The ferritic stainless steels have higher yield strengths and lower ductilities than austenitic stainless steels. Like carbon steels, and unlike austenitic stainless steels, the ferritic stainless alloys exhibit a transition from ductile-to-brittle behavior (Fig. 2) as the temperature is reduced, especially in notched impact tests. The ductile-to-brittle transition temperature (DBTT) for the ultrahigh-purity ferritic stainless steels is lower than that for type 430 and similar standard ferritic stainless steels. It is typically well below room temperature. The DBTT for the intermediate-purity stabilized ferritic stainless steels is similar to that for type 430 stainless steel. Since the DBTF decreases as metal thickness decreases, the stabilized alloys are available commercially only as sheet, strip, and light-wall tubing. The high-purity grades, because of their superior toughness, are available in thicker gages up to at least 6 mm ($\frac{1}{4}$ in.) plate. Nickel additions lower the DBTT and somewhat increase the thicknesses compatible with high toughness; nonetheless, with or without nickel, the ferritic stainless steels are not suitable for heavy-section applications.

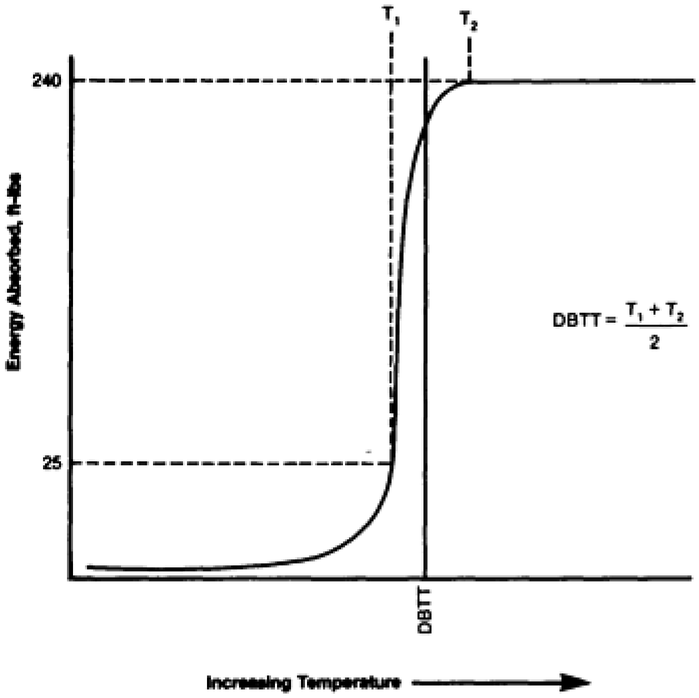

FIG. 2 SCHEMATIC OF THE DUCTILE-TO-BRITTLE IMPACT TOUGHNESS BEHAVIOR IN FERRITIC STAINLESS STEELS

In addition to the significant influence of section size (that is, thickness or mass) on the impact toughness of ferritic stainless steels, these alloys are cooling-rate sensitive (Ref 3). This effect is illustrated in Fig. 3. Depending on the cooling rate, precipitates of carbides and nitrides will tend to align along the grain boundaries or be randomly distributed in the matrix. Thus, grain size is an important factor relative to ductility and toughness. The finer the grain size, the higher the impact energy, just as in carbon or low-alloy steels.

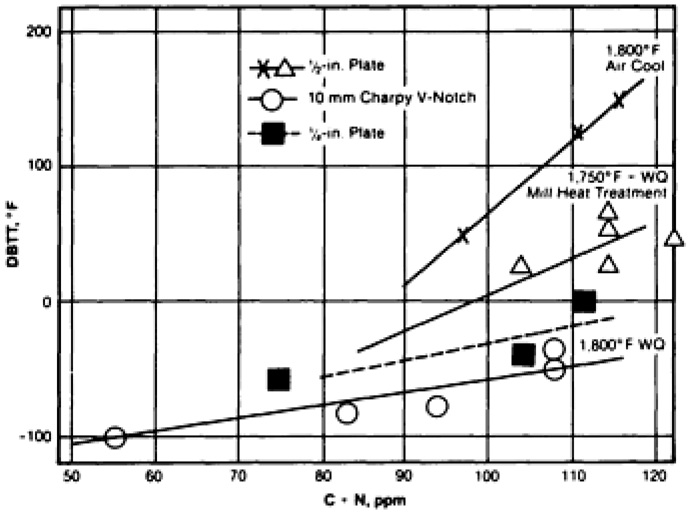

FIG. 3 THE DUCTILE-TO-BRITTLE TRANSITION TEMPERATURE (DBTT) AS A FUNCTION OF (C + N) CONTENT AND THERMAL TREATMENT. WC, WATER COOLED

Hunter and Eagar (Ref 11) investigated the mechanism of ductility loss in type 444 ferritic stainless steel (UNS S44400) welds stabilized with niobium and, in particular, titanium. It was found that stabilizing TiN or Nb(C,N) precipitates are dissolved during the autogenous gas-tungsten arc welding process resulting in a finer distribution of precipitates in the weld metal during solidification than in the base metal. The fracture appearance transition temperature (FATT) was studied using an Olsen cup test and was found to increase by almost 200 °C (360 °F) leading to decreased room-temperature ductility. Such an increase in FATT was not explainable solely in terms of grain growth. It was proposed that these finer precipitates were likely to contain chromium as well as titanium and niobium. The combination of chromium with the nitrogen releases more free titanium, which may have an embrittling effect.

Earlier, Pollard (Ref 12) studied the ductility of ferritic stainless steel weld metal in the 26 to 35% Cr range. He concluded that the most practical methods for maximizing the ductility of welds in ferritic stainless steels in gage thicknesses of at least 6 mm ($\frac{1}{4}$ in.) are:

- KEEP THE LEVEL OF BOTH SUBSTITUTIONAL AND INTERSTITIAL ALLOYING ELEMENTS AS LOW AS POSSIBLE
- MAKE ADDITIONS OF STRONG CARBONITRIDE-FORMING ELEMENTS (LOGICALLY AS LOW AS POSSIBLE)

References cited in this section

3. W.A. MATEJKA AND R.J. KNOTH, E-BRITE 26-1, THE HIGH-PURITY APPROACH TO HIGH CHROMIUM FERRITIC STAINLESS STEELS, *ASM SYMPOSIUM,* BAL HARBOUR, FLORIDA, ASTM, 6 DEC 1973
11. G.B. HUNTER AND T.W. EAGAR, DUCTILITY OF STABILIZED FERRITIC STAINLESS STEELS WELDS
12. B. POLLARD, DUCTILITY OF FERRITIC STAINLESS WELD METAL, *WELD. J.,* APR 1972, P 222S-230S

Selection of Wrought Ferritic Stainless Steels

K.F. Krysiak, Hercules, Inc.; J.F. Grubb, Allegheny Ludlum Corporation; B. Pollard; R.D. Campbell, Joining Services, Inc.

Weldability

The term "weldability" is defined here as meaning the ease with which sound welds can be made and the suitability of these welds to perform satisfactorily in service. It is necessary, therefore, that weldability include both the mechanical aspects, such as strength, ductility, and Charpy V-notch (CVN) impact toughness, and the corrosion aspects, such as resistance to intergranular attack (IGA), SCC, and general overall corrosion resistance.

The relative performance and weldability of ferritic alloys can best be determined by comparing the impact fracture behavior of these steels. Furthermore, impact testing of the HAZ as well as the weld should be included when qualifying both the welding procedure specification(s) and the welder for these materials. In addition to the CVN impact energy (which includes examination of the DBTT), the lateral expansion and percent shear fracture of CVN specimens also should be a part of a thorough analysis of weldability.

Over the years, two schools of thought have existed on how to weld the Group III ferritic stainless steels. One school recommends welding these alloys as one would the austenitics (this is not acceptable), and the other endorses using extra care and special techniques somewhat like those used to weld titanium. Extra care and special techniques are mandatory for Group III alloys. While the Group I alloys also require care, the welding of the Group II alloys is rather routine.

During the early to mid-1970s, a reluctance developed on the part of many users and fabricators to follow recommended procedures (Ref 13). As a result, many welds were produced that had inferior corrosion resistance, ductility, and impact toughness. Because of these problems, the weldability of the new ferritics was explored by many (Ref 14, 15, 16, 17, 18, 19, 20).

The unique as-welded properties of the Group III ferritics have been made possible by obtaining either very low levels of impurities, including carbon, nitrogen, and oxygen (ultrahigh-purity Grade III alloys), and a careful balance of niobium and/or titanium to match the carbon, nitrogen, and oxygen contents (intermediate-purity Group III alloys). For these reasons, every precaution must be taken and welding procedures must be selected that optimize gas shielding and cleanliness to avoid pickup of carbon, nitrogen, oxygen, and hydrogen.

Autogenous welds in ferritic stainless steels exhibit relatively simple microstructures. The grain size gradually increases from the edge of the heat-affected zone (HAZ) to the fusion boundary. In unstabilized or niobium-stabilized welds, columnar grains extend from the fusion boundary to meet at well-defined centerline (Fig. 4). In titanium-stabilized or dual-stabilized welds (niobium and titanium), there is typically a transition from columnar grains near the fusion boundary to equiaxed grains near the weld centerline (Ref 23, 24). The former structure shows a greater tendency for hot cracking (Ref 47).

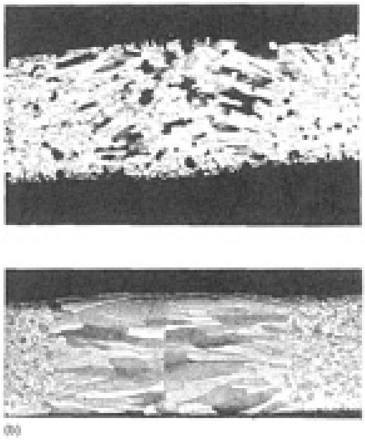

FIG. 4 PHOTOMICROGRAPH SHOWING CROSS SECTIONS OF WELDS IN 1.4 MM (0.06 IN.) SHEET. (A) 18CR-2MO-TI STEEL. (B) 18CR-2MO-NB STEEL. ETCHANT USED WAS X20 (HYDROCHLORIC-NITRIC-GLYCEROL SOLUTION). SOURCE: REF 21, 22

With lower chromium or high-carbon-content ferritic stainless alloys, such as types 409, 430, 434, 442, and 446, martensite formation during welding can occur as illustrated in Fig. 5. In the as-welded condition, the DBTT will easily be above room temperature with increased susceptibility to weld cracking during and after cooling from the molten state. Caution is advised when making welds under high restraint or in heavier section sizes. Preheat can be used under these conditions to slow the cooling rate and minimize stresses that can lead to cracking.

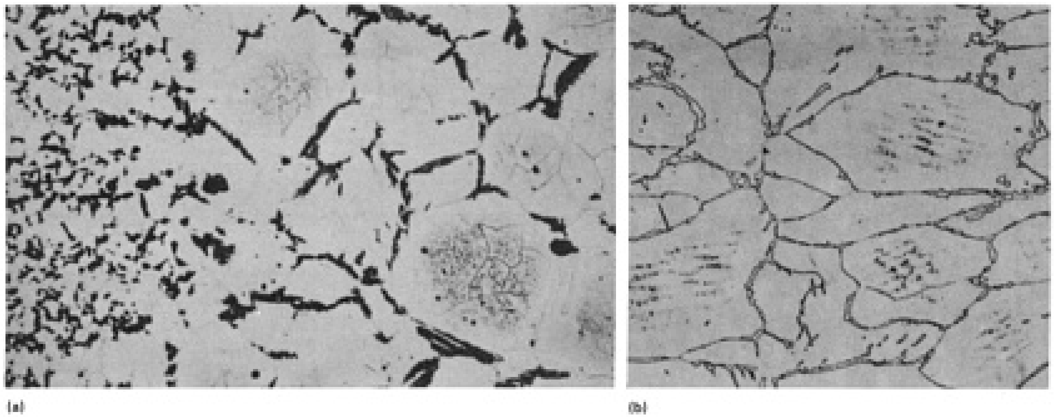

FIG. 5 GRAIN BOUNDARY MARTENSITE FORMATION IN A TYPE 430 FERRITIC STAINLESS STEEL GAS-TUNGSTEN ARC WELD. (A) FUSION ZONE. 100×. (B) HEAT-AFFECTED ZONE. 150×. SOURCE: REF 47, 48

A further word of caution is advised regarding selection of preheat and welding parameters. One must not forget that while higher heat inputs and preheats can reduce weld cracking in some ferritic stainless steels, grain growth in the weld HAZ can occur. Figure 6 shows grain growth in an ultrahigh-purity Fe-28Cr-2Mo alloy in the base metal and HAZ. Excessive grain growth can produce loss in fracture toughness, ductility, and corrosion resistance.

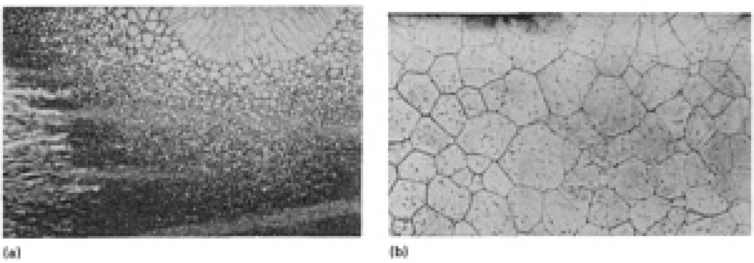

FIG. 6 CROSS SECTIONS OF PARTIAL PENETRATION GAS-TUNGSTEN ARC WELDS IN HIGH-PURITY FE-28CR-5MO FERRITIC STAINLESS STEEL. (A) WELD IN WARM-ROLLED SHEET. (B) WELD IN SHEET WHICH WAS PREWELD ANNEALED AT 1040 °C (1900 °F) FOR 60 MIN. ETCHED IN 40% NITRIC ACID ELECTROETCH. 11×

Under some conditions, annealing after welding can be performed to eliminate martensite, however the ferrite grains that have coarsened may grow even more. Generally annealing after welding is costly, can cause distortion, and usually is not practical in the field. Annealing of tubular products at the steel mill is the exception and is performed as standard procedure. This tube annealing is performed at high rates and is followed by quenching.

Regardless of which welding process is used, joint preparation and thorough solvent degreasing (using a solvent that does not leave a residue) on both sides of the joint (inside and outside) are of paramount importance.

Welding Processes

Gas-Tungsten Arc Welding (GTAW). When welding with the GTAW process, direct current electrode negative (DCEN), or straight polarity should be used. A tungsten-2% thoria electrode (AWS classification EWTh-2) with a tapered tip is recommended. Oxidation of the electrode tip or filler metal pickup are not acceptable as they reduce arc stability. When this occurs, welding must be stopped and the electrode redressed. To optimize weld quality, a gas lens collet body and the largest size ceramic nozzle should be used. With higher speed automatic processes, a trailing gas shield is recommended to prevent excessive oxidation of the weld bead.

Argon, helium, or mixtures of the two can be used for shielding. Only high-purity gas should be used. Standard welding grade gas may not be adequate because of cylinder contamination, and standard grade cylinders are not certified as to purity level. On full penetration welds or when welding thinner gages where oxidation of the back-side can occur, back gas shielding with a high-quality gas is mandatory. When welding the austenitic grades of stainless alloys, it is common to use nitrogen backing gas (usually to reduce costs, or enhance corrosion performance with these alloys). Nitrogen backing gas *must never be used* when welding ferritics because nitrogen pickup will cause severe embrittlement (loss of toughness and ductility) and loss of corrosion resistance. Inert backing gas should be used for at least two layers of deposited weld metal (3 mm, or $\frac{1}{8}$ in., minimum). Another common technique that is used when welding austenitics is to butt two pieces of metal together, weld from one side, then back grind the backside to sound metal and complete the joint. This procedure must *never* be attempted when welding ferritics because all of the contaminated metal may not be completely removed. The result is a defective, crack-sensitive ferritic weld.

To enhance weld quality and performance, especially for the Group III alloys, the following are recommended:

- THOROUGHLY DEGREASE THE WELD JOINT AND ADJACENT SURFACES BACK AT LEAST 50 MM (2 IN.) ON BOTH SIDES OF EACH JOINT MEMBER. THIS MUST ALSO INCLUDE THE WELD FILLER WIRE. WELDING SHOULD ONLY BE PERFORMED WITH CLEAN GLOVES. DIRTY AND SWEATY (DAMP) WELDING GLOVES MUST NEVER BE USED.
- KEEP THE WELDING TORCH AT RIGHT ANGLES TO THE WORK SURFACE WITHIN 15° OF VERTICAL TO PREVENT AIR ASPIRATION INTO THE INERT GAS SHIELD.
- GAS LENS SCREENS THAT BECOME CONTAMINATED WITH A DROP OR MORE OF WELD SPATTER MUST BE CHANGED IMMEDIATELY BECAUSE THE DISRUPTION OF GAS FLOW FROM THE TORCH NOZZLE CAUSES AIR ASPIRATION INTO THE MOLTEN WELD POOL. REMOVAL OF CONTAMINATED WELD METAL IS ADVISED. USUALLY, THE LENS MUST BE CHANGED.
- WEAVING THE ELECTRODE IS NOT PERMITTED, EXCEPT FOR A SMALL AMOUNT TO ALLOW GOOD SIDE-WALL FUSION AND TIE-IN. IT SHOULD BE NOTED THAT EXCESSIVE WEAVING CAUSES CONTAMINATION OF THE WELD POOL WITH A SUBSEQUENT LOSS IN FRACTURE TOUGHNESS, DUCTILITY, AND CORROSION RESISTANCE.
- WHEN WELDING, KEEP THE TIP OF THE FILLER WIRE WITHIN THE SHIELDING GAS ENVELOPE--ANOTHER REASON TO USE A GAS LENS AND AS LARGE A GAS CUP AS POSSIBLE. IF THE FILLER WIRE TIP BECOMES CONTAMINATED, EITHER BECAUSE IT TOUCHED THE TUNGSTEN ELECTRODE OR WAS WITHDRAWN OUTSIDE THE INERT GAS ENVELOPE, WELDING MUST STOP, THE TIP OF THE CONTAMINATED WIRE CUT BACK TO SOUND METAL, AND THE CONTAMINATED WELD, IF ANY, GROUND OUT, AND REPAIRED.
- ALL HOSE CONNECTIONS MUST BE OF GOOD QUALITY AND BE TIGHT TO PREVENT AIR INGESTION. CHECK ALL WATER SEALS IN THE TORCH. AIR DRAFTS IN THE WELDING AREA ARE TO BE AVOIDED TO PREVENT LOSS OF THE INERT GAS SHIELD. VERIFY THE INTEGRITY OF THE GAS SYSTEM WITH A WELD BUTTON TEST. THIS IS DONE BY STRIKING AN ARC ON A TEST PIECE OF E-BRITE PLATE OR OTHER HIGH-PURITY FERRITIC, ESTABLISHING A SMALL CIRCULAR POOL OF MOLTEN METAL APPROXIMATELY 13 MM ($\frac{1}{2}$ IN.) IN DIAMETER, EXTINGUISHING THE ARC, AND POSTPURGING AT LEAST SIX SECONDS OR MORE. THE WELD BUTTON MUST BE BRIGHT AND SHINY AND FREE OF ANY HEAT TINT. IF NOT, THE SOURCE OF CONTAMINATION MUST BE FOUND AND CORRECTED.
- PREPURGE THE TORCH BEFORE WELDING TO REMOVE AIR AND/OR MOISTURE WHICH HAS INFILTRATED INTO THE SYSTEM. POSTPURGE ALL WELD STOPS TO PREVENT CONTAMINATION. THE POSTPURGE TIME SHOULD BE SUFFICIENT SUCH THAT THE WELD STOP IS BRIGHT AND SHINY AND FREE OF ANY OXIDE HEAT TINT OR DISCOLORATION.
- A REMOTE CONTROL SHOULD BE USED TO ALLOW GRADUAL INCREASE AND DECREASE OF CURRENT WHEN INITIATING OR EXTINGUISHING THE ARC.
- AVOID OVERHEATING AND EMBRITTLING THE WELD BY MINIMIZING HEAT INPUT, AND IN MULTIPASS WELDS, BY KEEPING THE INTERPASS TEMPERATURE BELOW 95 °C (200 °F).
- WHEN HIGH-SPEED AUTOMATIC WELDING IS PERFORMED, A TRAILING GAS SHIELD IS RECOMMENDED.

Gas-Metal Arc Welding (GMAW). With the GMAW process, direct current electrode positive (DCEP), or reverse polarity, is used. The following methods of metal transfer across the welding arc are possible with GMAW:

- *SPRAY TRANSFER* IS CHARACTERIZED BY HIGH HEAT INPUT AND A FLUID WELD POOL. EXCESSIVE GRAIN GROWTH IS LIKELY IN THE WELD AND HAZ. SPRAY TRANSFER IS NOT USUALLY RECOMMENDED WHEN WELDING FERRITIC STAINLESS STEELS, ALTHOUGH SOME STABILIZED GRADES APPEAR TO BE LESS SENSITIVE TO HEAT INPUT. HEAT INPUT CAN BE REDUCED WITH THE PULSED GMAW PROCESS, THUS OFFERING OUT-OF-POSITION WELDING.

- *GLOBULAR TRANSFER* IS CHARACTERIZED BY LOWER HEAT INPUT, BUT IS NOT NORMALLY USED BECAUSE OF EXCESSIVE SPATTER. SPATTER CAN BE MINIMIZED WITH PROPER CHOICE OF HIGH-PURITY SHIELDING GAS AND THE PROPER CHOICE OF POWER SUPPLY (FOR EXAMPLE, RAPID ARC POWER SUPPLIES THAT FEATURE CAPACITIVE DISCHARGE ELECTRONICS).
- *SHORT-CIRCUITING TRANSFER* FEATURES THE LOWEST HEAT INPUT OF ALL THE GMAW PROCESSES. HOWEVER, LACK OF FUSION DEFECTS ARE COMMON WITH SHORT-CIRCUITING TRANSFER IF PROPER TECHNIQUES ARE NOT FOLLOWED.

While grain growth can be controlled to some degree with higher weld travel speeds, lower interpass temperatures, and heat sinks, care must be taken to avoid weld defects such as lack of fusion and insufficient penetration. In some situations, the use of an austenitic filler metal can overcome the problem of weld grain growth because of the formation of a two-phase grain structure. However, the HAZ grain growth must not be forgotten when using one of these filler metals.

When welding ferritic and austenitic stainless steels with the GMAW process using one of the standard power supplies, pure argon shielding gas is usually not used because of arc instability which causes air to be aspirated into the arc. As a result, the weld pool becomes enriched in nitrogen and oxygen. In the DCEP mode, the arc wanders, seeking spots which are emissive in character, usually oxides. This erratic arc behavior causes a disruption of the laminar flow of the inert gas, resulting in air being aspirated into the gas envelope.

One way to solve the arc wander problem has been to add an oxidizing element to the inert gas. Argon with 2% oxygen (O_2) or with 2% carbon dioxide (CO_2) typically used to weld plain carbon steel, low-alloy steel, and austenitic stainless steel, has been recommended for ferritic stainless steels as well. While this seems to be satisfactory for the stabilized ferritics in Group II where the bulk of the applications are for automobile exhaust systems (type 409), these active, oxidizing shielding gases should *never* be used for the Group III higher-purity ferritics, stabilized or not, because of the dramatic effect on degrading toughness and ductility.

Still other investigators have shown that the arc wander problems associated with the GMAW process can be resolved by using an argon-helium mix, typically 75Ar-25He, and by ensuring that the inert gas is purified and free of contaminants such as air, moisture, and hydrocarbons. To further enhance operability and weld quality, a device covered in a U.S. patent (No. 5,081,334, Improved Gas Shielding," 14 Jan 1992) enables air and moisture to be excluded from entering the welding arc. By taking the high-purity inert gas system approach rather than using a gas mixture containing reactive gases (O_2 and CO_2), critical reactive alloying elements such as titanium, niobium, aluminum, and chromium are not removed by oxidation. In this way, arc wander is virtually eliminated and wetting action is enhanced. With this new technology stable GMAW has been attained with pure argon. This patent also has possible application in any welding process that uses a wire feed system.

Flux-cored arc welding (FCAW) functions essentially the same as GMAW, except that instead of the filler wire being solid, it is fabricated from a carbon steel, stainless steel, or nickel sheath that surrounds a core of alloying additions formulated to the desired weld deposit chemistry. Some FCAW consumables contain active fluxing ingredients while others do not. In some instances, suppliers recommend the use of a supplementary gas shield, and sometimes the consumable is self-shielding. It is important to know the operating characteristics of the wire and when to use supplementary shielding gas which usually gives better results. The advantages of using shielding gas purification as discussed for GMAW may apply here as well.

The FCAW process is not recommended for the Group III ferritic stainless steels because of the difficulty of maintaining alloy purity and because of the pickup of carbon, nitrogen, and oxygen. Flux-cored arc weld deposits are never as clean as those produced by solid wire, inert gas welding processes. For less critical applications, or where an austenitic filler metal can usually be used, FCAW may be possible.

Because FCAW operates in the spray arc mode, higher heat inputs will result, making grain growth a potential problem. Although higher travel speeds can be employed with lower interpass temperature and heat sinks, care must be used to avoid lack-of-fusion defects.

Shielded Metal Arc Welding (SMAW). When welding with the SMAW process, DCEP (reverse polarity) is typically used. Shielded metal arc weld deposits made with covered electrodes are never as tough as welds made with an inert gas shielded process. For this reason, covered electrodes with matching composition are not available for the Group III ferritic stainless steels. Basically, the problem is one of purity, because carbon, nitrogen, and oxygen are difficult to suppress without seriously hampering welding electrode operability.

Where covered electrodes are available, proper storage is essential to avoid any pickup of moisture. Improper storage of SMAW consumables resulting in moisture pickup usually leads to weld porosity, and in some cases hydrogen cracking.

For some ferritic stainless steels that must be used in the as-welded condition, and where postweld annealing is not practical or cost effective, a dissimilar welding electrode can be used. The choice varies from the high-chromium austenitic grades such as a type E310 or E310 ELC (which are fully austenitic) to a high-nickel alloy such as an ENiCrFe-2. Type E309 coated electrode (welds contain ferrite for resistance to hot cracking) and its low-carbon and higher-silicon variations have been used to advantage.

High-alloy austenitic weld metals do not flow as readily as the ferritic stainless steels. High-nickel alloy weld metal is particularly sluggish with shallow penetration. This behavior requires that the welder make adjustments in technique. This means the welder must use arc force and electrode manipulation to physically move the molten weld pool against the sides of the joint. Ferritic stainless steels have significantly higher melting temperatures than do the austenitic stainless steels or corrosion-resistant nickel-base alloys. When using austenitic filler metals to join ferritic stainless steel base metals, care must be exercised to ensure that sufficient heat has been applied to fuse the ferritic stainless steel base metal. Failure to follow these recommendations usually results in lack-of-fusion defects.

Depending on weld bead profile, slag entrapment can be a problem with some covered electrodes. If the slag does not chip out readily, grinding will be necessary. Trying to float out residual slag with successive molten weld passes is rarely successful. The net result is slag entrapment which readily shows up on radiographic inspection. It should be remembered that grinding, if it becomes excessive, can result in more heat input via additional weld passes and ultimately will cause grain growth. The net result could be a loss in HAZ toughness and ductility.

When welding ferritic stainless steels with the SMAW process, the following techniques are recommended:

- AS WITH ALL OTHER WELDING PROCESSES, WELD JOINT SURFACES SHOULD BE CAREFULLY CLEANED WITH SOLVENT (BOTH SIDES OF EACH JOINT MEMBER) TO REMOVE ANY CONTAMINANT SUCH AS OIL, GREASE, OR DIRT WHICH WOULD DEGRADE WELD PROPERTIES AND CORROSION RESISTANCE.
- THE ARC SHOULD BE STRUCK IN THE BEVEL OF THE WELD JOINT, AND NOT ON THE ADJACENT BASE METAL.
- A SHORT ARC SHOULD BE MAINTAINED TO AVOID THE FORMATION OF POROSITY AND TO MINIMIZE OXIDATION EFFECTS AND PICKUP OF OXYGEN AND NITROGEN.
- AVOID EXCESSIVE INCLINATION OF THE ELECTRODE (A MAXIMUM INCLINE OF 20° FROM THE VERTICAL IS GENERALLY SATISFACTORY).
- A SLIGHT WEAVE FOR TIE-IN AND FUSION IS ALLOWED PROVIDED IT IS NOT WIDER THAN THREE TIMES THE DIAMETER OF THE ELECTRODE CORE WIRE.
- CRATER CRACKS AT WELD STOPS MUST BE GROUND OUT TO SOUND METAL BEFORE PROCEEDING. DYE PENETRANT INSPECTION IS THE BEST WAY TO BE SURE ALL CRACKED MATERIAL HAS BEEN REMOVED.

Plasma arc welding (PAW) of ferritic stainless steels is very similar to GTAW and the same precautions and recommendations apply. Concern has been voiced about the use of hydrogen gas with PAW because of the possibility of weld metal cracking caused by the interaction of hydrogen in the weld metal and in high stress conditions. Gas-tungsten arc welding experience with type 409 indicates that hydrogen additions will cause embrittlement immediately after welding but most of the hydrogen diffuses out of the weld metal within 48 h and ductility is restored. The situation, however, is more critical with higher-chromium alloys.

An advantage of the PAW process is that the unique high-intensity arc produces a weld bead characterized by deep penetration and a HAZ narrower than that produced by the GTAW process. The net result is that less HAZ grain growth is possible with PAW than with GTAW.

Resistance Welding. Spot welding, seam welding, and flash welding processes can be used provided the loss in toughness can be tolerated. Seam welding has been used to fabricate panel coil-type heat exchanger surfaces. Whenever possible argon inert gas should be used to protect the weld joint or seam (back side included) to minimize oxidation and the pickup of nitrogen. Titanium-stabilized alloys appear to be more resistant to the effects of nitrogen pickup due to poor shielding than are the high-purity alloys. This is apparently due to the formation of a protective TiN film which covers the weld pool (Ref 26).

Other Welding Processes. Ferritic stainless steels can be readily welded using the electron beam process with hard vacuum. Friction welding and laser welding are also possible. Where practical, the molten zone should be protected with an inert gas shield.

Submerged arc welding (SAW) can be used to join some of the ferritic stainless grades; however, the high heat input with the process becomes troublesome because of undesirable grain growth in the weld HAZ. Choice of fluxes and control of residues become critical. Dilution can exceed 50%. Submerged arc welding is not suitable for the Group III alloys and generally would also not be suitable for most of the other alloys.

Selection of Welding Consumables

Welds in ferritic stainless steel base metals can be produced several ways: (1) autogenously (that is, without the addition of filler metal) (Ref 27), (2) with a matching filler metal composition, (3) with an austenitic stainless steel filler metal (Ref 25), or (4) using a high-nickel filler alloy (Ref 17). Table 5 lists some base metal-filler metal combinations for several ferritic stainless steels.

TABLE 5 TYPICAL FERRITIC BASE METAL-FILLER METAL COMBINATIONS

BASE METAL	COVERED ELECTRODE, BARE ROD, OR FILLER WIRE	APPLICATION
405	405 NB	AS-WELDED
	430	ANNEALED
	308(L), 309(L), 310(ELC)	AS-WELDED
409	430	ANNEALED
	308(L), 409, 409NB, 309LSI	AS-WELDED
429	208(L), 309(L), 310(ELC)	AS-WELDED
430	430	ANNEALED
	308(L), 309(L), 310(ELC)	ANNEALED OR AS-WELDED
439	430, 430TI	ANNEALED OR AS-WELDED
442	442	ANNEALED
	308(L), 309(L), 310(ELC)	AS-WELDED
446	446	ANNEALED
	308(L), 309(L), 310(ELC)	AS-WELDED
E-BRITE 26-1	ER26-1	AS-WELDED
	INCONEL 112, HASTELOY C-276, OR HASTELOY C-22	AS-WELDED
AL 29-4C	AL 29-4-2	AS-WELDED
AL 29-4	AL 29-4-2	AS-WELDED
AL-29-4-2	AL 29-4-2	AS-WELDED

Before selecting a welding consumable, consideration must be given not only to mechanical property requirements, especially weld ductility and toughness (including the weld HAZ), but also to the level of corrosion resistance required for the application. Autogenous welding may be considered for some of the high-purity Group III alloys such as E-Brite 26-1. The mill annealed product will have good mechanical properties and corrosion resistance, but if the interstitial levels of carbon, oxygen, and nitrogen are on the high side of the specification limits, the autogenous weld will not only exhibit poor ductility and toughness, but also poor corrosion resistance. To avoid this kind of a problem, material should be purchased with interstitial element levels on the low side of the specification range.

Filler metals of matching composition for the more common grades of Groups I and II ferritic stainless steels, such as types 409 and 430, are available as solid wire, metal cored, or fluxcored wire (both gas shielded and self-shielded), covered electrodes, and bare rod. Because of the reactivity of titanium and aluminum and poor transfer of these elements across the arc, the availability of consumables containing these elements is usually limited to bare wire or rod.

When nonstandard grade ferritic filler metals are not available, or because poor wire drawing characteristics make production prohibitive, metal-cored or flux-cored filler metals offer a solution. Weld metal properties should be qualified before purchase.

Welding consumables for the Group III alloys are available as matching composition in bare wire or rod form. Because Group III weldments are usually always used in the as-welded condition, it is strongly recommended that these filler metals be purchased having interstitials at the very low end of the alloy specification. For best ductility, toughness and corrosion performance, the (C + N + O) content should not exceed 120 ppm for the ultrahigh-purity grades (such as E-Brite 26-1, 29-4, and 29-4-2).

Because the Group III ferritics of intermediate purity rely on stabilization with titanium or titanium and niobium, welds have such poor toughness that welding with these filler metals is limited to very thin section sizes such as thin-walled pipe, tubing, or sheet. When better toughness is required of stabilized grades with a ferritic consumable, welding should be performed with an ultrahigh-purity filler metal containing very low levels of (C + N + O). It should be noted that ultrahigh-purity welding consumables require careful handling and processing to keep interstitials at these very low levels. Keeping them clean at the job site is also mandatory.

Austenitic stainless steel welding consumables are often used to join ferritic stainless steels to overcome poor weld ductility and toughness deficiencies and they do not display a DBTT like ferritics. Although weld metal grain growth is usually not a problem (because a two-phase mixture of austenite and ferrite forms), grain growth in the weld HAZ of the ferritic base metal must not be overlooked. Sometimes the application of an austenitic filler can be a problem because of the rather large discrepancy in thermal expansion between the two alloys. In this regard, problems can range from (1) cracking due to thermal stress, (2) sensitization and loss of corrosion resistance if the austenitic filler metal is not a low-carbon grade or stabilized with titanium or niobium (corrosion can also result because of the heat from multipass welds or some subsequent thermal treatment after welding), or (3) stress relief which may not be fully effective because of the thermal expansion differences.

High-nickel alloy welding consumables have also been used successfully to join ferritics, especially the Group III alloys to themselves as well as to carbon steel, 300-series stainless steels, duplex stainless steels, superaustenitic stainless steels, and nickel-base alloys. Examples of nickel-base filler metals that have been used successfully are ENiCrFe-2, ENiCrMo-3, ERNiCrMo-1, and ERNiCrMo-3. In addition to their outstanding corrosion resistance, these high-nickel filler metals have coefficients of thermal expansion that are similar to the ferritics they are joining. While they have worked well with E-Brite alloy, attempts to use them with the 3 to 4% Mo grades of ultrahigh-purity ferritics have occasionally been less successful (Ref 28). Formation of brittle intermetallic compounds at the fusion boundary has been encountered in some cases, but good welds can be produced (Fig. 7). Sound welds have been produced using duplex stainless steel filler metal, but corrosion resistance of such joints may not be adequate in some applications.

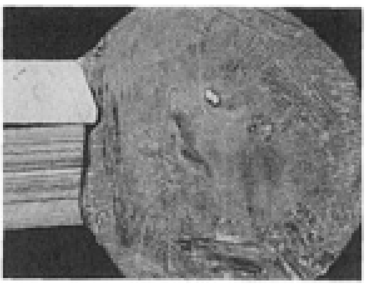

FIG. 7 WELD JOINING UNS S43735 SUPERFERRITIC STAINLESS STEEL (TOP) TO N08367 SUPERAUSTENITIC ALLOY (BOTTOM) USING N06625 NICKEL ALLOY FILLER METAL. NOTE DIFFERENCE BETWEEN MELT-BACK OF SUPERFERRITIC AND SUPERAUSTENITIC MATERIALS AND ISLAND OF UNMELTED SUPERFERRITIC STAINLESS STEEL NEAR CENTER OF WELD DEPOSIT. BOTH CONDITIONS INDICATE THE HIGH MELTING TEMPERATURE OF SUPERFERRITIC STAINLESS STEELS RELATIVE TO AUSTENITIC STAINLESS STEELS OR NICKEL ALLOYS. ETCHED WITH MIXED NITRIC/HYDROCHLORIC/ACETIC ACID SOLUTION. 14×

Welding Procedures and Weld Properties

Effect of Shielding Gas Purity on Weld Toughness (Ref 29, 30, 31, 32, 33, 34). The primary reasons for using shielding and backing gas are to protect the molten filler wire tip, the nonconsumable electrode (when used), the weld pool, and the solidified weld bead from atmospheric contamination. Researchers have shown that gas contaminants can cause cracking, varying degrees of weld bead oxidation, arc instability, and degradation of mechanical and corrosion properties. A particular gas or gas mixture might be used for enhanced arc stability, a particular mode of weld metal transfer, enhanced penetration or bead profile, its availability, or easier arc ignition.

Certainly, there is no question that quality welds require quality shielding. What is essential is that clean gas be maintained and delivered to the point of use. From a practical standpoint, some factors that have been identified as relating to "bad gas" and welding problems are:

- CONTAMINATED GAS CYLINDERS (BY MOISTURE/AIR)
- CONTAMINATED AND/OR LEAKING GAS MANIFOLD SYSTEMS
- DAMAGED, DEFECTIVE, OR LOOSE SHIELDING/BACKING GAS LINE FITTINGS
- INTRUSION OF CONTAMINANTS WHEN MIXING GASES

Achieving gas quality at the point of use can be elusive. Welding problems can show up with a single cylinder of gas or a series of cylinders. A manifold system can be particularly difficult to deal with because of its long length and numerous welding stations. As a result of poor gas quality, defects are produced which require repair, job completion can be delayed, and job quality may be compromised. What is needed is a portable, cost-effective system that will remove contaminants from the shielding gas and backing gas, not only when working in the shop, but also in the field.

Because ferritic stainless steels are sensitive to interstitial element contaminants, it is important that these elements be removed so as not to degrade properties such as ductility, toughness, and corrosion resistance. Impurities commonly found in gas cylinder or manifold systems are moisture, oxygen, hydrocarbons (such as oils from compressors), and carbon dioxide (when not intentionally added). Moisture in CO_2 gas is also a common problem.

While there are systems on the market that can eliminate gas contaminants to levels below one part per million (ppm), they have to rely on multicomponent, energy-intensive gettering. These systems are heavy, flow rates are usually restrictive, and they are not very portable. Another system relies on a resin that is a single rechargeable compound that can provide effective removal of a variety of impurities to less than 10 parts per billion (ppb), operates at room temperature, and is light and portable. The effects of using such a system on the impact properties of 6 mm ($\frac{1}{4}$ in.) thick E-Brite 26-1 alloy plate are shown in Fig. 8.

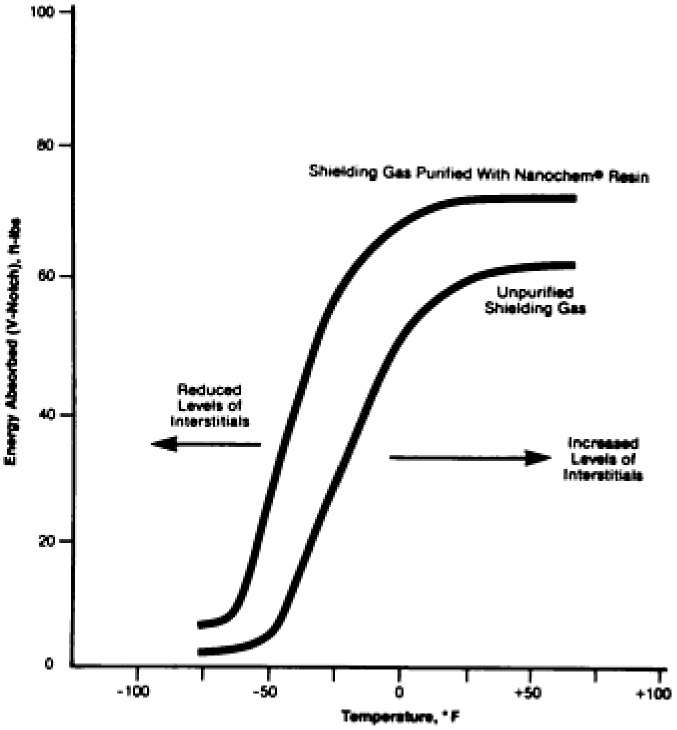

FIG. 8 CHARPY V-NOTCH IMPACT TEST RESULTS FROM GAS-TUNGSTEN ARC WELDED E-BRITE 26-1 PLATE. WELD CONDITIONS: PLATE THICKNESS, 6 MM ($\frac{1}{4}$ IN.); SHIELDING/BACKING GAS, ARGON; SHIELDING/BACKING GAS IMPURITY LEVELS, H_2O 40 PPM AND O_2 20 PPM; SHIELDING GAS FLOW RATE, 28 L/MIN (60 FT^3/H); VOLTAGE, 14V; CURRENT (DCEN), 150 A; TRAVEL SPEED, 1.5 TO 2.5 MM/S (3.5 TO 6 IN./MIN). NANOCHEM IS A REGISTERED TRADEMARK OF HERCULES INC. SOURCE: REF 35

Hydrogen Embrittlement. Like other ferritic materials, the ferritic stainless steels are susceptible to hydrogen embrittlement. For this reason, the use of hydrogen-containing shielding gas should be avoided. Hydrogen can also be created from water, water vapor, or oils, and these should be rigorously excluded from the weld region. The presence of hydrogen in the weld can lead to cracking of the weld bead shortly after welding. It can also reduce ductility in uncracked welds. Hydrogen is often spontaneously outgassed from these alloys at room temperature and is readily lost during low-temperature (90 to 200 °C, or 200 to 400 °F) heat treatments (Ref 36). Hydrogen outgassing at room temperature, if it will occur, requires a few days to a few weeks. Surface films of oxides and/or nitrides can inhibit hydrogen outgassing, and it may be necessary to remove such films to facilitate hydrogen outgassing and restoration of ductility. Sensitivity to hydrogen is a function of alloy composition, microstructure, and strength. In general, the stabilized, high-chromium plus molybdenum materials, such as AL 29-4C alloy, are more sensitive than the high-purity, lower-chromium plus molybdenum materials, such as the E-Brite alloy.

Preheating. One of the main reasons for preheating is to reduce the temperature gradient between the weld area and the unaffected base metal. Any austenite which forms in nominally ferritic stainless steels will air harden to martensite even in heavy (>300 mm, or 12 in., diam) sections. Preheating light-gage sheet will not prevent the formation of martensite. It is also unlikely that enough preheat could be applied to produce any significant post-transformation tempering. Hydrogen may be a factor in situations involving high stress and restraint. When additional preheat time is allowed, the hydrogen can diffuse out safely. Ferritic stainless steels that would benefit from preheat would be the Group I types (430, 434, 442, and 446).

When required, a preheat in the range of 150 to 230 °C (300 to 450 °F) is usually sufficient. Section sizes of ≤ 6 mm ($\leq \frac{1}{4}$ in.) generally do not need preheat, but such parameters as joint design, restraint, welding process, cooling rate, and dissimilar metal combinations (different coefficients of thermal expansion), and base metal composition should be reviewed before making a final decision.

Preheat should not be used on the Group III alloys, especially the ultra-high purity steels. These alloys do not form martensite when welded, so no preheat is necessary. Furthermore, because these alloys have such low levels of impurities, excessive grain growth would occur, compared with the Group I and Group II alloys.

Postweld annealing is normally performed on the Group I ferritic stainless steels. The purpose is to transform any martensite that may have formed during welding into a wholly ferritic structure (usually containing some spheroidized carbides), or to remove effects of high-temperature embrittlement and improve corrosion resistance. Annealing under these conditions does not refine coarsened ferrite grains (some form of metal working is needed to achieve recrystallization and grain refinement).

When postweld annealing is required, the recommended temperature range is 790 to 850 °C (1450 to 1550 °F). At these temperatures consideration must be given to oxidation of the metal surface and distortion effects. Fixturing to prevent metal deformation may be required. Sigma, σ, and chi, χ, formation is typically not a problem with the Group I alloys.

Cooling from the annealing temperature must be done carefully. Furnace cooling to 600 °C (1100 °F) is usually specified to minimize distortion from handling. Rapid cooling through the temperature range of 565 to 400 °C (1050 to 750 °F) is necessary to avoid 475 °C (885 °F) embrittlement. Depending on section thickness, forced air cooling or water spray quenching is usually performed.

Group III ferritic stainless steels (ultralow interstitials and the stabilized grades), particularly the high-chromium, high-molybdenum alloys, are typically not annealed after welding because of the danger of precipitation of σ and χ phases in the temperature range of 550 to 900 °C (1020 to 1650 °F) unless the anneal is followed by a rapid quench. However, postweld annealing of welded tubes is common.

Corrosion Resistance. Investigators have determined that the loss of corrosion resistance of ferritic stainless steels after welding is caused by chromium depletion upon cooling from exposure to temperatures of 925 °C (1700 °F) (Ref 37, 38, 39, 40, 41, 42). Sensitization[*] due to chromium depletion occurs by holding or cooling through the temperature range of 425 to 700 °C (800 to 1300 °F). In the temperature range from 700 to 925 °C (1300 to 1700 °F), rapid diffusion of chromium in ferrite occurs, replenishing the chromium-depleted zones. For this reason, weldments which are annealed at 790 to 850 °C (1450 to 1550 °F) show good intergranular corrosion resistance despite the presence of intergranular precipitates of chromium-rich carbides and nitrides. Below 700 °C (1300 °F), little chromium diffusion occurs and corrosion resistance is not completely restored.

Conventional (Group I) 400-series ferritic stainless steels such as types 430, 434, and 446 are susceptible to intergranular corrosion and to embrittlement in the as-welded condition. Corrosion in the weld area generally encompasses both the weld metal and weld HAZ. Early attempts to avoid some of these problems involved the use of austenitic stainless steel filler metals; however, failure by corrosion of the HAZ usually occurred even when exposure was to rather mild media for relatively short periods of time.

Figure 9 shows an example of a saturator tank used to manufacture carbonated water at room temperature that failed by leakage through the weld HAZ of the type 430 base metal after being in service for only 2 months. This vessel, fabricated by welding with a type 308 stainless steel welding electrode, was placed in service in the as-welded condition. Figure 10 shows a photomicrograph of the weld/base metal interface at the outside surface of the vessel; corrosion initiated at the inside surface. Postweld annealing--at 785 °C (1450 °F) for 4 h in the case of type 430 stainless steel--restores weld area ductility and resistance to corrosion to that of the unwelded base metal.

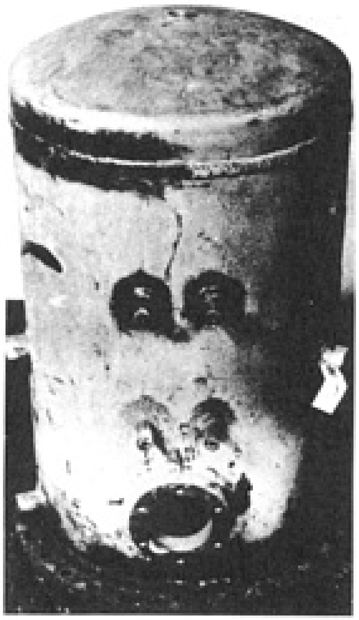

FIG. 9 AS-WELDED TYPE 430 STAINLESS STEEL SATURATOR TANK USED IN THE MANUFACTURE OF CARBONATED WATER THAT FAILED AFTER 2 MONTHS OF SERVICE. THE TANK WAS SHIELDED METAL ARC WELDED USING TYPE E308 FILLER METAL. SOURCE: REF 34

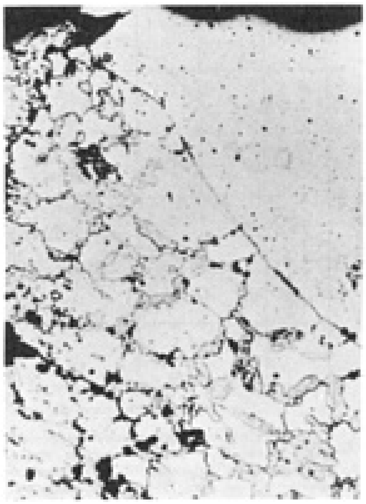

FIG. 10 PHOTOMICROGRAPH OF THE OUTSIDE SURFACE OF THE SATURATOR TANK SHOWN IN FIG. 9 SHOWING CORROSION AT THE FUSION LINE. SOURCE: REF 34

To overcome some of these earlier difficulties and to improve weldability, several of the standard grade ferritic stainless steels were modified. For example, type 405, containing nominally 12% Cr, is made with lower carbon and a small aluminum addition of 0.20% to restrict the formation of austenite at high temperature so that hardening is reduced during welding. For maximum ductility and corrosion resistance, however, postweld annealing is necessary. Recommendations for welding include either a 430- or a 309-type filler metal, the latter being used when increased weld ductility is desired.

As mentioned earlier in this article, researchers recognized that the high-chromium, high-molybdenum ferritic stainless steels possessed a desirable combination of good mechanical properties and resistance to corrosion. These properties made them attractive alternatives to the austenitic stainless steels, which often failed due to chloride SCC.

Reducing the total interstitial element levels in ferritic stainless steels improves weldability and corrosion resistance. Figure 11 shows the relationship of chromium content and (C + N) content on the combined properties of as-welded corrosion resistance and ductility. Table 6 lists this relationship separately for each property. As chromium content increases, the amount of (C + N) that can be tolerated for intergranular corrosion resistance increases. Conversely, for aswelded ductility, the amount of tolerable (C + N) is drastically reduced. Thus, at low chromium levels, as-welded corrosion resistance is the controlling factor; at high chromium levels, as-welded ductility is the factor that limits the use of high-chromium stainless steels.

TABLE 6 LIMITS OF INTERSTITIAL ELEMENT (C + N) CONTENT FOR ACCEPTABLE AS-WELDED INTERGRANULAR CORROSION RESISTANCE AND AS-WELDED DUCTILITY

Sample thickness: 2.5 mm (0.1 in.)

CHROMIUM CONTENT, WT%	**INTERGRANULAR CORROSION RESISTANCE(A), PPM (C + N)**	**DUCTILITY(B), PPM (C + N)**
19	60-80	>700
26	100-130	200-500
30	130-200	80-100
35	<250	<20

(A) INTERGRANULAR CORROSION RESISTANCE IN BOILING 50% FERRIC SULFATE-50% SULFURIC ACID SOLUTION.

(B) NO CRACKING AS DETERMINED BY SLOW BENDING AROUND A 5 MM (0.2 IN.) MANDREL

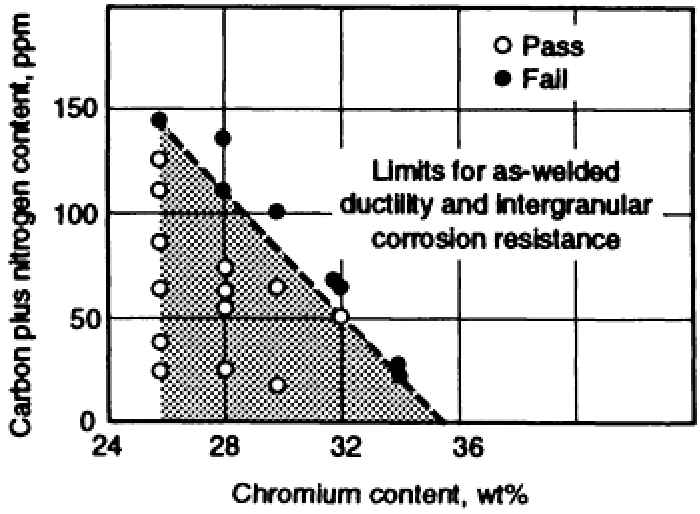

FIG. 11 THE EFFECT OF INTERSTITIAL LEVELS AND CHROMIUM CONTENT ON AS-WELDED DUCTILITY AND INTERGRANULAR CORROSION RESISTANCE. SOURCE: REF 43

To improve the corrosion resistance of ferritic stainless steels, two practical methods have been applied:

- *LOWER INTERSTITIAL LEVELS.* BECAUSE THE SOLUBILITY OF INTERSTITIAL ELEMENTS IS LOW IN FERRITE, NEW MELTING AND REFINING PRACTICES HAVE BEEN UTILIZED TO PRODUCE ULTRAHIGH-PURITY FERRITIC (GROUP III) STAINLESS ALLOYS LIKE THOSE SHOWN IN TABLE 3. WITH THESE LOW LEVELS OF CARBON AND NITROGEN, CHROMIUM CARBIDE AND NITRIDE PRECIPITATION CANNOT OCCUR TO A HARMFUL DEGREE, THUS AS-WELDED DUCTILITY AND CORROSION RESISTANCE ARE MAINTAINED

- *STABILIZATION OF CARBON AND NITROGEN.* THIS IS ACHIEVED BY THE ADDITION OF CARBIDE- AND NITRIDE-FORMING ELEMENTS WHICH ARE MORE STABLE THAN CHROMIUM CARBIDES AND CHROMIUM NITRIDES. TITANIUM AND/OR NIOBIUM ARE USED TO TIE UP THE CARBON AND NITROGEN IN THE MATRIX. STABILIZATION HAS PROVEN EFFECTIVE IN PREVENTING SENSITIZATION OF FERRITIC ALLOYS SUCH AS 18CR-2MO AND 26CR-1MO, AS WELL AS THE HIGHER CHROMIUM-MOLYBDENUM GRADES (REF 44). INVESTIGATORS HAVE SHOWN, HOWEVER, THAT IF EXCESSIVE AMOUNTS OF STABILIZER ARE USED BEYOND WHAT IS REQUIRED FOR CORROSION RESISTANCE, DUCTILITY AND ROOM-TEMPERATURE IMPACT PROPERTIES ARE DEGRADED (REF 15, 44, 45, 46).

It should also be pointed out that corrosion resistance can also be improved by following proper welding procedures and recommendations. As described previously, preheating (except to remove moisture) or postweld heat treating should not be performed. Postweld heat treatment is used only with the conventional ferritic stainless alloys. The following example illustrates the results of not following proper procedures.

Example 1. Leaking Welds in a Ferritic Stainless Steel Wastewater Vaporizer. A nozzle in a wastewater vaporizer began leaking after approximately 3 years of service with acetic and formic acid wastewaters at 105 °C (225 °F) and 414 kPa (60 psig).

Investigation. The shell of the vessel was gas-tungsten arc welded in 1972 from 6.5 mm ($\frac{1}{4}$ in.) E-Brite stainless steel plate. The shell measured 1.5 m (58 in.) in diameter and 8.5 m (28 ft) in length. Nondestructive examination included 100% radiography, dye-penetrant inspection, and hydrostatic testing of all E-Brite welds.

An internal inspection of the vessel revealed that portions of the circumferential and longitudinal seam welds, in addition to the leaking nozzle weld, displayed intergranular corrosion. At the point of leakage, there was a small intergranular crack. Figure 12 shows a typical example of the corroded weld. A transverse cross section through this weld characteristically displays intergranular corrosion with grains dropping out (Fig. 13). The HAZ next to the weld fusion line also experiences intergranular corrosion a couple of grains deep as a result of sensitization (Fig. 14).

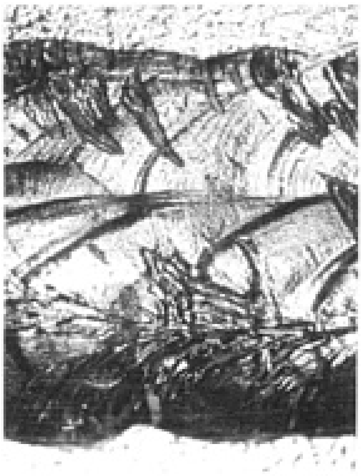

FIG. 12 TOP VIEW OF A LONGITUDINAL WELD IN 6 MM ($\frac{1}{4}$ IN.) THICK E-BRITE STAINLESS STEEL PLATE SHOWING INTERGRANULAR CORROSION. THE WELD WAS MADE WITH MATCHING FILLER METAL. ABOUT 4×. SOURCE: REF 47

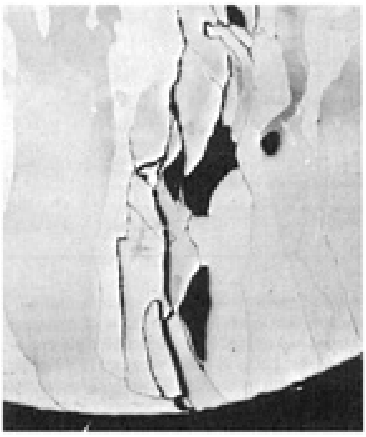

FIG. 13 INTERGRANULAR CORROSION OF A CONTAMINATED E-BRITE STAINLESS STEEL WELD. ELECTROLYTICALLY ETCHED WITH 10% OXALIC ACID. 200×. SOURCE: REF 47

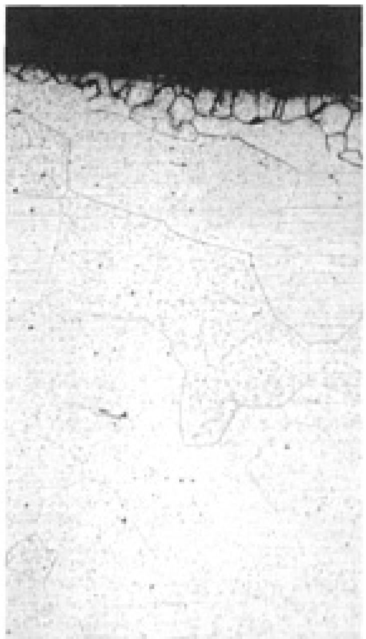

FIG. 14 INTERGRANULAR CORROSION OF THE INSIDE SURFACE HAZ OF E-BRITE STAINLESS STEEL ADJACENT TO THE WELD FUSION LINE. ELECTROLYTICALLY ETCHED WITH 10% OXALIC ACID. 100×. SOURCE: REF 47

The evidence indicated weldment contamination; therefore, effort was directed at finding levels of carbon, nitrogen, and oxygen in the various components present before and after welding. The following are the average results:

ELEMENT	**CONTENT, PPM**
E-BRITE BASE PLATE	
C	6
N	108 (C + N = 114)
O	57
CORRODED LONGITUDINAL WELD	

C	133
N	328 (C + N = 461)
O	262
CORRODED CIRCUMFERENTIAL WELD	
C	34
N	169 (C + N = 203)
O	225
E-BRITE WELD WIRE	
C	3
N	53 (C + N = 56)
O	55
SOUND LONGITUDINAL WELD	
C	10
N	124 (C + N = 134)
O	188
SOUND CIRCUMFERENTIAL WELD	
C	20
N	106 (C + N = 126)
O	85

These results confirmed suspicions that failure was due to excessive amounts of nitrogen, carbon, and oxygen. To characterize the condition of the vessel further, subsize CVN impact tests were run on the unaffected base metal, the HAZ, and the uncorroded (sound) weld metal. These tests showed the following ductile-to-brittle transition temperatures:

SPECIMEN	**DUCTILE-TO-BRITTLE TRANSITION TEMPERATURE**	
	°C	**°F**
BASE METAL	40 ± 3	105 ± 5
HAZ	85 ± 3	180 ± 5
WELD	5 ± 3	40 ± 5

Comparison of the interstitial levels of the corroded welds, sound welds, base metal, and filler wire suggested that insufficient joint cleaning and faulty gas shielding were probably the main contributing factors that caused this weld corrosion failure. Discussions with the vendor uncovered two discrepancies. First, the welder was using a large, 19 mm ($\frac{3}{4}$ in.) inside diameter ceramic nozzle with a gas lens, but was flowing only 19 L/min (40 ft^3/h) of argon; this was the flow rate previously used with a 13 mm ($\frac{1}{2}$ in.) inside diameter gas lens nozzle. Second, a manifold system, found to be contaminated, was used to distribute pure argon welding gas from a large liquid argon tank to various satellite welding stations in the welding shop. The exact cause for the carbon pickup was not determined.

Conclusions. Failure of the nozzle weld was the result of intergranular corrosion caused by the pickup of interstitial elements and subsequent precipitation (sensitization) of chromium carbides and nitrides. Carbon pickup was believed to have been caused by inadequate joint cleaning prior to welding. The increase in the weld nitrogen level was a direct result of inadequate argon gas shielding of the molten weld puddle. Two areas of inadequate shielding were identified:

- IMPROPER GAS FLOW RATE FOR A 19 MM ($\frac{3}{4}$ IN.) DIAM GAS LENS NOZZLE
- CONTAMINATION OF THE MANIFOLD GAS SYSTEM (LEAKS)

In order to preserve the structural integrity and corrosion performance of the new Group III ferritic stainless steels, it is important to avoid the pickup of the interstitial elements carbon, nitrogen, oxygen, and hydrogen. In this particular case, the vendor used a flow rate of 23 to 28 L/min (50 to 60 ft^3/h) of argon for a 19 mm ($\frac{3}{4}$ in.) gas lens nozzle. As was mentioned previously in this article, the gas lens collet body is an important and necessary part of the torch used to weld these alloys. Failure to use a gas lens will result in a flow condition that is turbulent enough to aspirate air into the gas stream, thus contaminating the weld and destroying its mechanical and corrosion properties.

The manifold gas system also contributed to this failure. When a system such as this is first used, it is necessary to purge the contents of the manifold of any air to avoid oxidation and contamination. When that is done, the system functions satisfactorily; however, when it is shut down overnight or for repairs, air infiltrates back in, and a source of contamination is reestablished. Manifold systems are never fully purged, and leaks are common.

The contaminated welds were removed, and the vessels were rewelded and put back into service. Some rework involved the use of covered electrodes of dissimilar composition. No problems have been reported to date.

Recommendations. First, to ensure proper joint cleaning, solvent washing and wiping with a lint-free cloth should be performed immediately before welding. Also, a word of caution: Solvents are generally flammable and can be toxic. Ventilation should be adequate. Cleaning should continue until cloths are free of any residues.

Second, when GTAW, a 19 mm ($\frac{3}{4}$ in.) diameter ceramic nozzle with gas lens collet body is recommended. An argon gas flow rate of 28 L/min (60 ft^3/h) is optimum. Smaller nozzles are not recommended. Argon back gas shielding is mandatory at a slight positive pressure to avoid disrupting the flow of the welding torch.

Third, the tip of the filler wire should be kept within the torch shielding gas envelope to avoid contamination and pickup of nitrogen and oxygen (they embrittle the weld). If the tip becomes contaminated, welding should be stopped, the contaminated weld area ground out, and the tip of the filler wire that has been oxidized should be snipped off before proceeding with welding.

Fourth, a manifold gas system can be a problem when supplying shielding and backing gas. Individual argon gas cylinders (or individual liquid dewars) were found to provide good performance. A weld button spot test should be performed to confirm the integrity of the argon cylinder and all hose connections. In this test, the weld button sample must be absolutely bright and shiny. Any cloudiness is an indication of contamination. It is then necessary to check for leaks or to replace the cylinder. The best insurance is to purify the gas.

Fifth, it is important to remember that corrosion resistance is not the only criterion when evaluating these new ferritic stainless steels. Welds must also be tough and ductile, and these factors must be considered when making welds.

Lastly, dissimilar weld filler metals can be successfully used. To avoid premature failure, the dissimilar combination should be corrosion tested to ensure suitability for the intended service.

Weld toughness is an important property, and of great interest to the design engineer as well as the end user because of the influence of weld defects or notches on the performance of the material such as in a piping system or pressure vessel, especially under high strain rate conditions. Because ferritic stainless steels undergo a transition from ductile-to-brittle behavior as the temperature decreases (Fig. 2), researchers have been able to show rather good correlation between the DBTT and weldability. The common method used to measure DBTT is the CVN test.

Section size and thickness have a significant effect on the impact toughness of ferritic stainless steels (and all ferritic alloys). The effect of section size on DBTT is shown in Fig. 15. The thinner the material, the lower the DBTT.

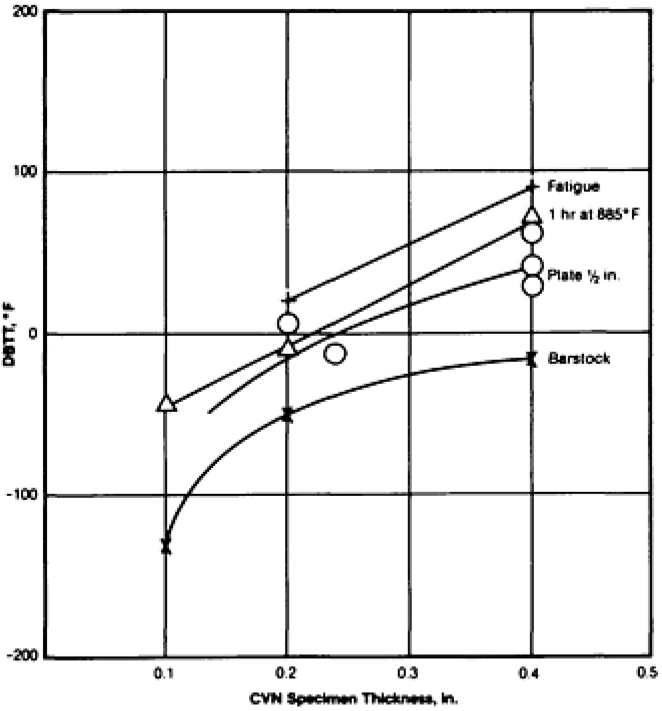

FIG. 15 IMPACT PROPERTIES AS A FUNCTION OF SECTION THICKNESS FOR XM-27 FERRITIC STAINLESS STEEL. SOURCE: REF 3

The cooling rate from peak welding temperatures is very important because toughness appears to be strongly related to the precipitation and morphology of carbides and nitrides. Even in ultrahigh-purity E-Brite 26-1, chromium nitrides have been observed (Ref 3). Depending on cooling rate, these precipitates will align along the grain boundaries or intragranularly within the matrix. For this reason, grain size is important. The finer the grain size, the more uniform the properties and the tougher the material.

Figure 16 illustrates the effect of carbon, nitrogen, and stabilizing elements on 26Cr-1Mo alloys. As shown in this figure, low interstitial levels correspond to lower DBTT values. Similar results are reported in Ref 4, 6, 44, 46, 48, and 49. It was further observed that notch brittleness was the result of the total interstitial element content (C + N), rather than the absolute value of each. Reducing the total interstitial element levels in ferritic stainless steels improves weldability and notch toughness. Another comparison is shown in Fig. 17. Here, the higher the purity of a 26Cr-1Mo alloy, the better the notch toughness and the lower the DBTT when compared to a 26Cr-1Mo alloy stabilized with titanium. Although the titanium-stabilized alloy was 3 mm ($\frac{1}{8}$ in.) thick compared to the 6 mm ($\frac{1}{4}$ in.) thick E-Brite plate, the ultrahigh-purity alloy had decidedly superior notch toughness (DBTT = -40 °C, or -40 °F, for E-Brite versus DBTT = +80 °C, or +175 °F, for 26-1 with titanium).

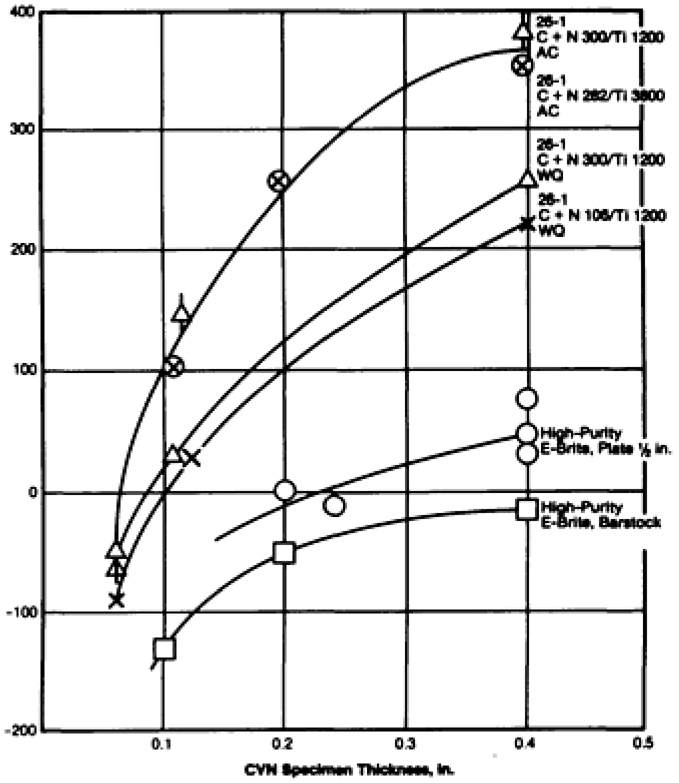

FIG. 16 EFFECT OF ALLOY PURITY (INTERSTITIAL ELEMENT CONTENT) AND STABILIZER CONTENT (VALUES GIVEN IN PARTS PER MILLION) ON THE DBTT OF FERRITIC STAINLESS STEELS. AC, AIR COOL; WQ, WATER QUENCHED. SOURCE: REF 3

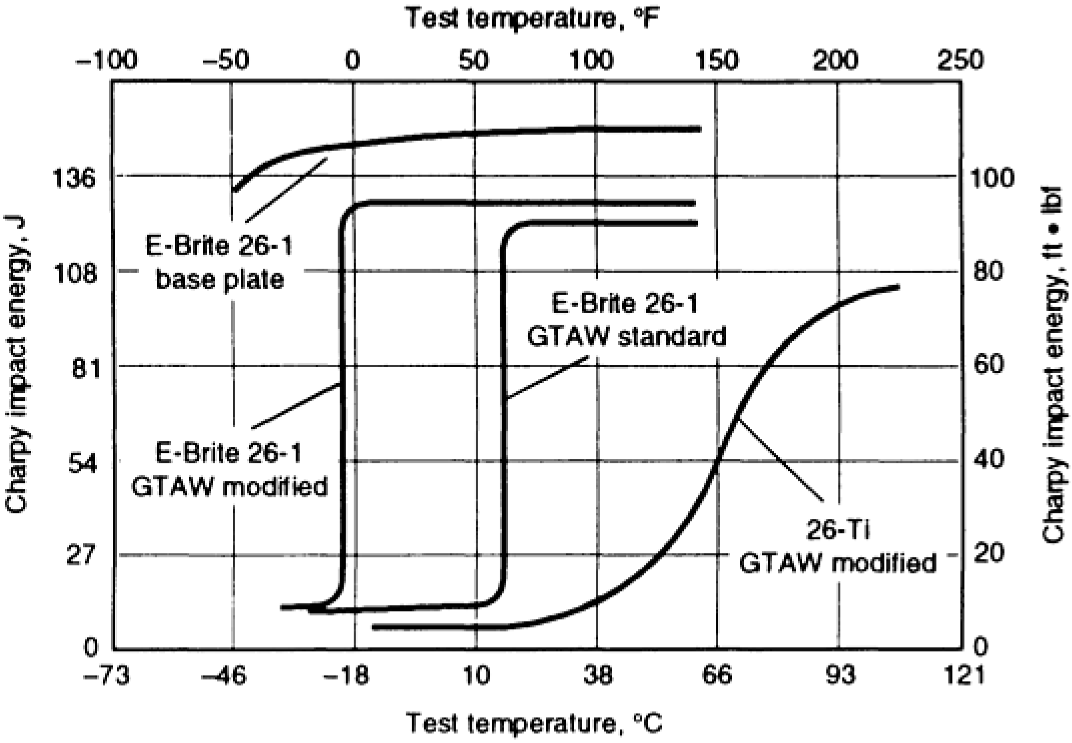

FIG. 17 NOTCH TOUGHNESS OF A GAS-TUNGSTEN ARC WELDED HIGH-PURITY FERRITIC STAINLESS STEEL (6 MM, OR $\frac{1}{4}$ IN., THICK E-BRITE 26-1 PLATE) VERSUS A TITANIUM-STABILIZED ALLOY (3 MM, OR $\frac{1}{8}$ IN., THICK

26-1 TI PLATE). SOURCE: REF 17

If the alloy thickness is reduced below 3 mm ($\frac{1}{8}$ in.), the stabilized grades start to display somewhat improved toughness.

Because of the difficulty of measuring CVN impact toughness in those thin gages, weld ductility as measured by a bend test (see Table 6) or bend angle, was determined to be an acceptable method of measuring weldability. Under these conditions, stabilized ferritic stainless steels have excellent aswelded ductility when stabilizers are present in the minimum amounts necessary for stabilization, but the room-temperature impact properties suffer. When excessive stabilizing element(s) are present, weld ductility decreases as excess stabilizer content increases (Ref 50). If welding techniques are poor and shielding gas is contaminated, the amounts of stabilizer added to prevent sensitization and embrittlement may be insufficient to combine with all of the carbon and nitrogen. Oxygen pickup can also be detrimental. The result is loss of ductility, toughness, and corrosion resistance. Shielding gas purity must not be overlooked.

Dissimilar austenitic filler metals can be used to overcome the difficulties of higher than desired weld metal DBTT. Typical CVN toughness data are shown in Fig. 18.

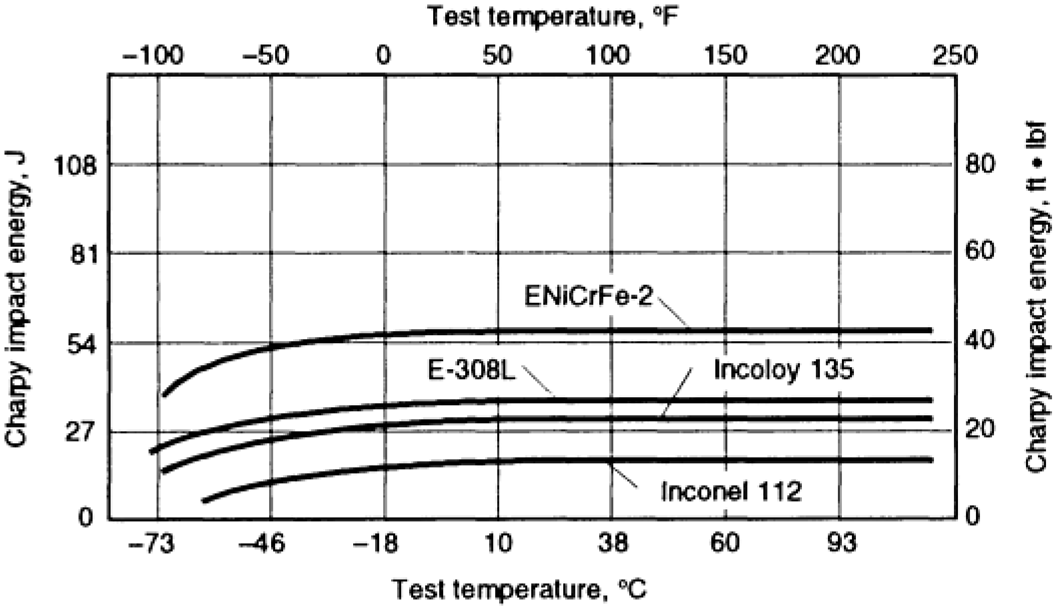

FIG. 18 CHARPY V-NOTCH TOUGHNESS OF SHIELDED METAL ARC WELDS MADE IN 6 MM ($\frac{1}{4}$ IN.) THICK E-BRITE 26-1 PLATE WITH DIFFERENT FILLER METALS. SOURCE: REF 17

Hot Cracking. Weld and HAZ hot cracking in the ferritic stainless steels is not as common a problem as with austenitic stainless steels. This is a result of the lower coefficient of thermal expansion of the ferritic stainless steels, as well as the greater solubility of sulfur and phosphorus in ferrite. However, excessive amounts of stabilizing elements have been shown to cause hot cracking as well as reduced ductility (Ref 25). Figure 19 shows hot cracks in the top surfaces of autogenous gas-tungsten arc welds in two type 444 ferritic stainless steel subscale varestraint test specimens. The stabilized alloys were found to be more susceptible to hot cracking than an alloy of similar composition with ultralow levels of impurities.

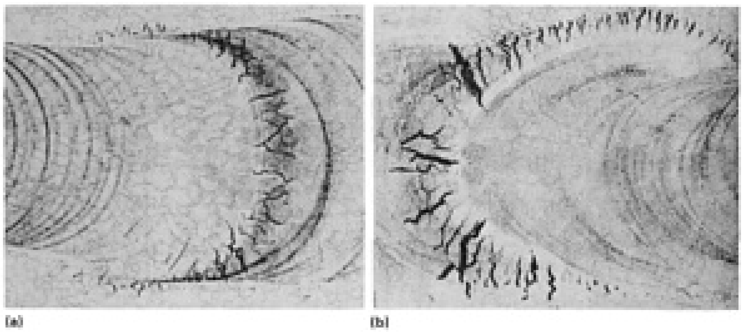

FIG. 19 TOP SURFACES OF SUBSCALE VARESTRAINT TEST GAS-TUNGSTEN WELDS IN TYPE 444 STAINLESS STEELS. (A) TITANIUM-STABILIZED (0.333% TI) TYPE 444 ALLOY. (B) NIOBIUM-STABILIZED (0.401% NB) TYPE 444 ALLOY. SOURCE: REF 25

The alloy which contained 0.40% Nb as a stabilizing element cracked more severely than the alloy which contained 0.333% Ti, but the worst cracking occurred in an alloy which contained both niobium (0.394%) and titanium (0.171%). The high-purity alloy which contained a total (C + N) content of 47 ppm (and no stabilizing elements), produced no hot cracks when tested under the same conditions. Although sulfur and phosphorus can also cause hot cracking in the ferritic stainless steels, the excess levels of niobium and titanium (above that necessary to tie up the carbon and nitrogen) are a more common cause of hot cracking. Thus, the stabilizing elements must be kept to the minimum practical levels to avoid hot cracking and reduced ductility.

Along with hot cracking, which is not often a problem for ferritic stainless steel welds, autogenous welds in alloys stabilized only with niobium often suffer from low ductility and poor toughness. In highly restrained situations, cracks may propagate in the weld along its length. For this reason, ferritic stainless steels of high or intermediate interstitial content stabilized only with niobium are rarely welded. (Titanium-free, stabilized ferritic stainless steels do, however, provide better brazeability and may be selected for brazed rather than welded applications.) Reduction of total interstitial element content to below approximately 0.02% largely eliminates this problem (Ref 51). Thus, the Group III ultra high purity alloys that contain niobium additions can be welded without fear of cracking. Addition of approximately 0.05% or greater titanium causes equiaxed grains to form near the weld centerline and yields improved weld properties (Ref 22). Thus, dual stabilized, intermediate purity Group III ferritic stainless steels are used to provide toughness in both the base metal and the weld regions.

References cited in this section

3. W.A. MATEJKA AND R.J. KNOTH, E-BRITE 26-1, THE HIGH-PURITY APPROACH TO HIGH CHROMIUM FERRITIC STAINLESS STEELS, *ASM SYMPOSIUM,* BAL HARBOUR, FLORIDA, ASTM, 6 DEC 1973
4. R.N. WRIGHT, MECHANICAL BEHAVIOR AND WELDABILITY OF HIGH-CHROMIUM FERRITIC STAINLESS STEELS AS A FUNCTION OF PURITY, *WELD. J.,* OCT 1971, P 434S-440S
6. J.J. DEMO, WELDABLE AND CORROSION-RESISTANT FERRITIC STAINLESS STEELS, *METALL. TRANS.,* VOL 5, NO. 11, NOV 1974, P 2253-2256
13. G. ROTHSCHILD, WELDING OF E-BRITE 26-1, PROCEDURE NO. W-2, WELDING SEMINAR HELD IN HOUSTON, TEXAS, 13 SEPT 1973
14. K.E. DORSCHU, WELDABILITY OF A NEW FERRITIC STAINLESS STEEL, *WELD. J.,* SEPT 1971
15. R.N. WRIGHT, "THE TOUGHNESS OF FERRITIC STAINLESS STEELS," MATERIALS ENGINEERING DEPARTMENT, RENSSELAER POLYTECHNIC INSTITUTE, TROY, NY, 1979
16. D.H. KAH AND D.W. DICKINSON, WELDABILITY OF FERRITIC STAINLESS STEEL, *WELD. J.,* AUG 1981, P 135S-142S
17. K.F. KRYSIAK, WELDABILITY OF THE NEW GENERATION OF FERRITIC STAINLESS STEELS--

PART I, PAPER NO. 92 PRESENTED AT THE NATIONAL ASSOCIATION OF CORROSION ENGINEERS CONFERENCE (TORONTO, CANADA), 14-18 APRIL 1975

18. K.F. KRYSIAK, WELDABILITY OF THE NEW GENERATION OF FERRITIC STAINLESS STEELS--UPDATE, *SYMPOSIUM ON FERRITIC STAINLESS STEELS* (SAN FRANCISCO, CA), 23-24 MAY 1979
19. R. LOWRIE, WELDING E-BRITE 26-1 TO OTHER ALLOYS, *WELD. J.,* NOV 1973, P 500S-506S
20. Y. OKAZAKI, R. TODOROKI, T. SAKAMOTO, AND T. ZAIZEN, ON THE PROPERTIES OF HIGH-PURITY 19 CR-2 MO FERRITIC STAINLESS STEEL WELDS, PAPER NO. 118, PRESENTED AT THE NATIONAL ASSOCIATION OF CORROSION ENGINEERS CONFERENCE (TORONTO, CANADA), 6-10 APRIL 1981
21. J.M. SAWHILL, JR. AND A.P. BOND, DUCTILITY AND TOUGHNESS OF STAINLESS STEEL WELDS, *WELD. J.,* VOL 55 (NO. 10), 1976, P 33S-41S
22. J.F. GRUBB AND S.D. WASHKO, THE EFFECT OF NIOBIUM AND TITANIUM DUAL STABILIZATION ON THE WELDABILITY OF 11% CHROMIUM FERRITIC STAINLESS STEELS, *PROC. INT. CONF. STAINLESS STEELS* (CHIBA, JAPAN), 10-13 JUNE 1991, ISIJ, P 1062-1068
23. J.C. VILLAFUERTE AND H.W. KERR, ELECTROMAGNETIC STIRRING AND GRAIN REFINEMENT IN STAINLESS STEEL GTA WELDS, *WELD. J.,* VOL 69 (NO. 1), 1990, P 1S-13S
24. J.C. VILLAFUERTE, E. PARDO, AND H.W. KERR, THE EFFECT OF ALLOY COMPOSITION AND WELDING COMPOSITIONS ON COLUMNAR-EQUIAXED TRANSITIONS IN FERRITIC STAINLESS STEEL GAS-TUNGSTEN ARC WELDS, *METALL. TRANS. A,* VOL 21 (NO. 7), JULY 1990, P 2009-2019
25. R.D. CAMPBELL, FERRITIC STAINLESS STEEL WELDING METALLURGY, *KEY ENGINEERING MATERIALS,* VOL 69 & 70, TRANS TECH PUBLICATIONS, SWITZERLAND, 1992, P 167-216
26. J.D. GATES AND R.A. JAGO, ABSORPTION OF GASEOUS CONTAMINANTS BY WELDS AND WELD SIMULATIONS IN FERRITIC STAINLESS STEELS, *MATER. SCI. TECHNOL.,* VOL 3, MAY 1987, P 386-393
27. J.M. SAWHILL, "DUCTILITY AND TOUGHNESS OF AUTOGENOUS GAS TUNGSTEN-ARC WELDS IN 18CR-2MO STEELS," RP-33-73-21, CLIMAX MOLYBDENUM CO. OF MICHIGAN, 24 FEB 1975
28. J.F. GRUBB, HYDROGEN EMBRITTLEMENT OF SUPERFERRITIC STAINLESS STEELS, *NEW DEVELOPMENTS IN STAINLESS STEEL TECHNOLOGY,* R.A. LULA, ED., AMERICAN SOCIETY FOR METALS, 1985, P 175-187
29. K.F. KRYSIAK *ET AL.,* CORROSION OF WELDMENTS, *METALS HANDBOOK,* VOL 13, 9TH ED., *CORROSION,* ASM INTERNATIONAL, 1987, P 344-368
30. E.J. MORGAN-WARREN, ATMOSPHERE CONTROL CRITERIA FOR WELDING TITANIUM, *ADVANCES IN WELDING PROCESS,* VOL 1, THE WELDING INSTITUTE, ABINGTON HALL, ABINGTON, CAMBRIDGE, 1978
31. G.G. PSARES, THE EFFECTS OF MOISTURE CONTENTS OF SHIELDING GASES ON THE FORMATION OF PORES WHEN ALUMINUM IS WELDED, *AUTOMAT. WELD.,* VOL 32 (NO. 10), P 6-8
32. D. NORTHCOTE, THE GAS TUNGSTEN ARC WELDING PROCESSES--ELECTRODES AND SHIELDING GASES--CRITERIA FOR OPTIMUM WELD QUALITY, *AUST. WELD. J.,* VOL 29 (NO. 1), AUTUMN 1984, P 18-20
33. K.F. KRYSIAK AND P.M. BHADHA, SHIELDING GAS PURIFICATION IMPROVES WELD QUALITY, *WELD. J.,* VOL 69 (NO. 11), NOV 1990, P 47-49
34. R.H. ESPY, HOW CORROSION AND WELDING CONDITIONS AFFECT CORROSION RESISTANCE OF WELDMENTS IN TYPE 430 STAINLESS STEEL, PAPER NO. 22, PRESENTED AT *CORROSION 68* (HOUSTON, TX), NATIONAL ASSOCIATION OF CORROSION ENGINEERS, 1968
35. K.F. KRYSIAK, WELDABILITY OF FERRITIC STAINLESS STEELS--AN OVERVIEW, *WELD. J.,* APRIL 1986
36. L. BERMUDEZ DE HELGUERO, WELDING OF HIGHLY ALLOYED DISSIMILAR STAINLESS

STEELS (USING STABILIZED AUSTENITIC OR FERRITIC STAINLESS STEEL FILLERS), *SIDERURGIA,* VOL 48, DEC 1988, P 55-60

37. J.J. DEMO, *CORROSION,* VOL 27 (NO. 3), 1971, P 531
38. R.H. HODGES, *CORROSION,* VOL 27 (NO. 3), 1971, P 119
39. R.H. HODGES, *CORROSION,* VOL 27 (NO. 3), 1971, P 164
40. E. BAERLECKEN, W.A. FISCHER, AND K. LORENZ, UNTERSUCHUNGEN ÜBER DAS UMWANDLUNGSVERHALTEN, DIE KERBSCHLAGZÄHIGKEIT UND DIE NEIGUNG ZUR INTERKRISTALLINEN KORROSION VON EISEN-CHROM-LEGIERUNGEN MIT CHROMGEHALTEN BIS 30% [INQUIRY INTO THE TRANSFORMATION, NOTCHED IMPACT TOUGHNESS, AND THE SUSCEPTIBILITY TOWARD INTERCRYSTALLINE CORROSION OF IRON-CHROMIUM ALLOYS WITH CHROMIUM CONTENTS UP TO 30%], *STAHL EISEN,* VOL 81 (NO. 12), JUNE 1961, P 768-778
41. G. HERBSLEB AND W. SCHEWENK, *WERKST. KORROS.,* VOL 19, 1968, P 103
42. J.J. DEMO AND A.P. BOND, INTERGRANULAR CORROSION AND EMBRITTLEMENT OF FERRITIC STAINLESS STEELS, VOL 31 (NO. 1), JAN 1975, P 21-22
43. J.J. DEMO, *METALL. TRANS.,* VOL 5 (NO. 11), NOV 1974, P 2253
44. M. SEMCHYSHEN, A.P. BOND, AND J.H. DUNDAS, *PROC. SYMP. TOWARD IMPROVING DUCTILITY AND TOUGHNESS* (KYOTO, JAPAN), 1971, P 239
45. B. POLLARD, *MET. TECHNOL.,* VOL 1, PART 1, JAN 1974, P 31
46. J. HOCHMANN, "THE INFLUENCE OF VACUUM FUSION ON THE PROPERTIES OF 25% CR FERRITIC STEELS," PH.D. THESIS, PARIS, FRANCE, 1950
47. *METALS HANDBOOK,* 9TH ED., VOL 13, *CORROSION,* ASM INTERNATIONAL, 1987, P 355-358
48. J.F. GRUBB, PH.D. THESIS, RENSSELAER POLYTECHNIC INSTITUTE, TROY, NY, MAY 1982
49. W.O. BINDER AND H.R. SPINDELOW, *TRANS. ASM,* VOL 43, 1951, P 759
50. J.F. GRUBB, STABILIZATION OF HIGH-CHROMIUM, FERRITIC STAINLESS STEELS, *PROC. INT. CONF. STAINLESS STEELS* (CHIBA, JAPAN), 10-13 JUNE 1991, ISIJ, P 541-548
51. K.E. PINNOW AND J.P. BRESSANELLI, "WELDED FERRITIC STAINLESS STEEL ARTICLES," U.S. PATENT 4,119,765, 10 OCT 1978

Note cited in this section

* * SENSITIZATION OF FERRITIC STAINLESS STEELS DIFFERS FROM THAT OF AUSTENITIC STAINLESS STEELS BECAUSE OF THE LOWER SOLUBILITY OF (C + N) IN FERRITE AND THE MUCH GREATER (APPROXIMATELY 10^3 TIMES) DIFFUSION RATE OF (C + N) AND CHROMIUM IN FERRITE THAN IN AUSTENITE.

Selection of Wrought Ferritic Stainless Steels

K.F. Krysiak, Hercules, Inc.; J.F. Grubb, Allegheny Ludlum Corporation; B. Pollard; R.D. Campbell, Joining Services, Inc.

References

1. *METALS HANDBOOK,* VOL 6, 8TH ED., *WELDING AND BRAZING,* AMERICAN SOCIETY FOR METALS, 1971, P 248
2. WORKING DATA, CARPENTER STAINLESS STEELS, 1971
3. W.A. MATEJKA AND R.J. KNOTH, E-BRITE 26-1, THE HIGH-PURITY APPROACH TO HIGH CHROMIUM FERRITIC STAINLESS STEELS, *ASM SYMPOSIUM,* BAL HARBOUR, FLORIDA, ASTM, 6 DEC 1973

4. R.N. WRIGHT, MECHANICAL BEHAVIOR AND WELDABILITY OF HIGH-CHROMIUM FERRITIC STAINLESS STEELS AS A FUNCTION OF PURITY, *WELD. J.,* OCT 1971, P 434S-440S
5. I.A. FRANSON, MECHANICAL PROPERTIES OF HIGH-PURITY FE-26CR-1MO FERRITIC STAINLESS STEEL, *METALL. TRANS.,* VOL 5, NOV 1974, P 2257-2264
6. J.J. DEMO, WELDABLE AND CORROSION-RESISTANT FERRITIC STAINLESS STEELS, *METALL. TRANS.,* VOL 5, NO. 11, NOV 1974, P 2253-2256
7. K.E. PINNOW, J.P. BRASSANELLI, AND A. MOSKOWITZ, A NEW HIGH-CHROMIUM FERRITIC STAINLESS STEEL, *MET. ENG. QUART.,* AUG 1975, P 32-41
8. R.F. STEIGERWALD, LOW INTERSTITIAL FE-CR-MO FERRITIC STAINLESS, *SOVIET-AMERICAN SYMPOSIUM ON NEW DEVELOPMENTS IN THE FIELD OF MOLYBDENUM-ALLOYED CAST IRON AND STEEL,* MOSCOW, JAN 1973
9. A.P. BOND, *TRANS. TMS-AIME,* VOL 245, 1969, P 2127
10. M.A. STREICHER, DEVELOPMENT OF PITTING-RESISTANT FE-CR-MO ALLOYS, *CORROSION,* VOL 30 (NO. 3), 1974, P 77-91
11. G.B. HUNTER AND T.W. EAGAR, DUCTILITY OF STABILIZED FERRITIC STAINLESS STEELS WELDS
12. B. POLLARD, DUCTILITY OF FERRITIC STAINLESS WELD METAL, *WELD. J.,* APR 1972, P 222S-230S
13. G. ROTHSCHILD, WELDING OF E-BRITE 26-1, PROCEDURE NO. W-2, WELDING SEMINAR HELD IN HOUSTON, TEXAS, 13 SEPT 1973
14. K.E. DORSCHU, WELDABILITY OF A NEW FERRITIC STAINLESS STEEL, *WELD. J.,* SEPT 1971
15. R.N. WRIGHT, "THE TOUGHNESS OF FERRITIC STAINLESS STEELS," MATERIALS ENGINEERING DEPARTMENT, RENSSELAER POLYTECHNIC INSTITUTE, TROY, NY, 1979
16. D.H. KAH AND D.W. DICKINSON, WELDABILITY OF FERRITIC STAINLESS STEEL, *WELD. J.,* AUG 1981, P 135S-142S
17. K.F. KRYSIAK, WELDABILITY OF THE NEW GENERATION OF FERRITIC STAINLESS STEELS--PART I, PAPER NO. 92 PRESENTED AT THE NATIONAL ASSOCIATION OF CORROSION ENGINEERS CONFERENCE (TORONTO, CANADA), 14-18 APRIL 1975
18. K.F. KRYSIAK, WELDABILITY OF THE NEW GENERATION OF FERRITIC STAINLESS STEELS--UPDATE, *SYMPOSIUM ON FERRITIC STAINLESS STEELS* (SAN FRANCISCO, CA), 23-24 MAY 1979
19. R. LOWRIE, WELDING E-BRITE 26-1 TO OTHER ALLOYS, *WELD. J.,* NOV 1973, P 500S-506S
20. Y. OKAZAKI, R. TODOROKI, T. SAKAMOTO, AND T. ZAIZEN, ON THE PROPERTIES OF HIGH-PURITY 19 CR-2 MO FERRITIC STAINLESS STEEL WELDS, PAPER NO. 118, PRESENTED AT THE NATIONAL ASSOCIATION OF CORROSION ENGINEERS CONFERENCE (TORONTO, CANADA), 6-10 APRIL 1981
21. J.M. SAWHILL, JR. AND A.P. BOND, DUCTILITY AND TOUGHNESS OF STAINLESS STEEL WELDS, *WELD. J.,* VOL 55 (NO. 10), 1976, P 33S-41S
22. J.F. GRUBB AND S.D. WASHKO, THE EFFECT OF NIOBIUM AND TITANIUM DUAL STABILIZATION ON THE WELDABILITY OF 11% CHROMIUM FERRITIC STAINLESS STEELS, *PROC. INT. CONF. STAINLESS STEELS* (CHIBA, JAPAN), 10-13 JUNE 1991, ISIJ, P 1062-1068
23. J.C. VILLAFUERTE AND H.W. KERR, ELECTROMAGNETIC STIRRING AND GRAIN REFINEMENT IN STAINLESS STEEL GTA WELDS, *WELD. J.,* VOL 69 (NO. 1), 1990, P 1S-13S
24. J.C. VILLAFUERTE, E. PARDO, AND H.W. KERR, THE EFFECT OF ALLOY COMPOSITION AND WELDING COMPOSITIONS ON COLUMNAR-EQUIAXED TRANSITIONS IN FERRITIC STAINLESS STEEL GAS-TUNGSTEN ARC WELDS, *METALL. TRANS. A,* VOL 21 (NO. 7), JULY 1990, P 2009-2019
25. R.D. CAMPBELL, FERRITIC STAINLESS STEEL WELDING METALLURGY, *KEY ENGINEERING MATERIALS,* VOL 69 & 70, TRANS TECH PUBLICATIONS, SWITZERLAND, 1992, P 167-216
26. J.D. GATES AND R.A. JAGO, ABSORPTION OF GASEOUS CONTAMINANTS BY WELDS AND

WELD SIMULATIONS IN FERRITIC STAINLESS STEELS, *MATER. SCI. TECHNOL.,* VOL 3, MAY 1987, P 386-393

27. J.M. SAWHILL, "DUCTILITY AND TOUGHNESS OF AUTOGENOUS GAS TUNGSTEN-ARC WELDS IN 18CR-2MO STEELS," RP-33-73-21, CLIMAX MOLYBDENUM CO. OF MICHIGAN, 24 FEB 1975
28. J.F. GRUBB, HYDROGEN EMBRITTLEMENT OF SUPERFERRITIC STAINLESS STEELS, *NEW DEVELOPMENTS IN STAINLESS STEEL TECHNOLOGY,* R.A. LULA, ED., AMERICAN SOCIETY FOR METALS, 1985, P 175-187
29. K.F. KRYSIAK *ET AL.,* CORROSION OF WELDMENTS, *METALS HANDBOOK,* VOL 13, 9TH ED., *CORROSION,* ASM INTERNATIONAL, 1987, P 344-368
30. E.J. MORGAN-WARREN, ATMOSPHERE CONTROL CRITERIA FOR WELDING TITANIUM, *ADVANCES IN WELDING PROCESS,* VOL 1, THE WELDING INSTITUTE, ABINGTON HALL, ABINGTON, CAMBRIDGE, 1978
31. G.G. PSARES, THE EFFECTS OF MOISTURE CONTENTS OF SHIELDING GASES ON THE FORMATION OF PORES WHEN ALUMINUM IS WELDED, *AUTOMAT. WELD.,* VOL 32 (NO. 10), P 6-8
32. D. NORTHCOTE, THE GAS TUNGSTEN ARC WELDING PROCESSES--ELECTRODES AND SHIELDING GASES--CRITERIA FOR OPTIMUM WELD QUALITY, *AUST. WELD. J.,* VOL 29 (NO. 1), AUTUMN 1984, P 18-20
33. K.F. KRYSIAK AND P.M. BHADHA, SHIELDING GAS PURIFICATION IMPROVES WELD QUALITY, *WELD. J.,* VOL 69 (NO. 11), NOV 1990, P 47-49
34. R.H. ESPY, HOW CORROSION AND WELDING CONDITIONS AFFECT CORROSION RESISTANCE OF WELDMENTS IN TYPE 430 STAINLESS STEEL, PAPER NO. 22, PRESENTED AT *CORROSION 68* (HOUSTON, TX), NATIONAL ASSOCIATION OF CORROSION ENGINEERS, 1968
35. K.F. KRYSIAK, WELDABILITY OF FERRITIC STAINLESS STEELS--AN OVERVIEW, *WELD. J.,* APRIL 1986
36. L. BERMUDEZ DE HELGUERO, WELDING OF HIGHLY ALLOYED DISSIMILAR STAINLESS STEELS (USING STABILIZED AUSTENITIC OR FERRITIC STAINLESS STEEL FILLERS), *SIDERURGIA,* VOL 48, DEC 1988, P 55-60
37. J.J. DEMO, *CORROSION,* VOL 27 (NO. 3), 1971, P 531
38. R.H. HODGES, *CORROSION,* VOL 27 (NO. 3), 1971, P 119
39. R.H. HODGES, *CORROSION,* VOL 27 (NO. 3), 1971, P 164
40. E. BAERLECKEN, W.A. FISCHER, AND K. LORENZ, UNTERSUCHUNGEN ÜBER DAS UMWANDLUNGSVERHALTEN, DIE KERBSCHLAGZÄHIGKEIT UND DIE NEIGUNG ZUR INTERKRISTALLINEN KORROSION VON EISEN-CHROM-LEGIERUNGEN MIT CHROMGEHALTEN BIS 30% [INQUIRY INTO THE TRANSFORMATION, NOTCHED IMPACT TOUGHNESS, AND THE SUSCEPTIBILITY TOWARD INTERCRYSTALLINE CORROSION OF IRON-CHROMIUM ALLOYS WITH CHROMIUM CONTENTS UP TO 30%], *STAHL EISEN,* VOL 81 (NO. 12), JUNE 1961, P 768-778
41. G. HERBSLEB AND W. SCHEWENK, *WERKST. KORROS.,* VOL 19, 1968, P 103
42. J.J. DEMO AND A.P. BOND, INTERGRANULAR CORROSION AND EMBRITTLEMENT OF FERRITIC STAINLESS STEELS, VOL 31 (NO. 1), JAN 1975, P 21-22
43. J.J. DEMO, *METALL. TRANS.,* VOL 5 (NO. 11), NOV 1974, P 2253
44. M. SEMCHYSHEN, A.P. BOND, AND J.H. DUNDAS, *PROC. SYMP. TOWARD IMPROVING DUCTILITY AND TOUGHNESS* (KYOTO, JAPAN), 1971, P 239
45. B. POLLARD, *MET. TECHNOL.,* VOL 1, PART 1, JAN 1974, P 31
46. J. HOCHMANN, "THE INFLUENCE OF VACUUM FUSION ON THE PROPERTIES OF 25% CR FERRITIC STEELS," PH.D. THESIS, PARIS, FRANCE, 1950
47. *METALS HANDBOOK,* 9TH ED., VOL 13, *CORROSION,* ASM INTERNATIONAL, 1987, P 355-358

48. J.F. GRUBB, PH.D. THESIS, RENSSELAER POLYTECHNIC INSTITUTE, TROY, NY, MAY 1982
49. W.O. BINDER AND H.R. SPINDELOW, *TRANS. ASM,* VOL 43, 1951, P 759
50. J.F. GRUBB, STABILIZATION OF HIGH-CHROMIUM, FERRITIC STAINLESS STEELS, *PROC. INT. CONF. STAINLESS STEELS* (CHIBA, JAPAN), 10-13 JUNE 1991, ISIJ, P 541-548
51. K.E. PINNOW AND J.P. BRESSANELLI, "WELDED FERRITIC STAINLESS STEEL ARTICLES," U.S. PATENT 4,119,765, 10 OCT 1978

Selection of Wrought Austenitic Stainless Steels

John A. Brooks, Sandia National Laboratories; John C. Lippold, Edison Welding Institute

---

Introduction

AUSTENITIC STAINLESS STEELS exhibit a single-phase, face-centered cubic (fcc) structure that is maintained over a wide range of temperatures. This structure results from a balance of alloying additions that stabilize the austenite phase from elevated to cryogenic temperatures. Because these alloys are predominantly single phase, they can only be strengthened by solid-solution alloying or by work hardening. The exceptions are the precipitation-strengthened austenitic stainless steels that are discussed in the article "Selection of Wrought Precipitation-Hardening Stainless Steels" in this Section of the Volume.

The austenitic stainless steels were developed for use in both mild and severe corrosive conditions. They are also used at temperatures that range from cryogenic temperatures, where they exhibit high toughness, to elevated temperatures of nearly 600 °C (1110 °F), where they exhibit good oxidation resistance. Because the austenitic materials are nonmagnetic, they are sometimes used in applications where magnetic materials are not acceptable.

The most common types of austenitic stainless steels are the UNS S20000 and S30000 alloys (AISI 200 and 300 series). Within these two grades, the alloying additions, which are chosen to provide the desired properties at reasonable cost, can be considerably different. Furthermore, alloying additions and specific alloy composition can have a major effect on weldability and the as-welded microstructure. The AISI 300 series of alloys typically contain from 8 to 20% Ni and from 16 to 25% Cr. Minor alloying additions are approximately 1% Si (maximum), which is used as a deoxidizer; 0.02 to 0.08% C, which is used as an austenite stabilizer; and approximately 1.5% Mn, which is used both as an austenite stabilizer and as a sulfur and silicon compound former. A more-complete compositional range of common AISI 200- and 300-series alloys is shown in Table 1. The less common alloys, including some of the superaustenitic stainless steels, are identified in Table 2.

TABLE 1 COMPOSITIONS OF STANDARD AUSTENITIC STAINLESS STEELS

TYPE	UNS NO.	COMPOSITION, %$^{(A)}$							
		C	Mn	Si	Cr	Ni	P	S	Other
201	S20100	0.15	5.5-7.5	1.00	16.0-18.0	3.5-5.5	0.06	0.03	0.25 N
202	S20200	0.15	7.5-10.0	1.00	17.0-19.0	4.0-6.0	0.06	0.03	0.25 N
205	S20500	0.12-0.25	14.0-15.5	1.00	16.5-18.0	1.0-1.75	0.06	0.03	0.32-0.40 N
301	S30100	0.15	2.0	1.00	16.0-18.0	6.0-8.0	0.045	0.03	. . .
302	S30200	0.15	2.0	1.00	17.0-19.0	8.0-10.0	0.045	0.03	. . .
302B	S30215	0.15	2.0	2.0-3.0	17.0-19.0	8.0-10.0	0.045	0.03	. . .

303	S30300	0.15	2.0	1.00	17.0-19.0	8.0-10.0	0.20	0.15 MIN	0.6 MO(B)
303SE	S30323	0.15	2.0	1.00	17.0-19.0	8.0-10.0	0.20	0.06	0.15 MIN SE
304	S30400	0.08	2.0	1.00	18.0-20.0	8.0-10.5	0.045	0.03	. . .
304H	S30409	0.04-0.10	2.0	1.00	18.0-20.0	8.0-10.5	0.045	0.03	. . .
304L	S30403	0.03	2.0	1.00	18.0-20.0	8.0-12.0	0.045	0.03	. . .
304LN	S30453	0.03	2.0	1.00	18.0-20.0	8.0-12.0	0.045	0.03	0.10-0.16 N
302CU	S30430	0.08	2.0	1.00	17.0-19.0	8.0-10.0	0.045	0.03	3.0-4.0 CU
304N	S30451	0.08	2.0	1.00	18.0-20.0	8.0-10.5	0.045	0.03	0.10-0.16 N
305	S30500	0.12	2.0	1.00	17.0-19.0	10.5-13.0	0.045	0.03	. . .
308	S30800	0.08	2.0	1.00	19.0-21.0	10.0-12.0	0.045	0.03	. . .
309	S30900	0.20	2.0	1.00	22.0-24.0	12.0-15.0	0.045	0.03	. . .
309S	S30908	0.08	2.0	1.00	22.0-24.0	12.0-15.0	0.045	0.03	. . .
310	S31000	0.25	2.0	1.50	24.0-26.0	19.0-22.0	0.045	0.03	. . .
310S	S31008	0.08	2.0	1.50	24.0-26.0	19.0-22.0	0.045	0.03	. . .
314	S31400	0.25	2.0	1.5-3.0	23.0-26.0	19.0-22.0	0.045	0.03	. . .
316	S31600	0.08	2.0	1.00	16.0-18.0	10.0-14.0	0.045	0.03	2.0-3.0 MO
316F	S31620	0.08	2.0	1.00	16.0-18.0	10.0-14.0	0.20	0.10 MIN	1.75-2.5 MO
316H	S31609	0.04-0.10	2.0	1.00	16.0-18.0	10.0-14.0	0.045	0.03	2.0-3.0 MO
316L	S31603	0.03	2.0	1.00	16.0-18.0	10.0-14.0	0.045	0.03	2.0-3.0 MO
316LN	S31653	0.03	2.0	1.00	16.0-18.0	10.0-14.0	0.045	0.03	2.0-3.0 MO; 0.10-0.16 N
316N	S31651	0.08	2.0	1.00	16.0-18.0	10.0-14.0	0.045	0.03	2.0-3.0 MO; 0.10-0.16 N
317	S31700	0.08	2.0	1.00	18.0-20.0	11.0-15.0	0.045	0.03	3.0-4.0 MO
317L	S31703	0.03	2.0	1.00	18.0-20.0	11.0-15.0	0.045	0.03	3.0-4.0 MO
321	S32100	0.08	2.0	1.00	17.0-19.0	9.0-12.0	0.045	0.03	5 × %C MIN TI
321H	S32109	0.04-0.10	2.0	1.00	17.0-19.0	9.0-12.0	0.045	0.03	5 × %C MIN TI
330	N08330	0.08	2.0	0.75-1.5	17.0-20.0	34.0-37.0	0.04	0.03	. . .
347	S34700	0.08	2.0	1.00	17.0-	9.0-	0.045	0.03	10 × %C MIN NB

					19.0	13.0			
347H	S34709	0.04-0.10	2.0	1.00	17.0-19.0	9.0-13.0	0.045	0.03	8 × %C MIN - 1.0 MAX NB
348	S34800	0.08	2.0	1.00	17.0-19.0	9.0-13.0	0.045	0.03	0.2 CO; 10 × %C MIN NB; 0.10 TA
348H	S34809	0.04-0.10	2.0	1.00	17.0-19.0	9.0-13.0	0.045	0.03	0.2 CO; 10 × %C MIN - 1.0 MAX NB; 0.10 TA
384	S38400	0.08	2.0	1.00	15.0-17.0	17.0-19.0	0.045	0.03	. . .

(A) SINGLE VALUES ARE MAXIMUM VALUES UNLESS OTHERWISE INDICATED.

(B) OPTIONAL

TABLE 2 COMPOSITIONS OF NONSTANDARD AUSTENITIC STAINLESS STEELS

DESIGNATION(A)	**UNS NO.**	**COMPOSITION, %**(B)							
		C	**Mn**	**Si**	**Cr**	**Ni**	**P**	**S**	**Other**
GALL-TOUGH	S20161	0.15	4.00-6.00	3.00-4.00	15.0-18.0	4.00-6.00	0.040	0.040	0.08-0.20 N
203 EZ (XM-I)	S20300	0.08	5.0-6.5	1.00	16.0-18.0	5.0-6.5	0.040	0.18-0.35	0.5 MO; 1.75-2.25 CU
NITRONIC 50 (XM-19)	S20910	0.06	4.0-6.0	1.00	20.5-23.5	11.5-13.5	0.040	0.030	1.5-3.0 MO; 0.2-0.4 N; 0.1-0.3 NB; 0.1-0.3 V
TENELON (XM-31)	S21400	0.12	14.5-16.0	0.3-1.0	17.0-18.5	0.75	0.045	0.030	0.35 N
CRYOGENIC TENELON (XM-14)	S21460	0.12	14.0-16.0	1.00	17.0-19.0	5.0-6.0	0.060	0.030	0.35-0.50 N
ESSHETE 1250	S21500	0.15	5.5-7.0	1.20	14.0-16.0	9.0-11.0	0.040	0.030	0.003-0.009 B; 0.75-1.25 NB; 0.15-0.40 V
TYPE 216 (XM-17)	S21600	0.08	7.5-9.0	1.00	17.5-22.0	5.0-7.0	0.045	0.030	2.0-3.0 MO; 0.25-0.50 N
TYPE 216 L (XM-18)	S21603	0.03	7.5-9.0	1.00	17.5-22.0	7.5-9.0	0.045	0.030	2.0-3.0 MO; 0.25-0.50 N
NITRONIC 60	S21800	0.10	7.0-9.0	3.5-4.5	16.0-18.0	8.0-9.0	0.040	0.030	0.08-0.18 N
NITRONIC 40 (XM-10)	S21900	0.08	8.0-10.0	1.00	19.0-21.5	5.5-7.5	0.060	0.030	0.15-0.40 N
21-6-9 LC	S21904	0.04	8.00-10.00	1.00	19.00-21.50	5.50-7.50	0.060	0.030	0.15-0.40 N
NITRONIC 33 (18-3 MN)	S24000	0.08	11.50-14.50	1.00	17.00-19.00	2.50-3.75	0.060	0.030	0.20-0.40 N
NITRONIC 32 (18-2 MN)	S24100	0.15	11.00-14.00	1.00	16.50-19.50	0.50-2.50	0.060	0.030	0.20-0.40 N
18-18 PLUS	S28200	0.15	17.0-19.0	1.00	17.5-19.5	. . .	0.045	0.030	0.5-1.5 MO; 0.5-1.5 CU; 0.4-0.6 N
303 PLUS X (XM-5)	S30310	0.15	2.5-4.5	1.00	17.0-19.0	7.0-10.0	0.020	0.25 MIN	0.6 MO
MVMA(C)	S30415	0.05	0.60	1.30	18.5	9.50	. . .	. . .	0.15 N; 0.04

									CE
304BI(D)	S30424	0.08	2.00	0.75	18.00-20.00	12.00-15.00	0.045	0.030	0.10 N; 1.00-1.25 B
304 HN (XM-21)	S30452	0.04-0.10	2.00	1.00	18.0-20.0	8.0-10.5	0.045	0.030	0.16-0.30 N
CRONIFER 1815 LCSI	S30600	0.018	2.00	3.73-4.3	17.0-18.5	14.0-15.5	0.020	0.020	0.2 MO
RA 85 H(C)	S30615	0.20	0.80	3.50	18.5	14.50	. . .	. . .	1.0 AL
253 MA	S30815	0.05-0.10	0.80	1.4-2.0	20.0-22.0	10.0-12.0	0.040	0.030	0.14-0.20 N; 0.03-0.08 CE; 1.0 AL
TYPE 309 S CB	S30940	0.08	2.00	1.00	22.0-24.0	12.0-15.0	0.045	0.030	10 × %C MIN TO 1.10 MAX NB
TYPE 310 CB	S31040	0.08	2.00	1.50	24.0-26.0	19.0-22.0	0.045	0.030	10 × %C MIN TO 1.10 MAX NB + TA
254 SMO	S31254	0.20	1.00	0.80	19.50-20.50	17.50-18.50	0.030	0.010	6.00-6.50 MO; 0.50-1.00 CU; 0.180-0.220 N
TYPE 316 TI	S31635	0.08	2.00	1.00	16.0-18.0	10.0-14.0	0.045	0.030	5 × %(C + N) MIN TO 0.70 MAX TI; 2.0-3.0 MO; 0.10 N
TYPE 316 CB	S31640	0.08	2.00	1.00	16.0-18.0	10.0-14.0	0.045	0.030	10 × %C MIN TO 1.10 MAX NB + TA; 2.0-3.0 MO; 0.10 N
TYPE 316 HQ	. . .	0.030	2.00	1.00	16.00-18.25	10.00-14.00	0.030	0.015	3.00-4.00 CU; 2.00-3.00 MO
TYPE 317 LM	S31725	0.03	2.00	1.00	18.0-20.0	13.5-17.5	0.045	0.030	4.0-5.0 MO; 0.10 N
17-14-4 LN	S31726	0.03	2.00	0.75	17.0-20.0	13.5-17.5	0.045	0.030	4.0-5.0 MO; 0.10-0.20 N
TYPE 317 LN	S31753	0.03	2.00	1.00	18.0-21.0	11.0-15.0	0.030	0.030	0.10-0.22 N
TYPE 370	S37000	0.03-0.05	1.65-2.35	0.5-1.0	12.5-14.5	14.5-16.5	0.040	0.010	1.5-2.5 MO; 0.1-0.4 TI; 0.005 N; 0.05 CO
18-18-2 (XM-15)	S38100	0.08	2.00	1.5-2.5	17.0-19.0	17.5-18.5	0.030	0.030	. . .
19-9 DL	S63198	0.28-0.35	0.75-1.50	0.03-0.8	18.0-21.0	8.0-11.0	0.040	0.030	1.0-1.75 MO; 0.1-0.35 TI; 1.0-1.75 W; 0.25-0.60 NB
20CB-3	N08020	0.07	2.00	1.00	19.0-	32.0-	0.045	0.035	2.0-3.0 MO;

					21.0	38.0			3.0-4.0 CU; 8 × %C MIN TO 1.00 MAX NB
20MO-4	N08024	0.03	1.00	0.50	22.5-25.0	35.0-40.0	0.035	0.035	3.50-5.00 MO; 0.50-1.50 CU; 0.15-0.35 NB
20MO-6	N08026	0.03	1.00	0.50	22.00-26.00	33.00-37.20	0.03	0.03	5.00-6.70 MO; 2.00-4.00 CU
SANICRO 28	N08028	0.02	2.00	1.00	26.0-28.0	29.5-32.5	0.020	0.015	3.0-4.0 MO; 0.6-1.4 CU
AL-6X	N08366	0.035	2.00	1.00	20.0-22.0	23.5-25.5	0.030	0.030	6.0-7.0 MO
AL-6XN	N08367	0.030	2.00	1.00	20.0-22.0	23.50-25.50	0.040	0.030	6.0-7.0 MO; 0.18-0.25 N
JS-700	N08700	0.04	2.00	1.00	19.0-23.0	24.0-26.0	0.040	0.030	4.3-5.0 MO; 8 × %C MIN TO 0.5 MAX NB; 0.5 CU; 0.005 PB; 0.035 S
TYPE 332	N08800	0.01	1.50	1.00	19.0-23.0	30.0-35.0	0.045	0.015	0.15-0.60 TI; 0.15-0.60 AL
904L	N08904	0.02	2.00	1.00	19.0-23.0	23.0-28.0	0.045	0.035	4.0-5.0 MO; 1.0-2.0 CU
CRONIFER 1925 HMO	N08925	0.02	1.00	0.50	24.0-26.0	19.0-21.0	0.045	0.030	6.0-7.0 MO; 0.8-1.5 CU; 0.10-0.20 N
CRONIFER 2328	. . .	0.04	0.75	0.75	22.0-24.0	26.0-28.0	0.030	0.015	2.5-3.5 CU; 0.4-0.7 TI; 2.5-3.0 MO

(A) XM DESIGNATIONS IN THIS COLUMN ARE ASTM DESIGNATIONS FOR THE LISTED ALLOY.

(B) SINGLE VALUES ARE MAXIMUM VALUES UNLESS OTHERWISE INDICATED.

(C) NOMINAL COMPOSITIONS.

(D) UNS DESIGNATION HAS NOT BEEN SPECIFIED; THIS DESIGNATION APPEARS IN ASTM A 887 AND MERELY INDICATES THE FORM TO BE USED.

Depending on alloy composition, the austenitic stainless steels may solidify with a microstructure containing some retained ferrite at room temperature, as a result of welding. Weld cracking, of which the most common form is solidification cracking, can be another consequence of welding. Cracks can occur in various regions of the weld with different orientations, such as centerline cracks, transverse cracks, and microcracks in the underlying weld metal or adjacent heat-affected zone (HAZ). These cracks are due, primarily, to low-melting liquid phases, which allow boundaries to separate under the thermal and shrinkage stresses resulting from weld solidification and cooling. Other possible metallurgical consequences of welding are:

- THE PRECIPITATION OF INTERGRANULAR, CHROMIUM-RICH $M_{23}C_6$ CHROMIUM CARBIDES IN THE WELD HAZ, WHICH CAN LEAVE THESE REGIONS SENSITIVE TO CORROSION
- THE TRANSFORMATION OF WELD FERRITE TO SIGMA PHASE DURING ELEVATED-

TEMPERATURE SERVICE, WHICH CAN REDUCE DUCTILITY AND TOUGHNESS

These phenomena can be either minimized or eliminated through alloy selection, process optimization, or postweld heat treatment. In general, these alloys are very weldable. With an understanding of the evolution of the weld microstructure and the ability to control it, one can produce welds to meet a wide variety of service requirements.

Alloy Types. One of the most important decisions of engineering design is the selection of the appropriate stainless steel type for the desired application and service. The factors for selection are covered more thoroughly in *Properties and Selection: Irons, Steels, and High-Performance Alloys,* Volume 1 of the *ASM Handbook.* However, when the engineering design includes welding as a fabrication process, then alloy selection must consider weldability and weld performance.

An important criterion for alloy selection in welding applications may be to use low-carbon grades to prevent chromium carbide precipitation and sensitization to intergranular corrosion. On the other hand, additions of sulfur or selenium may be used to improve machinability, although these elements can result in severe weld hot-cracking problems and, thus, may not be suitable when welding is required.

In reviewing the composition of the alloys listed in Tables 1 and 2, one can see that a wide range of alloy additions exists, as well as a fairly wide range in composition within the specification of individual alloy types. These two factors can have major effects on the welding behavior of the different alloy types, in addition to heat-to-heat variations in weldability. One goal of this article is to provide an understanding of the effects of composition on welding behavior, and another is to provide guidance in the selection and welding of these materials.

General Welding Characteristics. The austenitic stainless steels are generally considered the most weldable of the stainless steels. Because of their physical properties, their welding behavior may be considerably different than those of the ferritic, martensitic, and duplex stainless steels. For example, the thermal conductivity of typical austenitic alloys is only approximately half that of the ferritic steels. Therefore, the weld heat input that is required to achieve the same penetration is considerably reduced. In contrast, the coefficient of thermal expansion (CTE) of austenite is 30 to 40% greater than that of ferrite, which can result in increases in both distortion and residual stress, because of welding. The molten weld pool of the austenitic stainless steels also tends to be more viscous, or "sluggish," than ferritic and martensitic grades. This impedes the metal flow and wettability of welds in these materials, which may promote lack-of-fusion defects.

A major concern, when welding the austenitic stainless steels, is the susceptibility to solidification and liquation cracking. As discussed below, these materials can be very resistant to these forms of high-temperature cracking, if the material is compositionally balanced such that the solidification behavior and microstructure is controlled to ensure that the weld metal contains more than 3 vol% ferrite.

In cases where fully austenitic welds are required, such as when the weld must be nonmagnetic or when it is placed in corrosive environments that selectively attack the ferrite phase, the welds will solidify as austenite and the propensity for weld cracking will increase. In some alloys, such as AISI 310 and the superaustenitic grades, all the allowable compositions within the specification range solidify as austenite when welded. To minimize cracking in these welds, it is generally advisable to weld with low heat input and under low constraint conditions.

Welds that are made at slower speeds and produce elliptical-, rather than teardrop-, shaped pools are also generally less susceptible to cracking. This effect is particularly pronounced when welding thin sheet, as in the production of thin-walled tubing. Residual elements, which form low-melting liquid phases that promote cracking, should be kept to a minimum. These elements include phosphorus, sulfur, boron, selenium, niobium, silicon and titanium. Small additions of oxygen and nitrogen are somewhat beneficial and are thought to affect the wetting characteristics of the liquid phases. However, high concentrations of these elements may promote porosity. Manganese also can reduce cracking susceptibility, primarily by tying up sulfur and silicon that would otherwise be available to form low-melting phases.

Microstructural Development

Although austenitic stainless steels are predominantly austenitic, they often contain small amounts of body-centered cubic (bcc) ferrite, particularly in the weld metal. This ferrite is often described as "delta" ferrite, because it forms at elevated temperatures and is distinguished from "alpha" ferrite, which is the low-temperature form in iron-base alloys. In this article, the term ferrite will refer to high-temperature delta ferrite, unless noted otherwise. These alloys also may contain martensite, although the presence of this phase is unusual and limited to special composition and temperature ranges, forming only as a result of plastic deformation (Ref 1).

The solidification behavior and microstructural development of the weld metal of these alloys are quite complex, and, until recently, were not well understood. In this class of alloys, weldability and subsequent weld performance are often directly related to weld microstructure. Thus, when welding these alloys, a general understanding of the welding metallurgy can be very beneficial in optimizing weld performance. Because chemical composition has the greatest influence on weld microstructure, a number of empirical relationships and constitution diagrams have been developed to predict microstructure based on actual or approximated composition. In all cases, the concept of chromium equivalence (Cr_{eq}) and nickel equivalence (Ni_{eq}) has been used to normalize the effect of various alloying additions on the ferrite-forming and austenite-forming potency, respectively. Considerable disagreement still exists regarding these equivalency relationships (Table 3) for austenitic stainless steels.

TABLE 3 CHROMIUM- AND NICKEL-EQUIVALENCY RELATIONSHIPS FOR AUSTENITIC STAINLESS STEELS

AUTHOR	**YEAR**	**CHROMIUM EQUIVALENT, WT%**	**NICKEL EQUIVALENT, WT%**
SCHAEFFLER	1949	CR + MO + 1.5SI + 0.5NB	NI + 0.5MN + 30C
DELONG *ET AL.*	1956	CR + MO + 1.5SI + 0.5NB	NI + 0.5MN + 30C +30N
HULL	1973	CR + 1.21MO + 0.48SI + 0.14NB + 2.27V + 0.72W + 2.20TI + 0.21TA + 2.48AL	NI +(0.11MN - 0.0086MN^2) + 24.5C + 14.2N + 0.41CO + 0.44CU
HAMMAR AND SVENNSON	1979	CR + 1.37MO + 1.5SI + 2NB + 3TI	NI + 0.31MN + 22C + 14.2N + CU
SIEWERT *ET AL.*	1992	CR + MO + 0.7NB	NI + 35C + 20N + 0.25CU

The Schaeffier diagram (Ref 2) was developed in the late 1940s in an attempt to predict weld microstructure for a wide range of stainless steels. This diagram was refined as the DeLong diagram (Ref 3), which pertained to the range of compositions associated with 300-series stainless steels and filler metals. In 1988, the Welding Research Council introduced a further improvement, when compared with the DeLong diagram, which was called the WRC-1988 diagram (Ref 4). This diagram was further modified in 1992 by adding copper to the Ni_{eq} formula. The WRC-1992 diagram (Fig. 1) allows the amount of weld-metal ferrite to be predicted using newly developed Cr_{eq} and Ni_{eq} formulas. The use of this diagram will be described later.

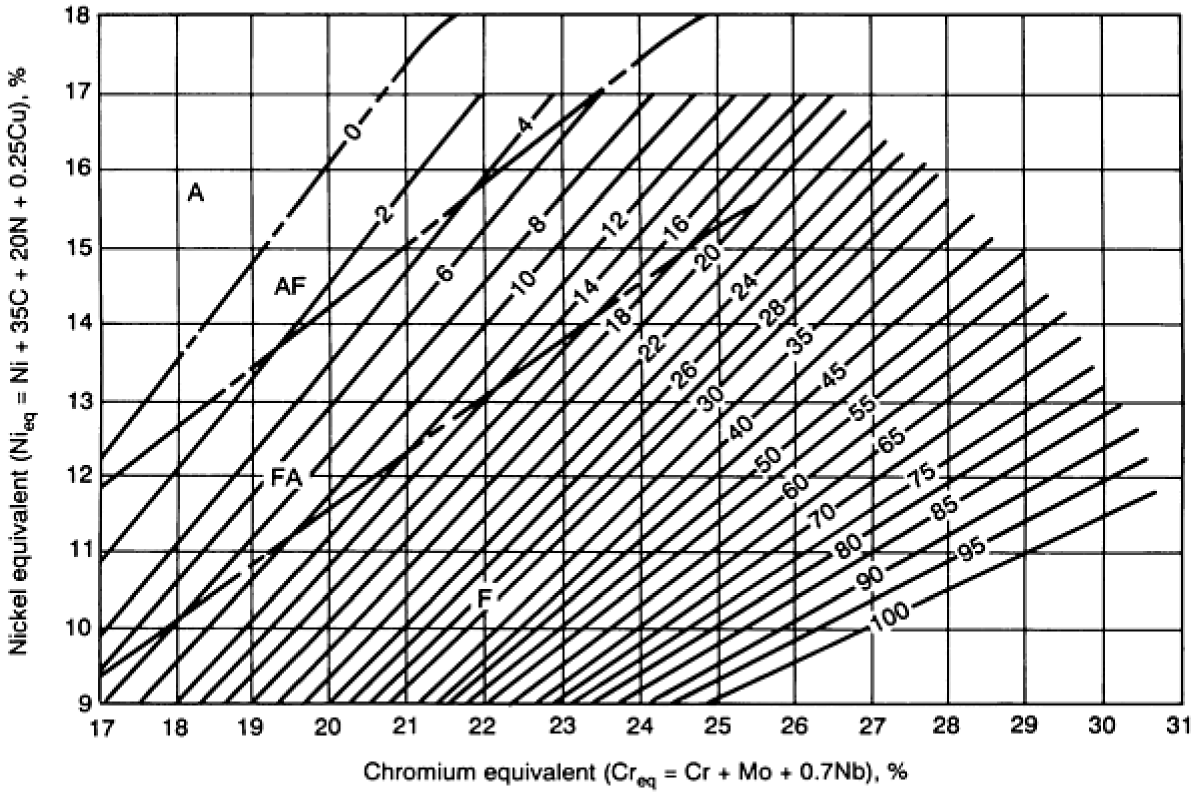

FIG. 1 WRC-1992 DIAGRAM FOR PREDICTING WELD FERRITE CONTENT AND SOLIDIFICATION MODE. SOURCE: REF 4

The following sections in this article summarize the important aspects of solidification behavior and microstructural evolution that dictate weld-metal ferrite content and morphology. An understanding of microstructural evolution and its ramifications in austenitic stainless steel welds is crucial, because it will dictate base- and filler-metal selection, as well as welding conditions.

References cited in this section

1. J.C. LIPPOLD, M.C. JUHAS, AND E.N.C. DALDER, *METALL. TRANS. A,* VOL 16, 1985, P 1835-1848
2. A.L. SCHAEFFLER, *MET. PROG.,* VOL 26, 1949, P 680
3. C.J. LONG AND W.T. DELONG, *WELD. J.,* VOL 52, 1973, P 281-S TO 297-S
4. D. KOTECKI AND T.A. SIEWERT, *WELD. J.,* VOL 71, 1992, P 171S-179S

Selection of Wrought Austenitic Stainless Steels

John A. Brooks, Sandia National Laboratories; John C. Lippold, Edison Welding Institute

Solidification Behavior

In austenitic stainless steels, the primary phase of solidification may be either austenite or ferrite. This is dictated predominantly by composition. It has been well documented (Ref 5, 6, 7) that welds that solidify as primary austenite are inherently more susceptible to weld solidification cracking than welds that solidify as primary ferrite. Thus, control of weld solidification behavior is critical to ensuring crack-free welds.

The iron-nickel-chromium ternary diagram assembled by Speich (Ref 8) can be used to discuss, in general terms, the solidification process. Figure 2 shows both liquidus and solidus surfaces. A eutectic trough extends to the ternary eutectic composition of roughly 49Cr-43Ni-8Fe. A peritectic trough at Fe-4Ni extends into the diagram to about 75% iron, at which location the liquidus crosses the nickel-rich solidus, changing from peritectic to eutectic behavior (Ref 9).

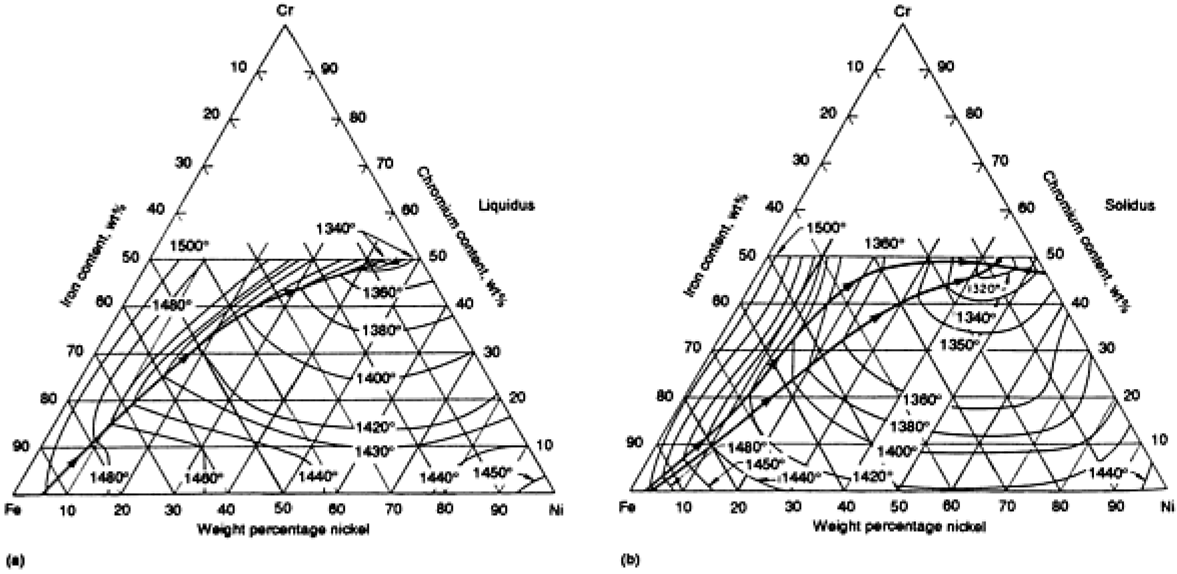

FIG. 2 PROJECTIONS OF (A) LIQUIDUS AND (B) SOLIDUS SURFACES, AS ISOTHERMAL CONTOURS ON IRON-NICKEL-CHROMIUM TERNARY DIAGRAM. TEMPERATURES SHOWN ARE IN °F. SOURCE: REF 8

In Fig. 2, it can be seen that for approximately 60 to 70% Fe, which is the iron content of many of the austenitic stainless steels, the liquidus lies along the chromium/nickel ratio of approximately 1.5. Compositions on the nickel-rich side of the liquidus will solidify as primary austenite, sometimes referred to as A-type solidification. Depending on composition, it may solidify with a small fraction of eutectic ferrite (AF-type solidification). Alloys on the chromium-rich side of the liquidus will solidify as primary ferrite with the peritectic or eutectic solidification of austenite (FA-type solidification). However, the solidification boundary between ferrite and austenite can be shifted, because of the nonequilibrium solidification conditions of welding. For compositions that are further from the peritectic/eutectic liquidus, the welds solidify completely as ferrite at higher chromium/nickel ratios (F-type solidification). The chromium/nickel ratio for single-phase ferrite solidification is often reported to range from 1.9 to 2.0.

A number of diagrams can be used to predict the primary phase of solidification based on the weld-metal composition. The WRC-1992 diagram (Fig. 1) can be used to predict solidification behavior, in addition to ferrite content. Note that this diagram contains four solidification regimes, designated as A, AF, FA, and F. The austenite-ferrite microstructural morphologies associated with these regimes are described below.

References cited in this section

5. H. THIER, *DVS-BER.*, VOL 41, 1976, P 100-104
6. I. MATSUMOTO, K. TAMAKI, AND M. MATSUMA, *J. JPN. WELD. SOC.*, VOL 41 (NO. 11), 1972, P 1306-1314
7. J.C. LIPPOLD AND W.F. SAVAGE, *WELD. J.*, VOL 61, 1982, P 388S-396S
8. G.R. SPEICH, *METALS HANDBOOK*, 8TH ED., VOL 8, AMERICAN SOCIETY FOR METALS, 1973, P 424
9. S.E. SCHURMANN AND I. BRAUCKMANN, *ARCH. EISENHUTTEN.*, VOL 48, 1977, P 3-8

Selection of Wrought Austenitic Stainless Steels

John A. Brooks, Sandia National Laboratories; John C. Lippold, Edison Welding Institute

Solid-State Transformations and Ferrite Morphologies

The final weld microstructure is controlled by both the solidification behavior and the subsequent solid-state transformations. A number of schematics have been published (Ref 10, 11, 12, 13), such as the one by Brooks *et al.* (Ref 11) shown in Fig. 3, which depicts some of the general characteristics of microstructural development. However, a much wider variety of microstructures exists in actual welds. The vertical sections of the iron-nickel-chromium ternary diagram for 70% and 60% Fe (Fig. 4) are useful in discussing the solid-state transformations. As indicated by these diagrams, during the cooling of primary austenite welds (AF-type), some of the eutectic ferrite may transform to austenite. The primary ferrite-solidified welds that exhibit peritectic/eutectic behavior during the final stages of solidification (FA-type) typically solidify with a large fraction of ferrite (Fig. 3). During cooling through the $\delta + \gamma$ two-phase field, much of the ferrite transforms to austenite, with chromium partitioning to the ferrite and nickel partitioning to the austenite, leaving ferrite only within the very cores of the original dendrites. When only several percent of the original ferrite is retained, which is often referred to as skeletal, or vermicular, ferrite, it may be difficult to distinguish between it and the eutectic ferrite that results from primary austenite solidification, unless good metallographic techniques are used (Fig. 3a and b). However, the location of the ferrite is different. It resides within the core of the dendrites, in the case of FA-type solidification, rather than at the dendrite boundaries, in the case of AF-type solidification.

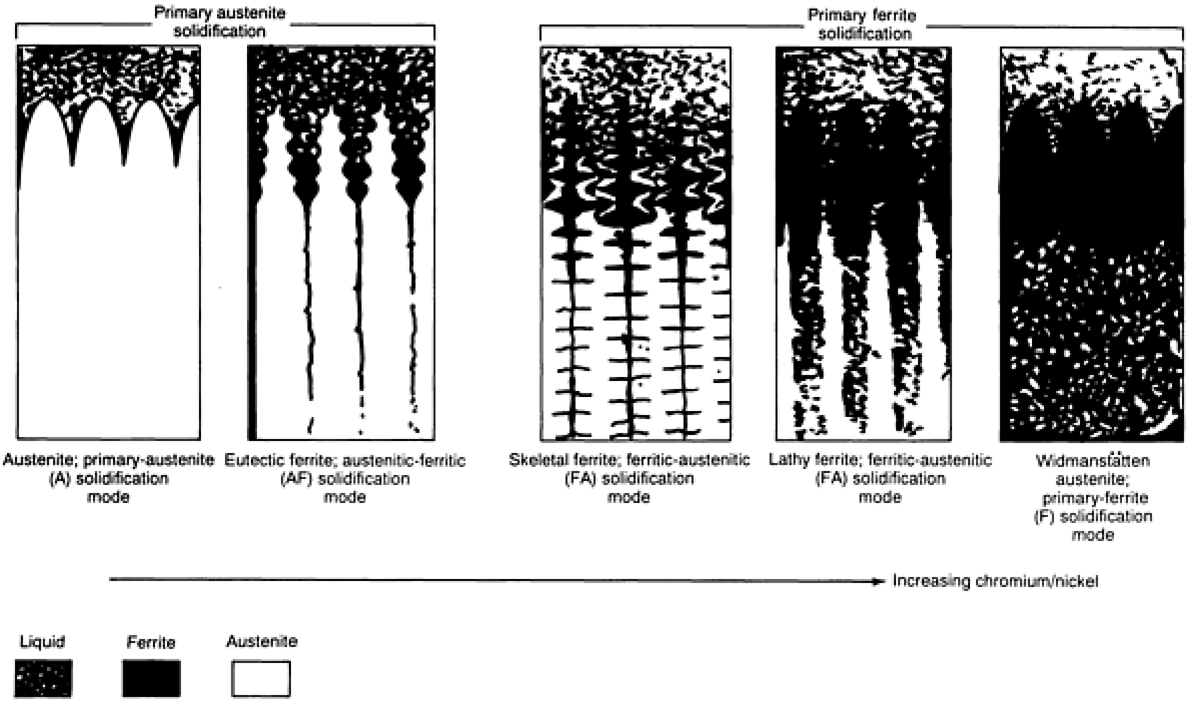
Primary austenite solidification
Primary ferrite solidification
Austenite; primary-austenite (A) solidification mode
Eutectic ferrite; austenitic-ferritic (AF) solidification mode
Skeletal ferrite; ferritic-austenitic (FA) solidification mode
Lathy ferrite; ferritic-austenitic (FA) solidification mode
Widmanstätten austenite; primary-ferrite (F) solidification mode
Increasing chromium/nickel
Liquid
Ferrite
Austenite

FIG. 3 SCHEMATICS SHOWING SOLIDIFICATION AND SOLID-STATE TRANSFORMATION BEHAVIOR OF WELDS WITH INCREASING CR_{EQ}/NI_{EQ} RATIOS. SOURCE: REF 11

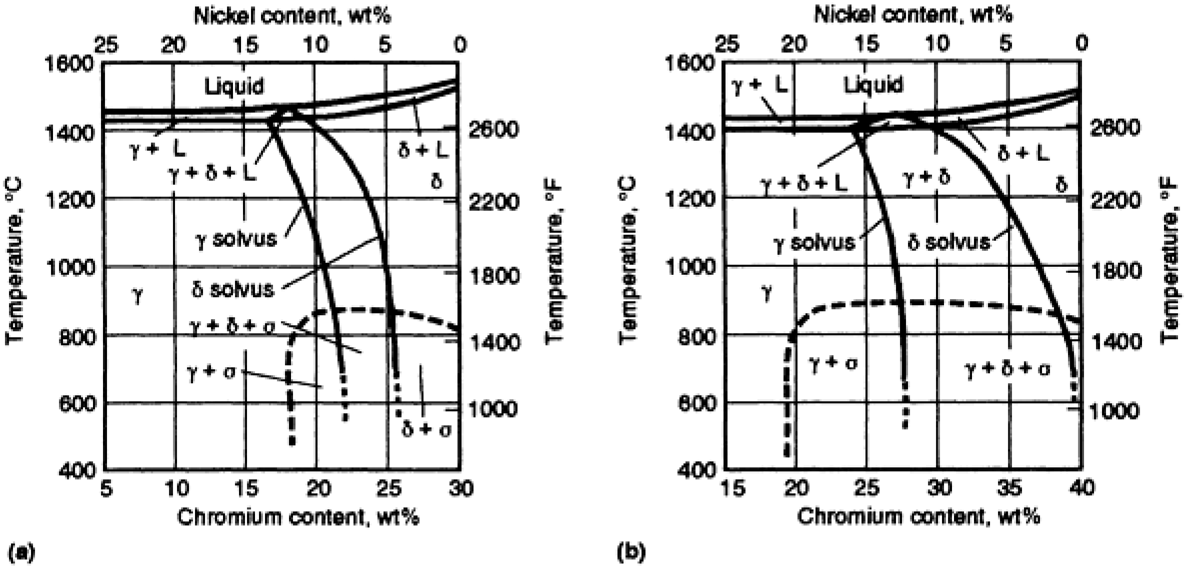

FIG. 4 VERTICAL SECTIONS OF IRON-NICKEL-CHROMIUM TERNARY DIAGRAM, AT CONSTANT IRON CONTENT. (A) 70% FE. (B) 60% FE. SOURCE: REF 14

At higher chromium/nickel ratios, other ferrite morphologies are often observed, such as the lathy structure shown in Fig. 5(c). It is generally reported that the skeletal and lathy ferrite morphologies are indicative of primary ferrite with the secondary solidification of austenite, but these ferrite morphologies can also exist in welds that solidify as single-phase ferrite, as shown in Fig. 3(c) and (d). At higher chromium/ nickel ratios, a Widmanstätten austenite structure can also form in welds that solidify as single-phase ferrite. It should be noted that the solidification and transformation behavior shown in Fig. 3 is dependent not only on composition, but also on cooling rate. Extremely high solidification and cooling rates, such as those encountered during electron-beam or laser-beam welding, may significantly alter the microstructure relative to that predicted using the WRC-1992 diagram. This behavior is described in the section "Solidification Behavior and Microstructure of High-Energy-Density Welds."

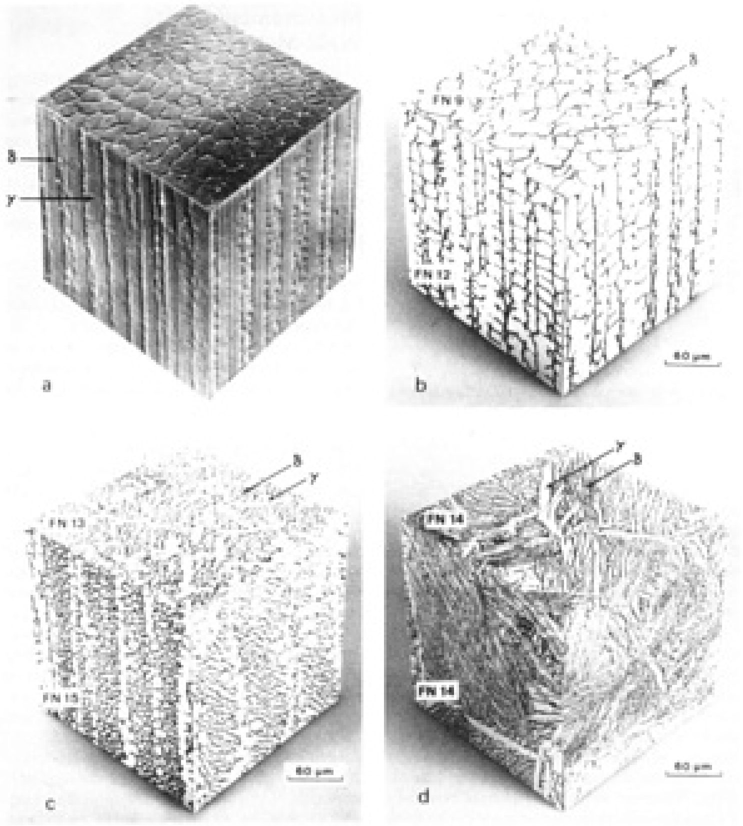

FIG. 5 THREE-DIMENSIONAL VIEWS OF TYPICAL AUSTENITIC STAINLESS STEEL WELD MORPHOLOGIES; SOLIDIFICATION DIRECTION, VERTICAL. IN EACH PHOTOGRAPH, FERRITE IS DARKER PHASE. (A) CELLULAR PRIMARY AUSTENITE SOLIDIFICATION STRUCTURE WITH INTERCELLULAR EUTECTIC FERRITE IN AISI 310. (B) SKELETAL FERRITE, 9-12 FN, IN AISI 308. (C) LATHY OR LACY FERRITE, 13-15 FN IN AISI 308 MULTIPASS WELD. (D) ACICULAR OR WIDMANSTÄTTEN AUSTENITE, 14 FN, IN UPPER PORTION OF AISI 308 WELD. WELDS IN (B)-(D) SOLIDIFIED AS PRIMARY FERRITE. SOURCE: REF 15

Micrographs of the three-dimensional structure of several weld ferrite morphologies compiled by David (Ref 15) are shown in Fig. 5. Small amounts of ferrite along solidification cell boundaries of a primary austenite structure are shown in Fig. 5(a). This very cellular-appearing solidification structure is characteristic of welds that solidify as austenite and results primarily from the microsegregation of chromium and, to a lesser extent, nickel, along the solidification cell boundaries. Depending on alloy composition and welding conditions, final solidification may occur as a eutectic of ferrite and austenite. Figure 5(b) and (c) show ferrite along cell cores of primary ferrite structures, whereas Fig. 5(d) shows acicular ferrite or Widmanstätten austenite in a structure that solidified completely as ferrite. Additional details on solidification behavior and microstructural development are provided in Ref 11 and 16.

Alloying Effects on Solidification Behavior, Weld Microstructure, and Cracking Susceptibility. The relationship between solidification behavior and a composition factor, in terms of Cr_{eq}/Ni_{eq}, was established by Suutala and Moisio (Ref 17) using coefficients of: $Ni_{eq} = Ni + 0.3Mn + 22C + 14.2N + Cu$; and $Cr_{eq} = Cr + 1.37Mo + 1.5Si + 2Nb + 3Ti$ (see Table 3). A diagram (Fig. 6) was developed that delineates the four solidification types, A, AF, FA, and F, as described previously. Using their data, the demarcation between single-phase austenite solidification (A) and austenite solidification with the eutectic solidification of ferrite (FA) occurs at a Cr_{eq}/Ni_{eq} ratio of approximately 1.35. The transition to solidification as primary ferrite with the eutectic/peritectic solidification of austenite (FA) occurs at a Cr_{eq}/Ni_{eq} of approximately 1.5. This value has been found to be valid for most conventional 300-series alloys welded under normal arc-welding conditions. This ratio increases with increasing solidification velocity. At the lower bounds of primary ferrite, some regions may solidify as ferrite and others as austenite (Ref 18). In this case, the region where primary austenite is most commonly located is along the weld centerline, where the solidification velocity is maximum. This transition between primary austenite and primary ferrite is especially important, because of the strong relationship between solidification behavior and solidification-cracking susceptibility.

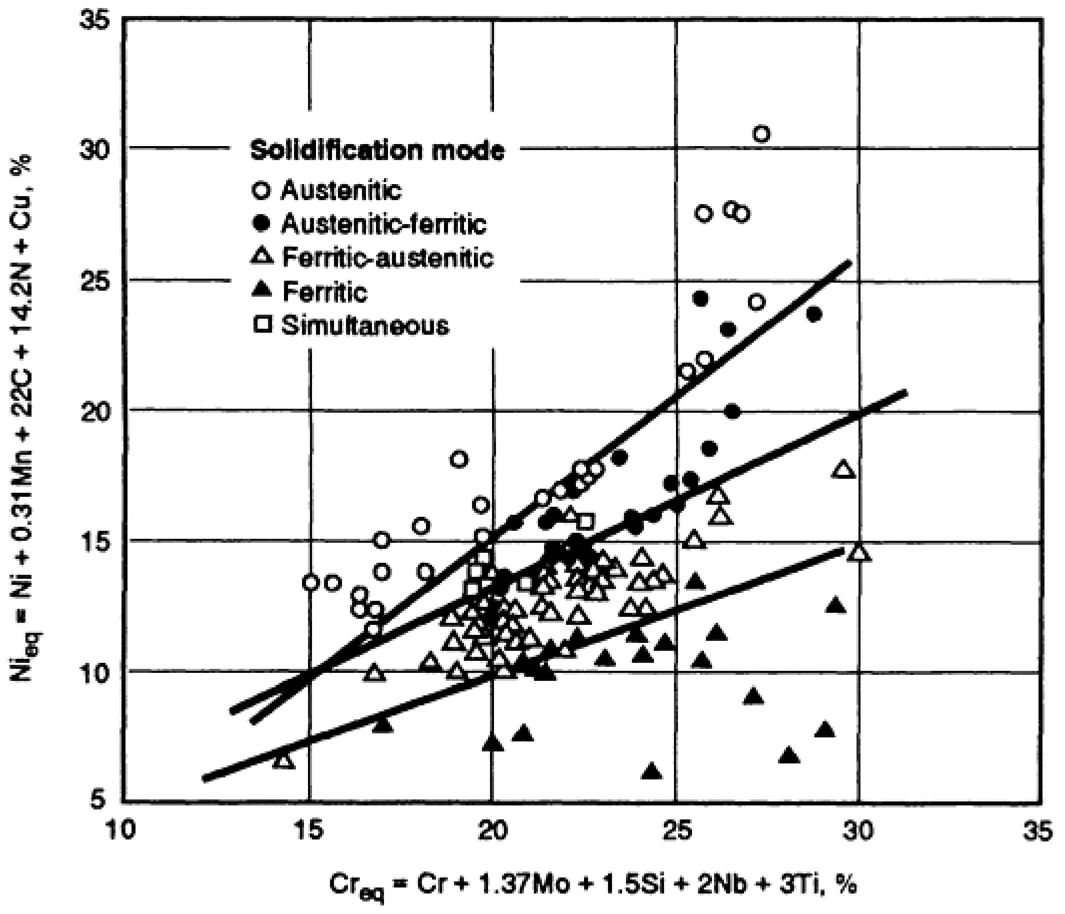

FIG. 6 SOLIDIFICATION MODE (EXPERIMENTAL) FOR VARIETY OF COMPOSITIONS, COMPARED ON BASIS OF CR_{EQ}/NI_{EQ} RATIO. SOURCE: REF 17

As indicated in Fig. 6, above a Cr_{eq}/Ni_{eq} ratio of approximately 1.95, solidification occurs as single-phase ferrite. However, other studies (Ref 13) indicate that single-phase ferrite solidification occurs at considerably lower Cr_{eq}/Ni_{eq} ratios, and that the ratio of approximately 1.95 may more closely correlate to microstructures containing Widmanstätten austenite.

Note that the Cr_{eq}/Ni_{eq} ratios defined above are valid only when using the Suutala and Moisio equivalency relationships. Corresponding ratios for delineating the solidification behavior by using the WRC-1992 equivalencies are somewhat different. However, discrepancies between the two diagrams exist, as is apparent for lean alloys or pure iron-nickel-chromium ternaries. For example, the transition from primary austenite to primary ferrite is reported to occur at a Cr_{eq}/Ni_{eq} ratio of 1.5 by Suutala and Moisio (Ref 17), versus 1.4 in the WRC diagram.

Takalo, Suutala, and Moisio (Ref 18) also described the solidification-cracking susceptibility of austenitic stainless steels by plotting their Cr_{eq}/Ni_{eq} ratios versus impurity content in terms of sulfur plus phosphorus. Their results are shown in Fig. 7. Note that a dramatic transition in cracking susceptibility occurs at a value of approximately 1.5. This transition represents the shift in solidification behavior from primary austenite to primary ferrite, as described above, and reinforces the importance of controlling solidification behavior in order to avoid weld solidification cracking.

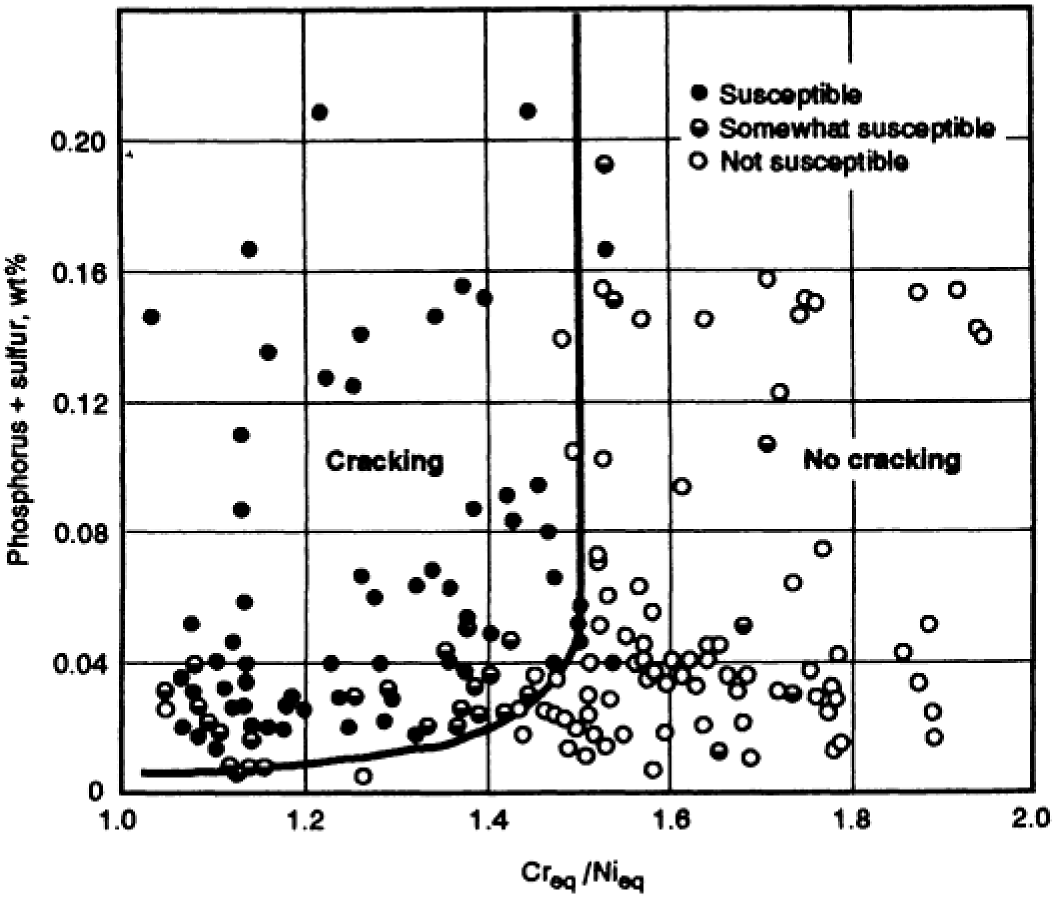

FIG. 7 RELATIONSHIP BETWEEN SOLIDIFICATION CRACKING SUSCEPTIBILITY AND CR_{EQ}/NI_{EQ} RATIO. BOUNDARY BETWEEN CRACKING AND NO CRACKING AT CR_{EQ}/NI_{EQ} = 1.5 CORRESPONDS TO CHANGE IN SOLIDIFICATION MODE FROM PRIMARY AUSTENITE BELOW 1.5 TO PRIMARY FERRITE ABOVE 1.5. SOURCE: REF 18

As noted previously, Schaeffler developed one of the first diagrams for predicting weld microstructures. This diagram is especially useful for predicting microstructures between dissimilar alloys using a tie line between the two alloy compositions and the appropriate alloy dilution. The diagram was developed using the shielded metal arc welding (SMAW) process, where the nominal nitrogen level was about 0.06 wt% and, as such, was not shown as an alloying addition. This diagram described the behavior of most of the alloys in use at that time, and is relatively accurate for most 300-series alloys when using conventional welding processes. The use of this diagram is discussed in more detail in the article "Welding of Stainless Steels." Other diagrams that have been developed, often using limited alloy compositional ranges, are reviewed in detail in Ref 19.

The WRC-1992 diagram, developed primarily with data from welds made by SMAW, is the most accurate diagram and is preferred in terms of the composition ranges for which it was established. In this diagram, the alloying elements have been limited to nickel, carbon, nitrogen, and copper in the Ni_{eq}, and to chromium, molybdenum, and niobium in the Cr_{eq}. However, the diagram is applicable for manganese contents up to 10%, molybdenum contents up to 3%, nitrogen contents up to 0.2%, and silicon contents up to 1%.

When the diagram is used for alloy systems containing significant amounts of alloying elements that have not been included in the diagram, predictions may be inaccurate. As discussed above, boundaries between the different solidification modes are also included on the diagram, although some disagreement exists with the boundaries established by other investigators. However, the most important boundary by far is the one separating primary ferrite from primary austenite solidification, in which some discrepancy still exists. The diagram does predict that the maximum amount of eutectic ferrite will not exceed a few percent, which is consistent with that observed in practice.

When using these diagrams to predict weld behavior, one must be aware that uncertainties in compositional analysis, as well as welding process and parameters, can affect predicted solidification behavior and ferrite content. One common occurrence in the welding of stainless steels is the possibility of nitrogen adsorption during welding when inadequate shielding is used. It is not uncommon that nitrogen levels of approximately 0.02% in the base material can be increased to values as high as 0.1%. This will significantly decrease ferrite content and may also result in a change in solidification mode from primary ferrite to primary austenite, which can have a drastic effect on weld solidification-cracking susceptibility. Thus, the nitrogen content of the deposited weld material must be used in making the predictions. However, with the use of the diagrams, magnetic measurements of ferrite number, and an understanding of weld microstructural evolution, one can reasonably predict the nature of the weld microstructure.

References cited in this section

10. N. SUUTALA, T. TAKALO, AND T. MOISIO, *METALL. TRANS. A,* VOL 10, 1979, P 512-514
11. J.A. BROOKS, J.C. WILLIAMS, AND A.W. THOMPSON, IN *TRENDS IN WELDING RESEARCH,* S.A. DAVID, ED., AMERICAN SOCIETY FOR METALS, 1982, P 331-355
12. S. KATAYAMA AND A. MATSUNAWA, *PROC. INT. CONF. ON ADVANCED LASER ELECTRO-OPTICS (ICALEO)* (SAN FRANCISCO), 1985
13. J.A. BROOKS, N.C.Y. YANG, AND J.S. KRAFCIK, ON THE ORIGIN OF FERRITE MORPHOLOGIES OF PRIMARY FERRITE SOLIDIFIED STAINLESS STEEL WELDS, *TRENDS IN WELDING RESEARCH,* ASM INTERNATIONAL, 1993
14. J.W. PUGH AND J.D. NISBET, *TRANS. TMS-AIME,* VOL 188, 1950, P 268-276
15. S.A. DAVID, *WELD. J.,* VOL 60, 1981, P 63-S TO 71-S
16. J.A. BROOKS AND A.W. THOMPSON, *LNT. MATER. REV.,* VOL 36 (NO. 1), 1991, P 16-44
17. N. SUUTALA AND T. MOISIO, THE USE OF CHROMIUM AND NICKEL EQUIVALENTS IN CONSIDERING SOLIDIFICATION PHENOMENA IN AUSTENITIC STAINLESS STEELS, *SOLIDIFICATION TECHNOLOGY IN THE FOUNDRY AND CASTHOUSE,* THE METALS SOCIETY PREPRINT, LONDON, 1980
18. T. TAKALO, N. SUUTALA, AND T. MOISIO, *METALL. TRANS. A* , VOL 10, 1979, P 1173-1181
19. D.L. OLSON, *WELD. J.,* VOL 64, 1985, P 281S-295S

Selection of Wrought Austenitic Stainless Steels

John A. Brooks, Sandia National Laboratories; John C. Lippold, Edison Welding Institute

---

Measurement of Weld-Metal Ferrite

As described previously, weld-metal ferrite content significantly influences both the weldability and service performance of austenitic stainless steels. Weld-metal ferrite content can either be predicted using constitution diagrams, such as the WRC-1992 diagram (Fig. 1), or measured using instruments that take advantage of the ferromagnetic characteristics of the ferrite phase.

The term *Ferrite Number*, designated FN, has been adopted as a relative measure for quantifying ferrite content using standardized magnetic techniques (Ref 20). The FN approach was developed in order to reduce the large variation in ferrite levels determined on a given specimen when measured using different techniques by different laboratories. FN approximates the "volume percent ferrite" at levels below 8 FN. Above this level, deviation occurs, where the FN value exceeds the actual volume percent ferrite. For example, a weld metal with 16 FN contains approximately 13.8 vol%.

A number of instruments are commercially available for determining the ferrite content of welds, including the Magne gage, Severn gage, and ferrite scope (Ref 20). The Severn gage and ferrite scope are particularly applicable for use in the field or on the production floor. The ferrite scope is also useful in measuring ferrite on welds of small cross section, as is often the case with electron-beam welds. Calibration procedures for magnetic measurement techniques have been recommended by the American Welding Society in AWS A4.2 (Ref 20). These procedures use either the thickness standards of the National Institute of Standards and Technology (NIST) or actual weld-metal standards that include certified levels of weldmetal ferrite.

However, it is important to understand that it is impossible to accurately determine the absolute ferrite content of austenitic stainless steel weld metals. Even on undiluted weld pads, ferrite variations from pad to pad must be expected, because of slight changes in welding and measuring variables. On a large group of pads made from one heat or lot, using a standard pad welding and preparation procedure, two sigma values indicate that 95% of the tests are expected to be within a range of approximately ±2.2 FN at about 8 FN. If different pad welding and preparation procedures are used, then these variations will increase. Variations are also introduced based on the composition of the ferrite. The magnetic attraction of the ferrite may vary significantly with chemical composition. Thus, for the same volume percent ferrite, the FN determined by magnetic instruments may vary for weld pads made using different filler materials.

Even larger variations in ferrite content may be encountered if the welding technique allows excessive nitrogen pickup, in which case the ferrite can be much lower than otherwise expected. A nitrogen pickup of 0.10% typically will decrease the FN by about 8.

Austenitic stainless steel base materials are intentionally balanced compositionally to produce inherently lower ferrite content than matching weld metals. This is done, in part, to facilitate the hot workability of the material during forging, extrusion, and rolling operations. For this reason, special attention is required when making autogenous welds if primary ferrite solidification is desired to prevent solidification cracking. In general, weld metal diluted with base metal will also be somewhat lower in ferrite than the undiluted weld metal. For example, this effect is commonly observed when welding types 304 or 304L base metals with ER308 or ER309 filler materials.

The agreement between the predicted and measured weld-metal ferrite content is also strongly dependent on the accuracy of the chemical analysis. Variations in the results of the chemical analyses encountered from laboratory to laboratory can have significant effects on the predicted ferrite level, particularly if these variations occur in the detection of carbon and nitrogen. It is not uncommon for predicted and measured ferrite levels to differ by as much as 4 to 8 FN.

Nitrogen-Strengthened Stainless Steels. The addition of nitrogen to austenitic stainless steels significantly improves the strength and pitting resistance of the alloy, and may also improve cryogenic toughness. These alloys may contain relatively high levels of manganese, because manganese increases the solubility of nitrogen in the austenite matrix, in addition to substituting for nickel as an austenite stabilizer.

In general, the weldability of these alloys is similar to the 300-series materials. The presence of ferrite in the as-deposited weld microstructure reduces weld-cracking susceptibility for the same reasons described previously. The prediction of weld-metal ferrite content in these alloys has historically been problematic, because the DeLong diagram did not adequately account for large concentrations of nitrogen and manganese. Both Espy and Hull developed equivalency relationships and predictive diagrams that better coped with high levels of nitrogen. The Espy diagram (Ref 21) represented a modification of the DeLong diagram and accounted for higher nitrogen and manganese levels. Hull (Ref 22) developed equivalencies (see Table 3) to predict ferrite levels in high-nitrogen, high-manganese stainless steels. The Ni_{eq} of 14.2 developed by Hull for nitrogen is significantly lower than the value of 20.0 used in the WRC-1992 diagram. However, the omission of manganese in the diagram is in agreement with the small Ni_{eq} assigned by Hull. Limited data suggest that the Hull and WRC-1992 diagrams are in fair agreement when predicting weld ferrite contents in these alloys with nitrogen contents as high as approximately 0.25%. At the higher nitrogen contents, one may expect that the WRC-1992 diagram would predict higher ferrite contents than that of the Hull diagram. From Table 3, it can be seen that Hammer and Svennson (Ref 23) used the same nickel equivalent developed by Hull.

Because of the high nitrogen contents of these steels, weld porosity may sometimes be encountered. The critical level of nitrogen, in terms of porosity problems, is lower for electron-beam welds made in vacuum (approximately 0.25%) than it is for welds made by gas-tungsten arc welding (GTAW), where porosity problems are usually not encountered at nitrogen levels less than approximately 0.35%.

References cited in this section

20. "STANDARD PROCEDURES FOR CALIBRATING MAGNETIC INSTRUMENTS TO MEASURE THE DELTA FERRITE CONTENT OF AUSTENITIC AND DUPLEX AUSTENITIC-FERRITIC STEEL

WELD METAL," AWS 4.2-91, AWS, 1991

21. R.H. ESPY, *WELD. J.,* VOL 61, 1982, P 1492-156S
22. F.C. HULL, *WELD. J.,* VOL 46, 1967, P 399-S TO 409-S
23. O. HAMMER AND U. SVENNSON, *SOLIDIFICATION AND CASTING OF METALS,* THE METALS SOCIETY, LONDON, 1979, P 401-410

Selection of Wrought Austenitic Stainless Steels

John A. Brooks, Sandia National Laboratories; John C. Lippold, Edison Welding Institute

---

Solidification Behavior and Microstructure of High-Energy-Density Welds

Differences exist between the solidification behavior and microstructures of high-energy-density welds, such as electron-beam and laser-beam welds, and more-conventional welds, such as shielded metal arc and gas-tungsten arc welds (Ref 12, 16, 24). These differences have been attributed to the rapid solidification velocities and cooling rates of the high-energy-density welds and, if not recognized, can have a drastic effect on weld quality and performance.

One characteristic of the high cooling rates during high-energy-density welding is a fine solidification cell size. However, changes in solidification behavior and solid-state transformation, compared with the behavior and transformation observed in gas-tungsten arc welds of the same alloy composition, can also occur. One of the most important effects of the rapid solidification rates of high-energy-density welds is a shift in the solidification transition from primary austenite to primary ferrite to higher $\mathrm{Cr_{eq}/Ni_{eq}}$ ratios than those encountered in conventional arc welds. A change in solidification mode from primary ferrite to primary austenite, which resulted in centerline cracking of an electron-beam weld in type 304L, is shown in Fig. 8. It must be recognized that similar changes in solidification mode from primary ferrite to primary austenite can occur along the centerline of welds made with the more-conventional welding processes, such as the GTAW process (Ref 25). This is especially common at higher welding speeds in materials with $\mathrm{Cr_{eq}/Ni_{eq}}$ ratios (Suutala and Moisio equivalents) slightly above 1.5.

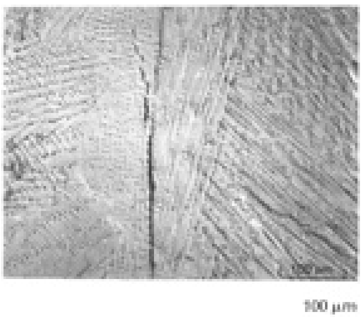

FIG. 8 SOLIDIFICATION CRACK IN ELECTRON-BEAM WELD ALONG WELD CENTER IN REGION WHERE SOLIDIFICATION OCCURRED AS PRIMARY AUSTENITE AS A RESULT OF HIGHER SOLIDIFICATION AND COOLING RATES

Several microstructures can also exist in high-energy-density welds that have not been reported in welds of more-moderate cooling rates. Basically, these microstructures consist of a single phase of either ferrite or austenite. An example of the single-phase austenite structure is shown on the right side of Fig. 9(a), where a carbon dioxide laser-beam weld was made over a gas-tungsten arc weld in an alloy of 23Cr-12Ni-balance Fe with a $\mathrm{Cr_{eq}/Ni_{eq}}$ ratio of 1.9 (Ref 26). These welds solidify as single-phase ferrite, but subsequently transform to approximately 100% austenite during cooling via a massive transformation. The gas-tungsten arc weld of this composition, shown to the left in Fig. 9(a), exhibits a typical two-phase microstructure with a FN of approximately 21.

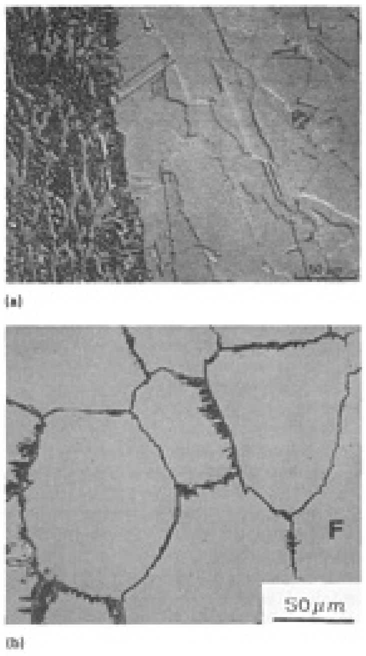

FIG. 9 MICROSTRUCTURES OF LASER-BEAM-WELDED AUSTENITIC STAINLESS STEELS. (A) GAS-TUNGSTEN ARC WELD SHOWN ON LEFT, WITH CO_2 LASER-BEAM WELD SHOWN ON RIGHT, IN ALLOY OF CR_{EQ}/NI_{EQ} = 1.8. LASER-BEAM WELD ON RIGHT IS SINGLE-PHASE AUSTENITE FORMED AS A PRODUCT OF MASSIVE TRANSFORMATION. (B) FERRITIC STRUCTURE FORMED IN LASER-BEAM WELD OF CR_{EQ}/NI_{EQ} = 2.0. SOURCE: REF 26, 27

Characteristics of the massively transformed austenitic structure are fairly small-faceted-appearing grain boundaries and little or no evidence of the cellular solidification that would result from microsegregation. This apparent lack of microsegregation may be due to the higher solute diffusivities in ferrite, as compared with austenite, which results in homogenization occurring during solidification and cooling and/or by reduced alloy partitioning occurring at the rapid solidification velocities (Ref 26, 27).

At high Cr_{eq}/Ni_{eq} ratios, high-energy-density welds can solidify completely as ferrite. However, because of suppression of the solid-state transformation of ferrite to austenite during the rapid cooling, a large fraction of the ferrite is retained at room temperature. This behavior has been reported in laser-beam welds of AISI 312. Microsegregation is also minimal in these welds, and the grain boundary region sometimes exhibits Widmanstätten austenites rather than the faceted grain structure of the massively transformed austenite. An example of this microstructure is shown in Fig. 9(b). In some applications, such as the containment of high-pressure hydrogen, for which the austenitics are highly suited, a weld that completely comprises ferrite would be unacceptable. With the massively transformed ferrite and the retention of ferrite at high Cr_{eq}/Ni_{eq} ratios, the composition range in which two-phase $\delta + \gamma$ microstructures exists in the high-energy-density welds is greatly reduced.

The shift from primary ferrite to primary austenite solidification at high solidification rates often leads to problems with weld cracking, as shown in the electron-beam weld in Fig. 8. Alloy selection based on solidification behavior predicted by the WRC-1992 diagram or that of Moisio and coworkers (Ref 17) may not be valid.

For example, a pulsed Nd:YAG laser-beam weld in a 316L alloy with a Cr_{eq}/Ni_{eq} (Suutala and Moisio equivalents) of approximately 1.6 is shown in Fig. 10. Solidification of this weld has occurred as primary austenite, with resultant solidification cracking. Reference to the Suutala and WRC-1992 diagrams indicate that solidification would occur as primary ferrite with conventional gas-tungsten arc welds and that the weld should contain approximately 6 FN.

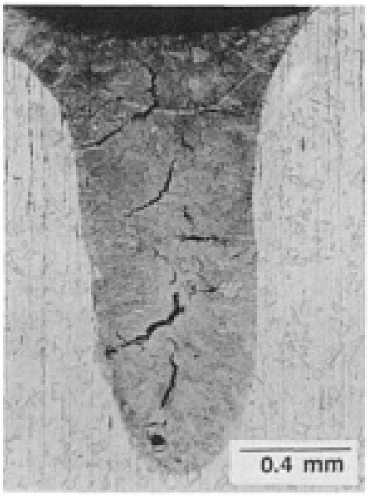

FIG. 10 PULSED ND:YAG LASER-BEAM WELD EXHIBITING SEVERE SOLIDIFICATION CRACKING AS A CONSEQUENCE OF PRIMARY AUSTENITE SOLIDIFICATION, CR_{EQ}/NI_{EQ} = 1.6. SOURCE: REF 28

A modified Suutala diagram developed by Pacary *et al.* (Ref 28) (Fig. 11) shows the shift in solidification mode and associated cracking behavior of pulsed laser-beam welds. The critical cracking/no cracking demarcation shifts from the Cr_{eq}/Ni_{eq} ratio of approximately 1.48 (Suutala and Moisio equivalents) for conventional arc welding to approximately 1.7 for the pulsed laser-beam welds.

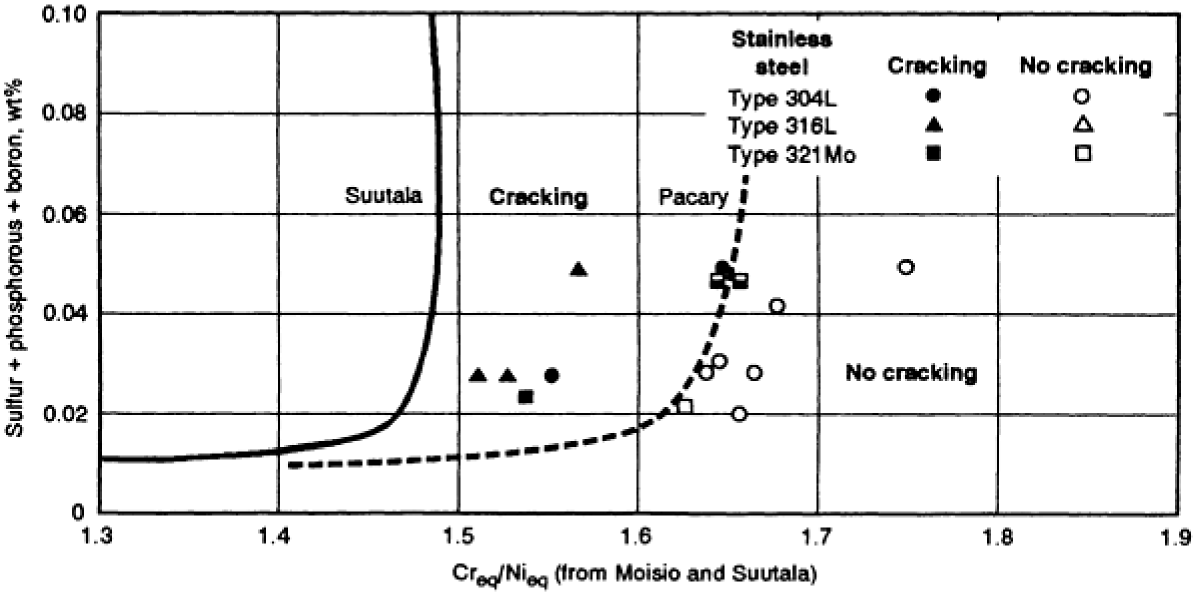

FIG. 11 DIAGRAM FOR PREDICTING WELD SOLIDIFICATION CRACKING SUSCEPTIBILITY OF PULSED LASER WELDS IN AUSTENITIC STAINLESS STEELS. NOTE WRC EQUIVALENTS ARE USED. SOURCE: REF 28

Katayama and Matsunawa (Ref 12) established the change in solidification using Schaeffler equivalents to also occur at the value of 1.7. These values are dependent on the welding process (for example, electron-beam and pulsed-Nd:YAG and carbon dioxide laser-beam welds), as well as on weld parameters (Ref 27). Because the pulsed-YAG technique produces the most rapid solidification interface velocities and cooling rates, the transition for this process should occur at the highest ratio. Thus, the selection of alloy compositions above the Cr_{eq}/Ni_{eq} ratio of approximately 1.7 (Suutala and Moisio equivalents) ensures solidification as primary ferrite and significantly reduces the risk of cracking, although higher accompanying ferrite content in the base material must also be expected.

References cited in this section

12. S. KATAYAMA AND A. MATSUNAWA, *PROC. INT. CONF. ON ADVANCED LASER ELECTRO-OPTICS (ICALEO)* (SAN FRANCISCO), 1985
16. J.A. BROOKS AND A.W. THOMPSON, *LNT. MATER. REV.*, VOL 36 (NO. 1), 1991, P 16-44
17. N. SUUTALA AND T. MOISIO, THE USE OF CHROMIUM AND NICKEL EQUIVALENTS IN CONSIDERING SOLIDIFICATION PHENOMENA IN AUSTENITIC STAINLESS STEELS, *SOLIDIFICATION TECHNOLOGY IN THE FOUNDRY AND CASTHOUSE,* THE METALS SOCIETY PREPRINT, LONDON, 1980
24. S.A. DAVID, J.M. VITEK, AND T.L. HEBBLE, *WELD. J.,* VOL 66 (NO. 10), 1987, P 289S-300S
25. N. SUUTALA, EFFECT OF SOLIDIFICATION CONDITIONS ON THE SOLIDIFICATION MODE IN AUSTENITIC STAINLESS STEELS, *TECHNOLOGY IN THE FOUNDRY AND CASTHOUSE,* THE METALS SOCIETY PREPRINT, LONDON, 1980
26. J.A. BROOKS, M.I. BASKES, AND F.A. GREULICH, *METALL. TRANS. A,* VOL 22A, 1991, P 915-925
27. J.W. ELMER, S.M. ALLEN, AND T.W. EAGAR, *METALL. TRANS. A,* VOL 20A, 1989, P 2117-2131
28. G. PACARY, M. MOLINE, AND J.C. LIPPOLD, EWI RESEARCH BRIEF NO. B9008, EDISON WELDING INSTITUTE, 1990

Selection of Wrought Austenitic Stainless Steels

John A. Brooks, Sandia National Laboratories; John C. Lippold, Edison Welding Institute

Weld Defect Formation

Solidification cracking, which results from the formation of low-melting liquid films along grain boundaries during the last stages of solidification, is probably of greater concern to the welding engineer than is any other metallurgical defect in austenitic stainless steel welds. As discussed above, welds that solidify as primary austenite may be very susceptible to solidification cracking. This susceptibility is increased by elements that promote the formation of liquid phases. These elements include phosphorus and sulfur, which are often present as impurities, and silicon, niobium, titanium, and boron, which are sometimes present as minor alloying additions. In these welds, high manganese levels may be beneficial in reducing cracking, for example, by tying up sulfur and silicon. Several percent of eutectic ferrite may be somewhat beneficial in counteracting the detrimental effect of phosphorus and sulfur. However, because the amount of eutectic ferrite is alloy dependent and seldom exceeds several percent (see Fig. 1), it is not a practical solution for improving cracking resistance.

Cracking susceptibility is drastically reduced in welds that solidify as primary ferrite with the peritectic/eutectic solidification of austenite (FA-type), even at high levels of phosphorus and sulfur, as shown in Fig. 7. Welding electrodes such as 308L are specifically designed to produce a desired range of ferrite, and thus promote this solidification mode. In applications where no such filler metal is used and cracking is a concern, it may be necessary to limit composition ranges or to select heats of material that solidify in this mode. However, at the higher Cr_{eq}/Ni_{eq} ratios, where welds solidify completely as ferrite, and when the transformation of ferrite to austenite occurs at lower temperatures, cracking susceptibility again increases, but not to the extent of welds solidifying as austenite. The solidification-cracking behavior can thus be directly related to the structures shown in Fig. 3 and 5. The welds that solidify as ferrite and exhibit skeletal and lathy morphologies are very resistant to cracking, whereas cracking increases in welds that exhibit microstructures of Widmanstätten austenite.

Many rationales have been proposed to explain the cracking behavior of the austenitic stainless steel welds. The two that seem to be most consistent in explaining this behavior are related to the nature of the weld microstructure during solidification and the early stages of cooling.

During single-phase solidification, the fairly planar ferrite/ferrite or austenite/austenite grain boundaries that form are easily wet by low-melting liquids, often involving eutectics with phosphides and sulfides. If sufficient stresses and strains are developed before these liquids are solidified, then solidification cracks can easily form and propagate. However, when solidification occurs as primary ferrite with the secondary solidification of austenite (FA-type), or if austenite is nucleated at the grain boundaries near the solidus temperature of welds solidifying completely as ferrite (F-type), then very irregular grain-boundary structures are formed. These ferrite/austenite boundaries that become potential crack paths are of lower energy than the grain boundaries of either of the single-phase structures. The resulting tortuous crack paths and/or the low-energy boundary configurations not readily wetted by the low-melting liquids are very resistant to the nucleation and propagation of cracks (Ref 29).

Thus, it can be seen that understanding and being able to identify solidification behavior are important in preventing solidification-cracking problems. Because increased solidification velocities and cooling rates shift the transition between primary austenite and primary ferrite to higher Cr_{eq}/Ni_{eq} ratios, this effect on weld cracking must be considered, as discussed in the section "Solidification Behavior and Microstructure of High-Energy-Density Welds." .

Heat-affected zone liquation cracking is occasionally observed in austenitic stainless steels, particularly in the stabilized grades (containing titanium and/or niobium). This cracking occurs as a result of the constitutional liquation of carbides, such as niobium carbide, in AISI type 347 alloy, resulting in grain-boundary liquation. HAZ cracking can also occur in other alloy types as a result of impurity segregation along HAZ grain boundaries. This form of cracking is always intergranular and located immediately adjacent to the weld fusion boundary. Occasionally, cracking is continuous across the fusion boundary into the weld metal.

Analogous to weld solidification cracking, alloys with a ferrite potential (predicted FN) above approximately 2 to 3 FN are generally resistant to HAZ liquation cracking (Ref 30). This results from the formation of ferrite along HAZ grain boundaries. HAZ cracking that is due to grain-boundary liquation also can be limited by minimizing heat input. Lower weld heat inputs result in steeper HAZ thermal gradients, thereby limiting the spatial extent over which cracking is possible.

Weld-metal liquation cracking is defined as HAZ cracking that occurs in weld metal. By definition, this form of cracking is associated with multipass welds. The cracking that occurs is sometimes referred to as "microfissuring," or underbead cracking, because the cracks are typically small and located adjacent to the interpass fusion boundary in the previously deposited weld metal. These defects are almost always associated with fully austenitic weld metal and are located along solidification grain boundaries or migrated grain boundaries.

Analogous to weld solidification cracking, weld-metal liquation cracking can generally be avoided by reducing the impurity levels in fully austenitic microstructures or by selecting filler metals that produce weld deposits containing more than approximately 3 FN.

Copper contamination cracking adjacent to austenitic stainless steel welds can also occur as a result of contact with copper or copper-bearing alloys, such as brass. This type of cracking is a form of liquid-metal embrittlement, whereby molten copper embrittles austenitic grain boundaries. Because copper melts at 1083 °C (1981 °F), this type of cracking caused by pure copper is only observed in regions of the weld (typically, the HAZ) that are heated near or above this temperature. Cracking is always intergranular and, normally, traces of copper or copper-rich material can be observed along the grain boundary.

This contamination can often be traced to copper weld fixturing, contact tips, or fabrication tools. The use of alternate fixturing, often incorporating chromium plating, and the prevention of copper abrasion normally eliminates the problem.

Ductility dip cracking occurs in the solid state and is typically associated with the HAZ or weld metal in highly constrained welds or thick sections. Such cracking occurs as a result of strain-induced grain-boundary precipitation, which limits mid-temperature ductility, but is much more common in precipitation-strengthened nickel-base alloys. However, this type of cracking has been observed in type 347 alloy, resulting from strain-induced precipitation of niobium carbides.

Weld porosity can result either from alloying additions (especially in the high-nitrogen alloys) or by contamination. In the Nitronic alloys, for example, nitrogen levels near the upper limit of 0.4% can result in porosity in gas-tungsten arc weld deposits, whereas lower levels (above approximately 0.25%) can result in excessive porosity in electron-beam welds (Ref 31). Porosity can also occur as a result of surface contamination, such as by oil and greases or undetected leaks in water-cooled welding torches or weld fixtures. Thus, surfaces should be thoroughly cleaned prior to welding. If wire brushing is used, then the brushes should be made of stainless steel and dedicated to stainless usage, rather than made of steel, which can result in iron deposition and in-service corrosion.

References cited in this section

29. J.A. BROOKS, A.W. THOMPSON, AND J.C. WILLIAMS, *WELD. J.,* VOL 63, 1984, P 71-S TO 83-S
30. J.C. LIPPOLD, I. VAROL, AND W.A. BAESLACK III, *WELD. J.,* VOL 71, 1992, P 1S-14S
31. J.A. BROOKS, *WELD. J.,* VOL 54, 1975, P 189S-195S

Selection of Wrought Austenitic Stainless Steels

John A. Brooks, Sandia National Laboratories; John C. Lippold, Edison Welding Institute

Sigma-Phase Embrittlement

Sigma is a hard, brittle phase that can form in stainless steels when they are held in the temperature range from approximately 600 to 800 °C (1110 to 1470 °F). The sigma-phase compositional range shown in the vertical sections of Fig. 4 reveals that sigma can form over a wide range of compositions. The tendency for sigma-phase formation increases with increasing amounts of chromium, molybdenum, and silicon, and is reduced somewhat by nitrogen, nickel, and carbon. As a consequence of alloy partitioning during both solidification and the solid-state transformation of ferrite to austenite, ferrite is enriched in the alloying elements, which promotes the sigma phase, and depleted in the elements that retard its formation. Thus, ferrite in austenitic stainless welds is especially susceptible to sigma-phase formation. The sigma phase has been observed to nucleate preferentially at the ferrite-austenite interface in weld metal.

The transformation kinetics of ferrite to sigma has been studied by Vitek and David (Ref 32) in type 308 gas-tungsten arc welds (Fig. 12). They found that sigma started to form after approximately 15 h at 725 °C (1335 °F), and all the ferrite had transformed to sigma and $M_{23}C_6$ after approximately 3000 h.

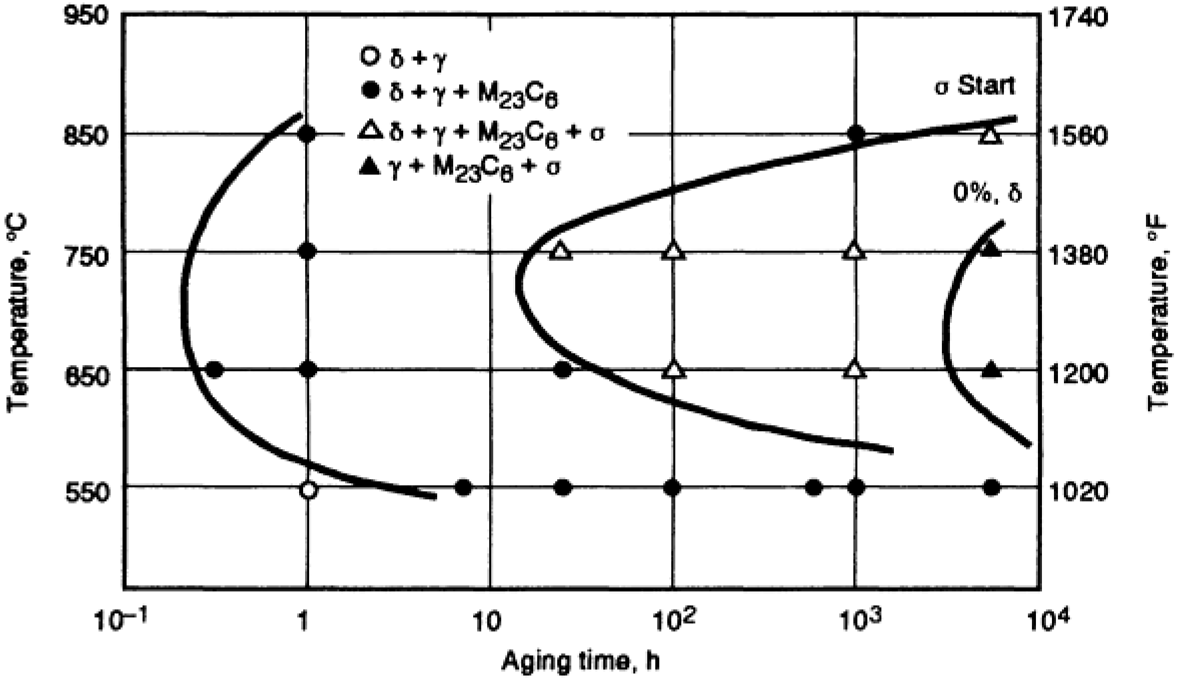

FIG. 12 TIME-TEMPERATURE TRANSFORMATION DIAGRAM FOR WELDED AND AGED TYPE 308L ALLOY GAS-

TUNGSTEN ARC WELDS. SOURCE: REF 32

The notorious effect of sigma phase is a decrease in ductility and toughness. When materials are to be used in elevated-temperature service, the effects of sigma formation must be considered. The sigma phase can be minimized by using filler materials that produce fully austenitic structures. However, even in these compositions, sigma formation can still occur, initiating in chromium-enriched interdendritic boundaries.

In cases where solidification cracking is a concern, some ferrite in the weld deposit is required. Weld-metal ferrite content should be controlled in the range from 3 to 8 FN to prevent the formation of a continuous network of ferrite. Because the chromium content of sigma is nearly double that of the weld ferrite, the potential maximum level of the sigma phase is always less than that of the original weld-metal ferrite.

Reference cited in this section

32. J.M. VITEK AND S.A. DAVID, *WELD. J.*, VOL 63, 1984, P 246S-253S

Selection of Wrought Austenitic Stainless Steels

John A. Brooks, Sandia National Laboratories; John C. Lippold, Edison Welding Institute

Corrosion Behavior

Austenitic stainless steel weldments are often subject to corrosive attack. The nature of this attack is a function of weld thermal history, service temperature and environment, and stress level (both residual and applied). Four general types of corrosive attack have been reported:

- INTERGRANULAR ATTACK
- STRESS-CORROSION CRACKING
- PITTING AND CREVICE CORROSION
- MICROBIOLOGICALLY INFLUENCED CORROSION

Each phenomena and its effect on service integrity is discussed below.

Carbide Precipitation and Intergranular Corrosion. Austenitic stainless steel weldments have often been subject to intergranular attack (IGA). This behavior is particularly prevalent in chloride-beating aqueous environments, and it is usually typified by accelerated attack in the HAZ. The basis for this attack is related to a phenomenon known as sensitization, whereby exposure or slow cooling in the temperature range from 400 to 850 °C (750 to 1550 °F) promotes the formation of chromium-rich carbides ($M_{23}C_6$) along the austenite grain boundaries. This temperature range is commonly described as the "sensitization" temperature range, because exposure within it makes austenitic stainless steels particularly sensitive to IGA. An example of intergranular corrosion in the HAZ of a weld in type 316L NaOH reactor vessel is shown in Fig. 13.

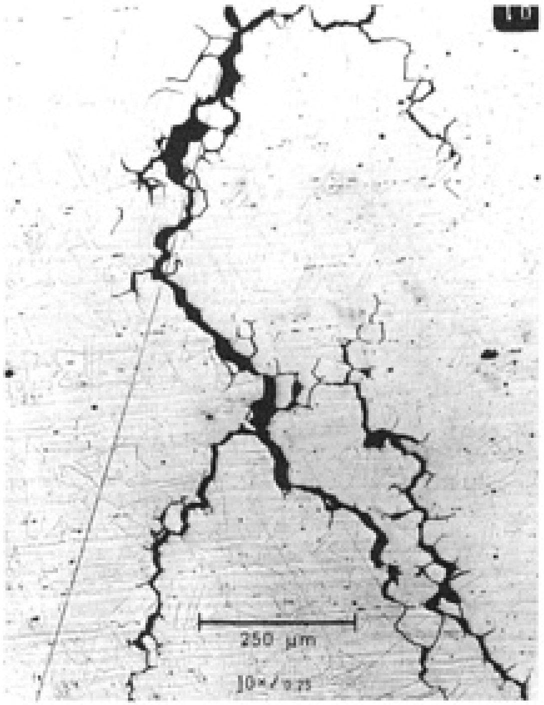

FIG. 13 CAUSTIC SCC IN THE HAZ OF A TYPE 316L STAINLESS STEEL NAOH REACTOR VESSEL. CRACKS ARE BRANCHING AND INTERGRANULAR.

Mechanistically, localized attack occurs immediately adjacent to the grain boundary, because of precipitation of the chromium-rich carbides in the boundary. Carbide formation depletes the adjacent matrix in chromium to a level generally estimated to be less than 12 wt% Cr. The severity of attack depends on the corrosive media, the time and temperature of exposure, and the composition and metallurgical condition of the steel.

Unfortunately, sensitization can be a by-product of the fusion-welding process. During welding, the weld HAZ experiences peak temperatures, ranging from the melting point to the ambient base-metal temperature, depending on the distance from the weld fusion boundary. Because heating and cooling in the weld HAZ is relatively rapid, regions of the HAZ that are heated to peak temperatures below about 650 °C (1200 °F) are immune to sensitization.

In addition, the HAZ region immediately adjacent to the fusion boundary is also resistant to sensitization, because carbon solubility is high at the peak temperatures achieved and cooling through the sensitization range is generally rapid enough to suppress carbide precipitation. This leaves a narrow region in the HAZ, somewhat removed from the fusion boundary, that will be heated into the sensitization temperature range. If the time within this temperature range is sufficient, then intergranular carbides will nucleate along austenite grain boundaries.

For a given thermal history, the level of precipitation and the associated degree of sensitization are approximately proportional to the carbon content of the austenitic stainless steel. Conventional austenitic stainless steels can retain a maximum of about 0.02 wt% C in solid solution over their service temperature range (up to 600 °C, or 1110 °F). Thus, as carbon content increases above this level, the driving force for carbide precipitation increases.

This can be illustrated by the time-temperature-precipitation (TTP) curves shown in Fig. 14 for an 18Cr-9Ni alloy. Because of this, low-carbon grades of stainless steel (the so-called L-grades) are preferred for corrosive service where weldments must be used in the as-welded condition. The use of "stabilized" alloys, such as types 321 and 347, which contain additions of titanium and niobium, respectively, are also effective in reducing susceptibility to IGA. Both titanium and niobium form stable carbides in the base metal that resist dissolution in the weld HAZ and/or at service temperatures in the sensitization range.

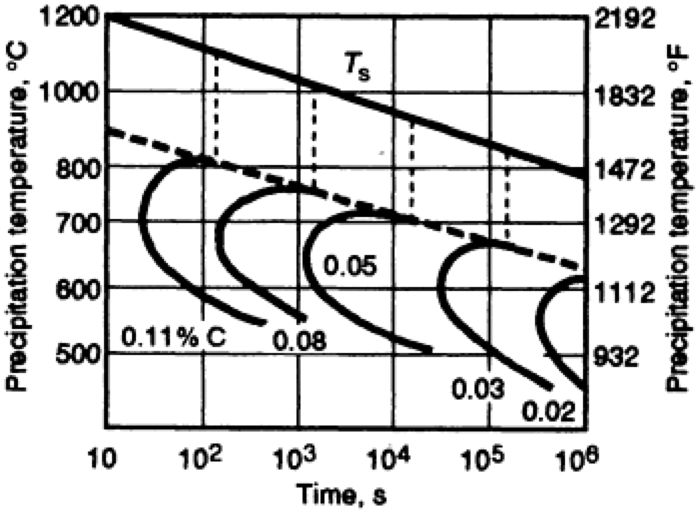

FIG. 14 TIME-TEMPERATURE PRECIPITATION CURVES FOR AN 18CR-9NI ALLOY WITH VARIOUS CARBON CONTENTS. SOURCE: REF 33

A phenomenon known as "knifeline" attack has been observed in the stabilized grades. However, its occurrence is associated with localized HAZ attack adjacent to the weld fusion boundary (Ref 34). Where temperatures exceed 1300 °C (2370 °F), niobium carbide and titanium carbide precipitates may dissolve and, upon reheating into the sensitization temperature range, chromium carbide ($M_{23}C_6$-type) precipitation occurs preferentially to the reformation of either niobium carbide or titanium carbide. Subsequent exposure to a corrosive medium may result in IGA that is extremely localized. In severe cases, it may often appear as if the weld has been cut out by a knife.

Weld metals containing even small amounts of delta ferrite are not generally susceptible to such severe sensitization as wrought material of similar composition. In ferrite-containing weld metal, carbide precipitation occurs more uniformly throughout the structure at the austenite-ferrite interphase boundaries. Because the ferrite phase has a relatively high chromium content (on the order of 25 wt% in a 18Cr-8Ni alloy), chromium depletion caused by carbide formation is not so significant. Consequently, weld metal is not generally susceptible to intergranular corrosion. However, fully austenitic weld metals may be susceptible to IGA when exposed to severely corrosive environments.

In general, it is possible to either minimize or eliminate IGA in austenitic stainless steel weldments by using one or more of these methods:

- USE A LOW-CARBON-CONTENT ALLOY, SUCH AS "L-GRADE" ALLOYS WITH CARBON CONTENTS ON THE ORDER OF 0.03 WT%.
- USE ALLOYS THAT ARE STABILIZED BY ADDITIONS OF NIOBIUM AND TITANIUM (SUCH AS TYPES 347 AND 321).
- ANNEAL THE MATERIAL PRIOR TO WELDING TO REMOVE ANY PRIOR COLD WORK (COLD WORK ACCELERATES CARBIDE PRECIPITATION).
- USE LOW WELD HEAT INPUTS AND LOW INTERPASS TEMPERATURES TO INCREASE WELD COOLING RATES, THEREBY MINIMIZING THE TIME IN THE SENSITIZATION TEMPERATURE RANGE.
- IN PIPE WELDING, WATER COOL THE INSIDE DIAMETER AFTER THE ROOT PASS, WHICH WILL ELIMINATE SENSITIZATION OF THE INSIDE DIAMETER CAUSED BY SUBSEQUENT PASSES.
- SOLUTION HEAT TREAT AFTER WELDING. HEATING THE STRUCTURE INTO THE TEMPERATURE RANGE FROM 900 TO 1100 °C (1650 TO 2010 °F) DISSOLVES ANY CARBIDES THAT MAY HAVE FORMED ALONG GRAIN BOUNDARIES IN THE HAZ. FOR LARGE STRUCTURES, THIS APPROACH IS USUALLY IMPRACTICAL.

The IGA of the weld HAZ has also been observed at service temperatures well below the classic sensitization temperature range. This behavior, often called low-temperature sensitization (LTS), typically occurs after years of service exposure at temperatures below 400 °C (750 °F), even in low-carbon grades. Mechanistically, it is proposed (Ref 35) that carbide embryos form along grain boundaries in the weld HAZ and then grow to form chromium-rich carbides in service. Corrosive attack occurs via the chromium-depletion mechanism described previously. In order to combat LTS, either stabilized alloys or alloys containing higher nitrogen contents (such as type 316LN) are substituted for conventional or L-grade materials.

The stress-corrosion cracking (SCC) of austenitic stainless steels may occur when an alloy is subjected simultaneously to a tensile stress and a specific corrosive medium. The important variables affecting SCC are temperature, environment, material composition, stress level, and microstructure. Crack propagation may be either transgranular or intergranular, depending on the interaction of these variables. Intergranular stress-corrosion cracking (IGSCC) can occur even though the alloy is insensitive to intergranular corrosion. The presence of residual tensile stresses in the HAZ may accelerate corrosion attack and cracking, particularly along sensitized grain boundaries.

Transgranular SCC is also observed in austenitic stainless steels. This form of cracking is usually indigenous to chloride environments (seawater), but may also occur in caustic media. The ions of the halogen family (fluorine, chlorine, bromine, and iodine) are largely responsible for promoting transgranular SCC. Of the halides, the chloride ion causes the greatest number of failures. Figure 15 shows an example of transgranular SCC in type 316 stainless steel. Note that cracking has initiated at the toe of the weld, probably because of the high residual stress level at that location.

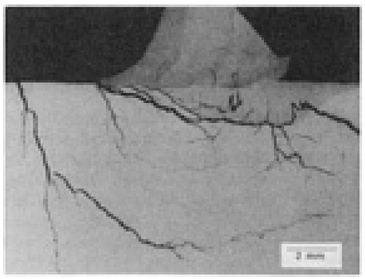

FIG. 15 TRANSGRANULAR SCC THAT HAS INITIATED NEAR THE TOE OF A WELD IN TYPE 316L ALLOY

Methods for controlling SCC include:

- A STRESS-RELIEVING OR ANNEALING HEAT TREATMENT TO REDUCE WELD RESIDUAL STRESSES
- SUBSTITUTION OF MORE-RESISTANT HIGH-NICKEL ALLOYS, HIGH-CHROMIUM FERRITIC STAINLESS STEELS, OR DUPLEX STAINLESS STEELS, THE LATTER OF WHICH HAVE BEEN DESIGNED SPECIFICALLY TO MINIMIZE SCC
- REDUCTION OF CHLORIDE AND OXYGEN CONTENTS FROM THE ENVIRONMENT, BECAUSE THESE TWO ELEMENTS, IN COMBINATION, ARE RESPONSIBLE FOR MOST SCC FAILURES IN AUSTENITIC STAINLESS STEELS

Pitting, Crevice Corrosion, and Microbiologically Influenced Corrosion. Localized attack in the weld metal or HAZ may occur in the form of pitting or crevice corrosion, particularly in aqueous, chlorine-bearing environments.

Microbiologically influenced corrosion (MIC) often occurs preferentially in the weld metal of austenitic stainless steels (Ref 36, 37). This curious form of attack is restricted to aqueous environments containing microbes that attach to the surface of the weld. In pipe welds, pitting from MIC often initiates at the weld fusion line and proceeds into the weld metal. Sensitized HAZs also are common regions of attack, although the most preferred site of attack seems to be the weld metal. Attack is a strong function of pitting resistance of the material and is thus more common in type 304 alloy, than in type 316.

The surface condition seems to have a large effect on MIC. Machining and grinding conducted on the inside of piping is thought to increase susceptibility to MIC by introducing rough surfaces for easy attachment by microbes and residual stresses, as well as localized metallurgical changes and corrosion pitting.

References cited in this section

33. V. CIHAL, *PROT. MET.* (RUSSIA), VOL 4 (NO. 6), 1968, P 653
34. C.J. NOVAK, SENSITIZATION, *HANDBOOK OF STAINLESS STEELS,* D. PECKHER AND I.M. BERNSTEIN, ED., MCGRAW-HILL, 1977, P 4-35 TO 4-53
35. M.C. JUHAS AND B.E. WILDE, *CORROSION,* VOL 46, 1990, P 812-822
36. D.H. POPE, "MICROBIOLOGICAL ASPECTS OF C INFLUENCED CORROSION," ER-6345, ELECTRIC POWER RESEARCH INSTITUTE, 1988
37. S.W. BORENSTEIN, MICROBIOLOGICALLY INDUCED CORROSION OF AUSTENITIC STAINLESS STEEL WELDMENTS, *MATER. PERFORM.,* VOL 1, 1991, P 52-54

Selection of Wrought Austenitic Stainless Steels

John A. Brooks, Sandia National Laboratories; John C. Lippold, Edison Welding Institute

Superaustenitic Stainless Steels

The superaustenitic stainless steels represent a special class of austenitic stainless steels that contain high levels of nickel, molybdenum, and, in some cases, nitrogen. These alloys are designed for severely corrosive environments. They provide improved resistance to SCC, pitting, and crevice corrosion, relative to the standard 300 series of austenitic alloys. Higher nickel contents improve chloride SCC resistance, whereas molybdenum and nitrogen provide improved pitting and crevice corrosion resistance. In general, these alloys are compositionally balanced to ensure fully austenitic microstructures. The compositions of selected commercially available superaustenitic alloys are shown in Table 2.

The weldability of most of these alloys is similar to that of fully austenitic grades of conventional stainless steel. Because these welds solidify as austenite and exhibit fully austenitic (no ferrite) microstructures, solidification cracking may be a problem in highly restrained joints. Typically, these alloys contain extremely low impurity levels, thus minimizing the likelihood of cracking.

The segregation of molybdenum in 6% Mo alloys during weld solidification has been reported to reduce corrosion resistance (Ref 38), particularly in severe pitting environments. The cell or dendrite cores in these alloys may exhibit molybdenum contents as low as 4.2 wt%, resulting in localized pitting attack at these sites. As a result, autogenous welding or the use of matching filler materials is normally avoided when severe service environments are anticipated. The use of molybdenum-containing, nickel-base filler materials, such as alloy 625 (AWS ERNi-CrMo-3), alloy C-276 (AWS ERNiCrMo-4), or alloy C-22 (AWS ERNiCrMo-10), is often recommended in order to avoid pitting. Precautions must be taken to prevent the formation of an unmixed zone adjacent to these overalloyed weld metals, because this region may have the same susceptibility to corrosive attack as autogenous weld metal does. Much more detail is provided in Ref 39.

References cited in this section

38. T.G. GOOCH, *PROC. CORROSION '92* (NASHVILLE), PAPER 296, NACE, 1992
39. CORROSION OF AUSTENITIC STAINLESS WELDMENTS, *CORROSION,* VOL 13, *ASM HANDBOOK,* ASM INTERNATIONAL, 1987, P 347-361

Selection of Wrought Austenitic Stainless Steels

John A. Brooks, Sandia National Laboratories; John C. Lippold, Edison Welding Institute

Mechanical Properties

Austenitic stainless steels exhibit moderate strength and excellent toughness and ductility. Because of these properties, austenitic filler materials are often used in dissimilar combinations with carbon steels and ferritic and martensitic stainless steels. The weld properties of austenitic stainless steels are essentially similar to those of the annealed base material. Typical base-metal properties for austenitic stainless steels are listed in Table 4. Weld-metal properties are described in the article "Welding of Stainless Steels" in this Volume.

TABLE 4 MINIMUM ROOM-TEMPERATURE MECHANICAL PROPERTIES OF SELECTED AUSTENITIC STAINLESS STEELS

DESIGNATION	PRODUCT FORM	CONDITION	TENSILE STRENGTH		0.2% YIELD STRENGTH		ELONGATION, %	REDUCTION IN AREA, %	HARDNESS, HRB	ASTM SPECIFICATION
			MPA	KSI	MPA	KSI				
TYPE 301 (UNS S30100)	BAR	ANNEALED	620	90	205	30	40	. . .	95 MAX	A 666
TYPE 302 (UNS S30200)	BAR, FORGINGS	HOT FINISHED AND ANNEALED	515	75	205	30	40	50	. . .	A 276, A 473
TYPES 303 (UNS S30300) AND 303SE (UNS S30323)	FORGINGS	ANNEALED	515	75	205	30	40	50	. . .	A 473
TYPE 304 (UNS S30400)	BAR, FORGINGS(A)	HOT FINISHED AND ANNEALED	515	75	205	30	40	50	. . .	A 276, A 473
TYPE 304L (UNS S30403)	BAR	HOT FINISHED AND ANNEALED	480	70	170	25	40	50	. . .	A 276
TYPE 305 (UNS S30500)	BAR, FORGINGS	HOT FINISHED AND ANNEALED	515	75	205	30	40	50	. . .	A 276, A 473
TYPE 308 (UNS S30800)	BAR, FORGINGS	HOT FINISHED AND ANNEALED	515	75	205	30	40	50	. . .	A 276, A 473
TYPE 309 (UNS S30900), 309S (UNS S30908), 310 (UNS S31000) AND 310S (UNS S31008)	BAR, FORGINGS	HOT FINISHED AND ANNEALED	515	75	205	30	40	50	. . .	A 276, A 473
310CB (UNS S31040)	BAR, SHAPES	HOT FINISHED	515	75	205	30	40	50	. . .	A 276

		AND ANNEALED								
TYPE 316 (UNS S31600)	BAR, FORGINGS[(A)]	HOT FINISHED AND ANNEALED	515	75	205	30	40	50	. . .	A 276, A 473
TYPE 316L (UNS S31603)	BAR	HOT FINISHED AND ANNEALED	480	70	170	25	40	50	. . .	A 276
TYPE 317 (UNS S31700)	BAR, FORGINGS	HOT FINISHED AND ANNEALED	515	75	205	30	40	50	. . .	A 276, A 473
TYPE 317L (UNS S31703)	PLATE, SHEET, STRIP	ANNEALED	515	75	205	30	40	. . .	95 MAX	A 167
TYPES 321 (UNS S32100) AND 321H (UNS 32109)	BAR, FORGINGS	HOT FINISHED AND ANNEALED	515	75	205	30	40	50	. . .	A 276, A 473
TYPES 347 (UNS S34700) AND 348 (UNS S34800)	BAR, FORGINGS	HOT FINISHED AND ANNEALED	515	75	205	30	40	50	. . .	A 276, A 473
20CB-3 (UNS N08020), 20MO-4 (UNS N08024), AND 20MO-6 (UNS N08026)	BAR, WIRE	ANNEALED	550	80	240	35	30	50	. . .	B 473
TYPE 330 (UNS N08330)	BAR	ANNEALED	485	70	210	30	30	. . .	. . .	B 511
AL-6X (UNS N08366)	BAR, WIRE	ANNEALED	515	75	210	30	30	. . .	. . .	B 691
TYPE 332 (UNS N08800)	PIPE, TUBE	ANNEALED	515	75	210	30	30	. . .	. . .	B 163, B 407, B 514, B 515
TYPE 904L (UNS N08904)	BAR	ANNEALED	490	71	220	31	35	. . .	. . .	B 649

(A) FOR FORGED SECTIONS 127 MM (5 IN.) AND OVER, THE TENSILE STRENGTH SHALL BE 485 MPA (70 KSI).

Weld geometry and alloy composition also affect weld properties. Most weld tensile properties are determined using specimens transverse to the weld direction. Because failure often occurs outside the weld metal, only minimum tensile strength values can be determined and no weld-metal ductility data are generated. Impact toughness data are much more accurate, because a notch can be placed in the weld metal. HAZ mechanical property data are essentially nonexistent. Because the HAZ undergoes high temperatures during welding where carbide dissolution and grain growth can occur, it would be expected that this region may exhibit slightly lower strength than the annealed base metal.

In welds containing ferrite, the ferrite provides second-phase strengthening with a concomitant decrease in ductility. Thus, for an autogenous weld, the actual properties may vary somewhat from heat to heat in the same manner as the ferrite content. Typically, the tensile strength of welds is fractionally higher than the annealed wrought properties of similar compositions. However, the elongation and impact properties are reduced on the order of 30%. In multipass welds of thicker sections, warm working that results from the large strains encountered during cooling of subsequent passes can also result in significant increases in yield strengths and variations in strength from the top to the bottom of the weld deposit.

Nitrogen is a strong solid-solution strengthener and, thus, its increase or decrease in the welding process may also affect weld properties. In wrought materials, each addition of approximately 0.01% N results in an approximate increase of 5.5 to 7 MPa (0.8 to 1 ksi) in yield strength. Thus, from a conservative standpoint, weld strength is often considered to be the same as that of the annealed base material.

Weld properties may be significantly influenced by service temperature. At elevated temperatures, carbide, nitride, and intermetallic formation may reduce mechanical and corrosion properties. At cryogenic temperatures, the strain-induced martensite transformation from austenite can result in reduced toughness and ductility, relative to those properties in the homogeneous single-phase base material. In cases where filler metal is added, ASTM specifications provide minimum requirements. The reader is referred to the article "Welding for Cryogenic Service" in this Volume to obtain information on mechanical properties at cryogenic temperatures.

Weld penetration characteristics of austenitic stainless steels are generally poor, relative to ferritic and martensitic grades, because of the low thermal conductivity of the austenite. In addition, large heat-to-heat variations in weld penetration can occur when welding austenitic stainless, especially in autogenous welds or the root pass of multiple-pass gas-tungsten arc welds. These penetration variations have been attributed primarily to minor elements affecting the surface-tension-driven fluid flow within the weld pool (known as the Marongoni effect) (Ref 40).

The elements that influence penetration tend to be surface active and include sulfur, selenium, and oxygen. When present in sufficient quantity, these elements tend to increase weld penetration. However, other elements can have synergistic effects. Aluminum combined with oxygen can mitigate the effect of oxygen. Thus, high levels of aluminum can decrease penetration. For example, heats of type 304L alloy with low sulfur contents (0.001 to 0.003 wt%) may have a weld depth-to-width ratio of only 0.25, whereas the same material, but with a sulfur content of 0.01 wt%, welded under identical conditions, will exhibit a depth-to-width ratio higher than 0.5. Increased surface-oxygen content that results from wire brushing may also result in increased depth-to-width ratios in materials that otherwise have low penetration characteristics. However, this method is not often practiced, because of problems with reproducibility.

In autogenous welds, penetration control is best achieved by the proper selection of base materials. The addition of hydrogen and helium to argon shielding gas may help to compensate for poor penetration. For example, $98Ar-2H_2$ and $Ar-5H_2$ are commonly used in the manufacture of welded tubing to increase penetration at high travel speeds. Trimixes of argon, helium, and H_2 may also be used to increase weld penetration.

Reference cited in this section

40. C.R. HEIPLE AND J.R. ROPER, *WELD. J.,* VOL 61, 1982, P 97S

Selection of Wrought Austenitic Stainless Steels

John A. Brooks, Sandia National Laboratories; John C. Lippold, Edison Welding Institute

Weld Thermal Treatments

Preheat and Interpass Heat Treatments. Because the austenitic stainless steels do not experience a martensitic transformation upon cooling, there is generally no benefit derived from the use of preheat or interpass temperature control during multipass welding. In fact, these thermal treatments may actually increase the degree of sensitization by reducing cooling rates and allowing more time for carbide precipitation. Preheat and interpass heating can also increase distortion and cracking susceptibility.

Postweld heat treatment (PWHT) is often required to relieve residual stresses in austenitic stainless steel weldments, particularly in thick sections. Because the CTE value and the elevated-temperature yield and creep strengths of austenitic materials are significantly greater than for ferritic materials, the magnitude of residual stresses is generally larger. Although the effect of residual stresses from welding is typically not as severe as when less-ductile materials are used, they may still affect mechanical properties and corrosion behavior. PWHT is particularly critical when machining must be performed after welding, because significant distortion may occur.

Perhaps of greatest concern, however, is the influence of residual stress on SCC resistance. As described in the section "Corrosion Behavior" in this article, high levels of residual stress associated with the weld may severely reduce SCC resistance in chlorine-beating and caustic environments.

Stress relieving can be performed over a wide range of temperatures, depending on the amount of stress relaxation required. Time at temperature ranges from about 1 h per inch of section thickness at temperatures above 650 °C (1200 °F) to 4 h per inch of thickness at temperatures below 650 °C (1200 °F). Because of the high CTE value and low thermal conductivity of austenitic stainless steels, cooling from the stress-relieving temperature must be slow. Because the stress-relieving temperature range overlaps the sensitization temperature range, the stress-relieving temperature must be compatible with the extent of acceptable carbide precipitation and with the required amount of corrosion resistance. Nonstabilized stainless steels cannot be stress relieved in the sensitizing range without sacrificing corrosion resistance. Extra-low-carbon grades are affected much less, because sensitization is more sluggish. Stabilized grades exhibit minimal chromium carbide precipitation tendencies.

Higher stress-relief temperatures can be used for fully austenitic weld metals. For weld metals that contain ferrite, stress-relief temperatures above 650 °C (1200 °F) may result in weld embrittlement via sigma formation.

For austenitic stainless steels, the estimated percentages of residual stresses relieved at various temperatures at the times previously noted are 85% stress relief at 845 to 900 °C (1550 to 1650 °F) and 35% stress relief at 540 to 650 °C (1000 to 1200 °F). The data compiled by Cole and Jones (Ref 41) in Table 5 show specific time-temperature combinations that represent a compromise between the degree of stress relieving and carbide precipitation, which can lead to reduced corrosion resistance and ductility.

TABLE 5 EFFECT OF STRESS-RELIEF TREATMENTS ON AUSTENITIC STAINLESS STEELS

STAINLESS STEEL	**CONDITION(A)**	**STRESS-RELIEF HEAT TREATMENT(B)**			
		200-400 °C (392-752 °F), ~40% PEAK	**550-650 °C (1022-1202 °F), ~35%**	**850-900 °C (1562-1652 °F), ~85%**	**950-1050 °C (1742-1922 °F), ~95%**
TYPE 304	SHT, SA(C)	NME	IGP	IGP	IGP(D)
	WELDMENTS(C)	NME	IGP AND SIP	IGP AND SIP	IGP(D) AND SIP(E)
TYPE 321	SHT, SA	NME	NME	NME	NME
	WELDMENTS		SIP	SIP	SIP(E)
TYPE 304L	SHT, SA, AND WELDMENTS	NME	NME	NME	NME
TYPE 316	SHT, SA(C), AND WELDMENTS(C)	NME	IGP	IGP	IGP(D)
TYPE 316 + TI	SHT, SA, AND WELDMENTS	NME	NME	NME	NME
TYPE 316L	SHT, SA, AND WELDMENTS	NME	NME	NME	NME

(A) SHT, SOLUTION HEAT TREATMENT, RAPIDLY COOLED FROM ~1050 °C (1922 °F); SA, SOLUTION ANNEALED, SLOWLY COOLED FROM ~1050 °C (1922 °F).

(B) IGP, INTERGRANULAR PRECIPITATION OF CHROMIUM CARBIDE (SENSITIZATION); NME, NO METALLURGICAL EFFECT; SIP, STRAIN-INDUCED PRECIPITATION OF CARBIDES, IN WELDMENTS OVER 19 MM (0.75 IN.).

(C) LIKELY TO BE SENSITIZED PRIOR TO STRESS RELIEF, UNLESS OF SMALL SECTION SIZE AND/OR LOW CARBON CONTENT.

(D) DURING COOLING (ASSUMED ALWAYS TO BE SLOW).

(E) DURING HEATING

Reference cited in this section

41. C.L. COLE AND J.D. JONES, PUBLICATION 117, IRON AND STEEL INSTITUTE, LONDON, 1969, P 74

Selection of Wrought Austenitic Stainless Steels

John A. Brooks, Sandia National Laboratories; John C. Lippold, Edison Welding Institute

References

1. J.C. LIPPOLD, M.C. JUHAS, AND E.N.C. DALDER, *METALL. TRANS. A,* VOL 16, 1985, P 1835-1848
2. A.L. SCHAEFFLER, *MET. PROG.,* VOL 26, 1949, P 680
3. C.J. LONG AND W.T. DELONG, *WELD. J.,* VOL 52, 1973, P 281-S TO 297-S
4. D. KOTECKI AND T.A. SIEWERT, *WELD. J.,* VOL 71, 1992, P 171S-179S
5. H. THIER, *DVS-BER.,* VOL 41, 1976, P 100-104
6. I. MATSUMOTO, K. TAMAKI, AND M. MATSUMA, *J. JPN. WELD. SOC.,* VOL 41 (NO. 11), 1972, P 1306-1314
7. J.C. LIPPOLD AND W.F. SAVAGE, *WELD. J.,* VOL 61, 1982, P 388S-396S
8. G.R. SPEICH, *METALS HANDBOOK,* 8TH ED., VOL 8, AMERICAN SOCIETY FOR METALS, 1973, P 424
9. S.E. SCHURMANN AND I. BRAUCKMANN, *ARCH. EISENHUTTEN.,* VOL 48, 1977, P 3-8
10. N. SUUTALA, T. TAKALO, AND T. MOISIO, *METALL. TRANS. A,* VOL 10, 1979, P 512-514
11. J.A. BROOKS, J.C. WILLIAMS, AND A.W. THOMPSON, IN *TRENDS IN WELDING RESEARCH,* S.A. DAVID, ED., AMERICAN SOCIETY FOR METALS, 1982, P 331-355
12. S. KATAYAMA AND A. MATSUNAWA, *PROC. INT. CONF. ON ADVANCED LASER ELECTRO-OPTICS (ICALEO)* (SAN FRANCISCO), 1985
13. J.A. BROOKS, N.C.Y. YANG, AND J.S. KRAFCIK, ON THE ORIGIN OF FERRITE MORPHOLOGIES OF PRIMARY FERRITE SOLIDIFIED STAINLESS STEEL WELDS, *TRENDS IN WELDING RESEARCH,* ASM INTERNATIONAL, 1993
14. J.W. PUGH AND J.D. NISBET, *TRANS. TMS-AIME,* VOL 188, 1950, P 268-276
15. S.A. DAVID, *WELD. J.,* VOL 60, 1981, P 63-S TO 71-S
16. J.A. BROOKS AND A.W. THOMPSON, *LNT. MATER. REV.,* VOL 36 (NO. 1), 1991, P 16-44
17. N. SUUTALA AND T. MOISIO, THE USE OF CHROMIUM AND NICKEL EQUIVALENTS IN CONSIDERING SOLIDIFICATION PHENOMENA IN AUSTENITIC STAINLESS STEELS, *SOLIDIFICATION TECHNOLOGY IN THE FOUNDRY AND CASTHOUSE,* THE METALS SOCIETY PREPRINT, LONDON, 1980
18. T. TAKALO, N. SUUTALA, AND T. MOISIO, *METALL. TRANS. A* , VOL 10, 1979, P 1173-1181
19. D.L. OLSON, *WELD. J.,* VOL 64, 1985, P 281S-295S
20. "STANDARD PROCEDURES FOR CALIBRATING MAGNETIC INSTRUMENTS TO MEASURE THE DELTA FERRITE CONTENT OF AUSTENITIC AND DUPLEX AUSTENITIC-FERRITIC STEEL WELD METAL," AWS 4.2-91, AWS, 1991
21. R.H. ESPY, *WELD. J.,* VOL 61, 1982, P 1492-156S
22. F.C. HULL, *WELD. J.,* VOL 46, 1967, P 399-S TO 409-S
23. O. HAMMER AND U. SVENNSON, *SOLIDIFICATION AND CASTING OF METALS,* THE METALS SOCIETY, LONDON, 1979, P 401-410

24. S.A. DAVID, J.M. VITEK, AND T.L. HEBBLE, *WELD. J.,* VOL 66 (NO. 10), 1987, P 289S-300S
25. N. SUUTALA, EFFECT OF SOLIDIFICATION CONDITIONS ON THE SOLIDIFICATION MODE IN AUSTENITIC STAINLESS STEELS, *TECHNOLOGY IN THE FOUNDRY AND CASTHOUSE,* THE METALS SOCIETY PREPRINT, LONDON, 1980
26. J.A. BROOKS, M.I. BASKES, AND F.A. GREULICH, *METALL. TRANS. A,* VOL 22A, 1991, P 915-925
27. J.W. ELMER, S.M. ALLEN, AND T.W. EAGAR, *METALL. TRANS. A,* VOL 20A, 1989, P 2117-2131
28. G. PACARY, M. MOLINE, AND J.C. LIPPOLD, EWI RESEARCH BRIEF NO. B9008, EDISON WELDING INSTITUTE, 1990
29. J.A. BROOKS, A.W. THOMPSON, AND J.C. WILLIAMS, *WELD. J.,* VOL 63, 1984, P 71-S TO 83-S
30. J.C. LIPPOLD, I. VAROL, AND W.A. BAESLACK III, *WELD. J.,* VOL 71, 1992, P 1S-14S
31. J.A. BROOKS, *WELD. J.,* VOL 54, 1975, P 189S-195S
32. J.M. VITEK AND S.A. DAVID, *WELD. J.,* VOL 63, 1984, P 246S-253S
33. V. CIHAL, *PROT. MET.* (RUSSIA), VOL 4 (NO. 6), 1968, P 653
34. C.J. NOVAK, SENSITIZATION, *HANDBOOK OF STAINLESS STEELS,* D. PECKHER AND I.M. BERNSTEIN, ED., MCGRAW-HILL, 1977, P 4-35 TO 4-53
35. M.C. JUHAS AND B.E. WILDE, *CORROSION,* VOL 46, 1990, P 812-822
36. D.H. POPE, "MICROBIOLOGICAL ASPECTS OF C INFLUENCED CORROSION," ER-6345, ELECTRIC POWER RESEARCH INSTITUTE, 1988
37. S.W. BORENSTEIN, MICROBIOLOGICALLY INDUCED CORROSION OF AUSTENITIC STAINLESS STEEL WELDMENTS, *MATER. PERFORM.,* VOL 1, 1991, P 52-54
38. T.G. GOOCH, *PROC. CORROSION '92* (NASHVILLE), PAPER 296, NACE, 1992
39. CORROSION OF AUSTENITIC STAINLESS WELDMENTS, *CORROSION,* VOL 13, *ASM HANDBOOK,* ASM INTERNATIONAL, 1987, P 347-361
40. C.R. HEIPLE AND J.R. ROPER, *WELD. J.,* VOL 61, 1982, P 97S
41. C.L. COLE AND J.D. JONES, PUBLICATION 117, IRON AND STEEL INSTITUTE, LONDON, 1969, P 74

Selection of Wrought Duplex Stainless Steels

David N. Noble, ARCO Alaska, Inc.

Introduction

WROUGHT DUPLEX STAINLESS STEELS (DSS) are two-phase alloys based on the iron-chromium-nickel system. These materials typically comprise approximately equal proportions of the body-centered cubic (bcc) ferrite and face-centered cubic (fcc) austenite phases in their microstructure and are characterized by their low carbon content (<0.03 wt%) and additions of molybdenum, nitrogen, tungsten, and copper (Ref 1). Typical chromium and nickel contents are 20 to 30% and 5 to 10%, respectively. The specific advantages offered by DSS over conventional 300 series stainless steels are strength (about twice that of austenitic stainless steels), chloride stress-corrosion cracking (Cl SCC) resistance, and pitting corrosion resistance (Ref 2). These materials are used in the intermediate temperature range (about -60 to 300 °C, or -75 to 570 °F) where resistance to acids and aqueous chlorides is required. The weldability and welding characteristics of DSS are somewhat better than those of ferritic stainless steels, but generally not as good as austenitic materials.

Wrought DSS development in the 1970s and 1980s was particularly rapid, although the steels had been available in a number of compositions for several decades, particularly in the form of castings (Ref 3). Early grades were alloyed with approximately 18% Cr, about 4 to 6% Ni, and sometimes with molybdenum. Current commercial grades contain between 22 and 26% Cr, 4 to 7% Ni, up to 4.5% Mo, and 0.7% Cu and W, and are alloyed with 0.08 to 0.35% N (Ref 4). Continual modifications to the alloy compositions have been made to improve corrosion resistance, workability, and weldability. In particular, nitrogen additions have been effective in improving pitting corrosion resistance and weldability.

The current commercial DSS can be loosely divided into four generic types. With increasing corrosion resistance these alloy groups are:

- FE-23CR-4NI-0.1N
- FE-22CR-5.5NI-3MO-0.15N
- FE-25CR-5NI-2.5MO-0.17N-CU
- FE-25CR-7NI-3.5MO-0.25N-W-CU

As with 18-8 austenitic stainless steels, DSS are also frequently referred to by their chromium and nickel contents to describe the alloy class. The alloys listed above are then described as 2304, 2205, 25% Cr, and 2507, respectively. The last alloy class is also frequently termed "super" duplex stainless steel.

The performance of DSS can be significantly affected by welding. Due to the importance of maintaining a balanced microstructure and avoiding the formation of undesirable metallurgical phases, the welding parameters and filler metals employed must be accurately specified and closely monitored. The balanced microstructure of the base material (that is, equal proportions of austenite and ferrite) will be affected by the welding thermal cycle. If the balance is significantly altered and the two phases are no longer in similar proportions, the loss of material properties can be acute. Because the steels derive properties from both austenitic and ferritic portions of the structure, many of the single-phase base material characteristics are also evident in duplex materials. Austenitic stainless steels have excellent weldability and low-temperature toughness, whereas their Cl SCC resistance and strength are comparatively poor. Ferritic stainless steels have high resistance to Cl SCC but have poor toughness, especially in the welded condition. A duplex microstructure with high ferrite content can therefore have poor low-temperature notch toughness, whereas a structure with high austenite content can possess low strength and reduced resistance to Cl SCC (Ref 5). The high alloy content of DSS also renders them susceptible to formation of intermetallic phases from extended exposure to high temperatures. Extensive intermetallic precipitation may lead to a loss of corrosion resistance and sometimes to a loss of toughness (Ref 6).

References

1. H.D. SOLOMON AND T.M. DEVINE, JR., DUPLEX STAINLESS STEELS--A TALE OF TWO PHASES, *DUPLEX STAINLESS STEELS CONF. PROC.*, AMERICAN SOCIETY FOR METALS, 1983
2. B. LARSSON AND B. LUNDQVIST, FABRICATING FERRITIC-AUSTENITIC STAINLESS STEELS, SANDVIK STEEL TRADE LITERATURE, PAMPHLET S-51-33-ENG, OCT 1987
3. S. BERNHARDSSON, THE CORROSION RESISTANCE OF DUPLEX STAINLESS STEELS, *DUPLEX STAINLESS STEELS CONF. PROC.*, LES EDITIONS DE PHYSIQUE, LES ULIS CEDEX, FRANCE, OCT 1991, P 185-210
4. L. VAN NASSAU, H. MEELKER, AND J. HILKES, WELDING DUPLEX AND SUPER-DUPLEX STAINLESS STEELS--A GUIDE FOR INDUSTRY, DOCUMENT 01463, LINCOLN NORWELD, JULY 1992
5. T.G. GOOCH, WELDABILITY OF DUPLEX FERRITIC-AUSTENITIC STAINLESS STEELS, *DUPLEX STAINLESS STEELS CONF. PROC.*, AMERICAN SOCIETY FOR METALS, 1983, P 573-602
6. J. CHARLES, SUPER DUPLEX STAINLESS STEEL: STRUCTURE AND PROPERTIES, *DUPLEX STAINLESS STEELS CONF. PROC.*, VOL 1, LES EDITIONS DE PHYSIQUE, LES ULIS CEDEX, FRANCE, OCT 1991, P 3-48

Selection of Wrought Duplex Stainless Steels

David N. Noble, ARCO Alaska, Inc.

Base Material Properties

Microstructure. Duplex stainless steel parent materials have approximately equal proportions of austenite and ferrite, with ferrite comprising the matrix. Representative base material microstructures are shown in Fig. 1 and 2. Both exhibit a ferritic matrix with austenite islands of various morphologies. The wrought structure in Fig. 1 (a) parallel to the rolling direction has a pronounced orientation. The transverse view in Fig 1 (b) shows the austenite islands "end-on." The plan view in Fig 1 (c) also has some directionality but it is not as pronounced as in Fig. 1(a). A cast microstructure (Fig. 2) is coarser and displays a different morphology of austenite than that observed in the wrought plate.

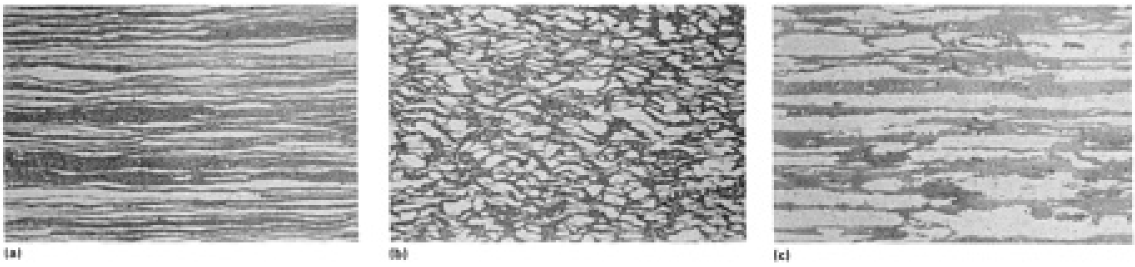

FIG. 1 EFFECT OF ORIENTATION PLANE ON THE MICROSTRUCTURE OF FE-22CR-5.5NI-3MO-0.15N WROUGHT DUPLEX STAINLESS STEEL BASE MATERIAL ELECTROLYTICALLY ETCHED IN 40% NAOH. (A) PARALLEL TO ROLLING DIRECTION. (B) TRANSVERSE TO ROLLING DIRECTION. (C) PLAN VIEW. 100X. COURTESY OF THE WELDING INSTITUTE

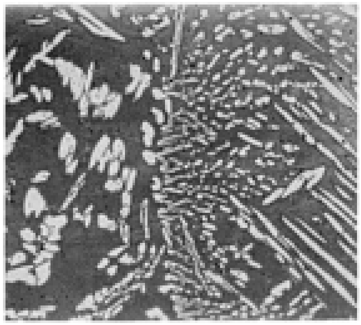

FIG. 2 MICROSTRUCTURE OF CAST FE-22CR-5.5NI-3MO-0.15N DUPLEX STAINLESS STEEL ETCHED ELECTROLYTICALLY IN 40% NAOH. 25×. COURTESY OF THE WELDING INSTITUTE

Alloy Grades. There are several types of DSS that have slightly different chemical compositions to cover the range of possible applications. The effects of chromium, nickel, molybdenum, and nitrogen on corrosion resistance are well understood and apply to most grades of stainless steels. The additional requirement of the alloying elements in the DSS base metal is to ensure that a balanced microstructure at room temperature is developed.

The DSS alloying additions are either austenite or ferrite formers. As the names suggest, certain elements will favor a higher proportion of austenite and others will favor ferrite. This is achieved by extending the temperature range over which the phase is stable. Among the major alloying elements in DSS, chromium and molybdenum are ferrite formers, whereas nickel, carbon, nitrogen, and copper are austenite formers (Ref 7, 8). The balance of austenite and ferrite formers will dictate the base material microstructure. Constitutional diagrams, such as the WRC-1992 diagram (Fig. 3) can be used for estimating the weld metal ferrite content in welds and developing appropriate base material compositions (Ref 8).

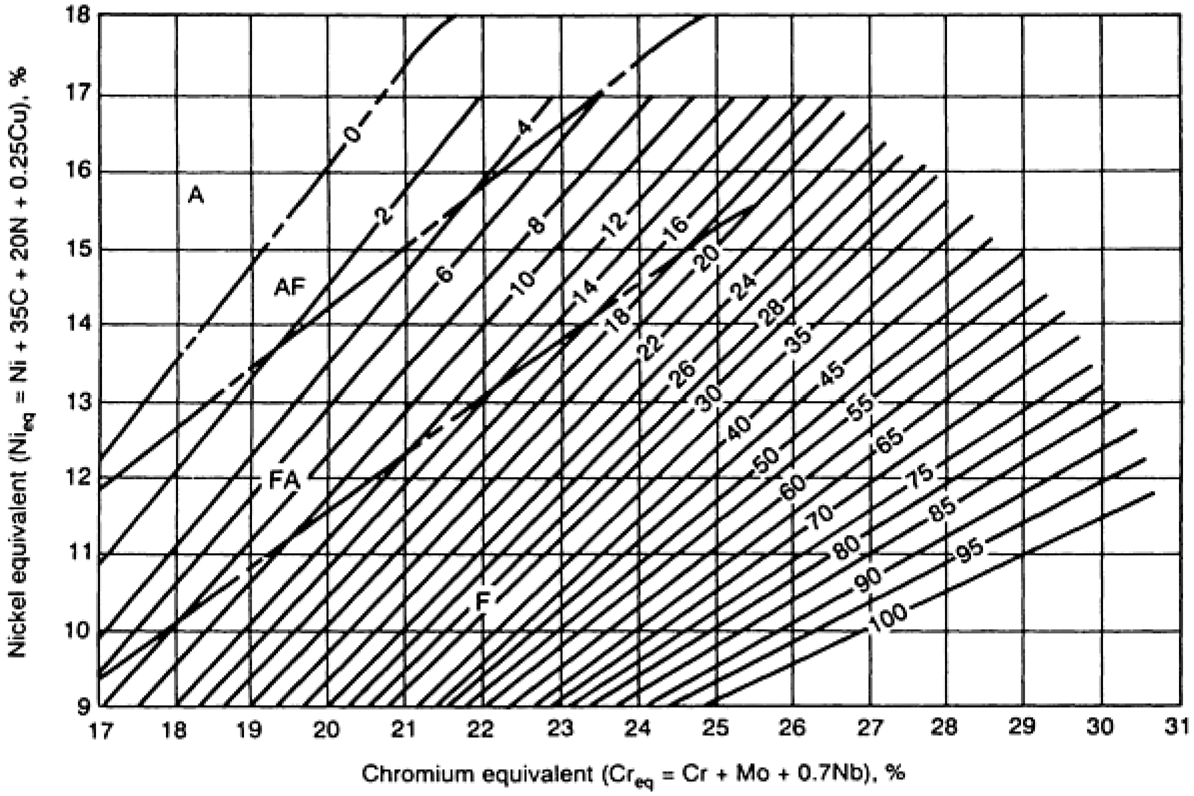

FIG. 3 WRC-1992 DIAGRAM FOR PREDICTING FERRITE CONTENT IN STAINLESS STEELS. FERRITE CONTENT IS GIVEN BY THE FERRITE NUMBER (FN), WHERE 100 FN IS APPROXIMATELY EQUAL TO 65 VOL% FERRITE. BOUNDARIES DENOTING A CHANGE IN SOLIDIFICATION MODE (A, PRIMARY AUSTENITE; AF, AUSTENITIC-FERRITIC; FA, FERRITIC-AUSTENITIC; F, PRIMARY FERRITE) ARE ALSO SHOWN (INDICATED BY DOTTED LINES). SOURCE: REF 8

Composition also plays a major role in the corrosion resistance of DSS. Pitting corrosion resistance is most easily affected. To determine the extent of pitting corrosion resistance offered by the material, the pitting resistance equivalent (PRE) is commonly used. This concept was proposed in the 1970s to characterize alloy resistance to pitting in chloride solutions (Ref 9). The PRE is calculated by adding the weight percentages of elements that affect pitting corrosion resistance, namely chromium, molybdenum, and nitrogen, then normalizing them with respect to the effect of 1% Cr. There are several formulas, which principally vary in the coefficient used for nitrogen, from 13 to 30. The PRE may also include negative coefficients for sulfur and phosphorus (Ref 10, 11). The most commonly used formula for pitting resistance equivalent is:

$$\text{PRE} = \%\text{CR} + 3.3(\%\text{MO}) + 16(\%\text{N}) \qquad \textbf{(EQ 1)}$$

The four types of DSS can also be classified by their PRE values. These range from 24 for the Fe-23Cr-4Ni-0.1N steel without molybdenum to over 40 for the Fe-25Cr-7Ni-3.5Mo-0.25N-Cu-W "super" grades. A minimum PRE value of 40 is often used to define the "super" DSS grades.

The chemical composition ranges of wrought DSS are given in Table 1, with the corresponding UNS numbers and PRE range for each. ASTM product specifications, in which the UNS numbers are referenced, are reported in Table 2. Some commercial grades of DSS, relevant composition ranges, and closest UNS numbers are given in Table 3.

TABLE 1 COMPOSITION AND PITTING RESISTANCE EQUIVALENT (PRE) VALUES OF SELECTED WROUGHT DUPLEX STAINLESS STEELS

UNS	COMPOSITION, %(A)

NUMBER	C	Mn	S	P	Si	Cr	Ni	Mo	N_2	OTHER	PRE RANGE$^{(B)}$
S31200	0.03	2.00	0.03	0.045	1.00	24.0-26.0	5.5-6.5	1.2-2.0	0.14-0.20	. . .	30.2-35.8
S31260	0.03	1.00	0.030	0.030	0.75	24.0-26.0	5.5-7.5	2.5-3.5	0.10-0.30	0.10-0.50 W, 0.20-0.80 CU	33.9-42.4
S31500	0.03	1.2-2.0	0.03	0.03	1.4-2.0	18.0-19.0	4.25-5.25	2.5-3.0	0.05-0.10	. . .	27.1-30.5
S31803	0.03	2.00	0.02	0.03	1.00	21.0-23.0	4.5-6.5	2.5-3.5	0.08-0.20	. . .	30.5-37.8
S32304	0.03	2.5	0.04	0.04	1.0	21.5-24.5	3.0-5.5	0.05-0.60	0.05-0.20	0.05-0.60 CU	22.5-29.7
S32550	0.03	1.5	0.03	0.04	1.0	24.0-27.0	4.5-6.5	2.9-3.9	0.10-0.25	1.5-2.5 CU	35.2-43.9
S32750	0.03	1.2	0.02	0.035	1.0	24.0-26.0	6.0-8.0	3.0-5.0	0.24-0.32	0.5 CU	37.7-47.6
S32760	0.03	1.0	0.01	0.03	1.0	24.0-26.0	6.0-8.0	3.0-4.0	0.30	0.5-1.0 CU, 0.5-1.0 W	40$^{(C)}$
S32900	0.06	1.00	0.03	0.04	0.75	23.0-28.0	2.5-5.0	1.0-2.0	$^{(D)}$	. . .	26.3-34.6
S32950	0.03	2.00	0.01	0.035	0.60	26.0-29.0	3.5-5.2	1.0-2.5	0.15-0.35	. . .	. . .

(A) SINGLE VALUES ARE MAXIMUM.

(B) PRE = %CR + 3.3(%MO) + 16(%N).

(C) MINIMUM VALUE.

(D) NOT SPECIFIED

TABLE 2 ASTM STANDARDS INCORPORATING WROUGHT DUPLEX STAINLESS STEELS OR THEIR EQUIVALENTS

ASTM STANDARD	PRODUCT APPLICATION	UNS NUMBERS INCLUDED
A 790/A 790M-91	SEAMLESS AND WELDED FERRITIC/AUSTENITIC STAINLESS STEEL PIPE	S31803, S31500, S32550, S31200, S31260, S32304, S32750, S32900, S32950
A 240-91A	HEAT-RESISTING CHROMIUM AND CHROMIUM-NICKEL STAINLESS STEEL PLATE, SHEET, AND STRIP FOR PRESSURE VESSELS	S31200, S31260, S31803, S32304, S32550, S32750, S32900, S32950
A 182/A 182-90B	FORGED OR ROLLED ALLOY-STEEL PIPE FLANGES, FORGED FITTINGS, AND VALVES AND PARTS FOR HIGH-TEMPERATURE SERVICE	S31200, S31803
A 789/A 789M-90	SEAMLESS AND WELDED FERRITIC/AUSTENITIC STAINLESS STEEL TUBING FOR GENERAL SERVICE	S31803, S31500, S32550, S31200, S31260, S32304, S32750, S32900, S32950
A 276-91	STAINLESS AND HEAT-RESISTING STEEL BARS AND SHAPES	S31803

TABLE 3 COMPOSITION AND PITTING RESISTANCE EQUIVALENT (PRE) VALUES FOR SELECTED COMMERCIAL DUPLEX STAINLESS STEELS

STEEL PRODUCER	**GRADE**	**COMPOSITION, %**							
		Cr	**Ni**	**Mo**	**N**	**Cu**	**OTHER**	**TYPICAL PRE VALUE**	**APPLICABLE UNS NO.**
FE-23CR-4NI-0.1N									
AVESTA	2304	23	4	. . .	0.10	. . .	. . .	25	S32304
CLI	UR 35 N	23	4	. . .	0.12	. . .	. . .	25	S32304
SANDVIK	SAF 2304	23	4	. . .	0.10	. . .	. . .	25	S32304
FE-22CR-5.5NI-3MO-0.15N									
ALLEGHENY LUDLUM	AL 2205	22	5.5	3.0	0.16	. . .	. . .	33-35	S31803
AVESTA	2205	22	5.5	3.0	0.16	. . .	. . .	33-35	S31803
BOHLER	A 903	22	5.5	3.0	0.16	. . .	. . .	33-35	S31803
CLI	UR 45 N	22	5.5	3.0	0.16	. . .	. . .	33-35	S31803
KRUPP	FALC 223	22	5.5	3.0	0.16	. . .	. . .	33-35	S31803
MANNESMANN	AF22	22	5.5	3.0	0.16	. . .	. . .	33-35	S31803
NIPPON KOKAN	NKCR22	22	5.5	3.0	0.16	. . .	. . .	33-35	S31803
SANDVIK	SAF 2205	22	5.5	3.0	0.16	. . .	. . .	33-35	S31803
SUMITOMO	SM22CR	22	5.5	3.0	0.16	. . .	. . .	33-35	S31803
THYSSEN	REMANIT 4462	22	5.5	3.0	0.16	. . .	. . .	33-35	S31803
VALOUREC	VS22	22	5.5	3.0	0.16	. . .	. . .	33-35	S31803
BRITISH STEEL	HYRESIST 22/5	22	5.5	3.0	0.16	. . .	. . .	33-35	S31803
FE-25CR-5NI-2.5MO-0.17N-CU									
BOHLER	A 905	25.5	3.7	2.3	0.37	. . .	5.8 MN	39	. . .
CARPENTER	7-MO PLUS	27.5	4.5	1.5	0.25	. . .	. . .	37	. . .
CLI	UR 47N	25	7	3.0	0.16	0.2	. . .	38	S31260
CLI	UR 52N	25	7	3.0	0.16	1.5	. . .	38	S32550
LANGLEY ALLOYS	FERRALIUM 255	26	5.5	3.3	0.17	2.0	. . .	39	S32550
MATHER AND PLATT	ZERON 25	25	4	2.5	0.15	. . .	. . .	36	. . .
SUMITOMO	DP-3	25	6.5	3.0	0.2	0.5	0.3 W	38	. . .
FE-25CR-7NI-3.5MO-0.25N-CU-W									
KRUPP-VDM	FALC 100	25	7	3.5	0.25	0.7	0.7 W	41	S32760
AVESTA	2507	25	7	4	0.28	. . .	. . .	43	S32750
SANDVIK	SAF 2507	25	7	4	0.28	. . .	. . .	43	S32750
WEIR MATERIALS LTD.	ZERON 100	25	6.5	3.7	0.25	0.7	0.7 W	41	S32760
CLI	UR 52N +	25	6.5	3.7	0.24	1.6	. . .	41	S32550

Mechanical and Physical Properties. The high yield strength of DSS offers designers the use of thin-wall material with adequate pressure-containing and load-bearing capacity. This can lead to major reductions in weight and welding time. The yield strength of the Fe-22Cr-5.5Ni- 3Mo-0.15N grade is 450 MPa (65 ksi) minimum, which is double that of austenitic stainless steels and significantly higher than that of ferritic stainless steels. Base material hardness varies dependent upon processing route and extent of cold working, but it generally averages about 260 HV 10 (25 HRC) for the Fe-22Cr-5.5Ni-3Mo-0.15N grade (Ref 2). Details of room-temperature tensile properties and hardnesses for each grade are given in Table 4, and the physical properties are reported in Table 5.

TABLE 4 ROOM-TEMPERATURE MECHANICAL PROPERTIES OF SELECTED DUPLEX STAINLESS

STEELS PER ASTM A 790

UNS NO.	**MINIMUM YIELD STRENGTH**		**MINIMUM TENSILE STRENGTH**		**ELONGATION (MINIMUM), %**	**HARDNESS**	
	MPA	**KSI**	**MPA**	**KSI**		**HB**	**HRC**
S31200	450	65	690	100	25	280	. . .
S31500	440	64	630	92	30	290	30.5
S31803	450	65	620	90	25	290	30.5
S32304	400	58	600	87	25	290	30.5
S32550	550	80	760	110	15	297	31.5
S32750	550	80	800	116	15	310	32
S32760(A)	550	80	750	109	25	200-270	. . .
S32900	485	70	620	90	20	271	28
S32950	480	70	690	100	20	290	30.5

(A) NOT LISTED IN ASTM A 790

TABLE 5 TYPICAL PHYSICAL PROPERTIES OF SELECTED DUPLEX STAINLESS STEELS

PROPERTY(A)	**GRADE**		
	UNS S32750	**UNS S32550**	**UNS S31803**
DENSITY, G/CM^3 (LB/IN.^{3})	7.8 (0.28)	7.8 (0.28)	7.8 (0.28)
THERMAL CONDUCTIVITY, W/M · K (BTU/FT · H · °F)			
AT 20 °C (70 °F)	16 (9.2)	17 (9.8)	. . .
AT 100 °C (210 °F)	17 (9.8)	18 (10.4)	. . .
AT 200 °C (390 °F)	19 (11.0)	19 (11.0)	. . .
AT 300 °C (570 °F)	20 (11.6)	. . .	. . .
AT 400 °C (750 °F)	21 (12.1)	. . .	. . .
THERMAL EXPANSION, 10^{-6}/K (10^{-6}/°F)			
AT 20-100 °C (70-210 °F)	13 (7.2)	. . .	. . .
AT 20-200 °C (70-390 °F)	13.5 (7.5)	. . .	. . .
AT 20-300 °C (70-570 °F)	14.0 (7.8)	. . .	. . .
AT 20-400 °C (70-750 °F)	14.5 (8.1)	. . .	. . .
YOUNG'S MODULUS, GPA (10^6 PSI)			
AT 20 °C (70 °F)	200 (29.0)	. . .	. . .
AT 100 °C (210 °F)	190 (27.6)	. . .	. . .
AT 200 °C (390 °F)	180 (26.1)	. . .	. . .
AT 300 °C (570 °F)	170 (24.7)	. . .	. . .
AT 400 °C (750 °F)	160 (23.2)	. . .	. . .
SPECIFIC HEAT CAPACITY AT 20 °C (70 °F), J/KG · K (BTU/LB · °F)	470 (112)	475 (113)	450 (107)

(A) UNLIKELY TO CHANGE SIGNIFICANTLY RELATIVE TO GRADE

The parent metal toughness down to approximately -100 °C (-150 °F) is good, but the weld metal and heat-affected zone (HAZ) are commonly much less tough and are frequently the limiting factor in DSS application. The parent metal will undergo a slight ductile-to-brittle transition (Fig. 4) but the change is not as marked as in other steels with bcc matrices. These alloys are not suitable for cryogenic applications.

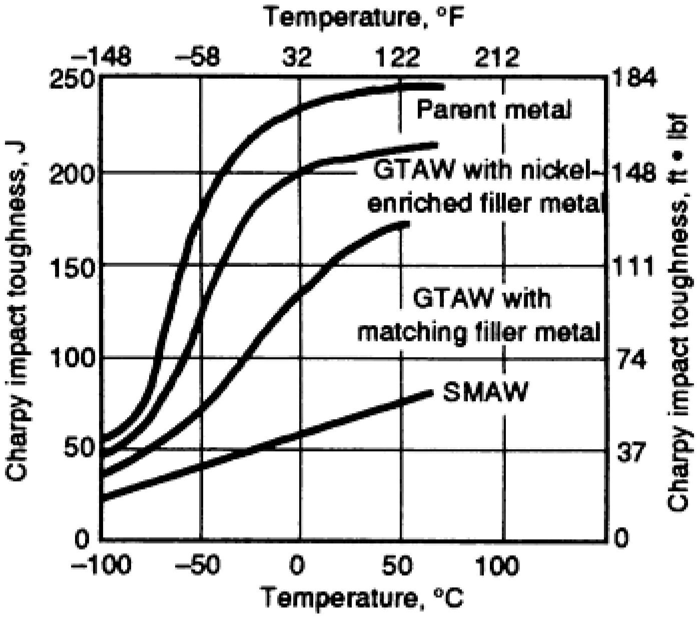

FIG. 4 PLOT OF CHARPY IMPACT TOUGHNESS VERSUS TEMPERATURE FOR FE-22CR-5.5NI-3MO-0.15N DSS BASE METAL AND WELDS PRODUCED WITH GAS-TUNGSTEN ARC WELDING (GTAW) AND SHIELDED METAL ARC WELDING (SMAW) PROCESSES. SOURCE: REF 2

The high alloy content and the presence of a ferritic matrix render DSS susceptible to embrittlement and loss of mechanical properties, particularly toughness, through prolonged exposure to elevated temperatures. This is caused by the precipitation of intermetallic phases, principally σ phase, χ phase, and η phase (Laves phase) (Ref 6). For this reason, the upper temperature of application is typically approximately 280 °C (535 °F) for parent material and 250 °C (480 °F) for welded structures (Ref 4). The more highly alloyed steels (for example, the Fe-25Cr-5Ni-2.5Mo-0.17N-Cu and Fe-25Cr-7Ni-3.5Mo-0.25N-Cu-W grades) are the most susceptible to the formation of these detrimental phases. Figure 5 shows the phases that can be formed in DSS, the approximate temperature range over which they will develop, and the effects of alloying elements on the transformation kinetics.

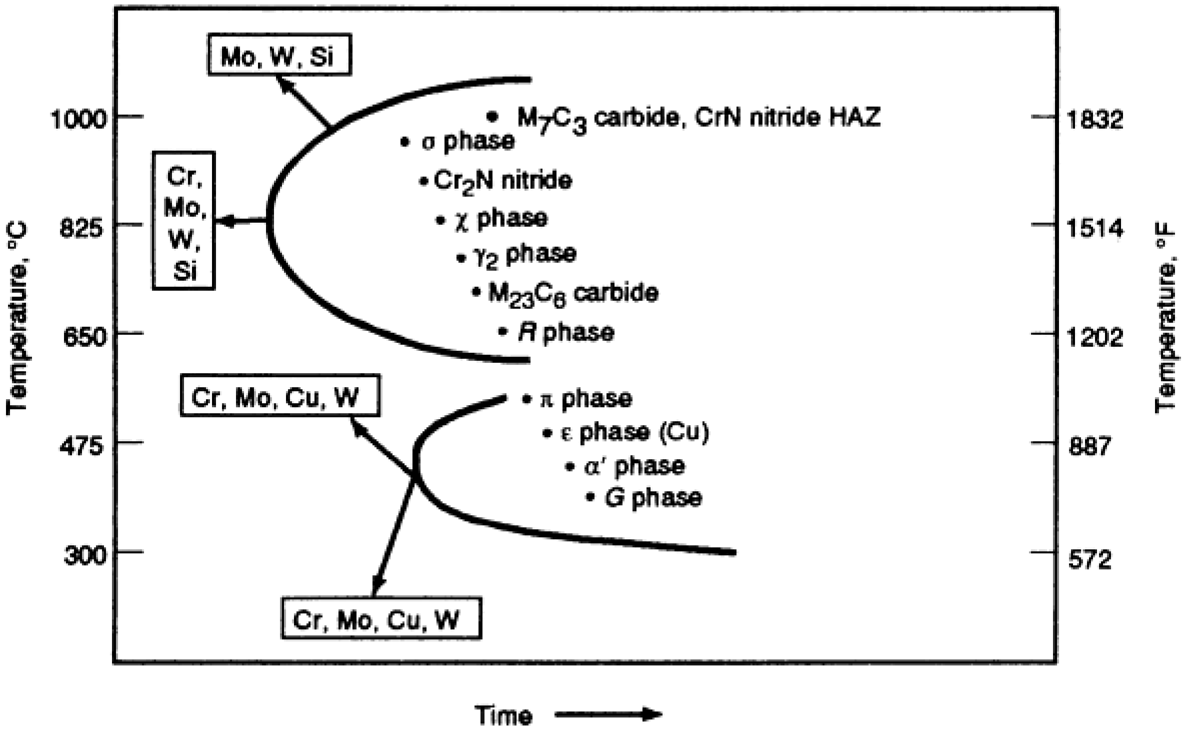

FIG. 5 TIME-TEMPERATURE TRANSFORMATION DIAGRAM SHOWING EFFECT OF ALLOYING ELEMENTS ON PRECIPITATION REACTIONS. SOURCE: REF 6

Corrosion Resistance. Duplex stainless steels were principally developed for their improved resistance to Cl SCC. Compared with conventional austenitic grades, they are clearly superior (Fig. 6). The more highly alloyed DSS are more resistant to Cl SCC than those with lower alloy contents.

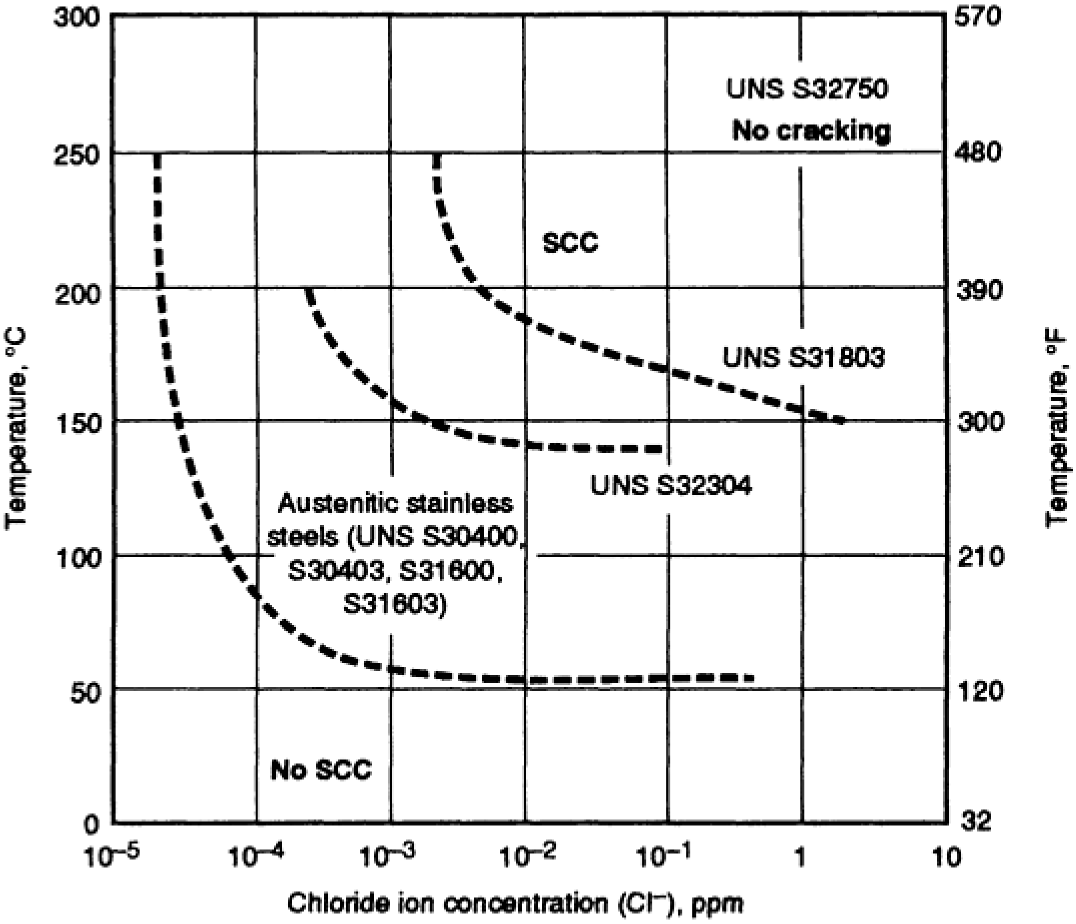

FIG. 6 STRESS-CORROSION CRACKING RESISTANCE OF SELECTED DUPLEX STAINLESS STEELS (S31803, S32304, AND S32750) RELATIVE TO AUSTENITIC STAINLESS STEELS (S30400, S30403, S31600, AND S31603) AS A FUNCTION OF TEMPERATURE AND CHLORIDE CONCENTRATION IN NEUTRAL O_2-BEARING SOLUTIONS (APPROXIMATELY 8 PPM). TEST DURATION WAS 1000 H. APPLIED STRESS WAS EQUAL TO YIELD STRENGTH. SOURCE: REF 2

Pitting corrosion resistance and general corrosion in acids are also generally acceptable (Fig. 7, 8) but these materials are susceptible to cracking in H_2S/Cl^- environments (Ref 12, 13, 14, 15). The tolerable H_2S limit for each grade depends on additional environmental factors (such as pH, temperature, chloride ion concentration, and so on) and is therefore not easily identified. Nevertheless, H_2S partial pressures of typically from 3.5 to 10 kPa (0.5 to 1.5 psi), dependent upon grade, have been reported as maxima for sour (H_2S-containing) service limits (Ref 16, 17). Regardless of H_2S partial pressure and chloride ion concentration, resistance to H_2S/Cl cracking appears to reach a minimum at about 60 to 100 °C (140 to 212 °F) (Ref 18).

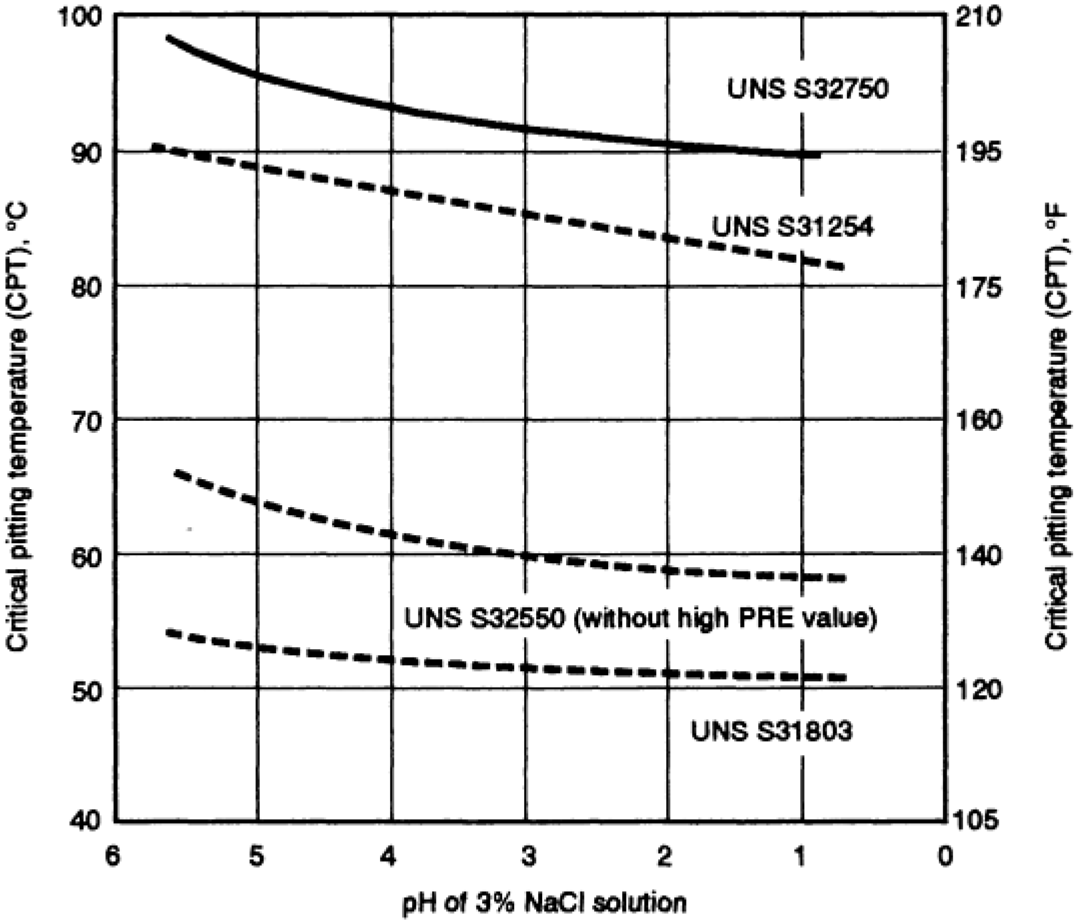

FIG. 7 CRITICAL PITTING TEMPERATURES FOR SELECTED DUPLEX STAINLESS STEELS (S31803, S32750, AND S32550) RELATIVE TO AUSTENITIC STAINLESS STEEL (S31254) IN 3% NACL SOLUTION WITH VARYING PH HELD AT AN ANODIC POTENTIAL OF +600 MV SATURATED CALOMEL ELECTRODE (SCE). SPECIMENS WERE SURFACE GROUND TO 600 GRIT. SOURCE: REF 2

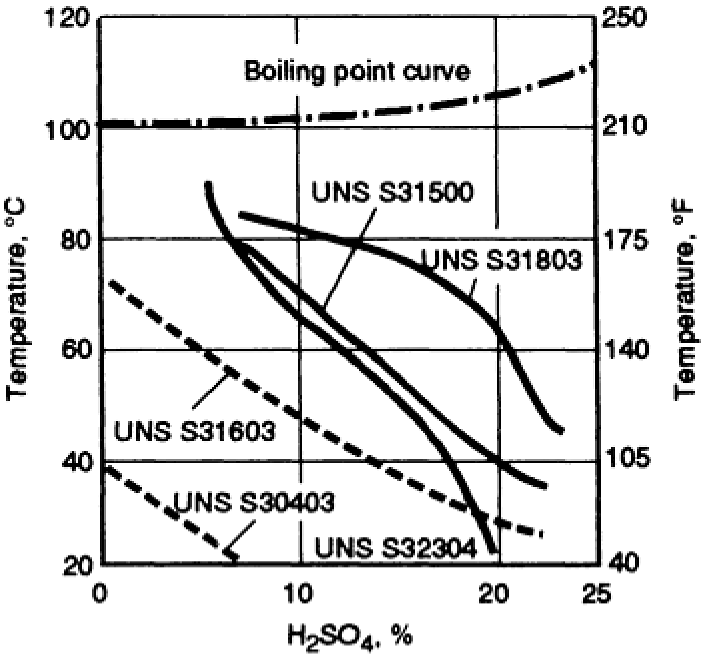

FIG. 8 ISOCORROSION DIAGRAM FOR SELECTED DUPLEX STAINLESS STEELS (S31803, S32304, AND S31500)

RELATIVE TO AUSTENITIC STAINLESS STEELS (S30403 AND S31603) IN STAGNANT SULFURIC ACID. CURVES REPRESENT CORROSION RATE OF 0.1 MM/YR (4 MILS/YR). SOURCE: REF 2

When considering corrosion performance, the effect of alloy partitioning within the microstructure is important. Even in base material that has undergone a soaking heat treatment at over 1000 °C (1830 °F), the austenite and ferrite phases will contain different amounts of the alloying elements. The ferrite will generally be highest in chromium and molybdenum, whereas the austenite will be richer in nickel and nitrogen (Ref 19). This imbalance in composition between the phases may translate into a difference in corrosion behavior, particularly pitting resistance. Significant differences in the PRE values have been reported between the ferrite and austenite phases in base material (Ref 19).

References cited in this section

2. B. LARSSON AND B. LUNDQVIST, FABRICATING FERRITIC-AUSTENITIC STAINLESS STEELS, SANDVIK STEEL TRADE LITERATURE, PAMPHLET S-51-33-ENG, OCT 1987
4. L. VAN NASSAU, H. MEELKER, AND J. HILKES, WELDING DUPLEX AND SUPER-DUPLEX STAINLESS STEELS--A GUIDE FOR INDUSTRY, DOCUMENT 01463, LINCOLN NORWELD, JULY 1992
6. J. CHARLES, SUPER DUPLEX STAINLESS STEEL: STRUCTURE AND PROPERTIES, *DUPLEX STAINLESS STEELS CONF. PROC.,* VOL 1, LES EDITIONS DE PHYSIQUE, LES ULIS CEDEX, FRANCE, OCT 1991, P 3-48
7. R. CASTRO AND J.J. DE CADENET, *WELDING METALLURGY OF STAINLESS AND HEAT-RESISTING STEELS,* CAMBRIDGE UNIVERSITY PRESS, 1974
8. D.J. KOTECKI AND T.A. SIEWERT, WRC-1992 CONSTITUTIONAL DIAGRAM FOR STAINLESS STEEL WELD METALS: A MODIFICATION OF THE WRC 1988 DIAGRAM, *WELD. J.,* MAY 1992, P 171S-178S
9. R.J. BRIGHAM AND E.W. TOZER, EFFECT OF ALLOYING ADDITIONS ON THE PITTING RESISTANCE OF 18% CR AUSTENITIC STAINLESS STEEL, *CORROSION,* VOL 30 (NO. 5), MAY 1974, P 161-166
10. N. SUUTALA AND M. KURKELA, *STAINLESS STEELS CONF. PROC.,* THE INSTITUTE OF METALS, 1985, P 240-247
11. R. DOLLING, V. NEUBERT, AND P. KNOLL, THE CORROSION BEHAVIOUR OF SUPER DUPLEX STEEL CAST ALLOYS WITH A PREN>41, *DUPLEX STAINLESS STEELS CONF. PROC.,* VOL 2, LES EDITIONS DE PHYSIQUE, LES ULIS CEDEX, FRANCE, OCT 1991, P 1341-1351
12. H. ERIKSSON AND S. BERNHARDSSON, THE APPLICABILITY OF DUPLEX STAINLESS STEELS IN SOUR ENVIRONMENTS, *CORROSION,* VOL 47 (NO. 8), SEPT 1991, P 719-727
13. A. MLYASAKA AND H. OGAWA, CRITICAL STRESS FOR SCC OF CORROSION RESISTANT ALLOY IN SOUR ENVIRONMENTS, PAPER 5, *CORROSION '89,* NATIONAL ASSOCIATION OF CORROSION ENGINEERS, 1989
14. S. BERNHARDSSON, THE CORROSION RESISTANCE OF DUPLEX STAINLESS STEELS, *DUPLEX STAINLESS STEELS CONF. PROC.,* AMERICAN SOCIETY FOR METALS, 1983, P 185-210
15. S. AZUMA, H. TSUGE, T. KUDO, AND T. MOROISHI, CREVICE CORROSION OF DUPLEX STAINLESS STEELS IN SIMULATED SOUR GAS ENVIRONMENTS, PAPER 306, *CORROSION '87,* NATIONAL ASSOCIATION OF CORROSION ENGINEERS, 1987
16. "SULFIDE STRESS CRACKING RESISTANT METALLIC MATERIALS FOR OILFIELD EQUIPMENT," MR-01-75-93, NATIONAL ASSOCIATION OF CORROSION ENGINEERS
17. L. SMITH, CORROSION RESISTANT ALLOYS OFFER PIPELINE ALTERNATIVES, *STAINL. STEEL EUR.,* JAN/FEB 1993, P 32-42
18. M.J. SCHOFIELD, AND R.D. KANE, DEFINING SAFE USE LIMITS FOR DUPLEX STAINLESS STEELS, *DUPLEX STAINLESS STEEL CONF. PROC.,* LES EDITIONS DE PHYSIQUE, LES ULIS CEDEX, FRANCE, OCT 1991, P 241-255

19. T. OGAWA AND T. KOSEKI, EFFECT OF COMPOSITION PROFILES ON METALLURGY AND CORROSION BEHAVIOUR OF DUPLEX STAINLESS STEEL WELD METALS, *WELD. J.,* VOL 88 (NO. 5), MAY 1989, P 181S-191S

Selection of Wrought Duplex Stainless Steels

David N. Noble, ARCO Alaska, Inc.

Applications

Duplex stainless steels have found widespread use in a range of industries, particularly the oil and gas, petrochemical, pulp and paper, and pollution control industries. They are commonly used in aqueous, chloride-containing environments and as replacements for austenitic stainless steels that have suffered from either Cl SCC or pitting during service (Ref 20). The more highly alloyed grades (that is, the "super" DSS) are resistant to oxygenated or chlorinated sea water (Ref 21). The leaner alloy materials are not resistant to live sea water and will suffer from pitting. Several grades have been used for transportation of wet, CO_2-containing gas where carbon would have suffered general corrosion (Ref 22). Table 6 gives a list of DSS grades and typical applications.

TABLE 6 TYPICAL APPLICATIONS OF DUPLEX STAINLESS STEELS

GRADE(S)	APPLICATION
S31803, S32760, S32750, S32550	TUBING FOR HEAT EXCHANGERS IN REFINERIES, CHEMICAL INDUSTRIES, PROCESS INDUSTRIES, AND OTHER INDUSTRIES USING WATER AS A COOLANT; KRAFT PAPER DIGESTERS
ALL	HEAT EXCHANGERS, CHEMICAL TANKERS, CHEMICAL REACTOR VESSELS, FLUE GAS FILTERS, ACETIC AND PHOSPHORIC ACID HANDLING SYSTEMS, OIL AND GAS INDUSTRY EQUIPMENT (MULTIPHASE FLOW LINES, DOWNHOLE PRODUCTION TUBULARS, COMMONLY COLD WORKED)
S32304	DOMESTIC WATER HEATERS AND WHERE PITTING RESISTANCE IS NOT OF OVERRIDING IMPORTANCE
S32750, S32760, S32550	PIPE FOR SEAWATER HANDLING AND FIREFIGHTING SYSTEMS, OIL AND GAS SEPARATORS, SALT EVAPORATION EQUIPMENT, DESALINATION PLANTS, GEOTHERMAL WELL HEAT EXCHANGERS, HUMAN BODY IMPLANTS (S32550 MAY SUFFER SLIGHT PITTING AND CREVICE CORROSION IN SEAWATER SERVICE)

References cited in this section

20. D. FRUYTIER, INDUSTRIAL EXPERIENCES WITH DUPLEX STAINLESS STEEL, *STAINL. STEEL EUR.,* VOL 3 (NO. 13), DEC 1991
21. B. WALLEN AND S. HENRIKSON, EFFECT OF CHLORINATION ON STAINLESS STEELS IN SEAWATER, PAPER 403, *CORROSION '86,* NATIONAL ASSOCIATION OF CORROSION ENGINEERS, 1986
22. J. HARSTON, E. HUTCHINS, AND S. SWEENEY, THE DEVELOPMENT AND CONSTRUCTION OF DUPLEX STAINLESS STEEL PIPELINES FOR USE OFFSHORE IN THE SOUTHERN NORTH SEA, *PROC. 3RD LNT. CONF. WELDING AND PERFORMANCE OF PIPELINES,* THE WELDING INSTITUTE, 1986

Microstructural Development

Factors Controlling Microstructure. By assuming a duplex stainless steel contains 68% Fe and 32% alloying elements (comprising nickel and chromium, but representing austenite and ferrite formers, respectively), the 68% Fe pseudo-binary phase diagram can be used to explain the microstructural transformations that occur during heating and cooling (Fig. 9).

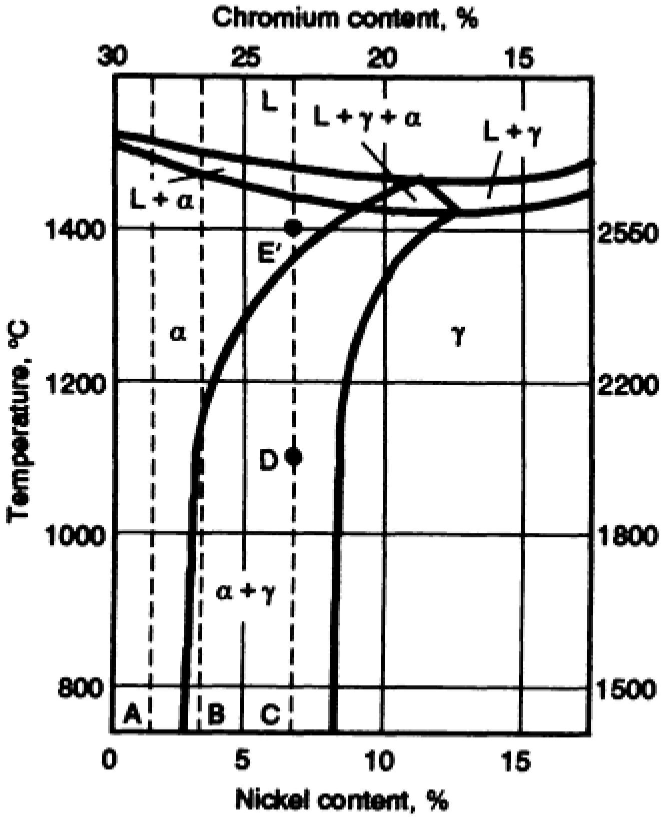

FIG. 9 PSEUDO-BINARY DIAGRAM AT CONSTANT IRON SECTION (68 WT% FE) OF THE FE/CR/NI TERNARY SYSTEM. SEE TEXT FOR DEFINITION OF LINES A, B, AND C AND POINTS D AND E. SOURCE: REF 9

Composition. With approximately 30% Cr and 2% Ni (composition indicated by line A), the steel will solidify completely to ferrite and remain ferritic down to room temperature. As the chromium content decreases and the nickel content increases, an alloy containing 28% Cr and 4% Ni, indicated by line B, will solidify as ferrite, remain fully ferritic until about 1200 to 1300 °C (2190 to 2370 °F), and then enter the dual-phase field (denoted by $\alpha + \gamma$). A two-phase microstructure will be developed and maintained down to room temperature. Because the composition of alloy B lies to the far left of the dual-phase field, a microstructure rich in ferrite would be anticipated. Current commercial DSS with 25% Cr are more likely to contain about 7% Ni, indicated by line C, which upon cooling enters the dual-phase field at a much higher temperature, thereby allowing more time for the diffusion-driven solid-state transformation from ferrite to austenite. An alloy with composition C will solidify as ferrite and be fully ferritic for a narrow temperature range prior to entering the dual-phase field. It should be noted that alloys A, B, and C will solidify primarily as ferrite and will have ferritic matrices.

Heat Treatment. The sequence of phase transformation in DSS upon cooling is (Fig. 9):

$$\text{LIQUID} \rightarrow \text{LIQUID + FERRITE} \rightarrow \text{FERRITE} \rightarrow \text{FERRITE + AUSTENITE} \qquad \textbf{(EQ 2)}$$

The solid-state transformation from ferrite to austenite is considered to be diffusional (Ref 23) and therefore affected by time and temperature. Very rapid cooling from a temperature at which the steel is fully ferritic will result in suppression of the ferrite-to-austenite transformation and retention of a predominantly ferritic structure. The temperature range over which the transformation from ferrite to austenite occurs is highly dependent upon composition and grade (Fig. 9), but is typically from 1200 to 800 °C (2190 to 1470 °F). For this reason, a $\Delta T_{12\text{-}8}$ cooling rate term has been used when characterizing a DSS thermal cycle (Ref 24, 25). A longer $\Delta T_{12\text{-}8}$ will result in more transformation from ferrite to austenite. In a weld of given thickness, longer $\Delta T_{12\text{-}8}$ values can be achieved by a higher heat input weld. The $\Delta T_{12\text{-}8}$ term has been employed in an effort to identify critical cooling rates necessary to develop acceptable austenite content in weld heat-affected zones to achieve the desired mechanical and corrosion resistance properties. The critical $\Delta T_{12\text{-}8}$ for phase transformation is independent of material thickness, unlike heat input, but is highly dependent on alloy composition.

Low-temperature thermal cycles will not affect the austenite/ferrite balance, but they will control the extent of intermetallic phase precipitation. These steels should undergo rapid cooling (that is, water quenching) through the precipitation temperature range to develop optimum properties. For welds, water quenching is usually not necessary or practical, but some means of forced cooling may be beneficial for productivity and quality reasons, particularly for multipass welds in the high-alloy grades (Ref 26, 27).

To produce the necessary base material microstructure, an isothermal heat treatment is carried out at a temperature at which the steel is dual phase, approximately 1050 to 1100 °C (1920 to 2010 °F), or point D in Fig. 9 (Ref 28). To avoid the formation of intermetallic phases during cooling, water quenching from the soaking temperature is necessary.

Morphology. Wrought DSS base material microstructures have a pronounced orientation of austenite islands in the ferritic matrix, parallel and transverse to the rolling direction, as a result of the hot and cold working following casting (Fig. 1). The cast microstructure (Fig. 2) is very different, and more similar to that observed in weld metal and the HAZ.

During the weld thermal cycle, the HAZ adjacent to the fusion line will become fully ferritic at temperatures above the austenite-to-ferrite transition temperature (point E in Fig. 9). Grain growth will occur and, regardless of the extent of austenite transformation that occurs during subsequent cooling, the coarse prior ferrite grain boundaries will still be apparent (Fig. 10). Very rapid cooling from the fully ferritic phase field will retard austenite reformation in the HAZ, and result in a predominantly ferritic structure, also shown in Fig. 10. When cooled more slowly, the austenite will form initially on the ferrite grain boundaries and then within the ferrite grains along preferred crystallographic planes in a manner similar to the cast material.

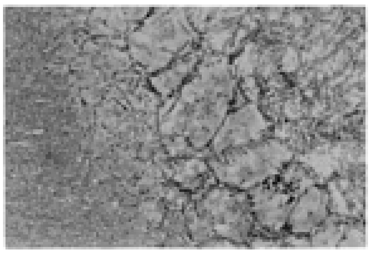

FIG. 10 PHOTOMICROGRAPH OF A FE-22CR-5.5NI-3MO-0.15N DSS WELD AREA SHOWING ELONGATED PARENT PLATE MORPHOLOGY, A COARSE GRAIN HEAT-AFFECTED ZONE HIGH IN FERRITE, AND COLUMNAR WELD METAL GRAINS. SPECIMEN ELECTROLYTICALLY ETCHED IN 40% NAOH. 30×. COURTESY OF THE WELDING INSTITUTE

The HAZ in Fig. 10 is relatively wide and has a high ferrite content. Such a structure may lead to poor low-temperature notch toughness and reduced corrosion resistance. Nitrogen levels in the base steel are being steadily increased to improve weldability and to avoid the formation of structures similar to that shown in Fig. 10; some DSS contain as much as 0.35 wt% N (Ref 4). A higher nitrogen level will increase the temperature at which ferrite begins to transform to austenite on cooling (Ref 6). This reduces the width of the transformed HAZ by limiting to a narrow temperature range the extent of the fully ferritic phase field. It also encourages more austenite reformation by starting the transformation at a much higher temperature. The recent grades of DSS, which typically have minimum 0.20 wt% N, develop a very narrow HAZ in which the transformed region is often difficult to distinguish.

The sequence of transformation reactions in the weld metal is the same as for the HAZ, but due to the solidification pattern, the macroappearance is different and columnar grains are more typical (this can be seen on the top left of Fig.

10). Duplex weld metal grain growth is epitaxial, with the large prior ferrite grains in the HAZ dictating weld metal grain size and orientation. Austenite again forms preferentially on the grain boundaries, and intragranular austenite forms along preferred crystallographic planes as the weld metal continues to cool (Fig. 11).

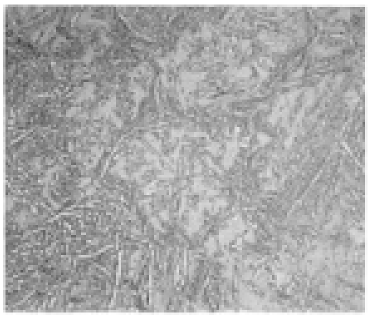

FIG. 11 PHOTOMICROGRAPH OF DSS WELD METAL WITH AUSTENITE LATHS (WHITE) AT PRIOR FERRITE GRAIN BOUNDARIES AND ALONG PREFERRED CRYSTALLOGRAPHIC PLANES IN A FERRITIC MATRIX. SAMPLE ELECTROLYTICALLY ETCHED IN 40% NAOH. 100×. COURTESY OF THE WELDING INSTITUTE

Precipitates are often difficult to observe and require precise etching techniques and experience with the type of microstructure. Due to its highly detrimental effect on corrosion resistance and toughness and relative ease of formation, the most commonly investigated precipitate in DSS is σ phase, an Fe(Cr,Mo) intermetallic. Sigma will initially precipitate at phase boundaries in the ferrite phase and will then grow into the ferrite phase, which is richer in chromium and molybdenum. Only a small amount of σ precipitation is required to reduce mechanical and corrosion properties.

References cited in this section

4. L. VAN NASSAU, H. MEELKER, AND J. HILKES, WELDING DUPLEX AND SUPER-DUPLEX STAINLESS STEELS--A GUIDE FOR INDUSTRY, DOCUMENT 01463, LINCOLN NORWELD, JULY 1992
6. J. CHARLES, SUPER DUPLEX STAINLESS STEEL: STRUCTURE AND PROPERTIES, *DUPLEX STAINLESS STEELS CONF. PROC.,* VOL 1, LES EDITIONS DE PHYSIQUE, LES ULIS CEDEX, FRANCE, OCT 1991, P 3-48
9. R.J. BRIGHAM AND E.W. TOZER, EFFECT OF ALLOYING ADDITIONS ON THE PITTING RESISTANCE OF 18% CR AUSTENITIC STAINLESS STEEL, *CORROSION,* VOL 30 (NO. 5), MAY 1974, P 161-166
23. J.E. LIPPOLD AND W.A. BAESLACK, *MET. CONSTR.,* VOL 20 (NO. L), 1988, P 21
24. H. HOFFMEISTER AND R. MUNDT, *ARCH. EISENHUTT.,* VOL 52 (NO. 4), P 159-164
25. D.N. NOBLE AND T.G. GOOCH, THE EFFECTS OF HEAT INPUT AND CHEMICAL COMPOSITION ON MICROSTRUCTURE IN ARC WELDED UNS S31803 TYPE DUPLEX STAINLESS STEELS, *ASM MATERIALS WEEK '87,* ASM INTERNATIONAL, OCT 1987
26. B. JOSEFSSON, J.-O. NILSSON, AND A. WILSON, PHASE TRANSFORMATION IN DUPLEX STEELS AND THE RELATION BETWEEN CONTINUOUS COOLING AND ISOTHERMAL HEAT TREATMENTS, *DUPLEX STAINLESS STEELS CONF. PROC.,* AMERICAN SOCIETY FOR METALS, 1983
27. L. ITRURGOYEN, J. ALCALA, AND M. ANGLADA, THE INFLUENCE OF AGEING AT 475 C ON THE FRACTURE RESISTANCE OF A DUPLEX STAINLESS STEEL, *CORROSION,* VOL 47 (NO. 8), SEPT 1991, P 757-764
28. "SANDVIK SAF 2507--A HIGH PERFORMANCE DUPLEX STAINLESS STEEL," SANDVIK STEEL TRADE LITERATURE, PAMPHLET S-1875-ENG, SEPT 1987

Selection of Wrought Duplex Stainless Steels

David N. Noble, ARCO Alaska, Inc.

Welding and Weldability

Duplex stainless steel weldability is generally good, although it are not as forgiving as austenitic stainless steels or as prone to degradation of properties as fully ferritic stainless steels. The current commercial grades are low in carbon (less than 0.03 wt%), thereby essentially eliminating the risk of sensitization and intergranular corrosion from carbide precipitation. The base material and filler metals also have low sulfur and phosphorus levels (less than 0.03 wt%), which in combination with the ferritic solidification reduce the likelihood of solidification cracking (hot cracking). Hydrogen cracking (cold cracking) resistance is also good due to the high hydrogen solubility in the austenite (Ref 29) and the high percentage of austenite in the matrix. Nevertheless, solidification cracking and hydrogen cracking can occur in duplex alloys, and some precautions are necessary.

Preheat is generally not recommended for DSS, but may sometimes be specified in low-nitrogen grades, because thick sections and low heat input welding processes may, in combination, develop highly ferritic heat-affected zones (Ref 30). For the more highly alloyed DSS, a preheat can be highly detrimental and reduce corrosion resistance and mechanical properties.

Postweld heat treatment (PWHT) is not commonly used except in autogenous welds or welds with a filler metal composition that exactly matches the base steel. Although not always necessary, particularly if a nickel-enriched filler metal is used, it is common to postweld heat treat DSS welded pipe after longitudinal seam welding. The PWHT will largely be for the purpose of restoring the correct phase balance and redissolving unwanted precipitates. Postweld heat treatment temperatures of about 1050 to 1100 °C (1920 to 2010 °F) are used, depending on grade, followed by the same heat treatment applied to the base material during solution annealing--usually water quenching (Ref 31). The heat treatments commonly used for structural steels (for example, 550 to 600 °C, or 1020 to 1110 °F) (Ref 32) are totally inappropriate for duplex alloys and should never be considered.

Interpass Temperature Control. The need for interpass temperature control depends on alloy grade and composition. The more highly alloyed grades should have tighter interpass control. For the Fe-22Cr-5.5Ni-3Mo-0.15N steels, interpass temperatures of 150 to 200 °C (300 to 390 °F) are typical, whereas for the Fe-25Cr-7Ni-3.5Mo-0.25N-(Cu-W) steels, temperatures as low as 70 °C (160 °F) have been recommended (Ref 33).

Welding practices employed for austenitic stainless steels (for example, cleanliness, use of a backing gas, avoidance of contamination with carbon steel, and so on) should also be adopted for DSS. The backing gas most commonly used is pure argon. However, the use of a backing gas without nitrogen can lead to nitrogen loss from the molten weld nugget (Ref 34) with probable loss in corrosion performance and toughness. Because of this, argon/nitrogen mixed gases for shielding and backing have been explored (Ref 35), and generally the corrosion resistance of welds produced with these gases has been superior to that of welds made with nitrogen-free gases. Pure nitrogen may also be used as a backing gas (Ref 2), but a concern exists that too much nitrogen may be picked up, thereby significantly affecting the phase balance locally by promoting the formation of austenite.

Open root gaps are commonly used in manual DSS welding to limit the extent of dilution from the base material, because extensive base metal dilution will negate the effect of the nickel over-alloying in the filler metal. To control the problems of too little austenite and intermetallic phase formation, associated with too-low and too-high heat inputs, respectively, DSS welding procedures typically specify a "window" of allowable heat inputs. For the highly alloyed and heat-sensitive super DSS, stringer beads are commonly specified, particularly in the hot pass (that is, weld pass immediately following the root), and in some cases in split hot passes (that is, two hot passes side by side). The technique of using a low heat input hot pass has led to the use of the term "cold" pass (Ref 33).

Welding procedure qualification generally follows the mechanical property test requirements of a major code, for example, ASME Section IX (Ref 36) or API 1104 (Ref 37). There are also several tests (specifically phase balance assessment, microstructural examination, and a pitting corrosion test) used for DSS weldment evaluation that are not typically employed for other stainless steels.

Phase-Balance Assessment. The phase balance test measures the amount of austenite and ferrite present in the microstructure and is similar to tests used to measure ferrite in austenitic weld metals. However, the testing requirements are usually more exhaustive for duplex alloys and will typically call for assessment of the ferrite content in the HAZ and weld metal. and possibly at several through-thickness locations (Ref 38). The amount of ferrite in a duplex structure can be specified and measured in two ways. Percentage ferrite is measured by manual point counting or image analysis. These methods measure the area fraction of ferrite and austenite exposed on the surface of a sample under examination. The values are reported as a volume fraction, thereby linking the surface area measurement on one plane to the overall volume fraction in the sample. Percentage ferrite requirements are typically 35 to 65% and are applied to weld metal and the HAZ.

The alternative method uses the ferromagnetic properties of the ferrite to measure its proportion in the matrix. The Ferrite Number (FN) system used for austenitic stainless steel weld metals (where the ferrite content is measured by attraction of a spring-loaded magnet) was not originally developed for high ferrite levels, such as those in duplex alloys, and was only calibrated to about 28 FN (Ref 39). Furthermore, the variations in iron content between the various duplex grades affected the magnetic attraction for a given phase balance (Ref 40). Because no reliable magnetic attraction system had been fully developed and proven at the time duplex alloys were seeing widespread industrial application, most specifications required ferrite percentage measurements. However, changes to the FN measuring system have now been made so that all ferrite levels in duplex alloys can be recorded, and the system has been shown to be more reliable. To achieve the extended range, counterweights were added to the spring-loaded magnet (Ref 40). Specification requirements for weld metals are typically 50 to 90 FN. A rule-of-thumb conversion from FN to percentage is:

$$0.70(\text{FN}) = \%\ \text{FERRITE} \qquad \textbf{(EQ 3)}$$

Measurement of ferrite content in DSS, however, is seldom currently performed with a magnetic measuring technique and is more commonly referred to in terms of percentage.

There are also some limitations to the application of the FN measuring system in certain regions of the weld. The measurement tool applies well to large areas of finely dispersed ferrite, such as in the weld metal, but it cannot accurately measure small areas that may have locally high ferrite levels (for example, the weld HAZ). It also may provide inaccurate readings in base material measurements due to the effect of magnet deflection toward the relatively coarse base material ferritic matrix.

Microstructural Examination. Optical microstructural examination is included to determine whether extensive phase precipitation has occurred (for example, σ-phase, carbides, nitrides, and so on), and specifications frequently call for the absence of any deleterious phases. However, the interpretation of DSS microstructures is so difficult that one cannot make a definitive assessment, especially without experience in studying similar structures. Detecting small proportions of precipitates by optical means is also relatively difficult.

Pitting Corrosion Test. The pitting corrosion test most commonly used for qualifying DSS is a modification of the ASTM G 48 Practice A ferric chloride test (Ref 41). Other tests are available (for example, potentiostatic critical pitting temperature tests in 3 or 15% NaCl, and others involving the use of CO_2), but the ASTM G 48 test has been shown to detect material with substandard properties and is preferred to most other common test techniques (Ref 42). The test solution is not representative of most service conditions in that it is highly acidic, oxidizing, and has a high chloride concentration, but the test is relatively easy to perform and can distinguish between good and bad welds. As used on DSS, the test involves the immersion of a welded coupon into a 10% $FeCl_3 \cdot 6H_2O$ solution for a duration of between 24 and 72 h at a temperature from 15 to 70 °C (60 to 160 °F), depending on material grade and specification requirements. Test assessment can be by direct weight loss, visual observation of pitting, weight loss per unit of exposed area, or corrosion rate. The procedure as laid out by ASTM was not designed for DSS weldment testing, and although it has been adapted in a number of ways, its interpretation still presents difficulties (Ref 43).

Filler Metal Requirements. For most DSS grades there are two types of filler metals:

- MATCHING COMPOSITION FILLER METALS
- FILLER METALS THAT ARE SLIGHTLY OVERALLOYED, PRINCIPALLY WITH RESPECT TO NICKEL

The matching filler metal is used where a PWHT is performed, whereas welds made with the filler metal enriched with nickel are used in the as-welded condition. The weld metal microstructure from a composition exactly matching that of the parent steel will contain a high ferrite content. The increase in nickel is made to improve the as-welded phase balance and increase austenite content. The ferrite content of a weld made with a nickel-enriched consumable would decrease significantly if it underwent a PWHT. It may suffer from slightly reduced weld metal strength and could also be more susceptible to σ phase formation during heat treatment (Ref 44). The nickel level in the enriched weld metal will be approximately 2.5 to 3.5% greater than in the base material (for example, for the Fe-22Cr-5.5Ni-3Mo-0.15N DSS base material containing 5.5% Ni, the filler metal will comprise 8.0 to 9.0% Ni, depending on consumable manufacturer and form). Nominal filler metal compositions for the Fe-22Cr-5.5Ni-3Mo-0.15N grade of DSS, proprietary names, and the American Welding Society classifications, where appropriate, are reported in Table 7.

TABLE 7 TYPICAL COMPOSITION OF SELECTED COMMERCIAL NICKEL-OVERALLOYED WELDING CONSUMABLES AND AWS SPECIFICATIONS FOR FE-22CR-5.5NI-3MO-0.15N DUPLEX STAINLESS STEEL

MANUFACTURER	TRADE NAME	COMPOSITION, %				
		C	Cr	Ni	Mo	N
COVERED ELECTRODES						
METRODE	SUPERMET 2205	0.02	25	9.5	3.5	0.17
SANDVIK	22.9.3.L	0.03	22	9.5	3.0	0.15
AVESTA	2205-PW	0.025	22	9.5	3.0	0.13
LINCOLN NORWELD	AROSTA 4462	0.025	22.0	9.0	3.0	0.14
THYSSEN	THERMANIT 22/09	<0.04	22.5	9.0	3.0	. . .
AWS A5.4	E2209	0.04(A)	21.0-24.0	8.0-10.0	2.5-4.0	0.08-0.20
BARE WIRE						
SANDVIK	22.8.3.L	<0.02	22.5	8.0	3.0	0.14
METRODE	ER329N	0.015	22.3	8.3	3.0	0.15
THYSSEN	22/09/SG	<0.025	23.0	9.0	3.0	. . .
LINCOLN NORWELD	LNT/LNM 4462	0.015	22.5	8.5	3.0	0.15
AWS A5.9(B)	ER 2209	0.04(A)	21.0-24.0	8.0-10.0	2.5-4.0	0.084-0.20

(A) MAXIMUM.

(B) PENDING

The higher alloy filler metals are sometimes used for welding a less alloyed base material (for example, a DSS filler metal with 25% Cr could be used for the root run in a Fe-22Cr-5.5Ni-3Mo-0.15N base metal). This is usually done to improve root weld metal corrosion resistance and thereby pass the qualification test requirements. In most cases, this does not lead to loss of mechanical properties; indeed, the more highly alloyed filler metal in the case above is likely to have greater strength.

To avoid all the requirements for weld metal phase balance and microstructural control necessary with duplex filler metals, nickel-base consumables (for example, AWS A 5.14 ERNi-CrMo-3) have been used. The yield strength, however, is slightly below that of the more highly alloyed grades, and the lack of nitrogen and the presence of niobium in the filler metal may contribute to unfavorable metallurgical reactions and the formation of intermetallic precipitates and areas of high ferrite content in the HAZ (Ref 45, 46).

Cracking Behavior. Duplex stainless steels can suffer from weld metal hydrogen cracking and solidification cracking (Ref 47, 48), but HAZ cracking has not been reported in practice and is considered highly unlikely to develop. Hydrogen cracking from welding and in-service hydrogen pickup has been observed (Ref 49). The duplex microstructure provides a combination of a ferritic matrix, where hydrogen diffusion can be fairly rapid, with intergranular and intragranular austenite, where the hydrogen diffusion is significantly slower, thereby acting as a barrier to hydrogen diffusion. The net effect appears to be that hydrogen can be "trapped" within ferrite grains by the surrounding austenite, particularly where it decorates the prior ferrite grain boundaries. Due to these characteristics, low-temperature hydrogen release treatments are not effective, and the hydrogen is likely to remain in the structure for a long period (Ref 49). Whether cracking actually develops will depend upon a number of factors, including the total amount of trapped hydrogen, the applied strain, and the amount of ferrite and austenite in the structure. Weld metal hydrogen content from covered electrodes can be relatively high, and levels up to 25 ppm have been reported (Ref 50).

The problem of weld metal hydrogen cracking in practice must not be overstated. The reported incidences of hydrogen cracking in DSS have been restricted to cases in which the alloy has been heavily cold worked or weld metals have seen high levels of restraint or possessed very high ferrite contents in combination with very high hydrogen levels as a result of poor control of covered electrodes or the use of hydrogen-containing shielding gas. Indeed, other studies have shown how resistant DSS weld metals are to hydrogen cracking, even with consumables intentionally humidified (Ref 51), and that hydrogen-containing backing gases can be employed without producing cracking (Ref 2). There is no doubt an effect of hydrogen on the ductility of DSS, and to avoid fabrication-related cracking problems, high-hydrogen-potential welding processes, such as shielded metal arc welding, should be controlled by careful storage and use of electrodes, and by ensuring that the weld metal phase balance is within acceptable limits.

Solidification cracking in DSS has not been reported as a commercial problem, but it has been witnessed in laboratory-made high heat input submerged arc welds (Ref 48). Cracking occurs at the solidification grain boundaries and has the appearance of running through the austenite phase. This is due to the subsequent solid-state transformation that occurs, in which the austenite forms predominantly on the prior ferrite grain boundaries. The solidification cracking susceptibility is not great, however, and weldability cracking tests have shown that the performance of DSS is similar to that of austenitic stainless materials with approximately 3 to 5 FN (Ref 52). Resistance to HAZ liquation cracking is also reportedly similar to type 304 stainless steel (Ref 53) and unlikely to develop in practical welding situations.

Loss of Properties. The weld is usually the part of a system with reduced corrosion resistance and low-temperature toughness, and therefore in many cases it is the limiting factor in material application. Welding primarily affects pitting corrosion, stress corrosion, and Charpy toughness.

Pitting corrosion resistance can be affected by many features of the welding operation, including:

- LOCALIZED SEGREGATION OF ALLOYING ELEMENTS TO THE DIFFERENT CONSTITUENT PHASES IN THE MICROSTRUCTURE, PRODUCING AREAS LEAN IN MOLYBDENUM AND CHROMIUM
- INCORRECT FERRITE/AUSTENITE PHASE BALANCE
- FORMATION OF NITRIDES OR INTERMETALLIC PHASES
- LOSS OF NITROGEN FROM THE ROOT PASS
- PRESENCE OF AN OXIDIZED SURFACE ON THE UNDERSIDE OF THE ROOT BEAD

The extent to which the reduction occurs depends on which of these factors are active and to what degree. Partitioning of alloying elements between the austenite and ferrite occurs in the weld metal, with chromium, molybdenum, and silicon partitioning to the ferrite and carbon, nickel, and nitrogen to the austenite (Ref 19, 54). The effect is not so apparent in as-deposited weld metals, but it becomes more significant as a result of reheating a previously deposited weld pass. The effect is also exacerbated by higher welding heat inputs (Ref 4, 54).

Weld metal and HAZ microstructures with very high ferrite contents are also less resistant to pitting attack than am balanced structures. This is largely because predominantly ferritic structures are more prone to chromium nitride precipitation, which locally denudes the chromium concentration and lowers resistance to pitting attack.

Nitrogen loss in the root pass may reduce weld metal corrosion resistance. Up to 20% loss of nitrogen has been reported for gas-tungsten arc welds (Ref 34), and nitrogen-bearing backing gases have been explored and used in limited applications (Ref 35).

Cleanliness of the root side purge gas may also affect pitting resistance. Figure 12 shows the effect of reducing oxygen content in an otherwise pure argon purge gas and its beneficial effect on pitting resistance. Also shown is the apparent benefit of using a reducing gas (NH_{10}), which would significantly reduce the tendency for oxide formation and leave the underbead appearance very shiny (Ref 55).

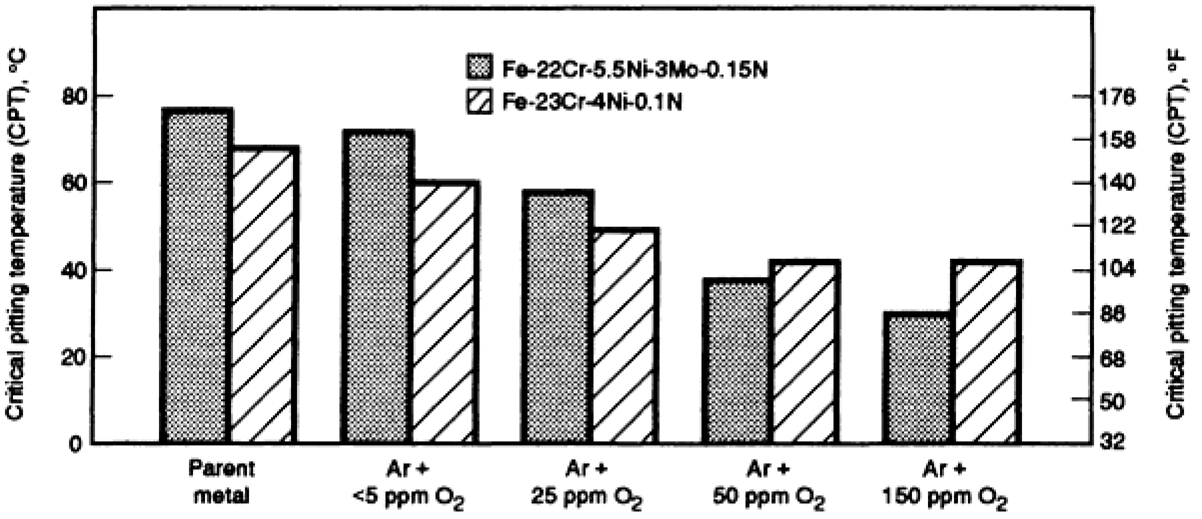

FIG. 12 PLOT OF PITTING TEMPERATURE VERSUS OXYGEN CONTENT OF BACKING GAS FOR FE-22CR-5.5NI-3MO-0.15N AND FE-23CR-4NI-0.1N DUPLEX STAINLESS STEELS TESTED IN 3% NACL AND 0.1% NACI SOLUTIONS, RESPECTIVELY, BOTH AT ANODIC POTENTIAL OF + 300 MV. SOURCE: REF 55

The net effect on pitting corrosion resistance may be observed by applying the ASTM G 48 pitting corrosion test to welds and base material with the same PRE value, then assessing the reduction in critical pitting temperature (that is, the temperature at which pitting in the ferric chloride solution is first observed). The difference is approximately 20 °C (35 °F), as reported in Fig. 13, thereby quantifying the effect of reduced weld metal properties. Figure 13 also shows that the use of a super DSS filler metal with a PRE value of about 40 on a Fe-22Cr-5.5Ni-3Mo-0.15N parent steel (which typically has a PRE value of about 33 to 35) will improve the weld metal pitting corrosion resistance, as assessed by the ASTM G 48 test, to approximately match that of the base material.

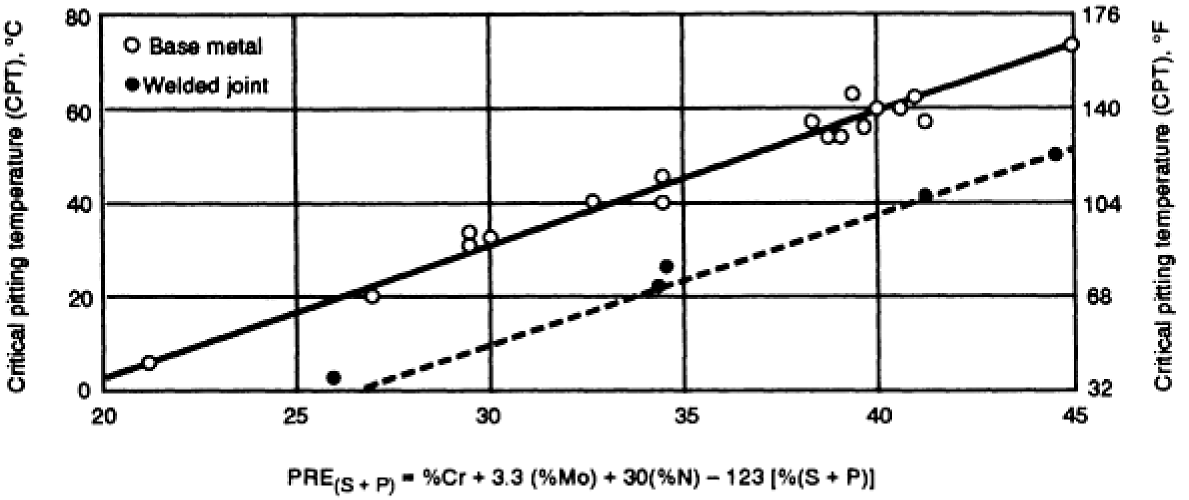

FIG. 13 PITTING CORROSION RESISTANCE OF BASE METAL RELATIVE TO WELD METAL PLACED IN 6 WT% $FECl_3$ SOLUTION FOR 24 H DURATION PER ASTM 6 48 (METHOD A). SOURCE: REF 11

Resistance to CI SCC does not appear to be affected significantly by welding per se (Ref 56). Nevertheless, welds are likely regions of attack for CI SCC due to the presence of high stresses and the structural inhomogeneity present at the weld. If localized pitting is a necessary precursor for CI SCC, the effects described above will also ultimately affect Cl SCC resistance.

There are three main causes of poor weld metal toughness:

- VERY HIGH FERRITE CONTENT
- PRESENCE OF INTERMETALLIC PHASES OR NITRIDES
- HIGH WELD METAL OXYGEN CONCENTRATION

Predominantly ferritic microstructures with a coarse grain size, as may be developed in a weld HAZ, have poor low-temperature notch toughness. Their performance is similar to that of ferritic stainless steels, and they may undergo low energy cleavage fracture. The Fe-25Cr-5Ni-2.5Mo-0.17N-Cu and super DSS are more susceptible to embrittlement through intermetallic phase precipitation, due to their increased kinetics of formation (Ref 57), and heat input and interpass temperature control are advised. Welding processes that produce high weld metal oxygen contents also produce welds with lower toughness, mainly due to the size and distribution of nonmetallic inclusions. Gas-shielded processes have the least oxygen potential and generally exhibit superior Charpy toughness properties (Fig. 14).

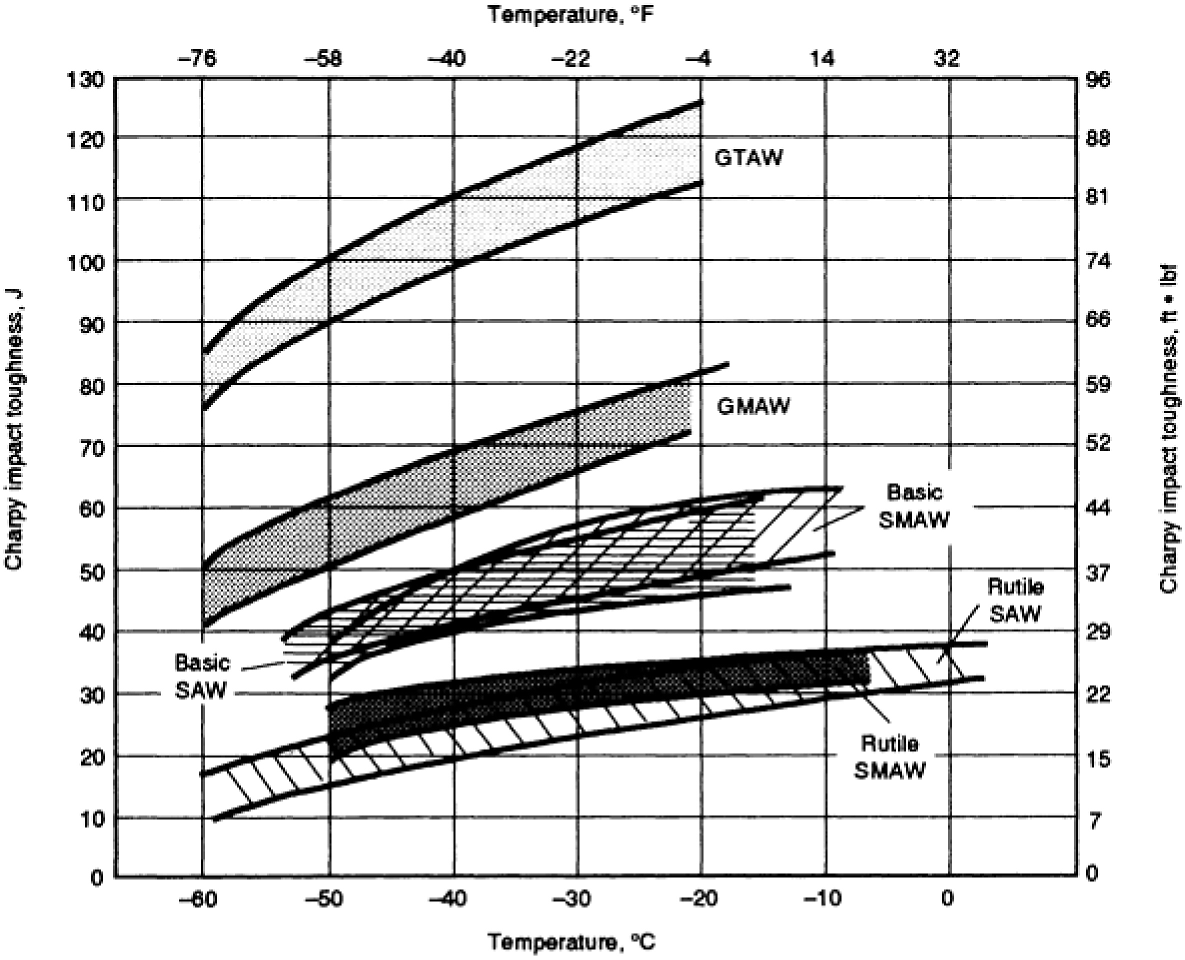

FIG. 14 PLOT OF CHARPY V-NOTCH TOUGHNESS VERSUS TEMPERATURE AS A FUNCTION OF WELDING PROCESS FOR UNS S32760 ALLOY. THE TERMS "BASIC" AND "RUTILE" REFER TO LIME-BASED AND TITANIA-BASED ELECTRODE COATINGS USED WITH SAW AND SMAW. SOURCE: REF 58

References cited in this section

2. B. LARSSON AND B. LUNDQVIST, FABRICATING FERRITIC-AUSTENITIC STAINLESS STEELS, SANDVIK STEEL TRADE LITERATURE, PAMPHLET S-51-33-ENG, OCT 1987
4. L. VAN NASSAU, H. MEELKER, AND J. HILKES, WELDING DUPLEX AND SUPER-DUPLEX

STAINLESS STEELS--A GUIDE FOR INDUSTRY, DOCUMENT 01463, LINCOLN NORWELD, JULY 1992

11. R. DOLLING, V. NEUBERT, AND P. KNOLL, THE CORROSION BEHAVIOUR OF SUPER DUPLEX STEEL CAST ALLOYS WITH A PREN>41, *DUPLEX STAINLESS STEELS CONF. PROC.,* VOL 2, LES EDITIONS DE PHYSIQUE, LES ULIS CEDEX, FRANCE, OCT 1991, P 1341-1351
19. T. OGAWA AND T. KOSEKI, EFFECT OF COMPOSITION PROFILES ON METALLURGY AND CORROSION BEHAVIOUR OF DUPLEX STAINLESS STEEL WELD METALS, *WELD. J.,* VOL 88 (NO. 5), MAY 1989, P 181S-191S
29. F.R. COE, *WELDING STEELS WITHOUT HYDROGEN CRACKING,* TWI PUBLICATIONS, 1978
30. "HOW TO WELD 2205," AVESTA WELDING TRADE LITERATURE
31. "SEAMLESS AND WELDED FERRITIC/AUSTENITIC STAINLESS STEEL PIPE," ASTM A 790-91
32. "CHEMICAL PLANT AND PETROLEUM REFINERY PIPING," ASME CODE B31.3, P 66-67
33. A.W. STEPHENSON, P.C. GOUGH, AND J.C.M. FARRAR, "THE WELDABILITY OF SUPER DUPLEX ALLOYS--WELDING CONSUMABLES AND PROCEDURE DEVELOPMENT FOR ZERON 100, *PROC. LNT. INST. WELDING ANNUAL ASSEMBLY CONFERENCE,* JULY 1991
34. T.G. GOOCH, CORROSION RESISTANCE OF WELDS IN DUPLEX STAINLESS STEELS, *DUPLEX STAINLESS STEELS CONF. PROC.,* AMERICAN SOCIETY FOR METALS, 1983, P 325-346
35. O. JONSSON, M. LILJAS, AND P. STENVALL, THE ROLE OF NITROGEN IN LONGITUDINAL WELDING OF TUBING IN DUPLEX STAINLESS STEELS, *DUPLEX STAINLESS STEELS CONF. PROC.,* AMERICAN SOCIETY FOR METALS, 1983, P 461-468
36. "WELDING AND BRAZING QUALIFICATIONS," ASME 1989 BOILER AND PRESSURE VESSEL CODE SECTION IX
37. "WELDING OF PIPELINES AND RELATED FACILITIES," STANDARD 1104, AMERICAN PETROLEUM INSTITUTE
38. "SPECIFICATION FOR CRA LINEPIPE," STANDARD 5LC, AMERICAN PETROLEUM INSTITUTE
39. "STANDARD PROCEDURES FOR CALIBRATING MAGNETIC INSTRUMENTS TO MEASURE THE DELTA FERRITE CONTENT OF AUSTENITIC STAINLESS STEEL WELD METALS," AWS A4.2-74, AMERICAN WELDING SOCIETY
40. D.J. KOTECKI, EXTENSION OF THE WRC FERRITE NUMBER SYSTEM, *WELD. J.,* NOV 1982, P 352S-361S
41. "TESTING FOR PITTING AND CREVICE CORROSION RESISTANCE OF STAINLESS STEELS AND RELATED ALLOYS BY THE USE OF FERRIC CHLORIDE SOLUTION," G 48 PRACTICE A, ASTM
42. R.M. DAVISON AND J.D. REDMOND, DEVELOPMENT OF QUALIFICATION TESTS FOR DUPLEX STAINLESS STEEL MILL PRODUCTS, PAPER 302, *CORROSION '91,* NATIONAL ASSOCIATION OF CORROSION ENGINEERS, 1991
43. R.A. CORBETT, PROBLEMS IN UTILIZING ASTM G 48 TO EVALUATE HIGH ALLOY STAINLESS STEELS, PAPER 298, *CORROSION '92,* NATIONAL ASSOCIATION OF CORROSION ENGINEERS, 1992
44. D.J. KOTECKI, HEAT TREATMENT OF DUPLEX STAINLESS STEEL WELD METALS, *WELD. J.,* NOV 1989, P 431S-441S
45. L. ODEGARG AND S.-A. FAGER, CHOICE OF RIGHT FILLER METAL FOR JOINING SANDVIK SAF 2507 TO A SUPER AUSTENITIC STAINLESS 6 MO STEEL, *ASM MATERIALS WEEK '87,* ASM INTERNATIONAL, OCT 1987, P 441-449
46. R.N. GUNN, THE INFLUENCE OF COMPOSITION AND MICROSTRUCTURE ON THE CORROSION BEHAVIOUR OF COMMERCIAL DUPLEX ALLOYS, *RECENT DEVELOPMENTS IN THE JOINING OF STAINLESS STEELS AND HIGH ALLOYS,* EDISON WELDING INSTITUTE, COLUMBUS, OHIO, OCT 1992
47. G.A. CLARK AND P. GUHA, *BR. WELD. J.,* VOL 13 (NO. 5), MAY 1981, P 269-273

48. J. HONEYCOMBE AND T.G. GOOCH, *WELD. J.,* VOL 56 (NO. 11), NOV 1977, P 339S-353S
49. P. SENTANCE, THE BRAE FIELD, *STAINL. STEEL EUR.,* VOL 4 (NO. 20), SEPT 1992, P 38-41
50. C.D. LUNDIN, K. KIKUCHI, AND K.K. KAHN, PHASE 1 REPORT; "MEASUREMENT OF DIFFUSIBLE HYDROGEN CONTENT AND HYDROGEN EFFECTS ON THE CRACKING POTENTIAL OF DUPLEX STAINLESS STEEL WELDMENTS," THE WELDING RESEARCH COUNCIL, MARCH 1991
51. R.A. WALKER AND T.G. GOOCH, HYDROGEN CRACKING OF WELDS IN DUPLEX STAINLESS STEEL, *CORROSION,* VOL 47 (NO. 8), SEPT 1991, P 1053-1063
52. I. VAROL, W.A. BAESLACK, AND J.C. LIPPOLD, *METALLOGRAPHY,* VOL 23 (NO. 1), 1989
53. J.C. LIPPOLD, W.A. BAESLACK, AND I. VAROL, WELDING OF DUPLEX STAINLESS STEELS, *WELD. J.,* IN PRESS
54. R.A. WALKER AND T.G. GOOCH, PITTING RESISTANCE OF WELD METAL FOR 22CR-5NI FERRITIC-AUSTENITIC STAINLESS STEEL, *BR. CORROS. J.,* VOL 26 (NO. 1), 1991, P 51-59
55. L. ODEGARD AND S.-A. FAGER, THE ROOT SIDE PITTING CORROSION RESISTANCE OF STAINLESS STEEL WELDS, *SANDVIK STEEL WELD. REP.,* NO. 1, 1990
56. T.G. GOOCH, *WELDING IN THE WORLD,* VOL 24 (NO. 7/8), 1986, P 148-167
57. J. CHARLES, SUPER DUPLEX STAINLESS STEEL: STRUCTURE AND PROPERTIES, *DUPLEX STAINLESS STEELS CONF. PROC.,* AMERICAN SOCIETY FOR METALS, 1983, P 151-168
58. P.C. GOUGH AND J.C.M. FARRAR, FACTORS AFFECTING WELD ROOT RUN CORROSION PERFORMANCE IN DUPLEX AND SUPER DUPLEX PIPEWORK, *ASM MATERIALS WEEK '87,* ASM INTERNATIONAL, OCT 1987, P 1009-1025

Selection of Wrought Duplex Stainless Steels

David N. Noble, ARCO Alaska, Inc.

Applicable Welding Processes

Fusion Welding. Nearly all of the arc welding processes that are employed for other stainless steels can be used with duplex alloys, except where the process characteristic is to weld autogenously, such as electron-beam welding and laser-beam welding. In such circumstances a PWHT is nearly always necessary to restore the correct phase balance to the weld metal and remove any undesirable precipitates. There are few reported differences in corrosion resistance between welding processes, but the nonmetallic inclusion distribution would be anticipated to have an effect. In most instances of pipe welding, where access is from one side only, gastungsten arc welding is almost exclusively employed for the root pass (Ref 22). This provides a controllable, high-quality root bead that dictates the final corrosion performance of the weld. The inert gas on the backside of the weld can also be more closely controlled with this process. Aside from this preference, process selection will probably be dictated more by the availability of consumable form and economic and logistic considerations than by desired properties for the particular welding process. Welds in DSS have been made by all the major fusion welding processes and have performed satisfactorily. However, in a very few cases, the final application may necessitate the stipulation of a particular welding process. Where exceptional low-temperature toughness is required, gasshielded processes may be specified because they produce higher weld metal toughness properties than flux-shielded processes (Fig. 14). Some consumable forms have not yet been fully developed for all grades of DSS. Nevertheless, gas-tungsten arc welding, shielded metal arc welding, submerged arc welding, flux-cored arc welding, and gas-metal arc welding are commonly used with success for most alloy classes.

Additional information is available in the article "Welding of Stainless Steels" in this Volume.

Solid-State Welding Processes. Few studies have been made of the solid-state welding characteristics of DSS. Explosive welding has been applied to overlaying of tube sheets and similar items (Ref 2). Corrosion and mechanical properties are presumed to be similar to parent material. Friction welding of DSS has shown that the corrosion resistance of a friction weld zone can be significantly better than that of a corresponding arc weld in the same material (Ref 34). This is attributed to the absence of fusion and the related effects that can result (that is, segregation and nitrogen loss).

The bond line will, however, experience high temperatures, and some degree of transformation of the austenite to ferrite would be anticipated. Very rapid cooling will then probably suppress reformation of austenite, and a structure high in ferrite could result. There is ongoing effort to establish whether solid-state welding processes offer significant advantages over fusion processes. The lack of a fused zone is potentially attractive to the welding of alloys that are prone to localized segregation effects.

References cited in this section

2. B. LARSSON AND B. LUNDQVIST, FABRICATING FERRITIC-AUSTENITIC STAINLESS STEELS, SANDVIK STEEL TRADE LITERATURE, PAMPHLET S-51-33-ENG, OCT 1987
22. J. HARSTON, E. HUTCHINS, AND S. SWEENEY, THE DEVELOPMENT AND CONSTRUCTION OF DUPLEX STAINLESS STEEL PIPELINES FOR USE OFFSHORE IN THE SOUTHERN NORTH SEA, *PROC. 3RD LNT. CONF. WELDING AND PERFORMANCE OF PIPELINES,* THE WELDING INSTITUTE, 1986
34. T.G. GOOCH, CORROSION RESISTANCE OF WELDS IN DUPLEX STAINLESS STEELS, *DUPLEX STAINLESS STEELS CONF. PROC.,* AMERICAN SOCIETY FOR METALS, 1983, P 325-346

Selection of Wrought Duplex Stainless Steels

David N. Noble, ARCO Alaska, Inc.

References

1. H.D. SOLOMON AND T.M. DEVINE, JR., DUPLEX STAINLESS STEELS--A TALE OF TWO PHASES, *DUPLEX STAINLESS STEELS CONF. PROC.,* AMERICAN SOCIETY FOR METALS, 1983
2. B. LARSSON AND B. LUNDQVIST, FABRICATING FERRITIC-AUSTENITIC STAINLESS STEELS, SANDVIK STEEL TRADE LITERATURE, PAMPHLET S-51-33-ENG, OCT 1987
3. S. BERNHARDSSON, THE CORROSION RESISTANCE OF DUPLEX STAINLESS STEELS, *DUPLEX STAINLESS STEELS CONF. PROC.,* LES EDITIONS DE PHYSIQUE, LES ULIS CEDEX, FRANCE, OCT 1991, P 185-210
4. L. VAN NASSAU, H. MEELKER, AND J. HILKES, WELDING DUPLEX AND SUPER-DUPLEX STAINLESS STEELS--A GUIDE FOR INDUSTRY, DOCUMENT 01463, LINCOLN NORWELD, JULY 1992
5. T.G. GOOCH, WELDABILITY OF DUPLEX FERRITIC-AUSTENITIC STAINLESS STEELS, *DUPLEX STAINLESS STEELS CONF. PROC.,* AMERICAN SOCIETY FOR METALS, 1983, P 573-602
6. J. CHARLES, SUPER DUPLEX STAINLESS STEEL: STRUCTURE AND PROPERTIES, *DUPLEX STAINLESS STEELS CONF. PROC.,* VOL 1, LES EDITIONS DE PHYSIQUE, LES ULIS CEDEX, FRANCE, OCT 1991, P 3-48
7. R. CASTRO AND J.J. DE CADENET, *WELDING METALLURGY OF STAINLESS AND HEAT-RESISTING STEELS,* CAMBRIDGE UNIVERSITY PRESS, 1974
8. D.J. KOTECKI AND T.A. SIEWERT, WRC-1992 CONSTITUTIONAL DIAGRAM FOR STAINLESS STEEL WELD METALS: A MODIFICATION OF THE WRC 1988 DIAGRAM, *WELD. J.,* MAY 1992, P 171S-178S
9. R.J. BRIGHAM AND E.W. TOZER, EFFECT OF ALLOYING ADDITIONS ON THE PITTING RESISTANCE OF 18% CR AUSTENITIC STAINLESS STEEL, *CORROSION,* VOL 30 (NO. 5), MAY 1974, P 161-166
10. N. SUUTALA AND M. KURKELA, *STAINLESS STEELS CONF. PROC.,* THE INSTITUTE OF METALS, 1985, P 240-247
11. R. DOLLING, V. NEUBERT, AND P. KNOLL, THE CORROSION BEHAVIOUR OF SUPER DUPLEX

STEEL CAST ALLOYS WITH A PREN>41, *DUPLEX STAINLESS STEELS CONF. PROC.,* VOL 2, LES EDITIONS DE PHYSIQUE, LES ULIS CEDEX, FRANCE, OCT 1991, P 1341-1351

12. H. ERIKSSON AND S. BERNHARDSSON, THE APPLICABILITY OF DUPLEX STAINLESS STEELS IN SOUR ENVIRONMENTS, *CORROSION,* VOL 47 (NO. 8), SEPT 1991, P 719-727
13. A. MLYASAKA AND H. OGAWA, CRITICAL STRESS FOR SCC OF CORROSION RESISTANT ALLOY IN SOUR ENVIRONMENTS, PAPER 5, *CORROSION '89,* NATIONAL ASSOCIATION OF CORROSION ENGINEERS, 1989
14. S. BERNHARDSSON, THE CORROSION RESISTANCE OF DUPLEX STAINLESS STEELS, *DUPLEX STAINLESS STEELS CONF. PROC.,* AMERICAN SOCIETY FOR METALS, 1983, P 185-210
15. S. AZUMA, H. TSUGE, T. KUDO, AND T. MOROISHI, CREVICE CORROSION OF DUPLEX STAINLESS STEELS IN SIMULATED SOUR GAS ENVIRONMENTS, PAPER 306, *CORROSION '87,* NATIONAL ASSOCIATION OF CORROSION ENGINEERS, 1987
16. "SULFIDE STRESS CRACKING RESISTANT METALLIC MATERIALS FOR OILFIELD EQUIPMENT," MR-01-75-93, NATIONAL ASSOCIATION OF CORROSION ENGINEERS
17. L. SMITH, CORROSION RESISTANT ALLOYS OFFER PIPELINE ALTERNATIVES, *STAINL. STEEL EUR.,* JAN/FEB 1993, P 32-42
18. M.J. SCHOFIELD, AND R.D. KANE, DEFINING SAFE USE LIMITS FOR DUPLEX STAINLESS STEELS, *DUPLEX STAINLESS STEEL CONF. PROC.,* LES EDITIONS DE PHYSIQUE, LES ULIS CEDEX, FRANCE, OCT 1991, P 241-255
19. T. OGAWA AND T. KOSEKI, EFFECT OF COMPOSITION PROFILES ON METALLURGY AND CORROSION BEHAVIOUR OF DUPLEX STAINLESS STEEL WELD METALS, *WELD. J.,* VOL 88 (NO. 5), MAY 1989, P 181S-191S
20. D. FRUYTIER, INDUSTRIAL EXPERIENCES WITH DUPLEX STAINLESS STEEL, *STAINL. STEEL EUR.,* VOL 3 (NO. 13), DEC 1991
21. B. WALLEN AND S. HENRIKSON, EFFECT OF CHLORINATION ON STAINLESS STEELS IN SEAWATER, PAPER 403, *CORROSION '86,* NATIONAL ASSOCIATION OF CORROSION ENGINEERS, 1986
22. J. HARSTON, E. HUTCHINS, AND S. SWEENEY, THE DEVELOPMENT AND CONSTRUCTION OF DUPLEX STAINLESS STEEL PIPELINES FOR USE OFFSHORE IN THE SOUTHERN NORTH SEA, *PROC. 3RD LNT. CONF. WELDING AND PERFORMANCE OF PIPELINES,* THE WELDING INSTITUTE, 1986
23. J.E. LIPPOLD AND W.A. BAESLACK, *MET. CONSTR.,* VOL 20 (NO. L), 1988, P 21
24. H. HOFFMEISTER AND R. MUNDT, *ARCH. EISENHUTT.,* VOL 52 (NO. 4), P 159-164
25. D.N. NOBLE AND T.G. GOOCH, THE EFFECTS OF HEAT INPUT AND CHEMICAL COMPOSITION ON MICROSTRUCTURE IN ARC WELDED UNS S31803 TYPE DUPLEX STAINLESS STEELS, *ASM MATERIALS WEEK '87,* ASM INTERNATIONAL, OCT 1987
26. B. JOSEFSSON, J.-O. NILSSON, AND A. WILSON, PHASE TRANSFORMATION IN DUPLEX STEELS AND THE RELATION BETWEEN CONTINUOUS COOLING AND ISOTHERMAL HEAT TREATMENTS, *DUPLEX STAINLESS STEELS CONF. PROC.,* AMERICAN SOCIETY FOR METALS, 1983
27. L. ITRURGOYEN, J. ALCALA, AND M. ANGLADA, THE INFLUENCE OF AGEING AT 475 C ON THE FRACTURE RESISTANCE OF A DUPLEX STAINLESS STEEL, *CORROSION,* VOL 47 (NO. 8), SEPT 1991, P 757-764
28. "SANDVIK SAF 2507--A HIGH PERFORMANCE DUPLEX STAINLESS STEEL," SANDVIK STEEL TRADE LITERATURE, PAMPHLET S-1875-ENG, SEPT 1987
29. F.R. COE, *WELDING STEELS WITHOUT HYDROGEN CRACKING,* TWI PUBLICATIONS, 1978
30. "HOW TO WELD 2205," AVESTA WELDING TRADE LITERATURE
31. "SEAMLESS AND WELDED FERRITIC/AUSTENITIC STAINLESS STEEL PIPE," ASTM A 790-91
32. "CHEMICAL PLANT AND PETROLEUM REFINERY PIPING," ASME CODE B31.3, P 66-67

33. A.W. STEPHENSON, P.C. GOUGH, AND J.C.M. FARRAR, "THE WELDABILITY OF SUPER DUPLEX ALLOYS--WELDING CONSUMABLES AND PROCEDURE DEVELOPMENT FOR ZERON 100, *PROC. LNT. INST. WELDING ANNUAL ASSEMBLY CONFERENCE,* JULY 1991
34. T.G. GOOCH, CORROSION RESISTANCE OF WELDS IN DUPLEX STAINLESS STEELS, *DUPLEX STAINLESS STEELS CONF. PROC.,* AMERICAN SOCIETY FOR METALS, 1983, P 325-346
35. O. JONSSON, M. LILJAS, AND P. STENVALL, THE ROLE OF NITROGEN IN LONGITUDINAL WELDING OF TUBING IN DUPLEX STAINLESS STEELS, *DUPLEX STAINLESS STEELS CONF. PROC.,* AMERICAN SOCIETY FOR METALS, 1983, P 461-468
36. "WELDING AND BRAZING QUALIFICATIONS," ASME 1989 BOILER AND PRESSURE VESSEL CODE SECTION IX
37. "WELDING OF PIPELINES AND RELATED FACILITIES," STANDARD 1104, AMERICAN PETROLEUM INSTITUTE
38. "SPECIFICATION FOR CRA LINEPIPE," STANDARD 5LC, AMERICAN PETROLEUM INSTITUTE
39. "STANDARD PROCEDURES FOR CALIBRATING MAGNETIC INSTRUMENTS TO MEASURE THE DELTA FERRITE CONTENT OF AUSTENITIC STAINLESS STEEL WELD METALS," AWS A4.2-74, AMERICAN WELDING SOCIETY
40. D.J. KOTECKI, EXTENSION OF THE WRC FERRITE NUMBER SYSTEM, *WELD. J.,* NOV 1982, P 352S-361S
41. "TESTING FOR PITTING AND CREVICE CORROSION RESISTANCE OF STAINLESS STEELS AND RELATED ALLOYS BY THE USE OF FERRIC CHLORIDE SOLUTION," G 48 PRACTICE A, ASTM
42. R.M. DAVISON AND J.D. REDMOND, DEVELOPMENT OF QUALIFICATION TESTS FOR DUPLEX STAINLESS STEEL MILL PRODUCTS, PAPER 302, *CORROSION '91,* NATIONAL ASSOCIATION OF CORROSION ENGINEERS, 1991
43. R.A. CORBETT, PROBLEMS IN UTILIZING ASTM G 48 TO EVALUATE HIGH ALLOY STAINLESS STEELS, PAPER 298, *CORROSION '92,* NATIONAL ASSOCIATION OF CORROSION ENGINEERS, 1992
44. D.J. KOTECKI, HEAT TREATMENT OF DUPLEX STAINLESS STEEL WELD METALS, *WELD. J.,* NOV 1989, P 431S-441S
45. L. ODEGARG AND S.-A. FAGER, CHOICE OF RIGHT FILLER METAL FOR JOINING SANDVIK SAF 2507 TO A SUPER AUSTENITIC STAINLESS 6 MO STEEL, *ASM MATERIALS WEEK '87,* ASM INTERNATIONAL, OCT 1987, P 441-449
46. R.N. GUNN, THE INFLUENCE OF COMPOSITION AND MICROSTRUCTURE ON THE CORROSION BEHAVIOUR OF COMMERCIAL DUPLEX ALLOYS, *RECENT DEVELOPMENTS IN THE JOINING OF STAINLESS STEELS AND HIGH ALLOYS,* EDISON WELDING INSTITUTE, COLUMBUS, OHIO, OCT 1992
47. G.A. CLARK AND P. GUHA, *BR. WELD. J.,* VOL 13 (NO. 5), MAY 1981, P 269-273
48. J. HONEYCOMBE AND T.G. GOOCH, *WELD. J.,* VOL 56 (NO. 11), NOV 1977, P 339S-353S
49. P. SENTANCE, THE BRAE FIELD, *STAINL. STEEL EUR.,* VOL 4 (NO. 20), SEPT 1992, P 38-41
50. C.D. LUNDIN, K. KIKUCHI, AND K.K. KAHN, PHASE 1 REPORT; "MEASUREMENT OF DIFFUSIBLE HYDROGEN CONTENT AND HYDROGEN EFFECTS ON THE CRACKING POTENTIAL OF DUPLEX STAINLESS STEEL WELDMENTS," THE WELDING RESEARCH COUNCIL, MARCH 1991
51. R.A. WALKER AND T.G. GOOCH, HYDROGEN CRACKING OF WELDS IN DUPLEX STAINLESS STEEL, *CORROSION,* VOL 47 (NO. 8), SEPT 1991, P 1053-1063
52. I. VAROL, W.A. BAESLACK, AND J.C. LIPPOLD, *METALLOGRAPHY,* VOL 23 (NO. 1), 1989
53. J.C. LIPPOLD, W.A. BAESLACK, AND I. VAROL, WELDING OF DUPLEX STAINLESS STEELS, *WELD. J.,* IN PRESS
54. R.A. WALKER AND T.G. GOOCH, PITTING RESISTANCE OF WELD METAL FOR 22CR-5NI FERRITIC-AUSTENITIC STAINLESS STEEL, *BR. CORROS. J.,* VOL 26 (NO. 1), 1991, P 51-59

55. L. ODEGARD AND S.-A. FAGER, THE ROOT SIDE PITTING CORROSION RESISTANCE OF STAINLESS STEEL WELDS, *SANDVIK STEEL WELD. REP.*, NO. 1, 1990
56. T.G. GOOCH, *WELDING IN THE WORLD,* VOL 24 (NO. 7/8), 1986, P 148-167
57. J. CHARLES, SUPER DUPLEX STAINLESS STEEL: STRUCTURE AND PROPERTIES, *DUPLEX STAINLESS STEELS CONF. PROC.,* AMERICAN SOCIETY FOR METALS, 1983, P 151-168
58. P.C. GOUGH AND J.C.M. FARRAR, FACTORS AFFECTING WELD ROOT RUN CORROSION PERFORMANCE IN DUPLEX AND SUPER DUPLEX PIPEWORK, *ASM MATERIALS WEEK '87,* ASM INTERNATIONAL, OCT 1987, P 1009-1025

Selection of Wrought Precipitation-Hardening Stainless Steels

B. Pollard

Introduction

PRECIPITATION-HARDENING (PH) STAINLESS STEELS are iron-chromium-nickel alloys with corrosion resistance superior to that of the hardenable 400 series stainless steels and yield strengths of 585 to 1795 MPa (85 to 260 ksi). These high strengths are obtained by precipitation hardening a martensitic or austenitic matrix with one or more of the following elements: copper, aluminum, titanium, niobium (columbium), and molybdenum. Precipitation-hardening steels can be grouped into three types--martensitic, semiaustenitic, and austenitic--based on their martensite start and finish (M_s and M_f) temperatures and resultant behavior upon cooling from a suitable solution treatment temperature.

The martensitic PH steels, such as 17-4PH, have M_f temperatures just above room temperature so that they transform completely to martensite upon air cooling from the solution treatment temperature and thus are martensitic in the annealed condition. Hardening is accomplished by a single aging treatment of 1 to 4 h at 480 to 620 °C (900 to 1150 °F).

The semiaustenitic PH steels, such as 17-7PH, have compositions that are balanced so that their M_s temperatures are well below room temperature. Therefore, they are predominantly austenitic upon cooling from the solution treatment temperature and are highly ductile and readily formed in that condition. After forming, transformation to martensite is accomplished by a conditioning treatment, which raises their M_s and M_f temperatures by precipitating carbon and alloying elements from solution. If a low conditioning temperature is used (730 to 760 °C, or 1350 to 1400 °F), the M_f temperature is raised to the vicinity of room temperature and transformation to martensite is complete upon cooling. If a high conditioning temperature (930 to 955 °C, or 1710 to 1750 °F) is used, less carbon is precipitated, the M_f temperature remains below zero, and refrigeration is required to accomplish transformation to martensite. However, since the martensite produced in this manner contains more carbon, it is of higher strength than that produced by transformation at lower temperatures. Transformation may also be accomplished by cold working. In all cases, the martensite structure is then hardened by aging at a temperature in the range of 455 to 565 °C (850 to 1050 °F) for 1 to 3 h.

The austenitic PH steels, such as A-286, have M_s temperatures so low that they cannot be transformed to martensite. Strengthening is obtained by the precipitation of intermetallic compounds in an austenitic matrix.

The chemical compositions of the PH steels presently produced in North America are presented in Table 1. A detailed treatment of aging in PH stainless steels is beyond the scope of this article because many different precipitation hardening phases are involved. Table 2, which was compiled from Ref 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, and 15, lists the phases produced in martensitic, semiaustenitic, and austenitic PH grades. However, certain aging characteristics are common to most PH stainless steels. For martensitic and semiaustenitic types, maximum strengthening is obtained by aging at 455 to 510 °C (850 to 950 °F). Higher temperatures increase ductility and toughness but reduce both the maximum strength level and the time required to attain it. Typical aging curves for selected martensitic and semiaustenitic PH steels aged at 480 °C (900 °F) and 510 °C (950 °F) are shown in Fig. 1(a) and 1(b), respectively (Ref 16 17, 18, 19, 20).

TABLE 1 COMPOSITIONS OF PRECIPITATION-HARDENING STAINLESS STEELS

ALLOY	**UNS NO.**	**COMPOSITION, %[A]**								
		C	**Mn**	**Si**	**Cr**	**Ni**	**Mo**	**P**	**S**	**OTHER**
MARTENSITIC TYPES										
PH13-8 MO	S13800	0.05	0.10	0.10	12.25-13.25	7.5-8.5	2.0-2.5	0.01	0.008	0.90-1.35 AL; 0.01 N
15-5PH	S15500	0.07	1.00	1.00	14.0-15.5	3.5-5.5	. . .	0.04	0.03	2.5-4.5 CU; 0.15-0.45 NB
17-4PH	S17400	0.07	1.00	1.00	15.0-17.5	3.0-5.0	. . .	0.04	0.03	3.0-5.0 CU; 0.15-0.45 NB

CUSTOM 450	S45000	0.05	1.00	1.00	14.0-16.0	5.0-7.0	0.5-1.0	0.03	0.03	1.25-1.75 CU; 8 × %C MIN NB
CUSTOM 455	S45500	0.05	0.50	0.50	11.0-12.5	7.5-9.5	0.50	0.04	0.03	1.5-2.5 CU; 0.8-1.4 TI; 0.1-0.5 NB
SEMIAUSTENITIC TYPES										
PH15-7 MO	S15700	0.09	1.00	1.00	14.0-16.0	6.50-7.75	2.0-3.0	0.04	0.04	0.75-1.50 AL
17-7PH	SI7700	0.09	1.00	1.00	16.0-18.0	6.50-7.75	. . .	0.04	0.04	0.75-1.50 AL
AM-350	S35000	0.07-0.11	0.50-1.25	0.50	16.0-17.0	4.0-5.0	2.50-3.25	0.04	0.03	0.07-0.13 N
AM-355	S35500	0.10-0.15	0.50-1.25	0.50	15.0-16.0	4.0-5.0	2.50-3.25	0.04	0.03	0.07-0.13 N
AUSTENITIC TYPES										
A-286	S66286	0.08	2.00	1.00	13.5-16.0	24.0-27.0	1.0-1.5	0.025	0.025	1.90-2.35 TI; 0.35 MAX AL; 0.10-0.50 V; 0.0030-0.0100 B
JBK-75(B)	. . .	0.015	0.05	0.02	14.5	29.5	1.25	0.006	0.002	2.15 TI; 0.25 AL; 0.27 V; 0.0015 B

(A) SINGLE VALUES ARE MAXIMUM VALUES UNLESS OTHERWISE INDICATED.

(B) TYPICAL VALUES

TABLE 2 PRECIPITATION-HARDENING PHASES IN PH STEELS

ALLOY	**PRECIPITATION-HARDENING PHASE**	**REF**
MARTENSITIC TYPES		
PH13-8 MO	COHERENT NIAL + FINE γ PARTICLES	1
15-5PH	FACE-CENTERED CUBIC CU-RICH PHASE(A)	
17-4PH	FACE-CENTERED CUBIC CU-RICH PHASE	2, 3
CUSTOM 450	LAVES-TYPE PHASE CONTAINING FE, NB, MO	4, 5
CUSTOM 455	HEXAGONAL CLOSE-PACKED ORDERED COHERENT NI_3TI	6
SEMIAUSTENITIC TYPES		
17-7PH	ORDERED BODY-CENTERED CUBIC	7, 8, AND 9
PH15-7 MO	β-NIAL + NI_3AL	10, 11
AM-350	CR_2N	12
AM-355	CR_2N	12
AUSTENITIC TYPES		
A-286	NI_3 (AL,TI)	13, 14
JBK-75	NI_3 (AL,TI)	15

(A) ASSUMED TO BE THE SAME PRECIPITATE AS FORMED IN 17-4PH

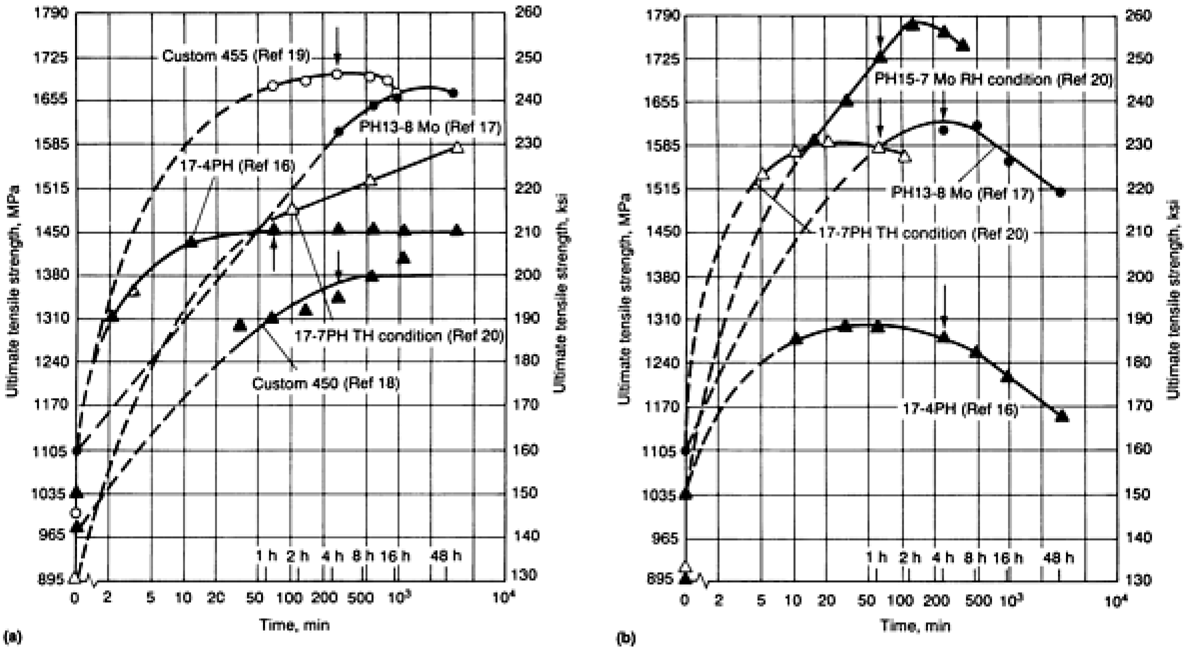

FIG. 1 AGING CURVES FOR SELECTED PH STAINLESS STEELS. (A) AGED AT 480 °C (900 °F). (B) AGED AT 510 °C (950 °F). ARROWS INDICATE STANDARD AGING TIME FOR EACH STEEL

For any alloy, a wide range of properties can be obtained by varying the heat treatment practice. However, certain standard heat treatments for which mechanical property data are available am generally used. These are summarized in Table 3. The corresponding mechanical properties are presented in Table 4.

TABLE 3 HEAT TREATMENT OF PRECIPITATION-HARDENING STAINLESS STEELS

ALLOY	CONDITION CODE	SOLUTION ANNEAL	CONDITIONING AND HARDENING TREATMENT	AGING AND/OR TEMPERING TREATMENT
MARTENSITIC TYPES				
PH13-8 MO	A	925 °C (1700 °F) FOR 15-30 MIN; OIL OR AIR COOL (AC) BELOW 15 °C (60 °F)	. . .	. . .
	RH950	925 °C (1700 °F) FOR 15-30 MIN; OIL OR AC BELOW 15 °C (60 °F)	REFRIGERATE AT -75 °C (-100 °F) FOR 8 H	AGE 510 °C (950 °F) FOR 4 H; AC
	H950	925 °C (1700 °F) FOR 15-30 MIN; OIL OR AC BELOW 15 °C (60 °F)	. . .	AGE 510 °C (950 °F) FOR 4 H; AC
	H1000	925 °C (1700 °F) FOR 15-30 MIN; OIL OR AC BELOW 15 °C (60 °F)	. . .	AGE 535 °C (1000 °F) FOR 4 H; AC
	H1050	925 °C (1700 °F) FOR 15-30 MIN; OIL OR AC BELOW 15 °C (60 °F)	. . .	AGE 565 °C (1050 °F) FOR 4 H; AC
	H1100	925 °C (1700 °F) FOR 15-30 MIN; OIL OR AC BELOW 15 °C (60 °F)	. . .	AGE 595 °C (1100 °F) FOR 4 H; AC
	H1150	925 °C (1700 °F) FOR 15-30 RAIN; OIL OR AC BELOW 15 °C (60 °F)	. . .	AGE 620 °C (1150 °F) FOR 4 H; AC
15-5PH	A	1035 °C (1900 °F) FOR 30	. . .	. . .

		MIN; OIL OR AC BELOW 30 °C (90 °F)		
	H900	1035 °C (1900 °F) FOR 30 MIN; OIL OR AC BELOW 30 °C (90 °F)	. . .	AGE 480 °C (900 °F) FOR 1 H; AC
	H925	1035 °C (1900 °F) FOR 30 MIN; OIL OR AC BELOW 30 °C (90 °F)	. . .	AGE 495 °C (925 °F) FOR 4 H; AC
	H1025	1035 °C (1900 °F) FOR 30 MIN; OIL OR AC BELOW 30 °C (90 °F)	. . .	AGE 550 °C (1025 °F) FOR 4 H; AC
	H1075	1035 °C (1900 °F) FOR 30 MIN; OIL OR AC BELOW 30 °C (90 °F)	. . .	AGE 580 °C (1075 °F) FOR 4 H; AC
	H1150	1035 °C (1900 °F) FOR 30 MIN; OIL OR AC BELOW 30 °C (90 °F)	. . .	AGE 620 °C (1150 °F) FOR 4 H; AC
17-4PH	A	1035 °C (1900 °F) FOR 30 MIN; OIL OR AC BELOW 30 °C (90 °F)	. . .	. . .
	H900	1035 °C (1900 °F) FOR 30 MIN; OIL OR AC BELOW 30 °C (90 °F)	. . .	AGE 480 °C (900 °F) FOR 1 H; AC
	H925	1035 °C (1900 °F) FOR 30 MIN; OIL OR AC BELOW 30 °C (90 °F)	. . .	AGE 495 °C (925 °F) FOR 4 H; AC
	H1025	1035 °C (1900 °F) FOR 30 RAIN; OIL OR AC BELOW 30 °C (90 °F)	. . .	AGE 550 °C (1025 °F) FOR 4 H; AC
	H1075	1035 °C (1900 °F) FOR 30 RAIN; OIL OR AC BELOW 30 °C (90 °F)	. . .	AGE 580 °C (1075 °F) FOR 4 H; AC
	H1100	1035 °C (1900 °F) FOR 30 MIN; OIL OR AC BELOW 30 °C (90 °F)	. . .	AGE 595 °C (1100 °F) FOR 4 H; AC
	H1150	1035 °C (1900 °F) FOR 30 MIN; OIL OR AC BELOW 30 °C (90 °F)	. . .	AGE 620 °C (1150 °F) FOR 4 H; AC
CUSTOM 450	A	1035 °C (1900 °F) FOR 1 H; WATER QUENCH	. . .	. . .
	H900	1035 °C (1900 °F) FOR 1 H; WATER QUENCH	. . .	AGE 480 °C (900 °F) FOR 4 H; AC
	H1000	1035 °C (1900 °F) FOR 1 H; WATER QUENCH	. . .	AGE 535 °C (1000 °F) FOR 4 H; AC
	H1150	1035 °C (1900 °F) FOR 1 H; WATER QUENCH	. . .	AGE 620 °C (1150 °F) FOR 4 H; AC
CUSTOM 455	A	830 °C (1525 °F) FOR 1 H; WATER QUENCH	. . .	. . .
	H900	830 °C (1525 °F) FOR 1 H; WATER QUENCH	. . .	AGE 480 °C (900 °F) FOR 4 H; AC
	H950	830 °C (1525 °F) FOR 1 H; WATER QUENCH	. . .	AGE 510 °C (950 °F) FOR 4 H; AC
	H1000	830 °C (1525 °F) FOR 1 H; WATER QUENCH	. . .	AGE 535 °C (1000 °F) FOR 4 H; AC
SEMIAUSTENITIC TYPES				
17-7PH AND PH15-7 MO	A	1065 °C (1950 °F) FOR 30 MIN/25 MM (1 IN.) OF THICKNESS; AC	. . .	. . .
	T	1065 °C (1950 °F) FOR 30 MIN/25 MM (1 IN.) OF THICKNESS; AC	760 °C (1400 °F) $1\frac{1}{2}$ H; AC TO 15 °C (60 °F) FOR 30 MIN	. . .
	TH1050	1065 °C (1950 °F) FOR 30 MIN/25 MM (1 IN.) OF	760 °C (1400 °F) $1\frac{1}{2}$ H; AC TO 15 °C (60	AGE 565 °C (1050

		THICKNESS; AC	°F) FOR 30 MIN	°F) FOR $1\frac{1}{2}$ H; AC
	A1750	1065 °C (1950 °F) FOR 30 MIN/25 MM (1 IN.) OF THICKNESS; AC	955 °C (1750 °F) FOR 10 MIN; AC	. . .
	R100	1065 °C (1950 °F) FOR 30 MIN/25 MM (1 IN.) OF THICKNESS; AC	955 °C (1750 °F) FOR 10 MIN; AC; REFRIGERATE AT -75 °C (-100 °F) FOR 8 H	. . .
	RH950	1065 °C (1950 °F) FOR 30 MIN/25 MM (1 IN.) OF THICKNESS; AC	955 °C (1750 °F) FOR 10 MIN; AC; REFRIGERATE AT -75 °C (-100 °F) FOR 8 H	AGE 510 °C (950 °F) FOR 1 H; AC
	C	1065 °C (1950 °F) FOR 30 MIN/25 MM (1 IN.) OF THICKNESS; AC	COLD REDUCE	. . .
	CH900	1065 °C (1950 °F) FOR 30 MIN/25 MM (1 IN.) OF THICKNESS; AC	COLD REDUCE	AGE 480 °C (900 °F) FOR 1 H; AC
AM-350	H	1065 °C (1950 °F) FOR 90 MIN/25 MM (1 IN.) OF THICKNESS; RAPID COOL	. . .	. . .
	L	1065 °C (1950 °F) FOR 90 MIN/25 MM (1 IN.) OF THICKNESS; RAPID COOL	930 °C (1710 °F) FOR 90 MIN/25 MM (1 IN.) OF THICKNESS; AC	. . .
	SCT850	1065 °C (1950 °F) FOR 90 MIN/25 MM (1 IN.) OF THICKNESS; RAPID COOL	930 °C (1710 °F) FOR 90 MIN/25 MM (1 IN.) OF THICKNESS; AC; REFRIGERATE AT -75 °C (-100 °F) FOR 3 H	TEMPER 455 °C (850 °F) FOR 3 H; AC
	SCT1000	1065 °C (1950 °F) FOR 90 MIN/25 MM (1 IN.) OF THICKNESS; RAPID COOL	930 °C (1710 °F) FOR 90 MIN/25 MM (1 IN.) OF THICKNESS; AC; REFRIGERATE AT -75 °C (-100 °F) FOR 3 H	TEMPER 535 °C (1000 °F) FOR 3 H; AC
AM-355	H	1025 °C (1875 °F) FOR 90 MIN/25 MM (1 IN.) OF THICKNESS; RAPID COOL	. . .	. . .
	L	1025 °C (1875 °F) FOR 90 MIN/25 MM (1 IN.) OF THICKNESS; RAPID COOL	(1710 °F) FOR 90 MIN/25 MM (1 IN.) OF THICKNESS; AC(A)	. . .
	SCT850	1025 °C (1875 °F) FOR 90 MIN/25 MM (1 IN.) OF THICKNESS; RAPID COOL	REFRIGERATE AT -75 °C (-100 °F); 930 °C (1710 °F) FOR 90 MIN/25 MM (1 IN.) OF THICKNESS; AC; REFRIGERATE AT -75 °C (-100 °F) FOR 3 H	TEMPER 455 °C (850 °F) FOR 3 H; AC
	SCT1000	1025 °C (1875 °F) FOR 90 MIN/25 MM (1 IN.) OF THICKNESS; RAPID COOL	REFRIGERATE AT -75 °C (-100 °F); 930 °C (1710 °F) FOR 90 MIN/25 MM (1 IN.) OF THICKNESS; AC; REFRIGERATE AT -75 °C (-100 °F) FOR 3 H	TEMPER 535 °C (1000 °F) FOR 3 H; AC
AUSTENITIC TYPE				
A-286	ST1650	900 °C (1650 °F) FOR 2 H; OIL OR WATER QUENCH	. . .	. . .
	ST1650A	900 °C (1650 °F) FOR 2 H; OIL OR WATER QUENCH	. . .	AGE 730 °C (1350 °F) FOR 16 H; AC
	ST1800	980 °C (1800 °F) FOR 1 H; OIL OR WATER QUENCH	. . .	. . .
	ST1800A	980 °C (1800 °F) FOR 1 H; OIL OR WATER QUENCH	. . .	AGE 730 °C (1350 °F) FOR 16 H; AC

(A) FOR BARS AND FORGING BILLETS, EMPLOY AN EQUALIZING TREATMENT OF 3 H AT 745-800 °C (1375-1475 °F) + 3 H AT 535-620 °C (1000-1150 °F) (CONDITION E + OT) BEFORE THE RE-SOLUTION TREATMENT AT 930 °C (1710 °F).

TABLE 4 MINIMUM MECHANICAL PROPERTIES OF PRECIPITATION-HARDENING STAINLESS STEELS PER SPECIFICATION NOTED

ALLOY AND PRODUCT FORM(A)	CONDITION	TENSILE STRENGTH		YIELD STRENGTH		ELONGATION, %	REDUCTION IN AREA, %	HARDNESS, HRC		ASTM SPECIFICATION
		MPA	KSI	MPA	KSI			MIN	MAX	
MARTENSITIC TYPES										
PH13-8 MO (UNS S13800)										
B, F	H950	1520	220	1410	205	10	45; 35(B)	45	. . .	A 564, A 705
B, F	H1000	1410	205	1310	190	10	50; 40(B)	43	. . .	A 564, A 705
B, F	H1025	1275	185	1210	175	11	50; 45(B)	41	. . .	A 564, A 705
B, F	H1050	1210	175	1140	165	12	50; 45(B)	40	. . .	A 564, A 705
B, F	H1100	1030	150	930	135	14	50	34	. . .	A 564, A 705
B, F	H1150	930	135	620	90	14	50	30	. . .	A 564, A 705
B, F	H1150M	860	125	585	85	16	55	26	. . .	A 564, A 705
P, SH, ST	H950	1520	220	1410	205	6-10(C)	. . .	45	. . .	A 693
P, SH, ST	H1000	1380	200	1310	190	6-10(C)	. . .	43	. . .	A 693
15-5PH (UNS S15500)										
B, F	H900	1310	190	1170	170	10; 6(B)	35; 15(B)	40	. . .	A 564, A 705
B, F	H925	1170	170	1070	155	10; 7(B)	38; 20(B)	38	. . .	A 564, A 705
B, F	H1025	1070	155	1000	145	12; 8(B)	45; 27(B)	35	. . .	A 564, A 705
B, F	H1075	1000	145	860	125	13; 9(B)	45; 28(B)	32	. . .	A 564, A 705
B, F	H1100	965	140	795	115	14; 10(B)	45; 29(B)	31	. . .	A 564, A 705
B, F	H1150	930	135	725	105	16; 11(B)	50; 30(B)	28	. . .	A 564, A 705
B, F	H1150M	795	115	515	75	18; 14(B)	55; 35(B)	24	. . .	A 564, A 705
P, SH, ST	H900	1310	190	1170	170	5-10(C)	. . .	40	48	A 693
P, SH, ST	H1100	965	140	790	115	5-14(C)	. . .	29	40	A 693
17-4PH (UNS S17400)										
B, F	H900(D)	1310	190	1170	170	10	40; 35(E)	40	. . .	A 564, A 705
B, F	H925(D)	1170	170	1070	155	10	44; 38(E)	38	. . .	A 564, A 705
B, F	H1025(D)	1070	155	1000	145	12	45	35	. . .	A 564, A 705
B, F	H1075(D)	1000	145	860	125	13	45	32	. . .	A 564, A 705
B, F	H1100(D)	965	140	795	115	14	45	31	. . .	A 564, A 705
B, F	H1150(D)	930	135	725	105	16	50	28	. . .	A 564, A 705
B, F	H1150M(D)	795	115	515	75	18	55	24	. . .	A 564, A 705
P, SH, ST	H900	1310	190	1170	170	5-10(C)	. . .	40	48	A 693
P, SH, ST	H1100	965	140	790	115	5-14(C)	. . .	29	40	A 693

CUSTOM 450 (UNS S45000)										
B, SHAPES	ANNEALED	895(F)	130(F)	655	95	10	40	. . .	32	A 564(F)
F, SHAPES	ANNEALED	860(F)	125(F)	655	95	10	40	. . .	33	A 705(F)
B, F, SHAPES	H900	1240(G)	180(G)	1170	170	6; 10(B)	20; 40(B)	39	. . .	A 564(G), A 705(G)
B, F, SHAPES	H950	1170(G)	170(G)	1100	160	7; 10(B)	22; 40(B)	37	. . .	A 564(G), A 705(G)
B, F, SHAPES	H1000	1100	160(G)	1030	150	8; 12(B)	27; 45(B)	36	. . .	A 564(G), A 705
B, F, SHAPES	H1025	1030(G)	150(G)	965	140	12	45	34	. . .	A 564(G), A 705
B, F, SHAPES	H1050	1000(G)	145(G)	930	135	9; 12(B)	30; 45(B)	34	. . .	A 564(G), A 705
B, F, SHAPES	H1100	895(G)	130(G)	725	105	11; 16(B)	30; 50(B)	30	. . .	A 564(G), A 705
B, F, SHAPES	H1150	860(G)	125(G)	515	75	12-18(H)	35-55(H)	26	. . .	A 564(G), A 705
P, SH, ST	ANNEALED	895-1205	130-165	620-1035	90-150	4 MIN	. . .	25	33	A 693
P, SH, ST	H900	1240	180	1170	170	3-5(C)	. . .	40	. . .	A 693
P, SH, ST	H1000	1105	160	1035	150	5-7(C)	. . .	36	. . .	A 693
P, SH, ST	H1150	860	125	515	75	8-10(C)	. . .	26	. . .	A 693
MARTENSITIC TYPES										
CUSTOM 455 (UNS S45500)										
B, F, SHAPES	H900(I)	1620	235	1520	220	8	30	47	. . .	A 564(G), A 705(G)
B, F, SHAPES	H950(I)	1520	220	1410	205	10	40	44	. . .	A 564(G), A 705(G)
B, F, SHAPES	H1000(I)	1410	205	1280	185	10	40	40	. . .	A 564(G), A 705(G)
P, SH, ST	H950	1530	222	1410	205	≤ 4	. . .	44		A 693
SEMIAUSTENITIC TYPES										
PH15-7 MO (UNS S15700)										
B, F	RH950	1380	200	1210	175	7	25	. . .	. . .	A 564, A 705
B, F	TH1050	1240	180	1100	160	8	25	. . .	. . .	A 564, A 705

P, SH, ST	ANNEALED	1035 MAX	150 MAX	450 MAX	65 MAX	25 MIN	. . .	. . .	. . .	A 693
P, SH, ST	RH950(D)	1550	225	1380	200	1-4(C)	. . .	45-46	. . .	A 693
P, SH, ST	TH1050(D)	1310	190	1170	170	2-5(C)	. . .	40	. . .	A 693
P, SH, ST	COLD ROLLED CONDITION C	1380	200	1210	175	1	. . .	41	. . .	A 693
P, SH, ST	CH900	1650	240	1590	230	1	. . .	46	. . .	A 693
17-7PH (UNS S17700)										
B, F	RH950(D)	1275	185	1030	150	6	10	41	. . .	A 564, A 705
B, F	TH1050(D)	1170	170	965	140	6	25	38	. . .	A 564, A 705
P, SH, ST	RH950	1450(C)	210(C)	1310(C)	190(C)	1-6(C)	. . .	43(C)	44(C)	A 693
P, SH, ST	TH1050	1240(C)	180(C)	1030(C)	150(C)	3-7(C)	. . .	38	. . .	A 693
P, SH, ST	COLD ROLLED CONDITION C	1380	200	1210	175	1	. . .	41	. . .	A 693
P, SH, ST	CH900	1650	240	1590	230	1	. . .	46	. . .	A 693
W	COLD DRAWN CONDITION C	1400-2035(C)	203-295(C)	. . .	. . .	. . .	. . .	. . .	. . .	A 313
W	CH900	1585-2515(C)	230-365(C)	. . .	. . .	. . .	. . .	. . .	. . .	A 313
AM-350 (UNS S35000)										
P, SH, ST	ANNEALED	1380 MAX	200 MAX	585-620 MAX(C)	85-90 MAX(C)	8-12(C)	. . .	. . .	30	A 693
P, SH, ST	SCT850	1275	185	1030	150	2-8(C)	. . .	42	. . .	A 693
P, SH, ST	SCT1000	1140	165	1000	145	2-8(C)	. . .	36	. . .	A 693
AM-355 (UNS 35500)										
F	SCT1000	1170	170	1070	155	12	25	37	. . .	A 705
P, SH, ST	SCT850	1310	190	1140	165	10	. . .	. . .	. . .	A 693
P, SH, ST	SCT1000	1170	170	1030	150	12	. . .	37	. . .	A 693
AUSTENITIC TYPE										
A-286 (UNS S66286)										
B, F	ST 1650	724 MAX	105 MAX	. . .	. . .	. . .	. . .	. . .	201 HB	AMS NO. 5734, 5737
B, F	ST 1650A	965	140	655	95	12	15	277 HB	363 HB	AMS NO. 5734, 5737
B, F	ST 1800	724 MAX	105 MAX	. . .	. . .	. . .	. . .	. . .	201 HB	AMS NO. 5731, 5732
B, F	ST 1800A	895	130	585	85	15	20	248	341	AMS NO. 5731,

								HB	HB	5732
P, SH, ST	ST 1800	724 MAX	105 MAX	. . .	. . .	10-25	. . .	. . .	90 HRB	AMS NO. 5525, 5858
P, SH, ST	ST 1800A	862-965	125-140	655	95	4-15	. . .	24	35	AMS NO. 5525, 5858

(A) B, BAR; F, FORGINGS, P, PLATE; SH, SHEET; ST, STRIP; W, WIRE.

(B) HIGHER VALUE IS LONGITUDINAL; LOWER VALUE IS TRANSVERSE.

(C) VALUES VARY WITH THICKNESS OR DIAMETER.

(D) LONGITUDINAL PROPERTIES ONLY.

(E) HIGHER VALUES ARE FOR SIZES UP TO AND INCLUDING 75 MM (3 IN.); LOWER VALUES ARE FOR SIZES OVER 75 MM (3 IN.) UP TO AND INCLUDING 200 MM (8 IN.).

(F) TENSILE STRENGTHS OF 860 TO 140 MPA (125 TO 165 KSI) FOR SIZES UP TO 13 MM ($\frac{1}{2}$ IN.).

(G) TENSILE STRENGTH ONLY APPLICABLE UP TO SIZES OF 13 MM ($\frac{1}{2}$ IN.).

(H) VARIES WITH SECTION SIZE AND TEST DIRECTION.

(I) UP TO AND INCLUDING 150 MM (6 IN.)

The standard aging times used for aging temperatures of 480 and 510 °C (900 and 950 °F) are such that most steels so heat treated are in the fully aged or slightly overaged condition, whichever gives the best combination of strength and ductility. PH 15-7 Mo stainless steel is unusual in that it has the best combination of strength and ductility when slightly underaged. Material aged at temperatures above 510 °C (950 °F) and, therefore, in a significantly overaged condition, is used when greater ductility or toughness is required. Precipitation in austenitic PH steels is markedly slower than in martensitic or semiaustenitic PH steels. For example, 16 h at 720 °C (1325 °F) is required to produce near-maximum hardening in A-286.

The problems likely to be encountered in welding PH steels are different for each group, but similar for different steels within a group. Therefore, the weldability of the steels within each group can be best understood by first considering the welding of a typical steel from that group, which is invariably also the steel for which the most welding data are available. The weldability of other steels within the same group can then be understood by comparison with the "typical" steel.

References

1. V. SEETHARAMAN ET AL., PRECIPITATION HARDENING IN A PH13-8 MO STAINLESS STEEL, *MATER. SCI. ENG.,* VOL 47, JAN 1981, P 1-11
2. H.J. RACK AND D. KALISH, THE STRENGTH, FRACTURE TOUGHNESS AND LOW CYCLE FATIGUE BEHAVIOR OF 17-4PH STAINLESS STEEL, *METALL. TRANS. A,* VOL 5, JULY 1974, P 1595-1605
3. K.C. ANTHONY, AGING REACTIONS IN PRECIPITATION HARDENABLE STAINLESS STEEL, *J. MET.,* VOL 15, DEC 1963, P 922-927
4. M. HENTHORNE, T.A. DEBOLD, AND R.J. YINGER, "CUSTOM 450-A NEW HIGHER STRENGTH STAINLESS STEEL," PAPER NO. 53, CORROSION '72, NATIONAL ASSOCIATION OF CORROSION ENGINEERS CONFERENCE, 1972
5. J.E. MCBRIDE, JR., AND G.N. MANIAR, *METALLOGRAPHY AS A QUALITY CONTROL TOOL,* PLENUM PUBLISHING, 1980, P 279
6. S. WIDGE, "A STUDY OF THE EFFECTS OF VARYING SOLUTION TREATMENT ON THE TOUGHNESS OF A STAINLESS MARAGING ALLOY," PH.D. DISSERTATION, LEHIGH UNIVERSITY, 1984
7. E.E. UNDERWOOD, A.E. AUSTIN, AND G.K. MANNING, THE MECHANISM OF HARDENING IN 17-7 NI-CR PRECIPITATION-HARDENING STAINLESS STEEL, *JISI,* VOL 200, AUG 1962, P 644-651
8. H.C. BURNETT, R.H. DUFF, AND H.C. VACHER, IDENTIFICATION OF METALLURGICAL REACTIONS AND THEIR EFFECT ON MECHANICAL PROPERTIES OF 17-7PH STEEL, *J. RES. NATL. BUR. STAND.--ENG. INSTRUMENTATION,* VOL 66C, 1962, P 113-119
9. E.G. FEL'DGANDLER AND M.V. PRIDANTSEV, PHASE TRANSFORMATION IN KH 17 N7 YU STAINLESS STEEL, *METALLOVED. TERM. OBRAB. MET.,* NOV 1960, P 2-7
10. J.C. WILKENS AND R.E. PENCE, A STUDY OF THE MICROSTRUCTURE OF PRECIPITATION HARDENING STAINLESS STEEL SHEET, *ADVANCES IN ELECTRON METALLOGRAPHY AND ELECTRON PROBE MICROANALYSIS,* STP 317, ASTM, 1962, P 140-149
11. J.C. WILKENS, PRIVATE COMMUNICATION, IN PRECIPITATION FROM IRON-BASE ALLOYS, PROCEEDINGS OF AN AIME SYMPOSIUM, CLEVELAND, OH, 21 OCT 1963, P 98
12. G. AGGEN, "PHASE TRANSFORMATIONS AND HEAT TREATMENT STUDIES OF A CONTROLLED TRANSFORMATION STAINLESS STEEL ALLOY," DR. ENG. SC. THESIS, RENSSELAER POLYTECHNIC INSTITUTE, AUG 1963
13. D.R. MUZYKA, *THE METALLURGY OF NICKEL-IRON ALLOYS, THE SUPERALLOYS,* C.T. SIMS AND W.C. HAGEL, ED., JOHN WILEY & SONS, 1972, P 113-143
14. A.W. THOMPSON AND J.A. BROOKS, THE MECHANISM OF PRECIPITATION STRENGTHENING IN AN IRON-BASE SUPERALLOY, *ACTA METALL.,* VOL 30, DEC 1982, P 2197-2203
15. T.J. HEADLEY, M.M. KAMOUSKY, AND W.R. SORENSON, EFFECT OF COMPOSITION AND

HIGH ENERGY RATE FORGING ON THE ONSET OF PRECIPITATION IN AN IRON-BASE SUPERALLOY, *METALL. TRANS. A*, VOL 13, MARCH 1982, P 345-353

16. "ARMCO 17-4PH STAINLESS STEEL BAR AND WIRE," PRODUCT DATA BULLETIN NO. S-24, ARMCO, INC., MIDDLETOWN, OH, SEPT 1966
17. "ARMCO PH 13-8MO STAINLESS STEEL," PRODUCT DATA BULLETIN NO. S-24, ARMCO, INC., MIDDLETOWN, OH, OCT 1986
18. "CARPENTER CUSTOM 450," PRODUCT DATA BULLETIN, CARPENTER TECHNOLOGY CORP., READING, PA, 1971
19. "CARPENTER CUSTOM 455," PRODUCT DATA BULLETIN, CARPENTER TECHNOLOGY CORP., READING, PA, 1971
20. "ARMCO 17-7PH AND PH15-7MO STAINLESS STEEL SHEET AND STRIP," PRODUCT DATA BULLETIN, ARMCO, INC., MIDDLETOWN, OH, JAN 1975

Selection of Wrought Precipitation-Hardening Stainless Steels

B. Pollard

Welding Martensitic PH Steels

Following a brief description of the solidification characteristics and microstructures of martensitic PH stainless steels, this section reviews welding parameters for types 17-4PH, 15-5PH, PH13-8 Mo, Custom 450, and Custom 455. Tables 1, 2, 3, and 4 list properties and heat treatment recommendations for these materials. Additional information can be found in the article "Welding of Stainless Steels" in this Volume.

Microstructural Evolution

There have been no systematic studies of the solidification of welds in PH stainless steels, but since the mechanisms involved are the same as for the standard austenitic stainless steels, the primary solidification mode can be predicted for a specific composition by using equations that were originally developed by Suutala and have since been modified by the addition of an aluminum term and higher values for the carbon and nitrogen equivalents. These equations are (Ref 21):

$$\begin{aligned}CR_{EQ} = {} & \%Cr + 1.4(\%Mo) + 1.5(\%Si) + 2(\%Nb) \\ & + 2.5(\%A1) + 3(\%Ti)\end{aligned}$$

$$\begin{aligned}NI_{EQ} = {} & \%Ni + 0.3(\%Mn) + 35(\%C) + 20(\%N) \\ & + \%Cu\end{aligned}$$

Calculated values of the Cr_{eq}/Ni_{eq} ratios for typical PH steels are given in Table 5. According to Suutala, a Cr_{eq}/Ni_{eq} greater than 1.55 is required for primary ferrite solidification (Ref 22). Therefore, the calculated values for the Cr_{eq}/Ni_{eq} ratio indicate that all the martensitic PH stainless steels except Custom 455 solidify as primary δ-ferrite. Upon cooling, nearly all this ferrite first undergoes a transformation to austenite, which subsequently transforms to martensite at temperatures close to ambient. Custom 455 solidifies as primary austenite, which transforms to martensite upon cooling to ambient temperature. Although the as-welded fusion zone microstructures of welds in martensitic PH steels made without filler metal or with a matching filler metal are predominantly martensitic, they may also contain a few percent of δ-ferrite in the untempered martensitic matrix (Fig. 2a). The morphology of this δ-ferrite is summarized in Table 5. The δ-ferrite in welds made with moderate-to-high heat inputs may be intercellular or vermicular. Intercellular ferrite is formed as part of a divorced δ-γ eutectic, which is a result of segregation during the last stages of primary austenite solidification, while vermicular ferrite constitutes the untransformed cores of the primary ferrite dendrites. Very high cooling rates, such as may be experienced by welds made at high speeds in sheet material, suppress the transformation of δ-ferrite to austenite and promote the formation of acicular austenitic-ferritic structures, the austenite content of which subsequently transforms to martensite.

TABLE 5 EFFECT OF COMPOSITION ON THE PRIMARY SOLIDIFICATION MODE AND FINAL AS-WELDED MICROSTRUCTURE OF WELDS IN PH STEELS

ALLOY	**CR_{EQ}/NI_{EQ}(A)**	**OBSERVED WELD-METAL δ-FERRITE, %**	**PREDICTED PRIMARY SOLIDIFICATION MODE(B)**	**FINAL AS-WELDED MICROSTRUCTURE**
AUSTENITIC TYPES				
JBK-75	0.77	0	A	100% AUSTENITE
A-286	0.88	0	A	100% AUSTENITE
MARTENSITIC TYPES				
CUSTOM 455	1.33	~0	A	~100% MARTENSITE
15-5PH	1.60	0-5	F	MARTENSITE + INTERCELLULAR OR VERMICULAR **δ** -FERRITE
17-4PH	1.74	3-8	F	MARTENSITE + VERMICULAR **δ** -FERRITE OR ACICULAR STRUCTURE
CUSTOM 450	1.82	0-5	F	MARTENSITE + VERMICULAR **δ** -FERRITE OR ACICULAR STRUCTURE
PH13-8 MO	1.98	0	F	100% MARTENSITE
SEMIAUSTENITIC TYPES				
AM-355	1.77	~5	F	AUSTENITE + VERMICULAR **δ** -FERRITE OR ACICULAR STRUCTURE
17-7PH	1.92	~25	F	WIDMANSTÄTTEN AUSTENITE IN FERRITIC MATRIX
PH15-7 MO	2.01	25-40	F	WIDMANSTÄTTEN AUSTENITE IN FERRITIC MATRIX
AM-350	2.22	>15	F	WIDMANSTÄTTEN AUSTENITE IN FERRITIC MATRIX

(A) CR_{EQ} = %CR + 1.4(%MO) + 1.5(%SI) + 2(%NB) + 2.5(%AL) + 3(%TI); NI_{EQ} = %NI + 0.3(%MN) + 35(%C) + 20(%N) + %CU. VALUES ARE FOR TYPICAL COMPOSITIONS.

(B) A, AUSTENITIC; F, DELTA FERRITIC

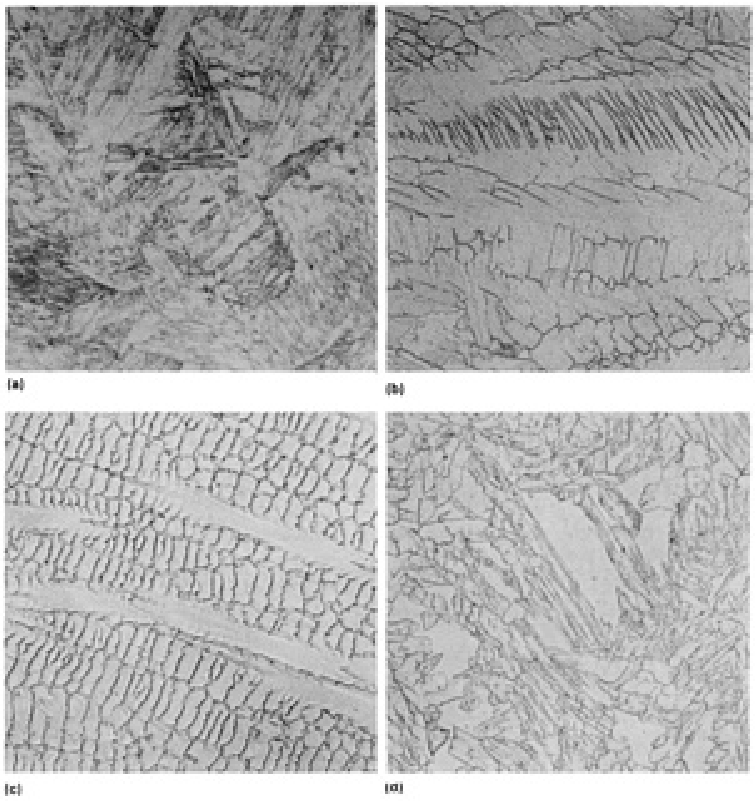

FIG. 2 MICROSTRUCTURES OF AUTOGENOUS GTA WELD METAL IN THE AS-WELDED CONDITION. (A) 17-4PH, δ -FERRITE IN A MATRIX OF UNTEMPERED MARTENSITE. ETCHANT: FRY'S REAGENT. 440×. (B) PH13-8 MO, UNTEMPERED MARTENSITE. ETCHANT: VILELLA'S REAGENT. 440×. (C) 17-7PH, AUSTENITE PLUS ABOUT 20% δ -FERRITE. ETCHANT: ELECTROLYTIC NITRIC ACID-ACETIC ACID FOLLOWED BY ELECTROLYTIC OXALIC ACID. 440×. (D) A-286, FULLY AUSTENITIC STRUCTURE. ETCHANT: 90 ML HYDROCHLORIC ACID, 5 ML NITRIC ACID, 5 ML SULFURIC ACID, 1 ML HYDROFLUORIC ACID. 440×

In the as-welded condition, the heat-affected zones of welds in martensitic PH steels are distinguishable from the base metal by grain growth and an increased amount of ferrite in the high-temperature region close to the fusion line.

Postweld solution treatments reduce the residual δ-ferrite content and refine the prior-austenite grain size of both the fusion and heat-affected zones (HAZ). Aging, likewise, simultaneously tempers the martensite and precipitates the age-hardening particles in the fusion zone, HAZ, and base metal. However, these particles are too small to observe by optical microscopy. Aging may also reform up to 10% austenite at temperatures of 540 to 620 °C (1000 to 1150 °F), the amount increasing with aging temperature.

17-4PH Stainless Steel

Weldability. 17-4PH stainless steel is typical of the martensitic PH stainless steels. Its weldability is superior to that of the standard austenitic stainless steels because, due to its low carbon content, it does not require preheat to prevent cracking in thicknesses up to 100 mm (4.0 in.) (Ref 23 , 24, 25). The welding conditions used for arc welding 17-4PH stainless steel are essentially the same as those used for joining the standard austenitic stainless steels. However, neither welds nor base metal have the high ductility of the austenitic stainless steels, so care must be taken to avoid the presence of stress raisers (Ref 23 24, 25). One common form of stress raiser is the built-in notch at the root of a partial penetration weld. If design requirements dictate partial-penetration welds, the initiation of cracks at the root of the weld can be avoided by making the root pass with a ductile, low-strength filler metal such as type 308L, and then completing the balance of the weld with a matching high-strength heat-treatable filler metal.

Arc Welding. Gas-tungsten arc, gas-metal arc, and shielded metal arc welding processes have all been used extensively for joining 17-4PH stainless steel (Ref 23, 24, 25). Typical mechanical properties of arc welds in 17-4PH are presented in Table 6. The choice of welding process has little effect on the strength of welds after postweld heat treatment (PWHT), but does affect the weld metal cleanliness and resultant ductility and toughness (Ref 23). Gas-tungsten arc welds have the highest and shielded metal arc welds the lowest quality. However, shielded metal arc welds are generally satisfactory for all but the most critical applications.

TABLE 6 MECHANICAL PROPERTIES OF 17-4PH ARC WELDS

THICKNESS		POSTWELD HEAT TREATMENT(B)	TENSILE STRENGTH		0.2% YIELD STRENGTH		ELONGATION, %			REDUCTION IN AREA, %	FAILURE LOCATION(C)	IZOD IMPACT ENERGY(D)	
MM	IN.		mpa	ksi	mpa	ksi	in 13 mm (0.5 in.)	in 25 mm (1 in.)	in 50 mm (2 in.)			J	FT · LBF
GTA WELDS(A)													
1.3	0.050	AGED 480 °C (900 °F), 1 H	1340	194	1295	188	7	. . .	3.0	. . .	FL	. . .	. . .
1.3	0.050	AGED 540 °C (1000 °F), 4 H	1235	179	1170	170	. . .	. . .	5.0	. . .	BM	. . .	. . .
1.3	0.050	AGED 595 °C (1100 °F), 4 H	1140	165	1105	160	14	. . .	5.5	. . .	FL	. . .	. . .
2.3	0.090	AGED 480 °C (900 °F), 1 H	1395	202	1290	187	16	. . .	5.0	. . .	FL	. . .	. . .
2.3	0.090	AGED 540 °C (1000 °F), 4 H	1240	180	1160	168	17	. . .	7.0	. . .	WM	. . .	. . .
2.3	0.090	AGED 595 °C (1100 °F). 4 H	1125	163	1060	154	. . .	. . .	9.0	. . .	BM	. . .	. . .
4.8	0.188	ST + AGED 480 °C (900 °F), 1 H	1370	199	1260	183	. . .	17.0	9.0	35	WM	. . .	. . .
4.8	0.188	ST + AGED 510 °C (950 °F), 1 H	1345	195	1255	182	. . .	18.0	9.0	42	WM	. . .	. . .
4.8	0.188	ST + AGED 550 °C (1025 °F), 1 H	1215	176	1145	166	. . .	17.0	9,2	45	WM	. . .	. . .
GMA WELDS(A)													
4.8	0.188	ST + AGED 480 °C (900 °F), 1 H	1400	203	1250	181	. . .	14.0	7.0	. . .	WM	. . .	. . .
4.8	0.188	ST + AGED 510 °C (950 °F), 1 H	1275	185	1200	174	. . .	14.0	7.5	. . .	WM	. . .	. . .
4.8	0.188	ST + AGED 550 °C (1025 °F), 1 H	1195	173	1125	163	. . .	15.0	7.5	. . .	WM	. . .	. . .
25	1.00	ST + AGED 480 °C (900 °F), 1 H	1305	189	1150	167	. . .	. . .	7.3	21	WM	15	11
SMA WELDS(A)													
4.8	0.188	ST + AGED 480 °C (900 °F), 1 H	1380	200	1275	185	. . .	13.0	6.5	29	. . .	. . .	. . .
4.8	0.188	ST + AGED 510 °C (950 °F), 1 H	1280	186	1205	175	. . .	12.0	6.0	30	. . .	. . .	. . .
4.8	0.188	ST + AGED 550 °C (1025 °F), 1 H	1195	173	1115	162	. . .	12.0	6.0	32	. . .	. . .	. . .
9.5	0.375	AGED 480 °C (900 °F), 1 H	1225	178	1060	154	. . .	10.0	. . .	32	BM	. . .	. . .
9.5	0.375	ST + AGED 480 °C (900 °F), 1 H	1290	187	1185	172	. . .	11.0	. . .	48	WM	. . .	. . .
9.5	0.375	AGED 540 °C (1000 °F), 1 H	1105	160	940	136	. . .	11.0	. . .	29	WM	. . .	. . .
9.5	0.375	ST + AGED 540 °C (1000 °F), 1 H	1170	170	1110	161	. . .	12.0	. . .	48	WM	. . .	. . .
9.5	0.375	AGED 595 °C (1100 °F), 1 H	1095	159	770	112	. . .	13.5	. . .	50	BM	. . .	. . .
9.5	0.375	ST + AGED 595 °C (1100 °F), 1 H	1090	158	925	134	. . .	12.0	. . .	54	WM	. . .	. . .
25	1.00	AGED 480 °C, (900 °F), 1 H	1090	158	815	118	. . .	. . .	8.0	21	WM	38	28
25	1.00	ST + AGED 480 °C (900 °F), 1 H	1280	186	1125	163	. . .	. . .	9.5	36	WM	8	6
25	1.00	AGED 540 °C (1000 °F), 1 H	1070	155	745	108	. . .	. . .	9.5	27	WM	46	34
25	1.00	ST + AGED 540 °C (1000 °F), 1 H	1150	167	1005	146	. . .	. . .	9.5	39	WM	23	17

25	1.00	AGED 595 °C (1100 °F), 1 H	1090	158	765	111	. . .	. . .	7.5	18	WM	12	9
25	1.00	ST + AGED 595 °C (1100 °F), 1 H	1090	158	890	129	. . .	. . .	11.0	38	WM	30	22

Source: Ref 25

(A) ALL WELDS MADE IN CONDITION A WITH A 17-4PH COATED ELECTRODE OR FILLER WIRE.

(B) ST = SOLUTION TREATED AT 1035 °C (1900 °F) FOR 30 MIN; AIR COOLED TO 15 °C (60 °F) AND HELD FOR 1 H.

(C) WM, WELD METAL; FL, FUSION LINE; BM, BASE METAL.

(D) NOTCH IN TRANSVERSE FACE OF WELD METAL.

Base-Metal Condition for Welding. Depending on the thickness and level of restraint, 17-4PH stainless steel can be welded in the annealed (A) or overaged (H1100 or H1150) conditions. The welding of material in conditions H900 through H1075 is generally not recommended. Material up to 25 mm (1.0 in.) thick can usually be welded in condition A provided that the weld is not very heavily restrained. Highly restrained welds or welds in heavier sections are best welded in conditions H1100 or H1150.

Filler Metals. Type 630 covered electrodes and filler wires, which deposit weld metal that closely matches 17-4PH base metal in composition and heat treatment response, are used when high-strength welds are required. Austenitic filler metals, such as type 308L, can be used for joining 17-4PH when high strength is not required in the joint. Welds made with austenitic filler metals are less susceptible to hydrogen embrittlement than those made with martensitic filler metals and may be used in the as-welded condition.

Postweld Heat Treatment. In the as-welded condition, the weld metal and the high-temperature regions of the HAZs of welds in 17-4PH stainless steel have structures consisting primarily of untempered martensite plus a small amount of ferrite (Fig. 2a). Weldments exhibit an aging peak in the HAZ and a weld metal hardness that is only slightly less than that of the base metal in condition A (Fig. 3a). Weldments in 17-4PH stainless steel are not usually put into service in the as-welded condition except for repair welds where PWHT is impractical. In order to obtain weld properties approximating those of the base metal, PWHT is necessary. For single-pass welds made with the base metal in condition A, a simple aging treatment of 1 to 4 h at 480 to 620 °C (900 to 1150 °F) is usually sufficient. It simultaneously hardens the weld metal, HAZ, and base metal and lowers the residual stresses associated with the weld. Because only slight overaging occurs in the portion of the HAZ that is heated into the aging temperature range during welding (Fig. 3a), joint efficiencies of 97 to 100% are obtained (Ref 25).

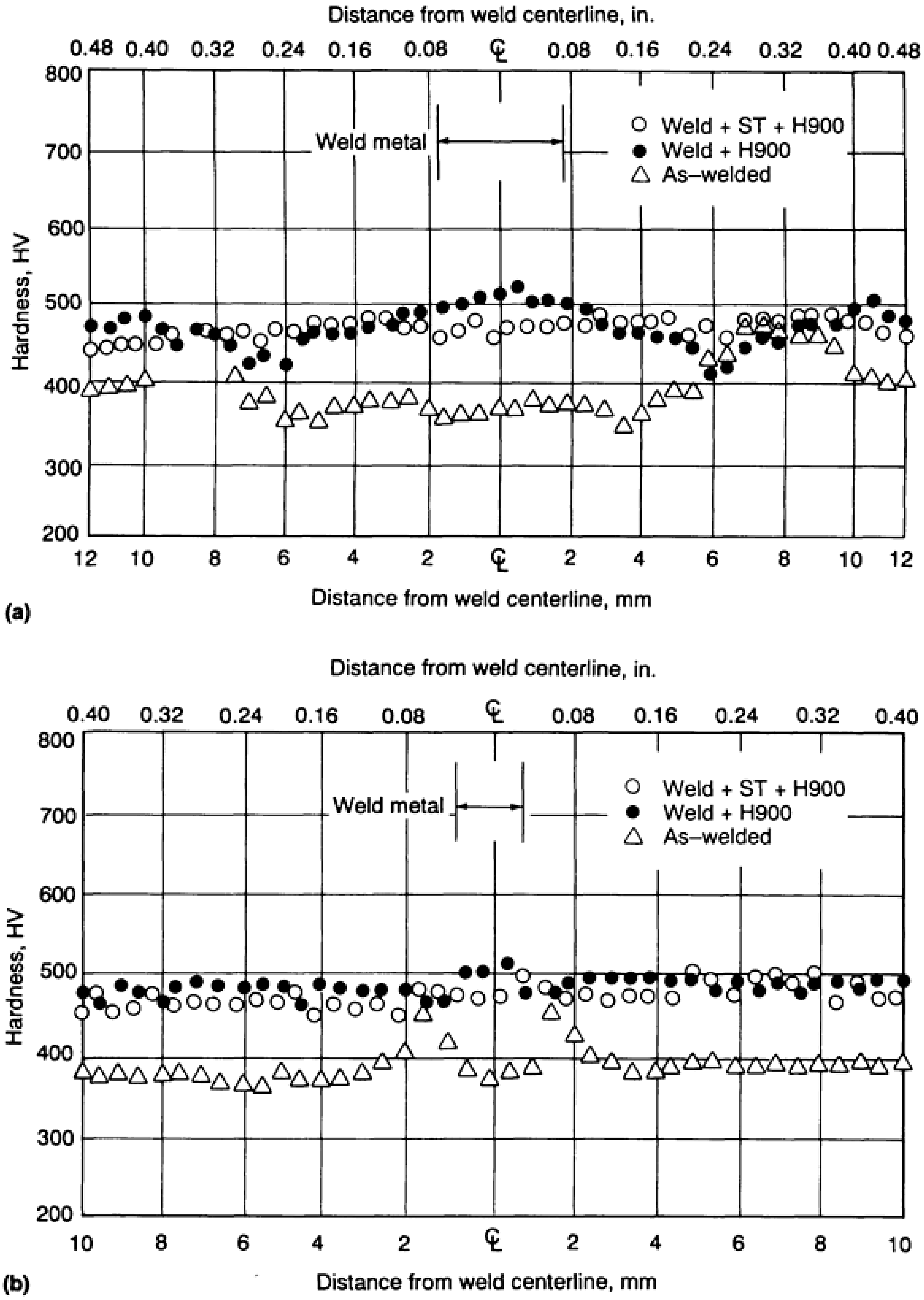

FIG. 3 MICROHARDNESS PROFILES OF AS-WELDED AND POSTWELD HEAT-TREATED WELDS IN 17-4PH MADE WITH (A) PAW AND (B) EBW PROCESSES. SOURCE: REF 26

In multipass welds, the repeated heating from the deposition of successive beads may leave a variation in structure from bead to bead that will result in nonuniform response to the aging treatment. Consequently, in the aged condition, weld yield and ultimate tensile strengths are only about 65% and 80 to 90%, respectively, of the base-metal values (Ref 25). Solution treating the weld before hardening reduces the weld metal and HAZ ferrite contents and improves weld metal uniformity and response to heat treatment. As a result, weld strength increases to 80 to 90% and 90 to 95%, respectively, of the base-metal yield and ultimate tensile strengths. For welds made with the base metal in the overaged condition, solution treatment is required if it is desired to heat treat the weldment to a higher strength level. In general, if the weld deposit is less than 13 mm (0.5 in.) in thickness, fairly good tensile properties can be obtained even if the solution treatment is omitted prior to aging. However, the toughness of the weld metal decreases with aging temperature above 540 °C (1000 °F), probably due to an unfavorable carbide morphology. Therefore, if weld deposits are 13 mm (0.5 in.) or greater in thickness and a postweld solution treatment is not feasible, an age-hardening temperature of 550 °C (1025 °F) or lower is suggested (Ref 25).

Electron-beam (EB) welding is used for joining 17-4PH stainless steel when precision welds with maximum strength and minimum distortion are required. Welding parameters are similar to those used for welding the standard austenitic stainless steels and the same considerations as in arc welding apply to the choice of base-metal condition for welding. The most important characteristic of an EB weld is the total heat input, which is only about 20% of that of a comparable arc weld (Ref 26). As a result, the HAZ is much narrower (Fig. 3b) and the weld metal is finer grained and has a lower ferrite content. Furthermore, since an EB weld is usually a single-pass weld, the weld-metal structure does not exhibit the inhomogeneities of a multipass arc weld. Consequently, the mechanical properties of EB welds are generally superior to those of arc welds, particularly in regard to fracture toughness (Ref 27). Typical mechanical properties of EB welds in 17-4PH are presented in Table 7. Postweld aging alone results in joint efficiencies of 95 to 100% (Ref 26, 28) and total elongation values only slightly lower than those of the base metal; however, the toughness of welds in the H900 condition is low (Ref 26). Re-solution treating before aging has little effect on weld strength but markedly increases the toughness of welds aged at 480 °C (900 °F). However, it produces only a small increase in toughness for welds aged at higher temperatures.

TABLE 7 MECHANICAL PROPERTIES OF ELECTRON-BEAM WELDS IN MARTENSITIC PH STEELS

THICKNESS		PREWELD CONDITION	POSTWELD HEAT TREATMENT(A)	TENSILE STRENGTH		0.2% YIELD STRENGTH		ELONGATION IN 50 MM (2 IN.), %	FAILURE LOCATION(B)	CVN IMPACT ENERGY		FRACTURE TOUGHNESS		REF
mm	in.			mpa	ksi	mpa	ksi			j	ft · lbf	mpa $\sqrt{m}$	ksi $\sqrt{in}$	
17-4 PH														
3.2	0.125	A	AGED 480 °C (900 °F), 1 H	1435	208	1365	198	11.0	BM	2.4	1.8	. . .	. . .	26
3.2	0.125	A	ST + AGED 480 °C (900 °F), 1 H	1400	203	1270	184	12.2	BM	14.0	10.3	. . .	. . .	26
3.2	0.125	A	AGED 550 °C (1025 °F), 4 H	1165	169	1125	163	12.0	BM	15.6	11.5	. . .	. . .	26
3.2	0.125	A	ST + AGED 550 °C (1025 °F), 4 H	1140	165	1090	158	12.0	BM	18.3	13.5	. . .	. . .	26
3.2	0.125	A	AGED 620 °C (1150 °F), 4 H	1025	149	985	143	15.4	BM	17.6	13.0	. . .	. . .	26
3.2	0.125	A	ST + AGED 620 °C (1150 °F), 4 H	1020	148	985	143	14.6	BM	19.4	14.3	. . .	. . .	26
3.2	0.125	A	AGED 760 °C (1400 °F), 2 H + 620 °C (1150 °F), 4 H	965	140	745	108	16.4	BM	20.6	15.2	. . .	. . .	26
9.5	0.375	A	AGED 480 °C (900 °F), 1 H	1315	191	1235	179	12.6(C)	. . .	. . .	. . .	. . .	. . .	28
(E)	(E)	H1150	NONE	885	128	705	102	11(D) TO 22	WM/BM	. . .	. . .	. . .	. . .	27
(E)	(E)	H1150	ST + AGED 595 °C (1 100 °F), 4 H	1055	153	. . .	. . .	. . .	. . .	. . .	. . .	K_Q = 187	K_Q = 170	27
15-5PH														
(E)	(E)	1150	ST + AGED 565 °C (1050 °F), 4 H	. . .	. . .	1125	163	. . .	. . .	. . .	. . .	K_Q = 194	K_Q = 176	27
(E)	(E)	1150	ST + AGED 595 °C (1100 °F), 4 H	. . .	. . .	1020	148	. . .	. . .	. . .	. . .	K_{IC} = 262	K_{IC} = 238	27
PH13-8 MO														
(E)	(E)	H1150M	ST + AGED 540 °C (1000 °F), 4 H	. . .	. . .	1415	205	. . .	. . .	. . .	. . .	K_{IC} = 110	K_{IC} = 100	31

(A) ST = SOLUTION TREATED: 17-4PH AND 15-5PH AT 1035 °C (1900 °F), PH 13-8 MO AT 925 °C (1700 °F) FOR 30 MIN; AIR COOL.

(B) WM, WELD METAL; BM, BASE METAL.

(C) IN 25 MM (1 IN.).

(D) LOW VALUE IS FOR FAILURE IN THE WELD METAL.

(E) EXACT VALUE NOT REPORTED. IN RANGE OF 23 TO 60 MM (0.90 TO 2.37 IN.)

15-5PH Stainless Steel

15-5PH stainless steel differs from 17-4PH in that its composition is balanced so that its structure is ferrite-free and material for bar and forging billets is vacuum arc remelted. As a result, the transverse ductility and toughness of 15-5PH stainless steel are superior to those of 17-4PH. However, its weldability is essentially the same as that of 17-4PH and the same welding processes and procedures are used. A matching 15-5PH filler wire (AMS 5826) is available for critical applications but it is most often welded with type 630 (17-4PH) covered electrodes or filler wire. Postweld heat treatment and weld mechanical properties are essentially identical to those for 17-4PH.

Autogenous welds in 15-5PH stainless steel contain less δ-ferrite than those in 17-4PH. Sufficient δ-ferrite is normally formed to prevent weld solidification cracking in GTA welds. However, the rapid cooling rates that are characteristic of EB and laser welding can produce welds with little or no δ-ferrite, which are susceptible to weld solidification cracking if welded with excessively high welding speeds (Ref 29). The strengths of defect-free EB welds in 15-5PH are similar to those in 17-4PH, but the fracture toughness is higher (Table 7).

PH13-8 Mo Stainless Steel

PH13-8 Mo stainless steel is vacuum induction melted plus vacuum arc remelted. It is ferrite-free and is used where high fracture toughness is required in heavy sections. Weldability is similar to 17-4PH stainless steel and the same procedures are used. It can be welded with the same processes, although covered electrodes for SMA welding are not available. Gas-tungsten arc and GMA welds can be made using a matching filler wire (AMS 5840). Although the final weld metal microstructure is ferrite-free (Fig. 2b), PH13-8 Mo has excellent resistance to weld solidification cracking because it solidifies as δ-ferrite and has low sulfur and phosphorus contents (Ref 30). After postweld solution treating and aging, the yield strengths and Charpy V-notch (CVN) impact energies of multipass GTA welds in PH13-8 Mo stainless steel are 96 to 100% and 50 to 100%, respectively, of the base-metal values (Fig. 4a), the difference between weld- and base-metal properties being greatest for aging temperatures above 540 °C (1000 °F). In heavy sections, PH13-8 Mo stainless steel is often joined by EB welding. After PWHT, EB welds have joint efficiencies of virtually 100% and their fracture toughness is approximately the same as that of the base metal (Ref 31).

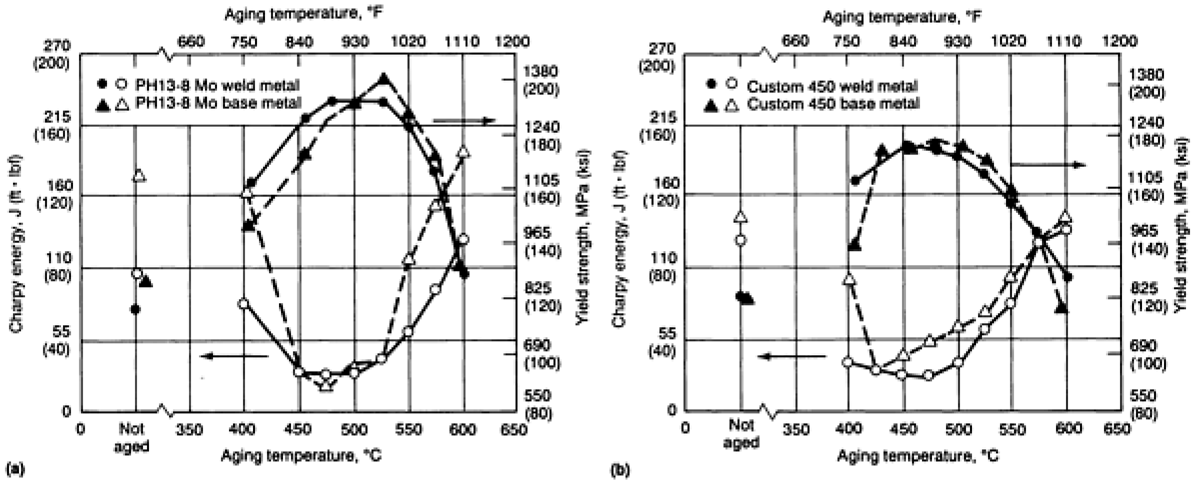

FIG. 4 EFFECT OF AGING TEMPERATURE ON THE YIELD STRENGTH AND CHARPY V-NOTCH IMPACT ENERGY OF (A) PH13-8 MO AND (B) CUSTOM 450 BASE METAL AND WELD METAL. GTA WELDS MADE IN 13 MM (0.50 IN.) THICK PLATE (CONDITION A) WITH MATCHING FILLER METALS. CLOSED SYMBOLS REPRESENT YIELD STRENGTH VALUES; OPEN SYMBOLS REPRESENT IMPACT VALUES. SOURCE: REF 30

Custom 450 Stainless Steel

The composition of Custom 450 stainless steel is balanced so that base metal, HAZ, and weld metal are essentially fully martensitic. In most respects, its weldability is similar to that of 17-4PH, and it can be joined using the same welding processes and procedures (Ref 32). A matching filler wire (AMS 5763) is usually used for GTA and GMA welding Custom 450 and is essential if the weldment is to be used without PWHT. A type 630 (17-4PH) filler wire or coated electrode can be used for joining Custom 450 if the weldment is postweld aged or solution treated and aged.

Custom 450 stainless steel is often used in condition A when only moderate strength is required. Material welded in this condition has acceptable weld properties in the as-welded condition, but CVN impact energy is substantially increased by a postweld solution anneal at 1040 °C (1900 °F) (Ref 32). Higher strength levels are obtained by postweld aging or solution treating and aging. After postweld solution treating and aging, the yield strengths of multipass GTA welds in Custom 450 are 98 to 100% of the base-metal values. Weld metal CVN impact energy increases with aging temperature from about 50% of the base-metal value for welds aged at 480 °C (900 °F) to 100% for welds aged at 565 °C (1050 °F) as shown in Fig. 4(b).

Custom 455 Stainless Steel

Custom 455 stainless steel is vacuum induction melted plus vacuum arc remelted and is completely martensitic. It is joined primarily by GTA welding and a Custom 455 filler metal is available for multipass welds. Procedures are similar to those used for 17-4PH, but a lower postweld solution treatment temperature (815 to 845 °C, or 1500 to 1550 °F) is used. In small sections, joint efficiencies of 80% are obtained by postweld aging only and 100% by solution treating and aging the weldment (Ref 33, 34). In large sections, joint efficiencies in excess of 85% are possible (Ref 19).

References cited in this section

19. "CARPENTER CUSTOM 455," PRODUCT DATA BULLETIN, CARPENTER TECHNOLOGY CORP., READING, PA, 1971
21. B. POLLARD, UNPUBLISHED WORK
22. N. SUUTALA, EFFECT OF SOLIDIFICATION CONDITIONS ON THE SOLIDIFICATION MODE IN STAINLESS STEELS, *METALL. TRANS. A,* VOL 14, FEB 1983, P 191-197
23. G.E. LINNERT, WELDING PRECIPITATION-HARDENING STAINLESS STEELS, *WELD. J.,* VOL 36, JAN 1957, P 9-27
24. F.J. HAWKINS, WELDING OF AGE-HARDENABLE STAINLESS STEELS, *WELD. RES. COUNC. BULL.,* NO. 103, FEB 1965
25. "ARMCO FABRICATING DATA," BULLETIN SF-2, ARMCO, INC., MIDDLETOWN, OH, FEB 1966
26. C. CHEN AND J.S. YEH, ELECTRON BEAM AND PLASMA ARC WELDING OF 17-4PH STAINLESS STEEL, *PROC. CONF. POWER BEAM PROCESSING: ELECTRON, LASER, PLASMA ARC* (SAN DIEGO, CA), ASM INTERNATIONAL, 1988, P 45-54
27. A.J. TURNER, ELECTRON BEAM WELDING THICK SECTION PRECIPITATION-HARDENING STEEL, *WELD. J.,* VOL 60, JAN 1981, P 18-26
28. M.U. ISLAM, G. CAMPBELL, AND R. HSU, FATIGUE AND TENSILE PROPERTIES OF EB WELDED 17-4PH STEEL, *WELD. J.,* VOL 68, SEPT 1989, P 45-50
29. S. MATSUI *ET AL.,* ELECTRON BEAM WELDING OF THE PRECIPITATION HARDENING STAINLESS STEEL, *REVUE DE LA SOUDURE-LASTIJDSCRIFT,* VOL 46, JAN 1990, P 52-57
30. W.R. CIESLAK, J.A. BROOKS, AND W.M. GARRISON, JR., THE WELDABILITY, MICROSTRUCTURE AND PROPERTIES OF PRECIPITATION STRENGTHENED MARTENSITIC STAINLESS STEELS, *PROC. CONF. ADVANCES IN WELDING SCIENCE AND TECHNOLOGY* (GATLINBURG, TN), ASM INTERNATIONAL, 1986
31. A.J. TURNER, ELECTRON BEAM WELDING OF HIGH VOLUME AIRCRAFT COMPONENTS, *PROC. CONF. FUNDAMENTAL AND PRACTICAL APPROACHES TO THE RELIABILITY OF WELDED STRUCTURES* (OSAKA, JAPAN), VOL 1, JAPAN WELDING SOCIETY, 1982, P 139-144
32. "WELDING CARPENTER CUSTOM 450 MULTIPLE-STRENGTH STAINLESS STEEL," PRODUCT DATA BULLETIN, CARPENTER TECHNOLOGY CORP., READING, PA, JAN 1973
33. *ALLOY DIG.,* FILING CODE SS-193, APRIL 1967
34. R. KALTENHAUSER, WELDABILITY OF PRECIPITATION HARDENING STEELS, *MET. ENG. QUART.,* VOL 9 (NO. 1), FEB 1969, P 44-57

Welding Semiaustenitic PH Steels

As shown in Tables 1, 2, 3, and 4, semiaustenitic PH stainless steels include types 17-7PH, PH15-7 Mo, AM-350, and AM-355. As with the previous section on martensitic PH stainless steels, both microstructural evolution and weld parameters associated with semiaustenitic PH steels are described. The article "Welding of Stainless Steels" in this Volume should also be consulted.

Microstructural Evolution

The semiaustenitic PH steels have a primary δ-ferritic solidification mode and, on cooling, the δ-ferrite partially transforms to austenite by a Widmanstätten reaction. This austenite is stable down to ambient temperature so the as-welded structure of the fusion zone consists of austenite in a ferrite matrix (Fig. 2c). With the exception of AM-355, the percentage of retained ferrite in semiaustenitic PH welds is substantially higher than for welds in martensitic PH steels.

Postweld conditioning treatments destabilize the austenite so that it transforms to martensite at ambient or subzero temperatures. Subsequent aging simultaneously tempers the martensite and precipitates the strengthening particles, which are not visible by optical microscopy. The final fusion zone microstructure, therefore, consists of areas of retained ferrite in a tempered martensitic matrix. If a low conditioning temperature is used, carbide particles may also be observed at the ferrite-martensite interfaces.

In the as-welded and, to a lesser degree, the postweld heat-treated conditions, the HAZ is distinguishable from the base metal by an increase in grain size and ferrite content in the high-temperature region close to the fusion line.

17-7PH and PH15-7 Mo Stainless Steels

Weldability. 17-7PH and PH15-7 Mo stainless steels are typical of the semiaustenitic PH steels. As-deposited welds made without filler metal have weld metal and HAZ structures that are predominantly austenitic (Fig. 2c). This ductile structure is not susceptible to cold cracking, so 17-7PH and PH15-7 Mo are readily welded in any condition without preheat and without the need for critical control of interpass temperature or cooling rate (Ref 24, 35). However, autogenous 17-7PH weld metal is not fully austenitic, but rather contains about 25% δ-ferrite versus 10 to 15% in the base metal. Likewise, autogenous PH15-7 Mo weld metal may contain up to 40% δ-ferrite versus 15% in the base metal. The δ-ferrite content of the weld metal is determined primarily by composition, but high values are also promoted to a lesser degree by thicker weld beads, higher welding currents, and higher welding speeds. A large volume fraction of δ-ferrite can make the weld metal susceptible to hot cracking and reduce the ductility of the weld metal in the heat-treated condition. If necessary, the δ-ferrite content of the weld metal can be reduced by the addition of an appropriate filler metal.

Filler Metals. Filler wires are available for 17-7PH (AMS 5824) and PH15-7 Mo (AMS 5812) stainless steels. The compositions of these wires have been adjusted so as to produce deposits with lower δ-ferrite contents than would be produced by the base-metal compositions while still responding correctly to the age-hardening treatment (Ref 24, 35). Covered electrodes are not available for 17-7PH and PH15-7 Mo stainless steels because excessive oxidation of aluminum would occur in the arc. However, type 630 (17-4PH) covered electrodes have been used successfully for joining 17-7PH. To obtain adequate heat treat response with this filler metal, welding conditions that produce high dilution and a compromise PWHT practice are required (Ref 24). A heat treatment of 870 °C (1600 °F) for 90 min is used to simultaneously solution treat the type 630 weld metal and condition the 17-7PH base metal (Ref 35). The weld is then refrigerated at -75 °C (-100 °F) for 8 h and aged at 510 to 595 °C (950 to 1050 °F) for 1 h, depending on the properties required. Standard iron-chromium-nickel filler metals, such as types 308, 308L, 309, and 347, can also be used for joining 17-7PH and PH15-7 Mo, but the weld metal will not respond to heat treatment in the same manner as the base metal. Response to heat treatment will depend on the ratio of austenite-to-ferrite-forming elements in the filler metal and the amount of dilution by the base metal. In most instances, weld strength will be much lower than that obtained with a matching filler metal, but ductility and toughness will be higher.

Gas-Tungsten Arc Welding. Because 17-7PH and PH15-7 Mo stainless steels are used primarily as sheet and strip, they are most frequently joined by GTA welding. Preheating and postheating are not required to prevent cracking, but the selection of welding conditions requires more care than for the standard austenitic stainless steels. Helium gas is preferred over argon for shielding the arc because helium minimizes the formation of a continuous aluminum oxide film on the surface of the weld pool (Ref 23). Particular care should be taken to provide efficient shielding of the weld pool in order to minimize loss of aluminum. Inert gas shielding should also be used on the underside of full penetration welds to protect the weld root. Welding speeds typically are of the order of 5 mm/s (12 in./min) (Ref 35). Higher welding speeds can cause porosity at the fusion line. Welding currents and weld-pass sequence should be selected to minimize the weld cross-sectional area.

Gas-metal arc welding is used for joining 17-7PH and PH15-7 Mo in heavy sheet and plate thicknesses. A mixture of 75He-25Ar is the most widely used shielding gas. Welding conditions are otherwise similar to those used for the standard austenitic stainless steels.

Postweld Heat Treatment. As-deposited welds in 17-7PH and PH15-7 Mo steels have a structure of austenite plus ferrite that is of much lower strength than the hardened base metal. A PWHT is, therefore, essential if a weld strength similar to the base-metal strength is required. The effects of PWHT on the tensile properties of GTA welds in 17-7PH and PH15-7 Mo steels are presented in Table 8; corresponding welding conditions are given in Table 9. The joint efficiencies of welds postweld hardened to the RH950 or TH1050 conditions without reannealing are 100% with weld reinforcement and 95 to 98% without it. Postweld reannealing before hardening has negligible effect on weld strength but increases weld ductility.

TABLE 8 TENSILE PROPERTIES OF GTA WELDS IN 1.6 MM (0.062 IN.) THICK 17-7PH AND PH15-7 MO STEELS

See also Table 9

ALLOY	**PREWELD CONDITION(A)**	**POSTWELD HEAT TREATMENT(A)**	**REINFORCEMENT**	**TENSILE STRENGTH**		**0.2% YIELD STRENGTH**		**ELONGATION, %**		**FAILURE LOCATION(B)**
				MPA	**KSI**	**MPA**	**KSI**	**IN 13 MM (0.5 IN.)**	**IN 50 MM (2 IN.)**	
HARDENED AND TESTED AS-WELDED										
PH15-7 MO	TH1050	NONE	OFF	840	122	605	88	25.0	7.0	WM
PH15-7 MO	R950	NONE	OFF	795	115	600	87	22.0	6.0	WM
WELDED AND HARDENED										
17-7PH	A	TH1050	ON	1415	205	1345	195	6.0	1.5	BM
17-7PH	A	RH950	ON	1515	220	1415	205	6.0	2.0	BM
PH15-7 MO	A	TH1050	ON	1450	210	1415	205	. . .	5.0	BM
PH15-7 MO	A	TH1050	OFF	1415	205	1345	195	8.0	2.0	WM
PH15-7 MO	A	RH950	ON	1655	240	1550	225	. . .	5.0	BM
PH15-7 MO	A	RH950	OFF	1585	230	1480	215	5.0	2.0	WM
WELDED, REANNEALED, AND HARDENED										
17-7PH	A	A + TH1050	ON	1435	208	1345	195	8.0	2.5	BM
PH15-7 MO	A	A + TH1050	ON	1450	210	1415	205	. . .	5.0	BM
PH15-7 MO	A	A + TH1050	OFF	1380	200	1345	195	10.0	2.5	WM
PH15-7 MO	A	A + RH950	ON	1620	235	1480	215	. . .	4.0	BM
PH15-7 MO	A	A + RH950	OFF	1585	230	1450	210	7.0	2.0	WM
PH15-7 MO	A	R + A + RH950	OFF	1660	241	1505	218	9.0	3.5	WM

(A) PREWELD AND POSTWELD HEAT TREATING CONDITIONS. A (CONDITION A): ANNEAL AT 1065 °C (1950 °F) FOR 10 MIN. AIR COOL. R (REFRIGERATION TREATMENT FOR TRANSFORMATION): -75 °C (-100 °F) FOR 8 H. TH1050: 760 °C (1400 °F) FOR 90 MIN, AIR COOL + 565 °C (1050 °F) FOR 90 MIN, AIR COOL. RH950: 955 °C (1750 °F) FOR 10 MIN; AIR COOL + -75 °C (-100 °F) FOR 8 H + 510 °C (950 °F) FOR 60 MIN; AIR COOL.

(B) WM, WELD METAL; BM, BASE METAL

TABLE 9 WELDING CONDITIONS FOR GTA WELDS DESCRIBED IN TABLE 8

WELDING CONDITION	ALLOY	
	17-7PH	PH15-7 MO
THICKNESS, MM (IN.)	1.59 (0.0625)	1.57 (0.062)
SHIELDING GAS	HELIUM	HELIUM
GAS FLOW RATE, L/RAIN (FT3/H)	28 (60)	24 (50)
WELDING SPEED, MM/S (IN./MIN)	6.3 (15)	5.1 (12)
CURRENT	AS REQUIRED FOR FULL PENETRATION IN A SINGLE PASS	AS REQUIRED FOR FULL PENETRATION IN A SINGLE PASS
FILLER WIRE	1.6 MM (0.062 IN.) DIAM W17-7PH	1.6 MM (0.062 IN.) DIAM WPH15-7MO
WIRE FEED RATE, MM/S (IN./MIN)	12 (28)	1 TO 1

Resistance Welding. Spot and seam welding have been used extensively for joining 17-7PH and PH15-7 Mo steels. For good results, particular attention should be given to the surface cleanliness of the steel. Surfaces to be joined should be free of oil, grease, dirt, and heat treating scale. Welding schedules should use longer weld times, lower weld currents, and electrode forces 10 to 20% higher than those normally used for the standard austenitic stainless steels (Ref 24). The best mechanical properties are obtained by welding just before or after final hardening. This practice produces weld nuggets with a tough austenitic microstructure. The tensile strength normal to the weld interface is a maximum for this condition, and the shear strength is not significantly lower than for nuggets in the hardened condition (Ref 35).

Sound flash butt welds can be produced in 17-7PH and PH15-7 Mo stainless steels, but ductility may be low as a result of ferrite stringers in the base metal being bent parallel to the weld interface (Ref 35).

AM-350 and AM-355 Stainless Steels

Weldability. AM-350 and AM-355 differ from 17-7PH and PH15-7 Mo stainless steels in that they do not contain aluminum. Martensitic transformation provides their major strengthening mechanism with an additional but smaller contribution from the precipitation of Cr_2N or an isostructurally similar phase during tempering (Ref 12). The absence of any reactive elements makes the weldability of AM-350 and AM-355 as good as that of the standard austenitic stainless steels and superior to that of 17-7PH. Although these steels are most often joined by GTA and GMA welding, they can be joined by any of the processes used for the standard austenitic stainless steels, including submerged arc, electron-beam, laser-beam, and resistance welding (Ref 34, 36, 37). Like 17-7PH and PH15-7 Mo, these steels may be welded in the annealed or hardened conditions and neither preheat nor special precautions with regard to interpass temperature or cooling rate are necessary. However, PWHT must be used if maximum weld strength is required.

Annealed AM-350 base metal contains 10 to 15% δ-ferrite, and welding produces even higher levels in autogeneous weld metal and the high-temperature portion of the HAZ. In the subzero cooled and tempered (SCT) condition, the final structure consists of islands of ferrite surrounded by carbides in a tempered martensitic matrix. In contrast, the composition of AM-355 is balanced so that the base metal contains less than 5% δ-ferrite in condition H (solution treated at 1065 °C, or 1950 °F) and 0% ferrite in the equalized plus overaged and tempered (EOT+) condition, and welding does not increase the ferrite content to any significant extent.

Filler Metals. An AM-355 filler wire (AMS 5780) with a composition identical to that of the base metal is available for joining AM-355. Likewise, a matching AM-350 filler wire (AMS 4774) is available for welding AM-350, but the AM-355 filler wire is usually preferred for this material because it gives slightly higher weld strengths after PWHT (Ref 36). The AM-355 filler wire can also be used with a proprietary flux for submerged arc welding. AM-355 covered electrodes have been developed for SMA welding but may not be readily available.

Postweld Heat Treatment. Joint efficiencies of 90 to 100% can be obtained by PWHT of AM-350 and AM-355 welds (Ref 36). The PWHT for welds in AM-350 consists of annealing at 930 °C (1710 °F) (condition L) and cooling rapidly to room temperature, followed by subzero cooling to -75 °C (-100 °F) and tempering at 455 or 540 °C (850 or 1000 °F) (condition SCT).

Unlike AM-350, AM-355 must have a suitably prepared microstructure prior to the condition L treatment in order to minimize intergranular precipitation during this treatment. The required microstructure is produced by a prior heat treatment which involves carbide solution annealing for 1 to 3 h at 1040 °C (1900 °F), cooling rapidly enough to room temperature to avoid intergranular carbide precipitation in the temperature range 870 to 650 °C (1600 to 1200 °F), and subzero cooling to -75 °C (-100 °F). The subzero cooling causes a partial transformation to martensite and, during subsequent heat treatments in the carbide precipitation range, carbides precipitate in the former martensitic areas in preference to the grain boundaries. After this treatment, AM-355 weldments are annealed at 915 to 970 °C (1675 to 1775 °F), subzero cooled to -75 °C (-100 °F), and tempered at 455 or 540 °C (850 or 1000 °F) (condition SCT).

Resistance Welding. AM-350 and AM-355 can be spot welded either before or after the hardening treatment using similar conditions as for standard austenitic stainless steels. Postweld heat treatments (for example, condition SCT) produce a moderate increase in tensile shear strength, but also a corresponding decrease in cross-tension strength. Consequently, the best combination of tensile shear and cross-tension strength is obtained by welding in the hardened condition and not postweld heat treating (Ref 24, 37).

References cited in this section

12. G. AGGEN, "PHASE TRANSFORMATIONS AND HEAT TREATMENT STUDIES OF A CONTROLLED TRANSFORMATION STAINLESS STEEL ALLOY," DR. ENG. SC. THESIS, RENSSELAER POLYTECHNIC INSTITUTE, AUG 1963
23. G.E. LINNERT, WELDING PRECIPITATION-HARDENING STAINLESS STEELS, *WELD. J.,* VOL 36, JAN 1957, P 9-27
24. F.J. HAWKINS, WELDING OF AGE-HARDENABLE STAINLESS STEELS, *WELD. RES. COUNC. BULL.,* NO. 103, FEB 1965
34. R. KALTENHAUSER, WELDABILITY OF PRECIPITATION HARDENING STEELS, *MET. ENG. QUART.,* VOL 9 (NO. 1), FEB 1969, P 44-57
35. "ARMCO FABRICATING DATA," BULLETIN SF-5, ARMCO, INC., MIDDLETOWN, OH, DEC 1967
36. "AM-350/AM-355 PRECIPITATION HARDENING STAINLESS STEELS," PRODUCT DATA BULLETIN, ALLEGHENY LUDLUM CORP., PITTSBURGH, PA, 1963
37. H.W. MISHLER, R.E. MONROE, AND P.J. RIEPPEL, "WELDING OF HIGH STRENGTH STEELS FOR AIRCRAFT AND MISSILE APPLICATIONS," DMIC REPORT 118, DEFENSE METALS INFORMATION CENTER, COLUMBUS, OH, 12 OCT 1959

Selection of Wrought Precipitation-Hardening Stainless Steels

B. Pollard

Welding Austenitic PH Steels

Weldability and welding recommendations for A-286 and JBK-75 austenitic PH stainless steels are discussed below. Tables 1, 2, 3, and 4 list properties and heat treatments for these steels. As noted previously, additional information can be found in the article "Welding of Stainless Steels" in this Volume.

A-286 Stainless Steel

Weldability. A-286 stainless steel is highly alloyed and since welds solidify entirely as austenite (Fig. 2d), they are susceptible to hot cracking in both the fusion zone and HAZ (Ref 38, 39, 40, 41, 42). Solidification cracking in the fusion zone results from the formation of a low-melting-point Laves-phase eutectic at the grain boundaries, which is produced by the segregation of titanium and other solutes during the last stages of solidification (Ref 40, 42). This phase is enriched in titanium, nickel, silicon, and molybdenum and depleted in iron and chromium (Ref 40). The same phase is also responsible for HAZ cracking (Ref 39, 40, 41, 42). However, in the HAZ, the Laves-phase eutectic is formed by the constitutional liquation of (Ti,Mo)(C,N) inclusions (Ref 41, 42). Constitutional liquation is a result of the rapid, nonequilibrium heating conditions typical of arc welding, which cause the carbonitrides to decompose more rapidly than their constituent elements can diffuse into the matrix. The result is the formation of solute-rich zones of eutectic composition (and hence low melting point) around the decomposing carbonitrides. This liquid then wets the grain boundaries, and the grains are pulled apart by shrinkage stresses as the weld cools. Of the residual elements normally present in A-286, silicon and boron have been found to have the greatest effect on hot cracking in both the fusion zone and HAZ (Ref 40, 41, 42).

The ease with which crack-free arc welds can be produced in A-286 depends on the grade of material, weld-section thickness, amount of restraint, the welding process used, the welding procedure, and the choice of filler metal. Both regular and welding grades of A-286 are produced. The welding grades (AMS 5858 and AMS 5895) have lower limits on manganese, silicon, sulfur, and phosphorus than the regular grades and exhibit greater resistance to hot cracking. Whenever possible, they should be specified in preference to the regular grades for applications that involve welding. Although A-286 is classified as a difficult-to-weld material, it can be joined successfully by GTA, EB, and resistance welding if the correct procedures are used.

Gas-Tungsten Arc Welding. When GTA welding A-286, the best results are obtained with the base metal in the solution treated condition, low-heat inputs, and minimum restraint (Ref 24, 34). Autogenous crack-free GTA welds can be produced in thin sheet, but welds in thick sheet and plate require filler metal. A stringer bead technique should be used for multipass welds in heavy sections.

The choice of filler metal depends on the application. The regular A-286 filler wires (AMS 5804) have lower limits on sulfur and phosphorus contents but are otherwise similar in composition to the base metal. Deposits made with these wires are the most prone to hot cracking, so they have been used primarily for single-pass welds (Ref 34). Premium quality A-286 filler wires (AMS 5805) with lower carbon, manganese, silicon, phosphorus, and boron contents and limits on oxygen, nitrogen, and hydrogen levels are also available. These wires are produced from vacuum induction melted material by wire rolling instead of drawing (Ref 43, 44). They produce high-purity deposits that are less susceptible to hot cracking than those made using the regular filler wires and are used for critical applications such as jet engine repairs.

Dissimilar metal filler wires are the usual choice for multipass welds if a matching filler metal is not essential because they exhibit greater resistance to hot cracking than A-286 weld metal. Hastelloy W has been used extensively for joining A-286 (Ref 24, 43). It has excellent resistance to hot cracking, but its stress-rupture strength is substantially lower than that of A-286 weld metal and it is not suitable for many high-temperature environments. Type 310 filler metal has been used for joining A-286 plate (Ref 24). It has very good corrosion resistance in many media but is nonhardening. AMS 5675 (AWS A5.14 Class ERNiCrFe-6) filler metal is significantly less crack sensitive than A-286 and is age hardenable, but it does not quite match the strength of A-286.

Postweld Heat Treatment. Typical tensile properties of welds in A-286 are presented in Table 10. A-286 weld metal does not age harden as rapidly or to the same extent as the base metal. A PWHT of 725 °C (1335 °F) for 48 h is required to increase the yield and ultimate tensile strengths to 93% and 89%, respectively, of the base-metal values (Ref 45). However, aging is much more rapid if the weld is first solution treated (Ref 46). For example, solution treating at 1080 °C (1975 °F) for 2 h and aging at 725 °C (1335 °F) for 16 h increases the yield and ultimate tensile strengths to 94% and 97%, respectively, of the base-metal values. If a high-temperature solution treatment is not feasible, weld strength can be increased by using a low-temperature multistep heat treatment consisting of 725 °C (1335 °F), 4 h plus 650 °C (1202 °F), 5 h plus 725 °C (1335 °F), 44 h (Ref 45). This heat treatment raises the yield and ultimate tensile strengths to 99% and 93%, respectively, of the base-metal values.

TABLE 10 TENSILE PROPERTIES OF WELDS IN AUSTENITIC PH STEELS

THICKNESS		FILLER METAL	PREWELD CONDITION(A)	SPECIMEN(B)	POSTWELD HEAT TREATMENT	TENSILE STRENGTH		0.2% YIELD STRENGTH		ELONGATION, %		REDUCTION IN AREA, %	COMMENTS	REF
mm	in.					mpa	ksi	mpa	ksi	in 25 mm (1 in.)	in 50 mm (2 in.)			
A-286 GTA WELDS														
3.0	0.118	NONE	ST	BM	AGED 725 °C (1335 °F), 48 H	952	138	710	103	16	. . .	29	. . .	45
3.0	0.118	NONE	ST	W	AGED 725 °C (1335 °F), 48 H	850	123	659	96	8	. . .	35	NO CRACKING	45
3.0	0.118	NONE	ST	BM	AGED 725 °C (1335 °F), 4 H + 650 °C (1200 °F), 5 H + 725 °C 1335 °F, 44 H	1020	148	777	113	20	. . .	28	. . .	45
3.0	0.118	NONE	ST	W	AGED 725 °C (1335 °F), 4 H + 650 °C (1200 °F), 5 H + 725 °C (1335 °F), 44	952	138	769	112	. . .	. . .	45	NO CRACKING	45
35	1.38	A-286	ST	BM	ST 980 °C (1795 °F), 2 H + AGED 725 °C (1335 °F), 16 H	1069	155	708	103	30	. . .	43	. . .	46
35	1.38	A-286	ST	W	NONE	724	105	611	89	15	. . .	34	. . .	46
35	1.38	A-286	ST	W	ST 980 °C (1795 °F), 2 H + AGED 725 °C (1335 °F), 16 H	1033	150	665	96	31	. . .	48	SCATTERED MICROFISSURES IN WELD METAL	46
A-286 EB WELDS														
3.2	0.125	NONE	ST	BM	AGED 730 °C (1350 °F), 16 H	1047	152	793	115	. . .	23.0	. . .	. . .	47
3.2	0.125	NONE	ST	W	AGED 730 °C (1350 °F), 16 H	919	133	775	112	. . .	5.6	. . .	MICROFISSURES IN HAZ	47
3.2	0.125	NONE	ST + AGED	W	REAGED 730 °C (1350 °F),	889	130	. . .	. . .	. . .	9.4	. . .	MICROFISSURES IN HAZ	47

					16 H									
JBK-75 GTA WELDS														
3.2	0.125	NONE	ST	BM	AGED 725 °C (1335 °F), 24 H	L130(C)	164(C)	. . .	. . .	. . .	. . .	. . .	. . .	52
3.2	0.125	NONE	ST	W	AGED 725 °C (1335 °F), 24 H	800(C)	116(C)	. . .	. . .	. . .	. . .	. . .	NO CRACKING	52
3.2	0.125	NONE	ST	BM	AGED 725 °C (1335 °F), 24 H + 650 °C (1200 °F), 24 H	1215	176	965	140	17	. . .	41	. . .	52
3.2	0.125	NONE	ST	W	AGED 725 °C (1335 °F), 24 H + 650 °C (1200 °F), 24 H	945	137	841	122	4	. . .	26	NO CRACKING	52
3.2	0.125	NONE	ST	BM	ST 950 °C (1740 °F), 24 H + 725 °C (1335 °F), 24 H	980(C)	142(C)	. . .	. . .	. . .	. . .	. . .	. . .	52
3.2	0.125	NONE	ST	W	ST 950 °C (1335 °F), 24 H + 650 °C (1200 °F), 24 H	980(C)	142(C)	. . .	. . .	. . .	. . .	. . .	NO CRACKING	52

(A) ST = SOLUTION TREATED.

(B) TRANSVERSE TENSILE SPECIMENS: BM. BASE METAL; W, WELDMENT.

(C) ESTIMATED FROM HARDNESS

Electron-Beam Welding. Although EB welding is a low-heat input process compared to arc welding, the use of EB welding does not guarantee freedom from hot cracking. However, crack-free EB welds can be made in A-286 if heat input, welding speed, and restraint are low. A-286 can be EB welded in both the solution treated and hardened conditions. However, resistance to hot cracking is greater in the solution treated condition. EB welds in A-286 are usually made without filler metal. In the as-welded condition, weld strength is about the same as for GTA welds in the same condition, but postweld aging is more effective in increasing weld strength than for GTA welds in A-286. For example, the yield and ultimate tensile strengths of EB welds made in 3.2 mm (0.128 in.) thick A-286 in the solution treated condition has yield and ultimate tensile strengths of 98% and 88%, respectively, of the base-metal values after aging at 720 °C (1330 °F) for 16 h (Ref 47).

Resistance Welding. A-286 stainless steel can be joined by resistance spot and seam welding in both the soft and hardened conditions, but the welds, as with those made by other fusion welding processes, are susceptible to hot cracking. However, some hot cracking can be tolerated in spot and seam welds since cracking of up to 5 vol% of the weld nugget has little effect on static or cyclic weld strength if it is concentrated at the center of the weld nugget (Ref 48). Hot cracking can be reduced or eliminated by using electrode forces significantly higher than those used for the standard austenitic stainless steels (Ref 24, 48). Welding conditions that slow down the cooling rate (for example, downsloping the weld current) have also been recommended as a means of eliminating cracking (Ref 48). Limited data suggest that longer weld times and lower weld currents than those used for standard austenitic stainless steels may also be beneficial.

Spot welds made in hardened material without PWHT have the best combination of shear and cross-tension strengths (Ref 37).

Seam welds in A-286 are also susceptible to hot cracking. Limited data indicate that the best results are obtained with a lower wheel speed, lower weld current and longer (on plus off) time than are generally used for the standard austenitic stainless steels. Increased electrode force and continuous (rather than interrupted) weld current may also be used to prevent cracking (Ref 49).

Little information is available on welding conditions for flash butt welding A-286 because these conditions vary from machine to machine and application to application. However, A-286 is routinely joined by flash butt welding with excellent results. After postweld aging at 720 °C (1325 °F), weld properties are virtually equal to those of the base metal (Ref 50).

JBK-75 Stainless Steel

JBK-75 is a modification of A-286 with the low carbon, manganese, silicon, and boron contents of welding grade A-286 plus a higher nickel content. The beneficial effect of the higher nickel content is explained by the concept of shrinkage brittleness (Ref 51). According to this concept, during the last stages of solidification, a minimum amount of eutectic liquid is required to fill all the dendritic fissures formed by the eutectic films. If a critical, subminimum amount of eutectic forms, severe cracking occurs. On the other hand, if only a very small amount of eutectic forms, no cracking occurs because enough solid boundaries are present to withstand the shrinkage stresses. Therefore, the beneficial effect of the high nickel content of JBK-75 is due to it increasing the amount of eutectic formed so that any microfissures which nucleate are self-healing. Consequently, JBK-75 has greater resistance to hot cracking than A-286 and can be GTA welded in heavier sections without cracking.

Mechanical properties of GTA welds in JBK-75 are presented in Table 10. Hardness measurements indicate that the weld ultimate tensile strength (UTS) is only about 70% of the base-metal value after postweld aging at 725 °C (1335 °F) for 24 h (Ref 52). Higher weld- and base-metal strengths and a joint efficiency, based on the UTS, of about 78% are obtained by a double aging treatment consisting of 725 °C (1335 °F), 24 h plus 650 °C (1200 °F), for 24 h. Homogenizing at 950 °C (1740 °F) before aging also increases weld-metal strength but simultaneously reduces the strength of the base metal. For example, a heat treatment consisting of 950 °C (1740 °F) for 24 h plus 725 °C (1335 °F) for 24 h produces a hardness of 30 HRC, which is equivalent to a UTS of about 965 MPa (140 ksi) in both the weld and base metal.

References cited in this section

24. F.J. HAWKINS, WELDING OF AGE-HARDENABLE STAINLESS STEELS, *WELD. RES. COUNC. BULL.,* NO. 103, FEB 1965

34. R. KALTENHAUSER, WELDABILITY OF PRECIPITATION HARDENING STEELS, *MET. ENG. QUART.,* VOL 9 (NO. 1), FEB 1969, P 44-57
37. H.W. MISHLER, R.E. MONROE, AND P.J. RIEPPEL, "WELDING OF HIGH STRENGTH STEELS FOR AIRCRAFT AND MISSILE APPLICATIONS," DMIC REPORT 118, DEFENSE METALS INFORMATION CENTER, COLUMBUS, OH, 12 OCT 1959
38. J.J. VAGI AND D.C. MARTIN, WELDING OF HIGH-STRENGTH STAINLESS STEELS FOR ELEVATED TEMPERATURE USE, *WELD. J.,* VOL 35, MARCH 1956, P 137S-144S
39. B.S. BLUM AND R.H. WITT, HEAT AFFECTED ZONE CRACKING IN A-286 WELDMENTS, *WELD. J.,* VOL 42, AUG 1963, P 365S-370S
40. J.A. BROOKS AND R.W. KRENZER, PROGRESS TOWARD A MORE WELDABLE A-286, *WELD. J.,* VOL 53, JUNE 1974, P 242S-245S
41. J.A. BROOKS, EFFECT OF ALLOY MODIFICATIONS ON HAZ CRACKING OF A-286 STAINLESS STEEL, *WELD. J.,* VOL 53, NOV 1974, P 517S-523S
42. J.A. BROOKS, A.W. THOMPSON, AND J.C. WILLIAMS, WELD CRACKING OF AUSTENITIC STAINLESS STEELS--EFFECTS OF IMPURITIES AND MINOR ELEMENTS, *PROC. CONF. PHYS. METALLURGY OF METAL JOINING* (ST. LOUIS, MO), TMS/AIME, 1980, P 117-136
43. W.H. SIMON, S. BANERJEE, AND R.S. CREMISIO, METALLURGICALLY CONTROLLED WELDING FILLER WIRE--TRACE ELEMENT CONTROLLED TO PRODUCE CONSISTENTLY SOUND WELDS, *PROC. INT. CONF.: THE EFFECTS OF RESIDUAL IMPURITY AND MICRO-ALLOYING ELEMENTS ON WELDABILITY AND WELD PROPERTIES* (LONDON, ENGLAND), WELDING INSTITUTE, PAPER 14, 1983
44. D.W. HARVEY, METALLURGICALLY CONTROLLED WELDING FILLER WIRE, *WELD. REV.,* MAY 1988, P 80, 82, 122, 123
45. L.T. SUMMERS, "EFFECT OF CHEMICAL SEGREGATION ON THE MICROSTRUCTURE AND PROPERTIES OF A-286 WELDMENTS," PH.D. DISSERTATION, UNIVERSITY OF CALIFORNIA, BERKELEY, 1984
46. A. INOUE *ET AL.,* FATIGUE STRENGTH OF ALLOY A-286 AT CRYOGENIC TEMPERATURE, *PROC. CONF. FATIGUE '87,* VOL 1 (CHARLOTTESVILLE, VA), ENGINEERING MATERIALS ADVISORY SERVICES, LTD., 1987, P 547-555
47. R.E. ROTH AND N.F. BRATKOVICH, CHARACTERISTICS AND STRENGTH DATA OF ELECTRON BEAM WELDS IN FOUR REPRESENTATIVE MATERIALS, *WELD. J.,* VOL 41, MAY 1962, P 229S-240S
48. I.F. SQUIRES, DETERMINATION OF RESISTANCE SPOT WELDING CONDITIONS FOR SOME HEAT-RESISTING STAINLESS STEELS, BR. *WELD. J.,* VOL 9, MARCH 1962, P 149-157
49. P.M. KNOWLSON, DETERMINATION OF RESISTANCE SEAM WELDING CONDITIONS FOR SOME HEAT RESISTING STEELS, *BR. WELD. J.,* VOL 9, 1962, P 436-445
50. J.J. VAGI, R.M. EVANS, AND D.C. MARTIN, "WELDING PRECIPITATION-HARDENING STAINLESS STEELS," NASA TECHNICAL MEMORANDUM X-53582, 28 FEB 1967
51. B.I. MEDOVAR, ON THE NATURE OF WELD HOT CRACKING, TRANSLATED FROM *AVTOM. SVARKA,* VOL 7 (NO. 4), 1954, P 12-28
52. M.J. STRUM, L.T. SUMMERS, AND J.W. MORRIS, JR., THE AGING RESPONSE OF A WELDED IRON-BASED SUPERALLOY, *WELD. J.,* VOL 62, SEPT 1983, P 235S-242S

Selection of Wrought Precipitation-Hardening Stainless Steels

B. Pollard

References

1. V. SEETHARAMAN ET AL., PRECIPITATION HARDENING IN A PH13-8 MO STAINLESS STEEL, *MATER. SCI. ENG.,* VOL 47, JAN 1981, P 1-11
2. H.J. RACK AND D. KALISH, THE STRENGTH, FRACTURE TOUGHNESS AND LOW CYCLE FATIGUE BEHAVIOR OF 17-4PH STAINLESS STEEL, *METALL. TRANS. A,* VOL 5, JULY 1974, P 1595-1605
3. K.C. ANTHONY, AGING REACTIONS IN PRECIPITATION HARDENABLE STAINLESS STEEL, *J. MET.,* VOL 15, DEC 1963, P 922-927
4. M. HENTHORNE, T.A. DEBOLD, AND R.J. YINGER, "CUSTOM 450-A NEW HIGHER STRENGTH STAINLESS STEEL," PAPER NO. 53, CORROSION '72, NATIONAL ASSOCIATION OF CORROSION ENGINEERS CONFERENCE, 1972
5. J.E. MCBRIDE, JR., AND G.N. MANIAR, *METALLOGRAPHY AS A QUALITY CONTROL TOOL,* PLENUM PUBLISHING, 1980, P 279
6. S. WIDGE, "A STUDY OF THE EFFECTS OF VARYING SOLUTION TREATMENT ON THE TOUGHNESS OF A STAINLESS MARAGING ALLOY," PH.D. DISSERTATION, LEHIGH UNIVERSITY, 1984
7. E.E. UNDERWOOD, A.E. AUSTIN, AND G.K. MANNING, THE MECHANISM OF HARDENING IN 17-7 NI-CR PRECIPITATION-HARDENING STAINLESS STEEL, *JISI,* VOL 200, AUG 1962, P 644-651
8. H.C. BURNETT, R.H. DUFF, AND H.C. VACHER, IDENTIFICATION OF METALLURGICAL REACTIONS AND THEIR EFFECT ON MECHANICAL PROPERTIES OF 17-7PH STEEL, *J. RES. NATL. BUR. STAND.--ENG. INSTRUMENTATION,* VOL 66C, 1962, P 113-119
9. E.G. FEL'DGANDLER AND M.V. PRIDANTSEV, PHASE TRANSFORMATION IN KH 17 N7 YU STAINLESS STEEL, *METALLOVED. TERM. OBRAB. MET.,* NOV 1960, P 2-7
10. J.C. WILKENS AND R.E. PENCE, A STUDY OF THE MICROSTRUCTURE OF PRECIPITATION HARDENING STAINLESS STEEL SHEET, *ADVANCES IN ELECTRON METALLOGRAPHY AND ELECTRON PROBE MICROANALYSIS,* STP 317, ASTM, 1962, P 140-149
11. J.C. WILKENS, PRIVATE COMMUNICATION, IN PRECIPITATION FROM IRON-BASE ALLOYS, PROCEEDINGS OF AN AIME SYMPOSIUM, CLEVELAND, OH, 21 OCT 1963, P 98
12. G. AGGEN, "PHASE TRANSFORMATIONS AND HEAT TREATMENT STUDIES OF A CONTROLLED TRANSFORMATION STAINLESS STEEL ALLOY," DR. ENG. SC. THESIS, RENSSELAER POLYTECHNIC INSTITUTE, AUG 1963
13. D.R. MUZYKA, *THE METALLURGY OF NICKEL-IRON ALLOYS, THE SUPERALLOYS,* C.T. SIMS AND W.C. HAGEL, ED., JOHN WILEY & SONS, 1972, P 113-143
14. A.W. THOMPSON AND J.A. BROOKS, THE MECHANISM OF PRECIPITATION STRENGTHENING IN AN IRON-BASE SUPERALLOY, *ACTA METALL.,* VOL 30, DEC 1982, P 2197-2203
15. T.J. HEADLEY, M.M. KAMOUSKY, AND W.R. SORENSON, EFFECT OF COMPOSITION AND HIGH ENERGY RATE FORGING ON THE ONSET OF PRECIPITATION IN AN IRON-BASE SUPERALLOY, *METALL. TRANS. A,* VOL 13, MARCH 1982, P 345-353
16. "ARMCO 17-4PH STAINLESS STEEL BAR AND WIRE," PRODUCT DATA BULLETIN NO. S-24, ARMCO, INC., MIDDLETOWN, OH, SEPT 1966
17. "ARMCO PH 13-8MO STAINLESS STEEL," PRODUCT DATA BULLETIN NO. S-24, ARMCO, INC., MIDDLETOWN, OH, OCT 1986
18. "CARPENTER CUSTOM 450," PRODUCT DATA BULLETIN, CARPENTER TECHNOLOGY CORP., READING, PA, 1971
19. "CARPENTER CUSTOM 455," PRODUCT DATA BULLETIN, CARPENTER TECHNOLOGY CORP., READING, PA, 1971
20. "ARMCO 17-7PH AND PH15-7MO STAINLESS STEEL SHEET AND STRIP," PRODUCT DATA BULLETIN, ARMCO, INC., MIDDLETOWN, OH, JAN 1975
21. B. POLLARD, UNPUBLISHED WORK
22. N. SUUTALA, EFFECT OF SOLIDIFICATION CONDITIONS ON THE SOLIDIFICATION MODE IN

STAINLESS STEELS, *METALL. TRANS. A,* VOL 14, FEB 1983, P 191-197

23. G.E. LINNERT, WELDING PRECIPITATION-HARDENING STAINLESS STEELS, *WELD. J.,* VOL 36, JAN 1957, P 9-27
24. F.J. HAWKINS, WELDING OF AGE-HARDENABLE STAINLESS STEELS, *WELD. RES. COUNC. BULL.,* NO. 103, FEB 1965
25. "ARMCO FABRICATING DATA," BULLETIN SF-2, ARMCO, INC., MIDDLETOWN, OH, FEB 1966
26. C. CHEN AND J.S. YEH, ELECTRON BEAM AND PLASMA ARC WELDING OF 17-4PH STAINLESS STEEL, *PROC. CONF. POWER BEAM PROCESSING: ELECTRON, LASER, PLASMA ARC* (SAN DIEGO, CA), ASM INTERNATIONAL, 1988, P 45-54
27. A.J. TURNER, ELECTRON BEAM WELDING THICK SECTION PRECIPITATION-HARDENING STEEL, *WELD. J.,* VOL 60, JAN 1981, P 18-26
28. M.U. ISLAM, G. CAMPBELL, AND R. HSU, FATIGUE AND TENSILE PROPERTIES OF EB WELDED 17-4PH STEEL, *WELD. J.,* VOL 68, SEPT 1989, P 45-50
29. S. MATSUI *ET AL.,* ELECTRON BEAM WELDING OF THE PRECIPITATION HARDENING STAINLESS STEEL, *REVUE DE LA SOUDURE-LASTIJDSCRIFT,* VOL 46, JAN 1990, P 52-57
30. W.R. CIESLAK, J.A. BROOKS, AND W.M. GARRISON, JR., THE WELDABILITY, MICROSTRUCTURE AND PROPERTIES OF PRECIPITATION STRENGTHENED MARTENSITIC STAINLESS STEELS, *PROC. CONF. ADVANCES IN WELDING SCIENCE AND TECHNOLOGY* (GATLINBURG, TN), ASM INTERNATIONAL, 1986
31. A.J. TURNER, ELECTRON BEAM WELDING OF HIGH VOLUME AIRCRAFT COMPONENTS, *PROC. CONF. FUNDAMENTAL AND PRACTICAL APPROACHES TO THE RELIABILITY OF WELDED STRUCTURES* (OSAKA, JAPAN), VOL 1, JAPAN WELDING SOCIETY, 1982, P 139-144
32. "WELDING CARPENTER CUSTOM 450 MULTIPLE-STRENGTH STAINLESS STEEL," PRODUCT DATA BULLETIN, CARPENTER TECHNOLOGY CORP., READING, PA, JAN 1973
33. *ALLOY DIG.,* FILING CODE SS-193, APRIL 1967
34. R. KALTENHAUSER, WELDABILITY OF PRECIPITATION HARDENING STEELS, *MET. ENG. QUART.,* VOL 9 (NO. 1), FEB 1969, P 44-57
35. "ARMCO FABRICATING DATA," BULLETIN SF-5, ARMCO, INC., MIDDLETOWN, OH, DEC 1967
36. "AM-350/AM-355 PRECIPITATION HARDENING STAINLESS STEELS," PRODUCT DATA BULLETIN, ALLEGHENY LUDLUM CORP., PITTSBURGH, PA, 1963
37. H.W. MISHLER, R.E. MONROE, AND P.J. RIEPPEL, "WELDING OF HIGH STRENGTH STEELS FOR AIRCRAFT AND MISSILE APPLICATIONS," DMIC REPORT 118, DEFENSE METALS INFORMATION CENTER, COLUMBUS, OH, 12 OCT 1959
38. J.J. VAGI AND D.C. MARTIN, WELDING OF HIGH-STRENGTH STAINLESS STEELS FOR ELEVATED TEMPERATURE USE, *WELD. J.,* VOL 35, MARCH 1956, P 137S-144S
39. B.S. BLUM AND R.H. WITT, HEAT AFFECTED ZONE CRACKING IN A-286 WELDMENTS, *WELD. J.,* VOL 42, AUG 1963, P 365S-370S
40. J.A. BROOKS AND R.W. KRENZER, PROGRESS TOWARD A MORE WELDABLE A-286, *WELD. J.,* VOL 53, JUNE 1974, P 242S-245S
41. J.A. BROOKS, EFFECT OF ALLOY MODIFICATIONS ON HAZ CRACKING OF A-286 STAINLESS STEEL, *WELD. J.,* VOL 53, NOV 1974, P 517S-523S
42. J.A. BROOKS, A.W. THOMPSON, AND J.C. WILLIAMS, WELD CRACKING OF AUSTENITIC STAINLESS STEELS--EFFECTS OF IMPURITIES AND MINOR ELEMENTS, *PROC. CONF. PHYS. METALLURGY OF METAL JOINING* (ST. LOUIS, MO), TMS/AIME, 1980, P 117-136
43. W.H. SIMON, S. BANERJEE, AND R.S. CREMISIO, METALLURGICALLY CONTROLLED WELDING FILLER WIRE--TRACE ELEMENT CONTROLLED TO PRODUCE CONSISTENTLY SOUND WELDS, *PROC. INT. CONF.: THE EFFECTS OF RESIDUAL IMPURITY AND MICRO-ALLOYING ELEMENTS ON WELDABILITY AND WELD PROPERTIES* (LONDON, ENGLAND), WELDING INSTITUTE, PAPER 14, 1983

44. D.W. HARVEY, METALLURGICALLY CONTROLLED WELDING FILLER WIRE, *WELD. REV.*, MAY 1988, P 80, 82, 122, 123
45. L.T. SUMMERS, "EFFECT OF CHEMICAL SEGREGATION ON THE MICROSTRUCTURE AND PROPERTIES OF A-286 WELDMENTS," PH.D. DISSERTATION, UNIVERSITY OF CALIFORNIA, BERKELEY, 1984
46. A. INOUE *ET AL.*, FATIGUE STRENGTH OF ALLOY A-286 AT CRYOGENIC TEMPERATURE, *PROC. CONF. FATIGUE '87*, VOL 1 (CHARLOTTESVILLE, VA), ENGINEERING MATERIALS ADVISORY SERVICES, LTD., 1987, P 547-555
47. R.E. ROTH AND N.F. BRATKOVICH, CHARACTERISTICS AND STRENGTH DATA OF ELECTRON BEAM WELDS IN FOUR REPRESENTATIVE MATERIALS, *WELD. J.*, VOL 41, MAY 1962, P 229S-240S
48. I.F. SQUIRES, DETERMINATION OF RESISTANCE SPOT WELDING CONDITIONS FOR SOME HEAT-RESISTING STAINLESS STEELS, BR. *WELD. J.*, VOL 9, MARCH 1962, P 149-157
49. P.M. KNOWLSON, DETERMINATION OF RESISTANCE SEAM WELDING CONDITIONS FOR SOME HEAT RESISTING STEELS, *BR. WELD. J.*, VOL 9, 1962, P 436-445
50. J.J. VAGI, R.M. EVANS, AND D.C. MARTIN, "WELDING PRECIPITATION-HARDENING STAINLESS STEELS," NASA TECHNICAL MEMORANDUM X-53582, 28 FEB 1967
51. B.I. MEDOVAR, ON THE NATURE OF WELD HOT CRACKING, TRANSLATED FROM *AVTOM. SVARKA*, VOL 7 (NO. 4), 1954, P 12-28
52. M.J. STRUM, L.T. SUMMERS, AND J.W. MORRIS, JR., THE AGING RESPONSE OF A WELDED IRON-BASED SUPERALLOY, *WELD. J.*, VOL 62, SEPT 1983, P 235S-242S

Selection of Cast Stainless Steels

Michael J. Cieslak, Sandia National Laboratories

Introduction

THE DEVELOPMENT OF CASTING TECHNOLOGY has been driven by two distinct engineering perspectives. The first involves the potential economy of cast hardware relative to totally machined wrought parts. Complex, nearnet shapes can be produced by a variety of casting processes, including sand casting, investment casting, and centrifugal casting (used almost exclusively for making tubular products). Casting also can be an extremely effective way of producing very large parts.

The second perspective is associated with metallurgical realities. Many alloys have been developed for which there are no true wrought counterparts. The chemical composition and microstructure of these alloys is such that deformation processing of large ingots into wrought product forms would be problematic. In addition, the very nature of the cast microstructure, which in general is composed of large grains, can provide benefits in high-temperature creep applications.

Some common examples of stainless steel castings include valve and pump bodies, compressor wheels, propellers, turbine impellers, large pipe elbows, heat treating fixtures, and food, paper, and chemical processing equipment. Extensive reviews on the properties and applications of cast stainless steels can be found in Ref 1, 2, 3, and 4.

Acknowledgement

This work was performed at Sandia National Laboratories, supported by the U.S. Department of Energy under contract number DE-AC04-76DP00789.

References

1. J.M. SVOBODA, HIGH-ALLOY STEELS, *CASTING,* 9TH ED., VOL 15, *METALS HANDBOOK,* ASM INTERNATIONAL, 1988, P 722-735
2. M. BLAIR, CAST STAINLESS STEELS, *PROPERTIES AND SELECTION: IRONS, STEELS, AND HIGH-PERFORMANCE ALLOYS,* VOL 1, *ASM HANDBOOK,* ASM INTERNATIONAL, 1990, P 908-929
3. R.W. MONROE AND S.J. PAWEL, CORROSION OF CAST STEELS, *CORROSION,* 9TH ED., VOL 13, *METALS HANDBOOK,* ASM INTERNATIONAL, 1987, P 573-582
4. P.R. WEISER, *ED., STEEL CASTINGS HANDBOOK,* 5TH ED., STEEL FOUNDERS' SOCIETY OF AMERICA, 1980

Selection of Cast Stainless Steels

Michael J. Cieslak, Sandia National Laboratories

Stainless Steel Casting Alloys

Categories. Unlike the case of wrought alloy designators, cast stainless steels are not grouped according to microstructural types, but by the intended service environment. The general classification of stainless steel castings places these alloys in one of two categories: corrosion-resistant alloys and heat-resistant alloys. The corrosion-resistant alloys have been designed primarily for service environments involving aqueous or liquid-vapor corrosives at temperatures generally less than 315 °C (600 °F). Use at extended temperatures up to 650 °C (1200 °F) in selected environments may also be considered.

The heat-resistant alloys have been designed for use in service environments exceeding 650 °C (1200 °F) where aqueous corrosives do not exist. At these elevated temperatures, oxidation, sulfidation, and carburization/decarburization are the environmental effects of consequence. High-temperature mechanical strength, creep and stress-rupture resistance, and microstructural stability are the engineering requirements for these alloys. The high-temperature alloys are most readily distinguished from the corrosion-resistant alloys on the basis of carbon content (although other alloying elements may also vary considerably in response to service needs) with the high-temperature alloys containing 0.20 to 0.70 wt% C and the corrosion-resistant alloys generally containing less than 0.20 wt% C.

Nomenclature. Cast stainless steels are most often specified on the basis of composition using the designation system of the High Alloy Product Group of the Steel Founders' Society of America. (The High Alloy Product Group has replaced the Alloy Casting Institute, or ACI, which formerly administered these designations.) The first letter of the designation indicates whether the alloy is intended primarily for liquid corrosion service (C) or high-temperature service (H). The second letter in either group of alloys indicates an approximate ratio of nickel-to-chromium with the ratio increasing in alphabetical order. (As the nickel content increases, the second letter of the designation progresses through the alphabet from A to Z.) The number immediately following the hyphen indicates the carbon content of the alloy in hundredths of a weight percent (also known as "points" of carbon). For the corrosion-resistant alloys, the carbon content indicated is the maximum allowable for the alloy. For the heat-resistant alloys, the carbon content identified is the midpoint with an allowable range of ± 0.05 wt%. Finally, if further alloying elements are present, these are indicated by the addition of one or more letters as a suffix. Thus, the designation of CF-8M refers to an alloy for corrosion-resistant service (C) of the 19Cr-9Ni type with a maximum carbon content of 0.08% that contains molybdenum (M). Figure 1 graphically shows the range of nickel and chromium contents in several grades of corrosion-resistant and heat-resistant alloys identified by the aforementioned designation system. Tables 1 and 2 list the chemical compositions of the common corrosion-resistant and heat-resistant alloys, respectively.

TABLE 1 COMPOSITIONS AND TYPICAL MICROSTRUCTURES OF CORROSION-RESISTANT STAINLESS STEEL CASTING ALLOYS

ALLOY DESIGNATION	**WROUGHT ALLOY TYPE(A)**	**ASTM SPECIFICATIONS**	**MOST COMMON END-USE MICROSTRUCTURE**	**COMPOSITIONS, %(B)**					
				C	**Mn**	**Si**	**Cr**	**Ni**	**OTHERS(C)**
CHROMIUM STEELS									
CA-15	410	A 743, A 217, A 487	MARTENSITE	0.15	1.00	1.50	11.5-14.0	1.0	0.50 MO(D)
CA-15M	. . .	A 743	MARTENSITE	0.15	1.00	0.65	11.5-14.0	1.0	0.15-1.00 MO
CA-40	420	A 743	MARTENSITE	0.40	1.00	1.50	11.5-14.0	1.0	0.5 MO(D)
CA-40F	. . .	A 743	MARTENSITE	0.2-0.4	1.00	1.50	11.5-14.0	1.0	. . .
CB-30	431, 442	A 743	FERRITE AND CARBIDES	0.30	1.00	1.50	18.0-22.0	2.0	. . .
CC-50	446	A 743	FERRITE AND CARBIDES	0.30	1.00	1.50	26.0-30.0	4.0	. . .
CHROMIUM-NICKEL STEELS									
CA-6N	. . .	A 743	MARTENSITE	0.06	0.50	1.00	1.0-12.5	6.0-8.0	. . .
CA-6NM	. . .	A 743, A 487	MARTENSITE	0.06	1.00	1.00	11.5-14.0	3.5-4.5	0.4-1.0 MO
CA-28MWV	. . .	A 743	MARTENSITE	0.20-0.28	0.50-1.00	1.00	11.0-12.5	0.50-1.00	0.9-1.25 MO; 0.9-1.25 W; 0.2-0.3 V
CB-7CU-1	. . .	A 747	MARTENSITE, AGE HARDENABLE	0.07	0.70	1.00	15.5-17.7	3.6-4.6	2.5-3.2 CU; 0.20-0.35 NB; 0.05 N MAX
CB-7CU-2	. . .	A 747	MARTENSITE, AGE HARDENABLE	0.07	0.70	1.00	14.0-15.5	4.5-5.5	2.5-3.2 CU; 0.20-0.35 NB; 0.05 N MAX
CD-4MCU	. . .	A 351, A 743, A 744, A 890	AUSTENITE IN FERRITE, AGE HARDENABLE	0.04	1.00	1.00	25.0-26.5	4.75-6.0	1.75-2.25 MO; 2.75-3.25 CU
CE-30	312	A 743	FERRITE IN AUSTENITE	0.30	1.50	2.00	26.0-30.0	8.0-11.0	. . .
CF-3(E)	304L	A 351, A 743, A 744	FERRITE IN AUSTENITE	0.03	1.50	2.00	17.0-	8.0-	. . .

							21.0	12.0	
CF-3M(E)	316L	A 351, A 743, A 744	FERRITE IN AUSTENITE	0.03	1.50	2.00	17.0-21.0	8.0-12.0	2.0-3.0 MO
CF-3MN	. . .	A 743	FERRITE TN AUSTENITE	0.03	1.50	1.50	17.0-21.0	9.0-13.0	2.0-3.0 MO; 0.10-0.20 N
CF-8(E)	304	A 351, A 743, A 744	FERRITE IN AUSTENITE	0.08	1.50	2.00	18.0-21.0	8.0-11.0	. . .
CF-8C	347	A 351, A 743, A 744	FERRITE IN AUSTENITE	0.08	1.50	2.00	18.0-21.0	9.0-12.0	NB(F)
CF-8M	316	A 351, A 743, A 744	FERRITE IN AUSTENITE	0.08	1.50	2.00	18.0-21.0	9.0-12.0	2.0-3.0 MO
CF-10	. . .	A 351	FERRITE IN AUSTENITE	0.04-0.10	1.50	2.00	18.0-21.0	8.0-11.0	. . .
CF-10M	. . .	A 351	FERRITE IN AUSTENITE	0.04-0.10	1.50	1.50	18.0-21.0	9.0-12.0	2.0-3.0 MO
CF-10MC	. . .	A 351	FERRITE IN AUSTENITE	0.10	1.50	1.50	15.0-18.0	13.0-16.0	1.75-2.25 MO
CF-10SMNN	. . .	A 351, A 743	FERRITE IN AUSTENITE	0.10	7.00-9.00	3.50-4.50	16.0-18.0	8.0-9.0	0.08-0.18 N
CF-12M	316	. . .	FERRITE IN AUSTENITE OR AUSTENITE	0.12	1.50	2.00	18.0-21.0	9.0-12.0	2.0-3.0 MO
CF-16F	303	A 743	AUSTENITE	0.16	1.50	2.00	18.0-21.0	9.0-12.0	1.50 MO MAX; 0.20-0.35 SE
CF-20	302	A 743	AUSTENITE	0.20	1.50	2.00	18.0-21.0	8.0-11.0	. . .
CG-6MMN	. . .	A 351, A 743	FERRITE IN AUSTENITE	0.06	4.00-6.00	1.00	20.5-23.5	11.5-13.5	1.50-3.00 MO; 0.10-0.30 NB; 0.10-30 V; 0.20-40 N
CG-8M	317	A 351, A 743, A 744	FERRITE IN AUSTENITE	0.08	1.50	1.50	18.0-21.0	9.0-13.0	3.0-4.0 MO
CG-12	. . .	A 743	FERRITE IN AUSTENITE	0.12	1.50	2.00	20.0-23.0	10.0-13.0	. . .
CH-8	. . .	A 351	FERRITE IN AUSTENITE	0.08	1.50	1.50	22.0-26.0	12.0-15.0	. . .
CH-10	. . .	A 351	FERRITE IN AUSTENITE	0.04-0.10	1.50	2.00	22.0-26.0	12.0-15.0	. . .

CH-20	309	A 351, A 743	AUSTENITE	0.20	1.50	2.00	22.0-26.0	12.0-15.0	. . .
CK-3MCUN	. . .	A 351, A 743, A 744	FERRITE IN AUSTENITE	0.025	1.20	1.00	19.5-20.5	17.5-19.5	6.0-7.0 V; 0.18-0.24 N; 0.50-1.00 CU
CK-20	310	A 743	AUSTENITE	0.20	2.00	2.00	23.0-27.0	19.0-22.0	. . .
NICKEL-CHROMIUM STEEL									
CN-3M	. . .	A 743	AUSTENITE	0.03	2.00	1.00	20.0-22.0	23.0-27.0	4.5-5.5 MO
CN-7M	. . .	A 351, A 743, A 744	AUSTENITE	0.07	1.50	1.50	19.0-22.0	27.5-30.5	2.0-3.0 MO; 3.0-4.0 CU
CN-7MS	. . .	A 743, A 744	AUSTENITE	0.07	1.50	3.50(G)	18.0-20.0	22.0-25.0	2.5-3.0 MO; 1.5-2.0 CU
CT-15C	. . .	A 351	AUSTENITE	0.05-0.15	0.15-1.50	0.50-1.50	19.0-21.0	31.0-34.0	0.5-1.5 V

(A) TYPE NUMBERS OF WROUGHT ALLOYS ARE LISTED ONLY FOR NOMINAL IDENTIFICATION OF CORRESPONDING WROUGHT AND CAST GRADES. COMPOSITION RANGES OF CAST ALLOYS ARE NOT THE SAME AS FOR CORRESPONDING WROUGHT ALLOYS; CAST ALLOY DESIGNATIONS SHOULD BE USED FOR CASTINGS ONLY.

(B) MAXIMUM UNLESS A RANGE IS GIVEN. THE BALANCE OF ALL COMPOSITIONS IS IRON.

(C) SULFUR CONTENT IS 0.04% IN ALL GRADES EXCEPT: CG-6MMN, 0.030% S (MAX); CF-10SMNN, 0.03% S (MAX); CT-15C, 0.03% S (MAX); CK-3MCUN, 0.010% S (MAX); CN-3M 0.030% S (MAX), CA-6N, 0.020% S (MAX); CA-28MWV, 0.030% S (MAX); CA-40F, 0.20-0.40% S; CB-7CU-1 AND -2, 0.03% S (MAX). PHOSPHORUS CONTENT IS 0.04% (MAX) IN ALL GRADES EXCEPT; CF-16F, 0.17% P (MAX); CF-10SMNN, 0.060% P (MAX); CT-15C, 0.030% P (MAX); CK-3MCUN, 0.045% P (MAX); CN-3M, 0.030% P (MAX); CA-6N, 0.020% P (MAX); CA-28MWV, 0.030% P (MAX); CB-7CU-1 AND -2, 0.035% P (MAX).

(D) MOLYBDENUM NOT INTENTIONALLY ADDED.

(E) CF-3A, CF-3MA, AND CF-8A HAVE THE SAME COMPOSITION RANGES AS CF-3, CF-3M, AND CF-8, RESPECTIVELY, BUT HAVE BALANCED COMPOSITIONS SO THAT FERRITE CONTENTS ARE AT LEVELS THAT PERMIT HIGHER MECHANICAL PROPERTY SPECIFICATIONS THAN THOSE FOR RELATED GRADES. THEY ARE COVERED BY ASTM A 351.

(F) NB, 8 × %C MIN (1.0% MAX); OR NB + TA × %C (1.1% MAX).

(G) FOR CN-7MS, SILICON RANGES FROM 2.50 TO 3.50%.

TABLE 2 COMPOSITIONS OF HEAT-RESISTANT STAINLESS STEEL CASTING ALLOYS

ALLOY	**UNS**	**ASTM SPECIFICATIONS(A)**	**COMPOSITION, %(B)**

DESIGNATION	**NO.**		**C**	**Cr**	**Ni**	**Si (max)**
HA	. . .	A217	0.20 MAX	8-10	. . .	1.00
HC	J92605	A 297, A 608	0.50 MAX	26-30	4 MAX	2.00
HD	J93005	A 297, A 608	0.50 MAX	26-30	4-7	2.00
HE	J93403	A 297, A 608	0.20-0.50	26-30	8-11	2.00
HF	J92603	A 297, A 608	0.20-0.40	19-23	9-12	2.00
HH	J93503	A 297, A 608, A 447	0.20-0.50	24-28	11-14	2.00
HI	J94003	A 297, A 567, A 608	0.20-0.50	26-30	14-18	2.00
HK	J94224	A 297, A 351, A 567, A 608	0.20-0.60	24-28	18-22	2.00
HK30	. . .	A 351	0.25-0.35	23.0-27.0	19.0-22.0	1.75
HK40	. . .	A 351	0.35-0.45	23.0-27.0	19.0-22.0	1.75
HL	J94604	A 297, A 608	0.20-0.60	28-32	18-22	2.00
HN	J94213	A 297, A 608	0.20-0.50	19-23	23-27	2.00
HP	. . .	A 297	0.35-0.75	24-28	33-37	2.00
HP-50WZ(C)	. . .	. . .	0.45-0.55	24-28	33-37	2.50
HT	J94605	A 297, A 351, A 567, A 608	0.35-0.75	13-17	33-37	2.50
HT30	. . .	A 351	0.25-0.35	13.0-17.0	33.0-37.0	2.50
HU	. . .	A 297, A 608	0.35-0.75	17-21	37-41	2.50
HW	. . .	A 297, A 608	0.35-0.75	10-14	58-62	2.50
HX	. . .	A 297, A 608	0.35-0.75	15-19	64-68	2.50

(A) ASTM DESIGNATIONS ARE THE SAME AS ACI DESIGNATIONS.

(B) BALANCE IRON IN ALL COMPOSITIONS. MANGANESE CONTENT: 0.35 TO 0.65% FOR HA. 1% FOR HC. 1.5% FOR HD AND 2% FOR THE OTHER ALLOYS. PHOSPHORUS AND SULFUR CONTENTS: 0.04% (MAX) FOR ALL BUT HP-50WZ. MOLYBDENUM IS INTENTIONALLY ADDED ONLY TO HA, WHICH HAS 0.90 TO 1.20% MO; MAXIMUM FOR OTHER ALLOYS IS SET AT 0.5% MO. HH ALSO CONTAINS 0.2% N (MAX).

(C) ALSO CONTAINS 4 TO 6% W, 0.1 TO 1.0% ZR, AND 0.035% S (MAX) AND P (MAX)

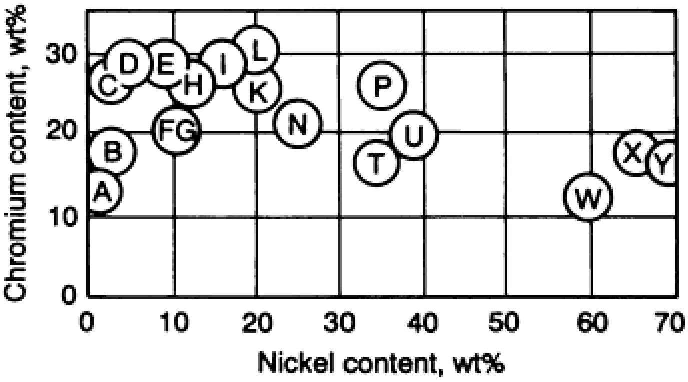

FIG. 1 SCHEMATIC OF THE CHROMIUM AND NICKEL CONTENT RANGES FOR CORROSION-RESISTANT AND HEAT-RESISTANT STAINLESS STEEL ALLOYS. THE LETTERS IN CIRCLES REPRESENT THE SECOND LETTER IN THE ALLOY DESIGNATION SYSTEM. SOURCE: REF 4

Reference cited in this section

4. P.R. WEISER, *ED., STEEL CASTINGS HANDBOOK*, 5TH ED., STEEL FOUNDERS' SOCIETY OF AMERICA, 1980

Selection of Cast Stainless Steels

Michael J. Cieslak, Sandia National Laboratories

Welding and Weldability of Cast Stainless Steels

Welding and Welding Processes. In general terms, there are two situations in which welding of stainless steel castings is required. One of these occurs before the casting ever leaves the foundry. Casting defects, such as hot tears, shrinkage cavities, and cold shuts, are often observed in stainless steel castings. Certain specifications preclude the acceptance of these castings for service use without repair. Fusion welding is an appropriate method for repair of these casting defects. The defects, identified through one of several possible procedures (for example, dye penetrant, radiography, or ultrasonic inspection), are removed by machining or grinding, leaving a clean joint surface onto which weld filler metal is added to produce a sound final product. For small, near-surface defects, autogenous welds (that is, no filler metal added) may be sufficient to eliminate the defects. A special case of repair welding is known by the term "upgrading." An "upgraded" casting is one that has been selectively welded in order to achieve a higher level of product specification acceptance.

The second situation requiring welding involves joining of the casting to other hardware, cast or wrought, as part of an engineering design. In this circumstance, the design rules for welding are similar to those for wrought alloys, as discussed in the articles on wrought stainless steels that appear in this Section of the Handbook. Any of the fusion welding processes can be successfully employed in the welding of castings. Most common are the arc welding processes, especially autogenous and filler-metal-added gas-tungsten arc welding, shielded metal arc welding, and gas-metal arc welding. Fabrication involving the use of extremely large castings may employ electroslag welding as the only reasonable economic choice. Other processes, such as electron-beam welding, laser-beam welding, and plasma arc welding, are also appropriate choices in many instances. A wide range of filler metals exists for joining of the various casting alloys. The choice depends on the specifics of the situation and whether the weld will be to a similar or dissimilar alloy. The use of matching filler metal is generally suggested when welding like alloys. Where a matching consumable is not available, common choices are 308L and 316L for the low-nickel-content corrosion-resistant austenitic alloys. For the higher-nickel-content corrosion-resistant and heat-resistant alloys, appropriate choices may be consumables (bare electrodes for gas-tungsten arc welding and gas-metal arc welding) such as 20Cb-3 (AWS ER320/ER320LR) and nickel alloys such as C-276 (AWS ERNiCrMo-4), Alloy 625 (ERNiCrMo-3), and Inconel alloys 82 (AWS ERNiCr-3), 92 (AWS ERNiCrFe-6), and 182 (AWS ERNiCrFe-3).

Metallurgical Considerations. Castings are metallurgically more complex than wrought alloys. Segregation during dendritic solidification can lead to local variations in both chemical composition and microstructure. The vast majority of the heat-resistant alloys will terminate solidification with the formation of a eutectic-like carbide constituent that remains upon cooling of the casting to room temperature. This microstructural constituent helps to augment the creep resistance of this group of alloys. Subsequent heat treatment of these alloys is not common.

Many of the lower-nickel-content corrosion-resistant alloys (for example, CF-3, CF-8M, and CB-7Cu) will have a duplex microstructure consisting of ferrite islands in a matrix of either austenite or martensite. Postcasting solution heat treatments used to prepare corrosion-resistant alloys for service are insufficient to homogenize the microstructure. The inhomogeneous microstructure of both groups of alloys will tend to promote relatively wide areas of partial melting adjacent to the fusion zone when compared with similar wrought alloys.

Martensitic Stainless Steel Castings. Special considerations exist for the welding of martensitic stainless steel castings (CA-6NM, CA-15, CA40, and CB-7Cu-1, and CB-7Cu-2). Quench cracking and a marked reduction in mechanical properties (especially ductility and impact toughness) may result from the welding of these alloys. Corrosion resistance may also be detrimentally affected by the thermal cycle that the material undergoes in the fusion zone and heat-affected-zone (HAZ) of welds.

Suggested welding practice often calls for the preheating and postheating of most of the martensitic alloys. Alloys CA-15 and CA-40 require preheating temperatures in the range of 200 to 315 °C (390 to 600 °F) and a postweld heat treatment (PWHT) temperature range of 610 to 760 °C (1130 to 1400 °F). Alloy CA-6NM requires a preheat temperature range of 100 to 150 °C (210 to 300 °F) and a PWHT temperature range of 590 to 620 °C (1095 to 1150 °F). The higher preheating temperatures should be used when welding thicker sections and higher carbon compositions of these alloys.

The PWHT for these alloys is a tempering operation intended to restore ductility and toughness at the expense of strength and hardness. Because of the broad range of potential tempering temperatures, a wide range of PWHT properties are possible.

Alloy CB-7Cu (1 and 2) has no suggested preheat temperature, but a range (480 to 590 °C, or 895 to 1095 °F) of PWHT temperatures is recommended. In the case of this alloy, the PWHT is an aging operation. In order to recover properties in the welded product, a full solution anneal (1040 °C, or 1095 °F, oil quench), followed by aging, is required. Direct aging without solution annealing will not fully restore preweld properties. The higher aging temperatures result in the lowest strength but also provide the highest ductility and toughness.

Welding Defects. The two most serious problems encountered in the welding of stainless steel castings, either during the repair of casting defects or for subsequent attachment to other structures, are solidification hot cracking and heat-affected zone (HAZ) hot cracking. Figure 2 shows a prototypical hot crack extending both into the fusion zone and the HAZ of a casting alloy. Both of these phenomena involve separation of grain boundaries due to the presence of a wetting liquid phase and sufficient mechanical imposition.

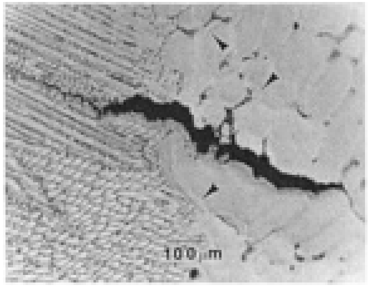

FIG. 2 OPTICAL MICROGRAPH OF THE FUSION LINE REGION IN A CAST CORROSION-RESISTANT AUSTENITIC STAINLESS STEEL. NOTE THE HOT CRACK EXTENDING BOTH INTO THE FUSION ZONE (LEFT) AND BACK INTO THE PARTIALLY MELTED REGION OF THE HAZ. ARROWS INDICATE REGIONS OF PARTIAL MELTING.

Studies (Ref 5, 6, 7, 8, 9) have indicated that wholly austenitic stainless steel casting alloys are more susceptible to the formation of these types of defects than are duplex ferrite plus austenite alloys. In particular, Cieslak (Ref 7, 8) and Cadden (Ref 9) have shown that CF-8M alloys (being prototypic of a wide range of low-nickel corrosion-resistant casting alloys) having welds that solidify with ferrite as the primary solidification phase are much more resistant to hot cracking than are alloys that solidify with austenite as the primary solidification phase.

Table 3 lists the chemical compositions and ferrite content of a series of CF-8M castings used in a study of weldability (Ref 8). Figure 3 shows the fusion zone hot cracking susceptibility of these alloys determined using the Varestraint test at $\frac{1}{2}$ % applied strain. The alloys that solidified as primary austenite (γ) were more susceptible to cracking than those which solidified as primary ferrite (δ). With the exception of heat number 7, the primary ferrite alloys did not experience any cracking under this set of experimental conditions. The high level of sulfur (0.032 wt%) and phosphorus (0.046 wt%) were identified as being responsible for the greater cracking sensitivity in heat number 7. This particular observation was important for establishing that ferrite content alone was not a sufficient criterion to preclude the possibility of hot cracking in duplex alloys.

TABLE 3 COMPOSITIONS OF ALLOYS IN CF-8M WELDABILITY STUDY

HEAT NO.	COMPOSITION, WT%$^{(A)}$									FERRITE, VOL%
	C	Mn	Si	Cr	Ni	Mo	S	P	N	
1	0.08	0.60	1.05	18.32	13.20	2.26	0.016	0.035	0.041	0.17
2	0.06	0.44	1.08	21.10	9.55	2.52	0.008	0.021	0.048	22.8
3	0.05	0.26	0.83	17.65	12.03	2.05	0.017	0.021	0.06	1.48
4	0.06	1.17	0.60	18.18	12.08	2.48	0.016	0.027	0.05	4.16

5	0.07	0.57	0.99	18.21	9.57	2.39	0.019	0.024	0.06	7.60
6	0.04	0.21	0.38	19.55	15.38	2.88	0.025	0.036	0.04	2.00
7	0.10	0.30	0.69	21.31	10.62	2.34	0.032	0.046	0.05	14.24

Source: Ref 8

(A) THE BALANCE IS IRON.

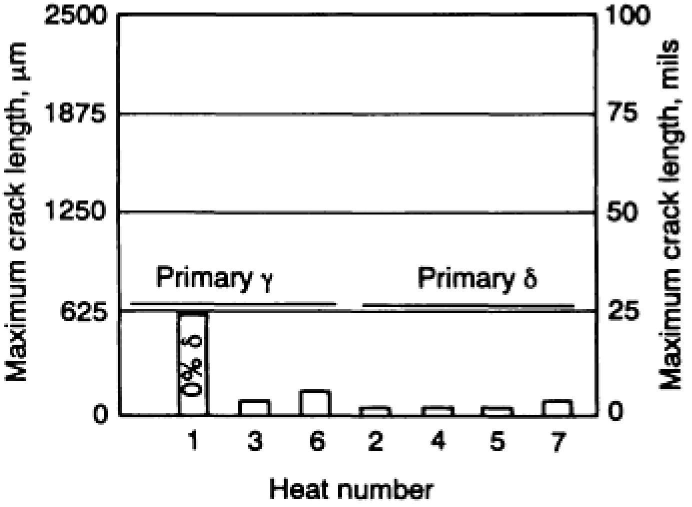

FIG. 3 VARESTRAINT TEST RESULTS ON SEVERAL HEATS OF CF-8M CORROSION-RESISTANT ALLOY. TESTS WERE RUN AT $\frac{1}{2}$ % AUGMENTED STRAIN USING AN ENERGY INPUT OF APPROXIMATELY 2.7 KJ/MM (69 KJ/IN.). SEE TABLE 3 FOR HEAT NUMBER CHEMICAL COMPOSITIONS.

In the same investigation, the hot cracking sensitivity of the heat-resistant alloy HK-40 was also examined. This alloy solidifies as primary austenite and terminates solidification with a eutectic-like Cr_7C_3 carbide constituent. This is prototypical solidification behavior of a wide range of heat-resistant stainless steel casting alloys. It was found that cracking was exacerbated by low carbon content (0.38 wt%), and high levels of combined sulfur and phosphorus (>0.026 wt%). Silicon addition also had a detrimental effect on cracking susceptibility when either carbon content was low or sulfur or phosphorus was high. It was observed that hot cracks could be "healed" by the carbide eutecticlike constituent, suggesting that higher carbon contents in alloys of this type could be helpful in minimizing this type of defect. Similar results have been found for 16Cr-35Ni (HT type) (Ref 10) and high-carbon-content alloys such as 25Cr-20Ni (the cast equivalent of wrought type 310 containing up to 0.2 wt% C) (Ref 11).

A study of the weldability of alloy CN-7M identified sulfur, phosphorus, and silicon as being detrimental to the hot cracking resistance of this high-nickel-content (29 wt% Ni) corrosion-resistant alloy (Ref 12). Specifically, silicon was found to promote the formation of an M_6C carbide eutecticlike constituent associated with intergranular hot cracked regions. For a discussion of the hot cracking susceptibility of a wide range of corrosion-resistant and heat-resistant casting alloys, along with microstructural analysis of the cracking behavior, a review of the work of Cadden (Ref 9) is suggested.

In many studies examining both the fusion zone and HAZ hot cracking behavior of stainless steel casting alloys, the presence of sulfur and phosphorus has universally been identified with increased cracking susceptibility. To a somewhat lesser extent, so have silicon and niobium. Carbon can also be detrimental, although not all studies have reached the same conclusion. It can be said that if there is sufficient carbon present to result in the formation of terminal-stage, eutectic-like constituents, then the hot cracking susceptibility will increase up to the point where the level of carbon promotes eutectic healing. Eutectic healing may eliminate the hot cracking problem, but the reduction in ductility with increasing carbon content can limit the utility of the weld joint for an engineering application (Ref 10).

Each of these elements appears to increase cracking susceptibility in two possible ways. The first is by promoting the formation of low-melting-point constituents, and the second is by acting as a surfactant which promotes wetting of solidifying grain boundaries. As discussed elsewhere in this Volume (see, for example, the articles in this Section of the

Handbook that describe wrought stainless steels), either phenomenon will promote an increased tendency toward fusion zone or HAZ hot cracking.

A specific concern in the welding of corrosion-resistant stainless steel castings is the production of a sensitized microstructure. Sensitization occurs in stainless steel alloys when chromium-rich phases--$M_{23}C_6$ carbides, sigma (σ), chi (χ), and Laves phases--precipitate during exposure to temperatures between 400 and 900 °C (750 and 1650 °F). The precipitation of these constituents can result in a local depletion of chromium (and also molybdenum), lowering the corrosion resistance to a point where corrosion-induced failure of the casting can result. Primary among the phases of concern is the $M_{23}C_6$ carbide (with M being primarily chromium). Precipitation of carbides at grain boundaries can occur during the fusion welding cycle at sites in both the HAZ and the fusion zone. Because the kinetics of precipitation are proportional to the carbon concentration, low-carbon grades (such as CF-3 or CF-3M) have been developed that can experience welding without producing a sensitized microstructure.

A final metallurgical consideration involves the welding of castings that have seen extended service at relatively high temperatures. For a variety of reasons, castings may have to be rewelded sometime during their service lifetime. This can pose a difficult problem if the temperature environment has modified the alloy microstructure so that additional constituents are present that adversely affect alloy ductility. It is always preferable to heat treat the casting to remove these unwanted microstructural features. This process may require solutionizing at temperatures above 1000 °C (1830 °F) and may be impractical for large castings. If solution heat treatment is not an option, then special care must be taken to ensure that the new weld joint design will have the lowest possible level of inherent restraint.

Weld Parameters. From an understanding of the metallurgical behavior of stainless steel casting alloys given above, general rules for fusion welding can be inferred. Low-heat-input processes are almost universally suggested. For alloys that are far from homogeneous, a high-heat-input process would produce flatter thermal gradients outside of the fusion zone, creating a wider zone of partial melting and a greater microstructural sensitivity to HAZ hot cracking. The greater total shrinkage strain associated with the solidification of a large (high-energy-input) weld can create a restraint situation more likely to result in fusion zone solidification cracking. The flatter thermal gradients of higher-energy-input welds reduce the cooling rates through the carbide precipitation temperature range, resulting in a higher likelihood of a sensitized structure in the corrosion-resistant alloys.

For the same reasons, low interpass temperatures are generally employed in the welding of these alloys. Interpass temperatures below 200 °C (400 °F) are common. For critical applications (especially among corrosion-resistant alloys), interpass temperatures of less than 150 °C (300 °F) are often specified. Spraying castings with water between passes is not uncommon for the austenitic corrosion-resistant alloys (Ref 4).

Preheating of castings prior to welding is almost never required and generally not suggested, at least for the corrosion-resistant alloys, because the sensitization impact of such processing is similar to the use of high-energy-input welding parameters. In the case of heat-resistant alloy castings, preheating can be used to reduce the magnitude of thermal shrinkage strains, but again this benefit comes at the expense of a wider zone of partial melting. There is no known formula for preheating temperature and welding parameters for a given alloy that can ensure the preclusion of welding defects. Because of this technological limitation, the use of preheating prior to welding of these alloys is questionable.

Postweld solutionizing of corrosion-resistant alloys is always recommended. For the heat treatable alloys (for example, CB-7Cu), a complete solutionizing and aging heat treatment is recommended to reestablish baseline properties. Where PWHT is not possible, it is even more important to weld using low-heat-input parameters.

Where direct fusion welding is not appropriate, as in the case of an extremely hot-crack-sensitive casting, other approaches might prove useful. One option is to "butter" the area of the crack-sensitive casting to be joined, using a consumable of a less crack-sensitive material. This "buttering" involves surfacing the casting locally with a consumable before fabricating the actual weld joint. With no joint present during the buttering operation, the restraint on the solidifying surface pass and the underlying HAZ of the casting can be quite small and defects (fusion zone and HAZ hot cracks) may be avoided. This surfaced area can be built up slowly to a thickness where it can then be prepared as a joint for subsequent welded attachment. Additional information on buttering can be found in the article "Hardfacing, Weld Cladding, and Dissimilar Metal Joining" in this Volume. As another option, an insert of a less hot-crack-sensitive alloy can be inertia or friction welded onto the mating surface of the hot-crack-sensitive casting. This insert could then be prepared for subsequent fusion welding to the required structure. These options are contingent upon the availability of consumable materials that are acceptable in the service environment.

References cited in this section

4. P.R. WEISER, *ED., STEEL CASTINGS HANDBOOK,* 5TH ED., STEEL FOUNDERS' SOCIETY OF AMERICA, 1980
5. H.S. AVERY, CAST HEAT-RESISTANT ALLOYS FOR HIGH-TEMPERATURE SERVICE, *WELD. RES. COUNC. BULL.,* NO. 143, AUG 1969
6. "THE WELDING OF CAST AUSTENITIC STEELS," WELDING INSTITUTE REPORT SERIES, THE WELDING INSTITUTE, SEPT 1970, PHASES 1-4
7. M.J. CIESLAK AND W.F. SAVAGE, *WELD. J.,* MAY 1980, P 136-S TO 146-S
8. M.J. CIESLAK AND W.F. SAVAGE, "WELDABILITY AND SOLIDIFICATION PHENOMENA OF CAST HIGH ALLOY HEAT RESISTANT ALLOY HK-40 AND CORROSION RESISTANT 19CR-9NI-2.5MO STEELS," REPORT A-69, STEEL FOUNDERS' SOCIETY OF AMERICA, 1980
9. C.H. CADDEN, "WELDABILITY OF HIGH ALLOY CASTINGS," M.S. THESIS, RENSSELAER POLYTECHNIC INSTITUTE, TROY, NEW YORK, MAY 1981
10. D. ROZET, H.C. CAMPBELL, AND R.D. THOMAS, *WELD. J.,* OCT 1949, P 481-S TO 491-S
11. Y. ARATA, F. MATSUDA, AND S. KATAYAMA, *TRANS. JWRI,* JUNE 1977, P 105-116
12. M.J. CIESLAK AND W.F. SAVAGE, *WELD. J.,* MAY 1985, P 119-S TO 126-S

Selection of Cast Stainless Steels

Michael J. Cieslak, Sandia National Laboratories

References

1. J.M. SVOBODA, HIGH-ALLOY STEELS, *CASTING,* 9TH ED., VOL 15, *METALS HANDBOOK,* ASM INTERNATIONAL, 1988, P 722-735
2. M. BLAIR, CAST STAINLESS STEELS, *PROPERTIES AND SELECTION: IRONS, STEELS, AND HIGH-PERFORMANCE ALLOYS,* VOL 1, *ASM HANDBOOK,* ASM INTERNATIONAL, 1990, P 908-929
3. R.W. MONROE AND S.J. PAWEL, CORROSION OF CAST STEELS, *CORROSION,* 9TH ED., VOL 13, *METALS HANDBOOK,* ASM INTERNATIONAL, 1987, P 573-582
4. P.R. WEISER, *ED., STEEL CASTINGS HANDBOOK,* 5TH ED., STEEL FOUNDERS' SOCIETY OF AMERICA, 1980
5. H.S. AVERY, CAST HEAT-RESISTANT ALLOYS FOR HIGH-TEMPERATURE SERVICE, *WELD. RES. COUNC. BULL.,* NO. 143, AUG 1969
6. "THE WELDING OF CAST AUSTENITIC STEELS," WELDING INSTITUTE REPORT SERIES, THE WELDING INSTITUTE, SEPT 1970, PHASES 1-4
7. M.J. CIESLAK AND W.F. SAVAGE, *WELD. J.,* MAY 1980, P 136-S TO 146-S
8. M.J. CIESLAK AND W.F. SAVAGE, "WELDABILITY AND SOLIDIFICATION PHENOMENA OF CAST HIGH ALLOY HEAT RESISTANT ALLOY HK-40 AND CORROSION RESISTANT 19CR-9NI-2.5MO STEELS," REPORT A-69, STEEL FOUNDERS' SOCIETY OF AMERICA, 1980
9. C.H. CADDEN, "WELDABILITY OF HIGH ALLOY CASTINGS," M.S. THESIS, RENSSELAER POLYTECHNIC INSTITUTE, TROY, NEW YORK, MAY 1981
10. D. ROZET, H.C. CAMPBELL, AND R.D. THOMAS, *WELD. J.,* OCT 1949, P 481-S TO 491-S
11. Y. ARATA, F. MATSUDA, AND S. KATAYAMA, *TRANS. JWRI,* JUNE 1977, P 105-116
12. M.J. CIESLAK AND W.F. SAVAGE, *WELD. J.,* MAY 1985, P 119-S TO 126-S

Dissimilar Welds With Stainless Steels

Joel G. Feldstein, Foster Wheeler Energy Corporation

Introduction

THE WELDING OF DISSIMILAR AUSTENITIC STAINLESS STEELS to each other is fairly common practice. When suitable welding procedures and filler metals are employed, most austenitic stainless steels can also be welded satisfactorily to several other classes of weldable steel, including ferritic and precipitation-hardening stainless steels, carbon steels, and low-alloy steels.

As a general rule, the weld metal should be at least equal in strength and corrosion properties to the poorest component in the joint. Additionally, there should be no intermetallic compounds or other phases to degrade the properties of the weld metal. Further, the microstructure must be resistant to cracking. An austenitic microstructure with a Ferrite Number (FN) of 5 to 10 has been shown to be highly resistant to cracking. If a fully austenitic weld metal occurs, satisfactory resistance to cracking can be achieved in certain applications by using a filler metal of high manganese content or a nickel-base alloy.

Dissimilar Welds With Stainless Steels

Joel G. Feldstein, Foster Wheeler Energy Corporation

Welding Dissimilar Austenitic Stainless Steels

In general, the deposited weld metal composition should nearly match the base metal composition when welding different austenitic stainless steels to themselves. However, it is not uncommon to weld different austenitic stainless steels, such as types 304 and 347, to each other or to a nickel-base alloy using a filler metal that differs in chemical composition from both base metals provided the weld metal will have suitable corrosion resistance or mechanical properties (or both) for the required service. Generally, the alternate filler metal should be more highly alloyed than the base metal, but the ferrite content of the resulting weld metal should be considered for corrosion or high-temperature service. Some filler metals may give too high a ferrite content, particularly the high-chromium, low-nickel types, such as type 312. Recommended filler metals for welding dissimilar austenitic stainless steels are given in Table 1. Additional information on stainless steel filler metals can be found in the article "Welding of Stainless Steels" in this Volume.

TABLE 1 STAINLESS STEEL FILLER METALS FOR WELDING SIMILAR AND DISSIMILAR AUSTENITIC STAINLESS STEELS

BASE METALS	**FILLER METALS**									
	201, 202, 301, 302, 302B, 303, 304, 305, 308	**304L**	**309 309S**	**310 310S, 314**	**316**	**316L**	**317**	**317L**	**321, 347, 348**	**330**
201,202, 301,302, 302B, 303,304, 305,308	E308(A)	E308(A)	E308(A)	E308(A)	E308(A)	E308(A)	E308(A)	E308(A)	E308(A)	E309
340L		E308L	E308(A)	E308(A)	E308(A)	E308(A)	E308(A)	E308(A)	E308(A)	309
309, 309S			E309	E309	E309	E309	E309	E309	E309	E309
310, 310S, 314				310	E316(B)	E316(B)	E317	E317	E308(A)	310
316					E316(B)	E316(B)	E316(B)	E316(B)	E308(A)	E309MO
316L						E316L	E316(B)	E316L	E316L	E309MO
317							E317	E317	E308(A)	E309MO
317L								E417L	308L	E309MO
321,347,348									E347(A)	E309
330										E330

(A) FOR SERVICE TEMPERATURE OF <370 °C (700 °F), E308L CAN BE USED.

(B) FOR SERVICE TEMPERATURE OF <370 °C (700 °F), E316L CAN BE USED.

Dissimilar Welds With Stainless Steels

Joel G. Feldstein, Foster Wheeler Energy Corporation

Welding Austenitic Stainless Steels to Carbon or Low-Alloy Steels

In joining austenitic stainless steel to carbon or low-alloy steel for service applications involving exposure to temperatures not exceeding 370 °C (700 °F), it is good practice to use a stainless steel filler metal with a total alloy content high enough to prevent the formation of martensite in the weld after dilution by the base metal and to preserve residual amounts of ferrite to minimize the possibility of hot cracking resulting from welding under severe restraint. Dilution is the change in chemical composition of a welding filler metal caused by the admixture of the base metal or previously deposited weld metal in the deposited weld bead. It is normally measured by the percentage of base metal or previously deposited weld metal in the weld bead. Dilution control, which is of primary concern when joining dissimilar metal combinations, is described in detail in the article "Hardfacing, Weld Cladding, and Dissimilar Metal Joining" in this Volume.

The Schaeffler diagram is particularly useful when examining what filler metal is suitable for joining dissimilar metals. The example in Fig. 1 shows the joining of a carbon steel (point A) to a type 304 austenitic stainless steel (point B) using type 309 as the filler metal (point D). Point C shows what the composition of the weld metal would be if these items were joined without a filler metal. If we assume that each base metal is fused to the same extent, point C will lie halfway between A and B. Because the welding is done using type 309 filler metal, the composition of the weld metal will lie along the line CD, depending on the degree of dilution. At point E a suitable weld metal composition is obtained, that is, an austenitic structure with 8 to 9% ferrite (Ferrite Number). This weld metal will be crack resistant, in contrast to that obtained at point C, which is very sensitive to cracking because of martensite formation. Types 309 and 309L (25Cr-12Ni) filler metals are most widely used for joining carbon or low-alloy steel to austenitic stainless steel; they normally contains about 8 to 15 FN. Types 304Cb, 309Mo, 309MoL, and 312 (29Cr-9Ni) are progressively more strongly ferritic. Satisfactory welds are also obtained with nickel-chromium-iron filler metal and these filler metals allow the service temperature to exceed 370 °C (700 °F) and minimize some stress relieving problems.

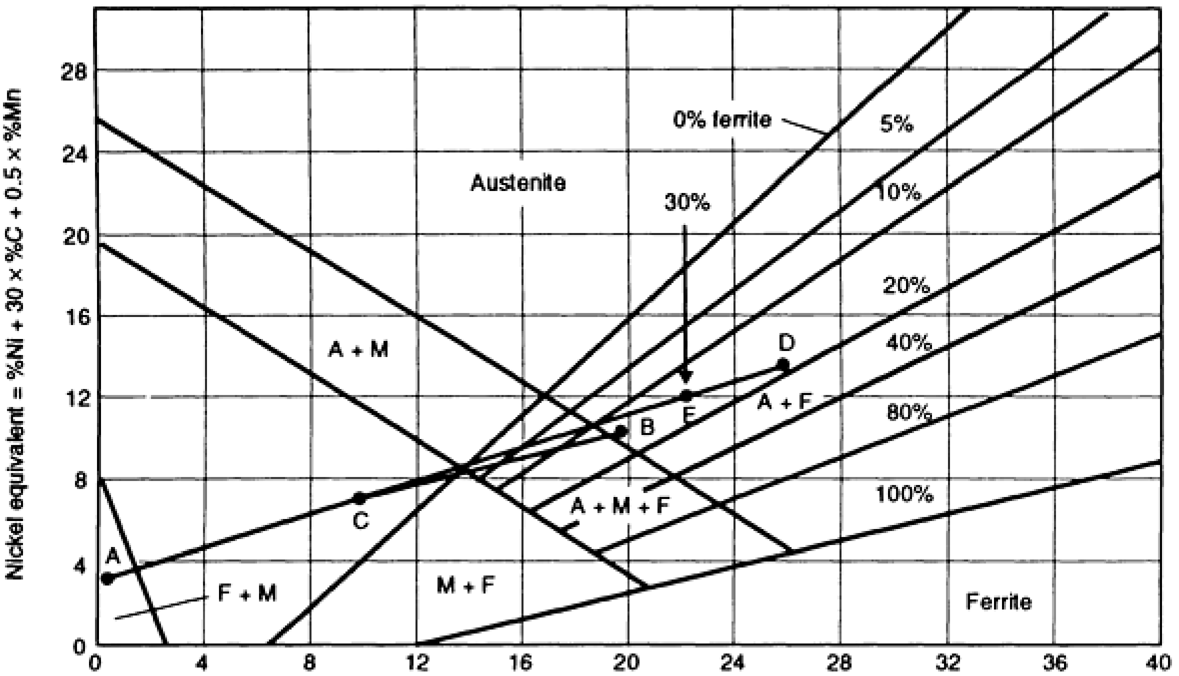

FIG. 1 THE SCHAEFFLER DIAGRAM USED FOR PREDICTING WELD MICROSTRUCTURE USING TYPE 309 TO JOIN

TYPE 304 STAINLESS STEEL TO CARBON STEEL. SEE TEXT FOR DETAILS.

Although the Schaeffler diagram is still widely used to predict the ferrite content of dissimilar metal weld deposits, more recently developed diagrams have increased the scope and accuracy of FN prediction in stainless steel weld metal and related dissimilar metal joints. These include the DeLong diagram, which is described later in this article, and the Welding Research Council 1988 and 1992 diagrams (WRC-1988 and WRC-1992). A review of each of the aforementioned diagrams can be found in the article "Hardfacing, Weld Cladding, and Dissimilar Metal Joining" in this Volume.

Specific Welding Practices. In joining austenitic stainless steel to carbon steel, it is good practice to first surface, or "butter," the carbon steel with a layer of type 309 or other suitable stainless steel weld metal, so that the portion of the weld (on the carbon steel member) where difficulties are most likely to occur is deposited with stainless steel filler metal while there is little restraint on the weld metal. Thus, following deposition and inspection of the buttered surface, the joint between the stainless steel member and the surfaced layer becomes a stainless-to-stainless joint that can be welded using a conventional stainless steel filler metal. The deposition of carbon steel or low-alloy steel filler metal on austenitic stainless steel should be avoided whenever possible because it can result in hard, brittle weld deposits.

A procedure for joining stainless steel to carbon steel or stainless steel to stainless-clad carbon steel is shown in Fig. 2. This procedure is widely used in the welding of stainless steel pipe to carbon steel, low-alloy steel, and stainless-clad carbon steel. Weld cladding procedures are described later in this article.

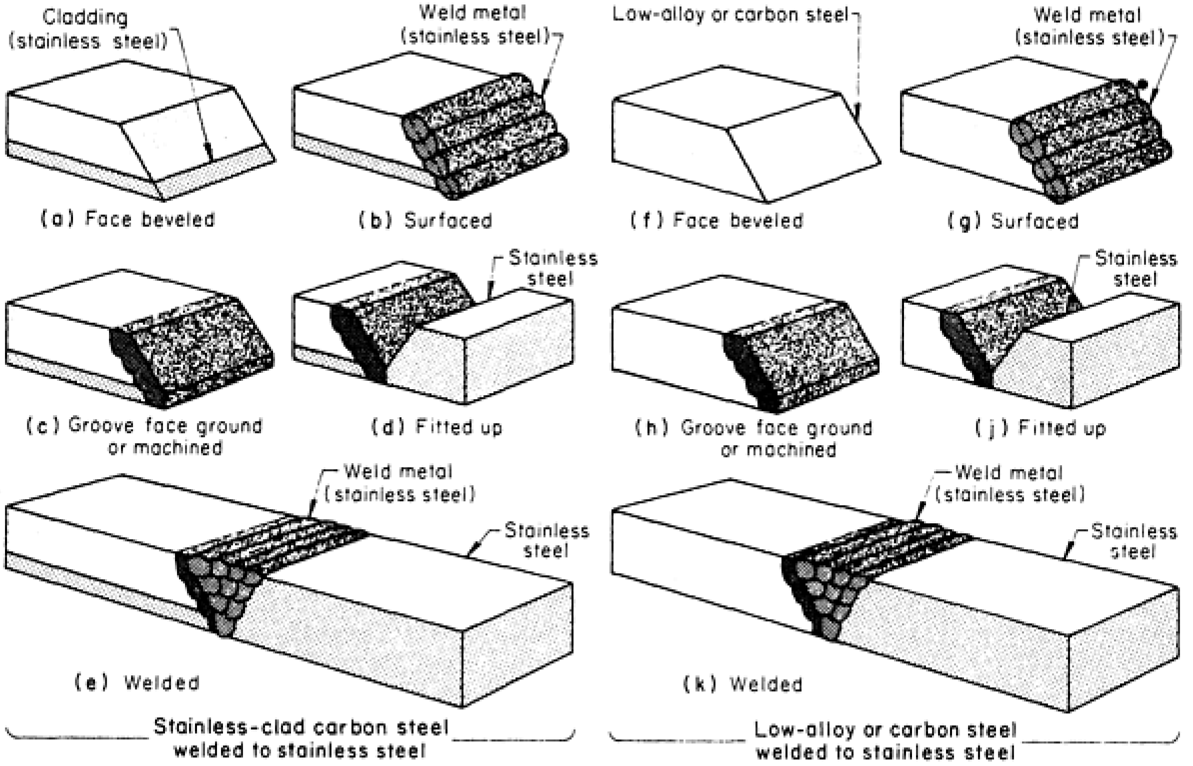
Cladding (stainless steel)
Weld metal (stainless steel)
Low-alloy or carbon steel
Weld metal (stainless steel)
(a) Face beveled
(b) Surfaced
(f) Face beveled
(g) Surfaced
Stainless steel
Stainless steel
(c) Groove face ground or machined
(d) Fitted up
(h) Groove face ground or machined
(j) Fitted up
Weld metal (stainless steel)
Stainless steel
Weld metal (stainless steel)
Stainless steel
(e) Welded
(k) Welded
Stainless-clad carbon steel welded to stainless steel
Low-alloy or carbon steel welded to stainless steel

FIG. 2 PROCEDURES FOR JOINING AUSTENITIC STAINLESS STEEL TO STAINLESS-CLAD CARBON STEEL, CARBON STEEL, AND LOW-ALLOY STEEL. STAINLESS-CLAD, LOW-ALLOY, OR CARBON STEEL EDGES ARE BEVELED FOR WELDING (A AND F). AN OVERLAY (OR "BUTTERING" LAYER) OF STAINLESS STEEL FILLER METAL IS APPLIED TO THE BEVELED EDGE (B AND G). THE LAYER IS MACHINED OR GROUND TO REQUIRED DIMENSIONS AND STRESS RELIEVED, IF REQUIRED (C AND H). THE COMPONENTS ARE FITTED UP FOR WELDING (D AND J). THE WELD IS COMPLETED, USING THE FILLER METAL NORMALLY USED FOR WELDING THE STAINLESS STEEL MEMBER OR THAT USED FOR THE SURFACING PASS ON STAINLESS-CLAD LOW-ALLOY OR CARBON STEEL (E AND K).

In depositing the surface layer, dilution of the stainless steel weld metal by carbon steel, which can cause cracking of the weld, can be minimized with the following practices. If the cladding on stainless-clad steel is type 304 stainless steel, types 309 and 309L should be used for the surfacing layer. After depositing the surfacing layer, the component is stress relieved, if necessary. The final weld between the solid stainless steel and the surfaced stainless-clad steel or carbon steel can be made either with the filler metal normally employed for welding the solid stainless steel member or with the one that was used for the overlay.

In another method, a short stainless steel transition member is welded to the carbon steel or stainless-clad carbon steel member prior to stress relieving and final welding. This method ensures protection of the carbon steel base metal from the effect of final welding, and stress relieving can be done while there is little restraint on the joint. This method, however, is more costly than the procedure for producing a surfacing layer.

In a third method, the workpiece interface (stainless steel and carbon steel or stainless steel and stainless-clad carbon steel) is beveled and fitted, leaving a root opening at the joint that is then welded with a filler metal of sufficiently high alloy content to prevent cracking of the weld after dilution by the carbon steel. The welding procedure should be controlled to hold penetration to a minimum. A disadvantage of this method is that the joint is under restraint both during welding and local stress relieving. Although often used, this method is the least desirable of the three.

Dissimilar Welds With Stainless Steels

Joel G. Feldstein, Foster Wheeler Energy Corporation

Welding of Ferritic and Martensitic Stainless Steels to Carbon and Low-Alloy Steels

When welding ferritic or martensitic stainless steel to carbon or low-alloy steel for general service (not high-temperature service), the use of austenitic stainless steel or nickel-base (NiCrFe) filler metal can produce welds of suitable quality if correct welding procedures are followed. (If an austenitic stainless steel or NiCrFe filler metal cannot be used, a filler metal similar to the stainless steel base metal is usually specified.) Two methods are commonly employed. The first involves surfacing both members of the joint, using suitable preheat and postheat treatments as required, and then making a weld between the buttered surfaces without preheat or postheat. Filler metals (such as type 309) that are sufficiently high in alloy content to minimize the effects of dilution by carbon steel, or straight-chromium grades of stainless steel (400 series) are widely used with this method. Penetration into the base metal should be kept to a minimum. The second method involves depositing the filler metal directly on the surfaces of the two members of the joint without using a separate surfacing layer. Dilution of the weld metal by both base metals must be kept under close control while the restrained joint is being welded.

Dissimilar Welds With Stainless Steels

Joel G. Feldstein, Foster Wheeler Energy Corporation

Welding Austenitic-Stainless-Clad Carbon or Low-Alloy Steels

To preserve its desirable properties, stainless-clad plate can be welded by either of the two following methods, depending on plate thickness and service conditions.

- THE UNCLAD SIDES OF THE PLATE SECTIONS ARE BEVELED AND WELDED WITH CARBON OR LOW-ALLOY STEEL FILLER METAL. A PORTION OF THE STAINLESS STEEL CLADDING IS REMOVED FROM THE BACK OF THE JOINT AND STAINLESS STEEL FILLER METAL IS DEPOSITED.
- THE ENTIRE THICKNESS OF THE STAINLESS-CLAD PLATE IS WELDED WITH STAINLESS STEEL FILLER METAL.

When the non-stainless portion of the plate is comparatively thick, as in most pressure vessel applications, it is more economical to use the first method. When the non-stainless portion of the plate is thin, the second method is often preferred. When welding components for applications involving elevated or cyclic temperatures, the differences in the coefficients of thermal expansion of the base plate and the weld should be taken into consideration.

All stainless steel deposits on carbon steel should be made with filler metal of sufficiently high alloy content to ensure that normal amounts of dilution by carbon steel will not result in a brittle weld. In general, filler metals of type 308, 316, or 347 should not be deposited directly on carbon or low-alloy steel. Deposits of types 309, 309L, 309Cb, 309Mo, 310, or 312 are usually acceptable, although type 310 is fully austenitic and is susceptible to hot cracking when there is high restraint in a welded joint. Thus, welds made with type 310 filler metal should be carefully inspected. Welds made with types 309 and 312 filler metals are partially ferritic and therefore are highly resistant to hot cracking.

The procedure most commonly used for making welded joints in stainless-clad carbon or low-alloy steel plate is shown in Fig. 3. Stainless steel filler metal is deposited only in that portion of the weld where the stainless steel cladding has been removed, and carbon or low-alloy steel filler metal is used for the remainder. The backgouged portion of the stainless steel cladding should be filled with a minimum of two layers of stainless steel filler metal (Fig. 3e); an additional layer is recommended if a high weld reinforcement at the cladding surface can be tolerated.

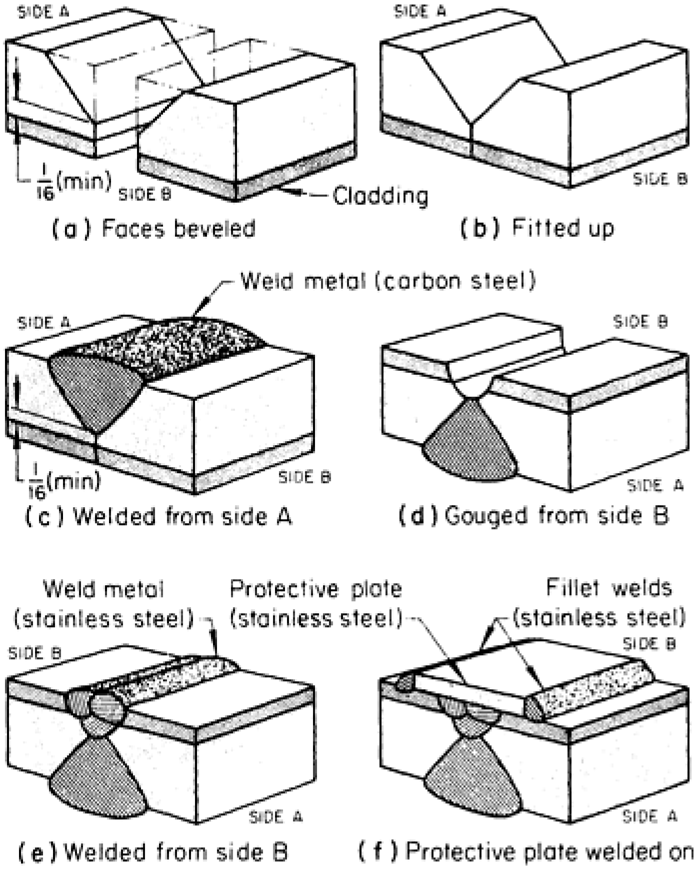

(a) Faces beveled

(b) Fitted up

(c) Welded from side A

(d) Gouged from side B

(e) Welded from side B

(f) Protective plate welded on

FIG. 3 PROCEDURE FOR WELDING STAINLESS-CLAD CARBON AND LOW-ALLOY STEEL, USING STAINLESS STEEL FILLER METAL ONLY IN PORTION OF JOINT FROM WHICH CLADDING WAS REMOVED. (A) AND (B) THE CLAD PLATES ARE MACHINED FOR A TIGHT FIT-UP, WITH THE BOTTOM OF THE WELD GROOVE NOT LESS THAN 1.6 MM ($\frac{1}{16}$ IN.) ABOVE THE STAINLESS STEEL CLADDING. (C) CARBON STEEL FILLER METAL IS DEPOSITED FROM SIDE A (A LOW-HYDROGEN FILLER METAL IS USED FOR THE FIRST PASS), TAKING CARE NOT TO PENETRATE CLOSER THAN 1.6 MM ($\frac{1}{16}$ IN.) TO THE CLADDING. (D) STAINLESS STEEL CLADDING ON SIDE B IS BACKGOUGED UNTIL SOUND CARBON STEEL WELD METAL IS REACHED. (E) THE BACKGOUGED GROOVE IS FILLED WITH STAINLESS STEEL WELD METAL IN A MINIMUM OF TWO LAYERS. (F) WHEN REQUIRED FOR SEVERELY CORROSIVE SERVICE, A PROTECTIVE STRIP OF STAINLESS STEEL PLATE MAY BE

FILLET WELDED TO THE CLADDING TO COVER THE WELD ZONE.

If the cladding is of type 304 stainless steel, the first layer of stainless steel weld metal should be of type 309 or 312. Subsequent layers of weld metal can be of type 308. If the cladding is of type 316, the first layer is deposited with type 309 Mo filler metal and the subsequent layers with type 316. When the cladding is of type 304L or type 347, the welding procedure must be carefully controlled to obtain the desired weld metal composition in the outer layers of the weld. Chemical analysis of sample welds should be made before joining of clad plates intended for use under severely corrosive conditions.

In some applications, a narrow protective plate of wrought stainless steel of the same composition as the cladding is welded over the completed weld (Fig. 3f) to ensure uniformity of corrosion resistance. The fillet welds joining the protective plate to the cladding should be carefully inspected after deposition. These welds, of course, are made with stainless steel filler metal.

Figure 4 illustrates an alternative method (method A) of welding clad plate in which a carbon or low-alloy steel weld joins the carbon steel portion of the plate and the use of stainless steel filler metal is limited to replacement of the cladding that was removed prior to making the carbon or low-alloy steel weld. This method is more expensive than the method described in Fig. 3 because of the cost of removing a larger portion of the cladding and depositing more stainless steel filler metal. Because there is no danger of alloy contamination from the cladding layer, method A in Fig. 4 permits the use of faster welding processes, such as submerged arc welding in depositing the carbon steel weld.

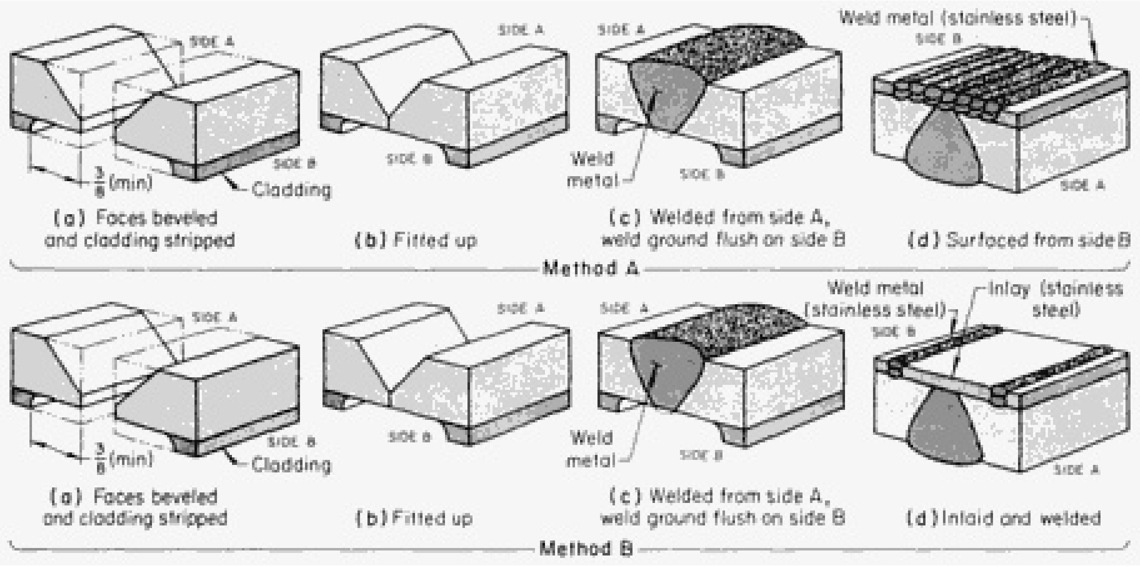

FIG. 4 ALTERNATIVE PROCEDURES FOR JOINING STAINLESS-CLAD CARBON AND LOW-ALLOY STEEL PLATE INVOLVING DIFFERENT TECHNIQUES FOR REPLACING PORTIONS OF THE STAINLESS STEEL CLADDING REMOVED BEFORE WELDING THE CARBON OR LOW-ALLOY STEEL SIDE. THE JOINT IS PREPARED BY BEVELING SIDE A AND REMOVING A PORTION OF THE STAINLESS STEEL CLADDING FROM SIDE B TO A MINIMUM WIDTH OF 9.5 MM ($\frac{3}{8}$ IN.) FROM EACH SIDE OF THE JOINT, AND THE JOINT IS FITTED UP IN POSITION FOR WELDING. USE OF A ROOT GAP (NOT SHOWN) IS PERMISSIBLE (A AND B, METHODS A AND B). CARBON STEEL FILLER METAL IS DEPOSITED AND THE ROOT OF THE WELD IS GROUND FLUSH WITH THE UNDERSIDE OF THE CARBON STEEL PLATE (C, METHODS A AND B). THE AREA FROM WHICH CLADDING WAS REMOVED IS SURFACED WITH AT LEAST TWO LAYERS OF STAINLESS STEEL WELD METAL (D, METHOD A), OR AN INLAY OF WROUGHT STAINLESS STEEL CAN BE WELDED IN PLACE (D, METHOD B).

In depositing the stainless steel weld metal, the first layer must be sufficiently high in alloy content to avoid cracking as a result of normal dilution by the carbon steel base metal. A stringer bead technique should be employed; penetration must be held to a minimum. If the proper weld metal composition is not achieved after the second layer has been deposited, a portion of the second layer should be ground off and additional filler metal should be deposited to obtain the desired composition. Figure 4(d) of method B shows an alternative procedure in which the exposed carbon steel weld on side B is covered by welding an inlay of wrought stainless steel to the edges of the cladding.

The most common method of joining stainless steel-clad carbon or low-alloy steel plate with a weld that consists entirely of stainless steel is shown in Fig. 5. This method is most frequently used for joining thin sections of stainless-clad plate. The same basic welding procedure is followed for both the butt and comer joints shown in Fig. 5. After the plate has been beveled and fitted up for welding, a stainless steel weld is deposited from the carbon steel side, using a filler metal sufficiently high in alloy content to minimize difficulties (such as cracking) resulting from weld dilution and joint restraint. Types 309 and 312 filler metals are suitable for this application.

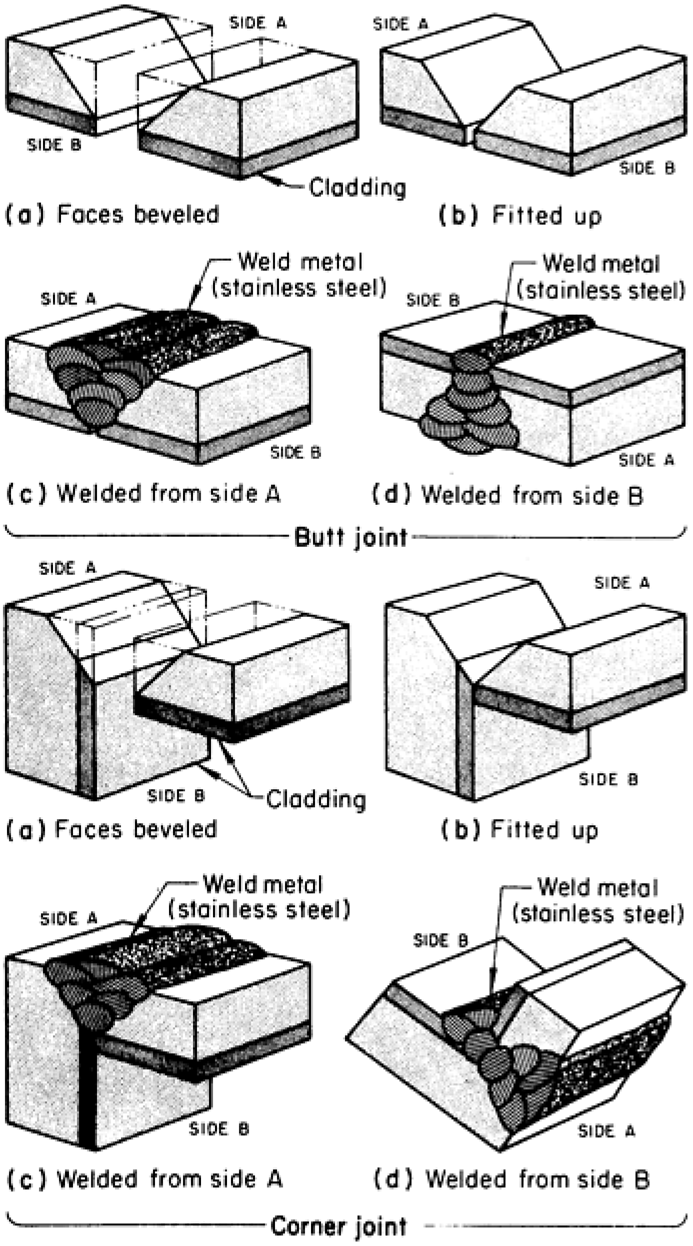
SIDE A
SIDE B
Cladding
(a) Faces beveled
SIDE A
SIDE B
(b) Fitted up
Weld metal (stainless steel)
SIDE A
SIDE B
(c) Welded from side A
Weld metal (stainless steel)
SIDE B
SIDE A
(d) Welded from side B
Butt joint
SIDE A
SIDE B
Cladding
(a) Faces beveled
SIDE A
SIDE B
(b) Fitted up
Weld metal (stainless steel)
SIDE A
SIDE B
(c) Welded from side A
Weld metal (stainless steel)
SIDE B
SIDE A
(d) Welded from side B
Corner joint

FIG. 5 PROCEDURES FOR WELDING V-GROOVE BUTT AND CORNER JOINTS IN STAINLESS-CLAD CARBON OR LOW-ALLOY STEEL PLATE, USING STAINLESS STEEL FILLER METAL EXCLUSIVELY. THE CLAD PLATES ARE BEVELED AND FITTED UP (A AND B, BUTT AND CORNER JOINTS). THE ROOT OF THE WELD IS CLEANED AND GOUGED, IF NECESSARY, BEFORE DEPOSITING STAINLESS WELD METAL FROM THE STAINLESS STEEL SIDE (D, BUTT AND CORNER JOINTS).

After the stainless steel weld has been deposited from the carbon steel side (Fig. 5c), the root of the weld is cleaned by brushing, chipping, or grinding, as required, and one or more layers of stainless steel filler metal are deposited (Fig. 5d).

The filler metal composition should correspond to that normally employed to weld the type of stainless steel used for the cladding. If the cladding is type 304, the final layer of weld metal should be type 308. If the cladding is type 316, it may be necessary to backgouge before deposition of the final weld metal layers to ensure that the proper weld metal composition is obtained at the surface of the weld.

Dissimilar Welds With Stainless Steels

Joel G. Feldstein, Foster Wheeler Energy Corporation

Cladding Austenitic Stainless Steel to Carbon or Low-Alloy Steels

Weld cladding involves the deposition of a relatively thick layer of weld metal by an arc welding process to impart a corrosion- or oxidation-resistant surface (see the article "Hardfacing, Weld Cladding, and Dissimilar Metal Joining" in this Volume). Materials of interest for cladding carbon or low-alloy steels include types 308, 309, 310, 316, 320, and 347. The exact chemical composition of a weld overlay depends on the composition of the cladding metal (determined from undiluted weld metals), the composition of the base metal, and dilution, and can be calculated if these variables are known. The chromium-nickel austenitic stainless steels, however, often contain more than one phase in the as-deposited structure. These materials are primarily austenitic with varying amounts of ferrite, martensite, and σ phases determined by the composition and cladding conditions.

The amounts of ferrite and martensite present in the austenitic structures can be estimated from a DeLong constitutional diagram similar to the one shown in Fig. 6. Chromium (ferrite formers) and nickel (austenite formers) equivalents are calculated from the following equations:

$$\text{CHROMIUM EQUIVALENT} = \%\text{CR} + \%\text{MO} + 1.5(\%\text{SI}) + 0.5(\%\text{NB})$$

$$\text{NICKEL EQUIVALENT} = \%\text{NI} + 30(\%\text{C}) + 0.5(\%\text{MN}) + 30(\%\text{N})$$

These values are then plotted and the percentages of phases present are read directly from the diagram. Care should be taken in this technique, however, because the as-clad microstructure also depends on cooling rate. Cooling rates can vary significantly, depending on cladding conditions. Another method of quickly determining the phases present is by using magnetic measuring devices. Austenite is nonmagnetic, whereas ferrite and martensite are ferromagnetic. When properly calibrated in accordance with AWS A4.2, these instruments can quickly give an indication of the amount of magnetic phase present. However, one should be aware that excessive dilution can result in martensite in the cladding, which will give a false high FN value when using a magnetic instrument.

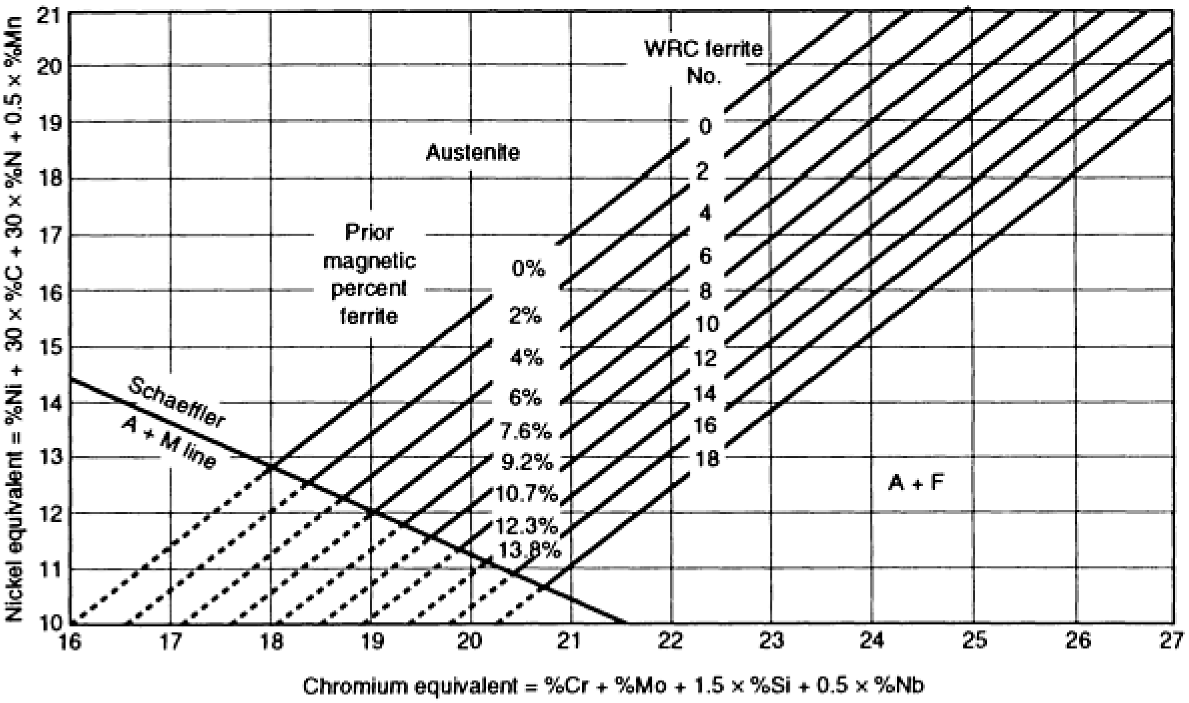

FIG. 6 THE DELONG DIAGRAM USED FOR PREDICTING WELD MICROSTRUCTURE. SEE TEXT FOR DETAILS.

Sigma phase is not common in as-clad structures but can occur where high-heat-input processes are used. It forms most readily from ferrite phase, but can also form from austenite in high-chromium areas. The transformation to σ phase occurs in the temperature range 500 to 900 °C (930 to 1650 °F), with longer times at these temperatures producing larger amounts of σ phase. The percentage of σ phase can be determined only by metallographic techniques.

Fully austenitic overlays are generally susceptible to hot cracking, which can occur upon solidification or reheating by a subsequent overlay pass (microfissures). Solidification cracks are generally very large and can be found by visual or dye-penetrant examination. However, microfissures are often subsurface and difficult to detect by conventional nondestructive examination techniques. For the austenitic stainless steels, the tendency for hot cracking is greatly reduced by the presence of a small amount of δ-ferrite in the structure. The amount of ferrite required to eliminate hot cracking varies from material to material, but it is generally recognized that a 4 or 5 FN minimum in the as-clad structure is sufficient for most austenitic stainless steel cladding alloys.

Finally, stainless steel cladding materials are susceptible to sensitization (precipitation of chromium carbides in grain boundaries that leaves chromium-depleted adjacent areas). Sensitization can occur when unstabilized stainless steels are heated in the range of 430 to 820 °C (800 to 1500 °F) either during stress relief or in service. The best way to inhibit these problems is to use low-carbon grades (such as type 308L or 309L) or stabilized grades (such as type 347) of austenitic stainless steels that are not readily sensitized. Also, finer microstructures produced by the lower-heat-input processes improve corrosion resistance for all grades. Additional material on weld sensitization can be found in the article "Corrosion of Weldments" in this Volume as well as in *Corrosion,* Volume 13 of the *ASM Handbook.*

Dissimilar Welds With Stainless Steels

Joel G. Feldstein, Foster Wheeler Energy Corporation

Selected References

- W.T. DELONG, FERRITE IN AUSTENITIC STAINLESS STEEL WELD METAL, *WELD. J.,* VOL 53 (NO. 7), 1974, P 273S-286S
- D.L. OLSON, PREDICTION OF AUSTENITIC WELD METAL MICROSTRUCTURES AND

PROPERTIES, *WELD. J.* VOL 64 (NO. 10), 1985, P 281S-295S
- A.L. SCHAEFFLER, CONSTITUTION DIAGRAM FOR STAINLESS STEEL WELD METAL, *MET. PROG.*, VOL 56 (NO. 11), 1949, P 680-688

Selection and Weldability of Conventional Titanium Alloys

W.A. Baeslack III, The Ohio State University; J.R. Davis, Davis and Associates; C.E. Cross, Martin Marietta Astronautics Group

Introduction

TITANIUM is a low-density element (approximately 60% of the density of steel) that can be highly modified by alloying and deformation processing. Titanium is nonmagnetic and has good heat transfer properties. Its coefficient of thermal expansion is somewhat lower than that of steels and less than half that of aluminum. Titanium and its alloys have melting points higher than those of steels, but maximum useful temperatures for structural applications generally range from 425 to 540 °C (800 to 1000 °F). More recently developed alloys such as IMI-834 or Ti-1100 have temperature limits approaching 600 °C (1100 °F). Titanium has the ability to passivate, and thereby exhibits a high degree of immunity to attack by most mineral acids and chlorides. Titanium is nontoxic and generally biologically compatible with human tissues and bones. The combination of high strength, stiffness, good toughness, low density, and good corrosion resistance provided by various titanium alloys at very low to moderately elevated temperatures allows weight savings in aerospace structures and other high-performance applications. The excellent corrosion resistance and biocompatibility coupled with good strengths make titanium and its alloys useful in chemical and petrochemical applications, marine environments, and biomaterials applications.

Titanium and its alloys are used primarily in two areas of application where the unique characteristics of these metals justify their selection: corrosion-resistant service and specific strength-efficient structures. For these two diverse areas, selection criteria differ markedly. Corrosion applications normally utilize low-strength unalloyed titanium mill products fabricated into tanks, heat exchangers, or reactor vessels for chemical-processing, desalination, or power-generation plants. In contrast, high-performance applications typically utilize high-strength titanium alloys in a very selective manner that depends on factors such as thermal environment, loading parameters, available product forms, fabrication characteristics, and inspection and/or reliability requirements. As a result of their specialized usage, alloys for high-performance applications normally are processed to more stringent and costly requirements than unalloyed titanium for corrosion service.

Historically, titanium alloys have been used instead of iron or nickel alloys in aerospace applications because titanium saves weight in highly loaded components that operate at low to moderately elevated temperatures. Many titanium alloys have been custom designed to have optimum tensile, compressive, and/or creep strength at selected temperatures and, at the same time, to have sufficient workability to be fabricated into mill products suitable for specific applications. During the past four decades, various compositions have had transient usage, but one alloy, Ti-6Al-4V, has been consistently responsible for about 50% of industry application. Ti-6Al-4V is unique in that it combines attractive properties with inherent workability (which allows it to be produced in all types of mill products, in both large and small sizes), good weldability (which allows the mill products to be made into complex hardware), and the production experience and commercial availability that lead to reliable and economic usage. Thus Ti-6Al-4V has become the standard alloy against which other alloys must be compared when selecting a titanium alloy for a specific application.

Strength, fatigue-crack growth rate, and fracture toughness, plus manufacturing considerations such as welding and forming requirements, normally provide the criteria that determine the alloy composition, structure (alpha, near-alpha, alpha-beta, or metastable beta), heat treatment (some variant of either annealing or solution treating and aging), and level of process control selected or prescribed for structural titanium alloy applications. Detailed information on the selection and properties of wrought, cast, and powder metallurgy titanium alloys can be found in ***Properties and Selection: Nonferrous Alloys and Special-Purpose Materials***, Volume 2 of the *ASM Handbook.*

In the present article, emphasis will be placed on the physical metallurgy of titanium and titanium alloys and their microstructural response to fusion welding conditions. A companion article, which immediately follows in this Section of the Handbook, describes welding characteristics of advanced high-temperature and high fracture toughness alloys, titanium-matrix composites, and ordered intermetallic (titanium aluminides) alloys. Additional information on filler metal and shielding gas selection can be found in the article "Welding of Titanium Alloys" in this Volume.

Selection and Weldability of Conventional Titanium Alloys

W.A. Baeslack III, The Ohio State University; J.R. Davis, Davis and Associates; C.E. Cross, Martin Marietta Astronautics Group

Physical Metallurgy (Ref 1)

Titanium undergoes an allotropic phase transformation at about 885 °C (1625 °F), changing from a close-packed hexagonal crystal structure (alpha phase) to a body-centered cubic crystal structure (beta phase). The transformation temperature (beta transus, or the completion of transformation to beta on heating) is strongly influenced by the interstitial elements oxygen, nitrogen, and carbon (alpha stabilizers), which raise the transformation temperature; by hydrogen (beta stabilizer), which lowers the transformation temperature; and by metallic impurity or alloying elements, which may either raise or lower the transformation temperature.

Depending on their microstructure, titanium alloys fall into one of four classes: alpha, near-alpha, alpha-beta, or metastable beta. These classes, which are described below, denote the general type of microstructure after processing. An alpha alloy will contain only trace amounts of beta phase. A near-alpha alloy will contain limited beta phase and may appear microstructurally similar to an alpha alloy. An alpha-beta alloy will consist of alpha and retained or transformed beta. A metastable beta alloy will tend to retain the beta phase on initial cooling to room temperature but will precipitate fine alpha phase during heat treatment.

Effects of Alloying Elements. The role of the interstitial elements oxygen, nitrogen, and carbon was described above. The principal effect of an alloying element in titanium alloys is its effect on the alpha-to-beta transformation temperature and the resulting microstructure and properties (Fig. 1).

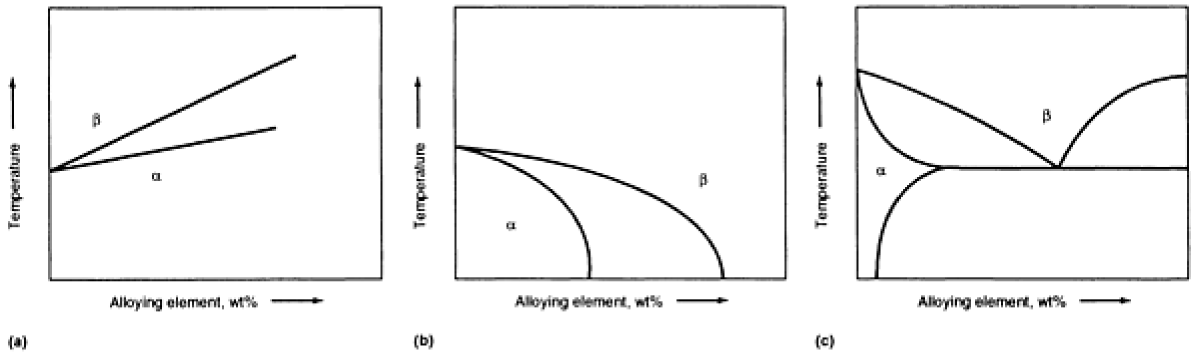

FIG. 1 EFFECT OF ALLOYING ELEMENTS ON TITANIUM MICROSTRUCTURE. (A) ALPHA STABILIZER (FOR EXAMPLE, ALUMINUM AND OXYGEN). (B) BETA STABILIZER (FOR EXAMPLE, VANADIUM AND NIOBIUM). (C) BETA STABILIZER PLUS (α + β) STRENGTHENER (FOR EXAMPLE, COPPER AND IRON)

Tantalum, vanadium, niobium, and molybdenum are beta isomorphous with body-centered cubic titanium. Titanium does not form intermetallic compounds with the beta isomorphous elements. Eutectoid systems are formed with chromium, iron, copper, nickel, palladium, cobalt, manganese, and certain other transition metals. These elements have low solubility in alpha titanium and decrease the transformation temperature. They usually are added to alloys in combination with one or more of the beta isomorphous elements to stabilize the beta phase and prevent or minimize formation of intermetallic compounds, which might occur during service at elevated temperature.

Zirconium and hafnium are unique in that they are isomorphous with both the alpha and beta phases of titanium. Tin and aluminum have significant solubility in both alpha and beta phases. Aluminum increases the transformation temperature significantly, whereas tin lowers it slightly. Aluminum, tin, and zirconium are commonly used together in alpha and near-alpha alloys. In alpha-beta alloys, these elements are distributed approximately equally between the alpha and beta phases. Almost all commercial titanium alloys contain one or more of these three elements because they are soluble in both alpha and beta phases, and particularly because they improve creep strength of the alpha phase. Many more elements are soluble in beta titanium than in alpha. Beta isomorphous alloying elements are preferred as additions because they do not form intermetallic compounds. However, iron, chromium, manganese, and other compound formers sometimes are used in beta-rich alpha-beta alloys or in metastable beta alloys, because they are strong beta stabilizers and improve hardenability and response to heat treatment. Nickel, molybdenum, and palladium improve corrosion resistance of unalloyed titanium in certain media. The approximate chemical composition ranges and effects of some alloying elements used in titanium can be summarized as follows (Ref 2):

ALLOYING ELEMENT	CONTENT, WT%	EFFECT ON STRUCTURE
ALUMINUM	2-7	α STABILIZER
TIN	2-6	α STABILIZER
VANADIUM	2-15	β STABILIZER
MOLYBDENUM	2- 13	β STABILIZER
CHROMIUM	2-12	β STABILIZER
COPPER	-3	β STABILIZER (α AND β STRENGTHENER)
ZIRCONIUM	2-8	IMPROVES CREEP STRENGTH
SILICON	0.05-1	IMPROVES CREEP STRENGTH

Secondary Phases and Martensitic Transformations. The basis for microstructural manipulation during elevated-temperature processing of titanium alloys centers around the beta-to-alpha transformation that occurs in these alloys during cooling. This transformation can occur by nucleation and growth, or it can occur martensitically, depending on the alloy composition and the cooling rate. The martensitic product is usually hexagonal close-packed and is designated alpha prime (α'). There also is an orthorhombic martensite, designated alpha double-prime (α"), which forms in alloys that contain higher concentrations of refractory elements such as molybdenum, tantalum, or niobium. Essentially all thermomechanical processing is conducted above the M_s temperature for either alpha prime or alpha double-prime. Alloys that contain enough beta-stabilizing elements to depress the M_s temperature below room temperature can be rapidly cooled to retain the metastable beta phase. More detailed information on the physical metallurgy and phase transformations associated with titanium alloys can be found in Ref 3.

A graphic illustration of the microstructural transformations that can occur in an alpha-beta alloy (Ti-6Al-4V) is shown in Fig. 2. As can be seen, the morphology (shape/location) of the phase changes with prior treatments. The alpha phase may remain relatively globular (equiaxed), but the transformed beta (martensites or alpha) may be very acicular or elongated. The amount of equiaxed alpha and the coarseness or fineness of the transformed beta products will affect alloy properties. For example, the fine, predominantly alpha-prime structures in rapidly cooled near-alpha and alpha-beta alloy weldments exhibit high strength levels that can become even higher after low-temperature postweld heat treatment. However, as the microstructure coarsens due to a reduced weld cooling rate or increased postweld heat treatment temperature, yield and ultimate tensile strengths decrease. See the section "Effect of Postweld Heat Treatment" in this article for additional information on the effects of PWHT on structure-property relationships.

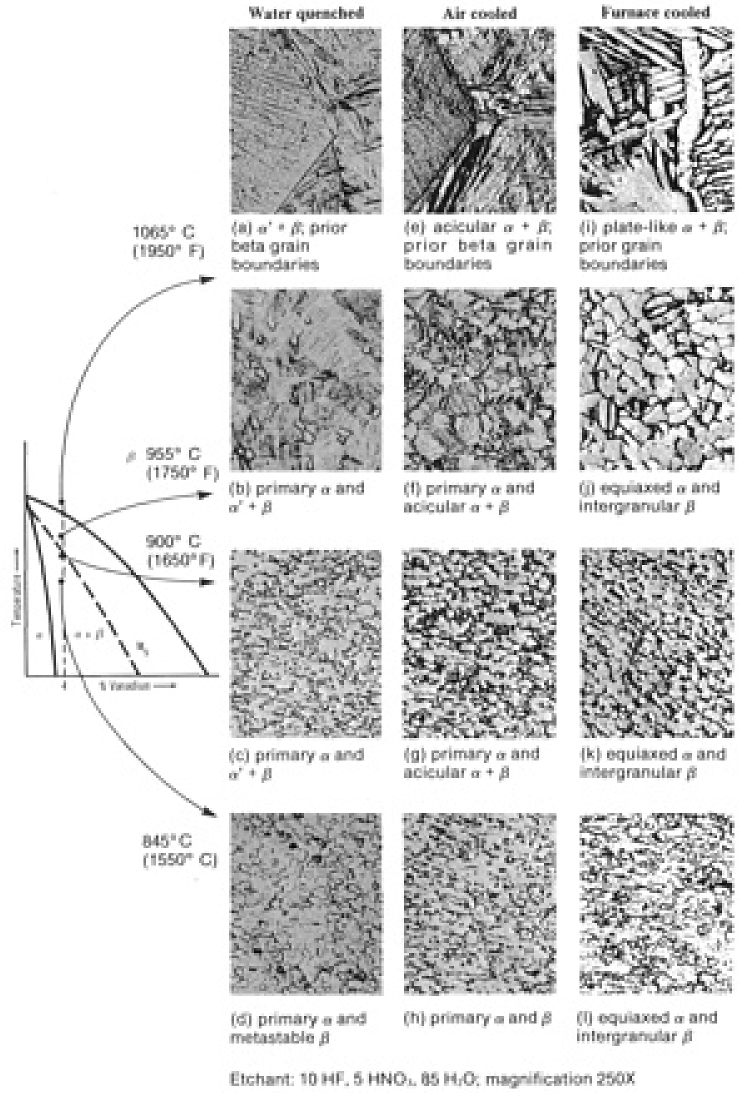
Water quenched
Air cooled
Furnace cooled
1065° C
(1950° F)
(a) α' + β; prior beta grain boundaries
(e) acicular α + β; prior beta grain boundaries
(i) plate-like α + β; prior grain boundaries
β 955° C
(1750° F)
(b) primary α and α' + β
(f) primary α and acicular α + β
(j) equiaxed α and intergranular β
900° C
(1650° F)
(c) primary α and α' + β
(g) primary α and acicular α + β
(k) equiaxed α and intergranular β
845° C
(1550° C)
(d) primary α and metastable β
(h) primary α and β
(l) equiaxed α and intergranular β
Temperature
α
α + β
M_s
4
% Vanadium
Etchant: 10 HF, 5 HNO₃, 85 H₂O; magnification 250X

FIG. 2 MICROSTRUCTURAL TRANSFORMATIONS THAT OCCUR IN TI-6AL-4V ON COOLING

References cited in this section

1. M.J. DONACHIE, JR., TITANIUM, *METALS HANDBOOK DESK EDITION,* AMERICAN SOCIETY FOR METALS, 1985, P 9-1 TO 9-3
2. S. LAMPMAN, WROUGHT TITANIUM AND TITANIUM ALLOYS, *PROPERTIES AND SELECTION: NONFERROUS ALLOYS AND SPECIAL-PURPOSE MATERIALS,* VOL 2, *ASM HANDBOOK,* ASM INTERNATIONAL, 1990, P 592-633
3. E.W. COLLINGS, *THE PHYSICAL METALLURGY OF TITANIUM ALLOYS,* AMERICAN SOCIETY FOR METALS, 1984

Selection and Weldability of Conventional Titanium Alloys

W.A. Baeslack III, The Ohio State University; J.R. Davis, Davis and Associates; C.E. Cross, Martin Marietta Astronautics Group

Classification of Titanium Alloys

When classifying unalloyed and alloyed titanium, it is common to group the materials as:

- UNALLOYED, OR COMMERCIALLY PURE (CP) TITANIUM
- ALPHA AND NEAR-ALPHA ALLOYS
- ALPHA-BETA ALLOYS
- METASTABLE BETA ALLOYS

Table 1 lists compositions and tensile properties of some of the more common grades. Of the alloys listed in Table 1, Ti-6Al-4V is the most widely used titanium alloy, accounting for about 50% of total titanium production. Unalloyed grades comprise about 25% of production, and all other alloys combined comprise the remaining 25%.

TABLE 1 PROPERTIES AND COMPOSITIONS OF TITANIUM AND TITANIUM ALLOYS

DESIGNATION	**TENSILE STRENGTH (MIN)**		**0.2% YIELD STRENGTH (MIN)**		**IMPURITY LIMITS (MAX), WT%**					**NOMINAL COMPOSITION, WT%**				
	MPA	**KSI**	**MPA**	**KSI**	**N**	**C**	**H**	**Fe**	**O**	**Al**	**Sn**	**Zr**	**Mo**	**OTHERS**
UNALLOYED GRADES														
ASTM GRADE 1	240	35	170	25	0.03	0.10	0.015	0.20	0.18	. . .	. . .	. . .	. . .	. . .
ASTM GRADE 2	340	50	280	40	0.03	0.10	0.015	0.30	0.25	. . .	. . .	. . .	. . .	. . .
ASTM GRADE 3	450	65	380	55	0.05	0.10	0.015	0.30	0.35	. . .	. . .	. . .	. . .	. . .
ASTM GRADE 4	550	80	480	70	0.05	0.10	0.015	0.50	0.40	. . .	. . .	. . .	. . .	. . .
ASTM GRADE 7	340	50	280	40	0.03	0.10	0.015	0.30	0.25	. . .	. . .	. . .	. . .	0.2 PD
ASTM GRADE 11	240	35	170	25	0.03	0.10	0.015	0.20	0.18	. . .	. . .	. . .	. . .	0.12-0.25 PD
ASTM GRADE 12	480	70	380	55	0.03	0.10	0.015	0.30	0.25	. . .	. . .	. . .	0.3	0.6-0.9 NI
ALPHA AND NEAR-ALPHA ALLOYS														
TI-0.3MO-0.8NI	480	70	380	55	0.03	0.10	0.015	0.30	0.25	. . .	. . .	. . .	0.3	0.8 NI
TI-5AL-2.5SN	790	115	760	110	0.05	0.08	0.02	0.50	0.20	5	2.5	. . .	. . .	. . .
TI-5AL-2.5SN-ELI	690	1130	620	90	0.07	0.08	0.0125	0.25	0.12	5	2.5	. . .	. . .	. . .
TI-8AL-1MO-1V	900	130	830	120	0.05	0.08	0.015	0.30	0.12	8	. . .	. . .	1	1 V
TI-6AL-2SN-4ZR-2MO	900	130	830	120	0.05	0.05	0.0125	0.25	0.15	6	2	4	2	. . .
TI-6AL-2NB-1TA-0.8MO	790	115	690	100	0.02	0.03	0.0125	0.12	0.10	6	. . .	. . .	1	2 NB, 1 TA
TI-2.25AL-11SN-5ZR-1MO	1000	145	900	130	0.04	0.04	0.008	0.12	0.17	2.25	11.0	5.0	1.0	0.2 SI
TI-5AL-5SN-2ZR-2MO(A)	900	130	830	120	0.03	0.05	0.0125	0.15	0.13	5	5	2	2	0.25 SI
ALPHA-BETA ALLOYS														
TI-6AL-4V(B)	900	130	83	120	0.05	0.10	0.0125	0.30	0.20	6.0	. . .	. . .	. . .	4.0 V
TI-6AL-4V-ELI(B)	830	120	760	110	0.05	0.08	0.0125	0.25	0.13	6.0	. . .	. . .	. . .	4.0 V
TI-6AL-6V-2SN(B)	1030	150	970	140	0.04	0.05	0.015	1.0	0.20	6.0	2.0	. . .	. . .	0.75 CU, 6.0 V
TI-8MN(B)	860	125	760	110	0.05	0.08	0.015	0.50	0.20	. . .	. . .	. . .	. . .	8.0 MN
TI-7AL-4MO(B)	1030	150	970	140	0.05	0.10	0.013	0.30	0.20	7.0	. . .	. . .	4.0	. . .
TI-6AL-2SN-4ZR-6MO(C)	1170	170	1100	160	0.04	0.04	0.0125	0.15	0.15	6.0	2.0	4.0	6.0	. . .
TI-5AL-2SN-2ZR-4MO-4CR(A)(C)	1125	163	1055	153	0.04	0.05	0.0125	0.30	0.13	5.0	2.0	2.0	4.0	4.0 CR
TI-6AL-2SN-2ZR-2MO-2CR(A)(B)	1030	150	970	140	0.03	0.05	0.0125	0.25	0.14	5.7	2.0	2.0	2.0	2.0 CR, 0.25 SI
TI-3AL-2.5V(D)	620	90	520	75	0.015	0.05	0.015	0.30	0.12	3.0	. . .	. . .	. . .	2.5 V
TI-1100	1000	145	900	130	. . .	. . .	. . .	0.02	0.07	6.0	2.75	4.0	0.4	0.45 SI

TI-4.5AL-5MO-1.5CR (CORONA 5)	900	130	830	120	. . .	. . .	. . .	. . .	. . .	4.5	. . .	. . .	5.0	1.5 CR
METASTABLE BETA ALLOYS														
TI-10V-2FE-3AL(A)(C)	1170	170	1100	160	0.05	0.05	0.015	2.5	0.16	3.0	. . .	. . .	. . .	10.0 V
TI-13V-11CR-3AL(C)	1170	170	1100	160	0.05	0.05	0.025	0.35	0.17	3.0	. . .	. . .	. . .	11.0 CR, 13.0 V
TI-8MO-8V-2FE-3AL(A)(C)	1170	170	1100	160	0.05	0.05	0.015	2.5	0.17	3.0	. . .	. . .	8.0	8.0 V
TI-3AL-8V-6CR-4MO-4ZR(A)(B)	900	130	830	120	0.03	0.05	0.020	0.25	0.12	3.0	. . .	4.0	4.0	6.0 CR, 8.0 V
TI-11.5MO-6ZR-4.5SN (BETA III)(B)	690	100	620	90	0.05	0.10	0.020	0.35	0.18	. . .	4.5	6.0	11.5	. . .
TI-15V-3AL-3SN-3CR	790	115	775	115	0.03	0.03	0.015	0.30	0.13	3.0	3.0	. . .	. . .	15.0 V, 3.0 CR
TI-15MO-2.7NB-3AL-0.2SI (BETA 21S)(C)	1100	160	1030	150	. . .	. . .	. . .	. . .	. . .	3.0	. . .	. . .	15.0	2.7 NB, 0.2 SI

(A) SEMICOMMERCIAL ALLOY; MECHANICAL PROPERTIES AND COMPOSITION LIMITS SUBJECT TO NEGOTIATION WITH SUPPLIERS.

(B) MECHANICAL PROPERTIES GIVEN FOR ANNEALED CONDITION; MAY BE SOLUTION TREATED AND AGED TO INCREASE STRENGTH.

(C) MECHANICAL PROPERTIES GIVEN FOR SOLUTION TREATED AND AGED CONDITION; ALLOY NOT NORMALLY APPLIED IN ANNEALED CONDITION. PROPERTIES MAY BE SENSITIVE TO SECTION SIZE AND PROCESSING.

(D) PRIMARILY A TUBING ALLOY; MAY BE COLD DRAWN TO INCREASE STRENGTH.

Unalloyed titanium products, which have minimum titanium contents ranging from about 98 to 99.5 wt%, are usually selected for their excellent corrosion resistance, especially in applications where high strength is not required. These materials are also used because they are highly formable and weldable. Yield strengths of unalloyed (commercially pure) grades (see Table 1) vary from 170 MPa (25 ksi) to 480 MPa (70 ksi) simply as a result of variations in the interstitial and impurity levels. Oxygen and iron are the primary variants in these grades; strength increases with increasing oxygen and iron contents. If the iron content is above about 0.05%, preferential corrosive attack of weld metal can occur in nitric acid solutions (Ref 4). Welds are particularly vulnerable because of the acicular nature of any retained beta phase that is stabilized by the iron. Galvanic cells between the beta and the contiguous alpha phase initiate corrosion of the weld metal. This behavior is not true for the base metal, where retained beta is finely divided and discontinuous, and corrosive attack is slight. Filler metal with low iron content should be used, and all sources of iron contamination during preparation and welding should be avoided.

Alpha and Near-Alpha Alloys. Alpha alloys that contain aluminum, tin, and/or zirconium are preferred for high-temperature as well as cryogenic applications. Alpha-rich alloys generally are more resistant to creep at high temperature than alpha-beta or metastable beta alloys. The extra-low-interstitial alpha alloys (ELI grades) retain ductility and toughness at cryogenic temperatures, and Ti-5Al-2.5Sn-ELI has been used extensively in such applications.

Unlike alpha-beta and metastable beta alloys, alpha alloys cannot be strengthened by heat treatment. Generally, alpha alloys are annealed or recrystallized to remove residual stresses induced by cold working. Alpha alloys have good weldability because they are insensitive to heat treatment. They generally have poorer forgeability and narrower forging-temperature ranges than alpha-beta or metastable beta alloys, particularly at temperatures below the beta transus. This poorer forgeability is manifested by a greater tendency for center bursts or surface cracks to occur, which means that small reduction steps and frequent reheats must be incorporated in forging schedules.

Alpha alloys that contain small additions of beta stabilizers (Ti-8Al-1Mo-1V or Ti-6Al-2Nb-1Ta-0.8Mo, for example) sometimes have been classed as "super-alpha" or "near-alpha" alloys. Although they contain some retained beta phase, these alloys consist primarily of alpha and behave more like conventional alpha alloys than alpha-beta alloys.

The alpha and near-alpha alloys have good weldability because of their good ductility (Ref 4). Welding operations have little effect on the mechanical properties of annealed material in the heat-affected zone. However, the strength of cold-worked material in the weld heat-affected zone is decreased as a result of heating. Therefore, these alloys are normally welded in the annealed condition. Stress relieving of weldments is commonly recommended.

Alpha-beta alloys contain one or more alpha stabilizers or alpha-soluble elements plus one or more beta stabilizers. These alloys retain more beta phase after solution treatment than do near-alpha alloys, the specific amount depending on the quantity of beta stabilizers present and on the heat treatment.

Alpha-beta alloys can be strengthened by solution treating and aging. Solution treating usually is done at a temperature high in the two-phase alpha-beta field, and is followed by quenching in water, oil, or other suitable quenchant. As a result of quenching, the beta phase present at the solution treating temperature may be retained or may be partly transformed during cooling by either martensitic transformation or nucleation and growth. The specific response depends on alloy composition, solution treating temperature (beta-phase composition at the solution temperature), cooling rate, and section size. Solution treatment is followed by aging, normally at 480 to 600 °C (900 to 1100 °F) to precipitate alpha and produce a fine mixture of alpha in the retained or transformed beta phase. Transformation kinetics, transformation products, and specific response of a given alloy can be quite complex; a detailed review is provided in Ref 3.

Solution treating and aging can increase the strength of alpha-beta alloys 30 to 50%, or more, over the annealed or over-aged condition. Response to solution treating and aging depends on section size; alloys relatively low in beta stabilizers (Ti-6Al-4V, for example) have poor hardenability and must be quenched rapidly to achieve significant strengthening. For Ti-6Al-4V, the cooling rate of a water quench is not rapid enough to significantly harden sections thicker than about 25 mm (1 in.). As the content of beta stabilizers increases, hardenability increases. For some alloys of intermediate beta-stabilizer content, the surface of a relatively thick section can be strengthened, but the core may be 10 to 20% lower in hardness and strength. The strength that can be achieved by heat treatment is also a function of the volume fraction of beta phase present at the solution treating temperature. Alloy composition, solution temperature, and aging conditions must be carefully selected and balanced to produce the desired mechanical properties in the final product. Table 2 summarizes typical heat treatments for alpha-beta alloys.

TABLE 2 SUMMARY OF HEAT TREATMENTS FOR ALPHA-BETA TITANIUM ALLOYS

HEAT TREATMENT DESIGNATION	HEAT TREATMENT CYCLE(A)	MICROSTRUCTURE
DUPLEX ANNEAL	SOLUTION TREAT AT 50-75 °C (90-135 °F) BELOW T β, AIR COOL AND AGE FOR 2-8 H AT 540-675 °C (1000-1250 °F)	PRIMARY α, PLUS WIDMANSTÄTTEN α - β REGIONS
SOLUTION TREAT AND AGE	SOLUTION TREAT AT ~40 °C (70 °F) BELOW T β, WATER QUENCH AND AGE FOR 2-8 H AT 535-675 °C (995-1250 °F)	PRIMARY α, PLUS TEMPERED α ' OR A β - α MIXTURE
BETA ANNEAL	SOLUTION TREAT AT ~15 °C (30 °F) ABOVE T β, AIR COOL STABILIZE AT 650-760 °C (1200-1400 °F) FOR 2 H	WIDMANSTÄTTEN α - β COLONY MIXTURE
BETA QUENCH	SOLUTION TREAT AT ~15 °C (30 °F) T β, WATER QUENCH AND TEMPER AT 650-760 °C (1200-1400 °F) FOR 2 H	TEMPERED α '
RECRYSTALLIZATION-ANNEAL	925 °C (1700 °F) FOR 4 H, COOL AT 50 °C/H (90 °F/H) TO 760 °C (1400 °F), AIR COOL	EQUIAXED α WITH β AT GRAIN-BOUNDARY TRIPLE JOINTS
MILL ANNEAL	α - β HOT WORK PLUS ANNEAL AT 705 °C (1300 °F) FOR 30 MIN TO SEVERAL HOURS AND AIR COOL	INCOMPLETELY RECRYSTALLIZED α WITH A SMALL VOLUME FRACTION OF SMALL β PARTICLE

(A) T β IS THE β TRANSUS TEMPERATURE FOR THE PARTICULAR ALLOY IN QUESTION. IN MORE HEAVILY β -STABILIZED ALLOYS SUCH AS TI-6AL-2SN-4ZR-6MO OR TI-6AL-6V-2SN, SOLUTION TREATMENT IS FOLLOWED BY AIR COOLING. SUBSEQUENT AGING CAUSES PRECIPITATION OF α PHASE TO FORM AN α - β MIXTURE.

Welding alpha-beta alloys may significantly change their strength, ductility, and toughness characteristics as a result of the thermal cycle to which the alloy is exposed (Ref 4). The low ductility of most alpha-beta alloy welds is caused by phase transformations in the weld metal or the heat-affected zone, or both (see the section "Weld Microstructure" below). Alpha-beta alloys can be welded with unalloyed titanium or alpha-titanium alloy filler metal to produce a weld metal that is low in beta phase. This improves weld ductility. However, this procedure does not overcome the low ductility of the heat-affected zone in alloys that contain large amounts of beta stabilizers.

Ti-6Al-4V has the best weldability of the alpha-beta alloys. This weldability can be attributed to two principal factors. First, the alpha-prime martensite which forms in Ti-6Al-4V is not as hard and brittle as that exhibited by more heavily beta-stabilized alloys, such as Ti-6Al-6V-2Sn. Secondly, Ti-6Al-4V exhibits a relatively low hardenability, which allows the formation of higher proportions of the more desirable Widmanstätten alpha plus retained beta microstructure even at relatively high weld cooling rates.

Due to its single-phase mode of solidification (i.e., absence of low-melting-point eutectics), Ti-6Al-4V is also highly resistant to solidification- and liquation-related cracking. However, the occurrence of solid-state cracking and the formation of porosity can be encountered during welding. Fortunately, these defects are not metallurgically inherent in Ti-6Al-4V, but rather originate from readily correctable deficiencies in preweld cleaning of the workpieces and shielding of the weld zone from atmospheric contamination.

Alpha-beta alloys that are highly beta stabilized, such as Ti-6Al-2Sn-4Zr-6Mo and Ti-6Al-6V-2Sn, have limited weldability. They tend to crack when welded under high restraint or when minor defects are present in the weld zone (Ref

4). The resistance to cracking may be improved by preheating in the range of 150 to 175 °C (300 to 350 °F) and then stress-relieving immediately after welding.

Metastable beta alloys are sufficiently rich in beta stabilizers (and lean in alpha stabilizers) that the beta phase can be completely retained with appropriate cooling rates. These alloys are metestable, and precipitation of alpha phase in the metestable beta is a method used to strengthen the alloys. Metestable beta alloys contain small amounts of alpha-stabilizing elements as strengthening agents.

As a class, metestable beta alloys offer increased fracture toughness over alpha-beta alloys at a given strength level (Ref 2). However, metastable beta and beta alloys require close control of processing and fabrication steps to achieve optimal properties. In the past, metestable beta alloys had rather limited applications, such as springs and fasteners, where very high strength was required.

In recent years, however, metestable beta alloys have received closer attention because their fracture toughness characteristics respond to the increased need for damage tolerance in aerospace structures. In addition, some metastable beta alloys containing molybdenum have good corrosion characteristics. Metastable beta alloys also exhibit:

- BETTER ROOM-TEMPERATURE FORMING AND SHAPING CHARACTERISTICS THAN ALPHA-BETA ALLOYS
- HIGHER STRENGTH THAN ALPHA-BETA ALLOYS AT TEMPERATURES WHERE YIELD STRENGTH (INSTEAD OF CREEP STRENGTH) IS THE APPLICABLE CRITERION
- BETTER RESPONSE TO HEAT TREATMENT (SOLUTION TREATMENT, QUENCHING, AND AGING) IN HEAVIER SECTIONS THAN THE ALPHA-BETA ALLOYS

Most metastable beta alloys are weldable in either the annealed or the heat-treated condition (Ref 4). These alloys include Ti-13V-11Cr-3Al, Ti-8Mo-8V-3A1-2Fe, and Ti-15V-3Al-3Cr-3Sn alloys. Welded joints have good ductility but relatively low strength in the as-welded condition. They are used most often in this condition because the welded joint can respond differently to heat treatment than the base metal and have low ductility after aging. Aging can take place if the welds are exposed to elevated temperatures in service.

References cited in this section

2. S. LAMPMAN, WROUGHT TITANIUM AND TITANIUM ALLOYS, *PROPERTIES AND SELECTION: NONFERROUS ALLOYS AND SPECIAL-PURPOSE MATERIALS,* VOL 2, *ASM HANDBOOK,* ASM INTERNATIONAL, 1990, P 592-633
3. E.W. COLLINGS, *THE PHYSICAL METALLURGY OF TITANIUM ALLOYS,* AMERICAN SOCIETY FOR METALS, 1984
4. TITANIUM, ZIRCONIUM, HAFNIUM, TANTALUM, AND COLUMBIUM, *WELDING HANDBOOK,* 7TH ED., VOL 4, AMERICAN WELDING SOCIETY, 1982, P 434-470

Selection and Weldability of Conventional Titanium Alloys

W.A. Baeslack III, The Ohio State University; J.R. Davis, Davis and Associates; C.E. Cross, Martin Marietta Astronautics Group

General Welding Considerations (Ref 5)

Titanium has a strong chemical affinity for oxygen, and a stable, tenacious oxide layer forms rapidly on a clean surface, even at room temperature. This behavior leads to a natural passivity that provides a high degree of corrosion resistance (Ref 6). The strong affinity of titanium for oxygen increases with temperature, and the surface oxide layer increases in thickness at elevated temperatures (Fig. 3). At temperatures exceeding 500 °C (930 °F), its oxidation resistance decreases rapidly, and the metal becomes highly susceptible to embrittlement by oxygen, nitrogen, and hydrogen, which dissolve interstitially in titanium. Therefore, the melting, solidification, and solid-state cooling associated with fusion welding must be conducted in completely inert or vacuum environments. The fusion welding processes most widely used for joining titanium are gas-tungsten arc welding (GTAW), gas-metal arc welding (GMAW), plasma arc welding (PAW), laser-beam welding (LBW), and electron-beam welding (EBW); each of these processes are discussed later in this article. With the arc and laser welding processes, protection of the weld zone can be provided by localized inert-gas shielding. Complete enclosure in a protective chamber of the high vacuum environment associated with the electron-beam welding process inherently provides better atmospheric protection. In addition to proper shielding, welded component cleanliness (including filler metals) is necessary to avoid weld contamination.

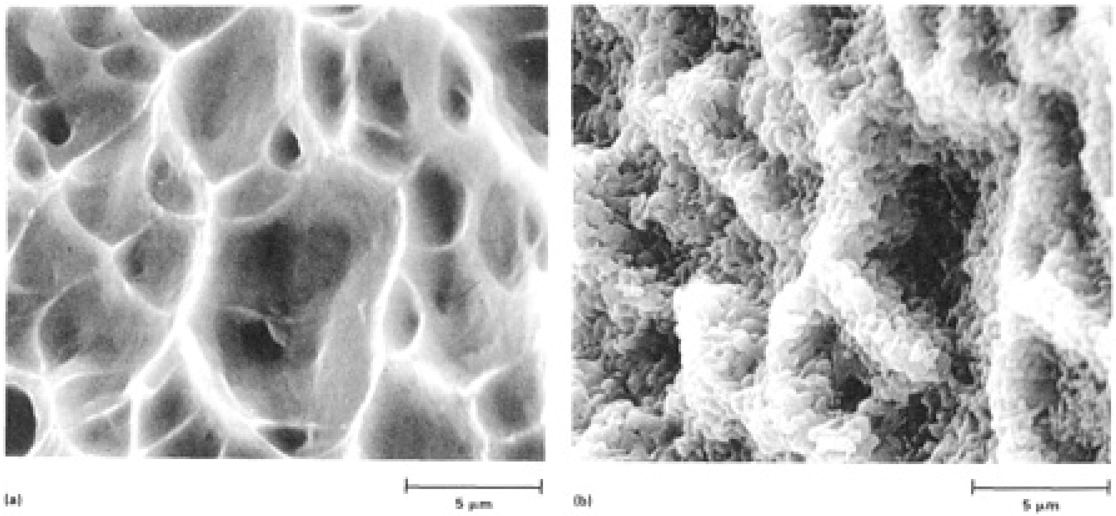

FIG. 3 EFFECT OF A 15 MIN 800 °C (1470 °F) AIR EXPOSURE ON A FRACTURE SURFACE OF AN ANNEALED TI-6AL-6V-2SN ALLOY. (A) FRACTURE APPEARANCE BEFORE EXPOSURE. (B) THE IDENTICAL FRACTURE SURFACE AFTER EXPOSURE; THE OXIDE BUILDUP IS SO GREAT THAT IT IS IMPOSSIBLE TO IDENTIFY THE FRACTURE MODE. COURTESY OF V. KERLINS, MCDONNELL DOUGLAS ASTRONAUTICS COMPANY

The welding of thin-to-moderate section thicknesses in titanium alloys can be accomplished using all of the aforementioned processes. The GTAW process offers the greatest flexibility for both manual and automatic application at minimum capital investment. If production volume is large, the high capital investment required for LBW and EBW systems can be acceptable, based on higher welding rates and improved productivity. For titanium plate thickness exceeding about 5 mm (0.20 in.), the high-energy density processes are the most efficient. Plasma are welding, for producing welds up to about 15 mm (0.6 in.) thick, and EBW, which can readily generate single-pass welds in plates over 50 mm (2 in.) thick, are used in current aerospace practice. In addition to thickness requirements, complex geometries may require manual welding or the use of extensive fixturing. The automatic welding of extremely large components may also prove difficult, particularly with the EBW process.

Residual stresses in titanium welds can greatly influence the performance of a fabricated aerospace component by degrading fatigue properties, while distortion can cause difficulties in the final assembly and operation of high-tolerance aerospace systems. Thus, the use of high-energy-density welding processes to produce full-penetration, single-pass autogenous welds, rather than multipass conventional arc welding, is desirable to minimize these difficulties.

Several important and interrelated aspects of welding phenomena contribute to the overall understanding of titanium alloy welding metallurgy. These factors include melting and solidification effects on weld microstructure, postweld heat treatment effects, and structure/mechanical property/fracture relationships. Each of these are described below as well as in several excellent review articles (Ref 5, 7, 8, 9).

References cited in this section

5. W.A. BAESLACK III, D.W. BECKER, AND F.H. FROES, ADVANCES IN TITANIUM ALLOY WELDING METALLURGY, *J. MET.*, MAY 1984, P 46-58

6. R.W. SCHUTZ AND D.E. THOMAS, CORROSION OF TITANIUM AND TITANIUM ALLOYS, *CORROSION*, VOL 13, *ASM HANDBOOK*, ASM INTERNATIONAL, 1987, P 669-706
7. D.W. BECKER, W.A. BAESLACK III, AND R.W. MESSLER, JR., TITANIUM WELDING, *TITANIUM '80 SCIENCE AND TECHNOLOGY*, VOL 1, H. KIMURA AND O. IZUMI, ED., THE METALLURGICAL SOCIETY OF AIME, 1980, P 255-275
8. D.F. HASSON, "A REVIEW OF TITANIUM WELDING PROCESSES," U.S. NAVY REPORT, ANNAPOLIS, MD, 1985
9. J.R. HARTMAN AND W.A. BAESLACK III, ADVANCES IN ARC WELDING OF TITANIUM ALLOYS, *WELDABILITY OF MATERIALS*, R.A. PATTERSON AND K.W. MAHIN, ED., ASM INTERNATIONAL, 1990, P 319-330

Selection and Weldability of Conventional Titanium Alloys

W.A. Baeslack III, The Ohio State University; J.R. Davis, Davis and Associates; C.E. Cross, Martin Marietta Astronautics Group

Weld Microstructure

The microstructure of the weldment and the extent to which it differs from the thermomechanically processed base material is strongly influenced by the thermal cycle of welding. Two characteristics that are vitally important are the size and shape of the prior-beta grains and the phase transformations that take place during weld cooling. These characteristics are described for both alpha-beta (Ti-6Al-4V) and metastable beta alloys in this section. Also described are weld defects (macrosegregation, microsegregation, and solidification defect formation) and the effects of postweld heat treatment on welded titanium structures.

Alpha-Beta Alloys

Prior-Beta Grain Size. Mechanical properties of composite weld structures in titanium alloys depend on structural characteristics of each weld region, which in turn depend on the specific thermal cycle(s) imposed during welding and on subsequent postweld heat treatment. As the weld macrographs in Fig. 4, 5, 6, and 7 illustrate, the weld fusion zone in titanium alloys is characterized by coarse, columnar prior-beta grains that originate during weld solidification. The size and morphology of these grains depend on the nature of the heat flow that occurs during weld solidification.

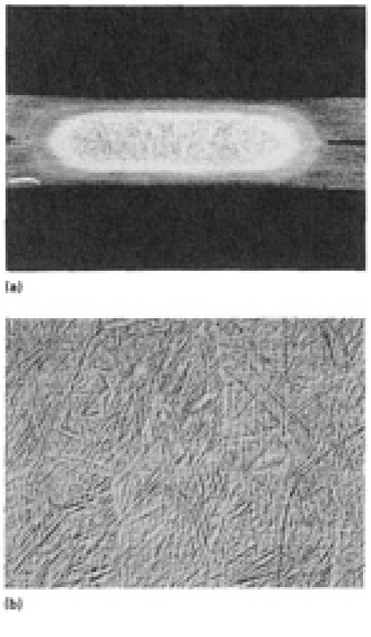

FIG. 4 COLUMNAR BETA GRAINS IN A TI-6AL-4V SPOT WELD. (A) 10×. (B) 240×. COURTESY OF THE

WELDING INSTITUTE

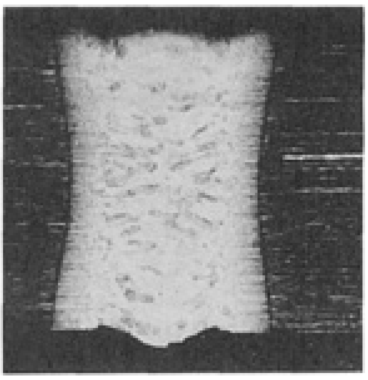

FIG. 5 MACROGRAPH SHOWING COLUMNAR BETA GRAINS IN A TI-6AL-4V LASER-BEAM WELD. 13×. COURTESY OF THE WELDING INSTITUTE

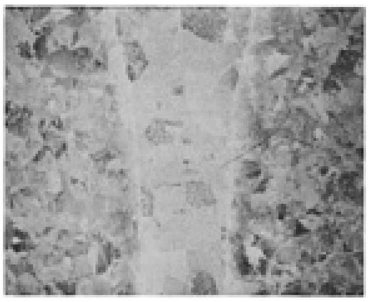

FIG. 6 MACROGRAPH SHOWING COARSE PRIOR-BETA GRAIN SIZE IN AN ELECTRON-BEAM WELDED TI-6AL-4V FORGING. COURTESY OF THE WELDING INSTITUTE

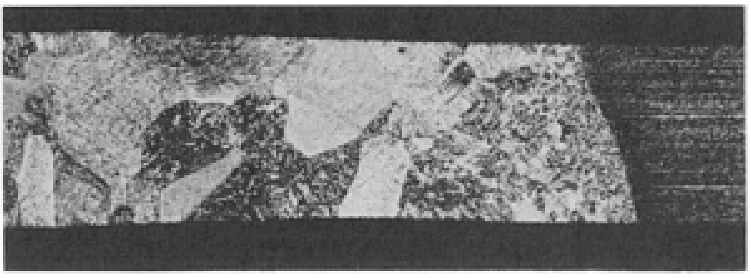

FIG. 7 MACROGRAPH SHOWING MULTIDIRECTIONAL BETA GRAIN MORPHOLOGY IN A TI-6AL-4V GAS-TUNGSTEN ARC WELD. 30×. COURTESY OF THE WELDING INSTITUTE

Under simple, uniaxial heat flow, such as occurs in a spot weld, the beta grains nucleate epitaxially on beta grains in the base-metal substrates and solidify preferentially in a direction parallel to the maximum temperature gradient (i.e., parallel to the welding electrodes) until they impinge at a horizontal weld centerline (Fig. 4). Under two-dimensional heat-flow conditions characteristic of full-penetration plasma-arc, laser-beam, and electron-beam welds, the columnar beta grains solidify inward from the base metal in a direction nearly parallel to the sheet or plate surface, ultimately impinging to form a vertical grain boundary at the weld centerline (Fig. 5 and 6). Three-dimensional or mixed two-dimensional/three-dimensional heat-flow conditions, such as those present in single-pass and multipass gas-tungsten arc and gas-metal arc weldments, promote the formation of more complex, multidirectional beta grain morphologies (Fig. 7).

The fusion-zone beta grain size depends primarily on the weld energy input, with a higher energy input promoting a larger grain size. Due to epitaxial grain growth, the fusion-zone beta grain size may also depend on the beta grain size in the near heat-affected zone (HAZ) directly adjacent to the fusion line. This latter effect of base metal grain size is most significant in the welding of extremely coarse-grained cast or beta-annealed alloys, as shown in Fig. 6 for an electron-beam weld produced in a coarse-grained forging. Because weld mechanical properties, particularly ductility, can be degraded by a coarse prior-beta grain size, it is important to maintain as fine a grain structure as possible by minimizing the weld energy input.

As indicated above, appreciable beta grain growth occurs in the near HAZ directly adjacent to the weld fusion line, where peak temperatures range from between the alloy solidus down to the beta transus (approximately 995 °C, or 1825 °F, for Ti-6Al-4V). As in the fusion zone, the extent of this growth increases with energy input into the weld zone. Consequently, this region can vary markedly in width, being almost unresolvable in electron-beam and laser-beam welds and yet being several beta grains wide in gas-tungsten arc welds (Fig. 5 and 7). Further from the fusion line, temperatures below the beta transus are encountered, which promote transformation of alpha phase in the mill-annealed microstructure to various proportions of the high-temperature beta phase. The presence of even small quantities of alpha phase at peak temperatures in the weld thermal cycle is sufficient to prevent beta grain growth, thereby contributing to the improved ductility of this region as compared with the coarser-grained fusion zone and near HAZ.

During the past two decades, numerous investigators in the former Soviet Union and in the United States have evaluated methodologies for promoting beta grain refinement in titanium alloy fusion welds. This work, which has recently been reviewed (Ref 10), utilized principally electromagnetic stirring techniques and inoculation with rare-earth (RE) elements and titanium "microcooler" particles to promote nonepitaxial heterogenous nucleation of beta grains during solidification and thereby provide a refined fusion zone grain structure. Several of these approaches have been effective in refining the beta grain structure, and correspondingly improving fusion zone mechanical properties. Figure 8 shows the equiaxed prior-beta grain structure in the fusion zone of a gas-tungsten arc weld produced in a Ti-6Al-4V joint using titanium powder microcooler additions. Despite the success of these techniques, particularly in thin sheet materials, their commercial application has been extremely limited, particularly in the United States.

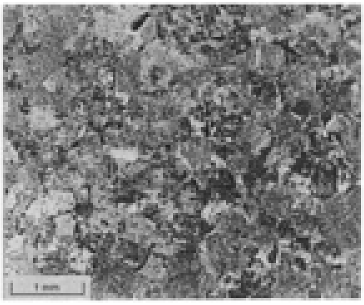

FIG. 8 PHOTOMICROGRAPH SHOWING EQUIAXED PRIOR-BETA MICROSTRUCTURE IN THE FUSION ZONE OF A TI-6AL-4V WELD JOINT. GTAW PROCESS UTILIZED TITANIUM MICROCOOLER ADDITIONS. 12×

Phase Transformations. In addition to prior-beta grain size, weld zone mechanical properties in Ti-6Al-4V are significantly influenced by the manner in which the high-temperature, body-centered-cubic beta phase transforms on cooling to the low-temperature, hexagonal-close-packed phase. Characteristics of this "transformed-beta" microstructure depend principally on the cooling rate from above the beta transus temperature, which is correspondingly influenced by the welding process, process parameters, and other welding conditions (such as workpiece geometry and fixturing). In the near-HAZ regions, high cooling rates associated with low-energy-input welding processes such as laser-beam, electron-beam, and resistance welding (100 to 10,000 °C/s, or 180 to 18,000 °F/s), promote transformation of beta to alpha-prime martensite. This extremely fine, acicular transformation product exhibits high strength and hardness but relatively low ductility and toughness. At the lower cooling rates associated with gas-tungsten arc or plasma-arc welding (10 to 100 °C/s, or 18 to 180 °F/s), a coarser structure of Widmanstätten alpha plus retained beta, or a mixture of this structure and alpha-prime, is produced, which exhibits yield and tensile strengths superior to those of the mill-annealed base metal and a ductility and toughness greater than those of an entirely martensitic microstructure. In the far HAZ, microstructures are comprised of primary alpha phase originating from the base-metal microstructure in a matrix of transformed beta.

The weld microstructure and mechanical properties may also be influenced by postweld heat treatment, with specific postweld heat treatment effects depending on the heat treatment time and temperature and on the as-welded microstructure. See the section on the "Effect of Postweld Heat Treatment" in this article.

Metastable-Beta Alloys

The weld HAZ and fusion zone beta grain macrostructures in metastable-beta titanium alloys are essentially identical to those observed in alpha-beta alloys. Due to the appreciable beta-stabilizer content in metastable-beta titanium alloys, and the relatively slow diffusivity of these elements, the high-temperature beta phase is retained on weld cooling to room temperature (although athermal transformation to omega phase may occur in certain alloys) (Ref 11, 12). Subsequent postweld aging heat treatment at temperatures ranging from about 450 to 650 °C (840 to 1200 °F) promotes the precipitation of fine alpha phase both intragranularly and along prior-beta grain boundaries. Through proper control of the aging temperature and time, a wide range of base metal and weld zone strength/ductility combinations can be achieved. The relatively uniform precipitation of extremely fine alpha phase during low-temperature aging (482 °C, or 900 °F, for 24 h), as shown in Fig. 9(a) and (b) for gas-tungsten arc weld fusion zone in Beta C, promotes high fusion zone hardness and strength (UTS, ultimate tensile strength, of 1382 MPa, or 200 ksi) and low ductility (2.5% elongation). An increased aging temperature promotes coarser alpha precipitation, softening (UTS of 1121 MPa, or 163 ksi), and an increase in ductility (8.0% elongation), as shown in Fig. 9(c) and 9(d) for the gas-tungsten arc weld fusion zone in Beta C aged at 593 °C (1100 °F) for 8 h. It should be noted that in contrast to alpha-beta alloys, residual microsegregation from weld solidification can influence alpha precipitation on aging, particularly for low-temperature aging heat treatments (Fig. 9).

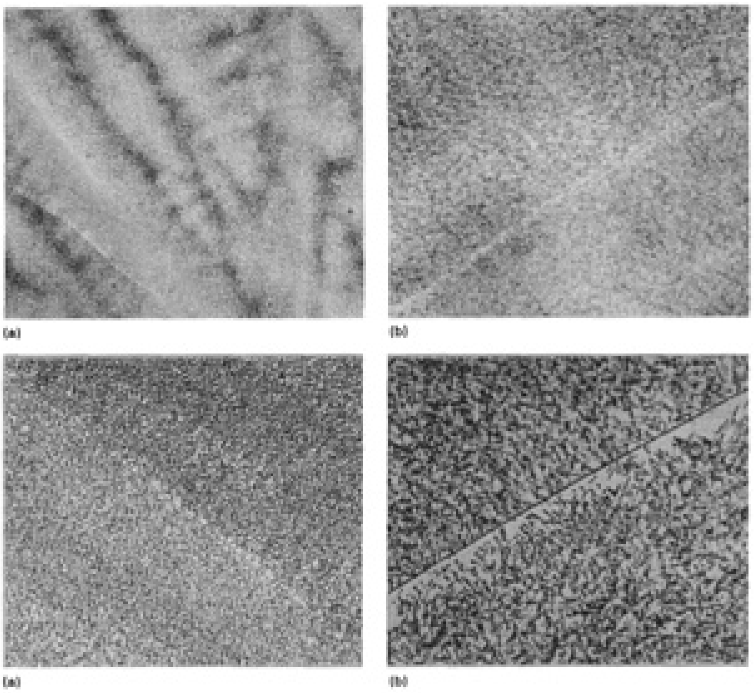

FIG. 9 POSTWELD HEAT TREATED GAS-TUNGSTEN ARC WELD FUSION ZONE IN BETA-C SHEET. (A) AGED AT 482 °C (900 °F) FOR 24 H; 275×. (B) SAME HEAT TREATMENT AS (A); 690×. (C) AGED AT 593 °C (1100 °F) FOR 8 H; 275×. (D) SAME HEAT TREATMENT AS (C); 690×. SOURCE: REF 12

Welding Defects

In many alloy systems, weldability is determined by the capability of the alloy to produce a weld that is free of discontinuities or defects. Defects that may be encountered when welding titanium alloys include:

- SOLIDIFICATION SEGREGATION (MACROSEGREGATION AND MICROSEGREGATION)
- SOLIDIFICATION CRACKING
- CONTAMINATION CRACKING
- HYDROGEN EMBRITTLEMENT

- SUBSOLIDUS (DUCTILITY DIP) CRACKING
- POROSITY

Macrosegregation in weldments is defined as segregation that extends over distances of several grain diameters. In titanium alloy fusion weldments, macrosegregation occurs primarily in the form of transverse solute banding (Ref 5). These solute-enriched or solute-depleted bands normally appear as curvilinear contours on the polished and etched weldment surface and are attributed to thermal variations in the weld pool, which periodically change the solid-liquid interface velocity. Transverse solute banding has occurred using both arc and high-energy welding processes. Studies on solute redistribution in Ti-6Al-4V gas-tungsten arc welds found distinct metallographic evidence of transverse solute banding. Electron microprobe analysis determined the bands resulted from vanadium and aluminum segregation (Ref 13).

The effect of macrosegregation on subsequent beta phase decomposition during cooling depends on the extent of segregation residual at the transformation temperature and the local alloy chemistry. Solute diffusion that occurs after solidification and prior to solid-state phase transformations reduces the magnitude of macrosegregation. In near-alpha and alpha-beta alloys, little effect of macrosegregation has been reported. However, in the more heavily beta-stabilized alloys, strong macrosegregation effects have been observed (Ref 14). This effect is shown in Fig. 10 for a gas-tungsten arc weld fusion zone produced between the alpha-beta alloy Ti-6Al-4V and the metastable-beta alloy Ti-15V-3Cr-3Al-3Sn. Transverse solute bands in the weld fusion zone, which were found to be depleted in the beta stabilizers vanadium and chromium, transformed to martensite while the remaining more heavily beta-stabilized regions were retained as beta phase during weld cooling. A similar effect of macrosegregation due to transverse-solute banding and incomplete mixing was recently observed for dissimilar-alloy laser welds produced between Ti-6Al-4V and the metastable-beta alloys Beta-C and Beta-21S (Ref 15, 16).

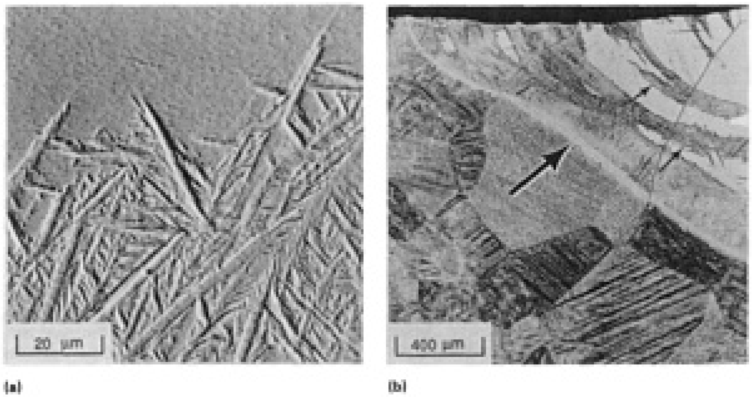

FIG. 10 LIGHT (A) AND SCANNING ELECTRON (B) MICROGRAPHS SHOWING MARTENSITICALLY-TRANSFORMED TRANSVERSE-SOLUTE BANDS (SMALL ARROWS) IN DISSIMILAR ALLOY GAS-TUNGSTEN ARC WELD BETWEEN TI-6AL-4V AND TI-15V-3AL-3CR-3SN SHEETS. LARGE ARROW IN (A) INDICATES FUSION LINE.

It should be pointed out, however, that the degradation of weld integrity, structure, or properties due to macro- and/or microsegregation is generally not significant in titanium alloys due to the limited extent of segregation of common alloying elements and the appreciable diffusional homogenization of alloying elements during weld cooling through the beta-phase field. This significant effect of alloying element homogenization on cooling through the beta-phase field has been recently demonstrated by lnoue and Ogawa (Ref 17) for gas-tungsten arc welds in Ti-6Al-4V using liquid tin quenching experiments that allowed analysis of the "as-solidified" fusion zone solidification structure. They observed in gas-tungsten arc welds produced in Ti-6Al-4V that the segregation of vanadium and iron to dendrite interstices disappeared at a very early stage of weld cooling following the completion of weld solidification due to their high diffusivity in the beta phase at high temperatures.

Microsegregation. The nonequilibrium solidification conditions experienced during fusion welding result in a breakdown of the advancing solid-liquid interface into cellular, cellular-dendritic, and dendritic substructures (Ref 5). This interface breakdown in commercial titanium alloy welds is associated with the segregation of alloying elements on an intercellular or interdendritic scale. Microsegregation is much more evident in metastable-beta titanium alloy weldments than in alpha-beta weldments. This behavior is due primarily to the appreciably higher alloying levels (and therefore higher absolute compositional differences in segregated regions) and the absence of a transformed structure in the completely retained-beta fusion zone. As indicated earlier, local variations in beta-stabilizer content due to microsegregation can affect the alpha-phase precipitation response during subsequent postweld heat treatment.

Microsegregation effects have also become significant in the joining of advanced, elevated-temperature titanium alloys that contain rare earth elements for dispersoid strengthening (Ref 17). Figure 11 shows the fusion zone of a laser weld in a rapidly-solidified, powder metallurgy Ti-8Al-2.8Sn-5.4Hf-3.6Ta-1Y-0.2Si (wt%) alloy. The segregation of yttrium to dendrite interstices during weld solidification promotes the nucleation and growth of yttrium-enriched dispersoids in these regions both during the final states of solidification and possibly in the solid-state during cooling. Similar segregation of yttrium and dispersoid formation would be expected in gas-tungsten arc welds, but with a coarser grain and dispersoid size.

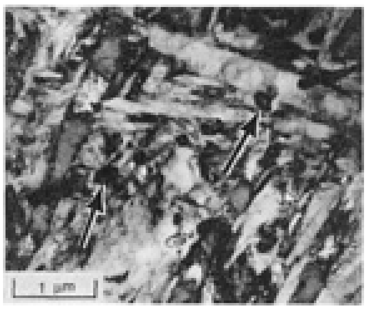

FIG. 11 TRANSMISSION ELECTRON MICROGRAPH OF LASER WELD FUSION ZONE IN RAPIDLY-SOLIDIFIED/POWDER METALLURGY TI-8AL-2.8SN-5.4HF-3.6TA-1Y-0.2SI ALLOY SHOWING YTTRIUM-ENRICHED DISPERSOIDS (INDICATED BY ARROWS) LOCATED ALONG SOLIDIFICATION SUBSTRUCTURE BOUNDARIES. SOURCE: REF 17

Solidification Cracking (Ref 9). Relative to many other structural alloys, such as aluminum alloys and many austenitic stainless steels, titanium alloys are generally not considered susceptible to fusion zone solidification cracking. Studies have demonstrated, however, that under severe conditions of restraint, solidification cracking along columnar beta grain boundaries can occur (Ref 18). Centerline cracking in a gas-tungsten arc welded alpha-beta alloy Ti-6Al-6V-2Sn is shown in Fig. 12. Recent work by Prokhorov *et al.* (Ref 19) has further evaluated the solidification cracking of Ti-6Al-4V using a weldability test in which a gas-tungsten arc weld is strained during the welding operation. They reported the formation of solidification cracks oriented along the welding direction and a quantitative susceptibility comparable to some steels, and concluded that solidification cracking must be considered when designing joints in Ti-6Al-4V structural components. However, this observation generally conflicts with the well-known excellent cracking resistance of this alloy, and suggests that their test imposed an excessive strain condition not representative of that experienced during conventional arc welding.

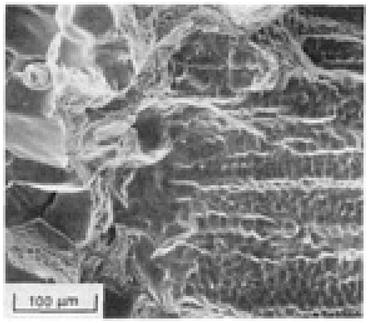

FIG. 12 SEM FRACTOGRAPH SHOWING SURFACE OF FUSION ZONE SOLIDIFICATION CRACK IN GAS-

TUNGSTEN ARC WELDED TI-6AL-6V-2SN. SOURCE: REF 18

Due to an absence of second-phase dispersoids or precipitate particles, or impurities at grain boundaries, conventional titanium alloys are also generally very resistant to HAZ or weld metal liquation cracking. However, recent welding trials on selected rare earth dispersion-strengthened titanium alloys have indicated that the liquation of these particles in the weld HAZ can occur, with this liquid subsequently wetting adjacent beta grain boundaries. Although cracking along these boundaries has not been observed, it may be expected in the arc welding of these alloys under conditions of high weld restraint.

Contamination Cracking. When titanium is exposed to air, moisture, or hydrocarbons at temperatures exceeding 500 °C (930 °F), it will readily pick up oxygen, nitrogen, carbon, and hydrogen. These small atoms of the interstitial elements will enter the crystal lattice in monatomic form and will migrate to interstitial sites (sites located between titanium atoms). The presence of these interstitial elements serves to inhibit plastic deformation and increase strength, but with a substantial loss in ductility. If contamination levels exceed a certain amount, cracking may ensue from the stresses generated during welding. Levels of oxygen on the order of 3000 ppm in the weld have been known to result in transverse cracking (Ref 20). The alpha phase is particularly susceptible to contamination cracking. The hexagonal crystal lattice of alpha-titanium is distorted by the presence of interstitial atoms. In particular, the *c*/*a* ratio is increased (Fig. 13), which results in a reduced number of active slip planes from three (prism, pyramidal, and basal) down to one (prism or pyramidal) (Ref 21). Slip planes are crystallographic planes upon which dislocations move to achieve plastic deformation.

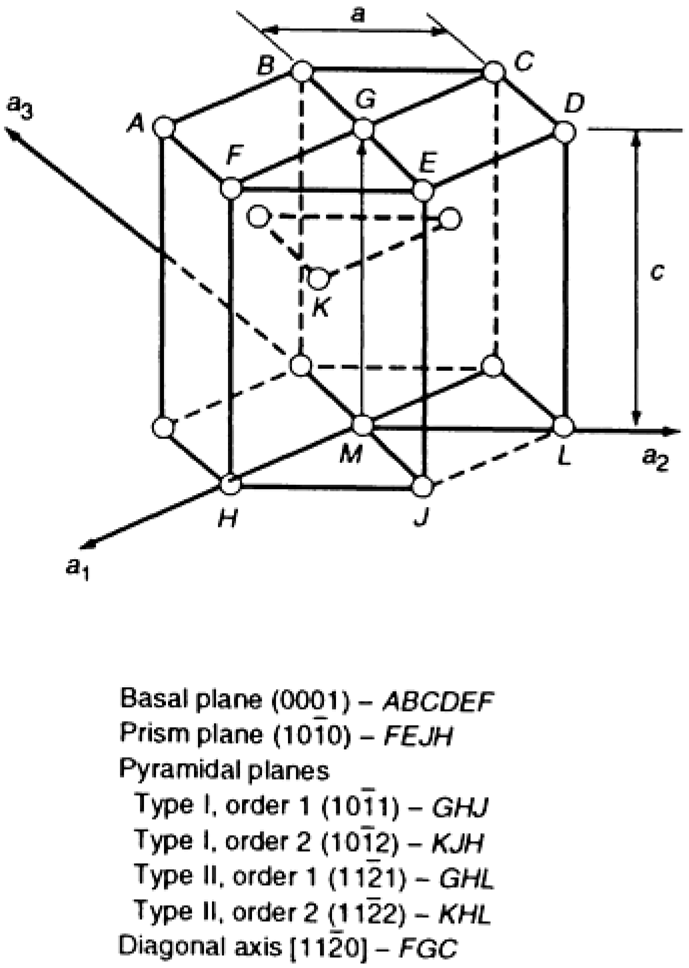
a
B
C
G
D
a3
A
F
E
K
c
M
L
a2
H
J
a1
Basal plane (0001) – ABCDEF
Prism plane (10$\bar{1}$0) – FEJH
Pyramidal planes
Type I, order 1 (10$\bar{1}$1) – GHJ
Type I, order 2 (10$\bar{1}$2) – KJH
Type II, order 1 (11$\bar{2}$1) – GHL
Type II, order 2 (11$\bar{2}$2) – KHL
Diagonal axis [11$\bar{2}$0] – FGC

FIG. 13 DIAGRAM OF THE UNIT CELL FOR A HEXAGONAL CLOSE-PACKED CRYSTAL SHOWING BASAL, PRISM, AND PYRAMIDAL PLANES. THE *C*/*A* RATIO OF THE UNIT CELL FOR ALPHA-TITANIUM IS AFFECTED BY THE PRESENCE OF INTERSTITIAL ATOMS.

Contamination cracking can be avoided by minimizing exposure of the molten weld pool and the heated weld region to interstitial elements. This practice is accomplished by:

- THOROUGHLY DEGREASING THE JOINT PRIOR TO WELDING
- PROVIDING SUFFICIENT INERT GAS SHIELDING TO THE TORCH, TRAILING SHOE, AND BACK SIDE OF THE JOINT
- USING SHIELDING GAS WITH A SUFFICIENTLY LOW DEW POINT

When titanium is oxidized, it assumes different colors depending on alloy content and the degree of oxidation. These colors range from silver to straw, to blue, to white (severe oxidation). While the presence of a blue or white oxide serves as a good indicator that unacceptable contamination has occurred, its absence is no guarantee that the weld is interstitial free.

Hydrogen Embrittlement. While hydrogen can play a role in contamination cracking as discussed above, it can also lead to a more insidious type of defect known as hydrogen embrittlement, or hydrogen delayed cracking. One mechanism proposed for this problem involves the precipitation of titanium hydrides in the vicinity of cracks (Ref 22). If it is assumed that a microcrack is present in the weldment, it follows that hydrogen atoms will diffuse to the crack tip, where a triaxial stress state produces large interstitial sites (holes) in the lattice. These large interstitial sites can more easily accommodate the hydrogen atoms.

Once a critical concentration of hydrogen is reached, titanium hydrides will precipitate ahead of the crack tip (Fig. 14). Hydrogen may also diffuse to the crack to form diatomic gas, increasing pressure within the crack and increasing the load at the crack tip. Eventually the hydrides will rupture, allowing the crack to advance. Upon rupture, the hydrides will redissolve and hydrogen will diffuse to the new crack tip, thus reinitiating the sequence. This cycle may repeat many times, each cycle resulting in a small advancement of the crack.

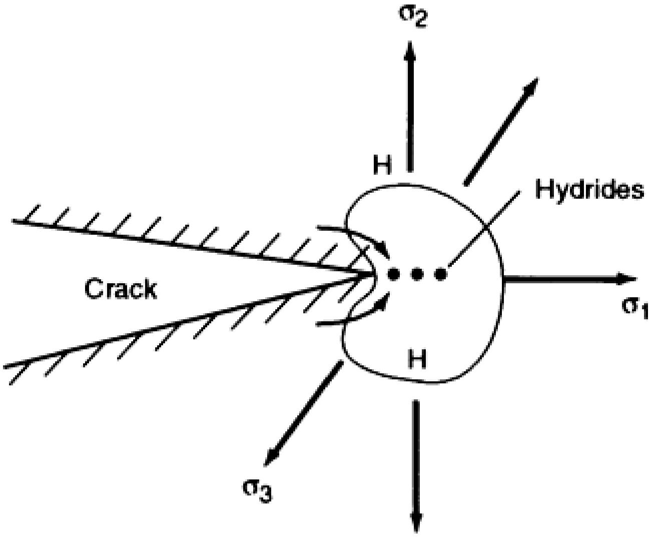

FIG. 14 SCHEMATIC DIAGRAM SHOWING HOW HYDROGEN ATOMS MAY DIFFUSE TO A CRACK TIP IN A TITANIUM WELDMENT TO FORM TITANIUM HYDRIDES

Because the hydrogen embrittlement sequence requires diffusion of hydrogen, the rate of crack advancement depends on time and temperature. Hours, days, or months may pass following welding before significant cracking is observed. This problem can be avoided by minimizing exposure of the weld to hydrogen (hydrocarbons and moisture). Hydrogen concentrations on the order of 200 ppm are known to cause cracking problems (Ref 15), although the combined presence of other interstitials (for example, oxygen or nitrogen) may lower this tolerance limit. The inert shielding gas should have a dew point of at least -50 °C (-60 °F) (34 ppm H_20). Welding-grade argon or helium gas (grade 4.5 with 50 ppm impurities) may not always satisfy this requirement. Also, the use of postweld heat treatment is beneficial because it relieves residual stress, thus reducing the driving force for crack advancement.

Subsolidus Cracking. Certain alloy compositions and microstructures have been found to be susceptible to a severe ductility loss (ductility dip) in a temperature range from about 750 to 850 °C (1380 to 1560 °F) on cooling from temperatures above the beta transus. In particular, alpha-beta alloys that contain niobium rather than vanadium as a beta-stabilizing element are more susceptible to this form of cracking. Microstructures having alpha films at prior-beta grain boundaries, with coarse Widmanstätten alpha colonies, are also more prone to subsolidus cracking.

Lewis *et al.* (Ref 23) utilized a Gleeble system and a typical gas-tungsten arc weld thermal cycle to evaluate the sensitivity of several near-alpha and alpha-beta titanium alloys to this phenomenon. The results of their analysis indicate that Ti-6211 (Ti-6Al-2Nb-1Ta-0.8Mo) exhibits the largest ductility dip and will be the most likely to crack during weld cool down, provided that the welding stresses are high enough (Fig. 15). Ductility dip cracks are intergranular, but show signs of microplasticity (ductile dimples) when examined at high magnification. They are believed to occur due to volumetric differences as a result of the beta-to-alpha transformation. It has been proposed that when the prior-beta grains transform to alpha, the grain boundaries are loaded in tension, causing rupture within grain boundary alpha films (Ref 24).

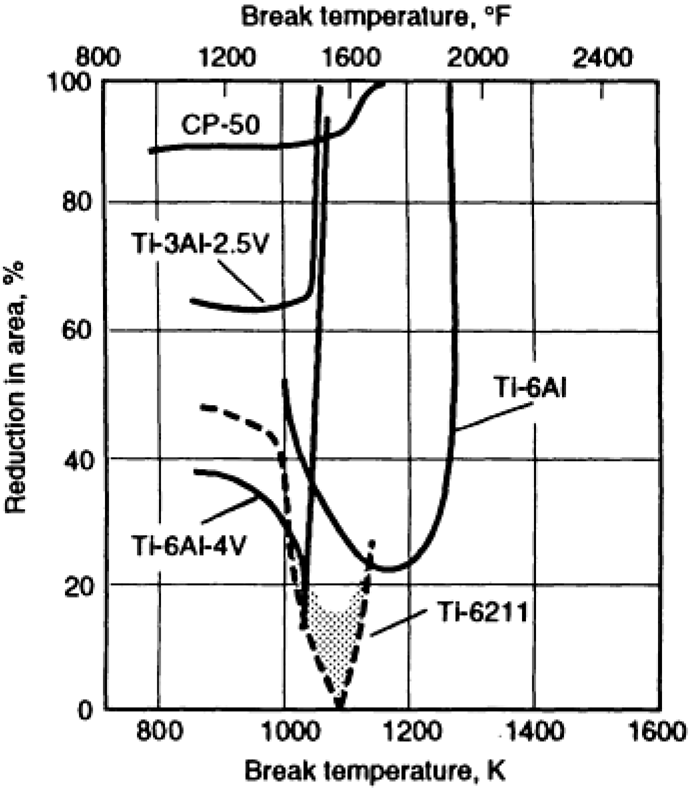

FIG. 15 DUCTILITY VERSUS TEMPERATURE FOR FIVE TITANIUM ALLOYS WITH VARYING ALUMINUM CONTENTS. SHADED AREAS INDICATE EXCESSIVE DATA SCATTER. SOURCE: REF 23

Considerable additional work has been performed in order to understand the origin of ductility-dip weld cracking in titanium alloys (Ref 25, 26, 27, 28). It is important to note, however, that difficulties with the occurrence of such defects in actual welds have not been reported in the literature, and therefore the potential for their formation is quite limited and may be encountered only under conditions of extreme restraint and for certain cooling rates.

Weld metal porosity originates at the trailing edge of the weld pool, where interstitial elements (oxygen or hydrogen) are partitioned between dendrites during solidification (Fig. 16). Partitioning occurs because of the large drop in oxygen (or hydrogen) solubility when transforming from liquid to solid.

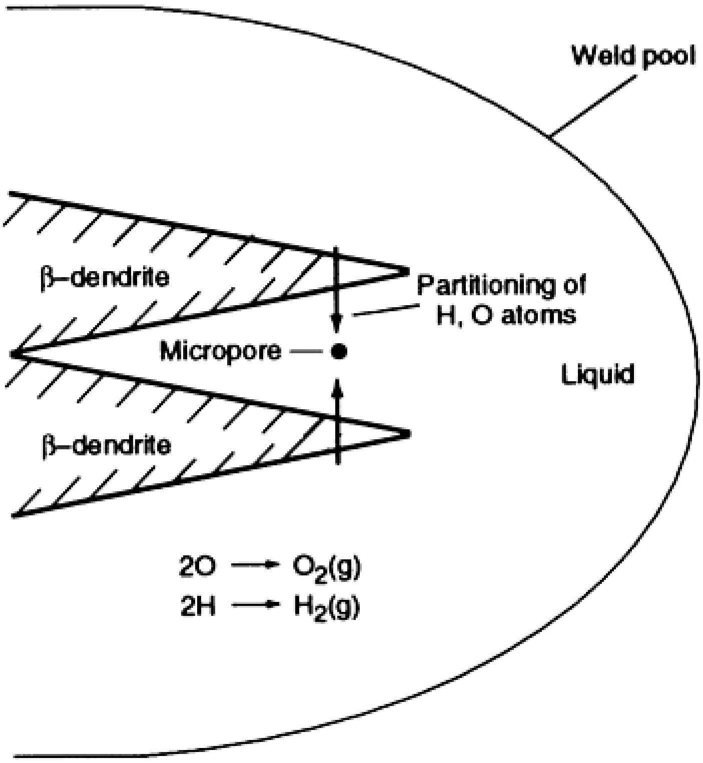

FIG. 16 SCHEMATIC DIAGRAM SHOWING HOW MICRO-PORES ARE NUCLEATED BETWEEN BETA-TITANIUM DENDRITES DURING WELD SOLIDIFICATION

As the concentration of monatomic oxygen (or hydrogen) in the interdendritic liquid reaches a critical concentration, micropores (~1 μm diam) of diatomic gas are nucleated. In order to nucleate, these micropores must overcome atmospheric pressure, liquid head pressure, and liquid/gas surface tension (Ref 29). Once nucleated, these micropores will either become entrapped interdendritically or will coalesce into larger pores (up to 3 mm, or 0.12 in., diam). These macropores will either become enveloped by the advancing solid/liquid interface or they will work their way to the weld pool surface and escape under buoyancy forces.

Avoiding porosity requires that the molten weld pool be protected from exposure to oxygen or hydrogen during welding. This means that the joint must be thoroughly degreased prior to welding and that adequate inert gas shielding with a low dew point must be provided. Porosity can occur at lower impurity levels than is required for contamination cracking, although the existence of porosity implies the possibility of contamination or hydrogen cracking. Porosity is also affected by process parameters and welding position. Welding at slow travel speeds favors the escape of gas pores, as does welding in the flat or vertical-up position.

Effect of Postweld Heat Treatment

Titanium alloy weldments are postweld heat treated for (Ref 5):

- RELIEF OF RESIDUAL STRESSES IMPARTED DURING THE WELDING OPERATION
- MODIFICATION OF THE METASTABLE WELD ZONE MICROSTRUCTURE
- COMPLETION OF THE SECOND STEP IN A TWO-STEP BASE METAL HEAT TREATMENT
- STABILIZATION OF THE MICROSTRUCTURE FOR WELD STRUCTURE APPLICATIONS AT ELEVATED TEMPERATURES

Numerous investigators have studied the influence of postweld heat treatment on the structure and properties of titanium alloy welds (Ref 30, 31, 32, 33, 34, 35). The structures commonly encountered include: alpha and hexagonal martensite in near-alpha and linear alpha-beta alloys, alpha and orthorhombic martensite in heavily beta-stabilized alpha-beta alloys, and retained beta in metastable-beta alloys.

Alpha and Alpha-Beta Alloys. Heat treatment of rapidly cooled welds in near-alpha and alpha-beta alloys at relatively low temperatures (between about 450 and 600 °C, or 840 and 1110 °F), which are commonly employed for stress relief, age the martensitic structure and can further increase its strength and decrease its ductility and toughness. At higher postweld heat treatment temperatures, which are limited to about 760 °C (1400 °F) for large fabricated structures, tempering of the martensite and subsequent coarsening of the resultant alpha phase promote softening of the fusion zone with concurrent increases in ductility and toughness. Microstructures comprised of Widmanstätten alpha plus retained beta, which are produced at slower weld cooling rates, are influenced to a lesser extent by postweld heat treatment at these low-to-moderate temperatures. However, welds exhibiting these microstructures are still normally postweld treated to relieve residual stresses.

Metastable Beta Alloys. In contrast to near-alpha and alpha-beta alloys, metastable-beta alloys can undergo appreciable structural changes at the lower temperatures. In the as-welded condition, the alloy exhibits a retained beta microstructure that contains athermal omega, an extremely fine phase that is unresolvable with light microscopy. On aging isothermally at 480 °C (900 °F), extremely fine alpha forms intergranularly and in a continuous manner at beta grain boundaries. As the heat treatment temperature is increased to 565 °C (1050 °F) and to 650 °C (1200 °F), both alpha structures coarsen appreciably. From a practical standpoint, the post-weld heat treatment of welded structures is difficult and expensive because inert shielding is required at temperatures above about 500 °C (930 °F). Additionally, the loss of strength at temperatures approaching the beta transus limits the capability to maintain weld tolerances unless complex fixturing is employed.

Table 3 lists the temperature ranges used for stress-relieving titanium alloy weldments. Postweld heat treatments that employ high temperatures (exceeding 800 °C, or 1470 °F) with water quenching followed by an aging treatment between 350 to 500 °C (660 to 930 °F) have also been developed for near-alpha and alpha-beta alloys (Ref 36). Tensile properties of postweld heat-treated titanium weldments are given in Table 4.

TABLE 3 RECOMMENDED STRESS-RELIEF TREATMENTS FOR TITANIUM AND TITANIUM ALLOYS

Parts can be cooled from stress relief by either air cooling or slow cooling.

ALLOY	TEMPERATURE		TIME, H
	°C	°F	
COMMERCIALLY PURE TITANIUM (ALL GRADES)	480-595	900-1100	$\frac{1}{4}$ -4
α OR NEAR- α TITANIUM ALLOYS			
TI-5AL-2.5SN	540-650	1000-1200	$\frac{1}{4}$ -4
TI-8AL-1MO-1V	595-705	1100-1300	$\frac{1}{4}$ -4
TI-2.5CU (IMI 230)	400-600	750-1110	0.5-24
TI-6AL-2SN-4ZR-2MO	595-705	1100-1300	$\frac{1}{4}$ -4
TI-6AL-5ZR-0.5MO-0.2SI (IMI 685)	530-570	980-1050	24-48
TI-5.5AL-3.5SN-3ZR-1NB-0.3MO-0.3SI (IMI 829)	610-640	1130-1190	1-3
TI-5.8AL-4SN-3.5ZR-0.7NB-0.5MO-0.3SI (IMI 834)	625-750	1160-1380	1-3
TI-6AL-2CB-1TA-0.8MO	595-650	1100-1200	$\frac{1}{4}$ -2
TI-0.3MO-0.8NI (TI CODE 12)	480-595	900-1100	$\frac{1}{4}$ -4
α -β TITANIUM ALLOYS			
TI-6AL-4V	480-650	900-1200	1-4
TI-6AL-7NB (IMI 367)	500-600	930-1110	1-4
TI-6AL-6V-2SN (CU + FE)	480-650	900-1200	1-4

TI-3AL-2.5V	540-650	1000-1200	$\frac{1}{2}$ -2
TI-6AL-2SN-4ZR-6MO	595-705	1100-1300	$\frac{1}{4}$ -4
TI-4AL-4MO-2SN-0.5SI (IMI 550)	600-700	1110-1290	2-4
TI-4AL-4MO-4SN-0.5SI (IMI 551)	600-700	1110-1290	2-4
TI-5AL-2SN-4MO-2ZR-4CR (TI-17)	480-650	900-1200	1-4
TI-7AL-4MO	480-705	900-1300	1-8
TI-6AL-2SN-2ZR-2MO-2CR-0.25SI	480-650	900-1200	1-4
TI-8MN	480-595	900-1100	$\frac{1}{4}$-2
β OR NEAR- β TITANIUM ALLOYS			
TI-13V-11CR-3AL	705-730	1300-1350	$\frac{1}{12}$ - $\frac{1}{4}$
TI-11.5MO-6ZR-4.5SN (BETA III)	720-730	1325-1350	$\frac{1}{12}$ - $\frac{1}{4}$
TI-3AL-8V-6CR-4ZR-4MO (BETA C)	705-760	1300-1400	$\frac{1}{6}$ - $\frac{1}{2}$
TI-10V-2FE-3AL	675-705	1250-1300	$\frac{1}{2}$ -2
TI-15V-3AL-3CR-3SN	790-815	1450-1500	$\frac{1}{12}$ - $\frac{1}{4}$

TABLE 4 FUSION ZONE TENSILE PROPERTIES OF GAS-TUNGSTEN ARC-WELDED TITANIUM ALLOYS

ALLOY	**CONDITION$^{(A)}$**	**YIELD STRENGTH**		**ULTIMATE TENSILE STRENGTH**		**ELONGATION, %**
		MPA	**KSI**	**MPA**	**KSI**	
TI-6AL-2SN-4ZR-2MO-0.1SI	BASE METAL (ANNEALED)	915	133	1020	148	13
	AS-WELDED	1050	152	1185	172	3
	PWHT: 595 °C (1100 °F)/6 H AC	1104	160	1254	182	2
	PWHT: 705 °C (1300 °F)/6 H AC	1116	162	1171	170	2
	PWHT: 900 °C (1650 °F)/ $\frac{1}{2}$ H AC + 650 C (1200 °F)/6 H AC	930	135	1020	148	6
CORONA 5	BASE METAL (ANNEALED)	1048	152	1158	168	16
	AS-WELDED	1262	183	1268	184	1
	PWHT: 650 °C (1250 °F)/4 H AC	1179	171	1220	177	1-2
	PWHT: 705 °C (1300 °F)/4 H AC	1144	166	1165	179	2-3
	PWHT: 775 °C (1425 °F)/ $\frac{1}{2}$ H AC + 650 °C (1200 °F)/4 H AC	1048	152	1082	157	5
	PWHT: 725 °C (1695 °F)/ $\frac{1}{2}$ H AC + 650 °C (1200 °F)/4 H AC	1027	149	1055	153	7
TI-15V-3CR-3AL-3SN	BASE METAL (SOLUTION ANNEALED)	730	106	760	110	24

	AS-WELDED	725	105	750	109	20
	PWHT: 565 °C (1050 °F)/4 H AC	1007	146	1063	154	8
	PWHT: 620 °C (1150 °F)/4 H AC	896	130	985	143	9
	PWHT: 675 °C (1250 °F)/4 H AC	794	115	854	124	18

Weld cooling rate, ~100 °C/s (180 °F/s).

Source: Ref 9

(A) PWHT, POSTWELD HEAT TREATMENT; AC, AIR COOLED.

References cited in this section

5. W.A. BAESLACK III, D.W. BECKER, AND F.H. FROES, ADVANCES IN TITANIUM ALLOY WELDING METALLURGY, *J. MET.,* MAY 1984, P 46-58
9. J.R. HARTMAN AND W.A. BAESLACK III, ADVANCES IN ARC WELDING OF TITANIUM ALLOYS, *WELDABILITY OF MATERIALS,* R.A. PATTERSON AND K.W. MAHIN, ED., ASM INTERNATIONAL, 1990, P 319-330
10. W.A. BAESLACK III AND D.L. HALLUM, NATURE OF GRAIN REFINEMENT IN TITANIUM ALLOY WELDS BY MICROCOOLER INOCULATION, *WELD. J. RES. SUPPL.,* VOL 69 (NO. 12), DEC 1990, P 326S-336S
11. D.W. BECKER AND W.A. BAESLACK III, PROPERTY-MICROSTRUCTURE RELATIONSHIPS IN METASTABLE-BETA TITANIUM ALLOY WELDMENTS, *WELD. J. RES. SUPPL.,* VOL 59, 1980, P 85S-92S
12. W.A. BAESLACK III, P.S. LIU, D.P. BARBIS, J.R. SCHLEY, AND J.R. WOOD, POSTWELD HEAT TREATMENT OF GTA WELDS IN A HIGH-STRENGTH, METASTABLE-BETA TITANIUM ALLOY--BETA-C, *PROC. OF 1992 INT. CONF. ON TITANIUM ALLOYS* (SAN DIEGO), 1992
13. A.T. D'ANNESSA, REDISTRIBUTION OF SOLUTE IN FUSION WELDING, *WELD. J.,* VOL 45, 1966, P 569S-576S
14. W.A. BAESLACK III, EFFECT OF SOLUTE BANDING ON TRANSFORMATIONS IN DISSIMILAR TITANIUM ALLOY WELDMENTS, *J. MAT. SCI. LET.,* VOL 1, 1982, P 229-231
15. P.S. LIU, W.A. BAESLACK III, AND J. HURLEY, DISSIMILAR ALLOY LASER WELDING OF TITANIUM: TI-6AL-4V TO BETA-C, *WELD. J. RES. SUPPL.,* IN PRESS
16. P.S. LIU, W.A. BAESLACK III, AND J. HURLEY, LASER WELDING OF TI-6AL-4V TO BETA-21S, *PROC. OF TMS CONF. ON BETA TITANIUM ALLOYS* (DENVER, CO), TMS, 1993
17. W.A. BAESLACK III, S. CHIANG, AND C.A. ALBRIGHT, *J. MAT. SCI. LET,* VOL 9, 1990, P 698-702
18. W.A. BAESLACK ILL, *METALLOGRAPHY,* VOL 13, 1980, P 277-281
19. N.N. PROKHOROV, *WELD. PROD.,* VOL 9, 1986, P 30-32
20. P.W. HOLSBERG, "CRACKING AND POROSITY IN TITANIUM ALLOY WELDMENTS," U.S. NAVY DTNSRDC REPORT, 1979
21. F.D. ROSI, D.A. DUBE, AND B.H. ALEXANDER, *TRANS. AIME,* VOL 197, 1953, P 257
22. N.E. PATON AND J.C. WILLIAMS, THE EFFECT OF HYDROGEN ON TITANIUM AND ITS ALLOYS, *HYDROGEN IN METALS,* AMERICAN SOCIETY FOR METALS, 1974, P 409-431
23. R.E. LEWIS, I.L. CAPLAN, AND W.C. COONS, THE ELEVATED TEMPERATURE DUCTILITY DIP PHENOMENON IN ALPHA, NEAR-ALPHA AND ALPHA-BETA TITANIUM ALLOYS, *TITANIUM SCIENCE AND TECHNOLOGY,* VOL 2, G. LÜTJERING, U. ZWICKER, AND W. BUNK, ED., DEUTSCHE GESELLSCHAFT FÜR METALLKUNDE E.V., WEST GERMANY, 1985
24. B.K. DAMKROGER, G.R. EDWARDS, AND B.B. RATH, A MODEL FOR HIGH TEMPERATURE

DUCTILITY LOSS IN ALPHA-BETA TITANIUM ALLOYS, *METALL. TRANS. A,* VOL 18, 1987, P 483-485

25. D.M. BOWDEN AND E.A. STARKE, JR., HOT CRACKING SUSCEPTIBILITY OF TITANIUM 6211, *TITANIUM SCIENCE AND TECHNOLOGY,* VOL 2, G. LÜTJERING, U. ZWICKER, AND W. BUNK, ED., DEUTSCHE GESELLSCHAFT FÜR METALLKUNDE E.V., WEST GERMANY, 1985, P 853
26. D.M. BOWDEN AND E.A. STARKE, JR., *METALL. TRANS. A,* VOL 15, 1984, P 1687-1698
27. D. HAYDUK, B.K. DAMKROGER, G.R. EDWARDS, AND D.L. OLSON, *WELD. J.,* VOL 65, 1986, P 251S-260S
28. B.K. DAMKROGER, G.R. EDWARDS, AND B.B. RATH, *WELD. J.,* VOL 68, 1989, P 290S-302S
29. R.E. TREVISAN, D.D. SCHWEMMER, AND D.L. OLSON, THE FUNDAMENTALS OF WELD PORE FORMATION, *WELDING: THEORY AND PRACTICE,* ELSEVIER, 1990, P 79-115
30. K. BORGGREEN AND I. WILSON, USE OF POSTWELD HEAT TREATMENTS TO IMPROVE DUCTILITY IN THIN SHEETS OF TI-6AL-4V, *WELD. J.,* VOL 59, 1980, P 1S-8S
31. S.L. BOSTON AND W.A. BAESLACK III, "MICROSTRUCTURE/MECHANICAL PROPERTY RELATIONSHIPS IN GTA WELDED TI-10V-2FE-3AL," AFML-TM-LL-80-1, WRIGHT PATTERSON AIR FORCE BASE, DAYTON, OH, 1980
32. M.A. GREENFIELD AND C.M. PIERCE, POSTWELD AGING OF A METASTABLE BETA TITANIUM ALLOY, *WELD. J.,* VOL 52, 1973, P 524S-527S
33. W.A. BAESLACK III, EVALUATION OF TRIPLEX HEAT TREATMENTS FOR TITANIUM ALLOYS, *WELD. J.,* VOL 61, 1982, P 193S-199S
34. D.W. BECKER AND W.A. BAESLACK III, PROPERTY/MICROSTRUCTURE RELATIONSHIPS IN METASTABLE-BETA TITANIUM ALLOY WELDMENTS, *WELD. J.,* VOL 59, 1980, P 850-920
35. W.A. BAESLACK III AND F.D. MULLINS, RELATIONSHIP OF MICROSTRUCTURE, MECHANICAL PROPERTIES AND FRACTOGRAPHIC FEATURES IN WELDED HIGH TOUGHNESS TITANIUM ALLOYS, *TRENDS IN WELDING RESEARCH,* S.A. DAVID, ED., AMERICAN SOCIETY FOR METALS, 1982, P 541
36. V.N. ZAMKOV *ET AL.,* WAYS OF PROVIDING THE HIGH SERVICEABILITY OF HEAT-HARDENED WELDED JOINTS OF ALPHA + BETA TITANIUM ALLOYS, *TITANIUM SCIENCE AND TECHNOLOGY,* VOL 2, G. LÜTJERING, U. ZWICKER, AND W. BUNK, ED., DEUTSCHE GESELLSCHAFT FÜR METALLKUNDE, E.V., WEST GERMANY, 1985, P 877

Selection and Weldability of Conventional Titanium Alloys

W.A. Baeslack III, The Ohio State University; J.R. Davis, Davis and Associates; C.E. Cross, Martin Marietta Astronautics Group

Titanium Weldment Structure/Property Relationships (Ref 5, 9)

Macro- and microstructural characteristics of weldments in titanium alloys, as influenced by weld solidification, cooling rate, and postweld heat treatment, significantly influence weldment mechanical properties. In the welding of thin sheet materials for aerospace applications, the principal properties of interest are room- and elevated-temperature tensile properties and *S-N* fatigue characteristics. For welds produced in plate materials for either aerospace or marine applications, fracture toughness and fatigue-crack growth rate properties also become important.

Tensile Properties. The strength across an arc weld is determined by the strength of the weakest link in the composite weld region, which consists of the fusion zone, the near HAZ (where the peak temperature exceeded the beta transus temperature), and the far HAZ (where peak temperatures were below the beta transus). In the United States, sheet products of near-alpha or alpha-beta titanium alloys are most commonly welded in a mill-annealed condition (Table 2). In all but extremely slowly cooled manual arc welds, the fine alpha or alpha-prime martensite structures produced in the weld fusion zone and near-HAZ regions exhibit hardnesses and strengths above that of the affected base metal, and transverse weld failures consequently occur in the base metal (Table 4). Although these welds exhibit 100% joint efficiency, the ductility of the weld zone as measured by longitudinal-weld-oriented bend or all-weld-metal tensile tests may be unacceptably low, particularly in rapidly cooled arc welds in the more heavily beta-stabilized alloys (for example, Ti-6Al-6V-2Sn, CORONA 5, and Ti-6Al-2Sn-4Zr-6Mo), which contain hard martensitic microstructures within a coarse prior-beta grain macrostructure. As shown in Table 4 for gas-tungsten arc welds in the near-alpha alloy Ti-6Al-2Sn-4Zr-2Mo-0.1Si and the alpha-beta alloy CORONA 5, the tempering of martensite in the weld zone and ultimate coarsening of the alpha + beta microstructure result in a decrease in strength but an increase in ductility with increasing heat treatment temperature.

Arc welds produced on near-alpha and alpha-beta base metals that have been solution heat treated and aged to high strength may exhibit weld zones exhibiting strengths below that of the base metal; consequently, transverse tensile fracture would occur preferentially in these regions. For example, the strength and ductility of weldments produced in Ti-6Al-4V that has been solution heat treated and aged to high strength levels are highly dependent on the weld zone structure. The completely martensitic microstructures exhibited in electron-beam and laser-beam welds having high depth-to-width ratios are sufficiently strong to promote base-metal fracture in the solution heat treated and aged base metal. However, the softer Widmanstätten plus retained-beta microstructures exhibited by the more slowly cooled gas-tungsten arc and plasma arc welds may promote preferential fracture in the weld zone at low macroscopic elongation.

As expected, the tensile properties of weldments in metastable-beta titanium alloys differ considerably from those of near-alpha and alpha-beta alloys. If arc welds are produced in solution-annealed base material, joints commonly fracture in the base metal because of a slightly higher strength in the weld metal due to oxygen pickup or the initiation of second-phase precipitation if cooling rates are sufficiently slow. Following postweld heat treatment, the presence of a coarse beta grain structure in the fusion zone combined with the formation of alpha preferentially along these grain boundaries (in addition to precipitation homogeneously within the beta grains) can promote fracture in the fusion zone at lower ductilities. As shown in Table 4 for the metastable-beta titanium alloy Ti-15V-3Al-3Sn-3Cr, postweld heat treatment significantly increases the fusion zone strength versus the as-welded condition. As indicated, this strength is maximum for low aging temperatures and decreases with an increase in aging temperature.

Figure 17 plots longitudinal weld tensile data for a variety of alloys. The most desirable weldment characteristic is the highest strength level that exhibits adequate ductility. Therefore, alloys and welding conditions would be selected that offer properties near the upper boundary of the scatter band shown in Fig. 17. Within the high-strength regions shown on such diagrams, the metastable beta and richer alpha-beta alloys show the most promise (Ref 7).

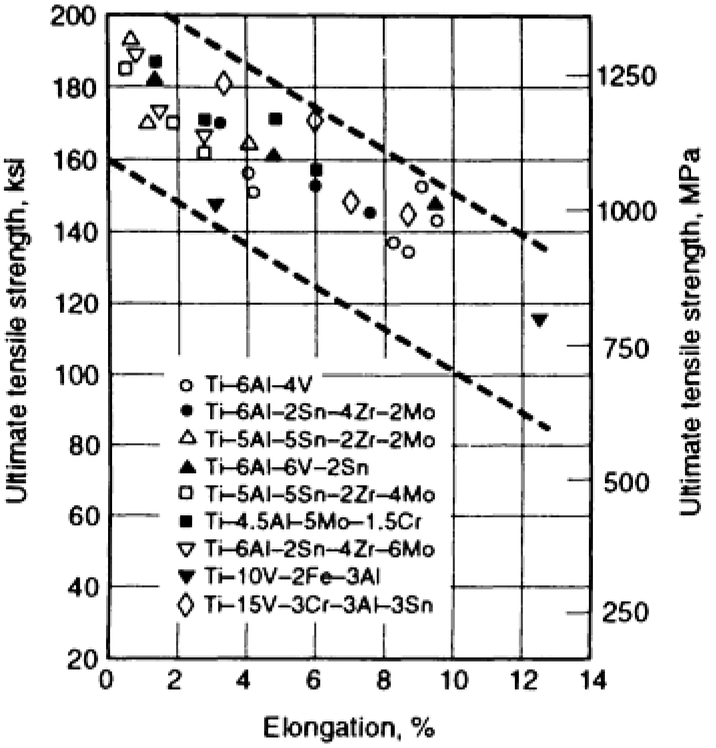

FIG. 17 LONGITUDINAL WELDMENT TENSILE DATA FOR COMMERCIAL TITANIUM ALLOYS. SOURCE: REF 7

The fracture toughness of suitably heat-treated titanium alloy weldments is generally equivalent or superior to that of the alpha-beta processed base material. The good fracture toughness of the fine Widmanstätten plus grain-boundary alpha microstructure in the weld fusion zone and HAZ is consistent with the good toughness of microstructurally similar beta-processed titanium alloy microstructures. Figure 18 shows fracture toughness versus ultimate-tensile strength for heat-treated electron-beam and plasma arc welds in wrought CORONA 5 base plate. Note that the values fall well within the scatterband of data for base-metal alloys exhibiting a lenticular alpha morphology (Ref 37).

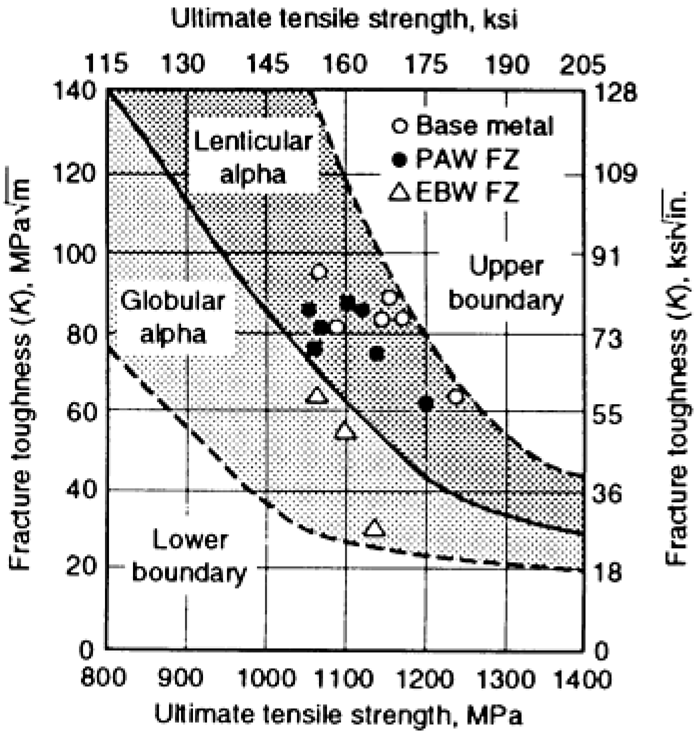

FIG. 18 FRACTURE TOUGHNESS VERSUS ULTIMATE TENSILE STRENGTH FOR CORONA 5 BASE METAL AND POSTWELD HEAT TREATED ELECTRON-BEAM AND PLASMA-ARC WELD FUSION ZONE REGIONS. SOURCE: REF 37

Fracture toughness is also markedly influenced by the interstitial content of the weld metal. When selecting a titanium alloy for a fracture-toughness-critical application, the use of extra-low interstitial (ELI) grades of base metal and consumable are recommended.

Lewis *et al.* (Ref 38) investigated the influence of boron on the dynamic tear toughness of gas-tungsten arc weldments in Ti-6Al-2Nb-1Ta-1Mo and found that welds containing a high boron content of 0.005 wt% exhibited toughness levels 40% below those containing 0.001 wt% B (boron can be introduced as an impurity in master alloys or during processing of welding wire). They attributed this difference to the occurrence of a beta + TiB to alpha peritectoid reaction during weld cooling in the alloy with a high boron content, which provided nucleation sites for alpha platelets. The absence of this transformation in low-boron weldments allowed greater undercooling of the beta phase and transformation to a martensitic structure.

Fatigue Properties. Generally, *S-N* fatigue properties of weld structures in the alpha-beta alloys such as Ti-6Al-4V and Ti-6Al-6V-2Sn are similar to that of the base material if the weld zones are free of defects such as porosity and undercut. Failure in transverse-oriented specimens occurs in the base metal remote from the weld zone. This greater resistance to crack initiation of these composite weld structures relates to the higher strength levels of the weld regions because the large prior-beta grain/transformed beta microstructures, which characterize these regions, generally exhibit reduced resistance to cyclic crack initiation compared to equiaxed-alpha type base metal structures. Studies of *S-N* fatigue properties of electron-beam-welded ELI CORONA 5 indicated that fatigue properties of all-fusion-zone specimens and transverse-weld specimens are comparable to those of the base metal over a range of stresses, even when fracture was initiated by weld porosity (Ref 35). Results from these studies are shown in Fig. 19.

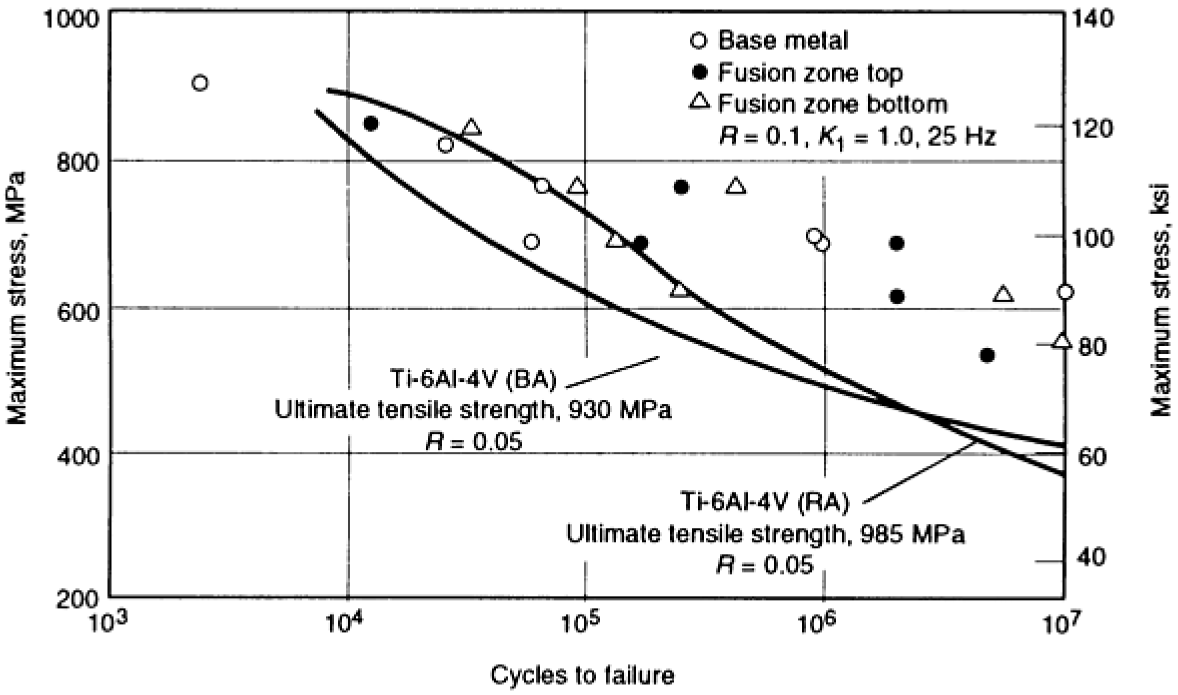

FIG. 19 AXIAL TENSION FATIGUE DATA FOR POSTWELD HEAT TREATED ELI CORONA 5 BASE METAL AND ELECTRON-BEAM-WELDED SPECIMENS COMPARED TO BETA-ANNEALED (BA) AND RECRYSTALLIZED ANNEALED (RA) TI-6AL-4V. SOURCE: REF 35

Although, as indicated above, equivalent weld zone and base-metal fatigue properties can be obtained, lower weld zone axial fatigue properties are commonly observed in actual practice. This is particularly true with high-energy-density welding processes (Fig. 20). This behavior is attributed to the presence of fine porosity, often unresolvable by radiography, and of other defects such as undercuts or underfills in fusion welds, or lack of bonding regions at the weld interface in solid-state diffusion or friction welds. As with tensile properties, the axial fatigue behavior of welds produced in solution heat treated and aged base metals depends on both the weld integrity and the strength of the weld zone.

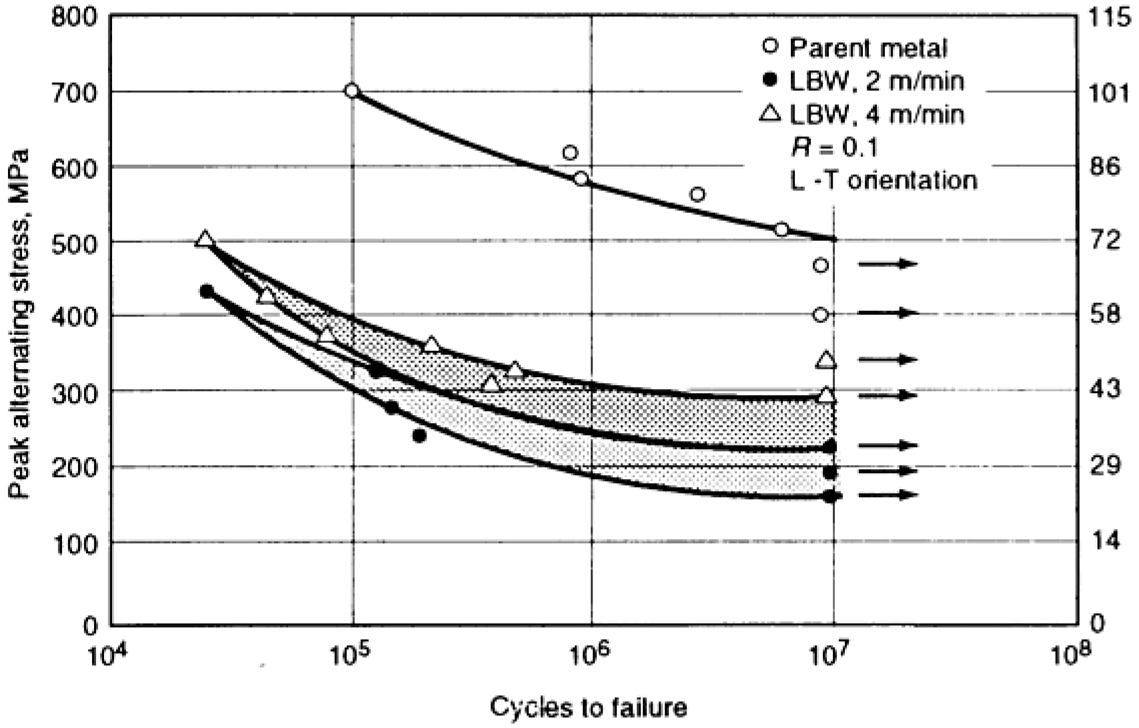

FIG. 20 ***S/N*** **FATIGUE CURVES FOR LASER-BEAM WELDS IN 4 MM (0.16 IN.) TI-6AL-4V SHEET PRODUCED AT 2 AND 4 M/MIN (0.6 AND 1.2 FT/MIN). TESTED IN AS-WELDED CONDITION. FRACTURE INITIATED AT WELD UNDERCUT. BASE METAL PROPERTIES ARE PROVIDED FOR COMPARATIVE PURPOSES. SOURCE: REF 39**

It should be noted that poor axial fatigue properties also can limit the application of lap joints produced in titanium using resistance, diffusion, and ultrasonic welding. In the case of resistance welds, the mechanical notch at the perimeter of the weld nugget, the presence of a brittle martensitic microstructure, and residual stresses can promote early fatigue crack initiation and low fatigue life. Although the metallurgical structure and residual stresses can be modified through process control, the presence of a notch is inherent to a lap joint and subject to less modification.

With the increased application of damage-tolerant design criteria, fatigue crack growth rate data for weldments are of increased importance. A comparison of fatigue-crack propagation rates between gas-tungsten arc, laser-beam, and electron-beam welds in Ti-6Al-4V and Ti-15V-3Al-3Cr-3Sn was made by Pao *et al.* (Ref 40). Their results indicated that the average fatigue crack growth rates through the postweld heat treated weld zones did not differ appreciably with the welding process (Fig. 21), although the scatter bands were narrower for the high-energy density versus the arc welding processes (Fig. 22). For Ti-6Al-4V, they found the average crack growth rates to be lower than that in the base metal, while in Ti-15V-3Al-3Cr-3Sn they found nearly equivalent base metal and weld metal crack growth rates (Fig. 21).

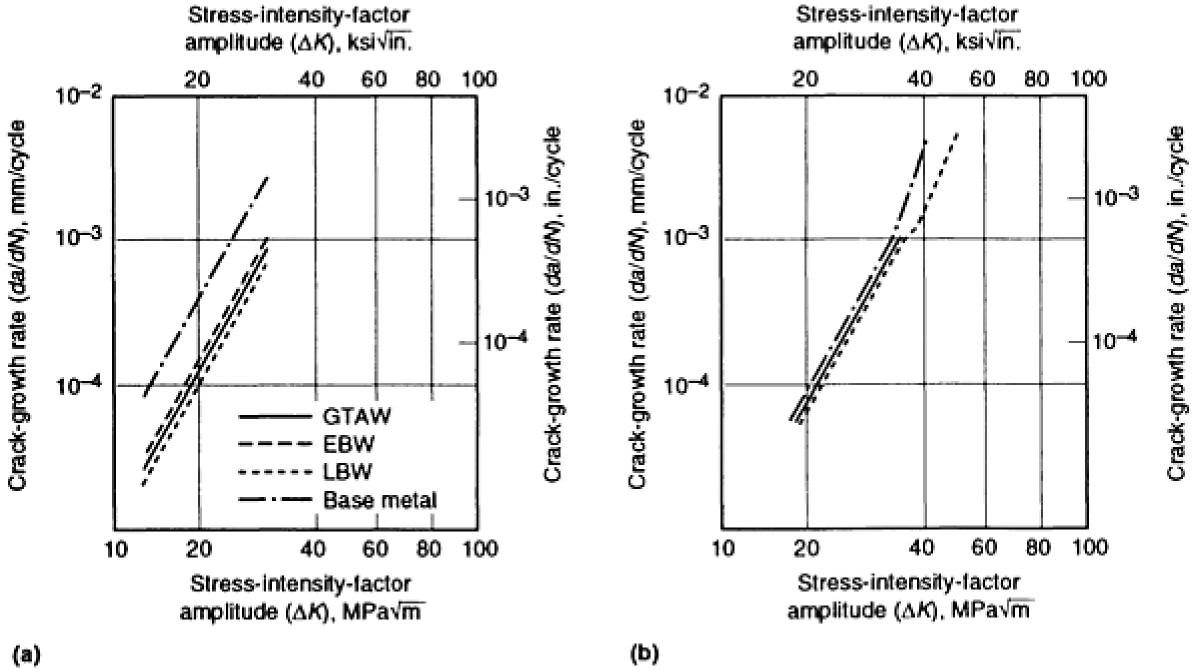

FIG. 21 EFFECT OF WELDING PROCESSES ON THE FATIGUE-CRACK-GROWTH RATE OF LONGITUDINALLY ORIENTED (A) TI-6AL-4V AND (B)TI-15V-3CR-3AL-3SN. SOURCE: REF 40

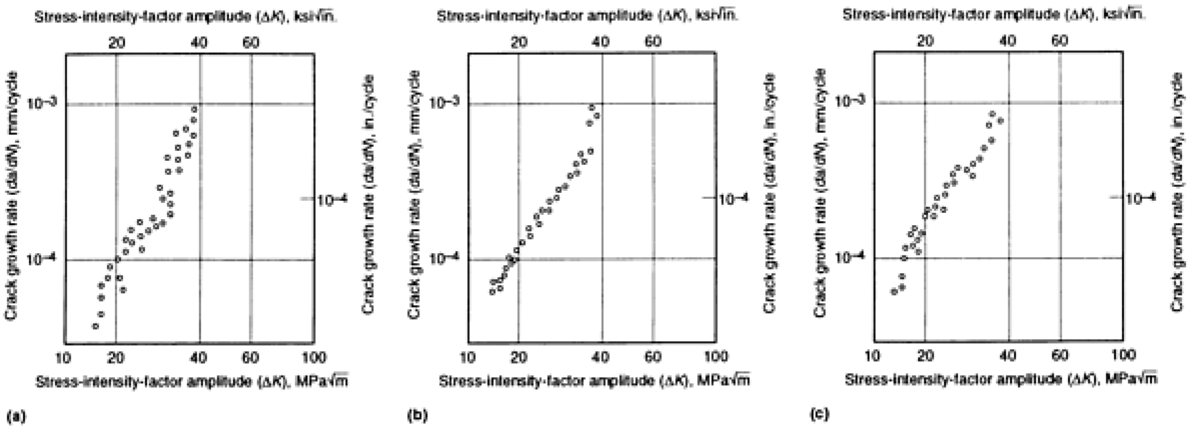

FIG. 22 EFFECT OF WELDING PROCESS ON THE FATIGUE-CRACK-GROWTH RATE SCATTERING IN TI-6AL-4V. (A) GAS-TUNGSTEN ARC WELDING. (B) ELECTRON-BEAM WELDING. (C) LASER-BEAM WELDING. SOURCE: REF 40

References cited in this section

5. W.A. BAESLACK III, D.W. BECKER, AND F.H. FROES, ADVANCES IN TITANIUM ALLOY WELDING METALLURGY, *J. MET.,* MAY 1984, P 46-58
7. D.W. BECKER, W.A. BAESLACK III, AND R.W. MESSLER, JR., TITANIUM WELDING, *TITANIUM '80 SCIENCE AND TECHNOLOGY,* VOL 1, H. KIMURA AND O. IZUMI, ED., THE METALLURGICAL SOCIETY OF AIME, 1980, P 255-275
9. J.R. HARTMAN AND W.A. BAESLACK III, ADVANCES IN ARC WELDING OF TITANIUM ALLOYS, *WELDABILITY OF MATERIALS,* R.A. PATTERSON AND K.W. MAHIN, ED., ASM INTERNATIONAL, 1990, P 319-330

35. W.A. BAESLACK III AND F.D. MULLINS, RELATIONSHIP OF MICROSTRUCTURE, MECHANICAL PROPERTIES AND FRACTOGRAPHIC FEATURES IN WELDED HIGH TOUGHNESS TITANIUM ALLOYS, *TRENDS IN WELDING RESEARCH,* S.A. DAVID, ED., AMERICAN SOCIETY FOR METALS, 1982, P 541
37. W.A. BAESLACK III, D.W. BECKER, AND F.H. FROES, WELDABILITY OF A HIGH TOUGHNESS TITANIUM ALLOY, A METALLURGICAL EVALUATION, *TITANIUM '80 SCIENCE AND TECHNOLOGY,* VOL 2, H. KIMURA AND O. IZUMI, ED., THE METALLURGICAL SOCIETY OF AIME, 1980, P 2369
38. R.E. LEWIS *ET AL.*, THE EFFECT OF BORON ON WELDMENT MICROSTRUCTURES IN TI-6AL-2NB-1TA-1MO ALLOY, *TITANIUM SCIENCE AND TECHNOLOGY,* VOL 2, G. LÜTJERING, U. ZWICKER, AND W. BUNK, ED., DEUTSCHE GESELLSCHAFT FÜR METALLKUNDE E.V., WEST GERMANY, 1985, P 799
39. T.S. PARKER AND P.G. PARTRRIDGE, FATIGUE PROPERTIES OF 4 MM THICK LASER WELDED TI-6AL-4V SHEET, *FATIGUE '84,* C.J. BEEVERS, ED., ENGINEERING MATERIALS ADVISORY SERVICES LTD., LONDON, 1984
40. P.S. PAO, T.C. PENG, AND J.E. O'NEAL, FATIGUE CRACK PROPAGATION THROUGH TITANIUM WELDS, *TITANIUM SCIENCE AND TECHNOLOGY,* VOL 2, G. LÜTJERING, U. ZWICKER, AND W. BUNK, ED., DEUTSCHE GESELLSCHAFT FÜR METALLKUNDE, E.V., WEST GERMANY, 1985, P 887

Selection and Weldability of Conventional Titanium Alloys

W.A. Baeslack III, The Ohio State University; J.R. Davis, Davis and Associates; C.E. Cross, Martin Marietta Astronautics Group

Welding Process Application

Titanium and its alloys can be welded by a wide variety of conventional fusion and solid-state processes, although titanium's chemical reactivity typically requires special procedures and precautions. In the United States, fusion welding of titanium is performed principally by inert gas-shielded arc welding and high-energy-beam welding processes. To date, the commercial joining of titanium alloys using flux-based processes (for example, submerged arc welding, flux-cored arc welding, and electroslag welding) has not been shown to be competitive due to requirements for costly high-purity fluoride-base fluxes. However, recent studies on submerged arc welding of 25 mm (1 in.) thick titanium alloy plate indicate that it is possible to weld heavy sections using a flux composed of 85% calcium fluoride, 10% strontium chloride, and 3% lithium chloride (Ref 41). Tensile and corrosion properties of the weld were equal to that of the base plate material; however, the coarser weld grain structure did not provide adequate toughness.

Additional information on the welding processes described below can be found in the Section "Joining Processes" in this Volume.

Gas-tungsten arc welding (GTAW), in which the heat of welding is provided by an arc maintained between a nonconsumable tungsten electrode and the workpiece, is the most widely used process for joining titanium sheet. In general, gas-tungsten arc welds can be produced manually (in complex joint geometries or for weld repair) and automatically in sheet up to about 3 mm (0.125 in.) in thickness without joint preparation or filler metal additions. For welding of greater thicknesses, suitable joint preparation and the cold or hot-wire addition of a filler metal are required. Although conventional GTAW equipment can be used for welding of titanium alloys, the heated metal in the weld zone must be shielded from the atmosphere to prevent pickup of oxygen and nitrogen, which can be deleterious to weld mechanical properties. Such protection is best provided by welding within a rigid or collapsible plastic chamber that has been evacuated and purged with argon, although local shielding is also suitable for many welding operations. Typical parameters for welding of Ti-6Al-4V by GTAW are shown in Table 5.

TABLE 5 WELDING PARAMETERS FOR GAS-TUNGSTEN ARC WELDING OF TI-6AL-4V

MATERIAL THICKNESS(A)		TUNGSTEN ELECTRODE DIAMETER		FILLER ROD DIAMETER		NOZZLE SIZE ID		SHIELDING GAS FLOW		WELDING CURRENT(B)	NUMBER OF PASSES	TRAVEL SPEED(C)	
mm	**in.**	**mm**	**in.**	**mm**	**in.**	**mm**	**in.**	**m^3/h**	**ft^3/h**			**mm/min**	**in./min**
SQUARE-GROOVE AND FILLET WELDS													
0.61	0.024	1.6	1/16	. . .	. . .	9.5	3/8	0.51	18	20-35	1	150	6
1.60	0.063	1.6	1/16	. . .	. . .	16	5/8	0.51	18	85-140	1	150	6
2.36	0.093	2.4	3/32	1.6	1/16	16	5/8	0.71	25	170-215	1	205	8
3.18	0.125	2.4	3/32	1.6	1/16	16	5/8	0.71	25	190-235	1	205	8
4.78	0.188	2.4	3/32	3.2	1/8	16	5/8	0.71	25	220-280	2	205	8
V-GROOVE AND FILLET WELDS													
6.35	0.25	3.2	1/8	3.2	1/8	16	5/8	0.76	30	275-320	2	205	8
9.52	0.375	3.2	1/8	3.2	1/8	19	3/4	0.99	35	300-350	2	150	6
12.7	0.50	3.2	1/8	4.0	5/32	19	3/4	1.13	40	325-425	3	150	6

Note: Tungsten used for the electrode; first choice 2% thoriated EWTh2, second choice 1% thoriated EWTh1. Use filler metal one or two grades lower in strength than the base metal. Adequate gas shielding is essential not only for the arc, but for heated material also. Backing gas is recommended at all times. A trailing gas shield is also recommended. Argon is preferred. For higher heat input, on thicker material use argon-helium mixture. Without backing or chill bar, decrease current 20%.

(A) OR FILLET SIZE.
(B) DIRECT CURRENT ELECTRODE NEGATIVE.
(C) PER PASS

Gas-metal arc welding (GMAW), in which the heat of welding is provided by an arc maintained between a consumable electrode and the workpiece, is more cost-effective than GTAW for the welding of titanium plate over 13 mm (0.5 in.) in thickness. However, the application of GMAW to titanium has been restricted by poor arc stability, which promotes appreciable spatter during welding, thereby reducing weld quality and overall process efficiency. Typical welding parameters for welding of Ti-6Al-4V by GMAW are given in Table 6.

TABLE 6 TYPICAL PARAMETERS FOR GAS-METAL ARC WELDING OF TITANIUM ALLOY PLATE

THICKNESS		WELDING POSITION(A)	SHIELDING GAS	ELECTRODE					ARC VOLTAGE, V	WELDING CURRENT			TRAVEL SPEED	
				TYPE OF TRANSFER(B)	DIAMETER		FEED							
mm	in.				mm	in.	m/min	in./min		A(C)	BACKGROUND	MEAN PULSE	mm/min.	in./min.
3.18	0.125	F	75AR-25HE	S	0.16	0.062	5.1-5.7	200-225	20	250-260	. . .	. . .	380	15
6.35	0.250	F	75AR-25HE	S	0.16	0.062	7.6-8.1	300-320	30	300-320	. . .	. . .	380	15
9.53	0.375	F	85HE-15AR	S	0.11	0.045	22.0	865	33	350-360	. . .	. . .	640	25
12.7	0.500	F	75AR-25HE	S	0.16	0.062	9.5-10.2	375-400	40	340-360	. . .	. . .	380	15
15.9	0.625	F	75AR-25HE	S	0.16	0.062	10.2-10.8	400-425	45	350-370	. . .	. . .	380	15
25.4	1.00	F	AR	S	0.16	0.062	9.7	380	36-37	320-350	. . .	. . .	510-585	20-23
		F	AR	S	0.16	0.062	14.0	550	31-33	315-340	. . .	. . .	330	13
		F	75AR-25HE	PS	0.16	0.062	4.3	170	22-24	. . .	75	80	100	4
		V-UP	75AR-25HE	PS	0.16	0.062	3.8	150	21-23	. . .	35	100	64	2.5
		F	75AR-25HE	SC(D)	0.08	0.030	16.5	650	17-21	140-185	. . .	. . .	180	7
		V-UP	75AR-25HE	SC(D)	0.08	0.030	14.0	550	15-16	100-135	. . .	. . .	46	1.8

Source: Ref 4

(A) F, FLAT; V-UP, VERTICAL UP.

(B) S, SPRAY (ELECTRODE POSITIVE); PS, PULSED SPRAY (ELECTRODE POSITIVE); SC, SHORT-CIRCUITING (ELECTRODE NEGATIVE).

(C) DIRECT CURRENT.

(D) FOR GROOVE WELDS, THE GROOVE ANGLE AND ROOT OPENING SHOULD BE LARGER THAN THOSE USED WITH SPRAY TRANSFER TO PROVIDE GOOD ACCESSIBILITY TO THE ROOT AND FACES OF THE JOINT.

Plasma arc welding (PAW) is an extension of the GTAW process in which the arc plasma is constricted by a nozzle, thereby increasing its temperature and energy density as compared with the more diffuse GTAW arc. This higher energy density provides greater penetration capabilities, allowing the production of full-penetration, keyhole square-butt welds in thicknesses up to 13 mm (0.5 in.). Typical welding parameters for PAW of Ti-6Al-4V are given in Table 7.

TABLE 7 TYPICAL OPERATING CONDITIONS FOR PLASMA ARC WELDING OF TI-6AL-4V

Backing gas and trailing shield required; keyhole technique used with orifice-to-work distance of 4.8 ($\frac{3}{16}$ in.)

THICKNESS		TRAVEL SPEED		CURRENT (DCEN), A	ARC VOLTAGE, V	GAS	ORIFICE GAS		SHIELDING GAS		JOINT TYPE
mm	in.	mm/min	in./min				m^3/h	ft^3/h	m^3/h	ft^3/h	
3.18	0.125	510	20	185	21	ARGON	0.23	8	1.7	60	SQUARE BUTT
4.74	0.187	330	13	175	25	ARGON	0.51	18	1.7	60	SQUARE BUTT
9.91	0.390	255	10	225	38	75HE-25AR	0.91	32	1.7	60	SQUARE BUTT
12.7	0.500	255	10	270	36	50HE-50AR	0.76	27	1.7	60	SQUARE BUTT
15.2	0.600	180	7	250	39	50HE-50AR	0.85	30	1.7	60	V-GROOVE(A)

(A) 30° INCLUDED ANGLE; 9.5 MM ($\frac{3}{8}$ IN.) ROOT FACE

Electron-beam welding (EBW) involves melting of the base metals to be joined by the impingement of a focused beam of high-energy electrons. This process has been widely used in the aerospace industry for producing high-quality welds in titanium plate ranging from 6 to more than 75 mm (0.25 to more than 3 in.) in thickness. Advantages of EBW welding include high depth-of-penetration, minimum joint preparation, a narrow weld HAZ, low distortion, and excellent weld cleanliness as welding is performed in a high vacuum. Disadvantages of the EBW process include high capital equipment cost and high weld cooling rates, the latter of which promote formation of undesirable martensite in the weld zone. Table 8 lists typical welding parameters used in the EBW of Ti-6Al-4V. Tensile and fracture toughness data for Ti-6Al-4V are shown in Table 9.

TABLE 8 TYPICAL OPERATING PARAMETERS FOR ELECTRON-BEAM WELDING OF TI-6AL-4V

MATERIAL THICKNESS		ACCELERATING VOLTAGE, KV	BEAM CURRENT, MA	TRAVEL RATE	
mm	in.			mm/min	in./min
3.18	0.125	20	95	760	30
5.08	0.20	28	170	2490	98
16.0	0.63	30	260	1525	60
25.4	1.0	36.5	375	1120	44
50.8	2.0	45	450	660	26
88.9	3.5	90.0	500	510	20

TABLE 9 PROPERTIES OF ELECTRON-BEAM-WELDED TI-6AL-4V

CONDITION(A)	0.2% OFFSET YIELD STRENGTH		ULTIMATE TENSILE STRENGTH		ELONGATION, %	REDUCTION IN AREA, %	FRACTURE TOUGHNESS (K_Q)	
	MPA	KSI	MPA	KSI			MPA $\sqrt{m}$	KSI $\sqrt{in}$
WELDED + 700 °C	944	137	1032	150	12	32(B)	63.7	58.0

(1290 °F)/1 H AC								
WELDED + 500 °C (930 °F)/6 H AC	965	140	1055	153	12	37(B)	58.7	53.5
AS-WELDED	935	136	1030	149	12	34(B)	67.2	61.2
BASE METAL	927	134	991	144	12	39	44.5	40.5

Source: Ref 41

(A) AC. AIR COOLED.

(B) FRACTURE IN BASE METAL.

Laser-Beam Welding (LBW). During the past decade, LBW has become increasingly competitive as a joining process for titanium sheet and thin plate. In LBW, melting of the workpieces is produced by the impingement of a high-intensity, coherent beam of light. Because the laser beam can be readily transmitted through air, the LBW process offers significant practical advantages over conventional EBW, which requires welding in a high vacuum. However, the LBW process is more limited than EBW from the standpoint of joining thick titanium plate, with 15 kW required to produce a full-penetration weld in 13 mm (0.5 in.) thick Ti-6Al-4V. As expected, considering similarities in the heat sources, the metallurgical and mechanical properties of laser-beam and electron-beam welds are quite similar. Table 10 presents typical parameters for LBW of Ti-6Al-4V. Weld properties of laser-beam welded Ti-6Al-4V are given in Table 11.

TABLE 10 DIMENSIONS AND PROCESS VARIABLES FOR SUCCESSFUL LASER WELDS IN TI-6AL-4V

POWER, W	WELDING SPEED		THICKNESS		MELT WIDTH AT SURFACE		WIDTH OF HAZ AT SURFACE		MELT WIDTH AT BOTTOM		WIDTH OF HAZ AT BOTTOM		NUGGET AREA		HAZ AREA		MELTING EFFICIENCY	JOINING, AREA MM^2/KJ
	mm/s	in./s	mm	in.	mm	in.	mm	in.	mm	in.	mm	in.	mm^2	in.^{2} × 10^{-3}	mm^3	in.^{2} × 10^{-3}		
1000	23	0.91	1.00	0.039	1.85	0.073	3.75	0.148	0.90	0.035	2.50	0.098	1.38	2.14	1.75	2.71	21.37	23
1000	34	1.34	1.00	0.039	1.65	0.065	3.30	0.130	0.80	0.031	2.50	0.098	1.22	1.89	1.67	2.59	28.14	34
1000	15	0.59	1.00	0.039	2.15	0.085	4.70	0.185	1.60	0.063	3.40	0.134	1.88	2.91	2.17	3.36	19.00	15
1000	11	0.43	1.00	0.039	2.65	0.104	6.40	0.252	2.00	0.078	4.60	0.181	2.33	3.61	3.17	4.91	17.28	11
1480	60	2.36	0.89	0.035	1.00	0.039	2.20	0.087	0.60	0.024	1.50	0.059	0.71	1.10	0.93	1.44	19.50	36
1480	50	1.97	0.89	0.035	1.35	0.053	3.00	0.118	0.90	0.035	2.00	0.079	1.00	1.55	1.22	1.89	22.86	30
1480	23	0.91	0.89	0.035	1.60	0.063	3.10	0.122	1.35	0.053	2.75	0.108	1.31	2.03	1.29	2.00	13.78	13.83
1500	60	2.36	0.81	0.032	1.00	0.039	2.00	0.079	0.90	0.035	1.50	0.059	0.77	1.19	0.65	1.01	20.80	32.4
1500	50	1.96	0.81	0.032	1.20	0.047	2.20	0.087	0.90	0.035	1.75	0.069	0.85	1.32	0.75	1.16	19.16	27
1500	23	0.91	0.81	0.032	1.35	0.053	5.60	0.220	1.25	0.049	4.00	0.157	1.05	1.63	2.82	4.37	10.91	12.42
1500	7	0.28	2.03	0.080	4.30	0.169	10.00	0.394	3.00	0.118	5.50	0.217	7.41	11.48	8.32	12.90	25.03	10.15
1500	9	0.35	2.03	0.080	3.85	0.152	8.35	0.329	2.15	0.085	4.00	0.157	6.09	9.44	6.45	10.00	26.06	12.86
1500	13	0.51	2.03	0.080	3.25	0.128	6.25	0.246	2.00	0.079	4.00	0.157	5.33	8.26	5.08	7.87	31.20	17.6
1110	6	0.24	2.03	0.080	3.75	0.148	8.25	0.324	0.75	0.030	8.25	0.324	4.57	7.08	12.18	18.88	16.68	10.97
1500	23	0.91	1.00	0.039	1.75	0.069	3.00	0. 118	1.45	0.057	2.75	0.108	1.60	2.48	1.27	1.97	16.58	15.33
1540	50	1.96	1.00	0.039	1.45	0.057	2.50	0.098	0.95	0.037	2.45	0.096	1.20	1.86	1.27	1.97	26.32	32.47
1540	60	2.36	1.00	0.039	1.15	0.045	2.25	0.089	0.85	0.033	2.00	0.079	1.00	1.55	1.12	1.74	26.32	38.96

Source: Ref 42

TABLE 11 TYPICAL MECHANICAL PROPERTIES OF LASER-BEAM-WELDED TITANIUM ALLOYS

MATERIAL	ULTIMATE TENSILE STRENGTH		0.2% PROOF STRESS		ELONGATION, %
	MPA	KSI	MPA	KSI	
TI-6AL-4V					
AS-WELDED	~860-923	~125-134	~800-860	~116-125	11-14

BASE MATERIAL	895-1004	130-146	834-895	121-130	10-15
COMMERCIALLY PURE TITANIUM					
AS-WELDED	~530-573	~77-83	460-503	~67-73	~27
BASE MATERIAL	>494	>72	>416	>60	27-28

Ref 42

Resistance welding is a process in which heat is produced by resistance to the flow of high electrical currents (5 to 10 kA) across the interface between two contacting surfaces. At sufficiently high current levels, melting is induced at the interface and a weld fusion zone (i.e., weld nugget) is created. At lower current levels, welding of the faying surfaces occurs entirely by a solid-state deformation or diffusion welding mechanism. Resistance welding is most commonly used to join sheet and can produce either discrete, individual spot welds or continuous, overlapping seam welds. The resistance welding parameters employed for titanium are generally comparable to those used for austenitic stainless steels, due to similarities in thermal conductivity between the two materials. As expected, the production of high-quality resistance welds requires careful cleaning of the sheet surfaces prior to welding. However, the rapid thermal cycles employed and the self-contained nature of the weld zone generally preclude the necessity for inert-gas shielding during welding. Table 12 presents typical parameters for resistance spot welding of Ti-6Al-4V.

TABLE 12 TYPICAL OPERATING CONDITIONS FOR RESISTANCE SPOT WELDING OF TI-6AL-4V

MATERIAL THICKNESS		**WELDING CURRENT, KA**	**ELECTRODE FORCE$^{(A)}$**		**WELD TIME, CYCLES (60 HZ)**	**NUGGET DIAMETER**	
mm	**in.**		**kn**	**lbf**		**mm**	**in.**
0.89	0.035	5.5	2.7	600	7	6.35	0.25
1.57	0.062	10.6	6.7	1500	10	9.15	0.36
2.36	0.093	12.5	10.7	2400	16	10.9	0.43

Source: Ref 4

(A) RWMA CLASS 2 COPPER ALLOY ELECTRODES, 75 MM (3 IN.) FACE RADIUS.

Diffusion welding is a solid-state joining process that is finding increased application for the fabrication of complex titanium alloy components. Diffusion welding (or bonding) of titanium involves four basic steps: (1) development of intimate physical contact between the faying surfaces through deformation of asperity peaks (surface roughness) by yielding and creep at low pressures and high temperatures, (2) formation of a metallic bond, (3) diffusion across the faying surface, and (4) grain growth across the original weld interface. Prerequisites for accomplishing these steps include a clean and smooth surface combined with a low applied pressure to promote steps 1 and 2, and moderate-to-high temperatures to promote steps 3 and 4. Titanium alloys are ideal for diffusion welding because titanium exhibits a high solubility for oxygen, which promotes dissolution of surface oxides at elevated temperatures. In addition, the yield and creep strengths of titanium are low at welding temperatures, thereby enhancing stages 1 and 2 of the diffusion welding process. Diffusion welding of Ti-6Al-4V is generally performed at temperatures ranging from about 900 to 950 °C (1650 to 1740 °F), at pressures ranging from 1.3 to 13.8 MPa (200 to 2000 psi), and for times ranging from about 1 to 6 h. In combination with superplastic forming, diffusion welding has been utilized to produce a wide variety of complex Ti-6Al-4V shapes, such as honeycomb panels for aerospace applications.

Other Weld Processes. A wide variety of more specialized welding processes have been used to join titanium, including flash-butt welding, continuous-drive and inertia friction welding, ultrasonic welding, and explosive welding. In general, these processes have been effective, producing defect-free joints with properties that depend principally on the cooling rates from high temperatures and the characteristics of the resulting transformed-beta microstructure.

Special Welding Precautions and Requirements. Production of high-quality welds in titanium and titanium alloys requires meticulous preweld cleaning of the workpiece and consumables and adequate gas shielding to prevent atmospheric contamination. Mill-annealed titanium sheet and plate are generally supplied in a condition that requires only the removal of surface grease, oil, and dirt. Such cleaning is normally accomplished by wiping with or dipping in a nonchlorinated solvent such as acetone or methyl-ethyl-ketone (MEK). Materials that exhibit a light oxide scale as a result of heat treatment below about 600 °C (1110 °F) or that contain entrapped oil from machining operations should be pickled for 5 to 10 min in a solution of 30 to 40% nitric acid and 4 to 5% hydrofluoric acid in water at a temperature between 20 and 70 °C (68 and 160 °F). Heavier oxide scales resulting from high-temperature heat treatment or thermomechanical processing may require cleaning in a molten salt bath or mechanical removal, followed by pickling. Pickled pieces should be rinsed in water, dried, and either directly welded or wrapped in clean plastic. All handling should be performed with clean white gloves. Weld contamination can also result from filler metal that has become dirty due to inadequate storage or careless handling prior to or during the welding operation. Degreasing the filler rod or wire is generally sufficient to eliminate contamination problems originating from these sources.

Effective shielding of the weld zone from the atmosphere during welding is extremely important in ensuring maximum weld ductility and toughness and in reducing the potential for solid-state weld cracking. Optimum shielding conditions are provided by welding within a permanent or collapsible clear plastic enclosure that has been evacuated and purged with argon (dew point, -24 °C, or -11 °F). Alternatively, localized trailing and backing shields can be used, particularly for automatic welding of simple joint geometries (for example, automatic butt welding of sheet). Use of localized shielding for manual welding of complex parts, although feasible, requires specially designed shielding fixtures and generally is not recommended. Fortunately, evidence of titanium weld contamination is readily apparent by discoloration of the weld surface, with a change from bright silver to straw, to magenta, to blue, with a gradual increase in the level of contamination. Severe contamination is associated with a white or gray powdery-appearing weld surface, and is often accompanied by solid-state cracking in the weld metal. In many high-performance applications, a bright silver color is required, although a light straw color may be accepted in less-critical applications.

Because titanium alloy welds are commonly used in fatigue-critical applications, a stress relief operation is generally required following welding. Specific stress-relief temperatures and times depend on the base-metal microstructure (Table 3). In order to prevent property degradation, stress relieving should be performed on components free of surface contamination and in an inert or vacuum environment.

Weld Design Criteria and Limitations. Designers of titanium welded structures must consider both welding process applicability and the physical characteristics and mechanical properties of the welds. From a welding process stand-point, an efficient design will utilize a welding process optimally suited to a particular material thickness and joint configuration. Such process suitability is also based on component size and shape (for example, whether the component will fit in an available EBW chamber, or whether the entire part can be suitably protected from the atmosphere during GTAW or PAW), the cost of producing the weld (including both capital equipment and operating costs), and postweld processing requirements.

The design of titanium welded structures is also influenced by the characteristics and mechanical properties of a joint produced by a particular welding process. Physical characteristics of the weld, including undercut or underfill in fusion welds, upset in solid-state welds, and weld distortion are important considerations, because these characteristics affect not only the physical dimensions of the component but also joint mechanical properties. Obviously, the mechanical properties of the joint, as influenced by the integrity and metallurgical structure of the weld zone, are the principal considerations in weld design. As indicated above, low fusion zone ductility and toughness in rapidly cooled welds, and poor axial fatigue behavior in welds containing defects, are important characteristics of welds that may influence the design of the welded component.

Welding Specifications. The American Welding Society has published two specifications related to welding of titanium. AWS Specification 5.16, "Specification for Titanium and Titanium Alloy Bare Welding Rods and Electrodes," provides requirements for titanium and titanium-alloy welding rods for use with GTAW and GMAW processes (see the article "Welding of Titanium Alloys" in this Volume for filler metal compositions). Although the basic specification principally covers filler metal chemical composition requirements, a rather lengthy appendix presents additional general information on welding of titanium, including toughness data for alpha-beta titanium alloy welds. AWS Specification D10.6, "Recommended Practices for Gas Tungsten-Arc Welding of Titanium Pipe and Tubing," is specific to single-phase alpha alloys.

Numerous additional specifications developed by the government and by industry discuss the procedural aspects of the welding of titanium and set forth minimum weld quality and mechanical property requirements. These specifications may be quite general or very specific, describing in detail the fabrication or weld repair of an individual component.

References cited in this section

4. TITANIUM, ZIRCONIUM, HAFNIUM, TANTALUM, AND COLUMBIUM, *WELDING HANDBOOK,* 7TH ED., VOL 4, AMERICAN WELDING SOCIETY, 1982, P 434-470
41. R.F. VAUGHAN, PROPERTIES OF WELDED TITANIUM ALLOYS AND THEIR APPLICATION IN THE AEROSPACE INDUSTRY, *TITANIUM '80 SCIENCE AND TECHNOLOGY,* VOL 4, H. KIMURA AND O. IZUMI, ED., THE METALLURGICAL SOCIETY OF AIME, 1983, P 2423
42. J. MAZUMDER AND W.M. STEEN, MICROSTRUCTURE AND MECHANICAL PROPERTIES OF LASER WELDED TI-6AL-4V, *METALL. TRANS. A,* VOL 13, 1982, P 865-871

Selection and Weldability of Conventional Titanium Alloys

W.A. Baeslack III, The Ohio State University; J.R. Davis, Davis and Associates; C.E. Cross, Martin Marietta Astronautics Group

References

1. M.J. DONACHIE, JR., TITANIUM, *METALS HANDBOOK DESK EDITION,* AMERICAN SOCIETY FOR METALS, 1985, P 9-1 TO 9-3
2. S. LAMPMAN, WROUGHT TITANIUM AND TITANIUM ALLOYS, *PROPERTIES AND SELECTION: NONFERROUS ALLOYS AND SPECIAL-PURPOSE MATERIALS,* VOL 2, *ASM HANDBOOK,* ASM INTERNATIONAL, 1990, P 592-633
3. E.W. COLLINGS, *THE PHYSICAL METALLURGY OF TITANIUM ALLOYS,* AMERICAN SOCIETY FOR METALS, 1984
4. TITANIUM, ZIRCONIUM, HAFNIUM, TANTALUM, AND COLUMBIUM, *WELDING HANDBOOK,* 7TH ED., VOL 4, AMERICAN WELDING SOCIETY, 1982, P 434-470
5. W.A. BAESLACK III, D.W. BECKER, AND F.H. FROES, ADVANCES IN TITANIUM ALLOY WELDING METALLURGY, *J. MET.,* MAY 1984, P 46-58
6. R.W. SCHUTZ AND D.E. THOMAS, CORROSION OF TITANIUM AND TITANIUM ALLOYS, *CORROSION,* VOL 13, *ASM HANDBOOK,* ASM INTERNATIONAL, 1987, P 669-706
7. D.W. BECKER, W.A. BAESLACK III, AND R.W. MESSLER, JR., TITANIUM WELDING, *TITANIUM '80 SCIENCE AND TECHNOLOGY,* VOL 1, H. KIMURA AND O. IZUMI, ED., THE METALLURGICAL SOCIETY OF AIME, 1980, P 255-275
8. D.F. HASSON, "A REVIEW OF TITANIUM WELDING PROCESSES," U.S. NAVY REPORT, ANNAPOLIS, MD, 1985
9. J.R. HARTMAN AND W.A. BAESLACK III, ADVANCES IN ARC WELDING OF TITANIUM ALLOYS, *WELDABILITY OF MATERIALS,* R.A. PATTERSON AND K.W. MAHIN, ED., ASM INTERNATIONAL, 1990, P 319-330
10. W.A. BAESLACK III AND D.L. HALLUM, NATURE OF GRAIN REFINEMENT IN TITANIUM ALLOY WELDS BY MICROCOOLER INOCULATION, *WELD. J. RES. SUPPL.,* VOL 69 (NO. 12), DEC 1990, P 326S-336S
11. D.W. BECKER AND W.A. BAESLACK III, PROPERTY-MICROSTRUCTURE RELATIONSHIPS IN METASTABLE-BETA TITANIUM ALLOY WELDMENTS, *WELD. J. RES. SUPPL.,* VOL 59, 1980, P 85S-92S
12. W.A. BAESLACK III, P.S. LIU, D.P. BARBIS, J.R. SCHLEY, AND J.R. WOOD, POSTWELD HEAT TREATMENT OF GTA WELDS IN A HIGH-STRENGTH, METASTABLE-BETA TITANIUM ALLOY--BETA-C, *PROC. OF 1992 INT. CONF. ON TITANIUM ALLOYS* (SAN DIEGO), 1992
13. A.T. D'ANNESSA, REDISTRIBUTION OF SOLUTE IN FUSION WELDING, *WELD. J.,* VOL 45, 1966, P 569S-576S
14. W.A. BAESLACK III, EFFECT OF SOLUTE BANDING ON TRANSFORMATIONS IN DISSIMILAR

TITANIUM ALLOY WELDMENTS, *J. MAT. SCI. LET.,* VOL 1, 1982, P 229-231

15. P.S. LIU, W.A. BAESLACK III, AND J. HURLEY, DISSIMILAR ALLOY LASER WELDING OF TITANIUM: TI-6AL-4V TO BETA-C, *WELD. J. RES. SUPPL.,* IN PRESS
16. P.S. LIU, W.A. BAESLACK III, AND J. HURLEY, LASER WELDING OF TI-6AL-4V TO BETA-21S, *PROC. OF TMS CONF. ON BETA TITANIUM ALLOYS* (DENVER, CO), TMS, 1993
17. W.A. BAESLACK III, S. CHIANG, AND C.A. ALBRIGHT, *J. MAT. SCI. LET,* VOL 9, 1990, P 698-702
18. W.A. BAESLACK ILL, *METALLOGRAPHY,* VOL 13, 1980, P 277-281
19. N.N. PROKHOROV, *WELD. PROD.,* VOL 9, 1986, P 30-32
20. P.W. HOLSBERG, "CRACKING AND POROSITY IN TITANIUM ALLOY WELDMENTS," U.S. NAVY DTNSRDC REPORT, 1979
21. F.D. ROSI, D.A. DUBE, AND B.H. ALEXANDER, *TRANS. AIME,* VOL 197, 1953, P 257
22. N.E. PATON AND J.C. WILLIAMS, THE EFFECT OF HYDROGEN ON TITANIUM AND ITS ALLOYS, *HYDROGEN IN METALS,* AMERICAN SOCIETY FOR METALS, 1974, P 409-431
23. R.E. LEWIS, I.L. CAPLAN, AND W.C. COONS, THE ELEVATED TEMPERATURE DUCTILITY DIP PHENOMENON IN ALPHA, NEAR-ALPHA AND ALPHA-BETA TITANIUM ALLOYS, *TITANIUM SCIENCE AND TECHNOLOGY,* VOL 2, G. LÜTJERING, U. ZWICKER, AND W. BUNK, ED., DEUTSCHE GESELLSCHAFT FÜR METALLKUNDE E.V., WEST GERMANY, 1985
24. B.K. DAMKROGER, G.R. EDWARDS, AND B.B. RATH, A MODEL FOR HIGH TEMPERATURE DUCTILITY LOSS IN ALPHA-BETA TITANIUM ALLOYS, *METALL. TRANS. A,* VOL 18, 1987, P 483-485
25. D.M. BOWDEN AND E.A. STARKE, JR., HOT CRACKING SUSCEPTIBILITY OF TITANIUM 6211, *TITANIUM SCIENCE AND TECHNOLOGY,* VOL 2, G. LÜTJERING, U. ZWICKER, AND W. BUNK, ED., DEUTSCHE GESELLSCHAFT FÜR METALLKUNDE E.V., WEST GERMANY, 1985, P 853
26. D.M. BOWDEN AND E.A. STARKE, JR., *METALL. TRANS. A,* VOL 15, 1984, P 1687-1698
27. D. HAYDUK, B.K. DAMKROGER, G.R. EDWARDS, AND D.L. OLSON, *WELD. J.,* VOL 65, 1986, P 251S-260S
28. B.K. DAMKROGER, G.R. EDWARDS, AND B.B. RATH, *WELD. J.,* VOL 68, 1989, P 290S-302S
29. R.E. TREVISAN, D.D. SCHWEMMER, AND D.L. OLSON, THE FUNDAMENTALS OF WELD PORE FORMATION, *WELDING: THEORY AND PRACTICE,* ELSEVIER, 1990, P 79-115
30. K. BORGGREEN AND I. WILSON, USE OF POSTWELD HEAT TREATMENTS TO IMPROVE DUCTILITY IN THIN SHEETS OF TI-6AL-4V, *WELD. J.,* VOL 59, 1980, P 1S-8S
31. S.L. BOSTON AND W.A. BAESLACK III, "MICROSTRUCTURE/MECHANICAL PROPERTY RELATIONSHIPS IN GTA WELDED TI-10V-2FE-3AL," AFML-TM-LL-80-1, WRIGHT PATTERSON AIR FORCE BASE, DAYTON, OH, 1980
32. M.A. GREENFIELD AND C.M. PIERCE, POSTWELD AGING OF A METASTABLE BETA TITANIUM ALLOY, *WELD. J.,* VOL 52, 1973, P 524S-527S
33. W.A. BAESLACK III, EVALUATION OF TRIPLEX HEAT TREATMENTS FOR TITANIUM ALLOYS, *WELD. J.,* VOL 61, 1982, P 193S-199S
34. D.W. BECKER AND W.A. BAESLACK III, PROPERTY/MICROSTRUCTURE RELATIONSHIPS IN METASTABLE-BETA TITANIUM ALLOY WELDMENTS, *WELD. J.,* VOL 59, 1980, P 850-920
35. W.A. BAESLACK III AND F.D. MULLINS, RELATIONSHIP OF MICROSTRUCTURE, MECHANICAL PROPERTIES AND FRACTOGRAPHIC FEATURES IN WELDED HIGH TOUGHNESS TITANIUM ALLOYS, *TRENDS IN WELDING RESEARCH,* S.A. DAVID, ED., AMERICAN SOCIETY FOR METALS, 1982, P 541
36. V.N. ZAMKOV *ET AL.,* WAYS OF PROVIDING THE HIGH SERVICEABILITY OF HEAT-HARDENED WELDED JOINTS OF ALPHA + BETA TITANIUM ALLOYS, *TITANIUM SCIENCE AND TECHNOLOGY,* VOL 2, G. LÜTJERING, U. ZWICKER, AND W. BUNK, ED., DEUTSCHE GESELLSCHAFT FÜR METALLKUNDE, E.V., WEST GERMANY, 1985, P 877
37. W.A. BAESLACK III, D.W. BECKER, AND F.H. FROES, WELDABILITY OF A HIGH TOUGHNESS

TITANIUM ALLOY, A METALLURGICAL EVALUATION, *TITANIUM '80 SCIENCE AND TECHNOLOGY,* VOL 2, H. KIMURA AND O. IZUMI, ED., THE METALLURGICAL SOCIETY OF AIME, 1980, P 2369

38. R.E. LEWIS *ET AL.*, THE EFFECT OF BORON ON WELDMENT MICROSTRUCTURES IN TI-6AL-2NB-1TA-1MO ALLOY, *TITANIUM SCIENCE AND TECHNOLOGY,* VOL 2, G. LÜTJERING, U. ZWICKER, AND W. BUNK, ED., DEUTSCHE GESELLSCHAFT FÜR METALLKUNDE E.V., WEST GERMANY, 1985, P 799
39. T.S. PARKER AND P.G. PARTRRIDGE, FATIGUE PROPERTIES OF 4 MM THICK LASER WELDED TI-6AL-4V SHEET, *FATIGUE '84,* C.J. BEEVERS, ED., ENGINEERING MATERIALS ADVISORY SERVICES LTD., LONDON, 1984
40. P.S. PAO, T.C. PENG, AND J.E. O'NEAL, FATIGUE CRACK PROPAGATION THROUGH TITANIUM WELDS, *TITANIUM SCIENCE AND TECHNOLOGY,* VOL 2, G. LÜTJERING, U. ZWICKER, AND W. BUNK, ED., DEUTSCHE GESELLSCHAFT FÜR METALLKUNDE, E.V., WEST GERMANY, 1985, P 887
41. R.F. VAUGHAN, PROPERTIES OF WELDED TITANIUM ALLOYS AND THEIR APPLICATION IN THE AEROSPACE INDUSTRY, *TITANIUM '80 SCIENCE AND TECHNOLOGY,* VOL 4, H. KIMURA AND O. IZUMI, ED., THE METALLURGICAL SOCIETY OF AIME, 1983, P 2423
42. J. MAZUMDER AND W.M. STEEN, MICROSTRUCTURE AND MECHANICAL PROPERTIES OF LASER WELDED TI-6AL-4V, *METALL. TRANS. A,* VOL 13, 1982, P 865-871

Selection and Weldability of Advanced Titanium-Base Alloys

David H. Phillips, Edison Welding Institute

Introduction

ADVANCED TITANIUM-BASE ALLOYS can generally be categorized as having conventional (terminal phase), intermetallic, or titanium-matrix composite structures. Conventional alloys can be further subdivided, on the basis of the amount of beta phase that is present, into four families: near-alpha, alpha-beta, near-beta, or metastable-beta (Table 1). Intermetallics that are based on titanium most commonly include alpha-2, gamma, and orthorhombic systems. Although the phrase titanium-matrix composite covers a potentially wide range of matrix materials and either continuous or discontinuous reinforcements, the most common composite used is continuous silicon-carbide-reinforced Ti-6Al-4V. This article will describe the physical metallurgy of each alloy category and its families, followed by a brief discussion on the weldability of each.

TABLE 1 ADVANCED TITANIUM ALLOY FAMILIES

CONVENTIONAL ALLOYS
NEAR-ALPHA: TI-1100; IMI-834 ALPHA-BETA: CORONA 5 NEAR-BETA: TI-10V-2FE-3AL METASTABLE-BETA: BETA 215; BETA C
INTERMETALLIC ALLOYS
ALPHA-2: TI24AL-11NB; TI25AL-10NB-3V-1MO GAMMA: TI-AL-NB-CR ORTHORHOMBIC: TI-AL-NB
TITANIUM-MATRIX COMPOSITES
TI-6AL-4V WITH SIC CONTINUOUS FIBERS

Selection and Weldability of Advanced Titanium-Base Alloys

David H. Phillips, Edison Welding Institute

Physical Metallurgy

Conventional alloys can usually be described by one of the four families shown in Table 1. The compositions and maximum use temperatures of selected alloys from these families are shown in Table 2, whereas basic mechanical properties are shown in Table 3.

TABLE 2 MAXIMUM USEFUL TEMPERATURES AND CHEMICAL COMPOSITIONS OF SELECTED HIGH-TEMPERATURE TITANIUM ALLOYS

ALLOY DESIGNATION	MAXIMUM USEFUL TEMPERATURE		COMPOSITION, WT%								
	°C	°F	Al	Sn	Zr	Mo	Nb	V	Si	Cr	C
TI-1100	595	1100	6	2.75	4	0.4	. . .	. . .	0.45	. . .	. . .
IMI-834	595	1100	5.5	4	4	0.3	1	. . .	0.5	. . .	0.06
CORONA 5	. . .	. . .	4.5	. . .	. . .	5	. . .	. . .	. . .	1.5	. . .
TI-10V-2FE-3AL	315	600	3	. . .	. . .	. . .	. . .	10	. . .	. . .	. . .
BETA 21S	. . .	. . .	3	. . .	. . .	15	2.7	. . .	0.2	. . .	. . .
BETA C	. . .	. . .	3	. . .	4	4	. . .	8	. . .	6	. . .

TABLE 3 BASIC MECHANICAL PROPERTIES OF ADVANCED TITANIUM-BASE ALLOYS

COMMON NAME	ULTIMATE TENSILE STRENGTH		YIELD STRENGTH		REDUCTION IN AREA, %	ELONGATION, %	FRACTURE TOUGHNESS	
	MPa	Ksi	MPa	Ksi			MPa $\sqrt{m}$	Ksi $\sqrt{in}$
TI-1100	1010	145	910	130	21	. . .	67	61
IMI 834	1050	150	950	140	24	. . .	44	40
CORONA 5	1200	175	1170	170	22	. . .	. . .	. . .
TI-10V-2FE-3AL	1000	145	1000	140	50	. . .	55	50
BETA 21S	1100	160	1000	140	. . .	9	. . .	. . .
BETA C	1300	190	1200	175	. . .	9	88	80
ALPHA-2	690	100	590	85	4	. . .	22	20
GAMMA	480	70	340	50	2	. . .	16	15
ORTHORHOMBIC	1100	160	1030	150	5	. . .	22	20

The physical metallurgies of selected alloys from each family are described below.

Near-Alpha Alloys. The alloy Ti-1100 was developed in response to a requirement for higher creep properties in Ti-6242S. It was determined that, by adjusting the tin and molybdenum compositions and adding silicon, significant improvements to the creep properties could be realized (Ref 1). Because it was developed for a maximum use temperature of 595 °C (1100 °F), it was named Ti-1100. The processing of this alloy usually involves a final beta forge, followed by an 8-h stabilization at 600 °C (1110 °F). This results in a fully transformed acicular structure, which is the proper microstructure for most applications.

The near-alpha alloy IMI-834 was developed for applications requiring a better balance of creep and fatigue properties. Because fully transformed microstructures are not ideal, in terms of fatigue, IMI-834 was developed to facilitate processing into a duplex microstructure. This was achieved by adding carbon, which stabilizes the alpha phase, thus increasing the size of the alpha-beta phase field. As a result, although careful control is required, a duplex microstructure can be achieved through subsequent alpha-beta processing (Ref 2). The microstructure typically contains 15% primary alpha. Niobium was added to increase oxidation resistance during high-temperature exposures, and silicon was added to improve strength and creep resistance.

Alpha-Beta Alloys. The need for titanium alloys with higher levels of toughness led to the development of Corona 5. This alpha-beta titanium alloy has intermediate strength and exceptionally high fracture toughness. In order to achieve the latter property, this alloy is typically processed high in the alpha-beta phase field, which results in an alpha and beta microstructure containing moderately high aspect ratio primary alpha particles. One of the principal attractions of this alloy is that its strength level can easily be adjusted by modifying the dispersion of the fine alpha (transformed beta) phase (Ref 3).

Near-Beta Alloys. Ti-10V-2Fe-3Al, which is considered to be a relatively high-strength near-beta alloy, can be processed for either high strength and medium toughness or moderate strength and high toughness. It is typically solution treated and aged, which results in a mostly transformed alpha and beta microstructure.

Metastable-Beta Alloys. The Beta 21S alloy was developed primarily as a material that had acceptable high-temperature properties and oxidation resistance, could be readily cold rolled, and was age hardenable. These features (primarily the cold rollability) make it an excellent candidate as a matrix material for titanium-matrix composites. Base-metal processing usually involves the hot and cold working of the ingot to form billet or strip product, followed by solution annealing, cold forming to near-net shape, resolution annealing, and then aging. The fine alpha platelets that precipitate during the aging process provide increased strength in the alloy. The resultant microstructure contains the fine alpha platelets in a beta-phase matrix.

Beta C is a metastable-beta alloy that can be processed and heat treated to provide an excellent combination of yield and tensile strengths, ductility, fracture toughness, and fatigue strength (Ref 4). The alloy contains the beta-isomorphous elements molybdenum and vanadium, as well as the sluggish beta-eutectoid element chromium at levels that depress the beta transus temperature to approximately 750 °C (1380 °F) and promote retention of the high-temperature beta phase down to room temperature upon air cooling. A solution heat treatment, followed by aging, promotes the precipitation of a fine distribution of alpha phase in the beta-phase matrix.

Intermetallics. Titanium intermetallics consist primarily of the titanium aluminides alpha-2 and gamma, as well as the orthorhombic plus beta types of alloys. The aluminides, as shown on the titanium-aluminum phase diagram of Fig. 1, consist of two alloy families: alpha-2, which is based on the Ti_3Al intermetallic, and gamma, which is based on the titanium-aluminum intermetallic. These alloys offer greater stiffness and improved specific creep resistance, and they are capable of extending the use temperature range of conventional titanium alloys. The orthorhombic alloys contain large amounts of niobium, which promotes formation of the orthorhombic phase, Ti_2AlNb.

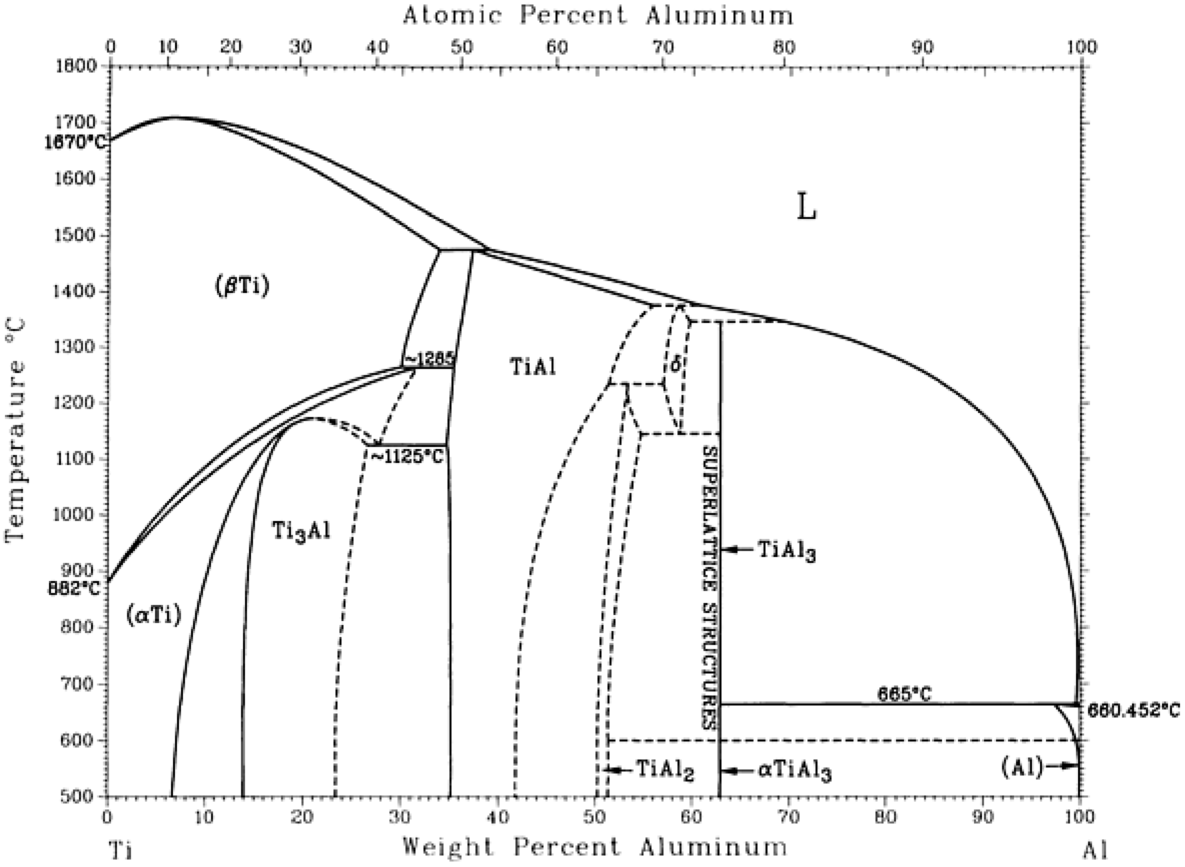

FIG. 1 TITANIUM-ALUMINUM PHASE DIAGRAM. SOURCE: REF 5

Alpha-2 alloys represent a growing family of intermetallics that contain from 14 to 25 at.% aluminum, which places them in the Ti_3Al phase field of Fig. 1. They usually contain beta-stabilizing elements, such as niobium and molybdenum, to promote beta-phase retention and ductility. These alloys are typically alpha-beta processed, resulting in a microstructure similar to that of conventional alloys, which consists of islands of equiaxed primary alpha-2 in a matrix of transformed beta. When compared with conventional titanium alloys, alpha-2 alloys offer higher temperature capabilities (from 100 to 150 °C, or 180 to 270 °F, higher), but have reduced room-temperature toughness and ductility properties (3 to 6% elongation).

The gamma alloy family, which is growing even more rapidly than the alpha-2 alloy family, typically contains from 45 to 52 at.% aluminum, which places these alloys primarily in the gamma phase field, with some overlap in the gamma/alpha-2 phase field. This results in a range of alloy microstructures, including single-phase gamma (TiAl) and duplex structures containing varying amounts of alpha-2 phase (Ti_3Al). Duplex microstructures are generally preferred, because of their higher ductility. Alloying elements such as niobium, chromium, vanadium, tantalum, manganese, and molybdenum are added in small amounts (generally from 2 to 4 at.%) to improve properties such as room-temperature ductility and oxidation resistance. The gamma alloys are more difficult to thermomechanically process than are the alpha-2 alloys, primarily because of their very high strengths at elevated temperatures. Accordingly, compared to alpha-2 alloys, the gamma alloys offer temperature capabilities that are from 100 to 150 °C (180 to 270 °F) higher. They also are known to suffer from low fracture toughness and ductility, although aggressive alloy development programs are currently attempting to improve these properties.

Orthorhombic alloys represent a new class of alloys based on the orthorhombic and beta phases. These alloys offer temperature capabilities similar to those of the alpha-2 alloys, but with higher ductilities and toughness at low temperatures. Although these alloys are similar to alpha-2 alloys, a higher niobium content results in an orthorhombic extension of the hexagonal lattice. The large amounts of niobium also result in a highly stabilized beta phase, compared with alpha-2 alloys. Processing is similar to that used with other titanium alloys. Forging is typically conducted below the beta transus temperature, resulting in a two-phase orthorhombic plus transformed-beta microstructure. A direct age then can be subsequently applied to develop properties.

Titanium-Matrix Composites. Although many titanium matrix/fiber combinations are currently being evaluated by aerospace companies and titanium producers, the most popular composite consists of a conventional Ti-6Al-4V or Ti-6242 matrix with layers of continuous silicon-carbide fibers. The layers can be either unidirectional or cross-ply. Two manufacturing techniques can be used: a foil/fiber/foil approach, in which layers of monolithic foil and fiber sets are either vacuum hot pressed or hot isostatically pressed to form the composite, or a plasma-spray approach, in which the composite is built up by sequentially winding individual layers of fibers and then spraying the monolithic layers, followed by consolidation using either vacuum hot pressing or hot isostatic pressing.

References cited in this section

1. P. BANIA, NEXT GENERATION TITANIUM ALLOYS FOR ELEVATED TEMPERATURE SERVICE, *ISIJ INT.*, VOL 31, 1991, P 840-847
2. OPTIMIZATION OF CREEP AND FATIGUE RESISTANCE IN HIGH TEMPERATURE TI ALLOYS IMI 829 AND IMI 834, *PROC. 5TH INT. CONFERENCE ON TITANIUM* (MUNICH, GERMANY), 1984, P 2419
3. D.W. BECKER, W.A. BAESLACK, AND F.H. FROES, POWDER METALLURGY OF TITANIUM ALLOYS, *PROC. CONF. AIME*, AIME, 1980
4. W.A. BAESLACK, P.S. LIU, D.B. BARBIS, J.R. SCHLEY, AND J.R. WOOD, POSTWELD HEAT TREATMENT OF GTA WELDS IN A HIGH-STRENGTH METASTABLE-BETA TITANIUM ALLOY--BETA-C, INTERNATIONAL TITANIUM CONFERENCE (SAN DIEGO), 1992
5. T.B. MASSALSKI, ED., *BINARY ALLOY PHASE DIAGRAMS*, 2ND ED., VOL 1, ASM INTERNATIONAL, 1990, P 226

Selection and Weldability of Advanced Titanium-Base Alloys

David H. Phillips, Edison Welding Institute

Weldability

There are many weldability (fusion welding) concerns associated with advanced titanium alloys. These differ appreciably from the weldability concerns that are characteristic of steels, such as hydrogen cracking, or of superalloys, such as strain age or liquation cracking. The weldability problems associated with advanced titanium alloys are related to their microstructural sensitivity to high heat input and rapid cooling rates. These thermal excursions, which are characteristic of many welding processes, can result in excessive beta grain growth and the formation of undesirable solid-state transformation products. The affected microstructures typically suffer from poor ductility and toughness. Such effects can be minimized, however, through proper welding practices, such as those discussed below.

Conventional Alloys. Each of the four families previously identified are described below in terms of weldability.

The near-alpha alloy Ti-1100 is generally considered to be a very weldable alloy, because of its high ductility. However, the excellent creep resistance of this alloy also means that high residual stresses will be developed when welding, especially with thick sections. These stresses may have to be relieved by a postweld heat treatment following welding, in order to avoid cracking.

The near-alpha alloy IMI 834 is also considered to be weldable, but the weld fusion zone and heat-affected zone (HAZ) may be void in primary alpha phase. These regions may not retain a good balance of creep and fatigue properties, for which IMI 834 is designed.

Alpha-Beta Alloys. The Corona 5 high-toughness alloy can be successfully welded, but it requires carefully controlled cooling rates, as well as postweld heat-treatment conditions. Excessively high cooling rates can result in the formation of orthorhombic martensite, a structure that exhibits inferior strength and fracture toughness. The aging of this martensite at low and intermediate postweld heat-treatment temperatures results in high strength, but low ductility and fracture toughness. However, through proper cooling and postweld heat treating, a fine alpha-beta fusion zone and HAZ microstructure that exhibits good mechanical properties can be created.

Near-Beta Alloys. The Ti-10V-2Fe-3Al alloy is very weldable in the heat-treated condition. Tensile and fracture toughness values that are near those of the base metal can be achieved with as-welded Ti-10V-2Fe-3Al.

Metastable-Beta Alloys. Although metastable-beta titanium alloys, such as Beta 21S, are generally considered to be weldable, cooling rates should be kept relatively high, and the weld fusion zone size should be minimized to avoid the formation of coarse columnar beta fusion zone grains and large beta HAZ grains, both of which can degrade ductility (Ref 6). If a relatively fine as-welded grain structure can be achieved, then moderate weld strength and high ductility can be achieved. Postweld heat treating to precipitate the alpha phase can significantly increase weld tensile strengths, but at the expense of ductility. Higher postweld aging temperatures may result in the formation of significant amounts of alpha phase along the beta grain boundaries, which changes the fracture mode, but does not necessarily degrade the properties.

Like Beta 21S, the Beta C alloy is considered to be highly weldable, but the fusion zone and HAZ beta grain size should be minimized (Ref 4). As-welded structures are characterized by low strength and high ductility. Relatively low postweld aging treatments can significantly increase strength levels, but ductility levels will be reduced, compared with those of the base metal, because of the large beta grain size and fine alpha phase distribution. Figure 2 shows a Beta C gas-tungsten arc weld microstructure, which reveals the large beta grain size. Higher postweld aging cycles coarsen the alpha phase, which reduces the strength but increases the ductility. As with Beta 21S, the higher aging temperatures can result in significant grain boundary alpha formation, which changes the fracture mode, but does not necessarily degrade the properties.

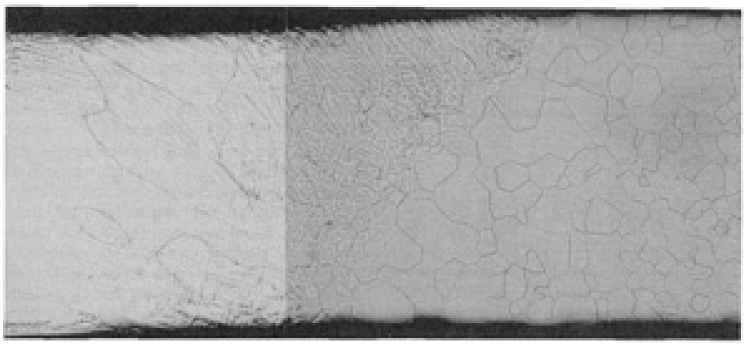

FIG. 2 MICROSTRUCTURE OF AS-WELDED GAS-TUNGSTEN ARC WELD IN BETA C SHEET

The intermetallic alloys are currently in the early stages of development. As a result, only limited weldability studies have been conducted to date. However, cursory fundamental studies have provided some indication of their weld sensitivities, and are discussed below.

Alpha-2 alloys form a similar weld and HAZ to many conventional alpha-beta titanium alloys. However, because of limited ductility and notch sensitivity, alpha-2 alloys are susceptible to solid-state fusion zone and HAZ cracking. Preheating to reduce residual stress and promote a more-ductile fusion zone and HAZ microstructure may be required in order to eliminate cracking. Weldability can also be improved by minimizing the size of the weld fusion zone and maximizing the cooling rate through the use of high-energy-density welding processes, such as laser-beam or electron-beam welding. Although this has the effect of retaining the metastable-beta phase, which can then be postweld heat treated, cracking may also occur during the heat treatment. Figure 3 shows cracking in the root of an alpha-2 electron-beam weld.

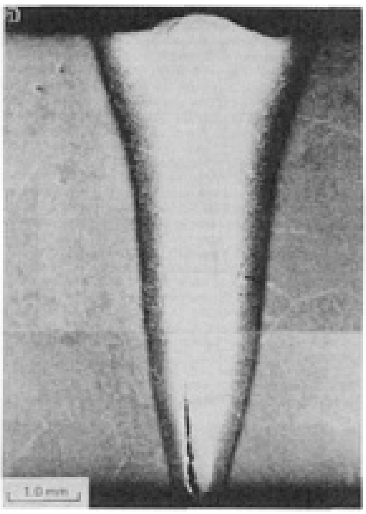

FIG. 3 CRACKING IN ALPHA-2 ELECTRON-BEAM WELD

Gamma Alloys. The gamma fusion zone and weld HAZ microstructures are significantly different from those of other titanium alloys. Depending on the weld process and metallographic technique used, the HAZ adjacent to the fusion zone can be nearly indistinguishable from the base-metal microstructure. Excessive grain growth does not appear to be an issue with gamma weldments. Because of low ductility and a high modulus, gamma alloys are very susceptible to weld fusion zone and HAZ cracking, resulting in the requirement for a weld preheat or controlled cooling.

Orthorhombic alloys are more ductile than the alpha-2 and gamma alloys, and are therefore easier to weld. Because of high beta-phase stabilities, a large retained-beta-phase HAZ will result from welding. Postweld heat treatments may be required to achieve an optimum combination of properties.

Titanium-Matrix Composites. The weldability issues relative to the monolithic titanium alloys also apply to titanium-matrix composites, with the additional complication of fiber/matrix interaction. Although some fiber damage from the welding process should not significantly degrade the properties, the damage should, of course, be minimized, where possible. Joint design is especially critical with titanium-matrix composites. The selection of a joining process is also critical, where processes such as brazing, solid-state diffusion welding, or transient liquid phase bonding may be required to minimize fiber damage. These processes eliminate or minimize base-metal melting that would damage the fibers.

References cited in this section

4. W.A. BAESLACK, P.S. LIU, D.B. BARBIS, J.R. SCHLEY, AND J.R. WOOD, POSTWELD HEAT TREATMENT OF GTA WELDS IN A HIGH-STRENGTH METASTABLE-BETA TITANIUM ALLOY--BETA-C, INTERNATIONAL TITANIUM CONFERENCE (SAN DIEGO), 1992
6. P.S. LIU, K.H. HOU, J. HURLEY, AND W.A. BAESLACK, LASER WELDING OF AN OXIDATION RESISTANT METASTABLE-BETA TITANIUM ALLOY--BETA-215, INTERNATIONAL TITANIUM CONFERENCE (SAN DIEGO), 1992

Selection and Weldability of Advanced Titanium-Base Alloys

David H. Phillips, Edison Welding Institute

References

1. P. BANIA, NEXT GENERATION TITANIUM ALLOYS FOR ELEVATED TEMPERATURE SERVICE, *ISIJ INT.*, VOL 31, 1991, P 840-847
2. OPTIMIZATION OF CREEP AND FATIGUE RESISTANCE IN HIGH TEMPERATURE TI ALLOYS IMI 829 AND IMI 834, *PROC. 5TH INT. CONFERENCE ON TITANIUM* (MUNICH, GERMANY), 1984, P 2419
3. D.W. BECKER, W.A. BAESLACK, AND F.H. FROES, POWDER METALLURGY OF TITANIUM ALLOYS, *PROC. CONF. AIME,* AIME, 1980
4. W.A. BAESLACK, P.S. LIU, D.B. BARBIS, J.R. SCHLEY, AND J.R. WOOD, POSTWELD HEAT TREATMENT OF GTA WELDS IN A HIGH-STRENGTH METASTABLE-BETA TITANIUM ALLOY--BETA-C, INTERNATIONAL TITANIUM CONFERENCE (SAN DIEGO), 1992
5. T.B. MASSALSKI, ED., *BINARY ALLOY PHASE DIAGRAMS,* 2ND ED., VOL 1, ASM INTERNATIONAL, 1990, P 226
6. P.S. LIU, K.H. HOU, J. HURLEY, AND W.A. BAESLACK, LASER WELDING OF AN OXIDATION RESISTANT METASTABLE-BETA TITANIUM ALLOY--BETA-215, INTERNATIONAL TITANIUM CONFERENCE (SAN DIEGO), 1992

Selection and Weldability of Heat-Treatable Aluminum Alloys

Richard P. Martukanitz, Applied Research Laboratory, Pennsylvania State University

Introduction

THE HEAT-TREATABLE ALUMINUM ALLOYS provide good strength and toughness in engineering applications while maintaining the low density and corrosion resistance of aluminum. These attributes allow the heat-treatable alloys to be used in a wide variety of applications, which include the aerospace, transportation, shipbuilding, tankage, piping, and appliance industries. The majority of these alloys are easily welded by the conventional arc welding processes (gas-metal arc and gas-tungsten arc welding), resistance spot and seam welding processes, as well as the high-energy processes (laser-beam and electron-beam welding).

However, these alloys also possess certain characteristics inherent to all aluminum alloys that must be considered during welding (Ref 1). Aluminum possesses a tenacious oxide that not only prevents corrosion, but may also impair welding if not minimized. This is achieved by mechanically or chemically removing heavy oxides prior to welding and using cathodic cleaning during direct current electrode positive (DCEP) gas-metal arc welding (GMAW) or alternating current gas-tungsten arc welding (GTAW). The high thermal conductivity of aluminum alloys requires the use of higher heat input for welding when compared to ferrous alloys. This requires greater welding current during arc welding of aluminum alloys. The high coefficient of thermal expansion of aluminum may also result in greater distortion when compared to ferrous alloys. Hence, aluminum is typically arc welded using high currents and high travel speeds. When in the liquid state, aluminum alloys possess a high solubility for hydrogen and very low solubility after solidification. Hence, proper cleaning and shielding must be utilized to ensure the production of welds free of gas porosity. Finally, aluminum alloys, and particularly the heat-treatable alloys, are sensitive to weld cracking. This condition is most easily avoided by proper filler and base alloy selection and adequate filler metal dilution. Anticipation of these characteristics and general knowledge of these materials allow the heat-treatable aluminum alloys to be readily welded.

Reference

1. P.B. DICKERSON AND B. IRVING, WELDING ALUMINUM: IT'S NOT AS DIFFICULT AS IT SOUNDS, *WELD. J.,* VOL 71 (NO. 4), 1992, P 45

Selection and Weldability of Heat-Treatable Aluminum Alloys

Richard P. Martukanitz, Applied Research Laboratory, Pennsylvania State University

Description of Heat-Treatable Aluminum Alloys

Designation Systems. The mechanical, physical, and chemical properties of aluminum alloys depend upon composition and microstructure. The addition of selected elements to pure aluminum greatly enhances its properties and usefulness. Because of this, most applications for aluminum utilize alloys having one or more elemental additions. The major alloying additions used with aluminum are copper, manganese, silicon, magnesium, and zinc. The total amount of these elements can constitute up to 10% of the alloy's composition (all percentages given in weight percent unless otherwise noted). Impurity elements are also present, but their total percentage is usually less than 0.15% in aluminum alloys. The two most commonly used methods of increasing the strength of aluminum alloys are to disperse alloying elements in solid solution and cold work the alloy, or to dissolve the alloying elements into solid solution and precipitate them throughout the alloy's matrix as strengthening particles (Ref 2). Alloys in which the first method applies are referred to as the work-hardenable alloys, and the latter method describes the precipitation-hardenable or heat-treatable alloys.

The designations used throughout North America for wrought aluminum alloys involve four digits, with the first digit indicating the major alloying additions, such as 6061 alloy representing the 6*xxx* alloy series. Aluminum casting alloy designations also involve four digits and may be proceeded by a letter to distinguish impurity levels or modifications of an original classification, such as A356.0 casting alloy. It should be pointed out that the fourth digit, which is separated from the others by a decimal point, indicates the product form (ingot or casting). The designations representing the heat-treatable wrought aluminum alloys are the 2*xxx* series (A1-Cu), the 6*xxx* series (Al-Mg-Si), and the 7*xxx* series (Al-Zn). The various combinations of alloying additions and strengthening mechanisms used for wrought aluminum alloys are shown in Table 1. The 8*xxx* aluminum alloy series denotes "other" major alloying constituents and may be found in both the work-hardenable and precipitation-hardenable classes of alloys. Table 2 provides the nominal composition of some of the common heat-treatable aluminum alloys (Ref 3). Included in Table 2 are compositions of selected heat-treatable casting alloys (Ref 4) and the recently commercialized aluminum-lithium alloys (Ref 5). Welding characteristics of aluminum-lithium alloys are described in the article "Selection and Weldability of Aluminum-Lithium Alloys" in this Volume.

TABLE 1 ALLOY DESIGNATIONS OF WROUGHT ALUMINUM ALLOYS

ALLOY SYSTEM	ALUMINUM SERIES
WORK-HARDENABLE ALLOYS	
PURE AL	1*XXX*
AL-MN	3*XXX*
AI-SI	4*XXX*
AL-MG	5*XXX*
AL-FE	8*XXX*
AL-FE-NI	8*XXX*
PRECIPITATION-HARDENABLE ALLOYS	
AL-CU	2*XXX*
AL-CU-MG	2*XXX*
AL-CU-LI	2*XXX*
AL-MG-SI	6*XXX*
AL-ZN	7*XXX*
AL-ZN-MG	7*XXX*
AL-ZN-MG-CU	7*XXX*
AL-LI-CU-MG	8*XXX*

TABLE 2 NOMINAL COMPOSITION OF SELECTED COMMON HEAT-TREATABLE ALUMINUM

ALLOYS

ALLOY	COMPOSITION, WT%						
	Si	Cu	Mn	Mg	Cr	Zn	OTHER
WROUGHT ALLOYS							
2014	0.8	4.4	0.8	0.5	. . .		
2024	. . .	4.4	0.6	1.5	. . .	. . .	. . .
2090	. . .	2.7	. . .	0.2	. . .	. . .	2.3 LI, 0.12 ZR
2095	. . .	4.0	. . .	0.4	. . .	. . .	1.0 LI, 0.4 AG
2219	. . .	6.3	0.3	. . .	. . .	. . .	0.06 TI, 0.18 ZR
2519	. . .	5.8	0.3	0.2	. . .	0.06	0.06 TI, 0.17 ZR
6005	0.8	. . .	. . .	0.5	. . .	. . .	. . .
6009	0.8	0.4	0.5	0.6	0.1	0.25	. . .
6010	1.0	0.4	0.5	0.8	0.1	0.25	. . .
6013	0.8	0.9	0.4	1.0	. . .	. . .	. . .
6061	0.6	0.3	. . .	1.0	0.2	. . .	. . .
6063	0.4	. . .	. . .	0.7	. . .	. . .	. . .
6262	0.6	0.3	. . .	1.0	0.1	. . .	0.6 PB, 0.6 BI
7005	. . .	. . .	0.5	1.4	0.1	4.5	0.04 TI, 0.14 ZR
7039	. . .	. . .	0.3	2.8	0.2	4.0	. . .
7050	. . .	2.3	. . .	2.2	. . .	6.2	0.12 ZR
7075	. . .	1.6	. . .	2.5	0.2	5.6	. . .
7178	. . .	2.0	. . .	2.8	0.2	6.8	. . .
8090	. . .	1.0	. . .	1.0	. . .	. . .	2.5 LI, 0.10 ZR
CASTING ALLOYS							
242.0	. . .	4.0	. . .	1.5	. . .	. . .	2.0 NI
319.0	6.0	3.5	. . .	. . .	. . .	. . .	. . .
355.0	5.0	1.2	. . .	0.5	. . .	. . .	. . .
356.0	7.0	. . .	. . .	0.4	. . .	. . .	. . .
357.0	7.0	. . .	. . .	0.6	. . .	. . .	. . .

Source: Ref 3, 4, 5

In commercial product forms, such as aluminum sheet, plate, extrusions, and forgings, these alloys have also undergone varying degrees of mechanical working at ambient and elevated temperatures. It is the combination of alloying additions and thermal-mechanical processing that imparts the desired characteristics to the product. Common considerations in designing alloy compositions and processing practices are strength, toughness, and resistance to corrosion.

The temper designation, which follows the alloy designation, is used to describe the thermal and mechanical processing history of the material. The common temper designations for aluminum alloys utilize an H to denote work-hardening, a T to describe various heat-treatment sequences, or an O to represent annealed material. The most common product forms of the heat-treatable alloys dictate the use of either the T or O tempers, such as 6061-T6 or 6061-O. One or more numbers may follow the T to indicate the specific sequence of thermal treatments. Table 3 summarizes the numerical designations used to describe the specific heat-treating sequences of the T-tempers (Ref 6). Additional information on systems for designating aluminum and aluminum alloys can be found in the article "Alloy and Temper Designation Systems for Aluminum and Aluminum Alloys" in Volume 2 of the *ASM Handbook.*

TABLE 3 NUMERICAL DESIGNATIONS FOR THE HEAT-TREATED TEMPERS OF ALUMINUM ALLOYS

TEMPER	DESCRIPTION
T1	COOLED FROM AN ELEVATED-TEMPERATURE SHAPING PROCESS AND NATURALLY AGED TO A SUBSTANTIALLY STABLE CONDITION

T2	COOLED FROM AN ELEVATED-TEMPERATURE SHAPING PROCESS, COLD WORKED, AND NATURALLY AGED TO A SUBSTANTIALLY STABLE CONDITION
T3	SOLUTION HEAT TREATED, COLD WORKED, AND NATURALLY AGED TO A SUBSTANTIALLY STABLE CONDITION
T4	SOLUTION HEAT TREATED AND NATURALLY AGED TO A SUBSTANTIALLY STABLE CONDITION
T5	COOLED FROM AN ELEVATED-TEMPERATURE SHAPING PROCESS AND THEN ARTIFICIALLY AGED
T6	SOLUTION HEAT TREATED AND THEN ARTIFICIALLY AGED
T7	SOLUTION HEAT TREATED AND OVERAGED/STABILIZED
T8	SOLUTION HEAT TREATED, COLD WORKED, AND THEN ARTIFICIALLY AGED
T9	SOLUTION HEAT TREATED, ARTIFICIALLY AGED, AND THEN COLD WORKED
T10	COOLED FROM AN ELEVATED-TEMPERATURE SHAPING PROCESS, COLD WORKED, AND THEN ARTIFICIALLY AGED

Source: Ref 6

Metallurgy. Wrought alloys that constitute heat-treatable (precipitation-hardenable) aluminum alloys include the 2*xxx,* 6*xxx,* 7*xxx,* and some of the 8*xxx* series alloys. They develop their maximum strength from appropriate solution, quenching, and precipitation treatments. The precipitation mechanism requires alloying elements with appreciable solid solubility in aluminum at elevated temperatures, but with limited solubility at lower temperatures. The solution heat treating operation, which involves heating the alloy to an elevated temperature, allows the alloyed elements to form a solid solution with aluminum. The temperature is then quickly reduced by quenching the material in water to create a supersaturated solid solution. A lower-temperature precipitation process, also referred to as artificial aging, is then used to precipitate strengthening phases throughout the metal in a controlled fashion. The driving force for precipitation is the supersaturation present after quenching.

This procedure is illustrated using the partial phase diagram for the aluminum-copper system in Fig. 1. Superimposed on the phase diagram are the temperature ranges used for solution heat treating and precipitation heat treating for a composition representing an aluminum alloy containing 4.0% Cu. Increasing the temperature of the 4.0% Cu alloy to 525 °C (980 °F) causes the copper to form a solid solution. When quenched to room temperature, the solid solution becomes supersaturated. If allowed to become stable, this condition is referred to as a T4 temper. Aging the alloy at a temperature of 175 °C (350 °F) causes nucleation and growth of strengthening precipitate. After solution heat treating and aging, the alloy condition is described as being in the T6 temper.

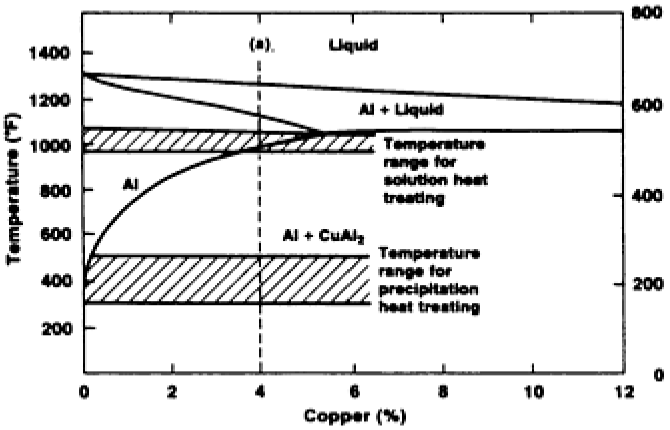

FIG. 1 PARTIAL PHASE DIAGRAM FOR THE ALUMINUM-COPPER SYSTEM SHOWING TEMPERATURE RANGES FOR SOLUTION AND PRECIPITATION HEAT TREATING

Many times these strengthening precipitates are not the equilibrium phases, such as $CuAl_2$ for the case of the 4.0% Cu alloy, but metastable phases. The metastable precipitates that form are determined by the time and temperature of the precipitation heat treating practice. The precipitation sequence for the aluminum-copper system is shown below:

$$\begin{array}{l}\text{SOLID SOLUTION} \rightarrow \text{GP1} \rightarrow \text{GP2} + \theta'' \rightarrow \theta' \rightarrow \theta \\ \qquad\quad (\text{CuAl}_2)\end{array} \tag{EQ 1}$$

In this system, strengthening is due to the formation of localized concentrations of copper atoms forming Guinier-Preston zones (GP1 and GP2) that are structurally coherent with the aluminum matrix. At longer aging times and higher aging temperatures the number of Guinier-Preston zones increase and lead to increased strength; the maximum strength is regarded as the peakaged condition. At still higher temperatures or longer times the Guinier-Preston zones are replaced by the noncoherent, metastable precipitates (θ'' and θ'). The decrease in the amount of Guinier-Preston zones formed and the loss of particle coherency result in a subsequent reduction of strength, termed the overaged condition. Figure 2 shows the relationship between strength and precipitate formation during aging of an aluminum 4.0% Cu alloy at 190 °C (375 °F) and 130 °C (265 °F) (Ref 7).

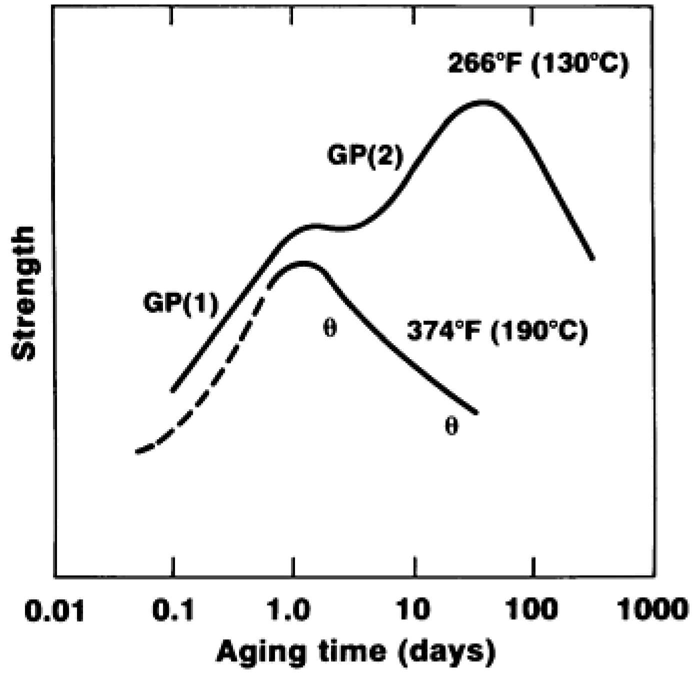

FIG. 2 RELATIONSHIP BETWEEN STRENGTH AND PRECIPITATE FORMATION DURING AGING OF AN AL-4CU ALLOY. SOURCE: REF 7

Characteristics and Applications. Alloying additions and subsequent thermal and mechanical processing dictate characteristics such as strength, toughness, and corrosion resistance of the heat-treatable aluminum alloys. However, it is usually the combination of these properties, that is, strength versus toughness, that is controlled, depending on application requirements. The utility of the heat-treatable aluminum alloys is greatly enhanced by the wide variety of product forms available. This includes sheet, plate, rod, bar, extruded shapes, forgings, and castings.

The 2*xxx* series alloys, having copper as the primary alloying addition, possess high strength but somewhat lower corrosion resistance than most other aluminum alloys. Many of these alloys also possess relatively good elevated-temperature strength. Magnesium is also added to the 2*xxx* series alloys for increased strength but results in greater weld crack sensitivity. Alloy 2024 is one of the highest strength 2*xxx* alloys and is used extensively in the aircraft industry; however, the 1.5% nominal Cu content of 2024 hinders its weldability due to increased weld crack sensitivity. Alloys having controlled levels of magnesium, such as 2014, 2219, and 2519, are utilized for applications requiring good weldability. Generally, these alloys provide good machinability, but less formability, than other heat-treatable aluminum alloys. Some applications of the 2*xxx* alloys are aircraft skin sheet, automotive panels, ballistic armor, and forged and machined components.

The 6*xxx* series alloys containing magnesium and silicon provide moderate strengths and good corrosion resistance in relation to other heat-treatable aluminum alloys. Because they are easily extruded, they are available in a wide range of structural shapes, as well as sheet and plate products. The versatility of these alloys is represented in 6061, which is one of the most commonly used aluminum alloys. Typically, the 6*xxx* alloys have good formability and good weldability. The 6*xxx* series alloys are used in autobody sheet, structural members, architectural panels, piping, marine applications, screw machine stock, and many other applications.

The 7*xxx* series provides the highest strength of all aluminum alloys, although this claim is being challenged by the recently developed aluminum-lithium alloys. High strength in the 7*xxx* alloys is achieved by alloying additions of zinc, magnesium, and often copper, combined with controlled thermal and mechanical processing. Copper, in combination with zinc and magnesium in the 7*xxx* series alloys, increases strength but hampers weldability due to increased susceptibility to weld cracking. Alloy 7075, containing nominally 5.6% Zn, 2.5% Mg, and 1.6% Cu, is a common alloy of this system having a propensity for weld cracking. Alloys 7004, 7005, and 7039 were designed for applications requiring high strength and good weldability. These alloys have limited amounts of copper that reduce the sensitivity to weld cracking. One important characteristic of the 7*xxx* series is the ability of the heat-affected zone (HAZ) in these alloys to naturally age, or become precipitation-strengthened at room temperature, after welding. Because of this, weld properties continue to improve for up to 30 days after welding. These alloys typically provide good machinability and adequate formability. Applications for the 7*xxx* alloys include aircraft construction, truck trailers, railcars, and armor plate.

References cited in this section

2. *ALUMINUM: PROPERTIES AND PHYSICAL METALLURGY,* J.E. HATCH, ED., AMERICAN SOCIETY FOR METALS, 1984, P 222
3. *ALUMINUM STANDARDS AND DATA,* THE ALUMINUM ASSOCIATION, INC., 1988, P 15
4. *ALUMINUM: PROPERTIES AND PHYSICAL METALLURGY,* J.E. HATCH, ED., AMERICAN SOCIETY FOR METALS, 1984, P 320
5. R.S. JAMES, ALUMINUM-LITHIUM ALLOYS, VOL 2, *ASM HANDBOOK,* ASM INTERNATIONAL, 1990, P 178-199
6. *ALUMINUM STANDARDS AND DATA,* THE ALUMINUM ASSOCIATION, INC., 1988, P 11
7. J.M. SILCOCK, T.J. HEAL, AND H.K. HARDY, STRUCTURAL AGEING CHARACTERISTICS OF BINARY ALUMINUM-COPPER ALLOYS, *J. INST. MET.,* VOL 82, 1953-1954, P 239

Selection and Weldability of Heat-Treatable Aluminum Alloys

Richard P. Martukanitz, Applied Research Laboratory, Pennsylvania State University

Welding of Heat-Treatable Aluminum Alloys

As stated earlier, the heat-treatable aluminum alloys possess characteristics inherent to all aluminum alloys that must be considered before successful welding can be employed. These characteristics include a tenacious oxide, high thermal conductivity, high coefficient of thermal expansion, solidification shrinkage almost twice that of ferrous alloys, relatively wide solidification-temperature ranges, and high solubility of hydrogen when in the molten state. The first two items may be addressed by weld pretreatment and processing conditions--proper weld pretreating to minimize the oxide and welding at high currents and high travel speeds to ensure that adequate penetration is utilized. However, the remaining characteristics, if not carefully considered, may result in the production of welds of unacceptable quality. The high coefficient of thermal expansion and solidification shrinkage, combined with the relatively wide solidification-temperature range, makes these alloys susceptible to weld cracking. The high solubility of hydrogen in molten aluminum can result in gas porosity unless proper precautions are heeded. In addition, the degree of degradation of the HAZ for these alloys may differ markedly, depending upon weld processing conditions. An understanding of these aspects yields welds in the heat-treatable aluminum alloys having excellent joint properties.

Crack Sensitivity during Welding

Weld cracking in aluminum alloys is of concern due to aluminum's relatively high thermal expansion, large change in volume upon solidification, and wide solidification-temperature range. The weld crack sensitivity of heat-treatable aluminum alloys is especially of concern due to the greater amounts of alloying additions used for these alloys. Because of the detrimental effect of weld cracks on joint properties, the weldability of aluminum alloys is defined as its resistance to weld cracking. Weld cracking in aluminum alloys may be classified into two primary categories based on the mechanism responsible for cracking and crack location. Solidification cracking takes place within the weld fusion zone and typically appears along the center of the weld or at termination craters. Liquation cracks occur adjacent to the fusion zone and may or may not be readily apparent.

Solidification cracking, or hot tearing, occurs when high levels of thermal stress and solidification shrinkage are present while the weld pool is undergoing various degrees of solidification. The hot tearing sensitivity of any given aluminum alloy is influenced by a combination of mechanical, thermal, and metallurgical factors (Ref 8). The degree of restraint of welded assemblies plays a significant role in crack sensitivity by increasing the stress on the solidifying weld. Hot tearing occurs within the weld fusion zone and is affected by weld-metal composition and welding parameters. High heat inputs, such as high currents and slow welding speeds, are believed to contribute to weld solidification cracking (Ref 9). It follows that processes that result in minimal heat input, such as electron-beam welding, reduce weld crack sensitivity. The primary method for eliminating cracking in aluminum welds is to control weld-metal composition through filler alloy additions. Crack sensitivity, determined experimentally as a function of weld composition, is shown in Fig. 3 for various binary aluminum systems (Al-Li, Al-Si, Al-Cu, Al-Mg, and the quasibinary Al-Mg_2Si) (Ref 10, 11, 12, 13, 14, 15).

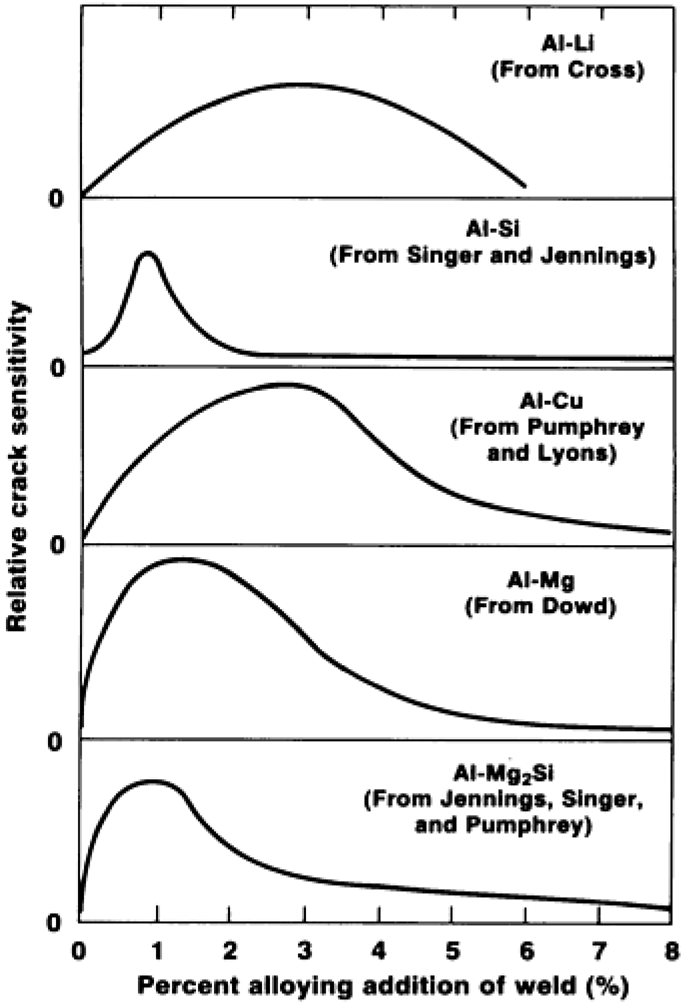

FIG. 3 RELATIVE CRACK SENSITIVITY VERSUS WELD COMPOSITION FOR VARIOUS BINARY ALUMINUM SYSTEMS. SOURCE: REF 10, 11, 12, 13, 14, 15

Although most commercial aluminum alloys are not strictly binary systems, crack-sensitivity curves based on binary systems are still useful in selecting a filler alloy to minimize weld crack sensitivity. In complex alloy systems, minor alloying additions may impact crack sensitivity by widening the temperature range in which a coherent solidifying structure is formed. This effect is shown in Fig. 4, which describes the weld crack sensitivity of two commercial quaternary aluminum systems (Al-Zn-Mg-Cu and Al-Mg-Si-Cu) and a ternary system (Al-Cu-Mg) (Ref 16, 17, 18). In all cases, the curves show the dramatic effect of minor alloying additions, such as copper or magnesium, on weld crack sensitivity. Several high-strength, precipitation-hardenable aluminum alloys rely on complex alloying additions to develop properties; however, these additions dramatically increase weld crack sensitivity. The most notable examples are alloys 2024 (Al-4.4Cu-1.5Mg) and 7075 (Al-5.6Zn-2.5Mg-1.6Cu). Increased weld crack sensitivity may also result when welding dissimilar aluminum alloys to produce weld compositions that widen the coherency temperature range, such as welding 2*xxx* series to 5*xxx* series alloys (Al-Cu alloys to Al-Mg alloys).

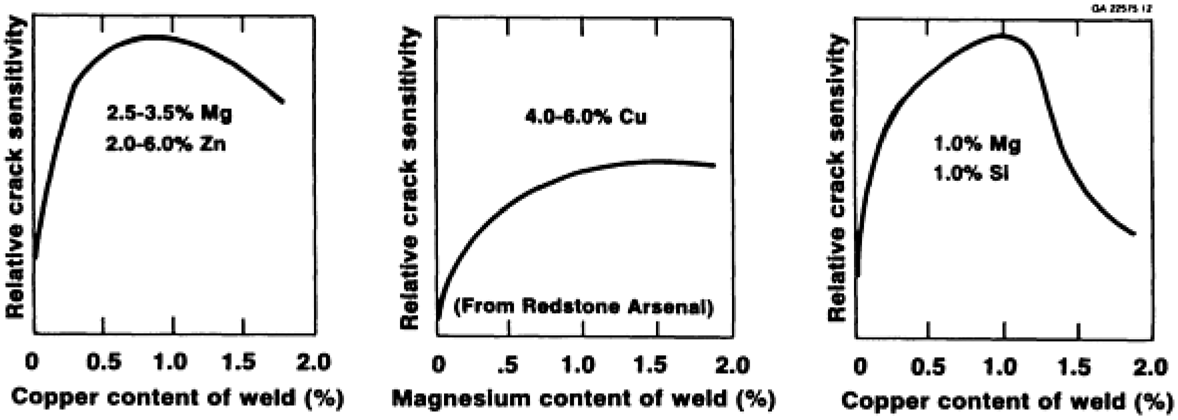

FIG. 4 WELD CRACK SENSITIVITY OF TWO COMMON QUATERNARY ALUMINUM ALLOY SYSTEMS (AL-MG-ZN-CU AND AL-MG-SI-CU) AND A COMMON TERNARY SYSTEM (AL-CU-MG). SOURCE: REF 16, 17, 18

Liquation Cracking. An important element of the HAZ for precipitation-hardenable alloys is the thin boundary layer adjacent to the fusion zone that is referred to as the partially melted region. This region is produced when eutectic phases or constituents that have low melting points (melting points below the melting point of the bulk material) liquate, or melt, at grain boundaries during welding (Ref 19). It occurs in precipitation-hardenable alloys because of the relatively large amount of alloying additions available to form eutectic phases. During welding, these phases liquate and, if sufficient stress is present, may be accompanied by tears. Under extreme conditions, continuous cracks may form along the fusion zone interface. As expected, higher heat input widens the partially melted region and makes it more prone to cracking. An example of cracks formed in this region for alloy 6013-T6 welded with 5356 filler alloy is shown in Fig. 5 (Ref 20). Filler alloy composition greatly affects whether cracks form in this liquated region (Ref 21 and 22). Filler alloys that have low-solidification temperatures provide less susceptibility to liquation cracks, since the solidification shrinkage strains occur at lower temperatures. This allows healing, or solidification, of the partially melted region to occur prior to solidification shrinkage strains.

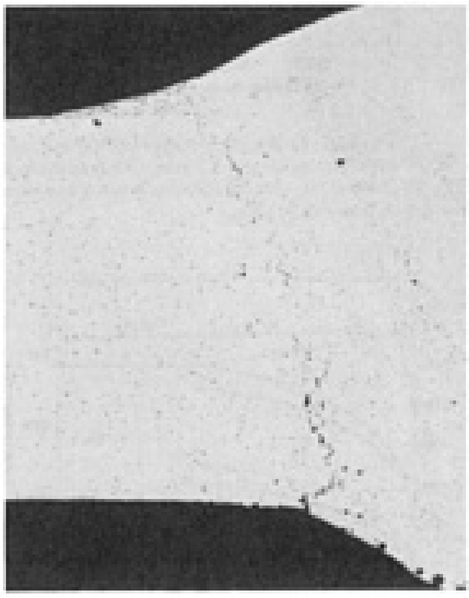

FIG. 5 MICROGRAPH OF A GAS-TUNGSTEN ARC WELD ON 2.0 MM (0.080 IN.) THICK 6013-T6 BASE METAL WITH 5356 FILLER ALLOY SHOWING CRACKS IN THE PARTIALLY MELTED REGION. SOURCE: REF 20

Porosity during Welding

Porosity in aluminum weldments occurs when hydrogen gas is entrapped during solidification (Ref 23 and 24). Hydrogen has an appreciable solubility in molten aluminum and a low solubility in the solid. Figure 6 shows the solubility of hydrogen in pure aluminum at temperatures representing the solid and liquid states (Ref 25).

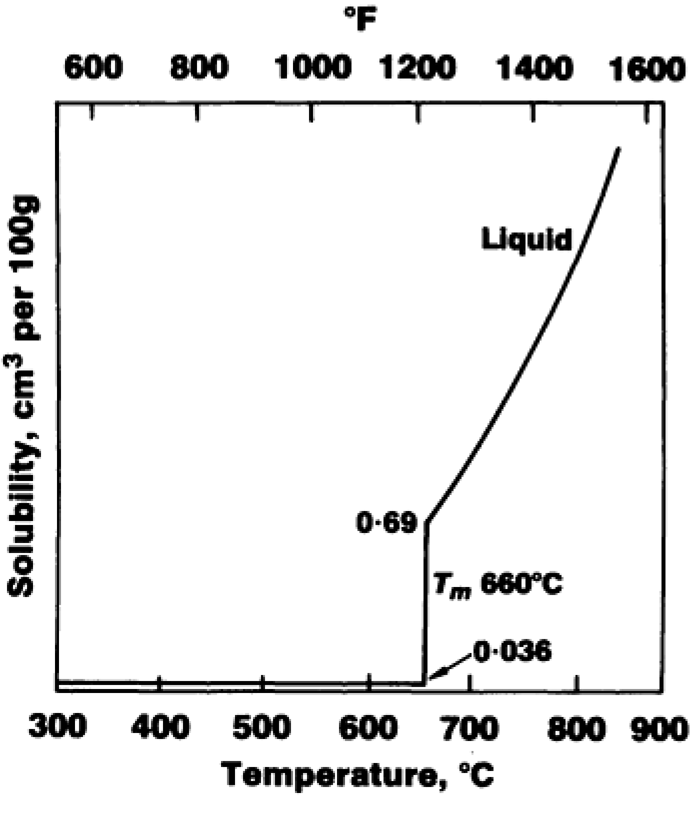

FIG. 6 HYDROGEN SOLUBILITY IN PURE ALUMINUM. SOURCE: REF 25

Hydrogen is absorbed into the molten pool during welding because of its high solubility, and it forms gas pores upon solidification due to the decrease in solubility. It is this difference in solubility that is the driving force for porosity formation (Ref 26). Once a gas bubble forms, it may be expelled from the weld by natural buoyancy or forced convection within the molten pool. Redistribution of hydrogen during solidification may also act to push pores ahead of the solidifying front. Therefore, the welding position and weld parameters may dramatically alter the amount of porosity found in solidified welds.

The expulsion of hydrogen pores may be described by the pore's buoyancy velocity and welding speed. If the solidification front is moving at a lower velocity than the pore's buoyancy velocity, gas pores may escape (Ref 27). Lower welding speeds create slower solidification fronts and favor expulsion of gas pores. Welding position may either help or hinder the expulsion of pores. In general, vertical-up welding of aluminum produces the least porosity because the solidifying pool provides easy escape of gas pores. Converse]y, overhead welding produces the greatest amount of porosity because float direction and pool convection patterns cause pores to be entrapped in the weld root during solidification.

The sources of hydrogen present in the welding system depend on the particular welding process. In arc welding these sources are: (1) hydrogen from the base metal, (2) hydrogen from the filler metal, and (3) hydrogen within the shielding gas (Ref 28). This is illustrated in Fig. 7 for GMAW. Hydrogen from the base metal may be attributed to the surface in the form of hydrocarbons (lubricants) and hydrated oxides, or it may be internal, as with castings and aluminum powder-metallurgy (PM) parts. Filler metal hydrogen is in the form of hydrocarbons (lubricants) or hydrated oxides on the surface of the filler. Hydrogen in the shielding gas is due to moisture contamination within the gas or aspiration of moist air into the gas system due to defective gas lines and couplings. A dew-point determination of the shielding gas may provide an indication of moisture contamination. Normally, a dew-point measurement of -50 °C (-60 °F) or below, measured at the gas cylinder outlet, is considered sufficient gas quality (Ref 29).

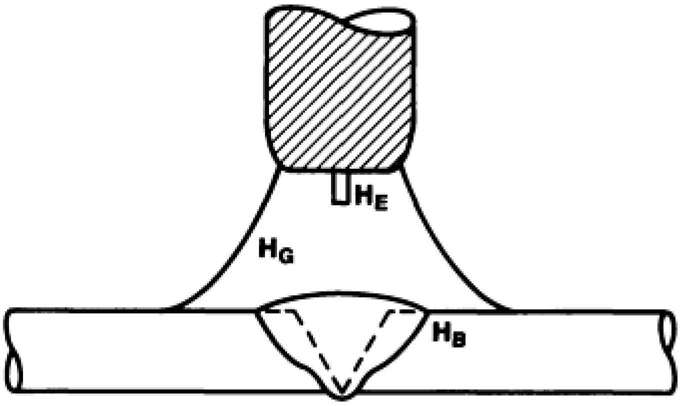

FIG. 7 SOURCES OF HYDROGEN IN GAS-METAL ARC WELDING. H_G, HYDROGEN FROM SHIELDING GAS; H_E, HYDROGEN FROM ELECTRODE; H_B, HYDROGEN FROM BASE METAL. SOURCE: REF 28

In the case of GMAW of aluminum wrought products, the filler electrode, shielding gas, and base-metal surface contaminants may contribute to gas porosity. In the case of welding aluminum castings or PM parts, internal hydrogen of the base metal typically governs pore formation (Ref 30). Because porosity in aluminum welds is attributed to a source of hydrogen contamination within the welding system, preventative measures may be taken to ensure sound welds. This begins with proper precleaning of the parts to be welded. Mill and machining lubricant must be removed by solvent degreasing before welding. Stainless steel wire brushing of the joint to remove thick mill oxides is recommended just prior to welding. During GTAW of aluminum, alternating current with sufficient electrode-positive polarity provides excellent arc cleaning action to remove surface oxides. For GMAW, procurement of quality electrodes and proper electrode storage are necessary to minimize porosity. In all cases, maintenance of gas hoses, regulators, and auxiliary equipment is required to produce porosity-free welds.

HAZ Degradation

A HAZ is created adjacent to the fusion zone and results in a degradation of base-metal properties. This degradation is caused by microstructural modifications associated with elevated temperatures experienced in this zone. For heat-treatable aluminum alloys, the HAZ is distinguished by dissolution or growth of precipitates. The response in the HAZ of 2*xxx* aluminum alloys follows dissolution, whereas the primary modification for the 6*xxx* alloys is growth of precipitates (Ref 31, 32, 33). Although the nature of these HAZs may differ, they are all diffusion-controlled and thermally dependent (Ref 34). A common method of determining the width and extent of the HAZ is by measuring the hardness across this zone. Hardness profiles for gas-tungsten arc welds of two common heat-treatable alloys, 2219-T87 and 6061-T6, are shown in Fig. 8 (Ref 35). A hardness profile for the work-hardenable alloy 5456-H116 is also shown in Fig. 8 for comparison.

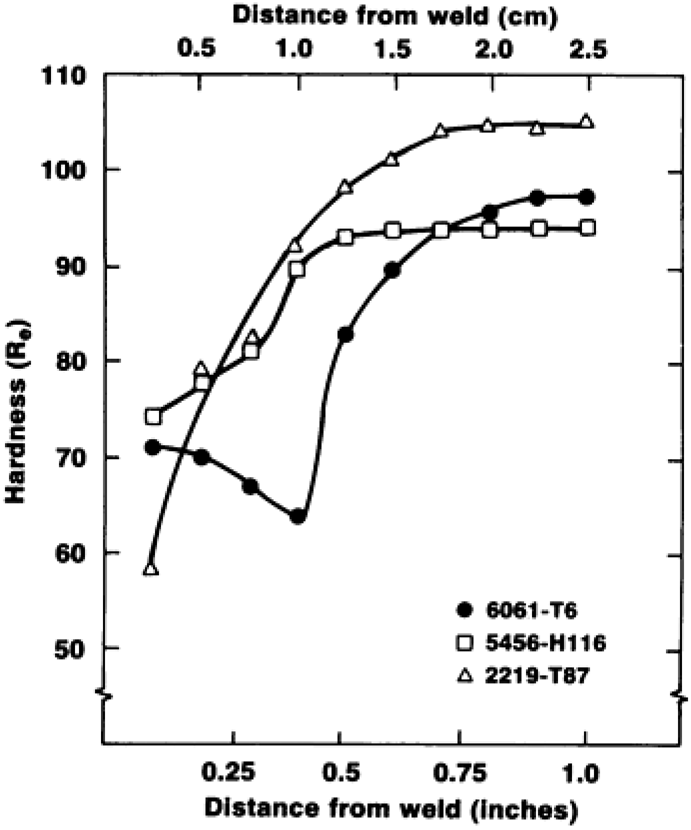

FIG. 8 HARDNESS PROFILES ACROSS THE HAZ OF GAS-TUNGSTEN ARC WELDS ON 3.2 MM (0.125 IN.) THICK 2219-T87, 5456-H116, AND 6061-T6 MADE USING CONSTANT HEAT INPUT. SOURCE: REF 35

The hardness profile for 2219-T87 exhibits dissolution of precipitates. The strengthening particles for these alloys are Guinier-Preston zones, which are metastable precipitates. The solvus curves for these phases are superimposed on the partial aluminum-copper phase diagram in Fig. 9 (Ref 31). When the temperatures in the HAZ exceed the various solvi, the respective phases are dissolved. At positions close to the fusion zone, higher temperatures are experienced and greater dissolution of strengthening phases occurs. This results in a continual decrease in strength in the HAZ of aluminum-copper alloys and is illustrated in Fig. 8 for 2219-T87 alloy.

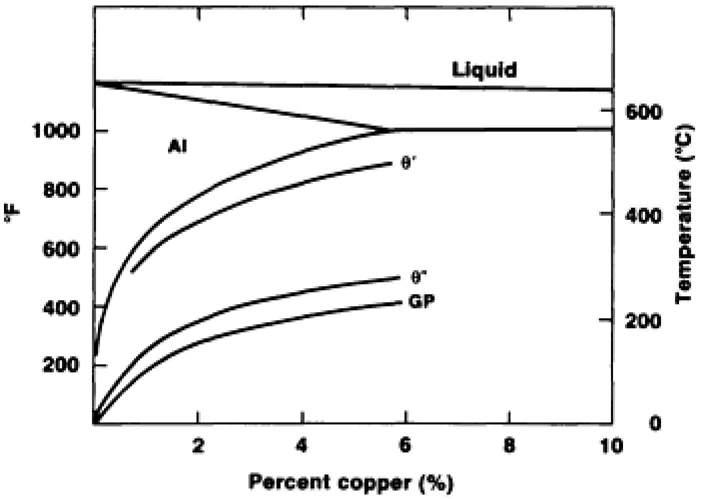

FIG. 9 PARTIAL PHASE DIAGRAM FOR THE ALUMINUM-COPPER SYSTEM SHOWING METASTABLE PHASE SOLVI. SOURCE: REF 31

The hardness profile for the HAZ of 6061-T6 is representative of the 6*xxx* series alloys and is governed by a growthlike transformation of precipitate. This system also exhibits a precipitation sequence of metastable precipitates, with the principal strengthening phases occurring early in the sequence in the form of Guinier-Preston zones. The transformation of these precipitates at elevated temperatures occurs in a smooth, continuous fashion and is closely akin to growth (Ref 31). The result is the formation of nonstrengthening phases. Due to the competing nature of the diffusion rate and supersaturation with increasing temperatures, the transformation proceeds at the greatest rate when temperatures are between 290 and 425 °C (550 and 800 °F) (Ref 36, 37, 38). The minimum hardness value for 6061-T6 in Fig. 8 represents the position in the HAZ that experiences temperatures within this range. This minimum has also been referred to as the overaged zone for these alloys. At higher temperatures, indicative of positions close to the fusion zone, particles are dissolved into solid solution and upon cooling may precipitate as strengthening phases. This contributes to the slight increase in hardness that typically occurs near the fusion zone for the 6*xxx* series alloys.

Effect of Heat Input on the HAZ. Because the metallurgical transformations of the HAZ of aluminum alloys are thermally dependent, the welding process and parameters determine the extent of base-metal degradation. High heat input and preheating increase both the degree and width of the HAZ. This is especially true for the precipitation-hardenable aluminum alloys and is illustrated for 6061-T6 welds made using various heat inputs in Fig. 10 (Ref 39). Thus, the HAZ degradation of the heat-treatable alloys can be reduced by employment of multipass welding, close control of interpass temperature, and elimination of preheating. The highest weld strengths for these alloys are achieved when interpass temperatures do not exceed 65 °C (150 °F) and welding heat input is minimized.

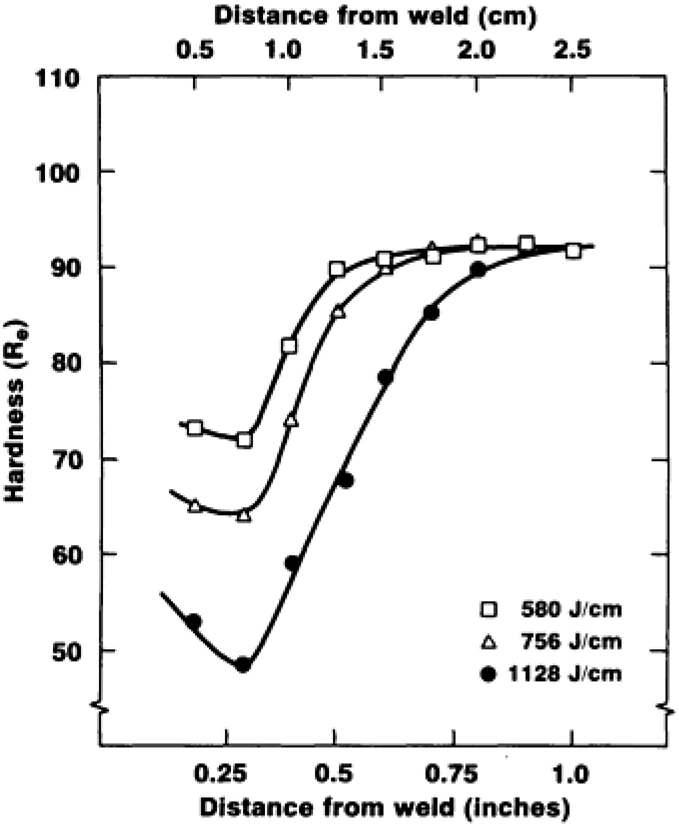

FIG. 10 HARDNESS PROFILES OF THE HAZ OF GAS-TUNGSTEN ARC WELDS ON 6061-T6 USING VARIOUS HEAT INPUTS. SOURCE: REF 39

Postweld Heat Treatment and the HAZ Postweld thermal heat treatments can also be used to improve the strength of the HAZ for heat-treatable alloys. This may involve complete postweld solution heat treating and aging or postweld aging only. Although the recovery of strength in the HAZ after postweld aging is less than postweld solution heat treating and aging, there are advantages to postweld aging only. Temperatures used for postweld aging are much lower than postweld solution heat treating, and postweld aging does not involve water quenching that imposes residual stresses and distortion to welded assemblies. However, great strides have been made in the development and application of polymer quenchants that reduce the amount of distortion observed after quenching and may complement the use of postweld solution heat treating and aging.

Figure 11 depicts hardness profiles of the HAZ for 6061-T4 and 6061-T6 starting material in the as-welded and postweld aged conditions (Ref 40). A dramatic recovery of strength in the HAZ is observed when welding 6061-T4 and postweld aging. Postweld thermal treatments may be used to recover strength in the HAZ of the heat-treatable alloys; however, when 6061-T6 is postweld aged, strength in the HAZ is increased while strength in the unaffected base metal is decreased. This is due to overaging of the 6061-T6 base metal; therefore, T4 or T3 material should be utilized when postweld aging the precipitation-hardenable alloys. The 7*xxx* series alloys possess the ability to naturally age (age at room temperature) in the HAZ. The recovery of strength is similar to postweld artificial aging, but because of the relatively low-temperature aging, the time is substantial. Usually, a period of 14 days is required for improving the HAZ strength of the 7*xxx* series alloys by natural aging.

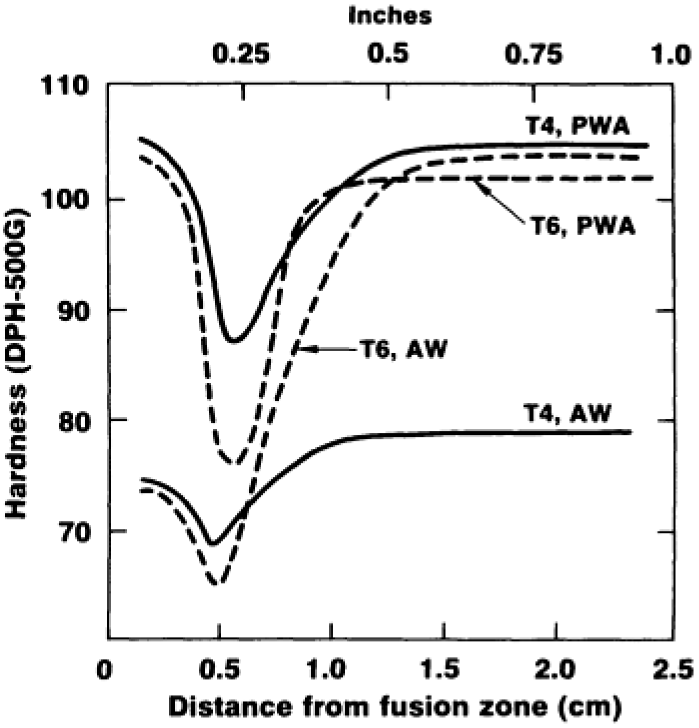

FIG. 11 HARDNESS PROFILES OF THE HAZ FOR 6061-T4 AND T6 STARTING MATERIALS IN THE AS-WELDED (AW) AND POSTWELD AGED (PWA) CONDITIONS. SOURCE: REF 40

References cited in this section

8. C.E. CROSS, W.T. TACK, L.W. LOECHEL, AND L.S. KRAMER, ALUMINUM WELDABILITY AND HOT TEARING THEORY, *WELDABILITY OF MATERIALS,* ASM INTERNATIONAL, 1990, P 275
9. R.A. CHIHOSKI, THE CHARACTER OF STRESS FIELDS AROUND A WELD ARC MOVING ON ALUMINUM SHEET, *WELD J.,* VOL 51 (NO. 1), 1972, P 9-S
10. C.L. CROSS, "WELDABILITY OF ALUMINUM-LITHIUM ALLOYS: AN INVESTIGATION OF HOT TEARING MECHANISMS," PH.D. THESIS, COLORADO SCHOOL OF MINES, GOLDEN, CO, 1986, P 144
11. A.R.E. SINGER AND P.H. JENNINGS, HOT SHORTNESS OF THE ALUMINUM SILICON ALLOYS OF COMMERCIAL PURITY, *J. INST. MET.,* VOL 73, 1947, P 197
12. W.I. PUMPHREY AND J.V. LYONS, CRACKING DURING THE CASTING AND WELDING OF THE MORE COMMON BINARY ALUMINUM ALLOYS, *J. INST. MET.,* VOL 74, 1948, P 439
13. J.D. DOWD, WELD CRACKING OF ALUMINUM ALLOYS, *WELD. J.,* VOL 31 (NO. 10), 1952, P 448-S
14. P.H. JENNINGS, A.R.E. SINGER, AND W.I. PUMPHREY, HOT-SHORTNESS OF SOME HIGH-PURITY ALLOYS IN THE SYSTEMS ALUMINUM-COPPER-SILICON AND ALUMINUM-MAGNESIUM-SILICON, *J. INST. MET.,* 1948, P 227
15. J.H. DUDAS AND F.R. COLLINS, PREVENTING CRACKS IN HIGH-STRENGTH ALUMINUM ALLOYS, *WELD. J.,* VOL 45 (NO. 6), 1966, P 241-S
16. W.I. PUMPHREY AND D.C. MOORE, CRACKING DURING AND AFTER SOLIDIFICATION IN SOME ALUMINUM-COPPER-MAGNESIUM ALLOYS OF HIGH PURITY, *J. INST. MET.,* VOL 74, 1948, P 425
17. J.K. DAWSON, WELD CRACK SENSITIVITY OF ALUMINUM ALLOYS, *MET. PROG.,* VOL 76 (NO. 1), 1959, P 116

18. R.P. MEISTER AND D.C. MARTIN, "WELDING OF ALUMINUM AND ALUMINUM ALLOYS," DEFENSE METALS INFORMATION CENTER, BATTELLE MEMORIAL INSTITUTE, 1967, P 51
19. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987, P 239
20. R.P. MARTUKANITZ, ALCOA LABORATORIES, UNPUBLISHED RESEARCH, 1985
21. N.F. GITTOS AND M.H. SCOTT, HEAT AFFECTED ZONE CRACKING OF AL-MG-SI ALLOYS, *WELD. J.,* VOL 60 (NO. 6), 1981, P 95-S
22. H.W. KERR AND M. KATOH, INVESTIGATION OF HEAT-AFFECTED ZONE CRACKING OF GMA WELDS OF AL-MG-SI ALLOYS USING THE VARESTRAINT TEST, *WELD. J.,* VOL 66 (NO. 9), 1987, P 251-S
23. D.G. HOWDEN AND D.R. MILNER, *BRIT. WELD. J.,* VOL 10, 1963, P 304
24. Z.P. SAPERSTEIN, G.R. PRESCOTT, AND E.W. MONROE, POROSITY IN ALUMINUM WELDS, *WELD. J.,* VOL 43 (NO. 10), 1964, P 443-S
25. C. RANSLEY AND J. NEUFELD, *J. INST. MET.,* VOL 74, 1948, P 617
26. M. UDA AND S. OHNO, *TRANS. NATL. RES. INST. MET.* (JPN), VOL 16, 1974, P 2
27. R.P. MARTUKANITZ, ALCOA LABORATORIES, UNPUBLISHED RESEARCH, 1986
28. R.P. MARTUKANITZ AND P.R. MICHNUK, SOURCES OF POROSITY IN GAS METAL ARC WELDING OF ALUMINUM, *ALUMINIUM,* VOL 58 (NO. 5), 1982, P 276
29. P.B. DICKERSON, ALCOA LABORATORIES, PRIVATE COMMUNICATION, 1987
30. A.N. O'SHANSKII AND V.V. D'YACHENCHO, EVALUATION OF THE SUSCEPTIBILITY OF ALLOYS TO THE FORMATION OF PORES IN WELDING, *WELD. PROD.,* VOL 24 (NO. 7), 1977, P 52
31. S.D. DUMOLT, "METALLURGICAL TRANSACTIONS IN THE HEAT AFFECTED ZONE OF ALUMINUM ALLOYS BY TRANSMISSION ELECTRON MICROSCOPY," PH.D. THESIS, CARNEGIE-MELLON UNIVERSITY, PITTSBURGH, PA, 1983
32. T. ENGO AND T. KURODA, MICROSTRUCTURE IN WELD HEAT-AFFECTED ZONE OF AL-MG-SI ALLOYS, *TRANS. JPN. WELD. RES. INST.,* VOL 11 (NO. 1), 1982, P 61
33. R.P. MARTUKANITZ, "MODELING THE HEAT AFFECTED ZONES OF ALUMINUM ARC WELDS, *ADVANCES IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID, ED., ASM INTERNATIONAL, 1986, P 193
34. S. KOU, WELDING METALLURGY AND WELDABILITY OF HIGH STRENGTH ALUMINUM ALLOYS, *WELD. RES. COUNC. BULL.,* NO. 320, WELDING RESEARCH COUNCIL, 1986
35. R.P. MARTUKANITZ, ALCOA LABORATORIES, UNPUBLISHED RESEARCH, 1985
36. S.D. DUMOLT, PH.D. THESIS, CARNEGIE-MELLON UNIVERSITY, PITTSBURGH, PA, 1983
37. W.L. BURCH, THE EFFECT OF WELDING SPEED ON STRENGTH OF 6061-T4 ALUMINUM JOINTS, *WELD. J.,* VOL 37 (NO. 7), 1958, P 361-S
38. R.P. MARTUKANITZ, "MODELING THE HEAT AFFECTED ZONES OF ALUMINUM ARC WELDS, ADVANCES IN WELDING SCIENCE AND TECHNOLOGY, S.A. DAVID, ED., ASM INTERNATIONAL, 1986, P 195
39. R.J. BRUNGRABER AND F.G. NELSON, EFFECT OF WELDING VARIABLES ON ALUMINUM ALLOY WELDMENTS, *WELD. J.,* VOL 52 (NO. 3), 1973, P 97-S
40. G.E. METZGER, SOME MECHANICAL PROPERTIES OF WELDS IN 6061 ALUMINUM ALLOY SHEET, *WELD. J.,* VOL 56 (NO. 10), 1967, P 457-S

Selection and Weldability of Heat-Treatable Aluminum Alloys

Richard P. Martukanitz, Applied Research Laboratory, Pennsylvania State University

---

Filler Alloy Selection and Properties of Welds

Although the selected base alloy and welding process greatly influence the characteristics of aluminum joints, the filler alloy plays a significant role in establishing a number of important weld properties for hem-treatable aluminum alloys. Of foremost concern is the weldability of a particular base alloy and filler alloy combination, which is defined as its susceptibility to weld cracking. Shear strength of a fillet weld, response to postweld heat treatments, and color match after anodizing are other attributes that are highly governed by the filler alloy used to produce the weld in heat-treatable aluminum alloys. Properties such as joint strength, ductility, toughness, and corrosion resistance across the weld are also dictated by the base alloy and filler alloy combination selected. The filler alloy selection chart in the article "Welding of Aluminum Alloys" in this Volume (see Table 14 in that article) provides a comprehensive guide for selecting the proper filler metal. Recently a knowledge-based system using computer software has also been constructed to allow greater interaction in selecting aluminum base metal and filler metal alloys for welding (Ref 41).

Filler Alloy Selection and Weld Crack Sensitivity

Weld crack sensitivity is most easily controlled by influencing the composition of the weld. This is achieved by selecting the proper filler alloy and ensuring adequate dilution of the filler alloy to the weld. Generally, joint designs that increase filler metal dilution (for example, square groove with gap as opposed to no gap) are beneficial in reducing the tendency for cracking. The principal filler alloys for welding aluminum are the 1*xxx* (1100), 2*xxx* (2319), 4*xxx* (4043, 4047, 4145, and 4643), and 5*xxx* (5154, 5183, 5356, 5554, 5556, and 5654) series. Nominal compositions of aluminum filler alloys are shown in Table 4 (Ref 42). Most of these alloys are high in solute to reduce crack sensitivity. Because of their narrow solidification temperature range, the 4*xxx* series filler alloys provide excellent insensitivity to weld cracking but are not applicable for welding all aluminum-base alloys. Because of the formation of large amounts of brittle magnesium-silicide (Mg_2Si), the 4*xxx* series filler alloys are not applicable for welding the 7*xxx* series alloys containing appreciable amounts of magnesium. Table 5 shows the filler alloys recommended for welding various base alloys to minimize sensitivity to weld cracking.

TABLE 4 NOMINAL COMPOSITION OF ALUMINUM FILLER METAL ALLOYS

ALLOY	COMPOSITION, WT%						
	Si	Cu	Mn	Mg	Cr	Ti	OTHER
1100	. . .	0.12	. . .	. . .	. . .	. . .	99.00 AL MIN
1188	. . .	. . .	. . .	. . .	. . .	. . .	99.88 AL MIN
2319	. . .	6.3	0.30	. . .	. . .	0.15	0.17 ZR, 0.10 V
4043	5.2	. . .	. . .	. . .	. . .	. . .	. . .
4047	12.0	. . .	. . .	. . .	. . .	. . .	. . .
4145	10.0	4.0	. . .	. . .	. . .	. . .	. . .
4643	4.1	. . .	. . .	0.2	. . .	. . .	. . .
5183	. . .	. . .	0.75	4.7	0.15	. . .	. . .
5356	. . .	. . .	0.12	5.0	0.12	0.13	. . .
5554	. . .	. . .	0.75	2.7	0.12	0.12	. . .
5556	. . .	. . .	0.75	5.1	0.12	0.12	. . .
5654	. . .	. . .	. . .	3.5	0.25	0.10	. . .
C355.0	5.0	1.25	. . .	0.5	. . .	. . .	. . .
A356.0	7.0	. . .	. . .	0.4	. . .	. . .	. . .
357.0	7.0	. . .	. . .	0.5	. . .	. . .	. . .
A357.0	7.0	. . .	. . .	0.6	. . .	0.12	. . .

Source: Ref 42

TABLE 5 FILLER ALLOYS RECOMMENDED FOR WELDING SELECTED ALUMINUM-BASE ALLOYS TO MINIMIZE SENSITIVITY TO WELD CRACKING

BASE METAL	**319.0, 333.0, 354.0, 355.0, C355.0, 380.0**	**356.0, A356.0, A357.0, 359.0, 413.0, A444,0, 443.0**	**511.0, 512.0, 513.0, 514.0**	**7005$^{(A)}$, 7039, 710.0, 711.0, 712.0**	**6070**	**6061, 6463, 6101, 6201, 6151, 6351, 6951**	**5456**	**5454**	**5154, 5254$^{(B)}$**	**5086**	**5083**	**5052, 5652$^{(B)}$**	**5005 5050**	**3004 ALCLAD 3004**	**2219, 2519**	**2014, 2036**	**1100, 1060, 3004, ALCLAD 3003**	**1060 1070, 1080, 1350**
1060, 1070, 1080, 1350	4145$^{(C)(D)}$	4043$^{(D)(C)}$	4043$^{(D)(F)}$	4043$^{(D)}$	4043$^{(D)}$	4043$^{(D)}$	5356$^{(C)}$	4043$^{(D)}$	4043$^{(D)(F)}$	5356$^{(C)}$	5356$^{(C)}$	4043$^{(D)}$	1100$^{(C)}$	4043	4145	4145	1100$^{(C)}$	1188$^{(G)}$
1100, 3003, ALCLAD 3003	4145$^{(C)(D)}$	4043$^{(D)(E)}$	4043$^{(D)(F)}$	4043$^{(D)}$	4043$^{(D)}$	4043$^{(D)}$	5356$^{(C)}$	4043$^{(D)(F)}$	4043$^{(D)(F)}$	5356$^{(C)}$	5356$^{(C)}$	4043$^{(D)(F)}$	4043$^{(F)}$	4043$^{(F)}$	4145	4145	1100$^{(C)}$	. . .
2014, 2036	4145$^{(H)}$	4145	. . .	. . .	4145	4145	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	4145$^{(H)}$	4145$^{(H)}$	. . .	. . .
2219, 2519	4145$^{(C)(D)(H)}$	4145$^{(C)(D)}$	4043$^{(D)}$	4043$^{(D)}$	4043$^{(D)(E)}$	4043$^{(D)(E)}$	4043	4043$^{(D)}$	4043$^{(D)}$	4043	4043	4043$^{(D)}$	4043	4043	2319$^{(C)(D)(E)}$	. . .	. . .	. . .
3004, ALCLAD 3004	4043$^{(D)}$	4043$^{(D)}$	5654$^{(I)}$	5356$^{(F)}$	4043$^{(F)}$	4043$^{(I)}$	5356$^{(F)}$	5654$^{(I)}$	5654$^{(I)}$	5356$^{(F)}$	5356$^{(F)}$	4043$^{(D)(F)}$	4043$^{(F)}$	4043$^{(F)}$	. . .	. . .	. . .	. . .
5005, 5050	4043$^{(D)}$	4043$^{(D)}$	5654$^{(I)}$	5356$^{(F)}$	4043$^{(F)}$	4043$^{(I)}$	5356$^{(F)}$	5654$^{(I)}$	5654$^{(I)}$	5356$^{(F)}$	5356$^{(F)}$	4043$^{(D)(F)}$	4043$^{(F)(J)}$	. . .	. . .	. . .	. . .	. . .
5052, 5652$^{(B)}$	4043$^{(D)}$	4043$^{(D)(I)}$	5654$^{(I)}$	5356$^{(F)}$	5356$^{(C)(I)}$	5356$^{(C)(I)}$	5356$^{(I)}$	5654$^{(I)}$	5654$^{(I)}$	5356$^{(F)}$	5356$^{(F)}$	5654$^{(B)(C)(I)}$	. . .	. . .	. . .	. . .	. . .	. . .
5083		5356$^{(C)(D)(F)}$	5356$^{(F)}$	5183$^{(F)}$	5356$^{(F)}$	5356$^{(F)}$	5183$^{(F)}$	5356$^{(F)}$	5356$^{(F)}$	5356$^{(F)}$	5183$^{(F)}$	. . .	. . .	. . .	. . .	. . .	. . .	. . .
5086		5356$^{(C)(D)(F)}$	5356$^{(F)}$	5356$^{(F)}$	5356$^{(F)}$	5356$^{(F)}$	5356$^{(F)}$	5356$^{(I)}$	5356$^{(I)}$	5356$^{(F)}$	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
5154, 5254$^{(B)}$		043$^{(D)(I)}$	5654$^{(I)}$	5356$^{(I)}$	5356$^{(C)(I)}$	5356$^{(C)(I)}$	5356$^{(I)}$	5654$^{(I)}$	5354$^{(B)(I)}$	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
5454	4043$^{(D)}$	4043$^{(D)(I)}$	5654$^{(I)}$	5356$^{(I)}$	5356$^{(C)(I)}$	5356$^{(C)(I)}$	5356$^{(I)}$	5554$^{(C)(F)}$	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
5456		4043$^{(D)(I)}$	5356$^{(F)}$	5556$^{(F)}$	5356$^{(F)}$	5356$^{(F)}$	5556$^{(F)}$	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
6061,6063, 6351, 6101, 6201, 6151, 6951	4145$^{(C)(D)}$	5356$^{(C)(D)(F)}$	5356$^{(C)(I)}$	5356$^{(C)(D)(I)}$	4043$^{(D)(I)}$	4043$^{(D)(I)}$	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
6070	4145$^{(C)(D)}$	4043$^{(D)(F)}$	5356$^{(C)}$	5356$^{(C)(D)}$	4043$^{(D)}$	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .

			$^{(F)}$	$^{(F)}$	$^{(F)}$													
7005$^{(A)}$, 7039, 710.0, 711.0, 712.0	4043$^{(D)}$	4043$^{(D)(I)}$	5356$^{(I)}$	5356$^{(F)}$	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
511.0, 512.0, 513.0, 514.0		4043$^{(D)(I)}$	5654$^{(I)(J)}$	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
356.0, A356.0, A357.0, 359.0, 413.0, A444.0, 443.0	4145$^{(C)(D)}$	4043$^{(D)(J)}$	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
319.0, 333.0, 354.0, 355.0, C355.0 380.0	4145$^{(C)(D)(J)}$	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .

Notes: (1) Service conditions such as immersion in fresh or salt water, exposure to specific chemicals, or a sustained high temperature (higher than 65 °C, or 150 °F) may limit choice of filler metals Filler alloys 5356, 5183, 5556, and 5654 are not recommended for sustained elevated-temperature service. (2) Recommendations in this table apply to gas-shielded arc welding processes. For gas welding, only 1100, 1188, and 4043 filler metals are ordinarily used. (3) All filler metals are listed in AWS specification A5.10. (4) Where no filler metal is listed, the base metal combination is not recommended for welding.

Source: Ref 43

(A) REFERS TO 7005 EXTRUSIONS ONLY (X-PREFIX STILL APPLIES TO SHEET AND PLATE).

(B) BASE METAL ALLOYS 5652 AND 5654 ARE USED FOR HYDROGEN PEROXIDE SERVICE. 5654 FILLER METAL IS USED FOR WELDING BETH ALLOYS FOR LOW-TEMPERATURE SERVICE (65 °C, OR 150 °F, OR BELOW).

(C) 4043 MAY BE USED FOR SOME APPLICATIONS.

(D) 4047 MAY BE USED FOR SOME APPLICATIONS.

(E) 4145 MAY BE USED FOR SOME APPLICATIONS.

(F) 5183, 5356, OR 5556 MAY BE USED.

(G) 1100 MAY BE USED FOR SOME APPLICATIONS.

(H) 2319 MAY BE USED FOR SOME APPLICATIONS.

(I) 5183, 5356, 5554, 5556. OR 5654 MAY BE USED. IN SOME CASES THEY PROVIDE IMPROVED COLOR MATCH AFTER ANODIZING TREATMENT, HIGHEST WELD QUALITY, AND HIGHER WELD STRENGTH 5554 IS SUITABLE FOR ELEVATED TEMPERATURE SERVICE.

(J) FILLER METAL WITH THE SAME ANALYSIS AS THE BASE METAL IS SOMETIMES USED.

Weld Strength and Ductility

The strength and ductility of welds in aluminum alloys are determined by a number of parameters. These may include base alloy and temper, filler alloy, welding process and parameters, use of postweld heat treatments, joint type, principal loading condition, and the presence of weld defects. As described earlier, metallurgical transformations in the HAZ of heat-treatable aluminum alloys result in a degradation of strength in this region. The degree of degradation may be related to heat input used to produce the weld. Processes that provide higher travel speeds minimize the degradation in the HAZ and typically result in higher weld strength. Postweld heat treatments may be used to recover the loss of strength in the HAZ. Postweld solution heat treating and aging provides the greatest improvement in joint strength; however, this practice requires the use of a cold-water quench that may impart unacceptable part distortion. Postweld aging, which is a lower-temperature treatment, results in a moderate improvement in joint strength and does not require water quenching. It is also appropriate to comment on the effects of loading that distinguish the properties of welded joints. The transverse tensile strength of a butt weld provides a relative indication of joint strength in service, although butt welds may also undergo multiaxial loading that hinders direct comparison to tabular data representing uniaxial behavior. In the case of fillet welds, the transverse or longitudinal shear strength (depending on the principal loading condition) obtained from shear test specimens may be used to ascertain joint performance. The strength and ductility of groove welds in the as-welded condition, the postweld aged only condition, and the solution heat treated and aged (artificially and naturally aged) condition are described in the article "Welding of Aluminum Alloys" in this Volume. Transverse and longitudinal shear strengths of fillet welds are also listed in the aforementioned article. Additional information on properties of aluminum weldments can be found in Ref 44 and 45.

Figure 12 displays strength profiles across gas-tungsten arc welds of 2219-T87 and 6061-T6 alloys. The 2219-T87 alloy was welded with 2319 filler alloy and the 6061-T6 was welded with 4043 alloy. The strength profile of the work-hardenable alloy 5456-H116 welded with filler alloy 5356 is also shown in the figure for comparison. The strength values in Fig. 12 are approximations obtained by converting hardness measurements using the empirically derived relationship for aluminum alloys (Ref 46):

$$UTS = 10,66 - 19,42 \ln\left(1 - \frac{HRE}{109}\right) \quad \textbf{(EQ 2)}$$

where UTS is the ultimate tensile strength in ksi, and HRE is the Rockwell hardness using the E-scale.

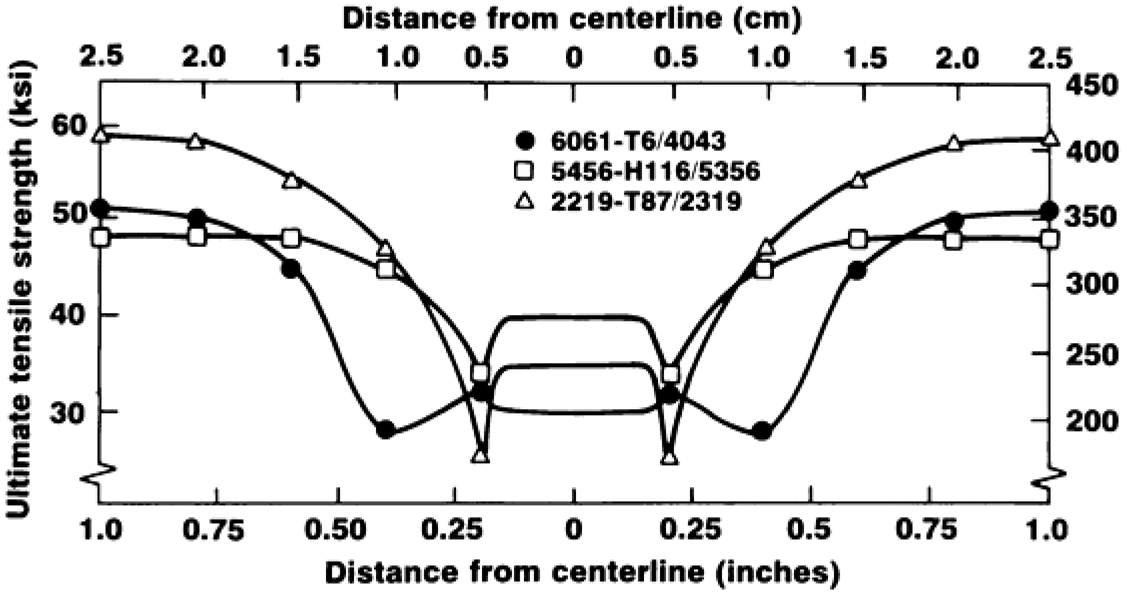

FIG. 12 STRENGTH PROFILES ACROSS ALTERNATING CURRENT GAS-TUNGSTEN ARC WELDS ON 3.2 MM (0.125 IN.) THICK 2219-T87 USING 2319 FILLER ALLOY, 6061-T6 USING 4043 FILLER ALLOY, AND 5456-H116 USING 5356 FILLER ALLOY. SOURCE: REF 35

It is evident from Fig. 12 that the degradation of strength associated with the HAZ usually dictates the strength of the joint for aluminum alloys. This is not true when weldments have been postweld heat treated, because the fusion zone may not respond to these treatments to the same degree as the HAZ.

The 2*xxx* series alloys exhibit the lowest ductility of the precipitation-hardenable alloys. This is due to the formation of brittle, copper-bearing eutectic phases at grain boundaries near the fusion zone interface. The 6*xxx* and 7*xxx* series alloys display good weld ductility. Postweld solution heat treating and aging of the heat-treat-able alloys or postweld aging only, usually decreases weld ductility.

Corrosion Resistance of Welds

Many of the heat-treatable aluminum alloys can be welded without reducing the resistance of the assembly to corrosion. However, because welding generates residual tensile stresses and modifies the parent metal structure, some alloys may have their resistance to corrosion lowered by welding. Corrosion of aluminum weldments is usually localized or preferential in nature and depends on the base alloy, filler alloy, and HAZ structure. The corrosion potential variations determined by electrode potential measurements across 2219 and 7039 weldments are shown in Fig. 13. The corrosion potential variations across the weld of the work-hardenable alloy 5456 are also shown in Fig. 13. The corrosion potentials of various base and filler alloys are also described in the article "Welding of Aluminum Alloys" in this Volume.

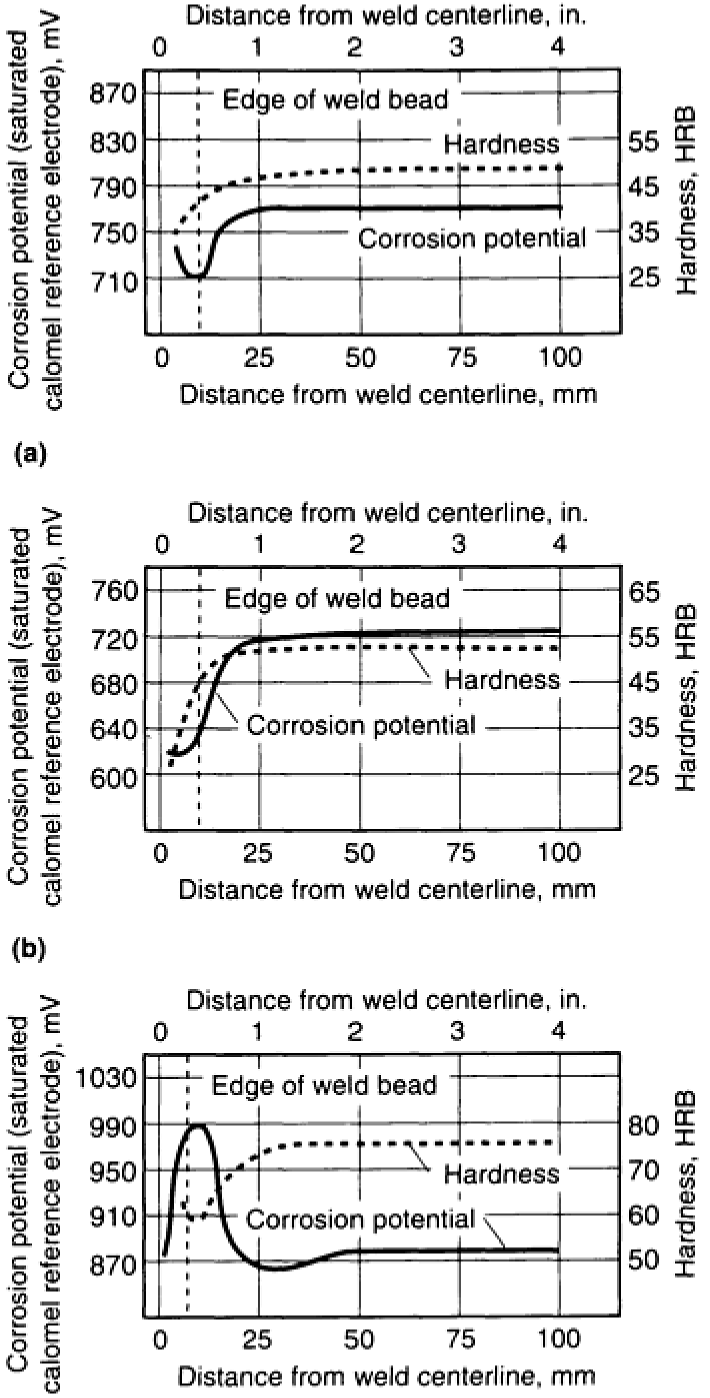

FIG. 13 PLOT OF CORROSION POTENTIAL AND HARDNESS LEVELS VERSUS DISTANCE FROM WELD CENTERLINE FOR SELECTED ALUMINUM ALLOYS. (A) ALLOY 5456-H321 (5556 FILLER METAL; DCRP-MIG WELD PRODUCED IN THREE PASSES). (B) 2219-T87 (2319 FILLER METAL; DCSP-TIG WELD PRODUCED IN TWO PASSES). (C) 7039-T651 (5183 FILLER METAL; DCRP-TIG WELD PRODUCED IN TWO PASSES). DCRP, DIRECT CURRENT REVERSE POLARITY; DCSP, DIRECT CURRENT STRAIGHT POLARITY. COURTESY OF ALCOA LABORATORIES

The 6*xxx* series alloys are highly resistant to localized corrosion after welding. The weldable alloy 2219 of the 2*xxx* series also exhibits good corrosion resistance after welding. The corrosion potential for 2219-T87 weldments produced using 2319 filler metal is shown in Fig. 13 and depicts a lower corrosion potential in the weld area as compared to the unaffected base metal. The weldable 7*xxx* series alloys produce greater sensitivity to corrosion after welding (Ref 47). In this case, the structural transformation of the HAZ due to the heat of welding makes the HAZ anodic to either the 5*xxx* weld metal or the unaffected base metal. This mechanism is also shown in Fig. 13 for 7089-T651 welded with 5183 filler metal. The highly anodic, small area of the HAZ will pit and corrode in the presence of an electrolyte to protect both the weld metal and the much larger unaffected base metal. Postweld solution heat treating and aging of the 7*xxx* series alloys will improve their corrosion resistance; however, aluminum-magnesium filler alloys containing more than 3.5% Mg should not be used when postweld heat treating because the fusion zone can be sensitized to stress-corrosion cracking (Ref 48).

Wrought alloys usually have greater resistance to stress-corrosion cracking in the longitudinal orientation (direction of working) than in the transverse orientation or in the short-transverse orientation (through the thickness). Because of this, welding of the 7*xxx* series alloys near a base-metal edge may result in a tensile stress in the short-transverse direction sufficient to cause stress-corrosion cracking in the exposed edge. "Buttering" the edge with weld metal provides compressive stress at the edge and overcomes the stress-corrosion cracking problem. Finally, paint protection or metallizing with selected aluminum alloys can be used to reduce the general corrosion sensitivity of heat-treated aluminum weldments.

Other Joint Characteristics

Other joint characteristics may be of importance, depending upon the service requirements of the welded assembly. These may include characteristics that can affect the structural performance of the assembly, such as strength and toughness at the service temperature, as well as aesthetic qualities, such as joint color match after anodizing. Aluminum alloys are compatible with all known cryogenics; base-metal strength and weld strength generally increase at cryogenic temperatures (Ref 49 and 50). The 2*xxx* series heat-treatable alloys, as well as the work-hardenable 5*xxx* series alloys, display a good combination of strength and ductility at low temperatures. Alloy 2219 is used extensively for cryogenic applications requiring high strength and good weldability. Toughness of welds in aluminum alloys is defined as the ability of the joint to absorb energy by deformation rather than fracture catastrophically. Fracture toughness in aluminum alloy welds is typically not determined by Charpy impact tests but may be reasonably obtained by tear or tension tests involving notched specimens (Ref 51, 52, 53). Based on results of these tests, welds of 2219 and 7039 possess good notch toughness, although it must also be mentioned that welds representing the work-hardenable alloys of the 5*xxx* series also exhibit good toughness.

Although the weld and base-metal strengths of the heat-treatable aluminum alloys, particularly the 2*xxx* series, fare well at elevated temperatures up to 150 °C (300 °F), caution must be observed when selecting filler alloys for welded assemblies functioning at these temperatures. Elevated-temperature service for aluminum welds is usually defined as service at sustained temperatures above 65 °C (150 °F). Sustained elevated-temperature service of welds containing high amounts of magnesium may result in increased sensitivity to stress-corrosion cracking due to the formation of Mg_2Al_3 at grain boundaries. Therefore, most welded applications involving sustained elevated-temperature service dictate the use of filler alloys that do not contain magnesium. An exception is the use of filler alloy 5554 (containing approximately 2.7% Mg) when welds of 6*xxx* series alloys at sustained elevated temperatures also require good color match after anodizing. Typically, 2319 filler alloy is used in welds on 2*xxx* series base alloys that require good color match after anodizing. When good color match after anodizing is desired for welds in the 6*xxx* or 7*xxx* series base alloys, filler alloys of the 5*xxx* series containing magnesium are utilized. High levels of silicon in the weld when using the 4*xxx* series filler alloys result in darkening of the fusion zone in comparison to the base metal after anodizing.

References cited in this section

35. R.P. MARTUKANITZ, ALCOA LABORATORIES, UNPUBLISHED RESEARCH, 1985
41. R.J. NAPOLITANO, H. KULLUK, M.L. NAGURKA, R.P. MARTUKANITZ, AND P.B. DICKERSON, DEVELOPMENT OF KNOWLEDGE-BASED SYSTEMS FOR ALUMINUM WELDING, *WELD. J.*, VOL 71 (NO. 12), 1992, P 43
42. P.B. DICKERSON AND B. IRVING, WELDING ALUMINUM: IT'S NOT AS DIFFICULT AS IT SOUNDS, *WELD. J.*, VOL 71 (NO. 4), 1992, P 47
43. *WELDING ALUMINUM: THEORY AND PRACTICE,* H.L. SAUNDERS, ED., THE ALUMINUM ASSOCIATION, 1991, P 3.8
44. *WELDING ALUMINUM,* AMERICAN WELDING SOCIETY, 1972, P 69.94-69.106
45. F.G. NELSON AND R.L. ROLF, SHEAR STRENGTH OF AL ALLOY FILLET WELDS, *WELD. J.*,

VOL 45 (NO. 2), 1966, P 82-S

46. B.S. SHABEL AND R.F. YOUNG, *PROC. CONF. NONDESTRUCTIVE CHARACTERIZATION OF MATERIALS II,* J.F. BUSSIERE, J.P. MONCHALIN, C.O. RUUD, AND R.E. GREEN, JR., ED., PLENUM, 1987, P 335
47. R.P. MEISTER AND D.C. MARTIN, "WELDING OF ALUMINUM AND ALUMINUM ALLOYS," DMIC REPORT 236, DEFENSE METALS INFORMATION CENTER, 1967, P 65
48. "ALCOA ALUMINUM ALLOY 7005 GREEN LETTER," ALUMINUM COMPANY OF AMERICA, 1974
49. R.P. MEISTER AND D.C. MARTIN, "WELDING OF ALUMINUM AND ALUMINUM ALLOYS," DMIC REPORT 236, DEFENSE METALS INFORMATION CENTER, 1967, P 55
50. J.F. CAMPBELL, ALUMINUM ALLOYS FOR CRYOGENIC SERVICE, *MATER. RES. STAND.,* VOL 4 (NO. 10), 1964, P 540
51. F.G. NELSON AND J.G. KAUFMAN, FRACTURE CHARACTERISTICS OF WELDS IN ALUMINUM ALLOYS, *WELD. J.,* VOL 45 (NO. 7), 1966, P 321-S
52. ALUMINUM FOR CRYOGENIC ENVIRONMENT, PART I: GENERAL CONSIDERATIONS, *CRYOG. ENG. NEWS,* BUSINESS COMMUNICATIONS PUBLISHERS, MARCH 1966
53. ALUMINUM FOR CRYOGENIC ENVIRONMENT, PART II: ALLOY GROUPS, *CRYOG. ENG. NEWS,* BUSINESS COMMUNICATIONS PUBLISHERS, APRIL 1966

Selection and Weldability of Heat-Treatable Aluminum Alloys

Richard P. Martukanitz, Applied Research Laboratory, Pennsylvania State University

---

References

1. P.B. DICKERSON AND B. IRVING, WELDING ALUMINUM: IT'S NOT AS DIFFICULT AS IT SOUNDS, *WELD. J.,* VOL 71 (NO. 4), 1992, P 45
2. *ALUMINUM: PROPERTIES AND PHYSICAL METALLURGY,* J.E. HATCH, ED., AMERICAN SOCIETY FOR METALS, 1984, P 222
3. *ALUMINUM STANDARDS AND DATA,* THE ALUMINUM ASSOCIATION, INC., 1988, P 15
4. *ALUMINUM: PROPERTIES AND PHYSICAL METALLURGY,* J.E. HATCH, ED., AMERICAN SOCIETY FOR METALS, 1984, P 320
5. R.S. JAMES, ALUMINUM-LITHIUM ALLOYS, VOL 2, *ASM HANDBOOK,* ASM INTERNATIONAL, 1990, P 178-199
6. *ALUMINUM STANDARDS AND DATA,* THE ALUMINUM ASSOCIATION, INC., 1988, P 11
7. J.M. SILCOCK, T.J. HEAL, AND H.K. HARDY, STRUCTURAL AGEING CHARACTERISTICS OF BINARY ALUMINUM-COPPER ALLOYS, *J. INST. MET.,* VOL 82, 1953-1954, P 239
8. C.E. CROSS, W.T. TACK, L.W. LOECHEL, AND L.S. KRAMER, ALUMINUM WELDABILITY AND HOT TEARING THEORY, *WELDABILITY OF MATERIALS,* ASM INTERNATIONAL, 1990, P 275
9. R.A. CHIHOSKI, THE CHARACTER OF STRESS FIELDS AROUND A WELD ARC MOVING ON ALUMINUM SHEET, *WELD J.,* VOL 51 (NO. 1), 1972, P 9-S
10. C.L. CROSS, "WELDABILITY OF ALUMINUM-LITHIUM ALLOYS: AN INVESTIGATION OF HOT TEARING MECHANISMS," PH.D. THESIS, COLORADO SCHOOL OF MINES, GOLDEN, CO, 1986, P 144
11. A.R.E. SINGER AND P.H. JENNINGS, HOT SHORTNESS OF THE ALUMINUM SILICON ALLOYS OF COMMERCIAL PURITY, *J. INST. MET.,* VOL 73, 1947, P 197
12. W.I. PUMPHREY AND J.V. LYONS, CRACKING DURING THE CASTING AND WELDING OF THE MORE COMMON BINARY ALUMINUM ALLOYS, *J. INST. MET.,* VOL 74, 1948, P 439

13. J.D. DOWD, WELD CRACKING OF ALUMINUM ALLOYS, *WELD. J.,* VOL 31 (NO. 10), 1952, P 448-S
14. P.H. JENNINGS, A.R.E. SINGER, AND W.I. PUMPHREY, HOT-SHORTNESS OF SOME HIGH-PURITY ALLOYS IN THE SYSTEMS ALUMINUM-COPPER-SILICON AND ALUMINUM-MAGNESIUM-SILICON, *J. INST. MET.,* 1948, P 227
15. J.H. DUDAS AND F.R. COLLINS, PREVENTING CRACKS IN HIGH-STRENGTH ALUMINUM ALLOYS, *WELD. J.,* VOL 45 (NO. 6), 1966, P 241-S
16. W.I. PUMPHREY AND D.C. MOORE, CRACKING DURING AND AFTER SOLIDIFICATION IN SOME ALUMINUM-COPPER-MAGNESIUM ALLOYS OF HIGH PURITY, *J. INST. MET.,* VOL 74, 1948, P 425
17. J.K. DAWSON, WELD CRACK SENSITIVITY OF ALUMINUM ALLOYS, *MET. PROG.,* VOL 76 (NO. 1), 1959, P 116
18. R.P. MEISTER AND D.C. MARTIN, "WELDING OF ALUMINUM AND ALUMINUM ALLOYS," DEFENSE METALS INFORMATION CENTER, BATTELLE MEMORIAL INSTITUTE, 1967, P 51
19. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987, P 239
20. R.P. MARTUKANITZ, ALCOA LABORATORIES, UNPUBLISHED RESEARCH, 1985
21. N.F. GITTOS AND M.H. SCOTT, HEAT AFFECTED ZONE CRACKING OF AL-MG-SI ALLOYS, *WELD. J.,* VOL 60 (NO. 6), 1981, P 95-S
22. H.W. KERR AND M. KATOH, INVESTIGATION OF HEAT-AFFECTED ZONE CRACKING OF GMA WELDS OF AL-MG-SI ALLOYS USING THE VARESTRAINT TEST, *WELD. J.,* VOL 66 (NO. 9), 1987, P 251-S
23. D.G. HOWDEN AND D.R. MILNER, *BRIT. WELD. J.,* VOL 10, 1963, P 304
24. Z.P. SAPERSTEIN, G.R. PRESCOTT, AND E.W. MONROE, POROSITY IN ALUMINUM WELDS, *WELD. J.,* VOL 43 (NO. 10), 1964, P 443-S
25. C. RANSLEY AND J. NEUFELD, *J. INST. MET.,* VOL 74, 1948, P 617
26. M. UDA AND S. OHNO, *TRANS. NATL. RES. INST. MET.* (JPN), VOL 16, 1974, P 2
27. R.P. MARTUKANITZ, ALCOA LABORATORIES, UNPUBLISHED RESEARCH, 1986
28. R.P. MARTUKANITZ AND P.R. MICHNUK, SOURCES OF POROSITY IN GAS METAL ARC WELDING OF ALUMINUM, *ALUMINIUM,* VOL 58 (NO. 5), 1982, P 276
29. P.B. DICKERSON, ALCOA LABORATORIES, PRIVATE COMMUNICATION, 1987
30. A.N. O'SHANSKII AND V.V. D'YACHENCHO, EVALUATION OF THE SUSCEPTIBILITY OF ALLOYS TO THE FORMATION OF PORES IN WELDING, *WELD. PROD.,* VOL 24 (NO. 7), 1977, P 52
31. S.D. DUMOLT, "METALLURGICAL TRANSACTIONS IN THE HEAT AFFECTED ZONE OF ALUMINUM ALLOYS BY TRANSMISSION ELECTRON MICROSCOPY," PH.D. THESIS, CARNEGIE-MELLON UNIVERSITY, PITTSBURGH, PA, 1983
32. T. ENGO AND T. KURODA, MICROSTRUCTURE IN WELD HEAT-AFFECTED ZONE OF AL-MG-SI ALLOYS, *TRANS. JPN. WELD. RES. INST.,* VOL 11 (NO. 1), 1982, P 61
33. R.P. MARTUKANITZ, "MODELING THE HEAT AFFECTED ZONES OF ALUMINUM ARC WELDS, *ADVANCES IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID, ED., ASM INTERNATIONAL, 1986, P 193
34. S. KOU, WELDING METALLURGY AND WELDABILITY OF HIGH STRENGTH ALUMINUM ALLOYS, *WELD. RES. COUNC. BULL.,* NO. 320, WELDING RESEARCH COUNCIL, 1986
35. R.P. MARTUKANITZ, ALCOA LABORATORIES, UNPUBLISHED RESEARCH, 1985
36. S.D. DUMOLT, PH.D. THESIS, CARNEGIE-MELLON UNIVERSITY, PITTSBURGH, PA, 1983
37. W.L. BURCH, THE EFFECT OF WELDING SPEED ON STRENGTH OF 6061-T4 ALUMINUM JOINTS, *WELD. J.,* VOL 37 (NO. 7), 1958, P 361-S
38. R.P. MARTUKANITZ, "MODELING THE HEAT AFFECTED ZONES OF ALUMINUM ARC WELDS, ADVANCES IN WELDING SCIENCE AND TECHNOLOGY, S.A. DAVID, ED., ASM

INTERNATIONAL, 1986, P 195

39. R.J. BRUNGRABER AND F.G. NELSON, EFFECT OF WELDING VARIABLES ON ALUMINUM ALLOY WELDMENTS, *WELD. J.,* VOL 52 (NO. 3), 1973, P 97-S
40. G.E. METZGER, SOME MECHANICAL PROPERTIES OF WELDS IN 6061 ALUMINUM ALLOY SHEET, *WELD. J.,* VOL 56 (NO. 10), 1967, P 457-S
41. R.J. NAPOLITANO, H. KULLUK, M.L. NAGURKA, R.P. MARTUKANITZ, AND P.B. DICKERSON, DEVELOPMENT OF KNOWLEDGE-BASED SYSTEMS FOR ALUMINUM WELDING, *WELD. J.,* VOL 71 (NO. 12), 1992, P 43
42. P.B. DICKERSON AND B. IRVING, WELDING ALUMINUM: IT'S NOT AS DIFFICULT AS IT SOUNDS, *WELD. J.,* VOL 71 (NO. 4), 1992, P 47
43. *WELDING ALUMINUM: THEORY AND PRACTICE,* H.L. SAUNDERS, ED., THE ALUMINUM ASSOCIATION, 1991, P 3.8
44. *WELDING ALUMINUM,* AMERICAN WELDING SOCIETY, 1972, P 69.94-69.106
45. F.G. NELSON AND R.L. ROLF, SHEAR STRENGTH OF AL ALLOY FILLET WELDS, *WELD. J.,* VOL 45 (NO. 2), 1966, P 82-S
46. B.S. SHABEL AND R.F. YOUNG, *PROC. CONF. NONDESTRUCTIVE CHARACTERIZATION OF MATERIALS II,* J.F. BUSSIERE, J.P. MONCHALIN, C.O. RUUD, AND R.E. GREEN, JR., ED., PLENUM, 1987, P 335
47. R.P. MEISTER AND D.C. MARTIN, "WELDING OF ALUMINUM AND ALUMINUM ALLOYS," DMIC REPORT 236, DEFENSE METALS INFORMATION CENTER, 1967, P 65
48. "ALCOA ALUMINUM ALLOY 7005 GREEN LETTER," ALUMINUM COMPANY OF AMERICA, 1974
49. R.P. MEISTER AND D.C. MARTIN, "WELDING OF ALUMINUM AND ALUMINUM ALLOYS," DMIC REPORT 236, DEFENSE METALS INFORMATION CENTER, 1967, P 55
50. J.F. CAMPBELL, ALUMINUM ALLOYS FOR CRYOGENIC SERVICE, *MATER. RES. STAND.,* VOL 4 (NO. 10), 1964, P 540
51. F.G. NELSON AND J.G. KAUFMAN, FRACTURE CHARACTERISTICS OF WELDS IN ALUMINUM ALLOYS, *WELD. J.,* VOL 45 (NO. 7), 1966, P 321-S
52. ALUMINUM FOR CRYOGENIC ENVIRONMENT, PART I: GENERAL CONSIDERATIONS, *CRYOG. ENG. NEWS,* BUSINESS COMMUNICATIONS PUBLISHERS, MARCH 1966
53. ALUMINUM FOR CRYOGENIC ENVIRONMENT, PART II: ALLOY GROUPS, *CRYOG. ENG. NEWS,* BUSINESS COMMUNICATIONS PUBLISHERS, APRIL 1966

Selection and Weldability of Non-Heat-Treatable Aluminum Alloys

C.E. Cross, Martin Marietta Astronautics Group; M.L. Kohn, FMC Corporation

Introduction

NON-HEAT-TREATABLE ALUMINUM ALLOYS constitute a group of alloys that rely solely upon cold work and solid solution strengthening for their strength properties. They differ from heat-treatable alloys in that they are incapable of forming second-phase precipitates for improved strength (see the article "Selection and Weldability of Heat-Treatable Aluminum Alloys" in this Volume). As such, the non-heat-treatable alloys cannot achieve the high strengths characteristic of precipitation-hardened alloys. For example, the yield strength of precipitation-hardened alloy 7075-T6 (505 MPa, or 73 ksi) is significantly greater than that of cold-worked and solution-strengthened alloy 5083-H32 (230 MPa, or 33 ksi).

The absence of precipitate-forming elements in these low- to moderate-strength non-heat-treatable alloys becomes a positive attribute when considering weldability, because many of the alloy additions needed for precipitation hardening (for example, copper plus magnesium, or magnesium plus silicon) can lead to liquation or hot cracking during welding. In addition, joint efficiencies are higher in non-heat-treatable alloys because the heat-affected zone (HAZ) is not compromised by the coarsening or dissolution of precipitates. This obviates the need for thick joint lands or postweld heat treatment and favors the use of welded structures in the as-welded condition.

Selection and Weldability of Non-Heat-Treatable Aluminum Alloys

C.E. Cross, Martin Marietta Astronautics Group; M.L. Kohn, FMC Corporation

Alloy Classification and Typical Applications

Non-heat-treatable wrought aluminum alloys can be placed into one of three groups using standard Aluminum Association designations:

ALLOY NUMBER	ALLOY ADDITION
1*XXX*	AL (99% MINIMUM PURITY)
3*XXX*	AL-MN
4*XXX*	AL-SI
5*XXX*	AL-MG

Nominal compositions for common non-heat-treatable alloys are given in Table 1.

TABLE 1 COMPOSITION OF WROUGHT NON-HEAT-TREATABLE ALUMINUM ALLOYS

ALLOY DESIGNATION	COMPOSITION, WT%(A)											
	Si	**Fe**	**Cu**	**Mn**	**Mg**	**Cr**	**Ni**	**Zn**	**Ti**	**OTHERS**		AL MIN.(D)
										EACH(B)	**TOTAL(C)**	
1050	0.25	0.40	0.05	0.05	0.05	. . .	. . .	0.05	0.03	0.03(E)	. . .	**99.50**
1060	0.25	0.35	0.05	0.03	0.03	. . .	. . .	0.05	0.03	0.03(E)	. . .	**99.60**

1100	0.95 (SI + FE)		0.05-0.20	0.05	. . .	. . .	. . .	0.10	. . .	0.05(F)	0.15	**99.00**
1145(G)	0.55 (SI + FE)		0.05	0.05	0.05	. . .	. . .	0.05	0.03	0.03(E)	. . .	**99.45**
1175(H)	0.15 (SI + FE)		0.10	0.02	0.02	. . .	. . .	0.04	0.02	0.02(I)	. . .	**99.75**
1188	0.00	0.06	0.005	0.01	0.01	. . .	. . .	0.03	0.01	. . .	. . .	**99.88**
1200	1.00 (SI + FE)		0.05	0.05	. . .	. . .	. . .	0.10	0.05	0.05	0.15	**99.00**
1230(H)	0.70 (SI + FE)		0.10	0.05	0.05	. . .	. . .	0.10	0.03	0.03(E)	. . .	**99.30**
1235	0.65 (SI + FE)		0.05	0.05	0.05	. . .	. . .	0.10	0.06	0.03(E)	. . .	**99.35**
1345	0.03	0.40	0.10	0.05	0.05	. . .	. . .	0.05	0.03	0.03(E)	. . .	**99.45**
1350(J)	0.10	0.40	0.05	0.01	. . .	0.01	. . .	0.05	. . .	0.03(K)	0.10	**99.50**
3003	0.6	0.7	0.05-0.20	1.0-1.5	. . .	. . .	. . .	0.10	. . .	0.05	0.15	**BAL**
3004	0.30	0.7	0.25	1.0-1.5	0.8-1.3	. . .	. . .	0.25	. . .	0.05	0.15	**BAL**
3005	0.6	0.7	0.30	1.0-1.5	0.20-0.6	0.10	. . .	0.25	0.10	0.05	0.15	**BAL**
3105	0.6	0.7	0.30	0.03-0.8	0.20-0.8	0.20	. . .	0.40	0.10	0.05	0.15	**BAL**
4032	11.0-13.5	1.0	0.50-1.3	. . .	0.8-1.3	0.10	0.50-1.3	0.25	. . .	0.05	0.15	**BAL**
4043	4.5-6.0	0.8	0.30	0.05	0.05	. . .	. . .	0.10	0.20	0.05(F)	0.15	**BAL**
4045(L)	9.0-11.0	0.8	0.30	0.05	0.05	. . .	. . .	0.10	0.20	0.05	0.15	**BAL**
4047(L)	11.0-13.0	0.8	0.30	0.15	0.10	. . .	. . .	0.20	. . .	0.05(F)	0.15	**BAL**
4145(L)	9.3-10.7	0.8	3.3-4.7	0.15	0.15	0.15	. . .	0.20	. . .	0.05(F)	0.15	**BAL**
4343(L)	6.8-8.2	0.8	0.25	0.10	. . .	. . .	. . .	0.20	. . .	0.05(F)	0.15	**BAL**
4643	3.6-4.6	0.8	0.10	0.05	0.10-0.30	. . .	. . .	0.10	0.15	0.05(F)	0.15	**BAL**
5005	0.30	0.7	0.20	0.20	0.50-1.1	0.10	. . .	0.25	. . .	0.05	0.15	**BAL**
5050	0.40	0.7	0.20	0.10	1.1-1.8	0.10	. . .	0.25	. . .	0.05	0.15	**BAL**
5052	0.25	0.40	0.10	0.10	2.2-2.8	0.15-0.35	. . .	0.10	. . .	0.05	0.15	**BAL**
5056	0.30	0.40	0.10	0.05-0.20	4.5-5.6	0.05-0.20	. . .	0.10	. . .	0.05	0.15	**BAL**

5083	0.40	0.40	0.10	0.40-1.0	4.0-4.9	0.05-0.25	. . .	0.25	0.15	0.05	0.15	**BAL**
5086	0.40	0.50	0.10	0.20-0.7	3.5-4.5	0.05-0.25	. . .	0.25	0.15	0.05	0.15	**BAL**
5154	0.25	0.40	0.10	0.10	3.1-3.9	0.15-0.35	. . .	0.20	0.20	0.05	0.15	**BAL**
5183	0.40	0.40	0.10	0.50-1.0	4.3-5.2	0.05-0.25	. . .	0.25	0.15	0.05(F)	0.15	**BAL**
5252	0.08	0.10	0.10	0.10	2.2-2.8	. . .	. . .	0.05	. . .	0.03(E)	0.10	**BAL**
5254	0.45 (SI + FE)		0.05	0.01	3.1-3.9	0.15-0.35	. . .	0.20	0.05	0.05	0.15	**BAL**
5356	0.25	0.40	0.10	0.05-0.20	4.5-5.5	0.05-0.20	. . .	0.10	0.06-0.20	0.05(F)	0.15	**BAL**
5454	0.25	0.40	0.10	0.50-1.0	2.4-3.0	0.05-0.20	. . .	0.25	0.20	0.05	0.15	**BAL**
5456	0.25	0.40	0.10	0.50-1.0	4.7-5.5	0.05-0.20	. . .	0.25	0.20	0.05	0.15	**BAL**
5457	0.08	0.10	0.20	0.15-0.45	0.8-1.2	. . .	. . .	0.05	. . .	0.03(E)	0.10	**BAL**
5554	0.25	0.40	0.10	0.50-1.0	2.4-3.0	0.05-0.20	. . .	0.25	0.05-0.20	0.05(F)	0.15	**BAL**
5556	0.25	0.40	0.10	0.50-1.0	4.7-5.5	0.05-0.20	. . .	0.25	0.05-0.20	0.05(F)	0.15	**BAL**
5652	0.40 (SI + FE)		0.04	0.01	2.2-2.8	0.15-0.35	. . .	0.10	. . .	0.05	0.15	**BAL**
5654	0.45 (SI + FE)		0.05	0.01	3.1-3.9	0.15-0.35	. . .	0.20	0.05-0.15	0.05(F)	0.15	**BAL**
5657	**0.08**	**0.10**	**0.10**	**0.03**	**0.6-1.0**	**. . .**	**. . .**	**0.05**	**. . .**	**0.02(I)**	**0.05**	**BAL**

Source: Ref 1

(A) COMPOSITION IN MAXIMUM WT% UNLESS SHOWN AS A RANGE OR A MINIMUM.

(B) A 0.0008 WT% BE MAXIMUM IS APPLICABLE TO ANY ALLOY TO BE USED AS WELDING ELECTRODE OR WELDING ROD.

(C) THE SUM OF THOSE "OTHERS" METALLIC ELEMENTS 0.010% OR MORE EACH, EXPRESSED TO THE SECOND DECIMAL BEFORE DETERMINING THE SUM.

(D) THE ALUMINUM CONTENT FOR UNALLOYED ALUMINUM NOT MADE BY A REFINING PROCESS IS THE DIFFERENCE BETWEEN 100.00% AND THE SUM OF ALL OTHER METALLIC

ELEMENTS PRESENT IN AMOUNTS OF 0.010% OR MORE EACH, EXPRESSED TO THE SECOND DECIMAL BEFORE DETERMINING THE SUM.

(E) 0.05% V MAXIMUM.

(F) 0.0008% BE MAX FOR WELDING ELECTRODE AND WELDING ROD ONLY.

(G) FOIL.

(H) CLADDING ALLOY.

(I) 0.03% GA MAXIMUM; 0.05% V MAXIMUM.

(J) ELECTRIC CONDUCTOR (EC).

(K) 0.02% (V + TI) MAXIMUM; 0.05% B MAX; 0.03% GA MAXIMUM.

(L) BRAZING ALLOY.

1*xxx* Alloys. The 1*xxx*-series alloys are of commercial purity (>99% Al) and are used where thermal/electrical conduction or corrosion resistance becomes paramount over strength in design considerations (for example, alloy 1100 is used for sheet metal work, fin stock and chemical equipment). Alloys with purity levels greater than 99.5% are used for electrical conductors (for example, alloy 1350).

3*xxx* Alloys. The 3*xxx*-series alloys are used in applications where added additional strength and formability are needed, in addition to maintaining excellent corrosion resistance (for example, alloy 3004 is used for sheet metal work, storage tanks, and beverage containers). Typical applications include cooking utensils, pressure vessels, and building products (siding, gutters, and so on). These alloys get their strength from cold work and fine $(Mn,Fe)Al_6$ dispersoids that pin grain and subgrain boundaries. There is also a small degree of solid solution strengthening from both manganese and magnesium.

4*xxx* Alloys. Apart from their use as welding filler material (see the section "Filler Alloy Selection" in this article), the 4*xxx*-series alloys have limited industrial application in wrought form.

5*xxx* Alloys. The 5*xxx*-series alloys are used in cases where still higher strengths are required; this strength is achieved from large quantities of magnesium in solid solution. More importantly, magnesium promotes work hardening by lowering the stacking fault energy, thus reducing the tendency for dynamic recovery. Applications for 5*xxx*-series alloys include automobile and appliance trim, pressure vessels, armor plate, and components for marine and cryogenic service. The ballistic analysis of alloy 5083 is given in the section "Weld Properties" of this article.

While these alloys normally exhibit good corrosion resistance, care must be taken during processing to avoid formation of continuous β-Mg_3Al_2 precipitates at grain boundaries, which can lead to intergranular corrosion. This can occur in heavily cold-worked, high-magnesium alloys exposed to temperatures from 120 to 200 °C (250 to 390 °F). Alloy 5454 possesses the highest magnesium content suitable for sustained elevated temperatures and has become the standard alloy used for truck bodies for hot oil or asphalt applications, and storage tanks for heated products.

Temper Designations. Appropriate heat treatments and temper designations for non-heat-treatable alloys include a solution anneal (O temper) and a cold-worked condition (H1 temper). When the cold work achieved during processing is more than desired, a partial anneal can be used to reduce strength (H2 temper). When dealing with 5*xxx*-series alloys, it is common to use a low-temperature heat treatment (H3 temper) to stabilize the microstructure and prevent intergranular corrosion. This treatment involves heating to 120 °C (250 °F), resulting in partial recovery and making properties less susceptible to change with time.

When working with H-type tempers, a second digit is used to indicate the degree of cold work:

RELATIVE HARDNESS	TEMPER DESIGNATION
FULL HARD	**H18**
$\frac{3}{4}$ HARD	**H16**

$\frac{1}{2}$ HARD	**H14**
$\frac{1}{4}$ **HARD**	**H12**

A more thorough and complete description of non-heat-treatable alloys and tempers is given in Volume 2 of the *ASM Handbook* (Ref 2).

References cited in this section

1. *ALUMINUM STANDARDS AND DATA*, ALUMINUM ASSOCIATION, 1984, P 93
2. PROPERTIES AND SELECTION NONFERROUS ALLOYS AND SPECIAL-PURPOSE MATERIALS, VOL 2, *ASM HANDBOOK*, ASM INTERNATIONAL, 1990, P 15-28

Selection and Weldability of Non-Heat-Treatable Aluminum Alloys

C.E. Cross, Martin Marietta Astronautics Group; M.L. Kohn, FMC Corporation

---

Filler Alloy Selection

Filler alloys used to join non-heat-treatable alloys can be selected from one of three alloy groups:

ALLOY NUMBER	ALLOY ADDITION
1*XXX*	**AL (99% MINIMUM PURITY)**
4*XXX*	**A1-SI**
5*XXX*	**AL-MG**

Commonly used filler alloys include 1100, 1188, 4043, 4047, 5554, 5654, 5183, 5356, and 5556. Nominal compositions for selected filler alloys are included in Table 1. Selecting the best filler alloy for a given application depends on the desired performance relative to weldability, strength, ductility, and corrosion resistance. In general, the filler alloy selected should be similar in composition to the base metal alloy. Thus, a 1*xxx* filler alloy is recommended for joining 1*xxx*- or 3*xxx*-series base metal alloys. Similarly, 5*xxx* filler alloys are used to join 5*xxx*-series base metal alloys. An exception to this rule is encountered when weldability becomes an issue. Weldability of non-heat-treatable aluminum alloys can be measured in terms of resistance to hot cracking and porosity.

Hot Cracking. Problems with hot cracking are encountered when welding under highly constrained conditions or when welding certain alloys that are highly susceptible to cracking. Such is the case when welding 5*xxx* alloys that have a low-range magnesium content ($\frac{1}{2}$ to $2\frac{1}{2}$ wt% Mg) where, according to the crack susceptibility curve (Fig. 1), a peak in susceptibility occurs. To avoid cracking, use of a high-magnesium filler alloy is recommended.

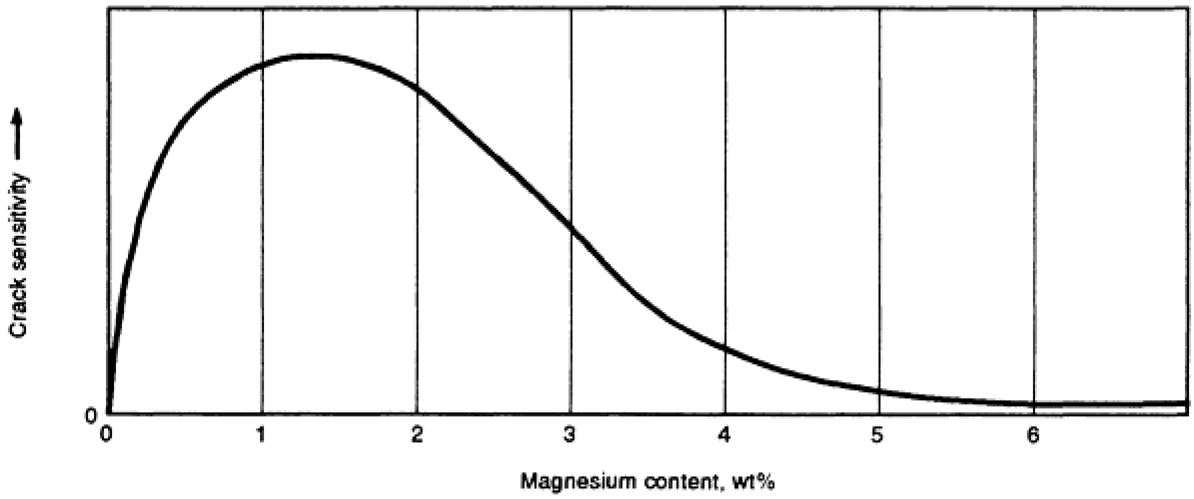

FIG. 1 PLOT OF HOT CRACKING SENSITIVITY VERSUS MAGNESIUM CONTENT TO SHOW PEAK IN SUSCEPTIBILITY OBTAINED FOR ALUMINUM-MAGNESIUM (5*XXX*) ALLOYS. SOURCE: REF 3

Similar problems may be encountered when 1*xxx* fillers are used to join 5*xxx* alloys (or vice versa), or when welding dissimilar metal alloys such as alloys 1100 and 5083, where mutual dilution may result in low magnesium levels. Electron-beam welding or laser-beam welding can also result in cracking when magnesium, a high-vapor-pressure alloying element, is boiled off. The problem is aggravated when welding in a vacuum environment. An example of a center-line hot crack produced when electron-beam welding alloy 5083 is shown in Fig. 2, where magnesium levels were measured to be approximately 3 wt% in the vicinity of the crack, based on microprobe analysis. This was an autogenous weld made in a hard vacuum (0.13 mPa, or 10^{-6} torr).

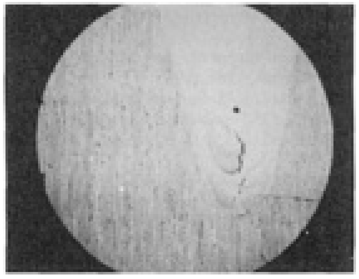

FIG. 2 CENTERLINE HOT CRACK PRODUCED IN AUTOGENOUS ELECTRON-BEAM WELDED ALLOY 5083

Another approach to be taken when hot cracking persists is to use 4*xxx* fillers. These aluminum-silicon alloys have exceptional resistance to cracking, due in part to their abundance of liquid eutectic available for back-filling (Ref 4). However, their use should be avoided when welding high-magnesium alloys (>3 wt%) because of embrittlement from excessive Mg_2Si precipitation. Other drawbacks include low joint ductility and nonmatching color when anodized. Weldability will improve with increased silicon content (for example, alloy 4047 versus alloy 4043), but the disadvantages discussed above will likewise become more predominant.

Porosity. Non-heat-treatable aluminum alloys are susceptible to hydrogen-induced weld metal porosity, as are all aluminum alloys in general. This porosity forms during solidification due to the abrupt drop in hydrogen solubility when going from liquid to solid. Porosity can best be avoided by minimizing hydrogen pickup during welding. This can be accomplished through proper joint preparation (that is, removal of hydrocarbons), use of high-grade (low-dew-point) shielding gas, and careful storage of filler wire (that is, protection from exposure to moisture and oil). It has been determined that welding filler wire is often the primary source of hydrogen contamination. The 5*xxx*-series filler alloys, in particular, are susceptible to the hydration of surface oxides, which can result in porosity (Ref 5).

In summary, guidelines for filler alloy selection when welding non-heat-treatable alloys are:

BASE METAL	FILLER METAL
1*XXX*	**1*XXX*, 4*XXX***
3*XXX*	**1*XXX*, 4*XXX***
5*XXX* (LOW MAGNESIUM)	**5*XXX*, 4*XXX***
5*XXX* (HIGH MAGNESIUM)	**5*XXX***

References cited in this section

3. J.H. DUDAS AND F.R. COLLINS, PREVENTING WELD CRACKS IN HIGH-STRENGTH ALUMINUM ALLOYS, *WELD. J.*, VOL 45, 1966, P 241S-249S
4. C.E. CROSS, L.S. KRAMER, W.T. TACK, AND L.W. LOECHEL, ALUMINUM WELDABILITY AND HOT TEARING THEORY, *WELDABILITY OF MATERIALS*, ASM INTERNATIONAL, 1990, P 275-282
5. R.P. MARTUKANITZ AND P.R. MICHNUK, SOURCES OF POROSITY IN GAS METAL ARC WELDING OF ALUMINUM, *TRENDS IN WELDING RESEARCH*, ASM INTERNATIONAL, 1982, P 315-330

Selection and Weldability of Non-Heat-Treatable Aluminum Alloys

C.E. Cross, Martin Marietta Astronautics Group; M.L. Kohn, FMC Corporation

Weld Properties

When non-heat-treatable alloys are welded, microstructural damage is incurred in the HAZ. Unlike the case of heat-treatable alloys, whose strengthening precipitates may dissolve or coarsen, the HAZ damage in non-heat-treatable alloys is limited to recovery, recrystallization, and grain growth. Thus, loss in strength in the HAZ is not nearly as severe as that experienced in heat-treatable alloys. (An example of this is shown in the hardness traverses of Fig. 8 in the article "Selection and Weldability of Heat-Treatable Aluminum Alloys" in this Volume, where alloy 5456 is compared against alloys 6061 and 2219.) For this reason, 5*xxx*-series alloys are popular for use in welded pressure vessels where reasonable joint strengths can be obtained in the as-welded condition without the need for post-weld heat treatment.

The weld metal of non-heat-treatable aluminum alloys is typically the weakest part of the joint and is the location of failure when the joint is loaded in tension. This is in contrast to most heat-treatable aluminum alloys, where the heat-affected zone often is the weakest link. The weld metal microstructure of the non-heat-treatable alloys consists of columnar, epitaxial grains with a cellular or columnar-dendritic substructure that has interdendritic eutectic constituents-primarily $(Fe,Mn)Al_6$ for 1*xxx* and 3*xxx* alloys; Si for 4*xxx* alloys; and Mg_3Al_2 for 5*xxx* alloys.

Joint efficiencies for several non-heat-treatable alloys are compared in Table 2. It should be noted that joint efficiency is highest for annealed plate (approaching 100%) and decreases continuously with the degree of cold working. Also shown in Table 2 is the effect of composition and cold work on plate properties. The two 5*xxx*-series alloys shown exhibit significant improvement in strength over 1*xxx*- and 3*xxx*-series alloys, a reflection on the strengthening ability of magnesium. Likewise, 5*xxx*-series alloys with higher magnesium content have higher strength (for example, alloy 5083 versus alloy 5050). This same trend is observed for weld metal strength. For example, a weld made on high-magnesium alloy 5083 (4% Mg) is much stronger than a similar weld made on alloy 5050 (1% Mg).

TABLE 2 EFFECT OF VARIATIONS IN TEMPER AND FILLER ALLOY ON NON-HEAT-TREATABLE ALUMINUM GAS-METAL ARC WELDMENT PROPERTIES

BASE ALLOY AND TEMPER	**PLATE PROPERTIES**					**FILLER ALLOY**	**WELD PROPERTIES**					JOINT EFFICIENCY, %(A)
	YIELD STRENGTH		**ULTIMATE TENSILE STRENGTH**		**ELONGATION IN 50 MM (2 IN.), %**		**YIELD STRENGTH**		**ULTIMATE TENSILE STRENGTH**		**ELONGATION IN 50 MM (2 IN.), %**	
	MPA	**KSI**	**MPA**	**KSI**			**MPA**	**KSI**	**MPA**	**KSI**		
1100-O	34	5	90	13	35	1100	41	6	93	13.5	23	**100**
						4043	41	6	90	13	21	**100**
3003-O	41	6	110	16	30	1100	41	6	90	13	20	**81**
						4043	48	7	91	13.2	17	**83**
3003-H18	186	27	200	29	4	1100	59	8.5	110	16	15	**55**
5050-O	55	8	145	21	24	5356	55	8	145	21	20	**100**
5050-H32	145	21	172	25	9	5356	97	14	159	23	14.5	**92**
						4043	90	13	152	22	15	**88**
5050-H38	200	29	221	32	6	5356	97	14	162	23.5	14.5	**73**
5083-O	145	21	290	42	22	5356	145	21	283	41	17	**98**
5083-H32	228	33	317	46	16	5356	145	21	276	40	16	**87**
						5183	**152**	**22**	**300**	**43.5**	**12**	**95**

Source: Ref 6

(A) JOINT EFFICIENCY = (UTS OF WELD)/(UTS OF PLATE), WHERE UTS STANDS FOR ULTIMATE TENSILE STRENGTH.

The correspondence between plate and joint strength does not apply to temper. For the high-magnesium alloy 5083 plate, no increase in joint strength is observed when welding an annealed versus a cold-worked temper (see Table 2). Likewise, filler metal selection has little effect on joint strength, as shown in Table 2. However, it is clear that use of a 4*xxx* filler does result in ductility loss. Also, filler alloys high in iron, silicon, or manganese (for example, filler alloy 5183) will have greater amounts of eutectic constituents, which will result in lower weld ductility. Even so, non-heat-treatable alloy welds exhibit significantly better ductility in comparison to similar heat-treatable welds.

Example 1: Use of Welded Alloy 5083 in Ballistic Applications.

An important application for heavy-gage alloy 5083 is the construction of tactical military vehicles. The hulls and turrets of Army vehicles such as the M113 armored personnel carrier, the M2/M3 infantry and cavalry fighting vehicles, the M109 self-propelled howitzer, and the Marine Corps AAV7A amphibians all consist of welded 5083 aluminum structures. There are also a myriad of brackets, clips, and so on, welded to the hulls and turrets, although not normally fabricated to ballistic requirements.

Aluminum alloys considered for use as armor include alloys 5083, 7039, and 2519. Alloys 7039 and 2519 are weldable and heat-treatable, capable of yield strengths exceeding 330 MPa (48 ksi). However, alloy 2519 weldments have failed rigorous ballistic shock tests (MIL-STD-1946, discussed below), even though the base material exhibits excellent ballistic and mechanical properties. Alloy 7039 possesses superior ballistic properties compared to alloy 5083, but it suffers from a tendency to delaminate and to exhibit stress-corrosion cracking. Alloy 5083 weldments have passed ballistic weld tests, and they possess the necessary resistance to corrosion.

In the MIL-STD-1946 ballistic test, a weldment is subjected to impact by a blunt-nose 75 mm (3 in.) aluminum projectile, fired at a velocity specified according to material thickness and alloy composition. Butt-welded specimens are struck no more than 13 mm ($\frac{1}{2}$ in.) from the weld centerline. Corner joints are struck so that the center of the impact projectile is within 50 ± 13 mm (2 ± $\frac{1}{2}$ in.) from the weld toe on its back side. Welds are considered acceptable when the total length of cracking generated is less than 305 mm (12 in.).

An offset double-V configuration of the butt joint is commonly used in ballistic applications (Fig. 3). This results in solid metal backing for both welds and allows for a slip fit between adjoining plates, thereby simplifying the fit-up of large, heavy plates.

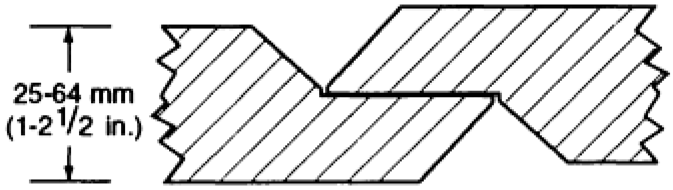

FIG. 3 OFFSET DOUBLE-V JOINT CONFIGURATION USED TO WELD THICK SECTION ALUMINUM ARMOR PLATE

Reference cited in this section

6. "WELDING KAISER ALUMINUM," KAISER ALUMINUM & SALES, INC., OAKLAND, CA, 1978, P 3.25-3.26

Selection and Weldability of Non-Heat-Treatable Aluminum Alloys

C.E. Cross, Martin Marietta Astronautics Group; M.L. Kohn, FMC Corporation

References

1. *ALUMINUM STANDARDS AND DATA*, ALUMINUM ASSOCIATION, 1984, P 93
2. PROPERTIES AND SELECTION NONFERROUS ALLOYS AND SPECIAL-PURPOSE MATERIALS, VOL 2, *ASM HANDBOOK*, ASM INTERNATIONAL, 1990, P 15-28
3. J.H. DUDAS AND F.R. COLLINS, PREVENTING WELD CRACKS IN HIGH-STRENGTH ALUMINUM ALLOYS, *WELD. J.*, VOL 45, 1966, P 241S-249S
4. C.E. CROSS, L.S. KRAMER, W.T. TACK, AND L.W. LOECHEL, ALUMINUM WELDABILITY AND HOT TEARING THEORY, *WELDABILITY OF MATERIALS*, ASM INTERNATIONAL, 1990, P 275-282
5. R.P. MARTUKANITZ AND P.R. MICHNUK, SOURCES OF POROSITY IN GAS METAL ARC WELDING OF ALUMINUM, *TRENDS IN WELDING RESEARCH*, ASM INTERNATIONAL, 1982, P 315-330
6. "WELDING KAISER ALUMINUM," KAISER ALUMINUM & SALES, INC., OAKLAND, CA, 1978, P 3.25-3.26

Selection and Weldability of Dispersion-Strengthened Aluminum Alloys

K. Sampath, Concurrent Technologies Corporation; W.A. Baeslack III, The Ohio State University

Introduction

CONVENTIONAL HIGH-STRENGTH ALUMINUM ALLOYS produced via ingot metallurgy (I/M) processing contain precipitate particles that interact with moving dislocations, thus providing high strength at room temperature. However, when these alloys are exposed to elevated temperatures, the precipitate particles undergo dissolution and/or coarsening, thereby reducing the strength of the base alloy. This behavior precludes the use of these alloys for elevated-temperature applications. In recent years, advances in two broad powder metallurgy (P/M) technologies--rapid solidification (RS) and mechanical alloying (mechanical attrition)--have enabled the development of a new family of high-strength aluminum alloys for elevated-temperature applications (Ref 1).

These new high-strength P/M aluminum alloys depend primarily on dispersion strengthening. Unlike the precipitate particles in I/M aluminum alloys, the dispersoid particles resist dissolution and/or coarsening when the alloys are exposed to elevated temperatures. In addition to high strength, these dispersion-strengthened aluminum alloys exhibit light weight, appreciable ductility, excellent fracture toughness, and resistance to stress-corrosion cracking (Ref 2). These alloys also exhibit coefficients of thermal expansion and galvanic properties comparable to those of conventional structural aluminum alloys (Ref 2). In view of the above properties, these dispersion-strengthened aluminum alloys are considered candidate materials to replace conventional high-strength aluminum alloys and titanium alloys used in the manufacture of selected aerospace components such as fan and compressor cases; vanes and blades in gas turbine engines; and fins, winglets, and rocket motor cases of missiles (Ref 3). Use of these alloys is expected to increase the thrust-to-weight ratio of gas turbine engines and allow significant improvements in the performance characteristics of advanced aircraft and weapons systems.

High-strength aluminum alloys produced via RS-P/M technology am primarily based on hypereutectic aluminum-iron compositions, although other alloys based on aluminum-chromium-zirconium, aluminum-manganese, aluminum-titanium, and aluminum-beryllium systems are also under various stages of development and/or evaluation (Ref 1). The dispersion-strengthened aluminum alloys produced via mechanical alloying contain a fine distribution of oxides, oxynitrides, and/or carbides in a fine-grain α-aluminum matrix. The α-aluminum matrix may additionally be strengthened by solid-solution strengthening or work hardening (Ref 1). Although many dispersion-strengthened aluminum alloys produced via mechanical alloying do not exhibit strength comparable to that of the conventional high-strength I/M aluminum alloys at room temperature, these alloys exhibit a minimal decrease in strength with increasing temperature (Ref 1).

The twofold purpose of this article is: (1) to guide users in the selection and application of welding processes and conditions for the joining of dispersion-strengthened aluminum alloys, based primarily on an understanding of the weldability of the material; and (2) to provide specific examples of material responses to welding conditions, chiefly by highlighting the microstructural development in the weld zone. In view of the limited number of publications available on the weldability and metallurgical response of this broad class of alloys, this article refers exclusively to the metallurgy and weldability of alloys based on the aluminum-iron system that are produced using various RS-P/M processing techniques.

References

1. J.R. PICKENS, HIGH STRENGTH ALUMINUM P/M ALLOYS, 10TH ED., VOL 2, *METALS HANDBOOK, PROPERTIES AND SELECTION: NON FERROUS ALLOYS AND SPECIAL-PURPOSE MATERIALS,* ASM INTERNATIONAL, 1990, P 200-215
2. S.K. DAS AND L.A. DAVIES, HIGH PERFORMANCE AEROSPACE ALLOYS VIA RAPID SOLIDIFICATION PROCESSING, *MATER. SCI. ENG.,* VOL 98, 1988, P 1-12
3. C.M. ADAM, RAPID SOLIDIFICATION PROCESSING METHODS, *MECHANICAL BEHAVIOR OF RAPIDLY SOLIDIFIED MATERIALS,* S.M.L. SASTRY AND B.A. MACDONALD, ED., THE METALLURGICAL SOCIETY OF AIME, 1985, P 21-39

Selection and Weldability of Dispersion-Strengthened Aluminum Alloys

K. Sampath, Concurrent Technologies Corporation; W.A. Baeslack III, The Ohio State University

Metallurgy of RS-P/M Aluminum-Iron-Base Alloys

Table 1 shows a list of selected dispersion-strengthened aluminum-iron-base alloys produced via RS-P/M processing. These alloys contain ternary and quaternary additions of transition and rare earth metals such as molybdenum, nickel, cerium, vanadium, and silicon. These alloy additions exhibit low solid-solubility and low solid-state diffusivity in α aluminum (Ref 4), and consequently they tend to form thermally stable dispersoid particles. However, these alloy compositions are characterized by a wider freezing range (typically in excess of 300 °C, or 540 °F) than most aluminum alloys. Conventional I/M processing of these compositions produces equilibrium or near-equilibrium microstructures containing coarse Al_3Fe-type primary intermetallic particles in a coarse α-aluminum matrix. Such microstructures exhibit poor ductility and cannot be used for structural engineering applications. In contrast, RS processing of these novel chemistries suppresses the formation of Al_3Fe-type primary intermetallic particles. It instead produces particulates exhibiting refined metastable microstructures, such as an Al-Al_6Fe-type microeutectic and/or a solute-supersaturated dendritic aluminum. In other words, RS processing conditions produce refined hypoeutectic-type microstructures from hypereutectic melts. The occurrence of these metastable hypoeutectic-type microstructures is attributed to the following two characteristics of RS processing techniques: (1) subdivision of the melt into particulates isolates potential nucleants and produces nucleant-free particulates that undergo nonequilibrium solidification (Ref 5); and (2) the rapid cooling rate does not provide sufficient reaction time for a potential nucleant to produce the equilibrium microstructure (Ref 6).

TABLE 1 CHEMICAL COMPOSITIONS OF DISPERSION-STRENGTHENED ALUMINUM-BASE ALLOYS PRODUCED BY RS-P/M

NOMINAL CHEMICAL COMPOSITION, WT%	MANUFACTURER
AL-8FE-1.7NI	**ALCOA**
AL-8.4FE-3.7CE	**ALCOA**
AL-9FE-4CE (AA8019 ALLOY)	**ALCOA**
AL-9FE-7CE	**ALCOA**
AL-10FE-5CE	**ALCOA**
AL-8.5FE-1.3V-1.7SI (AA8009 ALLOY)	**ALLIEDSIGNAL INC.**
AL-11.7FE-1.2V-2.4SI (FVS 1212 ALLOY)	**ALLIEDSIGNAL INC.**
AL-8FE-2.3MO	**PRATT & WHITNEY**
AL-8.7FE-2.8MO-1V	**PRATT & WHITNEY**

Subsequent thermomechanical processing of the rapidly solidified particulates produces a consolidated product with a unique "engineered" microstructure consisting of submicron-size dispersoid particles in an extremely fine-grain (<1 μm) α-aluminum matrix. During thermomechanical processing, the as-solidified hypoeutectic-type microstructures undergo solid-state phase transformations and produce fine dispersoid particles in an α-aluminum matrix. These particles result from the low-equilibrium solid-solubility limit of the alloy addition, and they consist of both the stable, "non-shearable," incoherent θ(Al_3Fe-type) dispersoid particles and metastable, "shearable," partially coherent θ' (Al_3Fe-type) dispersoid particles. The ternary and quaternary alloy additions promote the formation of complex metastable and stable dispersoid particles, and they reduce the surface free energy or the lattice misfit between the dispersoid particles and the α-aluminum matrix (Ref 2, 4, 7, 8). The partially coherent particles exhibit a spherical morphology and provide excellent resistance to fracture or crack initiation. These microstructural features provide excellent thermal stability and superior mechanical properties, including high strength and ductility, both at room temperature and at elevated temperatures up to 345 °C (650 °F). Because RS-P/M processing results in microstructures finer than the resolution limit of light microscopy, transmission electron microscopy (TEM) is often required to resolve the distribution and morphology of the microstructural constituents. Figures 1(a) and 1(b) compare a light micrograph of alloy AA8009 with a TEM micrograph. Figures 1(c) and 1(d) show the TEM micrographs of Al-8Fe-4Ce and Al-9Fe-3Mo-1V alloys. The volume fraction of the dispersoid particles in these representative alloys typically ranges from 15 to 35%.

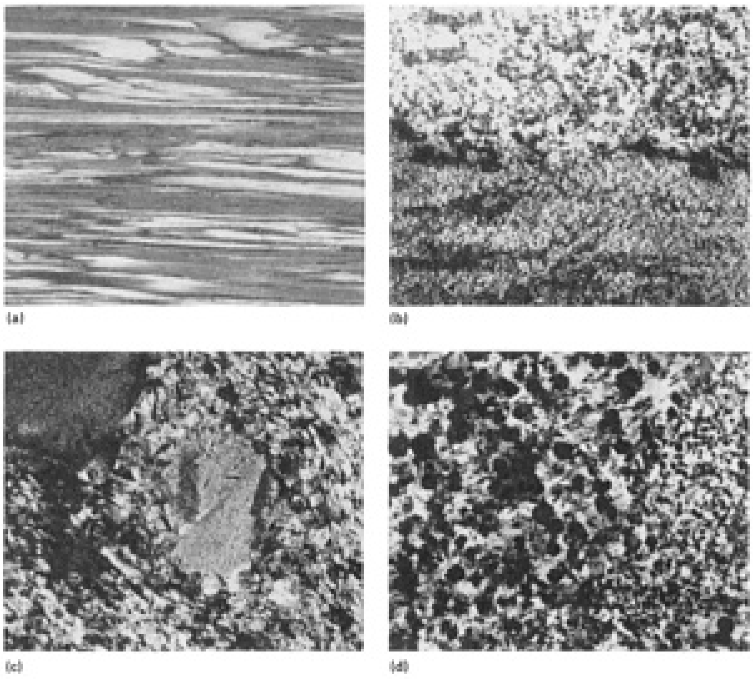

FIG. 1 MICROSTRUCTURE OF THREE REPRESENTATIVE RS-P/M ALUMINUM-IRON-BASE ALLOYS. (A) LIGHT MICROGRAPH OF AA8009 ALLOY. (B) TEM MICROGRAPH OF AA8009 ALLOY. SOURCE: REF 9. (C) AL-8.4FE-3.7CE ALLOY. SOURCE: REF 10. (D) AL-9FE-3MO-1V ALLOY. SOURCE: REF 11

References cited in this section

2. S.K. DAS AND L.A. DAVIES, HIGH PERFORMANCE AEROSPACE ALLOYS VIA RAPID SOLIDIFICATION PROCESSING, *MATER. SCI. ENG.,* VOL 98, 1988, P 1-12
4. C.M. ADAM AND R.G. BOURDEAU, TRANSITION AND REFRACTORY ELEMENT ADDITIONS TO RAPIDLY SOLIDIFIED ALUMINUM ALLOYS, *RAPID SOLIDIFICATION PROCESSING: PRINCIPLES AND TECHNOLOGIES II,*R. MEHRABIAN, B.H. KEAR, AND M. COHEN, ED., CLAITORS PUBLISHING, 1980, P 246-259
5. I.E. ANDERSON AND M.P. KEMPAINEN, UNDERCOOLING EFFECTS IN GAS-ATOMIZED POWDERS, *UNDERCOOLED ALLOY PHASES,* E.W. COLLINGS AND C.C. KOCH, ED., THE METALLURGICAL SOCIETY OF AIME, 1988, P 269-285
6. C.M. ADAM, RAPID SOLIDIFICATION TECHNOLOGY: THE POTENTIAL FOR INNOVATIVE ALLOY PROPERTIES, *SCIENCE AND TECHNOLOGY OF UNDERCOOLED MELTS,* P.R. SAHM, H. JONES, AND C.M. ADAMS, ED., NATO-ASI SERIES, MARTINUS-NIJHOFF PUBLISHERS, 1986, P 186-209
7. C.M. ADAM, STRUCTURE/PROPERTY RELATIONSHIPS AND APPLICATIONS OF RAPIDLY SOLIDIFIED ALUMINUM ALLOYS, *RAPIDLY SOLIDIFIED AMORPHOUS AND CRYSTALLINE ALLOYS,* VOL 8, B.H. KEAR, B.C. GIESSEN, AND M. COHEN, ED., ELSEVIER SCIENCE PUBLISHING CO., 1982, P 411-422
8. S.L. LANGENBECK, W.M. GRIFFITH, G.J. HILDEMAN, AND J.W. SIMON, DEVELOPMENT OF DISPERSION-STRENGTHENED ALUMINUM ALLOYS, *RAPIDLY SOLIDIFIED ALUMINUM ALLOYS,* STP 890, M.E. FINE AND E.A. STARKE, JR., ED., ASTM, 1986, P 410-422
9. H.H. KOO, S. KRISHNASWAMY, AND W.A. BAESLACK III, CHARACTERIZATION OF AN INERTIA-FRICTION WELD IN A HIGH-TEMPERATURE RS/PM AI-FE-V-SI ALLOY-AA8009, *MATER. CHARACT.,* VOL 26 (NO. 3), 1991, P 123-136
10. W.A. BAESLACK III, K.H. HOU, AND J.H. DEVLETIAN, ELECTRON MICROSCOPY OF RAPIDLY SOLIDIFIED WELDMENTS IN A POWDER METALLURGY AL-FE-CE ALLOY, *J. MATER. SCI.*

LETT., VOL 8 (NO. 6), 1989, P 716-720

11. H.H. KOO, K. SAMPATH, AND W.A. BAESLACK III, CHARACTERIZATION OF INERTIA-FRICTION WELDS BETWEEN A RAPIDLY-SOLIDIFIED AL-FE-MO-V ALLOY AND IM 2024-T351, *J. MATER. SCI.*, VOL 27, 1992, P 3266-3280

Selection and Weldability of Dispersion-Strengthened Aluminum Alloys

K. Sampath, Concurrent Technologies Corporation; W.A. Baeslack III, The Ohio State University

---

Weldability Issues

Although a variety of fusion welding and solid-state welding processes and techniques are available to join the conventional I/M aluminum alloys (Ref 12), several weldability issues, including the absence of compatible filler metals, preclude the application of many conventional welding processes and methodologies to the joining of dispersion-strengthened aluminum alloys. In particular, weldability issues related to weld solidification behavior, formation of hydrogen-induced porosity in the weld zone, and high-temperature deformation behavior of these alloys affect the selection and application of both fusion and solid-state welding processes.

Weld Solidification Behavior

The application of high-strength dispersion-strengthened aluminum alloys in the manufacture of specific aerospace structural components, where weight savings are critical for mission range and performance, will require the application of welding processes and procedures that optimize weld joint efficiency and minimize weld reinforcement. Because the high strength of the base alloy is due to the unique engineered microstructure obtained through RS-P/M processing, welding processes and procedures must re-create high-performance microstructures similar to the base alloy microstructures, and/or retain the superior base alloy microstructure in the weld zone. Since the wide freezing range of these alloys will tend to promote hypereutectic solidification and equilibrium or near-equilibrium microstructures, fusion welding conditions that promote RS or hypoeutectic-type solidification within the fusion zone will be required to re-create high-performance microstructures in the fusion zone. Welding conditions such as high energy density and minimum energy input, that increase the temperature gradients in both the fusion zone and the heat-affected zone (HAZ) and increase the overall weld cooling rate, will be required to minimize the widths of the partially melted region and the HAZ. Such welding conditions can be expected to facilitate the occurrence of plane front solidification, promote microstructural refining in the fusion zone, and minimize the extent of microstructural coarsening in the HAZ.

Alternatively, fusion welding conditions that provide minimum energy density and a shallow temperature gradient in the fusion zone can be expected to produce a wider HAZ and a wider partially melted region near the fusion boundary with the base metal. Depending on the local solidification conditions, this partially melted region can solidify to an equilibrium or near-equilibrium structure containing coarse intermetallic particles in a coarse-grain α-aluminum matrix. Coarse intermetallic particles can be expected to localize stress concentrations, thereby reducing the ductility and fracture toughness of the weldment. A distribution of coarse intermetallic particles can also be expected to be less effective at strengthening the matrix, especially at elevated temperatures.

Hydrogen-Induced Fusion Zone Porosity

The weldability of the dispersion-strengthened aluminum alloys is also limited by the hydrogen content of the base alloy. Hydrogen content in excess of 1 mL/100 g of aluminum is believed to produce excessive fusion zone porosity. Porosity in the fusion zone often reduces weld static strength and fatigue properties. Porosity occurs in aluminum alloy fusion welds mainly due to the rejection of hydrogen during weld solidification. It is related to the presence of hydrogen in the base alloy, filler metal, and the atmosphere and to the high ratio of solubility of hydrogen absorbed in the weld to the solubility limit of hydrogen at the melting point. The dispersion-strengthened aluminum alloys produced via RS-P/M processing often exhibit a residual hydrogen content in excess of 1 to 5 mL/100 g of aluminum and therefore show an increased tendency to form fusion zone porosity. Depending on the specific RS processing technique employed to produce the dispersion-strengthened aluminum alloy, the surface of the rapidly solidified particulates forms oxide layers to varying degrees (Ref 4, 7). During subsequent consolidation and handling, these surface oxides tend to hydrate. Decomposition of these hydrated oxides during fusion welding produces porosity in the weld fusion zone. Additionally, these oxides may serve as heterogeneous nucleation sites during fusion zone solidification, thereby limiting the extent of undercooling and promoting hypereutectic solidification conditions and the formation of equilibrium microstructures.

Controlling the Hydrogen Content. Three approaches are available to control the residual hydrogen content in RS-P/M aluminum alloys:

- USE VACUUM OR OTHER COMPLETELY INERT PROCESSING STEPS DURING RS PROCESSING AND CONSOLIDATION.
- DEGAS THE RAPIDLY SOLIDIFIED PARTICULATES PRIOR TO CONSOLIDATION, EITHER IN VACUUM OR IN OTHER INERT CARRIER GASES (DEPURITIVE TREATMENTS).
- DEGAS THE CONSOLIDATED PRODUCT (DIFFUSION CONSIDERATIONS NORMALLY RESTRICT THIS TREATMENT TO SHEETS AND PLATES) PRIOR TO FUSION WELDING.

The use of vacuum or other complete inert processing steps during RS processing and consolidation will minimize the formation of surface oxides and thereby control consequent hydrate formation. Degassing of either the rapidly solidified particulates or the consolidated products is aimed at removing or reducing the residual hydrogen to an acceptable level, preferably below 1 mL/100 g of aluminum. The critical factors that contribute to effective degassing are temperature, time, environment, and the product size and shape. Degassing treatments are effective only at temperatures in excess of 530 to 550 °C (990 to 1020 °F). Unfortunately, the use of such high temperatures often decomposes the metastable base alloy microstructure and/or promotes microstructural coarsening, thereby irreversibly degrading the mechanical properties of the base material (Ref 13).

Alternatively, despite a higher hydrogen content in the RS-P/M aluminum-base alloy, the occurrence of fusion zone porosity can be minimized by using welding techniques that increase the total pressure during welding. Since the equilibrium solubility of hydrogen in molten aluminum is governed by Sievert's law, the solubility of hydrogen in aluminum near the boiling point of aluminum can be decreased by reducing the partial pressure of hydrogen (that is, by increasing the total pressure). Depending on the specific characteristics of the fusion welding process, welding techniques that provide increased total pressure will reduce the partial pressure of hydrogen in the welding atmosphere and thereby reduce the level of fusion zone porosity. Increased total pressure is also believed to minimize the nucleation of hydrogen bubbles (Ref 14). However, such welds will contain excessive hydrogen, either in solution or as fine-size pores not detectable using conventional nondestructive evaluation techniques.

Applicable Welding Processes. The above weldability considerations show that RS-P/M aluminum-base alloys characterized by a low residual hydrogen content (<1 mL/100 g of aluminum) can be joined using fusion welding methods. As mentioned before, the absence of compatible fillers restricts the choice to essentially autogenous welding processes, while a need to recreate high-performance microstructures further limits that choice to high-energy-density welding methods such as capacitor-discharge welding, electron-beam welding, and laser welding. In fact, considering the effects of laser wavelength on the weldability of aluminum alloys, the application of the pulsed Nd:YAG laser welding process is preferred because this process provides 1.06 μm wavelength radiation, which is more efficiently absorbed by the base alloys than the 10.6 μm wavelength CO_2 laser radiation.

Fusion welding of base alloys with higher residual hydrogen will require the use of weld pretreatment or other special techniques to control hydrogen content in the weld atmosphere. In fact, one company that manufactures high-strength aluminum alloys containing iron, vanadium, and silicon additions has developed material pretreatments that can enhance joining and provide acceptable welding procedures (with minimal porosity) for resistance spot welding, gas-tungsten arc welding (GTAW), laser-beam welding, and vacuum brazing of these alloys (Ref 15).

The RS-P/M aluminum-base alloys characterized by a higher residual hydrogen content (>1 mL/100 g of aluminum) will require the use of solid-state welding methods such as conventional friction welding, inertia friction welding, or linear friction welding. These solid-state welding processes employ relative motion of the parts to be joined under an axial force to generate frictional heat at the weld interface. During the initial stages of welding, the interface becomes plastic (softened) due to frictional heating. When adequate heating (softening) of the interface is obtained, the relative motion is slowed down and a forging force is applied to produce the weld joint. The forging force produces localized plastic deformation and expels the heated material from the interface (as a flash around the weld joint), thereby enabling the formation of a weld joint between nearly unaffected base metals. In this regard, the weldability of the dispersion-strengthened aluminum alloys is influenced by the effects of welding process conditions and parameters on the high-temperature deformation characteristics, and subsequent microstructural development in the weld zone of these alloys.

Since these solid-state welding processes preclude melting and solidification, these welds tolerate higher levels of residual hydrogen. Further, the absence of melting at the faying surfaces during these processes prevents solidification-related discontinuities, such as cracking, and enables joining with dissimilar materials. However, despite the inherent ability of solid-state welding processes to produce high-integrity weld joints, such processes are only selectively applicable in view of specific limitations on component size, cross-sectional shape, weld joint design, and difficulties in evaluating the weld joint integrity using conventional nondestructive evaluation techniques. Application of joining processes such as brazing or diffusion welding using interlayers may potentially circumvent the aforementioned weldability issues related to the dispersion-strengthened aluminum alloys.

References cited in this section

4. C.M. ADAM AND R.G. BOURDEAU, TRANSITION AND REFRACTORY ELEMENT ADDITIONS TO RAPIDLY SOLIDIFIED ALUMINUM ALLOYS, *RAPID SOLIDIFICATION PROCESSING: PRINCIPLES AND TECHNOLOGIES II,*R. MEHRABIAN, B.H. KEAR, AND M. COHEN, ED., CLAITORS PUBLISHING, 1980, P 246-259
7. C.M. ADAM, STRUCTURE/PROPERTY RELATIONSHIPS AND APPLICATIONS OF RAPIDLY SOLIDIFIED ALUMINUM ALLOYS, *RAPIDLY SOLIDIFIED AMORPHOUS AND CRYSTALLINE ALLOYS,* VOL 8, B.H. KEAR, B.C. GIESSEN, AND M. COHEN, ED., ELSEVIER SCIENCE PUBLISHING CO., 1982, P 411-422
12. *SOURCE BOOK ON SELECTION AND FABRICATION OF ALUMINUM ALLOYS,* AMERICAN SOCIETY FOR METALS, 1978, P 233-263
13. G.E. METZGER, GAS TUNGSTEN ARC WELDING OF A POWDER METALLURGY ALUMINUM ALLOY, *WELD. J.,* VOL 71 (NO. 8), 1992, P 297S-304S
14. M.C. FLEMINGS, *SOLIDIFICATION PROCESSING,* MCGRAW-HILL INC., 1974, P 207-208
15. P.S. GILMAN, ALLIED-SIGNAL INC., PRIVATE COMMUNICATION, 2 FEB 1993

Selection and Weldability of Dispersion-Strengthened Aluminum Alloys

K. Sampath, Concurrent Technologies Corporation; W.A. Baeslack III, The Ohio State University

Fusion Welding

Fusion welding studies on dispersion-strengthened aluminum alloys have focused on the use of GTAW, with and without the addition of filler wire; resistance spot welding; and autogenous welding using capacitor-discharge, electron-beam (EB), and pulsed Nd:YAG laser welding processes. These studies have related welding conditions to their effects on weld microstructure and mechanical properties.

Gas-Tungsten Arc Welding. Autogenous butt welding of 1.6 mm (0.06 in.) thick Al-10Fe-5Ce alloy sheets using the GTAW process has resulted in excessive porosity in the fusion zone (Ref 13). However, vacuum pretreating the base alloy sheets for 20 h at 400 °C (750 °F) prior to GTAW using direct current electrode negative and helium gas shielding conditions reduced the level of porosity in the weld zone significantly. Unfortunately, this pretreatment also resulted in an 8% decrease in the tensile strength of the base alloy due to microstructural coarsening.

The same investigation evaluated the application of three conventional, unmatched filler wire additions--ER4043 (Al-5Si), ER4047 (Al-12Si), and ER5356 (Al-5Mg)--to the gas-tungsten arc weld pool. The results indicated that the ER5356 filler metal provided virtually porosity-free weld deposits following the vacuum pretreatment of the base alloy. Transverse weld tensile testing of the gas-tungsten arc welds produced using the ER5356 filler wire additions and vacuum pretreatment showed the joint efficiency to increase from 55% at room temperature to about 81% at 315 °C (600 °F). Despite an acceptable weld joint efficiency at elevated temperature, the transverse weld tensile testing revealed low (1%) ductility. Weld microstructure characterization showed that the minimal ductility was related to the presence of brittle, acicular intermetallics near the fusion boundary with the base metal. The formation of these brittle intermetallics is evidently related to the occurrence of near-equilibrium weld solidification conditions.

Capacitor-Discharge Welding. Compared to the gas-tungsten arc weld in the Al-10Fe-5Ce alloy, the microstructural characterization of autogenous capacitor-discharge welds produced in 6.4 mm (0.25 in.) cylindrical specimens of an Al-8Fe-4Ce alloy showed that the extremely rapid heating and cooling rates obtainable in this process can allow the generation of rapidly solidified, size-refined fusion zone microstructures (Ref 10, 16). In addition, this investigation revealed that increasing the forge pressure during the final bonding stage minimized fusion zone porosity. Transmission electron microscopy characterization of a moderately cooled capacitor-discharged weld in Al-8Fe-4Ce alloy showed fine-size acicular intermetallic particles in an extremely fine dendritic α-aluminum matrix, along with a minimal distribution of spherical intermetallic particles (Fig. 2). Although the microstructural features of the weld zone were similar to those of the RS-P/M base alloy microstructure, the relatively fine weld zone microstructures showed a higher hardness than the base alloy.

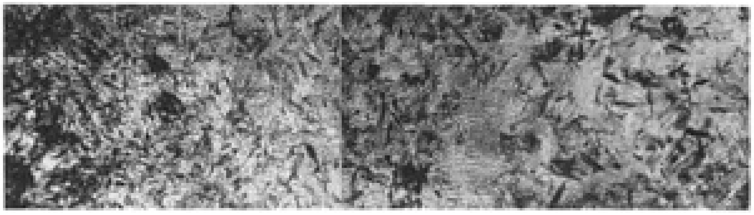

FIG. 2 A TEM MICROSTRUCTURE TRAVERSE ACROSS A MODERATELY COOLED CAPACITOR-DISCHARGE WELD IN AL-8FE-4CE ALLOY. SOURCE: REF 10

Electron-Beam and Laser-Beam Welding. Table 2 shows the weld parameters and weldment characteristics of autogenous EB and pulsed Nd:YAG laser welds produced in a low-hydrogen (<1 mL/100 g of aluminum) Al-8Fe-2Mo alloy. These welds showed only occasional evidence of fusion zone porosity (Ref 17, 18, 19). Evaluation of room-temperature weld mechanical properties of the EB welds produced in 1.27 mm (0.05 in.) thick Al-8Fe-2Mo alloy sheet showed a high weld joint efficiency of 73%, corresponding to 19.3 J/mm (490 J/in.) weld energy input, and preferential failure along the weld interface with the base material at minimal weld ductility. In comparison, transverse weld tensile testing of the EB weld, produced in 0.65 mm (0.026 in.) thick Al-8Fe-2Mo alloy sheet using 4.5 J/mm (114 J/in.) energy input determined 88% weld joint efficiency, appreciable weld ductility, and preferential failure in the weld HAZ. These observations indicate that reducing the energy input shifted the area most susceptible to failure from the fusion boundary region to the base metal. In fact, scanning electron microscopy examination of the fracture surfaces showed that the preferential failure along the weld interface was invariably related to the presence of acicular intermetallic particles in the fusion boundary region (Fig. 3).

TABLE 2 WELDING PARAMETERS AND WELDMENT CHARACTERISTICS OF AL-8FE-2.3MO ALLOY WELDS

WELDMENT NO.	TRAVEL SPEED		ENERGY INPUT		THRESHOLD BENDING STRAIN, %	TRANSVERSE WELD TENSILE TEST RESULTS				
						ULTIMATE TENSILE STRENGTH		ELONGATION, %	FAILURE LOCATION(A)	JOINT EFFICIENCY, %(B)
	MM/S	IN./MIN	J/MM	J/IN.		MPA	KSI			
ELECTRON-BEAM WELDING OF 1.27 MM (0.05 IN.) SHEET (REF 17)										
1(C)	16.9	41	23.6	599	4.0	330	47.8	1.6	FBR	**65**
2(C)	23.3	56	19.3	490	4.0	370	53.6	5.6	FBR	**73**
ELECTRON-BEAM WELDING OF 0.65 MM (0.026 IN.) SHEET (REF 18)										
3(D)	27.5	66	4.5	114	14.0	360	52.2	1.6	HAZ	**85**
PULSED ND:YAG LASER WELDING OF 1.27 MM (0.05 IN.) SHEET (REF 1)										
4(E)	**2.5**	**6**	**41.3(F)**	**1049(F)**	**11.0**	**374**	**54.2**	**11**	**UBM**	**100**

(A) FBR, FUSION BOUNDARY REGION; HAZ, HEAT AFFECTED ZONE; UBM, UNAFFECTED BASE METAL.

(B) JOINT EFFICIENCY IS A RATIO OF WELD STRENGTH TO BASE METAL STRENGTH IN THE SAME ORIENTATION.

(C) WELD PARAMETERS: ACCELERATION VOLTAGE 100 KV; CURRENT 3.5 MA.

(D) WELD PARAMETERS: ACCELERATION VOLTAGE 100 KV; CURRENT 1.25 MA.

(E) WELD PARAMETERS: 10 HZ PULSE RATE; 5.0 MS PULSE DURATION; 105 W PULSE POWER (10.5 J/PULSE); 2.1 KW PEAK POWER; 102 MM (4 IN.) FOCAL LENGTH; 10^6 W/MM^2 POWER DENSITY; 16.5 L/MIN (35 FT^3/H) ARGON SHIELDING GAS FLOW RATE.

(F) AVERAGE ENERGY INPUT

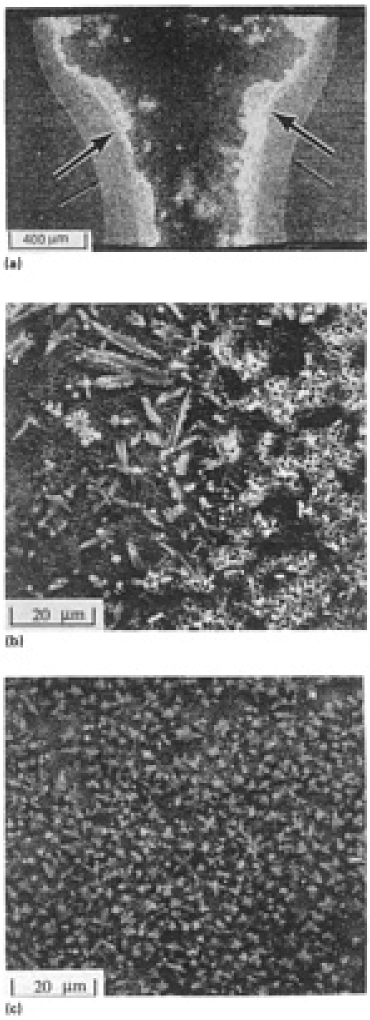

FIG. 3 ELECTRON-BEAM WELD (ENERGY INPUT 15 J/MM, OR 38 J/IN.) IN 1.27 MM (0.05 IN.) THICK AL-8FE-2MO SHEET. (A) WELD MACROSTRUCTURE (LARGE ARROWS INDICATE FUSION BOUNDARY; SMALL ARROWS INDICATE HAZ BOUNDING FUSION ZONE). (B) SEM MICROSTRUCTURE NEAR THE FUSION BOUNDARY. (C) SEM MICROSTRUCTURE NEAR WELD CENTER. SOURCE: REF 17

Microscopic examination of the low-energy-input EB weld showed that reducing the weld energy input minimized or eliminated the occurrence of acicular intermetallic particles within the fusion zone and produced refined α-aluminum dendrites (Ref 18). Coincident with the elimination of the acicular intermetallic particles near the fusion boundary, a relatively light-etching region was observed within the weld zone. Transmission electron microscopy examination of these regions revealed α-aluminum growth structures around coarse-size intermetallic particles (Fig. 4). In fact, TEM examination of a keyhole-type EB weld produced in a 2.5 mm (0.10 in.) thick RS-P/M Al-8Fe-1.7Ni alloy showed a similar microstructure at the weld center, one that contained primary Al_3Fe-type intermetallic particles nucleating an α-aluminum sheath around the primary particles (Ref 21). The α-aluminum sheath structure changed to a lamellar eutectic structure. In contrast to this structure, the fusion boundary region of this weld showed an epitaxially grown, extremely fine Al/Al_6Fe-type coupled-eutectic structure.

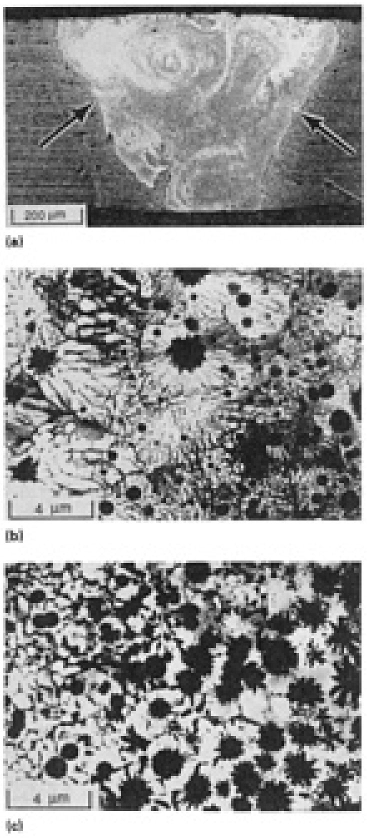

FIG. 4 ELECTRON-BEAM WELD (ENERGY INPUT 4.5 J/MM, OR 114 J/IN.) IN 0.65 MM (0.026 IN.) THICK AL-8FE-2MO SHEET. (A) WELD MACROSTRUCTURE (LARGE ARROWS INDICATE FUSION BOUNDARY, SMALL ARROWS INDICATE HAZ BOUNDING FUSION ZONE). (B) TEM MICROSTRUCTURE OF THE LIGHT-ETCHING REGIONS. (C) TEM MICROSTRUCTURE OF THE DARK-ETCHING REGIONS NEAR WELD CENTER. SOURCE: REF 20

Compared with the EB welds, autogenous pulsed Nd:YAG laser welding studies on 1.27 mm (0.05 in.) thick Al-8Fe-2Mo alloy sheet using a higher-energy-density (power density >10^6 W/mm^2) welding condition showed that the transverse weld joint efficiency could be improved to 100% (Ref 19). The longitudinal weld bend ductility was also comparable with that of the RS-P/M base alloy. Transmission electron microscopy examination of the fusion boundary region showed that the pulsed Nd:YAG laser welding conditions produced a matrix consisting of a columnar-dendritic α-aluminum structure with a uniform distribution of fine Al_6Fe-type spherical intermetallic particles and solute-rich particles at the dendrite interstices (Fig. 5). The pulse overlap region of the weld zone showed microstructures very similar to those observed in the EB welds, albeit on a structurally finer scale. The low overall weld energy input resulted in a narrow HAZ with the base metal and only minimal structural coarsening (Ref 21, 22).

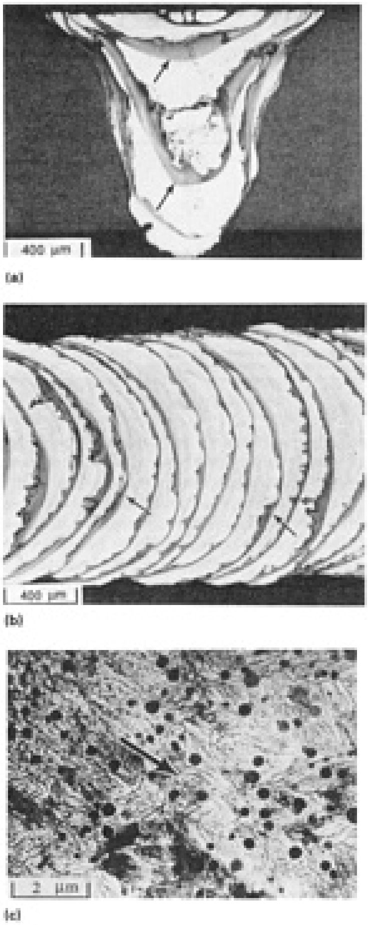

FIG. 5 PULSED ND:YAG LASER WELD IN AL-8FE-2MO SHEET. (A) TRANSVERSE SECTION. (B) PLAN VIEW. (C) TEM MICROSTRUCTURE OF THE LIGHT-ETCHING REGIONS NEAR THE FUSION BOUNDARY WITH BASE ALLOY. ARROWS IN (A) AND (B) INDICATE CURVILINEAR BANDS BOUNDING SUCCESSIVE MELT ZONES. ARROW IN (C) INDICATES COLUMNAR-DENDRITIC GROWTH DIRECTION. SOURCE: REF 19

Development of Weld Microstructure. The occurrence of different types of weld zone microstructures and significant differences in weld mechanical properties underscore the effects of welding process conditions and parameters on weld microstructure development. Specifically, distinct variations in the morphology and population density of the intermetallic particles within the weld zone occur as the weld energy input is decreased and/or as the energy density (power density) is increased. An analysis based on peak temperature distribution during fusion welding indicated that despite the use of high-energy-density welding processes, these welds likely contained undissolved base metal dispersoid particles within the bulk fusion zone (Ref 20). A weld solidification analysis based on this premise showed that the subsequent microstructure development within the fusion zone occurs as one of the following three different types, depending on local thermal conditions--principally cooling rate and temperature gradient--that depend on welding process condition and parameters:

- *TYPE A:* AL_3FE-TYPE INTERMETALLIC PARTICLES IN A MATRIX OF DENDRITIC A ALUMINUM WITH INTERDENDRITIC LAMELLAR OR DIVORCED EUTECTIC
- *TYPE B:* AL_3FE-TYPE INTERMETALLIC PARTICLES THAT ACT AS GROWTH CENTERS FOR DENDRITIC α ALUMINUM WITH INTERDENDRITIC LAMELLAR OR DIVORCED EUTECTIC
- *TYPE C:* AL_6FE-TYPE INTERMETALLIC PARTICLES ENTRAPPED IN A MATRIX OF DENDRITIC α ALUMINUM WITH INTERDENDRITIC DIVORCED EUTECTIC

Depending on local weld solidification conditions, the above three types of microstructures can be observed at different regions within the same weld.

Figures 6, 7, and 8 show Al-Al_3Fe pseudobinary phase diagrams to illustrate possible solidification events that promote the formation of type A, type B, and type C microstructures, respectively. In these diagrams, C_O represents the nominal composition of the alloy, and C_O' represents the effective composition of the melt (containing undissolved base metal dispersoid particles) prior to the onset of weld solidification, T_E is the equilibrium (Al-Al_3Fe) eutectic temperature and T'_E is the metastable (Al-Al_3Fe) eutectic temperature. In Fig. 6 and 7, C''_O represents the composition of the melt at which the solidification mode changes from hypereutectic to hypoeutectic with respect to the metastable eutectic composition.

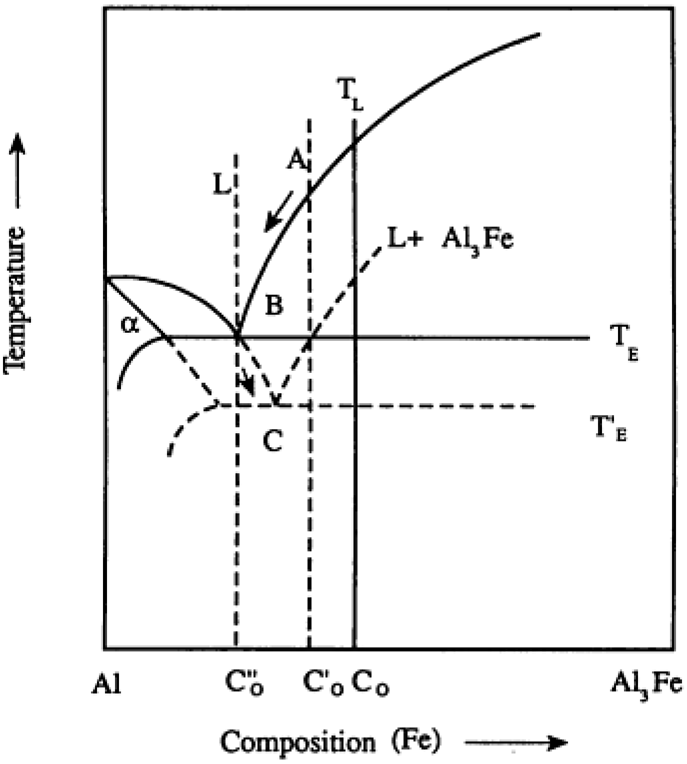

FIG. 6 AL-AL_3FE PSEUDOBINARY PHASE DIAGRAM ILLUSTRATING THE SEQUENCE OF SOLIDIFICATION EVENTS LEADING TO THE FORMATION OF TYPE A MICROSTRUCTURE. SOURCE: REF 20

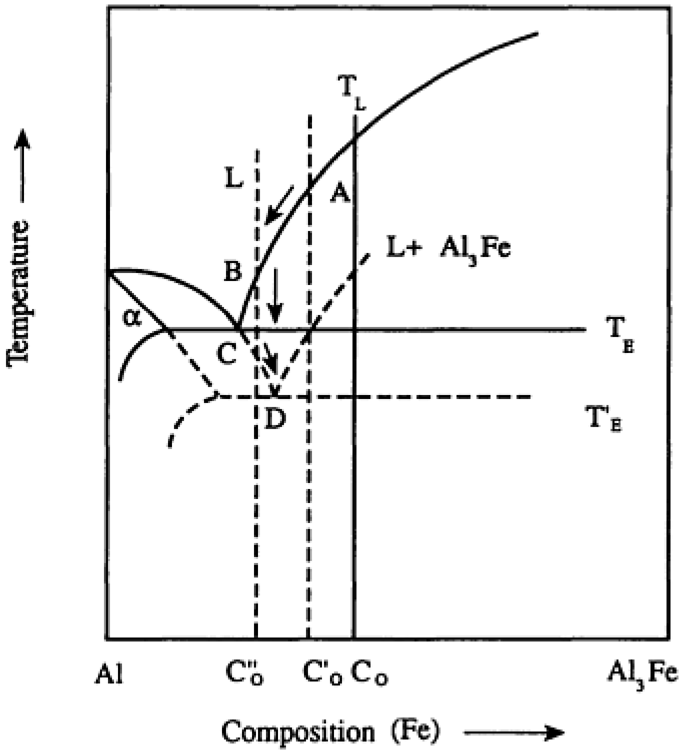
Temperature
T$_L$
L
A
L+ Al$_3$Fe
B
α
C
T$_E$
D
T'$_E$
Al
C"$_o$
C'$_o$
C$_o$
Al$_3$Fe
Composition (Fe)

FIG. 7 AL-AL$_3$FE PSEUDOBINARY PHASE DIAGRAM ILLUSTRATING THE SEQUENCE OF SOLIDIFICATION EVENTS LEADING TO THE FORMATION OF TYPE B MICROSTRUCTURE. SOURCE: REF 20

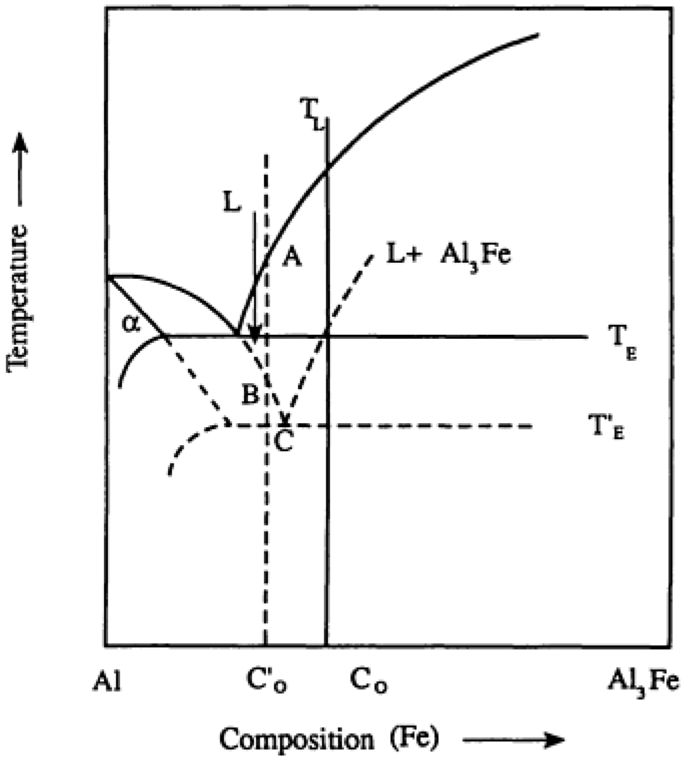
Temperature
T$_L$
L
A
L+ Al$_3$Fe
α
T$_E$
B
C
T'$_E$
Al
C'$_o$
C$_o$
Al$_3$Fe
Composition (Fe)

FIG. 8 AL-AL$_3$FE PSEUDOBINARY PHASE DIAGRAM ILLUSTRATING THE SEQUENCE OF SOLIDIFICATION EVENTS LEADING TO THE FORMATION OF TYPE C MICROSTRUCTURE. SOURCE: REF 20

The Type A Microstructure is a near-equilibrium microstructure and occurs within the fusion boundary region of high-energy-input welds. As shown in Fig. 6, the type A microstructure appears to occur in three stages. Initially, the undissolved base metal dispersoid particles grow in a relatively stagnant liquid region into randomly oriented, primary intermetallic particles and promote solute depletion in the remaining melt (from A to B). Since the primary intermetallic particles have a poor nucleation potency for α aluminum, solute depletion in the melt promotes the epitaxial growth of columnar-dendritic α aluminum from base metal α(from B to C). Further, since the α-aluminum solidification initiates from the fusion boundary at relatively low solidification velocities, the local solidification conditions promote appreciable structural coarsening and produce relatively coarse α-aluminum dendrites. Solute partitioning accompanying the α-aluminum growth promotes a solute-rich melt at the interdendritic regions, which undergoes terminal solidification to a metastable Al/Al_6Fe-type lamellar eutectic at the α-aluminum dendrite interstices (at C). These types of microstructures are commonly observed throughout the fusion zone in gas-tungsten arc weldments produced in these alloy types (Ref 13).

The Type B Microstructure contains both stable and metastable microstructural constituents. This structure is associated with high-energy-input welding conditions and is commonly observed within the central weld zone. As shown in Fig. 7, this microstructure appears to form with the growth of Al_3Fe-type intermetallic particles (from A to B). Due to high weld cooling rates, considerable undercooling occurs (from B to C) subsequent to the initial growth of the primary intermetallic particles, which allows the growth of supersaturated α-aluminum dendrites (from C to D). The occurrence of a metastable α-aluminum growth structure also indicates that the composition of the melt surrounding the equiaxed intermetallic particles was essentially hypoeutectic relative to the metastable eutectic composition. With the growth of the α-aluminum, the composition of the remaining melt in the dendrite interstices is enriched with solute elements. It ultimately reaches the metastable eutectic composition and undergoes terminal (metastable) eutectic solidification (at D). In view of the fine size of the interdendritic region, solute-rich particles are formed in this region instead of a regular lamellar eutectic (Ref 23).

The formation of the metastable α-aluminum growth structure around the primary intermetallic particles is attributed to catalytic nucleation. Catalytic nucleation of α aluminum occurs when the lattice disregistry between the potential nucleant and αaluminum is reduced below 10% (Ref 24). Since the supersaturation of solute elements in aluminum shows a negative deviation from Vegard's law and reduces the lattice parameter of α aluminum (Ref 25), improved lattice registry between the primary intermetallic particles and the metastable α aluminum, along with shallow temperature gradients in the central fusion zone, is considered responsible for the occurrence of the equiaxed-dendritic α-aluminum growth structure.

The Type C Microstructure is essentially a metastable growth structure and is associated with the extremely high-energy-density, low-energy-input welding conditions. Similar to the type A near-equilibrium microstructure, the type C metastable microstructure starts from the fusion boundary region with the base metal, by epitaxial solidification of α(or coupled eutectic) from base metal α in the fusion boundary. As shown in Fig. 8, the epitaxial growth of the α-aluminum grains appears to occur at significantly high undercoolings, possibly at temperatures slightly above the metastable eutectic temperature. Evidently, the coupled eutectic growth structures occur when the undercooling is below the metastable eutectic temperature.

The epitaxially grown α-aluminum grains entrap the undissolved dispersoid particles as solidification proceeds along a relatively steep temperature gradient across the advancing solid-liquid interface. The spherical morphology of the intermetallic particles is attributed to the high fusion zone cooling rate (in excess of 10^5 °C/s, or 1.8×10^5 °F/s), which does not provide adequate time for the growth of the undissolved base metal dispersoid particles. The steep temperature gradients in this region promote solidification of the melt to a columnar-dendritic substructure with fine intermetallics or solute-rich particles partitioning to the dendrite interstices. Coincident with the growth of the metastable α aluminum, the remaining melt is enriched with solute elements, which undergo final solidification to fine particles due to the relatively fine size of the interdendritic regions (Ref 23).

Epitaxial solidification generally indicates local wetting of the fusion boundary and minimal melt undercooling. However, epitaxially grown type C microstructures in the fusion boundary suggest that despite a readily wettable substrate, metastable growth structures in the weld fusion zone can likely occur from the combined effects of a minimal nucleation barrier for the metastable phase, a steep temperature gradient that restricts the growth of the stable phase, and perhaps a minimal difference in the level of undercooling required to allow the growth of metastable structures. In fact, although the nucleation of Al_3Fe occurs at less than 8 °C (14 °F) undercooling in aluminum-iron melts, the nucle tion of Al_6Fe is favored at undercoolings in excess of 10 °C (18 °F) (Ref 26).

References cited in this section

1. J.R. PICKENS, HIGH STRENGTH ALUMINUM P/M ALLOYS, 10TH ED., VOL 2, *METALS HANDBOOK, PROPERTIES AND SELECTION: NON FERROUS ALLOYS AND SPECIAL-PURPOSE MATERIALS,* ASM INTERNATIONAL, 1990, P 200-215
10. W.A. BAESLACK III, K.H. HOU, AND J.H. DEVLETIAN, ELECTRON MICROSCOPY OF RAPIDLY SOLIDIFIED WELDMENTS IN A POWDER METALLURGY AL-FE-CE ALLOY, *J. MATER. SCI. LETT.,* VOL 8 (NO. 6), 1989, P 716-720
13. G.E. METZGER, GAS TUNGSTEN ARC WELDING OF A POWDER METALLURGY ALUMINUM ALLOY, *WELD. J.,* VOL 71 (NO. 8), 1992, P 297S-304S
16. W.A. BAESLACK III, K.H. HOU, AND J.H. DEVLETIAN, RAPID SOLIDIFICATION JOINING OF A POWDER METALLURGY AL-FE-CE ALLOY, *J. MATER. SCI. LETT.,* VOL 7 (NO. 9), 1988, P 944-948
17. W.A. BAESLACK III AND S. KRISHNASWAMY, ELECTRON BEAM WELDABILITY OF A RAPIDLY SOLIDIFIED ALUMINUM ALLOY, *ADVANCES IN WELDING SCIENCE & TECHNOLOGY,* S.A. DAVID, ED., ASM INTERNATIONAL, 1986, P 357-362
18. S. KRISHNASWAMY AND W.A. BAESLACK III, STRUCTURE, PROPERTIES AND FRACTURE OF EB WELDS IN AL-8FE-2MO, *MATER. SCI. ENG.,* VOL 98, 1988, P 137-141
19. S. KRISHNASWAMY AND W.A. BAESLACK III, ND:YAG LASER WELDABILITY OF RS/PM AL-8FE-2MO SHEET, *RECENT TRENDS IN WELDING SCIENCE & TECHNOLOGY,* S.A. DAVID AND J.M. VITEK, ED., ASM INTERNATIONAL, 1990, P 631-636
20. S. KRISHNASWAMY, WELDABILITY OF A RAPID SOLIDIFICATION POWDER METALLURGY AL-8FE-2MO ALLOY, PH.D. DISSERTATION, THE OHIO STATE UNIVERSITY, MARCH 1989
21. W.A. BAESLACK III, MICROSTRUCTURAL TRANSITIONS IN A WELDED POWDER METALLURGY AL-8FE-1.7NI ALLOY, *METALLOGRAPHY,* VOL 18, 1985, P 73-82
22. S. KRISHNASWAMY AND W.A. BAESLACK III, METALLOGRAPHY OF A PULSED ND:YAG LASER WELD IN A RS/PM AL-8FE-2MO SHEET, *MATER. CHARACT.,* VOL 24 (NO. 4), 1990, P 331-353
23. H. JONES, *RAPID SOLIDIFICATION OF METALS AND ALLOYS,* MONOGRAPH 8, INSTITUTION OF METALLURGISTS, LONDON, 1982, P 45
24. J.H. PEREPEZKO AND S.E. LEBEAU, NUCLEATION OF AL DURING SOLIDIFICATION, *ALUMINUM TRANSFORMATION TECHNOLOGY AND APPLICATIONS--1981,* C.A. PAMPILLO, H. BILONI, L.F. MONDOLFO, AND F. SACCHI, ED., AMERICAN SOCIETY FOR METALS, 1982, P 309-346
25. H. JONES, OBSERVATIONS ON A STRUCTURAL TRANSITION IN ALUMINUM ALLOYS HARDENED BY RAPID SOLIDIFICATION, *MATER. SCI. ENG.,* 1969-70, P 1-18
26. L.F. MONDOLFO, ALUMINUM-IRON SYSTEM, *ALUMINUM ALLOYS: STRUCTURE AND PROPERTIES,* BUTTERWORTHS, LONDON, 1976, P 283-289

Selection and Weldability of Dispersion-Strengthened Aluminum Alloys

K. Sampath, Concurrent Technologies Corporation; W.A. Baeslack III, The Ohio State University

---

Solid-State Welding

The solid-state welding studies on RS-P/M aluminum-iron-base alloys are mostly limited to the application of conventional friction welding, inertia friction welding, or linear friction welding methods to produce similar or dissimilar metal welds in cylindrical specimens. In general, these studies evaluated the effects of weld axial force on the structure, mechanical properties, and fracture behavior of the friction welds. These studies include similar welds between Al-9Fe-4Ce and Al-9Fe-7Ce alloys, dissimilar welds between Al-9Fe-4Ce and I/M 2024-T351 (Ref 27), similar welds between AA8009 alloy (Ref 9), similar welds between FVS1212 alloy (Ref 28), dissimilar welds between FVS1212 alloy and I/M

2024-T351 (Ref 29), similar welds between Al-9Fe-3Mo-1V alloys (Ref 30), and dissimilar welds between Al-9Fe-3Mo-1V alloy and I/M 2024-T351 (Ref 11). The similar welds exhibit a uniform flash around the weld joint, the amount of the flash increasing with the weld axial force. The dissimilar welds, however, show a nonuniform flash around the weld joint. The alloy with the higher strength at elevated temperatures (RS-P/M base alloy) forms a minimal flash compared to the alloy with a higher strength at room temperature (I/M 2024-T351 alloy).

Light and analytical electron microscopy examinations of the interface region (weld center) in inertia friction welds have shown extensive high-temperature plastic deformation in the heat-and-deformation zone. This plastic deformation was generally beneficial in effectively homogenizing the base metal microstructure. The plastic deformation often promoted the fracture of coarse, acicular dispersoid particles present in selected base alloys (Ref 11, 30). However, the fine spherical dispersoid particles commonly observed in AA8009 alloy were relatively unaffected by the level of plastic deformation occurring in the weld zone and usually did not show any evidence of coarsening or fracture (Ref 9). Further, in welds produced using a low axial force, nonuniform deformation at the weld outer periphery resulted in the formation of dispersoid-lean, coarse-grain regions with a hardness appreciably lower than that of the base alloy. The application of a sufficiently higher axial force extruded out this softened region as weld flash, thereby preventing the hardness degradation across the weld zone (Fig. 9). Results of transverse weld tensile testing were generally consistent with the microstructural observations, showing that weld joint efficiency up to 95% of the base metal tensile strength can be obtained in welds produced using high axial force.

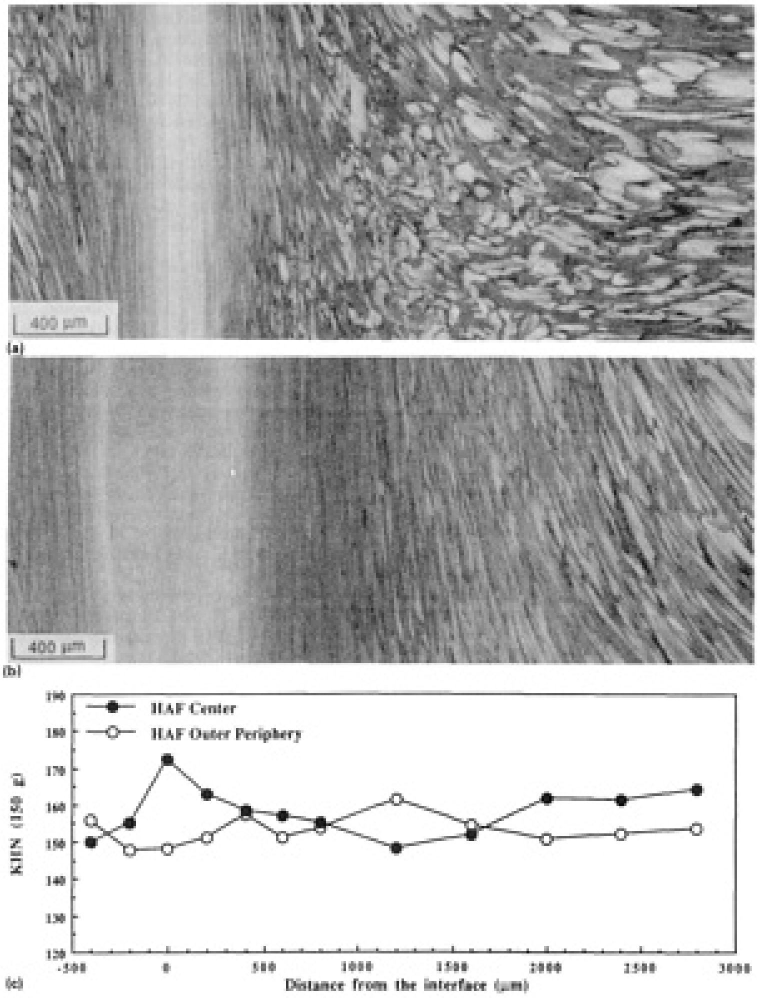

FIG. 9 LIGHT MICROGRAPHS OF AN INERTIA-FRICTION WELD IN AA8009 ALLOY PRODUCED USING HIGH AXIAL FORCE. (A) CENTER. (B) OUTER PERIPHERY. (C) CORRESPONDING KNOOP HARDNESS TRAVERSE. SOURCE: REF 9

Unlike the microstructure of the inertia friction welds, linear friction welds (produced in similar welds between FVS1212 alloy and dissimilar welds between FVS1212 alloy and I/M 2024-T351) showed appreciable coarsening of the α-aluminum grains and the base metal dispersoid particles resulting from a somewhat slower weld thermal cycle and less forging force (Ref 29). Appropriate optimization of welding parameters to provide rapid weld thermal cycles is expected to minimize the microstructural coarsening in the weld zone.

References cited in this section

9. H.H. KOO, S. KRISHNASWAMY, AND W.A. BAESLACK III, CHARACTERIZATION OF AN INERTIA-FRICTION WELD IN A HIGH-TEMPERATURE RS/PM AI-FE-V-SI ALLOY-AA8009, *MATER. CHARACT.,* VOL 26 (NO. 3), 1991, P 123-136
11. H.H. KOO, K. SAMPATH, AND W.A. BAESLACK III, CHARACTERIZATION OF INERTIA-FRICTION WELDS BETWEEN A RAPIDLY-SOLIDIFIED AL-FE-MO-V ALLOY AND IM 2024-T351, *J. MATER. SCI.,* VOL 27, 1992, P 3266-3280
27. W.A. BAESLACK III AND K. HAGEY, INERTIA FRICTION WELDING OF RAPIDLY SOLIDIFIED POWDER METALLURGY ALUMINUM, *WELD. J.,* VOL 67 (NO. 8), 1988, P 139S-149S
28. H.H. KOO, S. KRISHNASWAMY, AND W.A. BAESLACK III, SOLID-PHASE WELDING OF A DISPERSION-STRENGTHENED AL-FE-V-SI ALLOY-FVS1212, *LOW DENSITY, HIGH STRENGTH POWDER METALLURGY ALLOYS,* W.E. FRAZIER, M.J. KOCZAK, AND P.W. LEE, ED., THE METALLURGICAL SOCIETY OF AIME, 1991, P 183-196
29. H.H. KOO AND W.A. BAESLACK III, FRICTION WELDING OF A RAPIDLY SOLIDIFIED AL-FE-V-SI ALLOY, *WELD. J.,* VOL 71 (NO. 5), 1992, P 147S-169S
30. K.H. HOU AND W.A. BAESLACK III, ELECTRON MICROSCOPY OF INERTIA-FRICTION WELDMENTS IN A RAPIDLY SOLIDIFIED AL-FE-MO-V ALLOY, *J. MATER. SCI.,* VOL 25, 1990, P 2642-2653

Diffusion Welding

Diffusion welding of RS-P/M aluminum-iron-cerium alloy was investigated using electroplated silver interlayers (Ref 31). Temperature-time and temperature-pressure diffusion welding parameter zones were developed for an A1-9Fe-4Ce alloy and showed that it was possible to perform diffusion welding operations at temperatures up to 600 °C (1110 °F) without structural support. Although poor joint efficiencies were obtained for welds produced at low pressures, an increase in pressure and the promotion of deformation at the weld interface promoted joint efficiencies of up to 70%. The weld interface regions were characterized by the presence of a silver-aluminum intermetallic and occasional Kirkendall porosity.

Reference cited in this section

31. V. ANANTHANARAYANAN, "DIFFUSION WELDING OF AN RS/PM ALUMINUM ALLOY," PH.D. DISSERTATION, THE OHIO STATE UNIVERSITY, 1988

References

1. J.R. PICKENS, HIGH STRENGTH ALUMINUM P/M ALLOYS, 10TH ED., VOL 2, *METALS HANDBOOK, PROPERTIES AND SELECTION: NON FERROUS ALLOYS AND SPECIAL-PURPOSE MATERIALS,* ASM INTERNATIONAL, 1990, P 200-215
2. S.K. DAS AND L.A. DAVIES, HIGH PERFORMANCE AEROSPACE ALLOYS VIA RAPID SOLIDIFICATION PROCESSING, *MATER. SCI. ENG.,* VOL 98, 1988, P 1-12
3. C.M. ADAM, RAPID SOLIDIFICATION PROCESSING METHODS, *MECHANICAL BEHAVIOR OF RAPIDLY SOLIDIFIED MATERIALS,* S.M.L. SASTRY AND B.A. MACDONALD, ED., THE METALLURGICAL SOCIETY OF AIME, 1985, P 21-39
4. C.M. ADAM AND R.G. BOURDEAU, TRANSITION AND REFRACTORY ELEMENT ADDITIONS TO RAPIDLY SOLIDIFIED ALUMINUM ALLOYS, *RAPID SOLIDIFICATION PROCESSING: PRINCIPLES AND TECHNOLOGIES II,*R. MEHRABIAN, B.H. KEAR, AND M. COHEN, ED., CLAITORS PUBLISHING, 1980, P 246-259
5. I.E. ANDERSON AND M.P. KEMPAINEN, UNDERCOOLING EFFECTS IN GAS-ATOMIZED POWDERS, *UNDERCOOLED ALLOY PHASES,* E.W. COLLINGS AND C.C. KOCH, ED., THE METALLURGICAL SOCIETY OF AIME, 1988, P 269-285
6. C.M. ADAM, RAPID SOLIDIFICATION TECHNOLOGY: THE POTENTIAL FOR INNOVATIVE ALLOY PROPERTIES, *SCIENCE AND TECHNOLOGY OF UNDERCOOLED MELTS,* P.R. SAHM, H. JONES, AND C.M. ADAMS, ED., NATO-ASI SERIES, MARTINUS-NIJHOFF PUBLISHERS, 1986, P 186-209
7. C.M. ADAM, STRUCTURE/PROPERTY RELATIONSHIPS AND APPLICATIONS OF RAPIDLY SOLIDIFIED ALUMINUM ALLOYS, *RAPIDLY SOLIDIFIED AMORPHOUS AND CRYSTALLINE ALLOYS,* VOL 8, B.H. KEAR, B.C. GIESSEN, AND M. COHEN, ED., ELSEVIER SCIENCE PUBLISHING CO., 1982, P 411-422
8. S.L. LANGENBECK, W.M. GRIFFITH, G.J. HILDEMAN, AND J.W. SIMON, DEVELOPMENT OF DISPERSION-STRENGTHENED ALUMINUM ALLOYS, *RAPIDLY SOLIDIFIED ALUMINUM ALLOYS,* STP 890, M.E. FINE AND E.A. STARKE, JR., ED., ASTM, 1986, P 410-422
9. H.H. KOO, S. KRISHNASWAMY, AND W.A. BAESLACK III, CHARACTERIZATION OF AN INERTIA-FRICTION WELD IN A HIGH-TEMPERATURE RS/PM AI-FE-V-SI ALLOY-AA8009, *MATER. CHARACT.,* VOL 26 (NO. 3), 1991, P 123-136
10. W.A. BAESLACK III, K.H. HOU, AND J.H. DEVLETIAN, ELECTRON MICROSCOPY OF RAPIDLY SOLIDIFIED WELDMENTS IN A POWDER METALLURGY AL-FE-CE ALLOY, *J. MATER. SCI. LETT.,* VOL 8 (NO. 6), 1989, P 716-720
11. H.H. KOO, K. SAMPATH, AND W.A. BAESLACK III, CHARACTERIZATION OF INERTIA-FRICTION WELDS BETWEEN A RAPIDLY-SOLIDIFIED AL-FE-MO-V ALLOY AND IM 2024-T351, *J. MATER. SCI.,* VOL 27, 1992, P 3266-3280
12. *SOURCE BOOK ON SELECTION AND FABRICATION OF ALUMINUM ALLOYS,* AMERICAN SOCIETY FOR METALS, 1978, P 233-263
13. G.E. METZGER, GAS TUNGSTEN ARC WELDING OF A POWDER METALLURGY ALUMINUM ALLOY, *WELD. J.,* VOL 71 (NO. 8), 1992, P 297S-304S
14. M.C. FLEMINGS, *SOLIDIFICATION PROCESSING,* MCGRAW-HILL INC., 1974, P 207-208
15. P.S. GILMAN, ALLIED-SIGNAL INC., PRIVATE COMMUNICATION, 2 FEB 1993
16. W.A. BAESLACK III, K.H. HOU, AND J.H. DEVLETIAN, RAPID SOLIDIFICATION JOINING OF A POWDER METALLURGY AL-FE-CE ALLOY, *J. MATER. SCI. LETT.,* VOL 7 (NO. 9), 1988, P 944-948
17. W.A. BAESLACK III AND S. KRISHNASWAMY, ELECTRON BEAM WELDABILITY OF A RAPIDLY SOLIDIFIED ALUMINUM ALLOY, *ADVANCES IN WELDING SCIENCE & TECHNOLOGY,* S.A. DAVID, ED., ASM INTERNATIONAL, 1986, P 357-362
18. S. KRISHNASWAMY AND W.A. BAESLACK III, STRUCTURE, PROPERTIES AND FRACTURE OF EB WELDS IN AL-8FE-2MO, *MATER. SCI. ENG.,* VOL 98, 1988, P 137-141
19. S. KRISHNASWAMY AND W.A. BAESLACK III, ND:YAG LASER WELDABILITY OF RS/PM AL-8FE-2MO SHEET, *RECENT TRENDS IN WELDING SCIENCE & TECHNOLOGY,* S.A. DAVID AND

J.M. VITEK, ED., ASM INTERNATIONAL, 1990, P 631-636

20. S. KRISHNASWAMY, WELDABILITY OF A RAPID SOLIDIFICATION POWDER METALLURGY AL-8FE-2MO ALLOY, PH.D. DISSERTATION, THE OHIO STATE UNIVERSITY, MARCH 1989
21. W.A. BAESLACK III, MICROSTRUCTURAL TRANSITIONS IN A WELDED POWDER METALLURGY AL-8FE-1.7NI ALLOY, *METALLOGRAPHY,* VOL 18, 1985, P 73-82
22. S. KRISHNASWAMY AND W.A. BAESLACK III, METALLOGRAPHY OF A PULSED ND:YAG LASER WELD IN A RS/PM AL-8FE-2MO SHEET, *MATER. CHARACT.,* VOL 24 (NO. 4), 1990, P 331-353
23. H. JONES, *RAPID SOLIDIFICATION OF METALS AND ALLOYS,* MONOGRAPH 8, INSTITUTION OF METALLURGISTS, LONDON, 1982, P 45
24. J.H. PEREPEZKO AND S.E. LEBEAU, NUCLEATION OF AL DURING SOLIDIFICATION, *ALUMINUM TRANSFORMATION TECHNOLOGY AND APPLICATIONS--1981,* C.A. PAMPILLO, H. BILONI, L.F. MONDOLFO, AND F. SACCHI, ED., AMERICAN SOCIETY FOR METALS, 1982, P 309-346
25. H. JONES, OBSERVATIONS ON A STRUCTURAL TRANSITION IN ALUMINUM ALLOYS HARDENED BY RAPID SOLIDIFICATION, *MATER. SCI. ENG.,* 1969-70, P 1-18
26. L.F. MONDOLFO, ALUMINUM-IRON SYSTEM, *ALUMINUM ALLOYS: STRUCTURE AND PROPERTIES,* BUTTERWORTHS, LONDON, 1976, P 283-289
27. W.A. BAESLACK III AND K. HAGEY, INERTIA FRICTION WELDING OF RAPIDLY SOLIDIFIED POWDER METALLURGY ALUMINUM, *WELD. J.,* VOL 67 (NO. 8), 1988, P 139S-149S
28. H.H. KOO, S. KRISHNASWAMY, AND W.A. BAESLACK III, SOLID-PHASE WELDING OF A DISPERSION-STRENGTHENED AL-FE-V-SI ALLOY-FVS1212, *LOW DENSITY, HIGH STRENGTH POWDER METALLURGY ALLOYS,* W.E. FRAZIER, M.J. KOCZAK, AND P.W. LEE, ED., THE METALLURGICAL SOCIETY OF AIME, 1991, P 183-196
29. H.H. KOO AND W.A. BAESLACK III, FRICTION WELDING OF A RAPIDLY SOLIDIFIED AL-FE-V-SI ALLOY, *WELD. J.,* VOL 71 (NO. 5), 1992, P 147S-169S
30. K.H. HOU AND W.A. BAESLACK III, ELECTRON MICROSCOPY OF INERTIA-FRICTION WELDMENTS IN A RAPIDLY SOLIDIFIED AL-FE-MO-V ALLOY, *J. MATER. SCI.,* VOL 25, 1990, P 2642-2653
31. V. ANANTHANARAYANAN, "DIFFUSION WELDING OF AN RS/PM ALUMINUM ALLOY," PH.D. DISSERTATION, THE OHIO STATE UNIVERSITY, 1988

Selection and Weldability of Aluminum-Lithium Alloys

C.E. Cross and W.T. Tack, Martin Marietta Astronautics Group

Introduction

ALUMINUM-LITHIUM alloys constitute the most recently developed group of high-performance wrought aluminum alloys intended for use primarily in aircraft and aerospace structures. Because aluminum alloys account for a large portion of the structural weight of most aerospace structures, designers can use the higher elastic modulus, lower density, and, in some cases, higher strength and fracture toughness of aluminum-lithium alloys to reduce structural weight. Another benefit is that the use of aluminum-lithium alloys does not require changes in manufacturing equipment, design methods, or fabrication techniques. In particular, aluminum-lithium alloys have been found to be readily weldable, a characteristic that is uncommon with other high-strength aluminum alloys such as 7075 and 2024.

The purpose of this article is to provide a guide to the welding of aluminum-lithium alloys that are now available commercially. Tabular data are provided for alloy composition, density, and modulus; tensile properties of plate; and tensile properties of weldments. There is limited information in the open literature regarding many of the newer alloys; therefore, some of these tables are incomplete. Weld properties are also characterized in terms of aging response, microstructure, and corrosion resistance.

A detailed and more thorough discussion of aluminum-lithium alloy weldability can be found in two review articles by Pickens (Ref 1, 2). An attempt is made here to give a very concise overview of the subject, including pertinent information from more recent conference proceedings not available for publication in previous reviews. With the fast-accumulating bank of information regarding aluminum-lithium weldments, design engineers can now begin to select and utilize aluminum-lithium alloys based on well-established behavior and property data.

References

1. J.R. PICKENS, THE WELDABILITY OF LITHIUM-CONTAINING ALUMINUM ALLOYS, *J. MATER. SCI.*, VOL 20, 1985, P 4247-4258
2. J.R. PICKENS, RECENT DEVELOPMENTS IN THE WELDABILITY OF LITHIUM-CONTAINING ALUMINUM ALLOYS, *J. MATER. SCI.*, VOL 25, 1990, P 3035-3047

Selection and Weldability of Aluminum-Lithium Alloys

C.E. Cross and W.T. Tack, Martin Marietta Astronautics Group

Commercial Alloys

Classification. Aluminum-lithium alloys can be divided into two groups: aluminum-lithium-magnesium ternary and aluminum-lithium-copper-magnesium quaternary alloy systems. These alloy groups are shown schematically in Fig. 1 with accompanying alloy designations and primary strengthening phases associated with each alloy. In addition to the solid-solution strengthening effects of alloying additions, the important strengthening phases δ' (Al_3Li), S' (Al_2CuMg), and T_1 (Al_2CuLi) are formed upon aging and are dependent on alloy composition (Ref 3, 4).

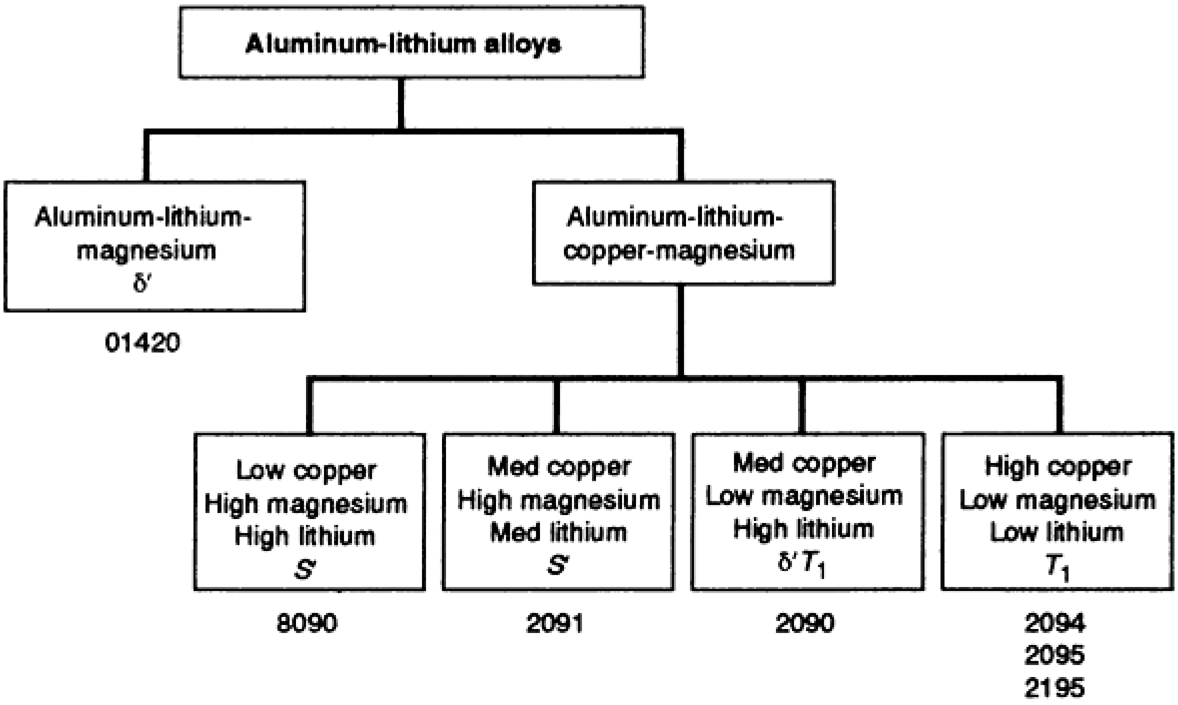
Aluminum-lithium alloys
Aluminum-lithium-magnesium δ'
01420
Aluminum-lithium-copper-magnesium
Low copper High magnesium High lithium S'
8090
Med copper High magnesium Med lithium S'
2091
Med copper Low magnesium High lithium δ' T1
2090
High copper Low magnesium Low lithium T1
2094
2095
2195

FIG. 1 SCHEMATIC OF TERNARY AND QUATERNARY ALUMINUM-LITHIUM ALLOY SYSTEMS SHOWING PRIMARY STRENGTHENING PHASES AND CORRESPONDING COMMERCIAL ALLOY DESIGNATIONS

Properties. Of the specific alloy compositions shown in Table 1, the high copper/magnesium ratio of alloy 2090 and the Weldalite 049 family of alloys (2094, 2095, 2195) results in the precipitation of T_1, which is the most potent strengthening phase in this alloy system (Ref 4, 5, 6). Accordingly, these alloys possess higher yield strengths than other aluminum-lithium alloys (Table 2). At lower copper/magnesium ratios, the *S*' phase is favored, and lower strengths are achieved (e.g., alloys 8090 and 2091). In the absence of copper, the δ' phase becomes predominant, resulting in even lower strength (e.g., Soviet alloy 01420). Density increases with copper content and decreases with lithium and magnesium contents; thus, alloy 01420 has the lowest density and the Weldalite 049 alloys have the highest densities. The elastic modulus tends to increase with lithium content; alloys 8090 and 2090 have the highest moduli in the aluminum-lithium-copper-magnesium alloy group.

TABLE 1 NOMINAL COMPOSITIONS AND SELECTED PROPERTIES OF ALUMINUM-LITHIUM ALLOYS

ALLOY	**COMPOSITION, WT%$^{(A)}$**					**DENSITY**		**MODULUS OF ELASTICITY**		SPECIFIC MODULUS, GPA · CM^3/G
	Li	**Cu**	**Mg**	**Zr**	**Other**	**gr/cm^3**	**lb/in^3**	**gpa**	**10^6 psi**	
01420	2.0	. . .	5.3	. . .	0.5 MN	2.47	0.089	73.8	10.7	**29.9**
8090	2.5	1.0	1.0	0.10	. . .	2.54	0.091	79.3	11.5	**31.2**
2091	2.0	2.2	1.5	0.10	. . .	2.58	0.093	77.9	11.3	**30.2**
2090	2.3	2.7	0.2	0.12	. . .	2.57	0.093	78.6	11.4	**30.6**
2094	1.3	4.7	0.4	0.14	0.4 AG	2.71	0.098	77.9	11.3	**28.7**
2095	1.3	4.3	0.4	0.14	0.4 AG	. . .	. . .	. . .	. . .	**. . .**
2195	**1.0**	**4.0**	**0.4**	**0.14**	**0.4 AG**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**

Source: Ref 3, 4, 5

(A) BALANCE ALUMINUM.

TABLE 2 LONGITUDINAL TENSILE PROPERTIES OF ALUMINUM-LITHIUM ALLOY PLATE

ALLOY AND TEMPER	**YIELD STRENGTH**		**ULTIMATE STRENGTH**		REF
	mpa	**ksi**	**mpa**	**ksi**	
01420-T6	280	41	460	67	**7**
8090-T6	429	62	504	73	**8**
2091-T8	440	64	480	70	**9**
2090-T8	518	75	559	81	**10**
2195-T8	**585**	**86**	**612**	**90**	**11**

Zirconium is added to most aluminum-lithium alloys to form β'-Al_3Zr or α'-$Al_3(Zr,Li)$ dispersoids; these dispersoids serve to inhibit recrystallization and nucleate the δ' phase giving improved strength and toughness (Ref 3, 12, 13). However, the presence of these dispersoids also leads to an inherent tendency toward texturing and subsequent anisotropic behavior (Ref 14). The β' phase may also act to promote grain refinement during solidification in both ingots and welds (Ref 15). Silver is added to the Weldalite 049 alloys to assist in precipitate nucleation (Ref 2).

Applications. Because of their good weldability, high strength, and accompanying high fracture toughness, the Weldalite 049 alloys have been considered as candidates to replace alloy 2219 in welded propellant tank applications, for example, the Space Shuttle external tank (Ref 14, 16). In this case, the weight savings gained from using a high-strength alloy (enabling the use of thinner sections) allows for a larger launch payload, which, in turn, offsets any additional cost in material. Weight savings can also be achieved by taking advantage of the high specific modulus of aluminum-lithium alloys (8090 and 2090 in particular) for use in stiffened or compression-loaded structures. Alloy 8090 has been successfully used in a welded tubular structure constructed with over 400 manual welds (Ref 17). Future applications may include welded skirt assemblies for space launch vehicles and other compression-loaded structures.

References cited in this section

2. J.R. PICKENS, RECENT DEVELOPMENTS IN THE WELDABILITY OF LITHIUM-CONTAINING ALUMINUM ALLOYS, *J. MATER. SCI.*, VOL 25, 1990, P 3035-3047
3. W.E. QUIST AND G.H. NARAYANAN, ALUMINUM-LITHIUM ALLOYS, *TREATISE MATER. SCI. TECH.*, VOL 31, MATERIALS AND COMPONENTS ENGINEERING PUBLISHING, LTD., 1989, P 219-254
4. T.J. LANGAN AND J.R. PICKENS, IDENTIFICATION OF STRENGTHENING PHASES IN THE AL-CU-LI ALLOY WELDALITE^{TM} 049, *PROC. 5TH INT. AL-LI CONF.*, 1989, P 691-700
5. REGISTRATION RECORD OF ALUMINUM ASSOCIATION DESIGNATIONS AND CHEMICAL COMPOSITION LIMITS FOR WROUGHT ALUMINUM ALLOYS, ALUMINUM ASSOCIATION, WASHINGTON, D.C., 1993
6. J.C. HUANG AND A.J. ARDELL, MICROSTRUCTURAL EVOLUTION IN TWO AL-LI-CU ALLOYS, *PROC. 3RD INT. AL-LI CONF.*, INSTITUTE OF METALS, 1986, P 455-469
7. I.N. FRIDLYANDER, AL-LI WELDABLE ALLOY 1420, *PROC. 5TH INT. AL-LI CONF.*, 1989, P 1359-1364
8. M.R. EDWARDS AND V.E. STONEHAM, THE FUSION WELDING OF AL-LI-CU-MG (8090) ALLOY, *J. PHYS.*, VOL 48, 1987, P 293-299
9. J. GLAZER AND J.W. MORRIS, THE STRENGTH-TOUGHNESS RELATIONSHIP AT CRYOGENIC TEMPERATURES IN AL-LI PLATE, *PROC. 5TH INT. AL-LI CONF.*, MATERIALS AND COMPONENTS ENGINEERING PUBLISHING, LTD., 1989, P 1471-1480
10. J.R. KERR AND R.E. MERINO, CRYOGENIC PROPERTIES OF VPPA WELDED AL-LI ALLOYS, *PROC. 5TH INT. AL-LI CONF.*, 1989, P 1491-1500
11. ALLOY 2195 DATA SHEET, REYNOLDS METALS CO., RICHMOND, VA, 1993
12. F.W. GAYLE AND J.B. VANDERSANDE, AL_3(LI,ZR) OR α' PHASE IN AL-LI-ZR SYSTEM, *PROC. 5TH INT. AL-LI CONF.*, INSTITUTE OF METALS, 1986, P 375-385
13. T.H. SANDERS AND E.A. STARKE, THE PHYSICAL METALLURGY OF ALUMINUM-LITHIUM ALLOYS--A REVIEW, *PROC. 5TH INT. AL-LI CONF.*, MATERIALS AND COMPONENTS ENGINEERING PUBLISHING, LTD., 1989, P 1-37
14. W.T. TACK AND L.W. LOECHEL, WELDALITE^{TM} 049: APPLICATION OF A NEW HIGH STRENGTH WELDABLE AL-LI-CU ALLOY, *PROC. 5TH INT. AL-LI CONF.*, MATERIALS AND COMPONENTS ENGINEERING PUBLISHING, LTD., 1989, P 1457-1467
15. L.F. MONDOLFO, GRAIN REFINEMENT IN THE CASTING OF NON-FERROUS ALLOYS, *PROC. CONF. GRAIN REFINEMENT IN CASTINGS AND WELDS*, TMS-AIME, 1982, P 3-50
16. W.T. TACK AND L.W. LOECHEL, USE OF APPLIED FRACTURE MECHANICS PRINCIPLES TO EVALUATE THE CRYOGENIC FRACTURE TOUGHNESS OF ALLOY 2090, *PROC. 5TH INT. AL-LI CONF.*, DEUTSCHE GESELLSCHAFT FUR MATERIALKUNDE, 1991, P 415-420
17. E. WILLNER AND D.L. YANEY, LOCKHEED MISSILES & SPACE CO., PRIVATE COMMUNICATION, 1993

Selection and Weldability of Aluminum-Lithium Alloys

C.E. Cross and W.T. Tack, Martin Marietta Astronautics Group

Weldability

Porosity. Weld metal porosity can form in aluminum alloys when monatomic hydrogen is partitioned interdendritically during solidification. A sufficient amount of hydrogen must be partitioned so that the interdendritic liquid becomes supersaturated, thus increasing the drive for pore nucleation. In order for diatomic gas pores to form, pore nuclei must overcome atmospheric pressure, surface tension, and liquid head pressure (Ref 18). Lithium-containing aluminum alloys exhibit a higher than normal propensity for weld metal porosity, although this does not appear to be due to any inherent thermodynamic property. Lithium has been shown to increase the solubility of hydrogen in aluminum for both solid and liquid phases such that less hydrogen is partitioned during solidification in comparison with other aluminum alloys (Ref 19). The observed tendency to form porosity in aluminum-lithium alloys may, however, be due to higher initial amounts of hydrogen in the weld pool, making it easier to reach the threshold concentration needed for pore nucleation.

Hydrogen can enter the weld pool from the base plate, from hydrated oxides or hydrocarbons on the plate surface, and from moisture in the shielding gas. Similar in behavior to aluminum-magnesium alloys, aluminum-lithium alloys are particularly susceptible to the formation of hydrated oxides (Ref 20). Thus, preweld joint preparation requires special attention to avoid porosity. A thorough removal of mill scale at the weld joint, by means of dry machining or chemical milling, is required. In a review of Soviet literature (Ref 1), it was reported that a marked drop in porosity in alloy 01420 weldments could be achieved by removing 0.05 mm (0.002 in.) from the joint surface prior to welding. It has become common practice in the welding of all aluminum-lithium alloys to dry machine approximately 0.25 mm (0.01 in.) off abutting surfaces as well as top and bottom plate surfaces, 25 to 50 mm (1 to 2 in.) from the centerline, on both sides of the joint. It is also considered good practice to weld within a few hours of machining and to degrease immediately prior to welding.

Another factor affecting weld quality and porosity entrapment is the weld cover gas, which is typically argon, helium, or a mixture of the two. It has been determined that low levels of oxygen impurity (less than 300 ppm) must be maintained to produce acceptable welds in alloy 2090 (Ref 21). In this same study, normal welding-grade bottled gas was found to vary from between 100 and 300 ppm O. Thus, in-line gas purifiers (getter or sieve type) or oxygen sensors at the welding torch may be necessary if shielding gas of consistently high quality cannot be supplied. Also, when making through-thickness welds, as in a keyhole welding process, backside shielding must be used to prevent excessive oxidation and the occurrence of porosity.

Hot cracking occurs when low-melting, eutectic liquid films are torn apart at the trailing edge of the weld pool during solidification. The hot cracking susceptibility of any aluminum-lithium alloy, or any aluminum alloy in general, appears to be related to its copper and magnesium contents and may involve an aluminum-copper-magnesium eutectic (Ref 22). Alloys with high copper and low magnesium (e.g., alloy 2219) are weldable, as are alloys with low copper and high magnesium (e.g., alloy 5083). It is alloys containing large amounts of both copper and magnesium--for example, conventional high-strength aerospace alloys such as 2024 and 7075--that encounter problems with weldability. With aluminum-lithium alloys, the formation of strengthening precipitates such as T_1 does not require large quantities of both copper and magnesium, and thus weldability can be maintained without compromising strength.

The relationship between copper and magnesium contents and weldability is demonstrated in the weldability index of Fig. 2. This index was obtained by superimposing modern alloy compositions on an existing crack susceptibility (castability) diagram generated by Pumphrey and Moore in 1948 for aluminum-copper-magnesium ternary alloys (Ref 23). In Fig. 2, it is predicted that aluminum-lithium alloys with high copper and low magnesium contents (e.g., alloys 2094 and 2090) should have good weldability. On the other hand, alloy 2091, an aluminum-lithium alloy with intermediate amounts of both copper and magnesium, is predicted to be less weldable. Weldability predictions made using this index do not account for the effect lithium has on copper or magnesium solubility, nor do they account for the use of filler alloys to modify weld metal composition. Nevertheless, this index has proven useful in aluminum-lithium base metal and filler alloy development by predicting trends in weldability behavior with variations in alloy composition.

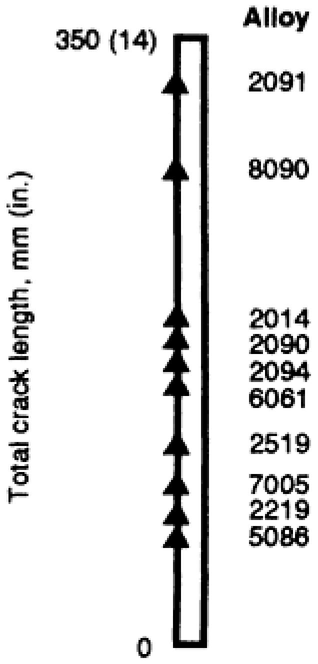

FIG. 2 PUMPHREY-MOORE CRACKING INDEX BASED ON CASTABILITY DATA OBTAINED FROM ALUMINUM-COPPER-MAGNESIUM ALLOYS. SOURCE: REF 25

Filler Metal Selection. Filler metal alloys used to join aluminum-lithium alloys are selected to give acceptable weldability and, at the same time, to optimize mechanical properties. Alloy 2319 has typically been used to join the high-copper alloys (e.g., alloys 2090, 2094, 2095, and 2195). Low-copper, high-magnesium alloys (e.g., alloy 8090) call for the use of a high-magnesium filler metal such as alloy 5356. Alloys that are difficult to weld (e.g., alloy 2091) require the use of an aluminum-silicon filler metal such as alloy 4043. Aluminum-silicon filler metal alloys provide excellent weldability, but are not normally used due to their characteristic low toughness and low ductility and their inability to respond to aging. Experimental aluminum-lithium filler metal alloys have been developed and show promise for future applications where high filler metal dilutions are required (Ref 24).

In a recent weldability study, a trans-Varestraint test was used to compare aluminum-lithium alloys with conventional aluminum alloys (Ref 25). Tests were performed using an applied strain of 0.5%, appropriate filler metal alloys at 30% filler dilution, and a gas-tungsten arc welding process operated at welding currents selected to achieve constant penetration for each alloy tested. Results of this study are shown in Fig. 3; alloy 2219 appears to be the most weldable and alloys 2090 and 2094 the least weldable of those alloys investigated. However, extensive weld development programs (Ref 10, 14, 26) have demonstrated that these aluminum-lithium alloys can be easily welded in practice with little concern for cracking. Weldability tests can provide a relative rating among alloys; they cannot be used to predict real-world behavior.

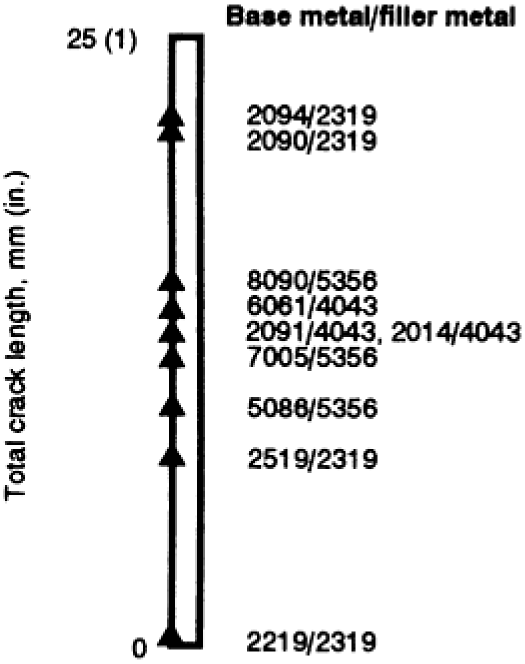

FIG. 3 TRANS-VARESTRAINT WELDABILITY DATA FOR ALUMINUM ALLOYS. SOURCE: REF 25

As shown in Fig. 4, the weldability of alloys 2090 and 2094 can be improved using high-silicon filler metal alloys (e.g., alloy 4047), although this is not normally deemed necessary or desirable (Ref 25, 27). Figure 4(b) shows that the use of a 2319 filler metal results in less cracking than a 4043 filler metal when welding alloy 2090. The same is not true when welding alloy 2094 (Fig. 4a). Alloy 2090 is much leaner in copper than alloy 2094, and thus it responds more to the use of a high-copper filler metal alloy. It should be noted that results from two different weldability tests are displayed in Fig. 4, and the magnitude of cracking is different for each test.

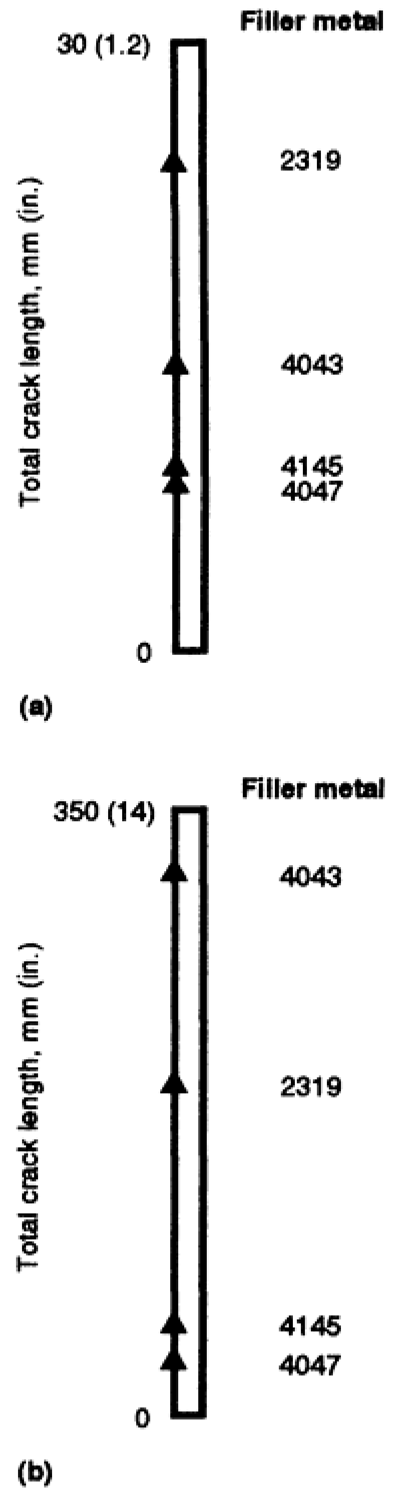

FIG. 4 WELDABILITY DATA SHOWING THE IMPROVED RESISTANCE TO HOT CRACKING OBTAINED WHEN USING AN ALUMINUM-SILICON FILLER ALLOY. (A) TRANS-VARESTRAINT TEST DATA FOR ALLOY 2094 WELDMENTS. SOURCE: REF 25. (B) INVERTED-TEE TEST DATA FOR ALLOY 2090 WELDMENTS. SOURCE: REF 27

References cited in this section

1. J.R. PICKENS, THE WELDABILITY OF LITHIUM-CONTAINING ALUMINUM ALLOYS, *J. MATER. SCI.*, VOL 20, 1985, P 4247-4258

10. J.R. KERR AND R.E. MERINO, CRYOGENIC PROPERTIES OF VPPA WELDED AL-LI ALLOYS, *PROC. 5TH INT. AL-LI CONF.*, 1989, P 1491-1500
14. W.T. TACK AND L.W. LOECHEL, WELDALITE^{TM} 049: APPLICATION OF A NEW HIGH STRENGTH WELDABLE AL-LI-CU ALLOY, *PROC. 5TH INT. AL-LI CONF.*, MATERIALS AND COMPONENTS ENGINEERING PUBLISHING, LTD., 1989, P 1457-1467
18. R.E. TREVISAN, D.D. SCHWEMMER, AND D.L. OLSON, THE FUNDAMENTALS OF WELD METAL PORE FORMATION, *WELDING*: *THEORY AND PRACTICE*, ELSEVIER, 1990
19. D.E.J. TALBOT, INFLUENCES OF HYDROGEN ON QUALITY OF ALUMINUM AND ALUMINUM ALLOY INGOTS, *PROC. 3RD INT.* SOLIDIFICATION PROCESSING CONF., INSTITUTE OF METALS, 1987, P 29-31
20. D.J. FIELD, OXIDATION OF ALUMINUM AND ITS ALLOYS, *TREATISE MATER. SCI. TECH.*, VOL 31, 1989, P 523-537
21. L.F. MARINEZ, J.C. MCCLURE, AND A.C. NUNES, THE EFFECT OF GAS CONTAMINATION DURING WELDING OF ALUMINUM-LITHIUM ALLOYS, *PROD. ENG. DES.* ASME, VOL 51, 1991, P 235-243
22. C.E. CROSS, L.S. KRAMER, W.T. TACK, AND L.W. LOECHEL, ALUMINUM WELDABILITY AND HOT TEARING THEORY, *WELDABILITY OF MATERIALS*, ASM INTERNATIONAL, 1990, P 275-282
23. W.I. PUMPHREY AND D.C. MOORE, CRACKING DURING AND AFTER SOLIDIFICATION IN SOME ALUMINUM-COPPER-MAGNESIUM ALLOYS OF HIGH PURITY, *J. INST. MET.*, VOL 73, 1948, P 425-438
24. L.S. KRAMER, C.E. CROSS, AND J.R. PICKENS, THE EFFECT OF COMPOSITION ON THE WELDABILITY OF AL-CU-LI-AG-MG ALLOYS IN THE HIGH CU:LI REGIME, *PROC. 6TH INT. AL-LI CONF.*, DEUTSCHE GESELLSCHAFT FUR MATERIALKUNDE, 1991, P 1197-1202
25. C.E. CROSS AND W.T. TACK, "FACTORS AFFECTING THE WELDABILITY EVALUATION OF ALUMINUM ALLOYS," PRESENTED AT THE 73RD AMERICAN WELDING SOCIETY CONVENTION (CHICAGO), 1992
26. E.F. SCHOLZ, L.W. LOECHEL, AND M.O. ROBERTS, ADVANCED CRYOGENIC PROPELLENT TANK DEVELOPMENT STATUS, *PROC. 28TH AIAA PROPULSION CONF.*, 1992, P 1-11
27. R.P. MARTUKANITZ, C.A. NATALIE, AND J.O. KNOEFEL, THE WELDABILITY OF AN AL-LI-CU ALLOY, *J. MET.*, VOL 39, 1987, P 38-42

Selection and Weldability of Aluminum-Lithium Alloys

C.E. Cross and W.T. Tack, Martin Marietta Astronautics Group

Weld Characterization

Joint Strength. Although the tensile properties of many wrought aluminum-lithium alloys are exceptionally high, the strength of the welded joints in these alloys is limited by the strength of the solidified weld metal. Even though joint strengths are higher than those found in conventional aluminum alloys, joint efficiencies on the order of 50% are to be expected for the high-strength aluminum-lithium alloys.

A comparison of gas-tungsten arc, variable polarity plasma arc, and gas-metal arc weldment properties for aluminum-lithium alloys, taken from numerous sources, is given in Table 3. Alloy 2094 is found to have the highest joint strength (372 MPa, or 54 ksi, ultimate strength), which corresponds to a 38% improvement over alloy 2219 weldments (269 MPa, or 39 ksi, ultimate strength) (Ref 10). Alloy 01420 is found to have the lowest weld strength (241 MPa, or 35 ksi, ultimate strength).

TABLE 3 TRANSVERSE TENSILE PROPERTIES FOR ARC WELDED ALUMINUM-LITHIUM WELDMENTS

IN THE AS-WELDED CONDITION

ALLOY AND TEMPER	WELDING PROCESS$^{(A)}$	FILLER METAL	YIELD STRENGTH		ULTIMATE STRENGTH		REF
			MPA	KSI	MPA	KSI	
01420-T6	GTAW	5356	103	15	241	35	**28**
2090-T8	GTAW	2319	165	24	248	36	**29**
	VPPAW	2319	165	24	283	41	**10**
	GMAW	2319	207	30	234	34	**27**
8090-T8	GTAW	5356	179	26	310	45	**29**
	VPPAW	5356	193	28	296	43	**29**
	GMAW	5356	186	27	276	40	**30**
2094-T8	GTAW	2319	. . .	. . .	352	51	**31**
	VPPAW	2319	269	39	372	54	**32**
	GMAW	**2319**	**. . .**	**. . .**	**283**	**41**	**31**

(A) GTAW, GAS-TUNGSTEN ARC WELDING; VPPAW, VARIABLE-POLARITY PLASMA ARC WELDING; GMAW, GAS-METAL ARC WELDING

High-energy-density welding processes have also been used successfully to join aluminum-lithium alloys. Electron-beam welding has been used to join alloys 8090, 2090, and 2094; tensile strength results for electron-beam weldments are given in Table 4. Typically, a slight improvement in strength above arc welding properties is to be expected because of a finer microstructure, narrower fusion zone, and better response to natural aging. Similarly, good results for laser-beam-welded alloy 2090 have been reported (Ref 34, 35).

TABLE 4 TRANSVERSE TENSILE PROPERTIES FOR ELECTRON-BEAM-WELDED ALUMINUM-LITHIUM ALLOYS IN THE AS-WELDED CONDITION

ALLOY AND TEMPER	ULTIMATE STRENGTH		REF
	MPA	KSI	
8090-T6	317	46	**33**
2090-T8	324	47	**27**
2094-T8	**434**	**63**	**2**

Postweld Aging. The Weldalite 049 alloys have been observed to naturally age in both the weld metal and the heat-affected zone (HAZ). It is important to wait a sufficient time (typically 30 days) between welding and proof or tensile testing to allow for precipitate stabilization, thereby ensuring accurate characterization of joint properties. In addition, postweld solution heat treatment and artificial aging, although not practical for large structures, have been found to significantly increase the strength of high-copper alloy gas-tungsten arc weldments; by using these postweld treatments, an ultimate tensile strength of 400 MPa (58 ksi) has been obtained with alloy 2090 (Ref 36) and 510 MPa (73 ksi) with alloy 2094 (Ref 37).

Ductility and Toughness. If the weld metal is significantly lower in strength than the base metal, when the weldment is plastically deformed, most of the strain will be concentrated in the weld metal. Such is the case for high-strength aluminum-lithium alloys, where hardness traverses for both alloy 2090 and 2094 have confirmed a steep gradient in strength in the HAZ between weld metal and base metal (Ref 31, 38). Typically, joint ductility measured over a 50 mm (2 in.) gage length will be 2 to 4% even though the actual weld metal ductility approaches 10 to 12%.

Fracture toughness is an important criterion in the design of welded tanks for aerospace applications. Also, an increase in toughness with decreasing temperature is particularly desirable for cryogenic tanks because it allows proof testing to be performed at room temperature, a much more feasible and economical procedure. Weld toughness values have been measured for alloy 2090 gas-tungsten arc (dual torch, variable polarity) weldments using a surface crack tension test. K_{Ie} values increasing from 30 MPa/ $\sqrt{m}$ at room temperature to 34 MPa/ $\sqrt{m}$ at 73 K have been reported, which is an improvement over similar measurements for alloy 2219 (19 to 24 MPa/ $\sqrt{m}$) (Ref 31). It should be noted that these are not plane-strain toughness values, and their magnitude is sensitive to precrack geometry.

Fatigue. Initial results from an ongoing study involving alloy 2095 weldments (Ref 39) have shown that fatigue growth resistance is higher in both the weld metal and HAZ when compared with the base metal. For precracks placed in the HAZ, there is a tendency for cracks to grow to the fusion line, where the fatigue crack growth rate is greater. In both the weld metal and HAZ (unconstrained condition), the general tendency is for the fatigue crack growth rate to increase with increasing stress ratio ($\sigma_{min}/\sigma_{max}$).

Microstructure. The high strength of aluminum-lithium weld metal is due to a combination of fine grain size and precipitate formation. This unexplained tendency for fine grains, as small as 10 μm in diameter (equiaxed), has been noted elsewhere (Ref 40). Constitutional undercooling by lithium has been ruled out as a possible cause (Ref 41). The tensile fracture of alloy 2094 weldments has been found to be intergranular, associated with a band of fine grains located along the fusion line (Ref 31, 42). Based onphase equilibria, two eutectic constituents are likely present in aluminum-lithium-copper weld metal: $\theta(Al_2Cu)$ and T_B ($Al_{15}Cu_8Li_2$) (Ref 43). Weld metal precipitates δ' and β' have been identified in alloy 2090 in the as-welded condition, and platelike precipitates, believed to be T_1, have been observed upon aging (Ref 36).

Corrosion. A study involving corrosion potential measurements taken across weldments (through base metal, HAZ, and weld metal) has identified certain aluminum-lithium alloy compositions as being susceptible to galvanic attack in a saline environment (Ref 44). In particular, two experimental alloys with high lithium content (2.9 wt% Li and 3.0 wt% Cu), welded with either 2319 or 4043 fillers, displayed a narrow region within the HAZ that was highly anodic to both base metal and weld metal. This behavior was attributed to the formation of the equilibrium δ(AlLi) phase and resulted in severe pitting in the HAZ. In contrast, a 2090-type alloy showed a continuously increasing (cathodic) potential when going from base metal to weld metal and was resistant to pitting attack. This behavior was attributed to the absence of the δ phase due to a higher copper-lithium ratio.

References cited in this section

2. J.R. PICKENS, RECENT DEVELOPMENTS IN THE WELDABILITY OF LITHIUM-CONTAINING ALUMINUM ALLOYS, *J. MATER. SCI.*, VOL 25, 1990, P 3035-3047
10. J.R. KERR AND R.E. MERINO, CRYOGENIC PROPERTIES OF VPPA WELDED AL-LI ALLOYS, *PROC. 5TH INT. AL-LI CONF.*, 1989, P 1491-1500
27. R.P. MARTUKANITZ, C.A. NATALIE, AND J.O. KNOEFEL, THE WELDABILITY OF AN AL-LI-CU ALLOY, *J. MET.*, VOL 39, 1987, P 38-42
28. J.R. PICKENS, T.J. LANGAN, AND E. BARTA, THE WELDABILITY OF AL-5MG-2LI-0.1ZR ALLOY 01420, *PROC. 5TH INT. AL-LI CONF.*, INSTITUTE OF METALS, 1985, P 137-147
29. C.C. GRIFFEE, G.A. JENSEN, AND T.L. REINHART, FACTORS INFLUENCING THE QUALITY AND PROPERTIES OF AL-LI ALLOY WELDS, *PROC. 5TH INT. AL-LI CONF.*, MATERIALS AND COMPONENTS ENGINEERING PUBLISHING LTD., 1989, P 1425-1434
30. M.H. SKILLINGBERG, FUSION WELDING OF AL-LI-CU-(MG)-ZR PLATE, PROC. CONF. ALUMINUM TECHNOLOGY, INSTITUTE OF METALS, 1986, P 509-515
31. C.E. CROSS, L.W. LOECHEL, AND G.F. BRAUN, WELDALITE™ 049 WELD DEVELOPMENT FOR AEROSPACE TANKAGE, *PROC. 5TH INT. AL-LI CONF.*, DEUTSCHE GESELLSCHAFT FUR MATERIALKUNDE, 1991, P 1165-1170
32. M.O. ROBERTS AND L.W. LOECHEL, "WELDING PROCESS PROCEDURE DEVELOPMENT AND OPTIMIZATION USING DESIGN OF EXPERIMENTS," PRESENTED AT THE AMERICAN WELDING SOCIETY CONVENTION (HOUSTON), 1993
33. P. LE POAC, A.M. NOMINE, AND D. MIANNAY, MECHANICAL PROPERTIES OF BEAM ELECTRON WELDINGS IN 8090 ALLOY, *J. PHYS.*, VOL 48, 1987, P 301-306
34. P.A. MOLIAN AND T.S. STRIVATSAN, WELDABILITY OF AI-LI-CU ALLOY 2090, USING LASER WELDING, *PROC. 5TH INT. AL-LI CONF.*, MATERIALS AND COMPONENTS ENGINEERING PUBLISHING LTD., 1989, P 1435-1445
35. T.A. MARSICO AND R. KOSSOWSKY, PHYSICAL PROPERTIES OF LASER WELDED AL-LI ALLOY 2090, *PROC. 5TH INT. AL-LI CONF.*, 1989, P 1447-1456

36. A.J. SUNWOO, "WELDMENT MECHANICAL PROPERTIES OF AL-LI-CU ALLOY 2090 AT AMBIENT AND CRYOGENIC TEMPERATURES," PH.D. THESIS, UNIVERSITY OF CALIFORNIA-BERKELEY, 1990
37. L.S. KRAMER AND J.R. PICKENS, MICROSTRUCTURE AND PROPERTIES OF A WELDED AL-CU-LI ALLOY, *WELD. J.*, VOL 71, 1992, P 115S-121S
38. R.P. MARTUKANITZ, R.H. STEVENS, AND L.R. JOHNSTON, MICROSTRUCTURAL AND MICROHARDNESS CHARACTERIZATIONS OF A GAS METAL ARC WELDED ALUMINUM-LITHIUM ALLOY, *MICROSTRUCT. SCI.*, VOL 14, 1985, P 53-64
39. G.O. RADING, "FATIGUE CRACK GROWTH IN WELDED AL-LI ALLOY 2095 (WELDALITE^{TM} 049)," PH.D. THESIS PROGRESS REPORT, UNIVERSITY OF ALABAMA, 1992
40. C.E. CROSS, D.L. OLSON, G.R. EDWARDS, AND J.F. CAPES, WELDABILITY OF ALUMINUM-LITHIUM ALLOYS, *PROC. 2ND INT. AL-LI CONF.*, TMS-AIME, 1983, P 676-682
41. C.E. CROSS, G.R. EDWARDS, D.L. OLSON, AND R.H. FROST, INTRINSIC NUCLEATION OF WELD METAL GRAINS, *PROC. 3RD INT.* CONF. SOLIDIFICATION PROCESSING, INSTITUTE OF METALS, 1987, P 388-391
42. S.R. SHAH, J.E. WITTIG, AND G.T. HAHN, MICROSTRUCTURE ANALYSIS OF A HIGH STRENGTH AL-CU-LI (WELDALITE^{TM} 049) ALLOY WELD, *PROC. 3RD INT.* CONF. TRENDS IN WELDING RESEARCH, ASM INTERNATIONAL, 1992, P 281-285
43. K.A. MONTOYA, F.H. HEUBAUM, K.S. KUMAR, AND J.R. PICKENS, COMPOSITIONAL EFFECTS ON THE SOLIDUS TEMPERATURE OF AN AL-CU-LI-AG-MG ALLOY, *SCR. MET.,* VOL 25, 1991, P 1489-1494
44. G. BEVERINI, "INVESTIGATIONS OF THE CORROSION CHARACTERISTICS OF AL-LI AND AL-LI-CU WELDMENTS IN A 3.5% NACL SOLUTION," M.SC. THESIS T-3508, COLORADO SCHOOL OF MINES, 1988

Selection and Weldability of Aluminum-Lithium Alloys

C.E. Cross and W.T. Tack, Martin Marietta Astronautics Group

References

1. J.R. PICKENS, THE WELDABILITY OF LITHIUM-CONTAINING ALUMINUM ALLOYS, *J. MATER. SCI.*, VOL 20, 1985, P 4247-4258
2. J.R. PICKENS, RECENT DEVELOPMENTS IN THE WELDABILITY OF LITHIUM-CONTAINING ALUMINUM ALLOYS, *J. MATER. SCI.*, VOL 25, 1990, P 3035-3047
3. W.E. QUIST AND G.H. NARAYANAN, ALUMINUM-LITHIUM ALLOYS, *TREATISE MATER. SCI. TECH.,* VOL 31, MATERIALS AND COMPONENTS ENGINEERING PUBLISHING, LTD., 1989, P 219-254
4. T.J. LANGAN AND J.R. PICKENS, IDENTIFICATION OF STRENGTHENING PHASES IN THE AL-CU-LI ALLOY WELDALITE^{TM} 049, *PROC. 5TH INT. AL-LI CONF.,* 1989, P 691-700
5. REGISTRATION RECORD OF ALUMINUM ASSOCIATION DESIGNATIONS AND CHEMICAL COMPOSITION LIMITS FOR WROUGHT ALUMINUM ALLOYS, ALUMINUM ASSOCIATION, WASHINGTON, D.C., 1993
6. J.C. HUANG AND A.J. ARDELL, MICROSTRUCTURAL EVOLUTION IN TWO AL-LI-CU ALLOYS, *PROC. 3RD INT. AL-LI CONF.,* INSTITUTE OF METALS, 1986, P 455-469
7. I.N. FRIDLYANDER, AL-LI WELDABLE ALLOY 1420, *PROC. 5TH INT. AL-LI CONF.*, 1989, P 1359-1364
8. M.R. EDWARDS AND V.E. STONEHAM, THE FUSION WELDING OF AL-LI-CU-MG (8090) ALLOY, *J. PHYS.*, VOL 48, 1987, P 293-299

9. J. GLAZER AND J.W. MORRIS, THE STRENGTH-TOUGHNESS RELATIONSHIP AT CRYOGENIC TEMPERATURES IN AL-LI PLATE, *PROC. 5TH INT. AL-LI CONF.*, MATERIALS AND COMPONENTS ENGINEERING PUBLISHING, LTD., 1989, P 1471-1480
10. J.R. KERR AND R.E. MERINO, CRYOGENIC PROPERTIES OF VPPA WELDED AL-LI ALLOYS, *PROC. 5TH INT. AL-LI CONF.*, 1989, P 1491-1500
11. ALLOY 2195 DATA SHEET, REYNOLDS METALS CO., RICHMOND, VA, 1993
12. F.W. GAYLE AND J.B. VANDERSANDE, AL$_3$(LI,ZR) OR α' PHASE IN AL-LI-ZR SYSTEM, *PROC. 5TH INT. AL-LI CONF.*, INSTITUTE OF METALS, 1986, P 375-385
13. T.H. SANDERS AND E.A. STARKE, THE PHYSICAL METALLURGY OF ALUMINUM-LITHIUM ALLOYS--A REVIEW, *PROC. 5TH INT. AL-LI CONF.*, MATERIALS AND COMPONENTS ENGINEERING PUBLISHING, LTD., 1989, P 1-37
14. W.T. TACK AND L.W. LOECHEL, WELDALITE$^{\text{TM}}$ 049: APPLICATION OF A NEW HIGH STRENGTH WELDABLE AL-LI-CU ALLOY, *PROC. 5TH INT. AL-LI CONF.*, MATERIALS AND COMPONENTS ENGINEERING PUBLISHING, LTD., 1989, P 1457-1467
15. L.F. MONDOLFO, GRAIN REFINEMENT IN THE CASTING OF NON-FERROUS ALLOYS, *PROC. CONF. GRAIN REFINEMENT IN CASTINGS AND WELDS*, TMS-AIME, 1982, P 3-50
16. W.T. TACK AND L.W. LOECHEL, USE OF APPLIED FRACTURE MECHANICS PRINCIPLES TO EVALUATE THE CRYOGENIC FRACTURE TOUGHNESS OF ALLOY 2090, *PROC. 5TH INT. AL-LI CONF.*, DEUTSCHE GESELLSCHAFT FUR MATERIALKUNDE, 1991, P 415-420
17. E. WILLNER AND D.L. YANEY, LOCKHEED MISSILES & SPACE CO., PRIVATE COMMUNICATION, 1993
18. R.E. TREVISAN, D.D. SCHWEMMER, AND D.L. OLSON, THE FUNDAMENTALS OF WELD METAL PORE FORMATION, *WELDING: THEORY AND PRACTICE*, ELSEVIER, 1990
19. D.E.J. TALBOT, INFLUENCES OF HYDROGEN ON QUALITY OF ALUMINUM AND ALUMINUM ALLOY INGOTS, *PROC. 3RD INT.* SOLIDIFICATION PROCESSING CONF., INSTITUTE OF METALS, 1987, P 29-31
20. D.J. FIELD, OXIDATION OF ALUMINUM AND ITS ALLOYS, *TREATISE MATER. SCI. TECH.*, VOL 31, 1989, P 523-537
21. L.F. MARINEZ, J.C. MCCLURE, AND A.C. NUNES, THE EFFECT OF GAS CONTAMINATION DURING WELDING OF ALUMINUM-LITHIUM ALLOYS, *PROD. ENG. DES.* ASME, VOL 51, 1991, P 235-243
22. C.E. CROSS, L.S. KRAMER, W.T. TACK, AND L.W. LOECHEL, ALUMINUM WELDABILITY AND HOT TEARING THEORY, *WELDABILITY OF MATERIALS*, ASM INTERNATIONAL, 1990, P 275-282
23. W.I. PUMPHREY AND D.C. MOORE, CRACKING DURING AND AFTER SOLIDIFICATION IN SOME ALUMINUM-COPPER-MAGNESIUM ALLOYS OF HIGH PURITY, *J. INST. MET.*, VOL 73, 1948, P 425-438
24. L.S. KRAMER, C.E. CROSS, AND J.R. PICKENS, THE EFFECT OF COMPOSITION ON THE WELDABILITY OF AL-CU-LI-AG-MG ALLOYS IN THE HIGH CU:LI REGIME, *PROC. 6TH INT. AL-LI CONF.*, DEUTSCHE GESELLSCHAFT FUR MATERIALKUNDE, 1991, P 1197-1202
25. C.E. CROSS AND W.T. TACK, "FACTORS AFFECTING THE WELDABILITY EVALUATION OF ALUMINUM ALLOYS," PRESENTED AT THE 73RD AMERICAN WELDING SOCIETY CONVENTION (CHICAGO), 1992
26. E.F. SCHOLZ, L.W. LOECHEL, AND M.O. ROBERTS, ADVANCED CRYOGENIC PROPELLENT TANK DEVELOPMENT STATUS, *PROC. 28TH AIAA PROPULSION CONF.*, 1992, P 1-11
27. R.P. MARTUKANITZ, C.A. NATALIE, AND J.O. KNOEFEL, THE WELDABILITY OF AN AL-LI-CU ALLOY, *J. MET.*, VOL 39, 1987, P 38-42
28. J.R. PICKENS, T.J. LANGAN, AND E. BARTA, THE WELDABILITY OF AL-5MG-2LI-0.1ZR ALLOY 01420, *PROC. 5TH INT. AL-LI CONF.*, INSTITUTE OF METALS, 1985, P 137-147
29. C.C. GRIFFEE, G.A. JENSEN, AND T.L. REINHART, FACTORS INFLUENCING THE QUALITY AND PROPERTIES OF AL-LI ALLOY WELDS, *PROC. 5TH INT. AL-LI CONF.*, MATERIALS AND

COMPONENTS ENGINEERING PUBLISHING LTD., 1989, P 1425-1434

30. M.H. SKILLINGBERG, FUSION WELDING OF AL-LI-CU-(MG)-ZR PLATE, PROC. CONF. ALUMINUM TECHNOLOGY, INSTITUTE OF METALS, 1986, P 509-515
31. C.E. CROSS, L.W. LOECHEL, AND G.F. BRAUN, WELDALITE™ 049 WELD DEVELOPMENT FOR AEROSPACE TANKAGE, *PROC. 5TH INT. AL-LI CONF.*, DEUTSCHE GESELLSCHAFT FUR MATERIALKUNDE, 1991, P 1165-1170
32. M.O. ROBERTS AND L.W. LOECHEL, "WELDING PROCESS PROCEDURE DEVELOPMENT AND OPTIMIZATION USING DESIGN OF EXPERIMENTS," PRESENTED AT THE AMERICAN WELDING SOCIETY CONVENTION (HOUSTON), 1993
33. P. LE POAC, A.M. NOMINE, AND D. MIANNAY, MECHANICAL PROPERTIES OF BEAM ELECTRON WELDINGS IN 8090 ALLOY, *J. PHYS.*, VOL 48, 1987, P 301-306
34. P.A. MOLIAN AND T.S. STRIVATSAN, WELDABILITY OF AI-LI-CU ALLOY 2090, USING LASER WELDING, *PROC. 5TH INT. AL-LI CONF.*, MATERIALS AND COMPONENTS ENGINEERING PUBLISHING LTD., 1989, P 1435-1445
35. T.A. MARSICO AND R. KOSSOWSKY, PHYSICAL PROPERTIES OF LASER WELDED AL-LI ALLOY 2090, *PROC. 5TH INT. AL-LI CONF.*, 1989, P 1447-1456
36. A.J. SUNWOO, "WELDMENT MECHANICAL PROPERTIES OF AL-LI-CU ALLOY 2090 AT AMBIENT AND CRYOGENIC TEMPERATURES," PH.D. THESIS, UNIVERSITY OF CALIFORNIA-BERKELEY, 1990
37. L.S. KRAMER AND J.R. PICKENS, MICROSTRUCTURE AND PROPERTIES OF A WELDED AL-CU-LI ALLOY, *WELD. J.*, VOL 71, 1992, P 115S-121S
38. R.P. MARTUKANITZ, R.H. STEVENS, AND L.R. JOHNSTON, MICROSTRUCTURAL AND MICROHARDNESS CHARACTERIZATIONS OF A GAS METAL ARC WELDED ALUMINUM-LITHIUM ALLOY, *MICROSTRUCT. SCI.*, VOL 14, 1985, P 53-64
39. G.O. RADING, "FATIGUE CRACK GROWTH IN WELDED AL-LI ALLOY 2095 (WELDALITE™ 049)," PH.D. THESIS PROGRESS REPORT, UNIVERSITY OF ALABAMA, 1992
40. C.E. CROSS, D.L. OLSON, G.R. EDWARDS, AND J.F. CAPES, WELDABILITY OF ALUMINUM-LITHIUM ALLOYS, *PROC. 2ND INT. AL-LI CONF.*, TMS-AIME, 1983, P 676-682
41. C.E. CROSS, G.R. EDWARDS, D.L. OLSON, AND R.H. FROST, INTRINSIC NUCLEATION OF WELD METAL GRAINS, *PROC. 3RD INT.* CONF. SOLIDIFICATION PROCESSING, INSTITUTE OF METALS, 1987, P 388-391
42. S.R. SHAH, J.E. WITTIG, AND G.T. HAHN, MICROSTRUCTURE ANALYSIS OF A HIGH STRENGTH AL-CU-LI (WELDALITE™ 049) ALLOY WELD, *PROC. 3RD INT.* CONF. TRENDS IN WELDING RESEARCH, ASM INTERNATIONAL, 1992, P 281-285
43. K.A. MONTOYA, F.H. HEUBAUM, K.S. KUMAR, AND J.R. PICKENS, COMPOSITIONAL EFFECTS ON THE SOLIDUS TEMPERATURE OF AN AL-CU-LI-AG-MG ALLOY, *SCR. MET.,* VOL 25, 1991, P 1489-1494
44. G. BEVERINI, "INVESTIGATIONS OF THE CORROSION CHARACTERISTICS OF AL-LI AND AL-LI-CU WELDMENTS IN A 3.5% NACL SOLUTION," M.SC. THESIS T-3508, COLORADO SCHOOL OF MINES, 1988

Selection and Weldability of Aluminum Metal-Matrix Composites

Thomas J. Lienert, The Ohio State University; Charles T. Lane, Duralcan USA; Jerry E. Gould, Edison Welding Institute

Introduction

THE PURPOSE of this article is to review the current understanding of the joinability of aluminum metal-matrix composites (Al-MMCs).* Unfortunately, the joining of Al-MMCs is not a mature technology, and many important details are not well understood. Relatively little work has been directed toward furthering the understanding of Al-MMC joining. Much of the work has been either proprietary or protected for national security reasons and is not available in the open literature.

Although much of the early work on Al-MMCs, including that on joining, concentrated on continuous fiber types, most of the present work is focused on discontinuously reinforced (particle or whisker) Al-MMCs, because of their greater ease of manufacture, lower production costs, and relatively isotropic properties. The most commonly used reinforcement materials in discontinuously reinforced aluminum (DRA) composites are silicon carbide (SiC) and aluminum oxide (Al_2O_3). Consequently, this article focuses on composites that incorporate these specific materials. A review of Al-MMC joining technology up to 1975 is provided in Ref 1.

Reference

1. G.E. METZGER, JOINING OF METAL-MATRIX FIBER-REINFORCED COMPOSITE MATERIALS, *WELD. RES. COUNC. BULL.,* NO. 207, 1975

Note

* * AL-MMCS ARE NORMALLY SPECIFIED ACCORDING TO A NOMENCLATURE DEVELOPED BY THE ALUMINUM ASSOCIATION (AA) AND LATER ADOPTED BY THE AMERICAN NATIONAL STANDARD INSTITUTE (ANSI 35.5-1992). IN THIS SPECIFICATION, THE MATRIX ALLOY IS LISTED FIRST, FOLLOWED BY THE REINFORCEMENT MATERIAL, AND THEN THE VOLUME PERCENT AND SHAPE OF REINFORCEMENT, EACH SEPARATED BY A SLASH. THE HEAT TREATMENT DESIGNATION, IF USED, IS PLACED AFTER THE ENTRY FOR PERCENT AND SHAPE OF REINFORCEMENT. FOR EXAMPLE, $6061/AL_2O_3/15_P$-T6 REFERS TO AN AL-MMC WITH A 6061 AL ALLOY MATRIX THAT HAS BEEN HEAT-TREATED TO THE T6 CONDITION AND IS REINFORCED WITH 15 VOL% AL_2O_3 PARTICLES. A 15_W DESIGNATION WOULD INDICATE REINFORCEMENT IN WHISKER (W) FORM. THIS NOMENCLATURE WILL BE USED THROUGHOUT THE ARTICLE.

Selection and Weldability of Aluminum Metal-Matrix Composites

Thomas J. Lienert, The Ohio State University; Charles T. Lane, Duralcan USA; Jerry E. Gould, Edison Welding Institute

General Joining Considerations

The effective integration of Al-MMCs into useful structures and devices often requires that they be joined to similar and dissimilar materials. As a result, the ultimate utility of Al-MMCs hinges on the ability to effectively join them within reasonable economic constraints. As with all aluminum products, appropriate cleaning before joining and proper shielding during joining are necessary to avoid hydrogen contamination and the subsequent formation of porosity. However, the weldability of Al-MMCs differs markedly from that of monolithic aluminum alloys, because of the presence of the reinforcement. The addition of reinforcement to the aluminum matrix acts to modify several physical and chemical properties, which in turn greatly affect the weldability of Al-MMCs. Thus, an understanding of the weldability of Al-MMCs must include a thorough knowledge of the effects of the various interactions between matrix and reinforcement.

Viscosity Effects. The welding of Al-MMCs differs from the welding of monolithic aluminum alloys in that the weld pool of the composite contains a solid phase, because the reinforcement does not necessarily melt at welding temperatures. Thus, in the strictest sense, the composite weld pool is actually a partially melted zone. The most obvious effect of the presence of the reinforcement is the increased viscosity of the Al-MMC weld pool (Ref 2). This weld pool does not flow and wet as readily as that of monolithic aluminum alloys. Mass transfer is limited, because of the viscous nature of the MMC weld pool. Therefore, heat flow by convection in the weld pool is believed to be less effective than it is in aluminum alloys. Consequently, conductive heat flow through the aluminum alloy matrix is thought to play a larger than normal role in determining the temperature distributions and cooling rates in and around the weld pool of Al-MMCs.

The differences in heat flow in the MMC, relative to monolithic aluminum alloys, can affect the resulting microstructures and the stress distributions in the MMC weld.

Chemical Reactions. From a thermodynamic standpoint, Al-MMCs are nonequilibrium materials. Because a chemical potential difference exists between the matrix and the reinforcement, chemical reactions may occur. The contact between liquid aluminum and the reinforcement during joining can lead to the formation of unplanned or undesirable phases. The new phases may form at the matrix-reinforcement interface or they may entirely consume the original reinforcement.

For example, Al_4C_3 will spontaneously form when joining silicon carbide- or graphite-reinforced Al-MMCs by one of the following reactions:

$$3\mathrm{SIC}(S) + 4\mathrm{AL}(L) \;\; \text{€} \;\; \mathrm{AL_4C_3}(S) + 3\mathrm{SI_{AL}} \qquad \textbf{(EQ 1)}$$

$$3\mathrm{C}(S) + 4\mathrm{AL}(L) \;\; \text{€} \;\; \mathrm{AL_4C_3}(S) \qquad \textbf{(EQ 2)}$$

where (*s*) and (*l*) represent solid and liquid, respectively.

Studies of the thermodynamics of Eq 1 and 2 show that the standard free energy, $\Delta G°$, of the reaction becomes increasingly negative (that is, the reaction proceeds further to the right) with increasing temperature and that the addition of silicon to the aluminum alloy matrix, which increases the activity of the silicon, aids in limiting the formation of Al_4C_3 during joining (Ref 3). The formation of Al_4C_3 creates concern for two reasons. First, its presence (in a very coarse or clustered form) is associated with a decrease in the mechanical properties of Al-MMCs. Second, it will readily dissolve in aqueous environments, resulting in formation of porosity and a loss of integrity to the weld.

Some control of the Al_4C_3 microstructure and, thus, the joint properties of silicon carbide- or graphite-reinforced Al MMCs can be achieved by tailoring the joining process. For these two types of composites, the integrated time at temperature above the matrix liquidus experienced by the composite during joining should be minimized, in order to limit the formation of Al_4C_3. Thus, joining should be performed using either a solid-state joining process or a fusion-welding process that can provide rapid thermal cycles with low heat inputs. Alternatively, a matrix alloy (or filler metal) with a high silicon content should be chosen, if possible, for the MMC.

A second type of chemical reaction occurs in aluminum-magnesium composites reinforced with Al_2O_3, where a magnesium spinel ($MgAl_2O_4$) is formed by the following reaction:

$$3\mathrm{MG} + 4\mathrm{AL_2O_3} \;\; \text{€} \;\; 3\mathrm{MGAL_2O_4} + 2\mathrm{AL} \qquad \textbf{(EQ 3)}$$

The formation of copper spinel ($CuAl_2O_4$) by similar reactions has also been reported in aluminum-copper alloy composites reinforced with Al_2O_3. The formation of these spinels is known to enhance wetting of the reinforcement by the aluminum, which limits or mitigates the agglomeration of the aluminum oxide into large clusters. The presence of the spinels at the interface does not appear to markedly change the mechanical properties of the weld.

Effects on Solidification. The third type of matrix-reinforcement interaction occurs during the solidification of discontinuously reinforced aluminum composites (Ref 4). Below a certain critical solidification rate, the reinforcement is pushed ahead of the solidification front. Because the reinforcement is pushed to the last areas to solidify, the uniformity of its distribution is governed by solidification conditions. The collection of reinforcement in front of the interface locally hinders heat flow and mass transport and causes a decrease in the extent of microsegregation, which diminishes the amount of terminal solidification phase(s) that forms. Furthermore, the modification of the amount and distribution of the terminal solidification phase(s) appears to improve the resistance of at least some Al-MMCs to solidification cracking. Modifications to microsegregation can also occur in continuous fiber reinforced Al-MMCs when solidification occurs along the small capillaries formed between the fibers. Both the pushing of reinforcement and the decrease in microsegregation can have an effect on the final microstructure and the properties of Al-MMC welds.

Weld Preparation. The standard methods of weld preparation are generally suitable for Al-MMCs, although a few modifications should be adopted. Because of their ceramic reinforcement, most MMCs are very wear resistant. Thus, standard steel-cutting tools and saw blades are ineffective for preparing the joint. Solid carbide (speed <100 m/min, or 330 ft/min) or diamond-tipped and diamond-plated tools (speed >400 m/min, or 1310 ft/min) are recommended for all beveling and back gouging. Feeds that are greater than 0.3 mm/tooth (0.012 in./tooth) should be used for cutting tools, whereas constant, medium-heavy pressures should be used for saws.

Although the lower coefficient of thermal expansion (CTE) of the MMC will reduce the tendency of the joint to close ahead of the arc, the root gap and included angle of the joint should still be increased slightly to allow for the reduced fluidity of the composite weld pool.

The welding processes used for Al-MMCs can be characterized by the state of the matrix during welding. Several fusion-welding processes have been used to join these composites, including gas-tungsten arc welding (GTAW), gas-metal arc welding (GWAW), laser-beam welding (LBW), electron-beam welding (EBW), and resistance welding (RW). In addition, friction welding (FRW), transient liquid phase (TLP) bonding, and capacitor discharge welding (CDW) also have been used. Details on each process are provided below.

References cited in this section

2. B. ALTSHULLER, W. CHRISTY, AND B. WISKEL, GMA WELDING OF AL-AL_2O_3 METAL MATRIX COMPOSITE, *WELDABILITY OF MATERIALS, PROC. MATERIALS WELDABILITY SYMP.*, ASM INTERNATIONAL, 1990, P 305-309
3. T. ISEKI, T. KAMEDA, AND T. MARUYAMA, INTERFACIAL REACTIONS BETWEEN SIC AND ALUMINUM DURING JOINING, *J. MATER. SCI.*, VOL 19, 1984, P 1692-1698
4. P.K. ROHATGI, R. ASTHANA, AND S. DAS, SOLIDIFICATION, STRUCTURES, AND PROPERTIES OF CAST METAL-CERAMIC PARTICLE COMPOSITES, *INT. MET. REV.*, VOL 31 (NO. 3), 1986, P 115-139

Selection and Weldability of Aluminum Metal-Matrix Composites

Thomas J. Lienert, The Ohio State University; Charles T. Lane, Duralcan USA; Jerry E. Gould, Edison Welding Institute

---

Fusion-Welding Processes

Arc Welding. The metallurgy of arc welding, more so than other joining processes, is akin to a casting process. Aluminum is melted, undergoes mixing, and then resolidifies. The microstructure of the weld, including its defects, is not dissimilar to that of an aluminum casting. Unfortunately, this "recasting" process can significantly degrade the microstructure and properties of most MMCs.

In terms of equipment and consumables, the arc-welding practice for Al-MMCs is substantially similar to that for aluminum alloys. When one considers that from 75 to 90 vol% of the MMC is aluminum, it becomes apparent that the basic rules for equipment and filler-metal selection still apply. The constant-current power supply is still recommended for most welding, although constant voltage is preferred for some automated welding. Argon is generally the preferred shielding gas, except in the usual cases where it is either mixed with or replaced by helium. Filler metals are usually matched to the matrix alloy of the composite and the service requirements of the weld.

For example, the alloy 6061 is typically welded with ER4043 or ER5356. These filler metals would be appropriate for a composite that has 6061 as its matrix. In the case of $6061/Al_2O_3/10_p$, the ER5356 is used for its ability to keep the reinforcement wetted. In the case of $6061/SiC/15_w$, the ER4043 would be preferred for its assistance in suppressing the Al_4C_3 reaction. The extra silicon in ER4045 or ER4047 may be even more effective in this regard. However, the near-eutectic composition of the resulting weld metal is unsuitable for postweld solution heat treatments because of the coarsening of the silicon phase (Ref 5).

Primary silicon will precipitate in significant amounts only if the silicon content of the melt is greater than 12.6%, or the cooling rate is sufficiently high at a lower silicon level to prevent eutectic formation. Although 4047 contains 11 to 13% Si, the weld pool contains less, owing to dilution by the base material being welded (6061 has only 4% Si). Arc welds do not cool rapidly enough to precipitate primary silicon from a hypoeutectic aluminum-silicon solution. Finally, primary silicon *enhances* properties if it is small and well-refined versus large and coarse.

The GTAW process is the original inert-gas arc-welding process developed for aluminum. Its low, controlled heat input is ideal for thin sections and complex shapes. The control of heat input is especially important for welding silicon carbide-reinforced wrought alloys. For most applications, the GTAW process is used with alternating current (ac), and the arc balance is adjusted to favor either cleaning or penetration. A cleaning arc can help minimize reactions between the reinforcement and matrix by reducing heat input. However, dilution of the composite into the weld pool is also decreased. Thus, some welds do not have homogeneous microstructures and do not reflect the properties of the parent material.

An arc set for cleaning is capable of breaking up the thin surface oxides that exist on the surface of a typical aluminum weld pool. However, in the case of a metal-matrix composite, the skin is stabilized by the ceramic particles (diluted from the base material), which are held at the pool surface by buoyancy forces and surface tension. A cleaning arc does not register properly on this "shielded" surface and tends to "skitter" along the joint. Formation and advancement of the weld pool is difficult, and some surface melting adjacent to the weld bead may occur.

For most Al-MMCs, a balanced, or penetrating, arc is preferred, because it is more focused and forces heat into the weld pool and joint, rather than melting the adjacent base material. The increased specific heat makes feeding the filler wire significantly easier and results in an improved bead profile. The degree of penetrating balance will depend on the geometry of the joint and the preferred technique of the welder.

Another effect of the particle-stabilized oxide skin is the dull appearance of the weld-pool surface. In fact, the initial formation of the weld pool is difficult to discern, and welders may overmelt the area while waiting for a bright pool to form. This is further complicated by the fact that the work area is not as well illuminated, because of the reduction in reflected light from the arc. Welders should watch the area under the arc for a "slumping" of the surface or another indication that subsurface melting is present. Feeding the filler rod into the weld pool early will accelerate pool formation, increase its fluidity, and partially dissipate the oxide skin. However, the pool will not become bright, like an unreinforced aluminum weld pool does, and the welder should begin advancing the weld as soon as he can feed in the filler.

An example of this procedure is a single-pass butt weld on 6.35 mm (0.25 in.) thick plate of Al-MMC. A series of different Al-MMC extrusions was prepared with a 30° edge bevel (60° included angle) and a 1.6 mm (0.06 in.) root land. After degreasing and scrubbing with a clean stainless steel brush, the plates were fixtured to a grooved stainless steel backing bar using a 1.6 mm (0.06 in.) root gap. The power supply was set for 22 V and 325 A, maximum, with the ac balance towards penetration (a setting of 5, where 3 equals balance for a range of 0 to 10). A 2.4 mm (0.094 in.) filler rod of ER5356 was fed into the weld at 23.3 mm/s (0.917 in./s) as the automatic cold-wire feed traveled at 25.8 mm/s (1.02 in./s). The shielding gas of 100% argon was metered at 16.5 L/min (35 ft^3/h).

The resulting welds passed the bead appearance, radiography, and tensile requirements for 6061, as described in AWS D1.2-90. Typical as-welded properties are listed in Table 1, and a cross section of a representative weld is shown in Fig. 1. Note the smooth bead contour, minimal porosity, and even distribution of particles in the weld.

TABLE 1 PROPERTIES OF SELECTED ALUMINUM ALLOYS AND MMCS JOINED USING THE GAS-TUNGSTEN ARC WELDING PROCESS

BASE MATERIAL (T6 TEMPER)	0.2% YIELD STRENGTH		ULTIMATE TENSILE STRENGTH		ELONGATION, %	FAILURE LOCATION
	mpa	**ksi**	**mpa**	**ksi**		
6061$^{(A)}$	131	19	207	30	11	. . .
6061/AL_2O_3/10_P	145	21	218	32	9.8	**HEAT-AFFECTED ZONE**
6061/AL_2O_3/20_P	160	23	221	32	5.2	**HEAT-AFFECTED ZONE**
7005/AL_2O_3/10_P	190	28	283	41	7.3	**WELD**
7005$^{(B)}$	**207**	**30**	**317**	**46**	**10**	**. . .**

(A) TYPICAL HANDBOOK VALUES FOR ER5356 FILLER (REF 6).

(B) TYPICAL HANDBOOK VALUES FOR ER5556 FILLER (REF 6)

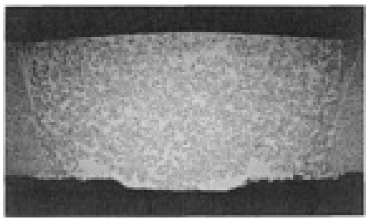

FIG. 1 MICROSTRUCTURE OF GAS-TUNGSTEN ARC WELDED 6061/$AL_2O_3/10_P$ (6.35 MM, OR 0.25 IN., PLATE)

The GMAW process is preferred for most production welding, because of its high deposition rate and ease of automation. A root gap of 1 to 2 mm (0.04 to 0.08 in.), in conjunction with a temporary grooved backing bar, is recommended for single-sided welds, because of the lower fluidity and reduced penetration of the composite weld pool. A root gap is also preferred for double-sided welds; the root should be thoroughly back gouged before welding the second side. Joint openings range from 60° for single-V to 90° for double-V butt welds with root lands of 10 to 20% of the plate thickness. For multipass welds, it is recommended that the weld be vigorously scrubbed with a clean stainless steel wire brush between passes to remove any dewetted reinforcing phase, condensed magnesium oxide (a black, powdery residue), or other contaminants.

This procedure is illustrated by a multipass butt weld on a 19 mm (0.75 in.) plate of 6061/$Al_2O_3/20_p$-T6 (Ref 7). The extrusion was prepared with a 30° edge bevel (60° included angle) and a 3.2 mm (0.16 in.) root land. After degreasing and scrubbing with a clean stainless steel brush, the plates were fixtured to a grooved, stainless steel backing bar using a 1.6 mm (0.06 in.) root gap. A constant-current, dc, inverter-type power supply was set for 305 A and 26 V_{dc}, electrode positive (DCEP). The four passes were made at different travel speeds. The root pass was made at 384 mm/min (15 in./min) to keep the arc at the leading edge of the weld pool and ensure good root penetration. The speed was slowed to 252 mm/min (10 in./min) during the second pass to ensure good side-wall fusion. Capping passes (third and fourth) were made at 354 mm/min (14 in./min). A 1.6 mm (0.06 in.) filler wire of ER5356 was used with 100% argon shielding metered at 23.6 L/min (50 ft^3/h).

Typical as-welded properties are listed in Table 2, and a representative microstructure from the fusion zone is shown in Fig. 2. In this figure, the base material is to the right and the weld is on the left. There are no visible reaction products at the particle interface, and the dilution of the particles from the base material is relatively even, with minimal agglomeration.

TABLE 2 PROPERTIES OF ALUMINUM ALLOY 6061 AND ALUMINUM MMC 6061/$AL_2O_3/20_P$ JOINED USING THE GAS-METAL ARC WELDING PROCESS

BASE MATERIAL	**0.2% YIELD STRENGTH**		**ULTIMATE TENSILE STRENGTH**		**ELONGATION, %**	FAILURE LOCATION
	Mpa	**Ksi**	**Mpa**	**Ksi**		
6061, AS-WELDED(A)	131	19	207	30	11	**. . .**
6061-T6(B)	276	40	304	44	5	**. . .**
6061/$AL_2O_3/20_P$, AS-WELDED	132	19	228	33	6.6	**HEAT-AFFECTED ZONE**
6061/$AL_2O_3/20_P$-T6	**189**	**27**	**283**	**41**	**3.9**	**WELD**

(A) TYPICAL HANDBOOK VALUES FOR ER5356 FILLER (REF 6).

(B) TYPICAL HANDBOOK VALUES FOR ER4043 FILLER; NO DATA AVAILABLE FOR ER5356 (REF 6)

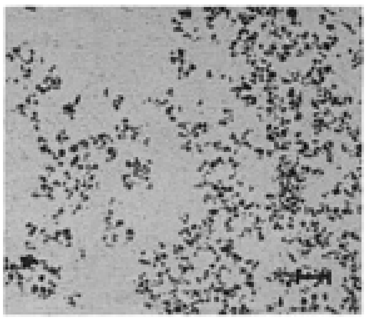

FIG. 2 MICROSTRUCTURE OF GAS-METAL ARC WELDED 6061/AL_2O_3/20_P

Laser-Beam Welding of Silicon Carbide-Reinforced Al-MMCs. Several researchers have attempted to limit the formation of aluminum carbide during fusion welding by employing welding processes, such as LBW, that can produce very rapid thermal cycles. Fully dense, crack-free welds have been produced with both continuous wave (CW) and pulsed carbon dioxide LBW processes, as well as the pulsed Nd:YAG process. The rapid thermal cycles made possible by these processes result in uniform distribution of reinforcement and a very fine grain size.

In one study, CW carbon dioxide laser-beam welds were produced on cast A356 (Al-7Si-0.3Mg)/SiC/15_p discontinuously reinforced aluminum at 1.0 and 1.5 kW and travel speeds ranging from 2.3 to 5.8 m/min (90 to 230 in./min) (Ref 8). These conditions resulted in heat inputs of 15.7 to 39.4 kJ/m (400 to 1000 J/in.). Pulsed Nd:YAG laser-beam welds were made on the same composite at an average power of 200 W with pulse energies that varied from 5 to 20 J. All welds were autogenous bead-on-plate welds.

The fusion zone of the CO_2 laser-beam welds made in this study exhibited three distinct regions (Fig. 3). The upper central region (region 1) contained large amounts of Al_4C_3 and silicon in the aluminum alloy matrix, with little remaining SiC. This region was spike-shape and encompassed less than 25% of the weld cross section. The Al_4C_3 in this region was not continuous with any silicon carbide, and it was relatively long (up to 45 μm, or 1.8 mil). The width of region 1 did not change with variations in heat input, and it was shown to be approximately the same size as the focused beam diameter. This suggests that region I is associated with the high temperatures of the weld keyhole, where the beam and material interact.

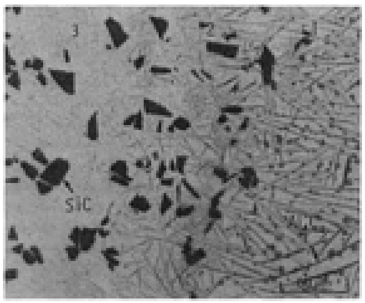

FIG. 3 MICROSTRUCTURE OF CO_2 LASER-BEAM WELD ON A356/SIC/15_P (UNETCHED)

The central region of the CO_2 laser-beam welds was surrounded by a very narrow region (region 2) in which the SiC had partially reacted to form Al_4C_3 and silicon, in accordance with Eq 1. The nucleation and growth of tiny needles of Al_4C_3 from the SiC in the partially reacted zone are evident. These needles, which are straight and relatively short, show an orientation relation with the silicon carbide.

Surrounding the partially reacted zone was a large region (region 3) in which the SiC appeared to be unreacted. This unreacted zone constituted approximately 70% of the total weld volume. Similar phenomena were also observed in the pulsed Nd:YAG laser-beam welds.

Other experiments showed that the extent of aluminum carbide formation in pulsed carbon dioxide laser-beam welds on a cast A356/SiC/15_p composite could be modified by changes in welding parameters. In one study, the size and quantity of Al_4C_3 was observed to increase proportionally with laser specific energy (Ref 9). In another experiment, measurements were made of the strength and ductility of pulsed carbon dioxide laser-beam welds made at a peak power of 3.2 kW and at duty cycles ranging from 50 to 91% (Ref 10). The laser processing resulted in increased strength in most of the welds, relative to the coarse-grained base material, despite the formation of a large amount of Al_4C_3. The presence of other phases, such as $Al_4Si_2C_5$ and MC-type carbides, in the pulsed CO_2 laser-beam welds was also reported.

Whereas LBW experiments have shown that the Al_4C_3 microstructure can be modified by varying process parameters, it appears that LBW is incapable of producing welds that are free of Al_4C_3. Experimental observations and physical data suggest that the laser energy is preferentially absorbed by the SiC, because of its greater absorptivity, relative to aluminum, at the laser wavelengths. The increased energy absorption results in local superheating of the SiC and surrounding areas and promotes Al_4C_3 formation. In summation, it appears that the LBW processes may have limited potential in joining silicon carbide-reinforced Al-MMCs.

Laser-Beam Welding of Aluminum Oxide-Reinforced Al-MMCs. Few LBW studies on aluminum oxide-reinforced Al-MMCs have been reported. It appears that the Al_2O_3 reinforcement also preferentially absorbs the impinging laser energy. The carbon dioxide LBW process was used on a 6061/Al_2O_3/20_p composite produced by powder metallurgy (Ref 11). Initial attempts at welding resulted in cutting of the samples. However, acceptable welds were ultimately produced when the shims of 5083 and 4047 filler wire were placed in the joints.

Similar problems were reported by other researchers on a 6061/aluminum oxide/20_p composite produced via ingot metallurgy (Ref 12). They attributed the difficulty to an unstable plasma above the weld pool and attempted to improve weldability with plasma control. Sound welds with a uniform distribution of reinforcement were produced by placing a small gas nozzle close to the weld area to blow away the plasma material. The quality of the welds was found to strikingly depend on the direction of the suppression gas flow, relative to the welding direction, and on nozzle positioning. Optimum welding conditions were realized when the flow was opposite to the welding direction and the end of the nozzle was inclined upward by 5° to the horizontal. The gas pressures at the nozzle ranged from 70 to 275 kPa (10 to 40 psi). Based on this study, LBW offers some potential for joining aluminum oxide-reinforced Al-MMCs.

Electron-Beam Welding of Silicon Carbide-Reinforced Al-MMCs. The EBW process appears to have potential for joining at least some silicon carbide-reinforced Al-MMCs. A follow-up study to the LBW work described in Ref 6 determined that EBW can be used successfully to join the same cast discontinuously reinforced aluminum (A356/SiC/15_p) without the formation of large amounts of Al_4C_3 (Ref 13). Like LBW, the EBW process is capable of producing rapid thermal cycles. However, the physics of beam-material interaction differs for the two processes. Heating during LBW results from the absorption of photons by the substrate, whereas heating during EBW occurs by the transfer of kinetic energy to the atoms of the substrate via collisions with the high-energy electrons of the beam. In this study, CO_2 laser-beam and electron-beam welds were produced at identical energy inputs, travel speeds, and focused beam diameters. The use of EBW resulted in sound welds with a uniform distribution of reinforcement and a fine grain size. The electron-beam welds exhibited orders of magnitude less Al_4C_3 than the laser-beam welds made using the same parameters. The electron-beam welds made at 85 mm/s (200 in./min) with sharp beam focus contained almost no Al_4C_3. Electron-beam welds made at slower speeds and/or with defocused beams contained somewhat more Al_4C_3.

The different mechanisms of energy transfer in EBW and LBW appear to affect the final weld microstructures of silicon carbide-reinforced Al-MMCs and, thus, the weld properties, despite the fact that both welding processes allow for rapid thermal cycles with low overall heat input.

Unfortunately, EBW cannot be successfully applied to all silicon carbide-reinforced Al-MMCs. For example, attempts at EBW on a wrought 2014/SiC/15_p MMC resulted in cutting of the sample. It is not known whether the problem resulted from the manufacturing process (powder metallurgy) or was inherent to the material.

Transient liquid-phase (TLP) bonding has proven fairly successful in joining Al-MMCs. The process employs a filler material or interlayer to produce a transient liquid layer at the interface to be joined. Solidification of the joint occurs isothermally, by diffusion. The advantage of TLP bonding of Al-MMCs involves the lower peak-processing temperatures, compared with the temperatures used in other fusion-welding processes. The lower processing temperatures result in less damage to the engineering properties of the composite in and around the joint. However, there is a concern for damage to the properties of the composite base material during processing, because the thermal cycles for TLP bonding hold the entire assembly at temperatures near the solidus for a longer time, relative to the other fusion-welding processes.

Transient liquid-phase bonding processes for Al-MMCs are controlled by several parameters, including surface finish, type and thickness of interlayer or filler metal, processing time and temperature, and clamping pressure on the joint. The surface aluminum oxides must be removed from the samples to facilitate wetting of the substrate by the interlayer. The choice of the correct combination of processing parameters is critical in avoiding such problems as reinforcement enriched or depleted zones and Kirkendall porosity. In general, the thinnest interlayer possible should be used to allow for the shortest processing times and a narrow liquid zone. Additionally, the lowest possible temperatures should be used to limit damage to the material properties of the composite. Although the application of pressure can aid in improving interlayer-substrate contact, excess pressure can lead to expulsion of the liquid zone, which results in reinforcement-rich areas at the joint line.

Silicon Carbide-Reinforced Al-MMCs. Transient liquid-phase bonding with gold and aluminum-silicon-magnesium interlayers has been used to join two 6061/SiC/25_p composite sheets (Ref 14). Based on thermomechanical simulation studies, a process window of time-temperature combinations that would not produce damage to the properties of the base material was developed. The window ranged from 30 min at 565 °C (1050 °F) to 10 min at 580 °C (1075 °F). A 0.025 mm (0.001 in.) gold interlayer was shown to produce better results than the aluminum-silicon-magnesium material. Optimum conditions of 30 min at 567 to 580 °C (1053 to 1075 °F) were reported for the gold interlayer. Three distinct zones were observed in the vicinity of the bond interface: a particle-enriched zone, a diffuse flow zone that showed a pattern of material flow, and the undisturbed base material. No mention was made of the formation of Al_4C_3 for these conditions. Low tensile joint efficiencies (<30%) were attributed to voids at the joint interface.

Aluminum Oxide-Reinforced Al-MMCs. The joining of a 6061/Al_2O_3/15_p Al-MMC using TLP bonding also has been reported (Ref 15). Three types of interlayers demonstrated an ability to join the material: gold, copper, and Al-12Si. The bonding experiments were performed in vacuum (1.3×10^{-8} Pa, or 10^{-10} torr) on mechanically cleaned samples using a clamping pressure of less than 70 kPa (10 psi). Joining with a 25 μm (1 mil) thick interlayer of gold at 580 °C (1075 °F) for 130 min produced the strongest bonds. The joints had a 323 MPa (47 ksi) yield strength and a 341 MPa (49 ksi) ultimate tensile strength with a joint efficiency of 95%. Joining with a 125 μm (0.005 in.) thick layer of Al-12Si at 585 °C for 20 min yielded bonds with nearly identical properties; however, the bond line contained residual filler material, as well as voids. A 25 μm (0.001 in.) thick layer of copper produced bonds with much lower strengths. An increase in volume fraction of the reinforcement was observed near the bond lines of joints made with gold and copper. Increased loading of reinforcement was not observed for bonds made with the Al-12Si.

The brazing of a 6061 MMC reinforced with 5, 10, and 15 vol% short Al_2O_3 fibers using aluminum-silicon and aluminum-manganese interlayers has been studied (Ref 16). However, the reported bonding temperature range of 580 to 610 °C (1075 to 1130 °F) suggests that TLP bonding may have actually occurred. A 150 μm (0.006 in.) thick sheet of Al-10Si was used for joining, as well as a 140 μm (0.0056 in.) thick sheet of Al-1.5Mn clad with 20 μm (800 μin.) of Al-10Si on each side. The bonding time for each test was 10 min at a vacuum of 6.7 mPa (5×10^{-5} torr). Bonding the composite to itself proved easier than joining the composite to a monolithic 6061 alloy. The tensile strengths for the similar material bonds of 5% aluminum oxide fiber material exceeded 230 MPa (33 ksi). The joint efficiencies of these bonds were not reported.

References cited in this section

5. S.H. LO, S. DIONNE, M. POPESCU, S. GEDEON, AND C.T. LANE, EFFECTS OF PRIOR- AND POST-WELD HEAT TREATMENTS ON THE PROPERTIES OF SIC/AL-SI COMPOSITE GAS METAL ARC WELDED JOINTS, *PROC. ADVANCES IN PRODUCTION AND FABRICATION OF LIGHT METALS AND MMCS,* M. AVEDESIAN, L. LAROUCHE, AND J. MASOUNAVE, ED., CANADIAN INSTITUTE OF MINING, METALLURGY AND PETROLEUM, 1992, P 575-588
6. J.E. HATCH, *ALUMINUM--PROPERTIES AND PHYSICAL METALLURGY,* AMERICAN SOCIETY FOR METALS, 1984, P 359
7. KAISER ALUMINUM, *WELDING,* 1978, P 3-24 TO 3-31
8. M.J. COLA, T.J. LIENERT, J.E. GOULD, AND J.P. HURLEY, LASER WELDING OF A SIC PARTICULATE REINFORCED ALUMINUM METAL MATRIX COMPOSITE, *WELDABILITY OF MATERIALS, PROC. MATERIALS WELDABILITY SYMP.,* ASM INTERNATIONAL, 1990, P 297-303
9. N.B. DAHOTRE, T.D. MCCAY, AND M.H. MCCAY, LASER PROCESSING OF A SIC/AL-ALLOY METAL MATRIX COMPOSITE, *J. APPL. PHYS.,* VOL 65 (NO. 12), 1989, P 5072-5077

10. N.B. DAHOTRE, M.H. MCCAY, T.D. MCCAY, AND S. GOPINATHAN, PULSE LASER PROCESSING OF A SIC/AL-ALLOY METAL MATRIX COMPOSITE, *J. MATER. RES.*, VOL 6 (NO. 3), 1991, P 514-529
11. R.P. MARTUKANITZ AND R.B. BHAGAT, LASER PROCESSING OF DISCONTINUOUSLY REINFORCED, ALUMINUM MATRIX COMPOSITES, *THE METAL SCIENCE OF JOINING,* ASM INTERNATIONAL, 1992, P 241-247
12. S.M. KAWALI, G.L. VIEGELAHN, AND R. SCHEUERMAN, LASER WELDING OF ALUMINA REINFORCED 6061 ALUMINUM ALLOY COMPOSITE, *PROC. LASER MATERIALS PROCESSING SYMPOSIUM,* ICALEO '91, LASER INSTITUTE OF AMERICA, 1991, P 156-167
13. T.J. LIENERT, E.D. BRANDON, AND J.C. LIPPOLD, LASER AND ELECTRON BEAM WELDING OF SIC$_P$ REINFORCED ALUMINUM A356 METAL MATRIX COMPOSITE, *SCR. METALL. MATER.*, VOL 28, JUNE 1993, P 1341-1346
14. K. SUDHAKAR, "JOINING OF ALUMINUM BASED PARTICULATE-REINFORCED METAL-MATRIX COMPOSITES," DISSERTATION, THE OHIO STATE UNIVERSITY, 1990
15. R. KLEHN, "JOINING OF 6061 ALUMINUM MATRIX CERAMIC PARTICLE REINFORCED COMPOSITES," M.S. THESIS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, SEPT 1991
16. K. SUGANUMA, T. OKAMOTO, AND N. SUZUKI, JOINING OF ALUMINA SHORT-FIBRE REINFORCED AA6061 ALLOY TO AA6061 ALLOY AND TO ITSELF, *J. MATER. SCI.,* 1987, P 1580-1584

Selection and Weldability of Aluminum Metal-Matrix Composites

Thomas J. Lienert, The Ohio State University; Charles T. Lane, Duralcan USA; Jerry E. Gould, Edison Welding Institute

Resistance Welding and Forge Welding

Both resistance welding and forge welding are quite adaptable to a range of discontinuously reinforced aluminum composites. As described previously, problems associated with welding these materials typically have two causes. For the silicon carbide-reinforced Al-MMCs, the high temperatures and relatively long times at temperatures permit dissolution of the carbides and subsequent precipitation of detrimental phases (Al_4C_3). For the aluminum oxide-reinforced Al-MMCs, similar conditions allow the individual reinforcing particles to agglomerate, which reduces material performance. Both resistance and forge welding processes have rapid thermal cycles that may limit these problems. Particular aspects of these processes are described separately below.

Resistance Welding. Virtually all available work concerning the resistance welding of MMCs has been done using the resistance spot welding technique. An example of a cross section of a spot weld made on an A359 (Al-9Si-0.3Mg)/SiC/10_p is shown in Fig. 4. Aluminum MMCs are generally spot welded using conditions similar to those for conventional aluminum alloys. However, these materials typically require substantially less current (about half) for welding. This is likely due to an increase in bulk resistivity associated with the addition of the particles. The spot weld in Fig. 4, which shows a full-size nugget, does not show any of the Al_4C_3 plates seen in the fusion zones of the laser welds. Rather, a uniform distribution of the SiC particles is noted.

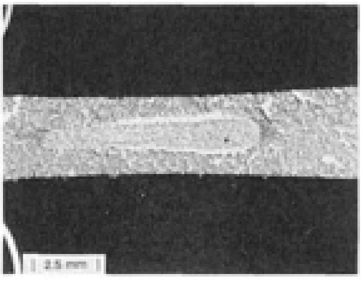

FIG. 4 RESISTANCE SPOT WELD ON A359/SIC/10$_{\text{P}}$

It is believed that the minimal reaction of the SiC to form Al_4C_3 is due to two factors. First, in resistance welds, the temperature of the liquid nugget only barely exceeds the melting point of the material (Ref 17), providing little, if any, driving force for the reaction. Second, temperature excursions above the liquidus are relatively short, typically less than 0.1 s, which minimizes any potential reaction.

Spot welds on discontinuously reinforced aluminum composites do show some redistribution of the reinforcement in the fusion zone (Fig. 4). Typically, the fusion zone is relatively free of the reinforcement near the fusion line, with an increased density of the reinforcement near the center of the weld. It is possible that this redistribution of the composite is the result of macroscopic liquid flow during solidification, but the effect is still under study. The effect of this reinforcement-free zone on the mechanical performance of the weld is not known.

Friction Welding. Both silicon carbide- and aluminum oxide-reinforced Al-MMCs have been friction welded using continuous-drive and inertia machines (Ref 18, 19). Friction welding is a solid-state process and does not require melting. Thus, it is not hampered by the melting and solidification concerns described previously. An examination of friction welds on these materials typically shows a distribution of reinforcement particles similar to that in the base material. There is, however, concern with the development of high in-process torques associated with these processes. High torques, particularly during stopping, can cause debonding of the matrix and the reinforcement, resulting in cracking near the bond line. In addition, stringers, if present in the material, can "open up" under the influence of these resultant torques.

Compared to conventional aluminum alloys, discontinuously reinforced aluminum composites require about double the applied force. This is because the reinforcement particles substantially increase the flow stress of the composite, even at high temperatures, preventing metal flow.

Inertia friction welding studies on an A356/SiC/20$_p$ and an A359/SiC/20$_p$ have been performed (Ref 18). Both similar and dissimilar material welds to a monolithic 6061-T651 aluminum alloy were produced over a range of flywheel speeds and axial pressures. High joint efficiencies (greater than 70%) were realized when high flywheel speeds and high axial pressures, relative to those used for monolithic aluminum alloys, were employed for the similar material welds. The joint efficiencies of the dissimilar welds were less than 70%, relative to the composite base material.

Flash-butt welding, another nominally solid-state process, has also been found to be effective for joining these materials (Ref 20). The flash welding of conventional aluminum alloys is typically characterized by very fast flashing rates and upset velocities. Similar approaches are recommended for the flash welding of Al-MMCs. Other process features that contribute to weld quality are the use of shielding gases during flashing (minimizing bond-line oxidation) and dc power for flashing. The dc power appears to provide a finer flash, resulting in higher thermal gradients in the workpiece.

Capacitor discharge welding is a generic term relating to all resistance-welding processes that supply energy to the workpiece by the discharge of charged capacitors. Variations on this process range from spot welding to small-scale butt welding. Generally, capacitive discharge welding processes are characterized by extremely rapid discharge rates (on the order of 5 to 25 ms) and subsequent rapid thermal cycles. In terms of suppressing chemical reactions in silicon carbide-reinforced Al-MMCs, the capacitive discharge processes offer considerable potential.

Work on the capacitive discharge welding of Al-MMCs has been reported (Ref 21). The capacitive system used in this work was basically a forge-butt welding process. Specifically, the system used oriented parts in a butt configuration, with the ends separated by a projection on one of the surfaces. A force was then applied to the workpieces and the energy from the discharging capacitors introduced. The result was surface melting, followed by impact and forging of the components. This work has shown that Al_4C_3 formation can be precluded on several types of silicon carbide-reinforced Al-MMCs. Consistent with the other resistance-welding processes, the short exposure times and rapid cooling rates (approximately 10^6 °C/s, or 2×10^6 °F/s) were responsible for the retention of the SiC.

References cited in this section

17. J.E. GOULD, AN EXAMINATION OF NUGGET DEVELOPMENTS DURING RESISTANCE SPOT

WELDING USING EXPERIMENTAL AND ANALYTICAL TECHNIQUES, *WELD. J. RES. SUPP.*, VOL 68 (NO. 1), 1987, P 1S-10S

18. M.J. COLA, "INERTIA-FRICTION WELDING OF PARTICULATE-REINFORCED ALUMINUM MATRIX COMPOSITES," THESIS, THE OHIO STATE UNIVERSITY, 1992
19. M.J. COLA, G.S. MARTIN, AND C.E. ALBRIGHT, "INERTIA-FRICTION WELDING OF A 6061-T6/$AL_2O_3/15_P$ METAL-MATRIX COMPOSITE," RESEARCH REPORT MR9108, EDISON WELDING INSTITUTE, JUNE 1991
20. D.S. SCHWARTZ, J.H. DEVLETIAN, S.J. CHEN, AND J.E. GOULD, CAPACITOR DISCHARGE AND FLASH WELDING OF SIC-REINFORCED HIGH TEMPERATURE AL ALLOYS, JOINING AND ADHESION OF INORGANIC MATERIALS, PROCEEDINGS OF MATERIALS RESEARCH SOCIETY SPRING 1993 MEETING, VOL 314
21. J.H. DEVLETIAN, SIC/AL METAL MATRIX COMPOSITE WELDING BY A CAPACITOR DISCHARGE PROCESS, *WELD. J.*, 1987, P 33-39

Selection and Weldability of Aluminum Metal-Matrix Composites

Thomas J. Lienert, The Ohio State University; Charles T. Lane, Duralcan USA; Jerry E. Gould, Edison Welding Institute

---

References

1. G.E. METZGER, JOINING OF METAL-MATRIX FIBER-REINFORCED COMPOSITE MATERIALS, *WELD. RES. COUNC. BULL.*, NO. 207, 1975
2. B. ALTSHULLER, W. CHRISTY, AND B. WISKEL, GMA WELDING OF AL-AL_2O_3 METAL MATRIX COMPOSITE, *WELDABILITY OF MATERIALS, PROC. MATERIALS WELDABILITY SYMP.*, ASM INTERNATIONAL, 1990, P 305-309
3. T. ISEKI, T. KAMEDA, AND T. MARUYAMA, INTERFACIAL REACTIONS BETWEEN SIC AND ALUMINUM DURING JOINING, *J. MATER. SCI.*, VOL 19, 1984, P 1692-1698
4. P.K. ROHATGI, R. ASTHANA, AND S. DAS, SOLIDIFICATION, STRUCTURES, AND PROPERTIES OF CAST METAL-CERAMIC PARTICLE COMPOSITES, *INT. MET. REV.*, VOL 31 (NO. 3), 1986, P 115-139
5. S.H. LO, S. DIONNE, M. POPESCU, S. GEDEON, AND C.T. LANE, EFFECTS OF PRIOR- AND POST-WELD HEAT TREATMENTS ON THE PROPERTIES OF SIC/AL-SI COMPOSITE GAS METAL ARC WELDED JOINTS, *PROC. ADVANCES IN PRODUCTION AND FABRICATION OF LIGHT METALS AND MMCS*, M. AVEDESIAN, L. LAROUCHE, AND J. MASOUNAVE, ED., CANADIAN INSTITUTE OF MINING, METALLURGY AND PETROLEUM, 1992, P 575-588
6. J.E. HATCH, *ALUMINUM--PROPERTIES AND PHYSICAL METALLURGY*, AMERICAN SOCIETY FOR METALS, 1984, P 359
7. KAISER ALUMINUM, *WELDING*, 1978, P 3-24 TO 3-31
8. M.J. COLA, T.J. LIENERT, J.E. GOULD, AND J.P. HURLEY, LASER WELDING OF A SIC PARTICULATE REINFORCED ALUMINUM METAL MATRIX COMPOSITE, *WELDABILITY OF MATERIALS, PROC. MATERIALS WELDABILITY SYMP.*, ASM INTERNATIONAL, 1990, P 297-303
9. N.B. DAHOTRE, T.D. MCCAY, AND M.H. MCCAY, LASER PROCESSING OF A SIC/AL-ALLOY METAL MATRIX COMPOSITE, *J. APPL. PHYS.*, VOL 65 (NO. 12), 1989, P 5072-5077
10. N.B. DAHOTRE, M.H. MCCAY, T.D. MCCAY, AND S. GOPINATHAN, PULSE LASER PROCESSING OF A SIC/AL-ALLOY METAL MATRIX COMPOSITE, *J. MATER. RES.*, VOL 6 (NO. 3), 1991, P 514-529
11. R.P. MARTUKANITZ AND R.B. BHAGAT, LASER PROCESSING OF DISCONTINUOUSLY REINFORCED, ALUMINUM MATRIX COMPOSITES, *THE METAL SCIENCE OF JOINING*, ASM INTERNATIONAL, 1992, P 241-247

12. S.M. KAWALI, G.L. VIEGELAHN, AND R. SCHEUERMAN, LASER WELDING OF ALUMINA REINFORCED 6061 ALUMINUM ALLOY COMPOSITE, *PROC. LASER MATERIALS PROCESSING SYMPOSIUM,* ICALEO '91, LASER INSTITUTE OF AMERICA, 1991, P 156-167
13. T.J. LIENERT, E.D. BRANDON, AND J.C. LIPPOLD, LASER AND ELECTRON BEAM WELDING OF SIC_P REINFORCED ALUMINUM A356 METAL MATRIX COMPOSITE, *SCR. METALL. MATER.*, VOL 28, JUNE 1993, P 1341-1346
14. K. SUDHAKAR, "JOINING OF ALUMINUM BASED PARTICULATE-REINFORCED METAL-MATRIX COMPOSITES," DISSERTATION, THE OHIO STATE UNIVERSITY, 1990
15. R. KLEHN, "JOINING OF 6061 ALUMINUM MATRIX CERAMIC PARTICLE REINFORCED COMPOSITES," M.S. THESIS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, SEPT 1991
16. K. SUGANUMA, T. OKAMOTO, AND N. SUZUKI, JOINING OF ALUMINA SHORT-FIBRE REINFORCED AA6061 ALLOY TO AA6061 ALLOY AND TO ITSELF, *J. MATER. SCI.,* 1987, P 1580-1584
17. J.E. GOULD, AN EXAMINATION OF NUGGET DEVELOPMENTS DURING RESISTANCE SPOT WELDING USING EXPERIMENTAL AND ANALYTICAL TECHNIQUES, *WELD. J. RES. SUPP.*, VOL 68 (NO. 1), 1987, P 1S-10S
18. M.J. COLA, "INERTIA-FRICTION WELDING OF PARTICULATE-REINFORCED ALUMINUM MATRIX COMPOSITES," THESIS, THE OHIO STATE UNIVERSITY, 1992
19. M.J. COLA, G.S. MARTIN, AND C.E. ALBRIGHT, "INERTIA-FRICTION WELDING OF A 6061-$T6/AL_2O_3/15_P$ METAL-MATRIX COMPOSITE," RESEARCH REPORT MR9108, EDISON WELDING INSTITUTE, JUNE 1991
20. D.S. SCHWARTZ, J.H. DEVLETIAN, S.J. CHEN, AND J.E. GOULD, CAPACITOR DISCHARGE AND FLASH WELDING OF SIC-REINFORCED HIGH TEMPERATURE AL ALLOYS, JOINING AND ADHESION OF INORGANIC MATERIALS, PROCEEDINGS OF MATERIALS RESEARCH SOCIETY SPRING 1993 MEETING, VOL 314
21. J.H. DEVLETIAN, SIC/AL METAL MATRIX COMPOSITE WELDING BY A CAPACITOR DISCHARGE PROCESS, *WELD. J.,* 1987, P 33-39

General Welding Characteristics of High-Temperature Materials

A.C. Lingenfelter, Lawrence Livermore National Laboratory

Introduction

HIGH-TEMPERATURE ALLOYS are generally recognized as being those alloys that are used at temperatures of 540 °C (1000 °F) and above. These alloys exhibit some combination of high strength at temperature; resistance to environmental attack (including nitridation, carbonization, oxidation, and sulfidation), excellent creep resistance, creep rupture strength, toughness, and metallurgical stability; useful thermal expansion characteristics; and resistance to thermal fatigue and corrosion. This article addresses the general welding characteristics common to both solid-solution-strengthened and precipitation-hardened nickel-, iron-, and cobalt alloys. The remaining articles in this Section focus on characteristics of nonferrous high-temperature materials.

Table 1 is a reasonably inclusive list of the alloys that have found application as high-temperature materials and that have demonstrated some degree of weldability. (It should be noted that the refractory metals and alloys that are applied at high temperature are not listed in Table 1; information about these materials can be found in the article "Special Metallurgical Welding Considerations for Refractory Metals" in this Section of the Handbook.) Many of the alloys listed in Table 1 are general-purpose materials that find application in a full temperature range, from cryogenic temperatures on up. In addition, certain of these alloys may be used for their corrosion resistance or some other property not associated with high-temperature application. These considerations make it difficult to precisely estimate the amount of material used specifically for high-temperature applications, although a rough estimate would be of the order of 30,000 tons/year (27,000 tonnes/year). Most of the alloys listed are produced as wrought products; however, some cast materials appear on the list, and cast versions of many of the wrought alloys are produced.

TABLE 1 NOMINAL COMPOSITIONS OF HIGH-TEMPERATURE ALLOYS

ALLOY	COMPOSITION, %									
	Ni	**Cr**	**Co**	**Fe**	**Mo**	**Ti**	**W**	**Nb**	**Al**	**C**	OTHER
SOLID-SOLUTION NICKEL-BASE ALLOYS											
HASTELLOY N	72.0	7.0	. . .	5.0 MAX	16.0	0.5 MAX	. . .	. . .	. . .	0.06	**. . .**
HASTELLOY S	67.0	15.5	. . .	1.0	15.5	. . .	. . .	. . .	0.2	0.02 MAX	**0.02 LA**
HASTELLOY X	49.0	22.0	1.5 MAX	18.5	9.0	. . .	0.6	. . .	2.0	0.15	**. . .**
HAYNES 230	BAL	22	0-5	0-3	2	. . .	14	. . .	0.3	0.1	**0.5 MN, 0.4 SI, 0.02 LA, 0.005 B**
INCONEL 600	76.0	15.5	. . .	8.0	. . .	. . .	. . .	. . .	. . .	0.08	**0.25 CU MAX**
INCONEL 601	60.5	23.0	. . .	14.1	. . .	. . .	. . .	. . .	1.35	0.05	**0.5 CU MAX**
INCONEL 617	55.0	22.0	12.5	. . .	9.0	. . .	. . .	. . .	1.0	0.07	**. . .**
INCONEL 625	61.0	21.5	. . .	2.5	9.0	0.2	. . .	3.6	0.2	0.05	**. . .**
PRECIPITATION-HARDENABLE NICKEL-BASE ALLOYS											
GMR-235	63.0	15.5	. . .	10.0	5.25	2.0	. . .	. . .	3.0	0.15	**0.06 B**
INCONEL 702	79.5	15.5	. . .	0.4	. . .	0.7	. . .	. . .	3.4	0.04	**. . .**
INCONEL 706	41.5	16.0	. . .	37.5	. . .	1.75	. . .	2.9(A)	0.2	0.03	**. . .**
IN-713C	74.0	12.5	. . .	. . .	4.2	0.8	. . .	2.0	6.0	0.12	**0.012 B, 0.10 ZR**
INCONEL 718	52.5	19.0	. . .	18.5	3.0	0.9	. . .	5.0	0.5	0.08 MAX	**0.15 CU MAX**
INCONEL 722	75.0	15.5	. . .	7.0	. . .	2.5	. . .	. . .	0.7	0.04	**. . .**
INCONEL X-750	73.0	15.5	. . .	7.0	. . .	2.5	. . .	1.0	0.7	0.04	**0.25 CU MAX**
INCOLOY 901	42.5	12.5	. . .	36.2	6.0	2.7	. . .	. . .	. . .	0.10 MAX	**. . .**
M-252	56.5	19.0	10.0	<0.75	10.0	2.6	. . .	. . .	1.0	0.15	**0.005B**
RENÉ	55.0	19.0	11.0	<0.3	10.0	3.1	. . .	. . .	1.5	0.09	**0.01 B**
UDIMET 700	53.0	15.0	18.5	<1.0	5.0	3.4	. . .	. . .	4.3	0.07	**0.03 B**
WASPALOY	57.0	19.5	13.5	2.0 MAX	4.3	3.0	. . .	. . .	1.4	0.07	**0.006 B, 0.09 ZR**
SOLID-SOLUTION IRON-BASE ALLOYS											
16-25-6	25.0	16.0	. . .	50.7	6.0	. . .	. . .	. . .	. . .	0.06	**1.35 MN, 0.70 SI**
17-14 CUMO	14.0	16.0	. . .	62.4	2.5	0.3	. . .	0.4	. . .	0.12	**0.75 MN, 0.50 SI, 3.0 CU**
19-9 DL	9.0	19.0	. . .	66.8	1.25	0.3	1.25	0.4	. . .	0.30	**1.10 MN, 0.60 SI**
INCOLOY 800	32.5	21.0	. . .	45.7	. . .	0.38	. . .	. . .	0.38	0.05	**. . .**
INCOLOY 800H	33.0	21.0	. . .	45.8	. . .	. . .	. . .	. . .	. . .	0.08	**. . .**
INCOLOY 800HT	32.5	21	. . .	46	. . .	0.4	. . .	. . .	0.4	0.08	**0.8 MN, 0.008 S, 0.5**

											SI, 0.4 CU
INCOLOY 801	32.0	20.5	. . .	46.3	. . .	1.13	. . .	. . .	. . .	0.05	**. . .**
INCOLOY 802	32.5	21.0	. . .	44.8	. . .	0.75	. . .	. . .	0.58	0.35	**. . .**
MULTIMET (N-155)	20.0	21.0	20.0	32.2	3.0	. . .	2.5	1.0	. . .	0.15	**0.15 N, 0.02 LA, 0.02 ZR**
RA 330	36.0	19.0	. . .	45.1	. . .	. . .	. . .	. . .	. . .	0.05	**. . .**
PRECIPITATION-HARDENABLE IRON-BASE ALLOYS											
A-286	26.0	15.0	. . .	55.2	1.25	2.0	. . .	. . .	0.2	0.04	**0.005 B, 0.3 V**
DISCOLOY	26.0	14.0	. . .	55.0	3.0	1.7	. . .	. . .	0.25	0.06	**. . .**
HAYNES 556	21.0	22.0	20.0	29.0	3.0	. . .	2.5	0.1	0.3	0.10	**0.50 TA, 0.02 LA**
INCOLOY 903	0.1 MAX	38.0	15.0	41.0	0.1	1.4	. . .	3.0	0.7	0.04	**. . .**
INCOLOY 909	38	. . .	13	42	. . .	1.5	. . .	4.7	0.03	0.01	**0.4 SI**
INCOLOY 925	44	21	. . .	28	3	2.1	. . .	. . .	0.3	0.01	**1.8 CU**
SOLID-SOLUTION COBALT-BASE ALLOYS											
HAYNES 25 (-605)	10.0	20.0	50.0	3.0	. . .	. . .	15.0	. . .	. . .	0.10	**1.5 MN**
HAYNES 188	22.0	22.0	37.0	3.0 MAX	. . .	. . .	14.5	. . .	. . .	0.10	**0.04 LA**
S-816	20.0	20.0	42.0	4.0	4.0	. . .	4.0	4.0	. . .	0.38	**. . .**
STELLITE 6B	1.0	30.0	61.5	1.0	. . .	. . .	4.5	. . .	. . .	1.0	**. . .**
UMCO-50	. . .	28.0	49.0	21.0	. . .	. . .	. . .	. . .	. . .	0.12 MAX	**. . .**
PRECIPITATION-HARDENABLE COBALT-BASE ALLOYS											
AR-213	0.5 MAX	19.0	65.0	0.5 MAX	. . .	. . .	4.5	. . .	3.5	0.17	**6.5 TA, 0.15 ZR, 0.1 Y**
MP-35N	35.0	20.0	35.0	. . .	10.0	. . .	. . .	. . .	. . .	. . .	**. . .**
MP-159	**25.0**	**19.0**	**36.0**	**9.0**	**7.0**	**3.0**	**. . .**	**0.6**	**0.2**	**. . .**	**. . .**

(A) NB + TA

Applications of these materials include:

- HEAT TREATMENT FIXTURES AND FURNACE PARTS
- HEATING ELEMENTS USED FOR DOMESTIC OR INDUSTRIAL APPLICATIONS
- NUCLEAR AND FOSSIL FUEL POWER PLANT COMPONENTS (BOTH ROTATING PARTS AND STRUCTURAL COMPONENTS SUCH AS PIPING AND PUMP HARDWARE)
- GAS TURBINE ENGINE COMPONENTS (AGAIN, BOTH ROTATING PARTS AND THE STATIC STRUCTURAL MEMBERS)
- PYROCHEMICAL AND PYROMETALLURGICAL PROCESSING EQUIPMENT, PARTICULARLY WHERE HIGH TEMPERATURE IS COMBINED WITH HIGH STRESS AND/OR AN AGGRESSIVE ENVIRONMENT
- ENVIRONMENTAL REMEDIATION EQUIPMENT SUCH AS THAT USED TO PROCESS STACK GASES, RECUPERATORS USED TO RECOVER WASTE HEAT, AND INCINERATORS OR OTHER HARDWARE USED FOR HIGH-TEMPERATURE DESTRUCTION OF WASTE

A.C. Lingenfelter, Lawrence Livermore National Laboratory

General Welding Characteristics of High-Temperature Alloys

As with all materials, successful welding of high-temperature alloys requires a thorough understanding of the metallurgy of the individual alloys involved. Strengthening of these alloys is achieved by the precipitation of a finely dispersed age-hardening constituent in the matrix of the alloy and/or by the addition of solid-solution-strengthening alloy additions in the matrix. In addition, the strength of these materials can be significantly increased by careful control of their thermomechanical history. Hot-working finishing temperatures can be lowered below the recrystallization temperature of the alloy, resulting in the retention of the hot/cold work introduced in the final hot-working operations. This retained strain energy significantly increases the strength of the material, particularly when combined with subsequent aging heat treatments. Yield strengths in excess of 1380 MPa (200 ksi) are possible with at least a few of these alloys.

Effect of Welding on Performance and Properties. The thermal cycle(s) introduced by the welding process can significantly alter the physical properties of the base material in the region of the weld. The heat of welding can result in the solutioning of age-hardening constituents on other precipitates such as delta phase or carbides, which define the strength, ductility, corrosion resistance, grain size, or other properties of the material. Temperatures immediately adjacent to the weld will reach and exceed the recrystallization temperature, resulting in grain growth and elimination of the retained strain energy from thermomechanical processing. These effects can significantly alter the properties of the weld heat-affected zone (HAZ).

The interaction of the alloy with the welding process can introduce micro- and, in some cases, macrofissuring in the base metal, either in the as-welded condition or as a result of subsequent heat treatment. Variables such as grain size, chemical composition, and level of restraint have been linked to the sensitivity of materials to base metal fissuring.

Applications that require good creep strength and resistance to creep rupture frequently require that the material be in a grain-coarsened condition. Applying a high-energy-input welding process to a coarse-grain material can frequently result in HAZ micro- or macrofissures. The addition of elements such as boron and zirconium to enhance stress rupture life and ductility has been shown to increase fissuring sensitivity. Fortunately, many of these materials have inherently good ductility, and the micro- or macrofissures, although undesirable, may not detract from the useful service of the material. Each case needs to be evaluated on the basis of the fitness-for-service requirements of the particular application.

Heat Treatment. Postweld heat treatment is required for most age-hardenable materials so that they can develop their full strength and physical property potential. Yield-strength-level residual stresses are introduced by the welding thermal cycle. Direct age hardening of some of these alloys after welding can result in strain age cracking; this is particularly a problem with the aluminum/titanium-hardened alloys, which exhibit a very rapid age-hardening response. Because of the relaxation resistance of these alloys, the yield-strength-level stresses can result in short-time stress rupture failures. Preweld heat treatments, which put the material in a condition that promotes stress relaxation (generally an overaged condition), as well as postweld heat treatment cycles, which stress relieve the structure prior to the age-hardening treatment, can serve to minimize or eliminate the strain age cracking problem.

Alloys with very high hardener contents, such as cast IN-713C, were once considered unweldable because of their susceptibility to strain age cracking. However, IN-713C has been successfully repair welded by preheating the material into or slightly above the age-hardening temperature range before welding and by applying the appropriate postweld heat treatment to avoid cracking.

The sluggish aging response of the niobium-hardened alloys, such as Inconel 718, helps them avoid the strain age cracking problem. Direct aging of such alloys, while possible in most instances, needs to be evaluated where highly restrained conditions are imposed.

In some cases, non-age-hardenable solution-strengthened materials exposed to high postweld temperatures--whether from a postweld heat treatment or the application environment--experienced base metal macrofissuring. These instances generally occurred with highly restrained heavy-section weldments made in grain-coarsened base materials. Again, the apparent cause is the inability of the base metal to relax the yield-strength-level residual stress prior to suffering a short-time stress rupture failure.

For additional information about the heat treatment response of high-temperature materials, see the article "Postweld Heat Treatment of Nonferrous High-Temperature Materials" in this Section of the Handbook.

Effect of Contaminants on Weld Soundness. High-temperature alloys must be clean and free of machining oils and other contaminants if welds free of defects are to be achieved. Introducing contaminants to the fusion zone can lead to fissuring and porosity.

The presence of low-melting elements and alloys in the HAZ can lead to fissuring. This problem has been noted with the presence of copper, brass, and lead, to name a few. The low-melting material becomes liquid in the HAZ as the arc passes and the liquid penetrates the grain boundary. This grain boundary penetration reduces the strength of the boundary and, coupled with yield-strength-level stress, leads to microfissuring. Care must be taken when using copper chills and tooling not to deposit the copper on the surface of the base material. Lead can cause weld metal fissuring if introduced even in minute quantities to the fusion zone. Unfortunately, lead/brass/copper hammers are commonly used in shops to "adjust" the position of metal parts in the tooling. Instances have been noted where this practice resulted in weld heat-affected zone and/or weld fusion zone fissuring.

Most of the high-temperature alloys have excellent oxidation resistance because they develop a tightly adhering refractory oxide on the surface of the material. This is a highly useful property from an application standpoint, but it can lead to problems such as trapped oxide in the weld metal and lack-of-fusion defects at the weld metal/parent metal interface if the surface of the material to be welded is not free of oxide. The oxide cannot be removed by simple wire brushing; the metal surface may appear to be clean after brushing, but the oxide has only been polished. An aggressive abrasive grinding operation is needed to positively remove the oxide. Less-than-ideal inert gas protection can allow the formation of an oxide film on the surface of the deposited weld metal. Care needs to be taken to remove this oxide before multipass welding to avoid the problems of entrapped oxide and lack-of-fusion defects.

Welding Metallurgy of Nonferrous High-Temperature Materials

R.G. Thompson, University of Alabama at Birmingham

Introduction

THE PHYSICAL METALLURGY of nonferrous high-temperature materials that affects weldability centers on the precipitates used for age hardening (γ' and γ'') and those associated with solidification and solidification segregation (primarily, Laves and carbides). Age hardening creates problems associated with strain-age (that is, reheat) cracking, while carbides and Laves phases can initiate melting and cause cracking of heat-affected zone (HAZ) grain boundaries. The weldability of these alloys generally affects both the fabricator, who must weld complicated assemblies, and the foundry, which must make repair welds to salvage expensive castings. Thus, weldability problems are not solved by designating some alloys weldable and others unweldable. Weldability solutions are often needed for both weldable and unweldable alloys.

Strain-Age Cracking

Cracking is sometimes found in age-hardenable alloys that are slowly cooled or reheated through the hardening temperature range in the presence of residual or applied stress in a constrained condition. The rate of hardening is of primary importance relative to the heating or cooling rate through the hardening temperature range. For cracking to occur, the thermal cycle must allow sufficient hardening for the imposed stress to cause cracking. Of equal importance is the imposed stress, which must be sufficient in magnitude to initiate cracking. When these two phenomena occur simultaneously, they can result in severe cracking.

The root of the strain-age cracking problem from the metallurgical viewpoint is γ' precipitation. The nominal composition of γ' is Ni_3Al. However, the precise composition is even more complex. A typical γ' precipitate in alloy 731 is:

$$(\mathrm{NI}_{0.884}\mathrm{CO}_{0.070}\mathrm{CR}_{0.032}\mathrm{MO}_{0.008}\ \mathrm{V}_{0.003})_3(\mathrm{AL}_{0.632}\mathrm{TI}_{0.347} \\ \mathrm{V}_{0.013}\mathrm{CR}_{0.006}\mathrm{MO}_{0.002})$$

A typical range of heat treatments for hardening nickel superalloys is given in Table 1.

TABLE 1 TYPICAL HEAT TREATMENT PARAMETERS FOR HARDENING NICKEL-BASE SUPERALLOYS

TREATMENT	**TEMPERATURE**		TIME, H
	°C	**°F**	
HOMOGENIZE AND HOT ISOSTATIC PRESSING	NEAR SOLIDUS		**1-20**
SOLUTION ANNEAL	925-1150	1700-2100	**1-5**
UPPER AGE HARDENING	815-980	1500-1800	**10-25**
LOWER AGE HARDENING	**650-815**	**1200-1500**	**10-20**

After welding, the residual stress is relieved, and the maximum strength is obtained by a solution anneal and aging heat treatment. Problems arise when the welded structure is heated through the aging temperature range on its way to the solution temperature. Strain-age weldability is dependent on both the rate and the magnitude of γ' precipitation (Ref 1). Figure 1 shows a plot of weldability as a function of the (Ti + Al) content. Titanium and aluminum are the γ' strengtheners in many superalloys. When the (Al + Ti) level exceeds some critical value, strain-age cracking becomes a significant problem. The base metal can be effectively protected against strain-age cracking by welding in the overaged condition. This prevents aging during reheating.

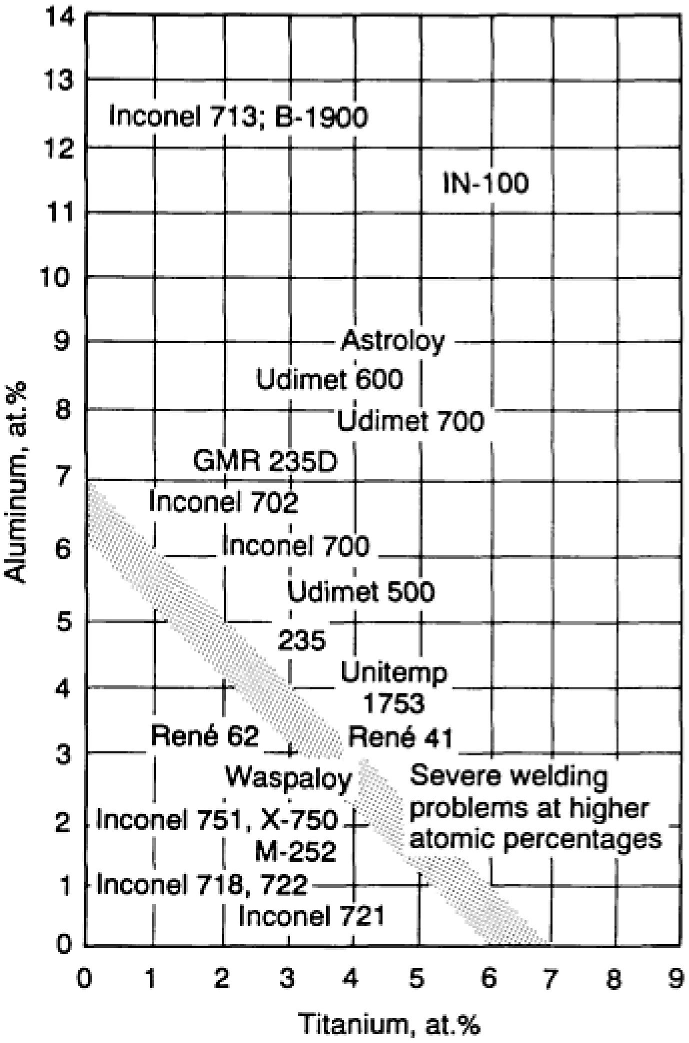

FIG. 1 PLOT OF ALUMINUM CONTENT VERSUS TITANIUM CONTENT TO INDICATE COMMERCIAL NICKEL-BASE ALLOYS THAT PRESENT WELDABILITY PROBLEMS WHEN (AL + TI) CONTENT EXCEEDS 6%. SOURCE: REF 1

Unfortunately, the HAZ cannot be protected in a similar manner. The welding process invariably produces a HAZ thermal cycle, which puts some of the age-hardening constituents into solution. During slow cooling or reheating these constituents can reprecipitate, age harden, and produce a crack-susceptible condition in the HAZ. This will be the case regardless of the preweld heat treatment.

Alloy 718 does not undergo strain-age cracking (Ref 2). The age hardening develops around a Ni_3Nb, γ'', precipitate. The γ'' precipitates at a much slower rate than the γ'. This allows alloys to be heated into the solution temperature range without suffering aging and the resultant strain-age cracking. Figure 2 compares the aging rates of several γ' alloys with those of alloy 718, which is strengthened with the γ'' precipitate. Hardening is retarded in the γ'' alloys, and they offer great latitude in methods for heating and cooling parts.

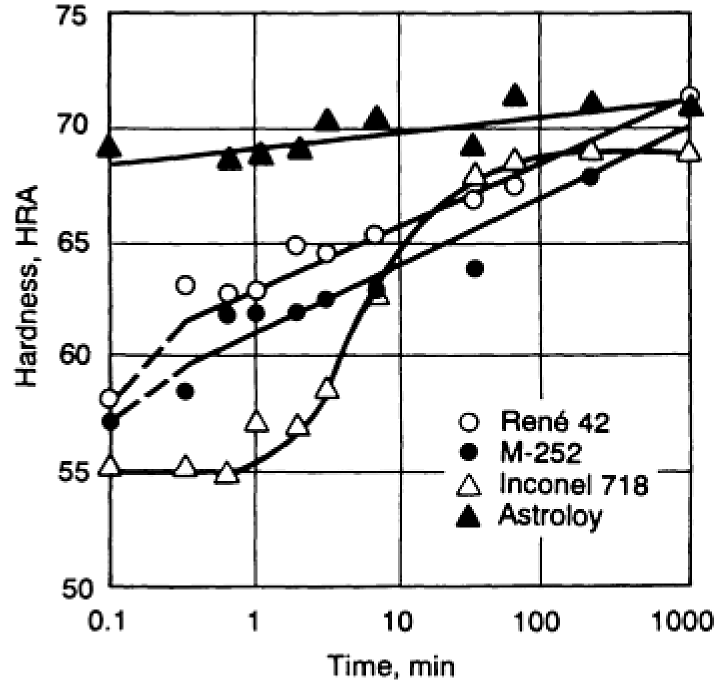

FIG. 2 PLOT OF HARDNESS VERSUS TIME TO SHOW AGE-HARDENING KINETICS OF SELECTED NICKEL-BASE ALLOYS. SOURCE: REF 3

Welded structures that require corrosion resistance but not high strength often will not need postweld heat, and can thus avoid strain-age cracking. However, if they are placed in service at elevated temperature, they may still crack during the initial heat-up cycle.

Strain-Age Cracking Susceptibility Curves. The circular-patch test is a high-restraint weld test that can be used to evaluate the sensitivity of an alloy to strain-age cracking. The test, when used with thermal treatments, can be used to produce crack-susceptibility C-curves. These are so named because of the characteristic "C" shape of the temperature-time space that separates the cracked from the uncracked behavior of this rate process. Figure 3 shows that, for a given C-curve behavior, different heating rates can produce cracked or uncracked plates. Figure 4 shows alloy chemistry variations that can affect the crack sensitivity of a given alloy.

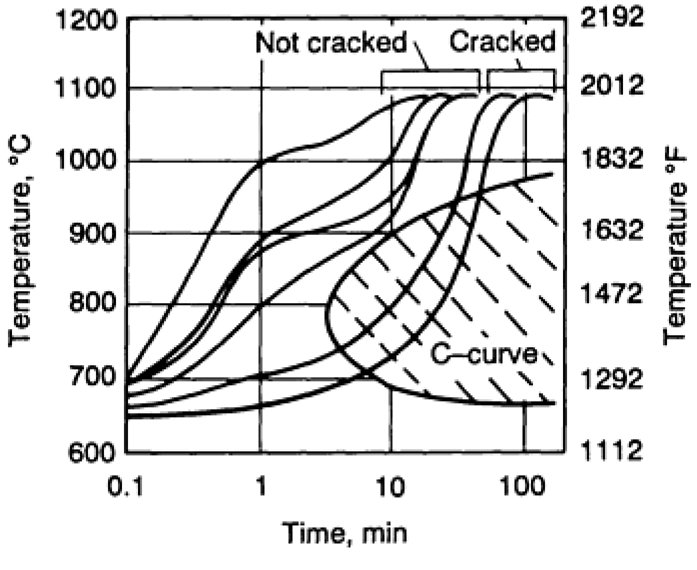

FIG. 3 EFFECT OF HEATING RATE ON THE CRACKING TENDENCY OF SOLUTION-ANNEALED (BEFORE WELDING) RENÉ 41 DURING POSTWELD HEAT TREATMENT. SOURCE: REF 4

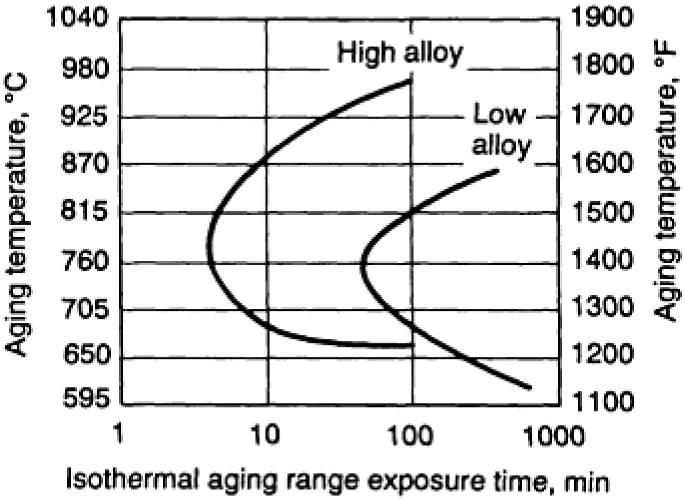

FIG. 4 EFFECT OF ALLOY COMPOSITION (HIGH ALLOY VERSUS LOW ALLOY CONCENTRATIONS OF IRON, SILICON, MANGANESE, AND SULFUR TAKEN COLLECTIVELY) ON THE CRACKING TENDENCY OF POSTWELD HEAT-TREATED RENÉ 41. SOURCE: REF 4

HAZ Liquation Cracking. Phases that form during solidification, such as MC carbides and Laves phases, have the potential to initiate melting in the HAZ during welding and spread along the grain boundaries (Fig. 5). The melting, often termed liquation, occurs because of a reaction between the dissolving precipitate and the matrix. When this melting is accompanied by sufficient thermal stress, cracks can form along the HAZ grain boundaries, and extend into the fusion zone. Such cracking may be termed liquation cracking, hot cracking, or microfissuring. A number of alloy systems are known to experience liquation cracking; some are listed in Table 2.

TABLE 2 CONSTITUTIONAL LIQUATION IN MULTICOMPONENT SYSTEMS

ALLOY SYSTEM	LIQUATING PHASE
HASTELLOY X	**$M_6C^{(A)}$**
INCONEL 600	**CR_7C_3; TI(CN)**
18% MARAGING STEEL	**TITANIUM SULFIDE$^{(B)}$**
AUSTENITIC ALLOY A-286	**TIC OR TI(CN)**
HIGH-STRENGTH MAGNESIUM-BASE	**ND; Y-CONTAINING PHASE ALLOY**

Source: Ref 5.

Source: Ref 7

(A)

(B)

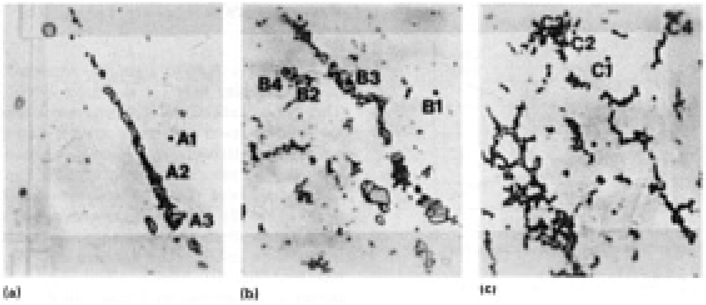

FIG. 5 LIQUATION OF A NIOBIUM CARBIDE STRINGER IN A NICKEL-BASE ALLOY 718 WIRE. (A) STRINGER BEFORE ONSET OF LIQUATION. (B) STRINGER IN INITIAL STAGE OF LIQUATION AS IT IS HEATED BY WELDING THERMAL CYCLE. (C) MOVEMENT OF STRINGER LIQUATION INTO GRAIN BOUNDARIES OF ALLOY. 500×

Liquid metal is invariably associated with the HAZ except in the unusual case of very pure, single-element metals. This is so because the HAZ stretches from the base metal to the edge of the fusion zone and will include all or part of the mushy zone (a pure metal has no mushy zone). This liquid is almost always found along the grain boundaries (Fig. 6). The few exceptions are alloys that have miscibility gaps, such as leaded brass. These alloys can have liquid from the low-melting-point phase, such as lead, which does not wet the grain boundaries. Such nonwetting behavior is associated with alloys where the solute is insoluble in the parent phase. It can be concluded that most engineering alloys contain a HAZ during welding that has a mushy zone full of intergranular liquid.

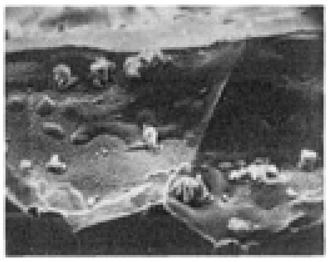

FIG. 6 INTERGRANULAR FRACTURE SURFACE SHOWING PARTIAL INTERGRANULAR WETTING WITH BOTH SOLID-SOLID AND LIQUID-LIQUID BONDING. 500×

The mushy-zone liquid generally does not contribute to poor weldability, because during the normal course of solidification it is always open to the fusion zone. The fusion zone acts as a source of liquid to backfill or heal shrinkage or cracks that might form.

Metallurgically, liquation cracking occurs when an alloy containing a susceptible second phase is heated at such a rate that the second phase cannot dissolve before the alloy reaches a system solidus that causes melting. This melting will invariably occur at the interface between the precipitate and the matrix, as can be shown with an analysis of the appropriate phase diagram. The molten liquid will spread along the grain boundaries if its solute is soluble in the matrix or if there is sufficient impurity to enhance wetting (Ref 10). This liquation can occur below the system solidus and thus be located remotely from the HAZ mushy zone. Such a location prevents backfilling and promotes liquation cracking. The fact that liquation from precipitates can extend well away from the fusion zone not only increases the chance of liquation cracking but also increases the potential length of the crack, accentuating the stress intensity of the crack.

Metallurgical evaluation of HAZ liquation reveals three stages of microstructural and compositional evolution (Fig. 7). An initial stage develops where the liquating precipitate is in contact with the liquid. In a later stage of evolution, the precipitate is completely liquated, but the intergranular liquid is stable and may even be growing. Finally, the intergranular liquid begins to solidify and may retract from the grain boundary.

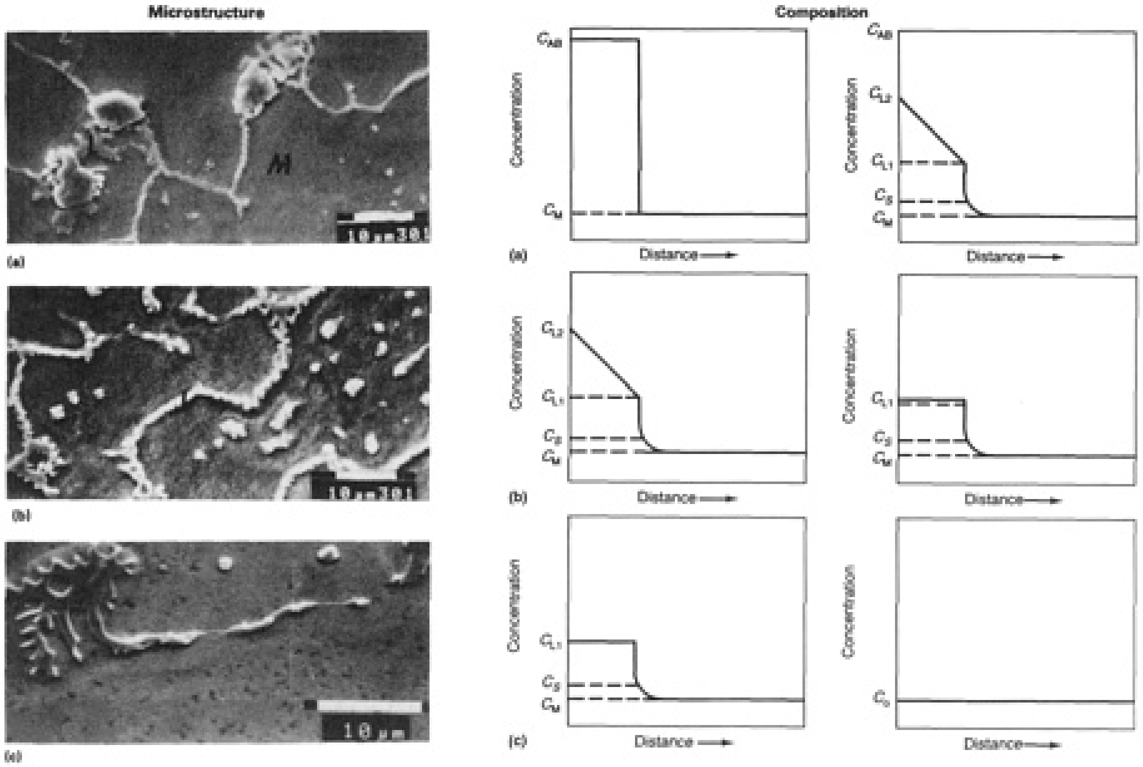

FIG. 7 THREE STAGES OF HAZ LIQUATION SHOWN RELATIVE TO THE TYPICAL TRANSFORMATION IN MICROSTRUCTURE AND COMPOSITION OF NICKEL-BASE ALLOY 718. (A) STAGE 1: PRECIPITATE LIQUATION. (B) STAGE 2: ELIMINATION OF COMPOSITION GRADIENT. (C) STAGE 3: SOLIDIFICATION

References cited in this section

1. M. PRAGER AND C.S. SHIRA, WELDING OF PRECIPITATION-HARDENING NICKEL-BASE ALLOYS, *WELD. RES. COUNC. BULL.,* NO. 128, 1968
2. H.L. EISELSTEIN, *ADVANCES IN THE TECHNOLOGY OF STAINLESS STEELS AND RELATED ALLOYS,* STP 369, ASTM, P 62-79
3. R.M. WILSON JR. AND L.W.G. BURCHFIELD, *WELD. J.,* VOL 35, 1956, P 32S
4. T.F. BERRY AND W.P. HUGHES, *WELD. J.,* VOL 46, 1969, P 505S
5. D.S. DUVALL AND W.A. OWCZARSKI, FURTHER HEAT-AFFECTED ZONE STUDIES IN HEAT RESISTANT NICKEL ALLOYS, *WELD. J.,* VOL 46, 1967, P 423S
7. J.J. PEPE AND W.F. SAVAGE, EFFECTS OF CONSTITUTIONAL LIQUATION IN 18-NI MARAGING STEEL WELDMENTS, *WELD. J.,* VOL 45, 1967, P 411S-422S
10. C.S. SMITH, GRAINS, PHASES AND INTERFACES: AN INTERPRETATION OF MICROSTRUCTURE, *TRANS. AIME,* VOL 175, 1948, P 15-51

Welding Metallurgy of Nonferrous High-Temperature Materials

R.G. Thompson, University of Alabama at Birmingham

Parameters Affecting Liquation Cracking

Grain Size. A large grain size promotes liquation cracking (Ref 11), as shown in Fig. 8. This can be understood in several ways. When the HAZ begins to accumulate strain from residual stress, grain boundary sliding is one mechanism that operates to accommodate the strain. Large grain size does not accommodate the strain as readily as smaller grain size, and the potential for crack initiation at grain boundary triple points, and therefore liquation cracking, is increased (Ref 13). Another effect of grain size occurs when precipitates liquate and the liquid spreads along the boundaries. Material of large grain size will accumulate a thicker layer of liquid for a given amount (volume fraction) of precipitate, because larger grains have less grain boundary surface area (Ref 14). The thicker liquid layer takes longer to resolidify during cooling, as has been shown in alloy 718 (Ref 11) and through computer simulations (Ref 14, 15). Anything that increases the life of the intergranular liquid relative to the onset of residual stress accentuates liquation cracking.

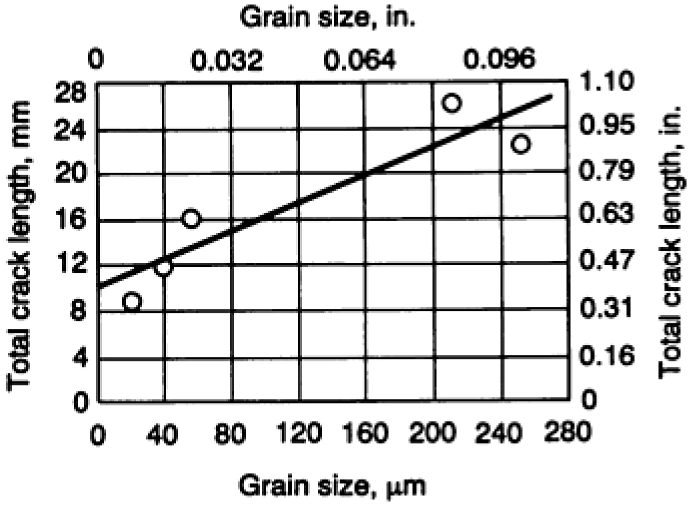

FIG. 8 PLOT OF TOTAL CRACK LENGTH VERSUS GRAIN SIZE TO SHOW EFFECT OF MICROFISSURING IN ALLOY 718 WHEN EVALUATING NICKEL-BASE ALLOY WITH SPOT VARESTRAINT TEST. SOURCE: REF 12

Liquation cracking is sensitive to the amount and location of second-phase precipitates. The size of precipitates in the HAZ, as well as the location relative to the position of grain boundaries, changes during the welding thermal cycle. Precipitates, especially those susceptible to constitutional liquation, dissolve during the thermal cycle (Ref 16).

Constitutional liquation can be illustrated with the simple binary eutectic shown in Fig. 9(a). Alloy C_O has a certain volume fraction of second-phase particles, AB, at room temperature. When the alloy is rapidly heated to a temperature, T, above the binary eutectic temperature, T_E, there may be insufficient time for complete dissolution of AB. Therefore, at temperatures above T_E, the AB phase is still present. Dissolution of AB above T_E takes place by the formation of a liquid phase that subsequently resolidifies in accordance with phase diagram requirements. The concentration gradients in the matrix adjacent to the AB particle are shown in Fig. 9(b). The terms C_{L1} and C_{L2} are the concentrations of liquid in equilibrium with γ and AB, respectively, at temperature T. The term C_S is the concentration of γ in equilibrium with liquid, and C_M is the concentration of the γ matrix in equilibrium with the AB particle at room temperature.

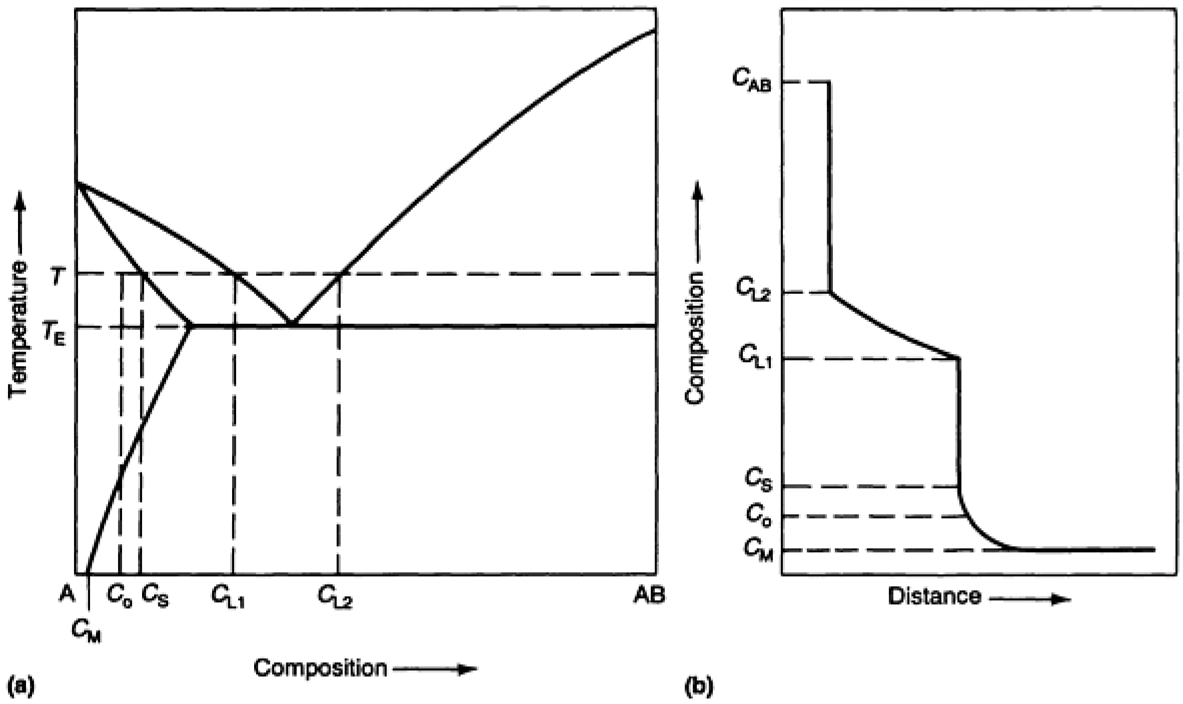
Temperature
T
T_E
A
C_M
C_o
C_S
C_{L1}
C_{L2}
AB
Composition
(a)
Composition
C_{AB}
C_{L2}
C_{L1}
C_S
C_o
C_M
Distance
(b)

FIG. 9 EFFECT OF TEMPERATURE AND DISTANCE ON LIQUATION IN A BINARY SYSTEM. (A) CONSTITUTIONAL LIQUATION OF AB PARTICLES. (B) CONCENTRATION GRADIENTS IN FRONT OF THE LIQUATING AB PARTICLE

Behavior of NbC Dissolved in Nickel (Ref 16). A computer simulation of NbC dissolved in nickel is shown in Fig. 10. Some precipitates dissolve completely and disappear, while others become small and lose their attachment to the grain boundary. This is shown by computer simulation in Fig. 11, where *R* is the fraction of initial grain boundary precipitates remaining on the HAZ grain boundaries during heating.

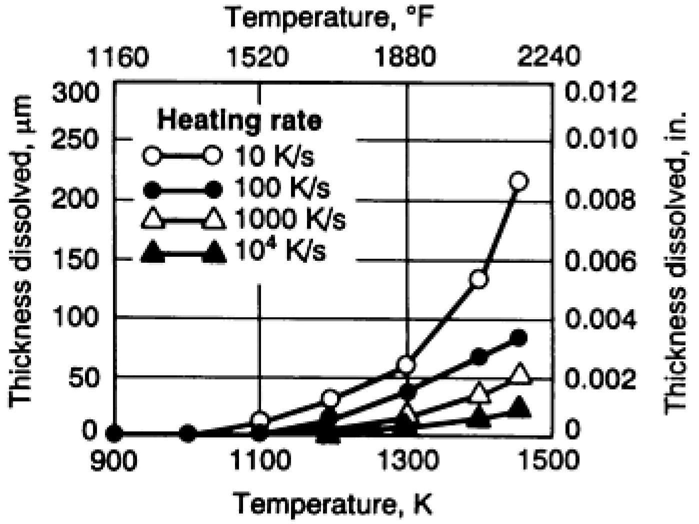

FIG. 10 DISSOLUTION KINETICS OF NBC PRECIPITATES IN NICKEL. SOURCE: REF 16

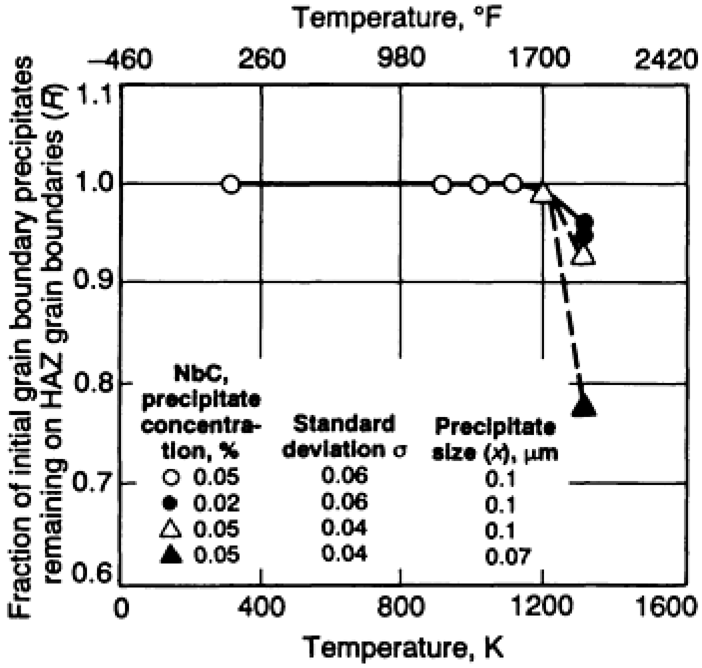

FIG. 11 PLOT OF *R* VERSUS TEMPERATURE AS A FUNCTION OF NBC PRECIPITATE CONCENTRATION, σ, AND *X* DURING CONTINUOUS 100 K/S (180 °F/S) HEATING. THE RATIO, *R*, HAS A STEEP DROP IN VALUE AS THE TEMPERATURE RISES ABOVE 1250 K (1790 °F). SOURCE: REF 16

Amount of Liquating Precipitate. A large amount of liquating precipitate can have both positive and negative effects relative to liquation cracking. On the negative side, a large amount of liquation produces a thicker layer of intergranular liquid. Just as for large-grain-size metal, a thicker layer takes longer to solidify and accentuate liquation cracking. However, on the positive side, a larger amount of liquation produces a wider path for backfilling from the mushy zone and thus can help heal cracking. This would be especially effective for liquation that occurs at a temperature near the bulk solidus. Because wrought alloys and many cast alloys are metallurgically designed to minimize liquation-type precipitates, the initial amount of precipitates is low (0.5 to 2.0 vol% for both wrought and cast alloy 718). Increasing an initially small amount of precipitate increases liquation cracking (Ref 11). Thus, using large amounts of precipitate to reduce liquation cracking would seem prudent only under closely guarded conditions where the precipitate is well connected to the HAZ mushy zone.

Finally, a more sophisticated analysis of the effect of precipitate volume fraction on liquation cracking must take into account the rate at which the precipitates dissolve before liquation (because all precipitates that liquate below the solidus try to dissolve first). While a large volume fraction of precipitates may be present in the HAZ before welding, those precipitates may disappear quickly during heating. Thus, the rate of precipitate dissolution and the welding thermal cycle must be known to predict the effect of precipitate volume fraction on cracking.

Type of Precipitate. Two types of precipitates liquate in the HAZ:

- CONSTITUTIONALLY LIQUATING PRECIPITATES (REF 7)
- EQUILIBRIUM MELTING PRECIPITATES (REF 12)

Constitutionally liquating precipitates can begin melting below the bulk solidus because their liquation is tied to the fact that the alloy has been forced to support their existence beyond some equilibrium limit. Such nonequilibrium processing can result from two factors: (1) a rapid heating rate that takes a normally low-temperature phase above its equilibrium stability temperature or (2) coring-type segregation during casting. Coring-type segregation causes the system to create a nonequilibrium phase (such as Laves phase in alloy 718) or an overabundance of an equilibrium phase (such as MC carbides). Both the nonequilibrium phase and the excessive fraction of the equilibrium phase can constitutionally liquate (Ref 4).

Equilibrium Melting Precipitates. The equilibrium phase melts when heated above its solidus temperature, producing HAZ liquid.

Precipitates Versus Liquation Cracking. The two types of precipitates described above are different relative to liquation cracking because their melting ranges are different relative to the mushy zone. Because equilibrium precipitates will always liquate at or above the bulk solidus, they will be connected with the mushy zone and have access to large volumes of liquid for backfilling and healing. The nonequilibrium precipitates will have the potential to liquate below the bulk solidus and thus be separated from backfilling liquid from the larger mushy zone. These nonequilibrium precipitates will have a greater potential for liquation cracking.

Solidification carbides such as MC carbides can be present in overabundance due to coring during solidification. Thus, they can be present in the nonequilibrium form but with a melting range that spans the subsolidus to the bulk liquidus. The liquation of the carbides should produce a continuous path to the mushy zone and thus provide a network for backfilling and healing cracks. However, the liquation of an MC carbide produces such great local solute enrichment at the groin boundary that the resolidification of these boundaries can be greatly delayed. This delay in resolidification separates them from the bulk solidification of the mushy zone and makes these boundaries potential sites for liquation cracking.

Impurities and Heat Treatment. Impurity metallic elements have long been the subject of intergranular-cracking investigations (Ref 17, 18, 19, 20, 21). Some elements (for example, sulfur, phosphorus, antimony, arsenic, bismuth, and tin) concentrate in the grain boundaries of various systems. Other elements (for example, boron), affect intergranular properties (for example, stress rupture strength).

Elements that have been identified as potential sources of liquation cracking are sulfur, phosphorus, boron, and lead. It has been suggested that if the impurity level restrictions shown below are observed, increased liquation cracking will not be observed (Ref 20):

ELEMENT	COMPOSITION, WT% (MAXIMUM)
SULFUR	**0.015**
PHOSPHORUS	**0.015**
SILICON	**0.02**
OXYGEN	**0.005**
NITROGEN	**0.005**

Some researchers separate the behavior of impurities into two groups: (1) those that are present as precipitates (for example, borides, sulfides, and so on) and (2) those that are not precipitated but are enriched at the grain boundary and the precipitate interfaces due to segregation. The latter category would also include elements such as boron and sulfur. However, the state in which the impurities reside is often debatable, and the issue here is their effect on HAZ hot cracking. Thus, no attempt will be made to differentiate between impurity precipitates or segregation.

Liquation is not necessarily caused by impurities. Impurities magnify the problem by:

- INCREASING THE WETTING OF THE INTERGRANULAR LIQUID (REF 11)
- DEPRESSING THE SOLIDUS TEMPERATURE
- FORMING LOW-MELTING-POINT PRECIPITATES/EUTECTICS
- INCREASING THE AMOUNT OF LIQUID PRESENT ON THE GRAIN BOUNDARIES

Sulfur and carbon provide a unique case. When the carbon content is increased in alloy 718, the amount of carbide and the potential for liquation cracking both increase (Fig. 12). If the sulfur content is increased, the potential for liquation cracking also increases about the same amount for both carbon levels (Fig. 12). One reason for the increased cracking susceptibility with sulfur is that the sulfur becomes trapped between the carbide and the matrix during casting (Table 3).

A high sulfur level reduces the solidification temperature of the carbide and appears to increase the wetting of the intergranular liquid in the HAZ during cooling (Ref 22).

TABLE 3 INTERFACIAL SULFUR LEVELS IN EXCESS OF SPUTTER-CLEANED SURFACES FOR HEAT-TREATED ALLOY 718 SPECIMENS RELATIVE TO AS-CAST ALLOY 718 FROM THE SAME HEAT

HEAT-TREATED CONDITION	SULFUR CONTENT, %	
	CARBIDE SURFACE	GRAIN BOUNDARY
AS-CAST	4.3	**0.11**
HOMOGENIZED	1.8	**0.03**
SOLUTION ANNEALED	1.1	**0.00**
STEP-COOLED	**0.1**	**2.20**

Source: Ref 21

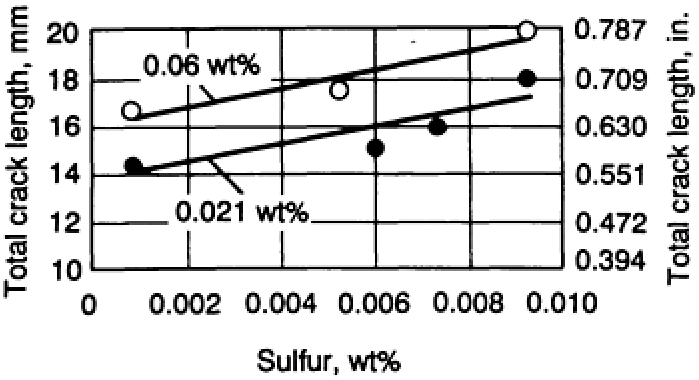

FIG. 12 PLOT OF TOTAL CRACK LENGTH VERSUS SULFUR CONTENT AS A FUNCTION OF TWO SELECTED BULK CARBON LEVELS TO SHOW MICROFISSURING SUSCEPTIBILITY OF CAST ALLOY 718. SOURCE: REF 11

Impurities may also affect the response of liquation cracking to heat treatment. Heat treatment has been shown to have either an adverse or a beneficial effect on liquation cracking (Ref 5, 23, 24). This has been shown for alloy 718 (Fig. 13) to be a function of carbon content. This effect might be partially explained by the change in impurity distribution (in this example, sulfur) during heat treatment (Ref 21, 25). As given in Table 3, sulfur moves from the carbide matrix interface onto the grain boundaries during age-hardening treatment. Simultaneously, liquation cracking increases during heat treatment. Conversely, the solution-anneal heat treatment cleans the grain boundary of any residual sulfur and reduces the potential for liquation cracking. Many other metallurgical events occur simultaneously during these heat treatments, and it has not been shown conclusively which of these events (if not all) affect cracking.

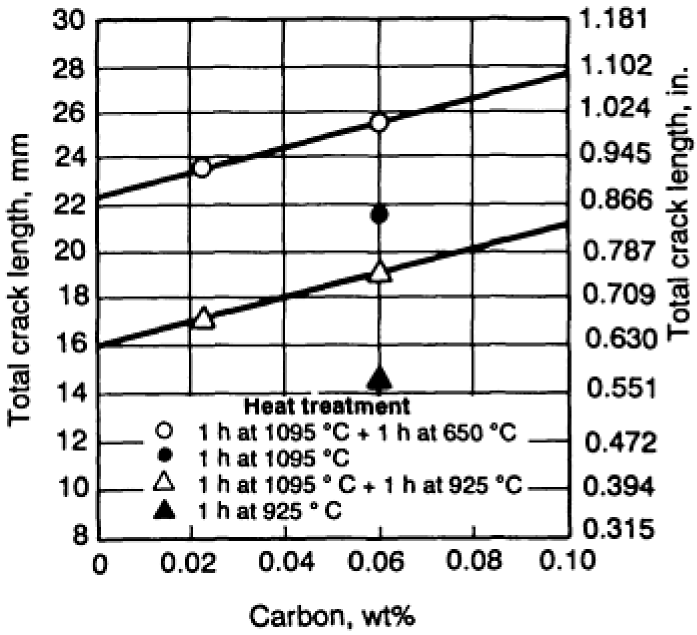

FIG. 13 PLOT OF TOTAL CRACK LENGTH VERSUS CARBON CONTENT AS A FUNCTION OF SELECTED HEAT TREATMENTS FOR CAST ALLOY 718 CONTAINING 0.009% S TO SHOW MICROFISSURING SUSCEPTIBILITY OF THE MATERIAL

Reducing Susceptibility to Liquation Cracking. Welding parameters and fabrication sequence can often be adjusted to reduce the possibility of liquation cracking. In addition, there are several metallurgical conditions that minimize HAZ liquation cracking:

- GRAIN SIZE SHOULD BE MINIMIZED
- IF PRECIPITATES ARE DESIRED FOR GRAIN SIZE CONTROL, SOLIDIFICATION PRECIPITATES SUCH AS MC CARBIDES SHOULD BE USED. THESE WILL MINIMIZE LIQUID PENETRATION INTO THE HAZ AND PROMOTE BACKFILLING AND HEALING OF CRACKS, BECAUSE THEIR LIQUATION TEMPERATURE IS SIMILAR TO THAT OF THE MUSHY ZONE
- THE AMOUNT OF PRECIPITATE THAT LIQUATES SHOULD BE MINIMIZED. IF A PARTICULARLY DIFFICULT CRACKING PROBLEM CANNOT BE RESOLVED ANY OTHER WAY, THEN INCREASING THE AMOUNT OF PRECIPITATE MAY PRODUCE ENOUGH LIQUATION TO INITIATE BACKFILLING FROM THE MUSHY ZONE AND PROMOTE HEALING OF CRACKS
- IMPURITY CONTENT SHOULD BE MINIMIZED
- WELDING SHOULD TAKE PLACE WHEN THE ALLOY HAS UNDERGONE SOME COMBINATION OF THE SOLUTIONED AND/OR HOMOGENIZED CONDITIONS, FOLLOWED BY RAPID COOLING (DIRECT QUENCHING SHOULD BE AVOIDED). THESE HEAT TREATMENTS MINIMIZE IMPURITY CONCENTRATION ON THE GRAIN BOUNDARIES

References cited in this section

4. T.F. BERRY AND W.P. HUGHES, *WELD. J.*, VOL 46, 1969, P 505S
5. D.S. DUVALL AND W.A. OWCZARSKI, FURTHER HEAT-AFFECTED ZONE STUDIES IN HEAT RESISTANT NICKEL ALLOYS, *WELD. J.*, VOL 46, 1967, P 423S
7. J.J. PEPE AND W.F. SAVAGE, EFFECTS OF CONSTITUTIONAL LIQUATION IN 18-NI MARAGING STEEL WELDMENTS, *WELD. J.*, VOL 45, 1967, P 411S-422S

11. R.G. THOMPSON, D.E. MAYO, AND B. RADHAKRISHNAN, ON THE RELATIONSHIP BETWEEN CARBON CONTENT, MICROSTRUCTURE, AND INTERGRANULAR HOT CRACKING IN CAST NICKEL ALLOY 718, *METALL. TRANS. A*, VOL 22, 1991, P 557-567
12. B. RADHAKRISHNAN AND R.G. THOMPSON, MODELING OF MICROSTRUCTURAL EVOLUTION IN THE WELD HAZ, *METAL SCIENCE OF JOINING*, M.J. CIESLAK *ET AL.*, ED., TMS, OCT 1991, P 31-40
13. R.G. THOMPSON, J.J. CASSIMUS, D.E. MAYO, AND J.R. DOBBS, THE RELATIONSHIP BETWEEN GRAIN SIZE AND MICROFISSURING IN ALLOY 718, *WELD. J.*, VOL 63, 1984, P 91S
14. B. RADHAKRISHNAN AND R.G. THOMPSON, THE KINETICS OF INTERGRANULAR LIQUATION IN THE HAZ OF ALLOY 718, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY*, PROC. 2ND INT. CONF. TRENDS IN WELDING RESEARCH, ASM INTERNATIONAL, 1989, P 637-648
15. B. RADHAKRISHNAN AND R.G. THOMPSON, A MODEL FOR THE SOLIDIFICATION OF GRAIN BOUNDARY LIQUID IN THE HAZ OF WELDS, *METALL. TRANS.*, JUNE 1992
16. Y. SHEN, B. RADHAKRISHNAN, AND R.G. THOMPSON, MONTE CARLO SIMULATION OF GRAIN GROWTH IN THE HAZ, *PROC. 3RD LNT. CONF. TRENDS IN WELDING RESEARCH*, ASM INTERNATIONAL, JUNE 1992
17. T.J. MORRISON, C.S. SHIRA, AND L.A. WEISENBERG, EFFECTS OF MINOR ELEMENTS ON THE WELDABILITY OF HIGH NICKEL ALLOYS, *PROC. WELD. RES. SYMP.*, AWS, VOL 93, 1967
18. W.A. OWCZARSKI, *EFFECTS OF MINOR ELEMENTS ON THE WELDABILITY OF HIGH-NICKEL ALLOYS*, WELDING RESEARCH COUNCIL, 1969, P 6
19. W. YENISCAVICH AND C.W. FOX, *EFFECTS OF MINOR ELEMENTS ON THE WELDABILITY OF HIGH-NICKEL ALLOYS*, WELDING RESEARCH COUNCIL, 1969, P 24
20. D.A. CANONICO *ET AL.*, *EFFECTS OF MINOR ELEMENTS ON THE WELDABILITY OF HIGH-NICKEL ALLOYS*, WELDING RESEARCH COUNCIL, 1969, P 68
21. R.G. THOMPSON, M.C. KOOPMAN, AND B.H. KING, GRAIN BOUNDARY CHEMISTRY OF ALLOY 718-TYPE ALLOYS, *SUPERALLOY 718, 625 AND DERIVATIVES*, E. LORIA, ED., TMS, 1991, P 53-70
22. C. CHEN, R.G. THOMPSON, AND D.W. DAVIS, A STUDY OF EFFECTS OF PHOSPHORUS, SULFUR, BORON AND CARBON ON LAVES AND CARBIDE FORMATION IN ALLOY 718, *SUPERALIOY 718, 625 AND DERIVATIVES*, E. LORIA, ED., TMS, 1991, P 81-96
23. R.G. THOMPSON AND S. GENCULU, MICROSTRUCTURAL EVALUATION IN THE HAZ OF INCONEL 718 AND CORRELATIONS WITH THE HOT DUCTILITY TEST, *WELD. J.*, VOL 62, 1983, P 337S-345S
24. R.G. THOMPSON, J.R. DOBBS, AND D.E. MAYO, THE EFFECT OF HEAT TREATMENT ON THE MICROFISSURING IN ALLOY 718, *WELD. J.*, VOL 65, 1986, P 299S-304S
25. R.G. THOMPSON, B. RADHAKRISHNAN, AND D.E. MAYO, GRAIN BOUNDARY CHEMISTRY CONTRIBUTIONS TO INTERGRANULAR HOT CRACKING, *J. PHYSIQUE-COLLOQUE C5*, SUPPL. NO. 10, 1988, P 471-482

Welding Metallurgy of Nonferrous High-Temperature Materials

R.G. Thompson, University of Alabama at Birmingham

References

1. M. PRAGER AND C.S. SHIRA, WELDING OF PRECIPITATION-HARDENING NICKEL-BASE ALLOYS, *WELD. RES. COUNC. BULL.*, NO. 128, 1968
2. H.L. EISELSTEIN, *ADVANCES IN THE TECHNOLOGY OF STAINLESS STEELS AND RELATED*

ALLOYS, STP 369, ASTM, P 62-79

3. R.M. WILSON JR. AND L.W.G. BURCHFIELD, *WELD. J.,* VOL 35, 1956, P 32S
4. T.F. BERRY AND W.P. HUGHES, *WELD. J.,* VOL 46, 1969, P 505S
5. D.S. DUVALL AND W.A. OWCZARSKI, FURTHER HEAT-AFFECTED ZONE STUDIES IN HEAT RESISTANT NICKEL ALLOYS, *WELD. J.,* VOL 46, 1967, P 423S
6. B. WEISS, G.E. GROTKE, AND R. STICKLER, *WELD. J.,* VOL 49, 1970, P 471S-487S
7. J.J. PEPE AND W.F. SAVAGE, EFFECTS OF CONSTITUTIONAL LIQUATION IN 18-NI MARAGING STEEL WELDMENTS, *WELD. J.,* VOL 45, 1967, P 411S-422S
8. J.R. BROOKS, *WELD J.,* VOL 53, 1974, P 517-523
9. W.A. BAESLACK III, S.J. SAVAGE, AND F.H. FROES, *J. MATER. SCI. LETT.,* 1986, P 935-939
10. C.S. SMITH, GRAINS, PHASES AND INTERFACES: AN INTERPRETATION OF MICROSTRUCTURE, *TRANS. AIME,* VOL 175, 1948, P 15-51
11. R.G. THOMPSON, D.E. MAYO, AND B. RADHAKRISHNAN, ON THE RELATIONSHIP BETWEEN CARBON CONTENT, MICROSTRUCTURE, AND INTERGRANULAR HOT CRACKING IN CAST NICKEL ALLOY 718, *METALL. TRANS. A,* VOL 22, 1991, P 557-567
12. B. RADHAKRISHNAN AND R.G. THOMPSON, MODELING OF MICROSTRUCTURAL EVOLUTION IN THE WELD HAZ, *METAL SCIENCE OF JOINING,* M.J. CIESLAK *ET AL.,* ED., TMS, OCT 1991, P 31-40
13. R.G. THOMPSON, J.J. CASSIMUS, D.E. MAYO, AND J.R. DOBBS, THE RELATIONSHIP BETWEEN GRAIN SIZE AND MICROFISSURING IN ALLOY 718, *WELD. J.,* VOL 63, 1984, P 91S
14. B. RADHAKRISHNAN AND R.G. THOMPSON, THE KINETICS OF INTERGRANULAR LIQUATION IN THE HAZ OF ALLOY 718, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* PROC. 2ND INT. CONF. TRENDS IN WELDING RESEARCH, ASM INTERNATIONAL, 1989, P 637-648
15. B. RADHAKRISHNAN AND R.G. THOMPSON, A MODEL FOR THE SOLIDIFICATION OF GRAIN BOUNDARY LIQUID IN THE HAZ OF WELDS, *METALL. TRANS.,* JUNE 1992
16. Y. SHEN, B. RADHAKRISHNAN, AND R.G. THOMPSON, MONTE CARLO SIMULATION OF GRAIN GROWTH IN THE HAZ, *PROC. 3RD LNT. CONF. TRENDS IN WELDING RESEARCH,* ASM INTERNATIONAL, JUNE 1992
17. T.J. MORRISON, C.S. SHIRA, AND L.A. WEISENBERG, EFFECTS OF MINOR ELEMENTS ON THE WELDABILITY OF HIGH NICKEL ALLOYS, *PROC. WELD. RES. SYMP.,* AWS, VOL 93, 1967
18. W.A. OWCZARSKI, *EFFECTS OF MINOR ELEMENTS ON THE WELDABILITY OF HIGH-NICKEL ALLOYS,* WELDING RESEARCH COUNCIL, 1969, P 6
19. W. YENISCAVICH AND C.W. FOX, *EFFECTS OF MINOR ELEMENTS ON THE WELDABILITY OF HIGH-NICKEL ALLOYS,* WELDING RESEARCH COUNCIL, 1969, P 24
20. D.A. CANONICO *ET AL., EFFECTS OF MINOR ELEMENTS ON THE WELDABILITY OF HIGH-NICKEL ALLOYS,* WELDING RESEARCH COUNCIL, 1969, P 68
21. R.G. THOMPSON, M.C. KOOPMAN, AND B.H. KING, GRAIN BOUNDARY CHEMISTRY OF ALLOY 718-TYPE ALLOYS, *SUPERALLOY 718, 625 AND DERIVATIVES,* E. LORIA, ED., TMS, 1991, P 53-70
22. C. CHEN, R.G. THOMPSON, AND D.W. DAVIS, A STUDY OF EFFECTS OF PHOSPHORUS, SULFUR, BORON AND CARBON ON LAVES AND CARBIDE FORMATION IN ALLOY 718, *SUPERALIOY 718, 625 AND DERIVATIVES,* E. LORIA, ED., TMS, 1991, P 81-96
23. R.G. THOMPSON AND S. GENCULU, MICROSTRUCTURAL EVALUATION IN THE HAZ OF INCONEL 718 AND CORRELATIONS WITH THE HOT DUCTILITY TEST, *WELD. J.,* VOL 62, 1983, P 337S-345S
24. R.G. THOMPSON, J.R. DOBBS, AND D.E. MAYO, THE EFFECT OF HEAT TREATMENT ON THE MICROFISSURING IN ALLOY 718, *WELD. J.,* VOL 65, 1986, P 299S-304S
25. R.G. THOMPSON, B. RADHAKRISHNAN, AND D.E. MAYO, GRAIN BOUNDARY CHEMISTRY

CONTRIBUTIONS TO INTERGRANULAR HOT CRACKING, *J. PHYSIQUE-COLLOQUE C5*, SUPPL. NO. 10, 1988, P 471-482

Postweld Heat Treatment of Nonferrous High-Temperature Materials

Steven C. Ernst, Haynes International, Inc.

Introduction

THE NONFERROUS HIGH-TEMPERATURE ALLOYS discussed in this article are based on nickel and on cobalt. These alloys can be generally categorized as either solid-solution strengthened or precipitation strengthened. The former category is not appreciably strengthened by heat treatment and is considered non-heat-treatable, whereas the latter category is typically used in the age-hardened or heat-treated, condition. It is important to distinguish between these two classes of alloys when considering postweld heat treatment. In general, alloys that contain significant amounts of age-hardening elements, such as aluminum, titanium, and niobium, are heat treatable. The compositions of selected nickel- and cobalt-base alloys in these two categories are shown in Table 1.

TABLE 1 NOMINAL COMPOSITION OF SELECTED NICKEL-BASE AND COBALT-BASE HIGH-TEMPERATURE ALLOYS

ALLOY	**UNS NO.**	COMPOSITION, WT%							
		C	**Cr**	**Ni**	**Co**	**Fe**	**Mo**	**W**	OTHERS
NICKEL-BASE SOLID-SOLUTION ALLOYS									
LNCONEL 600	N06600	0.08 (MAX)	15.5	BAL	. . .	8.0	. . .	. . .	**. . .**
LNCONEL 601	N06601	0.10 (MAX)	23.0	BAL	. . .	14.4	. . .	. . .	**1.4 AL**
INCONEL 617	N06617	0.07	22.0	BAL	12.5	1.5	9.0	. . .	**1.2 AL**
INCONEL 625	N06625	0.10 (MAX)	21.5	BAL	. . .	2.5	9.0	. . .	**3.6 NB**
HAYNES 230	N06230	0.1	22.0	BAL	. . .	3.0 (MAX)	2.0	14.0	**0.02 LA, 0.015 B (MAX)**
HAYNES HR-160	. . .	0.05	28.0	BAL	29.0	1.5	. . .	. . .	**2.75 SI**
HASTELLOY X	N06002	0.1	22.0	BAL	1.5	18.5	9.0	0.6	**. . .**
HASTELLOY W	N10004	0.12 (MAX)	5.0	BAL	2.5	6.0	24.0	. . .	**. . .**
HASTELLOY S	N06635	0.02 (MAX)	15.5	BAL	. . .	3.0 (MAX)	14.5	. . .	**0.05 LA, 0.015 B (MAX)**
RA 333	N06333	0.05	25.0	BAL	3.0	18.0	3.0	3.0	**. . .**
CHROMEL A (OR NICHROME 80)	. . .	. . .	20.0	BAL	. . .	. . .	. . .	. . .	**1.0 SI**
NIMONIC 75	. . .	0.10	19.5	BAL	. . .	. . .	. . .	. . .	**. . .**
NIMONIC 86	. . .	. . .	25.0	BAL	. . .	. . .	10.0	. . .	**0.03 CE**
COBALT-BASE SOLID-SOLUTION ALLOYS									
HAYNES 188	R30188	0.10	22.0	22.0	BAL	3.0 (MAX)	. . .	14.0	**0.04 LA**
HAYNES 25 (L-605)	R30605	0.10	20.0	10.0	BAL	3.0 (MAX)	. . .	15.0	**. . .**
HAYNES 150 (UMCO-50)	. . .	0.06	27.0	. . .	BAL	18.0	. . .	. . .	**. . .**

S-816	R30816	0.38	20.0	20.0	BAL	4.0	4.0	4.0	**4.0 NB**
MAR-M 918	. . .	0.05	20.0	20.0	BAL	. . .	. . .	. . .	**7.5 TA, 0.10 ZR**
MP35-N	R30035	. . .	20.0	35.0	BAL	. . .	10.0	. . .	**. . .**
MP159	. . .	. . .	19.0	25.5	BAL	9.0	7.0	. . .	**0.6 NB, 0.2 AL, 3.0 TI**
NICKEL-BASE PRECIPITATION-HARDENING ALLOYS									
LNCONEL 706	N09706	0.03	16.0	BAL	. . .	37.0	. . .	. . .	**1.8 TI, 0.2 AL, 2.9 NB**
INCONEL 718	N07718	0.04	18.0	BAL	. . .	18.5	3.0	. . .	**5.1 NB, 0.9 TI, 0.5 AL**
LNCONEL X-750	N07750	0.04	15.5	BAL	. . .	7.0	. . .	. . .	**2.5 TI, 0.7 AL, 1.0 NB**
LNCONEL 751	N07751	0.05	15.0	BAL	. . .	7.0	. . .	. . .	**2.5 TI, 1.1 AL, 1.0 NB**
NIMONIC 80A	N07080	0.06	19.5	BAL	. . .	. . .	. . .	. . .	**1.4 AL, 2.4 TI**
NIMONIC 90	N07090	0.07	19.5	BAL	16.5	. . .	. . .	. . .	**1.5 AL, 2.5 TI**
NIMONIC 91	. . .	. . .	28.5	BAL	20.0	. . .	. . .	. . .	**1.2 AL, 2.3 TI**
NIMONIC 105	. . .	0.08	15.0	BAL	20.0	. . .	5.0	. . .	**4.7 AL, 1.3 TI, 0.005 B**
NIMONIC 115	. . .	0.15	15.0	BAL	15.0	. . .	4.0	. . .	**5.0 AL, 4.0 TI**
RENÉ 41	N07041	0.09	19.0	BAL	11.0	5.0 (MAX)	10.0	. . .	**1.5 AL, 3.0 TI, 0.006 B**
RENÉ 95	. . .	0.15	14.0	BAL	8.0	. . .	3.5	3.5	**3.5 NB, 3.5 AL, 2.5 TI, 0.05 ZR**
RENÉ 100	. . .	0.16	9.5	BAL	15.0	. . .	3.0	. . .	**5.5 AL, 4.2 TI, 0.06 ZR, 0.015 B**
UDIMET 400	. . .	0.06	17.5	BAL	14.0	. . .	4.0	. . .	**0.5 NB, 1.5 AL, 2.5 TI, 0.06 ZR, 0.008 B**
UDIMET 500	. . .	0.08	18.0	BAL	18.5	. . .	4.0	. . .	**2.9 AL, 2.9 TI, 0.05 ZR, 0.006 B**
UDIMET 520	. . .	0.05	19.0	BAL	12.0	. . .	6.0	1.0	**2.0 AL, 3.0 TI, 0.005 B**
UDIMET 630	. . .	0.03	18.0	BAL	. . .	18.0	3.0	3.0	**6.5 NB, 0.5 AL, 1.0 TI**
UDIMET 700	. . .	0.03	15.0	BAL	18.5	. . .	5.2	. . .	**5.3 AL, 3.5 TI, 0.03 B**
UDIMET 710	. . .	0.07	18.0	BAL	15.0	. . .	3.0	1.5	**2.5 AL, 5.0 TI**
UDIMET 720	. . .	0.03	17.9	BAL	14.7	. . .	3.0	1.3	**2.5 AL, 5.0 TI, 0.03 ZR, 0.033 B**
WASPALOY ALLOY	N07001	0.08	19.0	BAL	14.0	. . .	4.3	. . .	**1.5 AL, 3.0 TI, 0.05 ZR, 0.006 B**
263	. . .	0.06	20.0	BAL	20.0	. . .	5.8	. . .	**0.5 AL, 2.2 TI**
UNITEMP AF2-1DA	. . .	0.35	12.0	BAL	10.0	. . .	3.0	6.0	**1.5 TA, 4.6 AL, 3.5 TI, 0.10 ZR**
UNITEMP AF2-1DA6	. . .	0.04	12.0	BAL	10.0	. . .	2.7	6.5	**1.5 TA, 4.0 AL, 2.8 TI, 0.1 ZR, 0.015 B**
ASTROLOY	. . .	0.06	15.0	BAL	17.0	. . .	5.3	. . .	**4.0 AL, 3.5 TI, 0.03 B**
D-979	N09979	0.05	15.0	BAL	. . .	27.0	4.0	4.0	**1.0 AL, 3.0 TI**
LNCONEL 100	N13100	0.15	10.0	BAL	15.0	. . .	3.0	. . .	**5.5 AL, 4.7 TI,**

									0.06 ZR, 1.0 V, 0.015 B
INCONEL 102	N06102	0.06	15.0	BAL	. . .	7.0	3.0	3.0	**3.0 NB, 0.4 AL, 0.6 TI, 0.02 MG, 0.03 ZR**
M252	N07252	0.15	20.0	BAL	10.0	. . .	10.0	. . .	**1.0 AL, 2.6 TI, 0.005 B**
PYROMET 31	N07031	0.04	22.5	BAL	. . .	15.0	2.0	. . .	**1.4 AL, 2.3 TI, 0.9 CU, 0.005 B**
PYROMET 860	. . .	0.05	13.0	BAL	4.0	28.9	6.0	. . .	**1.0 AL, 3.0 TI, 0.01 B**
REFRACTORY 26	. . .	0.03	18.0	BAL	20.0	16.0	3.2	. . .	**0.2 AL, 2.6 TI, 0.015 B**
ALLOY 625	N07716	0.02	20.0	BAL	. . .	5.0	9.0	. . .	**3.1 NB, 0.2 AL, 1.3 TI**
INCONEL 100 GATORIZED	**. . .**	**0.07**	**12.4**	**BAL**	**18.5**	**. . .**	**3.2**	**. . .**	**5.0 AL, 4.3 TI, 0.06 ZR, 0.02 B, 0.8 V**

Postweld Heat Treatment of Nonferrous High-Temperature Materials

Steven C. Ernst, Haynes International, Inc.

Types of Postweld Heat Treatment

Stress relieving is a process used to reduce residual welding stresses prior to further fabrication or service. An important result of a good stress-relief heat treatment is dimensional stability of a weldment during high-temperature service. Typical stress-relieving temperature ranges are less than those required for annealing or recrystallization. Accordingly, stress relieving can result in deleterious effects to the microstructure and corresponding mechanical properties. When considering a postweld stress-relief heat treatment, it is important to recognize these effects so that a suitable compromise can be made between dimensional stability and mechanical properties.

Annealing is a process performed above the recrystallization temperature to promote maximum softness in an alloy. This heat treatment is also effective at relieving residual welding stresses. Annealing should be conducted after welding age-hardenable alloys in highly restrained conditions. This annealing requires high heating rates to reduce the probability of strainage cracking.

Solution annealing is a high-temperature anneal carried out above the secondary carbide solvus. Because that temperature is significantly higher than the recrystallization temperature, significant grain coarsening can occur. The carbide dissolution and coarse grain structure can be beneficial to high-temperature rupture life.

Solution treating and quenching is a process used with precipitation-strengthened alloys to dissolve the principle aging phases and carbides formed during welding and to produce a supersaturated solid solution at room temperature. This heat treatment will put the weldment in optimum metallurgical condition for subsequent age hardening.

Aging treatments are performed in the intermediate temperature range, allowing precipitation of the strengthening phase(s) from the supersaturated solid solution. The aging response may not necessarily be uniform in all regions of a weldment because of segregation effects that occur during weld solidification.

Postweld Heat Treatment of Nonferrous High-Temperature Materials

Steven C. Ernst, Haynes International, Inc.

Postweld Heat Treatment of Solid-Solution-Strengthened Alloys

Solid-solution-strengthened alloys normally do not require postweld heat treatment to achieve or restore optimum mechanical properties. Postweld heat treatment is sometimes used to provide stress relief. Complete stress relief can be achieved with full solution anneal, which also gives the alloy an optimal metallurgical condition. Solution annealing will eliminate any prior cold work by recrystallization and will dissolve secondary $M_{23}C_6$-type carbides that may have formed upon cooling during welding. Typical solution annealing temperatures for solid-solution-strengthened alloys are given in Table 2. The time required can range from a few minutes to about an hour, followed by rapid quenching in water or air. A popular calculated time is 140 s/mm (1 h/in.) of cross-section thickness.

TABLE 2 TYPICAL SOLUTION ANNEALING AND MILL ANNEALING TEMPERATURES FOR SOLID-SOLUTION-STRENGTHENED ALLOYS

ALLOY	**SOLUTION ANNEALING TEMPERATURE RANGE**		MINIMUM TEMPERATURE FOR MILL ANNEALING	
	°C	**°F**	**°C**	°F
HASTELLOY X	1165-1190	2125-2175	1010	**1850**
HASTELLOY S	1050-1135	1925-2075	955	**1750**
ALLOY 625	1095-1205	2000-2200	925	**1700**
RA 333	1175-1205	2150-2200	1040	**1900**
INCONEL 617	1165-1190	2125-2175	1040	**1900**
HAYNES 230	1165-1245	2125-2275	1120	**2050**
HAYNES 188	1165-1190	2125-2175	1120	**2050**
HAYNES 25 (L-605)	**1175-1230**	**2150-2250**	**1040**	**1900**

In cases where stress relief is needed and a fine grain size is required, mill annealing is effective. The mill annealing heat treatment is similar to solution annealing. However, the lower temperatures involved are not sufficient to dissolve secondary carbides that may have precipitated intergranularly during welding. The mill annealing treatment may not develop the stress rupture properties in the weldment to their full potential. Examples of minimum mill annealing temperatures are given in Table 3.

TABLE 3 TYPICAL SOLUTION TREATMENT AND AGING CYCLES FOR PRECIPITATION-STRENGTHENED HIGH-TEMPERATURE ALLOYS

ALLOY	**SOLUTION TREATMENT**				AGING			
	TEMPERATURE		**TIME, H**	**COOLING PROCEDURE**	**TEMPERATURE**		**TIME, H**	COOLING PROCEDURE
	°C	**°F**			**°C**	**°F**		
INCONEL 718	980	1800	1	AIR COOL	720	1325	8	**FURNACE COOL**
					620	1150	8	**AIR COOL**
INCONEL X-750	1150	2100	2	AIR COOL	845	1550	24	**AIR COOL**
					705	1300	20	**AIR COOL**
NIMONIC 80A	1080	1975	8	AIR COOL	705	1300	16	**AIR COOL**
NIMONIC 90	1080	1975	8	AIR COOL	705	1300	16	**AIR COOL**

RENÉ 41	1065	1950	$\frac{1}{2}$	AIR COOL	760	1400	16	**AIR COOL**
UDIMET 500	1080	1975	4	AIR COOL	845	1550	24	**AIR COOL**
					760	1400	16	**AIR COOL**
WASPALOY	1080	1975	4	AIR COOL	845	1550	24	**AIR COOL**
					760	**1400**	**16**	**AIR COOL**

In terms of optimizing metallurgical structure, mechanical properties, and stress relief, annealing treatments are the best postweld heat treatments for solid-solution-strengthened alloys. When large or complex structures are involved, the high annealing temperatures can result in distortion problems during the heat treatment. In many cases, the best decision may be to completely avoid postweld heat treatment. If stress relief is mandatory, but the high annealing temperatures are not tolerable, a lower temperature may be used at the expense of the metallurgical condition. Some fabricators conduct stress relieving at a temperature about 100 °C (180 °F) above the intended service temperature to avoid distortion during service. A partial stress relief can be accomplished for many alloys in the range from 450 to 900 °C (840 to 1650 °F), concomitant with embrittlement caused by extensive precipitation of carbides and topologically close-packed (TCP) phases. Additionally, the weldment may display poor creep-rupture characteristics in service following a stress relief heat treatment in the 450 to 900 °C (840 to 1650 °F) temperature range. The magnitude of stress relief and property degradation vary for each alloy and stress-relief temperature. These factors should be carefully considered if stress relief is to be performed below the minimum mill annealing temperature.

Postweld Heat Treatment of Nonferrous High-Temperature Materials

Steven C. Ernst, Haynes International, Inc.

Postweld Heat Treatment of Precipitation-Strengthened Alloys

The precipitation-strengthened alloys are typically used in the solution-treated and aged condition. Solution treatment can be performed either above or below the γ' solvus temperature, depending on the desired microstructure. In wrought material, aging results in a homogeneous distribution of γ' and intergranular, secondary-type carbides. The γ' distribution in weld metal may or may not be homogeneous because of the segregational effects of weld solidification. In this situation, the weld metal may not strengthen to the same degree as the base metal. A two-step aging treatment that involves a stabilization followed by a lower-temperature aging treatment can be used in some cases. The stabilization heat treatment involves the formation of $M_{23}C_6$ carbides plus γ', which forms from a reaction between metal carbides (MC) and the matrix. Examples of typical solution treatment and aging cycles are shown in Table 3. For further information, see the article "Heat Treating of Superalloys" in *Heat Treating,* Volume 4 of the *ASM Handbook.*

The major problem that occurs when postweld heat treating the precipitation-hardenable alloys is strain-age cracking. The susceptibility of an alloy to strain-age cracking tends to increase with an increasing aluminum and titanium content. Alloys that contain niobium as a hardening element exhibit a sluggish aging response and are therefore less sensitive to strain-age cracking. Strain-age cracking is more prevalent in highly restrained weldments or in the presence of high residual stresses. Cracking is caused by the combination of stress, precipitation strengthening, and the volumetric contraction associated with γ' precipitation. Guidelines to follow to avoid strain-age cracking include:

- MINIMIZE RESIDUAL AND THERMALLY INDUCED STRESSES BY USING THE APPROPRIATE JOINT DESIGN, WELDING PROCESS, AND FILLER METAL.
- HEAT AS RAPIDLY AS POSSIBLE THROUGH THE REGION OF AGING TEMPERATURES WHEN STRESS-RELIEF ANNEALING.
- MINIMIZE HEAT INPUT DURING WELDING TO AVOID PARTIAL MELTING OF GRAIN BOUNDARIES ADJACENT TO THE FUSION LINE.
- HEAT TREAT IN AN INERT ATMOSPHERE IF POSSIBLE; THE PRESENCE OF OXYGEN TENDS TO INCREASE INTERGRANULAR EMBRITTLEMENT, WHICH CAN RESULT IN CRACKING.
- TO AVOID POSTWELD STRAIN-AGE CRACKING, ALWAYS STRESS-RELIEF ANNEAL BEFORE AGING. FOR HEAVY-SECTION WELDMENTS IN ALUMINUM-TITANIUM AGED

MATERIALS, MULTIPLE STRESS-RELIEF ANNEALS MAY BE NECESSARY DURING WELDING.
- AVOID WELDING ON MATERIAL WITH ROUGH OR POOR SURFACE CONDITIONS.

Special Metallurgical Welding Considerations for Nickel and Cobalt Alloys and Superalloys

S.D. Kiser, Inco Alloys International, Inc.

Introduction

THE DESIGN ENGINEER who wishes to use a welded structure for demanding service faces a challenging dilemma. On the one hand, the materials involved and the deposited weld metal must exhibit sufficient ductility to withstand the severe thermal cycle imposed by fusion welding. On the other hand, many applications demand that specimens taken from a qualification welded assembly be capable of passing 2*t* 180° side-bend testing (where *t* is the material thickness). This test requires 20% elongation of the outer fibers of the bend specimen. After welding, some applications for which the weldment is designed may demand different properties.

Most solid-solution (nonprecipitation-hardening) nonferrous alloys exhibit sufficient ductility to meet the above requirements. The weld fabrication of these materials is straightforward in that they usually do not require preheat or postheat, and interpass temperature control during welding normally is not critical.

The precipitation-hardenable materials are different from the above alloys in that they generate a second phase when exposed to temperatures for specified times in a particular range. They further distinguish themselves by exhibiting superior mechanical properties after being precipitation treated (aged). These materials are usually welded in the annealed (or solution-annealed) condition and are heat treated to precipitate the second phase as a final or near-final production step.

Both the solid-solution and the precipitation-hardening groups of alloys are typically austenitic (that is, the atoms of their basic structures fit together in a face-centered-cubic arrangement, often referred to as γphase). The precipitation-hardening alloys form a second phase, γ', as a result of controlled heat treatment. The development and dispersion of this second phase are made possible by additions of aluminum, titanium, and niobium. The γ' phase imparts great stiffness to these materials at room temperature and up to the upper limit of their aging-temperature ranges. When used at service temperatures above their aging ranges, these materials lose strength rapidly as a result of the agglomeration of their second phase. This condition, commonly referred to as overaging, results in relatively large γ' particles.

A subset of the precipitation-hardenable high-temperature materials is the secondary-carbide-strengthened materials. These alloys have an austenitic structure like the above groups, but do not contain sufficient alloy content to precipitate γ' phase. Instead, they contain generous additions of carbide-forming elements such as chromium, niobium (also forms γ''), and a supersaturating amount of carbon. These materials are primarily used as centrifugally cast tubes. They enjoy reasonably good weldability in the as-cast condition, but usually require special precautions for welding after they have been in service.

A final group of exceptionally strong high-temperature metals is the mechanically alloyed materials. These metals combine the high strength of the precipitation-hardened alloys by use of γ' strengthening up to about 705 °C (1300 °F) with dispersion strengthening with yttrium oxide (Y_2O_3) up to 1315 °C (2400 °F). These materials are produced by a combination of the mechanical forging of powder particles in a ball mill, followed by hot working and a 1315 °C (2400 °F) thermal treatment (Fig. 1). Joining these materials requires special precautions.

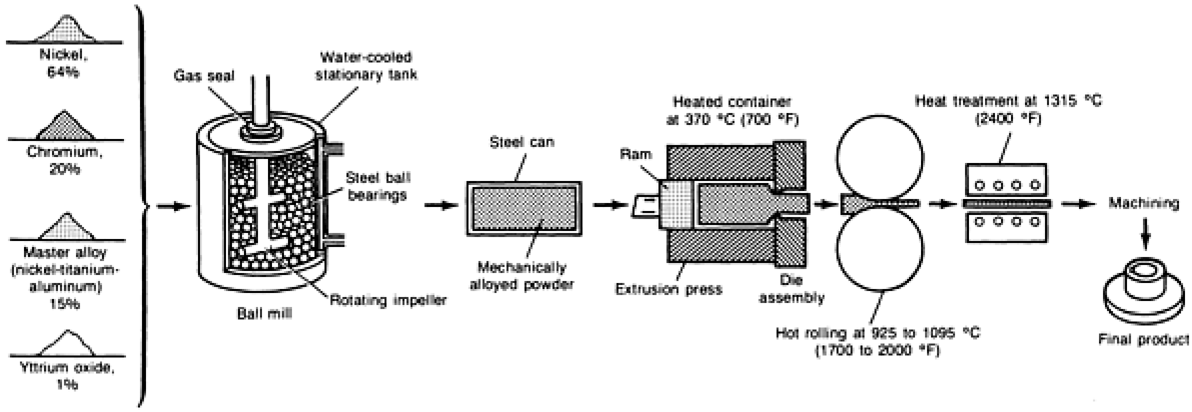
Nickel,
64%
Chromium,
20%
Master alloy
(nickel-titanium-
aluminum)
15%
Yttrium oxide,
1%
Gas seal
Water-cooled
stationary tank
Steel ball
bearings
Rotating impeller
Ball mill
Steel can
Mechanically
alloyed powder
Heated container
at 370 °C (700 °F)
Ram
Extrusion press
Die
assembly
Hot rolling at 925 to 1095 °C
(1700 to 2000 °F)
Heat treatment at 1315 °C
(2400 °F)
Machining
Final product

FIG. 1 SCHEMATIC SHOWING TYPICAL PROCESS OPERATIONS USED IN THE PRODUCTION OF MECHANICALLY ALLOYED OXIDE-DISPERSION-STRENGTHENED PRODUCTS

In many demanding applications, it is necessary to design components made of different materials. The process of making such assemblies often requires the welding of dissimilar metals, the welding of diffusion-bonded materials, and sometimes weld overlay cladding and even thermal spraying. These types of unusual welding and spraying applications require special knowledge and treatments that may have been developed specifically for each material.

Finally, in the service life of nearly every welded structure, there arises a need for alteration or repair. In many cases it becomes necessary to splice new materials into old. Most of the above materials require special conditioning before repair welding.

Special Metallurgical Welding Considerations for Nickel and Cobalt Alloys and Superalloys

S.D. Kiser, Inco Alloys International, Inc.

Solid-Solution Nickel Alloys

These materials have good weldability and are often used in the as-welded condition. The alloys are usually formed by additions to nickel of chromium, cobalt, molybdenum, iron, and sometimes small amounts of aluminum, silicon, and niobium (Table 1). While they do not develop abnormally high strength levels like the other groups, these alloys provide attractive creep resistance in conjunction with good thermal-fatigue resistance. Under moderate loading, alloys of this group can be used up to temperatures approaching 1150 °C (2100 °F). Typical applications are welded containers and fixtures for thermal processing, and headers and manifolds for chemical and petrochemical processing.

TABLE 1 COMPOSITION OF SELECTED SOLID-SOLUTION NICKEL ALLOYS FOR HIGH-TEMPERATURE APPLICATIONS UP TO 1150 °C (2100 °F)

ALLOY	UNS NO.	COMPOSITIONS, WT%									
		Ni	**Cr**	**Fe**	**Cu**	**Co**	**Mo**	**Ti**	**Al**	**Nb**	OTHERS
600	N06600	76.0	15.5	8.0	. . .	. . .	. . .	. . .	. . .	. . .	**. . .**
601	N06601	60.5	23.0	14.4	. . .	. . .	. . .	. . .	1.4	. . .	**. . .**
617	N06617	52.0	22.0	1.5	. . .	12.5	9.0	. . .	1.2	. . .	**. . .**
625	N06625	61.0	21.5	2.5	. . .	. . .	9.0	. . .	. . .	3.6	**. . .**
690	N06690	61.5	29.0	9.0	. . .	. . .	. . .	. . .	. . .	. . .	**. . .**
HX	. . .	47.5	21.8	18.5	. . .	1.5	9.0	. . .	. . .	. . .	**0.6 W**
800	N08800	32.5	21.0	46.0	. . .	. . .	. . .	. . .	. . .	. . .	**0.05 C**

800H	N08810	32.5	21.0	46.0	. . .	. . .	. . .	. . .	. . .	. . .	**0.08 C**
800HT	N08811	32.5	21.0	46.0	. . .	. . .	. . .	. . .	. . .	. . .	**0.08 C; 1.0 (AL + TI)**
802	N08802	32.5	21.0	46.0	. . .	. . .	. . .	. . .	. . .	. . .	**0.4 C**
DS	. . .	37.0	18.0	41.0	. . .	. . .	. . .	. . .	. . .	. . .	**1.0 MN; 2.3 SI**
330	N08330	35.5	18.5	44.0	. . .	. . .	. . .	. . .	. . .	. . .	**1.1 SI**
330HC	**. . .**	**31.5**	**18.5**	**48.0**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**1.2 SI; 0.4 C**

The potential problem in welding these materials is grain size. As grain size increases, weldability decreases. In some applications, a minimum grain size is specified for maximum creep resistance. The compromise normally implemented is to limit welding processes to those methods that use lower heat input (Ref 1).

Reference cited in this section

1. *WELDING--IAI-14,* INCO ALLOYS INTERNATIONAL, JULY 1991, TABLE 18, P 23

Special Metallurgical Welding Considerations for Nickel and Cobalt Alloys and Superalloys

S.D. Kiser, Inco Alloys International, Inc.

Precipitation-Hardenable Nickel Alloys

These materials are characterized by their distinctively high strength at room temperature through about 705 °C (1300 °F). They range in alloy content from the virtually pure nickel-aluminum Alloy 301 to nickel-copper Alloy K-500 through the nickel-chromium-iron Alloy 718 through the low-expansion nickel-iron Alloy 909. They enjoy the commonality of γ' strengthening by heat treatment because of additions of aluminum, titanium, and niobium. Table 2 lists selected precipitation-hardening alloys and their chemistries.

TABLE 2 COMPOSITION OF SELECTED PRECIPITATION-HARDENABLE NICKEL ALLOYS THAT RETAIN HIGH-STRENGTH PROPERTIES TO 705 °C (1300 °F)

ALLOY	**UNS NO.**	COMPOSITIONS, WT%									
		Ni	**Cr**	**Fe**	**Cu**	**Co**	**Mo**	**Ti**	**Al**	**Nb**	OTHERS
K-500	N05500	65.5	. . .	1.0	29.5	. . .	. . .	0.6	2.7	. . .	**. . .**
706	N07702	41.5	16.0	37.0	. . .	. . .	. . .	1.8	0.2	2.9	**. . .**
718	N07718	54.0	18.0	18.5	. . .	. . .	3.0	. . .	. . .	5.1	**. . .**
X-750	N07750	73.0	15.5	7.0	. . .	. . .	. . .	2.5	0.7	1.0	**. . .**
751	N07751	73.0	15.0	7.0	. . .	. . .	. . .	2.5	1.1	1.0	**. . .**
903	N19903	38.0	. . .	42.0	. . .	15.0	. . .	1.4	0.9	3.0	**. . .**
904	. . .	32.5	. . .	51.0	. . .	14.5	. . .	1.6	. . .	. . .	**. . .**
907	. . .	38.0	. . .	42.0	. . .	13.0	. . .	1.5	0.03	4.7	**0.15 SI**
909	. . .	38.0	. . .	42.0	. . .	13.0	. . .	1.5	0.03	4.7	**0.4 SI**
925	**. . .**	**44.0**	**21.0**	**28.0**	**1.8**	**. . .**	**3.0**	**2.1**	**0.3**	**. . .**	**0.01 C**

Most of these alloys have good weldability in the annealed condition. Because of this, most are formed, machined, and welded in the annealed condition. They are then reannealed after welding, and then aged, to obtain the desired properties.

Postweld Strain-Age Cracking (PWSAC). Because all these types of materials are at least somewhat prone to PWSAC, little welding is performed in the aged condition if the weldment is to be re-heat treated or put in elevated-temperature service. Postweld strain-age cracking occurs in precipitation-hardenable alloys when cold-work stresses, weld-induced residual stresses, and the stress imparted by aging exceed the yield strength (and available ductility) of the material. The result is an instantaneous, catastrophic failure by cracking (Fig. 2).

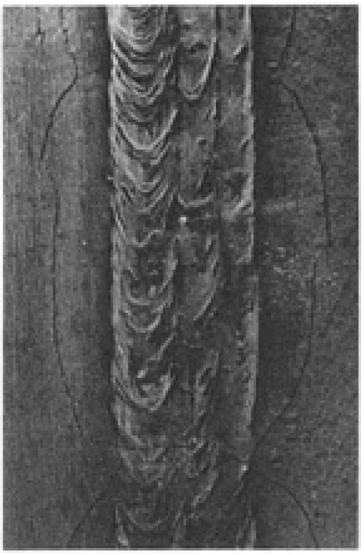

FIG. 2 PWSAC FAILURE IN ALLOY X-750 INDUCED BY RESIDUAL WELDING STRESS AT A 50 MM (2 IN.) THICK JOINT. ALLOY WAS WELDED IN THE AGE-HARDENED CONDITION AND RE-AGED AT 705 °C (1300 °F) FOR 20 H AFTER WELDING.

Those alloys whose γ' phase is formed predominantly by nickel-aluminum-titanium age more rapidly than those using niobium. As a result, these alloy are more prone to PWSAC.

Alloy 718. In an effort to improve the weldability of precipitation-hardenable alloys, niobium was substituted for large amounts of aluminum and titanium. The result was nickel-chromium-iron Alloy 718, which showed greatly improved resistance to PWSAC. With Alloy 718, thicknesses up to 3.2 mm ($\frac{1}{8}$ in.) could be welded in the aged condition and directly re-aged without cracking. This combination of very high mechanical properties and fabricability has earned Alloy 718 a prominent spot in the aerospace industry.

Alloy 713C. Some precipitation-hardenable alloys often require welding in the aged condition. Alloy 713C, for example, is a nickel-chromium alloy with very high levels of aluminum and titanium that is used for gas-turbine vanes. It is normally produced as a casting that precipitation hardens during cool down. It is considered nonweldable by fusion processes. However, experience has shown that blade and vane tips can be restored by fusion welding with a nonhardenable filler metal (ERNiCrMo-3) without cracking (Ref 2). Thus, crack-free welds can sometimes be accomplished in very-difficult-to-weld materials where residual shrinkage stresses can be kept low and cooling rates are slow. Both conditions occur naturally when welding on an edge.

Alloy 301. Preliminary work has shown similar results when welding on nickel-aluminum Alloy 301 in the aged condition. This alloy exhibits high mechanical properties and very good thermal conductivity. It finds use in cooled glass molds and plastic extrusion equipment. It can be readily welded in the aged condition--fully restrained--when the recommended ductile ERNi-1 filler metal is added. Also, autogenous, low-stress welds (on an edge) can be made at high travel speeds without cracking. However, when fully restrained, autogenous welds are made in forging stock at 200 to 250 mm/min (8 to 10 in./min) travel speed, the cooling rate is so rapid that the aluminum-rich weld pool contains numerous large solidification cracks. Decreasing the travel speed to decrease the cooling rate allows crack-free autogenous welds to be made.

Low-Coefficient-of-Expansion Alloys. For the next group of precipitation-hardening alloys, the same guidelines apply (that is, to weld in the annealed condition). In this group of alloys, the chromium has been removed in order to maintain a low coefficient of thermal expansion. The fabrication sequence includes welding, annealing, heat treating, and in some cases coating. The coating applied is usually an oxidation-resisting ceramic coating.

Alloy 909. The most prominent member of this group is nickel-iron Alloy 909. It is primarily used for seals and rings in aircraft turbines. Its low coefficient of expansion allows close-tolerance designs, which increase the fuel efficiency of gas-turbine engines.

Mechanically Alloyed Products. The final group of precipitation-hardening alloys to be discussed merits the name superalloys. Mechanically alloyed products have a remarkable advantage over conventional materials in elevated-temperature stress-rupture strength. This advantage is gained by the combination of precipitation hardening and dispersion strengthening. Neither concept is revolutionary, but the combination of the two using the ball mill is. As shown in Fig. 1, various ingredients are added to the ball mill as metallic and intermetallic powders. Extended processing times of approximately 10 h in the mill cause the particles to become metallurgically homogeneous. This process is similar to the step of incorporation in Du Pont's early approach to making homogenous particles of gunpowder, each containing the correct ratio of charcoal, sulfur, and potassium nitrate.

Homogenization in the mill is followed by either hot rolling or extrusion. The form is then thermally treated at 1315 °C (2400 °F) to produce the exceptional strength characteristics. Tables 3 and 4 list the chemical compositions and selected properties for mechanically alloyed products.

TABLE 3 COMPOSITION OF SELECTED MECHANICALLY ALLOYED MATERIALS

ALLOY	COMPOSITIONS, WT%							
	Fe	**Cr**	**Al**	**Ti**	**Ni**	**C**	**Y_2O_3**	MO
MA 754	1.0	20.0	0.3	0.5	BAL	0.05	0.6	**. . .**
MA 956	BAL	20.0	4.5	0.5	. . .	. . .	0.5	**. . .**
MA 957	BAL	14.0	. . .	1.0	. . .	. . .	0.25	**0.3**
MA 6000	**. . .**	**15.0**	**4.5**	**2.5**	**BAL**	**0.5**	**0.6**	**. . .**

TABLE 4 PROPERTIES OF SELECTED MECHANICALLY ALLOYED MATERIALS

ALLOY	**DENSITY**		**YOUNG'S MODULUS**		**TENSILE STRENGTH(A)**		**YIELD STRENGTH(A)**		**ELONGATION AT FRACTURE, %(A)**	STRESS TO PRODUCE RUPTURE AT 1095 °C (2000 °F) FOR INDICATED ELAPSED TIME			
										100 H		1000 H	
	G/CM^3	**LB/IN.^{3}**	**GPA**	**10^6 PSI**	**MPA**	**KSI**	**MPA**	**KSI**		**MPA**	**KSI**	**MPA**	KSI
MA 754(B)	8.3	0.30	149	21.6	148	21.5	134	19.5	12.5	102	14.8	94	**13.6**
MA 956(B)	7.2	0.26	269	39.0	91	13.2	85	12.3	3.5	57	8.3	51	**7.4**
MA 957(C)	7.67	0.277	210(D)	30.0(D)	1150	167	1100	160	17	345	50	. . .	**. . .**
MA 6000(B)	**8.11**	**0.293**	**203**	**29.4**	**222**	**32.2**	**192**	**27.8**	**9**	**131**	**19**	**127**	**18.5**

(A) LONGITUDINAL.

(B) TEST TEMPERATURE, 1095 °C (2000 °F).

(C) TEST TEMPERATURE, 600 °C (1100 °F), WHICH IS THE INTENDED APPLICATION TEMPERATURE FOR ALLOY IN NUCLEAR REACTORS.

(D) AT 20 °C (70 °F)

Figure 3 shows the dramatic advantage in strength of mechanically alloyed products over the conventional dispersion-strengthened and cast nickel-base alloys. Unfortunately, making these materials weldable is difficult. While they can undergo diffusion bonding or brazing with little loss in properties, fusion welding produces a softened heat-affected zone (HAZ). This behavior is caused by the agglomeration of the dispersoid Y_2O_3. In addition, it is not possible to produce a weld that has properties similar to those of the alloys. Thus, for most applications, the mechanically alloyed products are joined mechanically, or diffusion bonded, or brazed.

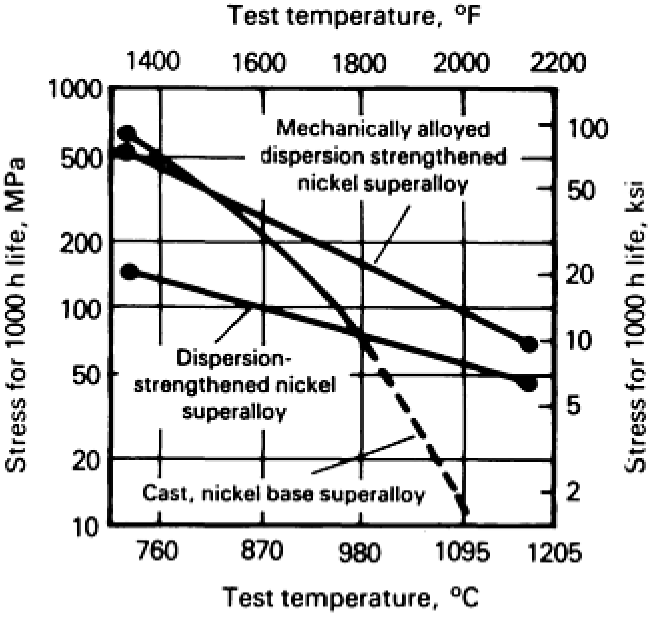

FIG. 3 STRESS-RUPTURE STRENGTH OF MECHANICALLY ALLOYED DISPERSION-STRENGTHENED NICKEL SUPERALLOY RELATIVE TO CONVENTIONAL DISPERSION-STRENGTHENED NICKEL SUPERALLOYS AND CAST NICKEL-BASE ALLOYS

Reference cited in this section

2. D.S. DUVALL AND J.R. DOYLE, *REPAIR OF TURBINE BLADES AND VANES,* PUBLICATION 73-GT-44, AMERICAN SOCIETY OF MECHANICAL ENGINEERS, 1973

Special Metallurgical Welding Considerations for Nickel and Cobalt Alloys and Superalloys

S.D. Kiser, Inco Alloys International, Inc.

Creep-Resistant Secondary Carbide-Strengthened Alloys Welded with Nickel Alloys

This group of alloys has become extremely popular in the chemical and petrochemical industries for the processing of hydrocarbon compounds. Such alloys are typically fabricated in the as-cast condition after being produced in a rotating, water-cooled mold. This production method, which incorporates rapid cooling, allows a supersaturated solution of carbon (0.40 to 0.45%) in nickel, chromium, and iron to retain the carbon in solution. This condition allows sufficient ductility in the as-cast condition to provide good weldability by most fusion processes.

Formation of Secondary Carbides. Once the assembly has been fabricated, it is normally put into service at temperatures of 760 to 1095 °C (1400 to 2000 °F). At these temperatures, the chromium (and sometimes niobium) combines with the carbon in a solid-state diffusion process that results in the precipitation of secondary carbides. These secondary carbides provide a tremendous strengthening effect for the intended high-temperature service. Figure 4 compares the stress-rupture strength of a typical centrifugal-cast carbide-strengthened alloy with that of a widely used nickel-alloy welding product that is needed to attain approximately equal strength. It is interesting to note that welds made in these alloys with near-matching composition welding products suffer failure at less than half the rupture life of welds made with ERNiCrCoMo-1 type welding products (Ref 3). (See Table 1 for the approximate chemical composition of ERNi-CrCoMo-1, identified as N06617.) In addition, transverse stress-rupture specimens welded with near-matching composition products routinely fail in the weld itself, while those welded with ERNiCrCoMo-1 products typically fail in the alloy casting.

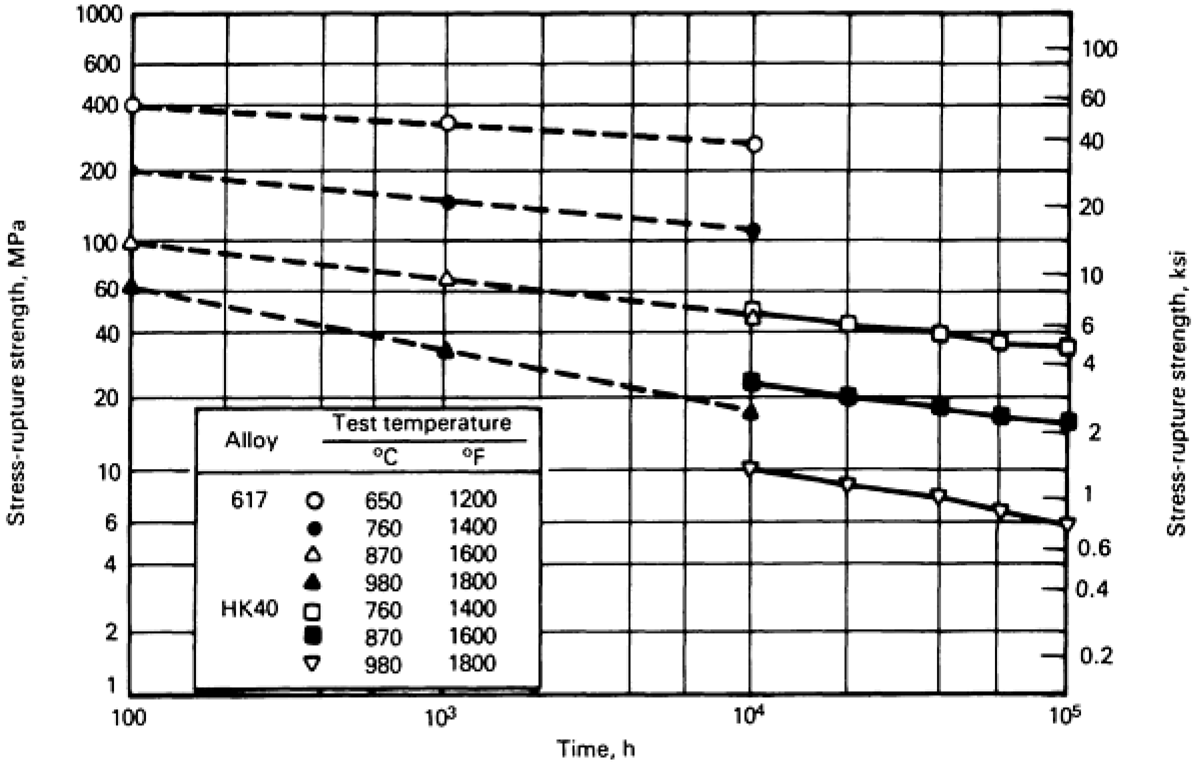

FIG. 4 STRESS-RUPTURE STRENGTH OF ALLOY 617 FILLER METAL RELATIVE TO CENTRIFUGAL-CAST CARBIDE-STRENGTHENED ALLOY HK40

Effect on Weldability. As might be expected, once these alloys have precipitated their secondary-phase carbides, their weldability decreases. Like the precipitation-hardenable, γ'-strengthened alloys, these materials undergo a drastic reduction in room-temperature ductility as a result of carbide precipitation. This result usually means that a secondary carbide-strengthened alloy cannot be welded by fusion processes immediately after service. It is customary to use a resolutioning heat treatment for used tubes in preparation for welding. Parts are exposed to 1095 to 1205 °C (2000 to 2200 °F) for several hours, usually followed by a forced-air cool (Ref 4). This practice restores ductility to near asproduced levels, and allows successful fusion welding without cracking of the base metal.

Castings. Most of the above-mentioned carbide-strengthened castings have good weldability in the proper thermally treated condition. Likewise, most of the solid-solution and precipitation-hardened nonferrous materials in previous sections can be produced as castings. Further, the weldability of these castings may be good if they are produced using clean foundry conditions and clean charge (raw) materials. However, castings are usually made to near-finished dimensions and do not require further hot working as wrought materials do. As a result, many castings do not have the good hot ductility and malleability of wrought materials, characteristics that contribute greatly to good weldability. Thus, a casting may be produced and delivered as a finished part if it simply solidifies from the liquid state without cracking.

In many cases, the customer may attempt to weld fabricate the casting into a larger design only to encounter HAZ cracking in the casting. Analysis usually discloses objectionable levels of low-melting-temperature elements as contaminants. Once the casting has been delivered, and welding difficulties discovered, options to produce good weldments are limited. Best results are obtained by using very low heat input, some-times coupled with buttering or overlaying joint surfaces before welding them together.

An alternate approach that generally yields successful results is for the buyer to specify in purchase documents that the casting must be weldable. To avoid sharply increased prices in quotations for castings, the buyer should not tightly specify chemical limits for damaging contaminating elements. Rather, a simple fusion weldability test can be written into the purchase specifications. Such a test is usually not seen as rigorously expensive to the foundry, and it ensures that the fabricator will get a weldable casting. The weldability test is usually a gas tungsten-arc weld made with no filler metal added on a ground surface of the casting (or testpiece). Amperage and travel speed are usually specified, along with length of weld. A testing method, usually liquid-penetrant inspection, is employed to detect weld and HAZ discontinuities. Generally, the start and stop are excluded from inspection, and cause for rejection is any linear indication.

References cited in this section

3. C.W. COX AND S.D. KISER, FUSION WELDING OF DISSIMILAR METALS FOR HIGH TEMPERATURE STRENGTH, *WELD. J.*, VOL 5, MAY 1991, FIGURE 1, P 68
4. R.E. AVERY AND C.M. SCHILLMOLLER, REPAIR WELDING HIGH ALLOY FURNACE TUBES,*HYDROCARBON PROCESS.*, JAN 1988

Special Metallurgical Welding Considerations for Nickel and Cobalt Alloys and Superalloys

S.D. Kiser, Inco Alloys International, Inc.

Special Welded Product Conditions

Dissimilar Metals. The welding of dissimilar metals has been amply discussed (Ref 5, 6, 7). The materials discussed in this article have reasonably good dissimilar weldability provided the dissimilar alloy is not too different in liquidus and solidus temperatures and other characteristics important to sound weldments. For instance, nearly all the materials previously mentioned can be readily joined to steel, high-strength low-alloy (HSLA) steels, stainless steels, and other nickel alloys. A high-temperature nickel alloy filler metal is often used to make such dissimilar joints. Design and filler metal selection criteria usually include strength, corrosion resistance, and dilution tolerance of the filler metal for the elements present in the metals to be joined. In some cases, special design consideration is given to selecting filler metals with a specific coefficient of thermal expansion, thermal conductivity, electrical resistivity, and thermal fatigue resistance (Ref 8).

An example of a filler-metal-added dissimilar weld for high-temperature service can be found in hydrogen furnaces. In many applications, centrifugal-cast, secondary-carbide-strengthened furnace tubes are welded to collection headers of UNS 08800 using ERNiCrCoMo-1, to achieve a combination of strength, ductility, and corrosion resistance. An example of a successful autogenous dissimilar weld is the weld used in the manufacture of internal combustion engine turbochargers. In some of these, the difficult-to-weld Alloy 713C is friction-welded to an HSLA steel shaft such as 4340. Such a design economically uses the high-strength precipitation-hardenable alloy only where its high-temperature strength and corrosion resistance are needed.

Clad Metals and Overlays. A similar set of considerations also applies to filler metal selection for welding clad materials and for overlay cladding by welding. First, the materials selected must be capable of being mixed during welding without resulting cracks, and second, the resulting weldment must perform well in the intended service.

In overlay cladding and the welding of clad materials, the alloy contribution of the substrate to the final weld composition must be considered. For instance, if a one-layer overlay on steel is desired to have a minimum of 50 wt% Ni and 15 wt% Cr, and the chosen welding process yields 20 wt% dilution, then a filler metal with approximately 70 wt% Ni and 20 wt% Cr must be selected. In this case, ERNiCr-3 filler metal, which has 18 to 22 wt% Cr and 67 wt% (min) Ni, would yield a single-layer overlay chemistry of 16 wt% Cr and 54 wt% Ni.

Likewise, if two clad plates are to be welded together, it is usually desirable that the weld have a slightly richer alloy content than the cladding. In order to achieve this condition, a welding product of considerably higher alloy content than the cladding must be selected to allow for the dilution by the substrate. For example, refineries often use steel tubing with the inside surface clad with UNS08825 for elevated-temperature resistance to H_2S. This alloy cladding contains 42 wt% Ni, 20 wt% Cr, 3 wt% Mo, and 1 wt% Ti. In order to achieve a slightly enriched weld composition, it is welded using ERNiCrMo-3, which contains 60 wt% Ni, 22 wt% Cr, and 9 wt% Mo.

Outside-Diameter Clad Tubing. A more complex application is needed for butt welding boiler tubing made from UNS08800 that had been outside-diameter clad with a 45Cr-55Ni nickel alloy for resistance to molten salt and fuel-ash corrosion. The substrate was welded using ERNiCr-3 to develop sufficient stress-rupture strength, and the weld was capped with two layers of a proprietary near-matching filler metal using a low-dilution-rate process. (A crack-free higher-chromium filler metal was unavailable.) In this application, the filler metal chosen for welding the substrate had to provide the necessary strength after having been diluted by the substrate and had to be compatible with the capping weld layers. Because both materials were nickel-chromium types, the procedure went smoothly, and the tubing has given nearly 20 years of service in a demanding corrosive environment.

Repair welding, or welding after service, has been briefly addressed in some previous sections of this article. It has also been thoroughly discussed in the literature (Ref 9). Primarily, the same fabrication requirements must be met; namely, a sound weld must be produced that is capable of giving usable service in a given environment. If service conditions such as corrosion have rendered the material unweldable, the corrosion layers must be removed or neutralized. The precipitation-hardenable alloys and the secondary-carbide-strengthened alloys usually must have weldability restored by thermal treatment before repair is attempted.

In some cases, the usefulness of a part can be restored without a 100% joint-efficient weld design. In the example of worn turbine blades, greatly extended service can be attained by repair welding without reannealing by using a lower-strength but more ductile filler metal.

In yet other applications only a small part of a welded assembly may have become inoperative due to corrosion or tertiary creep. In these situations, it is highly desirable to remove the damaged portions, remove corrosion deposits, restore weldability if necessary, and select a suitable filler metal. Then, the assembly can be successfully repaired with a proven welding procedure to provide the desired extension in service life.

Alternatively, corrosion attack such as oxidation or carburization may have completely penetrated the full thickness of all parts. In such cases, attempts to repair the weld will be unsuccessful. In addition, the formation of extensive creep voids or thermal-fatigue cracking may have permeated the materials to such an extent that repair welding is impossible. In these cases, good diagnostic analysis is important for preventing wasted effort when attempting to repair parts that cannot be salvaged.

References cited in this section

5. B. IRVING, DISSIMILAR METAL WELDING PAVES THE WAY TO NEW VENTURES, *WELD. J.*, VOL 5, MAY 1992, P 27-32
6. S.D. KISER, DISSIMILAR WELDING WITH NICKEL ALLOYS, *CAN. WELD. FABRICAT.*, JAN 1980
7. *PROCEEDINGS: SEMINAR ON DISSIMILAR WELDS IN FOSSIL-FIRED BOILERS*, EPRI-CS-3623, PROJECT 1874-1, JULY 1985
8. S.D. KISER, NICKEL ALLOYS CONSUMABLE SELECTION FOR SEVERE SERVICE CONDITIONS, *WELD. J.*, VOL 11, NOV 1990, P 30-35
9. S.D. KISER, REPAIR WELDING OF NICKEL ALLOYS, *CAN. WELD. FABRICAT.*, VOL 12, DEC 1980

Special Metallurgical Welding Considerations for Nickel and Cobalt Alloys and Superalloys

S.D. Kiser, Inco Alloys International, Inc.

References

1. *WELDING--IAI-14,* INCO ALLOYS INTERNATIONAL, JULY 1991, TABLE 18, P 23
2. D.S. DUVALL AND J.R. DOYLE, *REPAIR OF TURBINE BLADES AND VANES,* PUBLICATION 73-GT-44, AMERICAN SOCIETY OF MECHANICAL ENGINEERS, 1973
3. C.W. COX AND S.D. KISER, FUSION WELDING OF DISSIMILAR METALS FOR HIGH TEMPERATURE STRENGTH, *WELD. J.,* VOL 5, MAY 1991, FIGURE 1, P 68
4. R.E. AVERY AND C.M. SCHILLMOLLER, REPAIR WELDING HIGH ALLOY FURNACE

TUBES,*HYDROCARBON PROCESS.*, JAN 1988
5. B. IRVING, DISSIMILAR METAL WELDING PAVES THE WAY TO NEW VENTURES, *WELD. J.*, VOL 5, MAY 1992, P 27-32
6. S.D. KISER, DISSIMILAR WELDING WITH NICKEL ALLOYS, *CAN. WELD. FABRICAT.*, JAN 1980
7. *PROCEEDINGS: SEMINAR ON DISSIMILAR WELDS IN FOSSIL-FIRED BOILERS,* EPRI-CS-3623, PROJECT 1874-1, JULY 1985
8. S.D. KISER, NICKEL ALLOYS CONSUMABLE SELECTION FOR SEVERE SERVICE CONDITIONS, *WELD. J.,* VOL 11, NOV 1990, P 30-35
9. S.D. KISER, REPAIR WELDING OF NICKEL ALLOYS, *CAN. WELD. FABRICAT.,* VOL 12, DEC 1980

Special Metallurgical Welding Considerations for Refractory Metals

Bryan A. Chin, Probal Banerjee, Jiayan Liu, and Shaofeng Chen, Auburn University

Introduction

THE REFRACTORY METALS, which include niobium (also called columbium), tantalum, molybdenum, tungsten, and rhenium, have the highest melting temperatures and lowest vapor pressures of all metals, except osmium and iridium. The refractory metals are readily degraded by oxidizing environments at moderately low temperatures, a property that has restricted the applicability of the metals in low-temperature or nonoxidizing high-temperature environments.

This article is limited to a discussion of special metallurgical considerations during the fusion welding of refractory metal alloys. These alloys are discussed in order of decreasing weldability: tantalum, niobium, rhenium, molybdenum, and tungsten.

Bryan A. Chin, Probal Banerjee, Jiayan Liu, and Shaofeng Chen, Auburn University

Tantalum Alloys

Tantalum is a refractory metal that readily reacts with atmospheric gases, especially at high temperatures. This metal melts at 3000 °C (5430 °F), has a body-centered cubic (bcc) structure, and does not undergo an allotropic transformation. It also exhibits excellent corrosion resistance to a wide spectrum of reagents. However, it oxidizes readily at temperatures above 343 °C (650 °F).

Carbon, oxygen, nitrogen, and hydrogen actively react with tantalum, typically forming intermetallics that lead to weld embrittlement. Although tantalum and its alloys are generally considered to have the best weldability of the refractory metals, the formation of these brittle intermetallics makes it difficult to weld to most structural metals. Processes used for joining tantalum alloys include gas-tungsten arc welding (GTAW), solid-state diffusion bonding, friction welding, resistance welding, brazing, and explosion bonding (Ref 1). Trailing shields help reduce atmospheric contamination due to the active nature of tantalum.

Effect of Microstructure

The fusion zones generally contain large grains and, in the absence of hafnium, cellular substructures (Ref 2). The proportion of large grains is higher in alloys with low alloy contents, because of a narrow freezing range (Ref 3) and a single-phase structure. Metallographic sections along the cell growth direction typically consist of nearly parallel grain boundaries. However, transverse sections show hexagonal cells. Increasing alloy content produces a more delineated substructure (Ref 2). Extensive carbide precipitation at grain boundaries can lead to severe embrittlement after a silicide coating procedure in certain weld configurations in Ta-10W alloy (Ref 4).

Ductility. Differences in grain size or cell size do not appear to strongly affect ductility, but material quality, or impurity concentration, does have an influence. High interstitial contents in the alloys can lead to a large concentration of micropores. Only in special carbon-containing Ta-5W-2.5Mo alloys does weld hardness influence ductility (Ref 2). Ductility is reduced in surface-ground welded specimens. Welds in Ta-5W-2.5Mo and Ta-10W-2.5Mo have excellent bend ductility at room temperature.

The GTAW process results in an increase in sheet bend transition temperature in Ta-8W-2Hf, Ta-10W, and Ta-9.6W-2.4Hf-0.01C (T-222), with the highest increase being in the T-222 material (Ref 3). With an increasing tungsten content, alloys become brittle at room temperature (Ref 2). A postweld heat treatment of 1 h at 816 °C (1500 °F) can significantly reduce the 2*t* bend transition temperature for welds produced by the GTAW process. The bend transition temperature is the lowest temperature below which cracking develops in a weld bent at a radius of 2*t* (twice the thickness of the sheet) with an angle of at least 90 °. Welds produced by electron-beam welding (EBW), however, have slightly higher transition temperatures (Ref 2). Under elevated-temperature exposure, there is considerable grain growth, which reduces the cellular fusion zone substructure. This leads to an increased impurity concentration at grain boundaries. In alloys susceptible to impurities, elevated-temperature exposure leads to a decrease in ductility. Generally, total metallic alloy contents of less than 10 at.% produce ductile welds, whereas alloy contents of greater than 13 at.% produce brittle welds (Ref 2).

Tensile Properties. A postweld heat treatment of 1 h at 816 °C (1500 °F) reduces the susceptibility of tantalum-tungsten-molybdenum alloys to premature tensile failure (Ref 2). Joint efficiencies of welds in wrought material are lower than those of welds in recrystallized materials. However, room-temperature tensile strengths of welds in wrought material are higher than those of welds in recrystallized material. The tensile strengths of welds drop very sharply above 1649 °C (3000 °F). Tensile strengths decrease with an increase in tungsten content, whether testing is conducted at room temperature or at elevated temperatures. In general, tensile strengths decrease exponentially with an increase in testing temperature (Ref 3) (Fig. 1).

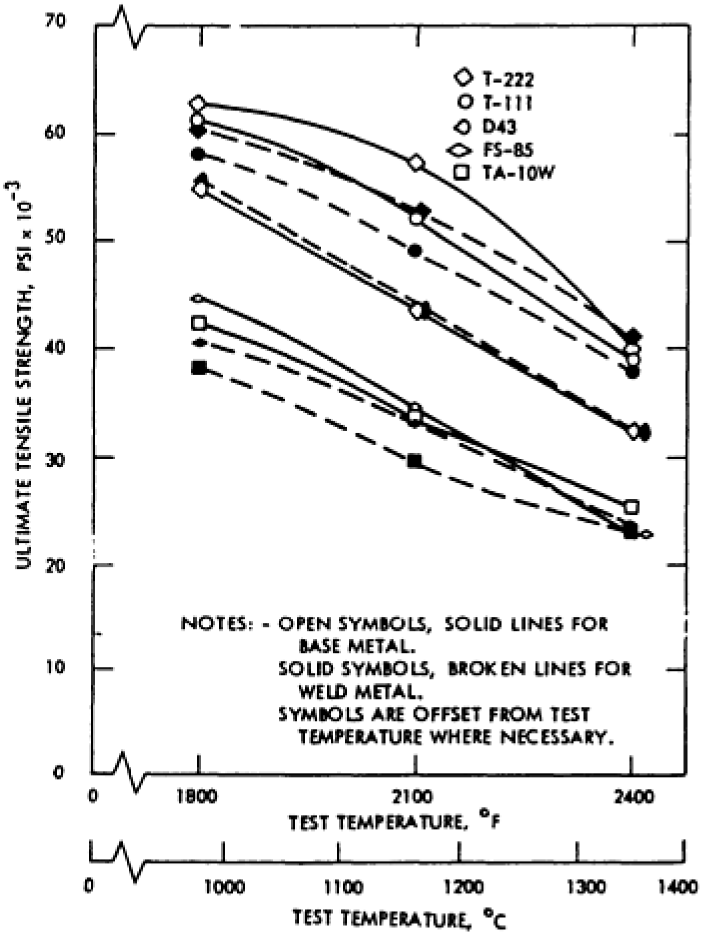

FIG. 1 ELEVATED-TEMPERATURE TENSILE STRENGTH OF ANNEALED BASE METAL AND TANTALUM ALLOY ARC WELDS. SOURCE: REF 3

Effect of Interstitial Impurities

Interstitial impurities can strongly influence the properties of welded tantalum materials. The average weld bend ductility is lower for gas-tungsten arc welds in low-interstitial-content materials. Generally, tantalum alloy sheet must have a low interstitial level (less than 100 ppm) to be weldable. Carbon contents of less than 100 ppm or hydrogen contents of less than 17 ppm do not significantly affect weld ductility (Ref 2).

Effect of Welding Conditions

Welding conditions can also be used to control properties in tantalum and its alloys. In low-interstitial-content sheets, arc travel speed and energy input do not appear to influence ductility. In high-interstitial-content sheets, ductility is better at high travel speeds (12.7 to 31.75 mm/s, or 30 to 75 in./min) and low energy inputs (98 to 118 J/mm, or 2500 to 3000 J/in.) (Ref 2).

Effect of Arc Oscillations

Arc oscillation has been found to reduce the columnar structure in the fusion zone (Ref 5), which promotes ductility. However, there is an optimum frequency that produces an equiaxed structure. If this frequency (10 Hz) is exceeded, then elongated grains result (Ref 5, 6).

Increased energy input results in larger heat-affected zone (HAZ) grains and fewer nucleation sites for fusion zone grains (Ref 2). There is an increase in hardness at the areas of metal deposition and at the HAZ in gas-tungsten arc welded tantalum (Ref 7).

References cited in this section

1. R.T. WEBSTER, JOINING OF TANTALUM AND NIOBIUM, *J. MET.*, VOL 36 (NO. 8), AUG 1984, P 43
2. P.A. KAMMER, R.E. MONROE, AND D.C. MARTIN, WELDABILITY OF TANTALUM ALLOYS, *WELD. J.*, JUNE 1972, P 304S-320S
3. G.G. LESSMAN, THE COMPARATIVE WELDABILITY OF REFRACTORY METAL ALLOYS, *WELD. J.*, DEC 1966, P 540S-560S
4. J. WADSWORTH AND C.M. PACKER, WELD EMBRITTLEMENT IN A SILICIDE-COATED TANTALUM ALLOY, *INT. J. REFRACT. HARD MET.*, VOL 2 (NO. 4), DEC 1983, P 164-169
5. A. GRILL, EFFECT OF ARC OSCILLATIONS ON THE TEMPERATURE DISTRIBUTION AND MICROSTRUCTURE IN GTA TANTALUM WELDS, *METALL. TRANS. B,* DEC 1981, P 667-674
6. Y. SHARIR, J. PELLEG, AND A. GRILL, EFFECT OF ARC VIBRATION AND CURRENT PULSES ON MICRO-STRUCTURE AND MECHANICAL PROPERTIES OF TIG TANTALUM WELDS, *MET. TECHNOL.*, JUNE 1978, P 190-196
7. H. HIROSE, K. SATO, AND T. HAYASHI, WELDING OF TANTALUM AND NIOBIUM, *WELD. LNT.*, AUG 1989, P 672-677

Special Metallurgical Welding Considerations for Refractory Metals

Bryan A. Chin, Probal Banerjee, Jiayan Liu, and Shaofeng Chen, Auburn University

Niobium Alloys

Niobium and its alloys are often selected for high-temperature use in fabricated systems. They can be easily welded by the GTAW, EBW, and laser-beam welding (LBW) processes, either autogenously or with matching filler metals in inert atmospheres. In special cases, resistance welding (RW), plasma arc welding (PAW), hot pressure welding (HPW), and explosion welding (EXW) can be considered.

Atmospheric contamination is a major cause of the embrittlement of niobium alloy welded joints. To avoid significant increases in the ductile-to-brittle transition temperature (DBTT) of welds (caused by interstitial contamination), welding atmospheres should have less than 600 ppm oxygen, 300 ppm nitrogen, and 150 ppm hydrogen.

Effect of Welding Atmosphere. Niobium and its alloys can be welded using an inert gas shield or a vacuum chamber. A vacuum-purged chamber backfilled with argon and helium is recommended for better-quality welds that are free from air contamination.

Effect of Microstructure. The DBTT of welds is also influenced by microstructural changes resulting from weld thermal cycles. The size of the HAZ and grain growth in this zone are major factors that affect the ductility of welds. Welding conditions that tend to increase the size of the HAZ, as well as the grains within it, will result in an upward shift in the transition temperature of welds. Furthermore, alloys with different strengthening mechanisms respond differently to weld heat input.

The solid-solution alloys, such as Nb-1Zr and Nb-10W-10Ta (SCb-291) show a continuous ductility loss with an increasing HAZ size and fusion zone grain size. A heat input limit, above which ductility impairment occurs, exists for solid-solution plus dispersion-strengthened alloys, such as Nb-10W-2.5Zr (Cb-752), Nb-27Ta-10W-1Zr (FS-85), Nb-10W-1Zr-0.2C (D-43), and Nb-5Mo-5V-1Zr (B-66). To reduce the tendency of grain growth in the HAZ and fusion zones, a high-energy-density welding process, such as EBW, is often more advantageous than GTAW.

Effect of Welding Conditions. The alloy Nb-1Zr can be easily welded using arc welding and EBW processes. The alloy B-66 has a hot tearing tendency during arc welding, which makes both manual and automatic welding more difficult. High-speed arc welding of Nb-10W-10Hf + Y (C-129Y) will cause gross shrinkage defects in the welds. Welds of D-43Y have the tendency to hot tear along the weld centerline and crack during welding. This factor, in conjunction with unmodified D-43, tends to produce high porosity in the weld.

Several alloys are temperature limited with respect to thermal stability. Welds of D-43 tend to lose strength with increasing aging time and temperature. Design stresses need to be reset for long-term applications of welded joints at temperatures above 1095 °C (2000 °F). During high-temperature aging, SCb-291 and B-66 welds gradually lose ductility, because of grain growth in both the weld and the HAZ.

Special Metallurgical Welding Considerations for Refractory Metals

Bryan A. Chin, Probal Banerjee, Jiayan Liu, and Shaofeng Chen, Auburn University

Rhenium Alloys

The weldability of rhenium is better than that of tungsten and molybdenum, but is slightly less than that of tantalum. Rhenium welds typically demonstrate a large degree of ductility. The only problem that can be encountered is porosity, which often occurs from welding a powder metallurgical product. It has been proposed that tungsten and molybdenum be alloyed with rhenium to improve both room-temperature ductility and weldability. The binary alloy of molybdenum with 50% rhenium can be easily welded using the GTAW and EBW processes. The welds retain nearly the same strength as the base metal and have good ductility.

Special Metallurgical Welding Considerations for Refractory Metals

Bryan A. Chin, Probal Banerjee, Jiayan Liu, and Shaofeng Chen, Auburn University

Molybdenum Alloys

Molybdenum and its alloys can be welded by the GTAW and EBW processes (Fig. 2), as well as most other welding methods. However, the weldability of molybdenum and its alloys is marginal, because of a high DBTT in the welded and recrystallized zones. Furthermore, the DBTT of molybdenum welds is significantly increased after the GTAW and EBW processes.

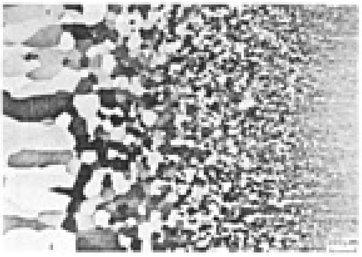

FIG. 2 WELD AND HAZ IN ELECTRON-BEAM WELDED TZM; ARROWS INDICATE FUSION LINE. SOURCE: REF 8

Effect of Welding Conditions. Heat input is a major factor influencing the ductility of welds. Electron-beam welding with relatively low heat input results in a lower DBTT than that obtained from the GTAW process. A high melting point, in conjunction with low ductility, also make molybdenum susceptible to failure by thermal impact during welding. Therefore, weldability can be improved by a weld preheat. If their DBTT values are below the ambient temperature, then molybdenum and most of its alloys with low DBTT values can be welded without a preheat. Alloys with high DBTT values must be preheated above their respective transition temperatures to avoid cracking during welding. Postweld annealing for stress relief helps improve weld ductility. The brittle behavior is principally the result of the large grain size in the HAZ and fusion zone. By reducing the grain size difference between the parent and weld metals through heat treatment, significant increases in ductility can be achieved. TZM (Mo-0.5Ti-0.1Zr) welds are not inherently as brittle as other molybdenum alloys (Ref 8).

Effect of Interstitial Impurities. Molybdenum and tungsten both have extremely low solubility limits for interstitials such as oxygen, nitrogen, and carbon (1 ppm or less). Of these three, oxygen is particularly harmful to ductility. Just several parts per million are enough to raise the DBTT from well below the ambient temperature to well above it. To prevent interstitial contamination, welding in vacuum or in an inert-gas-filled chamber is recommended. A high-purity shield gas is effective in promoting ductile joints.

Hot cracking and weld porosity represent other problems that can arise when welding powder metallurgy alloys. These problems often occur even when the welds are made in high vacuum. Large amounts of oxygen and other gases that are generally contained in the powder metallurgy base metal will react with the molybdenum during welding, leading to embrittlement.

Reference cited in this section

8. J. WADSWORTH, G.R. MORSE, AND P.M. CHEWEY, THE MICROSTRUCTURE AND MECHANICAL PROPERTIES OF A WELDED MOLYBDENUM ALLOY, *MATER. SCI. ENG.,* JUNE 1983, VOL 59, P 257-273

Special Metallurgical Welding Considerations for Refractory Metals

Bryan A. Chin, Probal Banerjee, Jiayan Liu, and Shaofeng Chen, Auburn University

Tungsten Alloys

Tungsten and tungsten alloys can be joined by welding (GTAW and EBW), brazing, diffusion bonding, and other joining techniques. Tungsten and its alloys are generally considered to have poor weldability, when compared with other refractory metals, such as niobium and tantalum. Welding problems encountered with tungsten and its alloys are generally traceable to physical metallurgy and properties. Chemical changes that are due primarily to atmospheric contamination during heating and cooling and microstructural changes that are due to thermal cycling can result in grain growth, hardening, and embrittlement. Embrittlement can often be traced to recrystallization and oxygen ingression along grain boundaries.

Of the refractory metals, tungsten has the highest modulus and melting point. This high melting point results in greater than normal cooling time intervals after joining and a large range of cooling temperatures over which sizeable thermal strains can be generated, because of differential thermal expansion and contraction. High elastic modulus, high yield strength, and resistance to plastic flow lead to a rapid rise in residual stress with an increase in thermal strain. High residual stresses are commonly encountered in tungsten welds. If the welds are made under restraint, then they may crack during cooling.

Effect of Alloy Production Technique. The weldability and properties of tungsten strongly depend on the process used to make the tungsten. Porosity and entrained gases in powder metallurgy tungsten products will result in weld porosity. This porosity usually concentrates along the fusion boundary. To a lesser extent, welds of chemical vapor deposition (CYD) tungsten products will show porosity along the fusion line and in the HAZ (Ref 9). Welds of arc cast tungsten materials can be made porosity free if proper welding procedures are followed. Tungsten weld embrittlement can be caused by contamination and/or recrystallization.

Effect of Interstitial Impurities. Tungsten, which is composed of a bcc crystal lattice, is sensitive to the presence of interstitial atoms such as carbon, oxygen, and nitrogen. Because the solubility of these elements in tungsten is very low, their embrittling effects are much greater. Tungsten has a particularly high affinity for carbon and oxygen, readily forming carbides (WC) and oxides (WO_2).

Carbon, oxygen, and nitrogen dramatically increase the bend ductility transition temperature and DBTT, as shown in Fig. 3. Typical sources of these contaminants are moisture/impurities in the shielding gas, surface films on the base metals and/or filler metals, and oils on the components being welded. Use of either a vacuum or an inert atmosphere chamber consisting of specially purified shielding gas (argon, helium, or combinations of the two) is necessary to control contamination.

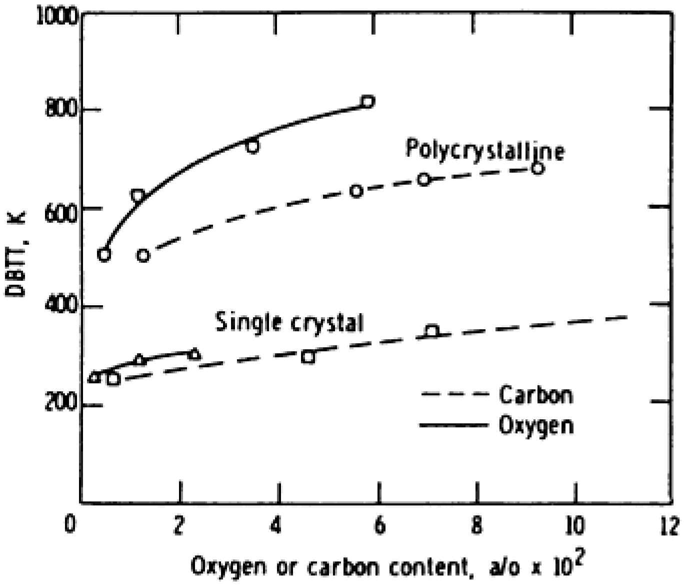

FIG. 3 EFFECTS OF OXYGEN AND CARBON ON DBTT OF TUNGSTEN. SOURCE: REF 10

Effect of Microstructure. Recrystallization and grain growth in the HAZ (Fig. 4) is another effect that can also contribute to embrittlement. Increasing grain size reduces the tensile strength, ductility, and toughness of tungsten alloys. Coarse grains in the HAZ have been found to increase the DBTT. The effect of grain size on DBTT is shown in Fig. 5. The effect of coarse grains on weldment ductility may be minimized by using a low heat input and by choosing proper welding procedures and parameters that limit recrystallization and grain growth effects. The EBW process can be used to obtain an extremely narrow HAZ. This technique has been shown to minimize embrittlement that is caused by recrystallization and grain growth.

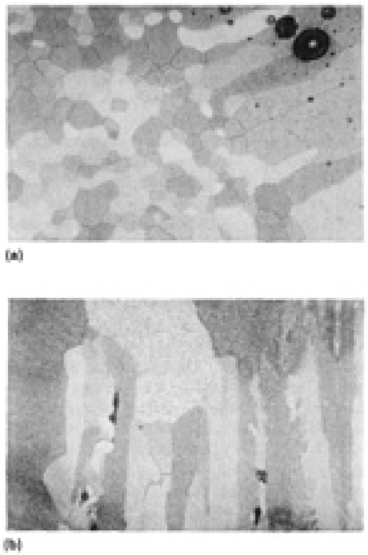

FIG. 4 COARSE GRAINS IN HAZ. 65×. (A) POWDER METALLURGY TUNGSTEN; POROSITY SHOWN ALONG FUSION LINE. (B) CVD TUNGSTEN

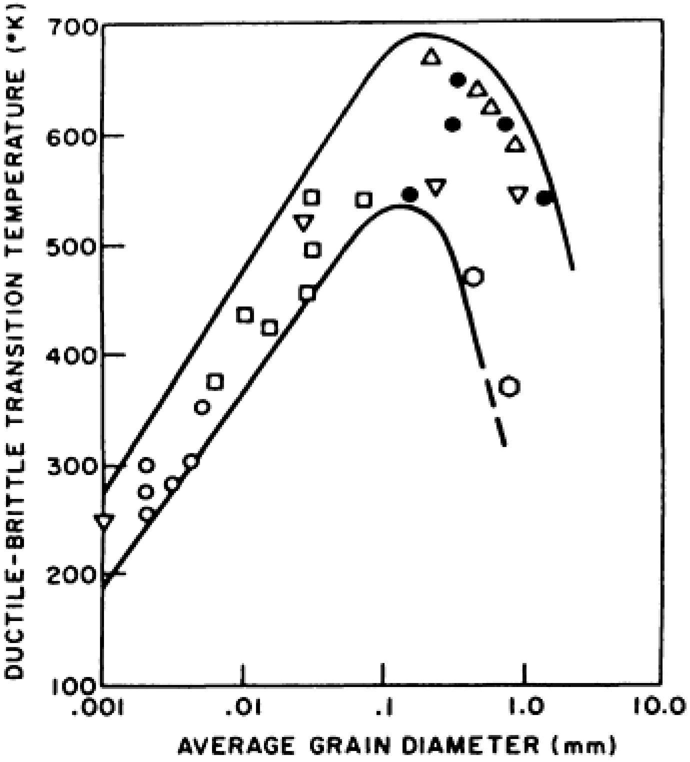

FIG. 5 EFFECT OF GRAIN SIZE ON DBTT OF TUNGSTEN. SOURCE: REF 10

Weld cracking usually can be avoided by preheating to above the DBTT. Welding tungsten-rhenium alloys or using the EBW process makes it possible to avoid preheat requirements. When CVD tungsten is welded, cracks are usually found in the columnar grain boundaries in the HAZ, having been caused by the growth and coalescence of small pores on the grain boundaries.

References cited in this section

9. K. FARRELL, J.T. HOUSTON, AND J.W. CHUMLEY, HOT CRACKING IN FUSION WELDS IN TUNGSTEN, *WELD. J.,* MARCH 1970, P 132S-137S
10. S.W.H. YIH AND C.T. WANG, *TUNGSTEN,* PLENUM PRESS, 1979

Special Metallurgical Welding Considerations for Refractory Metals

Bryan A. Chin, Probal Banerjee, Jiayan Liu, and Shaofeng Chen, Auburn University

References

1. R.T. WEBSTER, JOINING OF TANTALUM AND NIOBIUM, *J. MET.,* VOL 36 (NO. 8), AUG 1984, P 43
2. P.A. KAMMER, R.E. MONROE, AND D.C. MARTIN, WELDABILITY OF TANTALUM ALLOYS, *WELD. J.,* JUNE 1972, P 304S-320S
3. G.G. LESSMAN, THE COMPARATIVE WELDABILITY OF REFRACTORY METAL ALLOYS, *WELD. J.,* DEC 1966, P 540S-560S
4. J. WADSWORTH AND C.M. PACKER, WELD EMBRITTLEMENT IN A SILICIDE-COATED TANTALUM ALLOY, *INT. J. REFRACT. HARD MET.,* VOL 2 (NO. 4), DEC 1983, P 164-169
5. A. GRILL, EFFECT OF ARC OSCILLATIONS ON THE TEMPERATURE DISTRIBUTION AND MICROSTRUCTURE IN GTA TANTALUM WELDS, *METALL. TRANS. B,* DEC 1981, P 667-674
6. Y. SHARIR, J. PELLEG, AND A. GRILL, EFFECT OF ARC VIBRATION AND CURRENT PULSES ON MICRO-STRUCTURE AND MECHANICAL PROPERTIES OF TIG TANTALUM WELDS, *MET. TECHNOL.,* JUNE 1978, P 190-196
7. H. HIROSE, K. SATO, AND T. HAYASHI, WELDING OF TANTALUM AND NIOBIUM, *WELD. LNT.,* AUG 1989, P 672-677
8. J. WADSWORTH, G.R. MORSE, AND P.M. CHEWEY, THE MICROSTRUCTURE AND MECHANICAL PROPERTIES OF A WELDED MOLYBDENUM ALLOY, *MATER. SCI. ENG.,* JUNE 1983, VOL 59, P 257-273
9. K. FARRELL, J.T. HOUSTON, AND J.W. CHUMLEY, HOT CRACKING IN FUSION WELDS IN TUNGSTEN, *WELD. J.,* MARCH 1970, P 132S-137S
10. S.W.H. YIH AND C.T. WANG, *TUNGSTEN,* PLENUM PRESS, 1979

Special Metallurgical Welding Considerations for Refractory Metals

Bryan A. Chin, Probal Banerjee, Jiayan Liu, and Shaofeng Chen, Auburn University

Selected References

- N.C. COLE, R.G. GILLILAND, AND G.M. SLAUGHTER, WELDABILITY OF TUNGSTEN AND ITS ALLOYS, *WELD. J.,* SEPT 1971, P 419S-426S
- E.S. ROBERT, *REFRACTORY METALS AND THEIR INDUSTRIAL APPLICATIONS,* STP 849, ASTM
- E.M. SAVITSKII, M.A. TYLKINA, AND K.B. POVAROVA, *RHENIUM ALLOYS,* ISRAEL PROGRAM FOR SCIENTIFIC TRANSLATIONS, 1970
- J. WADSWORTH, T.G. NIEH, AND J.J. STEPHENS, RECENT ADVANCES IN AEROSPACE REFRACTORY METAL ALLOYS, *INT. MATER. REV.,* VOL 33 (NO.3), 1988, P 131-150

Introduction to the Selection of Nonferrous Corrosion-Resistant Materials

S.J. Matthews, Haynes International, Inc.

Introduction

CORROSION has a tremendous cost impact on our industrial society, estimated to be on the order of $170 billion/year. There are many different forms of corrosion, and the interested reader is referred to *Corrosion,* Volume 13 of the *ASM Handbook,* for a detailed description of the subject. Improvement in corrosion performance of a component or structure can be achieved through proper design, surface protection, proper material selection, or combinations of all three parameters. This Section of the Handbook will address weldability and material selection issues associated with nonferrous corrosion-resistant alloys.

Stainless steel, as its generic name implies, has been a traditional material employed to resist corrosion. However, some corrosion service is so harsh that even stainless steels do not perform adequately. In the chemical process industry, for example, the manufacture of phosphoric, acetic, hydrochloric, and sulfuric acids requires highly corrosion-resistant materials. A boiling 1% HCI solution, for example, can totally dissolve type 316 stainless steel during laboratory corrosion testing. For many applications, nickel-base alloys are called upon to resist certain severe corrosion environments.

Pure nickel has good resistance to reducing environment corrosion and is an excellent base upon which to build corrosion-resistant alloys. The atomic size, outer-shell electron configuration, and face-centered cubic (fcc) crystal structure of nickel make this element amenable to solid-solution alloying with elements such as iron, copper, chromium, tungsten, and molybdenum. Each of these alloying elements can contribute unique corrosion resistance to specific forms of corrosion attack, such as chloride pitting, galvanic corrosion, and stress-corrosion cracking. A wealth of empirical evidence and experience is available regarding the effects of alloying elements, individually and cumulatively, on corrosion resistance in aqueous environments (Ref 1).

Because industrial corrosion environments are diverse and complex, very few alloys can be expected to resist all forms of corrosion. Hence, numerous different nickel-base alloys have evolved over the last 50 years, each with a different combination of alloying elements to serve a particular niche in the corrosion market. It is not the intent of this Section to cover all nonferrous corrosion-resistant alloys on the market today. Rather, this Section has been organized into three different groupings of corrosion-resistant alloys; the information contained in these articles should also apply to similar alloys not listed. The first article addresses a family of materials that includes pure nickel, nickel-copper alloys, nickel-chromium alloys, and nickel-chromium-iron alloys. Monel and Inconel alloys are well-known members of this group of alloys. The next article addresses families of nickel-base alloys containing appreciable amounts of molybdenum. Molybdenum additions to nickel substantially improve resistance to nonoxidizing acid environments. Hastelloy alloys are typical of alloys belonging to this category of materials. In the last article, weldability and material selection issues associated with corrosion-resistant materials not based on nickel are discussed. These materials include cobalt-base corrosion-resistant alloys and commercially pure titanium, zirconium, and tantalum.

Reference

1. N. SRIDHAR, J.B.C. WU, AND P.E. MANNING, CORROSION RESISTANT NI-CR-MO ALLOYS, *J. MET.,* NOV 1985

Corrosion Performance

Information presented in these articles about corrosion performance should be treated as generalized claims. Potential material users are advised to seek actual laboratory corrosion data or field test experience from the material producer before selecting alloys for actual use. When communicating with a corrosion material specialist for the purposes of selecting alloys, it is important to identify five environmental parameters:

- CHEMICAL NATURE OF THE AQUEOUS (OR GASEOUS) ENVIRONMENT, INCLUDING IMPURITIES
- TEMPERATURE
- PH
- HALIDE CONCENTRATION (THAT IS, PRESENCE OF CHLORIDE OR FLUORIDE IONS)
- REDOX POTENTIAL (IF KNOWN)

Redox potential refers to the electrolytic nature of the aqueous environment. It determines whether the system is "oxidizing" or "reducing" to the exposed material and has an important influence on proper material selection.

Introduction to the Selection of Nonferrous Corrosion-Resistant Materials

S.J. Matthews, Haynes International, Inc.

Reference

1. N. SRIDHAR, J.B.C. WU, AND P.E. MANNING, CORROSION RESISTANT NI-CR-MO ALLOYS, *J. MET.,* NOV 1985

Selection of Nickel, Nickel-Copper, Nickel-Chromium, and Nickel-Chromium-Iron Alloys

Donald J. Tillack, D.J. Tillack and Associates

Introduction

THE NICKEL-BASE FAMILY OF ALLOYS was developed in the early 1900s, and significant amounts of nickel began to be used in engineering materials in the 1920s. Although nickel was discovered by a Swedish scientist in 1751, them were no useful applications of the element until the late 1800s.

In nature, nickel is found in two basic forms: in lateritic deposits (as oxides) and in a nickel-sulfide ore. Usually, nickel is found in combination with copper and numerous other elements. The ratio of nickel to copper in these ore bodies is normally 2:1. Therefore, it is not too surprising that the first major alloy based on a nickel-containing ore was composed of 67% nickel and 33% copper. This is also the same composition, roughly, of Monel alloy 400, which was patented in 1906 and first melted and cast commercially in the United States in 1921 by the International Nickel Company. Subsequently, numerous alloys that could be used in a wide variety of applications were developed.

The Monel composition was used extensively in corrosion-resisting applications early in its history, because stainless steel had not yet been invented. Kitchen and food preparation equipment, in particular, were made from "Monel metal," because of its corrosion resistance, ease of cleaning, and attractive appearance.

The consumption of high-nickel-content corrosion-resistant alloys grow at an impressive rate during the postwar years. The cost of nickel alloys was high after World War I and high cost per unit weight continues to be a primary concern in the present use of these alloys. Although these alloys are more expensive than some other materials, their high performance generally makes them cost effective in the long run. Although the current annual consumption of nickel alloys is a small fraction of the steel consumption total, the numbers are still impressive. In 1990, it was estimated that the Western nations consumed approximately 40 × 10^9 g (89 × 10^6 lb) of solid-solution corrosion-resistant nickel, nickel-copper, nickel-chromium, and nickel-chromium-iron alloys. Most (85%) were wrought products.

Because of their relatively high cost, these alloys are usually only specified when lower-cost alloys are unable to withstand the environment or are not strong enough. Often, the combination of these two criteria is the reason that nickel alloys are specified. No other alloy system combines the qualities of strength, environmental resistance, weldability, formability, and availability more completely than the nickel-base alloy family.

The compositions of some of the popular, commercially available, solid-solution wrought nickel, nickel-copper, nickel-chromium, and nickel-chromium-iron alloys are given in Table 1. This list does not include the molybdenum-rich alloys nor the precipitation-hardened alloys. The former are covered in the article "Selection of Nickel-Base Corrosion-Resistant Alloys Containing Molybdenum" in this Section of the Volume. The latter are described in the Section "Selection of Nonferrous High-Temperature Materials" in this Volume.

TABLE 1 COMPOSITIONS OF SELECTED NICKEL-BASE CORROSION-RESISTANT ALLOYS

ALLOY	COMPOSITION, WT%								
	Ni	Cr	Fe	Cu	C0	Si	Al	Ti	Mo
NICKEL ALLOYS									
200	99.5	. . .	. . .	. . .	0.08	0.2	. . .	. . .	. . .
201	99.5	. . .	. . .	. . .	0.01	0.2	. . .	. . .	. . .
NICKEL-COPPER ALLOY									
400	66.5	. . .	. . .	31.5	0.15	0.2	. . .	. . .	. . .
NICKEL-CHROMIUM ALLOYS									
600	76.0	15.5	8	. . .	0.08	0.2	. . .	. . .	. . .
601	60.5	23	14	. . .	0.08	0.2	1.3	. . .	. . .
690	60.0	30	9.3	. . .	0.03	0.2	0.3	0.2	. . .
NICKEL-CHROMIUM-IRON ALLOYS									
800	32.5	21	46	. . .	0.05	0.5	0.3	0.3	. . .
825	42	21	30	. . .	0.03	0.2	0.1	0.9	3

The commercially pure nickel grades are used extensively in applications that utilize their corrosion resistance, such as caustic service areas and food handling equipment, and in applications that utilize their electronic or magnetic characteristics, such as transducers. Alloy 201 has a low carbon content and is used in applications above 315 °C (600 °F), where long time exposure would cause the precipitation of embrittling graphite in alloy 200.

The nickel-copper alloys are used extensively in seawater applications, because of their excellent resistance to corrosion, cavitation, and erosion. Other significant applications include their use in the presence of chlorinated solvents, sulfuric acid, and alkalies. Typical industrial applications include valves and pumps, propeller shafts, marine fixtures and fasteners, chemical processing equipment, process vessels and piping, and heat exchangers.

The nickel-chromium alloys are used in severely corrosive environments at elevated temperatures, as well as at lower temperatures. Typical applications include furnace muffles and retorts, carburizing baskets, nuclear reactor components, thermocouple protection tubes, radiant furnace tubes, and chemical and food processing equipment.

The nickel-chromium-iron alloys are typically used in oxidizing and carburizing applications because of their resistance to these environments and their high strength. Typical applications include furnace components, heat exchangers and process piping, electric range heating-element sheathing, extruded tubing for ethylene and steam methane reforming furnaces, and chemical processing equipment.

All of the nickel-base alloys are resistant to sigma formation (that is, a hard, brittle, non-magnetic intermetallic intermediate phase with a tetragonal crystal structure) and to stress-corrosion cracking. They are austenitic throughout the range of temperatures in which they are normally used. This metallurgical stability enables designers to specify these alloys with the confidence that exposure to service temperatures will not cause premature structural damage to the materials.

The nickel and nickel-copper alloys are generally easier to fabricate than steels of the same thickness. The nickel-chromium and nickel-chromium-iron alloys are stronger than most steels, but can be formed with modest changes in technique. For example, their spring-back is greater, but if these alloys are overbent slightly more than normal, they can usually be formed on the same equipment that is used for steels. The nickel-base alloys can be work hardened more quickly than steels and require anneals more frequently during forming operations.

Selection of Nickel, Nickel-Copper, Nickel-Chromium, and Nickel-Chromium-Iron Alloys

Donald J. Tillack, D.J. Tillack and Associates

General Welding Characteristics

Thermal Conductivity. Nickel is quite conductive and copper is extremely conductive. However, the addition of chromium or iron decreases the thermal conductivity of nickel or nickel-copper. Conductivity, whether thermal or electrical, always decreases when the solvent accepts solute atoms, as in alloying. The data in Table 2 illustrate this fact. Note that nickel has a thermal conductivity value of 75 W/m · K (45 Btu/ft · h · °F) and copper (not shown in Table 2) has a value of 391 W/m · K (225 Btu/fi · h · °F), but alloy 400, which contains roughly two-thirds nickel and one-third copper, has a lower thermal conductivity value (22 W/m · K, or 13 Btu/ft · h · °F) than either nickel or copper.

TABLE 2 PROPERTIES OF SELECTED NICKEL-BASE CORROSION-RESISTANT ALLOYS

ALLOY	DENSITY, G/CM^3	MELTING RANGE		COEFFICIENT OF THERMAL EXPANSION(A), 10^{-6}/K	THERMAL CONDUCTIVITY		ELECTRICAL RESISTIVITY, 10^{-8} Ω/M
		°C	°F		W/M^2·K	BTU/FT^2 · H · °F	
1020 CARBON STEEL(B)	7.9	1470-1530	2680-2790	12	46	27	19
304 STAINLESS STEEL(B)	8.0	1390-1450	2535-2640	17	16	9.3	72
200	8.9	1435-1502	2615-2735	14	75	44	9
201	8.9	1435-1502	2615-2735	14	79	46	8
400	8.8	1299-1349	2370-2460	16	22	13	51
600	8.4	1354-1413	2470-2575	14	15	8.7	103
601	8.1	1301-1368	2375-2495	15	11	6.4	119
690	8.2	1343-1377	2450-2510	14	12	7	115
800	7.9	1357-1385	2475-2525	16	12	7	99
825	8.1	1371-1399	2500-2550	15	11	6.4	113

(A) IN RANGE FROM 20 TO 260 °C (68 TO 500 °F).

(B) DATA PROVIDED FOR PURPOSE OF COMPARISON

Thermal conductivity plays an important function during welding because it influences heat flow and the amount of heat required to melt material into the weld puddle. A base material composition with a high level of thermal conductivity will conduct heat away from the fusion zone and therefore requires slightly more heat input during welding than would an alloy with low thermal conductivity, which retains the local heat of welding for a longer time.

The nickel-base alloys vary considerably in their thermal conductivity characteristics. The nickel-chromium and nickel-chromium-iron alloys have values lower than those of both the carbon steels and the austenitic stainless steels, whereas nickel and the nickel-copper alloys have values considerably higher than those of the steels.

The electrical resistivity of a material is especially important in the welding process itself, because the melting rate of an electrode or wire is influenced by the electrical characteristics of the material (Table 2). Usually, solid-solution alloys are not very good electrical conductors because the solute and solvent atoms are of different sizes, which distorts the crystal lattice structure of the material and impedes the passage of electrons. The alloys that are composed primarily of one element, such as 1020 steel and nickel-base alloys 200 and 201, have the lowest electrical resistivity values of the alloys listed in Table 2.

The thermal expansion characteristics of a material are important because the heat that is applied to weld joints causes localized expansion and/or contraction during heating and/or cooling. A material with a greater thermal expansion value will expand and contract more during the welding operation, thus creating a greater amount of stress on the weld joint. This is particularly important when the weld deposit is marginally crack sensitive. It is also a consideration when dissimilar metals are welded together, especially when them is a large difference in expansion characteristics. The section "Dissimilar Welding" in this article provides more details on this topic. The nickel-base alloys have thermal expansion values that are between those of carbon steel and the austenitic stainless steels, as shown in Table 2.

Liquidus-Solidus Temperature. The cracking resistance of a material is often associated with the difference in the liquidus and solidus temperatures, particularly when stress is involved. When a highly restrained structure is welded, considerable stress is generated during the welding operation. If a material has a relatively large spread between its liquidus and solidus temperatures, then there is more time that the material is in a vulnerable situation in terms of cracking. However, a material with a relatively narrow liquidus-solidus range will cool down fairly rapidly through the temperature range, and there will be less time for stress-related cracking to occur. This example assumes that other factors that influence cracking, such as grain size or structural mass, are equal. Table 2 lists the liquidus-solidus temperatures, or melting ranges, of several solid-solution nickel alloys.

Selection of Nickel, Nickel-Copper, Nickel-Chromium, and Nickel-Chromium-Iron Alloys

Donald J. Tillack, D.J. Tillack and Associates

Welding Metallurgy

The heat-affected zone (HAZ) is discussed below in terms of grain boundary precipitation, grain growth, and hot cracking.

Grain Boundary Precipitation. Grain boundaries are the weak link in a microstructure that is exposed to elevated temperature, whether that exposure occurs during service or, in the case of welding, during the thermal cycles that are part of the welding operation. Precipitates that form in the grain boundaries can vary in size and composition and can be either harmful or beneficial. Intermetallic compounds, such as carbides, act as barriers to grain growth and can be very beneficial. Intergranular carbides in alloys 600 and 690 can also deter stress-corrosion cracking during high-purity water exposure in nuclear reactor steam generator primary side environments (Ref 1).

Grain Growth. Generally, it is easier to make defect-free welds in material with a relatively fine grain size, because there is less of an undesirable element concentration that contributes to grain boundary cracking in a fine-grain network. With larger grains, there are fewer grain boundaries and the deleterious elements are more concentrated in them. However, it is not always possible to have a fine-grain structure to weld. Most high-temperature applications, for example, need a large grain size material to ensure elevated-temperature creep and rupture strength.

Even when a fine-grain material is welded, grain growth will occur in the HAZ. The width of this grain growth band depends on numerous variables, such as the welding process used, the heat input and travel speed, and the heat transfer characteristics of the base material. The electron-beam (EB) welding process, for example, will produce a narrower HAZ and a smaller band of grain growth than the gas-metal arc welding (GMAW) process.

Hot Cracking. In terms of constitutional liquation, HAZ fissures are seldom a problem in the nickel or nickel-copper alloys, but they have been the subject of considerable study in nickel-chromium and nickel-chromium-iron alloys.

Such fissures have been observed in alloy 600, for example, when relatively high heat inputs puts (2.4 MJ/m, or 45×10^3 J/in., and higher) were used. These fissures are typically narrow, linear intergranular separations that are perpendicular to the direction of weld travel. Fissures have been found in the HAZ of welds made by the GMAW, gas-tungsten arc welding (GTAW), and EB welding processes. An example of a typical fissure found in a nickel-chromium HAZ is shown in Fig. 1.

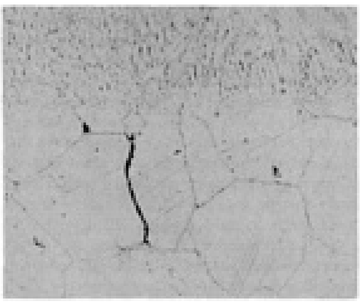

FIG. 1 PHOTOMICROGRAPH SHOWING A HEAT-AFFECTED ZONE MICROFISSURE IN A NICKEL-CHROMIUM ALLOY. SPECIMEN IS A GMAW 44.4 MM (1.75 IN.) PLATE. ELECTROLYTIC PHOSPHORIC ACID ETCHANT, 75×

Grain boundary liquation is one of the proposed mechanisms that causes fissuring (Ref 2). Grain boundary liquation occurs when a phase or particle causes the grain boundary to melt at a temperature below the bulk melting point. Because the liquated grain boundary has practically no strength, separation at the liquated grain boundary can occur, causing the fissure.

Constitutional liquation was proposed as a mechanism for hot cracking by Savage in 1959 and was described in detail in subsequent papers (Ref 3). In this theory, a relatively stable second-phase particle begins to dissolve during nonequilibrium, or rapid, heating and forms an intermediate phase with the matrix, which has a solidus temperature that is lower than that of either the second-phase particle or the matrix. In alloy 600, some experimenters have identified the precipitation of Cr_7C_3 at the grain boundary as the second-phase particle around which constitutional liquation occurs.

Chemical segregation at a grain boundary can also cause grain boundary liquation. Boron and sulfur are two elements that form lower melting point eutectics that can contribute to HAZ fissuring. Manganese segregation in alloy 600 has been identified as forming a binary compound (Mn-38Ni) with a 1000 °C (1832 °F) solidus temperature, which is considerably below the bulk melting range of the alloy.

Several approaches can be made to minimize the occurrence of HAZ fissuring, including the control of:

- HEAT INPUT DURING WELDING
- MICROSTRUCTURE, TO OBTAIN FAIRLY SMALL GRAIN SIZES
- CHEMISTRY, TO MINIMIZE THE AMOUNT OF DELETERIOUS MINOR ELEMENTS THAT ARE KNOWN TO CONTRIBUTE TO FISSURING

There are many minor elements that can contribute to HAZ fissuring in nickel-base alloys, including some that are intentionally added to enhance properties. Although these elements are considered minor because they are present in small amounts, they are far from being minor in their effect on fissuring. These elements include, but are not limited to, sulfur, phosphorus, lead, boron, zirconium, tin, zinc, selenium, tellurium, antimony, bismuth, silver, and gold. Boron and zirconium are two elements that are intentionally added to nickel-base alloys to enhance hot workability and elevated-temperature strength, but they can be detrimental during welding operations. More details on the effects of many of these elements are provided later in this article.

The fusion zone of autogenous welds is discussed below in terms of sensitivity to porosity formation, hot-cracking susceptibility, and microsegregation.

In terms of sensitivity to porosity, the nickel-base alloys can be divided into two groups: those that contain chromium and those that do not. Chromium has a natural affinity for the gases that form during the welding operation, notably oxygen, nitrogen, and hydrogen. Because the nickel and nickel-copper alloys lack chromium, they are sensitive to porosity formation during autogenous welding. Therefore, the protection of the weld puddle during melting and solidification is critical to the achievement of porosity-free autogenous welds in these two alloy classes. Dry torch gases and adequate gas protection are necessities. Although the use of filler metals helps in porosity control, because of the additions of gas-absorbing elements such as titanium and aluminum, proper precautions are still necessary.

Porosity control is much easier in the nickel-chromium and nickel-chromium-iron alloys because of their high chromium levels. This is not meant to imply that gas coverage is not important when autogenously welding the chromium-containing alloys, but it certainly is not as critical, compared with nickel and nickel-copper alloys.

Hot-Cracking Susceptibility. Hot cracking, or fissuring, is a grain boundary related mechanism that can occur, primarily, in the nickel-chromium and nickel-chromium-iron families of alloys. Although the nickel and nickel-copper alloys are not immune to this phenomenon, there have been more occurrences in the chromium-bearing nickel-base alloy materials.

A study of the weldability of alloy 800 (Ref 4) found that reducing the aluminum plus titanium content to extremely low levels (<0.06%) essentially eliminated hot cracking. It was noted that when normal levels of these elements were present, segregation of titanium and aluminum resulted in grain-boundary embrittlement and banding that was sufficiently severe to exhibit liquation during heating through a weld cycle. There also appeared to be a strong association between sulfur and titanium in the discrete inclusions observed in sulfur-bearing alloys.

Microsegregation. When a weld solidifies, an inhomogeneous dendritic structure forms. Solidification begins with the highest-melting-point liquid and continues until the whole structure is solid. A condition called microsegregation, or coring, occurs during this solidification process, and areas of the solidified microstructure can have great differences in composition. This becomes a problem in corrosion situations where an alloy was chosen for its particular resistance to corrosion. For example, because the dendritic weld material can have vastly different chromium and molybdenum levels, preferential corrosion may initiate and propagate at those areas of the microstructure that are lean in the element responsible for providing the corrosion resistance.

Microsegregation cannot be eliminated in a weld, but it can be minimized by judicious selection of welding parameters, such as the control of heat input. High-temperature heat treatments for several hours can partially equalize the composition gradients by diffusion. Cold working the weld structure, followed by annealing, can also break down this microsegregation, if the component lends itself to this type of treatment. Thin-sheet welds can be planished, for example. In addition to improving the corrosion resistance of the welds, cold working and annealing the weld area also increases weld ductility. Figure 2 illustrates how a dendritic weld structure can be reduced by cold working and annealing. After a 20% cold reduction followed by an anneal, the cast structure of the weld has been refined substantially and the wrought grain structure can be seen easily.

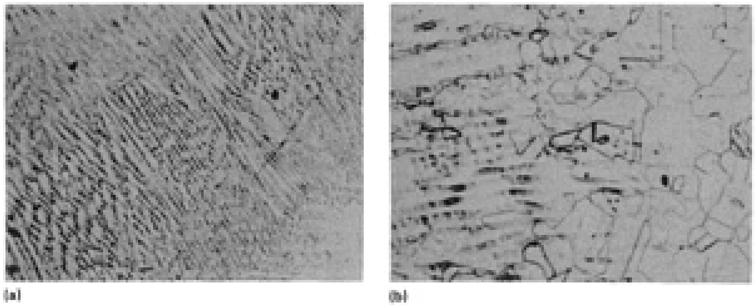

FIG. 2 MICROSTRUCTURES OF ALLOY 400 (UNS N04400) WELDED WITH FILLER METAL 60. (A) AS WELDED; CYANIDE PERSULFATE ETCHANT, 70×. (B) WELDED, PLUS 20% COLD REDUCTION, PLUS ANNEAL AT 871 °C (1600 °F)/2 H; CYANIDE PERSULFATE ETCHANT, 150×. SOURCE: REF 5

Unmixed Zone. For many years, it was assumed that there was complete mixing in the weld pool and that the composition of a fusion weld was homogeneous throughout. Now it is known that there are areas adjacent to the fusion line where mixing does not occur. This unmixed zone has been the subject of considerable research and is of concern to corrosion engineers because of the adverse effect that a cast base-metal structure can have on weld-associated corrosion. These zones are base material undiluted by filler metal and are, essentially, narrow bands of autogenous welds situated on the fusion line. Figure 3 shows an unmixed zone (Ref 6). There is a greater tendency for unmixed zones to be present in nickel-base welds, compared with steel and stainless steel welds. The sluggish nature of the molten nickel alloy weld metal tends to deter complete mixing during welding.

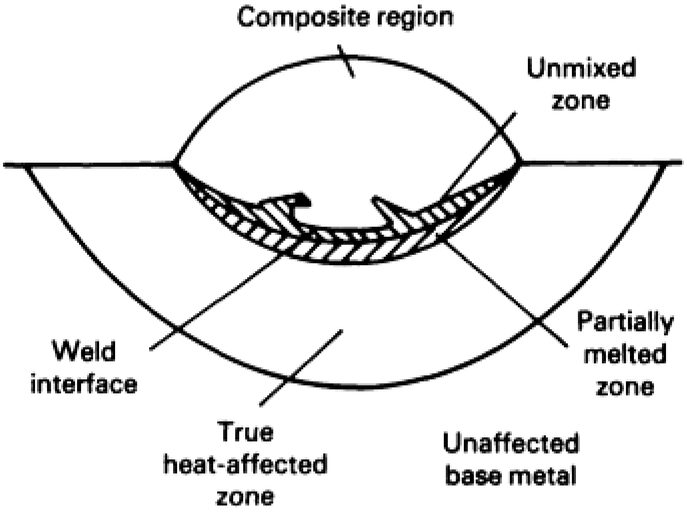

FIG. 3 SCHEMATIC SHOWING LOCATION OF UNMIXED ZONE IN A SOLIDIFIED WELDMENT. SOURCE: REF 6

These unmixed zones can be reduced or practically eliminated by controlling the welding procedures to ensure adequate mixing of the weld pool during welding. As an example, excessively fast travel speed promotes the formation of these zones. Slowing the travel speed and adjusting the other welding parameters to compensate for the slower travel speed will do a great deal to minimize the formation of unmixed zones.

The effects of alloying elements on the weldability of nickel-base alloys are described below.

Copper forms a complete series of solid-solution alloys with nickel and has little effect on the weldability of nickel. Metallurgically, the binary high-nickel-content alloys that contain from 15 to 40% Cu behave much like commercially pure nickel when welded. The nickel-copper alloys have about the same susceptibility to embrittlement by elements such as sulfur as does commercially pure nickel. The need for alloying to control gaseous porosity is not significantly reduced in the nickel-copper alloys, compared to that required for commercially pure nickel. As copper is increased up to the range of the copper-nickel alloys, such as 70Cu-30Ni and 80Cu-20Ni, the alloy behaves more like a copper alloy and reduced amounts of deoxidizers can be used. The presence of copper changes the high-temperature solubility of carbon to the point where graphitization is not encountered.

The chromium addition in alloys of commercial importance does not exceed the limit of solid-solution alloys and presents an attractively narrow freezing range. Chromium has the ability to form stable oxides and nitrides, which in turn greatly reduce the tendency for weld porosity, compared to that encountered in commercially pure nickel and nickel-copper welds. The chromium-bearing alloys are more susceptible than other high-nickel-content alloys to hot cracking when other elements, notably silicon, are present. For example, the hot-cracking susceptibility of a nickel-base weld metal containing 15% Cr is sharply increased as silicon increases from a few tenths of a percent to 1% and higher, whereas levels on the order of 1% Si are relatively innocuous in nickel and nickel-copper alloys.

Iron is commonly present in the nickel-chromium alloys, such as alloy 600, in amounts up to 8 wt%. It is not intended to enhance mechanical properties or weldability, but is added as a ferroalloy, along with other additives, and helps keep the cost of these alloys down. Up to this level, iron is not believed to appreciably effect welding behavior. As the amount of iron increases to the composition range of nickel-chromium-iron alloys (such as alloy 800, which contains about 46% Fe), the alloy becomes increasingly hot-crack sensitive. When steel is melted into a high-nickel-content alloy weld, such as in an overlay or in dissimilar metal welding, there is some reason to believe that residual elements such as sulfur and phosphorus, which are normally present in more abundant quantities in steel than in nickel-base alloys, are more damaging to the properties of the weld than the iron itself.

Carbon is usually present in nickel and high-nickel-content weld metals in amounts ranging from 0.01 to 0.15 or 0.20%. The effect of carbon is a major concern with commercially pure nickel. The limit of carbon solubility in the temperature range from 371 to 649 °C (700 to 1200 °F) is about 0.02 to 0.03%. During welding, any graphitic carbon present in the HAZ is dissolved and held in a super-saturated solution as the metal rapidly cools. Subsequent exposure to the temperature range from 371 to 649 °C (700 to 1200 °F) causes carbon to precipitate in the form of intergranular graphite. Unwelded high-carbon commercially pure nickel would be similarly embrittled in this temperature range by carbon migration to the grain boundaries, but it would require a much longer time. Therefore, when the wrought commercially pure alloys are to be used at temperatures above 371 °C (700 °F), a maximum of 0.02% carbon is used, as in alloy 201. In weld metal, titanium is added to tie carbon up in the form of titanium carbide, thus preventing the formation of graphitic carbon.

In the nickel-base alloys, copper alters solubility to the point where graphitic embrittlement is not encountered up to carbon levels of 0.15 to 0.20%. The carbon content in nickel-copper alloys does have an effect on hot-cracking tendencies when iron is present. For example, nickel-copper alloys with over 0.20% C can tolerate only about 5% Fe; hot cracking may result at higher levels. Low-carbon nickel-copper alloys with less than 0.10% C can tolerate up to about 30 wt% Fe.

In nickel-chromium alloys, carbon combines with chromium to form chromium carbides, unless titanium or columbium are present to form their respective carbides. Although chromium depletion may occur in the HAZ of welds, as it does with the iron-base alloys in the absence of titanium or columbium, it is only in extremely corrosive media that corrosion resistance is impaired.

Manganese was considered to be relatively unimportant to welding metallurgy for many years. It is now well established that in both the nickel-chromium and nickel-copper alloys, manganese is beneficial in minimizing fissuring and in preventing hot cracking in welds of heavy sections, as well as in cases where excessive iron dilution may be encountered. Additions of up to 9 wt% Mn have been used in some commercially available welding electrodes.

Magnesium in wrought materials forms a stable high-melting-point magnesium sulfide preferentially to the extremely low-melting-point nickel-nickel sulfide eutectic. In welding, magnesium in the base metal serves to prevent hot cracking in the HAZ.

Silicon may be present in nickel-base alloys in amounts ranging from 0.01% to 4.0%. However, it is generally considered to have an adverse effect on welding because it increases the tendency for hot cracking. Therefore, wrought materials generally have less than 1% Si. The adverse behavior is associated with a eutectic melting that is more damaging to weld metal than to base metal. Pure nickel is most resistant, followed by nickel-copper alloys. Nickel-chromium alloys are the least resistant to the effects of silicon.

Both titanium and aluminum can be added to the weld metal and the base material to promote age hardening. In smaller amounts, the addition of titanium to nickel and nonchromium-bearing nickel-base alloys helps control gas porosity. Aluminum is added to wrought alloys as a deoxidizer. If substantial amounts of titanium and aluminum are added, the sensitivity of the weld metal to hot cracking increases.

Boron amounts that range from 0.003 to 0.10% improve the high-temperature mechanical properties of certain wrought and cast nickel-base alloys. These alloys are usually intended for elevated-temperature service, where grain boundary strengths are enhanced by boron additions. Unfortunately, boron contents of 0.03% or higher have a grossly damaging effect on weldability. At higher boron levels, the weld metal and HAZ are extremely hot-crack sensitive and considerable care must be exercised with respect to heat input and stress levels.

Zirconium acts similarly to boron in nickel-base alloys by promoting hot cracking in the weld metal and HAZ. However, the tolerance for zirconium is somewhat higher than it is for boron.

Sulfur is perhaps the most damaging of all elements that are involved in the welding metallurgy of nickel and high-nickel-content alloys. Sulfur is particularly damaging to nickel because of its extremely limited solid solubility and because of a nickel-nickel sulfide eutectic reaction that occurs at 649 °C (1200 °F).

Sulfur is different from most other elements that adversely affect weldability because it affects both the HAZ and the weld metal. Penetration occurs at the weld HAZ area that is heated above a threshold temperature while being in contact with sulfur-bearing material. The threshold temperature varies from 316 °C (600 °F) for pure nickel to about 649 °C (1200 °F) for chromium-bearing alloys. Magnesium and manganese are alloying additions that are used to control the effects of sulfur. Unusual diligence is required when handling the nickel-base alloys to avoid contamination from sulfur-containing compounds, which can include such commonly used materials as machining lubricants, marking crayons, temperature-indicating sticks, and shop dirt. Figure 4 shows the type of cracking that can occur when a sheet of alloy 200 is wiped with a dirty cloth containing sulfur compounds and then welded.

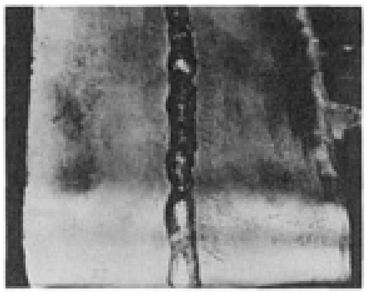

FIG. 4 SULFUR EMBRITTLEMENT OF A JOINT IN ALLOY 200 (UNS N02200). PRIOR TO WELDING, THE CRACKED SIDE OF JOINT WAS WIPED WITH DIRTY CLOTH ON WHICH SULFUR-CONTAINING RESIDUE HAD ACCUMULATED.

Lead causes hot shortness in high-nickel-content weld metal at about the same level as sulfur does. The metallurgical explanation of the subversive effect of lead is very similar to that of sulfur. In practice, lead contamination is less common than sulfur contamination, but probably only because there are fewer common sources of lead-containing contaminants in fabricating shops than there are sources of sulfur-containing contaminants.

Phosphorus can exhibit an effect similar to lead and sulfur in high-nickel-content weld metals. The solubility of phosphorus is quite limited and undergoes a eutectic reaction at about 871 °C (1600 °F). Severe weld metal cracking has been attributed to as little as a few hundredths of a percent of phosphorus.

References cited in this section

1. J.M. SARVER, J.R. CRUM, AND W.L. MANKINS, "CARBIDE PRECIPITATION AND THE EFFECT ON THERMAL TREATMENTS ON THE STRESS CORROSION CRACKING BEHAVIOUR OF INCONEL ALLOY 690," PAPER PRESENTED AT CORROSION '87, NACE, 1987
2. B.B. WEISS, G.E. ROTKE, AND R. STICKLER, PHYSICAL METALLURGY OF HOT DUCTILITY TESTING, *WELD. J.,* OCT 1970, P 471S-422S
3. W.F. SAVAGE AND J.J. PEPE, EFFECTS OF CONSTITUTIONAL LIQUATION IN 18-NI MARAGING STEEL WELDMENTS, *WELD. J.,* SEPT 1967, P 411S-422S
4. D.A. CANONICO, W.F. SAVAGE, W.J. WERNER, AND G.M. GOODWIN, EFFECTS OF MINOR ADDITIONS ON WELDABILITY OF INCOLOY 800, PROC. *WELDING RESEARCH COUNCIL SYMPOSIUM ON EFFECTS OF MINOR ELEMENTS ON THE WELDABILITY OF HIGH-NICKEL ALLOYS,* WELDING RESEARCH COUNCIL, OCT 1967
5. "WELDING TECHNICAL BULLETIN," INCO ALLOYS INTERNATIONAL, JULY 1991, P 29
6. W.F. SAVAGE, WELDMENTS, PHYSICAL METALLURGY AND FAILURE PHENOMENA, *PROC. FIFTH BOLTON LANDING CONF.,* AUG 1978, P 16

Selection of Nickel, Nickel-Copper, Nickel-Chromium, and Nickel-Chromium-Iron Alloys

Donald J. Tillack, D.J. Tillack and Associates

Postweld Heat Treatment

Postweld heat treatments are usually not required for the non-precipitation-hardenable nickel and nickel-base alloy weldments. For example, the ASME boiler and pressure vessel code for unfired pressure vessels does not require postweld heat treatment, except as agreed upon between the user and the manufacturer. However, a dissimilar weldment between a nickel-base alloy and a steel is postweld heat treated because the steel side of the joint requires it. In this case, the heat treatment will normally have little consequence on the nickel-base alloy. Table 3 lists the typical annealing temperatures for many of the solid-solution nickel-base alloys.

TABLE 3 TYPICAL ANNEALING TEMPERATURES FOR CORROSION-RESISTANT NICKEL, NICKEL-COPPER, NICKEL-CHROMIUM, AND NICKEL-CHROMIUM-IRON ALLOYS

All alloys air cooled or water quenched

ALLOY	UNS NO.	TEMPERATURE(A)	
		°C	°F
200	N02200	830-871	1525-1600
201	N02201	830-871	1525-1600
400	N04400	871	1600
600	N06600	925-1040	1700-1900
601	N06601	1175	2150
690	. . .	1040	1900
800	N08800	1010-1175	1850-2150
825	N08825	940	1725

(A) ±14 °C (±25 °F)

In certain environments, it has been beneficial to perform a postweld heat treatment on some of the nickel-base alloy weldments. Alloys 200, 201, and 600 receive this treatment for caustic service applications, as does alloy 400 for hydrofluoric acid service applications. The recommended heat treatment procedures for these materials in these service environments are given in Table 4.

TABLE 4 RECOMMENDED HEAT TREATMENT PROCEDURES FOR SELECTED NICKEL-BASE CORROSION-RESISTANT ALLOYS

ALLOYS 200 AND 201 FOR CAUSTIC SERVICE
1. HOLD AT 704 °C (1300 °F) FOR A MINIMUM OF 0.5 H/25.4 MM (L IN.) OF THICKNESS
2. DETERMINE THE HEATING AND COOLING RATES, WHICH WILL VARY WITH THE SHAPE OF THE PART. COMPLEX SHAPES WITH NONUNIFORM THICKNESSES SHOULD HAVE HEATING AND COOLING RATES OF 111 °C (200 °F)/H. SHAPES WITH UNIFORM THICKNESSES CAN USE FAST HEATING AND AIR COOLING
ALLOY 600 FOR CAUSTIC SERVICE
1. HOLD AT 899 °C (1650 °F) FOR A MINIMUM OF I H/25.4 MM (1 IN.) OF THICKNESS OR AT 788 °C (1450 °F) FOR A MINIMUM OF 4 H/25.4 MM (1 IN.) OF THICKNESS
2. DETERMINE HEATING AND COOLING RATES PER GUIDELINES FOR ALLOYS 200 AND 201
ALLOY 400 FOR HYDROFLUORIC ACID SERVICE
1. HOLD AT 593 °C (1100 °F) FOR A MINIMUM OF 0.5 H/25.4 MM (1 IN.) OF THICKNESS
2. DETERMINE HEATING AND COOLING RATES PER GUIDELINES FOR ALLOYS 200 AND 201

Although determining the rates of heating and cooling after postweld heat treating is seldom a problem with the solid-solution nickel-base alloys, the cooling rate should generally be at least as fast as the rate of an air cool. When there are large differences in cross-sectional areas in a welded structure, the heating and cooling rates may have to be controlled in order to minimize the distortion that might occur with a rapid rate of heating or cooling. Such structures could experience enough distortion to cause fit-up or assembly problems, particularly if a too-rapid cooling rate causes excessive stresses that are due to the mass differences in the structure.

The usual precautions regarding furnace atmosphere and temperature control should be followed when postweld heat treating these alloys. Because cracking caused by contaminants is always a concern, all surfaces should be thoroughly cleaned prior to being postweld heat treated.

Selection of Nickel, Nickel-Copper, Nickel-Chromium, and Nickel-Chromium-Iron Alloys

Donald J. Tillack, D.J. Tillack and Associates

Special Metallurgical Welding Considerations

Dissimilar welding of plate-to-plate, clad, and weld overlay cladding specimens are described below.

Plate-to-Plate. There can be drastic consequences related to complex metallurgical considerations when butt welding plates of different composition. The composition of the weld deposit is controlled not only by the filler metal, but by the amount of dilution from the two base metals. Many different dissimilar-metal combinations are possible, and the amount of dilution varies with the welding process, the operator technique, and the joint design.

In many cases, more than one welding product will satisfy the requirement of metallurgical compatibility. Selection should be based on the strength required, service environment, or the cost of the welding product. Table 5 lists the recommended filler metals to use when welding numerous dissimilar metal combinations.

TABLE 5 WELDING PRODUCTS FOR DISSIMILAR-METAL JOINTS BETWEEN NICKEL-BASE CORROSION-RESISTANT ALLOYS AND OTHER METALS

NICKEL-BASE ALLOY	**WELDING PRODUCT FOR DISSIMILAR-METAL JOINT WITH**									
	STAINLESS STEEL		**CARBON AND LOW-ALLOY STEEL**		**5-9% NI STEEL**		**COPPER**		**NICKEL-COPPER**	
	ELECTRODE	**FILLER METAL**	**ELECTRODE**	**FILLER METAL**	**ELECTRODE**	**FILLER METAL**	**ELECTRODE**	**FILLER METAL**	**ELECTRODE**	**FILLER METAL**
200, 201	ENICRFE-2 (LNCO WELD A)	ERNICR-3 (FM 82)	ENICRFE-2 (LNCO WELD A)	ERNICR-3 (FM 82)	ERNICRFE-2 (LNCO WELD A)	ERNICR-3 (FM 82)	ENICU-7 (WE 190)	ERNICU-7 (FM 60)	ENICU-7 (WE 190)	ERNICU-7 (FM 60)
400	ENICRFE-2 (INCO WELD A)	ERNICR-3 (FM 82)	ENICU-7 (WE 190)	ERNI-1 (FM 61)	ENICU-7 (WE 190)	ERNI-1 (FM 61)	ENICU-7 (WE 190)	ERNICU-7 (FM 60)	ENICU-7 (WE 190)	ERNICU-7 (FM 60)
600, 601	ENICRFE-2 (INCO WELD A)	ERNICR-3 (FM 82)	ENICRFE-2 (INCO WELD A)	ERNICR-3 (FM 82)	ENICRFE-2 (INCO WELD A)	ERNICR-3 (FM 82)	ENI-1 (WE 741)	ERNI-1 (FM 61)	ENI-1 (WE 141)	ERNI-1 (FM 61)
690	ENICRFE-2 (INCO WELD A)	ERNICR-3 (FM 82)	ENICRFE-2 (INCO WELD A)	ERNICR-3 (FM 82)	ENICRFE-2 (INCO WELD A)	ERNICR-3 (FM 82)	ENI-1(A) (WE 141)	ERNI-1(A) (FM 61)	ENI-1(A) (WE 141)	ERNI-1(A) (FM 61)
800	ENICRFE-2 (INCO WELD A)	ERNICR-3 (FM 82)	ENICRFE-2 (INCO WELD A)	ERNICR-3 (FM 82)	ENICRFE-2 (INCO WELD A)	ERNICR-3 (FM 82)	ENI-1 (WE 141)	ERNI-1 (FM 61)	ENI-1 (WE 141)	ERNI-1 (FM 61)
825	ENICRMO-3 (WE 112)	ERNICRMO-3 (FM 625)	ENICRMO-3 (WE 112)	ERNICRMO-3 (FM 625)	(WE 113)	ERNICRMO-3 (FM 625)	ENI-1 (WE 141)	ERNI-1 (FM 61)	ENI-1 (WE 141)	ERNI-1 (FM 61)

(A) PRECAUTION IS REQUIRED TO MINIMIZE COPPER DILUTION IN THE WELD, BECAUSE OF THE HIGH CHROMIUM LEVEL OF ALLOY 690. "BUTTERING" THE COPPER OR COPPER-NICKEL SIDE OF THE JOINT WITH NICKEL WELD METAL PRIOR TO WELDING MAY BE ADVISABLE.

The selection of the welding product involves extra considerations when dissimilar materials are being welded. In addition to dilution factors, the thermal expansion of the two metals being joined and of the filler metal should be reviewed in order to achieve a weld that will not fail prematurely.

In the power industry, for example, an austenitic stainless steel and a low-alloy steel represent a common joint. The expansion rate of the stainless steel side is higher than that of the low-alloy side (see Table 2). If this joint were welded with a stainless steel electrode, then both the weld and the stainless steel side of the joint would expand and contract during heating and cooling at the same rate, thus putting stress on the low-alloy side of the weld. Because the low-alloy steel has less strength than the stainless steel, the shifting of stress to this side of the joint is undesirable and could lead to premature fatigue or stress failure.

On the other hand, if a nickel-base alloy filler metal were used to weld this combination, both the weld and the low-alloy side of the joint would expand and contract together during temperature swings in the service environment and any stresses induced would be shifted to the stainless steel side of the joint, which is stronger than the low-alloy side. This would result in a longer service life for this weld.

Clad Material. Most clad specimens consist of a highly alloyed material, such as a nickel-base alloy, clad to a less-expensive material, usually steel. The nickel-base alloy provides corrosion resistance, whereas the steel, which usually constitutes at least 80% of the total thickness, provides backing strength and reduced costs. Because clad steels are frequently joined by welding, it is important to ensure that any weld joint surface that is exposed to corrosive media is at least as corrosion resistant as the cladding. This requirement influences both joint design and welding technique.

Butt joints should be used whenever possible. Figure 5 shows recommended designs for two thickness ranges. Both designs include a small edge of unbeveled steel above the cladding to protect the cladding when the steel is welded. The steel side should be welded first, using a low-hydrogen welding product. It is important to avoid cladding penetration during the first welding pass. Dilution of the steel weld with the nickel-alloy cladding can cause the deposit to crack. The clad side of the joint should be prepared by grinding or chipping, and it should be welded using the welding product that is recommended for solid sections of the cladding alloy. The weld metal will be diluted with steel. To maintain corrosion resistance, at least two layers should be applied, although three or more layers are preferable.

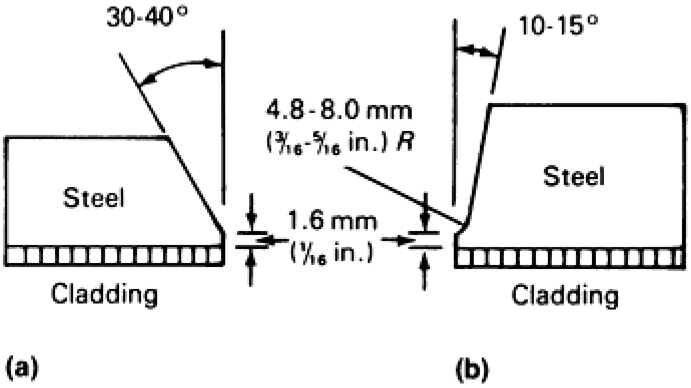

FIG. 5 JOINT DESIGNS FOR CLAD STEEL. (A) MATERIAL OF 4.8 TO 16 MM ($\frac{3}{16}$ TO $\frac{5}{8}$ IN.) THICKNESS. (B) MATERIAL OF 16 TO 25 MM ($\frac{5}{8}$ TO 1 IN.) THICKNESS. SOURCE: REF 7

The strip-back method is sometimes used instead of the procedure described above. The cladding is removed from the vicinity of the joint, as shown in Fig. 6. The remaining steel is then welded using a standard steel joint design and technique, and the nickel-alloy cladding is reapplied by weld overlaying. The advantage of the strip-back method is that it eliminates the possibility of cracking caused by steel weld metal penetration into the cladding (Ref 7).

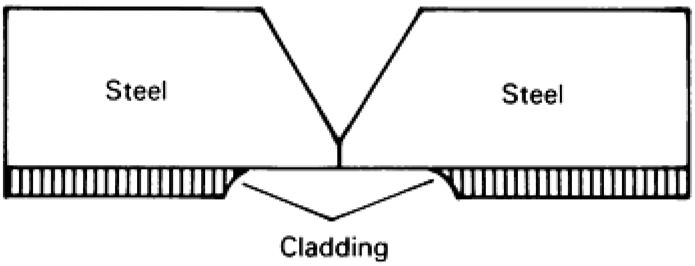

FIG. 6 STRIP-BACK METHOD OF JOINT PREPARATION. SOURCE: REF 7

Some joints, such as those in closed vessels or tubular products, are accessible only from the steel side. In such cases, a standard steel joint design is used, and the cladding at the bottom of the joint is welded first with nickel-base alloy weld metal. After the cladding is welded, the joint can either be completed with nickel-base alloy weld metal or a barrier layer of carbon-free iron can be applied and the joint completed with steel weld metal. If the steel thickness is 8.0 mm ($\frac{5}{16}$ in.) or less, it is generally more economical to complete the joint with nickel-base alloy weld metal.

Weld Overlay Cladding. Nickel-base alloy weld metals are readily applied as overlays on carbon steels, low-alloy steels, and other materials. The cleanliness of the surface to be overlaid is extremely important. All oxides and foreign materials must be removed prior to welding to ensure crack-free results and to minimize the possibility of oxide inclusions in the deposited weld metal. If the material being overlaid has a high sulfur or phosphorus content, then cracking is possible in the deposited nickel-base alloy weld layer. If this occurs, then the cracked layer should be ground or machined away and another layer deposited.

This sacrificial layer approach is just one solution to this situation. Other approaches would be to deposit the first layer with a more-tolerant weld metal composition or, if possible, to design the structure with a lower sulfur/phosphorus base material. Using a welding process and welding parameters that minimize dilution also is very helpful in avoiding problems in this type of situation.

Welding of Castings. Many of the wrought solid-solution nickel alloys discussed in this article have equivalent cast compositions that are modified slightly to enhance castability and properties. The largest modification usually is the increase in silicon content, which unfortunately has an adverse effect on weldability. It is not unusual to find silicon levels that exceed 1.5% in cast nickel-base alloys. Depending on other influences, such as restraint, amount of other minor elements, and the welding process used, this level of silicon may result in weld cracking. When welding surface defects that occurred during the casting process, which is a common welding procedure with castings, most welding methods can be used with most of the solid-solution nickel-base alloy castings.

Repair Welding. The solid-solution nickel-base alloy system does not undergo drastic metallurgical changes during either thermal cycling or exposure to commonly encountered service temperatures, as do many of the ferrous alloys. This is a definite benefit when it becomes necessary to repair weld a component. Except for exposure to surface-modifying environments, such as carburization, nickel-base alloys generally retain their excellent weldability, even after prolonged service exposure. In those cases where a general condition renders a component unusable, such as overall carburization or oxidation attack, very little can be done to repair weld the part. However, if a localized area fails, then there usually is a good chance that a repair weld can salvage the component.

There are occasions when a carburized area is the site of a failure, and if the carburization was not the primary reason for the failure, then sometimes the localized area can be ground free of the carburized layer and a successful repair weld made. If the carburization did-cause the failure, then localized repair welding would probably result in only a short extension of component life, because failure in another area of the carburized surface would undoubtedly occur quickly. This example illustrates how inspection and examination of the component is an important aspect when deciding whether or not to make a weld repair. Once the decision has been made that repair welding is feasible, an area of the component should be ground clean and a test weld made to determine whether the material can be successfully welded. If this test is positive, then the component should be thoroughly prepared by grinding the surface, if necessary, and cleaning it prior to actual repair welding. If groove welding is necessary, then the previously described precautions concerning joint design should be followed in order to provide room for electrode manipulation.

Reference cited in this section

7. "WELDING TECHNICAL BULLETIN," INCO ALLOYS INTERNATIONAL, JULY 1991, P 23

Selection of Nickel, Nickel-Copper, Nickel-Chromium, and Nickel-Chromium-Iron Alloys

Donald J. Tillack, D.J. Tillack and Associates

References

1. J.M. SARVER, J.R. CRUM, AND W.L. MANKINS, "CARBIDE PRECIPITATION AND THE EFFECT ON THERMAL TREATMENTS ON THE STRESS CORROSION CRACKING BEHAVIOUR OF INCONEL ALLOY 690," PAPER PRESENTED AT CORROSION '87, NACE, 1987
2. B.B. WEISS, G.E. ROTKE, AND R. STICKLER, PHYSICAL METALLURGY OF HOT DUCTILITY TESTING, *WELD. J.,* OCT 1970, P 471S-422S
3. W.F. SAVAGE AND J.J. PEPE, EFFECTS OF CONSTITUTIONAL LIQUATION IN 18-NI MARAGING STEEL WELDMENTS, *WELD. J.,* SEPT 1967, P 411S-422S
4. D.A. CANONICO, W.F. SAVAGE, W.J. WERNER, AND G.M. GOODWIN, EFFECTS OF MINOR ADDITIONS ON WELDABILITY OF INCOLOY 800, PROC. *WELDING RESEARCH COUNCIL SYMPOSIUM ON EFFECTS OF MINOR ELEMENTS ON THE WELDABILITY OF HIGH-NICKEL ALLOYS,* WELDING RESEARCH COUNCIL, OCT 1967
5. "WELDING TECHNICAL BULLETIN," INCO ALLOYS INTERNATIONAL, JULY 1991, P 29
6. W.F. SAVAGE, WELDMENTS, PHYSICAL METALLURGY AND FAILURE PHENOMENA, *PROC. FIFTH BOLTON LANDING CONF.,* AUG 1978, P 16
7. "WELDING TECHNICAL BULLETIN," INCO ALLOYS INTERNATIONAL, JULY 1991, P 23

Selection of Nickel-Base Corrosion-Resistant Alloys Containing Molybdenum

S.J. Matthews, Haynes International, Inc.

Introduction

THE FAMILY OF nickel-base corrosion-resistant (CR) alloys containing molybdenum are described in this article both in terms of a historical perspective and an alloy utilization viewpoint. Table 1 gives the nominal chemical compositions of the most widely used materials in this family, but is not all inclusive. Although other molybdenum-bearing nickel-base alloys exist, the information presented below is relative to the joining of similar alloys.

TABLE 1 NOMINAL COMPOSITION OF SELECTED MOLYBDENUM-BEARING NICKEL-BASE CORROSION-RESISTANT ALLOYS

ALLOY TYPE	ALLOY	UNS NO.	COMPOSITION, WT %										
			Ni	Co	Fe	Cr	Mo	W	Mn	Si	C	Cu	OTHERS
NI-MO	B-2	N10665	69	1 MAX	2 MAX	1 MAX	28	. . .	1 MAX	0.1 MAX	0.01 MAX	. . .	. . .
LOW C,	C-4	N06455	65	2	3	16	16	. . .	1	0.08	0.01	. . .	0.7 TI

NI-MO-CR				MAX	MAX				MAX	MAX	MAX		MAX
	C-22	N06022	56	2.5 MAX	3	22	13	3	0.5 MAX	0.08 MAX	0.010	. . .	0.35 V MAX
	C-276	N10276	57	2.5 MAX	5	16	16	4	1	0.08 MAX	0.01 MAX	. . .	0.35 V MAX
	G-3	N06985	44	5 MAX	19.5	22	7	1.5 MAX	1 MAX	1 MAX	0.015 MAX	2.5	0.5 NB MAX
NI-CR-FE-MO-CU	G-30	N06030	43	5 MAX	15	30	5.5	2.5	1.5 MAX	1 MAX	0.03 MAX	2	1.5 NB MAX
NI-MO-CR-FE	N	N10003	71	0.2 MAX	5 MAX	7	16	0.5 MAX	0.8 MAX	1 MAX	0.08 MAX	0.35 MAX	0.5 (AL + TI) MAX

Alloy C Family. Hastelloy alloy C, developed in the early 1930s, is the grandfather of the nickel-chromium-molybdenum family of corrosion-resistant materials. Advances in process metallurgy and corrosion research has prompted the evolution of various C-type alloys. The minimum carbon content of the original alloy C composition was limited to about 0.05% C by the arc melting practices of the time.

The nickel-chromium-molybdenum alloys proved to be extremely important in the chemical processing industry, because they provided corrosion resistance over a wide range of reducing and oxidizing environments. However, grain boundary precipitation of carbides in the as-welded condition required a postweld solution heat treatment to restore optimum corrosion resistance. The deleterious effect of both carbon and silicon on corrosion resistance led to the development of alloy C-276.

An important factor in this new alloy was the development of a reliable quantitative method for determining susceptibility to intergranular attack in the original C-type alloy and the newer C-276 and C-4 alloys (Ref 1). The introduction of the argon-oxygen decarburization (AOD) process in the late 1960s made the production of alloy C-276 commercially viable. The maximum carbon content of alloy C-276 is 0.01% C with typical production compositions of about 0.005% C (Ref 2). Unfortunately, the optimum corrosion resistance of alloy C-276 was found to be hampered by the precipitation of an intermetallic compound rich in molybdenum and tungsten. The alloy is particularly susceptible to μ-phase formation when exposed to temperatures ranging from 650 to 1095 °C (1200 to 2000 °F).

Advances in alloy development theory, specifically the concept of electron vacancy number (N_v), led to the evolution of alloy C-4 in 1975. Alloy C-4 is essentially a tungsten-free, low-carbon, low-silicon ternary alloy of nickel, chromium, and molybdenum. Secondary carbide precipitation was further controlled by the addition of titanium to stabilize residual carbon that was not totally removed by AOD refining. Unfortunately, the removal of tungsten in order to reduce μ-phase precipitation also reduced the resistance of the alloy to localized corrosion attack, namely pitting and crevice corrosion (Ref 2).

A relatively recent (1985) development in the alloy C family is alloy C-22. Because it was recognized that molybdenum and tungsten had an opposing role to chromium in reducing versus oxidizing acid environments, alloy C-22 was developed with the proper balance of both elements to achieve versatility in a wide range of corrosive environments. The reduction of the molybdenum and tungsten content in alloy C-22, versus alloy C-276, also resulted in improved thermal stability (that is, resistance to intermetallic compound formation). An example of an alloy C-type application is in the lining of inlet and outlet ducts that are associated with the flue gas desulfurization (FGD) systems being installed by electric power utility companies.

Alloy B Family. Hastelloy alloy B-2 is unique in that it contains essentially no chromium, an element common to nearly all other CR alloys (Ref 3). It was originally developed to resist hydrochloric acid in all concentrations up to the boiling point and is used in many applications involving the production of this acid, as well as acetic acid and other chemicals. The alloy also resists sulfuric acid and pure phosphoric acid. Without chromium, however, this alloy is vulnerable to corrosion attack in reducing acids when oxidizing salts such as ferric or cupric chloride are present, even in the parts-per-million range.

Not unlike the old alloy C, which required postweld heat treatment to restore optimum corrosion properties, alloy B (the predecessor to alloy B-2) was improved by the evolution of a low-carbon (0.02% C maximum) and low-silicon version (0.08% Si maximum). Alloy B-2 is close to being a commercially "pure" binary alloy of nickel and molybdenum. It has a face-centered cubic (fcc) crystal structure and is quite ductile in the solution-annealed condition (that is, 60% room-temperature tensile elongation). However, the alloy forms intermetallic phases, Ni_3Mo and Ni_4Mo, with modest exposure to the intermediate temperature range from 595 to 850 °C (1100 to 1560 °F). Precipitation of these intermetallic compounds can render alloy B-2 quite brittle. For example, its room-temperature ductility can drop to less than 5% elongation if it is exposed to 760 °C (1400 °F) for a period of 10 min.

A third-generation B-type alloy has recently been developed. Its composition was specially formulated to provide improved thermal stability while retaining the corrosion resistance of alloy B-2. This means that the reaction kinetics that drive the formation of Ni_3Mo and Ni_4Mo have been significantly retarded, greatly improving the fabricability of this family of alloys.

The alloy G family of alloys has excellent resistance to phosphoric acid and has been used in wet-process phosphoric acid evaporators, agitator shafts, pumps, and in the handling of superphosphoric acid (76% P_2O_5). Alloys G-3 and G-30 are newer versions of predecessor alloy G. Alloy G-3 is a lower-carbon version (0.015% C maximum versus 0.05% C maximum) and, hence, has improved resistance to heat-affected zone (HAZ) carbide precipitation. Improved resistance to hot cracking and improved weld metal bend ductility were also achieved by reducing the niobium plus tantalum content from nominally 2.0% (alloy G) to 0.5% maximum (alloy G-3). Alloy G-30 is a high-chromium (30% Cr) alloy designed to possess even greater resistance to HAZ corrosion attack. Recognizing that the phenomenon of sensitization involves not only the type and amount of HAZ precipitation, but also the depleted zones associated with the precipitates, the resistance of alloy G-30 to HAZ corrosion degradation is achieved by higher available chromium, even in the depleted zones. Alloy G-30 resists highly oxidizing acids such as the nitric plus hydrofluoric acid mixtures used in metal pickling baths and in nuclear waste reprocessing.

References

1. M.A. STREICHER, THE RELATIONSHIP OF HEAT TREATMENT AND MICROSTRUCTURE TO CORROSION RESISTANCE IN WROUGHT NI-CR-MO ALLOYS, *CORROSION*, VOL 19, 1963, P 272T-284T
2. M.A. STREICHER, THE EFFECT OF COMPOSITION AND STRUCTURE ON CREVICE, INTERGRANULAR, AND STRESS CORROSION OF SOME WROUGHT NI-CR-MO ALLOYS, *CORROSION*, VOL 32, 1976, P 79-93
3. A.I. ASPHAHANI AND F.G. HODGE, NICKEL-BASE MOLYBDENUM-CONTAINING ALLOYS FOR THE CHEMICAL PROCESS INDUSTRY, *ALLOYS FOR THE EIGHTIES*, R.Q. BARR, ED., CLIMAX MOLYBDENUM CO., 1980

Selection of Nickel-Base Corrosion-Resistant Alloys Containing Molybdenum

S.J. Matthews, Haynes International, Inc.

General Welding Characteristics

Because of similarities in heat capacity, density, and other physical properties, the welding characteristics of the alloys described in this article are quite similar to austenitic stainless steels. The major differences between these alloys and stainless steel are lower thermal expansion, lower thermal conductivity, and higher electrical resistivity. Generally, solidus temperatures are about 55 °C (100 °F) lower than type 304 stainless. Table 2 summarizes the comparative physical properties relevant to welding and joining operations.

TABLE 2 PHYSICAL PROPERTIES OF NICKEL-BASE MOLYBDENUM-BEARING CORROSION-RESISTANT ALLOYS RELATIVE TO CARBON STEEL AND STAINLESS STEELS

MATERIAL	MELTING TEMPERATURE RANGE	DENSITY	THERMAL CONDUCTIVITY	COEFFICIENT OF THERMAL EXPANSION	ELECTRICAL RESISTIVITY, $10^{-8}\Omega \cdot M$

	°C	°F	g/cm^3	lb/in.^{3}	w/m · k	btu in./ft^2 · h · °f	10^{-6}/K	10^{-6}/ °F	
CARBON STEEL	1470-1530	2670-2785	7.9	0.285	0.46	3.2	11.7	6.51	19
AUSTENITIC STAINLESS STEEL	1390-1450	2535-2640	8.0	0.289	0.16	1.1	17.0	9.45	72
NICKEL-BASE CORROSION-RESISTANT ALLOYS									
ALLOY B-2	1350-1390	2460-2535	9.2	0.332	0.11	0.76	10.3	5.73	137
ALLOY C-22	1350-1390	2460-2535	8.7	0.314	0.10	0.69	12.4	6.89	114
ALLOY G-30	1300-1350	2370-2460	8.2	0.296	0.10	0.69	12.8	7.12	116

Welding Processes. Gas-tungsten arc welding (GTAW), gas-metal arc welding (GMAW), and shielded metal arc welding (SMAW) processes are commonly used to join this family of CR alloys.

A filler metal that matches the composition of the base material is usually recommended when welding the nickel-chromium-molybdenum CR alloys. Other processes, such as plasma arc welding (PAW), resistance spot welding (RSW), laser-beam welding (LBW), electron-beam welding (EBW), and friction welding, can be used. The plasma arc cutting process is commonly used to cut alloy plate into desired shapes and to prepare weld angles. Oxyacetylene cutting is not recommended, nor is the use of oxyacetylene welding (OAW) and submerged arc welding (SAW). Oxyacetylene welding is not recommended because of the possibility of carbon pickup from the flame, whereas submerged arc welding is not recommended because of its high heat input, chromium loss across the arc, and silicon pickup from the welding flux.

Generally, welding heat input is controlled in the low-to-moderate range. Wide weave beads are not recommended. Stringer bead welding techniques, with some electrode manipulation, are preferred. The nickel-chromium-molybdenum alloy weld metal is not as fluid as carbon steel and does not flow out as readily to "wet" the sidewalls. Therefore, the welding arc and filler metal must be manipulated so as to place the molten metal where it is needed. In addition to sluggishness, the penetration pattern of this alloy type is less than that of a typical carbon or stainless steel weld, and incomplete fusion is more likely to occur. Therefore, care must be taken to ensure that the groove opening is wide enough to allow proper torch or electrode manipulation and proper placement of the weld bead.

Cleaning, Edge Preparation, and Fit-Up. Proper penetration of the weld angles is a very important part of welding these nickel-base CR alloys. It is necessary to condition all thermal cut edges to bright, shiny metal prior to welding. This is particularly important if air arc gouging is being used, because of the possibility of carbon pickup from the carbon electrode.

In addition to the weld angle, a 25 mm (1 in.) wide band on both the top and bottom (face and root) surface of the weld zone should be conditioned to bright metal with about an 80-grit flapper wheel or disk. This procedure is particularly important for the SMAW of alloy B-2. If the mill scale is not removed, then the B-2 alloy welding flux can interact with it and cause cracking at the toe of the weld in the base material.

The welding surface and adjacent regions should be thoroughly cleaned with an appropriate solvent prior to any welding operation. All greases, oils, cutting oils, crayon marks, machining solutions, corrosion products, paints, scale, dye-penetrant solutions, and other foreign matter should be completely removed.

Stainless steel wire brushing is normally sufficient for interpass cleaning of GTAW and GMAW weldments. The grinding of starts and stops is recommended for all fusion welding processes. When the GMAW process is used, light grinding may be necessary between passes prior to wire brushing, depending on the composition of the shielding gas. Slag removal during SMAW will require chipping and grinding, followed by wire brushing.

Preheating the nickel-molybdenum and nickel-chromium-molybdenum alloys is not required. A preheat is generally specified as room temperature (typical shop conditions). Interpass temperature should be maintained below 95 °C (200 °F). Auxiliary cooling methods can be used to control the interpass temperature. Water quenching is acceptable. However, care must be taken to avoid contaminating the weld zone with traces of oil from shop air lines, grit/dirt from soiled water-soaked rags, or mineral deposits from hard water used to cool the weld joint. The safest way to maintain a low interpass temperature is to allow the assembly to cool naturally. When attaching hardware to the outside of a thin-walled vessel, it is good practice to provide auxiliary cooling to the inside (process side) to minimize the extent of the HAZ.

Selection of Nickel-Base Corrosion-Resistant Alloys Containing Molybdenum

S.J. Matthews, Haynes International, Inc.

Welding Metallurgy

Heat-Affected Zone. The phenomenon of HAZ grain boundary sensitization is addressed in the article "Selection of Austenitic Stainless Steels" in this Volume. However, it should be noted here that the solid solubility of carbon is considerably lower in nickel-base alloys than in stainless steels. The net effect is that during the welding of low-carbon nickel-molybdenum and nickel-chromium-molybdenum CR alloys, HAZ grain boundary precipitation is still a potential reality, despite the relatively low carbon content of these alloys. The amount and severity of precipitation will depend on the cooling rate through the intermediate temperature range from 1000 to 600 °C (1830 to 1110 °F). Some heat-to-heat variations in grain boundary precipitation have also been observed during simulated HAZ studies in alloy B-2 (Ref 4). The adverse effects of minor grain boundary precipitation depend on the severity of the corrosive environment. For example, the maximum corrosion penetration after testing in boiling 10% hydrochloric acid (HCl) varied little for simulated heat-affected zones in alloy B-2. However, when harsher autoclave testing environments were used (20% HCl at 150 °C, or 300 °F), penetration by corrosion increased as the simulated HAZ cooling rate decreased from 20 to 120 s through the critical intermediate temperature range (Ref 4). Despite these laboratory findings, most low-carbon nickel-molybdenum and nickel-chromium-molybdenum alloys have been put into corrosion service in the as-welded condition with good results. Occasionally, poor results are encountered. Metallographic examination of the corroded part will usually reveal evidence of grain boundary precipitation in the heat-affected zone (Fig. 1).

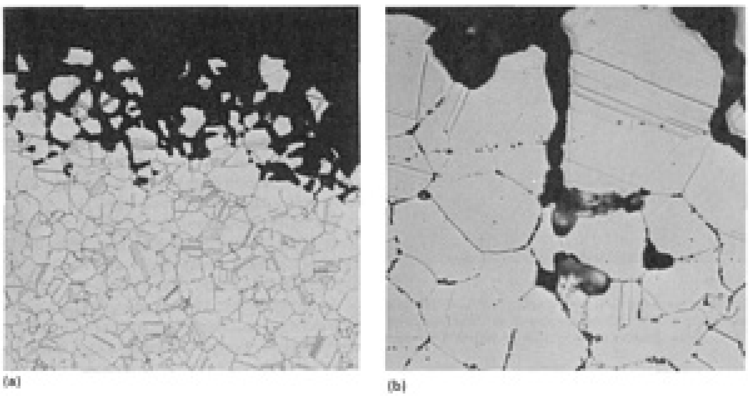

FIG. 1 PHOTOMICROGRAPHS SHOWING CORROSION ATTACK IN HASTELLOY C-276 (UNS N10276) CAUSED BY GRAIN BOUNDARY PRECIPITATION IN THE HEAT-AFFECTED ZONE OF THE WELD. THE SAMPLE WAS TAKEN FROM A PIPE REMOVED AFTER 18 MONTHS OF SERVICE IN A HYDROCHLORIC ACID VAPOR ENVIRONMENT IN A CHEMICAL PLANT. SAMPLE WAS ETCHED IN HCL AND OXALIC ACID. (A) 75×. (B) 375×

Aside from the possibility of grain growth, few other metallurgical HAZ problems have been reported for these alloys. HAZ hot cracking, especially of the type created by constitutional liquation of primary carbide particles, is not a problem with these alloys.

Fusion zone welding metallurgy of the nickel-molybdenum and nickel-chromium-molybdenum alloy family is important, because base materials are usually welded with filler metals of matching composition and because these alloys can be welded autogeneously (no filler added), as in the case of welded tubular products. Three issues should be considered in terms of the fusion zone:

- EFFECT OF WELD METAL SEGREGATION ON CORROSION RESISTANCE
- PROPENSITY TO FORM POROSITY
- SOLIDIFICATION HOT-CRACK SENSITIVITY

Segregation. Because of the segregation of solute elements upon solidification (principally molybdenum, which segregates to the cellular dendritic boundaries of the fusion zone), it is generally accepted that the corrosion resistance of the weld metal will be marginally less than that of the more homogeneous wrought base material. Differences in corrosion performance, however, depend heavily on the severity of the corrosion environment. In mild corrosion environments, little difference is observed. In severe environments, the fusion zone may be preferentially attacked, as illustrated in Fig. 2, a cross section of an as-welded alloy B-2 coupon exposed for about 1 year in a chemical plant process stream. Figure 3(a) is a high-magnification view of the surface attack on alloy B-2 weld metal when tested for 96 h in 20% HCl at 150 °C (300 °F) in an autoclave. Under these conditions, corrosion attack appears to be the most aggressive at the dendrite cores (molybdenum lean). Figure 3(b) shows the lack of corrosion attack on weld metal from a specially devised overalloyed B-type filler material (Ref 5). This alloy consists of 42% Mo, rather than the nominal 28% Mo. The higher-molybdenum nickel alloy solidifies as a nonequilibrium hypoeutectic structure consisting of a metastable α phase, with Ni_3Mo and NiMo in the interdendritic regions. In this case, despite solidification segregation, the dendrite core regions are sufficiently saturated with molybdenum to resist preferential fusion zone attack.

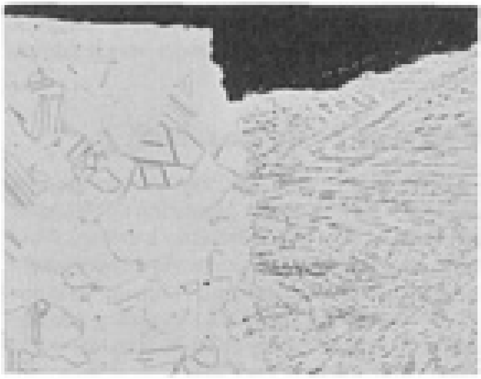

FIG. 2 METALLOGRAPHIC CROSS SECTION SHOWING PREFERENTIAL FUSION ZONE ATTACK IN HASTELLOY B-2 (UNS N10665). SAMPLE WAS A WELDED COUPON PLACED IN A CHEMICAL PLANT PROCESS STREAM FOR APPROXIMATELY 1 YEAR. HYDROCHLORIC AND CHROMIC ACID ETCH. 75×

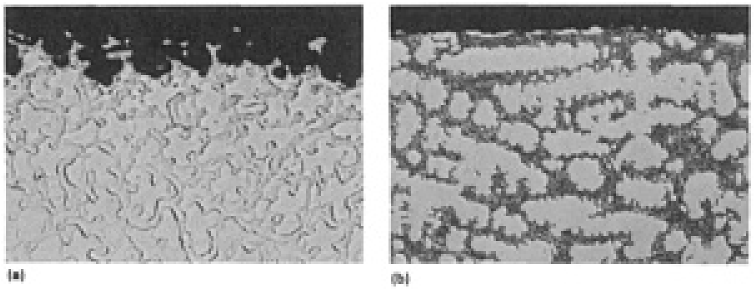

FIG. 3 METALLOGRAPHIC CROSS SECTIONS OF THE CORRODED SURFACE OF WELD DEPOSITS ON HASTELLOY B-2 (UNS N10665) AFTER TESTING IN AN AUTOCLAVE AT 150 °C (300 °F) FOR 96 H IN A 20% HCL ENVIRONMENT. (A) NI-28MO (UNS N10665) ALLOY FILLER METAL. 375×. (B) NI-42MO FILLER METAL. 375×. BOTH SAMPLES WERE ETCHED IN HYDROCHLORIC AND CHROMIC ACID. SEE TEXT FOR EXPLANATION.

It should be noted that although the microstructure of the Ni-42Mo material in Fig. 3(b) offers high corrosion resistance, the alloy is characterized by limited weld ductility (Ref 6). However, the use of overalloyed filler metals as a solution to preferential weld metal corrosion attack has been demonstrated elsewhere. For example, alloy C-22 (22% Cr) has been used successfully to refurbish corroded welds in an alloy C-276 (16% Cr) pulp and paper mill bleach-mixing device. Alloy G-30 (30% Cr) is also a logical choice as an overalloyed filler for certain alloy G-3 components.

Porosity. Recognizing that the solubility of gas is greater in the liquid state than in the solid state, the subject of weld porosity is relevant to fusion zone discussion. The three most logical origins of gas-generated porosity are carbon monoxide, hydrogen, and nitrogen. However, in welding the low-carbon nickel-molybdenum and nickel-chromium-molybdenum alloys, the cause of porosity is not believed to be from the formation of carbon monoxide. As long as there is a residual silicon or aluminum content in the weld deposit, the oxygen potential is too low for the formation of carbon monoxide. In order to produce a carbon monoxide "carbon boil" during welding, it would be necessary first to deplete all the aluminum, silicon, and any other deoxidizers that may be present.

Gas hole formation can be caused by hydrogen evolution. As is well known, even the smallest amount of water will be reduced by metal deoxidizers, releasing a sufficient volume of gas to create porosity. Hydrocarbons can also dissociate and go into solution. The lower the carbon content of the alloy, the more likely it is to pick up hydrogen from hydrocarbons. These reactions are favored as the carbon content of the alloy decreases. Gas hole formation that is due to hydrogen can be minimized by the well-known methods of keeping the joint area and the weld filler metal dry and oil-free. This approach is especially true when welding with coated electrodes. The ceramic coating ingredients can absorb sufficient moisture in humid weather to cause potential problems. It is because of this behavior that unused electrodes should be stored in a warm oven (110 °C, or 230 °F, minimum).

The third and only other possible cause of gas hole formation is nitrogen. As is well documented in the literature, a large quantity of nitrogen is capable of being picked up by the molten metal during arc welding processes whenever the gas in the arc contains any nitrogen. Nitrogen pickup from air, for example, might prove to be a problem whenever the weld pool is improperly shielded. Different alloys would be expected to have different susceptibilities to nitrogen-induced porosity. The ability of the solidified matrix plus secondary solid phases to hold or carry nitrogen is the primary difference between alloys. Alloys with a greater nitrogen-carrying capacity are apparently less susceptible to nitrogen-induced porosity. For example, a nickel-chromium alloy would be less susceptible to nitrogen porosity than pure nickel because of the increased solubility of nitrogen. Furthermore, because alloy B-2 contains no chromium, it possesses a lower nitrogen solubility than alloy C-22, and, thus, would be more prone to nitrogen porosity.

Hot-Crack Sensitivity. The final fusion zone issue is hot-crack sensitivity. Hot cracking of C-type alloys during solidification has been studied in depth by researchers equipped with sophisticated microanalysis equipment (Ref 7). Solidification hot cracking, as generated by the varestraint hot-cracking device, was found to increase as the tendency to form intermetallic secondary solidification constituents increased. In terms of general weldability ranking, alloy C-4 was found to be the least sensitive to solidification cracking, followed by alloy C-22, and then alloy C-276. It should be noted that all of the low-carbon nickel-chromium-molybdenum alloys of the C family are characterized by high weld metal ductility. Hot-cracking problems have rarely been reported under actual fabrication conditions.

References cited in this section

4. S.J. MATTHEWS, SIMULATED HEAT-AFFECTED ZONE STUDIES OF HASTELLOY ALLOY B-2, *WELD. J. RES. SUPPL.,* MARCH 1979
5. A.I. ASPHAHANI AND S.J. MATTHEWS, "HIGH MOLYBDENUM NICKEL-BASE ALLOY," U.S. PATENT 4,846,885, 11 JULY 1989
6. E.M. HORN, P. MATTERN, H. SCHLECKE, J. KORKHAUS, A.I. ASPHAHANI, AND S.J. MATTHEWS, HASTELLOY ALLOY B-42 WELD OVERLAY WELDING, *WERKST. KORROS.,* VOL 43, 1992, P 396-401
7. M.J. CIESLAK, T.J. HEADLEY, AND J. ROMIG, THE WELDING METALLURGY OF HASTELLOY ALLOYS C-4, C-22, AND C-276, *METALL. TRANS. A,* VOL 17A, NOV 1986, P 2035-2047

Selection of Nickel-Base Corrosion-Resistant Alloys Containing Molybdenum

S.J. Matthews, Haynes International, Inc.

Postweld Heat Treatment

Because of their very low carbon contents (<0.02% C), the newer-generation CR alloys do not require heat treatment after welding. However, for certain critical corrosion environments, proper postweld annealing will maximize corrosion resistance of the HAZ. Annealing can also be performed to reduce residual stresses, which play a role in the susceptibility to stress-corrosion cracking in some environments. Annealing can be used to increase the corrosion resistance of the weld fusion zone by promoting diffusion of some elements.

Corrosion-resistant alloys are usually supplied in the mill-annealed condition, unless otherwise specified. The mill-annealing procedure has been designed to place the material in the optimum condition with respect to ductility and corrosion resistance. During fabrication of a vessel or component, several operations or conditions may require an anneal. For example, following any hot-forming operation, a reanneal of the material should be done to restore ductility and corrosion properties. Annealing is also performed after cold working to restore ductility. Generally, annealing is not required if the total cold work is below 7% outer fiber elongation. In general, the only heat treatment that is acceptable for these alloys is a full solution anneal. Table 3 lists the proper heat-treating temperatures and type of quench. Intermediate temperatures commonly used for stress-relieving steels, stainless steels, and other alloys are not effective and could promote precipitation that would be detrimental to corrosion resistance.

TABLE 3 SOLUTION ANNEALING TEMPERATURES FOR CORROSION-RESISTANT NICKEL-BASE ALLOYS CONTAINING MOLYBDENUM

Data are applicable for either water quench (WQ) or rapid air cool (RAC) quench.

ALLOY	**UNS NO.**	**TEMPERATURE**	
		°C ± 14°	**°F ± 25°**
B-2	N10665	1065	1950
C-4	N06455	1065	1950
C-22	N06022	1120	2050
C-276	N10276	1120	2150
G-3	N06905	1150	2100
G-30	N06030	1175	2150
N	N10003	1175	2150

Before heat treatment, grease, graphite, and other foreign materials must be removed from all surfaces. Tube products may require special care for complete removal of these contaminants from interior surfaces. Carburization of these materials at the heat-treating temperatures can reduce corrosion resistance, and sulfur embrittlement can cause severe cracking. Proper control of the temperature and time cycle are also critical. A flexible set of rules governing soaking-annealing time is generally followed, because of the many variations in furnace type, furnace operation, facilities for loading and unloading the furnace, and so on. The temperature should be monitored using thermocouples attached to the piece being annealed. The actual holding time should begin to be measured when the entire section is at the specified annealing temperature. It is important to remember that the center of a section does not reach the solutioning temperature as soon as the surface.

Normally, hold time should be in the range from 10 to 30 min, depending on section thickness. Thin-sheet components are held at the shorter time, whereas heavier sections are held at the longer time. Shorter annealing times can be used if the sole purpose is to remove the effect of cold work. Longer annealing times might be required if the objective is to promote homogenization of the fusion zone.

Rapid cooling is essential after solution heat treatment to prevent the precipitation of secondary phases and the resultant lowering of the corrosion resistance of these alloys. Water quenching is recommended on material thicker than 9.5 mm

($\frac{3}{8}$ in.). Although rapid air cooling can be used on sections thinner than 9.5 mm ($\frac{3}{8}$ in.), water quenching is preferred. The time from the furnace to the quench tank or to the start of rapid air cooling must be as short as possible (less than 3 min).

The requirements for proper annealing of alloy B-2 (UNS 10665) are quite demanding. In addition to the guidelines established above, the following factors should be considered very carefully. It is strongly recommended that the part being heat treated be charged into a furnace that is at the annealing temperature, and that the heating rate be as fast as possible. In addition, the thermal capacity of the furnace should be large to allow the temperature of the furnace to recover quickly after the part is charged into it. These steps are designed to shorten the time in the intermediate temperature range from 595 to 815 °C (1100 to 1500 °F). Extended exposure (greater than 15 min) at these intermediate temperatures can cause intergranular cracking because of phase transformations, especially if the part is in the cold-worked condition. Under certain conditions, shot peening prior to heat treatment may be beneficial to the elimination of residual tensile stresses. For example, it has been observed that shot peening the knuckle radius and straight flange regions of a cold-formed head prior to heat treatment can help reduce intermediate temperature intergranular cracking by lowering the residual stress patterns at the surface of the cold-formed component.

Other Welding Situations

Welding of dissimilar metals to nickel-base CR alloys does not present a problem in most cases. The nature of nickel as a "forgiving" solvent element renders most industrially encountered dissimilar metal combinations as metallurgically compatible. Filler materials designed to join the nickel-base side of the joint are usually recommended.

When joining a nickel-base alloy to a ferrous alloy (for example, mild steel or a stainless steel), electrodes of the nickel-base alloy should be used as filler metal, rather than steel or stainless electrodes, because excess superheat caused by melting temperature differences (Table 2) will tend to aggravate the HAZ on the nickel-base side of the joint.

The joining of clad plate (that is, steel plate that has been explosion clad or roll clad with a thinner nickel-base CR alloy) has been successfully performed. However, proper welding procedures need to be developed to ensure minimum iron dilution in the weld zone facing the corrosion environment.

The welding of cast CR alloys presents no unusual problems, as long as the integrity of the casting is sound (that is, free from shrinkage cavity and inclusions). Welding is often used to repair casting defects. Generally, the defective area is thoroughly removed by grinding, which eventually forms a cavity. The prepared cavity should be dye-penetrant inspected to ensure that all objectionable defects have been removed and then thoroughly cleaned prior to welding repair. Because these alloys have low-penetration characteristics, the ground cavity must be broad enough and have sufficient sidewall clearance in the weld groove to allow weld rod/weld bead manipulation. It is not recommended to heal cracks or wash out defects by autogeneously remelting weld beads or by depositing additional filler metal over the defect.

Repair welding of wrought alloy components and structures is accomplished using a similar technique. The good thermal stability of low-carbon CR alloys is such that solution annealing of material that has been in corrosion service for many years is generally not required prior to repair welding. In these situations, the repair is usually required because of accelerated corrosion of a localized area. The selection of higher-alloyed filler metals for such repairs should be considered.

References

1. M.A. STREICHER, THE RELATIONSHIP OF HEAT TREATMENT AND MICROSTRUCTURE TO CORROSION RESISTANCE IN WROUGHT NI-CR-MO ALLOYS, *CORROSION,* VOL 19, 1963, P 272T-284T
2. M.A. STREICHER, THE EFFECT OF COMPOSITION AND STRUCTURE ON CREVICE, INTERGRANULAR, AND STRESS CORROSION OF SOME WROUGHT NI-CR-MO ALLOYS, *CORROSION,* VOL 32, 1976, P 79-93
3. A.I. ASPHAHANI AND F.G. HODGE, NICKEL-BASE MOLYBDENUM-CONTAINING ALLOYS FOR THE CHEMICAL PROCESS INDUSTRY, *ALLOYS FOR THE EIGHTIES,* R.Q. BARR, ED., CLIMAX MOLYBDENUM CO., 1980
4. S.J. MATTHEWS, SIMULATED HEAT-AFFECTED ZONE STUDIES OF HASTELLOY ALLOY B-2, *WELD. J. RES. SUPPL.,* MARCH 1979
5. A.I. ASPHAHANI AND S.J. MATTHEWS, "HIGH MOLYBDENUM NICKEL-BASE ALLOY," U.S. PATENT 4,846,885, 11 JULY 1989
6. E.M. HORN, P. MATTERN, H. SCHLECKE, J. KORKHAUS, A.I. ASPHAHANI, AND S.J. MATTHEWS, HASTELLOY ALLOY B-42 WELD OVERLAY WELDING, *WERKST. KORROS.,* VOL 43, 1992, P 396-401
7. M.J. CIESLAK, T.J. HEADLEY, AND J. ROMIG, THE WELDING METALLURGY OF HASTELLOY ALLOYS C-4, C-22, AND C-276, *METALL. TRANS. A,* VOL 17A, NOV 1986, P 2035-2047

Selection of Nickel-Base Corrosion-Resistant Alloys Containing Molybdenum

S.J. Matthews, Haynes International, Inc.

Selected References

- P.E. MANNING, J.D. SMITH, AND J.L. NICKERSON, NEW VERSATILE ALLOYS FOR THE CHEMICAL PROCESS INDUSTRY, *MATER. SELECT. DES.,* JUNE 1988, P 67-73

Selection of Cobalt-, Titanium-, Zirconium-, and Tantalum-Base Corrosion-Resistant Alloys

S.J. Matthews, Haynes International, Inc.

Introduction

A GREAT VARIETY of nonferrous corrosion-resistant (CR) alloys are available in the market. The common nickel-base corrosion-resistant alloys are covered in the preceding articles in this Section of the Handbook. This article will provide information about non-nickel-base alloys; however, no attempt will be made here to cover all such materials. The interested reader is referred to *Corrosion,* Volume 13 of the *ASM Handbook,* for information on the corrosion of specific alloy systems. The four corrosion-resistant materials of commercial significance that are briefly addressed here are based on:

- COBALT
- COMMERCIALLY PURE TITANIUM
- ZIRCONIUM
- TANTALUM

Selection of Cobalt-, Titanium-, Zirconium-, and Tantalum-Base Corrosion-Resistant Alloys

S.J. Matthews, Haynes International, Inc.

Cobalt-Base CR Alloys

Cobalt behaves much like nickel in terms of active-passive behavior in aqueous corrosion systems. Chromium additions reportedly are more effective in imparting passivity to cobalt than to nickel (Ref 1). Cobalt-chromium CR alloys typically contain sufficient additions of nickel or iron, which serve to stabilize the face-centered cubic (fcc) crystal structure and increase the stacking, fault energy of the fcc phase. Molybdenum and tungsten additions act as solid-solution strengtheners and enhance the corrosion resistance of cobalt-base alloys in nonoxidizing media. Cobalt-base CR alloys have been used in the pulp and paper, agrichemical, and chemical process industries in applications such as valves, pumps, nozzles, screw conveyors, and other components that are characterized by harsh environments aggravated by both corrosion and wear. Electrogalvanizing rolls are an example of a component exposed to both of these conditions. Cobalt-base CR alloys represent the premium materials for this application. The nominal composition of some cobalt-base corrosion-resistant alloys are given in Table 1.

TABLE 1 NOMINAL COMPOSITION OF SELECTED COBALT-BASE CORROSION-RESISTANT ALLOYS

ALLOY	UNS NO.	COMPOSITION, WT %							
		CO	Cr	Ni	Fe	Mo	W	C	N
ULTIMET	R31233	BAL	26	9	3	5	2	0.06	0.08
ALLOY 25	R30605	BAL	20	10	3 MAX	. . .	15	0.10	. . .
ALLOY 150	. . .	BAL	28	. . .	21	. . .	. . .	0.05 MAX	. . .
MP35-N MULTIPHASE	R30035	BAL	20	35	. . .	10	. . .	. . .	. . .

Weldability Characteristics. The fusion-welding characteristics of fcc-stabilized cobalt-base alloys are virtually identical to those of nickel-base alloys, because of similarities in melting range and physical properties. The only welding procedural idiosyncrasy of cobalt-base alloys relative to nickel is the potential for heat-affected zone (HAZ) cracking caused by copper surface contamination. Even minuscule amounts of copper inadvertently transferred to the surface of a sheet or plate will cause cracking if the contaminated surface corresponds to a HAZ of a subsequent weld (Ref 2). Cracking is caused by liquid-metal embrittlement (LME). Copper and cobalt are relatively insoluble in each other and form a classic LME couple (Ref 3). Nickel and copper exhibit 100% solid solubility; hence, nickel-base alloys are not prone to copper contamination cracking in the HAZ. Copper jigs and fixtures, especially those used in sheet welding, are a potential source of copper contamination. Chromium or nickel plating of copper fixtures will minimize the potential for inadvertent copper transfer when welding cobalt alloys.

Cobalt-base CR alloys are generally welded using a filler-metal composition that matches the composition of the base material. Like low-carbon nickel-base CR alloys, cobalt alloys have relatively good resistance to fusion-zone (FZ) solidification cracking. The lower the carbon content, the greater the resistance to hot cracking. Boron is particularly deleterious as a residual trace element and can cause welding problems.

Cobalt-chromium CR alloys with relatively low nickel or iron content (for example, UNS alloy R31233, Table 1) exhibit unique mechanical property characteristics in the as-welded condition. Weldments are characterized by high strength and low ductility, because of stress-induced localized transformation from a fcc crystal structure to a hexagonal close-packed (hcp) structure (Ref 4). The transformation products are not visible in optical photomicrographs (Fig. 1). Instead, thin-film electron microscopy is required to detect the hcp phase.

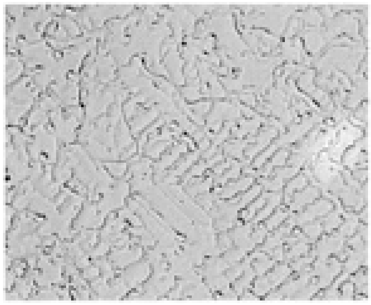

FIG. 1 TYPICAL WELD METAL MICROSTRUCTURE OF ULTIMET ALLOY (UNS R31233) OBTAINED WITH GMAW DEPOSIT WHEN SHORT-CIRCUITING MODE OF TRANSFER IS USED. SAMPLE WAS ETCHED WITH HCL AND OXALIC ACID. 500×

Tensile data comparing room-temperature elongation of weldments versus castings and annealed base material are given in Table 2. Castings generally have less residual stress in the as-cast condition than do weldments; hence, the observed difference in ductility. Table 2 also shows that a full postweld anneal at 1120 °C (2050 °F) will reduce the yield strength and improve the ductility of a welded structure. A postweld heat treatment is recommended, especially if the cold forming of a weldment after welding is a required part of the fabrication sequence.

TABLE 2 TENSILE PROPERTIES OF WELDED COBALT-BASE CORROSION-RESISTANT ALLOY UNS R31233 RELATIVE TO WROUGHT AND CAST PRODUCTS AT ROOM TEMPERATURE

PRODUCT FORM	CONDITION	YIELD STRENGTH AT 0.2% OFFSET		ULTIMATE TENSILE STRENGTH		ELONGATION, %
		MPa	ksi	MPa	ksi	
PLATE	ANNEALED	545	79	1020	148	36
CASTINGS$^{(A)}$	AS CAST	395	57	770	112	35
WELD METAL$^{(B)}$	AS WELDED	585	85	850	123	18
	ANNEALED (1120 °C, OR 2050 °F)	460	67	825	120	22

Source: Ref 4

(A) INVESTMENT CASTING.

(B) GMAW SPRAY TRANSFER.

References cited in this section

1. H.H. UHLIG AND A.I. ASPHAHANI, CORROSION BEHAVIOR OF COBALT BASE ALLOYS IN AQUEOUS MEDIA, *MATER. PERFORM.*, VOL 18 (NO. 11), NOV 1979
2. S.J. MATTHEWS, M.O. MADDOCK, AND W.F. SAVAGE, HOW COPPER SURFACE CONTAMINATION AFFECTS WELDABILITY OF COBALT SUPERALLOYS, *WELD. J.*, MAY 1972
3. W.F. SAVAGE AND M. MUSHALA, COPPER CONTAMINATION CRACKING IN THE WELD HEAT AFFECTED ZONE, *WELD. J.*, MAY 1978, P 145
4. S.J. MATTHEWS, P. CROOK, L.H. FLASCHE, AND J.W. TACKETT, WELDABILITY CHARACTERISTICS OF A NEW CORROSION- AND WEAR-RESISTANT COBALT ALLOY, *WELD. J.*, DEC 1991, P 331-337

Titanium-Base CR Alloys

Commercially pure (CP) titanium offers excellent corrosion resistance in most environments, except those media that contain fluoride ions. Titanium is more corrosion resistant than stainless steel in many industrial environments, and its use in the chemical process industry has been gradually increasing (Ref 5). Titanium has a density approximately one-half that of nickel- or cobalt-base alloys, and its cost effectiveness relative to these alloys is generally calculated using the cost per unit weight, the difference in density, the corrosion allowances, and other performance factors.

Titanium has moderate strength, good ductility, and excellent weldability relative to the more complex aerospace α-β titanium alloys. Various grades of CP titanium are classified on the basis of mechanical properties and interstitial impurities. In general, variations in strength are produced by differences in impurity content, primarily the interstitial elements (that is, oxygen, nitrogen, and carbon) and iron. As the content of these elements increases, strength also increases.

Titanium that is intentionally alloyed with trace amounts of palladium or molybdenum and nickel is also used to provide corrosion-resistant service. For example, the Ti-0.2Pd alloy and the Ti-0.3Mo-0.8Ni alloy are characterized by improved crevice corrosion resistance, as well as improved corrosion resistance in reducing environments.

Weldability Characteristics. Corrosion-resistant titanium materials are generally welded with matching filler-metal compositions. Welded joints will have essentially the same corrosion resistance as the base metal. However, any iron contamination of the weld metal that exceeds even 0.05% Fe will cause accelerated corrosion attack in nitric acid or in other strongly oxidizing solutions. Therefore, it is imperative that all potential sources of iron contamination be avoided. Steel wire brushes should not be used for interpass cleaning of titanium weldments.

Titanium is a single (α) phase material with a hcp crystal structure. However, an allotropic transformation occurs at temperatures above approximately 870 °C (1600 °F), where α transforms to a body-centered cubic (bcc) β phase. Accordingly, the heat-affected zone of CP titanium weldments will be characterized by large β grains, which form at high temperatures during welding but transform back to an acicular and serrated α phase upon cooling to room temperature.

Commercially pure titanium alloys are not heat treatable, that is, they do not develop significant differences in properties after heat treatment. However, a postweld anneal in the temperature range from 650 to 760 °C (1200 to 1400 °F) or a postweld stress relief from 480 to 595 °C (900 to 1100 °F) might be warranted for purposes of maintaining dimensional stability during subsequent machining operations. A postweld heat treatment may also improve resistance to stress-corrosion cracking (SCC), depending on the configuration of the weldment and the intended corrosion-resistant service.

The welding metallurgy of titanium and titanium alloys is discussed in greater depth in the Section "Selection of Nonferrous Low-Temperature Materials" in this Volume. Additional information can be found in the article "Welding of Titanium Alloys" in this Volume.

Reference cited in this section

5. F.E. KATEAK, I.D. SERVI, AND J.C. AGARWAL, NONAEROSPACE APPLICATIONS: TITANIUM'S SOLE OPPORTUNITY FOR GROWTH, *MET. BULL. MON.*, AUG 1991

Zirconium- and Tantalum-Base CR Alloys

The metals and alloys based on zirconium and tantalum have weldability characteristics similar to those of titanium. They readily react with oxygen and form very stable oxides. Like titanium, they have high solubilities for oxygen, nitrogen, and hydrogen at elevated temperatures. Small amounts of dissolved oxygen and nitrogen significantly increase the hardness of the metal, whereas dissolved hydrogen reduces toughness and increases notch sensitivity. Therefore, these metals must be welded or brazed in a shield of high-purity inert gas or in a vacuum (Ref 6).

Unalloyed zirconium is similar in characteristics to titanium, except that its density is about 50% higher. Like titanium, zirconium has a hcp crystal structure (α phase) up to approximately 865 °C (1590 °F), and it transforms to a bcc structure (β phase) above this temperature.

Zirconium is resistant to corrosive attack by most acids, strong alkalies, and some molten salts. Its corrosion-resistant behavior is due to the formation of a stable, dense, adherent, and self-healing zirconium oxide film on the metal surface. Zirconium is used for corrosion applications in the petrochemical, chemical process, and food processing industries. Zirconium alloys are also used in the nuclear industry for pressurized water reactor components, because of their low thermal neutron absorption characteristics.

Zirconium is readily welded using the same processes and procedures as titanium. Residual welding stresses are generally low because of its low modulus of elasticity. Zirconium also has a low coefficient of thermal expansion, which contributes to low distortion during welding. In addition, inclusions in welds are usually not a problem, because zirconium has a high solubility for its own oxide. Although some porosity is possible in zirconium welds, it can usually be minimized by properly adjusting welding parameters and by following sound welding practices with respect to precleaning and proper shielding during the welding process.

Additional information is available in the article "Welding of Zirconium Alloys" in this Volume.

Tantalum is a very dense (over 3.5 times the density of titanium) CR material. It is an inherently soft, fabricable metal that has a very high melting temperature, approximately 3000 °C (5400 °F). It has a bcc crystal structure, with no allotropic transformation at elevated temperature. Unlike many bcc metals, tantalum retains good ductility to very low temperatures, and exhibits a ductile-to-brittle transition temperature (DBTT) of approximately -250 °C (-420 °F).

Tantalum has excellent corrosion resistance to a wide variety of acids, alcohols, chlorides, sulfates, and other chemicals. For this reason, it is widely used for chemical equipment that operates at ambient temperatures.

Tantalum oxidizes in air above approximately 300 °C (570 °F) and is attacked by hydrofluoric, phosphoric, and sulfuric acids, and by chlorine and fluorine gases at temperatures above 150 °C (300 °F). Tantalum also reacts with carbon, hydrogen, and nitrogen at elevated temperatures. Because of the high per unit weight cost of tantalum and its high density, the cost effectiveness of this material has been increased by using clad, explosion clad, or resistance-welded clad products.

Like zirconium, tantalum is also readily welded by the processes and procedures used for titanium. The joining of tantalum-clad plate requires special procedures and precautions and is beyond the scope of this article.

Reference cited in this section

6. J.M. GERKIN *ET AL.*, TITANIUM, ZIRCONIUM, HAFNIUM, TANTALUM, AND COLUMBIUM, *WELDING HANDBOOK*, 7TH ED., VOL 4, *METALS AND THEIR WELDABILITY*, AWS, 1982

References

1. H.H. UHLIG AND A.I. ASPHAHANI, CORROSION BEHAVIOR OF COBALT BASE ALLOYS IN AQUEOUS MEDIA, *MATER. PERFORM.*, VOL 18 (NO. 11), NOV 1979
2. S.J. MATTHEWS, M.O. MADDOCK, AND W.F. SAVAGE, HOW COPPER SURFACE CONTAMINATION AFFECTS WELDABILITY OF COBALT SUPERALLOYS, *WELD. J.*, MAY 1972
3. W.F. SAVAGE AND M. MUSHALA, COPPER CONTAMINATION CRACKING IN THE WELD HEAT AFFECTED ZONE, *WELD. J.*, MAY 1978, P 145
4. S.J. MATTHEWS, P. CROOK, L.H. FLASCHE, AND J.W. TACKETT, WELDABILITY CHARACTERISTICS OF A NEW CORROSION- AND WEAR-RESISTANT COBALT ALLOY, *WELD. J.*, DEC 1991, P 331-337
5. F.E. KATEAK, I.D. SERVI, AND J.C. AGARWAL, NONAEROSPACE APPLICATIONS: TITANIUM'S SOLE OPPORTUNITY FOR GROWTH, *MET. BULL. MON.*, AUG 1991
6. J.M. GERKIN *ET AL.*, TITANIUM, ZIRCONIUM, HAFNIUM, TANTALUM, AND COLUMBIUM, *WELDING HANDBOOK*, 7TH ED., VOL 4, *METALS AND THEIR WELDABILITY*, AWS, 1982

Weldability Testing

Richard D. Campbell, Joining Services, Inc.; Daniel W. Walsh, California Polytechnic State University

Introduction

THIS ARTICLE describes many of the various weldability tests that have been developed through the years. Some of these tests are used extensively, while others find limited application. The purposes of this article are to briefly describe these tests, to show what material characteristics they provide, and to present some advantages and limitations of each. This discussion is limited to tests used to evaluate the weldability of base materials.

Additional Tests Used in Welding (Service Tests). This article does not cover the basic tests used for weld qualification, as specified in the numerous welding codes and standards. The tests for welding procedure qualification are used to establish the properties of the welded assembly (including any filler material) and determine that the weldment is capable of providing the required properties for the intended applications. Tests for welder performance qualification are intended to establish the ability of the welder to deposit sound weld metal. Examples of these tests are given in Table 1; information about these tests can be found in the Section "Joint Evaluation and Quality Control" in this Volume. These are often termed service tests because they measure the properties of the weldment under servicelike conditions.

TABLE 1 TYPICAL TESTS REQUIRED IN VARIOUS WELDING CODES AND STANDARDS FOR WELDING PROCEDURE QUALIFICATION AND WELDER PERFORMANCE QUALIFICATION

DESTRUCTIVE TESTS
MECHANICAL TESTS
TENSILE TESTS
GUIDED BEND TESTS
CHARPY V-NOTCH IMPACT TOUGHNESS TESTS
DROP WEIGHT NOTCH TOUGHNESS TESTS
FILLET WELD BREAK, SHEAR, OR FRACTURE TESTS
STUD WELD BEND OR TORQUE TESTS
METALLOGRAPHIC TESTS
MACROEXAMINATION TESTS
MACROETCH TESTS
MICROEXAMINATION TESTS
NONDESTRUCTIVE TESTS
VISUAL INSPECTION

LIQUID PENETRANT INSPECTION
MAGNETIC PARTICLE INSPECTION
RADIOGRAPHIC INSPECTION

Tests to evaluate the mechanical or corrosion properties of weldments are described in detail in several sources (Ref 1, 2). Other Volumes of the *ASM Handbook* present the basics of mechanical testing (Ref 3), metallographic techniques and etchants (Ref 4), and nondestructive testing methods (Ref 5).

This article does not cover tests for weld metal or weld wire. These tests, such as chemical analysis or mechanical testing of the deposited weld metal, are used to evaluate the properties of the weld filler material. Welding tests of filler material, such as spark tests, are performed to evaluate the welding characteristics of the wire.

References

1. "WELDING TECHNOLOGY," CHAPTER 12, *WELDING HANDBOOK,* VOL 1, 8TH ED., AMERICAN WELDING SOCIETY, 1987
2. "STANDARD METHODS FOR MECHANICAL TESTING OF WELDS," ANSI/AWS B4.0-85, AMERICAN WELDING SOCIETY, 1985
3. *MECHANICAL TESTING,* VOL 8, *ASM HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1985
4. *METALLOGRAPHY AND MICROSTRUCTURES,* VOL 9, *ASM HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1985
5. *NONDESTRUCTIVE EVALUATION AND QUALITY CONTROL,* VOL 17, *ASM HANDBOOK,* ASM INTERNATIONAL, 1989

Weldability Testing

Richard D. Campbell, Joining Services, Inc.; Daniel W. Walsh, California Polytechnic State University

General Characteristics of Weldability Tests

Weldability is defined as "*the capacity of a material to be welded under the imposed fabrication conditions into a specific, suitably designed structure and to perform satisfactorily in the intended service*" (Ref 6). The weldability tests described in this article are used to evaluate the effects of welding on such properties and characteristics as:

- BASE-METAL AND WELD-METAL CRACKING
- BASE-METAL AND WELD-METAL DUCTILITY
- WELD PENETRATION
- WELD POOL SHAPE AND FLUID FLOW

Because weldability testing is used to evaluate the welding characteristics of the base materials, many of these weldability tests are laboratory or research tests rather than "production" tests. Weldability tests are used extensively during alloy development. However, many are also used during weld procedure development to ensure the weldability of base materials before production commences. Each test is designed to evaluate a susceptibility to a specific weldability problem.

Obviously, the best tests evaluate the actual service conditions. However, because this is expensive and often impossible, standardized weldability tests and test methods have been developed. These tests evaluate various responses of the base material to different welding conditions or simulated welding conditions. Most of these tests utilize simple geometries and configurations, and loading conditions are typically simple and uniform. Thus, none of these tests duplicates the actual service conditions of a real weldment. Instead they provide a baseline for comparisons of base material responses.

The various weldability tests provide information on alloys, welding processes, and welding procedures. Data generated can be compared to data produced under the same testing conditions on other heats of material or to data produced with other welding processes or procedures (Ref 7). All of these tests provide qualitative information in that they are used to sort different materials or heats of the same materials, processes, procedures, and thermal cycles. None of these tests have any quantitative "acceptance" limits, such as the yield strength requirements of a specific alloy when tensile tested (Ref 7).

All weldability tests involve either actual welding or simulated welding. Those involving actual welding require that the test coupons be produced using an actual welding process. An example of this type of test is Varestraint hot crack testing using the gas-tungsten arc welding (GTAW) process. Simulated welding tests involve the application of heat through means other than welding, with or without loading. An example is Gleeble testing, in which a test specimen is resistance heated to simulate a weld heat-affected zone (Ref 7). Excellent reviews of various weldability tests are provided in Ref 8 and 9.

References cited in this section

6. "STANDARD WELDING TERMS AND DEFINITIONS," ANSI/AWS A3.0-89, AMERICAN WELDING SOCIETY, 1989
7. "WELDABILITY TESTING," CHAPTER 4, *WELDING HANDBOOK,* VOL 1, 8TH ED., AMERICAN WELDING SOCIETY, 1987
8. R.D. STOUT AND W.D. DOTY, *WELDABILITY OF STEELS,* WELDING RESEARCH COUNCIL, 1978
9. R.D. STOUT, *WELDABILITY OF STEELS,* WELDING RESEARCH COUNCIL, 1987

Weldability Testing

Richard D. Campbell, Joining Services, Inc.; Daniel W. Walsh, California Polytechnic State University

Weldability Tests for Evaluating Cracking Susceptibility

The majority of weldability tests have been developed to evaluate the susceptibility of the base materials (and any added filler material) to cracking. The purpose of these tests is to reduce or to eliminate the formation of these defects during fabrication or service. Cracking can be in the form of hot cracking, cold cracking, or lamellar tearing. Detailed information about these defects is available in the article "Cracking Phenomena Associated With Welding" in this Volume. Table 2 presents a comparison of several weldability tests for cracking susceptibility, their applications, and resultant types of data.

TABLE 2 WELDABILITY TESTS USED TO EVALUATE SUSCEPTIBILITY TO CRACKING

TEST	FIELDS OF USE	CONTROLLABLE VARIABLES	TYPE OF DATA	SPECIALIZED EQUIPMENT	RELATIVE COST
LEHIGH RESTRAINT TEST	WELD METAL HOT AND COLD CRACKING, ROOT CRACKING, HAZ HYDROGEN CRACKING, STRESS-RELATED CRACKING	JOINT GEOMETRY, PROCESS, FILLER METAL, RESTRAINT LEVEL, HEAT INPUT, PREHEAT	CRITICAL RESTRAINT, OR % HINDERED CONTRACTION	NONE	COSTLY MACHINING
SLOT TEST	HAZ HYDROGEN CRACKING	FILLER METAL, INTERPASS TIME, PREHEAT	TIME TO CRACK, CRITICAL PREHEAT	NONE	LOW COST
RIGID RESTRAINT (RRC) TEST	WELD METAL HOT AND COLD CRACKING, ROOT CRACKING, HAZ HYDROGEN CRACKING	JOINT GEOMETRY, PROCESS RESTRAINT LEVEL, FILLER METAL, HEAT INPUT, PREHEAT	CRITICAL RESTRAINT	RESTRAINING JIG	COSTLY MACHINING AND SET-UP
TEKKEN TEST	WELD METAL ROOT CRACKING, HAZ HYDROGEN CRACKING	JOINT GEOMETRY, PROCESS, FILLER METAL, HEAT INPUT, PREHEAT	CRITICAL PREHEAT	NONE	LOW-COST
CIRCULAR GROOVE TEST	WELD METAL HOT AND COLD CRACKING, HAZ HYDROGEN CRACKING	PROCESS, FILLER METAL, PREHEAT	GO-NO GO	NONE	COSTLY PREPARATION
IMPLANT TEST	HAZ HYDROGEN CRACKING, STRESS-RELIEF CRACKING	PROCESS, FILLER METAL, PREHEAT, PWHT	CRITICAL FRACTURE STRESS, CRITICAL PREHEAT	LOADING JIG	MEDIUM COST
TENSION RESTRAINT CRACKING (TRC) TEST	HAZ HYDROGEN CRACKING	PROCESS, FILLER METAL, HEAT INPUT, PREHEAT	CRITICAL FRACTURE STRESS, CRITICAL PREHEAT	LOADING JIG	COSTLY MACHINING AND SET-UP
VARESTRAINT TEST	WELD METAL AND HAZ HOT CRACKING	PROCESS, FILLER METAL, HEAT INPUT	CRACK LENGTH, % STRAIN	LOADING JIG	COSTLY PREPARATION AND ANALYSIS
LONGITUDINAL, BEAD-ON-PLATE TEST	HAZ HYDROGEN CRACKING	CURRENT, HEAT INPUT	% CRACKING	NONE	LOW COST

CONTROLLED THERMAL SEVERITY TEST	HAZ HYDROGEN CRACKING IN FILLET WELDS	CURRENT, COOLING RATE, PREHEAT	GO-NO GO (AT 2 COOLING RATES)	NONE	COSTLY PREPARATION
CRUCIFORM TEST	HAZ HYDROGEN CRACKING, WELD METAL ROOT CRACKS	PROCESS, HEAT INPUT, PREHEAT, FILLER METAL	GO-NO GO	NONE	COSTLY PREPARATION
LEHIGH CANTILEVER TEST	LAMELLAR TEARING	PROCESS, FILLER INPUT, HEAT INPUT, PREHEAT	CRITICAL RESTRAINT, STRESS AND STRAIN	LOADING JIG	COSTLY SPECIMEN PREPARATION
CRANFIELD TEST	LAMELLAR TEARING	FILLER METAL	NUMBER OF PASSES TO CRACK	NONE	LOW COST
NICK BEND TEST	WELD METAL SOUNDNESS	FILLER METAL	GO-NO GO	NONE	LOW COST

Source: Ref 9

Hot and Cold Cracking

Hot cracking occurs during weld solidification and can occur in the weld metal or in the heat-affected zone (HAZ). Hot cracking is caused by low melting temperature constituents, in addition to tensile stress on the weld. Other names for hot cracking include microfissuring, solidification cracking, reheat cracking, and liquation cracking. In actual weldments, the tensile stress may come from external loads on the structure, the weight of the structure producing stress on the weld, solidification shrinkage stresses, or combinations of these.

Cold cracking (also known as hydrogen-induced cracking) is a cracking phenomenon that occurs after the weld has solidified (Ref 6). It is also referred to as delayed cracking, because cracking is often "delayed" for minutes, hours, or even longer, after the weld has solidified. Cold cracking is most often associated with the presence of hydrogen in hardenable steels. The ductility and toughness of the welded structure are greatly reduced by cold cracking. Requisites for cold cracking include tensile stress (external loading or residual stress from welding), a crack-susceptible microstructure (martensite), and the presence of hydrogen. Cold cracking can occur in the weld fusion zone or in the HAZ.

Several weldability tests for evaluating cracking susceptibility are described in this article. Most tests are used to evaluate susceptibility to one type of cracking: cold cracking, hot cracking, underbead cracking, or lamellar tearing. Some tests are useful for both hot and cold cracks. The applicabilities of each test are discussed within the descriptions of the test. The weldability tests discussed in this section are classified as self-restraint tests or externally loaded tests.

Self-Restraint Tests

Self-restraint tests utilize the restraint, or stress, within the specimen to cause weld metal or base metal cracking. No external loading of the specimen occurs. The specimen is designed to produce variable restraint on the weld joint, thus causing cracking.

Lehigh Restraint Test (Hot and Cold Cracks). The Lehigh restraint test is illustrated in Fig. 1. This test, developed at Lehigh University, uses a plate with slots machined into the sides and ends. A groove weld joint is machined along the centerline of the plate, and a single-pass weld is produced along this groove. The restraint from the plate and slots produces a weld with various levels of cracking.

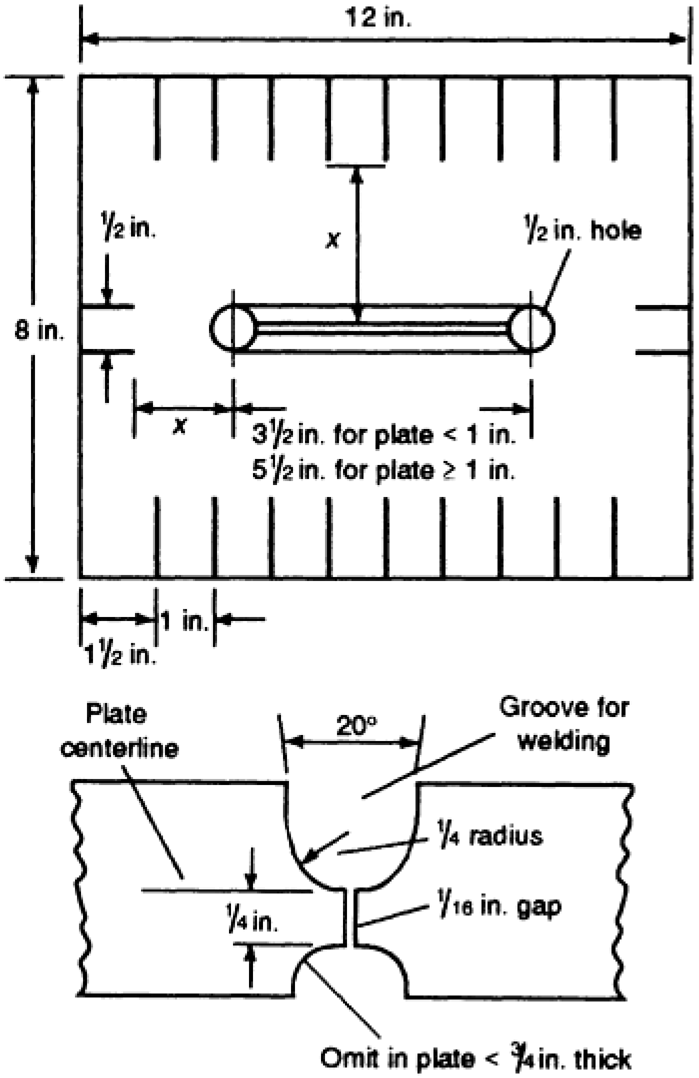
12 in.
8 in.
½ in.
x
½ in. hole
x
3½ in. for plate < 1 in.
5½ in. for plate ≥ 1 in.
1 in.
1½ in.
Plate centerline
20°
Groove for welding
¼ radius
¼ in.
1/16 in. gap
Omit in plate < ¾ in. thick

FIG. 1 SCHEMATIC SHOWING SLOTTED PLATE SPECIFICATIONS FOR LEHIGH RESTRAINT TEST. SOURCE: REF 8

This test compares quantitatively the degree of restraint at which cracking occurs in welds during cooling. The level of restraint is altered by changing the length of the slots (Ref 7). The restraint is quantified as 2*x*, where *x* is the distance from the centerline of the weld groove to the bottom of the slots. Numerous samples with different slot lengths are welded, and a threshold level of restraint is determined as the width that is just sufficient to cause cracking. Cracking can be evaluated by metallographic cross sections of the weld or by heat tinting and breaking open the weld to expose the cracked region. This test has also been used to evaluate reheat cracking in steels (Ref 7, 8, 9).

Keyhole Restraint Cracking Test (Hot and Cold Cracks). The Naval Research Laboratory (NRL) keyhole restraint crack test (Fig. 2) is a simplified version of the Lehigh test. The specimen is welded along the groove beginning at the open end and progressing toward the hole. This imposes a varying degree of restraint along the weld, with a maximum at the hole and a minimum at the edge where welding commenced. Cracks form at the hole and extend outward to a point where the restraint is low enough to arrest the crack growth. The crack length is the measure of crack susceptibility (Ref 8).

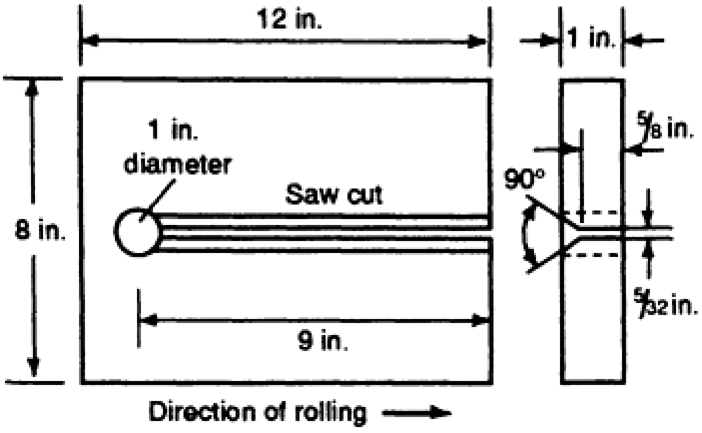
12 in.
1 in.
1 in.
diameter
Saw cut
5/8 in.
90°
8 in.
5/32 in.
9 in.
Direction of rolling

FIG. 2 PLATE SPECIFICATIONS FOR NRL KEYHOLE RESTRAINT TEST. SOURCE: REF 8

Houldcroft Crack Susceptibility Test (Hot Crack Test). The Houldcroft crack susceptibility test utilizes a design similar to the Lehigh test, but the slots vary in length from one end of the specimen to the other (Fig. 3). The Houldcroft test was developed for sheet steels. No weld joint is machined into the specimen, but rather a bead-on-plate, complete penetration weld (usually using the GTAW process) is made. Duplicate samples are produced, and the mean crack length of these specimens is used as an index of hot crack susceptibility (Ref 8).

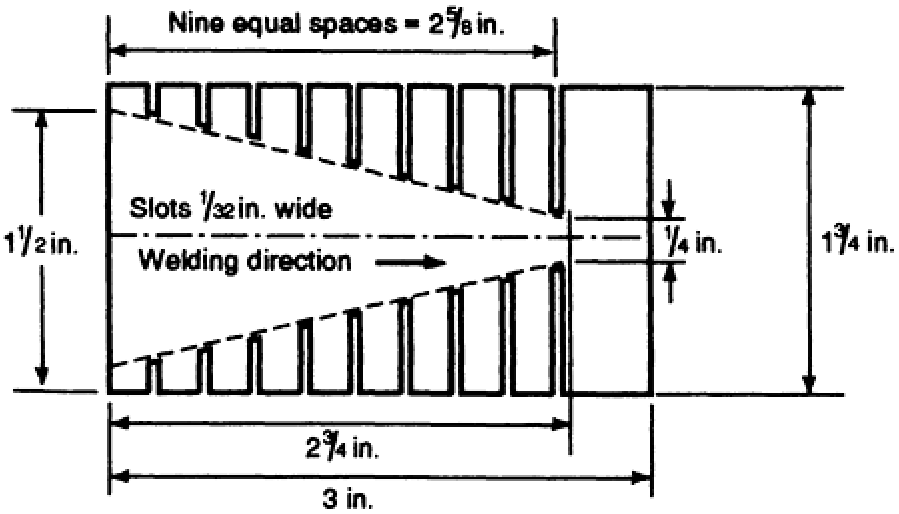
Nine equal spaces = 2 5/8 in.
Slots 1/32 in. wide
1 1/2 in.
Welding direction
1/4 in.
1 3/4 in.
2 3/4 in.
3 in.

FIG. 3 SPECIMEN DIMENSIONS FOR HOULDCROFT CRACK SUSCEPTIBILITY TEST. SOURCE: REF 8

Keyhole Slotted-Plate Restraint Test (Hot Crack Test). A variation of the Houldcroft test for plate material is the keyhole slotted-plate restraint test (Fig. 4). This test was developed at Battelle and utilizes a groove weld, which begins at the low restraint end. Crack sensitivity is determined, in a manner similar to the Lehigh test, by the uncut width of the specimen at which the crack propagation ceases (Ref 8).

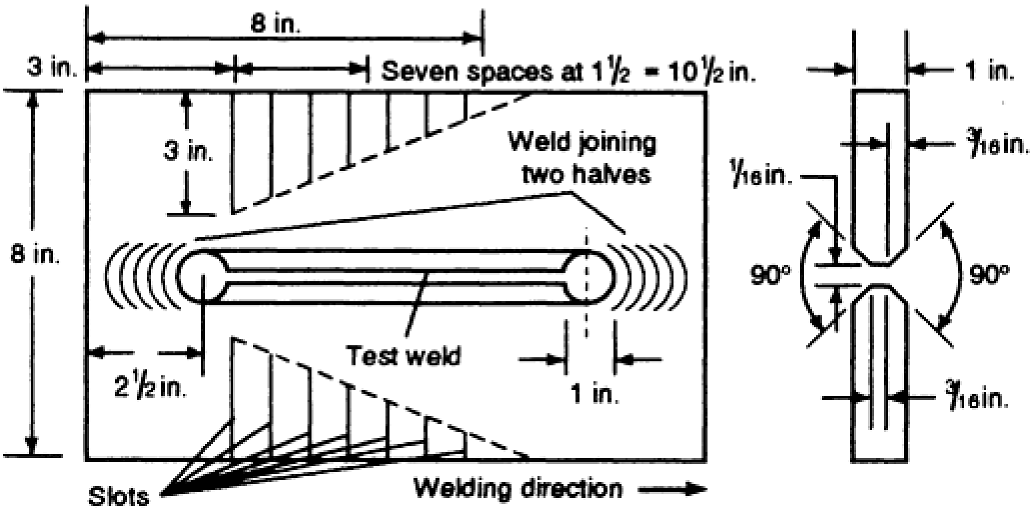
8 in.
3 in.
Seven spaces at 1½ = 10½ in.
1 in.
3 in.
Weld joining two halves
$^1\!/_{16}$ in.
$^3\!/_{16}$ in.
8 in.
90°
90°
Test weld
2½ in.
1 in.
$^3\!/_{16}$ in.
Slots
Welding direction

FIG. 4 SPECIMEN REQUIREMENTS FOR KEYHOLE SLOTTED-PLATE RESTRAINT TEST. SOURCE: REF 8

Tekken Test (Cold Crack Test). The Tekken Test, developed in Japan, uses several specimen dimensions and joint designs (Fig. 5). The Y joint provides more restraint than the U- or double-U joint designs. In these tests, preheat and welding parameters are varied to alter the stress state, rather than changing the specimen design. This test is used to evaluate HAZ cracks and underbead cracks.

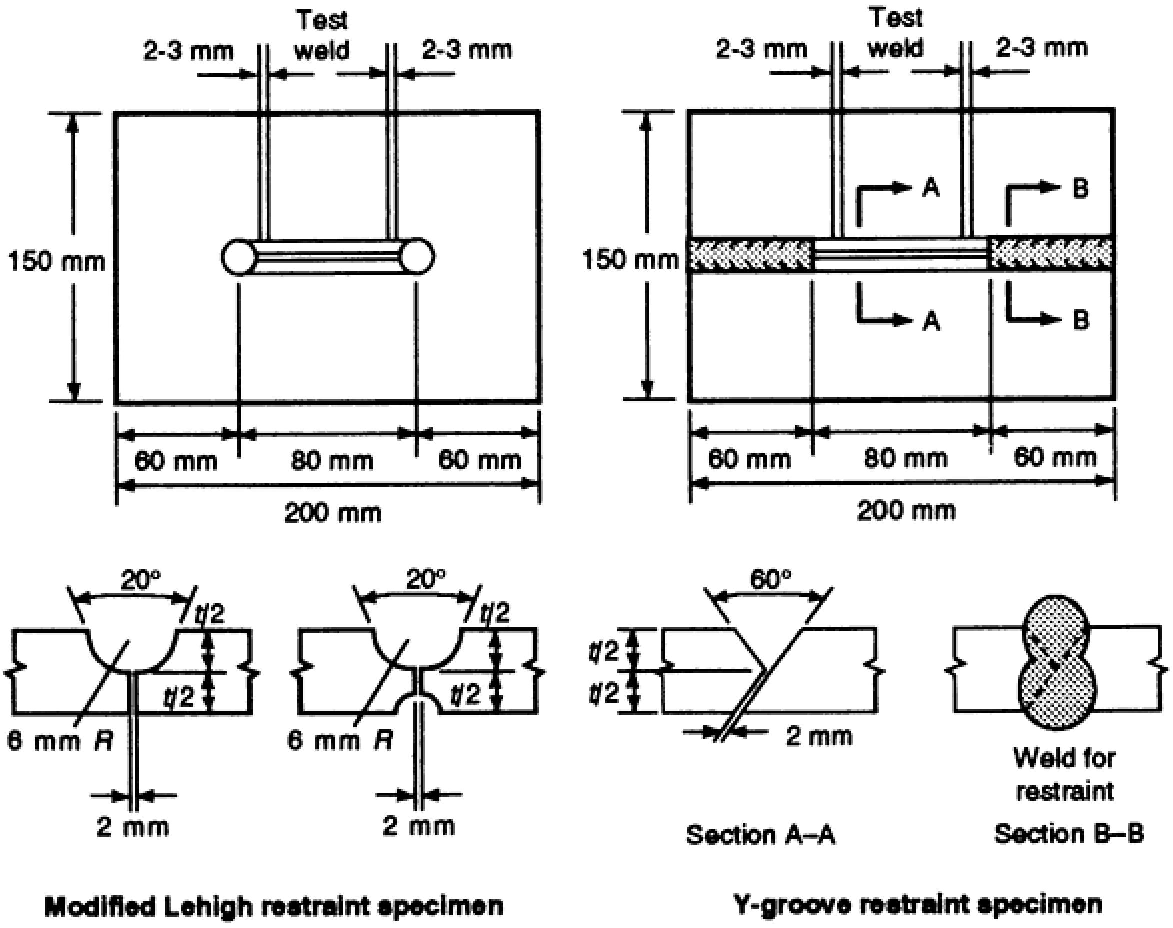
Test weld
2-3 mm
2-3 mm
150 mm
60 mm
80 mm
60 mm
200 mm
Test weld
2-3 mm
2-3 mm
A
B
150 mm
A
B
60 mm
80 mm
60 mm
200 mm
20°
t/2
t/2
6 mm R
2 mm
20°
t/2
t/2
6 mm R
2 mm
60°
t/2
t/2
2 mm
Section A–A
Weld for restraint
Section B–B
Modified Lehigh restraint specimen
Y-groove restraint specimen

FIG. 5 SAMPLE SPECIFICATIONS FOR TEKKEN TEST. SOURCE: REF 9

Circular Patch Test (Cold and Hot Crack Test). The circular patch test (Fig. 6) utilizes a plate with a circular hole cut out and replaced by a patch. Both the plate and the patch have a weld joint machined into them. One or more weld passes are made in this circular groove, and cracking is detected by visual, radiographic, and/or liquid penetrant inspection. This test is used to generate and evaluate both hot cracks and cold cracks (Ref 8, 9).

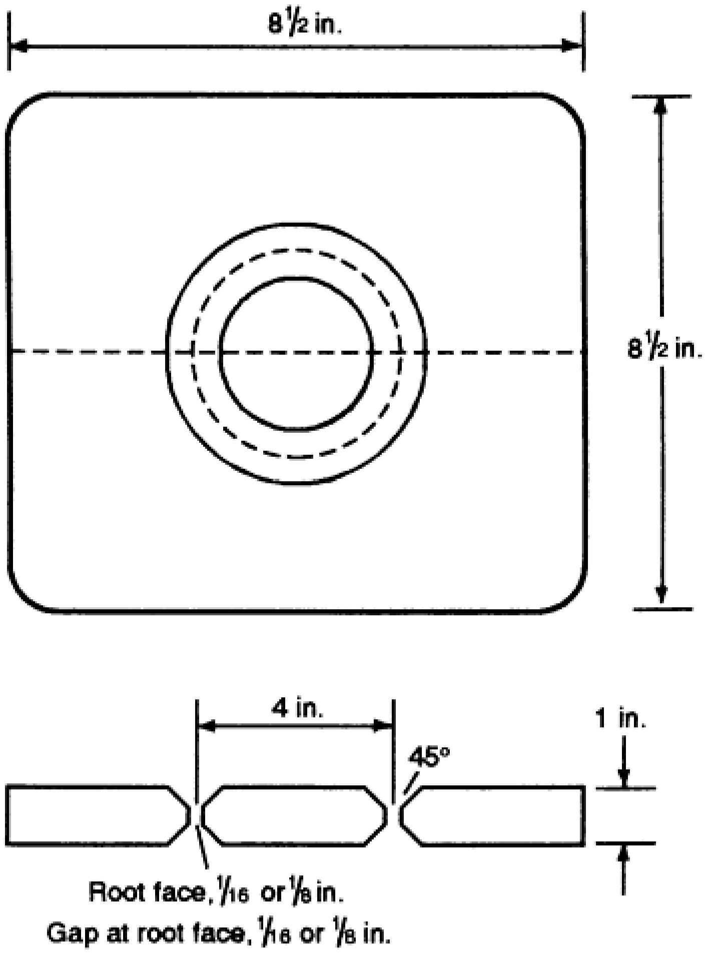

FIG. 6 PLATE SPECIFICATIONS FOR CIRCULAR PATCH TEST. SOURCE: REF 8

Navy Circular Patch Test (Cold and Hot Crack Test). The Navy circular patch test utilizes a much larger specimen (Fig. 7). This test is used for evaluating different electrodes on various base metals. A circular hole is cut in the plate, into which a patch is placed, with a backing ring to hold the plate and patch together. A multipass weld is made, one quadrant at a time, as shown in Fig. 7. Cracking evaluation is by visual, radiographic, and/or liquid penetrant inspection. Both hot cracks and cold cracks can be generated in this test (Ref 8).

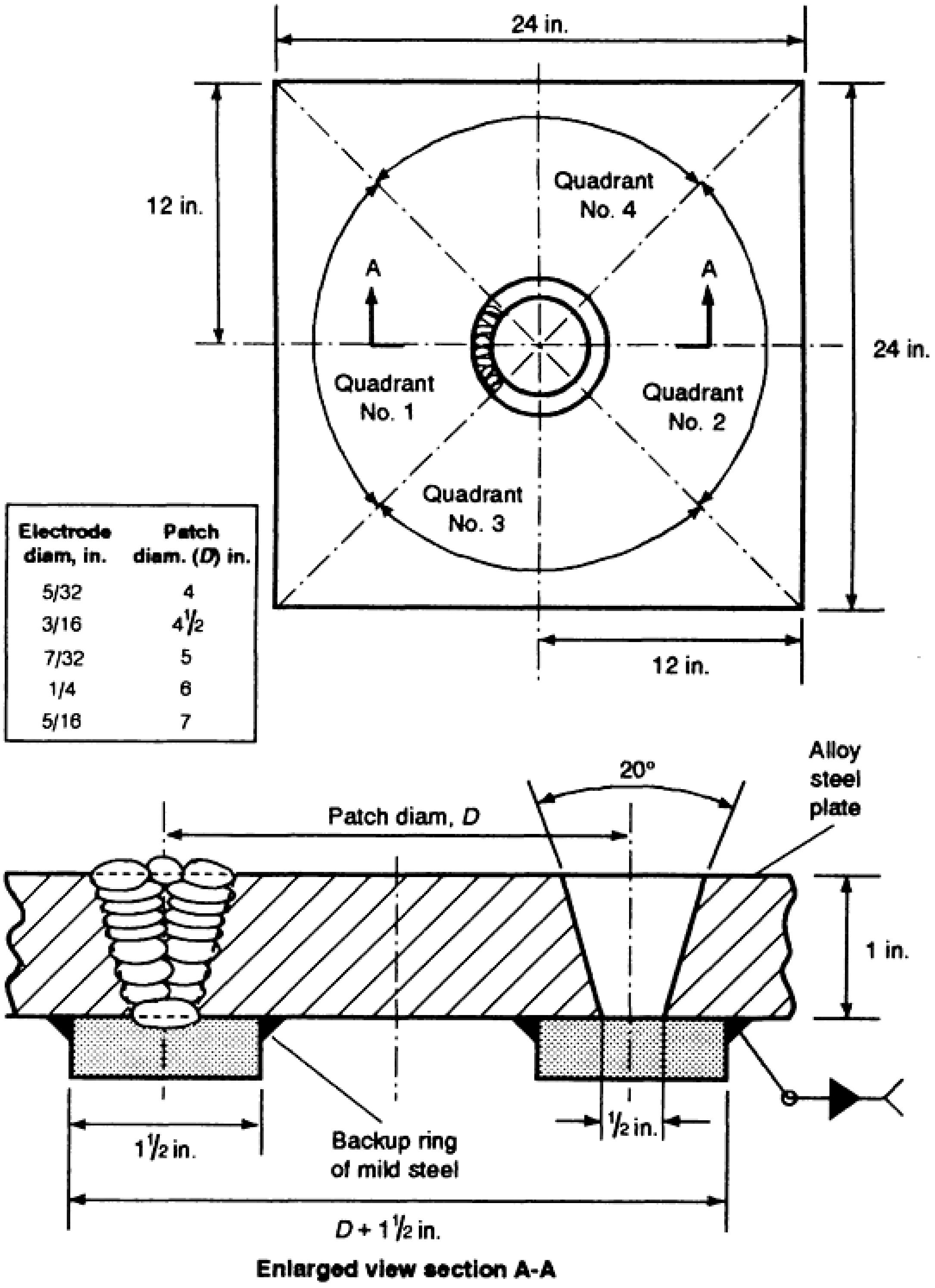
24 in.
12 in.
Quadrant No. 4
A
A
24 in.
Quadrant No. 1
Quadrant No. 2
Quadrant No. 3
12 in.
20°
Alloy steel plate
Patch diam, D
1 in.
1½ in.
Backup ring of mild steel
½ in.
D + 1½ in.
Enlarged view section A-A

Electrode diam, in.	Patch diam. (D) in.
5/32	4
3/16	4½
7/32	5
1/4	6
5/16	7

FIG. 7 PLATE REQUIREMENTS FOR NAVY CIRCULAR PATCH TEST. SOURCE: REF 8

Controlled-Thermal-Severity Test (Cold Crack Test). The controlled-thermal-severity test specimen consists of a square plate bolted and anchor welded to a larger rectangular plate (Fig. 8). After the anchor welds have cooled to room temperature, two test welds are made on the specimen. The fillet weld along the plate edges is made first. This weld, called the bithermal weld, has only two-dimensional heat flow. After this weld has cooled, the lap weld is made on the other edge of the top plate, near the center of the bottom plate. Because this weld has nearly three-dimensional heat flow, it is called a trithermal weld. The specimen is held for 72 h at room temperature, then the degree of cracking is determined by measuring crack length on three metallographic specimens cut from cross sections of the two welds.

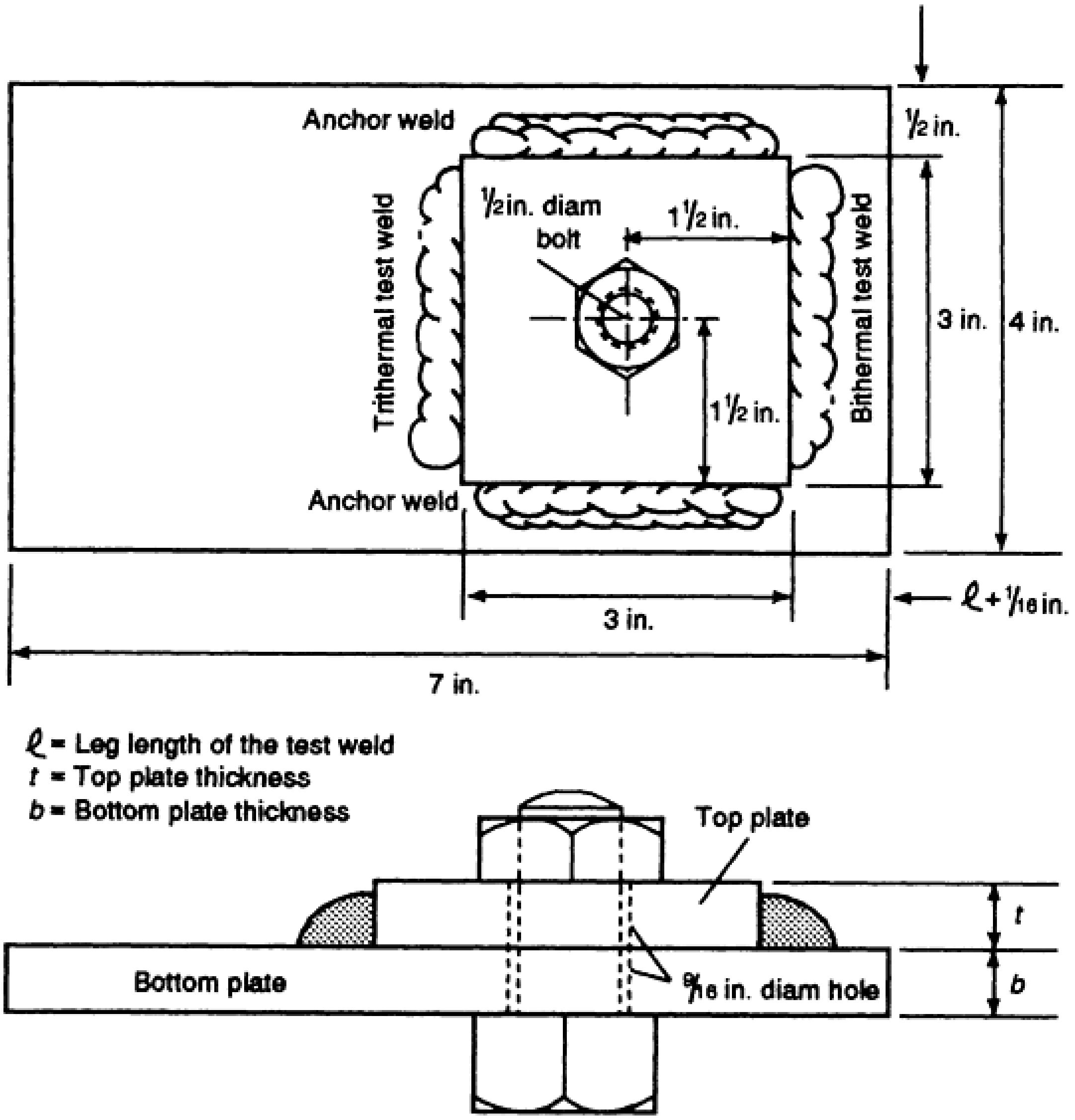
Anchor weld
½ in.
½ in. diam bolt
1½ in.
Trithermal test weld
Bithermal test weld
3 in.
4 in.
1½ in.
Anchor weld
3 in.
ℓ + 1/16 in.
7 in.
ℓ = Leg length of the test weld
t = Top plate thickness
b = Bottom plate thickness
Top plate
t
Bottom plate
9/16 in. diam hole
b

FIG. 8 PLATE DIMENSIONS FOR CONTROLLED THERMAL SEVERITY TEST. SOURCE: REF 8

Each specimen is assigned a thermal severity number (*TSN*) according to plate thickness using Eq 1 and 2. For bithermal welds:

$$TSN = 4(T + B) \qquad \textbf{(EQ 1)}$$

where *t* and *b* are the thicknesses of the top and bottom plates, respectively. For trithermal welds:

$$TSN = 4(T + 2B) \qquad \textbf{(EQ 2)}$$

A series of plate thicknesses, which provided varying cooling rates, are tested. The crack susceptibility of the base metal-filler material (welding wire or electrode) combination is determined by the minimum *TSN* that produces cracking. Controlled-thermal-severity testing is used to measure the cold crack sensitivity of steels under cooling rates controlled by the thickness of the plates and the differences in cooling rates between bithermal and trithermal welds. This test is primarily used to evaluate the crack sensitivity of hardenable steels (Ref 7, 9).

Cruciform Cracking Test (Cold Crack Test). The cruciform cracking test (Fig. 9) uses three plates that are surface ground along their interfaces to ensure consistent and reproducible fit-up. The ends are tack welded together to form a double T-joint. Four test fillet welds are deposited in the sequence shown in Fig. 9, with complete cooling between passes. The specimen is held at room temperature for at least two days before stress relieving at about 620 °C (1150 °F) for 2 h (Ref 7). Cracking is determined by evaluation of metallographic cross sections of the four welds. The test data can be erratic and are very susceptible to sample preparation. Five to ten samples are typically required to provide reliable results for cracking sensitivity level (Ref 7, 8).

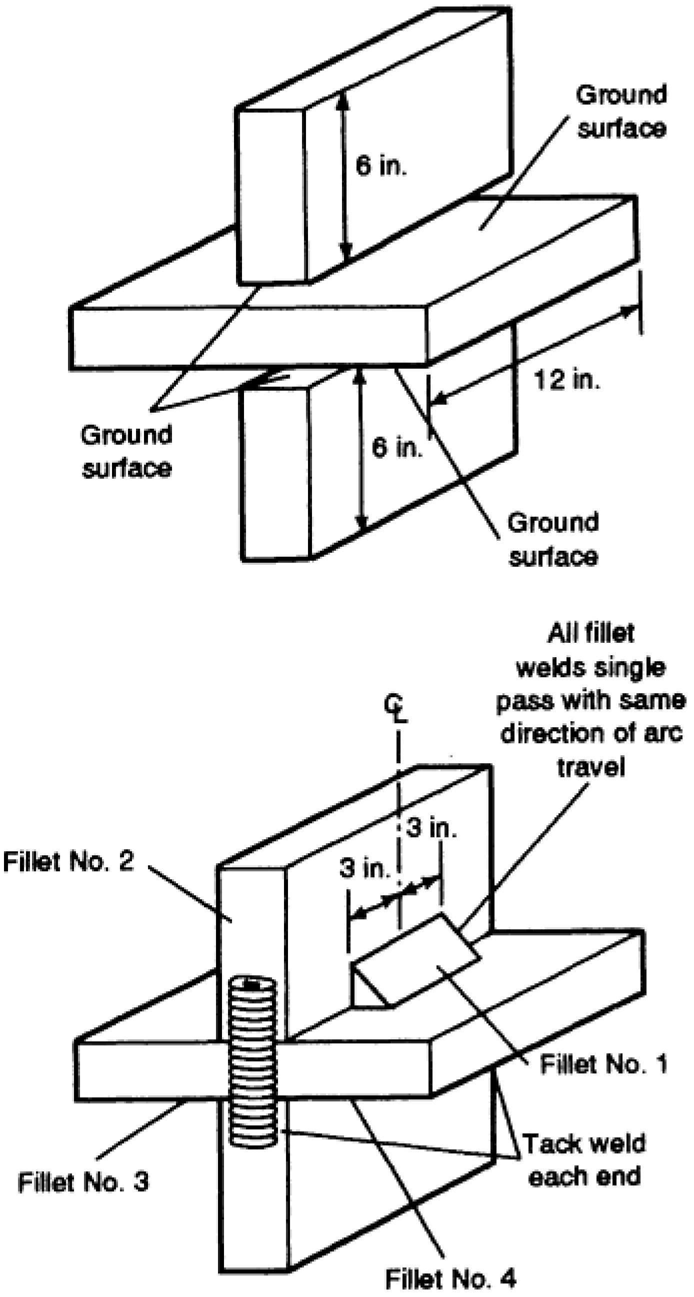
Ground surface
6 in.
12 in.
Ground surface
6 in.
Ground surface
All fillet welds single pass with same direction of arc travel
℄
3 in.
3 in.
Fillet No. 2
Fillet No. 1
Tack weld each end
Fillet No. 3
Fillet No. 4

FIG. 9 SPECIFICATIONS FOR CRUCIFORM CRACKING TEST SPECIMENS. SOURCE: REF 8

Externally Loaded Tests

Externally loaded tests involve the application of an external load to the specimen during weld testing. This external load is typically varied to alter the stress state and thus the severity of cracking. This allows a more quantitative measure of stress than can be obtained from the self-restraint tests.

Implant Testing (Cold Crack Test). The implant test is used to evaluate hydrogen-induced cracking (HIC). A rod of the steel base metal to be tested is machined to the dimensions shown in Fig. 10, the end of the rod having either a circular groove or a helical groove machined into it (Ref 8). The rod is placed inside a hole in the center of a plate so the top of the rod is flush with the top of the plate. A weld bead is made on the top surface of the plate and passes directly over the top of the rod. By making the weld over the rod, the groove in the rod is thus located in the coarse-grained HAZ, which is most susceptible to HIC.

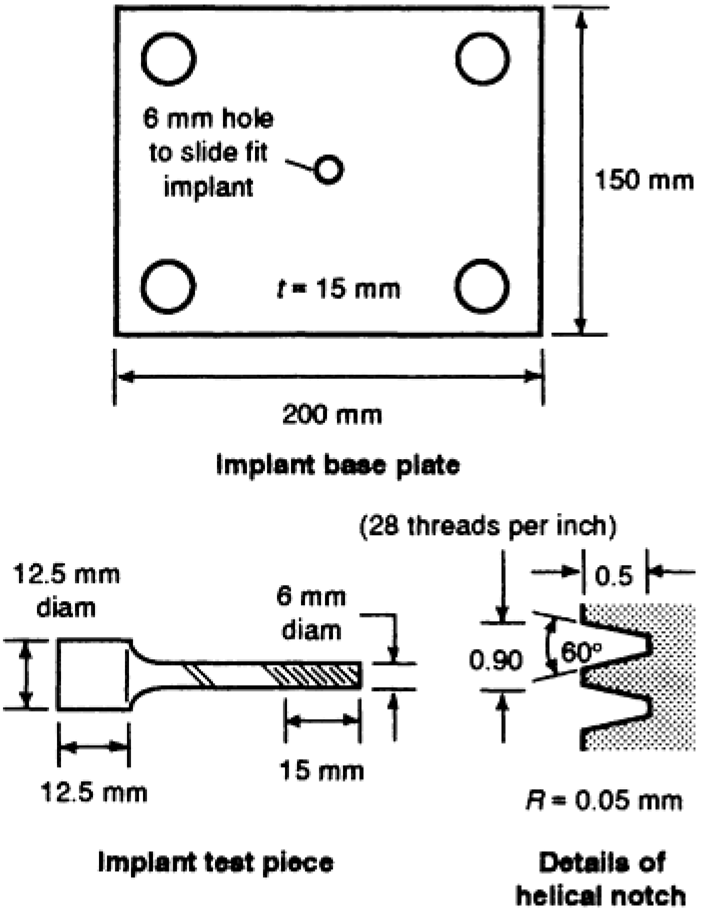

FIG. 10 SAMPLE SPECIFICATIONS FOR IMPLANT TEST SPECIMENS. SOURCE: REF 8

After welding (but within 60 s after weld completion), the test setup is placed in an apparatus that places a tensile stress on the rod, and this is maintained for 24 h, or until the weldment fails. Numerous samples are welded, then tested under different loads. The time to fail is plotted as a function of the loading stress. The crack susceptibility of the base metal is measured by the degree to which the failure stress is reduced with time.

Varestraint Test (Hot Crack Test). The Varestraint (variable restraint) test was developed at Rensselaer Polytechnic Institute (Ref 10). It is the most common test used to evaluate hot crack sensitivity. In this test, one end of a rectangular bar (typically 50 by 205 by 6.4 mm, or 2 by 8 by $\frac{1}{4}$ in.) is firmly mounted in a fixed position while the opposite end is attached to a hydraulic or pneumatic plunger (Fig. 11). A gas-tungsten arc weld is produced on the top side of the plate, along its longitudinal centerline, beginning at the unsupported end and moving toward the fixed end. When the welding arc reaches a predetermined location over a die block, the plate is bent to conform to the radius of the die block. This induces an augmented longitudinal strain on the welded surface of the specimen. The value of the strain is given approximately by:

$$e = \frac{t}{2R} \qquad \textbf{(EQ 3)}$$

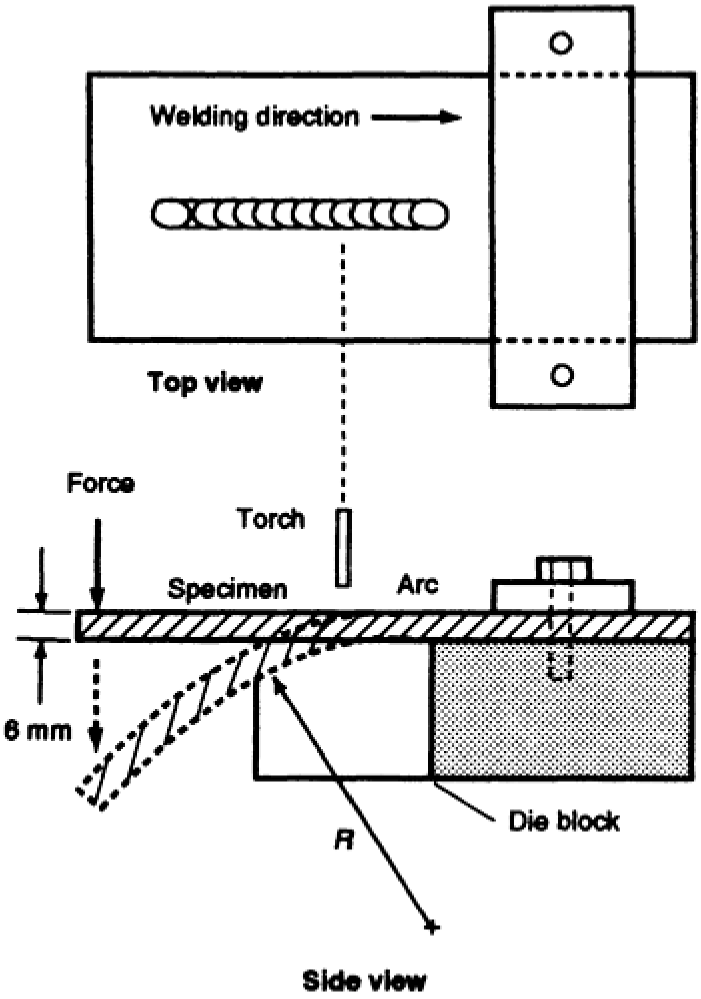

FIG. 11 SETUP AND SAMPLE SPECIFICATIONS FOR VARESTRAINT TEST. SOURCE: REF 7

Various die blocks with different radii can be used on samples of the same material to characterize the strain level at which cracking begins. Typically, for alloy evaluation, a single die block of known radius is used which will produce cracking in all samples. Hot cracks typically form radially along the trailing edge of the weld pool and in the HAZ. Duplicate samples of each alloy are tested, and mean values of total crack length, number of cracks, or maximum crack length are used as a quantitative means of base metal weldability (Ref 7).

The Varestraint test also has been used to test the hot crack susceptibility of filler material. By depositing filler wire on a plate (for example, with the gas-metal arc welding process) and then machining the specimen surface flush, a Varestraint test can be made using the GTAW process. This method has been used to evaluate the effect of base metal-filler material dilution on hot crack susceptibility.

The subscale Varestraint test uses a smaller sample (Fig. 12). This test uses the same testing equipment, but was developed for sheet material (Ref 11). Typical specimen size is 25 by 150 by 3.2 mm (1 by 6 by $\frac{1}{8}$ in.).

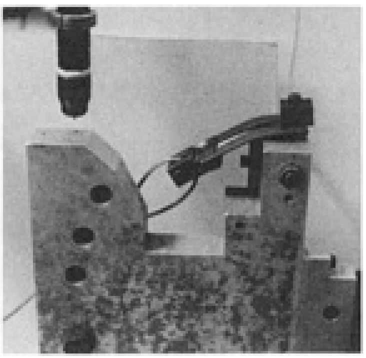

FIG. 12 PHOTO OF A TYPICAL SUBSCALE VARESTRAINT TEST APPARATUS. SOURCE: REF 11

Advantages of the Varestraint and subscale Varestraint tests are that they produce top surface cracks, both in the weld and the HAZ. Disadvantages are the cost of manufacturing or procuring the test equipment.

Another variation of this test is the trans Varestraint test, in which the sample is bent in a direction normal to the welding direction instead of the longitudinal direction. The loading in this test is tensile loading normal to the welding direction. This usually causes centerline cracks to form. The length of centerline cracks is used as a measure of hot crack susceptibility.

Spot Varestraint Test (Hot Crack Test). The spot Varestraint test is a modification of the Varestraint test; it is sometimes referred to as the TIG-A-MA-JIG test (Ref 10). In this test, a stationary welding torch is used to produce a gas-tungsten arc spot weld in the center of a rectangular sample (Fig. 13). The specimen size is typically 140 by 25 by 6.4 mm (5.5 by 1 by 0.25 in.). After the weld is established for a predetermined time, the test specimen is forced around a curved die block and the arc is extinguished. Hot cracks or hot tears form radially around the weld pool in the HAZ. The total crack length, maximum crack length, or strain level necessary to cause cracking is used as the measure of crack sensitivity. The spot Varestraint test is used mainly for evaluating susceptibility to HAZ hot cracking or liquation cracking (Ref 12).

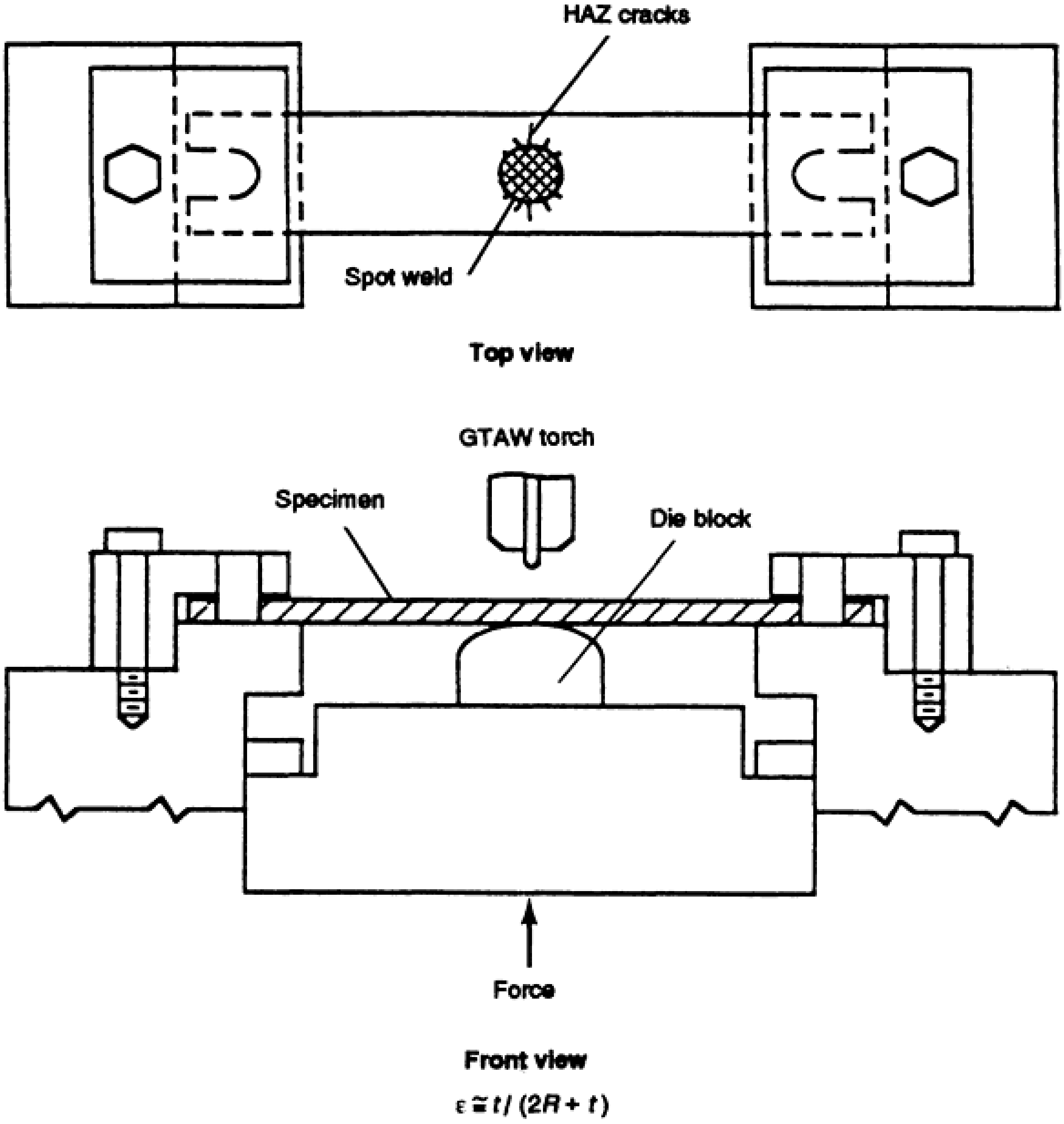

FIG. 13 SCHEMATIC SHOWING TOP AND FRONT VIEWS OF A SETUP USED IN SPOT VARESTRAINT TESTING. SOURCE: REF 12

One potential problem with the spot Varestraint test is that crater cracks often form, and they are difficult to distinguish from the other hot cracks. Also, extensive subsurface cracking often occurs, but cannot be detected without examination of metallographic cross sections. Reproducibility is difficult because slight variations often occur in welding conditions with the spot weld. A modified spot Varestraint test, developed at the University of Tennessee, moves the torch away from the pool prior to bending. This helps eliminate problems associated with crater cracks.

Sigmajig Test (Hot Crack Test). This test, developed at Oak Ridge National Laboratory, involves application of a transverse normal stress, σ, to a 50 by 50 mm (2 by 2 in.) sheet specimen (Fig. 14). An autogenous gas-tungsten arc weld is produced along the longitudinal centerline of the specimen. Numerous specimens are welded, with the stress being increased with each specimen, until centerline cracking occurs. A threshold stress level necessary to cause 100% cracking is used to quantify hot crack susceptibility (Ref 13). The advantages of this test are its usefulness for thin sheet, up to approximately 2.5 mm (0.1 in.).

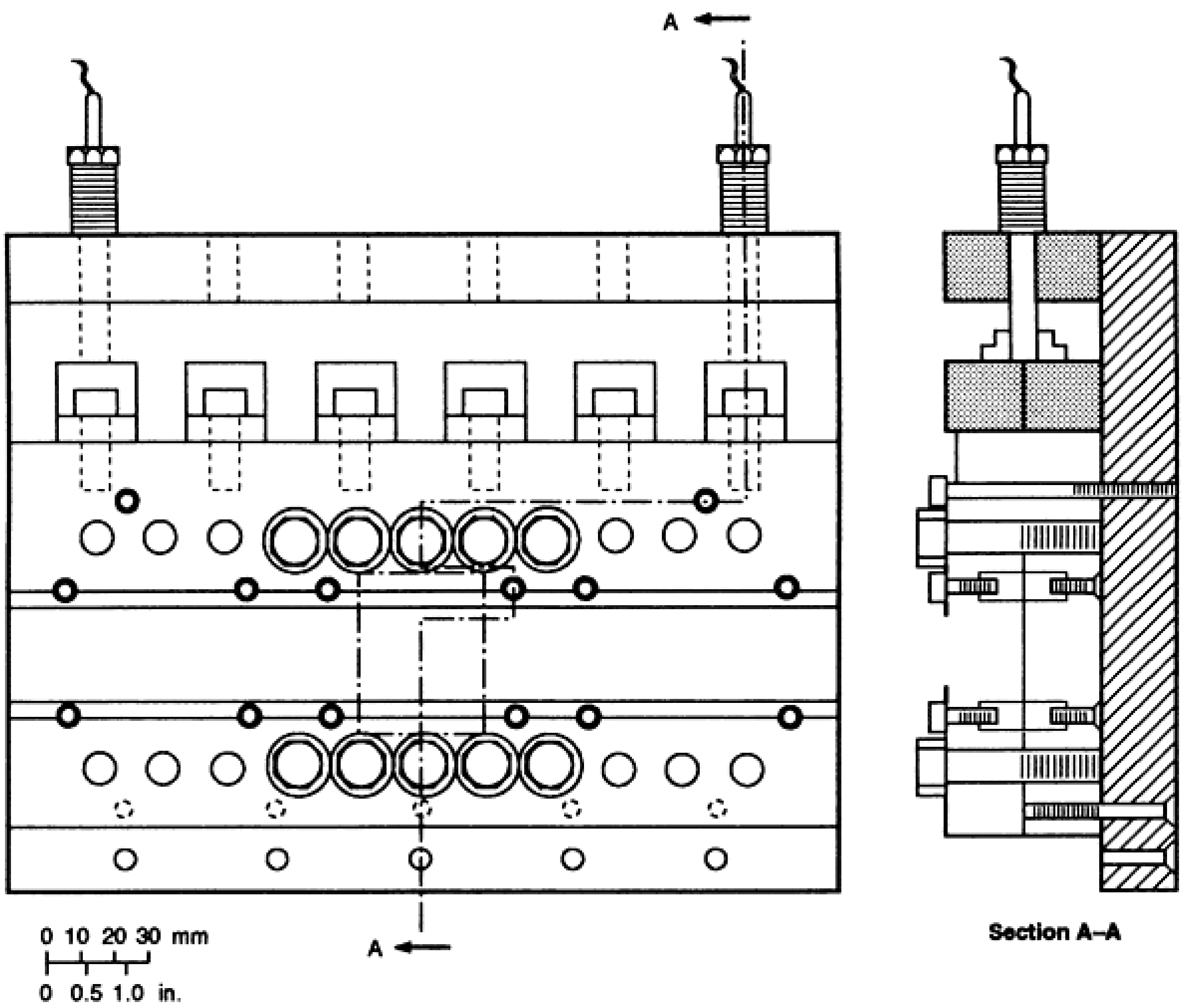

FIG. 14 RELATIONSHIP OF 50 BY 50 MM (2 BY 2 IN.) SHEET SPECIMEN TO APPARATUS USED IN SIGMAJIG TEST. SOURCE: REF 13

References cited in this section

6. "STANDARD WELDING TERMS AND DEFINITIONS," ANSI/AWS A3.0-89, AMERICAN WELDING SOCIETY, 1989
7. "WELDABILITY TESTING," CHAPTER 4, *WELDING HANDBOOK,* VOL 1, 8TH ED., AMERICAN WELDING SOCIETY, 1987
8. R.D. STOUT AND W.D. DOTY, *WELDABILITY OF STEELS,* WELDING RESEARCH COUNCIL, 1978
9. R.D. STOUT, *WELDABILITY OF STEELS,* WELDING RESEARCH COUNCIL, 1987
10. C.D. LUNDIN, A.C. LINGENFELTER, G.E. GROTKE, G.G. LESSMANN, AND S.J. MATTHEWS, THE VARESTRAINT TEST, *WELD. RES. COUNC. BULL.,* NO. 280, AUG 1982
11. R.D. CAMPBELL, "AN INVESTIGATION INTO THE PHYSICAL METALLURGY, WELDING METALLURGY, HOT CRACKING AND WELD POOL SHAPE OF FERRITIC STAINLESS STEELS," PH.D. THESIS, RENSSELAER POLYTECHNIC INSTITUTE, TROY, NEW YORK, 1987
12. W. LIN, J.C. LIPPOLD, AND W.A. BAESLACK, AN EVALUATION OF HEAT-AFFECTED ZONE LIQUATION CRACKING SUSCEPTIBILITY, PART I: DEVELOPMENT OF A METHOD FOR QUANTIFICATION, *WELD. J.,* VOL 72 (NO. 4), APRIL 1993, P 135-S TO 153-S
13. G.M. GOODWIN, DEVELOPMENT OF A NEW HOT-CRACKING TEST--THE SIGMAJIG, *WELD. J.,* VOL 66 (NO. 2), FEB 1987, P 33-S TO 38-S

Richard D. Campbell, Joining Services, Inc.; Daniel W. Walsh, California Polytechnic State University

Weldability Tests for Evaluating Weld Pool Shape, Fluid Flow, and Weld Penetration

Weld pool shape, fluid flow, and the subsequent weld penetration and weld profile are important characteristics in industrial welding applications. Improvements in material processing have resulted in alloys with much "cleaner" chemical analyses. Small changes in the amounts of certain trace elements have been shown to cause dramatic variations in weld pool shape and penetration (Ref 14). Predictable and reproducible weld shape and weld penetration are critical in many applications. As a result, several weldability tests have been developed to evaluate base material response characteristics. Although these tests are not as standardized as the others described above, they are becoming increasingly important, both for alloy development and for production welding applications.

Weld Penetration Tests. Individual companies have developed tests for evaluating the penetration behavior of base metals. A simple weld test utilizes a bar stock sample on which an autogenous gas-tungsten arc, bead-on-plate, partial-penetration weld is produced (Fig. 15). A single set of welding parameters and the same heat sinking are used to test different heats of the same base materials. After welding, the sample is metallographically cross sectioned and etched. Measurements are made of the weld bead width (*W*) and depth of penetration (*D*); from these dimensions, a depth-to-width ratio (*D*/*W*) is calculated.

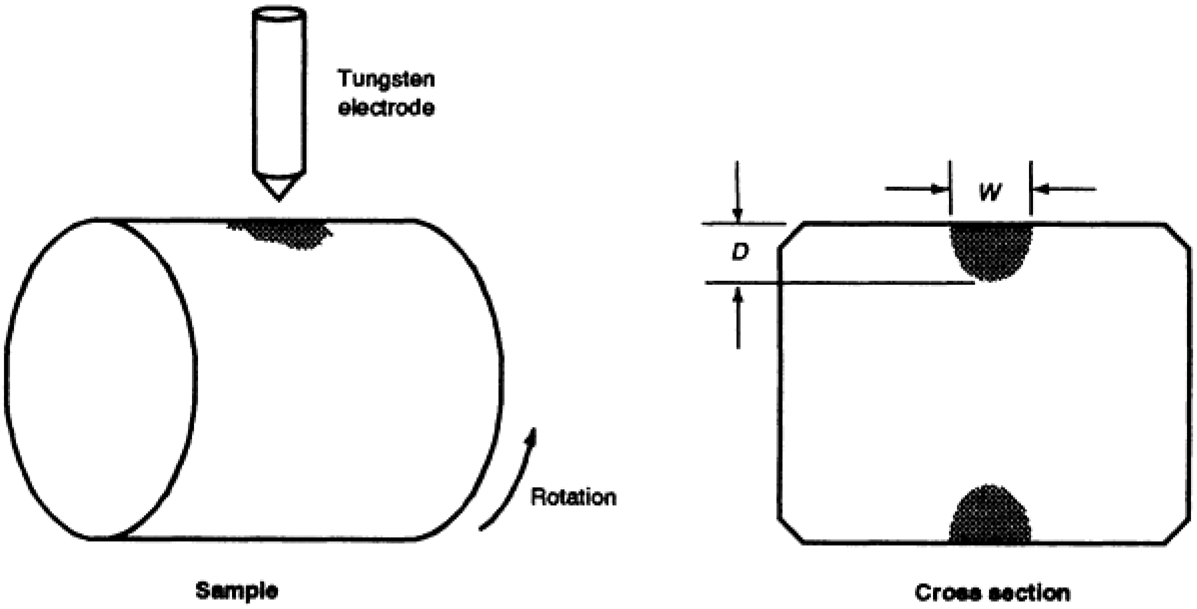

FIG. 15 POSITION OF WELD RELATIVE TO SOLID BAR STOCK IN PARTIAL PENETRATION WELD TEST. SOURCE: REF 14

The *D*/*W* ratio can be used to characterize the weld penetration characteristics of base materials and different heats of the same materials. For production applications, this provides information about the material response before components reach the production welding area. The advantage of this partial-penetration test is that measurements of the solidified weld bead are made. These data characterize the response of the material to all forces that affect weld bead shape, without the influence of a backside or underbead.

Full-penetration weld tests can also be used. These tests are applicable when the weldability test specimen is similar in geometry to the production weldment. However, full-penetration weld shape is complicated because, in some cases, three-dimensional heat flow applies, while in other cases, two-dimensional heat flow applies. Heat buildup on the backside, weld joint shape, and-heat sinking dramatically affect the final weld shape. As a result, full-penetration weldability tests are useful only when the production geometry is used in the test (Ref 14).

Weld Pool Shape Tests. Weld pool shape is determined by a complex set of factors that include process and material variables (Ref 15, 16, 17). Pool shape determination is of interest for basic research and for production applications. Walsh, Campbell, and Savage (Ref 18) pointed out that determinations of pool shape based on metallographic cross sections are incomplete, and subject to error even in stationary welds. Any arc or weld pool motion that sweeps out a width or length greater than the instantaneous pool dimensions cannot be detected on metallographic cross sections. Many critical geometric features of the weld pool could be overlooked if a true three-dimensional measurement was not made.

Some work has been done to characterize the three-dimensional shape of weld pools (Ref 19, 20). However, techniques based on gravity and gunshots have drawbacks, principally lack of speed and contamination, respectively. A new weldability test, called an impulse decanting test, is shown in Fig. 16. This test utilizes a high-pressure jet of argon gas that is impulse-released through the bottom of a modified GTAW torch to decant, or expel, the liquid weld pool. A control box contains the necessary timing circuitry and system relays to extinguish the welding arc and fire the solenoid valve on an accumulator tank that holds the argon.

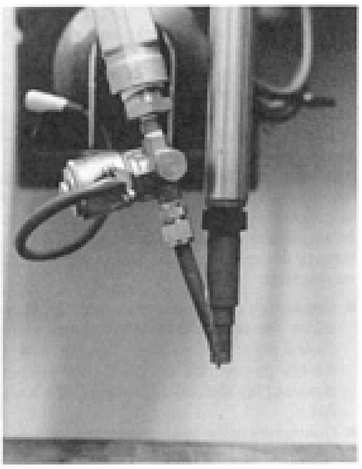

FIG. 16 TYPICAL SETUP USED TO MEASURE WELDABILITY USING THE IMPULSE DECANTING TEST. SOURCE: REF 11

The decanted weld pool shape can be evaluated by two methods:

- USING SILICONE RUBBER TO REPLICATE THE WELD POOL CAVITY
- VISUALLY MEASURING THE WELD POOL CAVITY DIMENSIONS ALONG THE SAMPLE TOP SURFACE, AND ALSO ON A METALLOGRAPHIC TRANSVERSE CROSS SECTION OF THE CAVITY

The silicone rubber replicas are the best method (Fig. 17), because they accurately provide the weld pool length, width, depth, volume, and surface area. This test method is extremely valuable to investigators of weld pool modeling.

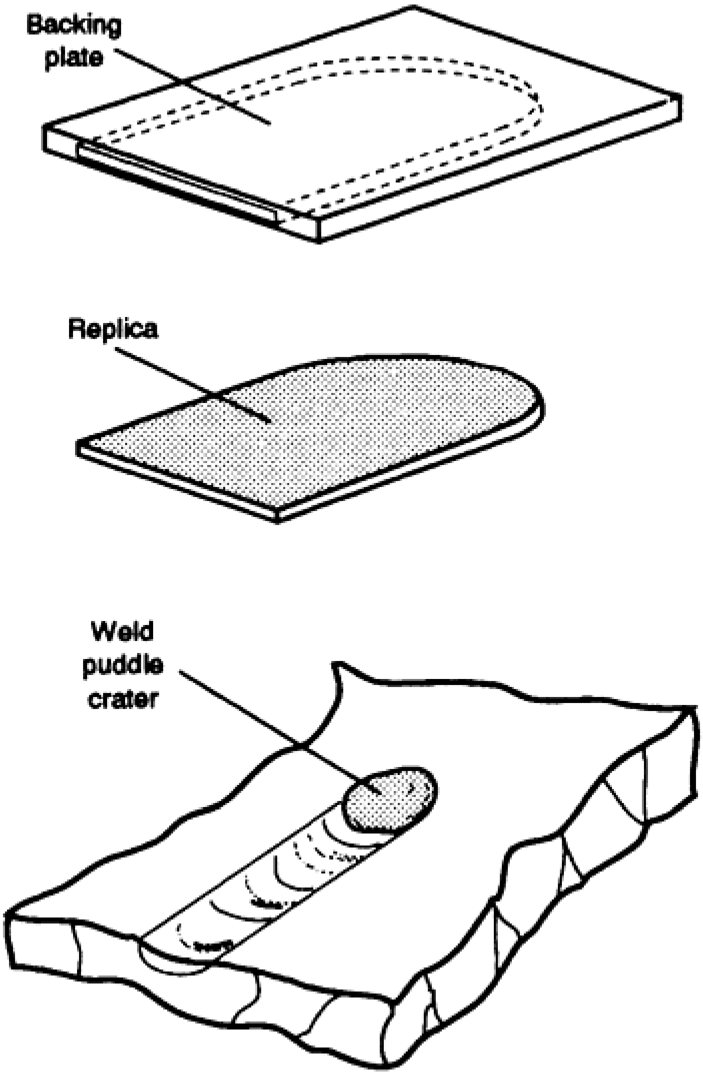

FIG. 17 SILICONE RUBBER REPLICATION TECHNIQUE USED TO EVALUATE DECANTED WELD POOL SHAPE. SOURCE: REF 21

References cited in this section

11. R.D. CAMPBELL, "AN INVESTIGATION INTO THE PHYSICAL METALLURGY, WELDING METALLURGY, HOT CRACKING AND WELD POOL SHAPE OF FERRITIC STAINLESS STEELS," PH.D. THESIS, RENSSELAER POLYTECHNIC INSTITUTE, TROY, NEW YORK, 1987
14. P. BURGARDT AND R.D. CAMPBELL, CHEMISTRY EFFECTS ON STAINLESS STEEL WELD PENETRATION, CHAPTER 13, *FERROUS ALLOY WELDMENTS,* TRANS TECH PUBLICATIONS, SWITZERLAND, 1992
15. D.W. WALSH AND W.F. SAVAGE, THE MECHANISM OF MINOR ELEMENT INTERACTION IN AUTOGENOUS WELD POOLS, *ADVANCES IN WELDING TECHNOLOGY INTERNATIONAL CONF. PROC.* (GATLINBURG, TN), 18 MAY 1986, AMERICAN SOCIETY FOR METALS, P 59-63
16. C.R. HEIPLE AND J.R. ROPER, EFFECTS OF MINOR ELEMENTS ON GTAW FUSION ZONE SHAPE, *TRENDS IN WELDING RESEARCH IN THE UNITED STATES,* CONF. PROC. (NEW ORLEANS, LA), NOV 1981
17. S.S. GLICKSTEIN AND W. YENISCAVICH, "A REVIEW OF MINOR ELEMENT EFFECTS ON THE WELDING ARC AND WELD PENETRATION," BULLETIN 226, WELDING RESEARCH COUNCIL, MAY 1977

18. D.W. WALSH, R.D. CAMPBELL AND W.F. SAVAGE, UNPUBLISHED RESEARCH, RENSSELAER POLYTECHNIC INSTITUTE, TROY, NY, 1986
19. F.J. ZANNER, PH.D. THESIS, RENSSELAER POLYTECHNIC INSTITUTE, TROY, NEW YORK, JUNE 1967
20. D.W. WALSH, "MINOR ELEMENT EFFECTS ON GAS TUNGSTEN ARC WELDS," PH.D. THESIS, RENSSELAER POLYTECHNIC INSTITUTE, TROY, NEW YORK, JUNE 1986
21. W.F. SAVAGE, E.F. NIPPES, AND F.J. ZANNER, *WELD. J.*, VOL 57 (NO. 7), JULY 1978, P 201

Weldability Testing

Richard D. Campbell, Joining Services, Inc.; Daniel W. Walsh, California Polytechnic State University

Gleeble Testing

The Gleeble, a thermomechanical testing device, has been used in a wide range of applications involving study of a variety of material problems. These include, but are not limited to, studies of hot ductility (Ref 22, 23, 24, 25), mechanical properties of the weld HAZ (Ref 26, 27, 28, 29, 30, 31, 32, 33, 34, 35), continuous cooling transformation diagrams (Ref 36, 37, 38), stress relief cracking (Ref 39), precipitation kinetics (Ref 40), basic weldability (Ref 41, 42, 43, 44, 45), strain age cracking (Ref 46, 47), constitutional liquation (Ref 48), liquid metal embrittlement (LME) (Ref 49, 50), and solidification of austenitic stainless steel weldments (Ref 51, 52). The Gleeble system permits simulation of virtually any thermomechanical process. The device is used in a variety of applications and is in no way limited by its historical application to welding.

The Gleeble test method is shown schematically in Fig. 18. Stainless steel samples being tested under load and no-load conditions are shown in Fig. 19(a) and 19(b), respectively. The Gleeble is a fully computer interfaced system, readily programmed to provide reference signals for closed loop control of both thermal and mechanical operations. Heating is accomplished by the flow of low-frequency alternating current in the specimen, and heating rates up to 20,000 °C/s (36,000 °F/s) per second can be obtained (Ref 54). The feedback signal necessary for closed-loop control is normally obtained from a fine wire thermocouple percussion welded to the specimen surface. However, for temperatures above the operating limits of thermocouples, a radiation pyrometer can be employed at some sacrifice in the overall response time of the system.

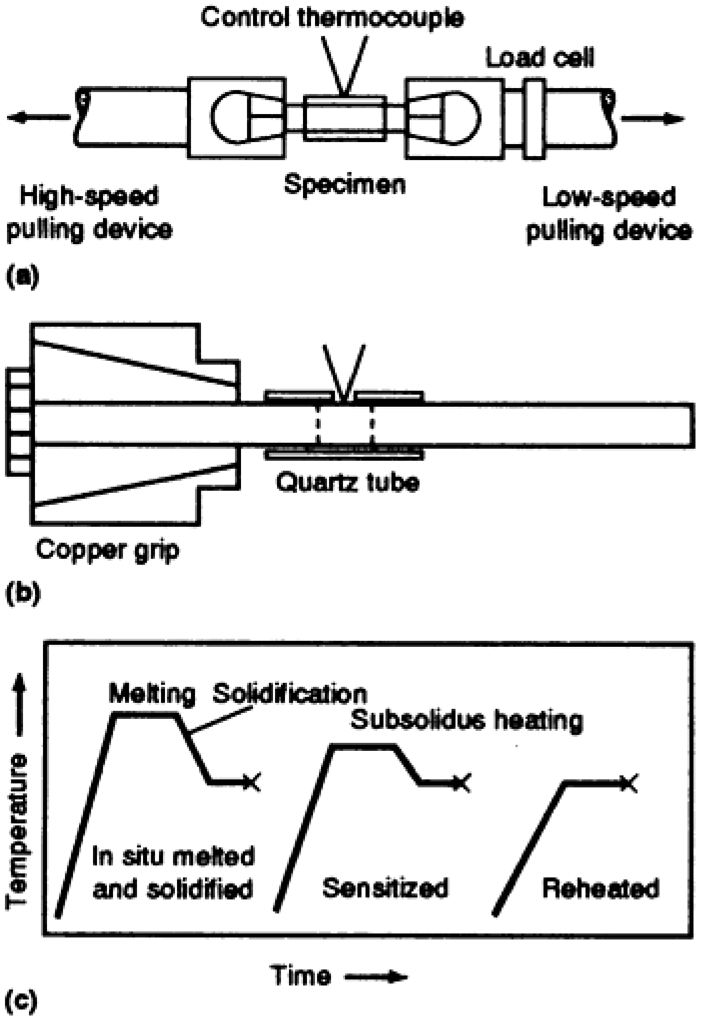
Control thermocouple
Load cell
High-speed pulling device
Specimen
Low-speed pulling device
(a)
Quartz tube
Copper grip
(b)
Temperature
Melting
Solidification
Subsolidus heating
In situ melted and solidified
Sensitized
Reheated
Time
(c)

FIG. 18 GLEEBLE TEST METHOD. (A) PRIMARY COMPONENTS. (B) CLOSE-UP VIEW OF RESISTANCE HEATER. (C) PROGRAMMED THERMAL CYCLE. SOURCE: REF 53

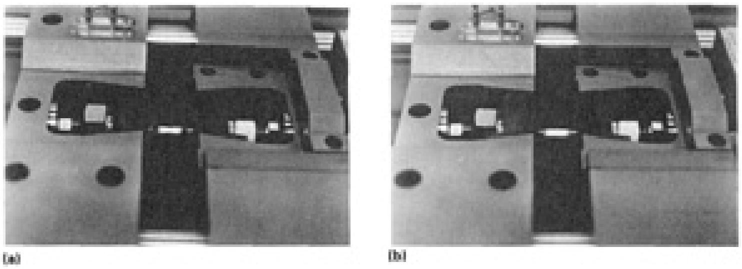

FIG. 19 GLEEBLE TESTING OF A 6 MM (0.236 IN.) DIAMETER STAINLESS STEEL BAR AT 1100 °C (2010 °F). (A) SAMPLE UNDER NO-LOAD CONDITION (NOTE UNIFORMITY OF REPRODUCIBLE THERMAL GRADIENT). (B) SAMPLE UNDER LOAD AT ELEVATED TEMPERATURES TO EVALUATE DUCTILITY

The Gleeble has a long and proven history as a tool for both the study of metallurgical phenomena at the research level and for materials testing to predict service behavior at the production level. A major advantage of the Gleeble is that it generates large volumes of microstructure that simulate small, hard-to-study regions in actual weldments.

Gleeble test samples take on an unlimited variety of sizes and shapes, based on material availability and project application. The following is a discussion of round bar samples, as they are the most typical samples studied. The concepts discussed can be extended to a variety of geometries.

A given test sample is a system in dynamic thermal equilibrium with its surroundings. Thus, to maintain any given temperature profile, heat is supplied by the electrical current passing through the specimen, and heat is lost by three mechanisms:

- CONDUCTION ALONG THE SPECIMEN ITSELF, IN WHICH THE HEAT IS TRANSFERRED TO THE WATER-COOLED JAWS. THE RATE OF THIS HEAT FLOW IS DIRECTLY PROPORTIONAL TO BOTH THE THERMAL GRADIENT AND THE THERMAL DIFFUSIVITY OF THE MATERIAL.
- CONVECTION IN THE GASEOUS ATMOSPHERE, WHICH REMOVES HEAT AND DISSIPATES IT TO THE SURROUNDINGS BY MIXING.
- RADIATION EMITTED BY THE SPECIMEN. THE AMOUNT OF HEAT LOST THROUGH THIS MODE IS PROPORTIONAL TO THE FOURTH POWER OF THE ABSOLUTE TEMPERATURE, AND THEREFORE BECOMES MORE IMPORTANT AT HIGH TEMPERATURES.

Thus, the thermal profile generated by a given peak temperature, jaw separation, specimen diameter, and material type is determined by a complex interplay of these mechanisms in establishing a thermal balance.

The Gleeble specimen may be treated as a uniform cylindrical bar with internal generation of heat. If convection and radiation losses are neglected, the principle of the conservation of energy leads to the following relationships for the instantaneous temperature distribution in the bar:

$$\frac{\partial T}{\partial t} = \frac{K}{rC_p}\nabla^2 T + \frac{t^2}{rC_p w^2 s} \qquad \textbf{(EQ 4)}$$

where C_p is the heat capacity at constant pressure; ∇ is the del operator; I is the current; ω is the area in cross section; σ is the electrical conductivity; K is the thermal conductivity; T is the temperature; ρ is the density; and t is the time.

If the system is simplified by limiting the heat flow to one dimension, Eq 4 becomes:

$$\frac{\partial T}{\partial t} = \frac{K}{rC_p}\frac{\partial^2 T}{\partial x^2} + \frac{t^2}{rC_p w^2 s} \qquad \textbf{(EQ 5)}$$

This relationship closely approximates the behavior of a resistance-heated specimen. If the temperature distribution within the bar is assumed to be steady state, Eq 5 reduces to the Poisson relation:

$$\frac{\partial^2 T}{\partial x^2} = \frac{-I^2}{Kw^2 s} \qquad \textbf{(EQ 6)}$$

Therefore, if radiation and convection losses are ignored, the temperature distribution in the bar will be a parabolic one. This can be demonstrated by double integration of Eq 6 to obtain:

$$T(x) = \frac{-I^2}{Kw^2 s}x^2 + c' + c'' \qquad \textbf{(EQ 7)}$$

The constants of c' and c" can be determined for specific boundary conditions. Solutions to this equation exist in the literature (Ref 55) and have been applied to the determination of the lengths over which the temperature of the bar lies, between the maximum, $T_{\max}$, and some arbitrary lower temperature, T_0. This length is given by the following:

$$L = L_o \left(\frac{\Delta T}{T_{\max} - T_o} \right)^{1/2} \qquad \textbf{(EQ 8)}$$

where L is the length at or above arbitrary temperature; L_0 is the jaw separation; T_0 is the temperature at the jaw; and ΔT is the difference between peak temperature and arbitrary temperature.

However, problems exist for this simplified treatment because experimental results do not fit the model well. Several factors contribute to this. The thermal conductivity, K, is not a constant, and being itself a function of temperature can be presented as:

$$K = K_0(1 - BT) \qquad \textbf{(EQ 9)}$$

where K_0 is the thermal conductivity at absolute zero if conductivity is treated linearly.

However, because the value of b is small, an average value of K can be used over a small range of temperatures without generating an excessively large error.

The mathematics of the system become more complex when treatments for either convection or radiation heat transfer are included. Heat loss by either mechanism will tend to flatten the parabolic profile predicted by Eq 7. Convection losses can be represented by adding a single term to the Poisson equation to obtain:

$$\frac{K}{rC_p}\frac{\partial^2 T}{\partial x^2} + \frac{-I^2}{rC_p w^2 s} - \frac{hg}{rC_p w}(T - T_o) = 0 \qquad \textbf{(EQ 10)}$$

where h is the convective heat transfer coefficient and γ is the dependent constant relative to geometry.

The inclusion of another temperature-dependent term makes the solution to the differential equation more complex, and leads to solutions involving hyperbolic functions. The addition of a term for radiation losses further complicates the solution because the radiative heat loss, Q_{12}, is proportional to the fourth power of absolute temperature:

$$Q_{12} = F_{12}\alpha(T_1^4 - T_2^4) \qquad \textbf{(EQ 11)}$$

where Q_{12} is the heat transfer from body 1 to body 2; F_{12} is the view factor, body 1 to body 2; T_1 is the temperature of body 1; and α is the Stefan-Boltzmann constant. The constants h, in the convective solution, and α, in the radiative solution, are small, and high temperatures are required before convection and radiation play a major role in determining the temperature profile in the bar.

Walsh, Savage and Cieslak (Ref 56) reported an empirical study of carbon and stainless steel that related test parameters to the extent of a uniform temperature region in Gleeble samples. They developed polynomial equations that describe longitudinal and radial thermal profiles that exist in test samples under a wide variety of operator-controlled conditions. At any particular peak temperature, a span in which the temperature lies within a specified percentage of the peak temperature can be generated by a judicious choice of operating conditions and specimen geometries.

Gleeble testing has been used to study a variety of phenomenon in electrically conductive materials. However, the discussion of its application to weldability testing will be limited to studies of hot cracking and hot ductility. Information about other applications of Gleeble testing is available in the References cited in this article.

Hot Ductility Testing. The development of high-performance materials and increasingly rigid standards for quality, coupled with the jux-taposition of very different materials in engineering assemblies, has led to significant increases in the difficulty associated with producing satisfactory weldments. The temperatures and steep thermal gradients associated with welding produce high stresses and complex transient stress patterns. These factors often combine to produce cracking in the weld and HAZ region. The rapid thermal cycle associated with the welding process produces a plethora of microstructures in the weld/HAZ region, and weld understood metallurgical transformations associated with heating and cooling become complex under the nonequilibrium conditions associated with the welding process.

Goodwin (Ref 57) has pointed out that hot cracking is an often-studied, yet poorly understood phenomena. Hot cracking results when a susceptible microstructure is subjected to tensile stresses, and it occurs while liquid and solid phases coexist. A successful hot cracking test must reproduce combinations of restraint and microstructure that can be used to differentiate the crack sensitivity of different materials (Ref 57). Several hot cracking tests are described in the section "Weldability Tests for Evaluating Cracking Susceptibility" in this article.

Borland (Ref 58) described features required of an ideal hot crack test--it should be inexpensive, simple, brief, directly correlatable to fabrication parameters and service conditions, reproducible, free from operator error, and applicable to all welding processes. In addition to these criteria, the test should also produce easily interpretable results, be amenable to modeling, be quantitative, and above all, be universally accepted (Ref 57). It is very difficult to conceive of a single test that could satisfy all these criteria; however, many investigators have found that the combination of Varestraint or Varestraint-like testing and Gleeble testing yields results that can be used to provide a state-of-the-art characterization of weldability. (The Varestraint test is described in the section "Varestraint Test (Hot Crack Test)" in this article.)

Gleeble testing provides a complete characterization of hot ductility related to hot crack susceptibility. In the Gleeble test, small cylindrical tensile samples are fractured during heating or during cooling from specified temperatures. Reduction-in-area and ultimate tensile strength data are used to generate strength and ductility curves upon cooling and upon heating. A number of factors are important to the test, including heating and cooling rates, peak temperatures, hold times, and stroke rates. A number of investigators have examined the effects of changes to these parameters and found that the peak temperature of the test is by far the most crucial. The magnitude of effects that changes to other parameters produce is a strong function of material and material condition.

Weiss, Grotke, and Stickler (Ref 25) discuss the physical metallurgy of hot ductility testing and the use of the Gleeble in studying alloy weldability. Hot cracking is caused by the presence of persistent liquid films on grain boundaries. These films are unable to accommodate strains that are produced by mechanical restraint and/or transient thermal distributions during cooling. Hot cracking will not occur unless a susceptible microstructure is exposed to critical level of restraint. Several characteristic temperatures that can be used to classify material weldability have been identified using the Gleeble. In hot ductility testing, the nil temperature (NDT) is the temperature upon heating at which the material ductility goes to zero; in essence, the temperature at which all surfaces are coated by a thin liquid layer. The nil strength temperature (NST) is the temperature upon heating at which the strength of a material drops essentially to zero; in essence, the temperature at which liquid layers thicken to the point that minimal capillarity exists. The ductility recovery temperature is the temperature upon cooling at which an arbitrary ductility is observed.

Investigators have used a wide variety of criteria to interpret hot ductility curves produced using the Gleeble. Nippes, Savage, and Grotke (Ref 23) published a classification of hot ductility curves encountered during testing of stainless steels and nickel-base alloys. The criteria most widely used include minimum arbitrary ductility, recovery rate of ductility, recovery rate of ultimate strength, zero-ductility range, and the presence of ductility dips (Ref 31). Kreischer (Ref 24) published a critical review of hot ductility testing and suggested that:

- WHEN THE NST IS USED AS THE PEAK TEMPERATURE IT PROVIDES MUCH MORE ACCURATE INDICATION OF HOT CRACK SUSCEPTIBILITY THAN THE NDT
- ALLOYS EXHIBITING HIGH DUCTILITIES UPON HEATING AND RAPID RECOVERY OF DUCTILITY UPON COOLING FROM THE NST ARE CRACK RESISTANT
- ALLOYS WITH HIGH TENSILE STRENGTH AT ELEVATED TEMPERATURES CAN BE CRACK RESISTANT DESPITE LOW DUCTILITIES
- ALLOYS WITH RELATIVELY LOW TENSILE STRENGTHS AT ELEVATED TEMPERATURE, IN CONCERT WITH ONLY MODERATE DUCTILITIES, ARE NORMALLY CRACK SENSITIVE

Lin, Lippold, and Baeslack (Ref 59, 60) have detailed a methodology for quantifying crack susceptibility using the Gleeble hot ductility test.

References cited in this section

22. E.F. NIPPES AND W.F. SAVAGE, AN INVESTIGATION OF THE HOT DUCTILITY OF HIGH-TEMPERATURE ALLOYS, *WELD. J.*, VOL 34 (NO. 4), 1955, P 183-S TO 196-S

23. E.F. NIPPES, W.F. SAVAGE, AND G. GROTKE, "FURTHER STUDIES OF THE HOT DUCTILITY OF HIGH-TEMPERATURE ALLOYS," BULLETIN SERIES VOL 33 (NO. 2), WELDING RESEARCH COUNCIL, 1957
24. C.H. KREISCHER, A CRITICAL ANALYSIS OF THE WELD HEAT-AFFECTED ZONE HOT DUCTILITY TEST, *WELD. J.,* VOL 42 (NO. 2), 1963, P 49-S TO 59-S
25. B. WEISS, B.E. GROTKE, AND R. STICKLER, PHYSICAL METALLURGY OF HOT DUCTILITY TESTING, *WELD. J.,* VOL 49 (NO. 10), 1970, P 471-S TO 487-S
26. E.F. NIPPES, W.F. SAVAGE, AND J.M. PAEZ, IMPACT CHARACTERISTICS OF HEAT-AFFECTED ZONES IN MN-MO ARMOR STEELS, *WELD. J.,* VOL 39 (NO. 1), 1960, P 31-S TO 39-S
27. K.C. WU, A STUDY OF THE WELD HEAT-AFFECTED ZONE OF CENTRIFUGALLY CAST 5% CHROMIUM STEEL, *WELD. J.,* VOL 42 (NO. 9), 1963, P 392-S TO 396-S
28. E.F. NIPPES AND E.W. EMMERICH, HEAT-AFFECTED ZONE STUDY OF AN EUTECTOID ULTRA-HIGH STRENGTH STEEL, *WELD. J.,* VOL 42 (NO. 12), 1963, P 547-S TO 556-S
29. W.A. PETERSON, WELD HEAT-AFFECTED ZONE OF 18% NICKEL MARAGING STEEL, *WELD. J.,* VOL 43 (NO. 9), 1964, P 428-S TO 432-S
30. W.F. SAVAGE AND B.M. KRANTZ, AN INVESTIGATION OF HOT CRACKING IN HASTELLOY X, *WELD. J.,* VOL 45 (NO. 1), 1966, P 13-S TO 25-S
31. W.A. OWCZARSKI, D.S. DUVAL, AND C.P. SULLIVAN, A MODEL FOR HEAT-AFFECTED ZONE CRACKING IN NICKEL-BASE SUPERALLOYS, *WELD. J.,* VOL 45 (NO. 4), 1966, P 145-S TO 155-S
32. L.P. CONNOR, A.M. RATHBONE, AND J.H. GROSS, EFFECTS OF COMPOSITION ON THE HEAT-AFFECTED ZONE TOUGHNESS OF CONSTRUCTIONAL ALLOY STEELS, *WELD. J.,* VOL 46 (NO. 5), 1967, P 217-S TO 234-S
33. R.A. MUELLER, D.G. HOWDEN, AND F.B. SIMMONS, HAZ THERMOCYCLE AND STRUCTURE SIMULATION, *WELD. J.,* VOL 52 (NO. 10), 1973, P 411-S TO 416-S
34. N.E. HANNERY, EFFECT OF CB ON HAZ DUCTILITY IN CONSTRUCTIONAL HT STEELS, *WELD. J.,* VOL 54 (NO. 5), 1975, P 162-S TO 168-S
35. S.J. MATTHEWS, SIMULATED HEAT-AFFECTED ZONE STUDIES OF HASTELLOY ALLOY B-2, *WELD. J.,* VOL 58 (NO. 3), 1979, P 91-S TO 95-S
36. R.F. FASTINI AND M. SEMCHYSHEN, THE EFFECT OF MOLYBDENUM ON THE CONTINUOUS COOLING TRANSFORMATION OF A 0.40% CARBON STEEL, *ASM TRANS. QUART.,* SEPT 1964
37. D.G. HOWDEN, ISOTHERMAL TRANSFORMATION CHARACTERISTICS OF A LOW-ALLOY STEEL USING A WELD THERMAL CYCLE SIMULATOR, *CAN. METALL. QUART.,* VOL 6 (NO. 3), 1967
38. K.C. WU, CONTINUOUS-COOLING TRANSFORMATION DIAGRAMS OF TI-8A1-MO-1V, *TRANS. ASM,* VOL 61, 1968, P 621-626
39. C.F. MEITZNER AND A.W. PENSE, STRESS-RELIEF CRACKING IN LOW-ALLOY STEEL WELDMENTS, *WELD. J.,* VOL 48 (NO. 10), 1969, P 431-S TO 440-S
40. A.W. DIX AND W.F. SAVAGE, SHORT TIME AGING CHARACTERISTICS OF INCONEL X-750, *WELD. J.,* VOL 52 (NO. 3), 1973, P 135-S TO 144-S
41. S.J. MATTHEWS, WELDABILITY OBSERVATIONS IN 70/30 COPPER-NICKEL TUBESHEET MATERIAL, *WELD. J.,* VOL 47 (NO. 4), 1968, P 155-S TO 161-S
42. K.C. WU AND T.A. KRINKE, WELDABILITY STUDIES FOR AFC-77--PART 1: MECHANICAL PROPERTIES, *WELD. J.,* VOL 47 (NO. 6), 1968, P 332-S TO 336-S
43. W.A. PETERSON, WELDABILITY OF A CHROMIUM STRENGTHENED COPPER-NICKEL ALLOY, *WELD. J.,* VOL 48 (NO. 10), 1969, P 425-S TO 430-S
44. J. GORDINE, THE WELDABILITY OF SOME ARTIC-GRADE LINE-PIPE STEELS, *WELD. J.,* VOL 56 (NO. 7), 1977, P 201-S TO 210-S
45. W.F. SAVAGE, E.F. NIPPES, AND J.E. CASTERAS, EFFECT OF ALLOYING ADDITIONS ON THE WELDABILITY OF 70 CU-30 NI, *WELD. J.,* VOL 57 (NO. 12), 1978, P 375-S TO 382-S

46. W.P. HUGHES AND T.F. BERRY, A STUDY OF THE STRAIN-AGE CRACKING CHARACTERISTICS IN WELDED RENE 41--PHASE 1, *WELD. J.,* VOL 46 (NO. 8), 1967, P 361-S TO 370-S
47. A.W. DIX AND W.F. SAVAGE, FACTORS INFLUENCING STRAIN-AGE CRACKING IN INCONEL X-750, *WELD. J.,* VOL 50 (NO. 6), 1971, P 247-S TO 252-S
48. J.J. PEPE AND W.F. SAVAGE, EFFECTS OF CONSTITUTIONAL LIQUATION IN 18-NI MARAGING STEEL WELDMENTS, *WELD. J.,* VOL 46 (NO. 9), 1967, P 411-S TO 422-S
49. W.F. SAVAGE, E.F. NIPPES, AND M.C. MUSHALA, COPPER-CONTAMINATION CRACKING IN THE WELD HEAT-AFFECTED ZONE, *WELD. J.,* VOL 57 (NO. 5), 1978, P 237-S TO 245-S
50. W.F. SAVAGE, E.F. NIPPES, AND M.C. MUSHALA, LIQUID-METAL EMBRITTLEMENT OF THE HEAT-AFFECTED ZONE BY COPPER CONTAMINATION, *WELD. J.,* VOL 57 (NO. 8), 1978, P 237-S TO 245-S
51. J.C. LIPPOLD AND W.F. SAVAGE, SOLIDIFICATION OF AUSTENITIC STAINLESS STEEL WELDMENTS: PART 1--A PROPOSED MECHANISM, *WELD. J.,* VOL 58 (NO. 12), 1979, P 362-S TO 374-S
52. J.C. LIPPOLD AND W.F. SAVAGE, SOLIDIFICATION OF AUSTENITIC STAINLESS STEEL WELDMENTS: PART 2--THE EFFECT OF ALLOY COMPOSITION ON FERRITE MORPHOLOGY, *WELD. J.,* VOL 59 (NO. 2), 1980, P 48-S TO 58-S
53. COMPUTERIZED THERMAL AND MECHANICAL TESTING MACHINE A SIGNIFICANT ASSET IN RESEARCH, DEVELOPMENT AND PRODUCTION, *INDUSTRIAL HEATING,* VOL 53 (NO. 12), DEC 1986
54. H. FERGUSON, METALWORKING PROCESS SIMULATIONS SAVE PRODUCTION HEADACHES, *MET. PROG.,* SEPT 1986, P 39-42
55. D.J. WIDGERY, THE DESIGN AND USE OF RESISTANCE HEATED WELD THERMAL SIMULATORS, *PROC. CONF. WELD THERMAL SIMULATORS FOR RESEARCH AND PROBLEM SOLVING,* THE WELDING INSTITUTE, CAMBRIDGE, ENGLAND, APRIL 1972
56. D.W. WALSH, M.J. CIESLAK, AND W.F. SAVAGE, TEMPERATURE MEASUREMENTS IN RESISTANCE HEATED SPECIMENS: LONGITUDINAL GRADIENTS, *WELD. J.,* VOL 65 (NO. 8), 1986, P 184-S TO 192-S
57. G.M. GOODWIN, *PROC. 1ST JAPAN-U.S. SYMPOSIUM ON ADVANCES IN WELDING METALLURGY* (SAN FRANCISCO, CA), JUNE 1990, P 27-39
58. J.C. BORLAND, *BR. WELD. J.,* VOL 7 (NO. 8), 1961, P 508-512
59. W. LIN, W.A. BAESLACK, AND J.C. LIPPOLD, HOT DUCTILITY TESTING OF HIGH-STRENGTH, LOW-EXPANSION SUPERALLOYS, *RECENT TRENDS IN WELDING AND TECHNOLOGY LNT. CONF. PROC.* (GATLINBURG, TN), 14 MAY 1989, ASM INTERNATIONAL, P 609-614
60. W. LIN, W.A. BAESLACK, AND J.C. LIPPOLD, A METHOD FOR QUANTIFYING HAZ LIQUATION CRACKING SUSCEPTIBILITY USING THE GLEEBLE HOT-DUCTILITY TEST, *PROC. INT. SYMP. PHYSICAL SIMULATION,* DELFT, 1992, P 93-99

Weldability Testing

Richard D. Campbell, Joining Services, Inc.; Daniel W. Walsh, California Polytechnic State University

---

References

1. "WELDING TECHNOLOGY," CHAPTER 12, *WELDING HANDBOOK,* VOL 1, 8TH ED., AMERICAN WELDING SOCIETY, 1987
2. "STANDARD METHODS FOR MECHANICAL TESTING OF WELDS," ANSI/AWS B4.0-85, AMERICAN WELDING SOCIETY, 1985

3. *MECHANICAL TESTING,* VOL 8, *ASM HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1985
4. *METALLOGRAPHY AND MICROSTRUCTURES,* VOL 9, *ASM HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1985
5. *NONDESTRUCTIVE EVALUATION AND QUALITY CONTROL,* VOL 17, *ASM HANDBOOK,* ASM INTERNATIONAL, 1989
6. "STANDARD WELDING TERMS AND DEFINITIONS," ANSI/AWS A3.0-89, AMERICAN WELDING SOCIETY, 1989
7. "WELDABILITY TESTING," CHAPTER 4, *WELDING HANDBOOK,* VOL 1, 8TH ED., AMERICAN WELDING SOCIETY, 1987
8. R.D. STOUT AND W.D. DOTY, *WELDABILITY OF STEELS,* WELDING RESEARCH COUNCIL, 1978
9. R.D. STOUT, *WELDABILITY OF STEELS,* WELDING RESEARCH COUNCIL, 1987
10. C.D. LUNDIN, A.C. LINGENFELTER, G.E. GROTKE, G.G. LESSMANN, AND S.J. MATTHEWS, THE VARESTRAINT TEST, *WELD. RES. COUNC. BULL.,* NO. 280, AUG 1982
11. R.D. CAMPBELL, "AN INVESTIGATION INTO THE PHYSICAL METALLURGY, WELDING METALLURGY, HOT CRACKING AND WELD POOL SHAPE OF FERRITIC STAINLESS STEELS," PH.D. THESIS, RENSSELAER POLYTECHNIC INSTITUTE, TROY, NEW YORK, 1987
12. W. LIN, J.C. LIPPOLD, AND W.A. BAESLACK, AN EVALUATION OF HEAT-AFFECTED ZONE LIQUATION CRACKING SUSCEPTIBILITY, PART I: DEVELOPMENT OF A METHOD FOR QUANTIFICATION, *WELD. J.,* VOL 72 (NO. 4), APRIL 1993, P 135-S TO 153-S
13. G.M. GOODWIN, DEVELOPMENT OF A NEW HOT-CRACKING TEST--THE SIGMAJIG, *WELD. J.,* VOL 66 (NO. 2), FEB 1987, P 33-S TO 38-S
14. P. BURGARDT AND R.D. CAMPBELL, CHEMISTRY EFFECTS ON STAINLESS STEEL WELD PENETRATION, CHAPTER 13, *FERROUS ALLOY WELDMENTS,* TRANS TECH PUBLICATIONS, SWITZERLAND, 1992
15. D.W. WALSH AND W.F. SAVAGE, THE MECHANISM OF MINOR ELEMENT INTERACTION IN AUTOGENOUS WELD POOLS, *ADVANCES IN WELDING TECHNOLOGY INTERNATIONAL CONF. PROC.* (GATLINBURG, TN), 18 MAY 1986, AMERICAN SOCIETY FOR METALS, P 59-63
16. C.R. HEIPLE AND J.R. ROPER, EFFECTS OF MINOR ELEMENTS ON GTAW FUSION ZONE SHAPE, *TRENDS IN WELDING RESEARCH IN THE UNITED STATES,* CONF. PROC. (NEW ORLEANS, LA), NOV 1981
17. S.S. GLICKSTEIN AND W. YENISCAVICH, "A REVIEW OF MINOR ELEMENT EFFECTS ON THE WELDING ARC AND WELD PENETRATION," BULLETIN 226, WELDING RESEARCH COUNCIL, MAY 1977
18. D.W. WALSH, R.D. CAMPBELL AND W.F. SAVAGE, UNPUBLISHED RESEARCH, RENSSELAER POLYTECHNIC INSTITUTE, TROY, NY, 1986
19. F.J. ZANNER, PH.D. THESIS, RENSSELAER POLYTECHNIC INSTITUTE, TROY, NEW YORK, JUNE 1967
20. D.W. WALSH, "MINOR ELEMENT EFFECTS ON GAS TUNGSTEN ARC WELDS," PH.D. THESIS, RENSSELAER POLYTECHNIC INSTITUTE, TROY, NEW YORK, JUNE 1986
21. W.F. SAVAGE, E.F. NIPPES, AND F.J. ZANNER, *WELD. J.,* VOL 57 (NO. 7), JULY 1978, P 201
22. E.F. NIPPES AND W.F. SAVAGE, AN INVESTIGATION OF THE HOT DUCTILITY OF HIGH-TEMPERATURE ALLOYS, *WELD. J.,* VOL 34 (NO. 4), 1955, P 183-S TO 196-S
23. E.F. NIPPES, W.F. SAVAGE, AND G. GROTKE, "FURTHER STUDIES OF THE HOT DUCTILITY OF HIGH-TEMPERATURE ALLOYS," BULLETIN SERIES VOL 33 (NO. 2), WELDING RESEARCH COUNCIL, 1957
24. C.H. KREISCHER, A CRITICAL ANALYSIS OF THE WELD HEAT-AFFECTED ZONE HOT DUCTILITY TEST, *WELD. J.,* VOL 42 (NO. 2), 1963, P 49-S TO 59-S
25. B. WEISS, B.E. GROTKE, AND R. STICKLER, PHYSICAL METALLURGY OF HOT DUCTILITY TESTING, *WELD. J.,* VOL 49 (NO. 10), 1970, P 471-S TO 487-S

26. E.F. NIPPES, W.F. SAVAGE, AND J.M. PAEZ, IMPACT CHARACTERISTICS OF HEAT-AFFECTED ZONES IN MN-MO ARMOR STEELS, *WELD. J.,* VOL 39 (NO. 1), 1960, P 31-S TO 39-S
27. K.C. WU, A STUDY OF THE WELD HEAT-AFFECTED ZONE OF CENTRIFUGALLY CAST 5% CHROMIUM STEEL, *WELD. J.,* VOL 42 (NO. 9), 1963, P 392-S TO 396-S
28. E.F. NIPPES AND E.W. EMMERICH, HEAT-AFFECTED ZONE STUDY OF AN EUTECTOID ULTRA-HIGH STRENGTH STEEL, *WELD. J.,* VOL 42 (NO. 12), 1963, P 547-S TO 556-S
29. W.A. PETERSON, WELD HEAT-AFFECTED ZONE OF 18% NICKEL MARAGING STEEL, *WELD. J.,* VOL 43 (NO. 9), 1964, P 428-S TO 432-S
30. W.F. SAVAGE AND B.M. KRANTZ, AN INVESTIGATION OF HOT CRACKING IN HASTELLOY X, *WELD. J.,* VOL 45 (NO. 1), 1966, P 13-S TO 25-S
31. W.A. OWCZARSKI, D.S. DUVAL, AND C.P. SULLIVAN, A MODEL FOR HEAT-AFFECTED ZONE CRACKING IN NICKEL-BASE SUPERALLOYS, *WELD. J.,* VOL 45 (NO. 4), 1966, P 145-S TO 155-S
32. L.P. CONNOR, A.M. RATHBONE, AND J.H. GROSS, EFFECTS OF COMPOSITION ON THE HEAT-AFFECTED ZONE TOUGHNESS OF CONSTRUCTIONAL ALLOY STEELS, *WELD. J.,* VOL 46 (NO. 5), 1967, P 217-S TO 234-S
33. R.A. MUELLER, D.G. HOWDEN, AND F.B. SIMMONS, HAZ THERMOCYCLE AND STRUCTURE SIMULATION, *WELD. J.,* VOL 52 (NO. 10), 1973, P 411-S TO 416-S
34. N.E. HANNERY, EFFECT OF CB ON HAZ DUCTILITY IN CONSTRUCTIONAL HT STEELS, *WELD. J.,* VOL 54 (NO. 5), 1975, P 162-S TO 168-S
35. S.J. MATTHEWS, SIMULATED HEAT-AFFECTED ZONE STUDIES OF HASTELLOY ALLOY B-2, *WELD. J.,* VOL 58 (NO. 3), 1979, P 91-S TO 95-S
36. R.F. FASTINI AND M. SEMCHYSHEN, THE EFFECT OF MOLYBDENUM ON THE CONTINUOUS COOLING TRANSFORMATION OF A 0.40% CARBON STEEL, *ASM TRANS. QUART.,* SEPT 1964
37. D.G. HOWDEN, ISOTHERMAL TRANSFORMATION CHARACTERISTICS OF A LOW-ALLOY STEEL USING A WELD THERMAL CYCLE SIMULATOR, *CAN. METALL. QUART.,* VOL 6 (NO. 3), 1967
38. K.C. WU, CONTINUOUS-COOLING TRANSFORMATION DIAGRAMS OF TI-8A1-MO-1V, *TRANS. ASM,* VOL 61, 1968, P 621-626
39. C.F. MEITZNER AND A.W. PENSE, STRESS-RELIEF CRACKING IN LOW-ALLOY STEEL WELDMENTS, *WELD. J.,* VOL 48 (NO. 10), 1969, P 431-S TO 440-S
40. A.W. DIX AND W.F. SAVAGE, SHORT TIME AGING CHARACTERISTICS OF INCONEL X-750, *WELD. J.,* VOL 52 (NO. 3), 1973, P 135-S TO 144-S
41. S.J. MATTHEWS, WELDABILITY OBSERVATIONS IN 70/30 COPPER-NICKEL TUBESHEET MATERIAL, *WELD. J.,* VOL 47 (NO. 4), 1968, P 155-S TO 161-S
42. K.C. WU AND T.A. KRINKE, WELDABILITY STUDIES FOR AFC-77--PART 1: MECHANICAL PROPERTIES, *WELD. J.,* VOL 47 (NO. 6), 1968, P 332-S TO 336-S
43. W.A. PETERSON, WELDABILITY OF A CHROMIUM STRENGTHENED COPPER-NICKEL ALLOY, *WELD. J.,* VOL 48 (NO. 10), 1969, P 425-S TO 430-S
44. J. GORDINE, THE WELDABILITY OF SOME ARTIC-GRADE LINE-PIPE STEELS, *WELD. J.,* VOL 56 (NO. 7), 1977, P 201-S TO 210-S
45. W.F. SAVAGE, E.F. NIPPES, AND J.E. CASTERAS, EFFECT OF ALLOYING ADDITIONS ON THE WELDABILITY OF 70 CU-30 NI, *WELD. J.,* VOL 57 (NO. 12), 1978, P 375-S TO 382-S
46. W.P. HUGHES AND T.F. BERRY, A STUDY OF THE STRAIN-AGE CRACKING CHARACTERISTICS IN WELDED RENE 41--PHASE 1, *WELD. J.,* VOL 46 (NO. 8), 1967, P 361-S TO 370-S
47. A.W. DIX AND W.F. SAVAGE, FACTORS INFLUENCING STRAIN-AGE CRACKING IN INCONEL X-750, *WELD. J.,* VOL 50 (NO. 6), 1971, P 247-S TO 252-S
48. J.J. PEPE AND W.F. SAVAGE, EFFECTS OF CONSTITUTIONAL LIQUATION IN 18-NI MARAGING STEEL WELDMENTS, *WELD. J.,* VOL 46 (NO. 9), 1967, P 411-S TO 422-S

49. W.F. SAVAGE, E.F. NIPPES, AND M.C. MUSHALA, COPPER-CONTAMINATION CRACKING IN THE WELD HEAT-AFFECTED ZONE, *WELD. J.,* VOL 57 (NO. 5), 1978, P 237-S TO 245-S
50. W.F. SAVAGE, E.F. NIPPES, AND M.C. MUSHALA, LIQUID-METAL EMBRITTLEMENT OF THE HEAT-AFFECTED ZONE BY COPPER CONTAMINATION, *WELD. J.,* VOL 57 (NO. 8), 1978, P 237-S TO 245-S
51. J.C. LIPPOLD AND W.F. SAVAGE, SOLIDIFICATION OF AUSTENITIC STAINLESS STEEL WELDMENTS: PART 1--A PROPOSED MECHANISM, *WELD. J.,* VOL 58 (NO. 12), 1979, P 362-S TO 374-S
52. J.C. LIPPOLD AND W.F. SAVAGE, SOLIDIFICATION OF AUSTENITIC STAINLESS STEEL WELDMENTS: PART 2--THE EFFECT OF ALLOY COMPOSITION ON FERRITE MORPHOLOGY, *WELD. J.,* VOL 59 (NO. 2), 1980, P 48-S TO 58-S
53. COMPUTERIZED THERMAL AND MECHANICAL TESTING MACHINE A SIGNIFICANT ASSET IN RESEARCH, DEVELOPMENT AND PRODUCTION, *INDUSTRIAL HEATING,* VOL 53 (NO. 12), DEC 1986
54. H. FERGUSON, METALWORKING PROCESS SIMULATIONS SAVE PRODUCTION HEADACHES, *MET. PROG.,* SEPT 1986, P 39-42
55. D.J. WIDGERY, THE DESIGN AND USE OF RESISTANCE HEATED WELD THERMAL SIMULATORS, *PROC. CONF. WELD THERMAL SIMULATORS FOR RESEARCH AND PROBLEM SOLVING,* THE WELDING INSTITUTE, CAMBRIDGE, ENGLAND, APRIL 1972
56. D.W. WALSH, M.J. CIESLAK, AND W.F. SAVAGE, TEMPERATURE MEASUREMENTS IN RESISTANCE HEATED SPECIMENS: LONGITUDINAL GRADIENTS, *WELD. J.,* VOL 65 (NO. 8), 1986, P 184-S TO 192-S
57. G.M. GOODWIN, *PROC. 1ST JAPAN-U.S. SYMPOSIUM ON ADVANCES IN WELDING METALLURGY* (SAN FRANCISCO, CA), JUNE 1990, P 27-39
58. J.C. BORLAND, *BR. WELD. J.,* VOL 7 (NO. 8), 1961, P 508-512
59. W. LIN, W.A. BAESLACK, AND J.C. LIPPOLD, HOT DUCTILITY TESTING OF HIGH-STRENGTH, LOW-EXPANSION SUPERALLOYS, *RECENT TRENDS IN WELDING AND TECHNOLOGY LNT. CONF. PROC.* (GATLINBURG, TN), 14 MAY 1989, ASM INTERNATIONAL, P 609-614
60. W. LIN, W.A. BAESLACK, AND J.C. LIPPOLD, A METHOD FOR QUANTIFYING HAZ LIQUATION CRACKING SUSCEPTIBILITY USING THE GLEEBLE HOT-DUCTILITY TEST, *PROC. INT. SYMP. PHYSICAL SIMULATION,* DELFT, 1992, P 93-99

Brazeability and Solderability of Engineering Materials

S.D. Brandi, Escola Politécnica da Universidade de São Paulo, Brazil; S. Liu, Colorado School of Mines; J.E. Indacochea and R. Xu, University of Illinois at Chicago

Introduction

ANALYSIS of the brazeability and solderability of engineering materials requires the following considerations:

- JOINING PROCESS CHARACTERISTICS (TYPE AND CHARACTERISTICS OF HEAT SOURCE)
- CHEMICAL COMPOSITION OF BASE METAL
- CLEANING PREPARATION AND AFTER-PROCESS CLEANING
- CHEMICAL COMPOSITION OF FILLER METAL
- JOINT PROTECTION AGAINST OXIDATION (FLUX, BATH COMPOSITION, PROTECTIVE ATMOSPHERES)
- JOINT TEMPERATURE AND TIME
- JOINT DESIGN (JOINT GEOMETRY, JOINT CLEARANCE, JOINING POSITION)

Each one of these topics influences wetting and spreading behavior, joint mechanical properties, corrosion resistance, and residual stress levels.

Brazeability and Solderability of Engineering Materials

S.D. Brandi, Escola Politécnica da Universidade de São Paulo, Brazil; S. Liu, Colorado School of Mines; J.E. Indacochea and R. Xu, University of Illinois at Chicago

Process Applications and Requirements

The use of high-performance materials in advanced engineering and manufacturing places demanding requirements on joining processes such as brazing and soldering. Appropriate design of the joints and proper selection of joining processes and filler metals can increase the application of brazed and soldered joints in sophisticated mechanical assemblies, such as aerospace equipment, chemical reactors, and high-value electronic assemblies.

Soldering in the electronics field has undergone many dramatic changes since the first radios were built. Between the 1920s and 1940s, all interconnections were made using the point-to-point wiring method, and the solder joints were made with soldering irons. With the end of World War II and a greater demand for consumer electronics, mass soldering techniques such as dip soldering were developed. By the late 1950s, wave soldering was developed in England and introduced to the United States soon after; this technique is still widely used on printed wiring boards.

Brazing has also evolved significantly. The introduction of torch brazing, furnace brazing, and vacuum brazing has expanded the application of brazing to different material systems such as aluminum, stainless steel, superalloys, titanium, and ceramics.

Recent developments in aluminum brazing in the automotive industry include the application of noncorrosive fluxes in conjunction with a dry nitrogen atmosphere in a continuous mesh belt furnace. Under these conditions, only a small amount of flux is required for joint protection, so flux removal after brazing is not necessary. Many other automotive parts are furnace brazed, such as pollution control tubing, water pump impellers, torque converters, and so on. While the torque converter is usually brazed with a copper filler metal, a nickel filler metal is used when additional strength is required (for example, for dynamometer testing or other high-stress applications). Stainless steel automotive assemblies are brazed in a low-dew-point atmosphere such as cryogenic nitrogen, with an addition of 20 to 50 vol% of dry hydrogen.

The potential applications of brazing and soldering have increased both in volume and significance with the introduction of advanced structural ceramics. Much emphasis has been placed on the application of these ceramics in aerospace, automotive, and electronics applications. A complete characterization and understanding of the existing brazing and soldering techniques for joining dissimilar materials will contribute to the further development of these techniques and to the joining of ceramic materials reliably. Improved brazing and soldering technology will also result in the establishment of optimized design criteria and high-yield, low-cost manufacturing processes.

Aerospace Applications

Jet engine components such as turbines blades and stators with very thin section are assembled by brazing. The vanes can be hollow, with passageways for dynamic cooling during operation. Yet the brazements must be designed to withstand loads that the part will experience throughout its service life. In the modem aircraft engine industry, most brazing is carried out in vacuum or hydrogen, using specialized brazing cycles. Typically, large components are brazed in vacuum furnaces at 0.13 to 0.013 Pa (10^{-3} to 10^{-4} torr). Other parts of engine systems are brazed in continuous hydrogen furnaces and batch hydrogen furnaces.

The interest of the aerospace industries in structural ceramics centers on the potential application in gas turbine engines. The severe service conditions encountered within a gas turbine, and the desire of design engineers to increase operating temperatures for higher efficiency, make several of the properties of structural ceramics particularly attractive. High-temperature strength, low thermal expansion coefficient, dimensional accuracy, and corrosion resistance are some of these properties.

The silicon-base ceramics, such as Si_3N_4, SiC, or Sialon ($Si_{5.5}Al_{0.5}O_{0.5}N_{7.5}$), have been considered for gas turbine parts (including bearings, shroud rings, and turbine blades). The bearings may be found either as ceramic balls or rollers in conventional bearings or shells. The use of ceramics for shroud rings allows for high-temperature operation while maintaining tight control of shroud clearances on the blades, because of their low thermal expansion coefficients. In the case of present turbine blades technology, the operating temperatures are already close to the maximum temperature limit. Therefore, the high-temperature strength of these ceramics is particularly desirable.

Ceramic materials are more difficult to braze because they are not easily wet by the standard commercial filler metals. When joined to metals, the different thermal expansion coefficients may result in residual stress generation and cracking of the ceramic component. Ceramic-to-metal joints must be specially designed to alleviate the residual stress problem. Techniques such as metallization and ductile reactive metal fillers can be utilized to promote wettability and minimize stress-induced cracking in the ceramic component.

Braze joints for heat engine applications require good mechanical properties at high temperatures. Thus, the combination of base metal and filler metal in the joint must be thermally stable. The presence of intermetallics at the interface of base metal and filler metal should be minimized, and the filler metal should not undergo any type of strengthening or softening mechanism, such as precipitation hardening, order/disorder reactions, or precipitate coarsening. Wetting and spreading are important to avoid lack of filling, which will create stress concentrations and, consequently, the premature failure of the component.

Nuclear Applications

Because of the need of structural integrity at high temperatures for components used in nuclear applications, refractory metals are targeted for those applications. Joints of significant dimensions with properties that are adequate to meet requirements of strength, ductility, and corrosion resistance are required. Brazing has been widely accepted as the joining method to satisfy these requirements.

For the past 30 years engineers have been attracted to the inherent properties of ceramic materials for potential use in nuclear reactors, which involve high temperatures and corrosive environments. Fuel element structures were one of the first applications for which experimental assemblies were designed and fabricated using ceramic materials. To demonstrate the use of ceramic brazing in advanced fuel elements, one element was made entirely of Al_2O_3 and compartmented ceramic plates were brazed after loading with fuel (Fig. 1).

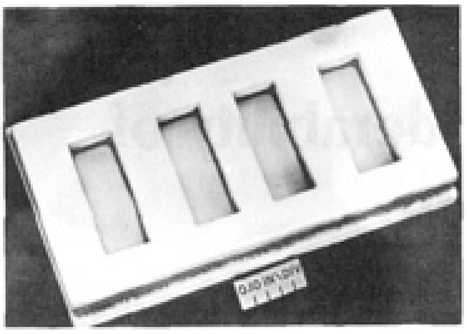

FIG. 1 ALUMINUM OXIDE COMPARTMENTED ASSEMBLY VACUUM BRAZED WITH 49TI-49CU-2BE (IN WT%) BRAZING FILLER METAL AT 980 °C (1795 °F) FOR 10 MIN

A BeO tubular assembly was also brazed, simulating a fuel bundle with ferrule spacers. The larger BeO tube diameters represent the fuel tubes, while the smaller tubes simulate the spacers. This assembly (Fig. 2) was brazed with the same brazing filler metal as the flat-plate assembly in Fig. 1 (vacuum, 1000 °C, or 1830 °F, for l0 min).

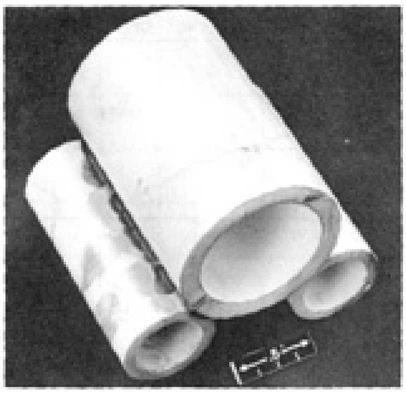

FIG. 2 BERYLLIUM OXIDE TUBULAR ASSEMBLY BRAZED WITH 49TI-49CU-2BE FILLER METAL

Another application is the advanced Tokamak fusion test reactor, being developed at Princeton University. The effort is aimed at achieving a commercially feasible magnetic-fusion reactor. In this reactor, large ceramic insulator rings are used to electrically insulate the vacuum-pumping and natural-beam injector system from the stainless steel vacuum vessel, in which the plasma is confined by surrounding magnetic fields.

The insulating rings for the fusion reactor are composed of ceramics, copper, and stainless steel, all of which must be brazed together successfully. To accomplish the task, titanium hydride was first applied to the ceramic mating surfaces, followed by silver-copper foil of 0.076 mm (0.003 in.) thick. The assembled parts, wrapped in titanium foil that acted as a getter to obtain a higher purity atmosphere, were then successfully brazed.

Heat Exchangers

Heat exchangers, particularly in the automotive industry. are fabricated from many different materials. such as copper. aluminum, copper-nickel alloys, steels, and titanium. Depending on the base metal selected, joint strength requirements, and assembly design, different brazing processes, such as torch, furnace, resistance brazing, and so on, can be used.

Temperature and humidity control within the atmospheres of space vehicles, space stations, and space suits can be realized with condensing heat exchangers. By use of the plate-fin design, units with high efficiency at minimum mass and volume may be tailored close to actual system requirements. Stainless steel cores, brazed together in vacuum with nickel-base filler metals, are used to meet these requirements.

A variety of ceramic materials have been considered for use in heat exchange systems, such as the main heat transfer surfaces (tubes, plates, and so on), seals, tube sheets, and insulation. They have been used primarily in these critical zones in an effort to minimize cost over metallic counterparts. Heat exchanger designs have ranged from all-ceramic to ceramic-metal hybrids with varying numbers of ceramic components. These parts are commonly brazed with ductile active metal fillers.

Brazeability/solderability in this application is related mainly to the corrosion resistance of the joint. The filler metal should be more noble than the base metal to minimize galvanic corrosion in the joint. Wetting and spreading are also important to prevent crevices that may generate crevice corrosion, especially when the environment has chloride ions in solution. The flux should be easily removed to avoid residues that can accelerate the corrosion rate in the joint.

Electronic Packaging

Filler metal compositions and forms for soldering have experienced significant changes to meet the needs of a growing number of process techniques and applications, from simple wires to foils, preforms, and pastes. The first electronic solder was 50Sn-50Pb wire, traditionally used by plumbers. Today, the selection of materials has expanded to eutectic tin-lead solders, neareutectic (60Sn-40Pb) solders, and ternary (62.5Sn-36.1Pb-l.4Ag) alloys in the form of foils, preforms, and pastes. Although the main emphasis has remained on the Sn-Pb eutectic and near-eutectic alloys, other alloys, such as the indium and gold-tin eutectics, have also been developed for more specialized applications.

Flux technology has also evolved. Environmental regulations have placed emphasis on the development of new flux types. Extensive work has been done with organic acids, that can be removed with aqueous media or low-residue fluxes which have eliminated cleaning altogether.

The use of mass soldering techniques has increased the need for adequate solderability of substrates. The large throughput of high-production-rate processes prohibits the repair of defective solder joints. In addition, modern components are more heat-sensitive. Lower soldering temperatures and less mechanical working have made it necessary to maintain the highest solderability to ensure a metallurgically sound and reliable joint on the first pass.

The service performance of a solder joint is a strong consideration in modem electronics. Many environments, such as aircraft avionics, telecommunications systems, dishwashers, and automobiles, impose a regular thermal cycle on the electronic equipment that results in thermal-mechanical strains. Failure by fatigue or creep becomes inevitable if the joint design is incorrect or if the joint has defects due to poor manufacturing control.

Solderability in this application describes the formation of a joint by adequate wetting of the entire surface by the molten solder to achieve good electrical contact. Base-metal dissolution by the molten solder should be minimized by processing variables (time, temperature, filler metal composition). The flux should promote the oxide removal and protect the molten solder against oxidation to prevent "cold weld," which may cause a misfunction of the electronic component.

Dissimilar Material Joints

Industrial machinery construction and electronic components fabrication require the joining of materials that exhibit highly different properties. An example is the joining of ceramics with metals that have quite different melting points and coefficients of thermal expansion. Various processes have been developed to join dissimilar materials. Usually the selection of the joining technique will depend on the materials used and on the specified joint properties. For instance, to obtain high mechanical strength or good heat resistance, the reaction layer must consist of high-melting-temperature products. For good dielectric properties, the reaction layer must contain oxide compounds. The use of vitreous phases in the interlayer may provide chemical resistance and high puncture-proofness. Presently, processes such as sintering of interfaces, spraying and sintering of coatings, and depositing of layers by chemical vapor deposition, soldering (metal-metal and metal-ceramic), and infiltration (joints of composite materials) are applied in joining dissimilar materials. In developing processes to join dissimilar materials, one must consider and solve such problems as wettability, chemical reactivity of the base materials and filler metals to be bonded under specific processing conditions, and the effect of the joining processes on the stress state in the joint during production and during operation.

Brazeability and Solderability of Engineering Materials

S.D. Brandi, Escola Politécnica da Universidade de São Paulo, Brazil; S. Liu, Colorado School of Mines; J.E. Indacochea and R. Xu, University of Illinois at Chicago

Materials Selection

Brazing and soldering are joining processes performed at temperatures below the base material solidus temperature. To fill the joint, the liquid filler metal must spread through the joint gap by capillarity or coat the substrate by spreading. Capillarity and spreading are related to the balance of surface tensions, which determines the wetting and spreading behavior of a liquid on a solid. Both wetting and spreading are results of interfacial reactions (physical and chemical reactions). The influence of material selection, joint design, and surface conditions on the mechanical properties and corrosion resistance of a brazed joint will be discussed below.

There is no precise definition of brazeability and solderability of a material. These properties are related to how easily a material can be joined, how closely a joint can fit its design function (mechanics, corrosion, electrical contact, and so on), and the most suitable joining process. As an example, in electronic applications solderability involves not only the soldering process characteristics but also the effect of heating on the electronic component characteristics.

Wetting and Spreading. Wetting is a surface phenomenon related to the balance of the solid-liquid surface tension, γ_{sl}, the solid-vapor surface tension, γ_{sv}, and the liquid-vapor surface tension, γ_{lv}. The contact angle, θ, between the solid and the liquid is defined in equilibrium conditions as the angle between the vectors that represents the liquid-vapor surface tension and the solid-liquid surface tension. The relationship between the contact angle and the surface tensions is given by the Young-Dupré equation:

$$\gamma_{LV} \cdot \mathrm{COS}\ \theta = \gamma_{SV} - \gamma_{SL} \qquad \textbf{(EQ 1)}$$

A liquid is said to wet a solid when the contact angle is less than 90°. In this case, for constant γ_{lv} the solid-vapor surface tension is greater than the solid-liquid surface tension. On the other hand, if the solid-vapor surface tension is lower than the solid-liquid surface tension, the contact angle is greater than 90° and the liquid will not wet the solid. Figure 3 shows a wetting and a nonwetting system in a joint. In neither case did a reaction layer form between the solid and the liquid.

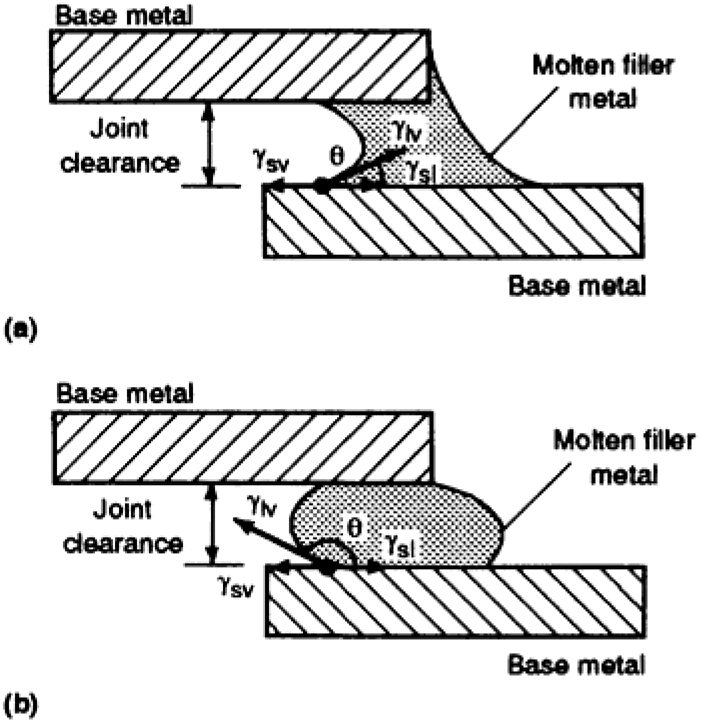
Base metal
Molten filler metal
Joint clearance
γ_{sv}
θ
γ_{lv}
γ_{sl}
Base metal
(a)
Base metal
Molten filler metal
Joint clearance
γ_{lv}
θ
γ_{sl}
γ_{sv}
Base metal
(b)

FIG. 3 SCHEMATIC SHOWING RELATIONSHIP OF CONTACT ANGLE TO SURFACE TENSION. (A) WETTING SYSTEM. (B) NONWETTING SYSTEM

Spreading is a dynamic process related to the contact angle θ and can be described as an increase in the liquid surface and the interfacial area, ΔA, with time. Based on Fig. 4, a liquid is said to completely spread over a solid surface when the contact angle tends to 0° (Fig. 4a). Alternatively, spreading is considered to proceed when the liquid surface area increases to a maximum, even if θis greater than 0° (Fig. 4b). Rate of spreading can be inhibited by the liquid inertia and/or the liquid viscosity.

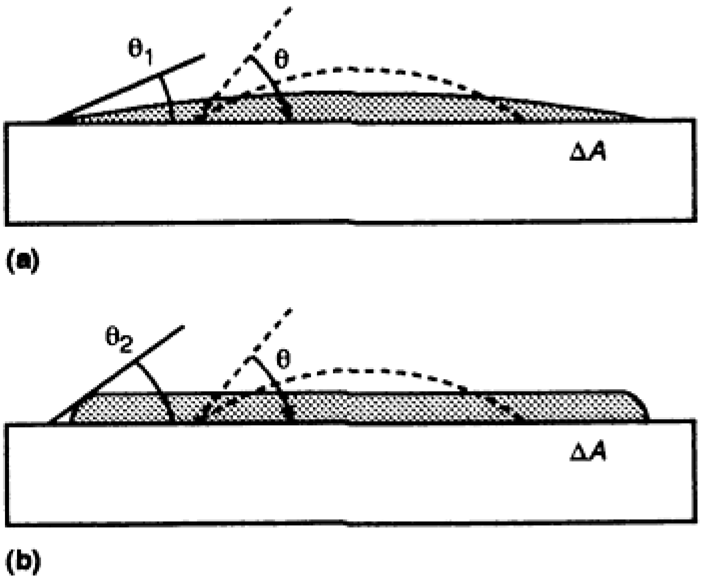

FIG. 4 SCHEMATIC SHOWING A LIQUID SPREADING THROUGH A SOLID SURFACE WITH $\theta < 90°$. (A) θ_1 0°. (B) $\theta_2 = \theta_{EQUIL}$

From Eq 1, $\gamma_{lv} \cdot \cos \theta$can be considered as a resistance force that opposes a liquid to wet a solid, and $(\gamma_{sv} - \gamma_{sl})$ is a driving force for wetting. If $\gamma_{lv} \cdot \cos \theta$ is greater than $(\gamma_{sv} - \gamma_{sl})$, no wetting or spreading will occur. So it is important to decrease $\gamma_{lv} \cdot \cos \theta$and/or to increase $(\gamma_{sv} - \gamma_{sl})$ by means of a flux, a filler metal, or a suitable brazing atmosphere. γ_{lv} and γ_{sv} decrease when a deoxidation reaction takes place at the liquid and solid surfaces. γ_{sl} can be reduced by promoting a deoxidation reaction via alloying elements in the filler metal (such as phosphorus in BCuP filler in contact with a copper substrate), formation of a reaction layer, or adsorption of a chemical element at the solid-liquid interface. Figure 5 shows the effect of the atmosphere and filler metal composition on the surface tensions for a Cu-BCuP system.

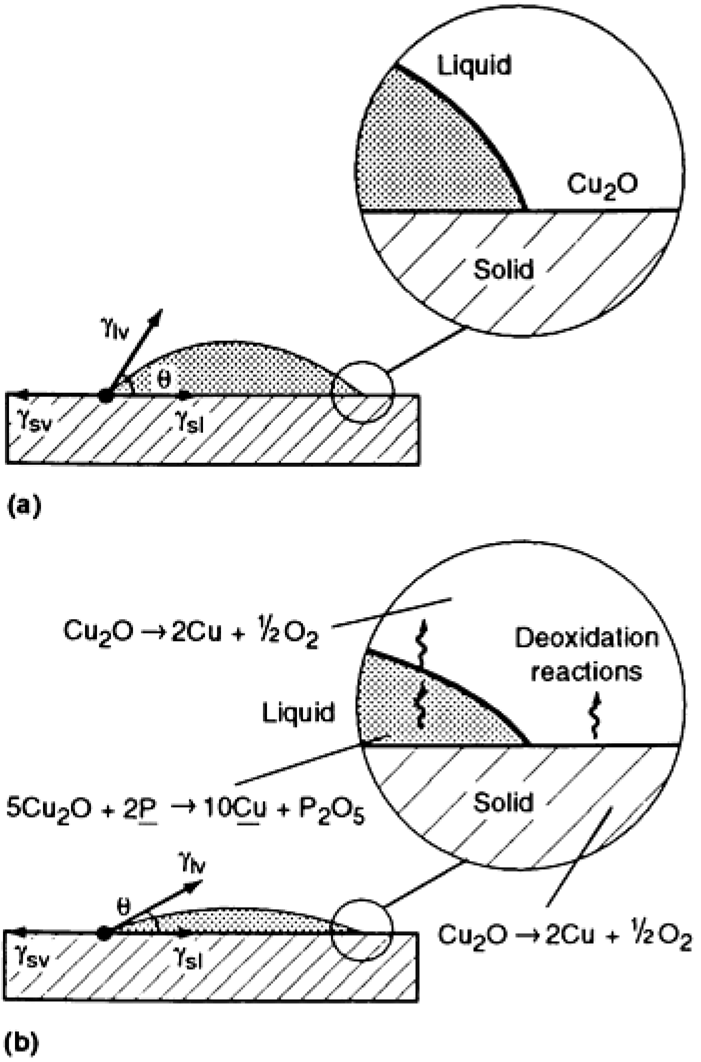
Liquid
Cu2O
Solid
γlv
θ
γsv
γsl
(a)
Cu2O → 2Cu + ½O2
Deoxidation reactions
Liquid
5Cu2O + 2P → 10Cu + P2O5
Solid
γlv
θ
γsv
γsl
Cu2O → 2Cu + ½O2
(b)

FIG. 5 SCHEMATIC SHOWING EFFECT OF THE SURROUNDING MEDIUM AND THE COMPOSITION OF THE FILLER METAL ON THE WETTING AND SPREADING BEHAVIOR OF COPPER-BCUP SYSTEM. (A) NEUTRAL OR OXIDIZING MEDIUM. (B) DEOXIDIZING MEDIUM

Interfacial Reactions. During brazing, the molten filler metal often reacts with the base metal. The product is a reaction layer that is sometimes too thin to be observed by optical microscopy. In fact, this layer modifies the wetting and spreading behavior of the liquid filler metal in contact with the base metal. Figure 6 shows the interfacial reaction during the brazing process.

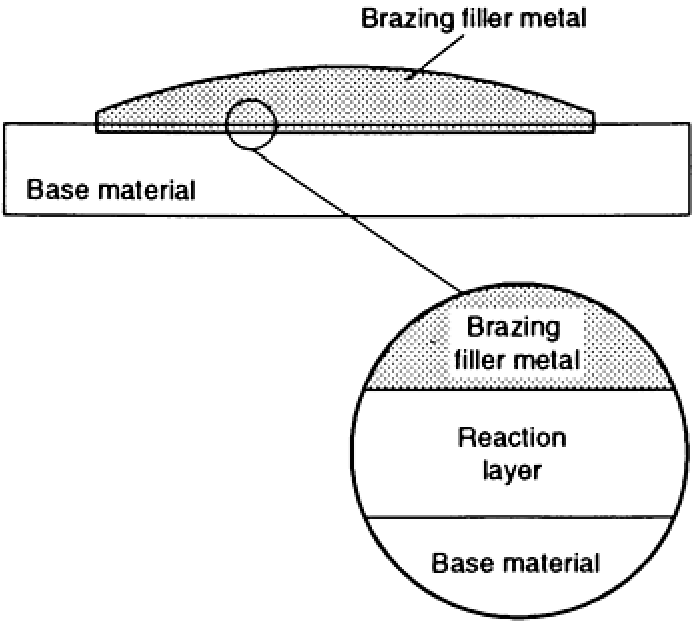
Brazing filler metal
Base material
Brazing filler metal
Reaction layer
Base material

FIG. 6 REACTION LAYER FORMATION DURING THE BRAZING PROCESS

When an interfacial reaction occurs between the base metal and the liquid filler, the free energy of the reaction per unit area and unit time (ΔG_r) should be included in Eq 1, as follows:

$$\gamma_{SV} - (\gamma'_{SL} - \Delta G_R) = \gamma_{LV} \cdot \mathrm{COS}\ \theta \qquad \textbf{(EQ 2)}$$

Due to a change in the solid-liquid interface by a chemical reaction, γ'_{sl} is lower than γ_{sl} of Eq 1. The variables γ'_{sl} and ΔG_r will increase the driving force for wetting. If the driving force is greater than $\gamma_{lv} \cdot \cos\theta$, spreading will occur until the liquid has reacted completely with the solid. This concept can be illustrated by the results of a sessile drop experiment in copper-silver systems, shown in Fig. 7.

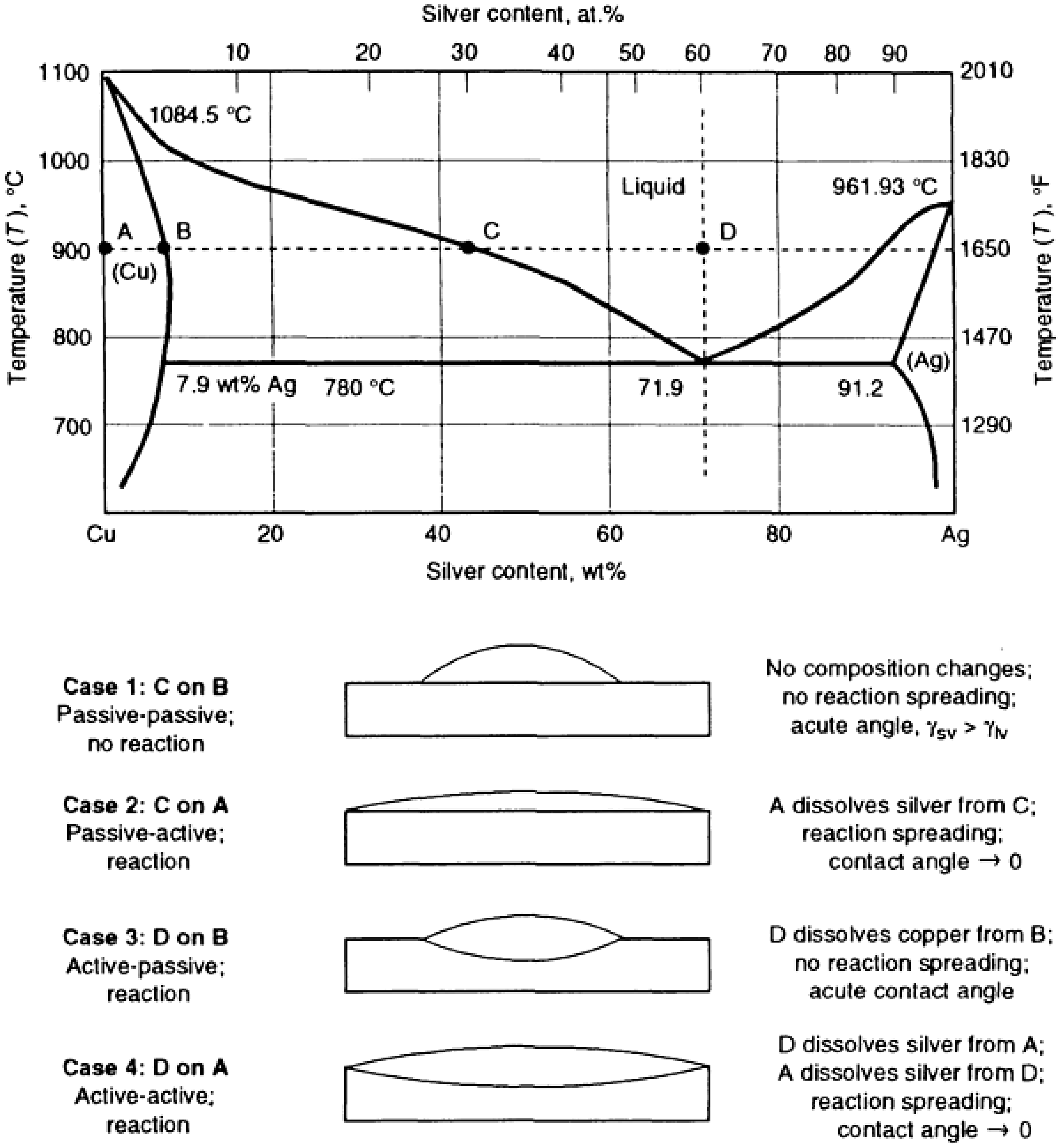
Silver content, at.%
10
20
30
40
50
60
70
80
90
1100
1000
900
800
700
2010
1830
1650
1470
1290
Temperature (T), °C
Temperature (T), °F
1084.5 °C
Liquid
961.93 °C
A
B
C
D
(Cu)
(Ag)
7.9 wt% Ag
780 °C
71.9
91.2
Cu
20
40
60
80
Ag
Silver content, wt%
Case 1: C on B
Passive-passive;
no reaction
No composition changes;
no reaction spreading;
acute angle, $\gamma_{sv} > \gamma_{lv}$
Case 2: C on A
Passive-active;
reaction
A dissolves silver from C;
reaction spreading;
contact angle → 0
Case 3: D on B
Active-passive;
reaction
D dissolves copper from B;
no reaction spreading;
acute contact angle
Case 4: D on A
Active-active;
reaction
D dissolves silver from A;
A dissolves silver from D;
reaction spreading;
contact angle → 0

FIG. 7 RESULTS OF A SESSILE DROP TEST TO ANALYZE SPREADABILITY AS A FUNCTION OF THE SOLID AND LIQUID COMPOSITION. SOURCE: REF 1

Cases 1 and 2 in Fig. 7 show that no spreading occurred. In case 1 the solid and the liquid are in equilibrium (points B and C), and no composition change can occur as long as the brazing temperature is maintained. In case 3 the liquid dissolves copper from the solid to reach the equilibrium composition (point C). In both cases the chemical composition of the solid remains the same. This means that in cases 1 and 3 the solid has a passive behavior with respect to the liquid. On the other hand, cases 2 and 4 show the solid as the active element in the system. In other words, the chemical composition of solid A changes up to B, based on the phase diagram. In case 2, a spreading reaction occurs (that is, silver from the liquid diffuses into the solid and changes the composition at the solid-liquid interface). In case 4, both solid and liquid are active and a spreading reaction is expected to occur. It is important to point out that wetting will occur in all four cases, but only in cases 2 and 4 is spreading expected to take place.

Depending on the mechanical properties of the base metal and the filler metal, the mechanical behavior of the reaction layer may determine the strength of the braze joint. If a solid solution is formed, the reaction layer may be as tough as the base metal and the filler metal. On the other hand, if a brittle intermetallic compound is produced, the interfacial region may degrade the joint properties. Phase diagrams are good tools to estimate the compatibility between the filler metal and the base metal.

Liquid Filler Flowability. Joint clearance is an important factor in the brazeability or solderability of a joint. Table 1 shows the recommended joint clearance at brazing temperatures for different filler metal groups. The BAlSi and BMg filler metals require larger clearances, while the BCu, BAu, BAg, and BCuP filler metals require smaller gaps. This difference can be related to the difficulty with which the first two filler metal groups wet and spread over a solid surface. The flowability of a liquid filler metal on a solid surface is related to the equilibrium conditions expressed through Eq 2, the dynamic properties such as liquid viscosity, and interfacial reaction kinetics, if applicable. So it is desirable that a filler metal has:

- HIGH WETTABILITY AS DESCRIBED BY EQ 2
- LOW VISCOSITY
- FASTER INTERFACE REACTION KINETICS

Figure 8 shows the effect of the chemical composition of a binary alloy on its viscosity. The eutectic composition will have the lowest viscosity. Filler metals with large solidification ranges tend to have higher viscosities. Notice that the chemical composition of the filler metal may change during brazing or soldering due to a dissolution reaction or an intermetallic formation. This will also affect the liquid flowability.

TABLE 1 RECOMMENDED JOINT CLEARANCE FOR SELECTED FILLER METALS AT BRAZING TEMPERATURE PER AWS B2.2 SPECIFICATION

AWS FILLER METAL CLASSIFICATION	FLUX OR BRAZING ATMOSPHERE	JOINT CLEARANCE(A)	
		mm	**in.**
BAG, BVAG	FLUX	0.05-0.12	0.002-0.005
	BRAZING ATMOSPHERE	0.00-0.05	0.000-0.002
BAU, BVAU	FLUX	0.05-0.12	0.002-0.005
	BRAZING ATMOSPHERE	0.00-0.05	0.000-0.002
BALSI	FLUX	0.05-0.21(B)	0.002-0.008(B)
	BRAZING ATMOSPHERE	0.21-0.25(C)	0.008-0.010(C)
BCUP	FLUX OR FLUXLESS	0.025-0.12	0.001-0.005
BCU, BVCU	BRAZING ATMOSPHERE	0.00-0.05	0.000--0.002
BCUZN, RBCUZN	FLUX	0.05-0.12	0.002-0.005
BNI1 THROUGH BNI5, BNI8, BCO1	FLUX OR BRAZING ATMOSPHERE	0.05-0.12	0.002-0.005
BNI6, BNI7	BRAZING ATMOSPHERE	0.00-.05	0.000-0.002
BMG	FLUX	0.10-0.25	0.004-0.010

(A) CLEARANCE ON THE RADIUS WHEN RINGS, PLUGS, OR TUBULAR MEMBERS ARE INVOLVED. ON SOME APPLICATIONS IT MAY BE NECESSARY TO USE THE RECOMMENDED CLEARANCE ON THE DIAMETER TO PREVENT EXCESSIVE CLEARANCE WHEN ALL THE CLEARANCE IS ON ONE SIDE. AN EXCESSIVE CLEARANCE WILL PRODUCE VOIDS. THIS IS PARTICULARLY TRUE WHEN BRAZING IS ACCOMPLISHED IN A HIGH-QUALITY ATMOSPHERE (GAS PHASE FLUXING).

(B) FOR A LAP LENGTH OF $\le$6.35 MM ($\le \frac{1}{4}$ IN.).

(C) FOR A LAP LENGTH OF >6.35 MM ($>\frac{1}{4}$IN.)

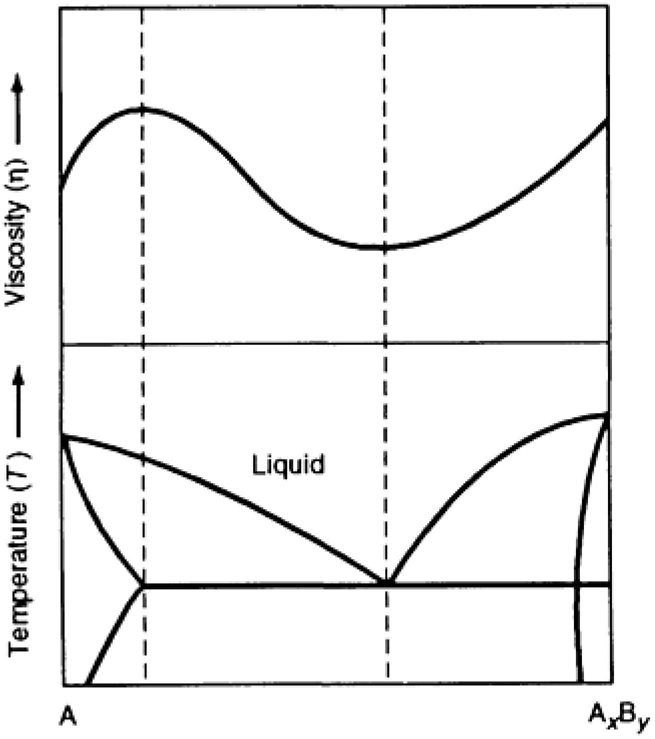

FIG. 8 EFFECT OF THE CHEMICAL COMPOSITION OF A BINARY ALLOY ON ALLOY VISCOSITY. SOURCE: REF 2

Joint design also affects filler metal flow. Figure 9 shows two different groove geometries for brazing with preplaced filler metals. In the case of a groove with sharp edges (Fig. 9a), the filler metal melts and starts to wet and spread through the groove surface. At the corner, however, the abrupt change in wall angle produces a barrier to the liquid's continued flow. Nearly all the liquid metal will accumulate at the corner, so that only a smaller amount of the molten filler is available to form the braze joint. On the other hand, the groove in Fig. 9(b) is designed to assist liquid filler flow to the outer joint surface by specifying a radius at the corner. Smooth changes in angles are preferred, to promote a more uniform flow of the liquid filler metal.

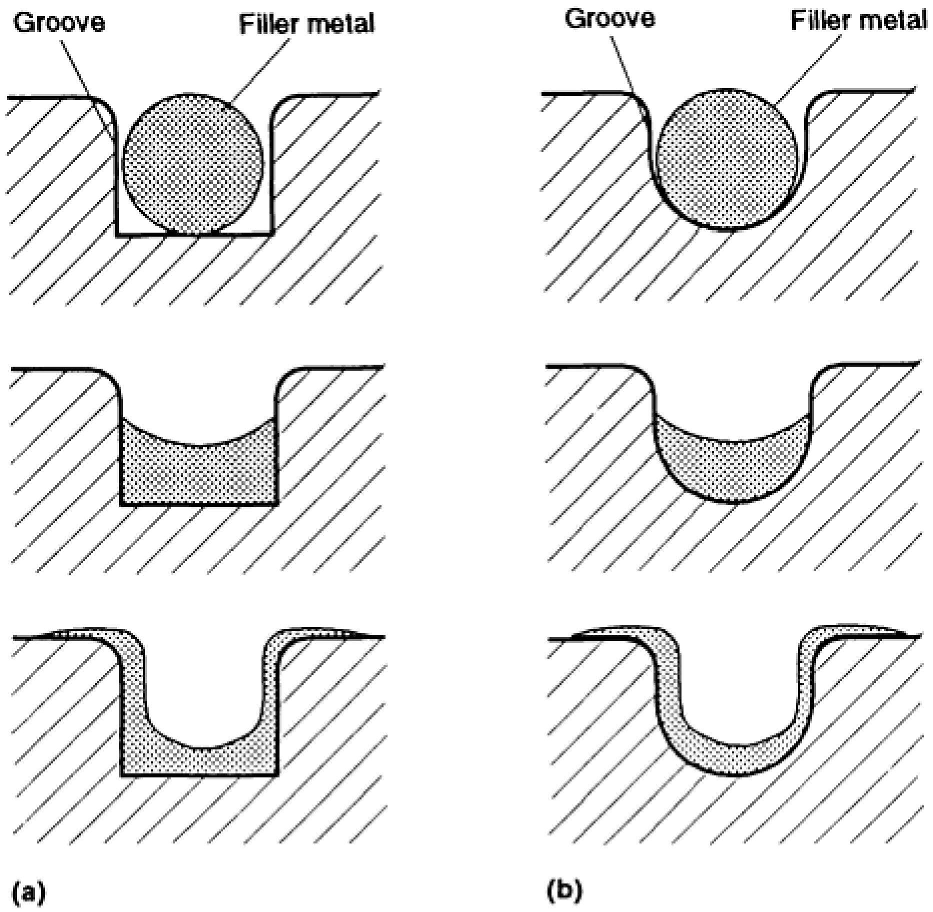

FIG. 9 EFFECT OF GROOVE GEOMETRY ON THE SPREADABILITY OF A PREPLACED FILLER METAL. (A) GROOVE WITH SHARP EDGES AND RECTANGULAR BOTTOM. (B) GROOVE WITH RADIUSED CORNERS AND SEMICIRCULAR BOTTOM WALL

Brazing position will also affect the flow rate of the liquid filler metal. A horizontal joint position is expected to offer a better flow characteristic than a vertical joint design. This is due to the gravitational field action in the vertical joint. The flow equation for the vertical and horizontal gaps are, respectively:

$$\frac{dh}{dt}=\frac{1}{4}\sqrt{\frac{g_{lv}.d.\cos q}{h.t}} \qquad \textbf{(EQ 3)}$$

$$\frac{dh}{dt}=\frac{1}{2}\sqrt{\frac{g_{lv}.d.\cos q}{h.t}} \qquad \textbf{(EQ 4)}$$

where *h* is the filler metal penetration in the joint, *d* is the joint clearance, η is the molten filler metal viscosity, *t* is brazing time, γ_{1v} is the surface tension between liquid and vapor, and θ is the wetting angle.

Joint Clearance. The effect of joint clearance on the behavior of a joint is about the same whether the joint is brazed or soldered. When shear strength is plotted as a function of joint clearance (Fig. 10), three different regions can be identified. At small clearances, region 1, the joint strength decreases as the clearance diminishes. The number of pores and the lack of joint filling caused by improper flow of the molten filler metal may be responsible for the lower shear strength observed. Surface discontinuity and/or entrapped flux are other factors that contribute to the joint performance. If intermetallic compounds form at the interface, a small joint clearance will invariably result in a higher volume fraction of brittle compounds, which degrade the joint properties.

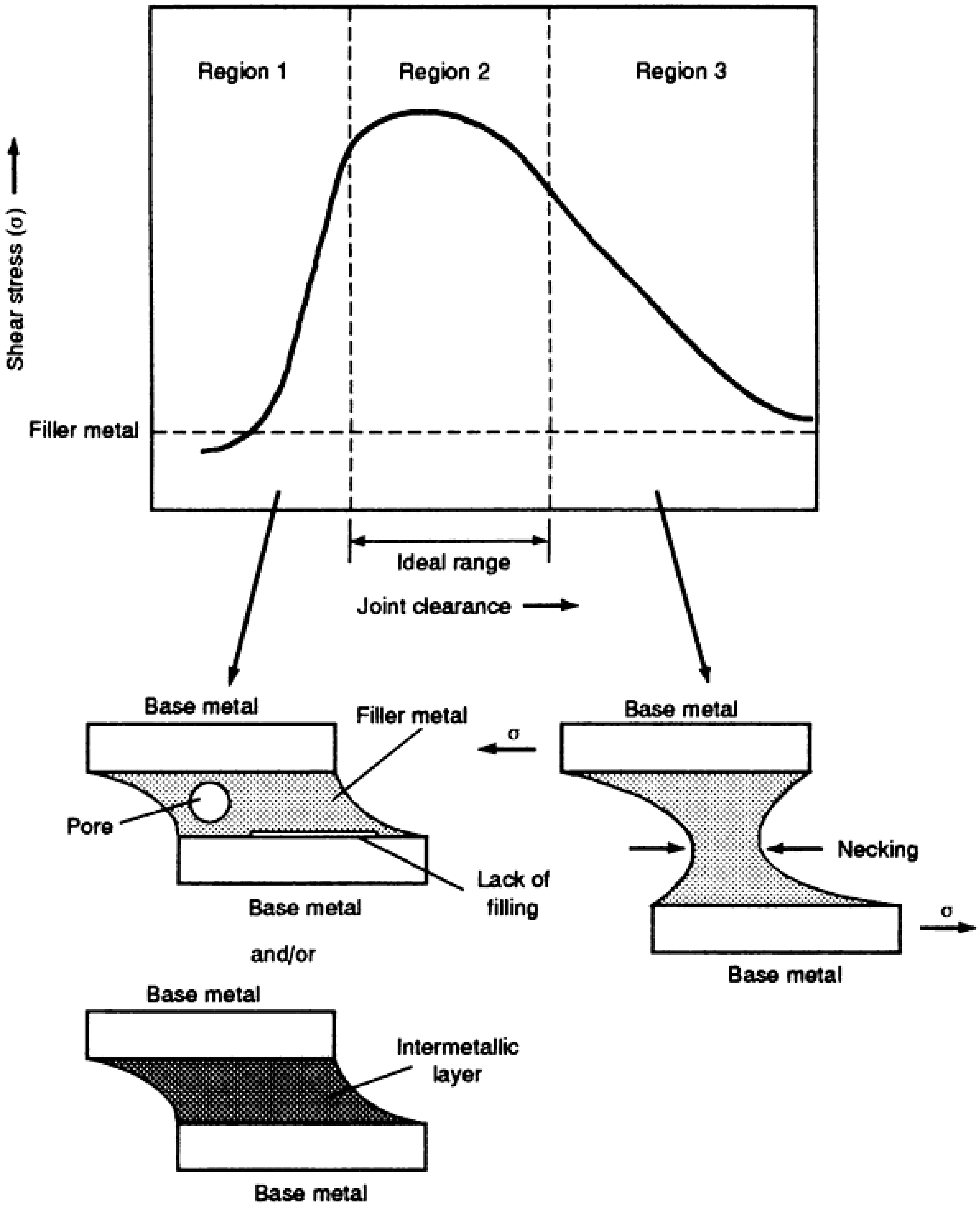
Region 1
Region 2
Region 3
Shear stress (σ)
Filler metal
Ideal range
Joint clearance
Base metal
Filler metal
σ
Base metal
Pore
Necking
Lack of filling
Base metal
σ
and/or
Base metal
Base metal
Intermetallic layer
Base metal

FIG. 10 EFFECT OF JOINT CLEARANCE ON THE STRENGTH OF A BRAZEMENT

With clearances indicated in region 2, the joint strength increases to a value close to that of the base metal, but only in the case of brazing. This increase in strength is due to a constraint generated by the base metal in the filler metal. As the clearance further increases, region 3, the constraint in the filler metal decreases and the mechanical properties of the filler metal will determine the strength of the joint. The lack of joint filling and the presence of voids can also affect the properties in this region.

Mutual Dissolution and Erosion. In the presence of the liquid filler metal, base metal dissolution can occur. Depending upon the mutual solubility of the base metal and the filler metal, excessive dissolution or erosion can result. Additionally, the alloying elements found in the filler metal may migrate into the base metal through solid-state diffusion or liquid metal grain boundary penetration. Figure 11(a) shows a phase diagram that suggests complete solubility of the filler metal (fm) and the base metal (bm). The amount of base metal dissolution expected in the system represented by Fig. 11(a) will be higher than that expected in the eutectic phase diagram shown in Fig. 11(b). The effect of dissolution depends on the kinetics of the reaction. One example of mutual solubility is the brazing of nickel-base alloys with pure copper. The copper-nickel phase diagram is similar to Fig. 9(a).

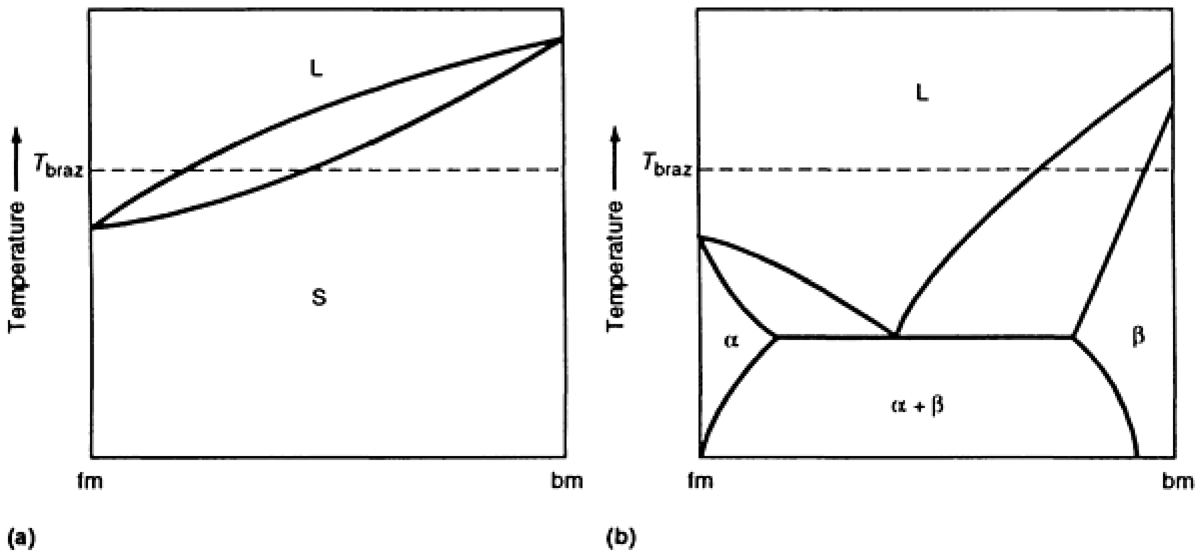

FIG. 11 HYPOTHETICAL PHASE DIAGRAMS BETWEEN FILLER METAL AND BASE METAL. (A) COMPLETE SOLUBILITY OF FILLER METAL AND BASE METAL. (B) EUTECTIC PHASE

Liquid penetration through grain boundaries may increase with decreasing mutual solubility. Copper is known to cause hot shortness in steel because of its limited solubility in steel (0.3 wt% at 590 °C, or 1095 °F) which favors the grain boundary liquid penetration.

Vapor Pressure. The vapor pressure of an element increases with increasing temperature. Furthermore, decreasing the surrounding pressure, such as in a vacuum, further increases the vaporization loss. This will cause variations in the chemical composition of the filler metal. As a consequence, the flowing behavior of the filler metal can be impaired. Metal vaporization may also promote porosity formation in the joint. Figure 12 summarizes the relationship between joining temperature, furnace pressure, and metal vapor pressure. Using silver as an illustration, the equilibrium vapor pressure at 1000 °C (1830 °F) is 13 Pa (10^{-1} mm Hg). If the furnace pressure is lower than 13 Pa (10^{-1} mm Hg), silver will vaporize. The loss due to vaporization is critical for soldering, especially when joined under vacuum.

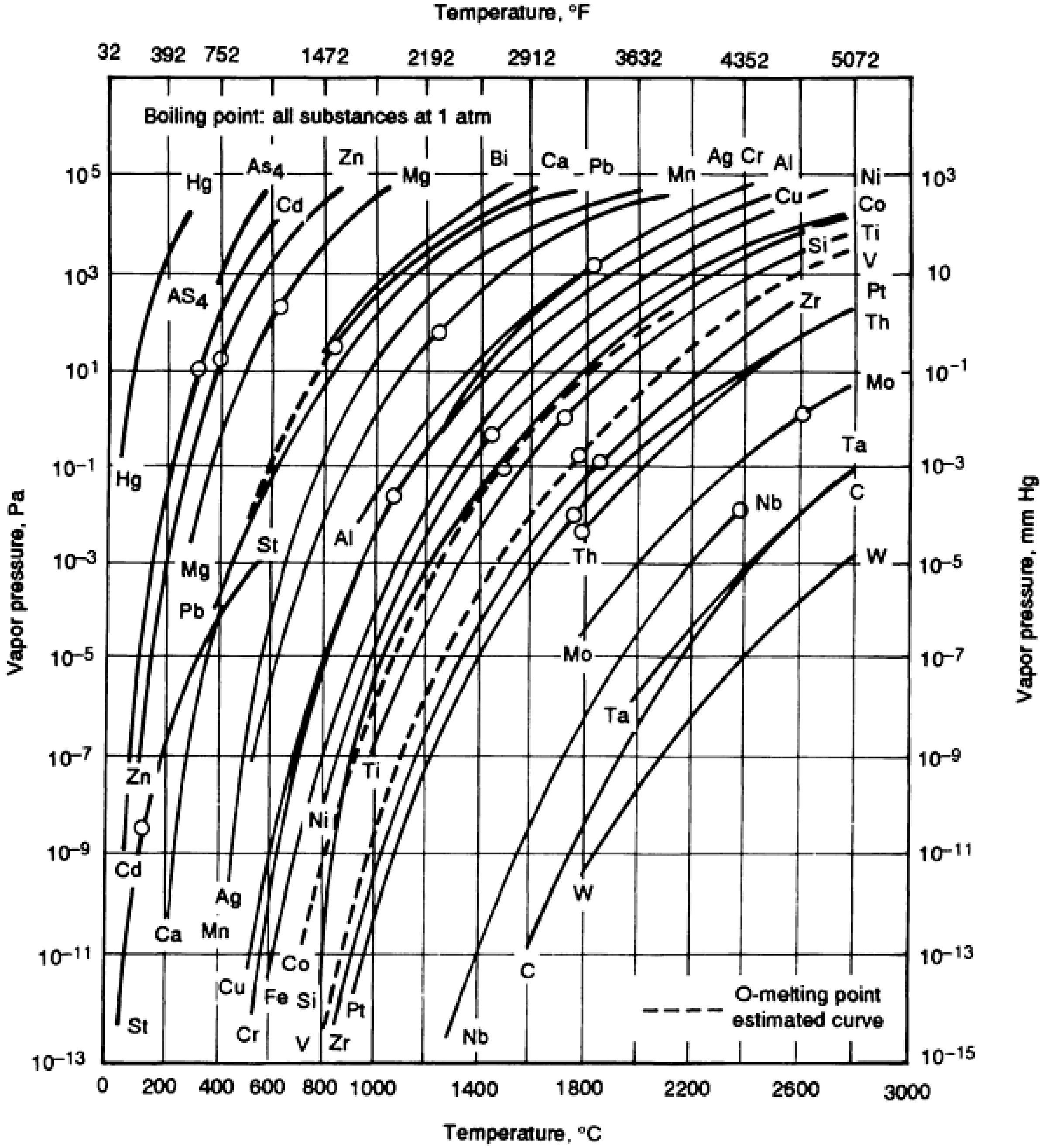

FIG. 12 VAPOR PRESSURE AS A FUNCTION OF TEMPERATURE. SOURCE: REF 3

Vapor pressure of metallic elements is particularly important in vacuum tube applications. The presence of high vapor pressure elements in the filler metal is restricted because of potential contamination inside the tube. The American Welding Society (AWS) filler metal specification for vacuum brazing limits the amount of high vapor pressure elements in the alloys. For elements that have vapor pressures higher than 13 μPa (10^{-7} torr) at 500 °C (930 °F), such as magnesium, antimony, potassium, sodium, lithium, thallium, sulfur, cesium, rubidium, selenium, tellurium, strontium, and calcium, the maximum allowed concentrations are 0.001 and 0.002 wt% for grades 1 and 2, respectively. In addition, the total sum of high vapor pressure elements, including zinc, cadmium, and lead, is limited to 0.010 wt%.

Fluxes and Atmospheres Selection. The major functions of fluxes in brazing and soldering are to remove oxides from the base metal surface by a reduction and/or dissolution reaction, to protect the clean surface from reoxidation, and to modify the surface tension of the molten metal. These functions can be achieved through chemical reactions that are, in general, thermally activated.

Chemical reactions typically occur between the surface oxide and the active components of the flux. The surface oxide can undergo a reduction reaction that results in the formation of a soluble compound, as exemplified below:

$$(\mathrm{MO})_{\mathrm{SURF}} + (\mathrm{FLUX}) \rightarrow \mathrm{M}_{\mathrm{SURF}} + \mathrm{O} + \ldots \qquad \textbf{(EQ 5)}$$

$$\begin{aligned}&(\mathrm{MO})_{\mathrm{SURF}} + (\mathrm{FLUX})\\ &\rightarrow \mathrm{M}_{\mathrm{SURF}} + (\mathrm{MO} \cdot \mathrm{FLUX})_{\mathrm{VEHICLE\ SOLUB.}}\end{aligned} \qquad \textbf{(EQ 6)}$$

$$\rightarrow \mathrm{M}_{\mathrm{SURF}} + (\mathrm{MO} \cdot \mathrm{VEHICLE})_{\mathrm{FLUX\ SOLUB.}} \qquad \textbf{(EQ 6A)}$$

where MO is the metal oxide, M is the metal, and O is oxygen. The subscript *surf* indicates that the reaction is located at the surface. (MO · flux) and (MO · vehicle) represent metal oxides dissolved in the flux or vehicle. Equation 5 represents a reduction of a surface oxide by a flux, resulting in a metal at the surface plus the evolution of oxygen. Equations 6 and 6a(a) represent a reduction of a surface oxide by a flux succeeding in a metal at surface and the oxygen soluble in the vehicle flux (Eq 6) or in the flux (Eq 6aa).

The chemical activity of a flux is strongly dependent on temperature and on the properties of a flux, such as thermal stability, activation temperature, and deactivation temperature, that characterize the activation temperature range. Suitable joining processes require that the flux components be stable throughout the joining cycle. They should not evaporate or decompose. For example, in applications in which the boiling point of the flux vehicle is lower than the joining temperature, the flux is thermally stable and able to protect the clean surface from reoxidation. On the other hand, the flux vehicle may have a high boiling point and remain intact on the part surface. The liquid flux vehicle assists the flux solder to act as a physical barrier that ensures protection of the clean surface against reoxidation.

The activation temperature range is defined as the temperature range at which a flux reaches its maximum chemical activity. The deactivation temperature is the temperature at which the flux breaks down and is no long active. It is important that the joining temperature remains slightly higher than the minimum activation temperature for maximum activity. Figure 13 shows two temperature ranges. The first one corresponds to the maximum flux chemical activity and dissolution and reduction reactions. The second corresponds to the breakdown of the fluxes and a loss of flux chemical activity.

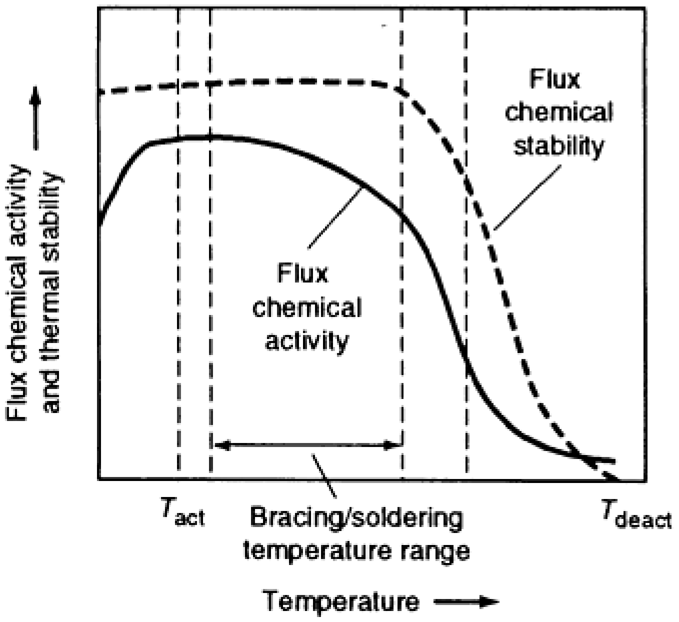

FIG. 13 EFFECT OF TEMPERATURE ON FLUX CHEMICAL ACTIVITY AND FLUX THERMAL STABILITY

Fluxes are not designed to remove grease, oil, dirt, or paint from the part surface. Thus, the joint surface must be free of these contaminants prior to joining.

Another common practice in brazing is the use of protective atmospheres. Combustion gases of different compositions, reducing gases such as NH_3 and hydrogen, and inert gases such as argon and nitrogen are typical examples of the gases used. In the case of combustion gases, the oxygen and carbon potentials of the gases determine the oxidation and reduction behavior of the alloying elements and surface oxides. Thermodynamic calculations can also be carried out to relate these potentials to the partial pressures of CO and CO_2. The water vapor content, similar to that of CO_2, is usually expressed as the dew point of the gas.

In the case of a hydrogen-rich atmosphere, it is the hydrogen-oxygen reaction alone that controls the oxidation-reduction behavior of the filler metal and base metal. For a particular brazing atmosphere and dew point, a threshold temperature must be reached to start to reduce the surface oxides that cover both the base metal and filler metal. This temperature is specific for each metal-metal oxide system and is related to the oxide stability.

Figure 14 shows the metal-metal oxide equilibria in pure hydrogen atmospheres. In this diagram the region under a metal-metal oxide curve represents the conditions under which the metal oxide can be reduced by hydrogen. For example, brazing a stainless steel at 925 °C (1700 °F) in a hydrogen-rich atmosphere can be successfully performed only if the dew point is lower than -43 °C (-45 °F). In this case the thermodynamic conditions are favorable to reduce the chromium oxide which will promote good wetting and spreading.

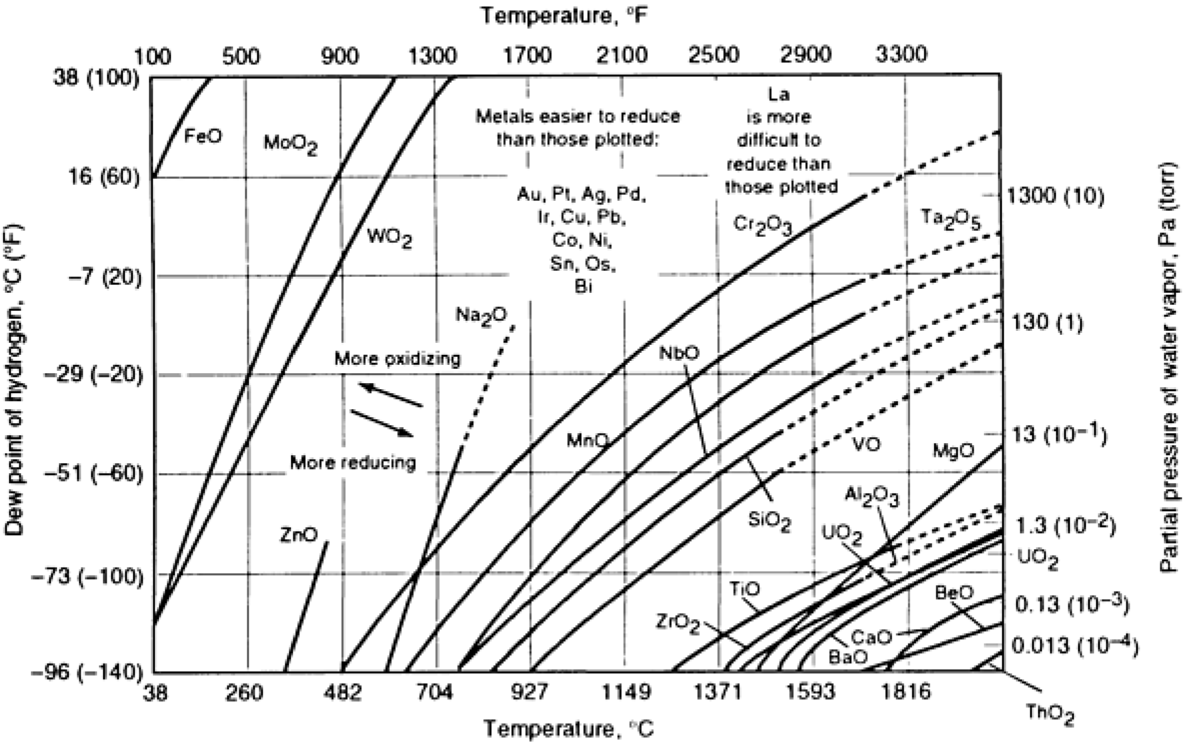

FIG. 14 METAL-METAL OXIDE EQUILIBRIA IN PURE HYDROGEN ATMOSPHERES. SOURCE: REF 4

Tables 2, 3, and 4 present the classifications of brazing fluxes, brazing atmospheres, and soldering fluxes. These tables are intended for general use only.

TABLE 2 BRAZING FLUX TYPES PER AWS SPECIFICATION B2.2

AWS BRAZING FLUX NO.	RECOMMENDED BASE METALS$^{(A)}$	RECOMMENDED FILLER METALS	RECOMMENDED USEFUL TEMPERATURE RANGE		CONSTITUENTS
			°C	°F	
1	ALL BRAZEABLE	BALSI	370-	700-	CHLORIDES,

	ALUMINUM ALLOYS		645	1190	FLUORIDES
2	ALL BRAZEABLE MAGNESIUM ALLOYS	BMG	480-650	900-1200	CHLORIDES, FLUORIDES
3A	(B)	BCUP, BAG	565-870	1050-1600	BORIC ACID, BORATES, FLUORIDES, FLUOBORATES, WETTING AGENT
3B	(B)	BCU, BCUP, BAG, BAU, RBCUZN, BNI	730-1150	1350-2100	BORIC ACID, BORATES, FLUORIDES, FLUOBORATES, WETTING AGENT
4	ALUMINUM BRONZE, ALUMINUM BRASS, AND IRON-BASE OR NICKEL-BASE ALLOYS CONTAINING ALUMINUM OR TITANIUM	BAG, BCUP(C)	565-870	1050-1600	CHLORIDES, FLUORIDES, BORATES, WETTING AGENT
5	(B)	(D)	760-1205	1400-2200	BORAX BORIC ACID BORATES, WETTING AGENT

(A) NOT APPLICABLE, IN SOME CASES, TO TITANIUM-BASE AND ZIRCONIUM-BASE METALS.

(B) ALL EXCEPT THOSE LISTED UNDER 1, 2, AND 4.

(C) COPPER-BASE ALLOYS ONLY.

(D) SAME AS 3B (EXCLUDING BAG-1 THROUGH BAG-7)

TABLE 3 BRAZING ATMOSPHERE TYPES (PER AWS SPECIFICATION B2.2)

AWS BRAZING ATMOSPHERE NO.	SOURCE	DEW POINT MAXIMUM(A)		PRESSURE		COMPOSITION, WT%				
		°C	°F	PA	TORR	H_2	N_2	CO	CO_2	OTHER
1	COMBUSTED FUEL GAS	RT	RT	. . .	. . .	1-5	87	1-5	11-12	. . .
2	COMBUSTED FUEL GAS	RT	RT	. . .	. . .	14-15	70-71	9-10	5-6	. . .
3	COMBUSTED FUEL GAS	-40	-40	. . .	. . .	15-16	73-75	10-11	. . .	. . .
4	COMBUSTED FUEL GAS	-40	-40	. . .	. . .	38-40	41-45	17-19	. . .	. . .
5	DISSOCIATED AMMONIA	-54	-65	. . .	. . .	75	25	. . .	. . .	. . .
6	HYDROGEN	RT	RT	. . .	. . .	97-100	. . .	. . .	. . .	. . .
7	HYDROGEN	-59	-74	. . .	. . .	100	. . .	. . .	. . .	. . .
8	HEATED VOLATILE MATERIALS	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	INORGANIC VAPOR (E.G., ZN, CD, LI, MG, FLUORIDES)

9	PURIFIED INERT GAS	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
10	VACUUM	. . .	. . .	>265	>2	. . .	. . .	. . .	. . .	. . .
10A	VACUUM	. . .	. . .	>65-265	>0.5-2	. . .	. . .	. . .	. . .	. . .
10B	VACUUM	. . .	. . .	>0.13-65	>0.001-0.5	. . .	. . .	. . .	. . .	. . .
10C	VACUUM	. . .	. . .	≤0.13	≤0.001	. . .	. . .	. . .	. . .	. . .

(A) RT, ROOM TEMPERATURE

TABLE 4 CLASSIFICATION OF SOLDERING FLUXES

FLUX TYPE	**USES**	**CONSTITUENTS**	**TEMPERATURE STABILITY**	**RECOMMENDED TEMPERATURE RANGE**	
				°C	**°F**
CORROSIVE (INORGANIC)					
ACIDS	STRUCTURAL	HYDROCHLORIC, HYDROFLUORIC, ORTHOPHOSPHORIC	GOOD	150-450	300-840
SALTS	STRUCTURAL	ZINC CHLORIDE, AMMONIUM CHLORIDE, TIN CHLORIDE	EXCELLENT	150-450	300-840
INTERMEDIATE (ORGANIC)					
ACIDS	STRUCTURAL, ELECTRICAL	LACTIC, OLEIC, STEARIC, GLUTAMIC, PHTHALIC	FAIRLY GOOD	180-280	355-500
HALOGENS	STRUCTURAL, ELECTRICAL	ANILINE HYDROCHLORIDE, GLUTAMIC HYDROCHLORIDE, BROMIDE DERIVATIVES OF PALMITIC ACID, HYDRAZINE HYDROCHLORIDE (OR HYDROBROMIDE)	FAIRLY GOOD	180-280	355-500
AMINES AND AMIDES	STRUCTURAL, ELECTRICAL	UREA, ETHYLENE DIAMINE, MONOTRIETHANOL AMINE, TRIETHANOLAMINE	FAIR	180-280	355-500
NONCORROSIVE (ORGANIC)					
ROSIN NONACTIVATED (R)	ELECTRICAL	WATER-WHITE ROSIN	POOR	140-270	285-520
ROSIN MILDLY ACTIVATED (RMA)	ELECTRICAL	ROSIN WITH ACTIVATOR (HALIDE-BASE, ORGANIC ACID, AMINES, AMIDES . . .)	POOR	140-270	285-520
ROSIN ACTIVATED (RA)	ELECTRICAL, STRUCTURAL	ROSIN WITH MORE ACTIVATOR THAN RMA FLUX	FAIR	140-270	285-520

Source: Ref 5

Additional Metallurgical Considerations. Exposure to brazing temperatures can also result in base metal embrittlement. These embrittling conditions must be taken into consideration in the design of a brazed joint and brazing process parameters. Soldering occurs at lower temperatures than brazing, so embrittlement seldom happens in the base metal of a soldered joint because it is thermally activated. Embrittlement in the base metal can be caused by second-phase precipitation, hydrogen embrittlement, or the brazing of dissimilar metals.

Second-Phase Precipitation. The most common embrittlement phenomenon is second-phase precipitation of carbides at grain boundaries. Plain steels with low carbon contents can be embrittled by the precipitation of a cementite (Fe_3C) film at the grain boundary. For a steel with 0.02 wt% C, this phenomenon occurs at approximately 400 °C (750 °F).

Ferritic and austenitic stainless steels can have their corrosion behavior and toughness impaired by chromium carbide precipitation between 450 to 850 °C (750 to 1560 °F). This precipitation forms a chromium-depleted zone around the carbides and along the grain boundaries, which decreases the corrosion resistance of the alloy. This phenomenon is called sensitization. For the austenitic stainless steels, a low-carbon steel or a stabilized steel should be chosen to avoid the loss in corrosion resistance. Higher chromium diffusivity in the ferritic stainless steels results in a narrower chromium-depleted zone than that observed in the austenitic stainless steels. Thus, a heat treatment can remove the chromium-depleted zone and restore corrosion resistance. However, the toughness of the ferritic stainless steels will not be recovered by heat treatment.

Aluminum-magnesium alloys can experience an Al_3Mg_2 grain boundary precipitate that decreases the corrosion resistance of the alloy. In this case the difference between the corrosion potentials of the precipitate and the matrix increases the corrosion rate at the matrix-precipitate interface. Depending on the conditions, intergranular corrosion may take place, followed by stress-corrosion cracking.

Hydrogen Embrittlement. Hydrogen can cause severe embrittlement in many different materials. It can diffuse faster at room temperature and may cause a delayed embrittlement, mainly in steels. This type of embrittlement is not important for soldering.

In steels, hydrogen embrittlement occurs when four conditions are met: favorable microstructure (martensite, bainite, ferrite), hydrogen level, residual stress, and low temperature (<<100 °C, or 212 °F). These four conditions should be kept in mind when choosing a brazing procedure for steels. For example, steel brazing in a hydrogen-rich atmosphere may be inappropriate due to hydrogen pickup during joining.

Two mechanisms have been proposed for hydrogen embrittlement in nonferrous alloys: water vapor buildup and hydride precipitation. If an alloy has oxides precipitated along the grain boundaries, hydrogen will diffuse toward the oxides and may react with them to form water vapor. The pressure built up because of water vapor at the grain boundaries can separate the boundaries and promote embrittlement of the alloy. This is the case with electrolytic tough pitch copper that is brazed in a hydrogen-rich atmosphere.

Refractory metals such as titanium, zirconium, niobium, tantalum, tungsten, and molybdenum are very susceptible to the presence of interstitial elements (hydrogen, nitrogen, oxygen, and carbon). These interstitials distort the metal lattice, causing a strong interstitial solid solution hardening that diminishes the toughness of these metals. In a body-centered cubic (bcc) lattice, the solubility of the interstitial elements is reduced, which may result in the precipitation of the carbides and nitrides in these metals. In the case of hydrogen, a hydride will precipitate. The brazing conditions should be carefully selected to avoid interstitial pickup and the formation of a second interstitial-based phase during brazing.

Brazing of Dissimilar Materials. When different materials are brazed together, the possibility of intermetallics formation and the effect of the mismatch between the coefficients of thermal expansion must be considered. The formation of intermetallics was described above. The mismatch between the coefficients of thermal expansion generates residual stresses that influence the mechanical properties of the joint. Ultimately, fracture may occur, depending on the residual stress level. This effect is particularly important in the case of metal-to-ceramic brazing, in which the thermal expansion coefficients are very different. Table 5 presents some physical properties for selected materials.

TABLE 5 GENERAL PROPERTIES OF SELECTED MATERIALS

MATERIAL	**MELTING POINT**		**DENSITY**		**EXPANSION COEFFICIENT**		**YOUNG'S MODULUS**		**TENSILE STRENGTH**	
	°C	**°F**	**g/cm^3**	**lb/in.^{3}**	**10^{-6}/k**	**10^{-6}/ °f**	**GPa**	**10^6 psi**	**MPa**	**ksi**
CERAMIC										
OXIDE										
AL$_2$O$_3$ (SAPPHIRE)	2050	3720	3.985	0.144	5.8	3.2	380	55.1	620	90
BEO	2530	4585	3.01	0.109	8.4-9.0	4.7-5.0	311	45.1	172-275	24.9-39.9
SIO$_2$ (VITREOUS)	1710	3110	2.19	0.079	0.54	0.30	69	10.0	110	16.0
ZRO$_2$ (Y$_2$O$_3$)	2960	5360	5.56	0.200	10	5.6	138	20.0	>300	>44
CARBIDE										
B$_4$C	2350	4260	2.51	0.091	5.6	3.1	440-470	64-68	310-350(B)	45-51(B)
SIC	2300-2580(A)	4170-4675(A)	3.17	0.115	4.3-4.6	2.4-2.6	414	60.0	450-520(C)	65-75(C)
TIC	3140	5685	4.93	0.178	7.7	4.3	462	67.0	275-450(C)	40-65(C)
WC	2770	5020	15.70	0.567	4.9	2.7	814	118	790-825	115-120
NITRIDE										
ALN	2400	4350	3.25	0.117	5.3	2.9	350	50.8	270	39(B)
SI$_3$N$_4$	1750-1900(A)	3180-3452(A)	3.19	0.115	3.3	1.8	304	44.1	400-580(B)	58-72.5
METAL										
AL	660	1220	2.70	0.098	23.5	13.1	69	10	50-195	7.3-28.3
CU	1083	1981	8.96	0.324	17.0	9.5	180	19	224-314	32.5-45.5
BRASS (70-30)	910-965	1670-1770	8.55	0.309	19-20	10.6-11.1	100-115	14.5-16.7	300-700	43.5-102
FE	1535	2795	7.87	0.284	12.1	6.7	197	28.5	180-210	26-30.5
AISI 304	1400-1455	2550-2650	7.93	0.286	18.0	10.0	190-210	27.6-30.4	460-1100	67-160
AISI 410	1480-1530	2695-2785	7.73	0.279	10-12	5.6-6.7	190-210	27.6-30.4	480-1500	70-220
NI	1453	2647	8.9	0.322	13.3	7.4	199	28.9	660	96
MO	2617	4743	10.22	0.369	5.1	2.8	324.8	47.1	458-690	66.4-100
NB	2468	4474	8.57	0.310	7.2	4.0	104.9	15.2	330-585	48-85
ZR	1852	3366	6.49	0.234	5.9	3.3	98	14.2	350-390	51-57
W	3410	6170	19.3	0.697	4.5	2.5	411	59.6	550-620	80-90
GRAPHITE										
GRAPHITE	3650	6600	2.26	0.082	0.6-4.3	0.3-2.4	6.9	10.0	28	41

(A) SUBLIMATION POINT.
(B) 5% POROSITY.
(C) 2% POROSITY

The strain developed during heating and cooling by mismatch of thermal expansion coefficients can be evaluated using Eq 7.

$$\Delta\varepsilon= \Delta\alpha\cdot \Delta T \qquad \textbf{(EQ 7)}$$

where $\Delta\varepsilon$ is the strain variation, $\Delta\alpha$ is the thermal expansion coefficient mismatch, and ΔT is the temperature variation below solidus temperature. However, more in-depth computations using finite-element analysis should be performed to identify stress concentrations as a function of the joint configuration.

In the case of a silicon nitride (Si_4N_3) part being joined to an austenitic stainless steel component at 1325 K (1925 °F) using a filler metal with a solidus temperature of 1298 K (1877 °F), the value of the strain is estimated at 16.7×10^{-3} (thermal expansion coefficient mismatch = 16.7×10^{-6} K). Correspondingly, a stress of 5.078 GPa (much higher than the strength of silicon nitride) is generated. Based on this estimation, it is very likely that cracks will nucleate and propagate in the brittle ceramic material. The actual value of the stress is, however, lower than the value calculated by Eq 7 once the metal deforms and the stress developed in the joint is alleviated.

One technique for decreasing the residual stress level in the joint is to perform brazing with a soft, ductile metal interlayer. In this case the interlayer will deform and will accommodate the residual stresses generated by the mismatch between the coefficients of thermal expansion. Figure 15 shows this technique.

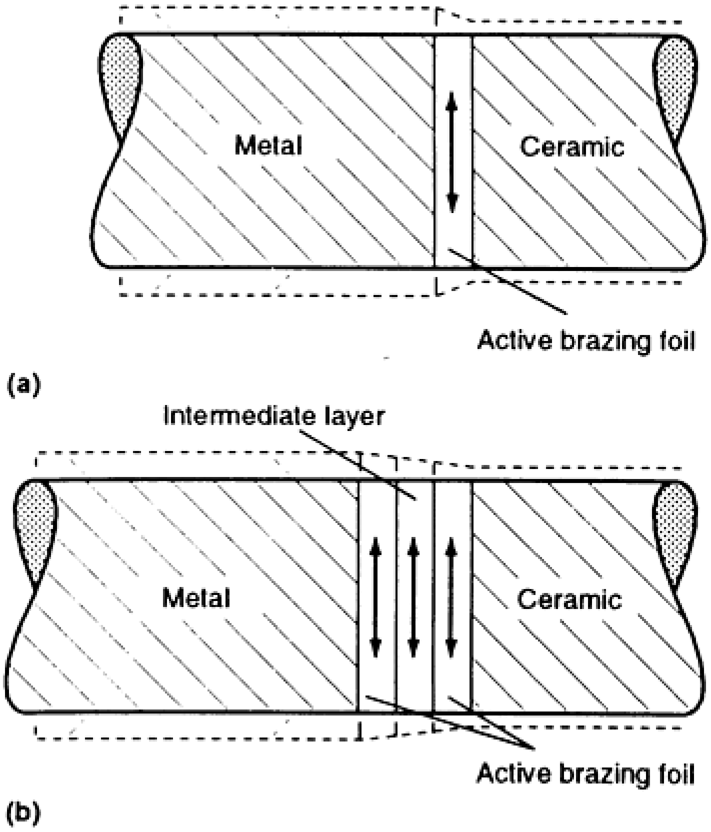

FIG. 15 ADDITION OF SOFT DUCTILE METAL INTERLAYER TO REDUCE RESIDUAL STRESSES IN A JOINT. (A) WITHOUT INTERLAYER. (B) WITH INTERLAYER

The mismatch in the strength may also generate a stress concentration in the joint under applied loads. In this case, assuming the same load and a metal-ceramic joint, the deformation in the ceramic is lower than that in the metal. Thus, as a result of the different deformation ($\varepsilon_{metal} - \varepsilon_{ceramic}$), a stress concentration is created at the interface, as shown in Fig. 16.

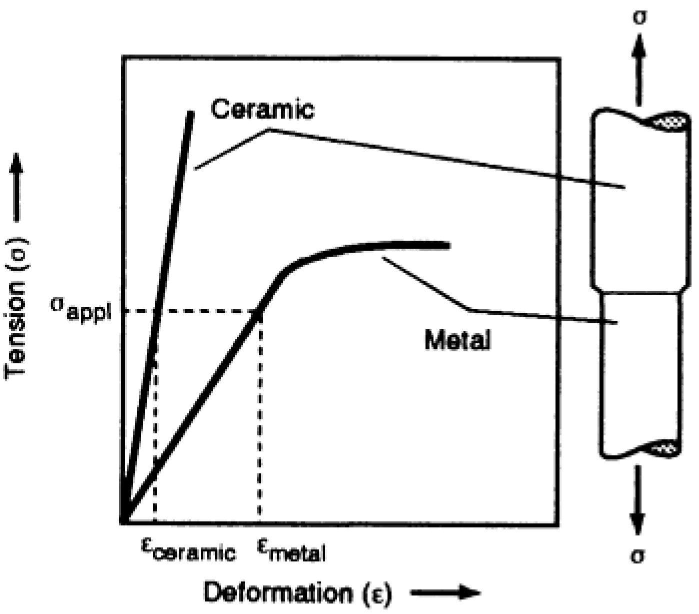

FIG. 16 STRESS CONCENTRATION AS A FUNCTION OF MISMATCHED STRENGTH

References cited in this section

1. P.R. SHARPS, A.P. TOMSIA, AND J.A. PASK, WETTING AND SPREADING IN THE CU-AG SYSTEM, *ACTA METALL.,* VOL 29 (NO. 5), 1981, P 855-865
2. V. KONDIC, *METALLURGICAL PRINCIPLES OF FOUNDING,* AMERICAN ELSEVIER, 1973
3. M.G. NICHOLAS AND D.A. MORTIMER, CERAMIC/METAL JOINING FOR STRUCTURAL APPLICATIONS, *MATER. SCI. TECHNOL.,* VOL 1 (NO. 9), 1985, P 657-665
4. D.R. MILNER AND R.L. APPS, *INTRODUCTION TO WELDING AND BRAZING,* PERGAMON PRESS, 1969
5. *SOLDERING MANUAL,* AMERICAN WELDING SOCIETY, 1977

Brazeability and Solderability of Engineering Materials

S.D. Brandi, Escola Politécnica da Universidade de São Paulo, Brazil; S. Liu, Colorado School of Mines; J.E. Indacochea and R. Xu, University of Illinois at Chicago

Brazing and Soldering Characteristics of Engineering Materials

To continue the discussion on brazeability and solderability of engineering materials, the joining process characteristics will be discussed. More detailed information about practices for brazing and soldering specific materials is available in the Section "Procedure Development and Practice Considerations for Brazing and Soldering" in this Volume.

All brazing and soldering processes consist of a cleaning operation, a filler metal addition, and a heating and cooling cycle. In some processes the material is heated faster than in others, which may require different temperature and time control. The faster the joining process, the lower the probability that intermetallics will be formed and the narrower the heat-affected zone (HAZ) in the base metal. On the other hand, fast heating may lead to high residual stress levels.

Depending on the joining process characteristics (heating and cooling cycles, joint design, protection of the joint, and so on), the mechanical and corrosion behavior could be different for brazed or soldered joints of the same base metal and filler metal. Figure 17 presents the thermal experience of a joint prepared using three different heating methods. Curve A represents induction or resistance heating, characterized by faster heating and cooling. Torch heating is generally more diffused and performed at a slower rate. As a result, base metal heating is more extensive, as shown by curve B. The thermal performance of furnace heating is described by curve C, where uniform heat in both base metal and joint is achieved.

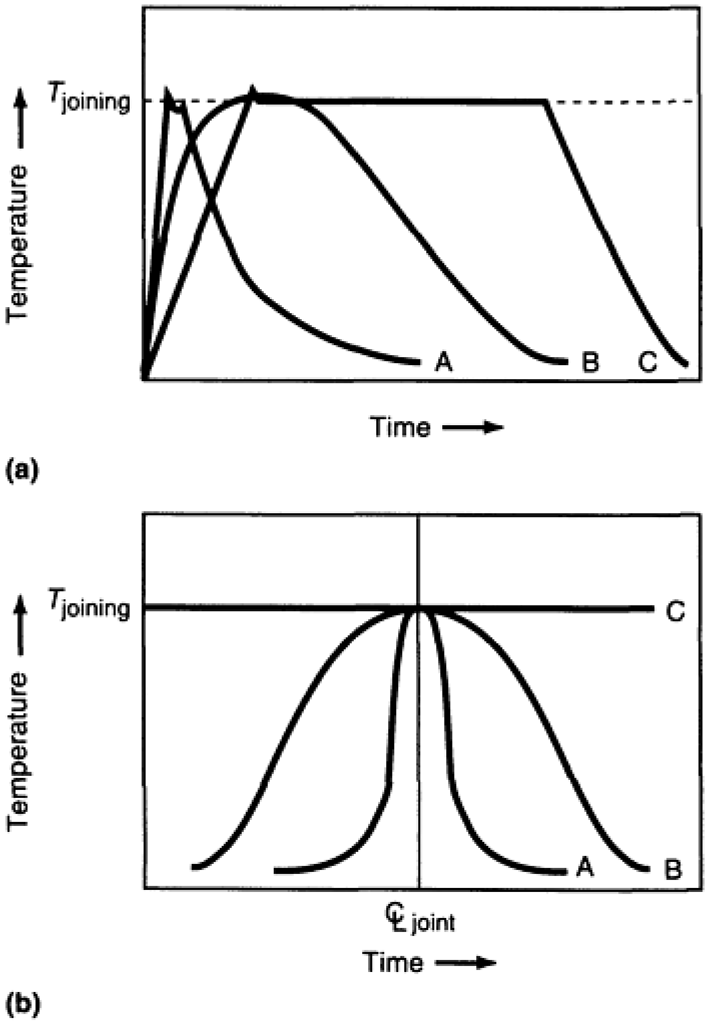

FIG. 17 THERMAL PERFORMANCE OF A JOINT AS A FUNCTION OF HEATING METHOD. (A) THERMAL CYCLE. (B) TEMPERATURE DISTRIBUTION. A, INDUCTION OR RESISTANCE HEATING; B, TORCH HEATING; C, FURNACE HEATING

The heating method is chosen according to base metal and filler metal characteristics, base metal-filler metal interaction, joint geometry, quality, cost, and productivity. Tables 6 and 7 summarize the common heating methods for brazing and soldering different materials.

TABLE 6 RELATIONSHIP OF TYPICAL HEATING METHOD, BASE METALS, AND FILLER METALS USED IN BRAZING

AWS FILLER-METAL CLASSIFICATION(A)	APPLICATION	HEATING METHOD(A)	AWS FLUX CLASSIFICATION(B)
BAG-1; BAG-1A; BAG-2; BAG-2A	GENERAL-PURPOSE (EXCEPT AL AND MG ALLOYS)	T, I, R, F(A), F(H)	FB3-A; FB3-C; FB3-D; FB3-E; FB3-F; FB3-G; FB3-H; FB3-I, FB3-J; FB3-K; FB4-A

BAG-3	GENERAL-PURPOSE (EXCEPT AL AND MG ALLOYS)	T, I, R	FB3-A; FB3-C; FB3-D; FB3-E; FB3-F; FB3-G; FB3-H; FB3-I, FB3-J; FB3-K; FB4-A
BAG-4; BAG-5; BAG-6; BAG-7; BAG-9; BAG-10; BAG-13; BAG-20; BAG-22; BAG-24; BAG-25; BAG-26; BAG-27; BAG-28	SPECIAL-PURPOSE	T, I, R, F(A), F(H)	FB3-A; FB3-C; FB3-D; FB3-E; FB3-F; FB3-G: FB3-H; FB3-I, FB3-J; FB3-K; FB4-A
BAG-8; BAG-8A; BAG-13A; BAG-19	SPECIAL-PURPOSE	F(V), F(H), F(A)	FB3-A; FB3-C; FB3-D; FB3-E; FB3-F; FB3-G; FB3-H; FB3-I, FB3-J; FB3-K; FB4-A
BAG-18; BAG-21; BAG-23	SPECIAL-PURPOSE	T, R, I, F(V), F(H), F(A)	FB3-A; FB3-C; FB3-D; FB3-E; FB3-F; FB3-G; FB3-H; FB3-I, FB3-J; FB3-K; FB4-A
BAU-1; BAU-2; BAU-3; BAU-4	SPECIAL-PURPOSE (STEEL, NI, CO = RESISTANCE TO CORROSION AND OXIDATION)	R, I, F(V), F(H), F(A)	FB3-D; FB3-I; FB3-J
BAU-5; BAU-6; BAU-7; BAU-8	SPECIAL-PURPOSE (STEEL, NI, CO = RESISTANCE TO CORROSION AND OXIDATION)	F(V), F(A)	FB3-D, FB3-I, FB3-J
BPD-I	SPECIAL-PURPOSE	F(V), F(A)	. . .
BALSI-2; BALSI-5	SPECIAL-PURPOSE	F(A), D	FB1-A; FB1-B; FB1-C
BALSI-3; BALSI-4	GENERAL-PURPOSE (AL ALLOYS)	T, D, I, R, F(V)	FB1-A; FB1-B; FB1-C
BALSI-6; BALSI-7; BALSI-8; BALSI-9; BALSI-10; BALSI-11	SPECIAL-PURPOSE	F(V)	FB1-A; FB1-B; FB1-C
BCUP-1	SPECIAL-PURPOSE (COPPER AND COPPER ALLOYS)	R, I	FB3-A; FB3-C; FB3-E; FB3-F; FB3-G; FB3-H; FB3-K; FB4-A
BCUP-2; BCUP-3; BCUP-4; BCUP-5; BCUP-6; BCUP-7	SPECIAL-PURPOSE (COPPER AND COPPER ALLOYS)	T, I, R, F(A), F(H)	FB3-A; FB3-C; FB3-E; FB3-F; FB3-G; FB3-H; FB3-K; FB4-A
BCU-1; BCU-1A; BCU-2	SPECIAL-PURPOSE (FERROUS AND NONFERROUS ALLOYS)	F(H), F(V)	FB3-D; FB3-I; FB3-J
RBCUZN-A, RBCUZN-C	SPECIAL-PURPOSE (FERROUS AND NONFERROUS ALLOYS)	T, F(H), F(A), I, R	FB3-D; FB3-I; FB3-J; FB3-K
RBCUZN-D	SPECIAL-PURPOSE (FERROUS AND NONFERROUS ALLOYS)	T, I, R	FB3-D; FB3-I; FB3-J; FB3-K
RBCUZN-E; RBCUZN-F; RBCUZN-G; RBCUZN-H	SPECIAL-PURPOSE (FERROUS AND	T, I, R	FB3-D; FB3-I; FB3-J; FB3-K

	NONFERROUS ALLOYS)		
BNI-1; BNI-1A; BNI-2; BNI-5; BNI-6; BNI-7	SPECIAL-PURPOSE (CORROSION AND HEAT RESISTANCE MATERIALS)	T, I, R, F(H)	FB3-D, FB3-I, FB3-J
BNI-3; BNI-4	SPECIAL-PURPOSE (CORROSION AND HEAT RESISTANCE MATERIALS)	T, I, R, F(H), F(V)	FB3-D; FB3-I; FB3-J
BNI-8	SPECIAL-PURPOSE (CORROSION AND HEAT RESISTANCE MATERIALS)	F(H), F(A), F(V)	FB3-D; FB3-I; FB3-J
BCO-1	SPECIAL-PURPOSE (HIGH-TEMPERATURE SERVICE)	F(V)	. . .
BMG-1	SPECIAL-PURPOSE (AZ1OA,K1A, AND M1A MG ALLOYS)	T, F(A), D	FB2-A

Source: Ref 3

(A) T, TORCH; I, INDUCTION; R, RESISTANCE; F(H), FURNACE WITH H_2 ATMOSPHERE; F(A), FURNACE WITH ARGON ATMOSPHERE; F(V), FURNACE WITH VACUUM; D, DIP BRAZING.

(B) IN SOME CASES, A FLUX IS NOT NECESSARY.

TABLE 7 RELATIONSHIP OF COMMON HEATING METHODS, BASE METALS, AND FILLER METALS USED IN SOLDERING

BASE METAL	**FILLER METAL**	**HEATING METHOD(A)**	**FLUX TYPE(B)**
COPPER AND COPPER ALLOYS	SN-PB	ALL	NC (CU, CU-SN, CU-ZN), IN (CU, CU-SN, CU-ZN), CO (ALL CU-BASE ALLOYS, ESPECIALLY CU-BE, SILICON BRONZE, AND ALUMINUM BRONZE)
STEEL	SN-PB (20-50%) (PREDOMINANTLY 40SN-60PB)	ALL	CO
AL COATED STEEL	40SN-60PB, 50SN-50PB, 60SN-40PB	I, R, US	CO
CD COATED STEEL	40SN-60PB, 50SN-50PB, 60SN-40PB	IR	NC
SN COATED STEEL	40SN-60PB, 50SN-50PB, 60SN-40PB	IR, I, D, T	NC
ZN COATED STEEL	40SN-60PB, 50SN-50PB, 60SN-40PB	DIFFICULT TO SOLDER	CO (SAME AS STAINLESS STEEL)
STAINLESS STEEL	SN-PB (SN > 50%)	ALL	CO
CAST IRON	SN-PB (20-50%)	ALL (AVOID OVERHEATING)	CO

NI AND NI ALLOYS	60SN-40PB, 50SN-50PB, 95SN-5SB	ALL	CO (CHLORIDE FLUX FOR NI, NI-CU), CO (HCI FLUX FOR NI-CR)
AL AND AL ALLOYS	SN-ZN, PB-BI, CD-ZN, ZN-AL	ALL (T, IR, AVOID OVERHEATING)	CO
MG AND MG ALLOYS	CD-ZN-SN, CD-ZN, SN-CD, SN-ZN	ALL	. . .
TIN AND TIN ALLOYS	63SN-37PB, 60SN-40PB	T, IR (AVOID OVERHEATING)	NC
PB AND PB ALLOYS	40SN-58PB-2SB	IR (WIPING)	NC
SILVER COATINGS AND FILMS	62SN-36PB-2AG	ALL	NC
GOLD COATINGS AND FILMS	53SN-29PB- 17IN-0.5ZN; 951N-5BI	ALL	NC
PLATINUM, PALLADIUM, RHODIUM COATINGS AND FILMS	60SN-40PB	ALL	NC
PRINTED CIRCUITS	63SN-37PB; 60SN-40PB	IR, D, W	NC

Source: Ref 5

(A) IR, SOLDERING IRONS; T, TORCH; D, DIP; I, INDUCTION; R, RESISTANCE; F, FURNACE; US, ULTRASONIC SOLDERING; FIR, FOCUSED INFRARED RADIATION; HG, HOT GAS; W, WAVE SOLDERING; VPC, VAPOR PHASE CONDENSATION; LB, LASER-BEAM SOLDERING.

(B) NC, NONCORROSIVE ORGANIC FLUX (BASE: ROSIN): IN. INTERMEDIATE ORGANIC FLUX (BASE: ORGANIC ACIDS, ORGANIC HYDROHALIDES, AMINES, AND AMIDES); CO, CORROSIVE INORGANIC FLUX (BASE: CHLORIDES AND ACIDS).

Low-Carbon Steels, Low-Alloy Steels, and Tool Steels

Brazing of low-carbon steels, low-alloy steels, and tool steels is a common practice. In this article, low-carbon steel is defined as a carbon steel with a maximum carbon content of 0.1 wt% and with silicon and manganese within the usual range. Low-alloy steels include not only the steels with a maximum total alloy content of 5.0 wt%, but also the high-strength low-alloy steels (microalloyed steels). Tool steels are high-alloy steels designed for high hardness and wear resistance. Tool steels can be further divided into two groups: carbon tool steels and high-speed tool steels. The mechanical properties of carbon tool steels depend mainly on the carbon content, which lies between 0.6 and 1.4 wt%. The high-speed tool steels are generally lower in carbon content, and their mechanical properties are mainly achieved by alloying elements such as tungsten, molybdenum, chromium, and vanadium. These elements are carbide formers. These carbides may affect the wetting and spreading behaviors of the braze alloy at the microstructural level. They also change the characteristics of iron oxide, making wetting and spreading more difficult.

Because each type of steel has a different set of mechanical properties, the brazing procedure for these steels should be determined as a function of the process characteristics and possible deterioration in material properties due to heating. In particular, the type of heating and heating cycle control (temperature and time), the kind of quenching medium (oil, water, air, and so on), and the optimal brazing conditions (filler metal type, flux or atmosphere type, cleaning conditions, joint geometry, fixtures, and so on) must be considered.

Low-carbon and low-alloy steels can be brazed by all brazing processes, but torch, furnace, induction, and dip brazing are most commonly used. Torch brazing of low-alloy steels should be performed with a neutral or slightly reducing flame, to minimize thermal effects on the base metal, filler metal, and flux. However, this process does not offer accurate temperature control, and excessive grain growth may occur. On the other hand, furnace brazing generally offers more accurate temperature and brazing cycle control. The chemical composition of the atmosphere, if not adequately controlled, may lead to decarburization or hydrogen pickup, which eventually will degrade the steel properties.

The presence of surface oxides on steels affects the wetting of the molten filler metal. In general, surface oxides on low-carbon and low-alloy steel can be easily removed by means of a flux or atmosphere during brazing. AWS specification B2.2 recommends different brazing conditions for steels that have more than 1.0 wt% Cr and for those that contain aluminum and/or titanium. These three elements change the characteristics of the iron oxide on the surface. They increase the adherence of the oxides and decrease the diffusivity of oxygen in the oxide layer. Thus, the flux or brazing atmosphere must be able to reduce these oxides and promote wetting.

Carbon and low-alloy steels can be brazed with BCu, BAg, and RBCuZn filler metals. These filler metal groups present brazing temperature ranges of 1095 to 1150 °C (2000 to 2100 °F), 620 to 900 °C (1145 to 1650 °F), and 910 to 982 °C (1670 to 1800 °F), respectively. Based on the brazing temperature range, two different brazing techniques can be performed (Fig. 18). If the brazing temperature is higher than the austenitization temperature, A_1, of the steel (Fig. 18a and b), brazing can be done together with the first part of the heat treatment. In other cases, the brazing process may require lower temperatures so it can be performed simultaneously with the first or secondary tempering of the steel (Fig. 18c).

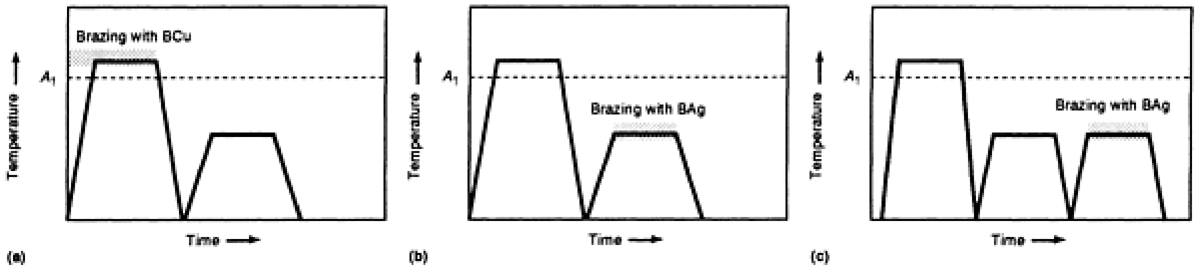

FIG. 18 INCORPORATION OF BRAZING OPERATION INTO HEAT TREATMENT CYCLE. (A) DURING AUSTENITIZATION PROCESS (BCU FILLER METAL). (B) DURING AUSTENITIZATION PROCESS (BAG FILLER METAL). (C) DURING TEMPERING PROCESS. A_1, AUSTENITIZATION TEMPERATURE

As a lower temperature process, soldering is related more to wetting phenomena than to microstructural changes in the steels. As with brazing, surface preparation and flux selection are fundamental to achieving a sound joint. Low-carbon and low-alloy steels present good solderability if the surfaces of the steels are properly cleaned and an aggressive flux is used.

Common filler metals used to solder steel belong to the tin-lead alloy group, with 20 to 50 wt% Sn. Because iron forms two intermetallics with tin, FeSn and Fe_2Sn, the amount and distribution of these intermetallics may impair the soundness of the joint. To minimize the presence of intermetallics, a solderable protective layer, such as copper and nickel, should be used. These layers present a limited intermetallic formation and can improve the solderability. The nickel layer must be protected from excessive oxidation by an overcoat of gold.

Tool steels having higher hardenability than the low-carbon, low-alloy steels, and the usual microstructure is tempered martensite with a fine dispersion of carbides. This microstructure is susceptible to thermal gradients, which may result in thermal cracking. During cooling, depending on the size of the component, the outer part transforms to martensite with volumetric expansion while the core remains untransformed and at a higher temperature. Together with different thermal expansion coefficients, this phase transformation may induce residual stresses and cracking in the transition region (between transformed and untransformed microstructure). Therefore, the heating and cooling of tool steels should be carefully controlled to minimize thermal gradients throughout the component.

Tool steels can be brazed by most brazing processes. However, torch brazing should be used carefully because of the lack of control in heating. As a result, thermal cracking is quite common. Induction brazing can be performed if a ductile metal insert is placed in the joint. The soft metal reduces the residual stresses generated and minimizes the cracking tendency. A composite filler metal (for example, copper-silver-copper sandwich) can be used to balance the shear stresses generated during brazing. As indicated previously, furnace brazing presents the best temperature control, but the composition of the furnace atmosphere must be regulated to avoid decarburization of the base metal.

The filler metal groups are similar to those used to braze low-carbon and low-alloy steels. Some tool steels may have a higher chromium content, so the fluxes must be specially designed to promote wetting in spite of the oxidation potential of chromium.

Soldering of tool steel is not commonly done due to a lack of relevant applications.

Stainless steels are designed for better corrosion and oxidation resistance than ordinary steels. Chromium is responsible for this improvement. However, as the amount of chromium increases, the surface oxide changes from an $FeO/Fe_3O_4/Fe_2O_3$ to a more stable Cr_2O_3, which is difficult to wet. The presence of titanium, aluminum, manganese, and silicon in the stainless steels further modifies the nature of the surface oxide. The flux or the joining atmosphere must be able to promote reduction or dissolution of the oxide layer and enhance wetting and spreading behavior.

Stainless steels are often divided into five groups: austenitic, ferritic, martensitic, precipitation-hardened, and duplex. Filler metal selection will depend on the particular application:

- SERVICE CONDITIONS SUCH AS TEMPERATURE, STRESSES, AND ENVIRONMENT
- HEAT TREATMENT (ESPECIALLY FOR MARTENSITIC AND PRECIPITATION-HARDENING STAINLESS STEELS)
- BRAZING PROCESS CHARACTERISTICS (INCLUDING THERMAL CYCLE CONTROL)
- WETTING AND SPREADING BEHAVIOR
- METAL DISSOLUTION AND EROSION OF THE BASE METAL
- SECOND-PHASE PRECIPITATION (CARBIDES, σ PHASE)
- COST

Austenitic stainless steels exhibit good corrosion resistance, except in chloride-content environments. According to AWS specification B2.2, the austenitic and austenitic-alloyed-with-titanium stainless steels are classified into two groups. All other stainless steel groups are classified as one.

Corrosion resistance of an austenitic stainless steel is impaired by sensitization, which is a precipitation of chromium carbides along the grain boundaries when slowly heating or cooling the steel through the temperature range of 450 to 850 °C (750 to 1560 °F). The amount of precipitation depends on the temperature, time, carbon content, grain size, degree of plastic deformation, and the amount of stabilizers such as titanium and niobium. If not controlled, the heating cycle during brazing may sensitize the base metal. Based on Fig. 17, the best thermal experience to avoid sensitization is induction or resistance brazing (curve A). Furnace brazing (curve C) can give good results if joining is carried out at a temperature higher than the sensitization range, followed by a fast cooling.

Often the base metal that is susceptible to sensitization can be substituted with a low-carbon or a niobium-stabilized stainless steel, without the need to qualify a new brazing procedure (according to AWS B2.2).

Galvanic corrosion may also occur because of the potential difference between the base metal and the filler metal. In the case that the filler metal is less noble than the base metal, it will corrode faster and result in a crevice, which will further accelerate the corrosion process. A lack of joint filling (promoted by poor wetting) or void formation may also create a crevice that will corrode with or without galvanic action.

The chromium-nickel stainless steels are prone to stress-corrosion cracking (SCC). During brazing, the filler metal may penetrate along the grain boundaries of the base metal and form a grain boundary network of the filler metal. This grain boundary penetration, associated with tensile stresses and/or cyclic stresses and an aggressive medium, may promote stress-corrosion cracking. An example of stress-corrosion cracking at room temperature occurs when a sensitized austenitic stainless steel is exposed to fluorides (residual from brazing fluxes). Thus, cleaning after brazing is as important as the brazing operation to increase the life of brazed components.

Austenitic stainless steels have considerably higher coefficients of thermal expansion and lower thermal conductivities than other steels. The combination of these physical properties generally results in large distortions in a brazement. In the case of localized heating, the broad heat source that heats up a large region may also induce plastic deformation in a large region, which generates residual stresses during cooling. For example, torch brazing of stainless steels should be used carefully when distortion is an important factor.

Stainless steels can be brazed using BCu, BAg, BNi, BCo, BPt, BPd, and BAu filler metal groups. Table 8 shows some filler metals and respective service temperatures of the brazement.

TABLE 8 BRAZING TEMPERATURES AND SERVICE TEMPERATURES FOR FILLER METALS USED IN THE BRAZING OF STAINLESS STEELS

FILLER METAL	BRAZING TEMPERATURE RANGE		SERVICE TEMPERATURE	
	°C	°F	°C	°F
BAG-1 THROUGH BAG-8A, BAG-24	620-900	1145-1650	205	400
BAG-13; BAG-19; BAG-21	800-980	1475-1800	370	700
BCU	1095-1150	2000-2100	425	800
CU-MN-NI	. . .	. . .	425-540	800-1000
BNI	925-1205	1700-2200	[GES]540	[GES]1000
BCO	1150-1230	2100-2250	[GES]540	[GES]1000
BAU	890-1230	1635-2250	[GES]540	[GES]1000
BPD	1220-1250	2225-2285	[GES]540	[GES]1000

Source: Ref 6

Stainless steels can be soldered using a nickel solderable coating and a gold protection finish to eliminate the need for harsh cleaning treatment.

Ferritic stainless steels usually have less chromium than austenitic grades. Sensitization is faster and at a slightly lower temperature range when compared with austenitic grades. So, the same brazing procedure should be followed to braze ferritic grades. A heating at the sensitization temperature range may eliminate the chromium-depleted zone, during or after brazing.

Ferritic stainless steel brazements are subjected to interfacial corrosion when brazed with some BAg alloys. A filler metal containing nickel can be used to minimize the corrosion rate. A special filler metal, BAg-21, was designed to braze AISI 403 ferritic stainless steel and reduce this corrosion attack.

Because the strength of ferritic grades is drastically reduced above 815 °C (1500 °F), special fixtures must be designed to maintain the shape of the component if brazing is to be performed at higher temperatures.

Martensitic and Precipitation-Hardening Stainless Steels. The brazing of a martensitic or a martensitic-type precipitation-hardening steel is similar to that of a martensitic low-alloy steel. The techniques shown in Fig. 18 can also be applied to these steels. The difference between the coefficients of thermal expansion of a transformed and an untransformed microstructure in a component must be considered to avoid residual stresses and subsequent cracking during cooling.

Some martensitic and precipitation-hardening stainless steels contain aluminum and titanium, which require special fluxes or a suitable atmosphere (dry H_2) to promote wetting. The hydrogen-rich atmosphere may cause hydrogen embrittlement of the martensitic stainless steels.

Duplex stainless steels generally contain a mixed microstructure of 50% ferrite and 50% austenite. They have better corrosion resistance than other stainless steels, and they have better mechanical properties than the austenitic and ferritic grades.

Few published data related to duplex stainless steel brazing are available. Compared with the austenitic grades, however, the duplex grades have higher chromium contents, so wetting and spreading problems may be expected.

Duplex stainless steels are susceptible to several embrittlement mechanisms, particularly chromium carbide formation, σ-phase precipitation, and 475 °C (885 °F) embrittlement. They all impair the corrosion resistance and the toughness of these steels. The effects of chromium carbide precipitation were discussed above. In many aspects, σ-phase precipitation is similar to sensitization in that it generates a chromium-depleted zone and occurs at almost the same temperature range (450 to 900 °C, or 840 to 1650 °F). Small amounts of σ phase (~1%) can reduce Charpy V-notch absorbed energy by 50%. Sigma phase also increases by at least eight times the corrosion rate of duplex stainless steels. To avoid σ-phase precipitation, brazing should be performed at high temperatures, greater than 950 °C (1740 °F).

The 475 °C (885 °F) embrittlement occurs in the ferrite phase at a temperature range from 300 and 500 °C (570 to 930 °F). It is characterized by the formation of a chromium-rich phase (α'), which causes a precipitation-hardening effect in the ferritic phase and, as a consequence, decreases the toughness of the steel. The presence of α' may also impair the corrosion resistance of the steel.

Duplex stainless steels also experience strength loss at temperatures near 1000 °C (1830 °F). Special fixtures must be designed to maintain the shape of the component during brazing.

Cast Irons

The brazing of gray, ductile, and malleable cast irons differs from the brazing of steel in two principal respects: special precleaning methods are necessary to remove graphite from the surface of the iron; and the brazing temperature is kept as low as feasible to avoid reduction in the hardness and strength of the iron. The processes used for brazing cast irons are the same as those used for steel--furnace, torch, induction, and dip brazing.

Because most cast irons are brazed at relatively low temperatures, the filler metals used are almost exclusively silver brazing alloys. Of these silver alloys, BAg-1 is most often used for brazing of cast iron, principally because it has the lowest brazing temperature range. A fluoride-type flux such as AWS type 3A is usually used with BAg-1 filler metal. Soldering of cast iron is not usual.

Relatively high silicon content and sand inclusions on as-cast surfaces have some adverse effects on the brazeability of cast iron. These effects, however, are less significant than the adverse effect of graphite, which is present in all gray, ductile, and malleable cast irons. Graphite has essentially the same effect on machined joint surfaces as on as-cast surfaces. Although gray, ductile, and malleable irons all have lower brazeabilities than carbon or low-alloy steels, the three types of iron are not equal in brazeability.

Malleable iron is generally considered the most brazeable of the three types of cast iron, largely because the total carbon content is somewhat lower (seldom over 2.70 wt%) and, therefore, graphitization is less. Brazeability is also enhanced because the graphite occurs in the form of approximately spherical nodules that are easier to remove or cover up (as by abrasive blasting). Also, malleable iron is lower in silicon than the other types of cast iron and thus is less graphitized, which makes it better suited for brazing.

Ductile iron and gray iron can have nearly the same composition. However, the graphite particles in ductile iron are spheroidal rather than flake-shaped. The spheroidal shape is more favorable for brazing. Gray iron, which is characterized by large flakes of graphite, is the most difficult type of cast iron to braze. Until the development of electrolytic salt-bath cleaning, brazing of gray iron was considered impractical.

A number of methods have been tried for preparing cast iron surfaces for brazing; most of them have been only partly successful. Abrasive blasting with steel shot or grit has proved reasonably successful for preparing the surfaces of ductile and malleable iron castings, but it is seldom suitable for preparing surfaces of gray iron casting. Electrolytic treatment in a molten salt bath, alternatively reducing and oxidizing, has been the most successful method for surface preparation and is applicable to all graphitic cast irons. Ordinary chemical cleaning methods, such as degreasing, detergent washing, or acid pickling, have the distinct disadvantage of not removing surface carbon, which interferes with bonding.

Before any procedure for cleaning is adopted, tests should be made by cleaning samples of the iron intended for use in the castings to be brazed, fluxing the samples, and applying filler metal (preferably on a smooth, flat surface). The samples are then heated to the pre-established brazing temperature, cooled, and examined visually. If the samples show that the filler metal has not uniformly wetted the test piece, the surface is not sufficiently clean.

Aluminum Alloys

Brazing of aluminum alloys was made possible by the development of fluxes that disrupt the oxide film on aluminum without harming the underlying metal and filler metals (aluminum alloys) that have suitable melting ranges and other desirable properties, such as corrosion and mechanical resistance. The aluminum-base filler metals used for brazing aluminum alloys have liquidus temperatures much closer to the solidus temperature of the base metal than those for brazing most other metals. For this reason, close temperature control is required in brazing aluminum. The brazing temperature should be approximately 40 °C (70 °F) below the solidus temperature of the base metal.

The non-heat-treatable wrought aluminum alloys, such as the 1*xxx,* 3*xxx,* and 5*xxx* (low-magnesium) series, have been brazed successfully. Alloys that contain higher magnesium contents are more difficult to braze by the usual flux methods because of poor wetting and excessive penetration by the filler metal. Filler metals that melt below the solidus temperatures of most commercial, non-heat-treatable wrought alloys are available.

The most commonly brazed heat-treatable wrought alloys are those of the 6*xxx* series. The 2*xxx* and 7*xxx* series of aluminum alloys have low melting points and therefore are not normally brazeable. Alloys 7072 (used as a cladding material only) and 7005 are the few exceptions. Alloys that have solidus temperatures above 595 °C (1100 °F) are easily brazed with commercial binary aluminum-silicon filler metals. Higher-strength, lower-melting-point alloys can be brazed with proper attention to filler metal selection and temperature control, but the brazing cycle must be short to minimize penetration by the molten filler metal. Sand and permanent mold casting alloys with high solidus temperatures are brazeable.

Commercial filler metals for brazing aluminum are aluminum-silicon alloys containing 7 to 12 wt% Si. Brazing fillers with lower melting points are attained, with some sacrifice in resistance to corrosion, by adding copper and zinc. Filler metals for vacuum brazing of aluminum usually contain magnesium.

The optimum brazing temperature range for an aluminum-base filler metal is determined by the melting range of the filler metal, the amount of molten filler metal needed to fill the joint, and the mutual solubility between the filler metal and the base metal.

Filler metals for separate application from the base metal to be brazed are available as wire and sheet.

Most filler metals are used for any of the common brazing processes and methods. Two alloys, 4004 and 4104, have been developed exclusively for use in fluxless vacuum brazing. They contain additions of magnesium and magnesium-bismuth, respectively, with a brazing temperature range of approximately 590 to 605 °C (1090 to 1120 °F). Similarly, a proprietary mixture of filler metal BAlSi-4 (alloy 4047) in powder form and a chemical compound is used exclusively with dip brazing. The manufacture of filler metal in sheet and wire forms becomes more difficult as the silicon content increases. Only BAlSi-2 (alloy 4343), BAlSi-4 (alloy 4047), and alloy 4004 are available as sheet.

Brazing of aluminum to other alloys should be performed only after considering the differences between:

- COEFFICIENTS OF THERMAL EXPANSION, WHICH CAN AFFECT THE RESIDUAL STRESS LEVEL AND CHANGE THE CLEARANCE OF THE JOINT
- THERMAL CONDUCTIVITIES, WHICH CAN GENERATE DIFFERENT HEATING RATES
- CORROSION RESISTANCES OF ALUMINUM AND THE OTHER MATERIAL, WHICH CAN GENERATE A GALVANIC CORROSION
- MECHANICAL PROPERTIES

Brazing of Aluminum to Ferrous Alloys. The steel component should be protected from oxidation during preheating and brazing to aluminum. In dip brazing, oxidation can be prevented by dipping unheated parts into molten flux, but this procedure has limited application because it is likely to cause warping and misalignment of the components.

Plated or coated steel can be brazed to aluminum more readily than bare steel. Copper, nickel, or zinc electroplates and aluminum, silver, tin, or hot dip zinc coatings are used to promote wetting of the steel and to minimize formation of brittle aluminum-iron compounds, thus producing a more ductile joint.

Brazing of aluminum to copper is difficult, because of the low melting temperature, 548 °C (1018 °F), of the aluminum-copper eutectic and its extreme brittleness. The eutectic is formed due to the dissolution of aluminum during brazing. By heating and cooling rapidly, however, it is possible to make reasonably ductile joints for applications such as copper inserts in aluminum castings for electrical conductors. The usual filler metals and fluxes for aluminum-to-aluminum brazing can be used. Silver alloy filler metals, BAg-1 and BAg-1a, can also be used if heating and cooling are rapid (to minimize interdiffusion). Hot dipping the copper surfaces with solder or silver alloy filler metal improves wetting and permits shorter time at brazing temperatures. An example of brazing aluminum tubing to copper tubing is to use a transition tubing of steel with aluminum coating on one end of the tubing. The aluminum-coated end of the tubing is brazed to the aluminum, and the other end is silver brazed to the copper tubing.

Brazing of Aluminum to Other Nonferrous Metals. Aluminum-silicon filler metals are unsuitable for brazing aluminum to uncoated titanium because of the formation of brittle intermetallic compounds (titanium-aluminum and titanium-silicon). Titanium can be hot dip coated with aluminum, however, and then brazed to aluminum with the usual aluminum filler metals.

Under adequate conditions, nickel and nickel alloys are no more difficult to braze to aluminum than ferrous alloys. They can be brazed directly or precoated with aluminum. Although Monel alloys can be wetted directly, brazed joints are likely to be brittle; thus, Monel alloys are preferably precoated with aluminum.

Beryllium can be wetted directly by aluminum brazing alloys. Magnesium alloys can be brazed to aluminum, but the brazed joints have limited usefulness because of the extremely brittle aluminum-magnesium phases that form at the interface.

Soldering of aluminum and aluminum alloys is relatively simple. Compared with brazing, soldering presents advantages such as little loss of base metal temper, minimal distortion, and easy removal of flux. Aluminum is soldered at a minimum of 110 °C (200 °F) below the solidus temperature of the base metal.

Wetting and spreading are affected by the presence of an oxide at the faying surface of the joint. The nature of the aluminum oxide is different from alloy to alloy. The heat-treatable alloys present a more tenacious oxide that is more adherent and more difficult to remove. The oxides of the non-heat-treatable alloys are less tenacious and easier to remove. The action of a corrosive flux is enough to disrupt and displace the oxide layer in a non-heat-treatable aluminum alloy. However, the oxide layer of a heat-treatable alloy must be removed by chemical or mechanical means before soldering.

Solderability of aluminum alloys is influenced by alloying elements of the base metal. Generally, purer aluminum alloys are easily soldered. Table 9 presents the solderabilities of aluminum alloys. Analyzing this table, one can see that alloys of groups 1*xxx,* 3*xxx,* and 6*xxx* have good solderabilities.

TABLE 9 SOLDERABILITY OF ALUMINUM ALLOYS

ALLOY GROUP	TYPICAL ALLOY	SOLDERABILITY
1*XXX* (COMMERCIAL PURITY OR HIGHER)	1060	GOOD
	1100	GOOD
2*XXX* (AL-CU)	2014	FAIR(A)
3*XXX* (AL-MN)	3003	GOOD
4*XXX* (AL-SI)	4043	POOR(B)
5*XXX* (AL-MG OR AL-MG-MN)(C)	5005	GOOD
	5050, 5154	FAIR(A)
	5456, 5083	FAIR(A)
6*XXX* (AL-MG-SI)	6061	GOOD(A)
7*XXX* (AL-ZN)(C)	7072	GOOD
	7075	POOR
8*XXX* (AL-OTHER)(C)	8112	GOOD

Source: Ref 7

(A) SUSCEPTIBLE TO INTERGRANULAR PENETRATION BY SOLDER.

(B) SOLDERABLE ONLY WITH ABRASION OR ULTRASONIC TECHNIQUES.
(C) SOLDERABILITY GREATLY AFFECTED BY COMPOSITION.

Solderability is primarily affected by two alloying elements, magnesium and silicon. The presence of magnesium in an aluminum alloy not only reduces the wettability (more than 1 wt% Mg) but also increases the intergranular penetration (more than 0.5 wt% Mg). Magnesium content up to 1.0 wt% does not reduce the flux effectiveness, so it does not affect wetting and spreading of the molten filler metal. Alloys with less than 1.0 wt% Mg can be soldered with all flux types. Between 1.0 and 1.5 wt% Mg, the low-temperature organic-type flux is not effective to remove the surface oxide of the faying surface. When the amount of magnesium exceeds 1.5 wt%, the corrosive flux does not work either. Silicon plays the same role. Thus, if the amount of silicon is higher than 4.0 wt%, all flux types are ineffective. To solve this problem, a fluxless technique such as abrasion or ultrasonic soldering should be utilized.

Typical solders used for aluminum are shown in Table 10. The corrosion resistances of the filler metals are rated in this table. Solders that melt below 260 °C (500 °F) are called low-melting-point solders. Those that melt between 260 and 370 °C (500 and 700 °F) are called intermediate-melting-point solders. Those that melt between 370 and 440 °C (700 and 820 °F) are called high-melting-point solders.

TABLE 10 COMPOSITION AND TYPICAL PROPERTIES OF SOLDERS USED WITH ALUMINUM

SOLDER TYPE	COMPOSITION, WT%						MELTING RANGE, SOLIDUS - LIQUIDUS(A)		WETTING ABILITY(B)	FLUX TYPE(C)	CORROSION RESISTANCE(B)
	Sn	Zn	Al	Cd	Pb	Cu	°C	°F			
ZINC	. . .	100	. . .	. . .	. . .	. . .	419	787	G	C	VG
	. . .	94	4	. . .	. . .	2	380-395	720-740	G	C	VG
	. . .	95	5	. . .	. . .	. . .	375	710	G	C	VG
	. . .	90	5	. . .	. . .	5	380	720	G	C	VG
	2	79.6	10	0.4	3	5	275-400	527-750	. . .	. . .	. . .
ZINC-CADMIUM	. . .	90	. . .	10	. . .	. . .	265-405	509-760	G	C	F
	. . .	60	. . .	40	. . .	. . .	265-335	509-635	VG	C	F
	. . .	17.5	. . .	82.5	. . .	. . .	265	509	. . .	. . .	. . .
TIN-ZINC	20	15	0.8	64.2	. . .	. . .	110-275	230-530	. . .	. . .	. . .
	70	30	. . .	. . .	. . .	. . .	200-310	390-592	F	C	F
	30	70	. . .	. . .	. . .	. . .	200-190	390-710	G	C	G
TIN-LEAD	40	. . .	. . .	. . .	60	. . .	185-240	361-460	. . .	. . .	. . .
TIN-ZINC	60	39.4	. . .	. . .	0.1	0.5	200-340	390-645	G	C	G
TIN-LEAD	63	. . .	. . .	. . .	37	. . .	185-215	361-420	. . .	. . .	. . .
TIN-ZINC	69.3	28	0.7	. . .	2.0	. . .	195-335	385-635	. . .	. . .	. . .
	80	20	. . .	. . .	. . .	. . .	200-275	390-530	. . .	. . .	. . .
	91	9	. . .	. . .	. . .	. . .	205	400	F	OC	F

TIN-LEAD	36.9	. . .	. . .	3.8	59.3	. . .	145-230	290-450	. . .	. . .	. . .
	34	3	. . .	. . .	63	. . .	195-255	383-492	P	OC	P
	31.6	9	. . .	8	51	0.4	140-250	282-485	. . .	. . .	. . .
	40	15	0.8	. . .	44.2	. . .	170-355	335-675	. . .	. . .	. . .
TIN-CADMIUM	20	15	0.8	64.2	. . .	. . .	110-275	230-530	. . .	. . .	. . .
	50	. . .	. . .	. . .	50	. . .	180-215	360-420	P	OC	P
TIN-ZINC	91	9	. . .	. . .	. . .	. . .	200	391	F	O	P

Source: Ref 8

(A) SOLDERS WITH ONLY ONE TEMPERATURE GIVEN ARE EUTECTICS.
(B) VG, VERY GOOD; G, GOOD; F, FAIR; P, POOR.
(C) C, CORROSIVE; OC, ORGANIC CORROSIVE; O, ORGANIC.

Aluminum alloys can be soldered to all usual metals and nonmetallic materials. However, loss of corrosion resistance should be expected. The high-melting-point solders are suitable for soldering mild steel, stainless steel, nickel, copper, brass, zinc, and silver directly to aluminum. Magnesium, titanium, zirconium, niobium, tantalum, molybdenum, and tungsten may be soldered if they are plated with a solderable metal coating such as silver.

Copper and Copper Alloys

Most coppers and copper alloys can be brazed satisfactorily using one or more of the conventional brazing processes: furnace, torch, induction, resistance, and dip brazing. Their brazeabilities are rated from good to excellent.

Brazing of Tough Pitch Coppers. Tough pitch coppers are subject to embrittlement when heated at temperatures above 480 °C (900 °F) in reducing atmospheres containing hydrogen. Phosphorus-deoxidized and oxygen-free coppers can be brazed without flux in hydrogen-containing atmospheres without risk of embrittlement, provided self-fluxing filler metals (BCuP series) are used.

The coppers, including those that contain small additions of silver, lead, tellurium, selenium, or sulfur (generally no more than 1 wt%), are readily brazed with the self-fluxing BCuP filler metals, but wetting action is improved when a flux is used and when a sliding motion between components is provided while the filler metal is molten. Precipitation-hardenable copper alloys that contain beryllium, chromium, or zirconium form oxide films that impede the flow of filler metal. To ensure proper wetting action of the joint surface by the filler metal, beryllium copper parts, for example, should be freshly machined or mechanically abraded before being brazed. Removal of beryllium oxide from joint surfaces requires the use of a high-fluoride-content flux.

Brazing of Red and Yellow Brasses. Red and yellow brasses are readily brazed with a variety of filler metals. Flux is normally required for best results, especially when the zinc content is above 15 wt%. Low-melting filler metals should be used to avoid dezincification of the yellow brasses. If added to red brass or yellow brass, lead forms a dross on heating that can seriously impede wetting and the flow of filler metal. Consequently, in brazing leaded brasses, the use of a flux is mandatory to prevent dross formation in the joint area.

Brazing of Tin-Containing Brasses. Tin-containing brasses, which include admiralty brass, naval brass, and leaded naval brass, contain up to 1 wt% Sn and may contain other alloying elements, such as lead, manganese, arsenic, nickel, and aluminum. Except for the aluminum-containing alloys, these brasses are readily brazed; they have greater resistance to thermal shock and are less susceptible to hot cracking than the high-lead brasses. For proper wetting, brasses that contain aluminum require a special flux.

Brazing of Phosphor Bronzes. Phosphor bronzes contain small amounts of phosphorus, up to approximately 0.25 wt%, added as a deoxidizer. Although susceptible to hot cracking in the coldworked condition, alloys in this group have good brazeability and are adaptable to brazing with any of the common filler metals that have melting temperatures lower than that of the base metal. The use of a flux is generally preferred. To avoid cracking, parts made from phosphor bronzes should be stress-relieved at approximately 290 to 345 °C (550 to 650 °F) before brazing. Silicon bronzes, which contain up to about 3.25 wt% Si, when applied in a highly stressed condition, are susceptible to hot shortness and stress cracking by molten filler metal. To avoid cracking, the alloys should be stress-relieved at about 290 to 345 °C (550 to 650 °F) before brazing.

Because of the formation of aluminum oxide on the surface, aluminum bronzes are generally considered difficult to braze. The oxide, which inhibits the flow of filler metal, cannot be reduced in dry hydrogen. However, alloys containing 8 wt% Al or less are brazeable, provided AWS type 4 flux is used to dissolve the aluminum oxide. Use of the low-melting, high-silver filler metals is recommended for these bronzes.

Brazing of Copper Alloys to Other Alloys. Copper and copper alloys can be brazed to other alloys. Some characteristics to consider are: coefficients of thermal expansion, thermal conductivities, corrosion behavior, and mechanical properties.

Copper nickels, which may contain from about 5 to 40 wt% Ni, are susceptible both to hot cracking and to stress cracking by molten filler metal. The silver alloy filler metals (BAg series) are preferred for brazing these alloys. In general, the use of filler metals containing phosphorus should be avoided, because the copper nickels are susceptible to the formation of brittle nickel phosphides at the interface. Nickel silvers (brasses that contain up to about 20 wt% Ni but do not contain silver) are highly susceptible to hot cracking and should be stress-relieved at about 290 °C (550 °F) before being brazed. They should be heated and cooled uniformly because of their low thermal conductivity.

Most of the alloys belonging to any one of the above groups can be brazed to an alloy of another group. However, to achieve compatibility, some compromise may be required in the selection of brazing temperature, filler metal, and flux. For example, if a copper component is to be brazed to a component made of aluminum bronze, the brazing temperature should be predicated on the lower melting temperature of the bronze, and a suitable flux should be selected to accommodate the bronze.

Soldering of Copper and Copper Alloys. Copper and copper alloys are among the most frequently soldered engineering materials. The copper oxide is easily disrupted and displaced by most flux types. The presence of alloying elements such as beryllium, chromium, silicon, and aluminum modifies the nature of the oxide, making it more tenacious. For these alloys, a special flux is recommended to remove the oxide from the surface and enhance the solderability of these base metal groups.

The most common solders for copper are tin- or lead-base solders. Tin can react with copper and form two intermetallic phases, Cu_6Sn_5 and Cu_3Sn, at the solid-liquid interface. As the thickness of this reaction layer increases, the mechanical behavior of the joint became impaired. Therefore, a thin intermetallic layer is desirable. The kinetics of intermetallic phase growth are controlled by the amount of tin and other elements in the solder, the soldering temperature, and time. Thus, to minimize the thickness of the intermetallic phase, a low-tin solder or a short soldering time should be utilized. Another way to minimize the intermetallic formation is to use a barrier coating, such as nickel, silver, or gold. These copper-tin intermetallics may also cause a dewetting phenomenon. Dewetting is defined as the withdrawal of molten solder from a surface that was previously wetted. If the solderable surface is not well protected during soldering, the intermetallic can oxidize and the dewetting phenomenon takes place.

Nickel-Base Alloys

In the selection of a brazing process for a nickel-base alloy, the characteristics of the alloy must be carefully considered. The nickel-base alloy family includes alloys that differ significantly in physical metallurgy (such as precipitation-strengthened versus solid-solution strengthened) and in process history (such as cast versus wrought). These characteristics can have a profound effect on their brazeability.

Precipitation-hardenable alloys present several difficulties not normally encountered with solid-solution alloys. Precipitation-hardenable alloys often contain appreciable (greater than 1 wt%) quantities of aluminum and titanium. The oxides of these elements are almost impossible to reduce in a controlled atmosphere (vacuum, hydrogen). Therefore, nickel plating or the use of a flux is necessary to obtain a surface that allows wetting by the filler metal.

Because these alloys are hardened at temperatures of 540 to 815 °C (1000 to 1500 °F), brazing at or above these temperatures may alter the alloy properties. This frequently occurs when using silver-copper (BAg) filler metals, which occasionally are used on heat-resistant alloys.

Liquid metal embrittlement is another difficulty encountered in brazing precipitation-hardenable alloys. Many nickel-base, iron-base, and cobalt-base alloys crack when subjected to tensile stresses in the presence of molten metals. However, liquid-metal-induced cracking of these alloys is usually confined to the silver-copper (BAg) filler metals. If precipitation-hardenable alloys are brazed in the hardened condition, residual stresses are often high enough to initiate cracking.

Oxide dispersion-strengthened alloys (ODS) are powder metallurgy (P/M) alloys that contain stable oxide, evenly distributed throughout the matrix. The oxide does not go into solution in the alloy, even at the liquidus temperature of the matrix. However, once the oxide particles are rejected into the molten matrix (for example, during fusion welding), they cannot be redistributed evenly in the matrix on solidification. Therefore, these alloys are usually joined by brazing. There are two commercial alloy classes of ODS alloys: the dispersion-strengthened nickel, and the mechanically alloyed Inconel MA 754, Inconel MA 6000, and Incoloy MA 956.

Inconel MA 754, dispersion-strengthened nickel alloys, and dispersion-strengthened nickel-chromium alloys are the easiest to braze of the ODS alloys. Vacuum, hydrogen, or inert atmospheres can be used for brazing. Prebraze cleaning consists of grinding or machining with a solvent that evaporates without leaving a residue. Generally, brazing temperatures should not exceed 1315 °C (2400 °F) unless demanded by a specific application that has been well examined and tested. The brazing filler metals for use with these ODS alloys usually are not classified by AWS. In most cases, the brazing filler metals used with these alloys have brazing temperatures in excess of 1230 °C (2250 °F); these filler metals include proprietary alloys based on nickel, cobalt, gold, or palladium.

Inconel MA 6000 is a nickel-base ODS alloy that is also γ'-strengthened. The number of alloying elements plus the γ' precipitation presents a formidable challenge to the joining of this alloy. Inconel MA 6000 has a solidus temperature of 1300 °C (2372 °F); therefore, the brazing temperature should be no higher than 1250 °C (2280 °F). Additionally, because 1230 °C (2250 °F) is the γ' solution treatment temperature, it becomes important to carefully select the brazing filler metal and to heat treat the assembly after brazing. The BNi, BCo, and specially formulated filler metals have been used in this alloy. Inconel MA 6000 is used for its high-temperature strength and corrosion resistance; unfortunately, the passive oxide scale that provides good corrosion resistance also prevents wetting and flow of brazing filler metal. Therefore, correct cleaning procedures are very important.

The brazeability of these alloys is affected by the chromium content, oxide dispersion, and the manufacturing process of the component. Chromium forms a stable and adherent oxide that impairs the wetting and spreading of the molten filler metal. The oxide dispersion also may affect the wetting and spreading behavior since an oxide is not easily wettable. The P/M manufacturing process can introduce some porosity in the component, which will act as a barrier to molten metal flow.

References cited in this section

3. M.G. NICHOLAS AND D.A. MORTIMER, CERAMIC/METAL JOINING FOR STRUCTURAL APPLICATIONS, *MATER. SCI. TECHNOL.,* VOL 1 (NO. 9), 1985, P 657-665
5. *SOLDERING MANUAL,* AMERICAN WELDING SOCIETY, 1977
6. J.A. PASK, FROM TECHNOLOGY TO THE SCIENCE OF GLASS/METAL AND CERAMIC/METAL SEALING, *AM. SOC. CERAM. BULL.,* VOL 66 (NO. 11), 1987, P 1587-1592
7. H.H. MANKO, *SOLDERS AND SOLDERING,* 3RD ED., MCGRAW-HILL, 1992
8. *ALUMINUM SOLDERING HANDBOOK,* ALUMINUM ASSOCIATION, 1974

S.D. Brandi, Escola Politécnica da Universidade de São Paulo, Brazil; S. Liu, Colorado School of Mines; J.E. Indacochea and R. Xu, University of Illinois at Chicago

Heat-Resistant Alloys

Heat-resistant alloys are frequently referred to as superalloys because of their strength, oxidation resistance, and corrosion resistance at elevated temperatures. Superalloys can be subdivided into two categories: conventional cast and wrought alloys, and P/M products. Powder metallurgy products may be produced as conventional alloy compositions and as ODS alloys. Almost any metal or nonmetallic can be brazed to these heat-resistant alloys if it can withstand the heat of brazing.

The American Welding Society has classified several gold-base, nickel-base, and cobalt-base brazing filler metals that can be used for elevated-temperature service. In addition, there are many that are not classified by AWS. It should be noted that for lower service temperatures, copper (BCu) and silver (BAg) brazing filler metals have been used for many successful applications.

Generally, heat-resistant alloys are brazed with nickel-base or cobalt-base alloys containing boron and/or silicon, which serve as melting-point depressants and as oxide-reducing agents. In many commercial brazing filler metals, the levels are 2 to 3.5 wt% B and 3 to 10 wt% Si. Phosphorus is another effective melting-point depressant for nickel and is used in filler metals from 0.02 to 15 wt%. It is also used where good flow is important in applications of low stress, where temperatures do not exceed 760 °C (1400 °F). Chromium is often present to provide oxidation and corrosion resistance. The amount may be as high as 20 wt%, depending on the service conditions. Higher amounts, however, tend to lower brazement strength.

Cobalt-base filler metals are used mainly for brazing cobalt-base components, such as first-stage turbine vanes for jet engines. Most cobalt-base filler metals are proprietary. In addition to boron and silicon, these alloys usually contain chromium, nickel, and tungsten to provide corrosion and oxidation resistance and to improve strength. As discussed in the section "Stainless Steels" of this article, the presence of chromium in higher amounts affects the wetting and spreading behavior of the flux during brazing.

Titanium and Titanium Alloys

Titanium is one of the chemical elements that reacts readily with oxygen to form an adherent and stable oxide. This oxide gives titanium and titanium alloys excellent corrosion resistance. Properties such as corrosion resistance, light weight, and high strength make titanium especially attractive in aerospace and chemical applications.

On cooling, pure titanium undergoes an allotropic transformation from β(bcc) to α(hcp) at 882.5 °C (1621 °F). Some alloying elements stabilize the α phase, others, the β phase. Figure 19 shows the αand β stabilizers and their effects on a binary phase diagram. In this case, α and β represent a solid solution and γ represents an intermetallic compound. Table 11 shows the classification for some commercial titanium alloys according to the prevailing microstructure.

TABLE 11 CLASSIFICATION OF SELECTED COMMERCIAL TITANIUM ALLOYS

ALLOY	CLASSIFICATION
TI-5AL-2.5SN	α
TI-8AL-1MO-1V(A)	α + β
TI-6AL-2SN-4ZR-2MO(A)	α + β
TI-6AL-4V	α + β
TI-6AL-2SN-6V	α + β
TI-3AL-2.5V	α + β
TI-6AL-2SN-4ZR-6MO(B)	α + β
TI-5AL-2SN-2ZR-4CR-4MO(B)	α + β

TI-3AL-10V-2FE(B)	α + β
TI-13V-11CR-3AL	β
TI-15V-3CR-3AL-3SN	β
TI-4MO-8V-6CR-4ZR-3AL	β
TI-8MO-8V-2FE-3AL(C)	β
TI-11.5MO-6ZR-4.5SN	β

Source: Ref 9

(A) NEAR- α; THE TERMS "LEAN- β " AND "SUPER- α " MAY ALSO BE USED.

(B) NEAR- β.

(C) OBSOLETE ALLOY.

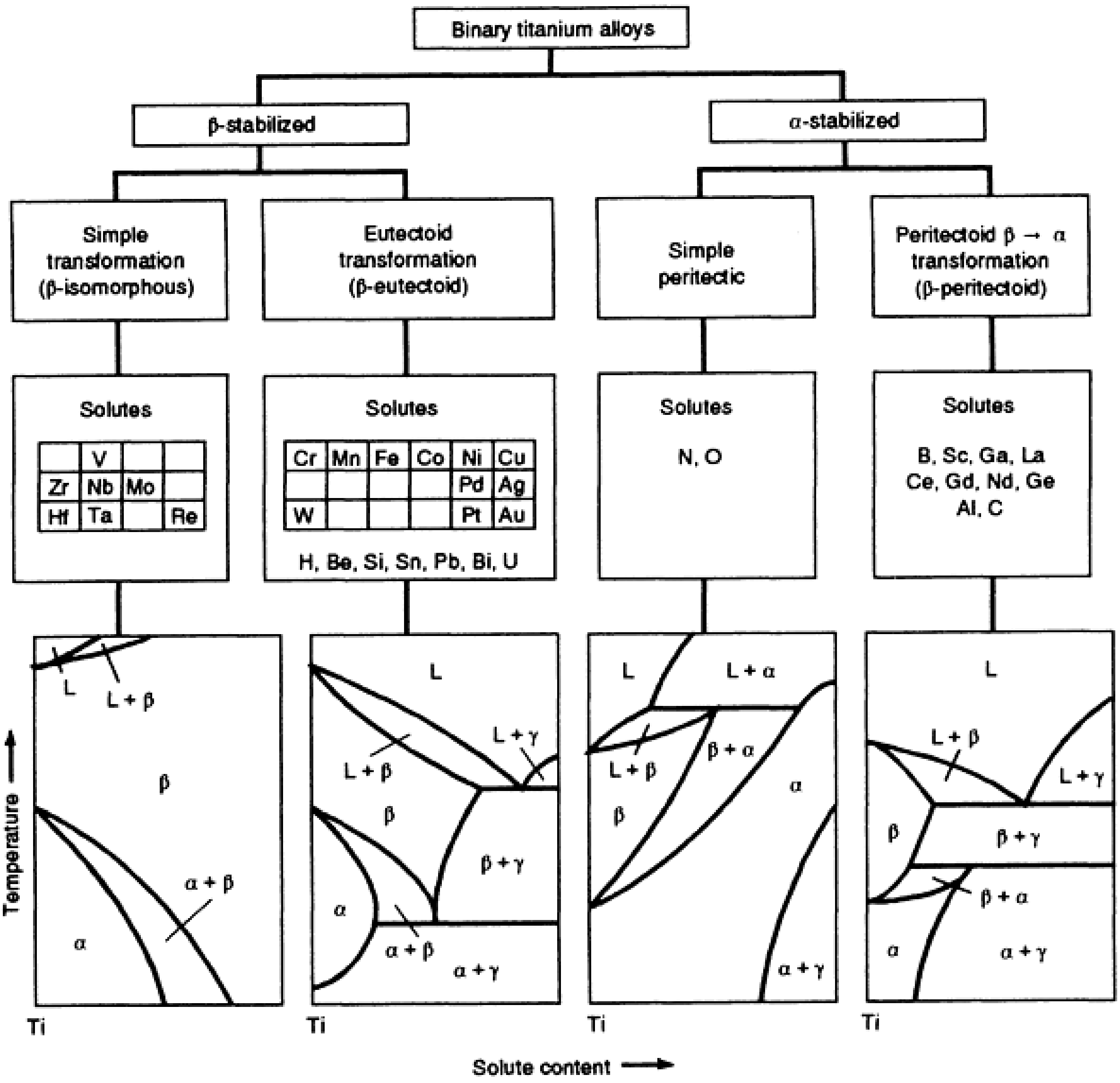
Binary titanium alloys
β-stabilized
α-stabilized
Simple transformation (β-isomorphous)
Eutectoid transformation (β-eutectoid)
Simple peritectic
Peritectoid β → α transformation (β-peritectoid)
Solutes
V
Zr Nb Mo
Hf Ta Re
Cr Mn Fe Co Ni Cu
Pd Ag
W Pt Au
H, Be, Si, Sn, Pb, Bi, U
N, O
B, Sc, Ga, La
Ce, Gd, Nd, Ge
Al, C
L
L + β
β
α + β
α
L + γ
β + γ
α + γ
L + α
β + α
Temperature
Ti
Solute content

FIG. 19 FLOWCHART SHOWING EFFECT OF α' AND β STABILIZERS ON A BINARY PHASE DIAGRAM. SOURCE: REF 9

The β alloys present a ductile-to-brittle transition temperature (DBTT). Thus, depending on the temperature, β alloys can be brittle. On the other hand, α alloys have no DBTT. These characteristics determine some applications of the two alloys. The α alloys have, in general, good strength, toughness, and creep resistance, and they are suitable for cryogenic applications. The β alloys are not designed for low-temperature applications, but they present good formability and a high strength-to-weight ratio. The α+ βalloys are also formable with high room-temperature strength and moderate elevated-temperature strength. The volumetric fraction of β in these alloys is between 10 to 50%, but alloys with over 20 vol% β phase are not joinable.

Titanium and titanium alloys are not directly solderable due to the formation of an adherent stable oxide.

Commercially Pure Titanium, α and Near-α Titanium Alloys. Because these alloys are not strengthened by heat treatment, the brazing thermal cycle has little effect on their mechanical properties. The filler metals used to braze titanium and titanium alloys are shown in Table 12. During brazing there is a dissolution of the base metal and a diffusion of the alloying elements toward the base metal and/or the filler metal. These phenomena will form a reaction layer, which may consist of brittle titanium intermetallics. Titanium can combine with copper, nickel, aluminum, and silver to form intermetallics. The first three intermetallics (titanium-copper, titanium-nickel, and titanium-aluminum) are harder and have a lower ductility than a titanium-silver intermetallic. Therefore, the filler metal should have a lower amount of these elements, or the brazing process should be as fast as possible, to minimize the intermetallic formation.

TABLE 12 TEMPERATURE SPECIFICATIONS FOR FILLER METALS USED TO BRAZE TITANIUM ALLOYS

FILLER METAL	SERVICE TEMPERATURE		BRAZING TEMPERATURE	
	°C	°F	°C	°F
AG	425	800	925	1700
AG-5AL, AG-7.5CU	425	800	870-925	1600-1700
AG-5AL-0.5MN	425	800	870-900	1600-1650
AL, AL-SI (4040), AL-MN (3003)	260	500	650-690	1200-1275
TI-48ZR-4BE, TI-43ZR-12NI-2BE	540	1000	870-1095	1600-2000
AG-9PD-9GA	. . .	. . .	880-920	1615-1690

Source: Ref 4

Titanium and titanium alloys are very resistant to corrosion. The filler metal should be selected carefully to avoid galvanic corrosion.

Titanium and its alloys are usually brazed by induction or furnace processes with a protective atmosphere. The brazing atmosphere is usually a vacuum with less than 13 mPa (10^{-4} torr) or an inert atmosphere with dew point lower than -55 °C (-70 °F). Vacuum brazing with filler metals that contain silver or gallium must be performed with an argon back pressure to avoid the vaporization loss of these elements.

In induction brazing, the chemical elements of the filler metal must alloy readily with titanium, due to the fast heating cycle. However, furnace brazing requires a filler metal with a chemical composition that does not alloy excessively with titanium, due to longer periods at elevated temperatures. Torch brazing is not usually done because special fluxes and a skilled operator are required.

β and α+ β titanium alloys can be strengthened by heat treatment. The basic heat treatment consists of heating in a single-phase field (β alloys) or in a two-phase field (α + βalloys), followed by a quenching. The aging is performed between 480 and 650 °C (900 and 1200 °F). During aging, α or other compounds will precipitate in a β matrix, increasing the strength and toughness of the alloy. Thus, the thermal cycle during brazing may affect the mechanical properties of the base metal. The β alloys should be brazed at a temperature close to the solubility temperature. Brazing at higher temperatures will reduce the ductility of these alloys. The α + β alloys should be brazed at a temperature below the α-to-β transition temperature. Table 13 shows the α-to-β transition temperature for some titanium alloys.

TABLE 13 α-TO-β TRANSITION TEMPERATURE FOR TITANIUM AND TITANIUM ALLOYS

ALLOY	α -TO- β TRANSITION

	TEMPERATURE	
	°C ± 15°	**°F ± 25°**
COMMERCIALLY PURE TI, 0.25 MAX O_2	915	1675
COMMERCIALLY PURE TI, 0.40 MAX O_2	945	1735
α AND NEAR- α ALLOYS		
TI-5AL-2.5SN	1050	1925
TI-8AL-1MO-1V	1040	1900
TI-6AL.4ZR-2MO-2SN	995	1820
TI-6AL-2CB- 1TA-0.8MO	1015	1860
TI-0.8NI-0.3MO	880	1615
α - β ALLOYS		
TI-6AL-4V	1000(A)	1830(B)
TI-6AL-6V-2SN	945	1735
TI-3AL-2.5V	935	1715
TI-6AL-6MO-4ZR-2SM	940	1720
TI-7AL-4MO	1005	1840
TI-8MN	800(C)	1475(D)
β OR NEAR- β ALLOYS		
TI-13V-11CR-3AL	720	1330
TI- 11.5MO-6ZR-4.5SN	760	1400
TI-3AL-8V-6CR-4ZR-4MO	795	1460

Source: Ref 4

(A)	±20°.
(B)	±30°.
(C)	±35°.
(D)	±50°.

The fixture material for brazing a titanium alloy should be chosen carefully to avoid contamination or solid-state bonding between them. For example, nickel forms an eutectic with titanium at 940 °C (1725 °F) and can be bonded to titanium at approximately 816 °C (1500 °F). Thus, a physical contact between titanium alloys and nickel-containing alloys, such as austenitic stainless steels, should be avoided.

Cobalt-Base Alloys

The brazing of cobalt-base alloys is readily accomplished with the same techniques used for nickel-base alloys. Because most of the popular cobalt-base alloys do not contain appreciable amounts of aluminum or titanium, brazing atmosphere requirements are less stringent. Several cobalt-base alloys can be brazed in either a hydrogen atmosphere or a vacuum. Filler metals are usually nickel-base alloys, cobalt-base alloys, or gold-palladium compositions. Silver or copper brazing filler metals may not have sufficient strength and oxidation resistance in many high-temperature applications. Although cobalt-base alloys do not contain appreciable amounts of aluminum or titanium, an electroplate or flash of nickel is often used to better promote wetting of the brazing filler metal.

Cobalt alloys, much like nickel alloys, may become susceptible to liquid metal embrittlement or stress-corrosion cracking when brazed under residual or dynamic stresses. This frequently is observed when using silver or silver-copper (BAg) filler metals. Liquid metal embrittlement of cobalt-base alloys by copper (BCu) filler metals occurs with or without the application of stress; therefore, BCu filler metals should be avoided when brazing cobalt.

Refractory Metals

The mechanical properties of the refractory metals are affected markedly by their DBTT behavior, recrystallization temperature, and reactions with carbon and selected gases. These characteristics must be considered when procedures for brazing refractory metals are established. The strength and ductility of the refractory metals are adversely affected by microstructural changes that occur when the recrystallization temperatures of these metals are exceeded.

The recrystallization temperature range also varies with alloying additions, interstitial content, fabrication method (including degree of cold working), and time at temperature. Some applications permit brazing with brazing filler metals that melt below the recrystallization temperature range. Other applications require the use of brazing filler metals that melt above this temperature range. As a result, a joint must be designed to accommodate the loss in mechanical properties associated with recrystallization.

The environment in which the refractory metals are brazed is determined by the reactivity of these metals with oxygen, hydrogen, carbon, and nitrogen, and the effect of these elements on the mechanical properties of the refractory metals. All of the refractory metals react with oxygen at moderately elevated temperatures, but they form different types of oxides. Niobium and tantalum form hard adherent oxides at temperatures above 205 and 400 °C (400 and 750 °F), respectively. On the other hand, molybdenum and tungsten form volatile oxides at temperatures above 400 and 510 °C (750 and 950 °F), respectively. In either case, the surfaces of the refractory metals must be protected from oxidation during brazing to ensure wetting by the braze filler metal. Also, these metals must be coated with an oxidation-resistant material, such as nickel or silver plating, if they are exposed to air at elevated temperatures. For such service conditions, the brazing filler metal must be compatible with both the base metal and the coating.

Niobium is used mainly for nuclear and aerospace applications. As a result, research has been directed toward the development of brazing filler metals with characteristics that are compatible with the base metal and its intended use. For example, brazing filler metals have been developed to produce joints that are resistant to liquid alkali metals, such as sodium, and that possess useful properties at 705 to 815 °C (1300 to 1500 °F). The base metal Nb-lZr has been vacuum brazed at 1250 °C (2280 °F) with the Ti-28V-4Be brazing filler metal.

Other niobium alloys--for example, D-43 (Nb-10W-IZr-0.1C), Cb-752 (Nb-10W-2.5Zr), and C-129Y (Nb-10W-11Hf-0.07Y)--have been successfully brazed with two brazing filler metals (B120VCA and Ti-8.5si) at 1455 °C (2650 °F).

Molybdenum. Copper-base and silver-base brazing filler metals can be used to braze molybdenun for low-temperature service. For high-temperature applications, molybdenum can be brazed with gold, palladium, and platinum filler metals, nickel-base filler metals, reactive metals, and refractory metals that melt at lower temperatures than molybdenum. It should be noted, however, that nickel-base alloys have limited applicability for high-temperature service, because nickel and molybdenum form an intermetallic compound that melts at 1315 °C (2400 °F). Two binary brazing filler metals (V-35Nb and Ti-30V) have been evaluated for use with the Mo-0.5Ti molybdenum base metal.

Molybdenum alloy TZM (0.5Ti-0.08Zr-Mo) has also been successfully vacuum brazed at 1400 °C (2550 °F) with molybdenum powder added to the Ti-8.5Si brazing filler metal. Molybdenum has an extremely low coefficient of thermal expansion, which should be considered in the design of brazed joints, particularly when molybdenum is joined to other metals.

Tantalum. Nickel-base brazing filler metals (such as the nickel-chromium-silicon filler metals) have been used to braze tantalum. Tantalum forms a homogeneous solid solution with nickel at concentrations up to 36 wt% with the liquidus temperature reducing from 1450 to 1350 °C (2640 to 2460 °F). These brazing filler metals are satisfactory for service temperatures below 982 °C (1800 °F). Copper-gold alloys with less than 40 wt% Au can be used as brazing filler metals. When gold is added in amounts between 40 to 90 wt%, the alloys tend to form brittle age-hardening compounds. Because tantalum and its alloys are usually used for elevated-temperature applications (1650 °C, or 3000 °F, and above), only a few brazing filler metals have been developed. Most of the brazing filler metals currently available for tantalum are in powder form, which is difficult to work with at elevated temperatures. New powder brazing filler metals (for example, Hf-7Mo, Hf-40Ta, and Hf-19Ta-2.5Mo) are being developed for tantalum as foil-type brazing filler metals.

Tungsten can be brazed in much the same manner as molybdenum and its alloys, using many of the same brazing filler metals. Brazing can be accomplished in a vacuum or in a dry argon, helium, or hydrogen atmosphere. To some extent, the selection of the brazing atmosphere depends on the brazing filler metal used. For example, brazing filler metals that contain elements with high vapor pressures at the brazing temperature cannot be used effectively in a high vacuum.

Ceramic Materials

Ceramic materials can be joined by two different methods: by metallization of the ceramic surface or by using a ductile active filler metal. In the first method, a coating of a metal such as molybdenum and manganese is used as a transition between the ceramic and metal. It may be followed by a nickel coating to improve the wetting and spreading of the filler metal. The disadvantage of this method is that the heating, cycles used to sinter these coatings can crack the ceramic substrate. In the second method, an active metal in the filler reduces the ceramic to promote the bond between a metal and a ceramic. This filler metal should be ductile to relieve the residual stresses built up during cooling. A uniform heating of all components, such as in furnace joining, should be used to minimize the presence of cracking due to thermal stresses.

Graphite. Brazeability of graphite is strongly influenced by the presence of impurities, such as oxygen and moisture, and by pore size and pore distribution. Data from Table 5 indicate that graphite has a low coefficient of thermal expansion and low strength; these properties make joining graphite to other engineering materials a difficult task. A metal with a coefficient of thermal expansion close to that of graphite (for example, tungsten, molybdenum, tantalum, and zirconium), should be chosen as a transition piece. Graphite is also difficult to wet, so a coating or a filler metal with a strong tendency to form carbide must be used to form the joint. Sometimes, a ductile metal such as copper is used to decrease the residual stress developed in the joint.

At least two commercially available filler metals are used to join graphite. One has a nominal chemical composition of 68Ag-27Cu-4.5Ti (in wt%) with a solidus temperature of 830 °C (1525 °F) and a liquidus temperature of 849 °C (1560 °F). The other filler metal has a chemical composition of 70Ti- 15Cu- 15Ni (in wt%) with a solidus temperature of 910 °C (1670 °F) and a liquidus temperature of 960 °C (1760 °F).

Figure 20 shows a sandwich technique for brazing graphite to an austenitic stainless steel. The filler metal is 68Ag-27Cu-4.5Ti with alternate layers of copper and molybdenum to match as closely as possible the coefficients of thermal expansion of graphite and stainless steel. A special fixturing system was designed to create compression in the joint to compensate for the mismatch in the coefficients of thermal expansion.

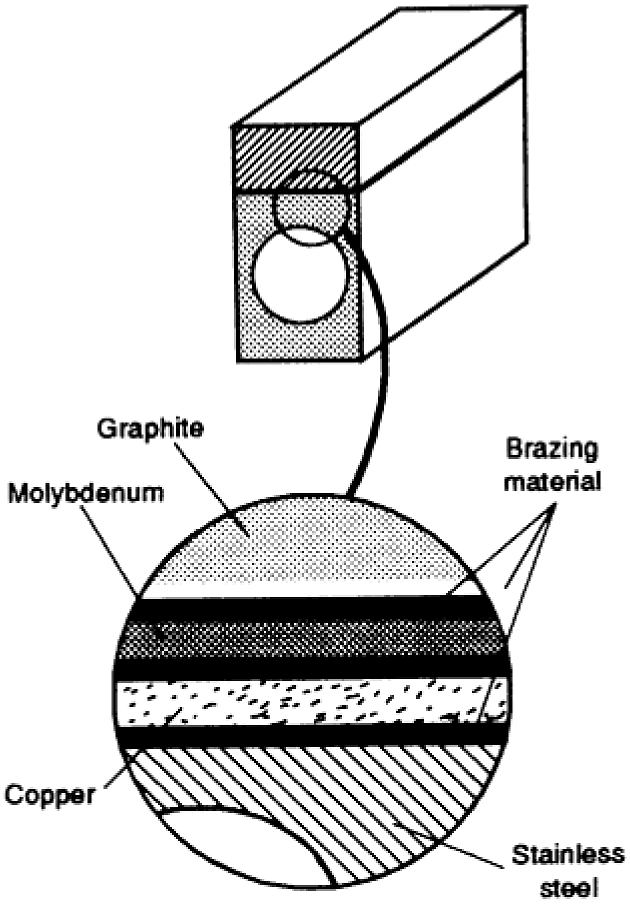

FIG. 20 APPLICATION OF SANDWICH TECHNIQUE TO BRAZE GRAPHITE TO AN AUSTENITIC STAINLESS STEEL WITH 68AG-27CU-4.5TI FILLER METAL. SOURCE: REF 10

Another example of graphite brazing also used transition inserts of gradient composition. In this case, graphite was joined to a nickel-base alloy, and the transition piece consisted of seven rings of tungsten-nickel-iron alloys. The first ring contained 97.5 wt% W to match the coefficient of thermal expansion of graphite, and the last ring contained 40 wt% W to match the coefficient of thermal expansion of the nickel-base alloy.

Carbides. The discussion of carbide brazing will be divided into two parts: brazing of carbide tools and brazing of silicon carbide (ceramic).

Carbide Tools. This group includes tungsten carbide with cobalt binder (3 to 25 wt%); tungsten carbide plus titanium and tantalum carbides, or niobium carbide with cobalt binder; titanium or tantalum carbides plus tungsten carbide with nickel or cobalt binder; chromium carbides with nickel or cobalt binders, and other carbide combinations.

Filler metals from the BAg group have been used to perform the carbide-to-steel joining. The filler metals that contain nickel (BAg-3, BAg-4, and BAg-22) promote a better wettability, because nickel improves the wettability in carbide-metal systems. RBCuZn-D and BCu have also been used, particularly where a post-braze heat treatment is required. Filler metals BAg-23 or 52.5Cu-38Mn-9.5Ni have been used when the brazements will be subjected to elevated temperature or additional wetting, as in the case of brazing titanium-base or chromium-base carbides.

In the case of tungsten-base carbide, the use of composite filler metal (consisting of a sandwich of two layers of a silver-base filler metal and a copper shim) has been reported. The ductile copper layer deforms during brazing to decrease the residual stress generated in the joint. A wider joint requires a thicker copper shim. Figure 21 shows this technique.

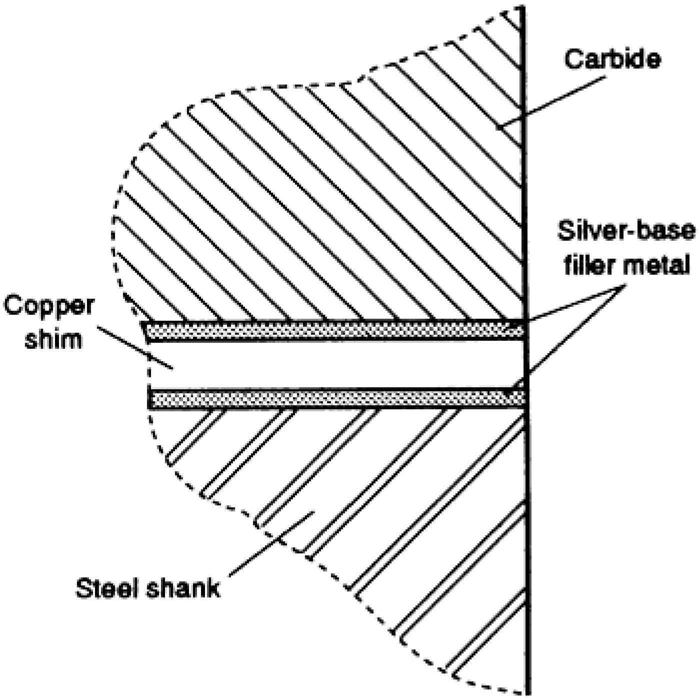

FIG. 21 USE OF COMPOSITE BRAZING FILLER METAL (TWO LAYERS OF SILVER-BASE FILLER METAL AND A COPPER SHIM) TO JOIN TUNGSTEN CARBIDE TO STEEL

Joint design is important in carbide tool brazing. The mismatch in the coefficients of thermal expansion must be taken into account and to avoid cracking during brazing, which decreases tool life. Figure 22 shows some optimum tool joint designs. In each set, the right-most drawing represents an improved design with longer expected tool life.

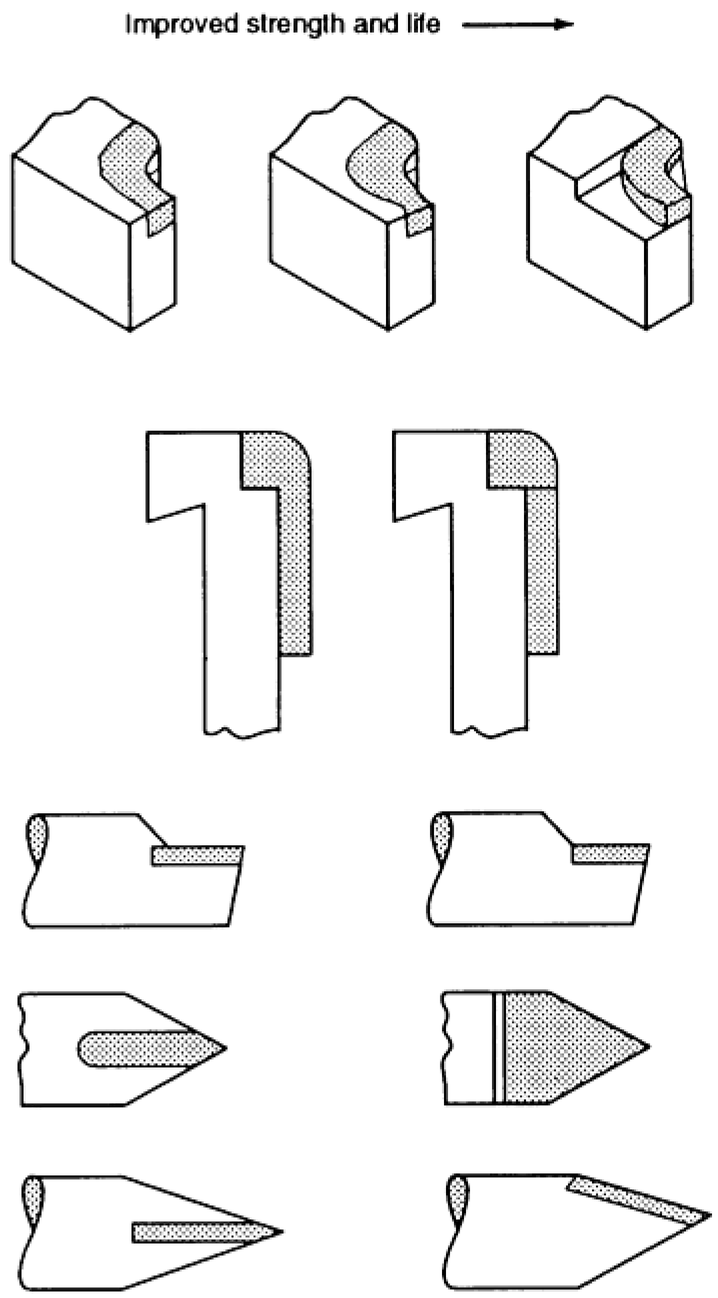

FIG. 22 JOINT DESIGNS TO OPTIMIZE STRENGTH AND TOOL LIFE OF BRAZED CARBIDE TOOL ASSEMBLIES. IN EACH SET, THE RIGHT-MOST DRAWING REPRESENTS AN IMPROVED DESIGN.

Silicon carbide can usually be brazed using a titanium-base filler metal according to the following reaction:

$$SiC + Ti_{fm} \rightarrow TiC_i + \underline{Si}_{fm} \qquad \textbf{(EQ 8)}$$

The subscripts *fm* and *i* denote the filler metal and interface, respectively. Commercial silver-titanium, silver-copper-titanium, and silver-copper-indium-titanium alloys can be used as filler metals to join silicon carbide to silicon carbide or to other alloys. The amount of copper in the filler metal must be controlled, because copper dissolves silicon from silicon carbide to form a copper-silicon alloy and graphite and changes the characteristics of the ceramic material close to the joint.

Some experimental results show that a Ni-13.4Cr-40Si (in at.%) filler metal can also be used to braze silicon carbide to silicon carbide components with good results. On the other hand, an experimental filler metal Fe-34Si (in at.%) used to join silicon carbide to Inconel 600 showed extensive chemical reaction with formation of low-melting-point silicides.

To increase the work temperature range of brazed silicon carbide components, a filler metal with higher solidus temperature is needed. Ni-50Ti (in at.%) has been reported to perform well. Table 14 presents some filler metals commonly used to braze silicon carbide.

TABLE 14 FILLER METALS COMMONLY USED IN SILICON CARBIDE BRAZING APPLICATIONS

BRAZING FILLER METAL SYSTEM	COMPOSITION, WT%					SOLIDUS TEMPERATURE		LIQUIDUS TEMPERATURE		BRAZING TEMPERATURE	
	Cu	In	Ti	Ag	Ni	°C	°F	°C	°F	°C	°F
SILVER-COPPER-TITANIUM	27.5	. . .	2	70.7	. . .	780	1435	795	1465	840	1545
	26.5	. . .	3	70.5	. . .	803	1475	857	1575	950	1740
	6	. . .	3	91	. . .	875	1605	917	1685	970	1780
SILVER-COPPER-INDIUM-TITANIUM	23.5	14.5	1.25	60.75	. . .	605	1120	715	1320	760	1400
	19.5	5	3	72.5	. . .	732	1350	811	1490	950	1740
NICKEL-TITANIUM	. . .	. . .	55.1	. . .	44.9	985	1805	130	2390	1550	2820

Oxide ceramics such as Al_2O_3, ZrO_2, and MgO are not easily wetted. Consequently, an active element must be present in the filler metal to reduce the oxide. Titanium is the active element most frequently added in the filler metals, although zirconium, niobium, aluminum, chromium, and vanadium have also been used. The oxide reduction by titanium is represented by:

$$M_xO_y + y\underline{Ti}_{fm} \rightarrow yTiO_i + x\underline{M}_{fm} \qquad \textbf{(EQ 9)}$$

Depending on the brazing system, a dissolution reaction may occur instead of the reduction of the oxide ceramics. In the case of the Al_2O_3-Cu-Cu_2O system, a spinel $CuAlO_2$ is reported to form at the interface. This spinel makes the metal/ceramic transition to promote a good bond between both types of material.

The nature of the oxide formed at the ceramic oxide-filler metal interface is very important to the wetting and spreading of the molten filler metal and to the integrity of the joint. The metallic-type conductivity and metallic-ionic bonding of titanium gives it a metallic characteristic. On the other hand, Ti_2O_3 has only ionic bonding, giving it a typical ceramic characteristic.

To protect the titanium in the filler metal against "premature" oxidation or other reactions during the joining, a high vacuum or an inert atmosphere (that is, non-nitrogen) is recommended. Table 15 shows active brazing conditions for some oxide ceramics.

TABLE 15 TYPICAL ACTIVE BRAZING PARAMETERS FOR OXIDE CERAMICS

JOINT	FILLER METAL	BRAZING TEMPERATURE		STRENGTH	
		°C	°F	MPA	KSI
AL_2O_3/AL_2O_3	CU-44AG-4SN-4TI	800-900	1470-1650	80-120	12-17
AL_2O_3/AL	AL INTERLAYERS (40 µM, OR 0.0016 IN., THICK)	1000	1830	73-89	11-13
AL_2O_3/FE-29NI-18CO	AG-27CU-3TI	900	1650	185	27
	AG-25CU-15IN-1TI	900	1650	162	23

AL_2O_3/TI-6AL-4V	CU-40AG-5TI	850-890	1560-1635	. . .	. . .
PSZ ZRO_2/PSZ ZRO_2	CU-44AG-4SN-4TI	800-850	1470-1560	>400	>58
	CU-44AG-4SN-4TIK	900	1650	248	36
TZP ZRO_2/STEEL(B)	AG-4TI	1050	1920	151	22

(A) (A) PSZ, PARTIALLY STABILIZED ZIRCONIA.

(B) TZP, TETRAGONAL ZIRCONIA POLYCRYSTALS

Nitride Ceramics. Nonoxide ceramics are more easily reduced by chemical reactions than oxide ceramics. Thus, the brazing techniques developed for oxide ceramics should be used judiciously in joining nonoxide ceramics.

Joining of nitrides using a glass bonding agent or by active metal brazing has been reported. A glass bonding agent such as $70SiO_2$-$27MgO$-$3Al_2O_3$ (in wt%) is used to join Si_3N_4. Si_2N_2O is formed at the ceramic-glass interface to promote the joining.

Active brazing of nitrides follows the same principles as outlined in the section "Oxide Ceramics" in this article. An active element reduces the ceramic at the interface between the ceramics and the molten filler metal. The general equation for a nitride reduced by a titanium-rich filler metal is:

$$M_xN_y + y\underline{Ti}_{fm} \rightarrow yTiN_i + x\underline{M}_{fm} \qquad \textbf{(EQ 10)}$$

TiN is formed at the filler metal-nitride interface, and it has a character similar to TiO in promoting wetting. Depending on which active metal is contained in the filler metal, a silicide or an aluminide may form first at the interface. In the case of copper-niobium filler metal for silicon nitride joining, niobium silicide may form before the niobium nitride. Table 16 shows some typical brazing conditions for the nitride ceramics.

TABLE 16 TYPICAL PARAMETERS FOR BRAZING OF NITRIDE CERAMICS

JOINT	FILLER METAL	BRAZING TEMPERATURE		STRENGTH	
		°C	°F	MPA	KSI
BN/BN	AL	1000	1830	6(A)	0.9(A)
AIN/MO	CU-42TI	1100	2010	167(A)	24.2(A)
	CU-71TI	1100	2010	49(A)	7.1(A)
	AG-35CU-4TI	980	1795	192(A)	27.8(A)
SI_3N_4/SI_3N_4	AL	1000	1830	450-500(B)	65-72.5(B)
	AL-SI (0.06-10.6)	800	1470	>400(B)	>58(B)
	AG-27CU-2TI	850	1560	270(A)	39(A)
	AG-CU-IN-1.25TI	700	1290	219-427(B)	31.8-61.9(B)
	$70SIO_2$-27MGO-$3AL_2O_3$	1500	2730	450(B)	65(B)
SI_3N_4/MO	AG-35CU-4TI	980	1795	195(A)	28.3(A)
SI_3N_4/FE-29NI-18CO	AG-28CU-2TI	800	1470	95(B)	14(B)
	AG-25CU-15IN-1TI	900	1650	51(B)	7.4(B)
SIALON/SIALON	AL	1000	1830	61(A)	8.8(A)

(A) SHEAR TEST.

(B) FOUR-POINT BEND TEST

The presence of these compounds, as a reaction layer, is necessary to promote wetting and spreading and to make a good bond between ceramic and metal. The growth of the reaction layer is affected by brazing temperature, time, reactivity, and the amount of active element in the filler metal. Thus, the joining conditions should be carefully controlled to produce a thin layer and promote the optimum mechanical behavior of the joint.

References cited in this section

4. D.R. MILNER AND R.L. APPS, *INTRODUCTION TO WELDING AND BRAZING,* PERGAMON PRESS, 1969
9. E.W. COLLINGS, *THE PHYSICAL METALLURGY OF TITANIUM ALLOYS,* AMERICAN SOCIETY FOR METALS, 1984
10. M.M. SCHWARTZ, *CERAMIC JOINING,* ASM INTERNATIONAL, 1990

S.D. Brandi, Escola Politécnica da Universidade de São Paulo, Brazil; S. Liu, Colorado School of Mines; J.E. Indacochea and R. Xu, University of Illinois at Chicago

---

References

1. P.R. SHARPS, A.P. TOMSIA, AND J.A. PASK, WETTING AND SPREADING IN THE CU-AG SYSTEM, *ACTA METALL.,* VOL 29 (NO. 5), 1981, P 855-865
2. V. KONDIC, *METALLURGICAL PRINCIPLES OF FOUNDING,* AMERICAN ELSEVIER, 1973
3. M.G. NICHOLAS AND D.A. MORTIMER, CERAMIC/METAL JOINING FOR STRUCTURAL APPLICATIONS, *MATER. SCI. TECHNOL.,* VOL 1 (NO. 9), 1985, P 657-665
4. D.R. MILNER AND R.L. APPS, *INTRODUCTION TO WELDING AND BRAZING,* PERGAMON PRESS, 1969
5. *SOLDERING MANUAL,* AMERICAN WELDING SOCIETY, 1977
6. J.A. PASK, FROM TECHNOLOGY TO THE SCIENCE OF GLASS/METAL AND CERAMIC/METAL SEALING, *AM. SOC. CERAM. BULL.,* VOL 66 (NO. 11), 1987, P 1587-1592
7. H.H. MANKO, *SOLDERS AND SOLDERING,* 3RD ED., MCGRAW-HILL, 1992
8. *ALUMINUM SOLDERING HANDBOOK,* ALUMINUM ASSOCIATION, 1974
9. E.W. COLLINGS, *THE PHYSICAL METALLURGY OF TITANIUM ALLOYS,* AMERICAN SOCIETY FOR METALS, 1984
10. M.M. SCHWARTZ, *CERAMIC JOINING,* ASM INTERNATIONAL, 1990

S.D. Brandi, Escola Politécnica da Universidade de São Paulo, Brazil; S. Liu, Colorado School of Mines; J.E. Indacochea and R. Xu, University of Illinois at Chicago

---

Selected References

- *BRAZING MANUAL,* AMERICAN WELDING SOCIETY, 1992
- J.A. DEVORE, SOLDERING AND MOUNTING TECHNOLOGY, *ELECTRONIC MATERIALS HANDBOOK,* VOL 1, *PACKAGING,* ASM INTERNATIONAL, 1987
- G. ELSSNER AND G. PETZOW, METAL/CERAMIC JOINING, *ISIJ LNT.,* VOL 30 (NO. 12), 1990, P 1011-1032

- H.H. MANKO, *SOLDERING HANDBOOK FOR MOUNTED CIRCUITS AND SURFACE MOUNTING,* VAN NOSTRAND REINHOLD, 1986
- Y. NAKAO, RESEARCH ON JOINING OF CERAMICS IN JAPAN, *1ST U.S.-JAPAN SYMP. ADVANCES IN WELDING METALLURGY,* AWS/JWS, JWES, 1990, P 79-101
- M.G. NICHOLAS, REACTIVE METAL BRAZING, *JOINING CERAMICS, GLASS AND METAL,* W. KRAFT, ED., DGM INFORMATIONSGESELLSCHAFT-VERLAG, 1989
- R.L. PEASLEE, BRAZING: YESTERDAY'S ART HAS BECOME TODAY'S SCIENCE, *WELD. J.,* VOL 71 (NO. 10), NOV 1992, P 25-31
- S.L. RICHLEN AND W.P. PARKS, JR., HEAT EXCHANGERS, *ENGINEERED MATERIALS HANDBOOK,* VOL 4, *CERAMICS AND GLASSES,* ASM INTERNATIONAL, 1991
- M.M. SCHWARTZ, *BRAZING,* ASM INTERNATIONAL, 1987
- R.J. KLEIN WASSINK, *SOLDERING IN ELECTRONICS,* ELECTROCHEMICAL PUBLICATIONS, 1989
- WELDING, *BRAZING, AND SOLDERING, METALS HANDBOOK,* 9TH ED., VOL 6, AMERICAN SOCIETY FOR METALS, 1982
- *WELDING PROCESSES, WELDING HANDBOOK,* 8TH ED., VOL 2, AMERICAN WELDING SOCIETY, 1991
- W. WLOSINSKI, INTERFACES IN DISSIMILAR MATERIAL JOINTS, *2ND INTERNATIONAL KOLLOQUIM BAD NAUHEIM,* MARCH 1985, P 27-29

Arc Welding of Carbon Steels

Ronald B. Smith, The ESAB Group, Inc.

Introduction

CARBON STEELS are defined as those steels containing up to 2% C, 1.65% Mn, 0.60% Si, and 0.60% Cu, with no deliberate addition of other elements to obtain a desired alloying effect. Tables 1(a) and 1(b) list some of the common grades of carbon steel that are covered in this article. The weldability of these steels greatly depends on their carbon and manganese contents and impurity levels.

TABLE 1(A) COMPOSITION OF SELECTED CARBON STEELS USED IN ARC WELDING APPLICATIONS

SAE-AISI NO.	COMPOSITION, WT %(A)				
	C	Mn	P	S	OTHER
1006	0.08	0.45	0.040	0.050	. . .
1010	0.08-0.13	0.30-0.60	0.040	0.050	. . .
1020	0.17-0.23	0.30-0.60	0.040	0.050	. . .
1030	0.27-0.34	0.60-0.90	0.040	0.050	. . .
1040	0.36-0.44	0.60-0.90	0.040	0.050	. . .
1050	0.47-0.55	0.60-0.90	0.040	0.050	. . .
1060	0.55-0.66	0.60-0.90	0.040	0.050	. . .
1070	0.65-0.73	0.60-0.90	0.040	0.050	. . .
1080	0.74-0.88	0.60-0.90	0.040	0.050	. . .
1095	0.90-1.04	0.30-0.50	0.040	0.050	. . .
MANGANESE-CARBON					
1513	0.10-0.16	1.10-1.40	0.040	0.050	. . .
1527	0.22-0.29	1.20-1.50	0.040	0.050	. . .
1541	0.36-0.44	1.35-1.65	0.040	0.050	. . .
1566	0.60-0.71	0.85-1.15	0.040	0.050	. . .
FREE-MACHINING					
1108	0.08-0.13	0.50-0.80	0.040	0.08-0.13	. . .
1139	0.35-0.43	1.35-1.65	0.040	0.13-0.20	. . .
1151	0.48-0.55	0.70-1.00	0.040	0.08-0.13	. . .
1212	0.13	0.70-1.00	0.07-0.12	0.16-0.23	. . .
12L14	0.15	0.85-1.15	0.04-0.09	0.26-0.35	0.15-0.35 PB

(A) SINGLE VALUES ARE MAXIMUMS.

TABLE 1(B) COMPOSITION AND MECHANICAL PROPERTIES OF SELECTED CARBON STEELS USED IN ARC WELDING APPLICATIONS

ASTM DESIGNATION		COMPOSITION, WT%(A)			ULTIMATE TENSILE STRENGTH(B)		YIELD STRENGTH(B)		COMMENTS(C)
SPECIFICATION	GRADE OR TYPE	c	mn	si(d)	mpa	ksi	mpa	ksi	
STRUCTURAL STEELS									
A 36	. . .	0.29	0.80-1.20	0.15-0.40	400-550	58-80	220-250	32-36	. . .

A 131	B	0.21	0.80-1.10	0.35	400-500	58-71	220	32	. . .
	E	0.18	0.70-1.35	0.10-0.35	400-500	58-71	220	32	NORMALIZED
	CS	0.16	1.00-1.35	0.10-0.35	400-500	58-71	220	32	NORMALIZED
A 283	A	0.14	0.90	0.04	310-415	45-60	165	24	. . .
	B	0.17	0.90	0.04	345-450	50-65	185	27	. . .
	C	0.24	0.90	0.04	380-485	55-70	205	30	. . .
	D	0.27	0.90	0.04	415-515	60-75	230	33	. . .
A 284	C	0.36	0.90	0.15-0.40	415	60	205	30	. . .
	D	0.35	0.90	0.15-0.40	415	60	230	33	. . .
A 573	58	0.23	0.60-0.90	0.10-0.35	400-500	58-71	220	32	. . .
	65	0.26	0.85-1.20	0.15-0.40	450-530	65-77	240	35	. . .
	70	0.28	0.85-1.20	0.15-0.40	485-620	70-90	290	42	. . .
A 285	A	0.17	0.90	0.35	310-450	45-65	165	24	. . .
	B	0.22	0.90	0.35	345-485	50-70	185	27	. . .
	C	0.28	0.90	0.35	380-515	55-75	205	30	. . .
A 442	55	0.24	0.80-1.10	0.15-0.40	380-515	55-75	205	30	. . .
	60	0.27	0.80-1.10	0.15-0.40	415-550	60-80	220	32	. . .
A 515	55	0.28	0.90	0.15-0.40	380-515	55-75	205	30	NORMALIZED IF T > 38 MM (1.5 IN.)
	60	0.31	0.90	0.15-0.40	415-550	60-80	220	32	NORMALIZED IF T > 38 MM (1.5 IN.)
	65	0.33	0.90	0.15-0.40	450-585	65-85	240	35	NORMALIZED IF T > 38 MM (1.5 IN.)
	70	0.35	1.20	0.15-0.40	485-620	70-90	260	38	NORMALIZED IF T > 38 MM (1.5 IN.)
A 516	55	0.26	0.60-1.20	0.15-0.40	380-515	55-75	205	30	NORMALIZED IF T > 38 MM (1.5 IN.)
	60	0.27	0.60-1.20	0.15-0.40	415-550	60-80	220	32	NORMALIZED IF T > 38 MM (1.5 IN.)
	65	0.29	0.85-1.20	0.15-0.40	450-585	65-85	240	35	NORMALIZED IF T > 38 MM (1.5 IN.)
	70	0.31	0.85-1.20	0.15-0.40	485-620	70-90	260	38	NORMALIZED IF T > 38 MM (1.5 IN.)
A 537	C1.1	0.24	0.70-1.60	0.15-0.50	450-585	65-85	310	45	NORMALIZED
	C1.2	0.24	0.70-	0.15-	515-	75-	380	55	QUENCHED AND

			1.60	0.50	655	95			TEMPERED
A 662	A	0.14	0.90-1.50	0.15-0.40	400-540	58-78	275	40	NORMALIZED
	B	0.19	0.85-1.60	0.15-0.40	450-585	65-85	275	40	NORMALIZED IF *T* > 38MM (1.5 IN.)
	C	0.20	1.00-1.60	0.15-0.50	485-620	70-90	295	43	NORMALIZED IF *T* > 38MM (1.5 IN.)
PIPE STEELS									
A 53	A	0.25	0.95	. . .	330	48	205	30	. . .
	B	0.30	1.20	. . .	415	60	240	35	. . .
A 106	A	0.25	0.27-0.93	0.10	330	48	205	30	. . .
	B	0.30	0.29-1.06	0.10	415	60	240	35	. . .
	C	0.35	0.29-1.06	0.10	485	70	275	40	. . .
A 381	Y42	0.26	1.40	. . .	415	60	290	42	. . .
	Y52	0.26	1.40	. . .	495	72	360	52	. . .
	Y60	0.26	1.40	. . .	540	78	415	60	. . .
CAST STEELS									
A 27	60-30	0.30	0.60	0.80	415	60	205	30	HEAT-TREATED
	70-40	0.25	1.20	0.80	485	70	275	40	HEAT-TREATED
A 216	WCA	0.25	0.70	0.60	415-585	60-85	205	30	HEAT-TREATED
	WCB	0.30	1.00	0.60	485-655	70-95	250	36	HEAT-TREATED
	WCC	0.25	1.20	0.50	485-655	70-95	275	40	HEAT-TREATED
SAE J435C	0025	0.25	0.75	0.80	415	60	205	30	HEAT-TREATED
	0050A	0.40-0.50	0.50-0.90	0.80	585	85	310	45	HEAT-TREATED

(A) SINGLE VALUES ARE MAXIMUMS.
(B) SINGLE VALUES ARE MINIMUMS.
(C) *T*, PLATE THICKNESS.
(D) SILICON CONTENT VARIES WITH DEOXIDATION PRACTICE USED.

At low carbon levels (less than 0.15% C), the steels are nonhardening and weldability is excellent. The bulk of the steels in this carbon range are used for flat-rolled products (sheet and strip), which may contain up to 0.5% Mn. Most of these steels are now aluminum-killed, continuous-cast product supplied in the cold-rolled and annealed condition. The lower available oxygen in the killed sheet makes it easier to arc weld without porosity formation. Typical uses for these steels are automobile body panels, appliances, and light-walled tanks.

In the range of 0.15 to 0.30% C, the steels are generally easily welded, but because hardening is a possibility, precautions such as preheating may be required at higher manganese levels, in thicker sections, or at high levels of joint restraint. Much of the steel in this carbon range is used for rolled structural plate and tubular products. These steels are generally killed or semi-killed and are usually supplied in the hot-rolled condition. The presence of surface scale (iron oxide) from the high-temperature rolling process increases the likelihood of porosity formation during welding and may require the use of welding electrodes with higher levels of deoxidizers, or removal of the scale prior to welding. Some carbon steel plate may be heat-treated (normalized or quenched and tempered) to enhance properties. In such cases, careful selection of the welding consumable must be made in order to match the base metal properties. Welding heat input may have to be controlled so as not to diminish the properties in the heat-affected zone (HAZ) of the base metal, or postweld heat treatment may be necessary to restore the strength and/or toughness of the HAZ.

Medium-carbon steel--steel that contains 0.30 to 0.60% C--can be successfully welded by all of the arc welding processes, provided suitable precautions are taken. The higher carbon content of these steels, along with manganese from 0.6 to 1.65%, makes these steels more hardenable. For this reason, they are commonly used in the quenched and tempered condition for such applications as shafts, couplings, gears, axles, crankshafts, and rails. Because of the greater likelihood of martensite formation during welding, and the higher hardness of the martensite formed, preheating and postheating treatments are necessary. Low-hydrogen consumables and procedures should also be used to reduce the likelihood of hydrogen-induced cracking. The higher strength level of these steels may require the use of an alloyed electrode to match the base metal properties. It may also be necessary to postweld heat treat the part in order to restore the strength and/or toughness of the HAZ.

Higher-carbon steel--steel that contains 0.60 to 2.00% C--has poor weldability because of the likelihood of formation of a hard, brittle martensite upon weld cooling. Steels of this type are used for springs, cutting tools, and abrasion-resistant applications. Low-hydrogen consumables and procedures, preheating, interpass control, and stress relieving are essential if cracking is to be avoided. Austenitic stainless steel electrodes are sometimes used to weld high-carbon steels. These electrodes will reduce the risk of hydrogen-induced cracking but may not match the strength of the high-carbon steel base metal.

The 11*xx* and 12*xx* series steels contain large amounts of sulfur, phosphorus, or lead for improved machinability. Both series are difficult to weld because of solidification cracking and porosity formation.

Carbon steel castings can be made in the same wide range of compositions as the wrought types. The major differences are that castings are always fully killed (0.3 to 1.0% Si) and are generally heat-treated. For a given composition, the weldabilities of the cast and wrought carbon steels are similar. Casting defects are often repaired by welding. Care in the selection of a welding electrode is required if the casting is to be heat-treated. Low-alloy steel electrodes are frequently used because they match the casting properties better after heat treatment. Low-hydrogen electrodes and procedures are commonly used, depending on casting size and carbon content.

Arc Welding of Carbon Steels

Ronald B. Smith, The ESAB Group, Inc.

Weldability Considerations for Carbon Steels

As a group, the carbon steels are among the most weldable of materials, yet they are susceptible, to one degree or another, to:

- HYDROGEN-INDUCED CRACKING
- SOLIDIFICATION CRACKING
- LAMELLAR TEARING
- WELD METAL POROSITY
- WELD METAL AND HAZ MECHANICAL PROPERTY VARIATIONS

Additional information about the weldability of carbon steel is available in the Section "Selection of Carbon and Low-Alloy Steels" in this Volume.

Hydrogen-Induced Cracking (Ref 1, 2)

Hydrogen-induced cracking (also referred to as cold cracking, delayed cracking, or underbead cracking) is the most serious problem affecting weldability. Any hardenable carbon steel is susceptible. This type of cracking results from the combined effects of four factors:

- A SUSCEPTIBLE ("BRITTLE") MICROSTRUCTURE
- THE PRESENCE OF HYDROGEN IN THE WELD METAL

- TENSILE STRESSES IN THE WELD AREA
- A SPECIFIC TEMPERATURE RANGE, -100 TO 200 °C (-150 TO 390 °F)

Hydrogen-induced cracking occurs after weld cooling (hence the term cold cracking) and is often delayed for many hours while atomic hydrogen diffuses to areas of high tensile stress. At microstructural flaws in a tensile stress field, the hydrogen changes to its molecular form, causing cracking. Cracking may occur in the HAZ or weld metal, and it may be longitudinal or transverse (Fig. 1). For carbon steels, cracking is more likely to occur in the HAZ because carbon steel electrodes are usually low in carbon and the weld metal is generally not hardenable. Exceptions would be if a highly alloyed electrode were being used, if the weld metal were made more hardenable by dilution of carbon from the base material, or in certain submerged arc welds where the use of excessive arc voltage and active fluxes results in high manganese and/or silicon pickup from the flux (see the section "Submerged Arc Welding" in this article).

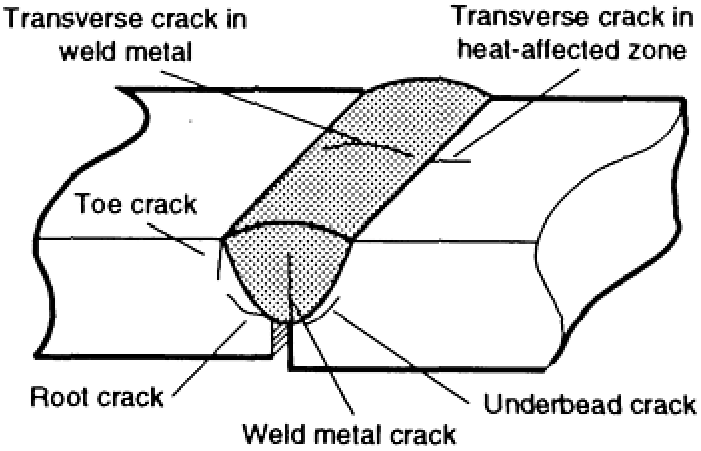
Transverse crack in weld metal
Transverse crack in heat-affected zone
Toe crack
Root crack
Weld metal crack
Underbead crack

FIG. 1 SCHEMATIC SHOWING LOCATION OF HYDROGEN-INDUCED CRACKS IN CARBON STEEL WELDMENTS. SOURCE: REF 1

Cracks in the HAZ are most often longitudinal. Underbead cracks lie more or less parallel to the fusion line. They do not normally extend to the surface and may therefore be difficult to detect. Underbead cracks will form at relatively low stress levels in martensite when high levels of hydrogen are present.

Toe cracks and root cracks start in areas of high stress concentration. Cracking may therefore occur in less susceptible microstructures or at relatively low hydrogen levels. This type of cracking is often delayed while the necessary hydrogen diffuses to the area.

Transverse cracking in the HAZ is less common. It will occur in high-carbon martensite under conditions of high longitudinal stresses (for example, outside fillet welds on heavy plate).

Weld metal cracks may be longitudinal or transverse. Longitudinal cracks start due to stress concentrations at the root of the weld. Transverse cracking starts at hydrogen-containing defects subject to longitudinal stresses. Weld metal cracks do not always extend to the surface. In submerged arc weld metal made with damp fluxes, a unique crack morphology known as chevron cracking can occur. Here the cracks lie at 45° to the weld axis.

One of the serious problems with hydrogen-induced cracking is the difficulty in detecting the presence of a crack. The delayed nature of some of the cracks demands that inspection not be carried out too soon, especially in welds that will have external stresses applied when put in service. Because some of the cracks do not extend to the surface, they are not detectable by visual inspection methods (for example, liquid penetrant, or magnetic particle inspection, which requires the defect to be near the surface). Radiography is most sensitive to volumetric flaws, and it may not detect cracks that are too fine or of the wrong orientation. Ultrasonic inspection is capable of detecting the crack if the operator knows where to look. Given the difficulty in detecting hydrogen-induced cracks and the possibility of this cracking leading to in-service failure, it is prudent to observe the precautions necessary to avoid cracking in the first place.

Susceptible Microstructures. A susceptible microstructure is any hard, brittle transformation product formed in the HAZ or weld metal. Martensite is by far the most susceptible microstructure found in carbon steels. In all but high-carbon steel weld zones, the martensite, when formed, is usually mixed with other, less susceptible transformation products (for example, pearlite). Nevertheless, these mixed microstructures are susceptible to cracking, depending on the level of hardening. Martensite itself may be more or less susceptible, depending on its carbon content. Higher-carbon martensite is harder, more brittle, and more susceptible to cracking. Figure 2 shows the effect of carbon content on the hardness of carbon steel for various percentages of martensite formed upon rapid cooling.

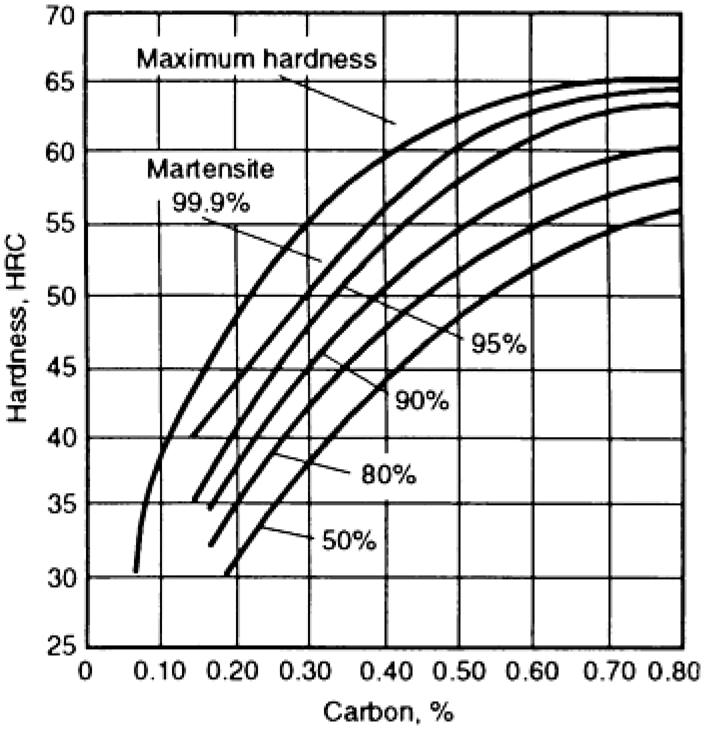

FIG. 2 PLOT OF HARDNESS VERSUS CARBON CONTENT AS A FUNCTION OF MARTENSITE FORMATION IN CARBON STEEL THAT HAS COOLED RAPIDLY

The hardenability of the steel also increases as the carbon content increases. Figure 3 presents hardenability curves for five carbon steels (as determined by end-quench tests). As the cooling rate (top scale) decreases, the lower-carbon steels show a more rapid decline in hardness (that is, less martensite is being formed at the slower cooling rates). Higher carbon steels are more likely to form martensite upon cooling, and the martensite formed will be harder and thus more susceptible to hydrogen-induced cracking.

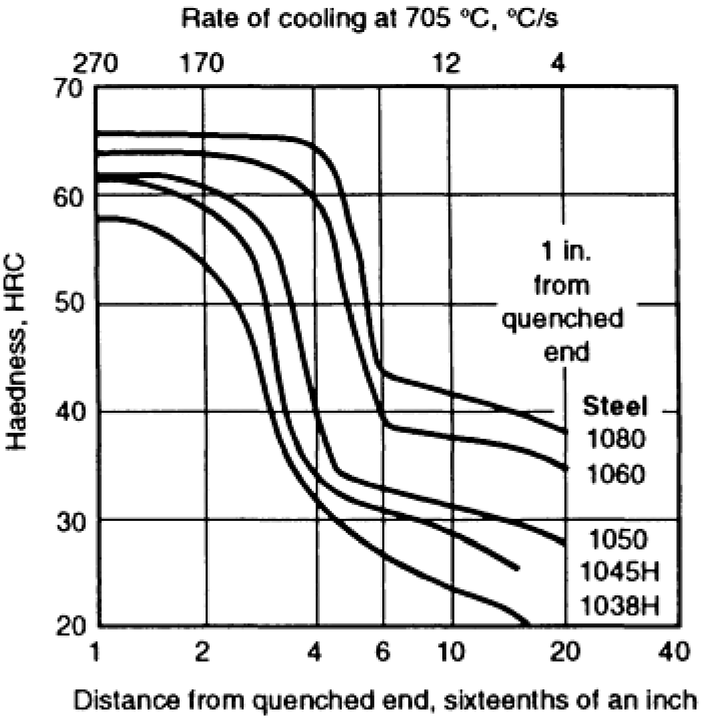

FIG. 3 HARDENABILITY CURVES FOR FIVE CARBON STEELS AS DETERMINED BY END-QUENCH TESTING

Martensite formation is also a function of grain size. The HAZ adjacent to the fusion line will have large austenite grains, which more readily form martensite than smaller grains.

Presence of Hydrogen in Weld Metal. The exact role that hydrogen plays in cold cracking is unknown, but its presence is necessary for this type of cracking to occur. Hydrogen is generally introduced into the weld area during welding. The sources are:

- MOISTURE IN THE FLUX COATING OF SHIELDED METAL ARC WELDING ELECTRODES, IN SUBMERGED ARC FLUXES, OR IN THE CORE OF FLUX-CORED ELECTRODES
- HYDROGEN-CONTAINING LUBRICANTS LEFT ON THE SURFACE OF WIRE ELECTRODES OR IN THE SEAMS OF CORED ELECTRODES
- HYDROGEN-CONTAINING COMPOUNDS OR RESIDUES LEFT ON THE PLATE SURFACE (THESE CAN INCLUDE GREASE, OIL, PAINT, RUST, AND SO ON)
- LEAKING WATER-COOLED TORCHES, BROKEN GAS LINES, OR HIGH MOISTURE IN THE SHIELDING GAS
- ASPIRATION OF MOISTURE-LADEN AIR INTO THE WELD AREA

The principal source of hydrogen is the welding consumable.

Hydrogen, ionized to its atomic form by the intense heat of the arc, is very soluble in liquid iron. This solubility decreases with decreasing temperature. Rapid cooling (in a weld, for example) can result in significant supersaturation of atomic hydrogen in the solidified metal. Because atomic hydrogen diffuses relatively easily through ferrite, a certain amount of the initial hydrogen present when the weld has cooled to ambient temperatures will subsequently diffuse out of the weld metal. This diffusible hydrogen, whose exit from the weld metal is accelerated by higher temperatures, may be measured using standardized test methods (such as those outlined in AWS A4.3). The results, measured in mL of hydrogen per 100 g of weld metal, give an indication of the amount of hydrogen potentially available to induce cracking. These tests are now widely used to measure diffusible hydrogen and to classify arc welding electrodes.

The distribution of hydrogen in the weld metal and HAZ and the subsequent susceptibility to cold cracking is a function of the phase transformations that take place upon cooling. Three examples are described below.

HAZ Hardens but Weld Metals Do Not. If the HAZ hardens (transforms to martensite) but the weld metal does not (for example, welding a medium-carbon steel with a low-carbon, mild-steel electrode), the weld metal undergoes transformation to a ferritic structure at relatively high temperatures. Because hydrogen diffuses easily in ferrite, it will diffuse to and migrate across the fusion boundary into the still-austenitic HAZ, which has a higher solubility for hydrogen but a lower diffusivity. At a much lower temperature, the HAZ transforms to martensite. The area near the fusion boundary is highly concentrated in hydrogen. Cracking may then occur.

Weld Metal More Hardenable than Base Material. If the weld metal is more hardenable than the base material (for example, highly alloyed electrode picks up substantial carbon from a carbon steel base plate), the HAZ transforms first. The weld metal, which is austenitic, will retain the hydrogen. When the weld metal transforms, it will be in the presence of high hydrogen, which may lead to cracking in the weld metal. The HAZ, which is low in hydrogen, will not crack.

No Lower Temperature Transformation in Austenitic Weld Metal. If an austenitic filler metal is used to weld a hardenable base metal (for example, high-carbon steel), there is no lower temperature transformation in the weld metal, which remains austenitic with a high tolerance for hydrogen. Little hydrogen migrates across the fusion boundary, and so the HAZ transforms to martensite with a low hydrogen content and cracking is avoided.

Once hydrogen is present in a hardenable microstructure, the exact mechanism by which it engenders cracking is unknown. The hydrogen is thought to diffuse to defects and areas of triaxial stress, where it interacts with tensile stress to enlarge and extend discontinuities.

Tensile Stresses in Weld Area. The stresses that a weldment is subjected to may be either externally or internally generated. External stresses are those applied when the weld is put in service. Internal, or residual, stresses are those that arise from the welding process, due mainly to thermal gradients, unequal thermal expansion and contraction of the base metal and weld metal, and volume changes resulting from phase transformations during cooling. The residual stresses can be reduced by unrestrained movement of the parts being welded, thermal treatments, and proper weld design.

Movement of the parts can be actual physical displacement or an elastic or plastic deformation. An example of the former would be offsetting the parts prior to welding and allowing them to rotate into the desired final position. An example of the latter is the use of a lower strength ferritic filler metal or of an austenitic filler metal to weld a medium- or high-carbon steel. The weaker, more ductile weld metal of these deposits is able to absorb some of the stress (by elastic or plastic deformation) and reduce the level in the more vulnerable HAZ.

Any condition that limits the movement of welded parts will increase the residual stresses and the likelihood of hydrogen-induced cracking. Thicker, more massive plates, high clamping pressure, or rigid surrounding structures reduce the freedom of motion of the welded parts and result in high residual stresses. Thicker, more massive weld structures also cool faster because of the large heat sink, thus increasing the likelihood of martensite formation. The use of preheat to reduce the risk of martensite formation will also reduce and delay the onset of maximum residual stresses (and allow more time for hydrogen to escape). Postweld stress relief reduces residual stresses by heating the weldment to a temperature at which the yield point drops low enough for plastic flow to occur, thus relaxing the stresses.

Joint designs that reduce the volume of weld metal needed will result in lower residual stresses. The smallest groove angles and root openings that allow good access should be used. The smallest fillet weld size that meets shear strength requirements should be used. The presence of notches or gaps in the weld joint will also increase the incidence of cracking because they produce a localized concentration of stress. Accumulation of relatively low levels of hydrogen in these areas may be sufficient to cause cracking.

Temperature Range for Hydrogen-Induced Cracking. Hydrogen-induced cracking will not normally occur outside of the temperature range of -100 to 200 °C (-150 to 390 °F). Below -100 °C (-150 °F), the hydrogen probably diffuses too slowly to reach a critical concentration in a susceptible area, while above 200 °C (390 °F) the hydrogen diffuses so rapidly out of the weld area that it does not have time to reach a critical accumulation.

Prevention of Hydrogen-Induced Cracking. The major preventative measures to avoid cold cracking are:

- PREHEAT, INCLUDING MAINTENANCE OF PROPER INTERPASS TEMPERATURE

- HEAT INPUT CONTROL
- POSTWELD HEAT TREATMENT
- BEAD TEMPERING
- USE OF LOW-HYDROGEN PROCESSES AND CONSUMABLES
- USE OF ALTERNATE FILLER MATERIALS (FOR EXAMPLE, AUSTENITIC ELECTRODES)

Preheating of the weld area is the most effective and widely used method for avoiding hydrogen-induced cracking. Its primary function is to reduce the weld metal cooling rate so that transformation to martensite is avoided or reduced below a certain critical level. The slower cooling also gives hydrogen more time to diffuse out of the weld area and delays the onset of maximum residual stresses. Many specifications and codes require the use of specific preheat and interpass temperatures for welding hardenable steels. AWS D1.1 (Ref 3) and CSA W59 (Ref 4) specify minimum preheat and interpass temperatures for various thicknesses of structural carbon steels (Table 2). Table 3 gives suggested preheat and postheat treatments for steels. Various methods are also available that allow one to determine the required preheat by taking into account the steel composition, restraint level, and hydrogen level. These methods are discussed in the section "Methods for Determining Preheat/Heat Input" in this article. Some are available as software packages. For example, PREHEAT is a software package from The Welding Institute that is based on BS 5135 (Ref 6) and is available from the Edison Welding Institute.

TABLE 2 RECOMMENDED PREHEAT AND INTERPASS TEMPERATURES FOR SELECTED THICKNESSES OF STRUCTURAL CARBON STEELS

CATEGORY	ASTM DESIGNATIONS	WELDING PROCESS	THICKNESS OF THICKEST PART AT POINT OF WELDING (*T*)		MINIMUM TEMPERATURE(A)(B)(D)	
			MM	IN.	°C	°F
A	A 36(C); A 53, GRADE B; A 106, GRADE B; A 131, GRADES A, B, CS, D, DS, AND E; A 139, GRADE B; A 381, GRADE Y35; A 500, GRADES A AND B; A 501; A 516; A 524, GRADES I AND II; A 529; A 570(H); A 573, GRADE 65; A 709, GRADE 36(C); API 5L(F), GRADES B AND X42; ABS(G), GRADES A, B, D, CS, DS, AND E	SHIELDED METAL ARC WELDING WITH OTHER THAN LOW HYDROGEN ELECTRODES	≤ 19	$\le \frac{3}{4}$	(E)	(E)
			$19 < T \le 38.1$	$\frac{3}{4} < T \le 1\frac{1}{2}$	66	150
			$38.1 < T \le 63.5$	$1\frac{1}{2} < T \le 2\frac{1}{2}$	107	225
			>63.5	$>2\frac{1}{2}$	150	300
B	A 36(C); A 53, GRADE B; A 106, GRADE B; A 131, GRADES A, B, CS, D, DS, E, AH 32, AH 36, DH 32, DH 36, EH 32, AND EH 36; A 139, GRADE B; A 242; A 381, GRADE Y35; A 441; A 500, GRADES A AND B; A 501; A 516, GRADES 55, 60, 65, AND 70; A 524, GRADES I AND II; A 529; A 537, CLASSES 1 AND 2; A 570(H); A 572, GRADES 42 AND 50; A 573, GRADE 65; A 588; A 595, GRADES A, B, AND C; A 606, A 607, GRADES 45, 50, AND 55; A 618; A 633, GRADES A, B, C, AND D; A 709, GRADES 36, 50, AND 50W(C); A 808; API 5L(F), GRADES B AND X42; API 2H(F), GRADES 42 AND 50; ABS(G), GRADES AH 32, AH 36, DH 32, DH 36, EH 32, EH 36, A, B, D, CS, DS, AND E	SHIELDED METAL ARC WELDING WITH LOW-HYDROGEN ELECTRODES, SUBMERGED ARC WELDING, GAS-METAL ARC WELDING, FLUX-CORED ARC WELDING	≤ 19	$\le \frac{3}{4}$	(E)	(E)
			$19 < T < 38.1$	$\frac{3}{4} < T \le 1\frac{1}{2}$	10	50
			$38.1 < T < 63.5$	$1\frac{1}{2} < T \le 2\frac{1}{2}$	66	150
			>63.5	$>2\frac{1}{2}$	107	225
C	A 572, GRADES 60 AND 65; A 633, GRADE E; API 5L(F), GRADE X52	SHIELDED METAL ARC WELDING WITH LOW-HYDROGEN ELECTRODES, SUBMERGED ARC WELDING, GAS-METAL ARC WELDING, FLUX-CORED	≤ 19	$\le \frac{3}{4}$	10	50
			$19 < T < 38.1$	$\frac{3}{4} < T$	66	150

				$\le 1\frac{1}{2}$		
			$38.1 < T < 63.5$	$1\frac{1}{2} < T \le 2\frac{1}{2}$	107	225
			>63.5	$>2\frac{1}{2}$	150	300
D	A 514; A 517; A 709, GRADES 100 AND 100W	SHIELDED METAL ARC WELDING WITH LOW-HYDROGEN ELECTRODES, SUBMERGED ARC WELDING WITH CARBON OR ALLOY STEEL WIRE, NEUTRAL FLUX, GAS-METAL ARC WELDING, OR FLUX-CORED ARC WELDING	≤ 19	$\le \frac{3}{4}$	10	50
			$19 < T < 38.1$	$\frac{3}{4} < T \le 1\frac{1}{2}$	50	125
			$38.1 < T < 63.5$	$1\frac{1}{2} < T \le 2\frac{1}{2}$	80	175
			>63.5	$>2\frac{1}{2}$	107	225
E	A 710, GRADE A(I)	. . .	. . .	. . .	(J)	(J)

Source: Ref 3

(A) WELDING SHALL NOT BE DONE WHEN THE AMBIENT TEMPERATURE IS LOWER THAN -18 °C (O °F). TEMPERATURE OF -18 °C (0 °F) DOES NOT MEAN THE AMBIENT ENVIRONMENTAL TEMPERATURE BUT THE TEMPERATURE IN THE IMMEDIATE VICINITY OF THE WELD. THE AMBIENT ENVIRONMENTAL TEMPERATURE MAY BE BELOW -18 °C (0 °F) BUT A HEATED STRUCTURE OR SHELTER AROUND THE AREA BEING WELDED COULD MAINTAIN THE TEMPERATURE ADJACENT TO THE WELDMENT AT -18 °C (0 °F) OR HIGHER. WHEN THE BASE METAL IS BELOW THE TEMPERATURE LISTED FOR THE WELDING PROCESS BEING USED AND THE THICKNESS OF MATERIAL BEING WELDED, IT SHALL BE PREHEATED (EXCEPT AS OTHERWISE PROVIDED) IN SUCH MANNER THAT THE SURFACES OF THE PARTS ON WHICH WELD METAL IS BEING DEPOSITED ARE AT OR ABOVE THE SPECIFIED MINIMUM TEMPERATURE FOR A DISTANCE EQUAL TO THE THICKNESS OF THE PART BEING WELDED, BUT NOT LESS THAN 75 MM (3 IN.) IN ALL DIRECTIONS FROM THE POINT OF WELDING. PREHEAT AND INTERPASS TEMPERATURES MUST BE SUFFICIENT TO PREVENT CRACK FORMATION. TEMPERATURE ABOVE THE MINIMUM SHOWN MAY BE REQUIRED FOR HIGHLY RESTRAINED WELDS. FOR A 514, A 517, AND A 709, GRADES 100 AND 100W STEEL, THE MAXIMUM PREHEAT AND INTERPASS TEMPERATURE SHALL NOT EXCEED 205 °C (400 °F) FOR THICKNESS UP TO 38.1 MM ($1\frac{1}{2}$ IN.) INCLUSIVE, AND 230 °C (450 °F) FOR GREATER

THICKNESS. HEAT INPUT WHEN WELDING A 514, A 517, AND A 709 GRADES 100 AND 100W STEEL SHALL NOT EXCEED THE STEEL PRODUCER'S RECOMMENDATION.

(B) IN JOINTS INVOLVING COMBINATIONS OF BASE METALS, PREHEAT SHALL BE AS SPECIFIED FOR THE HIGHER STRENGTH STEEL BEING WELDED.

(C) ONLY LOW-HYDROGEN ELECTRODES SHALL BE USED WHEN WELDING A 86 OR A 709 GRADE 36 STEEL MORE THAN 25.4 MM (1 IN.) THICK FOR DYNAMICALLY LOADED STRUCTURES.

(D) WHEN THE BASE METAL TEMPERATURE IS BELOW 0 °C (32 °F), THE BASE METAL SHALL BE PREHEATED TO AT LEAST 21 °C (70 °F) AND THIS MINIMUM TEMPERATURE MAINTAINED DURING WELDING.

(E) NONE.

(F) AMERICAN PETROLEUM INSTITUTE (API) SPECIFICATION.

(G) AMERICAN BUREAU OF SHIPPING (ABS) SPECIFICATION.

(H) ALL GRADES.

(I) ALL CLASSES.

(J) PREHEAT IS NOT REQUIRED FOR THE BASE METAL. PREHEAT FOR E80XX-X FILLER METAL SHALL BE AS FOR GROUP C; FOR HIGHER STRENGTH FILLER METAL TREAT AS GROUP D.

TABLE 3 RECOMMENDED PREHEAT AND INTERPASS TEMPERATURES FOR SELECTED CARBON STEELS

AISI-SAE STEEL SPECIFICATIONS	RECOMMENDED WELDING CONDITIONS										
	CARBON RANGE, %(A)	THICKNESS RANGE		MINIMUM PREHEAT AND INTERPASS TEMPERATURE				POSTWELD HEAT TREATMENT			PEENING MAY BE NECESSARY
				LOW HYDROGEN		OTHER THAN LOW HYDROGEN		DESIRABLE OR OPTIONAL	TEMPERATURE RANGE		
		MM	IN.	°C	°F	°C	°F		°C	°F	
1008, 1010, 1011, 1012, 1013	WITHIN SPECIFICATION	≤50	≤2	AMBIENT(B)		AMBIENT(B)		OPTIONAL	590-675	1100-1250	. . .
		>50 TO 100	>2 TO 4	38	100	65	150	OPTIONAL	590-675	1100-1250	YES
1015, 1016	WITHIN SPECIFICATION	≤50	≤2	AMBIENT(B)		AMBIENT(B)		OPTIONAL	590-675	1100-1250	. . .
		>50 TO 100	>2 TO 4	38	100	93	200	OPTIONAL	590-675	1100-1250	YES
1017, 1018, 1019	WITHIN SPECIFICATION	≤50	≤2	AMBIENT(B)		AMBIENT(B)		OPTIONAL	590-675	1100-1250	. . .
		>50 TO 100	>2 TO 4	38	100	121	250	OPTIONAL	590-675	1100-1250	YES
1020, 1021, 1022, 1023	WITHIN SPECIFICATION	≤50	≤2	AMBIENT(B)		38	100	OPTIONAL	590-675	1100-1250	. . .
		>50 TO 100	>2 TO 4	93	200	149	300	OPTIONAL	590-675	1100-1250	YES
1024, 1027	0.18-0.25	≤25	≤1	AMBIENT(B)		38	100	OPTIONAL	590-675	1100-1250	. . .
		>25 TO 50	>1 TO 2	38	100	93	200	OPTIONAL	590-675	1100-1250	. . .
		>50 TO 100	>2 TO 4	93	200	149	300	OPTIONAL	590-675	1100-1250	YES
	0.25-0.29	≤13	$\le \frac{1}{2}$	10	50	65	150	OPTIONAL	590-675	1100-1250	. . .

		>13 TO 25	> $\frac{1}{2}$ TO 1	65	150	149	300	OPTIONAL	590-675	1100-1250	. . .
		>25 TO 50	>1 TO 2	121	250	149	300	OPTIONAL	590-675	1100-1250	. . .
		>50 TO 100	>2 TO 4	149	300	177	350	OPTIONAL	590-675	1100-1250	YES
1025	WITHIN SPECIFICATION	≤25	≤1	AMBIENT(B)		AMBIENT(B)		OPTIONAL	590-675	1100-1250	. . .
		>25 TO 50	>1 TO 2	AMBIENT(B)		38	100	OPTIONAL	590-675	1100-1250	. . .
		>50 TO 100	>2 TO 4	38	100	93	200	OPTIONAL	590-675	1100-1250	YES
1026, 1029, 1030	0.25-0.30	≤25	≤1	AMBIENT(B)		38	100	DESIRABLE	590-675	1100-1250	. . .
		>25 TO 50	>1 TO 2	38	100	93	200	DESIRABLE	590-675	1100-1250	YES
		>50 TO 100	>2 TO 4	93	200	149	300	DESIRABLE	590-675	1100-1250	YES
	0.31-0.34	≤13	≤ $\frac{1}{2}$	AMBIENT(B)		38	100	DESIRABLE	590-675	1100-1250	. . .
		>13 TO 25	> $\frac{1}{2}$ TO 1	38	100	93	200	DESIRABLE	590-675	1100-1250	. . .
		>25 TO 50	>1 TO 2	93	200	149	300	DESIRABLE	590-675	1100-1250	YES
		>50 TO 100	>2 TO 4	121	250	177	350	DESIRABLE	590-675	1100-1250	YES
1035, 1037	WITHIN SPECIFICATION	≤13	≤ $\frac{1}{2}$	AMBIENT(B)		38	100	DESIRABLE	590-675	1100-1250	. . .
		>13 TO 25	> $\frac{1}{2}$ TO 1	38	100	93	200	DESIRABLE	590-675	1100-1250	. . .

		>25 TO 50	>1 TO 2	93	200	149	300	DESIRABLE	590-675	1100-1250	YES
		>50 TO 100	>2 TO 4	149	300	205	400	DESIRABLE	590-675	1100-1250	YES
1036, 1041	0.30-0.35	≤13	≤ $\frac{1}{2}$	38	100	93	200	DESIRABLE	590-675	1100-1250	. . .
		>13 TO 25	> $\frac{1}{2}$ TO 1	65	150	149	300	DESIRABLE	590-675	1100-1250	. . .
		>25 TO 50	>1 TO 2	121	250	149	300	DESIRABLE	590-75	100-1250	YES
		>50 TO 100	>2 TO 4	149	300	177	350	DESIRABLE	590-675	1100-1250	YES
	0.36-0.40	≤13	≤ $\frac{1}{2}$	65	150	93	200	DESIRABLE	590-675	1100-1250	. . .
		>13 TO 25	> $\frac{1}{2}$ TO 1	93	200	149	300	DESIRABLE	590-675	1100-1250	. . .
		>25 TO 50	>1 TO 2	149	300	177	350	DESIRABLE	590-675	100-1250	YES
		>50 TO 100	>2 TO 4	177	350	205	400	DESIRABLE	590-675	1100-1250	YES
	0.41-0.44	≤13	≤ $\frac{1}{2}$	93	200	149	300	DESIRABLE	590-675	1100-1250	. . .
		>13 TO 25	> $\frac{1}{2}$ TO 1	149	300	177	350	DESIRABLE	590-675	1100-1250	. . .
		>25 TO 50	>1 TO 2	177	350	205	400	DESIRABLE	590-675	1100-1250	YES
		>50 TO 100	>2 TO 4	205	400	230	450	DESIRABLE	590-675	1100-1250	YES
1038, 1039, 1040	0.34-0.40	≤13	≤ $\frac{1}{2}$	38	100	93	200	DESIRABLE	590-675	1100-1250	. . .

		>13 TO 25	> $\frac{1}{2}$ TO 1	93	200	149	300	DESIRABLE	590-675	1100-1250	. . .
		>25 TO 50	>1 TO 2	121	250	177	350	DESIRABLE	590-675	1100-1250	YES
		>50 TO 100	>2 TO 4	149	300	205	400	DESIRABLE	590-675	1100-1250	YES
	0.41-0.44	≤13	≤ $\frac{1}{2}$	65	150	121	250	DESIRABLE	590-675	1100-1250	. . .
		>13 TO 25	> $\frac{1}{2}$ TO 1	93	200	149	300	DESIRABLE	590-675	1100-1250	. . .
		>25 TO 50	>1 TO 2	149	300	177	350	DESIRABLE	590-675	1100-1250	YES
		>50 TO 100	>2 TO 4	205	400	230	450	DESIRABLE	590-675	1100-1250	YES
1042, 1043	WITHIN SPECIFICATION	≤13	≤ $\frac{1}{2}$	93	200	149	300	DESIRABLE	590-675	1100-1250	YES
		>13 TO 25	> $\frac{1}{2}$ TO 1	121	250	149	300	DESIRABLE	590-675	1100-1250	YES
		>25 TO 50	>1 TO 2	149	300	177	350	DESIRABLE	590-675	1100-1250	YES
		>50 TO 100	>2 TO 4	205	400	230	450	DESIRABLE	590-675	1100-1250	YES
1044, 1045, 1046	WITHIN SPECIFICATION	≤13	≤ $\frac{1}{2}$	149	300	177	350	DESIRABLE	590-675	1100-1250	YES
		>13 TO 100	> $\frac{1}{2}$ TO 4	205	400	230	450	DESIRABLE	590-675	1100-1250	YES
1048, 1049, 1050, 1052, 1053	0.43-0.50	≤25	≤1	93	200	149	300	DESIRABLE	590-675	1100-1250	YES
		>25 TO 50	>1 TO 2	149	300	177	350	DESIRABLE	590-675	1100-1250	YES

		>50 TO 100	>2 TO 4	205	400	230	450	DESIRABLE	590-675	1100-1250	YES
1108, 1109, 1110	WITHIN SPECIFICATION	≤50	≤2	AMBIENT$^{(B)}$		NOT RECOMMENDED		OPTIONAL	590-675	1100-1250	. . .
		>50 TO 100	>2 TO 4	38	100	NOT RECOMMENDED		OPTIONAL	590-675	1100-1250	YES
1116, 1117, 1118, 1119	WITHIN SPECIFICATION	≤25	≤1	AMBIENT$^{(B)}$		NOT RECOMMENDED		OPTIONAL	590-675	1100-1250	. . .
		>25 TO 100	>1 TO 4	93	200	NOT RECOMMENDED		OPTIONAL	590-675	1100-1250	YES
1132, 1137, 1139, 1140, 1141, 1144, 1145, 1146, 1151	0.27-0.30	≤13	$\le \frac{1}{2}$	10	50	NOT RECOMMENDED		OPTIONAL	590-675	1100-1250	. . .
		>13 TO 25	$> \frac{1}{2}$ TO 1	38	100	NOT RECOMMENDED		DESIRABLE	590-675	1100-1250	. . .
		>25 TO 50	>1 TO 2	93	200	NOT RECOMMENDED		DESIRABLE	590-675	1100-1250	YES
		>50 TO 100	>2 TO 4	121	250	NOT RECOMMENDED		DESIRABLE	590-675	1100-1250	YES
	0.31-0.35	≤13	$\le \frac{1}{2}$	38	100	NOT RECOMMENDED		DESIRABLE	590-675	1100-1250	. . .
		>13 TO 25	$> \frac{1}{2}$ TO 1	65	150	NOT RECOMMENDED		OPTIONAL	590-675	1100-1250	. . .
		>25 TO 50	>1 TO 2	121	250	NOT RECOMMENDED		DESIRABLE	590-675	1100-1250	YES
		>50 TO 100	>2 TO 4	149	300	NOT RECOMMENDED		DESIRABLE	590-675	1100-1250	YES
	0.36-0.40	≤13	$\le \frac{1}{2}$	65	150	NOT RECOMMENDED		DESIRABLE	590-675	1100-1250	. . .
		>13 TO 25	$> \frac{1}{2}$	93	200	NOT RECOMMENDED		DESIRABLE	590-675	1100-1250	. . .

			TO 1							
		>25 TO 100	>1 TO 4	149	300	NOT RECOMMENDED	DESIRABLE	590-675	1100-1250	YES
	0.41-0.45	≤13	$\leq \frac{1}{2}$	93	200	NOT RECOMMENDED	DESIRABLE	590-675	1100-1250	. . .
		>13 TO 25	$> \frac{1}{2}$ TO 1	149	300	NOT RECOMMENDED	DESIRABLE	590-675	1100-1250	. . .
		>25 TO 100	>1 TO 4	177	350	NOT RECOMMENDED	DESIRABLE	590-675	1100-1250	YES
	0.45-0.50	≤13	$\leq \frac{1}{2}$	149	300	NOT RECOMMENDED	DESIRABLE	590-675	1100-1250	. . .
		>13 TO 25	$> \frac{1}{2}$ TO 1	205	400	NOT RECOMMENDED	DESIRABLE	590-675	1100-1250	. . .
		>25 TO 100	>1 TO 4	230	450	NOT RECOMMENDED	OPTIONAL	590-675	1100-1250	YES
1211, 1212, 1213, 1215, B1111, B1112, B1113	WITHIN SPECIFICATION	≤50	≤2	AMBIENT(B)		NOT RECOMMENDED	OPTIONAL	590-675	1100-1250	. . .
		>50 TO 100	>2 TO 4	38	100	NOT RECOMMENDED	OPTIONAL	590-675	1100-1250	YES
12L 13(C), 12L 14(C)	WITHIN SPECIFICATION	≤50	≤2	AMBIENT(B)		NOT RECOMMENDED	OPTIONAL	590-675	1100-1250	. . .
		>50 TO 100	>2 TO 4	38	100	NOT RECOMMENDED	OPTIONAL	590-675	1100-1250	YES
1330, 1335, 1340, 1345	0.27-0.33	≤13	$\leq \frac{1}{2}$	121	250	NOT RECOMMENDED	DESIRABLE	550-565	1025-1050	. . .
		>13 TO 25	$> \frac{1}{2}$ TO 1	149	300	NOT RECOMMENDED	DESIRABLE	550-565	1025-1050	. . .
		>25 TO 50	>1 TO 2	177	350	NOT RECOMMENDED	DESIRABLE	550-565	1025-1050	YES

	0.33-0.38	≤13	≤ 1/2	149	300	NOT RECOMMENDED		DESIRABLE	550-565	1025-1050	. . .
		>13 TO 25	> 1/2 TO 1	205	400	NOT RECOMMENDED		DESIRABLE	550-565	1025-1050	. . .
		>25 TO 50	>1 TO 2	205	400	NOT RECOMMENDED		DESIRABLE	550-565	1025-1050	YES
	0.38-0.43	≤13	≤ 1/2	177	350	NOT RECOMMENDED		DESIRABLE	550-565	1025-1050	. . .
		>13 TO 25	> 1/2 TO 1	230	450	NOT RECOMMENDED		DESIRABLE	550-565	1025-1050	. . .
		>25 TO 50	>1 TO 2	290(D)	550(D)	NOT RECOMMENDED		DESIRABLE	550-565	1025-1050	YES
	0.43-0.49	≤13	≤ 1/2	205	400	NOT RECOMMENDED		DESIRABLE	550-565	1025 1050	. . .
		>13 TO 25	> 1/2 TO I	260(D)	500	NOT RECOMMENDED		DESIRABLE	550-565	1025-1050	. . .
		>25 TO 50	>1 TO 2	315	600(D)	NOT RECOMMENDED		DESIRABLE	550-565	1025-1050	YES
1513, 1518, 1522, 1525	≤0.20	≤25	≤1	AMBIENT(B)		AMBIENT(B)		OPTIONAL	590-675	1100-1250	. . .
		>25 TO 50	> 1 TO 2	AMBIENT(B)		38	100	OPTIONAL	590-675	1100-1250	. . .
		>50 TO 100	>2 TO 4	38	100	93	200	OPTIONAL	590-675	1100-1250	YES
	0.21-0.25	≤25	≤1	AMBIENT(B)		38	100	OPTIONAL	590-675	1100-1250	. . .
		>25 TO 50	>1 TO 2	38	100	93	200	OPTIONAL	590-675	1100-1250	. . .
		>50 TO 100	>2 TO 4	93	200	149	300	OPTIONAL	590-675	1100-1250	YES
	0.26-0.29	≤13	≤ 1/2	10	50	65	150	OPTIONAL	590-675	1100-1250	. . .

		>13 TO 25	> $\frac{1}{2}$ TO 1	65	150	149	300	OPTIONAL	590-675	1100-1250	. . .
		>25 TO 50	>1 TO 2	121	250	149	300	OPTIONAL	590-675	1100-1250	. . .
		>50 TO 100	>2 TO 4	149	300	177	350	OPTIONAL	590-675	1100-1250	YES
1524, 1526, 1527	≤0.25	≤25	≤1	10	50	38	100	OPTIONAL	590-675	1100-1250	. . .
		>25 TO 50	>1 TO 2	38	100	93	200	OPTIONAL	590-675	1100-1250	. . .
		>50 TO 100	>2 TO 4	93	200	149	300	OPTIONAL	590-675	1100-1250	YES
	0.26-0.29	≤25	≤1	65	150	121	250	DESIRABLE	590-675	1100-1250	. . .
		>25 TO 50	>1 TO 2	121	250	149	300	DESIRABLE	590-675	1100-1250	. . .
		>50 TO 100	>2 TO 4	149	300	177	350	DESIRABLE	590-675	1100-1250	YES
1536, 1541	≤0.35	≤13	≤ $\frac{1}{2}$	38	100	93	200	DESIRABLE	590-675	1100-1250	. . .
		>13 TO 25	> $\frac{1}{2}$ TO 1	65	150	149	300	DESIRABLE	590-675	1100-1250	. . .
		>25 TO 50	>1 TO 2	121	250	149	300	DESIRABLE	590-675	1100-1250	YES
		>50 TO 100	>2 TO 4	149	300	177	350	DESIRABLE	590-675	1100-1250	YES
	0.364-0.40	≤ 13	≤ $\frac{1}{2}$	65	150	93	200	DESIRABLE	590-675	1100-1250	. . .
		>13 TO 25	> $\frac{1}{2}$ TO 1	93	200	149	300	DESIRABLE	590-675	1100-1250	. . .

		>25 TO 50	>1 TO 2	149	300	177	350	DESIRABLE	590-675	1100-1250	YES
		>50 TO 100	>2 TO 4	177	350	205	400	DESIRABLE	590-675	1100-1250	YES
	0.414-0.44	≤ 13	≤ $\frac{1}{2}$	93	200	149	300	DESIRABLE	590-675	1100-1250	. . .
		>13 TO 25	> $\frac{1}{2}$ TO 1	149	300	177	350	DESIRABLE	590-675	1100-1250	. . .
		>25 TO 50	>1 TO 2	177	350	205	400	DESIRABLE	590-675	1100-1250	YES
		>50 TO 100	>2 TO 4	205	400	230	450	DESIRABLE	590-675	1100-1250	YES
1547, 1548	0.43-0.52	≤ 13	≤ $\frac{1}{2}$	149	300	177	350	DESIRABLE	590-675	1100-1250	. . .
		>13 TO 100	> $\frac{1}{2}$ TO 4	205	400	230	450	DESIRABLE	590-675	1100-1250	YES

Source: Ref 5

(A) RANGES ARE INCLUSIVE.

(B) AMBIENT ABOVE -12 °C (10 °F).

(C) DUE TO LEAD CONTENT, MANUFACTURING OPERATIONS INVOLVING ELEVATED TEMPERATURES IN THE RANGE OF THOSE ENCOUNTERED IN GAS CUTTING OR WELDING SHOULD BE CARRIED OUT WITH ADEQUATE VENTILATION.

(D) HOLD AT TEMPERATURE FOR 1 H AFTER WELDING IS COMPLETED.

Particular care must be exercised when dealing with weldments made at low heat inputs (high cooling rates). Tack welds are usually short, low-heat-input welds that cool rapidly. The first pass of a multipass weld in a large structure may be made at low heat input (to avoid bum-through) or on a relatively cold workpiece. Again, cooling rates may be high. It is essential that proper preheat be used under these circumstances, because the likelihood of martensite formation is much greater, given the rapid cooling rates.

Heat input control can also be used to reduce the cooling rate. The use of a higher heat input to slow cooling may not always be commensurate with good welding practice. For example, the resultant bead size may be too large, or the welding parameters (current, voltage, travel speed) needed to achieve the required heat input level may be outside the operating window of the electrode.

Postweld heat treatment may involve holding the completed weld at or near the interpass temperature for a period of time sufficient to allow hydrogen to escape the weld area. This temperature should be above the temperature range where cracking occurs (approximately 200 °C, or 390 °F).

If the interpass and postweld soak temperature are held above the martensite start temperature, M_s, for a sufficient length of time, isothermal transformation to a less susceptible microstructure will occur (for example, bainitic structure with some ferrite). A tempered martensite, however, will generally have better toughness than the isothermal transformation product, so this procedure should be used only when a stress relief is not possible.

The use of a proper postweld treatment may allow a reduction in the preheat temperature because both treatments serve the same purpose (that is, reducing the weld metal cooling rate). A lower preheat will increase operator comfort.

In the event that martensite formation cannot be avoided (as may be the case in a medium- to high-carbon steel), or when joint restraint is high, it is advisable to postweld heat treat the weldment to produce a less brittle, tempered martensite. The weldment should be held at or near the interpass temperature until postweld heat treatment can be done. This will reduce the possibility of cracking before the postweld heat treatment. Postweld heat treating of carbon steels is normally performed in the 600 to 650 °C (1110 to 1200 °F) range. The soak time is usually 1 h/in. of the thickest member. Many codes (for example, the ASME Boiler and Pressure Vessel Code) define when postweld heat treatment is to be done. Figure 4 can also be used as a guide for determining when postweld heat treatment should be done.

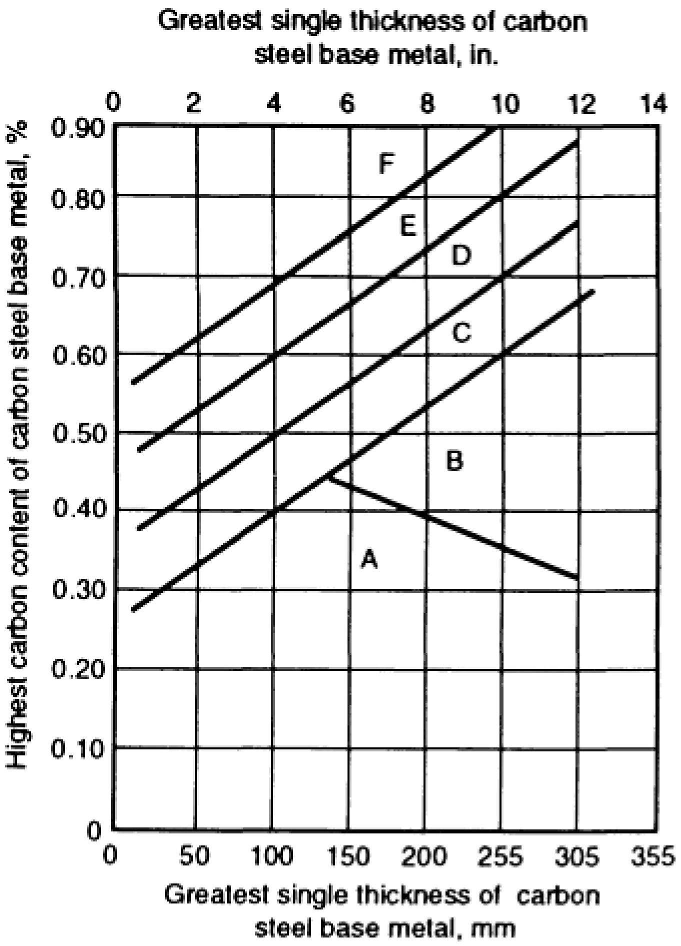

FIG. 4 PLOT OF BASE-METAL CARBON CONTENT VERSUS BASE-METAL THICKNESS OF CARBON STEELS TO SHOW DEGREE OF STRESS RELIEF REQUIRED. A, POSTWELD HEATING IS SELDOM REQUIRED. B, POSTWELD HEATING IS REQUIRED ONLY FOR DIMENSIONAL STABILITY, AS WHEN PARTS ARE TO BE FINISH MACHINED AFTER WELDING. C, POSTWELD STRESS RELIEF IS HIGHLY DESIRABLE FOR REPETITIVELY LOADED OR SHOCK-LOADED STRUCTURES AND FOR RESTRAINED JOINTS HAVING A THICKNESS GREATER THAN 25 MM (1 IN.). D, POSTWELD STRESS RELIEF IS REQUIRED FOR ALL REPETITIVELY LOADED OR SHOCK-LOADED STRUCTURES, FOR ALL RESTRAINED JOINTS, AND FOR ALL THICKNESSES OVER 50 MM (2 IN.); IT IS ALSO DESIRABLE FOR STATICALLY LOADED STRUCTURES. E, POSTWELD STRESS RELIEF IS RECOMMENDED FOR ALL APPLICATIONS; NO INTERMEDIATE COOLDOWN SHOULD BE PERMITTED FOR RESTRAINED STRUCTURES OR FOR BASE METAL HAVING A THICKNESS GREATER THAN 50 MM (2 IN.). F, SAME AS E, EXCEPT HAZARDS ARE GREATER

Bead Tempering. In a multipass weld, the underlying weld metal is subjected to repeated thermal cycles. This extends the time for hydrogen to diffuse out of the weld area, and it tempers any martensite that may have previously formed. Each bead also provides the preheat for the subsequent weld pass. The final bead and the HAZ it creates will not receive this same benefit. If the base metal is crack-sensitive, the final bead should be placed so that the HAZ does not form in the base metal (Fig. 5).

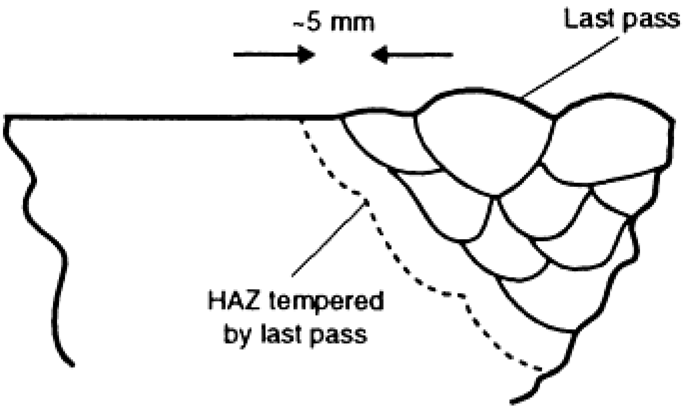

FIG. 5 BEAD TEMPERING. THE LAST BEAD IS PLACED SO ITS HAZ IS FORMED IN THE WELD METAL RATHER THAN IN THE MORE HARDENABLE BASE METAL. SOURCE: REF 1

Use of Low-Hydrogen Processes and Consumables. The principal source of weld metal hydrogen is the welding consumable(s). Careful selection of low-hydrogen consumables or low-hydrogen processes (such as gas-metal arc welding) can greatly reduce hydrogen pickup. This is especially important with medium- to high-carbon steels, in which the formation of a hard, brittle martensite is more likely. Many new specifications (for example, AWS A5.*x* filler metal specifications) now provide methods for consumable classification based on the diffusible hydrogen levels of their deposited weld metals. A standardized procedure is prescribed for measuring the diffusible hydrogen (for AWS specifications, this is AWS A4.3, while in many other parts of the world ISO 3690 is used). Designators are then used to indicate the maximum hydrogen level that an electrode will produce. The AWS uses a logarithmic scale; the International Institute of Welding method uses a linear system (Table 4).

TABLE 4 CLASSIFICATION OF CONSUMABLES RELATIVE TO THE DIFFUSIBLE HYDROGEN LEVELS OF THEIR DEPOSITED WELD METALS

SPECIFICATION	HYDROGEN LEVELS AT INDICATED DESIGNATIONS, ML/100 G		
	VERY LOW	LOW	HYDROGEN-CONTROLLED
AWS	<4(A)	<8(B)	<16(C)
IIW	<5	<10	<15

(A)	H4.
(B)	H8.
(C)	H16

The extension of the AWS series down to diffusible hydrogen levels of less than 2 mL/100 g is feasible (for example, for gas-metal arc welding, flux-cored arc welding, or submerged arc welding). In both systems, an electrode that produces hydrogen in excess of the maximum, 16 mL/100 g (or 15 mL/100 g), is not considered hydrogen-controlled.

Use of Alternate Filler Materials. Often, the use of a softer, more ductile weld metal can help to absorb some of the high residual stresses in the weld area. The use of a lower-strength carbon steel electrode with greater ductility will reduce the high residual stresses in the HAZ and reduce the likelihood of cracking in the HAZ. Care must be exercised to be sure the decreased weld metal strength does not compromise the usefulness of the welded structure.

In higher carbon steels, austenitic weld metals are often used. They not only help reduce the stress level in the HAZ, but they also absorb and hold weld metal hydrogen. Austenite has a higher solubility for hydrogen, and the relatively low diffusivity of hydrogen in austenite ensures that little hydrogen will diffuse into the more susceptible HAZ. Caution must also be exercised here. The strength of the austenitic deposit will be lower than that of a carbon or alloy steel deposit. The austenitic filler material should be selected carefully to avoid other problems (for example, solidification cracking).

Methods for Determining Preheat/Heat Input. It is common practice to find preheat information by using published tables (for example, WRC Bulletin No. 326) or tables in applicable codes (for example, AWS D1.1). There are also various methods available that allow the fabricator to determine the necessary preheat (and/or heat input) by taking into account the effect of important variables such as steel composition, amount of restraint, and hydrogen level. Many codes, such as AWS D1.1 (Ref 3), Canadian Standard W59 (Ref 4), and British Standard 5135 (Ref 6), provide methods that the fabricator can use to determine preheat and/or heat input. The method used depends on the hardenability of the steel.

For lower hardenability steels, a hardness control method is used. The basis of this method is that welding conditions can be chosen that will give cooling rates slow enough to avoid the formation of a hard, brittle microstructure in the HAZ. Critical hardness levels, below which cracking is unlikely to occur, have been established. For high-hydrogen consumables, this critical hardness value is 350 HV, while for low-hydrogen consumables (<10 mL/100 g of diffusible hydrogen) it increases to 400 HV.

The cooling rate required to keep the HAZ hardness below the critical value is a function of the composition of the steel (this can be seen in Fig. 2). A carbon equivalent (CE) is used as an indicator of the hardenability of the steel. If the CE is known, Fig. 6 can be used to determine the cooling rate necessary to avoid cracking. Then the heat input and, if necessary, preheat to produce that cooling rate can be determined by using published data for various plate thicknesses. This is done graphically in AWS D1.1 (Ref 3) and CSA W59 (Ref 4) for determining the heat inputs necessary to make crackfree fillet welds without the need to preheat. Figure 7 is an example for submerged arc fillets with 13 mm ($\frac{1}{2}$ in.) flanges. Reference 7 (upon which specification BS 5135 is based) provides a more thorough treatment that allows one to determine the required heat inputs and, if necessary, preheat. An example of the graphical presentation is given in Fig. 8. Knowing the carbon equivalent, one can determine an energy input/preheat combination for a variety of plate thicknesses. The hardness control method is used for carbon steels up to CE = 0.6.

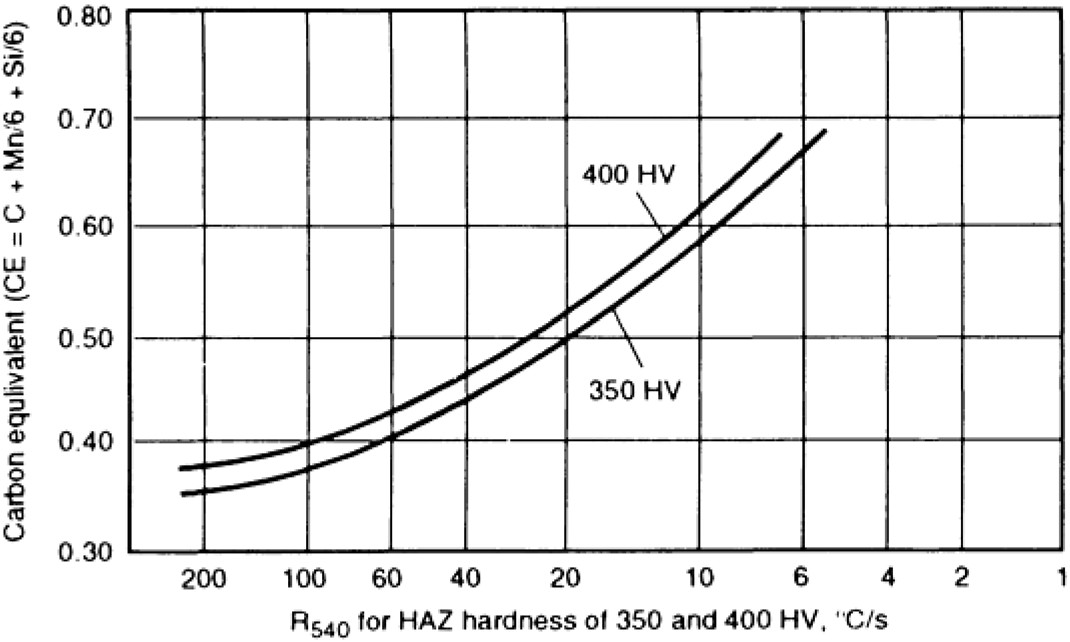

FIG. 6 CRITICAL COOLING RATES TO OBTAIN HARDNESS OF 350 HV (HIGH-HYDROGEN CONSUMABLES) AND 400 HV (LOW-HYDROGEN CONSUMABLES). R_{540} IS COOLING RATE AT 540 °C (1005 °F). SOURCE: REF 3

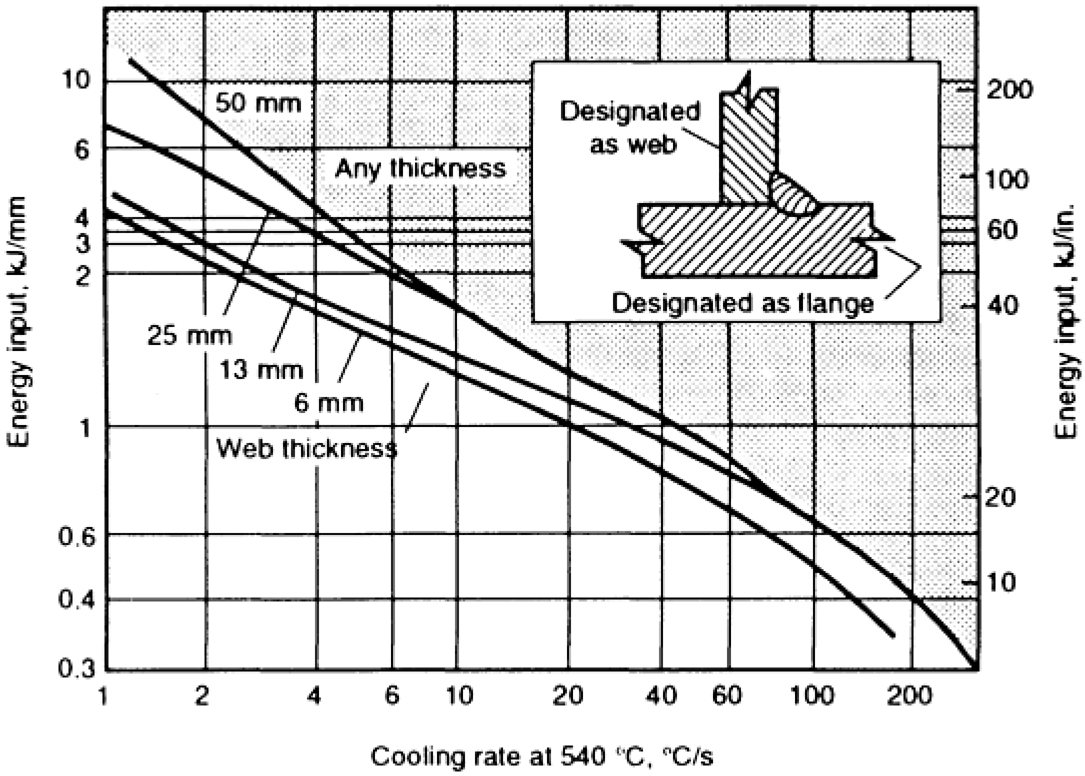

FIG. 7 PLOT OF MINIMUM ENERGY INPUT VERSUS CRITICAL COOLING RATE TO WELD SINGLE-PASS SAW FILLET WELD (WITH 13 MM, OR $\frac{1}{2}$ IN., FLANGES) WITHOUT PREHEAT. SOURCE: REF 3

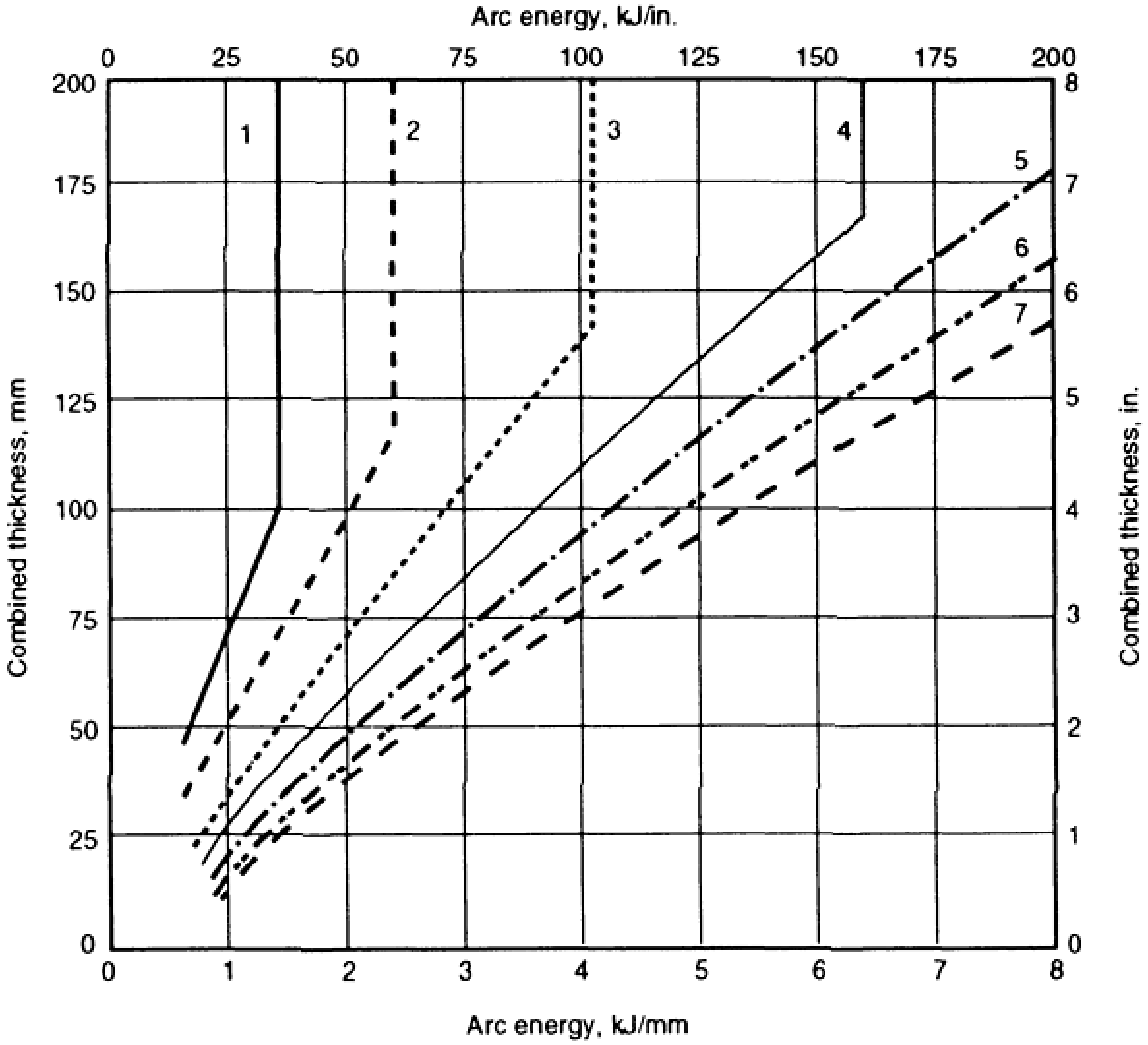

LINE	MINIMUM PREHEAT TEMPERATURE	
	°C	°F
1	175	**345**
2	150	**300**
3	125	**255**
4	100	**212**
5	75	**165**
6	20	**70**
7	**0**	32

CARBON EQUIVALENT OF STEEL	**AT HYDROGEN SCALE**	EQUIVALENT DIFFUSIBLE H_2, ML/100 G
≤0.47	A	**> 15**
≤0.49	B	**10-15**
≤0.51	C	**5-10**
≤**0.57**	**D**	**< 5**

FIG. 8 PLOT USED TO OBTAIN HEAT INPUT/PREHEAT COMBINATIONS FOR SPECIFIC CARBON EQUIVALENTS.

THE SCALES A, B, C, AND D REFER TO THE LEVEL OF DIFFUSIBLE HYDROGEN AS CERTIFIED BY THE ELECTRODE MANUFACTURER. SOURCE: REF 7

For higher hardenability steels, a hydrogen control method is used. Even slow cooling rates cannot prevent the formation of hard, brittle microstructures. In order to avoid cracking, a hydrogen control method is used. The hydrogen control method consists of holding the weld at an elevated temperature, above the range where cracking occurs, so that hydrogen can diffuse out of the weld area more rapidly. This method requires a preheat, an interpass temperature to maintain the heat, and a postweld soak to hold the heat after welding. Both AWS D1.1 (Appendix XI) (Ref 3) and CSA W59 (Appendix P) (Ref 4) provide a hydrogen control method that was designed for low-alloy steels and that may yield conservative results for carbon steels. Reference 7 provides a variety of data for different grades of steel. Figure 9 can be used for medium-carbon steels with up to 0.45% C.

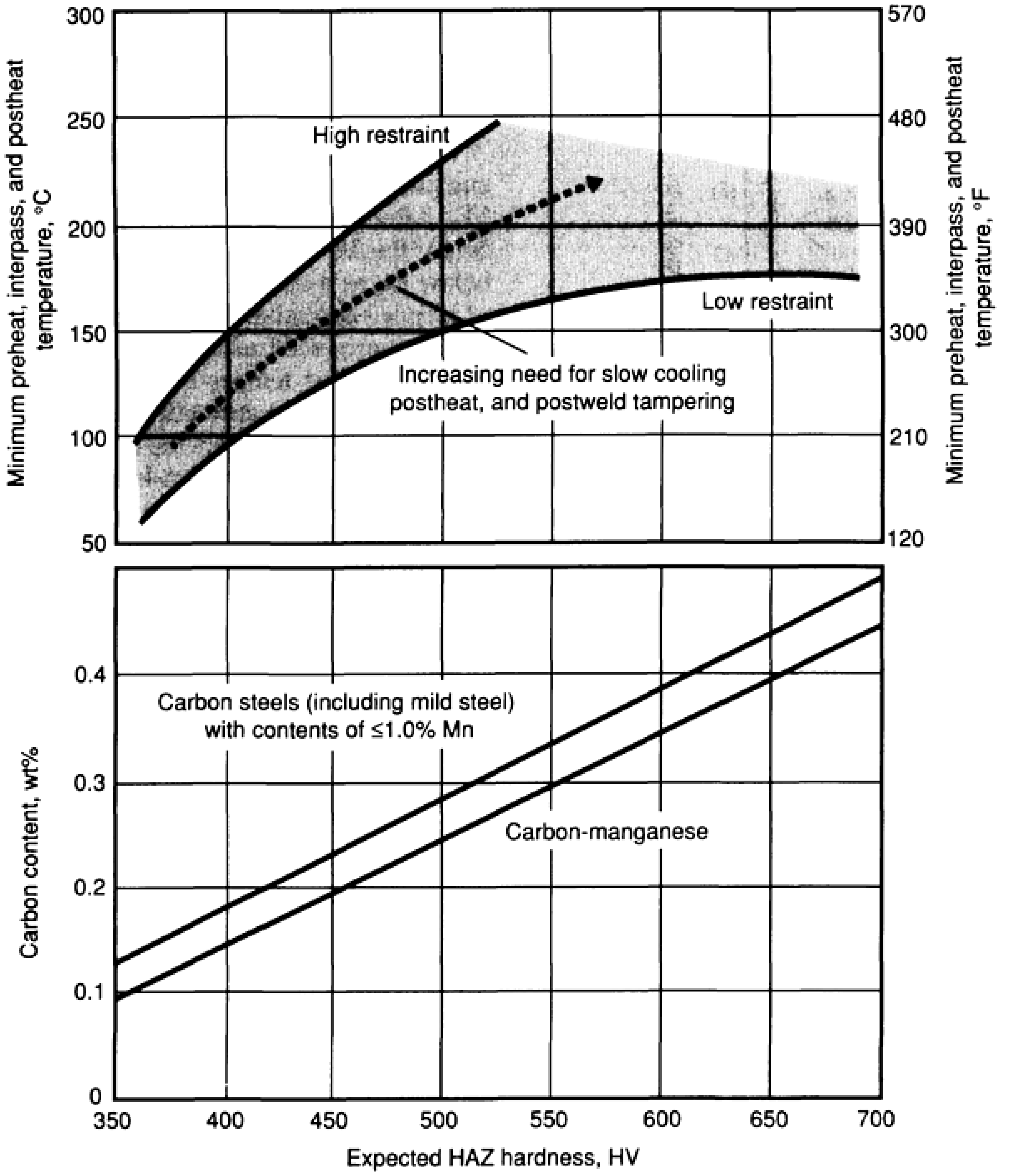

FIG. 9 PLOT FOR SELECTING PREHEAT, INTERPASS, AND POSTWELD SOAK TEMPERATURES FOR FULLY HARDENED CARBON STEELS USING HYDROGEN CONTROL METHOD. SOURCE: REF 7

Medium-carbon steels are often stress-relieved, especially at higher carbon levels, at higher restraint, and for welds exposed to shock or fatigue loading. When stress relief is not possible, an isothermal transformation method may be used. The use of this method requires knowledge about the isothermal transformation characteristics of the steel. By maintaining preheat, interpass, and postweld soak temperatures above the M_s for a sufficient length of time, a nonmartensitic microstructure is formed. This microstructure (usually bainite) will be less susceptible to hydrogen-induced cracking than martensite, but more susceptible than tempered martensite. For this reason, stress relieving is preferred when it can be done.

When even minimal preheat cannot be used with medium- or high-carbon steels, the use of an austenitic filler material should be considered. However, it should be noted that an austenitic filler cannot meet the yield strength requirements of a medium- or high-carbon steel. The high solubility of hydrogen in the austenitic weld metal and the slow diffusion rates limit the amount of hydrogen in the HAZ. At higher base-plate carbon levels (>0.4% C), a 150 °C (300 °F) preheat is recommended to reduce cracking tendency.

For high-carbon steels, the preheat, interpass, and postweld temperature should be equal to the M_s. After welding, the weld should be cooled below the M_s (60 to 70 °C, or 110 to 125 °F, below M_s as a minimum) to reduce the amount of retained austenite. A double temper treatment is then introduced as an additional heat treatment. The first temper treatment converts the remaining retained austenite to martensite. The second treatment tempers the martensite.

Overview of Hydrogen-Induced Cracking in Carbon Steels. The following guidelines apply to hydrogen-induced cracking in carbon steels.

Carbon steels with less than 0.2% C are generally not susceptible to cracking, and precautions are usually unnecessary.

Mild steels can be welded with hardness control methods. Up to CE of 0.40, cooling rates can be controlled by energy input, and preheat is usually not necessary except for thick plates. Above CE of 0.40, preheat is required except in thin sections. The use of low-hydrogen consumables is recommended at higher carbon and manganese levels and will reduce preheat requirements at all carbon equivalents. A heat-treated mild steel may require a low-alloy steel weld metal to match its properties. The resultant weld metal itself may be susceptible to hydrogen-induced cracking, and suitable precautions should be taken (such as using the hydrogen control method based on weld metal composition). Low-hydrogen consumables should be used.

Medium-carbon steels up to CE = 0.6 can be welded with hardness control methods. Above CE = 0.6, hydrogen control methods (preheat, interpass, and postweld soak) should be used with low-hydrogen consumables. Stress relief is recommended at high carbon levels (>0.45% C), for highly restrained welds, or for welds subjected to fatigue or shock loading. When stress relief is not possible, the isothermal transformation method should be considered.

High-carbon steels require preheat equal to the martensite start temperature, followed by a double temper. When tempering or isothermal transformation is not possible, the use of an austenitic filler material should be considered. Procedure testing for each specific application of high-carbon steel welding is essential.

Solidification Cracking

Solidification cracking occurs at high temperature, at the terminal stages of solidification. The cracks are easily recognized because they usually follow the weld centerline, although they do not always come to the surface. The cracks often appear towards the end of the weld, due to stresses arising from the thermal field. The crack surfaces are heavily oxidized, and dendritic structures can often be found upon examination with scanning electron microscopy.

Solidification cracks arise due to the formation of low melting compounds at grain boundaries. When a coherent solid network forms upon solidification, stresses start to build across the weld due to shrinkage. The low melting compounds remain molten, and the grain boundaries rupture under the shrinkage stresses. The conditions that control the extent of solidification cracking are the amount of impurities, the shape of the weld, and the stress level.

Impurities. Sulfur and phosphorus are the main causes of solidification cracking in carbon steel weld metal. Both will segregate to grain boundaries and form low melting compounds. In the case of sulfur, this is FeS, which tends to form a continuous film at the grain boundaries. The addition of a sufficient amount of manganese to the weld metal will cause MnS to form in preference to FeS. Manganese sulfide solidifies at a higher temperature and forms as globules, a morphology much less prone to cracking. Figure 10 shows the ratio of manganese to sulfur necessary to avoid cracking. Carbon also has a very strong effect, with the cracking tendency increasing with carbon content. At higher carbon levels, increasing the manganese-to-sulfur ratio is less effective, and it is more beneficial to lower the carbon content to avoid cracking. The effect of carbon is believed to be related to the phase, δ-ferrite or austenite, that forms upon solidification. At lower carbon contents, δ-ferrite, which has a relatively high solubility for sulfur and phosphorus, is the primary phase formed. As the carbon content increases, more austenite is formed upon solidification. Because austenite has a low solubility for sulfur and phosphorus, more of these impurities are rejected upon solidification and tend to build up at interdendritic boundaries.

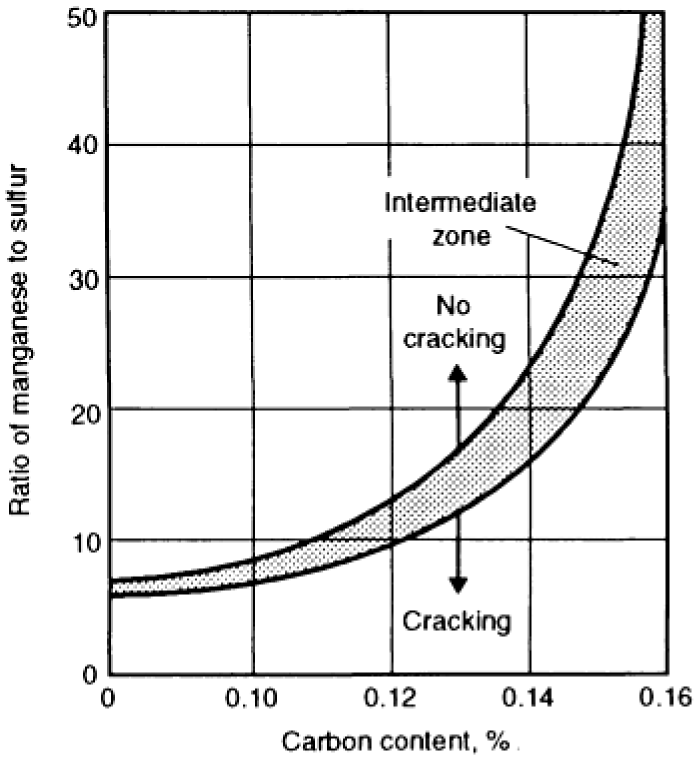

FIG. 10 EFFECT OF MANGANESE-TO-SULFUR RATIO AND CARBON CONTENT ON SOLIDIFICATION CRACKING TENDENCIES IN CARBON STEEL WELD METAL. SOURCE: REF 8

Welding consumables are kept deliberately low in sulfur and phosphorus to avoid problems. Base plates are the primary contributors of these impurities. Any high-dilution welding procedure or process (for example, submerged arc welding) will promote the pickup of impurities.

The shape of the weld bead and the solidification grain structure also affect solidification cracking (Fig. 11). Concave weld beads have a greater tendency to crack because their outer surface is stressed in tension. A slightly convex bead is always preferred. Deep narrow beads also have a greater tendency to crack. The depth-to-width ratio (*d*/*w*) should always be between 0.5 and 0.8.

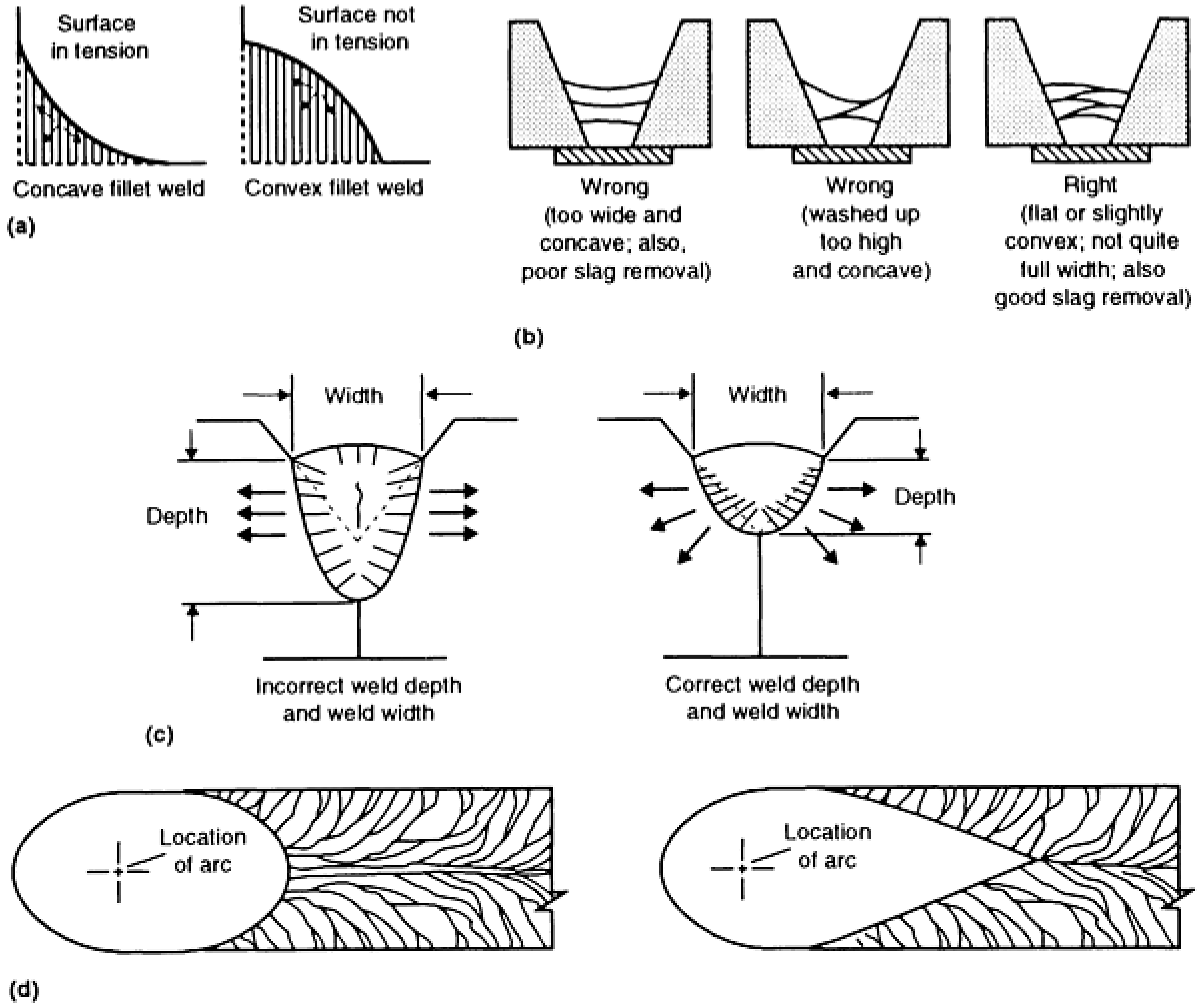
Surface
in tension
Surface not
in tension
Concave fillet weld
Convex fillet weld
(a)
Wrong
(too wide and
concave; also,
poor slag removal)
Wrong
(washed up
too high
and concave)
Right
(flat or slightly
convex; not quite
full width; also
good slag removal)
(b)
Width
Depth
Incorrect weld depth
and weld width
Width
Depth
Correct weld depth
and weld width
(c)
Location
of arc
Location
of arc
(d)

FIG. 11 EFFECT OF WELD BEAD SHAPE AND SOLIDIFICATION PATTERN ON SOLIDIFICATION CRACKING. (A) AND (B) CONCAVE VERSUS CONVEX BEAD SHAPES. SLIGHTLY CONVEX BEAD SHAPES ARE PREFERRED. (C) PREFERRED DEPTH AND WIDTH OF WELD BEAD. DEEP, NARROW BEADS SHOULD BE AVOIDED. (D) EFFECT OF TRAVEL SPEED ON CRACKING POTENTIAL. HIGH TRAVEL SPEEDS (TEARDROP-SHAPED PUDDLE, ON RIGHT) RESULT IN MORE CENTERLINE SEGREGATION AND A GREATER CHANCE OF CRACKING. SOURCE: REF 9, 10

High travel speeds tend to cause the dendrites growing from either side of the weld to meet head on. This results in heavy segregation of impurities at the weld centerline and a greater likelihood of centerline cracking. Teardrop-shaped puddles, which produce this type of solidification pattern, should be avoided if the weld metal is susceptible to hot cracking.

Weld Configuration. The greater the restraint in a joint, the greater the likelihood of solidification cracking. Particular care must be taken with root beads. They are generally small, highly diluted beads (a problem if the steel is high in carbon, sulfur, or phosphorus) that can be subjected to very high stresses. Concave weld beads should be avoided here also. A convex bead gives a better stress distribution and increases the throat dimension, thus providing more cross-sectional area to carry the stresses. A slight offset of the plates (Fig. 12) and freedom for them to move will reduce the residual stresses. If no offset is used, a gap at the root should be provided to allow for shrinkage movement. When welding higher-carbon steels, a low-carbon electrode will reduce the weld metal carbon level and decrease the likelihood of cracking.

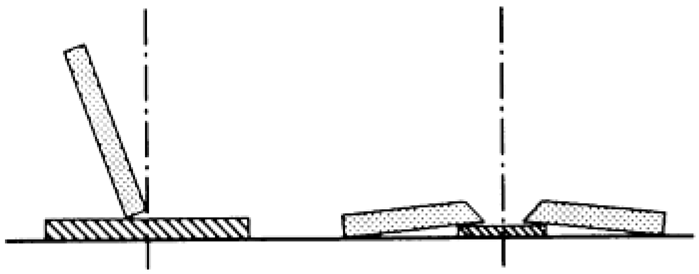

FIG. 12 DESIGN FOR OFFSETTING PLATES THAT ALLOWS THEM TO MOVE INTO FINAL POSITION, THEREBY REDUCING RESIDUAL STRESSES. SOURCE: REF 11

Welding of Free-Machining Steels. SAE-AISI series 11*xx* and 12*xx* steels are a particular problem because they contain high amounts of sulfur, phosphorus, or lead for improved machinability (see Tables 1(a) and 1(b)). The 11*xx* series steels are all resulfurized. The 12*xx* series steels are resulfurized and rephosphorized. The letter "L" after the first two numbers in any steel designation indicates the presence of lead.

All of these steels are susceptible to solidification cracking because of the high sulfur and phosphorus contents. High sulfur contents can also cause weld metal porosity. Lead can cause weld metal porosity or embrittlement, but the main concern is toxic fumes. Special precautions are required to ensure good ventilation during the welding of leaded steels.

When free-machining steels must be welded, a low-hydrogen consumable with a basic slag system is recommended. Low hydrogen will reduce the formation of hydrogen sulfide porosity, while the basic slag system reduces the sulfur and phosphorus content of the weld metal. Shielded metal arc welding electrodes such as E7015, E7016, and E7018 are preferred because they are the basic electrodes that produce the least base plate dilution. For gas-shielded flux-cored electrodes, E70T-5 and E71T-5 types are preferred over T-1 electrodes. Self-shielded flux-cored electrodes of the T-4, T-7 and T-8 types are recommended. Low currents should be used with all electrodes in order to minimize base plate dilution.

There is no guarantee that these precautions will eliminate the problem. Sometimes the best that can be achieved is a reduced level of cracking and/or porosity. The best advice is not to weld free-machining steel at all. Welding specifications can state that no component in a welded assembly should be produced from these materials. Overall, this will lead to the lowest total cost for the welded assembly.

Lamellar Tearing

Lamellar tearing is a form of cracking that occurs in the base metal beneath the weld (Fig. 13). It is caused by a combination of high, localized stresses from weld contraction and poor through-thickness ductility in the steel. The tearing is initiated by the separation of the interface between inclusions and metal (delamination) or by fracture of an inclusion itself. The cracks enlarge by linkage of delaminations in the same plane and by shear steps that join cracks on different planes. The controlling factors in lamellar tearing are:

- QUANTITY AND MORPHOLOGY OF NONMETALLIC INCLUSIONS (USUALLY SULFIDES OR SILICATES)
- MAGNITUDE OF THE INDUCED STRESSES NORMAL TO THE PLATE SURFACE

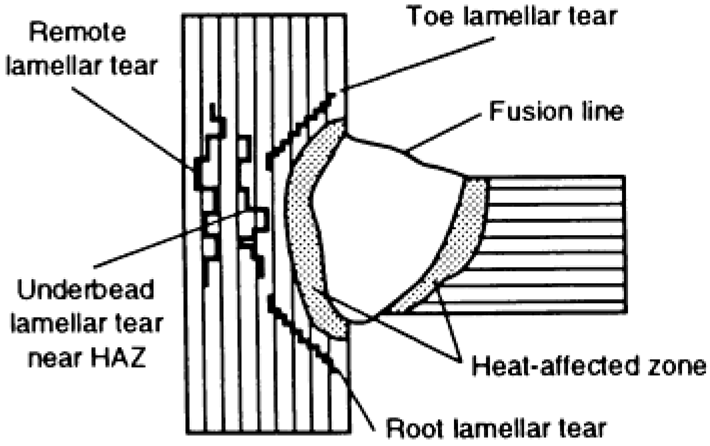

FIG. 13 ORIENTATION OF LAMELLAR TEARS IN A BASE METAL BELOW THE WELD. SOURCE: REF 4

During the hot rolling of plate, the inclusions become elongated in the rolling direction and tend to flatten out parallel to the surface of the plate. The ductility in the through-thickness direction is reduced because the bond between the inclusions and steel is much weaker than the steel itself. If the quantity of inclusions is high and if they have been sufficiently flattened during hot rolling, the steel may fail under high stresses in the thickness direction.

The magnitude of the induced stresses depends on the joint design, the degree of restraint imposed on the joint, the plate thickness, the orientation of the weld, and the size of the weld. Restrained corner or T-joints are the most susceptible to lamellar tearing because the fusion boundary is roughly parallel to the plate surface (Fig. 14) and through-thickness contraction stresses are high. Lamellar tearing in butt welds is rare. Thick-plate, high-imposed restraint (for example, rigid clamping) and large weld beads all increase residual stresses and the likelihood of tearing.

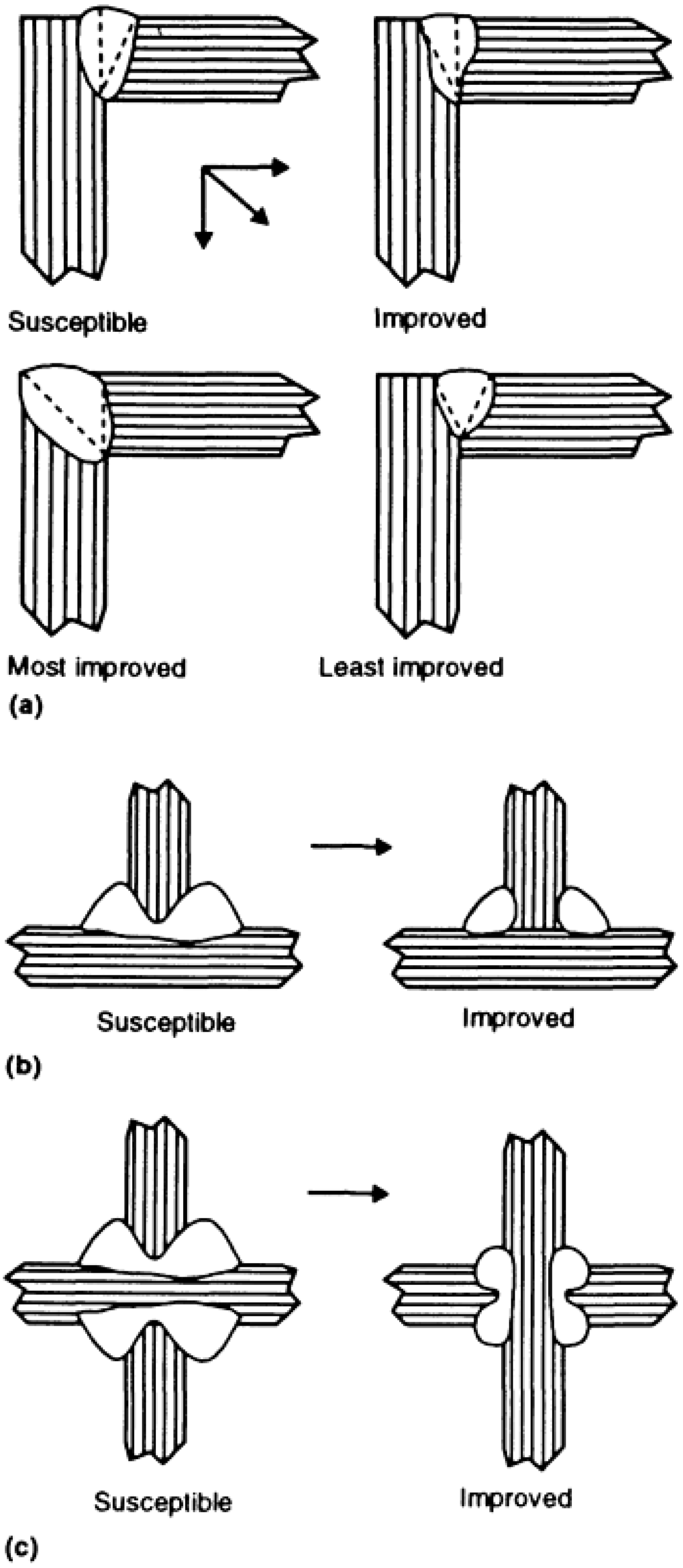
Susceptible
Improved
Most improved
Least improved
(a)
Susceptible
Improved
(b)
Susceptible
Improved
(c)

FIG. 14 METHODS USED TO REDUCE LAMELLAR TEARING. (A) ANGLING THE WELD FUSION LINE TO AVOID SHRINKAGE STRESSES IN THE THROUGH-THICKNESS DIRECTION. (B) USE OF SMALLER PARTIAL-PENETRATION WELDS TO REDUCE JOINT RESTRAINT. (C) PLACEMENT OF WELDING BEADS ON THE THINNER PLATE (WHEN WELDING PLATES OF DIFFERENT THICKNESSES) SO THAT SMALLER BEADS CAN BE USED. SOURCE: REF 4

Hydrogen has also been found to increase susceptibility to lamellar tearing. For this reason, a moderate preheat is sometimes beneficial in reducing the tendency to lamellar tearing. The use of low-hydrogen consumables is also beneficial.

Steps to avoid lamellar tearing include:

- USE OF PLATE MATERIAL WITH IMPROVED THROUGH-THICKNESS PROPERTIES. A REDUCTION IN OXYGEN CONTENT (BY ALUMINUM KILLING, FOR EXAMPLE) OR A REDUCTION IN SULFUR CONTENT WILL REDUCE THE AMOUNT OF INCLUSIONS IN THE STEEL. SULFUR UNDER 0.01% WILL GREATLY REDUCE THE RISK OF LAMELLAR TEARING. SULFIDE SHAPE CONTROL, THROUGH THE ADDITION OF RARE EARTHS OR CALCIUM TREATMENT DURING STEELMAKING, WILL ALTER THE INCLUSION MORPHOLOGY FROM STRINGERS TO DISCONTINUOUS GLOBULES.
- CHANGE THE JOINT DESIGN TO MINIMIZE THROUGH-THICKNESS CONTRACTION STRESSES (SEE FIG. 14 FOR EXAMPLES).
- USE LOWER-STRENGTH WELDING CONSUMABLES. THIS WILL ABSORB AND REDUCE SOME OF THE THROUGH-THICKNESS STRESS.
- IN EXTREME CASES, THE SURFACE OF THE PLATE MAY BE GROUND OR MACHINED TO A LEVEL BELOW WHERE LAMELLAR TEARING IS ANTICIPATED. THE AREA CAN THEN BE BUTTERED WITH WELD METAL AND THE ATTACHMENT WELD CAN BE MADE.

Porosity

Porosity in carbon steel weld metal may result from excessive amounts of carbon monoxide, nitrogen, hydrogen, or hydrogen sulfide remaining in the weld pool upon solidification.

Carbon monoxide is formed when oxygen combines with carbon in the weld puddle. The oxygen may come from the arc atmosphere, the welding consumables (especially fluxes), or the steel base material. Rimmed and capped steels contain low silicon (<0.05%) and high oxygen. When these types of steels are welded, the carbon and oxygen from the steel will react to form carbon monoxide, unless there is sufficient deoxidizer in the weld metal to prevent it. This deoxidizer (usually silicon) must come from the welding consumable. Higher levels of deoxidizer in the welding electrode are required when welding rimmed and capped steels. A killed steel is fully deoxidized during manufacture by the addition of silicon and/or aluminum or by vacuum treatment. Because it is the least costly, silicon (0.15 to 0.30%) is often used, although aluminum killing is becoming much more common. Killed steels are much less prone to carbon monoxide porosity, and consumables with a lower silicon level may often be used to weld them. A semikilled steel (0.05 to 0.10% Si) has a behavior that is intermediate to the rimmed and killed steels. It is important that the fabricator realize what deoxidation practice was used on the base material, especially for processes like autogenous gas-tungsten arc welding where no filler material is used.

The surface condition of the base material can also affect porosity levels. Millscale (iron oxide) or rust (hydrated iron oxide) adds additional oxygen to the weld puddle. Higher levels of deoxidizers in the consumables will be needed to prevent carbon monoxide porosity.

Nitrogen in the weld metal is most often the result of poor shielding. The amount of pickup will be increased by the use of excessive voltage (longer arc length), excessive travel speeds (effectively outrunning the shielding), broken electrode coatings, or unstable arc conditions (such as arc blow). A nitrogen-containing residue may be left on the surface of plasma cut plate when nitrogen is used as a cutting gas. Grinding of the surface may be necessary to avoid weld metal porosity.

Hydrogen has numerous sources, including atmospheric moisture, welding consumables (especially fluxes), and plate contaminants. The major source of hydrogen is the welding consumable. Hydrogen sources and pickup were discussed more fully in the section "Hydrogen-Induced Cracking" in this article.

Hydrogen sulfide is formed when high-hydrogen weld metal combines with high-sulfur (>0.04%) base plate. This type of porosity was discussed in the section "Welding of Free-Machining Steels" in this article.

There are many potential causes of porosity, too numerous to be covered in this article. Detailed information is available in the Section "Fusion Welding Processes" and the article "Overview of Weld Discontinuities" in this Volume.

References cited in this section

1. B.A. GRAVILLE, *THE PRINCIPLES OF COLD CRACKING CONTROL IN WELDS,* DOMINION BRIDGE COMPANY, 1975
2. H. GRANJON, *FUNDAMENTALS OF WELDING METALLURGY,* ABINGTON PUBLISHING, 1991, P 156-177
3. STRUCTURAL WELDING CODE, STEEL, AWS D1.1, AMERICAN WELDING SOCIETY, 1992
4. *WELDED STEEL CONSTRUCTION (METAL ARC WELDING),* CSA W59, CANADIAN STANDARDS ASSOCIATION, 1989
5. R.D. STOUT, *WELDABILITY OF STEELS,* 4TH ED., WELDING RESEARCH COUNCIL, 1987, P 420-426
6. *PROCESS OF ARC WELDING CARBON & CARBON MANGANESE STEELS,* BS5135, BRITISH STANDARDS INSTITUTE, 1984
7. F.R. COE, *WELDING STEELS WITHOUT HYDROGEN CRACKING,* THE WELDING INSTITUTE, 1973
8. J.F. LANCASTER, *THE METALLURGY OF WELDING,* 4TH ED., ALLEN AND UNWIN, 1987, P 157
9. *THE PROCEDURE HANDBOOK OF ARC WELDING,* LINCOLN ELECTRIC COMPANY, 1973, CHAPTERS 5 AND 12
10. W.F. SAVAGE AND A.H. ARONSON, *PREFERRED ORIENTATION IN THE WELD FUSION ZONE,* VOL 46, 1966
11. *WELDING HANDBOOK,* 8TH ED., VOL 1, AMERICAN WELDING SOCIETY, 1987, P 261

Arc Welding of Carbon Steels

Ronald B. Smith, The ESAB Group, Inc.

Process Selection

All of the arc welding processes are suitable for welding carbon steels. Choosing the best process to use for a particular application must take into account not only the material characteristics (sensitivity to porosity or cracking and the required mechanical properties), but also the details of the joint (plate thickness, joint design, welding position, and location) and weld economics (deposition rates and efficiencies, cost of labor, cost of consumables, capital cost of equipment, number of spare parts required, operator skills or training needed, set-up, and cleanup cost, and so on). There are many excellent reviews (Ref 9, 11, 12, 13) that discuss process selection and cost calculation details. Table 5 provides a brief guide that can be used to select a process for welding carbon steels.

TABLE 5 PROCESS SELECTION GUIDELINES FOR ARC WELDING CARBON STEELS

PARAMETERS	**SMAW**	**GMAW**	**FCAW**	**GTAW/PAW**	**SAW**	**EGW/ESW**	SW
USABILITY	VERY ADAPTABLE, ALL-POSITION PROCESS. CAN BE USED OUTDOORS. GIVES EXCELLENT JOINT ACCESSIBILITY. VERY PORTABLE. CAN BE USED ON CARBON STEELS DOWN TO 18 GAGE. JOINT PREPARATION IS REQUIRED ON THICKNESSES OVER 3.2 MM ($\frac{1}{8}$ IN.). UNLIMITED UPPER THICKNESS BUT OTHER PROCESSES (GMAW, FCAW, OR SAW) ARE USUALLY MORE ECONOMICAL.	AN ALL-POSITION PROCESS IN THE SHORT-ARC OR PULSED MODE. MODERATELY ADAPTABLE, BUT USE LIMITED OUTSIDE WHERE LOSS OF SHIELDING IS POSSIBLE. USABLE ON STEEL DOWN TO <0.25 MM (0.010 IN.) THICK. ABOVE 4.8 MM ($\frac{3}{16}$ IN.), REQUIRES JOINT PREPARATION. NO UPPER LIMIT OF PLATE THICKNESS.	ALL-POSITION PROCESS. EQUIPMENT SIMILAR TO GMAW, BUT SELF-SHIELDED VERSION HAS BETTER PORTABILITY AND IS USABLE OUTDOORS. MINIMUM PLATE THICKNESS IS 18 GAGE. FOR SELF-SHIELDED MATERIAL ABOVE 6.4 MM ($\frac{1}{4}$ IN.), REQUIRES JOINT PREPARATION. WITH CO_2 SHIELDING, MATERIALS ABOVE 13 MM ($\frac{1}{2}$ IN.) REQUIRE JOINT PREPARATION. THERE IS NO UPPER LIMIT ON PLATE THICKNESS.	AN ALL-POSITION PROCESS WITH FAIR JOINT ACCESSIBILITY. LIMITED OUTDOOR USE. CAN WELD VERY THIN MATERIAL IN ALL POSITIONS; GTAW--<0.13 MM (0.005 IN.) MIN. THICKNESS, JOINT PREP. OVER 3.2 MM ($\frac{1}{8}$ IN.) PAW--<0.13 MM (0.005 IN.) MIN. THICKNESS, JOINT PREP OVER 6.4 MM ($\frac{1}{4}$ IN.). NO UPPER LIMIT ON THICKNESS BUT OTHER PROCESSES ARE MORE EFFICIENT.	LIMITED TO FLAT AND HORIZONTAL POSITION. SEMI-AUTOMATIC VERSION HAS SOME ADAPTABILITY BUT PROCESS IS MOST OFTEN MECHANIZED AND HAS LIMITED PORTABILITY. MINIMUM MATERIAL THICKNESS IS 1.6 MM ($\frac{1}{16}$ IN.) (MECHANIZED). JOINT PREPARATION IS REQUIRED ON MATERIAL ABOVE 13 MM ($\frac{1}{2}$ IN.) THICK. THE PROCESS LENDS ITSELF TO WELDING THICK MATERIALS.	LIMITED TO VERTICAL OR NEAR-VERTICAL. O.K. OUTSIDE. FAIR PORTABILITY. EGW USED ON PLATE 9.5 TO 102 MM ($\frac{3}{8}$ TO 4 IN.) THICK. ESW USED ON PLATE 19 TO 900 MM ($\frac{3}{4}$ TO 36 IN.) HICK.	**ALL-POSITION. O.K. OUTSIDE. FAIR PORTABILITY. USED ON PLATE 1.0 MM (0.04 IN.) AND ABOVE. NO JOINT PREPARATION REQUIRED.**
COST FACTORS	A LOW-DEPOSITION-RATE PROCESS (UP TO 9 KG/H, OR 20 LB/H) WITH LOW DEPOSIT EFFICIENCY (TYPICALLY 65%).	DEPOSITION RATES (TO 16 KG/H, OR 35 LB/H) ARE HIGHER THAN SMAW. DEPOSITION EFFICIENCY (90-	DEPOSITION RATES (UP TO 18 KG/H, OR 40 LB/H) ARE HIGHER THAN GMAW. DEPOSITION EFFICIENCIES (80-90%) ARE LOWER	LOW DEPOSITION RATES (UP TO 5.4 KG/H, OR 12 LB/H) BUT HIGH DEPOSITION EFFICIENCY (99%). IF AUTOMATED,	DEPOSITION RATES ARE VERY HIGH (OVER 45 KG/H, OR 100 LB/H WITH MULTIWIRE SYSTEMS). DEPOSITION EFFICIENCY IS 99%	HIGH JOINT COMPLETION RATE. HIGH DEPOSITION EFFICIENCY (SOLID WIRES--99%, CORED WIRES LOWER).	**EQUIPMENT COST IS MODERATE.**

	LOW OPERATOR FACTOR. EQUIPMENT COST IS LOW AND SPARE PARTS ARE MINIMAL. WELDING SPEEDS ARE GENERALLY LOW. HOUSEKEEPING IS REQUIRED TO DESLAG AND DISPOSE OF FLUX AND ELECTRODE STUBS.	95%) AND OPERATOR FACTOR (TYPICAL 50%) ARE ALSO HIGHER. EQUIPMENT AND SPARE PARTS COST ARE MODERATE TO HIGH (PULSED-ARC POWER SUPPLIES ARE HIGHER COST). WELDING SPEEDS ARE MODERATE TO HIGH (BURIED ARC CO_2 CAN WELD AT OVER 2540 MM/MIN (100 IN./MIN). CLEANUP IS MINIMAL.	BUT OPERATOR FACTORS (50%) ARE SIMILAR TO GMAW. EQUIPMENT COST IS MODERATE. GOOD OUT-OF-POSITION DEPOSITION RATES CAN BE ACHIEVED WITH CONVENTIONAL POWER SOURCES. WELDING SPEEDS ARE MODERATE TO HIGH. SLAG AND SPATTER REMOVAL AND DISPOSAL IS REQUIRED.	THE OPERATOR FACTOR IS HIGH. EQUIPMENT AND SPARE PARTS COST IS MODERATE TO HIGH, DEPENDING ON COMPLEXITY (PAW IS SLIGHTLY MORE EXPENSIVE). CLEANUP IS GENERALLY UNNECESSARY DUE TO NO FLUX OR SPATTER.	BUT THAT DOES NOT INCLUDE THE FLUX (A 1:1 FLUX-TO-WIRE RATIO IS COMMON). BECAUSE THE PROCESS IS USUALLY MECHANIZED, OPERATOR FACTORS ARE HIGH. COST IS MODERATE FOR SINGLE-WIRE SYSTEMS. HIGH WELDING SPEEDS ARE ACHIEVABLE. HIGHER HOUSEKEEPING COSTS RESULT FROM THE NEED TO HANDLE THE SLAG AND UNFUSED FLUX.	HIGH OPERATOR FACTOR. HIGH EQUIPMENT AND SET-UP COST.	
WELD METAL QUALITY	STRONGLY DEPENDENT ON THE SKILL OF THE WELDER. LACK OF FUSION OR SLAG INCLUSIONS ARE POTENTIAL PROBLEMS. THE RELATIVELY SMALL BEADS USUALLY RESULT IN A HIGH PERCENTAGE OF REFINING IN MULTIPASS WELDS, AND VERY GOOD TOUGHNESS IS ACHIEVABLE WITH SOME	VERY GOOD QUALITY. POROSITY OR LACK OF FUSION CAN BE A PROBLEM. LESS TOLERANT OF RUST AND MILLSCALE THAN FLUX-USING PROCESSES. VERY GOOD TOUGHNESS IS ACHIEVABLE.	GOOD QUALITY. WELD METAL TOUGHNESS IS FAIR TO GOOD (BEST WITH BASIC ELECTRODES). SLAG INCLUSIONS ARE A POTENTIAL PROBLEM.	HIGH QUALITY, BUT REQUIRES CLEAN PLATE. EXCELLENT TOUGHNESS POSSIBLE DUE TO SMALL BEAD SIZE (HIGH REFINED WELD METAL, FINER GRAIN SIZE).	VERY GOOD. GOOD WELD METAL TOUGHNESS IS POSSIBLE. HANDLES RUST AND MILLSCALE WELL WITH PROPER FLUX SELECTION. HIGH-DILUTION PROCESS.	LARGE COLUMNAR GRAIN STRUCTURE RESULTS IN ONLY FAIR TOUGHNESS. HIGH DILUTION OF BASEPLATE (UP TO 35% FOR EGW, UP TO 50% FOR ESW).	**MINIMAL WELD SIZE. USUALLY FINE-GRAINED DUE TO FAST COOLING. WELD STRUCTURE IS AS CAST.**

	ELECTRODES.						
EFFECT ON BASE METAL	LOW HEAT INPUTS CAN CAUSE RAPID HAZ COOLING. FLUX COATINGS ARE A POTENTIAL SOURCE OF HYDROGEN.	GENERALLY A LOW-HYDROGEN PROCESS.	FLUX CORE CAN CONTRIBUTE HYDROGEN.	A LOW-HYDROGEN PROCESS. LOW HEAT INPUTS CAN CAUSE RAPID HAZ COOLING.	HIGHER HEAT INPUTS CAN RESULT IN LARGE HAZ AND POSSIBLE DETERIORATION OF BASEPLATE PROPERTIES (ESPECIALLY QUENCHED AND TEMPERED PLATE). FLUX IS A SOURCE OF HYDROGEN.	HIGH HEAT INPUT RESULTS IN LARGE HAZ AND GRAIN GROWTH. NOT RECOMMENDED FOR QUENCHED AND TEMPERED STEELS UNLESS REHEAT TREATMENT IS TO BE DONE. SLOW WELD COOLING ALLOWS HYDROGEN TO ESCAPE WELD AREA.	**CAUSES RAPID COOLING OF HAZ.**
GENERAL COMMENTS	**VERY VERSATILE, LOW-COST PROCESS. ESPECIALLY STRONG ON NON-ROUTINE OR REPAIR JOBS. USUALLY NOT ECONOMICAL ON STANDARD THICK WELDS OR ON REPETITIVE JOBS THAT CAN BE MECHANIZED.**	**RELATIVELY VERSATILE. EQUIPMENT MORE COSTLY, COMPLEX, AND LESS PORTABLE THAN SMAW. EASILY MECHANIZED. A CLEAN PROCESS WITH HIGHER DEPOSITION RATES AND EFFICIENCIES THAN SMAW.**	**RELATIVELY VERSATILE (SELF-SHIELDED VERSION CAN BE USED OUTSIDE). HIGHER DEPOSITION RATES THAN GMAW (HIGHER FUMES ALSO). WELDS EASILY OUT-OF-POSITION. READILY MECHANIZED.**	**EXTREMELY GOOD WELDS IN ALL POSITIONS. REQUIRES MORE PREPARATION. LOW DEPOSITION RATES LIMIT PROCESS TO THINNER PLATES.**	**A HIGH-DEPOSITION, HIGH-PENETRATION PROCESS, BUT THIN MATERIAL CAN BE WELDED AT HIGH SPEEDS. EASILY MECHANIZED. NATURAL FOR WELDING THICKER PLATES. HOUSEKEEPING AND POSITION LIMITATION CAN BE A PROBLEM.**	**HIGH JOINT COMPLETION RATE, BUT LIMITED TO VERTICAL. USED MAINLY ON LONG VERTICAL WELDS. HIGH HEAT INPUT CAN LIMIT BASEPLATES.**	**SPECIALIZED PROCESS FOR ATTACHING STUDS TO STEEL PLATE. USUALLY MORE ECONOMICAL THAN DRILLING AND TAPPING.**

SMAW, shielded metal arc welding; GMAW, gas-metal arc welding; FCAW, flux-cored arc welding; GTAW, gas-tungsten arc welding; PAW, plasma arc welding; SAW, submerged arc welding; EGW, electrogas welding; ESW, electroslag welding; SAW, stud arc welding

The effect of the welding process on carbon steel base metal is a function of the heat input (and cooling rate) imposed by the process and the amount of hydrogen contributed. Low-heat-input processes, such as gas-tungsten arc welding (GTAW), plasma arc welding (PAW), shielded metal arc welding (SMAW), short-arc gas-metal arc welding (GMAW), and stud welding (SW), result in a narrow, rapidly cooled HAZ. Greater care needs to be exercised when using these processes to weld hardenable carbon steels. On the other hand, a higher-heat-input process, like electrogas welding (EGW), will produce a much wider HAZ, produce more grain coarsening, and may cause a deterioration in the properties of heat-treated carbon steels. Heat input limits or a process change may be necessary when welding quenched and tempered carbon steels.

The arc welding processes that employ flux or flux-containing consumables contribute hydrogen to the weld area if the consumables are not adequately dried or protected from moisture. Examples are SMAW, flux-cored arc welding (FCAW), and submerged arc welding (SAW). Each of these processes offers low-hydrogen consumable types that should be used when welding hardenable carbon steels. Procedures are available, or will soon be available, in AWS electrode specifications to allow classification of arc welding electrodes based upon the weld metal diffusible hydrogen produced. This is a safe approach to use when welding hardenable carbon steels, because there can be wide variations between electrodes, even in the so-called "low-hydrogen" processes.

Weld metal properties as a function of process are a bit more complex. Not only are the heat input and hydrogen contributions important, but also the amount of base metal dilution, rust and scale tolerance, and pickup of atmospheric gases. The high-heat-input processes, particularly EGW, and electroslag welding (ESW) will result in large columnar grains and resulting lower toughness. The lower-heat-input processes (particularly GTAW, PAW, and SMAW) produce finer grain size. For a given weld chemistry, the higher-heat-input processes will produce a weaker weld deposit because of the formation of coarser, high-temperature transformation products. Hydrogen-induced cracking is generally not as serious a problem in weld metal as it is in the HAZ, because weld metal is usually low in carbon. Excessive carbon pickup (by dilution) or the use of an alloyed electrode would make the weld metal more vulnerable.

High dilution processes (for example, SAW and EGW) will incorporate a large amount of base metal into the weld. In square-butt submerged arc welds, over 60% of the weld metal can be diluted base metal. Obviously, this has a major effect on the properties of the weld metal. It also illustrates the importance of conducting procedure qualification testing to determine the actual mechanical and chemical properties of the production joint. The pickup of impurities (for example, sulfur, phosphorus, or nitrogen) from the base plate could lead to problems such as solidification cracking, porosity, or toughness deterioration.

The tolerance of the process to plate contaminants, such as rust, scale, or organic compounds (oil, grease, and so on) can be important. The nonflux processes are generally less tolerant of rust and scale. Processes such as GTAW, GMAW, and PAW are particularly sensitive and require clean plate. The submerged arc process is sensitive to porosity caused by organic compounds. In the open arc processes, the compounds can be vaporized by the heat of the arc before steel melting occurs. In submerged arc welding, the flux blanket acts to block the radiant heat from the arc, and the compound becomes incorporated in the weld metal.

Pickup of nitrogen and oxygen can also cause problems. Porosity is one outcome; toughness deterioration is another. Figure 15 shows the levels of weld metal oxygen and nitrogen produced by several arc welding processes. The GTAW and PAW processes produce the lowest gaseous impurity level. The arc in both of these processes is short and extremely stable, and the shielding gases (for example, argon or argon-helium mixtures) are inert and make no contribution. GMAW deposits are higher in both oxygen and nitrogen. A longer, less stable arc allows a greater chance of pickup. The use of Ar-O_2, Ar-CO_2, or CO_2 shielding gases also raises the oxygen level of the weld metal.

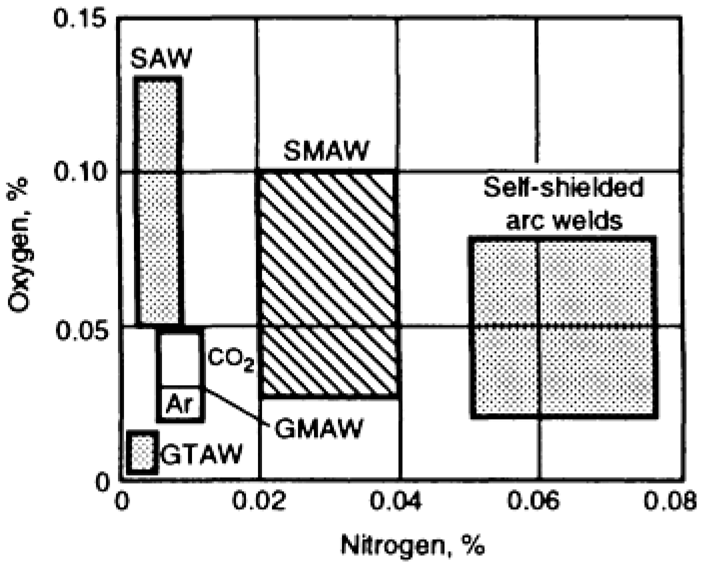

FIG. 15 TYPICAL OXYGEN AND NITROGEN LEVELS OBTAINED IN WELD METAL FROM SELECTED ARC WELDING PROCESSES. SOURCE: REF 14

The flux-using processes (that is, SMAW, FCAW, and SAW) can all pick up oxygen from decomposition of metal oxides contained within the flux. More basic flux systems produce lower oxygen levels. The two open arc, self-shielded processes (that is, SMAW and self-shielded FCAW) do not produce an even, well-directed gas shield. As a result, aspiration of air is higher than with the externally gas-shielded processes. Strong denitriders (for example, aluminum) are added to self-shielded flux-cored wires to counteract this pickup. Some submerged arc welds are more prone to nitrogen pickup when alternating current (ac) power is used.

All of the arc welding processes are susceptible to aspiration of air when long arc lengths (excessive arc voltages) are used. A longer arc length increases the time of exposure of the molten drops on transfer and increases the likelihood of gaseous impurity pickup.

References cited in this section

9. *THE PROCEDURE HANDBOOK OF ARC WELDING,* LINCOLN ELECTRIC COMPANY, 1973, CHAPTERS 5 AND 12
11. *WELDING HANDBOOK,* 8TH ED., VOL 1, AMERICAN WELDING SOCIETY, 1987, P 261
12. H.B. CARY, *MODERN WELDING TECHNOLOGY,* 2ND ED., PRENTICE-HALL, 1989, P 204-206
13. P.T. HOULDCROFT, *WHICH PROCESS?,* ABINGTON PUBLISHING, 1990
14. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987, P 63

Arc Welding of Carbon Steels

Ronald B. Smith, The ESAB Group, Inc.

Welding Consumable Selection and Procedure Development

When choosing a welding consumable, the following should be considered:

- MECHANICAL PROPERTIES REQUIRED

- WELDABILITY CONSIDERATIONS
- PERFORMANCE FEATURES NEEDED OR DESIRED

Codes or customer specifications may define the mechanical properties required. Some codes will specify the types of electrodes to be used, while others limit the selection to certain products from certain manufacturers who have performed tests to qualify their products. If this information is not available, then it is customary to try to match or slightly exceed the base plate strength. There may be conditions where an undermatching weld metal strength is actually preferred--this is a function of the weld design and type of loading. The toughness level and other properties (for example, hardness) are dictated by service conditions and are more difficult to define. Code requirements can be used as guidelines.

Weldability considerations (that is, potential cracking or porosity problems in carbon steel) should then be assessed. A medium-carbon steel, for instance, will require a low-hydrogen consumable, preheat, and perhaps a postweld heat treatment to avoid hydrogen-induced cracking. The effect of this treatment on candidate consumables should be considered.

Certain performance features may be necessary, or at least desired. They allow the welding to be performed more economically. These types of choices are more frequently made with the flux-using processes where the formulations are designed to do certain things (for example, weld out-of-position, at high travel speeds, or at high deposition rates). Even nonflux processes require choices. For example, for GTAW of very thin steel tubing an ER70S-2 electrode (which contains some aluminum) produces a more viscous puddle that is easier to control out-of-position.

Once a consumable is chosen, a suitable process and procedure must be specified. American Society for Quality Control (ASQC) Standard Q91 (or ISO 9001) treats welding as a special process because the results cannot be fully verified by subsequent inspection and testing of the product. Therefore, documented work instructions, monitoring and control of process, approval of process and equipment, and criteria for workmanship shall be stipulated. In addition, ASQC Standard Q94 requires verifying the results of this special processing and determining to what extent equipment, environments, and operators are able to affect these final desired results.

In welding, the documented work instruction is usually referred to as a welding procedure specification (WPS) or just welding procedure (Fig. 16). In some fabrication specifications (for example, AWS D1.1), the WPS can be prequalified if it is specified within certain guidelines, such as equipment settings and methods. In other fabrication specifications, the WPS must be qualified, usually with the rules of the qualification standard such as the ASME Boiler and Presser Vessel Code Section IX or MIL STD 248. In all cases, a welding procedure specification that addresses all the important process variables must be documented. If these variables change, then a new WPS may have to be prepared to reflect these changes. Whether the WPS must be requalified when revised always depends on which variables are being revised. The qualification standard will identify which changes require requalification.

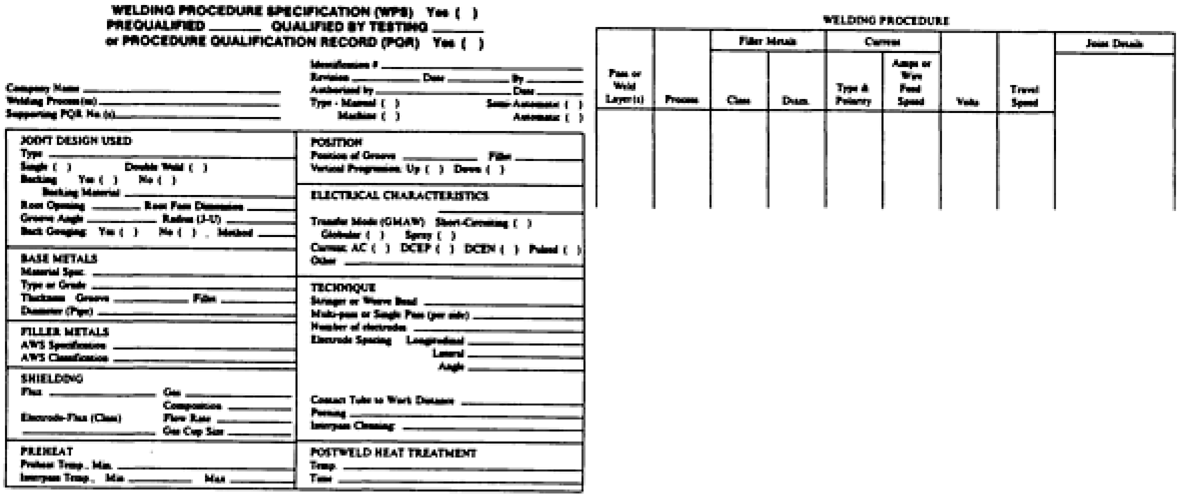

WELDING PROCEDURE SPECIFICATION (WPS) Yes ( )
PREQUALIFIED ______ QUALIFIED BY TESTING ______
or PROCEDURE QUALIFICATION RECORD (PQR) Yes ( )

Identification # ______
Revision ______ Date ______ By ______
Authorized by ______ Date ______
Type - Manual ( ) Semi-Automatic ( )
Machine ( ) Automatic ( )

Company Name ______
Welding Process(es) ______
Supporting PQR No.(s) ______

JOINT DESIGN USED
Type ______
Single ( ) Double Weld ( )
Backing Yes ( ) No ( )
Backing Material ______
Root Opening ______ Root Face Dimension ______
Groove Angle ______ Radius (J-U) ______
Back Gouging: Yes ( ) No ( ) Method ______

BASE METALS
Material Spec. ______
Type or Grade ______
Thickness: Groove ______ Fillet ______
Diameter (Pipe) ______

FILLER METALS
AWS Specification ______
AWS Classification ______

SHIELDING
Flux ______ Gas ______
Composition ______
Electrode-Flux (Class) Flow Rate ______
______ Gas Cup Size ______

PREHEAT
Preheat Temp., Min. ______
Interpass Temp., Min ______ Max ______

POSITION
Position of Groove ______ Fillet ______
Vertical Progression: Up ( ) Down ( )

ELECTRICAL CHARACTERISTICS
______
Transfer Mode (GMAW) Short-Circuiting ( )
Globular ( ) Spray ( )
Current: AC ( ) DCEP ( ) DCEN ( ) Pulsed ( )
Other ______

TECHNIQUE
Stringer or Weave Bead ______
Multi-pass or Single Pass (per side) ______
Number of electrodes ______
Electrode Spacing Longitudinal ______
Lateral ______
Angle ______
Contact Tube to Work Distance ______
Peening ______
Interpass Cleaning: ______

POSTWELD HEAT TREATMENT
Temp. ______
Time ______

WELDING PROCEDURE

Pass or Weld Layer(s)	Process	Filler Metals		Current		Volts	Travel Speed	Joint Details
		Class	Diam.	Type & Polarity	Amps or Wire Feed Speed			

FIG. 16 WELDING PROCEDURE SPECIFICATION FORM. SOURCE: REF 3

All welding procedure specifications must be qualified by actual mechanical testing of welds made in accordance with the WPS. If this testing fails to meet the requirements of a customer or code, then a new WPS must be prepared and tested. Detailed information about procedure qualification is available in the article "Weld Procedure Qualification" in this Volume and in Ref 15.

Once a welding procedure specification is prepared, then the welding operators who will be applying this process should establish through testing that they are proficient in its application. This testing of worker proficiency is usually termed a welder or welding operator performance qualification test. Again, the qualification specification will outline the guidelines in which the welder or welding operator can be qualified with each test. As long as the worker continues to perform within those guidelines on a periodic basis without significant faults, the worker's qualification remains in effect. If the worker does not use the process for an extended period of time, uses it outside the guideline limits for performance qualification, or fails to achieve acceptable quality, then disqualification may occur with further training and testing as a corrective action, if deemed necessary. Repeated disqualification may necessitate reassignment to a less-skilled job. Adequate records are required to show continuity of process application for each qualified worker within the certifying organization.

Shielded Metal Arc Welding

Shielded metal arc welding is widely used for joining all classes of carbon steels. Detailed information is available in the article "Shielded Metal Arc Welding" in this Volume.

A wide variety of electrode types exist, each designed to work in a specific way and each possessing unique characteristics. Proper electrode selection is essential to achieve good welding performance and weld quality. AWS A5.1 (covering carbon steel electrodes) (Ref 16) and A5.5 (covering low-alloy steel electrodes) (Ref 17) are widely used to classify SMAW electrodes suitable for welding carbon steels. The low-alloy steel versions are used to weld higher carbon steels when higher strength is required.

E60XX electrodes (415 MPa, or 60 ksi, minimum tensile strength) are suitable for welding carbon steels containing up to about 0.3% C. They are not made with low-hydrogen coatings and should not be used to weld hardenable steels, especially under conditions of high joint restraint.

E70XX electrodes (485 MPa, or 70 ksi, minimum tensile strength) are suitable for steels with a wide range of strength values. The E7015, E7016, E7018, E7018M, E7028, and E7048 types are low-hydrogen electrodes and may be used in situations where hydrogen-induced cracking is a possibility. Table 6 summarizes some of the features and characteristics of the different carbon steel electrode types (see the Section "Fusion Welding Procedures" in this Volume or the appendices of Ref 16 and 17 for a more thorough discussion).

TABLE 6 CARBON STEEL ELECTRODES FOR SMAW WELDING APPLICATIONS

ELECTRODE	**COATING TYPE**	**POSITION(A)**	**CURRENT TYPE**	**MINIMUM IMPACT ENERGY (CHARPY V-NOTCH)**		**GENERAL COMMENTS**	APPLICATIONS
				J	**J · LBF**		
E6010	HIGH-CELLULOSE SODIUM	ALL	DCEP	27(B)	20(B)	GOOD PENETRATION. DIRECT FORCEFUL ARC. HIGH WELD METAL HYDROGEN.	**USED ON ALL TYPES OF JOINTS**
E6011	HIGH-CELLULOSE POTASSIUM	ALL	DCEP, AC	27(B)	20(B)	SIMILAR TO 6010 EXCEPT THAT IT HAS A MORE STABLE ARC AND IS SLIGHTLY LESS PENETRATING	**USED ON ALL TYPES OF JOINTS**
E6012	HIGH-RUTILE SODIUM	ALL	DCEN, AC	(C)	(C)	LOWER PENETRATION, HIGHER TRAVEL SPEEDS	**THIN SHEET METAL**
E6013	HIGH-RUTILE POTASSIUM	ALL	DCEN, DCEP, AC	(C)	(C)	HIGHER TRAVEL SPEEDS, LOWER PENETRATION. BETTER ON AC THAN E6012.	**THIN SHEET METAL**
E7014	RUTILE, IRON POWDER	ALL	DCEN, DCEP, AC	(C)	(C)	HIGHER-DEPOSITION-RATE VERSION OF E6013	**INCLINED AND SHORT FILLETS**
E6019	IRON OXIDE, RUTILE POTASSIUM	ALL	DCEN, DCEP, AC	27(D)	20(D)	MEDIUM-PENETRATION FLUID SLAG, ONLY SMALLER DIAMETER USED OUT-OF-POSITION (NO VERTICAL DOWN)	**MULTIPASS WELDS (TO 25 MM, OR 1 IN.), SOME OUT-OF-POSITION USE**
E6020	HIGH-IRON OXIDE	HF FLAT	DCEN, AC DCEN, DCEP, AC	(C)	(C)	SPRAY-LIKE ARC, HEAVY FLUID SLAG	**DOWNHAND WELDING OF THICKER PLATE**
E7024	RUTILE, IRON POWDER	HF, FLAT	DCEN, DCEP, AC	(C)(E)	(C)(E)	VERY HIGH-DEPOSITION-RATE VERSION OF E7014. THICK COATING OF FLUX. GOOD ARC CHARACTERISTICS.	**FILLETS AT HIGH TRAVEL SPEEDS. GRAVITY WELDING.**
E6027	HIGH-IRON OXIDE, IRON POWDER	HF FLAT	DCEN, AC DCEN, DCEP, AC	27(B)	20(B)	MEDIUM PENETRATION. HIGH-DEPOSITION-RATE VERSION OF E6020.	**FLAT FILLETS OR GROOVE WELDS**
E7027	HIGH-IRON OXIDE, IRON POWDER	HF FLAT	DCEN, AC DCEN, DCEP, AC	27(B)	20(B)	HIGHER TENSILE VERSION OF 6027	**SAME AS 6027**
LOW-HYDROGEN ELECTRODES							
E7015	LOW-HYDROGEN SODIUM	ALL	DCEP	27(B)	20(B)	THE ORIGINAL LOW-HYDROGEN ELECTRODE, NOT WIDELY USED TODAY	**USED ON ALL TYPES OF JOINTS WHERE LOW HYDROGEN IS NEEDED**

E7016	LOW-HYDROGEN POTASSIUM	ALL	DCEP, AC	27(B)(F)	20(B)(F)	SIMILAR TO E7015, BUT A MORE STABLE ARC	**USED ON ALL TYPES OF JOINTS WHERE LOW HYDROGEN IS NEEDED**
E7018	LOW-HYDROGEN POTASSIUM, IRON POWDER	ALL	DCEP, AC	27(B)(G)	20(B)(G)	MODERN LOW-HYDROGEN ELECTRODE. BETTER ARC AND OPERATING CHARACTERISTICS.	**USED ON ALL TYPES OF JOINTS, ESPECIALLY WHEN LOW HYDROGEN IS NEEDED**
E7018M	LOW-HYDROGEN IRON POWDER	ALL	DCEP	68(B)	50(B)	A LOW-MOISTURE VERSION OF 7018	**JOINING CARBON STEELS TO HIGH-STRENGTH LOW-ALLOY STEELS**
E7028	LOW-HYDROGEN POTASSIUM, HIGH-IRON POWDER	FLAT, HF	DCEP, AC	27(D)	20(D)	E7018 WITH HIGH IRON POWDER, HIGH DEPOSITION RATES	**HIGH TRAVEL SPEEDS ON FILLETS AND GROOVE WELDS**
E7048	**LOW-HYDROGEN POTASSIUM, IRON POWDER**	**FLAT, OH, HORIZONTAL, V-DOWN**	**DCEP, AC**	**27(B)**	**20(B)**	**SIMILAR TO E7018 WITH EXCELLENT VERTICAL-DOWN WELDING CHARACTERISTICS**	**VERTICAL DOWN WELDING**

(A) HF, HORIZONTAL FLAT; OH, OVERHEAD; V, VERTICAL.

(B) AT -30 °C (-20 °F).

(C) NOT REQUIRED.

(D) AT - 20 °C (0 °F).

(E) 7024-1 SPECIFICATIONS: 27 J (20 FT · LBF) AT -20 °C (0 °F).

(F) 7016-1 SPECIFICATIONS: 27 J (20 FT · LBF) AT -45 °C (-50 °F).

(G) 7018-1 SPECIFICATIONS: 27 J (20 FT · LBF) AT -45 °C (-50 °F)

Low-Alloy Steel Electrodes. Most low-alloy steel electrodes have low-hydrogen coatings (EXX15, EXX16, and EXX18), but some may not. Because low-alloy steel electrodes are used to weld higher-carbon, more hardenable steels, it is recommended that only the low-hydrogen versions be used. Consideration should also be given to the hydrogen-induced cracking tendencies of the low-alloy steel weld metal, especially if it is heavily diluted with carbon from the baseplate. The low-hydrogen electrodes have been specially formulated and processed to give low coating moistures. It is important that they be stored and handled properly in order to avoid moisture absorption. Once removed from their storage cans, they should be held in a vented oven at 110 to 150 °C (230 to 300 °F). If the electrodes have absorbed excessive moisture, they can be restored by baking at 250 to 425 °C (480 to 800 °F) for 1 to 2 h (consult the manufacturer of the electrode for the exact temperature).

Because SMAW is a lower-heat-input process and cooling rates can be high, it is especially important to control hydrogen when welding higher-carbon steels. The newer revisions of the AWS specifications will have optional procedures for choosing electrodes on the basis of the weld metal diffusible hydrogen produced.

Gas-Metal Arc Welding

Detailed information about this process is in the article "Gas-Metal Arc Welding" in this Volume. The process is unique in that it operates in a number of different modes, each with its own distinct method of metal transfer, using different shielding gases, operating at different heat input levels, and having different penetration (dilution) characteristics. The resultant weldment properties may vary significantly, depending on the mode used. For this reason, it is important to be cautious about generalizations regarding the process and to perform procedure qualifications.

Electrodes for GMAW are available as solid-or metal-cored wires. AWS A5.18 (Ref 19) classifies carbon steel filler metals. AWS A5.28 (Ref 18) classifies low-alloy steel electrodes, used to weld medium- and high-carbon steels where a higher-strength deposit is needed. Compositions of GMAW electrodes are given in Table 7. Mechanical property requirements are given in Table 8.

TABLE 7 COMPOSITION REQUIREMENTS FOR GMAW ELECTRODES USED TO WELD CARBON STEELS

AWS CLASSIFICATION(A)	COMPOSITION, WT%												
	C	**Mn**	**Si**	**P**	**S**	**Ni**	**Cr**	**Mo**	**V**	**Cu(b)**	**Ti**	**Zr**	**Al**	TOTAL OTHER ELEMENTS(C)
ER70S-2	0.07	0.90-1.40	0.40-0.70	0.025	0.035	(D)	(D)	(D)	(D)	0.50	0.05-0.15	0.02-0.12	0.05-0.15	...
ER70S-3	0.06-0.15	0.90-1.40	0.45-0.70	0.025	0.035	(D)	(D)	(D)	(D)	0.50	...	...	...	...
ER70S-4	0.07-0.15	1.00-1.50	0.65-0.85	0.025	0.035	(D)	(D)	(D)	(D)	0.50	...	...	...	...
ER70S-5	0.07-0.19	0.90-1.40	0.30-0.60	0.025	0.035	(D)	(D)	(D)	(D)	0.50	...	...	0.50-0.90	...
ER70S-6	0.07-0.15	1.40-1.85	0.80-1.15	0.025	0.035	(D)	(D)	(D)	(D)	0.50	...	...	...	...
ER70S-7	0.07-0.15	1.50-2.00(E)	0.50-0.80	0.025	0.035	(D)	(D)	(D)	(D)	0.50	...	...	...	...
ER70S-G	**NO CHEMICAL REQUIREMENTS(F)**													
ER80S-NI1	...	...	...	...	...	0.80-1.10	0.15	0.35	0.05	...	...	...	...	...
ER80S-NI2	0.12	1.24	0.40-	0.025	0.025	2.0-2.75	...	...	...	0.35	...	...	...	**0.50**

			0.80											
ER80S-NI3	. . .	. . .	. . .	. . .	. . .	3.0-3.75	. . .	. . .	. . .	. . .	. . .	. . .	. . .	**. . .**
ER80S-D2	0.07-0.12	1.60-2.10	0.50-0.80	0.025	0.025	0.15	. . .	0.40-0.60	. . .	0.50	. . .	. . .	. . .	**0.50**
ER100S-1	0.08	1.25-1.30	0.20-0.50	0.010	0.010	1.40-2.10	0.30	0.25-0.55	0.05	0.25	. . .	. . .	. . .	**. . .**
ER100S-2	0.12	1.25-1.80	0.20-0.60	0.010	0.010	0.80-1.25	0.30	0.20-0.55	0.05	0.35-0.65	0.10	0.10	0.10	**0.50**
ER110S-1	0.09	1.40-1.80	0.20-0.55	0.010	0.020	1.90-2.60	0.50	0.25-0.55	0.04	0.25	. . .	. . .	. . .	**. . .**
ER120S-1	0.10	1.40-1.80	0.25-0.60	0.010	0.010	2.0-2.80	0.60	0.30-0.65	0.03	0.25	. . .	. . .	. . .	**. . .**
ERXXS-G	**SUBJECT TO AGREEMENT BETWEEN SUPPLIER AND PURCHASER**													

Note: Single values shown are maximums. Analysis shall be made for the elements for which specific values are shown in this table. If, however, the presence of other elements is indicated in the course of routine analysis, further analysis shall be made to determine that the total of these other elements, except iron, is not present in excess of the limits specified for "Total other elements" in the last column of this table.

Source: Ref 18

(A) ELECTRODES CLASSIFIED E70S-1B PREVIOUSLY ARE NOW CLASSIFIED ER80S-D2. THE SUFFIXES B2, NIL, AND SO ON, DESIGNATE THE CHEMICAL COMPOSITION OF THE ELECTRODE AND ROD CLASSIFICATION.

(B) THE MAXIMUM WEIGHT PERCENT OF COPPER IN THE ROD OR ELECTRODE DUE TO ANY COATING PLUS THE RESIDUAL COPPER CONTENT IN THE STEEL SHALL COMPLY WITH THE STATED VALUE.

(C) OTHER ELEMENTS, IF INTENTIONALLY ADDED, SHALL BE REPORTED.

(D) THESE ELEMENTS MAY BE PRESENT BUT ARE NOT INTENTIONALLY ADDED.

(E) IN THIS CLASSIFICATION, THE MAXIMUM MANGANESE MAY EXCEED 2.0%. IF IT DOES, THE MAXIMUM CARBON MUST BE REDUCED 0.01% FOR EACH 0.05% INCREASE IN MANGANESE OR PART THEREOF.

(F) FOR THIS CLASSIFICATION, THERE ARE NO CHEMICAL REQUIREMENTS FOR THE ELEMENTS LISTED, WITH THE EXCEPTION THAT THERE SHALL BE NO INTENTIONAL ADDITION OF NICKEL, CHROMIUM, MOLYBDENUM, OR VANADIUM.

(G) TO MEET THE REQUIREMENTS OF THE G CLASSIFICATION, THE ELECTRODE MUST HAVE AS A MINIMUM ONE OF EITHER 0.50% NI, 0.30% CR, OR 0.20% MO.

TABLE 8 MECHANICAL-PROPERTY REQUIREMENTS FOR WELD-METAL DEPOSITS OF GMAW ELECTRODES

AWS CLASSIFICATION	**SHIELDING GAS**	**CURRENT AND POLARITY**	**MINIMUM TENSILE STRENGTH**		**MINIMUM YIELD STRENGTH 0.2% OFFSET**		**ELONGATION IN 50 MM (2 IN.) MIN, %**	**CONDITION**	MINIMUM REQUIRED IMPACT PROPERTIES	
			MPA	**KSI**	**MPA**	**KSI**			**J**	FT · LBF
ER70S-2	CO_2(A)	DCEP	495(B)	72(B)	415(B)	60(B)	22	AS WELDED	27(F)	**20(F)**
ER70X-3	CO_2(A)	DCEP	495(B)	72(B)	415(B)	60(B)	22	AS WELDED	27(G)	**20(G)**
ER70S-4	CO_2(A)	DCEP	495(B)	72(B)	415(B)	60(B)	22	AS WELDED	(H)	**(H)**

ER70S-5	CO$_2$(A)	DCEP	495(B)	72(B)	415(B)	60(B)	22	AS WELDED	(H)	**(H)**
ER70S-6	CO$_2$(A)	DCEP	495(B)	72(B)	415(B)	60(B)	22	AS WELDED	27(F)	**20(F)**
ER70S-7	CO$_2$(A)	DCEP	495(B)	72(B)	415(B)	60(B)	22	AS WELDED	27(F)	**20(F)**
ER70S-G	(C)	DCEP	495(B)	72(B)	415(B)	60(B)	22	AS WELDED	(C)	**(C)**
ER80S-NI1	ARGON + 1 TO 5% OXYGEN	DCEP	550	80	470	68	24	AS WELDED	27(I)(J)(K)	**20(I)(J)(K)**
ER80S-NI2	ARGON + 1 TO 5% OXYGEN	DCEP	550	80	470	68	24	PWHT(D)	27(J)(L)(M)	**20(J)(L)(M)**
ER80S-NI3	ARGON + 1 TO 5% OXYGEN	DCEP	550	80	470	68	24	PWHT(D)	27(J)(M)(N)	**20(J)(M)(N)**
ER80S-D2	CO$_2$	DCEP	550	80	470	68	17	PWHT(D)	27(F)	**20(F)**
ER100S-1	ARGON + 2% OXYGEN	DCEP	670	100	605-705	88-102	16	AS WELDED	27(F)(J)(K)	**20(F)(J)(K)**
ER100S-2	ARGON + 2% OXYGEN	DCEP	670	100	605-705	88-102	16	AS WELDED	68(O)	**50(O)**
ER110S-1	ARGON + 2% OXYGEN	DCEP	760	110	655-740	95-107	15	AS WELDED	68(O)	**50(O)**
ER120S-1	ARGON + 2% OXYGEN	DCEP	830	120	725-840	105-122	14	AS WELDED	68(O)	**50(O)**
ERXXS-G	(C)	DCEP	(E)	(E)	(C)	(C)	(C)	AS WELDED	(C)	**(C)**
EXXC-G	**(C)**	**DCEP**	**(E)**	**(E)**	**(C)**	**(C)**	**(C)**	**AS WELDED**	**(C)**	**(C)**

Source: Ref 18

(A) CARBON DIOXIDE SHIELDING GAS. THE USE OF CO$_2$ FOR CLASSIFICATION PURPOSES SHALL NOT BE CONSTRUED TO RESTRICT THE USE OF ARGON-CO$_2$ OR ARGON-O$_2$ SHIELDING GAS MIXTURES. A FILLER METAL CLASSIFIED WITH CO$_2$ WILL ALSO MEET THE REQUIREMENTS OF THIS SPECIFICATION WHEN USED WITH ARGON-CO$_2$ OR ARGON-O$_2$ MIXTURES.

(B) FOR EACH INCREASE OF ONE PERCENTAGE POINT IN ELONGATION OVER THE MINIMUM, THE YIELD STRENGTH, OR TENSILE STRENGTH, OR BOTH MAY DECREASE 7 MPA (1 KSI) TO A MINIMUM OF 485 MPA (70 KSI) FOR THE TENSILE STRENGTH AND 400 MPA (58 KSI) FOR THE YIELD STRENGTH.

(C) SUBJECT TO AGREEMENT BETWEEN SUPPLIER AND PURCHASER.

(D) PWHT, POSTWELD HEAT TREATED IN ACCORDANCE WITH AWS A5.28-79.

(E) TENSILE STRENGTH SHALL BE CONSISTENT WITH THE LEVEL PLACED AFTER THE "ER" OR "E" PREFIX; E.G., ER90S-G SHALL HAVE 620 MPA (90 KSI) MINIMUM ULTIMATE TENSILE STRENGTH.

(F) AT -30 °C (-20 °F).

(G) AT -20 °C (0 °F).

(H) NOT REQUIRED.

(I) AT -45 °C (-50 °F).

(J) THE LOWEST AND THE HIGHEST VALUES OBTAINED SHALL BE DISREGARDED FOR THIS TEST. TWO OF THE THREE REMAINING VALUES SHALL BE GREATER THAN THE SPECIFIED 27 J (20 FT · LBF) ENERGY LEVEL; ONE OF THE THREE MAY BE LOWER BUT

SHALL NOT BE LESS THAN 20 J (15 FT · LBF). THE COMPUTED AVERAGE VALUE OF THE THREE VALUES SHALL BE EQUAL TO OR GREATER THAN THE 68 J (20-FT · LBF) ENERGY LEVEL.

(K) AS-WELDED IMPACT PROPERTIES.

(L) AT -60 °C (-80 °F).

(M) REQUIRED IMPACT PROPERTIES AFTER POSTWELD HEAT TREATMENT (SEE AWS A5.28-79).

(N) AT -75 °C (-100 °F).

(O) IMPACT PROPERTIES FOR THE ER100S-1, ER100S-2, ER110S-1, AND ER120S-1 SHALL BE OBTAINED AT TEST TEMPERATURE SPECIFIED IN THE TABLE ±1.7 °C (± 3 °F). THE LOWEST AND THE HIGHEST IMPACT VALUES THUS OBTAINED SHALL BE DISREGARDED FOR THIS TEST. TWO OF THE THREE REMAINING VALUES SHALL BE GREATER THAN THE SPECIFIED 68 J (50-FT · LBF) ENERGY LEVEL; ONE OF THE THREE MAY BE LOWER BUT SHALL NOT BE LESS THAN 54 J (40 FT · LBF); THE COMPUTED AVERAGE VALUES OF THESE THREE SHALL BE GREATER THAN THE SPECIFIED 68 J (50-FT · LBF) LEVEL.

Carbon steel electrodes are welded using CO_2, argon-oxygen, or argon-CO_2 shielding gases. Concern about welding fume levels has led to the increased use of the argon-base shielding gases. The low-alloy steel electrodes are commonly welded with argon-oxygen (2 to 5% O) or argon-CO_2 (5 to 12% CO_2) to avoid the loss of alloy during arc transfer. For carbon steel metal-cored electrodes, argon-CO_2 or argon-oxygen mixtures are used. The performance of welds is generally enhanced at higher argon levels.

The major difference between the carbon steel electrodes is the deoxidizer content. As the oxygen level of the molten weld pool increases, so must the deoxidizer level in the welding electrode if excessive porosity is to be avoided. The sources of oxygen are CO_2 or O_2 in the shielding gas, rust or millscale on the plate surface, and the baseplate oxygen (left from the deoxidation practice employed during steel manufacture. The ER70S-3 (1.1Mn-0.55Si) electrodes are the most widely used type, because of low cost, but they also have the lowest deoxidizer level. When welding rimmed steels with CO_2 shielding gas, porosity may be a problem. In that case an ER70S-4 (1.3Mn-0.75Si) electrode would do a better job. When heavier rest or scale is present or when high current welds are made (more baseplate dilution), it may be necessary to use an ER70S-6 (1.7Mn-0.9Si) electrode, which will tolerate most conditions, including high-speed welding. The ER70S-7 (1.8Mn-0.6Si) electrodes also have a higher deoxidizer level, better bead shape, and higher strength compared to ER70S-3 electrodes. The ER70S-5 (1.2Mn-0.5Si-0.7Al) electrodes, which contain high aluminum for deoxidation, are not readily available and are rarely used.

The ER70S-2 (low carbon, 1.2% Mn, 0.6% Si) electrodes contain small amounts of aluminum, titanium, and zirconium. Because these are strong deoxidizers, the ER70S-2 electrodes also produce good-quality welds at high weld puddle oxygen levels. The aluminum content makes the weld puddle more viscous, and this electrode is suitable for out-of-position welding because of ease of control. The low carbon content, and the fact that aluminum, titanium, and zirconium are strong denitriders, also make this electrode ideal for root passes. The highest weld quality is obtained when shielding gases higher in oxygen or CO_2 content are used. The viscous nature of the weld puddle can lead to an entrapment of gas that creates weld porosity. Higher oxygen and carbon dioxide components in the shielding gas improve puddle fluidity and minimize porosity levels.

Low-alloy electrodes contain sufficient deoxidizers to avoid porosity on relatively clean plate, but the silicon levels are kept low so as not to decrease toughness. An exception is ER80S-D2 electrodes, which contain higher silicon and can therefore be used with CO_2 or Ar-25CO_2 shielding.

Metal-cored electrodes may also be used. They produce higher deposition rates for a given heat input than do solid electrodes. Some users also claim that they provide a better bead profile and less chance of burn-through in root passes.

The GMAW process is considered a low-hydrogen process, but hydrogen can vary, depending on wire lubricant levels. Revisions of AWS A5.18 (Ref 18) and A5.28 (Ref 19) will include optional provisions for specification based on weld metal diffusible hydrogen.

Flux-Cored Arc Welding

Flux-cored arc welding may be used with or without an external shielding gas (see the article "Flux-Cored Arc Welding" in this Volume). Electrodes for flux-cored arc welding are classified under AWS A5.20 (covering carbon-steel electrodes) (Ref 20) and AWS A5.29 (covering low-alloy steel electrodes) (Ref 21).

Gas-shielded electrodes are designed to be used only with an external shielding gas (usually CO_2 or Ar-CO_2). The characteristics of these electrodes are given in Table 9. The majority of these electrodes are rutile-base and have excellent operating characteristics. The larger diameter versions (2.0 mm, or $\frac{5}{64}$ in., and larger) are used for flat and horizontal fillet welding, usually with CO_2 shielding gas. High penetration (high dilution), high heat inputs, and high deposition rates are characteristic of these electrodes. The T-1 group of large-diameter wires are used for single- and multiple-pass welding. The T-2 group contains higher levels of deoxidizer (manganese and silicon) to handle rust, scale, or rimmed steels, and is designed primarily for single-pass welding.

TABLE 9 ELECTRODES FOR FCAW APPLICATIONS

AWS ELECTRODE TYPE	**POLARITY**	**WELDING POSITION**	CHARACTERISTICS
GAS-SHIELDED ELECTRODES			
EXOT-1	DCEP	FLAT, HORIZONTAL	**RUTILE-BASE SPRAY-LIKE ARC, LOW SPATTER, FLAT BEAD SHAPE, EASY-PEELING SLAG, DEEP PENETRATION, HIGH DEPOSITION RATE. SINGLE- AND MULTIPASS WELDING. CO_2 NORMALLY USED.**
EXIT-1	DCEP	ALL	**SMALL-DIAMETER VERSIONS OF EXOT-1. CO_2 OR AR-CO_2 SHIELDING USED. SOME MICROALLOY VERSIONS CAPABLE OF GOOD LOW-TEMPERATURE TOUGHNESS.**
EXOT-2	DCEP	FLAT, HORIZONTAL	**OPERATE SIMILARLY TO EXOT-1 ELECTRODES BUT CONTAIN HIGHER LEVELS OF SILICON AND MANGANESE FOR IMPROVED PERFORMANCE OVER RUST AND MILLSCALE**
EXOT-5	DCEP	FLAT, HORIZONTAL	**BASIC ELECTRODE, MORE GLOBULAR ARC, ROUNDED BEAD, THIN SLAG COVER. PRODUCES VERY GOOD LOW-TEMPERATURE TOUGHNESS AND VERY LOW WELD-METAL HYDROGEN. USED FOR SINGLE-PASS AND MULTIPASS WELDS WHERE LOW HYDROGEN AND GOOD TOUGHNESS ARE REQUIRED. CO_2 SHIELDING NORMALLY USED.**
EXIT-5	DCEN	ALL	**SMALL-DIAMETER VERSIONS OF EXOT-5. USE CO_2 OR AR-CO_2 SHIELDING. OPERATING CHARACTERISTICS NOT AS GOOD AS T-1 ELECTRODES.**
SELF-SHIELDED ELECTRODES			
EXOT-3	DCEP	FLAT, HORIZONTAL	**DESIGNED FOR HIGH WELDING SPEEDS. SPRAY-LIKE ARC. USED ON THIN GAGE MATERIAL ONLY (<5 MM, OR 0.2 IN.).**
EXOT-4	DCEP	FLAT, HORIZONTAL	**HIGH-DEPOSITION-RATE ELECTRODE. GLOBULAR ARC. GOOD RESISTANCE TO SOLIDIFICATION CRACKING. USED FOR SINGLE-PASS AND MULTIPASS WELDING.**

EXOT-6	DCEP	FLAT, HORIZONTAL	**SPRAY-LIKE ARC WITH GOOD ROOT PENETRATION AND SLAG PEELING. PRODUCES GOOD LOW-TEMPERATURE TOUGHNESS. SINGLE-PASS AND MULTIPASS USE.**
EXOT-7	DCEN	FLAT, HORIZONTAL	**DESIGNED FOR SINGLE-PASS OR MULTIPASS WELDING IN ALL POSITIONS**
EXIT-7	DCEN	ALL	**GOOD CRACK-RESISTANCE WELD METAL**
EXOT-8	DCEN	FLAT, HORIZONTAL	**DESIGNED FOR SINGLE-PASS OR MULTIPASS WELDS IN ALL POSITIONS. GOOD LOW-TEMPERATURE TOUGHNESS.**
EXIT-8	DCEN	ALL	
EXOT-10	DCEN	FLAT, HORIZONTAL	**HIGH-SPEED SINGLE-PASS WELDS ON ANY THICKNESS PLATE**
EXOT-11	DCEN	FLAT, HORIZONTAL	**VERY WELL-OPERATING, GENERAL PURPOSE, ALL-POSITION ELECTRODE. TENSILE DUCTILITY MAY BE POOR ON WELDS IN PLATE >19 MM ($\frac{3}{4}$ IN.).**
EXIT-11	**DCEN**	**ALL**	

Source: Ref 20, 21

The smaller diameter (1.6 mm, or $\frac{1}{16}$ in., and under) rutile-base electrodes are used in all positions. They are widely used because of the ease of use and the high deposition rates possible when out-of-position. Carbon dioxide shielding or Ar-CO_2 shielding is used. The Ar-CO_2 mixture (typically with 20 to 25% CO_2) gives better operating characteristics and less loss of elements during arc transfer. Many electrodes are specifically designed to work in one gas or the other, so manufacturer recommendations should be followed. The weld metal toughness of these electrodes can vary from fair to very good. Newer, microalloyed versions, both carbon and low-alloy steel, are capable of achieving good low-temperature toughness and low-hydrogen characteristics. Small quantities of titanium and boron are added to modify transformation characteristics and promote the formation of acicular ferrite. The quantities of titanium and boron transferred to the weld metal are critical. The use of the recommended shielding gas and tight control of welding variables (especially arc voltage) are necessary to achieve optimum results. There is the possibility of solidification cracking with excessive amounts of boron in the weld metal.

The basic electrodes (T-5 types) do not have operating characteristics as good as those of the rutile types, but they are capable of producing excellent low-temperature toughness and low diffusible hydrogen. They are mainly used for multipass welding or where resistance to cracking (both hydrogen-induced and solidification) is required. Some smaller-diameter versions can be used for out-of-position welding, but the performance features are inferior to those of rutile-base electrodes.

Self-shielded electrodes are designed to be used only without a shielding gas. The weld metal is protected from atmospheric contamination by denitriders, deoxidizers, and gas formers in the core. Compared to the gas-shielded types, the self-shielded electrodes require closer control of welding parameters to achieve consistent results. This is because the balance of denitriders/deoxidizers to atmospheric gases, N_2 and O_2, is critical to performance and weld metal mechanical properties. An increase in arc length, for instance, will reduce the transfer of aluminum and titanium to the weld metal, but it will increase the amount of nitrogen and oxygen absorbed. This interaction with the atmosphere also makes weld metal hydrogen levels more dependent on atmospheric moisture. Self-shielded electrodes can be classified as being either controlled-hydrogen or low-hydrogen, provided that welding parameters, including proper electrode stickout parameters are maintained. These electrodes also have the special ability to make quality welds, even in windy conditions.

T-3 and T-10 Types. Self-shielded electrodes are differentiated according to the intended application and the mechanical properties achieved. The T-3 types are single-pass electrodes for use on sheet metal and plate under 5 mm (0.2 in.) thick. The fast travel speeds and high cooling rate of a thicker material result in poor weld metal ductility. The T-

10 types are also designed for high-speed, single-pass welding. They operate at lower travel speeds, but the deposit is less hardenable and they are not limited by plate thickness.

The T-4, T-7, and T-11 types are intended for single-pass or multipass applications, but none of them is required to meet minimum weld metal toughness levels. The T-4 and large diameter T-7 types are for horizontal and flat use. Both produce high deposition rates, but the T-7 is capable of higher travel speeds. The T-11 electrodes have very good operating characteristics and are frequently used for out-of-position welding of thin plate and gage materials. The higher hardenability of the T-11 deposit may require preheat when welding plate over 19 mm ($\frac{3}{4}$ in.) thick.

T-6 and T-8 Types. T-6 electrodes are intended for single-pass and multipass welding where good toughness is needed in the flat and horizontal position. T-8 electrodes provide good toughness single-pass or multipass welds in all positions. The T-8 slag system forms the basis for many of the higher-toughness, low-alloy steel electrodes.

Gas-Tungsten Arc and Plasma Arc Welding

These two processes are so similar that they can be treated together. Their main use on carbon steels is for welding thinner material components (especially thin-walled tubing) and for making root passes in pipe joints. (Detailed information is available in the articles "Gas-Tungsten Arc Welding" and "Plasma Arc Welding" in this Volume.) Both processes produce very clean, high-quality weld metal because of the use of inert shielding gases and a tightly-controlled arc. Each is considered a low-hydrogen process, with the major contribution coming from the filler material, if one is used. Heat inputs are relatively low, so care must be exercised when welding hardenable carbon steels.

A choice must be made whether to use a filler material or to weld autogenously (without a filler material). Generally, thin carbon steels are welded autogenously up to 3 mm ($\frac{1}{8}$ in.) thick (for GTAW) or 6 mm ($\frac{1}{4}$ in.) thick (for PAW) unless a metallurgical problem precludes the use of either process. If the baseplate sulfur or phosphorus levels are high, then a low-carbon, high-manganese filler material may have to be used to prevent solidification cracking. High oxygen levels in the steel (rimmed) will require a high-deoxidizer-containing filler material to counteract porosity. Consumable inserts are sometimes used in root-pass welds. These inserts are of a specific steel composition (some have compositions equal to the filler wires) and are preplaced in the root of weld. They are melted and become part of the root-pass weld metal. Their use generally results in more consistent weld-bead fusion and underbead formation. Consumable inserts are covered in AWS A5.30-79 (Ref 22).

Filler metals for GTAW or PAW joining of carbon steels are the same as those used for GMAW (see Tables 7 and 8). The same guidelines prevail (that is, higher oxygen weld puddles require higher deoxidizer levels in the filler material). Because much of the welding is done-out-of-position or in root passes, the ER70S-2 filler rod is very popular. This is due to its viscous, easier-to-control weld puddle and to the low carbon level that reduces the likelihood of solidification cracking in root beads. Manual GTAW or PAW processing is normally done with straight length rods. Spooled or coiled product is used for mechanized setups.

Submerged Arc Welding

Submerged arc welding is a high-current process capable of high heat input, high penetration, high rate of travel, and high deposition rates (see the article "Submerged Arc Welding" in this Volume).

The high heat inputs associated with this process lead to slow weld metal cooling rates. The presence of a flux blanket tends to slow down the rate of cooling even more. This promotes formation of microstructures less susceptible to hydrogen-induced cracking and allows longer times for hydrogen to escape. Excessive heat inputs, however, can result in low strength and perhaps low toughness in the weld metal and HAZ.

High penetration results in high baseplate dilution. It is not uncommon to have over 60% dilution in one- or two-pass, square butt welds. This addition of base material must be taken into consideration when selecting procedures and consumables to meet specific mechanical property requirements. Procedure tests should always be run. Impurities in the baseplate (for example, sulfur, phosphorus, oxygen, and nitrogen) can lead to problems in heavily diluted weld metal.

High travel speeds increase the probability of solidification cracking, especially in small, more diluted root passes. Under such circumstances, special attention should be paid to baseplate chemistry (carbon, sulfur, and phosphorus) and bead shape control.

Consumable selection involves choosing both a flux and an electrode. They are an interactive pair that together determine the properties of the weld. Fluxes and electrodes for welding carbon steels are classified under AWS A5.17 (carbon steels) (Ref 23) and AWS A5.23 (low-alloy steels) (Ref 24). Solid electrodes are classified on the basis of chemistry. The flux/electrode combination is classified on the basis of weld metal mechanical properties (tensile and Charpy V-notch) and, for low-alloy electrodes only, deposit chemistry. Preliminary selection based on mechanical properties is thus fairly easy. Procedure tests should always be run, because production welding conditions rarely match the AWS test procedures (which are low-dilution, moderate-heat-input welds).

Flux Selection. Selection for performance features is more difficult. Performance characteristics are largely determined by the flux. Past experience or manufacturer recommendations can be used. Caution must be exercised in the selection of a flux.

Active fluxes (those that contain, and contribute to the weld metal, significant amounts of silicon and manganese) are designed to tolerate high weld puddle oxygen. Their use is mainly for single- or limited-pass welding on scaled or rusty plate. In the absence of oxides, the deoxidizers enter the weld metal and may cause excessive hardness. Therefore, their use may be limited on multipass welds. Transfer of metals from the flux to the weld metal is enhanced at higher arc voltages, where more flux is consumed per unit of electrode. Close control of arc voltage is required with active fluxes, or weld metal cracking might occur. Low-alloy steel electrodes should not be used with active fluxes.

Neutral fluxes (those that contain and contribute no or low amounts of silicon and manganese) are not as tolerant to rust and millscale and are intended for welding on clean plate or in multipass welding.

Intermediate Activity Fluxes. Some fluxes have intermediate activity. This provides them with some tolerance to rust and millscale, and it allows their use for multipass welding with less probability of excessive manganese and silicon pickup.

Wall Neutrality Number Method. The appendices of AWS A5.17 (Ref 23) and A5.23 (Ref 24) describe a method, the wall neutrality number, that may be used to quantify the activity of a flux.

Basic fluxes are generally used for multipass welding where good low-temperature toughness is required.

Acid fluxes generally have better performance features than basic fluxes. Acid fluxes deslag more easily and may give better visual bead appearance. However, overall performance may be lacking due to silicon pickup or carbon depletion in heavier thicknesses.

Flux Storage. Submerged arc welding can be a low-hydrogen process if the fluxes are properly stored and handled. Flux exposed to the atmosphere can pick up moisture, especially under humid conditions. This moisture can be removed by drying the flux. The manufacturer should be consulted for proper procedures. Heated flux hoppers, which hold the flux above 120 °C (250 °F), are frequently used to maintain low moisture during usage, especially when welding crack-sensitive steels.

Electrogas Welding

From a metallurgical viewpoint, the primary feature of the electrogas welding process is the protracted thermal cycle that results from the relatively slow progression rate of the welding arc (see the article "Electroslag and Electrogas Welding" in this Volume). The slow heat-up and cool-down results in a wide, grain-coarsened HAZ and large columnar grains in the weld metal. Low toughness can be a problem in some applications. Control of welding conditions to limit heat input will improve HAZ properties. Low-alloy electrodes are often used when low-temperature weld-metal toughness is required. Table 10 gives typical applications of steels based on grade. Electrogas welding of quenched and tempered carbon steels can result in serious degradation of strength and toughness properties in the HAZ. A postweld heat treatment may be necessary to restore the properties. Some codes (for example, AWS D1.1, Ref 3) prohibit the use of EGW on quenched and tempered steels.

TABLE 10 SELECTED STEEL GRADES COMMONLY JOINED BY ELECTROGAS WELDING PROCESS

GRADE	
AISI	ASTM
PLAIN CARBON STEELS	
1010	**. . .**
1018	**. . .**
1020	**. . .**
STRUCTURAL STEELS	
. . .	**A 36**
. . .	**A 131**
. . .	**A 242**
. . .	**A 283**
. . .	**A 441**
. . .	**A 572**
. . .	**A 573**
. . .	**A 588**
PRESSURE VESSEL STEELS	
. . .	**A 36**
. . .	**A 285**
. . .	**A 515**
. . .	**A 516**
. . .	**A 537**

Source: Ref 25

On the positive side, the formation of hard, crack-sensitive microstructures is less likely to occur because of the slow cooling rates, and the elevated time at temperature allows more time for hydrogen to escape (electrogas welding is a low-hydrogen process). Preheating is not normally required.

Baseplate dilution can vary with welding parameters. High voltage and low current will increase sidewall penetration. Up to about 35% dilution is possible. Because this will have a major effect on weld metal properties, it is very important to run procedure qualification tests.

Electrodes for electrogas welding are classified in accordance with AWS A5.26 (Ref 26). The process utilizes either solid electrodes (Table 11) or composite electrodes (Table 12). The composite electrodes can be flux-cored or metal-cored and require either shielding gas or self shielding. Typical properties of carbon steel grades in the as-welded condition are given in Table 13. Carbon dioxide or Ar-CO_2 mixtures are used for the shielded flux-cored and solid-wire electrodes. The solid electrodes used are similar to those used for GMAW. The flux-cored electrodes used for electrogas welding are not the same as those used for FCAW. They have a lower level of slag formers. Flux-cored electrodes offer an advantage over solid electrodes in that they have a higher deposition rate (approximately 20%). This will result not only in faster joint completion rates, but also in lower heat inputs.

TABLE 11 CHEMICAL COMPOSITION FOR SOLID CARBON STEEL ELECTRODES USED IN ELECTROGAS WELDING

AWS CLASSIFICATION$^{(B)}$	COMPOSITION, %$^{(A)}$											
	C	**Mn**	**P**	**S**	**Si**	**Ni**	**Cr**	**Mo**	**Cu**$^{(c)}$	**Ti**	**Zr**	Al
EGXXS-1	0.07-0.19	0.90-1.40	0.025	0.035	0.30-0.50	. . .	. . .	. . .	0.35	. . .	. . .	**. . .**
EGXXS-D2$^{(D)}$	0.07-0.12	1.60-2.10	0.025	0.035	0.50-0.80	0.15	. . .	0.40-0.60	0.35	. . .	. . .	**. . .**

EGXXS-G	**NO CHEMICAL REQUIREMENTS(E)**											
EGXXS-2	0.06	0.90-1.40	0.025	0.035	0.40-0.70	. . .	. . .	. . .	0.35	0.05-0.15	0.02-0.12	**0.05-0.15**
EGXXS-3	0.06-0.15	0.90-1.40	0.025	0.035	0.45-0.70	. . .	. . .	. . .	0.35	. . .	. . .	**. . .**
EGXXS-5	0.07-0.19	0.90-1.40	0.025	0.035	0.30-0.60	. . .	. . .	. . .	0.35	. . .	. . .	**0.50-0.90**
EGXXS-6	**0.07-0.15**	**1.40-1.85**	**0.025**	**0.035**	**0.80-1.15**	**. . .**	**. . .**	**. . .**	**0.35**	**. . .**	**. . .**	**. . .**

(A) SINGLE VALUES ARE MAXIMUMS.

(B) THE LETTERS "XX" (AS IN 6Z, 60, 62, 7Z, 70, 72, AND 8Z, 80, AND 82) ARE MECHANICAL PROPERTY DESIGNATIONS.

(C) THE COPPER LIMIT INCLUDES COPPER THAT MAY BE APPLIED AS A COATING ON THE ELECTRODE.

(D) FORMERLY EGXXS-1B.

(E) COMPOSITION SHALL BE REPORTED; THE REQUIREMENTS ARE THOSE AGREED TO BY THE PURCHASER AND THE SUPPLIER.

TABLE 12 COMPOSITION OF WELD METAL OBTAINED WHEN USING COMPOSITE FLUX-CORED AND METAL-CORED ELECTRODES

AWS CLASSIFICATION (A)	**SHIELDING GAS**	**COMPOSITION, WT%(B)**										OTHER ELEMENTS (TOTAL)
		C	**Mn**	**P**	**S**	**Si**	**Ni**	**Cr**	**Mo**	**Cu**	**V**	
EGXXT-1	NONE	(C)	1.7	0.03	0.03	0.50	0.30	0.20	0.35	0.35	0.08	**0.50**
EGXXT-2	CO_2	(C)	2.0	0.03	0.03	0.90	0.30	0.20	0.35	0.35	0.08	**0.50**
EGXXT-NIL (FORMERLY EGXXT-3)	CO_2	0.10	1.O-1.8	0.03	0.03	0.50	0.7-1.1	. . .	0.30	0.35	. . .	**0.50**
EGXXT-NM1 (FORMERLY EGXXT-4)	AR/CO_2 OR CO_2	0.12	1.0-2.0	0.02	0.03	0.15-0.50	1.5-2.0	0.20	0.40-0.65	0.35	0.05	**0.50**
EGXXT-NM2 (FORMERLY EGXXT-6)	CO_2	0.12	1.10-2.10	0.03	0.03	0.20-0.60	1.1-2.0	0.20	0.10-0.35	0.35	0.05	**0.50**
EGXXT-W (FORMERLY EGXXT-5)	CO_2	0.12	0.50-1.3	0.03	0.03	0.30-0.80	0.40-0.80	0.45-0.70	. . .	0.30-0.75	. . .	**0.50**
EGXXT-G	**NOT SPECIFIED(D)**											

Source: Ref 25

(A) THE LETTERS "XX" (FOR EXAMPLE, 6Z, 60, 62, 7Z, 70, 72, 8Z, 80 OR 82) REFER TO MECHANICAL PROPERTY DESIGNATIONS.

(B) SINGLE VALUES ARE MAXIMUMS.

(C) COMPOSITION RANGE FOR CARBON NOT SPECIFIED FOR THESE CLASSIFICATIONS, BUT THE AMOUNT SHALL BE DETERMINED AND REPORTED.

(D) COMPOSITION SHALL BE REPORTED; THE REQUIREMENTS ARE THOSE AGREED TO BETWEEN THE PURCHASER AND SUPPLIER.

TABLE 13 TYPICAL AS-WELDED MECHANICAL PROPERTIES OF ELECTROGAS WELDS OBTAINED WHEN USING FLUX-CORED ELECTRODES

ASTM GRADE	**APPLICATION**	**THICKNESS**		**ELECTRODE CLASS (A5.26)**	**SHIELDING GAS**	**TENSILE PROPERTIES**					**TEST TEMPERATURE**		IMPACT ENERGY (CHARPY V-NOTCH)			
						YIELD STRENGTH		**TENSILE STRENGTH**		**ELONGATION IN 50 MM (2 IN.), %**			**WELD METAL**		HAZ	
		MM	**IN.**			**MPA**	**KSI**	**MPA**	**KSI**		**°C**	**°F**	**J**	**FT · LBF**	**J**	FT · LBF
1020[(A)]	STRUCTURAL	12	$\frac{15}{32}$	EG72T1	. . .	496	72	552	80	28	-29	-20	58	43	. . .	**. . .**
		25	1	EG72T1	. . .	441	64	545	79	27	-29	-20	45	33	. . .	**. . .**
A 36	STRUCTURAL	25	1	EG72T1	. . .	448	65	545	79	27	-29	-20	56	41	. . .	**. . .**
		25	1	EG72T4	CO_2	483	70	634	92	26	-30	-22	47	35	. . .	**. . .**
		75	3	EG72T4	CO_2	462	67	614	89	26	-30	-22	42	31	. . .	**. . .**
		75	3	EG72T4	80AR-20CO_2	. . .	. . .	600	87	23	-18	-0.4	72	53	. . .	**. . .**
A 131-C	SHIPBUILDING	38	$1\frac{1}{2}$	EG72T3	CO_2	. . .	. . .	490	71	30	-34	-29	61	45	41	**30**
A 441	STRUCTURAL	19	$\frac{3}{4}$	EG72T1	. . .	510	74	565	82	23	-29	-20	56	41	. . .	**. . .**
		25	1	EG72T1	. . .	455	66	558	81	23	-29	-20	49	36	. . .	**. . .**
		50	2	EG72T1	. . .	448	65	558	81	24	-29	-20	41	30	. . .	**. . .**
		50	2	EG72T4	CO_2	468	68	634	92	24	-18	-0.4	41	30	. . .	**. . .**
A 572	STRUCTURAL	25	1	EG72T4	80AR-20CO_2	. . .	. . .	. . .	. . .	. . .	-10	14	83	61	46	**34**
A 572-50	STRUCTURAL	38	$1\frac{1}{2}$	EG72T1	. . .	393	57	531	77	23	-18	-0.4	19	14	. . .	**. . .**
A 588	STRUCTURAL	76	3	EG72T4	CO_2	. . .	. . .	655	95	23	-18	-0.4	76	56	. . .	**. . .**
A 203	PRESSURE VESSEL	41	$1\frac{5}{8}$	EG72T4	80AR-20CO_2	365	53	496	72	32	-40	-40	31	23	. . .	**. . .**

A 516	PRESSURE VESSEL	25	1	EG72T1	. . .	400	58	510	74	26	-29	-20	38	28	. . .	**. . .**
		38	$1\frac{1}{2}$	EG72T4	80AR-20CO$_2$	538	78	621	90	29	-29	-20	56	41	87	**64**
A 537-1	PRESSURE VESSEL	19	$\frac{3}{4}$	EG72T1	. . .	483	70	593	86	24	-20	-4	58	43	. . .	**. . .**
		25	1	EG72T4	CO$_2$	427	62	572	83	29	-29	-20	46	34	53	**39**
		28	$1\frac{1}{8}$	EG72T4	80AR-20CO$_2$	510	74	689	100	26	-30	-22	62	46	. . .	**. . .**
A 633	**PRESSURE VESSEL**	**25**	**1**	**EG72T1**	**. . .**	**510**	**74**	**614**	**89**	**25**	**-18**	**-0.4**	**84**	**62**	**. . .**	**. . .**

Source: Ref 25

(A) AMERICAN IRON AND STEEL INSTITUTE (AISI) DESIGNATION.

Electroslag Welding (Ref 27, 28)

Electroslag welding (ESW) is similar in many respects, to electrogas welding (see the article "Electroslag and Electrogas Welding" in this Volume). Both are used to complete large, vertical welds in a single pass. High heat inputs and protracted thermal cycles are characteristic of these processes. Weld-metal hydrogen levels are low, and preheat is usually not required. The residual stresses and angular distortion are generally low. Weld-metal soundness is excellent provided the welding can be completed without stopping. The slow heating and cooling cycles, however, result in a wide, coarse-grained heat-affected zone and coarse, columnar weld metal. Low toughness and strength can be a problem in certain applications. Some codes, such as AWS D1.1, have restrictions on the use of these processes.

Electroslag welding is not an arc welding process. An arc is used only to initiate melting of a flux in order to form a molten slag. The molten slag, which is conductive, then extinguishes the arc and the heat necessary to complete the weld is generated by electrical resistance to the current passing through the slag. The molten slag also acts to shield the weld metal from the atmosphere. The fluxes used for electroslag welding are unique to the process. Most SAW fluxes will not work in ESW applications. Unlike electrogas welding, electroslag welding may use multiple electrodes. This allows thicker materials (up to 900 mm, or 36 in.) to be welded. Dilution from the baseplates can be quite high (up to 50%), so procedure qualifications are essential.

Process Variations. There are two variations of the process in general use. The conventional method uses a wire electrode (either solid or cored) and a nonconsumable guide to direct the electrode into the molten bath. The welding head moves up the plate as the weld progresses, and sliding copper shoes hold the molten puddle in position.

The consumable guide method employs a consumable guide tube that extends down the length of the joint. The welding head remains stationary at the top of the joint, and a wire electrode (either solid or cored) is fed down the center of the guide tube into the molten puddle. The steadily rising weld puddle consumes both the electrode and the guide tube. Some guide tubes are coated with flux. This provides electrical insulation and supplements the slag, some of which plates out onto the weld surface. Copper dams are used to hold the molten pool in place.

Metallurgically, there is little difference between the variations except that the consumable guide tubes (which comprise 5 to 10% of the weld metal in a 50 mm, or 2 in., plate weld) may not be available in the same range of chemistries as the wire electrodes.

Electrode chemistries and required properties for ESW are given in Ref 29. Typical mechanical properties of electroslag welds in carbon and alloy steels and in structural steels are given in Tables 14 and 15, respectively.

TABLE 14 TYPICAL MECHANICAL PROPERTIES OF WELD METAL FROM ELECTROSLAG WELDS IN CARBON AND ALLOY STEELS

BASE METAL (ASTM SPECIFICATION)	THICKNESS		ELECTRODE		HEAT TREATMENT(A)	YIELD STRENGTH		TENSILE STRENGTH		ELONGATION IN 50 MIN (2 IN.), %	IMPACT ENERGY (CHARPY V-NOTCH) AT 12.2 °C (10 °F) FOR INDICATED NOTCH LOCATION					
											WMFG(B)		WMCG(C)		BM HAZ(D)	
	IN.	MM	TYPE	QUANTITY		MPA	KSI	MPA	KSI		J	FT · LBF	J	FT · LBF	J	FT · LBF
A 204-A	$3\frac{1}{2}$	89	MN-MO	1	NT	382	55.5	562	81.5	27	46	34	39	29	18	**13**
A 515 GR70	$1\frac{1}{2}$	38	MN-MO	1	SR	358	52.0	579	84.1	26	61	45	35	26	10	**7**
	2	51	MN-MO	1	SR	469	68.1	587	85.2	23	63	46	29	21	7	**5**
	$3\frac{3}{8}$	86	MN-MO	2	NT	396	57.5	537	78.0	29	46	34	45	33	30	**22**
	$6\frac{3}{4}$	171	MN-MO	2	NT	313	45.5	512	74.3	31	33	24	29	21	16	**12**
A 302-B	3	76	MN-MO-NI	1	NT	393	57.0	565	82.0	28	72	53	71	52	86	**63**
A 387-C	3	76	$1\frac{1}{4}$CR-$\frac{1}{2}$MO	2	NT	320	46.5	503	73.0	29	95	70	103	76	78	**57**
A 387-D	$3\frac{1}{4}$	83	$2\frac{1}{4}$CR-1MO	1	SR	396	57.5	565	82.0	25	63	46	68	50	65	**48**
	$7\frac{1}{2}$	**191**	**$2\frac{1}{4}$CR-1MO**	**2**	**SR**	**551**	**80.0**	**658**	**95.5**	**20**	**84(E)**	**62(E)**	**102(E)**	**75(E)**	**113(E)**	**83(E)**

Source: Ref 28

(A) NT, NORMALIZED AND TEMPERED; SR, STRESS-RELIEVED.

(B) WMFG, WELD METAL, FINE GRAIN.

(C) WMCG, WELD METAL, COARSE GRAIN.

(D) BM HAZ, BASE METAL, HEAT-AFFECTED ZONE.

(E) IMPACTS AT 10 °C (50 °F).

TABLE 15 TYPICAL MECHANICAL PROPERTIES OF WELD METAL (IN AS-WELDED CONDITION) OBTAINED WITH ELECTROSLAG WELDS IN STRUCTURAL STEELS

Using solid-wire electrodes and applying consumable guide method

BASE METAL, ASTM	THICKNESS		ELECTRODE		YIELD STRENGTH		TENSILE STRENGTH		ELONGATION IN 50 MM (2 IN.), %	REDUCTION IN AREA, %	IMPACT ENERGY$^{(A)}$	
	MM	IN	TYPE (AWS)	QUANTITY	MPA	KSI	MPA	KSI			J	FT · LBF
A 441	25	1	EM13K-EW	1	344	49.9	523	75.8	28.0	59.0	23	**17**
	64	$2\frac{1}{2}$	EM13K-EW	1	316	45.8	505	73.3	26.5	66.0	37	**27**
A 36	152	6	EM13K-EW	2	317	46.0	548	79.5	28.5	52.8	. . .	**. . .**
	305	12	EM13K-EW	2	255	37.0	464	67.3	33.5	71.0	. . .	**. . .**
A 572 GRADE 42	203	8	EM13K-EW	2	400	58.2	585	84.8	25.0	67.6	38	**28**
GRADE 60	**57**	$2\frac{1}{4}$	**EH10MO-EW**	**1**	**423**	**61.5**	**680**	**98.5**	**18.0**	**35.6**	**. . .**	**. . .**

Source: Ref 28

(A) AT -17.8 °C (0 °F).

Stud Arc Welding

Stud arc welding is commonly used to attach carbon steel studs to carbon steel plate (see the article "Stud Arc Welding" in this Volume). A primary characteristic of this process is the rapid heating and cooling cycle. Both the HAZ and the weld-metal area are relatively small. The rapid cooling rate can be a problem with medium-carbon and high-carbon steels because the likelihood of martensite formation is high. Preheat is essential for welding medium-carbon and high-carbon steels but is not required for low-carbon steels. Procedure tests should be run on high-carbon steels. Stud arc welding is a low-hydrogen process if the plate surface and studs are free of hydrogen-containing materials.

Studs for attachment to carbon steels are chosen based on their strength. Low-carbon steel studs are made from cold-drawn bar stock in grades 1006 to 1020 (killed or semi-killed). A minimum tensile strength of 380 to 415 MPa (55 to 60 ksi) is required. A stronger heat-treated steel stud (830 MPa, or 120 ksi, tensile strength) is also available. Most carbon steel studs require a flux (affixed to the end of the stud) for deoxidation. Aluminum is commonly used for this purpose.

References cited in this section

3. STRUCTURAL WELDING CODE, STEEL, AWS D1.1, AMERICAN WELDING SOCIETY, 1992
15. *WELDING HANDBOOK*, 8TH ED., VOL 1, AMERICAN WELDING SOCIETY, 1987, P 437-464
16. SPECIFICATION FOR CARBON STEEL FOR SHIELDED METAL ARC WELDING, A5.1-91, AMERICAN WELDING SOCIETY, 1991
17. SPECIFICATION FOR LOW-ALLOY STEEL COVERED ARC WELDING ELECTRODES, A5.5-81, AMERICAN WELDING SOCIETY, 1981
18. SPECIFICATION FOR LOW-ALLOY STEEL FILLER METALS FOR GAS-SHIELDED ARC WELDING, A5.28-79, AMERICAN WELDING SOCIETY, 1979
19. SPECIFICATION FOR CARBON STEEL FILLER METALS FOR GAS-SHIELDED ARC WELDING, A5.18-79, AMERICAN WELDING SOCIETY, 1979
20. SPECIFICATION FOR CARBON STEEL ELECTRODES FOR FLUX-CORED ARC WELDING, A5.20-79, AMERICAN WELDING SOCIETY, 1979

21. SPECIFICATION FOR LOW-ALLOY STEEL ELECTRODES FOR FLUX-CORED ARC WELDING, A5.29-80, AMERICAN WELDING SOCIETY, 1980
22. SPECIFICATION FOR CONSUMABLE INSERTS, A5.30-79, AMERICAN WELDING SOCIETY, 1979
23. SPECIFICATION FOR CARBON STEEL ELECTRODES AND FLUXES FOR SUBMERGED ARC WELDING, A5.17-89, AMERICAN WELDING SOCIETY, 1989
24. SPECIFICATION FOR LOW-ALLOY STEEL ELECTRODES AND FLUXES FOR SUBMERGED ARC WELDING, A5.23-90, AMERICAN WELDING SOCIETY, 1990
25. *WELDING HANDBOOK,* 8TH ED., VOL 2, AMERICAN WELDING SOCIETY, 1991, P 234-269
26. SPECIFICATION FOR CARBON AND LOW-ALLOY STEEL ELECTRODES FOR ELECTROGAS WELDING, A5.26-91, AMERICAN WELDING SOCIETY, 1991
27. *WELDING HANDBOOK,* 8TH ED., VOL 1, AMERICAN WELDING SOCIETY, 1987, P 13-16
28. *WELDING HANDBOOK,* 8TH ED., VOL 2, AMERICAN WELDING SOCIETY, 1991, P 272-294
29. SPECIFICATION FOR CARBON AND LOW-ALLOY STEEL ELECTRODES AND FLUXES FOR ELECTROSLAG WELDING, A5.25-91, AMERICAN WELDING SOCIETY, 1991

Arc Welding of Carbon Steels

Ronald B. Smith, The ESAB Group, Inc.

References

1. B.A. GRAVILLE, *THE PRINCIPLES OF COLD CRACKING CONTROL IN WELDS,* DOMINION BRIDGE COMPANY, 1975
2. H. GRANJON, *FUNDAMENTALS OF WELDING METALLURGY,* ABINGTON PUBLISHING, 1991, P 156-177
3. STRUCTURAL WELDING CODE, STEEL, AWS D1.1, AMERICAN WELDING SOCIETY, 1992
4. *WELDED STEEL CONSTRUCTION (METAL ARC WELDING),* CSA W59, CANADIAN STANDARDS ASSOCIATION, 1989
5. R.D. STOUT, *WELDABILITY OF STEELS,* 4TH ED., WELDING RESEARCH COUNCIL, 1987, P 420-426
6. *PROCESS OF ARC WELDING CARBON & CARBON MANGANESE STEELS,* BS5135, BRITISH STANDARDS INSTITUTE, 1984
7. F.R. COE, *WELDING STEELS WITHOUT HYDROGEN CRACKING,* THE WELDING INSTITUTE, 1973
8. J.F. LANCASTER, *THE METALLURGY OF WELDING,* 4TH ED., ALLEN AND UNWIN, 1987, P 157
9. *THE PROCEDURE HANDBOOK OF ARC WELDING,* LINCOLN ELECTRIC COMPANY, 1973, CHAPTERS 5 AND 12
10. W.F. SAVAGE AND A.H. ARONSON, *PREFERRED ORIENTATION IN THE WELD FUSION ZONE,* VOL 46, 1966
11. *WELDING HANDBOOK,* 8TH ED., VOL 1, AMERICAN WELDING SOCIETY, 1987, P 261
12. H.B. CARY, *MODERN WELDING TECHNOLOGY,* 2ND ED., PRENTICE-HALL, 1989, P 204-206
13. P.T. HOULDCROFT, *WHICH PROCESS?,* ABINGTON PUBLISHING, 1990
14. S. KOU, *WELDING METALLURGY,* JOHN WILEY & SONS, 1987, P 63
15. *WELDING HANDBOOK,* 8TH ED., VOL 1, AMERICAN WELDING SOCIETY, 1987, P 437-464
16. SPECIFICATION FOR CARBON STEEL FOR SHIELDED METAL ARC WELDING, A5.1-91, AMERICAN WELDING SOCIETY, 1991
17. SPECIFICATION FOR LOW-ALLOY STEEL COVERED ARC WELDING ELECTRODES, A5.5-81, AMERICAN WELDING SOCIETY, 1981

18. SPECIFICATION FOR LOW-ALLOY STEEL FILLER METALS FOR GAS-SHIELDED ARC WELDING, A5.28-79, AMERICAN WELDING SOCIETY, 1979
19. SPECIFICATION FOR CARBON STEEL FILLER METALS FOR GAS-SHIELDED ARC WELDING, A5.18-79, AMERICAN WELDING SOCIETY, 1979
20. SPECIFICATION FOR CARBON STEEL ELECTRODES FOR FLUX-CORED ARC WELDING, A5.20-79, AMERICAN WELDING SOCIETY, 1979
21. SPECIFICATION FOR LOW-ALLOY STEEL ELECTRODES FOR FLUX-CORED ARC WELDING, A5.29-80, AMERICAN WELDING SOCIETY, 1980
22. SPECIFICATION FOR CONSUMABLE INSERTS, A5.30-79, AMERICAN WELDING SOCIETY, 1979
23. SPECIFICATION FOR CARBON STEEL ELECTRODES AND FLUXES FOR SUBMERGED ARC WELDING, A5.17-89, AMERICAN WELDING SOCIETY, 1989
24. SPECIFICATION FOR LOW-ALLOY STEEL ELECTRODES AND FLUXES FOR SUBMERGED ARC WELDING, A5.23-90, AMERICAN WELDING SOCIETY, 1990
25. *WELDING HANDBOOK,* 8TH ED., VOL 2, AMERICAN WELDING SOCIETY, 1991, P 234-269
26. SPECIFICATION FOR CARBON AND LOW-ALLOY STEEL ELECTRODES FOR ELECTROGAS WELDING, A5.26-91, AMERICAN WELDING SOCIETY, 1991
27. *WELDING HANDBOOK,* 8TH ED., VOL 1, AMERICAN WELDING SOCIETY, 1987, P 13-16
28. *WELDING HANDBOOK,* 8TH ED., VOL 2, AMERICAN WELDING SOCIETY, 1991, P 272-294
29. SPECIFICATION FOR CARBON AND LOW-ALLOY STEEL ELECTRODES AND FLUXES FOR ELECTROSLAG WELDING, A5.25-91, AMERICAN WELDING SOCIETY, 1991

Arc Welding of Carbon Steels

Ronald B. Smith, The ESAB Group, Inc.

Selected References

- S. GANESH AND R.D. STOUT, MATERIAL VARIABLES AFFECTING LAMELLAR TEARING SUSCEPTIBILITY IN STEELS, *WELD. J.,* NOV 1976
- P.F. HOULDCROFT, *SUBMERGED ARC WELDING,* 2ND ED., ABINGTON PUBLISHING, 1989
- D.J. KOTECKI, HYDROGEN RECONSIDERED, *WELD. J.,* AUG 1992
- G.E. LINNERT, *WELDING METALLURGY,* VOL 2, AMERICAN WELDING SOCIETY, 1967
- W. LUCAS, *TIG AND PLASMA WELDING,* ABINGTON PUBLISHING, 1990
- *ASM HANDBOOK,* VOL 1, ASM INTERNATIONAL, 1990
- *MIG WELDING HANDBOOK,* L-TEC WELDING AND CUTTING SYSTEMS, 1987
- *SUBMERGED ARC WELDING HANDBOOK,* L-TEC WELDING AND CUTTING SYSTEMS, 1987
- *TIG WELDING HANDBOOK,* L-TEC WELDING AND CUTTING SYSTEMS, 1987
- *WELDING HANDBOOK,* 7TH ED., VOL 4, AMERICAN WELDING SOCIETY, 1982

Welding of Low-Alloy Steels

F.J. Winsor, Welding Consultant

Introduction

THE DETERMINATION of appropriate procedures and practices for low-alloy steels (those containing up to 10% Cr) involves a systematic evaluation of factors such as:

- DESIGN REQUIREMENTS
- AVAILABILITY AND PRACTICABILITY OF USING A PARTICULAR ARC WELDING PROCESS
- PRODUCTIVITY REQUIREMENTS
- CHARACTERISTICS OF FILLER METALS AND OTHER CONSUMABLES
- PROPERTIES OF WELD DEPOSITS
- PREHEATING AND POSTWELD HEAT-TREATMENT REQUIREMENTS DETERMINED BY THE METALLURGICAL CHARACTERISTICS OF THE LOW-ALLOY STEEL

When compared with other steels, the arc welding of low-alloy steels can require greater emphasis on some of these factors, because of unique mechanical property or service requirements, as well as metallurgical characteristics, such as hardenability and hydrogen embrittlement.

This article discusses factors involved in selecting welding processes and consumables and establishing procedures and practices for the arc welding of low-alloy steels. Because of their different design or service requirements and their often different welding characteristics, the low-alloy steels discussed in this article are categorized as:

- HIGH-STRENGTH LOW-ALLOY (HSLA) STRUCTURAL STEELS
- HIGH-STRENGTH LOW-ALLOY QUENCHED AND TEMPERED (HSLA Q&T) STRUCTURAL STEELS
- LOW-ALLOY STEELS FOR PRESSURE VESSELS AND PIPING
- MEDIUM-CARBON HEAT-TREATABLE (QUENCHED AND TEMPERED) LOW-ALLOY (HTLA) STEELS
- ULTRAHIGH-STRENGTH LOW-ALLOY STEELS
- LOW-ALLOY TOOL AND DIE STEELS

Welding Processes. All of the common arc welding processes can be used with low-alloy steels. Shielded metal arc, gas-tungsten arc, gas-metal arc, flux-cored arc, and submerged arc welding processes are used for most applications.

Electroslag and electrogas processes can be used to weld some of the low-alloy structural and pressure vessel steels, although restrictions may be imposed by some codes.

Welding of Low-Alloy Steels

F.J. Winsor, Welding Consultant

Welding Consumables

General information is provided below in terms of filler metals and fluxes and shielding gases.

Filler Metals and Fluxes

The selection of consumables for the arc welding of low-alloy steels is greatly facilitated by reference to American National Standards Institute/American Welding Society (ANSI/AWS) Filler Metal Specifications (Ref 1). These specifications define the requirements for classifying welding consumables in accordance with established parameters for chemical composition, mechanical properties, and usability. The ANSI/AWS specifications are often incorporated into the materials requirements of fabrication codes.

Shielding Gases

Gas-Tungsten Arc Welding (GTAW) and Plasma Arc Welding (PAW). The shielding gas most commonly used in the GTAW and PAW of low-alloy steels is 100% welding grade (99.995%) argon. Although argon is commonly used when gas backing is specified, nitrogen can also be used satisfactorily as a backing gas for many applications.

Gas-Metal Arc Welding (GMAW), Flux-Cored Arc Welding (FCAW), and Electrogas Welding (EGW). The selection of shielding gases used in these processes often depends on filler metal characteristics. Information should be obtained from the filler metal manufacturer or supplier regarding the proper shielding gas or mixture to be used. Use of a shielding gas other than that which is recommended can result in inferior welding performance or unsatisfactory weld metal chemical composition and/or properties.

Argon-oxygen and argon-carbon dioxide mixtures are commonly used for the GMAW and FCAW of low-alloy steels. Either argon-helium-carbon dioxide mixtures or 100% carbon dioxide are used for low-heat-input welding that utilizes small-diameter electrodes and a short-circuiting mode of metal transfer. Carbon dioxide shielding can also be used with larger-diameter electrodes, when its characteristic high spatter level can be tolerated and its resulting weld metal mechanical properties are acceptable.

The type of shielding gas has a significant effect on the mechanical properties of welds in low-alloy steels. A change in shielding from 100% CO_2 to a mixture of 75% Ar and 25% CO_2 generally results in a 10% increase in tensile strength of the weld metal in the as-deposited condition. This is due to greater retention of strengthening elements, such as manganese, silicon, and chromium, in the weld metal when the oxidizing effect of the arc atmosphere is lowered by the decrease in percentage of carbon dioxide. Depending on the original carbon content of the electrode, a carbon dioxide shielding atmosphere can result in either an increase or decrease in the carbon content of undiluted weld metal. If the carbon content of the electrode is less than 0.05%, then the weld metal tends to pick up carbon from the CO_2 shielding gas. If the carbon content of the electrode is greater than 0.10%, then the weld metal tends to lose carbon.

Reference cited in this section

1. "ANSI/AWS FILLER METAL SPECIFICATIONS," AWS, 1979-1992

Welding of Low-Alloy Steels

F.J. Winsor, Welding Consultant

High-Strength Low-Alloy (HSLA) Structural Steels

The HSLA structural steels typically contain less than 0.2% C and less than 2% alloy. They are different from other alloy steels in the sense that they are designed to meet specific mechanical property requirements, rather than a specified chemical composition. The steels have minimum yield strengths that range from 275 to 550 MPa (40 to 80 ksi) and tensile strengths that range from 415 to 655 MPa (60 to 95 ksi). ASTM specifications for typical HSLA as-rolled pearlitic structural steels include A 242, A 440, A 441, A 572, A 588, A 606, A 607, A 618, A 633, A 656, A 690, A 710, and A 715. Other HSLA steels are covered by Society of Automotive Engineers (SAE) Recommended Practice J410c. Some HSLA structural steels contain copper, phosphorus, nickel, chromium, and silicon in various combinations and amounts that provide a level of corrosion resistance that is up to four times better than that of carbon steel. The chemical compositions of typical ASTM HSLA structural steels are given in Table 1.

TABLE 1 TYPICAL COMPOSITION AND STRENGTH PROPERTIES OF HSLA STRUCTURAL STEELS

ASTM SPECIFICATION	**TYPE OR GRADE**	**COMPOSITION, %(A)**										**MINIMUM TENSILE STRENGTH**		MINIMUM YIELD STRENGTH	
		C	**Mn**	**P**	**S**	**Si**	**Cr**	**Ni**	**Mo**	**V**	**OTHER**	**MPA**	**KSI**	**MPA**	KSI
A 242	1	0.15	1.00	0.15	0.05	. . .	. . .	. . .	. . .	. . .	0.20 MIN CU	435-485	63-70	290-345	**42-50**
A 441	1	0.22	0.85-1.25	0.04	0.05	0.30	. . .	. . .	. . .	. . .	0.20 MIN CU; 0.02 MIN V	415-485	60-70	275-345	**40-50**
A 572	42(B)	0.21	1.35	0.04	0.05	0.30	. . .	. . .	. . .	. . .	0.20 MIN CU	415	60	290	**42**
	50(B)	0.23	1.35	0.04	0.05	0.30	. . .	. . .	. . .	. . .	0.20 MIN CU	450	65	345	**50**
	60(B)	0.26	1.35	0.04	0.05	0.30	. . .	. . .	. . .	. . .	0.20 MIN CU	515	75	415	**60**
	65(B)	0.26	1.65	0.04	0.05	0.30	. . .	. . .	. . .	. . .	0.20 MIN CU	550	80	450	**65**
A 588	A	0.10-0.19	0.90-1.25	0.04	0.05	0.15-0.30	0.40-0.65	. . .	. . .	0.02-0.10	0.25-0.40 CU	435-485	63-70	290-345	**42-50**
	B	0.20	0.75-1.25	0.04	0.05	0.15-0.30	0.40-0.70	0.25-0.50	. . .	0.01-0.10	0.20-0.40 CU	435-485	63-70	290-345	**42-50**
	C	0.15	0.80-1.35	0.04	0.05	0.15-0.30	0.30-0.50	0.25-0.50	. . .	0.01-0.10	0.20-0.50 CU	435-485	63-70	290-345	**42-50**
	D	0.10-0.20	0.75-1.25	0.04	0.05	0.50-0.90	0.50-0.90	. . .	. . .	. . .	0.30 CU; 0.05-0.15 ZR; 0.04 NB	435-485	63-70	290-345	**42-50**
	E	0.15	1.20	0.04	0.05	0.15-0.30	. . .	0.75-1.25	0.10-0.25	0.05	0.50-0.80 CU	435-485	63-70	290-345	**42-50**
	F	0.10-0.20	0.50-1.00	0.04	0.05	0.30	0.30	0.40-1.10	0.10-0.20	0.01-0.10	0.30-1.00 CU	435-485	63-70	290-345	**42-50**
	G	0.20	1.20	0.04	0.05	0.25-0.70	0.50-1.00	0.80	0.10	. . .	0.30-0.50 CU; 0.07 TI	435-485	63-70	290-345	**42-50**
	H	0.20	1.25	0.035	0.040	0.25-0.75	0.10-0.25	0.30-0.60	0.15	0.02-0.10	0.20-0.35 CU; 0.005-0.030 TI	435-485	63-70	290-345	**42-50**
	J	0.20	0.60-1.00	0.04	0.05	0.30-0.50	. . .	0.50-0.70	. . .	. . .	0.30 MIN CU; 0.03-0.05 TI	435-485	63-70	290-345	**42-50**
A 633	A	0.18	1.00-1.35	0.04	0.05	0.15-0.50	. . .	. . .	. . .	. . .	0.05 NB	435-570	63-83	290	**42**
	C	0.20	1.15-1.50	0.04	0.05	0.15-0.50	. . .	. . .	. . .	. . .	0.01-0.05 NB	450-620	65-90	315-345	**46-50**

	D	0.20	0.70-1.60	0.04	0.05	0.15-0.50	0.25	0.25	0.08	. . .	0.35 CU	450-620	65-90	315-345	**56-50**
	E	0.22	1.15-1.50	0.04	0.05	0.15-0.50	. . .	. . .	. . .	0.04-0.11	0.01-0.03 N	515-700	75-100	380-415	**55-50**
A 710	A	0.07	0.40-0.70	0.025	0.025	0.35	0.60-0.90	0.70-1.00	0.15-0.25	. . .	1.00-1.30 CU; 0.02 MIN NB	450-620	65-90	380-585	**55-85**
	B	**0.06**	**0.40-0.65**	**0.025**	**0.025**	**0.20-0.35**	**. . .**	**1.20-1.50**	**. . .**	**. . .**	**1.00-1.30 CU; 0.02 MIN NB**	**605-620**	**88-90**	**515-585**	**75-85**

Source: Ref 2

(A) SINGLE VALUES ARE MAXIMUM UNLESS OTHERWISE NOTED.

(B) THESE GRADES MAY CONTAIN NIOBIUM, VANADIUM, OR NITROGEN.

Filler Metal Selection

The predominant factor in most applications of HSLA steels is tensile strength; the alloy content is incidental. Consequently, filler metal is selected, most frequently, on the basis of providing a weld metal with minimum yield and tensile strength values that are at least equal to those of the base metal. However, HSLA steels are also often used because of their resistance to atmospheric corrosion. Such applications require that the alloy content and corrosion resistance of the weld deposit be considered.

Shielded Metal Arc Welding (SMAW). Some types of carbon steel electrodes (classified in ANSI/AWS specification A5.1) used in this process will deposit weld metal that has mechanical properties that are adequate to match those of some of the HSLA structural steels. Low-alloy steel covered electrodes are classified in ANSI/AWS specification A5.5. Table 2 shows the welding consumables for HSLA structural steels, based on ANSI/AWS D1.1 (Ref 3). This specification requires the use of low-hydrogen filler metals for welding HSLA structural steels in order to satisfy the conditions established for the prequalification of welding procedures.

TABLE 2 GUIDELINES FOR SELECTING CONSUMABLES TO WELD HSLA STRUCTURAL STEELS

PROCESS	**FILLER METAL**		BASE MATERIAL			
			ASTM SPECIFICATION	**THICKNESS**		GRADE
	ANSI/AWS SPECIFICATION	**CONSUMABLE(A)**		**MM**	**IN.**	
SMAW	A5.1	E7015; E7016; E7018; E7028; E7048	A 242	≤100	≤4	**. . .**
			A 441	ALL	ALL	**. . .**
			A 572	. . .	. . .	**42; 50**
			A 588	≤100	≤4	**. . .**
			A 633	≤65	≤2 $\frac{1}{2}$	**A, B, C, D**
	A5.5	E8015-XX; E8016-XX; E8018-XX(B)	A 572	. . .	. . .	**60, 65**
			A 633	. . .	. . .	**E**
SAW	A5.17	F7XXX-EXXX	A 242	≤100	≤4	**. . .**
			A 441	ALL	ALL	**. . .**
			A 572	. . .	. . .	**42; 50**
			A 588	≤100	≤4	**. . .**
			A 633	≤65	≤2 $\frac{1}{2}$	**A, B, C, D**
	A5.23	F8XX-EXXX-XX(B)	A 572	. . .	. . .	**60; 65**
			A 633	. . .	. . .	**E**
GMAW	A5.18	ER70S-X	A 242	≤100	≤4	**. . .**
			A 441	ALL	ALL	**. . .**
			A 572	. . .	. . .	**42; 50**
			A 588	≤100	≤4	**. . .**
			A 633	≤65	≤2 $\frac{1}{2}$	**A, B, C, D**
	A5.28	ER80S-XX(B)	A 572	. . .	. . .	**60; 65**
			A 633	. . .	. . .	**E**
FCAW	A5.20	E7XT-1; E7XT-4; E7XT-5; E7XT-6; E7XT-7; E7XT-8; E7XT-11; E7XT-G	A 242	≤100	≤4	**. . .**
			A 441	ALL	ALL	**. . .**
			A 572	. . .	. . .	**42; 50**
			A 588	≤100	≤4	**. . .**
			A 633	≤65	≤2 $\frac{1}{2}$	**A, B, C, D**
	A5.29	E8XTX-XX(B)	A 572	. . .	. . .	**60; 65**

			A 633	**. . .**	**. . .**	**E**

Source: Ref 3

(A) CONSULT THE APPROPRIATE ANSI/AWS FILLER METAL SPECIFICATION FOR INFORMATION ON THE SIGNIFICANCE OF "X" USED IN THE FILLER METAL DESIGNATIONS. OTHER CLASSIFICATIONS OF FILLER METAL (FOR EXAMPLE, A5.5; E80XX-XX, E7018-W, OR E8018-W COVERED ELECTRODES), AND CORRESPONDING ELECTRODES FOR SAW, GMAW, AND FCAW PROCESSES MAY BE REQUIRED TO ACHIEVE NOTCH TOUGHNESS REQUIREMENTS OR TO MATCH ATMOSPHERIC CORROSION AND WEATHERING CHARACTERISTICS.

(B) FOR BRIDGES, THE WELD METAL SHALL HAVE A MINIMUM CHARPY V-NOTCH IMPACT STRENGTH OF 27 J (20 FT · LBF) AT -20 °C (0 °F).

The low-temperature impact properties, as well as the tensile strength of the weld deposit, can be an important factor when selecting filler metals for the welding of HSLA steels. Various electrodes in ANSI/AWS specifications A5.1 and A5.5 have minimum Charpy impact test requirements at temperatures ranging from -20 to -100 °C (0 to -150 °F). In addition to relying on either the requirements of electrode specifications or typical test results furnished by electrode producers, it is often necessary to run procedure qualification tests to determine the impact properties of the weld metal and heat-affected zones (HAZs), because of code requirements or other considerations.

As indicated in footnote (a) of Table 2, low-alloy steel weld metal compositions are available to match the resistance to atmospheric weathering afforded by some HSLA structural steels. Because of the risk of hydrogen cracking when welding HSLA structural steels, most steel producers recommend the use of low-hydrogen electrodes. However, the use of electrodes that are not low in hydrogen may be acceptable in some applications, if the parts are relatively thin and the joints are not restrained or if a higher preheating temperature can be used.

Gas-Metal Arc Welding (GMAW) and Gas-Tungsten Arc Welding (GTAW). Suggested filler metals for the GMAW of low-alloy structural steels are given in Table 2. Similar filler metals can also be employed for GTAW, although this process is not often used for structural welding. The chemical compositions of the low-alloy steel filler metals included in ANSI/AWS specification A5.28 are generally comparable to the weld metal compositions provided by low-alloy steel SMAW electrodes. Some of the filler metals identified in this specification have minimum impact property requirements. This information is useful when selecting filler metals for application where low-temperature notch ductility is important.

Flux-Cored Arc Welding (FCAW). Flux-cored electrodes used for welding HSLA structural steels are classified in ANSI/AWS specifications A5.20 and A5.29. Weld metal compositions are generally comparable to those available from SMAW electrodes. Several classifications of flux-cored electrodes also have specified impact test requirements. Some types of flux-cored electrodes (self-shielding) require no external gas shielding. E80T1-W electrodes provide weld metal that has weathering characteristics similar to those of some low-alloy structural steels. Suggested electrodes for the FCAW of low-alloy structural steels are given in Table 2.

Submerged Arc Welding (SAW). Electrodes and fluxes used for the SAW of HSLA structural steels are classified in ANSI/AWS specifications A5.17 and A5.23. Inasmuch as numerous different electrode/flux combinations can be suitable for a specific application, it is not possible to suggest in this article anything other than broad categories, as shown in Table 2. When selecting electrodes and fluxes, it is often necessary to refer to producer literature for information on available electrode/flux combinations, performance characteristics of fluxes, and typical properties of weld deposits, as well as the electrode/flux specifications. In addition, procedure development tests may be necessary in order to determine the results obtained under representative fabricating conditions.

For the welding of HSLA steels exposed to outdoor environments in the unpainted condition, SAW compositions that match the weathering characteristics, as well as the mechanical properties of several HSLA steels, are available. Producers can offer recommendations on consumables for such applications.

Electroslag Welding (ESW) and Electrogas Welding (EGW). Solid and metal-cored electrodes used in the ESW of HSLA structural steels are classified in ANSI/AWS specification A5.25. Weld metal compositions are similar to those of many of the electrodes available for GMAW and SAW. Chromium-molybdenum, nickel, or nickel-chromium-copper low-alloy steel weld metals are recommended for exposed, bare, and unpainted applications of ASTM A 588 steel. Classifications EWT2, EWT4, and EWS-EW are useful for other weather-resistant applications.

Flux-cored and solid electrodes classified in ANSI/AWS specification A5.26 for the EGW of HSLA structural steels are similar in composition to many of the flux-cored electrodes classified in ANSI/AWS specification A5.29 and the solid electrodes classified in ANSI/AWS specification A5.28 for GMAW.

The higher-energy input associated with ESW and EGW can result in weld metal and HAZ properties that are significantly different from those obtained with other arc welding processes. To select appropriate consumables, a consultation with producers and welding procedure development tests are recommended.

Welding Procedures and Practices

Preheating. The successful welding of HSLA structural steels requires consideration of a preheat. Generally, the need for preheating increases with increasing carbon and alloy content and with increasing steel thickness. The recommended minimum preheat and interpass temperatures for several HSLA structural steels are given in Table 3.

TABLE 3 MINIMUM PREHEAT AND INTERPASS TEMPERATURES TO WELD HSLA STRUCTURAL STEELS USING LOW-HYDROGEN WELDING PROCESSES

TEMPERATURE AT INDICATED THICKNESS								BASE METAL	
AT ≤20 MM (≤ $\frac{3}{4}$ IN.)		**AT 20-40 MM ($\frac{3}{4}$-1$\frac{1}{2}$ IN.)**		**AT 40-65 MM (1$\frac{1}{2}$-2$\frac{1}{2}$ IN.)**		**AT >65 MM (2$\frac{1}{2}$ IN.)**			
°C	**°F**	**°C**	**°F**	**°C**	**°F**	**°C**	**°F**	**ASTM SPECIFICATION**	GRADE
0	32	10	50	65	150	105	225	A 242	. . .
								A 441	. . .
								A 572	**42**
									50
								A 588	. . .
								A 633	**A**
									B
									C
10	50	65	150	105	225	150	300	A 572	**60**
									65
								A 633	**E**

Source: Ref 3

Information on recommended preheat temperatures for other HSLA structural steels can be obtained from steel producers. If the preheat temperature is not sufficiently high, weld cracking can occur, particularly with highly restrained joints. When different preheat temperatures are specified for the joining of different thicknesses or types of HSLA steels, the higher preheat temperature should be used.

Postweld heat treatment is seldom required for welds in HSLA structural steels. However, it may be necessary to consider postweld heat treatment if:

- THE WELDMENT IS EXPOSED TO AN ENVIRONMENT THAT CAN CAUSE STRESS-CORROSION CRACKING
- CONTROL OF DIMENSIONS DURING MACHINING IS IMPORTANT
- IMPROVED DUCTILITY IS REQUIRED

When postweld heat treatment is required, the vanadium content of the weld metal should be no greater than 0.05%. The recommended postweld heat-treatment range for many HSLA structural steels is from 595 to 675 °C (1100 to 1250 °F). For proprietary steels, the postweld heat-treatment recommendations of the producer should be followed.

References cited in this section

2. CHAPTERS 1, 3, 4, *WELDING HANDBOOK*, 7TH ED., VOL 4, *METALS AND THEIR WELDABILITY*, AWS, 1982
3. "STRUCTURAL WELDING CODE--STEEL," ANSI/AWS D1.1-92, AWS, 1992

Welding of Low-Alloy Steels

F.J. Winsor, Welding Consultant

High-Strength Low-Alloy Quenched and Tempered (HSLA Q&T) Structural Steels

HSLA Q&T structural steels typically contain less than 0.25% C and less than 5% alloy. These steels combine high yield and tensile strengths with good ductility, notch toughness, atmospheric corrosion resistance, and weldability. They are strengthened primarily by quenching and tempering to produce microstructures that contain martensite, bainite, and, in some compositions, ferrite. Some types are produced with a precipitation-hardening treatment that follows hot rolling or quenching. Yield strength values range from approximately 345 to 895 MPa (50 to 130 ksi), and tensile strength values range from 485 to 1035 MPa (70 to 150 ksi), depending on chemical composition, thickness, and heat treatment.

Many of the steels are included in ASTM specifications such as A 514, A 517, and A 543. However, a few steels, such as HY-80, HY-100, and HY-130, are covered by military specifications. The HSLA Q&T steels are furnished primarily as plate, although some are produced as either forgings or castings. The chemical compositions of typical steels are given in Table 4. Heat treatments used by the producers of these steels and the resulting microstructures are shown in Table 5.

TABLE 4 NOMINAL COMPOSITION AND TENSILE PROPERTIES OF SELECTED QUENCHED AND TEMPERED LOW-ALLOY STRUCTURAL STEELS

ASTM SPECIFICATION OR COMMON DESIGNATION	GRADE	COMPOSITION, %(A)										MINIMUM TENSILE STRENGTH(C)		MINIMUM YIELD STRENGTH(C)	
		C	Mn	P	S	Si	Cr	Ni	Mo	Cu	OTHER	MPA	KSI	MPA	KSI
A 514; A 517	A	0.15-0.21	0.80-1.10	0.035	0.04	0.40-0.80	0.50-0.80	. . .	0.18-0.28	. . .	ZR, 0.05-0.15; B, 0.0025	700-895; 725-930	100-130; 105-135	620-700	**90-100**
	B	0.12-0.21	0.70-1.00	0.035	0.04	0.20-0.35	0.40-0.65	. . .	0.15-0.25	. . .	V, 0.03-0.08; TI, 0.01-0.03; B, 0.0005-0.005	700-895; 725-930	100-130; 105-135	620-700	**90-100**
	C	0.10-0.20	1.10-1.50	0.035	0.04	0.15-0.30	. . .	. . .	0.20-0.30	. . .	B, 0.001-0.005	700-895; 725-930	100-130; 105-135	620-700	**90-100**
	D	0.13-0.20	0.40-0.70	0.035	0.04	0.20-0.35	0.85-1.20	. . .	0.15-0.25	0.20-0.40	TI, 0.004-0.10(B); B, 0.0015-0.005	700-895; 725-930	100-130; 105-135	620-700	**90-100**
	E	0.12-0.20	0.40-0.70	0.035	0.04	0.20-0.35	1.40-2.00	. . .	0.40-0.60	0.20-0.40	TI, 0.04-0.10(A); B, 0.0015-0.005	700-895; 725-930	100-130; 105-135	620-700	**90-100**
	F	0.10-0.20	0.60-1.00	0.035	0.04	0.15-0.35	0.40-0.65	0.70-1.00	0.40-0.60	0.15-0.50	V, 0.03-0.08; B, 0.0005-0.006	700-895; 725-930	100-130; 105-135	620-700	**90-100**
	G	0.15-0.21	0.80-1.10	0.035	0.04	0.50-0.90	0.50-0.90	. . .	0.40-0.60	. . .	ZR, 0.05-0.15; B, 0.0025	700-895; 725-930	100-130; 105-135	620-700	**90-100**
	H	0.12-0.21	0.95-1.30	0.035	0.04	0.20-0.35	0.40-0.65	0.30-0.70	0.20-0.30	. . .	V, 0.03-0.08; B, 0.0005-0.005	700-895; 725-	100-130; 105-	620-700	**90-100**

											930	135		
	J	0.12-0.21	0.45-0.70	0.035	0.04	0.20-0.35	. . .	. . .	0.50-0.65	. . .	B, 0.001-0.005	700-895; 725-930	100-130; 105-135	620-700	**90-100**
	K	0.10-0.20	1.10-1.50	0.035	0.04	0.15-0.30	. . .	. . .	0.45-0.55	. . .	B, 0.001-0.005	700-895; 725-930	100-130; 105-135	620-700	**90-100**
	L	0.13-0.20	0.40-0.70	0.035	0.04	0.20-0.35	1.15-1.65	. . .	0.25-0.40	0.20-0.40	TI, 0.04-0.10(B); B, 0.0015-0.005	700-895; 725-930	100-130; 105-135	620-700	**90-100**
	M	0.12-0.21	0.45-0.70	0.035	0.04	0.20-0.35	. . .	1.20-1.50	0.45-0.60	. . .	B, 0.001-0.005	700-895; 725-930	100-130; 105-135	620-700	**90-100**
	N	0.15-0.21	0.80-1.10	0.035	0.04	0.40-0.90	0.50-0.80	. . .	0.25	. . .	ZR, 0.05-0.15; B, 0.0005-0.0025	700-895; 725-930	100-130; 105-135	620-700	**90-100**
	P	0.12-0.21	0.45-0.70	0.035	0.04	0.20-0.35	0.85-1.20	1.20-1.50	0.45-0.60	. . .	B, 0.001-0.005	700-895; 725-930	100-130; 105-135	620-700	**90-100**
	Q	0.14-0.21	0.95-1.30	0.035	0.04	0.15-0.35	1.00-1.50	1.20-1.50	0.40-0.60	. . .	V, 0.03-0.08	700-895; 725-930	100-130; 105-135	620-700	**90-100**
A 543	B	0.23	0.40	0.020	0.020	0.20-0.40	1.50-2.00	2.60-4.00	0.45-0.60	. . .	V, 0.03	620-930	90-135	480-700	**70-100**
	C	0.23	0.40	0.020	0.020	0.20-0.40	1.20-1.50	2.25-3.50	0.45-0.60	. . .	V, 0.03	620-930	90-135	480-700	**70-100**
HY-80	. . .	0.12-0.18	0.10-0.40	0.025	0.025	0.15-0.35	1.00-1.80	2.00-3.25	0.20-0.60	0.25	V, 0.03; TI, 0.02	. . .	. . .	550	**80**
HY-100	. . .	0.12-0.20	0.10-0.40	0.025	0.025	0.15-0.35	1.00-1.80	2.25-3.50	0.20-0.60	0.25	V, 0.03; TI, 0.02	. . .	. . .	700	**100**

HY-130	**. . .**	**0.12**	**0.60-0.90**	**0.010**	**0.015**	**0.15-0.35**	**0.40-0.70**	**4.75-5.25**	**0.30-0.65**	**. . .**	**V, 0.05-0.10**	**. . .**	**. . .**	**895**	**130**

Source: Ref 2

(A) WHEN A SINGLE VALUE IS SHOWN, IT IS A MAXIMUM LIMIT.

(B) VANADIUM MAY BE SUBSTITUTED FOR PART OR ALL OF TITANIUM CONTENT ON A ONE-FOR-ONE BASIS.

(C) LIMITING VALUES VARY WITH THE PLATE THICKNESS.

TABLE 5 TYPICAL HEAT TREATMENTS FOR QUENCHED AND TEMPERED LOW-ALLOY STRUCTURAL STEELS THAT YIELD BAINITE AND MARTENSITE MICROSTRUCTURES

BASE METAL ASTM SPECIFICATION OR COMMON DESIGNATION	**AUSTENITIZING TEMPERATURE**		**QUENCHING MEDIUM**	TEMPERING TEMPERATURE	
	°C	**°F**		°C	°F
A 514; A 517	900	1650	WATER OR OIL	620	**1150**
A 543, TYPE B	900	1650	WATER	540	**1000**
HY-80; HY-100	900	1650	WATER	650	**1200**
HY-130	**815**	**1500**	**WATER**	**540**	**1000**

Welding Processes

Any of the common arc welding processes can be used to join the quenched and tempered low-alloy structural steels that have minimum yield strengths of ≤ 1035 MPa (≤ 150 ksi). However, the GMAW and GTAW processes are preferred for welding steels with minimum yield strengths of [ges]700 MPa ([ges]100 ksi), such as HY-130, because they afford a higher level of notch toughness in the weld metal.

Filler Metals

Welding filler metals and electrode/flux combinations are available for welding most of the HSLA Q&T structural steels. Many of these consumables are classified in ANSI/AWS filler metal specifications for low-alloy steels. The selection of consumables is usually based on factors such as tensile strength, composition, and notch toughness of the weld metal. For some steels, more than one type or classification of welding consumable may satisfy the fabrication requirements.

The selection of welding consumables for a specific application should be based on a thorough evaluation of the design requirements and knowledge of the available weld metal properties, such as tensile strength, ductility, and impact strength. In general, the deposited weld metal should have properties that either match or exceed those of the base metal. However, for welds that are subject to relatively low stresses, the use of a weld metal that has a lower strength than that of the base metal may be acceptable. Typical examples are fillet welds stressed in longitudinal shear, welds carrying low secondary stresses, and welds that join a lower-strength steel to a HSLA Q&T steel. In cases where the design permits, lower-strength weld metal can be advantageous, allowing highly restrained welds to resist the tendency for toe cracking (that is, cracking at the toe of a fillet weld) or lamellar tearing in the steel, especially with corner and T-joints.

Weld Metal Hydrogen

The weld metal hydrogen content must be maintained at very low levels when welding HSLA Q&T steels, in order to avoid hydrogen-induced cracking. The maximum acceptable moisture, or hydrogen content, of covered electrodes, bare filler metal, and fluxes decreases as the strength of the weld metal increases. Welding consumables must be produced, packaged, stored, and used in a manner that limits the weld metal hydrogen content to a low level.

Weld Design

In order to employ HSLA Q&T structural steels advantageously, it is necessary to use appropriate joint designs, careful workmanship, and adequate inspection. Weld joints should be designed and located to avoid abrupt changes in cross section or other stress raisers in regions of high stress. Joints should permit good accessibility for welding and inspection. Groove welds are preferred over fillet welds for many applications, because of inherent stress concentrations associated with fillet welds and because groove welds are more amenable to volumetric methods of examination, such as radiographic or ultrasonic testing. The weld metal should blend smoothly into the base metal, with no abrupt ridges or valleys, and excessive reinforcement should be avoided.

Preheat and Interpass Temperature Control

The successful welding of HSLA Q&T structural steels often requires preheating in order to prevent cracking. However, caution must be exercised, because the preheat temperature affects the cooling rate during welding, thereby possibly affecting the microstructures and properties of the weld metal and HAZs. A preheat or interpass temperature that is too high can decrease the rate of cooling of the weld metal and HAZs sufficiently to result in the transformation of austenite to either ferrite (containing regions of high-carbon martensite) or coarse bainite. Both structures lack high strength and good toughness. Next to the reaustenitized portion of the HAZ, a region of previously tempered base metal may be overtempered, causing a loss in strength.

Suggested preheat and interpass temperature ranges for some HSLA Q&T steels are given in Table 6. Information on preheating can be obtained from producers of HSLA Q&T steels. Exceeding the recommended maximum temperature should be avoided. If a preheat temperature of 40 °C (100 °F) or less is used, precautions must be taken to ensure that the surfaces to be welded are dry.

TABLE 6 TYPICAL PREHEAT AND INTERPASS TEMPERATURE RANGES TO WELD QUENCHED AND TEMPERED HSLA STRUCTURAL STEELS

BASE METAL ASTM SPECIFICATION OR COMMON DESIGNATION	**THICKNESS**		PREHEAT AND INTERPASS TEMPERATURE RANGE	
	MM	**IN.**	**°C**	°F
A 514; A517	≤20	≤ $\frac{3}{4}$	10-95	**50-200**
	>20-40	> $\frac{3}{4}$-1$\frac{1}{2}$	50-135	**125-275**
	>40-65	>1$\frac{1}{2}$-2$\frac{1}{2}$	80-165	**175-325**
	>65	>2$\frac{1}{2}$	105-190	**225-375**
A 543	≤12.5	≤ $\frac{1}{2}$	40-120	**100-250**
	>12.5-20	> $\frac{1}{2}$-$\frac{3}{4}$	50-135	**125-275**
	>20-25	> $\frac{3}{4}$-1	65-150	**150-300**
	>25-50	>1-2	95-175	**200-350**
	>50	>2	150-230	**300-450**
HY-80, HY-100	≤12.5	≤ $\frac{1}{2}$	15-150	**60-300**
	>12.5-28.7	> $\frac{1}{2}$-1.13	50-150	**125-300**
	>28.9	>1.14	95-150	**200-300**
HY-130	≤16	≤0.63	25-65	**75-150**
	16.3-22.4	0.64-0.88	50-95	**125-200**
	22.6-35.1	0.89-1.38	95-135	**200-275**
	>35.1	**>1.38**	**105-150**	**225-300**

Heat Input

In addition to preheat and interpass temperature, the heat (energy) input during welding and the base metal thickness are factors that affect the cooling rate, and thereby affect the microstructures and mechanical properties of the weld metal and HAZs. As is the case with too high a preheat and interpass temperature, excessive heat input can decrease the rate of cooling sufficiently to produce unfavorable microstructures and properties. Welding heat input is of greater concern in steels with a minimum yield strength of 700 MPa (100 ksi) than it is in lower-strength HSLA Q&T steels.

The welding heat input can be calculated as follows:

$$H = \frac{E.I.60}{S.1000} \qquad \textbf{(EQ 1)}$$

where H is heat input (arc energy), in units of kJ/in.; E is arc voltage, in V; I is welding current, in A; and S is arc travel speed, in units of in./min.

Table 7 shows the maximum permissible welding heat input for butt joints in ASTM A 514/A 517, grade F steel, depending on the preheat temperature and section thickness. Exceeding the maximum specified heat input can result in undesirable microstructures and unacceptable mechanical properties. Because of the different metallurgical characteristics that are possible, other types of steel may require different heat input limitations.

TABLE 7 MAXIMUM WELDING HEAT INPUT FOR BUTT JOINTS IN ASTM SPECIFICATION A 514 AND A 517 GRADE F QUENCHED AND TEMPERED HSLA STEEL

PREHEAT TEMPERATURE		MAXIMUM HEAT INPUT FOR INDICATED SECTION THICKNESS(A)															
		AT 4.8 MM (0.19 IN.)		**AT 6.35 MM (0.25 IN.)**		**AT 12.7 MM (0.5 IN.)**		**AT 19.1 MM (0.75 MIN.)**		**AT 25.4 MM (1.0 IN.)**		**AT 31.8 MM (1.25 IN.)**		**AT 38.1 MM (1.5 IN.)**		AT 50.8 MM (2.0 IN.)	
°C	**°F**	**KJ/MM**	**KJ/IN.**	**KJ/MM**	**KJ/IN.**	**KJ/MM**	**KJ/IN.**	**KJ/MM**	**KJ/IN.**	**KJ/MM**	**KJ/IN.**	**KJ/MM**	**KJ/IN.**	**KJ/MM**	**KJ/IN.**	**KJ/MM**	KJ/IN.
20	70	1.1	27	1.4	36	2.8	70	4.8	121	(B)	(B)	(B)	(B)	(B)	(B)	(B)	**(B)**
95	200	0.83	21	1.1	29	2.2	56	3.9	99	6.8	173	(B)	(B)	(B)	(B)	(B)	**(B)**
150	300	0.67	17	0.94	24	1.9	47	3.2	82	5.0	126	6.9	175	(B)	(B)	(B)	**(B)**
205	**400**	**0.51**	**13**	**0.75**	**19**	**1.6**	**40**	**2.6**	**65**	**3.7**	**93**	**5.0**	**127**	**6.5**	**165**	**(B)**	**(B)**

Source: Ref 2

(A) FOR INTERMEDIATE PREHEAT TEMPERATURES AND THICKNESSES, INTERPOLATE TO OBTAIN HEAT INPUT LIMIT. THIS LIMIT CAN BE INCREASED 25% FOR FILLET WELDS.

(B) NO HEAT INPUT LIMIT FOR MOST ARC WELDING PROCESSES. HIGH HEAT INPUT PROCESSES SHOULD NOT BE USED.

Heat input limitations apply to each weld pass and are not considered cumulatively. They also limit the size of weld beads. Large weld beads typically have poor notch toughness. Therefore, a good welding practice for Q&T steels is to deposit many small stringer beads. This technique produces a weld metal that has good notch toughness as a result of the grain refining and tempering effect of successive weld passes.

These heat input limitations are applicable only for single-arc welding processes. Multiple-arc processes, with the arcs in tandem, generally do not allow sufficient time for the metal deposited by the first arc to cool before the second arc passes over it, supplying additional heat. This can significantly influence the transformation of austenite in the weld metal and HAZs and adversely affect the weld properties. Multiple-arc processes should not be used to weld HSLA Q&T steels prior to conducting welding procedure development tests to prove their suitability.

The cooling rates of welds made using the SMAW, GMAW, FCAW, and GTAW processes are usually sufficiently rapid, under normal conditions, so that the mechanical properties of the HAZ approach those of the unaffected quenched and tempered base metal. The use of high heat input welding processes, such as ESW, EGW, or multiple-wire SAW, generally should be avoided when welding HSLA Q&T steels. Such processes may require complete heat treatment (quenching and tempering) of the weldment in order to obtain acceptable mechanical properties.

Postweld Heat Treatment

Postweld heat treatment is seldom required for welds in HSLA Q&T structural steels, except where heavy sections are involved. The mechanical properties of either the base metal or the weld metal can be adversely affected by postweld heat treatment. When weldments of some HSLA Q&T structural steels are heated to above 510 °C (950 °F), intergranular cracking can occur in the coarse-grained region of the weld HAZ. Cracking occurs by stress rupture, usually early in the heat-treatment cycle.

In situations where postweld heat treatment is necessary, some practices may be used singly or in combination to minimize cracking during postweld heat treatment. These practices involve:

- SELECTING A WELD LOCATION, WELD JOINT DESIGN, AND ASSEMBLY SEQUENCE THAT WILL MINIMIZE RESIDUAL STRESSES AND STRESS CONCENTRATIONS
- USING A TYPE OF WELD METAL THAT HAS SIGNIFICANTLY LOWER STRENGTH THAN THAT OF THE HAZ AT THE POSTWELD HEAT-TREATMENT TEMPERATURE
- PEENING EACH WELD LAYER, EXCEPT THE ROOT AND FINAL LAYERS
- HEATING TO THE POSTWELD HEAT-TREATMENT TEMPERATURE RANGE AS RAPIDLY AS POSSIBLE

In the postweld heat treatment of quenched and tempered steels, the mill tempering temperature used by the steel producer should not be exceeded. To avoid unnecessary risks, the steel producer should be consulted prior to reaching a decision regarding postweld heat treatment of a HSLA Q&T structural steel.

Of the available HSLA Q&T structural steels, only ASTM A 514 and A 517 are customarily given postweld heat treatments when the section thickness exceeds a specified amount.

Reference cited in this section

2. CHAPTERS 1, 3, 4, *WELDING HANDBOOK*, 7TH ED., VOL 4, *METALS AND THEIR WELDABILITY*, AWS, 1982

Welding of Low-Alloy Steels

F.J. Winsor, Welding Consultant

Low-Alloy Steels for Pressure Vessels and Piping

A variety of low-alloy steels containing from 0.10 to 0.30% C and up to 10% total alloy are used for welded pressure vessels and piping. Alloy additions include manganese, silicon, molybdenum, chromium, nickel, vanadium, copper, and niobium. Table 8 gives the nominal chemical composition and minimum specified tensile strength values of representative low-alloy steels used in pressure vessels and piping. These steels are usually furnished in the as-rolled, as-forged, or annealed condition. Some steels, such as those shown in Table 9, also may be obtained in the normalized, normalized and tempered, or quenched and tempered condition to provide enhanced mechanical properties.

TABLE 8 TYPICAL AS-ROLLED, AS-FORGED, OR ANNEALED LOW-ALLOY STEELS USED FOR PRESSURE VESSELS AND PIPING

PRODUCT TYPE	**BASE METAL ASTM SPECIFICATION**	**GRADE**	**NOMINAL COMPOSITION, WT%**			MAXIMUM SPECIFIED TENSILE STRENGTH	
			Cr	**Mo**	**Ni**	**MPA**	KSI
PLATE	A 203	A	. . .	. . .	2.25	450	**65**
		B	. . .	. . .	2.25	480	**70**
		D	. . .	. . .	3.50	450	**65**
		E	. . .	. . .	3.50	480	**70**
		F (≤50 MM, OR 2 IN.)	. . .	. . .	3.50	550	**80**
		(>50 MM, OR 2 IN.)	. . .	. . .	3.50	515	**75**
	A 204	A	. . .	0.5	. . .	450	**65**
		B	. . .	0.5	. . .	480	**70**
		C	. . .	0.5	. . .	515	**75**
	A 387	2	0.5	0.5	. . .	380	**55**
		12	1	0.5	. . .	380	**55**
		11	1.25	0.5	. . .	415	**60**
		22	2.25	1	. . .	415	**60**
		21	3	1	. . .	415	**60**
		5	5	0.5	. . .	415	**60**
		7	7	0.5	. . .	415	**60**
		9	9	1	. . .	415	**60**
PIPE	A 369	FP1	. . .	0.5	. . .	380	**55**
		FP2	0.5	0.5	. . .	380	**55**
		FP5	5	0.5	. . .	415	**60**
		FP7	7	0.5	. . .	415	**60**
		FP9	9	1	. . .	415	**60**
		FP11	1.25	0.5	. . .	415	**60**
		FP12	1	0.5	. . .	415	**60**
		FP21	3	1	. . .	415	**60**
		FP22	**2.25**	**1**	**. . .**	**415**	**60**

TABLE 9 SELECTED NORMALIZED, NORMALIZED AND TEMPERED, OR QUENCHED AND TEMPERED LOW-ALLOY STEELS USED FOR PRESSURE VESSEL AND PIPING APPLICATIONS

PRODUCT TYPE	**ASTM SPECIFICATION**	**GRADE**	**COMPOSITION, WT%**	MAXIMUM SPECIFIED TENSILE

								STRENGTH	
			Ni	**Cr**	**Cu**	**Mn**	**Mo**	**MPA**	KSI
PIPE FITTINGS	A 420	WPL 9	1.9	. . .	1	. . .	. . .	435	**63**
		WPL 3	3.5	. . .	. . .	. . .	. . .	450	**65**
		WPL 8	9	. . .	. . .	. . .	. . .	690	**100**
FORGINGS	A 508	CLASS 2	0.75	0.35	. . .	. . .	0.65	550	**80**
		CLASS 3	0.70	. . .	. . .	. . .	0.50	550	**80**
		CLASS 4	3.50	1.75	. . .	. . .	0.50	725	**105**
		CLASS 5	3.50	1.75	. . .	. . .	0.50	725	**105**
	A 541	CLASS 2	0.75	0.35	. . .	. . .	0.65	550	**80**
		CLASS 3	0.70	. . .	. . .	. . .	0.50	550	**80**
		CLASS 5	. . .	1.25	. . .	. . .	0.50	550	**80**
		CLASS 6	. . .	2.25	. . .	. . .	1	725	**105**
		CLASS 7	3.35	1.6	. . .	. . .	0.50	725	**105**
		CLASS 8	3.35	1.6	. . .	. . .	0.50	725	**105**
FLAGES, FITTINGS, AND VALVES	A 522	TYPE I	9	. . .	. . .	. . .	. . .	725	**100**
		TYPE II	8	. . .	. . .	. . .	. . .	725	**100**
PLATES	A 533	TYPE A	. . .	. . .	. . .	1.3	0.5	550(A) 620(B) 690(C)	**80(A)** **90(B)** **100(C)**
		TYPE B	0.55	. . .	. . .	1.3	0.5	550(A) 620(B) 690(C)	**80(A)** **90(B)** **100(C)**
		TYPE C	0.85	. . .	. . .	1.3	0.5	550(A) 620(B) 690(C)	**80(A)** **90(B)** **100(C)**
		TYPE D	**0.30**	**. . .**	**. . .**	**1.3**	**0.5**	**550(A)** **620(B)** **690(C)**	**80(A)** **90(B)** **100(C)**

(A)	CLASS 1.
(B)	CLASS 2.
(C)	CLASS 3

The choice of low-alloy steel to be used for a pressure vessel or piping system is usually made on the basis of one or more desired properties in service, such as tensile strength, resistance to hydrogen attack, resistance to creep and oxidation at elevated temperatures, or low-temperature notch ductility. Weldability is an important consideration, as well.

Heat-resisting low-alloy steels used for pressure vessels and pipe generally contain up to 0.25% C, plus alloy additions such as chromium, molybdenum, vanadium, and silicon. When compared with plain carbon steels, the heat-resisting steels offer greater tensile and creep-rupture strengths at elevated temperatures and greater resistance to oxidation and hydrogen attack. Steels with from 1 to 9% Ni provide greater resistance to notch embrittlement than do carbon steels or many other low-alloy steels and are widely used in low-temperature service.

Welding Processes and Practices

All of the common arc welding processes can be used to join low-alloy steels used for pressure vessels and piping. SMAW, GTAW, GMAW, FCAW, and SAW are used for most applications. Successful welding requires consideration of preheating and control of weld metal hydrogen content. The use of low-hydrogen consumables and practices is essential in welding most low-alloy steels that contain more than minor amounts of alloying elements. When notch toughness requirements are important, control of heat input must be considered.

ESW and EGW can be used for many low-alloy steels. However, careful evaluation is required when good notch toughness in the weld metal and HAZ is important. A postweld grain refinement heat treatment may be required by some codes and specifications. Preheating is not normally necessary with these high heat input processes.

Filler Metal Selection

In order to provide weld metals that match or exceed the service performance characteristics of low-alloy steel base metals used in pressure vessels and piping, it is frequently necessary to base the filler metal selection on considerations of chemical composition and mechanical properties. In most cases, the content of critical elements in the weld metal must either match or exceed that of the base metal.

Shielded Metal Arc Welding. Electrodes for the SMAW of low-alloy steels used for pressure vessels and piping are classified in ANSI/AWS specification A5.5. In addition, 5%Cr-0.5%Mo and 9%Cr-1%Mo electrodes are classified in ANSI/AWS specification A5.4. Nearly all welding of low-alloy steels is done using low-hydrogen electrodes. However, electrodes that are not low in hydrogen can be used satisfactorily in a few applications where base metals are relatively thin and/or where higher preheat temperatures are feasible. Examples include the welding of C-Mo steel pipe and tubing and the welding of the root pass of groove welds in low Cr-Mo pipe and tubing. All-position electrodes that are not low in hydrogen may have better usability than low-hydrogen types for out-of-position welding, especially where frequent changes in welding position are necessary, as in the welding of pipe in the fixed horizontal position.

At present, SMAW electrodes that will deposit weld metal that has a nominal content of 9% Ni, to match that of ASTM A 353 or A 333, grade 8, base metal, for example, are not commercially available. To weld this material, it is common industry practice to use either an austenitic stainless steel electrode classified in ANSI/AWS specification A5.4 or a nickel-chromium-iron or nickel-chromium-molybdenum alloy electrode classified in ANSI/AWS specification A5.11. The selection of dissimilar alloy covered electrodes or other consumables for the welding of 9% Ni steels must be done carefully, based on the design tensile strength and low-temperature impact requirements.

For cyclic temperature service, consideration should be given to the possible adverse effect of the greater coefficient of linear thermal expansion of austenitic stainless steel, compared to that of nickel-base alloy weld metals.

Gas-Tungsten Arc Welding and Gas-Metal Arc Welding. Bare filler metals for the gas-shielded arc welding of low-alloy steels used for pressure vessels and piping are classified in ANSI/AWS specifications A5.28 and A5.9 (classifications ER502 and ER505). The chemical compositions of low-alloy steel bare filler metals are generally comparable to the deposits produced by covered electrodes for SMAW. As with covered electrodes, there is no standard matching composition of bare welding filler metal for joining 9% Ni steel. To weld this material, it is common practice to use a stainless steel filler metal classified in ANSI/AWS specification A5.9 or a nickel-chromium or nickel-chromium-iron alloy filler metal classified in ANSI/AWS specification A5.14, selected to provide the required tensile strength and impact properties in the weld metal.

ANSI/AWS specification A5.9 classifies matching ER502 and ER505 filler metals for the welding of 5Cr-0.5Mo and 9Cr-lMo base metals, respectively. Filler metal of 9Cr-1Mo is also used to weld 7Cr-0.5Mo base metal.

Flux-Cored Arc Welding. Low-alloy steel electrodes for the FCAW of low-alloy steels used in pressure vessels and piping are classified in ANSI/AWS specification A5.29. Also, ANSI/AWS specification A5.22 classifies E502T-X and E505T-X flux-cored electrodes for welding 5Cr-0.5Mo, 7Cr-0.5Mo, and 9Cr-1Mo base metals.

Submerged Arc Welding. Electrodes and fluxes used for the SAW of low-alloy steels for pressure vessels and piping are classified in ANSI/AWS specification A5.23. Low-alloy steels are most often welded using a neutral flux and an alloy steel electrode. Occasionally, a carbon steel electrode can be used with a particular alloy flux designed to yield weld metal of the desired alloy composition.

When selecting electrodes and fluxes for submerged arc welding, it is often necessary or desirable to consult producers of electrodes and fluxes for information on the characteristics of their fluxes, recommendations on electrode/flux combinations, and data on typical properties of weld deposits, in addition to referring to ANSI/AWS specification A5.23. Such information can often be of great assistance in narrowing the search for suitable SAW consumables. Most pressure vessel and piping applications require procedure qualification testing to demonstrate that a particular electrode/flux combination will deposit weld metal that has the required chemical composition and mechanical properties.

To avoid hydrogen-induced cracking, care should be taken to maintain the flux and electrode in a dry condition. Producers should be consulted for recommendations concerning the storage and reconditioning of submerged arc welding flux.

Electroslag and Electrogas Welding. Low-alloy steel electrodes for ESW are classified in ANSI/AWS specification A5.25, whereas low-alloy steel electrodes for EGW are classified in ANSI/AWS A5.26. Only a few of the low-alloy steel compositions in these specifications satisfy requirements for matching chemical compositions of low-alloy steels used for pressure vessels. However, in some situations, nonmatching compositions may be acceptable. Fortunately, electrodes classified in various other filler metal specifications are useful for electroslag and electrogas welding. Thus, solid and metal-cored electrodes classified in ANSI/AWS A5.23 (for SAW) and in ANSI/AWS A5.28 (for GMAW) may also be useful for ESW and EGW. Appropriate testing is required to ensure that the required weld metal chemical composition and mechanical properties are obtained.

Welding Procedures and Practices

Preheating. The successful welding of low-alloy steels used for pressure vessels and piping requires the consideration of preheating. The suggested minimum preheat and interpass temperatures for several low-alloy steels are given in Table 10. The appropriate sections of the ASME Boiler and Pressure Vessel Code (Ref 4) or the American Society of Mechanical Engineers (ASME) B31.1 Power Piping Code and B31.3 Chemical Plant and Petroleum Piping Code should be consulted for preheat recommendations or requirements regarding boilers, pressure vessels, and code piping.

TABLE 10 RECOMMENDED MINIMUM PREHEAT AND INTERPASS TEMPERATURES TO WELD LOW-ALLOY STEEL PRESSURE VESSELS AND PIPES

PRODUCT TYPE AND ALLOY CLASSIFICATION	ASTM SPECIFICATION	GRADE	THICKNESS		MINIMUM PREHEAT AND INTERPASS TEMPERATURE		MAXIMUM TENSILE STRENGTH	
			mm	in.	°C	°F	MPA	KSI
PRESSURE VESSEL PLATE								
NICKEL	A 203	A	(A)	(A)	120	250	(B)	**(B)**
		B	(A)	(A)	120	250	(B)	**(B)**
		D	(A)	(A)	150	300	(B)	**(B)**
		E	(A)	(A)	150	300	(B)	**(B)**
		F	(A)	(A)	150	300	(B)	**(B)**
MOLYBDENUM	A 204	. . .	<12.5	<0.5	10	50	485	**70**
					80	175	>485	**>70**
			[GES]12.5	[GES]0.5	80	175	(B)	**(B)**
MANGANESE-	A 302	A; B	<12.5	<0.5	10	50	485	**70**

MANGANESE-MOLYBDENUM NICKEL			[GES]12.5	[GES]0.5	80	175	(B)	**(B)**
9% NICKEL	A 353	. . .	(A)	(A)	10	50	(B)	**(B)**
CHROMIUM-NICKEL	A 387	11; 12	≤12.5	≤0.5	10	50	415	**60**
			>12.5	>0.5	120	250	415	**60**
			(A)	(A)	120	250	>415	**>60**
		5; 22	(A)	(A)	150	300	415	**60**
					205	400	>415	**>60**
PIPE								
SEAMLESS AND WELDED FOR LOW-TEMPERATURE SERVICE	A 333	3	(A)	(A)	150	300	(B)	**(B)**
		7	(A)	(A)	120	250	(B)	**(B)**
		8	(A)	(A)	10	50	(B)	**(B)**
SEAMLESS FERRITIC FOR HIGH-TEMPERATURE SERVICE	A 335	P1	<12.5	<0.5	10	50	485	**70**
					80	175	>485	**>70**
			[GES]12.5	[GES]0.5	80	175	(B)	**(B)**
		P5	(A)	(A)	150	300	415	**60**
					205	400	>415	**>60**
		P11;P12	≤12.5	≤0.5	10	50	415	**60**
			>12.5	>0.5	120	250	415	**60**
			(A)	(A)	120	250	>415	**>60**
		P22	(A)	(A)	150	300	415	**60**
					205	400	>415	**>60**
ELECTRIC-FUSION WELDED FOR HIGH-PRESSURE SERVICE AT MODERATE TEMPERATURES	A 672	H75; H80	<12.5	<0.5	10	50	485	**70**
					80	175	>485	**>70**
			[GES]12.5	**[GES]0.5**	**80**	**175**	**(B)**	**(B)**

(A) ALL AVAILABLE SIZES PER ASTM SPECIFICATIONS.

(B) VALUES VARY WITH SIZE.

Postweld heat treatment is often essential for welds in low-alloy steels used for pressure vessels and piping. Various sections of the ASME Boiler and Pressure Vessel Code (Ref 4), the ASME Piping Codes, and military specifications specify requirements for postheat treatment of welds in low-alloy steels. Frequently, purchasers of equipment refer to proprietary specifications that contain postweld heat-treatment requirements. For some materials with certain categories and thicknesses of joints, the codes or other fabrication specifications may allow exemptions from postweld heat-treatment requirements. Because of the differences in various codes and specifications, it is not feasible to provide a universally adequate and reliable summary here of all postweld heat-treatment requirements and practices for low-alloy steel pressure vessel and pipe welding.

Table 11 presents a general view of some of the different postweld heat-treatment temperatures involved in welding low-alloy steels for pressure vessels and piping. The applicable section of the ASME Boiler and Pressure Vessel Code (Ref 4), the ASME Piping Codes, military specifications, or purchaser specifications should be consulted for specific detailed postweld heat-treatment requirements. When dealing with applications that may not fall within the scope of a code or specification, it is a good policy to consult various codes for information to use in determining postweld heat-treatment practices.

TABLE 11 POSTWELD HEAT TREATMENT TEMPERATURES FOR LOW-ALLOY STEELS USED IN

PRESSURE VESSEL AND PIPE APPLICATIONS

PRODUCTS TYPE AND ALLOY CLASSIFICATION	**ASTM CLASSIFICATION**	**GRADE**	POSTWELD HEAT TREATMENT TEMPERATURE RANGE	
			°C	°F
PRESSURE VESSEL PLATE				
NICKEL	A 203	A; B; D; E; F	595-635	**1100-1175**
MOLYBDENUM	A 204	. . .	595-720	**1100-1325**
MANGANESE-MOLYBDENUM; MANGANESE-MOLYBDENUM NICKEL	A 302	A; B	595-720	**1100-1325**
9% NICKEL	A 353	. . .	550-585	**10265-1085**
CHROMIUM-NICKEL	A 387	11; 12	595-745	**1100-1375**
		5; 22	675-760	**1250-1400**
PIPE				
SEAMLESS AND WELDED FOR LOW-TEMPERATURE SERVICE	A 333	3; 7	595-635	**1100-1175**
		8	550-585	**1025-1085**
SEAMLESS FERRITIC FOR HIGH-TEMPERATURE SERVICE	A 335	P1	595-720	**1100-1325**
		P5; P22	675-760	**1250-1400**
		P11; P12	595-745	**1100-1375**
ELECTRIC-FUSION WELDED FOR HIGH-PRESSURE SERVICE AT MODERATE TEMPERATURES	**A 672**	**H75; H80**	**595-720**	1100-1325

Reference cited in this section

4. "ASME BOILER AND PRESSURE VESSEL CODE," ASME, 1992

Welding of Low-Alloy Steels

F.J. Winsor, Welding Consultant

Heat-Treatable Low-Alloy (HTLA) Steels

The term heat-treatable low-alloy steels applies to medium-carbon HSLA Q&T steels, which typically contain up to 5% of total alloy content and from 0.25 to 0.50% C. The carbon content of these steels is significantly higher than the 0.10 to 0.25% C range that is found in quenched and tempered structural low-alloy steels or low-alloy steels used for pressure vessels and piping.

Medium-carbon low-alloy steels are strengthened by quenching to produce a martensitic structure, followed by tempering to obtain the desired levels of hardness, strength, and ductility. The combination of high carbon content, which promotes high maximum hardness levels, and high hardenability, which is produced by the carbon and alloy content, results in a greater susceptibility to weld cracking.

The high hardness of many of these steels in their final hardened and tempered condition often precludes their being welded in that condition, because of the strong tendency for cold cracking (that is, cracking that develops after solidification is complete). Except for weld repairs, these steels are normally welded in the annealed or over-tempered condition, and then the complete weldment is heat treated to obtain the desired level of hardness or strength. The compositions of several typical HTLA steels are given in Table 12.

TABLE 12 COMPOSITION OF SELECTED HEAT-TREATABLE LOW-ALLOY (HTLA) STEELS

AISI-SAE OR OTHER DESIGNATION	COMPOSITION, WT%						
	C	**Mn**	**Si**	**Cr**	**Ni**	**Mo**	V
1330	0.28-0.33	1.60-1.90	0.15-0.30	. . .	. . .	. . .	. . .
1340	0.38-0.43	1.60-1.90	0.15-0.30	. . .	. . .	. . .	. . .
4023	0.20-0.25	0.70-0.90	0.15-0.30	. . .	. . .	0.20-0.30	. . .
4028	0.25-0.30	0.70-0.90	0.15-0.30	. . .	. . .	0.20-0.30	. . .
4047	0.45-0.50	0.70-0.90	0.15-0.30	. . .	. . .	0.20-0.30	. . .
4118	0.18-0.23	0.70-0.90	0.15-0.30	0.40-0.60	. . .	0.08-0.15	. . .
4130	0.28-0.33	0.40-0.60	0.15-0.30	0.80-1.10	. . .	0.15-0.25	. . .
4140	0.38-0.43	0.75-1.00	0.15-0.30	0.80-1.10	. . .	0.15-0.25	. . .
4150	0.48-0.53	0.75-1.00	0.15-0.30	0.80-1.10	. . .	0.15-0.25	. . .
4320	0.17-0.22	0.45-0.65	0.15-0.30	0.40-0.60	1.65-2.00	0.20-0.30	. . .
4340	0.38-0.43	0.60-0.80	0.15-0.30	0.70-0.90	1.65-2.00	0.20-0.30	. . .
4620	0.17-0.22	0.45-0.65	0.15-0.30	. . .	1.65-2.00	0.20-0.30	. . .
5120	0.17-0.22	0.70-0.90	0.15-0.30	0.70-0.90	. . .	. . .	. . .
5145	0.40-0.48	0.70-0.90	0.15-0.30	0.70-0.90	. . .	. . .	. . .
6150	0.48-0.53	0.70-0.90	0.20-0.35	0.80-1.10	. . .	. . .	**0.15-0.25**
8620	0.18-0.33	0.70-0.90	0.15-0.30	0.40-0.60	0.40-0.70	0.15-0.25	. . .
8630	0.28-0.33	0.70-0.90	0.15-0.30	0.40-0.60	0.40-0.70	0.15-0.25	. . .
8640	**0.38-0.43**	**0.70-0.90**	**0.15-0.30**	**0.40-0.60**	**0.40-0.70**	**0.15-0.25**	**. . .**

Welding Processes and Practices

Any of the common arc welding processes, such as SMAW, SAW, GMAW, GTAW, and FCAW, can be used to join the HTLA steels. It is important to use low-hydrogen welding consumables and processes in order to minimize cold cracking.

Because of its versatility, simple equipment requirements, and availability of electrode choices, the SMAW process is very often used in the fabrication or repair welding of HTLA steels. GTAW is capable of producing welds with the lowest hydrogen content, and this process is often preferred for critical applications. Both GTAW and GMAW processes provide good control of weld metal composition and cleanliness. For some applications, automatic and semiautomatic GTAW, GMAW, FCAW, and SAW processes offer the advantages of greater productivity and higher weld quality, compared with manual welding.

To avoid porosity and to minimize cracking problems, all materials involved in the welding operation must be clean and free of moisture, oil, grease, dirt, paint, or other contaminants. This includes the base metal, weld filler metal, submerged arc fluxes, gloves, fixtures, and handling devices.

The successful welding of HTLA steels requires not only the selection of the proper welding process and filler metal, but the consideration of several other factors:

- PREHEAT AND INTERPASS TEMPERATURE REQUIREMENTS
- ACHIEVEMENT OF A LOW HYDROGEN CONTENT IN THE WELD METAL, ESPECIALLY WHEN THE PREHEAT TEMPERATURE IS LESS THAN 205 °C (400 °F)
- NEED TO MAINTAIN PREHEAT AFTER WELDING IS COMPLETED UNTIL THE COMMENCEMENT OF POSTWELD HEAT TREATMENT

Filler Metals

Shielded Metal Arc Welding. Low-hydrogen electrodes are nearly always recommended for the welding of HTLA steels to minimize weld metal and HAZ cracking. Electrodes that deposit weld metal that has a response to heat treatment matching that of the base metal are available from some electrode manufacturers for use in welding a few HTLA steels that are quenched and tempered after welding. The chemical composition of the weld metal deposited by these electrodes sometimes, but not always, matches that of the base metal for which they are intended. The main objective is to achieve a weld deposit that responds to a quench and temper heat treatment in a manner similar to that of the base metal. Sometimes this objective can be achieved with weld metal that has a lower carbon and higher alloy content than the base metal. This is advantageous in reducing weld-cracking tendencies. Medium-carbon HTLA steel welding electrodes are not classified in ANSI/AWS specification A5.5 or other standard filler metal specifications. Rather, they are proprietary types designed for welding steels such as American Iron and Steel Institute-Society of Automotive Engineers (AISI-SAE) types 4130, 4140, and 8630. Electrode manufacturers should be consulted for information on the availability of HTLA steel electrodes and applications.

Some applications of HTLA steels do not require that the hardness and strength of the weld match those of the base metal. The location of weld joints and the orientation of applied stresses can be such that weld properties need not match base metal properties after a quench and temper heat treatment. In some repair situations, it may be possible to use the steel in the annealed or over-tempered condition. In this case, after welding the material in that condition, the weldment is placed in service after a stress-relieving or tempering treatment. For such applications, a variety of low-hydrogen carbon steel, low-alloy steel, stainless steel, and nickel alloy covered electrodes may be useful.

Some types of electrodes that deposit non-heat-treatable weld metal can be used to a limited extent for the welding of HTLA steels in the hardened condition, as in repair welding, where it may not be feasible to anneal or over-temper the steel prior to welding nor to stress relieve the weld upon completion. Weld metal from austenitic stainless steel electrodes, such as ANSI/AWS A5.4 E309 (25Cr-12Ni), E310 (25Cr-20Ni), and E312 (29Cr-9Ni), and from nickel alloy electrodes, such as ANSI/AWS A5.11 ENiCrFe-2 and ENiCrFe-3, can tolerate a rather large amount of dilution from the medium-carbon low-alloy steel base metal without cracking. Additionally, these weld deposits have lower strength and greater ductility than the quenched and tempered base metal. Consequently, high shrinkage stresses during welding can result in plastic deformation of the weld deposit, rather than in cracking of the base metal.

Table 13 shows classifications of carbon steel and low-alloy steel covered electrodes commonly used for the welding of HTLA steels in applications that do not require a properties match between the weld metal and the base metal in the quenched and tempered condition. In some instances, the weld metal compositions include one or more of the principal alloying elements in the associated base metal. Also, a weld metal that has a higher strength in the as-welded condition is associated with steels that have higher alloy and carbon contents, which are often heat treated to higher hardness and strength levels. However, it should be noted that all of the weld deposits listed have a low carbon content. Therefore, even though the alloy content of the weld deposit may match that of the base metal, it is unlikely that the carbon content of the weld, even with considerable dilution from the base metal, will be high enough to result in weld metal hardness and strength values that match those of the base metal after a post-weld quench and temper heat treatment. It may be necessary to use one of the higher-strength weld metals in order to satisfy a higher design strength requirement for the weld in the as-welded or stress-relieved condition. If this is not necessary, then the use of a lower-strength weld metal is recommended to decrease the susceptibility to cracking.

TABLE 13 COVERED ELECTRODES COMMONLY USED FOR SHIELDED METAL ARC WELDING OF HTLA STEELS

ELECTRODE CLASSIFICATION	TYPE OF STEEL(A)
E7018	**1330**
E7018-A1	**4023; 4028; 4118; 4320; 8620**
E8016-C1	**4620**
E8016-B2	**5120**
E9016-B3	**5145**
E10016-D2	**1340; 4047; 4130**
E11018-M	**8630**
E12018-M	**4140; 4150; 4340; 4640; 8640**

(A) WELD METAL PROPERTIES DO NOT MATCH BASE METAL PROPERTIES IN THE QUENCHED AND TEMPERED CONDITION.

Gas-Tungsten Arc Welding and Gas-Metal Arc Welding. There are no standard filler metal specifications that classify medium-carbon, heat-treatable bare electrodes and rods for gas-shielded arc welding processes. Bare filler metal of a composition similar to the base metal can be used for the GTAW and GMAW processes. It is desirable to have the lowest possible phosphorus and sulfur contents, and, in some instances, it may be permissible to use weld metal that has a lower carbon content than the base metal, with the objectives of reducing weld metal hot cracking tendencies and improving weld metal ductility. Many of the standard AISI-SAE steel compositions are available in wire form that is suitable for gas-shielded arc welding. The results of transverse tensile tests of GMAW and GTAW joints in some representative low-alloy steels after quenching and tempering are shown in Table 14.

TABLE 14 TYPICAL TRANSVERSE TENSILE PROPERTIES OF ARC WELDED JOINTS IN SELECTED LOW-ALLOY STEELS AFTER QUENCHING AND TEMPERING

WELDING PROCESS AND STEEL DESIGNATION	**THICKNESS**		**FILLER METAL**	**TEMPERING TEMPERATURE**		**WELDED JOINT**					APPROXIMATE BASE METAL TENSILE STRENGTH	
						TENSILE STRENGTH		**YIELD STRENGTH**		**ELONGATION IN 2 IN., %**		
	MM	**IN.**		**°C**	**°F**	**MPA**	**KSI**	**MPA**	**KSI**		**MPA**	KSI
GAS-METAL ARC WELDING (GMAW)												
4130	6.35	0.25	(A)	510	950	1170	170	1145	166	7	1170	**170**
4140	12.7	0.50	4140	480	900	1305	189	1225	178	8(E)	1310	**190**
4340	25.4	1.0	4340	510	950	1307	189.5	1251	181.5	11	1310	**190**
			(B)	510	950	1320	191.5	1224	177.5	8	1310	**190**
GAS-TUNGSTEN ARC WELDING (GTAW)												
4335V	6.35	0.25	4340	205	400	1760	255	1530	222	9	1785	**259**
D6	2.36	0.093	(C)	315	600	1860	270	1635	237	6(E)	1895	**275**
			D6	315	600	1780	258	1505	218	6	1825	**265**
D6	**12.7**	**0.50**	**(D)**	**540**	**1000**	**1545**	**224**	**1425**	**207**	**7**	**1585**	**230**

Source: Ref 2

(A) COMPOSITION: 0.18 C-1.50 MN-0.44 SI-1.2 NI-0.34 MO-0.65 CR.

(B) COMPOSITION: 0.25 C-1.17 MN-0.65 SI-1.8 NI-0.80 MO-1.17 CR-0.21 V.

(C) COMPOSITION: 0.25 C-0.28 MN-0.03 SI-1.29 MO-0.98 CR-0.56 V.

(D) COMPOSITION: 0.25 C-0.55 MN-0.65 SI-0.50 SI-1.25 CR-0.30 V.

(E) ELONGATION IN 25 MM (1 IN.) GAGE LENGTH.

The surface condition of the filler metal can have a critical effect on arc welding operations. Welding filler metal producers or specialty wire suppliers who are familiar with welding requirements usually are the best sources of bare welding filler metals.

In cases where service requirements do not dictate the need for weld deposits that are heat treatable by quenching and tempering after welding, the carbon steel filler metals classified in ANSI/AWS specification A5.18 or the low-alloy steel filler metals classified in ANSI/AWS A5.28 can be used. Also, nonheat-treatable austenitic stainless steel or nickel alloy filler metals may be acceptable choices for some applications.

Flux-Cored Arc Welding. There are no standard specifications for FCAW electrodes that provide weld deposits that respond similarly to heat treatment as do medium-carbon low-alloy steel-base metals. However, FCAW electrodes of special composition can often be obtained more readily than solid filler metals of similar composition. For example, to achieve weld strength requirements equivalent to a quenched and tempered AISI 4130 steel (0.30C-0.95Cr-0.20Mo) weldment, a FCAW electrode was specially formulated to produce a weld deposit with a composition of 0.15% C, 1.50 to 1.70% Mn, 0.50 to 0.60% Si, 0.60 to 0.80% Cr, and 0.60 to 0.90% Mo. Producers of flux-cored electrodes should be consulted regarding the availability of special electrodes.

Where service requirements do not dictate the need for weld deposits that are heat treatable by quenching and tempering after welding, ANSI/AWS A5.20 carbon steel flux-cored electrodes or ANSI/AWS A5.29 low-alloy steel flux-cored electrodes can be used. A T5 slag system is preferred to T1, because weld deposits often have a higher toughness and lower hardness. Carbon steel deposits can fail, because of low strength, if the fillet size is small or high restraint exists. When selecting FCAW electrodes, it may be helpful to refer to commonly used SMAW electrodes (shown in Table 13). Also, either the non-heat-treatable austenitic stainless steel flux-cored electrodes listed in ANSI/AWS specification A5.22 or proprietary nickel alloy flux-cored electrodes may be suitable choices for some applications.

Submerged Arc Welding. Some low-alloy steel weldments that require a quench and temper heat treatment after welding can be welded by the SAW process. Submerged arc weld metal usually has lower toughness than welds made by GTAW, particularly after heat treatment to a high strength level. Neutral or basic fluxes generally produce welds with better notch toughness. Manufacturers of electrodes and fluxes, as well as steel producers, should be consulted to determine the availability of consumables, recommended welding parameters and heat treatment, and information on mechanical properties for specific applications of SAW. If such information is not available, then welding procedure development tests will be needed.

Preheating

In low-alloy steels, the carbon and alloy contents determine the transformation characteristics (microstructure) of the steel upon cooling from a high temperature. The higher the carbon content, the higher the maximum hardness of as-quenched martensite. If the weld metal and HAZs are allowed to cool rapidly to room temperature during welding of low-alloy steels that have high carbon and high alloy content, then hard, crack-sensitive martensite is likely to form. The risk of cracking is greatly increased by the presence of hydrogen. By slowing the cooling rate, preheating may lower the risk of cracking through the reduction of stresses imposed on the weld metal and HAZs during the transformation to martensite or to martensite and bainite. In some steels, preheating may promote the formation of softer bainite instead of hard martensite. Preheating also promotes the evolution of dissolved hydrogen. Steels with an identical alloy content, but different carbon contents, usually require an increase in preheat temperature with increasing carbon content.

The minimum preheat and interpass temperature required to prevent cracking during the welding of HTLA steels depends on the:

- CARBON AND ALLOY CONTENT OF THE BASE METAL AND WELD METAL
- CONDITION OF HEAT TREATMENT PRIOR TO WELDING
- BASE METAL THICKNESS
- AMOUNT OF RESTRAINT
- AMOUNT OF AVAILABLE HYDROGEN

Figure 1, an isothermal transformation diagram for AISI 4340 steel, shows a M_s temperature (the temperature at which martensite starts to form from austenite upon cooling) of approximately 290 °C (550 °F). At 315 °C (600 °F), the diagram shows about 30 min of elapsed time, and, at 370 °C (700 °F) about 3 h time, required for isothermal transformation of austenite (A) to bainite (ferrite and carbides, F + C). When welding this steel by using a preheat and interpass temperature ranging from 315 to 345 °C (600 to 650 °F), and maintaining this temperature for a minimum of 3 h after completion of welding, followed by slow cooling, a bainitic microstructure can be obtained in the HAZ and in weld metal of matching composition. It is essential to maintain the preheat after welding for a sufficient time to allow for variations that could result in a longer transformation time. Maintaining preheat after welding also promotes the evolution of hydrogen and reduces the risk of hydrogen-induced cracking.

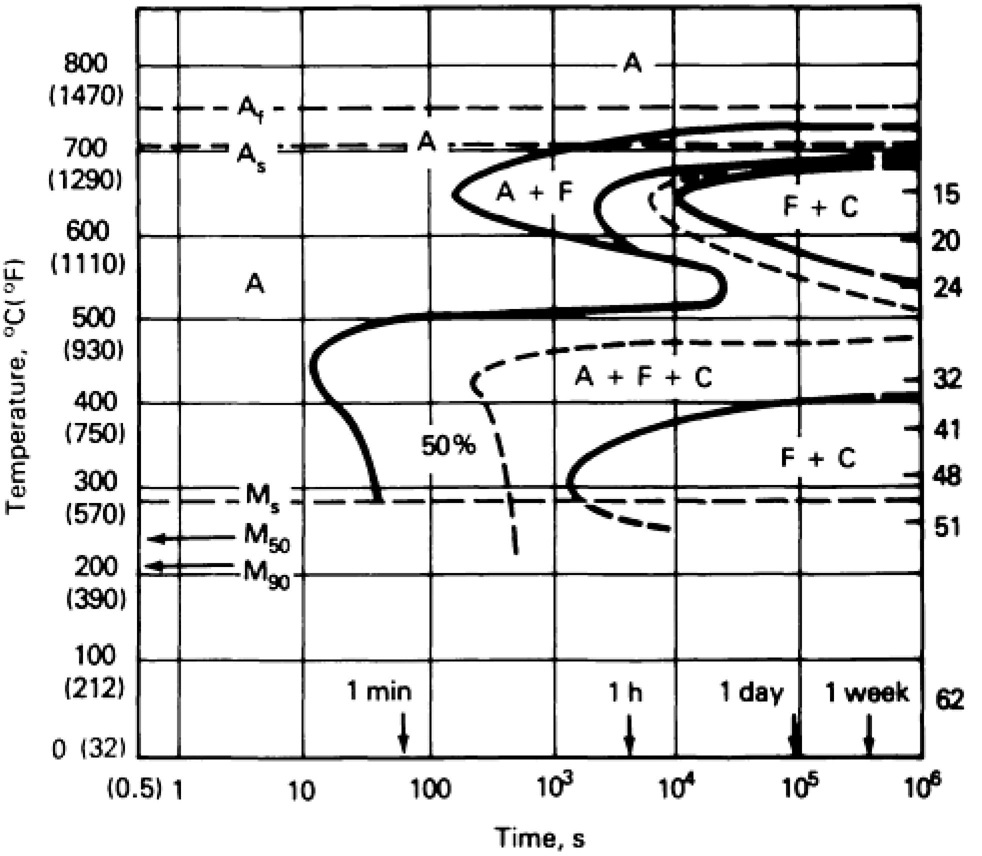

FIG. 1 ISOTHERMAL TRANSFORMATION DIAGRAM FOR AISI 4340 STEEL. SOURCE: REF 5

Unfortunately, preheat and interpass temperatures above about 290 °C (550 °F) promote the formation of a thin oxide layer on the joint surfaces and can also contribute to welder discomfort. These factors can have an adverse effect on weld quality and should be considered when planning the welding operation.

The isothermal transformation diagram in Fig. 2 for AISI 8620 steel, which has a lower carbon and lower total alloy content than AISI 4340, indicates a considerably higher M_s temperature of 405 °C (760 °F). M_s temperatures for HTLA steels with relatively low carbon and alloy contents are usually so high that it is often not feasible to employ a preheat and interpass temperature that is above the M_s temperature for the particular steel. However, with a lower carbon and lower alloy content, these steels have lower hardenability. This means that, with moderate rates of cooling from the welding temperature, there is a possibility for obtaining at least partial transformation to ferrite and carbides, rather than complete transformation to martensite. When martensite does form, its hardness will be lower, because of the lower carbon content. In addition, for steels that have a high martensite formation temperature range, the use of a preheat and interpass temperature below this range can permit partial tempering of the martensite. Because of these factors, lower preheat and interpass temperatures can be used satisfactorily when welding lower carbon and lower alloy content steels.

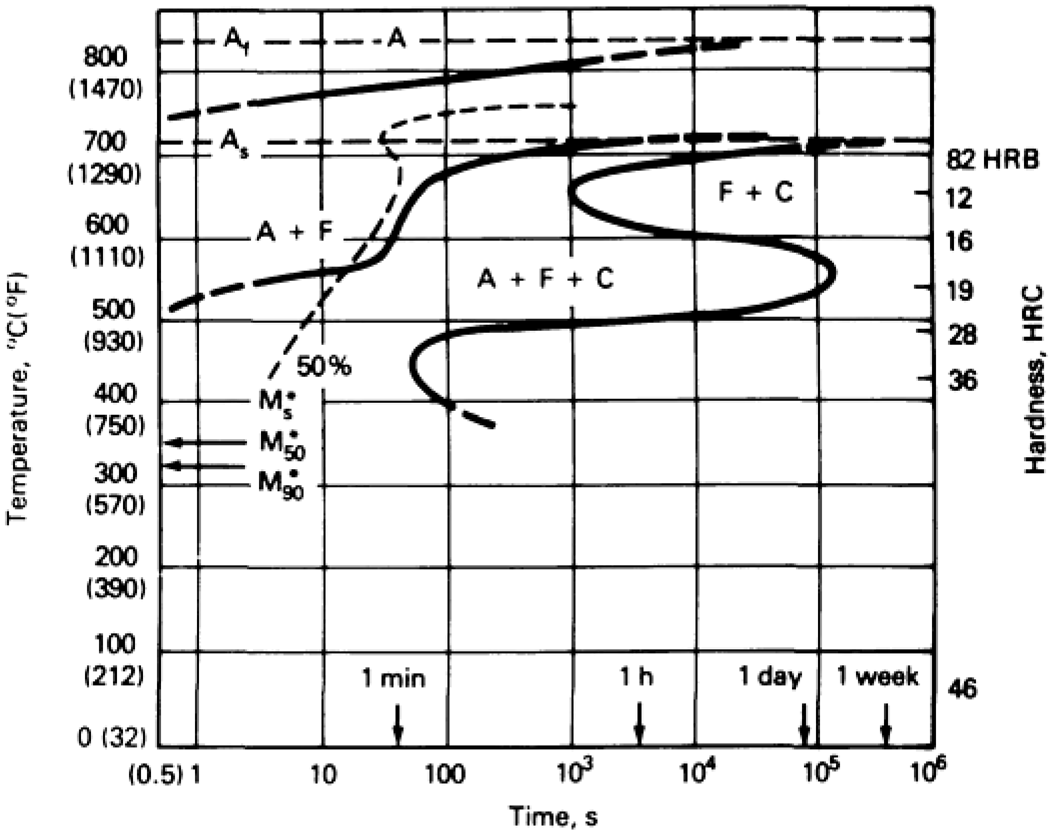

FIG. 2 ISOTHERMAL TRANSFORMATION DIAGRAM FOR AISI 8620 STEEL. * INDICATES ESTIMATED TEMPERATURE. SOURCE: REF 5

For a given initial base metal temperature (preheat) and level of energy input during welding, increasing the section thickness will increase the cooling rate. Therefore, when welding steels of intermediate hardenability, the likelihood of martensite formation also increases with increasing thickness, and it may be necessary to increase the preheat temperature as the section thickness is increased.

Suggested preheat and interpass temperatures for several HTLA steels are given in Table 15. Other than the inherent metallurgical characteristics of the steel, some factors that may permit the use of lower preheat and interpass temperatures than might otherwise be the case are:

- WELDING THE STEEL IN THE FULLY ANNEALED OR OVER-TEMPERED CONDITION
- LOCATING THE WELD IN AREAS OF LEAST SECTION THICKNESS OR WHERE RESTRAINT CONDITIONS MAY BE LEAST SEVERE
- SELECTING PROCESSES AND FILLER METALS THAT OFFER THE LOWEST AMOUNT OF AVAILABLE HYDROGEN
- USING FILLER METALS THAT HAVE THE LOWEST CARBON AND ALLOY CONTENT CONSISTENT WITH OBTAINING THE REQUIRED WELD METAL MECHANICAL PROPERTIES

TABLE 15 RECOMMENDED PREHEAT AND INTERPASS TEMPERATURES FOR SELECTED HTLA STEELS

STEEL DESIGNATION	PREHEAT AND INTERPASS TEMPERATURE RANGE FOR INDICATED SECTION THICKNESS

	AT ≤13 MM (≤ $\frac{1}{2}$ IN.)		**AT 13-25 MM ($\frac{1}{2}$-1 IN.)**		AT 25-50 MM (1-2 IN.)	
	°C	**°F**	**°C**	**°F**	°C	°F
1330	175-230	350-450	205-260	400-500	230-290	**450-550**
1340	205-260	400-500	260-315	500-600	315-370	**600-700**
4023	[GES]40	[GES]100	95-150	200-300	120-170	**250-350**
4028	95-150	200-300	120-175	250-350	205-260	**400-500**
4047	205-260	400-500	230-290	450-550	260-315	**500-600**
4118	95-150	200-300	175-230	350-450	205-260	**400-500**
4130	150-205	300-400	205-260	400-500	230-290	**450-550**
4140	175-230	350-450	230-290	450-550	290-345	**550-650**
4150	205-260	400-500	260-315	500-600	315-370	**600-700**
4320	95-150	200-300	175-230	350-450	205-260	**400-500**
4340	290-345	550-650	315-370	600-700	315-370	**600-700**
4620	[GES]40	[GES]100	95-150	200-300	120-175	**250-350**
4640	175-230	350-450	205-260	400-500	230-290	**450-550**
5120	[GES]40	[GES]100	95-150	200-300	120-175	**250-350**
5145	205-260	400-500	230-290	450-550	260-315	**500-600**
8620	[GES]40	[GES]100	95-150	200-300	120-175	**250-350**
8630	95-150	200-300	120-175	250-350	175-230	**350-450**
8640	**120-175**	**250-350**	**175-230**	**350-450**	**205-260**	**400-500**

Note: Data are for low-hydrogen welding processes and low-hydrogen filler metals.

Heat Input

Low-to-moderate levels of heat input are usually preferred for the welding of HTLA steels. High heat input levels tend to promote wide HAZs with enlarged grain size and reduced notch ductility. High heat input also increases the risk of hot cracking (solidification cracking) in the weld metal and the HAZs.

Postweld Heat Treatment

The type of heat treatment required after welding the HTLA steels (that is, postweld stress relieving or tempering or a quench and temper heat treatment of the completed weldment) depends on the metallurgical characteristics of the steel and the mechanical property requirements. Postweld heat treatment details often depend on the preheat and interpass temperature used in welding.

For the higher carbon and higher alloy content steels, welding with low-to-moderate heat input and using preheat and interpass temperatures that are below the M_s temperature of the particular steel will result in the partial or complete transformation to martensite of some or all of the austenite in the weld metal and HAZs. The amount of martensite that forms depends on the preheat temperature and its relation to the martensite transformation temperature range (essentially between M_s and M_{90}, as shown in Fig. 2). If the martensite transformation is incomplete at the preheat temperature, subsequent cooling from that temperature will result in additional martensite formation. Rapid cooling to room temperature can result in weld or HAZ cracking in higher-carbon HTLA steels. If partial or complete transformation to martensite has occurred, then heating the part to a higher temperature for a sufficient time will temper the martensite.

Steels of lower alloy and carbon content have higher temperature ranges of martensite formation (compare AISI 8620 in Fig. 2 to AISI 4340 in Fig. 1), and they are less susceptible to weld cracking than higher carbon and higher alloy content steels. Preheat and interpass temperatures for steels with lower carbon and lower alloy content usually are within or below the martensite-forming temperature ranges, and weldments are given a tempering heat treatment after welding, unless a quench and temper heat treatment is required.

In order to avoid weld metal or HAZ cracking in the higher carbon and higher alloy content steels, it is necessary, upon completion of welding, to proceed immediately with postweld heat treatment without interrupting the preheating. Even with less crack-sensitive steels, it is preferable to start the postweld heat treatment as soon as possible after welding. If a stress relief or tempering heat treatment is specified immediately after welding (usually 595 to 675 °C, or 1100 to 1250

°F), and if the preheat temperature used in welding is within the martensite-formation temperature range for the steel, then the weldment should first be allowed to cool slowly and uniformly to just below the M_{90} temperature, where transformation to martensite will be essentially complete.

Next, the weldment should be immediately heated to the stress-relieving temperature where the martensite will be softened by tempering. After holding at temperature for a sufficient time (usually ~1 h/in. of joint thickness), the weldment can be cooled to room temperature without danger of cracking. It may not always be practical to start the postweld heat treatment immediately after completion of welding. When cooling to room temperature is necessary after welding, it should be done slowly and uniformly to avoid cracking.

References cited in this section

2. CHAPTERS 1, 3, 4, *WELDING HANDBOOK*, 7TH ED., VOL 4, *METALS AND THEIR WELDABILITY*, AWS, 1982
5. "ISOTHERMAL TRANSFORMATION DIAGRAMS," UNITED STATES STEEL CORP., 1963

Welding of Low-Alloy Steels

F.J. Winsor, Welding Consultant

Ultrahigh-Strength Low-Alloy Steels

The ultrahigh-strength alloy steels used for structural applications have a minimum yield strength of 1240 MPa (180 ksi) and minimum tensile strength of 1380 MPa (200 ksi), combined with good fracture toughness. General types of ultrahigh-strength steels include the nickel-chromium-molybdenum-vanadium steels, nickel-cobalt steels, and the 5% chromium-molybdenum-vanadium steels. The latter are also known as hot-work die steels and are included among the steels discussed in the section "Tool and Die Steels" in this article.

Welded joints in the nickel-cobalt steels generally do not require postweld heat treatment, except for stress relieving in certain applications. Steels in the Ni-Cr-Mo-V series are represented by Aerospace Material Specifications (AMS) 6434, 300M, and D-6a. Some medium-carbon low-alloy steels, such as AISI 4140 and 4340, discussed in the section "Heat-Treatable Low-Alloy (HTLA) Steels," also may be included among the ultrahigh-strength steels. Compositions of several typical ultrahigh-strength steels are given in Table 16.

TABLE 16 TYPICAL COMPOSITION OF SELECTED ULTRAHIGH-STRENGTH LOW-ALLOY STEELS

STEEL DESIGNATION	COMPOSITION, WT%							
	C	**Mn**	**Si**	**Cr**	**Ni**	**Mo**	**V**	**Co**	OTHER
AF1410	0.16	0.05 MAX	0.1 MAX	2 MAX	10	1	0.1	14	**0.15 AL; 0.01 TI (MAX)**
AMS 6434	0.31-0.38	0.60-0.80	0.20-0.35	0.65-0.90	1.65-2.00	0.30-0.40	0.17-0.20	. . .	**. . .**
300M	0.40-0.46	0.65-0.90	1.45-1.80	0.70-0.95	1.65-2.00	0.30-0.45	0.05 MIN	. . .	**. . .**
D-6A	0.42-0.48	0.60-0.90	0.15-0.30	0.90-1.20	0.40-0.70	0.90-1.10	0.05-0.10	. . .	**. . .**
H11 MOD	0.37-0.43	0.20-0.40	0.80-1.00	4.75-5.25	. . .	1.20-1.40	0.40-0.60	. . .	**. . .**
H13	0.32-0.45	0.20-0.50	0.80-1.20	4.75-5.50	. . .	1.10-1.75	0.80-1.20	. . .	**. . .**

HP 9-4-20	0.16-0.23	0.20-0.40	0.20 MAX	0.65-0.85	8.50-9.50	0.90-1.10	0.06-0.12	4.25-4.75	. . .
HP 9-4-30	0.29-0.34	0.10-0.35	0.20 MAX	0.90-1.10	7.0-8.0	0.90-1.10	0.06-0.12	4.25-4.75	. . .
HY-180	**0.09-0.13**	**0.05-0.25**	**0.15 MAX**	**1.8-2.2**	**9.5-10.5**	**0.90-1.10**	**. . .**	**7.5-8.5**	**0.025 AL (MAX); 0.02 TI (MAX)**

Welding Processes and Filler Metals

For many important applications of ultrahigh-strength alloy steels, the mechanical properties of the weld metal and the HAZs must be equal to those of the base metal. The required weld metal strength and toughness can only be achieved with very low levels of impurities such as oxygen, hydrogen, nitrogen, carbon, sulfur, and phosphorus.

The preferred process for achieving maximum strength and toughness in both weld metal and HAZs is gas-tungsten arc welding. Plasma arc welding (PAW) can also be used. Dry and pure argon, helium, or mixtures of both can be used for shielding. Shielding gases should be essentially free from oxygen, nitrogen, and hydrogen.

Premium-quality filler metals are produced from selected raw materials by vacuum induction melting (VIM) and vacuum arc remelting (VAR). Filler metals of matching chemical composition are generally recommended for the welding of nickel-cobalt steels. For the 5Cr-Mo-V steels and the medium-carbon low-alloy steels, the filler metal can either have the same composition as the base metal or, for some applications, lower alloy and lower carbon content filler metal can be used, with some sacrifice in strength.

Welding Procedures and Practices

Joint Design, Workmanship, and Inspection. To utilize the properties of ultrahigh-strength steels and to achieve good serviceability of weldments, it is necessary to provide appropriate joint designs, good workmanship, and adequate inspection. In order to minimize cracking tendencies, welded joints in these steels should not be located in areas of abrupt changes in cross section or high stress concentration. Joints should be located to provide good accessibility for welding and inspection. Groove welds are preferable to fillet welds for many applications, because of the inherent stress concentrations usually associated with fillet welds and because groove welds are more amenable to volumetric methods of examination, such as radiographic or ultrasonic methods. Weld metal should blend smoothly with the adjacent base metal, and excessive weld reinforcement should be avoided.

Nickel-cobalt alloy steels are often welded in the quenched and tempered condition. Preferably, the 5% chromium-molybdenum-vanadium steels and the medium-carbon low-alloy steels are welded in the annealed condition. Successful welding of ultrahigh-strength alloy steels requires careful control of welding conditions. Low nitrogen, oxygen, and hydrogen contents in the weld metal are essential. The filler metal, base metal, weld joint area, and deposited weld metal all must be exceptionally clean. An important monitor of weld metal cleanliness is total oxygen content. The oxygen level must be very low to achieve maximum weld metal properties.

In HY-180 weldments, for example, an oxygen content of less than 50 ppm is necessary. Only GTAW and PAW processes are capable of achieving the required low oxygen levels in the weld metal. The oxygen content of weld metals deposited by GMAW is seldom less than 150 ppm, and it is far higher for slag shielded welding processes.

Welding heat input must be controlled to a low and consistent level to produce grain refinement in previously deposited weld beads and to minimize degradation of the properties of the HAZ. GTAW or PAW with cold wire feed is generally recommended for these steels in preference to GMAW or to GTAW or PAW with hot wire feed. For some applications, high-energy welding processes can be used successfully, if the weld metal oxygen content is kept low and some degradation of mechanical properties can be tolerated.

Preheating

Suggested preheat and interpass temperatures for welding ultrahigh-strength steels are given in Table 17. Preheating generally is not required for welding the nickel-cobalt alloy steels, but it is essential in welding the medium-carbon low-alloy steels and the 5Cr-Mo-V alloy steels in order to prevent cracking in the weld metal and HAZs.

TABLE 17 RECOMMENDED PREHEAT AND INTERPASS TEMPERATURES TO WELD SELECTED ULTRAHIGH-STRENGTH STEELS

STEEL DESIGNATION	PREHEAT AND INTERPASS TEMPERATURE RANGE FOR INDICATED SECTION THICKNESS					
	AT ≤13 MM (≤ $\frac{1}{2}$ IN.)		**AT 13-25 MM ($\frac{1}{2}$-1 IN.)**		AT 25-50 MM (1-2 IN.)	
	°C	**°F**	**°C**	**°F**	**°C**	°F
AMS 6434	175-230	350-450	230-290	450-550	260-315	**500-600**
300M	290-345	550-650	290-345	550-650	315-370	**600-700**
D-6A	230-290	450-550	260-315	500-600	290-345	**550-650**
H11 MOD	315-540	600-1000	315-540	600-1000	315-540	**600-1000**
H13	315-540	600-1000	315-540	600-1000	315-540	**600-1000**
HP 9-4-20	(A)	(A)	(A)	(A)	(A)	**(A)**
HP 9-4-30	(A)	(A)	(A)	(A)	(A)	**(A)**
AF1410	(A)	(A)	(A)	(A)	(A)	**(A)**
HY-180	**(A)**	**(A)**	**(A)**	**(A)**	**(A)**	**(A)**

(A) PREHEATING NOT REQUIRED

Preferred preheat and interpass temperatures for the 5Cr-Mo-V steels (H11 mod and H13) are at least 30 to 55 °C (50 to 100 °F) above the M_s temperature for the steel. The M_s temperature is about 290 °C (550 °F) for H11 mod and about 345 °C (650 °F) for H13 steel. Maintaining the preheat and interpass temperature above the M_s temperature will prevent martensite formation in the weld metal and HAZs during welding. However, high preheat and interpass temperatures can cause welding problems, because of oxidation of the joint faces. The oxide can lead to occasional weld metal porosity or lack of fusion. Although the use of a lower preheat and interpass temperature will overcome this problem, it will allow some of the austenite in the weld zone to transform to martensite. A preheat temperature of about 230 °C (450 °F) can be used without excessive oxidation. For thin sections welded by the GTAW process, a preheat and interpass temperature of 150 °C (300 °F) can be suitable.

If the completed weldment is cooled from the preheat temperature to room temperature, then the remaining austenite will transform to martensite. The volumetric expansion accompanying the martensite transformation will produce high local stresses and can result in cracking of the weld metal or HAZs, especially if the cooling is rapid or nonuniform. In order to avoid this, heat treatment immediately after welding is recommended.

Postweld Heat Treatment

The nickel-cobalt high-strength steels generally do not require postweld heat treatment. Weldments of HP9-4-20 and HP9-4-30 steels can be stress relieved at 540 °C (1000 °F) for 24 h. This postweld heat treatment has no adverse effect on the strength or toughness of either weld metal or base metal. Aging of weldments in HY-180 and AF 1410 steels for a short time at 480 to 510 °C (900 to 950 °F) appears to improve tensile properties and toughness. However, these properties decrease with longer aging times.

The problem of weld metal or HAZ cracking that is due to martensite formation when welding the 5Cr-Mo-V steels and the medium-carbon low-alloy steels can be avoided by postweld heat treatment immediately after welding, rather than allowing the weldment to cool to room temperature. One approach is to raise the temperature after welding to about 55 °C (100 °F) above the M_s temperature and hold it for a sufficient time (indicated by the isothermal transformation diagram) to allow the austenite to transform to a reasonably ductile bainitic structure. The weldment can then be cooled to room temperature without much risk of cracking. A subsequent tempering treatment is required to produce a desirable microstructure in the weld metal and HAZs.

An alternative approach to postweld heat treatment is to allow the weldment to cool slowly and uniformly from the preheat temperature to approximately 95 °C (200 °F), which allows the retained austenite to transform to martensite. The weldment should then be heated to the recommended tempering temperature for the steel to relieve residual welding stresses and to temper the martensite.

Tool and Die Steels

Welding applications for tool and die steels are generally divided into five work operations:

- ASSEMBLY OF COMPONENTS INTO A SINGLE TOOL OR DIE
- FABRICATION OF A COMPOSITE TOOL BY DEPOSITING AN OVERLAY OF TOOL STEEL WELD METAL ON SPECIFIC AREAS OF A CARBON STEEL OR LOW-ALLOY STEEL BASE METAL
- REBUILDING OF WORN SURFACES AND EDGES
- ALTERATION OF A TOOL OR DIE TO MEET A DIFFERENT APPLICATION REQUIREMENT
- REPAIR OF CRACKED, BROKEN, OR OTHERWISE DAMAGED TOOLS, DIES, OR OTHER COMPONENTS

The assembly of components by welding permits the use of less-expensive steels for parts of the assembly that do not require the special properties of tool steel, such as hardness or wear resistance. This practice also permits the use of a tougher steel to provide support for tool steel components or weld overlays. Applications covered by the second and third bullets listed above include most of the welding that is done on tools or dies by hard surfacing, often using special hard-facing alloys, rather than tool steel weld metals. The use of arc welding has significant application for the repair of crocked or broken tools and dies, restoration of worn areas to original dimensions, and alteration of tools and dies to produce modified parts.

Because the manufacturing of complicated tools and dies involves a high material cost, and because premature failures incur the cost of downtime, welding procedures and operations should be carefully planned and controlled to ensure that the required properties are achieved in the completed weldment.

Description of Steels

Tool steels are categorized by AISI and the SAE into seven major groups (listed in Table 18) according to the normal hardening medium or the general applications of the alloys. Tool and die steels are produced in small tonnage quantities by processes that yield clean, homogeneous material with close control of chemical composition.

TABLE 18 COMPOSITION OF TOOL AND DIE STEEL GROUPS

GROUP	**STEEL TYPE**	**LETTER SYMBOL**	PRINCIPAL ALLOYING ELEMENTS
WATER-HARDENING	PLAIN CARBON,	W	**. . .**
	HIGH-CARBON, LOW-ALLOY	W	**. . .**
SHOCK-RESISTING	MEDIUM-CARBON, LOW-ALLOY	S	**CR, MN, MO, SI, W**
COLD-WORK	HIGH-CARBON, LOW-ALLOY	S	**CR, MN, MO, SI, W**
	HIGH-CARBON, MEDIUM-	A	**CR, MN, MO, NI, V, W**

	ALLOY		
	HIGH-CARBON, HIGH-CHROMIUM	D	**CO, CR, MO, V**
HOT-WORK	CHROMIUM	H	**CO, CR, MO, V, W**
	TUNGSTEN	H	**CR, V, W**
	MOLYBDENUM	H	**CR, MO, V, W**
HIGH-SPEED	TUNGSTEN	T	**CO, CR, MO, V, W**
	MOLYBDENUM	M	**CO, CR, MO, V, W**
MOLD STEELS	LOW-TO-MEDIUM CARBON, LOW-ALLOY	P	**CR, MO, NI**
SPECIAL PURPOSE	**MEDIUM-TO-HIGH CARBON, LOW-ALLOY**	**L**	**CR, MO, NI, V**

Welding of Low-Alloy Steels

F.J. Winsor, Welding Consultant

Welding Processes

All of the common arc welding processes can be used to fabricate or repair tools and dies. Process selection primarily depends on the size and complexity of the work, the requirements for preheat and interpass temperature control and postweld heat treatment, and the availability of filler metal.

The SMAW process is most widely used for these reasons: it often affords more flexibility with regard to accessibility to the joint area and position of welding, it requires relatively simple and portable equipment, and it offers a variety of weld metal compositions based on readily available covered electrodes.

The SMAW, GTAW, GMAW, FCAW, and PAW processes are widely used to repair damaged or worn areas or to fabricate composite tools. Both GTAW and PAW are particularly well suited for use on small areas, thin sections, or sharp edges where greater control of the process is advantageous. To deposit large volumes of weld metal, as in building up large die blocks, filling large cavities, or repairing deep cracks, SMAW that utilizes special large electrodes of up to 19 mm ($\frac{3}{4}$ in.) in diameter can be used, as well as SAW, GMAW, FCAW, and hot-wire PAW. Filler metals for semiautomatic and automatic processes can either be solid wires or composite metal-cored or flux-cored tubular products. SAW process modifications can include the use of alloy-enriched fluxes and the direct addition of metal powders into the weld pool.

Welding of Low-Alloy Steels

F.J. Winsor, Welding Consultant

Filler Metals

Factors that must be considered when selecting a filler metal for a specific application include:

- COMPOSITION OF THE BASE METAL
- HEAT-TREATED CONDITION OF THE BASE METAL (ANNEALED OR HARDENED AND TEMPERED)
- SERVICE REQUIREMENTS OF THE WELD DEPOSIT
- POSTWELD HEAT TREATMENT

For many tool steel welds, it is essential that the weld metal and the base metal respond similarly to heat treatment, so that the weld metal hardness will be equal to that of the base metal after a quench and temper heat treatment. Although this often means that the weld metal chemical composition must match that of the base metal, that is not always the case. It is sometimes possible to develop the same final hardness or strength in the weld using a weld deposit that has a lower carbon and higher alloy content than the base metal. Where possible, lowering the carbon content of the weld metal generally minimizes the risk of weld cracking and improves the toughness of the weld deposit.

For some applications, either when fabricating a tool or die from components or when repairing a damaged tool or die, the region of welding may not involve the working surface of the tool or die. In those cases, matching the composition or heat-treatment response of the weld metal to that of the base metal may not be essential. A weld deposit composition that offers greater resistance to weld cracking and improved ductility and toughness often is a better choice.

When selecting filler metals for the welding of tools or dies in the hardened condition, the location of the required weld with respect to the working area of the part should be considered. If the location of the weld deposit is in a functional area of the tool, such as a cutting edge or a wear surface, then it is necessary to select a weld metal that will have the desired hardness or strength in the as-welded condition or, possibly, in the stress-relieved or tempered condition, without a hardening heat treatment. If the weld deposit is located away from the working area of the tool or die, then selection of the filler metal can often be based on achieving maximum crack resistance, ductility, and toughness in the weld deposit.

Filler metals for welding tool steels can be divided into three main categories. The first includes those that produce weld deposits that correspond in hardenability or composition to those of some of the basic tool steel types: water-, oil-, and air-hardening; hot-work; and high-speed. Table 19 shows proprietary compositions of typical tool steel weld metals available from covered electrodes and bare solid and cored filler metals. Such weld deposits will be hard in the as-welded condition. These types of filler metal are not included in standard filler metal specifications.

TABLE 19 COMPOSITION OF SELECTED PROPRIETARY HEAT-TREATABLE WELD DEPOSITS USED TO WELD NUMEROUS TYPES OF TOOL STEEL

AISI STEEL TYPE	**FILLER-METAL TYPE**	NOMINAL WELD DEPOSIT COMPOSITION, WT%$^{(A)}$							
		C	**Mn**	**Si**	**Cr**	**Ni**	**V**	**W**	Mo
W1; W2	WATER-HARDENING	0.95	0.30	0.20	. . .	. . .	0.20	. . .	**. . .**
S1; S5; S7	HOT-WORK	0.33	0.40	1.00	5.00	. . .	. . .	1.25	**1.35**
O1; O6	OIL-HARDENING	0.92	1.28	0.30	0.50	. . .	. . .	0.50	**. . .**
A2; A4; D2	AIR-HARDENING	0.95	0.40	0.30	5.25	. . .	0.25	. . .	**1.10**
H11; H12; H13	HOT-WORK	0.33	0.40	1.00	5.00	. . .	. . .	1.25	**1.35**
M1; M2; M10; T1; T2; T4	**HIGH-SPEED**	**0.90**	**0.30**	**0.35**	**4.00**	**. . .**	**1.00**	**1.50**	**8.00**

(A) COMPOSITIONS OF SIMILAR TYPES OF WELD METAL FROM DIFFERENT PRODUCERS CAN VARY.

The second category of filler metals includes carbon and low-alloy steels that have moderate hardness and toughness properties at room temperature in the as-welded or stress-relieved condition. Although some deposit compositions can respond to a quench and temper heat treatment, most of the filler metals in this category are intended for applications where the properties of the weld deposit are not required to match those of the tool steel in the hardened condition. These filler metals are included in ANSI/AWS specifications for carbon steel and low-alloy steel filler metals.

The third category of filler metals includes austenitic stainless steels and the nickel-chromium-iron, nickel-copper, and copper-nickel alloys. These nonhardenable filler metals are often used for the repair welding of tools in the hardened condition, where high weld metal ductility is advantageous in resisting weld metal and base metal cracking. They can also be useful in providing ductile deposits for building up worn tools. Some types of nonhardenable deposit can be overlaid with a hardfacing deposit to provide a hard, wear-resistant surface. Nonhardenable filler metals are included in ANSI/AWS filler metal specifications for stainless steels, nickel alloys, and copper alloys.

Welding of Low-Alloy Steels

F.J. Winsor, Welding Consultant

Welding Procedures and Practices

It is preferable to conduct the welding of tool and die steels with the part in the annealed condition whenever feasible. Tool steels welded in the hardened and tempered condition are more susceptible to cracking. However, many items are successfully repaired in the hardened and tempered condition. When this can be done safely, it avoids the need for annealing before the repair, hardening and tempering afterwards, and the possible costly correction of dimensional changes caused by such heat treatment. If a tool or die has previously been hardened but not yet tempered, then it cannot be safely welded. Annealing is necessary prior to welding.

Preheating. In order to avoid cracking, tools and dies should never be welded without preheating, regardless of composition or heat-treatment condition (annealed or hardened). The preheat and interpass temperature depends on the composition and thickness of the steel and, sometimes, on the heat-treatment condition. Typical preheat and interpass temperatures for welding different types of tool steel are given in Table 20.

TABLE 20 TYPICAL PREHEAT AND INTERPASS TEMPERATURES TO WELD SELECTED TOOL STEELS

AISI STEEL TYPE	PREHEAT AND INTERPASS TEMPERATURE RANGE			
	ANNEALED BASE METAL		HARDENED BASE METAL	
	°C	**°F**	**°C**	°F
W1; W2	120-230	250-450	120-230	**250-450**
S1; S5; S7	150-260	300-500	150-260	**300-500**
O1; O6	150-205	300-400	150-205	**300-400**
A2; A4	150-260	300-500	150-205	**300-400**
D2	370-480	700-900	370-480	**700-900**
H11; H12; H13	480-650	900-1200	370-540	**700-1000**
M1; M2; M10; T1; T2; T4	**510-595**	**950-1100**	**510-565**	**950-1050**

Source: Ref 2

Preheat and interpass temperatures for welding should not be confused with temperatures used for preheating a tool prior to austenitizing during a hardening treatment. The preheat and interpass temperature for welding a hardened tool steel should not exceed the recommended tempering temperature for the steel. Use of a higher preheat temperature will over-temper and soften the tool. If the tempering temperature is unknown, the preheat and interpass temperature should be at the low end of the range recommended for the particular steel.

To weld annealed tool steels in massive sections, the preheat and interpass temperatures should be near the upper end of the recommended range. For thin sections, the lower portion of the range is satisfactory.

Postweld Heat Treatment. *Heat Treating,* Volume 4 of the *ASM Handbook,* contains information on the heat treatment of tool steel materials. The producers of tool and die steels also can be consulted regarding heat-treatment procedures for their products.

Depending on the type of tool steel and weld metal, the weld metal and HAZ can be very hard and brittle in the as-welded condition and susceptible to cracking if the tools are allowed to cool completely to room temperature after welding and before postweld heat treatment.

With tool steels in the annealed condition prior to welding, it is usually necessary to harden and temper the part after welding in order to provide desired performance characteristics. It is often preferable to heat the part to the hardening temperature as soon as welding is completed, without allowing the part to cool. With some steels, it can be necessary to heat to the hardening temperature in stages in order to avoid cracking. With less-crack-sensitive tool steels, it may not be necessary to proceed with the hardening and tempering heat treatments immediately after completion of welding. In those cases, after welding is completed, the part can be allowed to cool slowly and uniformly to room temperature.

For steels welded in the hardened and tempered condition, reference to isothermal transformation diagrams (Ref 5) can assist in developing postweld heat-treatment practices. For example, with some steels it may be possible to use a preheat and interpass temperature that, if maintained for a reasonable length of time after welding, allows isothermal transformation of austenite to a bainitic microstructure rather than to martensite. Postweld heat treatment involving isothermal transformation of AISI 4340 steel is discussed in the section on welding structural HTLA steels, with reference to Fig. 1, the isothermal transformation diagram for AISI 4340 steel. If the preferred welding preheat and interpass temperature (see Table 20) does not coincide with the desired isothermal transformation temperature, then the temperature of the part should be raised or lowered, as necessary, after welding is completed.

An alternative postweld heat-treatment practice for some tool steels welded in the hardened and tempered condition is to allow the part to cool slowly and uniformly after welding to a temperature from 30 to 55 °C (50 to 100 °F) lower than the M_{90} temperature shown in Fig. 1, permitting transformation to martensite. Then the part should be immediately heated to the recommended tempering temperature.

References cited in this section

2. CHAPTERS 1, 3, 4, *WELDING HANDBOOK*, 7TH ED., VOL 4, *METALS AND THEIR WELDABILITY*, AWS, 1982
5. "ISOTHERMAL TRANSFORMATION DIAGRAMS," UNITED STATES STEEL CORP., 1963

Welding of Low-Alloy Steels

F.J. Winsor, Welding Consultant

Repair Practices for Tools and Dies

Base Metal. The type of tool steel to be repaired and its heat-treatment condition are best determined from the original specifications for the tool or die. If this information is not available, then a chemical analysis must be made to determine the carbon and alloy content of the steel. The heat-treatment condition (either hardened and tempered or annealed) can be determined by a hardness test.

Surface Preparation for Welding. Surfaces of tool and die materials must be clean and dry before welding. Prior to repair welding, all cracks and the tool marks produced in removing damaged areas should be removed. Metal removal can be accomplished by machining, grinding, or thermal cutting, depending on the hardness. Preheating is essential before thermal cutting and gouging, and is often desirable before grinding on hardened tools, as well.

Preparation of Repair Area. For the partial repair of cutting edges or working surfaces, the damaged areas should be machined or ground to a uniform depth that will permit a deposit thickness with the required hardness and wear characteristics. Dilution of the weld metal by fusion with the base metal must be considered, particularly when repairing a composite tool that has a carbon steel or a lower alloy steel base supporting the working surface. A groove depth of about 3.2 mm ($\frac{1}{8}$ in.) is commonly used. When it is necessary to repair the entire cutting or working edge, the edge should be chamfered at 45° for a sufficient depth.

Cracks are generally ground in order to remove the damaged base metal and produce a suitable groove for repair welding. Thermal gouging is also used where extensive excavation is required.

Whether the repair requires a full- or partial-penetration weld depends on many factors, such as the location of the repair area, the stresses that will be imposed on the area in service, the type of weld metal required, and the condition of heat treatment. If the repair requires a hardenable alloy steel weld deposit, or if it is necessary to harden and temper the part after welding, then a partial-penetration weld is generally unsatisfactory, because cracks are likely to propagate from the unwelded root of the repair area.

Welding Processes. The SMAW process is very versatile for the repair welding of small areas. Some types of SMAW electrodes that are available in very large sizes are also used to repair deep cracks or to rebuild large surface areas in massive parts. Some large repairs can be accomplished more economically with semiautomatic or automatic welding processes, such as GMAW, FCAW, SAW, and hot-wire PAW.

Manual GTAW is generally used to repair small areas or edges of parts. Automatic GTAW, using hot or cold wire feed, and hot-wire PAW processes can be used for larger areas.

Repair Practices. If a repair is in a working area of a tool in the annealed condition and is shallow in depth, then a heat-treatable type of deposit that will respond to heat treatment in a manner similar to that of the base metal will be required. If the area has sufficient depth, it can be partially built up with a nonhardenable filler metal and then completed with a tool steel filler metal. The built-up layer should have sufficient strength to support the tool steel deposit under working conditions of the tool. The tool steel deposit should be thick enough to compensate for dilution with the built-up deposit, and it should have adequate wear properties after the quench and temper heat treatment.

A weld deposit used to repair the working surface of a tool in the hardened and tempered condition should have adequate properties after tempering. Alternatively, a hardfacing deposit that does not require heat treatment can be used.

In some cases, the required repairs of either hardened or annealed tools can be in noncritical areas, and a nonhardenable type of filler metal can be used.

When repair welding a cutting edge, the weld metal should be allowed to flow or roll over the edge so that the repaired area can be machined or ground to produce a finished edge without notches or depressions.

Peening is often required to reduce distortion and prevent the cracking of weld deposits. In order to be effective, peening must be done properly; generally, when the deposit is red. Insufficient peening will not relieve stresses, whereas excessive peening can cause cracking.

Welding of Low-Alloy Steels

F.J. Winsor, Welding Consultant

References

1. "ANSI/AWS FILLER METAL SPECIFICATIONS," AWS, 1979-1992
2. CHAPTERS 1, 3, 4, *WELDING HANDBOOK*, 7TH ED., VOL 4, *METALS AND THEIR WELDABILITY*, AWS, 1982
3. "STRUCTURAL WELDING CODE--STEEL," ANSI/AWS D1.1-92, AWS, 1992
4. "ASME BOILER AND PRESSURE VESSEL CODE," ASME, 1992
5. "ISOTHERMAL TRANSFORMATION DIAGRAMS," UNITED STATES STEEL CORP., 1963

Welding of Stainless Steels

D.J. Kotecki, The Lincoln Electric Company

Introduction

STAINLESS STEELS can be defined as alloys that contain at least 10.5% Cr, no more than 1.5% C, and more iron than any other single element. There are five major families of stainless steels, based on microstructure and properties:

- MARTENSITIC STAINLESS STEELS
- FERRITIC STAINLESS STEELS
- AUSTENITIC STAINLESS STEELS
- PRECIPITATION-HARDENING (PH) STAINLESS STEELS
- DUPLEX FERRITIC-AUSTENITIC STAINLESS STEELS

Each family requires different weldability considerations, because of the varied phase transformation behavior upon cooling from solidification to room temperature or below.

Often, the optimal filler metal is not the one that most closely matches the base-metal composition. The most successful procedures for one family are often markedly different from those that are appropriate for another family.

This article addresses consumable selection and procedure development for the welding of stainless steels. Additional information about the selection and weldability of these alloys is available in the Section "Selection of Stainless Steels" in this Volume.

Stainless steel base metals and, therefore, the welding filler metals used with them, are almost invariably chosen on the basis of adequate corrosion resistance for the intended application. This usually means that the welding filler metal must at least match (and sometimes overmatch) the contents of the base metal in terms of specific alloying elements, such as chromium, nickel, and molybdenum.

After considering corrosion resistance, the avoidance of cracking becomes the unifying theme in filler-metal selection and procedure development for the welding of stainless steels. Cracking can occur at temperatures that are just below the bulk solidus temperature of the alloy(s) being welded. This hot cracking, as it is called, can appear as large weld-metal cracks, usually along the weld centerline. However, it can also appear as small, short cracks (microfissures) in the weld metal or in the heat-affected zone (HAZ) at the fusion line and, usually, perpendicular to it. Hot cracking in stainless steel welds is of most concern in austenitic weld metals, although it can occur in all types of stainless steel weldments.

Cracking can also occur at rather low temperatures of, typically, 150 °C (300 °F) or below, because of the interaction of high weld stresses, high-strength metal, and diffusible hydrogen. This cold cracking commonly occurs in martensitic weld metals, as well as HAZs, including those of PH stainless steels. Cold cracking can also occur in ferritic stainless steel weldments that have become embrittled by grain coarsening and/or second-phase particles.

In many instances of cold cracking, where the resulting weld has acceptable properties, the substitution of a mostly austenitic filler metal (with appropriate corrosion resistance) for a martensitic or ferritic filler metal serves as a remedy. When hot cracking occurs in an austenitic weld metal, a common remedy is to use a mostly austenitic filler metal that includes a small amount of ferrite. However, another approach toward avoiding hot cracking may be necessary in situations that require low magnetic permeability, high toughness at cryogenic temperatures, resistance to media that selectively attack ferrite (such as urea), or postweld heat treatments (PWHT) that embrittle ferrite, because their requirements may severely limit the amount of ferrite that is acceptable.

Welding engineers who are responsible for filler-metal selection and procedure development for stainless steel welds will often estimate the microstructure of the weld metal, taking the dilution effects into consideration. This task is best accomplished with a constitution diagram for the stainless steel weld metals. Several diagrams that have been developed over the years are described below.

Welding of Stainless Steels

D.J. Kotecki, The Lincoln Electric Company

Constitution Diagrams

In 1949, Anton Schaeffler (Ref 1) published what has become known as the Schaeffler diagram (Fig. 1). This diagram proposed a relationship among alloy elements that promote the formation of ferrite (chromium-equivalent elements) and elements that promote the formation of austenite and the suppression of ferrite (nickel-equivalent elements). To use this diagram, both the chromium and nickel equivalents are first calculated from the composition of a given weld bead. Next, these equivalents are plotted as coordinates on the Schaeffler diagram. This allows an estimated weld-metal microstructure to be determined from the boundaries given in the diagram. Experience has shown that the Schaeffler diagram is reasonably accurate for conventional "300 series" stainless steel weld deposits from covered electrodes. However, it is of limited use when less-conventional compositions are used and when a high level of nitrogen is present.

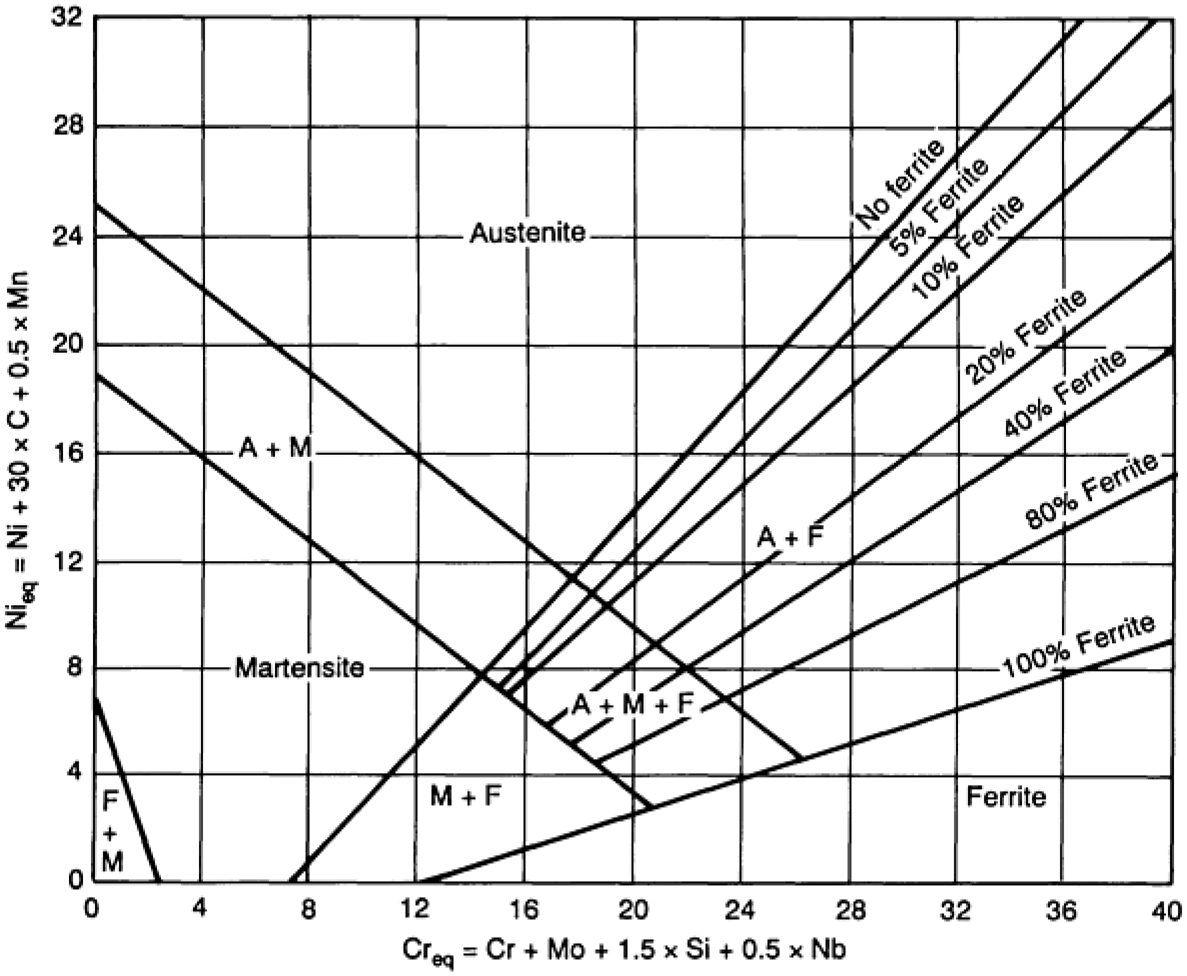

FIG. 1 SCHAEFFLER CONSTITUTION DIAGRAM FOR STAINLESS STEEL WELD METAL

W.T. DeLong recognized the effect of nitrogen in promoting austenite at the expense of ferrite, and he developed a new diagram (Ref 2) that covered a more-restricted composition range and included the effect of nitrogen (Fig. 2). Specifically, a nickel equivalent of 30 × %N was added. To use this diagram, the nickel and chromium equivalents are calculated from the weld-metal analysis. If a nitrogen analysis of the weld metal is not available, then a value of 0.06% should be assumed for the gas-tungsten arc welding (GTAW) process and covered electrodes and a value of 0.08% assumed for gas-metal arc welded metals. If the chemistry is accurate, then the diagram predicts the Welding Research Council (WRC) Ferrite Number within ±3 in approximately 90% of the tests for the alloy 308, 309, 316, and 317 families.

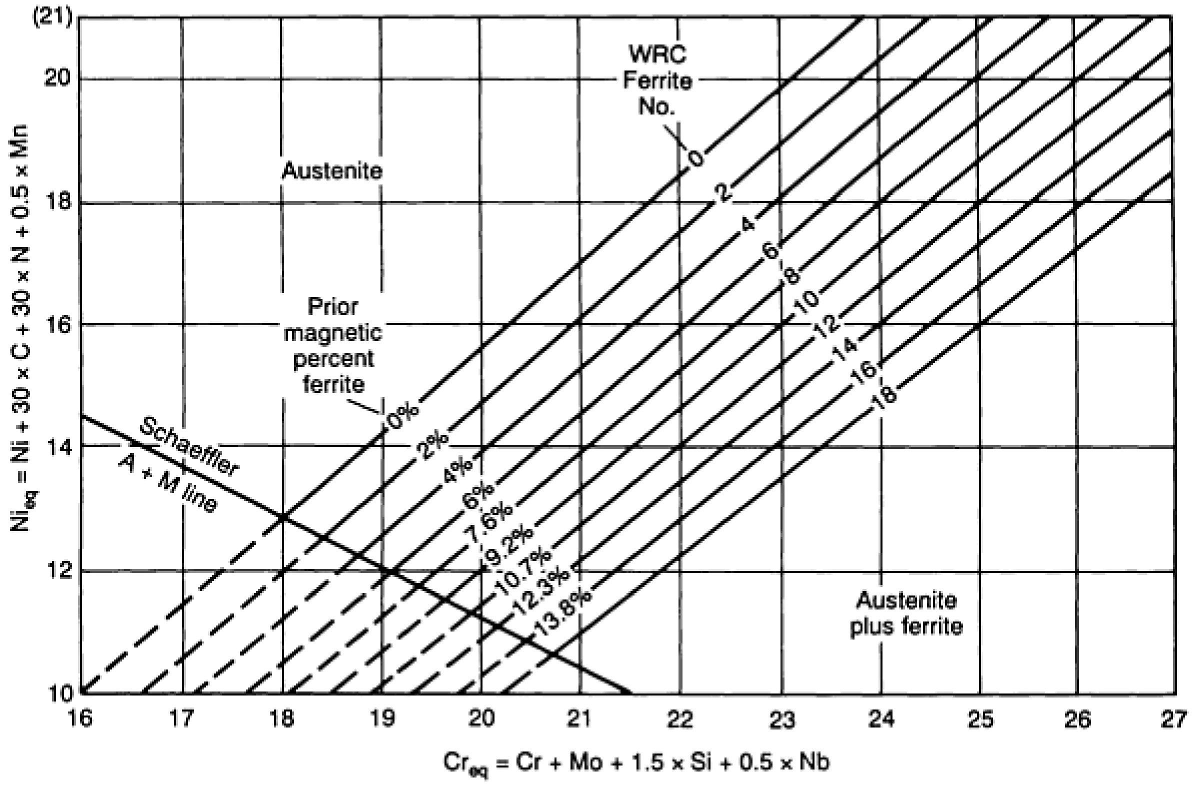

FIG. 2 DELONG CONSTITUTION DIAGRAM FOR STAINLESS STEEL WELD METAL. THE SCHAEFFLER AUSTENITE-MARTENSITE BOUNDARY IS INCLUDED FOR REFERENCE.

In the DeLong diagram, the Ferrite Numbers for alloy 308, 308L, and 347 covered electrodes are similar to those in the Schaeffler diagram, but the alloy 309, 316, and 317 families have Ferrite Numbers that are about two to four times higher. Generally, the DeLong diagram correlates better with GTAW and gas-metal arc welding (GMAW) weld metals than does the Schaeffler diagram, because it allows for nitrogen pickup.

In 1974, at about the time the DeLong diagram was being published, the measurement of ferrite in nominally austenitic stainless steel weld metals was standardized by the ANSI/AWS A4.2 specification in terms of magnetically determined Ferrite Numbers (FN), rather than the metallographically determined "percent ferrite" used by the Schaeffler diagram. The DeLong diagram became part of the "Boiler and Pressure Vessel Code" of the American Society of Mechanical Engineers (ASME).

The DeLong diagram was subsequently discovered to seriously underestimate the ferrite content of weld metals with high manganese contents and overestimate the FN of highly alloyed weld metals, such as alloy 309. Consequently, the Welding Research Council both funded and collected data for the development of a new, more-accurate diagram using computer mapping techniques (Ref 3, 4). A joint effort was conducted by the Colorado School of Mines and the U.S. National Institute of Standards and Technology (NIST). The product of this effort, known as the WRC-1988 diagram (Fig. 3), covers a broader range of compositions than does the DeLong diagram and removes the two errors noted above.

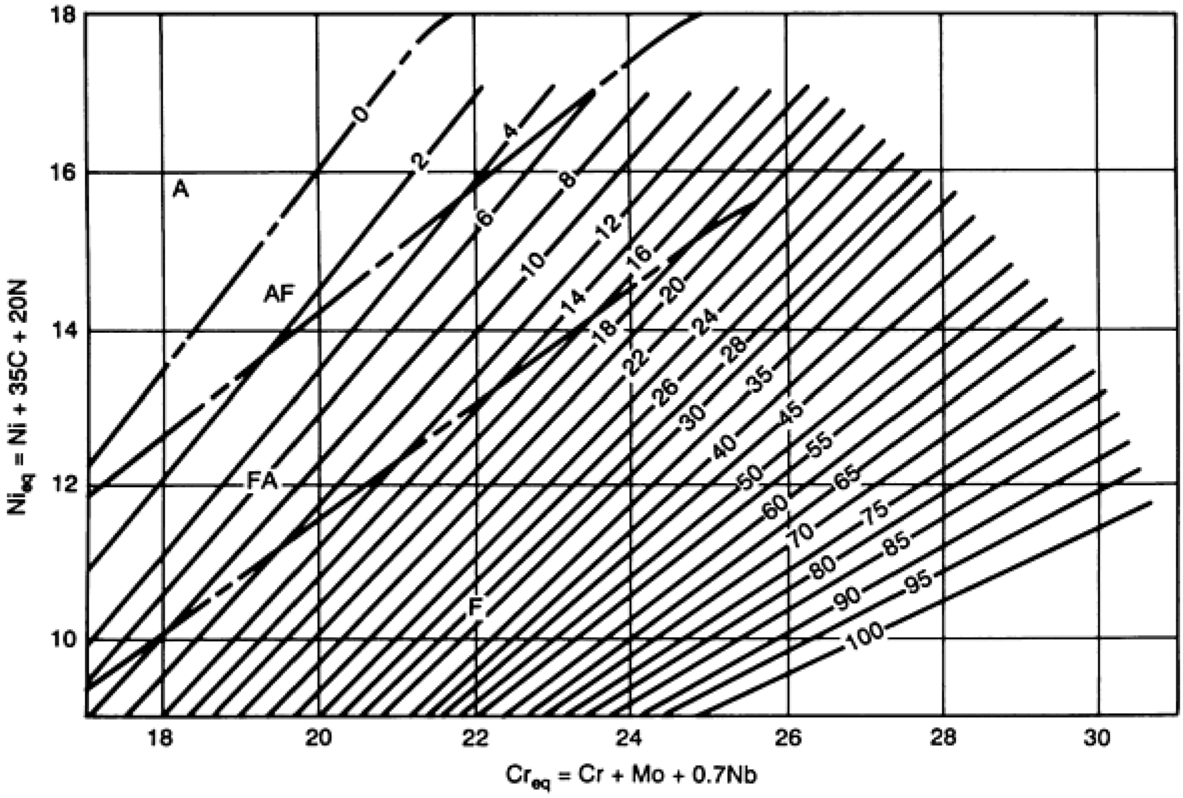

FIG. 3 WRC-1988 DIAGRAM, INCLUDING SOLIDIFICATION-MODE BOUNDARIES

A modification of the WRC-1988 diagram, which allowed the nickel equivalent to include a coefficient for copper, was first proposed by F.B. Lake in 1990 (Ref 5). This modification and an extension of the chromium-equivalent and nickel-equivalent axes were incorporated into the most recent constitution diagram, the WRC-1992 diagram (Ref 6) shown in Fig. 4. Its extended axes permit graphical estimation of the ferrite content of weld metal that comprises very different base metal(s) and filler metal, as could be done less precisely with the Schaeffler diagram.

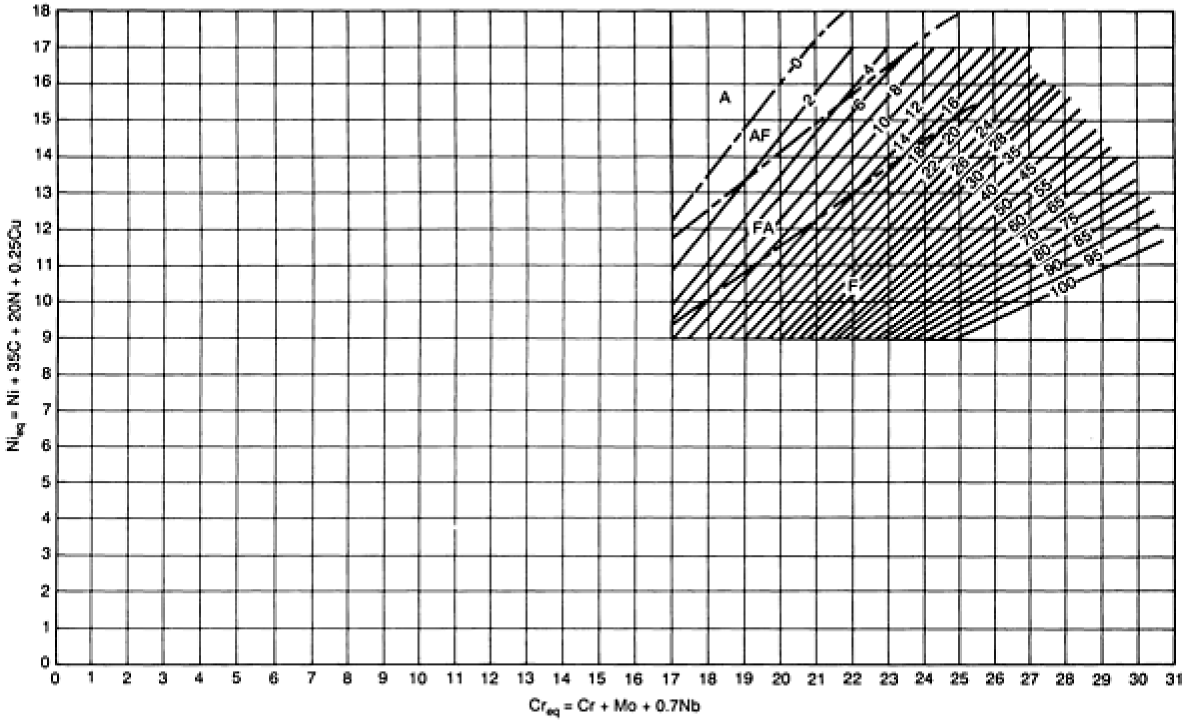

FIG. 4 WRC-1992 DIAGRAM WITH EXPANDED SCALE FOR DILUTION CALCULATIONS. ALTHOUGH EXTENDED AXES ALLOW A WIDE RANGE OF COMPOSITIONS TO BE LOCATED, THE FN PREDICTION IS ONLY ACCURATE FOR WELD COMPOSITIONS THAT FALL WITHIN THE BOUNDS OF THE ISO-FN LINES (0 TO 100 FN). LIMITS OF WRC DIAGRAM WERE DETERMINED BY EXTENT OF DATA BASE, AND EXTENSION OF THE LINES COULD RESULT IN ERRONEOUS PREDICTIONS. SOURCE: REF 6

Although the WRC-1992 diagram is more accurate in predicting ferrite content for many weld metals, the Schaeffler diagram still retains some utility because it can offer reasonably accurate predictions in terms of martensite in lean stainless steel compositions. This is because the WRC-1992 diagram does not include manganese. Manganese has been found (Ref 7) to have no effect on the high-temperature transformation of ferrite to austenite during cooling. Therefore, it does not figure in the nickel equivalent of the WRC-1992 diagram. However, it does have a significant effect on the low-temperature transformation of austenite to martensite during further cooling. It tends to stabilize austenite at low temperatures (Ref 8). Without a manganese effect, it is not possible to put a boundary for the martensite phase in the WRC-1992 diagram.

If the weld deposit is austenitic, then a Ferrite Number around 4 or 5, minimum, typically suffices to prevent hot cracking (Ref 9). In addition, a deposit that is primarily austenitic will not cold crack. These two considerations often influence filler-metal selection and procedure development. In the following sections of this article, the WRC-1992 diagram and, occasionally, the Schaeffler diagram, are used to illustrate the rationale behind many filler-metal choices.

It is appropriate to keep in mind that the ferrite that survives at room temperature after solidification is only a fraction of what existed during solidification. In terms of the utility of constitution diagrams like the WRC-1992 diagram, it is currently thought (Ref 10) that finding ferrite at room temperature is only an indirect indication of the solidification mode. According to this theory, if the weld metal solidifies as ferrite first, with austenite appearing only in the latter stages of solidification, if at all, then the weld metal will be crack resistant. However, if the weld metal solidifies as austenite first, with ferrite only appearing in the latter stages of solidification, if at all, then the weld metal is at risk for cracking.

The dividing line between compositions that solidify as austenite first and compositions that solidify as ferrite first is approximated by the dashed line between the fields labeled "AF" (primary austenite solidification) and "FA" (primary ferrite solidification) in Fig. 3 and 4. This line is not parallel to the isoferrite lines. Rather, it is at a small angle to them, which means that more ferrite is needed at room temperature to represent evidence of primary ferrite solidification in higher-alloyed stainless steel weld metals than in lower-alloyed stainless steel weld metals.

References cited in this section

1. A.L. SCHAEFFLER, CONSTITUTION DIAGRAM FOR STAINLESS STEEL WELD METAL, *MET. PROG.,* VOL 56 (NO. 11), 1949, P 680-680B
2. W.T. DELONG, FERRITE IN AUSTENITIC STAINLESS STEEL WELD METAL, *WELD. J.,* VOL 53 (NO. 7), 1974, P 273S-286S
3. T.A. SIEWERT, C.N. MCCOWAN, AND D.L. OLSON, FERRITE NUMBER PREDICTION TO 100 FN IN STAINLESS STEEL WELD METAL, *WELD J.,* VOL 67 (NO. 12), 1988, P 289S-298S
4. C.N. MCCOWAN, T.A. SIEWERT, AND D.L. OLSON, "STAINLESS STEEL WELD METAL: PREDICTION OF FERRITE CONTENT," WELDING RESEARCH COUNCIL BULLETIN 342, WRC, 1989
5. F.B. LAKE, "EFFECT OF CU ON STAINLESS STEEL WELD METAL FERRITE CONTENT," PAPER PRESENTED AT 1990 AMERICAN WELDING SOCIETY ANNUAL MEETING, AS YET UNPUBLISHED
6. D.J. KOTECKI AND T.A. SIEWERT, WRC-1992 CONSTITUTION DIAGRAM FOR STAINLESS STEEL WELD METALS: A MODIFICATION OF THE WRC-1988 DIAGRAM, *WELD. J.,* VOL 71 (NO. 5), 1992, P 171S-178S
7. E.R. SZUMACHOWSKI AND D.J. KOTECKI, MANGANESE EFFECT ON STAINLESS STEEL WELD METAL FERRITE, *WELD. J.,* VOL 71 (NO. 5), 1984, P 156S-161S
8. J.A. SELF, D.K. MATLOCK, AND D.L. OLSON, AN EVALUATION OF AUSTENITIC FE-MN-NI WELD METAL FOR DISSIMILAR METAL JOINING, *WELD. J.,* VOL 63 (NO. 9), 1984, P 282S-288S
9. C.D. LUNDIN, W.T. DELONG, AND D.F. SPOND, FERRITE-FISSURING RELATIONSHIP IN AUSTENITIC STAINLESS STEEL WELD METALS, *WELD. J.,* VOL 54 (NO. 8), 1975, P 241S-246S
10. J.C. LIPPOLD AND W.F. SAVAGE, SOLIDIFICATION OF AUSTENITIC STAINLESS STEEL WELDMENTS: PART III--THE EFFECT OF SOLIDIFICATION BEHAVIOR ON HOT CRACKING SUSCEPTIBILITY, *WELD. J.,* VOL 61 (NO. 12), 1982, P 388S-396S

Welding of Stainless Steels

D.J. Kotecki, The Lincoln Electric Company

Martensitic Stainless Steels

Basic Metallurgy. At high temperatures, the equilibrium microstructure of martensitic stainless steels is either entirely or almost entirely austenite. At room temperature, the equilibrium microstructure of these steels is a mixture of ferrite and carbides. The reformation of austenite upon heating occurs very rapidly, but the transformation back to ferrite upon cooling is extremely slow. As a result, these steels have a very high tendency to transform to martensite upon cooling from temperatures at which austenite is the stable phase. In fact, it is very difficult to avoid martensite in these steels. They are air hardening, which means that even upon slow cooling in heavy sections, they form martensite.

The binary iron-chromium phase diagram shown in Fig. 5 indicates that above a level of approximately 12% Cr, binary alloys will not form austenite at any temperature, so they cannot be hardened by heating and cooling. However, the addition of carbon to iron-chromium alloys increases the range of chromium contents over which austenite can be formed at elevated temperatures (Fig. 6). It should be noted that nitrogen, nickel, copper, and, possibly, manganese, also expand this range. Even with 17% Cr, a fully austenitic microstructure can be achieved with about 0.4% C at 1250 °C (2280 °F). It can then be fully hardened. When the corrosion-resisting properties of iron-chromium alloys were first discovered early in the 20th century, it was extremely difficult to remove carbon from melts of iron-chromium alloys. Thus, the first stainless steels, which were discovered more or less simultaneously in Germany and England, were martensitic.

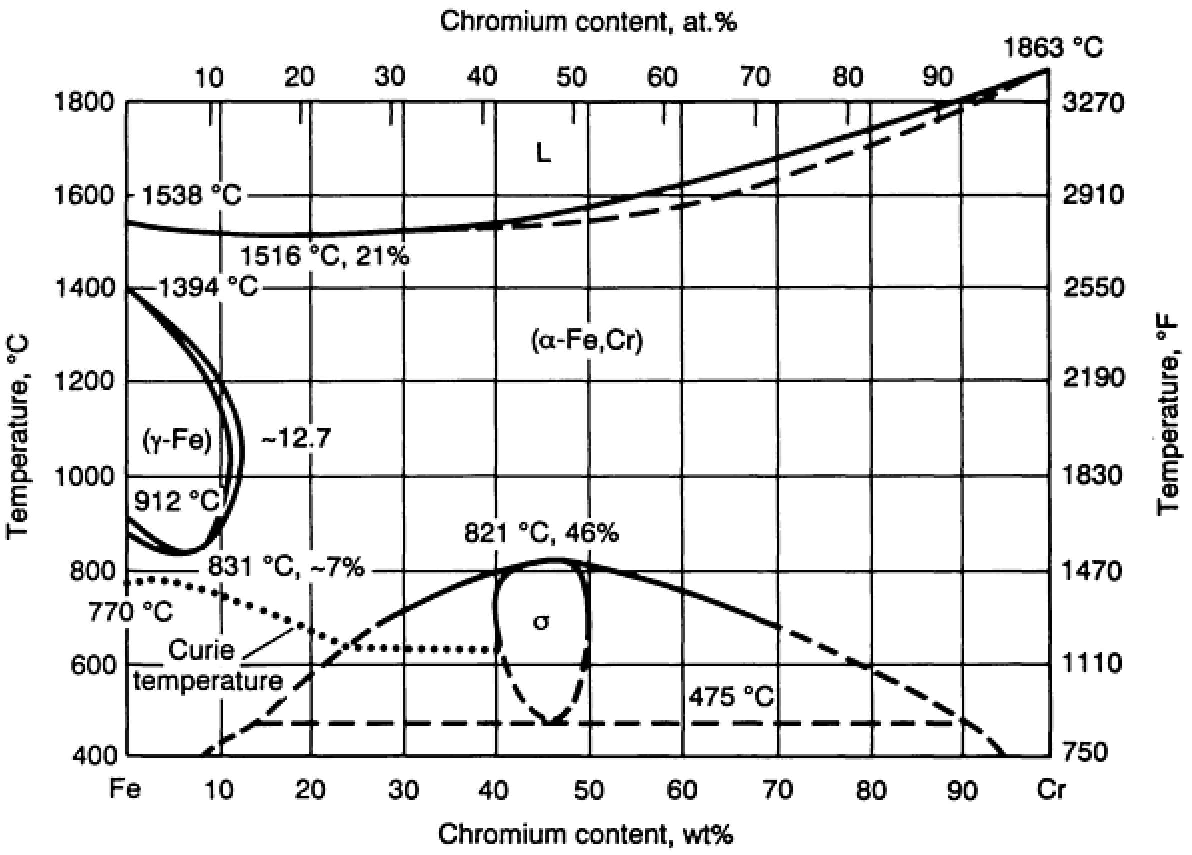

FIG. 5 BINARY IRON-CHROMIUM EQUILIBRIUM PHASE DIAGRAM. SOURCE: REF 11

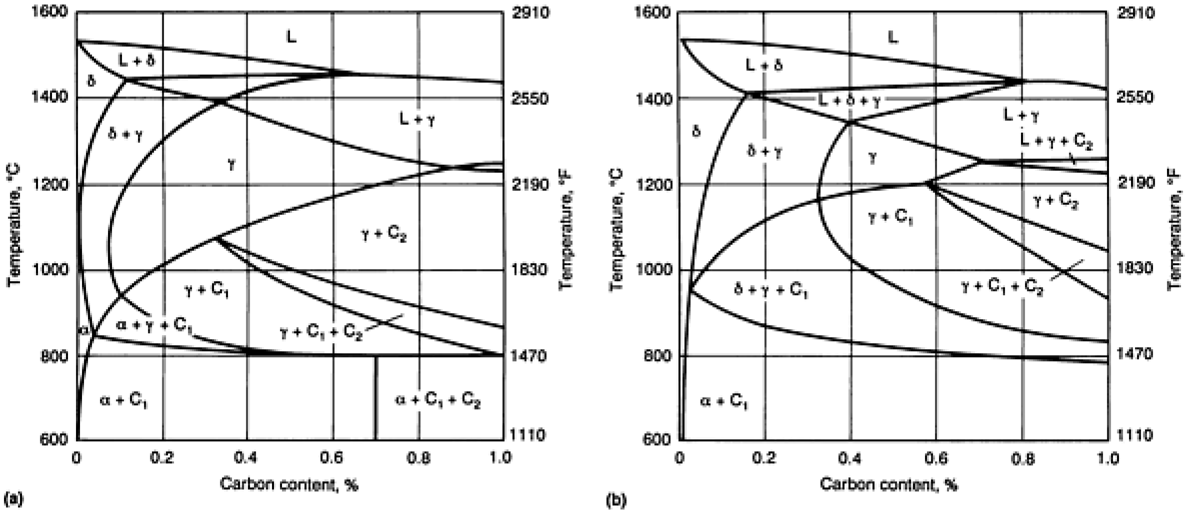

FIG. 6 CONCENTRATION PROFILE IN TERNARY IRON-CHROMIUM-CARBON CONSTITUTION DIAGRAM. (A) 13% CR. (B) 17% CR. SOURCE: REF 12

Carbon content almost entirely determines the hardness of martensite. If virtually 100% martensite can be achieved in a steel, then Fig. 7 shows several relationships that have been found experimentally between carbon and hardness. In particular, it shows that as little as 0.10% C in martensite will result in a hardness value of approximately 35 HRC. The hardness of martensite increases to over 60 HRC at approximately 0.5% C, and it does not change significantly at higher levels of carbon, at least on a macroscale.

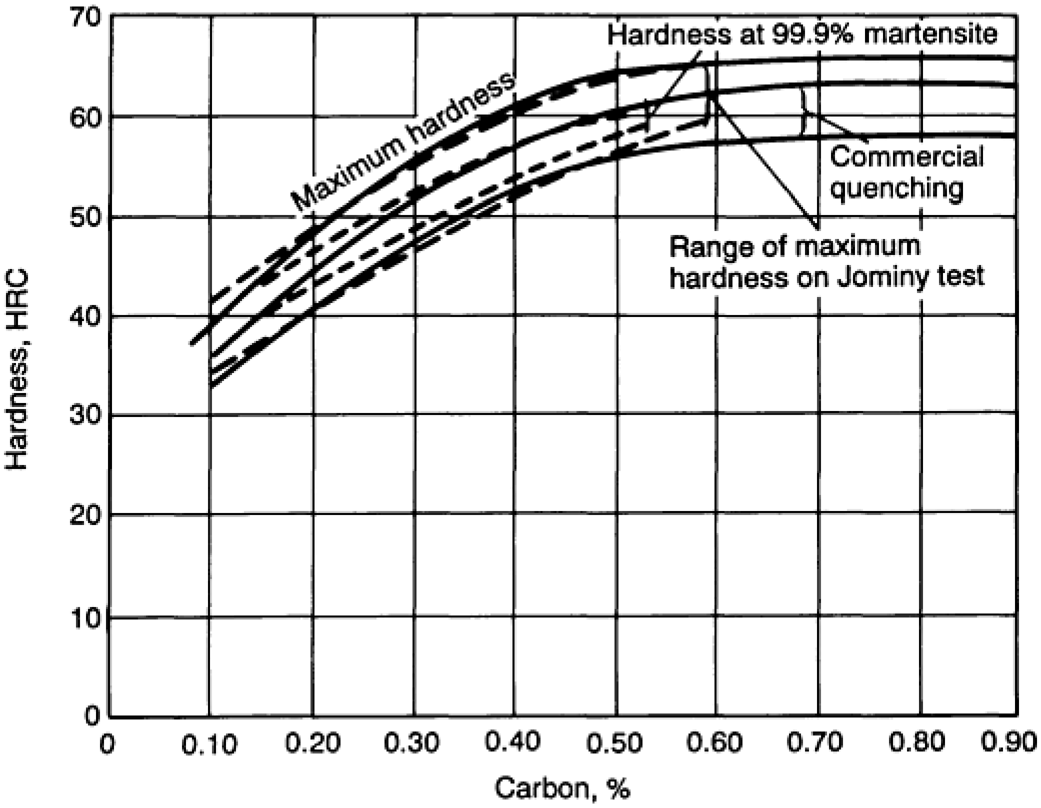

FIG. 7 HARDNESS VALUES OF FULLY HARDENED STEELS WITH A RANGE OF CARBON CONTENTS. SOURCE: REF 13

Figure 7 presents the situation if essentially all martensite is obtained. In carbon and low-alloy steels, rapid cooling is usually necessary to avoid the formation of ferrite. However, as more alloy is added, the avoidance of ferrite becomes easier. At a 12% Cr level, nearly any cooling rate will result in virtually 100% martensite. The prototype of all martensitic stainless steels is type 410 (approximately 12% Cr and 0.10% C). The isothermal transformation diagram for type 410 stainless steel (Fig. 8) indicates that, after cooling from high temperatures and being held at approximately 700 °C (1290 °F), it will take approximately 2 min before any ferrite forms, and several hours will be required before the transformation to ferrite with carbides is complete. Cooling to a holding temperature of 400 °C (750 °F) will require more than 1 week before any ferrite forms.

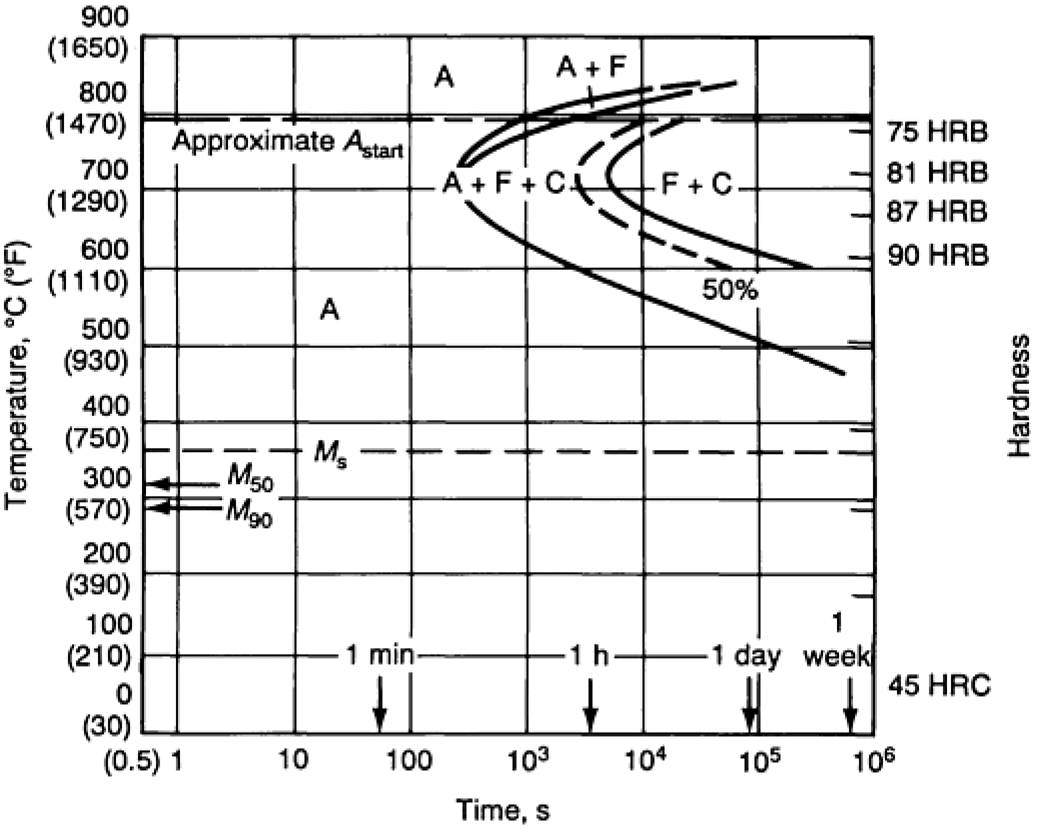

FIG. 8 ISOTHERMAL TRANSFORMATION DIAGRAM FOR TYPE 410 STAINLESS STEEL (12CR-0.1C) AUSTENITIZED AT 980 °C (1800 °F), WITH GRAIN SIZE OF 6 TO 7. A, AUSTENITE; F, FERRITE; C, CARBIDE; M, MARTENSITE; B, BAINITE; P, PEARLITE. SOURCE: REF 14

Figure 8 further shows that martensite begins to form at about 350 °C (660 °F) on cooling type 410 stainless steel and is finished when cooling reaches about 250 °C (480 °F). Alloying at levels that are either above or in addition to 12% Cr serves to further delay ferrite formation and further reduces the martensite-start (M_s) temperature. In the extreme case, some austenite will exist indefinitely, even at room temperature, and refrigeration or complex heat treatments (such as double tempering) may be necessary to induce complete martensite transformation in highly alloyed martensitic stainless steels. Therefore, in any fusion welding process, it must be assumed that regardless of the condition of heat treatment of the base metal, the HAZ and any nearly matching filler metal for martensitic stainless steels in the as-welded condition will be freshly formed martensite. The martensite will be hard and, unless it is very low carbon martensite, it will also be quite brittle and crack-sensitive. Therefore, a standard rule to follow when welding martensitic stainless steels is to weld hot, that is, a high preheat, high interpass temperature, and high heat input are, in most cases, beneficial to accomplishing the weld without incurring cracking.

Martensitic Stainless Steel Base Metals. A number of martensitic stainless steel base metals are identified in Table 1(a). Welding filler metals that have approximately matching compositions, as classified by the American Welding Society (AWS), are also listed in Table 1(b), but only for selected base metals.

TABLE 1(A) COMPOSITIONS OF NOMINALLY MARTENSITIC STAINLESS STEELS

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	COMPOSITION, WT%									
			C	**Mn**	**P (max)**	**S**	**Si (max)**	**Cr**	**Ni**	**Mo**	**N**	OTHER
XM-32	K64152	A 565	0.08-0.15	0.50-0.90	0.025	0.025 MAX	0.35	11.00-12.50	2.00-3.00	1.50-2.00	0.01-0.05	**V, 0.25-0.40**
403	S40300	A 479	0.15 MAX	1.00 MAX	0.040	0.030 MAX	0.50	11.50-13.00	. . .	. . .	. . .	**. . .**
410	S41000	A 240	0.15 MAX	1.00 MAX	0.040	0.030 MAX	1.00	11.50-13.50	0.75 MAX	. . .	. . .	**. . .**
410S	S41008	A 240	0.08 MAX	1.00 MAX	0.040	0.030 MAX	1.00	11.50-13.50	0.60 MAX	. . .	. . .	**. . .**
XM-30	S41040	A 479	0.18 MAX	1.00 MAX	0.040	0.030 MAX	1.00	11.50-13.50	. . .	. . .	. . .	**NB + TA, 0.05-0.30**
. . .	S41041	A 565	0.13-0.18	0.40-0.60	0.030	0.030 MAX	1.50	11.50-13.00	0.50 MAX	0.20 MAX	. . .	**NB + TA, 0.15-0.45; AL, 0.05 MAX**
. . .	S41050	A 240	0.040 MAX	1.00 MAX	0.040	0.030 MAX	1.00	10.50-12.50	0.60-1.10	. . .	0.10 MAX	**. . .**
414	S41400	A 479	0.15 MAX	1.00 MAX	0.040	0.030 MAX	1.00	11.50-13.50	1.25-2.50	. . .	. . .	**. . .**
410NIMO	S41500	A 240	0.05 MAX	0.5-1.0	0.030	0.030 MAX	0.60	11.5-14.0	3.5-5.5	0.5-1.0	. . .	**. . .**
416	S41600	A 473	0.15 MAX	1.25 MAX	0.06	0.15 MIN	1.00	12.00-14.00	. . .	0.60 MAX	. . .	**. . .**
XM-6	S41610	A 582	0.15 MAX	1.50-2.50	0.06	0.15 MIN	1.00	12.00-14.00	. . .	. . .	. . .	**. . .**
416SE	S41623	A 473	0.15 MAX	1.25 MAX	0.06	0.06 MAX	1.00	12.00-14.00	. . .	. . .	. . .	**SE, 0.15 MIN**
615	S41800	A 565	0.15-0.20	0.50 MAX	0.040	0.030 MAX	0.50	12.00-14.00	1.80-2.20	0.50 MAX	. . .	**W, 2.50-3.50**
420	S42000	A 276	0.15 MIN	1.00 MAX	0.040	0.030 MAX	1.00	12.00-14.00	. . .	. . .	. . .	**. . .**
. . .	S42010	A 276	0.5-0.30	1.00 MAX	0.040	0.030 MAX	1.00	13.50-15.00	0.35-0.85	0.40-0.85	. . .	**. . .**

420F	S42020	A 565	0.30-0.40	1.25 MAX	0.06	0.15 MIN	1.00	12.00-14.00	0.50 MAX	. . .	. . .	**CU, 0.60 MAX**
420FSE	S42023	A 565	0.20-0.40	1.25 MAX	0.06	0.06 MAX	1.00	12.00-14.00	0.50 MAX	. . .	. . .	**CU, 0.60 MAX; SE, 0.15 MIN**
616	S42200	A 565	0.20-0.25	0.50-1.00	0.025	0.025 MAX	0.50	11.00-12.50	0.50-1.00	0.90-1.25	. . .	**V, 0.20-0.30; W, 0.90-1.25**
619	S42300	A 565	0.27-0.32	0.95-1.35	0.025	0.025 MAX	0.50	11.00-12.00	0.50 MAX	2.50-3.00	. . .	**V, 0.20-0.30**
431	S43100	A 479	0.20 MAX	1.00 MAX	0.040	0.030 MAX	1.00	15.00-17.00	1.25-2.50	. . .	. . .	**. . .**
440A	S44002	A 276	0.60-0.75	1.00 MAX	0.040	0.030 MAX	1.00	16.00-18.00	. . .	0.75	. . .	**. . .**
440B	S44003	A 276	0.75-0.95	1.00 MAX	0.040	0.030 MAX	1.00	16.00-18.00	. . .	0.75	. . .	**. . .**
440C	S44004	A 276	0.95-1.20	1.00 MAX	0.040	0.030 MAX	1.00	16.00-18.00	. . .	0.75	. . .	**. . .**
440F	S44020	A 582	0.95-1.20	1.25 MAX	0.06	0.15 MIN	1.00	16.00-18.00	0.50 MAX	. . .	. . .	**CU, 0.60 MAX**
440FSE	S44023	A 582	0.95-1.20	1.25 MAX	0.06	0.06 MAX	1.00	16.00-18.00	0.50 MAX	. . .	. . .	**CU, 0.60 MAX; SE, 0.15 MIN**
CA-15	. . .	A 743	0.15 MAX	1.00 MAX	0.04	0.04 MAX	1.50	11.5-14.0	1.00 MAX	0.50 MAX	. . .	**. . .**
CA-15M	. . .	A 743	0.15 MAX	1.00 MAX	0.040	0.040 MAX	0.65	11.5-14.0	1.0 MAX	0.15-1.0	. . .	**. . .**
CA-40	. . .	A 743	0.20-0.40	1.00 MAX	0.04	0.04 MAX	1.50	11.5-14.0	1.0 MAX	0.5 MAX	. . .	**. . .**
CA-40F	. . .	A 743	0.20-0.40	1.00 MAX	0.04	0.20-0.40	1.50	11.5-14.0	1.0 MAX	0.5 MAX	. . .	**. . .**
CA6N	. . .	A 743	0.06 MAX	0.50 MAX	0.02	0.02 MAX	1.00	10.5-12.5	6.0-8.0	. . .	. . .	**. . .**
CA-6NM	. . .	A 743	0.06 MAX	1.00 MAX	0.04	0.03 MAX	1.00	11.5-14.0	3.5-4.5	0.40-1.0	. . .	**. . .**
CA-28MWV	**. . .**	**A 743**	**0.20-0.28**	**0.50-1.00**	**0.030**	**0.030 MAX**	**1.0**	**11.0-12.5**	**0.50-1.00**	**0.90-1.25**	**. . .**	**V, 0.20-0.30; W, 0.90-1.25**

TABLE 1(B) PROPERTIES AND FILLER METALS FOR NOMINALLY MARTENSITIC STAINLESS STEELS

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	**TENSILE STRENGTH**		**YIELD STRENGTH**		**ELONGATION, %**	MATCHING FILLER METALS		
			MPA	**KSI**	**MPA**	**KSI**		**A5.4**	**A5.9(A)**	A5.22
XM-32	K64152	A 565	1000	145	790	115	15	. . .	. . .	**. . .**
403	S40300	A 479	480	70	275	40	20	E410-XX	ER410	**E410T-X**
410	S41000	A 240	450	65	205	30	20	E410-XX	ER410	**E4101-X**
410S	S41008	A 240	415	60	205	30	22	E410-XX	ER410	**E410T-X**
XM-30	S41040	A 479	480	70	275	40	13	. . .	. . .	**. . .**
. . .	S41041	A 565	790	115	520	75	15	. . .	. . .	**. . .**
. . .	S41050	A 240	415	60	205	30	22	. . .	. . .	**. . .**
414	S41400	A 479	790	115	620	90	15	. . .	. . .	**. . .**
410NIMO	S41500	A 240	790	115	620	90	15	E410NIMO-XX	ER410NIMO	**E410NIMOT-X**
416	S41600	A 473	480	70	275	40	20	NCW	NCW	**NCW**
XM-6	S41610	A 582	NS	NS	NS	NS	NS	NCW	NCW	**NCW**
416SE	S41623	A 473	480	70	275	40	20	NCW	NCW	**NCW**
615	S41800	A 565	970	140	760	110	115	. . .	. . .	**. . .**
420	S42000	A 276	NS	NS	NS	NS	NS	. . .	ER420	**. . .**
. . .	S42010	A 276	NS	NS	NS	NS	NS	. . .	. . .	**. . .**
420F	S42020	A 565	NS	NS	NS	NS	NS	NCW	NCW	**NCW**
420FSE	S42023	A 565	NS	NS	NS	NS	NS	NCW	NCW	**NCW**
616	S42200	A 565	970	140	760	110	13	. . .	. . .	**. . .**
619	S42300	A 565	970	140	760	110	8	. . .	. . .	**. . .**
431	S43100	A 479	790	115	620	90	15	. . .	. . .	**. . .**
440A	S44002	A 276	NS	NS	NS	NS	NS	. . .	. . .	**. . .**
440B	S44003	A 276	NS	NS	NS	NS	NS	. . .	. . .	**. . .**
440C	S44004	A 276	NS	NS	NS	NS	NS	. . .	. . .	**. . .**
440F	S44020	A 582	NS	NS	NS	NS	NS	NCW	NCW	**NCW**
440FSE	S44023	A 582	NS	NS	NS	NS	NS	NCW	NCW	**NCW**
CA-15	. . .	A 743	620	90	450	65	18	E410-XX	ER4109	**E410T-X**
CA-15M	. . .	A 743	620	90	450	65	18	. . .	. . .	**. . .**
CA-40	. . .	A 473	690	100	480	70	15	. . .	ER420	**. . .**
CA-40F	. . .	A 743	690	100	480	70	12	NCW	NCW	**NCW**

CA6N	. . .	A 743	970	140	930	135	15	. . .	. . .	**. . .**
CA-6NM	. . .	A 743	760	110	550	80	15	E410NIMO-XX	ER410NIMO	**E410NIMOT-X**
CA-28MWV	**. . .**	**A 743**	**970**	**140**	**760**	**110**	**10**	**. . .**	**. . .**	**. . .**

NS, not specified; NCW, not considered weldable.

(A) METAL-CORED ELECTRODES, INDICATED BY A "C" IN PLACE OF THE "R" IN THE CLASSIFICATION, ARE ALSO INCLUDED.

Of these base metals, the lower carbon content alloys, especially type 410 and CA-6NM (410NiMo), are often used for turbines in hydroelectric and similar types of facilities. They provide resistance to mild corrosion, cavitation, and moderate abrasion, and they possess relatively high strength. CA-6NM is also used in fittings, valves, and pumps for sour (H_2S-containing) oil and gas service. In this application, the alloy usually must be heat treated to a hardness value that does not exceed 22 HRC. This usually requires a double temper treatment consisting of 2 h at 675 °C (1250 °F) (intercritical temperature range), an air cool to room temperature, and then 4 h at 615 °C (1140 °F) (below lower critical temperature).

Type 420 and similar alloys are used in cutlery, valve parts, gears, shafts, and rollers. One extensive application is a weld overlay on rolls for steel mill continuous casters. This type of alloy is quite brittle in the freshly hardened condition and usually must be tempered to obtain useful toughness properties.

The higher carbon content martensitic stainless steels, such as types 440A, 440B, and 440C, can be hardened to values near 60 HRC. Then, they can be tempered at temperatures up to 480 °C (900 °F) to toughen without experiencing much loss of hardness. In this condition, they serve well as cutting edges and as bearings.

Martensitic stainless steels generally develop full hardness upon air cooling from about 1000 °C (1830 °F). They can be softened by tempering at temperatures ranging from approximately 500 to 750 °C (930 to 1380 °F), unless they contain significant amounts of nickel, in which case the maximum temperature for tempering is reduced. After tempering at 650 to 750 °C (1200 to 1380 °F), their hardness values generally drop to about 30 HRC or lower. This can be useful if it is necessary to soften a martensitic stainless steel before welding to allow enough ductility in the bulk of the material to accommodate shrinkage stresses associated with welding. However, such high-temperature tempering produces rather coarse chromium carbides, which damage the corrosion resistance of the metal. Then, it may be necessary after welding to austenitize, air cool to room temperature, and temper at a low temperature (less than 450 °C, or 840 °F) to restore corrosion resistance.

Engineering for Use in the As-Welded Condition. Except for very small weldments or very low carbon base metals, martensitic stainless steels are not usually used in the as-welded condition. This is due to the very brittle weld area that normally results. This area includes the weld metal itself, if it has a matching or near-matching composition to that of the base metal, and the heat-affected zone. However, repair situations can necessitate that the engineer work with these circumstances.

If a weldment of martensitic stainless steel must be used in the as-welded condition, then it is usually best to avoid both autogenous welds (no filler metal) and welds with matching filler metals. Small parts that are welded by laser-beam welding (LBW), electron-beam welding (EBW), or GTAW processes are an exception, because residual stresses can be very low and the welding processes generate almost no diffusible hydrogen.

If at all possible, an austenitic filler metal, such as type 309 or 309L stainless, or a duplex ferritic-austenitic stainless filler metal, such as type 312 stainless, should be chosen, depending on the base metal. The filler-metal choice should provide for a small amount of ferrite in the weld metal, in order to avoid hot cracking. This can be anticipated by using the extended WRC-1992 diagram (Fig. 4). If the weld metal is austenite with a little ferrite, then the weld metal itself will have appreciable ductility, and only the HAZ will be at risk for cold cracking. This possibility can be minimized by using high preheat temperatures (200 °C, or 390 °F, minimum for type 410 base metal and 350 °C, or 660 °F, minimum for types 440A, 440B, and 440C base metals) and slow cooling after welding.

Example 1: Type 410 base metal (0.10C-12.5Cr-0.04N), Cr_{eq} = 12.5 and Ni_{eq} = 4.3; to be welded with E309-16 manual electrode (0.05C-23.5Cr-13.0Ni-0.06N all-weld metal composition), Cr_{eq} = 23.5 and Ni_{eq} = 15.95.

If the type 410 base metal is plotted in Fig. 9 (point A), along with the all-weld metal composition of the E309-16 electrode (point B), then all possible mixtures of these two metals must lie along the line connecting these two points. If the root pass is diluted with 30% base metal (a typical result to expect), then the actual root pass composition will lie along the tie line, 30% of the distance from point B to point A. This composition is shown as point C in Fig. 9. It would correspond to slightly over 8 FN, which is quite safe in terms of the likelihood of hot cracking. Furthermore, because the tie line is nearly parallel to the isoferrite lines in the diagram, major changes in dilution would have practically no effect on weld-metal ferrite content. Additional weld passes on top of the root pass would have compositions that are also along the tie line, but closer to the E309-16 all-weld metal composition. Therefore, E309-16 would be a very good filler metal choice if a nonhardenable weld deposit is acceptable. By similar reasoning, type 309 (or 309L) would be an excellent choice for gas-metal arc weld wire, flux-cored wire, or submerged arc wire.

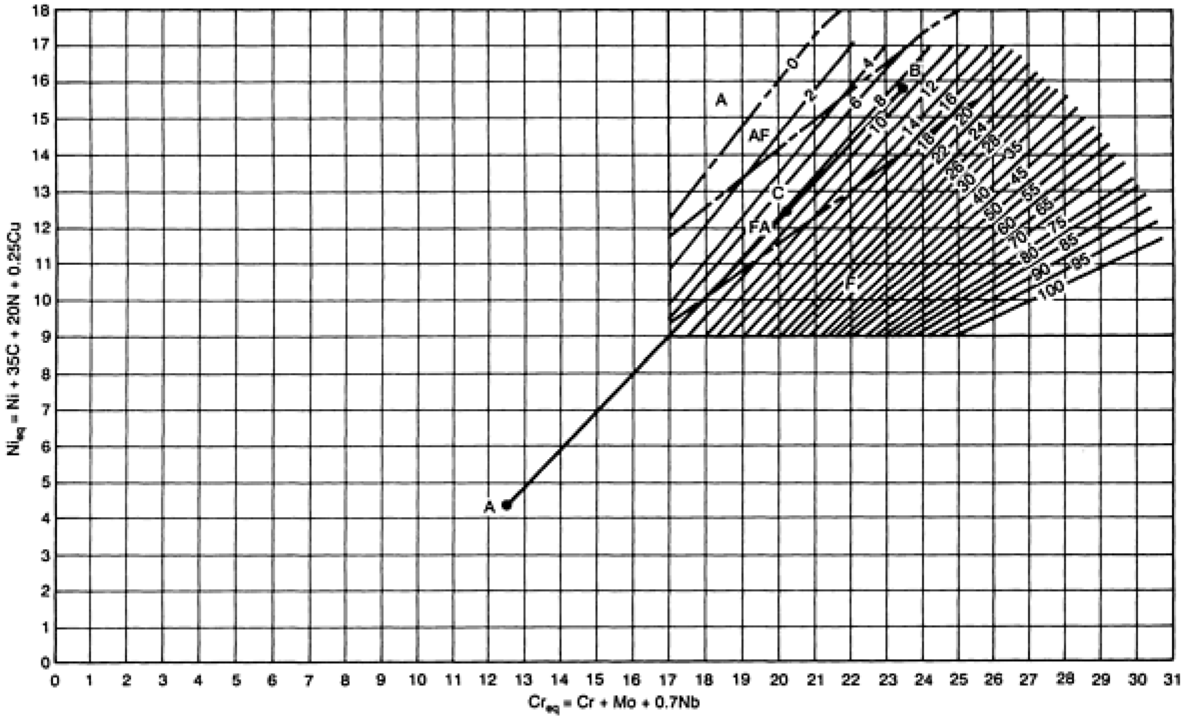

FIG. 9 WRC-1992 DIAGRAM WITH EXPANDED SCALE FOR DILUTION CALCULATIONS, IN WHICH FERRITE IS CALCULATED FOR TYPE 410 STAINLESS STEEL WELDED WITH E309-16 (REFER TO EXAMPLE 1)

Example 2: Type 440A base metal (0.7C-17.0Cr-0.04N), Cr_{eq} = 17.0 and Ni_{eq} = 25.3; to be welded with E309-16 manual electrode (0.05C-23.5Cr-13.0Ni-0.06 N all-weld metal composition), Cr_{eq} = 23.5 and Ni_{eq} = 15.95.

Because the type 440A base metal cannot be plotted on the WRC-1992 diagram, it is necessary to calculate the weld Cr_{eq} and Ni_{eq}. Assuming a 30% dilution, the weld Cr_{eq} = 0.7 (E309-16 Cr_{eq}) + 0.3(440A Cr_{eq}) = 0.7(23.5) + 0.3(17.0) = 21.55; and the weld Ni_{eq} = 0.7(E309-16 Ni_{eq}) + 0.3(440A Ni_{eq}) = 0.7(15.95) + 0.3(25.3) = 18.8. Because the weld-metal composition would clearly lie above and to the left of the 0 FN line in Fig. 4, there would be no ferrite in such a weld deposit. Therefore, E309-16 would not be a good choice for welding type 440A base metal.

Example 3: Type 440A base metal (0.7C-17.0Cr-0.04N), Cr_{eq} = 17.0 and Ni_{eq} = 25.3; to be welded with E312-16 manual electrode (0.10C-29.0Cr-8.8Ni-0.06N), Cr_{eq} = 29.0 and Ni_{eq} = 13.3.

Because the type 440A base metal cannot be plotted on the WRC-1992 diagram, it is necessary to calculate the weld Cr_{eq} and Ni_{eq}. Assuming a 30% dilution, the weld Cr_{eq} = 0.7(E312-16 Cr_{eq}) + 0.3(440A Cr_{eq}) = 0.7(29.0) + 0.3(17) = 25.4; and the weld Ni_{eq} = 0.7(E312-16 Ni_{eq}) + 0.3(440A Ni_{eq}) = 0.7(13.3) + 0.3(25.3) = 16.9. Plotting the weld Cr_{eq} and Ni_{eq} on the WRC-1992 diagram shows this composition corresponds to about 12 FN. Therefore, E312-16 would be a good choice for welding type 440A base metal, if a nonhardenable weld is acceptable.

However, unlike the situation in Example 1, the tie line from the E312-16 electrode all-weld metal composition on the WRC-1992 diagram to the type 440A base-metal composition is nearly perpendicular to the isoferrite lines. This means that large changes in dilution would have a strong effect on weld-metal ferrite content. In particular, it can be calculated that an increase in dilution to just over 35% would put the weld composition into the primary austenite solidification mode and, therefore, at risk of hot cracking. Because higher dilution is normal in submerged arc welding (SAW) and spray-transfer GMAW processes, the avoidance of these processes would be suggested by this example. However, the GMAW process using short-circuiting transfer with type 312 filler wire should be acceptable.

Although the welding filler metals of Examples 1 and 3 seem appropriate choices for avoiding hot cracking in joints in the martensitic stainless steels used in these examples, the weld metals would considerably undermatch the strengths of the base metals, unless the base metals were annealed. The base metals could easily have tensile strength values that exceed 1000 MPa (145 ksi), depending on their condition of heat treatment before welding, whereas the weld metals are likely to have tensile strengths on the order of 550 to 600 MPa (80 to 90 ksi). Furthermore, the austenitic weld metals cannot be hardened by heat treatment. If higher-strength weld metal is essential, then a martensitic filler metal is necessary, which almost always requires a PWHT to obtain reasonable ductility in the weld metal.

Engineering for Use after Postweld Heat Treatment. Because of the major weld-metal strength (hardness) undermatch that often occurs when martensitic stainless steel base metals are welded with austenitic filler metals, these base metals are usually welded with martensitic stainless steel filler metals, and the weldment is subsequently heat treated to improve the properties of the weld area. Heat treatment after welding usually takes one of two forms. The weldment could be either tempered by heat treatment below the austenitic-start temperature or heated above the austenitic-finish temperature to austenitize the entire mass, and then cooled back to nearly room temperature to fully harden it, and then heated again, but to a temperature below the austenitic-start temperature in order to temper the metal to desired properties. Tempering heat treatments at temperatures between the austenitic-start and austenitic-finish temperatures are generally avoided, because they result in a mixture of heavily tempered martensite and fresh untempered martensite, although it is possible to follow such a temper (after cooling to near-room temperature) with a second temper just below the austenitic-start temperature. Such a "double temper" has been successfully used to obtain low hardness and high toughness properties in weldments of CA-6NM (410NiMo).

Tempering at temperatures up to approximately 450 °C (840 °F) has little effect on the hardness of martensitic stainless steels, compared with the hardness of the same steel in the untempered condition. Tempering above 450 °C (840 °F) causes the hardness to fall rapidly with increasing tempering temperatures (Fig. 10), but at the same time, ductility increases. The tempering temperature upper limit is the austenite-start temperature, which is about 800 °C (1470 °F) for the nickel-free alloys, such as types 410, 420, and 440 (Fig. 6 and 8). The addition of nickel to a martensitic alloy reduces the austenite-start temperature appreciably. For CA-6NM, it is about 640 °C (1180 °F).

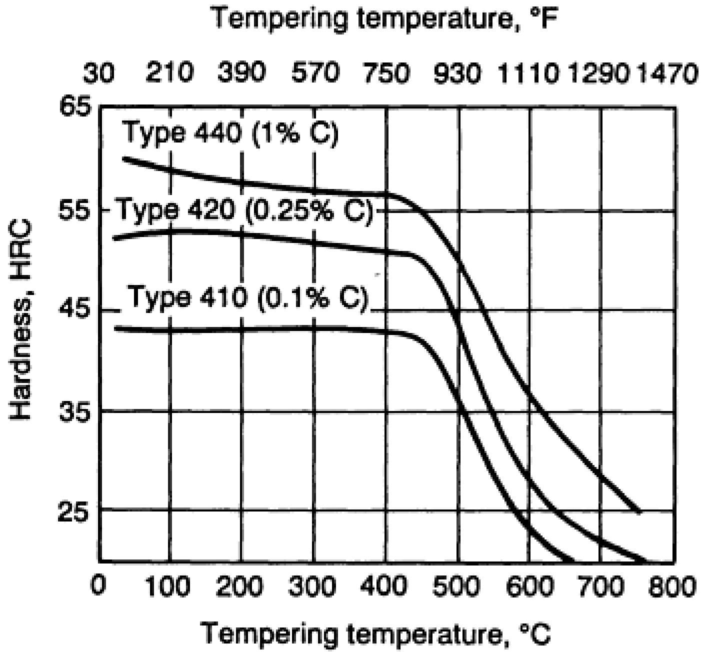

FIG. 10 TEMPERING RESPONSE OF MARTENSITIC STAINLESS STEELS. SOURCE: REF 15

Tempering at temperatures above 450 °C (840 °F), but below the austenitic-start temperature, also can be useful *before* welding, because the base metal is softened and made more ductile, which better accommodates the shrinkage strains that accompany cooling of the weld metal. This helps to avoid weldment cracking during cooling.

With all but the lowest carbon content martensitic stainless steel base metals, high preheat and high interpass temperatures are usually necessary to avoid cracking when the weld metal is also martensitic stainless steel. The alloy CA-6NM only requires approximately 140 °C (280 °F), minimum, whereas types 410 and 420 require approximately 200 °C (390 °F), and type 440 requires approximately 260 °C (500 °F), minimum. However, for a sizable weldment, a safer procedure is to perform all welding at temperatures above the martensite-start temperature of the weld metal, so that no martensite forms in the weld until welding is completed. Preheat and interpass temperatures ranging from 300 to 500 °C (570 to 930 °F) are successfully used. In either case, the weldment must be very slowly cooled to below 100 °C (212 °F) to complete martensite transformation of the weld metal and HAZ without cracking, once welding is finished.

Only after martensite transformation is completed (when the weldment is below 100 °C, or 212 °F) should tempering of the weldment be undertaken. If untransformed austenite is still present, tempering followed by cooling will result in fresh martensite formation when the weldment reaches room temperature, and unexpected hardness (possibly accompanied by cracking) is likely to be encountered.

The availability of matching filler metals is a problem for some martensitic stainless steels. For covered electrodes, only E410-XX and E410NiMo-XX types are commercially significant, as indicated by the ANSI/AWS A5.4 classification. The same compositions are also the only entries in the flux-cored electrode standard, ANSI/AWS A5.22. These compositions, as well as the ER420 classification, are given as solid wires in the ANSI/AWS A5.9 standard. Therefore, Table 1(b), which designates matching filler metals along with selected base metals, has many voids in the filler-metal listings. A few filler-metal producers can produce special orders, such as tubular metal-cored electrodes, in compositions to match the other base metals. Otherwise, the fabricator may be forced to choose a filler metal that does not match the base metal.

One special case deserves mention. Literally thousands of tons of filler metal in the form of tubular wires have been deposited by submerged arc overlay of type 420 and modified-composition weld metals on low-alloy steel rolls for continuous casters. In service, these rolls typically crack from thermal shock. After a period ranging from 6 months to 1 year, their surfaces are sufficiently damaged so that they must be removed from service. If cracking is not too deep, then the cracks can be machined off and the roll rewelded as before.

A very successful welding procedure has been to maintain preheat and interpass temperatures of at least 350 °C (660 °F) from the start to the finish of welding. Because this temperature is above the martensite-start temperature of the weld metal, the weld metal remains as austenite during welding, which may last 1 day or longer. As many as three welding heads, each using 400 to 900 A, can be used on a roll that is about 2 m (6.5 ft) long, which helps to keep the roll hot. The welded roll is covered with insulating material and left to slow cool to nearly room temperature. Then, it is tempered, typically at temperatures of 450 °C (840 °F) or higher, to soften the overlay enough to make it machinable. Molybdenum, vanadium, and, sometimes, niobium, can be added to the basic type 420 composition to modify the tempering response (as well as service performance). Figure 11 illustrates the change in tempering response that results from the alloy modifications.

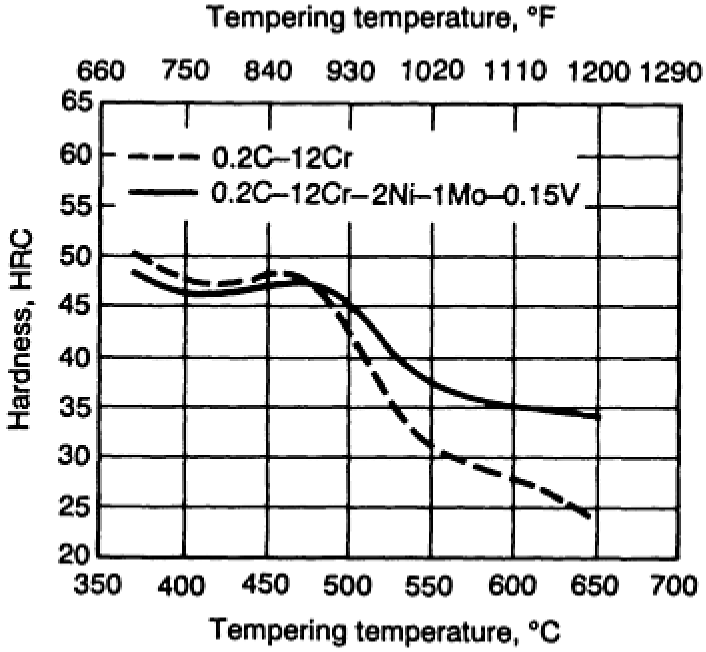

FIG. 11 TEMPERING RESPONSE (2 H AT TEMPERATURE) OF TYPES 420 AND MODIFIED 420 CONTINUOUS CASTER ROLL OVERLAY. SOURCE: REF 16

References cited in this section

11. *METALS HANDBOOK,* 8TH ED., VOL 8, AMERICAN SOCIETY FOR METALS, 1973, P 291
12. E. FOLKHARD, *WELDING METALLURGY OF STAINLESS STEELS,* SPRINGER-VERLAG, 1984, P 17
13. M.A. GROSSMANN AND E.C. BAIN, *PRINCIPLES OF HEAT TREATMENT,* 5TH ED., AMERICAN SOCIETY FOR METALS, 1964, P 49
14. *ATLAS OF ISOTHERMAL TRANSFORMATION AND COOLING TRANSFORMATION DIAGRAMS,* AMERICAN SOCIETY FOR METALS, 1977, P 324
15. "CARPENTER STAINLESS STEELS--WORKING DATA," CARPENTER TECHNOLOGY CORPORATION, 1983
16. "ENGINEERING REPORT ES-489," THE LINCOLN ELECTRIC COMPANY, 1991

Welding of Stainless Steels

D.J. Kotecki, The Lincoln Electric Company

Ferritic Stainless Steels

Basic Metallurgy. There are essentially three generations of ferritic stainless steels. The first--in which carbon is not very low, so that chromium needs to be high--dates from the early decades of the 20th century, when the decarburization of iron-chromium alloys was quite inefficient. The prototype alloy is type 430 stainless, typically 0.08C-17Cr. Figure 6(b) indicates that this alloy would be fully ferritic at temperatures above approximately 1250 °C (2280 °F). At lower temperatures, ferrite and austenite would co-exist to about 1030 °C (1890 °F). Further cooling would cause carbides to appear as well. Then, at about 920 °C (1690 °F), the austenite would disappear and only ferrite and some carbides would remain at room temperature, under equilibrium conditions. Type 446 stainless (typically, 0.1C-25Cr) is also of this first generation, and it behaves similarly.

It is something of a misnomer to call these first-generation ferritic stainless steels "ferritic" because some austenite appears at certain high temperatures. This austenite has a beneficial effect in that it retards grain growth, which has an important embrittling effect when it occurs. However, because it also tends to transform to martensite under welding conditions, first-generation ferritic stainless steels generally have hard spots after welding (see Fig. 12). In addition, normal weld cooling will cause some supersaturation of the ferrite with carbon, because of declining carbon solubility in the ferrite with falling temperatures below 950 °C (1740 °F). A PWHT at about 760 °C (1400 °F) simultaneously tempers the martensite where austenite existed at higher temperatures, and causes the excess carbon in the ferrite to precipitate as carbides. Both of these effects act to significantly improve the ductility and toughness of type 430 after welding. Type 446 is not ductile at room temperature, even before welding.

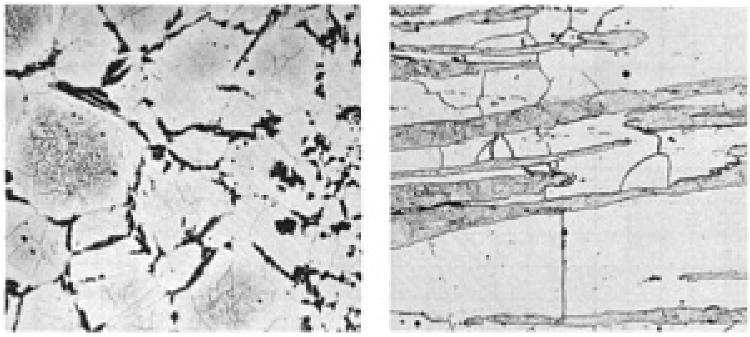

FIG. 12 MICROSTRUCTURE OF TYPE 430 FERRITE STAINLESS STEEL. (A) BASE METAL, 25 MM (1 IN.) THICK PLATE, AS HOT ROLLED; SPECIMEN FROM LONGITUDINAL DIRECTION. FERRITE MATRIX CONTAINS ELONGATED LAYERS OF MARTENSITE AND TRANSFORMATION PRODUCTS. PICRAL AND HYDROCHLORIC ACID ETCH, 100×. (B) WELD HEAT-AFFECTED ZONE. PROGRESSIVE COARSENING OF FERRITE GRAINS AND DISTRIBUTION OF MARTENSITE (DARK) AT GRAIN BOUNDARIES AS DISTANCE FROM WELD ZONE (RIGHT) INCREASES. TEMPERED 10 MIN AT 593 °C (1100 °F) TO CAUSE MARTENSITE TO BE DARK-ETCHING. PICRIC ACID (5%) AND HYDROCHLORIC ACID (3%) IN ALCOHOL, 150×. SOURCE: REF 17

The second generation of ferritic stainless steels have lower-carbon and low-nitrogen contents, to which a stabilizer is added to tie up whatever carbon and/or nitrogen is present. The prototype of this second generation is type 409, typically 0.04C-11Cr-0.5Ti. The titanium ties up both carbon and nitrogen, leaving all of the chromium free. Surplus titanium is also a ferritizer. Therefore, the alloy is ferritic at all temperatures. Type 405 is a similar alloy, but it is stabilized with aluminum, which ties up nitrogen, but not carbon. Aluminum is also a powerful ferritizer. The various carbides and/or nitrides produced by the addition of stabilizers help to resist embrittling grain growth in second-generation ferritic stainless steels during welding.

The third generation of ferritic stainless steels arose circa 1970, with the advent of more-efficient decarburization techniques in steelmaking. Carbon and nitrogen levels are typically 0.02% or less, and stabilizers, such as titanium and/or niobium, are often added to tie up any free interstitials. The prototype alloy is type 444 (18Cr-2Mo). There are also a number of proprietary alloys that are not yet covered by an ASTM designation.

The third-generation ferritic stainless steel alloys are austenite-free at all temperatures. However, they can be embrittled by the formation of intermetallic phases at elevated temperatures. They are sensitive to "475 °C (885 °F) embrittlement," because of precipitation of the chromium-rich α' phase, and to embrittlement by σ, χ, and other phases at higher temperatures. Figure 5 indicates that the αphase is approximately FeCr, whereas α' is a chromium-rich ferrite that precipitates from the iron-rich ferrite. The χ phase is a complex iron-chromium-molybdenum intermetallic, which can only appear in molybdenum-bearing steels. Except in the case of α' embrittlement, which disappears rapidly upon exposure to temperatures above approximately 575 °C (1070 °F), temperatures that are high enough to dissolve all of these phases are also high enough to cause severe grain growth in the ferrite. Because this grain growth also has an embrittling effect, heat treatment alone is not usually desirable. In general, hot working to reduce grain size is necessary to restore full ductility and toughness once the alloy is embrittled by σ, χ, or other high-temperature intermetallic compounds.

Ferritic stainless steels of the third generation can be severely embrittled by interstitial contamination during welding, because of improper or contaminated gas shielding. Although nitrogen contamination is most common, carbon, hydrogen, and oxygen also have detrimental effects on ductility and toughness. Cleanliness, inert backing gas, and high-purity gas shielding conducted with precautions to avoid disturbances are essential if the best weldment properties are to be obtained.

Ferritic Stainless Steel Base Metals. Tables 2(a) and 2(b) list selected ferritic stainless steel base metals, their minimum mechanical property requirements, and some corresponding welding filler metals. Ferritic stainless steels cannot offer the range of mechanical properties (depending on heat treatment) that martensitic stainless steels offer. Whereas the martensitic stainless steels offer only moderate corrosion resistance, that of the ferritic stainless steels can range from moderate, for types 430 or 409, to outstanding, for the third-generation "super ferritics" such as type 444 and UNS Nos. S44627, S44635, S44660, S44700, and S44800. Types 405 and 409 have resistance to atmospheric corrosion. Type 430 also resists mild oxidizing acids and organic acids, and is therefore used in food-handling equipment, sinks, and such. The super ferritics offer exceptional resistance to localized corrosion induced by exposure to aqueous chlorides. Localized corrosion, such as pitting, crevice corrosion, and stress-corrosion cracking, are problems that plague many austenitic stainless steels. Therefore, the super ferritics are often used in heat exchangers and piping systems for chloride-bearing aqueous solutions and seawater.

TABLE 2(A) COMPOSITIONS OF NOMINALLY FERRITIC STAINLESS STEELS

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	COMPOSITION, WT%									
			C	**Mn**	**P (max)**	**S**	**Si (max)**	**Cr**	**Ni**	**Mo**	**N**	OTHER
XM-34	S18200	A 582	0.08 MAX	2.50 MAX	0.04	0.15 MIN	1.00	17.50-19.50	. . .	1.50-2.50	. . .	
. . .	S18235	A 582	0.025 MAX	0.50 MAX	0.030	0.15-0.30	1.00	17.50-18.50	1.00 MAX	2.00-2.50	0.025 MAX	**C + N, 0.035 MAX; TI, 0.30-1.00**
405	S40500	A 240	0.08 MAX	1.00 MAX	0.040	0.030 MAX	1.00	11.50-14.50	0.60 MAX	. . .	. . .	**AL, 0.10 TO 0.30**
409	S40900	A 240	0.08 MAX	1.00 MAX	0.045	0.030 MAX	1.00	10.50-11.75	0.50 MAX	. . .	. . .	**TI, 6 × C MIN TO 0.75 MAX**
429	S42900	A 240	0.12 MAX	1.00 MAX	0.040	0.030 MAX	1.00	14.00-16.00	0.75 MAX	. . .	. . .	**. . .**
430	S43000	A 240	0.12 MAX	1.00 MAX	0.040	0.030 MAX	1.00	16.00-18.00	0.75 MAX	. . .	. . .	**. . .**
430F	S43020	A 473	0.12 MAX	1.25 MAX	0.06	0.15 MAX	1.00	16.00-18.00	0.75 MAX	0.60 MAX	. . .	
430F SE	S43023	A 473	0.12 MAX	1.25 MAX	0.06	0.06 MAX	1.00	16.00-18.00	0.75 MAX	. . .	. . .	**SE, 0.15 MIN**
439	S43035	A 240	0.07 MAX	1.00 MAX	0.040	0.030 MAX	1.00	17.00-19.00	0.50 MAX	. . .	0.04 MAX	**TI, 0.20 + 4 (C + N) MIN TO 1.10 MAX; AL, 0.15 MAX**
444	S44400	A 240	0.025 MAX	1.00 MAX	0.040	0.030 MAX	1.00	17.5-19.5	1.00 MAX	1.75-2.50	0.035 MAX	**TI + NB, 0.20 + 4 (C + N) MIN TO 0.80 MAX**
446	S44600	A 276	0.20 MAX	1.50 MAX	0.040	0.030 MAX	1.00	23.00-27.00	. . .	. . .	0.25 MAX	**. . .**
XM-27	S44627	A 240	0.010	0.40	0.020	0.020 MAX	0.40	25.00-27.50	0.50 MAX	0.75-1.50	0.015 MAX	**NB + TA, 0.05-0.20; AL, 0.20 MAX; NI + CU, 0.50 MAX**
XM-33	S44626	A 240	0.06	0.75	0.040	0.020	0.75	25.00-	0.50	0.75-	0.04	**TI, 7 × (C + N)**

			MAX	MAX		MAX		27.00	MAX	1.50	MAX	**MIN AND 0.02-1.00; CU, 0.20 MAX**
. . .	S44635	A 240	0.025 MAX	1.00 MAX	0.040	0.030 MAX	0.75	24.5-26.0	3.5-4.5	3.5-4.5	0.035 MAX	**TI + NB, 0.20 + 4 × (C + N) MIN TO 0.80 MAX**
. . .	S44660	A 240	0.030 MAX	1.00 MAX	0.040	0.030 MAX	1.00	25.0-28.0	1.0-3.50	3.00-4.00	0.040 MAX	**TI + NB, 6 × (C + N) MIN AND 0.20 TO 1.00**
29-4	S44700	A 240	0.010 MAX(A)	0.30 MAX	0.025	0.020 MAX	0.20	28.0-30.0	0.15 MAX	3.5-4.2	0.020 MAX(A)	
29-4-2	S44800	A 240	0.010 MAX(A)	0.30 MAX	0.025	0.020 MAX	0.20	28.0-30.0	2.0-2.5	3.5-4.2	0.020 MAX(A)	
. . .	S44735	A 240	0.030 MAX	1.00 MAX	0.040	0.030 MAX	1.00	28.00-30.00	1.00 MAX	3.60-4.20	0.045 MAX	**TI + NB, 6 × (C + N) MIN AND 0.20 TO 1.00**
CB-30	. . .	A 743	0.30 MAX	1.00 MAX	0.04	0.04 MAX	1.50	18.0-21.0	2.00 MAX	. . .	. . .	**. . .**
CC-50	**. . .**	**A 743**	**0.50 MAX**	**1.00 MAX**	**0.04**	**0.04 MAX**	**1.50**	**26.0-30.0**	**4.00 MAX**	**. . .**	**. . .**	**. . .**

(A) (C + N) = 0.025 MAX, CU = 0.015 MAX

TABLE 2(B) PROPERTIES AND FILLER METALS FOR NOMINALLY FERRITIC STAINLESS STEELS

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	**TENSILE STRENGTH**		**YIELD STRENGTH**		**ELONGATION, %**	MATCHING FILLER METALS		
			MPA	**KSI**	**MPA**	**KSI**		**A5.4**	**A5.9**$^{(A)}$	A5.22
XM-34	S18200	A 582	NS	NS	NS	NS	NS	NCW	NCW	**NCW**
. . .	S18235	A 582	NS	NS	NS	NS	NS	NCW	NCW	**NCW**
405	S40500	A 240	415	60	170	25	20	. . .	. . .	**E409T-X**
409	S40900	A 240	380	55	170	25	20	. . .	ER409	**E409T-X**
429	S42900	A 240	450	65	205	30	22	. . .	. . .	**. . .**
430	S43000	A 240	450	65	205	30	22	E430-XX	ER430	**E430T-X**
430F	S43020	A 473	480	70	275	40	20	NCW	NCW	**NCW**
430F SE	S43023	A 473	480	70	275	40	20	NCW	NCW	**NCW**
439	S43035	A 240	415	60	205	30	22	. . .	. . .	**. . .**
444	S44400	A 240	415	60	275	40	20	. . .	. . .	**. . .**
446	S44600	A 276	450	65	275	40	20	. . .	. . .	**. . .**
XM-27	S44627	A 240	450	65	275	40	22	. . .	ER446LMO	**. . .**
XM-33	S44626	A 240	470	68	310	45	20	. . .	. . .	**. . .**
. . .	S44635	A 240	620	90	520	75	20	. . .	. . .	**. . .**
. . .	S44660	A 240	585	85	450	65	18	. . .	. . .	**. . .**
29-4	S44700	A 240	550	80	415	60	20	. . .	. . .	**. . .**
29-4-2	S44800	A 240	550	80	415	60	20	. . .	. . .	**. . .**
. . .	S44735	A 240	550	80	415	60	18	. . .	. . .	**. . .**
CB-30	. . .	A 743	450	65	205	30	. . .	. . .	. . .	**. . .**
CC-50	**. . .**	**A 743**	**380**	**55**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**

NS, not specified; NCW, not considered weldable.

(A) METAL-CORED ELECTRODES, INDICATED BY A "C" IN PLACE OF THE "R" IN THE CLASSIFICATION, ARE ALSO INCLUDED.

Because of 475 °C (885 °F) embrittlement and σ-phase formation, long-term service temperatures for most ferritic stainless steels are usually limited to 250 °C (480 °F), maximum. However, 475 °C (885 °F) embrittlement does not seem to be a serious problem in the low-chromium types 409 and 405, which have been extensively used in automobile exhaust systems (where temperatures can exceed 575 °C, or 1070 °F, at times, which can dissolve α'). Another exception to the maximum service temperature limit is type 446, which offers outstanding resistance to scaling in air at high temperatures, because of its high chromium content. It is used in compressive loading, as the supports in heat-treat furnaces and the like, where brittleness at room temperature is not an important issue. Because ferritic stainless steels lose toughness and ductility at low temperatures, their use below ambient temperatures is very limited.

The ferritic stainless steels generally require rapid cooling from hot-working temperatures to avoid grain growth and embrittlement from the α phase. As a result, most ferritic stainless steel is used in relatively thin gages, especially in alloys that are high in chromium. The super ferritics are limited to thin plate, sheet, and tube forms. To avoid embrittlement in welding, the general rule is to "weld cold," which means to weld with either no or low preheating; a low interpass temperature; and a low level of welding heat input, commensurate with adequate fusion, to avoid cold laps and other defects.

Engineering for Use in the As-Welded Condition. In contrast to martensitic stainless steels, weldments in ferritic stainless steels (at least in the stabilized grades or super ferritics, that is, second- and third-generation alloys), are usually used in the as-welded condition. If the "weld cold" rule is followed, then embrittlement that is due to grain coarsening in the HAZ can be avoided. However, it is not as easy to avoid coarse grains in the fusion zone if the weld metal is fully ferritic. As a result, a considerable amount (but not all) of the filler metal used to join ferritic stainless steels is austenitic, usually with considerable ferrite.

If an austenitic filler metal is contemplated, then consideration should be given to differences in corrosion resistance (especially localized corrosion resistance) across the joint, as well as to differences in properties, such as the coefficient of thermal expansion (CTE) between austenitic filler metal and ferritic base metal. Nickel-base alloy filler metals are another possible alternative, especially for the third-generation ferritic stainless steels, which have the highest corrosion resistance requirements. High molybdenum content nickel-base alloy filler metals of the ANSI/AWS A5.11 specification, covered-electrode classifications ENiCrMo-3, ENiCrMo-4, and ENiCrMo-10, as well as the ANSI/AWS A5.14 specification, solid-wire classifications ERNiCrMo-3, ERNiCrMo-4, and ERNiCrMo-10, have been successfully used for the most corrosion resistant ferritic stainless steels.

Undoubtedly, the greatest quantity of ferritic stainless steel welded at the present time is type 409, which is used for automobile exhaust system components. The welds are virtually all single-pass fillet or lap welds. The most common filler metal is a matching composition supplied as a 1.1 mm (0.045 in.) diameter metal-cored wire (ANSI/AWS A5.9-92 classification EC409 or EC409Cb), which is used for gas-metal arc welding with argon-oxygen gas mixtures, such as the popular 98Ar-2O_2 mixture. This filler metal is often used in robotic welding, at high travel speeds, in the vertical-down position on rotating pieces that are 18 gage or less in thickness. Welding is done in a short-circuiting transfer mode, or nearly so.

Example 4:

An alternative procedure for welding type 409 base metal in automotive exhaust applications is to use 0.9 mm (0.035 in.) diameter solid austenitic stainless wire, in the short-circuiting transfer mode, with an argon-oxygen gas mixture. Because of the thin gage of the base metal, dilution that is on the order of 40% can be anticipated. Because the filler metal ER308LSi would exceed the corrosion resistance of the type 409 base metal, it could be considered. If the type 409 composition were 0.04C-0.5Mn-0.4Si-11.0Cr-0.5Ti (WRC-1988 Cr_{eq} = 11.0 and Ni_{eq} = 1.4; FN, off the diagram), and if the ER308LSi composition were 0.02C-1.5Mn-0.8Si-19.7Cr-10.2Ni-0.2Mo-0.04N (WRC-1988 Cr_{eq} = 19.9 and Ni_{eq} = 11.7; FN = 9.7), then the 40% dilution weld metal would have Cr_{eq} = 16.34, Ni_{eq} = 7.58, and FN would be off the diagram. Recalculating the chromium and nickel equivalents according to Schaeffler (Fig. 1), the weld metal would be deeply into the A + M + F field, that is, on the order of 50% martensite could be expected in the weld metal. This could be a problem for cracking or forming after welding. Therefore, ER308LSi may not be a good choice for welding type 409 base metal.

Example 5:

Assume that ER309LSi is substituted for the ER308LSi of Example 4. If the ER309LSi composition were 0.02C-1.5Mn-0.8Si-23.5Cr-13.5Ni-0.2Mo-0.04N (WRC-1988 Cr_{eq} = 23.7 and Ni_{eq} = 15.0; FN = 12.8), then the 40% dilution weld metal would have Cr_{eq} = 18.6, Ni_{eq} = 9.56, and FN = 12.6. Recalculating the chromium and nickel equivalents according to Schaeffler (Fig. 1) puts the weld metal just at the upper edge of the A + M + F field, that is, a trace of martensite, at most, would be present in the weld metal. Furthermore, because the calculated weld-metal FN is almost exactly the same as the calculated wire FN, the tie line from the wire composition to the base-metal composition must be nearly parallel to the isoferrite lines in the WRC-1988 diagram. This means that the weld-metal FN will be immune to changes in dilution, although excess dilution should be avoided because it would lead to more martensite. Thus, the ER309LSi filler metal would be a good choice for welding type 409 base metal.

The third generation of ferritic stainless steels, the super ferritics, are often supplied in the form of thin-walled tubing and are used in heat exchangers that handle chloride solutions. Tube-to-tube or tube-to-tubesheet weld joints are often made by orbital GTAW machines at relatively high speeds, without filler metal. The tube-to-tube weld is, of course, fully ferritic, and the high-speed "weld cold" procedure avoids excessive grain growth in materials such as type 444 ferritic stainless steel (18Cr-2Mo). However, the tubesheet in such a heat exchanger is often not type 444, but type 316L.

Example 6:

Type 444 tube welded to type 316L tube sheet by the GTAW process, without filler metal, makes an interesting material combination. If the Type 444 is 0.01C-0.5Mn-0.2Si-18.0Cr-2.0Mo-0.01N-0.4Nb (WRC-1988 Cr_{eq} = 20.28 and Ni_{eq} = 0.55; FN, out of range), and if the type 316L is 0.02C-1.5Mn-0.4Si-17.0Cr-2.2Mo-0.03N (WRC-1988 Cr_{eq} = 19.2, Ni_{eq} = 12.8, and FN = 3.4), and if each base metal is assumed to contribute equally to the weld metal, then the weld metal will have Cr_{eq} = 19.7, Ni_{eq} = 6.7, and FN will be out of range. Recalculating chromium and nickel equivalents for the Schaeffler diagram indicates approximately 40% ferrite, with the remainder being primarily austenite, with some martensite. This would not be the best situation, in terms of weld-metal ductility, but many such combinations have been welded this way and have provided satisfactory service.

Engineering for Use in the Postweld Heat-Treated Condition. In general, only the first-generation ferritic stainless steels are postweld heat treated, especially type 430. If it were welded with matching filler metal (or no filler metal), then both the weld metal and the HAZ would contain fresh martensite in the as-welded condition. In addition, some carbon would have gone into solution in the ferrite at elevated temperatures, and the rapid cooling after welding would result in ferrite in both the weld metal and the HAZ being supersaturated with carbon. Thus, the joint would be quite brittle, compared with the unwelded condition. Ductility could be significantly improved by a PWHT at 760 °C (1400 °F) for 1 h, followed by rapid cooling to avoid the 475 °C (885 °F) embrittlement phenomenon.

Example 7:

Type 430 could also be welded with type 308 or 308L filler metals, which are readily available and would exceed the corrosion resistance of the type 430 base metal in most environments. Then, only the HAZ would suffer embrittlement that is due to welding. For a type 430 composition of 0.06C-0.5Mn-0.2Si-17Cr-0.02N (WRC-88 Cr_{eq} = 17.0 and Ni_{eq} = 2.5; FN, off the diagram), and an E308L-16 electrode with all-weld metal composition of 0.03C-1.0Mn-0.4Si-19Cr-10Ni-0.08Mo-0.06N (WRC-1988 Cr_{eq} = 19.08 and Ni_{eq} = 12.25; calculated FN = 4.4), with 30% dilution, the first weld pass would have Cr_{eq} = 18.46, Ni_{eq} = 9.33, and the calculated FN = 12.9. Further, the weldment could be postweld heat treated for 1 h at 760 °C (1400 °F) to improve the HAZ ductility without harming the weld metal, and a very satisfactory joint would result.

Reference cited in this section

17. *METALS HANDBOOK,* 8TH ED., VOL 7, AMERICAN SOCIETY FOR METALS, 1972, P 144-145

Welding of Stainless Steels

D.J. Kotecki, The Lincoln Electric Company

Austenitic Stainless Steels

Basic Metallurgy. Nominally austenitic stainless steels cover a very broad spectrum of compositions. Many, but by no means all, solidify at least partly as ferrite, which transforms under equilibrium cooling, or hot working, to austenite at high temperatures (above 1000 °C, or 1830 °F) by a diffusion-controlled reaction. The austenite is stable, or at least metastable, to room temperature and below. Some of the leaner austenitic alloys can be induced to transform to martensite, at least in part, by cold work and/or by refrigeration to cryogenic temperatures.

In order to understand the weldability of nominally austenitic stainless steels, it is necessary to first understand their solidification and high-temperature transformation behavior. Austenitic stainless steel compositions are based on a balance between alloy elements that promote ferrite formation and those that promote austenite formation. The prototype ferritizing element is chromium, but molybdenum, niobium, titanium, aluminum, tungsten, and vanadium also promote ferrite. The prototype austenitizing element is nickel, but carbon, nitrogen, and copper all promote transformation of ferrite to austenite at high temperatures. In addition, while manganese does not seem to promote transformation of ferrite to austenite at high temperatures, it clearly does tend to stabilize austenite with respect to transformation to martensite at low temperatures. Further, manganese promotes the solubility of nitrogen in steel (as does chromium), making possible a low-nickel family of austenitic stainless steels that are high in manganese and nitrogen.

The solidification and high-temperature transformation behavior of austenitic stainless steels can be understood by examining sections of the ternary iron-chromium-nickel phase diagram. Figure 13 presents three levels of iron: 70 wt%, 60 wt%, and 55 wt%. Each diagram has a eutectic triangle (labeled "γ+δ+L"), which indicates compositions in which austenite, ferrite, and liquid metal can all coexist.

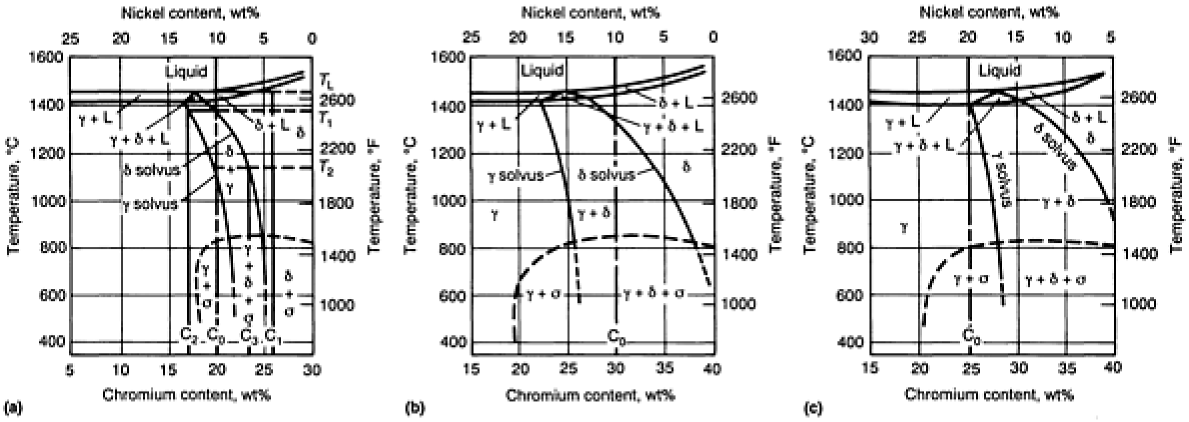

FIG. 13 IRON-CHROMIUM-NICKEL PSEUDO-BINARY DIAGRAM. (A) 70 WT% FE. (B) 60 WT% FE. (C) 55 WT% FE. SOURCE: REF 10

On the left (nickel-rich) side of each diagram are compositions that solidify entirely as austenite and remain austenite to room temperature, for example, an alloy of 10Cr-20Ni in the 70% iron diagram. Close to, and within the left side of, the eutectic triangle are compositions (for example, 17Cr-13Ni) that solidify as austenite in the early stages of solidification, but in which some ferrite will appear in the last stages of solidification. This is called "primary austenite solidification." Further cooling can cause that ferrite to transform to austenite under equilibrium conditions, but under the relatively rapid cooling conditions of welding, some of this ferrite may survive to room temperature.

Close to, and within the right side of, the eutectic triangle are compositions (for example, 19Cr-11Ni) that solidify as ferrite in the early stages of solidification, but in which some austenite will appear in the last stages of solidification. This is called "primary ferrite solidification." Further cooling can cause that ferrite to transform to austenite under equilibrium conditions, but under the relatively rapid cooling conditions of welding, some of this ferrite may, and usually does, survive to room temperature.

Further to the right in Fig. 13 are compositions that solidify entirely as ferrite, but, upon further cooling under equilibrium conditions, transform either partly or entirely to austenite. An alloy of 20Cr-10Ni would be entirely ferrite when solidification was completed, but would begin to transform to austenite at a temperature (T_1 in Fig. 13) slightly below the end of solidification, and would be completely transformed to austenite at a still lower temperature (T_2 in Fig. 13). However, an alloy of 23Cr-7Ni would only partially transform to austenite, that is, some ferrite would be stable at room temperature. Under the relatively rapid cooling conditions of welding, not as much ferrite transforms to austenite, as Fig. 13 indicates, because the diffusion of atoms must occur. As a result, the 20Cr-10Ni alloy weld metal would still contain some ferrite at room temperature, and the 23Cr-7Ni alloy would contain more ferrite than Fig. 13 indicates.

Finally, there are compositions (still further to the right of the eutectic triangle in Fig. 13) that will not form any austenite at any temperature. These are ferritic stainless steels.

Considerable research has shown that alloys that solidify as primary ferrite are the most resistant to hot cracking. Alloys that solidify entirely as austenite are the most susceptible to hot cracking. Alloys that solidify as primary austenite are more susceptible than all alloys in this system, except those that solidify entirely as austenite. Alloys that solidify entirely as ferrite are more resistant to hot cracking than any other alloys, except those that solidify as primary ferrite. Therefore, from a welding viewpoint, the preferred order of solidification is primary ferrite, fully ferritic, primary austenite, and, last, fully austenitic. In the WRC-1988 diagram (Fig. 3), the dashed lines divide the range of chromium equivalents and nickel equivalents into those that solidify fully austenitic (Region A), those that solidify as primary austenite (Region AF), those that solidify as primary ferrite (Region FA), and those that solidify fully ferritic (Region F).

Figures 3 and 13, when considered together, indicate that the ferrite that remains in a nominally austenitic weld metal when it reaches room temperature is only a fraction of that which existed during solidification. Most of the initial ferrite is transformed to austenite. Thus, the more-fundamental piece of information about a given weld metal is its solidification mode. However, this is difficult and time-consuming to assess, whereas the Ferrite Number, measured at room temperature after the fact, is easy to measure using any number of instruments calibrated to the ANSI/AWS A4.2 specification, and the interlaboratory reproducibility of measurement is very high. The conventional wisdom accumulated over numerous years indicates that a 4 FN minimum ill ensure freedom from hot cracking problems in most nominally austenitic stainless steel weld metals. An examination of Fig. 3 indicates that 4 FN is comfortably within the FA region for lower-alloyed weld metals and barely within the FA region for highly alloyed weld metals.

Several caveats need to be remembered when applying the "4 FN minimum" convention. First, it is based on measurement *along the top centerline of an as-welded weld bead.* This means that applying such a rule, either after PWHT or at all points across a multipass overlay, is inappropriate because heat treatment or reheating by a subsequent weld pass reduces ferrite toward its equilibrium value (which may be zero).

Second, this convention derives from experience with normal impurity levels (primarily sulfur and phosphorus). Extraordinarily high impurities, as in free-machining stainless steels, often renders the metal susceptible to hot cracking at any ferrite level. On the other hand, extraordinarily low impurity levels can make low or zero ferrite weld metals fairly crack resistant.

Third, one cannot determine, from a ferrite measurement on unwelded base metal, what its hot cracking susceptibility in autogenous (no filler metal) welding is likely to be. Because hot working erases ferrite, hot-rolled plate is likely to indicate zero ferrite. On the other hand, cold working causes lean alloys, such as types 201, 301, 302, and 304, to transform to martensite, which is ferromagnetic and gives a false ferrite reading with instruments based on magnetic principles. Ferrite can be either estimated from the composition of a base metal (including nitrogen), using the WRC-1988 diagram, or measured by first making a gas-tungsten arc melt of the base metal without using filler metal, and then using a calibrated instrument along the weld centerline.

Fourth, a certificate of test from a filler metal manufacturer that indicates a FN greater than 4 only indicates that the filler metal was properly designed. It cannot take into account the ability of a careless welder to change the Ferrite Number by permitting nitrogen from the air to enter the welding arc. Practices such as long arc length in shielded metal arc welding (SMAW) or self-shielded flux-cored arc welding (FCAW); the disturbance of shielding gas by drafts in gas-shielded welding; or the loss of flux cover, as evidenced by arc "flashing" in the SAW process, can produce enough nitrogen pickup to reduce the ferrite of an otherwise satisfactory weld deposit to an unacceptably low level, which makes hot cracking possible.

Fifth, very lean alloy weld metals, such as ER16-8-2, can solidify as crack-resistant primary ferrite, even when the weld metal measures nearly zero FN.

Sixth, the minimum 4 FN convention is really only applicable to weld metals that solidify under cooling rates normal for conventional SMAW, GMAW, GTAW, and single-electrode SAW. The slow cooling rates of strip cladding can produce lower than normal FN measurements when the weld metal solidifies as primary ferrite. Furthermore, the very rapid cooling rates associated with EBW or LBW can produce marked departures from the expected solidification mode and ferrite content (see Ref 18). Nevertheless, the minimum 4 FN convention has been a very successful guideline in normal welded fabrication of nominally austenitic stainless steels.

Arc welding heats the base metal adjacent to the weld deposit, which can produce undesirable metallurgical reactions. In austenitic stainless steels, a principal concern is grain-boundary carbide precipitation. Whatever carbides were initially present before welding can be dissolved at temperatures above approximately 1000 °C (1830 °F). Then, subsequent slow cooling through the temperature range from approximately 950 °C (1740 °F) down to approximately 500 °C (900 °F) can cause chromium carbides to precipitate again, usually on austenite grain boundaries. This precipitation can leave the metal adjacent to the grain boundaries depleted in chromium and, therefore, susceptible to intergranular corrosion. The steel is said to be "sensitized." Improper heat treatment can do this to an entire base metal.

Figure 14 gives the times and temperatures that produce sensitization. In particular, it indicates that carbides, although they form quickly, must be quite extensive in order to make the metal susceptible to intergranular corrosion. Generally, a selection of low-carbon (less than 0.03%) base metal is nearly immune to sensitization that is due to welding. Another alternative is to choose a base metal to which a strong carbide-forming element (titanium or niobium) is added, which ties up the available carbon and leaves the chromium free to provide corrosion resistance.

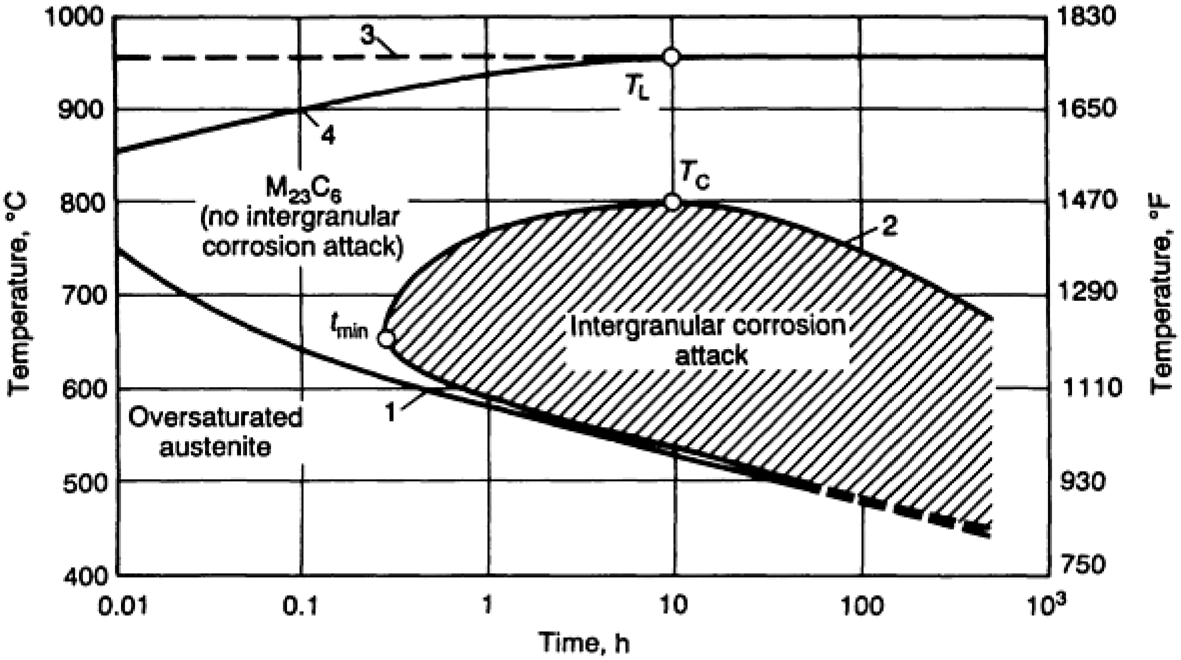

FIG. 14 PRECIPITATION OF CARBIDE $M_{23}C_6$ AND AREA OF INTERGRANULAR CORROSION ATTACK OF STAINLESS STEEL GRADE AISI 304 (0.042C-0.59SI-1.23MN-17.46CR-10.58NI-0.046N). SOURCE: REF 19

The σ phase precipitates slowly in austenite, but more rapidly in ferrite. In nominally austenitic weld metal with some ferrite, cooling rates are still high enough that σ is seldom encountered in the as-welded condition. However, a PWHT at 700 to 900 °C (1290 to 1650 °F) for even 1 h will convert some ferrite to the σ phase, as indicated by Fig. 15. High levels of chromium, molybdenum, silicon, titanium, and/or niobium can greatly accelerate σ formation. The segregation of these elements during solidification can promote σ formation. Because of sensitization and the σ phase, it is best to avoid the heat treatment of nominally austenitic weld metals at temperatures between 500 and 900 °C (930 and 1650 °F).

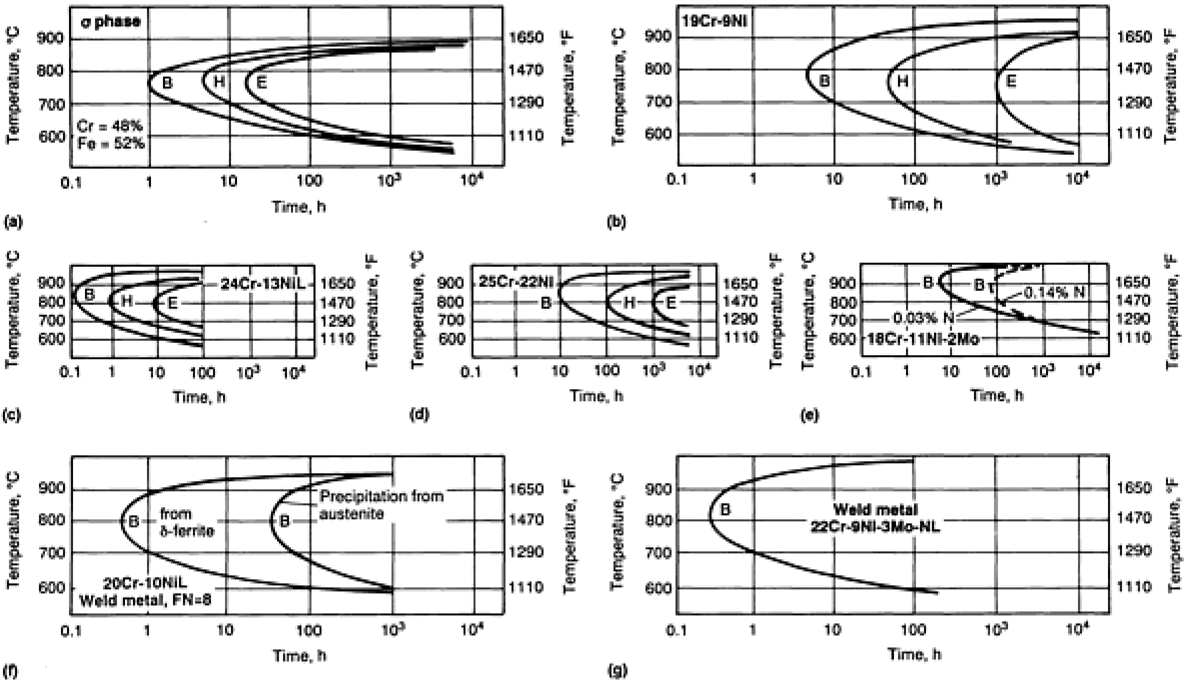

FIG. 15 DIAGRAM OF σ-PHASE PRECIPITATION IN SOLUTION ANNEALED AND QUENCHED STAINLESS STEELS AND IN WELD METAL (AS-WELDED STATE). B, BEGINNING OF PRECIPITATION; H, HALF-WAY THROUGH PRECIPITATION; E, END OF PRECIPITATION. (A) PURE σPHASE. (B) AUSTENITIC 19CR-8NI STEEL (AISI 304). (C) AUSTENITIC 24CR-13NI STEEL (AISI 309S) WITH APPROXIMATELY 15% DELTA FERRITE. (D) AUSTENITIC 25CR-22NI STEEL (AISI 310). (E) AUSTENITIC 18CR-11NI-2MO STEELS (AISI 316) WITH VARYING NITROGEN CONTENTS. (F) AUSTENITIC WELD METAL WITH LOW CONTENT OF DELTA FERRITE (FN = 8) OF THE TYPE 20CR-10NI-L (AWS E308L). (G) AUSTENITIC-FERRITIC DUPLEX WELD METAL WITH 24 TO 27% DELTA FERRITE OF THE TYPE 22CR-9NI-3MO-LN. SOURCE: REF 20

Austenitic Stainless Steel Base Metals. Tables 3(a) and 3(b) list numerous nominally austenitic stainless steel base metals and their minimum mechanical property requirements, as well as corresponding welding filler metals. The yield strengths of these materials are rather modest, and are comparable to those of mild steels. Austenitic stainless steels cannot be hardened by heat treatment, but they do work harden rapidly. Therefore, high strength is possible in cold-worked forms, especially in drawn wire, in which a tensile strength of 1200 MPa (175 ksi) or higher is achievable. The heat of welding will soften the HAZ of heavily cold-worked austenitic stainless steel.

TABLE 3(A) COMPOSITIONS OF NOMINALLY AUSTENITIC STAINLESS STEELS

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	COMPOSITION, WT%									
			C	**Mn**	**P**	**S**	**Si**	**Cr**	**Ni**	**Mo**	**N**	OTHER
. . .	S01815	A 167	0.018 MAX	2.0 MAX	0.02 MAX	0.02 MAX	3.7-4.3	17.0-18.5	14.0-15.5	0.20 MAX	. . .	**. . .**
201	S20100	A 240	0.15 MAX	5.50-7.50	0.060 MAX	0.030 MAX	1.00 MAX	16.00-18.00	3.50-5.50	. . .	0.25 MAX	**. . .**
201	S20100	A 666	0.15 MAX	5.50-7.50	0.060 MAX	0.030 MAX	0.75 MAX	16.00-18.00	3.50-5.50	. . .	0.25 MAX	**. . .**
. . .	S20161	A 479	0.15 MAX	4.00-6.00	0.040 MAX	0.040 MAX	3.00-4.00	15.00-18.00	4.00-6.00	. . .	0.08-0.20	**. . .**
202	S20200	A 240	0.15 MAX	7.50-10.0	0.060 MAX	0.030 MAX	1.00 MAX	17.00-19.00	4.00-6.00	. . .	0.25 MAX	**. . .**
202	S20200	A 666	0.15 MAX	7.50-10.0	0.060 MAX	0.030 MAX	0.75 MAX	17.00-19.00	4.00-6.00	. . .	0.25 MAX	**. . .**
XM-1	S20300	A 582	0.08 MAX	5.00-6.50	0.04 MAX	0.18-0.35	1.00 MAX	16.00-18.00	5.00-6.50	. . .	. . .	**CU, 1.75-2.25**
205	S20500	A 666	0.12-0.25	14.00-15.00	0.060 MAX	0.030 MAX	0.75 MAX	16.50-18.00	1.00-1.75	. . .	0.32-0.40	**. . .**
XM-19	S20910	A 240	0.06 MAX	4.00-6.00	0.040 MAX	0.030 MAX	0.75 MAX	20.50-23.50	11.50-13.50	1.50-3.00	0.20-0.40	**NB + TA, 0.10-0.30;V, 0.10-0.30**
XM-31	S21400	A 240	0.12 MAX	14.00-16.00	0.045 MAX	0.030 MAX	0.30-1.00	17.00-18.50	1.00 MAX	. . .	0.35 MIN	**. . .**
XM-14	S21460	A 666	0.12 MAX	14.00-16.00	0.060 MAX	0.030 MAX	0.75 MAX	17.00-19.00	5.00-6.00	. . .	0.35-0.50	**. . .**
XM-17	S21600	A 240	0.08 MAX	7.50-9.00	0.045 MAX	0.030 MAX	0.75 MAX	17.50-22.00	5.00-7.00	2.00-3.00	0.25-0.50	**. . .**

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	COMPOSITION, WT%									
			C	**Mn**	**P**	**S**	**Si**	**Cr**	**Ni**	**Mo**	**N**	OTHER
XM-18	S21603	A 240	0.03 MAX	7.50-9.00	0.045 MAX	0.030 MAX	0.75 MAX	17.50-22.00	5.00-7.00	2.00-3.00	0.25-0.50	**. . .**
. . .	S21800	A 240	0.10 MAX	7.00-9.00	0.060 MAX	0.030 MAX	3.50-4.50	16.00-18.00	8.00-9.00	. . .	0.08-0.18	**. . .**
XM-10	S21900	A 276	0.08 MAX	8.00-10.00	0.060 MAX	0.030 MAX	1.00 MAX	19.00-21.00	5.50-7.50	. . .	0.15-0.40	**. . .**
XM-11	S21904	A 666	0.04 MAX	8.00-10.00	0.060 MAX	0.030 MAX	0.75 MAX	19.00-21.50	5.50-7.50	. . .	0.15-0.40	**. . .**
XM-29	S24000	A 240	0.08 MAX	11.50-14.50	0.045 MAX	0.030 MAX	0.75 MAX	17.00-19.00	2.25-3.75	. . .	0.20-0.40	**. . .**
XM-28	S24100	A 276	0.15 MAX	11.00-14.00	0.060 MAX	0.030 MAX	1.00 MAX	16.50-19.00	0.50-2.50	. . .	0.20-0.45	**. . .**
. . .	S28200	A 276	0.15 MAX	17.00-19.00	0.045 MAX	0.030 MAX	1.00 MAX	17.00-19.00	. . .	0.75 1.25	0.40-0.60	**CU, 0.75-1.25**
301	S30100	A 666	0.15 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	16.00-18.00	6.00-8.00	. . .	. . .	**. . .**
302	S30200	A 240	0.15 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	17.00-19.00	8.00-10.00	. . .	0.10 MAX	**. . .**
302	S30200	666	0.15 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	17.00-19.00	8.00-10.00	. . .	. . .	**. . .**
302B	S30215	A 167	0.15 MAX	2.00 MAX	0.045 MAX	0.030 MAX	2.00-3.00	17.00-19.00	8.00-10.00	. . .	0.10 MAX	**. . .**
303	S30300	A 473	0.15 MAX	2.00 MAX	0.20 MAX	0.15 MIN	1.00 MAX	17.00-19.00	8.00-10.00	0.60 MAX	. . .	**. . .**
XM-5	S30310	A 582	0.15 MAX	2.50-4.50	0.20 MAX	0.25 MIN	1.00 MAX	17.00-19.0	7.00-10.0	. . .	. . .	**. . .**

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	COMPOSITION, WT%									
			C	**Mn**	**P**	**S**	**Si**	**Cr**	**Ni**	**Mo**	**N**	OTHER
								0	0			
303SE	S30323	A 473	0.15 MAX	2.00 MAX	0.20 MAX	0.06 MAX	1.00 MAX	17.00-19.00	8.00-10.00	. . .	. . .	**SE, 0.15 MIN**
XM-2	S30345	A 582	0.15 MAX	2.00 MAX	0.05 MAX	0.11-0.16	1.00 MAX	17.00-19.00	8.00-10.00	0.40-0.60	. . .	**. . .**
XM-3	S30360	A 582	0.15 MAX	2.00 MAX	0.04 MAX	0.12-0.15	1.00 MAX	17.00-19.00	8.00-10.00	. . .	. . .	**PB, 0.12-0.30**
304	S30400	A 240, A 666	0.08 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	18.00-20.00	8.00-10.50	. . .	0.10 MAX	**. . .**
TP304	. . .	A 376	0.08 MAX	2.00 MAX	0.040 MAX	0.030 MAX	0.75 MAX	18.0-20.0	8.00-11.0	. . .	. . .	**. . .**
304L	S30403	A 240, A 666	0.030 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	18.00-20.00	8.00-12.00	. . .	0.10 MAX	**. . .**
304H	S30409	A 240	0.04-0.10	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	18.00-20.00	8.00-10.50	. . .	. . .	**. . .**
TP304H	. . .	A 376	0.04-0.10	2.00 MAX	0.040 MAX	0.030 MAX	0.75 MAX	18.0-20.0	8.00-11.0	. . .	. . .	**. . .**
. . .	S30415	A 240	0.04-0.06	0.30 MAX	0.045 MAX	0.030 MAX	1.00-2.00	18.00-19.00	9.00-10.00	. . .	0.12-0.18	**CE, 0.03 TO 0.08**
XM-7	S30430	A 276	0.10 MAX	2.00 MAX	0.045 MAX	0.030 MAX	1.00 MAX	17.00-19.00	8.00-10.00	. . .	. . .	**CU, 3.00-4.00**
304N	S30451	A 240	0.08 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	18.00-20.00	8.00-10.50	. . .	0.10-0.16	**. . .**
304N	. . .	A 376	0.08 MAX	2.00 MAX	0.040 MAX	0.030 MAX	0.57 MAX	18.0-20.0	8.00-11.0	. . .	. . .	**. . .**
XM-21	S30452	A 240	0.08 MA	2.00 MA	0.045	0.030	0.75 MA	18.00-	8.00-	. . .	0.16-	**. . .**

DESIGNATION	UNS NO.	ASTM SPECIFICATION	COMPOSITION, WT%									
			C	Mn	P	S	Si	Cr	Ni	Mo	N	OTHER
			X	X	MAX	MAX	X	20.00	10.50		0.30	
304LN	S30453	A 240	0.030 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	18.00-20.00	8.00-12.00	. . .	0.10-0.16	**. . .**
304LN	S30453	A 666	0.030 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	18.00-20.00	8.00-10.50	. . .	0.10-0.15	**. . .**
TP304LN	. . .	A 376	0.035 MAX	2.00 MAX	0.040 MAX	0.030 MAX	0.75 MAX	18.0-20.0	8.00-11.0	. . .	0.10-0.16	**. . .**
. . .	S30454	A 276	0.03 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	18.00-20.00	8.00-10.50	. . .	0.16-0.30	**. . .**
305	S30500	A 240	0.12 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	17.00-19.00	10.50-13.00	. . .	. . .	**. . .**
306	S30600	A 240	0.018 MAX	2.00 MAX	0.020 MAX	0.020 MAX	3.7-4.3	17.0-18.5	14.0-15.5	0.20 MAX	. . .	**. . .**
RA85H	S30615	A 240	0.16-0.24	2.00 MAX	0.030 MAX	0.030 MAX	3.20-4.00	17.00-19.00	13.50-16.00	. . .	. . .	**AL, 0.80-1.50**
308	S30800	A 167	0.08 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	19.00-21.00	10.00-12.00	. . .	. . .	**. . .**
. . .	S30815	A 167, A 240	0.05-0.10	0.80 MAX	0.040 MAX	0.030 MAX	1.40-2.00	20.00-22.00	10.00-12.00	. . .	0.14-0.20	**CE, 0.03 TO 0.08**
309	S30900	A 167	0.20 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	22.00-24.00	12.00-15.00	. . .	. . .	**. . .**
309S	S30908	A 240	0.08 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	22.00-24.00	2.00-15.00	. . .	. . .	**. . .**
309H	S30909	A 240	0.04-0.10	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	22.00-24.00	12.00-15.00	. . .	. . .	**. . .**
309CB	S309	A 240	0.08	2.00	0.04	0.03	0.75	22.0	12.0	. . .	. . .	**NB + TA, 10 ×**

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	COMPOSITION, WT%									
			C	**Mn**	**P**	**S**	**Si**	**Cr**	**Ni**	**Mo**	**N**	OTHER
	40		MAX	MAX	5 MAX	0 MAX	MAX	0-24.00	0-16.00			**C TO 1.10**
309HCB	S30949	A 240	0.04-0.10	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	22.00-24.00	12.00-16.00	. . .	. . .	**NB + TA, 10 × C TO 1.10**
310	S31000	167	0.25 MAX	2.00 MAX	0.045 MAX	0.030 MAX	1.50 MAX	24.00-26.00	19.00-22.00	. . .	. . .	**. . .**
310S	S31008	A 240	0.08 MAX	2.00 MAX	0.045 MAX	0.030 MAX	1.50 MAX	24.00-26.00	19.00-22.00	. . .	. . .	**. . .**
310H	S31009	A 240	0.04-0.10	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	24.00-26.00	19.00-22.00	. . .	. . .	**. . .**
310CB	S31040	A 240	0.08 MAX	2.00 MAX	0.045 MAX	0.030 MAX	1.50 MAX	24.00-26.00	19.00-22.00	. . .	. . .	**NB + TA, 10 × C TO 1.10**
310HCB	S31049	A 240	0.04-0.10	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	24.00-26.00	19.00-22.00	. . .	. . .	**NB + TA, 10 × C TO 1.10**
310MOLN	S31050	A 240	0.030 MAX	2.00 MAX	0.030 MAX	0.010 MAX	0.50 MAX	24.00-26.00	21.00-23.00	2.00-3.33	0.10-0.16	**. . .**
254SMO	S31254	A 240	0.020 MAX	1.00 MAX	0.030 MAX	0.010 MAX	0.80 MAX	19.50-20.50	17.50-18.50	6.00-6.50	0.18-0.22	**CU, 0.50-1.00**
314	S31400	A 276	0.25 MAX	2.00 MAX	0.045 MAX	0.030 MAX	1.50-3.00	23.00-26.00	19.00-22.00	. . .	. . .	**. . .**
316	S31600	A 240	0.08 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	16.00-18.00	10.00-14.00	2.00-3.00	0.10 MAX	**. . .**
316	S31600	A 666	0.08 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	16.00-18.00	10.00-14.00	2.00-3.00	. . .	**. . .**
TP316	. . .	A 376	0.08 MAX	2.00 MAX	0.040 MAX	0.030 MAX	0.75 MAX	16.0-18.0	11.0-14.0	2.00-3.00	. . .	**. . .**

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	COMPOSITION, WT%									
			C	**Mn**	**P**	**S**	**Si**	**Cr**	**Ni**	**Mo**	**N**	OTHER
316L	S31603	A 240	0.030 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	16.00-18.00	10.00-14.00	2.00-3.00	0.10 MAX	**. . .**
316L	S31603	A 666	0.030 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	16.00-18.00	10.00-14.00	2.00-3.00	. . .	**. . .**
316H	S31609	A 240	0.04-0.10	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	16.00-18.00	10.00-14.00	2.00-3.00	. . .	**. . .**
TP316H	. . .	A 376	0.04-0.10	2.00 MAX	0.040 MAX	0.030 MAX	0.75 MAX	16.0-18.0	11.0-14.0	2.00-3.00	. . .	**. . .**
316TI	S31635	A 240	0.08 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	16.00-18.00	10.00-14.00	2.00-3.00	. . .	**TI, 5 × (C + N) TO 0.70**
316CB	S31640	A 240	0.08 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	16.00-18.00	10.00-14.00	2.00-3.00	0.10 MAX	**NB + TA, 10 × C TO 1.10**
316N	S31651	A 240	0.08 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	16.00-18.00	10.00-14.00	2.00-3.00	0.10-0.16	**. . .**
316N	S31651	A 666	0.08 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	16.00-18.00	10.00-14.00	2.00-3.00	0.10-0.16	**. . .**
316N	. . .	A 376	0.08 MAX	2.00 MAX	0.040 MAX	0.030 MAX	0.75 MAX	16.0-18.0	11.0-14.0	2.00-3.00	0.10-0.16	**. . .**
316LN	S31635	A 240	0.030 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	16.00-18.00	10.00-14.00	2.00-3.00	0.10-0.16	**. . .**
T316LN		A 376	0.035 MAX	2.00 MAX	0.040 MAX	0.030 MAX	0.75 MAX	16.0-18.0	11.0-14.0	2.00-3.00	0.10-0.16	**. . .**
	S31654	A 276	0.03 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	16.00-18.00	10.00-14.00	2.00-3.00	0.16-0.30	**. . .**
317	S31700	A 240	0.08 MAX	2.00 MAX	0.045 MA	0.030 MA	0.75 MAX	18.00-20.0	11.00-15.0	3.00-4.00	0.10 MAX	**. . .**

DESIGNATION	UNS NO.	ASTM SPECIFICATION	COMPOSITION, WT%									
			C	Mn	P	S	Si	Cr	Ni	Mo	N	OTHER
					X	X		0	0			
317L	S31703	A 240	0.030 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	18.00-20.00	11.00-15.00	3.00-4.00	0.10 MAX	**. . .**
317LM	S31725	A 240	0.030 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	18.00-20.00	13.50-17.50	4.0-5.0	0.10 MAX	**. . .**
317LM	S31725	A 376	0.03 MAX	2.00 MAX	0.040 MAX	0.030 MAX	0.75 MAX	18.0-20.0	13.5-17.5	4.0-5.0	0.10 MAX	**CU, 0.75 MAX**
317LMN	S31726	A 240	0.030 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	18.00-20.00	13.50-17.50	4.0-5.0	0.10-0.20	**. . .**
317LMN	S31726	A 376	0.03 MAX	2.00 MAX	0.040 MAX	0.030 MAX	0.75 MAX	17.0-20.0	13.5-17.5	4.0-5.0	0.10-0.20	**CU, 0.75 MAX**
317LN	S31753	A 240	0.030 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	18.00-20.00	11.00-15.00	3.00-4.00	0.10-0.22	**. . .**
321	S32100	A 240	0.08 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	17.00-19.00	9.00-12.00	. . .	0.10 MAX	**TI, 5 × (C + N) TO 0.70**
TP321		A 376	0.08 MAX	2.00 MAX	0.040 MAX	0.030 MAX	0.75 MAX	17.0-20.0	9.00-13.0	. . .	. . .	**TI, 5 × C TO 0.60**
321H	S32109	A 240	0.04-0.10	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	17.00-19.00	9.00-12.00	. . .	0.10 MAX	**TI, 4 × (C + N) TO 0.70**
TP321H	. . .	A 376	0.04-0.10	2.00 MAX	0.040 MAX	0.030 MAX	0.75 MAX	17.0-20.0	9.00-13.0	. . .	. . .	**TI, 4 × C TO 0.60**
. . .	S32615	A 240	0.07 MAX	2.00 MAX	0.045 MAX	0.030 MAX	4.8-6.0	16.5-19.5	19.0-22.0	0.30-1.5	. . .	**CU, 1.5-2.5**
347	S34700	A 240	0.08 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	17.00-19.00	9.00-13.00	. . .	. . .	**NB + TA, 10 × C TO 1.00**
TP347	. . .	A 376	0.08 MA	2.00 MA	0.040	0.030	0.75 MA	17.0-	9.00-	. . .	. . .	**NB + TA, 10 × C TO 1.00**

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	COMPOSITION, WT%									
			C	**Mn**	**P**	**S**	**Si**	**Cr**	**Ni**	**Mo**	**N**	OTHER
			X	X	MAX	MAX	X	20.0	13.0			
347H	S34709	A 240	0.04-0.10	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	17.00-19.00	9.00-13.00	. . .	. . .	**NB + TA, 8 × C TO 1.00**
TP347H	. . .	A 376	0.04-0.10	2.00 MAX	0.040 MAX	0.030 MAX	0.75 MAX	17.0-20.0	9.00-13.0	. . .	. . .	**NB + TA, 8 × C TO 1.00**
348	S34800	A240	0.08 MAX	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	17.00-19.00	9.00-13.00	. . .	. . .	**NB + TA, 10 × C TO 1.00; CO, 0.20 MAX; TA, 0.10 MAX**
TP348	. . .	A 376	0.08 MAX	2.00 MAX	0.040 MAX	0.030 MAX	0.75 MAX	17.0-20.0	9.00-13.0	. . .	. . .	**NB + TA, 10 × C TO 1.00**
348H	S34809	A 240	0.04-0.10	2.00 MAX	0.045 MAX	0.030 MAX	0.75 MAX	17.00-19.00	9.00-13.00	. . .	. . .	**NB + TA, 8 × C TO 1.00; CO, 0.20 MAX; TA, 0.10 MAX**
16-8-2H	. . .	A 376	0.05-0.10	2.00 MAX	0.040 MAX	0.030 MAX	0.75 MAX	14.5-16.5	7.50-9.50	1.5 2.0	. . .	**. . .**
XM-15	S38100	A 240	0.08 MAX	2.00 MAX	0.030 MAX	0.030 MAX	1.50-2.50	17.00-19.00	17.50-18.50	. . .	. . .	**. . .**
320	N08020	B 463	0.07 MAX	2.00 MAX	0.045 MAX	0.035 MAX	1.00 MAX	19.00-21.00	32.00-38.00	2.00-3.00	. . .	**NB + TA, 8 × C TO 1.00; CU, 3.00-4.00**
SANICRO 28	N08028	B 709	0.030 MAX	2.50 MAX	0.030 MAX	0.030 MAX	1.00 MAX	26.0-38.0	30.0-34.0	3.0-4.0	. . .	**CU, 0.6-1.4**
330	N08330	B 536	0.10 MAX	2.00 MAX	0.03 MAX	0.03 MAX	0.75-1.50	17.0-20.0	34.0-37.0	. . .	. . .	**CU, 1.00 MAX**
AL-6XN	N08367	B 688	0.030 MAX	2.00 MAX	0.040 MAX	0.030 MAX	1.00 MAX	20.00-22.00	23.50-25.50	6.00-7.00	. . .	**CU, 0.75 MAX**
904L	N08904	B 625	0.020 MAX	2.00 MAX	0.045 MAX	0.035 MAX	1.00 MAX	19.00-2300	23.00-28.00	4.0-5.0	. . .	**CU, 1.0-2.0**

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	COMPOSITION, WT%									
			C	**Mn**	**P**	**S**	**Si**	**Cr**	**Ni**	**Mo**	**N**	OTHER
25-6MO	N08925	B 625	0.020 MAX	1.00 MAX	0.045 MAX	0.030 MAX	0.50 MAX	19.00-21.00	24.00-26.00	6.0-7.0	0.10-0.20	**CU, 0.8-1.5**
CF-8	. . .	A 743	0.08 MAX	1.50 MAX	0.04 MAX	0.04 MAX	2.0 MAX	18.0-21.0	8.0-11.0	. . .	. . .	**. . .**
CG-12	. . .	A 743	0.12 MAX	1.50 MAX	0.04 MAX	0.04 MAX	2.00 MAX	20.0-23.0	10.0-13.0	. . .	. . .	**. . .**
CF-20	. . .	A 743	0.20 MAX	1.50 MAX	0.04 MAX	0.04 MAX	2.00 MAX	18.0-21.0	0.8-11.0	. . .	. . .	**. . .**
CF-8M	. . .	A 743, A 744	0.08 MAX	1.50 MAX	0.04 MAX	0.04 MAX	2.00 MAX	18.0-21.0	9.0-12.0	2.0-3.0	. . .	**. . .**
CF-8C	. . .	A 743, A 744	0.08 MAX	1.50 MAX	0.04 MAX	0.04 MAX	2.00 MAX	18.0-21.0	9.0-12.0	. . .	. . .	**NB + TA = 8 × C TO 1.0 (9 × C TO 1.1 AT HI TA)**
CF-16F	. . .	A 743	0.16 MAX	1.50 MAX	0.04 MAX	0.04 MAX	2.00 MAX	18.0-21.0	9.0-12.0		. . .	**S, MO, P, SE ADDED ABOVE LIMITS GIVEN FOR MACHINABILITY**
CH-20	. . .	A 743	0.20 MAX	1.50 MAX	0.04 MAX	0.04 MAX	2.00 MAX	22.0-26.0	12.0-15.0	. . .	. . .	**. . .**
CK-20	. . .	A 743	020 MAX	2.00 MAX	0.04 MAX	0.04 MAX	2.00 MAX	23.0-27.0	19.0-22.0	. . .	. . .	**. . .**
CE-30	. . .	A 743	0.30 MAX	1.50 MAX	0.04 MAX	0.04 MAX	2.00 MAX	26.0-30.0	8.0-11.0	. . .	. . .	**. . .**
CF-3	. . .	A 743, A 744	0.03 MAX	1.50 MAX	0.04 MAX	0.04 MAX	2.0 MAX	17.0-21.0	8.0-11.0	. . .	. . .	**. . .**
CF10SMNN	. . .	A 743	0.10 MAX	7.00-9.00	0.060 MAX	0.030 MAX	3.50-4.50	16.0-18.0	8.0-9.0	. . .	0.08-0.18	**. . .**
CF-3M	. . .	A 743, A 744	0.03 MAX	1.50 MAX	0.04 MAX	0.04 MAX	1.50 MAX	17.0-21.0	9.0-13.0	2.0-3.0	. . .	**. . .**
CF-3MN	. . .	A 743	0.03 MAX	1.50 MAX	0.040 MAX	0.040 MAX	1.50 MAX	17.0-22.0	9.0-13.0	2.0-3.0	0.10-0.20	**. . .**
CG6MM	. . .	A 743	0.06	4.00	0.04	0.03	1.00	20.5	11.5	1.50	0.20	**NB + TA, 0.10-**

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	COMPOSITION, WT%									
			C	**Mn**	**P**	**S**	**Si**	**Cr**	**Ni**	**Mo**	**N**	OTHER
N			MAX	-6.00	MAX	MAX	MAX	-23.5	-13.5	-3.00	-0.40	**0.30; V, 0.10-0.30**
CG-8M	. . .	A 743, A 744	0.08 MAX	1.50 MAX	0.04 MAX	0.04 MAX	1.50 MAX	18.0-21.0	9.0-13.0	3.0-4.0	. . .	**. . .**
CN-3M	. . .	A 743	0.03 MAX	2.0 MAX	0.03 MAX	0.03 MAX	1.0 MAX	20.0-22.0	23.0-27.0	4.5-5.5	. . .	**. . .**
CN-7M	. . .	A 743, A 744	0.07 MAX	1.50 MAX	0.04 MAX	0.04 MAX	1.50 MAX	19.0-22.0	27.5-30.5	2.0-3.0	. . .	**CU, 3.0-4.0**
CN-7MS	. . .	A 743, A 744	0.07 MAX	1.00 MAX	0.04 MAX	0.03 MAX	2.50-3.50	18.0-20.0	22.0-25.0	2.5-3.0	. . .	**CU, 1.5-2.0**
CK-3MCUN	**. . .**	**A 743, A 744**	**0.025 MAX**	**1.20 MAX**	**0.045 MAX**	**0.010 MAX**	**1.00 MAX**	**19.5-20.5**	**17.5-19.5**	**6.0-7.0**	**0.180-0.240**	**CU, 0.50-1.00**

TABLE 3(B) PROPERTIES AND FILLER METALS FOR NOMINALLY AUSTENITIC STAINLESS STEELS

DESIGNATION	UNS NO.	ASTM SPECIFICATION	COMMENTS	TENSILE STRENGTH		YIELD STRENGTH		ELONGATION, %	MATCHING FILLER METALS		
				MPA	KSI	MPA	KSI		A5.4	A5.9(A)	A5.22
. . .	S01815	A 167	. . .	540	78	240	35	40	. . .	. . .	**. . .**
201	S20100	A 240	201-1, BOTH SPECIFICATIONS	655	95	260	38	40	E240-XX	ER240	**. . .**
201	S20100	A 666	201-2, BOTH SPECIFICATIONS	655	95	310	45	40	E240-XX	ER240	**. . .**
. . .	S20161	A 479	. . .	860	125	345	50	40	E240-XX	ER240	**. . .**
202	S20200	A 240	. . .	620	90	260	38	40	E240-XX	ER240	**. . .**
202	S20200	A 666	. . .	620	90	260	38	40	E240-XX	ER240	**. . .**
XM-1	S20300	A 582	. . .	NS	NS	NS	NS	NCW	NCW	NCW	**. . .**
205	S20500	A 666	. . .	790	115	450	65	40	E240-XX	ER240	**. . .**
XM-19	S20910	A 240	NITRONIC 50	690	100	380	55	35	E209-XX	ER209	**. . .**
XM-31	S21400	A 240	SHEET HIGHER STRENGTH	720	105	380	55	40	E240-XX	ER240	**. . .**
XM-14	S21460	A 666	. . .	720	105	380	55	40	E240-XX	ER240	**. . .**
XM-17	S21600	A 240	SHEET HIGHER STRENGTH	620	90	345	50	40	E209-XX	ER209	**. . .**
XM-18	S21603	A 240	SHEET HIGHER STRENGTH	620	90	345	50	40	E209-XX	ER209	**. . .**
. . .	S21800	A 240	NITRONIC 60	655	95	345	50	35	E240-XX	ER218	**. . .**
XM-10	S21900	A 276	NITRONIC 40	620	90	345	50	45	E219-XX	ER219	**. . .**
XM-11	S21904	A 666	SHEET HIGHER STRENGTH	620	90	345	50	45	E219-XX	ER219	**. . .**
XM-29	S24000	A 240	NITRONIC 33	690	100	380	55	40	E240-XX	ER240	**. . .**
XM-28	S24100	A 276	NITRONIC 32	690	100	380	55	30	. . .	. . .	**. . .**
. . .	S28200	A 276	. . .	760	110	415	60	35	. . .	. . .	**. . .**
301	S30100	A 666	LOWER STRENGTH IN A 167	620	90	205	30	40	. . .	. . .	**. . .**
302	S30200	A 240	. . .	515	75	205	30	40	E308-XX	ER308	**E308T-X**
302	S30200	666	. . .	515	75	205	30	40	E308-XX	ER308	**E308T-X**
302B	S30215	A 167	. . .	515	75	205	30	40	. . .	. . .	**. . .**
303	S30300	A 473	. . .	515	75	205	30	40	NCW	NCW	**NCW**
XM-5	S30310	A 582	. . .	NS	NS	NS	NS	NS	NCW	NCW	**NCW**

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	**COMMENTS**	**TENSILE STRENGTH**		**YIELD STRENGTH**		**ELONGATION, %**	MATCHING FILLER METALS		
				MPA	**KSI**	**MPA**	**KSI**		**A5.4**	**A5.9(A)**	A5.22
303SE	S30323	A 473	. . .	515	75	205	30	40	NCW	NCW	**NCW**
XM-2	S30345	A 582	. . .	NS	NS	NS	NS	NS	NCW	NCW	**NCW**
XM-3	S30360	A 582	. . .	NS	NS	NS	NS	NS	NCW	NCW	**NCW**
304	S30400	A 240, A 666	. . .	515	75	205	30	40	E308-XX	ER308	**E308T-X**
TP304	. . .	A 376	. . .	515	75	205	30	35	E308-XX	ER308	**E308T-X**
304L	S30403	A 240, A 666	. . .	480	70	170	25	40	E308L-XX	ER308L	**E308LT-X**
304H	S30409	A 240	. . .	515	75	205	30	40	E308H-XX	ER308H	**E308T-X**
TP304H	. . .	A 376	. . .	515	75	205	30	35	E308H-XX	ER308H	**E308-X**
. . .	S30415	A 240	. . .	600	87	290	42	40	. . .	. . .	**. . .**
XM-7	S30430	A 276	. . .	480	70	170	25	40	. . .	. . .	**. . .**
304N	S30451	A 240	. . .	550	80	240	35	30	. . .	. . .	**. . .**
304N	. . .	A 376	. . .	550	80	240	35	35	. . .	. . .	**. . .**
XM-21	S30452	A 240	SHEET HIGHER STRENGTH	585	85	275	40	30	. . .	. . .	**. . .**
304LN	S30453	A 240	. . .	515	75	205	30	40	. . .	. . .	**. . .**
304LN	S30453	A 666	. . .	550	80	240	35	40	. . .	. . .	**. . .**
TP304LN	. . .	A 376	. . .	515	75	205	30	35	. . .	. . .	**. . .**
. . .	S30454	A 276	. . .	620	90	245	50	30	. . .	. . .	**. . .**
305	S30500	A 240	LOWER STRENGTH IN A 167	515	75	205	30	40	E308-XX	ER308	**E308T-X**
306	S30600	A 240	. . .	540	78	240	35	40	. . .	. . .	**. . .**
RA85H	S30615	A 240	. . .	620	90	275	40	35	. . .	. . .	**. . .**
308	S30800	A 167	. . .	515	75	205	30	40	E308-XX	ER308	**E308T-X**
. . .	S30815	A 167, A 240	. . .	600	87	310	45	40	. . .	. . .	**. . .**
309	S30900	A 167	. . .	515	75	205	30	40	E309-XX	ER309	**E309T-X**
309S	S30908	A 240	. . .	515	75	205	30	40	E309-XX	ER309	**E309T-X**
309H	S30909	A 240	. . .	515	75	205	30	40	E309-XX	ER309	**E309T-X**
309CB	S30940	A 240	. . .	515	75	205	30	40	E309CB-XX	. . .	**. . .**
309HCB	S30949	A 240	. . .	515	75	205	30	40	E309CB-XX	. . .	**. . .**
310	S31000	.167	. . .	515	75	205	30	40	E310-XX	ER310	**E310T-X**
310S	S31008	A 240	. . .	515	75	205	30	40	E310-XX	ER310	**E310T-X**

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	**COMMENTS**	**TENSILE STRENGTH**		**YIELD STRENGTH**		**ELONGATION, %**	MATCHING FILLER METALS		
				MPA	**KSI**	**MPA**	**KSI**		**A5.4**	**A5.9$^{(A)}$**	A5.22
310H	S31009	A 240	. . .	515	75	205	30	40	E310-XX	ER310	**E310T-X**
310CB	S31040	A 240	. . .	515	75	205	30	40	E310CB-XX	. . .	**. . .**
310HCB	S31049	A 240	. . .	515	75	205	30	40	E310CB-XX	. . .	**. . .**
310MOLN	S31050	A 240	. . .	550	80	240	35	30	. . .	. . .	**. . .**
254SMO	S31254	A 240	. . .	650	94	305	44	35	. . .	. . .	**. . .**
314	S31400	A 276	. . .	515	75	205	30	40	. . .	. . .	**. . .**
316	S31600	A 240	. . .	515	75	205	30	40	E316-XX	ER316	**E316T-X**
316	S31600	A 666	. . .	515	75	205	30	40	E316-XX	ER316	**E316T-X**
TP316	. . .	A 376	. . .	515	75	205	30	35	E316-XX	ER316	**E316T-X**
316L	S31603	A 240	. . .	480	70	170	25	40	E316L-XX	ER316L	**E316LT-X**
316L	S31603	A 666	. . .	480	70	170	25	40	E316L-XX	ER316L	**E316LT-X**
316H	S31609	A 240	. . .	515	75	205	30	40	E316H-XX	ER316H	**E316T-X**
TP316H	. . .	A 376	. . .	515	75	205	30	35	E316H-XX	ER316H	**E316T-X**
316TI	S31635	A 240	. . .	515	75	205	30	40	E318-XX	ER318	**. . .**
316CB	S31640	A 240	. . .	515	75	205	30	40	E318-XX	ER318	**. . .**
316N	S31651	A 240	. . .	550	80	240	35	35	. . .	. . .	**. . .**
316N	S31651	A 666	. . .	550	80	240	35	40	. . .	. . .	**. . .**
316N	. . .	A 376	. . .	550	80	240	35	35	. . .	. . .	**. . .**
316LN	S31653	A 240	. . .	515	75	205	30	40	. . .	. . .	**. . .**
TP316LN	. . .	A 376	. . .	515	75	205	30	35	. . .	. . .	**. . .**
. . .	S31654	A 276	. . .	620	90	345	50	30	. . .	. . .	**. . .**
317	S31700	A 240	. . .	515	75	205	30	35	E317-XX	ER317	**E317LT-X**
317L	S31703	A 240	. . .	515	75	205	30	40	E317L-XX	ER317L	**E317LT-X**
317LM	S31725	A 240	. . .	515	75	205	30	40	. . .	. . .	**. . .**
317LM	S31725	A 376	. . .	515	75	205	30	35	. . .	. . .	**. . .**
E17LMN	S31726	A 240	. . .	550	80	240	35	40	. . .	. . .	**. . .**
317LMN	S31726	A 376	. . .	550	80	240	35	35	. . .	. . .	**. . .**
317LN	S31753	A 240	. . .	550	80	240	35	40	. . .	. . .	**. . .**

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	**COMMENTS**	**TENSILE STRENGTH**		**YIELD STRENGTH**		**ELONGATION, %**	MATCHING FILLER METALS		
				MPA	**KSI**	**MPA**	**KSI**		**A5.4**	**A5.9**$^{(A)}$	A5.22
321	S32100	A 240	. . .	515	75	205	30	40	E347-XX	ER321	**E347T-X**
TP321	. . .	A 376	. . .	515	75	205	30	35	E347-XX	ER321	**E347T-X**
321H	S32109	A 240	. . .	515	75	205	30	40	E347-XX	ER321	**E347T-X**
TP321H	. . .	A 376	. . .	515	75	205	30	35	E347-XX	ER321	**E347T-X**
. . .	S32615	A 240	. . .	550	80	220	32	25	. . .	. . .	**. . .**
347	S34700	A 240	. . .	515	75	205	30	40	E347-XX	ER347	**E347T-X**
TP347	. . .	A 376	. . .	515	75	205	30	35	E347-XX	ER347	**E347T-X**
347H	S34709	A 240	. . .	515	75	205	30	40	E347-XX	ER347	**E347T-X**
TP347H	. . .	A 376	. . .	515	75	205	30	35	E347-XX	ER347	**E347T-X**
348	S34800	A 240	. . .	515	75	205	30	40	E347-XX	ER347	**E347T-X**
TP348	. . .	A 376	. . .	515	75	205	30	35	E347-XX	ER347	**E347T-X**
348H	S34809	A 240	. . .	515	75	205	30	40	E347-XX	ER347	**E347T-X**
16-8-2H	. . .	A 376	. . .	515	75	205	30	35	E16-8-2-XX	ER16-8-2	**. . .**
XM-15	S38100	A 240	. . .	515	75	205	30	40	. . .	. . .	**. . .**
320	N08020	B 463	. . .	550	80	240	35	30	E320-XX	ER320	**. . .**
SANICRO 28	N08028	B 709	. . .	505	73	215	31	40	E383-XX	ER383	**. . .**
330	N08330	B 536	. . .	480	70	205	30	30	E330-XX	ER330	**. . .**
AL-6XN	N08367	B 688	. . .	690	100	310	45	30	. . .	. . .	**. . .**
904L	N08904	B 625	. . .	490	71	215	31	35	E385-XX	ER385	**. . .**
25-6MO	N08925	B 625	. . .	600	87	295	43	40	. . .	. . .	**. . .**
CF-8	. . .	A 743	. . .	480	70	205	30	35	E308-XX	ER308	**E308T-X**
CG-12	. . .	A 743	. . .	480	70	195	28	35	E309-XX	ER309	**E309T-X**
CF-20	. . .	A 743	. . .	480	70	205	30	30	E308-XX	ER308	**E308T-X**
CF-8M	. . .	A 743, A 744	. . .	480	70	205	30	30	E316-XX	ER316	**E316T-X**
CF-8C	. . .	A 743, A 744	. . .	480	70	205	30	30	E347-XX	ER321	**E347T-X**
CF-16F	. . .	A 743	. . .	480	70	205	30	25	. . .	. . .	**. . .**
CH-20	. . .	A 743	. . .	480	70	205	30	30	E309-XX	ER309	**E309T-X**
CK-20	. . .	A 743	. . .	450	65	195	28	30	E310-XX	ER310	**E310T-X**
CE-30	. . .	A 743	. . .	550	80	275	40	10	. . .	. . .	**. . .**
CF-3	. . .	A 743, A 744	. . .	480	70	205	30	35	E308L-XX	ER308L	**E308LT-X**
CF10SMNN	. . .	A 743	. . .	585	85	290	42	30	. . .	. . .	**. . .**

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	**COMMENTS**	**TENSILE STRENGTH**		**YIELD STRENGTH**		**ELONGATION, %**	MATCHING FILLER METALS		
				MPA	**KSI**	**MPA**	**KSI**		**A5.4**	**A5.9**$^{(A)}$	A5.22
CF-3M	. . .	A 743 , A 744	. . .	480	70	205	30	30	E316L-XX	ER316L	**E316LT-X**
CF-3MN	. . .	A 743	. . .	515	75	255	37	35	. . .	. . .	**. . .**
CGMMN	. . .	A 743	. . .	585	85	290	42	30	. . .	. . .	**. . .**
CG-8M	. . .	A 743 , A 744	. . .	515	75	240	35	25	E317-XX	ER317	**E317LT-X**
CN-3M	. . .	A 743	. . .	435	63	170	25	30	E385-XX	. . .	**. . .**
CM-7M	. . .	A 743, A 744	. . .	425	62	170	25	35	E320-XX	ER320	**. . .**
CM-7MS	. . .	A 743, A 744	. . .	480	70	205	30	35	. . .	. . .	**. . .**
CK-3MCUN	**. . .**	**A 743, A 744**	**. . .**	**550**	**80**	**260**	**38**	**35**	**. . .**	**. . .**	**. . .**

NS, not specified; NCW, not considered weldable.

(A) ELECTRODES WITH HIGHER SILICON (0.65-1.00%), INDICATED BY "SI" ADDED AT THE END OF THE CLASSIFICATION WERE INCLUDED. METAL-CORED ELECTRODES, INDICATED BY A "C" IN PLACE OF THE "R" IN THE CLASSIFICATION, ARE ALSO INCLUDED.

Even the leanest austenitic stainless steels (such as types 302 and 304, which can be considered the base alloys of the austenitic stainless steels) offer general corrosion resistance in the atmosphere, in many aqueous media, in the presence of foods, and in oxidizing acids such as nitric acid. Types 321 and 347 are essentially type 304 with additions of either titanium or niobium, respectively, which stabilize carbides against sensitization.

The addition of molybdenum in type 316L provides pitting resistance in phosphoric and acetic acids and dilute chloride solutions, as well as corrosion resistance in sulfurous acid (H_2SO_3). An even higher molybdenum content, as in type 317L (3%), and even richer alloys further enhance pitting resistance.

Nitrogen is added to enhance strength at room temperature and, especially, at cryogenic temperatures (as in type 304LN, for example). Nitrogen is also added to reduce the rate of chromium carbide precipitation and, therefore, the susceptibility to sensitization. Nitrogen is also added to molybdenum-containing alloys to increase resistance to chloride-induced pitting and crevice corrosion.

Higher amounts of chromium and/or nickel are used to enhance high-temperature oxidation resistance (as in types 309, 310, and 330, for example). Copper and nickel can be added to enhance resistance to reducing acids, such as sulfuric acid (type 320). Nickel and molybdenum, when present in sufficient amounts, promote resistance to chloride-induced stress-corrosion cracking. As indicated by this discussion and Tables 3(a) and 3(b), there is an enormous number of austenitic stainless steel types.

Engineering for Use in the As-Welded Condition. Most weldments in nominally austenitic stainless steel base metals are put into service in the as-welded condition. In contrast to the situations with martensitic and ferritic stainless steel base metals, many of the nominally austenitic stainless steel base metals have matching, or near-matching, welding filler metals (see Tables 3(a) and 3(b)). In many cases, a nearly matching filler metal is a good choice.

Filler-metal selection and welding procedure development for austenitic stainless steel base metals primarily depend on whether or not ferrite is possible and acceptable in the weld metal. In general, if ferrite is both possible and acceptable, then the suitable choices of filler metals and procedures are broad. If the weld metal solidifies as primary ferrite, then the range of acceptable welding procedures can be as broad as it would be for ordinary mild steel. However, if ferrite in the weld metal is neither possible nor acceptable, then the filler metal and procedure choices are much more restricted, because of hot-cracking considerations.

Most filler-metal compositions are available in a variety of forms. When the presence of ferrite is possible, some filler-metal manufacturers will tailor it to meet specific needs, at least in filler metals that are composites (for SMAW, FCAW, and metal-cored wires for GMAW or SAW processes). For example, in type 308 or 308L filler metals, which are typically selected for joining base-metal types 302, 304, 304L, 305, CF-8, or CF-3, the filler metal can be designed for zero ferrite, more than 20 FN, or any FN value between these extremes, and still be within the composition limits for the classification in the relevant AWS filler-metal specification.

In solid wires of type 308 or 308L, which are used for GMAW, GTAW, or SAW processes, the potential ferrite content of all-weld metal is often designed by the steel producer to provide maximum yield from the heat of steel. Stainless steel producers have generally found that the yield from a heat of steel, which will be reduced to "green rod" for supply to a filler-metal producer, is maximized when the ferrite content ranges from 3 to 8 FN, based on the WRC-1988 diagram. Weld rod demand is usually for a minimum of 8 FN. Poor yields make the production of metal with more than 10 FN uneconomical and, thus, limit its availability. A lack of demand limits production and, thus, the availability of green rod with less than 8 FN. Therefore, most solid-wire compositions are designed to achieve approximately 8 to 10 FN. Solid wires with compositions that are much beyond this range (as in austenitic grades that generally contain some ferrite) usually require special order, heat quantities (of 18 Mg, or 40,000 lb), long lead times, and they command premium prices.

With composite filler metals, the manufacturer can adjust both composition and ferrite content via alloying in either the electrode coating of SMAW electrodes or the core of flux-cored and metal-cored wires. Tailoring composition and ferrite for special applications is much more easily accomplished than it is for solid wires. For most off-the-shelf composite filler metals, the filler-metal manufacturer has a target all-weld metal FN in mind when designing the filler metal. For a given alloy type, such as 308L, the target FN is likely to depend on the economics of filler-metal design, the appropriate Ferrite Number for the specific type of filler metal to avoid hot cracking, and market requirements.

For example, the ANSI/AWS A5.4 classification system recognizes three all-position coating types. In general, E308L-15 electrodes provide better hot-cracking resistance at low or no levels of ferrite than do E308L-16 electrodes, which, in turn, provide better hot-cracking resistance than do E308L-17 electrodes when little or no ferrite is present in the deposit. This difference in cracking resistance without ferrite primarily stems from bead shape. The E308L-15 electrodes tend to produce a convex bead shape. Convexity provides something of a riser, as in a casting, which tends to resist hot cracking. The E308L-16 electrodes tend to produce a nearly flat profile, which provides little or no riser. The E308L-17 electrodes tend to produce a concave profile (as well as a high silicon deposit), which makes this the least crack resistant coating type when there is little or no ferrite in the deposit. Further, the sensitivity to nitrogen pickup with a long arc (and the consequent loss of ferrite) is least with E308L-15 and greatest with E308L-17 electrodes.

As a result, filler-metal manufacturers generally aim their off-the-shelf E308L-15 electrodes toward a lower FN than they do their E308L-16 electrodes, which, in turn, are aimed toward lower FN values than are their E308L-17 electrodes. Of course, the same tendencies apply to other alloy compositions with the same coating types, which can be generalized as EXXX-15, EXXX-16, or EXXX-17, where "XXX" stands for 308, 308L, 309, 309L, 316, 316L, and so on.

Until 1992, the EXXX-17 coating type was not recognized by the ANSI/AWS A5.4 specification as being a distinct type, although the electrodes have been available for many years. Because electrodes that are properly classified as EXXX-17 generally weld satisfactorily on both direct current (dc) and alternating current (ac), they had to be classified as EXXX-16 electrodes until the publication of ANSI/AWS A5.4-92 introduced the EXXX-17 classifications. Using these electrodes and meeting the vertical-up fillet weld size and convexity requirements of the EXXX-16 classifications were difficult to achieve, at best.

Although the EXXX-17 classifications have relaxed vertical-up fillet size and convexity requirements, some electrode manufacturers have been reluctant to change the classification of certain brand names (originally classified as EXXX-16) in order to more correctly reflect the true vertical-up capabilities of the electrodes. The knowledgeable fabricator can detect an electrode that is classified by its manufacturer as EXXX-16, when it properly ought to be classified as EXXX-17, either by attempting to meet the vertical-up fillet requirements of the EXXX-16 classification or by examining the all-weld metal composition that the electrode produces. The EXXX-17 electrodes all produce deposits of low manganese and high silicon. In general, a nominally austenitic electrode with an all-weld metal silicon content that is nearly as high as, or higher than, its manganese content (that is, silicon content approaches the 0.9% maximum of the specification) is in all probability properly classified as an EXXX-17 electrode.

Flux-cored and metal-cored electrodes do not have the clear distinctions that exist among the three main coating types. In general, both off-the-shelf flux-cored and metal-cored electrodes tend to have higher ferrite numbers, more like those of EXXX-17 electrodes, because the higher heat input that normally accompanies the use of these products tends to make the weld metal less crack resistant when little or no ferrite is present.

If a fully austenitic stainless steel as-welded deposit, or one with almost no ferrite, is necessary, and solidification as primary austenite, or fully austenitic solidification, is expected, then the choices of both filler metals and procedures are more restricted. When designing filler metals with low or no ferrite, the producer tries to obtain raw materials with minimal sulfur and phosphorus contents. Generally, high-silicon compositions are avoided, because of their hot-cracking tendencies. Concave beads and high heat input are best avoided in procedure development. As a result, little or no filler metals are even offered commercially as EXXX-17 SMAW electrodes without ferrite. Submerged arc joining with fully austenitic filler metals is usually avoided, because of cracking problems, although submerged arc cladding is more successful. The GMAW process is often restricted to a short-circuiting transfer mode.

A general rule to follow in procedure development for fully austenitic weld metals, in order to avoid or minimize cracking problems, is to "weld ugly." In particular, this means that if cracking is encountered, then welding procedures that produce markedly convex beads can overcome it. For compositions that are very crack sensitive, it may be necessary to make individual weld beads that are so convex that part of the weld bead must be ground away before the next bead is deposited. For the most crack sensitive steels, such as type 320, crack-free welding with matching filler metal may be possible only with either small-diameter (1.6 through 2.4 mm, or 0.06 through 0.10 in.) E320-15 electrodes that are welded on the low side of the recommended current range or small-diameter ($\leq$0.9 mm, or 0.04 in.) GMAW wire in the short-circuiting transfer mode.

Niobium, phosphorus, and sulfur in such alloys generally tend to make hot-cracking problems more severe. As a result, for type 320 stainless steel, both SMAW electrodes (E320LR-15) and GMAW wires (ER320LR) are available with reduced residual elements, as well as low carbon, so that niobium can be minimized without loss of the stabilization of carbides.

Additional welding procedure approaches to eliminating hot cracking with weld metals that solidify either fully austenitic or as primary austenite include those that reduce heat buildup in the steel. Low welding heat input and low preheat and interpass temperatures are all beneficial.

Weld puddle shape has a strong influence on hot-cracking tendencies. Narrow, deep puddles crack more readily than wide, shallow puddles. In particular, it is helpful to adjust welding conditions to obtain a weld pool that is elliptical in surface shape and to avoid tear-drop-shaped weld pools, the sharp tail of which tends to promote hot cracking. Low current and slow travel speed favor an elliptical pool shape. Copper chill blocks also can help to extract heat and speed cooling.

With fully austenitic or primary austenite solidification, it is essential to backfill weld craters to avoid crater cracks, which are a form of localized hot cracking. This means stopping the travel of the arc while continuing to add filler metal, with a down-sloping current, if possible, before breaking the arc, so that the crater becomes convex, rather than concave.

It was noted earlier that the usual choice for a filler metal for austenitic stainless steel is a matching or near-matching composition to that of the base metal. However, one situation generally requires an exception. This situation occurs when a high-molybdenum base metal is chosen for pitting or crevice corrosion resistance in a chloride-containing aqueous solution. Although the high molybdenum content of the weld metal provides resistance to localized corrosion, it invariably has segregation of molybdenum on a microscopic scale. The center of each dendrite is leaner in molybdenum than are the interdendritic spaces. The center of a dendrite then becomes a preferential site for localized corrosion.

It has been found that this preferential localized corrosion in the weld metal can be overcome by using a filler metal that is higher in molybdenum than is the base metal. Thus, a 3% Mo base metal (such as type 317L) requires a filler metal of at least 4% Mo. A 4% Mo base metal (such as type 904L) requires a filler metal of at least 6% Mo. A 6% Mo base metal requires a filler metal of at least 8% Mo. However, stainless steel weld metal that contains more than 5% Mo often has poor mechanical properties. Therefore, a nickel-base alloy of at least 8% Mo is usually chosen for filler metal when joining stainless steels with contents of 4% Mo and higher. The most common filler metals selected are the ENiCrMo-3 or ENiCrMo-10 covered-electrode classifications in ANSI/AWS A5.11 and the ERNiCrMo-3 or ERNiCrMo-10 bare-wire classifications in the ANSI/AWS A5.14 specification.

Engineering for Use in the Postweld Heat-Treated Condition. Austenitic stainless steel weldments are usually put into service in the as-welded condition, except in the situations described below. When non-low-carbon grades are welded and sensitization that is due to chromium carbide precipitation cannot be tolerated, then annealing at 1050 to 1150 °C (1920 to 2100 °F), followed by a water quench, can be used to dissolve all carbides (and any intermetallic compounds, such as σ phase) and to maintain a solid solution. This annealing heat treatment also causes much of the ferrite that is present to transform to austenite.

Autogenous welds in high-molybdenum stainless steel, such as longitudinal seams in pipe, represent another exception. Annealing is used to cause the diffusion of molybdenum to erase the microsegregation of the as-welded condition, thus allowing the weld to match the pitting and crevice corrosion resistance of the base metal. Because these welds generally contain no ferrite in the as-welded condition, no ferrite is lost.

A third exception is the joints between austenitic stainless steel and carbon or low-alloy steel or the weld overlay of carbon or low-alloy steel with austenitic stainless steel filler metal. Type 309 filler metal often has been used for such a joint or overlay, so that the weld metal (especially in a multipass weld where type 309 is used throughout) contains an appreciable amount of ferrite. The HAZ of the carbon or low-alloy steel may require a PWHT at various temperatures from 600 to 700 °C (1110 to 1290 °F), depending on the specific alloy. These temperatures produce both carbide precipitation in the austenitic base metal and weld metal and transformation of ferrite to σ phase in the weld metal.

In the weld metal, carbide precipitation works against σ-phase formation because the carbides tend to precipitate where chromium is highest: at the ferrite-austenite interface. Because carbide precipitation reduces the chromium in the adjacent ferrite, some of it becomes unstable and transforms to austenite. This reduces the amount of ferrite available for σ formation. In a joint, but not in an overlay, corrosion resistance of the weld metal is seldom of concern because of the proximity of carbon or low-alloy steel. At the same ferrite content in the as-welded condition, less ferrite will be available for σ formation in type 309 than in type 309L weld metal. Therefore, the higher-carbon filler metal is often preferred for such applications.

At the same time, carbon from the mild steel or low-alloy steel immediately adjacent to the fusion line migrates to the higher-chromium weld metal, producing a layer of carbides along the fusion line in the weld metal and a carbon-depleted layer in the HAZ of the base metal. This carbon-depleted layer is weak at elevated temperatures. If the weldment is in service at suitable temperatures, then creep failure can occur there. There is also a serious mismatch in the CTE between austenitic stainless steel filler metal and carbon steel or low-alloy steel base metal. Therefore, thermal cycling produces strain accumulations along the interface and can lead to premature failure in creep, as has occurred in joints of chromium-molybdenum low-alloy steels to stainless steels. As a result, such dissimilar joints intended for elevated-temperature service are often welded with nickel-base alloy filler metals.

Although PWHT can be used for stress relief in austenitic stainless steel weldments, the yield strength of austenitic stainless steels falls much more slowly with rising temperature than does the yield strength of carbon or low-alloy steel. In addition, carbide precipitation and σ-phase embrittlement occur in the temperature range (600 to 700 °C, or 1110 to 1290 °F) where carbon steels and low-alloy steels are normally stress relieved. Significant improvement in relieving residual stresses without damaging corrosion resistance is only accomplished with a full anneal at temperatures ranging from 1050 to 1150 °C (1920 to 2100 °F), followed by rapid cooling to avoid carbide precipitation in unstabilized grades.

In stabilized grades, heat treatment at these temperatures causes niobium or titanium carbides to precipitate in part, which accomplishes much of the stabilization. However, rapid cooling is still necessary to obtain full corrosion resistance. In either case, the rapid cooling often reintroduces residual stresses, which means that this approach has to be used carefully. At the annealing temperatures, significant surface oxidation occurs in air. This oxide is tenacious on stainless steels and must be removed by pickling, followed by a water rinse, and then passivation, in order to restore full corrosion resistance.

References cited in this section

10. J.C. LIPPOLD AND W.F. SAVAGE, SOLIDIFICATION OF AUSTENITIC STAINLESS STEEL WELDMENTS: PART III--THE EFFECT OF SOLIDIFICATION BEHAVIOR ON HOT CRACKING SUSCEPTIBILITY, *WELD. J.,* VOL 61 (NO. 12), 1982, P 388S-396S
18. S.A. DAVID, J.M. VITEK, AND T.L. HEBBLE, EFFECT OF RAPID SOLIDIFICATION ON STAINLESS STEEL WELD METAL MICROSTRUCTURES AND ITS IMPLICATIONS ON THE SCHAEFFLER DIAGRAM, *WELD. J.,* VOL 66 (NO. 10), 1987, P 289S-300S
19. E. FOLKHARD, *WELDING METALLURGY OF STAINLESS STEELS,* SPRINGER-VERLAG, 1984, P 105
20. E. FOLKHARD, *WELDING METALLURGY OF STAINLESS STEELS,* SPRINGER-VERLAG, 1984, P 126

Precipitation-Hardening Stainless Steels

Basic Metallurgy. PH stainless steels can be divided into three groups, based on microstructure and precipitation behavior. Martensitic PH stainless steels are solutionized at a high temperature and transform to low-carbon martensite during cooling to room temperature. Then, they are further hardened by aging at an intermediate temperature to precipitate a second phase, which can be an intermetallic compound or a single component, such as copper. Both martensite and the precipitates harden the steel.

Semiaustenitic PH stainless steels can be cooled from their solutionizing temperature to room temperature without martensite formation, remaining rather soft austenite as they cool. Then, they are aged at an intermediate temperature, which causes carbides or an intermetallic compound to precipitate. This precipitation removes alloying elements and, possibly, carbon, from the solid solution, which destabilizes the austenite, allowing it to transform to martensite upon cooling from the aging temperature. Again, both martensite and the precipitates harden the steel.

Austenitic PH stainless steels contain enough alloys to be stable austenite upon cooling from their solutionizing temperature to room temperature. They are still stable austenite upon cooling from their aging temperature after intermetallic compounds are precipitated. In the case of austenitic PH stainless steels, only the precipitate hardens the alloy. As a result, the austenitic PH stainless steels cannot achieve tensile strengths as high as can be achieved with the martensitic or semiaustenitic grades.

Some of the PH stainless steels (some martensitic and semiaustenitic grades) can solidify as primary ferrite and have very good resistance to hot cracking. Because martensite is also ferromagnetic, a magnetic Ferrite Number loses meaning in terms of ensuring freedom from hot cracking. In the highly alloyed austenitic PH stainless steels, no ferrite is possible, and it is difficult to weld these without hot cracking.

Precipitation-Hardening Base Metals. Table 4(a) lists several PH stainless steels. Most of their applications are in the aerospace and other high-technology industries. Only one composition, the martensitic type 630 (17-4PH), is made as a welding filler metal in sufficient volume to warrant AWS classification of a matching filler metal, as can be seen in Table 4(b).

TABLE 4(A) COMPOSITIONS OF PRECIPITATION-HARDENING STAINLESS STEELS

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	COMPOSITION, WT%									
			C	**MN**	**P (MAX)**	**S (MAX)**	**SI (MAX)**	**CR**	**NI**	**MO**	**N**	OTHER
630	S17400	A 564	0.07 MAX	1.00 MAX	0.040	0.030	1.00	15.00-17.50	3.00-5.00	. . .	. . .	**NB + TA, 0.15-0.45; CU, 3.00-5.00**
631	S17700	A 564	0.09 MAX	1.00 MAX	0.040	0.030	1.00	16.00-18.00	6.50-7.75	. . .	. . .	**AL, 0.75-1.50**
632	S15700	A 564	0.09 MAX	1.00 MAX	0.040	0.030	1.00	14.00-16.00	6.50-7.75	2.00-3.00	. . .	**AL, 0.75-1.50**
633	S35000	A 693	0.07-0.11	0.50-1.25	0.040	0.030	0.50	16.00-17.00	4.00-5.00	2.50-3.25	0.07-0.13	**. . .**
634	S35500	564	0.10-0.15	0.50-1.25	0.040	0.030	0.50	15.00-16.00	4.00-5.00	2.50-3.25	0.07-0.13	**. . .**
635	S17600	A 564	0.08 MAX	1.00 MAX	0.040	0.030	1.00	16.00-17.50	6.00-7.50	. . .	. . .	**TI, 0.40-1.20; AL, 0.40 MAX**
XM-9	S36200	A 564	0.05 MAX	0.50 MAX	0.030	0.030	0.30	14.00-14.50	6.25-7.00	0.3 MAX	. . .	**TI, 0.60-0.90 AL, 0.10 MAX**
XM-12	S15500	A 564	0.07 MAX	1.00 MAX	0.040	0.030	1.00	14.00-15.50	3.50-5.50	. . .	. . .	**NB + TA, 0.15-0.45; CU, 2.50-4.50**
XM-13	S13800	A 564	0.05 MAX	0.20 MAX	0.010	0.008	0.10	12.25-13.25	7.50-8.50	2.00-2.50	0.01 MAX	**AL, 0.90-1.35**
XM-16	S45500	A 564	0.03 MAX	0.50 MAX	0.015	0.015	0.50	11.00-12.50	7.50-9.50	0.50 MAX	. . .	**NB + TA, 0.10-0.50; TI, 0.90-1.40; CU, 1.50-2.50**
. . .	S45503	A 564	0.010 MAX	0.50 MAX	0.010	0.010	0.20	11.00-12.50	7.50-9.50	0.50 MAX	. . .	**NB + TA, 0.10-0.50; TI, 1.00-1.35; CU, 1.50-2.50**
XM-25	S45000	A 564	0.05 MAX	1.00 MAX	0.030	0.030	1.00	14.00-16.00	5.00-7.00	0.50-1.00	. . .	**NB + TA, 8 × C MIN; CU, 1.25-1.75**

660 (A286)	S66286	A 638, A 453	0.08 MAX	2.00 MAX	0.040	0.030	1.00	13.50-16.00	24.00-27.00	1.00-1.50	. . .	**B, 0.0010-0.010; TI, 1.90-2.35; AL, 0.35 MAX; V, 0.10-0.50**
662	**K66220**	**A 638**	**0.08 MAX**	**1.50 MAX**	**0.040**	**0.030**	**1.00**	**12.00-15.00**	**24.00-28.00**	**2.50-3.50**	**. . .**	**B, 0.0010-0.010; TI, 1.55-2.00; CU, 0.50 MAX; AL, 0.35 MAX**

TABLE 4(B) PROPERTIES AND FILLER METALS FOR PRECIPITATION-HARDENING STAINLESS STEELS

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	**COMMENTS**	**TENSILE STRENGTH**		**YIELD STRENGTH**		**ELONGATION, %**	MATCHING FILLER METALS		
				MPA	**KSI**	**MPA**	**KSI**		**A5.4**	**A5.9**[(A)]	A5.22
630	S17400	A 564	480 °C (900 °F) AGE	1310	190	1170	170	10	E630-XX	ER630	. . .
631	S17700	A 564	510 °C (950 °F) AGE	1280	185	1030	150	6	. . .	. . .	. . .
632	S15700	A 564	510 °C (950 °F) AGE	1380	200	1210	175	7	. . .	. . .	. . .
633	S35000	A 693	450 °C (850 °F) AGE	1280	185	1030	150	8	. . .	. . .	. . .
634	S35500	564	540 °C (1000 °F) AGE	1170	170	1070	155	12	. . .	. . .	. . .
635	S17600	A 564	510 °C (950 °F) AGE	1310	190	1170	170	8	. . .	. . .	. . .
XM-9	S36200	A 564	AGE NOT SPECIFIED	1240	180	1210	175	10	. . .	. . .	. . .
XM-12	S15500	A 564	480 °C (900 °F) AGE	1310	190	1170	170	10	. . .	. . .	. . .
XM-13	S13800	A 564	510 °C (950 °F) AGE	1520	220	1410	205	10	. . .	. . .	. . .
XM-16	S45500	A 564	480 °C (900 °F) AGE	1620	235	1520	220	8	. . .	. . .	. . .
. . .	S45503	A 564	480 °C (900 °F) AGE	1620	235	1520	220	8	. . .	. . .	. . .
XM-25	S45000	A 564	480 °C (900 °F) AGE	1240	180	1170	170	10	. . .	. . .	. . .
660 (A286)	S66286	A 638, A 453	730 °C (1350 °F) AGE	900	130	590	85	15	. . .	. . .	. . .
662	**K66220**	**A 638**	**650 °C (1200 °F) AGE**	**860**	**125**	**550**	**80**	**15**	**. . .**	**. . .**	**. . .**

(A) METAL-CORED ELECTRODES, INDICATED BY A "C" IN PLACE OF THE "R" IN THE CLASSIFICATION, ARE ALSO INCLUDED.

Type 630 base metal (17Cr-4Ni-3Cu) is also the prototype martensitic PH stainless steel. In the solution-annealed (martensitic) condition, its yield strength is already above 750 MPa (110 ksi). The PH phase is copper, which, after aging at 480 °C (900 °F), is so fine that it can only be detected with a high-powered electron microscope. A yield strength of more than 1200 MPa (170 ksi) is obtained after this aging treatment. Higher aging temperatures coarsen the precipitate and reduce the strength.

Type 632 base metal (15Cr-7Ni-2Mo-1Al) is an example of a semiaustenitic PH stainless steel. In the solution-annealed (austenitic) condition, its yield strength is only on the order of 380 MPa (55 ksi), which is not very different from that of an ordinary austenitic stainless steel. Heat treatment at 955 °C (1750 °F) precipitates carbides and destabilizes the austenite, which transforms to martensite upon subsequent cooling. Then, aging at 510 °C (950 °F) precipitates nickel-aluminum intermetallic compounds. The aged martensite can exhibit a strength of approximately 1500 MPa (220 ksi). Again, aging at a higher temperature coarsens the precipitates and softens the steel.

Type 660 (A286) is the prototype austenitic PH stainless steel. Its high alloy content (15Cr-25Ni-1Mo-2Ti) provides higher corrosion resistance than that of the martensitic or semiaustenitic types. Furthermore, it contains no magnetic phases, which makes it suitable for use in high magnetic fields, such as those associated with superconducting magnets used for fusion energy research. Aging at about 730 °C (1350 °F) causes nickel-titanium intermetallic compound particles to precipitate, hardening the alloy. Its fully hardened yield strength (approximately 590 MPa, or 86 ksi), however, is considerably lower than that available from the martensitic or semiaustenitic grades.

Engineering for Use in the As-Welded Condition. Because PH stainless steels achieve their high strength by heat treatment, it is not reasonable to expect the weld metal to match the properties of the base metal in the as-welded condition. Therefore, the design of a weldment to be used in the as-welded condition generally must assume that the weld metal will under-match the strength of the base metal. If that is acceptable, then austenitic filler metals, such as types 308 and 309, can be suitable filler metals for the martensitic and semiaustenitic PH stainless steels. As with many of the other filler-metal selections in this article, consideration should be given to obtaining some ferrite in the weld metal to avoid hot cracking.

Example 8: Type 630 base metal (0.04C-16.5Cr-4Ni-4Cu-0.3Nb-0.02N, WRC-1992 $Cr_{eq} = 16.71$ and $Ni_{eq} = 6.80$; FN, off the diagram) to be welded with E308-16 electrode (0.05C-19.8Cr-9.6Ni-0.06N, WRC-1992 $Cr_{eq} = 19.8$ and $Ni_{eq} = 12.55$; FN = 6.2).

Assuming 30% dilution, the first weld bead would have $Cr_{eq} = 18.87$, $Ni_{eq} = 10.83$, and the calculated FN = 8.6. This is comfortably safe to ensure freedom from cracking during welding, and the weld composition is sufficiently rich in alloy that the weld metal will be stable as regards transformation of austenite to martensite. Thus, E308-16 is a good filler-metal selection, if its low strength, when compared with that of type 630 base metal, is acceptable. Clearly, type 309 filler metal would do as well.

Type A286 PH stainless steel, which is fully austenitic, presents a different situation if it is to be used in the as-welded condition. If nonmagnetic weld metal is essential, then type 310 filler metal is often the choice. Type 310 probably has the best cracking resistance of all of the fully austenitic filler metals. However, the first bead on A286 would be most at risk for hot cracking, because of the elements picked up from the A286 in the weld metal. Because a joint always presents a greater risk of hot cracking than does an overlay, a possible strategy to adopt is to first butter the A286 with type 310 weld metal. Then, welding type 310 to itself is less risky. In any event, the "weld ugly" practice of low heat input, convex beads, and so on, is always appropriate for crack-sensitive fully austenitic welds.

On the other hand, if a partially magnetic weld joint is acceptable and the weldment will not experience elevated temperatures where embrittlement could occur, then type 312 filler metal (at least for the weld deposits directly on A286) will provide the weld metal with ferrite and minimal hot-cracking risk. Once the A286 is covered with type 312 filler metal, the joint could be finished with a filler metal having a much lower ferrite content, such as type 308 or 309. However, type 309 will not provide ferrite when deposited directly on A286 with normal weld dilution.

Engineering for Use in the Postweld Heat Treated Condition. Normally, PWHT is done on a weldment of PH stainless steel because a weld-metal strength that is comparable to that of the base metal is desired. This, in turn, generally means that the weld metal must also be a PH stainless steel. Many PH stainless steel weldments are light-gage materials that readily lend themselves to autogeneous GTAW or LBW processes. Then, the weld metal matches the base metal and responds similarly to heat treatment.

Except for type 630, matching filler metals for the PH stainless steels are not very easy to locate and do not have AWS classifications. Some are covered by Aerospace Material Specifications (AMS) as bare wires only, which match base-metal compositions. These wires, used for GTAW or GMAW processes, are most likely to be available from suppliers to the aerospace industry. Alternatively, the shearing of base metal into narrow strips can provide filler metal for the GTAW process.

It is a common practice to use type 630 as a filler metal with several of the martensitic and semiaustenitic PH stainless steel base metals. The best weldment properties can then be obtained by a full-solution heat treatment, rapid cooling to room temperature, reheating to age harden at about 480 °C (900 °F), and cooling to room temperature. Because type 630 filler metal will transform to martensite upon cooling from solidification, it is also possible to obtain nearly full-strength weld metal, when using this filler metal, by simply cooling to room temperature and then aging at 480 °C (900 °F) to harden the weld metal. This presupposes that the weld metal does not pick up enough alloying elements by dilution to stabilize austenite and therefore does transform to martensite. When the weldment is simply aged after welding, without first using a solution heat treatment, some portion of the HAZ will have been overaged by the heat of welding and will become the weak link in the weldment. Then, 100% joint efficiency cannot be expected. A procedure qualification test becomes necessary to determine if the weldment strength will be adequate.

Welding of Stainless Steels

D.J. Kotecki, The Lincoln Electric Company

Duplex Ferritic-Austenitic Stainless Steels

Basic Metallurgy. Duplex ferritic-austenitic stainless steels solidify as essentially 100% ferrite. At high temperatures (on the order of 1300 °C, or 2370 °F), austenite nucleates and grows, first, at ferrite grain boundaries and, later, along preferred crystallographic directions within the ferrite grains. Diffusion of alloying elements must occur as the transformation of ferrite to austenite proceeds, with austenite-stabilizing elements (carbon, nickel, nitrogen, and copper) concentrating in the austenite, and ferrite-stabilizing elements (chromium, molybdenum, and tungsten) concentrating in the ferrite. The extent of the transformation depends not only on the balance between austenite-stabilizing elements and ferrite-stabilizing elements in the alloy, but on the time available for diffusion and on the actual diffusion of specific elements. It is possible to quench a duplex stainless steel from near the solidification temperature and to obtain nearly 100% ferrite at room temperature. Slow cooling, annealing, and hot working promote transformation of ferrite to austenite by promoting diffusion.

In a duplex ferritic-austenitic stainless steel, the optimum phase balance is usually approximately equal amounts of ferrite and austenite. Base-metal compositions are usually adjusted to obtain this phase balance as the equilibrium structure at approximately 1040 °C (1900 °F) after hot working and/or annealing. Of the alloying elements that affect the phase balance, carbon is generally undesirable for reasons of corrosion resistance. All of the other elements are slow to diffuse substitutional alloy elements, except for nitrogen. The other alloy elements contribute to determining the equilibrium phase balance, but nitrogen is most important in determining the relative ease of achieving a near-equilibrium phase balance. In developing the earliest duplex stainless steels (such as types 329 and CD-4MCu), nitrogen was not a deliberate alloying element. Therefore, under normal weld cooling conditions, the weld HAZs and matching weld metals reached room temperature with very little austenite, and had poor mechanical properties and corrosion resistance. To obtain useful properties, the welds required an annealing heat treatment, followed by quenching to avoid embrittlement of the ferrite by σ and other phases.

Overalloying of the weld filler metal with nickel causes the transformation to begin at a higher temperature, where diffusion is more rapid. Using this approach, a better phase balance can be obtained in the as-welded condition in the weld metal. However, this approach does nothing for the HAZ. Alloying with nitrogen, in the newer duplex stainless steels, usually solves the HAZ problem. With normal welding heat input and about 0.15% Ni, a reasonable phase balance is achieved in the HAZ. In addition, the nitrogen, when it diffuses to the austenite, imparts improved pitting resistance. On the other hand, if the cooling rate is too rapid and nitrogen is trapped in ferrite, then chromium nitride precipitates, damaging corrosion resistance. Therefore, low welding heat inputs should be avoided with duplex stainless steels.

The ferrite in the duplex stainless steels behaves much like a fully ferritic stainless steel in terms of formation of intermetallic compounds during heat treatment. The α, χ, and other phases can form in the temperature range from 500 to

1000 °C (930 to 1830 °F) in the base metals. High-nickel weld metals can extend the σ range to 1050 °C (1920 °F) or even higher (Ref 21). An α' embrittlement (475 °C, or 885 °F, embrittlement) can occur at temperatures as low as 250 °C (480 °F) to as high as 550 °C (1020 °F), depending on composition and time at temperature. An α' embrittlement occurs extremely slowly below approximately 350 °C (660 °F). Therefore, the American Society of Mechanical Engineers (ASME) code allows the use of UNS No. S31803 alloy at service temperatures up to 315 °C (600 °F). But, generally, the PWHT of duplex stainless steels needs to be done at temperatures above 1040 °C (1900 °F), or not at all.

It has been shown that duplex stainless steel weldments can suffer from hydrogen-induced cracking at very high hydrogen levels (Ref 22), and that lower levels damage ductility (Ref 23). This presents a special problem in duplex stainless steel weld metals, because austenite that occurs along original ferrite grain boundaries acts as a barrier to hydrogen escape. Therefore, the damage to ductility, caused by hydrogen, can remain indefinitely, in contrast to the easy escape of hydrogen from ferritic weld metals. Fortunately, α' embrittlement at 250 °C (480 °F) requires over 1000 h at temperature, whereas hydrogen escape from a 12 mm (0.5 in.) thick weldment can be achieved in about 100 h at this temperature (Ref 22). A postweld soak at this temperature can restore ductility to a hydrogen-damaged weldment.

Duplex Ferritic-Austenitic Base Metals. Tables 5(a) and 5(b) list selected duplex ferritic-austenitic stainless steel base metals and their compositions and mechanical property requirements, along with a limited number of "matching" filler metals. It is noteworthy that the yield strength of these steels is on the order of twice that of ordinary austenitic stainless steels, such as type 304 or 316L. The duplex stainless steels provide resistance to stress-corrosion cracking (SCC) in aqueous chloride-containing media. The level of resistance is generally superior to that of ordinary austenitic stainless steels, but may not be as good as that of the third-generation ferritic stainless steels described above. The newer duplex stainless steels, alloyed with nitrogen and molybdenum, also provide pitting resistance in aqueous chloride-containing media.

TABLE 5(A) COMPOSITIONS OF DUPLEX FERRITIC-AUSTENITIC STAINLESS STEELS

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	COMPOSITION, WT%									
			C (max)	**Mn**	**P (max)**	**S (max)**	**Si**	**Cr**	**Ni**	**Mo**	**N**	OTHER
XM-26	S31100	A 276	0.06	1.00 MAX	0.040	0.030	1.00 MAX	25.00-27.00	6.00-7.00	. . .	. . .	**TI, 0.25 MAX**
. . .	S31200	A 240	0.030	2.00 MAX	0.045	0.030	1.00 MAX	24.0-26.0	5.5-6.5	1.2-2.0	0.14-0.20	**. . .**
. . .	S31260	A 240	0.03	1.00 MAX	0.030	0.030	0.75 MAX	24.0-26.0	5.50-7.50	2.50-3.50	0.10-0.30	**CU, 0.20-0.08; W, 0.10-0.50**
. . .	S31500	A 790	0.030	1.20-2.00	0.030	0.030	1.40-2.00	18.0-19.0	4.25-5.25	2.50-3.00	0.05-0.1	**. . .**
2205	S31803	A 240, A 790	0.030	2.00 MAX	0.030	0.020	1.00 MAX	21.0-23.0	4.50-6.50	2.50-3.50	0.08-0.20	**. . .**
2304	S32304	A 790	0.030	2.50 MAX	0.040	0.040	1.0 MAX	21.5-24.5	3.0-5.5	0.05-0.60	0.05-0.20	**CU, 0.05-0.60**
255	S32550	A 240, A 790	0.04	1.5 MAX	0.040	0.030	1.0 MAX	24.0-27.0	4.5-6.5	2.9-3.9	0.10-0.25	**CU, 1.5-2.5**
2507	S32750	A 790	0.030	1.2 MAX	0.035	0.020	0.8 MAX	24.0-26.0	6.0-8.0	3.0-5.0	0.24-0.32	**CU, 0.5 MAX**
329	S32900	A 240, A 790	0.08	1.00 MAX	0.040	0.030	0.75 MAX	23.0-28.0	2.50-5.00	1.0-2.0	. . .	**. . .**
. . .	S32950	A 240, A 790	0.03	2.00 MAX	0.035	0.010	0.60 MAX	26.00-29.00	3.50-5.20	1.00-2.50	0.15-0.35	**. . .**
CD-4MCU	**. . .**	**A 743**	**0.40**	**1.00 MAX**	**0.04**	**0.04**	**1.00 MAX**	**24.5-26.5**	**4.75-6.00**	**1.75-2.25**	**. . .**	**CU, 2.75-3.35**

TABLE 5(B) PROPERTIES AND FILLER METALS FOR DUPLEX FERRITIC-AUSTENITIC STAINLESS STEELS

DESIGNATION	**UNS NO.**	**ASTM SPECIFICATION**	**TENSILE STRENGTH**		**YIELD STRENGTH**		**ELONGATION, %**	MATCHING FILLER METALS		
			MPA	**KSI**	**MPA**	**KSI**		**A5.4**	**A5.9**[(A)]	A5.22
XM-26	S31100	A 276	620	90	450	65	20	. . .	. . .	**. . .**
. . .	S31200	A 240	690	100	450	65	25	. . .	. . .	**. . .**
. . .	S31260	A 240	690	100	480	70	20	. . .	. . .	**. . .**
. . .	S31500	A 790	635	92	440	64	30	. . .	. . .	**. . .**
2205	S31803	A 240, A 790	620	90	450	65	25	E2209-XX	ER2209	**. . .**
2304	S32304	A 790	600	87	400	58	25	. . .	. . .	**. . .**
255	S32550	A 240, A 790	760	110	550	80	15	E2553-XX	ER2553	**. . .**
2507	S32750	A 790	800	116	550	80	15	. . .	. . .	**. . .**
329	S32900	A 240, A 790	620	90	480	70	15	. . .	. . .	**. . .**
. . .	S32950	A 240, A 790	690	100	480	70	15	. . .	. . .	**. . .**
CD-4MCU	**. . .**	**A 743**	**690**	**100**	**480**	**70**	**16**	**E2553-XX**	**ER2553**	**. . .**

(A) METAL-CORED ELECTRODES, INDICATED BY A "C" IN PLACE OF THE "R" IN THE CLASSIFICATION, ARE ALSO INCLUDED.

It is common to estimate pitting resistance of these alloys according to a pitting resistance equivalent (PRE_N). Various expressions have been used to calculate pitting resistance. The most common is (Ref 24):

$$PRE_N = \%CR + 3.3(\%MO) + 16(\%N)$$

When the PRE_N exceeds 40, the duplex stainless steel is said to be a "super duplex." A critical pitting temperature (CPT) of approximately 80 °C (180 °F) in a 6% ferric chloride solution is typically observed with super duplex stainless grades. New alloy developments in this area have led to several proprietary grades that are not yet addressed in ASTM standards.

Of the duplex stainless steels, the most common is probably type 2205. It is used in pressure vessels, piping, valves, fittings, and pumps for offshore oil collection and processing; chemicals handling; desalinization; and paper mills. Its CPT is on the order of 30 °C (85 °F). More-severe environments may require the use of a duplex stainless steel with 25% Cr, of which alloy 255 is the most common. It generally offers a CPT of 35 °C (95 °F) or higher. For the most severe environments that promote SCC, pitting, and crevice corrosion, the super duplex stainless steels are finding increasing application.

Because of their high yield strengths, duplex stainless steels are more difficult to cold form than are the austenitic stainless steels. Further, because of the various high-temperature embrittlements that can occur in these alloys, forging and other hot-working processes require more care than do the same processes with austenitic stainless steels.

Engineering for Use in the As-Welded Condition. As noted above, a weld metal with a matching composition to that of duplex stainless steels often has inferior ductility and toughness, because of a high ferrite content. This problem is less critical with inert gas-shielded weld metals (GTAW or GMAW process), when compared with flux-shielded weld metals (SMAW, SAW, or FCAW processes), but it is still significant. The safest procedure for the as-welded condition is to use a filler metal that matches the base metal in all respects, except that is has a higher nickel content. It is usually better to avoid autogenous welds, especially with the very rapid cooling rates that are typical of EBW or LBW processes. With the GTAW process, especially for the root pass, it is important to design the welding procedure to limit dilution of the weld metal by the base metal. This often means using a wider than normal root opening and more filler metal in the root than one would use for an austenitic stainless steel joint.

The AWS specifications only classify two filler-metal compositions for the duplex stainless steels. The 2209 filler metal composition, used as a covered electrode for the SMAW process or as a bare wire for SAW, GTAW, GMAW, plasma arc welding (PAW), and other welding processes, is a match for type 2205 base metal, except that it has a higher nickel content, which is necessary for proper ferrite control, ductility, and toughness. It can also be used for any lower-alloyed duplex stainless. The 2553 filler metal composition, used as a covered electrode or as a bare wire, is a match for alloy 255 base metal. In general, it can also be used for all lower-alloyed base metals. This leaves only the super duplex alloys, such as 2507, without an acceptable matching or overmatching filler metal in the AWS specifications. The supplier of such base metals should be contacted for welding filler-metal recommendations.

For submerged arc welding of the duplex stainless steels, high-basicity fluxes generally provide the best results, especially if weld-metal toughness is a concern. The toughness of duplex stainless steel weld metal is strongly sensitive to oxygen content, and basic fluxes provide the lowest weld-metal oxygen content with a given wire. Because fluxes for submerged arc welding of stainless steels are not classified by the AWS, the manufacturer is the best source of information regarding flux selection.

Welding procedure development for use in the as-welded condition, in addition to filler-metal selection, involves hydrogen considerations, preheat, interpass temperature, and welding heat input. In the past, argon-hydrogen gas mixtures have been utilized to some extent in the GTAW process to achieve better wetting and bead shape. However, instances of significant hydrogen embrittlement of GTAW weldments made with argon-hydrogen gas mixtures have occurred, and this practice is best avoided for weldments used in the as-welded condition.

Covered electrodes for use in the SMAW process should be treated as low-hydrogen electrodes for low-alloy steels, that is, they should be kept in sealed containers provided by their manufacturer until they are ready for use. Alternatively, they should be held in heated ovens, typically at 125 to 175 °C (260 to 350 °F), and exposure to a humid atmosphere should be limited to a few hours, at most.

In contrast to the situation with low-alloy steels, preheat and interpass temperature control are not practiced with duplex stainless steel weldments for the purpose of hydrogen cracking protection, but for HAZ microstructure control. A minimum preheat of 100 °C (212 °F) and a maximum interpass temperature of 200 °C (390 °F) can be recommended for most duplex stainless steel weldments. This modest preheat/interpass temperature helps to slow the cooling rate to promote more-extensive austenite formation in the HAZ, without posing potential problems with α' embrittlement.

A limitation on welding heat input also serves to promote austenite formation in the HAZ, as well as in the weld metal. Heat input below approximately 1.2 kJ/mm (30 kJ/in.) tends to result in undesirably high ferrite content in the HAZ and, sometimes, in the weld metal. Further, heat input levels above approximately 4 kJ/mm (100 kJ/in.) tend to produce excessive grain growth in the HAZ and, sometimes, in the weld metal, which reduces toughness and may also promote chromium nitride precipitation (Ref 25).

Engineering for Use in the Postweld Heat Treated Condition. Longitudinal seams in pipe lengths, welds in forgings, and repair welds in castings are often annealed after welding. This involves heating to 1040 °C (1900 °F) or higher, followed by rapid cooling, usually a water quench, to avoid σ phase. Annealing permits the use of either an exactly matched filler metal or no filler metal at all, because proper annealing adjusts the phase balance to near equilibrium. Slow heating to annealing temperature permits σand other phases to precipitate during heating, because the precipitation of σ phase occurs in only a few minutes at 800 °C (1470 °F). The production of pipe lengths often includes very rapid induction heating of the pipe, so that the σ phase is completely avoided during heating, and only a brief hold near 1040 °C (1900 °F) is necessary to accomplish phase balance control.

Furnace annealing can produce slow heating. Therefore, σ phase can be expected to form during heating. It is necessary to hold at the annealing temperature for longer times (perhaps an hour or more) to ensure that all σ is dissolved. However, properly mn continuous furnaces can provide high heating rates, especially for light-wall tubes and other thin sections. If the precipitation of σ phase during heating can be avoided, long anneals are not necessary. The extremely low creep strength of duplex stainless steels at their annealing temperatures can result in severe distortion during annealing. Rapid cooling to avoid σ phase also can contribute to distortion that is due to annealing.

High-nickel filler metals can present a special problem when the weldment is to be annealed. A higher nickel content raises the minimum temperature needed to dissolve the σ phase; temperatures as high as 1100 °C (2010 °F) may be needed at levels of 9 to 10% nickel (Ref 21), especially when the molybdenum content exceeds about 3%.

A very significant loss in weld metal yield strength occurs during the annealing of duplex stainless steel weld metals (Ref 21). In the as-welded condition, the yield strength of duplex stainless weld metal, whether matching nickel (approximately 5 to 6% Ni in the weld, as in the base metal) or high nickel (approximately 8 to 10% Ni in the weld), far exceeds base-metal minimum requirements. But after annealing, especially a furnace anneal, the yield strength of the weld metal usually falls to near (or even below) base-metal minimum requirements. It is thought that the strength loss is due to loss of coherence at the austenite-ferrite interfaces. Wrought base metal does not seem to be as badly affected, partly because hot working markedly textures it, that is, it aligns ferrite and austenite in stringers parallel to the rolling direction. Higher nitrogen contents (over 0.2%) in the weld metal may be necessary to maintain weld metal-yield strength after annealing (Ref 21).

Annealing in air causes a tenacious oxide to form on the surface of the duplex stainless steels. This must be removed to obtain optimal corrosion resistance of the weldment. Pickling with nitric plus hydrofluoric acid solutions, rinsing with clear water, and passivation with a nitric acid solution usually produces the optimum corrosion resistance with duplex stainless steels. The steel supplier can provide optimal solution concentrations for each alloy.

References cited in this section

21. D.J. KOTECKI, HEAT TREATMENT OF DUPLEX STAINLESS STEEL WELD METALS, *WELD. J.*, VOL 68 (NO. 11), 1989, P 431S-441S
22. C.D. LUNDIN, K. KIKUCHI, AND K.K. KHAN, "PHASE 1 REPORT: MEASUREMENT OF DIFFUSIBLE HYDROGEN CONTENT AND HYDROGEN EFFECTS ON THE CRACKING POTENTIAL OF DUPLEX STAINLESS STEEL WELDMENTS," WELDING RESEARCH COUNCIL, MARCH 1991
23. U. FEKKEN, L. VAN NASSAU, AND M. VERWEY, "HYDROGEN INDUCED CRACKING IN

AUSTENITIC/FERRITIC DUPLEX STAINLESS STEEL," PAPER 26, INTERNATIONAL CONFERENCE ON DUPLEX STAINLESS STEEL (THE HAGUE), NEDERLANDS INSTITUUT VOOR LASTECHNIEK, 26-28 OCT 1986

24. L. VAN NASSAU, H. MEELKER, AND J. HILKES, "WELDING DUPLEX AND SUPER-DUPLEX STAINLESS STEELS," DOCUMENT II-C-893-92, INTERNATIONAL INSTITUTE OF WELDING, 1992
25. I. VAROL, W.A. BAESLACK III, AND J.C. LIPPOLD, "FINAL REPORT: WELD CRACKING AND SOLID-STATE PHASE TRANSFORMATION CHARACTERISTICS IN DUPLEX STAINLESS STEELS," WELDING RESEARCH COUNCIL, 30 JUNE 1989

Welding of Stainless Steels

D.J. Kotecki, The Lincoln Electric Company

Welding Process and Parameter Selection for Stainless Steels

Stainless steels of all types are weldable by virtually all welding processes. In part, process selection is often dictated by available equipment. Perhaps the simplest and most universal welding process is manual SMAW with coated electrodes. It has been applied to material as thin as 1.2 mm (0.05 in.), and there is no upper limit on thickness. Other very commonly used processes for stainless steels are GTAW, GMAW, SAW, FCAW, and several forms of resistance welding. The PAW, LBW, and EBW processes are also used.

Detailed discussion of all of the possible processes and parameters is beyond the scope of this article. However, the common arc welding processes are briefly discussed below, especially with regard to procedure and technique errors that can lead to loss of ferrite control with the common austenitic stainless steel weld metals that are designed to contain a small amount of ferrite (that is, solidify as primary ferrite) for protection from hot cracking.

Shielded-Metal Arc Welding. The ANSI/AWS A5.4-92 specification provides for five different electrode coating types. If the alloy to be deposited were type 308, then the five coating types would be typified by the following five electrode classifications: E308-15, E308-16, E308-17, E308-25, and E308-26. Of these, the E308-25 and E308-26 are typically very heavily coated electrodes, often made with mild-steel or lower-alloyed core wire, with all or most of the alloying elements in the coating. For these two coating types (-25 and -26), the various commercial products are not sufficiently alike to permit a statement to be made about typical current ranges for the various sizes. The -25 coating type is recommended only for direct current, electrode positive (DCEP). The -26 coating type is recommended for alternating current, as well as DCEP. Both the -25 and -26 coating types are generally limited to welding in the flat and horizontal positions. The remaining three coating types are very well standardized in existing commercial products, although the -17 type has been classified separately from the -16 type only since the 1992 edition of the ANSI/AWS A5.4 specification.

The -15 coating type is recommended for welding on DCEP only. Its slag system, which generally consists of lime and fluorspar, permits the easiest out-of-position welding of all of the stainless steel electrode coating types, especially for fixed-position pipe welding. It tends to produce a convex weld surface profile, which provides the best cracking resistance when primary ferrite solidification is not achieved. It usually requires more welder skill to make attractive-appearing welds, but it also has the greatest resistance to porosity when the electrode is exposed to moisture or when the careless welder draws a long arc. Its arc is relatively harsh, and surface ripple is coarse. In general, it produces more spatter than the other coating types.

The -16 coating type is recommended for welding both DCEP and alternating current. Its slag system, which is very high in titanium dioxide, permits fairly easy out-of-position welding, but it is generally necessary to weave slightly for vertical-up welding, which is not necessary for the -15 coating type. It tends to produce a nearly flat surface profile. The characteristics of the -16 coating that are not quite as good as those of the -15 coating type, but are more than adequate for most purposes, are its weld metal cracking resistance when no ferrite is formed; the resistance of the coating to porosity, which is due to either moisture pickup or the drawing of a long arc; and its out-of-position weldability. Its weld surface appearance, spatter level, and slag removal are generally more attractive than those of the -15 coating type.

The -17 coating type is a modification of the -16 coating type, produced by substituting silica and silicates for some of the titanium dioxide, which welds both DCEP and alternating current. The use of this coating type and a comparison to the -16 coating type were described earlier in the section on engineering for use of austenitic stainless steels in the as-welded condition.

Table 6 lists typical welding current ranges for the three more-standardized electrode coating types. Welding voltage is determined by the arc length that the welder holds. Some -16 coated electrodes, and most -17 electrodes, can be used by dragging the coating on the base-metal surface, which maintains a short arc length. This length is generally desirable, because a longer arc length permits more air to enter the arc and the weld metal. Nitrogen from the air can be picked up by the weld metal and become an alloying element. This means that the careless welder who draws a long arc with an electrode designed to produce primary ferrite solidification can add enough nitrogen to change the solidification mode to primary austenite (refer to Fig. 3, which shows that nitrogen promotes austenite and suppresses ferrite). This can result in hot cracks in a weld metal, caused by electrodes designed to produce sufficient ferrite to avoid hot cracking. The -15 coating type is the least sensitive to this error, whereas the -17 coating type is generally the most sensitive.

TABLE 6 TYPICAL CURRENT RANGES FOR STAINLESS STEEL SMAW ELECTRODES

ELECTRODE DIAM		CURRENT (DCEP), A		
MM	**IN.**	**FOR E3XX-15**	**FOR E3XX-16$^{(A)}$**	FOR E3XX-17$^{(A)}$
1.6	$\frac{1}{16}$	20-40	20-40	**. . .**
2.0	$\frac{5}{64}$	30-50	30-50	**35-55**
2.4	$\frac{3}{32}$	30-70	40-70	**40-80**
3.2	$\frac{1}{8}$	45-95	60-100	**75-110**
4.0	$\frac{5}{32}$	75-130	90-140	**95-150**
4.8	$\frac{3}{16}$	95-165	120-185	**130-200**
6.4	$\frac{1}{4}$	**150-225**	**200-300**	**. . .**

(A) CURRENT RANGES ARE TYPICALLY 10% HIGHER WHEN ALTERNATING CURRENT IS USED FOR E3XX-16 OR E3XX-17 ELECTRODES.

Figure 16 provides the suggested procedures for the SMAW of butt joints in stainless steel sheet and plate in the flat position. Figure 17 provides the suggested procedures for butt welds in the vertical and overhead positions. Figure 18 addresses fillet welds in the flat or horizontal positions. Figure 19 considers lap joints, and Fig. 20 considers comer joints. The information in these figures is meant only as a starting point in procedure development. Although originally developed for austenitic stainless steels, the procedures are also reasonable starting points for other stainless steels within the guidelines given elsewhere in this article.

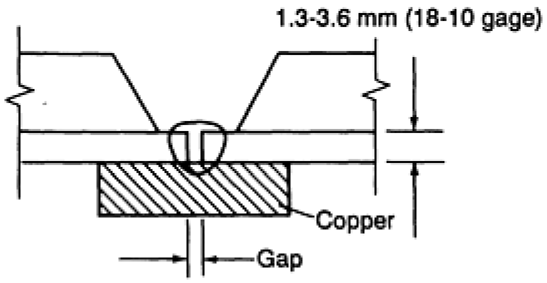

Joint A

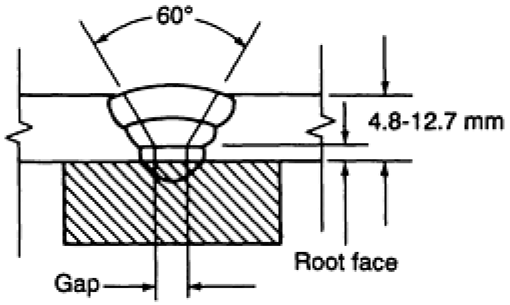

Joint B

PLATE THICKNESS			ELECTRODE DIAMETER				CURRENT (DCEP)(A), A		ARC SPEED				AMOUNT OF ELECTRODE REQUIRED		TOTAL TIME OF WELD		GAP		ROOT FACE	
			FIRST PASS		ADDITIONAL PASSES				FIRST PASS		ADDITIONAL PASSES									
mm	**in.**	**gage**	**mm**	**in.**	**mm**	**in**	**First pass**	**Additional passes**	**mm/s**	**in./min**	**mm/s**	**in./min**	**kg/m**	**lb/ft**	**h/m**	**h/ft**	**mm**	**in.**	**mm**	in.
JOINT A																				
1.3	0.050	18	2.0	$\frac{5}{64}$	. . .	. . .	40(B)	. . .	5.9-6.8	14-16	. . .	. . .	0.030	0.020	0.0436	0.0133	0	0	0	**0**
2.0	0.078	14	2.4	$\frac{3}{32}$	. . .	. . .	60	. . .	4.9-5.3	11.5-12.5	. . .	. . .	0.057	0.038	0.0548	0.0167	0.8	$\frac{1}{32}$	0	**0**
3.6	0.140	10	3.2	$\frac{1}{8}$	. . .	. . .	85	. . .	3.6-4.0	8.5-9.5	. . .	. . .	0.119	0.080	0.0728	0.0222	0.8	$\frac{1}{32}$	0	**0**
JOINT B																				
4.8	0.19	. . .	4.0	$\frac{5}{32}$	. . .	. . .	125	. . .	2.8-3.1	6.7-7.3	. . .	. . .	0.223	0.150	0.0938	0.0286	1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$
6.4	0.25	. . .	4.0	$\frac{5}{32}$	4.8(C)	$\frac{3}{16}$ (C)	125	160(C)	2.4-2.7	5.7-6.3	3.2-3.6(C)	7.6-8.5(C)	0.506	0.340	0.1913	0.0583	2.4	$\frac{3}{32}$	1.6	$\frac{1}{16}$
9.5	0.38	. . .	4.0	$\frac{5}{32}$	4.8(D)	$\frac{3}{16}$ (D)	125	160(D)	2.4-2.7	5.7-6.3	2.4-2.7(D)	5.7-6.3(D)	0.968	0.650	0.3281	0.100	2.4	$\frac{3}{32}$	1.6	$\frac{1}{16}$
12.7	**0.50**	**. . .**	**4.0**	$\frac{5}{32}$	**4.8(E)**	$\frac{3}{16}$ **(E)**	**125**	**160(E)**	**2.4-2.7**	**5.7-6.3**	**2.4-2.7(E)**	**5.7-6.3(E)**	**1.579**	**1.06**	**0.5479**	**0.167**	**2.4**	$\frac{3}{32}$	**1.6**	$\frac{1}{16}$

(A) ALTERNATING CURRENT CAN BE USED WITH 10% INCREASE IN CURRENT; E3XX-15 ELECTRODE CAN BE USED WITH 10% DECREASE IN CURRENT.

(B) USE DCEN.
(C) PASS 2.
(D) PASSES 2 AND 3.
(E) PASSES 2 THROUGH 5

FIG. 16 SUGGESTED PROCEDURES FOR SMAW OF BUTT JOINTS IN AUSTENITIC STAINLESS STEEL FROM 1.3 MM (0.05 IN., OR 18 GAGE) TO 12.7 MM (0.5 IN.) IN THICKNESS IN THE FLAT POSITION. E3XX-16 ELECTRODE. SOURCE: REF 26

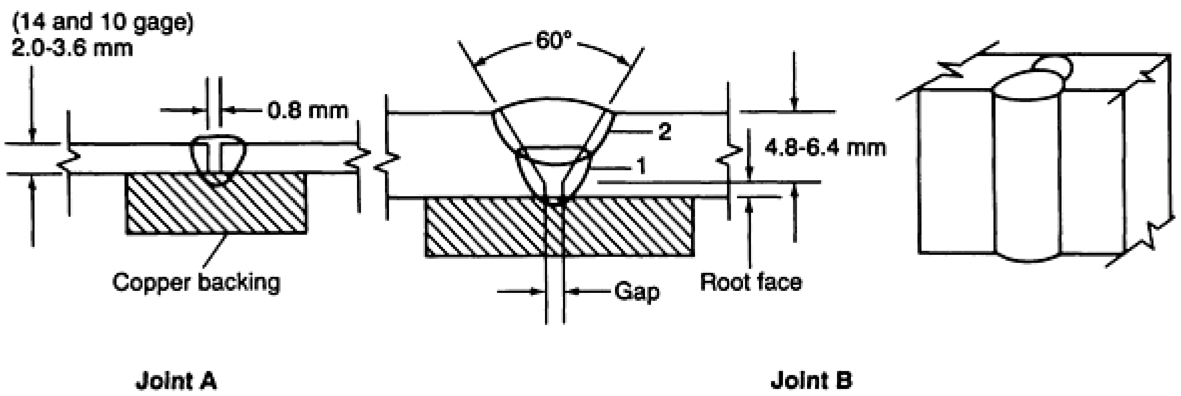
(14 and 10 gage)
2.0-3.6 mm
0.8 mm
Copper backing
60°
2
1
4.8-6.4 mm
Gap
Root face
Joint A
Joint B

PLATE THICKNESS			ELECTRODE DIAMETER		CURRENT (DCEP), A	ARC SPEED				AMOUNT OF ELECTRODE REQUIRED		TOTAL TIME OF WELD		GAP		ROOT FACE	
						FIRST PASS		SECOND PASS									
MM	IN.	GAGE	MM	IN.		mm/s	in./min	mm/s	in/min	kg/m	lb/ft	h/m	h/ft	mm	in.	mm	in.
JOINT A																	
2.0$^{(A)}$	0.078$^{(A)}$	14	2.4	$\frac{3}{32}$	50	5.9-6.8	14-16	. . .	. . .	0.045	0.030	0.0436	0.0133	0	0	0	**0**
3.6	0.140	10	3.2	$\frac{1}{8}$	75	2.8-3.1	6.7-7.3	. . .	. . .	0.136	0.091	0.0938	0.0286	0	0	0	**0**
JOINT B																	
4.8	$\frac{3}{16}$	. . .	4.0	$\frac{5}{32}$	110	2.2-2.5	5.2-5.8	. . .	. . .	0.238	0.160	0.1194	0.0364	1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$
6.4	$\frac{1}{4}$	**. . .**	**4.0**	$\frac{5}{32}$	**L10**	**2.2-2.5**	**5.2-5.8**	**1.8-2.0**	**4.3-4.7**	**0.551**	**0.370**	**0.2651**	**0.0808**	**2.4**	$\frac{1}{16}$	**1.6**	$\frac{1}{16}$

FIG. 17 SUGGESTED PROCEDURES FOR SMAW OF BUTT JOINTS IN AUSTENITIC STAINLESS STEEL FROM 2.0 MM (0.08 IN., OR 14 GAGE) TO 6.4 MM (0.25 IN.) IN THICKNESS IN THE VERTICAL AND OVERHEAD POSITIONS. E3XX-15 ELECTRODE. SOURCE: REF 26

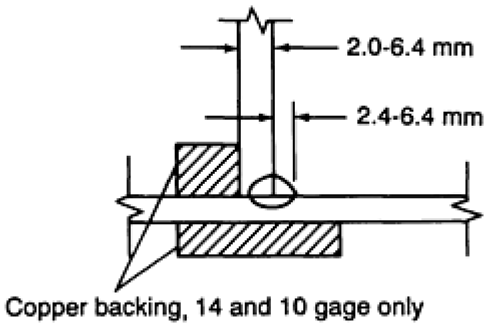

Joint A

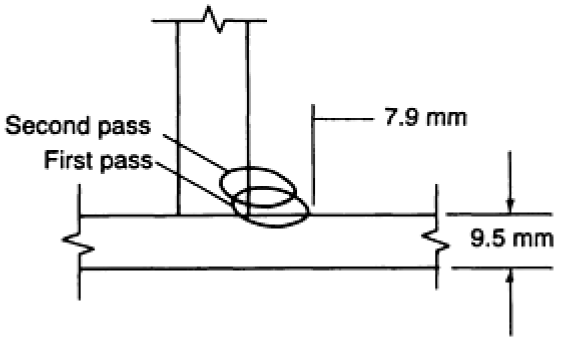

Joint B

PLATE THICKNESS			**WELD SIZE**		**ELECTRODE DIAMETER**		**CURRENT (DCEP)$^{(A)}$, A**	**ARC SPEED**				**AMOUNT OF ELECTRODE REQUIRED**		TOTAL TIME OF WELD	
								FIRST PASS		**SECOND PASS**					
mm	**in.**	**gage**	**mm**	**in.**	**mm**	**in.**		**mm/s**	**in./min**	**mm/s**	**in./min**	**kg/m**	**lb/ft**	**h/m**	h/ft
JOINT A															
2.0	0.078	14	2.4	3/32	2.4	3/32	60	5.3-5.7	12.5-13.5	. . .	. . .	0.054	0.036	0.051	**0.0154**
3.6	0.140	10	3.2	1/8	3.2	1/8	85	5.3-5.7	12.5-13.5	. . .	. . .	0.083	0.056	0.051	**0.0154**
4.8	3/16	. . .	4.8	3/16	4.0	5/32	120	3.6-4.0	8.6-9.4	. . .	. . .	0.178	0.120	0.073	**0.0222**
6.4	1/4	. . .	6.4	1/4	4.8	3/16	160	2.6-2.9	6.2-6.8	. . .	. . .	0.328	0.220	0.101	**0.0308**
JOINT B															
9.5	3/8	**. . .**	**7.9**	5/16	**4.8**	3/16	**170**	**2.6-2.9**	**6.2-6.8**	**2.8-3.1**	**6.7-7.3**	**0.640**	**0.430**	**0.195**	**0.0594**

Note: For vertical and overhead welding position, use same procedures as for vertical and overhead butt welds.

(A) ALTERNATING CURRENT CAN BE USED WITH A 10% INCREASE IN CURRENT; E3XX-15 ELECTRODE CAN BE USED WITH A 10% DECREASE IN CURRENT.

FIG. 18 SUGGESTED PROCEDURES FOR SMAW OF FILLET JOINTS IN AUSTENITIC STAINLESS STEEL FROM 2.0 MM (0.08 IN., OR 14 GAGE) TO 9.5 MM (0.38 IN.) IN THICKNESS IN THE FLAT OR HORIZONTAL POSITIONS. E3XX-16 ELECTRODE. SOURCE: REF 26

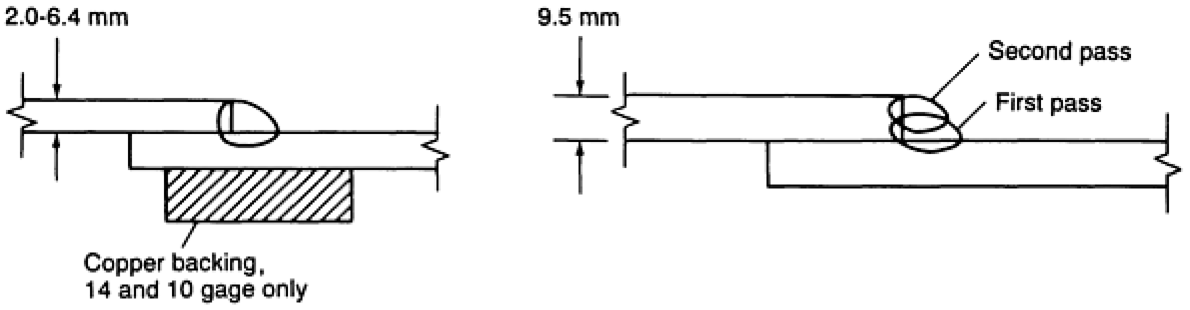

Joint A

Joint B

PLATE THICKNESS			ELECTRODE DIAMETER		CURRENT (DCEP)$^{(A)}$, A	ARC SPEED				AMOUNT OF ELECTRODE REQUIRED		TOTAL TIME OF WELD	
						FIRST PASS		SECOND PASS					
mm	in.	gage	mm	in.		mm/s	in./min	mm/s	in./min	kg/m	lb/ft	h/m	h/ft
JOINT A													
2.0	0.078	14	2.4	3/32	60	5.3-5.7	12.5-13.5	. . .	. . .	0.054	0.036	0.051	**0.0154**
3.6	0.140	10	3.2	1/8	90	5.3-5.7	12.5-13.5	. . .	. . .	0.083	0.056	0.051	**0.0154**
4.8	3/16	. . .	4.0	5/32	125	3.6-4.0	8.6-9.4	. . .	. . .	0.194	0.130	0.073	**0.0222**
6.4	1/4	. . .	4.8	3/16	170	2.6-2.9	6.2-6.8	. . .	. . .	0.357	0.240	0.101	**0.0308**
JOINT B													
9.5	3/8	**. . .**	**4.8**	3/16	**175**	**2.6-2.9**	**6.2-6.8**	**2.8-3.1**	**6.7-7.3**	**0.685**	**0.460**	**0.195**	**0.0594**

(A) ALTERNATING CURRENT CAN BE USED WITH A 10% INCREASE IN CURRENT; E3XX-15 ELECTRODE CAN BE USED WITH A 10% DECREASE IN CURRENT.

FIG. 19 SUGGESTED PROCEDURE FOR SMAW OF LAP JOINTS IN AUSTENITIC STAINLESS STEEL FROM 2.0 MM (0.08 IN., OR 14 GAGE) TO 9.5 MM (0.38

IN.) IN THICKNESS IN THE HORIZONTAL POSITION. E3XX-16 ELECTRODE. SOURCE: REF 26

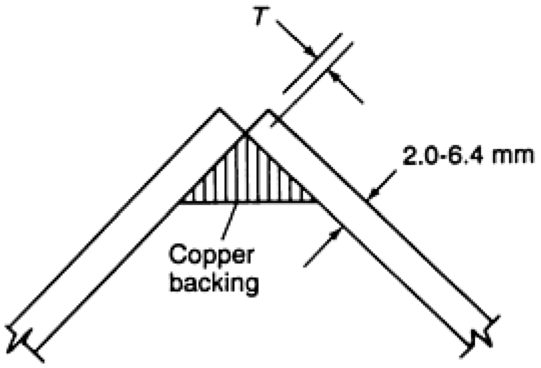

Joint A

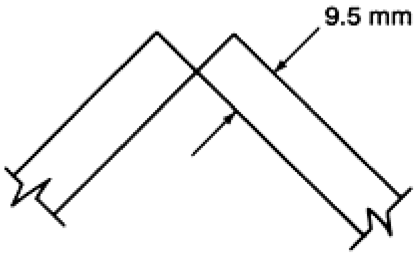

Joint B

PLATE THICKNESS			ELECTRODE DIAMETER		CURRENT (DCEP)$^{(A)}$, A		ARC SPEED FIRST PASS		ARC SPEED SECOND PASS		AMOUNT OF ELECTRODE REQUIRED		TOTAL TIME OF WELD		THICKNESS, *T*	
mm	**in.**	**gage**	**mm**	**in.**	**FIRST PASS**	**SECOND PASS**	**mm/s**	**in./min**	**mm/s**	**in./min**	**kg/m**	**lb/ft**	**h/m**	**h/ft**	**mm**	in.
JOINT A																
2.0	0.078	14	2.4	3/32	60	. . .	5.9-6.8	14-16	. . .	. . .	0.042	0.028	0.0436	0.0133	1.0	**0.04**
3.6	0.140	10	3.2	1/8	85	. . .	5.3-5.7	12.5-13.5	. . .	. . .	0.083	0.056	0.0505	0.0154	0.8	1/32
4.8	3/16	. . .	4.0	5/32	125	. . .	4.4-4.9	10.5-11.5	. . .	. . .	0.140	0.094	0.0597	0.0182	1.2	3/64
6.4	1/4	. . .	4.8	3/16	160	. . .	2.6-2.9	6.2-6.8	. . .	. . .	0.33	0.22	0.101	0.0308	1.6	1/16
JOINT B																
9.5	3/8	**. . .**	**4.8**	3/16	**160**	**175**	**2.6-2.9**	**6.2-6.8**	**2.4-2.7**	**5.7-6.3**	**0.67**	**0.45**	**0.210**	**0.0641**	**0**	**0**

FIG. 20 SUGGESTED PROCEDURE FOR SMAW OF CORNER JOINTS IN AUSTENITIC STAINLESS STEEL FROM 2.0 MM (0.08 IN., OR 14 GAGE) TO 9.5 MM (0.38 IN.) IN THICKNESS IN THE FLAT POSITION. E3XX-16 ELECTRODE. SOURCE: REF 26

Submerged Arc Welding. The ANSI/AWS A5.9-92 specification classifies both solid and composite (metal-cored tubular or stranded) wires that can be used for several welding processes, including submerged arc welding. Wires are classified only on the basis of their composition and whether they are solid or composite. For the ferrite-containing austenitic grades, the chromium content of a wire of a given alloy type is generally shifted to a higher level for the wires, compared with SMAW electrodes. For example, the chromium content for E308-XX electrodes ranges from 18.0 to 21.0%, but for ER308 wire, it is 19.5 to 22.0%. There is a good reason for this. The composition requirements for SMAW electrodes are for the all-weld metal deposit, that is, the finished weld, whereas the requirements for the wires are for the wire composition, the tool that is used to produce the finished weld. Some loss of chromium is to be expected in transferring the metal across the arc in SAW, as well as in gas-shielded processes. In the SAW process, chromium recovery depends strongly on the specific flux chosen and on the choice of welding parameters.

The AWS filler-metal specifications do not classify fluxes for the submerged arc welding of stainless steels. However, practical experience indicates that there are essentially three groupings of SAW fluxes with respect to chromium recovery. In submerged arc welding of carbon steels and low-alloy steels, it is common to describe fluxes as "neutral" or "active," depending on whether manganese and silicon content of the deposited weld metal vary appreciably with changes in arc voltage or flux-to-wire ratio. However, a flux that is neutral with respect to manganese and silicon recovery may be very aggressive in removing chromium from the weld metal. Therefore, "neutral" and "active" descriptions from the ANSI/AWS A5.17 (mild-steel submerged arc) or ANSI/AWS A5.23 (low-alloy submerged arc) specifications have no meaning with regard to stainless steel.

A somewhat quantitative grouping of fluxes for stainless steels relates all-weld metal chromium content to wire chromium content. A number of fluxes that are high in manganese silicate and the other oxides that are less stable than chromium oxide tend to deplete the chromium of the deposit on the order of 2%, or even more, in absolute terms. With such a flux, a wire of 21% Cr is likely to produce a deposit of only 19% Cr, which fits well with the higher chromium content of the wires of ANSI/AWS A5.9, when compared with the electrodes of ANSI/AWS A5.4. With such fluxes, a higher arc voltage, or any other change that tends to increase the ratio of flux melted to wire melted, will tend to reduce the chromium recovery further. These fluxes are sometimes referred to as "acid" fluxes.

A second group of fluxes for stainless steels are those composed primarily of oxides of calcium, magnesium, titanium, and aluminum, along with calcium fluoride. These compounds are more stable than chromium oxide. Fluxes of this group tend to produce only a small loss of chromium, typically on the order of 0.5 to 1.0%, in absolute terms, when compared with the wire compositions. With such fluxes, a higher voltage, or any other change that tends to increase the ratio of flux melted to wire melted, produces only a small increase in chromium loss. These fluxes are sometimes referred to as "basic" fluxes.

A third group of fluxes for stainless steels are those that include, in the individual flux particles, particles of chromium metal or alloy. The amount of this chromium is usually only slightly more than that needed to make up for the chromium that is lost to oxidation during welding. Therefore, a small chromium increase, typically on the order of 0.5 to 1%, is observed in the weld metal, when compared with the wire. This chromium gain may increase with increase in arc voltage, or with any other welding parameter change that increases the ratio of flux melted to wire melted. These fluxes are sometimes referred to as "chromium-compensating" fluxes.

Loss of chromium not only means loss of some corrosion resistance, but also loss of ferrite in nominally austenitic stainless steel weld metals that are intended to contain some ferrite. The use of an acid flux is likely to produce about 5 FN less than the Ferrite Number calculated from the wire composition using Fig. 3. The use of a basic flux is likely to produce almost the same Ferrite Number as that calculated from the wire composition using Fig. 3. The use of a chromium-compensating flux is likely to produce as much as 5 FN, or more, greater than the FN value calculated from the wire composition using Fig. 3.

A further consideration in flux selection is carbon recovery with martensitic stainless steels. The acid fluxes tend to remove considerable carbon from high-carbon stainless steels, so that the ability to achieve a given hardness level is affected. The basic fluxes tend to recover nearly all of the wire carbon. The chromium-compensating fluxes can do either, depending on the specific oxide balance in a given flux. The use of either an acid flux or a chromium-compensating flux with a martensitic stainless steel wire like ER410 may produce ferrite stringers in what ought to be a fully martensitic deposit. As a result, the mechanical properties may be inferior.

Submerged arc welding permits much broader ranges of current for a given electrode size than does the SMAW process. Current is largely determined by wire feed speed, although increased electrode extension ("stickout") will decrease current at a given wire feed speed. Figure 21 presents typical wire feed speed ranges for a variety of solid electrode sizes and the corresponding welding currents with austenitic filler metals. Other current ranges result when either composite or martensitic wires are used, because the electrical resistance of these wires differs from that of the austenitic wires. Nevertheless, Fig. 21 can provide a starting point. In general, as the wire feed speed is increased, a corresponding voltage increase is appropriate for the best bead shape. Typically, 26 to 28 V produce the best results at the bottom end of the wire feed speed range, and 32 to 36 V produce the best results at the top end of the wire feed speed range.

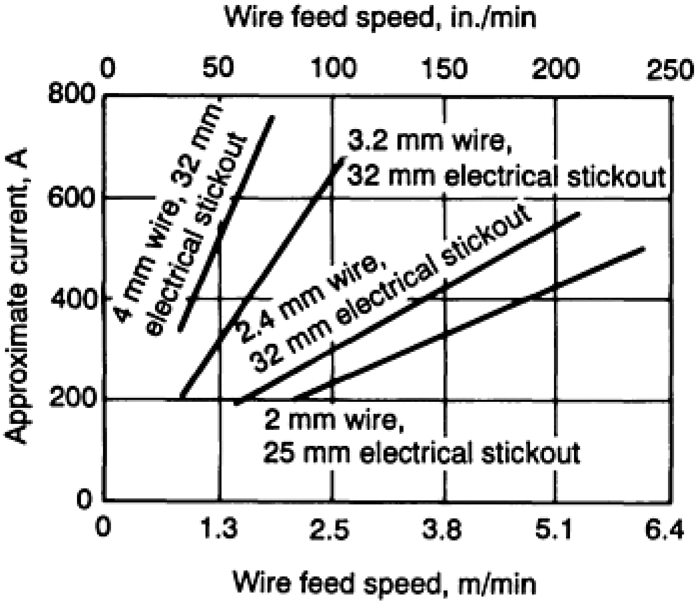

FIG. 21 CURRENT (DCEP) VERSUS WIRE FEED SPEED FOR SUBMERGED ARC WELDING WITH E3XX STAINLESS STEEL ELECTRODES. SOURCE: REF 27

Because there is no classification system for submerged arc fluxes for stainless steels, and because the broad range of chromium recovery that is possible depends on the specific flux chosen, either the advice or literature of the flux manufacturer should be obtained before using an unfamiliar flux.

Figures 22 and 23 provide joint preparations that have proven useful in submerged arc butt welding of stainless steel plate. Figure 24 provides copper backing information for one-side butt welding using the SAW process. Circumferential welding on tubular components is usually done by off-setting the welding head from top-dead-center, so that solidification occurs in the flat position, slightly uphill or slightly downhill. This affects the weld profile, as shown in Fig. 25.

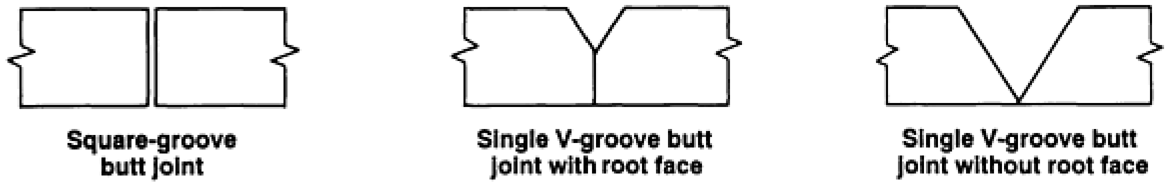

Square-groove butt joint

Single V-groove butt joint with root face

Single V-groove butt joint without root face

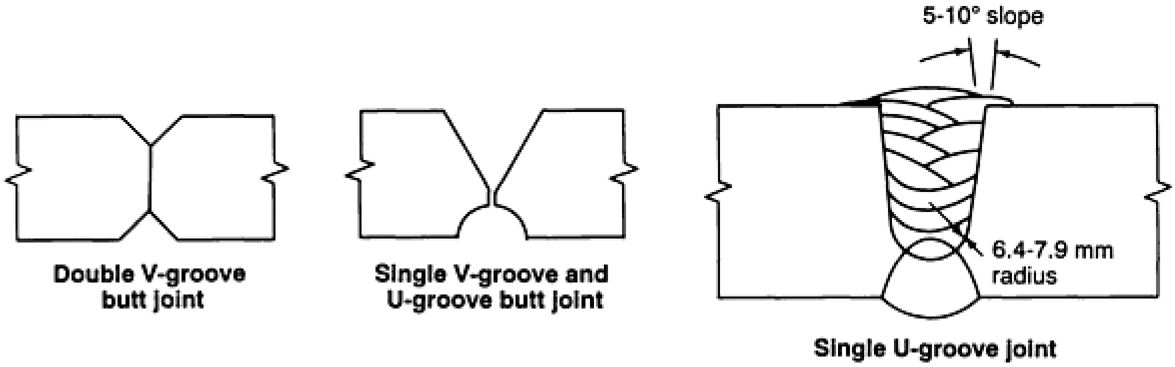

Double V-groove butt joint

Single V-groove and U-groove butt joint

Single U-groove joint

5-10° slope
6.4-7.9 mm radius

FIG. 22 BUTT JOINT DESIGNS FOR SUBMERGED ARC WELDING. SOURCE: REF 26

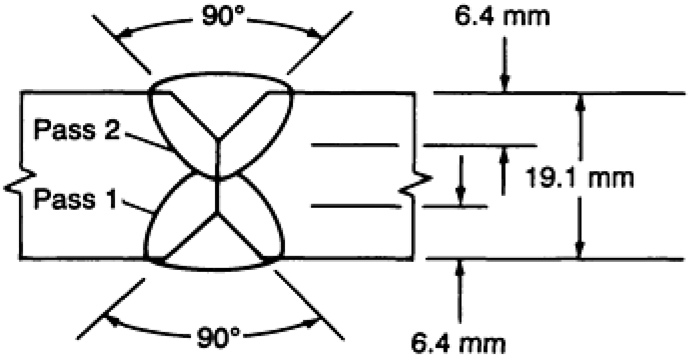
90°
6.4 mm
Pass 2
Pass 1
19.1 mm
90°
6.4 mm

FIG. 23 TYPICAL DOUBLE-V WELD IN TYPE 304 PLATE. PASS 1 WAS MADE AT 700 A AND 33 V, AT 6.8 MM/S (16 IN./MIN). PASS 2 WAS MADE AT 950 A AND 35 V, AT 5.1 MM/S (12 IN./MIN). POWER, DCEP; TYPE 308 ELECTRODE, 4.8 MM ($\frac{3}{16}$ IN.) DIAM; NEUTRAL FLUX. SOURCE: REF 26

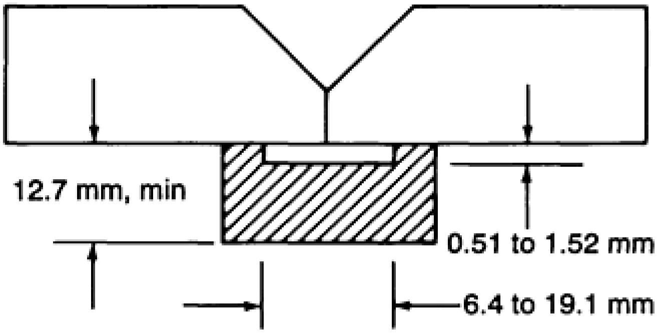
12.7 mm, min
0.51 to 1.52 mm
6.4 to 19.1 mm

FIG. 24 RECOMMENDED GROOVE DIMENSIONS FOR COPPER BACKING BARS IN THE SUBMERGED ARC WELDING OF STAINLESS STEELS. SOURCE: REF 26

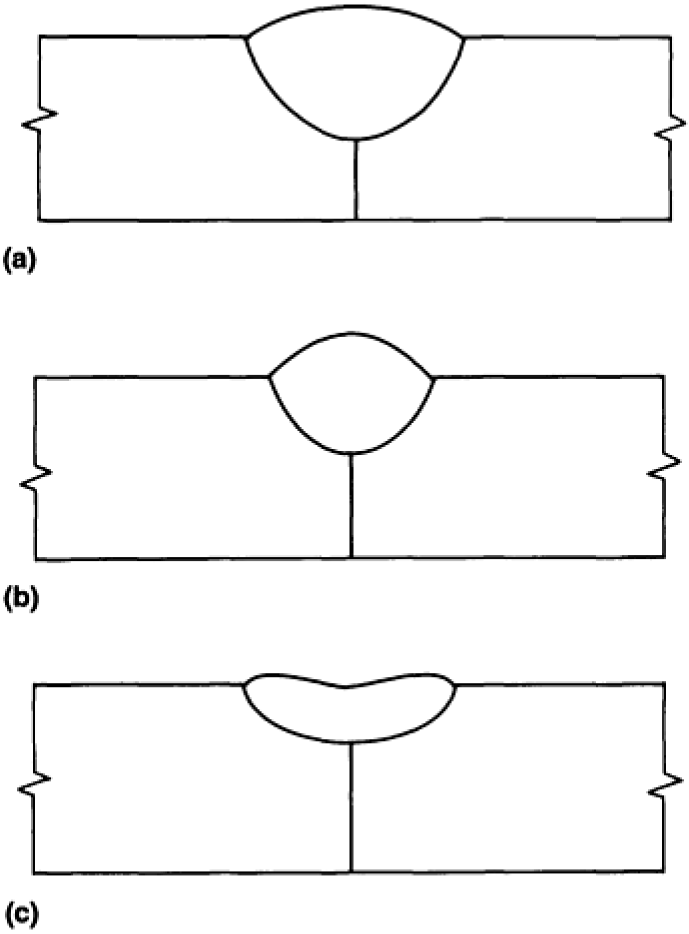

FIG. 25 WELD PROFILES. (A) CONTOUR OF A WELD BEAD IN THE FLAT POSITION WITH THE WORK HORIZONTAL. (B) WELDING SLIGHTLY UPHILL. (C) WELDING SLIGHTLY DOWNHILL. SOURCE: REF 26

Gas-Tungsten Arc Welding. The GTAW process is widely used for welding stainless steels, especially for full-penetration welds in thin-gage materials and root passes in thicker materials. For almost all gas-tungsten arc welding of stainless steels, thoriated tungsten electrodes (ANSI/AWS A5.12 classification EWTh-2) are preferred, and welding utilizes argon gas shielding with direct current, electrode negative (DCEN) polarity. For heavy sections, sometimes helium or argon-helium mixes can be used to achieve greater heat input or higher welding speeds.

Additions of hydrogen (on the order of 5%) to argon gas shielding have also been used, primarily for austenitic stainless steels, to provide higher heat input and a cleaner weld surface. Hydrogen additions are a possible source of embrittlement for all other types of stainless steels. Both manual and automatic procedures are suitable. In automatic welding, a voltage-controlled head can be used to maintain a constant arc length, and either cold-wire or hot-wire filler-metal additions are used.

Many welds in tubing and light-gage material are made without filler-metal additions, using square butt joints, lap joints, edge joints, corner joints, and the like. Then, the weld metal has nearly the same composition as the base metal. When welding austenitic stainless steels such as type 304 or 316, ferrite is usually desirable in the weld metal, but attempts to measure ferrite in the base metal before welding will generally detect no ferrite. This is not a cause for alarm, because the hot-working operations that were conducted to reduce the original stainless steel ingot to the final form generally cause all ferrite in the base metal to be transformed to austenite. Steelmakers generally prefer to adjust the composition of a stainless steel ingot to contain a small amount (typically 3 to 8 FN) of ferrite as cast, because this improves resistance to the ingot breaking up during the first reductions. Thus, their interest is similar to that of the welding engineer. Although no ferrite may be present after hot working, remelting during welding will return the original as-cast ferrite, which means that ferrite is usually found in the weld metal of autogenous joints in types 304, 316, and similar stainless steel alloys.

The potential for ferrite in an autogenous joint can be assessed from the base-metal certified material test report (CMTR) using the WRC-1988 diagram. To be truly useful for this purpose, the CMTR should include nitrogen. If the CMTR indicates that no ferrite is likely in the autogenous weld, then the addition of a filler metal that will provide ferrite ought to be considered, at least for base metals that normally can be expected to provide some ferrite in near-matching weld-metal compositions.

Filler metals for the GTAW process are classified according to the ANSI/AWS A5.9 specification, and many are listed in Tables 1(a), 1(b), 2(a), 2(b), 3(a), 3(b), 4(a), 4(b), 5(a), and 5(b). Similar compositions in the form of consumable inserts, classified to the ANSI/AWS A5.30 specification, are also available for full-penetration root-pass welding. The consumable inserts provide a little surplus filler metal in order to attain an underside profile that is slightly convex. Consumable inserts can also be purchased with a variety of special shapes that assist in joint alignment, especially for welding pipe.

Pipe and tube welds are often made with a backing gas inside in order to prevent oxidation of the root surface. Although argon backing gas is the safest choice, it is more expensive than nitrogen, which has also been used. Small amounts of nitrogen from backing gas can be picked up in the weld metal, and cases are known where such nitrogen pickup was responsible for loss of ferrite and consequent hot cracking. Nevertheless, many successful welds have been made with nitrogen backing gas.

In the gas-tungsten arc welding of fully austenitic stainless steels, it is often advisable to add filler metal even when it does not appear to be needed. The proper addition of filler metal can be used to obtain a convex bead shape, which has better resistance to hot cracking than does the flat or slightly concave weld shape that often results when no filler metal is added. This can be of critical importance when welding fully austenitic stainless steels without cracking.

The successful gas-tungsten arc welding of nominally austenitic stainless steels (the weld metals of which are designed to contain a little ferrite) can depend on the integrity of the gas shield. Drafts, low gas-flow rates, or excessively high gas-flow rates can cause air (which is about 80% nitrogen) to enter the arc. Nitrogen is then picked up by the weld metal, and a weld, which ought to contain ferrite, can be rendered ferrite-free and sensitive to hot cracking. Proper gas-flow rates, typically 0.3 to 0.8 m^3/h (10 to 30 ft^3/h), and barriers that prevent moving air from reaching the arc area can be used to prevent nitrogen pickup. Special attention to the integrity of both shielding gas and backing gas can be especially important in the successful welding of third-generation ferritic stainless steels, which can be embrittled by traces of nitrogen that are picked up during welding.

Gas-tungsten arc welding can be done with a broad range of welding currents for most applications, with travel speed adjusted to make the proper size of weld for the current chosen. Each individual tungsten electrode size has a corresponding proper current range. These current ranges are not particular to stainless steels. Figure 26 provides some welding procedure information for beginning procedure development in the gas-tungsten arc welding of stainless steels.

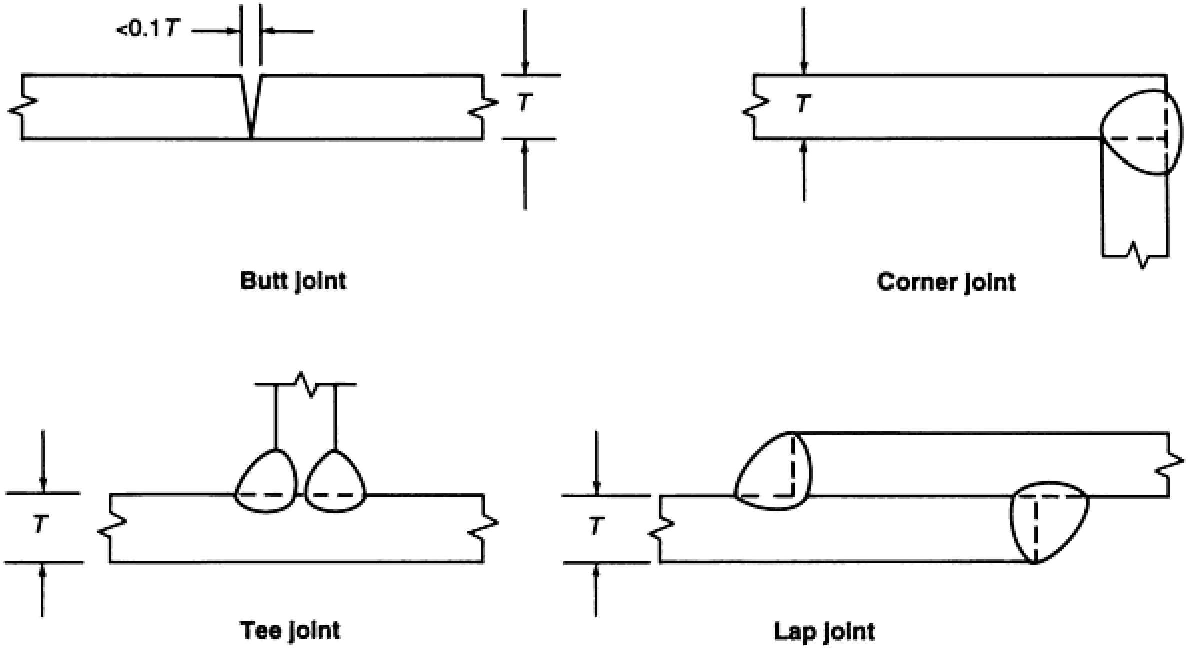

PLATE THICKNESS, *T*		**CURRENT DCEN(A), A**	**ELECTRODE DIAMETER**		**ARGON GAS FLOW**		**FILLER ROD DIAMETER**		**ARC SPEED**		TOTAL TIME OF WELD	
mm	**in.**		**mm**	**in.**	**m^3/h**	**ft^3/h**	**mm**	**in.**	**mm/s**	**in./min**	**h/m**	h/ft
BUTT AND CORNER JOINTS												
1.6	$\frac{1}{16}$	80-100	1.6	$\frac{1}{16}$	0.3	10	1.6	$\frac{1}{16}$	5.1	12	0.0548	**0.0167**
2.4	$\frac{3}{32}$	100-120	1.6	$\frac{1}{16}$	0.3	10	1.6	$\frac{1}{16}$	5.1	12	0.0548	**0.0167**
3.2	$\frac{1}{8}$	120-140	1.6	$\frac{1}{16}$	0.3	10	2.4	$\frac{3}{32}$	5.1	12	0.0548	**0.0167**
4.8	$\frac{3}{16}$	200-250	2.4	$\frac{3}{32}$	0.4	15	3.2	$\frac{1}{8}$	4.2	10	0.0656	**0.0200**
6.4	$\frac{1}{4}$	200-350	3.2	$\frac{1}{8}$	0.56	20	3.2	$\frac{1}{8}$	3.4	8	0.0820	**0.0250**
12.7	$\frac{1}{2}$	225-375	3.2	$\frac{1}{8}$	0.7	25	3.2	$\frac{1}{8}$	3.4	8	0.0820	**0.0250**
TEE AND LAP JOINTS												
1.6	$\frac{1}{16}$	90-110	1.6	$\frac{1}{16}$	0.3	10	1.6	$\frac{1}{16}$	4.2	10	0.0656	**0.0200**
2.4	$\frac{3}{32}$	110-130	1.6	$\frac{1}{16}$	0.3	10	1.6	$\frac{1}{16}$	4.2	10	0.0565	**0.0200**
3.2	$\frac{1}{8}$	130-150	1.6	$\frac{1}{16}$	0.3	10	2.4	$\frac{3}{32}$	4.2	10	0.0656	**0.0200**
4.8	$\frac{3}{16}$	225-275	2.4	$\frac{3}{32}$	0.4	15	3.2	$\frac{1}{8}$	3.4	8	0.0820	**0.0250**
6.4	$\frac{1}{4}$	225-350	3.2	$\frac{1}{8}$	0.56	20	3.2	$\frac{1}{8}$	3.4	8	0.0820	**0.0250**
12.7	$\frac{1}{2}$	**225-375**	**3.2**	$\frac{1}{8}$	**0.7**	**25**	**3.2**	$\frac{1}{8}$	**3.4**	**8**	**0.0820**	**0.0250**

(A) FOR VERTICAL-UP AND OVERHEAD POSITIONS, DECREASE CURRENT 10 TO 20%.

FIG. 26 SUGGESTED PROCEDURES FOR GTAW OF BUTT, CORNER, TEE, AND LAP JOINTS IN STAINLESS STEELS. SOURCE: REF 26

Special attention should be given to the role of sulfur and other surface-active elements in controlling penetration in full-penetration GTAW welds. It has been amply demonstrated that sulfur levels on the order of 0.005% and lower tend to produce penetration profiles with very low depth-to-width ratios, because of surface tension gradient effects. On the other hand, sulfur levels of 0.010% and higher seem to ensure a high depth-to-width ratio, which is desirable for full-penetration butt welds (Ref 28).

When searching for a small amount of filler metal for a GTAW application, especially in repair situations, it is not uncommon for the welder and/or engineer to be tempted to remove the coating from a covered electrode of nominally matching composition to a given base metal, and to use the core wire for GTAW filler metal. This can be a very risky approach, because the coating of many covered stainless steel electrodes provides alloy elements. The core wire may not be close to the weld-metal composition. For example, it is not uncommon to produce type 347 weld metal from a covered electrode whose core wire is type 308, and has all of the niobium in the coating. It is also not uncommon to produce type 317L covered electrodes using type 316L core wire. Some stainless steel electrodes are made from mild-steel core wire, and have all of the alloy in the coating. Therefore, before attempting a gas-tungsten arc weld with the core wire from a covered electrode, those concerned should contact the electrode manufacturer to determine what core wire was used to make the electrode.

Gas-Metal Arc Welding. The GMAW process is extensively used for joining both thin and thick stainless steel base metals. Filler metals are classified to the ANSI/AWS A5.9 specification, as are filler metals for submerged arc or gas-tungsten arc welding. For many applications with nominally austenitic filler metals that are intended to contain a little ferrite, high-silicon versions of the basic classifications are preferred for better wetting and bead shape. These high-silicon versions contain from 0.65 to 1.00% Si, versus 0.30 to 0.65% for the normal grades.

With thin materials, the short-circuiting transfer mode is primarily used, with 0.9 mm (0.035 in.) and 1.1 mm (0.045 in.) diameters being the most common. The preferred shielding gas is often a mixture of 90% He, 7.5% Ar, and 2.5% CO_2. Helium-rich gases provide higher voltage in short-circuiting transfer welding, resulting in better wetting and fewer problems with lack-of-fusion defects than occur with argon-rich gases. However, argon-rich gas may be preferred for welding on very thin materials. Typical welding conditions for short-circuiting transfer welding are given in Table 7 and Fig. 27. Short-circuiting transfer is usable both in the flat position and out of position.

TABLE 7 PROCEDURE RANGE FOR GAS-METAL ARC WELDING WITH ER3XXLSI ELECTRODES BY SHORT-CIRCUIT TRANSFER MODE

CHARACTERISTICS	**WIRE FEED SPEED**		**APPROXIMATE CURRENT, A**	**ARC VOLTAGE, V**[(A)]	DEPOSITION RATE	
	M/MIN	**IN./MIN**			**KG/H**	LB/H
FOR A 0.9 MM (0.035 IN.) DIAM ELECTRODE WEIGHING 5.11 G/M (0.003 LB/FT), USING DCEP AND A 13 MM ($\frac{1}{2}$ IN.) ELECTRICAL STICKOUT, WITH A SHIELDING GAS OF 90HE-7$\frac{1}{2}$AR-2$\frac{1}{2}$$CO_2$	3.0	120	55	19-20	0.9	**2.0**
	3.8	150	75	19-20	1.2	**2.5**
	4.6	180	85	19-20	1.4	**3.0**
	5.2	205	95	19-20	1.6	**3.4**
	5.8	230	105	20-21	1.8	**3.9**
	6.9	275	110	20-21	2.1	**4.6**
	7.6	300	125	20-21	2.3	**5.0**
	8.3	325	130	20-21	2.5	**5.4**
	8.9	350	140	21-22	2.7	**5.9**
	9.5	375	150	21-22	2.9	**6.3**
	10.2	400	160	22-23	3.1	**6.7**
	10.8	425	170	22-23	3.3	**7.1**
FOR A 1.1 MM (0.045 IN.) DIAM ELECTRODE WEIGHING 7.63 G/M (0.0045 LB/FT), USING DCEP AND A 13 MM ($\frac{1}{2}$ IN.) ELECTRICAL STICKOUT, WITH A SHIELDING GAS OF 90HE-7 $\frac{1}{2}$AR-2$\frac{1}{2}$$CO_2$	2.5	100	100	19-20	1.1	**2.8**
	3.2	125	120	19-20	1.5	**3.5**
	3.8	150	135	21	1.7	**4.2**
	4.4	175	140	21	2.0	**4.8**
	5.6	220	170	22	2.6	**6.1**
	6.4	250	175	22-23	2.9	**6.9**
	7.0	**275**	**185**	**22-23**	**3.2**	**7.6**

(A) ARC VOLTAGES ARE MEASURED FROM THE WIRE FEEDER GUN CABLE BLOCK TO WORK.

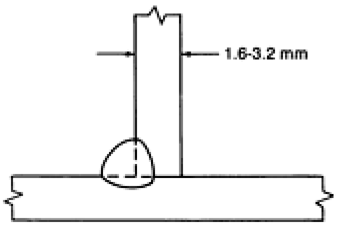

Tee joint

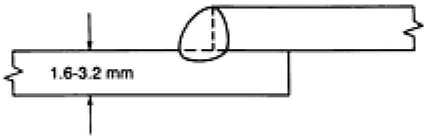

Lap joint

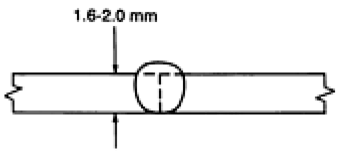

Butt joint

PLATE THICKNESS		**CURRENT (DCEP), A**	**VOLTAGE, V**[(A)]	**WIRE FEED SPEED**		**WELDING SPEED**		**AMOUNT OF ELECTRODE REQUIRED**		TOTAL TIME OF WELD	
mm	**in.**			**mm/s**	**in./min**	**mm/s**	**in./min**	**kg/m**	**lb/ft**	**h/m**	h/ft

TREE AND LAP JOINTS											
1.6	0.063	85	15	78	184	7.2-8.0	17-19	0.037	0.025	0.0364	**0.0111**
2.0	0.078	90	15	81	192	5.5-6.3	13-15	0.051	0.034	0.0469	**0.0143**
2.4	0.093	105	17	98	232	5.9-6.8	14-16	0.058	0.039	0.0436	**0.0133**
3.2	0.125	125	17	119	280	5.9-6.8	14-16	0.069	0.046	0.0436	**0.0133**
BUTT JOINT											
1.6	0.063	85	15	78	184	8.0-8.9	19-21	0.034	0.023	0.0328	**0.0100**
2.0	**0.078**	**90**	**15**	**81**	**192**	**4.9-5.3**	**11.5-12.5**	**0.058**	**0.039**	**0.0548**	**0.0167**

(A) FOR AR-25 CO_2 OR AR-2O; VOLTAGE IS 6 TO 7V HIGHER FOR HE-$7\frac{1}{2}$AR-$2\frac{1}{2}$$CO_2$.

FIG. 27 SUGGESTED PROCEDURES FOR GMAW OF BUTT FILLET AND LAP JOINTS IN 200 AND 300 SERIES STAINLESS STEELS USING THE SHORT-CIRCUITING TRANSFER MODE. SHIELDING GAS, ARGON PLUS 25% CO_2 OR ARGON PLUS 2% O_2; GAS FLOW, 0.4 TO 0.6 M^3/H (15 TO 20 FT^3/H); ELECTRODE, 0.8 MM (0.030 IN.) DIAM. SOURCE: REF 26

With thicker materials, the spray-transfer mode is most often used. In addition to the same two common diameters used in short-circuiting transfer, a 1.6 mm ($\frac{1}{16}$ in.) diameter is often used. The most common shielding gases are argon-rich mixtures, such as argon with from 1% to 5% O, argon with a few percent of carbon dioxide, or argon with both oxygen and carbon dioxide added. Typical welding conditions for spray transfer are given in Table 8 and Fig. 28. Spray-transfer welding is limited to the flat and horizontal positions.

TABLE 8 PROCEDURE RANGE FOR GAS-METAL ARC WELDING WITH ER3XXLSI ELECTRODES BY AXIAL SPRAY TRANSFER

CHARACTERISTICS	**WIRE FEED SPEED**		**APPROXIMATE CURRENT, A**	**ARC VOLTAGE, V$^{(A)}$**	DEPOSITION RATE	
	m/min	**in./min**			**kg/h**	lb/h
FOR A 0.9 MM (0.035 IN.) DIAM ELECTRODE WEIGHING 5.11 G/M (0.003 LB/FT), USING DCEP AND A 13 MM ($\frac{1}{2}$ IN.) ELECTRICAL STICKOUT, WITH A SHIELDING GAS OF 98AR-$2O_2$	10.2	400	180	23	3.1	**6.7**
	10.8	425	190	24	3.3	**7.1**
	11.4	450	200	24	3.5	**7.5**
	12.1	475	210	25	3.7	**8.0**
FOR A 1.1 MM (0.045 IN.) DIAM ELECTRODE WEIGHING 7.63 G/M (0.0045 LB/FT), USING DCEP AND A 19 MM ($\frac{3}{4}$ IN.) ELECTRICAL STICKOUT, WITH A SHIELDING GAS OF 98AR-$2O_2$	6.1	240	195	24	2.8	**6.6**
	6.6	260	230	25	3.0	**7.2**
	7.6	300	240	25	3.5	**8.3**
	8.3	325	250	26	3.8	**9.0**
	9.1	360	260	26	4.2	**10.0**
FOR A 1.6 MM ($\frac{1}{16}$ IN.) DIAM ELECTRODE WEIGHING 16.14 G/M (0.01 LB/FT), USING DCEP AND A 19 MM ($\frac{3}{4}$ IN.) ELECTRICAL STICKOUT, WITH A SHIELDING GAS OF 98AR-$2O_2$	4.4	175	260	26	4.3	**9.2**
	5.1	200	310	29	4.9	**10.5**
	6.4	250	330	29	6.2	**13.1**
	7.0	275	360	31	6.8	**14.4**
	7.6	**300**	**390**	**32**	**7.4**	**15.8**

(A) ARC VOLTAGES ARE MEASURED FROM THE WIRE FEEDER GUN CABLE BLOCK TO WORK.

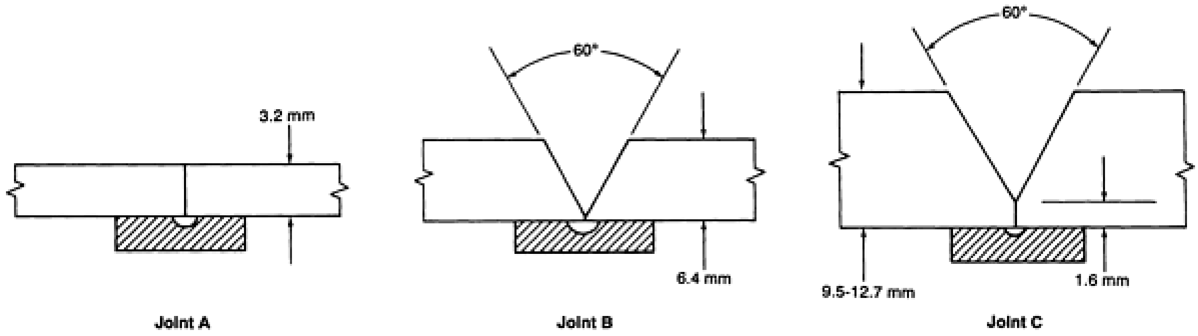

PLATE THICKNESS		ELECTRODE DIAMETER		NUMBER OF PASSES	CURRENT (DCEP), A	WIRE FEED SPEED		ARC SPEED		AMOUNT OF ELECTRODE REQUIRED		TOTAL TIME OF WELD	
mm	in.	mm	in			mm/s	in./min	mm/s	in./min	kg/m	lb/ft	h/m	h/ft
JOINT A													
3.2	$\frac{1}{8}$	1.6	$\frac{1}{16}$	1	225	60	140	8.0-8.9	19-21	0.112	0.075	0.033	**0.010**
JOINT B													
6.4	$\frac{1}{4}$	1.6	$\frac{1}{16}$	2	275	74	175	8.0-8.9	19-21	0.282	0.189	0.066	**0.020**
JOINT C													
9.5	$\frac{3}{8}$	1.6	$\frac{1}{16}$	2	300	85	200	6.3-7.2	15-17	0.405	0.272	0.082	**0.025**
12.7	$\frac{1}{2}$	**2.4**	$\frac{3}{32}$	**4**	**325**	**95**	**225**	**6.3-7.2**	**15-17**	**0.737**	**0.495**	**0.164**	**0.050**

FIG. 28 SUGGESTED PROCEDURES FOR GMAW OF BUTT JOINTS IN 200 AND 300 SERIES STAINLESS STEELS USING THE SPRAY-ARC TRANSFER MODE. SHIELDING GAS, ARGON PLUS 1% O_2; GAS FLOW, 1 M^3/H (35 FT^3/H). SOURCE: REF 26

With special pulsing power sources, a spray-like transfer can be produced at a much lower average welding current, more like that of the short-circuiting transfer mode. Pulsed gas-metal arc welding can permit out-of-position welding. So-called "synergic" power sources are claimed to automatically adjust pulse frequency to produce the transfer of a single drop with each pulse, despite variations in electrical stickout or wire feed speed. The pulsed GMAW process is an area of rapid technological development at the present time.

All gas-metal arc welding conducted with nominally austenitic filler metals that are intended to produce a little ferrite is sensitive to the integrity of the gas shielding, as is the GTAW process. Drafts and gas turbulence can permit air to enter the gas stream, resulting in nitrogen pickup and loss of ferrite. As with gas-tungsten arc welding, proper gas-flow rates (0.8 to 1.1 m^3/h, or 30 to 40 ft^3/h) and the use of barriers to exclude drafts from the arc area are essential to shielding gas integrity.

Flux-Cored Arc Welding. The ANSI/AWS A5.22 specification provides for three different classifications of a given alloy type, based on recommended shielding. Using type 308L as an example, an electrode can be classified as E308LT-1 (carbon dioxide shielding), E308LT-2 (Ar-2O shielding), or E308LT-3 (no gas shielding).

In practice, E308LT-1 (and other alloy wires designed for carbon dioxide gas shielding) are usually used in small diameters, that is, 1.6 mm ($\frac{1}{16}$ in.) and 1.1 mm (0.045 in.). Even 0.9 mm (0.035 in.) diameters are becoming available. Despite the use of carbon dioxide shielding gas, low-carbon-deposit compositions can be produced. Many of these wires, especially in 1.1 mm (0.045 in.) and smaller diameters, are welded in the vertical-up and overhead positions, as well as in the downhand position. Although the wires are classified with carbon dioxide gas shielding, in many cases they weld better in the vertical-up and overhead positions with Ar-25CO_2 shielding. The small-diameter wires are generally made from austenitic stainless steel strip (often type 304L), which work hardens very much during reduction to its final diameter. Thus, these wires are mechanically rather hard and tolerate high feed-roll pressures. Equipment for feeding such wires can be the same as for GMAW solid wires.

It is doubtful that there are any flux-cored stainless steel wires that are correctly classified with Ar-2O. The reason for this statement is that the scope of the ANSI/AWS A5.22 specification requires that a wire classified as flux cored must contain a minimum of 5% nonmetallics, by weight of the electrode. This much slag is not compatible with very high argon gas shielding. Certain metal-cored electrodes seem to be improperly classified as EXXXT-2, but they have almost no nonmetallic ingredients and are more properly classified as composite electrodes, according to the ANSI/AWS A5.9 specification.

In practice, E308LT-3 electrodes (and other stainless steel alloy compositions designed for welding without gas shielding) are commonly available in somewhat larger sizes than are the carbon dioxide classifications. The most popular size is still 2.4 mm ($\frac{3}{32}$ in.), although 2.0 mm ($\frac{5}{64}$ in.) and 1.6 mm ($\frac{1}{16}$ in.) sizes are also popular. These wires are generally usable only in the downhand and horizontal positions. In contrast to the carbon dioxide shielded wires, these wires generally produce rather shallow penetration. This feature can require somewhat higher welder skill levels to make sound x-ray quality welds, because of the dangers of slag entrapment. Higher travel speed is usually necessary when this defect is encountered. On the other hand, this shallow penetration is a major advantage for cladding operations, and these electrodes have been extensively used to achieve fully alloyed low-carbon overlays in a minimum number of layers.

The 2.4 mm (0.10 in.) electrodes, typified by E308LT-3, are generally made from thin, mild-steel strip, with all alloying elements in the core. This makes for a mechanically soft wire that often requires special U-grooved gear drive rolls for the wire feeder. Such drive rolls can grip the wire positively all around its circumference and do not tend to flatten the wire, which results in the best feeding characteristics. Special drive-roll conversion kits are offered for feeding these and other very soft wires. Some 2.0 mm (0.08 in.) diameter wires are also very soft, and similar drive rolls may be necessary.

The flux-cored wires designed for use without gas shielding are virtually immune to any effects of drafts on weld-metal ferrite content. They generally produce a fairly heavy fume plume, which means that fume extracting welding guns or other local exhaust systems can be desirable. Because of their immunity to drafts, these wires lend themselves very well to local exhaust.

On the other hand, the wires designed for use without shielding gas are sensitive to nitrogen pickup that is due to excessive arc length, as are SMAW coated electrodes. A proper balance between wire feed speed (current) and voltage will ensure proper ferrite control with such electrodes (Ref 29).

The wires designed for use with carbon dioxide (or Ar-25CO_2) are sensitive to loss of shielding gas, caused by drafts or inappropriate gas-flow rates, although less so than solid wires in the GMAW process. Protection from drafts is necessary for ferrite control, and local exhaust must be used with care so as not to disturb the gas shield. Otherwise, nitrogen can be picked up by the weld and ferrite can be lost. These wires also appear to have a greater tendency to pick up moisture from humid air if they are carelessly stored for long periods of time. Moisture pickup is often evidenced by "worm track" porosity (surface porosity elongated parallel to the solidification direction). This can often be eliminated by increasing the electrical stickout to preheat the wire, but care must be used to avoid loss of gas shielding.

A simple technique is to recess the contact tip approximately 5 to 10 mm (0.2 to 0.4 in.) behind the end of the shielding gas cup, and to then weld with normal distance between the gas cup and the workpiece. In extreme cases of moisture pickup, the wire may require a bake at 260 to 315 °C (500 to 600 °F) to remove the moisture. In humid weather, it is best to return gas-shielded flux-cored stainless steel wires to their packaging, or to store them in a tight plastic bag or in a low-temperature oven (100 °C, or 212 °F), if the wire is not to be used overnight.

References cited in this section

26. J.M. GERKEN AND D.J. KOTECKI, "STAINLESS STEELS," PUBLICATION NO. M250, THE LINCOLN ELECTRIC COMPANY, 1990
27. "ENGINEERING REPORT PROD/ES533," THE LINCOLN ELECTRIC COMPANY, 1992
28. C.R. HEIPLE AND J.R. ROPER, MECHANISM OF MINOR ELEMENT EFFECT ON GTA FUSION ZONE GEOMETRY, *WELD. J.,* VOL 61 (NO. 4), 1982, P 97S-102S
29. D.J. KOTECKI, WELDING PARAMETER EFFECTS ON OPEN-ARC STAINLESS STEEL WELD METAL FERRITE, *WELD. J.*, VOL 57 (NO. 4), 1978, P 109S-117S

Welding of Stainless Steels

D.J. Kotecki, The Lincoln Electric Company

References

1. A.L. SCHAEFFLER, CONSTITUTION DIAGRAM FOR STAINLESS STEEL WELD METAL, *MET. PROG.,* VOL 56 (NO. 11), 1949, P 680-680B
2. W.T. DELONG, FERRITE IN AUSTENITIC STAINLESS STEEL WELD METAL, *WELD. J.,* VOL 53 (NO. 7), 1974, P 273S-286S
3. T.A. SIEWERT, C.N. MCCOWAN, AND D.L. OLSON, FERRITE NUMBER PREDICTION TO 100 FN IN STAINLESS STEEL WELD METAL, *WELD J.,* VOL 67 (NO. 12), 1988, P 289S-298S
4. C.N. MCCOWAN, T.A. SIEWERT, AND D.L. OLSON, "STAINLESS STEEL WELD METAL: PREDICTION OF FERRITE CONTENT," WELDING RESEARCH COUNCIL BULLETIN 342, WRC, 1989
5. F.B. LAKE, "EFFECT OF CU ON STAINLESS STEEL WELD METAL FERRITE CONTENT," PAPER PRESENTED AT 1990 AMERICAN WELDING SOCIETY ANNUAL MEETING, AS YET UNPUBLISHED
6. D.J. KOTECKI AND T.A. SIEWERT, WRC-1992 CONSTITUTION DIAGRAM FOR STAINLESS STEEL WELD METALS: A MODIFICATION OF THE WRC-1988 DIAGRAM, *WELD. J.,* VOL 71 (NO. 5), 1992, P 171S-178S
7. E.R. SZUMACHOWSKI AND D.J. KOTECKI, MANGANESE EFFECT ON STAINLESS STEEL WELD METAL FERRITE, *WELD. J.,* VOL 71 (NO. 5), 1984, P 156S-161S
8. J.A. SELF, D.K. MATLOCK, AND D.L. OLSON, AN EVALUATION OF AUSTENITIC FE-MN-NI WELD METAL FOR DISSIMILAR METAL JOINING, *WELD. J.,* VOL 63 (NO. 9), 1984, P 282S-288S
9. C.D. LUNDIN, W.T. DELONG, AND D.F. SPOND, FERRITE-FISSURING RELATIONSHIP IN AUSTENITIC STAINLESS STEEL WELD METALS, *WELD. J.,* VOL 54 (NO. 8), 1975, P 241S-246S
10. J.C. LIPPOLD AND W.F. SAVAGE, SOLIDIFICATION OF AUSTENITIC STAINLESS STEEL WELDMENTS: PART III--THE EFFECT OF SOLIDIFICATION BEHAVIOR ON HOT CRACKING SUSCEPTIBILITY, *WELD. J.,* VOL 61 (NO. 12), 1982, P 388S-396S
11. *METALS HANDBOOK,* 8TH ED., VOL 8, AMERICAN SOCIETY FOR METALS, 1973, P 291
12. E. FOLKHARD, *WELDING METALLURGY OF STAINLESS STEELS,* SPRINGER-VERLAG, 1984, P 17
13. M.A. GROSSMANN AND E.C. BAIN, *PRINCIPLES OF HEAT TREATMENT,* 5TH ED., AMERICAN SOCIETY FOR METALS, 1964, P 49
14. *ATLAS OF ISOTHERMAL TRANSFORMATION AND COOLING TRANSFORMATION DIAGRAMS,* AMERICAN SOCIETY FOR METALS, 1977, P 324
15. "CARPENTER STAINLESS STEELS--WORKING DATA," CARPENTER TECHNOLOGY CORPORATION, 1983

16. "ENGINEERING REPORT ES-489," THE LINCOLN ELECTRIC COMPANY, 1991
17. *METALS HANDBOOK,* 8TH ED., VOL 7, AMERICAN SOCIETY FOR METALS, 1972, P 144-145
18. S.A. DAVID, J.M. VITEK, AND T.L. HEBBLE, EFFECT OF RAPID SOLIDIFICATION ON STAINLESS STEEL WELD METAL MICROSTRUCTURES AND ITS IMPLICATIONS ON THE SCHAEFFLER DIAGRAM, *WELD. J.,* VOL 66 (NO. 10), 1987, P 289S-300S
19. E. FOLKHARD, *WELDING METALLURGY OF STAINLESS STEELS,* SPRINGER-VERLAG, 1984, P 105
20. E. FOLKHARD, *WELDING METALLURGY OF STAINLESS STEELS,* SPRINGER-VERLAG, 1984, P 126
21. D.J. KOTECKI, HEAT TREATMENT OF DUPLEX STAINLESS STEEL WELD METALS, *WELD. J.,* VOL 68 (NO. 11), 1989, P 431S-441S
22. C.D. LUNDIN, K. KIKUCHI, AND K.K. KHAN, "PHASE 1 REPORT: MEASUREMENT OF DIFFUSIBLE HYDROGEN CONTENT AND HYDROGEN EFFECTS ON THE CRACKING POTENTIAL OF DUPLEX STAINLESS STEEL WELDMENTS," WELDING RESEARCH COUNCIL, MARCH 1991
23. U. FEKKEN, L. VAN NASSAU, AND M. VERWEY, "HYDROGEN INDUCED CRACKING IN AUSTENITIC/FERRITIC DUPLEX STAINLESS STEEL," PAPER 26, INTERNATIONAL CONFERENCE ON DUPLEX STAINLESS STEEL (THE HAGUE), NEDERLANDS INSTITUUT VOOR LASTECHNIEK, 26-28 OCT 1986
24. L. VAN NASSAU, H. MEELKER, AND J. HILKES, "WELDING DUPLEX AND SUPER-DUPLEX STAINLESS STEELS," DOCUMENT II-C-893-92, INTERNATIONAL INSTITUTE OF WELDING, 1992
25. I. VAROL, W.A. BAESLACK III, AND J.C. LIPPOLD, "FINAL REPORT: WELD CRACKING AND SOLID-STATE PHASE TRANSFORMATION CHARACTERISTICS IN DUPLEX STAINLESS STEELS," WELDING RESEARCH COUNCIL, 30 JUNE 1989
26. J.M. GERKEN AND D.J. KOTECKI, "STAINLESS STEELS," PUBLICATION NO. M250, THE LINCOLN ELECTRIC COMPANY, 1990
27. "ENGINEERING REPORT PROD/ES533," THE LINCOLN ELECTRIC COMPANY, 1992
28. C.R. HEIPLE AND J.R. ROPER, MECHANISM OF MINOR ELEMENT EFFECT ON GTA FUSION ZONE GEOMETRY, *WELD. J.,* VOL 61 (NO. 4), 1982, P 97S-102S
29. D.J. KOTECKI, WELDING PARAMETER EFFECTS ON OPEN-ARC STAINLESS STEEL WELD METAL FERRITE, *WELD. J.*, VOL 57 (NO. 4), 1978, P 109S-117S

Welding of Cast Irons*

Roger A. Bushey, The ESAB Group, Inc.

Introduction

THE FIRST STEP in selecting the type of welding process and consumable to use for welding cast iron is to determine what type or grade of casting needs to be welded. This discussion is not intended to cover all types of cast irons available on the market today (Table 1), but will discuss the major types of cast irons (that is, gray, ductile, malleable, white, and compacted graphite).

TABLE 1 CLASSIFICATION OF CAST IRON BY COMMERCIAL DESIGNATION, MICROSTRUCTURE, AND FRACTURE

COMMERCIAL DESIGNATION	CARBON-RICH PHASE	MATRIX(A)	FRACTURE	FINAL STRUCTURE AFTER
GRAY IRON	LAMELLAR GRAPHITE	P	GRAY	SOLIDIFICATION
DUCTILE IRON	SPHEROIDAL GRAPHITE	F, P, A	SILVER-GRAY	SOLIDIFICATION OR HEAT TREATMENT
COMPACTED GRAPHITE IRON	COMPACTED VERMICULAR GRAPHITE	F, P	GRAY	SOLIDIFICATION
WHITE IRON	FE_3C	P, M	WHITE	SOLIDIFICATION AND HEAT TREATMENT(B)
MOTTLED IRON	LAMELLAR GR + FE_3C	P	MOTTLED	SOLIDIFICATION
MALLEABLE IRON	TEMPER GRAPHITE	F, P	SILVER-GRAY	HEAT TREATMENT
AUSTEMPERED DUCTILE IRON	SPHEROIDAL GRAPHITE	AT	SILVER-GRAY	HEAT TREATMENT

(A) F, FERRITE; P, PEARLITE; A, AUSTENITE; M, MARTENSITE; AT, AUSTEMPERED (BAINITE).

(B) WHITE IRONS ARE NOT USUALLY HEAT TREATED, EXCEPT FOR STRESS RELIEF AND TO CONTINUE AUSTENITE TRANSFORMATION.

Cast iron can be described as an alloy of predominantly iron, carbon, and silicon. Commercially produced irons contain manganese and may be alloyed with nickel, chromium, copper, molybdenum, tin, antimony, vanadium, and various other elements either by themselves or in combinations.

Cast iron is produced by adding excess amounts of carbon to an austenite structure. During solidification, a portion of this carbon separates from the melt as either iron carbide or graphite. The form that the excess carbon takes is determined by the rate of cooling. If the cooling is rapid, the carbon will solidify as iron carbide. If the cooling is slow, the carbon will solidify as graphite. The type of carbon present and its shape will determine the type and, in particular, the properties of the cast iron. These properties are of prime importance when considering the welding procedure to be used. A flowchart illustrating the structure of the common types of commercial cast irons, as well as the processing required to obtain them, is shown in Fig. 1.

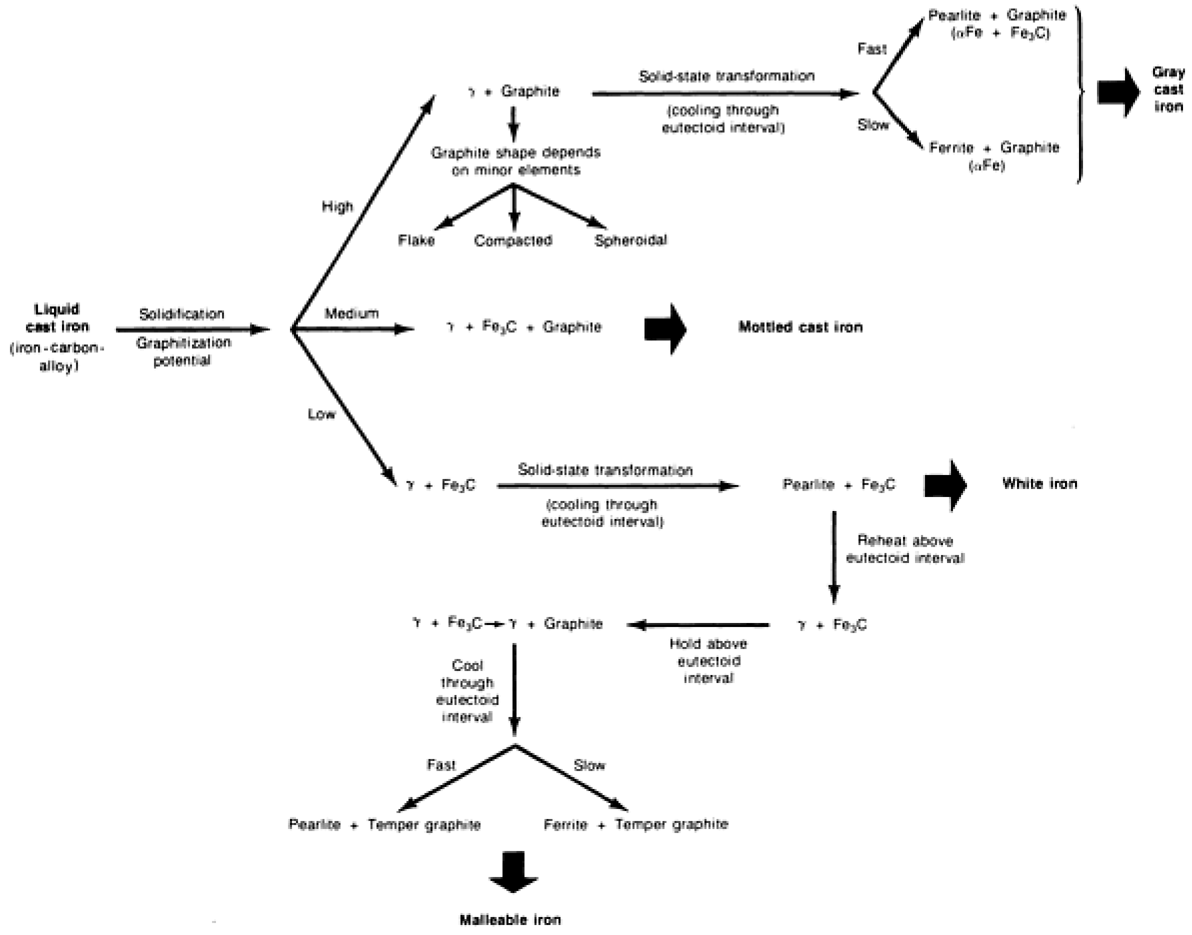
Liquid cast iron (iron-carbon-alloy)
Solidification
Graphitization potential
High
Medium
Low
γ + Graphite
Graphite shape depends on minor elements
Flake
Compacted
Spheroidal
Solid-state transformation
(cooling through eutectoid interval)
Fast
Slow
Pearlite + Graphite (αFe + Fe_3C)
Ferrite + Graphite (αFe)
Gray cast iron
γ + Fe_3C + Graphite
Mottled cast iron
γ + Fe_3C
Solid-state transformation
(cooling through eutectoid interval)
Pearlite + Fe_3C
White iron
Reheat above eutectoid interval
γ + Fe_3C
Hold above eutectoid interval
γ + Fe_3C → γ + Graphite
Cool through eutectoid interval
Fast
Slow
Pearlite + Temper graphite
Ferrite + Temper graphite
Malleable iron

FIG. 1 FLOWCHART SHOWING BASIC MICROSTRUCTURES AND PROCESSING STEPS REQUIRED TO OBTAIN COMMON COMMERCIAL CAST IRONS

Note

* PORTIONS OF THIS ARTICLE WERE ADAPTED FROM "GUIDE FOR WELDING IRON CASTINGS," ANSI/AWS D11.2-89, AMERICAN WELDING SOCIETY, 1989.

Welding of Cast Irons*

Roger A. Bushey, The ESAB Group, Inc.

Classification

Historically, the first classification of cast iron was based on its fracture. Two types of iron were initially recognized:

- *WHITE IRON:* EXHIBITS A WHITE, CRYSTALLINE FRACTURE SURFACE BECAUSE FRACTURE OCCURS ALONG THE IRON CARBIDE PLATES; IT IS THE RESULT OF METASTABLE SOLIDIFICATION (FE_3C EUTECTIC)
- *GRAY IRON:* EXHIBITS A GRAY FRACTURE SURFACE BECAUSE FRACTURE OCCURS ALONG THE GRAPHITE PLATES (FLAKES); IT IS THE RESULT OF STABLE SOLIDIFICATION (GR EUTECTIC)

With the advent of metallography, and as the body of knowledge pertinent to cast iron increased, other classifications based on microstructural features became possible:

- *GRAPHITE SHAPE:* LAMELLAR (FLAKE) GRAPHITE, SPHEROIDAL (NODULAR) GRAPHITE, COMPACTED (VERMICULAR) GRAPHITE, AND TEMPER GRAPHITE; TEMPER GRAPHITE RESULTS FROM A SOLID-STATE REACTION (MALLEABILIZATION)
- *MATRIX:* FERRITIC, PEARLITIC, AUSTENITIC, MARTENSITIC, BAINITIC (AUSTEMPERED)

This classification is seldom used by the floor foundryman. The most widely used terminology can be made in two categories:

- *COMMON CAST IRONS:* FOR GENERAL-PURPOSE APPLICATIONS, THEY ARE UNALLOYED OR LOW-ALLOY
- *SPECIAL CAST IRONS:* FOR SPECIAL APPLICATIONS, GENERALLY HIGH-ALLOY

The correspondence between commercial and microstructural classification, as well as the final processing stage in obtaining common cast irons, is given in Table 1. Special cast irons differ from the common cast irons mainly in the higher content of alloying elements (>3%). Higher alloy content promotes microstructures with special properties for elevated-temperature applications, corrosion resistance, and wear resistance.

Gray Cast Iron

Gray cast iron is the most common form of cast iron. The excess carbon solidifies as flakes of graphite. This type of cast iron has moderate strength, excellent damping and machinability, but poor ductility. It contains 2 to 4% C, 1 to 3% Si, and up to 1% Mn, and has a tensile strength of up to 275 MPa (40 ksi).

The properties and the microstructure of gray iron can be altered with additions of alloying elements. The tensile strength of these alloyed gray cast irons can be as high as 550 MPa (80 ksi). Because of these many variables and different alloying systems, special welding procedures may be required. Because of the alloys added, some gray iron can be considered unweldable.

White Cast Iron

If cooling is rapid, then the excess carbon remains in the metastable form of iron carbide. This material is referred to as white cast iron. Because it contains a large amount of this intermetallic compound, white cast iron is very brittle, greatly reducing its ductility and weldability. The formation of iron carbide does improve hardness and abrasion resistance in these cast irons.

If large amounts of this iron carbide (also referred to as cementite) are present, the material is prone to cracking and is considered unweldable. White cast iron contains 2.5 to 3.8% C, 0.2 to 2.8% Si, and as much as 5.5% Ni, 30% Cr, 6.5% Mo, and 30% Mn, if designed for wear resistance. Tensile strengths can range from 160 to 620 MPa (23 to 90 ksi).

Malleable Cast Iron

White iron, if given the proper heat treatment, can transform into a ductile form of cast iron known as malleable iron. The iron carbides present are dissociated during a high-temperature anneal above 870 °C (1600 °F) for more than 60 h. During this process, irregularly shaped graphite nodules are precipitated and grow in the solid iron matrix. If the atmosphere in the furnace is of an oxidizing nature, some of the carbon will be oxidized, producing a white-heart malleable iron that is considered weldable. If the atmosphere is of a reducing nature, the entire amount of carbon is retained, and the material is a blackheart malleable iron, considered unweldable.

Malleable cast iron contains 2 to 2.8% C and 1 to 1.7% Si. Tensile strengths can range from 310 to 620 MPa (45 to 90 ksi) and elongations from 2 to 20%.

Ductile Cast Iron

Adding nodularizers, such as manganese or rare earths, makes the carbon solidify as small spheroids instead of flakes. Graphite in spheroidal form causes minimal interruption in the continuity of the metallic matrix. This is the most favorable form in which the graphite can be dispersed in order to achieve the desired mechanical properties. Ductile cast irons possess the highest strength and ductility of the cast irons. They combine the principal advantages of gray iron (machinability) with the advantages of steel (high strength, toughness, ductility, hot workability, and hardenability).

Ductile iron (also known as spheroidal graphite or nodular iron) typically contains 3 to 4% C, 1.8 to 2.8% Si, and a maximum of 0.7% Mn. Manganese levels are sometimes restricted if impact properties are important. Higher manganese content in the heat-affected zone during welding can reduce the impact resistance of the weld. Ductile iron has tensile strengths of 415 to 830 MPa (60 to 120 ksi). Alloying additions are sometimes made to ductile irons to improve heat and corrosion resistance. These alloyed grades can contain 2 to 3% C, 1 to 6% Si, 0.7 to 2.4% Mn, 18 to 36% Ni, and up to 5.5% Cr.

The matrix of ductile irons can be varied by heat treatment, alloying, or casting practices. The common as-cast product is 65-45-12. This designation means it has a tensile strength of 65 ksi (450 MPa), a yield strength of 45 ksi (310 MPa), and a tensile elongation of 12%. If this matrix, which has a 10 to 20% pearlite structure, is annealed, the structure becomes ferritic, producing a 60-40-18 iron material. Many different types of higher-strength grades can be produced. In general, as the strength of the iron increases, so do the problems that can be experienced during welding.

Compacted Graphite

Compacted graphite, or vermicular graphite, irons have structures and properties that are in between those of gray irons and ductile irons. The graphite forms interconnected flakes as in a graphite iron, but the flakes are thicker. Compacted graphite iron has better machinability and damping characteristics than ductile iron, but has improved ductility.

Welding of Cast Irons*

Roger A. Bushey, The ESAB Group, Inc.

Weldability

Many microstructural and mechanical properties can be obtained in a cast iron. Therefore, there is a wide range to their weldability. White cast iron is generally considered unweldable, while a ductile iron is generally easier to weld than gray iron.

During welding, a full spectrum of thermal gradients is experienced across the weldment and into the heat-affected zone of the cast iron. This results in many different microstructures and properties. The various microstructures can be classified into different zones and regions, as shown in Fig. 2. The nature and relative size of these zones will be determined by the heat generated during welding, the composition of the iron casting being welded, and the welding consumable used. Each zone will be discussed separately to provide a better understanding of the zones themselves and how welding procedures can affect their characteristics.

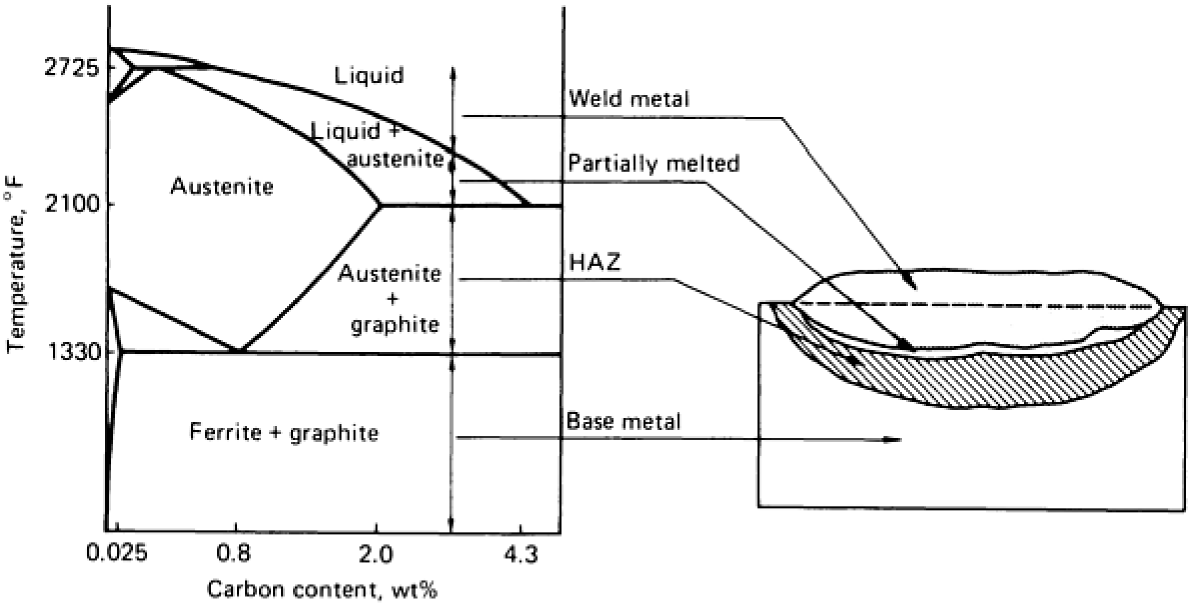

FIG. 2 SCHEMATIC REPRESENTATION OF THE TEMPERATURE ZONES IN A TYPICAL CAST IRON WELDING

Heat-Affected Zone (HAZ). The welding process by its nature is characterized by rapid cooling when compared to a casting procedure. Therefore, the properties of the weld--as well as the sections of the casting exposed to these high temperatures, known as the heat-affected zone--can be different from the rest of the casting.

During welding, the carbon can diffuse into the austenite. When cooling occurs, this austenite transforms to martensite. This martensitic structure is very brittle and susceptible to cracking. The amount of martensite formed depends on the chemistry or type of cast iron and on the thermal history. The brittle martensitic structure can be modified by tempering using preheat and interpass temperature control, by multiple-pass welding (stringer beads), or by a postweld heat treatment.

The partially melted region is an area next to the HAZ where the heat of the weld has been high enough to cause partial melting of the base metal, near the fusion line. This is the most critical zone in the weld, because upon cooling, this area freezes as white iron (that is, when held at high temperature long enough for carbon to dissolve) due to the high cooling rates. The region is very complex and may contain many different types of microstructures. If the amount of graphite dissolved during welding is high enough and forms a continuous molten matrix, it is likely that the carbide network will also be continuous, leading to problems. Because the partially melted region contains a great proportion of hard products, it is the hardest zone of the weld.

The high hardness and resulting low toughness make this region responsible for many of the mechanical problems experienced in the welding of cast iron. The most effective way to reduce the severity of cracking problems is to reduce the peak temperatures and the duration at the high temperatures. This can be controlled by heat input, preheat, interpass temperature, and proper filler metal selection.

The use of a low melting point filler metal will help reduce the peak temperature experienced in the partially remelted region. However, if a weldment is sufficiently small and a high heat input was used, the temperature of the base metal may be raised enough to cause severe fusion line problems even if no preheat was used. If, on the other hand, a high preheat temperature was used to prevent the formation of martensite in the HAZ and reduce the stresses, fusion line cracking may occur even with a low heat input. Some amount of trial-and-error procedure testing may be necessary. A welding procedure that can handle the partially melted region and the HAZ will likely produce a successful weld.

Fusion Zone. The properties and microstructure of the fusion zone are dictated by the welding consumable used. This area is made up of the melted filler metal consumable and some dilution from the iron casting. Because of the turbulence of the weld pool, the weld deposit has a relatively uniform composition. All efforts should be made to minimize the dilution.

Identification of Iron Casting. It is important to determine the type of casting to be welded in order to ensure the best weldment possible. If the type of casting is not known, it will be necessary to try and obtain this information. Records kept on the material purchased may indicate the type or manufacturer of the iron casting. Examination of a small piece under the microscope will indicate its structure. Some of the newer types of cast irons available today are difficult to distinguish from one another; much experience is usually needed. A detailed chemical analysis and hardness level of the cast iron will also help in classifying the material. If all else fails, the casting should be assumed to be a gray cast iron, and caution should be exercised when proceeding.

Weldability of a Casting. One attempt to classify the weldability of a casting is given in Ref 1. The document describes techniques to determine a no-crack temperature based on a carbon equivalent (CE) number. In order to obtain an accurate CE number, the following new CE formula was devised:

$$\begin{gathered} \mathrm{CE} = \mathrm{C} + 0.31\,\mathrm{SI} + 0.33\,\mathrm{P} + 0.45\,\mathrm{S} - 0.028 \\ \mathrm{Mn} + \mathrm{Mo} + \mathrm{Cr} - 0.02\,\mathrm{Ni} - 0.01\,\mathrm{Cu} \end{gathered}$$

The no-crack temperature can then be found for the specific casting under test by using Fig. 3. The procedure was used to show a correlation between the CE of a casting and a no-crack temperature. These data should be used only as a guide to establish a preheat temperature for a casting. Many factors other than preheat affect the ability to produce a sound weld. These include filler metal selection, peening, heat input, local contaminants, trapped sand, and absorbed lubricants in the casting.

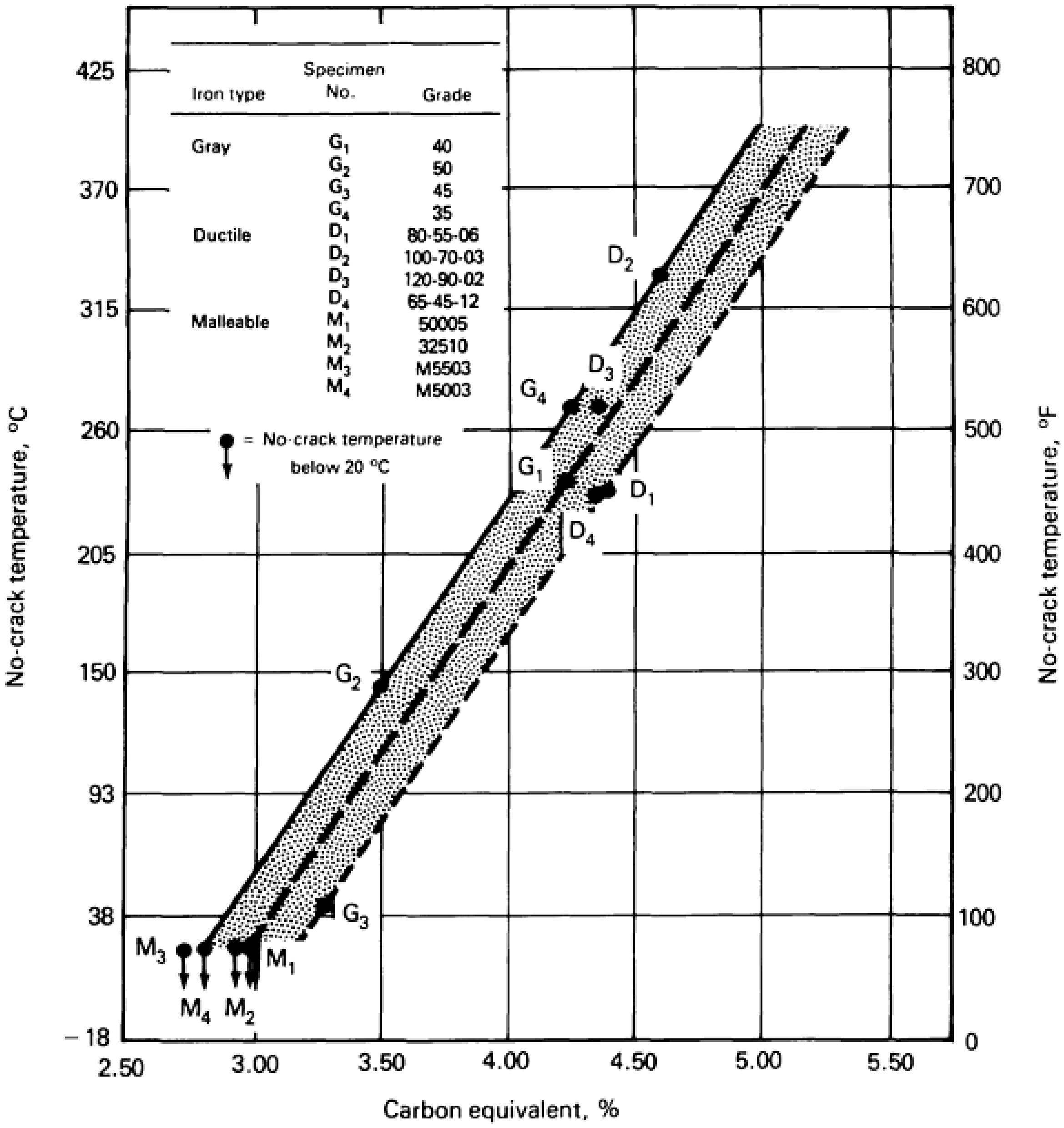

FIG. 3 EFFECT OF CARBON EQUIVALENT ON NO-CRACK TEMPERATURE FOR SELECTED GRADES OF IRON CASTINGS. SOURCE: REF 1

Preheat. One of the most important considerations in welding cast irons is the use of preheat. It:

- PREVENTS CRACKING CAUSED BY TEMPERATURE GRADIENTS AND THERMAL STRESSES
- REDUCES RESIDUAL STRESSES IN THE CASTING
- REDUCES DISTORTION
- REDUCES HARDNESS OF THE HAZ
- REDUCES TEMPERATURE GRADIENTS WHEN WELDING DISSIMILAR BASE METALS
- REDUCES HYDROCARBON CONTAMINANTS IN CASTINGS THAT HAVE BEEN IN SERVICE (CONTAMINANTS ARE BURNED OFF)

How much preheat to give a cast iron is debatable, but the following are some general guidelines:

- THE HIGHER THE CE, THE HIGHER THE PREHEAT TEMPERATURE
- IF THE TYPE OF CASTING IS KNOWN BUT THE SPECIFIC CHEMISTRY IS NOT, THEN THE PREHEAT SHOULD BE BASED ON THE MICROSTRUCTURE AND THE STRENGTH OF THE CASTING
- LOW-STRENGTH CAST IRONS USUALLY REQUIRE LOWER PREHEAT TEMPERATURE THAN HIGH-STRENGTH CAST IRONS
- A CASTING WITH A COMPLEX SHAPE WILL USUALLY REQUIRE A HIGHER PREHEAT TEMPERATURE TO CONTROL DISTORTION OR RESIDUAL STRESSES
- WHEN WELDING THIN METAL TO A THICK BASE METAL, IT IS RECOMMENDED TO PREHEAT THE THICK MEMBER TO DECREASE ITS HEAT SINK CAPACITY AND TO REDUCE THERMAL STRESSES
- MALLEABLE OR DUCTILE CAST IRONS USUALLY REQUIRE LOWER PREHEAT TEMPERATURES THAN GRAY OR WHITE CAST IRON

Preheat is applied to control the microstructure of the weld area. The influence of preheat on the microstructure in a typical cast iron is indicated below:

PREHEAT TEMPERATURE		RESULTING MICROSTRUCTURE
°C	**°F**	
22	72	**MARTENSITE**
100	212	**PEARLITE TRANSFORMATIONS OCCUR.**
200	392	**A GREATER PORTION OF MARTENSITE AND CARBIDES ARE REPLACED BY PEARLITE.**
300	572	**ALMOST ALL MARTENSITE IS REPLACED BY PEARLITE.**
400	**752**	**ALL MARTENSITE IS TRANSFORMED.**

Source: Ref 1

Preheating to temperatures above 315 °C (600 °F) slows the cooling rate to the point that formation of martensite is minimized, reducing the tendency of the cast iron to crack. Slow cooling also helps minimize residual stresses.

Preheating needs to be done properly. Preheat should be applied so that the weld joint is in compression, not tension (Fig. 4). If the proper method is used while the weld metal is contracting during cooling, the base metal will also contract.

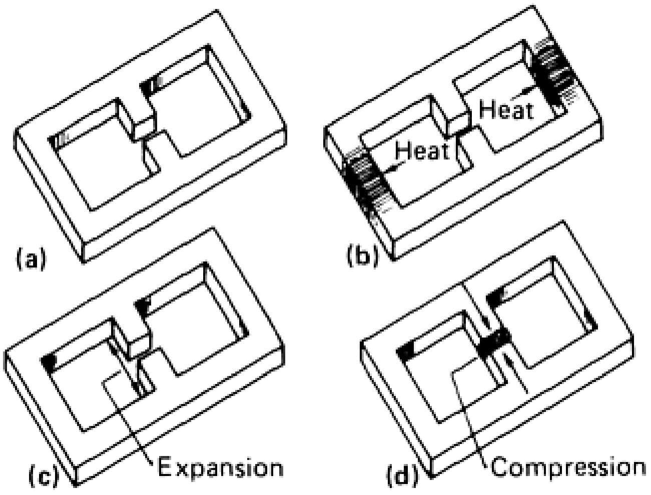

FIG. 4 PROPER METHOD OF PREHEATING CAST IRON CASTINGS TO PREPARE SAMPLES FOR WELDING. (A) CASTING AT ROOM TEMPERATURE. (B) HEAT APPLIED TO PERIMETER OF CASTING. (C) VERTICAL MEMBERS OF CASTING IN TENSION. (D) VERTICAL MEMBERS IN COMPRESSION READY TO BE WELDED

The need to weld complex shapes may necessitate preheating the entire structure. This helps eliminate stresses due to thermal gradients. The area to be welded should be brought to an even higher temperature if necessary. After welding, the entire structure should then be slow cooled to reduce stresses.

Preweld Testing. Cast irons, because they are porous, can collect many contaminants during service. Preheating the casting before welding can help burn out some of these contaminants (for example, grease, oils, or certain chemicals). Another technique uses the electrode selected for the repair. In the area to be repaired, a small stringer bead of 50 to 75 mm (2 to 3 in.) is deposited. The bead is then removed by grinding until the weld bead is back to the original surface of the casting. If a dense porosity-free weld exists, then the possibility is good that the casting contains limited amounts of contaminants. If, however, porosity does exist, then the contaminants must be removed prior to welding. Castings that have excessive amounts of contaminants will exhibit porosity throughout the entire weld. However, if the level of contaminants is lower, then the weld bead will appear to be sound. The only way to determine lack of fusion or porosity due to contaminants is to grind to the base of the weld metal. This procedure can be done without preheat, provided the area tested will not require subsequent machining.

Base-Metal Preparation. The only way to deposit a sound weld on a cast iron is to begin with a sound, clean base metal. When a known crack or defect is to be repaired, it should be completely removed before welding can begin. This also applies when repairing other defects, such as casting tears, porosity, shrinkage defects, and inclusions. The entire defect should be removed before attempting the repair.

Contaminants such as casting skin, coatings, sand, rust, paint, oil, grease, moisture, dirt, and other foreign materials should be removed prior to welding. Castings exposed to high temperatures may have surfaces that are heavily oxidized or carburized, and these surfaces should be removed before attempting a repair.

Baking the casting at temperatures from 370 to 480 °C (700 to 900 °F) will remove most of the organic materials, such as oil and grease. Castings that are impregnated with silt or sand cause additional problems. One method is to apply a weld, then grind, and examine the weld. If the weld metal is not free of porosity, the procedure should be repeated until a sound weld is obtained.

Many of the methods used for removing defects in the castings may make the crack or defect worse unless special procedures are used. One technique is to pin the defect before starting any procedure. This is done by locating the ends of the defect, either visually or with magnetic particle or liquid penetrant testing, followed by drilling holes at this location. The crack or defect will terminate in the hole and not propagate through the casting.

An alternative method involves welding a ductile weld bead perpendicular to the direction of the crack or defect. A nickel or nickel-iron electrode can be used. The defect can then be excavated or removed.

There are several methods of cutting and gouging the iron castings before repair. The air carbon arc process is the most commonly used. Either direct or alternating current can be used with cast irons. The arc length should be kept to a minimum, but the carbon electrode should not be allowed to touch the casting. This will prevent the formation of a high melting point slag on the surface of the casting, which will interfere with cutting. To prevent the formation of large HAZs in the castings, the heat input used for this procedure should be kept to a minimum. The HAZ should be removed by grinding or other mechanical means before welding.

An alternative method is the use of shielded metal arc cutting electrodes. These can be used provided they do not require supplemental compressed air or a special torch. The electrodes have special exothermic flux coatings that concentrate the heat into a very small crater at the tip of the electrode. When the electrode tip is placed in contact with the iron casting, the high concentration of heat from the arc causes the metal to melt.

Once the contaminants and defects are removed, the casting must be prepared for welding. This will require the casting to be machined or ground to achieve a proper joint configuration. The design of the weld groove is important in the casting in order to achieve a sound weld and to provide useful performance. Factors such as base metal thickness, type of base metal, filler metal, service application, and accessibility will determine the design of the groove.

Welds in thin cast irons can be made with single V or U grooves. Standard steel weld groove designs can be used for superficial repairs. However, if a nickel-base filler metal is required, the groove angles need to be increased to allow for the manipulation of the sluggish weld pool. Figure 5 illustrates some of the suggested joint preparations.

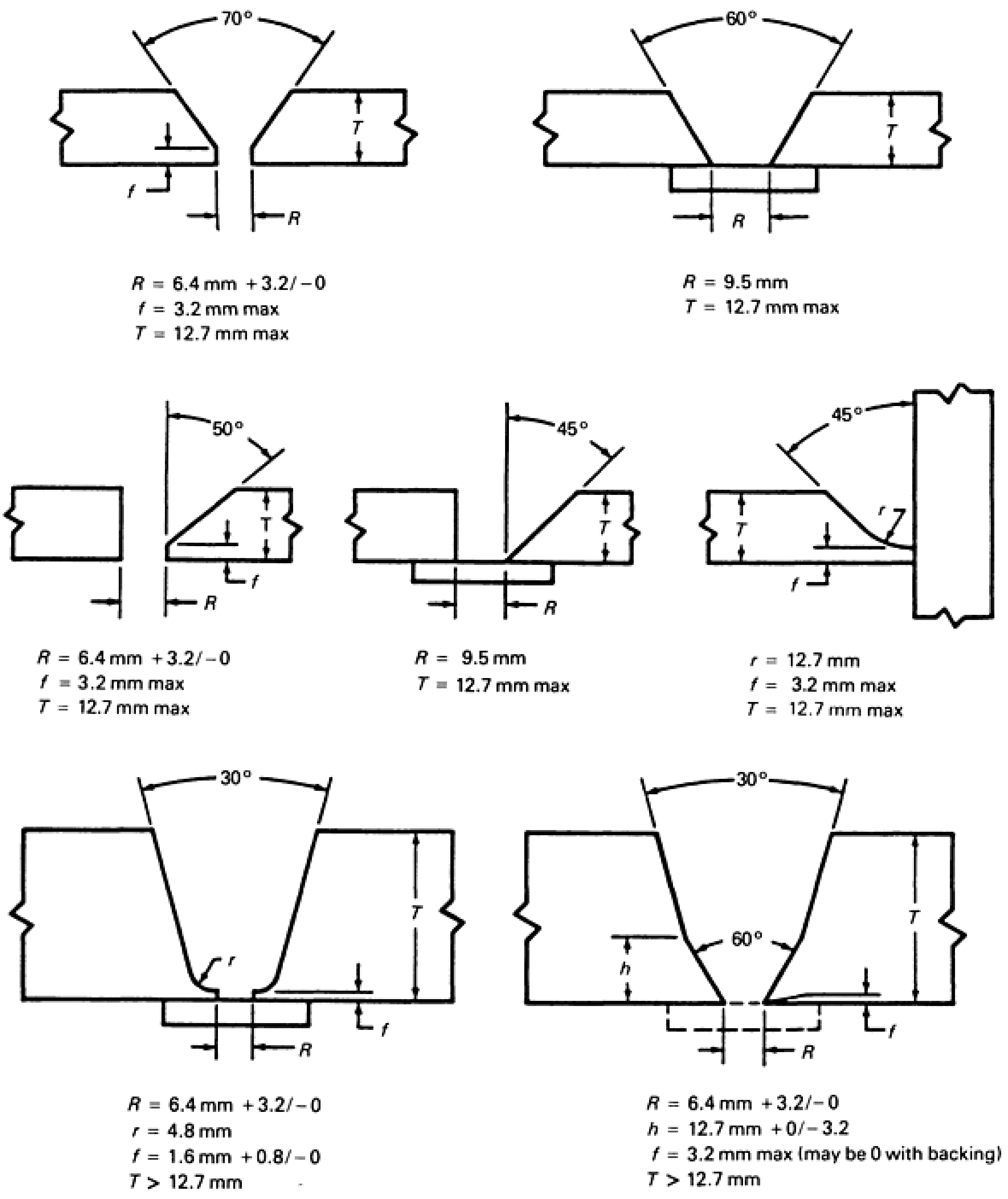
70°
T
f
R
R = 6.4 mm +3.2/−0
f = 3.2 mm max
T = 12.7 mm max
60°
T
R
R = 9.5 mm
T = 12.7 mm max
50°
T
f
R
R = 6.4 mm +3.2/−0
f = 3.2 mm max
T = 12.7 mm max
45°
T
R
R = 9.5 mm
T = 12.7 mm max
45°
r
T
f
r = 12.7 mm
f = 3.2 mm max
T = 12.7 mm max
30°
r
T
f
R
R = 6.4 mm +3.2/−0
r = 4.8 mm
f = 1.6 mm +0.8/−0
T > 12.7 mm
30°
60°
h
T
f
R
R = 6.4 mm +3.2/−0
h = 12.7 mm +0/−3.2
f = 3.2 mm max (may be 0 with backing)
T > 12.7 mm

FIG. 5 RECOMMENDED JOINT PREPARATION FOR WELDS USING SINGLE U OR V GROOVES. SOURCE: REF 1

For castings thicker than 13 mm ($\frac{1}{2}$ in.), the groove must be designed so that the stresses caused by welding can be distributed uniformly in the casting. Double-welded joints are used in such applications when possible. For thicknesses up to 19 mm ($\frac{3}{4}$ in.), a double-V groove or a double-bevel groove is recommended. For thicknesses over 19 mm ($\frac{3}{4}$ in.), a double-U groove or double-J groove is recommended. When only one side of the section is accessible, a modified U groove can be used. This type of joint allows access to the joint root and decreases the width of the weld face, helping to decrease the mass of the weld metal and shrinkage stresses. These joints are illustrated in Fig. 6. When welding thick material, the groove faces are sometimes buttered to help minimize stresses.

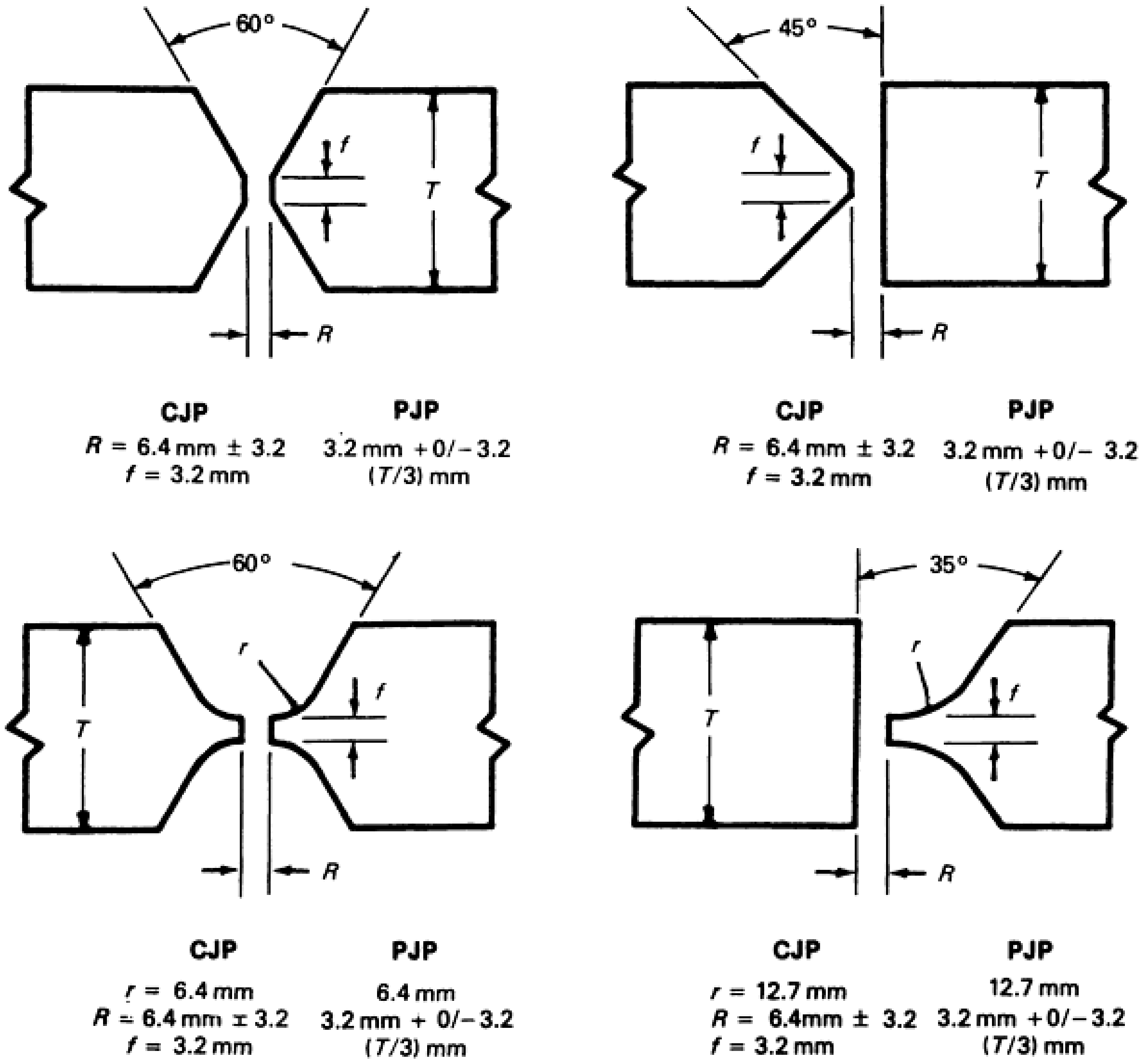

60°
f
T
R
CJP
PJP
R = 6.4 mm ± 3.2
f = 3.2 mm
3.2 mm +0/−3.2
(T/3) mm
45°
f
T
R
CJP
PJP
R = 6.4 mm ± 3.2
f = 3.2 mm
3.2 mm +0/− 3.2
(T/3) mm
60°
r
f
T
R
CJP
PJP
r = 6.4 mm
R = 6.4 mm ± 3.2
f = 3.2 mm
6.4 mm
3.2 mm + 0/−3.2
(T/3) mm
35°
r
f
T
R
CJP
PJP
r = 12.7 mm
R = 6.4mm ± 3.2
f = 3.2 mm
12.7 mm
3.2 mm +0/−3.2
(T/3) mm

FIG. 6 RECOMMENDED DIMENSIONS FOR COMPLETE JOINT PENETRATION (CJP) AND PARTIAL JOINT PENETRATION (PJP) GROOVES. SOURCE: REF 1

The joint penetrations for braze welding a cast iron are essentially the same as those for arc welding, except the edges are rounded (Fig. 7). This facilitates access, flow, and finishing.

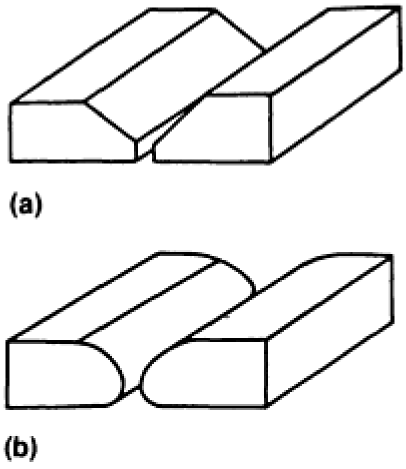

FIG. 7 RECOMMENDED GROOVE CONFIGURATIONS FOR JOINING CAST IRONS. (A) ARC WELDING. (B) BRAZE WELDING. SOURCE: REF 1

The degree of joint preparation required depends on several factors, such as the relative strength of the weld metal, the strength of the base plate, the stresses the casting will see in service, and the thickness of the casting being welded. If there is a question, then a complete joint penetration is preferred. The savings in time and filler metal should be weighed carefully against the conditions in the service environment.

Thick castings of gray iron are sometimes joined with welds having less than complete joint penetration. When nickel-base welding consumables are used, the welds are stronger than the gray cast iron, making partial-penetration welds acceptable. For castings over 13 mm ($\frac{1}{2}$ in.) thick, welding only two-thirds of the thickness should be adequate, and a double-U groove is recommended.

When welding other castings having higher strength levels, such as ductile iron, a complete joint penetration is recommended. Welds to be exposed to fatigue loading, or welds used for pipe applications, should always be complete joint penetration welds.

Special Techniques. The strength of a weld or its fitness for service can be improved by using various special techniques:

- GROOVE FACE GROOVING
- STUDDING
- JOINT DESIGN MODIFICATIONS
- PEENING

Groove Face Grooving. The face of the weld is grooved as shown in Fig. 8. The grooves are then filled with a weld, after which the joint is filled. As a result, the weld interface is irregular so that the path of a crack, if initiated, would not be straight.

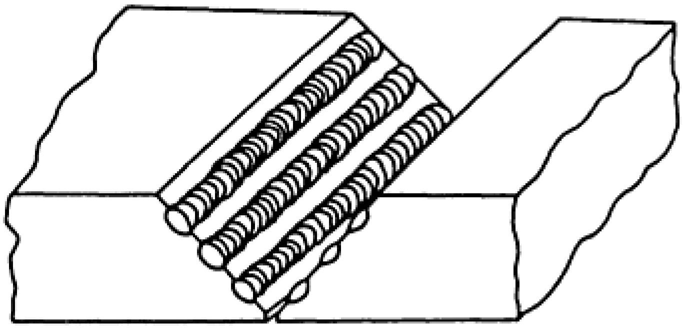

FIG. 8 GROOVE FACE GROOVING FOR WELDING OF CAST IRONS. SOURCE: REF 1

Studding is a mechanical means used to increase the strength and quality of a joint. The studs are either threaded or driven into the joint and must be made of a material compatible with the filler metal used (Fig. 9).

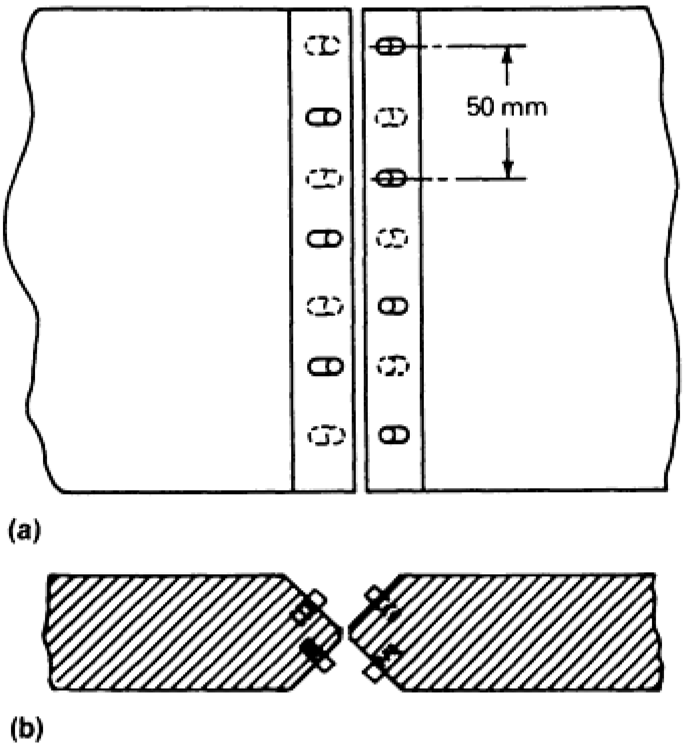

FIG. 9 APPLICATION OF STAGGERED STUD PATTERN TO PROVIDE INCREASED STRENGTH AT WELD INTERFACE. (A) TOP VIEW. (B) FRONT VIEW. SOURCE: REF 1

Joint design modifications in iron castings can improve the service of the welds. Figure 10 shows some samples of these improvements. Figure 10(a) shows a partial joint penetration weld (left) compared with a complete penetration weld (right). Figure 10(b) shows how to improve the casting design to help eliminate problems with changes in metal thickness, so that the weld is not in a highly stressed region. Figure 10(c) shows the addition of a fillet weld to help eliminate the flexing between the two members, again helping to reduce stress concentrations at the weld joint.

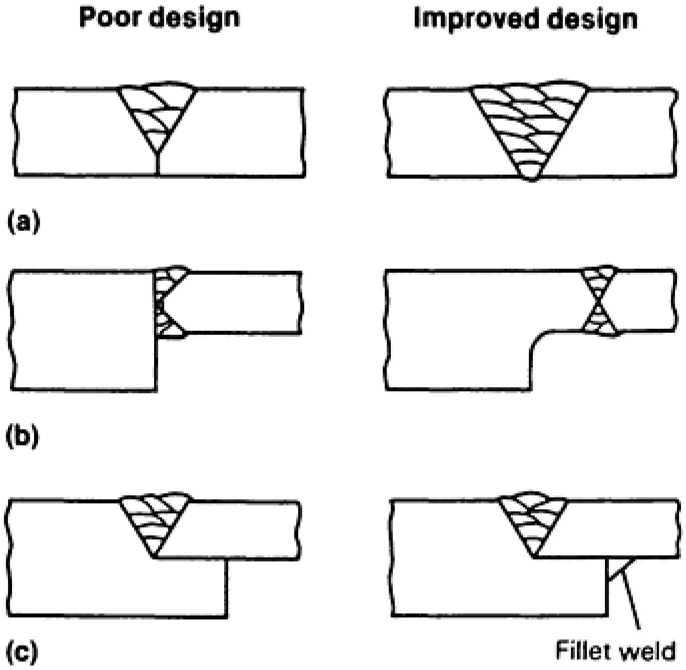

FIG. 10 MODIFICATION OF JOINTS TO PRODUCE OPTIMUM WELD. (A) PARTIAL JOINT PENETRATION WELD (LEFT) CHANGED TO COMPLETE JOINT PENETRATION WELD (RIGHT). (B) PART MODIFIED TO ALLEVIATE PROBLEMS CAUSED BY CHANGES IN METAL THICKNESS. (C) ADDITION OF SECONDARY FILLET WELD TO REDUCE STRESS CONCENTRATIONS OF PRIMARY WELD. SOURCE: REF 1

Peening can help distribute the stresses associated with a weld. A 13 to 19 mm ($\frac{1}{2}$ to $\frac{3}{4}$ in.) rounded ballpeen hammer is used to make repeated, moderate strikes perpendicular to the weld surface. The weld metal should be maintained at or near red hot and not allowed to cool below 540 °C (1000 °F). If a reciprocating air hammer is used, it should be operated at 620 kPa (90 psi) on welds up to 75 mm (3 in.) in length. The tool used should be no wider than the weld bead and should have a tip radius equivalent to one-half the tip width. Travel speed would be 750 to 1000 mm/min (30 to 40 in./min).

Reference cited in this section

1. "GUIDE FOR WELDING IRON CASTINGS," ANSI/AWS D11.2-89, AMERICAN WELDING SOCIETY, 1989

Welding of Cast Irons*

Roger A. Bushey, The ESAB Group, Inc.

Postwelding Treatment

The final stress level of a welded casting is affected by the rate at which the casting is cooled after welding. The casting should be allowed to cool slowly to minimize the residual stresses caused by welding. This can be accomplished by:

- BURYING THE ENTIRE CASTING IN SAND OR VERMICULITE
- POSTHEATING THE CASTING WITH A TORCH
- TRANSFERRING THE CASTING TO A FURNACE
- COVERING THE CASTING WITH AN INSULATING BLANKET

A postweld heat treatment may be necessary to:

- IMPROVE THE DUCTILITY OF THE HAZ
- IMPROVE THE MACHINABILITY OF THE WELD AND HAZ
- TRANSFORM THE MARTENSITIC MICROSTRUCTURE FORMED DURING WELDING TO A LESS BRITTLE PHASE
- RELIEVE RESIDUAL STRESSES IN THE CASTING

The need for postweld heat treatment will depend on the condition of the casting, possible distortion during subsequent machining, the desired finish of the machined surfaces, and prior heat treatment. Thermal stress cracking due to temperature variations in the casting can be avoided by keeping postweld heating and cooling rates below 55 °C/h (100 °F/h) for complex parts; slower rates may be necessary under some circumstances. Table 2 gives the recommended holding temperatures and cooling rates for the postweld heat treatment of gray and ductile cast irons.

TABLE 2 RECOMMENDED POSTWELD HEAT TREATMENT PRACTICE FOR SELECTED GRAPHITIC

CAST IRONS

HEAT TREATMENT	**TEMPERATURE**		**HOLDING TIME (PER WORKPIECE THICKNESS)**	COOLING RATE
	°C	**°F**		
DUCTILE IRON				
STRESS RELIEF	510-565(A)	950-1050(A)	1H/IN.	(F)
	565-595(B)	1050-1100(A)	1H/IN.	(F)
	540-650(C)	1000-1200(C)	1H/IN.	(F)
	620-675(D)	1150-1250(D)	1H/IN.	(F)
FERRITIZE ANNEAL	900-955	1650-1750	1H + 1 H/IN.	(G)
FULL ANNEAL	870-900	1600-1650	1H/IN.	(H)
GRAPHITIZING ANNEAL	. . .	. . .	. . .	. . .
NORMALIZING ANNEAL	. . .	. . .	. . .	. . .
NORMALIZING AND TEMPERING ANNEAL	900-940	1650-1725	2H/IN.(E)	(I)
GRAY IRON				
STRESS RELIEF	595-650	1100-1200	1H/IN.	(F)
FERRITIZE ANNEAL	705-760	1300-1400	1H/IN.	(F)
FULL ANNEAL	790-900	1450-1650	1H/IN.	(F)
GRAPHITIZING ANNEAL	900-955	1650-1750	1-3 H + 1 H/IN.	(F)
NORMALIZING ANNEAL	**870-955**	**1600-1750**	**1-3 H + 1 H/IN.**	(J)

(A) UNALLOYED.

(B) LOW ALLOY.

(C) HIGH ALLOY.

(D) AUSTENITIC.

(E) 2 H MINIMUM.

(F) FURNACE COOL TO 315 °C (600 °F) AT 55 °C/H (100 °F/H), AIR COOL TO ROOM TEMPERATURE (RT).

(G) FURNACE COOL TO 690 °C (1275 °F); HOLD AT 690 °C (1275 °F) FOR 5 H + 1 H/IN. OF THICKNESS, FURNACE COOL TO 345 °C (650 °F) AT 55 °C/H (100 °F/H); AIR COOL TO RT.

(H) FURNACE COOL TO 345 °C (650 °F) AT 55 °C/H (100 °F/H); AIR COOL TO RT.

(I) FAST COOL WITH AIR TO 540 TO 640 °C (1000 TO 1200 °F), FURNACE COOL TO 345 °C (650 °F) AT 55 °C (100 °F/H), AIR COOL TO RT.

(J) AIR COOL FROM ANNEALING TEMPERATURE TO BELOW 480 °C (900 ° F) TO RT.

Welding of Cast Irons*

Roger A. Bushey, The ESAB Group, Inc.

Welding Processes and Consumables

Oxyfuel welding (OFW) is any welding process that uses oxygen and a fuel gas as a heating medium. It involves melting the base plate and filler metal with a welding torch flame. The fuel gas and oxygen are mixed in the proper proportions in a mixing chamber. Molten metal from the groove faces and filler metal intermix and, upon cooling, form a continuous deposit.

The major advantage of the OFW process is the control the welder has over the rate of heat input, the temperature of the weld zone, and the oxidizing and reducing potential of the welding torch flame. The weld bead size and shape, as well as the viscosity of the weld pool, can be controlled because the filler metal is added independently of the welding flame. The OFW process is ideal for welding thin sections, tubes, and small-diameter pipe.

Acetylene is the preferred fuel gas. Other fuel gases (for example, methylacetylene propadiene, propylene, propane, natural gas, and several proprietary gases) are excessively oxidizing and do not develop sufficient heat to properly weld cast irons.

Oxyfuel welding is widely used to repair minor defects in gray cast iron and is used less frequently for ductile iron castings. The slow heating rate of OFW causes a large HAZ to develop but prevents the formation of the brittle martensitic structure.

The consumables for the oxyfuel welding of gray iron are normally cast with higher levels of carbon and silicon than the castings; this compensates for losses experienced during welding. The weld metal has a graphitic structure, making it machinable, and is a good color match. The chemical compositions for three gray iron welding rods are given in Table 3. The RCI filler material is used to join Class 20 through 35 gray irons, and RCI-A is used for gray castings with higher levels of nickel and molybdenum of Class 35 through 45.

TABLE 3 COMPOSITION OF WELDING RODS USED FOR OXYFUEL WELDING OF GRAY IRONS

GRAY IRON WELDING APPLICATION	**AWS FILLER METAL CLASSIFICATION(B)**	**UNS NO.**	COMPOSITION, WT%(A)									
			C	**Mn**	**Si**	**P**	**S**	**Fe**	**Ni(c)**	**Mo**	**Mg**	CE
CONVENTIONAL	RCI	F10090	3.2-3.5	0.60-0.75	2.7-3.0	0.50-0.75	0.10	BAL	TRACE	TRACE	. . .	. . .
HIGH-STRENGTH ALLOY	RCI-A	F10091	3.2-3.5	0.50-0.70	2.0-2.5	0.20-0.40	0.10	BAL	1.2-1.6	0.25-0.45	. . .	. . .
MALLEABLE AND NODULAR	**RCI-B**	**F10092**	**3.2-4.0**	**0.10-0.40**	**3.2-3.8**	**0.05**	**0.015**	**BAL**	**0.50**	**. . .**	**0.04-0.10**	**0.20**

Source: Ref 2

(A) SINGLE VALUES SHOWN ARE MAXIMUM.

(B) COPPER-BASE FILLER METALS FREQUENTLY USED IN THE BRAZE WELDING OF CAST IRONS ARE NO LONGER INCLUDED IN THIS SPECIFICATION. FOR INFORMATION PERTAINING TO THESE MATERIALS SEE ANSI/AWS A7.6.

(C) NICKEL PLUS INCIDENTAL COBALT.

These consumables, along with some proprietary filler metals, have been designed to furnish weld metal having adequate strength and a minimum level of iron carbide. This is achieved by alloying with elements that promote solid solution strengthening in combination with a graphitizing agent. Nickel, manganese, molybdenum, and silicon are the most common. The addition of phosphorus promotes fluidity of the molten iron. If the phosphorus level is high, then a hard, brittle, iron/iron phosphide eutectic can be formed.

The RCI-B filler material given in Table 3, along with additional proprietary filler metals, is designed for welding ductile iron castings. A ductile iron welding rod contains additions of nodularizing elements such as magnesium or cerium. These types of filler metals are capable of producing weld metal with a nodular graphite structure, provided the magnesium or cerium is high enough to compensate for losses incurred during welding. Cerium-containing rods may be the preferred nodularizing element to minimize porosity. Rods containing magnesium can develop a scum during welding unless enough flux is used. This scum is hard to remove and can cause contaminated and low-strength welds.

The filler metals listed in Table 3 have been used for welding malleable cast iron. However, the OFW process can create a wide HAZ of brittle white iron, lowering the ductility of the base metal and making it more prone to cracking. Some foundries repair the rough castings while still having a white iron microstructure before the malleabilizing heat treatment.

Fluxes are used to protect the molten metal from oxidation, to increase the fluidity of the slag formed, and to help dissolve oxides and other impurities. Fluxes used for OFW when utilizing acetylene fuel gas contain sodium borate, or boric acid, soda ash, and some iron oxide powder. Some fluxes are designed specifically for welding ductile iron by lowering the melting point; otherwise, the fluxes are used for both gray and ductile iron.

The weld preparation for OFW must be wider than that for the arc-welding processes to allow for manipulation of the welding rod and torch. Groove angles of 90 to 120° are common.

A preheating temperature of 540 to 650 °C (1000 to 1200 °F) is recommended for the oxyfuel welding of cast irons. This allows for a quick repair and provides slow cooling. If possible, the entire casting should be preheated to minimize local thermal stresses. Gray iron is especially susceptible to cracking due to thermal stresses introduced by localized preheating. If done carefully with a torch, localized preheating can be used successfully on small castings.

After the proper preheating temperature is reached, the repair should be done immediately before the base metal adjacent to the weld area is allowed to cool. The casting should not be allowed to cool below 315 to 425 °C (600 to 800 °F). If the casting cools below this temperature during welding, or if welding is interrupted, the casting should be reheated before welding is resumed. The casting should not be exposed to extreme temperatures during welding to help reduce thermal stresses. During OFW, the area being repaired should not exceed 675 °C (1250 °F). Excessive preheating and welding time may introduce additional stresses, may distort the casting, and may reduce tensile strength.

The temperature of the casting must be monitored during welding. The recommended practice is to preheat the entire casting in a furnace, maintain an interpass temperature during welding, and cool slowly in a furnace. If a permanent furnace is not available, a temporary furnace can be made from sheet metal and fire brick on top of several gas burners.

Several techniques are unique to OFW when welding cast irons:

- THE SIZE OF THE WELDING TORCH SHOULD BE SIMILAR TO THOSE USED FOR MILD STEEL.
- A NEUTRAL FLAME OR SLIGHTLY REDUCING FLAME IS RECOMMENDED. AN OXIDIZING FLAME IS NOT RECOMMENDED, BECAUSE IT CAN CAUSE LOSS OF SILICON, LEADING TO THE FORMATION OF WHITE CAST IRON AND POROSITY.
- AFTER THE CASTING IS AT THE PROPER PREHEAT TEMPERATURE, THE FLUX IS SPRINKLED ONTO THE GROOVE FACES TO BE WELDED. THE CONE OF THE FLAME SHOULD BE KEPT APPROXIMATELY 3.2 TO 6.4 MM ($\frac{1}{8}$ TO $\frac{1}{4}$IN.) FROM THE BASE METAL. CONTACTING THE CONE TO THE BASE METAL SHOULD BE AVOIDED, BECAUSE IT COULD CAUSE OXIDATION OF SILICON AND CARBON, RESULTING IN HARD SPOTS.
- WHEN THE BOTTOM OF THE GROOVE HAS MELTED, THE TORCH FLAME IS MOVED FROM SIDE TO SIDE TO MELT BOTH GROOVE FACES. AT THIS POINT, THE WELDING ROD IS ADDED SLOWLY TO FORM THE WELD POOL. FURTHER ADDITION OF THE ROD SHOULD BE DONE BY MELTING THE ROD IN THE WELD POOL, NOT IN THE FLAME.
- FURTHER ADDITION OF FLUX SHOULD BE DONE BY HEATING THE TIP OF THE WELDING ROD IN THE OUTER PORTION OF THE WELDING FLAME AND DIPPING THE ROD INTO THE FLUX. WHITE SPOTS AND GAS BUBBLES IN THE WELD POOL INDICATE THAT IMPURITIES ARE PRESENT AND THAT ADDITIONAL FLUX IS NEEDED.
- USING A STIRRING ACTION WITH THE WELDING ROD HELPS THE IMPURITIES FLOAT TO THE SURFACE IN THE SLAG AND HELPS MAINTAIN A GOOD FUSION WITH THE GROOVE FACES. EXCESSIVE STIRRING SHOULD BE AVOIDED IN ORDER TO MINIMIZE

OXIDATION.

One recommended practice for depositing the filler metal is called block sequence (Fig. 11). This procedure requires depositing welds approximately 25 mm (1 in.) long in different sections of the joint. The ends of each length should be staggered. Care must be taken to obtain complete fusion between each section and both groove faces. Excess molten metal that might have run ahead of the block must be remelted before welding over it. Another method, called the cascade method, involves depositing thin layers in succession, where each layer is longer than the previous one (Fig. 11). The repaired casting can then be given a postweld heat treatment. Large castings with complex shapes or castings that require excessive machining should be given a postweld stress relief immediately after welding. This requires holding the casting at 540 to 650 °C (1000 to 1200 °F) for 1 h per inch of thickness, followed by slow cooling.

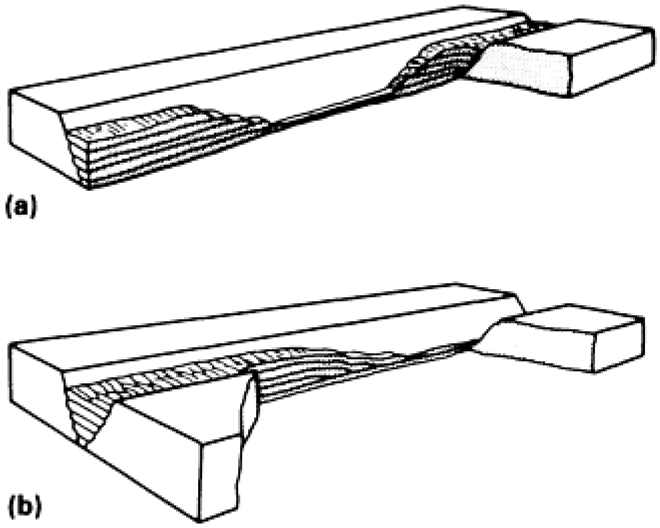

FIG. 11 METHODS APPLIED TO THE OFW PROCESS TO DEPOSIT FILLER METAL WITH MINIMUM WELD STRESS. (A) BLOCK SEQUENCE. (B) CASCADE SEQUENCE

Flame spraying (FLSP) of gray iron is mainly used to repair worn areas, defects, and broken corners on gray iron glass molds. The FLSP process is preferred because deposits as thin as 0.10 to 0.13 mm (0.004 to 0.005 in.) can be applied uniformly to a complex contour of a casting. Finishing time is minimal, and the deposits can be hand filed or machined back to the original contours of the casting. Flame spraying allows a high-strength bond with varying hardness values.

The FLSP process uses a special acetylene torch designed to store and inject powder particles into the gas stream ahead of the flame. These particles then become semimolten as they pass through the flame before striking the casting.

The powders used for FLSP are manufactured by an atomization process and then blended to become an aggregate of various-sized particles. The correct mesh size distribution is necessary in order to obtain an efficient flow of powder through the torch and to maintain reasonable efficiency in thermal spraying deposition.

Various powders have been developed to conform to the Aerospace Material Specification 4779 standard. The standard has the following specifications:

ELEMENT	CONTENT, WT%
CARBON	**0.6 MAX**
SILICON	**4.0-5.0**
IRON	**1.5 MAX**

BORON	**1.0-2.2**
NICKEL	**BAL**

Altering the percentage of each element causes variations in performance and hardness. The hardness range can vary from 165 to 392 HV (85 HRB to 40 HRC).

The area of the casting to be sprayed must be cleaned beforehand. All oxides must be removed. Sand blasting is not recommended, because small particles of silica can become embedded in the casting and cause bonding problems.

The preheat temperature for FLSP is 260 to 315 °C (500 to 600 °F). The preheated area is then sprayed completely with a layer 0.10 to 0.18 mm (0.004 to 0.007 in.) thick. With the torch flame only, the area is then heated to 1010 to 1150 °C (1850 to 2100 °F), starting at one edge until complete fusion is obtained. After the initial deposit has been fused, additional layers can be deposited using the same technique.

Braze welding is a welding process in which the filler metal has a melting point above 450 °C (840 °F) but below the melting point of the casting. Unlike brazing, the filler metal in braze welding is not distributed in the joint by capillary action.

Braze welding was originally developed to repair cracks or broken cast iron parts. Because the process does not require the casting to be melted, the temperature is lower, reducing the formation of brittle cementite, and thus reducing the tendency toward cracking.

The molten filler metal must wet the hot base casting in order for a proper metallic bond to be created between the filler metal and the unmelted casting. This heat is generated with an oxyfuel gas torch.

Flux must be used to properly clean and protect both the base metal and weld metal. The filler metal is fed into the heated groove face in one or more passes until the groove is completely filled. The joint design is similar to joints used for oxyfuel gas welding. Typical groove and joint designs are shown in Fig. 5, 6, 7, 8, 9, and 10.

Advantages. There are some distinct advantages of braze welding over fusion welding:

- BECAUSE LESS HEAT IS REQUIRED, THERE IS LESS DISTORTION AND SUSCEPTIBILITY TO CRACKING.
- THE DEPOSITED WELD METAL IS RELATIVELY SOFT AND DUCTILE, RESULTING IN DEPOSITS THAT HAVE EXCELLENT MACHINABILITY AND LOW RESIDUAL STRESSES.
- JOINTS CAN BE PRODUCED WITH ADEQUATE STRENGTH FOR MANY APPLICATIONS.
- EQUIPMENT REQUIRED IS VERY SIMPLE AND RELATIVELY EASY TO USE.
- CAST IRONS, WHICH ARE VERY SENSITIVE TO TEMPERATURE GRADIENTS OR THERMAL SHOCK, CAN BE BRAZE WELDED WITHOUT EXCESSIVE PREHEATING.

Limitations of braze welding include:

- THE JOINT STRENGTH IS CONTROLLED BY THE WELD METAL OR WELD INTERFACE.
- THE SERVICE TEMPERATURES WILL BE LOWER THAN THOSE IN FUSION WELDS DUE TO THE LOWER MELTING TEMPERATURES OF THE FILLER METAL. IF A COPPER ALLOY IS USED, IT IS LIMITED TO 260 °C (500 °F) OR LOWER.
- BRAZE-WELDED JOINTS WILL NOT MATCH THE COLOR OF THE CASTING.
- DEFECTS ARE HARDER TO LOCATE IN THE BRAZE-WELDED JOINTS, MAKING THEM MORE DIFFICULT TO INSPECT THAN FUSION-WELDED JOINTS.

Equipment and Techniques. The equipment used for braze welding consists of a welding torch and associated components. Braze welding can be used to join castings to other castings or castings to other metals (for example, steel or stainless steel), provided the proper filler metal is used.

A preferred technique of braze welding is to butter one or both faces of the groove prior to welding. The buttered layer can be the same alloy as the filler metal or a different filler metal that will act as a buffer between the two metals to be joined.

The filler metal used for this process is predominantly a 60Cu-40Zn brass alloy. Tin, iron, manganese, silicon, and nickel are added to improve flow characteristics, lower the melting temperature, decrease the volatilization of zinc, react with oxygen, and increase strength and hardness. Table 4 gives the chemical composition and properties of several filler metal rods. Even though the minimum tensile strengths of these alloys range from 275 to 485 MPa (40 to 70 ksi), they may decrease if the service temperature is above 260 °C (500 °F).

TABLE 4 COMPOSITION AND PROPERTIES OF RODS AND ELECTRODES USED FOR BRAZE WELDING OF CAST IRONS (WITH JOINT CONFIGURATIONS SIMILAR TO THOSE APPLIED TO OFW PROCESSING) PER AWS SPECIFICATION A5.27

FILLER METAL TYPE	**COMPOSITION, WT%**											**TENSILE STRENGTH(A)**		**LIQUIDUS TEMPERATURE**		APPLICATIONS
	CU	**ZN**	**SN**	**FE**	**NI**	**MN**	**AL**	**SI**	**PB**	**P**	**TOTAL OTHER**	**MPA**	**KSI**	**°C**	**°F**	
RBCUZN-A	57.0-61.0	BAL	0.25-1.00	. . .	. . .	. . .	0.01	0.04-41.15	0.05	. . .	0.50	275	40	900	1650	**A LOW-FUMING BRONZE USED ON COPPER SHEET, MILD TUBE STEEL, DEEP-DRAWING STEEL, AND CAST IRON**
RBCUZN-B	56.0-60.0	BAL	0.80-1.10	0.25-1.2	0.20-0.80	0.01-0.50	0.01	0.04-0.15	0.05	. . .	0.50	344	50	890	1630	**HIGHER STRENGTH, USED TO SURFACE OR BRAZE WELD COPPER, STEEL, CAST IRON, OR WROUGHT IRON**
RBCUZN-C	56.0-60.0	BAL	0.80-1.10	0.25-1.2	. . .	0.01-0.50	0.01	0.04-0.15	0.05	. . .	0.50	344	50	890	1630	**BETTER COLOR MATCH WITH MILD STEEL, CAST IRON, AND WROUGHT IRON**
RBCUZN-D	**46.0-50.0**	**BAL**	**. . .**	**. . .**	**9.0-11.0**	**. . .**	**0.01**	**0.40-0.25**	**0.05**	**0.25**	**0.50**	**413**	**60**	**935**	**1715**	**NICKEL-SILVER-BRONZE FILLER METAL PROVIDING BEST STRENGTH AND COLOR MATCH WITH CAST IRON AND WROUGHT IRON**

Source: Ref 1

(A) MINIMUM AT ROOM TEMPERATURE.

Fluxes for braze welding are basic fluxes of alkali fluoride and borax mixtures. They are designed to clean the base metal casting and weld metal, and are active over the full melting range of the filler metal. Braze-welding fluxes contain some iron oxide or manganese dioxide, which combines with the free carbon on the cast iron surface to promote wetting. Fluxes are applied by dipping the heated filler metal into the flux or by brushing onto the joint. They are also precoated on filler metal or applied through the oxyfuel gas flame.

The cleanliness of the castings to be braze welded is very important, because contaminants will inhibit the wetting of the filler metal. This wetting is necessary to accomplish the metallic bond between the filler metal and the casting. Both the groove faces and the surfaces adjacent to the joint must be cleaned of all contaminants and foreign materials. This includes graphite smears caused by machining. Graphite smears will adversely affect the wetting of the filler metal. Smears can be removed by quickly heating the cast iron to a dull red color with a slightly oxidizing flame and then wire brushing after cooling. Malleable iron, because of its lower carbon content and rough graphite nodules, can be cleaned by abrasive blasting.

The joints used for braze welding are similar to those used for oxyfuel welding. For thicknesses over 2.4 mm ($\frac{3}{32}$ in.), single and double-V grooves are prepared, with groove angles between 90 and 120°. This allows a large weld-interface area between the base metal and the weld metal. Square grooves can be used for thicknesses of 2.4 mm ($\frac{3}{32}$ in.) or less. In large casting repairs, studs may be inserted into the groove surfaces to increase the strength of the repair (Fig. 9).

The malleable cast iron grades are the easiest to braze weld. Ductile iron is more difficult to braze weld because of its higher total carbon content. The graphite flakes in gray cast iron make it the most difficult to braze weld; special groove preparation will be necessary. The groove surfaces need to be cleaned with caustic or decarburized with an oxidizing oxyfuel flame.

Preheating can be done locally, or the entire part can be heated, depending on the complexity of the part. The part should be preheated to 425 to 480 °C (800 to 900 °F) and slow cooled after welding.

Braze welding uses an oxyfuel gas torch with a slightly oxidizing flame. This is obtained by increasing the oxygen/fuel ratio until the feather disappears and the cone becomes slightly necked. Once the smeared graphite is removed from the groove surfaces, this type of flame is no longer required.

The technique for braze welding is as follows:

- THE BRAZE-WELDING FLUX IS APPLIED TO THE GROOVE FACES AND FILLER METAL TO BE USED.
- THE BASE METAL CASTING IS HEATED UNTIL THE FILLER METAL MELTS, WETS THE CASTING, AND FLOWS ONTO THE GROOVE FACES.
- THE GROOVE IS THEN FILLED WITH ONE OR MORE PASSES. THE CONE OF THE GAS FLAME SHOULD NOT BE DIRECTED ON THE COPPER-ZINC FILLER METAL OR ON IRON-BASE METALS.

References cited in this section

1. "GUIDE FOR WELDING IRON CASTINGS," ANSI/AWS D11.2-89, AMERICAN WELDING SOCIETY, 1989
2. "SPECIFICATION FOR WELDING RODS AND COVERED ELECTRODES FOR CAST IRON," ANSI/AWS A5.15-90, AMERICAN WELDING SOCIETY, 1990

Welding of Cast Irons*

Roger A. Bushey, The ESAB Group, Inc.

Arc-Welding Processes

The arc-welding processes used for welding cast irons are shielded metal arc welding, gas-metal arc welding, flux-cored arc welding, gas-tungsten arc welding, and submerged arc welding.

More than 90% of all industrial welding is done by arc welding. Arc welding has a lower heat input than oxyfuel gas welding because it has a faster welding speed and higher deposition rate. Arc welding may achieve temperatures in excess of 5000 °C (9000 °F) in the arc. Such intense heat allows necessary fusion while heating only a small portion of the weldment. This causes high cooling rates and results in large thermal expansion and contraction stresses. The high cooling rate may also cause more HAZ problems than normally encountered with oxyfuel gas welding. Arc-welding processes are capable of producing welds of high quality when the proper procedure and consumables are used.

Shielded Metal Arc Welding

Shielded metal arc welding (SMAW) is the most widely used process. It uses the heat of the arc to melt the base metal and the tip of the consumable-covered electrode. Protection of the weld pool is achieved by decomposition of the electrode covering.

The electrode and the part to be welded constitute a circuit that also includes the power source, welding leads, and electrode holder. Once an arc is established between the electrode and the base metal, globules of molten metal rapidly form on the tip of the electrode. The molten metal is then transferred through the arc stream into the weld pool. Filler metal is then deposited as this covered electrode is consumed.

The current used can be either an alternating current (ac) or direct current (dc). Any power source capable of supplying the proper range, type and polarity of current, and open circuit voltage may be used.

The advantages of this welding process include a wide range of consumables, good availability, low-cost power sources, and all-position capabilities.

The consumables used for welding cast irons are quite varied. Economic considerations and weld requirements will determine the appropriate electrode for each application.

Cast iron electrodes are covered with a flux to improve arc stability, to reduce porosity, and to add alloying elements. These electrodes produce a composition, mechanical properties, microstructure, and matching color similar to cast iron.

Steel electrodes are made with carbon-steel-core wire and are classified as ESt (Table 5). This classification comes from ANSI/AWS A5.15-90 specification. These steel electrodes are specifically designed for the welding of cast iron, and are different from ordinary mild steel electrodes. The welds deposited by these electrodes are not readily machinable, and it is almost impossible to prevent the formation of hard weld metal with low ductility.

TABLE 5 COMPOSITION OF ELECTRODES FOR THE SHIELDED METAL ARC AND FLUX-CORED ARC WELDING OF CAST IRONS

ELECTRODE PRODUCT FORM	**AWS CLASSIFICATIONS**	**UNS NO.**	COMPOSITION, WT%									
			C	**Mn**	**Si**	**P**	**S**	**Fe**	**Ni**	**Cu**	**Al**	OTHER ELEMENTS (TOTAL)
SHIELDED METAL ARC WELDING ELECTRODES												
UNDILUTED WELD	ENI-CI	W82001	2.0	2.5	4.0	. . .	0.03	8.0	85(A)	2.5	1.0	**1.0**

METAL	ENI-CI-A	W82003	2.0	2.5	4.0	. . .	0.03	8.0	85(A)	2.5	1.0-3.0	**1.0**
	ENIFE-CI	W82002	2.0	2.5	4.0	. . .	0.03	BAL	45-60	2.5	1.0	**1.0**
	ENIFE-CI-A	W82004	2.0	2.5	4.0	. . .	0.03	BAL	45-60	2.5	1.0-3.0	**1.0**
	ENIFEMN-CI	W82006	2.0	10-14	1.0	. . .	0.03	BAL	35-45	2.5	1.0	**1.0**
	ENICU-A	W84001	0.35-0.55	2.3	0.75	. . .	0.025	3.0-6.0	50-60	35-45	. . .	**1.0**
	ENICU-B	W84002	0.35-0.55	2.3	0.75	. . .	0.025	3.0-6.0	60-70	25-35	. . .	**1.0**
CORE WIRE	EST	K01520	0.15	0.60	0.15	0.04	0.04	BAL	. . .	. . .	. . .	**. . .**
FLUX-CORED ARC WELDING ELECTRODES												
UNDILUTED WELD METAL	**ENIFET3-CI**	**W82032**	**2.0**	**3.0-5.0**	**1.0**	**. . .**	**0.03**	**BAL**	**45-60**	**2.5**	**1.0**	**1.0**

(A) MINIMUM

This electrode is normally used to repair small pits, cracks, and other defects when the casting does not require machining. Because the shrinkage of steel is greater than that of cast iron, high tensile stresses develop upon cooling, possibly causing cracking. Groove preparation, as shown in Fig. 5, 6, 7, 8, and studs, as shown in Fig. 9, are recommended for any sizable casting.

Preheating the casting is recommended only when absolutely necessary in order to prevent excessive stresses. Steel electrodes are used at low current to limit the amount of dilution from the casting. The recommended polarity and current for this electrode is dc reverse (electrode positive). The weld beads deposited by these electrodes should be short and widely separated, followed by a light peening after each deposit. After welding, the slag should be completely removed, especially prior to painting.

The hardness level of the weld deposit of a steel electrode depends on dilution of the casting. Small single beads with high dilution have hardnesses as high as 500 HB. Heavier beads with less dilution and multiple passes will have lower hardness levels.

Low-hydrogen electrodes, such as E7016 and E7018, are sometimes used to weld cast irons. They are generally used for cosmetic repair in nonstructural applications where machining is not necessary and cracking is not a major concern. These electrodes are classified in ANSI/AWS A5.1, are readily available at low cost, and are easy to use. The color match of their weld deposit to the casting is good. Because of the dilution of the casting, the weld metal deposit of these electrodes will contain carbon levels of 0.8 to 1.5%. As a result, the deposit will have a high level of hardness and, with rapid cooling, can produce crack-sensitive martensitic microstructures. This can be minimized by using preheat, slow cooling, and postweld annealing.

Stainless steel electrodes are generally not used for welding cast irons. This is due to the formation of chromium carbides during the welding operation, resulting in a crack-prone brittle microstructure. Cracking in the HAZ is likely to occur as a result of differences in tensile strength and coefficient of expansion between the stainless steel weld deposit and the casting.

Nickel-base electrodes are widely used to weld cast irons. Nickel, unlike iron, does not form carbides and has a low solubility for carbon. As the weld metal solidifies and cools, carbon is rejected from solution as graphite. This increases the volume of the weld metal, reducing shrinkage stresses and reducing the likelihood of cracking.

Nickel-base electrodes can be classified into four categories:

- PREDOMINANTLY NICKEL
- 50NI-50FE
- NICKEL-IRON-MANGANESE
- NICKEL COPPER

Predominantly Nickel Electrodes. The electrodes classified as ENi-CI and ENi-CI-A (see Table 5 for composition) produce a weld deposit having a high nickel content; therefore, it is soft and machinable. The "A" designation is used for electrodes to which aluminum has been added to levels from 1 to 3%. These electrodes are used for surfacing applications and where the diluted weld deposit must be machined.

50Ni-50Fe Electrodes. The electrodes classified as ENiFe-CI and ENiFe-CI-A (see Table 5 for composition) have approximately the same level of nickel and iron. These electrodes are different from the high-nickel electrodes in that the:

- WELDS ARE STRONGER AND MORE DUCTILE.
- STRENGTH OF THE ELECTRODES MAKES THEM SUITABLE FOR WELDING DUCTILE AND HIGH-STRENGTH GRAY IRONS.
- ELECTRODES CAN BE USED FOR DISSIMILAR JOINTS, SUCH AS CAST IRON TO CARBON STEEL AND NICKEL-BASE ALLOYS. IF THE DISSIMILAR JOINT USES CAST IRON WELDED TO A CHROMIUM-CONTAINING STEEL SUCH AS STAINLESS STEEL, THEN A BUTTERING LAYER SUCH AS ENI-1 ELECTRODE (ANSI/AWS 5.11) SHOULD BE USED ON THE CAST IRON. THIS WILL HELP PREVENT THE FORMATION OF CHROMIUM CARBIDES IN THE WELD DEPOSIT.
- ELECTRODES CAN BE USED TO WELD GRAY IRONS HAVING HIGH LEVELS OF PHOSPHORUS BECAUSE THE IRON IN THE WELD METAL CAN TOLERATE PHOSPHORUS BETTER THAN NICKEL.

At a chemical composition of 30% Ni and 70% Fe, the nickel-iron alloy has its lowest coefficient of expansion. The weld deposit from ENiFe-CI and ENiFe-CI-A electrodes has approximately this 30Ni-70Fe composition because of base metal dilution. Therefore, these electrodes can be used for thicker sections of casting than the high-nickel electrodes.

Because of the formation of nickel-iron compositions, the weld hardness levels are higher than those made with high-nickel electrodes. Therefore, the use of ENi-CI-A is preferred when machinability is important. In multipass welds, each pass is tempered by each subsequent pass, reducing the hardness and resulting in improved machinability. Typical hardness levels are given in Table 6.

TABLE 6 MECHANICAL PROPERTIES OF WELDS OBTAINED WITH NICKEL-BASE ELECTRODES USED IN SMAW OF CAST IRONS

ELECTRODE	**HARDNESS FOR TYPE OF WELD METAL INDICATED**								**TENSILE STRENGTH**		**YIELD STRENGTH (0.2% OFFSET)**		ELONGATION IN 50 MM (2 IN.), %
	HIGHLY DILUTED BEAD		**MODERATELY DILUTED BEAD**		**MULTIPLE LAYER WELD(A)**								
					WELD FACE		**ROOT SURFACE**						
	HV	**HB**	**HV**	**HB**	**HV**	**HB**	**HV**	**HB**	**MPA**	**KSI**	**MPA**	**KSI**	
ENI-CI	220	209	182	173	. . .	. . .	. . .	. . .	276(B)	40(B)	263(B)	38(B)	**3-6(B)**
ENIFE-CI	364	345	209	199	183	174	187	174	400-579(B)	58-84(B)	296-434(B)	43-63(B)	**6-13(B)**
									449-500(C)	65-72(C)	310-358(C)	45-62(C)	**8-19(C)**
									544(D)	**79(D)**	**420-462(D)**	**61-67(D)**	**6-10(D)**

Source: Ref 1

(A) BUTT JOINT OF 12.7 MM ($\frac{1}{2}$ IN.) PLATE USING 60° V AND 4.8 MM ($\frac{3}{16}$ IN.) ROOT OPENING.

(B) AS-WELDED CONDITION.

(C) ANNEALED AT 900 °C (1650 °F) FOR 4 H, FURNACE COOLED.

(D) ANNEALED AT 845 TO 900 °C (1550 TO 1650 °F) FOR 3 H, FURNACE COOLED TO 690 °C (1275 °F) FOR 5 H, FURNACE COOLED SLOWLY TO 595 °C (1100 °F), AIR COOL.

Nickel-Iron-Manganese Electrodes. When manganese is added to the nickel-iron alloy, it improves strength, ductility, and resistance to cracking. The composition of this electrode is given in Table 5 as ENiFeMn-CI. The strength of this alloy is sufficient for joining nodular irons up to the 80-55-06 grades. It is also used where increased wear resistance is needed on surfacing or buildup applications.

Nickel-copper electrodes are also used for welding cast irons, but because of their sensitivity to iron dilution, they will have limited usage. They are to be used with low-penetration procedures because the dilution with iron may cause weld cracking.

Copper-base electrodes are mainly used for surfacing iron castings on bearing surfaces to improve corrosion and wear resistance, or for joining cast irons to dissimilar metals. As described in the section "Base-Metal Preparation" in this article, base metal cleanliness when welding cast iron is very important, especially prior to using copper-base electrodes.

Copper-base electrodes are classified in ANSI/AWS A5.6-84, and the composition and mechanical property requirements are given in Table 7. The coverings developed for these copper-base electrodes perform satisfactorily on cast iron.

TABLE 7 COMPOSITION AND MECHANICAL PROPERTIES FOR COVERED COPPER ALLOY WELDING ELECTRODES USED IN SMAW APPLICATIONS TO JOIN CAST IRONS

AWS CLASSIFICATION	**COMPOSITION, WT %(A)(B)**										MECHANICAL PROPERTIES			
											HARDNESS, HB	**TENSILE STRENGTH(F)**		ELONGATION IN 50 MM (2 IN.), %(F)
	Cu + Ag	**Sn**	**Mn**	**Fe**	**Si**	**Ni(c)**	**P**	**Al**	**Pb(d)**	**OTHER ELEMENTS TOTAL(E)**		**MPA**	**KSI**	
ECUSN-A	BAL	4.0-6.0	. . .	0.25	. . .	. . .	0.05-0.35	0.01	0.02	0.50	70-85(G)	240	35	**20**
ECUSN-C	BAL	7.0-9.0	. . .	0.25	. . .	. . .	0.05-0.35	0.01	0.02	0.50	85-100(G)	280	40	**20**
ECUA1-A2	BAL	. . .	. . .	0.5-5.0	1.0	. . .	. . .	7.0-9.0	0.02	0.60	130-150(H)	410	60	**20**
ECUMNNIAL	**BAL**	**. . .**	**11.0-13.0**	**2.0-6.0**	**1.5**	**1.0-2.5**	**. . .**	**5.5-7.5**	**0.02**	**0.60**	**160-200(H)**	**520**	**75**	**15**

Source: ANSI/AWS A5.6-84

(A) PER SPECIFICATION A5.6.

(B) SINGLE VALUES ARE MAXIMUM PERCENTAGES.

(C) INCLUDES COBALT.

(D) MUST BE INCLUDED IN DETERMINING THE VALUE FOR TOTAL OTHER ELEMENTS.

(E) MAXIMUM.

(F) MINIMUM.

(G) 4-9 KN (500 KGF) LOAD.

(H) 30 KN (3000 KGF) LOAD.

Phosphor-Bronze Electrodes. The electrodes classified as ECuSn-A and ECuSn-C are referred to as phosphor-bronze because of the residual phosphorus used for deoxidizing the molten metal during their manufacture. These electrodes, because of their chemical makeup, flow sluggishly and require preheat and interpass temperatures of at least 205 °C (400 °F) on thick base metal. Eliminating moisture from the surface of the metals requires that the electrodes be baked at 120 to 150 °C (250 to 300 °F) prior to use.

The ECuSn-C electrodes have a higher tin content than the ECuSn-A electrodes, resulting in a weld deposit with a higher hardness, tensile strength, and yield strength. Hardness levels for the ECuSn-C electrode deposit range from 85 to 100 HB; for the ECuSn-A electrode deposit, from 70 to 85 HB.

Aluminum-bronze electrodes are also used for welding cast irons. The ECuAl-2 electrode is classified in ANSI/AWS A5.6, and the composition and mechanical property requirements are given in Table 7. This electrode has a weld deposit that has a higher tensile strength than the tin-bronze electrodes (minimum of 415 MPa, or 60 ksi) and a hardness of 130 to 150 HB. Aluminum-bronze electrodes also require close attention to the cleanliness of the base plate prior to welding.

The ECuMnNiAl electrode is also listed in Table 7. This electrode is used for the repair of cast irons where a higher tensile strength is needed than that provided by the ECuAl-A2, ECuSn-A, or ECuSn-C electrode deposit. The ECuMnNiAl electrodes vary the level of aluminum and have higher levels of iron, manganese, and nickel than the ECuAl-A2 electrode. This results in higher minimum tensile strengths of 515 MPa (75 ksi) and hardness levels of 160 to 200 HB.

Gas-Metal Arc Welding

Gas-metal arc welding (GMAW) produces coalescence of metals by heating them with an arc between a continuously fed filler metal electrode and the cast iron to be welded. Shielding is obtained from an externally supplied gas.

This process can produce high deposition rates, high heat input, and deep-penetration welds by varying the arc voltage, gas shielding, and wire-feed speeds of the electrode. These electrodes will function in all positions if used at the proper settings.

The transfer of the molten metal from the electrode to the cast iron being welded can occur in three different modes:

- SPRAY
- GLOBULAR
- SHORT CIRCUITING

The mode of transfer is determined by the heat input and the type of shielding gas used.

Spray transfer takes place at the highest heat input of the three modes and is used in the flat position. An inert gas is required, such as argon or helium or a mixture of the two gases. Spray transfer gives the highest deposition rates and the greatest penetration of the three modes of transfer. It is also the most likely to cause cracking due to the high heat input, resulting in the largest HAZ. Additional information is available in the article "Transfer of Heat and Mass to the Base Metal in Gas-Metal Arc Welding" in this Volume.

The globular transfer mode gives a slightly lower heat input, with the metal transferring as large droplets. This lower heat input makes globular transfer less likely to cause cracking and results in welds having lower penetration than with the spray transfer mode. Globular transfer can be used in all positions, but may require a very experienced welder for the vertical and overhead positions.

The short-circuiting transfer mode is the lowest-heat-input and lowest-penetration mode. This is achieved by having the welding electrode actually contact the weld pool many times per second. As a result, short circuiting produces the smallest HAZ and the lowest crack sensitivity of the three transfer modes. Shielding gases for this process include argon, helium, carbon dioxide, or mixtures of these gases. Short circuiting should be used only on thin sections or with high preheat in order to avoid incomplete-fusion and incomplete-penetration defects.

Choosing the Right Transfer Mode. Selection of transfer mode is dictated by the size and complexity of the casting to be welded, the deposition required, and the weld position to be used. Generally, the low-heat-input processes are preferred because they minimize the melting of the cast iron and stresses introduced by arc welding.

Consumable Electrodes. The consumables of this welding process closely parallel those of SMAW. Steel, nickel-base, nickel-iron, nickel-iron-manganese, and copper-base electrodes are used.

Steel Electrodes. A typical application that uses steel electrodes is the repair of casting defects in noncritical areas of the casting that will not be machined. The color match that steel electrodes provide is excellent. Because the weld will be hard due to the increase in carbon level from dilution of the casting in the weld deposit, steel electrodes should not be used where machining is required. Areas to be welded should be preheated, and parameters should be chosen to minimize penetration--for example, short-circuiting transfer and the use of small-diameter electrodes with a carbon dioxide or C-25 (75Ar/25CO_2) shielding gas. The typical welding electrodes used that fall into this category are ER70S-3 and ER70S-6. These electrodes are designed with significant quantities of deoxidizers such as manganese and silicon.

Nickel-base electrodes containing 95% or higher levels of nickel are used to weld cast iron. These electrodes are used when there is highly diluted weld metal and the deposit must be machined. The most common is electrode containing 99% or greater nickel, with no additional deoxidizers. Another electrode that falls into this category is the ERNi-1 electrode classified in ANSI/AWS A5.14. This electrode contains approximately 4% Ti as a deoxidizer, which leads to the formation of titanium carbides, increasing the hardness of the deposit. Table 8 gives recommended shielding gas and welding parameters.

TABLE 8 GMAW PARAMETERS FOR NICKEL-BASE ELECTRODES USED TO WELD CAST IRONS

ELECTRODE DIAMETER		**SHIELDING GAS**	**WIRE-FEED SPEED**		**ARC VOLTAGE, V**	**CURRENT$^{(A)}$, A**	**ELECTRODE EXTENSION**		TRAVEL SPEED	
mm	**in.**		**in./min**	**mm/s**			**in.**	**mm**	**in./min**	mm/s
0.9	0.035	75AR/25CO_2	300	130	24	130	1/2	13	15	**6.5**
			400	170	25	160	1/2	13	15	**6.5**
			500	210	27	200	1/2	13	15	**6.5**
		CO_2	300	130	24	125	1/2	13	15	**6.5**
			400	170	25	155	1/2	13	15	**6.5**
			500	210	26	180	1/2	13	15	**6.5**
			575	245	27	200	1/2	13	15	**6.5**
		ARGON	300	130	26	150	1/2	13	15	**6.5**
			400	170	27	190	1/2	13	15	**6.5**
			500	210	28	220$^{(B)}$	1/2	13	15	**6.5**
			575	245	29	250$^{(B)}$	1/2	13	15	**6.5**
1.2	0.045	75AR/25CO_2	250	105	27	180	5/8	16	15	**6.5**
			300	130	28	190	5/8	16	15	**6.5**

			350	150	29	220	$\frac{5}{8}$	16	17	**7.0**
			400	170	30	240	$\frac{5}{8}$	16	20	**8.5**
			500	210	32	280	$\frac{5}{8}$	16	25	**10.5**
		CO_2	250	105	26	165	$\frac{5}{8}$	16	15	**6.5**
			300	130	27	180	$\frac{5}{8}$	16	15	**6.5**
			350	150	28	220	$\frac{5}{8}$	16	17	**7.0**
			400	170	29	230	$\frac{5}{8}$	16	20	**8.5**
			500	210	31	270	$\frac{5}{8}$	16	25	**10.5**
		ARGON	250	105	27	190	$\frac{5}{8}$	16	15	**6.5**
			300	130	28	220	$\frac{5}{8}$	16	15	**6.5**
			350	150	29	$240^{(B)}$	$\frac{5}{8}$	16	17	**7.0**
			400	170	30	$275^{(B)}$	$\frac{5}{8}$	16	20	**8.5**
			500	210	32	$310^{(B)}$	$\frac{5}{8}$	16	25	**10.5**
1.6	0.062	$75AR/25CO_2$	150	65	27	235	$\frac{5}{8}$	16	17	**7.0**
			175	75	28	235	$\frac{5}{8}$	16	17	**7.0**
			200	85	29	285	$\frac{5}{8}$	16	20	**8.5**
			225	95	30	320	$\frac{5}{8}$	16	25	**10.5**
		CO_2	150	65	27	220	$\frac{5}{8}$	16	17	**7.0**
			175	75	28	245	$\frac{5}{8}$	16	17	**7.0**
			200	85	29	265	$\frac{5}{8}$	16	20	**8.5**
			225	95	30	285	$\frac{5}{8}$	16	25	**10.5**
		ARGON	150	65	27	240	$\frac{5}{8}$	16	17	**7.0**
			175	75	28	265	$\frac{5}{8}$	16	17	**7.0**
			200	85	29	$300^{(B)}$	$\frac{5}{8}$	16	20	**8.5**
			225	**95**	**30**	**$350^{(B)}$**	$\frac{5}{8}$	**16**	**25**	**10.5**

Source: Ref 1

(A) POLARITY, DCEP.
(B) SPRAY TRANSFER.

Nickel-iron electrodes are also used for the GMAW process. They will deposit a harder weld metal and, therefore, are more sensitive to cracking than the high-nickel electrodes.

Nickel-Iron-Manganese Electrodes. As stated in the section "Shielded Metal Arc Welding" of this article, the addition of manganese increases the crack resistance and strength of the weld deposit. As a result, this consumable is suitable for welding the higher-strength grades of ductile iron, such as 65-45-12 and 80-55-06. The typical hardness level of the weld deposit is 80 to 95 HRB. The composition of this electrode is 12% Mn, 44% Ni, and the remainder iron.

Copper-Base Electrodes. Electrodes containing predominantly copper can be used to join or surface cast iron. The electrodes are classified as ERCuSn-A, ERCuAl-A2, and ERMnNiAl in ANSI/AWS A5.7. Table 9 gives the chemical composition and minimum mechanical properties of these electrodes.

TABLE 9 COMPOSITION AND MECHANICAL PROPERTIES OF COPPER-BASE ELECTRODES USED TO JOIN OR SURFACE CAST IRON IN GMAW APPLICATIONS

AWS CLASSIFICATION	**COMPOSITION, WT%$^{(A)}$**											MECHANICAL PROPERTIES		
	CU + AG	**ZN**	**SN**	**MN**	**FE**	**SI**	**NI + CO**	**P**	**AL**	**PB**	**OTHER ELEMENTS (TOTAL)$^{(B)}$**	**HARDNESS, HB**	TENSILE STRENGTH$^{(C)}$	
													MPA	KSI
ECUSN-A	93.5$^{(C)}$	$^{(D)}$	4.0-6.0	$^{(D)}$	$^{(D)}$	$^{(D)}$	$^{(D)}$	0.10-0.35	0.01$^{(D)}$	0.02$^{(D)}$	0.50	70-85$^{(E)}$	240	**35**
ERCUA1-A2	BAL	0.02	. . .	. . .	1.5	0.10	. . .	. . .	9.0-11.0	0.02$^{(D)}$	0.50	130-150$^{(F)}$	415	**60**
ERCUMNNIA 1	**BAL**	**0.15**	**. . .**	**11.0-14.0**	**2.0-4.0**	**0.10**	**1.5-3.0**	$^{(D)}$	**7.0-8.5**	**0.02$^{(D)}$**	**0.50**	**160-200$^{(F)}$**	**515**	**75**

Source: Ref 1

(A) SINGLE VALUES ARE MAXIMUMS.

(B) VALUES, INCLUDING ELEMENTS MARKED WITH (D), SHALL NOT EXCEED VALUE SPECIFIED.

(C) MINIMUM.

(D) TOTAL OF THESE ELEMENTS CANNOT EXCEED THOSE IN (B).

(E) 4.9 KN (500 KGF) LOAD.

(F) 30 KN (3000 KGF) LOAD.

Flux-Cored Arc Welding

Flux-cored arc welding (FCAW) produces coalescence of metals by heating them with an arc between a continuous filler metal electrode and the workpiece. The equipment used for GMAW can also be used for FCAW. Shielding may be provided by the decomposition of the flux fill (self-shielding) or as an external shielding gas (gas-shielded).

Core Composition. The electrode core contains slag formers, arc stabilizers, deoxidizers, and metal alloys. Depending on the makeup and the shielding gas, the electrode may achieve a deep- or shallow-penetration weld.

Electrode Types. The different types of FCAW electrodes include steel, iron-nickel, and nickel-iron-manganese.

Steel electrodes are classified in ANSI/AWS A5.20. They are classified based on type of slag system and gas-shielding requirements. The gas-shielded types EXXT-1 and EXXT-2 are futile-or acid-base, and EXXT-5 is basic. The EXXT-5 electrodes are preferred for welding cast iron, because they have lower penetration, resulting in less dilution and therefore greater crack resistance.

The EXXT-4, EXXT-7, and EXXT-8 electrodes are steel electrodes that are classified as self-shielding or require no external shielding gas. These electrodes can be used to weld cast iron.

The weld deposit of these steel flux-cored electrodes will become hardenable steel with the pickup of carbon from the cast iron. This will make these deposits hard to machine and susceptible to cracking.

Iron-Nickel Electrodes. Flux-cored electrodes having a 70/30 iron-nickel ratio are used to repair heavy cast iron vessels. These heavy castings include the pouring ladles and molds in foundries and steel plants. The high hardness of the weld metal is not detrimental, because all it is required to do in service is to fill in voids and maintain the proper contours.

Nickel-iron-manganese electrodes are flux-cored electrodes with the nickel-iron-manganese matrix described previously. The typical composition is 1% C, 5% Mn, 40% Fe, and the balance nickel. The addition of manganese improves the strength, ductility, and machinability of the weld deposit. These electrodes are used for upgrading or repairing castings, fabrication, surfacing, and welding dissimilar metals. Typical welding conditions for the nickel-iron-manganese electrodes are given in Table 10.

TABLE 10 TYPICAL WELDING PARAMETERS FOR FCAW OF CAST IRONS WITH NICKEL-IRON-MANGANESE ELECTRODES

ELECTRODE DIAMETER		**WELD TYPE**	**AWS POSITION SPECIFICATION**	**WIRE FEED SPEED**		**CURRENT**		VOLTAGE, V
MM	**IN.**			**MM/S**	**IN./MIN**	**A**	**POLARITY**	
2.0	$\frac{5}{64}$	FLAT BUTT OR FILLET	1G,F	101	240	300	DCEP	**30**
2.0	$\frac{5}{64}$	VERTICAL BUTT OR FILLET	3G,F	55	130	180	DCEP	**18-25**
2.4	$\frac{3}{32}$	FLAT BUTT OR FILLET	1G,F	90	215	320	DCEP	**32**
2.4	$\frac{3}{32}$	FLAT OVERLAY$^{(A)}$	. . .					
		FIRST LAYER	. . .	88	210	250	DCEN	**29-33**
		REMAINING LAYERS	**. . .**	**113**	**270**	**350**	**DCEP**	**28-30**

(A) OSCILLATION RECOMMENDED: 19 TO 25 MM ($\frac{3}{4}$ TO 1 IN.) WIDTH, 26 TO 40 CYCLES PER MINUTE. OVERLAP PREVIOUS BEAD 6.4 TO 9.6 MM ($\frac{1}{4}$ TO $\frac{3}{8}$ IN.).

Gas-Tungsten Arc Welding

Gas-tungsten arc welding (GTAW) produces the coalescence of metals by heating them with an arc between a tungsten electrode (nonconsumable) and the workpiece. Shielding is obtained from an inert gas such as argon or gas mixtures. This process can make autogenous welds or welds using a filler metal. The filler metals would be the same as those used for GMAW; however, the ENi-1 electrode containing titanium is not recommended.

The GTAW process is used for the welding of cast iron utilizing nickel or nickel-base filler metal with an argon shielding gas. The procedure for GTAW welding should minimize dilution of the weld by the base metal.

Again, it is of utmost importance to make sure that the groove surface to be welded is free of contaminants. The joint design must be wide enough to facilitate manipulation of the torch and retention of the external shielding gas.

Submerged Arc Welding

Submerged arc welding (SAW) produces the coalescence of metals by heating them with an arc or multiple arcs between a bare metal electrode(s) and the workpiece. The arc, as well as the molten metal, is protected by a granular flux. This welding process is capable of producing welds only in the flat position.

The electrode used for SAW can be either solid or flux-cored. The SAW process is designed and used for high-deposition welding. The equipment used is similar to that for GMAW. The flux can be added manually or by some automatic feeding system.

The only filler metals used for SAW are nickel-base materials. Further alloy additions can be made by variations in the flux, by using powdered metals, or by adding welding wire. The nickel-iron-manganese electrode can be used for groove, fillet, or surfacing applications. This alloy is more resistant to cracking and therefore is more widely used than other nickel-base electrodes.

The selection of the flux is also very important for ensuring a quality weld. The flux must be matched to the filler metal to be used, alloy requirements, strength requirements, and multiple-pass requirements. Welding conditions for SAW are given in Table 11.

TABLE 11 RECOMMENDED PARAMETERS FOR SAW OF CAST IRONS WITH NICKEL-BASE ELECTRODES

WELD TYPE	**CURRENT**	**VOLTAGE,V**	**ELECTRODE EXTENSION**		**TRAVEL SPEED**		OSCILLATION		
							WIDTH		FREQUENCY, CYCLES/MIN
			mm	**in.**	**mm/s**	**in./min**	**mm**	**in.**	
OVERLAY	250	32-34	25	1	2	$4\frac{1}{2}$	29	$1\frac{1}{8}$	**60**
GROOVE/FILLET	**250**	**32**	**25**	**1**	**4.3**	**10**	**. . .**	**. . .**	**. . .**

Source: Ref 1

Surfacing and Overlaying

Surfacing processes are used to restore surfaces or to improve the surface properties of iron castings damaged by corrosion, spalling, or wear. Any process used for the fabrication, maintenance, or repair of iron castings can be used successfully for surfacing.

Materials for surfacing can be any material normally used for the brazing or welding of iron castings. Ceramic powders and hardfacing alloys not normally associated with cast iron are gaining wide acceptance.

Oxyfuel gas welding was first used for buildup. The total heat input to the base metal is high, producing a large HAZ; however, the cooling rate is quite slow, reducing the residual stresses and the formation of hard martensitic microstructures. This process allows good control of the metal size, shape, and thickness, thus requiring less expensive equipment. However, it is a low-deposition process.

The arc-welding processes are the most commonly used for surfacing cast irons. Process selection depends on the application, grade or type of cast iron, degree of dilution allowable, thickness of the deposit desired, and the intended service.

Thermal spraying processes are employed for surfacing to improve wear or corrosion properties, to salvage mismachined parts, and to improve the service life of a cast iron.

Flame spraying uses the heat from a fuel gas and oxygen to melt surfacing material in the form of cord, powder, rod, or wire. Compressed air is used to propel the melted particles onto the cast iron to be surfaced. The fuel gas usually used for this process is acetylene. Flame spraying can produce bond strengths of 17 to 41 MPa (2.5 to 6.0 ksi), depending on the surfacing material used. Bond strengths of 70 MPa (10 ksi) can be obtained by using special techniques, such as high-energy compressed gas with detonation gun systems. Powders containing boron and silicon can be fused after spraying, producing a homogeneous metallurgical structure with the substrate.

Arc spraying uses the heat produced between two electrically conductive wires. Compressed air is then used to atomize and propel the molten surfacing metal particles to the surface of cast iron. Arc spraying has the highest deposition rate of all the surfacing processes. Bond strengths of 35 to 70 MPa (5 to 10 ksi) are possible, but strength depends on the surfacing metal used.

Plasma spraying involves the use of a water-cooled gun containing a tungsten cathode surrounded by a copper anode nozzle. The arc generated in the gun produces a plasma, which is forced through a small orifice in the copper nozzle by an ionized shielding gas. An additional envelope of inert or mixed carrier gas is added to the stream as it passes out of the gun. The metal or ceramic powders used for surfacing cast iron are pressure fed into the gas stream, melted, then propelled to the cast iron. A clean surface is of utmost importance for achieving a good bond between the thermal spray deposit and the cast iron substrate being surfaced.

Additional processes, such as laser and electron beam, can be used to melt surfacing material to produce layers of deposited metal on cast iron. The surfacing material can take many forms, such as powder, rod, wire, or strip, as long as they can be fed into the beam. Laser remelting is used on thermal spray deposits when a denser deposit with a metallic bond is required.

Surfacing Materials. Filler metals used for surfacing and overlaying are the same as those used in the other welding processes described. Copper alloys are used for corrosion resistance or to build up surfaces for wear. Alloys designed for corrosion resistance are covered under ANSI/AWS A5.6 and A5.7. Classifications, such as CuSn-A, CuSn-C, CuAl-A2, and Cu-NiAl, are used for this purpose.

Hardfacing alloys, as covered in ANSI/AWS A5.13, are used for surfacing cast irons. Copper alloys, such as the CuZn grades B, C, and D, are used for bearing surfaces. A buttering layer of CuAl-2 is recommended before trying to apply a surface of the CuZn grades. If several layers are required, then a stress relieving of 315 °C (600 °F) is recommended after each layer is added.

Stainless steels, not typically used for joining or repairing iron castings, are used for surfacing to increase corrosion and oxidation resistance. Carbon absorption into the stainless steel is a problem. It can cause the formation of chromium carbide and can lower the crack resistance by reducing the level of ferrite. This can be eliminated by first surfacing with a low-carbon nickel electrode before adding the stainless steel surface. The stainless steel deposit may increase in nickel level as a result of the buttering, thus lowering the volume of ferrite, but the formation of chromium carbides is eliminated. The amount of dilution can be minimized by overlapping the beads.

A deposit with a convex shape is recommended. This will help eliminate the tendency of the weld to separate from the casting due to differences in the coefficient of expansion between stainless steel and cast iron.

High-nickel alloys are used for surfacing to correct machining errors, to provide surface machinability in castings containing high phosphorus, and to improve resistance to caustic attack. Classifications such as ENi-CI and ENi-CI-A, which are discussed in the "Shielded Metal Arc Welding" of this article, are used. In addition, alloys covered in ANSI/AWS A5.11 and A5.14 are also used.

The corrosion resistance of a nickel overlay can be greatly reduced if it becomes diluted with iron and carbon from the cast iron. This dilution can be minimized by overlapping beads up to one-half the bead width. Stringer beads are the preferred method except for the SAW process. Widely oscillated beads with straight polarity (electrode negative) are used for SAW surfacing applications. Again, a buffer layer of nickel should be added to reduce the effects caused by differences in the coefficient of expansion between the surfacing material and the iron casting.

Cast iron alloys are used for buildup when comparable microstructure is required in the application. These electrodes are available in gray, ductile, and malleable iron compositions. The use of preheat, the proper interpass temperature, and a postweld heat treatment may be necessary to ensure a crack-free deposit.

Hardfacing alloys are specified in ANSI/AWS A5.13 and A5.21. They are used for over-laying when wear is the main concern. These alloys may increase the low-stress abrasion resistance, hot hardness, and corrosion resistance of the casting. For more demanding service conditions, nickel- or cobalt-base alloys may be used.

Special precautions must be taken when using hardfacing alloys containing high levels of chromium carbides, which are designed to stress crack in service. Several layers of an iron-nickel alloy, such as ENiFe-CI or ENiFe-CI-A are used to obtain a buffer zone between the hardfacing alloy and the iron casting. Preheating and techniques for minimizing dilution are required when welding these high-chromium alloys.

Ceramic materials or other refractory materials can be overlayed on cast iron when thermal distortion of the casting is important. Other materials, such as alumina and chromium oxide, are used in high-temperature wear and corrosion-resistance applications. Tungsten carbide is applied when surface temperatures up to 540 °C (1000 °F) will be encountered in service or when abrasion resistance is the main concern.

Minimizing dilution of the base iron casting into the weld deposit can be kept to a minimum by:

- MAXIMIZING THE AMOUNT OF METAL DEPOSITED PER UNIT LENGTH. THE BEADS BEING DEPOSITED SHOULD BE OVERLAPPED AS MUCH AS POSSIBLE
- MINIMIZING THE AMOUNT OF HEAT INPUT IN ORDER TO REDUCE THE AMOUNT OF BASE METAL MELTING. THE USE OF STRAIGHT-POLARITY DC CURRENT WILL HELP DECREASE THE PENETRATION
- USING MULTIPLE LAYERS INSTEAD OF ONLY ONE LAYER

Caution must be exercised when considering the reduction in penetration or melting of the cast iron substrate. The deposited weld metal must melt part of the casting to adhere properly. A rule of thumb is that a minimum of 5 to 10% dilution is needed for complete fusion.

Reference cited in this section

1. "GUIDE FOR WELDING IRON CASTINGS," ANSI/AWS D11.2-89, AMERICAN WELDING SOCIETY, 1989

Welding of Cast Irons*

Roger A. Bushey, The ESAB Group, Inc.

References

1. "GUIDE FOR WELDING IRON CASTINGS," ANSI/AWS D11.2-89, AMERICAN WELDING SOCIETY, 1989
2. "SPECIFICATION FOR WELDING RODS AND COVERED ELECTRODES FOR CAST IRON," ANSI/AWS A5.15-90, AMERICAN WELDING SOCIETY, 1990

Welding of Aluminum Alloys

Paul B. Dickerson, Aluminum Company of America (retired)

Introduction

ALUMINUM AND ITS ALLOYS can be joined by as many or more methods than any other metal. The primary methods used are the gas-shielded arc welding processes, which are described in detail in this article. Other methods are briefly reviewed.

Aluminum has several chemical and physical properties that need to be understood when using the various joining processes. The specific properties that affect welding are its oxide characteristics; the solubility of hydrogen in molten aluminum; its thermal, electrical, and nonmagnetic characteristics; its lack of color change when heated; and its wide range of mechanical properties and melting temperatures that result from alloying with other metals. Various property values are provided in Tables 1, 2, 3, and 4.

TABLE 1 COMPOSITION, PHYSICAL PROPERTIES, AND WELDABILITY OF SELECTED NON-HEAT-TREATABLE WROUGHT ALUMINUM ALLOYS

BASE ALLOY	**NOMINAL COMPOSITION, WT%**					**APPROXIMATE MELTING RANGE**		**THERMAL CONDUCTIVITY AT 25 °C (77 °F)**		**ELECTRICAL CONDUCTIVITY(A), %IACS**	WELDABILITY(B)				
	Al	**Cu**	**Mn**	**Mg**	**Cr**	**°C**	**°F**	**W/M · K**	**BTU/FT · H · °F**		**GAS**	**ARC WITH FLUX**	**ARC WITH INERT GAS**	**RESISTANCE**	PRESSURE
1060	[GES]99.6	. . .	. . .	. . .	. . .	646-657	1195-1215	234	135	62	A	A	A	B	**A**
1100	[GES]99.0	. . .	. . .	. . .	. . .	643-657	1190-1215	222	128	59	A	A	A	A	**A**
1350	[GES]99.5	. . .	. . .	. . .	. . .	646-657	1195-1215	234	135	62	A	A	A	B	**A**
3003	REM	0.12	1.2	. . .	. . .	643-654	1190-1210	193	112	50	A	A	A	A	**A**
3004	REM	. . .	1.2	1.0	. . .	629-654	1165-1210	163	94	42	B	A	A	A	**B**
5005	REM	. . .	. . .	0.8	. . .	632-654	1170-1210	200	116	52	A	A	A	A	**A**
5050	REM	. . .	. . .	1.4	. . .	624-652	1155-1205	193	112	50	A	A	A	A	**A**
5052, 5652	REM	. . .	. . .	2.5	. . .	607-649	1125-1200	138	80	35	A	A	A	A	**B**
5083	REM	. . .	0.7	4.4	0.15	574-638	1065-1180	117	67.5	29	C	C	A	A	**C**
5086	REM	. . .	0.45	4.0	0.15	585-641	1085-1185	125	72.5	31	C	C	A	A	**B**
5154, 5254	REM	. . .		3.5	0.25	593-643	1100-1190	125	72.5	32	C	C	A	A	**B**
5454	REM	. . .	0.8	2.7	0.12	602-646	1115-1195	134	77.5	34	B	B	A	A	**B**
5456	**REM**	**. . .**	**0.8**	**5.1**	**0.12**	**568-638**	**1055-1180**	**117**	**67.5**	**29**	**C**	**C**	**A**	**A**	C

(A) EQUAL VOLUME AT 20 °C (68 °F).

(B) WELDABILITY RATINGS: A, READILY WELDABLE; B, WELDABLE IN MOST APPLICATIONS, BUT MAY REQUIRE SPECIAL TECHNIQUE OR FILLER ALLOY; C, LIMITED WELDABILITY

TABLE 2 COMPOSITION, PHYSICAL PROPERTIES, AND WELDABILITY OF SELECTED HEAT-TREATABLE WROUGHT ALUMINUM ALLOYS

BASE ALLOY	**NOMINAL COMPOSITION OF ALLOYING ELEMENTS, WT%**								**APPROXIMATE MELTING RANGE**		**THERMAL CONDUCTIVITY AT 25 °C (77 °F)**		**ELECTRICAL CONDUCTIVITY (A), %IACS**	WELDABILITY(B)				
	CU	**SI**	**MN**	**MG**	**ZN**	**CR**	**AL**	**OTHERS**	**°C**	**°F**	**W/M · K**	**BTU/FT · H · °F**		**GAS**	**ARC WITH FLUX**	**ARC WITH INERT GAS**	**RESISTANCE**	PRESSURE
2014	4.4	0.8	0.8	0.50	. . .	. . .	REM	. . .	507-593	945-1100	154	89	40	X	C	B	B	**C**
2024	4.4	. . .	0.6	1.5	. . .	. . .	REM	. . .	502-638	935-1180	121	70	30	X	C	C	B	**C**
2090	2.7	. . .	. . .	. . .	. . .	. . .	REM	2.2 LI; 0.12 ZR	560-643	1040-1190	88	51	17	X	X	B	B	**C**
2219	6.3	. . .	0.30	. . .	. . .	. . .	REM	0.6 TI; 0.10 V; 0.18 ZR	543-643	1010-1190	121	70	30	X	C	A	B	**C**
2618	2.3	0.18	. . .	1.6	. . .	. . .	REM	1.1 FE; 1.0 NI; 0.07 TI	549-638	1020-1180	161	93	37	X	C	B	B	**C**
6009	0.40	0.8	0.50	0.6	0.25	0.10	REM	. . .	560-649	1040-1200	167	97	44	C	C	B	B	**B**
6013	0.9	0.25	0.35	0.95	. . .	. . .	REM	. . .	579-649	1075-1200	150	87	38	C	C	B	A	**B**
6061	0.28	0.6	. . .	1.0	. . .	0.20	REM	. . .	582-652	1080-1205	167	97	43	A	A	A	A	**B**
6063	. . .	0.40	. . .	0.7	. . .	. . .	REM	. . .	616-654	1140-1210	200	116	53	A	A	A	A	**B**
6101	. . .	0.50	. . .	0.6	. . .	. . .	REM	. . .	621-654	1150-1210	218	126	57	A	A	A	A	**B**
6262	0.28	0.6	. . .	1.0	. . .	0.09	REM	0.6 PB; 0.6 BI	582-652	1080-1205	172	99	44	C	C	B	A	**B**
6351	. . .	1.0	0.6	1.0	. . .	. . .	REM	. . .	596-652	1105-1205	176	102	46	A	A	A	A	**B**
6951	0.28	0.35	. . .	0.6	. . .	. . .	REM	. . .	616-654	1140-1210	198	114	52	A	A	A	A	**A**
7005	. . .	. . .	0.45	1.4	4.5	0.13	REM	0.15 ZR; 0.04 TI	607-646	1125-1195	. . .	. . .	. . .	X	X	A	A	**B**

7039	. . .	. . .	0.30	2.8	4.0	0.20	REM	. . .	577-638	1070-1180	154	89	34	X	X	A	A	**B**
7075	1.6	. . .	. . .	2.5	5.6	0.23	REM	. . .	477-635	890-1175	130	75	33	X	X	C	B	**C**
7079	0.6	. . .	0.20	3.3	4.3	0.20	REM	. . .	482-638	900-1180	125	72	32	X	X	C	B	**C**
7178	**2.0**	**. . .**	**. . .**	**2.8**	**6.8**	**0.23**	**REM**	**. . .**	**477-629**	**890-1165**	**125**	**72**	**31**	**X**	**X**	**C**	**B**	**C**

(A) EQUAL VOLUME AT 20 °C (68 °F).

(B) WELDABILITY RATINGS: A, READILY WELDABLE; B, WELDABLE IN MOST APPLICATIONS, BUT MAY REQUIRE SPECIAL TECHNIQUE OR FILLER ALLOY; C, LIMITED WELDABILITY; X, METHOD NOT RECOMMENDED

TABLE 3 COMPOSITION, PHYSICAL PROPERTIES, AND WELDABILITY OF SELECTED NON-HEAT-TREATABLE CAST ALUMINUM ALLOYS

BASE ALLOY	**NOMINAL COMPOSITION OF ALLOYING ELEMENTS, WT%**					**APPROXIMATE MELTING RANGE**		**THERMAL CONDUCTIVITY AT 25 °C (77 °F)**		**ELECTRICAL CONDUCTIVITY[(A)], %IACS**	WELDABILITY[(B)]				
	CU	**SI**	**MG**	**ZN**	**OTHERS**	**°C**	**°F**	**W/M · K**	**BTU/FT · H · °F**		**GAS**	**ARC WITH FLUX**	**ARC WITH INERT GAS**	**RESISTANCE**	PRESSURE
SAND CASTINGS															
208.0	4.0	3.0	. . .	. . .	. . .	521-632	970-1170	121	70	31	C	C	B	B	**X**
443.0	. . .	5.25	. . .	. . .	. . .	577-632	1070-1170	146	84	37	A	A	A	A	**X**
511.0	. . .	0.50	4.0	. . .	. . .	588-638	1090-1180	141	82	36	X	X	A	A	**X**
512.0	. . .	1.8	4.0	. . .	. . .	588-632	1090-1170	146	84	38	X	X	B	B	**X**
514.0	. . .	. . .	4.0	. . .	. . .	599-638	1110-1180	137	79	35	X	X	A	A	**X**
535.0	. . .	. . .	6.9	. . .	0.18 MN; 0.18 TI; 0.005 BE	549-632	1020-1170	99	58	23	X	X	A	A	**X**
710.0	0.50	. . .	0.7	6.5	. . .	599-649	1110-1200	137	79	35	C	C	B	B	**X**
712.0	. . .	. . .	0.6	5.8	0.50 CR; 0.20 TI	599-638	1110-1180	159	92	40	C	C	A	B	**X**
PERMANENT MOLD CASTINGS															
208.0	4.0	3.0	. . .	. . .	. . .	521-632	970-1170	121	70	31	C	C	B	B	**X**
238.0	10.0	4.0	0.25	. . .	. . .	510-599	950-1110	104	60	25	C	C	B	A	**X**
443.0	. . .	5.25	. . .	. . .	. . .	577-632	1070-1170	146	84	37	A	A	A	A	**X**
A444.0	. . .	7.0	. . .	. . .	. . .	577-632	1070-1170	159	92	41	A	A	A	A	**X**
513.0	. . .	. . .	4.0	1.8	. . .	582-638	1080-1180	133	77	34	X	X	A	A	**X**
711.0	0.50	. . .	0.35	6.5	1.0 FE	599-643	1110-1190	159	92	40	B	B	A	A	**X**
DIE CASTINGS															
360.0	. . .	9.5	0.50	. . .	. . .	571-588	1060-1090	146	84	37	C	X	C	B	**X**

380.0	3.5	8.5	. . .	. . .	. . .	521-588	970-1090	108	62	27	C	X	C	B	**X**
413.0	. . .	12.0	. . .	. . .	. . .	577-588	1070-1090	154	89	39	C	X	C	B	**X**
518.0	**. . .**	**. . .**	**8.0**	**. . .**	**. . .**	**538-621**	**1000-1150**	**99**	**58**	**24**	**X**	**X**	**C**	**B**	**X**

(A) EQUAL VOLUME AT 20 °C (68 °F).

(B) WELDABILITY RATINGS: A, READILY WELDABLE; B, WELDABLE IN MOST APPLICATIONS, BUT MAY REQUIRE SPECIFIC TECHNIQUE OR FILLER ALLOY; C, LIMITED WELDABILITY; X, JOINING METHOD NOT RECOMMENDED

TABLE 4 COMPOSITION, PHYSICAL PROPERTIES, AND WELDABILITY OF SELECTED HEAT-TREATABLE CAST ALUMINUM ALLOYS

BASE ALLOY	**NOMINAL COMPOSITION OF ALLOYING ELEMENTS, WT%**					**APPROXIMATE MELTING RANGE**		**THERMAL CONDUCTIVITY AT 25 °C (77 °F)**		**ELECTRICAL CONDUCTIVITY(A), %IACS**	WELDABILITY(B)				
	CU	**SI**	**MG**	**NI**	**OTHERS**	**°C**	**°F**	**W/M · K**	**BTU/FT · H · °F**		**GAS**	**ARC WITH FLUX**	**ARC WITH INERT GAS**	**RESISTANCE**	PRESSURE
ONLY SAND CASTINGS															
A201.0	4.5	. . .	0.25	. . .	0.7 AG; 0.30 MN; 0.25 TI	571-649	1060-1200	121	70	30	C	C	B	B	**X**
240.0	8.0	. . .	6.0	. . .	0.50 MN	516-604	960-1120	95	55	23	X	X	C	B	**X**
A242.0	4.1	. . .	1.5	2.0	0.20 CR; 0.13 TI	527-638	980-1180	146	84	38	X	X	B	B	**X**
295.0	4.5	1.1	. . .	. . .	. . .	521-643	970-1190	141	82	35	C	C	B	B	**X**
520.0	. . .	. . .	10.0	. . .	. . .	449-599	840-1110	87	50	21	C	C	B	C	**X**
ONLY PERMANENT MOLD CASTINGS															
332.0	3.0	9.5	1.0	. . .	1.0 ZN	521-582	970-1080	104	60	26	X	X	B	B	**X**
333.0	3.5	9.0	0.30	. . .	. . .	521-588	970-1090	117	68	29	X	X	B	B	**X**
336.0	1.0	12.0	1.0	2.5	. . .	538-571	1000-1060	117	68	29	C	C	B	B	**X**
354.0	1.8	9.0	0.50		. . .	538-599	1000-1110	125	72	32	C	C	B	B	**X**
SAND AND PERMANENT MOLD CASTINGS															
222.0	10.0	. . .	0.25	. . .	. . .	521-627	970-1160	130	75	33	X	X	B	B	**X**
242.0	4.0	. . .	1.5	2.0	. . .	527-638	980-1180	133	77	34	X	X	C	B	**X**
319.0	3.5	6.0	. . .	. . .	. . .	521-604	970-1120	112	65	27	C	C	B	B	**X**
355.0	1.3	5.0	0.50	. . .	. . .	549-621	1020-1150	150	87	39	B	B	B	B	**X**
C355.0	1.3	5.0	0.50	. . .	. . .	549-621	1020-1150	146	84	39	B	B	B	B	**X**
356.0	. . .	7.0	0.35	. . .	. . .	560-616	1040-1140	150	87	41	A	A	A	A	**X**

A356.0	. . .	7.0	0.35	. . .	. . .	560-610	1040-1130	150	87	40	A	A	A	A	**X**
A357.0	. . .	7.0	0.55	. . .	0.5 BE	554-610	1030-1130	159	92	40	B	B	A	A	**X**
359.0	**. . .**	**9.0**	**0.6**	**. . .**	**. . .**	**566-599**	**1050-1110**	**137**	**79**	**35**	**B**	**B**	**A**	**A**	**X**

(A) EQUAL VOLUME AT 20 °C (68 °F).

(B) WELDABILITY RATINGS: A, READILY WELDABLE; B, WELDABLE IN MOST APPLICATIONS, BUT MAY REQUIRE SPECIAL TECHNIQUE OR FILLER ALLOY; C, LIMITED WELDABILITY; X, JOINING METHOD NOT RECOMMENDED

Welding of Aluminum Alloys

Paul B. Dickerson, Aluminum Company of America (retired)

Properties of Aluminum

Oxide. Aluminum has a strong chemical affinity for oxygen and it will oxidize immediately upon exposure to air. Thermal treatments and moist storage conditions will increase oxide thickness, as will electrochemical "anodizing" treatments applied for ornamental, architectural, and wear-resistance applications. The natural oxide is thin enough to be removed by either inert-gas welding arcs or fluxes. However, the thicker oxide should be reduced by chemical or mechanical means prior to welding.

Aluminum oxide melts at about 2050 °C (3720 °F), which is much higher than the melting point of the base alloy. If the oxide is not removed or displaced, the result is incomplete fusion. The fluxes used with some joining processes in order to remove the oxide contain chlorides and fluorides that must be removed after the joining operation to avoid a possible corrosion problem in service. This has led to a wide acceptance of the inert-gas arc welding processes, which remove the oxide without use of a flux.

Aluminum oxide is an electrical insulator. If it is thick enough, as in anodic coatings, then it can prevent arc initiation. In this case, oxide reduction is required not only in the weld joint, but also at the location of the ground lead connection. A thick oxide on a gas-metal arc welding (GMAW) electrode (from a thermal treatment or storage in a humid atmosphere) can produce erratic electrical commutation in the contact tube of the gun, with poor welding results.

Hydrogen Solubility. Hydrogen dissolves very rapidly in molten aluminum. However, hydrogen has almost no solubility in solid aluminum and it has been determined to be the primary cause of porosity in aluminum welds. High temperatures of the weld pool allow a large amount of hydrogen to be absorbed, and as the pool solidifies, the solubility of hydrogen is greatly reduced. Hydrogen that exceeds the effective solubility limit forms gas porosity, if it does not escape from the solidifying weld.

Any source of hydrogen--such as lubricant on the base metal or filler, moisture on the surface or in the form of a hydrated oxide on the base or filler metals, moisture leaks or condensation inside the nozzle of a water-cooled torch, or moisture in the shielding gas--must be eliminated in order to produce sound welds. The aluminum-magnesium (5*xxx* series) alloys are most sensitive to forming a hydrated oxide, and spooled bare electrode must be stored in a dry, heated area. The moisture cannot be removed from the hydrated oxide by heating, as is done with flux-coated electrodes.

Electrical Conductivity. When arc welding, the ground connection can be attached anywhere on the part. This is preferred to the attachment of the ground cable to a steel work table, because the insulating oxide on the aluminum and rust on the steel can both cause a poor transfer of the electric current. Arcing at contact points can occur and can mar the aluminum surface. Aluminum alloys possess high electrical conductivity and pure aluminum has 62% that of pure copper, as shown in Tables 1, 2, 3, and 4.

High electrical conductivity permits the use of long contact tubes in GMAW guns, because resistance heating of the electrode does not occur, as is experienced with ferrous electrodes. Long contact tubes are desirable to provide multiple contact points for good current commutation. Arcing in short contact tubes is common with aluminum, because of its natural insulating oxide, when only a couple of contacts are available. This is especially true when the arc is being initiated using a constant-voltage power supply and a fast electrode "run-in" speed, which cause a high-current surge.

The nonmagnetic property of aluminum minimizes problems of "arc blow" and makes it a suitable material for clamps and fixtures when arc welding other materials. Vacuum attachments, rather than magnetic devices, are required for lifting or attaching portable welding tracks and the like.

Thermal Characteristics. The thermal conductivity of aluminum is about 6 times that of steel. Although the melting temperature of aluminum alloys is substantially below that of ferrous alloys, higher heat inputs are required to weld aluminum because of its high specific heat. Even so, if a slow welding pace is employed, then the heat can be conducted ahead of the arc, requiring continuous adjustments of welding parameters. Most uniform welding conditions dictate high heat inputs with moderate welding speeds.

High thermal conductivity makes aluminum very sensitive to fluctuations in heat input by the welding process. Specifically, a steady heat input is required with the faster welding processes, such as GMAW, to avoid variations in penetration and fusion. For instance, both of these discontinuities can result from the fluctuations in current that occur in the root pass of a groove or fillet weld when manipulating a GMAW gun powered by a constant-voltage machine. For assemblies subjected to radiographic examination, the use of a constant-current (drooping volt-ampere characteristic) power supply with the GMAW process is preferred in order to produce the soundest welds in aluminum.

The thermal expansion of aluminum is about twice that of steel, and aluminum welds shrink about 6% by volume during solidification. Thin material must be fixtured or "tack welded" closely to keep the edges of the joint in alignment. Shrinkage must be balanced by the welding sequence to control distortion. Weld termination craters must be given more attention and avoided in order to prevent shrinkage voids and crater cracking.

Aluminum does not change color as it is heated. Therefore, the welder must be in a position to see the melting under the arc in order to control the degree of melting as the weld progresses.

Forms of Aluminum. Most forms of aluminum can be welded. All the wrought forms (sheet, plate, extrusions, forgings, rod, bar, and impact extrusions), as well as sand and permanent mold castings, can be welded. Welding on conventional die castings produces excessive porosity in both the weld and the base metal adjacent to the weld because of internal gas. Vacuum die castings, however, have been welded with excellent results. Powder metallurgy (P/M) parts also may suffer from porosity during welding because of internal gas.

The alloy composition is a much more significant factor than the form in determining the weldability of an aluminum alloy. Compositions of the major wrought base alloys and their relative weldability ratings are given in Tables 1 and 2. Similar data are listed in Tables 3 and 4 for the common casting alloys.

Welding of Aluminum Alloys

Paul B. Dickerson, Aluminum Company of America (retired)

Filler Alloy Selection Criteria

When choosing the optimum filler alloy, the end use of the weldment and its desired performance must be prime considerations. Many alloys and alloy combinations can be joined using any one of several filler alloys, but only one filler may be optimal for a specific application. Table 5 lists the chemical composition and melting range of standard aluminum filler alloys.

TABLE 5 NOMINAL COMPOSITION AND MELTING RANGE OF STANDARD ALUMINUM FILLER ALLOYS

ALUMINUM ALLOY	**NOMINAL COMPOSITION, WT%**								APPROXIMATE MELTING RANGE	
	SI	**CU**	**MN**	**MG**	**CR**	**TI**	**AL**	**OTHERS**	**°C**	°F
1100	. . .	0.12	. . .	. . .	. . .	. . .	[GES]99.00	. . .	643-657	**1190-1215**
1188	. . .	. . .	. . .	. . .	. . .	. . .	[GES]99.88	. . .	657-660	**1215-1220**
2319	. . .	6.3	0.30	. . .	. . .	0.15	REM	0.18 ZR; 0.10 V	543-643	**1010-1190**
4009(A)	5.0	1.25	. . .	0.50	. . .	. . .	REM	. . .	546-621	**1015-1150**
4010(B)	7.0	. . .	. . .	0.35	. . .	. . .	REM	. . .	557-613	**1035-1135**
4011(C)	7.0	. . .	. . .	0.58	. . .	0.12	REM	0.55 BE	557-613	**1035-1135**
4043	5.25	. . .	. . .	. . .	. . .	. . .	REM	. . .	574-632	**1065-1170**
4047	12.0	. . .	. . .	. . .	. . .	. . .	REM	. . .	577-582	**1070-1080**
4145	10.0	4.0	. . .	. . .	. . .	. . .	REM	. . .	521-585	**970-1085**
4643	4.1	. . .	. . .	0.20	. . .	. . .	REM	. . .	574-635	**1065-1175**

5183	. . .	. . .	0.75	4.75	0.15	. . .	REM	. . .	579-638	**1075-1180**
5356	. . .	. . .	0.12	5.0	0.12	0.13	REM	. . .	571-635	**1060-1175**
5554	. . .	. . .	0.75	2.7	0.12	0.12	REM	. . .	602-646	**1115-1195**
5556	. . .	. . .	0.75	5.1	0.12	0.12	REM	. . .	568-635	**1055-1175**
5654	. . .	. . .	. . .	3.5	0.25	0.10	REM	. . .	593-643	**1100-1190**
C355.0	5.0	1.25	. . .	0.50	. . .	. . .	REM	. . .	546-621	**1015-1150**
A356.0	7.0	. . .	. . .	0.35	. . .	. . .	REM	. . .	557-613	**1035-1135**
A357.0	**7.0**	**. . .**	**. . .**	**0.58**	**. . .**	**0.12**	**REM**	**0.55 BE**	**557-613**	**1035-1135**

(A) WROUGHT ALLOY WITH COMPOSITION IDENTICAL TO CAST ALLOY C355.0.

(B) WROUGHT ALLOY WITH COMPOSITION IDENTICAL TO CAST ALLOY A356.0.

(C) WROUGHT ALLOY WITH COMPOSITION IDENTICAL TO CAST ALLOY A357.0

The primary factors commonly considered when selecting a welding filler alloy are:

- EASE OF WELDING OR FREEDOM FROM CRACKING
- TENSILE OR SHEAR STRENGTH OF THE WELD
- WELD DUCTILITY
- SERVICE TEMPERATURE
- CORROSION RESISTANCE
- COLOR MATCH BETWEEN THE WELD AND BASE ALLOY AFTER ANODIZING

Each of these factors is addressed below.

Sensitivity to Weld Cracking. Ease of welding is the first consideration for most welding applications. In general, the non-heat-treatable aluminum alloys can be welded with a filler alloy of the same basic composition as the base alloy. In some cases, as shown in Table 5, a small amount of grain refiner, such as titanium, is added to the filler alloy to minimize cracking during welding. For example, filler alloys 5554 and 5556 are identical to base alloys 5454 and 5456, respectively, except that titanium is added to the filler alloy.

The heat-treatable aluminum alloys are some-what more complex, metallurgically, and are more sensitive to "hot short" cracking, which results from heat-affected zone (HAZ) liquidation during the welding operation. Generally, a dissimilar alloy filler having higher levels of solute (for example, copper or silicon) is used in this case. At high solute levels, there is sufficient eutectic liquid available for the back-filling (healing) of tears that formed during solidification. Typically, alloys used as filler for welding heat-treatable aluminum alloys have a lower melting temperature than the base alloy. By allowing the low-melting-point constituents of the base alloy adjacent to the weld to solidify before the weld metal, stresses are minimized on the base metal during its hot short condition, and tendencies for intergranular cracking are greatly reduced. One heat-treatable base alloy that exhibits low sensitivity to hot short cracking is 2219, which can be welded with a similar filler alloy, 2319.

A simple inverted "T" fillet weld test (Ref 1, 2) can be used to determine the compatibility of a filler alloy with a base alloy. It is a measure of weldability in terms of resistance to weld metal hot cracking. This test requires that a continuous fillet weld be made joining 13 to 25 mm ($\frac{1}{2}$ to 1 in.) plate by the gas-tungsten arc welding (GTAW) process. The total length of cracking that results from a given combination of base and filler alloys provides a quantitative comparison with other alloy combinations. When no cracking is observed in the continuous fillet weld, a more sensitive discontinuous weld test can be used. In this case, a weld nugget is deposited and allowed to solidify before depositing an overlapping weld nugget. Typical weld cracking results obtained from the fillet weld test and similar tests (Ref 3, 4, 5, 6) are shown in Fig. 1 and 2. These tests, as well as field experience, have indicated that:

- THE HIGH-PURITY L*XXX* SERIES ALLOYS AND 3003 ARE EASY TO WELD WITH BASE ALLOY FILLER, 1100 ALLOY, OR AN ALUMINUM-SILICON ALLOY FILLER, SUCH AS 4043.
- ALLOY 2219 EXHIBITS THE BEST WELDABILITY OF THE 2*XXX* SERIES BASE ALLOYS AND IS EASILY WELDED WITH 2319, 4043, AND 4145 FILLERS.
- ALUMINUM-SILICON-COPPER FILLER ALLOY 4145 PROVIDES THE LEAST

SUSCEPTIBILITY TO WELD CRACKING WITH 2*XXX* SERIES WROUGHT COPPER BEARING ALLOYS, AS WELL AS ALUMINUM-COPPER AND ALUMINUM-SILICON-COPPER ALUMINUM ALLOY CASTINGS.

- THE CRACKING SENSITIVITY OF ALUMINUM-MAGNESIUM ALLOY WELDS DECREASES AS THE MAGNESIUM CONTENT OF THE WELD INCREASES ABOVE 2%. THE HIGH-MAGNESIUM CONTENT ALUMINUM FILLER ALLOYS 5356, 5183, AND 5556 CAN BE USED TO WELD THE ALUMINUM-MAGNESIUM WROUGHT AND CAST ALUMINUM ALLOYS WITH RELATIVE EASE, USING THE GAS-SHIELDED ARC WELDING PROCESSES.
- THE 6*XXX* SERIES BASE ALLOYS ARE MOST EASILY WELDED WITH THE ALUMINUM-SILICON TYPE FILLER ALLOYS, SUCH AS 4043 AND 4047. HOWEVER, THE ALUMINUM-MAGNESIUM TYPE FILLER ALLOYS CAN ALSO BE EMPLOYED SATISFACTORILY WITH THE LOW-COPPER BEARING 6*XXX* ALLOYS WHEN HIGHER ASWELDED SHEAR STRENGTH AND WELD METAL DUCTILITY ARE REQUIRED. THE 6*XXX* ALLOYS SHOULD NOT BE WELDED WITH BASE ALLOY FILLER OR WITHOUT A FILLER ADDITION, BECAUSE IT CAN RESULT IN CRACKING. WHEN 4043 FILLER ALLOY IS USED, THE WELD METAL SHOULD POSSESS A MINIMUM OF 50% 4043; WITH 5356 FILLER ALLOY, THE WELD METAL SHOULD CONSIST OF 70% 5356, MINIMUM (REF 7). JOINTS IN 6*XXX* SERIES ALLOYS ARE OFTEN BEVELED OR SPACED TO PERMIT THE EXCESS FILLER ALLOY DILUTION.
- THE 7*XXX* SERIES (ALUMINUM-ZINC-MAGNESIUM) ALLOYS EXHIBIT A WIDE RANGE OF CRACK SENSITIVITY DURING WELDING. ALLOYS 7005 AND 7039, WITH A LOW (<0.1%) COPPER CONTENT, HAVE A NARROW MELTING RANGE AND CAN BE READILY JOINED WITH THE HIGH-MAGNESIUM CONTENT FILLER ALLOYS 5356, 5183, AND 5556. THE 7*XXX* SERIES ALLOYS THAT POSSESS A SUBSTANTIAL AMOUNT OF COPPER, SUCH AS 7075 AND 7178, HAVE A VERY WIDE MELTING RANGE WITH A LOW SOLIDUS TEMPERATURE AND ARE EXTREMELY SENSITIVE TO WELD CRACKING WHEN ARC WELDED.
- THE RECENTLY INTRODUCED ALUMINUM-LITHIUM ALLOYS ARE WELDABLE WHEN CORRECT FILLER ALLOYS ARE SELECTED. ALLOYS 2090 AND 2095 CAN BE EASILY WELDED WITH 2319 ALLOY.

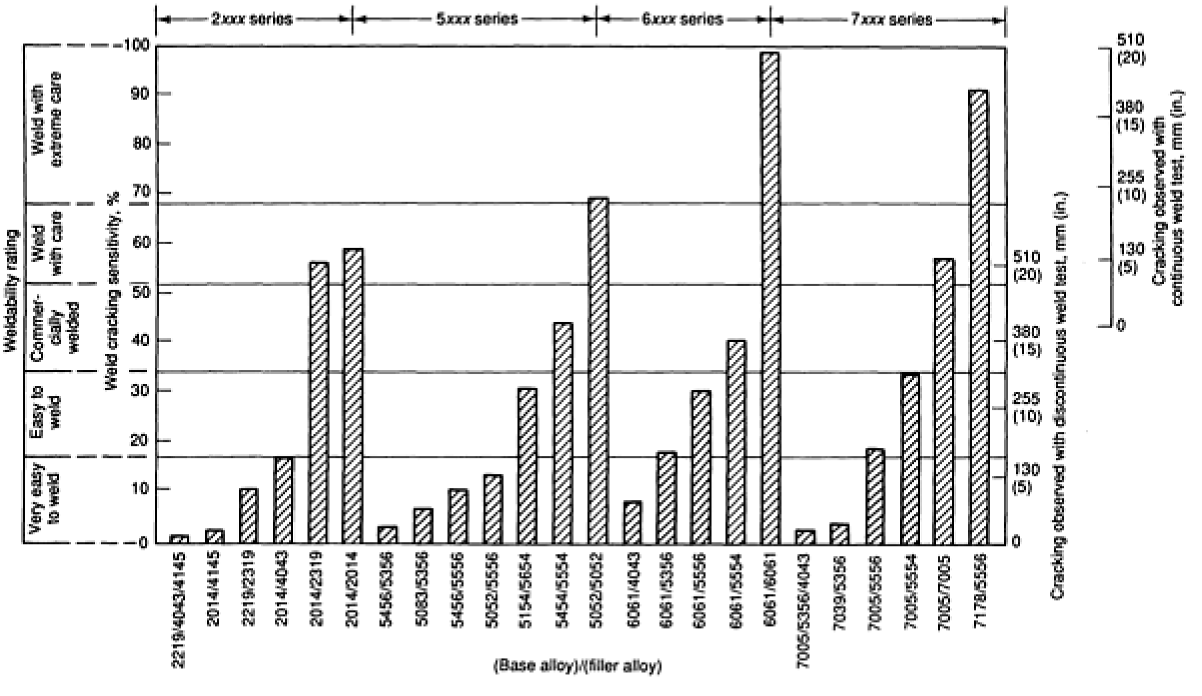

FIG. 1 RELATIVE CRACK SENSITIVITY RATINGS OF SELECTED ALUMINUM (BASE ALLOY/FILLER ALLOY) COMBINATIONS

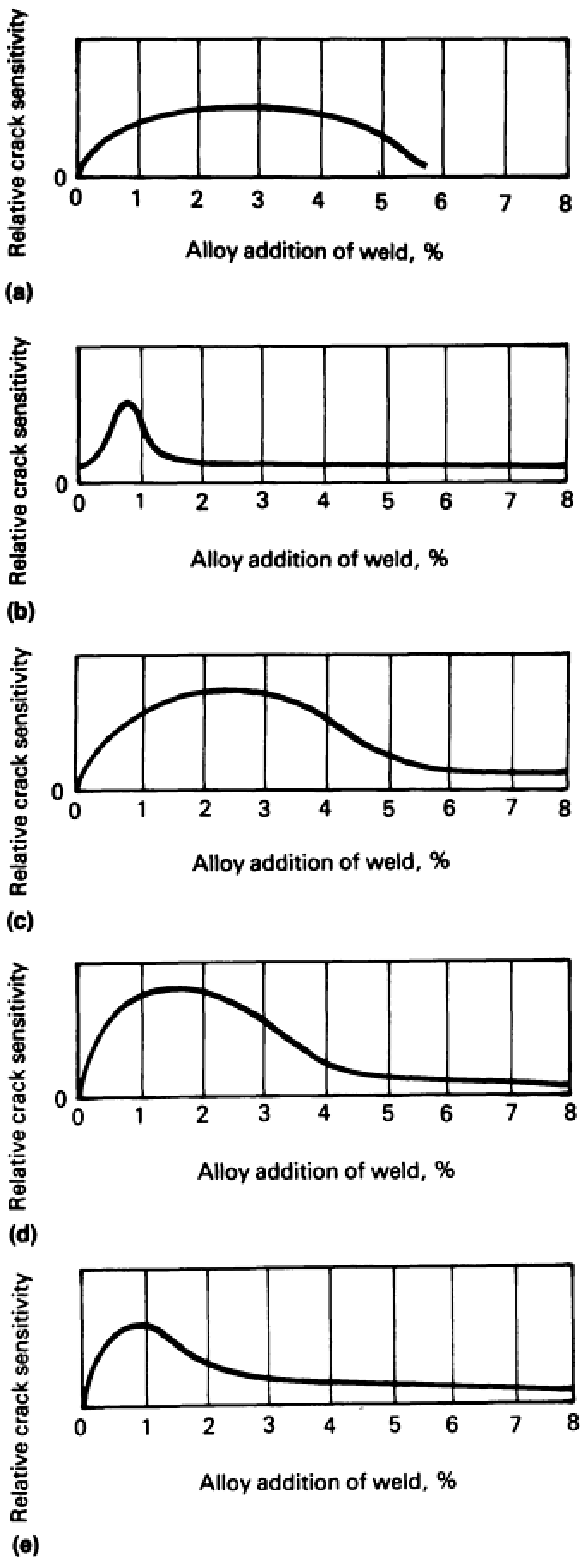

FIG. 2 EFFECT OF ALLOYING ADDITIONS ON SOLIDIFICATION CRACK SENSITIVITY OF SELECTED ALUMINUM

ALLOY SYSTEMS. (A) ALUMINUM-LITHIUM. (B) ALUMINUM-SILICON. (C) ALUMINUM-COPPER. (D) ALUMINUM-MAGNESIUM. (E) ALUMINUM-MAGNESIUM SILICIDE. SOURCE: REF 1, 3, 4, 5, AND 6

Groove weld strength is another major consideration when selecting a proper filler alloy. The heat of welding softens the aluminum alloys adjacent to the weld if they are in any temper other than annealed. In most groove welds, the HAZ of the base alloy will control the as-welded strength of the joint.

For non-heat-treatable aluminum alloys, the zone adjacent to the weld will be completely annealed, as illustrated in Fig. 3. A few seconds at a temperature above 343 °C (650 °F) will completely remove any effect of work hardening for 25 to 38 mm (1 to $1\frac{1}{2}$ in.) in any direction from the weld in these alloys. The welding procedure and choice of fusion welding process have little effect on the transverse ultimate tensile strength of groove welds; they simply affect the width of the HAZ. The mechanical properties of groove welds in common non-heat-treatable alloys are listed in Table 6.

TABLE 6 MECHANICAL PROPERTIES OF GAS-SHIELDED ARC WELDED BUTT JOINTS IN SELECTED NON-HEAT-TREATABLE ALUMINUM ALLOYS

BASE ALLOY	**FILLER ALLOY**	**ULTIMATE TENSILE STRENGTH**				**MINIMUM ULTIMATE YIELD STRENGTH(A)**		**TENSILE ELONGATION, % IN 50.8 MM (2 IN.)**	FREE BEND ELONGATION, %
		TYPICAL		**MINIMUM**					
		MPA	**KSI**	**MPA**	**KSI**	**MPA**	**KSI**		
1060	1188	69	10	55	8	17.2	2.5	29	**63**
1100	1100	90	13	75.8	11	31	4.5	29	**55**
1350	1188	69	10	55	8	17.2	2.5	29	**63**
3003	1100	110	16	96.5	14	48.3	7	24	**54**
5005	1100	110	16	96.5	14	48.3	7	15	**32**
5050	5356	159	23	124	18	55.2	8	18	**36**
5052	5356	193	28	172	25	90	13	19	**39**
5083	5183	296	43	276(B)	40(B)	165	24	16	**34**
5086	5356	262	38	241(C)	35(C)	117	17	17	**38**
5154	5654	228	33	207	30	103	15	17	**39**
5454	5554	234	34	214	31	110	16	17	**40**
5456	**5556**	**317**	**46**	**290(B)**	**42(B)**	**179**	**26**	**14**	**28**

(A) 0.2% OFFSET IN 254 MM (10 IN.) GAGE LENGTH.

(B) FOR THICKNESSES UP TO 38 MM (1.5 IN) PLATE

(C) FOR THICKNESSES UP TO 51 MM (2 IN.) PLATE

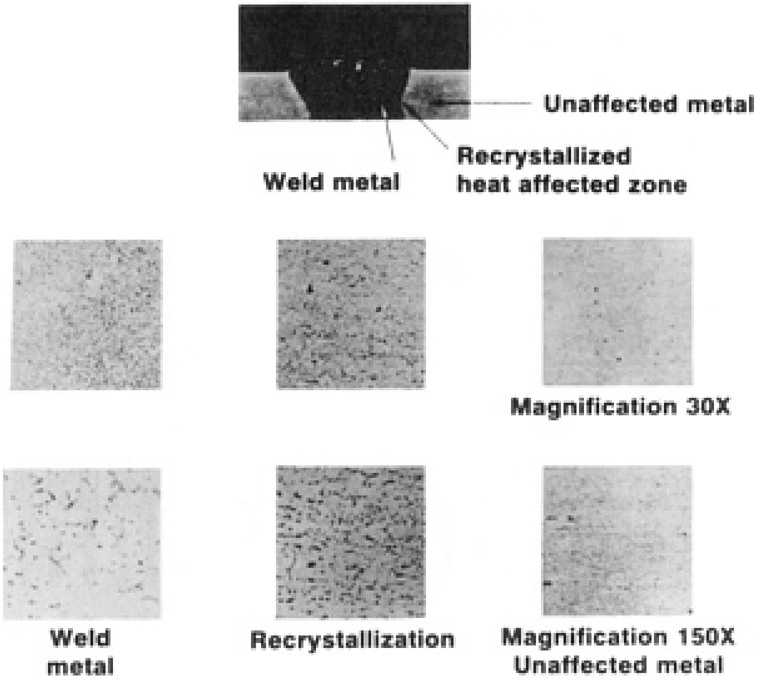
Unaffected metal
Recrystallized heat affected zone
Weld metal
Magnification 30X
Weld metal
Recrystallization
Magnification 150X
Unaffected metal

FIG. 3 PHOTOMICROGRAPHS SHOWING HAZ MICROSTRUCTURES OF A SINGLE-PASS GAS-METAL ARC WELD ON 6.35 MM (0.25 IN.) THICK 5083-H116 ALLOY. KELLER'S REAGENT USED AS ETCHANT

The heat-treatable alloys require 2 to 3 h at their annealing temperature, combined with slow cooling, for a complete anneal. This does not occur during welding, and the HAZ will consist of several stages of dissolution and varying degrees of precipitation, based on the thermal conditions, as shown in Fig. 4. The degree of softening in the HAZ is quite sensitive to the peak temperature reached at a specific location, as well as the time at temperature.

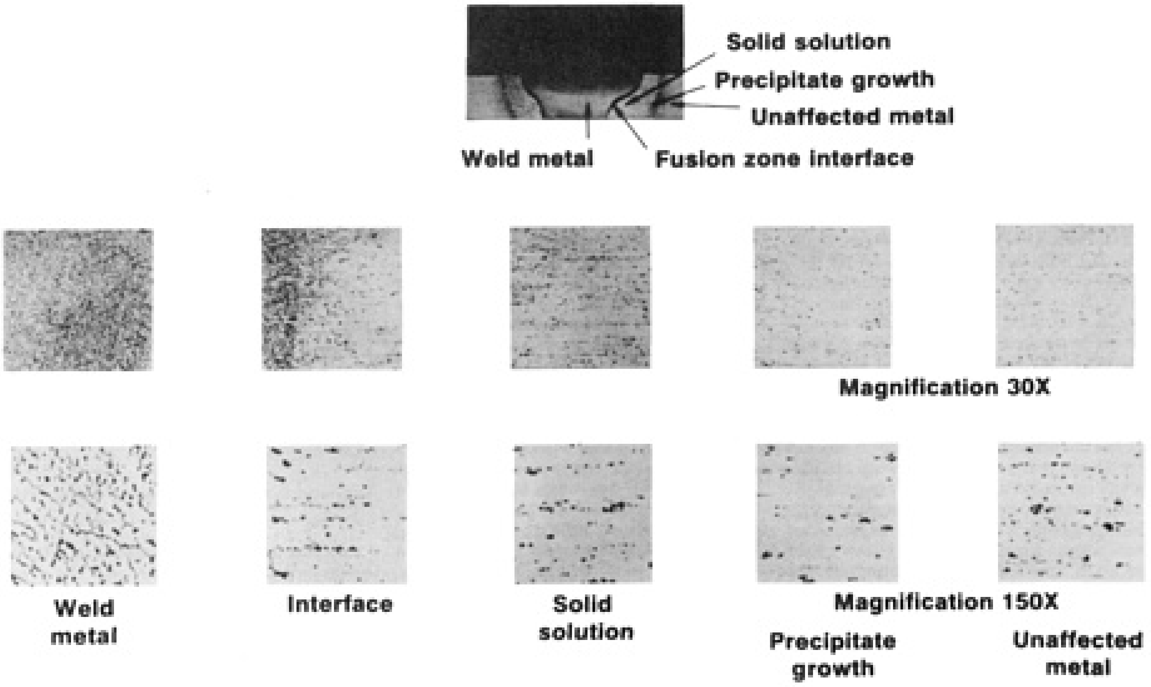
Solid solution
Precipitate growth
Unaffected metal
Weld metal
Fusion zone interface
Magnification 30X
Weld metal
Interface
Solid solution
Magnification 150X
Precipitate growth
Unaffected metal

FIG. 4 PHOTOMICROGRAPHS SHOWING HAZ MICROSTRUCTURES OF A SINGLE-PASS GAS-METAL ARC WELD ON 6.35 MM (0.25 IN.) THICK 6061-T6 ALLOY. KELLER'S REAGENT USED AS ETCHANT

Factors such as preheat temperature, interpass cooling, heat input of the joining process and welding technique, use of "chill" bars, and rate of cooling all have an effect on the degree of softening that results from welding the heat-treatable aluminum alloys. Preheat, lack of interpass cooling, and slow, weaving weld passes all increase peak temperatures and extend high temperatures farther from the weld, as well as increase the time at these high temperatures so as to reduce the mechanical properties in the HAZ. These factors alone, as well as the use of too small a specimen to provide an adequate "heat sink," can create sufficient overheating that the minimum values required for procedure qualification may not be met. The weldable 7*xxx* series alloys 7005 and 7039 are least sensitive to these variables and will naturally age (dissolved precipitate will reprecipitate) at room temperature for 2 to 4 weeks following the welding operation to provide high strength without a postweld thermal operation. Typical mechanical properties for groove welds in the heat-treatable aluminum alloys are listed in Table 7.

TABLE 7 TYPICAL MECHANICAL PROPERTIES OF GAS SHIELDED ARC WELDED BUTT JOINTS IN SELECTED WROUGHT HEAT-TREATABLE ALUMINUM ALLOYS

BASE ALLOY AND TEMPER	**FILLER ALLOY**	**AS WELDED**						POSTWELD HEAT TREATED AND AGED					
		TENSILE STRENGTH		**YIELD STRENGTH(A)**		**ELONGATION**		**TENSILE STRENGTH**		**YIELD STRENGTH(A)**		ELONGATION	
		MPA	**KSI**	**MPA**	**KSI**	**TENSILE, % IN 50.8 MM (2 IN.)**	**FREE BEND, %**	**MPA**	**KSI**	**MPA**	**KSI**	**TENSILE, % IN 50.8 MM (2 IN.)**	FREE BEND, %
2014-T6	4043	234	34	193	28	4	9	345	50	. . .	. . .	2	**5**
2014-T6	2319	241	35	193	28	5	. . .	414	60	317	46	5	**. . .**
2219-T81, T87	2319	241	35	179	26	3	15	379	55	262	38	7	**5**
2219-T31, T37	2319	241	35	179	26	3	15	276(B)	40(B)	228(B)	33(B)	2(B)	**12(B)**
6009-T4	4043	221	32	138	20	9	. . .	303(B)	44(B)	. . .	. . .	. . .	**. . .**
6061-T6	4043	186	27	124	18	8	16	303(C)	44(C)	276(C)	40(C)	5(C)	**11(C)**
6061-T6	5356	207	30	131	19	11	25	. . .	. . .	. . .	. . .	. . .	
6061-T4	4043	186	27	124	18	8	16	241(B)	35(B)	165(B)	24(B)	3(B)	**. . .**
6063-T6	4043	138	20	83	12	8	16	207	30	. . .	. . .	13	**11**
6063-T6	5356	138	20	83	12	12	25	. . .	. . .	. . .	. . .	. . .	**. . .**
7005-T53	5356	317(D)	46(D)	207(D)	30(D)	10(D)	33(D)	. . .	. . .	. . .	. . .	. . .	**. . .**
7039-T61	**5556**	**324(D)**	**47(D)**	**221(D)**	**32(D)**	**10(D)**	**21(D)**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**

(A) 0.2% OFFSET IN 50.8 MM (2 IN.) GAGE LENGTH.

(B) POSTWELD ARTIFICIALLY AGED ONLY.

(C) FOR THICKNESSES GREATER THAN 19 MM ($\frac{3}{4}$ IN.), 4643 FILLER IS REQUIRED.

(D) POSTWELD AGED AT ROOM TEMPERATURE FOR 30 DAYS

When heat-treatable alloy weldments are given a postweld solution heat treatment and artificial aging, the filler selection is limited. The heat-treatable filler alloy 2319 will provide the highest strength for 2014 and 2219 base alloys. Heat-treatable alloy castings are often repaired or welded with a filler of the same composition as the base alloy and can be postweld heat treated to the desired strength. In most cases, the filler is either not a heat-treatable alloy or is one that is only mildly responsive to strengthening thermal treatments. In these cases, dilution of the base alloy into the weld metal is necessary to obtain a weld metal alloy that can be postweld heat treated.

Filler alloy 4043 (5% Si, balance Al) is a non-heat-treatable alloy. When it is used to weld 6061 alloy, some magnesium from the 6061 must be alloyed with it to provide an aluminum-silicon-magnesium mixture in the weld metal. This new weld metal mixture will respond to a postweld solution heat treatment and produce a groove weld with a tensile strength equal to that of the original 6061-T6. When thick sections, that is, 19 mm ($\frac{3}{4}$ in.) and thicker, are welded, the bevel is normally so wide that the dilution effect may not reach the center of the weld. Filler alloy 4643, which contains some magnesium itself, has been used to provide the highest strength in these cases, as shown for a 76 mm (3 in.) groove weld in Table 8. This filler is also useful for fillet welds in thinner metal.

TABLE 8 EFFECT OF WELDING CONDITIONS ON 6061 ALLOY WELD STRENGTH

BASE METAL ALLOY AND TEMPER	**THICKNESS**		**WELDING PROCESS AND CONDITIONS**	**AS-WELDED**					**ONLY AGED AFTER WELDING**					POSTWELD SOLUTION HEAT TREATED AND AGED				
				TENSILE STRENGTH		**YIELD STRENGTH(A)**		**ELONGATION IN 51 MM (2 IN.), %**	**TENSILE STRENGTH**		**YIELD STRENGTH(A)**		**ELONGATION IN 51 MM (2 IN.), %**	**TENSILE STRENGTH**		**YIELD STRENGTH(A)**		ELONGATION IN 51 MM (2 IN.), %
	MM	**IN.**		**MPA**	**KSI**	**MPA**	**KSI**		**MPA**	**KSI**	**MPA**	**KSI**		**MPA**	**KSI**	**MPA**	**KSI**	
6061-T6(B)	3.2	$\frac{1}{8}$	DCEN-GTAW, 15 MM/S (35 IN./MIN)	248	36	165	24	6	. . .	. . .	. . .	. . .	. . .	303	44	276	40	**5**
6061-T4(B)	3.2	$\frac{1}{8}$	DCEN-GTAW, 8.5 MM/S (20 IN./MIN)	234	34	145	21	8	283	41	179	26	3	303	44	276	40	**5**
6061-T6(B)	6.4	$\frac{1}{4}$	AUTOMATIC GMAW, ONE PASS EACH SIDE 17 MM/S (40 IN./MIN)	255	37	138	20	6	. . .	. . .	. . .	. . .	. . .	296	43	276	40	**5**
6061-T4(B)	76	3	AUTOMATIC GMAW, MULTIPASS DOUBLE-V	172	25	90	13	L0	. . .	. . .	. . .	. . .	. . .	234	34	. . .	. . .	**4**
6061-T6(C)	**76**	**3**	**AUTOMATIC GMAW, MULTIPASS DOUBLE-V**	**186**	**27**	**97**	**14**	**13**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**310**	**45**	**276**	**40**	**4**

(A) 0.2% OFFSET IN 50.8 MM (2 IN.) GAGE LENGTH.

(B) 4043 FILLER WAS USED.

(C) 4643 FILLER WAS USED.

Complete re-solution heat treatment is not always practical, because the rapid quench from a high temperature can cause distortion of a welded assembly. An alternate method for increasing the welded strength of heat-treatable alloys is to weld them in the solution heat-treated temper (T4) and age them after welding is complete. To accomplish this effectively, a welding procedure that keeps the heat input relatively low and short in duration should be employed (Ref 8, 9, 10). Tables 7 and 8 compare mechanical properties for postweld solution heat treating and aging versus artificial aging only for 2219 and 6061 alloys. An example of this is highway light poles tapered in a very workable 6*xxx*-T4 alloy, welded to the base, and artificially aged to the T6 temper.

Filler alloys 5183, 5356, 5556, and 5654, which contain over 3% Mg, are not normally recommended for use when parts are to be postweld aged. These high-magnesium content aluminum alloys can be sensitized to stress-corrosion cracking (SCC) when subjected to a relatively long-time precipitation heat treatment (artificial aging) under conditions of high residual stress (Ref 11, 12).

Fillet weld strength represents a significant consideration when selecting a welding filler alloy. Minimum shear strength values for several common filler alloys are listed in Table 9 (Ref 13). Filler alloy 5556 produces the highest fillet weld strength in the as-welded condition. This is nearly twice the shear strength of 4043 fillet welds.

TABLE 9 MINIMUM SHEAR STRENGTHS OF SELECTED FILLET WELDS

FILLER ALLOY	SHEAR STRENGTH			
	LONGITUDINAL		TRANSVERSE	
	MPA	**KSI**	**MPA**	KSI
1100	51.7	7.5	51.7	**7.5**
2319[(A)(B)]	110	16.0	110	**16.0**
2319[(C)]	152	22.0	200	**29.0**
4043	79.3	11.5	103	**15.0**
4643	93.1	13.5	138	**20.0**
5183	128	18.5	193	**28.0**
5356	117	17.0	179	**26.0**
5554	117	17.0	159	**23.0**
5556	138	20.0	207	**30.0**
5654	**82.7**	**12.0**	**124**	**18.0**

(A) AS-WELDED.
(B) ARTIFICIALLY AGED AFTER WELDING.
(C) SOLUTION HEAT-TREATED AND AGED AFTER WELDING

Because several filler alloys can be used to weld many base alloys, the higher-strength weld metal can provide an economic advantage in some applications. For instance, 6061 is commonly welded with 4043 filler, although 5556 filler can also be used. For a symmetrical design requiring a force per unit length of 875 N/mm (5000 lbf/in.) of fillet weld length, an 11 mm ($\frac{7}{16}$ in.) fillet would be required for the 4043 filler versus a 6.4 mm ($\frac{1}{4}$ in.) fillet for 5556 filler. The 6.4 mm ($\frac{1}{4}$ in.) fillet can be made in a single weld pass, but three passes normally would be required for the 11 mm ($\frac{7}{16}$ in.) fillet size. Not only can the 5556 weld be made in one-third of the time, but distortion from the volume of weld metal shrinkage is also minimized by the smaller weld.

The ductility of aluminum welds is excellent for the non-heat-treatable alloys when like filler alloys are used. Using free-bend elongation values as the measure of weld metal ductility, the highest ductility is observed with the pure aluminum alloys. Excellent ductility is also experienced in 5*xxx* series alloys welded with 5*xxx* series fillers. Tables 6 and 7 list the free-bend elongation of butt joints in commonly welded base alloys.

Welds in the heat-treatable alloys do not exhibit as high ductility as the non-heat-treatable alloys. In addition, postweld heat treatments generally reduce ductility, compared to the as-welded condition. About a 50% higher ductility level is obtained when a 5*xxx* series filler is used with 6061 and 6063 base alloys, rather than an aluminum-silicon filler alloy, such as 4043.

Low weld ductility results when welding the aluminum-copper alloys with 4043 or 4145 filler alloys. However, moderate ductility can be obtained when the aluminum-copper alloys are welded with 2319 filler. In addition, 2319 produces the highest weld metal ductility with the aluminum-copper alloys after postweld heat treatments.

Welds in Al-Zn-Mg-Cu alloys, such as 7075 and 7178, possess very low ductility (1% or less). This, in addition to their high sensitivity to cracking when arc welded, generally discourages their use for arc welded applications. The low copper content 7*xxx* series alloys, 7005 and 7039, exhibit good weld ductility with 5183, 5356, and 5556 fillers.

Low ductility is experienced at the weld fusion zone when joining a high silicon content alloy (such as 356.0 casting) to a high magnesium content base alloy (such as 5083). No favorable filler selection is available. If an aluminum-magnesium filler compatible with 5083 is used, then a low-ductility interface of Mg_2Si occurs on the 356.0 side of the weld. Use of an aluminum-silicon filler compatible with the 356.0 casting moves the low-ductility interface to the 5083 side of the weld. A desirable solution is to weld either the 5083 to an aluminum-magnesium casting or the 356.0 (7 Si) casting to a wrought alloy, such as 6061, which is also welded with an aluminum-silicon filler alloy.

Temperature Versus Performance. Aluminum alloys lose a substantial portion of their strength as temperatures exceed 200 °C (390 °F). The aluminum-copper 2*xxx* series alloys exhibit the best elevated-temperature properties, as indicated in Table 10. The 5*xxx* series base and filler alloys with nominal magnesium contents that exceed 3% are not recommended for use at sustained temperatures above 66 °C (150 °F) because they can be sensitized to SCC (Ref 11). Base alloy 5454 and its companion filler alloy 5554 were developed specifically to provide the highest strength among the 5*xxx* alloys for sustained elevated-temperature service. Filler alloy selections suitable for welding common wrought and cast alloys for sustained elevated-temperature service are listed in Table 11.

TABLE 10 ULTIMATE TENSILE STRENGTH AT SELECTED TEMPERATURES FOR GAS-SHIELDED ARC WELDED GROOVE JOINTS IN ALUMINUM ALLOYS

ALLOY AND TEMPER	**FILLER ALLOY**	ULTIMATE TENSILE STRENGTH AT INDICATED TEMPERATURE											
		-184 °C (-300 °F)		**-129 °C (-200 °F)**		**-73 °C (-100 °F)**		**38 °C (100 °F)**		**149 °C (300 °F)**		260 °C (500 °F)	
		MPA	**KSI**	**MPA**	**KSI**	**MPA**	**KSI**	**MPA**	**KSI**	**MPA**	**KSI**	**MPA**	KSI
2219-T37(A)	2319	334	48.5	276	40.0	248	36.0	241	35.0	214	31.0	131	**19.0**
2219(B)	2319	445	64.5	410	59.5	379	55.0	345	50.0	262	38.0	152	**22.0**
3003	1100	190	27.5	148	21.5	121	17.5	97	14.0	66	9.5	34	**5.0**
5052	5356(C)	262	38.0	214	31.0	183	26.5	172	25.0	145	21.0	72	**10.5**
5083	5183	376	54.5	317	46.0	279	40.5	276	40.0	(D)	(D)	(D)	**(D)**
5086	5356	331	48.0	279	40.5	245	35.5	241	35.0	(D)	(D)	(D)	**(D)**
5454	5554	303	44.0	255	37.0	221	32.0	214	31.0	179	26.0	103	**15.0**
5456	5556	386	56.0	328	47.5	293	42.5	290	42.0	(D)	(D)	(D)	**(D)**
6061-T6(A)	4043	238	34.5	207	30.0	183	26.5	165	24.0	138	20.0	41	**6.0**
6061-T6(B)	**4043(E)**	**379**	**55.0**	**341**	**49.5**	**317**	**46.0**	**290**	**42.0**	**217**	**31.5**	**48**	**7.0**

(A) AS WELDED.

(B) POSTWELD SOLUTION HEAT TREATED AND ARTIFICIALLY AGED.

(C) 5554 FILLER ALLOY USED AT >38 °C (> 100 °F).

(D) ALLOY NOT RECOMMENDED FOR USE AT SUSTAINED OPERATING TEMPERATURES ABOVE 66 °C (150 °F).

(E) 4643 FILLER ALLOY FOR [GES] 19 MM ([GES] $\frac{3}{4}$ IN.) BASE METAL

TABLE 11 FILLER ALLOYS FOR WELDING SELECTED ALUMINUM ALLOYS USED FOR SUSTAINED ELEVATED-TEMPERATURE SERVICE

ALUMINUM ALLOYS	ACCEPTABLE

WROUGHT	**CAST**	FILLER ALLOYS
1*XXX* SERIES	. . .	**BASE ALLOY, 1100, 1188, 4043, 4047**
2014, 2219	222.0, 295.0, 319.0, 333.0	**2319, 4043, 4145**
3003, 5005, 5050	. . .	**1100, 4043, 4047**
5052, 5454, 6061, 6063, 7005	. . .	**4043, 4047, 5554**
. . .	354.0, 355.0, C355.0	**4009, 4145, C355.0**
. . .	356.0, A356.0, A357.0, 359.0	**4010, 4011, 4043, 4047, 4643, A356.0, A357.0**
. . .	**413.0, 443.0, A444.0**	**4043, 4047**

With decreasing temperature, aluminum alloys and their weldments gain strength (Table 10). Rather than becoming brittle, the aluminum alloys either maintain or increase in ductility as the temperature decreases below 0 °C. Typical properties for 5083 plate welded with 5183 alloy filler are shown in Fig. 5. The high magnesium content 5*xxx* series alloys have very good cryogenic properties and, because of their excellent weldability, have been used widely for the storage and transport of liquid oxygen and liquid natural gas. Additional information is available in the Section "Selection of Nonferrous Low-Temperature Materials" and the article "Welding for Cryogenic Service" in this Volume.

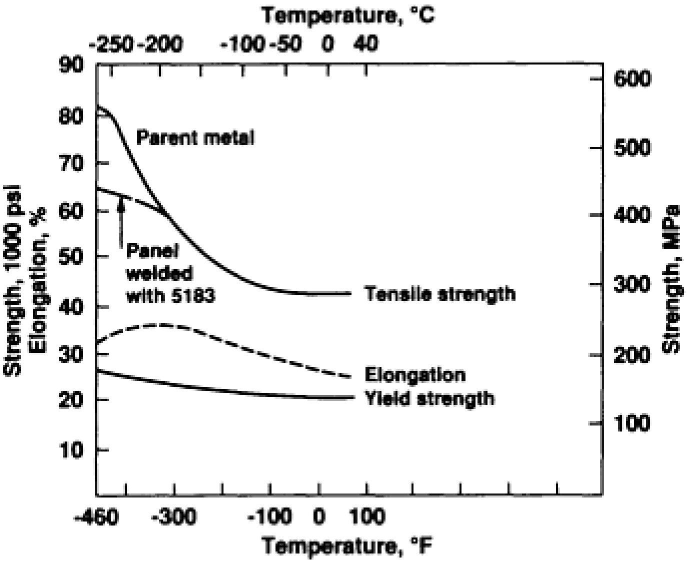

FIG. 5 TYPICAL ROOM-TEMPERATURE AND LOW-TEMPERATURE TENSILE PROPERTIES OF 5083-O BASE ALLOY PLATE WELDED WITH 5183 ALLOY FILLER

Corrosion resistance of the non-heat-treatable alloys is not altered significantly by the heat of welding. The aluminum-magnesium-silicon heat-treatable alloys, such as 6061 and 6063, also have high corrosion resistance in the welded condition. The 2*xxx* and 7*xxx* series heat-treatable alloys, which contain substantial amounts of copper and zinc, respectively, can have their resistance to corrosion altered by the heat of welding. For example, in the aluminum-copper alloys, the HAZ becomes cathodic, whereas in the aluminum-zinc alloys, it becomes anodic to the remainder of the weldment in the presence of water or other electrolytes (Ref 12). The solution potentials across the weld zone for a 5*xxx*, 2*xxx*, and 7*xxx* series weldment are shown in Fig. 6.

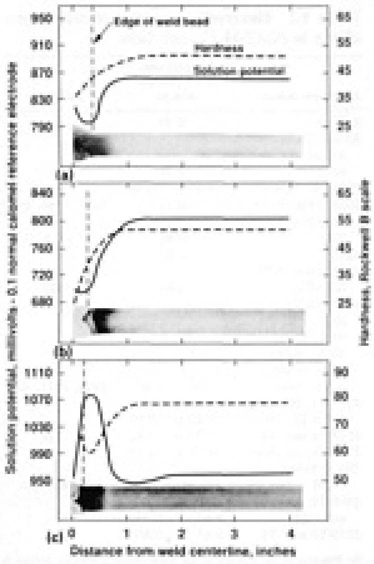

FIG. 6 PLOT OF SOLUTION POTENTIAL AND HARDNESS VERSUS DISTANCE FROM WELD CENTERLINE TO SHOW EFFECT OF THE HEAT OF WELDING ON METALLURGICAL CHANGES IN SELECTED ALUMINUM ALLOYS. (A) 5456-H321 BASE ALLOY WELDED WITH 5556 ALLOY FILLER. GMAW-DCEP, THREE PASSES. (B) 2219-T87 BASE ALLOY WELDED WITH 2319 ALLOY FILLER. GTAW-DCEN, TWO PASSES. (C) 7039-T651 BASE ALLOY WELDED WITH 5183 ALLOY FILLER. GTAW-DCEN, TWO PASSES

Different aluminum alloy compositions produce slightly different electrode potentials in the presence of various solutions. Selective corrosion can result in immersed service, where the base alloy and weld metal possess significant differences in potential. Table 12 lists the solution potentials for common aluminum alloys in a salt solution.

TABLE 12 ELECTRODE POTENTIAL OF ALUMINUM ALLOYS IN NACL-H_2O_2 SOLUTION

ALUMINUM ALLOY(A)	**POTENTIAL VOLTS 0.1 *N* CALOMEL SCALE**(B)	FILLER ALLOY
A712.0	-0.99	. . .
ALCLAD. 3003, ALCLAD. 6061, 7072	-0.96	. . .
7005-T6, 7039-T6	-0.93 TO -0.96	. . .
5083, 5456, 514.0	-0.87	**5183, 5356, 5556**
5154, 5254, 5454, 5086	-0.86	**5554, 5654**
5052	-0.85	. . .
1350, 3004, 5050, 7075-T73	-0.84	**1188**
1100, 3003, 5005, 6061-T6, 6063, ALCLAD 2014, ALCLAD 2024, 413.0, 443.0, A444.0	-0.82 TO -0.83	**1100, 4043, 4047**
6061-T4, 7075-T6, 356.0-T6, 360.0	-0.80 TO -0.81	. . .
2219-T6 AND -T8	-0.79 TO -0.82	**2319**
2014-T6, 355.0-T6	-0.78 TO -0.79	
380.0, 319.0, 333.0	-0.75	**4145**
2014-T4, 2024-T3 AND -T4	-0.68 TO -0.70(C)	. . .
2219-T3 AND -T4	**-0.63 TO -0.65**(C)	. . .

(A) POTENTIAL OF ALL TEMPERS IS THE SAME UNLESS A SPECIFIC TEMPER IS DESIGNATED.

(B) MEASURED IN AN AQUEOUS SOLUTION OF 53 G NACL + 3 G H_2O_2 PER LITER AT 25 °C (77 °F).

(C) POTENTIAL VARIES WITH QUENCHING RATE DURING FABRICATION.

The alloy with the more negative potential in the weldment will attempt to protect the other part. Thus, if the weld metal is anodic to the base metal (as is a 5356 weld in 6061-T6), the small weld can be attacked preferentially to protect the larger surface area of the base metal. The greater the area to be protected and the greater the difference in electrode potential, the more rapidly will corrosive action occur.

Optimum corrosion resistance is obtained when the solution potential of the filler is the same as that of the base alloy, as shown in Table 12 for 4043 filler alloy and 6061-T6. If this is not practical, then a preferred arrangement is to have the larger base alloy surface area be anodic to the weld metal, such as 7005-T6 welded with 5356 filler.

For the welds shown in Fig. 6, the HAZ in the 5*xxx* alloy is mildly cathodic, whereas the 2*xxx* alloy exhibits a greater cathodic differential. The 7*xxx* series HAZ is anodic to the unaffected material and would be of greatest concern. Fabrications in the 7*xxx* alloys are usually painted to avoid galvanic corrosion. However, as an additional safety precaution in some cases, the weld area is metallized with another aluminum alloy to prevent galvanic corrosion if a void occurs in the paint coating. Most unprotected aluminum-base and filler alloy combinations are very satisfactory for general atmospheric conditions.

In some cases, an alloy constituent can be formed by alloying components of the base and filler alloys to produce an anodic zone at the transition of the weld and base metal. If a 5*xxx* alloy is welded with an aluminum-silicon filler, or vice versa, then a magnesium silicide constituent can be formed. For certain immersed conditions, such as a mild acid condition, the magnesium silicide can be highly anodic to all other parts of the weldment (Ref 14, 15). A very selective knifelike corrosive attack can result from this immersed service.

Some chemical exposures or special circumstances can require special controls within the elements of an alloy. In the case of hydrogen peroxide exposure, the manganese and copper impurities have been controlled to low limits in 5652 and 5254 base alloys, as well as 5654 filler alloy. In some cases, a high-purity aluminum alloy is chosen for special exposure. A filler alloy of equal or higher purity to that of the base alloy is generally acceptable in these cases, and filler alloy 1188 would meet most of these requirements.

Weld and Base Metal Color Match. A good color match between the weld metal and the base metal is often desired for ornamental or architectural applications that are given either chemical or electrochemical finishes. The final color of the alloy depends on the chemical composition of the alloy. Ideally, a filler alloy should possess the same composition as the base alloy for an optimum color match. This can be accomplished for most of the non-heat-treatable alloys, but it is not practical with the majority of the heat-treatable aluminum alloys because of weld cracking sensitivity. Thus, dissimilar alloys with the same color response to the final surface treatment are used with these alloys. The two alloying elements of primary interest are silicon and chromium. Silicon in an alloy will create a gray-to-black color, depending on the percentage of silicon (Fig. 7). Thus, welds made with aluminum-silicon filler alloys would exhibit a sharp color contrast with all but the aluminum-silicon base alloys or those clad with an aluminum-silicon alloy. Chromium causes an alloy to take on a yellow or gold shading when anodically treated. Of lesser concern is manganese, which causes a faint gray color.

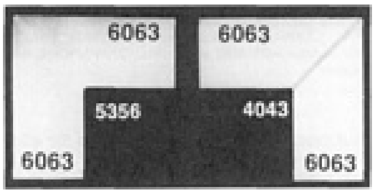

FIG. 7 CLEAR ANODIC COATING ON 6063 WELDS MADE WITH ALUMINUM-MAGNESIUM (5356) AND ALUMINUM-SILICON (4043) FILLERS

The filler alloys that are recommended to obtain a good color match with various base alloys are listed in Table 13. Base alloys 6061 and 6063, which are most commonly welded with 4043 alloy filler, are of the most concern. Because a black weld results when a 4043 weld is anodized, filler alloy 5356 is preferred for the best color match (Fig. 7).

TABLE 13 RECOMMENDED FILLER ALLOYS USED TO OBTAIN BEST COLOR MATCH WITH

ALUMINUM ALLOYS

ALUMINUM ALLOY	FILLER ALLOY
1100, 3003, 5005, 5050	**1188**
5052, 5154	**5654**
6061, 6063, 511.0, 514.0, 535.0	**5356**
5083, 5086, 5454, 5456	**5183 OR 5556**
356.0, A356.0, A357.0, 443.0, A444.0	**4010, 4043, OR 4047**

Filler Alloy Choices. The final determination of the optimal filler alloy for joining an aluminum alloy or combination of alloys should be made after a thorough analysis of the final performance desired from the weldment. Usually, a filler alloy can be selected to meet the service conditions with little or no compromise in the ease-of-welding characteristic. The criteria discussed for an alloy type can be applied to the cast aluminum alloys, as well as to the wrought aluminum alloys.

Table 14 provides a comprehensive matrix for selecting a filler alloy for the GTAW and GMAW processes, based on the various requirements or service conditions presented previously. Commonly welded wrought and cast aluminum alloys are included to assist in selecting a filler for welding an alloy to itself or to another aluminum alloy. The corrosion data are based on performance in fresh or salt water and do not necessarily apply to other exposure conditions. In addition, all ratings are based on as-welded performance and may not apply when the weldment is subjected to postweld thermal treatments.

TABLE 14 RELATIVE RATING OF SELECTED ALUMINUM FILLER ALLOYS USED TO FILLET WELD OR BUTT WELD TWO COMPONENT BASE ALLOYS

Data are for welded assemblies not heat treated after welding.

BASE ALLOYS TO BE JOINED		**FILLER ALLOYS**	FILLER ALLOY CHARACTERISTIC(A)					
ALLOY 1	**ALLOY 2**		**W**	**S**	**D**	**C**	**T**	M
319.0, 333.0, 354.0, 355.0, C355.0, 380.0	1060, 1350	4043	B	A	A	A	A	**A**
		4145	A	A	B	A	A	**A**
	1100	4043	B	A	A	A	A	**A**
		4145	A	A	B	A	A	**A**
	2014, 2036	2319	B	A	A	A	A	**A**
		4043	C	C	B	C	A	**A**
		4145	A	B	C	B	A	**A**
	2219	2319	B	A	A	A	A	**A**
		4043	C	C	B	C	A	**A**
		4145	A	B	C	B	A	**A**
	3003, ALCLAD 3003	4043	B	B	A	A	A	**A**
		4145	A	A	B	A	A	**A**
	3004	4043	B	B	A	A	A	**A**
		4145	A	A	B	A	A	**A**
	ALCLAD 3004	4043	B	B	A	A	A	**A**
		4145	A	A	B	A	A	**A**
	5005, 5050	4043	B	B	A	A	A	**A**
		4145	A	A	B	A	A	**A**
	5052, 5652	4043	A	A	A	A	A	**A**
	5083, 5456	4043	A	A	A	A	. . .	**A**
	5086, 5356	4043	A	A	A	A	. . .	**A**
	514.0, A514.0, B514.0, F514.0, 5154, 5254	4043	A	A	A	A	. . .	**A**
	5454	4043	A	A	A	A	A	**A**
	6005, 6063, 6101, 6151, 6201, 6351, 6951	4043	B	B	A	A	A	**A**
		4145	A	A	B	A	A	**A**
	6061, 6070	4043	B	B	A	A	A	**A**
		4145	A	A	B	A	A	**A**
	7005, 7021, 7039, 7046, 7146,	4043	B	B	A	A	A	**A**

	A712.0, C712.0	4145	A	A	B	A	A	**A**
	413.0, 443.0, 444.0, 356.0, A356.0, A357.0, 359.0	4043	B	B	A	A	A	**A**
		4145	A	A	B	A	A	**A**
	319.0, 333.0, 354.0, 355.0, C355.0, 380.0	2319	B	A	A	A	A	**A**
		4145	A	B	B	B	A	**A**
413.0, 443.0, 444.0, 356.0, A356.0, A357.0, 359.0	1060, 1350	4043	A	A	A	A	A	**A**
		4145	A	A	B	B	A	**. . .**
	1100	4043	A	A	A	A	A	**A**
		4145	A	A	B	B	A	**. . .**
	2014, 2036	4043	B	B	A	A	A	**A**
		4145	A	A	B	A	A	**. . .**
	2219	4043	B	B	A	A	A	**A**
		4145	A	A	B	A	A	**. . .**
	3003, ALCLAD 3003	4043	A	A	A	A	A	**A**
		4145	A	A	B	B	A	**. . .**
	3004	4043	A	A	A	A	A	**A**
	ALCLAD 3004	4043	A	A	A	A	A	**A**
	5005, 5050	4043	A	A	A	A	A	**A**
	5052, 5652	4043	A	B	A	A	A	**A**
		5356	B	A	B	B	. . .	**A**
	5083, 5456	4043	A	B	B	A	. . .	**A**
		5356	A	A	A	A	. . .	**A**
	5086, 5356	4043	A	B	B	A	. . .	**A**
		5356	A	A	A	A	. . .	**A**
	514.0, A514.0, B514.0, F514.0, 5154, 5254	4043	A	B	B	A	. . .	**A**
		5356	A	A	A	B	. . .	**A**
	5454	4043	A	B	B	A	A	**A**
		5356	A	A	A	B	. . .	**A**
	6005, 6063, 6101, 6151, 6201, 6351, 6951	4043	A	B	A	A	A	**A**
		4145	A	A	B	B	A	**. . .**
	6061, 6070	4043	A	B	A	A	A	**A**
		4145	A	A	B	B	A	**. . .**
	7005, 7021, 7039, 7046, 7146, A712.0, C712.0	4043	A	B	B	A	A	**A**
		4145	A	A	B	B	A	**. . .**
		5356	A	A	A	A	. . .	**B**
	413.0, 443.0, 444.0, 356.0, A356.0, A357.0, 359.0	4043	A	B	A	A	A	**A**
		4145	A	A	B	B	A	**. . .**
7005, 7021, 7039, 7046, 7146, A712.0, C712.0	1060, 1350	4043	A	A	C	A	A	**. . .**
		5183	B	A	B	A	. . .	**A**
		5356	B	A	A	A	. . .	**A**
		5556	B	A	B	A	. . .	**A**
	1100	4043	A	A	C	A	A	**. . .**
		5183	B	A	B	A	. . .	**A**
		5356	B	A	A	A	. . .	**A**
		5556	. . .	A	B	A	. . .	**A**
	2014, 2036	4043	B	B	A	A	A	**. . .**
		4145	A	A	B	A	A	**. . .**
	2219	4043	B	B	A	A	A	**. . .**
		4145	A	A	B	A	A	**. . .**
	3003, ALCLAD 3003	4043	A	B	C	A	A	**. . .**
		5183	B	A	B	A	. . .	**A**
		5356	B	A	A	A	. . .	**A**
		5556	B	A	B	A	. . .	**A**
	3004	4043	A	D	C	B	A	**. . .**
		5183	B	A	B	A	. . .	**A**
		5356	B	B	A	A	. . .	**A**
		5554	C	C	A	A	A	**A**
		5556	B	A	B	A	. . .	**A**
		5654	C	C	A	A	. . .	**B**
	ALCLAD 3004	4043	A	D	C	B	A	**. . .**
		5183	B	A	B	A	. . .	**A**

		5356	B	B	A	A	. . .	**A**
		5554	C	C	A	A	A	**A**
		5556	B	A	B	A	. . .	**A**
		5654	C	C	A	A	. . .	**B**
	5005, 5050	4043	A	B	C	B	A	**. . .**
		5183	B	A	B	A	. . .	**A**
		5356	B	A	A	A	. . .	**A**
		5554	C	A	A	A	A	**A**
		5556	B	A	B	A	. . .	**A**
		5654	C	A	A	A	. . .	**A**
	5052, 5652	4043	B	D	C	B	A	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5554	B	C	A	A	A	**A**
		5556	A	A	B	A	. . .	**A**
		5654	B	C	A	A	. . .	**A**
	5083, 5456	5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
	5086, 5356	5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
	514.0, A514.0, B514.0, F514.0, 5154, 5254	5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5554	B	C	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
		5654	B	C	A	A	. . .	**A**
	5454	5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5554	B	C	A	A	A	**A**
		5556	A	A	B	A	. . .	**A**
		5654	B	C	A	A	. . .	**A**
	6005, 6063, 6101, 6151, 6201, 6351, 6951	4043	A	D	C	B	A	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5554	B	C	A	A	A	**A**
		5556	A	A	B	A	. . .	**A**
		5654	B	C	A	A	. . .	**A**
	6061, 6070	4043	A	D	C	B	A	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5554	B	C	A	A	A	**A**
		5556	A	A	B	A	. . .	**A**
		5654	B	C	A	A	. . .	**A**
	7005, 7021, 7039, 7046, 7146, A712.0, C712.0	4043	B	D	C	B	A	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	A	**A**
		5554	B	C	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
		5654	B	C	A	A	. . .	**A**
6061, 6070	1060, 1350	4043	A	A	C	A	A	**. . .**
		4145	A	A	D	B	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	A	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	1100	4043	A	A	C	A	A	**. . .**
		4145	A	A	D	B	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	A	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	2014, 2036	4043	B	B	A	A	A	**. . .**

		4145	A	A	B	A	A	**. . .**
	2219	4043	B	B	A	A	A	**. . .**
		4145	A	A	B	A	A	**. . .**
	3003, ALCLAD 3003	4043	A	B	C	A	A	**. . .**
		4145	A	A	D	B	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	A	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	3004	4043	A	D	C	A	A	**. . .**
		4145	B	C	D	B	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	B	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	ALCLAD 3004	4043	A	D	C	A	A	**. . .**
		4145	B	C	D	B	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	B	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	5005, 5050	4043	A	B	C	A	A	**. . .**
		4145	A	B	D	B	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	A	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	5052, 5652	4043	A	D	C	A	A	**. . .**
		5183	B	A	B	C	. . .	**B**
		5356	B	B	A	C	. . .	**A**
		5554	C	C	A	B	A	**B**
		5556	B	A	B	C	. . .	**B**
		5654	C	C	A	B	. . .	**A**
	5083, 5456	4043	A	D	C	A	. . .	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5554	B	C	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
		5654	B	C	A	A	. . .	**B**
	5086, 5356	4043	A	D	C	A	. . .	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5554	B	C	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
		5654	B	C	A	A	. . .	**B**
	514.0, A514.0, B514.0, F514.0, 5154, 5254	4043	A	D	C	A	. . .	**. . .**
		5183	B	A	B	C	. . .	**B**
		5356	B	B	A	C	. . .	**A**
		5554	C	C	A	B	. . .	**B**
		5556	B	A	B	C	. . .	**B**
		5654	C	C	A	B	. . .	**A**
	5454	4043	A	D	C	B	A	**. . .**
		5183	B	A	B	C	. . .	**A**
		5356	B	B	A	C	. . .	**A**
		5554	C	C	A	A	A	**A**
		5556	B	A	B	C	. . .	**A**
		5654	C	C	A	B	. . .	**B**
	6005, 6063, 6101, 6151, 6201, 6351, 6951	4043	A	C	B	A	A	**. . .**
		5183	B	A	A	C	. . .	**A**
		5356	B	A	A	C	. . .	**A**
		5554	C	B	A	B	B	**A**
		5556	B	A	A	C	. . .	**A**
		5654	C	B	A	B	. . .	**B**
	6061, 6070	4043	A	C	B	A	A	**. . .**
		5183	B	A	A	C	. . .	**B**

		5356	B	B	A	C	. . .	**A**
		5554	C	B	A	B	B	**B**
		5556	B	A	A	C	. . .	**B**
		5654	C	B	A	B	. . .	**B**
6005, 6063, 6101, 6151, 6201, 6351, 6951	1060, 1350	4043	A	A	C	A	A	**. . .**
		4145	A	A	D	B	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	A	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	1100	4043	A	A	C	A	A	**. . .**
		4145	A	A	D	B	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	A	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	2014, 2036	4043	B	B	A	A	A	**. . .**
		4145	A	A	B	A	A	**. . .**
	2219	4043	B	B	A	A	A	**. . .**
		4145	A	A	B	A	A	**. . .**
	3003, ALCLAD 3003	4043	A	B	C	A	A	**. . .**
		4145	A	A	D	B	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	A	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	3004	4043	A	D	C	A	A	**. . .**
		4145	B	C	D	B	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	B	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	ALCLAD 3004	4043	A	D	C	A	A	**. . .**
		4145	B	C	D	B	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	B	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	5005, 5050	4043	A	B	C	A	A	**. . .**
		4145	A	B	D	B	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	A	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	5052, 5652	4043	A	D	C	A	A	**. . .**
		5183	B	A	B	C	. . .	**B**
		5356	B	B	A	C	. . .	**A**
		5554	C	C	A	B	A	**B**
		5556	B	A	B	C	. . .	**B**
		5654	C	C	A	B	. . .	**A**
	5083, 5456	4043	A	B	C	A	. . .	**. . .**
		5183	A	A	B	A	. . .	**A**
		5354	A	A	A	A	. . .	**A**
		5556	B	A	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
		5654	B	A	A	A	. . .	**B**
	5086, 5356	4043	A	B	C	A	. . .	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	A	A	A	. . .	**A**
		5554	B	A	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
		5654	B	A	A	A	. . .	**B**
	514.0, A514.0, B514.0, F514.0, 5154, 5254	4043	A	B	C	A	. . .	**. . .**
		5183	B	A	B	C	. . .	**A**
		5356	B	A	A	C	. . .	**A**
		5554	C	A	A	B	. . .	**A**
		5556	B	A	B	C	. . .	**A**

		5654	C	A	A	B	. . .	**B**
	5454	4043	A	B	C	B	A	**. . .**
		5183	B	A	B	C	. . .	**A**
		5356	B	A	A	C	. . .	**A**
		5554	C	A	A	A	A	**A**
		5556	B	A	B	C	. . .	**A**
		5654	C	A	A	B	. . .	**B**
	6005, 6063, 6101, 6151, 6201, 6351, 6951	4043	A	C	B	A	A	**. . .**
		5183	B	A	A	C		**A**
		5356	B	A	A	C	. . .	**A**
		5554	C	B	A	B	B	**A**
		5556	B	A	A	C	. . .	**A**
		5654	C	B	A	B	. . .	**B**
5454	1060, 1350	4043	A	B	C	C	A	**. . .**
		5183	B	A	B	B	. . .	**A**
		5356	B	A	A	B	. . .	**A**
		5554	C	A	A	A	A	**A**
		5556	B	A	B	B	. . .	**A**
	1100	4043	A	B	C	C	A	**. . .**
		5183	B	A	B	B	. . .	**A**
		5356	B	A	A	B	. . .	**A**
		5554	C	A	A	A	A	**A**
		5556	B	A	B	B	. . .	**A**
	2014, 2036	(B)	. . .	. . .	. . .	. . .	. . .	**. . .**
	2219	4043	A	A	A	A	A	**. . .**
	3003, ALCLAD 3003	4043	A	B	C	C	A	**. . .**
		5183	B	A	B	B	. . .	**A**
		5356	B	A	A	B	. . .	**A**
		5554	C	A	A	A	A	**A**
		5556	B	A	B	B	. . .	**A**
	3004	4043	A	D	C	C	A	**. . .**
		5183	B	A	B	B	. . .	**A**
		5356	B	B	A	B	. . .	**A**
		5554	C	C	A	A	A	**A**
		5556	B	A	B	B	. . .	**A**
	ALCLAD 3004	4043	A	D	C	C	A	**. . .**
		5183	B	A	B	B	. . .	**A**
		5356	B	B	A	B	. . .	**A**
		5554	C	C	A	A	A	**A**
		5556	B	A	B	B	. . .	**A**
	5005, 5050	4043	A	B	C	C	A	**. . .**
		5183	B	A	B	B	. . .	**A**
		5356	B	A	A	B	. . .	**A**
		5554	C	A	A	A	A	**A**
		5556	B	A	B	B	. . .	**A**
	5052, 5652	4043	A	D	C	C	A	**. . .**
		5183	A	A	A	B	. . .	**A**
		5356	A	B	A	B	. . .	**A**
		5554	C	C	A	A	A	**A**
		5556	A	A	B	B	. . .	**A**
		5654	B	C	A	B	. . .	**B**
	5083, 5456	5183	A	A	B	B	. . .	**A**
		5356	A	B	A	B	. . .	**A**
		5554	B	C	A	A	. . .	**A**
		5556	A	A	B	B	. . .	**A**
	5086, 5356	5183	A	A	B	B	. . .	**A**
		5356	A	B	A	B	. . .	**A**
		5554	B	C	A	A	. . .	**A**
		5556	A	A	B	B	. . .	**A**
	514.0, A514.0, B514.0, F514.0, 5154, 5254	5183	A	A	B	B	. . .	**A**
		5356	A	B	A	B	. . .	**A**

		5554	B	C	A	A	. . .	**A**
		5556	A	A	B	B	. . .	**A**
		5654	B	C	A	A	. . .	**B**
	5454	5183	A	A	B	B	. . .	**A**
		5356	A	B	A	B	. . .	**A**
		5554	B	C	A	A	A	**A**
		5556	A	A	B	B	. . .	**A**
		5654	B	C	A	B	. . .	**B**
514.0, A514.0, B514.0, F514.0, 5154, 5254	1060, 1350	4043	A	B	C	C	. . .	**. . .**
		5183	B	A	B	B	. . .	**A**
		5356	B	A	A	B	. . .	**A**
		5554	C	A	A	A	. . .	**A**
		5556	B	A	B	B	. . .	**A**
		5654	C	A	A	A	. . .	**B**
	1100	4043	A	B	C	C	. . .	**. . .**
		5183	B	A	B	B	. . .	**A**
		5356	B	A	A	B	. . .	**A**
		5554	C	A	A	A	. . .	**A**
		5556	B	A	B	B	. . .	**A**
		5654	C	A	A	A	. . .	**B**
	2014, 2036	(B)	. . .	. . .	. . .	. . .	. . .	**. . .**
	2219	4043	A	A	A	A	. . .	**. . .**
	3003, ALCLAD 3003	4043	A	B	C	C	. . .	**. . .**
		5183	B	A	B	B	. . .	**A**
		5356	B	A	A	B	. . .	**A**
		5554	C	A	A	A	. . .	**A**
		5556	B	A	B	B	. . .	**A**
		5654	C	A	A	A	. . .	**B**
	3004	4043	A	D	C	C	. . .	**. . .**
		5183	B	A	B	B	. . .	**A**
		5356	B	B	A	B	. . .	**A**
		5554	C	C	A	A	. . .	**A**
		5556	B	A	B	B	. . .	**A**
		5654	C	C	A	A	. . .	**B**
	ALCLAD 3004	4043	A	D	C	C	. . .	**. . .**
		5183	B	A	B	B	. . .	**A**
		5356	B	B	A	B	. . .	**A**
		5554	C	C	A	A	. . .	**A**
		5556	B	A	B	B	. . .	**A**
		5654	C	C	A	A	. . .	**B**
	5005, 5050	4043	A	B	C	C	. . .	**. . .**
		5183	B	A	B	B	. . .	**A**
		5356	B	A	A	B	. . .	**A**
		5554	C	A	A	A	. . .	**A**
		5556	B	A	B	B	. . .	**A**
		5654	C	A	A	A	. . .	**B**
	5052, 5652	4043	A	D	C	C	. . .	**. . .**
		5183	A	A	B	B	. . .	**B**
		5356	A	B	A	B	. . .	**A**
		5554	C	C	A	A	. . .	**B**
		5556	A	A	B	B	. . .	**B**
		5654	B	C	A	A	. . .	**A**
	5083, 5456	5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5554	B	C	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
		5654	B	C	A	A	. . .	**B**
	5086, 5356	5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5554	B	C	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**

		5654	B	C	A	A	. . .	**B**
	514.0, A514.0, B514.0, F514.0, 5154, 5254	5183	A	A	B	B	. . .	**B**
		5356	A	B	A	B	. . .	**A**
		5554	B	C	A	A	. . .	**B**
		5556	A	A	B	B	. . .	**B**
		5654	B	C	A	A	. . .	**A**
5086, 5356	1060, 1350	4043	A	B	C	B	. . .	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	A	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
	1100	4043	A	B	C	B	. . .	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	A	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
	2014, 2036	(B)	. . .	. . .	. . .	. . .	. . .	**. . .**
	2219	4043	A	A	A	A	. . .	**. . .**
	3003, ALCLAD 3003	4043	A	B	C	B	. . .	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	A	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
	3004	4043	A	C	C	B	. . .	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
	ALCLAD 3004	4043	A	C	C	B	. . .	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
	5005, 5050	4043	A	B	C	B	. . .	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	A	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
	5052, 5652	5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5554	C	C	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
		5654	B	C	A	A	. . .	**B**
	5083, 5456	5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
	5086, 5356	5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
5083, 5456	1060, 1350	4043	A	B	C	B	. . .	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	A	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
	1100	4043	A	B	C	B	. . .	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	A	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
	2014, 2036	(B)	. . .	. . .	. . .	. . .	. . .	**. . .**
	2219	4043	A	A	A	A	. . .	**. . .**
	3003, ALCLAD 3003	4043	A	B	C	B	. . .	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	A	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
	3004	4043	A	C	C	B	. . .	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**

	ALCLAD 3004	4043	A	C	C	B	. . .	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
	5005, 5050	4043	A	B	C	B	. . .	**. . .**
		5183	A	A	B	A	. . .	**A**
		5356	A	A	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
	5052, 5652	5183	A	A	B	A	. . .	**A**
		5356	A	B	A	A	. . .	**A**
		5554	C	C	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
		5654	B	C	A	A	. . .	**B**
	5083, 5456	5183	A	A	B	A	. . .	**A**
		5356	A	. . .	A	A	. . .	**A**
		5556	A	A	B	A	. . .	**A**
5052, 5652	1060, 1350	4043	A	B	C	A	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	A	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	1100	4043	A	B	C	A	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	A	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	2014, 2036	4043	A	A	A	A	A	**. . .**
	1219	4043	A	A	A	A	A	**. . .**
	3003, ALCLAD 3003	4043	A	B	C	A	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	A	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	3004	4043	A	C	C	A	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	B	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	ALCLAD 3004	4043	A	C	C	A	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	B	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	5005, 5050	4043	A	B	C	A	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	A	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	5052, 5652	4043	A	D	C	B	A	**. . .**
		5183	A	A	B	C	. . .	**B**
		5356	A	B	A	C	. . .	**A**
		5554	C	C	A	A	A	**B**
		5556	A	A	B	C	. . .	**B**
		5654	B	C	A	B	. . .	**A**
5005, 5050	1060, 1350	1100	C	B	A	A	A	**A**
		4043	A	A	C	A	A	**. . .**
		4145	B	A	D	B	A	**. . .**
		5183	C	A	B	. . .	. . .	**B**
		5356	C	A	B	. . .	. . .	**B**
		5556	C	A	B	. . .	. . .	**B**
	1100	1100	C	B	A	A	A	**A**
		4043	A	A	C	A	A	**. . .**
		4145	B	A	D	B	A	**. . .**
		5183	C	A	B	. . .	. . .	**B**
		5356	C	A	B	. . .	. . .	**B**
		5556	C	A	B	. . .	. . .	**B**
	2014, 2036	4043	B	B	A	A	A	**. . .**

		4145	A	A	B	A	A	**. . .**
	2219	4043	B	B	A	A	A	**. . .**
		4145	A	A	B	A	A	**. . .**
	3003, ALCLAD 3003	1100	C	C	A	A	A	**A**
		4043	A	B	C	A	A	**. . .**
		4145	B	B	D	B	A	**. . .**
		5183	C	A	B	C	. . .	**B**
		5356	C	A	B	C	. . .	**B**
		5556	C	A	B	C	. . .	**B**
	3004	4043	A	B	C	A	A	**. . .**
		5183	B	A	B	. . .	. . .	**A**
		5356	B	A	A	. . .	. . .	**A**
		5556	B	A	B	. . .	. . .	**A**
	ALCLAD 3004	4043	A	B	C	A	A	**. . .**
		5183	B	A	B	B	. . .	**A**
		5356	B	A	A	B	. . .	**A**
		5556	B	A	B	B	. . .	**A**
	5005, 5050	1100	B	. . .	A	A	A	**A**
		4043	A	B	D	A	A	**. . .**
		5183	B	A	C	. . .	. . .	**B**
		5356	B	A	B	. . .	. . .	**B**
		5556	B	A	C	. . .	. . .	**B**
ALCLAD 3004	1060, 1350	1100	D	B	A	A	A	**A**
		4043	A	A	C	A	A	**. . .**
		4145	B	A	D	B	A	**. . .**
		5183	C	A	B	C	. . .	**B**
		5356	C	A	B	C	. . .	**B**
		5556	C	A	B	C	. . .	**B**
	1100	1100	D	B	A	A	A	**A**
		4043	A	A	C	A	A	**. . .**
		4145	B	A	D	B	A	**. . .**
		5183	C	A	B	C	. . .	**B**
		5356	C	A	B	C	. . .	**B**
		5556	C	A	B	C	. . .	**B**
	2014, 2036	4043	B	B	A	A	A	**. . .**
		4145	A	A	B	A	A	**. . .**
	2219	4043	B	B	A	A	A	**. . .**
		4145	A	A	B	A	A	**. . .**
	3003, ALCLAD 3003	1100	C	C	A	A	A	**A**
		4043	A	B	C	A	A	**. . .**
		4145	B	B	D	B	A	**. . .**
		5183	C	A	B	C	. . .	**A**
		5356	C	A	B	C	. . .	**A**
		5556	C	A	B	C	. . .	**A**
	3004	4043	A	D	D	A	A	**. . .**
		5183	B	A	C	C	. . .	**A**
		5356	B	B	B	C	. . .	**A**
		5554	C	C	A	B	A	**A**
		5556	B	A	C	C	. . .	**A**
	ALCLAD 3004	4043	A	D	D	A	A	**. . .**
		5183	B	A	C	C	. . .	**A**
		5356	B	B	B	C	. . .	**A**
		5554	C	C	A	B	A	**A**
		5556	B	A	C	C	. . .	**A**
3004	1060, 1350	1100	D	B	A	A	A	**A**
		4043	A	A	C	A	A	**. . .**
		4145	B	A	D	B	A	**. . .**
		5183	C	A	B	. . .	. . .	**B**
		5356	C	A	B	. . .	. . .	**B**
		5556	C	A	B	. . .	. . .	**B**
	1100	1100	D	B	A	A	A	**A**

		4043	A	A	C	A	A	. . .
		4145	B	A	D	B	A	. . .
		5183	C	A	B	. . .	. . .	B
		5356	C	A	B	. . .	. . .	B
		5556	C	A	B	. . .	. . .	B
	2014, 2036	4043	B	B	A	A	A	. . .
		4145	A	A	B	A	A	. . .
	2219	4043	B	B	A	A	A	. . .
		4145	A	A	B	A	A	. . .
	3003, ALCLAD 3003	1100	C	C	A	A	A	A
		4043	A	B	C	A	A	. . .
		4145	B	B	D	B	A	. . .
		5183	C	A	B	C	. . .	A
		5356	C	A	B	C	. . .	A
		5556	C	A	B	C	. . .	A
	3004	4043	A	B	D	A	A	. . .
		5183	B	A	C	C	. . .	A
		5356	B	B	B	C	. . .	A
		5554	C	C	A	B	A	A
		5556	B	A	C	C	. . .	A
3003, ALCLAD 3003	1060, 1350	1100	B	B	A	A	A	A
		4043	A	A	B	A	A	. . .
		4145	A	A	C	B	A	. . .
	1100	1100	B	B	A	A	A	A
		4043	A	A	B	A	A	. . .
		4145	A	A	C	B	A	. . .
	2014, 2036	4043	B	A	A	A	A	. . .
		4145	A	A	B	A	A	. . .
	2219	4043	B	A	A	A	A	. . .
		4145	A	A	B	A	A	. . .
	3003, ALCLAD 3003	1100	B	B	A	A	A	A
		4043	A	A	B	A	A	. . .
		4145	A	A	C	B	A	. . .
2219	1060, 1350	4043	B	A	A	A	A	. . .
		4145	A	A	B	A	A	. . .
	1100	4043	B	A	A	A	A	. . .
		4145	A	A	B	A	A	. . .
	2014, 2036	2319	B	A	A	A	A	A
		4043	B	C	B	C	A	. . .
		4145	A	B	C	B	A	. . .
	2219	2319	A	A	A	A	A	A
		4043	B	C	B	C	A	. . .
		4145	A	B	C	B	A	. . .
2014, 2036	1060, 1350	4043	B	A	A	A	A	. . .
		4145	A	A	B	A	A	. . .
	1100	4043	B	A	A	A	A	. . .
		4145	A	A	B	A	A	. . .
	2014, 2036	2319	C	A	A	A	A	A
		4043	B	C	B	C	A	. . .
		4145	A	B	C	B	A	. . .
1100	1060, 1350	1100	B	B	A	A	A	A
		4043	A	A	B	A	A	. . .
	1100	1100	B	B	A	A	A	. . .
		4043	A	A	B	A	A	. . .
1060, 1350	1060, 1350	1100	B	B	A	A	A	B
		1188	C	C	A	A	A	A
		4043	A	A	B	A	A	. . .
		5554	C	C	A	B	A	A
		5556	**B**	**A**	**C**	**C**	**. . .**	**A**

Note: Combinations having no ratings are not recommended.

Source: Aluminum Company of America

(A) A, B, C, AND D REPRESENT RELATIVE RATINGS (WHERE A IS BEST AND D IS WORST) OF THE PERFORMANCE OF THE TWO COMPONENT BASE ALLOYS COMBINED WITH EACH GROUP OF SELECTED FILLER ALLOYS. W, EASE OF WELDING (RELATIVE FREEDOM FROM WELD CRACKING); S, STRENGTH OF WELDED JOINT IN AS-WELDED CONDITION (RATING APPLIES SPECIFICALLY TO FILLET WELDS, BUT ALL RODS AND ELECTRODES RATED WILL DEVELOP PRESENTLY SPECIFIED MINIMUM STRENGTHS FOR BUTT WELDS); D, DUCTILITY (RATING BASED ON FREE BEND ELONGATION OF THE WELD); C, CORROSION RESISTANCE IN CONTINUOUS OR ALTERNATE IMMERSION OF FRESH OR SALT WATER; T, PERFORMANCE IN SERVICE AT SUSTAINED TEMPERATURES >65 °C (>150 °F); M, COLOR MATCH AFTER ANODIZING.

(B) NO FILLER SUITABLE.

Quality control requirements for bare aluminum filler rod and electrode are stated in Ref 16. To maintain the quality level to which they are produced, the filler wire must be protected during shipment, storage, handling, and use. The spooled filler wire, which is most critical, is normally packaged in a sealed plastic bag and carton. The best storage place is a dry, heated area. If stored in an unheated area, then the packaged product should be brought to the production area and allowed to reach that ambient temperature before the carton is opened. If the temperature of the wire is more than 9 °C (16 °F) below the room temperature when the package is opened, moisture in the atmosphere can condense on the wire, creating a hydrated oxide. The 5*xxx* (aluminum-magnesium) series filler alloys are most sensitive to this situation and can be ruined after just a few seconds of exposure, because the hydrated oxide can cause excessive porosity in the weld. For the same reason, the spooled product should be removed from the welding apparatus at the completion of work and stored in a cabinet where the relative humidity is kept below 35%. A small resistance heating unit in an electrical socket on the inside of a sealed cabinet will accomplish this. If the product is not removed from the equipment, a small resistance heater inside the spool enclosure can also be effective.

References cited in this section

1. J.D. DOWD, WELD CRACKING OF ALUMINUM ALLOYS, *WELD. J.*, RESEARCH SUPPLEMENT, VOL 31, 1952, P 448S-456S
2. J.H. DUDAS AND F.R. COLLINS, PREVENTING WELD CRACKS IN HIGH-STRENGTH ALUMINUM ALLOYS, *WELD. J.*, RESEARCH SUPPLEMENT, VOL 45 (NO. 6), JUNE 1966, P 241S-249S
3. C.E. CROSS, PH.D. THESIS, COLORADO SCHOOL OF MINES, 1986, P 144
4. A.R.E. SINGER AND P.H. JENNINGS, HOT SHORTNESS OF THE ALUMINUM-SILICON ALLOYS OF COMMERCIAL QUALITY, *J. INST. MET.*, VOL 73, 1947, P 197-212
5. W.I. PUMPHREY AND J.V. LYONS, CRACKING DURING THE CASTING AND WELDING OF THE MORE COMMON BINARY ALUMINUM ALLOYS OF COMMERCIAL QUALITY, *J. INST. MET.*, VOL 74, 1948, P 439-455
6. P.H. JENNINGS, A.R.E. SINGER, AND W.I. PUMPHREY, HOT-SHORTNESS OF SOME HIGH PURITY ALLOYS IN THE SYSTEMS ALUMINUM-COPPER AND ALUMINUM-MAGNESIUM-SILICON, *J. INST. MET.*, VOL 74, 1948, P 227-248
7. N.F. GITTOS AND M.H. SCOTT, HEAT AFFECTED ZONE CRACKING OF AL-MG-SI ALLOYS, *WELD. J.*, RESEARCH SUPPLEMENT, JUNE 1981, P 95S-103S
8. W.L. BURCH, THE EFFECT OF WELDING SPEED ON THE STRENGTH OF 6061-T4 ALUMINUM JOINTS, *WELD. J.*, RESEARCH SUPPLEMENT, VOL 37 (NO. 8), AUG 1958, P 361S-367S
9. G.E. METZGER, SOME MECHANICAL PROPERTIES OF WELDS IN 6061 ALUMINUM ALLOY SHEET, *WELD. J.*, RESEARCH SUPPLEMENT, VOL 46 (NO. 10), OCT 1967, P 457S-469S
10. F.R. COLLINS, F.G. NELSON, AND I.B. ROBINSON, ALUMINUM 2219--NEW ALLOY FOR HIGH-STRENGTH WELDED STRUCTURES, *MET. PROG.*, NOV 1961
11. E.H. DIX, W.H. ANDERSON, AND M.B. SHUMAKER, INFLUENCE OF SERVICE TEMPERATURE

ON THE RESISTANCE OF WROUGHT ALUMINUM-MAGNESIUM ALLOYS TO CORROSION, *CORROSION*, VOL 15 (NO. 2), FEB 1959, P 55T-62T

12. M.B. SHUMAKER, R.A. KELSEY, D.O. SPROWLS, AND J.G. WILLIAMSON, "EVALUATION OF VARIOUS TECHNIQUES FOR STRESS-CORROSION TESTING WELDED ALUMINUM ALLOYS," PRESENTED AT ASTM STRESS-CORROSION TESTING SYMPOSIUM, JUNE 1966
13. F.G. NELSON AND R.L. ROLF, SHEAR STRENGTHS OF ALUMINUM ALLOY FILLET WELDS, *WELD. J.,* RESEARCH SUPPLEMENT, VOL 45 (NO. 2), FEB 1966, P 82S-84S
14. S.L. WÖHLER AND O. SCHLIEPHAKE, ÜBEREINEM GERS., DURCH POTENTIALMESSUNGER DAS AUFGREGEN DER VERB. MG_2SI ZU BESTÄTIGEN, *Z. ANORG. CHEM.*, 1962, P 151
15. G.W. AKIMOW AND A.S. OLESCHKO, GMELINS HANDBUCH DER ANORGANISCHEN CHEMIE, (NO. 8), AUFLAGE, 1942, P 393
16. "SPECIFICATION FOR BARE ALUMINUM AND ALUMINUM ALLOY WELDING ELECTRODE AND ROD," ANSI/AWS A5.10-92, AMERICAN WELDING SOCIETY, 1992

Welding of Aluminum Alloys

Paul B. Dickerson, Aluminum Company of America (retired)

Preparation for Welding

Joint Designs. Typical joint geometries for ac-GTAW and GMAW processes used with aluminum are shown in Fig. 8. Because good penetration control can be obtained with GTAW, welds can be made from a single side without backing. When mechanizing the GTAW process and when using the GMAW process by manual or mechanical means, the use of a temporary grooved backing is very helpful in controlling the uniformity of penetration for single-sided welds. When welding from both sides, a temporary support of the initial weld pass can be given by an angle or rod when a thin land is employed, or the land can be thickened to provide support. In either case, back gouging to sound weld metal is required before application of the initial root pass.

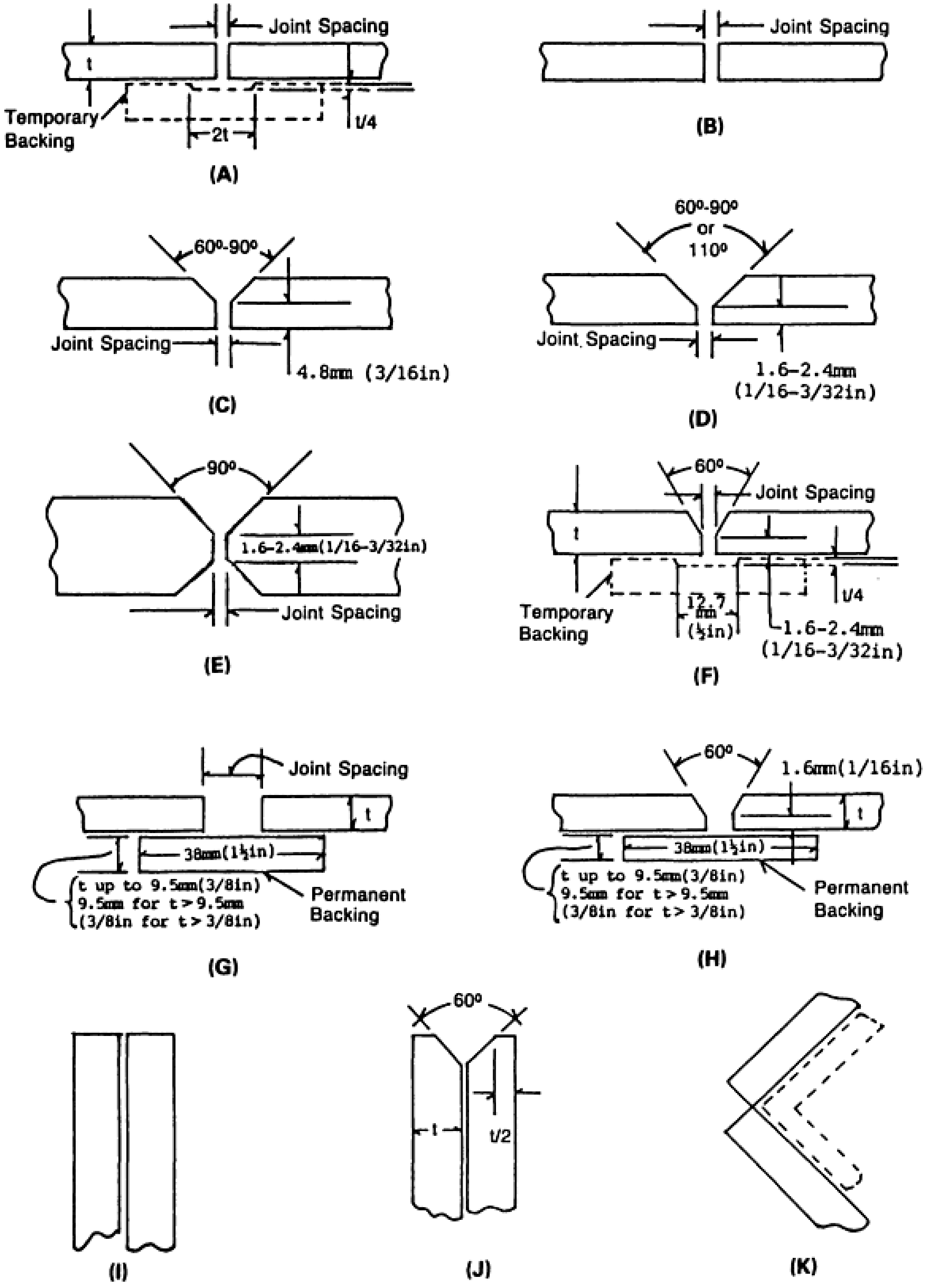
Joint Spacing
t
Temporary Backing
2t
t/4
(A)
Joint Spacing
(B)
60°-90°
Joint Spacing
4.8mm (3/16in)
(C)
60°-90° or 110°
Joint Spacing
1.6-2.4mm (1/16-3/32in)
(D)
90°
1.6-2.4mm(1/16-3/32in)
Joint Spacing
(E)
60°
Joint Spacing
t
Temporary Backing
12.7 mm (½in)
t/4
1.6-2.4mm (1/16-3/32in)
(F)
Joint Spacing
t
38mm(1½in)
t up to 9.5mm(3/8in)
9.5mm for t > 9.5mm
(3/8in for t > 3/8in)
Permanent Backing
(G)
60°
1.6mm(1/16in)
t
38mm(1½in)
t up to 9.5mm(3/8in)
9.5mm for t > 9.5mm
(3/8in for t > 3/8in)
Permanent Backing
(H)
(I)
60°
t
t/2
(J)
(K)

FIG. 8 TYPICAL JOINT GEOMETRIES USED FOR AC-GTAW AND GMAW OF ALUMINUM COMPONENTS. LETTERS CORRESPOND TO EDGE PREPARATIONS IN TABLES 15, 17, AND 18. SOURCE: AMERICAN WELDING SOCIETY

A 60° included "V" bevel is standard for GMAW. For GTAW, the bevel varies from 60° to 90°, and for overhead groove welds, a 110° minimum "V" bevel is required for root pass control. The large angles are needed to allow the gas nozzle access to the joint to provide adequate inert-gas shielding of the root passes. For proper nozzle access with 25 mm (1 in.)

and thicker material, single "U" and double "U" bevels are required. A special joint design used for aluminum pipe and other single-sided access joints, which permits uniform and complete penetration welds without backing, is shown in Fig. 9 (Ref 17).

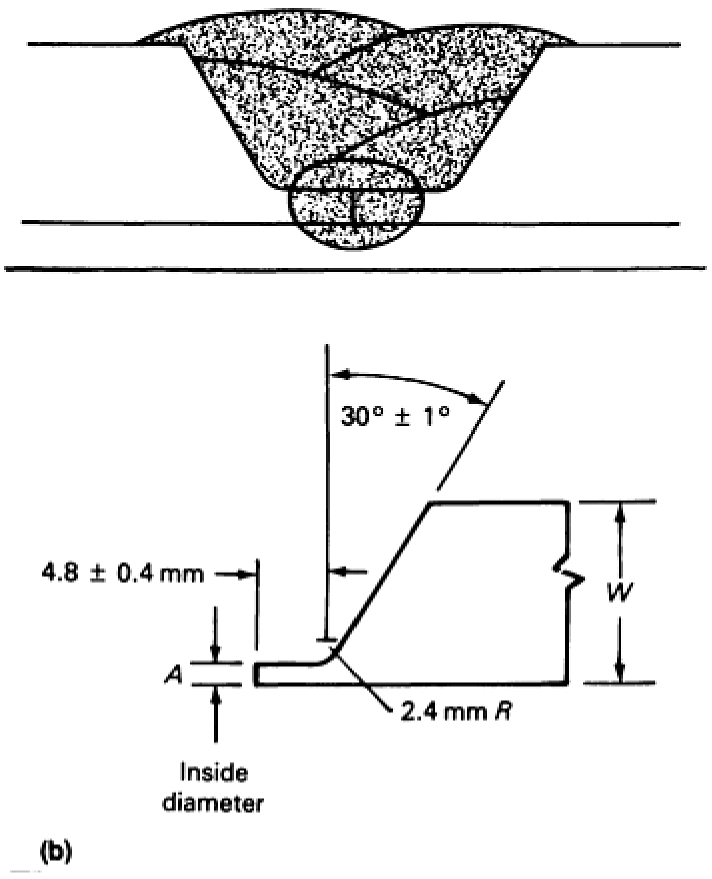

PIPE SIZE		*A*		W_{MAX}	
MM	IN.	MM	IN.	MM	IN.
3.2-64	$\frac{1}{8}$ -2 $\frac{1}{2}$	1.6 ± 0.4	$\frac{1}{16} \pm \frac{1}{64}$	7.01	**0.276**
76-305	**3-12**	**2.4 ± 0.4**	$\frac{3}{32} \pm \frac{1}{64}$	**12.7**	**0.500**

FIG. 9 EXTENDED-LAND BEVEL USED TO WELD ALUMINUM PIPE. (A) TYPICAL MULTIPLE-PASS WELD USED TO JOIN PIPE WITH EXTENDED-LAND BEVEL. (B) RECOMMENDED EXTENDED-LAND BEVEL SPECIFICATIONS FOR 3.2 TO 305 MM ($\frac{1}{8}$ TO 12 IN.) DIAM PIPE WITH 7.01 TO 12.7 MM (0.276 TO 0.500 IN.) WALL THICKNESS. SOURCE: AMERICAN WELDING SOCIETY

When welding from a single side, the oxide on the opposite side is not removed by the action of the arc and inert-gas shield. Therefore, it must be physically stretched (Fig. 10) out of the way to permit complete root fusion for full-penetration welds. To accomplish this, the depth of the penetration reinforcement needs to be greater than the levels normally allowed by specifications governing other metals. Practical values for the root reinforcement (*R*) are (Ref 18):

THICKNESS, T		ROOT REINFORCEMENT, R	
mm	**in.**	**mm**	in.
$T \le 1.6$	$T \le \frac{1}{16}$	$1.5T$	**$1.5T$**
$1.6 < T \le 6.4$	$\frac{1}{16} < T \le \frac{1}{4}$	2.4	$\frac{3}{32}$
$6.4 < T \le 12.7$	$\frac{1}{4} < T \le \frac{1}{2}$	3.2	$\frac{1}{8}$
$12.7 < T \le 25.4$	$\frac{1}{2} < T \le 1$	4.0	$\frac{5}{32}$
$T > 25.4$	**$T > 1$**	**4.8**	$\frac{3}{16}$

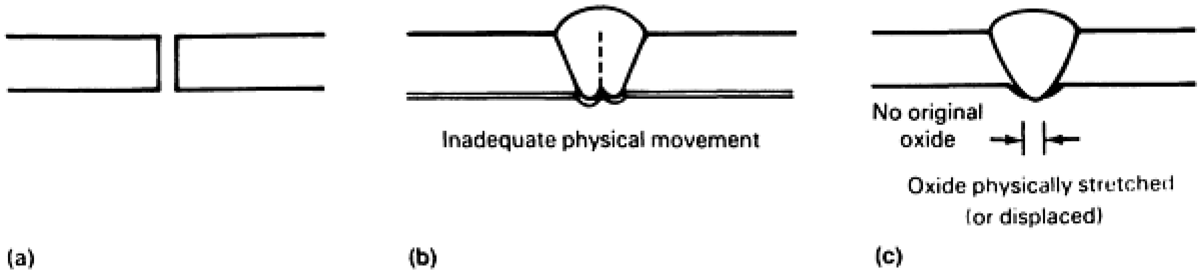
Inadequate physical movement
No original oxide
Oxide physically stretched (or displaced)
(a)
(b)
(c)

FIG. 10 EFFECT OF PHYSICAL STRETCHING ON ROOT FUSION. (A) ORIGINAL OXIDE SURFACES. (B) INCOMPLETE (POOR) ROOT FUSION. (C) COMPLETE (PROPER) ROOT FUSION

When an integral aluminum backing is used, the base metal edge spacing must be adequate to permit the arc cleaning action to remove the oxide from the backing surface and thus allow proper fusion. An opening of about 1*t* (one times the thickness) is required for thicknesses up to 9.5 mm ($\frac{3}{8}$ in.). For thicker material, the edge spacing should be sufficient to allow two root passes--one along each edge. Integral backings should either be flush with the base metal or possess a shallow groove and should be tack welded to the base metal. Any projections up into the joint to control edge spacing can result in poor fusion along each side of the projection because the oxide cannot be displaced.

Back gouging is normally done mechanically with pneumatic chisels, portable milling cutters, and routers. The gouge should possess a 3.2 mm ($\frac{1}{8}$ in.) minimum radius at its base. Sharp diamond-shaped chisels and sharp saw cuts do not permit proper fusion at their root when welded. Plasma arc cutting can be used to make square and beveled cuts that are free from surface contaminants. The plasma arc welding (PAW) process can also be used for back gouging. Argon-hydrogen gas mixtures are preferred for the best surface smoothness. Surface cracks can occur when arc cutting or gouging the heat-treatable aluminum alloys. These cracks can extend about 2.3 mm (0.090 in.) into the metal and must be mechanically removed before welding over that surface.

Surface Preparation. All surface lubricants on the base metal must be removed following metal working, sawing, and machining operations in order to avoid porosity. Vapor degreasers and any chlorinated solvent must be isolated from the welding area, because the arc radiation can alter these cleaning solutions chemically, resulting in toxic gases. Petroleum-base solvents can be used satisfactorily in the welding area.

Water stains, heat-treating oxides, chemical conversion coatings, and anodic finishes must be removed by chemical etching or mechanical means. Stainless steel brushes are preferred to avoid rust contamination of the joint. If a power brush is used, then a light pressure should be applied to remove the thick oxide. A high force can embed the oxide or contaminant into the surface and increase, rather than eliminate, weld discontinuities.

Surface moisture can be removed by a slight preheat of the joint prior to welding. A brief exposure at a maximum temperature of 120 °C (250 °F) should be controlled with the 5*xxx* series and heat-treatable aluminum alloys to avoid lowering corrosion resistance and reducing mechanical properties, respectively.

Optimal welding occurs when the metal is clean and dry, and possesses a minimum oxide thickness. In multiple-pass welds, it is common to wire brush the previously deposited surface in order to minimize the oxide or surface "smut" and thereby facilitate the subsequent weld pass.

References cited in this section

17. "RECOMMENDED PRACTICES FOR GAS SHIELDED ARC WELDING OF ALUMINUM AND ALUMINUM ALLOY PIPE," ANSI/AWS D10.7-86, AMERICAN WELDING SOCIETY, 1986
18. "STRUCTURAL WELDING CODE--ALUMINUM," ANSI/AWS D1.2-90, AMERICAN WELDING SOCIETY, 1990

Welding of Aluminum Alloys

Paul B. Dickerson, Aluminum Company of America (retired)

Welding Processes

The GTAW process has been used to weld thicknesses from 0.25 to 150 mm (0.010 to 6 in.) and can be used in all weld positions. Because it is relatively slow, it is highly maneuverable for welding tubing, piping, and variable shapes. It permits excellent penetration control and can produce welds of excellent soundness. Weld termination craters can be filled easily as the current is tapered down by a foot pedal or electronic control.

The ac-GTAW process provides an arc cleaning action to remove the surface oxide during the positive electrode half of the cycle and a penetrating arc when the electrode is operated at negative polarity. It is used primarily to weld thicknesses up to 6.3 mm ($\frac{1}{4}$ in.) and can be used for manual, mechanized, or automated applications. Typical manual welding procedures are listed in Tables 15, 16, and 17.

TABLE 15 TYPICAL AC-GTAW PROCEDURES FOR GROOVE WELDING OF ALUMINUM

See also AWS D10.7, "Recommended Practices for Gas Shielded Arc Welding of Aluminum and Aluminum Alloy Pipe"

BASE METAL THICKNESS		**WELD POSITION$^{(A)}$**	**EDGE PREPARATION$^{(B)}$**	**ROOT OPENING**		**WELD PASSES**	**FILLER DIAMETER**		**PURE TUNGSTEN ELECTRODE DIAMETER**		**GAS CUP INSIDE DIAMETER**		**ARGON GAS FLOW**		**BALANCED WAVE AC, A**	ARC TRAVEL SPEED	
MM	**IN.**			**MM**	**IN.**		**MM**	**IN.**	**MM**	**IN.**	**MM**	**IN.**	**L/MIN**	**FT^3/H**		**MM/S**	IN./MIN
1.6	$\frac{1}{16}$	F, V, H	A OR B	0-1.6	0-$\frac{1}{16}$	1	2.4	$\frac{3}{32}$	1.6-2.4	$\frac{1}{16}$-$\frac{3}{32}$	9.5	$\frac{3}{8}$	9	20	70-100	3-4	**8-10**
		O	A OR B	0-1.6	0-$\frac{1}{16}$	1	2.4	$\frac{3}{32}$	1.6	$\frac{1}{16}$	9.5	$\frac{3}{8}$	12	25	60-75	3-4	**8-10**
2.4	$\frac{3}{32}$	F	A OR B	0-2.4	0-$\frac{3}{32}$	1	3.2	$\frac{1}{8}$	2.4-3.2	$\frac{3}{32}$-$\frac{1}{8}$	9.5	$\frac{3}{8}$	9	20	95-115	3-4	**8-10**
		V, H	A OR B	0-2.4	0-$\frac{3}{32}$	1	2.4-3.2	$\frac{3}{32}$-$\frac{1}{8}$	2.4	$\frac{3}{32}$	9.5	$\frac{3}{8}$	9	20	85-110	3-4	**8-10**
		O	A OR B	0-2.4	0-$\frac{3}{32}$	1	2.4-3.2	$\frac{3}{32}$-$\frac{1}{8}$	2.4-3.2	$\frac{3}{32}$-$\frac{1}{8}$ $\frac{1}{8}$	9.5	$\frac{3}{8}$	12	25	90-110	3-4	**8-10**
3.2	$\frac{1}{8}$	F	A OR B	0-2.4	0-$\frac{3}{32}$	1-2	3.2-4.0	$\frac{1}{8}$-$\frac{5}{32}$	3.2	$\frac{1}{8}$	11.1	$\frac{7}{16}$	9	20	125-150	4-5	**10-12**
		V, H	A OR B	0-2.4	0-$\frac{3}{32}$	1-2	3.2	$\frac{1}{8}$	3.2	$\frac{1}{8}$	11.1	$\frac{7}{16}$ $\frac{7}{16}$	9	20	110-140	4	**10**
		O	A OR B	0-2.4	0-$\frac{3}{32}$	1-2	3.2-4.0	$\frac{1}{8}$-$\frac{5}{32}$	3.2	$\frac{1}{8}$	11.1	$\frac{7}{16}$	12	25	115-140	4-5	**10-12**

4.8	$\frac{3}{16}$	F	D-60°	0-3.2	0-$\frac{1}{8}$	2	4.0-4.8	$\frac{5}{32}$-$\frac{3}{16}$	4.0-4.8	$\frac{5}{32}$-$\frac{3}{16}$	11.1-12.7	$\frac{7}{16}$-$\frac{1}{2}$	12	25	170-190	4-5	**10-12**
		V	D-60°	0-2.4	0-$\frac{3}{32}$	2	4.0	$\frac{5}{32}$	4.0	$\frac{5}{32}$	11.1	$\frac{7}{16}$	12	25	160-175	4-5	**10-12**
		H	D-90°	0-2.4	0-$\frac{3}{32}$	2	4.0	$\frac{5}{32}$	4.0	$\frac{5}{32}$	11.1	$\frac{7}{16}$	12	25	155-170	4-5	**10-12**
		O	D-110°	0-2.4	0-$\frac{3}{32}$	2	4.0	$\frac{5}{32}$	4.0	$\frac{5}{32}$	11.1	$\frac{7}{16}$	14	30	165-180	4-5	**10-12**
6.3	$\frac{1}{4}$	F	D-60°	0-3.2	0-$\frac{1}{8}$	2	4.8	$\frac{3}{16}$	4.8-6.3	$\frac{3}{16}$-$\frac{1}{4}$	12.7	$\frac{1}{2}$	14	30	220-275	3-4	**8-10**
		V	D-60°	0-2.4	0-$\frac{3}{32}$	2	4.8	$\frac{3}{16}$	4.8	$\frac{3}{16}$	12.7	$\frac{1}{2}$	14	30	200-240	3-4	**8-10**
		H	D-90°	0-2.4	0-$\frac{3}{32}$	2-3	4.0-4.8	$\frac{5}{32}$-$\frac{3}{16}$	4.0-4.8	$\frac{5}{32}$-$\frac{3}{16}$	12.7	$\frac{1}{2}$	14	30	190-225	3-4	**8-10**
		O	D-110°	0-2.4	0-$\frac{3}{32}$	2	4.8	$\frac{3}{16}$	4.8	$\frac{3}{16}$	12.7	$\frac{1}{2}$	16	35	210-250	3-4	**8-10**
9.5(C)	$\frac{3}{8}$ (C)	F	D-60°	0-3.2	0-$\frac{1}{8}$	2	4.8-6.3	$\frac{3}{16}$-$\frac{1}{4}$	6.3	$\frac{1}{4}$	15.8	$\frac{5}{8}$	16	35	315-375	3-4	**8-10**
		F	E	0-2.4	0-$\frac{3}{32}$	2	4.8-6.3	$\frac{3}{16}$-$\frac{1}{4}$	6.3	$\frac{1}{4}$	15.8	$\frac{5}{8}$	16	35	340-380	3-4	**8-10**
		V	E-60°	0-2.4	0-$\frac{3}{32}$	3	4.8	$\frac{3}{16}$	4.8-6.3	$\frac{3}{16}$-$\frac{1}{4}$	15.8	$\frac{5}{8}$	16	35	260-300	3-4	**8-10**

		V, H, O	E	0-2.4	0-$\frac{3}{32}$	2	4.8	$\frac{3}{16}$	4.8-6.3	$\frac{3}{16}-\frac{1}{4}$	15.8	$\frac{5}{8}$	16	35	240-300	3-4	**8-10**
		H	D-90°	0-2.4	0-$\frac{3}{32}$	3	4.8	$\frac{3}{16}$	4.8-6.3	$\frac{3}{16}-\frac{1}{4}$	15.8	$\frac{5}{8}$	16	35	240-300	3-4	**8-10**
		O	**D-110°**	**0-2.4**	**0-**$\frac{3}{32}$	**3**	**4.8**	$\frac{3}{16}$	**4.8-6.3**	$\frac{3}{16}-\frac{1}{4}$	**15.8**	$\frac{5}{8}$	**19**	**40**	**260-300**	**3-4**	**8-10**

(A) F, FLAT; V, VERTICAL; H, HORIZONTAL; O, OVERHEAD.

(B) THESE ARE JOINTS SHOWN IN FIG. 8.

(C) PREHEATING IS OPTIONAL UP TO 177 °C (350 °F), HOWEVER, THIS MAY REDUCE THE WELD STRENGTH OF HEAT-TREATABLE ALLOYS.

TABLE 16 TYPICAL AC-GTAW PROCEDURES FOR FILLET AND LAP WELDING OF ALUMINUM COMPONENTS

BASE METAL THICKNESS		**WELD POSITION(A)**	**WELD PASSES**	**FILLER DIAMETER**		**PURE TUNGSTEN ELECTRODE DIAMETER**		**GAS CUP INSIDE DIAMETER**		**ARGON GAS FLOW**		**BALANCED WAVE AC, A**	ARC TRAVEL SPEED	
MM	**IN.**			**MM**	**IN.**	**MM**	**IN.**	**MM**	**IN.**	**L/MIN**	**FT^3/H**		**MM/S**	IN./MIN
1.6	$\frac{1}{16}$	F, H, V	1	2.4	$\frac{3}{32}$	1.6-2.4	$\frac{1}{16}-\frac{3}{32}$	9.5	$\frac{3}{8}$	8	16	70-100	3-4	**8-10**
		O	1	2.4	$\frac{3}{32}$	1.6-2.4	$\frac{1}{16}-\frac{3}{32}$	9.5	$\frac{3}{8}$	9	20	65-90	3-4	**8-10**
2.4	$\frac{3}{32}$	F	1	2.4-3.2	$\frac{3}{32}-\frac{1}{8}$	3.2-4.0	$\frac{1}{8}-\frac{5}{32}$	9.5	$\frac{3}{8}$	8	18	110-145	3-4	**8-10**
		H, V	1	2.4	$\frac{3}{32}$	2.4-3.2	$\frac{3}{32}-\frac{1}{8}$	9.5	$\frac{3}{8}$	8	18	90-125	3-4	**8-10**
		O	1	2.4	$\frac{3}{32}$	2.4-3.2	$\frac{3}{32}-\frac{1}{8}$	9.5	$\frac{3}{8}$	9	20	110-135	3-4	**8-10**
3.2	$\frac{1}{8}$	F	1	3.2	$\frac{1}{8}$	3.2-4.0	$\frac{1}{8}-\frac{5}{32}$	11.1	$\frac{7}{16}$	9	20	135-175	4-5	**10-12**
		H, V	1	3.2	$\frac{1}{8}$	2.4-3.2	$\frac{3}{32}-\frac{1}{8}$	9.5	$\frac{3}{8}$	9	20	115-145	3-4	**8-10**
		O	1	3.2	$\frac{1}{8}$	2.4-3.2	$\frac{3}{32}-\frac{1}{8}$	11.1	$\frac{7}{16}$	12	25	125-155	3-4	**8-10**
4.8	$\frac{3}{16}$	F	1	4.0	$\frac{5}{32}$	4.0-4.8	$\frac{5}{32}-\frac{3}{16}$	12.7	$\frac{1}{2}$	12	25	190-245	3-4	**8-10**
		H, V	1	4.0	$\frac{5}{32}$	4.0-4.8	$\frac{5}{32}-\frac{3}{16}$	12.7	$\frac{1}{2}$	12	25	175-210	3-4	**8-10**
		O	1	4.0	$\frac{5}{32}$	4.0-4.8	$\frac{5}{32}-\frac{3}{16}$	12.7	$\frac{1}{2}$	14	30	180-225	3-4	**8-10**
6.3	$\frac{1}{4}$	F	1	4.8	$\frac{3}{16}$	4.8-6.3	$\frac{3}{16}-\frac{1}{4}$	12.7	$\frac{1}{2}$	14	30	240-295	3-4	**8-10**
		H,V	1	4.8	$\frac{3}{16}$	4.8	$\frac{3}{16}$	12.7	$\frac{1}{2}$	14	30	220-265	3-4	**8-10**
		O	1	4.8	$\frac{3}{16}$	4.8	$\frac{3}{16}$	12.7	$\frac{1}{2}$	16	35	230-275	3-4	**8-10**
9.5(B)	$\frac{3}{8}$ (B)	F	2	4.8	$\frac{3}{16}$	6.3	$\frac{1}{4}$	16	$\frac{5}{8}$	16	35	325-375	3-4	**8-10**

		V	2	4.8	3/16	4.8-6.3	3/16 - 1/4	16	5/8	16	35	280-315	3-4	**8-10**
		H	3	4.8	3/16	4.8-6.3	3/16 - 1/4	16	5/8	16	35	270-300	3-4	**8-10**
		O	**3**	**4.8**	3/16	**4.8-6.3**	3/16 - 1/4	**16**	5/8	**19**	**40**	**290-335**	**3-4**	**8-10**

(A) F, FLAT; V, VERTICAL; H, HORIZONTAL; O, OVERHEAD.

(B) PREHEATING IS OPTIONAL UP TO 177 °C (350 °F); HOWEVER, THIS MAY REDUCE THE WELD STRENGTH OF HEAT-TREATABLE ALLOYS.

TABLE 17 TYPICAL AC-GTAW PROCEDURES FOR CORNER AND EDGE WELDING OF ALUMINUM COMPONENTS

BASE METAL THICKNESS		**WELD POSITION(A)**	**NUMBER OF WELD PASSES**	**FILLER DIAMETER**		**PURE TUNGSTEN ELECTRODE DIAMETER**		**GAS CUP INSIDE DIAMETER**		**ARGON GAS FLOW**		**BALANCED WAVE AC, A**	ARC TRAVEL SPEED	
MM	**IN.**			**MM**	**IN.**	**MM**	**IN.**	**MM**	**IN.**	**L/MIN**	**FT^3/H**		**MM/S**	IN./MIN
1.6	1/16	I, K	1	2.4	3/32	1.6	1/16	9.5	3/8	9	20	60-85	4-7	**10-16**
2.4	3/32	I, K	1	3.2	1/8	2.4	3/32	9.5	3/8	9	20	90-120	4-7	**10-16**
3.2	1/8	I, K	1	3.2-4.0	1/8 - 5/32	3.2	1/8	9.5	3/8	9	20	115-150	4-7	**10-16**
4.8	3/16	J, K	1	4.0	5/32	4.0	5/32	11.1	7/16	12	25	160-210	4-7	**10-16**
6.3	1/4	**J, K**	**2**	**4.8**	3/16	**4.8**	3/16	**12.7**	1/2	**14**	**30**	**200-250**	**3-5**	**8-12**

(A) JOINT DESIGNS AS SHOWN IN FIG. 8.

(B) HIGHER CURRENTS AND WELDING SPEEDS CAN BE EMPLOYED IF A TEMPORARY BACKING IS USED FOR COMER JOINTS.

(C) USE LOW SIDE OF CURRENT RANGE FOR HORIZONTAL AND VERTICAL JOINTS.

Optimal control of penetration in groove welds is obtained with a balanced ac wave, which permits a wide tolerance in the welding current or heat input between insufficient and excessive penetration. The fairly wide melting pattern does not penetrate readily into narrow grooves or corners and tends to bridge the root of fillet welds. When a wave balance control is available on the power supply, an increase in the percentage of the negative polarity can be beneficial, as long as adequate arc cleaning action is maintained.

The ac-GTAW process is particularly useful for welding aluminum pipe. An integral backing is suitable for structural and electrical bus applications; however, for fluid flow of water, gas, oil, and chemicals, crevice corrosion can result at the backing interface within the pipe. The joint configuration shown in Fig. 9 has been used for all-position GTAW of aluminum pipe and is considered standard for this purpose (Ref 17).

Argon shielding gas is normally used. However, for increased penetration capabilities and use on thick base metal, a mixture of argon with 25 to 50% helium can be used. Welding operations must be protected from wind gusts, which can disturb the inert-gas shielding, when welding near open doors and windows or when welding outdoors.

The alternating current arc (whether sinusoidal or square wave) requires a superimposed high-frequency power to maintain the ionized gas circuit when the current changes polarity. This avoids rectification, which reduces the arc cleaning action and arc stability.

Pure or zirconia-tungsten electrodes, which form a ball on the end during welding, should be used with ac-GTAW. Fine spitting of thoria-tungsten electrodes will occur when used with alternating current and should be limited to direct current (dc) operations.

The dc-GTAW Process. Negative electrode polarity direct current can be used to weld aluminum by manual and mechanized means. Helium gas shielding or a mixture of helium with 5 to 10% argon is used with a thoria-tungsten electrode. Because no arc cleaning occurs, the oxide must be minimized on the surface prior to welding. This is normally accomplished by chemical etching, followed by a mechanical "scraping" of the joint surfaces immediately prior to welding. Internal oxide stringers and associated poor fusion can occur if this cleaning is not done thoroughly. The deep, narrow penetration allows welding of square groove joints in thicknesses up to 12.7 mm ($\frac{1}{2}$ in.) and permits welding of heat-treatable alloys with a lower heat input than does ac-GTAW. This can result in higher as-welded mechanical properties with the heat-treatable alloys. The process is most widely used by the aerospace industry for this purpose.

Manual dc-GTAW using negative electrode polarity has been applied primarily to the repair of thick castings or weldments in thick plate. Agitation of the weld pool with the filler rod is sometimes required to break up the surface oxide for proper weld fusion. An excellent fabrication technique utilizing this process has been tack welding of fillet joints in 4.8 mm ($\frac{3}{16}$ in.) and thicker non-heat-treatable alloys, such as 5083 and 5456. A very small fillet reinforcement possessing deep penetration can be made, and there is little or no interference when the continuous weld is produced.

Variable Polarity PAW. This inert-gas process provides good arc cleaning action and sufficient penetration capabilities, such that a "keyhole" is created. It is best used in the vertical weld position so that when the keyhole is generated, the weld metal backfills the joint. The process is only used in mechanized or automated environments, and the equipment is very expensive. The weld is less sensitive to oxide conditions than those produced by either ac- or dc-GTAW, and high-quality welds can be made. Inert-gas backing is desirable for some of the aluminum alloys, such as the aluminum-copper-lithium type. Square groove joints up to 16 mm ($\frac{5}{8}$ in.) can be welded in a single pass.

GMAW is the major high-speed production process for arc welding aluminum. The equipment is portable, and the process can be easily used in all weld positions. It uses positive electrode dc power, which gives it a continuous cleaning action and concentrates the arc to produce rapid melting. Widely varying thicknesses can be welded without the need to preheat. Spooled aluminum wire is automatically fed through the welding gun, melted at the arc, and then sprayed across the arc. Because the electrode and filler are the same, the process is a one-handed operation that is easily mechanized or adapted for robotic operations. Manual travel speeds normally range from 7.6 to 15.2 mm/s (18 to 36 in./min). However, mechanized welds have been made at speeds that exceed 42 mm/s (100 in./min). Initially, GMAW was limited to 3.2 mm

($\frac{1}{8}$ in.) and thicker material. However, the availability of pull-type wire feed apparatus and pulsed-dc power supplies has extended its use with smaller electrodes and lower average currents to weld thicknesses of less than 1 mm (0.040 in.).

GMAW is best used in lap, fillet, or groove welds with integral or temporary backing. Typical welding procedures are listed in Tables 18 and 19. It is not recommended for corner or edge joints, unless the metal is thick enough to provide a substantial ledge or trough upon which to deposit the filler. Heat input needs to be constant for uniform penetration. A constant voltage power supply can be used for straight stringer passes. However, when the arc is oscillated transversely or in line with the joint, wide current excursions produce variable penetration and fusion. Weld starts with short contact tubes and fast electrode feed rates, as used in equipment designed for welding ferrous metals, and a constant voltage power supply often results in poorly fused starts, arcing in the contact tube, and excessive "burn backs" when used with aluminum.

TABLE 18 TYPICAL SEMIAUTOMATIC GMAW PROCEDURES FOR GROOVE WELDING OF ALUMINUM COMPONENTS

BASE METAL THICKNESS		**WELD POSITION$^{(A)}$**	**EDGE PREPARATION$^{(B)}$**	**JOINT SPACING**		**TYPE AND NUMBER OF WELD PASSES$^{(C)}$**	**ELECTRODE DIAMETER**		**CURRENT (DC, POSITIVE ELECTRODE)$^{(D)}$, A**	**ARC VOLTAGE$^{(D)}$, V**	**ARGON GAS FLOW**		ARC TRAVEL SPEED	
MM	**IN.**			**MM**	**IN.**		**MM**	**IN.**			**L/MIN**	**FT^3/H**	**MM/S**	IN./MIN
1.6	1/16	F	A	. . .	. . .	1	0.8	0.030	70-110	15-20	12	25	10-19	**25-45**
		F	G	2.4	3/32	1	0.8	0.030	70-110	15-20	12	25	10-19	**25-45**
2.4	3/32	F	A	. . .	. . .	1	0.8-1.2	0.030-3/64	90-150	18-22	14	30	10-19	**25-45**
		F, V, H, O	G	3.2	1/8	1	0.8	0.030	110-130	18-23	14	30	10-13	**23-30**
3.2	1/8	F, V, H	A	0-2.4	0-3/32	1	0.8-1.2	0.030-3/64	120-150	20-24	14	30	10-13	**24-30**
		F, V, H, O	G	4.8	3/16	1	0.8-1.2	0.030-3/64	110-135	19-23	14	30	8-12	**18-28**
4.8	3/16	F, V, H	B	0-1.6	0-1/16	1F, 1R	0.9-1.2	0.035-3/64	130-175	22-26	16	35	10-13	**24-30**
		F, V, H	F	0-1.6	0-1/16	1	1.2	3/64	140-180	23-27	16	35	10-13	**24-30**
		O	F	0-1.6	0-1/16	2F	1.2	3/64	140-175	23-27	28	60	10-13	**24-30**
		F, V	H	3.2-4.8	1/8-3/16	2	1.2-1.6	3/64-1/16	140-185	23-27	16	35	10-13	**24-30**
		H, O	H	4.8	3/16	3	0.9-1.2	0.035-3/64	130-175	23-27	28	60	10-15	**25-35**

6.3	$\frac{1}{4}$	F	B	0-2.4	0-$\frac{3}{32}$	1F, 1R	1.2-1.6	$\frac{3}{64}$-$\frac{1}{16}$	175-200	24-28	19	40	10-13	**24-30**
		F	F	0-2.4	0-$\frac{3}{32}$	2	1.2-1.6	$\frac{3}{64}$-$\frac{1}{16}$	185-225	24-29	19	40	10-13	**24-30**
		V, H	F	0-2.4	0-$\frac{3}{32}$	3F, 1R	1.2	$\frac{3}{64}$	165-190	25-29	21	45	10-15	**25-35**
		O	F	0-2.4	0-$\frac{3}{32}$	3F, 1R	1.2-1.6	$\frac{3}{64}$-$\frac{1}{16}$	180-200	25-29	28	60	10-15	**25-35**
		F, V	H	3.2-6.3	$\frac{1}{8}$-$\frac{1}{4}$	2-3	1.2-1.6	$\frac{3}{64}$-$\frac{1}{16}$	175-225	25-29	19	40	10-13	**24-30**
		O, H	H	6.3	$\frac{1}{4}$	4-6	1.2-1.6	$\frac{3}{64}$-$\frac{1}{16}$	170-200	25-29	28	60	11-17	**25-40**
9.5	$\frac{3}{8}$	F	C-90°	0-2.4	0-$\frac{3}{32}$	1F, 1R	1.6	$\frac{1}{16}$	225-290	26-29	23	50	8-13	**20-30**
		F	F	0-2.4	0-$\frac{3}{32}$	2F, 1R	1.6	$\frac{1}{16}$	210-275	26-29	23	50	10-15	**24-35**
		V, H	F	0-2.4	0-$\frac{3}{32}$	3F, 1R	1.6	$\frac{1}{16}$	190-220	26-29	26	55	10-13	**24-30**
		O	F	0-2.4	0-$\frac{3}{32}$	5F, 1R	1.6	$\frac{1}{16}$	200-250	26-29	38	80	11-17	**25-40**
		F, V	H	6.3-9.5	$\frac{1}{4}$-$\frac{3}{8}$	4	1.6	$\frac{1}{16}$	210-290	26-29	23	50	10-13	**24-30**
		O, H	H	9.5	$\frac{3}{8}$	8-10	1.6	$\frac{1}{16}$	190-260	26-29	38	80	11-17	**25-40**
19.0	$\frac{3}{4}$	F	C-60°	0-2.4	0-$\frac{3}{32}$	3F, 1R	2.4	$\frac{3}{32}$	340-400	26-31	28	60	6-8	**14-20**

		F	F	0-3.2	0-$\frac{1}{8}$	4F, 1R	2.4	$\frac{3}{32}$	325-375	26-31	28	60	7-8	**16-20**
		V, H, O	F	0-1.6	0-$\frac{1}{16}$	8F, 1R	1.6	$\frac{1}{16}$	240-300	26-30	38	80	10-13	**24-30**
		F	E	0-1.6	0-$\frac{1}{16}$	3F, 3R	1.6	$\frac{1}{16}$	270-330	26-30	28	60	7-10	**16-24**
		V, H, O	**E**	**0-1.6**	**0-**$\frac{1}{16}$	**6F, 6R**	**1.6**	$\frac{1}{16}$	**230-280**	**26-30**	**38**	**80**	**7-10**	**16-24**

(A) F, FLAT; V, VERTICAL; H, HORIZONTAL; O, OVERHEAD.

(B) JOINTS AS SHOWN IN FIG. 8.

(C) F, FACE PASSES; R, ROOT PASSES. BACK GOUGING SHOULD BE DONE PRIOR TO THE APPLICATION OF FIRST ROOT PASS.

(D) FOR 5*XXX* SERIES ELECTRODES, USE A WELDING CURRENT IN THE HIGH SIDE OF THE RANGE AND AN ARC VOLTAGE IN THE LOWER PORTION OF THE RANGE. FOR 1*XXX*, 2*XXX*, AND 4*XXX* SERIES ELECTRODES, USE THE LOWER CURRENTS AND THE HIGHER ARC VOLTAGES.

TABLE 19 TYPICAL SEMIAUTOMATIC GMAW PROCEDURES FOR FILLET AND LAP WELDING OF ALUMINUM COMPONENTS

BASE METAL THICKNESS		**WELD POSITION(A)**	**WELD PASSES(B)**	**ELECTRODE DIAMETER**		**CURRENT (DC, POSITIVE ELECTRODE)(C), A**	**ARC VOLTAGE(C), V**	**ARGON GAS FLOW**		ARE TRAVEL SPEED	
MM	**IN.**			**MM**	**IN.**			**L/MIN**	**FT^3/H**	**MM/S**	IN./MIN
2.4	$\frac{3}{32}$	F, V, H, O	1	0.8	0.030	100-130	18-22	14	30	10-13	**24-30**
3.2	$\frac{1}{8}$	F	1	0.8-1.2	0.030-$\frac{3}{64}$	125-150	20-24	14	30	10-13	**24-30**
		V, H	1	0.8	0.030	110-130	19-23	14	30	10-13	**24-30**
		O	1	0.8-1.2	0.030-$\frac{3}{64}$	115-140	20-24	19	40	10-13	**24-30**
4.8	$\frac{3}{16}$	F	1	1.2	$\frac{3}{64}$	180-210	22-26	14	30	10-13	**24-30**
		V, H	1	0.8-1.2	0.030-$\frac{3}{64}$	130-175	21-25	16	35	10-13	**24-30**
		O	1	0.8-1.2	0.030-$\frac{3}{64}$	130-190	22-26	21	45	10-13	**24-30**
6.3	$\frac{1}{4}$	F	1	1.2-1.6	$\frac{3}{64}$-$\frac{1}{16}$	170-240	24-28	19	40	10-13	**24-30**
		V, H	1	1.2	$\frac{3}{64}$	170-210	23-27	21	45	10-13	**24-30**
		O	1	1.2-1.6	$\frac{3}{64}$-$\frac{1}{16}$	190-220	24-28	28	60	10-13	**24-30**
9.5	$\frac{3}{8}$	F	1	1.6	$\frac{1}{16}$	240-300	26-29	23	50	8-10	**18-25**
		V, H	3	1.6	$\frac{1}{16}$	190-240	24-27	28	60	10-13	**24-30**
		O	3	1.6	$\frac{1}{16}$	200-240	25-28	40	85	10-13	**24-30**
19.0(D)	$\frac{3}{4}$ (D)	F	4	2.4	$\frac{3}{32}$	360-380	26-30	28	60	8-10	**18-25**
		V, H	4-6	1.6	$\frac{1}{16}$	260-310	25-29	33	70	10-13	**24-30**
		O	**10**	**1.6**	$\frac{1}{16}$	**275-310**	**25-29**	**40**	**85**	**10-13**	**24-30**

(A) F, FLAT; V, VERTICAL; H, HORIZONTAL; O, OVERHEAD.

(B) NUMBER OF WELD PASSES GIVEN IS FOR ONE SIDE ONLY.

(C) FOR 5*XXX* SERIES ELECTRODES, USE A WELDING CURRENT IN THE HIGH SIDE OF THE RANGE GIVEN AND AN ARC VOLTAGE IN THE LOWER PORTION OF THE RANGE; FOR 1*XXX*, 2*XXX*, AND 4*XXX* SERIES ELECTRODES, USE THE LOWER CURRENTS AND THE HIGHER VOLTAGES IN THEIR RANGES.

(D) METAL THICKNESSES OF [GES]19.0 MM ([GES] $\frac{3}{4}$ IN.) FOR FILLET WELDS CAN EMPLOY A DOUBLE-VEE BEVEL OF [GES]50° INCLUDED VEE WITH 2.4-3.2 MM ($\frac{3}{32}$-$\frac{1}{8}$ IN.) LAND THICKNESS IN THE ABUTTING MEMBER.

The optimal equipment for the GMAW of aluminum to meet radiographic standards and to overcome excessive downtime in maintaining the equipment consists of:

- A CONSTANT-ENERGY (CONSTANT-CURRENT, OR "DROOPER") TYPE POWER SUPPLY
- A WIRE FEEDER POSSESSING A "TOUCH-START" OR SLOW RUN-IN FEATURE AND RADIUSED TOP AND BOTTOM DIVE ROLLS
- A WELDING GUN POSSESSING A CONTACT TUBE OF 100 MM (4 IN.) OR MORE IN LENGTH
- NONMETALLIC GUIDES AND LINERS FOR THE ELECTRODE CIRCUITRY

For 1.2 mm ($\frac{3}{64}$ in.) and smaller diameter electrodes, as well as for robotic installations, the wire feeding apparatus should be a pull or push-pull type.

Argon gas shielding is most often used. However, when welding with 5*xxx* series electrode alloys and in facilities that possess highly loaded power circuits, a mixture of argon with 50 to 75% helium will provide improved weld quality.

Other arc welding processes include shielded metal arc welding (SMAW), as well as electroslag and electrogas welding (ESW, EGW). SMAW with flux-coated rods has been replaced to a very substantial degree by the GMAW process. SMAW can be effective on 9.5 mm ($\frac{3}{8}$ in.) and thicker aluminum where high heat inputs are used. However, it produces porous and erratic welds on thinner metal. Excessive spatter limits its use overhead, and the flux requires removal after welding. Because a flux is used, the alloys that can be welded are limited (refer to Tables 1, 2, 3, 4).

The ESW and EGW processes have been used experimentally, but the processes are not in commercial use with aluminum. GMAW can be used for the few applications appropriate for the ESW and EGW processes.

The oxyfuel gas welding (OFW) process uses a flux and either an oxyacetylene or oxyhydrogen gas flame. When the oxyacetylene flame is used, a slightly reduced flame is required, which causes a carbonaceous deposit that obscures the weld and slows the travel speed. The best visibility, control, and weld speed are obtained when an oxyhydrogen flame is used with aluminum alloys. The flux (composed of chlorides and fluorides) must be removed after welding to avoid corrosion in service. For this reason, GTAW has generally replaced OFW of aluminum alloys. The use of a flux limits the alloys for which it is suitable (Tables 1, 2, 3, 4) and produces the greatest heat input. This results in the lowest mechanical properties and highest sensitivity to weld cracking with heat-treatable aluminum alloys. The advantage of the OFW process is the low equipment investment.

Electron-beam welding (EBW) in a vacuum chamber produces a very deep, narrow penetration at high welding speeds. The low overall heat input produces the highest as-welded strengths in the heat-treatable alloys. The high thermal gradient from the weld into the base metal creates very limited metallurgical modifications and is least likely to cause intergranular cracking in butt joints when no filler is added. Fixturing to provide transverse compressive loading on the joint can be very helpful in avoiding intergranular cracking in the fusion zones or HAZs. Alloys are more prone to have porosity when welded in a vacuum.

For lap joints and butt joints with an integral backing, a shim of filler alloy is commonly placed at the interface of the heat-treatable aluminum alloys to prevent transverse cracking at the interface. For the 6*xxx* series and for 2014, 2024, and 7075 alloys, a filler alloy shim is often desirable between the abutting edges of square groove welds. This adds two additional oxide layers to the joint and can result in weld inclusions.

Laser-beam welding (LBW) is now considered to be a viable fusion joining process for aluminum with the advent of commercially available, stable, high-power laser systems. Because of aluminum's high reflectivity, effective coupling of the laser beam and aluminum requires a relatively high power density. High power densities are obtained by minimizing the focused spot size of the beam through optical considerations and higher power levels.

When sufficient power densities are obtained (that is, on the order of 10^6 W/cm^2, or 6×10^6 W/in.^{2}), successful laser-beam welds on aluminum can be produced, with minimal distortion, and at high processing speeds. Inert gas shielding is commonly used. Filler metal is required when welding heat-treatable aluminum alloys.

Reference cited in this section

17. "RECOMMENDED PRACTICES FOR GAS SHIELDED ARC WELDING OF ALUMINUM AND ALUMINUM ALLOY PIPE," ANSI/AWS D10.7-86, AMERICAN WELDING SOCIETY, 1986

Paul B. Dickerson, Aluminum Company of America (retired)

Joining Aluminum to Other Metals

Very brittle intermetallic compounds are formed when metals such as steel, copper, magnesium, or titanium are fusion welded to aluminum. Bimetallic transition materials in sheet, plate, and tubular forms are commercially available in combinations of aluminum to such other metals as steel, stainless steel, and copper. These are made by rolling, explosion welding, fiction welding, flash welding, or hot pressure welding and provide the easiest method for fusion welding aluminum to other metals. Conventional GTAW and GMAW methods, as well as resistance spot welding, are used to join the aluminum side of the transition piece to the intended component. The dissimilar metal is joined to the opposite side of the bimetallic transition.

A wide range of dissimilar metals can be joined directly to aluminum by such means as ultrasonic welding, pressure welding, friction welding, explosion welding, and soldering. Hot dip or electroplated coatings on the dissimilar metal facilitate diffusion bonding and the brazing of aluminum to these metals.

Paul B. Dickerson, Aluminum Company of America (retired)

References

1. J.D. DOWD, WELD CRACKING OF ALUMINUM ALLOYS, *WELD. J.,* RESEARCH SUPPLEMENT, VOL 31, 1952, P 448S-456S
2. J.H. DUDAS AND F.R. COLLINS, PREVENTING WELD CRACKS IN HIGH-STRENGTH ALUMINUM ALLOYS, *WELD. J.,* RESEARCH SUPPLEMENT, VOL 45 (NO. 6), JUNE 1966, P 241S-249S
3. C.E. CROSS, PH.D. THESIS, COLORADO SCHOOL OF MINES, 1986, P 144
4. A.R.E. SINGER AND P.H. JENNINGS, HOT SHORTNESS OF THE ALUMINUM-SILICON ALLOYS OF COMMERCIAL QUALITY, *J. INST. MET.,* VOL 73, 1947, P 197-212
5. W.I. PUMPHREY AND J.V. LYONS, CRACKING DURING THE CASTING AND WELDING OF THE MORE COMMON BINARY ALUMINUM ALLOYS OF COMMERCIAL QUALITY, *J. INST. MET.,* VOL 74, 1948, P 439-455
6. P.H. JENNINGS, A.R.E. SINGER, AND W.I. PUMPHREY, HOT-SHORTNESS OF SOME HIGH PURITY ALLOYS IN THE SYSTEMS ALUMINUM-COPPER AND ALUMINUM-MAGNESIUM-SILICON, *J. INST. MET.,* VOL 74, 1948, P 227-248
7. N.F. GITTOS AND M.H. SCOTT, HEAT AFFECTED ZONE CRACKING OF AL-MG-SI ALLOYS, *WELD. J.,* RESEARCH SUPPLEMENT, JUNE 1981, P 95S-103S
8. W.L. BURCH, THE EFFECT OF WELDING SPEED ON THE STRENGTH OF 6061-T4 ALUMINUM JOINTS, *WELD. J.,* RESEARCH SUPPLEMENT, VOL 37 (NO. 8), AUG 1958, P 361S-367S
9. G.E. METZGER, SOME MECHANICAL PROPERTIES OF WELDS IN 6061 ALUMINUM ALLOY

SHEET, *WELD. J.,* RESEARCH SUPPLEMENT, VOL 46 (NO. 10), OCT 1967, P 457S-469S

10. F.R. COLLINS, F.G. NELSON, AND I.B. ROBINSON, ALUMINUM 2219--NEW ALLOY FOR HIGH-STRENGTH WELDED STRUCTURES, *MET. PROG.,* NOV 1961
11. E.H. DIX, W.H. ANDERSON, AND M.B. SHUMAKER, INFLUENCE OF SERVICE TEMPERATURE ON THE RESISTANCE OF WROUGHT ALUMINUM-MAGNESIUM ALLOYS TO CORROSION, *CORROSION,* VOL 15 (NO. 2), FEB 1959, P 55T-62T
12. M.B. SHUMAKER, R.A. KELSEY, D.O. SPROWLS, AND J.G. WILLIAMSON, "EVALUATION OF VARIOUS TECHNIQUES FOR STRESS-CORROSION TESTING WELDED ALUMINUM ALLOYS," PRESENTED AT ASTM STRESS-CORROSION TESTING SYMPOSIUM, JUNE 1966
13. F.G. NELSON AND R.L. ROLF, SHEAR STRENGTHS OF ALUMINUM ALLOY FILLET WELDS, *WELD. J.,* RESEARCH SUPPLEMENT, VOL 45 (NO. 2), FEB 1966, P 82S-84S
14. S.L. WÖHLER AND O. SCHLIEPHAKE, ÜBEREINEM GERS., DURCH POTENTIALMESSUNGER DAS AUFGREGEN DER VERB. MG_2SI ZU BESTÄTIGEN, *Z. ANORG. CHEM.,* 1962, P 151
15. G.W. AKIMOW AND A.S. OLESCHKO, GMELINS HANDBUCH DER ANORGANISCHEN CHEMIE, (NO. 8), AUFLAGE, 1942, P 393
16. "SPECIFICATION FOR BARE ALUMINUM AND ALUMINUM ALLOY WELDING ELECTRODE AND ROD," ANSI/AWS A5.10-92, AMERICAN WELDING SOCIETY, 1992
17. "RECOMMENDED PRACTICES FOR GAS SHIELDED ARC WELDING OF ALUMINUM AND ALUMINUM ALLOY PIPE," ANSI/AWS D10.7-86, AMERICAN WELDING SOCIETY, 1986
18. "STRUCTURAL WELDING CODE--ALUMINUM," ANSI/AWS D1.2-90, AMERICAN WELDING SOCIETY, 1990

Welding of Nickel Alloys

Revised by E.B. Hinshaw, Inco Alloys International, Inc.

Introduction

NICKEL ALLOYS can be joined reliably by all types of welding processes or methods, with the exception of forge welding and oxyacetylene welding. Because the majority of consumables are used with arc welding processes, it is those processes that are discussed in this article. The wrought nickel alloys listed in Table 1 can be welded under conditions similar to those used to weld austenitic stainless steels. Cast nickel alloys, particularly those with a high silicon content, present difficulties in welding. The arc welding of heat-resistant and corrosion-resistant nickel alloys is described in the Sections "Selection of Nonferrous Corrosion-Resistant Materials" and "Selection of Nonferrous High-Temperature Materials" in this Volume.

TABLE 1 NOMINAL COMPOSITIONS OF SELECTED WELDABLE WROUGHT NICKEL AND NICKEL ALLOYS

UNS NO.	**ALLOY**	COMPOSITION, WT%											
		Ni	**C**	**Mn**	**Fe**	**S**	**Si**	**Cu**	**Cr**	**Al**	**Ti**	**Nb**	OTHER
N02200	200	99.5	0.08	0.18	0.2	0.005	0.18	0.13	. . .	. . .	. . .	. . .	**. . .**
N02201	201	99.5	0.01	0.18	0.2	0.005	0.18	0.13	. . .	. . .	. . .	. . .	**. . .**
N02205	205	99.5	0.08	0.18	0.10	0.004	0.08	0.08	. . .	. . .	0.03	. . .	**0.05 MG**
N0233	233	99.64	0.06	0.19	0.05	0.004	0.03	0.01	. . .	. . .	0.002	. . .	**0.01 MG**
N02270	270	99.98	0.01	<0.001	0.003	<0.001	<0.001	<0.001	<0.001	. . .	0.001	. . .	**<0.001 MG** **<0.001 CO**
N03300	300	98.67	0.30	0.11	0.05	0.002	0.06	0.01	. . .	. . .	0.46	. . .	**. . .**
N03301	301	94.18	0.17	0.25	0.06	0.001	0.42	0.01	. . .	4.48	0.43	. . .	**. . .**
N04400	400	66.5	0.15	1.0	1.25	0.012	0.25	31.5	. . .	. . .	. . .	. . .	**. . .**
N04401	401	42.5	0.05	1.6	0.38	0.008	0.13	BAL	. . .	. . .	. . .	. . .	**. . .**
N04404	404	54.5	0.08	0.05	0.25	0.012	0.05	44.0	. . .	0.03	. . .	. . .	**. . .**

N04405	R-405	66.5	0.15	1.0	1.25	0.043	0.25	31.5	. . .	. . .	. . .	. . .	**. . .**
N05500	K-500	66.5	0.13	0.75	1.00	0.005	0.25	29.5	. . .	2.73	0.60	. . .	**. . .**
N06600	600	76.0	0.08	0.5	8.0	0.008	0.25	0.25	15.5	. . .	. . .	. . .	**. . .**
N06601	601	60.5	0.05	0.5	14.1	0.007	0.25	0.50	23.0	1.35	. . .	. . .	
N06617	617	54.5	0.07	. . .	. . .	. . .	. . .	. . .	22.0	1.0	. . .	. . .	**12.5 CO, 9.0 MO**
N06022	622	59.4	0.01	. . .	2.3	<0.001	. . .	. . .	20.5	. . .	. . .	. . .	**3.2 W, 14.2 MO**
N06625	625	61.0	0.05	0.25	2.5	0.008	0.25	. . .	21.5	0.2	0.2	3.65	**9.0 MO**
N06686	686	57	<0.01	0.25	1.0	<0.001	. . .	. . .	20.5	0.2	0.03	. . .	**16.5 MO, 4.0 W**
N10276	C-276	57	0.01	. . .	6.0	<0.001	. . .	. . .	15.5	. . .	. . .	. . .	**16.0 MO, 3.9 W**
N06690	690	60.0	0.03	. . .	9.0	<0.001	0.20	. . .	29.0	0.25	0.25	. . .	**. . .**
N07702	702	79.5	0.05	0.50	1.0	0.005	0.35	0.25	15.5	3.25	0.63	. . .	**. . .**
N09706	706	41.5	0.03	0.18	40.0	0.008	0.18	0.15	16.0	0.20	1.75	2.9	**. . .**
N07718	718	52.5	0.04	0.18	18.5	0.008	0.18	0.15	19.0	0.50	0.90	5.13	**3.05 MO**
N07725	725	57.0	<0.01	. . .	8.0	. . .	. . .	. . .	21.5	0.2	1.4	3.5	**8.0 MO**
N07750	X-750	73.0	0.04	0.50	7.0	0.005	0.25	0.25	15.5	0.70	2.50	0.95	**. . .**
N07751	751	72.5	0.05	0.5	7.0	0.005	0.25	0.25	15.5	1.20	2.30	0.95	**. . .**
N08800	800	32.5	0.05	0.5	39.5	0.008	0.50	0.50	21.0	0.40	0.40	. . .	**. . .**
N08810	800H	32.5	0.075	0.5	39.5	0.008	0.50	0.50	21	0.40	0.37	. . .	**. . .**
N08811	800HT	32.5	0.08	0.5	39.5	0.008	0.50	0.50	21	0.40	0.37	. . .	**1.0 MAX (AL + TI)**
N08801	801	32.0	0.05	0.75	44.5	0.008	0.50	0.50	20.5	. . .	1.13	. . .	**. . .**
N08802	802	32.5	0.35	0.75	46.0	0.008	0.38	0.38	21.0	0.58	0.75	. . .	**. . .**
N08825	825	42.0	0.03	0.50	30.0	0.015	0.25	0.25	21.5	0.10	0.90	. . .	**3.0 MO**
N09902	902	42.25	0.03	0.40	48.5	0.02	0.50	0.50	5.33	0.55	2.58	. . .	**. . .**
N19903	903	38.2	0.03	0.13	41.13	0.002	0.19	. . .	. . .	0.86	1.38	3.00	**. . .**
N19907	907	37.42	0.02	0.05	41.78	0.004	0.09	. . .	0.13	0.05	1.55	4.72	**14.33 CO, 0.005 B**
. . .	908	50.0	0.15	0.05	40.5	<0.002	0.20	0.01	4.0	1.0	1.5	3.0	**0.15 CO**
N19909	909	38.0	<0.01	0.025	42.0	<0.001	0.30	0.10	0.10	0.07	1.60	4.8	**13.0 CO, 0.04 MO**
K09925	925	44.0	0.01	. . .	28.0	. . .	. . .	1.8	21.0	0.3	2.1	. . .	**3.0 MO**
K93600	36	35.8	0.05	0.40	63.2	0.010	. . .	. . .	. . .	. . .	. . .	. . .	**. . .**
K94100	42	40.63	0.03	0.36	58.9	0.004	0.21	. . .	0.02	. . .	. . .	. . .	**. . .**
K94800	48	48	0.03	0.40	52.0	0.010	. . .	. . .	. . .	. . .	. . .	. . .	**. . .**
N06333	RA333	45	0.05	. . .	18.0	. . .	. . .	. . .	25.0	. . .	. . .	. . .	**. . .**
N08020	20CB3	35.0	0.07	2.0	BAL	. . .	1.0	3.5	20.0	. . .	. . .	1.0	**. . .**
N08028	28	31.0	0.03	2.5	BAL	0.030	1.00	1.0	27.0	. . .	. . .	. . .	**3.5 MO**
N06030	G-30	BAL	0.03	1.5	15.0	0.02	0.8	1.6	29.7	. . .	. . .	0.9	**5.0 CO, 0.4 P, 2.7 W**
N10665	B-2	BAL	0.02	1.0	2.0	0.03	0.10	. . .	1.0	. . .	. . .	. . .	**1.0 CO, 28.0 MO, 0.04 P**
N08925	25-6MO	25.0	0.01	0.7	BAL	<0.001	0.30	1.0	20.0	. . .	. . .	. . .	**6.5 MO, 0.18 N**
N06985	**G-3**	**BAL**	**0.01**	**0.8**	**19.0**	**0.03**	**0.40**	**1.8**	**22.5**	**. . .**	**. . .**	**. . .**	**3.4 CO, 7.0 MO. 1.0 W**

The most widely employed processes for welding the non-age-hardenable (solid solution-strengthened) wrought nickel alloys are gas-tungsten arc welding (GTAW), gas-metal arc welding (GMAW), and shielded metal arc welding (SMAW). Submerged arc welding (SAW) and electroslag welding (ESW) have limited applicability, as does plasma arc welding (PAW). Although the GTAW process is preferred for welding the precipitation-hardenable alloys, both the GMAW and SMAW processes are also used.

Heat Treatment of Nickel Alloys

Preweld Heating and Heat Treating. Preweld heating of wrought alloys is not required, unless the base metal is below 16 °C (60 °F), in which case a path that is 255 to 305 mm (10 to 12 in.) wide on both sides of the joint should be warmed to 16 to 21 °C (60 to 70 °F) to avoid the condensation of moisture that can cause porosity in the weld metal.

Nickel alloys are usually welded in the solution-treated condition. Precipitation-hardenable (PH) alloys should be annealed before welding if they have undergone any operations that introduce high residual stresses (see the section "Welding of Precipitation-Hardenable Alloys" in this article).

Postweld Treatment. No postweld treatment, either thermal or chemical, is required to maintain or restore corrosion resistance, although in some cases a full solution anneal will improve corrosion resistance. Heat treatment may be necessary to meet specification requirements, such as stress relief of a fabricated structure to avoid age hardening or stress-corrosion cracking (SCC) of the weldment in hydrofluoric acid vapor or caustic soda. If welding induces moderate-to-high residual stresses, then the PH alloys would require a stress-relief anneal after welding and before aging (see the section "Welding of Precipitation-Hardenable Alloys" in this article).

Cleaning of Workpieces

Nickel and nickel alloys are susceptible to embrittlement by lead, sulfur, phosphorus, and other low-melting-point elements. These materials can exist in grease, oil, paint, marking crayons or inks, forming lubricants, cutting fluids, shop dirt, and processing chemicals. Work-pieces must be completely free of foreign material before they are heated or welded. Both sides of the workpiece should be cleaned in the area that will be heated by the welding operations. When no subsequent heating is involved, the cleaned area may be restricted to 50 mm (2 in.) on each side of the joint.

Shop dirt, oil, and grease can be removed by either vapor degreasing or swabbing with acetone or another nontoxic solvent. Paint and other materials that are not soluble in degreasing solvents may require the use of methylene chloride, alkaline cleaners, or special proprietary compounds. If alkaline cleaners that contain sodium carbonate are used, then the cleaners themselves must be removed prior to welding. Spraying or scrubbing with hot water is recommended. Marking ink can usually be removed with alcohol. Processing material that has become embedded in the work metal can be removed by grinding, abrasive blasting, and swabbing with a 10 vol% HCl solution, followed by a thorough water wash. Further information on the cleaning of nickel and nickel alloys is given in the article "Nickel Alloy Plating" in *Surface Engineering,* in Volume 5, *ASM Handbook.*

Oxides must also be removed from the area involved in the welding operation, primarily because of the difference between the oxide and base metal melting points. Oxides are normally removed by grinding, machining, abrasive blasting, or pickling.

Joint Design

Various joint designs are used in the arc welding of nickel alloys (Fig. 1). Although the same design can be used for GTAW and SMAW processes, the GMAW and SAW processes require special considerations. A joint design developed for other metals is not necessarily suitable for nickel alloys and should not be used unless it proves to be satisfactory, based on either experience or tests.

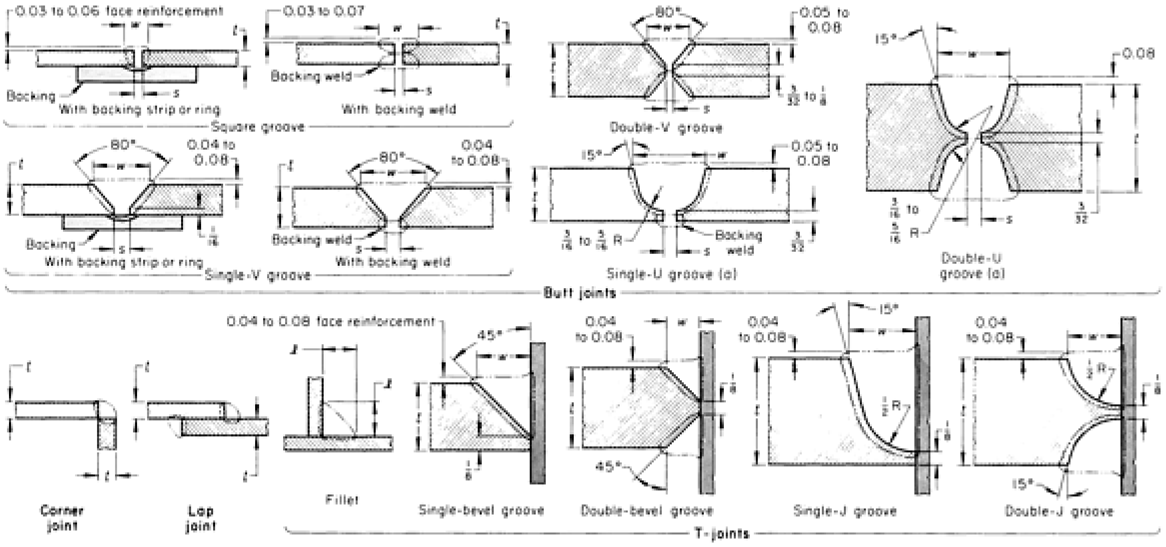
0.03 to 0.06 face reinforcement
w
t
Backing
s
With backing strip or ring
0.03 to 0.07
Backing weld
With backing weld
Square groove
80°
0.05 to 0.08
3/32 to 1/8
Double-V groove
15°
0.08
3/16 to 5/16 R
3/32
Double-U groove (a)
0.04 to 0.08
1/16
Single-V groove
Single-U groove (a)
Butt joints
0.04 to 0.08 face reinforcement
45°
1/8
1/2 R
Corner joint
Lap joint
Fillet
Single-bevel groove
Double-bevel groove
Single-J groove
Double-J groove
T-joints

BASE-METAL THICKNESS(*T*)		**WIDTH OF GROOVE OR BEAD (*W*)**		**MAXIMUM ROOT OPENING (*S*)**		**APPROXIMATE AMOUNT OF METAL DEPOSITED**		APPROXIMATE WEIGHT OF ELECTRODE	
mm	**in.**	**mm**	**in.**	**mm**	**in.**	**kg/m**	**lb/ft**	**kg/m**	lb/ft$^{(a)}$
SQUARE-GROOVE BUTT JOINT WITH BACKING STRIP OR RING									
0.94	0.037	3.2	$\frac{1}{8}$	0	0	0.03	0.02	0.038	**0.025**
1.27	0.050	4.0	$\frac{5}{32}$	0	0	0.06	0.04	0.07	**0.05**
1.57	0.062	4.8	$\frac{3}{16}$	0	0	0.06	0.04	0.09	**0.06**
2.36	0.093	4.8-6.4	$\frac{3}{16}$-$\frac{1}{4}$	0.8	$\frac{1}{32}$	0.09	0.06	0.12	**0.08**
3.18	0.125	6.4	$\frac{1}{4}$	1.6	$\frac{1}{16}$	0.10	0.07	0.13	**0.09**
SQUARE-GROOVE BUTT JOINT WITH BACKING WELD									
3.2	$\frac{1}{8}$	6.4	$\frac{1}{4}$	0.8	$\frac{1}{32}$	0.16	0.11	0.22	**0.15**
4.8	$\frac{3}{16}$	9.5	$\frac{3}{8}$	1.6	$\frac{1}{16}$	0.36	0.24	0.48	**0.32**
6.4	$\frac{1}{4}$	22	$\frac{7}{8}$	2.4	$\frac{3}{32}$	0.46	0.31	0.62	**0.42**
SINGLE-V-GROOVE BUTT JOINT WITH BACKING STRIP OR RING									
4.8	$\frac{3}{16}$	8.89	0.35	3.2	$\frac{1}{8}$	0.338	0.227	0.46	**0.31**
6.4	$\frac{1}{4}$	12.9	0.51	4.8	$\frac{3}{16}$	0.659	0.443	0.91	**0.61**

7.9	$\frac{5}{16}$	15.5	0.61	4.8	$\frac{3}{16}$	0.866	0.582	1.19	**0.80**
9.5	$\frac{3}{8}$	18.0	0.71	4.8	$\frac{3}{16}$	1.11	0.745	1.52	**1.02**
13	$\frac{1}{2}$	23.1	0.91	4.8	$\frac{3}{16}$	1.73	1.16	2.37	**1.59**
16	$\frac{5}{8}$	29.5	1.16	4.8	$\frac{3}{16}$	2.40	1.61	3.29	**2.21**
SINGLE-V-GROOVE BUTT JOINT WITH BACKING WELD									
6.4	$\frac{1}{4}$	10.4	0.41	2.4	$\frac{3}{32}$	0.62	0.42	0.86	**0.58**
7.9	$\frac{5}{16}$	12.9	0.51	2.4	$\frac{3}{32}$	0.80	0.54	1.10	**0.74**
9.5	$\frac{3}{8}$	16.5	0.65	3.2	$\frac{1}{8}$	1.09	0.73	1.49	**1.00**
13	$\frac{1}{2}$	21.6	0.85	3.2	$\frac{1}{8}$	1.80	1.21	2.48	**1.67**
16	$\frac{5}{8}$	26.9	1.06	3.2	$\frac{1}{8}$	2.17	1.46	2.98	**2.00**
DOUBLE-V-GROOVE BUTT JOINT									
13	$\frac{1}{2}$	10.2	0.40	3.2	$\frac{1}{8}$	1.32	0.89	1.73	**1.16**
16	$\frac{5}{8}$	12.4	0.49	3.2	$\frac{1}{8}$	1.61	1.08	2.20	**1.48**
19	$\frac{3}{4}$	15.7	0.62	3.2	$\frac{1}{8}$	2.17	1.46	2.98	**2.00**
25	1	20.6	0.81	3.2	$\frac{1}{8}$	3.60	2.42	4.97	**3.34**
32	1 $\frac{1}{4}$	26.2	1.03	3.2	$\frac{1}{8}$	4.34	2.92	5.95	**4.00**
SINGLE-BEVEL-GROOVE T-JOINT									
6.4	$\frac{1}{4}$	3.18	0.125	. . .	. . .	0.10	0.07	0.13	**0.09**
7.9	$\frac{5}{16}$	4.78	0.188	. . .	. . .	0.19	0.13	0.25	**0.17**
9.5	$\frac{3}{8}$	6.35	0.250	. . .	. . .	0.28	0.19	0.39	**0.26**
13	$\frac{1}{2}$	9.52	0.375	. . .	. . .	0.57	0.38	0.77	**0.52**
16	$\frac{5}{8}$	12.7	0.500	. . .	. . .	0.94	0.63	1.28	**0.86**
19	$\frac{3}{4}$	15.9	0.625	. . .	. . .	1.38	0.93	1.90	**1.28**
25	1	22.2	0.875	. . .	. . .	2.64	1.77	3.60	**2.42**
DOUBLE-BEVEL-GROOVE T-JOINT									
13	$\frac{1}{2}$	4.78	0.188	. . .	. . .	0.37	0.25	0.51	**0.34**
16	$\frac{5}{8}$	6.35	0.250	. . .	. . .	0.58	0.39	0.80	**0.54**
19	$\frac{3}{4}$	7.95	0.313	. . .	. . .	0.83	0.56	1.15	**0.77**
25	1	11.1	0.438	. . .	. . .	1.47	0.99	2.02	**1.36**
32	1 $\frac{1}{4}$	14.3	0.563	. . .	. . .	2.29	1.54	3.20	**2.15**
38	1 $\frac{1}{2}$	17.5	0.688	. . .	. . .	3.29	2.21	4.51	**3.03**
44	1 $\frac{3}{4}$	20.7	0.813	. . .	. . .	4.46	3.00	6.09	**4.09**

50	2	23.8	0.938	. . .	. . .	5.80	3.90	7.96	**5.35**
SINGLE-U-GROOVE BUTT JOINT(B)									
13	$\frac{1}{2}$	17.2	0.679	3.2	$\frac{1}{8}$	1.53	1.03	2.10	**1.41**
16	$\frac{5}{8}$	18.9	0.745	3.2	$\frac{1}{8}$	2.05	1.38	2.83	**1.90**
19	$\frac{3}{4}$	20.7	0.813	3.2	$\frac{1}{8}$	2.50	1.68	3.42	**2.30**
25	1	24.3	0.957	3.2	$\frac{1}{8}$	3.91	2.63	5.36	**3.60**
32	$1\frac{1}{4}$	27.3	1.073	3.2	$\frac{1}{8}$	5.39	3.62	7.38	**4.96**
38	$1\frac{1}{2}$	30.9	1.215	3.2	$\frac{1}{8}$	7.13	4.79	9.75	**6.55**
44	$1\frac{3}{4}$	34.3	1.349	3.2	$\frac{1}{8}$	8.90	5.98	12.2	**8.19**
50	2	37.7	1.485	3.2	$\frac{1}{8}$	11.0	7.40	15.1	**10.12**
DOUBLE-U-GROOVE BUTT JOINT(B)									
25	1	17.2	0.679	3.2	$\frac{1}{8}$	3.07	2.06	4.20	**2.82**
32	$1\frac{1}{4}$	18.9	0.745	3.2	$\frac{1}{8}$	4.11	2.76	5.65	**3.80**
38	$1\frac{1}{2}$	20.7	0.813	3.2	$\frac{1}{8}$	5.00	3.36	6.84	**4.60**
50	2	24.3	0.957	3.2	$\frac{1}{8}$	7.83	5.26	10.7	**7.20**
64	$2\frac{1}{2}$	27.3	1.073	3.2	$\frac{1}{8}$	10.8	7.24	14.8	**9.92**
CORNER AND LAP JOINT									
1.6	$\frac{1}{16}$	. . .	. . .	. . .	. . .	0.03	0.02	0.06	**0.04**
3.2	$\frac{1}{8}$	. . .	. . .	. . .	. . .	0.07	0.05	0.10	**0.07**
4.8	$\frac{3}{16}$	. . .	. . .	. . .	. . .	0.15	0.10	0.21	**0.14**
6.4	$\frac{1}{4}$	. . .	. . .	. . .	. . .	0.28	0.19	0.39	**0.26**
9.5	$\frac{3}{8}$	. . .	. . .	. . .	. . .	0.62	0.42	0.85	**0.57**
13	$\frac{1}{2}$	. . .	. . .	. . .	. . .	1.04	0.74	1.52	**1.02**
T-JOINT WITH FILLET									
. . .	. . .	. . .	. . .	3.2	$\frac{1}{8}$	0.04	0.03	0.06	**0.04**
. . .	. . .	. . .	. . .	4.8	$\frac{3}{16}$	0.10	0.07	0.15	**0.10**
. . .	. . .	. . .	. . .	6.4	$\frac{1}{4}$	0.18	0.12	0.24	**0.16**
. . .	. . .	. . .	. . .	7.9	$\frac{5}{16}$	0.28	0.19	0.39	**0.26**
. . .	. . .	. . .	. . .	9.5	$\frac{3}{8}$	0.40	0.27	0.55	**0.37**
. . .	. . .	. . .	. . .	13	$\frac{1}{2}$	0.70	0.47	0.95	**0.64**
. . .	. . .	. . .	. . .	16	$\frac{5}{8}$	1.10	0.74	1.50	**1.01**

. . .	. . .	. . .	. . .	19	$\frac{3}{4}$	1.59	1.07	2.17	**1.46**
. . .	. . .	. . .	. . .	25	1	2.83	1.90	3.87	**2.60**
SINGLE-J-GROOVE T-JOINT									
25	1	15.9	0.625	. . .	. . .	2.65	1.78	3.6	**2.4**
32	$1\frac{1}{4}$	18.3	0.719	. . .	. . .	3.72	2.50	5.1	**3.4**
38	$1\frac{1}{2}$	19.8	0.781	. . .	. . .	4.81	3.23	6.5	**4.4**
44	$1\frac{3}{4}$	22.2	0.875	. . .	. . .	6.09	4.09	8.3	**5.6**
50	2	24.6	0.969	. . .	. . .	7.34	4.93	10.1	**6.8**
57	$2\frac{1}{4}$	26.2	1.031	. . .	. . .	8.63	5.80	11.9	**8.0**
64	$2\frac{1}{2}$	27.8	1.094	. . .	. . .	10.3	6.94	14.1	**9.5**
DOUBLE-J-GROOVE T-JOINT									
25	1	12.7	0.500	. . .	. . .	2.20	1.48	3.0	**2.0**
32	$1\frac{1}{4}$	14.3	0.563	. . .	. . .	2.83	1.90	3.9	**2.6**
38	$1\frac{1}{2}$	15.1	0.594	. . .	. . .	3.81	2.56	5.2	**3.5**
44	$1\frac{3}{4}$	15.9	0.625	. . .	. . .	4.63	3.11	6.4	**4.3**
50	2	16.7	0.656	. . .	. . .	5.67	3.81	7.7	**5.2**
57	$2\frac{1}{4}$	17.5	0.688	. . .	. . .	6.71	4.51	9.2	**6.2**
64	$\mathbf{2}\frac{1}{2}$	**19.1**	**0.750**	**. . .**	**. . .**	**7.84**	**5.27**	**10.7**	**7.2**

(A) TO OBTAIN LINEAR FEET OF WELD PER POUND OF CONSUMABLE ELECTRODE, TAKE THE RECIPROCAL OF POUNDS PER LINEAR FOOT. IF THE UNDERSIDE OF THE FIRST BEAD IS CHIPPED OUT AND WELDED, ADD 0.31 KG/M (0.21 LB/FT) OF METAL DEPOSITED (EQUIVALENT TO 0.43 KG/M, OR 0.29 LB/FT, OF CONSUMABLE ELECTRODE).

(B) FOR GMAW (EXCEPT WITH THE SHORT-CIRCUITING ARC), ROOT RADIUS SHOULD BE ONE-HALF THE VALUE SHOWN AND BEVEL ANGLE SHOULD BE TWICE AS GREAT.

FIG. 1 TYPICAL JOINT CONFIGURATIONS AND DIMENSIONS FOR GTAW, GMAW, AND SMAW APPLICATIONS. DIMENSIONS IN SCHEMATIC GIVEN IN INCHES

Beveled Joints. Beveling is usually not required for metal that is 2.4 mm (0.093 in.) or less in thickness, although thinner sections of certain alloys are sometimes beveled. Metal thicker than 2.4 mm (0.093 in.) should be beveled to form a V-, U-, or J-shaped groove and should be welded using either a backing material or a gas, unless welding from both sides. Otherwise, erratic penetration can result, leading to crevices and voids that are potential areas for joint weakness and accelerated corrosion in the underside of the joint. To attain the best underbead contour on joints that cannot be welded from both sides, the GTAW process should be used for the root pass.

For metal that is more than 9.5 mm ($\frac{3}{8}$ in.) thick, a double U- or V-shaped groove design is preferred. The added cost of preparation is justified, because less welding time and material are required and less residual stress is developed than with a single-groove design.

Beveling is best accomplished by machine, usually a plate planer or other machine tool. Plasma arc and electric arc cutting can be used for joint penetration, but all oxidized metal must be removed from the joint area by grinding or chipping for a depth of 0.8 to 1.6 mm ($\frac{1}{32}$ to $\frac{1}{16}$ in.).

Corner and lap joints can be welded when service stresses will not be excessive. However, they should be avoided if service temperatures are high or if service conditions involve thermal or mechanical cycling. When comer joints are used, a full-penetration weld must be made, usually with a fillet weld on the root side. Lap joints should be welded on both sides.

Design Considerations. Molten nickel alloy weld metal does not flow as easily as steel weld metal. This characteristic cannot be compensated for by using extra heat input or pooling, because serious loss of residual deoxidizers may result. For joints in metal up to 16 mm ($\frac{5}{8}$ in.) thick, ample accessibility is provided by V-groove butt joints beveled to a 60° groove angle. For thicker metal, U-groove butt joints are machined to a 15° bevel angle and a 4.8 to 7.9 mm ($\frac{3}{16}$ to $\frac{5}{16}$ in.) radius. Single bevels used to form T joints should have an angle of 45°. The bottom radius of a J-shaped groove for T joints should be at least 9.5 mm ($\frac{3}{8}$ in.), and the bevel angle described above should be used for J- and U-shaped grooves. For full-penetration welds, the bevel and groove angles should be about 30% more than those used for carbon steels.

Welding of Nickel Alloys

Revised by E.B. Hinshaw, Inco Alloys International, Inc.

Welding Fixtures

Proper fixturing and clamping hold the workpieces firmly in place, minimize buckling, maintain alignment, and, when needed, provide compressive stress in the weld metal. A backing bar, or any portion of a fixture that might contact the workpiece, should be made of copper. This metal has such a high thermal conductivity that inadvertent welding against the copper will reduce the chance that the bar will fuse to the weldment. In addition, the grooves in copper bars are often used to contour, without sticking, the side underneath the weld.

Backing bars, when used, should incorporate a groove of suitable contour to permit penetration of weld metal and avoid the possibility of trapping gas or flux at the bottom of the weld. Grooves in backing bars used in the GMAW process usually have a shallow semielliptical shape that is 0.38 to 0.89 mm (0.015 to 0.035 in.) deep and 4.8 to 6.4 mm ($\frac{3}{16}$ to $\frac{1}{4}$ in.) wide. Square corner grooves are used with the GMAW and GTAW processes to accommodate the backing gas. A machined passageway is connected to the gas supply, and holes that are 1.6 mm ($\frac{1}{16}$ in.) in diameter are drilled 75 mm (3 in.) apart from the bottom of the groove to the passageway, permitting the backing gas to flow along the weld. The gas flows out of the groove at the ends of the bar. Figure 2 shows groove designs for backing bars.

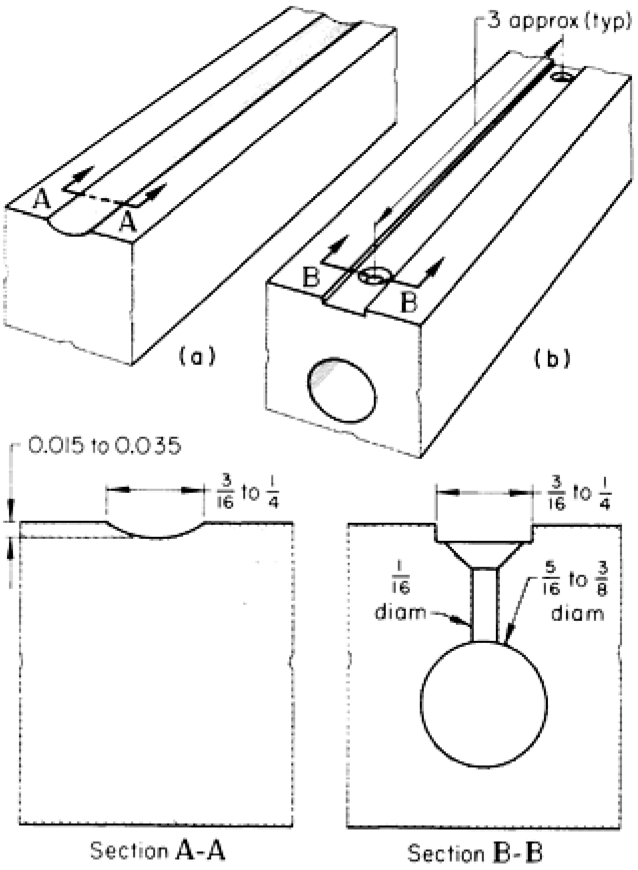

FIG. 2 GROOVE CONFIGURATIONS FOR BACKING BARS. (A) STANDARD GROOVE FOR USE WITHOUT A BACKING GAS. (B) SQUARE-CORNER GROOVE EMPLOYED WITH BACKING GAS (GROOVE DEPTH DEPENDS ON GAS FLOW REQUIRED AND LENGTH OF BAR). DIMENSIONS GIVEN IN INCHES

The clamping and restraint requirements for nickel alloys are about the same as for low-carbon steel. The hold-down bars should be located close enough to the line of the weld to maintain alignment and provide the proper degree of heat transfer. Generally, hold-down pressure should be just enough to maintain alignment, but high restraint can be advantageous when welding square groove joints in thin metals using the GMAW process without a filler metal. When the pieces to be welded are positioned with a zero root opening and hold-down bars are brought very close to the line of welding, and when a high hold-down pressure is used, the heat of welding creates an expansive force that results in compression in the line of the weld. This compression will have an upsetting effect on the hot weld metal and can cause the weld to develop a slight top and bottom crown, or weld reinforcement, without the use of filler metal.

Welding of Nickel Alloys

Revised by E.B. Hinshaw, Inco Alloys International, Inc.

Welding of Precipitation-Hardenable Alloys

The PH alloys require special welding procedures because of their susceptibility to cracking. Cracks can occur in the base-metal heat-affected zone (HAZ) upon aging or in service at temperatures above the aging temperature, as a result of residual welding stress and stress induced by precipitation. Before welding these alloys, a full-solution anneal is usually performed. After welding, the appropriate aging heat treatment is performed. To further improve alloy properties, a full anneal after welding, followed by a postweld heat treatment, can be incorporated in the welding procedure.

Preweld and Postweld Treatments. Any part that has been subjected to severe bending, drawing, or other forming operations should be annealed before welding. If possible, heating should be done in a controlled-atmosphere furnace to limit oxidation and minimize subsequent surface cleaning.

The aluminum and titanium hardened alloys must be stress relieved (solution treated) after welding and before precipitation hardening. To avoid prolonged exposure of the welded structure to temperatures within the precipitation-hardening range, rapid heating in a furnace preheated to the appropriate temperature is recommended. When the workpiece is large, compared with the furnace area, it may be necessary to preheat the furnace to a temperature that is 110 to 280 °C (200 to 500 °F) above the solution-treatment temperature and then to reset the furnace controls when the workpiece has reached the solution-treatment temperature. The stresses created by repair or alteration welding must be relieved in a similar fashion by rapid heating to the solution-treating temperature prior to re-aging. If satisfactory stress relief of the weldment is not feasible, particularly if the structure is complicated, then preweld overaging treatments may be helpful. Preheating, however, is not a satisfactory substitute for postweld heat treatment. Although PH alloys can be welded in the aged condition, if temperatures encountered in service are in the PH range, then the weldment must be solution treated and reaged.

General Welding Procedures. Precipitation-hardenable alloys are usually welded by the GTAW process, but SMAW and GMAW processes are also applicable. Heat input during the welding operations should be held to a moderately low level in order to obtain the highest possible joint efficiency and minimize the extent of the HAZ. For multiple-bead or multiple-layer welds, many narrow stringer beads should be used, rather than a few large, heavy beads.

Any oxides that form during welding should be removed by abrasive blasting or grinding. If such films are not removed as they accumulate on multiple-pass welds, then they can become thick enough to inhibit weld fusion and produce unacceptable laminar-type oxide stringers along the weld axis.

Welding of Cast Nickel Alloys

Cast nickel alloys can be joined by the GTAW, GMAW, and SMAW processes. For optimum results, castings should be solution annealed before welding to relieve some of the casting stresses and provide some homogenization of the cast structure.

Light peening of solidified metal after the first pass will relieve stresses and, thus, reduce cracking at the junction of the weld metal and the cast metal. The peening of subsequent passes is of little, if any, benefit. Stress relieving after welding is also desirable.

Gas-Tungsten Arc Welding

Nickel alloys, both cast and wrought and either solid-solution-strengthened or precipitation-hardenable, can be welded by the GTAW process. The addition of filler metal is usually recommended. Direct current electrode negative (DCEN) is recommended for both manual and machine welding. Parameters for manual GTAW of alloy 400 are given in Table 2.

TABLE 2 PARAMETERS FOR MANUAL GTAW OF 1.6 MM (0.062 IN.) THICK ALLOY 400

JOINT TYPE	**BEVELED BUTT, 1.6 MM ($\frac{1}{16}$ IN.) ROOT OPENING**
ELECTRODE	**2.4 MM ($\frac{3}{32}$ IN.) DIAM EWTH-2, TAPERED TO 0.4 MM ($\frac{1}{64}$ IN.) DIAM**
FILLER METAL	**1.6 MM ($\frac{1}{16}$ IN.) DIAM ERNICU-7**
NUMBER OF PASSES	**3**
WELDING CURRENT, A	**70-90 (DCEN)**
VOLTAGE, V	**10-12**
SHIELDING GAS:	
AT TORCH, M^3/H (FT^3/H)	**0.6 (20)**
BACKING GAS, M^3/H (FT^3/H)	**0.08-0.14 (3-5)**
PREHEAT AND POSTHEAT TREATMENT	**NONE**
INTERPASS TEMPERATURE, °C (°F)	**175 (350) MAXIMUM**

The welding torch should be set or held at an angle of 90° to the work. A slight deviation is permissible to provide a better view of the work, but an acute angle can lead to aspiration of the surrounding air and contamination of the shielding gas. The largest gas nozzle size applicable to the job should be used, and a minimum practical distance between the nozzle and the work should be maintained. A gas lens should be used to improve shielding gas flow and coverage (see the article "Gas-Tungsten Arc Welding" in this Volume).

Shielding Gas. Either argon or helium, or a mixture of the two, is used as a shielding gas for welding nickel and nickel alloys. Additions of oxygen, carbon dioxide, or nitrogen to argon gas will usually cause porosity or erosion of the electrode. Argon with small quantities of hydrogen (typically 5%) can be used and may help avoid porosity in pure nickel, as well as aid in reducing oxide formation during welding. The advantages and disadvantages of using helium, rather than argon, when welding thin metal without the addition of filler metal, are:

- *IMPROVED SOUNDNESS:* POROSITY-FREE WELDS ARE MORE EASILY OBTAINED IN ALLOY 400, AND THE POROSITY OF WELDS IN NICKEL 200 IS LESS.
- *INCREASED WELDING SPEED:* WITH DCEN, WELDING SPEED CAN BE INCREASED AS MUCH AS 40% OVER THAT ACHIEVED WITH ARGON AT THE SAME CURRENT SETTING, BECAUSE WELDING SPEED IS A FUNCTION OF HEAT INPUT, AND HEAT INPUT IS CONSIDERABLY GREATER WITH HELIUM.
- *DECREASED ARC STABILITY:* AT LOWER CURRENT LEVELS (<120 A), ARC STABILITY IS LOWERED AS HELIUM CONTENT IS INCREASED. MIXTURES OF ARGON AND HELIUM HELP STABILIZE THE ARC WHILE MAINTAINING THE BENEFITS OF THE HELIUM SHIELDING GAS.

Because the purity and dryness of the welding gas are important, welding-grade shielding gas should be used. Furthermore, the welding area should be screened to avoid drafts, because disruption of the shielding atmosphere and drawing in of air can cause porosity in the weld metal.

Electrodes. Pure-tungsten electrodes or tungsten alloyed with thoria or zirconia can be used. The alloyed electrodes are more economical, because of their lower vaporization loss. Overheating that is due to the use of excessive amperage must be avoided at all times.

The best arc stability and penetration control are achieved by tapering the electrode tip. The taper angle should be approximately 30°, with a flat land of about 0.38 mm (0.015 in.) diameter on the tip end. Larger taper angles are used to produce a narrower bead and deeper penetration. After its selection, the taper angle should be maintained, because changes in the angle can affect penetration and bead width.

Electrode extension (stick out) should be short, and the amount should be based on joint design. For example, a maximum extension of 4.8 mm ($\frac{3}{16}$ in.) can be used for butt welds in thin metal, but a 9.5 to 13 mm ($\frac{3}{8}$ to $\frac{1}{2}$in.) extension is required for some fillet welds or for reaching the bottom of the groove welds.

The electrode should be inclined slightly in the forehand position. Filler metal, when used, should be added carefully at the leading edge of the weld pool to avoid contact with and contamination of the electrode. If contamination occurs, then the electrode should be cleaned and reshaped.

Filler Metals. Compositions of the filler metals used with the GTAW process are, in general, similar to those of the base metals with which they are used. Because of high welding temperatures, filler metals are alloyed to resist porosity and hot cracking of the weld metal. Filler-metal additions and dilution ratios should be adjusted to ensure that the weld metal contains about 75% filler metal.

Table 3 lists the compositions of filler metals used for the gas-tungsten arc welding of nickel and nickel alloys. Filler metals of the ERNi-1 classification are used for welding the high-nickel-content alloys (for example, Nickel 200 and 201), whereas those of the ERNiCu-7 classification are used for welding nickel-copper alloys (for example, alloys 400 and 404).

TABLE 3 COMPOSITION OF FILLER METALS FOR GTAW, GMAW, AND SAW APPLICATIONS OF NICKEL ALLOYS

UNS NO.	**ELECTRODE**	COMPOSITION, WT%(A)													
		C MAX	**MN**	**FE**	**S**	**SI**	**CU**	**NI + CO**	**AL MAX**	**TI**	**P**	**NB + TA**	**MO**	**CR**	OTHER
N02061	ERNI-1	0.15	1.0	1.0	0.015	0.75	0.25	93.0 MIN	1.5	2.0-3.5	0.03	. . .	. . .	. . .	**. . .**
N04060	ERNICU-7	0.15	4.0	2.5	0.015	1.25	BAL	62.0-69.0	1.25	1.5-3.0	0.02	. . .	. . .	. . .	**. . .**
N08065	ERNIFECR-1	0.05	1.0	22.0 MIN	0.03	0.50	1.5-3.0	38.0-46.0	0.20	0.60-1.2	0.03	. . .	2.50-3.50	19.5-23.5	**. . .**
N07092	ERNICRFE-6	0.03	2.0-2.7	8.0	0.015	0.35	0.50	67.0 MIN	. . .	2.5-3.5	0.03	. . .	. . .	14.0-17.0	**. . .**
N06082	ERNICR-3	0.10	2.5-3.5	3.0	0.015	0.50	0.50	67.0 MIN	. . .	0.75	0.03	2.0-3.0	. . .	18.0-22.0	**. . .**
N06062	ERNICRFE-5	0.08	1.0	6.0-10.0	0.015	0.35	0.50	70.0 MIN	. . .	. . .	. . .	1.5-3.0	. . .	14.0-17.0	**. . .**
N06052	ERNICRFE-7	0.04	1.0	7.0-11.0	0.015	0.50	0.30	BAL	1.10	1.10	0.020	0.10	0.50	28.0-31.5	**. . .**
N06007	ERNICRMO-1	0.05	1.0-2.0	18.0-21.0	0.04	1.0	1.5-2.5	BAL	. . .	. . .	0.04	1.75-2.50	5.5-7.5	21.0-23.5	**. . .**
N06625	ERNICRMO-3	0.10	0.50	5.0	0.015	0.50	0.50	58 MM	0.40	0.40	0.02	3.15-4.15	8.0-10.0	20.0-23.0	**. . .**
N10276	ERNICRMO-4	0.02	1.0	4.0-7.0	0.03	0.08	0.50	BAL	. . .	. . .	0.04	2.5	15.0-17.0	14.5-16.5	**2.5 CO, 0.35 V, 3.0-4.5 W**
N06985	ERNICRMO-9	0.015	1.0	18.0-21.0	0.03	1.0	1.5-2.5	BAL	. . .	. . .	0.04	. . .	6.0-8.0	21.0-23.5	**1.5 W, 5.0 CO**
N06022	ERNICRMO-10	0.02	1.0	2.0-6.0	0.015	0.20	. . .	BAL	. . .	. . .	0.03	. . .	12.5-14.5	20-22.5	**2.5-3.5 W**
N07718	ERNIFECR-2	0.08	0.35	BAL	0.015	0.35	0.30	50-55	0.2-0.8	0.65-1.15	0.015	4.75-5.50	2.8-3.3	17.0-21.0	**0.006 B**
N06617	ERNICRCOMO-1	0.05-0.15	1.0	3.0 MAX	0.015	1.0	0.50	BAL	0.8-1.5	0.60	0.03	. . .	8.0-10.0	20.0-24.0	**. . .**
N10665	ERNIMO-7	0.02	8.0	2.0	0.03	0.10	0.50	BAL	. . .	. . .	0.04	. . .	26.0-30.0	1.0	**1.0 CO**
N06030	ERNICRMO-11	0.03	1.5	13.0-	0.02	0.80	1.0-	BAL	. . .	. . .	0.04	0.30-	4.0-	28.0-	**5.0 CO, 1.5-**

				17.0			2.4					1.50	6.0	31.5	**4.0 W**
N10004	ERNIMO-3	0.12	1.00	4.0-7.0	0.03	1.00	0.50	BAL	. . .	. . .	0.04	. . .	23.0-26.0	4.0-6.0	**2.5 CO, 0.60 V**
N10001	**ERNIMO-1**	**0.08**	**1.00**	**4.0-7.0**	**0.025**	**1.00**	**0.50**	**BAL**	**. . .**	**. . .**	**0.03**	**. . .**	**26.0-30.0**	**1.00**	**2.5 CO, 0.20-0.40 V**

(A) SINGLE VALUES ARE MAXIMUM (EXCEPT WHERE OTHERWISE SPECIFIED).

Joint Design. All joint designs shown in Fig. 1 can be used when welding nickel alloys by the GTAW process. If no filler metal is used, then the sections to be joined must be held tightly together to promote satisfactory fusion.

Welding Techniques. When a filler metal is used, the wire diameter should be related to the work-metal thickness. For example, a 1.6 mm ($\frac{1}{16}$ in.) diameter wire is generally used to weld 1.6 mm ($\frac{1}{16}$ in.) thick sheet. During welding, the hot end of the wire must remain under the shielding gas to avoid oxidation. A short arc length should be maintained to ensure that deoxidizing elements transfer to the weld pool.

To ensure a sound weld, the arc should be maintained at the shortest possible length. When no filler metal is added, the arc length should be 1.3 mm (0.05 in.), maximum, and preferably 0.5 to 0.8 mm (0.02 to 0.03 in.). When filler metal is used, the arc should be kept as short as possible, consistent with filler-metal diameter, to avoid loss of gas coverage. This loss can result in oxidation and reduced penetration. A short arc also reduces weld spatter with filler wire.

The speed of welding affects penetration, weld width, and weld soundness, especially when no filler metal is added. For a given thickness of metal, there is a range of welding speeds that results in minimum porosity. Speeds outside this range, either faster or slower, can result in increased porosity. When welding thinner-gage material, the required heat input is considerably lower. This contributes to increased porosity, which is due to the increased solidification rate of the molten weld pool.

Complete-penetration welds require the use of grooved backing bars that permit local gas shielding (Fig. 2b), or the purging of the inside of the workpiece. The surface condition affects the porosity level in thin sheet material to a much larger degree, because the surface layer makes up a larger percentage of the overall thickness of material.

Revised by E.B. Hinshaw, Inco Alloys International, Inc.

Gas-Metal Arc Welding

The solid-solution nickel-base alloys can be joined by the GMAW process. In addition, the use of special procedures allows PH alloys (for example, alloy K-500) to be gas-metal arc welded.

Spray, globular, and short-circuiting metal transfer methods are suitable. Varying the power input produces the different types of metal transfer. The pulsed arc process is also used. In these methods, the filler metal is a current-carrying consumable and is typically 0.89, 1.14, or 1.57 mm (0.035, 0.045, or 0.062 in.) in diameter. Table 4 lists typical conditions for the welding of nickel-base alloys using spray-type metal transfer (Table 4(a)), short-circuiting metal transfer (Table 4(b)), and pulsed-arc transfer (Table 4(c)) with various nickel-base filler metals.

TABLE 4 TYPICAL PARAMETERS FOR GMAW APPLICATIONS OF NICKEL 200 AND ALLOY 400

PARAMETER	**SPRAY METAL TRANSFER**(A)		**SHORT-CIRCUITING METAL TRANSFER**(B)		PULSED-ARC METAL TRANSFER(C)	
	NICKEL 200	**ALLOY 400**	**NICKEL 200**	**ALLOY 400**	**NICKEL 200**	ALLOY 400
ELECTRODE WIRE DESIGNATION	ERNI-1	ERNICU-7	ERNI-1	ERNICU-7	ERNI-1	**ERNICU-7**
AVERAGE VOLTAGE, V	29-31	28-30	20-21	16-18	21-22	**21-22**
PEAK VOLTAGE, V	. . .	. . .	. . .	. . .	46	**40**
AVERAGE CURRENT, A	375	290	160	130-135	150	**110**
WIRE FEED RATE, MM/S (IN./MIN)	**87 (205)**	**85 (200)**	**152 (360)**	**116-123 (275-290)**	**68 (160)**	**59 (140)**

(A) ARGON SHIELDING GAS FLOW RATE AT 1.7 M^3/H (60 FT^3/H); FLAT WELDING POSITION; 1.57

MM (0.062 IN.) DIAM ELECTRODE WIRE.

(B) ARGON-HELIUM SHIELDING GAS FLOW RATE AT 1.4 M^3/H (50 FT^3/H); VERTICAL WELDING POSITION; 0.89 MM (0.035 IN.) DIAM WIRE.

(C) ARGON OR ARGON-HELIUM SHIELDING GAS FLOW RATE AT 0.71 TO 0.99 M^3/H (25 TO 35 FT^3/H); VERTICAL WELDING POSITION 1.14 MM (0.045 IN.) DIAM ELECTRODE

TABLE 4(A) TYPICAL PARAMETERS FOR SPRAY TRANSFER GMAW OF NICKEL-BASE ALLOYS

Workpiece was flat welded with argon shielding gas (flow rate of 1.7 m^3/h, or 60 ft^3/h).

FILLER METAL	**WIRE DIAMETER**		**WIRE FEED RATE**		**VOLTAGE, V**	CURRENT, A
	MM	**IN.**	**M/MIN**	**IN./MIN**		
ERNI-1	0.9	0.035	10.8-13.2	425-520	26-32	**200-300**
	1.1	0.045	7.0-8.1	275-320	26-32	**250-325**
	1.6	0.062	4.4-5.6	175-220	27-33	**275-350**
ERNICU-7	0.9	0.035	12.1-13.2	475-520	26-32	**175-230**
	1.1	0.045	6.4-7.6	250-300	26-32	**225-300**
	1.6	0.062	3.8-5.1	150-200	27-33	**250-300**
ERCUNI	0.9	0.035	12.1-14.6	475-575	26-32	**200-300**
	1.1	0.045	6.4-8.1	250-320	26-32	**250-325**
	1.6	0.062	4.4-5.8	175-230	27-33	**275-350**
ERNICRFE-5	0.9	0.035	10.8-13.2	425-520	26-32	**175-240**
	1.1	0.045	6.4-7.9	250-310	26-32	**225-300**
	1.6	0.062	4.4-5.6	175-220	27-33	**250-330**
ERNICR-3, ERNIFECR-1, ERNICRMO-1	0.9	0.035	11.4-13.2	450-520	26-32	**175-240**
	1.1	0.045	6.4-7.9	250-310	26-32	**225-300**
	1.6	0.062	3.2-5.1	125-200	27-33	**250-330**
ERNICRFE-6	0.9	0.035	11.4-13.2	450-520	26-32	**175-240**
	1.1	0.045	6.4-8.1	250-320	26-32	**225-300**
	1.6	0.062	3.2-5.1	125-200	27-33	**250-330**
ERNICRCOMO-1, ERNICRMO-3, ERNICRMO-4 THROUGH -11	0.9	0.035	11.4-15.2	450-600	26-32	**180-245**
	1.1	0.045	6.4-8.9	250-350	26-32	**225-300**
	1.6	**0.062**	**3.2-5.7**	**125-225**	**27-33**	**250-330**

TABLE 4(B) TYPICAL PARAMETERS FOR SHORT-CIRCUITING TRANSFER GMAW OF NICKEL-BASE ALLOYS

Data applicable to vertical welding of 3.18 mm (0.125 in.) thick base metal using 65Ar-35H_2 shielding gas at a flow rate of 0.99 m^3/h (35 ft^3/h)

FILLER METAL(A)	OPEN-CIRCUIT VOLTAGE, V		WELDING VOLTAGE, V		WELDING CURRENT, A		WIRE FEED RATE				SECONDARY INDUCTANCE(B)		SLOPE(B)	
							M/MIN		IN./MIN					
	RANGE	TYPICAL	RANGE	TYPICAL	RANGE	TYPICAL	RANGE	TYPICAL	RANGE	TYPICAL	RANGE	TYPICAL	RANGE	TYPICAL
ERNI-1	28-34	30	20-24	22	90-150	110	5.1-11.4	6.0	200-450	235	5-12	10	$6\frac{1}{2}$-$8\frac{1}{2}$	**$7\frac{1}{2}$**
ERNICU-7	27-33	30	19-24	22	90-150	110	5.1-11.4	7.0	200-450	275	7-12	10	$6\frac{1}{2}$-$8\frac{1}{2}$	**8**
ERCUNI	27-36	28	19-25	20	100-175	130	5.7-11.4	7.0	225-450	275	5-14	$7\frac{1}{2}$	6-$7\frac{1}{2}$	**$6\frac{1}{2}$**
ERNICR-3	29-36	33	21-26	23	90-150	115	4.4-9.5	7.0	175-375	275	10-17	15	7-$9\frac{1}{2}$	**$9\frac{1}{2}$**
ERNICRMO-3	**29-36**	**33**	**21-26**	**23**	**90-150**	**110**	**4.4-9.5**	**7.0**	**175-375**	**275**	**10-17**	**15**	**7-$9\frac{1}{2}$**	**$9\frac{1}{2}$**

(A) ROD DIAMETER OF 0.9 MM (0.035 IN.).

(B) RELATIVE VALUES ONLY. HIGHER NUMBERS DENOTE GREATER SECONDARY INDUCTANCE OR SLOPE.

TABLE 4(C) TYPICAL PARAMETERS FOR PULSED-ARC TRANSFER GMAW OF NICKEL-BASE ALLOYS

Data applicable for vertical welding with 1.1 mm (0.045 in.) wire and argon shielding gas flowing at 0.8 m^3/h (30 ft^3/h)

FILLER METAL	TYPICAL WIRE FEED RATE		PEAK VOLTAGE, V		AVERAGE VOLTAGE, V		CURRENT, A	
	M/MIN	IN./MIN	RANGE	TYPICAL	RANGE	TYPICAL	RANGE	TYPICAL
ERNI- 1	4.1	160	45-46	46	21-22	21	90-150	**130**
ERNICU-7	3.6	140	39-40	40	21-22	21	90-150	**125**
ERNICR-3	**3.6**	**140**	**43-44**	**44**	**20-22**	**21**	**90-150**	**110**

Welding current is most often provided by constant-voltage power supplies, although all standard direct current (dc) power supplies are satisfactory. Direct current electrode positive (DCEP) is used because it is virtually impossible to transfer metal in a controlled manner when using DCEN. The amperage should be well within the maximum rating of the equipment, but it should be sufficient to obtain the desired melting rate.

The shielding gas used in the gas-metal arc welding of nickel alloys by spray or globular transfer can be argon, which yields good results. The addition of 15 to 20% He is beneficial when welding nickel alloys with short-circuiting arc or pulsed-arc transfer methods. As the helium content increases from 0 to 20%, the weld beads become progressively wider and flatter, and penetration decreases. The addition of oxygen or carbon dioxide to argon stabilizes the arc, but results in heavily oxidized and irregular bead surfaces. With standard constant-voltage GMAW power sources, helium alone has been used as a shielding gas, but it creates an unsteady arc with excessive weld spatter and is a difficult atmosphere in which to initiate an arc. With pulsed GMAW power sources in which the waveform can be controlled, helium can be very effective for overlaying, resulting in low base metal dilution.

Gas flow rates vary from 0.71 to 2.8 m^3/h (25 to 100 ft^3/h), depending on joint design and position. A flow rate of 1.4 m^3/h (50 ft^3/h) is used most frequently. The choice of gas used with short-circuiting metal transfer is influenced by the type of equipment available. Argon, without additions, is suitable for use with equipment that has both inductance and slope control. Argon gas produces convex beads, which may cause cold lapping (lack of fusion), but it also provides a pronounced pinch effect that can be controlled by inductance. The addition of helium is helpful when induction and slope cannot be varied. Helium imparts a wetting action, and the arc is hotter. These factors greatly decrease the possibility of cold lapping. As the percentage of helium increases, the gas flow rate must be increased to ensure adequate protection.

The size of the gas nozzle is important when using a short-circuiting arc. When a 1:1 mixture of argon and helium is used at a flow rate of 1.1 m^3/h (40 ft^3/h) through a 9.5 mm ($\frac{3}{8}$ in.) diameter nozzle, at a wire feed rate of 6.4 m/min (250 in./min), and a wire diameter of 1.1 mm (0.045 in.) the current can be increased to 160 to 180 A without adversely affecting weld quality. An argon-helium mixture is recommended as the shielding gas with the pulsed-arc GMAW process. The gas flow rate should range from 0.42 to 1.3 m^3/h (15 to 45 ft^3/h). An excessive flow rate can interfere with the arc.

Filler metals used in the GMAW process are listed in Table 3. The appropriate wire diameter depends on the method used and the thickness of the base metal. With globular or spray transfer, 0.89, 1.14, and 1.57 mm (0.035, 0.045, and 0.062 in.) diameter wire is used. The short-circuiting and pulsed-arc processes generally require wire that is 1.14 mm (0.045 in.) or less in diameter.

Joint designs recommended for the GMAW process are shown in Fig. 1. For U-shaped groove designs that use globular or spray transfer methods, the root radius should be decreased by about 50%, and the bevel angle should be doubled. With these transfer methods, the use of high amperage on small-diameter wire produces a high level of arc force. As a result, the arc cannot be deflected from a straight line, as is possible when the welding process utilizes covered electrodes. Because the arc must contact all areas to be fused, the joint design must provide an intersection with the arc force line. When a short-circuiting arc is used, the U-shaped groove designs are the same as those employed with the other arc welding processes.

Welding Techniques. Positioning the electrode holder vertically along the centerline of the joint yields the best results. Some inclination is permissible to allow a better view of the work, but excessive displacement can cause the surrounding atmosphere to be drawn into the shielding gas, causing porous or heavily oxidized welds. Arc length is also important. Too short an arc causes spatter, and too long an arc results in loss of control and reduced penetration.

Because arc length is directly proportional to voltage, particular care must be paid to the voltage setting to ensure properly controlled arc length. In addition, once a procedure has been qualified, consistent torch-to-work positioning should be maintained to prevent possible changes in transfer mode caused by loss (or gain) of energy available at the arc. The alteration in energy at the arc is caused by I^2R energy changes corresponding to wire extension changes.

Cold lapping (lack of fusion) can occur with the short-circuiting, globular, and pulsed-arc processes if manipulation is faulty. The torch should be advanced at a rate that will keep the arc in contact with the base metal, just ahead of the weld pool. The manipulation and angle used with the pulsed-arc torch are similar to those used in the SMAW process. To avoid undercutting, a slight pause should be made at the limit of the weave. Excessive welding speeds and improper torch positioning can cause oxides on the surface of underlying passes to become entrapped between weld passes. Slower welding speeds and good torch positioning allow the arc to reduce certain oxides and/or allow time to float the oxides to the surface of the molten weld puddle.

Welding of Nickel Alloys

Revised by E.B. Hinshaw, Inco Alloys International, Inc.

Plasma Arc Welding

Using the keyholing mode, the PAW process can produce acceptable welds in nickel alloys that are up to about 7.6 mm (0.3 in.) thick. Argon-hydrogen mixtures are used as orifice and shielding gases. Levels from 5 to 8% H_2 are optimum. The amount of current needed for keyholing decreases as hydrogen content increases (up to about 7% H_2). Above this level, torch starting becomes more difficult. If helium or hydrogen is used, then the arc starting ability is subsequently reduced.

Typical travel speed and current parameters for keyhole welding are:

TRAVEL SPEED		CURRENT, A	
MM/MIN	**IN./MIN**	**NICKEL 200**	ALLOY 400
205	8	185	**155**
255	L0	200	**175**
305	12	220	**195**
355	**14**	**235**	**215**

Typical conditions for welding 5.97 mm (0.235 in.) thick Nickel 200 using 95% Ar-5% H_2 gas are: current, 245 A; voltage, 31.5 V; feed rate, 355 mm/min (14 in./min); and gas flow, 0.28 m^3/h (10 ft^3/h) (orifice) and 1.3 m^3/h (45 ft^3/h) (shielding) (Ref 1, 2).

References cited in this section

1. A.C. LINGENFELTER, *WELD. ENG.,* JAN 1970, P 42-45
2. A.C. LINGENFELTER *ET AL., WELD. J.,* MAY 1966, P 417-422

Welding of Nickel Alloys

Revised by E.B. Hinshaw, Inco Alloys International, Inc.

Shielded Metal Arc Welding

The SMAW process can be used for welding nickel and nickel alloys. Although the minimum metal thickness is usually about 1.3 mm (0.050 in.), thinner metal can be welded when appropriate fixtures are provided. The types of joints used, as well as bead and groove dimensions, are given in Fig. 2.

Electrodes. The composition of electrodes used in the SMAW process are given in Table 5. Electrode composition is selected to be similar to that of the base metal with which the electrode is to be used. Electrodes of the ENi-1 classification are used to weld wrought and cast forms of nickel and nickel alloys to themselves and to steel. ENiCu electrodes can be used to weld nickel-copper alloys to themselves, to surface steel with a nickel-copper alloy, to weld the clad side of a nickel-copper-clad steel, and to weld nickel-copper alloys to steel. ECuNi electrodes can be used to weld nickel-copper, copper-nickel, and nickel wrought alloys when overlaying steel. A buttering layer of ENi-1 is usually placed onto the steel surface before a copper-nickel surface layer is made.

TABLE 5 COMPOSITION OF ELECTRODES USED FOR SMAW OF NICKEL-ALLOYS, HIGH-NICKEL CONTENT ALLOYS, AND NICKEL-COPPER ALLOYS

UNS NO.	**ELECTRODE**	COMPOSITION, WT %														
		C	**MN**	**FE**	**S**	**SI**	**CU**	**NI + CO**	**AL**	**TI**	**NB + TA**	**CR**	**MO**	**P**	**W**	OTHER
W82141	ENI-1	0.10	0.75	0.75	0.02	1.25	0.25	92.0 MIN	1.0	1.0-4.0	. . .	. . .	. . .	. . .	. . .	**. . .**
W84190	ENICU-7	0.15	4.00	2.5	0.015	1.5	BAL	62.0-69.0	0.75	1.0	. . .	. . .	. . .	0.02	. . .	**. . .**
W60715	ECUNI	0.15	1-2.5	0.4-.75	0.015	0.5	BAL	29 MIN	. . .	0.5	. . .	. . .	. . .	0.02	. . .	**. . .**
W86112	ENICRMO-3	0.10	1.0	7.0	0.02	0.75	0.50	55.0 MIN	. . .	. . .	3.15-4.15	20.0-23.0	8.0-10.0	0.03	. . .	**2.5 CO, 0.35 V**
W80276	ENICRMO-4	0.02	1.0	4.0-7.0	0.03	0.2	0.50	BAL	. . .	. . .	. . .	14.5-16.5	15.0-17.0	0.04	3.0-4.5	**. . .**
W86985	ENICRMO-9	0.02	1.0	18.0-21.0	0.03	1.0	1.5-2.5	BAL	. . .	. . .	0.5	21.0-23.5	6.0-8.0	0.04	1.5	**5.0 CO**
W86022	ENICRMO-10	0.02	1.0	2.0-6.0	0.015	0.2	0.5	BAL	. . .	. . .	. . .	20.0-22.5	12.5-14.5	0.03	2.5-3.5	**0.35 V, 2.5 CO**
W86030	ENICRMO-11	0.03	1.5	13.0-17.0	0.02	1.0	1.0-2.4	BAL	. . .	1.5-4.0	0.3-1.5	28.0-31.5	4.0-6.0	0.04	1.5-4.0	**5.0 CO**
W86040	ENICRMO-12	0.03	2.2	5.0	0.02	0.7	0.50	BAL	. . .	. . .	1.0-2.8	20.5-22.5	8.8-10.0	0.03	. . .	**. . .**
W86132	ENICRFE-1	0.08	3.5	11.0	0.015	0.75	0.50	62.0 MIN	. . .	. . .	1.5-4.0	13.0-17.0	. . .	0.03	. . .	**. . .**
W86133	ENICRFE-2	0.10	1.0-3.5	12.0	0.02	0.75	0.50	62.0 MIN	. . .	. . .	0.5-3.0	13.0-17.0	0.50-2.50	0.03	. . .	**0.50-2.50 V**
W86134	ENICRFE-3	0.10	5.0-9.5	10.0	0.015	1.0	0.50	59.0 MIN	. . .	1.0	1.0-2.5	13.0-17.0	. . .	0.03	. . .	**. . .**
W80001	ENIMO-1	0.07	1.0	4.0-7.0	0.03	1.0	0.50	BAL	. . .	. . .	. . .	1.0	26.0-30.0	0.04	1.0	**0.60 V, 2.5 CO**
W80665	ENIMO-7	0.02	1.75	2.0	0.03	0.2	0.50	BAL	. . .	. . .	. . .	1.0	26.0-30.0	0.04	1.0	**1.0 CO**
W86152	ENICRFE-7	0.05	5.0	7.0-12.0	0.015	0.75	0.5	BAL	0.50	0.50	1.0-2.5	28-31.5	0.50	0.030	. . .	**. . .**

W86117	**ENICRCOMO-1**	**0.05-0.15**	**0.3-2.5**	**5.0**	**0.015**	**0.75**	**0.50**	**BAL**	**. . .**	**. . .**	**1.0**	**21.0-26.0**	**8.0-10.0**	**0.03**	**. . .**	**9.0-15.0 CO**

(A) SINGLE VALUES ARE MAXIMUM (EXCEPT WHERE OTHERWISE SPECIFIED).

ENiCrMo-3 electrodes typically are used to join a wide range of pitting and crevice corrosion-resistant alloys. The lowest alloyed materials employed with these electrodes are the molybdenum-modified type 317 stainless steels. ENiCrMo-3 electrodes are used for richer alloy compositions of the nickel-chromium-iron-molybdenum families, up to and including alloy 625. Intermediate alloys welded with ENiCrMo-3 include alloy 825, alloy 20, alloy G, and alloy G-3. This type of electrode is widely used because its high nickel content allows good dissimilar weldability, whereas the high molybdenum content matches or exceeds the pitting resistance of the base alloys being welded. Because of the strength of the ENiCrMo-3 electrode, welding to many high-strength low-alloy (HSLA) grade steels can be performed without sacrificing strength.

Other electrodes in this alloy group, such as ENiCrMo-4 and ENiCrMo-10, are used to join alloys to themselves or to steel. They are also used to join alloys in the ENi-Cr-Mo group to each other. Usually, the base metal is welded with a higher alloyed welding electrode to counteract the microsegregation that occurs in the weld deposit. For example, in a pitting environment, such as the SO_2 environments found in flue gas desulfurization (FGD) applications, the choice would be a nickel-chromium-molybdenum alloy group with a higher molybdenum-containing welding electrode family.

Nickel-chromium-iron electrodes, such as ENiCrFe-2 and ENiCrFe-3, are designed to weld the same family of alloys, as well as dissimilar metal joints involving carbon steel, stainless steel, nickel, and nickel-base alloys. Other welding electrodes with the ENi-Cr-Fe electrode grouping are special-application chemistry electrodes, such as ENi-Cr-Fe-7, which are used to weld alloy 690 to itself, to alloy 600, and to butt weld alloy 690 to steel or to overlay certain grades of steel in nuclear boiler and pressure water reactor applications.

Nickel-chromium-cobalt-molybdenum electrodes, such as ENi-Cr-Co-Mo-1, are readily used to join nickel-chromium-iron alloys, especially when service temperatures will exceed 790 °C (1450 °F) and could reach up to 1150 °C (2100 °F) in oxidizing atmospheres.

Nickel-molybdenum electrodes, such as ENMo-1, are designed to weld nickel-molybdenum alloys to themselves and to other nickel-, cobalt-, and iron-base metals. They are normally used in the flat position only.

Covered electrodes should be kept in sealed, moisture-proof containers in a dry storage area prior to their use. All opened containers of unused electrodes should be stored in a cabinet equipped with a dessicant or heated to 6 to 8 °C (10 to 15 °F) above ambient temperature prior to use. If the electrodes are exposed to excessive moisture, then they can be reclaimed by rebaking at 260 °C (500 °F) for 2 h or at 315 °C (600 °F) for 1 h. The electrode diameter should be chosen on the basis of quality requirements and position of welding, rather than on speed of production.

Welding Current. The SMAW process utilizes DCEP. Each electrode size has an optimum amperage range in which good arcing characteristics prevail and outside of which the arc becomes unstable or the electrode overheats. Suggested electrode diameters and current settings for various metal thicknesses are shown in Table 6. Variables such as type of backing, tightness of clamping, and joint design influence the required amount of current density. Actual welding currents should be developed by making sample welds in metal of the same composition and thickness as the metal to be welded.

TABLE 6 RECOMMENDED ELECTRODE DIAMETER AND WELDING CURRENT FOR SMAW OF SELECTED THICKNESSES OF HIGH-NICKEL-CONTENT ALLOYS AND NICKEL-COPPER ALLOYS IN THE FLAT POSITION

BASE-METAL THICKNESS		ELECTRODE DIAMETER[(A)]		CURRENT[(B)], A
MM	IN.	MM	IN.	
HIGH-NICKEL-CONTENT ALLOYS				
0.94	0.037	2.4	$\frac{3}{32}$	[(C)]
1.09	0.043	2.4	$\frac{3}{32}$	[(C)]
1.27	0.050	2.4	$\frac{3}{32}$	[(C)]
1.57	0.062	2.4	$\frac{3}{32}$	**75**

1.98	0.078	2.4	$\frac{3}{32}$	**80**
2.36	0.093	2.4	$\frac{3}{32}$	**85**
2.77	0.109	3.2	$\frac{1}{8}$	**105**
3.18	0.125	3.2	$\frac{1}{8}$	**150**
3.18	0.125	2.4-4.0	$\frac{3}{32}$-$\frac{5}{32}$	**80-150**
3.56	0.140	4.0	$\frac{5}{32}$	**130**
3.96	0.156	4.0	$\frac{5}{32}$	**135**
[GES]4.75	[GES]0.187	4.0	$\frac{5}{32}$	**150**
NICKEL-COPPER ALLOYS				
0.94	0.037	2.4	$\frac{3}{32}$	**(C)**
1.09	0.043	2.4	$\frac{3}{32}$	**(C)**
1.27	0.050	2.4	$\frac{3}{32}$	**(C)**
1.57	0.062	2.4	$\frac{3}{32}$	**50**
1.98	0.078	2.4	$\frac{3}{32}$	**55**
2.36	0.093	2.4	$\frac{3}{32}$	**60**
2.77	0.109	2.4	$\frac{3}{32}$	**60**
2.77	0.109	3.2	$\frac{1}{8}$	**65**
3.18	0.125	2.4-4.0	$\frac{3}{32}$-$\frac{5}{32}$	**60-140**
3.56	0.140	2.4-4.0	$\frac{3}{32}$-$\frac{5}{32}$	**60-140**
3.96	0.156	2.4-4.0	$\frac{3}{32}$-$\frac{5}{32}$	**60-140**
6.35	0.250	2.4-4.0	$\frac{3}{32}$-$\frac{5}{32}$	**60-140**
9.53	0.375	2.4-4.8	$\frac{3}{32}$-$\frac{3}{16}$	**60-180**
[GES]12.7	**[GES]0.500**	**2.4-4.8**	$\frac{3}{32}$-$\frac{3}{16}$	**60-180**

(A) WHERE A RANGE IS SHOWN, THE SMALLER DIAMETER ELECTRODES ARE USED FOR THE FIRST PASSES AT THE BOTTOM OF THE GROOVE, AND THE JOINTS ARE COMPLETED WITH THE LARGER DIAMETER ELECTRODES.

(B) CURRENT SHOULD BE IN THE RANGE RECOMMENDED BY THE ELECTRODE MANUFACTURER.

(C) USE MINIMUM AMPERAGE AT WHICH ARC CONTROL CAN BE MAINTAINED.

Welding Position. Flat-position welding should be used whenever possible, because it is faster, more economical, and produces a weld of good quality. The recommended electrode position for flat-position welding is an inclination of about 20° from the vertical, ahead of the weld pool. This position facilitates control of the molten flux and prevents slag entrapment. A short arc must be maintained to prevent atmospheric contamination of the arc and weld and to prevent porosity.

For vertical welding, the arc should be slightly shorter than for flat-position welding, and the amperage should be 10 to 20% lower than the values in Table 6. The electrode should be approximately a 90° angle to the joint. For overhead welding, the arc should be slightly shorter than for flat-position welding, and the welding current should be reduced by 5 to 15 A from the values shown in Table 6.

Welding Techniques. Because molten nickel alloy weld metal is not as free flowing as iron-base alloys, it must be deposited where it is needed. Therefore, a slight weave is required. The amount of weave will depend on joint design, welding position, and type of electrode. Although a straight stringer bead made without weaving can be used for single-pass work, and is satisfactory at the bottom of a deep groove on thicker sections, a weave is generally desirable.

When a weave is used, it should be no wider than approximately three times the electrode diameter. Some deviation from this role may be necessary during vertical welding. A slight weave, if it is not excessive, can help prevent possible centerline solidification cracking problems.

Weld spatter should be avoided. Spatter indicates that the arc is too long, that excessive amperage is being used, or that the current is straight polarity (DCEN). Other causes of spatter are wet electrodes or base metal and slag that run under the arc during welding.

Arc blow may occur when one or both of the metals being welded is magnetic. The arc is deflected from its normal path by a magnetic force in the base metal. One method of overcoming arc blow is to change the location of the ground connection on the workpiece or to change the direction of the electrical path to the arc. A dissimilar joint (for example, steel welded to a nickel-base alloy) can cause considerable arc blow toward the steel side of the joint.

When the arc is to be broken, it should be shortened slightly, and the rate of travel should be increased to reduce the size of the weld pool. This practice reduces the probability of crater cracking, ensures that the crater does not develop a rolled leading edge, and prepares the way for restriking the arc. Another technique to prevent crater cracking is called crater filling, in which the weld pool is held stationary for a short period of time before the arc is broken. This technique is usually incorporated at the end of a weld.

The manner in which the arc is restruck has a significant effect on the soundness of the weld. A reverse (or T) restrike is recommended. The arc should be struck at the leading edge of the crater and carried back to the extreme rear of the crater at a normal welding speed. The direction is then reversed, weaving is commenced, and the weld is continued. Advantages of this procedure are:

- THE WELDER HAS AN OPPORTUNITY TO ESTABLISH THE CORRECT ARC LENGTH BEFORE ACTUAL WELDING COMMENCES.
- SOME PREHEAT IS APPLIED TO THE COLD CRATER.
- THE FIRST DROPS OF RAPIDLY COOLED WELD METAL ARE DEPOSITED WHERE THEY WILL BE REMELTED, THUS RESTRICTING THE POROSITY TO A MINIMUM.

Another technique is to make the restrike where the weld metal can be readily removed (for example, 13 to 25 mm, or $\frac{1}{2}$ to 1 in., behind the crater on top of the previous pass). Later, the restrike area can be ground to be level with the rest of the bead. This technique is used when the weld must meet rigid radiographic standards and calls for less welder skill than the reverse-restrike method.

Welding procedures should be qualified before production begins. In one plant where this was not done, qualifying tests of the welding procedure were conducted before further welds were made when cracks appeared in the weld metal of a joint in a 13 mm ($\frac{1}{2}$ in.) thick Nickel 200 cylinder.

The original procedure for this weld specified a V-shaped groove butt joint with a 60° included angle, a 3.2 mm ($\frac{1}{8}$ in.) root face, and zero root opening (Fig. 3a). An ENi-1 electrode that was 355 mm (14 in.) long by 48 mm ($\frac{3}{16}$ in.) wide was used with a welding current of 235 A. On the first pass, a weld bead that was 165 mm ($6\frac{1}{2}$ in.) long was deposited with each electrode. On the second pass, a weld bead that was 75 mm (3 in.) long was deposited with each electrode. A cross section of the completed weld is shown in Fig. 3(b). An investigation showed that cracks in the weld metal were caused by the high welding current and low travel speed, which had resulted in the weld pool being held in a superheated condition for an excessively long time.

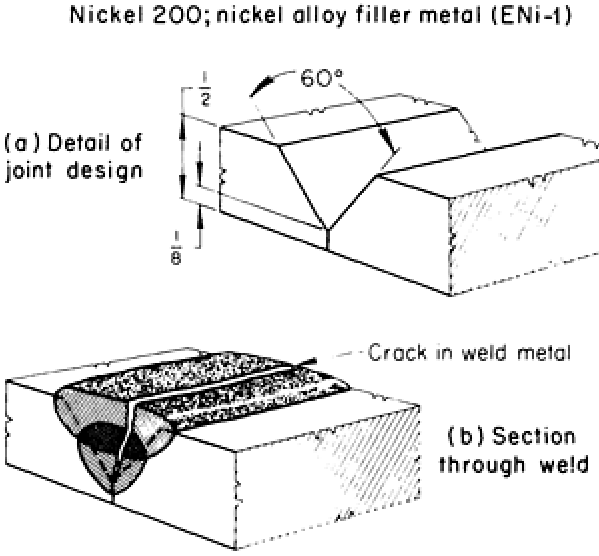
Nickel 200; nickel alloy filler metal (ENi-1)
60°
1/2
1/8
(a) Detail of joint design
Crack in weld metal
(b) Section through weld

FIG. 3 JOINT CONFIGURATIONS FOR WELDING 13 MM ($\frac{1}{2}$ IN.) NICKEL 200 AND SUBSEQUENT CRACK IN WELD METAL CAUSED BY EXCESSIVE HEAT INPUT. DIMENSIONS GIVEN IN INCHES

Therefore, a qualified procedure was established, specifying a welding current of 210 A and an increased travel speed, which resulted in a 305 mm (12 in.) long weld for each electrode on the first pass and 230 mm (9 in.) long welds per electrode on the second pass. Electrode type and length, and all other welding conditions, were unchanged. An included angle of 80° and a 1.6 to 2.4 mm ($\frac{1}{16}$ to $\frac{3}{32}$ in.) root face would have produced a better welded joint, but the workpieces had been machined and roll formed before the problem arose.

Cleaning the Weld Bead. In multiple-pass welding, all flux and slag must be removed before each succeeding bead is deposited. All slag should be completely removed from completed welds, especially if service involves high temperatures. The slag is easily removed with either hand tools or a hand or power wire brush.

Welding of Nickel Alloys

Revised by E.B. Hinshaw, Inco Alloys International, Inc.

Submerged Arc Welding

Submerged arc welding can be used to join solid-solution nickel alloys. Alloy 400, Nickel 200, and several Ni-Cr-Fe alloys are most frequently welded by this process. Joints in metal up to 75 mm (3 in.) thick have met American Society of

Mechanical Engineers (ASME) codes and other specification requirements. The SAW process cannot be used to weld the PH nickel alloys.

Joint Design. Some joint designs used in the submerged arc welding of nickel-alloy plate are shown in Fig. 4. Single compound-angle grooves, single U-grooves, and double U-grooves are used for metal that is 19 mm ($\frac{3}{4}$ in.) or more thick. The double U-groove is usually preferred, because it results in a lower level of residual stress and can be completed in less time, with less filler metal. A single V-groove is used for stock up to 13 mm (1 in.) thick.

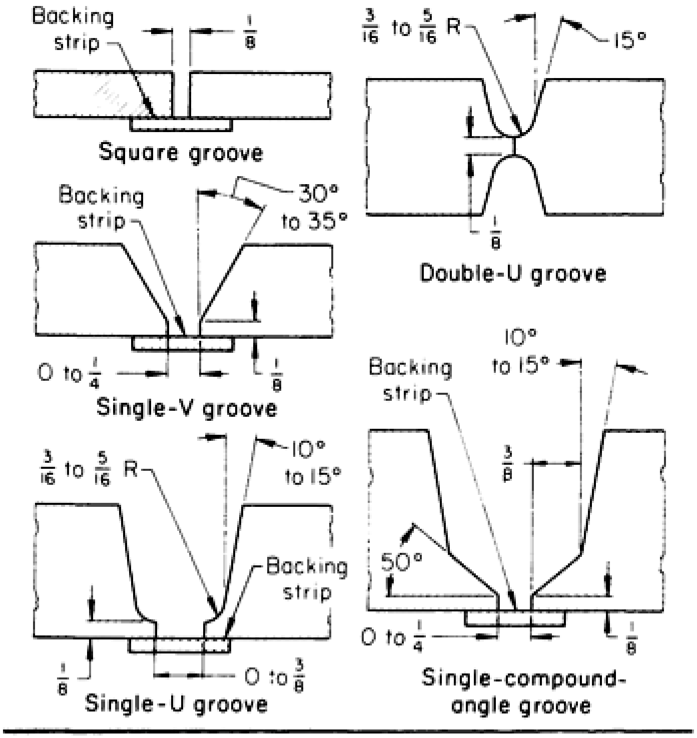
Backing strip
$\frac{1}{8}$
Square groove
$\frac{3}{16}$ to $\frac{5}{16}$ R
15°
$\frac{1}{8}$
Double-U groove
Backing strip
30° to 35°
0 to $\frac{1}{4}$
$\frac{1}{8}$
Single-V groove
$\frac{3}{16}$ to $\frac{5}{16}$ R
10° to 15°
Backing strip
$\frac{1}{8}$
0 to $\frac{3}{8}$
Single-U groove
10° to 15°
Backing strip
$\frac{3}{8}$
50°
0 to $\frac{1}{4}$
$\frac{1}{8}$
Single-compound-angle groove

FIG. 4 TYPICAL DIMENSIONS FOR SAW OF NICKEL ALLOYS. BACKING STRIPS ARE OPTIONAL WITH ZERO ROOT OPENINGS. DIMENSIONS GIVEN IN INCHES

Electrodes. The compositions of the electrode wires used in the SAW process are the same as those of the filler metals or electrode wires used in GTAW and GMAW processes (Table 3). Submerged arc welding is possible with the following electrodes: ERNi-1, ER-CuNi, ERNiCu-7, ERNiCr-3, ERNiCrFe-5, and ERNiCrMo-3.

Wire that ranges from 1.14 to 2.36 mm (0.045 to 0.093 in.) in diameter can be used for all nickel alloys. The 1.6 mm ($\frac{1}{16}$ in.) diameter wire is generally preferred. Small-diameter wires are used to weld metal that is up to 13 mm ($\frac{1}{2}$ in.) thick, and 2.4 mm ($\frac{3}{32}$ in.) diameter wires are used for heavier sections. The approximate amounts of filler metal needed for a single compound angle, a single U-groove, and a single V-groove are given in Table 7.

TABLE 7 QUANTITY OF FILLER METAL REQUIRED TO PRODUCE SUBMERGED ARC WELDS IN THREE TYPES OF GROOVES

JOINT DESIGN								APPROXIMATE AMOUNT OF METAL REQUIRED (IN POUNDS PER LINEAR FOOT) FOR INDICATED PLATE THICKNESS (IN INCHES) OF:								
ANGLE	**ANGLE β**	**RADIUS (*R*)**		**ROOT OPENING (*S*)**		**DIMENSIONS (S_1)**										
		MM	**IN.**	**MM**	**IN.**	**MM**	**IN.**	**1**	**1**$\frac{1}{4}$	**1**$\frac{1}{2}$	**1**$\frac{3}{4}$	**2**	**2**$\frac{1}{2}$	**3**	**3**$\frac{1}{2}$	4
SINGLE-COMPOUND-ANGLE GROOVE																
10°	40°	. . .	. . .	0	0	9.5	$\frac{3}{8}$	2.19	3.10	4.10	5.17	6.33	8.90	11.77	14.95	**18.50**
10°	40°	. . .	. . .	0	0	13	$\frac{1}{2}$	2.74	3.56	4.84	6.16	7.61	10.45	13.73	17.32	**21.00**
10°	50°	. . .	. . .	0	0	9.5	$\frac{3}{8}$	1.95	2.83	3.79	4.82	5.93	8.40	11.18	14.00	**17.70**
15°	50°	. . .	. . .	0	0	9.5	$\frac{3}{8}$	2.03	2.95	4.08	5.29	6.62	9.68	13.20	17.30	**21.78**
10°	50°	. . .	. . .	6.4	$\frac{1}{4}$	9.5	$\frac{3}{8}$	2.87	3.98	5.17	6.43	7.77	10.70	13.94	17.22	**21.38**
10°	40°	. . .	. . .	6.4	$\frac{1}{4}$	9.5	$\frac{3}{8}$	3.11	4.25	5.48	6.78	8.17	11.20	14.53	18.17	**22.18**
SINGLE U-GROOVE																
10°	. . .	4.8	$\frac{3}{16}$	0	0	. . .	. . .	1.46	2.07	2.76	3.53	4.37	6.33	8.60	11.20	**14.10**
10°	. . .	6.4	$\frac{1}{4}$	0	0	. . .	. . .	1.77	2.47	3.25	4.12	5.06	7.20	9.67	12.45	**15.60**
10°	. . .	7.9	$\frac{5}{16}$	0	0	. . .	. . .	2.03	2.86	3.68	4.71	5.76	8.10	10.70	13.70	**17.00**
15°	. . .	4.8	$\frac{3}{16}$	0	0	. . .	. . .	1.60	2.36	3.23	4.22	5.34	7.94	11.00	14.60	**18.68**
15°	. . .	6.4	$\frac{1}{4}$	0	0	. . .	. . .	1.90	2.73	3.68	4.76	5.96	8.74	12.00	15.78	**20.00**
15°	. . .	7.9	$\frac{5}{16}$	0	0	. . .	. . .	2.17	3.08	4.14	5.30	6.60	9.55	12.90	16.90	**21.40**
10°	. . .	4.8	$\frac{3}{16}$	6.4	$\frac{1}{4}$	. . .	. . .	2.38	3.22	4.14	5.14	6.21	8.63	11.36	14.42	**17.78**

10°	. . .	6.4	$\frac{1}{4}$	6.4	$\frac{1}{4}$	. . .	. . .	2.69	3.62	4.63	5.73	6.90	9.50	12.43	15.67	**19.28**
SINGLE V-GROOVE																
30°	. . .	. . .	. . .	0	0	. . .	. . .	1.63	. . .	. . .	. . .	. . .	. . .	. . .	. . .	**. . .**
35°	. . .	. . .	. . .	0	0	. . .	. . .	1.96	. . .	. . .	. . .	. . .	. . .	. . .	. . .	**. . .**
30°	**. . .**	**. . .**	**. . .**	**6.4**	$\frac{1}{4}$	**. . .**	**. . .**	**2.55**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**

Fluxes. The fluxes used in the submerged arc welding of carbon and stainless steels are not satisfactory when welding nickel alloys. Special proprietary fluxes are available and must be used. Loss of alloy elements (especially chromium), poor weld contour, flux entrapment, weld cracking, and inclusions can result when the wrong flux is used.

Only enough flux cover to prevent arc flashing should be used because excessive flux can cause a deformed bead surface. Slag entrapment can be prevented by appropriate joint design and by correct placement of beads. Slag is easily removed from welds in the bottom of grooves. Fused flux is self-lifting from exposed welds and should be discarded. Unfused flux can be recovered by means of a vacuum system, and it can be reused, if clean. Screening to adjust particle size is not required.

Fluxes should be stored in dry storage areas. Opened containers should be resealed to prevent moisture pickup. However, flux that has absorbed moisture can be reclaimed by drying at 315 °C (600 °F) for 1 h. Hoppers used for fluxes appropriate for the welding of steel and other metals should be thoroughly cleaned before being filled with flux appropriate for the welding of nickel alloys.

Welding Current. Direct current electrode negative (typically used for groove plus cladding applications) or direct current electrode positive (typically used for cladding applications) can be used. Reverse polarity (DCEP) is preferred for butt welding, because it produces a flatter bead with deeper penetration at a rather low arc voltage (30 to 33 V). Straight polarity (DCEN) results in a slightly higher rate (typically 30% higher) of deposition at increased voltage (over 35 V) and requires oscillation. The flux covering, however, must be appreciably deeper. Consequently, flux consumption and the risk of slag entrapment increase. Alternating current and the two wire series technique for multiple arc welding are not suitable for use with the available fluxes.

Bead Deposition. The bead location in a multiple-pass layer should provide an open, or reasonably wide, root area for the next bead. Slightly convex beads are preferred to concave beads. Bead contour is controlled by voltage, travel speed, and the position of the electrode.

Welding of Nickel Alloys

Revised by E.B. Hinshaw, Inco Alloys International, Inc.

Minimizing Weld Defects

The defects and metallurgical difficulties encountered in the arc welding of nickel alloys include:

- POROSITY
- SUSCEPTIBILITY TO HIGH-TEMPERATURE EMBRITTLEMENT BY SULFUR AND OTHER CONTAMINANTS
- CRACKING IN THE WELD BEAD, CAUSED BY HIGH HEAT INPUT AND EXCESSIVE WELDING SPEEDS
- STRESS-CORROSION CRACKING IN SERVICE

Porosity. Oxygen, carbon dioxide, nitrogen, or hydrogen can cause porosity in welds. In the SMAW and SAW processes, porosity can be minimized by using electrodes that contain deoxidizing or nitride-forming elements, such as aluminum and titanium. These elements have a strong affinity for oxygen and nitrogen and form stable compounds with them. Presence of deoxidizers in either type of electrode serves to reduce porosity. In addition, porosity is much less likely to occur in chromium-bearing nickel alloys than in non-chromium-bearing alloys.

In the GMAW and GTAW processes, porosity can be avoided by preventing the access of air to the molten weld metal. Gas backing on the underside of the weld is sometimes used.

In the GTAW process, the use of argon with up to 10% H_2 as a shielding gas helps to prevent porosity. Bubbles of hydrogen that form in the weld pool gather the diffusing nitrogen. Too much hydrogen (>15%) in the shielding gas can result in hydrogen porosity.

Cracking. Hot shortness of welds can result from contamination by sulfur, lead, phosphorus, cadmium, zinc, tin, silver, boron, bismuth, or any other low-melting-point elements, which form intergranular films and cause severe liquid-metal embrittlement at elevated temperatures. Many of these elements are found in soldering and brazing filler metals. Hot cracking of the weld metal usually results from such contamination. Cracking in the heat-affected zone is often caused by intergranular penetration of contaminants from the base-metal surface. Sulfur, which is present in most cutting oils used for machining, is a common cause of cracking in nickel alloys. The removal of foreign material from the surfaces of the work metal is imperative (see the section "Cleaning of Workpieces" in this article). Weld metal cracking also can be caused by heat input that is too high, as a result of high welding current and low welding speed (Fig. 3). Welding speeds have a large effect on the solidification pattern of the weld. High welding speeds create a tear-drop molten weld pool, which leads to uncompetitive grain solidification at the center of the weld. At the weld centerline, residual elements will collect and cause centerline hot cracking or lower transverse tensile properties.

In addition, cracking may result from undue restraint. When conditions of high restraint are present, as in circumferential welds that are self restraining, all bead surfaces should be slightly convex. Although convex beads are virtually immune to centerline splitting, concave beads are particularly susceptible to centerline cracking. In addition, excessive width-to-depth or depth-to-width ratios can result in cracking that may be internal (that is, subsurface cracking).

Stress-Corrosion Cracking. Nickel and nickel alloys do not experience any metallurgical changes, either in the weld metal or in the HAZ, that affect normal corrosion resistance. When the alloys are intended to contact substances such as concentrated caustic soda, fluorosilicates, and some mercury salts, however, the welds may need to be stress relieved to avoid stress-corrosion cracking. Nickel alloys have good resistance to dilute alkali and chloride solutions. Because resistance to stress-corrosion cracking increases with nickel content, the stress relieving of welds in the high-nickel-content alloys is not usually needed.

Effect of Slag on Weld Metal. Because fabricated nickel alloys are ordinarily used in high-temperature service and in aqueous corrosive environments, all slag should be removed from finished weldments. If slag is not removed in the latter type of application, then crevices and accelerated corrosion can result. Slag inclusions between weld beads reduce the strength of the weld. If the service temperature approximates the melting point of the slag, then severe corrosion can occur when slag is present on the weld surfaces, particularly in oxygen-containing atmospheres. Fluorides in the slag can react with moisture or elements in the environment to create highly corrosive compounds.

Slag also acts as an accumulator of sulfur, particularly in reducing atmospheres, which can lead to service failure in atmospheres that would be considered adequately low in sulfur. In one instance, although only 0.01% S was present in the atmosphere, the sulfur content of the slag on the weld surface rose from 0.02% to 2 or 3% after a one-month exposure. Sulfur pickup also depresses the melting temperature of the weld slag and, consequently, the maximum safe operating temperature of the weldment.

Joining of Dissimilar Metals

Nickel and nickel alloys have been successfully joined to other nickel alloys and carbon and low-alloy steels, stainless steel, and a number of copper alloys. Filler metal, or consumable electrode, must be selected to ensure a compatible metallurgical relationship between the two base metals. Several factors are involved:

- DIFFERENCES IN THERMAL EXPANSION OF THE BASE METALS
- POSSIBILITY OF PERMANENT CHANGES IN VOLUME AFTER EXTENDED SERVICE AT ELEVATED TEMPERATURE
- EFFECT OF WELD-METAL DILUTION AT THE INTERFACES WITH THE BASE METAL

All of these factors influence the choice of filler metal and welding process.

An example of a good metallurgical combination is the welding of nickel to alloy 400. Because these metals are completely compatible, they can be welded to each other by any welding process, using any compatible filler metal.

Dilution of Weld Metal. The composition of the weld metal can be expected to differ from that of the filler metal, or consumable electrode, and the base metals being joined, because some elements are lost through the arc when they are transferred from the filler-metal electrode to the weld metal during welding. Filler metal and base metal mixing results in the formation of a unique alloy. The composition of the weld metal is fairly uniform for a given bead, except in a narrow band at the edge of the weld, just adjacent to the fusion line in the near heat-affected zone of the weld. This area is sometimes referred to as the "unmixed zone." Although this region can never entirely be eliminated, its size can be minimized by lowering the overall heat input during welding. The amount of dilution varies from bead to bead. The welding process, current density, deposition rate, welding speed, weld-metal thickness, and welding technique have a significant influence on the amount of base-metal dilution that occurs in the weld deposit. Another region that forms in the near heat-affected zone of the weld is the "partially melted zone." This region can lower the overall corrosion resistance of the weldment, but can be controlled by lowering or limiting the heat input during welding.

The dilution of a nickel-base alloy by a dissimilar metal can be tolerated only to a limited degree. When a stainless steel filler metal is used to weld alloy 400 to an austenitic stainless steel, any significant amount of copper pickup from the alloy 400 causes the weld metal to become hot short and to crack. Thus, stainless steel filler metal should not be used when the welding process can cause a considerable amount of dilution. Neither nickel-copper nor copper-nickel types of filler metal can be used in this instance, because chromium from the stainless steel dilutes the weld metal and causes cracking. Although nickel or a nickel-chromium-iron or nickel-chromium-molybdenum filler metal is best, they should only be used after the welding procedure has been qualified by tests.

Filler Metals/Welding Electrodes. Selection of the correct filler metal helps produce a weld that meets the service requirements for strength, corrosion resistance, and soundness and tolerates dilution without causing cracking susceptibility in the weld metal. Crack sensitivity of the weld metal is proportional to the amount of dilution, especially when there is a considerable difference in the composition of a base metal and a filler metal.

Some combinations of base metal and filler metal produce undesirable weld-metal compositions:

- A FERRITIC WELD-METAL DEPOSIT DILUTED BY NICKEL, CHROMIUM, OR COPPER
- AN 18-8 STAINLESS STEEL WELD-METAL DEPOSIT DILUTED BY MORE THAN 3% CU
- AN 18-8 TYPE OF WELD-METAL DEPOSIT SUFFICIENTLY DILUTED BY NICKEL OR CHROMIUM TO RESULT IN THE FULLY AUSTENITIC CRACK-SENSITIVE WELD-METAL COMPOSITION (FOR EXAMPLE, 35% NI AND 15% CR)
- ANY NICKEL-COPPER OR COPPER-NICKEL ALLOY WELD-METAL DEPOSIT DILUTED BY MORE THAN 6 TO 8% CR

Manipulation of the welding arc so that it impinges primarily on the base metal nearest in composition to the filler metal helps reduce dilution. For most combinations of dissimilar metals, filler-metal suppliers should be consulted before a filler metal is selected. The addition of ample filler metal to the root pass helps lower the amount of base-metal dilution and also prevents the probability of producing a concave weld deposit profile, instead of the preferred convex shape.

The coefficients of thermal expansion (CTEs) of the base metals must be considered when selecting filler metals for joining dissimilar metal combinations. CTE differences are a major factor in situations where there is temperature cycling (for example, welding of 300 series stainless steel to carbon or low-alloy steel at temperatures above 425 °C, or 800 °F). A large difference in CTE values can induce stresses of sufficient magnitude to produce cracking in the weld metal. Some of the nickel-base electrodes (ENiCr, ENiCrFe, ERNiCr, and ERNiCrFe) and matching filler metals can be used to weld a wide range of base-metal combinations. These electrodes and filler metals can tolerate considerable dilution without loss of strength or ductility. Filler-metal ERNiMo-6 can be used to join nickel-base alloys, cobalt-base alloys, and stainless steels to themselves and to other metals and alloys. The low chromium content of this filler metal, however, should be carefully considered when welding oxidation-resistant materials.

Although GTAW and GMAW processes can be employed, the SMAW process is the most widely used to join nickel and nickel alloys to dissimilar metals. The welding current should be maintained near the middle of the recommended range for the electrode in order to control dilution. Dilution usually can be kept below 25% by manipulating the electrode to dissipate the arc force on already-deposited weld metal. When depositing the first bead, the arc should be directed toward the member from which dilution of the weld metal will be least detrimental. Welding procedures that reduce the amount of penetration, which can cause dilution of the weld metal by the base metal, will increase the probability of securing a good weld. Table 8 gives recommended welding consumables for joining selected nickel-base alloys.

TABLE 8 SELECTION OF WELDING CONSUMABLES FOR NICKEL-BASE ALLOY JOINING APPLICATIONS

WELDING ELETRODE OR FILLER METAL	PRIMARY APPLICATIONS
COATED ELECTRODES	
ENI-1	**NICKEL 200 AND NICKEL 201; THE CLAD SIDE OF NICKEL-CLAD STEEL; JOINING OF STEELS TO NICKEL ALLOYS**
ENICU-7	**ALLOY 400 TO ITSELF, TO LOW-ALLOY AND CARBON STEELS, TO COPPER AND COPPER-NICKEL ALLOYS; SURFACING OF STEELS**
ECUNI	**ALLOY 450; WELDABLE GRADES OF CAST AND WROUGHT 70/30, 80/20, AND 90/10 COPPER-NICKEL ALLOYS**
ENICRFE-1	**ALLOY 600; ALLOY 330**
ENICRFE-3	**ALLOYS 600 AND 601, SURFACING OF STEEL; DISSIMILAR COMBINATIONS OF STEELS AND NICKEL ALLOYS**
ENICRMO-3	**ALLOYS 625 AND 601; PIT-RESISTANT ALLOYS; DISSIMILAR COMBINATIONS OF STEELS AND NICKEL ALLOYS; SURFACING OF STEELS**
ENICRCOMO-1	**ALLOY 617; ALLOY 800HT; DISSIMILAR COMBINATIONS OF HIGH-TEMPERATURE ALLOYS**
ENICRMO-3	**ALLOY 825; ALLOY 686, ALLOY 622, ALLOY C-276; OTHER PIT-RESISTANT ALLOYS; SURFACING OF STEELS**
ENICRMO-3, ENICRMO-9	**ALLOYS G AND G-3; OTHER PIT-RESISTANT ALLOYS; DISSIMILAR COMBINATIONS OF STEELS AND NICKEL ALLOYS; SURFACING OF STEELS**
ENICRFE-3, ENICRFE-7	**ALLOY 600 AND ALLOY 690 TO THEMSELVES AND TO DISSIMILAR COMBINATIONS WITH STEEL; SURFACING OF STEELS**
FILLER METALS	
ERNI-1	**NICKEL 200 AND NICKEL 201; DISSIMILAR COMBINATIONS OF NICKEL ALLOYS AND STEELS; SURFACING OF STEELS**
ERNICU-7	**ALLOYS 400, R-405, AND K-500; SURFACING OF STEEL**
ERCUNI	**ALLOY 450; WELDABLE GRADES OF 70/30, 80/20, AND 90/10 COPPER-NICKEL ALLOYS**
ERNICR3, ERNICRFE-7	**ALLOY 600, ALLOY 690; DISSIMILAR COMBINATION TO STEELS; SURFACING OF STEELS**
ERNICR-3	**ALLOYS 600 AND 601; ALLOYS 800 AND 800HT; ALLOY 330; DISSIMILAR COMBINATIONS OF STEELS AND NICKEL ALLOYS; SURFACING OF STEELS**
ERNICR-3, ERNICRMO-3	**DISSIMILAR COMBINATIONS OF STEELS AND NICKEL ALLOYS**
ERNICRCOMO-1	**ALLOY 617; ALLOY 800HT; DISSIMILAR COMBINATIONS OF HIGH-TEMPERATURE ALLOYS**
ERNICRMO-3	**ALLOYS 625 AND 601; PIT-RESISTANT ALLOYS; DISSIMILAR COMBINATIONS OF STEELS AND NICKEL ALLOYS; SURFACING OF STEELS**
ERNIFECR-2	**ALLOYS 718 AND X-750**
ERNICRMO-3	**ALLOY 825**
ERNICRMO-4 THROUGH -10	**ALLOY 686, ALLOY 622, ALLOY C-276; OTHER PIT-RESISTANT ALLOYS; SURFACING OF STEELS**

Example 1: Elimination of Weld Cracking and Distortion.

The welding of an alloy 400 skirt box to a low-carbon steel (ASTM A 285) cover for a fluorine generator resulted in cracking of the weld metal and warping of the cover (Fig. 5). Cracks resulting from excessive dilution were eliminated by modifying the way in which the beads were deposited. Weld crater cracks were prevented by taking care when ending the weld bead and breaking the arc. The alloy 400 skirt boxes replaced low-carbon steel skirt boxes that had corroded.

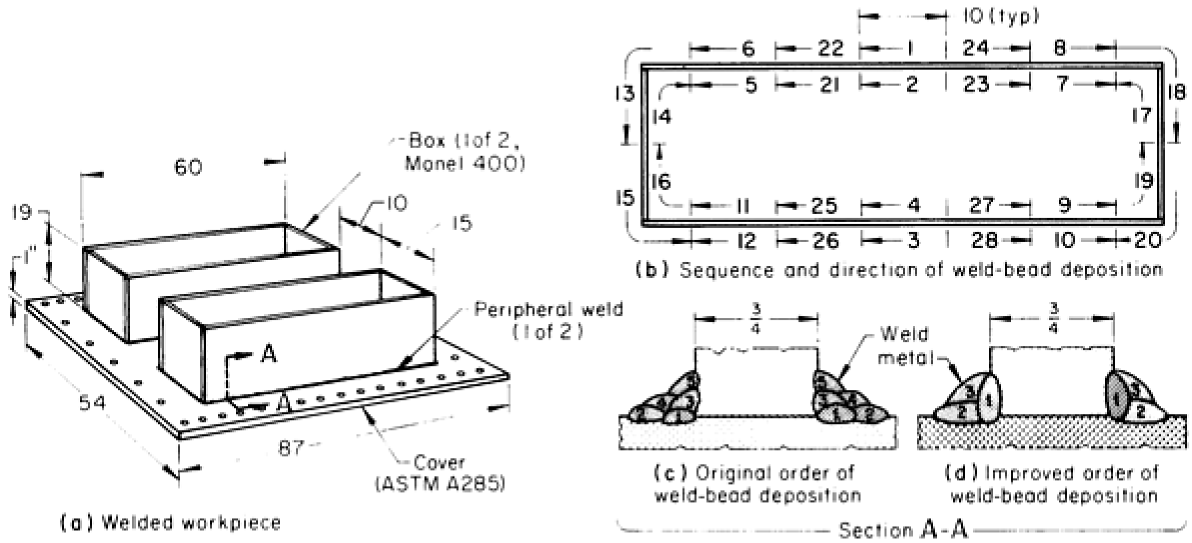
60
Box (1 of 2, Monel 400)
10
15
19
1"
A
A
54
87
Peripheral weld (1 of 2)
Cover (ASTM A285)
10 (typ)
6
22
1
24
8
13
5
21
2
23
7
18
14
17
16
19
15
11
25
4
27
9
12
26
3
28
10
20
3/4
Weld metal
3/4

(a) Welded workpiece

(b) Sequence and direction of weld-bead deposition

(c) Original order of weld-bead deposition

(d) Improved order of weld-bead deposition

Section A-A

PARAMETER	**ALLOY 400 WELDED TO ALLOY 400**		ALLOY 400 WELDED TO LOW-CARBON STEEL	
	ORIGINAL METHOD	**IMPROVED METHOD**	**ORIGINAL METHOD**	IMPROVED METHOD
WELDING PASSES	1	1	5	**3**
FIXTURES	NONE	CLAMPS	NONE	**CLAMPS**
ELECTRODE TYPE[(A)]	4 MM ($\frac{5}{32}$ IN.) DIAM ENICU-4	4 MM ($\frac{5}{32}$ IN.) DIAM ENICU-4	4 MM ($\frac{5}{32}$ IN.) DIAM ENICU-1	**4.8 MM ($\frac{3}{16}$ IN.) DIAM ENICU-1**
WELDING CURRENT	130 A (DCEP)	130 A (DCEP)	154 A (DCEP)	**145 TO 150 A (DCEP)**
LEAK TEST[(B)]	WELDS REJECTED	WELDS ACCEPTED	WELDS REJECTED	**WELDS ACCEPTED**
WELDING TIME PER COVER	40 H	25 H	40 H	**25 H**
WELD PENETRATION:				
BEAD 1	1.6 MM ($\frac{1}{16}$ IN.) INTO STEEL	3.2 MM ($\frac{1}{8}$ IN.) INTO ALLOY 400, 0.4 MM ($\frac{1}{64}$ IN.) INTO STEEL	1.6 MM ($\frac{1}{16}$ IN.) INTO STEEL	**3.2 MM ($\frac{1}{8}$ IN.) INTO ALLOY 400, 0.4 MM ($\frac{1}{64}$ IN.) INTO STEEL**
BEAD 2	1.6 MM ($\frac{1}{16}$ IN.) INTO STEEL	0.8 MM ($\frac{1}{32}$ IN.) INTO STEEL	1.6 MM ($\frac{1}{16}$ IN.) INTO STEEL	**0.8 MM ($\frac{1}{32}$ IN.) INTO STEEL**
BEAD 3	1.6 MM ($\frac{1}{16}$ IN.) INTO ALLOY 400	0.8 MM ($\frac{1}{32}$ IN.) INTO ALLOY 400	1.6 MM ($\frac{1}{16}$ IN.) INTO ALLOY 400	**0.8 MM ($\frac{1}{32}$ IN.) INTO ALLOY 400**
BEAD 4	NONE IN BASE METAL	. . .	NONE IN BASE METAL	**. . .**
BEAD 5	0.8 MM ($\frac{1}{32}$ IN.) INTO ALLOY 400	. . .	0.8 MM ($\frac{1}{32}$ IN.) INTO ALLOY 400	**. . .**
IRON IN WELD METAL:				
BEAD 1	30%	7%	30%	**7%**
BEAD 2	30%	10%	30%	**10%**

BEAD 3	20%	4%	20%	**4%**
BEAD 4 AND 5	**7%**	**. . .**	**7%**	**. . .**

(A) NEITHER ENICU-1 NOR ENICU-4 ELECTRODES ARE CURRENTLY PRODUCED; BOTH HAVE BEEN REPLACED BY ENICU-7 ELECTRODE WHICH HAS 2.5% FE CONTENT.

(B) HELIUM MASS SPECTROMETER WITH INTERNAL VACUUM OF 3.3 PA (0.025 MM HG)

FIG. 5 FLUORINE-GENERATOR COVER MADE BY SMAW OF SIMILAR (ALLOY 400 TO ALLOY 400) AND DISSIMILAR (ALLOY 400 TO LOW-CARBON STEEL) MATERIALS. (ASTM A 285). BOTH WELDS USED NICKEL ALLOY FILLER METAL. A CORNER JOINT AND DOUBLE FILLET WELD WAS UTILIZED TO JOIN THE SIMILAR ALLOYS. A T-JOINT AND DOUBLE FILLET WELD WAS UTILIZED TO JOIN THE DISSIMILAR MATERIALS. PARAMETERS COMMON TO BOTH SETUPS: POWER SUPPLY, 400 A (DC); STRESS RELIEF, 1 H AT 620 ± 14 °C (1150 ± 25 °F)

The original procedure for removing corroded skirt boxes, making new boxes, and welding the new boxes to the cover plate was:

1. REMOVE CORRODED LOW-CARBON STEEL SKIRT BOXES BY AIR-CARBON ARC CUTTING.
2. CLEAN SURFACES OF LOW-CARBON STEEL COVER BY SURFACE GRINDING.
3. CLEAN ALL JOINT AREAS ON THE ALLOY 400 PLATES WITH TRICHLOROETHYLENE AND REMOVE MILL SCALE WITH A SURFACE GRINDER.
4. ASSEMBLE THE ALLOY 400 SKIRT BOX AND WELD THE FOUR COMERS IN ONE PASS EACH WITH A 4.0 MM ($\frac{5}{32}$ IN.) DIAMETER ENICU-4 ELECTRODE, USING A WELDING CURRENT OF 130 A.
5. WELD THE ALLOY 400 SKIRT BOX TO THE LOW-CARBON STEEL COVER WITH INTERMITTENT WELDS, USING THE SEQUENCE OF WELD BEAD DEPOSITION SHOWN IN FIG. 5(B). (THE BOX AND COVER WERE FIRST TACK WELDED AT THE STARTING PLACE FOR EACH OF THE 255 MM, OR 10 IN., LONG WELDS.) DEPOSIT FIVE BEADS AS SHOWN IN FIG. 5(C), WITH A 4.0 MM ($\frac{5}{32}$ IN.) DIAMETER ENICU-1 ELECTRODE, USING A WELDING CURRENT OF 154 A.
6. STRESS RELIEVE AT 620 ± 14 °C (1150 ± 25 °F) FOR 1 H IN A REDUCING ATMOSPHERE. (THE ASSEMBLY WAS PLACED IN THE FURNACE WHEN THE FURNACE TEMPERATURE WAS LESS THAN 315 °C (600 °F), AND, AFTER HEATING, WAS ALLOWED TO COOL TO 315 °C, OR 600 °F, BEFORE IT WAS REMOVED.)
7. TEST THE WELDS WITH A HELIUM MASS SPECTROMETER UNDER AN INTERNAL VACUUM OF 3.3 PA (0.025 MM HG). (WELDS THAT FAILED THE TEST WERE REWELDED.)

During deposition of the first bead, considerable arc blow toward the alloy 400 skirt box was encountered. The ground connection was repositioned several times, but no improvement was obtained. When the welds were tested with a helium mass spectrometer, leakage caused by cracks in the weld beads was detected. An investigation revealed that some cracks had resulted from excessive dilution of the weld metal by iron (analysis showed 30% Fe in bead 1 in Fig. 5b), a result of the high welding heat, and that cracks in the weld craters had been caused by ending beads too abruptly. In addition to the cracking, there was excessive distortion of the cover, where the cover edges were warped upward toward the skirt boxes for 3.2 mm ($\frac{1}{8}$ in.) maximum, along the sides and about 6.4 mm ($\frac{1}{4}$ in.) across the ends. The distortion was not reduced during stress relieving (step 6).

Experimentation resulted in the adoption of a new welding procedure (step 5) that minimized distortion and eliminated weld-metal cracking. The remaining steps in the original procedure were retained, unchanged. In the new procedure, when welding the box to the cover plate, both parts were securely clamped to a welding surface plate. The sequence of weld-bead deposition was the same as for the original method (Fig. 5b), but only firm weld beads were deposited (Fig.

5d). Bead 1 was placed mostly on the alloy 400, which resulted in considerably less dilution by iron (7%, compared with 30%). The bead was ended very slowly to avoid the danger of subsequent crater cracking. The welds were made with an electrode that was slightly larger than the original (4.8 mm, or $\frac{3}{16}$ in., diameter instead of 4.0 mm, or $\frac{5}{32}$ in., diameter), using from 145 to 150 A welding current, instead of 154 A. These modifications reduced the current density and, thus, the heat input.

When the clamps were released, no significant distortion of the cover plate was observed. All joints passed the helium leak-detection test.

Welding of Nickel Alloys

Revised by E.B. Hinshaw, Inco Alloys International, Inc.

References

1. A.C. LINGENFELTER, *WELD. ENG.,* JAN 1970, P 42-45
2. A.C. LINGENFELTER *ET AL., WELD. J.,* MAY 1966, P 417-422

Welding of Nickel Alloys

Revised by E.B. Hinshaw, Inco Alloys International, Inc.

Selected References

- "WELDING," BROCHURE IAI-14-1, INCO ALLOYS INTERNATIONAL, MAY 1989, P 15-36

Welding of Copper Alloys

Revised by M. Ned Rogers, Batesville Casket Company

Introduction

COPPER AND COPPER ALLOYS offer a unique combination of material properties that makes them advantageous for many manufacturing environments. They are widely used because of their excellent electrical and thermal conductivities, outstanding resistance to corrosion, ease of fabrication, and good strength and fatigue resistance. Other useful characteristics include spark resistance, metal-to-metal wear resistance, low-permeability properties, and distinctive color.

Copper and copper alloys find their greatest use for electrical conductors and in the manufacture of electrical components. In its pure form, copper has a face-centered cubic crystal structure and a density of 8.968 g/cm^3 (0.32 lb/in.^{3}), which is about three times the density of aluminum. The electrical conductivity of copper is only slightly lower than that of silver, but it exceeds the electrical and thermal conductivity of aluminum by one-and-a-half times. The electrical conductivity reference standard of engineering materials is copper with a rating of 100% IACS (International Annealed Copper Standard). All other materials are compared on a conductivity basis to the IACS standard. Special processing of copper can produce some forms which reach 102% IACS.

Their excellent resistance to fresh water, salt water, and alkaline solutions makes copper alloys ideally suited for fabrication of tubing, valve fittings, heat exchangers, and chemical equipment. Copper reacts with sulfur and ammonia compounds. Ammonium hydroxide solutions will rapidly attack copper and its alloys, causing severe corrosion.

Many copper alloys are used for bearings and wear surfaces. Their excellent metal-to-metal wear resistance and lack of galling makes them excellent bearing surfaces. The good malleability and formability of copper and its alloys, as well as its aesthetically pleasing color, make it a prime candidate for architectural applications such as ornamentation and roofing.

In manufacturing, copper is often joined by welding. The arc welding processes are of prime concern. Arc welding can be performed using shielded metal arc welding (SMAW), gas-tungsten arc welding (GTAW), gas-metal arc welding (GMAW), plasma arc welding (PAW), and submerged arc welding (SAW). These joining processes will be discussed with reference to specific copper alloy groups. The effects of various alloying agents are also discussed relative to welding metallurgy and process considerations. Examples of standard joint geometry, process parameters, and specific applications are described.

Welding of Copper Alloys

Revised by M. Ned Rogers, Batesville Casket Company

Alloy Metallurgy and Weldability

Many common metals are alloyed with copper to produce the various copper alloys. The most common alloying elements are aluminum, nickel, silicon, tin, and zinc. Other elements and metals are alloyed in small quantities to improve certain material characteristics, such as corrosion resistance or machinability. Copper and its alloys are divided into nine major groups. These major groups are:

- *COPPERS,* WHICH CONTAIN A MINIMUM OF 99.3% CU
- *HIGH-COPPER ALLOYS,* WHICH CONTAIN UP TO 5% ALLOYING ELEMENTS
- *COPPER-ZINC ALLOYS (BRASSES),* WHICH CONTAIN UP TO 40% ZN
- *COPPER-TIN ALLOYS (PHOSPHOR BRONZES),* WHICH CONTAIN UP TO 10% SN AND 0.2% P
- *COPPER-ALUMINUM ALLOYS (ALUMINUM BRONZES),* WHICH CONTAIN UP TO 10% AL
- *COPPER-SILICON ALLOYS (SILICON BRONZES),* WHICH CONTAIN UP TO 3% SI
- *COPPER-NICKEL ALLOYS,* WHICH CONTAIN UP TO 30% NI
- *COPPER-ZINC-NICKEL ALLOYS (NICKEL SILVERS),* WHICH CONTAIN UP TO 27% ZN AND 18% NI
- *SPECIAL ALLOYS,* WHICH CONTAIN ALLOYING ELEMENTS TO ENHANCE A SPECIFIC PROPERTY OR CHARACTERISTIC, FOR EXAMPLE, MACHINABILITY

Many copper alloys have common names, such as oxygen-free copper (99.95% Cu min), beryllium copper (0.2 to 2.0% Be), Muntz metal (Cu-40Zn), Naval brass (Cu-39.25Zn-0.75Sn), and commercial bronze (Cu-10Zn).

A more standardized system of identification is the Unified Numbering System (UNS). In this system, wrought alloys of copper are designated by numbers 1*xxxx* to 7*xxxx*, and cast alloys are designated 8*xxxx* to 9*xxxx*; thus, the same alloy can be produced as a wrought and cast product.

Table 1 lists various copper alloys that are frequently arc welded, their UNS numbers, and their physical properties. Many of the physical properties of copper alloys are important to the welding processes, including melting temperature, coefficient of thermal expansion, and electrical and thermal conductivity. As shown in Table 1, certain alloying elements greatly decrease the electrical and thermal conductivities of copper and copper alloys. This will in turn significantly affect the weldability of the alloys. The copper and copper alloys listed in Table 1 are rated for their relative weldability using SMAW, GTAW, and GMAW.

TABLE 1 NOMINAL COMPOSITIONS, MELTING POINTS, RELATIVE THERMAL CONDUCTIVITIES, AND WELDABILITIES OF WROUGHT COPPERS AND COPPER ALLOYS THAT ARE COMMONLY ARC

WELDED

UNS NO.	**ALLOY NAME**	**NOMINAL COMPOSITION, %**	**MELTING POINT (LIQUIDUS)**		**RELATIVE THERMAL CONDUCTIVITY**[(A)]	WELDABILITY[(B)] BY		
			°C	**°F**		**GTAW**	**GMAW**	SMAW
OFC AND ETP COPPERS								
C10200	OXYGEN-FREE COPPER (OFC)	99.95 CU	1083	1981	100	G	G	**NR**
C11000	ELECTROLYTIC TOUGH PITCH (ETP) COPPER	99.90 CU, 0.04 O_2	1083	1981	100	F	F	**NR**
DEOXIDIZED COPPERS								
C12000	PHOSPHORUS-DEOXIDIZED COPPER, LOW-P	99.9 CU, 0.008 P	1083	1981	99	E	E	**NR**
C12200	PHOSPHORUS-DEOXIDIZED COPPER, HIGH-P	99.9 CU, 0.02 P	1083	1981	87	E	E	**NR**
BERYLLIUM COPPERS								
C17000	HIGH-STRENGTH BERYLLIUM COPPER	98.3 CU, 1.7 BE	982	1800	27-33[(C)]	G	G	**G**
C17200	HIGH-STRENGTH BERYLLIUM COPPER	98.1 CU, 1.9 BE	982	1800	27-33[(C)]	G	G	**G**
C17500	HIGH-CONDUCTIVITY BERYLLIUM COPPER	96.9 CU, 0.6 BE, 2.5 CO	1068	1955	53-66[(C)]	F	F	**F**
LOW-ZINC BRASSES								
C21000	GILDING	95 CU, 5 ZN	1065	1950	60	G	G	**NR**
C22000	COMMERCIAL BRONZE	90 CU, 10 ZN	1043	1910	48	G	G	**NR**
C23000	RED BRASS	85 CU, 15 ZN	1026	1880	41	G	G	**NR**
C24000	LOW BRASS	80 CU, 20 ZN	999	1830	36	G	G	**NR**
HIGH-ZINC BRASSES								
C26000	CARTRIDGE BRASS	70 CU, 30 ZN	954	1750	31	F	F	**NR**
C26800	YELLOW BRASS	65 CU, 35 ZN	932	1710	30	F	F	**NR**
C28000	MUNTZ METAL	60 CU, 40 ZN	904	1660	31	F	F	**NR**
TIN BRASSES								
C44300	ADMIRALTY BRASS	71 CU, 28 ZN, 1 SN[(D)]	937	1720	28	F	F	**NR**
C46400	NAVAL BRASS	60 CU, 39.25 ZN, 0.75 SN[(D)]	899	1650	30	F	F	**NR**
SPECIAL BRASSES								
C67500	MANGANESE BRONZE A	58.5 CU, 39 ZN, 1.4 FE, 1 SN, 0.1 MN	888	1630	27	F	F	**NR**
C68700	ALUMINUM BRASS, ARSENICAL	77.5 CU, 20.5 ZN, 2 AL (0.06 AS)	971	1780	26	F	F	**NR**
NICKEL SILVERS								
C74500	NICKEL SILVER	65 CU, 25 ZN, 10 NI	1021	1870	12	F	F	**NR**
C75200	NICKEL SILVER	65 CU, 17 ZN, 18 NI	1110	2030	8	F	F	**NR**
C75400	NICKEL SILVER	65 CU, 20 ZN, 15 NI	1076	1970	9	F	F	**NR**
C75700	NICKEL SILVER	65 CU, 23 ZN, 12 NI	1037	1900	10	F	F	**NR**
C77000	NICKEL SILVER	55 CU, 27 ZN, 18	1054	1930	8	F	F	**NR**

		NI						
PHOSPHOR BRONZES								
C50500	PHOSPHOR BRONZE, 1.25% E	98.7 CU, 1.3 SN (0.2 P)	1076	1970	53	G	G	**F**
C51000	PHOSPHOR BRONZE, 5% A	95 CU, 5 SN (0.2 P)	1049	1920	18	G	G	**F**
C52100	PHOSPHOR BRONZE, 8% C	92 CU, 8 SN (0.2 P)	1026	1880	16	G	G	**F**
C52400	PHOSPHOR BRONZE, 10% D	90 CU, 10 SN (0.2 P)	999	1830	13	G	G	**F**
ALUMINUM BRONZES								
C61300	ALUMINUM BRONZE D, SN-STABILIZED	89 CU, 7AL, 3.5 FE (0.35 SN)	1046	1915	14	G	E	**G**
C61400	ALUMINUM BRONZE D	91 CU, 6-8 AL, 1.5-3.5 FE, 1 MAX MN	1046	1915	17	G	E	**G**
C63000	ALUMINUM BRONZE E	82 CU, 10 AL, 5 NI, 3 FE	1054	1930	10	G	G	**G**
SILICON BRONZES								
C65100	LOW-SILICON BRONZE B	98.5 CU, 1.5 SI	1060	1940	15	E	E	**F**
C65500	HIGH-SILICON BRONZE A	97 CU, 3 SI	1026	1880	9	E	E	**F**
COPPER-NICKEL								
C70600	COPPER NICKEL	88.6 CU, 9-11 NI, 1.4 FE, 1.0 MN	1149	2100	12	E	E	**G**
C71500	**COPPER NICKEL**	**70 CU, 30 NI**	**1238**	**2260**	**8**	**E**	**E**	**E**

(A) BASED ON THE THERMAL CONDUCTIVITY OF ALLOY C10200 (2553.8 KJ/M^2 IN METERS PER HOUR AT 20 °C) AS 100. FOR COMPARISON, CARBON STEEL HAS A THERMAL CONDUCTIVITY OF 339 KJ/M^2 IN METERS PER HOUR AT 20 °C WHICH IS 13 ON THIS SCALE.

(B) E, EXCELLENT; G, GOOD; F, FAIR; NR, NOT RECOMMENDED.

(C) IN THE PRECIPITATION-HARDENED CONDITION.

(D) ALLOYS C44300 AND C46400 CONTAIN NOMINAL 0.06% AS; ALLOYS C44400 AND C46600, A NOMINAL 0.06% SB; ALLOYS C44500 AND 46700, A NOMINAL 0.06% P.

Several alloying elements have pronounced effects on the weldability of copper and copper alloys. Small amounts of volatile, toxic alloying elements are often present in copper and its alloys. As a result, the requirement of an effective ventilation system to protect the welder and/or the welding machine operator is more critical than when welding ferrous metals.

Zinc reduces the weldability of all brasses in relative proportion to the percent of zinc in the alloy. Zinc has a low boiling temperature, which results in the production of toxic vapors when welding copper-zinc alloys. Effective forced ventilation is mandatory, and source-capture systems are necessary to contend with the fumes.

Tin increases the hot-crack susceptibility during welding when present in amounts from 1 to 10%. These alloys are typically the phosphor bronzes and tin brasses. Tin, when compared with zinc, is far less volatile and toxic. During welding, tin may preferentially oxidize relative to copper. The results will be an oxide entrapment, which may reduce the strength of the weldment.

Beryllium, aluminum, and nickel form tenacious oxides that must be removed prior to welding. The formation of these oxides during the welding process must be prevented by shielding gas or by fluxing, in conjunction with the use of the appropriate welding current. The oxides of nickel interfere with arc welding less than those of beryllium or aluminum. Consequently, the nickel silvers and copper-nickel alloys are less sensitive to the type of welding current used during manufacture. Beryllium-containing alloys also produce toxic fumes during welding.

Silicon has a beneficial effect on the weldability of copper-silicon alloys because of its deoxidizing and fluxing actions. The combination of this effect and the low thermal conductivity makes silicon bronzes the most weldable of the copper alloys for any arc process.

Phosphorus is beneficial to certain copper and copper alloys as a strengthener and deoxidizer. When added to brass, phosphorus inhibits dezincification corrosion. In the amounts normally present in most copper alloys, phosphorus does not adversely affect or hinder welding.

Chromium, like beryllium and aluminum, can form a refractory oxide on the surface of the molten weld pool. Arc welding should be done in an inert protective atmosphere to prevent formation of chromium oxides.

Cadmium has no serious effect on the weldability of copper. However, the low boiling temperature of cadmium does result in evaporation of this alloy at welding temperatures, thereby creating a potential health hazard. Cadmium will also form an oxide in the molten weld metal, but the oxide can be easily reduced by fluxing agents.

Oxygen can cause porosity and reduce the strength of welds made in certain copper alloys that do not contain sufficient quantities of phosphorus or other deoxidizers. Oxygen may be found as a free gas or as cuprous oxide. Most commonly welded copper alloys contain deoxidizing elements--usually phosphorus, silicon, aluminum, iron, or manganese. These elements will readily combine with oxygen and eliminate the potential for porosity. The same deoxidizers are also included in the filler metals. The soundness and strength of arc welds made in commercial coppers depend upon the cuprous oxide content. As the oxide content decreases, weld soundness increases. Deoxidized coppers provide the best results because they are free from cuprous oxides and contain residual amounts of phosphorus.

Iron and manganese do not significantly affect the weldability of the alloys that contain them. Iron is typically present in some special brasses, aluminum bronzes, and copper-nickel alloys in amounts of 1.4 to 3.5%. Manganese is commonly used in these same alloys, but at lower concentrations than iron.

Free-Machining Additives. Lead, selenium, tellurium, and sulfur are added to copper alloys to improve machinability. Bismuth is beginning to be used for this purpose as well when lead-free alloys are desired. These minor alloying agents, while improving machinability, adversely affect the weldability of copper alloys by rendering the alloys hot-crack susceptible. The adverse effect on weldability begins to be evident at about 0.05% of the additive and is more severe with larger concentrations. Lead is the most harmful of the alloying agents with respect to hot-crack susceptibility. However, alloys that contain 0.5 to 4% Pb are not ordinarily welded and are not included in Table 1.

Welding of Copper Alloys

Revised by M. Ned Rogers, Batesville Casket Company

Factors Affecting Weldability

Besides the alloying elements that comprise a specific copper alloy, several other factors affect weldability. These factors are the thermal conductivity of the alloy being welded, the shielding gas, the type of current used during welding, the joint design, the welding position, and the surface condition and cleanliness. The effects of the shielding gas, type of current, and joint design used during welding are discussed under the sections for each individual process and alloy, described later in this article.

Effect of Thermal Conductivity. The behavior of copper and copper alloys during welding is strongly influenced by the thermal conductivity of the alloy. Table 1 shows the relative thermal conductivities based on the conductivity of alloy C10200 (oxygen-free copper). As shown in Table 1, the thermal conductivity varies greatly from this baseline value, which is 339 kJ/m^2 in meters per hour at 20 °C (226 Btu/ft^2 in feet per hour at 68 °F). The range is from 100, for alloys C10200 and C11000, to lows of 8 to 12 for nickel silvers and copper-nickel alloys and 9 for alloy C65500. In comparison, carbon steel has a thermal conductivity of 13 on this scale.

When welding commercial coppers and lightly alloyed copper materials with high thermal conductivities, the type of current and shielding gas must be selected to provide maximum heat input to the joint. This high heat input counteracts the rapid head dissipation away from the localized weld zone. Depending on section thickness, preheating may be required for copper alloys with lower thermal conductivities. The interpass temperature should be the same as for preheating. Copper alloys are not postweld head treated as frequently as steels, but some alloys may require controlled cooling rates to minimize residual stresses and hot shortness.

Welding Position. Due to the highly fluid nature of copper and its alloys, the flat position is used whenever possible for welding. The horizontal position is used in some fillet welding of corner joints and T-joints.

Vertical and overhead positions and the horizontal position are less frequently used in welding butt joints. These positions are ordinarily restricted to GTAW, GMAW, and PAW of the less conductive aluminum bronzes, silicon bronzes, and copper-nickel alloys. Small-diameter electrodes and filler-metal wires are used in conjunction with low welding currents for out-of-position welding. Current pulsation is used with GTAW, GMAW, and PAW, to control weld pool fluidity.

When using SMAW, out-of-position welding is usually limited to the joining of aluminum bronzes and copper-nickel alloys, but it can also be done on some phosphor bronzes and silicon bronzes.

Precipitation-Hardenable Alloys. The most important precipitation-hardening reactions are obtained with beryllium, chromium, boron, nickel, silicon, and zirconium. Care must be taken when welding precipitation-hardenable copper alloys to avoid oxidation and incomplete fusion. Whenever possible, the components should be welded in the annealed condition, and then the weldment should be given a precipitation-hardening heat treatment.

Hot Cracking. Copper alloys, such as copper-tin and copper-nickel, are susceptible to hot cracking at solidification temperatures. This characteristic is exhibited in all copper alloys with a wide liquidus-to-solidus temperature range. Severe shrinkage stresses produce interdendritic separation during metal solidification. Hot cracking can be minimized by reducing restraint during welding, preheating to slow the cooling rate and reduce the magnitude of welding stresses, and reducing the size of the root opening and increasing the size of the root pass.

Porosity. Certain elements (for example, zinc, cadmium, and phosphorus) have low boiling points. Vaporization of these elements during welding may result in porosity. When welding copper alloys containing these elements, porosity can be minimized by fast weld speeds and a filler metal low in these elements.

Surface Condition. Grease and oxide on work surfaces should be removed before welding. Wire brushing or bright dipping can be used. Millscale on the surfaces of aluminum bronzes and silicon bronzes is removed for a distance from the weld region of at least 13 mm (0.5 in.), usually by mechanical means. Grease, paint, crayon marks, shop dirt, and similar contaminants on copper-nickel alloys may cause embrittlement and should be removed before welding. Millscale on copper-nickel alloys must be removed by grinding or pickling; wire brushing is not effective.

Welding of Copper Alloys

Revised by M. Ned Rogers, Batesville Casket Company

Arc Welding Processes

Copper and most copper alloys can be joined by arc welding. The commercial alloys that are readily arc welded are listed in Table 1. Welding processes that use gas shielding are generally preferred, although SMAW can be used for many noncritical applications.

Argon, helium, or mixtures of the two are used as shielding gases for GTAW, PAW, and GMAW. Generally, argon is used when manually welding material that is less than 3 mm (0.13 in.) thick, has low thermal conductivity, or both. Helium or a mixture of 75% helium and 25% argon is recommended for machine welding of thin sections and for manual welding of thicker sections of alloys that have high thermal conductivity. Small amounts of nitrogen can be added to the argon shielding gas to increase the effective heat input.

Shielded metal arc welding can be used to weld a wide range of thicknesses of copper alloys. Covered electrodes for SMAW of copper alloys are available in standard sizes ranging from 2.4 to 4.8 mm ($\frac{3}{32}$ to $\frac{3}{16}$ in.). Other sizes are available in certain electrode classifications, and the reader is referred to ANSI/AWS A5.6 "Specification for Covered Copper and Copper Alloy Arc Welding Electrodes."

Higher thermal conductivity and thermal expansion of copper and copper alloys result in greater weld distortion than in comparable steel welds. The use of preheat, fixtures, proper welding sequence, and tack welds can minimize distortion or warping.

Gas-Tungsten Arc Welding

Gas-tungsten arc welding is well suited for copper and copper alloys because of its intense arc, which produces an extremely high temperature at the joint and a narrow heat-affected zone (HAZ). In welding copper and the more thermally conductive copper alloys, the intensity of the arc is important in completing fusion with minimum heating of the surrounding, highly conductive base metal. A narrow HAZ is particularly desirable in the welding of copper alloys that have been precipitation-hardened.

The most frequent use of GTAW for copper and copper alloys is on sections up to 3 mm (0.13 in.) thick that have been prepared with a square edge. Often no filler metal is used in joining these thicknesses. Filler metal is usually required in GTAW of sections thicker than 3 mm (0.13 in.). Gas-tungsten arc welding of sections thicker than 13 mm (0.5 in.) is performed only if GMAW equipment is not available or if special conditions exist. These conditions would include hot shortness of the base metal or adjacent heat-sensitive features that make it necessary to limit the heat input to the base material. Under certain conditions, pulsed GMAW would be the preferred process.

Type of Current. Direct current electrode negative (DCEN) is used for GTAW of most copper and copper alloys. This permits the use of an electrode that has a minimum diameter for a given welding current and that provides maximum penetration of the base material. Alternating current stabilized by high frequency is used on beryllium coppers and aluminum bronzes to prevent the buildup of tenacious oxide film on the base metals.

Electrodes. Many of the standard tungsten or alloyed tungsten electrodes can be used in GTAW of copper and copper alloys. The selection factors normally considered for tungsten electrodes apply in general to the copper and copper alloys. Except as noted for the specific classes of copper alloys, thoriated tungsten (usually EWTh-2) is preferred for its better performance, longer life, and greater resistance to contamination.

Gas-Metal Arc Welding

Gas-metal arc welding is used to join all of the coppers and copper alloys listed previously in Table 1. Gas-tungsten arc welding is preferred for thicknesses less than 3 mm (0.13 in.), while GMAW is preferred for section thicknesses above 3 mm (0.13 in.) and for the joining of aluminum bronzes, silicon bronzes, and copper-nickel alloys.

The major application of GMAW of copper alloys is in the joining of material thicknesses between 3 and 13 mm (0.13 and 0.5 in.) thick, and the process is almost invariably selected for arc welding sections of copper alloys thicker than 13 mm (0.5 in.). The high deposition rate for GMAW is a major advantage in this application when compared to GTAW or SMAW. The greater heat input to the weld, compared with that for GTAW, is a disadvantage in some applications due to the formation of a wider HAZ.

Direct current electrode positive (DCEP) is used exclusively for GMAW of copper alloys. Typically, argon is used as a shielding gas. Helium or mixtures of argon and helium are often used when hotter arc voltages are needed with standard operating current levels. A square-groove joint is not ordinarily used for welding thicknesses greater than 3 mm (0.13 in.), except for coppers. Single-V-grooves are used for thicknesses of 3 to 13 mm (0.13 to 0.5 in.). When section thicknesses exceed 13 mm (0.5 in.), double-V-grooves or double-U-grooves are used as joint preparations.

Welding Position. Most GMAW of copper alloys is done in the flat position with spray transfer; fillet welds acceptable for many applications can be produced in a horizontal position. Gas-metal arc welding is preferred to GTAW or SMAW when weld positions other than flat are required. Out-of-position welding is usually restricted to the less-fluid copper alloys, such as aluminum bronzes, silicon bronzes, and copper-nickel alloys. Small-diameter filler metals and low welding currents are preferred for such applications, and a globular or short-circuiting transfer is ordinarily used. The development of pulsed spray transfer has been quite advantageous for welding of coppers and copper alloys. Small- to medium-diameter wires can be readily used out-of-position for vertical and overhead welding of most alloys.

Shielded Metal Arc Welding

When compared with SMAW of low-carbon steels, SMAW of copper and copper alloys uses larger root openings, wider groove angles, more tack welds, higher preheat and interpass temperatures, and higher welding currents. Welding of coppers and copper alloys using SMAW is almost always restricted to flat-position welding. Out-of-position welding using this process is usually limited to the joining of phosphor bronzes and copper-nickel alloys.

Plasma Arc Welding

The welding of coppers and copper alloys using PAW is comparable to GTAW of these alloys. Argon, helium, or mixtures of the two are used for the welding of all alloys. Hydrogen gas should never be used when welding coppers. Plasma arc welding has two distinct advantages over GTAW: (1) the tungsten is concealed and entirely shielded, which greatly reduces contamination of the electrode, particularly for alloys with low-boiling-temperature constituents such as brasses, bronzes, phosphor bronzes, and aluminum bronzes; and (2) the constricted arc plume gives rise to higher arc energies while minimizing the growth of the HAZ. As with GTAW, current pulsation and current ramping may also be used. Plasma arc welding equipment has been miniaturized for intricate work, known as microplasma welding.

Plasma arc welding of coppers and copper alloys may be performed either autogenously or with filler metal. Filler metal selection is identical to that outlined for GTAW. Automation and mechanization of this process is readily performed and is preferable to GTAW where contamination can restrict production efficiencies. Welding positions for PAW are identical to those for GTAW. However, the plasma keyhole mode has been evaluated for thicker sections in a vertical-up position. Generally, all information presented for GTAW is applicable to PAW.

Submerged Arc Welding

The welding of thick gage material, such as seamed pipe formed from heavy plate, can be achieved by continuous metal-arc operation under a granular flux. Effective deoxidation and slag-metal reactions to form the required weld-metal composition are critical, and the SAW process is still under development for copper-base materials. A variation on this process can be used for weld cladding or hardfacing. Promising results have been seen with the copper-nickel alloys for section thicknesses greater than 13 mm (0.5 in.). V-groove and U-groove joint designs similar to those used in GMAW are satisfactory.

Commercially available fluxes should be used for the copper-nickel alloys. Welding conditions, which are greatly dependent upon the flux used, are provided by the flux manufacturer. Special attention should be given to the weld bead sequence when multi-pass welds are deposited, to ensure complete fusion while maintaining proper bead contour. X-ray quality results can be obtained when the technique is correctly performed.

Welding of Copper Alloys

Revised by M. Ned Rogers, Batesville Casket Company

Filler Metals

Both covered electrodes and bare electrode wire and rods are available for welding copper and copper alloys. Tables 2, 3, and 4 list the various American Welding Society (AWS) classifications for the filler metals for GTAW, GMAW, and SMAW, respectively.

TABLE 2 FILLER METALS FOR GTAW OF COPPER AND COPPER ALLOYS

FILLER METAL	**AWS CLASSIFICATION**	PRINCIPLE CONSTITUENTS[(B)]
COPPER	ERCU	**98.0 MIN CU + AG, 1.0 SN, 0.5 MU, 0.50 SI, 0.15 P**
PHOSPHOR BRONZE	ERCUSN-A	**93.5 MIN CU + AG, 4.0-6.0 SN, 0.10-0.35 P**
ALUMINUM BRONZE	ERCUAL-A2	**1.5 FE, 9.0-11.0 AL, REM CU + AG**
ALUMINUM BRONZE	ERCUAL-A3	**3.0-4.25 FE, 11.0-12.0 AL, REM CU + AG**
SILICON BRONZE	ERCUSI-A	**94.0 MIN CU + AG, 2.8-4.0 SI, 1.5 ZN, 1.5 SN, 1.5 MN, 0.5 FE**
COPPER-NICKEL	**ERCUNI**	**1.00 MN, 0.40-0.70 FE, 29.0-32.0 NI + CO, 0.20-0.50 TI, REM CU + AG**

(A) NOTE: THESE DATA ARE BASED ON AWS A5.27, A5.6, AND A5.7; SEE CURRENT EDITIONS OF THOSE SPECIFICATIONS FOR COMPLETE COMPOSITIONS AND QUALIFICATIONS.

(B) SINGLE PERCENTAGES ARE MAXIMUMS UNLESS OTHERWISE STATED. OPTIONAL ELEMENTS AND IMPURITIES HAVE BEEN OMITTED.

TABLE 3 FILLER METALS FOR GMAW OF COPPER AND COPPER ALLOYS

ELECTRODE DESIGNATION(A)	**COMMON NAME**	BASE-METAL APPLICATIONS
ERCU	COPPER	**COPPERS**
ERCUSI-A	SILICON BRONZE	**SILICON BRONZES, BRASSES**
ERCUSN-A	PHOSPHOR BRONZE	**PHOSPHOR BRONZES, BRASSES**
ERCUNI	COPPER-NICKEL	**COPPER-NICKEL ALLOYS**
ERCUAL-A2	ALUMINUM BRONZE	**ALUMINUM BRONZES, BRASSES, SILICON BRONZES, MANGANESE BRONZES**
ERCUAL-A3	ALUMINUM BRONZE	**ALUMINUM BRONZES**
ERCUNIAL	. . .	**NICKEL-ALUMINUM BRONZES**
ERCUMNNIAL	. . .	**MANGANESE-NICKEL-ALUMINUM BRONZES**
RBCUZN-A	NAVAL BRASS	**BRASSES, COPPER**
RCUZN-B	LOW-FUMING BRASS	**BRASSES, MANGANESE BRONZES**
RCUZN-C	**LOW-FUMING BRASS**	**BRASSES, MANGANESE BRONZES**

(A) SEE THE MOST RECENT EDITION OF AWS A5.7, "SPECIFICATION FOR COPPER AND COPPER ALLOY BARE WELDING RODS AND ELECTRODES," OR AWS A5.27, "SPECIFICATION FOR COPPER AND COPPER ALLOY GAS WELDING RODS."

TABLE 4 FILLER METAL FOR SMAW OF COPPER AND COPPER ALLOYS

COVERED ELECTRODE(A)	**COMMON NAME**	BASE-METAL APPLICATIONS
ECU	COPPER	**COPPERS**
ECUSI	SILICON BRONZE	**SILICON BRONZES, BRASSES**
ECUSN-A, ECUSN-C	PHOSPHOR BRONZE	**PHOSPHOR BRONZES, BRASSES**
ECUNI	COPPER-NICKEL	**COPPER-NICKEL ALLOYS**
ECUAL-A2	ALUMINUM BRONZE	**ALUMINUM BRONZES, BRASSES, SILICON BRONZES, MANGANESE BRONZES**
ECUAL-B	ALUMINUM BRONZE	**ALUMINUM BRONZES**
ECUNIAL	. . .	**NICKEL-ALUMINUM BRONZES**
ECUMNNIAL	**. . .**	**MANGANESE-NICKEL-ALUMINUM BRONZES**

(A) SEE THE MOST RECENT EDITION OF AWS A5.6, "SPECIFICATION FOR COVERED COPPER AND COPPER ALLOY ARC WELDING COVERED ELECTRODES."

Copper Filler Metals. The electrodes and rods used to weld the coppers, primarily deoxidized and electrolytic tough pitch (ETP) coppers, are designated as ERCu and have a minimum copper content of 98%. These filler metals, which can be used with GMAW, GTAW, and PAW processes, have electrical conductivity ratings of 25 to 40% IACS.

The covered electrodes used for SMAW are designated as ECu. These are normally used with DCEP. Compared to carbon steel electrodes of the same diameter, the required welding current for ECu electrodes is typically 30 to 40% greater.

Copper-Zinc Filler Metals. Copper-zinc (brass) weld rods are available in three classifications: RBCuZn-A (Naval Brass), RCuZn-B (Low-Fuming Brass), and RCuZn-C (Low-Fuming Brass). These particular welding rods are used for braze welding of copper, bronze, and nickel alloys. Their electrical conductivity is about 25% IACS; thermal-conductivity values are about 30% of those of copper.

Copper-zinc filler metals cannot be used as electrodes for the arc welding processes. The high zinc content tends to become volatile during arc welding and boils from the molten weld pool. This boiling of the low-melting-temperature constituents results in a porous weld.

Copper-Tin Filler Metals. The copper-tin (phosphor bronze) welding electrodes and rods are designated as ECuSn-A, ERCuSn-A, and ECuSn-C. The ECuSn-A and ECuSn-C electrodes contain approximately 5 and 8% Sn, respectively. Phosphorus is used as a deoxidizer for both electrodes. These electrodes are suitable for welding bronze, brass, and copper if the presence of tin in the weld metal is not objectionable. These filler metals are also commonly used for repair welding of castings. Greater weld-metal strength and hardness are obtained with the ECuSn-C electrodes than with the ECuSn-A electrodes.

The ERCuSn-A rods can be used with GTAW and PAW for joining phosphor bronzes. Preheat temperatures of about 200 °C (400 °F) are required when welding with these electrodes, particularly for heavy sections.

Copper-Silicon Filler Metals. Copper-silicon (silicon bronze) electrodes are used in barewire form for GMAW, GTAW, PAW, and sometimes for oxyfuel welding. The designation for these electrodes is ERCuSi-A. Chemical compositions contain between 2.8 and 4.0% Si with approximately 1.5% Mn, 1.0% Sn, and 1.0% Zn. These filler metals are frequently used to weld silicon bronzes and brasses. They are also used for braze welding of galvanized steel. The electrical conductivity is about 6.5% IACS, and the thermal conductivity is approximately 8.4% of that of copper. The covered electrodes in this filler metal classification are designated as ECuSi and are used primarily for welding copper-zinc alloys with the DCEP setting. The covered electrode can also be used for welding of silicon bronze, copper, and galvanized steel. Weld-metal mechanical properties are slightly higher than those of silicon bronze base metal.

Copper-Aluminum Filler Metals. There are a number of copper-aluminum (aluminum) bronze filler metals for both hardfacing and joining applications. The first of these is designated as ERCuAI-A and is an iron-free aluminum bronze. It is used as a surfacing alloy for wear-resistant surfaces with relatively light loads and for resistance to corrosive media, such as salt water and some commonly used acids. This alloy is not recommended for joining applications. Another hardfacing filler metal is ECuAl-B, which is a covered electrode containing 7.5 to 10% AI. These electrodes are used for surfacing applications and repair welding of aluminum bronze castings with similar compositions.

The covered electrodes for SMAW are designated as ECuAl-A2 and contain from 6.5 to 9% A1. The bare-wire electrodes for GTAW, GMAW, and PAW contain from 8.5 to 11% Al and are designated as ERCuAl-A2. This electrode has a higher strength than the covered electrode. Both filler metals can be used for joining aluminum bronzes, silicon bronzes, copper-nickel alloys, copper-zinc alloys, manganese bronzes, and many combinations of dissimilar metals.

ERCuAl-A3 electrodes and rods are used for repair welding of aluminum bronze castings using GMAW and GTAW. The high aluminum content produces welds with less tendency to crack in highly stressed cross sections.

Repair of both wrought and cast aluminum bronze materials containing nickel can be performed with copper-nickel-aluminum electrodes and bare wire. The designations for these metals are ECuNiA1 and ERCuNiA1. These electrodes offer good corrosion resistance as well as good cavitation resistance in both salt water and brackish water.

Joining of manganese-nickel-aluminum bronzes is performed with covered electrodes, designated as ECuMnNiAl, and bare filler metals, designated as ERCuMnNiAl. Like the copper-nickel-aluminum electrodes, these electrodes have good resistance to cavitation, erosion, and corrosion.

Copper-Nickel Filler Metals. Covered electrodes for this classification are designated as ECuNi, and bare electrode wire and rods are designated as ERCuNi. The nominal compositions are 70% Cu and 30% Ni. Titanium is used as a deoxidizer in these filler metals, which are used for the joining of copper-nickel alloys.

Welding of Copper Alloys

Revised by M. Ned Rogers, Batesville Casket Company

Welding of Coppers

As indicated in Table 1, coppers include oxygen-free, ETP, and deoxidized grades that contain [ges]99% Cu (UNS Numbers C10100 to C 15760). The coppers are welded using GTAW, GMAW, and SMAW. Additional information on the properties and applications of coppers can be found in Volume 2 of the *ASM Handbook.*

Gas-Tungsten Arc Welding

When using GTAW to join commercial coppers, weld quality will differ depending on the cuprous oxide content of the copper. The nominal welding conditions, however, for a given thickness and joint design are approximately the same. Table 5 lists several representative welding conditions for GTAW of commercial coppers.

TABLE 5 NOMINAL CONDITIONS FOR GTAW OF COMMERCIAL COPPERS USING EWTH-2 ELECTRODES, ERCU WELDING ROD, AND DCEN

WORKPIECE THICKNESS		**ROOT OPENING**(A)		**ELECTRODE DIAMETER**		**WELDING ROD DIAMETER**		**SHIELDING GAS**(B)	**GAS FLOW RATE**		**CURRENT, A**	**TRAVEL SPEED**		**PREHEAT**		NUMBER OF PASSES
MM	**IN.**	**MM**	**IN.**	**MM**	**IN.**	**MM**	**IN.**		**L/MIN**	**FT^3/H**		**M/MIN**	**IN./MIN**	**°C**	**°F**	
BUTT JOINTS--SQUARE GROOVE																
1.6	1/6	0	0	1.6	1/6	NONE USED		ARGON	7.1	15	110-114	0.30	12	NONE USED		**1**
3.2	1/8	0	0	2.4	3/32	NONE USED		ARGON	7.1	15	175-225	0.28	11	NONE USED		**1**
3.2	1/8	3.2	1/8	2.4	3/32	2.4	3/32	ARGON	7.1	15	175-225	0.28	11	NONE USED		**1**
4.8	3/16	4.8	3/16	3.2	1/8	3.2	1/8	HELIUM	15.0	30	190-225	0.25	10	95	200	**1**
BUTT JOINTS--60° SINGLE-V-GROOVE, 1.6 MM (1/16 IN.) ROOT FACE																
6.4	1/4	1.6	1/16	3.2	1/8	3.2	1/8	HELIUM	15.0	30	225-260	0.23	9	150	300	**1**
9.5	3/8	1.6	1/16	4.8	3/16	4.8	3/16	HELIUM	18.9	40	280-320	. . .	. . .	260	500	**2**
BUTT JOINTS--60° DOUBLE-V-GROOVE, 3.2 MM (1/8 IN.) ROOT FACE(C)																
12.7	1/2	1.6	1/16	6.4	1/4	6.4	1/4	HELIUM	18.9	40	375-525	. . .	. . .	260	500	**3**
LAP JOINTS--FILLET WELDED(D)																
1.6	1/16	0	0	1.6	1/16	1.6	1/16	ARGON	7.1	15	130-150	0.25	10	NONE USED		**1**
3.2	1/8	0	0	2.4	3/32	3.2	1/8	ARGON	7.1	15	200-250	0.23	9	NONE USED		**1**
4.8	3/16	0	0	3.2	1/8	3.2	1/8	HELIUM	15.0	30	205-250	0.20	8	95	200	**1**
6.4	1/4	0	0	3.2	1/8	3.2	1/8 1/8	HELIUM	15.0	30	250-280	0.18	7	150	300	**1**
9.5	3/8	0	0	4.8	3/16	4.8	3/16	HELIUM	18.9	40	300-340	. . .	. . .	260	500	**3**
OUTSIDE CORNER JOINTS--SQUARE GROOVE																

3.2	1/8	3.2	1/8	2.4	3/32	3.2	1/8	ARGON	7.1	15	170-225	0.28	11	NONE USED		**1**
4.8	3/16	4.8	3/16	3.2	1/8	3.2	1/8	HELIUM	15.0	30	190-225	0.25	10	95	200	**1**
6.4	1/4	4.8	3/16	3.2	1/8	3.2	1/8	HELIUM	15.0	30	225-260	0.23	9	150	300	**1**
9.5	3/8	6.4	1/4	4.8	3/16	4.8	3/16	HELIUM	18.9	40	280-320	. . .	. . .	260	500	**2**
OUTSIDE CORNER JOINTS--50° SINGLE-BEVEL-GROOVE, 1.6 MM (1/16 IN.) ROOT FACE																
4.8	3/16	1.6	1/16	3.2	1/8	3.2	1/8	HELIUM	15.0	30	205-250	0.20	8	95	200	**1**
6.4	1/4	1.6	1/16	3.2	1/8	3.2	1/8	HELIUM	15.0	30	250-280	0.18	7	150	300	**1**
9.5	3/8	1.6	1/16	4.8	3/16	4.8	3/16	HELIUM	18.9	40	300-340	. . .	. . .	260	500	**3**
INSIDE CORNER JOINTS--SQUARE GROOVE, FILLET WELDED																
3.2	1/8	3.2	1/8	2.4	3/32	3.2	1/8	ARGON	7.1	15	200-250	0.23	9	NONE USED		**1**
T-JOINTS--FILLET WELDED																
3.2	1/8	1.6	1/16	2.4	3/32	3.2	1/8	ARGON	7.1	15	200-250	0.23	9	NONE USED		**1**
4.8	3/16	1.6	1/16	3.2	1/8	3.2	1/8	HELIUM	15.0	30	205-250	0.20	8	95	200	**1**
6.4	1/4	1.6	1/16	3.2	1/8	3.2	1/8	HELIUM	15.0	30	250-280	0.18	7	150	300	**1**
9.5	3/8	**1.6**	1/16	**4.8**	3/16	**4.8**	3/16	**HELIUM**	**18.9**	**40**	**300-340**	**. . .**	**. . .**	**260**	**500**	**3**

Note: The data in this table are intended to serve as starting points for the establishment of optimum joint design and conditions for welding of parts on which previous experience is lacking. They are subject to adjustments necessary to meet the special requirements of individual applications.

(A) COPPER, CARBON, GRAPHITE, OR CERAMIC TAPE BACKING STIPS OR RINGS MAY BE USED (SEE TEXT).

(B) MIXTURES OF ARGON AND HELIUM ARE ALSO USED (SEE TEXT).

(C) DEPTH OF BACK V IS 9.5 MM (3/8 IN.) OF STOCK THICKNESS.

(D) USE OF FILLER METAL IS OPTIONAL FOR THICKNESSES OF 6.4 MM ($\frac{1}{4}$ IN.) OR LESS.

Effect of Cuprous Oxide. Cuprous oxide may be present within the base metal or introduced through oxidation of the molten weld pool during the arc welding process. Migration of cuprous oxides to the grain boundaries lowers the strength and ductility of the weld and adversely affects fatigue properties.

The best results in the arc welding of copper are obtained on the deoxidized coppers because these alloys are free of cuprous oxides and they contain residual phosphorus. This phosphorus combines with oxygen, which is absorbed during the heating or welding, and prevents the formation of cuprous oxide. The strength, ductility, and porosity of welds made in alloy C10200 (oxygen-free copper) are intermediate between the same properties for welds made in deoxidized coppers (alloys C12000 and C12200) and those made in alloy C11000 (ETP copper), which contains 0.02 to 0.5% oxygen.

While cuprous oxide does have an adverse effect on the mechanical properties of coppers, weld properties of arc welded coppers containing cuprous oxide are superior to those obtained from oxyacetylene welding of oxygen-bearing coppers. Welds made by oxyfuel welding are subject to gassing and embrittlement.

Shielding Gases. Argon is the preferred shielding gas when welding copper sections up to 1.6 mm (0.06 in.) thick. As sections grow thicker, slower travel speeds and higher preheat temperatures are required when using argon shielding gas. For sections in excess of 1.6 mm (0.06 in.), helium is the preferred gas. The risk of oxide entrapment in the weld pool is greatly reduced when using helium. Also, compared to argon, helium provides deeper penetration or higher travel speed for the same welding current.

Mixtures of argon and helium shielding gases result in intermediate welding characteristics. For welding positions other than the flat position, a mixture of 65 to 75% He with argon provides a good balance between the penetrating effect of helium and the ease of control for argon. Gas flow rate usually ranges from 7.1 to 19 L/min (15 to 40 ft^3/h), with higher values being used for high currents in the welding of thick sections.

Type of Current. As indicated in Table 5, DCEN is preferred for GTAW of commercial coppers.

Electrodes. In GTAW of coppers, best results for electrode life and maintenance are obtained using a thoriated tungsten electrode containing 2% thoria (EWTh-2). Proper electrode preparation for the welding of copper is usually a pointed electrode with an included angle of 60°. The tip of the point is truncated with a diameter approximately one-third the diameter of the electrode.

Welding Without Filler Metal. Square-groove butt joints on copper up to 3 mm (0.13 in.) thick can be gas-tungsten arc welded without the use of filler metal. However, filler metal is sometimes required as thicknesses approach this value. Copper sections greater than 3 mm (0.13 in.) may be gas-tungsten arc welded without the use of filler metal by making two passes, one from each side.

Filler Metals. As shown in Table 5, copper sections greater than 3 mm (0.13 in.) that are joined by GTAW typically require the use of filler metals. As discussed above, the adverse effect of the formation of cuprous oxide requires that the filler metal contain a residual deoxidizer. The deleterious effect of oxygen is even more severe within the filler metal than within the base metal, because the filler metal has greater exposure to welding heat.

In most cases, copper filler metal is selected that contains a maximum of 0.15% P and 0.50% Si as deoxidizers. This filler metal is known as ERCu, as shown in Table 2. Alloys such as C12000 and C12200 (deoxidized coppers) do not contain enough residual phosphorus to ensure sound welds and are not used as filler metal. The ERCu filler metal has other advantages, such as relatively high electrical conductivity (30 to 40% IACS) and good color match to the copper base metal. Any of the other filler metals listed in Table 2 can be used in GTAW of commercial coppers. Most of the alloys contain adequate amounts of deoxidizing elements such as phosphorus, silicon, iron, aluminum, or titanium. These filler metals offer the advantage of greater joint strengths, but they have lower electrical conductivity and poor color match.

Welding Technique. Either the forehand or backhand welding technique can be used when GTAW is used for coppers. Forehand welding is preferred for all welding positions. It produces a smaller and more uniform bead size than is possible with backhand welding; however, a greater number of passes may be required to fill the joint.

The weld joint should be filled using one or more stringer beads or narrow weave beads. Wide oscillation of the arc while welding should be avoided, lest the edges of the bead be intermittently exposed to the atmosphere and consequently to oxidation. The first bead should penetrate the root of the joint and be relatively thick, to provide time for deoxidation of the weld metal and to avoid cracking of the bead.

Joint Designs. Welding of butt joints with thicker cross sections requires the use of filler metal and a root opening, due to the high thermal conductivity of copper. Clearance is needed to prevent the base metal from conducting the heat away too rapidly and solidifying the filler metal, thus choking the joint before it is filled.

Because of the high fluidity of copper, backing rings or strips are ordinarily used in GTAW of butt joints. The backing rings are typically made of copper, carbon, graphite, or ceramic tape. Backing is needed both for tightly fitted butt joints used in thin metal and for loosely fitted butt joints in thick metal, to prevent the loss of the fluid weld pool. Backing may also be used when needed for this purpose with other joint geometries.

Preheating. For GTAW of sections greater than 3 mm (0.13 in.), preheating of the base material is typically required to maintain the base metal at welding temperature. Without preheating, the high thermal conductivity of coppers results in excessive loss of heat away from the weld zone. The small arc that is typically used in GTAW cannot maintain sufficient welding heat in thick sections of copper, even under the most favorable conditions. Preheat temperatures are also shown in Table 5.

Deoxidized Copper. When copper weldments require the strength of the base metal, deoxidized coppers are usually used. Even with the localized heat input of GTAW, special fixturing and other welding conditions may be needed to minimize distortion, as shown in the following example.

Example 1: Redesign of Joint and Fixture to Minimize Distortion in GTAW of Alloy C12000.

A wafer of nuclear fuel was encapsulated in a can and cover assembly constructed of alloy C12000 using automatic GTAW. The completed weldment is shown in Fig. 1(a). As shown in detail A in Fig. 1, the can has a machined recess for the wafer and a 0.13 mm (0.005 in.) thick bottom. The can-to-cover joint had raised lips, which eliminated the need for filler metal.

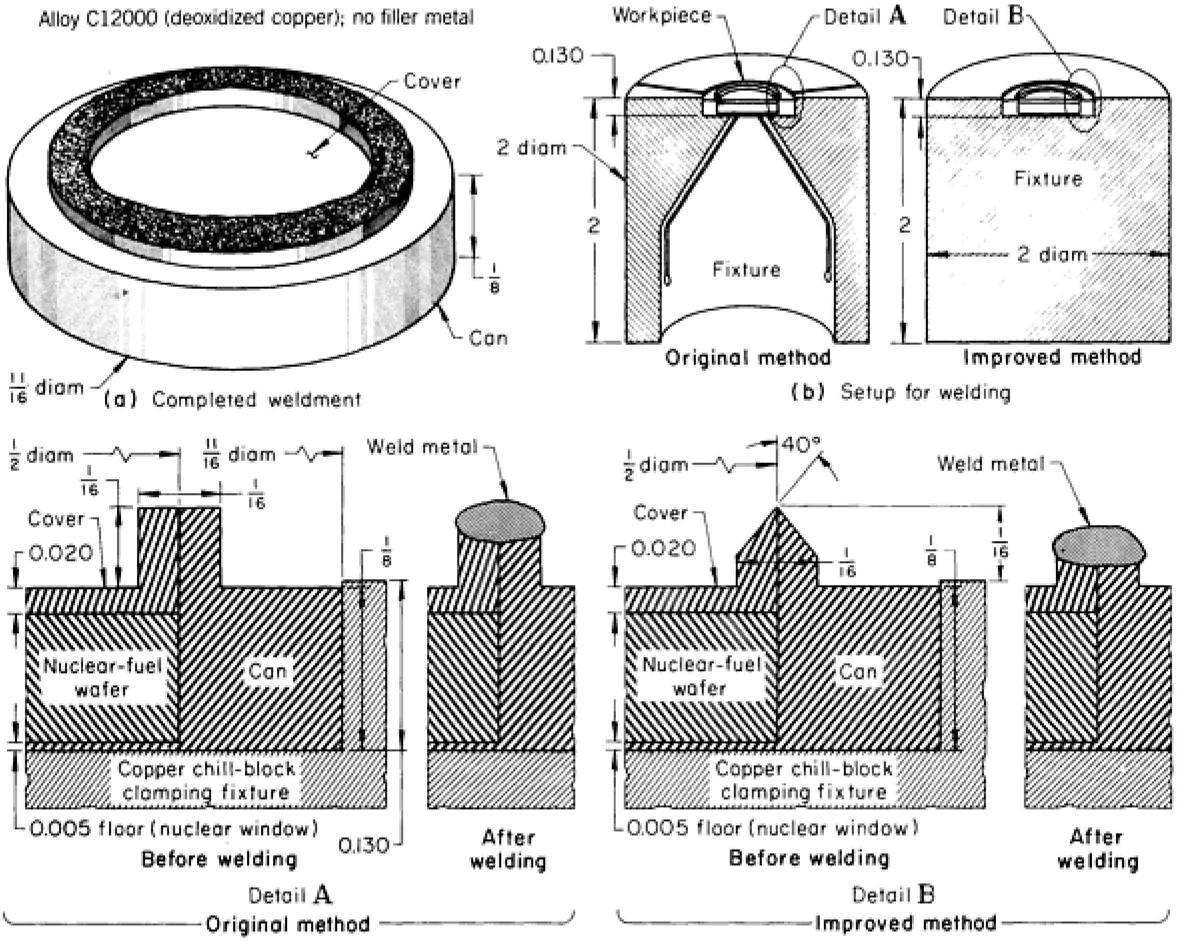
Alloy C12000 (deoxidized copper); no filler metal
Cover
$\frac{1}{8}$
Can
$\frac{11}{16}$ diam
(a) Completed weldment
Workpiece
Detail A
Detail B
0.130
0.130
2 diam
2
Fixture
2
Fixture
2 diam
Original method
Improved method
(b) Setup for welding
$\frac{1}{2}$ diam
$\frac{11}{16}$ diam
$\frac{1}{16}$
$\frac{1}{16}$
Weld metal
Cover
0.020
$\frac{1}{8}$
Nuclear-fuel wafer
Can
Copper chill-block clamping fixture
0.005 floor (nuclear window)
0.130
Before welding
After welding
Detail A
Original method
$\frac{1}{2}$ diam
40°
Weld metal
Cover
0.020
$\frac{1}{16}$
$\frac{1}{8}$
$\frac{1}{16}$
Nuclear-fuel wafer
Can
Copper chill-block clamping fixture
0.005 floor (nuclear window)
Before welding
After welding
Detail B
Improved method

Erro! •

AUTOMATIC GTAW	
JOINT TYPE	**CIRCUMFERENTIAL EDGE**
WELD	**EDGE FLANGE**
POWER SUPPLY	**250 A RECTIFIER**
ELECTRODE	**1.0 MM (0.040 IN.) DIAM EWTH-2**
TORCH	**110 A, AIR-COOLED**(A)
FILLER METAL	**NONE**
FIXTURE	(B)
CURRENT	**16 A (DCEN)**(C)
ARC STARTING	(D)
ARC LENGTH	**0.38 MM (0.015 IN.)**
SHIELDING GAS	**ARGON**(E)
WELDING POSITION	**FLAT**
TRAVEL SPEED	**0.14 M/MIN (5.5 IN./S)**

(A) MODIFIED BY ELIMINATION OF CERAMIC CUP AND BY USE OF BARE CONNECTION CABLE TO AVOID THE PRESENCE OF ORGANIC MATERIAL IN THE CHAMBER.

(B) COPPER CHILL-BLOCK CLAMPING FIXTURE HELD BY A SMALL CHUCK THAT WAS ROTATED BY A VARIABLE-SPEED DRIVE.

(C) WITH CONTINUOUS SUPERIMPOSED HIGH-FREQUENCY CURRENT, FOR ARC STABILITY.

(D) SUPERIMPOSED HIGH-FREQUENCY CURRENT, WHICH WAS CONTINUOUS AS DESCRIBED IN NOTE(C).

FIG. 1 CONDITIONS FOR GTAW OF A NUCLEAR FUEL CONTAINER. DISTORTION IN WELDING WAS MINIMIZED BY INCREASING THE MASS OF THE FIXTURE TO INCREASE HEAT WITHDRAWAL AND BY CHAMFERING JOINT EDGES TO PERMIT REDUCTION OF WELDING CURRENT.

The weld had to provide a hermetic seal without distortion of the assembly, including the thin can bottom that was only 4.8 mm (0.19 in.) from the welded area. A mass spectrometer was used to conduct leak testing of the weldment.

The welds were made in a controlled atmosphere chamber with the can mounted in a rotating copper chill-block clamping fixture underneath a stationary electrode. Welding conditions are given in the table accompanying Fig. 1.

Originally the fixture and joint shown in Fig. 1 and detail A were used. However, at currents high enough for fusion, the welds were porous, the joint leaked, and the can warped unacceptably.

The porosity that caused leaking resulted from the machining operation of the can. The lubricant was wiped under the surface of the metal by the cutting tool, and subsequent cleaning operations failed to remove it. Machining without a lubricant eliminated the difficulty. To minimize distortion, the joint and fixture were improved as follows. The square lips of the joint were chamfered (40°) to a sharp edge, as shown in detail B in Fig. 1. This was done to reduce the heat input required to make the weld. With the chamfered lips, the welding current could be lowered because a smaller mass of metal was being melted at any instant. The mass of the fixture was increased, as shown in Fig. 1(b), to increase heat withdrawal. The redesigned fixture also provided complete support for the thin can bottom against pressure during leak testing.

Before welding, the assembly was retained in the fixture, which was held by a small chuck with all of the 0.13 mm (0.005 in.) thick can bottom resting on the thick copper backing, as shown in detail B of Fig. 1. The fixture assembly was welded in a chamber that was first vacuum-purged and then filled with 100% Ar under slightly positive pressure.

The torch was rigidly mounted on a sliding base with locating stops with which the electrode tip was positioned over the seam and clamped into position. In the argon atmosphere it was not necessary to provide a shielding gas to the electrode tip or to use an electrode cup. The electrode was ground to a sharp point to reduce current flow and arc wander.

During welding, the chuck holding the fixture was rotated by a variable-speed drive. Arc stability was maintained by a superimposed high-frequency current that was continuous. Current upslope was not controlled. Two revolutions at 3.5 rev/min were used, because slight variations in pressure between the cover and the can caused variations in the thickness of the weld bead after the first revolution. The second revolution smoothed out these irregularities and ensured adequate penetration of the base metal along the circumference of the joint. After the second revolution, the current was tapered to 0 A in 5 s.

Distortion was minimized because of the low welding current, the high heat-sink efficiency, and rigid support by the fixture. Control of heat input was also helpful in preventing excessive internal gas expansion after the joint was closed.

After welding, the chamber was evacuated to 1.3×10^{-4} Pa (10^{-6} torr). The interior of the can remained at approximately 0.1 MPa (15 psi). Under the pressure differential, the can, still in the fixture, was inspected for leaks by use of a mass spectrometer. The fixture supported the can bottom against internal pressure.

Oxygen-Free and ETP Coppers. Compared to GMAW or other processes with nonlocalized heat input, GTAW is preferred for the joining of either oxygen-free copper (alloy C10200) or ETP copper (alloy C11000) for thicknesses up to approximately 13 mm (0.5 in.).

Because of the absence of a deoxidizer, gas-tungsten arc welds in alloy C 10200 have slightly lower strength and ductility and are slightly more porous. Gas-tungsten arc welds in alloy C11000 have somewhat lower tensile strength than those in deoxidized coppers, and they are also more porous (see the section "Effect of Cuprous Oxide" in this article). However, the properties of gas-tungsten arc welds in ETP copper are acceptable in many applications involving welded electrical conductors, particularly when tensile strength is relatively unimportant, as in the induction-coil weldment described in the following example.

Example 2: Automatic GTAW of Joints between Alloy C11000 Bar and Tubing.

A solid bar approximately 9 m (30 ft) long of alloy C11000 was gas-tungsten arc welded automatically to a comparable length of tubing of the same alloy. The weldment was used in making water-cooled induction coils for low frequency (60 to 180 Hz) induction heating that required more cross-sectional area for electrical conduction than was provided by the tubular section alone. Figure 2 shows a weldment with a cross section of typical size. The 9 m (30 ft) length was standard for inductor stock. The weldment was bent into coils with an inside diameter of 100 mm (4 in.). The bar stock was located to the inside of the coil. The requirement of the joint was to maintain good electrical conductivity.

Alloy C11000 (ETP copper), annealed to half hard; no filler metal

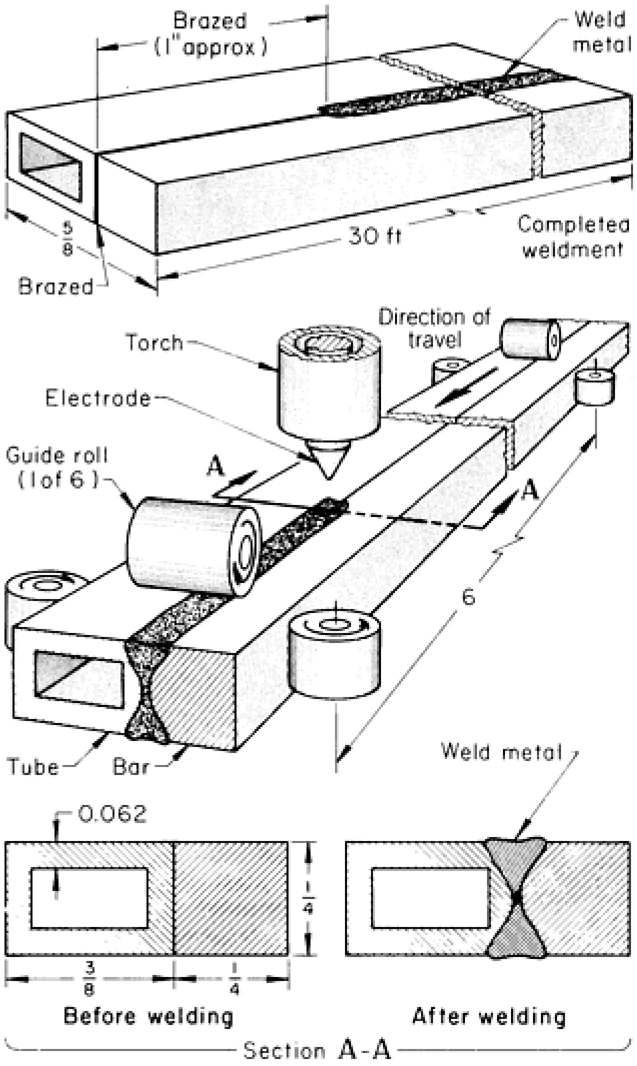
Brazed
(1" approx)
Weld metal
5/8
30 ft
Completed weldment
Brazed
Torch
Direction of travel
Electrode
Guide roll
(1 of 6)
A
A
6
Tube
Bar
Weld metal
0.062
1/4
3/8
1/4
Before welding
After welding
Section A-A

AUTOMATIC GTAW	
JOINT TYPE	**BUTT**
WELD TYPE	**SQUARE GROOVE, ZERO ROOT OPENING**
POWER SUPPLY	**500 A RECTIFIER**$^{(A)}$
ELECTRODE	**3.2 MM ($\frac{1}{8}$ IN) DIAM EWTH-2**$^{(B)}$
TORCH	**350 A, WATER-COOLED**$^{(C)}$
FILLER METAL	**NONE**
FIXTURE	**6 GUIDE ROLLS**$^{(D)}$
NUMBER OF PASSES	**2 (ONE PER SIDE)**
CURRENT	**220 A (DCEN)**
ARC STARTING	**TORCH START**
ARC LENGTH	**3.2 MM ($\frac{1}{8}$ IN.)**
SHIELDING GAS (HELIUM) FLOW RATE	**9.5 L/MIN (20 FT^3/H)**
WELDING POSITION	**FLAT**
TRAVEL SPEED	**0.33 M/MIN (13 IN./MIN)**

CUP-TO-WORK DISTANCE	**7.9 MM ($\frac{5}{16}$ IN.)**
PREHEAT/POSTHEAT	**NONE**

(A) CONSTANT VOLTAGE.

(B) TAPER GROUND.

(C) FIXED, ON AN ADJUSTABLE MOUNT.

(D) VARIABLE-SPEED DRIVE AND CHAIN USED TO PULL ASSEMBLY THROUGH ROLLS

FIG. 2 CONDITIONS FOR AUTOMATIC GTAW OF ALLOY C11000 TUBING AND BAR INDUCTOR COILS. THE POSITIONS OF THE ASSEMBLY, GUIDE ROLLS, AND TORCH DURING WELDING ARE SHOWN.

Originally, the bar and tubing were torch brazed along the entire length using a silver alloy filler metal. Not only was the filler metal expensive, but brazing was slow, requiring two operators to handle the lengths of tubing and bar. A faster and less costly method was developed by mechanizing a GTAW procedure in which no filler metal was used, thus increasing the production rate.

Welding conditions for the application are given in the table accompanying Fig. 2. The fixture, designed to accommodate the sizes of inductor bar and tubing regularly used, consisted of six guide rolls, as shown in Fig. 2. These guide rolls properly aligned the assembly during welding. A chain and variable-speed drive was used to pull the assembly through the rolls during welding. Positioning of the torch over the joint was achieved by using an adjustable mount.

The first 25 mm (1 in.) of the leading edge of the joint was brazed prior to welding, to prepare the assembly for feeding through the guide rolls. The position of the assembly, guide rolls, and electrode holder during welding is shown in Fig. 2. The relative travel speed of the assembly as pulled through the rolls was 5.5 mm/s (13 in./min). A single pass was used without coolant for each side of the weldment and the ends were trimmed after welding.

The high thermal conductivity of alloy C11000 caused rapid dissipation of the heat energy. However, this difficulty was overcome without preheating by the use of helium as the shielding gas. Helium provided a hotter arc than could be obtained with argon. Thus, it was possible to produce the relatively deep, narrow beads at the stated travel speed. Oxygen from the oxide in alloy C11000 caused some porosity, but this was minimized by the fast freezing promoted by the high travel speed. The required joint properties were obtained in spite of the potential for porosity.

Gas-Metal Arc Welding

As has been previously discussed, oxygen has a detrimental effect when copper is welded. The resulting porosity and reduced weld strength described for GTAW of copper are more pronounced in GMAW. This is due to the greater heat input associated with the process. Gas-metal arc welds of deoxidized coppers compare favorably in density and strength with the welds made using GTAW. However, the greater heat input and lesser localization of heat obtained with GMAW result in greater porosity and lower strength in the HAZ of the welds. This is especially true in coppers that contain insufficient amounts of deoxidizer, particularly ETP copper. Consequently, GMAW has less applicability than GTAW to the welding of ETP copper, oxygen-free copper, or low-phosphorus-deoxidized copper up to approximately 13 mm (0.5 in.) thick.

Nominal Conditions. Table 6 gives the nominal conditions for gas-metal arc butt welding of copper and copper alloys. In welding joints for which the root opening is between 2.5 and 3 mm (0.10 and 0.13 in.), mixtures of helium and argon are used as shielding gas instead of pure argon. This results in lower requirements for preheat, deeper penetration into the weld joint, and higher rates of filler metal deposition.

TABLE 6 NOMINAL CONDITIONS FOR GAS-METAL ARC BUTT WELDING OF COMMERCIAL COPPERS AND COPPER ALLOYS

WELD TYPES FOR BUTT JOINTS, AS DESCRIBED	**WORKPLACE THICKNESS**		**ROOT FACE**		**ROOT OPENING(B)**		**ELECTRODE TYPE**	**ELECTRODE DIAMETER**		**SHIELDING GAS(C)**	**GAS FLOW RATE**		**CURRENT (DCEP), A**	**VOLTAGE, V**	**TRAVEL SPEED**		**PREHEAT TEMPERATURE**		NUMBER OF PASSES
	MM	**IN.**	**MM**	**IN.**	**MM**	**IN.**		**MM**	**IN.**		**L/MIN**	**FT^3/H**			**M/MIN**	**IN./MIN**	**°C**	**°F**	
COMMERCIAL COPPERS																			
SQUARE GROOVE(B)	3.2	1/8	3.2	1/8	0	0	ERCU	1.6	1/16	ARGON	15	30	310	27	0.76	30	NONE USED		**1**
SQUARE GROOVE(C)	3.2	1/8	3.2	1/8	0-1.6	0-1/16	ERCU	1.6	1/16	ARGON(D)	15-16.5	30-35	325-350	28-33	. . .	. . .	NONE USED		**1**
SQUARE GROOVE	6.4	1/4	6.4	1/4	0	0	ERCU	2.4	3/32	ARGON	15	30	460	26	0.50	20	95	200	**2**
	6.4	1/4	6.4	1/4	0	0	ERCU	2.4	3/32	ARGON	15	30	500	27	0.50	20	95	200	**1**
75-90° SINGLE-V-GROOVE(C)	6.4	1/4	3.2	1/8	0-3.2	0-1/8	ERCU	1.6	1/16	ARGON(D)	15-16.5	30-35	400-425	32-36	. . .	. . .	200-260	400-500	**2**
	12.7	1/2	0-3.2	0-1/8	0-3.2	0-1/8	ERCU	1.6	1/16	ARGON(D)	15-16.5	30-35	425-450	35-40	. . .	. . .	425-480	800-900	**4**
90° SINGLE-V-GROOVE(C)	9.5	3/8	4.8	3/16	0	0	ERCU	2.4	3/32	ARGON	15	30	500	27	0.35	14	200	400	**(E)**
	9.5	3/8	4.8	3/16	0	0	ERCU	2.4	3/32	ARGON	15	30	550	27	0.35	14	200	400	**(E)**
	12.7	1/2	6.4	1/4	0	0	ERCU	2.4	3/32	ARGON	15	30	540	27	0.30	12	200	400	**(E)**
	12.7	1/2	6.4	1/4	0	0	ERCU	2.4	3/32	ARGON	15	30	600	27	0.25	10	200	400	**(E)**
ALLOYS C17000 AND C17200 (HIGH-STRENGTH BERYLLIUM COPPERS)(F)																			
90° SINGLE-V-GROOVE	6.4-9.5	1/4 -	0.8-1.6	1/32 -	. . .	. . .	ALLOY C17000, C17500	1.2	0.045	5AR/HE	21	45	175-200	. . .	. . .	. . .	150-200	300-400	**3-4(G)**

		$\frac{1}{2}$		$\frac{1}{16}$															
30° DOUBLE-U-GROOVE(I)	19-38.1	$\frac{3}{4}$-1 $\frac{1}{2}$	1.6	$\frac{1}{16}$	. . .	. . .	ALLOY C17000, C17500	1.6	$\frac{1}{16}$	AR/HE	28	60	325-350	. . .	. . .	. . .	150-200	300-400	**10-20**(H)
ALLOY C17500 (HIGH-CONDUCTIVITY BERYLLIUM COPPERS)(F)																			
90° SINGLE-V-GROOVE	6.4-9.5	$\frac{1}{4}$-$\frac{1}{2}$	0.8	$\frac{1}{32}$	. . .	. . .	ALLOY C17500	1.2	0.045	AR/HE	15	30	200-240	. . .	. . .	. . .	315	600	**3-4**(J)
	19.0	$\frac{3}{4}$	0.8	$\frac{1}{32}$	. . .	. . .	ALLOY C17500	1.2	0.045	AR/HE	15	30	200-240	. . .	. . .	. . .	480	900	**6**(J)
LOW-ZINC BRASSES																			
SQUARE GROOVE(C)	3.2	$\frac{1}{8}$	3.2	$\frac{1}{8}$	0	0	ERCUSI-A	1.6	$\frac{1}{16}$	ARGON	15	30	275-285	25-28	. . .	. . .	NONE USED		**1**
	3.2	$\frac{1}{8}$	3.2	$\frac{1}{8}$	0	0	ERCUSI-A	1.6	$\frac{1}{16}$	HELIUM	16.5	35	275-285	25-28	. . .	. . .	NONE USED		**1**
60° SINGLE-V-GROOVE(C)	9.5	$\frac{3}{8}$	0	0	3.2	$\frac{1}{8}$	ERCUSI-A	1.6	$\frac{1}{16}$	ARGON	15	30	275-285	25-28	. . .	. . .	NONE USED		**2**
	12.7	$\frac{1}{2}$	0	0	3.2	$\frac{1}{8}$	ERCUSI-A	1.6	$\frac{1}{16}$	ARGON	15	30	275-285	25-28	. . .	. . .	NONE USED		**4**
70° SINGLE-V-GROOVE(C)	9.5	$\frac{3}{8}$	0	0	3.2	$\frac{1}{8}$	ERCUSN-A	1.6	$\frac{1}{16}$	HELIUM	16.5	35	275-285	25-28	. . .	. . .	260(K)	500(K)	**2**
	12.7	$\frac{1}{2}$	0	0	3.2	$\frac{1}{8}$	ERCUSN-A	1.6	$\frac{1}{16}$	ARGON	15	30	275-285	25-28	. . .	. . .	260(K)	500(K)	**4**
HIGH-ZINC BRASSES, TIN BRASSES, SPECIAL BRASSES, NICKEL SILVERS																			
SQUARE GROOVE(C)	3.2	$\frac{1}{8}$	3.2	$\frac{1}{8}$	0	0	ERCUSN-A	1.6	$\frac{1}{16}$	ARGON	15	30	275-285	25-28	. . .	. . .	NONE USED		**1**
70° SINGLE-V-GROOVE(C)	9.5	$\frac{3}{8}$	0	0	3.2	$\frac{1}{8}$	ERCUSN-A	1.6	$\frac{1}{16}$	ARGON	15	30	275-285	25-28	. . .	. . .	NONE USED		**2**
	12.7	$\frac{1}{2}$	0	0	3.2	$\frac{1}{8}$	ERCNSN-A	1.6	$\frac{1}{16}$	ARGON	15	30	275-285	25-28	. . .	. . .	NONE USED		**4**
PHOSPHOR BRONZES(L)																			
90° SINGLE-	9.5	$\frac{3}{8}$	0	0	3.2	$\frac{1}{8}$	ERCUSN-A	1.6	$\frac{1}{16}$	HELIUM	16.5	35	275-285	25-28	. . .	. . .	95-150	200-300	**3-4**(M)

V-GROOVE(C)	12.7	$\frac{1}{2}$	0	0	3.2	$\frac{1}{8}$	ERCUSN-A	1.6	$\frac{1}{16}$	HELIUM	16.5	35	275-285	25-28	. . .	. . .	175-200	350-400	**5-6(M)**
ALUMINUM BRONZES(N)																			
SQUARE GROOVE(O)	3.2	$\frac{1}{8}$	3.2	$\frac{1}{8}$	0	0	ERCUAL-A2	1.6	$\frac{1}{16}$	ARGON	15	30	280-290	27-30	. . .	. . .	NONE USED		**1**
60-70° SINGLE V-GROOVE(C)	9.5	$\frac{3}{8}$	0	0	3.2	$\frac{1}{8}$	ERCUAL-A2	1.6	$\frac{1}{16}$	ARGON	15	30	280-290	27-30	. . .	. . .	NONE USED		**2**
	12.7	$\frac{1}{2}$	0	0	3.2	$\frac{1}{8}$	ERCUAL-A2	1.6	$\frac{1}{16}$	ARGON	15	30	280-290	27-30	. . .	. . .	NONE USED		**3**
SILICON BRONZES(P)																			
SQUARE GROOVE(Q)	3.2	$\frac{1}{8}$	3.2	$\frac{1}{8}$	0	0	ERCUSI-A	1.6	$\frac{1}{16}$	ARGON	15	30	260-270	27-30	. . .	. . .	NONE USED		**1**
70° SINGLE-V-GROOVE(C)	9.5	$\frac{3}{8}$	0	0	3.2	$\frac{1}{8}$	ERCUSI-A	1.6	$\frac{1}{16}$	ARGON	15	30	260-270	27-30	. . .	. . .	NONE USED		**2**
	12.7	$\frac{1}{2}$	0	0	3.2	$\frac{1}{8}$	ERCUSI-A	1.6	$\frac{1}{16}$	ARGON	15	30	260-270	27-30	. . .	. . .	NONE USED		**3**
COPPER NICKELS																			
SQUARE GROOVE(C)	3.2	$\frac{1}{8}$	3.2	$\frac{1}{8}$	0	0	ERCUNI	16	$\frac{1}{16}$	ARGON	15	30	280	27-30	. . .	. . .	NONE USED		**1**
60° SINGLE-V-GROOVE(C)	9.5	$\frac{3}{8}$	0	0	3.2	$\frac{1}{8}$	ERCUNI	1.6	$\frac{1}{16}$	ARGON	15	30	280	27-30	. . .	. . .	NONE USED		**2**
	12.7	$\frac{1}{2}$	**0**	**0**	**3.2**	$\frac{1}{8}$	**ERCUNI**	**16**	$\frac{1}{16}$	**ARGON**	**15**	**30**	**280**	**27-30**	**. . .**	**. . .**	**NONE USED**		**4**

Note: The data in this table are intended to serve as starting points for the establishment of optimum joint design and conditions for welding of parts on which previous experience is lacking. They are subject to adjustments necessary to meet the special requirements of individual applications.

(A) THICKNESSES UP TO 38.1 MM (1$\frac{1}{2}$ IN.) ARE SOMETIMES WELDED BY USE OF SLIGHTLY HIGHER CURRENT AND LOWER TRAVEL SPEED THAN SHOWN FOR A THICKNESS OF 12.7 MM ($\frac{1}{2}$ IN.).

(B) COPPER BACKING.

(C) GROOVED COPPER BACKING.

(D) OR 75AR-25HE.

(E) SPECIAL WELDING SEQUENCE IS USED (SEE TEXT).

(F) SEE TABLE 1 FOR COMPOSITIONS.

(G) THE FINAL PASS IS MADE ON THE ROOT SIDE AFTER BACK CHIPPING. WIRE BRUSH AFTER EACH PASS.

(H) SEVERAL PASSES ARE MADE ON THE FACE SIDE, THEN SEVERAL ON THE BACK SIDE, UNTIL THE WELD IS COMPLETED. BACK CHIP THE ROOT PASS BEFORE MAKING THE FIRST PASS ON THE BACK SIDE. WIRE BRUSH AFTER EACH PASS.

(I) SIMILAR TO THE DOUBLE V-GROOVE WELD, BUT WITH A GROOVE RADIUS OF 9.5 MM ($\frac{3}{8}$ IN.).

(J) THE FINAL PASS IS MADE ON THE ROOT SIDE AFTER BACK CHIPPING.

(K) SHOULD NOT BE OVERHEATED; AS LITTLE PREHEAT AS POSSIBLE SHOULD BE USED.

(L) WELDING CONDITIONS BASED ON ALLOYS C51000, C52100, AND C52400; CURRENT IS INCREASED OR SPEED IS DECREASED FOR ALLOY C50500.

(M) HOT PEENING BETWEEN PASSES IS RECOMMENDED FOR MAXIMUM STRENGTH.

(N) SLIGHT PREHEAT MAY BE NEEDED ON HEAVY SECTIONS; INTERPASS TEMPERATURE SHOULD NOT EXCEED 315 °C (600 °F).

(O) WITH ALUMINUM BRONZE BACKING.

(P) NO PREHEAT IS USED ON ANY THICKNESS; INTERPASS TEMPERATURE SHOULD NOT EXCEED 95 °C (200 °F).

(Q) WITH SILICON BRONZE BACKING

Filler Metals (Electrode Wires). The recommended ERCu electrode contains phosphorus, tin, silicon, and manganese as deoxidizers to minimize porosity. The electrode produces sound, trouble-free welds that also provide good color match to the copper base metal and have good electrical conductivity. The ERCu electrode was used as a basis for the conditions listed in Table 6.

Copper electrodes of higher purity than ERCu are seldom used in GMAW of copper, because these usually result in welds that are porous, due to the absence of deoxidizers. Most of the electrodes listed in Table 3 can be used. Any of the electrodes other than ERCu will produce dense, strong welds, but they will result in poor color match and unacceptable electrical conductivities. Filler metals should be deposited as stringer beads or narrow weave beads using the spray-transfer mode; weaving of the electrodes may result in oxidation at the edges of the weld.

Welding Technique. The forehand welding technique should be used in the flat position. For vertical-position welding, an upward progression should be utilized. Gas-metal arc welding of copper is not recommended for the overhead position due to poor bead contour. The preferred process for overhead-position welding is GTAW.

Joint Design. A square-groove joint is used for the single-pass welding of coppers up to 3 mm (0.13 in.) thick. A backing bar is used for zero root openings, and a grooved copper backing bar is used for root openings of 1.6 mm (0.06 in.) maximum. The square-groove joint can also be used for one-pass-per-side welding of coppers up to 6 mm (0.25 in.) thick.

Single-V-groove joints are used when welding coppers of sections above 6 mm (0.25 in.). The filler metal is deposited on one side using three or more passes, and the root pass is back gouged to sound metal before the last pass is applied to the back of the joint. In some applications, the root pass is applied using GTAW and subsequent passes are completed via GMAW for rapid buildup.

If heavy sections (thicker than 13 mm, or 0.5 in.) are to be welded, these should be prepared with a double-V- or a double-U-groove joint. Welding is performed by applying alternating passes to opposite sides of the joint, if readily accessible, to minimize distortion. In a small, closed vessel, limited access may prevent the use of this technique, and in addition, heat buildup often prevents welding on both sides.

Preheating. Because of the high thermal conductivity of copper, sections thicker than 6 mm (0.25 in.) are usually preheated. As shown in Table 6, single-V-groove joints in 6 mm (0.25 in.) thick copper, welded using 1.6 mm ($\frac{1}{16}$ in.) diameter filler metal, are also preheated.

Deoxidized Coppers. Heavy-wall copper pressure vessels (up to about 40 mm, or 1.5 in., wall thickness), which are usually made from phosphorus-deoxidized copper, are frequently gas-metal arc welded. Residual phosphorus helps obtain maximum weld strength while minimizing porosity. In order to meet the high heat demand in welding these heavy-walled vessels, welding currents and preheat temperatures may be higher than those shown in Table 6.

Similar operating conditions are employed in welding crucibles for arc melting of refractory metals in controlled atmospheres. These crucibles are typically heavy-wall deoxidized copper cylinders with wall thicknesses up to 32 mm (1.25 in.), diameters of 0.3 to 1.2 m (1 to 4 ft), and lengths of 1 to 7.5 m (3 to 25 ft). Formed and welded copper plate is used, because seamless tubes of this size are not available. Longitudinal seams are welded, and flanges are welded to each end of the cylinders by GMAW. Such cylinders are preheated to 650 °C (1200 °F) before welding. Welding is done automatically, to avoid exposing operators to such heavy sections at elevated temperatures.

Oxygen-Free and ETP Coppers. The GMAW process is also used for welding oxygen-free copper in the production of the crucibles previously described. These crucibles are also used for arc melting of refractory metals. However, the weld properties are inferior to those produced with deoxidized copper. Because oxygen-free copper contains no residual deoxidizer, heating and welding cycles must be kept as short as possible, and gas shielding must be completely effective to avoid excessive porosity. Although the strength and soundness of welds in oxygen-bearing coppers, such as ETP copper, are substantially less than in oxygen-free copper, the reduced properties are seldom important for welds in electrical conductors. Busbars made of ETP copper are frequently joined using the GMAW process. The use of filler metals such as ERCu or another of the standard filler metals listed in Table 3 significantly lowers the electrical conductivity of the joint. In order to minimize this adverse effect, large root faces are used to increase the contact area at the joint line. For improved electrical conductivity, a filler metal with chemical composition approximately 0.75% Sn, 0.25% Si, and 0.20% Mn is sometimes used for welding ETP copper busbars.

Shielded Metal Arc Welding

Problems with porosity and low weld strength, due to oxygen content of the base metal and oxygen absorption during welding, are more severe in joining coppers by SMAW than by the gas-shielded processes. The resulting lower mechanical properties greatly limit the usefulness of weldments produced using SMAW. Additionally, electrical conductivity suffers greatly.

Revised by M. Ned Rogers, Batesville Casket Company

Welding of High-Strength Beryllium Coppers

The wrought high-strength beryllium coppers include alloys C17000 and C17200, which contain nominally 1.7 and 1.9% Be, respectively, and up to 0.25% Co. A free-machining grade version of C17200 is available that is modified with a small lead addition, but this grade is not recommended for welding. The high-strength beryllium coppers are joined by GTAW and GMAW. Additional information on the properties and application of these materials can be found in Volume 2 of the *ASM Handbook.*

Gas-Tungsten Arc Welding

Copper alloys C17000 and C17200, high-strength beryllium coppers, are readily welded using GTAW. These alloys are more easily welded than the high-conductivity beryllium coppers, which exhibit higher melting temperature and less fluidity. Gas-tungsten arc welding can be used for section thicknesses in excess of 13 mm (0.5 in.) when the use of the preferred GMAW process is not practical.

Nominal Conditions. Table 7 lists the suggested starting points for nominal conditions for GTAW of high-strength beryllium coppers. The shielding gas is usually a mixture of argon and helium to obtain a hot arc, smooth and spatter-free welds, and maximum electrode life.

TABLE 7 NOMINAL CONDITIONS FOR GTAW OF BERYLLIUM COPPERS

For butt joints having zero root opening; welding with a zirconiated tungsten electrode, filler metal of the same composition as the base metal, argon-helium gas at 12 L/min (25 ft^3/h)

WORKPIECE THICKNESS		**BUTT-JOINT GROOVE**	**ELECTRODE DIAMETER**		**CURRENT(A), A**	**TRAVEL SPEED**		**PREHEAT TEMPERATURE**		NUMBER OF PASSES
MM	**IN.**		**MM**	**IN.**		**M/MIN**	**IN./MIN**	**°C**	**°F**	
ALLOY C17500 (HIGH-CONDUCTIVITY BERYLLIUM COPPER)(B)										
0-2.4	$0\text{-}\frac{3}{32}$	SQUARE	2.4	$\frac{3}{32}$	150	0.13-0.25	5-10	NONE USED		**1**
2.4-3.2	$\frac{3}{32}\text{-}\frac{1}{8}$	90° SINGLE-V(C)	4.8	$\frac{3}{16}$	250	0.13-0.25	5-10	NONE USED		**1-2**
6.4	$\frac{1}{4}$	90° SINGLE-V(C)	4.8	$\frac{3}{16}$	250	0.13-0.25	5-10	425	800	**4-5**
ALLOYS C17000 AND C172000 (HIGH-STRENGTH BERYLLIUM COPPERS(C)										
0-2.4	$0\text{-}\frac{3}{32}$	SQUARE	2.4	$\frac{3}{32}$	150	0.13-0.25	5-10	NONE USED		**1**
2.4-3.2	$\frac{3}{32}\text{-}\frac{1}{8}$	90° SINGLE-V(C)	2.4	$\frac{3}{32}$	180	0.13-0.25	5-10	NONE USED		**1**
6.4-12.7(D)	$\frac{1}{4}\text{-}\frac{1}{2}$ (D)	90° SINGLE-V(C)	4.8	$\frac{3}{16}$	250	0.13-0.25	5-10	150	300	**3-4**
12.7(D)	$\frac{1}{2}$ **(D)**	**90° SINGLE-V(C)**	**4.8**	$\frac{3}{16}$	**250**	**0.13-0.25**	**5-10**	**200**	**400**	**5-8**

Note: The data in this table are intended to serve as starting points for the establishment of optimum joint design and conditions for welding of parts on which previous experience is lacking. They are subject to adjustments necessary to meet the special requirements of individual applications.

(A) HIGH-FREQUENCY-STABILIZED ALTERNATING CURRENT IS PREFERRED; DCEN, WITH A THORIATED TUNGSTEN ELECTRODE, IS SUITABLE UNDER SOME CONDITIONS (SEE TEXT).

(B) FOR COMPOSITION, SEE TABLE 1.

(C) MAXIMUM ROOT FACE IS 1.6 MM ($\frac{1}{16}$ IN.).

(D) GAS-TUNGSTEN ARC WELDING IS USED ON THESE THICKNESSES ONLY WHEN GMAW CANNOT BE USED.

Electrodes. The preferred electrode when using alternating current is zirconiated tungsten (EWZr); noncritical applications can use the more economical pure tungsten (EWP). For DCEN, the thoriated electrodes (such as EWTh-2) are preferred.

Filler Metals. The addition of filler metals is almost always required to provide proper fill for the joint or to provide joint reinforcement in a V-groove weld. This is due to the high concern for strength when welding these alloys. Rods or strips of the same composition as the base metal are generally used as filler metals; the standard filler metals of other copper alloys are weaker and offer no advantages.

Joint Design. As Table 7 indicates, the typical joint designs for GTAW of high-strength beryllium copper are square-groove or 90° single-V-groove butt joints with a 1.5 mm (0.06 in.) maximum root face and zero root opening. The 90° single-V-groove butt welds can be used for thicknesses in excess of 13 mm (0.5 in.).

Preheating and Postweld Heat Treatment. As shown in Table 7, preheat temperatures of 150 to 200 °C (300 to 400 °F) are recommended for GTAW of section thicknesses greater than 3 mm (0.13 in.). Maximum weld strength is obtained by solution annealing and aging after welding. Aging treatments are 3 h at 315 °C (600 °F) for alloy C17000 and 3 h at 345 °C (650 °F) for alloy C17200. This postweld treatment does not consistently provide the full strength of solution-annealed and aged base metal, even for welds made under optimum conditions. Higher strength can be obtained by cold working the annealed metal to a higher temper and modifying the aging treatment.

For some applications, the intermediate weld metal strength obtained by aging after welding, without a solution anneal, is adequate. Omission of solution annealing avoids the expense and distortion associated with this high-temperature operation. Following are three applications of GTAW of high-strength beryllium coppers. Welding conditions for GTAW for three applications are shown in Table 8.

TABLE 8 THREE APPLICATIONS OF GTAW OF HIGH-STRENGTH BERYLLIUM COPPERS

	PRESSURE VESSEL (EXAMPLE 3)	**GENERATOR LINER (EXAMPLE 4)**	COVER TO HOUSING (EXAMPLE 5)
ALLOY	C17000	C17000 TO C17200	**CAST BE-CU (1.7% BE) TO C17000**
SIZE OF WORKPIECE	483 MM DIAM BY 1016 MM (19 IN. DIAM BY 40 IN.)	495 MM DIAM BY 1219 MM (19.5 IN. DIAM BY 48 IN.)	**295 MM (11.6 IN.)**
THICKNESS OF WORK METAL	19 MM ($\frac{3}{4}$ IN.)	(A)	**1.6 MM ($\frac{1}{16}$ IN.)**
FILLER METAL	1.6 MM ($\frac{1}{16}$ IN.) DIAM ALLOY C17200	NONE	**NONE**
CURRENT, A	100-150 DCEN	122, AC(B)	**275, AC(B)**
SHIELDING GAS	ARGON AND HELIUM	ARGON AND HELIUM	**ARGON AND HELIUM**
GAS FLOW RATE	12 L/MIN (25 FT^3/H)	12 L/MIN (25 FT^3/H)	**9.5 L/MIN (20 FT^3/H)**
TRAVEL SPEED	0.50 M/MIN (20 IN./MIN)	0.25 M/MIN (10 IN./MIN)	**0.30 M/MIN (12 IN./MIN)**
NUMBER OF PASSES	10	1	**1**
POSTWELD HEAT	345 °C (650 °F), 3 H	NONE	**NONE**
TESTING METHOD	**X-RAY(C)**	**. . .**	**(D)**

(A) REINFORCING RINGS, 40 BY 32 MM (1$\frac{1}{2}$ BY 1$\frac{1}{4}$ IN.) WELDED TO 6 BY 6 MM ($\frac{1}{4}$ × $\frac{1}{4}$IN.) LONGITUDINAL RIBS.

(B) HIGH-FREQUENCY-STABILIZED.

(C) WELD WAS TESTED BY DYE-PENETRATION TECHNIQUES; LEAK TESTING WAS BY HELIUM MASS SPECTROMETRY AND X-RAY RADIOGRAPHY.

(D) WELD DEPTH OF 1.6 MM ($\frac{1}{16}$ IN.) MINIMUM WAS DETERMINED ULTRASONICALLY; WELD

WAS LEAK-TESTED BY HELIUM MASS SPECTROMETRY AT 76 MPA (11 KSI).

Example 3: Gas-Tungsten Arc Welding of High-Strength Beryllium Coppers in an Alloy C17000 Pressure Vessel.

The pressure vessel was used to scavenge propane from a freon bubble chamber. Use of GTAW was dictated because the vessel could not be solution-annealed after welding because of danger of warpage. The service conditions for this vessel included pressure of 41 MPa (6 ksi) and rapid thermal cycling. This application is shown in Table 8.

Example 4: Gas-Tungsten Arc Welding of High-Strength Beryllium Copper in a Cylindrical Generator Liner.

A cylindrical generator liner was constructed using alloy C17000 and alloy C17200. It was back-extruded from a cast billet. Longitudinal rings about 6 by 6 mm (0.25 by 0.25 in.) were machined on the outer surface. Nine reinforcing rings were rolled from extruded alloy C17000, equally spaced along the longitudinal rings, and welded to them. The reinforcing rings measured 40 mm (1.5 in.) wide and 32 mm (1.25 in.) high. The liner, used in a magneto-hydrodynamic electrical power generator, contained plasma at 1650 to 1925 °C (3000 to 3500 °F). This application is shown in Table 8.

Example 5: Gas-Tungsten Arc Welding of High-Strength Beryllium Copper in a Cover-to-Housing Weldment.

For ocean-cable use, a cast beryllium copper cover made from alloy C17000 was welded to a housing of the same alloy. The cast cover was welded to an extruded housing to produce a joint that was watertight at 83 MPa (12 ksi). This application is shown in Table 8.

Gas-Metal Arc Welding

Alloys C17000 and C17200 are more easily welded using the GMAW process than is high-conductivity alloy C17500. These alloys have lower melting temperatures, greater fluidity, and 50% lower thermal conductivities. The GMAW process is generally preferred for the welding of precipitation-hardened, high-strength beryllium coppers in thicknesses of 6 mm (0.25 in.) and greater. This process is also preferred for thicknesses from 6 mm (0.25 in.) down to 2.5 mm (0.1 in.) if heat treatment is to be done after welding.

Nominal Conditions. Table 6 gives the nominal welding conditions for GMAW of high-strength beryllium coppers. The shielding gas used for GMAW is typically a mixture of argon and helium. Direct current electrode positive is used to prevent the buildup of oxides during welding. The filler metal is ordinarily of the same composition as the base metal, for joints requiring maximum strength. However, electrode wires ERCuSi-A and ERCuA1-A2 can be used when joint strength is less critical. Preheating, postweld heat treating, and joint designs are generally the same as discussed in the section on GTAW of high-strength beryllium coppers.

Welding of Copper Alloys

Revised by M. Ned Rogers, Batesville Casket Company

Welding of High-Conductivity Beryllium Copper

The traditional wrought high-conductivity alloys (C17500 and C17510) contain 0.2 to 0.7% Be and nominal 2.5% Co (or 2% Ni). The leanest and most recently developed high-conductivity alloy is C17410, which contains somewhat less than 0.4% Be and 0.6% Co. This section will describe the GTAW and GMAW of alloy C17500, the most frequently welded high-conductivity beryllium copper alloy. Additional information on the properties and applications of these materials can be found in Volume 2 of the *ASM Handbook.*

Gas-Tungsten Arc Welding

Gas-tungsten arc welding of precipitation-hardened copper alloy C17500 (high-conductivity beryllium copper) is performed for thicknesses up to 6 mm (0.25 in.). This process produces a narrow HAZ, which is desirable. Ordinarily, the maximum thickness that can be gas-tungsten arc welded without substantial decrease in strength is about 13 mm (0.5 in.). However, thicker sections have been gas-tungsten arc welded. For weldments that require postweld heat treatments, GTAW is generally used only for thicknesses up to 2.5 mm (0.1 in.), with GMAW being used for thicker sections.

High-conductivity beryllium copper is more difficult to weld than the high-strength beryllium coppers because of its higher thermal conductivity, which is about twice that of the high-strength beryllium coppers. A difficulty common to the beryllium coppers is the formation of surface oxides, both beryllium oxide and cuprous oxide. The beryllium oxide is a tenacious oxide that inhibits wetting and fusion during welding. Cleanliness of the joint faces and surrounding surfaces before and during welding is necessary to ensure weld soundness.

Nominal Conditions. Table 7 gives the nominal starting conditions for GTAW of high-conductivity beryllium copper alloy C17500. Shielding gas, type of current, and electrode type are the same as for C17000 and C17200 high-strength alloys.

Type of Current. Variation in arc length or welding speed during GTAW can result in the formation of tenacious films on beryllium coppers. For this reason, high-frequency stabilized alternating current is preferred in automatic welding and must be used in manual welding. The alternating current continually breaks up the oxide coating, and high-frequency adds to the cleaning action during welding. Direct current electrode negative can be employed in the automatic welding of beryllium coppers. The advantage is found with the high heat input to the work and the deep penetration of this polarity. Close control of the arc length and welding speed must be maintained to minimize oxide formation.

High-conductivity beryllium coppers are more susceptible to porosity and cracking than welds made in high-strength beryllium copper. This is particularly true in multi-pass welding. However, successfully welded joints exhibit improved mechanical properties in HAZ compared with similar joints in high-strength beryllium coppers.

Filler Metals. Material with the same composition as the base alloy (C17500) is generally used, since high electrical conductivity is desired in the welds. If maximum electrical conductivity through the weld joint is not a requirement, silicon bronze filler metal is satisfactory. Aluminum bronze filler metals can be used but are less satisfactory than silicon bronze.

Joint Design. As Table 7 indicates, the usual joint designs for GTAW of high-conductivity beryllium copper are square-groove or 90° single-V-groove butt joints with a 1.6 mm (0.06 in.) root face and zero root opening. All joints should be backed with grooved copper or graphite backing strips or rings.

Preheating. Preheating is not ordinarily needed for welding alloy C17500 up to about 3 mm (0.13 in.) thick. However, thicker stock that may require multipass welds is usually preheated to approximately 425 °C (800 °F).

Gas-Metal Arc Welding

Gas-metal arc welding is preferred to GTAW for joining alloy C17500 with section thicknesses greater than 2.5 mm (0.1 in.) when the weldment is to be heat treated to obtain maximum weld strength. It is also the process of choice for joining thicknesses in excess of 6 mm (0.25 in.) if the welding is done on precipitation-hardened material. Gas-metal arc welding can be used for thicknesses up to 19 mm (0.75 in.).

As pointed out in the discussion of GTAW of alloy C17500, the three important factors in arc welding this alloy are: (1) high thermal and electrical conductivities; (2) oxide-forming characteristics; and (3) responsiveness to heat treatment. Although compromise is necessary between weld strength and conductivity, high thermal conductivities are typically the prime objectives in welding assemblies using this alloy.

Nominal Conditions. Table 6 lists the typical conditions for GMAW of high-conductivity beryllium coppers. Typically, DCEP is used. Argon is a suitable shielding gas; however, greater heat input, which is usually required when welding this alloy, can be obtained through the use of an argon-helium gas mix.

Filler Metals. As with GTAW for high-conductivity beryllium coppers, GMAW utilizes a filler metal of similar composition to the base metal. If reduced conductivity is an acceptable result of the weld metal, wires made of high-strength beryllium copper alloys C17000 and C17200 can be used for greater welding ease. Because of the precipitation-hardening characteristics of beryllium coppers, the joint strength is always somewhat lower than that of the base material and depends on the initial condition of the base metal, the welding conditions, and the filler-metal selection.

Other electrodes, such as ERCuSi-A or ERCuAl-A2, can be used for GMAW of alloy C17500. Because of their lack of beryllium content, however, these alloys do not develop the high strength attained when beryllium copper filler metal is used.

Preheating and Postweld Aging. Section thicknesses of alloy C17500 greater than 3 mm (0.13 in.) are usually preheated at 315 to 480 °C (600 to 900 °F), depending on section thickness. When a beryllium copper filler metal is used, strength can be increased by aging after welding. The aging heat treatment for alloy C17500 is 480 °C (900 °F) for 3 h.

Properties of Weldments. Based on its high thermal conductivity and moderate to high strength, alloy C17500 is used for welded water-cooled assemblies such as tuyeres for blast furnaces, attrition mills for grinding beryllium chips to powder, and molds for the continuous casting of steel. While some strength is lost in the weld metal and the adjoining HAZ when alloy C17500 is welded, the magnitude of this decrease is affected by the condition and thicknesses of the base metal, and by joint design, welding process, and parameters.

Welding of Copper Alloys

Revised by M. Ned Rogers, Batesville Casket Company

Welding of Cadmium and Chromium Coppers

Typically, the welding procedures for GTAW of coppers are a satisfactory basis for determining initial welding parameters for cadmium and chromium coppers. These alloys have lower thermal and electrical conductivities than copper, and they can be welded with lower preheats and heat inputs than those required for copper. Section thickness limitations are the same as for GTAW of coppers. In addition to GTAW, GMAW can be used for the joining of cadmium and chromium coppers. The nominal conditions previously stated for GMAW welding of coppers are satisfactory for the welding of cadmium and chromium coppers.

The filler metal types, welding currents, polarities, and shielding gas mixtures for either GTAW or GMAW of coppers are suitable for the joining of cadmium and chromium coppers. These alloys should be welded using the standard joint geometries listed for GTAW and GMAW of coppers.

Welding of Copper Alloys

Revised by M. Ned Rogers, Batesville Casket Company

Welding of Copper-Zinc Alloys

Brasses are basically copper-zinc alloys and are the most widely used group of copper alloys. As shown in Table 1, brasses can be categorized as low-zinc brasses (5 to 20% Zn), high-zinc brasses (30 to 40% Zn), tin brasses, and special brasses. Copper-zinc alloys are joined by GTAW, GMAW, and SMAW. Additional information on the properties and applications of copper-zinc (brass) alloys can be found in Volume 2 of the *ASM Handbook.*

Gas-Tungsten Arc Welding

Table 1 lists the copper-zinc (brass) alloys and shows their relative weldability. The low-zinc brasses are shown to have good weldability using GTAW. High-zinc brasses, tin brasses, special brasses, and nickel silvers have only fair weldability. This is because of the high zinc content or the moderate zinc content in combination with other elements, such as oxide-forming aluminum or nickel.

Gas-tungsten arc welding is sometimes used for the welding of copper-zinc alloys with 20% Zn or less. Alloys containing up to 1% Pb can also be welded with this process. Gas-tungsten arc welding is used because of its ability to weld rapidly with a highly localized heat input. It should be noted that most leaded copper alloys are generally not recommended for arc welding.

Section thicknesses of copper-zinc alloys ordinarily welded using GTAW are limited to about 9.5 mm (0.4 in.). However, thick sections of cast alloys such as manganese bronze are sometimes repair welded in localized areas. Preheating is not normally required for the welding of these alloys.

Shielding Gas. Selection of the proper shielding gas for GTAW of copper-zinc alloys is influenced by the heat required to fuse the metals together. This requirement is directly related to the thermal conductivity of the base metal. Normally argon is used for GTAW of the less conductive materials. Helium or helium mixes are the preferred shielding gases for alloys having higher thermal conductivities. However, the use of helium and helium-rich gas mixes can reduce the fumes produced when welding the less conductive alloys, such as high-zinc brasses, tin brasses, special brasses, and nickel silvers.

Filler Metals. Filler metals used in GTAW of copper-zinc alloys should not contain zinc. The addition of filler metal is recommended when welding sections greater than 1.6 mm (0.06 in.) thick. The arc is concentrated on the filler metal, thus reducing the zinc loss and fuming of the base material. For the welding of low-zinc brass, ERCuSn-A is recommended; ERCuSi-A and ERCuAl-2 are recommended for high-zinc alloys.

The silicon found in ERCuSi-A helps to decrease the zinc fumes. For this reason and to provide reinforcement, ERCuSi-A is sometimes used, with alternating current, on copper-zinc alloys 1.3 mm (0.05 in.) or less. Thinner sections of high-zinc alloys are sometimes gas-tungsten arc welded using ERCuAl-A2. The addition of this filler metal produces sound welds but does not decrease zinc vaporization.

Thin brass sheets less than 1.6 mm (0.06 in.) thick can be welded without the addition of filler metal. Zinc fuming can be limited through the use of high electrode travel speed, which generates short weld pool arc times. Tube sheet joints can be made with brass members, using the same kind of joint preparation, if high electrode speed is maintained.

Gas-Metal Arc Welding

Unleaded brasses can be welded using GMAW. These brasses include both the low-zinc alloys (red brasses) and the high-zinc alloys (including yellow brasses, tin brasses, and special brasses). Copper-zinc alloys are not used as filler metals for GMAW due to the violent fuming and loss of zinc that accompanies arc welding of these zinc-bearing alloys.

Low-Zinc Brasses. Table 6 lists nominal conditions for butt welding of low-zinc brasses, using GMAW. These alloys include up to 20% Zn. Direct current electrode positive is always used.

Silicon bronze (ERCuSi-A) wire provides easy welding, since it has good fluidity at low welding currents. A 60° single-V-groove is used with this filler metal. When ERCuSn-A electrode wire is used, its sluggish flow characteristics make a 70° V-groove advisable for heavier sections, as shown in Table 6. ERCuSn-A is primarily used for a better color match on low-zinc copper alloys. The wider groove provided with the 70° V-groove joint allows more room for manipulation of the molten weld pool. Except for color, weld metal properties are comparable when using either of these two types of electrodes. A preheat in the range of 95 to 315 °C (200 to 600 °F) is sometimes required for low-zinc brasses due to their relatively high thermal conductivities.

High-Zinc Copper Alloys. Nonleaded copper alloys that have zinc contents from 20 to 40% or more can be gas-metal arc welded. However, the weldability is less than with nonleaded low-zinc brasses. The evolution of zinc fumes is more severe, and the welds experience greater porosity and lower strength than with low-zinc brasses. Both wrought and cast alloys can be joined by GMAW. Massive sections, such as for manganese bronze ship propellers, are regularly repair welded using this process.

The nominal conditions for gas-metal arc butt welding of high-zinc copper alloys (yellow brasses, tin brasses, special brasses, and nickel silvers) are given in Table 6. Welding parameters and variables are generally the same when using either ERCuAl-A2 or ERCuSn-A filler metals. The difference between the two filler metals is a higher weld-metal strength with ERCuAl-A2 and better color match with ERCuSn-A. Preheating is seldom necessary since these alloys have relatively low thermal conductivities. However, preheating does help to limit zinc fuming in some applications, because it permits the use of lower welding amperages.

Shielded Metal Arc Welding

Brasses can be welded using SMAW. A variety of covered electrodes can be used, including phosphor bronze, silicon bronze, or aluminum bronze (ECuSi, ECuSn-A, ECuSn-C, ECuAl-A2, ECuAl-B). Relatively large welding grooves are required for good joint penetration and to avoid the entrapment of slag. Welding is typically done using a backing strip of copper or brass.

Phosphor bronze electrodes, such as ECu-Sn-A and ECuSn-C, have been used for SMAW of low-zinc brasses. Preheating of the base metal from 200 to 260 °C (400 to 500 °F) is required. Application of filler metal should be in narrow, shallow stringer beads.

The high-zinc copper alloys can be welded with aluminum bronze (ECuAl-A2) electrodes. Preheat and interpass temperatures are 260 to 370 °C (500 to 700 °F). The arc is held directly on the molten weld pool rather than toward the base metal, and it is advanced slowly to minimize zinc volatilization and fuming.

Welding of Copper Alloys

Revised by M. Ned Rogers, Batesville Casket Company

Welding of Phosphor Bronzes

The phosphor bronzes (C50100-C52400) contain additions of tin (up to 10%) and phosphorus (0.2%), which produce good resistance to flowing seawater and to most nonoxidizing acids (except hydrochloric acid). Phosphor bronzes are joined by GTAW, GMAW, and SMAW. Additional information on the properties and applications of phosphor bronzes can be found in Volume 2 of the *ASM Handbook.*

Gas-Tungsten Arc Welding

Sheet and other wrought forms of phosphor bronze can be gas-tungsten arc welded up to approximately 13 mm (0.5 in.). The process can also be used to join or repair phosphor bronze castings. Copper-tin alloys solidify with large, weak, dendritic grain structures that are susceptible to cracking. Hot peening each layer of multi-pass welds reduces cracking and stresses and therefore the likelihood of weld metal cracking.

Nominal Conditions. The use of GTAW for joining nonleaded phosphor bronzes is described in Table 9. Square-groove butt joints are employed, and the current used is DCEN or a stabilized alternating current arc.

TABLE 9 NOMINAL CONDITIONS FOR GTAW OF SQUARE-GROOVE BUTT JOINTS IN PHOSPHOR BRONZES

WORKPIECE THICKNESS		**ROOT OPENING**		**ELECTRODE (EWTH-2) DIAMETER**		**WELDING ROD DIAMETER**[A]		**SHIELDING GAS**	**GAS FLOW RATE**		**CURRENT (DCEN), A**	**TRAVEL SPEED**		**PREHEAT TEMPERATURE**		**NUMBER OF PASSES**
MM	**IN.**	**MM**	**IN.**	**MM**	**IN.**	**MM**	**IN.**		**L/MIN**	**FT^3/H**		**M/MIN**	**IN./MIN**	**°C**	**°F**	
1.6	$\frac{1}{16}$	0	0	3.2	$\frac{1}{8}$	NONE USED		ARGON	15	30	90-150	1.78	70	NONE USED		1
3.2	$\frac{1}{8}$	1.6	$\frac{1}{16}$	3.2	$\frac{1}{8}$	3.2	$\frac{1}{8}$	ARGON	15	30	100-220	1.27	50	NONE USED		1

Notes: (1) The data in this table are intended to serve as starting points for the establishment of optimum joint design and conditions for welding of parts on which previous experience is lacking. They are subject to adjustments necessary to meet the special requirements of individual applications. (2) Higher welding current or lower welding travel speed is used in welding alloy C50500.

(A) ERCUSN-A, OR ROD OF COMPOSITION CLOSE TO THAT OF THE BASE METAL

The nominal conditions shown in Table 9 are based primarily on the three phosphor bronzes given in Table 1, which have low thermal conductivities. These include alloys C51000, C52100, and C52400. When welding alloy C50500, higher welding currents or slower welding travel speeds are required. This alloy, commonly known as phosphor bronze-1.25% E, contains approximately 98.7% Cu and has 3 to 4 times the thermal conductivity of the other alloys in this grouping.

Shielding Gas. The use of argon as a shielding gas is most common, since argon restricts the size of the HAZ. This minimizes the area where mechanical properties are decreased. When necessary for thicker sections, helium shielding gas can be used, along with an argon backup gas shield.

Filler metal may be added to the weld joint using either ERCuSn-A or wire of approximately the same composition as the base metal. When matching the composition of the base metal is not essential, ERCuSi-A filler metal can be used instead, producing stronger welds. When welding section thicknesses less than 3 mm (0.13 in.), phosphor bronze strip can be butt welded using autogenous GTAW.

Preheating is not normally required for GTAW of thin sections of phosphor bronzes. Thick sections usually require preheating to 175 or 200 °C (350 or 400 °F) and are slowly cooled after welding. Due to the hot-short nature of phosphor bronze, weld layers should be thin, and interpass temperature should not exceed 200 °C (400 °F). Hot peening is often used to refine grain size and minimize distortion.

Gas-Metal Arc Welding

Table 6 gives the nominal conditions for GMAW of butt joints using phosphor bronze. Lead-bearing or other free-machining types are not welded. Table 10 gives GMAW conditions that can be used as starting points for establishing welding conditions for joining phosphor bronzes. As with GTAW, the welding conditions are based on three poorly conductive phosphor bronze alloys (C51000, C52100, and C52400). When welding the more conductive alloy C50500, higher welding currents or slower welding speeds are needed. For joining thicknesses of 9.5 to 13 mm (0.4 to 0.5 in.), 90° single-V grooves are used rather than the narrow grooves used for most other poorly conductive alloys.

TABLE 10 SUGGESTED CONDITIONS FOR GMAW OF PHOSPHOR BRONZES

WORKPIECE THICKNESS		JOINT DESIGN			ELECTRODE DIAMETER[(A)]		ARC VOLTAGE, V	WELDING CURRENT (DCEP), A[(A)]
		GROOVE TYPE	ROOT OPENING					
MM	IN.		MM	IN.	MM	IN.		
1.6	$\frac{1}{16}$	SQUARE	1.3	0.05	0.8	0.030	25-26	130-140
3.2	$\frac{1}{8}$	SQUARE	2.4	0.09	0.9	0.035	26-27	140-160
6.4	$\frac{1}{4}$	V	1.6	0.06	1.2	0.045	27-28	165-185
12.7	$\frac{1}{2}$	V	2.4	0.09	1.6	0.062	29-30	315-335
19.0	$\frac{3}{4}$	DOUBLE-V OR DOUBLE-U	2.4	0.09	2.0	0.078	31-32	365-385
25.4	1.0	DOUBLE-V OR DOUBLE-U	2.4	0.09	2.4	0.094	33-34	440-460

(A) ERCUSN-A PHOSPHOR BRONZE ELECTRODES AND ARGON SHIELDING GAS

Filler Metals. For joining phosphor bronzes, ERCuSn-A electrode wire is generally used. The filler metal wire that contains about 0.5% Si is sometimes used to minimize porosity in the weld.

Preheating. As with other copper alloys, preheating of the phosphor bronzes helps in obtaining complete fusion. Also, porosity is minimized because the freezing rate of the weld pool is decreased, and more gas is permitted to evolve before solidification. However, preheating increases the susceptibility of the weld to the formation of large columnar grains and the risk of hot cracking. This susceptibility requires that welding be performed with a stringer bead technique and that each weld layer be peened. A small weld pool and rapid electrode travel speed are required. Interpass temperatures while welding should not exceed the preheat temperatures because of the hot shortness of the phosphor bronze alloys.

Shielded Metal Arc Welding

Shielded metal arc welding is done on a limited basis for phosphor bronzes. The covered electrodes typically used are ECuSn-A and ECu-Sn-C. These electrodes may be used interchangeably and are designed for use with DCEP.

Preheating is required in the range of 150 to 200 °C (300 to 400 °F), since the phosphor bronzes are very sluggish. However, because of the hot shortness of these alloys, the interpass temperature must not exceed the preheat temperature. This is extremely important in thicker sections. Stringer beads with rapid travel speed are normally used. In groove welding, the first two passes are made with a weaving technique. The width of the weave should not exceed two electrode diameters. The remaining passes are made without appreciable traverse weaving and with the use of narrow stringer beads. The development of the coarse, dendritic structure with low strength and ductility is minimized by the control of preheat and interpass temperatures. Hot peening is often used after welding to reduce grain size. For the maximum ductility, the welded assembly should be postweld heat treated to 480 °C (900 °F) and cooled rapidly. Groove joints should be wide, with an included angle of 80 to 90° to achieve proper washing and tie-in at the walls of the groove.

Welding of Copper Alloys

Revised by M. Ned Rogers, Batesville Casket Company

Welding of Aluminum Bronzes

As shown in Table 1, aluminum bronzes (C60600-C64400) are complex alloys based on the Cu-Al-Ni-Fe-Si-Sn system. They are welded using GTAW, GMAW, or SMAW. Some aluminum bronzes are also used as hardfacing alloys (see the article "Hardfacing, Weld Cladding, and Dissimilar Metal Joining" in this Volume). Additional information on the properties and applications of aluminum bronzes can be found in Volume 2 of the *ASM Handbook.*

Gas-Tungsten Arc Welding

Gas-tungsten arc welding is readily used for joining aluminum bronzes up to 9.5 mm (0.4 in.). Aluminum bronze castings are also repair welded by GTAW. Conditions differ somewhat from those of most other copper alloys. Porosity is minimized in the weld metal by the presence of iron, manganese, or nickel in the filler metal or base metal.

Nominal Conditions. Table 11 lists the nominal conditions for welding of aluminum bronze alloys. Conditions must be selected to avoid difficulties that result from the tenacious refractory aluminum oxide films that form almost instantaneously during any heating process such as welding unless oxygen is completely excluded from the metal. Heat input requirements are not high, since aluminum bronzes have a thermal conductivity near that of carbon steel.

TABLE 11 NOMINAL CONDITIONS FOR GTAW OF ALUMINUM BRONZES

WORKPIECE THICKNESS		ROOT OPENING		ELECTRODE DIAMETER(B)		WELDING ROD DIAMETER(C)		FLOW RATE OF ARGON		CURRENT (AC, HF-STABILIZED), A(D)	NUMBER OF PASSES
mm	in.	mm	in.	mm	in.	mm	in.	l/min	ft^3/h		
SQUARE-GROOVE BUTT JOINTS											
0-1.6	0-$\frac{1}{16}$	0	0	1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$ (E)	9.5-15	20-30	25-80	1
1.6-3.2	$\frac{1}{16}$-$\frac{1}{8}$	1.6 MAX	$\frac{1}{16}$ MAX	2.4	$\frac{3}{32}$	3.2	$\frac{1}{8}$	9.5-15	20-30	60-175	1

3.2	$\frac{1}{8}$	3.2 MAX	$\frac{1}{8}$ MAX	4.8	$\frac{3}{16}$	4.0	$\frac{5}{32}$	15	30	210	1
70° SINGLE-V-GROOVE BUTT JOINTS											
9.5	$\frac{3}{8}$	0	0	4.8	$\frac{3}{16}$	4.0	$\frac{5}{32}$	15	30	210-330	4
FILLET-WELDED T-JOINTS OR SQUARE-GROOVE INSIDE CORNER JOINTS											
9.5	$\frac{3}{8}$	(F)		4.8	$\frac{3}{16}$	4.0	$\frac{5}{32}$	15	30	225	3

Note: The data in this table are intended to serve as starting points for the establishment of optimum joint design and conditions for welding of parts on which previous experience is lacking. They are subject to adjustments necessary to meet the special requirements of individual applications.

(A) PREHEATING IS NOT ORDINARILY USED IN WELDING THE THICKNESSES SHOWN.

(B) ZIRCONIATED OR UNALLOYED ELECTRODES ARE RECOMMENDED WITH HIGH-FREQUENCY-STABILIZED ALTERNATING CURRENT.

(C) PREFERRED WELDING ROD IS ERCUAL-A2; OTHERWISE ERCUAL-A3 OR ROD OF THE SAME COMPOSITION AS THE BASE METAL.

(D) DIRECT CURRENT ELECTRODE NEGATIVE CAN ALSO BE USED IN MAKING SINGLE-PASS WELDS (SEE TEXT).

(E) USE OF WELDING ROD IS OPTIONAL FOR THICKNESSES UP TO 1.6 MM ($\frac{1}{16}$ IN.).

(F) ZERO ROOT OPENING FOR T-JOINTS; 9.5 MM ($\frac{3}{8}$ IN.) MAX FOR CORNER JOINTS

Shielding Gas. Argon is typically used with alternating current as a shielding gas for the welding of aluminum bronzes. For better penetration or faster travel speed, direct current can be used with argon, helium, or a mixture of these two gases. While the shielding gas protects the weld pool and adjacent base metal, the use of a special flux augments the shielding effect of the gas. The flux is applied to the edge of the weld joints to increase fluidity and help protect the base metal from oxide formation. Aluminum oxide forms even at room temperature, and the flux prevents air from reaching the prepared edges until the protective argon shielding gas can become effective in the localized area.

Type of Current. As stated above, alternating current is used for GTAW of aluminum bronzes. The alternating current is stabilized by continuous high frequency and is preferred to DCEN, since it prevents oxide formation on the surface of the weld pool. High-frequency-stabilized alternating current is particularly desirable when multi-pass welding of these alloys is performed. Direct current electrode negative can be used in a single-pass welding, particularly for automatic welding. Surfaces must be well cleaned and protected, and the arc must be closely controlled.

Electrodes. Gas-tungsten arc welding of aluminum bronzes typically uses zirconiated tungsten (EWZr) or unalloyed tungsten (EWP) electrodes. Thoriated tungsten electrodes are used with direct current; however, these electrodes cause the arc to wander when they are used with alternating current. The thoriated tungsten gives a longer tip life and is easier to start. Electrode preparation for use with direct current is a tapered cone-shaped tip. For alternating current, the end of the electrode is typically balled via an arc melting process.

Filler Metals. ERCuAl-A2 is ordinarily used for joining aluminum bronzes. ERCuAl-A3 or other aluminum bronze wire of matching chemical composition can be used when the composition and color of the weld metal must closely match that of the base metal.

Preheating of aluminum bronzes for GTAW of sections thinner than approximately 6 mm (0.25 in.) is not necessary. Aluminum bronzes have a low thermal conductivity but may require preheating for section thicknesses greater then 6 mm (0.25 in.). When preheating is required, the preheat and interpass temperature should not exceed 150 °C (300 °F) for alloys with less than 10% Al. In the following example, preheating was used for joining of a 9.5 mm (0.4 in.) thick aluminum bronze.

Example 6: Use of GTAW versus GMAW to Avoid Melt-Through of Neighboring Sections.

The propeller housing for the Kort nozzle shown in Fig. 3 was originally made as a one-piece sand casting from aluminum bronze alloy 9B (ASTM B148), which contains 10% Al and 1% Fe. Difficulties were experienced in the coring necessary to produce the thin outer shell. This led to the production of this outer shell as a separate piece that was welded to the casting to make a two-piece structure.

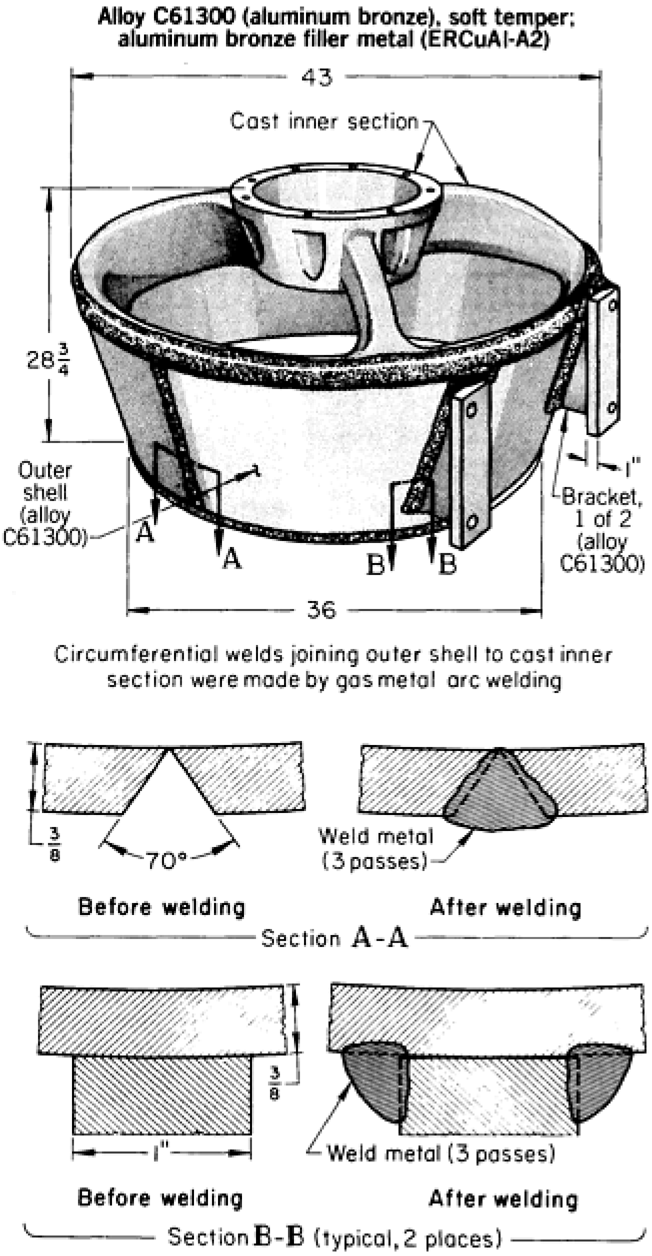
Alloy C61300 (aluminum bronze), soft temper;
aluminum bronze filler metal (ERCuAl-A2)
43
Cast inner section
$28\frac{3}{4}$
Outer shell (alloy C61300)
A
A
B
B
1"
Bracket, 1 of 2 (alloy C61300)
36
Circumferential welds joining outer shell to cast inner section were made by gas metal arc welding
$\frac{3}{8}$
70°
Weld metal (3 passes)
Before welding
After welding
Section A-A
$\frac{3}{8}$
1"
Weld metal (3 passes)
Before welding
After welding
Section B-B (typical, 2 places)

Erro! •

CONDITIONS FOR GTAW	
JOINT TYPES	BUTT AND T
WELD TYPES	SINGLE-V-GROOVE AND FILLET
WELDING POSITION	FLAT
NUMBER OF PASSES	3
POWER SUPPLY	300 A RECTIFIER, CONSTANT-CURRENT TYPE
TORCH	WATER-COOLED
ELECTRODE	2.4 MM ($\frac{3}{32}$ IN.) DIAM EWTH-2

FILLER METAL	2.4 MM ($\frac{3}{32}$ IN.) DIAM ERCUAL-A2
SHIELDING GAS	ARGON; FLOW RATE, 9.5 L/MIN (20 FT^3/H)
PREHEAT AND INTERPASS TEMPERATURE	150 °C (300 °F)
VOLTAGE	35 V
CURRENT	225 A (DCEN)
ARC STARTING	CONTACT AND HIGH FREQUENCY

FIG. 3 CONDITIONS FOR GTAW OF A PROPELLER HOUSING. THE WELDMENT REPLACED A SINGLE-PIECE CASTING, THEREBY REDUCING COST AND IMPROVING SERVICEABILITY.

The inner section was cast in propeller bronze as one intricate piece that included a central hub and shaft-mounting flange connected to the main body of the nozzle by four arms. The outer shell was cut from 9.5 mm (0.4 in.) thick alloy C61300 (aluminum bronze) with soft temper and was wrapped around the cast intersection after being cold formed into a truncated cone. Before cold forming, all edges of the shell were machined to a 35° bevel to provide a groove for subsequent welding.

The assembly was tack welded together, and then circumferential seams at the top and bottom were gas-metal arc welded. Next, to avoid melt-through of the cast intersection (and resultant poor effect on the propeller stream), GTAW was used for the longitudinal seam in the shell and for joining two brackets to the shell. The brackets were used for mounting the shell to the hull of the vessel. Figure 3 illustrates the weldment as constructed.

Not only was the cost of production reduced by eliminating the complex and expensive coring in the casting, but the weight of the welded housing was less than that of the completed cast housing. Furthermore, the cast housing had been made of an alloy that, although easy to cast, did not resist erosion, corrosion, or cavitation as well as did the alloys used in the welded assembly.

Gas-Metal Arc Welding

Gas-metal arc welding with aluminum bronze electrode wire is the preferred technique for welding aluminum bronzes. Because of the comparatively high surface tension of the molten weld pool and the relatively low thermal conductivity of the base metal, welding can be performed in all positions. Welding in the vertical and overhead positions is usually performed with either the globular or short-circuiting mode of metal transfer, using filler metal up to 1.6 mm ($\frac{1}{16}$ in.) in diameter.

Nominal Conditions. The nominal conditions for gas-metal arc butt welding of aluminum bronze are given in Table 6. Backing strips or rings of aluminum bronze are used when welding material up to 3 mm (0.13 in.) thick. These conditions make it necessary to use helium or argon-helium mixes for shielding gas instead of the argon shown in Table 6. These hotter shielding gases provide adequate heat input.

Filler Metals. ERCuAl-A2 ordinarily is used in the GMAW of aluminum bronze. The 1.5% Fe present in this wire reduces the hot-short cracking susceptibility of the weld metal. The use of ERCuAl-A3 also produces welds that are free from hot shortness. Joints made with ERCuAl-A3 are also stronger and harder, but lower in ductility.

Joint Design. The joining of aluminum bronzes thicker than 13 mm (0.5 in.) requires wider root openings and groove angles than those indicated in Table 6. This is necessary to avoid incomplete fusion and to improve the penetration resulting from the use of a thicker section, which increases the weld heat-sink characteristics.

Preheating. For most thin sections of aluminum bronzes, preheating is not required prior to welding. If preheat is used, the preheat and interpass temperatures should not exceed 150 °C (300 °F) for alloys that contain less than 10% Al. When aluminum content exceeds 10%, preheat temperatures should be elevated to 260 °C (500 °F). Again, interpass temperatures should not exceed preheat temperatures, due to the hot shortness of these alloys.

Shielded Metal Arc Welding

Aluminum bronzes are readily welded using SMAW for both the wrought and cast forms. The fluxing action of the electrode coating removes the aluminum oxides that form on the surface of the alloys.

Except for thin sections, a 70 to 90° V-groove joint is used. Typically, a backing strip of the same composition as the base metal is also utilized. Deposition technique and bead thickness are not critical, because the weld metal has excellent hot strength and ductility.

Electrodes. Aluminum bronze electrodes ECuAl-A2 and ECuAl-B are used for SMAW of aluminum bronze alloys C61300 and C61400. Preheating of thick sections of these alloys may be needed and is usually in the range of 200 °C (400 °F) with a controlled interpass temperature not exceeding 200 °C (400 °F). Depending upon section thickness and overall mass, preheat and interpass temperatures may vary from 65 to 425 °C (150 to 800 °F). Weldments made of aluminum bronze alloys containing 7% Al need not be heat treated after welding. Aluminum bronzes having an aluminum content greater than 7% are usually welded with electrodes that contain more aluminum than do ECuAl-A2 and ECuAl-B. These higher-aluminum bronze electrodes, ECuAl-C, ECuAl-D, and ECuAl-E, are best known as surfacing electrodes and have nominal aluminum contents of 12.5, 13.5, and 14.5%, respectively. These electrodes are described in AWS A5.13 and have strength corresponding to the aluminum content. In welding high-aluminum bronzes, thick sections may require preheating up to 620 °C (1150 °F), and fan cooling may be necessary to avoid cracking. These alloys may also require annealing at 620 °C (1150 °F), followed by fan cooling for stress relief.

Welding of Copper Alloys

Revised by M. Ned Rogers, Batesville Casket Company

Welding of Silicon Bronzes

Silicon bronzes (C64700-C66100), which are based on the Cu-Si-Sn system, are the most weldable of the copper alloys. Silicon bronzes can be welded by most arc welding processes, the most commonly utilized being GTAW, GMAW, and SMAW. Additional information on the properties and applications of silicon bronzes can be found in Volume 2 of the *ASM Handbook.*

Gas-Tungsten Arc Welding

Silicon bronze alloys that do not contain lead can be gas-tungsten arc welded in thin and moderately thick sections. These alloys are typically the most weldable of all copper alloys. Inherent characteristics of these bronzes that contribute to their weldability include low thermal conductivity, good deoxidation of the weld metal by silicon, and protection offered by the resulting slag. Silicon bronzes have a relatively narrow hot-short range just below the solidus, and they must be rapidly cooled through this range to avoid weld cracking.

Nominal Conditions. Table 12 lists the nominal welding conditions for GTAW of silicon bronzes. The data for manual welding include welding conditions for various positions. Because of their low fluidity, silicon bronzes are the only group of copper alloys where GTAW is applied extensively in the vertical and overhead positions. Direct current electrode negative with argon or helium shielding can be used for section thicknesses less than 1.6 mm (0.06 in.). Zirconiated (EWZr) or pure tungsten (EWP) electrodes are recommended.

TABLE 12 NOMINAL CONDITIONS FOR GTAW OF SILICON BRONZES USING ZERO ROOT OPENING, NO PREHEAT, EWTH-2 ELECTRODES, ERCUSI-A WELDING ROD, AND ARGON SHIELDING GAS

WORKPIECE THICKNESS(A)		ELECTRODE DIAMETER		WELDING ROD DIAMETER(B)		GAS FLOW RATE		CURRENT (DCEN), A	TRAVEL SPEED		NUMBER OF PASSES
mm	in.	mm	in.	mm	in.	l/min	ft^3/h		M/MIN	IN./MIN	
AUTOMATIC WELDING											
SQUARE-GROOVE BUTT JOINTS, FLAT POSITION											
0.3-1.3	0.012-0.050	3.2	$\frac{1}{8}$	NONE USED		7.1-16.5	15-35	80-140	1.52-2.03	60-80	1

1.6-3.2	$\frac{1}{16}$ - $\frac{1}{8}$	3.2	$\frac{1}{8}$	NONE USED		7.1-16.5	15-35	90-210	1.14-1.52	45-60	1
3.2	$\frac{1}{8}$	3.2	$\frac{1}{8}$	1.6	$\frac{1}{16}$	7.1-16.5	15-35	250	0.46-0.50	18-20	1
MANUAL WELDING											
SQUARE-GROOVE BUTT JOINTS, FLAT POSITION											
1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$	7.1	15	100-120	0.30	12	1
3.2	$\frac{1}{8}$	1.6	$\frac{1}{16}$	2.4	$\frac{3}{32}$	7.1	15	130-150	0.30	12	1
4.8	$\frac{3}{16}$	2.4	$\frac{3}{32}$	3.2	$\frac{1}{8}$	9.5	20	150-200	. . .	. . .	1
6.4	$\frac{1}{4}$	3.2	$\frac{1}{8}$	3.2	$\frac{1}{8}$	9.5	20	250-300	. . .	. . .	1
6.4	$\frac{1}{4}$	3.2	$\frac{1}{8}$	3.2	$\frac{1}{8}$	9.5	20	150-200	. . .	. . .	3
SQUARE-GROOVE BUTT JOINTS, VERTICAL AND OVERHEAD POSITIONS											
1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$	7.1	15	90-110	. . .	. . .	1
3.2	$\frac{1}{8}$	1.6	$\frac{1}{16}$	2.4	$\frac{3}{32}$	7.1	15	120-140	. . .	. . .	1
60° SINGLE-V-GROOVE BUTT JOINTS, FLAT POSITION											
9.5	$\frac{3}{8}$	3.2	$\frac{1}{8}$	3.2	$\frac{1}{8}$	9.5	20	230-280	. . .	. . .	3-4
12.7	$\frac{1}{2}$	3.2	$\frac{1}{8}$	4.8	$\frac{3}{16}$	9.5	20	250-300	. . .	. . .	4-5
19.0	$\frac{3}{4}$	3.2	$\frac{1}{8}$	4.8	$\frac{3}{16}$	9.5	20	300-350	. . .	. . .	9-10
25.4	1	3.2	$\frac{1}{8}$	6.4	$\frac{1}{4}$	9.5	20	300-350	. . .	. . .	13
FILLET-WELDED LAP JOINTS, FLAT POSITION											
1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$	7.1	15	110-130	0.25	10	1
3.2	$\frac{1}{8}$	2.4	$\frac{3}{32}$	2.4	$\frac{3}{32}$	7.1	15	140-160	0.25	10	1
4.8	$\frac{3}{16}$	2.4	$\frac{3}{32}$	3.2	$\frac{1}{8}$	9.5	20	175-225	. . .	. . .	1
6.4	$\frac{1}{4}$	2.4	$\frac{3}{32}$	3.2	$\frac{1}{8}$	9.5	20	175-225	. . .	. . .	3
9.5	$\frac{3}{8}$	3.2	$\frac{1}{8}$	4.8	$\frac{3}{16}$	9.5	20	250-300	. . .	. . .	3
12.7	$\frac{1}{2}$	3.2	$\frac{1}{8}$	4.8	$\frac{3}{16}$	9.5	20	275-325	. . .	. . .	6
19.0	$\frac{3}{4}$	3.2	$\frac{1}{8}$	4.8	$\frac{3}{16}$	9.5	20	300-350	. . .	. . .	12
25.4	1	3.2	$\frac{1}{8}$	6.4	$\frac{1}{4}$	9.5	20	325-350	. . .	. . .	16
FILLET-WELDED LAP JOINTS, VERTICAL AND OVERHEAD POSITIONS											
1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$	7.1	15	100-120	. . .	. . .	1
3.2	$\frac{1}{8}$	2.4	$\frac{3}{32}$	2.4	$\frac{3}{32}$	7.1	15	130-150	. . .	. . .	1
SQUARE-GROOVE OUTSIDE CORNER JOINTS, FLAT POSITION											

1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$	7.1	15	100-130	0.30	12	1
3.2	$\frac{1}{8}$	1.6	$\frac{1}{16}$	2.4	$\frac{3}{32}$	7.1	15	130-150	0.30	12	1
4.8	$\frac{3}{16}$	2.4	$\frac{3}{32}$	3.2	$\frac{1}{8}$	9.5	20	150-200	. . .	. . .	1
SQUARE-GROOVE OUTSIDE CORNER JOINTS, VERTICAL AND OVERHEAD POSITIONS											
1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$	7.1	15	90-110	. . .	. . .	1
3.2	$\frac{1}{8}$	1.6	$\frac{1}{16}$	2.4	$\frac{3}{32}$	7.1	15	120-140	. . .	. . .	1
50° SINGLE-BEVEL-GROOVE OUTSIDE CORNER JOINTS, FLAT POSITION(C)											
6.4	$\frac{1}{4}$	2.4	$\frac{3}{32}$	3.2	$\frac{1}{8}$	9.5	20	175-225	. . .	. . .	3
9.5	$\frac{3}{8}$	3.2	$\frac{1}{8}$	4.8	$\frac{3}{16}$	9.5	20	230-280	. . .	. . .	3
12.7	$\frac{1}{2}$	3.2	$\frac{1}{8}$	4.8	$\frac{3}{16}$	9.5	20	275-325	. . .	. . .	7
19.0	$\frac{3}{4}$	3.2	$\frac{1}{8}$	4.8	$\frac{3}{16}$	9.5	20	300-350	. . .	. . .	14
25.4	1	3.2	$\frac{1}{8}$	6.4	$\frac{1}{4}$	9.5	20	325-350	. . .	. . .	20
FILLET-WELDED SQUARE-GROOVE INSIDE CORNER JOINTS, FLAT POSITION(D)											
1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$	7.1	15	110-130	0.25	10	1
3.2	$\frac{1}{8}$	2.4	$\frac{3}{32}$	2.4	$\frac{3}{32}$ $\frac{3}{32}$	7.1	15	140-150	0.25	10	1
4.8	$\frac{3}{16}$	2.4	$\frac{3}{32}$	3.2	$\frac{1}{8}$	9.5	20	175-225	. . .	. . .	1
FILLET-WELDED T-JOINTS, FLAT POSITION											
1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$	1.6	$\frac{1}{16}$	7.1	15	110-130	0.25	10	1
3.2	$\frac{1}{8}$	2.4	$\frac{3}{32}$	2.4	$\frac{3}{32}$	7.1	15	140-160	0.25	10	1
4.8	$\frac{3}{16}$	2.4	$\frac{3}{32}$	3.2	$\frac{1}{8}$	9.5	20	175-225	. . .	. . .	1
6.4	$\frac{1}{4}$	2.4	$\frac{3}{32}$	3.2	$\frac{1}{8}$	9.5	20	175-225	. . .	. . .	3
9.5	$\frac{3}{8}$	3.2	$\frac{1}{8}$	4.8	$\frac{3}{16}$	9.5	20	230-280	. . .	. . .	3
12.7	$\frac{1}{2}$	3.2	$\frac{1}{8}$	4.8	$\frac{3}{16}$	9.5	20	275-325	. . .	. . .	7
19.0	$\frac{3}{4}$	3.2	$\frac{1}{8}$	4.8	$\frac{3}{16}$	9.5	20	300-350	. . .	. . .	14
25.4	1	3.2	$\frac{1}{8}$	6.4	$\frac{1}{4}$	9.5	20	325-350	. . .	. . .	20

Note: The data in this table are intended to serve as starting points for the establishment of optimum joint design and conditions for welding of parts on which previous experience is lacking. They are subject to adjustments necessary to meet the special requirements of individual applications.

(A) THICKNESSES GREATER THAN ABOUT 12.7 MM ($\frac{1}{2}$ IN.) ARE GAS-TUNGSTEN ARC WELDED ONLY WHEN IT IS NOT PRACTICAL TO USE GMAW.

(B) WIRE FEED RATE OF 2.9-3.2 M/MIN (115-125 IN./MIN).

(C) ROOT FACE IS 1.6 MM ($\frac{1}{16}$ IN.) FOR THICKNESSES OF 12.6 MM ($\frac{1}{2}$ IN.) OR LESS, AND 3.2 MM ($\frac{1}{8}$ IN.) FOR THICKNESSES GREATER THAN 12.7 MM ($\frac{1}{2}$ IN.)

(D) MAXIMUM ROOT OPENING IS EQUAL TO BASE METAL THICKNESS, *T*.

Filler Metals. The conventional silicon bronze filler metal, ERCuSi-A, which is similar in composition to alloy C65500 (high-silicon bronze A), can be used to weld any of the silicon bronzes. Alloy C65500 is the most commonly used of all the silicon bronzes. Thin sections of this alloy are readily welded using GTAW without the addition of filler metal.

Joint Design. On metal thicknesses greater than 6 mm (0.25 in.) a V-groove with a 60° included angle is used. Butt joints in thin sections can be welded without special preparation.

Preheating. Silicon bronzes have the lowest thermal conductivity of all copper alloys. The thermal conductivity of these alloys is less than that of carbon steels. Further, these alloys exhibit hot shortness. Based on this, preheating can be harmful to the welding of these alloys, and interpass temperature on multi-pass welds should not exceed 95 °C (200 °F).

Gas-Metal Arc Welding

The nonleaded silicon bronzes are readily welded using GMAW. Nominal conditions for GMAW of silicon bronzes in various thicknesses and joint designs are given in Table 6. This process is used for section thicknesses greater than 6 mm (0.25 in.). Rapid travel speeds (minimum of 3.5 mm/s, or 8 in./min) should be used to avoid excessive heat buildup. As shown in Table 6, the welding current used for these alloys is slightly lower than that used for other copper alloys. Because of the low thermal conductivity of silicon bronzes (Table 1), the resulting low heat input to the weld is adequate for complete fusion and good penetration. Argon shielding gas is preferred. A thin layer of oxide forms on the weld metal after each pass and must be removed by wire brushing prior to the next weld pass.

All silicon bronze alloys can be gas-metal arc welded using ERCuSi-A filler metal. The chemical composition of this wire is similar to that of alloy C65500 (high-silicon bronze A), which is the most frequently used silicon bronze. These wires may include up to 1.5% Sn instead of, or in addition to, the 1.5% Mn specified as a maximum for alloy C65500.

Joint Preparation. Metal thicknesses of 6 to 19 mm (0.25 to 0.75 in.) should utilize a single 60° V-groove. For silicon bronze alloys of section thicknesses greater than 19 mm (0.75 in.), U-grooves or 60° double-V-groove joints can be used.

Preheating. As with GTAW of silicon bronze alloys, preheating is not required and should not be used with GMAW of silicon bronze. Interpass temperatures must be held below 95 °C (200 °F) to prevent weld cracking. Stress relieving of silicon bronze weldment is recommended to prevent stress-corrosion cracking failures.

Shielded Metal Arc Welding

Shielded metal arc welding of silicon bronze alloys is usually done using ECuAl-A2 aluminum bronze electrodes. Welding temperatures are easily attained, due to the low thermal conductivity of silicon bronzes. As with other arc welding processes, preheat and interpass temperatures should not exceed 95 °C (200 °F) to prevent hot-short cracking.

Groove dimensions are similar to those used for steel joints. Metal thicknesses up to 4 mm (0.16 in.) can be welded with square grooves. Thicker sections can be welded with a single-V-or double-V-groove of 60° included angle.

Weld properties in silicon bronzes joined using SMAW are usually substantially lower than those of welds made by the gas-shielded processes. Based on this, these welds may not meet code or design requirements for strength. Peening of welds reduces residual stress and minimizes weld distortion.

Welding of Copper Alloys

Revised by M. Ned Rogers, Batesville Casket Company

Welding of Copper-Nickel Alloys

Copper-nickel alloys (C70000-C79900) are the most corrosion-resistant copper-base alloys. Alloy C71500 (Cu-30Ni) has the best general resistance to aqueous corrosion, but alloy C70600 (Cu-10Ni) is often selected because it offers good resistance at lower cost. Copper-nickel alloys are readily welded by GTAW, GMAW, and SMAW. Additional information on the properties and applications of copper-nickel alloys can be found in Volume 2 of the *ASM Handbook.*

Gas-Tungsten Arc Welding

Gas-tungsten arc welding is the preferred process for joining copper-nickel alloys with section thicknesses up to 1.6 mm (0.06 in.). This process may also be used for thicker sections. Automatic welding produces satisfactory results in thin sections of tube sheets. Manual welding is normally used for sheet and plate thicknesses up to 6 mm (0.25 in.).

Nominal Conditions. Conditions by which copper-nickel alloys can be manually and automatically welded using GTAW are given in Table 13. These include butt joints with square- and single-V-groove preparation.

TABLE 13 NOMINAL CONDITIONS FOR GAS-TUNGSTEN ARC BUTT WELDING OF COPPER-NICKEL ALLOYS

WORKPIECE THICKNESS		**BUTT-JOINT GROOVE**	**ELECTRODE DIAMETER(A)**		**ERCUNI WELDING ROD DIAMETER(B)**		**FLOW RATE OF ARGON**		**CURRENT (DCEN), A**	**TRAVEL SPEED**		**NUMBER OF PASSES**
mm	**in.**		**mm**	**in.**	**mm**	**in.**	**l/min**	**ft^3/h**		**M/MIN**	**IN./MIN**	
AUTOMATIC WELDING OF ALLOY C70600 (COPPER-NICKEL, 10%)												
3.2	$\frac{1}{8}$	SQUARE	4.8	$\frac{3}{16}$	1.6	$\frac{1}{16}$	15	30	310-320	0.38-0.46	15-18	1
MANUAL WELDING OF ALLOY C70600 (COPPER-NICKEL, 10%)												
3.2	$\frac{1}{8}$	SQUARE	4.8	$\frac{3}{16}$	3.2	$\frac{1}{8}$	15	30	300-310	0.13	5	1
9.5	$\frac{3}{8}$	70-80° SINGLE-V	4.8	$\frac{3}{16}$	4.8	$\frac{3}{16}$	15	30	300-310(C)	0.15	6	2-4
MANUAL WELDING OF ALLOY C71500 (COPPER-NICKEL, 30%)												
3.2	$\frac{1}{8}$	SQUARE	4.8	$\frac{3}{16}$	3.2	$\frac{1}{8}$	15	30	270-290	0.13	5	1
9.5	$\frac{3}{8}$	70-80° SINGLE-V	4.8	$\frac{3}{16}$	4.0	$\frac{5}{32}$	15	30	270-290(C)	0.15	6	4

Notes: (1) The data in this table are intended to serve as starting points for the establishment of optimum joint design and conditions for welding of parts on which previous experience is lacking. They are subject to adjustments necessary to meet the special requirements of individual applications. (2) Root opening is zero. Preheating is not needed.

(A) PREFERRED ELECTRODE MATERIAL IS EWTH-2.

(B) FILLER METAL (ERCUNI) MUST BE USED ON ALL WELDED JOINTS (SEE TEXT).

(C) CURRENT SHOULD BE INCREASED IN EQUAL INCREMENTS WITH EACH PASS, UP TO A MAXIMUM OF ABOUT 375 A, WITH LARGER WELDING RODS.

Preferred conditions include the use of argon shielding gas, DCEN, and thoriated tungsten electrodes. However, these variables are not critical for copper-nickel, since these alloys have low heat conductivities. Helium may be substituted for argon as a shielding gas, but argon provides better arc control and stability. Alternating current can be employed for automatic welding if there is adequate control of arc length.

During GTAW of alloy C70600, slightly higher welding currents or slower travel speeds should be used, due to the higher thermal conductivity of this alloy. Preheating is not necessary, and backing strips or rings can be used. Backing strips or rings should be made of copper or copper-nickel alloy, not carbon, graphite, or steel.

Filler Metals. ERCuNi is used as a filler metal in GTAW of copper-nickel alloys. This filler metal contains 0.20 to 0.50% Ti to minimize porosity and the possibility of oxygen embrittlement in either the weld metal or the HAZ. Because the standard copper-nickel alloys do not contain titanium or a comparable deoxidizer, filler metal should be used even in welding thin sheets of copper-nickel. This will avoid the formation of porosity. However, special compositions of alloys C70600 and C71500 are available that do contain titanium. Thin sheets of these special alloys can be welded without the use of a filler metal. In multi-pass welding of copper-nickel alloys, the filler metal diameter and welding current may be increased with successive passes to provide efficient deposition. Autogenous welds can sometimes be made on sheet thicknesses up to 1.6 mm (0.06 in.), but porosity may be a problem due to the absence of deoxidizers in the base metal.

Gas-Metal Arc Welding

Gas-metal arc welding is the preferred welding process for non-leaded copper-nickel alloys thicker than approximately 1.6 mm (0.06 in.). Table 14 lists welding conditions for GMAW of copper-nickel alloy plate. The preferred welding position is the flat position. Argon is the preferred shielding gas; however, argon-helium mixes give better penetration on thick sections. Direct current electrode positive is recommended.

TABLE 14 SUGGESTED CONDITIONS FOR GMAW OF COPPER-NICKEL ALLOY PLATE

WORKPIECE THICKNESS		ELECTRODE FEED(A)		ARC VOLTAGE, V	WELDING CURRENT (DCEP), A(A)
mm	in.	m/min	in./min		
6.4	$\frac{1}{4}$	4.5-5.6	180-220	25-26	130-140
9.5	$\frac{3}{8}$	5.1-6.1	200-240	26-27	140-160
12.7	$\frac{1}{2}$	5.6-6.1	220-240	27-28	165-185
19.0	$\frac{3}{4}$	5.6-6.1	220-240	29-30	315-335
25.4	1.0	5.6-6.1	220-240	31-32	365-385
>25.4	1.0	6.1-6.6	240-260	33-34	440-460

(A) ERCUNI ELECTRODE, 1.6 MM ($\frac{1}{16}$ IN.), ARGON SHIELDING GAS

Nominal conditions for gas-metal arc butt welding of non-leaded copper-nickel alloys for various joint designs and thicknesses are given in Table 6. These alloys can be satisfactorily welded using either spray or short-circuiting transfer. Spray transfer is normally used for sections greater than 6 mm (0.25 in.). Pulsed spray transfer or short-circuiting transfer can be used to weld thin sections. These transfer methods provide better control of the molten weld pool when welding in the vertical and overhead positions. Selection of processing conditions is not critical in GMAW of these poorly heat-conductive alloys. Alloy C70600 is usually welded at a slightly higher current or slower travel speed than alloy C71500.

Filler Metals. ERCuNi electrode wire is the common choice for GMAW of copper-nickel alloys. This wire resembles alloy C71500 in composition, since it has a 70 to 30 copper-to-nickel ratio. ERCuNi contains 0.15 to 1.00% Ti, which serves as a deoxidizer to minimize porosity and prevent oxygen embrittlement. Titanium also improves the fluidity of the weld pool.

Joint Design. For GMAW of butt joints in non-leaded copper nickels 3 mm (0.13 in.) thick, the joint design usually includes a square-groove weld and grooved copper backing. This is indicated in Table 6. For thicknesses between 3 and 13 mm (0.13 and 0.5 in.), single-V-grooves of 60 to 80° and grooved copper backing are typically used. For section thicknesses greater than 13 mm (0.5 in.), double-V-groove or double-U-groove joints are used. As with GTAW of copper-nickel alloys, backing material should be made of copper or copper-nickel rather than carbon, graphite, or steel.

Preheating and Postheating. The copper-nickel alloys have thermal conductivity equal to or lower than that of low-carbon steel. Therefore, no preheating or postheating is needed. However, interpass temperatures should be maintained below 65 °C (150 °F).

Shielded Metal Arc Welding

Both wrought and cast forms of copper-nickel alloys can be welded using SMAW. The weldability of these alloys is similar to that of low-carbon steels, because the thermal conductivity of these alloys is quite similar to that of steel. Copper-nickel electrode ECuNi is used in welding the copper-nickel alloys C70600 and C71500 with DCEP.

The weld deposits ordinarily have a high center crown, and the slag is viscous when molten and quite adherent when cold. Therefore, special care is needed to ensure complete slag removal before complete solidification of the weld metal, to prevent slag entrapment when cleaning between passes. Welding in both the vertical and overhead positions using SMAW provides good results in these alloys, although best results are obtained in the flat position. This process is preferred in some applications where access to the joint is limited, as in the butt welding of copper-nickel pipe.

Welding of Copper Alloys

Revised by M. Ned Rogers, Batesville Casket Company

Welding of Dissimilar Metals

Copper and many copper alloys can be welded to other (dissimilar) copper alloys, to carbon and alloy steels, to stainless steels, and to nickel and nickel alloys using GTAW and GMAW. Tables 15 and 16 provide guidelines for joining copper and copper alloys to dissimilar metals.

TABLE 15 ELECTRODES AND PREHEAT AND INTERPASS TEMPERATURES USED IN GTAW OF COPPER AND COPPER ALLOYS TO DISSIMILAR METALS

METAL TO BE WELDED	ELECTRODES (WITH PREHEAT AND INTERPASS TEMPERATURES) FOR WELDING METAL IN COLUMN 1 TO:				
	COPPERS	PHOSPHOR BRONZES	ALUMINUM BRONZES	SILICON BRONZES	COPPER-NICKEL ALLOYS
COPPER ALLOYS					
LOW-ZINC BRASSES	ERCUSN-A OR ERCU (540 °C, OR 1000 °F)	. . .	. . .	. . .	. . .
PHOSPHOR BRONZES	ERCUSN-A OR ERCU (540 °C, OR 1000 °F)	. . .	. . .	. . .	. . .
ALUMINUM BRONZES	ERCUAL-A2 (540 °C, OR 1000 °F)	ERCUAL-A2 OR ERCUSN-A (200 °C, OR 400 °F)	. . .	. . .	. . .
SILICON BRONZES	ERCUSN-A OR ERCU (540 °C, OR 1000	ERCUSI-A (65 °C MAX, OR 150 °F	ERCUAL-A2 (65 °C MAX, OR 150 °F MAX)	. . .	. . .

	°F)	MAX)			
COPPER-NICKEL ALLOYS	ERCUAL-A2 OR ERCUNI OR ERCU (540 °C, OR 1000 °F)	ERCUSN-A (65 °C MAX, OR 150 °F MAX)	ERCUAL-A2 (65 °C MAX, OR 150 °F MAX)	ERCUAL-A2 (65 °C MAX, OR 150 °F MAX)	. . .
NICKEL ALLOYS					
NICKEL AND NI-CU ALLOYS	ERCUNI OR ERCUNI-7 (540 °C, OR 1000 °F)	(A)	(A)	(A)	ERCUNI OR ERCUNI-7 (65 °C MAX, OR 150 °F MAX)
NI-CR, NI-FE, AND NI-CR-FE ALLOYS	ERCUNI-3 (540 °C, OR 1000 °F)	(A)	(A)	(A)	ERCUNI-3 (65 °C MAX, OR 150 °F MAX)
STEELS					
LOW-CARBON STEELS	ERCUAL-A2 OR ERCU OR ERCUNI-3 (540 °C, OR 1000 °F)	ERCUSN-A (200 °C, OR 400 °F)	ERCUAL-A2 (150 °C, OR 300 °F)	ERCUAL-A2 (65 °C MAX, OR 150 °F MAX)	ERCUAL-A2 OR ERCUNI-3 (65 °C MAX, OR 150 °F MAX)
MEDIUM-CARBON STEELS	ERCUAL-A2 OR ERCU OR ERCUNI-3 (540 °C, OR 1000 °F)	ERCUSN-A (200 °C, OR 400 °F)	ERCUAL-A2 (200 °C, OR 400 °F)	ERCUAL-A2 (65 °C MAX, OR 150 °F MAX)	ERCUAL-A2 OR ERCUNI-3 (65 °C MAX, OR 150 °F MAX)
HIGH-CARBON STEELS	ERCUAL-A2 OR ERCU OR ERCUNI-3 (540 °C, OR 1000 °F)	ERCUSN-A (260 °C, OR 500 °F)	ERCUAL-A2 (260 °C, OR 500 °F)	ERCUAL-A2 (200 °C, OR 400 °F)	ERCUAL-A2 OR ERCUNI-3 (65 °C MAX, OR 150 °F MAX)
LOW-ALLOY STEELS	ERCUAL-A2 OR ERCU OR ERCUNI-3 (540 °C, OR 1000 °F)	ERCUSN-A (260 °C, OR 500 °F)	ERCUAL-A2 (260 °C, OR 500 °F)	ERCUAL-A2 (200 °C, OR 400 °F)	ERCUAL-A2 OR ERCUNI-3 (65 °C MAX, OR 150 °F MAX)
STAINLESS STEELS	ERCUAL-A2 OR ERCU OR ERCUNI-3 (540 °C, OR 1000 °F)	ERCUSN-A (200 °C, OR 400 °F)	ERCUAL-A2 (65 °C MAX, OR 150 °F MAX)	ERCUAL-A2 (65 °C MAX, OR 150 °F MAX)	ERCUAL-A2 OR ERCUNI-3 (65 °C MAX, OR 150 °F MAX)

Notes: (1) Filler metal selections shown are based on weldability except where mechanical properties are usually more important. (2) Preheating usually is used only when at least one member is thicker than 3.2 mm ($\frac{1}{8}$ in.) or is highly conductive (see text). Preheat and interpass temperatures are subject to adjustment based on size and shape of weldment.

(A) THESE COMBINATIONS ARE SELDOM WELDED; AS A STARTING POINT IN DEVELOPING WELDING PROCEDURES, USE OF ERCUAL-A2 FILLER METALS IS RECOMMENDED, EXCEPT FOR COMBINATIONS INCLUDING PHOSPHOR BRONZE.

TABLE 16 ELECTRODES AND PREHEAT AND INTERPASS TEMPERATURES USED IN GMAW OF COPPER AND COPPER ALLOYS TO DISSIMILAR METALS

METAL TO BE WELDED	**ELECTRODES (WITH PREHEAT AND INTERPASS TEMPERATURES) FOR WELDING METAL IN COLUMN 1 TO:**						
	COPPERS	**LOW-ZINC**	**HIGH-ZINC**	**PHOSPHOR BRONZES**	**ALUMINUM BRONZES**	**SILICON BRONZES**	**COPPER-NICKEL**

		BRASSES	**BRASSES, TIN BRASSES, AND SPECIAL BRASSES**				**ALLOYS**
COPPER ALLOYS							
LOW-ZINC BRASSES	ERCUSN-A OR ERCU (540 °C, OR 1000 °F)	. . .	. . .	. . .	. . .	. . .	. . .
HIGH-ZINC, TIN, SPECIAL BRASSES	ERCUSI-A OR ERCUSN-A OR ERCU (540 °C, OR 1000 °F)	ERCUSN-A (315 °C, OR 600 °F)	. . .	. . .	. . .	. . .	. . .
PHOSPHOR BRONZES	ERCUSN-A OR ERCU (540 °C, OR 1000 °F)	ERCUSN-A (260 °C, OR 500 °F)	ERCUSN-A (315 °C. OR 600 °F)	. . .	. . .	. . .	. . .
ALUMINUM BRONZES	ERCUAL-A2 (540 °C, OR 1000 °F)	ERCUAL-A2 (315 °C, OR 600 °F)	ERCUAL-A2 (315 °C, OR 600 °F)	ERCUAL-A2 OR ERCUSN-A (200 °C, OR 400 °F)	. . .	. . .	. . .
SILICON BRONZES	ERCUSN-A OR ERCU (540 °C, OR 1000 °F)	ERCUAL-A2 OR ERCUSI-A (65 °C, OR 150 °F)	ERCUAL-A2 OR ERCUSI-A 65 °C, OR 150 °F)	ERCUSI-A (65 °C, OR 150 °F)	ERCUAL-A2 (65 °C, OR 150 °F)	. . .	. . .
COPPER-NICKEL ALLOYS	ERCUAL-A2 OR ERCUNI OR ERCU (540 °C, OR 1000 °F)	ERCUAL-A2 (65 °C, OR 150 °F)	ERCUAL-A2 (65 °C, OR 150 °F)	ERCUSN-A (65 °C, OR 150 °F)	ERCUAL-A2 (65 °C, OR 150 °F)	ERCUAL-A2 (65 °C, OR 150 °F)	. . .
NICKEL ALLOYS							
NICKEL AND NI-CU ALLOYS	ERCUNI OR ERCUNI-7 (540 °C, OR 1000 °F)	(A)	(A)	(A)	(A)	(A)	ERCUNI OR ERCUNI-7 (65 °C, OR 150 °F)
NI-CR, NI-FE, AND NI-CR-FE ALLOYS	ERCUNI-3 (540 °C, OR 1000 °F)	(A)	(A)	(A)	(A)	(A)	ERCUNI-3 (65 °C, OR 150 °F)
STEELS							
LOW-CARBON STEELS	ERCUAL-A2 OR ERCU OR ERCUNI-3 (540 °C, OR 1000 °F)	ERCUSN-A (315 °C, OR 600 °F)	ERCUAL-A2 (260 °C, OR 500 °F)	ERCUSN-A (200 °C, OR 400 °F)	ERCUAL-A2 (150 °C, OR 300 °F)	ERCUAL-A2 (65 °C, OR 150 °F)	ERCUAL-A2 OR ERCUNI-3 (65 °C, OR 150 °F)
MEDIUM-CARBON STEELS	ERCUAL-A2 OR ERCU OR ERCUNI-3 (540 °C, OR 1000 °F)	ERCUAL-A2 (315 °C, OR 600 °F)	ERCUAL-A2 (260 °C, OR 500 °F)	ERCUSN-A (200 °C, OR 400 °F)	ERCUAL-A2 (150 °C, OR 300 °F)	ERCUAL-A2 (65 °C, OR 150 °F)	ERCUAL-A2 OR ERCUNI-3 (65 °C, OR 150 °F)
HIGH-CARBON STEELS	ERCUAL-A2 OR ERCU	ERCUAL-A2 (315 °C,	ERCUAL-A2 (260 °C, OR	ERCUSN-A (260 °C, OR 500 °F)	ERCUAL-A2 (260 °C, OR 500 °F)	ERCUAL-A2 (200 °C, OR	ERCUAL-A2 OR ERCUNI-3

	OR ERCUNI-3 (540 °C, OR 1000 °F)	OR 600 °F)	500 °F)			400 °F)	(65 °C, OR 150 °F)
LOW-ALLOY STEELS	ERCUAL-A2 OR ERCU OR ERCUNI-3 (540 °C, OR 1000 °F)	ERCUAL-A2 (315 °C, OR 600 °F)	ERCUAL-A2 (315 °C, OR 600 °F)	ERCUSN-A (260 °C, OR 500 °F)	ERCUAL-A2 (260 °C, OR 500 °F)	ERCUAL-A2 (200 °C, OR 400 °F)	ERCUAL-A2 OR ERCUNI-3 (65 °C, OR 150 °F)
STAINLESS STEELS	ERCUAL-A2 OR ERCU OR ERCUNI-3 (540 °C, OR 1000 °F)	ERCUAL-A2 OR ERCUSN-A (315 °C, OR 600 °F)	ERCUAL-A2 (315 °C, OR 600 °F)	ERCUSN-A (200 °C, OR 400 °F)	ERCUAL-A2 (95 °C MAX, OR 200 °F MAX)	ERCUAL-A2 (95 °C MAX, OR 200 °F MAX)	ERCUAL-A2 OR ERCUNI-3 (65 °C MAX, OR 150 °F MAX)
CAST IRONS							
GRAY AND MALLEABLE IRONS	ERCUAL-A2 OR ERCU (540 °C, OR 1000 °F)	ERCUAL-A2 OR ERCUSN-A (315 °C, OR 600 °F)	ERCUAL-A2 (315 °C, OR 600 °F)	ERCUSN-A (200 °C, OR 400 °F)	ERCUAL-A2 (200 °C, OR 400 °F)	ERCUAL-A2 OR ERCUSI-A (150 °C, OR 300 °F)	ERCUAL-A2 OR ERCUNI (65 °C MAX, OR 150 °F MAX)
DUCTILE IRONS	ERCUAL-A2 OR ERCU (540 °C, OR 1000 °F)	ERCUAL-A2 (315 °C, OR 600 °F)	ERCUAL-A2 (315 °C, OR 600 °F)	ERCUSN-A (200 °C, OR 400 °F)	ERCUAL-A2 (95 °C MAX, OR 200 °F MAX)	ERCUAL-A2 OR ERCUSI-A (150 °C MAX, OR 300 °F)	ERCUAL-A2 OR ERCUNI (65 °C MAX, OR 150 °F MAX)

Notes: (1) Electrode selections in table are based on weldability, except where mechanical properties are usually more important. (2) Preheating usually is used only when at least one member is thicker than 3.2 mm ($\frac{1}{8}$ in.) or is highly conductive (see text). Preheat and interpass temperature are subject to adjustment based on size and shape of weldment.

(A) THESE COMBINATIONS ARE SELDOM WELDED; AS A STARTING POINT IN DEVELOPING WELDING PROCEDURES, USE OF ERCUAL-A2 ELECTRODES IS RECOMMENDED, EXCEPT FOR COMBINATIONS INCLUDING PHOSPHOR BRONZE.

Gas-Tungsten Arc Welding

Copper and many copper alloys can be gas-tungsten arc welded to other materials with good results. The use of a filler metal is typically required. Copper and its alloys have been readily gas-tungsten arc welded to steels, stainless steels, nickel, and nickel alloys. Gas-tungsten arc welding for the joining of these dissimilar metals is usually restricted to thin sections less than 3 mm (0.13 in.) thick. Preliminary surface welding, also known as buttering, is ordinarily not required for thin sections. For section thicknesses exceeding 3 mm (0.13 in.), the preferred joining process is GMAW or SMAW. Usually, when GTAW is used to join copper alloys to dissimilar metals, the arc is directed at the more conductive metal being welded. Table 15 shows combinations of dissimilar metals that are joined by GTAW with the aid of copper alloy or nickel alloy filler metals. The recommended filler metals, preheat temperature, and interpass temperature for each combination are given.

Except when one or both members of the joint being welded have high thermal conductivity, preheating is not required for GTAW of copper and dissimilar metals. This is restricted to section thicknesses of 3 mm (0.13 in.) or less.

Gas-Metal Arc Welding

The GMAW process can be used for joining nearly all combinations of the weldable copper alloys. Two major considerations when developing the GMAW procedures are: (1) the composition of the welding wire to be used as an electrode; and (2) the preheat temperatures required. As with the other arc welding processes, the arc is typically directed at the more conductive material to concentrate the heat energy on this base metal.

The electrodes as well as preheat and interpass temperatures for GMAW of copper and copper alloys to dissimilar metals are given in Table 16. This table includes all alloys listed in Table 15 for GTAW, plus the high-zinc brasses, tin brasses, and special brasses. As is shown in Table 16, ERCu-A1-A2 comes close to being the universal electrode for GMAW of dissimilar copper alloys. However, this filler metal is incompatible with phosphor bronzes. Silicon bronze (ERCuSi-A) and phosphor bronze (ERCuSn-A) electrodes are useful for many combinations of copper alloys that do not contain nickel. Copper electrodes (ERCu) are suitable for final welding passes (cap or cover passes) of coppers to any of the commonly welded copper alloys. Underlying or preliminary weld deposits are typically made with alloy electrodes. For joining copper to copper-nickel alloys, ERCuNi (copper-nickel) electrodes are generally used.

Generally, for welding dissimilar copper alloys, the electrode selected should be suitable for welding one of the materials within the combination to itself. The primary exception to this general rule is in the welding of silicon bronzes to copper-nickel alloys. The only electrode that can be used successfully for the joining of this combination is ERCuAl-A2.

As has been stated in the sections of this article dealing with specific alloy groups, alloys which have susceptibility to hot cracking, oxide formation, or other deterrents to weld strength characteristics should be noted when welding a dissimilar-metal combination. Typically, the restrictions for preheat, interpass temperature, and surface cleansing that apply to the individual alloy also apply to that alloy when joining it to other copper alloys or other metal groups. Of particular importance is preheat and interpass temperature for alloys that exhibit hot shortness.

For GMAW of copper or copper alloys to ferrous or nickel alloys, the primary consideration is dilution. Strength is not a factor because it is quite easy to generate weld metal whose strength is equal to that of the two base materials. The difficulty is in retaining the ductility demanded by the in-service requirements of the weldment. The most serious effect of excessive dilution is the formation of shrinkage cracks. These can start in the weld metal itself and propagate into the base metal.

Two methods are frequently used to control dilution in GMAW of combinations involving copper alloys. The first is to braze one side of the joint, thus promoting a minimum amount of dilution on that side, and weld the other side of the joint. The second technique is to overlay one or both joint surfaces with a buffer material. This technique, known as buttering, is typically used on thicker sections. Filler-metal selection again should be based upon the composition of the base alloy and the requirements for weld-metal strength and ductility. Additional information on buttering can be found in the article "Hardfacing, Weld Cladding, and Dissimilar Metal Joining" in this Volume.

Safe Welding Practices

Many of the elements used in producing copper alloys, as well as copper itself, have low or very low permissible exposure limits, as set by the American Conference of Governmental Industrial Hygienists. These would include such elements as arsenic, beryllium, cadmium, chromium, lead, manganese, and nickel. When welding, brazing, or soldering copper or copper alloys, adequate ventilation is required to ensure that contaminant levels do not exceed the human exposure limits. Precautions may include local exhaust ventilation, general area exhaust ventilation, operator respiratory protection, or a combination of these. The reader is referred to ANSI/AWS Z49.1, "Safety in Welding and Cutting," for detailed information.

Copper alloys that contain appreciable amounts of beryllium, cadmium, or chromium may present health hazards to the welder and others. Exposure to welding fumes containing these elements can produce adverse health defects. The user should consult the Occupational Safety and Health Administration guidelines for the specific element.

Fumes and dust of copper and zinc can cause irritation of the upper respiratory tract, nausea, metal-fume fever, skin irritation, dermatitis, and eye problems. It should be noted that cadmium and beryllium fumes are toxic when inhaled.

Fluxes used during welding, brazing, and soldering of certain copper alloys may contain fluorides and chlorides. Again, these fumes can be irritating to the eyes, nose, throat, and skin. Additionally, some fluorine compounds are toxic.

Proper personal hygiene must be emphasized, particularly before eating. Food and beverages should not be stored or consumed in and/or around the work area. Contaminated clothing should be changed. For more detailed information on safety considerations, see the article "Safe Practices" in this Volume.

Welding of Copper Alloys

Revised by M. Ned Rogers, Batesville Casket Company

Selected References

- H.B. CARY, *MODERN WELDING TECHNOLOGY,* 2ND ED., PRENTICE-HALL, 1991
- "GUIDE TO THE WELDING OF COPPER-NICKEL ALLOYS," INTERNATIONAL NICKEL COMPANY, 1979
- E.F. NIPPES, ED., 9TH ED., VOL 6, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1984
- "SAFETY IN CUTTING AND WELDING," ANSI/ IAWS Z49.1, AMERICAN WELDING SOCIETY, 1988
- "SPECIFICATION FOR COPPER AND COPPER ALLOY BARE WELDING RODS AND ELECTRODES," AWS A5.7, AMERICAN WELDING SOCIETY, 1984
- "SPECIFICATION FOR COPPER AND COPPER ALLOY ELECTRODES," AWS A5.6, AMERICAN WELDING SOCIETY, 1984
- "SPECIFICATION FOR COPPER AND COPPER ALLOY GAS WELDING RODS," AWS A5.27, AMERICAN WELDING SOCIETY, 1979
- *WELDING HANDBOOK,* 8TH ED., VOL 2, AMERICAN WELDING SOCIETY, 1991

Introduction

MOST MAGNESIUM ALLOYS can be joined by gas-tungsten arc welding (GTAW) and gas-metal arc welding (GMAW). Relative weldability ratings, joint design and surface preparation, and the use of filler metals and shielding gases suitable to arc welding are discussed in this article. For more information on arc welding, see the Section "Fusion Welding Processes" in this Volume.

Weldability

Table 1 lists some magnesium alloys that are weldable, along with their respective weldability ratings based on a scale of A (excellent) to D (limited). This rating is based largely on freedom from susceptibility to cracking, and to some extent on joint efficiency. Under optimum welding conditions, including favorable joint design, joint efficiencies of 60 to 100% can be obtained for virtually all of the magnesium alloys. Alloys rated A in Table 1 are likely to have high joint efficiency ratings.

TABLE 1 RELATIVE ARC WELDABILITY OF MAGNESIUM ALLOYS

A, excellent; B, good; C, fair; D, limited weldability

ALLOY	RATING
CASTING ALLOYS	
AM100A	B+
AZ63A	C
AZ81A	B+
AZ91C	B+
AZ92A	B
EK30A	B
EK41A	B
EQ21	B
EZ33A	A
HK31A	B+
HZ32A	B
K1A	A
QE22A	B
ZE41A	B
QH21A	B
WE43	B-
WE54	B-
ZC63	B-
ZH62A	C-
ZK51A	D
ZK61A	D
WROUGHT ALLOYS	
AZCOML	A
AZ10A	A
AZ31B,C	A

AZ80A	B
HK31A	A
HM21A	A
HM31A	A
ZE10A	A
ZK21A	B

In the magnesium-aluminum-zinc alloys (AZ31B, AZ61A, AZ63A, AZ80A, AZ81A, AZ91C, and AZ92A), aluminum content up to about 10% (Table 2) aids weldability by helping to refine the grain structure, while zinc content of more than 1% increases hot shortness, which may cause weld cracking. Alloys with high zinc content (ZH62A, ZK51A, ZK60A, and ZK61A) are highly susceptible to cracking and have poorer weldability. Thorium-containing alloys (HK31A, HM21A, and HM31A) have excellent arc weldability and are rated B + or A in Table 1.

TABLE 2 COMPOSITION OF SELECTED MAGNESIUM ALLOYS

DESIGNATION		**COMPOSITION, WT%**									
ASTM	**UNS NO.**	**Al**	**Li**	**Mn**	**Rare earths**	**Th**	**Zn**	**Zr**	**Ca**	**Cu**	**Fe**
9980A	M19980	. . .	. . .	0.10	. . .	. . .	. . .	. . .	. . .	0.02	. . .
9980B	M19981	. . .	. . .	0.10	. . .	. . .	. . .	. . .	. . .	0.021	. . .
9990A	M19990	0.003	. . .	0.004	. . .	. . .	. . .	. . .	. . .	. . .	0.04
9990B	M19991	0.005	. . .	0.004	. . .	. . .	. . .	. . .	. . .	. . .	0.011
9995A	M19995	0.010	. . .	0.004	. . .	. . .	. . .	. . .	. . .	. . .	0.003
9998A	M19998	0.004	. . .	0.002	. . .	. . .	. . .	. . .	. . .	0.0005	0.002
A3A	M10030	2.5-3.5	. . .	0.005	0.001	. . .	0.10	. . .	. . .	0.005	0.005
AM60A	M10600	5.5-6.5	. . .	0.13 MIN	. . .	. . .	0.22	. . .	. . .	0.35	. . .
AM80A	M10800	8.0-9.0	. . .	0.18 MIN	. . .	. . .	0.20	. . .	. . .	0.08	. . .
AM90A	M10900	8.5-9.5	. . .	0.15 MIN	. . .	. . .	0.20	. . .	. . .	0.10	0.008
AM100A	M10100	9.3-10.7	. . .	0.10 MIN	. . .	. . .	0.30	. . .	. . .	0.10	. . .
AM100B	M10102	9.4-10.6	. . .	0.13 MIN	. . .	. . .	. . .	. . .	. . .	0.08	. . .
AS41A	M10410	3.5-5.0	. . .	0.20-0.50	. . .	. . .	0.12	. . .	. . .	0.06	. . .
AZ10A	M11100	1.0-1.5	. . .	0.20 MIN	. . .	. . .	0.20-0.6	. . .	0.04	0.10	0.005
AZ21A	M11210	1.6-2.5	. . .	0.15	. . .	. . .	0.8-1.6	. . .	0.10-0.25	0.05	0.005
AZ31A	M11310	2.5-3.5	. . .	0.20 MIN	. . .	. . .	0.6-1.4	. . .	0.30	0.05	0.005
AZ31B	M11311	2.5-3.5	. . .	0.20 MIN	. . .	. . .	0.6-1.4	. . .	0.04	0.05	0.005
AZ31C	M11312	2.4-3.6	. . .	0.15 MIN	. . .	. . .	0.50-1.5	. . .	. . .	0.10	. . .
AZ61A	M11610	5.8-7.2	. . .	0.15 MIN	. . .	. . .	0.40-1.5	. . .	. . .	0.05	0.005

AZ63A	M11630	5.3-6.7	. . .	0.15 MIN	. . .	. . .	2.5-3.5	. . .	. . .	0.25	. . .
AZ80A	M11800	7.8-9.2	. . .	0.12 MIN	. . .	. . .	0.20-0.8	. . .	. . .	0.05	0.005
AZ81A	M11810	7.0-8.1	. . .	0.13 MIN	. . .	. . .	0.40-1.0	. . .	. . .	0.10	. . .
AZ90A	M11900	8.5-9.5	. . .	0.15 MIN	. . .	. . .	0.20	. . .	. . .	0.02	0.015
AZ91A	M11910	8.5-9.5	. . .	0.15 MIN	. . .	. . .	0.45-0.9	. . .	. . .	0.08	. . .
AZ91B	M11912	8.5-9.5	. . .	0.15 MIN	. . .	. . .	0.45-0.9	. . .	. . .	0.25	. . .
AZ91C	M11914	8.1-9.3	. . .	0.13 MIN	. . .	. . .	0.40-1.0	. . .	. . .	0.10	. . .
AZ92A	M11920	8.3-9.7	. . .	0.10 MIN	. . .	. . .	1.6-2.4	. . .	. . .	0.25	. . .
AZ101A	M11101	9.5-10.5	. . .	0.13 MIN	. . .	. . .	0.75-1.25	. . .	. . .	0.05	0.005
AZ125A	M11125	11.0-13.0	. . .	. . .	. . .	. . .	4.5-5.5	. . .	. . .	. . .	. . .
EK30A	M12300	. . .	. . .	. . .	2.5-4.0	. . .	0.30	0.20 MIN	. . .	0.10	. . .
EK41A	M12410	. . .	. . .	. . .	3.0-5.0	. . .	0.30	0.40-1.0	. . .	0.10	. . .
EZ33A	M12330	. . .	. . .	. . .	2.5-4.0	. . .	2.0-3.1	0.50-1.0	. . .	0.10	. . .
HK31A	M13310	. . .	. . .	. . .	. . .	2.5-4.0	0.30	0.40-1.0	. . .	0.10	. . .
HM21A	M13210	. . .	. . .	0.45-1.1	. . .	1.5-2.5	. . .	. . .	. . .	. . .	. . .
HM31A	M13312	. . .	. . .	1.2 MIN	. . .	2.5-3.5	. . .	. . .	. . .	. . .	. . .
HZ32A	M13320	. . .	. . .	. . .	0.10	2.5-4.0	1.7-2.5	0.50-1.0	. . .	0.10	. . .
K1A	M18010	. . .	. . .	. . .	. . .	. . .	. . .	0.40-1.0	. . .	. . .	. . .
LA141A	M14141	1.0-1.5	13.0-15.0	0.15 MIN	. . .	. . .	. . .	. . .	. . .	0.005	. . .
LS141A	M14142	0.05	12.0-15.0	0.15	. . .	. . .	. . .	. . .	. . .	0.05	0.005
LZ145A	M14145	0.05	12.0-15.0	0.15	. . .	. . .	4.5-5.0	. . .	. . .	0.05	0.005
M1A	M15100	. . .	. . .	1.2 MIN	. . .	. . .	. . .	. . .	0.30	0.05	. . .
M1B	M15101	. . .	. . .	1.3 MIN	. . .	. . .	. . .	. . .	. . .	0.08	. . .
M1C	M15102	0.01	. . .	0.9-1.2	. . .	. . .	. . .	. . .	. . .	0.02	0.03
TA54A	M18540	3.0-4.0	. . .	0.20 MIN	. . .	. . .	0.30	. . .	. . .	0.05	. . .
QE22A	M18220	. . .	. . .	. . .	1.8-2.5(A)	. . .	. . .	0.40-1.0	. . .	0.10	. . .
QH21A	M18210	. . .	. . .	. . .	0.6-	0.6-	0.20	0.40-	. . .	0.10	. . .

					1.5(B)(C)	1.6(C)		1.0(D)			
ZE10A	M16100	. . .	. . .	. . .	0.12-0.22	. . .	1.0-1.5	. . .	. . .	. . .	. . .
ZE41A	M16410	. . .	. . .	0.15	0.75-1.75	. . .	3.5-5.0	0.40-1.0	. . .	0.10	. . .
ZE63A	M16630	. . .	. . .	. . .	2.1-3.0	. . .	5.5-6.0	0.40-1.0	. . .	0.10	. . .
ZH62A	M16620	. . .	. . .	. . .	. . .	1.4-2.2	5.2-6.2	0.50-1.0	. . .	0.10	. . .
ZK21A	M16210	. . .	. . .	. . .	. . .	. . .	2.0-2.6	0.45-0.8	. . .	0.10	. . .
ZK40A	M16400	. . .	. . .	. . .	. . .	. . .	3.5-4.5	0.45 MIN	. . .	0.10	. . .
ZK51A	M16510	. . .	. . .	. . .	. . .	. . .	3.6-5.5	0.50-1.0	. . .	. . .	. . .
ZK60A	M16600	. . .	. . .	. . .	. . .	. . .	4.8-6.2	0.45 MIN	. . .	. . .	. . .
ZK60B	M16601	. . .	. . .	. . .	. . .	. . .	4.8-6.8	0.45 MIN	. . .	. . .	. . .
ZK61A	M16610	. . .	. . .	. . .	. . .	. . .	5.5-6.5	0.6-1.0	. . .	. . .	. . .

DESIGNATION		**COMPOSITION, WT%**					**OTHER ELEMENTS**			**Mg**
ASTM	**UNS NO.**	**Si**	**Ag**	**Cd**	**Ni**	**Pb**	**OTHER**	**EACH**	**TOTAL**	
9980A	M19980	. . .	. . .	. . .	0.001	0.01	0.01 SN	0.05	. . .	99.80 MIN
9980 B	M19981	. . .	. . .	. . .	0.005	0.01	0.01 SN	0.05	. . .	99.80 MIN
9990A	M19990	0.005	. . .	0.0001	0.001	. . .	0.00007 B	0.01	. . .	99.90 MIN
9990B	M19991	0.005	. . .	. . .	0.001	. . .	. . .	0.01	. . .	99.90 MIN
9995A	M19995	0.005	. . .	0.00005	0.001	. . .	0.00003 B 0.01 TI	0.005	. . .	99.95 MIN
9998A	M19998	0.003	. . .	0.00005	0.00005	0.001	0.00003 B 0.001 TI	0.005	. . .	99.98 MIN
A3A	M10030	. . .	0.001	0.001	0.001	. . .	. . .	. . .	0.30	BAL
AM60A	M10600	0.50	. . .	. . .	0.03	. . .	. . .	. . .	0.30	BAL
AM80A	M10800	0.20	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
AM90A	M10900	0.15	0.008	. . .	0.005	. . .	. . .	0.02	0.10	BAL
AM100A	M10100	0.30	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
AM100B	M10102	1.0	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
AS41A	M10410	0.50-1.5	. . .	. . .	0.03	. . .	. . .	. . .	0.30	BAL
AZ10A	M11100	0.10	. . .	. . .	0.005	. . .	. . .	. . .	0.30	BAL
AZ21A	M11210	0.05	. . .	. . .	0.002	. . .	. . .	. . .	0.30	BAL
AZ31A	M11310	0.30	. . .	. . .	0.005	. . .	. . .	. . .	0.30	BAL
AZ31B	M11311	0.10	. . .	. . .	0.005	. . .	. . .	. . .	0.30	BAL
AZ31C	M11312	0.10	. . .	. . .	0.03	. . .	. . .	. . .	0.30	BAL
AZ61A	M11610	0.10	. . .	. . .	0.005	. . .	. . .	. . .	0.30	BAL

AZ63A	M11630	0.30	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
AZ80A	M11800	0.10	. . .	. . .	0.005	. . .	. . .	. . .	0.30	BAL
AZ81A	M11810	0.30	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
AZ90A	M11900	0.20	. . .	. . .	0.005	. . .	. . .	0.07	0.30	BAL
AZ91A	M11910	0.20	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
AZ91B	M11912	0.20	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
AZ91C	M11914	0.30	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
AZ92A	M11920	0.30	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
AZ101A	M11101	0.05	. . .	. . .	0.005	. . .	. . .	0.0002-0.0008 BE	0.30	BAL
AZ125A	M11125	. . .	. . .	. . .	. . .	. . .	. . .	0.0002-0.0008 BE	0.30	BAL
EK30A	M12300	. . .	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
EK41A	M12410	. . .	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
EZ33A	M12330	. . .	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
HK31A	M13310	. . .	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
HM21A	M13210	. . .	. . .	. . .	. . .	. . .	. . .	. . .	0.30	BAL
HM31A	M13312	. . .	. . .	. . .	. . .	. . .	. . .	. . .	0.30	BAL
HZ32A	M13320	. . .	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
K1A	M18010	. . .	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
LA141A	M14141	0.004	. . .	. . .	0.005	. . .	0.005 NA	. . .	0.20	BAL
LS141A	M14142	0.50-0.6	. . .	. . .	0.005	. . .	0.005 NA	. . .	. . .	BAL
LZ145A	M14145	1.5-2.0	2.0-3.0	. . .	0.005	. . .	0.005 NA	. . .	0.20	BAL
M1A	M15100	0.10	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
M1B	M15101	0.10	. . .	. . .	0.01	. . .	. . .	. . .	0.20	BAL
M1C	M15102	. . .	. . .	. . .	0.001	. . .	. . .	0.05	0.30	BAL
TA54A	M18540	0.30	. . .	. . .	0.01	. . .	4.0-6.0 SN	. . .	0.30	BAL
QE22A	M18220	. . .	2.0-3.0	. . .	0.01	. . .	. . .	. . .	0.30	BAL
QH21A	M18210	. . .	2.0-3.0	. . .	0.01	. . .	. . .	. . .	0.30	BAL
ZE10A	M16100	. . .	. . .	. . .	. . .	. . .	. . .	. . .	0.30	BAL
ZE41A	M16410	. . .	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
ZE63A	M16630	. . .	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
ZH62A	M16620	. . .	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
ZK21A	M16210	. . .	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
ZK40A	M16400	. . .	. . .	. . .	0.01	. . .	. . .	. . .	0.30	BAL
ZK51A	M16510	. . .	. . .	. . .	. . .	. . .	. . .	. . .	0.30	BAL
ZK60A	M16600	. . .	. . .	. . .	. . .	. . .	. . .	. . .	0.30	BAL
ZK60B	M16601	. . .	. . .	. . .	. . .	. . .	. . .	. . .	0.30	BAL
ZK61A	M16610	. . .	. . .	. . .	. . .	. . .	. . .	. . .	0.30	BAL

Source: Ref 1

(A) RARE EARTH ELEMENTS ARE IN THE FORM OF DIDYMIUM.

(B) DIDYMIUM, NOT LESS THAN 70% NEODYMIUM ANY REMAINDER SUBSTANTIALLY PRAESODYMIUM.

(C) THORIUM AND DIDYMIUM TOTAL 1.5 TO 2.4.

(D) ZIRCONIUM SOLUBLE 0.40 MIN, BUT DETERMINATION NOT REQUIRED FOR ROUTINE ACCEPTANCE.

Welds in magnesium alloys are characterized by a fine grain size averaging less than 0.25 mm (0.01 in.). Magnesium alloys containing more than 1.5% Al are susceptible to stress corrosion, and residual welding stresses must be relieved.

Reference cited in this section

1. "STANDARD PRACTICE FOR CODIFICATION OF CERTAIN NONFERROUS METALS AND ALLOYS, CAST AND WROUGHT," B 275, *1989 ANNUAL BOOK OF STANDARDS,* VOL 2.02, *DIE-CAST METALS; ALUMINUM AND MAGNESIUM ALLOYS,* ASTM, 1989, P 325-326

Welding of Magnesium Alloys

Filler Metals

Compositions of the four most commonly used electrode wires for GMAW and filler metals (when used) for GTAW are given in Table 3. The choice of electrode wire or filler metal is governed by the composition of the base metal (Table 4).

TABLE 3 COMPOSITIONS OF ELECTRODES AND FILLER METALS USED IN GAS SHIELDED ARC WELDING OF MAGNESIUM ALLOYS PER AWS A5.19

FILLER METAL	COMPOSITION, WT%										
	Al	Be	Mn	Zn	Zr	Rare earth	Cu	Fe	Ni	Si	OTHERS (TOTAL)	MG
ER AZ61A	5.8-7.2	0.0002-0.0008	0.15 MIN	0.40-1.5	. . .	. . .	0.05 MAX	0.005 MAX	0.005 MAX	0.05 MAX	0.30 MAX	BAL
ER AZ101A	9.5-10.5	0.0002-0.0008	0.13 MIN	0.75-1.25	. . .	. . .	0.05 MAX	0.005 MAX	0.005 MAX	0.05 MAX	0.30 MAX	BAL
ER AZ92A	8.3-9.7	0.0002-0.0008	0.15 MIN	1.7-2.3	. . .	. . .	0.05 MAX	0.005 MAX	0.005 MAX	0.05 MAX	0.30 MAX	BAL
ER EZ33A	. . .	. . .	. . .	2.0-3.1	0.45-1.0	2.5-4.0	. . .	. . .	. . .	. . .	0.30 MAX	BAL

TABLE 4 SELECTION GUIDE FOR USING FILLER METAL TO JOIN TWO MAGNESIUM-BASE ALLOYS

BASE ALLOY	**FILLER METAL FOR USE WITH INDICATED BASE ALLOYS**																						
	AM100A	**AZ10A**	**AZ31, AZ31C**	**AZ61A**	**AZ63A**	**AZ80A**	**AZ81A**	**AZ91C**	**AZ92A**	**EK41A**	**EZ33A**	**HK31A**	**HM21A**	**HM31A**	**HZ32A**	**K1A**	**LA141A**	**M1A, MG1**	**QE22A**	**ZE10A**	**ZE41A**	**ZX21A**	**ZH62A, ZK51A, ZK60A, ZK61A**
AM100A	AZ92A, AZ101	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .
AZ10A	AZ92A	AZ61A, AZ32A	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .
AZ31B, AZ31C	AZ92A	AZ61A, AZ92A	AZ61A, AZ92A	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .
AZ61A,	AZ92A	AZ61A, AZ92A	AZ61A, AZ92A	AZ61A, AZ92A	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .
AZ63A	(A)	(A)	(A)	(A)	AZ92A	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .
AZ80A	AZ92A	AZ61A, AZ92A	AZ61A, AZ92A	AZ61A, AZ92A	(A)	AZ61A, AZ92A	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .
AZ81A	AZ92A	AZ92A	AZ92A	AZ92A	(A)	AZ92A	AZ92A, AZ101	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .
AZ91C	AZ92A	AZ92A	AZ92A	AZ92A	(A)	AZ92A	AZ92A	AZ92A, AZ101	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .
AZ92	AZ92	AZ9	AZ9	AZ9	(A)	AZ9	AZ9	AZ9	AZ1	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .

A	A	2A	2A	2A		2A	2A	2A	01														
EK41A	AZ92A	AZ92A	AZ92A	AZ92A	(A)	AZ92A	AZ92A	AZ92A	AZ92A	EZ33A	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
EZ33A	AZ92A	AZ92A	AZ92A	AZ92A	(A)	AZ92A	AZ92A	AZ92A	AZ92A	EZ33A	EZ33A	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
HK31A	AZ92A	AZ92A	AZ92A	AZ92A	(A)	AZ92A	AZ92A	AZ92A	AZ92A	EZ33A	EZ33A	EZ33A	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
HM21A	AZ92A	AZ92A	AZ92A	AZ92A	(A)	AZ92A	AZ92A	AZ92A	AZ92A	EZ33A	EZ33A	EZ33A	EZ33A	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
HM31A	AZ92A	AZ92A	AZ92A	AZ92A	(A)	AZ92A	AZ92A	AZ92A	AZ92A	EZ33A	EZ33A	EZ33A	EZ33A	EZ33A	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
HZ32A	AZ92A	AZ92A	AZ92A	AZ92A	(A)	AZ92A	AZ92A	AZ92A	AZ92A	EZ33A	EZ33A	EZ33A	EZ33A	EZ33A	EZ33A	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
K1A	AZ92A	AZ92A	AZ92A	AZ92A	(A)	AZ92A	AZ92A	AZ92A	AZ92A	EZ33A	EZ33A	EZ33A	EZ33A	EZ33A	EZ33A	EZ33A	. . .	. . .	. . .	. . .	. . .	. . .	. . .
M1A, MG1	AZ92A	AZ61A, AZ92A	AZ61A, AZ92A	AZ61A, AZ92A	(A)	AZ61A, AZ92A	AZ92A	AZ92A	AZ92A	AZ92A	AZ92A	AZ92A	AZ92A	AZ92A	AZ92A	AZ92A	(B)	AZ61A, AZ92A	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .
ZE41A	(B)	(B)	(B)	(B)	(A)	(B)	(B)	(B)	(B)	EZ33A	EZ33A	EZ33A	EZ33A	EZ33A	EZ33A	EZ33A	(B)	(B)	EZ33A	(B)	EZ33A	. . .	. . .
ZX21A	AZ92A	AZ61A, AZ92A	AZ61A, AZ92A	AZ61A, AZ92A	(A)	AZ61A, AZ92A	AZ92A	AZ92A	AZ92A	AZ92A	AZ92A	AZ92A	AZ92A	AZ92A	AZ92A	AZ92A	(B)	AZ61A, AZ92A	AZ92A	AZ61A, AZ92A	AZ92A	AZ61A, AZ92A	. . . . . .
ZH62A, ZK51A, ZK60A, ZK61A	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	(A)	EZ33A

(A) WELDING NOT RECOMMENDED.

(B) WELDING DATA NOT AVAILABLE

Electrode wires or filler metals having compositions conforming to ER AZ61A or ER AZ92A (Mg-Al-Zn) are considered satisfactory for welding alloys AZ10A, AZ31B, AZ31C, AZCOML, AZ61A, AZ80A, ZE10A, and ZK21A to themselves or to each other. ER AZ61A is usually preferred for welding aluminum-containing wrought products because of its tendency to resist crack sensitivity. The ER AZ92A filler metal shows less crack sensitivity for welding the cast magnesium-aluminum-zinc and magnesium-aluminum alloys. The same electrode wires or filler metals are used for joining any one of the above alloys to high-temperature alloys HK31A, HM21A, and HM31A. However, when the high-temperature alloys are joined to each other, ER EZ33A is recommended. Joints of wrought or cast alloys welded with ER EZ33A filler metal exhibit good mechanical properties at high temperatures.

The choice of electrode wire or filler metal for welding wrought alloys to cast alloys should be based on the recommendations outlined above, except that ER AZ101A may be used instead of ER AZ61A or ER AZ92A.

When aluminum-containing cast alloys are joined to aluminum-containing cast alloys, ER AZ101A or ER AZ92A electrode wire or filler metal is usually recommended. However, for joining HK31A and HZ32A to themselves or to each other, ER EZ33A is preferred; for joining HK31A and HZ32A to any of the other cast alloys, ER AZ101A is used. Filler rod of the same composition as the base metal should be used for most welds.

Shielding Gases

Only the inert gases are used for shielding in arc welding of magnesium alloys. Argon is the most widely used. Helium and various mixtures of argon and helium have also proved satisfactory, but because of the higher cost per unit volume of helium, and because two to three times more helium than argon is required for the same degree of shielding, the use of pure helium has gradually decreased. Pure helium is also undesirable because it raises the current required for spray arc transfer and increases weld spatter. The factors determining shielding-gas selection for magnesium alloys are similar to those for aluminum.

Joint Design

Typical joint designs for gas shielded arc welding of various thicknesses of magnesium alloy sheet and plate are presented in Table 5. The metal thickness ranges vary for GTAW in accordance with the type of current used (alternating current; direct current electrode negative, DCEN; and direct current electrode positive, DCEP) and vary for GMAW in accordance with the mode of metal transfer.

TABLE 5 TYPICAL JOINT DESIGNS USED FOR GAS SHIELDED ARC WELDING OF SELECTED THICKNESSES OF MAGNESIUM ALLOY SHEET AND PLATE

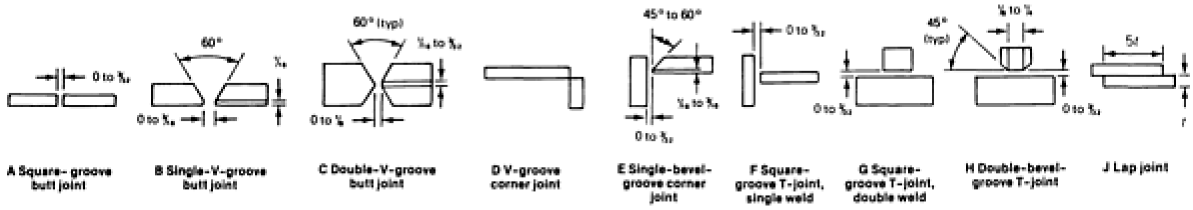

A Square-groove butt joint
B Single-V-groove butt joint
C Double-V-groove butt joint
D V-groove corner joint
E Single-bevel-groove corner joint
F Square-groove T-joint, single weld
G Square-groove T-joint, double weld
H Double-bevel-groove T-joint
J Lap joint

TYPE OF JOINT	APPLICABLE RANGE OF WORK-METAL THICKNESS[(A)]											
	GTAW[(B)]						GMAW[(C)]					
	AC		DCEN		DCEP		SHORT CIRCUITING		PULSED ARC		SPRAY ARC	
	mm	in.	mm	in.	mm	in.	mm	in.	mm	in.	mm	in.
A[(D)]	0.64-6.4	0.025-$\frac{1}{4}$	0.64-13	0.025-$\frac{1}{2}$	0.64	0.025-$\frac{3}{16}$	0.64-4.8	0.025-$\frac{3}{16}$	2.3-6.4	0.090-$\frac{1}{4}$	4.8-9.5	$\frac{3}{16}$-

												$\frac{3}{8}$
B$^{(E)}$	6.4-9.5	$\frac{1}{4}$ $\frac{3}{8}$ $\frac{3}{8}$	6.4-9.5	$\frac{1}{4}$-$\frac{3}{8}$	4.8-9.5	$\frac{3}{16}$-$\frac{3}{8}$	$^{(F)}$	$^{(F)}$	4.8-6.4	$\frac{3}{16}$-$\frac{1}{4}$	6.4-13	$\frac{1}{4}$-$\frac{1}{2}$
C$^{(G)}$	9.5$^{(H)}$	$\frac{3}{8}$ (H)	9.5$^{(H)}$	$\frac{3}{8}$ (H)	9.5$^{(H)}$	$\frac{3}{8}$ (H)	$^{(F)}$	$^{(F)}$	$^{(F)}$	$^{(F)}$	13$^{(H)}$	$\frac{1}{2}$ (H)
D$^{(I)}$	1.0-6.4	0.04-$\frac{1}{4}$	1.0-6.4	0.040-$\frac{1}{4}$	1.0-6.4	0.040-$\frac{1}{4}$	1.6-4.8	$\frac{1}{16}$-$\frac{3}{16}$	1.6-6.4	$\frac{1}{16}$-$\frac{1}{4}$	4.8-13	$\frac{3}{16}$-$\frac{1}{2}$
E$^{(J)}$	4.8$^{(H)}$	$\frac{3}{16}$ (H)	4.8$^{(H)}$	$\frac{3}{16}$ (H)	4.8$^{(H)}$	$\frac{3}{16}$ (H)	$^{(F)}$	$^{(F)}$	3.2-6.4	$\frac{1}{8}$-$\frac{1}{4}$	6.4$^{(H)}$	$\frac{1}{4}$ (H)
F$^{(K)}$	0.64-6.4	0.025-$\frac{1}{4}$	0.64-13	0.025-$\frac{1}{2}$	0.64-4.0$^{(H)}$	0.025-$\frac{5}{32}$ (H)	1.6-4.0	$\frac{1}{16}$-$\frac{5}{32}$	2.3-4.8	0.090-$\frac{3}{16}$	4.0-9.5	$\frac{5}{32}$-$\frac{3}{8}$
G$^{(L)}$	1.6-4.8	$\frac{1}{16}$-$\frac{3}{16}$	1.6-9.5	$\frac{1}{16}$-$\frac{3}{8}$	1.6-3.2	$\frac{1}{16}$-$\frac{1}{8}$	1.6-4.0	$\frac{1}{16}$-$\frac{5}{32}$	2.3-6.4	0.090-$\frac{1}{4}$	4.0-19	$\frac{5}{32}$-$\frac{3}{4}$
H$^{(M)}$	4.8$^{(H)}$	$\frac{3}{16}$ (H)	9.5$^{(H)}$	$\frac{3}{8}$ (H)	3.2$^{(H)}$	$\frac{1}{8}$ (H)	$^{(F)}$	$^{(F)}$	6.4-9.5	$\frac{1}{4}$-$\frac{3}{8}$	9.5$^{(H)}$	$\frac{3}{8}$ (H)
J$^{(N)}$	1.0$^{(H)}$	0.040$^{(H)}$	1.0$^{(H)}$	0.040$^{(H)}$	0.64$^{(H)}$	0.025$^{(H)}$	1.0-4.0	0.040-$\frac{5}{32}$	2.3-6.4	0.090-$\frac{1}{4}$	4.0$^{(H)}$	$\frac{5}{32}$ (H)

(A) SUGGESTED MINIMUM AND MAXIMUM THICKNESS LIMITS.

(B) USING 300 A AC OR DCEN, OR 125 A DCEP.

(C) USING 400 A DCEP.

(D) SINGLE-PASS FULL-PENETRATION WELD. SUITABLE FOR THIN MATERIAL.

(E) FULL-PENETRATION WELD. SUITABLE FOR THICK MATERIAL. ON MATERIAL THICKER THAN SUGGESTED MAXIMUM, USE DOUBLE-V-GROOVE BUTT JOINT TO MINIMIZE DISTORTION.

(F) NOT RECOMMENDED BECAUSE SPRAY ARC WELDING IS MORE PRACTICAL OR ECONOMICAL, OR BOTH.

(G) FULL-PENETRATION WELD. USED ON THICK MATERIAL. MINIMIZES DISTORTION BY EQUALIZING SHRINKAGE STRESS ON BOTH SIDES OF JOINT.

(H) NO MAXIMUM. THICKEST MATERIAL IN COMMERCIAL USE COULD BE WELDED IN THIS TYPE OF JOINT.

(I) SINGLE-PASS FULL-PENETRATION ESPECIALLY IF A SQUARE CORNER IS REQUIRED.

(J) SINGLE-PASS OR MULTIPLE-PASS FULL-PENETRATION WELD. USED ON THICK MATERIAL TO MINIMIZE WELDING. PRODUCES SQUARE JOINT CORNERS.

(K) SINGLE WELD T-JOINT. THICKNESS LIMITS ARE BASED ON 40% JOINT PENETRATION.

(L) DOUBLE WELD T-JOINT. SUGGESTED THICKNESS LIMITS BASED ON 100% JOINT PENETRATION.

(M) DOUBLE WELD T-JOINT. USED ON THICK MATERIAL REQUIRING 100% JOINT PENETRATION.

(N) SINGLE OR DOUBLE WELD JOINT. STRENGTH DEPENDS ON SIZE OF FILLET. MAXIMUM STRENGTH IN TENSION ON DOUBLE WELD JOINTS IS OBTAINED WHEN LAP EQUALS FIVE TIMES THE THICKNESS OF THE THINNER MEMBER.

Edges are prepared by milling, sawing, shearing, arc cutting, chipping, planing, routing, sanding, or filing. Sheet up to about 2.03 mm (0.080 in.) thick is generally single sheared; double shearing is preferred for sheet thicker than 2.03 mm (0.080 in.). Double-bevel edges are preferred over single-bevel edges because less distortion occurs during welding.

Fit-Up. Components of weldments should fit closely at abutting edges, preferably with no root opening, although a root opening up to 1.6 mm ($\frac{1}{16}$ in.) is usually permissible. If tack welds are used, the first tack should be a short distance from the end of a seam. The best practice in tack welding is to use 3.2 mm ($\frac{1}{8}$ in.) tacks spaced 25 to 50 mm (1 to 2 in.) on centers for sheet up to 1.65 mm (0.065 in.) thick, and 6.4 mm ($\frac{1}{4}$ in.) tacks spaced 100 to 125 mm (4 to 5 in.) on centers for sheet or plate from 1.65 to 6.4 mm (0.065 to $\frac{1}{4}$ in.) thick.

Welding fixtures must be rigid enough to resist movement of the workpieces during welding. Hold-down bars must exert enough pressure to keep the edges from overlapping or rising away from the backing bar or plate. A backing bar or strip normally is used when welding sheet metal components to help control joint penetration, root surface contour, and heat removal. The gas is usually supplied through holes in the backing strip. When a backing strip cannot be used because of space limitations, a chemical flux is sometimes painted on the root side of the joint to smooth the root surface and control joint penetration. Backing plates are made of mild steel, stainless steel, magnesium, aluminum, or copper and are cut with a small groove that is positioned directly under the seam. They prevent excessive metal drop-through and minimize distortion. Depth of the grooves in backing plates depends on base-metal thickness, the welding process used, and the absence or presence of a root opening in the joint. Typical groove depths are given in Table 6. When a temporary backing bar or strip is used, the root side of the joint should be shielded with inert gas to prevent oxidation of the root surface.

TABLE 6 TYPICAL DEPTH OF GROOVES IN BACKING BARS OR PLATES USED IN ARC WELDING OF MAGNESIUM ALLOYS

BASE-METAL THICKNESS		DEPTH OF GROOVE					
		GTAW		GMAW			
				NO ROOT OPENING$^{(A)}$		ROOT OPENING	
mm	in.	mm	in.	mm	in.	mm	in.
0.64	0.025	0.38	0.015	0.51	0.020	0.51	0.020
1.02	0.040	0.51	0.020	0.76	0.030	0.51	0.020
1.60	0.063	0.64	0.025	1.02	0.040	0.76	0.030
2.29	0.090	0.76	0.030	1.52	0.060	1.02	0.040
3.18	0.125	0.76	0.030	1.78	0.070	1.02	0.040
4.06	0.160	1.02	0.040	1.78	0.070	1.27	0.050
4.83	0.190	1.02	0.040	1.78	0.070	1.27	0.050
6.35	0.250	1.27	0.050	1.78	0.070	1.52	0.060
9.53	0.375	1.52	0.060	2.03	0.080	1.52	0.060

(A) WITH NO ROOT OPENING, GROOVES IN BACKING BARS OR PLATES ARE DEEPER TO PERMIT BETTER BALANCE BETWEEN TOP AND BOTTOM WELD REINFORCEMENT. USE OF A ROOT OPENING PERMITS BALANCING REINFORCEMENT WITH A SHALLOWER GROOVE DEPTH.

Welding of Magnesium Alloys

Surface Preparation

Magnesium alloys are usually supplied with either an oil coating, an acid pickled surface, or a chromate coated surface. Surfaces and edges must be cleaned just before welding to remove oxides and dirt picked up during forming, assembly, and fixturing. Electrode wire or filler metal must be mechanically or chemically cleaned. A partly used spool of electrode wire or filler metal should be properly stored to protect it from dirt.

Mechanical cleaning with aluminum or stainless steel wool, aluminum oxide abrasive cloth, or power wire brushes with stainless steel bristles is preferred for most production jobs. In shops where chemical finishing equipment is available, a cleaning bath of 190 mL/L (24 oz/gal) chromic acid, 40 mL/L ($5\frac{1}{3}$ oz/gal) ferric nitrate, 0.5 mL/L ($\frac{1}{16}$ oz/gal) potassium fluoride, and water to make 3.8 L (1 gal) can be used. Parts are dipped in this bath, which is maintained at 20 to 30 °C (70 to 90 °F), for about 3 min, then are rinsed thoroughly in hot water and dried in air. Ceramic or stainless steel tanks are preferred; tanks lined with lead, synthetic rubber, or a vinyl-based material are also acceptable.

Welding of Magnesium Alloys

Preheating

The need for preheating castings is determined largely by section thickness and amount of restraint. Thick sections, particularly if the magnitude of joint restraint is small, seldom need preheating. Thin sections and highly restrained joints often require preheating to prevent weld cracking, particularly in the high-zinc alloys. Maximum preheat temperatures for all of the common cast magnesium alloys are given in Table 7. In practice, if preheating is used, the temperature is usually less than the maximum shown in Table 7.

TABLE 7 PREHEAT TEMPERATURES AND POSTWELD HEAT TREATMENTS FOR MAGNESIUM ALLOY CASTINGS

ALLOY	TEMPER OF ALLOY(A)		MAXIMUM PREHEAT TEMPERATURE(B)		POSTWELD HEAT TREATMENT(C)
	BEFORE WELDING	AFTER TREATMENT	°C	°F	
AZ63A	T4	T4	380	720	$\frac{1}{2}$ H AT 390 °C (730 °F)
	T4 OR T6	T6	380	720	$\frac{1}{2}$ H AT 390 °C (730 °F) + 5 H AT 220 °C (425 °F)
	T5	T5	260(D)	500(D)	5 H AT 220 °C (425 °F)
AZ81A	T4	T4	400	750	$\frac{1}{2}$ H AT 415 °C (780 °F)
AZ91C	T4	T4	400	750	$\frac{1}{2}$ H AT 415 °C (780 °F)
	T4 OR T6	T6	400	750	$\frac{1}{2}$ H AT 415 °C (780 °F) + 4 H AT 215 °C (420 °F)(E)
AZ92A	T4	T4	400	750	$\frac{1}{2}$ H AT 410 °C (770 °F)
	T4 OR T6	T6	400	750	$\frac{1}{2}$ H AT 410 °C (770 °F) + 4 H AT 260 °C (500 °F)
AM100A	T6	T6	400	750	$\frac{1}{2}$ H AT 415 °C (780 °F) + 5 H AT 220 °C (425 °F)
EK30A	T6	T6	260(D)	500(D)	16 H AT 205 °C (400 °F)
EK41A	T4 OR T6	T6	260(D)	500(D)	16 H AT 205 °C (400 °F)
	T5	T5	260(D)	500(D)	16 H AT 205 °C (400 °F)
EQ21	T4 OR T6	T6	300	570	1 H AT 505 °C (940 °F)(F) + 16 H AT

					200 °C (390 °F)
EZ33A	F OR T5	T5	260(D)	500(D)	2 H AT 345 °C (650 °F)(G) + 5 H AT 215 °C (420 °F)
HK31A	T4 OR T6	T6	260	500	1 H AT 315 °C (600 °F)(G) + 16 H AT 205 °C (400 °F)
HZ32A	F OR T5	T5	260	500	16 H AT 315 °C (600 °F)
K1A	F	F	NONE	NONE	NONE
QE22A	T4 OR T6	T6	260	500	8 H AT 530 °C (985 °F)(F) + 8 H AT 205 °C (400 °F)
WE43	T4 OR T6	T6	300	570	1 H AT 510 °C (950 °F)(F) + 16 H AT 250 °C (480 °F)
WE54	T4 OR T6	T6	300	570	1 H AT 510 °C (950 °F)(F)(H) + 16 H AT 250 °C (480 °F)(I)
ZC63	F OR T4	T6	250	480	1 H AT 425 °C (795 °F)(F) + 16 H AT 180 °C (355 °F)(I)
ZE41A	F OR T5	T5	315	600	2 H AT 330 °C (625 °F) + 16 H AT 175 °C (350 °F)(G)
ZH62A	F OR T5	T5	315	600	2 H AT 330 °C (625 °F) + 16 H AT 175 °C (350 °F)
ZK51A	F OR T5	T5	315	600	2 H AT 330 °C (625 °F) + 16 H AT 175 °C (350 °F)(G)
ZK61A	F OR T5	T5	315	600	48 H AT 150 °C (300 °F)
	T4 OR T6	T6	315	600	2-5 H AT 500 °C (930 °F) + 48 H AT 130 °C (265 °F)

(A) T4, SOLUTION HEAT TREATED; T6, SOLUTION HEAT TREATED AND ARTIFICIALLY AGED; T5, ARTIFICIALLY AGED; F, AS CAST. "AFTER TREATMENT" MEANS AFTER POSTWELD HEAT TREATMENT.

(B) HEAVY AND UNRESTRAINED SECTIONS USUALLY NEED NO PREHEAT; THIN AND RESTRAINED SECTIONS MAY NEED PREHEATING, UP TO MAXIMUM TEMPERATURES SHOWN, TO AVOID WELD CRACKING. A SULFUR DIOXIDE OR CARBON DIOXIDE ATMOSPHERE IS RECOMMENDED WHEN TEMPERATURE EXCEEDS 370 °C (700 °F).

(C) TEMPERATURES SHOWN ARE MAXIMUM ALLOWABLE. FURNACE CONTROLS SHOULD BE SET SO TEMPERATURE DOES NOT EXCEED INDICATED MAXIMUM. A SULFUR DIOXIDE OR CARBON DIOXIDE ATMOSPHERE IS RECOMMENDED WHEN TEMPERATURE EXCEEDS 370 °C (700 °F).

(D) FOR $1\frac{1}{2}$ H MAX.

(E) 16 H AT 170 °C (335 °F) MAY BE USED INSTEAD OF 4 H AT 215 °C (420 °F).

(F) QUENCH IN WATER AT 60 TO 105 °C (140 TO 220 °F) BEFORE SECOND HEAT TREATMENT.

(G) THIS PHASE OF HEAT TREATMENT IS OPTIONAL AND SERVES TO INDUCE GREATER STRESS RELIEF. SOME LOSS IN HIGH-TEMPERATURE CREEP STRENGTH MAY OCCUR IN EZ33A DUE TO THE 345 °C (650 °F) STRESS RELIEF.

(H) AIR COOL AFTER QUENCHING.

(I) AIR COOL AFTER SECOND HEAT TREATMENT

Welding of Magnesium Alloys

Gas-Metal Arc Welding

Power supplies used for welding magnesium alloys furnish DCEP. Constant-voltage machines must be used for short circuiting welding and are generally preferred for spray arc welding. Constant-current (drooping volt-ampere output) machines can be used for spray arc welding and are sometimes advantageous for welding that is performed at a current level near the minimum for spray transfer because weld spatter is less. Special constant-voltage machines that are designed to pulse the current output must be used in pulsed arc welding. For detailed information on power supplies and modes of metal transfer, see the article "Power Sources" and "Transfer of Heat and Mass to the Base Metal in Gas-Metal Arc Welding" in this Volume.

Electrode holders, wire feeders, and related equipment used for GMAW of magnesium alloys are generally the same as those used for welding other metals.

Metal Transfer. Three modes of metal transfer are suitable for welding magnesium alloys: short circuiting, pulsed arc, and spray transfer. Pulsed arc transfer can be achieved only with a power supply designed to produce a pulsing secondary current. Without the pulsing feature, the type of transfer that would be obtained in the specific operating range of current, wire feed, and wire size for this process would be globular, which is not suitable for welding magnesium alloys.

Each of the three modes of transfer is best suited to a specific (sometimes overlapping) range of base-metal thickness and welding current, as shown in Table 8. Curves that show the relationship among current, electrode wire size, and wire-feed rate for each mode are presented in Fig. 1.

TABLE 8 TYPICAL OPERATING CONDITIONS FOR GAS-METAL ARC BUTT WELDING OF MAGNESIUM ALLOYS AT WELDING SPEED OF 610 TO 915 MM/MIN (24 TO 36 IN./MIN)

BASE-METAL THICKNESS		TYPE OF GROOVE	NO. OF PASSES	ELECTRODE						CURRENT, A	VOLTAGE, V	ARGON FLOW RATE	
				DIAMETER		FEED RATE		CONSUMPTION OF WELD					
mm	in.			mm	in.	m/min	in./min	kg/m	lb/ft			m^3/h	ft^3/h
SHORT CIRCUITING MODE OF METAL TRANSFER													
0.64	0.025	SQUARE(A)	1	1.02	0.040	3.6	140	0.009	0.006	25	13	1.1-1.7	40-60
1.02	0.040	SQUARE(A)	1	1.02	0.040	5.8	230	0.013	0.009	40	14	1.1-1.7	40-60
1.60	0.063	SQUARE(A)	1	1.60	0.063	4.7	185	0.027	0.018	70	14	1.1-1.7	40-60
2.29	0.090	SQUARE(A)	1	1.60	0.063	6.2	245	0.036	0.024	95	16	1.1-1.7	40-60
3.18	0.125	SQUARE(B)	1	2.39	0.094	3.4	135	0.045	0.030	115	14	1.1-1.7	40-60
4.06	0.160	SQUARE(B)	1	2.39	0.094	4.2	165	0.055	0.037	135	15	1.1-1.7	40-60
4.83	0.190	SQUARE(B)	1	2.39	0.094	5.2	205	0.068	0.046	175	15	1.1-1.7	40-60
PULSED ARE MODE OF METAL TRANSFER(C)													
1.60	0.063	SQUARE(A)	1	1.02	0.040	9.1	360	0.021	0.014	50	21	1.1-1.7	40-60
3.18	0.125	SQUARE(A)	1	1.60	0.063	7.1	280	0.042	0.028	110	24	1.1-1.7	40-60

4.83	0.190	SQUARE(A)	1	1.60	0.063	12.1	475	0.070	0.047	175	25	1.1-1.7	40-60
6.35	0.250	SINGLE-V, 60°(D)	1	2.39	0.094	7.4	290	0.097	0.065	210	29	1.1-1.7	40-60
SPRAY ARC MODE OF METAL TRANSFER(E)													
6.35	0.250	SINGLE-V, 60°(D)	1	1.60	0.063	13.5	530	0.062	0.042	240	27	1.4-2.3	50-80
9.53	0.375	SINGLE-V, 60°(D)	1	2.39	0.094	7.2-7.9	285-310	0.085	0.057	320-350	24-30	1.4-2.3	50-80
12.7	0.500	SINGLE-V, 60°(D)	2	2.39	0.094	8.1-9.1	320-360	0.158	0.106	360-400	24-30	1.4-2.3	50-80
15.9	0.625	DOUBLE-V, 60°(F)	2	2.39	0.094	8.4-9.4	330-370	0.186	0.125	370-420	24-30	1.4-2.3	50-80
25.4	1.000	DOUBLE-V, 60°(F)	4	2.39	0.094	8.4-9.4	330-370	0.310	0.208	370-420	24-30	1.4-2.3	50-80

(A) ZERO ROOT OPENING.

(B) ROOT OPENING, 2.3 MM (0.090 IN.).

(C) PULSE VOLTAGE OF 55 V, EXCEPT FOR METAL 4.83 MM (0.190 IN.) THICK, WHICH USES A PULSE VOLTAGE OF 52 V.

(D) 1.6 MM ($\frac{1}{16}$ IN.) ROOT FACE AND ZERO ROOT OPENING.

(E) SETTINGS ALSO APPLY TO FILLET WELDS IN THE SAME THICKNESS OF METAL.

(F) 3.2 MM ($\frac{1}{8}$ IN.) ROOT FACE AND ZERO ROOT OPENING

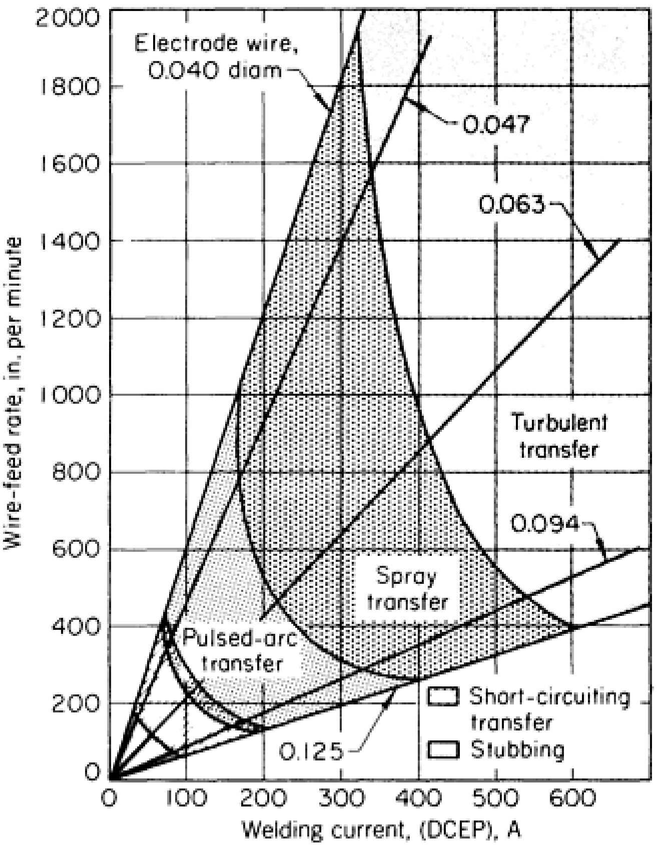

FIG. 1 PLOT OF WIRE-FEED RATE VERSUS WELDING CURRENT AS A FUNCTION OF ELECTRODE WIRE DIAMETER AND MODE OF METAL TRANSFER

Welding positions are restricted to flat, horizontal, and vertical-up by the high deposition rate and fluidity of the weld metal.

Typical operating conditions for gas metal arc butt welding of magnesium alloys in the thickness range from 0.64 to 25 mm (0.025 to 1 in.) are presented in Table 8. These operating conditions are intended as a guide and should be altered to the extent required by experience with specific applications.

Welding of Magnesium Alloys

Gas-Tungsten Arc Welding

Gas-tungsten arc welding is used more extensively than GMAW for joining magnesium alloys. It is well suited for welding thin sections. Control of heat input and the molten weld pool is better with GTAW than GMAW. Alternating current machines with a high-frequency current superimposed on the normal welding current for arc stabilization and direct current machines with continuous amperage control are used. For thin sheets, both alternating current and DCEP are used. On material over 4.8 mm ($\frac{3}{16}$ in.) thick, alternating current is preferred because it provides deeper penetration. Direct current electrode negative is seldom used on magnesium alloys because the arc lacks cathodic cleaning action. Alternating current machines should be equipped with a primary contactor, actuated by a switch on the torch or by a foot switch, for starting and stopping the arc. Otherwise, the arcing that occurs while the electrode approaches or draws away from the workpiece may result in burned spots on the work. Pure tungsten, zirconiated, and thoriated electrodes, from 0.25 to 6.35 mm (0.010 to 0.250 in.) in diameter, are used for GTAW of magnesium alloys. Additional information on power supplies, electrodes, and related items for GTAW is given in the article "Gas-Tungsten Arc Welding" in this Volume.

Manual Welding. Data on current settings, electrode diameter, shielding-gas flow, filler-metal diameter, and filler-metal consumption for manual GTAW of butt joints in magnesium alloys 1.0 to 12.7 mm (0.040 to 0.500 in.) thick are given in Table 9. For best results, the electrode should be held close to the work to produce an arc about 0.8 mm ($\frac{1}{32}$ in.) long. The preferred angles of torch, joint, and filler metal are given in Fig. 2. Welding should be done at a uniform speed in a straight line. A weaving or rotary motion should be used only for fillet welds or large corner joints. Forehand welding is preferred; a minimal number of stops produces the best results. If stops are required, the weld should be restarted on weld metal about 13 mm ($\frac{1}{2}$ in.) from the end of the previous weld.

TABLE 9 TYPICAL PARAMETERS FOR MANUAL GTAW OF BUTT JOINTS WITH MAGNESIUM ALLOYS

BASE-METAL THICKNESS		JOINT DESIGN(A)	NO. OF PASSES	ELECTRODE DIAMETER		AC CURRENT(B), A	ARGON FLOW RATE		FILLER METAL			
									DIAMETER		CONSUMPTION	
mm	in.			mm	in.		m^3/h	ft^3/h	mm	in.	kg/m	lb/ft
1.02	0.040	A	1	1.6	$\frac{1}{16}$	35	0.34	12	2.4	$\frac{3}{32}$	0.006	0.004
1.60	0.063	A	1	2.4	$\frac{3}{32}$	50	0.34	12	2.4	$\frac{3}{32}$	0.007	0.005
2.03	0.080	A	1	2.4	$\frac{3}{32}$	75	0.34	12	2.4	$\frac{3}{32}$	0.009	0.006
2.54	0.100	A	1	2.4	$\frac{3}{32}$	100	0.34	12	2.4	$\frac{3}{32}$	0.012	0.008
3.18	0.125	A	1	2.4	$\frac{3}{32}$	125	0.34	12	3.2	$\frac{1}{8}$	0.013	0.009
4.83	0.190	A	1	3.2	$\frac{1}{8}$	160	0.42	15	3.2	$\frac{1}{8}$	0.016	0.011
6.35	0.250	B	2	4.0	$\frac{5}{32}$	175	0.57	20	3.2	$\frac{1}{8}$	0.039	0.026
9.53	0.375	B	3	4.0	$\frac{5}{32}$	175	0.57	20	4.0	$\frac{5}{32}$	0.085	0.057
9.53	0.375	C	2	4.8	$\frac{3}{16}$	200	0.57	20	3.2	$\frac{1}{8}$	0.036	0.024
12.7	0.500	C	2	4.8	$\frac{3}{16}$	250	0.57	20	3.2	$\frac{1}{8}$	0.070	0.047

(A) A, SQUARE GROOVE BUTT JOINT, ZERO ROOT OPENING; B, 60° BEVEL, SINGLE-V-GROOVE BUTT JOINT 1.6 MM ($\frac{1}{16}$ IN.) ROOT FACE, ZERO ROOT OPENING; C, 60° BEVEL, DOUBLE-V-GROOVE BUTT JOINT 4 MM ($\frac{3}{32}$ IN.) ROOT FACE, ZERO ROOT OPENING.

(B) THORIUM-CONTAINING ALLOYS WILL REQUIRE ABOUT 20% HIGHER CURRENT. WITH HELIUM SHIELDING, REQUIRED WELDING CURRENT WILL BE REDUCED BY ABOUT 20 TO 30 A.

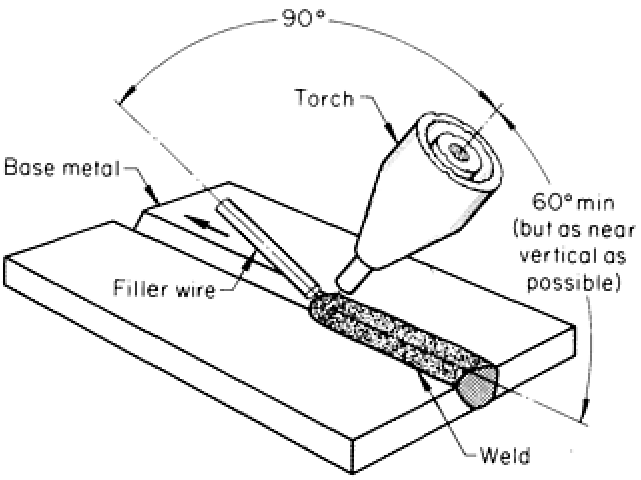
90°
Torch
Base metal
60° min
(but as near
vertical as
possible)
Filler wire
Weld

FIG. 2 RECOMMENDED GTAW SETUP FOR JOINING TWO MAGNESIUM ALLOY PLATES

To minimize the risk of weld cracking, either of the methods shown in Fig. 3 should be used: method A, in which starting and runoff plates (or tabs) are used to start and end the weld; or method B, in which the weld is made in two increments by passes that begin in the middle and progress toward opposite ends. Also, the base metal (and fixture, if used) should be preheated to at least 95 to 150 °C (200 to 300 °F). If the thickness of the two compositions to be welded differs by 6.4 mm ($\frac{1}{4}$ in.) or more, the thicker section should be preheated to approximately 150 °C (300 °F).

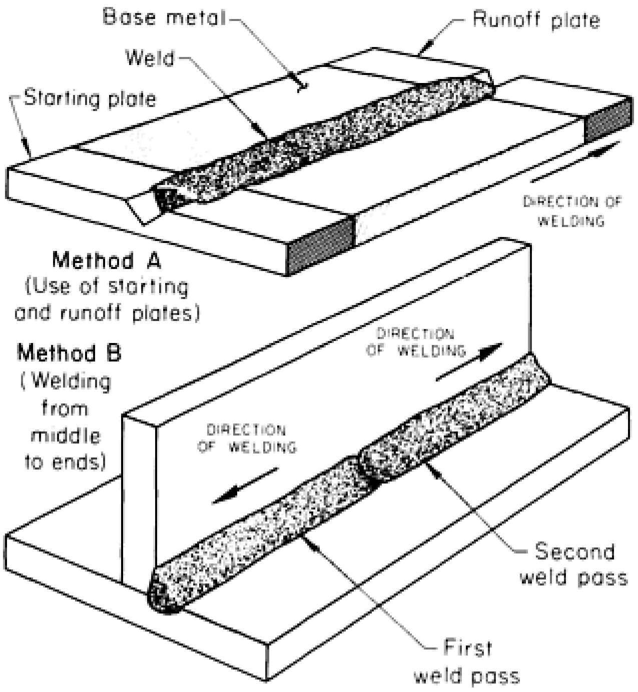
Base metal
Runoff plate
Weld
Starting plate
DIRECTION OF
WELDING
Method A
(Use of starting
and runoff plates)
Method B
(Welding
from
middle
to ends)
DIRECTION
OF WELDING
DIRECTION
OF WELDING
Second
weld pass
First
weld pass

FIG. 3 TECHNIQUES TO PREVENT CRACKING OF WELD WHEN JOINING MAGNESIUM ALLOY PLATES

Short-Run Production. In short production runs for which the cost of elaborate fixtures cannot be justified, tooling costs can be minimized by fabricating the components so that they can be tack welded into position.

Example 1: Tack Welding for Short-Run Production.

To minimize tooling costs on short production runs of electronic deck assemblies, which were essentially two rectangular boxes 50 by 50 by 100 mm (2 by 2 by 4 in.), as shown in Fig. 4, tack welds were used instead of fixturing to position some of the component pieces. Formed sheet sections of 1.3 mm (0.050 in.) thick AZ31B-H24 were tack welded into position by GTAW, using 1.6 mm ($\frac{1}{16}$ in.) diam ER AZ61A filler wire. The tack welds were 3.2 mm ($\frac{1}{8}$ in.) long and were spaced on 50 mm (2 in.) centers (starting at each comer). A tool plate and toggle clamps held the pieces for tack welding. Tack welds were not used to hold angle pieces.

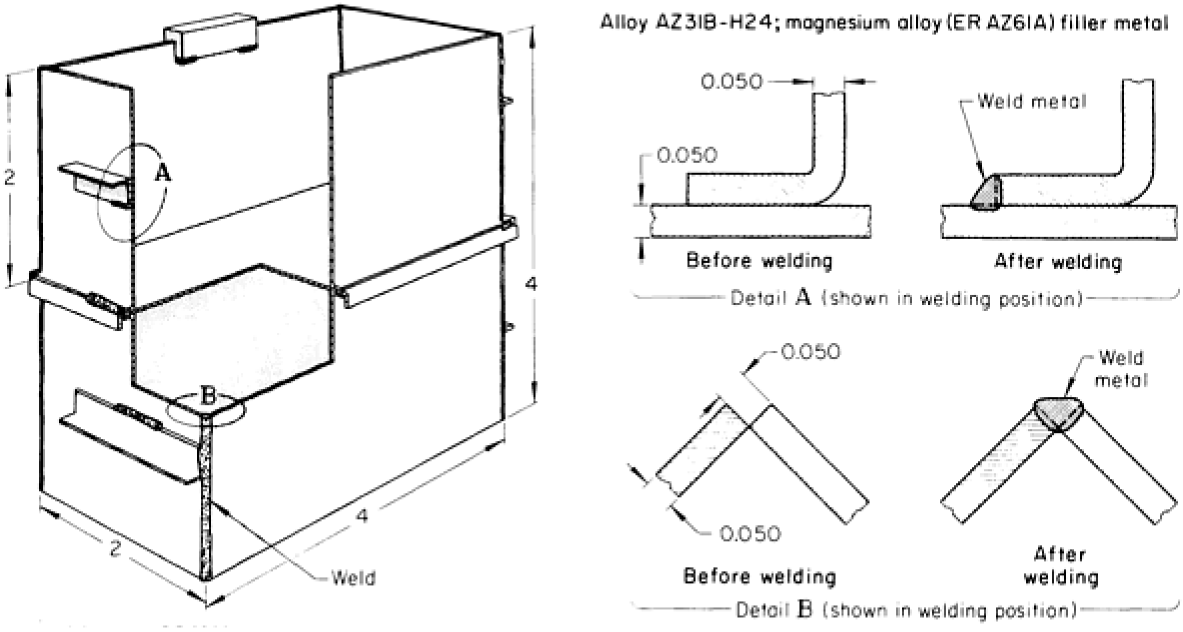
2
A
4
B
4
2
Weld
Alloy AZ31B-H24; magnesium alloy (ER AZ61A) filler metal
0.050
0.050
Weld metal
Before welding
After welding
Detail A (shown in welding position)
0.050
0.050
Weld metal
Before welding
After welding
Detail B (shown in welding position)

CONDITIONS FOR MANUAL GTAW	
JOINT TYPES	LAP AND CORNER
WELD TYPES	FILLET AND SINGLE-V-GROOVE
WELDING POSITIONS	HORIZONTAL AND FLAT
PREWELD CLEANING	WIRE BRUSHING
PREHEAT	NONE
FIXTURES	TOOL PLATE AND TOGGLE CLAMPS
SHIELDING GAS	HELIUM, AT 0.71 M^3/H (25 FT^3/H)
ELECTRODE	1.02 MM (0.040 IN.) DIAM EWP
FILLER METAL	1.6 MM ($\frac{1}{16}$ IN.) DIAM ER AZ61A(A)
TORCH	350 A, WATER COOLED(B)
POWER SUPPLY	300 A TRANSFORMER(C)
CURRENT, FILLET WELDS	25 A, AC
CURRENT, V-GROOVE WELDS	40 A, AC
POSTWELD HEAT TREATMENT	175 °C (350 °F) FOR $3\frac{1}{2}$ H

(A) 915 MM (36 IN.) LONG ROD.

(B) CERAMIC NOZZLE.

FIG. 4 MANUAL GTAW OF ELECTRONIC DECK ASSEMBLY

Welding of the assembled and tack welded components was completed by manual GTAW with 1.6 mm ($\frac{1}{16}$ in.) diam ER AZ61A filler wire. The comer joints were welded with continuous beads about 50 mm (2 in.) long, and the flanged bottom of the top part of the assembly was joined to the sides with 25 mm (1 in.) long fillet welds. Extruded angle sections were fillet welded to the ends of the boxes with welds about 25 mm (1 in.) long (Fig. 4, detail A).

The assembly was repositioned manually so that all welds could be made in either the flat or the horizontal position. A standard alternating current power supply with a high-frequency arc stabilizer was used. Helium was selected as the shielding gas because a hotter and more stable arc was produced than would have been possible with argon shielding gas. Preheating was not used, but after welding, the assemblies were stress relieved at 175 °C (350 °F) for $3\frac{1}{2}$ h to prevent stress-corrosion cracking. Welds were inspected visually.

Example 2: Manually Welded T-Joint.

The welding of a single 3.2 mm ($\frac{1}{8}$ in.) fillet in a 205 mm (8 in.) long T-joint between sheets of alloy AZ31B, 1.6 and 3.2 mm ($\frac{1}{16}$ and $\frac{1}{8}$ in.) thick (see Table 10), was used as a quick shop test for welder qualification. The welder adjusted the machine settings, gas flow, and welding speed to produce a sound weld with proper contour and adequate penetration, using manual GTAW. After welding, the joint was broken by striking a blow against the unwelded side of the vertical stem to open and expose the root of the weld. The fracture was then examined for depth of penetration, porosity, lack of fusion, and other defects. Welding conditions and a view of the joint are given in Table 10.

TABLE 10 MANUAL GTAW OF ALLOY AZ31B

Argon shielding gas; flow rate, 0.51 m^3/h (18 ft^3/h)

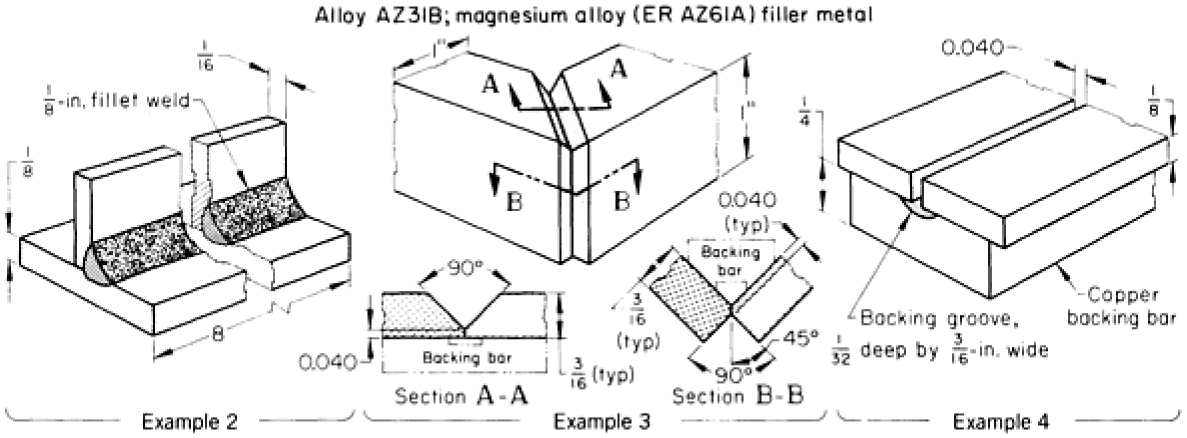
Alloy AZ31B; magnesium alloy (ER AZ61A) filler metal
1/16
1/8-in. fillet weld
1/8
8
Example 2
1"
A
A
1"
B
B
90°
0.040
Backing bar
3/16 (typ)
Section A-A
0.040 (typ)
Backing bar
3/16 (typ)
45°
90°
Section B-B
Example 3
0.040
1/4
1/8
Backing groove, 1/32 deep by 3/16-in. wide
Copper backing bar
Example 4

ITEM	EXAMPLE 2	EXAMPLE 3	EXAMPLE 4
JOINT TYPE	T	CORNER; BUTT	BUTT
WELD TYPE	SINGLE FILLET	SINGLE-V-GROOVE	SQUARE GROOVE
WELDING POSITION	HORIZONTAL	VERTICAL-UP; FLAT	FLAT
ELECTRODE (EWP) DIAMETER, MM (IN.)	2.4 ($\frac{3}{32}$)	3.2 ($\frac{1}{8}$)	3.2 ($\frac{1}{8}$)
FILLER METAL (ER AZ61A) DIAMETER, MM (IN.)	1.6 ($\frac{1}{16}$)	2.4 ($\frac{3}{32}$)	1.6 ($\frac{1}{16}$)
CURRENT (AC, HF-STABILIZED)[(A)]	110 A	125 A	135 A
WELDING SPEED	255 MM/MIN (10 IN./MIN)	5 UNITS/H[(B)]	255 MM/MIN (10 IN./MIN)

POSTWELD HEAT TREATMENT	260 °C (500 °F) FOR 15 MIN	175 °C (350 °F) FOR 1 $\frac{1}{2}$ H	175 °C (350 °F) FOR 1 $\frac{1}{2}$ H

(A) A 300 A AC/DC POWER SUPPLY WITH CONTINUOUS HIGH FREQUENCY WAS USED FOR ALL THREE APPLICATIONS, WITH A LIGHT-DUTY WATER-COOLED WELDING TORCH. ALL WORKPIECES WERE PREWELD CLEANED BY CHROMIC-SULFURIC PICKLING. NO PREHEAT WAS USED.

(B) INCLUDES FIXTURING, WELDING, AND UNLOADING

Example 3: Manually Welded Single-V-Groove Butt and Corner Joints.

Structural frames were fabricated by welding the mitered ends of 25 by 25 by 4.8 mm (1 by 1 by $\frac{3}{16}$ in.) angles extruded from alloy AZ31B. One of the four right angle corners of one of these structural frames is shown in Table 10. Manual GTAW was used because production quantity was less than 50 pieces per month.

The procedure consisted of mitering, edge beveling, cleaning, assembling on a fixture, and welding. Horizontal joint edges were beveled 45° with a 1.0 mm (0.040 in.) deep root face; vertical joint edges were beveled to form a similar root face for a 90° corner joint. After acid pickling and rinsing, the angles were assembled on a clamping fixture with flat backing bars for the horizontal joints and corner backing bars for the vertical joints. The backing bars for both the horizontal and vertical joints were provided with backing grooves that were 3.2 mm ($\frac{1}{8}$ in.) wide by 0.51 mm (0.020 in.) deep. The outside corner joint was single-pass welded, vertical-up; the grooved butt joint was single-pass welded in the flat position. No preheat was used, but the frames were postweld stress relieved at 175 °C (350 °F) for 1$\frac{1}{2}$ h. Welding conditions are given in Table 10.

Example 4: Square-Groove-Welded Butt Joints.

Gas-metal arc welding was to be used on the square-groove-welded butt joints (with 1.0 mm, or 0.040 in., root opening) in 3.2 mm ($\frac{1}{8}$ in.) thick AZ31B-H24 sheet, but contact tube arcing occurred and contamination of the filler metal resulted. Rather than delay production while the arcing problem was being eliminated, manual GTAW was used instead. Good mechanical properties were obtained in single-pass welds, using argon shielding on the torch side and a grooved copper backing bar, but no gas backing, on the underside. Joint alignment was maintained with bolted hold-down bars. Welding conditions are given in Table 10.

Automatic welding of magnesium alloys by GTAW is similar to manual welding, except that higher currents and welding speeds are used. Table 11 can be used as a guide to determine settings for automatic welding. Alternating current is best, although DCEP can be used. A balanced-wave alternating current machine, or a conventional alternating current machine equipped with a battery bias for wave balancing, should be used.

TABLE 11 OPERATING PARAMETERS FOR AUTOMATIC GTAW OF AZ31B MAGNESIUM ALLOY BUTT JOINTS

Helium shielding gas; flow rate, 0.6 m^3/h (20 ft^3/h)

TYPE OF CURRENT	CURRENT A	DIAMETER OF TUNGSTEN ELECTRODE		ARC LENGTH		ER AZ61A OR ER AZ92A FILLER METAL, 1.60 MM (0.063 IN.) DIAM				TRAVEL SPEED	
						FEED RATE		CONSUMPTION OF WELD			
		mm	in.	mm	in.	m/min	in./min	kg/m	lb/ft	mm/min	in./min
AZ31B, 1.60 MM (0.063 IN.) THICK											
AC,	55	2.39	0.094	0.64	0.025	0.89	35	0.010	0.007	255	12

WAVE	70	3.18	0.125	0.64	0.025	1.37	54	0.006	0.004	915	36
	95	3.18	0.125	0.64	0.025	2.44	96	0.007	0.005	1145	45
	170	4.76	0.188	0.64	0.025	4.06	160	0.007	0.005	1780	70
	195	4.76	0.188	0.64	0.025	4.83	190	0.009	0.006	2030	80
	200	4.76	0.188	0.64	0.025	5.16	203	0.007	0.005	2415(A)	95(A)
DCEN	75	3.18	0.125	0.64	0.025	2.03	80	0.006	0.004	1220	48
DCEP	120	6.35	0.250	0.51	0.020	4.67	184	0.007	0.005	2030(A)	80(A)
AZ31B, 4.83 MM (0.190 IN.) THICK											
AC, BALANCED WAVE	300		0.250	0.51	0.020	4.04	159	0.016	0.011	865	34
DCEN	170		0.125	0.76	0.030	1.78	70	0.012	0.008	510	20
DCEP	120		0.250	0.51	0.020	0.25	10(B)	0.012	0.008(B)	180(A)	7(A)

(A) MAXIMUM SPEED LIMITATION BECAUSE OF UNDERCUTTING OR ARC INSTABILITY.

(B) DIAMETER OF FILLER METAL, 2.39 MM (0.094 IN.)

Constant alignment of electrode and filler metal is important in automatic welding. Filler metal is fed into the arc at a low angle to the work so that the filler rod touches the weld surface just ahead of the electrode. The arc length shown in Table 11 should be preset and maintained. Travel speeds up to 2410 mm/min (95 in./min) can be used on thin sheet, but 610 to 915 mm/min (24 to 36 in./min) is more common. Automatic welding is not necessarily restricted to high-production applications; sometimes the high degree of control that can be attained with automatic welding is required regardless of quantity.

Example 5: Automatic Welding of Extrusions.

Airtight doors for an aerospace application were made by welding panels of alloy AZ31B-H24 sheet to frames extruded from AZ31B. The frames, which acted as stiffeners, also contained a groove for an air seal. Cross sections of similar offset butt joints in two designs of door assemblies are shown as joints A and B in Fig. 5. The offset lip of the extruded frames provided a single-bevel groove butt joint and supplied backing for the weld; the lap joint on the underside was not welded. Although production quantities were low, automatic GTAW was used because weld quality was good and the equipment was available. Automatic travel was obtained by mounting the welding equipment on the motorized carriage of a cutting machine. Differences in welding conditions for the two joints resulted from operator choice or judgment. Both procedures produced satisfactory welds, but the difference in welding speeds would have been significant had production quantities been large.

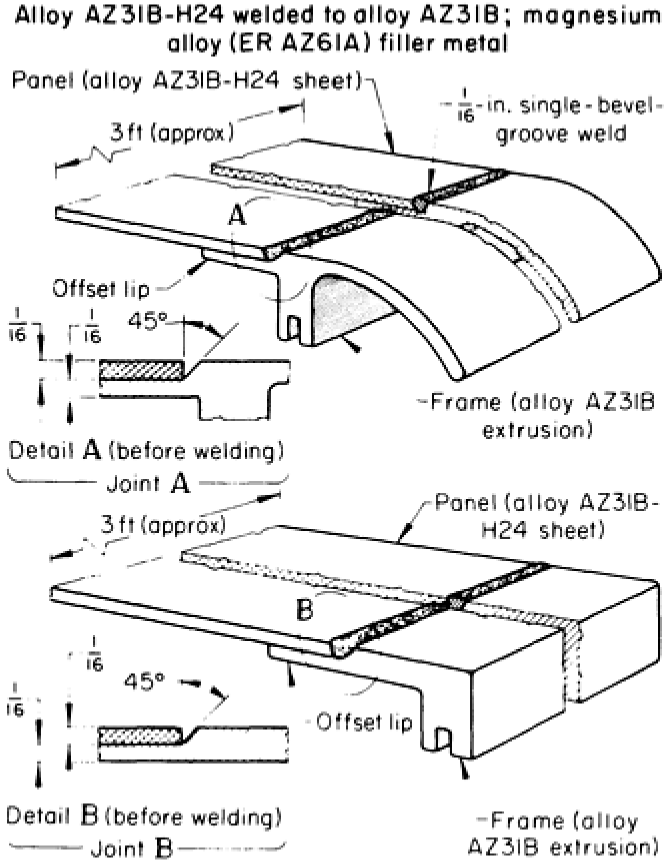
Alloy AZ31B-H24 welded to alloy AZ31B; magnesium alloy (ER AZ61A) filler metal
Panel (alloy AZ31B-H24 sheet)
3 ft (approx)
1/16-in. single-bevel-groove weld
A
Offset lip
1/16
1/16
45°
Frame (alloy AZ31B extrusion)
Detail A (before welding)
Joint A
3 ft (approx)
Panel (alloy AZ31B-H24 sheet)
B
1/16
45°
1/16
Offset lip
Detail B (before welding)
Joint B
Frame (alloy AZ31B extrusion)

AUTOMATIC GTAW	
JOINT TYPE	OFFSET BUTT
WELD TYPE	SINGLE-BEVEL GROOVE
PREWELD CLEANING	CHROMIC-SULFURIC PICKLE
WELDING POSITION	FLAT
PREHEAT	NONE
SHIELDING GAS	ARGON, 0.51 M^3/H (18 FT^3/H) FOR JOINT A; ARGON, 0.45 M^3/H (16 FT^3/H) FOR JOINT B
ELECTRODE	3.2 MM ($\frac{1}{8}$ IN.) DIAM EWP
FILLER METAL	1.6 MM ($\frac{1}{16}$ IN.) DIAM ER AZ61A
TORCH	WATER COOLED
POWER SUPPLY	300 A AC (HF-STABILIZED)
CURRENT (AC)	175 A FOR JOINT A; 135 A FOR JOINT B
WIRE-FEED RATE	1650 MM/MIN (65 IN./MIN)
TRAVEL SPEED	510 MM/MIN (20 IN./MIN) FOR JOINT A; 380 MM/MIN (15 IN./MIN FOR JOINT B)
POSTWELD HEAT TREATMENT	175 °C (350 °F) FOR $1\frac{1}{2}$ H

FIG. 5 AUTOMATIC GTAW OF AZ31B ALLOY SHEET AND EXTRUDED MATERIAL

Repair Welding of Castings

A significant portion of the total amount of welding that is done on magnesium alloy castings is for repair. Repair welding is usually limited to the repair of defects in clean metal, including broken sections, sand or blowholes, cracks, and cold shuts. Repair welding is not recommended in areas containing flux or oxide inclusions or excessive porosity. Castings that have been organically impregnated for pressure tightness or that may contain oil in pores should not be welded. Many magnesium castings are aircraft parts that are heat treated to meet strength requirements. These castings must be heat treated again after welding. The type, size, and location of defects in castings vary so much that each job presents its own welding problems; procedures cannot be standardized. However, to complete most repairs:

- CASTINGS MUST BE STRIPPED OF PAINT AND DEGREASED BEFORE WELDING; CONVERSION COATINGS SHOULD BE REMOVED FROM DEFECTIVE AREAS WITH STEEL WOOL OR WITH A POWER WIRE BRUSH HAVING STAINLESS STEEL BRISTLES; A ROTARY DEBURRING TOOL IS RECOMMENDED FOR REMOVING DEFECTS AND PREPARING THE WORKPIECE FOR WELDING.
- CASTINGS SHOULD BE CLAMPED IN PLACE, AND THE SURFACE ADJOINING THE FRACTURE SHOULD BE BEVELED FOR THE WELD (TABLE 5).
- PREHEATING, IF NECESSARY, SHOULD BE DONE AT THE TEMPERATURES RECOMMENDED IN TABLE 7. EITHER LOCAL HEATING WITH A TORCH OR FURNACE HEATING CAN BE USED. THE USE OF A PROTECTIVE ATMOSPHERE DURING FURNACE HEATING WILL REDUCE THE POSSIBILITY OF OXIDATION AT TEMPERATURES ABOVE 370 °C (700 °F).
- WELDING SHOULD BE DONE IMMEDIATELY AFTER PREHEATING. REHEATING MAY BE NECESSARY IF THE TEMPERATURE OF THE CASTING DROPS SIGNIFICANTLY BELOW THE PREHEAT TEMPERATURE.
- MEDIUM-SIZE WELD BEADS SHOULD BE USED.
- WELDING SHOULD PROGRESS FROM THE CENTER OF THE BREAK TO THE OUTSIDE EDGES. THE ARC SHOULD NOT BE ALLOWED TO DWELL TOO LONG IN ANY ONE AREA BECAUSE WELD CRACKING MAY RESULT.
- HIGH WELDING CURRENT, WHICH MAY CAUSE WELD CRACKING OR INCIPIENT MELTING IN THE HEAT-AFFECTED ZONE (HAZ), SHOULD BE AVOIDED; LOW WELDING CURRENTS, WHICH MAY CAUSE COLD LAPS, OXIDE CONTAMINATION, OR POROSITY, SHOULD ALSO BE AVOIDED.
- A FOOT CONTROL SHOULD BE USED TO FADE IN AND OUT THE ARC GRADUALLY TO MINIMIZE THERMAL SHOCK FROM ARC STOPS. THERMAL SHOCK MAY CAUSE CRACKING.

The examples that follow describe procedures that were used in successful repair welding of specific castings.

Example 6: Repair of an Aircraft Wheel Rim Casting.

After machining, aircraft wheel film castings of AZ91C-T4 were occasionally found to have small sand holes (Fig. 6). Defects, of a size that could be removed by cutting a groove about 3.2 mm ($\frac{1}{8}$ in.) deep and 13 mm ($\frac{1}{2}$ in.) wide, were repaired by welding. The defective area was cleaned by cutting a smoothly blended groove (see Fig. 6, Before welding), using a burring tool. After complete removal of defective material was verified by liquid-penetrant inspection, the groove and surrounding area were wire brushed to prepare for GTAW. The relatively thick section of the casting in the repair area and the wide groove for the deposit ensured that the weld would be under low restraint; therefore, preheating was not necessary. Ordinarily, a standard filler metal, such as ER AZ92A or ER AZ101A, would prove satisfactory for welding the AZ91C base metal; in this application, a filler metal of the same composition as the base metal was specified. The weld was made by striking the arc at the bottom of the groove and proceeding in a circular direction along the groove wall to fill the groove, as shown in the "After welding" view in Fig. 6.

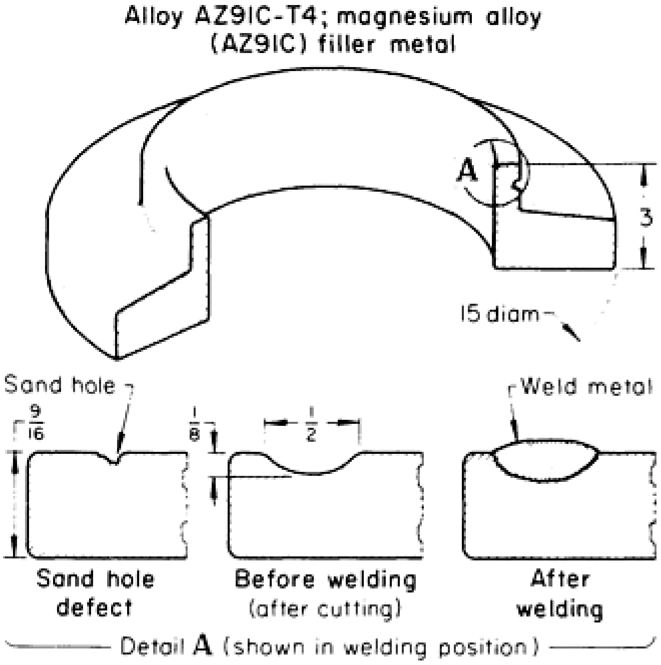
Alloy AZ91C-T4; magnesium alloy
(AZ91C) filler metal
A
3
15 diam
Sand hole
Weld metal
$\frac{9}{16}$
$\frac{1}{8}$
$\frac{1}{2}$
Sand hole defect
Before welding (after cutting)
After welding
Detail A (shown in welding position)

MANUAL GTAW	
WELD TYPE	SURFACING, FOR REPAIR
WELDING POSITION	FLAT
PREHEAT	NONE
SHIELDING GAS	HELIUM, 0.6 M^3/H (20 FT^3/H)
ELECTRODE	2.4 MM ($\frac{3}{32}$ IN.) DIAM EWP
FILLER METAL	3.2 MM ($\frac{1}{8}$ IN.) DIAM AZ91C
TORCH	300 A, WATER COOLED
POWER SUPPLY	300 A TRANSFORMER, WITH HIGH-FREQUENCY GENERATOR
CURRENT	60 A, AC
POSTWELD HEAT TREATMENT	415 °C (780 °F) FOR $\frac{1}{2}$ H PLUS 215 °C (420 °F) FOR 4 H

FIG. 6 REPAIR WELDING OF AZ91C ALLOY AIRCRAFT WHEEL RIM CASTING

After welding, the weld reinforcement was ground to 0.8 mm ($\frac{1}{32}$ in.) above the base-metal surface, and surface weld quality was verified by liquid-penetrant inspection. After passing inspection, the casting was heat treated to the T6 condition. The casting was accepted after radio-graphic inspection indicated satisfactory internal weld quality.

Example 7: Repair Welding of a Jet Engine Casting.

During an aircraft jet engine overhaul, fluorescent-penetrant inspection revealed a 64 mm ($2\frac{1}{2}$ in.) long crack near a rib in the cast AZ92A-T6 compressor housing shown in Fig. 7. The thickness of the section containing the crack ranged from

4.8 to 7.9 mm ($\frac{3}{16}$ to $\frac{5}{16}$ in.). Repair welding was permissible. The part was vapor degreased to remove surface grease and dirt and soaked in a commercial alkaline paint remover. The crack was then marked with a felt-tip marker, and the part was stress relieved at 205 °C (400 °F) for 2 h. The crack was removed by slotting the flange through to the periphery. Each side of the slot was beveled to approximately 30° from vertical to form a 60° double-V-groove. The area to be welded was cleaned with a power wire brush with stainless steel bristles. Welding was done by manual GTAW, without preheating.

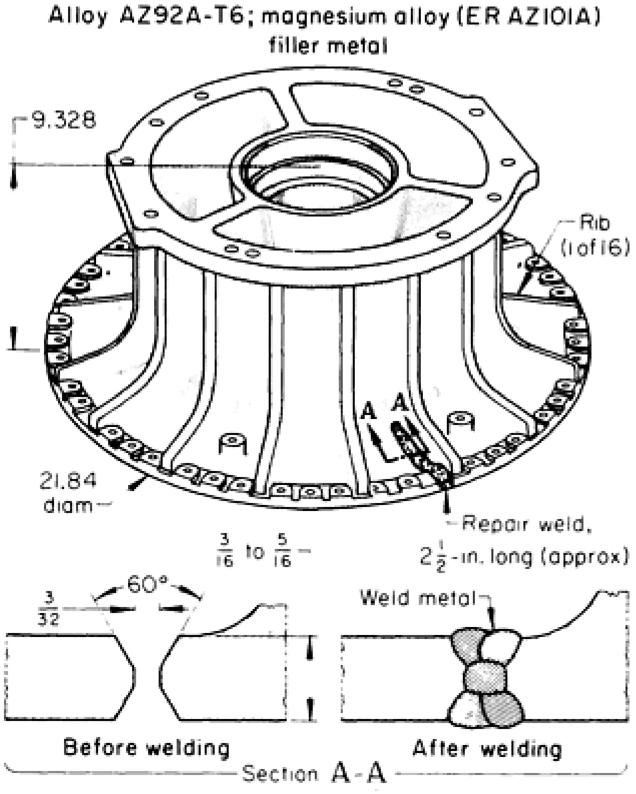
Alloy AZ92A-T6; magnesium alloy (ER AZ101A) filler metal
9.328
Rib (1 of 16)
A
A
21.84 diam
Repair weld, 2½-in. long (approx)
3/16 to 5/16
60°
3/32
Weld metal
Before welding
After welding
Section A-A

MANUAL GTAW	
JOINT TYPE	BUTT
WELD TYPE	60° DOUBLE-V-GROOVE REPAIR
SHIELDING GAS	ARGON, 0.6 M^3/H (20 FT^3/H)(A)
ELECTRODE	1.6 MM ($\frac{1}{16}$ IN.) DIAM EWTH-2
FILLER METAL	1.6 MM ($\frac{1}{16}$ IN.) DIAM ER AZ101A
TORCH	WATER COOLED
POWER SUPPLY	300 A TRANSFORMER, WITH HIGH-FREQUENCY STARTING
CURRENT	UNDER 70 A, AC(B)
POSTWELD STRESS RELIEF	205 °C (400 °F) FOR 2 H(C)
INSPECTION	FLUORESCENT PENETRANT

(A) ALSO USED FOR BACKING.

(B) CURRENT WAS REGULATED BY A FOOT SWITCH.

FIG. 7 REPAIR WELDING OF A CAST AZ92A ALLOY COMPRESSOR HOUSING

The welding technique maintained a low-amperage arc (less than 70 A) directed onto the base metal while filler metal was deposited on the sides of the groove, working from the innermost point outward. After a molten weld pool formed, the arc was weaved slightly while depositing a bead on the sides of the groove. During welding, heat input was adjusted by a foot-operated current-control rheostat to maintain a uniform weld pool. After welding was completed on one side of the slot, the casting was turned over. Excess drop-through and areas of incomplete penetration were removed by grinding. The underside was then welded by the same technique used for the first side. After welding, the casting was stress relieved at 205 °C (400 °F) for 2 h and inspected by the fluorescent-penetrant method.

Sand casting repairs can be made with manual GTAW. Figure 8 shows three repairs for the correction of surface flaws that occur in machining, defects resulting from the casting process, and deficiencies in design. Welding was done with high-frequency stabilized alternating current. The castings were preheated and maintained at temperature during welding to avoid cracking. To avoid distortion, the postweld heat treatments were used as a compromise between stress relieving and an aging treatment. However, when repair-welded areas were large enough so that the heat of welding caused a significant reduction in mechanical properties, welding was followed by a full heat treatment. Radiographic inspection was used in establishing the welding procedure, and all repair welds were subjected to fluorescent-penetrant inspection. Fillet welds were left as welded; other types of welds were finished to conform to the contour of the casting or to the desired shape.

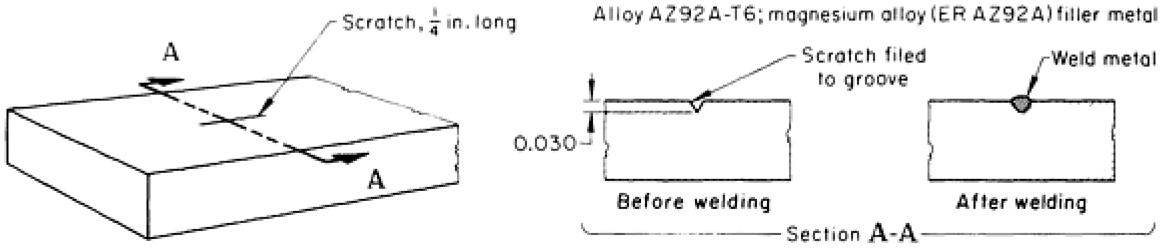
Scratch, $\frac{1}{4}$ in. long
A
A
Alloy AZ92A-T6; magnesium alloy (ER AZ92A) filler metal
Scratch filed to groove
Weld metal
0.030
Before welding
After welding
Section A-A

Example 8

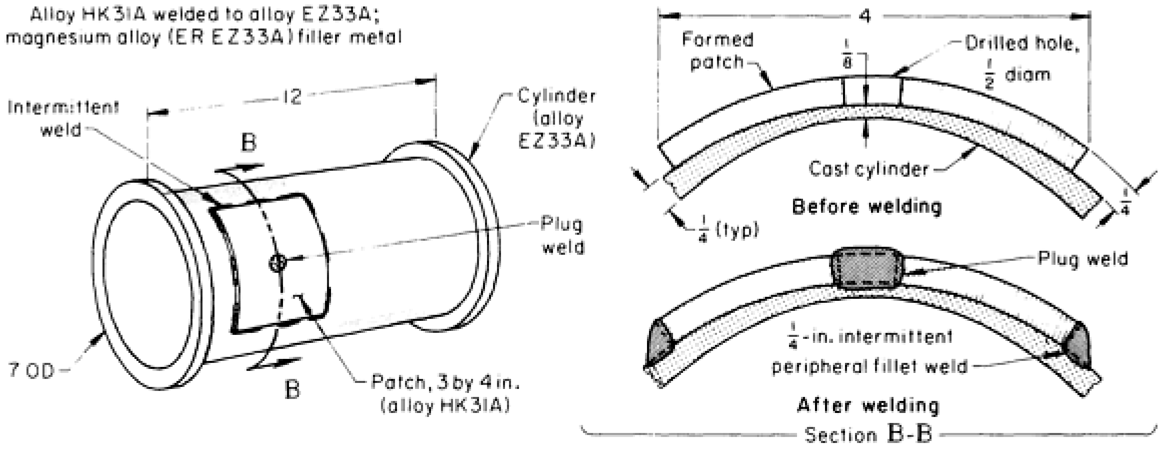
Alloy HK31A welded to alloy EZ33A; magnesium alloy (ER EZ33A) filler metal
Intermittent weld
12
B
Cylinder (alloy EZ33A)
Plug weld
7 OD
B
Patch, 3 by 4 in. (alloy HK31A)
4
Formed patch
$\frac{1}{8}$
Drilled hole, $\frac{1}{2}$ diam
Cast cylinder
Before welding
$\frac{1}{4}$ (typ)
$\frac{1}{4}$
Plug weld
$\frac{1}{4}$-in. intermittent peripheral fillet weld
After welding
Section B-B

Example 9

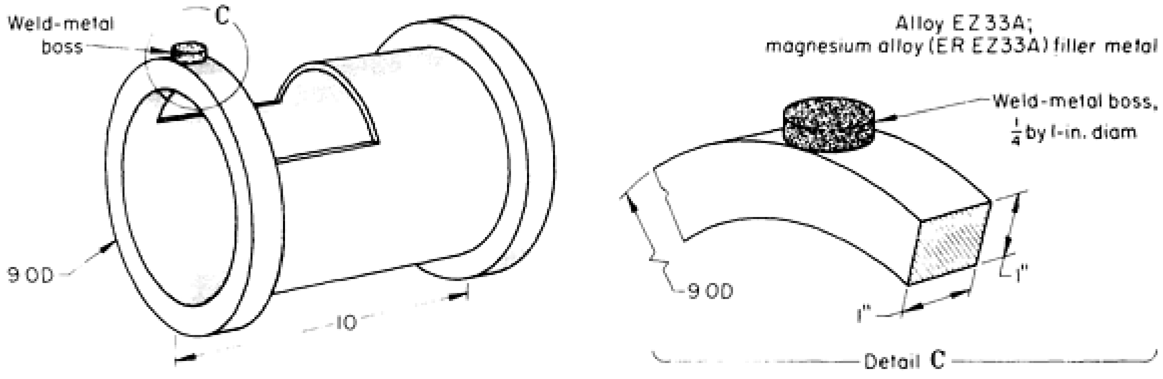
Weld-metal boss
C
9 OD
10
Alloy EZ33A; magnesium alloy (ER EZ33A) filler metal
Weld-metal boss, $\frac{1}{4}$ by 1-in. diam
9 OD
1"
1"
Detail C

Example 10

WELDING CONDITION	**EXAMPLE 8**	**EXAMPLE 9**	**EXAMPLE 10**
CONDITIONS FOR MANUAL GTAW$^{(A)}$			
FIXTURES	NONE	PLATES	NONE
WELDING POSITION	FLAT	FLAT	FLAT
SHIELDING GAS	ARGON; 0.57-0.68 M^3/H (20-24 FT^3/H)	ARGON; 0.68-0.79 M^3/H (24-28 FT^3/H)$^{(B)}$	ARGON; 0.68-0.79 M^3/H (24-28 FT^3/H)
ELECTRODE, 3.2 MM ($\frac{1}{8}$ IN.) DIAM	EWP	EWP	EWP
FILLER METAL, 1.6 MM ($\frac{1}{16}$ IN.) DIAM	ER AZ92A	ER EZ33A	ER EZ33A
CURRENT (AC, HF-STABILIZED)	80-140 A	160-180 A	160-180 A
PREHEAT AND INTERPASS	120 °C (250 °F)	150 °C (300 °F)	120-175 °C (250-

TEMPERATURE			350 °F)
POSTWELD HEAT TREATMENT	150 °C (300 °F) FOR 3 H	205 °C (400 °F) FOR 2 H	150 °C (300 °F) FOR 4 H
TOTAL TIME PER PIECE(C)	$\frac{1}{2}$ H	2 H	$\frac{1}{2}$ H

(A) FOR ALL THREE APPLICATIONS, PREWELD CLEANING WAS DONE WITH A WIRE BRUSH OR A ROTARY FILE, POWER SUPPLY WAS A 300 A TRANSFORMER WITH HIGH-FREQUENCY (BALANCED WAVE) STABILIZER, AND ELECTRODE HOLDER WAS A 300 A WATER-COOLED TYPE.

(B) HELIUM AT 0.23 M^3/H (8 FT^3/H) WAS USED FOR BACKING.

(C) NOT INCLUDING TIME FOR POSTWELD HEAT TREATMENT

FIG. 8 MANUAL GTAW REPAIR OF SAND CASTINGS

Example 8: Repair Welding of a Machining Scratch.

The views at the top in Fig. 8 show a portion of an alloy AZ92A-T6 sand casting with a machining scratch 6.4 mm ($\frac{1}{4}$ in.) long and 0.76 mm (0.030 in.) deep. This scratch was repaired by gas tungsten arc surface welding. Machining flaws up to 9.5 mm ($\frac{3}{8}$ in.) square in castings of various shapes and sizes have been corrected by this method.

Example 9: Repair Welding of Wall Thinning.

The middle views in Fig. 8 show a flanged cylindrical alloy EZ33A sand casting that was one of about 200 such castings in which core shift caused the 6.4 mm ($\frac{1}{4}$ in.) wall to thin to 3.2 mm ($\frac{1}{8}$ in.), beginning about 25 mm (1 in.) away from one of the flanges. To repair these thin areas without distortion of the cylinder, a 75 by 100 mm (3 by 4 in.) piece of 6.4 mm ($\frac{1}{4}$ in.) alloy HK31A plate was formed to fit the outside wall of the cylinder, drilled in the center with a 13 mm ($\frac{1}{2}$ in.) diam hole, and clamped in place over the thin area. The patch was tack welded at the four corners and then gas tungsten arc welded along the periphery with intermittent 6.4 mm ($\frac{1}{4}$ in.) fillet welds 25 mm (1 in.) long. During welding, the inside of the cylinder was flushed with helium to prevent oxidation of heated portions. After the patch had been fillet welded to the cylinder, the center hole was plug welded, using successive small beads to ensure fusion to the wall of the casting and to minimize the possibility of cracking, which can occur when large beads are deposited. The part was then finish machined using plates to hold the flanges.

Example 10: Application of an All-Weld-Metal Boss.

Because a design change was made after the first run of flanged cylindrical alloy EZ33A castings had been produced, a boss had to be added to one of the flanges on each casting, as shown in the views at the bottom in Fig. 8. Using a buildup welding procedure of successive small beads, a cylindrical boss 25 mm (1 in.) in diameter by 6.4 mm ($\frac{1}{4}$ in.) high was produced on the flange by manual GTAW. In later production, the pattern was changed, and the boss was cast integrally with the flange.

Defects in Repair Welds. Oxide inclusions are caused by (1) welding on unsound base metal, (2) inadequate cleaning of base metal or filler metal, (3) maintaining too long an arc, (4) insufficient flow of shielding gas, (5) leaky gas connections, and (6) defective shielding-gas hoses that allow the drawing in of air.

Tungsten inclusions are caused by maintaining too long an arc, using too high a current, and touching the weld pool, base metal, or filler metal with the electrode.

Porosity is usually caused by welding on unsound base metal, poor preweld cleaning of base metal or filler metal, and contamination of the shielding gas.

Microshrinkage, a defect consisting of interdendritic voids detectable on magnification, is caused by too rapid current decay at the end of a weld, which results in too rapid freezing of the weld pool. As a result, the weld pool is unable to serve as a riser to feed the solidifying weld metal below.

Base-metal cracks are usually caused by excessive heat input or by carrying the arc too far over onto the base-metal surface during repair welding.

In repair welds of magnesium alloys that contain rare earth elements or thorium, a light-colored area at the HAZ of the weld is often visible on a radiograph taken with the x-ray beam parallel and the x-ray film perpendicular to the weld edge. This is caused by eutectic-enriched material with high atomic numbers absorbing substantially more radiation. Eutectic enrichment is caused by the melting of eutectic material at a lower temperature and heat expansion of the parent grains, squeezing this material into a layer at the edge of the weld. Below this layer is a darker area on the radiograph where eutectic material has vacated. Above the eutectic-rich layer appears another eutectic-depleted zone due to the first solidification of higher melting constituents. The effect can be minimized by avoiding preheat and by rapid welding.

Welding of Magnesium Alloys

Heat Treatment After Welding

Heat treated castings are often heat treated again after welding. The heat treatments shown in Table 7 depend on the temper before welding and the temper required after welding. Only the minimum heat-treating time ($\frac{1}{2}$ h) for complete solution is used for welded AZ81A, AZ91C, AZ92A, and QE22A castings to avoid abnormal grain growth in the deposited weld metal. As noted in Table 7, some solution treatments require an SO_2 or CO_2 atmosphere.

If complete solution treatment is not required, magnesium alloys containing more than 1.5% Al should always be stress relieved to prevent stress-corrosion cracking in service. Postweld stress-relieving temperatures and times are given in Table 12.

TABLE 12 POSTWELD TREATMENTS TO OBTAIN 80 TO 95% STRESS RELIEF IN MAGNESIUM ALLOYS

ALLOY	TEMPERATURE		TIME
	°C	°F	
SHEET			
AZ31B-O(B)	260	500	15 MIN
AZ31B-H24(B)	150	300	1 H
HK31A-H24	315	600	30 MIN
HM21A-T8	370	700	30 MIN
HM21A-T81	400	750	30 MIN
ZE10A-O	230	450	30 MIN
ZE10A-H24	135	275	1 H
EXTRUSIONS			
AZ10A-F	260	500	15 MIN
AZ31B-F(B)	260	500	15 MIN
AZ61A-F(B)	260	500	15 MIN
AZ80A-F(B)	260	500	15 MIN
AZ80A-T5(B)	205	400	1 H
HM31A-T5(A)	425	800	1 H
AZCOML	260	500	15 MIN
CASTINGS(C)			

AM100A	260	500	1 H
AZ63A	260	500	1 H
AZ81A	260	500	1 H
AZ91C	260	500	1 H
AZ92A	260	500	1 H
EZ33	330	625	2-4 H
EQ21	505	940	1 H
QE22	505	940	1 H
ZE41	330	625	2-4 H
ZC63	425	795	1 H
WE43	510	950	1 H
WE54	510	950	1 H

(A) 70% STRESS RELIEF IS PROVIDED.

(B) REQUIRES POSTWELD HEAT TREATMENT TO AVOID STRESS-CORROSION CRACKING.

(C) REQUIRES POSTWELD HEAT TREATMENT FOR MAXIMUM STRENGTH (SEE TABLE 7 FOR POSTWELD HEAT TREATMENTS)

Welding of Magnesium Alloys

References

1. "STANDARD PRACTICE FOR CODIFICATION OF CERTAIN NONFERROUS METALS AND ALLOYS, CAST AND WROUGHT," B 275, *1989 ANNUAL BOOK OF STANDARDS,* VOL 2.02, *DIE-CAST METALS; ALUMINUM AND MAGNESIUM ALLOYS,* ASTM, 1989, P 325-326
2. H.B. CARY, *MODERN WELDING TECHNOLOGY,* 2ND ED., PRENTICE-HALL, 1989, P 526-527

Welding of Titanium Alloys

Revised by R. Terrence Webster, Consultant

Introduction

COMMERCIALLY PURE TITANIUM and most titanium alloys can be welded by procedures and equipment used in welding austenitic stainless steel and aluminum. Because of the high reactivity of titanium and titanium alloys at temperatures above 550 °C (1020 °F), additional precautions must be applied to shield the weldment from contact with air. Also, titanium base metal and filler metal must be clean to avoid contamination during welding.

Welding of Titanium Alloys

Revised by R. Terrence Webster, Consultant

Weldability

Unalloyed titanium and all alpha titanium alloys are weldable. Although the alpha-beta alloy Ti-6Al-4V and other weakly beta-stabilized alloys are also weldable, strongly beta-stabilized alpha-beta alloys are embrittled by welding. Most beta alloys can be welded, but because aged welds in beta alloys can be quite brittle, heat treatment to strengthen the weld by age hardening should be used with caution.

Unalloyed titanium is generally available in several grades, ranging in purity from 98.5 to 99.5% Ti. These grades are strengthened by variations in oxygen, nitrogen, carbon, and iron. Strengthening by cold working is possible but is seldom used. All grades are usually welded in the annealed condition. Welding of cold-worked alloys anneals the heat-affected zone (HAZ) and negates the strength produced by cold working.

Alpha alloys Ti-5Al-2.5Sn, Ti-6Al-2Sn-4Zr-2Mo, Ti-5Al-5Sn-2Zr-2Mo, Ti-6Al-2Nb-1Ta-1Mo, and Ti-8Al-1Mo-1V are always welded in the annealed condition.

Alpha-beta alloys of Ti-6Al-4V can be welded in the annealed condition or in the solution-treated and partially aged condition, with aging completed during postweld stress relieving. In contrast to unalloyed titanium and the alpha alloys, which can be strengthened only by cold work, the alpha-beta and beta alloys can be strengthened by heat treatment.

The low weld ductility of most alpha-beta alloys is caused by phase transformation in the weld zone or in the HAZ. Alpha-beta alloys can be welded autogeneously or with various filler metals. It is common to weld some of the lower-alloyed materials with matching filler metals. Filler metal of an equivalent grade or one grade lower is used to ensure good weld strength and ductility. Filler metal of matching composition is used to weld the Ti-6Al-4V alloy. This extra-low-interstitial (ELI) grade improves ductility and toughness. The use of filler metals that improve ductility may not prevent embrittlement of the HAZ in susceptible alloys. In addition, low-alloy welds can be embrittled by hydride precipitation. However, with proper joint preparation, filler-metal storage, and shielding, hydride precipitation can be avoided.

Sheet thicknesses 2.54 mm (0.100 in.) and thinner can be welded without filler metal additions by the fused-root welding technique. Filler metal may be added to repair unfused and sunken weld-metal areas. The lack of a joint line on the root face of the weld indicates 100% penetration.

Metastable beta alloys Ti-3Al-13V-11Cr, Ti-11.5Mo-6Zr-4.5Sn, Ti-8Mo-8V-2Fe-3Al, Ti-15V-3Cr-3Al-3Sn, and Ti-3Al-8V-6Cr-4Zr-4Mo are weldable in the annealed or solution heat treated condition. In the as-welded condition, welds are low in strength but ductile. Beta alloy weldments are sometimes used in the as-welded condition. Welds in the Ti-3Al-13V-11Cr alloy embrittle more severely when age hardened. To obtain full strength, the metastable beta alloys are welded in the annealed condition; the weld is cold worked by peening or planishing, and the weldment is then solution treated and aged. This procedure also obtains adequate ductility in the weld.

For additional information on the weldability of titanium alloys, see the Section "Selection and Weldability of Nonferrous Low-Temperature Materials" in this Volume.

Welding of Titanium Alloys

Revised by R. Terrence Webster, Consultant

Process Selection

The following fusion-welding processes are used for joining titanium and titanium alloys:

- GAS-TUNGSTEN ARC WELDING (GTAW)
- GAS-METAL ARC WELDING (GMAW)
- PLASMA ARC WELDING (PAW)
- ELECTRON-BEAM WELDING (EBW)
- LASER-BEAM WELDING (LBW)
- FRICTION WELDING (FRW)
- RESISTANCE WELDING (RW)

Fluxes cannot be used with these processes because they combine with titanium to cause brittleness and may reduce corrosion resistance. The welding processes that use fluxes are electroslag welding, submerged arc welding, and flux-cored arc welding. These processes have been used on a limited basis. However, they are not considered to be economical because they require high-cost, fluoride-base fluxes.

Gas-tungsten arc welding is the most widely used process for joining titanium and titanium alloys except for parts with thick sections. Square-groove butt joints can be welded without filler metal in base metals up to 2.5 mm (0.10 in.) thick. For thicker base metals, the joint should be grooved, and filler metal is required. The heated weld metal in the weld zone must be shielded from the atmosphere to prevent contamination with oxygen, nitrogen, and carbon, which will degrade the weldment ductility.

Welding may be done manually or automatically. Local shielding can be accomplished by the use of backup shields (Fig. 1) and torch-trailing shields (Fig. 2). For complex weld geometries, rigid or collapsible plastic bags purged and filled with inert gas (usually argon) may be used. For welds with high chemical purity requirements, the use of vacuum weld chambers is recommended.

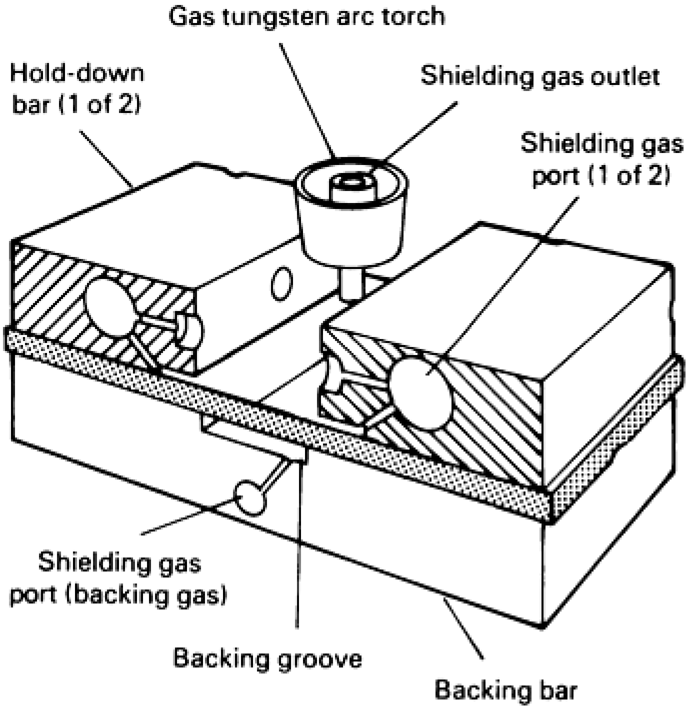
Gas tungsten arc torch
Hold-down bar (1 of 2)
Shielding gas outlet
Shielding gas port (1 of 2)
Shielding gas port (backing gas)
Backing groove
Backing bar

FIG. 1 SETUP FOR INERT GAS SHIELDING FOR GTAW OF TITANIUM ALLOYS OUTSIDE A WELDING CHAMBER. GAS SHIELDING IS FROM THE TORCH AND THROUGH PARTS IN HOLD-DOWN BARS, BACKING BARS, AND FROM TRAILING AND BACKUP SHIELDS.

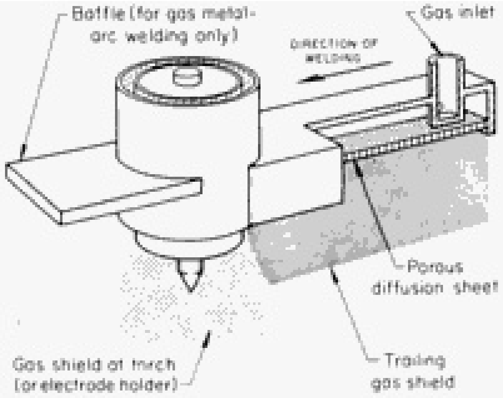
Baffle (for gas metal-arc welding only)
Gas inlet
DIRECTION OF WELDING
Porous diffusion sheet
Gas shield at torch (or electrode holder)
Trailing gas shield

FIG. 2 ARRANGEMENT FOR SHIELDED AUTOMATIC WELDING OF TITANIUM ALLOYS IN AIR. THE BAFFLE SHOWN ON THE LEADING SIDE OF THE TORCH (OR ELECTRODE HOLDER) IS SELDOM USED FOR GTAW, BUT

IS USED FOR GMAW.

Gas-metal arc welding is used to join titanium and titanium alloys more than 3.18 mm (0.125 in.) thick. It is applied using pulsed current or the spray mode and is less costly than GTAW, especially when the base metal thickness is greater than 13 mm (0.50 in.).

Plasma arc welding is also applicable to joining titanium and titanium alloys. It is faster than GTAW and can be used on thicker sections, such as one-pass welding of plate up to 13 mm (0.50 in.) thick, using square-butt and keyhole techniques.

Electron-beam welding is used in the aircraft and aerospace industries for producing high-quality welds in titanium and titanium alloy plates ranging from 6.4 mm (0.25 in.) to more than 76 mm (3 in.) thick. Because the welding is performed in a high-vacuum atmosphere, low contamination of the weldment is achieved.

Laser-beam welding is increasingly being used to join titanium and titanium alloys. Square-butt weld joint configurations can be used, and the welding process does not require the use of vacuum chambers; gas shielding is still required. This process is more limited than electron-beam welding regarding base metal thickness, which cannot usually exceed 13 mm (0.50 in.).

Friction welding is useful in joining tube, pipe, or rods, where joint cleanliness can be achieved without shielding.

Resistance welding is used to join titanium and titanium alloy sheet by either spot welds or continuous seam welds. The process is also used for welding titanium sheet to dissimilar metals, that is, cladding titanium to carbon or stainless steel plate.

Welding of Titanium Alloys

Revised by R. Terrence Webster, Consultant

Filler Metals

For welding titanium thicker than about 2.5 mm (0.10 in.) by the GTAW process, a filler metal must be used. For PAW, a filler metal may be used for welding metal less than 13 mm (0.5 in.) thick.

Composition. Thirteen titanium and titanium alloy filler-metal (or electrode) classifications are given in AWS A5.16. Five of these are essentially unalloyed titanium, and the remainder are titanium alloy filler metals. Maximums are set on carbon, oxygen, hydrogen, and nitrogen contents. Compositions for titanium and titanium alloy filler metals are given in Table 1.

TABLE 1 CHEMICAL COMPOSITION REQUIREMENTS FOR TITANIUM AND TITANIUM ALLOY ELECTRODES AND RODS

AWS CLASSIFICATION		UNS NUMBER	COMPOSITION, WT%								
1990	1970		C	O	H	N	Al	V	Sn	Fe	OTHER
ERTI-1	ERTI-1	R50100	0.03	0.10	0.005	0.015	. . .	. . .	. . .	0.10	. . .
ERTI-2	ERTI-2	R50120	0.03	0.10	0.008	0.020	. . .	. . .	. . .	0.20	. . .
ERTI-3	ERTI-3	R50125	0.03	0.10-0.15	0.008	0.020	. . .	. . .	. . .	0.20	. . .
ERTI-4	ERTI-4	R50130	0.03	0.15-0.25	0.008	0.020	. . .	. . .	. . .	0.30	. . .
ERTI-5	ERTI-6A1-	R56400	0.05	0.18	0.015	0.030	5.5-	3.5-	. . .	0.30	0.005 Y

	4V						6.7	4.5			
ERTI-5ELI	ERTI-6AL-4V-1	R56402	0.03	0.10	0.005	0.012	5.5-6.5	3.5-4.5	. . .	0.15	0.005 Y
ERTI-6	ERTI-5A 1-2.5SN	R54522	0.08	0.18	0.015	0.050	4.5-5.8	. . .	2.0-3.0	0.50	0.005 Y
ERTI-6ELI	ERTI-5AL-2.5SN-1	R54523	0.03	0.10	0.005	0.012	4.5-5.8	. . .	2.0-3.0	0.20	0.005 Y
ERTI-7	ERTI-0.2PD	R52401	0.03	0.10	0.008	0.020	. . .	. . .	. . .	0.20	0.12/0.25 PD
ERTI-9	ERTI-3AL-2.5V	R56320	0.03	0.12	0.008	0.020	2.5-3.5	2.0-3.0	. . .	0.25	0.005 Y
ERTI-9ELI	ERTI-3AL-2.5V-1	R56321	0.03	0.10	0.005	0.012	2.5-3.5	2.0-3.0	. . .	0.20	0.005 Y
ERTI-12	. . .	R53400	0.03	0.25	0.008	0.020	. . .	. . .	. . .	0.30	0.2/0.4 MO 0.6/0.9 NI
ERTI-15	ERTI-6AL-2CB-1TA-1MO	R56210	0.03	0.10	0.005	0.015	5.5-65	. . .	. . .	0.15	0.5/1.5 MO 1.5/2.5 NB 0.5/1.5 TA

Notes: Titanium constitutes the remainder of the composition. Single values are maximum. Analysis of the interstitial elements carbon, oxygen, hydrogen, and nitrogen shall be conducted on samples of filler metal taken after the filler metal has been reduced to its final diameter and all processing operations have been completed. Analysis of the other elements may be conducted on these same samples or it may have been conducted on samples taken from the ingot or the rod stock from which the filler metal is made. In case of dispute, samples from the finished filler metal shall be the referee method. Residual elements, total, shall not exceed 0.02%, with no single such element exceeding 0.05%. Residual elements need not be reported unless a report is specifically required by the purchaser. Residual elements are those elements (other than titanium) that are not listed in Table 1 for the particular classification, but which are inherent in the raw material or the manufacturing practice. Residual elements can be present only in trace amounts and they cannot be elements that have been intentionally added to the product.

Source: ANSI/AWS A5.16-90, American Welding Society, 1990

Filler-metal composition is usually matched to the grade of titanium being welded. For improved joint ductility in welding the higher strength grades of unalloyed titanium, filler metal of yield strength lower than that of the base metal is occasionally used. Because of the dilution that occurs during welding, the weld deposit acquires the required strength. Unalloyed filler metal is sometimes used to weld Ti-5Al-2.5Sn and Ti-6Al-4V for improved joint ductility. The use of unalloyed filler metals lowers the beta content of the weldment, thereby reducing the extent of the transformation that occurs and improving ductility. Engineering approval, however, is recommended when using pure filler metal to ensure that the weld meets strength requirements. Another option is filler metal containing lower interstitial content (oxygen, hydrogen, nitrogen, and carbon) or alloying contents that are lower than the base metal being used. The use of filler metals that improve ductility does not preclude embrittlement of the HAZ in susceptible alloys. In addition, low-alloy welds may enhance the possibility of hydrogen embrittlement.

Preparation. The filler metal, as well as the base metal, should be clean at the time of welding. As shown in Fig. 3, wires used for filler metals have a large surface-to-volume ratio. Therefore, if the wire surface is slightly contaminated, the weld may be severely contaminated. Some procedures require that the filler wire be cleaned immediately before use. The use of an acetone-soaked, lint-free cloth serves to assess surface contamination caused by the die lubricant used in the wire drawing operation, in addition to cleaning the filler wire. Pickling in nitric-hydrofluoric acid solution is also used for cleaning.

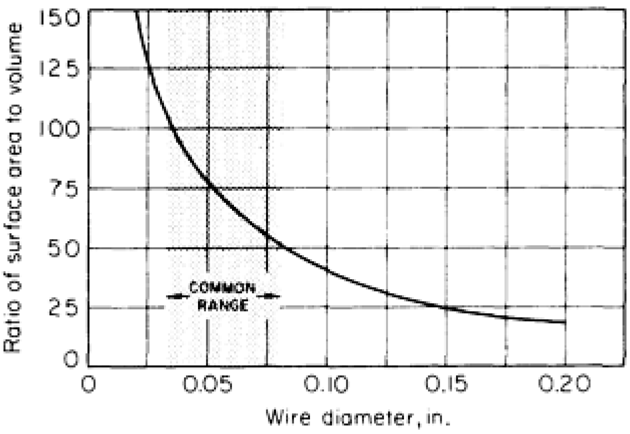

FIG. 3 RATIO OF FILLER-METAL (ELECTRODE) WIRE SURFACE AREA TO VOLUME FOR VARIOUS WIRE DIAMETERS. "COMMON RANGE" INDICATES RANGE OF WIRE DIAMETERS MOST OFTEN USED FOR WELDING TITANIUM.

Welding of Titanium Alloys

Revised by R. Terrence Webster, Consultant

Shielding Gases

In welding titanium and titanium alloys, only argon and helium, and occasionally a mixture of these two gases, are used for shielding. Because it is more readily available and less costly, argon is more widely used. Argon shielding gas was used in the examples given in this article.

Because of its high purity (99.985% min.) and low moisture content, a liquid argon supply is often preferred. The argon gas should have a dew point of -60 °C (-75 °F) or lower. The hose used for the shielding gas should be clean, nonporous, and flexible, and may be made of Tygon or vinyl plastic. Because rubber hose absorbs air, it should not be used. Excessive gas flow rates that cause turbulence should be avoided, and flowmeters are usually employed for all gas shields. Pressure gages may be used for trailing and backup shields.

Characteristics. The type of shielding gas used affects the characteristics of the arc. At a given welding current, the arc voltage is much greater with helium than with argon. Because the heat energy liberated in helium is about twice that in argon, higher welding speeds can be obtained, weld penetration is deeper, and thicker sections can be welded more rapidly using helium shielding. However, when using pure helium for welding, arc stability and weld-metal control are sacrificed.

Argon is used when welding thin and thick sections where the arc length can be altered without appreciably changing the heat input. Argon-helium mixtures are also used, particularly 75% argon, which improves arc stability, and 25% helium, which increases penetration. The 75Ar-25He mixture is also frequently utilized as the shielding gas at the torch in automatic operations. Furthermore, helium is used in shielding out-of-position welds.

Welding of Titanium Alloys

Revised by R. Terrence Webster, Consultant

Joint Preparation

If welding is done outside a controlled-atmosphere welding chamber, joints must be carefully designed so that both the top and the underside of the weld can be shielded (Fig. 1). Dimensions of typical joints are given in Table 2. For welding titanium alloys, joint fit-up should be better than for welding other metals, because of the possibility of entrapping air in the joint. The joint should be clamped to prevent separation during welding.

TABLE 2 DIMENSIONS OF TYPICAL JOINTS FOR WELDING TITANIUM AND TITANIUM ALLOYS

BASE-METAL THICKNESS		ROOT OPENING(A)	GROOVE ANGLE, DEGREES	WELD-BEAD WIDTH(A)
mm	in.			
SQUARE-GROOVE BUTT JOINT				
0.254-2.29	0.010-0.090	0	. . .	. . .
0.787-3.18	0.0314-0.125	0-0.10*T*	. . .	. . .
SINGLE-V-GROOVE BUTT JOINT				
1.57-3.18	0.062-0.125	0-0.10*T*	30-60	0.10-0.25*T*
2.29-3.18	0.090-0.125	(B)	90	. . .
3.18-6.35	0.125-0.250	0-0.10*T*	30-60	0.10-0.25*T*
DOUBLE-V-GROOVE BUTT JOINT				
6.35-12.7	0.250-0.500	0-0.20*T*	30-120	0.10-0.25*T*
SINGLE-U-GROOVE BUTT JOINT				
6.35-19.1	0.250-0.750	0-0.10*T*	15-30	0.10-0.25*T*
DOUBLE-U-GROOVE BUTT JOINT				
19.1-38.1	0.750-1.500	0-0.10*T*	15-30	0.10-0.25*T*
FILLET WELD				
0.787-3.18	0.031-125	0-0.10*T*	0-45	0-0.25*T*
3.18-12.7	0.125-0.500	0-0.10*T*	30-45	0.10-0.25*T*

Source: J.J. Vagi *et al.,* "Welding Procedures for Titanium and Titanium Alloys," NASA TMX 53432, 1965

(A) *T,* BASE-METAL THICKNESS.

(B) ROOT FACE, 0.76 MM (0.030 IN.).

Welding of Titanium Alloys

Revised by R. Terrence Webster, Consultant

Cleaning

To obtain a good weld, the joint and the surfaces of the workpiece at least 51 mm (2 in.) beyond the width of the gas trailing shield on each side of the weld groove must be meticulously cleaned. The cleaning procedure depends on whether the oxide layer in the joint area is light or heavy.

Degreasing. Grease and oil accumulated during forming and machining must be removed before welding to avoid weld contamination. Scale-free metal requires only degreasing. Degreasing precedes descaling for metal with an oxide scale. Methods of degreasing include steam cleaning, alkaline cleaning, vapor degreasing, and solvent cleaning.

For vapor degreasing, toluene rather than a chlorinated solvent should be used, because residues from chlorinated solvents (and also from silicated solvents) may contribute to cracking of titanium weldments. Solvent cleaning is frequently used, especially for large components that cannot conveniently be placed in a vapor degreaser or washer for alkaline cleaning. Solvents applied include methyl ethyl ketone, toluene, acetone, and other chlorine-free solvents. Because methyl alcohol has reportedly caused stress corrosion, it is prohibited for use on aerospace hardware.

In solvent cleaning, the joint areas are hand-wiped with the solvent just before welding. All wiping should be done with clean, lint-free cloths or a cellulose sponge. Plastic or lint-free gloves should be worn; rubber gloves are likely to leave traces of plasticizer that can cause porosity in the weld metal. Handprints are also a source of contamination.

After a lightly oxidized joint area has been degreased, it should be pickled for a short time. A typical mixture is 4 wt% hydrofluoric acid and 40 wt% nitric acid. Because hydrogen is detrimental to the properties of titanium, causing embrittlement and sometimes contributing to weld porosity, pickling should be performed cautiously. Industrial practice usually is to maintain the acid bath at a high oxidation potential of 30 wt% or more nitric acid, which simultaneously holds the ratio of nitric acid to hydrofluoric acid at 10 to 1 strictly as a factor of safety. If the nitric acid content falls below 30 wt% and the ratio of nitric acid to hydrofluoric acid falls below 10 to 1, excessive hydrogen pickup is possible.

Oxide Removal. Lightly oxidized joint areas may also be cleaned by brushing with a stainless steel wire brush or by draw filing. When weld corrosion resistance is important, however, these cleaning methods should be followed immediately by acid pickling. Steel wool or abrasives should never be used, because of the danger of contamination.

If grinding is required, the use of silicon carbide burrs is preferred, because wheels produce residues of rubber or resin on the surface that contaminate the weld. To prevent burning, excessive heat should be avoided while grinding, and low rotating speeds should be used.

If titanium has been exposed to temperatures above 540 °C (1000 °F), a more complex removal treatment, such as chemical, salt bath, mechanical, or combinations of these treatments is required. The salt baths are basically sodium hydroxide to which oxidizing agents or hydrogen has been added to form sodium hydride.

One procedure for removing heavy scale involves subjecting the parts to liquid abrasive blasting or salt-bath descaling after degreasing. These treatments are usually followed by pickling in nitric-hydrofluoric acid for the removal of light scale. When salt-bath descaling is used, oxide removal can be hastened by removing the workpieces from the bath, scrubbing them with brushes, and then re-immersing them. To prevent hydrogen pickup during salt-bath descaling, cycle time must be short (preferably no more than 2 min) and bath temperature must be carefully controlled.

Selection of cleaning method depends largely on the size and shape of the parts and on the cleaning methods available in a particular plant.

Welding of Titanium Alloys

Revised by R. Terrence Webster, Consultant

Equipment for Gas-Tungsten Arc Welding

Power Supplies. Inverters are the preferred power supply for welding titanium, because the current can be controlled more closely than with motor-generator sets; slight variations in welding current may cause variations in penetration. Transformer-rectifiers may also be used. Direct current electrode negative (DCEN) is always used for GTAW of titanium because deeper weld penetration and a narrower bead can be obtained than with direct current electrode positive (DCEP). Also, in manual welding, DCEN is easier to control.

The power supply should include accessories for arc initiation because of the danger of tungsten contamination of the weld if the arc is struck by torch starting. If welding is to be done in air, controls for extinguishing the arc without pulling the torch away from the workpiece are needed, so that shielding-gas flow continues and the hot weld metal is not contaminated by air.

Electrodes. The conventional thoriated tungsten types of electrodes (EWTh-1 or EWTh-2) are used for GTAW of titanium. Electrode size is governed by the smallest diameter able to carry the welding current. To improve arc initiation and control the spread of the arc, the electrode should be ground to a point. The electrode may extend one and a half times the size of the diameter beyond the end of the nozzle.

Shielding. To ensure a diffuse, nonturbulent flow of shielding gas, nozzles of torches for welding titanium are larger than those used for welding other metals. With a 0.8 mm ($\frac{1}{32}$ in.) diam electrode, a 14 mm ($\frac{9}{16}$ in.) ID nozzle is ordinarily used, and with a 1.6 mm ($\frac{1}{16}$ in.) diam electrode, a 19 mm ($\frac{3}{4}$ in.) ID nozzle is used.

Because titanium has low thermal conductivity, the area ahead of the arc does not exceed 540 °C (1000 °F); therefore, leading shields are seldom required when welding is done by the GTAW process. For welding operations where a trailing shield (Fig. 2) is not adaptable, the nozzle of the torch is fitted with a concentric outer shroud through which a supplementary supply of shielding gas is fed. In Fig. 4, which shows this type of torch, a shielding gas is diffused through copper shavings contained in the outer shroud that cool the gas substantially, helping to protect the metal near the weld.

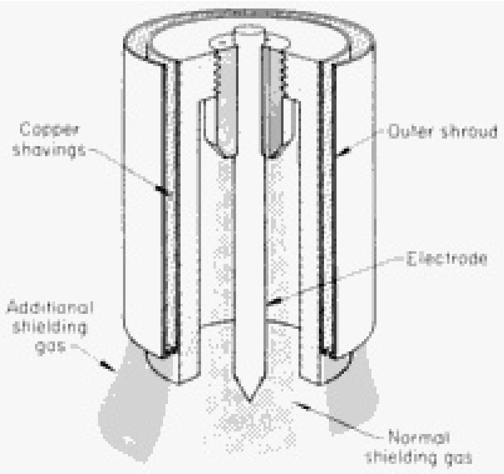

FIG. 4 SECTIONAL VIEW OF A TORCH NOZZLE EQUIPPED WITH AN OUTER SHROUD. COPPER SHAVINGS IN THE OUTER SHROUD PROVIDE ADDITIONAL GAS SHIELDING FOR MANUAL INERT GAS WELDING.

Shielding of the underside of a weld is provided by slotted backing bars, usually copper, through which a diffuse flow of argon or helium is maintained (Fig. 1). Gas channels in the clamping fixtures also provide diffuse flow of inert gas to the weld area. These fixtures are placed close to the weld to avoid the danger of air contamination.

Fixtures. Copper fixtures are usually employed for GTAW. Although other metals are sometimes used, they should be nonmagnetic; arc blow tends to occur with magnetic metal fixtures. Metal fixtures are sometimes water cooled, but this method introduces the possibility of moisture from the air condensing on the fixtures.

Welding of Titanium Alloys

Revised by R. Terrence Webster, Consultant

Stress Relieving

Most titanium weldments are stress relieved after welding to prevent weld cracking and susceptibility to stress-corrosion cracking in service. Stress relief also improves fatigue strength. An assembly subjected to a substantial amount of welding and severe fixturing restraint may require intermediate stress relieving of the partially welded structure, which should be done in an inert atmosphere. Otherwise, the unwelded joints may have to be recleaned before being welded.

With unalloyed titanium and alpha titanium alloys, time and temperature should be controlled to prevent grain growth. Stress relieving times and temperatures for several weldable titanium alloys are given in Table 3. For alloys not mentioned in this table, tests should be conducted to ensure that stress relieving does not reduce fracture toughness, creep strength, or other important properties.

TABLE 3 STRESS-RELIEVING TIMES AND TEMPERATURES FOR SELECTED TITANIUM ALLOYS

GRADE	TEMPERATURE		TIME, H
	°C	°F	
UNALLOYED TI-0.15PD	425	800	8
	480	900	0.75
	540	1000	0.5
TI-5AL-2.5SN(A)	480	900	20
	540	1000	6
	595	1100	2
	650	1200	1
TI-6AL-4V:			
ANNEALED(A)	480	900	20
	540	1000	2
	595	1100	1
SOLUTION-TREATED	480	900	15
	540	1000	4
AGED	480	900	15
	540	1000	5
TI-8AL-1MO-IV:			
SINGLE- AND DOUBLE-ANNEALED	595	1100	2
	650	1200	1.5
	705	1300	0.5
TRIPLE-ANNEALED	595	1100	5
	650	1200	2
	750	1300	0.5
TI-3AL-2.5V	540	1000	0.5
	595	1100	2
TI-13V-11CR-3AL	480	900	0.5
	540	1000	6

(A) DATA ALSO APPLY TO EXTRA-LOW-INTERSTITIAL (ELI) MODIFICATIONS.

Stress relieving of Ti-13V-11Cr-3Al weldments causes aging and subsequent embrittlement of the weld and HAZ and, therefore, is not recommended. Re-solution heat treatment (re-annealing) may be used to relieve stresses if the welded assembly is amenable to such treatment. All titanium surfaces should be free of dirt, fingerprints, grease, and residues before stress relieving. Contaminated surface metal must be removed from the entire weldment by machining or descaling and pickling to remove 0.025 to 0.050 mm (0.001 to 0.002 in.) per surface.

Welding of Titanium Alloys

Revised by R. Terrence Webster, Consultant

Repair Welding

Repair welds should follow the established specification requirements for the original welds and be made prior to final heat treatment. Manual or automatic GTAW is generally used for repairing butt and fillet welds. Repairs can also feature a combination of welding processes such as GTAW and the initial welding process (GMAW, PAW, or hot wire welding).

Repair welds must always be carefully executed, and all traces of liquid-penetrant inspection material must be removed. Generally, inspection is performed on both faces of the repair weld and several inches beyond the repaired area.

Welding of Zirconium Alloys

R. Terrence Webster, Consultant

Introduction

ZIRCONIUM is a corrosion-resistant structural metal that has many physical and mechanical properties similar to titanium and austenitic stainless steels. Zirconium alloys are weldable; procedures and equipment are similar to those used for welding titanium and austenitic stainless steels. Zirconium has a low coefficient of thermal expansion, which contributes to low distortion during welding.

Because of the reactivity of zirconium with oxygen, nitrogen, and hydrogen, the metal must be shielded during welding with high-purity inert gas or a good vacuum. Furthermore, zirconium should be free of oil, grease, and dirt to avoid the dissolving of carbon- and oxygen-containing materials, which can embrittle the metal or create porosity and may reduce the corrosion-resistant properties of the metal.

Zirconium and its alloys are available in two general categories: commercial grade and reactor grade. Commercial-grade zirconium designates zirconium that contains hafnium as an impurity. Reactor-grade zirconium designates zirconium from which most of the hafnium has been removed to make it suitable for nuclear reactor applications. Because pure zirconium has relatively low mechanical properties, various alloying elements are added to enhance its mechanical properties.

Zirconium and its alloys are available in plate, sheet, bar, rod, and tubing form in a variety of material specifications. The four most common grades of commercial zirconium alloys, as specified in ASTM B 550, are given in Table 1. For reactor grades, the hafnium content is reduced to a minimum and other impurities are closely controlled. Chemical compositions of the most common grades of nuclear-grade Zircaloys, as specified in ASTM B 351, are given in Table 2.

TABLE 1 CHEMICAL COMPOSITIONS OF COMMERCIAL ZIRCONIUM AND ZIRCONIUM ALLOYS PER ASTM SPECIFICATION B 550-92

ELEMENT	**COMPOSITION, WT%**		
	R60702	**R60704**	**R60705**
ZIRCONIUM + HAFNIUM, MIN	99.2	97.5	95.5
HAFNIUM, MAX	4.5	4.5	4.5
IRON + CHROMIUM	0.2 MAX	0.2-0.4	0.2 MAX
TIN	. . .	1.0-2.0	. . .
HYDROGEN, MAX	0.005	0.005	0.005
NITROGEN, MAX	0.025	0.025	0.025
CARBON, MAX	0.05	0.05	0.05
NIOBIUM	. . .	. . .	2.0-3.0
OXYGEN, MAX	0.16	0.18	0.18

TABLE 2 CHEMICAL COMPOSITIONS OF THE MOST COMMON GRADES OF NUCLEAR-GRADE ZIRCALOYS PER ASTM SPECIFICATION B 351-92

ELEMENT	**COMPOSITION WT%**		
	R60802 (ZIRCALOY-2)	**R60804 (ZIRCALOY-4)**	**R60901 (ZR-2.5NB)**
TIN	1.20-1.70	1.20-1.70	. . .
IRON	0.07-0.20	0.18-0.24	. . .
CHROMIUM	0.05-0.15	0.07-0.13	. . .

NICKEL	0.03-0.08	. . .	. . .
NIOBIUM	. . .	. . .	2.40-2.80
OXYGEN	. . .	. . .	0.09-0.13
IRON + CHROMIUM + NICKEL	0.28-038	. . .	. . .
IRON + CHROMIUM	. . .	0.28-0.37	. . .
HAFNIUM	0.010	0.010	0.010
ZIRCONIUM	BAL	BAL	BAL

Zirconium can only be welded to itself or to other reactive metal alloys, such as titanium, niobium, or tantalum. When welded to nonreactive metals, intermetallics that embrittle the weld metal are formed.

Welding of Zirconium Alloys

R. Terrence Webster, Consultant

Process Selection

Zirconium alloys are highly reactive to oxygen and nitrogen in air at high temperatures. Consequently, the selected welding processes and procedures must be capable of shielding the weldment and heat-affected zones (HAZ) from contamination. The use of fluxes is generally avoided, because reactivity with the chemicals in the fluxes causes brittleness and may reduce the corrosion resistance of zirconium weldments.

The following welding processes can be used for welding any of the zirconium alloys:

- GAS-TUNGSTEN ARC WELDING (GTAW)
- GAS-METAL ARC WELDING (GMAW)
- PLASMA ARC WELDING (PAW)
- ELECTRON-BEAM WELDING (EBW)
- LASER-BEAM WELDING (LBW)
- FRICTION WELDING (FRW)
- RESISTANCE WELDING (RW)
- RESISTANCE SPOT WELDING (RSW)
- RESISTANCE SEAM WELDING (RSEW)

The selection of a welding process depends on several factors: weld joint, tensile and corrosion-resistant property requirements, cost, and design configuration.

Gas-tungsten arc welding is the most widely used process for joining zirconium alloys. It employs techniques similar to those used for welding stainless steel, that is, the direct current power supply is connected for straight polarity (electrode negative, DCEN). Two desirable features are a contactor for making and breaking the arc and high-frequency arc starting. Recommended electrodes are thoriated tungsten, Type EWTh2.

Plasma arc welding is also commonly used, especially for autogenous welding of butt joint thicknesses from 3.18 to 1.52 mm (0.125 to 0.600 in.).

Gas-metal arc welding is occasionally used for joint thicknesses from 3.18 mm (0.125 in.) or more, because of its more-rapid weld time and the consequent savings in shielding gas and production time. Weld quality is more difficult to maintain, because of weld spatter and arc instability, which result in weld contamination and weld defects.

Electron-beam welding is rarely used, because of high equipment operating cost as well as weld chamber size limitations.

Laser-beam welding has had very limited use in joining zirconium and has been applied primarily in nuclear reactors.

Friction welding is used to join zirconium tubes to zirconium rods, as well as to dissimilar metal alloys (for example, zirconium to stainless steel) for heat-exchanger applications.

Resistance welding is especially useful for the seam or spot welding of thin sheets, because no shielding is required.

Welding of Zirconium Alloys

R. Terrence Webster, Consultant

Shielding Gases

High-purity argon (99.996% Ar; -65 °C, or -85 °F, dew point), helium, or a mixture of the two gases are used to weld zirconium alloys. Argon has the advantage over helium in terms of lower cost, and it provides good stability and weld metal control. Helium is used to obtain deeper penetration into the joint, and mixtures of argon and helium are frequently used in automatic operations. Helium is the gas most often used to fabricate nuclear reactor equipment.

Welding of Zirconium Alloys

R. Terrence Webster, Consultant

Filler-Metal Selection

Three zirconium and zirconium-alloy electrodes are identified in American .Welding Society (AWS) specification A5.24-90 (Table 3). The filler-metal composition is selected to match the base-metal composition. Type ERZr2 is used to weld commercially pure zirconium (grade R60702); type ERZr3 to weld Zr-1.5Sn alloy (grade R60704); and type ERZr4 to weld Zr-2.5Nb alloy (grades R60705 and R60706). For critical nuclear applications, the impurities are more closely controlled and the hafnium level is minimized. The nuclear-grade zirconium alloys (Table 1) are welded with filler wire of the same grade chemical composition.

TABLE 3 CHEMICAL COMPOSITIONS OF ZIRCONIUM AND ZIRCONIUM-ALLOY ELECTRODES

ELEMENT	COMPOSITION, WT%		ERZR4 (R60705)
	ERZR2 (R60702)	ERZR3 (R60704)	
CARBON	0.05	0.05	0.05
CHROMIUM + IRON	0.20	0.20-0.40	0.20
NIOBIUM	. . .	. . .	2.0-3.0
HAFNIUM	4.5	4.5	4.5
HYDROGEN	0.005	0.005	0.005
NITROGEN	0.025	0.025	0.025
TIN	. . .	1.00-2.00	. . .
ZIRCONIUM + HAFNIUM	99.01 MIN	97.5 MIN	95.5 MIN

Source: AWS A5.24-90

Welding of Zirconium Alloys

R. Terrence Webster, Consultant

Process Procedures

Joints and welding wire should be cleaned before welding. Most welding wire is adequately acid cleaned prior to spooling or packaging. When cut lengths of wire are found to be dirty, it is good practice to wipe each piece with a clean cloth and solvent prior to use. If it is necessary to clean spooled wire, then a wipe with solvents will usually be satisfactory. If wire is acid cleaned, then the acid bath should be fresh and should not contain iron or any other contaminant that will remain on the wire surface.

Heavy grease and machining oil on the work-piece can be removed using degreasing solvents or reagent methanol, acetone, methyl ethyl ketone (MEK), or the equivalent. If acetone is used, then an alcohol rinse should follow.

Heavy mill scale (usually gray in color) should be removed by grinding. The blue surface oxide that develops when the metal is heated above 315 to 425 °C (600 to 800 °F) can be removed either by wire brushing (with a clean austenitic stainless steel wire brush not used for other metals) or by grinding or acid pickling, which are more effective than wire brushing. A common treatment is to soak from 2 to 5 min in room-temperature solution of 35% HNO_3 and 5% HF in water, followed by a thorough clean-water rinse and air drying.

A joint cleaned by degreasing can still be covered by a very light oil film that is approximately the equivalent thickness of the residue left by fingerprints. This can be effectively removed by abrasive or detergent cleaning, using common household cleansers or detergents, followed by air drying. Air drying can be accelerated by rinsing with methanol, acetone, or other solvents.

Joints cleaned by abrasive, detergent, or acid pickling should be kept entirely clean of dirt, including fingerprints. Generally, it is good practice to clean and to weld a joint within the same work shift. When this is not possible, the joint should be covered with paper or bagged with plastic sheet or another material to minimize the accumulation of dust and dirt.

ASTM specification B 614 ("Descaling and Cleaning Zirconium and Zirconium Alloy Surfaces") should be consulted for cleaning details.

Preheating is not normally required for zirconium welding. The commercial alloys have enough ductility to prevent weld cracking (of uncontaminated welds). Preheating is an aid to moisture removal and is good practice where the shop temperature is low or the humidity is high (it is better to control temperature and humidity in the welding area). Heat can be applied by any conventional means. However, if torch heating is used, then the torch should be moved continuously to avoid local heating of the metal.

Interpass cleaning is not required if the weld deposit is bright and shiny. Light temper colors (straw to light blue) should be removed by brushing with a clean austenitic stainless steel brush.

To prevent weld contamination from the oxygen and nitrogen in air, the weld must be shielded during welding and cooling to below 480 °C (896 °F).

The commercial grades of zirconium will most often be welded in air, and the weldments will require gas shielding on the torch side and the back side of the weld. For sheet, plate, and pipe welds, this requires a gas trailing shield on the torch side and a back-up gas shield for the back side. The shielding gases must envelope the weldment during welding and cooling.

Complex assemblies that are difficult to shield in air can be welded in a chamber. The chamber can be capable of either being evacuated and back filled with argon or helium or it can be supply purged with shielding gas before and during the welding operation. Taping a plastic bag over the part to be welded and sealing the bag at the point where the welding cables enter the bag will simulate an enclosed chamber when helium is introduced into the bag system as a shielding gas.

Heat Treatment. Preweld heat treatment is not required. Postweld stress relief for commercial grades R60702 and R60704 can be used when high residual stresses occur. Grade R60705 must be stress relieved within 14 days after welding to prevent delayed hydride cracking (DHC). Commercial-grade zirconium alloys can be heat treated in air or in an inert atmosphere. Nuclear-grade zirconium alloys are usually heat treated in an inert atmosphere.

Hardfacing, Weld Cladding, and Dissimilar Metal Joining

J.R. Davis, Davis & Associates

Introduction

SURFACING, as applied to welding technology, refers to the deposition of a filler metal on a base metal (substrate) to impart some desired property to the surface that is not intrinsic to the underlying base metal. There are several types of surfacing: hardfacing, buildup, weld cladding, and buttering.

Hardfacing is a form of surfacing that is applied for the purpose of reducing wear, abrasion, impact, erosion, galling, or cavitation. As will be described in the first section of this article, hardfacing alloys may be deposited by oxyfuel welding, various arc welding methods, laser welding, and thermal spray processes.

The term buildup refers to the addition of weld metal to a base metal surface for the restoration of the component to the required dimensions. Buildup alloys are generally not designed to resist wear, but to return the worn part back to, or near, its original dimensions or to provide adequate support for subsequent layers of true hardfacing materials.

A weld clad is a relatively thick layer of filler metal applied to a carbon or low-alloy steel base metal for the purpose of providing a corrosion-resistant surface. Weld cladding is normally carried out by arc welding, the subject of the second section of this article. It should be pointed out, however, that cladding can also be accomplished by solid-state welding processes (roll cladding or explosion cladding) and vacuum brazing. These non-fusion techniques are described elsewhere in this Volume.

Buttering also involves the addition of one or more layers of weld metal to the face of the joint or surface to be welded. It differs from buildup in that the primary purpose of buttering is to satisfy some metallurgical consideration. It is used primarily for the joining of dissimilar metal base metals, as is described in the final section of this article.

Many of the principles associated with successful welding of dissimilar metals are also addressed in the sections on "Hardfacing" and "Weld Cladding" in this article, because these processes also involve a combination of materials with significantly different chemical, mechanical, and physical properties. Such combinations pose a number of challenges for the welding engineer, as evidenced by the discussions of dilution control found throughout this article.

Hardfacing

Hardfacing is the application of a hard, wear-resistant material to the surface of a component by welding, thermal spraying, or allied welding processes to reduce wear or loss of material by abrasion, impact, erosion, galling, and cavitation (Ref 1). The stipulation that the surface be modified by welding, thermal spraying, or allied welding processes excludes the use of heat treatment or surface modification processes such as flame hardening, nitriding, or ion implantation as a hardfacing process. The stipulation that the surface be applied for the main purpose of reducing wear excludes the application of materials primarily used for prevention or control of corrosion or high-temperature scaling. Weld overlays to control corrosion are described in the section on "Weld Cladding" in this article. Corrosion and/or high-temperature scaling may, however, have a major effect on the wear rate and, hence, may become a significant factor in the selection of proper materials for hardfacing.

Hardfacing applications for wear control vary widely, ranging from very severe abrasive wear service, such as rock crushing and pulverizing, to applications to minimize metal-to-metal wear, such as control valves where a few thousandths of an inch of wear is intolerable. Hardfacing is used for controlling abrasive wear, such as that encountered by mill hammers, digging tools, extrusion screws, cutting shears, parts of earthmoving equipment, ball mills, and crusher parts. It is also used to control the wear of unlubricated or poorly lubricated metal-to-metal sliding contacts such as control valves, undercarriage parts of tractors and shovels, and high-performance bearings.

Hardfacing is also used to control combinations of wear and corrosion, as encountered by mud seals, plows, knives in the food processing industry, valves and pumps handling corrosive liquids, or slurries. In most instances, parts are made of either plain carbon steel or stainless steel, materials that do not provide desirable wear on their own. Hardfacing alloys are applied to critical wear areas of original equipment or during reclamation of worn parts.

Hardfacing Materials

Hardfacing materials include a wide variety of alloys, carbides, and combinations of these materials. Conventional hardfacing materials, also referred to as weld overlays, are normally classified as steels or low-alloy ferrous materials, high-chromium white irons or high-alloy ferrous materials, carbides, nickel-base alloys, or cobalt-base alloys. A few copper-base alloys are sometimes used for hardfacing applications, but for the most part, hardfacing alloys are either iron-, nickel-, or cobalt-base. Out of a total hardfacing market of about 18×10^6 kg (40×10^6 lb) per year, it has been estimated that iron-base hardfacing alloys account for about 90% by weight, but only about 75% by dollar volume (Ref 2).

Microstructurally, hardfacing alloys generally consist of hard phase precipitates such as borides, carbides, or intermetallics bound in a softer iron-, nickel, or cobalt-base alloy matrix. Carbides am the predominant hard phases in iron- and cobalt-base hardfacing alloys. Carbon contents of iron- and cobalt-base hardfacing alloys generally range up to 4 wt%. Borides, as well as carbides, are the predominant hard phases in nickel-base hardfacing alloys. Combined carbon plus boron contents generally range up to 5 wt%. The specific carbide and/or boride phases that form are determined by matrix alloying additions.

The matrix alloys in most cobalt-, nickel-and high-alloy iron-base hardfacing alloys generally contain up to 35% Cr, up to 30% Mo, and up to 13% W, with smaller amounts of silicon and manganese.

Thermal spray materials, which encompass an extremely wide range of metals and alloys, carbides, ceramics, and polymers, are described in the section on "Thermal Spraying" in this article.

Hardfacing Alloy Categories (Ref 2)

Hardfacing (weld overlay) materials can be grouped into five general categories:

- BUILDUP ALLOYS
- METAL-TO-METAL WEAR ALLOYS
- METAL-TO-EARTH ABRASION ALLOYS
- TUNGSTEN CARBIDES (FOR EXTREME EARTH SLIDING AND CUTTING WEAR)
- NONFERROUS ALLOYS

The buildup alloys include low-alloy pearlitic steels and highly alloyed austenitic manganese steels. For the most part, these alloys are not designed to resist wear but to return a worn part back to, or near, its original dimensions and to provide adequate support for subsequent layers of true hardfacing materials. However, austenitic manganese steels are used as wear-resistant materials under mild wear conditions. Typical examples of applications where buildup alloys are used for wearing surfaces include tractor rails, railroad rail ends, steel mill table rolls, and large slow-speed gear teeth.

The metal-to-metal wear alloys are martensitic air-hardening steels that, with care, can be applied (without cracking) to wearing areas of machinery parts (hence, these materials are commonly referred to as machinery hardfacing alloys). Typical applications of this alloy family include undercarriage components of tractors and power shovels, steel mill work rolls, and crane wheels.

The metal-to-earth abrasion alloys are high-chromium white irons in which chromium carbides are formed during alloy solidification. The metal-to-earth alloys possess resistance to sliding and crushing (that is, low- and high-stress abrasion) on a moderate scale. Typical applications for these material types include shovel teeth, rock crusher parts, plowshares, and auger flights.

Tungsten carbides are actually composite materials, and their use involves the transfer of discrete tungsten carbide particles (which in the welding consumable forms are encased in a steel tube) across the welding arc and into the molten weld pool, where they are subsequently "frozen" into the overlay structure by the matrix formed from the melting of the steel tube. The tungsten carbides are intended for use under extreme sliding and cutting conditions. Typical applications for these tungsten carbides include cutting teeth and edge-holding surfaces of rock drill bits, and quarrying, digging, and earthmoving equipment.

The nonferrous hardfacing alloys are used in environments that are too aggressive for the ferrous hardfacing alloys or where high resistance to a specific type of wear (other than abrasion) is required. For the purpose of this article, they are classified as follows:

- COBALT-BASE, INCLUDING CARBIDE-CONTAINING ALLOYS AND LAVES PHASE ALLOYS
- NICKEL-BASE, INCLUDING BORIDE-CONTAINING ALLOYS, CARBIDE-CONTAINING ALLOYS, AND LAVES PHASE ALLOYS
- COPPER-BASE ALLOYS

Also included in this section is a discussion of several cobalt-substitute materials, some iron-base or nickel-base, which share with the cobalt-base alloys resistance to a variety of wear conditions and corrosive media.

Whereas the ferrous hardfacing alloys and tungsten carbide composites find extensive use in agriculture, mining, construction, and the steel industry, the nonferrous alloys are typically used in the chemical processing, power, automotive, and oil industries, all of which require resistance to a hostile environment in addition to resistance to wear. The cobalt-base alloys are especially resistant to deformation and chemical attack at high temperatures (500 °C < $T \leq 900$ °C, or 930 °F < $T \leq 1650$ °F) and are used to protect dies and guide rolls in the steel industry. Other common applications of the nonferrous hardfacing alloys include valve seating surfaces (both control valves and diesel exhaust valves), pump parts, extrusion screw flights, rock bit bearings, marine bearings, and glass molding hardware.

Iron-Base Alloys (Ref 1)

As described above, iron-base hardfacing alloys are more widely used than cobalt- and/or nickel-base hardfacing alloys and constitute the largest volume use of hardfacing alloys. Iron-base hardfacing alloys offer low cost and a broad range of desirable properties. Most equipment that undergoes severe wear, such as crushing and grinding equipment and earthmoving equipment, is usually very large and rugged. Parts subjected to wear usually require downtime for repair. For this reason, there is a general temptation to hardface them with the lowest cost and most readily available materials. As a result, literally hundreds of iron-base hardfacing alloys are in use today.

Due to the great number of alloys involved, iron-base hardfacing alloys are best classified by their suitability for different types of wear and their general microstructures rather than by chemical composition. Most iron-base hardfacing alloys can be divided into the following classes:

- PEARLITIC STEELS
- AUSTENITIC STEELS
- MARTENSITIC STEELS
- HIGH-ALLOY IRONS

Pearlitic steels are essentially low-alloy steels with minor adjustments in composition to achieve weldability. These alloys contain low carbon (<0.2% C) and low amounts of other alloying elements (for example, up to 2% Cr), resulting in a pearlitic structure. Pearlitic steels are useful as buildup overlays, primarily to rebuild carbon steel or low-alloy steel machinery parts back to size. Examples include shafts, rollers, and other parts in heavy machinery subjected to rolling, sliding, or impact loading. Typically, this group of alloys has high impact resistance and low hardness (in the range 25 to 37 HRC), as well as excellent weldability. Table 1 lists the composition and properties of a typical low-alloy pearlitic steel (EFe1) used for buildup applications.

TABLE 1 COMPOSITIONS, HARDNESS, AND ABRASION DATA FOR A PEARLITIC LOW-ALLOY STEEL (EFE1) AND AUSTENITIC MANGANESE STEELS (EFEMN-C AND EFEMN-CR) USED FOR BUILDUP OVERLAYS AND AIR-HARDENING OF MARTENSITIC STEELS USED FOR METAL-TO-METAL WEAR APPLICATIONS

ALLOY	COMPOSITION, WT%							HARDNESS, HRC	ABRASION, VOLUME LOSS			
									LOW-STRESS(C)		HIGH-STRESS(D)	
	Fe	Cr	C	Si	Mn	Mo	Ni		MM^3	$IN.^3 \times 10^{-3}$	MM^3	$IN.^3 \times 10^{-3}$
BUILDUP WELD OVERLAY												
EFE1(A)	BAL	2	0.1	1.0	1	1.5	. . .	37	88	5.4	49	3.0
EFEMN-C(A)	BAL	4	0.8	1.3	14	. . .	4	18	65	4.0	57	3.5
EFEMN-CR(A)	BAL	15	0.5	1.3	15	2.0	1	24	93	5.7	46	2.8
METAL-TO-METAL WELD OVERLAY												
EFE2(A)	BAL	3	0.2	1.0	1	1.0	1	48	54	3.3	66	4.0
EFE3(A)	BAL	6	0.7	1.0	1	1.0	. . .	59	60	3.7	68	4.1
ER420(B)	BAL	12	0.3	1.0	2	. . .	. . .	45	84	5.1	62	3.8

(A) TWO-LAYER SMA DEPOSIT PROCESS.

(B) TWO-LAYER SAW DEPOSIT PROCESS.

(C) DRY SAND/RUBBER WHEEL TEST (ASTM G 65, PROCEDURE B): LOAD 13.6 KG (30 LB); 2000 REV.

(D) SLURRY/STEEL WHEEL TEST (ASTM B 611, MODIFIED): LOAD 22.7 KG (50 LB); 250 REV

Austenitic steels are usually stabilized by manganese additions. Austenitic iron-base hardfacing alloys essentially are modeled after Hadfield steels. Most commercially available alloys in this category can be broadly subdivided into low-chromium and high-chromium alloys.

Low-chromium alloys usually contain up to 4% Cr and 12 to 15% Mn and some nickel or molybdenum (Table 1). Low-chromium austenitic steels generally are used to build up manganese steel machinery parts subjected to high impact (impact crusher or shovel lips).

High-chromium austenitic steels, which may normally contain 12 to 17% Cr in addition to about 15% Mn, are used for the buildup and joining of austenitic manganese steels as well as carbon and low-alloy steels. In addition, the as-deposited hardness of high-chromium steels is higher (~24 HRC) than that of low-chromium steels (~18 HRC).

The high strength of the austenitic manganese steels is the result of a synergism between manganese and carbon (Ref 3). In suppressing the formation of phases other than austenite, manganese not only increases carbon solubility at lower temperatures but also encourages carbon supersaturation of the structure. Both high inherent strength and a high work-hardening rate result from this.

Because the austenitic manganese steels are metastable, problems of carbide embrittlement tend to arise when the alloys are cooled slowly or reheated. Manganese steel components are therefore kept as cool as possible during the buildup (repair) process. Often, the bulk of the part is submerged in water during the welding process. Compositions and properties of austenitic manganese steel hardfacing alloys used for buildup applications are given in Table 1.

Martensitic steels are designed to form martensite on normal air cooling of the weld deposit. As a result, these steels are often termed "self-hardening" or "air-hardening," and they resemble tool steels with hardnesses in the range of 45 to 60 HRC. The carbon content of the martensitic steels ranges up to 0.7%. Other elements such as molybdenum, tungsten, nickel, and chromium (up to 12%), are added to increase hardenability and strength and to promote martensite formation. Manganese and silicon usually are added to aid weldability.

The major hardfacing applications for martensitic steels include unlubricated metal-to-metal rolling or sliding parts such as undercarriage parts of tractors. The impact resistance of martensitic steels is inferior to that of pearlitic or austenitic alloys, but there is a compensating increase in hardness and resistance to abrasive wear.

In hostile environments, a higher chromium content is beneficial. AWS ER420 and modified versions containing nickel, molybdenum, and niobium (or vanadium) are therefore the natural choice when high temperatures and mildly corrosive environments are encountered. For applications using steel mill hot-work rolls (which demand considerable hot hardness, resistance to oxidation, and resistance to thermal fatigue) both ER420 and EFe3 have been found suitable. Other applications for the metal-to-metal martensitic wear alloys in Table 1 include crane wheels (EFe2), pincer guide shoes (EFe3), and blast furnace bells (ER420).

The high-chromium irons encompass a wide range of compositions in which chromium may vary between approximately 6 and 35 wt%, and carbon may vary from about 2 to 6 wt%. Other possible alloying additions include molybdenum, manganese, and silicon.

The most important microstructural feature in the high-chromium irons, at least from a wear standpoint, is an M_7C_3 carbide, which forms in abundance during solidification and contains chromium, iron, and (if present) molybdenum. The matrix around these carbide particles can be austenitic, pearlitic, or martensitic (Ref 4). In general, the austenitic alloys rely on manganese for austenite stability.

Carbon content is a good indicator of abrasion resistance for this class of materials. To illustrate this fact, low- and high-stress data are plotted for selected alloys in Fig. 1. This information was generated by the use of six different alloys (open arc/flux-cored wires) and nine different sets of welding parameters (Ref 5). The carbon contents referred to in the figures are those of the second layer of the two-layer overlays tested. The use of a shaded zone, rather than a line, indicates that there was considerable scatter in the data (particularly at the lower carbon contents, in the case of high-stress abrasion). The test methods and parameters were identical to those described in the footnotes to Table 1 in this article.

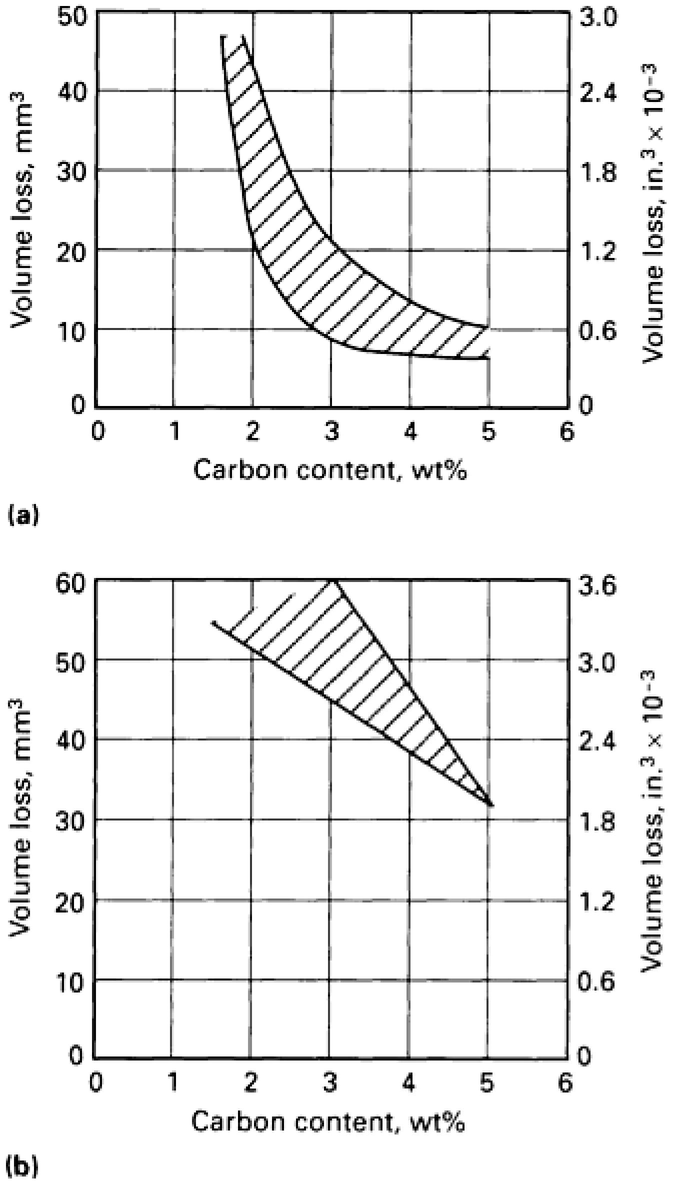

FIG. 1 PLOT OF VOLUME LOSS VERSUS CARBON CONTENT FOR HIGH-CHROMIUM WHITE IRON METAL-TO-EARTH ABRASION HARDFACING ALLOYS. (A) LOW-STRESS CONDITION. (B) HIGH-STRESS CONDITION. SOURCE: REF 5

At high carbon and chromium levels, the formation of a hypereutectic microstructure, containing large, spinelike carbide particles (with a hexagonal cross section), is favored. At lower carbon and chromium contents, the microstructure is hypoeutectic. The microstructures of three high-chromium irons (as deposited by the open arc welding process) are shown in Fig. 2. The nominal compositions of the same three alloys are presented in Table 2. As can be seen, ERFeCr-A3 (at a chromium content of 11 wt% and a carbon content of 2.6 wt%) exhibits a hypoeutectic (primary austenite) microstructure. The other two alloys, ERFeCr-A4(Mod) (29Cr-3.5C) and ERFeCr-A2 (28Cr-4.3C), possess a hypereutectic (primary carbide) structure. In Fig. 2, the large, spinelike carbides are shown in cross section.

TABLE 2 COMPOSITIONS OF HIGH-CHROMIUM WHITE IRONS USED FOR METAL-TO-EARTH ABRASION APPLICATIONS

ALLOY	COMPOSITION, WT%							
	Cr	C	Si	Mn	Mo	Ni	B	Fe

ERFECR-A3	11	2.6	1.3	1.8	1.5	. . .	. . .	BAL
ERFECR-A4(MOD)	29	3.5	1.1	0.9	. . .	2.6	0.7	BAL
ERFECR-A2	28	4.3	0.8	1.7	1.4	. . .	. . .	BAL

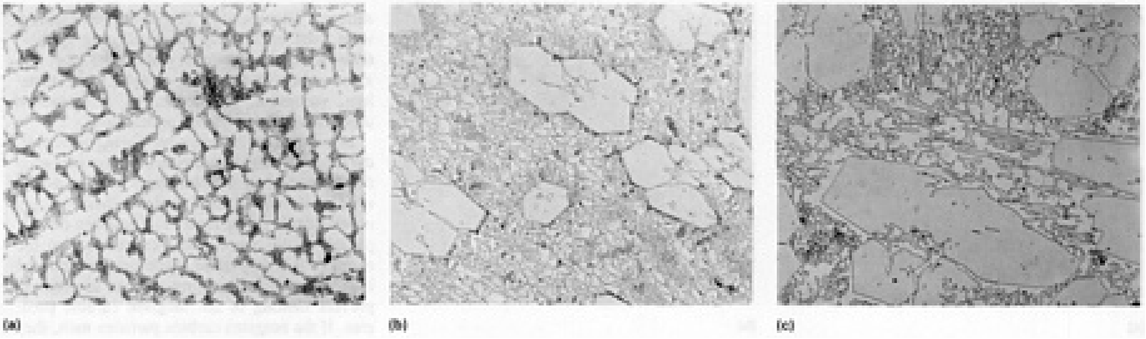

FIG. 2 MICROSTRUCTURES OF HIGH-CHROMIUM WHITE IRON METAL-TO-EARTH ABRASION ALLOYS HARDFACED WITH TWO-LAYER FLUX-COLORED OPEN ARC DEPOSIT. (A) ERFECR-A3. (B) ERFECR-A4(MOD). (C) ERFECR-A2. 300×. SOURCE: REF 2

In addition to M_7C_3, deposits of ERFeCr-A2 contain small quantities of M_6C, and deposits of ERFeCr-A4(Mod) contain small quantities of both M_6C and M_3C. With regard to the matrix, both ERFeCr-A3 and ERFeCr-A4(Mod) exhibit a face-centered cubic (fcc) austenitic structure as deposited. ERFeCr-A2 is largely austenitic but may also contain small quantities of ferrite or martensite. The presence of martensite in the matrix structure of a high-chromium iron is believed to be beneficial to its high-stress abrasion resistance, because of the additional support martensite provides to surface carbides.

For a given hardfacing consumable, deposit (overlay) microstructure is strongly influenced by the welding process and parameters used. Factors of concern, in this regard, include:

- OVERLAY/SUBSTRATE INTERMIXING
- CHANGES IN COMPOSITION DUE TO LOSSES IN THE ARC
- DEPOSIT COOLING RATE

With regard to industrial applications of the high-chromium metal-to-earth abrasion alloys, the low-carbon (2 to 3% C) hypoeutectic materials are usually selected for situations involving moderate abrasion and impact, whereas the higher-carbon (4 to 6% C) hypereutectic alloys are used in applications involving severe abrasion and little or no impact. Specific uses include crusher cones and pump casings (ERFeCr-A3), bulldozer blades and crusher hammers [ERFeCr-A4(Mod)], and coal pulverizer liners and gravel pumps (ERFeCr-A2).

Carbides (Ref 1, 2)

The amount of carbides used for hardfacing applications is small compared to iron-base hardfacing alloys, but carbides are extremely important for severe abrasion and cutting applications. Historically, tungsten-base carbides were used exclusively for hardfacing applications. Recently, however, carbides of other elements, such as titanium, molybdenum, tantalum, vanadium, and chromium, have proved useful in many hardfacing applications.

In contrast to the other hardfacing materials, the tungsten carbide composites do not rely upon the formation of suitable hard phases during weld pool solidification. Instead, these overlay materials rely on the transfer of tungsten carbide particles from the welding consumable to the overlay. It is important, therefore, to limit the heat input of the welding process in order to prevent melting of the tungsten carbide particles. If the tungsten carbide particles melt, they mix with iron to form much softer iron-tungsten carbides, thus reducing abrasion resistance. For this reason, oxyacetylene deposits usually exhibit higher abrasion resistance than arc-welded tungsten carbide overlays. An advantage of the tungsten carbide composites is that the size of the hard particles in the overlay can be controlled. This is important because abrasion resistance is dependent upon the size relationship between microstructural features (such as carbides) and the abrading particles. It is believed that if the abrading particles are large in comparison to the microstructural particles, then, after a running-in period (during which the softer matrix material at the surface is worn down), the abrading particles ride over the hard microstructural outcrops. On the other hand, if the abrading particles are small in comparison to the microstructural particles, it is believed that the opportunity exists for wear of the matrix around the microstructural particles. Eventually, these may drop out, having played only a small role in resisting abrasion.

Several tungsten carbide composites are available in a variety of tubular product forms. Popular compositions are 40, 50, 55, and 60 wt% tungsten carbide, with a carbon steel tube making up the balance. For each composition, several carbide size ranges are available. As an example, for the 60% tungsten carbide oxyacetylene welding consumable, four mesh size ranges are available:

AWS DESIGNATION	MESH SIZE RANGE
RWC-12/20	12-20
RWC-20/30	20-30
RWC-30/40	30-40
RWC-40/120	40-120

The same composition is also available in flux-coated form for shielded metal arc welding and as a continuous wire (with an internal flux) for open arc welding.

Tungsten carbide composites generally possess very high resistance to abrasion and very low impact strength. Performance in a given situation is dependent upon:

- CARBIDE VOLUME FRACTION
- SIZE RELATIONSHIP BETWEEN THE CARBIDES AND THE ABRASIVE MEDIUM
- WELDING TECHNIQUE APPLIED

Important factors are the distribution of carbides in the overlay (because the particles tend to sink, turbulence in the molten weld pool is an advantage), and the amount of carbide dissolution and reprecipitation in the steel matrix during welding. Impact strength generally decreases with increasing carbide volume fraction.

Table 3 relates tungsten carbide composites and the previously mentioned alloys on the basis of low-stress and high-stress abrasion data (generated under identical test conditions). These values should be compared with the values in Table 1 and Fig. 1. Micrographs of the test deposits are shown in Fig. 3. Noteworthy features include:

- THE DIFFERENCE IN SIZE OF THE TUNGSTEN CARBIDE PARTICLES IN THE TWO MATERIALS
- THE EXTENT TO WHICH SECONDARY CARBIDES HAVE PRECIPITATED WITHIN THE MATRIX

TABLE 3 ABRASION DATA FOR TUNGSTEN CARBIDE HARDFACING MATERIALS

MATERIAL		**ABRASION, VOLUME LOSS**			
		LOW-STRESS(A)		**HIGH-STRESS(B)**	
CARBIDE, WT%	**MESH SIZE**	**mm^3**	**$in.^3 \times 10^{-3}$**	**mm^3**	**$in.^3 \times 10^{-3}$**
60	20-30	7.3	0.45	28.7	1.75
61	100-250	10.6	0.65	24.4	1.49

(A) DRY SAND/RUBBER WHEEL TEST (ASTM G 65. PROCEDURE B): LOAD 13.6 KG (30 LB); 2000

REV.
(B) SLURRY/STEEL WHEEL TEST (ASTM B 611, MODIFIED): LOAD 22.7 KG (50 LB): 250 REV

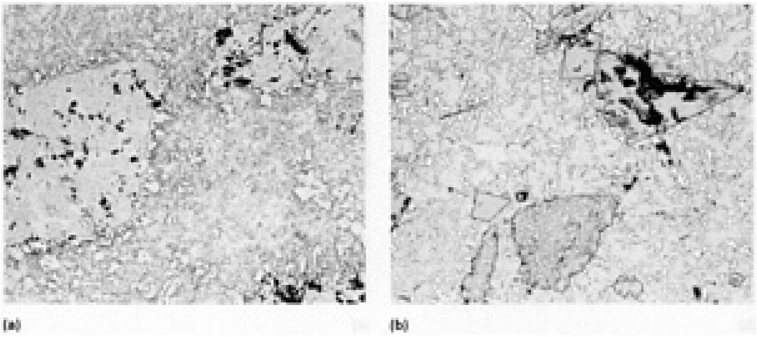

FIG. 3 MICROSTRUCTURES OF TUNGSTEN CARBIDE COMPOSITES WITH CARBIDES OF DIFFERENT SIZES IN THE HARDFACING MATERIAL. (A) 60% TUNGSTEN CARBIDE, 20 TO 30 MESH PARTICLES. (B) 61% TUNGSTEN CARBIDE, 100 TO 250 MESH PARTICLES. HARDFACING PROCESS APPLIED IS ONE-LAYER SHIELDED METAL ARC DEPOSIT. 120 ×. SOURCE: REF 2

The tungsten carbide composites have been used to solve a wide variety of industrial sliding and drilling abrasion problems. Some of the common applications are plowshares, ditchdigger teeth, ripper teeth, and oil well drilling tools. For extremely hostile environments, some nonferrous tungsten carbide products (cobalt- and nickel-base products in the form of bare cast rods) are available. Also, several alternative composite materials, utilizing other carbides (for example, vanadium, titanium, or niobium), are available that have the advantage of creating a more homogeneous deposit because of their lower densities.

Cobalt-Base Alloys (Ref 1, 2)

The two types of commercially available cobalt-base hardfacing alloys are carbide-containing alloys and alloys containing Laves phase. Compositions and properties of cobalt-base hardfacing alloys are given in Tables 4 and 5, respectively.

TABLE 4 COMPOSITIONS OF SELECTED COBALT-BASE HARDFACING ALLOYS

ALLOY	COMPOSITION, WT%					
	CR	MO	W	SI	C	CO
CARBIDE-CONTAINING ALLOYS						
ERCOCR-A	28	. . .	5	. . .	1.2	BAL
ERCOCR-B	29	. . .	8	. . .	1.5	BAL
ERCOCR-C	31	. . .	13	. . .	2.5	BAL
ERCOCR-E	27	6	. . .	. . .	0.2	BAL
LAVES PHASE ALLOYS						
T-400	9	29	. . .	2.5	. . .	BAL
T-800	18	29	. . .	3.5	. . .	BAL

TABLE 5 PROPERTIES OF SELECTED COBALT-BASE HARDFACING ALLOYS

PROPERTY	ALLOY 21[A]	RCOCRA	RCOCRB	RCOCRC	TRIBALOY T-800
DENSITY, G/CM^3 (LB/IN.^{3})	8.3 (0.30)	8.3 (0.30)	8.6 (0.31)	8.6 (0.31)	8.6 (0.31)
ULTIMATE COMPRESSIVE STRENGTH,	1295	1515	1765	1930	1780

MPA (KSI)	(188)	(220)	(256)	(280)	(258)
ULTIMATE TENSILE STRENGTH, MPA (KSI)	710 (103)	834 (121)	827 (120)	620 (90)	. . .
ELONGATION, %	8	1.2	1	1	
COEFFICIENT OF THERMAL EXPANSION, $°C^{-1}$ ($°F^{-1}$)	14.8×10^{-6}	15.7×10^{-6}	14×10^{-6}	13.1×10^{-6}	12.3×10^{-6}
	(8.2×10^{-6})	(8.7×10^{-6})	(7.8×10^{-6})	(7.3×10^{-6})	(6.8×10^{-6})
HOT HARDNESS, DPH AT:					
445 °C (800 °F)	150	300	345	510	659
540 °C (1000 °F)	145	275	325	465	622
650 °C (1200 °F)	135	260	285	390	490
760 °C (1400 °F)	115	185	245	230	308
WEAR					
UNLUBRICATED SLIDING WEAR^(B)^, MM^3 ($IN.^3 \times 10^3$) AT:					
670 N (150 LBF)	5.2 (0.32)	2.6 (0.16)	2.4 (0.15)	0.6 (0.04)	1.7 (0.11)
1330 N (300 LBF)	14.5 (0.90)	18.8 (1.17)	18.4 (1.14)	0.8 (0.05)	2.1 (0.13)
ABRASIVE WEAR(C), MM^3 ($IN.^3 \times 10^{-3}$)					
OAW	. . .	29 (1.80)	12 (0.75)	8 (0.50)	. . .
GTAW	86 (5.33)	64 (3.97)	57 (3.53)	52 (3.22)	24 (1.49)
UNNOTCHED CHARPY IMPACT STRENGTH, J (FT · LBF)	37 (27)	23 (17)	5 (4)	5 (4)	1.4 (1)
CORROSION RESISTANCE(D):					
65% NITRIC ACID AT 65 °C (150 °F)	U	U	U	U	S
5% SULFURIC ACID AT 65 °C (150 °F)	E	E	E	E	. . .
50% PHOSPHORIC ACID AT 400 °C (750 °F)	E	E	E	E	E

(A) CO-27CR-5MO-2.8NI-0.2C.

(B) WEAR MEASURED FROM TESTS CONDUCTED ON DOW-CORNING LFW-1 AGAINST 4620 STEEL RING AT 80 REV/MIN FOR 2000 REV VARYING THE APPLIED LOAD.

(C) WEAR MEASURED FROM DRY SAND RUBBER WHEEL ABRASION TESTS. TESTED FOR 2000 REV AT A LOAD OF 135 N (30 LBF) USING A 230 MM (9 IN.) DIAM RUBBER WHEEL AND AMERICAN FOUNDRYMEN'S SOCIETY TEST SAND.

(D) E, LESS THAN 2 MILS/YEAR; S, OVER 20 TO LESS THAN 50 MILS/YEAR; G, LESS THAN 20 MILS/YEAR; U, MORE THAN 50 MILS/YEAR

Carbide-containing cobalt-base alloys have been widely used since the early 1900s, when a cobalt-base alloy with the nominal composition of Co-28Cr-4W-1.1C was first developed. The chief difference between the various carbide-containing cobalt-base alloys is in carbon content (hence, carbide volume fraction, room-temperature hardness, and level of abrasion resistance). Chromium-rich M_7C_3 is the predominant carbide in these alloys, although tungsten-rich M_6C is evident in those alloys having a high tungsten content, and chromium-rich $M_{23}C_6$ is common in the low-carbon alloys. Figure 4 relates the carbon content and the resistance to low-stress abrasion for the carbide-containing cobalt-base alloys (ERCoCr-A, -B, -C, and -E). The low-stress abrasion data in this figure were generated using the ASTM G 65 dry sand/rubber wheel test.

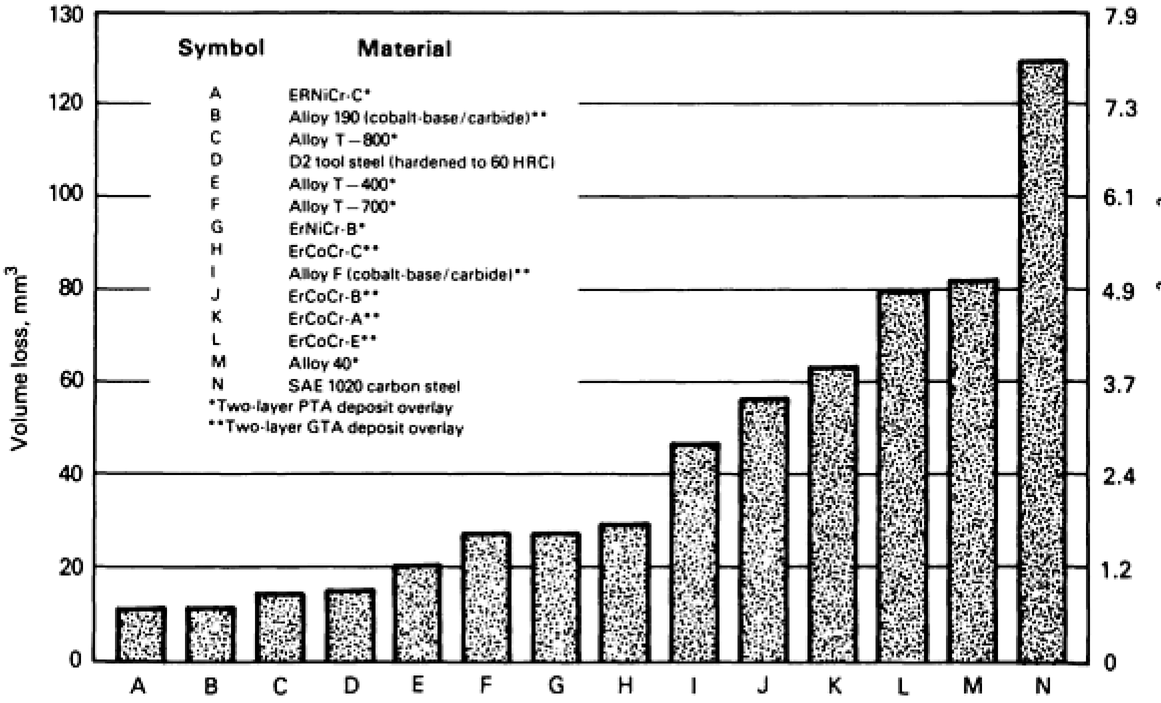

FIG. 4 COMPARISON OF NONFERROUS HARDFACING ALLOYS TO TOOL STEEL AND CARBON STEEL REFERENCE MATERIALS USING ASTM G 65 LOW-STRESS ABRASION TEST. G 65 TEST PARAMETERS: PROCEDURE B; ROOM TEMPERATURE; 13.6 KG (30 LBF) LOAD; QUARTZ GRAIN SAND DIAMETER OF 212 TO 300 μM; 2000 REV AT 200 REV/MIN; 390 G/MIN (0.86 LB/MIN) FEED RATE. SOURCE: REF 2

Micrographs of plasma-transferred arc (two-layer) deposits of three carbide-containing cobalt-base alloys are shown in Fig. 5. The differences in carbide volume fraction and geometry are evident from these micrographs.

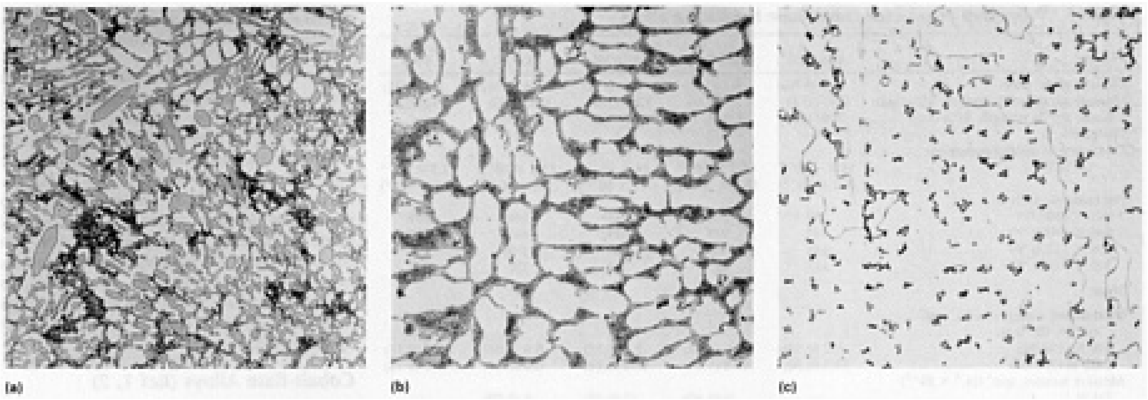

FIG. 5 MICROSTRUCTURES OF TWO-LAYER PLASMA TRANSFERRED ARC DEPOSIT CARBIDE-CONTAINING COBALT-BASE HARDFACING ALLOYS. (A) ERCOCR-C. (B) ERCOCR-A. (C) ERCOCR-E. 425×. SOURCE: REF 2

Laves Phase Alloys. Laves phase is a type of topologically closed-packed intermetallic compound. Historically, metallurgists have avoided the presence of Laves phase in most alloys because of its detrimental effect on mechanical properties. However, in the early 1960s, the usefulness of Laves phase in resisting metal-to-metal wear was discovered; subsequently, alloys containing Laves phase have become commercially available. As listed in Table 4, there are two different Laves phase-containing cobalt-base alloys commercially available for hardfacing applications: Co-29Mo-9Cr-2.5Si (T-400) and Co-29Mo-18Cr-3.5Si (T-800). Both of these alloys contain at least 50 vol% of Laves phase, bound in a cobalt-base matrix alloyed with chromium and molybdenum.

Because the Laves intermetallic phase is so abundant in these alloys, its presence governs all the material properties. Accordingly, the effects of the matrix composition in these alloys are less pronounced than is the case for the carbide-containing cobalt-base alloys. for example. The Laves phase is specifically responsible for outstanding abrasion resistance, but it severely limits the material ductility and the impact strength. In fact, it is difficult to attain crack-free overlays on all but the smallest components given adequate preheat. For this reason, these alloys have been more successful as thermal spray materials. Figure 4 compares the abrasion resistance of various nonferrous hardfacing alloys, including T-400 and T-800 Laves phase cobalt-base alloys. The microstructure of a two-layer plasma transferred arc (PTA) deposit of alloy T-800 is shown in Fig. 6.

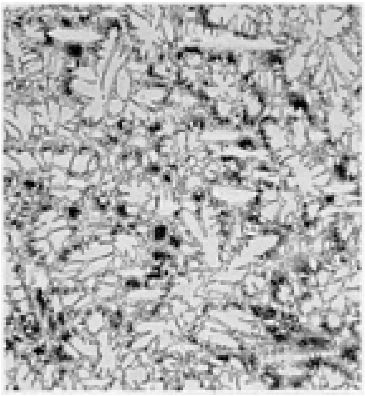

FIG. 6 MICROSTRUCTURE OF T-800 (TWO-LAYER PLASMA TRANSFERRED ARC DEPOSIT) COBALT-BASE LAVES PHASE HARDFACING ALLOY. 200×. SOURCE: REF 2

Nickel-Base Alloys (Ref 1, 2)

Most commercially available nickel-base hardfacing alloys can be divided into three groups: boride-containing alloys, carbide-containing alloys, and Laves phase-containing alloys. The compositions of some typical nickel-base hardfacing alloys are listed in Table 6. Selected properties of these alloys are listed in Table 7.

TABLE 6 COMPOSITION OF NICKEL-BASE AND HIGH-NICKEL-CONTENT HARDFACING ALLOYS

ALLOY	COMPOSITION, WT%								
	Fe	Cr	Mo	W	Si	C	B	Co	Ni
BORIDE-CONTAINING ALLOYS									
ALLOY 40	1.5	7.5	. . .	. . .	3.5	0.3	1.5	. . .	BAL
ERNICR-B	3	11	. . .	. . .	4	0.5	2.5	. . .	BAL
ERNICR-C	4	16	. . .	. . .	4	0.7	3.5	. . .	BAL
CARBIDE-CONTAINING ALLOYS									
ALLOY N-6	3	29	5.5	2	1.5	1.1	0.6	3	BAL
ALLOY 716	BAL	26	3	3.5	1.5	1.1	0.5	11	23
LAVES PHASE ALLOYS									
T-700	. . .	16	33	. . .	3.5	. . .	. . .	. . .	BAL

TABLE 7 PROPERTIES OF SELECTED NICKEL-BASE HARDFACING ALLOYS

PROPERTY	RNICRC	RNICRB	HASTELLOY C (NI-17CR-17MO-0.12C)	TRIBALOY T-700
DENSITY, G/CM^3 (LB/IN.^{3})	7.8 (0.28)	8.0 (0.29)	8.6 (0.32)	8.6 (0.32)
COEFFICIENT OF THERMAL EXPANSION, °C^{-1} (°F^{-1})	14.3 (8.0)	14.3 (8.0)	13.7 (7.6)	11.9 (6.6)
HOT HARDNESS, DPH AT: 425 °C (800 °F)				

425 °C (800 °F)	555	. . .	190	500
540 °C (1000 °F)	440	. . .	185	485
650 °C (1200 °F)	250	. . .	170	400
760 °C (1400 °F)	115	. . .	145	280
WEAR				
UNLUBRICATED SLIDING WEAR(A), MM^3 (IN.^{3} × 10^{-3}) AT:				
670 N (150 LBF)	0.15 (0.009)	0.3 (0.018)	0.4 (0.024)	0.1 (0.006)
1330 N (300 LBF)	0.3 (0.018)	0.4 (0.024)	. . .	0.3 (0.018)
ABRASIVE WEAR(B), MM^3 (IN.^{3} × 10^{-3})				
OAW	12 (0.73)	18 (1.10)	. . .	. . .
GTAW	11 (0.67)	12 (0.73)	105 (6.40)	43 (2.62)
UNNOTCHED CHARPY IMPACT STRENGTH, J (FT · LBF)	3 (2)	3 (2)	39 (29)	1.4 (1)
CORROSION RESISTANCE AT 65 °C (150 °F)(C): 65% NITRIC ACID				
65% NITRIC ACID	U	U	E	E
5% SULFURIC ACID	U	U	E	E
50% PHOSPHORIC ACID	U	U	E	E

(A) WEAR MEASURED FROM TESTS CONDUCTED ON DOW-CORNING LFW-1 AGAINST 4620 STEEL RING AT 80 REV/MIN FOR 2000 REV VARYING THE APPLIED LOADS.

(B) WEAR MEASURED FROM DRY SAND RUBBER WHEEL ABRASION TESTS. TESTED FOR 2000 REV AT A LOAD OF 135 N (30 LBF) USING A 230 MM (9-IN.) DIAM RUBBER WHEEL AND AMERICAN FOUNDRYMEN'S SOCIETY TEST SAND.

(C) E, LESS THAN 2 MILS/YEAR; S, OVER 20 TO LESS THAN 50 MILS/YEAR; G, LESS THAN 20 MILS/YEAR; U, MORE THAN 50 MILS/YEAR

The boride-containing nickel-base alloys were first commercially produced as spray-and-fuse powders. The alloys are currently available from most manufacturers of hardfacing products under various tradenames and in a variety of forms, such as bare cast rod, tube wires, and powders for plasma weld and manual torch. This group of alloys is primarily composed of nickel, chromium, boron, silicon, and carbon. Usually, the boron content ranges from 1.5 to 3.5%, depending on chromium content, which can be as high as 16%. The higher chromium alloys generally contain a large amount of boron, which forms very hard chromium borides with hardnesses of approximately 1800 DPH (kg/mm^2).

Of all the hardfacing alloys, the boride-containing nickel-base alloys are microstructurally the most complex. The complexity of this type of alloy structure is evident in Fig. 7, which shows a two-layer PTA deposit of ERNiCr-C (see Table 6 for composition). The alloy compositions represent a progression in terms of iron, chromium, boron, and carbon contents. Iron content is largely incidental, allowing the use of ferrocompounds during manufacture. Together with nickel, the other three elements determine the level and type of hard phase within the structure upon solidification, boron being the primary hard-phase forming element (for which nickel and chromium. compete) and carbon being the secondary hard-phase former. The actual phases that form in boride-containing nickel-base alloys are listed in Table 8 on the basis of chromium content.

TABLE 8 PHASES FORMED IN BORIDE-CONTAINING NICKEL-BASE HARDFACING ALLOYS

CHROMIUM CONTENT	**SECONDARY PHASES**		**DOMINANT HARD PHASE**
	COMPOUND FORMED	**REQUIRED CONDITIONS**	
LOW (~5 WT%)	NI_3SI	>3 WT% SI	NI_3B
MEDIUM (~15 WT%)	NI_3SI	>2.5 WT% SI	NI_3B AND CHROMIUM BORIDE (USUALLY CRB, ALTHOUGH CR_2B AND CR_3B_2 MAY ALSO BE PRESENT)
HIGH (~25	NI_3SI	>3 WT% SI	CRB AND CR_5B_3

WT%)			
ALL LEVELS	. . .	. . .	COMPLEX CARBIDES OF $M_{23}C_6$ AND M_7C_3 TYPES

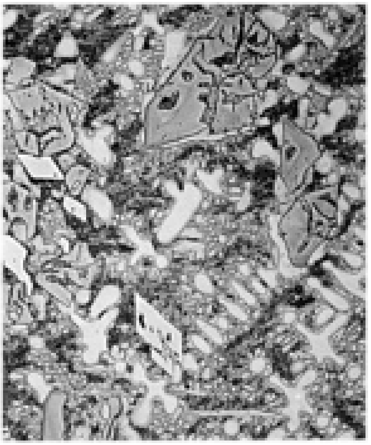

FIG. 7 MICROSTRUCTURE OF ERNICR-C (TWO-LAYER PLASMA TRANSFERRED ARC DEPOSIT) BORIDE-CONTAINING NICKEL-BASE HARDFACING ALLOY. 425×. SOURCE: REF 2

The chief purpose of silicon in the material is to provide, in conjunction with boron, self-fluxing characteristics. However, as an important matrix element and as a potential promoter of intermetallic precipitates, it also has a powerful influence on the wear properties of the alloys.

Boron content influences the level of silicon required for silicide (Ni_3Si) formation. The higher the boron content, the lower is the silicon content required to form silicides.

Because of the boride and carbide dispersions within their microstructures, the boride-containing nickel-base alloys exhibit excellent resistance to abrasion (ERNiCr-C, ERNiCr-B, and Alloy 40 in Fig. 4). Low-stress abrasion resistance generally increases with boron and carbon contents, hence the hard-phase volume fraction for these materials. Although their performance is not as good under self-mated sliding conditions as that of cobalt-base materials, the boride-containing nickel-base alloys possess moderate resistance to galling. Of the nonferrous materials, the boride-containing nickel-base alloys are the least resistant to corrosion. This is attributed to the lack of chromium in the matrix that follows boride and carbide formation.

Originally, the boride-containing nickel-base alloys were developed from brazing alloy compositions specifically for use with the spray-and-fuse process (for which their self-fluxing properties are ideal). During the fusing cycle of this process, it is believed that oxides within the sprayed coating combine with some of the silicon and boron to form a borosilicate slag (which floats to the surface of the deposit). The only other commercially significant spray-and-fuse alloys are some silicon-containing and boron-containing derivatives of the carbide-containing cobalt-base materials. Microstructurally, and in a property sense, these alloys bear little resemblance to their traditional counterparts because the presence of boron markedly changes the matrix chromium content and the types of hard phase that form during solidification. Additional information on the spray-and-fuse process can be found later in this article in the section on "Thermal Spraying."

Carbide-Containing Alloys. The use of carbide-containing nickel-base alloys has been somewhat limited to date. However, these alloys are gaining popularity as alternatives to cobalt-base hardfacing alloys. This is particularly applicable in the nuclear power industry, where it has been found that high-cobalt hardfacing alloys are sources of the highly radioactive ^{60}Co isotope that deposits in pipes, fittings, and valve bodies (Ref 6). Wear and corrosion of hardfacing alloys lead to the release of ^{59}Co into the primary coolant. The ^{59}Co deposits on the reactor core, where it is transmuted to the radioactive ^{60}Co isotope, which is then released back to the coolant. The most popular and widely used cobalt-free alloys are in the Ni-Cr-Mo-C and Ni-Cr-Fe-C systems. Table 6 lists several compositions of these alloys.

Generally, carbide-containing nickel-base alloys have poor oxyacetylene weldability, which has made then unsuitable for hardfacing applications in the past. Carbide-containing nickel-base alloys of the Ni-Cr-Mo-Co-Fe-W-C system are more readily weldable by oxyacetylene methods and are gaining popularity as low-cost alternatives to cobalt-base alloys. These alloys, depending on precise composition, contain M_7C_3- or M_6C-type carbides similar to cobalt-base alloys.

Laves Phase Alloys. Only one Laves phase-containing nickel-base alloy is commercially available: Ni-33Mo-16Cr-3.5Si (T-700 in Table 6). This alloy, like most nickel-base alloys, is difficult to weld using the oxyacetylene process, but it can be readily welded using gas-tungsten arc welding or the PTA process. It can also be applied using the plasma or high-velocity oxyfuel thermal spray techniques. Although it has excellent metal-to-metal wear resistance and moderate abrasive wear resistance, it possesses poor impact resistance.

High-Silicon Stainless Steels (Ref 2)

Recently, several new high-silicon stainless steels have been introduced that may also be considered as alternate materials to cobalt-base alloys. These exhibit identical anti-galling characteristics and possess equal resistance to cavitation erosion. Some of these have been developed for general use and others for specific applications (for example, hydroelectric turbines and nuclear valves). Some are based on type 200 (manganese- and nitrogen-containing) austenitic stainless steel; others are based on type 300 stainless steel. One of these latter compositions is as follows (Ref 7):

ELEMENT	COMPOSITION, WT%
CARBON	0.12
CHROMIUM	21.0
NICKEL	8.0
SILICON	5.0
MANGANESE	6.5
IRON	BAL

These high-silicon hardfacing alloys differ from the traditional carbide-containing cobalt-base alloys in terms of their corrosion resistance (they are better in some aqueous media and worse in others), their mechanical properties (particularly at high temperatures), and their thermal stability.

Copper-Base Alloys (Ref 2)

Copper-base hardfacing alloys fall in the aluminum bronze category. Examples include ECu-AlB (Cu-9Al-4Fe-1Si) and ECuAl-D (Cu-13.5Al-4Fe). Silicon and other types of bronze are also available in welding consumable form for hardfacing. Aluminum, silicon, and iron are added to the bronzes to strengthen the solid solution, and, in excess of their solubility limits (about 8%, in the case of aluminum), to provide precipitation hardening.

Wear tests have shown that the aluminum bronzes possess anti-galling characteristics (self-coupled) comparable to those of ER-CoCr-E; low-stress abrasion resistance of the aluminum bronzes, however, was very low. Components typically protected using the aluminum bronzes include gears, cams, bearings, and cold drawing/forming dies. These alloys are not recommended for elevated-temperature use because their mechanical properties considerably decrease at temperatures >200 °C (>390 °F).

Hardfacing Alloy Selection (Ref 1)

Hardfacing alloy selection is guided primarily by wear and cost considerations. However, other manufacturing and environmental factors must also be considered, such as base metal; deposition process; and impact, corrosion, oxidation, and thermal requirements. The factors affecting hardfacing process selection are discussed in the next section of this article. Usually, the hardfacing process dictates the hardfacing or filler-metal product form.

Hardfacing alloys usually are available as bare rod, flux-coated rod, long-length solid wires, long-length tube wires (with and without flux), or powders. The most popular processes, and the forms most commonly associated with each process, are (Ref 2):

HARDFACING PROCESS	**CONSUMABLE FORM**
OXYFUEL/OXYACETYLENE (OFW/OAW)	BARE CAST OR TUBULAR ROD
SHIELDED METAL ARC (SMAW)	COATED SOLID OR TUBULAR ROD (STICK ELECTRODE)
GAS-TUNGSTEN ARC (GTAW)	BARE CAST OR TUBULAR ROD
GAS-METAL ARC (GMAW)	TUBULAR OR SOLID WIRE
FLUX-CORED OPEN ARC	TUBULAR WIRE (FLUX CORED)
SUBMERGED ARC (SAW)	TUBULAR OR SOLID WIRE
PLASMA TRANSFERRED ARC (PTA)	POWDER
LASER BEAM	POWDER

In general, the impact resistance of hardfacing alloys decreases as the carbide content increases. As a result, in a situation where a combination of impact and abrasion resistance is desired, a compromise between the two must be made. In applications where impact resistance is extremely important, austenitic manganese steels are used to build up worn parts.

Frequently, wear is accompanied by aqueous corrosion from acids or alkalis, such as those encountered in the chemical processing or petroleum industries or in flue gas scrubbers. Few of the iron-base hardfacing alloys possess the necessary corrosion resistance in such aqueous media. As a result, nickel- or cobalt-base hardfacing alloys generally are recommended when corrosion resistance combined with wear resistance is required. For example, a knife used to cut tomatoes in a food processing plant can last many times longer than a tool steel knife if the edge is made of a cobalt-base alloy.

Oxidation and hot corrosion resistance of iron-base alloys is also generally poor. Typically, boride-containing nickel-base alloys do not contain sufficient chromium in the matrix to resist oxidation. Hence, Laves phase-containing or carbide-containing nickel- or cobalt-base alloys typically are recommended for applications in which wear resistance combined with oxidation or hot corrosion resistance is required.

The ability of an alloy to retain strength at elevated temperatures is important for wear applications such as hot forging dies or valves for service at 870 °C (1600 °F), as well as service in coal gasification/liquefaction applications. Iron-base alloys with martensitic structures lose hardness at elevated temperatures. Generally, the high-temperature strength retention of a hardfacing alloy increases with its tungsten or molybdenum content. In applications requiring elevated-temperature strength and wear resistance, cobalt-base alloys or Laves phase alloys are recommended.

Wear-resistant materials typically are classified by type and alloy content, which largely determines their properties. These materials have structures consisting of hard phases (carbides, borides, or Laves phases) in a matrix hardened to a level that depends on the desired alloy properties.

Table 9 provides an extensive list of hardfacing materials along with their specific advantages and applications. Such recommendations are also available from producers of hardfacing alloys. The following steps should be taken in selecting a hardfacing alloy:

- ANALYSIS OF THE SERVICE CONDITIONS TO DETERMINE THE TYPE OF WEAR AND ENVIRONMENTAL RESISTANCE REQUIRED
- SELECTION OF SEVERAL HARDFACING ALLOY CANDIDATES
- ANALYSIS OF THE COMPATIBILITY OF THE HARDFACING ALLOYS WITH THE BASE METAL, TAKING INTO CONSIDERATION THERMAL STRESSES AND POSSIBLE CRACKING
- FIELD TESTING OF HARDFACED PARTS
- SELECTION OF AN OPTIMUM HARDFACING ALLOY, CONSIDERING COST AND WEAR LIFE
- SELECTION OF THE HARDFACING PROCESS FOR PRODUCTION OF WEAR COMPONENTS, CONSIDERING DEPOSITION RATES, THE AMOUNT OF DILUTION, DEPOSITION,

EFFICIENCY, AND OVERALL COST, INCLUDING THE COST OF CONSUMABLES AND PROCESSING

TABLE 9 SELECTION GUIDE FOR HARDFACING ALLOYS

TYPES OF MATERIALS	**ADVANTAGES**	**APPLICATIONS**
LOW HARDNESS		
PEARLITIC STEELS	LOW COST, CRACK-RESISTANT	A BUILDUP MATERIAL TO RESTORE DIMENSION OR AS A LOW-COST BASE FOR SURFACING APPLICATIONS
MILD STEELS	AN EXCELLENT BASE FOR HARDFACING	
LOW-ALLOY STEELS	SOME TOUGHNESS FOR BUILDUP OF WORN AREAS	
AUSTENITIC STEELS	EXCELLENT FOR HEAVY IMPACT, TOUGH	FOR METAL-TO-METAL WEAR UNDER HEAVY IMPACT CONDITIONS
LOW-CARBON NI-CR STAINLESS	GOOD CORROSION RESISTANCE	CORROSION-RESISTANT SURFACING OF LARGE TANKS, NUCLEAR VESSELS
HIGH-CARBON NI-CR STAINLESS	OXIDATION AND HOT WEAR RESISTANT	FURNACE PARTS, RED HEAT FRICTIONAL WEAR
14% MN, CR-NI	HIGH YIELD STRENGTH AUSTENITIC	BUILDUP, CRACK REPAIR, JOINING MANGANESE ALLOYS OR MANGANESE TO MILD STEEL
14% MN, 1% MO	FAIR ABRASION AND CORROSION RESISTANCE, WORK HARDENS	RAILWAY TRACK BUILDUP
14% MN, 3% NI	WORK HARDENS	BUILDUP AND JOINING OF MANGANESE PARTS
SEMIAUSTENITIC STEELS	LOW-COST CRACK-RESISTANT MATERIALS	HARDFACING APPLICATIONS
COPPER ALLOYS	RESISTANCE TO SEIZING UNDER FRICTIONAL WEAR CONDITIONS	FOR BUILDUP OF BEARING SURFACES
MEDIUM HARDNESS		
NICKEL-BASE ALLOYS	GOOD CORROSION RESISTANCE AND EXCELLENT HOT HARDNESS	FOR APPLICATIONS WHERE HOT HARDNESS IS IMPORTANT
NICKEL-CHROMIUM	RESISTANCE TO OXIDATION	FOR CORROSIVE WEAR CONDITIONS
NICKEL-CHROMIUM-MOLYBDENUM	RESISTANCE TO EXHAUST GAS CORROSION	FOR HARDFACING TRUCK, BUS, AND AIRCRAFT ENGINE VALVES
NICKEL-CHROMIUM-MOLYBDENUM-TUNGSTEN	CORROSION-RESISTANT PROPERTIES	FOR SURFACING APPLICATIONS
NICKEL-CHROMIUM-BORON	CORROSION-RESISTANT AS WELL AS ABRASIVE-	PUMPING SERVICE APPLICATIONS, ESPECIALLY IN THE OIL FIELDS

	RESISTANT PROPERTIES	
MARTENSITIC ALLOY IRONS	EXCELLENT ABRASION RESISTANCE	FOR OVERCOMING ABRASIVE WEAR WHERE LIGHT OR NO IMPACT IS INVOLVED
CHROMIUM-MOLYBDENUM	ABRASION RESISTANCE WHILE WITHSTANDING LIGHT IMPACT	FOR THE BUILDUP OF MACHINE PARTS SUBJECT TO REPETITIVE METAL-TO-METAL WEAR WITH OR WITHOUT LIGHT IMPACT
CHROMIUM-TUNGSTEN	HIGH COMPRESSIVE STRENGTHS	
AUSTENITIC ALLOY IRONS	BETTER CRACK-RESISTANT PROPERTIES THAN THE MARTENSITIC ALLOY IRONS	WHERE EROSIVE WEAR WITH OR WITHOUT LIGHT IMPACT IS PRESENT
CHROMIUM-MOLYBDENUM	GOOD CRACK-RESISTANT QUALITIES	PUMP AND TURBINE BUILDUP
NICKEL-CHROMIUM	CRACK-RESISTANT, LIGHT IMPACT	FOR EROSIVE WEAR WITH LIGHT IMPACT
MARTENSITIC STEELS	ABRASION RESISTANCE WITH MEDIUM IMPACT	FOR A VARIETY OF ABRASIVE WEAR CONDITIONS WHERE THERE MAY ALSO BE INVOLVEMENT WITH MEDIUM IMPACT
LOW CARBON (UP TO 0.30% C)	LEAST EXPENSIVE, TOUGH	
MEDIUM CARBON (0.30-0.65% C)	GOOD RESISTANCE TO MEDIUM IMPACT	
HIGH CARBON (0.65-1.7% C)	FAIR ABRASION RESISTANCE	
HIGH HARDNESS		
HIGH-CHROMIUM IRONS	EXCELLENT EROSION-RESISTANT PROPERTIES	FOR GENERAL USE WHERE HOT GASES OR MATERIALS ARE INVOLVED
AUSTENITIC	GOOD ABRASION-RESISTANT QUALITIES	FOR FARM AND EARTH-MOVING EQUIPMENT WORKING IN SANDY SOILS
MARTENSITIC	ABILITY TO BE REHARDENED AFTER ANNEALING	AS A HARDFACING IN STEEL MILLS AND REFINERIES WHERE HOT EROSION (595 °C, OR 1100 °F) IS A WEAR PROBLEM
TUNGSTEN-MOLYBDENUM ALLOYS	HIGH RED HARDNESS	FOR HARDFACING OF COKE OVEN PARTS AS WELL AS OTHER HOT (425 TO 650 °C, OR 800 TO 1200 °F) STEEL MILL APPLICATIONS
CHROMIUM-TUNGSTEN-COBALT ALLOYS	HIGH RED HARDNESS, HIGH HOT STRENGTH AND CREEP-RESISTANT QUALITIES	FOR MOST HARDFACING APPLICATIONS, BUT BEST SUITED TO THOSE APPLICATIONS WHERE HOT WEAR AND ABRASION (ABOVE 650 °C, OR 1200 °F) ARE INVOLVED (JET ENGINE TURBINE BLADES, GAS ENGINE EXHAUST VALVE)
LOW CARBON (UP TO 1.0% C)	GOOD OXIDATION RESISTANCE, TOUGH	
MEDIUM	RESISTANT TO ABRASION AND	

CARBON (1.4% C)	OXIDATION	
HIGH CARBON (2.5% C)	GOOD ABRASION RESISTANCE, BRITTLE	
TUNGSTEN CARBIDES	THE ULTIMATE IN ABRASION-RESISTANT QUALITIES	A VARIETY OF MATERIALS USED FOR HARDFACING TO MEET AN EXTREMELY WIDE RANGE OF SEVERE ABRASIVE CONDITIONS, ESPECIALLY OIL WELL DRILL BITS, TOOL JOINTS, ROCK DRILL BITS
FINE-GRAIN TUBULAR RODS	A RELATIVELY SMOOTH, EXTREMELY WEAR-RESISTANT SURFACE, GAS WELD	
COARSE GRANULAR TUBE ROD	DEVELOPS A ROUGH CUTTING SURFACE AS THE DEPOSITED BINDER WEARS AWAY	
GRANULES OR INSERTS	DEVELOPS AN EXTREMELY ROUGH ABRASIVE CUTTING SURFACE AS GRANULES OR INSERTS BECOME MORE EXPOSED. HIGHLY ABRASIVE, BUT BRITTLE UNDER IMPACT CONDITIONS	

Source: Ref 8

Hardfacing Process Selection (Ref 1)

Hardfacing process selection may be as important as hardfacing alloy selection, depending on the engineering application. Service performance requirements not only dictate hardfacing alloy selection, but also have a strong influence on hardfacing process selection. Other technical factors involved in hardfacing process selection include (but are not limited to) hardfacing property and quality requirements, physical characteristics of the workpiece, metallurgical properties of the base metal, form and composition of the hardfacing alloy, and welder skill. Ultimately, economic considerations predominate, and cost is the determining factor in the final process selection.

The term "hardfacing" is sometimes used to describe true welding processes, in which both the filler material and a small amount of the base metal are molten and are bonded metallurgically. Such processes, which include oxyfuel, arc, and laser welding methods, are preferred for applications that require dense and relatively thick coatings with excellent bonding between the hardfacing material and the base metal. Thermal spray processes are used for applications that require a thin, hard coating applied with minimum work distortion and where a purely mechanical bond between the hardfacing material and the base metal is acceptable.

Property and Quality Requirements. Welded hardfacing deposits are, in effect, mini-castings characterized by variable composition (segregation) and solidification kinetics that influence deposit microstructure. It is not surprising, therefore, that the properties and quality of welded hardfacing deposits should depend on welding process and technique, as well as on alloy selection.

Composition variations also derive from base-metal dilution during welding, although carbon pickup in oxyfuel gas welding (OFW) processes and solute volatilization during arc welding processes are additional factors. Dilution is the inter-alloying of the hardfacing alloy and the base metal and is usually expressed as the percentage of base metal in the hardfacing deposit. A dilution of 10% means that the deposit contains 10% base metal and 90% hardfacing alloy (see the section on "Weld Cladding" in this article for more detailed information). The wear resistance and other desirable properties of the hardfacing alloy are generally thought to degrade as dilution increases. The maximum amount of allowable dilution depends on specific service requirements. Welding process and technique should, however, be selected so as to control dilution to less than 20% for the hardest layer. More dilution is acceptable for buildup layers.

Dilution is nil in thermal spraying processes and tends to be acceptably low in conventional OFW processes. Dilution is generally more of a problem in arc weld hardfacing processes, ranging from approximately 5% in plasma arc welding

(PAW) hardfacing processes to as much as 50% in submerged arc welding (SAW). Laser hardfacing dilution is also low, ranging generally from 1 to 10%. High-dilution welding processes can be tolerated in applications requiring relatively thick hardfacings that can be applied in multiple layers. The effects of dilution on the microstructure of a high-chromium, white iron hardfacing alloy deposited in multiple layers on low-carbon steel substrate using shielded metal arc welding (SMAW) are illustrated in Fig. 8. The first layer (Fig. 8a) microstructure in this particular case was hypoeutectic with interdendritic eutectic and exhibited a hardness of 55 HRC. The second layer (Fig. 8b) consisted of a fine hypereutectic microstructure with a large percentage of eutectic and exhibited a hardness of 57 HRC. The third (Fig. 8c) and fifth (Fig. 8d) layers were microstructurally equivalent, consisting of primary M_7C_3 carbides in eutectic matrices. The hardness in the third and subsequent layers ranged from 60 to 61 HRC.

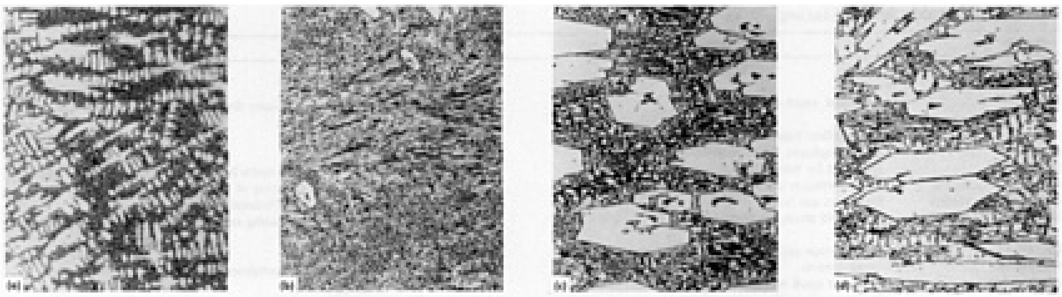

FIG. 8 EFFECT OF DILUTION ON THE MICROSTRUCTURE OF A FE-28CR-4MO-0.4MN-4.6C HARDFACING ALLOY. (A) FIRST LAYER OVER MILD STEEL SHOWING A HYPOEUTECTIC STRUCTURE OF METAL DENDRITES AND INTERDENDRITIC EUTECTIC, 55 HRC. (B) SECOND LAYER SHOWING A FINE HYPOEUTECTIC STRUCTURE WITH A LARGE PERCENTAGE OF EUTECTIC MATRIX, 57 HRC. (C) THIRD LAYER SHOWING A COARSE HYPEREUTECTIC STRUCTURE OF PRIMARY M_7C_3 CARBIDES IN A EUTECTIC MATRIX, 61 HRC. (D) FIFTH LAYER SHOWING A STRUCTURE SIMILAR TO BOTH THE THIRD AND FOURTH PASSES, 60 HRC. SOURCE: REF 1

The microstructures and mechanical properties of hardfacing deposits, like castings, vary depending on solidification kinetics as well as on dilution. Solidification generally is rapid in thermal spray processes, with nearly amorphous structures being achieved with some thermal spray processes. Solidification kinetics tend to be somewhat slower in conventional weld hardfacing processes. As a rule, hardfacing deposits produced by OFW processes tend to solidify more slowly than hardfacing deposits produced by arc welding processes. These differences in solidification rate produce widely different microstructures and widely different properties regardless of dilution. The effects of welding process on the microstructure of a cobalt-base (ERCoCr-A) hardfacing alloy deposited on a low-carbon steel substrate are illustrated in Fig. 9.

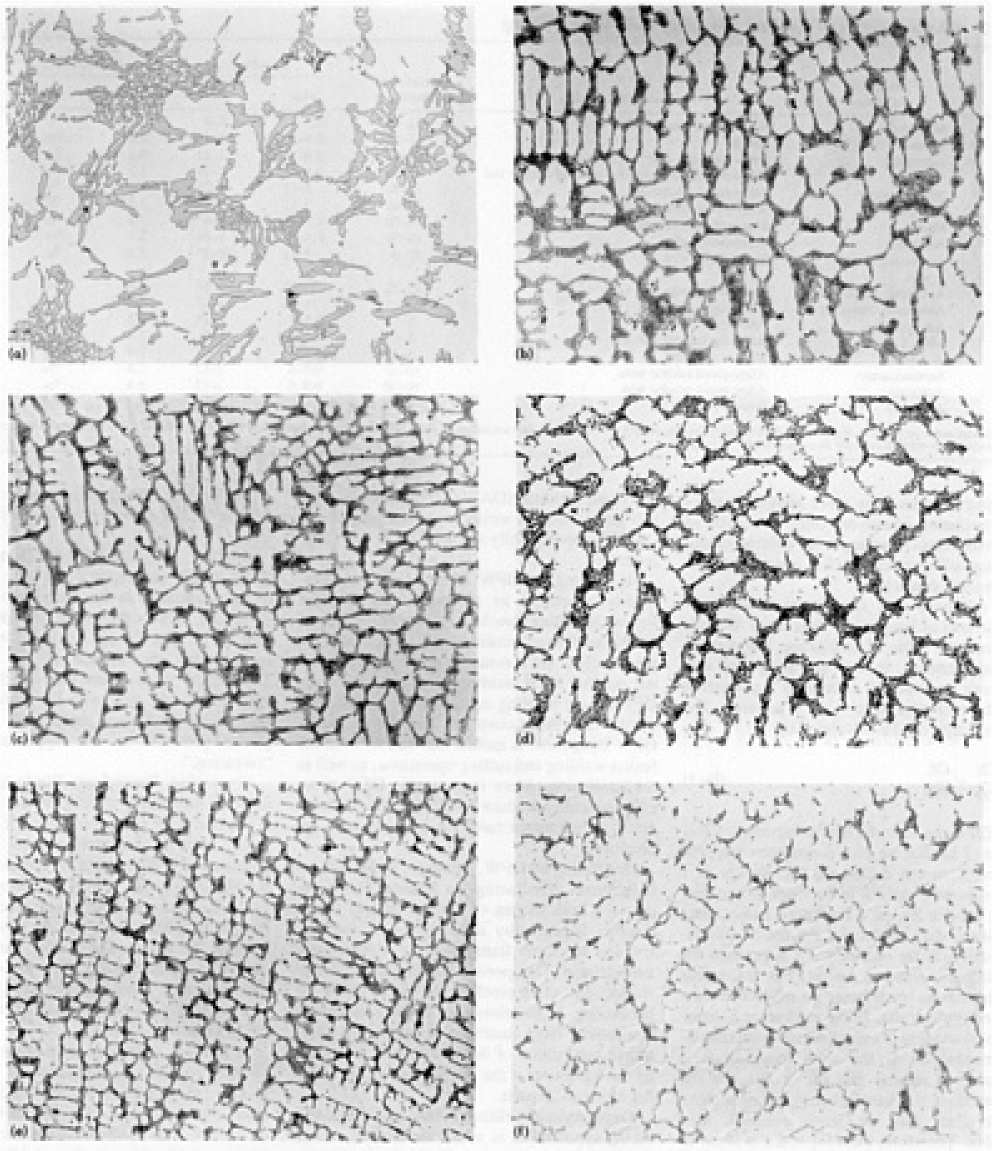

FIG. 9 EFFECT OF WELDING PROCESS ON THE MICROSTRUCTURE OF AN ERCOCR-A HARDFACING ALLOY. (A) OXYFUEL GAS WELDING. (B) PLASMA TRANSFERRED ARC WELDING. (C) GAS-TUNGSTEN ARC WELDING. (D) FLUX-CORED OPEN ARC WELDING. (E) SUBMERGED ARC WELDING. (F) SHIELDED METAL ARC WELDING. SOURCE: REF 1

Physical Characteristics of the Workpiece. The size, shape, and weight of the workpiece have a strong influence on hardfacing process selection. Large, heavy workpieces that are difficult to transport are most conveniently hardfaced using manual or semiautomatic hardfacing processes in which the hardfacing equipment can be moved to the workpiece. Gas shielded metal arc and open arc welding processes, for which portable equipment is readily available, are well suited to field applications involving inaccessible surfaces on large, difficult-to-transport workpieces. Thermal spraying processes and equipment also are available for *in situ* field hardfacing applications.

Hardfacing process selection is more involved in applications where the workpiece is sufficiently small so that it can be readily transported to the welding equipment. Thermal spraying processes and OFW, SAW, gas-tungsten arc welding (GTAW), PAW, and gas-tungsten arc welding (GMAW) are all well suited to in-plant hardfacing of small workpieces. Preference is usually given to thermal spraying and/or OFW or GTAW for small workpieces requiring thin, accurately placed hardfacing deposits. Large quantities of small workpieces can usually be hardfaced most economically using dedicated, specially designed, fully automatic hardfacing equipment and processes.

Metallurgical Characteristics of the Base Metal. As described below, base-metal surface preparation is important in thermal spray hardfacing processes. However, base-metal composition, melting temperature range, and thermal expansion and contraction characteristics have a significant effect on welding process selection.

Steels are generally suitable base metals for hardfacing. Low-alloy and medium-carbon steels with carbon contents up to 0.4% can be easily hardfaced by all welding processes with generally excellent results. Higher-carbon steels are similarly readily amenable to hardfacing, but preheating is necessary to minimize deleterious martensitic reactions in the heat-affected zone (HAZ). Austenitic stainless steels, with the exception of the free-machining grades, and most nickel-base alloys can also be easily hardfaced by all welding processes with excellent results. Martensitic stainless steels, tool steels, die steels, and cast irons are also amenable to hardfacing, but greater attention to preheat, interpass temperature, and postweld heat treatment is required. Age-hardenable base metals present special hardfacing problems and generally require solution and/or overaging heat treatments prior to hardfacing, as well as careful attention to preheat, interpass temperature, and postheat.

Hardfacing welding processes require that the base metal melting temperature range be higher than or at least equal to that of the hardfacing alloy. Thermal spraying processes, on the other hand, do not require a base metal melting temperature advantage.

Base metal thermal expansion and contraction characteristics affect solidification strain. Workpieces frequently have to be uniformly heated to relatively high temperatures during welding to enable the deposition of crack-free hardfacing alloys. Welding process selection is important in that it directly affects heat input and preheat requirements.

Thermal expansion and contraction differences between the hardfacing alloy and base metal are important in applications involving thermal cyclic service conditions. Large differences can result in shear failures in thermal spray deposits or thermal fatigue failures in welded deposits. Buffer layers are often deposited between the base metal and hardfacing alloy to counteract large differences in thermal expansion and contraction characteristics. Frequently, dilution, which is generally regarded to be detrimental, can be used to good advantage to form ***in situ*** buffer layers and to minimize the risk of thermal fatigue failure. Hardfacing process selection is an important consideration in this regard.

Hardfacing Product Forms. Hardfacing process selection is often limited by product form. Hardfacing alloys normally are available as bare cast rod, covered electrodes, wire, and powders. However, not all alloys are available in all product forms.

Virtually all hardfacing alloys can be produced in powder form, and most hardfacing alloys can be produced in rod form. Rods generally are available in bare cast form or in tubular form. Bare cast rod diameters are limited to a large extent by alloy characteristics and casting processes to a range of 3.2 to 8.0 mm ($\frac{1}{8}$ to $\frac{5}{16}$ in.). Bare cast rod lengths usually range from 250 to 700 mm (10 to 28 in.), but some hardfacing alloys can be cast in extra long, continuous lengths with diameters as small as 3.2 mm ($\frac{1}{8}$ in.). Tubular rods consist of highly alloyed powder cores encased in low-alloy sheaths. Tubular rods can be produced in discrete lengths or in spooled continuous coils. Discrete-length tube rods generally are available in the same diameter range as bare cast rod. Long-length tube rods, or tube wire as they are properly named, normally are available in diameters ranging from 4.0 to 1.6 mm ($\frac{5}{32}$ to $\frac{1}{16}$ in.). Some of the tougher, more ductile hardfacing alloys can be produced as solid drawn wire. Some of the more brittle hardfacing alloys can be produced, at a premium cost, as solid wire using extrusion techniques.

Welder Skill. It is essential to consider hardfacing quality requirements in relation to welder/operator skill, as well as in relation to the hardfacing process. As a rule, hardfacing by manual welding processes such as OFW or GTAW requires high welder proficiency, whereas hardfacing by automatic welding processes, such as SAW, requires a minimum of operator skill, assuming machine control settings have previously been established. Thermal spraying and related powder welding processes generally require intermediate welder/operator skills.

Manual GTAW can be used to obtain high-quality deposits on relatively small areas, such as the interlock regions of turbine blades. In this type of application, thin layers can be deposited with dilution as low as 10%, but relatively high welder skill and close control of the welding operation are necessary. In contrast, earthmoving and mining equipment can be hardfaced adequately in the field by relatively unskilled welders. Selection of process is usually based on maximum deposition rate: high dilution rates seldom significantly affect the suitability of the coating for service.

Cost is ultimately the determining factor in hardfacing process selection. Hardfacing costs basically consist of labor, materials, and in some instances, new equipment cost. Transportation cost also must be considered.

Labor costs depend primarily on the level of welder skill required and process deposition rates, but surface preparation and finishing costs also must be considered. Deposition rates are normally lowest with manual hardfacing processes. Considerably higher deposition rates are possible with automatic arc welding hardfacing processes. Deposition rates as high as 27 kg/h (60 lb/h) are possible with multiwire or powder and wire submerged arc hardfacing processes, but dilution is also high. An ideal welding process for hardfacing would be one with the ability to limit dilution while achieving a high deposition rate. Typical dilution percentages; deposition rates; and minimum deposit thicknesses for different welding processes and various forms, compositions, and modes of application of hardfacing alloys are given in Table 10.

TABLE 10 CHARACTERISTICS OF WELDING PROCESSES USED IN HARDFACING

WELDING PROCESS	MODE OF APPLICATION	FORM OF HARDFACING ALLOY	WELD-METAL DILUTION, %	DEPOSITION		MINIMUM THICKNESS$^{(A)}$		DEPOSIT EFFICIENCY, %
				kg/h	lb/h	mm	in.	
OAW	MANUAL	BARE CAST ROD, TUBULAR ROD	1-10	0.5-2	1-4	0.8	MATH OMITTED	100
	MANUAL	POWDER	1-10	0.5-2	1-4	0.8	$\frac{1}{32}$	85-95
	AUTOMATIC	EXTRA-LONG BARE CAST ROD, TUBULAR WIRE	1-10	0.5-7	1-15	0.8	$\frac{1}{32}$	100
SMAW	MANUAL	FLUX-COVERED CAST ROD, FLUX-COVERED TUBULAR ROD	10-20	0.5-5	1-12	3.2	$\frac{1}{8}$	65
OPEN ARC	SEMIAUTOMATIC	ALLOY-CORED TUBULAR WIRE	15-40	2-11	5-25	3.2	$\frac{1}{8}$	80-85
	AUTOMATIC	ALLOY-CORED TUBULAR WIRE	15-40	2-11	5-25	3.2	$\frac{1}{8}$	80-85
GTAW	MANUAL	BARE CAST ROD, TUBULAR ROD	10-20	0.5-3	1-6	2.4	$\frac{3}{32}$	98-100
	AUTOMATIC	VARIOUS FORMS$^{(B)}$	10-20	0.5-5	1-10	2.4	$\frac{3}{32}$	98-100
SAW	AUTOMATIC, SINGLE WIRE	BARE TUBULAR WIRE	30-60	5-11	10-25	3.2	$\frac{1}{8}$	95
	AUTOMATIC, MULTIWIRE	BARE TUBULAR	15-25	11-27	25-60	4.8	$\frac{3}{16}$	95

		WIRE						
	AUTOMATIC, SERIES ARC	BARE TUBULAR WIRE	10-25	11-16	25-35	4.8	$\frac{3}{16}$	95
PAW	AUTOMATIC	POWDER$^{(C)}$	5-15	0.5-7	1-15	0.8	$\frac{1}{32}$	85-95
	MANUAL	BARE CAST ROD, TUBULAR ROD	5-15	0.5-4	1-8	2.4	$\frac{3}{32}$	98-100
	AUTOMATIC	VARIOUS FORMS$^{(B)}$	5-15	0.5-4	1-8	2.4	$\frac{3}{32}$	98-100
GMAW	SEMIAUTOMATIC	ALLOY-CORED TUBULAR WIRE	10-40	0.9-5	2-12	1.6	$\frac{1}{16}$	90-95
	AUTOMATIC	ALLOY-CORED TUBULAR WIRE	10-40	0.9-5	2-12	1.6	$\frac{1}{16}$	90-95
LASER	AUTOMATIC	POWDER	1-10	$^{(D)}$	$^{(D)}$	0.13	0.005	85-95

(A) RECOMMENDED MINIMUM THICKNESS OF DEPOSIT.

(B) BARE TUBULAR WIRE; EXTRA-LONG (2.4 M, OR 8 FT) BARE CAST ROD; TUNGSTEN CARBIDE POWDER WITH CAST ROD OR BARE TUBULAR WIRE.

(C) WITH OR WITHOUT TUNGSTEN CARBIDE GRANULES.

(D) VARIES WIDELY DEPENDING ON POWDER FEED RATE AND LASER INPUT POWER

Hardfacing material costs depend on the prevailing price of raw materials and product form. Raw material cost is predominant in tungsten-, cobalt-, and nickel-base hardfacing alloys, whereas product form is the predominant cost in iron-base hardfacing alloys. As a rule, tubular rod and wire are the least expensive hardfacing product forms. In contrast, solid wire is the most expensive hardfacing product form. Cast rod and powder are generally intermediate in cost. Powder, depending on specific powder size requirements, tends to be less expensive than cast rod. However, deposit efficiency, measured as the percentage of hardfacing consumable retained on the workpiece surface, is generally slightly lower for hardfacing processes that utilize powder consumables than for hardfacing processes that utilize rod consumables. Thermal spraying processes are particularly poor in this regard, with typical deposit efficiencies as low as 70%.

Equipment costs vary from simple arc welding equipment to fully automated laser beam hardfacing systems. The more advanced automatic hardfacing equipment often costs 100 to 1000 times as much as simple arc welding equipment, depending on the level of instrumentation and computer control.

It is relatively easy to estimate or predetermine the cost of hardfacing. It is more difficult to estimate the value of hardfacing. For maintenance applications, it has been suggested that hardfacing is preferred over total replacement when the cost advantage (CA) is positive. The cost advantage can be calculated by:

$$CA=\frac{CN}{PN}-\frac{CR}{PR} \qquad \textbf{(EQ 1)}$$

where CN is the cost of a new component; CR is the cost of hardfacing plus downtime costs; PN is the work output during the service life of a new component; and PR is the work output during the service life of a hardfaced component. Most of the data needed for the above calculation usually can be obtained; PR is perhaps the most difficult value to assess, but reasonable estimates can be made based on relative laboratory wear test results. If the workpiece is hardfaced with an alloy of essentially the same chemical composition as the wom material, it is reasonable to assume that $PR = PN$. On the other hand, if the hardfacing material is more wear-resistant than the worn material, then $PR > PN$. Therefore, from Eq 1, CA is likely to be positive, which is the primary objective of hardfacing.

Hardfacing by Oxyfuel Gas Welding

Oxyfuel gas welding (OFW) is a process in which the heat from a gas flame is used to melt hardfacing materials onto workpiece surfaces. During hardfacing, the workpiece is only superficially melted. The gas flame is created by the combustion of a hydrocarbon gas and oxygen. Acetylene is the most widely used fuel for OFW. In fact, OFW is commonly referred to as oxyacetylene welding (OAW), even though other fuel gases, such as methylacetylene propadiene (MAPP), occasionally are used in certain OFW processes.

Hardfacing by OFW is a versatile process, readily adaptable to all hardfacing product forms, including powder. As described below, OFW hardfacing processes that use fillers in powder form are commonly referred to as manual powder torch welding or powder welding. Oxyfuel gas welding equipment is simple, and capital equipment costs are low. Furthermore, OFW equipment is mobile and can be used for fusion welding and cutting operations, as well as for hardfacing. More importantly, OFW processes generally produce hardfacing deposits with optimum microstructures for wear-resistant applications.

Hardfacing by OFW, however, is not without limitations. Hardfacing by manual OAW requires a high degree of welder skill to obtain smooth, high-quality deposits, as the welding rod and the torch flame have to be separately manipulated. The speed of welding and low deposition rates often preclude the use of OAW for hardfacing applications on large components that require large quantities of hardfacing alloy, unless automation of the operation is warranted by configuration of the surface and a large number of identical parts.

Oxyacetylene welding is best suited to hardfacing applications in which the surface area to be hardfaced is minimal. Thus, for hardfacing steam valves, automotive and diesel engine valves, cutter blades for wood and plastic, chain saw bars, and plowshares and other agricultural implements, it is often the most satisfactory process as it can be easily employed in field applications.

Most steel workpieces can be hardfaced by OFW processes. High-manganese and high-sulfur steels are exceptions. High-speed steel workpieces are difficult to hardface and should be fully annealed before hardfacing. Cast iron workpieces also can be hardfaced by OFW processes, but cast iron substrates generally require the use of fluxes and special welding precautions.

Dilution is generally low (1 to 10%) in OFW hardfacing deposits, as only the surface of the workpiece in the immediate area being hardfaced is brought to the melting temperature. The hardfacing rod is melted and spread over the surface of the workpiece. The melted rod does not mix or alloy with the base metal, but only with the melted surface. Superficial workpiece melting of steel workpieces produces a wet appearance, called "sweating."

Operating Procedures. Workpiece surfaces must be thoroughly cleaned before hardfacing. If impurities are present on surfaces of the workpiece, or if an excessive amount of oxide is produced in heating (usually, a neutral or reducing flame is used), it may be necessary to dislodge the solid contaminants by prodding or rubbing with a hardfacing rod tip. Oxide trapped beneath the hardfacing overlay can react with carbon in the hardfacing alloy to produce a gas and thus cause porosity. Hardfacing alloys usually contain a deoxidizer that controls a moderate amount of oxidation.

It is often necessary to preheat workpieces prior to hardfacing, regardless of the welding process, to prevent cracks in the hardfacing deposit. Preheating not only reduces tensile strain associated with thermal contraction differences between the hardfacing alloy and the workpiece, but in the case of OFW, also permits the use of a softer flame and results in less dilution. Small low-carbon steel workpieces can be preheated locally with a torch, but large workpieces are best preheated with gas burners in a brick muffle furnace, allowing sufficient soak time to ensure even heating.

With practice, one-layer hardfacing deposits up to 1.6 mm ($\frac{1}{16}$ in.) thick can be deposited, particularly if the workpiece is inclined so that the deposition is slightly uphill. Alternatively, thick deposits can be applied in several layers. Hardfacing deposits by OFW can be remelted as necessary to improve deposit quality or surface appearance.

Common Flaws. Hardfacing deposits applied by OFW generally are characterized by low dilution and high deposit quality. Discontinuities are possible, however, depending on the skill of the welder and the quality of the filler material.

Porosity is probably the most common flaw observed in hardfacing deposits applied by OFW. Porosity in hardfacing deposits, like porosity in castings, can result from inadequate feeding or gas evolution during solidification. Porosity resulting from inadequate feeding is sometimes referred to as microshrinkage. Microshrinkage is most common at the termination of the hardfacing deposit and usually results from the inability of the hardfacing deposit to accommodate solidification shrinkage after the flame is withdrawn. This type of flaw can be minimized or prevented by withdrawing the flame very slowly and allowing solidification to progressively occur from the bond line to the surface.

Porosity can also occur by gas evolution during solidification if the total gas content in the molten deposit is greater than the solid solubility limit. Gas-induced porosity generally derives from slag or non-metallic inclusions trapped in or on the surface of the base metal. However, the initial gas content of the filler materials and possible gas pickup during OFW are also contributing factors.

Manual Powder Torch Welding (Ref 9)

Also termed "powder welding," manual powder torch welding is a modified OFW process in which powder application and fusion occur in a single operation using a special oxyacetylene torch. Powder is fed from a small hopper mounted on a gas welding torch into the fuel gas supply and conveyed through the flame to the workpiece surface. By following general oxyacetylene surfacing practice, smooth, thin, dense deposits are obtained. Deposit thickness is controlled by powder flow rate and movement of the torch. A dilution of 1 to 5% in the deposit occurs from this process.

Deposits ranging from 0.75 to 3 mm (0.030 to 0.125 in.) in thickness can be applied at a rate of 1.3 mm (0.050 in.) per pass. Sweating of the base metal during surfacing generates a fusion-type bond with the hardfacing alloy. Ease of application and material recovery are enhanced because of close control of the powder flow into the welding flame.

Oxyacetylene torch flame adjustment is essential to produce good powder deposition. The amount of excess acetylene required for hardfacing varies depending on alloy composition. The amount of excess acetylene is conveniently measured in terms of the flame geometry as the ratio between the length of the oxyacetylene feather and the length of the inner cone. The length of the inner cone, indicated by the value X, is the distance from the torch tip to the extreme end of the inner cone.

The length of the oxyacetylene feather is measured from the torch tip to the extreme end of the intermediate zone, generally expressed in multiples of the length of the inner cone. A neutral flame is thus classified as 1×. Cobalt-base powders are usually applied using 2× to 3× flame. The oxyacetylene feather is two to three times the length of the inner cone.

Tungsten carbide-containing composite powders should be applied using a 1.5× flame. Nickel-base powders are welded with a neutral or slightly oxidizing flame. Because flame characteristics change somewhat when powder is being sprayed, the powder feed handle should be depressed and final flame adjustment made while the powder is being sprayed. Alloys commonly welded with this process contain boron and silicon, which combine with any oxide skin to form an easily removed slag and to provide a clean weld pool.

The manual torch process is well suited for small repairs or for hardfacing of small areas requiring a wide range of hardnesses (about 20 to 60 HRC). Typical applications of manual torch welding include glass dies and molds, brick molding equipment, components for the cement industry, cast iron parts for foundries, feed screws, hammer mill hammers, and pump parts.

Hardfacing by Arc Welding (Ref 1)

In hardfacing by arc welding, the heat from an electric arc is used to melt the hardfacing material onto the workpiece surface. The electric arc is developed by impressing a voltage between an electrode and the workpiece. The voltage required to sustain the arc varies with the distance between the electrode and workpiece and with the arc welding process. The filler material is provided by the electrode used to form the arc or can be externally introduced into the arc. Arc welding processes in which the filler derives directly from the electrode are sometimes referred to as consumable electrode processes. Arc welding processes in which the filler is externally introduced into the arc are sometimes referred to as nonconsumable electrode processes. As a rule, nonconsumable electrode processes are accomplished at lower power requirements than are consumable electrode processes, resulting in less dilution and lower deposition rates. Consumable and nonconsumable electrode hardfacing processes both require that the filler materials be protected from oxidation as they are melted and joined to the workpiece surfaces.

Filler materials in consumable electrode hardfacing processes are sometimes protected by fluxes as in SMAW, SAW, and flux-cored open arc welding, or by inert gases, as in GMAW. The selection of a particular consumable or nonconsumable electrode hardfacing process depends on consideration of the factors discussed in the section on "Hardfacing Process Selection" in this article.

Shielded Metal Arc Welding. The electrodes used in the SMAW consumable electrode hardfacing process generally consist of discrete lengths of filler material rod covered with specially formulated coatings. The filler material rod generally ranges in diameter from 3.2 to 8.0 mm ($\frac{1}{8}$ to $\frac{5}{16}$ in.) and may be cast or of tubular construction. The electrode coating protects the filler material during welding by chemical reaction as the electrode melts. The coating can be fortified to increase alloy content during melting, as well as to simply provide molten metal protection. Coating ingredients also determine or affect the shape of the weld, amount of penetration, cooling rate of the weld, stabilization of the arc, and refinement of the hardfacing deposit.

The selection of the proper power supply for SMAW consumable electrode hardfacing is not a serious problem. Motor-generators, direct current rectifiers, and alternating current transformers can be used. Direct current electrode positive (DCEP) normally is recommended for hardfacing. Direct current electrode negative (DCEN) can be used to increase deposition rates slightly, but there is a tendency for the weld beads to be high and narrow. Although direct current is recommended, alternating current has been used in many applications with reasonably good success. Arc stability and burnoff characteristics are not quite as good by this method.

Speed of travel, position of the electrode in relation to the weld pool, arc voltage, and amperage are modified to minimize penetration and the consequent dilution of the deposit. Typical welding currents for small 3.2 mm ($\frac{1}{8}$ in.) diam hardfacing electrodes range from 60 to 130 A, depending on the alloy. Welding currents for larger 6.4 mm ($\frac{1}{4}$ in.) diam hardfacing electrodes range from 150 to 330 A.

Depending on the surface area to be hardfaced, the bead-deposit pattern may be of the stringer or weave type, with or without staggering of beads, as dictated by service requirements. Selection of the bead pattern and sequence also depends on the accessibility of the surface to be hardfaced and the distortion limits of the workpiece.

Deposition rates vary considerably with welding conditions. Normally from 1.4 to 1.8 kg (3 to 4 lb) of metal per hour of arc time can be applied. The recovery of alloy in the deposit generally ranges from 65 to 75% of the electrode weight, excluding stub loss. The average stub loss is approximately 75 mm (3 in.), which corresponds to about 20% of a 350 mm (14 in.) long electrode.

There are many advantages to the SMAW consumable electrode process. The intense heat of the arc permits the hardfacing of large parts without preheat. Small areas of parts machined or fabricated to close tolerances can be hardfaced without distorting the entire part. Difficult-to-reach areas are best accessed with coated electrodes. The overall process is relatively fast, as preheat and high interpass temperatures do not have to be maintained. Another attractive feature of this process is the portability of the equipment. Gasoline engine-driven generators can be readily taken into the field.

There are also some disadvantages to the SMAW consumable electrode process that must be considered. Penetration tends to be high, and dilution rates of 20% or more are not uncommon. This results in lower hardness and less resistance to abrasion. Two or three layers of weld metal normally are required to obtain maximum wear properties. Coated electrodes also leave a slag covering on the surface of the weld deposit which must be removed prior to each succeeding pass. Because coated electrodes usually are deposited with no more than a small amount of preheat, cracking or cross checking may occur, especially with the higher alloy electrodes that produce hard overlays.

The hardfacing material is at its maximum temperature, and hence its maximum thermal expansion, as it is being deposited. The workpiece, by contrast, may be near ambient temperature during the welding process and may be considerably less expanded than the hardfacing material. Severe strains develop in the hardfacing deposit as it cools after welding. These strains are accommodated by plastic deformation, in the case of a ductile hardfacing material or by cracking and cross checking, in the case of more brittle hardfacing materials. Cracking or cross checking does not affect the wear resistance of the overlay in many applications.

Submerged arc welding is a consumable electrode hardfacing process in which the area between the electrode and workpiece is shielded by a blanket of granular, fusible flux material. Electrodes in the SAW hardfacing process consist of continuous lengths, or coils, of filler materials--generally in tubular wire form.

The fluxes used for hardfacing by SAW are mineral compositions specially formulated to protect the molten filler metal without appreciable gas evolution. Submerged arc fluxes are generally finely divided, free-flowing granules in the solid state, which are placed by gravity along the surface to be hardfaced in advance of the electrode. The flux becomes molten and highly conductive under the arc, creating localized conditions conducive to high deposition rates.

Hardfacing filler materials can be deposited over a wide range of welding currents, voltages, and travel speeds, each of which can be controlled independently. Generally, however, the current, voltage, and workpiece travel speed are adjusted for hardfacing applications to produce a bead approximately 15 mm ($\frac{5}{8}$ in.) wide with a slightly convex surface and sides tapering into the base metal. Welding current usually ranges from 300 to 500 A for most SAW hardfacing applications. However, it is not uncommon to increase current to as much as 900 A to increase deposition rate.

A number of features make SAW particularly attractive for hardfacing. Because it is an automatic process, SAW requires little skill on the part of the operator for efficient operation. High deposition rates are possible, usually ranging as high as 4.5 to 9.0 kg/h (10 to 20 lb/h), depending on the job. The process is capable of making porosity-free deposits that are smooth and even on the surface and that either can be used aswelded or require only a minimum of grinding to clean up. The filler metal is protected from the atmosphere by flux, which also completely surrounds the arc, requiring no shielding for the operator.

The major disadvantages of hardfacing by SAW are that: (1) it is generally limited to hardfacing simple cylindrical or flat workpieces; and (2) it is not particularly adaptable to the surfacing of small parts. Hardfacing in other than the flat position requires special flux retainers. Preheating the workpiece in excess of 315 °C (600 °F) usually makes slag removal extremely difficult unless the slag is designed for easy hot slag removal. Occasionally, small portable units are used, but normally SAW units are fairly large and are mounted on stationary tracks, thus limiting their use for field operations. It is imperative that the flux be kept dry prior to use; therefore, storage can be a problem. Dilution is relatively high. It is not uncommon to find interalloying with the base metal, reaching or exceeding levels of 30%. Consequently, two or three layers of weld metal may be required for maximum wear resistance. Dilution can be reduced by long stickout techniques and by heavy overlap of individual weld beads.

Flux-cored open arc welding is a semiautomatic or automatic process in which the arc between the electrode and workpiece is shielded by a self-generated or self-contained gas or flux. Flux-cored open arc electrodes are continuous lengths, or coils, of filler material in tubular wire form. Open arc tubular filler materials are generally of low-carbon steel sheath construction with alloying elements, deoxidants, arc stabilizers, and a shielding gas source and/or flux contained in the core as powder or granules. High-alloy filler materials are generally shielded by a self-generated CO_2 gas deriving from the reaction between carbonaceous materials contained in the core and the atmosphere, or by vaporization of fluorides in the corn. Arc shielding is often supplemented in low-alloy filler materials with flux and/or aluminum or magnesium additions, which serve as deoxidizers.

Low equipment cost is one of the primary advantages of the flux-cored open arc hardfacing process. Neither flux-handling nor gas-regulating equipment is required. Thus the flux-cored open arc hardfacing process is a simple semiautomatic welding process requiring only a torch and a device to feed the continuous electrode wire. Conventional power, supplied with drooping voltage-amperage characteristics, is well adapted to flux-cored open arc hardfacing equipment. Motor-generator units are frequently used for field applications because of their portability. In-plant operations normally make use of direct current rectifiers. Although flux-cored open arc hardfacing is a semiautomatic process, it is not necessary to have a constant-potential power supply for its operation.

Flux-cored open arc equipment generally is designed for small-diameter welding wires, usually 1.2 mm ($\frac{3}{64}$ in.), 1.6 mm ($\frac{1}{16}$ in.), or 2.8 mm ($\frac{7}{64}$ in.). Most ferrous metals can be hardfaced with ease. Cast iron and chilled iron should be handled cautiously. Open arc is a high-energy process capable of much higher deposition rates than SMAW. The 2.8 mm ($\frac{7}{64}$ in.) diam rod can be deposited with welding currents as high as 500 A. The flux-cored open arc hardfacing process is closely related to the submerged arc process, except that it does not require a flux burden over the bead; consequently, them is little or no slag to be removed from the overlay. Most hardfacing alloys can be deposited by the flux-cored open arc process without preheat, and there is good recovery of elements across the arc. Hardfacing by fluxcored open arc welding is an easy process to use, and only a short time is required to train an operator for the average application.

Arc stability and melting rates are excellent. The best hardfacing characteristics are obtained from DCEP. However, good results can be obtained with an alternating current supply at 26 to 30 V. Hardfacing deposits are comparable in soundness to deposits made with covered electrodes, at deposition rates three to five times faster than with covered electrodes.

Hardfacing by flux-cored open arc welding is not without disadvantages. Because the arc is not shielded with inert gas or a blanket of granular flux, considerable spatter and some porosity can be expected with this process. In addition, because of high welding currents, it is not particularly well adapted to the hardfacing of small parts. The recovery of weld metal from the tube wire to the deposit normally averages near 80 to 85% or higher. The amount of cross cracking and distortion is comparable to that of other arc welding processes.

Gas-metal arc welding is a consumable electrode hardfacing process in which the filler material and workpiece surfaces are protected by a flowing shielding gas--carbon dioxide, argon, or helium, either singly or in combination with a small amount of oxygen. The GMAW process is suitable for semiautomatic hardfacing, and the versatility of the process lends itself to the hardfacing of complex shapes. Hardfacing using GMAW can be fully mechanized. The hardfacing deposit is visible at all times in GMAW, thereby enabling high-quality deposits. Hardfacing by GMAW utilizes small-diameter (generally 1.6 mm, or $\frac{1}{16}$ in., or less) hardfacing filler wires or electrodes. Gas-metal arc hardfacing wires can be deposited in either spray arc or short arc modes. The spray arc mode produces a continuous stream of droplets about the same diameter as the hardfacing wire, which are projected by electromagnetic force from the tip of the wire. Deposition rates are high, as is dilution. In the short arc mode, lower voltages are used, and transfer of the molten hardfacing wire is more globular, usually resulting in more spatter.

Hardfacing by GMAW, unlike SMAW, requires a constant-potential power source with voltage, slope, and wire-feed rate controls. Best results for 1.6 mm ($\frac{1}{16}$ in.) diam cobalt-base hardfacing wires in the spray arc mode are generally obtained at 18 to 26 V, with wire-feed rates ranging from 45 to 65 mm/s (110 to 150 in./min.). Best results in the short arc mode are generally obtained at 16 to 18 V, with wire-feed rates ranging between 35 to 40 mm/s (80 to 100 in./min.). The use of auxiliary shielding gases adds to the cost of hardfacing by GMAW. However, the higher cost is generally offset by the higher deposit quality associated with GMAW.

Gas tungsten arc welding is a nonconsumable electrode arc process in which the heated area of the workpiece, molten hardfacing alloy, and the nonconsumable electrode are shielded from the atmosphere by a flow of shielding gas fed through the torch. Thoriated tungsten is the preferred nonconsumable electrode material. Conventional power supplies with drooping voltage-amperage characteristics are normally used for hardfacing. Motor-driven generators can be used, as can alternating current transformers with continuous high-frequency current. Direct current electrode negative is always used to minimize tungsten contamination in the hardfacing deposit. Argon generally is used as the shielding gas, but helium is also suitable.

Hardfacing by manual GTAW is a useful alternative to hardfacing by OAW techniques, particularly when large components or reactive base metals are involved. Gas-tungsten arc welding is preferred for hardfacing on reactive base metals such as titanium-stabilized stainless steels or aluminum-bearing nickel-base alloys. Gas-tungsten arc welding also is preferred over OAW for hardfacing applications in which carbon pickup is unacceptable or in applications involving filler materials that tend to boil under the influence of an oxyacetylene flame. The localized, intense heat of GTAW generally results in more base metal dilution than OAW. Base metal dilution can be minimized by torch oscillation, by using no more amperage than is necessary, and/or by concentrating the arc on the hardfacing deposit rather than on the workpiece. Hardfacing deposition rates generally are comparable in manual GTAW and OAW, despite the difference in base metal dilution.

Hardfacing by GTAW can be accomplished automatically by simply attaching the torch to an oscillating mechanism and using a mechanized device to feed the hardfacing filler material into the arc region. This, of course, requires long-length filler rod or continuous wire. High-quality, reproducible hardfacing deposits can be produced by automatic GTAW by controlling filler-metal feed rates, torch oscillation, and travel speed. Current-delay controls can be used to control the final stages of solidification in the hardfacing deposit to minimize shrinkage and crater cracking.

The arc action in GTAW is generally smoother, quieter, and freer of weld spatter than in other consumable electrode arc welding processes. The use of a separate filler rod or wire enables greater operator control of deposit shape. Gas-tungsten arc welding is more adaptable to hardfacing small intricate parts and generally produces higher-quality deposits than other hardfacing processes.

Plasma Arc Welding. Hardfacing by PAW is similar to hardfacing by GTAW in that both processes employ a gas-shielded arc between a nonconsumable tungsten electrode and the workpiece as the primary heat source for hardfacing. The processes differ in that the PAW process maximizes the use of plasma as a secondary heat source. The plasma is formed by ionizing the gas flowing in a nozzle surrounding the electrode. The electrode generally is recessed into the nozzle. The plasma gas generally emerges from a constricting orifice arrangement. The PAW process can use bare rod or wire as a hardfacing consumable, but more often powder is used as the consumable. When powder is used, the process is often referred to as the plasma transferred arc (PTA) process.

Plasma transferred arc surfacing, as shown schematically in Fig. 10, is a welding process in which the powder is introduced into a combined arc/plasma stream to form a molten pool on the workpiece. The resulting deposit is homogeneous and dense, with excellent metallurgical bonding with the base metal (workpiece). This process uses a constricted arc. The alloy is carried from a powder feeder to the plasma torch in a stream of argon gas, which forms the plasma and is directed away from the torch into the arc effluent, where it is melted and fusion-bonded to the base metal. A direct current power source connecting the tungsten electrode and the workpiece provides energy for the transferred arc.

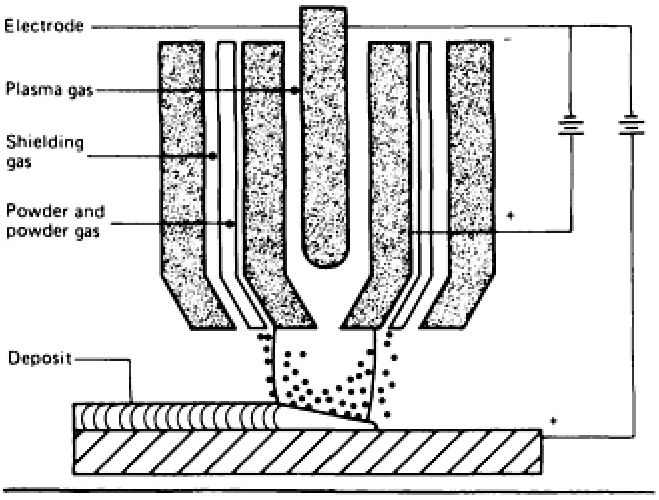

FIG. 10 SCHEMATIC OF THE PLASMA TRANSFERRED ARC HARDFACING PROCESS. SOURCE: REF 1

A second direct current power source connecting the tungsten electrode and the arc-constricting orifice supports a nontransferred arc. This nontransferred arc supplements the heat of the transferred arc and serves as a pilot arc to initiate the transferred arc. Argon is passed through a gas diffuser to provide a blanket of shielding gas in and around the arc zone.

Hardfacing with the PTA process has many advantages. Deposits ranging from 0.64 to 3.2 mm (0.025 to 0.125 in.) thick can be produced rapidly in a single pass. A dilution of 5 to 20% in the deposit is typical. Powder recoveries as high as 95%, with deposition rates up to 4.5 kg/h (10 lb/h), are possible, depending on the size and shape of the part being hardfaced. Still higher deposition rates ranging up to 18 kg/h (40 lb/h) are possible using the plasma arc hot wire process, which, as its name implies, uses wire hardfacing consumables and two independently controlled systems operating in one weld pool to melt and fuse a hardfacing alloy to a substrate. This process, however, is considered a weld cladding operation and is described later in this article.

Hardfacing by the PTA process has some disadvantages. The equipment is relatively expensive and is limited to straight-line or cylindrical parts, except when special tooling is provided. Usually, several parts are needed for setup, which makes the process somewhat unattractive to the low-volume job shop. The torch is mounted on a carriage and track that require permanent installation. There are two circuits employed in this process; therefore, either a special power supply or two separate conventional power units are required. Argon consumption is somewhat higher than with GTAW. Argon has to be supplied to the center of the torch to protect the electrode. Shielding gas is necessary to protect the weld metal, and argon is also used as a carrier for the powder. Large parts being faced with hard alloys normally require preheat. This is a serious potential problem with the PAW process, if excessive preheats and prolonged hardfacing times cause overheating of the torch.

Laser Hardfacing (Ref 10)

Laser hardfacing, also commonly referred to as laser cladding, differs little in principle from traditional forms of hardfacing; the primary difference is the use of a high-energy laser beam heat source rather than an arc or gas flame. Laser beams offer potential in applying thin overlays or when access to the surface to be hardfaced can be achieved more readily by a laser beam than with an electrode or torch.

Cobalt-, nickel-, and tungsten carbide-base hardfacing alloys are the usual cladding materials used for laser hardfacing. As with conventional hardfacing methods, the materials are used in applications involving metal-to-metal contact, impact, erosion, and abrasion wear resistance. Other laser hardfacing materials include titanium carbide, Fe-Cr-Ni-B alloys, aluminum bronzes, and ceramics. Substrates have included carbon and low-alloy steels, stainless steels, nickel-base alloys (including superalloys), aluminum, cast irons, and tool steels (Ref 11).

Hardfacing material selection depends on its metallurgical compatibility with the substrate material. Alloys that form brittle intermetallic phases with the substrate are undesirable. For example, cobalt forms harmful intermetallics with titanium, and cobalt-base hardfacing alloys would be unsuitable on titanium alloy substrates.

Processing. The hardfacing alloy is melted by a laser beam and allowed to spread freely and freeze over the substrate. The beam also melts a very thin layer of the substrate, which combines with the liquid weld metal to the least extent necessary and solidifies to form a strong metallurgical bond. A good fusion bond can be achieved with a dilution zone that is only 10 to 20 μm thick.

The hardfacing alloy can be in several forms, examples of which are a prealloyed powder that is applied to the sample surface with or without a binder, a self-fluxing powder that is flame-sprayed, a hardfacing alloy that is plasma-sprayed, or a chip that is preplaced. Laser consolidation of these coatings results in densification and smoothening, eliminates channels to the substrate, improves the bonding between coating and substrate, and reduces porosity, all of which contribute to the strength and integrity of the hardfacing layer. Processing parameters for laser cladding are a power density that ranges from 10 to 1000 MW/m^2 and an interaction time from 0.1 to 1 s. The shielding gas could be any of the inert gases or a combination of gases, such as He/Ar and H_2/Ar.

To date the most successful application of laser hardfacing involves a proprietary process that utilizes a specially designed powder-feed apparatus. Using such equipment, the powder and an assist gas are fed to the weld area through a ceramic nozzle; a shielding gas of helium and argon surrounds the powder-gas mixture as it leaves the nozzle (Fig. 11). The powder delivery nozzle is positioned in such a manner as to flood the entire melt pool with powder. The feeding angle is generally from 35 to 45° to the horizontal, and the feed tube, which typically has a 3 mm (0.12 in.) diam, is positioned 10 to 12 mm (0.4 to 0.5 in.) from the substrate. Typically, the powder flow rate is from 0.005 to 0.1 cm^3/s (0.0003 to 0.006 $in.^3/s$), particle velocity is from 1 to 2 m/s (3.3 to 6.6 ft/s), and carrier gas velocity is from 3 to 7 m/s (10 to 23 ft/s). Desirable dilutions are from 3 to 8%. Overlay thickness can be varied from 0.15 to 4 mm (0.005 to 0.15 in.). Uniform feeding of the powder ensures uniform surfacing layers. Unlike plasma-sprayed coatings, porosity and unmelted powder particles are almost never observed within a laser hardfacing layer.

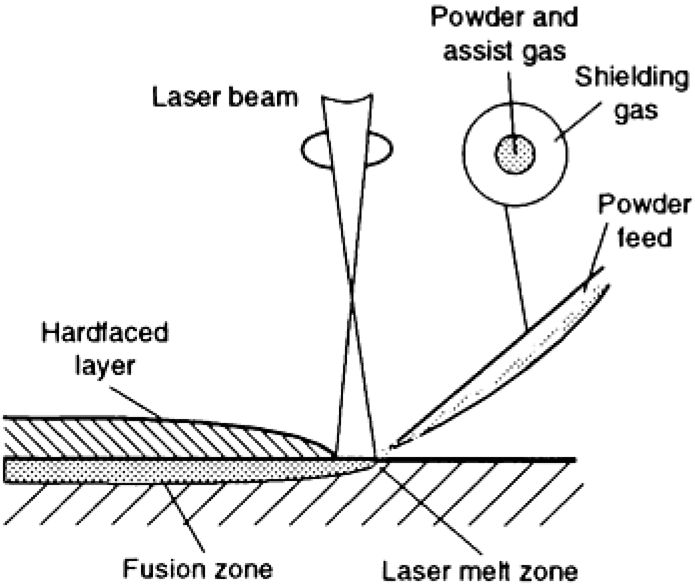

FIG. 11 SCHEMATIC OF THE LASER HARDFACING PROCESS USING DYNAMIC POWDER FEED. COURTESY OF QUANTUM LASER CORPORATION

Laser hardfacing deposits are commonly achieved by relatively large doughnut-shape beams, or by somewhat focused beams that are oscillated. These ensure a desirable deposit profile. Hardfacing material can also be added in rod, wire, or sheet form, but special procedures are needed because of reflectivity problems.

Thermal stresses in the weld metal can cause harmful cracking, but this can be eliminated by an appropriate preheating practice. Low power densities, large beam diameters, and slow sample translation rates tend to produce crack-free deposits.

Laser Hardfacing Applications. Because one of the primary goals of laser hardfacing is to improve the tribological properties of components, several applications have been found for the technique. Some of them are in the exploratory stage, others are in the pilot-plant stage, and a few have reached the production stage.

Ayers and Gnanamuthu (Ref 12) give several examples of laser hardfacing for improved wear resistance, such as Tribaloy T-800 on powder-coated ASTM A 387 steel, Haynes Stellite alloy No. 1 cast rods placed on AISI 4815 steel, and a mixture of WC and iron powder applied without binder on AISI 1018 steel.

In addition to stainless steel on mild steel, Steen (Ref 13) has reported successful laser hardfacing of nickel, bronze, Fe-B, and Stellite SF6 on mild steel; chromium on titanium; stainless on aluminum; mild steel on stainless; and Stellite on brass. Other examples of laser hardfacing are AISI 1016 steel clad with a mixture of chromium, manganese, and carbon powders (Ref 14) and gray cast iron hardfaced with a powder-fed mixture of molybdenum, chromium-carbide (Cr_3C_2), chromium, nickel, and silicon (Ref 15).

Bruck (Ref 16) demonstrated laser hardfacing within a confined space by coating the inside surface of small-bore pipes (inside diameter of 50 to 100 mm, or 2 to 4 in., and length of 0.3 to 1.2 m, or 1 to 4 ft). An oscillating mirror placed in the pipe formed a 12.5 mm (0.5 in.) wide melt pass, into which the hardfacing alloy powder was fed. The pipe was rotated and translated to cover the entire inner surface. Previously, chromium plating was used, but the laser technique appreciably improved galling resistance. Using suitable cooling techniques, the pipe temperature was maintained below 500 °C (930 °F) to avoid deterioration of the core properties, as well as distortion.

La Rocca (Ref 17) describes laser hardfacing of exhaust valves with a cobalt-base alloy at an Italian automotive manufacturer. The cobalt-base powder was fed into the laser melt pool. The laser hardfacing material was superior to GTAW material in terms of thickness uniformity, uniform microstructure and elemental distributions, and adhesion. When compared to GTAW, excess-metal removal was reduced by 10 to 15%. One British automotive manufacturer used production-stage laser hardfacing of a nickel-alloy turbine blade shroud interlock by powder feeding either a Tribaloy alloy or a Nimonic alloy (Ref 18). Previously, a manual arc melting technique was used. Laser hardfacing reduced hardfacing time from 14 min to 75 s, improved productivity and quality, reduced cost by 85%, and reduced powder consumption by 50%. Table 11 lists components that are laser hardfaced for commercial applications.

TABLE 11 TYPICAL APPLICATIONS FOR LASER HARDFACING

COMPONENT	**HARDFACING ALLOY**	**HARDFACING TECHNIQUE**	**COMPANY**
GATE AND SEAT OF STEEL VALVES FOR OIL-FIELD, GEOTHERMAL, AND NUCLEAR ENERGY PRODUCTION, AND CHEMICAL PROCESSING	STELLITE 6	POWDER FEED	W-K-M DIV. OF JOY MFG., HOUSTON, TX
DIESEL ENGINE VALVE	STELLITE SF6	PREPLACED POWDER PASTE	CABOT CORP., KOKOMO, IN
GATE VALVE SEAT	DELORO 60	PREPLACED POWDER PASTE	CABOT CORP., KOKOMO, IN
GAS TURBINE BLADE "Z" NOTCH	CO-BASE ALLOY	POWDER FEED	QUANTUM LASER, EDISON, NJ
LEADING EDGE OF TURBINE BLADES	CO-BASE ALLOY	POWDER FEED	QUANTUM LASER, EDISON, NJ
PUMP BUSHINGS AND IMPELLERS	CO-BASE ALLOY	POWDER FEED	QUANTUM LASER, EDISON, NJ
ENGINE VALVE BASES	NI-CR ALLOY	POWDER FEED	QUANTUM LASER, EDISON, NJ
OIL-FIELD VALVE GATES	WC AND CO ALLOY	POWDER FEED	QUANTUM LASER, EDISON, NJ
TRACTOR BUSHINGS	STELLITE 6	POWDER FEED	QUANTUM LASER, EDISON, NJ
TURBINE BLADES, PUMP VALVE SEATS	STELLITE, COLMOLOY, BLENDED POWDER	PREPLACED BEDS, GRAVITY FEED	WESTINGHOUSE, PITTSBURGH, PA
VALVE STEM, VALVE SEAT, ALUMINUM BLOCK	CRC_2, CR, NI, MO/CAST FE	PREPLACED POWDER	FIAT, TURIN, ITALY
OFF-SHORE DRILLING AND PRODUCTION PARTS, VALVE COMPONENTS, BOILER FIREWALL	STELLITE, COLMONOY, ALLOYS/CARBIDES	POWDER FEED	COMBUSTION ENG., OH
AEROSPACE COMPONENTS	STELLITE, TRIBALOY	POWDER FEED	ROCKWELL INT., CA
TURBINE BLADE SHROUD INTERLOCKS	PWA 694, NIMONICS	PREPLACED CHIP	PRATT & WHITNEY, WEST PALM BEACH, FL

Source: Ref 10

References cited in this section

1. K.C. ANTONY *ET AL.*, HARDFACING, *WELDING, BRAZING, AND SOLDERING,* 9TH ED., VOL 6, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1983, P 771-793
2. P. CROOK AND H.N. FARMER, FRICTION AND WEAR OF HARDFACING ALLOYS, *FRICTION,*

LUBRICATION, AND WEAR TECHNOLOGY, VOL 18, *ASM HANDBOOK,* ASM INTERNATIONAL, 1992, P 758-765

3. D.K. SUBRAMANYAM, A.E. SWANSIGER, AND H.S. AVERY, AUSTENITIC MANGANESE STEELS, *PROPERTIES AND SELECTION: IRONS, STEELS, AND HIGH-PERFORMANCE ALLOYS,* VOL 1, *ASM HANDBOOK,* ASM INTERNATIONAL, 1990, P 822-840
4. H.S. AVERY AND H.J. CHAPIN, *WELD J.,* VOL 31, 1952, P 917
5. R.D. ZORDAN, REPORT 10890, CABOT CORPORATION, KOKOMO, IN, 1982
6. H. OCKEN AND A.K. VELAN, ED., "LABORATORY EVALUATIONS OF COBALT-FREE, NICKEL-BASED HARDFACING ALLOYS FOR NUCLEAR APPLICATIONS," REPORT NP-4993, PROJECT 1935, ELECTRIC POWER RESEARCH INSTITUTE, PALO ALTO, CA, MARCH 1987
7. P.J. HOFMANN, B.C. FRIEDRICH, AND H. OCKEN, ED., "LABORATORY EVALUATIONS OF IRON-BASED HARDFACING ALLOYS," REPORT NP-5874, PROJECT 1935-11, ELECTRIC POWER RESEARCH INSTITUTE, PALO ALTO, CA, JUNE 1988
8. WELD DESIGN AND FABRICATION, DATA SHEET 420, *WELD. DES. FABR.,* VOL 48 (NO. 8), 1975, P 50
9. K.M. KULKARNI AND V. ANAND, METAL POWDERS USED FOR HARDFACING, *POWDER METALLURGY,* 9TH ED., VOL 7, *ASM HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1984, P 823-836
10. K.P. COOPER, LASER SURFACE PROCESSING, *FRICTION, LUBRICATION, AND WEAR TECHNOLOGY,* VOL 18, *ASM HANDBOOK,* ASM INTERNATIONAL, 1992, P 861-872
11. A.G. BLAKE, *ET AL.,* "LASER COATING TECHNOLOGY: A COMMERCIAL REALITY," BULLETIN OF QUANTUM LASER CORPORATION, EDISON, NJ
12. J.D. AYERS AND D.S. GNANAMUTHU, HARDENING OF METAL SURFACES BY LASER PROCESSING, *WELDING, BRAZING, AND SOLDERING,* 9TH ED., VOL 6, *ASM HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1983, P 793-803
13. W.M. STEEN, *PROC. NATO ADVANCED STUDY INSTITUTE ON LASER SURFACE TREATMENT OF METALS,* C.W. DRAPER AND P. MAZZOLDI, ED., MARTINUS NIJHOFF, 1986, P 369
14. J. MAZUMDER AND J. SINGH, *PROC. NATO ADVANCED STUDY INSTITUTE ON LASER SURFACE TREATMENT OF METALS,* C.W. DRAPER AND P. MAZZOLDI, ED., MARTINUS NIJHOFF, 1986, P 297
15. A. BELMONDO AND M. CASTAGNA, *SOURCEBOOK ON APPLICATIONS OF THE LASER IN METALWORKING,* E. METZBOWER, ED., AMERICAN SOCIETY FOR METALS, 1981, P 310
16. G.J. BRUCK, *J. MET.,* VOL 39 (NO. 2), 1987, P 10
17. A.V. LA ROCCA, *PROC. NATO ADVANCED STUDY INSTITUTE ON LASER SURFACE TREATMENT OF METALS,* C.W. DRAPER AND P. MAZZOLDI, ED., MARTINUS NIJHOFF, 1986, P 521
18. R.M. MACINTYRE, *PROC. NATO ADVANCED STUDY INSTITUTE ON LASER SURFACE TREATMENT OF METALS,* C.W. DRAPER AND P. MAZZOLDI, ED., MARTINUS NIJHOFF, 1986, P 545

Hardfacing, Weld Cladding, and Dissimilar Metal Joining

J.R. Davis, Davis & Associates

Thermal Spraying

Thermal spraying comprises a group of processes in which divided molten metallic or nonmetallic material is sprayed onto a prepared substrate to form a coating. The sprayed material is originally in the form of wire, rod, or powder. As the coating materials are fed through the spray unit, they are heated to a molten or plastic state and propelled by a stream of compressed gas onto the substrate. As the particles strike the surface, they flatten and form thin platelets that conform and adhere to the irregularities of the prepared surface and to each other. They cool and accumulate, particle by particle, into a lamellar, castlike structure. In general, the substrate temperature can be kept below approximately 200 °C (400 °F), eliminating metallurgical change of the substrate material. The spray gun generates the necessary heat for melting through combustion of gases, an electric arc, or a plasma. Figure 12 illustrates a general thermal spraying process.

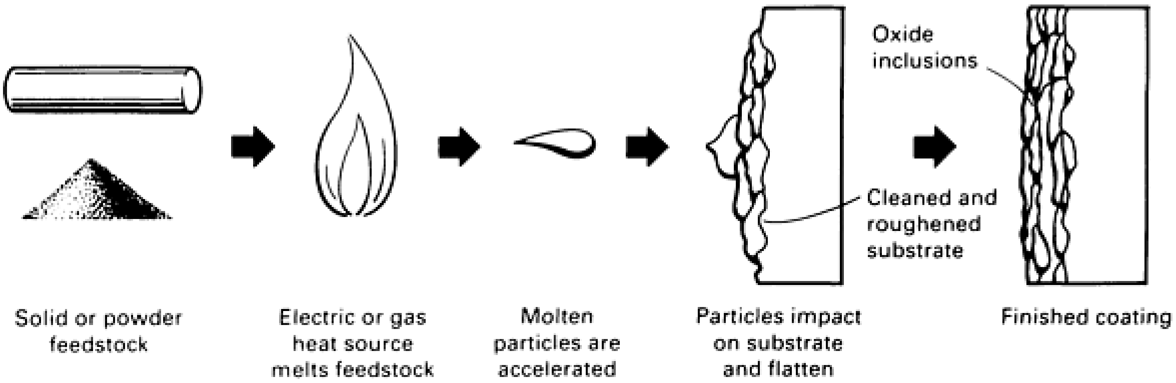
Oxide inclusions
Cleaned and roughened substrate
Solid or powder feedstock
Electric or gas heat source melts feedstock
Molten particles are accelerated
Particles impact on substrate and flatten
Finished coating

FIG. 12 SCHEMATIC OF THE GENERAL THERMAL SPRAY PROCESS. SOURCE: REF 20

The deposited structures of thermal spray coatings differ from those of the same material in the wrought form because of the incremental nature of the coating buildup and because the coating composition is often affected by reaction with the process gases and the surrounding atmosphere while the materials are in the molten state. For example, where air or oxygen is used as the process gas, oxides of the material applied may be formed and become part of the coating. The as-applied structures of all thermal spray coatings are similar in their lamellar nature; the variations in structure depend on the particular thermal spray process used, the processing parameters and techniques employed, and the material applied. Figure 13 illustrates the microstructure that results from the thermal spray process. As shown in this figure, the molten particles spread out and deform (splatter) as they impact the substrate, at first locking onto irregularities on the roughened surface, then interlocking with each other.

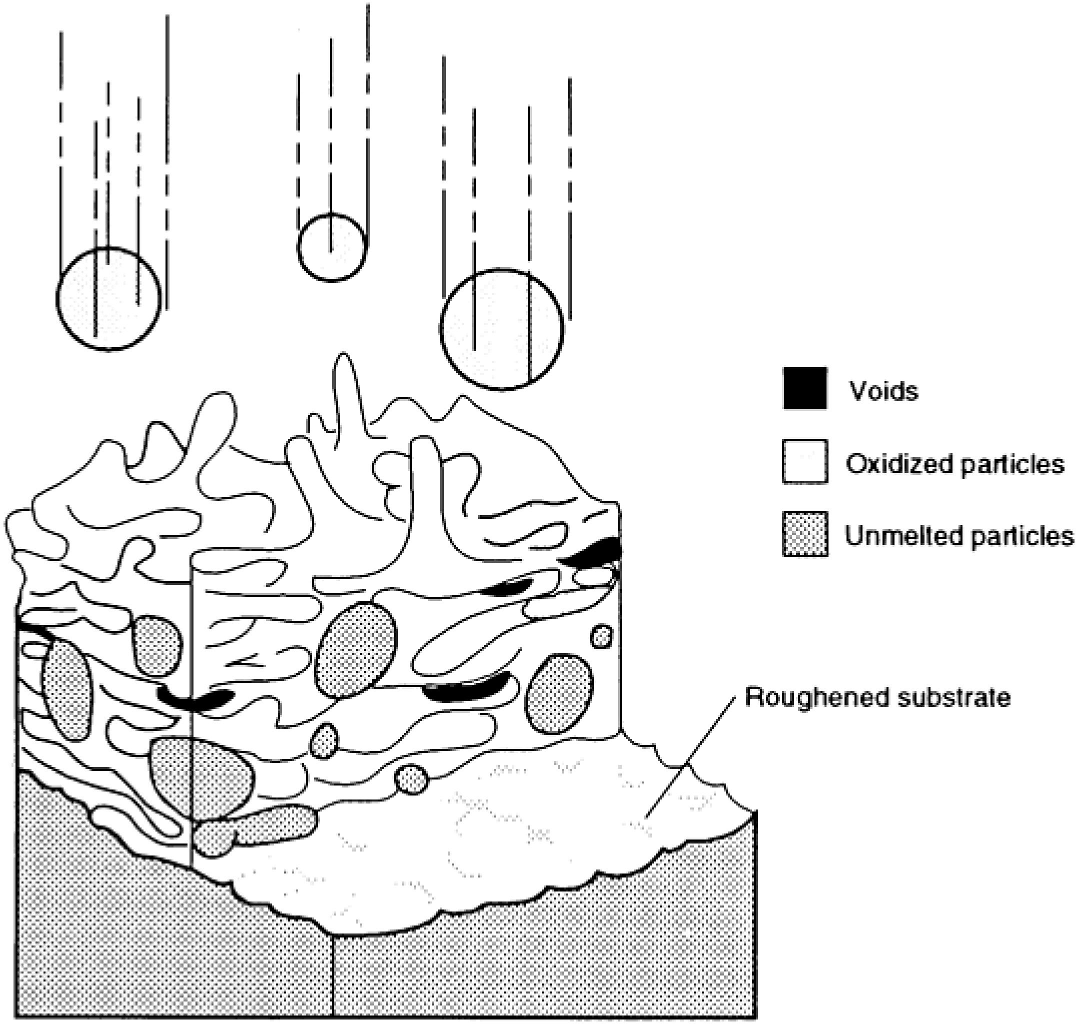

FIG. 13 SCHEMATIC SHOWING THE BUILDUP OF A THERMAL SPRAY COATING. MOLTEN PARTICLES SPREAD OUT AND DEFORM (SPLATTER) AS THEY STRIKE THE TARGET, AT FIRST LOCKING ONTO IRREGULARITIES ON THE SUBSTRATE, THEN INTERLOCKING WITH EACH OTHER. VOIDS CAN OCCUR IF THE GROWING DEPOSIT TRAPS AIR. PARTICLES OVERHEATED IN THE FLAME BECOME OXIDIZED. UNMELTED PARTICLES MAY SIMPLY BE EMBEDDED IN THE ACCUMULATING DEPOSIT. COURTESY OF THE MATERIALS ENGINEERING INSTITUTE, ASM INTERNATIONAL

The bond between the sprayed coating and the substrate is generally mechanical. Proper surface preparation of the substrate before spraying is often the most critical influence on the bond strength of the coating. Surface preparation techniques will be discussed later in this section.

Thermal spraying is a very versatile technique in that it can be used both in-shop or on-site. The thermal spray process permits rapid application of high-performance material in thicknesses from a few mils to more than 25 mm (1 in.) thick on parts of a variety of sizes and geometries. A list of applications would include coatings for:

- WEAR RESISTANCE
- OXIDATION RESISTANCE
- CORROSION RESISTANCE
- DIMENSIONAL RESTORATION
- THERMAL BARRIERS
- ELECTRICAL CONDUCTIVITY OR RESISTIVITY
- BIOMEDICAL DEVICES

Thermal Spray Processes (Ref 19, 20)

Currently, five different commercially available thermal spray methods are in use:

- OXYFUEL WIRE (OFW) SPRAY
- ELECTRIC ARC WIRE (EAW) SPRAY
- OXYFUEL POWDER (OFP) SPRAY
- PLASMA ARC (PA) POWDER SPRAY
- HIGH-VELOCITY OXYFUEL (HVOF) POWDER SPRAY

Of these, the plasma arc process is most commonly used. Market studies indicate that by the year 2000, plasma spraying will capture about 48% of the market, the HVOF process 25%, the electric-arc method 15%, and the combustion methods (OFW and OFP) a combined 12% (Ref 21). Table 12 compares important process characteristics associated with these techniques. Selection of the appropriate thermal spray method is typically determined by:

- DESIRED COATING MATERIAL
- COATING PERFORMANCE REQUIREMENTS
- ECONOMICS
- PART SIZE AND PORTABILITY

TABLE 12 COMPARISON OF THERMAL SPRAY PROCESSES

PROCESS	**GAS FLOW**		**FLAME OR EXIT PLASMA TEMPERATURE**		**PARTICLE IMPACT VELOCITY**		**RELATIVE ADHESIVE STRENGTH(A)**	**COHESIVE STRENGTH**	**OXIDE CONTENT, %**	**RELATIVE PROCESS COST(A)**	**MAXIMUM SPRAY RATE**		**POWER**		**ENERGY REQUIRED TO MELT**	
	M^3/H	**FT^3/H**	**°C**	**°F**	**M/S**	**FT/S**					**KG/H**	**LB/H**	**KW**	**HP**	**KW/KG**	**KW/LB**
FLAME POWDER: POWDER FEEDSTOCK, ASPIRATED INTO THE OXYGEN/FUEL-GAS FLAME, IS MELTED AND CARRIED BY THE FLAME ONTO THE WORKPIECE. PARTICLE VELOCITY IS RELATIVELY LOW, AND BOND STRENGTH OF DEPOSITS IS LOW. POROSITY IS HIGH AND COHESIVE STRENGTH IS LOW. SPRAY RATES ARE USUALLY IN THE 0.5 TO 9 KG/H (1 TO 20 LB/H) RANGE. SURFACE TEMPERATURES CAN RUN QUITE HIGH.	11	400	2200	4000	30	100	3	LOW	6	3	7	15	25-75	34-100	11-22	5-10
FLAME WIRE: IN FLAME WIRE SPRAYING, THE ONLY FUNCTION OF THE FLAME IS TO MELT THE MATERIAL. A STREAM OF AIR THEN DISINTEGRATES THE MOLTEN MATERIAL AND PROPELS IT ONTO THE WORKPIECE. SPRAY RATES FOR MATERIALS SUCH	71	2500	2800	5000	180	600	4	MEDIUM	4	3	9	20	50-100	70-135	11-22	5-10

AS STAINLESS STEEL ARE IN THE RANGE OF 0.5 TO 9 KG/H (1 TO 20 LB/H). SUBSTRATE TEMPERATURES ARE FROM 95 TO 205 °C (200 TO 400 °F) BECAUSE OF THE EXCESS ENERGY INPUT REQUIRED FOR FLAME MELTING.																
WIRE ARC: TWO CONSUMABLE WIRE ELECTRODES ARE FED INTO THE GUN, WHERE THEY MEET AND FORM AN ARC IN AN ATOMIZING AIR STREAM. THE AIR FLOWING ACROSS THE ARC/WIRE ZONE STRIPS OFF THE MOLTEN METAL, FORMING A HIGH-VELOCITY SPRAY STREAM. THE PROCESS IS ENERGY EFFICIENT; ALL INPUT ENERGY IS USED TO MELT THE METAL. SPRAY RATE IS ABOUT 2.3 KG/H/KW (5 LB/H/KW). SUBSTRATE TEMPERATURE CAN BE LOW BECAUSE ENERGY INPUT PER POUND OF METAL IS ONLY ABOUT ONE-EIGHTH THAT OF OTHER SPRAY METHODS.	71	2500	5500	10,000	240	800	6	HIGH	0.5-3	1	16	35	4-6	5-8	0.2-0.4	0.1-0.2
CONVENTIONAL PLASMA: CONVENTIONAL	4.2	150	5500	10,000	240	800	6	HIGH	0.5-1	5	5	10	30-80	40-110	13-22	6-10

PLASMA SPRAYING PROVIDES FREE-PLASMA TEMPERATURES IN THE POWDER HEATING REGION OF 5500 °C (10,000 °F) WITH ARGON PLASMA, AND 4400 °C (8000 °F) WITH NITROGEN PLASMA--ABOVE THE MELTING POINT OF ANY KNOWN MATERIAL. TO GENERATE THE PLASMA, AN INERT GAS IS SUPERHEATED BY PASSING IT THROUGH A *DC* ARC. POWDER FEEDSTOCK IS INTRODUCED AND IS CARRIED TO THE WORKPIECE BY THE PLASMA STREAM. PROVISIONS FOR COOLING OR REGULATION OF THE SPRAY RATE MAY BE REQUIRED TO MAINTAIN SUBSTRATE TEMPERATURES IN THE 95 TO 205 °C (200 TO 400 °F) RANGE. TYPICAL SPRAY RATE IS 0.1 KG/H/KW (0.2 LB/H/KW).																
DETONATION GUN: SUSPENDED POWDER IS FED INTO A 1 M (3 FT) LONG TUBE ALONG WITH OXYGEN AND FUEL GAS. A	11	400	3900	7000	910	3000	8	VERY HIGH	0.1	10	1	2	100-270	135-360	220	100

SPARK IGNITES THE MIXTURE AND PRODUCES A CONTROLLED EXPLOSION. THE HIGH TEMPERATURES AND PRESSURES (1 MPA, OR 150 PSI) THAT ARE GENERATED BLAST THE PARTICLES OUT OF THE END OF THE TUBE TOWARD THE SUBSTRATE.																
HIGH-VELOCITY OXYFUEL: IN HVOF SPRAYING, A FUEL GAS AND OXYGEN ARE USED TO CREATE A COMBUSTION FLAME AT 2500 TO 3100 °C (4500 TO 5600 °F). THE COMBUSTION TAKES PLACE AT VERY HIGH CHAMBER PRESSURE, EXITING THROUGH A SMALL-DIAMETER BARREL TO PRODUCE A SUPERSONIC GAS STREAM AND VERY HIGH PARTICLE VELOCITIES. THE PROCESS RESULTS IN EXTREMELY DENSE, WELL-BONDED COATINGS, MAKING IT ATTRACTIVE FOR MANY CORROSION-RESISTANT APPLICATIONS.	28-57	1000-2000	3100	5600	610-1060	2000-3500	8	VERY HIGH	0.2	5	14	30	100-270	135-360	22-200	10-90

EITHER POWDER OR WIRE FEEDSTOCK CAN BE SPRAYED, AT TYPICAL RATES OF 2.3 TO 14 KG/H (5 TO 30 LB/H).																
HIGH-ENERGY PLASMA: THE HIGH-ENERGY PLASMA PROCESS PROVIDES SIGNIFICANTLY HIGHER GAS ENTHALPIES AND TEMPERATURES, ESPECIALLY IN THE POWDER HEATING REGION, DUE TO A MORE STABLE, LONGER ARC AND HIGHER POWER DENSITY IN THE ANODE NOZZLE. THE ADDED POWER (TWO TO THREE TIMES THAT OF CONVENTIONAL PLASMA) AND GAS FLOW (TWICE AS HIGH) PROVIDE A LARGER, HIGHER TEMPERATURE POWDER INJECTION REGION AND REDUCED AIR ENTRAINMENT. ALL THIS LEADS TO IMPROVED POWDER MELTING, FEWER UNMELTS, AND HIGH PARTICLE IMPACT VELOCITY.	17-28	600-1000	8300	15,000	240-1220	800-4000	8	VERY HIGH	0.1	4	23	50	100-250	135-335	9-13	4-6
VACUUM PLASMA: VACUUM PLASMA USES A CONVENTIONAL PLASMA TORCH IN A CHAMBER AT	8.4	300	8300	15,000	240-610	800-2000	9	VERY HIGH	(B)	10	10	24	50-100	70-135	11-22	5-10

PRESSURES IN THE RANGE OF 10 TO 50 KPA (0.1 TO 0.5 ATM). AT LOW PRESSURES THE PLASMA IS LARGER IN DIAMETER, IS LONGER, AND HAS A HIGHER VELOCITY. THE ABSENCE OF OXYGEN AND THE ABILITY TO OPERATE WITH HIGHER SUBSTRATE TEMPERATURES PRODUCE DENSER, MORE ADHERENT COATINGS HAVING MUCH LOWER OXIDE CONTENTS.																

Source: Ref 21

(A)	1 (LOW) TO 10 (HIGH).
(B)	PPM LEVELS.

The oxyfuel wire spray process (also called wire flame spraying or the combustion wire process) is the oldest of the thermal spray coating methods and among the lowest in capital investment. The process utilizes an oxygen-fuel gas flame as a heating source and coating material in wire form. Any material in the form of wire and capable of being melted below 2480 °C (4500 °F) can be flame-sprayed. Solid-rod feed stock has also been used. During operation, the wire is drawn into the flame by drive rolls that are powered by an adjustable air turbine or electric motor. The tip of the wire is melted as it enters the flame, atomized into particles by a surrounding jet of compressed air, and propelled to the workpiece. Figure 14 shows a schematic of the OFW spray process.

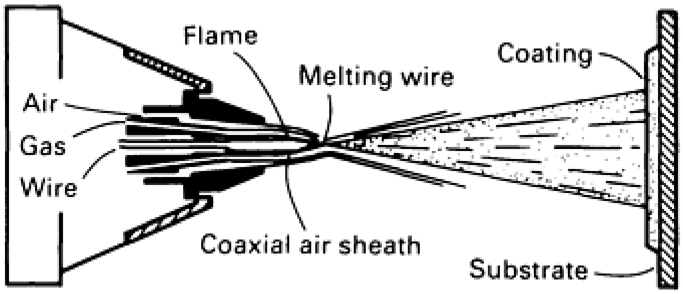

FIG. 14 SCHEMATIC OF THE OXYFUEL WIRE SPRAY PROCESS. SOURCE: REF 20

Spray rates for this process range from 0.5 to 10 kg/h (1 to 20 lb/h) and are dictated by the melting point of the material and the choice of fuel gas. Common fuel gases are acetylene, MAPP gas, propane, propylene, and natural gas, each combined with oxygen. The wire spray gun is most commonly used as a hand-held device for on-site application, although an electric-motor-driven gun is recommended for fixed-mounted use in high-volume, repetitive production work.

The OFW process is widely used for corrosion protection of large outdoor structures, such as bridges and storage tanks, and for restoration of dimension to worn machinery components. It is a good choice for all-purpose spraying. Coatings can be applied rapidly and at low cost, and a wide variety of metal coating materials are available. Typical spray materials include austenitic and martensitic stainless steels, nickel aluminide, nickel-chromium alloys, bronze, Monel, babbitt, aluminum, zinc, and molybdenum. In general, oxyfuel wire flame-sprayed coatings exhibit lower bond strengths, higher porosity, a narrower working temperature range, and higher heat transmittal to the substrate than plasma or electric arc-sprayed coatings.

Electric arc wire spraying also applies coatings of selected metals in wire form. Push-pull motors feed two electrically charged wires through the arc gun to contact tips at the gun head. An arc is created that melts the wires at temperatures above 5500 °C (10,000 °F). Compressed air atomizes the molten metal and projects it onto a prepared surface. Figure 15 shows a schematic of the EAW spray process.

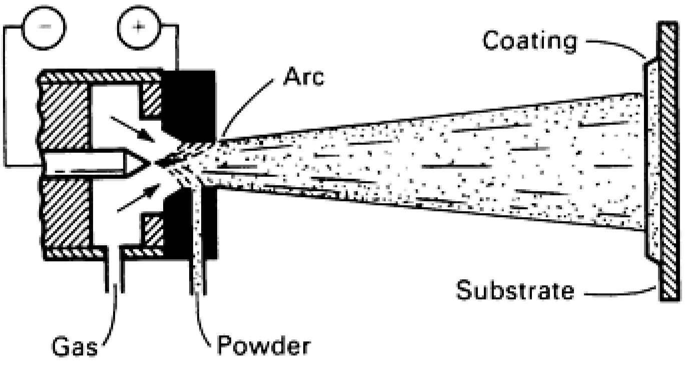

FIG. 15 SCHEMATIC OF THE ELECTRIC ARC WIRE SPRAY PROCESS. SOURCE: REF 20

The EAW process is excellent for applications that require heavy coating buildup or have large surfaces to be sprayed. The arc system can produce a spray pattern ranging from 50 to 300 mm (2 to 12 in.) and can spray at high speeds. It has built-in flexibility, allowing coating characteristics, such as hardness or surface texture, to be tailored to specific applications.

The EAW method is characterized by strong coating adhesion because of the high particle temperatures produced. Because the process uses only electricity and compressed air, it allows equipment to be moved relatively easily from one installation to another, and it eliminates the need to stock oxygen and fuel gas supplies. Materials applied by the EAW method are similar to those used in the OFW process.

The oxyfuel powder spray method, which also uses combustion gases as the heat source, extends the range of available coatings and subsequent applications to include ceramics, cermets, carbides, and fusible hardfacing coatings. Using either gravity flow or pressurized feed, powder is fed into the gun and carried to the gun nozzle, where it is melted and projected by the gas stream onto a prepared surface (Fig. 16). For general-purpose spraying, the gravity-flow system is used. When exacting coating consistency and/or high spray rates are desired, the pressurized feed system is used. Oxyfuel powder guns are the lowest-cost thermal spray equipment, are easiest to set up, and make it easy to change coating materials. The OFP method finds widest use in short-run machinery maintenance work and in the production spraying of abradable clearance-control seals for gas turbine engines. The OFP spray equipment is lighter and more compact than other forms of thermal spray equipment, but because of the lower particle velocities obtained, the coatings generally have lower bond strength, higher porosity, and a lower overall cohesive (interparticle) strength than the coatings produced by other thermal spray processes.

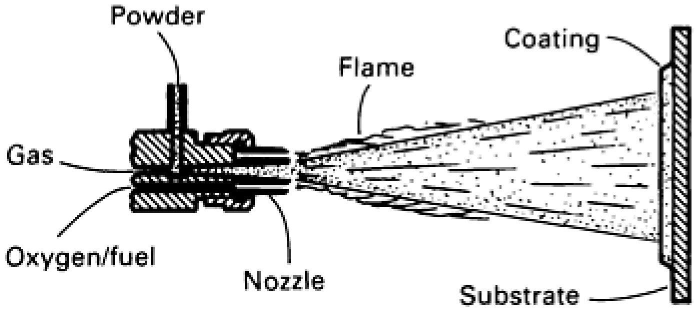

FIG. 16 SCHEMATIC OF THE OXYFUEL POWDER SPRAY PROCESS. SOURCE: REF 20

Flame spray and fuse is a modification of the OFP spray method. The materials used for coating are self-fluxing, fusible materials which require postspray heat treatment. In general, these are the boride-containing nickel-base alloys described earlier in this article. In practice, parts are prepared and coated as in other thermal spray processes and then fused. Fusing is accomplished using one of several techniques, such as flame or torch; induction; or in vacuum, inert, or hydrogen furnaces. These alloys generally fuse between 1010 to 1175 °C (1850 to 2150 °F), depending upon composition.

Spray-and-fuse coatings are used in applications where excessive wear is a problem. Coating hardnesses can be as high as 65 HRC. The use of spray-and-fuse coatings is limited to substrate materials that can tolerate the 1010 to 1175 °C (1850 to 2150 °F) fusing temperatures. Fusing temperatures may alter the heat-treated properties of some alloys. However, the coating will usually withstand additional heat treatment of the substrate.

Plasma arc spraying is one of the most sophisticated and versatile thermal spray methods. The PA spray process uses coating materials in the form of powders, which are heated with a plasma heat source (Fig. 17). Temperatures that can be obtained with commercial plasma equipment have been calculated to be greater than 11,000 °C (20,000 °F) and are far above the melting or even the vaporization point of any known material. Decomposition of materials during spraying is minimized because of the high gas velocities produced by the plasma, resulting in extremely short residence time in the thermal environment. The plasma process also provides a controlled atmosphere for melting and transport of the coating material, thus minimizing oxidation, and the high gas velocities produce coatings of high density.

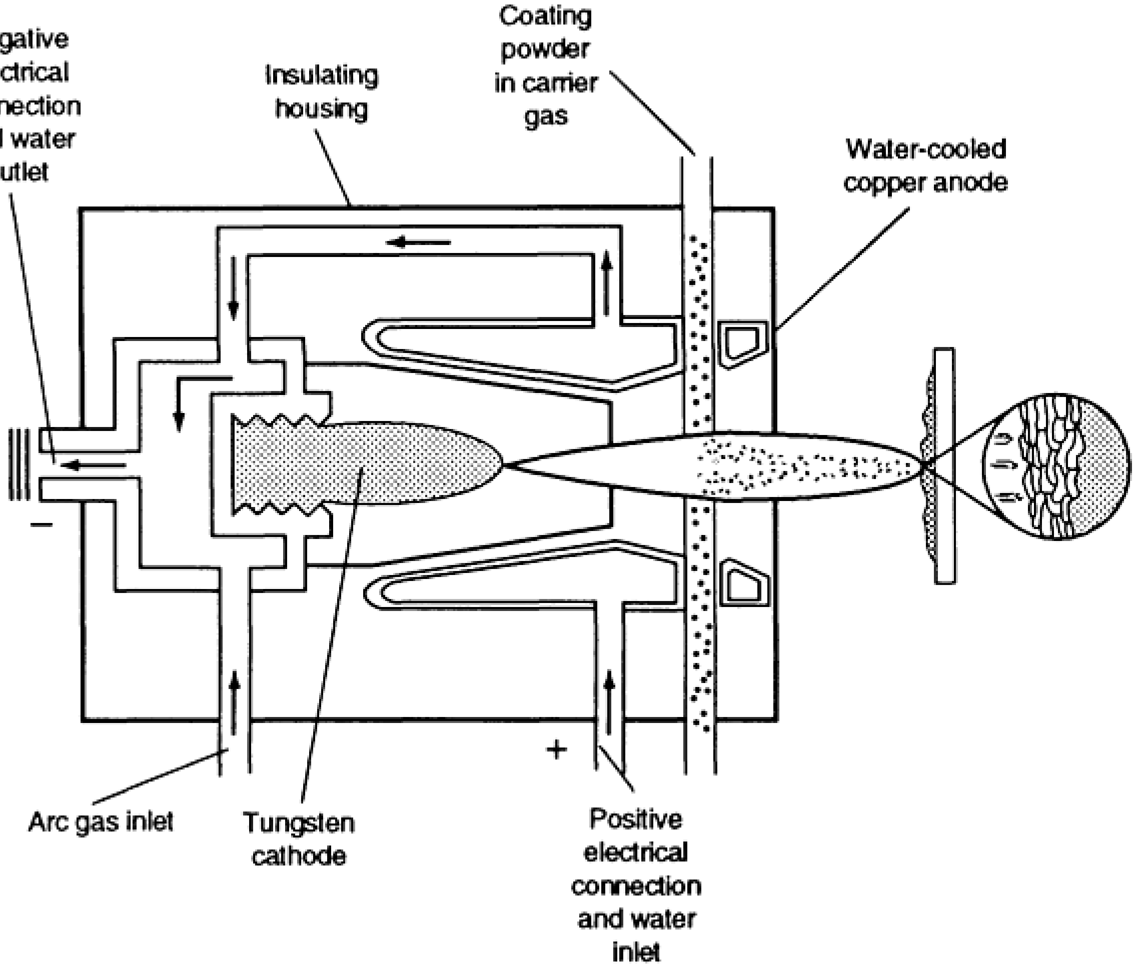
Negative electrical connection and water outlet
Insulating housing
Coating powder in carrier gas
Water-cooled copper anode
−
+
Arc gas inlet
Tungsten cathode
Positive electrical connection and water inlet

FIG. 17 SCHEMATIC OF THE PLASMA ARC SPRAY PROCESS. COURTESY OF D. BAXTER, *ADVANCED MATERIALS & PROCESSES*

The plasma gun operates on the principle of raising the energy state of a gas by passing it through an electric arc. The release of energy in returning the gas to its ground state results in exceedingly high temperatures. A gas such as nitrogen or argon enters a direct-current arc between a tungsten cathode and a copper anode that make up the nozzle (Fig. 17). Both components are cooled by a constant flow of water through internal passages. Here the plasma gas first dissociates (in the case of nitrogen, into two atoms), followed by ionization that releases free electrons. The electrons recombine outside the electric arc, and energy is released as heat and light. In addition, frequent collisions transfer energy from the electrons to the positive ions, accelerating them until the plasma reaches a state of equilibrium. The result is a thermal plasma in which the energy of the electrons has been turned into enthalpy, or heat content. At this point, powdered coating material suspended in a gas is injected into the plasma and is subsequently melted and propelled at high velocity to the workpiece. In practice, a small amount of a secondary gas, such as hydrogen or helium, is mixed with the primary plasma gas to increase operating voltage and thermal energy.

The high temperatures and high gas velocities produced by the plasma process result in coatings that are superior in mechanical and metallurgical properties to low-velocity OFW or OFP coatings. The plasma process is particularly efficient for spraying high-quality coatings of ceramic materials, such as zirconium oxide for turbine engine combustors and chromium oxide for printing rolls. The plasma process is also readily field-portable and is thus used for large on-site applications such as power utility plant boiler tubes.

Current plasma spray technology permits fully automatic start/stop operation and closed-loop computer control for power level, plasma gas flow, and powder feed rate. System problems can be diagnosed via computer modem to the equipment manufacturer, and system performance can be documented and stored on a digital recorder or data logger.

With any thermal spray powder method, the degree to which a given flame effectively melts and accelerates the powder depends on the type of coating material and the size and shape of the particles. Each particular coating material and gun combination has an optimum particle size. Particles much smaller than ideal will overheat and vaporize; much larger particles will not melt and may fall from the flame or rebound from the target. For the highest coating density, powders are sized between about 44 μm (325 mesh) and 10 μm (Ref 1, 9). The lower limit ensures free flow. Coarser powder may be sized up to 100 μm, with typical distribution falling between 88 μm (170 mesh) and 44 μm (325 mesh). Powder shape ranges from spherical to acicular; spherical is best because of reduced surface area and consistent flow.

Any material that melts without subliming and is available as a properly sized powder has potential as a plasma spray process coating. Plasma spray coatings may be elemental (aluminum, molybdenum, nickel, or chromium), alloyed (nickel-, iron-, or cobalt-base), ceramics and cermets (Al_2O_3, Cr_2O_3, or Cr_3C_2), composites (nickel-aluminum, cobalt-bonded tungsten carbide, or clad Al_2O_3), or mechanical blends.

In one process variation, the plasma gun and workpiece are enclosed in a vacuum chamber, and the entire operation is carried out in an inert atmosphere at low pressure (7 kPa, or 50 torr). Advantages of this low-pressure plasma technique include improved bonding and density of the deposit, improved control over coating thickness (even with an irregular work surface), and higher deposit efficiency. However, facility cost is higher, and workpiece size is limited by the vacuum chamber size. This modification is suitable for coating turbine airfoils, blade tips, and shroud segments with complex coatings such as NiCrAlY and other MCrAlY-type alloys. Very fine powder (-325 mesh) generally is used for low-pressure plasma spraying to achieve maximum density. If high-purity coatings are also desired, then the powder must be manufactured under inert cover and atomized in an inert gas to preserve purity.

The high-velocity oxyfuel powder spray process represents the state of the art for thermal spray coatings. The HVOF process uses extremely high kinetic energy and controlled thermal energy output to produce very-low-porosity coatings that exhibit high bond strength, fine as-sprayed surface finish, and low residual stresses. Both pulsed-combustion and continuous-combustion HVOF methods are used.

Pulsed-Combustion HVOF. The detonation gun (or D-gun) process was the first HVOF method introduced and was initially developed for the deposition of hard, wear-resistant materials such as oxides and carbides (Ref 1, 9). The extremely high particle velocities achieved in the detonation gun result in coatings with higher density, greater internal strength, and superior bond strength than can be achieved with conventional plasma spraying or oxyfuel powder spraying. Detonation gun coatings have been successfully applied to critical areas of precision components made from virtually all commercial alloys. The typical coating thickness is less than 0.75 mm (0.030 in.), but many applications require only 0.075 mm (0.003 in.) or less of finished coating.

This process, which is depicted schematically in Fig. 18, uses a controlled series of explosions of a mixture of oxygen and acetylene to blast the powder onto the surface of the workpiece. The detonation gun consists of a water-cooled barrel several feet long, with an inside diameter of about 25 mm (1 in.), and associated gas- and powder-metering equipment. In operation, a mixture of oxygen and acetylene is fed into the barrel along with a charge of powder. The gas is then ignited, and the detonation wave accelerates the powder to about 760 m/s (2500 ft/s), while heating it close to, or above, its melting point. Temperatures up to 3900 °C (7000 °F) are reached inside the gun. The distance that the powder is entrained in the high-velocity gas is much longer than in a plasma or flame spray device, which accounts for the high particle velocity. After the powder has exited the barrel, a pulse of nitrogen purges the barrel. The gun discharges four to eight times per second.

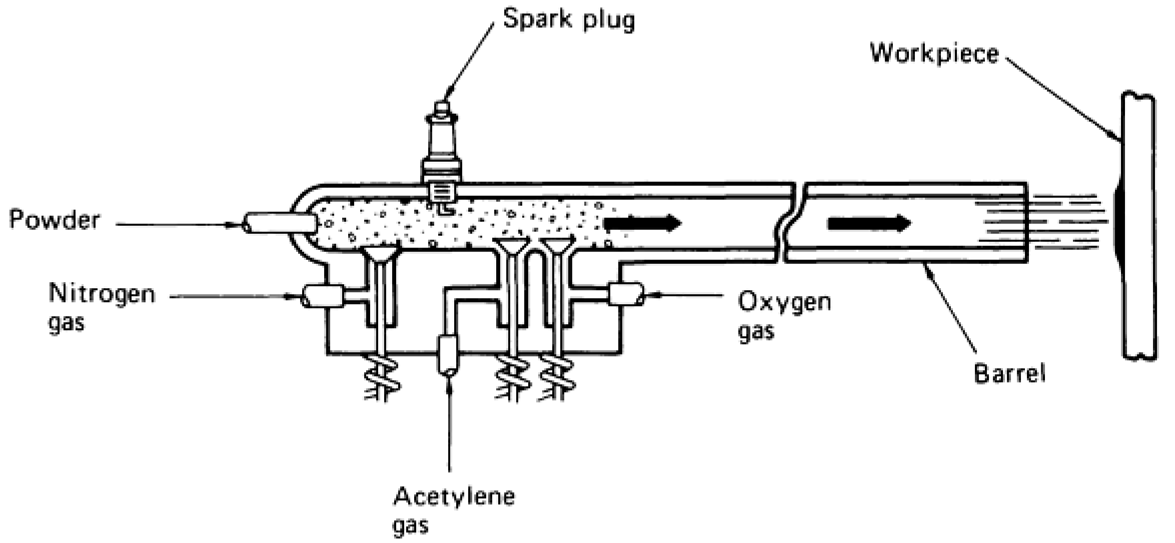

FIG. 18 SCHEMATIC OF THE DETONATION GUN (D-GUN) HVOF PROCESS. SOURCE: REF 1

Suitable materials for detonation gun coatings include pure metals, alloys, oxides, carbides, composites, and mechanical blends of two or more components. Historically, the detonation gun has proven particularly advantageous in applying hard wear-resistant carbide and oxide coatings. Detonation gun coatings have been successfully applied to critical areas of precision parts made from virtually all commercial alloys. The substrate surface is seldom heated above 150 °C (300 °F) during the coating process. As a result, a part can be fabricated and fully heat treated prior to coating. In addition, very high melting point materials (such as WC-Co) can be deposited on relatively low melting point substrates (such as aluminum). Usually, there is little or no distortion, and the substrate microstructure and strength are not changed by the coating process.

Continuous-combustion HVOF processes, which also feature internal, confined combustion, operate with an oxyfuel mixture consisting of oxygen and either propylene, propane, hydrogen, and even liquid fuels such as kerosene. High-velocity oxyfuel guns have combustor sections (air- or water-cooled), which combust the fuel/oxygen mixtures under pressure and accelerates the combusted gas streams down a confined cooled tube. Powders, with carrier gases, are fed into the nozzle where the particles are entrained with the confined, high-pressure combusting gases.

The high-velocity gases produce uniquely characteristic multiple shock diamond patterns, which are visible in the flame (Fig. 19). Combustion temperatures approach 2750 °C (5000 °F) and form a circular flame configuration. Powder is injected into the flame axially to provide uniform heating, and powder particles are accelerated by the high velocity gases, which typically approach speeds ranging from 1370 m/s (4500 ft/s) to as high as 2160 m/s (7100 ft/s). Particle velocities range from 760 m/s (2500 ft/s) to more than 900 m/s (3000 ft/s).

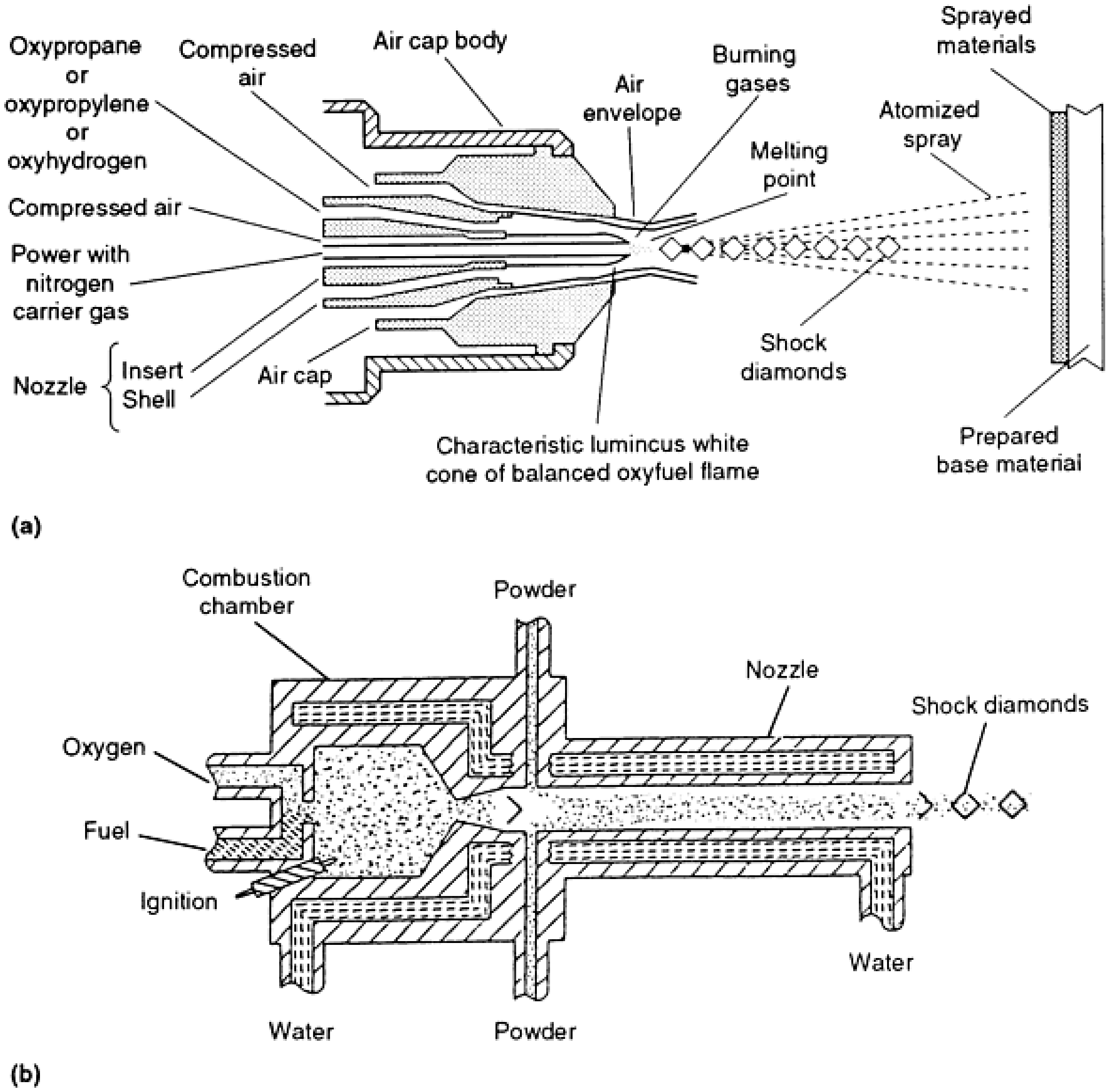
Oxypropane or oxypropylene or oxyhydrogen
Compressed air
Air cap body
Air envelope
Burning gases
Sprayed materials
Atomized spray
Melting point
Compressed air
Power with nitrogen carrier gas
Nozzle
Insert
Shell
Air cap
Shock diamonds
Characteristic lumincus white cone of balanced oxyfuel flame
Prepared base material
(a)
Combustion chamber
Powder
Nozzle
Shock diamonds
Oxygen
Fuel
Ignition
Water
Water
Powder
(b)

FIG. 19 SCHEMATICS OF (A) AIR-COOLED AND (B) WATER-COOLED CONTINUOUS COMBUSTION HVOF THERMAL SPRAY PROCESSES. (A) COURTESY OF METCO PERKIN ELMER. (B) SOURCE: REF 22

Continuous-combustion HVOF spray coating processes are separated into four major designs: (1) internal combustion, with the hot gases directed into a chamber where they mix with and heat a mixture of carrier gas and powder particles; (2) simultaneous injection of powder into a combustion chamber, where the mixture continuously bums and heats the particles; (3) confined combustion into which a stream of powder with carrier gas is injected; and (4) high-pressure liquid fuel systems with radial injection of powders downstream of the combustor exit. Figure 19 illustrates two of these variations. Common to all HVOF methods is that the gases are being burned under higher-than-atmospheric pressures, with the combustion gases, carrier gas, and an entrained distribution of powders being accelerated in the tube. The length of the nozzle, the design of heating chambers, the geometry/location of powder feed, and the cooling, either by water or air, are the major features which differ.

The low residual coating stress produced in the continuous-combustion HVOF process allows significantly greater thickness capability than the plasma method, while providing lower porosity, lower oxide content, and higher coating adhesion. Coatings produced by the continuous-combustion HVOF process also have much better machinability compared with other methods, and coating porosity has closely approached wrought materials, as verified by gas permeability testing. Continuous-combustion HVOF systems are available with closed-loop computer control and robotics capability. Figures 20 and 21 compare the bond strengths and hardness of coatings deposited with various thermal spray processes. The correlation between hardness and particle velocity is shown in Fig. 22.

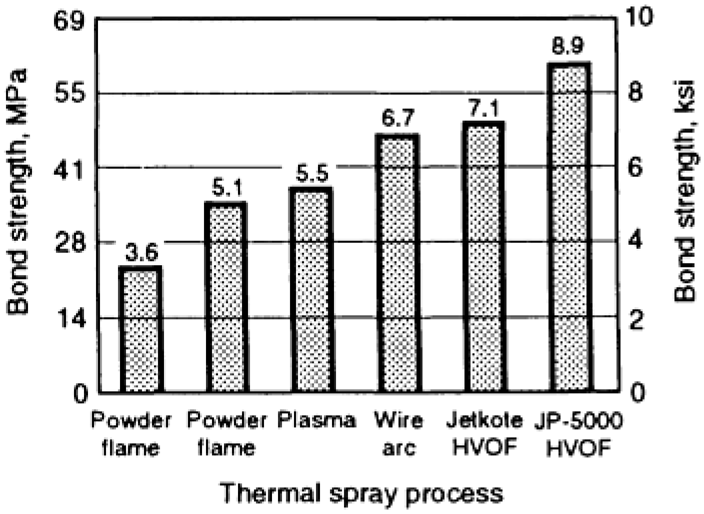

FIG. 20 BOND STRENGTHS OF TYPE 316 STAINLESS STEEL COATINGS DEPOSITED BY VARIOUS THERMAL SPRAY PROCESSES. JETKOTE IS A REGISTERED TRADENAME OF THERMODYNE, INC.; JP-5000 IS A REGISTERED TRADENAME OF HOBART TAFA TECHNOLOGIES, INC. SOURCE: REF 22

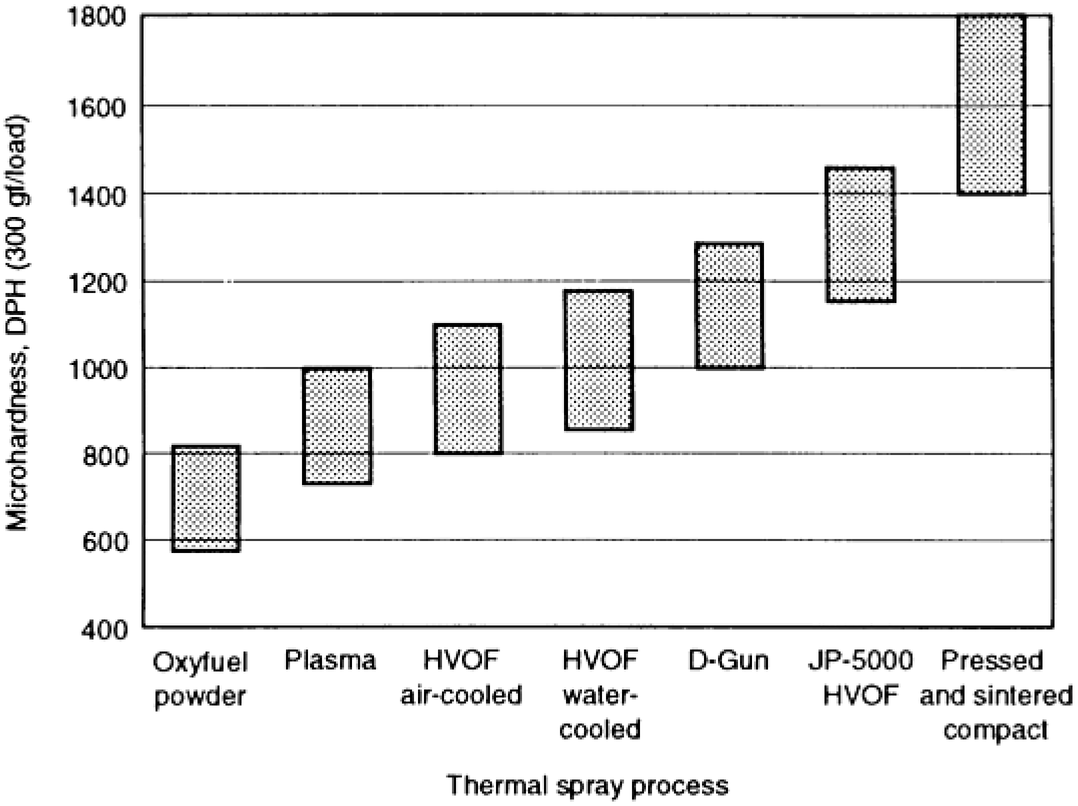

FIG. 21 HARDNESS OF WC-CO COATING VERSUS THERMAL SPRAY PROCESS. JP-5000 IS A REGISTERED TRADENAME OF HOBART TAFA TECHNOLOGIES, INC. SOURCE: REF 22

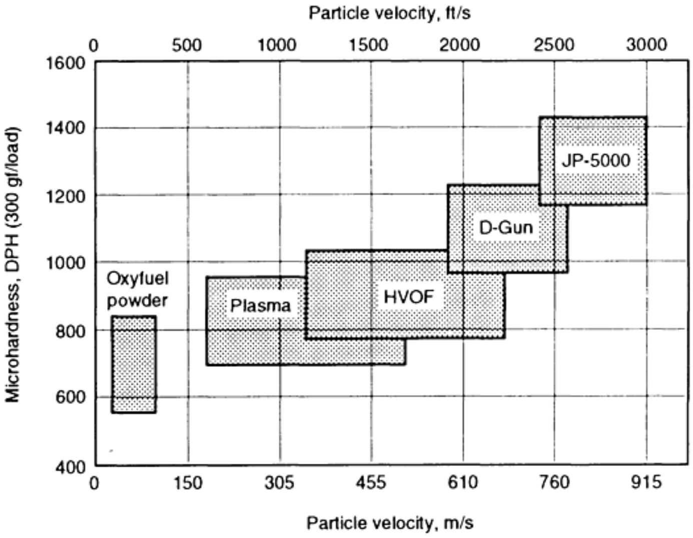

FIG. 22 HARDNESS OF WC-CO COATING VERSUS THERMAL SPRAY PROCESS PARTICLE VELOCITY. JP-5000 IS A REGISTERED TRADENAME OF HOBART TAFA TECHNOLOGIES, INC. SOURCE: REF 22

Surface Preparation (Ref 20)

Thermal spray coatings rely primarily on mechanical bonding to the substrate surface. Because bonding relies on this mechanical interlocking of spray particles with the substrate surface, preparation of the surface is critical to a successful coating. Preparation of the substrate before spraying is virtually the same for all thermal spray processes. More detailed information on surface preparation can be found in Ref 23.

Chemical Cleaning. Chemical cleaning is used on parts that are contaminated or impregnated with material that cannot be removed by other methods. Surfaces can be chemically cleaned by vapor degreasing, steam, hot detergent washing, or application of an industrial cleaning solvent. Degreasing is usually the most economical and a fast way to remove lubricants and oils.

Grit blasting is the most frequently used surface preparation procedure. The roughening cleans the surface of contamination that may inhibit bonding and creates an irregular profile of minute surface irregularities. This enhances the adhesion of the coating. Before application of the thermal spray coating, the cleaned substrate should be blasted to a white-metal condition (according to Steel Structures Painting Council specification SSPC-SP5-82) with a minimum surface profile of 6 μm (250 μm) R_a (arithmetic average surface roughness). Sharp angular grit (18 to 24 mesh average) should be used. Typical-grit blasting materials include chilled iron grit, alumina (Al_2O_3), silica sand, crushed flint, crashed garnet, silicon carbide, and crushed slag.

Rough threading is often used in conjunction with grit blasting for cylindrical surfaces. It consists essentially of machining a thread form on the area to be sprayed. This type of preparation provides excellent adhesion of the coating to the substrate.

Bond coats are often used when an extremely high bond strength is desired or when mechanical deformation of the substrate is not possible or practical. This involves spraying materials that are exothermic and self-bonding. Nickel-aluminum alloys (Ni-5 to 20Al) and molybdenum are the most commonly used bond materials. A thin coating (normally about 50 to 125 μm, or 2 to 5 mils) of bond material is generally applied before the thermally sprayed overcoat. Care should be taken to ensure that the bond material is compatible with the overcoat and that a galvanic cell will not be created, which may be detrimental to the performance of the coating system.

Preheating of the substrate is often beneficial. Preheating to 65 to 95 °C (150 to 200 °F) will eliminate surface condensation and will reduce shrink differentials between the coating and the substrate. It is generally employed only with the flame spray processes, which use their flame as the heat source for warming the substrate.

Finishing Treatment

Thermal spray coatings usually exhibit two common features in the as-deposited condition, a sandpaper-like surface finish and a structure with inherent porosity. A typical surface roughness of an as-sprayed coating ranges from 5.0 to 13 μm (200 to 500 μin.) R_a. The porosity usually ranges from 2 to 17 vol%, depending on the process by which the coating is deposited and the material sprayed. In many cases, thermal spray coatings are applied to machine elements where the coated component must conform to close dimensional tolerances or have a high surface finish. In many applications, coatings are exposed to corrosive solutions or hydraulic fluids, which can infiltrate the pores, resulting in fluid leakage or corrosion of the base material. These conditions can contribute to the premature failure of the coating. Many applications require the coating to be sealed before finishing by machining or grinding, to protect the thermal spray coating.

Sealing is the process by which the pores of a coating are filled to eliminate the possibility of infiltration by fluids or corrosive media that can contribute to premature failure. Sealing materials, which may be applied by brushing, spraying, or dipping, include petroleum-based waxes and organic materials such as phenolics, epoxy resins, polyesters, polyurethanes, and polyimides.

Coating Finishing Method. Thermal spray coatings are finished to dimension either by single-point machining or by grinding. Hard surfacing materials such as ceramics and carbides are restricted to grind finishing, typically with diamond wheels. Thermal spray coatings are sufficiently different from the same material in wrought form that different grinding wheel and finishing tool recommendations are almost always required. Therefore, tools and wheels should not be chosen based on experience with the parent material in wrought or cast form.

Most thermal spray coatings exhibit some degree of porosity. However, the HVOF method produces coatings closest in machining and grinding behavior to wrought materials. In general, thermal spray coatings can be machined to 1 to 2 μm (40 to 80 μin.) R_a and ground to 0.25 to 0.5 μm (10 to 20 μin.) R_a. Many plasma spray ceramic coatings can be lapped down to 0.025 to 0.05 μm (1 to 2 μin.) R_a.

References cited in this section

1. K.C. ANTONY *ET AL.*, HARDFACING, *WELDING, BRAZING, AND SOLDERING,* 9TH ED., VOL 6, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1983, P 771-793
9. K.M. KULKARNI AND V. ANAND, METAL POWDERS USED FOR HARDFACING, *POWDER METALLURGY,* 9TH ED., VOL 7, *ASM HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1984, P 823-836
19. B.A. KUSHNER AND E.R. NOVINSKI, THERMAL SPRAY COATINGS, *FRICTION, LUBRICATION, AND WEAR TECHNOLOGY,* VOL 18, *ASM HANDBOOK,* ASM INTERNATIONAL, 1992, P 829-833
20. R.H. UNGER, THERMAL SPRAY COATINGS, *CORROSION*, 9TH ED., VOL 13, *ASM HANDBOOK,* ASM INTERNATIONAL, 1987, P 459-462
21. M.L. THORPE, THERMAL SPRAY: INDUSTRY IN TRANSITION, *ADV. MATER. PROCESS.,* VOL 143 (NO. 5), 1993, P 50-56
22. G. IRONS, "HIGHER VELOCITY THERMAL SPRAY PROCESSES PRODUCE BETTER AIRCRAFT ENGINE COATINGS," SAE TECHNICAL PAPER 920947, PRESENTED AT THE 28TH ANNUAL AEROSPACE/AIRLINE PLATING AND METAL FINISHING FORUM AND EXPOSITION (SAN DIEGO, CA), APRIL 1992
23. F.N. LONGO, PROCESSING AND DESIGN, LESSON 3, COURSE 51, *THERMAL SPRAY TECHNOLOGY,* MATERIALS ENGINEERING INSTITUTE, ASM INTERNATIONAL, 1992

Hardfacing, Weld Cladding, and Dissimilar Metal Joining

J.R. Davis, Davis & Associates

Thermal Spray Materials and Applications

The most outstanding feature of thermal spray coatings is probably their diverse applicability. There are two primary reasons for this. First, materials selection is almost unlimited. In fact, more than 200 coating materials with different characteristics of toughness, coefficient of friction, hardness, and other properties are available. These materials can be grouped as follows:

- PURE METALS
- METAL ALLOYS
- CERMETS
- CERAMICS
- CARBIDES
- POLYMERS
- SPECIAL COMPOSITE MATERIALS SUCH AS NICKEL-ALUMINUM AND NICKEL-GRAPHITE COMPOSITE POWDERS

Second, the thermal spray process, properly controlled, imparts very little heat to the substrate (100 to 260 °C, or 200 to 500 °F), avoiding metallurgical change, distortion, and oxidation. This allows thermal spray coatings to be applied to practically any substrate (metals, plastics, composites, etc.). Also, such coatings can often be applied to the finish machined part, a practice unthinkable with welding, heat treating, or other high-temperature processes.

Thermal spray coatings provide the design engineer with many options. They offer material properties not available in wrought metals. Coatings can be applied to selected areas rather than treating the entire part. Manufacturing costs can be reduced by eliminating unnecessary processing steps. Extraordinary performance characteristics can be designed into a part to extend its useful life, creating a new, marketable product in the process.

Thermal spray coatings provide the solution to many mechanical, electrical, and corrosion resistance problems involving metal parts and assemblies. However, there are certain applications where such coatings should not be used. Before a thermal spray coating is specified, its suitability can usually be determined according to these criteria (Ref 24):

- NO STRENGTH IS IMPARTED TO THE BASE MATERIAL BY A SPRAYED DEPOSIT. THE COMPONENT TO BE SPRAYED MUST, IN ITS PREPARED FORM, BE ABLE TO WITHSTAND ANY MECHANICAL LOADING THAT WILL BE EXPERIENCED IN SERVICE. (IN A FEW APPLICATIONS, SOME STRENGTH CAN BE ADDED BY A THERMAL SPRAY COATING; HOWEVER, SUCH USES ARE UNUSUAL AND SHOULD BE CAREFULLY TESTED.)
- IF THE AREA ON A PART TO BE SPRAYED OR ANY SECTION OF THE TOTAL AREA WILL BE SUBJECTED TO SHEAR LOADING IN SERVICE, THE PART IS NOT A SUITABLE CANDIDATE FOR THERMAL SPRAYING. GEAR TEETH, SPLINES, AND THREADS ARE EXAMPLES.
- POINT LOADING WITH LINE CONTACT ON A SPRAYED METAL DEPOSIT WILL EVENTUALLY SPREAD THE DEPOSIT, CAUSING DETACHMENT. IF THE DEPOSIT IS ON A MOVING COMPONENT WITH SUCH LOADING, IT WILL FAIL RAPIDLY. FOR EXAMPLE, NEEDLE AND ROLLER BEARING SEATS, WHERE THE BEARING ELEMENTS ARE IN DIRECT CONTACT WITH THE SPRAYED DEPOSIT, MAY NOT BE GOOD THERMAL SPRAY CANDIDATES.
- IF THE BASE METAL OF A COMPONENT TO BE TREATED HAS BEEN NITRIDED, NON-HVOF THERMAL SPRAYING IS NOT RECOMMENDED UNLESS THE NITRIDING IS FIRST REMOVED. OTHER FORMS OF SUBSTRATE HARDENING REQUIRE SPECIAL TREATMENT, SUCH AS INTENSIVE GRIT BLASTING. THERMAL SPRAYING OF NITRIDED TURBINE ENGINE COMPONENTS USING HVOF METHODS IS DESCRIBED IN REF 25.

To date, the most common application area for thermal spraying has been advanced aircraft gas turbine components such as compressor blades, compressor stator vanes, beating housings, and labyrinth seals. Common coating materials include cobalt-base (Laves phase) alloys T-400 and T-800, tungsten carbide-cobalt materials, and chromium carbide-nickel chromium cermets (Cr_3C_2-20 to 30% NiCr). Other key application areas are listed below.

Wear coatings are used to resist abrasion, erosion, cavitation, and fretting, and to reduce friction. These coatings consist of a wide range of metals and their alloys, ceramics, cermets, carbides, and even low-friction plastics. Typical coating hardness ranges from 20 to 70 HRC. Metal matrices are hardened by rapid solidification, by dispersions introduced from the spray process, by the addition of separate hard phases (for example, carbides of chromium, tungsten, titanium and/or tantalum), or by oxide inclusions. Table 13 lists friction and wear applications for various thermal spray materials.

TABLE 13 THERMAL SPRAY COATINGS USED FOR HARDFACING APPLICATIONS

TYPE OF WEAR	COATING MATERIAL	COATING PROCESS(A)	APPLICATIONS
ADHESIVE WEAR	SOFT BEARING COATINGS:		
	ALUMINUM BRONZE	OFW, EAW, OFP, PA, HVOF	BABBITT BEARINGS, HYDRAULIC PRESS SLEEVES, THRUST BEARING SHOES, PISTON GUIDES, COMPRESSOR CROSSHEAD SLIPPERS
	TOBIN BRONZE	OFW, EAW	
	BABBITT	OFW, EAW, OFP	
	TIN	OFW, EAW, OFP	
	HARD BEARING COATINGS:		
	MO/NI-CR-B-SI BLEND	PA	BUMPER CRANKSHAFTS FOR PUNCH PRESS, SUGAR CANE GRINDING ROLL JOURNALS, ANTIGALLING SLEEVES, RUDDER BEARINGS, IMPELLER SHAFTS, PINION GEAR JOURNALS, PISTON RINGS (INTERNAL COMBUSTION); FUEL PUMP ROTORS
	MOLYBDENUM	OFW, EAW, PA	
	HIGH-CARBON STEEL	OFW, EAW	
	ALUMINA/TITANIA	OFP, PA	
	TUNGSTEN CARBIDE	OFP, PA, HVOF	
CO-MO-CR-SI	PA, HVOF		
FE-MO-C	PA		
ABRASIVE WEAR	ALUMINUM OXIDE	PA	SLUSH-PUMP PISTON RODS, POLISH ROD LINERS, AND SUCKER ROD COUPLINGS (OIL INDUSTRY); CONCRETE MIXER SCREW CONVEYORS; GRINDING HAMMERS (TOBACCO INDUSTRY); CORE MANDRELS (DRY-CELL BATTERIES); BUFFING AND POLISHING FIXTURES; FUEL-ROD MANDRELS
	CHROMIUM OXIDE	PA	
	TUNGSTEN CARBIDE	PA, HVOF	
	CHROMIUM CARBIDE	PA, HVOF	
	NI-CR-B-SIC/WC (FUSED)	OFP, HVOF	
	NI-CR-B-SIC (FUSED)	OFP, HVOF	
	NI-CR-B-SIC (UNFUSED)	HVOF	
SURFACE FATIGUE WEAR			
FRETTING:	MOLYBDENUM	OFW, PA	SERVOMOTOR SHAFTS, LATHE

			COMBUSTION), CYLINDER LINERS
FRETTING: SMALL-AMPLITUDE OSCILLATORY DISPLACEMENT APPLICATIONS:			
LOW TEMPERATURE (<540 °C, OR 1000 °F)	ALUMINUM BRONZE	OFW, EAW, PA, HVOF	AIRCRAFT FLAP TRACKS (AIRFRAME COMPONENT); EXPANSION JOINTS AND MID-SPAN SUPPORTS (JET ENGINE COMPONENTS)
	CU-NI-IN	PA, HVOF	
	CU-NI	PA, HVOF	
HIGH TEMPERATURE (>540 °C, OR 1000 °F)	CO-CR-NI-W	PA, HVOF	COMPRESSOR AIR SEALS, COMPRESSOR STATORS, FAN DUCT SEGMENTS AND STIFFENERS (ALL JET ENGINE COMPONENTS)
	CHROMIUM CARBIDE	PA, HVOF	
EROSION	CHROMIUM CARBIDE	PA, HVOF	EXHAUST FANS, HYDROELECTRIC VALVES, CYCLONE DUST COLLECTOR, DUMP VALVE PLUGS AND SEATS, EXHAUST VALVE SEATS
	TUNGSTEN CARBIDE	PA, HVOF	
	WC/NI-CR-B-SIC (FUSED)	OFP, HVOF	
	WC/NI-CR-B-SIC (UNFUSED)	OFP, HVOF	
	CHROMIUM OXIDE	PA	
CAVITATION	NI-CR-B-SIC-AL-MO	PA	WEAR RINGS (HYDRAULIC TURBINES), WATER TURBINE BUCKETS, WATER TURBINE NOZZLES, DIESEL ENGINE CYLINDER LINES PUMPS
	NI-AL/NI-CR-B-SIC	PA	
	TYPE 316 STAINLESS STEEL	PA	
	NI-CR-B-SIC (FUSED)	OFP, HVOF	
	NI-CR-B-SIC (UNFUSED)	HVOF	
	ALUMINUM BRONZE	PA, HVOF	
	CU-NI	PA, HVOF	

(A) OFW, OXYFUEL WIRE SPRAY; EAW, ELECTRIC ARC WIRE SPRAY; OFP, OXYFUEL POWDER SPRAY; PA, PLASMA ARC SPRAY; HVOF, HIGH-VELOCITY OXYFUEL POWDER SPRAY

Corrosion Resistance. Zinc and aluminum and their alloys are the metals most widely used for thermal spray anticorrosion coatings. They are extensively used for the corrosion protection of iron and steel in a wide range of environments and have been shown to provide very long-term protection (over 20 years) in both marine and industrial locations. Properties and applications of zinc and aluminum thermal spray coatings are described in the article "Thermal Spraying Coatings" in this Volume.

Other coating materials for corrosion-resistant applications include austenitic stainless steels, aluminum bronze, nickel-base alloys, and MCrAlY materials. Selection of a specific alloy depends on the particular environment. When using these alloys, it must be understood that these coatings will not galvanically protect the underlying steel. This can be a particular problem due to the porosity of thermal spray coatings. Great care must be taken to ensure that these coatings are properly sealed in order to prevent penetration of the corrosive medium to the underlying steel and subsequent corrosion at the coating/substrate interface.

Oxidation Protection. Thermal spray coatings are extensively used by industry to protect steel components and structures from heat oxidation at surface temperatures to 1095 °C (2000 °F). By ensuring long-term protection, thermal spray coatings show real economic advantages during the service lives of such items. Coatings are particularly effective in protecting low-alloy and carbon steels. Table 14 lists thermal spray coatings for use in high-temperature applications.

TABLE 14 THERMAL SPRAY COATINGS USED FOR ELEVATED-TEMPERATURE SERVICE

SERVICE TEMPERATURE	COATING METAL OR ALLOY	COATING THICKNESS	
		µm	mils
UP TO 550 °C (1020 °F)	ALUMINUM	175	7
UP TO 550 °C (1020 °F) IN THE PRESENCE OF SULFUROUS GASES	NI-43CR-2TI	375	15
550-900 °C (1020-1650 °F)	ALUMINUM OR ALUMINUM-IRON	175	7
900-1000 °C (1650-1830 °F)	NICKEL-CHROMIUM OR MCRALY	375	15
900-1000 °C (1650-1830 °F) IN THE PRESENCE OF SULFUROUS GASES	NICKEL-CHROMIUM, FOLLOWED BY ALUMINUM	375	15
		100	4

Source: Ref 20

Aluminum has been widely used to protect such steels in a number of applications involving high surface temperatures up to 550 °C (1020 °F). Nickel-chromium alloys and some of the MCrAlY alloys have also provided protection in severe environments (up to 1000 °C, or 1830 °F).

Thermal insulation provides low-thermal-conductivity zirconium and/or aluminum oxides that are deposited as coatings to reduce heat conduction to base metals. These insulating oxide coatings can also reduce heat loss, reduce thermal transient effects, and reduce the high-temperature oxidation/corrosion effects by lowering temperatures of the coated base metals. Thermal spray processes usually introduce a controlled level of porosity, which further enhances the coating insulation and increases its thermal shock resistance.

Thermal barrier coatings consist of a low-conductivity (thermal) ceramic deposited over an MCrAlY bond coat. The ceramic of choice is partially-stabilized zirconia (7 to 8 wt% Y_2O_3-ZrO_2) deposited at a thickness of 0.25 mm (0.010 to 0.040 in.) with 10 to 15% porosity. A bond coat of NiCrAlY or CoCrAlY is used at a thickness of 0.125 mm (0.005 in.). Both coatings are used on components used in gas turbine engines and adiabatic engines to improve efficiency and reduce metal temperatures or cooling requirements.

Electrically conductive coatings include silver, copper, aluminum, zinc, and bronze alloys. Typical conductivities of these sprayed metals are reported to be 40 to 90% of wrought materials, depending on the spray coating method. In addition, superconducting plasma-sprayed Y-Ba-Cu-O, Cu/Nb_3Sn, and Nb_3Sn materials that have zero resistance at low to intermediate temperatures of 100 K (-280 °F) have been deposited by thermal spraying. Conductive coatings designed to shield electronic components against radio frequency or electromagnetic interference such as aluminum, zinc, and tin coatings, are also sprayed.

Dimensional Restorative Coatings. Thermal spray is used as a coating to repair or resurface (up to 3 mm, or 0.12 in. thick) worn and/or corroded parts, as well as to repair new part errors (for example, undersize parts), to repair nicks or other blemishes, or to renew surfaces that are corroded or worn in use.

Medically Compatible Coatings. Interactive bonding coatings on implants use thermal spray coatings to provide biocompatible titanium (commercially pure Ti or Ti-6Al-4V alloy) and/or ceramic (calcium phosphates) surfaces on both dental and orthopedic devices, to improve wear surface and/or tissue adhesion. Coating porosity and high surface roughness can be produced in thermal spray coatings to improve further tissue adherence.

References cited in this section

20. R.H. UNGER, THERMAL SPRAY COATINGS, *CORROSION*, 9TH ED., VOL 13, *ASM HANDBOOK*, ASM INTERNATIONAL, 1987, P 459-462
24. M.L. THORPE, THERMAL SPRAYING BECOMES A DESIGN TOOL, *MACH. DES.*, 24 NOV 1983
25. P.O. BUZBY AND J. NIKITICH, HVOF THERMAL SPRAYING OF NITRIDED PARTS, *ADV. MATER. PROCESS.*, DEC 1991, P 35-36

Hardfacing, Weld Cladding, and Dissimilar Metal Joining

J.R. Davis, Davis & Associates

Weld Cladding

The term weld cladding usually denotes the application of a relatively thick layer ([ges]3 mm, or $\frac{1}{8}$ in.) of weld metal for the purpose of providing a corrosion-resistant surface. Hardfacing, as described earlier in this article, produces a thinner surface coating than a weld cladding and is normally applied for dimensional restoration or wear resistance. Typical base metal components that are weld-cladded include the internal surfaces of carbon and low-alloy steel pressure vessels, paper digesters, urea reactors, tubesheets, nuclear reactor containment vessels, and hydrocrackers. The cladding material is usually an austenitic stainless steel or a nickel-base alloy, although certain copper-base alloys are also sometimes used. Some specialized cladding is also carried out using silver filler metal. Weld cladding is usually performed using the submerged arc welding process (SAW). However, flux-cored (either self-shielded or gas-shielded), plasma arc, and electroslag welding methods can also produce weld claddings. Filler metals are available as covered electrodes, coiled electrode wire, and strip electrodes. Table 15 lists some of the filler metals for stainless steel weld claddings.

TABLE 15 STAINLESS STEEL FILLER METALS FOR WELD CLADDING APPLICATIONS

WELD OVERLAY TYPE	FIRST LAYER		SUBSEQUENT LAYERS	
	COVERED ELECTRODE(A)	BARE ROD OR ELECTRODE(B)	COVERED ELECTRODE(A)	BARE ROD OR ELECTRODE(B)
304	E309	ER309	E308	ER308
304L	E309L	ER309L	E308L	ER308L
	E309CB			
321	E309CB	ER309CB	E347	ER347
347	E309CB	ER309CB	E347	ER347
309	E309	ER309	E309	ER309
310	E310	ER310	E310	ER310
316	E309MO	ER309MO	E316	ER316
316L	E309MOL	ER309MOL	E316L	ER316L
	E317L	ER317L		
317	E309MO	ER309MO	E317	ER317
	E317	ER317		
317L	E309MOL	ER309MOL	E317L	ER317L
	E317L	ER317L		
20 CB	E320	ER320	E320	ER320

Note: Columbium (Cb) is also referred to as niobium (Nb).

(A) REFER TO AWS SPECIFICATION A5.4.

(B) REFER TO AWS SPECIFICATION A5.9.

Application Considerations (Ref 26)

The technique of weld cladding is an excellent method to impart properties to the surface of a substrate that are not available from that base metal, or to conserve expensive or difficult-to-obtain materials by using only a relatively thin surface layer on a less expensive or abundant base material. This technique has several inherent limitations or possible problems that must be considered when planning for weld cladding. The thickness of the required surface must be less than the maximum thickness of the overlay that can be obtained with the particular process and filler metal selected.

Welding position also must be considered when selecting an overlay material and process. Certain processes are limited in their available welding positions; for example, SAW can be used only in the flat position. In addition, when using a high-deposition-rate process that exhibits a large liquid pool, welding vertically or overhead may be difficult or impossible. Some alloys exhibit eutectic solidification, which leads to large molten pools that solidify instantly rather than with a "mushy" (liquid plus solid) transition. Such materials are also difficult to weld except in the flat position.

Composition Control of Stainless Steel Weld Overlays (Ref 26)

The economics of stainless steel weld cladding are dependent on achieving the specific chemistry at the highest practical deposition rate in a minimum number of layers. The fabricator selects the filler wire and welding process, whereas the purchaser specifies the surface chemistry and thickness, along with the base metal. The most outstanding difference between welding a joint and depositing an overlay is in the area of dilution. The percentage of dilution equals the amount of base metal melted (*x*) divided by the sum of base metal melted and filler metal added (*x* + *y*), the quotient of which is multiplied by 100.

$$\% dilution = \frac{x}{x+y} x100 \qquad \textbf{(EQ 2)}$$

For stainless steel cladding, a fabricator must understand how the dilution of the filler metal with the base metal affects the composition and metallurgical balance, such as the proper ferrite level to minimize hot cracking, absence of martensite at the interface for bond integrity, and carbon at a low level to ensure corrosion resistance. The prediction of the microstructures and properties (such as hot cracking and corrosion resistance) for the austenitic stainless steels has been the topic of many studies (Ref 27). During the last two decades, four microstructure prediction diagrams have found the widest application. These include the Schaeffler diagram, the DeLong diagram, and the Welding Research Council (WRC) diagrams (WRC-1988 and WRC-1992). Each of these is described below.

The Schaeffler diagram (Fig. 23), published in 1949, has been extensively used for estimating the ferrite content of stainless steel weld metals (Ref 28). The diagram contains phase fields and iso-ferrite lines that permit prediction of the weld structure from the composition. It is still widely used for predicting the ferrite content of dissimilar weld deposits. This involves calculating a "chromium equivalent" (Cr_{eq}) and a "nickel equivalent" (Ni_{eq}) for each base metal and for the proposed filler metal, plotting each equivalent on the Schaeffler diagram, and drawing tie lines between the plotted points, proportioned according to expected dilution, to obtain an estimate of the weld metal ferrite content.

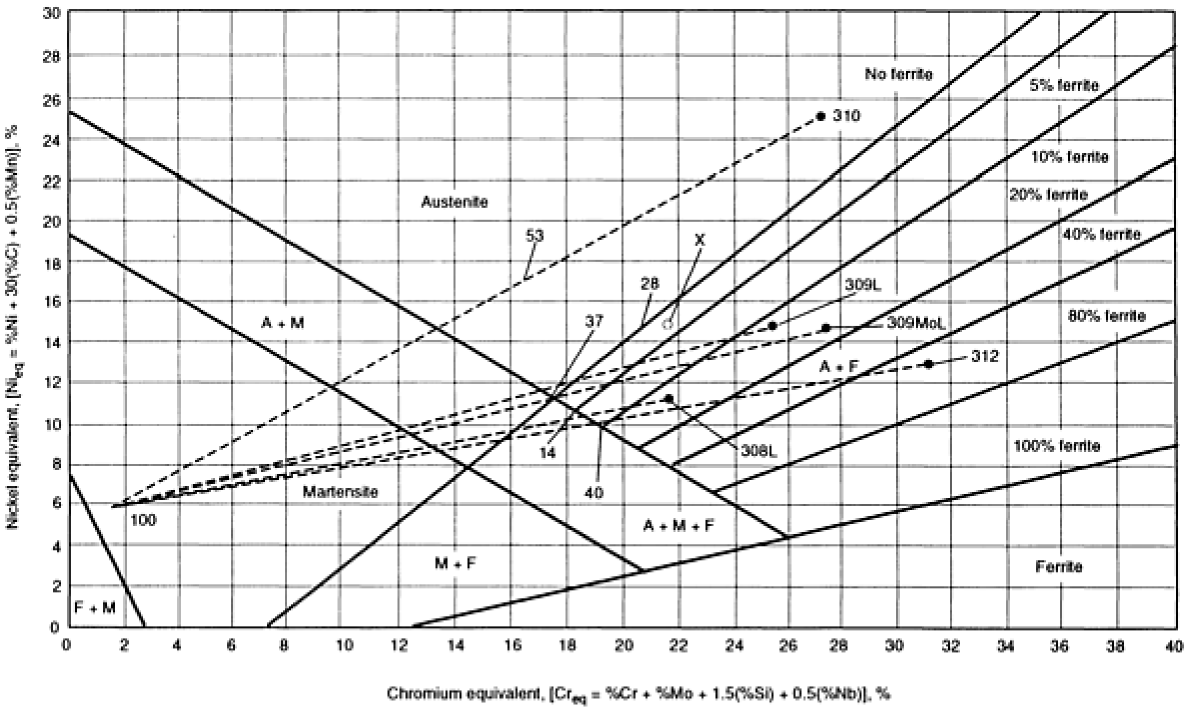

FIG. 23 SCHAEFFLER DIAGRAM FOR STAINLESS STEEL WELD METAL. A, AUSTENITE; F, FERRITE; M, MARTENSITE. POINT 100 IS ASTM A 387A STEEL. EXAMPLE: POINT X ON THE DIAGRAM INDICATES THE EQUIVALENT COMPOSITION OF TYPE 318 (316 CB) WELD DEPOSIT CONTAINING 0.07% C, 1.55% MN, 0.57% SI, 19.02% CR, 11.87% NI, 2.16% MO, AND 0.80% NB. EACH OF THESE PERCENTAGES WAS MULTIPLIED BY THE "POTENCY FACTOR" INDICATED FOR THE ELEMENT IN QUESTION ALONG THE AXES OF THE DIAGRAM TO DETERMINE THE CHROMIUM AND THE NICKEL EQUIVALENT. WHEN THESE WERE PLOTTED AS POINT X, THE CONSTITUTION OF THE WELD WAS INDICATED AS AUSTENITE PLUS FROM 0 TO 5% FERRITE; MAGNETIC ANALYSIS OF THE ACTUAL SAMPLE REVEALED AN AVERAGE FERRITE CONTENT OF 2%. FOR AUSTENITE-PLUS-FERRITE (A+F) STRUCTURES, THE DIAGRAM PREDICTS THE PERCENTAGE OF FERRITE WITHIN 4% FOR THE FOLLOWING STAINLESS STEELS: 308, 309, 309 CB, 310, 312, 316, 317, 318 (316 CB), AND 347. SOURCE: REF 26

As an example of the use of the Schaeffler diagram, suppose a type 304L clad surface is specified (shown as a triangle in Fig. 23). A desirable composition is approximately 20% Cr equivalent and 11% Ni equivalent. If an ASTM A 387A plate (point 100 on the diagram) is specified, all combinations of weld deposit lie on a line connecting point 100 with a proposed filler metal. As an example, note the line connecting point 100 with type 310 stainless steel. The diagram shows that weld-deposit compositions with up to 53% dilution (53% of the length of the line connecting the base metal and the filler metal) will be fully austenitic. Although sufficient alloy might be present, at least for a first layer, the absence of ferrite may make the deposit sensitive to fissuring. The same procedure can be used to estimate the usefulness of other filler metals.

Good metallurgical practice dictates that the first layer contain no martensite, shown as M in the diagram. This limits the dilution. If a process can be found that has less than 14% dilution, then a type 308L could be deposited with absence of martensite. On the other hand, if a first layer is made using type 312 stainless, any dilution factor less than 40% prevents the formation of martensite. Choosing a filler metal for the first layer that falls in the A + F region with 3 to 15% ferrite and no martensite meets the metallurgical requirements of a sound deposit. Proper composition and thickness may be achieved in a single layer, but more often a second layer is required.

Special consideration must be given to carbon levels for the austenitic stainless steels. Figure 24 predicts the carbon in the deposit once the base metal carbon and cladding process dilution factors have been established. To illustrate its use, suppose a base metal plate with 0.20% C and a welding process that results in 30% dilution are established. The diagram indicates that the deposit chemistry will contain approximately 0.07% C. If the specification requires 0.04% C maximum, a second layer would be required to achieve the necessary carbon level. Using the same diagram with the first layer composition of 0.07% C, and assuming the same dilution factor, the diagram predicts that a second layer will contain approximately 0.03% C. This would meet the specification of 0.04% C maximum.

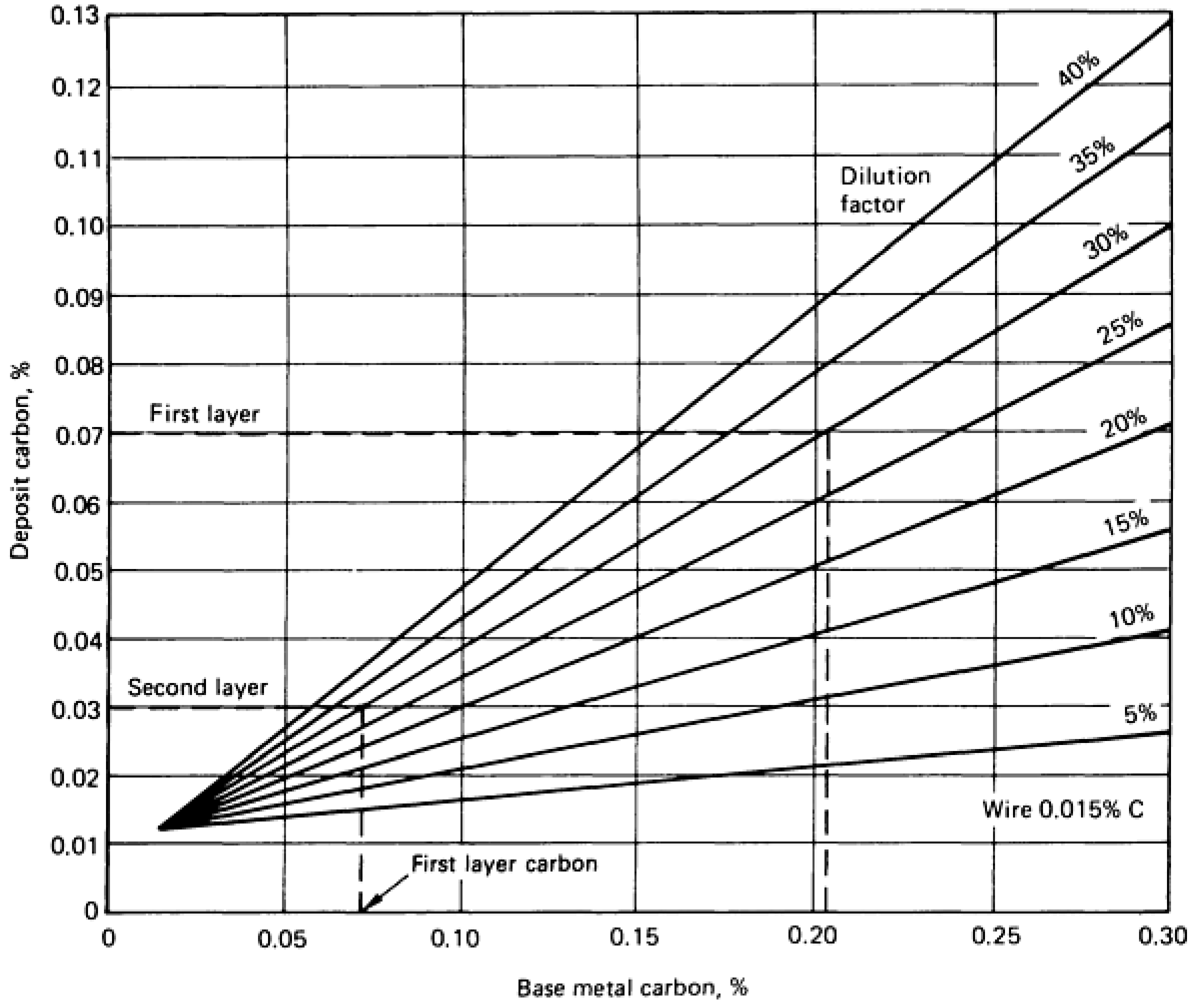

FIG. 24 PERCENTAGE OF CARBON IN THE DEPOSIT AFTER BASE METAL CARBON AND CLADDING DILUTION FACTORS HAVE BEEN ESTABLISHED. SOURCE: REF 26

Several problems have been identified when using the Schaeffler diagram for predicting the physical and mechanical properties of a weld. The Schaeffler diagram does not consider the effect of nitrogen in promoting austenite at the expense of ferrite, and it is incorrect in its treatment of manganese. Manganese does not promote the high-temperature formation of austenite at the expense of ferrite, as predicted by the diagram, although manganese does stabilize austenite in its low-temperature transformation to martensite. Also, the Schaeffler diagram makes its predictions in terms of percent ferrite. This has been found to be imprecise, and the magnetically based Ferrite Number (FN) unit was developed for the specification and determination of ferrite content. The basis for the FN scale is described in the American Welding Society standard AWS 4.2, "Standard Procedures for Calibrating Magnetic Instruments to Measure the Delta Ferrite Content of Austenitic and Duplex Austenitic Ferritic Steel."

The DeLong Diagram. A second widely used prediction diagram, the DeLong diagram (Fig. 25) was published in 1974 and incorporated some improvements (Ref 29). It has an FN scale and includes a coefficient for nitrogen in the Ni_{eq}, though the range of compositions is more limited than for the Schaeffler diagram.

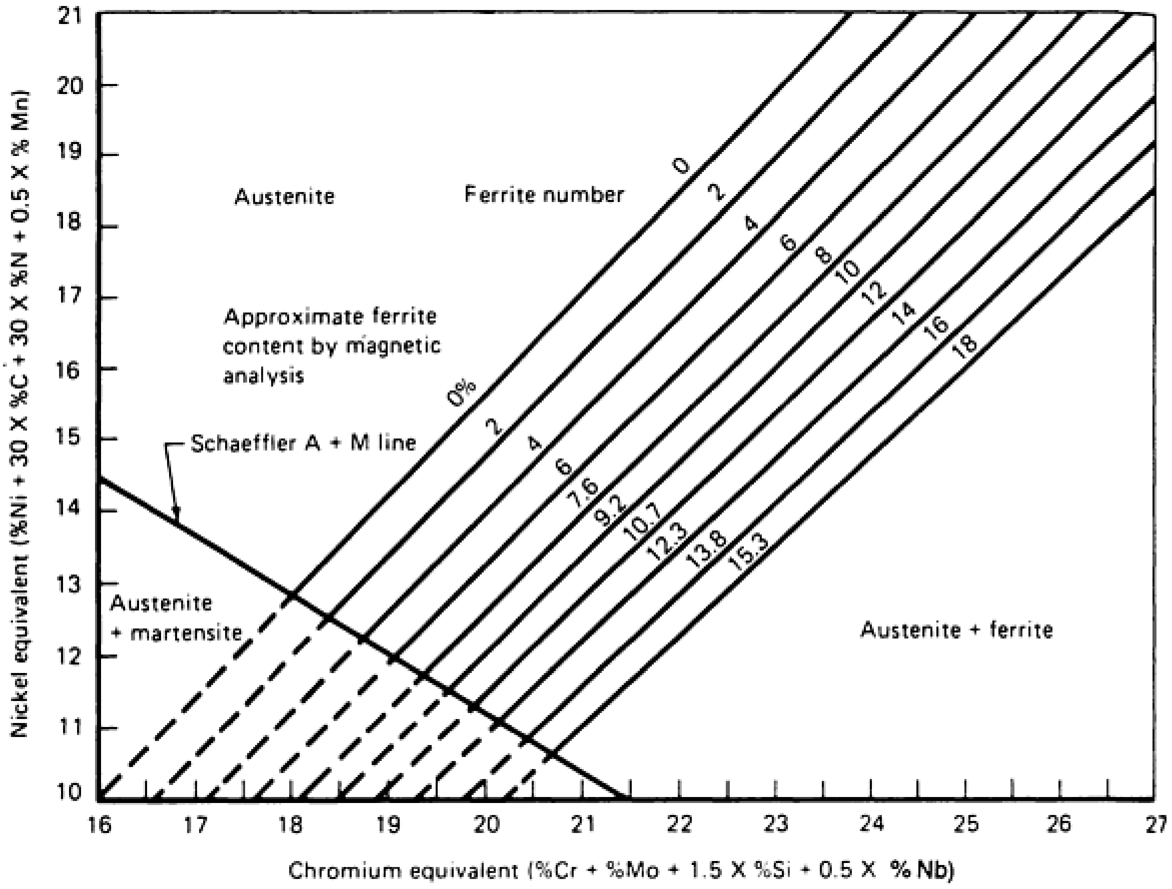

FIG. 25 THE DELONG CONSTITUTION DIAGRAM, REVISED JANUARY 1973 TO CONVERT IT TO THE WELDING RESEARCH COUNCIL'S FERRITE NUMBER (FN) SYSTEM FOR WELD METALS

The difference in the applicability of the Schaeffler and DeLong diagrams is substantial (Ref 30). The DeLong diagram is a finely tuned subset of the Schaeffler range, designed specifically for the 300-series stainless steel welds containing small amounts of ferrite. Its widely accepted use with these alloys is documented by its inclusion in the AWS filler metal specifications A5.4 and A5.9 and in the American Society of Mechanical Engineers (ASME) Boiler and Pressure Vessel Code. Prediction of FN for compositions that fall outside the range of the DeLong diagram (for example, duplex stainless steels) is the most quantitative use of the Schaeffler diagram; however, its extension to fully ferritic compositions and the inclusion of martensitic regions provides much qualitative information about the general effect of alloying elements and characteristic mechanical properties of alloy families.

In effect, the DeLong diagram traded improved prediction accuracy for a more limited composition range, but a range that included all the common austenitic stainless steel weld metals. However, it has been found to incorrectly handle manganese (as does the Schaeffler diagram) and to overestimate the FN of more highly alloyed compositions, such as type 309 (Ref 30). It also has had limited application to dissimilar metal joints.

The WRC-1988 diagram (Fig. 26) overcomes many of the problems associated with the Schaeffler and DeLong diagrams (Ref 30). It was developed with data measured by the most recent definition of the FN scale. It removed the erroneous manganese coefficient from the Ni_{eq}, and it eliminated the systematic overestimation of FN for highly alloyed weld metals. While it covers a much broader range of compositions than does the DeLong diagram, it has a narrower composition range than the Schaeffler diagram, because it extends only over the composition range of the commercial alloys with which it was developed.

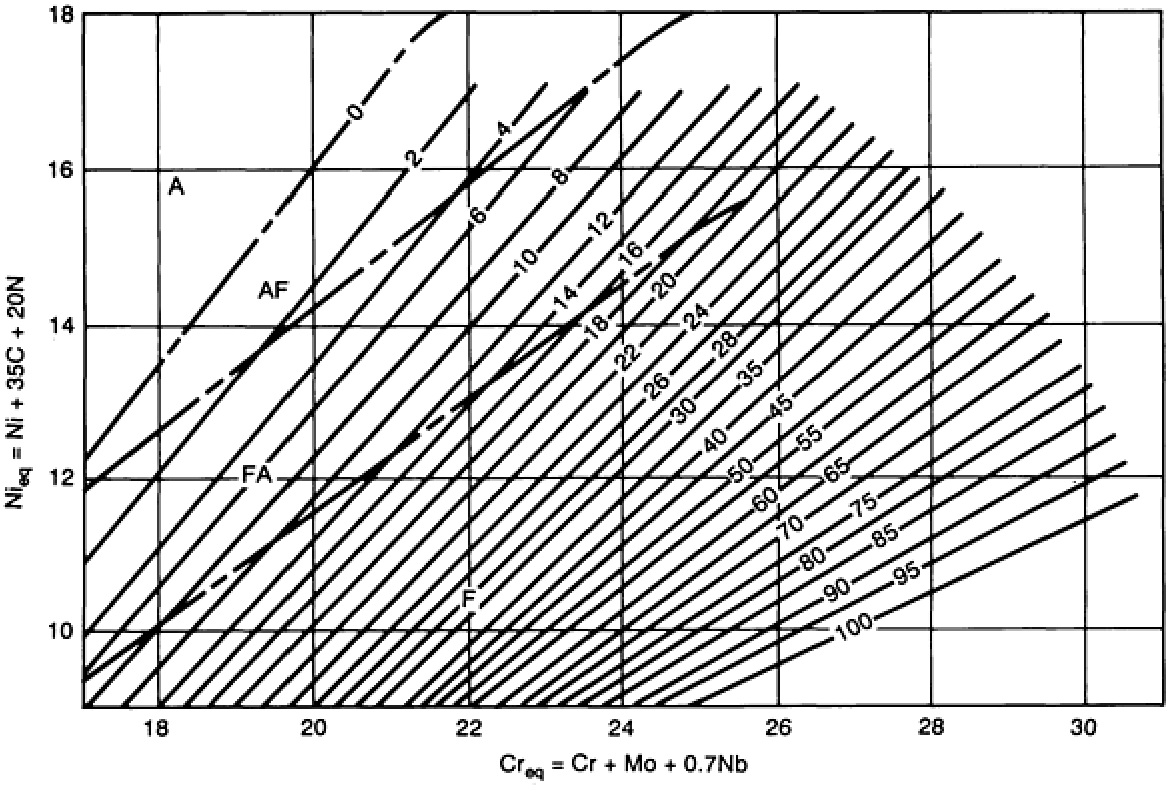

FIG. 26 THE WRC-1988 DIAGRAM, INCLUDING SOLIDIFICATION-MODE BOUNDARIES. SOURCE: REF 30

In the interval since the WRC-1988 diagram first appeared, it has been subjected to considerable evaluation and review. Independent data from more than 200 weld metals have confirmed the improved prediction accuracy of the WRC-1988 over that of its predecessors (Ref 31). The WRC-1988 diagram predicts the ferrite content of the conventional 300-series stainless steels with a higher degree of accuracy than the DeLong diagram, is more accurate than the Schaeffler diagram for predicting the ferrite content of duplex stainless steels, and is suitable for new alloys containing extended compositional ranges. It also includes primary solidification-mode data to reveal relationships to the FN response. Recent studies have emphasized that the solidification mode is more accurate than the FN for predicting the resistance to hot cracking (Ref 30). These studies indicate that the initial solidification structure (primary austenite or ferrite) directly determines the sensitivity to cracking by controlling the solubility of elements that promote cracking at these temperatures.

The WRC-1992 Diagram. In recent years, duplex ferritic-austenitic stainless steels have been used more frequently. Some of these steels and their welds contain significant amounts of copper (often about 2 wt%). While these compositions might be plotted on the WRC-1988 diagram, greater accuracy can be obtained with the modified WRC-1992 diagram (Ref 32). Whereas the Schaeffler diagram offers predictions for Cr_{eq} from 0 to 40 and Ni_{eq} from 0 to 32, the WRC-1992 diagram is more limited, considering Cr_{eq} from 17 to 31 and Ni_{eq} from 9 to 18 (Fig. 27). As indicated in the following example, the WRC-1992 diagram is useful for predicting the FN for dissimilar metal joints.

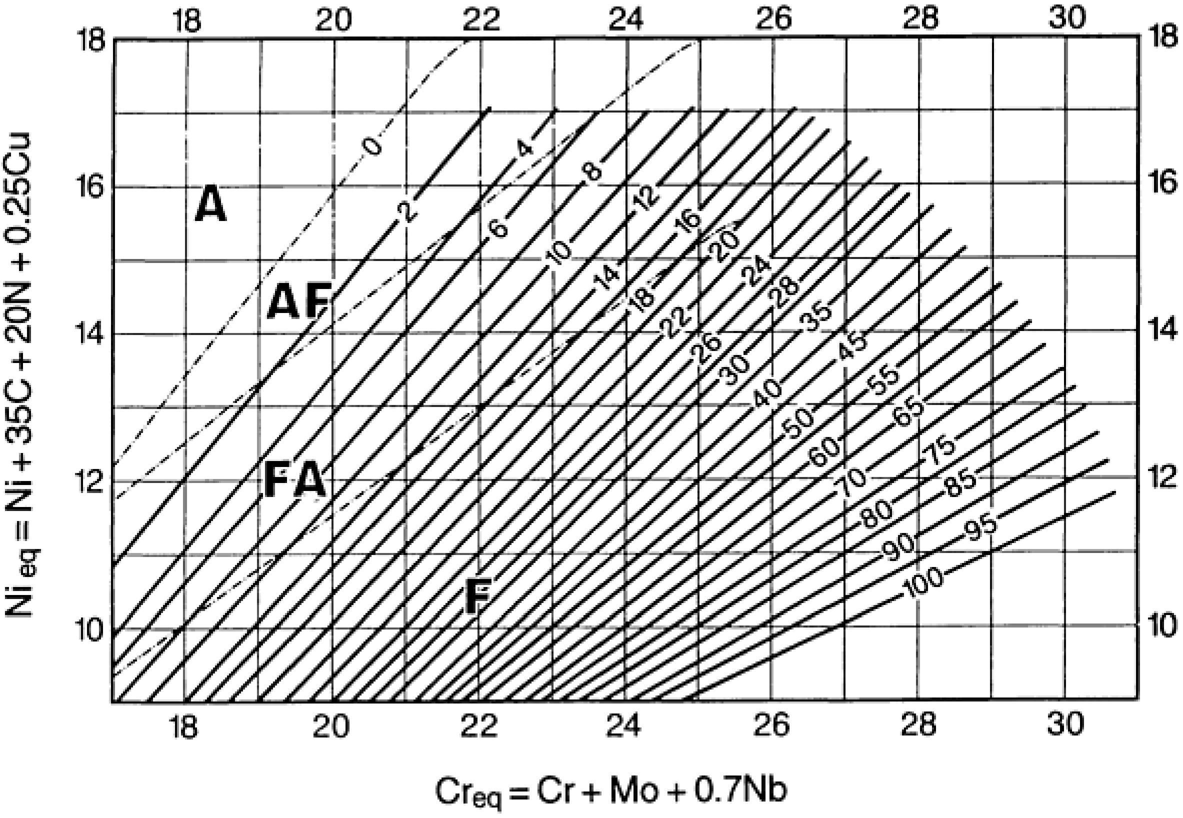

FIG. 27 THE WRC-1992 DIAGRAM. SOURCE: REF 32

Example 1: Weld Cladding of AISI 1050 Steel with AWS A5.4 Class E312-16 Electrode

(Ref 32). Table 16 lists typical plate composition and all-weld-metal composition produced with this electrode. The Cr_{eq} and Ni_{eq} for the AISI 1050 steel (0.0 and 17.50, respectively) do not permit an FN calculation for this material. The Cr_{eq} and Ni_{eq} for the E312-16 electrode all-weld-metal composition (29.0 and 11.9, respectively) result in a predicted 88.2 FN for this material. If the Cr_{eq} and Ni_{eq} for each composition is plotted in Fig. 28 and a line drawn between the point corresponding to the AISI 1050 composition (Point A in Fig. 28) and the point corresponding to the E312-16 composition (Point B in Fig. 28), then all possible mixtures of these two materials must lie along this line.

TABLE 16 BASE METAL AND WELD METAL COMPOSITIONS PRODUCED WHEN AISI 1050 CARBON STEEL WAS WELD CLAD USING E312-16 STAINLESS STEEL FILLER METAL

See Example 1 in text.

MATERIAL	AISI 1050	E312-16 ALL-WELD-METAL	30% DILUTION CLADDING
C, %	0.50	0.060	0.192
MN, %	0.30	1.20	0.93
SI, %	0.02	0.60	0.43
CR, %	. . .	29.00	20.30
NI, %	. . .	8.60	6.02
N, %	0.004	0.06	0.043
CR_{EQ}	0.00	29.00	20.30
NI_{EQ}	17.58	11.90	13.60
WRC-1988 FN	. . .	88.2	4.6

Source: Ref 32

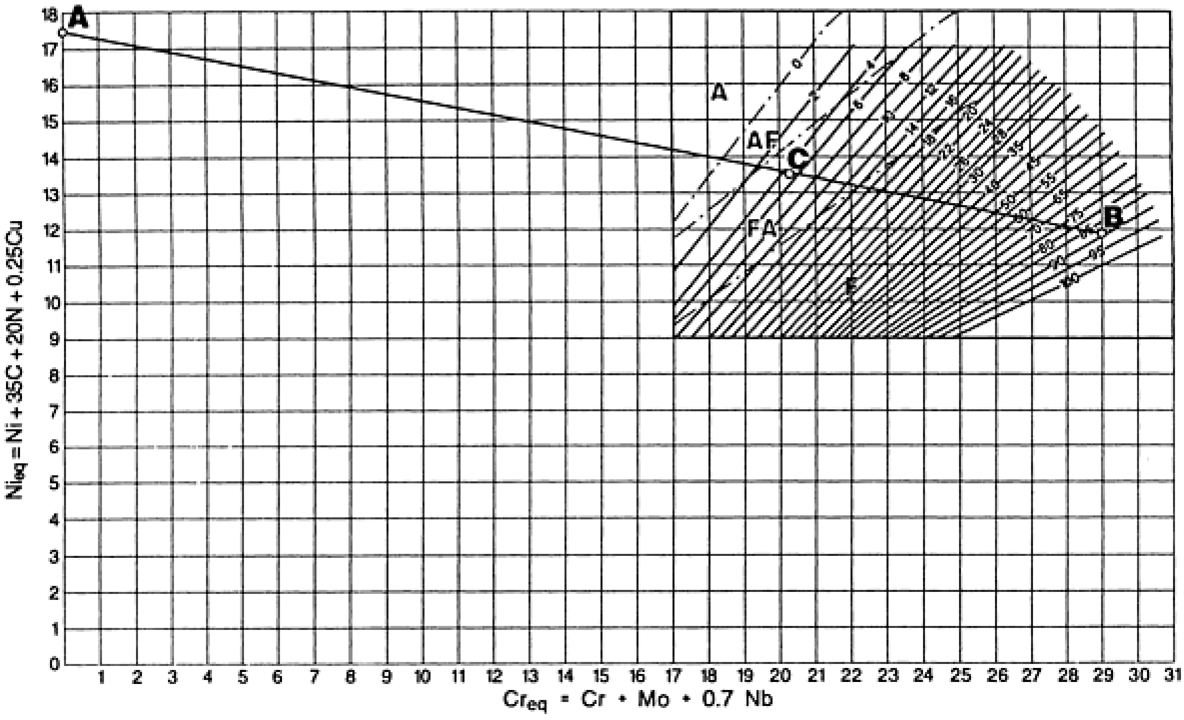

FIG. 28 DILUTION CALCULATION IN EXAMPLE 1 USING THE WRC-1992 DIAGRAM. SOURCE: REF 32

The lever rule is then used to estimate the position of the weld metal. In SMAW, typical dilution is 30%. A weld pass of E312-16 with 30% dilution from AISI 1050 steel would lie at a Point C, located 30% of the distance from Point B to Point A in Fig. 28. Mathematically, this can be calculated as:

$$\begin{aligned} CR_{EQ}(C) &= 0.7\ CR_{EQ}(B) + 0.3\ CR_{EQ}(A) \\ &= 0.7(29.00) + 0.3(0.00) = 20.30 \end{aligned} \quad \textbf{(EQ 3)}$$

$$\begin{aligned} NI_{EQ}(C) &= 0.7\ NI_{EQ}(B) + 0.3\ NI_{EQ}(A) \\ &= 0.7(11.90) + 0.3(17.58) = 13.60 \end{aligned} \quad \textbf{(EQ 4)}$$

The Cr_{eq} and Ni_{eq} of Point C correspond to 4.6 FN, which indicates that the weld pass will have sufficient FN to avoid fissuring (Ref 30). Any additional passes overlapping part of this first pass will also lie along the line between Point B and Point A, but they will be closer to Point B than is Point C, so all subsequent passes will contain even more ferrite than the first pass and the weld cladding should be crack-free.

Control of dilution plays an important part in the economics of the weld cladding process. Although each process has an expected dilution factor, experimenting with the welding parameters can minimize dilution. A value between 10 and 15% is generally considered optimum. Less than 10% raises the question of bond integrity, and greater than 15% increases the cost of the filler metal. Unfortunately, most welding processes have considerably greater dilution.

Because of the importance of dilution in weld cladding as well as hardfacing applications, each welding parameter must be carefully evaluated and recorded. Many of the parameters that affect dilution in weld cladding applications are not so closely controlled when arc welding a joint. These parameters include:

- *AMPERAGE:* INCREASED AMPERAGE (CURRENT DENSITY) INCREASES DILUTION. THE ARC BECOMES HOTTER, IT PENETRATES MORE DEEPLY, AND MORE BASE METAL MELTING OCCURS.
- *POLARITY:* DIRECT CURRENT ELECTRODE NEGATIVE (DCEN) GIVES LESS PENETRATION AND RESULTING LOWER DILUTION THAN DIRECT CURRENT ELECTRODE POSITIVE

(DCEP). ALTERNATING CURRENT RESULTS IN A DILUTION THAT LIES BETWEEN DCEN AND DCEP.

- *ELECTRODE SIZE:* THE SMALLER THE ELECTRODE, THE LOWER THE AMPERAGE, WHICH RESULTS IN LESS DILUTION.
- *ELECTRODE EXTENSION:* A LONG ELECTRODE EXTENSION FOR CONSUMABLE ELECTRODE PROCESSES DECREASES DILUTION. A SHORT ELECTRODE EXTENSION INCREASES DILUTION.
- *TRAVEL SPEED:* A DECREASE IN TRAVEL SPEED DECREASES THE AMOUNT OF BASE METAL MELTED AND INCREASES PROPORTIONALLY THE AMOUNT OF FILLER METAL MELTED, THUS DECREASING DILUTION.
- *OSCILLATION:* GREATER WIDTH OF ELECTRODE OSCILLATION REDUCES DILUTION. THE FREQUENCY OF OSCILLATION ALSO AFFECTS DILUTION--THE HIGHER THE FREQUENCY OF OSCILLATION, THE LOWER THE DILUTION.
- *WELDING POSITION:* DEPENDING ON THE WELDING POSITION OR WORK INCLINATION, GRAVITY CAUSES THE WELD POOL TO RUN AHEAD OF, REMAIN UNDER, OR RUN BEHIND THE ARC. IF THE WELD POOL STAYS AHEAD OF OR UNDER THE ARC, LESS BASE METAL PENETRATION AND RESULTING DILUTION WILL OCCUR. IF THE POOL IS TOO FAR AHEAD OF THE ARC, THERE WILL BE INSUFFICIENT MELTING OF THE SURFACE OF THE BASE METAL, AND COALESCENCE WILL NOT OCCUR.
- *ARC SHIELDING:* THE SHIELDING MEDIUM, GAS OR FLUX, ALSO AFFECTS DILUTION. THE FOLLOWING LIST RANKS VARIOUS SHIELDING MEDIUMS IN ORDER OF DECREASING DILUTION: GRANULAR FLUX WITHOUT ALLOY ADDITION (HIGHEST), HELIUM, CARBON DIOXIDE, ARGON, SELF-SHIELDED FCAW, AND GRANULAR FLUX WITH ALLOY ADDITION (LOWEST).
- *ADDITIONAL FILLER METAL:* EXTRA METAL (NOT INCLUDING THE ELECTRODE), ADDED TO THE WELD POOL AS POWDER, WIRE, STRIP, OR WITH FLUX, REDUCES DILUTION BY INCREASING THE TOTAL AMOUNT OF FILLER METAL AND REDUCING THE AMOUNT OF BASE METAL THAT IS MELTED.

Procedures for Stainless Steel Weld Cladding (Ref 26)

Submerged arc welding is by far the most commonly used process for weld cladding. The process is adaptable for use with single wires as the electrode filler metal, multiple wires, or strip. Alternating current is often used, but direct current, with either reverse or straight polarity, has been found preferable in numerous instances. Oscillation of the filler metal is frequently found desirable. Typical welding parameters for stainless steel weld cladding by SAW are given in Table 17.

TABLE 17 TYPICAL WELDING PARAMETERS FOR STAINLESS STEEL WELD CLADDING BY SAW

FORM OF ELECTRODE	SIZE OF ELECTRODE		POWER SOURCE	CURRENT, A	VOLTAGE, V	TRAVEL SPEED		OSCILLATION		PRODUCTIVITY		DILUTION(A), %
	mm	**in.**				**mm/min**	**in./min**	**mm**	**in.**	**kg/h**	**lb/h**	
SINGLE WIRE	1.6	$\frac{1}{16}$	DCEN(B)	240	34	130	5	19	$\frac{3}{4}$	4	9	15-20
	2.4	$\frac{3}{32}$	DCEN(B)	350	42	150	6	32	$1\frac{1}{4}$	9	20	15-20
	2.4	$\frac{3}{32}$	AC	350	30	280	11	NONE		5.0	11	20-25
	3.2	$\frac{1}{8}$	AC	450	32	280	11	NONE		5.4	12	20-25
	4.0	$\frac{5}{32}$	AC	500	34	280	11	NONE		5.9	13	20-25
SERIES ARC												
TWO WIRES	4.0	$\frac{5}{32}$	AC(B)	400	26	255	10	NONE		19	41	20-25
THREE WIRES	4.0	$\frac{5}{32}$	AC(B)	480	28	355	14	NONE		20	43	15-20
SIX WIRE												
FIRST WIRE	1.6	$\frac{1}{16}$	DCEP	255	27	130	5	16	$\frac{5}{8}$	28	83	15-20
OTHERS	1.6	$\frac{1}{16}$	DCEN	1125	30	130	5	16	$\frac{5}{8}$	28	83	15-20
STRIP	0.5 × 30	0.020 × 1.2	DCEP(B)	400	25	180	7	NONE		35	16	15-20
	0.5 × 61	0.020 × 2.4	DCEP(B)	700	27	180	7	NONE		15	32	15-20
	0.5 × 88	0.020 × 3.5	DCEP(B)	1250	27	130	5	NONE		27	60	15-20
	0.5 × 119	0.020 × 4.7	DCEP(B)	1550	27	130	5	NONE		35	78	15-20
	0.64 × 25	0.025 × 1	DCEP(B)	525	27	305	12	NONE		14	30	10-15
	0.64 ×	0.025 × 2	DCEP(B)	1050	27	305	12	NONE		27	60	10-15

	50										
	0.64 × 100	0.025 × 4	DCEP(B)	2100	27	305	12	NONE	54	120	10-15

(A) VARIATIONS IN CURRENT, VOLTAGE, AND TRAVEL SPEED AFFECT PERCENTAGE OF DILUTION.

(B) CONSTANT POTENTIAL

The single-wire process is largely used for overlaying limited areas or where it is necessary to limit the heat input. Wires from 1.6 to 4.8 mm (or $\frac{1}{16}$ to $\frac{3}{16}$ in.) diam are practical. The small-size wires are generally employed with oscillation to keep dilution low, to provide a wide (up to about 25 mm, or 1 in.) weld bead, and to give uniformly good tie-ins with the adjacent beads. Either alternating or direct current may be employed for single-wire overlays.

Productivity (deposition rate) is greatly increased by the use of multiple wires (Table 17). Two-wire series arc (each wire connected to the poles of an alternating current power source) is a frequently used method for producing type 347 stainless steel weld cladding of petrochemical pressure vessels. The wire size is usually either 3.2 or 4.0 mm ($\frac{1}{8}$ or $\frac{5}{32}$ in.) in diameter. Alternating current is essential to achieve uniform fusion of both wires. The height of the welding head over the base metal greatly influences the penetration and hence the dilution. If the base metal has a non-uniform surface, or if the head cannot be accurately controlled above the base metal, the process may not be acceptable because of variations in dilution, or even lack of penetration in spots.

"Shingling" is a technique used to minimize dilution in single- or two-wire weld claddings. By providing an overlap of as much as 70% of a bead width, it is possible to reduce the dilution from 30 or 35% to 15 or 20%. In some applications, this allows the cladding to be deposited in a single layer at considerable cost savings as compared with two-layer overlays.

To further improve productivity, fabricators have increased the number of wires deposited by a single welding head. Up to six wires have been used with oscillation to deposit beads 150 mm (6 in.) or more in width. Usually, the wire sizes are 1.6 to 4.0 mm ($\frac{1}{16}$ to $\frac{5}{32}$ in.) in diameter. Direct current electrode negative has been found best for relatively low dilution (around 15%). Sometimes the wire that overlaps the previously deposited bead is connected to the positive tap of the power source (DCEP), which gives somewhat greater penetration and ensures good tie-in with the previous bead.

For weld cladding the inside surfaces of large pressure vessels, as shown in Fig. 29(a)and 29(b), wide beads produced by oscillated multiple-wire systems or strip electrodes have become the means to improve productivity and minimize dilution while offering a uniformly smooth surface. Welding parameters for stainless steel strip weld overlays are illustrated in Fig. 30, 31, 32. In Fig. 30, the effect of increased current on dilution, penetration, and bead thickness of a 60 mm (2.4 in.) strip is shown. Figure 31 illustrates the effect of travel speed on dilution, penetration, and bead dimensions using a 60 mm (2.4 in.) strip. Productivity of different strip sizes (deposition rate versus arc current) is shown in Fig. 32.

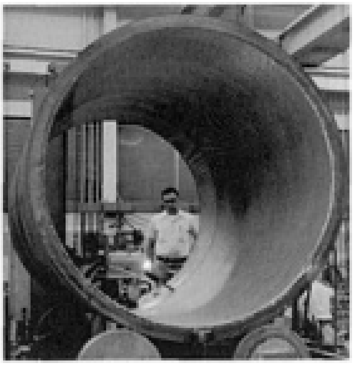

FIG. 29(A) WELD CLADDING OF A 1.8 M (6 FT) INNER DIAMETER PRESSURE VESSEL SHELL WITH 50 MM (2 IN.) WIDE, 0.64 MM (0.025 IN.) THICK STAINLESS STEEL STRIP. COURTESY OF J.J. BARGER, ABB COMBUSTION ENGINEERING

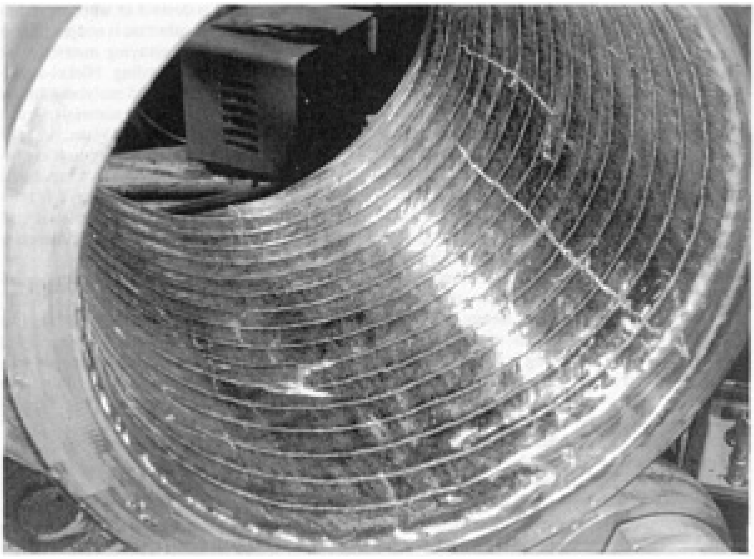

FIG. 29(B) CLOSEUP VIEW OF THE 25 MM (1 IN.) WIDE BY 0.64 MM (0.025 IN.) THICK STAINLESS STEEL STRIP USED TO CLAD A 300 MM (12 IN.) INNER DIAMETER PRESSURE VESSEL NOZZLE. COURTESY OF J.J. BARGER, ABB COMBUSTION ENGINEERING

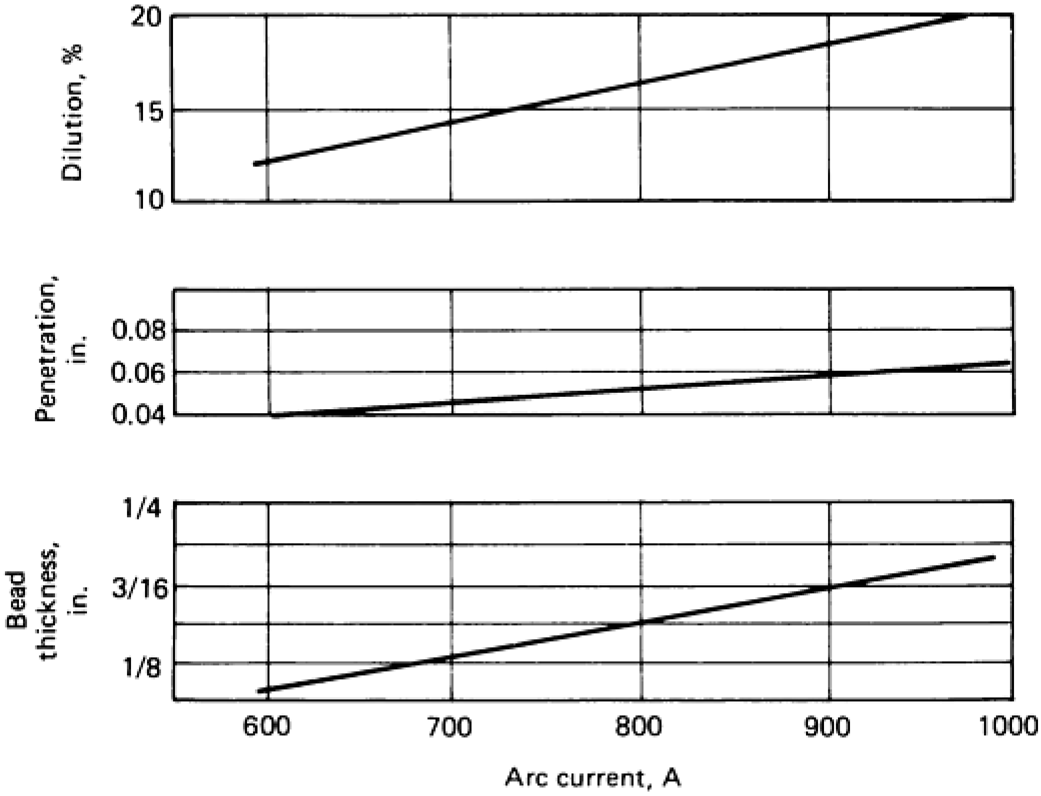

FIG. 30 DILUTION, PENETRATION, AND BEAD THICKNESS USING 60 MM (2.4 IN.) STRIP AT VARIOUS CURRENTS. SOURCE: REF 26

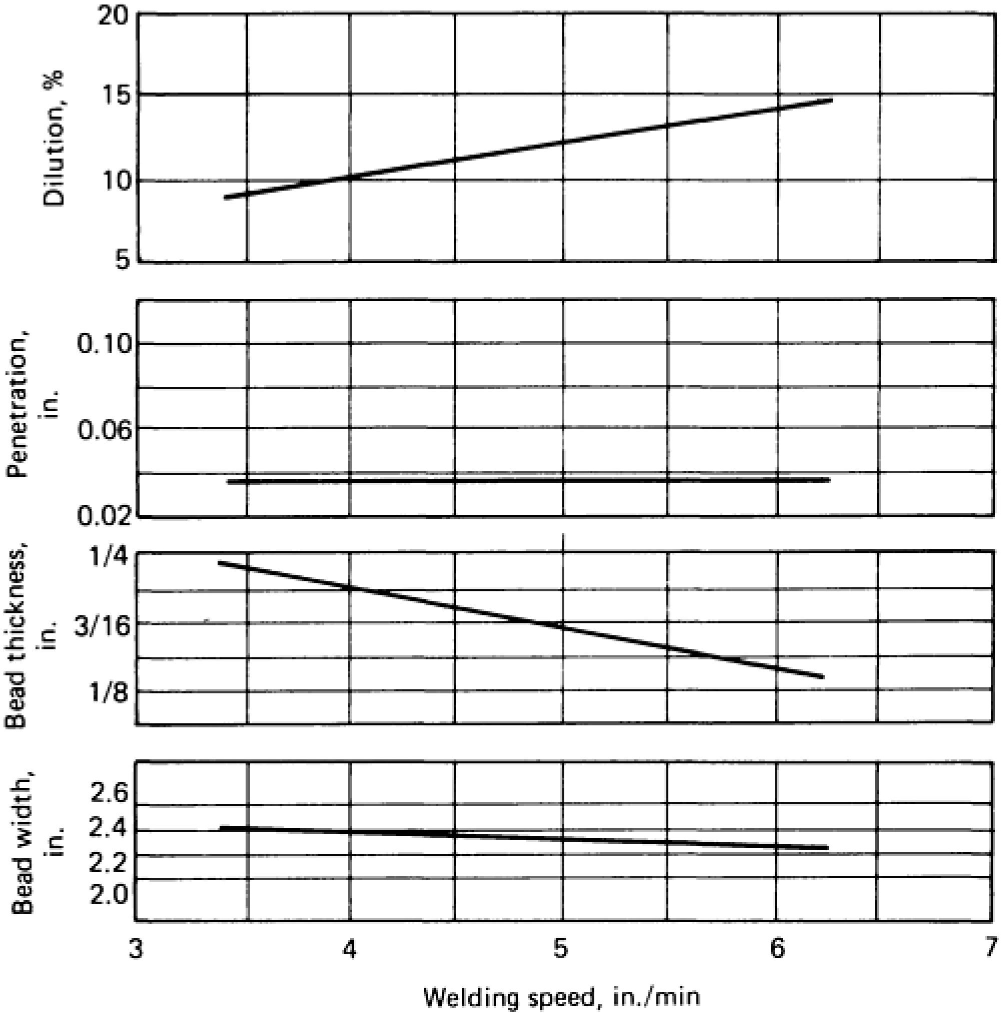

FIG. 31 INFLUENCE OF TRAVEL SPEED ON DILUTION, PENETRATION, AND BEAD THICKNESS USING 60 MM (2.4 IN.) STRIP. SOURCE: REF 26

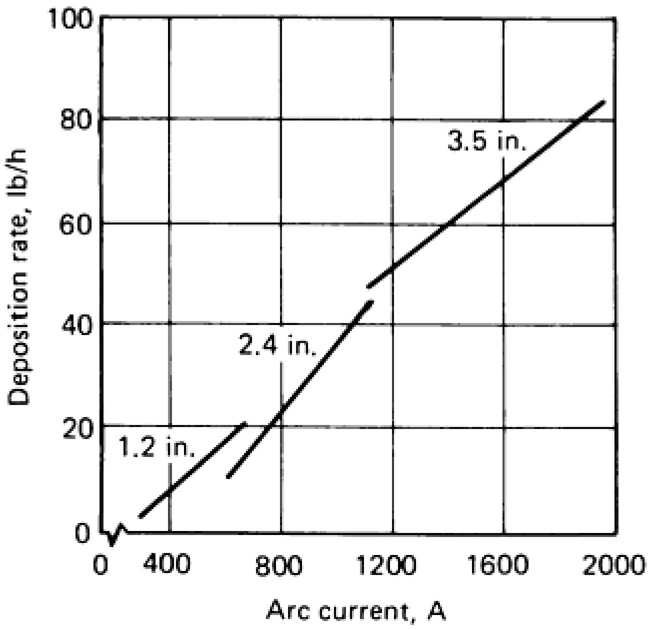

FIG. 32 DEPOSITION RATE (100% DUTY CYCLE) AS A FUNCTION OF CURRENT AND DIFFERENT STRIP SIZES. SOURCE: REF 26

The fluxes for submerged arc overlays should have low bulk density, especially for wide weld beads produced with multiple wire or strip electrodes. Standard fused fluxes, therefore, are usually unsuitable. Agglomerated fluxes, which have much lower bulk density, are commonly used for this purpose. Low-bulk-density fused fluxes have been successfully produced by a manufacturing process known as foaming. Such foamed fluxes have even lower bulk density than agglomerated fluxes.

Agglomerated fluxes (or bonded fluxes) have a distinct advantage over fused fluxes in that it is possible to introduce alloying elements by means of the flux. When this is done, they are known as reinforced, or alloyed, fluxes. For stainless steel weld claddings, chromium is commonly added to the fluxes to overcome the oxidation of chromium, which is characteristic of SAW of stainless alloys, and to compensate for the dilution with the non-chromium or low-chromium base metal. Other alloying elements often added to reinforced fluxes are manganese, nickel, molybdenum, and niobium. With reinforced fluxes, it is possible to produce various modified stainless weld overlays using a single filler metal as an electrode, for example, a type 347 overlay with a 308 or 309 filler metal using a niobium-alloyed flux. Table 18 lists variations in chemical compositions of series-arc weld deposits using an agglomerated flux and type 308 or 309 wire.

TABLE 18 CHEMICAL COMPOSITION VARIATIONS IN SERIES-ARC WELD CLADDING DEPOSITS USING AN AGGLOMERATED FLUX AND TYPE 308 OR 309 WIRE

WELDING CONDITIONS	**WIRE SIZE**		**ANGLE BETWEEN WIRES, DEGREES**	**CURRENT (AC), A**	**VOLTAGE, V**	**TRAVEL SPEED**	
	mm	**in.**				**mm/min**	**in./min**
TYPE 308 OVERLAY	4.0	$\frac{5}{32}$	45	500	20-25	255	10
TYPE 309 OVERLAY	4.0	$\frac{5}{32}$	45	540	26-30	330	13

CHEMICAL ANALYSIS	COMPOSITION, %						
	C	**Mn**	**Si**	**S**	**P**	**Cr**	Ni
308 FILLER WIRE	0.04	1.74	0.30	0.011	0.022	21.30	**9.83**
TOP OR THIRD LAYER	0.05	1.39	0.61	0.007	0.028	20.95	**9.57**
SECOND LAYER	0.06	1.36	0.63	0.010	0.029	20.29	**9.02**
FIRST LAYER	0.07	1.20	0.58	0.006	0.026	18.08	**8.16**
CARBON STEEL BASE	0.07	0.51	. . .	0.013	0.016	. . .	**. . .**
309 FILLER WIRE	0.05	1.81	0.32	0.016	0.022	24.68	**13.36**
TOP OR THIRD LAYER	0.06	1.49	0.93	0.005	0.043	24.43	**13.14**
SECOND LAYER	0.06	1.55	0.84	0.006	0.040	24.37	**11.08**
FIRST LAYER	0.07	1.36	0.84	0.008	0.039	21.79	**11.43**
CARBON STEEL BASE	**0.07**	**0.51**	**. . .**	**0.013**	**0.016**	**. . .**	**. . .**

Another method for providing the desired stainless steel cladding composition is to introduce a layer of powdered metal under the submerged arc flux. This procedure is sometimes known as bulk welding. It is suitable for use either with single-wire or strip electrodes. The powdered metal not only provides the alloying elements needed in the overlay, but it also greatly reduces the dilution. A schematic of the process is shown in Fig. 33.

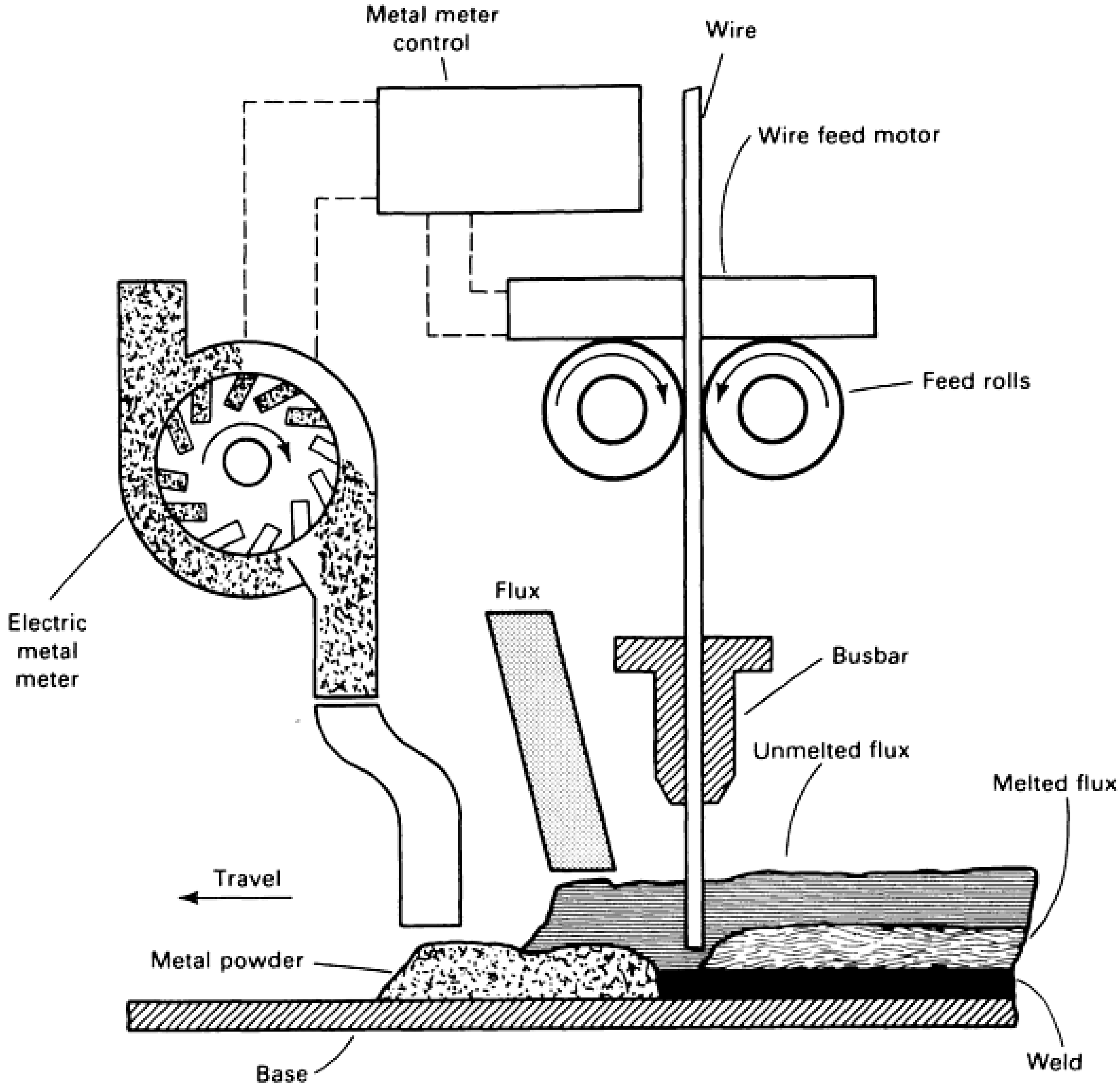

FIG. 33 SCHEMATIC OF THE BULKWELDING PROCESS USING METAL POWDER JOINT FILL. SOURCE: REF 9

Either method of adding alloying elements by using powdered alloys, either as an ingredient of the flux or separately underneath the flux, is subject to variability of the deposit composition. When reinforced fluxes are employed, the recovery of the alloying ingredient is dependent on the arc voltage, with a measurable difference in recovery by a change of as little as 1 or 2 V. This is not the case when the powdered alloys are separately fed below the flux, but in this case variability results if the height of the powder layer varies. Thus, it is unwise to rely on these methods of alloy adjustment of the weld overlay composition for more than three percentage points.

Still another method of providing the desired composition for the cladding is to use tubular wire for the filler metal. Whereas the solid wire producer is limited in the compositions that can be produced and still provide hot workability and cold drawing capability, the producer of tubular wire is virtually unlimited as to the compositions that can be produced. For example, the strip can be chromium or chromium-nickel steel, to which the core ingredients are added to give a weld deposit of austenitic stainless steel, rich enough in the essential alloys to compensate for both oxidation and dilution losses. Another distinct advantage of tubular wires is for relatively small production applications of compositions that are not available as "shelf items" in solid-wire form. Small quantities of tubular wires can also be prepared for experiments to determine the optimum composition for a particular application. The effective cross section of tubular wire for current-carrying ability is substantially less than solid wire of the same diameter. Welding parameters that have been established for solid wires will have to be adjusted to take this and other differences into account.

Self-Shielded Flux-Cored Wire. Tubular wires are produced that contain not only the alloying ingredients, but also the fluxing and gas-shielding ingredients. This permits the deposition of weld claddings without the need of an external flux, as is needed with SAW, or an external gas, as is needed with gas-shielded solid wire welding. Weld claddings are produced by manually directed welding guns for limited areas using a single tubular filler wire.

High productivity procedures have been developed using multiple wires in a manner similar to six-wire SAW. These wires have inherently low penetration characteristics, making them highly suitable for cladding. Deposition rates of 35 to 45 kg/h (80 to 100 lb/h) are often achieved. In addition, improved wire feed roll mechanisms have lessened the problems of kinking and crushing of these thin-wall electrodes during the wire-feeding process, thereby increasing the productivity of multiple-wire applications. Some fabricators who use self-shielded flux-cored wires prefer to limit the automatic cladding process to two or three wires, particularly because the gain in productivity with increasing numbers of wires diminishes. A three-wire welding head for overlay applications is shown in Fig. 34.

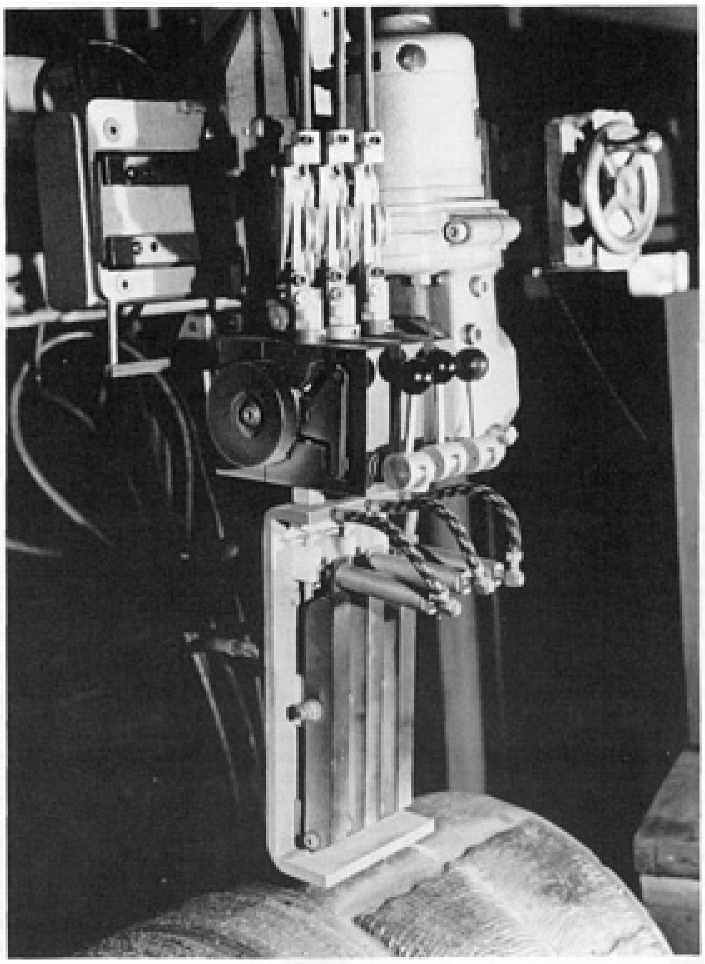

FIG. 34 THREE-WIRE WELDING HEAD FOR WELD CLADDING APPLICATIONS. COURTESY OF STOODY COMPANY

Metallurgically, flux-cored wires offer potentially higher alloyed claddings, permitting a considerable cost saving to produce the desired composition in a single layer. Also, dilution is minimized with this process. Furthermore, carbon content can be kept to much lower levels than is possible with solid wires, partly due to the ability to have very low-carbon ingredients in the flux-cored wire and partly due to the low dilution. Typical welding parameters for stainless steel weld cladding using self-shielded flux-cored wires are given in Table 19.

TABLE 19 TYPICAL WELDING PARAMETERS FOR STAINLESS STEEL WELD CLADDING USING SELF-SHIELDED FLUX-CORED WIRES

NO. OF 2.4 MM ($\frac{8}{32}$ IN.) DIAM WIRES	**CONSTANT POTENTIAL POWER SOURCE CAPACITY[(A)]**	**CURRENT, A**	**VOLTAGE, V**	**TRAVEL SPEED**		**OSCILLATION, CYCLES/MIN**	**PRODUCTIVITY**		DILUTION, %
				MM/MIN	**IN./MIN**		**KG/H**	**LB/H**	
1	400	300	27	510	20	NONE	5.4	12	**20**
2	800	600	27	115	4.5	20	14	30	**12**
3	1200	900	27	100	4.0	20	20	45	**12**
6	**1200 EACH (50 MM, OR 2 IN., PARALLEL)**	**1800**	**27**	**90**	**3.5**	**20**	**39**	**85**	**12**

(A) 100% DUTY CYCLE

The plasma arc hot wire process is used for depositing stainless steel claddings on pressure vessels subjected to corrosive, high-temperature, and high-pressure hydrogen environments. The setup for this process is illustrated in Fig. 35. Using this method, two filler-metal wires are fed at a constant speed through individual hot wire nozzles to intersect in the weld pool beneath the plasma arc. The filler metals are electrically connected in series and energized by an alternating current constant-potential power supply. Resistance heating occurs in each electrode between the wire contact tip and the weld pool. Input power is selected for a given filler-metal feed rate so that maximum resistance heating is induced into both electrodes without resulting in an open arc between them. While a single hot wire increases the deposition rate considerably over the cold wire or powder methods, a dual hot wire process permits deposition rates as high as 18 to 32 kg/h (40 to 70 lb/h), thus making the hot wire process commercially competitive with other weld cladding processes. A continuously made stainless steel weld overlay made by the hot wire process is shown in Fig. 36.

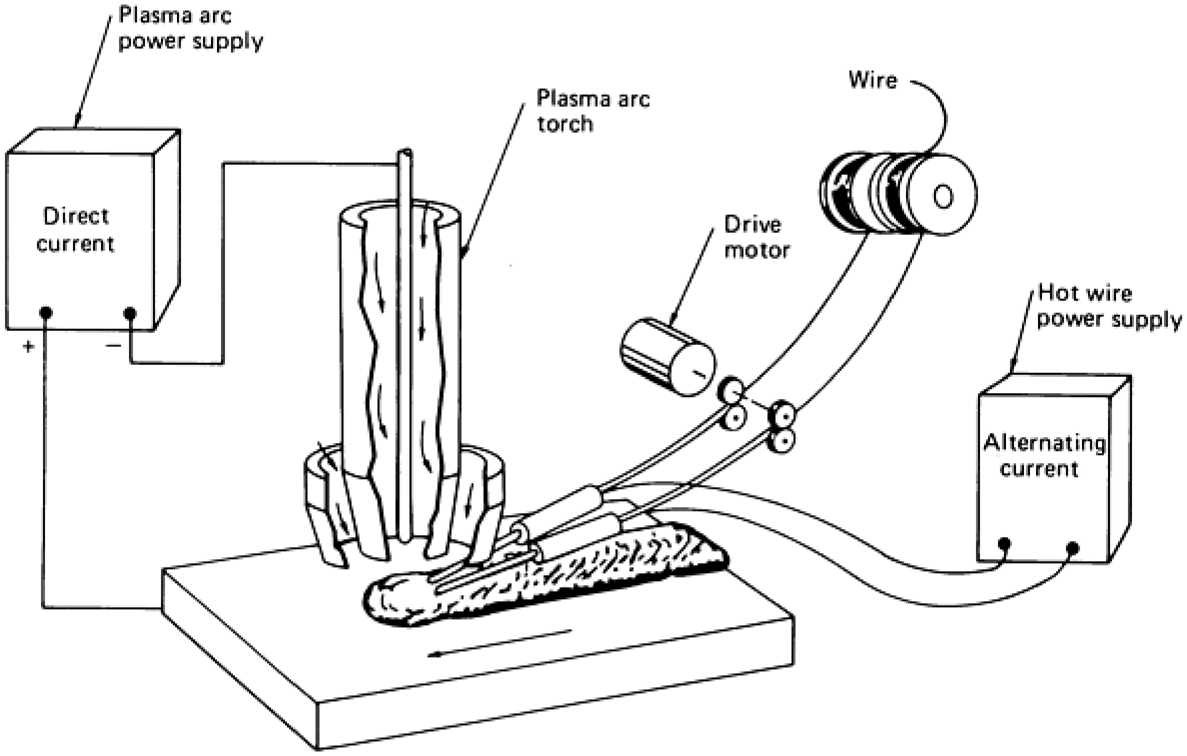

FIG. 35 SCHEMATIC OF THE PLASMA ARC HOT WIRE WELD CLADDING PROCESS. SOURCE: REF 1

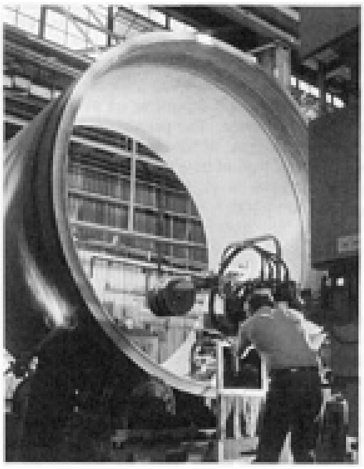

FIG. 36 PLASMA ARC HOT WIRE PROCESS FOR WELD CLADDING THE INTERNAL SURFACE OF A PRESSURE VESSEL SHELL. COURTESY OF NOOTER CORPORATION

Advantages of the hot wire process include:

- ACCURATE CONTROL OF DILUTION AND THUS COMPOSITION OF THE OVERLAY
- LITTLE OR NO CHANGES DUE TO OXIDATION OF THE ELEMENTS IN THE PROCESS
- PARAMETERS THAT CAN BE INDEPENDENTLY AND ACCURATELY CONTROLLED
- SINGLE-LAYER OVERLAYS THAT CAN BE PRODUCED IN THICKNESSES RANGING FROM 4.0 TO 8.0 MM ($\frac{5}{32}$ TO $\frac{5}{32}$ IN.)
- APPLICABILITY TO MANY CORROSION-RESISTANT ALLOYS, SUCH AS MONEL, INCONEL, AND HASTELLOY ALLOYS

Typical welding parameters for stainless steel weld cladding by the plasma arc process are given in Table 20.

TABLE 20 TYPICAL WELDING PARAMETERS FOR STAINLESS STEEL WELD CLADDING BY THE PLASMA ARC PROCESS

TORCH CONDITIONS				**FILLER-WIRE CONDITIONS**				**TRAVEL SPEED**		**PRODUCTIVITY**		DILUTION, %
CURRENT, A	**VOLTAGE, V**	**GAS FLOW RATES**		**NO. OF WIRES**	**SIZE**		**CURRENT, A**					
		mm^3/h × 10^6	**$in.^3$/h**		**mm**	**in.**		**mm/min**	**in./min**	**kg/h**	**lb/h**	
440	38	1.4	85	2	1.6	$\frac{1}{16}$	160	205	8	18-23	40-50	**8-12**
480	38	1.4	85	2	1.6	$\frac{1}{16}$	180	230	9	23-27	50-60	**8-12**
500	39	1.4	85	2	1.6	$\frac{1}{16}$	200	230	9	27-32	60-70	**8-15**
500	**39**	**1.4**	**85**	**2**	**2.4**	$\frac{3}{32}$	**240**	**255**	**10**	**27-32**	**60-70**	**8-15**

Electroslag Overlays. The electroslag welding process has been adapted to produce stainless steel weld claddings. This process offers the ability to use wider strip (up to 300 mm, or 12 in.) than the submerged arc process and more uniform penetration into, and lower dilution with, the base metal. Its principal disadvantages are operator exposure to the radiant heat from the molten slag and a sharper compositional change at the fusion line, which may give unsatisfactory service in certain high-temperature pressure vessels (for example, disbonding in hydrocracker or hydrosulfurizer service).

For large heavy-wall pressure vessels, the electroslag overlay process typically uses strip 4.0 mm (0.16 in.) thick by 150 mm (6 in.) wide. The effect of electroslag welding parameters on dilution using a 150 mm (6 in.) strip is shown in Fig. 37. As in SAW, energy is created by ohmic resistance as the electric current flows from the strip to the work through the molten slag; no arc exists. Unlike SAW, however, the slag bath is relatively thin. To provide the required electrical resistance, viscosity, and other properties, the fluxes must be significantly higher in fluorides than for SAW (Ref 26). A typical fused flux composition comprises 49% CaF_2, 21% CaO, 21% SiO_2, and 9% Al_2O_3.

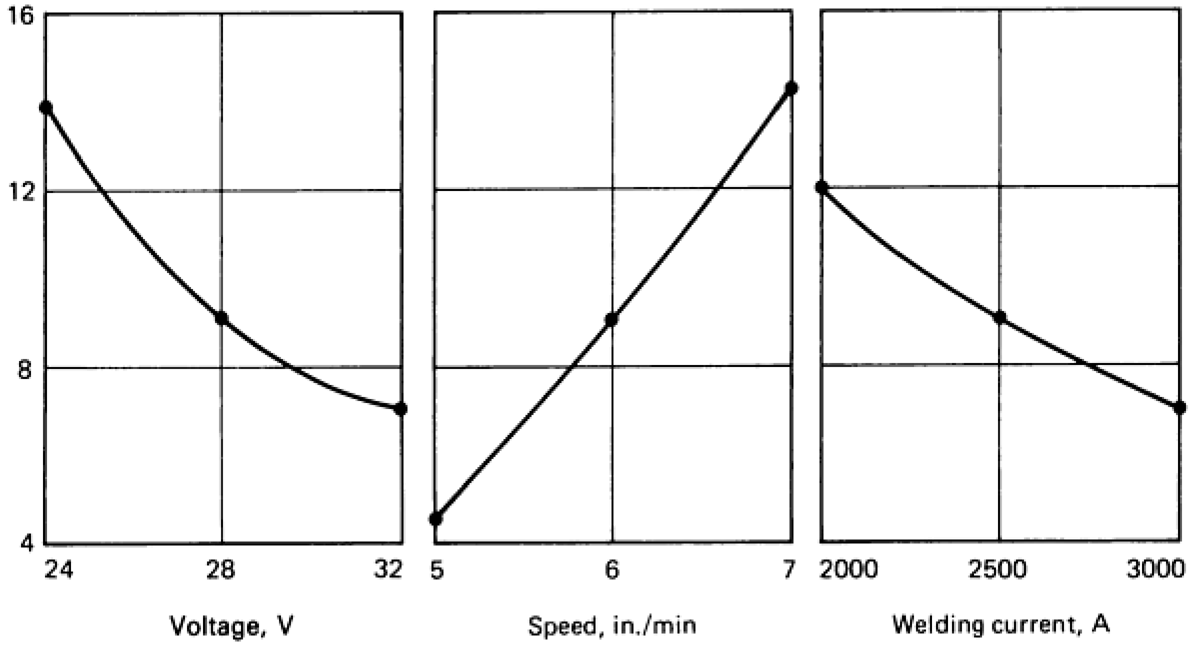

FIG. 37 EFFECT OF WELDING PARAMETERS ON DILUTION USING 150 MM (6 IN.) STRIP IN ELECTROSLAG WELD CLADDING PROCESS. SOURCE: REF 26

Because of disbonding of weld claddings deposited on pressure vessels in a hydrogen environment, electroslag weld claddings are not used for single-layer overlays; but where two-layer overlayers are specified, the electroslag process is used for the second layer. Higher productivity and a smoother surface are thus obtained, the first layer having been deposited by the SAW.

Extremely low carbon levels are possible because of low dilution with the base metal and a removal of carbon through oxidation in the process. Using a 0.18% C base metal and a 0.01% C strip, one investigator produced weld-metal carbon levels of 0.02% in the first layer and 0.01% in the second (Ref 26). The electroslag weld cladding process also can be applied for weld overlays of nickel-chromium (Inconel) alloys in the same manner as for stainless steel.

Weld Overlays Other Than Stainless Steel (Ref 26)

Many of the procedures described for stainless steel weld cladding are applicable to other alloys. Generally, nickel-base alloys can be produced successfully using essentially the same procedures as for stainless steels. Some of the common alloy filler metals used are given in Table 21. Figure 38 shows the use of both Inconel and stainless steel strip overlays used to clad a large pressure vessel shell.

TABLE 21 NONFERROUS FILLER METALS USED FOR WELD CLADDING OF STEELS

WELD OVERLAY TYPE	**COVERED ELECTRODE**		BARE ROD OR ELECTRODE	
	AWS SPECIFICATION NO.	**CLASSIFICATION NO.**	**AWS SPECIFICATION NO.**	CLASSIFICATION NO.
ALUMINUM-BRONZE	A5.6	ECUAL-A2	A5.7	**ERCUAL-AL**
SILICON-BRONZE	A5.6	ECUSI	A5.7	**ERCUSI**
COPPER-NICKEL	A5.6	ECUNI(A)	A5.7	**ERCUNI(A)**
NICKEL	A5.11	ENI-1	A5.14	**ERNI-1**
MONEL	A5.11	ENICU-7	A5.14	**ENICU-7**
LNCONEL	A5.11	ENICRFE-3	A5.14	**ERNICR-3**
LNCONEL 625	A5.11	ENICRMO-3	A5.14	**ERNICRMO-3**
HASTELLOY B	A5.11	ENIMO-7	A5.14	**ERNIMO-7**

HASTELLOY C	**A5.11**	**ENICRMO-4**	**A5.14**	**ERNICRMO-4**

(A) FIRST LAYER MUST BE EITHER NICKEL OR MONEL.

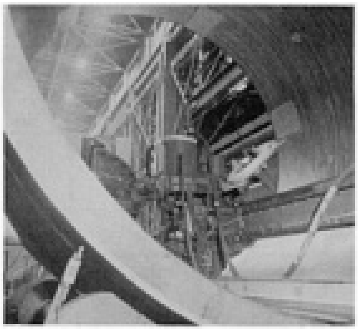

FIG. 38 WELD CLADDING OF A 1.8 M (6 FT) INNER DIAMETER PRESSURE VESSEL SHELL USING 100 MM (4 IN.) WIDE BY 0.64 MM (0.025 IN.) THICK STAINLESS STEEL STRIP. NOTE SMALL "WINDOWS" THAT HAVE BEEN CLAD WITH 50 MM (2 IN.) WIDE BY 0.64 MM (0.025 IN.) THICK INCONEL STRIP. COURTESY OF J.J. BARGER, COMBUSTION ENGINEERING

Metals and alloys that do not tolerate significant amounts of iron cannot be used in weld overlays on steels. Titanium, zirconium, magnesium, and aluminum are examples of incompatible metals for cladding steel. Pure copper is also unsatisfactory, but some copper alloys, such as aluminum-bronze and silicon-bronze, can be applied to steels. The incompatible metals are usually bonded by explosion cladding or roll bonding techniques, which are described elsewhere in this Volume.

References cited in this section

1. K.C. ANTONY *ET AL.*, HARDFACING, *WELDING, BRAZING, AND SOLDERING*, 9TH ED., VOL 6, *METALS HANDBOOK*, AMERICAN SOCIETY FOR METALS, 1983, P 771-793
9. K.M. KULKARNI AND V. ANAND, METAL POWDERS USED FOR HARDFACING, *POWDER METALLURGY*, 9TH ED., VOL 7, *ASM HANDBOOK*, AMERICAN SOCIETY FOR METALS, 1984, P 823-836
26. R.D. THOMAS *ET AL.*, WELD OVERLAYS, *WELDING, BRAZING, AND SOLDERING*, 9TH ED., VOL 6, *METALS HANDBOOK*, AMERICAN SOCIETY FOR METALS, 1983, P 804-819
27. D.L. OLSON, PREDICTION OF AUSTENITIC WELD METAL MICROSTRUCTURE AND PROPERTIES, *WELD. J.*, VOL 64 (NO. 10), 1985, P 2815-2955
28. A.L. SCHAEFFLER, CONSTITUTION DIAGRAM FOR STAINLESS STEEL WELD METAL, *MET. PROG.*, VOL 56 (NO. 11), 1949, P 680-680B
29. W.T. DELONG, FERRITE IN AUSTENITIC STAINLESS STEEL WELD METAL, *WELD. J.*, VOL 53 (NO. 7), 1974, P 2735-2865
30. T.A. SIEWERT, C.N. MCCOWAN, AND D.L. OLSON, FERRITE NUMBER PREDICTION TO 100 FN IN STAINLESS STEEL WELD METAL, *WELD. J.*, VOL 67 (NO. 12), 1988, P 2895-2985
31. D.J. KOTECKI, "VERIFICATION OF THE NBS-CSM FERRITE DIAGRAM," INTERNATIONAL INSTITUTE OF WELDING DOCUMENT II-C-834-88, 1988
32. D.J. KOTECKI AND T.A. SIEWERT, WRC-1992 CONSTITUTION DIAGRAM FOR STAINLESS STEEL WELD METALS: A MODIFICATION OF THE WRC-1988 DIAGRAM, *WELD. J.*, MAY 1992, P 171S-178S

Hardfacing, Weld Cladding, and Dissimilar Metal Joining

J.R. Davis, Davis & Associates

Dissimilar Metal Joining

Because various portions of a process system operate at different service conditions, different structural alloys are used in the design, and hence, dissimilar-metal welded joints may be required. Many factors must be considered when welding dissimilar metals, and the development and qualification of adequate procedures for the various metals and sizes of interest for a specific application must be undertaken.

Most combinations of dissimilar metals can be joined by solid-state welding (diffusion welding, explosion welding, friction welding, or ultrasonic welding), brazing, or soldering where alloying between the metals is normally insignificant. In these cases, only the differences in the physical and mechanical properties of the base metals and their influence on the serviceability of the joint should be considered. When dissimilar metals are joined by fusion welding processes, alloying between the base metals and a filler metal, when used, becomes a major consideration. The resulting weld metal can behave much differently from one or both base metals during subsequent processing or in service.

In this section, the principal factors that are responsible for failure (cracking) of dissimilar metal arc welds are described. These include:

- GENERAL ALLOYING PROBLEMS (BRITTLE PHASE FORMATION AND LIMITED MUTUAL SOLUBILITY) OF THE TWO METALS
- WIDELY DIFFERING MELTING POINTS
- DIFFERENCES IN COEFFICIENTS OF THERMAL EXPANSION
- DIFFERENCES IN THERMAL CONDUCTIVITY

In addition, service considerations that can result in dissimilar-metal weld failure by carbon migration or by corrosion/oxidation are addressed. Lastly, a brief review of specific dissimilar metal combinations and filler metal selection guidelines are presented. Additional information on welding dissimilar metals can be found throughout this Volume and in Ref 33 and 34.

Factors Influencing Joint Integrity (Ref 33)

Weld Metal. In the fusion welding of dissimilar-metal joints, the most important consideration is the weld metal composition and its properties. Its composition depends upon the compositions of the base metals; the filler metal, if used; and the relative dilutions of these. The weld metal composition is usually not uniform, particularly with multipass welds, and a composition gradient is likely to exist in the weld metal adjacent to each base metal.

These solidification characteristics of the weld metal are also influenced by the relative dilutions and the composition gradients near each base metal. These characteristics are important with respect to hot cracking of the weld metal during solidification.

The basic concepts of alloying, the metallurgical characteristics of the resultant alloy, and its mechanical and physical properties must be considered when designing a dissimilar-metal joint. For the fusion welding processes it is important to investigate the phase diagram of the two metals involved. If there is mutual solubility of the two metals, the joint can usually be made successfully. If there is little or no solubility between the two metals to be joined, the weld joint will not be successful. The intermetallic compounds that are formed between the dissimilar metals must be investigated to determine their crack sensitivity, ductility, susceptibility to corrosion, and so on. The microstructure of this intermetallic compound is extremely important. In some cases, it is necessary to use a third metal that is soluble with each metal in order to produce a successful joint.

Dilution. In dissimilar-metal welding, the filler metal must alloy readily with the base metals to produce a weld metal that has a continuous, ductile matrix phase. Specifically, the filler metal must be able to accept dilution (alloying) by the base metals without producing a crack-sensitive microstructure (dilution was defined and described earlier in this article in the sections on "Hardfacing" and "Weld Cladding"). The weld metal microstructure must also be stable under the expected service conditions. A successful weld between dissimilar metals is one that is as strong as the weaker of the two metals being joined, that is, possessing sufficient tensile strength and ductility so that the joint will not fail.

In multipass welding, the composition of each weld bead should be relatively uniform. However, definite compositional differences are likely in succeeding weld beads, especially between a root bead, the beads adjacent to the base metals, and the remaining fill beads. The average composition of the whole weld metal can be calculated when: (1) the ratio of the volumes of base metals melted to the entire weld metal volume can be determined; and (2) the compositions of the base and filler metals are known. The dilution can be based on area measurements on a transverse cross section through a test weld. Figure 39 illustrates how to determine the dilution by two base metals, A and B, when welding with filler metal F.

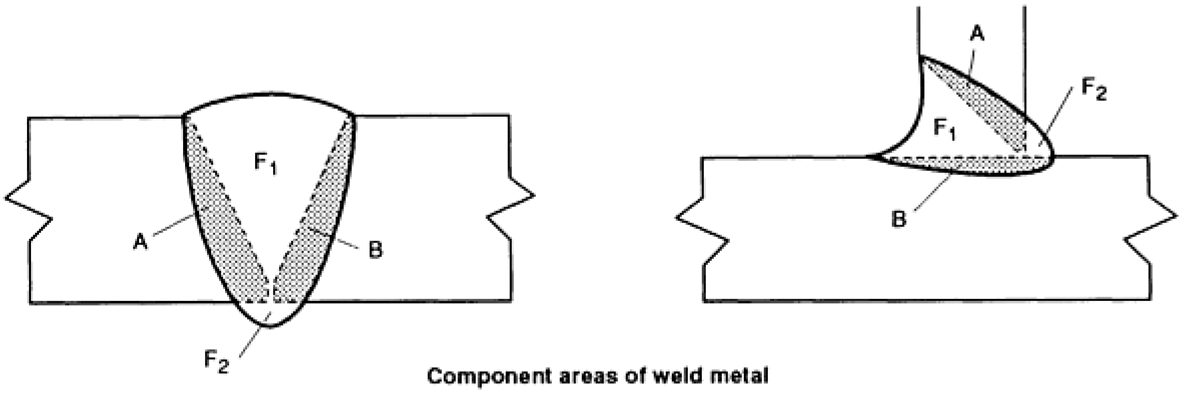

$$\text{Dilution by metal A, \%} = \frac{A}{A + B\,(F_1 + F_2)} \times 100$$

$$\text{Total dilution , \%} = \frac{A + B}{A + B\,(F_1 + F_2)} \times 100$$

$$\text{Dilution by metal B, \%} = \frac{A}{A + B\,(F_1 + F_2)} \times 100$$

FIG. 39 DILUTION IN A DISSIMILAR-METAL WELDED JOINT. SOURCE: REF 33

The average percentage of a specific alloying element in the diluted weld metal can be calculated using the following equation developed by the American Welding Society (Ref 33):

$$X_W = (D_A)(X_A) + (D_B)(X_B) + (1 - D_T)(X_F) \qquad \textbf{(EQ 5)}$$

where X_w is the average percentage of element X in the weld metal; X_A is the percentage of element X in base metal A; X_B is the percentage of element X in base metal B; X_F is the percentage of element X in the filler metal F; D_A is the percent dilution by base metal A, expressed as a decimal; D_B is the percent dilution by base metal B, expressed as a decimal; and D_T is the percent total dilution by base metals A and B, expressed as a decimal.

To illustrate the calculation of weld metal composition, assume that type 316 stainless steel is welded to a 2.25Cr-1Mo (UNS K21590) low-alloy ferritic pressure vessel steel with a nickel-chromium alloy filler metal ERNiCr3. The nominal chemical compositions of the three alloys are:

ALLOY	COMPOSITION, WT%			
	CR	**NI**	**MO**	FE
2.25CR-1MO	2.5	. . .	1.0	**95.5**
TYPE 316	17.0	12.0	2.5	**63.0**
ERNICR-3	**20.0**	**72.0**	**. . .**	**3.0**

Assuming that the total dilution is 35%, 15% by the Cr-Mo alloy steel and 20% from the type 316 stainless steel, the average percentages of Cr, Ni, and Mo in the weld metal are calculated as follows:

$$\mathrm{CR}, \% = 0.15(2.5) + 0.20(17) + 0.65(20) = 16.8 \quad \textbf{(EQ 6)}$$

$$\mathrm{NI}, \% = 0.20(12) + 0.65(72) = 49.2 \quad \textbf{(EQ 7)}$$

$$\mathrm{MO}, \% = 0.15(1) + 0.20(2.5) = 0.65 \quad \textbf{(EQ 8)}$$

Melting Temperatures. The difference in melting temperatures of the two metals that are to be joined must also be considered. This is of primary interest when a fusion welding process utilizing considerable heat is involved, since one metal may be molten long before the other when subjected to the same heat source. Significant difference between the melting temperatures of the two base metals or between those of the weld metal and a base metal can result in rupture of the metal having the lower melting temperature. Solidification and contraction of the metal with the higher melting temperature will induce stresses in the other metal while it is in a weak, partially solidified condition. This problem may be solved by depositing one or more layers of a filler metal of intermediate melting temperature on the face of the base metal with the higher melting temperature. This procedure is known as buttering. The weld is then made between the buttered face and the other base metal. The buttering layer should serve to reduce the melting temperature differential. Buttering may also be used to provide a transition between materials with substantially different coefficients of thermal expansion but which must endure cycling temperatures in service. Similarly, buttering may be used to provide a barrier layer that will slow the migration of undesirable elements from the base metal to the weld metal during postweld heat treatment or in service at elevated temperatures.

Thermal Conductivity. Most metals and alloys are relatively good conductors of heat, but some are much better than others. Rapid conduction of heat from the molten weld pool by an adjacent base metal may affect the energy input required to locally melt the base metal. When two dissimilar metals of significantly different thermal conductivities are welded together (for example, plain carbon steels and copper-base alloys), the welding procedure must provide for this difference. Often the welding heat source must be directed at the metal having the higher thermal conductivity to obtain the proper heat balance.

When welding dissimilar metals, heat loss to the base metals can be balanced somewhat by selectively preheating the metal having the higher thermal conductivity. Dilution is more uniform with balanced heating.

Preheating the base metal of higher thermal conductivity also reduces the cooling rate of the weld metal and the HAZ. The net effect of preheating is to reduce the heat needed to melt that base metal.

The coefficient of thermal expansion of the two dissimilar base metals is another important factor. Large differences in thermal expansion coefficients of adjacent metals during cooling will induce tensile stress in one metal and compressive stress in the other. The metal subject to tensile stress may hot crack during welding, or it may cold crack in service unless the stresses are relieved thermally or mechanically. This factor is particularly important in joints that will operate at elevated temperatures in a cyclic temperature mode. A common example of this is austenitic stainless steel/ferritic steel pipe butt joints used in energy conversion plants.

Ideally, the thermal expansion coefficient of the weld metal should be intermediate between those of the base metals, especially if the difference between those of the two base metals is large. If the difference is small, the weld metal may have an expansion coefficient equivalent to that of one of the base metals.

Welding Considerations (Ref 33)

Welding Process. The three most popular arc welding processes utilized for joining dissimilar metals are SMAW, GMAW, and GTAW. Selecting the welding process to make a given dissimilar-metal joint is almost as important as selecting the proper filler metal. As described in the articles in this Volume that cover the various fusion welding processes, the depth of fusion into the base metals and the resulting dilution may vary with different welding processes and techniques.

It is not uncommon with SMAW for the filler metal to be diluted up to 30% with base metal. The amount of dilution can be modified somewhat by adjusting the welding technique. For example, the electrode can be manipulated so that the arc impinges primarily on the previously deposited weld metal. The dilution rate can be kept below 25% with this technique. If dilution from one base metal is less detrimental than from the other, the arc should be directed toward that metal. This technique is also applicable to GTAW. Other means to control dilution are described in the section on "Weld Cladding" in this article.

Dilution rates with GMAW can range from 10 to 50%, depending upon the type of metal transfer and the welding gun manipulation. Spray transfer gives the greatest dilution; short-circuiting transfer, the least dilution. Penetration with SAW can be greater, depending on polarity, and can result in more dilution.

Regardless of the process, dilution is also affected by other factors, including joint design and fit-up. It is always best to have a minimum uniform dilution along the joint. Variations in dilution may produce inconsistent joint properties.

Selection of a suitable filler metal is an important factor in producing a dissimilar-metal joint that will perform well in service. One objective of dissimilar-metal welding is to minimize undesirable metallurgical interactions between the metals. The filler metal should be compatible with both base metals and be capable of being deposited with a minimum of dilution.

Two important criteria that should govern the selection of a proper filler metal for welding two dissimilar metals are as follows:

- THE CANDIDATE FILLER METAL MUST PROVIDE THE JOINT DESIGN REQUIREMENTS, SUCH AS MECHANICAL PROPERTIES OR CORROSION RESISTANCE.
- THE CANDIDATE FILLER METAL MUST FULFILL THE WELDABILITY CRITERIA WITH RESPECT TO DILUTION, MELTING TEMPERATURE, AND OTHER PHYSICAL PROPERTY REQUIREMENTS OF THE WELDMENT.

The Schaeffler diagram is commonly used to predict weld metal microstructure and subsequent filler metal selection when joining a stainless steel to a carbon or low-alloy steel. Figure 40 illustrates the procedure with an example of a single-pass weld joining mild steel to type 304 stainless steel with ER309 stainless steel filler metal. First a connecting line is drawn between the two points representing the base metal compositions, based on their chromium and nickel equivalents. Point X, representing the relative dilutions contributed by each base metal is then located on this line. If the relative dilutions are equal, point X is at the midpoint of the line. A second line is drawn between point X and the point representing the ER309 filler metal composition. The composition of the weld pass lies somewhere on this line, the exact location depending upon the total dilution. With 30% dilution, the composition would be at point Y and would be considered acceptable. If a succeeding pass joins the first pass to mild steel, the dilution with the mild steel should be kept to a minimum, to avoid martensite formation in the weld metal.

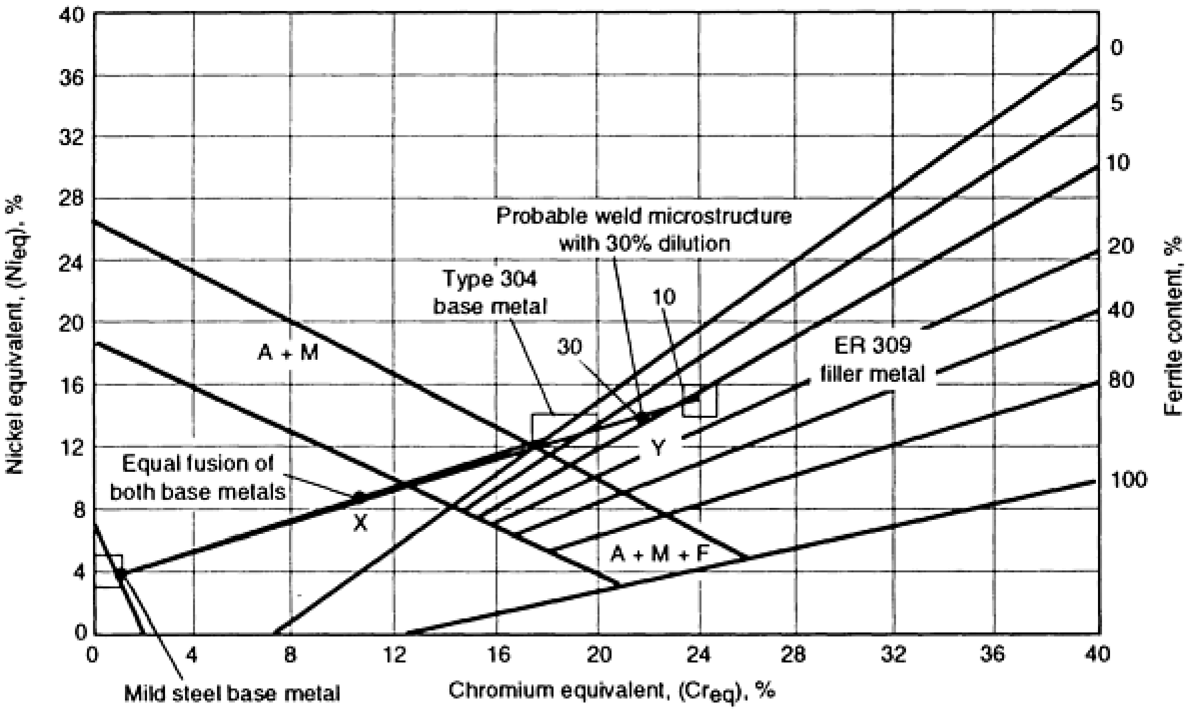

FIG. 40 PREDICTION OF WELD METAL COMPOSITION FROM THE SCHAEFFLER DIAGRAM. A, AUSTENITE; F, FERRITE, M, MARTENSITE. SEE TEXT FOR DETAILS. SOURCE: REF 33

For improved accuracy in predicting the microstructure of dissimilar-metal welds, the WRC-1988 or WRC-1992 diagrams should be used. The following example demonstrates the use of the WRC-1992 diagram for predicting the microstructure of a structural steel/stainless steel weld.

Example 2: Joining Type 304 Stainless Steel to ASTM A 36 Structural Steel with E309L-16 Electrode

(Ref 32). Table 22 lists typical plate compositions and all-weld-metal composition from the electrode. The Cr$_{eq}$ and Ni$_{eq}$ for the A 36 steel (0.00 and 7.08, respectively) do not permit an FN calculation for this material. The Cr$_{eq}$ and Ni$_{eq}$ for the type 304 stainless steel (18.83 and 12.45, respectively) result in a calculated 3.2 FN for this material. The Cr$_{eq}$ and Ni$_{eq}$ for these two base metals are shown in Fig. 41 and Point D (A 36) and Point E (type 304). Any mixture of these two base metals will lie along the line connecting Point D to E. If each base metal contributes equally to the weld metal, then the overall base metal contribution is given by the midpoint of the line between Point D and Point E, indicated in Fig. 41 as Point F. If one base metal contributed more than the other to the joint (for example, due to unequal plate thickness of a complex joint design), the Point F would slide along this line proportionately toward the greater contributor. In any case, the average base metal contribution to the weld pool would be along this line.

TABLE 22 BASE METAL AND WELD METAL COMPOSITIONS PRODUCED WHEN ASTM A 36 STRUCTURAL STEEL AND TYPE 304 STAINLESS STEEL WERE JOINED BY SMAW USING E309L-16 FILLER METAL

See Example 2 in text.

MATERIAL	**AISI 304**	**ASTM A 36**	**E309L-16 ALL-WELD-METAL**	70% E309L-16, 15% AISI 304, 15% A 36
C, %	0.05	0.20	0.03	**0.059**
MN, %	1.60	0.80	1.40	**1.34**
SI, %	0.40	0.20	0.60	**0.51**
CR, %	18.75	. . .	24.40	**19.89**

NI, %	9.90	. . .	12.70	**10.38**
MO, %	0.08	. . .	0.20	**0.15**
N, %	0.04	0.004	0.06	**0.049**
CR_{EQ}	18.83	0.00	24.60	**20.04**
NI_{EQ}	12.45	7.08	14.95	**13.39**
WRC-1988 FN	**3.2**	**. . .**	**17.4**	**4.3**

Source: Ref 32

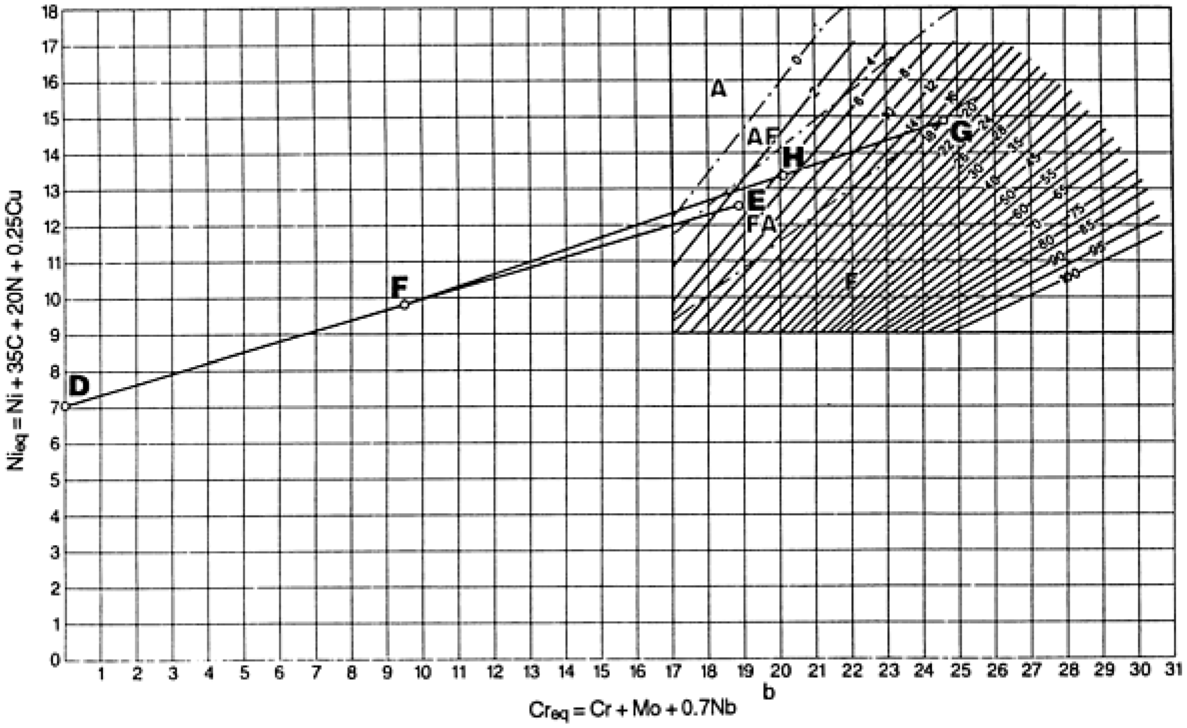

FIG. 41 DILUTION CALCULATION IN EXAMPLE 2 USING THE WRC-1992 DIAGRAM. SOURCE: REF 32

The all-weld-metal Cr_{eq} and Ni_{eq} (24.60 and 14.95, respectively) for the E309L-16 electrode is shown as Point G in Fig. 41, and a 17.4 FN would be predicted for that electrode. The root pass weld metal, consisting of the E309L-16 electrode and equal parts of the two base metals, must lie along the line from Point G to Point F in Fig. 41. Assuming the normal 30% base metal dilution with the shielded metal arc process, the root pass weld would lie 30% of the distance along the line from Point G to Point F. This is shown as Point H in Fig. 41. The calculations to reach Point H are as follows:

$$\begin{aligned} CR_{EQ}(F) &= 0.5\ CR_{EQ}(D) + 0.5\ CR_{EQ}(E) \\ &= 0.5(0.00) + 0.5(18.83) = 9.47 \end{aligned} \quad \textbf{(EQ 9)}$$

$$\begin{aligned} NI_{EQ}(F) &= 0.5\ NI_{EQ}(D) + 0.5\ NI_{EQ}(E) \\ &= 0.5(7.08) + 0.5(12.45) = 9.77 \end{aligned} \quad \textbf{(EQ 10)}$$

$$\begin{aligned} CR_{EQ}(H) &= 0.7\ CR_{EQ}(G) + 0.3\ CR_{EQ}(F) \\ &= 0.7(24.60) + 0.3(9.47) = 20.04 \end{aligned} \quad \textbf{(EQ 11)}$$

$$\begin{aligned} NI_{EQ}(H) &= 0.7\ NI_{EQ}(G) + 0.3\ NI_{EQ}(F) \\ &= 0.7(14.95) + 0.3(9.77) = 13.39 \end{aligned} \quad \textbf{(EQ 12)}$$

Point H corresponds to 4.3 FN, which would be expected to result in a crack-free root pass in this joint. Higher dilution, or excess dilution from the A 36 side of the joint, could reduce the ferrite content and increase the likelihood of cracking.

Buttering. If dilution of austenitic stainless steel filler metal is a problem, it may be controlled by first buttering the joint face of the carbon or low-alloy steel with one or two layers of type 309 or 310 stainless steel filler metal, as shown in Fig. 42(a) and 42(b). After machining and inspecting the buttered layer (Fig. 42c), the joint between the stainless steel component and the buttered steel part can be made using conventional welding procedures and the appropriate filler metal for welding the stainless steel base metal (Fig. 42d and 42e). A low-alloy steel component can be heat-treated after the buttering operation, then joined to the stainless steel part. This avoids a postweld heat treatment that might sensitize the austenitic stainless steel to intergranular corrosion. Additional information on buttering can be found in the article "Dissimilar Welds with Stainless Steels" in this Volume.

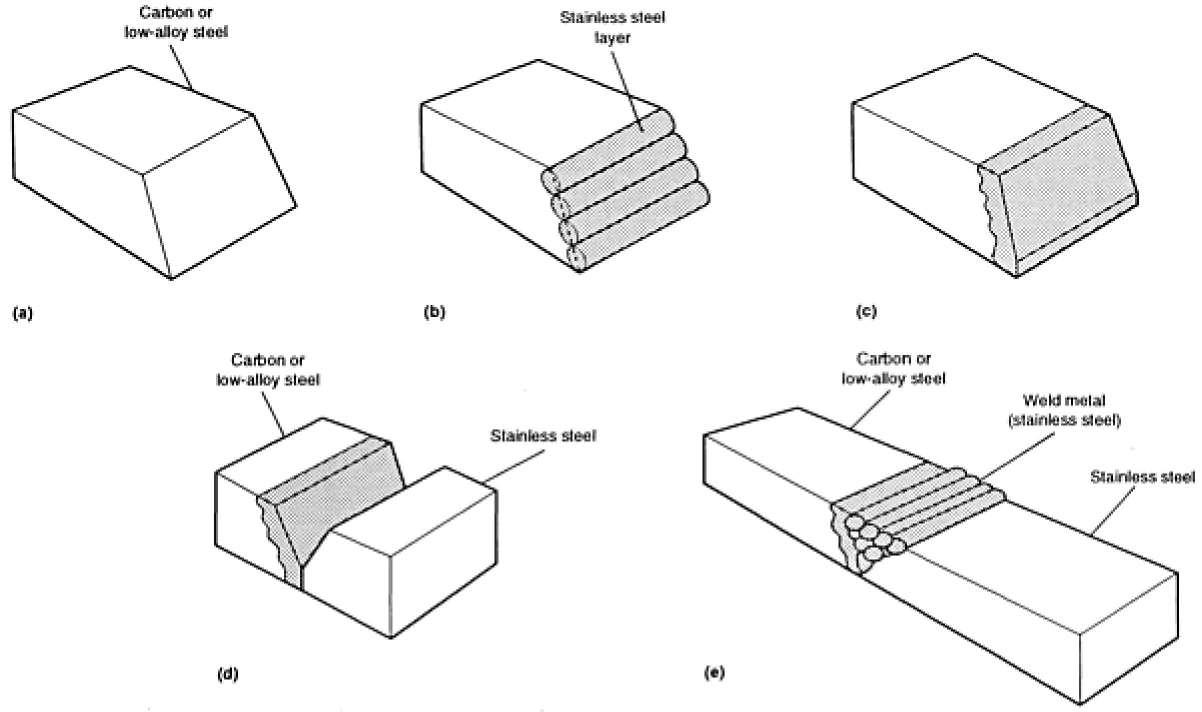
Carbon or low-alloy steel
Stainless steel layer
(a)
(b)
(c)
Carbon or low-alloy steel
Stainless steel
(d)
Carbon or low-alloy steel
Weld metal (stainless steel)
Stainless steel
(e)

FIG. 42 BUTTERING TECHNIQUE USED TO ASSIST WELDING STAINLESS STEEL TO CARBON OR LOW-ALLOY STEEL. (A) EDGE PREPARED FOR BUTTERING. (B) FACE BUTTERED WITH FILLER METAL. (C) BUTTERED FACE PREPARED FOR WELDING. (D) JOINT ALIGNED FOR WELDING. (E) JOINT WELDED WITH STAINLESS STEEL FILLER METAL. SOURCE: REF 33

Joint Design. When designing butt joints between dissimilar metals, consideration must be given to the melting characteristics of each base metal and the filler metal, as well as to dilution effect. Large grooves decrease dilution, permit better control of viscous weld metal, and provide room for better manipulation of the arc for good fusion.

The joint design should provide for appropriate dilution in the first few passes placed in the joint when welding from one side. Improper dilution could result in a layer of weld metal possessing mechanical properties that are inappropriate for the intended service, particularly when the joint will be exposed to cyclic stresses. When welding from both sides, back gouging of the first weld can provide better control of dilution in the first few passes of the second weld.

Dissimilar-metal joints are often made with groove geometries similar to those used in conventional similar-metal welds and are left in the as-welded condition. However, removing the excess penetration of the weld root eliminates possible defects that may be present in the root of the joint, and it eliminates notches and crevices associated with backing rings. Removing the weld root also improves the capabilities of nondestructive tests (radiography, liquid penetrant, and ultrasonic). Provision can also be made for removing the exterior reinforcement to eliminate stress concentrations associated with undercutting and with the change in section thickness across the as-welded joint. Again, inspection capabilities are improved.

Preheat and Postweld Heat Treatments. Selection of an appropriate preheat or postweld heat treatment (PWHT) for a welded joint can present a problem with some dissimilar metal combinations. The appropriate heat treatment for one component of the weldment may be deleterious to the other component for the intended service conditions. For example, if an age-hardenable nickel-chromium alloy is welded to a nonstabilized austenitic stainless steel, exposure of the weldment to the aging treatment for the nickel-chromium alloy would sensitize the stainless steel and decrease its resistance to intergranular corrosion (weld decay).

One solution is to use a stabilized austenitic stainless steel if that is acceptable. Another solution might be to butter the face of the age-hardenable, nickel-chromium alloy component with a similar alloy that is not age-hardenable. This component is then heat-treated to obtain the desired properties. Finally, the buttered surface is welded to the stainless steel component.

Service Considerations (Ref 33)

Property Considerations. A dissimilar metal joint normally contains weld metal having a composition different from that of one or both base metals. The properties of the weld metal depend on the filler metal composition, the welding procedures, and the relative dilution with each base metal. There also two different HAZs, one in each base metal adjacent to the weld metal. The mechanical and physical properties of the weld metal as well as those of the two HAZs, must be considered for the intended service.

Special considerations are generally given to dissimilar-metal joints intended for elevated-temperature service. A favorable situation exists when the joint will operate at constant temperatures. During elevated-temperature service, internal stresses can decrease by relaxation and reach an equilibrium. However, it is best to reduce the effect of large differences in coefficients of thermal expansion when large temperature fluctuations cannot be avoided in service. The problem can be avoided by selecting base metals with similar thermal expansion characteristics.

Figure 43 gives the mean linear thermal expansion coefficients as a function of temperature for several alloys commonly associated with transition butt joints for steam plant applications. The thermal expansion coefficient of 2.25Cr-1Mo steel is about 25% less than that of types 304 and 316 austenitic stainless steel. In certain applications, transition joints between austenitic stainless steel and low-alloy steel may experience numerous temperature changes during operation. For a given change in temperature, the stress imposed at the weld joint is proportional to the difference in their coefficients of thermal expansion. Stress analyses of welded joints between these two types of steel indicate that the stresses introduced by thermal change are nearly an order of magnitude greater than those produced by the operating pressures and any thermal gradients across the joint.

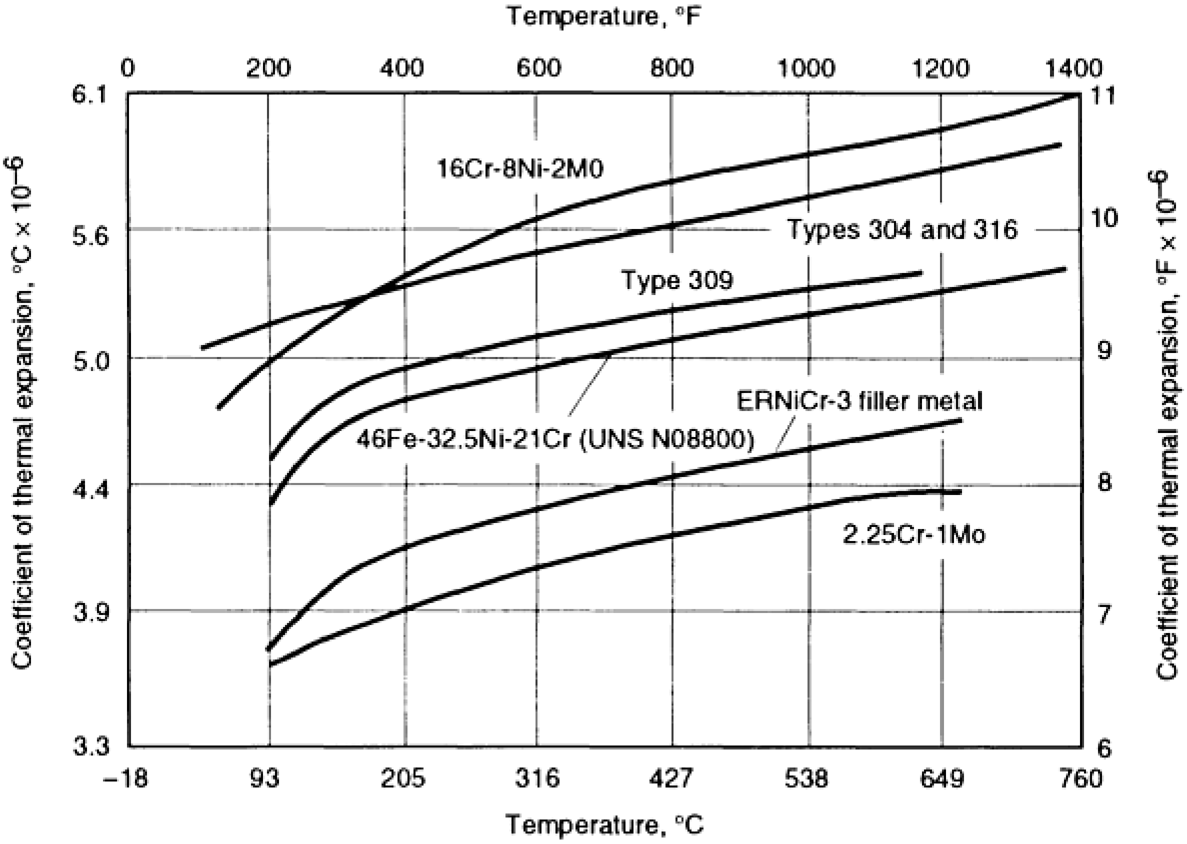

FIG. 43 MEAN COEFFICIENTS OF THERMAL EXPANSION AS A FUNCTION OF TEMPERATURE FOR TRANSITION BUTT-JOINT MATERIALS. SOURCE: REF 33

Experience with dissimilar-metal transition joints has shown a significant number of failures in less than the expected service life. The majority of the transition joint failures in austenitic steel/ferritic steel joints occur in the ferritic steel HAZ, adjacent to the weld interface. These failures are generally attributed to one or more of the following causes:

- HIGH STRESSES AND THE RESULTING CREEP AT THE INTERFACE AS A RESULT OF THE DIFFERENCES BETWEEN THE COEFFICIENTS OF THERMAL EXPANSION OF THE WELD AND BASE METALS
- CARBON MIGRATION FROM THE FERRITIC STEEL INTO THE STAINLESS STEEL, WHICH WEAKENS THE HAZ IN THE FERRITIC STEEL
- PREFERENTIAL OXIDATION AT THE INTERFACE, ACCELERATED BY THE PRESENCE OF STRESS

Carbon Migration. There is the likelihood of significant chemical composition gradients in the weld metal, particularly in the regions adjacent to the base metals. In addition, operation at elevated temperatures can cause interdiffusion between the weld metal and the base metal that, in turn, results in microstructure changes. A dissimilar-metal weld made between a low-alloy steel and an austenitic stainless steel, with an austenitic stainless steel filler metal, illustrates this problem. Chromium in steel has a greater affinity for carbon than does iron. When a carbon or low-alloy steel is welded with a steel filler metal containing a significant amount of chromium, carbon will diffuse from the base metal into the weld metal at temperatures above about 425 °C (800 °F). The diffusion rate is a function of temperature and exposure time, and it increases rather rapidly at 595 °C (1100 °F) and above. Carbon migration can take place during PWHT heat treatment or service at elevated temperatures.

Austenitic steel has a greater solubility for carbon than does ferritic steel. Therefore, carbon depletion in a carbon or low-alloy steel is greater when an austenitic stainless steel filler metal is used in preference to a ferritic steel filler metal. On the other hand, carbon migration is not as serious a problem when a nickel-chromium-iron filler metal is used.

If carbon migration is extensive, it will be indicated by a lightly etching, low-carbon band in the carbon steel HAZ and a dark, high-carbon zone in the stainless steel weld metal, as seen in a transverse metallographic section. The extent of carbon migration during PWHT or elevated-temperature service should be determined during welding procedure qualification. It is known to influence long-time stress-rupture strength during elevated-temperature service.

During cyclic temperature service, the HAZ will be subjected to varying shear stresses because of the differences in coefficients of thermal expansion of the base and weld metals. These stresses may produce fatigue failure in the decarburizing band next to the weld interface.

Corrosion and Oxidation Resistance. The weld metal and both base metals have specific corrosion behaviors that must be considered by the designer in the initial selection of materials. For example, with dissimilar metal weldments, the formation of galvanic cells can cause corrosion of the most anodic metal or phase in the joint. Also, the weld metal is usually composed of several microstructural phases, and very localized cells between phases can result in galvanic corrosion at the microstructural level. To minimize galvanic corrosion, the composition of the weld metal can be adjusted to provide cathodic protection to the base metal that is most susceptible to galvanic attack. However, other design requirements should not be seriously compromised to do this. Instead, some other form of protection should be used.

A galvanic cell associated with a high-strength steel may promote hydrogen embrittlement in the HAZ of that steel if it is the cathode of the cell. Hydrogen embrittlement must be considered if the service temperature of the weldment will be in the range of -40 to 95 °C (-40 to 200 °F), and the weld will be in a highly stressed area of the assembly. Residual stresses developed in the weld zone are often sufficient to promote hydrogen embrittlement and stress-corrosion cracking.

Chemical compositional differences in a dissimilar-metal weld can also cause high-temperature corrosion problems. Compositional variations at the interfaces between the different metals can result in selective oxidation when operating at high temperatures in air and formation of notches at these locations. Such notches are potential stress-raisers in the joint and can cause oxidation failure along the weld interface under cyclic thermal conditions.

Specific Dissimilar Metal Combinations (Ref 33)

The emphasis in the following section will be on joining carbon and low-alloy steels to various dissimilar materials (both ferrous and nonferrous) by arc welding. See other articles in this Section of the Handbook for additional information (for example, "Welding of Nickel Alloys" and "Welding of Copper Alloys").

Stainless Steels to Carbon or Low-Alloy Steels. Austenitic, ferritic, or martensitic stainless steels can be readily fusion welded to carbon or low-alloy steels using a filler metal that can tolerate dilution by both base metals without formation of flaws in the joint. An austenitic stainless steel or nickel alloy filler metal is commonly used. The choice depends on the application and the service conditions.

Generally, carbon or low-alloy steels are joined to 300-type stainless steels using austenitic filler metals, such as types E/ER 308, 309, 316, and 347. Such filler metals are used for service applications below 425 °C (800 °F). For higher temperatures or when high-temperature cyclic service is a concern, high-nickel filler metals such as ERNiCr-3 and ENiCrFe-3 are used. The coefficients of thermal expansion of high-nickel alloys approximate those of ferritic steels, and during cyclic-temperature service, the major differential expansion stresses are located primarily at the tough stainless steel/weld metal interface. Another advantage of a nickel-base weld metal is that it markedly reduces carbon migration from the ferritic steels to the weld metal. Extensive carbon migration into stainless steel weld metal weakens the HAZ of the carbon or low-alloy steel.

The problem of cracking between ferritic steels, such as the 2.25Cr-1Mo steel described earlier, and austenitic stainless steels is a recurring problem in such operations as the petrochemical industry and electric utility generating plants. Cracking has been repeatedly observed in HAZs of the ferritic portions of dissimilar-metal welds, particularly those that have been subjected to a large number of thermal cycles and extended time at elevated temperatures around 565 °C (1050 °F). Such cracking must be minimized, since repair maintenance and non-productive downtime are costly. High-nickel filler metals have proven useful in such applications.

Buttering is also used to improve the weldability between ferritic and austenitic stainless steel joints. A thick overlay (13 mm, or $\frac{1}{2}$ in., or greater) of a highly austenitic weld metal, such as type 309 stainless steel, can be deposited on carbon or low-alloy steel components so that stainless steel/stainless steel welds can be made subsequently in the field. This technique has limited reliability for high-temperature cyclic service and has not been used extensively for material combinations other than ferritic steels, stainless steels, and some high-nickel alloys.

A filler metal for joining a 4*xx*-series stainless steel to a carbon or low-alloy steel can be selected using the following general rules:

- FOR WELDING ONE HARDENABLE CHROMIUM STEEL TO ANOTHER WITH A HIGHER CHROMIUM CONTENT, FILLER METAL CONTAINING CHROMIUM EQUAL TO THAT OF EITHER STEEL MAY BE USED. FURTHERMORE, ANY FILLER METAL WHOSE CHROMIUM CONTENT LIES BETWEEN THESE LIMITS IS EQUALLY SATISFACTORY, PROVIDED THE WELDMENT IS PROPERLY HEAT-TREATED.
- A GENERAL RULE FOR WELDING ANY CHROMIUM STEEL TO ANY LOW-ALLOY STEEL IS TO USE A FILLER METAL THAT HAS THE SAME COMPOSITION AS THE LOW-ALLOY STEEL, PROVIDED THAT IT MEETS THE SERVICE REQUIREMENTS OF THE APPLICATION. WITH ANY LOW-ALLOY STEEL FILLER METAL, THE CHROMIUM THAT IS PICKED UP BY DILUTION WITH THE CHROMIUM STEEL BASE METAL MUST BE CONSIDERED.
- FOR WELDING ANY CHROMIUM STEEL TO A CARBON STEEL, CARBON STEEL FILLER METAL CAN BE USED. A CHROMIUM STEEL FILLER METAL CAN ALTERNATIVELY BE USED, BUT IT IS PREFERABLE TO USE A LESS HARDENABLE FILLER METAL.

When the properties of the HAZ in the chromium stainless steel are important, both the stainless steel and other steel can be buttered with type 309 or 310 austenitic stainless steel weld metal. An appropriate preheat or PWHT can be used to obtain desired properties in the buttered components. The stainless steel surfaces can then be welded together without preheat using a suitable austenitic stainless steel filler metal, such as type 308.

Nickel-Base Alloys to Steels. Nickel alloys can easily be welded to steels using a suitable filler metal and proper control of dilution. Nickel-base filler metals are generally used because of their good ductility and tolerance of dilution by iron. Suggested nickel-base filler metals for welding nickel alloys to steel or stainless steel are listed in Table 23.

TABLE 23 SUGGESTED FILLER METALS FOR WELDING NICKEL-BASE ALLOYS TO CARBON AND LOW-ALLOY STEELS

NICKEL ALLOY		**FILLER METAL FORM**	FILLER METAL FOR WELDING TO	
UNS NO.	**COMMON DESIGNATION**		**CARBON OR LOW-ALLOY STEEL**	STAINLESS STEEL
N02200	COMMERCIALLY PURE NICKEL	COVERED ELECTRODE	ENI-1, ENICRFE-2	**ENI-1, ENICRFE-2, ENICRFE-3**
		BARE WIRE	ERNI-1, ERNICR-3	**ERNI-1, ERNICR-3, ERNICRFE-6**
N04400	MONEL 400	COVERED ELECTRODE	ENICU-7, ENI-1	**ENICRFE-2, ENICRFE-3**
N05500	MONEL K-500	BARE WIRE	ERNI-1	**ERNICR-3, ERNICRFE-6**
N05502	MONEL 502			
N06600	INCONEL 600	COVERED ELECTRODE	ENICRFE-2, ENICRFE-3	**ENICRFE-2, ENICRFE-3**
N08800	INCOLOY 800	BARE WIRE	ERNICR-3, ERNICRFE-6	**ERNICR-3, ERNICRFE-6**
N08825	INCOLOY 825	COVERED ELECTRODE	ENICRMO-3	**ENICRMO-3**
		BARE WIRE	ERNICRMO-3	**ERNICRMO-3**

N10665	HASTELLOY B-2	COVERED ELECTRODE	ENIMO-7	**ENIMO-7**
		BARE WIRE	ERNIMO-7	**ERNIMO-7**
N10276	HASTELLOY C-276	COVERED ELECTRODE	ENICRMO-4	**ENICRMO-4**
		BARE WIRE	ERNICRMO-4	**ERNICRMO-4**
N06455	HASTELLOY C-4	COVERED ELECTRODE	ENICRMO-4	**ENICRMO-4**
		BARE WIRE	ERNICRMO-7	**ERNICRMO-7**
N06007	HASTELLOY G	COVERED ELECTRODE	ENICRMO-9	**ENICRMO-9**
		BARE WIRE	**ERNICRMO-1**	**ERNICRMO-1**

Source: Ref 33

Sulfur and phosphorus in nickel and nickel alloys cause hot cracking. The melting techniques used to produce nickel and its alloys are designed to keep the content of these elements to low levels. By contrast, the sulfur and phosphorus contents in some steels are typically higher. Consequently, dilution should be carefully controlled when joining a steel to a nickel alloy with a nickel alloy filler metal, to avoid hot cracking in the weld metal.

Most nickel-base weld metals can accept a substantial amount of iron dilution, but the dilution limit generally varies with the welding process. Weld metal deposited with nickel- or nickel-chromium-covered electrodes can tolerate up to approximately 40% iron dilution. With bare nickel or nickel-chromium filler metals, however, dilution should be kept to about 25%.

Acceptable limits of iron dilution for nickel-copper weld metal vary, depending on the weld process. With SMAW, iron dilution of up to about 30% can be tolerated. Submerged arc weld metal should not be diluted by more than 25%.

With the gas-shielding processes, nickel-copper weld metal is less tolerant of iron dilution, especially if the weld is to be thermally stress-relieved. The maximum limits for iron dilution in a welded joint are 10% when it will be used as-welded and 5% when it will be stress-relieved. A buttering layer of nickel or nickel-copper weld metal should be applied to the steel face in order to avoid exceeding these limits.

Nickel-copper weld metal has a maximum dilution tolerance for chromium of about 8%. Consequently, nickel-copper filler metal should not be used to join nickel-copper alloys to stainless steels (see Table 23).

Cobalt-Base Alloys to Steels. Metallurgically, cobalt-base alloys behave similarly to the high-temperature nickel-chromium alloys with respect to welding. When joining a cobalt alloy to a stainless steel, a filler metal with a composition similar to that of the cobalt alloy is recommended. A nickel alloy filler metal may also be suitable for some applications. In any case, the filler metal selection, welding process, and welding procedure for the application should be established by suitable tests.

Copper-Base Alloys to Steels. As described in the article "Welding of Copper Alloys" in this Volume, copper and many copper-base alloys can be joined to carbon, low-alloy, and stainless steels by GTAW, GMAW, and SMAW. Iron dilution can be minimized by the use of appropriate welding procedures or by placement of a buttering layer of nickel on the steel. Tables 24 and 25 show combinations of dissimilar metals by GTAW and GMAW with the aid of copper-base or nickel-base filler metals.

TABLE 24 FILLER METALS AND PREHEAT AND INTERPASS TEMPERATURES USED IN GAS-TUNGSTEN ARC WELDING OF COPPERS AND COPPER-BASE ALLOYS TO STEELS

METAL TO BE WELDED	FILLER METALS (AND PREHEAT AND INTERPASS TEMPERATURES) FOR WELDING METAL IN COLUMN 1 TO$^{(A)}$:				
	COPPERS	**PHOSPHOR BRONZES**	**ALUMINUM BRONZES**	**SILICON BRONZES**	COPPER NICKELS
LOW-CARBON STEEL	ERCUAL-A2 OR ERCU OR ERNI-3	ERCUSN-A (205 °C, OR 400 °F)	ERCUAL-A2 (150 °C, OR 300 °F)	ERCUAL-A2 (65 °C, OR 150 °F	**ERCUAL-A2 OR ERNI-3 (65 °C, OR**

	(540 °C, OR 1000 °F)			MAX)	**150 °F MAX)**
MEDIUM-CARBON STEEL	ERCUAL-A2 OR ERCU OR ERNI-3 (540 °C, OR 1000 °F)	ERCUSN-A (205 °C, OR 400 °F)	ERCUAL-A2 (205 °C, OR 400 °F)	ERCUAL-A2 (65 °C, OR 150 °F MAX)	**ERCUAL-A2 OR ERNI-3 (65 °C, OR 150 °F MAX)**
HIGH-CARBON STEEL	ERCUAL-A2 OR ERCU OR ERNI-3 (540 °C, OR 1000 °F)	ERCUSN-A (260 °C, OR 500 °F)	ERCUAL-A2 (260 °C, OR 500 °F)	ERCUAL-A2 (205 °C, OR 400 °F)	**ERCUAL-A2 OR ERNI-3 (65 °C, OR 150 °F MAX)**
LOW-ALLOY STEEL	ERCUAL-A2 OR ERCU OR ERNI-3 (540 °C, OR 1000 °F)	ERCUSN-A (260 °C, OR 500 °F)	ERCUAL-A2 (260 °C, OR 500 °F)	ERCUAL-A2 (205 °C, OR 400 °F)	**ERCUAL-A2 OR ERNI-3 (65 °C, OR 150 °F MAX)**
STAINLESS STEEL	**ERCUAL-A2 OR ERCU OR ERNI-3 (540 °C, OR 1000 °F)**	**ERCUSN-A (205 °C, OR 400 °F)**	**ERCUAL-A2 (65 °C, OR 150 °F MAX)**	**ERCUAL-A2 (65 °C, OR 150 °F MAX)**	**ERCUAL-A2 OR ERNI-3 (65 °C, OR 150 °F MAX)**

(A) FILLER-METAL SELECTIONS SHOWN IN TABLE ARE BASED ON WELDABILITY, EXCEPT WHERE MECHANICAL PROPERTIES ARE USUALLY MORE IMPORTANT. PREHEATING IS ORDINARILY USED ONLY WHEN AT LEAST ONE MEMBER IS THICKER THAN ABOUT 3.2 MM ($\frac{1}{8}$ IN.) OR IS HIGHLY CONDUCTIVE (SEE TEXT). PREHEAT AND INTERPASS TEMPERATURES ARE SUBJECT TO ADJUSTMENT ON THE BASIS OF THE SIZE AND SHAPE OF THE WELDMENT.

TABLE 25 FILLER METALS AND PREHEAT AND INTERPASS TEMPERATURES USED IN GAS-METAL ARC WELDING OF COPPERS AND COPPER-BASE ALLOYS TO STEELS

METAL TO BE WELDED	ELECTRODES (AND PREHEAT AND INTERPASS TEMPERATURES) FOR WELDING METAL IN COLUMN 1 TO(A):						
	COPPERS	**LOW-ZINC BRASSES**	**HIGH-ZINC BRASSES, TIN BRASSES, SPECIAL BRASSES**	**PHOSPHOR BRONZES**	**ALUMINUM BRONZES**	**SILICON BRONZES**	COPPER NICKELS
LOW-CARBON STEEL	ERCUAL-A2 OR ERCU OR ERNI-3 (540 °C, OR 1000 °F)	ERCUSN-A (315 °C, OR 600 °F)	ERCUAL-A2 (260 °C, OR 500 °F)	ERCUSN-A (205 °C, OR 400 °F)	ERCUAL-A2 (150 °C, OR 300 °F)	ERCUAL-A2 (65 °C, OR 150 °F MAX)	**ERCUAL-A2OR ERNI-3 (65 °C, OR 150 °F MAX)**
MEDIUM-CARBON STEEL	ERCUAL-A2 OR ERCU OR ERNI-3 (540 °C, OR 1000 °F)	ERCUAL-A2 (315 °C, OR 600 °F)	ERCUAL-A2 (260 °C, OR 500 °F)	ERCUSN-A (205 °C, OR 400 °F)	ERCUAL-A2 (205 °C, OR 400 °F)	ERCUAL-A2 (65 °C, OR 150 °F MAX)	**ERCUAL-A2 OR ERNI-3 (65 °C, OR 150 °F MAX)**
HIGH-CARBON STEEL	ERCUAL-A2 OR ERCU OR ERNI-3 (540 °C, OR 1000 °F)	ERCUAL-A2 (315 °C, OR 600 °F)	ERCUAL-A2 (260 °C, OR 500 °F)	ERCUSN-A (260 °C, OR 500 °F)	ERCUAL-A2 (260 °C, OR 500 °F)	ERCUAL-A2 (205 °C, OR 400 °F)	**ERCUAL-A2 OR ERNI-3 (65 °C, OR 150 °F MAX)**
LOW-ALLOY STEEL	ERCUAL-A2 OR ERCU OR ERNI-3 (540 °C, OR 1000 °F)	ERCUAL-A2 (315 °C, OR 600 °F)	ERCUAL-A2 (315 °C, OR 600 °F)	ERCUSN-A (260 °C, OR 500 °F)	ERCUAL-A2 (260 °C, OR 500 °F)	ERCUAL-A2 (205 °C, OR 400 °F)	**ERCUAL-A2, OR ERNI-3 (65 °C, OR 150 °F MAX)**
STAINLESS STEEL	**ERCUAL-A2 OR ERCU OR ERNI-3 (540 °C, OR 1000 °F)**	**ERCUAL-A2 OR ERCUSN-A (315 °C, OR 600 °F)**	**ERCUAL-A2 (315 °C, OR 600 °F)**	**ERCUSN-A (205 °C, OR 400 °F)**	**ERCUAL-A2 (65 °C, OR 150 °F MAX)**	**ERCUAL-A2 (65 °C, OR 150 °F MAX)**	**ERCUAL-A2 OR ERNI-3 (65 °C, OR 150 °F MAX)**

(A) ELECTRODE SELECTIONS IN TABLE ARE BASED ON WELDABILITY, EXCEPT WHERE MECHANICAL PROPERTIES ARE USUALLY MORE IMPORTANT. PREHEATING USUALLY IS USED ONLY WHEN AT LEAST ONE MEMBER IS THICKER THAN 3.2 MM ($\frac{1}{8}$ IN.) OR IS HIGHLY CONDUCTIVE (SEE TEXT). PREHEAT AND INTERPASS TEMPERATURES ARE SUBJECT TO ADJUSTMENT BASED ON SIZE AND SHAPE OF WELDMENT.

Aluminum-Base Alloys to Steels. With respect to fusion welding, iron and aluminum are not compatible metals. Their melting temperatures differ greatly: 660 °C (1220 °F) for aluminum versus 1538 °C (2800 °F) for iron. Both metals have almost no solubility for the other in the solid state, especially iron in aluminum, and several brittle intermetallic phases can form ($FeAl_2$, Fe_2Al_5, or $FeAl_3$). Consequently, fusion welds joining iron and aluminum are brittle. In addition, high welding stresses would be expected because of the significant differences in the thermal expansion coefficients, thermal conductivities, and specific heats of iron and aluminum. Aluminum can be joined to carbon or stainless steel by brazing, solid-state welding or by high-energy processes such as electron-beam welding.

References cited in this section

32. D.J. KOTECKI AND T.A. SIEWERT, WRC-1992 CONSTITUTION DIAGRAM FOR STAINLESS STEEL WELD METALS: A MODIFICATION OF THE WRC-1988 DIAGRAM, *WELD. J.,* MAY 1992, P 171S-178S
33. DISSIMILAR METALS, *WELDING HANDBOOK,* 7TH ED., VOL 4, AMERICAN WELDING SOCIETY, 1982, P 514-547
34. DISSIMILAR-WELD FAILURE ANALYSIS AND DEVELOPMENT PROGRAM, VOL 1-4, REPORT CS-4252, ELECTRIC POWER RESEARCH INSTITUTE, PALO ALTO, CA

J.R. Davis, Davis & Associates

References

1. K.C. ANTONY *ET AL.,* HARDFACING, *WELDING, BRAZING, AND SOLDERING,* 9TH ED., VOL 6, *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1983, P 771-793
2. P. CROOK AND H.N. FARMER, FRICTION AND WEAR OF HARDFACING ALLOYS, *FRICTION, LUBRICATION, AND WEAR TECHNOLOGY,* VOL 18, *ASM HANDBOOK,* ASM INTERNATIONAL, 1992, P 758-765
3. D.K. SUBRAMANYAM, A.E. SWANSIGER, AND H.S. AVERY, AUSTENITIC MANGANESE STEELS, *PROPERTIES AND SELECTION: IRONS, STEELS, AND HIGH-PERFORMANCE ALLOYS,* VOL 1, *ASM HANDBOOK,* ASM INTERNATIONAL, 1990, P 822-840
4. H.S. AVERY AND H.J. CHAPIN, *WELD J.,* VOL 31, 1952, P 917
5. R.D. ZORDAN, REPORT 10890, CABOT CORPORATION, KOKOMO, IN, 1982
6. H. OCKEN AND A.K. VELAN, ED., "LABORATORY EVALUATIONS OF COBALT-FREE, NICKEL-BASED HARDFACING ALLOYS FOR NUCLEAR APPLICATIONS," REPORT NP-4993, PROJECT 1935, ELECTRIC POWER RESEARCH INSTITUTE, PALO ALTO, CA, MARCH 1987
7. P.J. HOFMANN, B.C. FRIEDRICH, AND H. OCKEN, ED., "LABORATORY EVALUATIONS OF IRON-BASED HARDFACING ALLOYS," REPORT NP-5874, PROJECT 1935-11, ELECTRIC POWER RESEARCH INSTITUTE, PALO ALTO, CA, JUNE 1988
8. WELD DESIGN AND FABRICATION, DATA SHEET 420, *WELD. DES. FABR.,* VOL 48 (NO. 8), 1975, P 50
9. K.M. KULKARNI AND V. ANAND, METAL POWDERS USED FOR HARDFACING, *POWDER METALLURGY,* 9TH ED., VOL 7, *ASM HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1984, P 823-836
10. K.P. COOPER, LASER SURFACE PROCESSING, *FRICTION, LUBRICATION, AND WEAR TECHNOLOGY,* VOL 18, *ASM HANDBOOK,* ASM INTERNATIONAL, 1992, P 861-872
11. A.G. BLAKE, *ET AL.,* "LASER COATING TECHNOLOGY: A COMMERCIAL REALITY," BULLETIN OF QUANTUM LASER CORPORATION, EDISON, NJ

12. J.D. AYERS AND D.S. GNANAMUTHU, HARDENING OF METAL SURFACES BY LASER PROCESSING, *WELDING, BRAZING, AND SOLDERING,* 9TH ED., VOL 6, *ASM HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1983, P 793-803
13. W.M. STEEN, *PROC. NATO ADVANCED STUDY INSTITUTE ON LASER SURFACE TREATMENT OF METALS,* C.W. DRAPER AND P. MAZZOLDI, ED., MARTINUS NIJHOFF, 1986, P 369
14. J. MAZUMDER AND J. SINGH, *PROC. NATO ADVANCED STUDY INSTITUTE ON LASER SURFACE TREATMENT OF METALS,* C.W. DRAPER AND P. MAZZOLDI, ED., MARTINUS NIJHOFF, 1986, P 297
15. A. BELMONDO AND M. CASTAGNA, *SOURCEBOOK ON APPLICATIONS OF THE LASER IN METALWORKING,* E. METZBOWER, ED., AMERICAN SOCIETY FOR METALS, 1981, P 310
16. G.J. BRUCK, *J. MET.,* VOL 39 (NO. 2), 1987, P 10
17. A.V. LA ROCCA, *PROC. NATO ADVANCED STUDY INSTITUTE ON LASER SURFACE TREATMENT OF METALS,* C.W. DRAPER AND P. MAZZOLDI, ED., MARTINUS NIJHOFF, 1986, P 521
18. R.M. MACINTYRE, *PROC. NATO ADVANCED STUDY INSTITUTE ON LASER SURFACE TREATMENT OF METALS,* C.W. DRAPER AND P. MAZZOLDI, ED., MARTINUS NIJHOFF, 1986, P 545
19. B.A. KUSHNER AND E.R. NOVINSKI, THERMAL SPRAY COATINGS, *FRICTION, LUBRICATION, AND WEAR TECHNOLOGY,* VOL 18, *ASM HANDBOOK,* ASM INTERNATIONAL, 1992, P 829-833
20. R.H. UNGER, THERMAL SPRAY COATINGS, *CORROSION*, 9TH ED., VOL 13, *ASM HANDBOOK,* ASM INTERNATIONAL, 1987, P 459-462
21. M.L. THORPE, THERMAL SPRAY: INDUSTRY IN TRANSITION, *ADV. MATER. PROCESS.,* VOL 143 (NO. 5), 1993, P 50-56
22. G. IRONS, "HIGHER VELOCITY THERMAL SPRAY PROCESSES PRODUCE BETTER AIRCRAFT ENGINE COATINGS," SAE TECHNICAL PAPER 920947, PRESENTED AT THE 28TH ANNUAL AEROSPACE/AIRLINE PLATING AND METAL FINISHING FORUM AND EXPOSITION (SAN DIEGO, CA), APRIL 1992
23. F.N. LONGO, PROCESSING AND DESIGN, LESSON 3, COURSE 51, *THERMAL SPRAY TECHNOLOGY,* MATERIALS ENGINEERING INSTITUTE, ASM INTERNATIONAL, 1992
24. M.L. THORPE, THERMAL SPRAYING BECOMES A DESIGN TOOL, *MACH. DES.,* 24 NOV 1983
25. P.O. BUZBY AND J. NIKITICH, HVOF THERMAL SPRAYING OF NITRIDED PARTS, *ADV. MATER. PROCESS.,* DEC 1991, P 35-36
26. R.D. THOMAS *ET AL.,* WELD OVERLAYS, *WELDING, BRAZING, AND SOLDERING*, 9TH ED., VOL 6, *METALS HANDBOOK*, AMERICAN SOCIETY FOR METALS, 1983, P 804-819
27. D.L. OLSON, PREDICTION OF AUSTENITIC WELD METAL MICROSTRUCTURE AND PROPERTIES, *WELD. J.,* VOL 64 (NO. 10), 1985, P 2815-2955
28. A.L. SCHAEFFLER, CONSTITUTION DIAGRAM FOR STAINLESS STEEL WELD METAL, *MET. PROG.,* VOL 56 (NO. 11), 1949, P 680-680B
29. W.T. DELONG, FERRITE IN AUSTENITIC STAINLESS STEEL WELD METAL, *WELD. J.,* VOL 53 (NO. 7), 1974, P 2735-2865
30. T.A. SIEWERT, C.N. MCCOWAN, AND D.L. OLSON, FERRITE NUMBER PREDICTION TO 100 FN IN STAINLESS STEEL WELD METAL, *WELD. J.,* VOL 67 (NO. 12), 1988, P 2895-2985
31. D.J. KOTECKI, "VERIFICATION OF THE NBS-CSM FERRITE DIAGRAM," INTERNATIONAL INSTITUTE OF WELDING DOCUMENT II-C-834-88, 1988
32. D.J. KOTECKI AND T.A. SIEWERT, WRC-1992 CONSTITUTION DIAGRAM FOR STAINLESS STEEL WELD METALS: A MODIFICATION OF THE WRC-1988 DIAGRAM, *WELD. J.,* MAY 1992, P 171S-178S
33. DISSIMILAR METALS, *WELDING HANDBOOK,* 7TH ED., VOL 4, AMERICAN WELDING SOCIETY, 1982, P 514-547
34. DISSIMILAR-WELD FAILURE ANALYSIS AND DEVELOPMENT PROGRAM, VOL 1-4, REPORT

CS-4252, ELECTRIC POWER RESEARCH INSTITUTE, PALO ALTO, CA

Procedure Development and Practice Considerations for Resistance Welding

Introduction

RESISTANCE WELDING (RW) encompasses a group of processes in which the heat for welding is generated by the resistance to the flow of electrical current through the parts being joined. These processes are most commonly used to weld two overlapping sheets or plates that may have different thicknesses. A pair of electrodes clamp the workpieces under pressure to provide good electrical contact and to contain the molten metal in the joint. The electrodes conduct electrical current to the joint; resistance to the flow of current heats the faying surfaces, forming a weld (Ref 1).

The required heat generated through the electrical resistance of the two bodies to be joined is given by

$$H = I^2RT \qquad \textbf{(EQ 1)}$$

where H is the heat generated in joules; I is the current in amperes; R is the resistance in ohms; and t is the time of current flow, in seconds. The actual temperature rise at the joint depends on the interrelationship of the three component variables. These variables are affected by:

- COOLING EFFECT OF THE ELECTRODE
- EFFECT OF FOLLOW UP ON RESISTANCE
- EFFECT OF SHUNTING

Pressure is applied even before the current is activated to ensure a good bond. Accurate control and timing of the electric current and application of pressure is an essential feature of resistance welding.

The strength of the bond also depends on the cleanliness of the mating surfaces, although the process is more tolerant of the presence of oxide layers and contaminants than other welding processes. Resistance welding processes require specialized machinery, but they have the major advantages of not requiring filler metals, shielding gases, or flux.

The three major resistance welding processes are resistance spot welding (RSW), resistance seam welding (RSEW), and projection welding (PW). Flash welding (FW) is also classified as a resistance welding process, but it is unique in that the heat for welding is created at the faying surfaces of the joint not only by resistance to the flow of electric current, but also by arcs across the interface. Upset welding (UW) differs from flash welding in that no arc is generated by the process. Capacitor discharge stud welding incorporates a stud which serves as a weld member and is a process similar to flash welding. High-frequency resistance welding (HFRW) concentrates current of the surface of the part to heat the base metal and cause the coalescence of metal.

This article addresses procedure development and considerations for using these processes to join specific types of materials. Additional information about these processes can be found in the Section "Fusion Welding Processes" in this Volume.

Acknowledgements

Special thanks are due to David M. Beneteau of CenterLine (Windsor) Limited, who offered suggestions about revisions for portions of this article. His participation was arranged through the Resistance Welder Manufacturers Association.

Reference

1. L.P. CONNER, ED., *WELDING HANDBOOK*, 8TH ED., VOL 1, AWS, 1987, P 16

Procedure Development and Practice Considerations for Resistance Welding

Resistance Spot Welding

Resistance spot welding is the most common form of resistance welding. It is a process in which laying surfaces are joined in one or more overlapping spots by the heat generated by resistance to the flow of electric current through workpieces that are held together under force by electrodes. The contacting surfaces in the region of current concentration are heated by a short-time pulse of low-voltage, high-amperage current to form a fused nugget of weld metal. When the flow of current ceases, the electrode force is maintained while the weld metal rapidly cools and solidifies. The electrodes are retracted after each weld, which usually is completed in a fraction of a second.

The size and shape of the individually formed welds are limited primarily by the size and contour of the electrode faces. The weld nugget forms at the faying surfaces (Fig. 1), but does not extend completely to the outer surfaces. In a cross section, the nugget in a properly formed spot weld is obround or oval in shape; in plan view, it has the same shape as the electrode face (which usually is round) and approximately the same size. The spots should be at a sufficient distance from the edge of the workpiece (edge distance) so that there is enough base metal to withstand the electrode force and to ensure that the local distortion during welding does not allow expulsion of metal from the weld. Additional information is available in the article "Resistance Spot Welding" in this Volume.

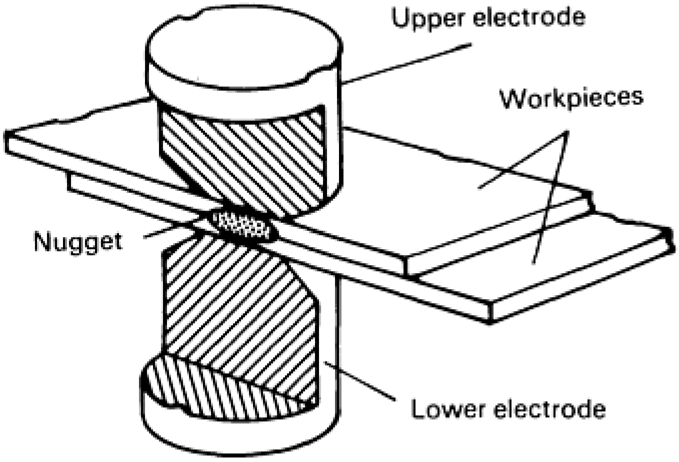

FIG. 1 CUTAWAY VIEW SHOWING LOCATION AND SHAPE OF WELD NUGGET RELATIVE TO THE INNER AND OUTER SURFACES OF WORKPIECES IN A RESISTANCE SPOT WELDING SETUP

RSW Applications

Weldability and the thickness of the metals joined are key factors in determining whether two materials can be resistance spot welded.

Material Selection (Ref 2). Practically all combinations of ductile metals and alloys can be spot welded (Table 1). Some, like copper to aluminum, and aluminum to magnesium, form alloys having little strength. Others, such as zinc and some of the high-chromium alloys, experience grain growth even during a very short welding period. Although welding times are relatively short, the material is adversely affected by the heat conducted through the sheet to the point that the physical properties such as the strength and corrosion resistance of the material surface are typically reduced in value. The short welding period is also necessary in stainless steel to prevent carbide precipitation when the carbon content is high enough to permit it. High-carbon steels weld readily, but the weld will be in the full-hardened state and will require subsequent heat treatment. This can be accomplished automatically by a control that applies preheating, postheating, or both as part of the welding cycle. To ensure that the control can apply the compensating quench or temper cycle, the electrodes must open immediately after the cycle is complete. Excessive hold time will quench the joint, negating the benefits associated with this control function.

TABLE 1 RELATIVE WELDABILITY RATINGS OF SELECTED METAL AND ALLOY COMBINATIONS USED FOR RESISTANCE SPOT WELDING OPERATIONS

A, excellent; B, good; C, fair; D, poor; E, very poor; F, impractical

METALS	**ALUMINUM**	**STAINLESS STEEL**	**BRASS**	**COPPER**	**GALVANIZED IRON**	**STEEL**	**LEAD**	**MONEL**	**NICKEL**	**NICHROME**	**TINPLATE**	**ZINC**	**PHOSPHOR BRONZE**	**NICKEL SILVER**	**TERNEPLATE**
ALUMINUM	B	E	D	E	C	D	E	D	D	D	C	C	C	F	C
STAINLESS STEEL	F	A	E	E	B	A	F	C	C	C	B	F	D	D	B
BRASS	D	E	C	D	D	D	F	C	C	C	D	E	C	C	D
COPPER	E	E	D	F	E	E	E	D	D	D	E	E	C	C	E
GALVANIZED IRON	C	B	D	E	B	B	D	C	C	C	B	C	D	E	B
STEEL	D	A	D	E	B	A	E	C	C	C	B	F	C	D	A
LEAD	E	F	F	E	D	E	C	E	E	E	. . .	C	E	E	D
MONEL	D	C	C	D	C	C	E	A	B	B	C	F	C	B	C
NICKEL	D	C	C	D	C	C	E	B	A	B	C	F	C	B	C
NICHROME	D	C	C	D	C	C	E	B	B	A	C	F	D	B	C
TINPLATE	C	B	D	E	B	B	. . .	C	C	C	C	C	D	D	C
ZINC	C	E	E	E	C	F	C	F	F	F	C	C	D	F	C
PHOSPHOR BRONZE	C	D	C	C	D	C	E	C	C	D	D	D	B	B	D
NICKEL SILVER	F	D	C	C	E	D	E	B	B	B	D	F	B	A	D
TERNEPLATE	C	B	D	E	B	A	D	C	C	C	C	C	D	D	B

Source: Ref 3

Zinc-alloy die castings can be welded with little loss of strength, but the ductility is reduced. Free-turning Bessemer screw-machine steel frequently refuses to weld, or results in brittle welds, and hence its use for welded parts should be avoided. Usually, however, when clean, properly prepared metal parts are spot welded, the strength of the welds is perfectly satisfactory. Table 1 lists combinations that have been successfully welded. Many other combinations may also be satisfactorily welded. Copper and silver are difficult to weld, but they may be welded by the use of low-conductivity electrodes (for example, refractory alloys such as Elkonite and Trodaloy). Copper alloys are more commonly resistance brazed. In many applications, the weldability of copper is increased if it has a tinned surface.

Thickness Parameters (Ref 4). Parts of widely different thicknesses may be spot welded. For example, 0.4 mm ($\frac{1}{64}$ in.) material can be welded to a 150 mm (6 in.) thick piece and would require only slightly more power input and pressure than to weld two pieces of 0.4 mm ($\frac{1}{64}$ in.) material, as it is not necessary to heat the 150 mm (6 in.) piece through, and the push-up is obtained on the thinner piece. The thickness of the thinner component is referred to as the governing metal thickness because it dictates the heat input. However, the thermal mass of the larger component may make RSW applications impractical because the thick section may not reach fusion temperature before the thin piece is melted. The limit to welding pieces of equal thickness with an uninterrupted flow of current seems to be approximately 3.2 mm ($\frac{1}{8}$ in.). Greater thicknesses may be welded by a technique known as pulsation welding. In the pulsation welding method, the current is turned off during the weld at intervals that allow the electrodes to cool the electrode/workpiece surface. This ensures that the heat will be generated as desired, primarily at the faying surface.

Sheet Steel Limitations. Spot welded lap joints are widely used in joining sheet steel up to about 3.2 mm ($\frac{1}{8}$ in.) thick (Table 2) and are used occasionally in joining steel 6.4 mm ($\frac{1}{4}$ in.) or more in thickness, *t*. Thicknesses of 25 mm (1 in.) or more have been joined by spot welding, but this requires special equipment and would not ordinarily be economical. Many assemblies of two or more sheet metal stampings that do not require gastight or liquid-tight joints can be more economically joined by high-speed RSW than by mechanical methods.

TABLE 2 MINIMUM DIMENSIONS FOR SINGLE-IMPULSE SPOT WELDS IN LOW-CARBON STEEL

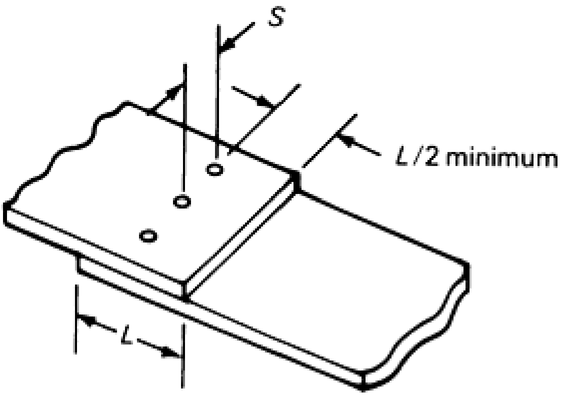

STOCK THICKNESS		**MINIMUM WELD SPACING, *S***		MINIMUM OVERLAP OF MATERIAL, *L*	
mm	**in.**	**mm**	**in.**	**mm**	in.
0.25	0.010	6.3	0.250	9.5	**0.375**
0.53	0.021	9.5	0.375	11	**0.433**
0.80	0.031	13	0.500	11	**0.433**
1.0	0.040	19	0.750	13	**0.500**
1.3	0.050	22	0.875	14	**0.551**
1.6	0.063	27	1.062	16	**0.625**

2.0	0.078	35	1.375	17	**0.669**
2.4	0.094	41	1.625	19	**0.750**
2.8	0.109	46	1.813	21	**0.827**
3.2	**0.125**	**51**	**2.000**	**22**	**0.875**

Source: Ref 5

For aluminum and magnesium welds, minimum spot spacing, *S,* should equal 8*t.* A spot spacing value of 16*t* would be even better if the strength requirements of the joint permit such spacing (Ref 5).

Joint Design Recommendations. Like other welds, resistance spot welds are subject to shrinkage on cooling. To minimize distortion, welds should be positioned to balance shrinkage forces (Fig. 2).

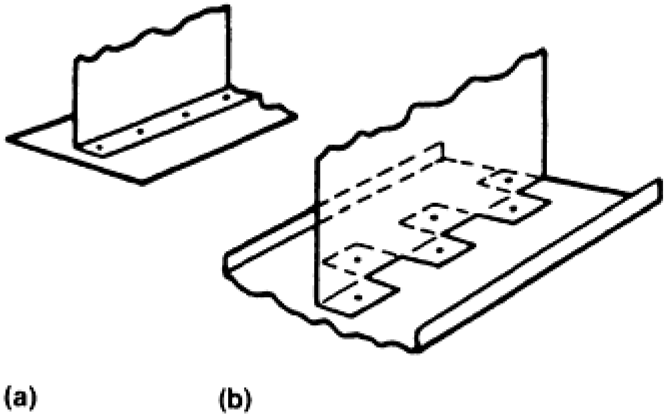

FIG. 2 PLACEMENT OF WELDS TO JOIN TWO SHEETS OF METAL IN A T-JOINT. (A) CONFIGURATION NOT RECOMMENDED DUE TO UNBALANCED SHRINKAGE FORCES. (B) PREFERRED SETUP THAT RESULTS IN BALANCED FORCES AND MINIMUM DISTORTION

Distortion can also result from electrode skidding. This is related to stress introduced at the time of electrode impact and machine deflection under high electrode force. The components must be designed to provide methods of accommodating variations (for example, by incorporating slip joints and minimizing material stackups).

Even when welding of multiple spot thicknesses is feasible, it can be more troublesome than simple welds of two sheet thicknesses. Scalloped edges permit alternate welding of spots in two thicknesses only (Fig. 3). However, scalloped edges are rarely used and the current trend is toward triple-metal-thickness welding as a method of reducing capital cost.

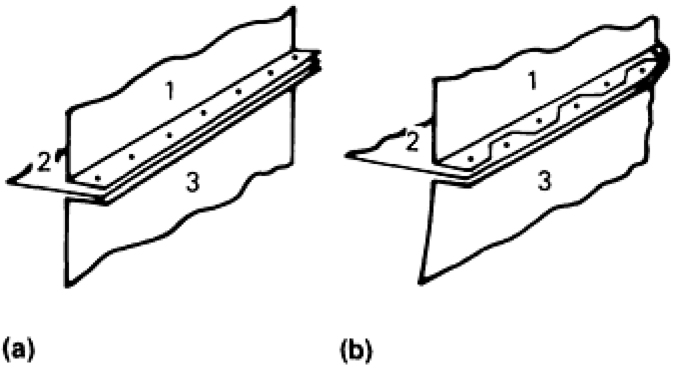

FIG. 3 PLACEMENT OF WELDS TO JOIN THREE SHEETS OF METAL IN A T-JOINT. (A) CONFIGURATION THAT CAN BE USED TO JOIN THREE STOCK THICKNESSES. (B) RECOMMENDED CONFIGURATION INCORPORATING

SCALLOPED EDGES TO PERMIT ALTERNATE WELDING OF SPOTS IN TWO THICKNESSES ONLY

Spot-welded assemblies should be designed so that joint areas are readily accessible to the electrode and simple, inexpensive electrodes can be used. When special electrode shapes are required, every effort should be made to avoid extremes of shape and to provide as accessible a joint as possible. Deep, narrow joint areas should be avoided. Figure 4 provides guidelines for accessibility of joints.

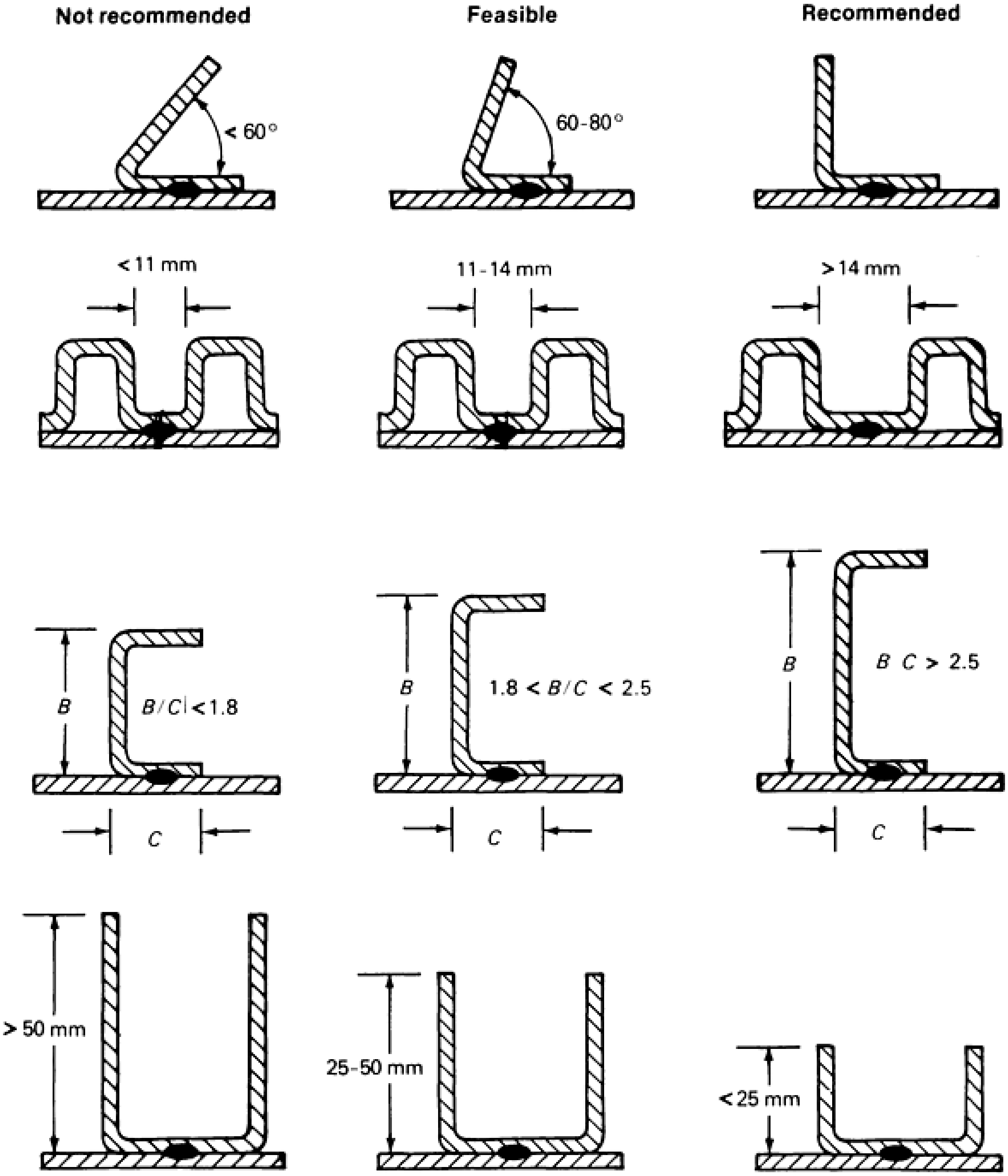

Not recommended
Feasible
Recommended
< 60°
60-80°
< 11 mm
11-14 mm
> 14 mm
B
B/C < 1.8
C
B
1.8 < B/C < 2.5
C
B
B C > 2.5
C
> 50 mm
25-50 mm
< 25 mm

FIG. 4 GRADUAL IMPROVEMENTS IN JOINT DESIGN THAT PROVIDE OPTIMUM ACCESS TO SELECTED JOINT CONFIGURATIONS USING STANDARD ELECTRODES

Weld Strength in RSW Joints. The strength of a single spot weld in shear depends on the cross-sectional area of the nugget in the plane of the faying surfaces.

Lap joints may fail either by shear through the nugget, or by tearing of the base metal adjacent to the weld nugget. Low weld strengths are typically associated with weld nugget shear failure, and high strengths are typically associated with base metal tearing. A minimum nugget diameter that is unique to the type of base material, surface condition, and metal gage used, is required in order to obtain failure by base metal tearing.

Increasing the nugget diameter above this minimum value can produce an increase in the weld strength. Figure 5 shows the slight increase in strength values obtained for low-carbon steel as a result of increases in nugget size.

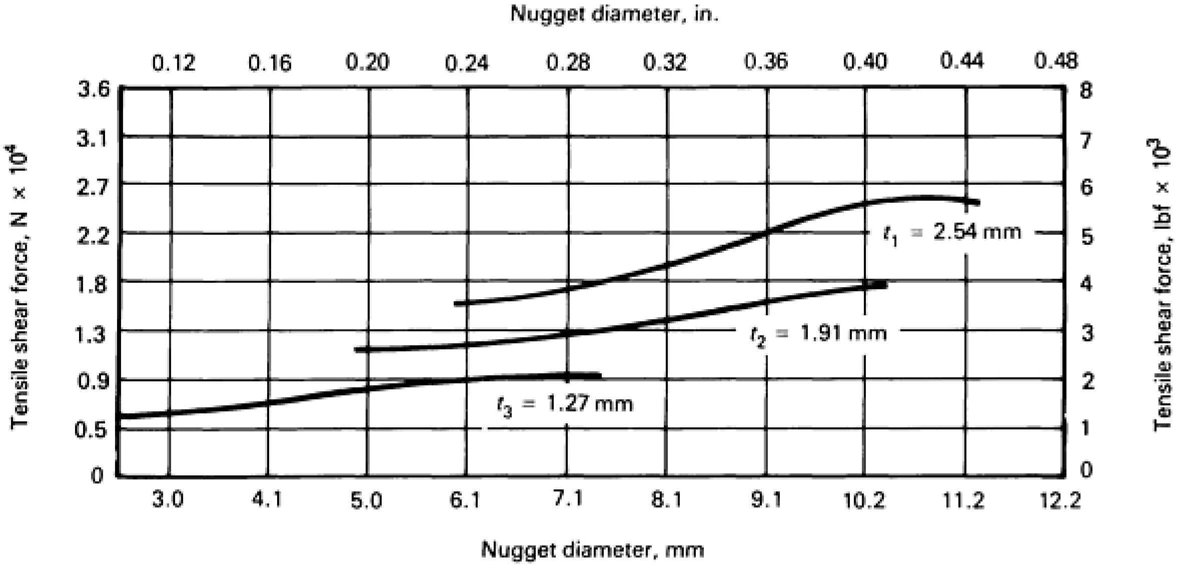

FIG. 5 PLOT OF TENSILE SHEAR STRENGTH VERSUS WELD NUGGET DIAMETER AS A FUNCTION OF SHEET THICKNESS, *T*, FOR 1008 LOW-CARBON STEEL

The strength of multiple spot-welded joints depends on material thickness, spot weld spacing, and weld pattern. The spacing between adjacent spot welds may alter the joint weld strength due to current shunting through previous welds. As the spacing between adjacent spot welds decreases, the joint shear strength may decrease.

Table 3 summarizes the relationship between resistance spot welding variables and joint strength for uncoated low-carbon steel.

TABLE 3 RECOMMENDED PARAMETERS FOR RESISTANCE SPOT WELDING OF UNCOATED LOW-CARBON STEEL SHEET PRODUCED FOR AUTOMOTIVE APPLICATIONS

Data are average values.

SHEET THICKNESS		**ELECTRODE**				**FORCE**[C]		**WELD TIME, CYCLES (60 HZ)**	**WELDING CURRENT, A**	**MINIMUM CONTACT OVERLAP**		**MINIMUM WELD SPACING**[D]				**MINIMUM SHEAR STRENGTH**		BUTTON DIAMETER	
		FACE DIAMETER		**SHAPE**[A]	**BEVEL ANGLE**[B]							**TWO STACK**		**THREE STACK**					
MM	**IN.**	**MM**	**IN.**			**KN**	**LBF**			**MM**	**IN.**	**MM**	**IN.**	**MM**	**IN.**	**KN**	**LBF**	**MM**	IN.
0.51	0.020	4.78	0.188	E,A,B	45 °	1.8	400	7	8,500	11.2	0.44	9.7	0.38	15.8	0.62	1.43	320	2.5	**0.10**
0.64	0.025	4.78	0.188	E,A,B	45 °	2.0	450	8	9,500	11.9	0.47	15.8	0.62	22.4	0.88	2.02	450	3.1	**0.12**
0.76	0.030	6.35	0.250	E,A,B	45 °	2.2	500	9	10,500	11.9	0.47	15.8	0.62	22.4	0.88	2.58	575	3.6	**0.14**
0.89	0.035	6.35	0.250	E,A,B	45 °	2.7	600	9	11,500	13.5	0.53	19.1	0.75	26.9	1.06	3.36	750	4.1	**0.16**
1.02	0.040	6.35	0.250	E,A,B	45 °	3.1	700	10	12,500	13.5	0.53	19.1	0.75	26.9	1.06	4.14	925	4.6	**0.18**
1.14	0.045	6.35	0.250	E,A,B	45 °	3.4	750	11	13,000	15.0	0.59	24.9	0.94	30.0	1.18	5.15	1150	4.8	**0.19**
1.27	0.050	7.92	0.312	E,A,B	30°	3.6	800	12	13,500	15.0	0.59	24.9	0.94	30.0	1.18	6.05	1350	5.1	**0.20**
1.40	0.055	7.92	0.312	E,A,B	30°	4.0	900	13	14,000	16.0	0.63	26.9	1.06	33.3	1.31	7.53	1680	5.3	**0.21**
1.52	0.060	7.92	0.312	E,A,B	30°	4.5	1000	14	15,000	16.0	0.63	26.9	1.06	33.3	1.31	8.29	1850	5.8	**0.23**
1.78	0.070	7.92	0.312	E,A,B	30°	5.4	1200	16	16,000	16.8	0.66	30.0	1.18	38.1	1.50	10.3	2300	6.4	**0.25**
2.03	0.080	7.92	0.312	E,A,B	30°	6.3	1400	18	17,000	18.3	0.72	35.1	1.38	40.6	1.60	12.1	2700	6.6	**0.26**
2.29	0.090	9.53	0.375	E,A,B	30°	7.2	1600	20	18,000	19.8	0.78	39.6	1.56	47.8	1.88	15.5	3450	6.9	**0.27**
2.67	0.105	9.53	0.375	E,A,B	30°	8.1	1800	23	19,500	21.3	0.84	42.7	1.68	50.8	2.00	18.6	4150	7.1	**0.28**
3.05	**0.120**	**9.53**	**0.375**	**E,A,B**	**30°**	**9.4**	**2100**	**26**	**21,000**	**22.4**	**0.88**	**46.0**	**1.81**	**63.5**	**2.50**	**22.4**	**5000**	**7.6**	0.30

Source: Ref 6

(A) SHAPE DEFINITIONS: E, TRUNCATED CONE; A, POINTED; B, 75 MM (3 IN.) RADIUS.

(B) APPLIES TO TRUNCATED CONE ELECTRODES ONLY AND IS MEASURED FROM THE PLANE OF THE ELECTRODE FACE.

(C) FOR INTERMEDIATE THICKNESSES, FORCE AND WELD TIME MAY BE INTERPOLATED.

(D) MINIMUM WELD SPACING IS MEASURED FROM CENTERLINE TO CENTERLINE.

Power Requirements. Nearly all RSW of low-carbon steel is direct-energy welding, in which single- or three-phase 60-cycle alternating current, drawn ordinarily from 220 or 440 V implant power lines and stepped down to about 2 to 20 V, is fed directly to the electrodes as each weld is made.

Direct and Series Welding

Spot welds are classified as either direct or series welds.

Direct Single-Spot Welding. Single-spot welds are usually made by direct welding. Figure 6 shows schematically three arrangements used for making this type of weld; these arrangements may be modified to meet special requirements. In all of the arrangements shown, one transformer secondary circuit makes one spot weld.

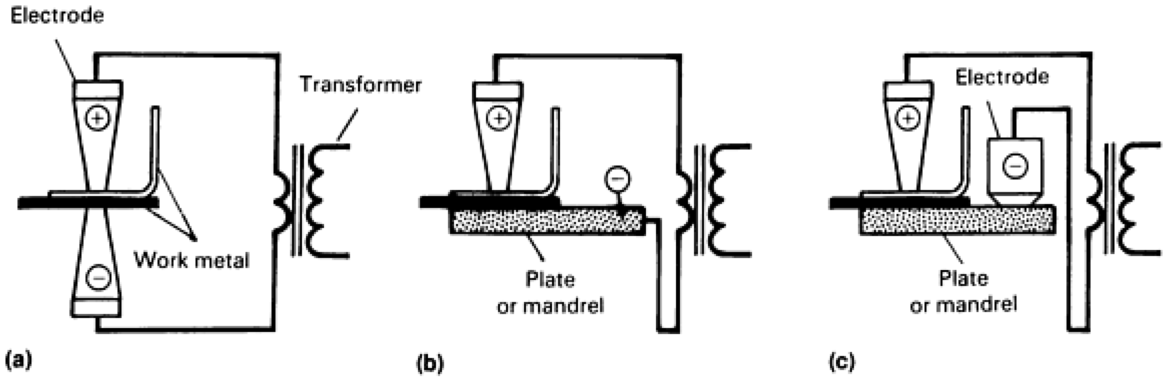

FIG. 6 RELATION OF WORK METAL TO ELECTRODES WHEN MAKING SINGLE-SPOT WELDS BY DIRECT WELDING

The simplest and most common arrangement in which two workpieces are sandwiched between opposing upper and lower electrodes, is shown in Fig. 6(a). In Fig. 6(b), a conductive plate or mandrel having a large contacting surface is used as the lower electrode; this reduces marking on the lower workpiece and conducts heat away from the weld more rapidly and may be necessary because of the shape of the workpiece. In the arrangement in Fig. 6(c), a conductive plate or mandrel beneath the lower workpiece is used for the same purposes but in conjunction with a second upper electrode.

Direct Multiple-Spot Welding. Two arrangements of the secondary circuit for making two or more spot welds simultaneously by direct welding are shown in Fig. 7(a) and 7(b). In direct multiple-spot welding (as well as in series multiple-spot welding, described below), tip contour and surface condition must be the same for each electrode. Also, the force exerted by all the electrodes on the workpieces must be equal, regardless of inequalities in work-metal thickness. The force can be equalized by using a spring-loaded electrode holder or a hydraulic equalizing system. The use of a conductive plate or mandrel, as in Fig. 7(b), minimizes weld marks on the lower workpiece.

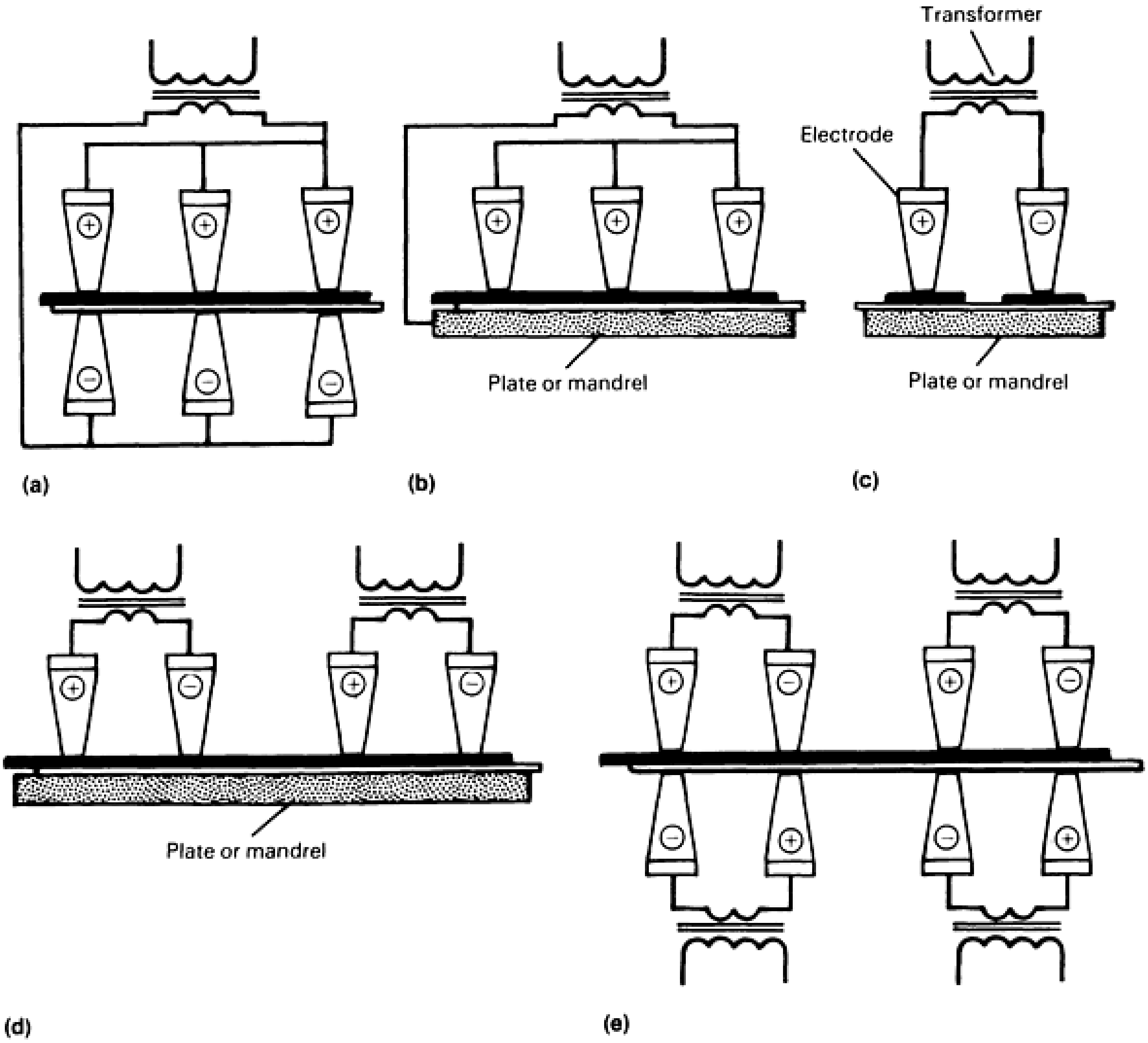

FIG. 7 RELATION OF WORK METAL TO ELECTRODES WHEN MAKING MULTIPLE-SPOT WELDS. (A) AND (B) DIRECT WELDING. (C), (D), AND (E) SERIES WELDING

Series Multiple-Spot Welding. Three arrangements for making a number of spot welds simultaneously by series welding are shown in Fig. 7(c), 7(d), and 7(e). In Fig. 7(d), each of the two transformer secondary circuits makes two spot welds. A portion of the current bypasses the weld nuggets through the upper workpiece.

Push-Pull Welding. Figure 7(e) is commonly referred to as push-pull welding. The advantage to this process is that the secondary loop area is quite small. This is common for components such as floor pans where a normal welding unit would have a throat several feet deep.

References cited in this section

2. J.F. YOUNG AND R.S. SHANE, ED., *MATERIALS AND PROCESSES, PART B: PROCESSES,* 3RD ED., MARCEL DEKKER, 1985, P 11257-1126
3. J.G. BRALLA, ED., *HANDBOOK OF PRODUCT DESIGN FOR MANUFACTURING,* MCGRAW-HILL, 1986, P 7-35
4. J.F. YOUNG AND R.S. SHANE, ED., *MATERIALS AND PROCESSES, PART B: PROCESSES,* 3RD ED., MARCEL DEKKER, 1985, P 1127

5. J.G. BRALLA, ED., *HANDBOOK OF PRODUCT DESIGN FOR MANUFACTURING,* MCGRAW-HILL, 1986, P 7-38
6. R.L. O'BRIEN, ED., *WELDING HANDBOOK,* 8TH ED., VOL 2, AWS, 1991, P 552

Procedure Development and Practice Considerations for Resistance Welding

Resistance Seam Welding

Resistance seam welding is a process in which heat caused by resistance to the flow of electric current in the work metal is combined with pressure to produce a welded seam. This seam, consisting of a series of overlapping spot welds, is normally gastight or liquid-tight. Two rotating, circular electrodes (electrode wheels), or one circular and one bar-type electrode, are used for transmitting the current to the work metal. When two electrode wheels are used, one or both wheels are driven either by means of a gear-driven shaft or by a knurl or friction drive that contacts the peripheral surface of the electrode wheel.

The series of spot welds is made without retracting the electrode wheels or releasing the electrode force between spots, although the electrode wheels may advance either continuously or intermittently.

Additional information is available in the article "Resistance Seam Welding" in this Volume.

Seam Welding Machines

A seam welding machine is similar in construction to a spot welding machine, except that one or two electrode wheels are substituted for the spot welding electrodes. Generally, seam. Welding is done in a press-type resistance welding machine.

There are four basic types of RSEW machines:

- *CIRCULAR:* AXIS OF ROTATION OF THE ELECTRODE WHEELS IS AT RIGHT ANGLES TO THE FRONT OF THE MACHINE (FIG. 8A).
- *LONGITUDINAL:* AXIS OF ROTATION OF THE ELECTRODE WHEELS IS PARALLEL TO THE FRONT OF THE MACHINE (FIG. 8B), AND THROAT DEPTH IS TYPICALLY 305 TO 915 MM (12 TO 36 IN.).
- *UNIVERSAL:* A SWIVEL-TYPE HEAD AND INTER-CHANGEABLE LOWER ARMS ALLOW THE AXIS OF ROTATION OF THE ELECTRODE WHEELS TO BE SET EITHER AT RIGHT ANGLES OR PARALLEL TO THE FRONT OF THE MACHINE.
- *PORTABLE:* WORK IS CLAMPED IN A FIXTURE, AND A PORTABLE WELDING HEAD IS MOVED OVER THE SEAM. THIS TYPE OF MACHINE IS USED FOR WORKPIECES THAT ARE TOO BULKY TO BE HANDLED BY REGULAR MACHINES

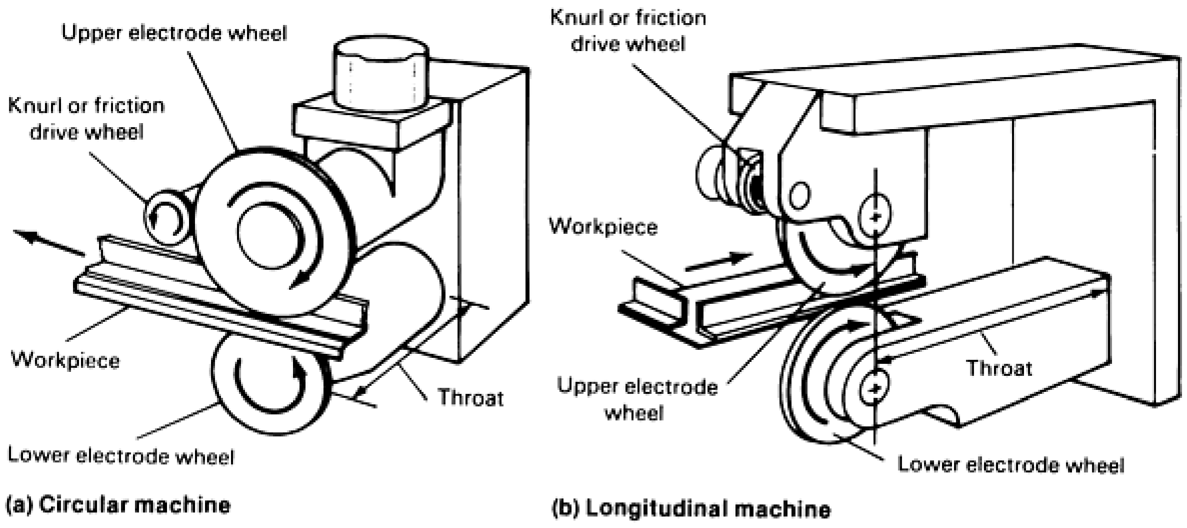
Upper electrode wheel
Knurl or friction drive wheel
Workpiece
Throat
Lower electrode wheel
(a) Circular machine
Knurl or friction drive wheel
Workpiece
Upper electrode wheel
Throat
Lower electrode wheel
(b) Longitudinal machine

FIG. 8 RELATION OF ELECTRODE WHEELS TO OTHER COMPONENTS IN CIRCULAR AND LONGITUDINAL RSEW MACHINES

Electrodes

Electrode wheels made of Resistance Welder Manufacturers Association (RWMA) class 1 material were once used for seam welding of aluminum and magnesium alloys, galvanized steel, and tin-plated steel. However, class 1 material is now uncommon because of the restrictions on cadmium exposure. Typically, class 2, 3, or 20 materials are used (similar to spot welding electrodes).

Electrode wheels range in diameter from 50 to 610 mm (2 to 24 in.). Table 4 shows the diameters and body widths of commonly used electrodes, in relation to the size of the welding machine. Narrower wheels are used in machines with knurl or friction drive; wider wheels, in gear drive and idler machines.

TABLE 4 COMMON SIZES OF ELECTRODE WHEELS USED FOR RESISTANCE SEAM WELDING APPLICATIONS

MACHINE SIZE	**WHEEL DIAMETER**		WHEEL WIDTH(A)	
	mm	**in.**	**mm**	in.
SMALL	18	7	9.5	$\frac{3}{8}$
MEDIUM	20	8	9.5-13	$\frac{3}{8}-\frac{1}{2}$
LARGE	**25-30**	**10-12**	**9.5-19**	$\frac{3}{8}-\frac{3}{4}$

(A) DATA ARE FOR BODY WIDTH.

A small-diameter electrode wheel (Fig. 9a), or a large-diameter wheel mounted on a canted axis (Fig. 9b), can be used to avoid interference with a sidewall when a narrow flange is being welded on an inside radius or on a re-entrant curve. Beads, fibs, and other extensions on the surfaces to be seam welded can be passed over by cutting notches in the electrode wheels (Fig. 10).

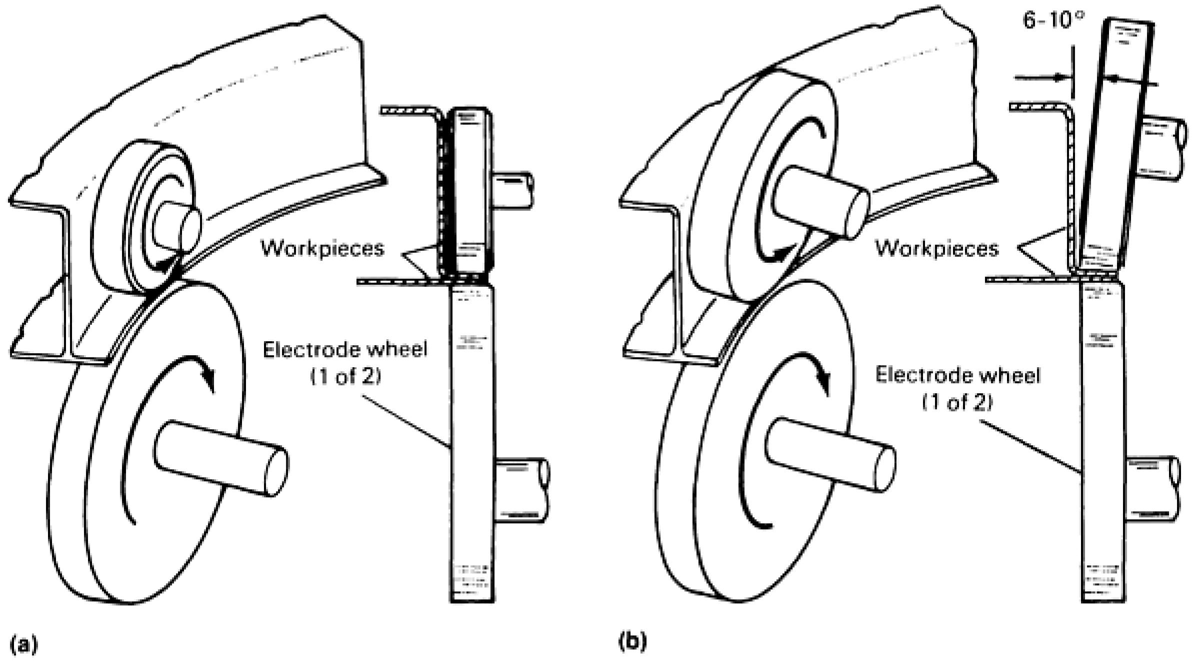
6-10°
Workpieces
Electrode wheel
(1 of 2)
Workpieces
Electrode wheel
(1 of 2)
(a)
(b)

FIG. 9 UPPER ELECTRODE WHEELS DESIGNED TO AVOID INTERFERENCE WITH A SIDEWALL. (A) SMALL-DIAMETER WHEEL. (B) CANTED LARGE-DIAMETER WHEEL

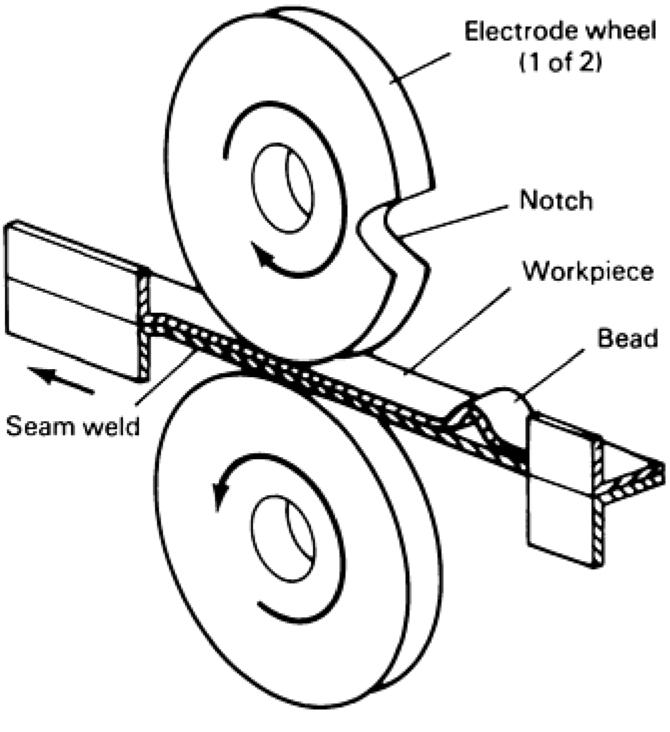
Electrode wheel
(1 of 2)
Notch
Workpiece
Bead
Seam weld

FIG. 10 NOTCH ELECTRODE WHEEL USED TO SEAM WELD A WORKPIECE HAVING AN OBSTRUCTION IN THE PATH OF THE WHEEL

Process Variations

Seam welding operations vary primarily in terms of travel speed and duration of the welding current. In continuous-motion welding, the electrodes (or workpieces) are driven at a constant speed, and welding current is either interrupted or flows continuously. In either case, a continuous, leaktight seam is generated. This method is generally used for metals less than about 4.8 mm ($\frac{3}{16}$ in.) thick.

In intermittent-motion welding, rotation of the electrode wheels is automatically stopped to produce a spot weld; the electrodes are then rotated to move the workpiece the required distance for the next weld. This method is often used for aluminum and steel alloys less than 4.8 mm ($\frac{3}{16}$ in.) thick and selected metals greater than 4.8 mm ($\frac{3}{16}$ in.) thick.

Lap Seam Welding. In this process (also known as stitch or roll welding), the on and off cycling of the current while the electrodes continuously rotate actually determines the weld spacing (Fig. 11a). The minimum joint overlap is the same as for spot welding. There is generally no forging (that is, mash-down) with lap seam welding.

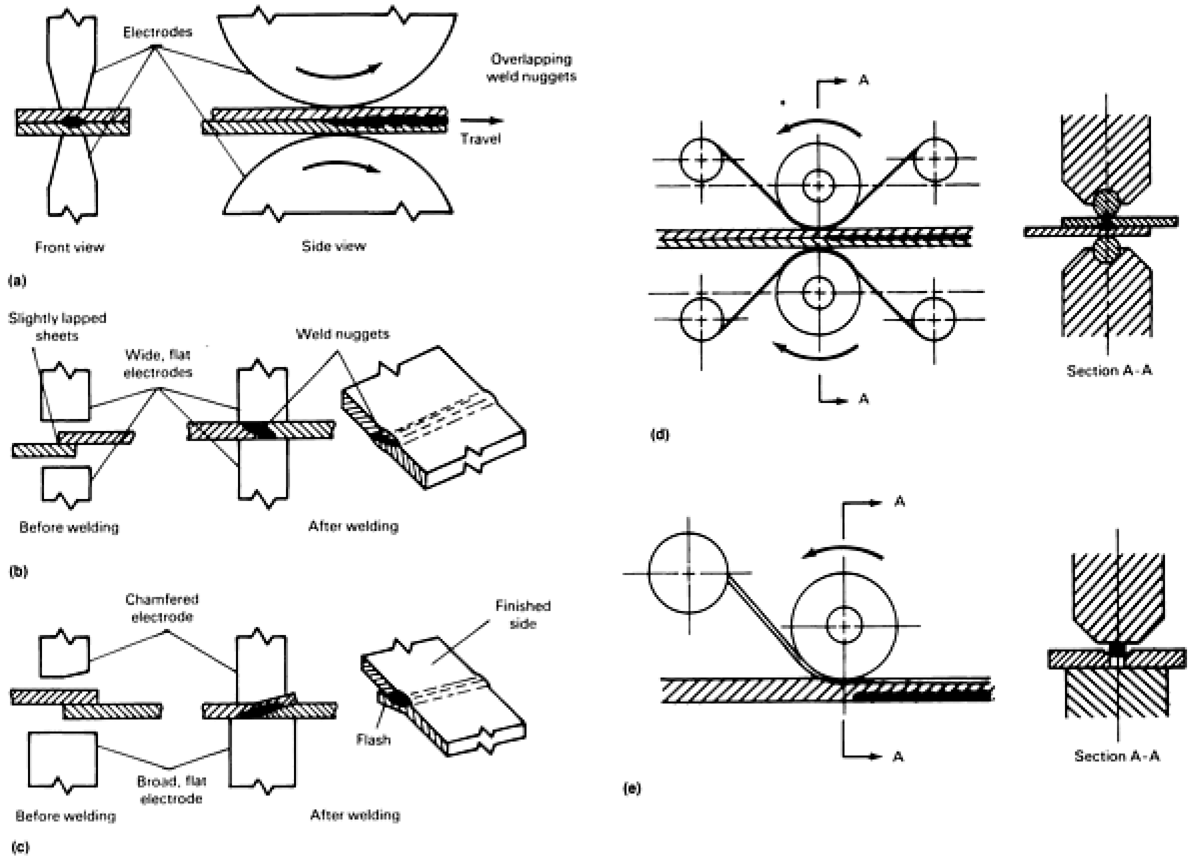
Electrodes
Overlapping weld nuggets
Travel
Front view
Side view
(a)
Slightly lapped sheets
Wide, flat electrodes
Weld nuggets
Before welding
After welding
(b)
Chamfered electrode
Finished side
Flash
Broad, flat electrode
Before welding
After welding
(c)
A
A
Section A-A
(d)
A
A
Section A-A
(e)

FIG. 11 PROCESS VARIATIONS OF SEAM WELDING. (A) LAP SEAM WELDING. (B) MASH SEAM WELDING. (C) METAL FINISH SEAM WELDING. (D) ELECTRODE WIRE SEAM WELDING. (E) FOIL BUTT SEAM WELDING

Mash seam welding generally has less overlap than lap seam welding and produces more mash-down (electrode force) with its continuous current (Fig. 11b). Joint thicknesses after welding usually vary from 120 to 150% of the sheet thickness. Continuous seams produced by this method have smooth appearance, but rigid workholding fixtures are required to minimize distortion.

Metal Finish Seam Welding. This process uses a chamfered electrode wheel, and one side of the joint is mashed down while the other side of the joint has a smooth, flat finish (Fig. 11c). Overlap is greater than for mash seam welding, and higher current amperages and electrode forces are also required.

Electrode wire seam welding is an internationally patented process that requires an intermediate wire electrode between each wheel electrode and the workpiece (Fig. 11d). The electrode wire seam welding process is used almost exclusively for seam welding of tin mill products to fabricate cans. The copper wire travels around the wheel electrodes at the welding speed. The circular or flat cross-section of copper wire electrode provides a continuously renewed surface, but it is not consumed in the welding operation. The tin build-up which would occur on a copper wheel electrode is thus avoided. The process requires specially designed welding systems. The seam welds may be made with two wheel electrodes, or one wheel electrode and a mandrel electrode.

Foil Butt Seam Welding. This process requires metal foil to be placed on one or both sides of a butt joint prior to seam welding (Fig. 11e). The foil distributes current to the edges of the sheet, provides additional electrical resistance, helps to contain the molten metal, and also serves as a filler.

Applications

Resistance seam welding can be applied to a variety of workpiece shapes. Girth welds can be made in round, square, or rectangular parts by using electrode wheels of suitable diameter. Longitudinal welds can also be made.

Advantages of RSEW, as compared with resistance spot and projection welding, are:

- GASTIGHT OR LIQUID-TIGHT JOINTS CAN BE PRODUCED
- OVERLAP CAN BE LESS THAN FOR RESISTANCE SPOT WELDS OR PROJECTION WELDS, AND SEAM WIDTH CAN BE LESS THAN THE DIAMETER OF SPOT OR PROJECTION WELDS (TABLE 5)
- ROLLED SECTIONS ARE FREQUENTLY WELDED IN-LINE, THEREBY ELIMINATING SUBSEQUENT OPERATIONS

TABLE 5 RECOMMENDED PRACTICE FOR SEAM WELDING OF 1010 LOW-CARBON STEEL

Data are for material free of scale, oxide, paint, grease, and oil. Welding conditions determined by thickness of thinnest outside piece.

THICKNESS,*T*, OF THINNEST OUTSIDE PIECE(A)		**ELECTRODE DIMENSIONS**				**NET ELECTRODE FORCE**		**HEAT TIME, CYCLES (60 HZ)**	**COOL TIME (PRESSURE TIGHT), CYCLES (60 HZ)**	**WELDING SPEED**		**WELDS PER MM**	**WELDS PER IN.**	**APPROXIMATE WELDING CURRENT, A**	MINIMUM CONTACTING OVERLAP(C)	
		WHEEL WIDTH (*W*)(B)		**FACE WIDTH (*W*)(B)**												
mm	**in.**	**mm**	**in.**	**mm**	**in.**	**kn**	**lbf**			**mm/min**	**in./min**				**MM**	IN.
0.25	0.010	9.5	$\frac{3}{8}$	4.8	$\frac{3}{16}$	1.8	400	2	1	200	80	0.60	15	8,000	9.5	$\frac{3}{8}$
0.53	0.021	9.5	$\frac{3}{8}$	4.8	$\frac{3}{16}$	2.5	550	2	2	190	75	0.48	12	11,000	11	$\frac{7}{16}$
0.80	0.031	13	$\frac{1}{2}$	6.4	$\frac{1}{4}$	3.1	700	3	2	185	72	0.40	10	13,000	13	$\frac{1}{2}$
1.02	0.040	13	$\frac{1}{2}$	6.4	$\frac{1}{4}$	4.0	900	3	3	170	67	0.36	9	15,000	13	$\frac{1}{2}$
1.27	0.050	13	$\frac{1}{2}$	7.9	$\frac{5}{16}$	4.7	1050	4	3	165	65	0.32	8	16,500	14	$\frac{9}{16}$
1.57	0.062	13	$\frac{1}{2}$	7.9	$\frac{5}{16}$	5.3	1200	4	4	160	63	0.28	7	17,500	16	$\frac{5}{8}$
1.98	0.078	16	$\frac{5}{8}$	9.5	$\frac{3}{8}$	6.7	1500	6	5	140	55	0.24	6	19,000	17.5	$\frac{11}{16}$
2.39	0.094	16	$\frac{5}{8}$	11	$\frac{7}{16}$	7.6	1700	7	6	125	50	0.24	6	20,000	19	$\frac{3}{4}$
2.77	0.109	19	$\frac{3}{4}$	13	$\frac{1}{2}$	8.7	1950	9	6	120	48	0.20	5	21,000	20.5	$\frac{13}{16}$
3.18	**0.125**	**19**	$\frac{3}{4}$	**13**	$\frac{1}{2}$	**9.8**	**2200**	**10**	**7**	**115**	**45**	**0.20**	**5**	**22,000**	**22**	$\frac{7}{8}$

(A) DATA FOR TOTAL THICKNESS OF PILE-UP NOT EXCEEDING 4*T*. MAXIMUM RATIO BETWEEN THICKNESSES, 3 TO 1.

(B) FOR RWMA CLASS 2 ELECTRODE MATERIAL (MINIMUM CONDUCTIVITY, 75% IACS; MAXIMUM HARDNESS, 75 HRB). WHEEL WIDTH IS OVERALL WIDTH OF A FIAT-FACE WHEEL OR OF A WHEEL WITH A 75 MM (3 IN.) RADIUS FACE; FACE WIDTH IS THAT OF THE CONTACTING SURFACE OF FLAT-FACE WHEELS.

(C) FOR LARGER ASSEMBLIES, MINIMUM CONTACTING OVERLAP INDICATED SHOULD BE INCREASED 30%.

Limitations of RSEW, apart from those it shares with spot and projection welding, are:

- THE WELD ORDINARILY MUST PROCEED IN A STRAIGHT OR UNIFORMLY CURVED LINE
- OBSTRUCTIONS ALONG WITH THE PATH OF THE ELECTRODE WHEEL MUST BE AVOIDED OR BE COMPENSATED FOR IN THE DESIGN OF THE WHEEL
- SHARP CORNER RADII OR ABRUPT CHANGES IN CONTOUR ALONG THE PATH OF THE ELECTRODE WHEEL MUST BE AVOIDED (TABLE 6)
- FATIGUE LIFE OF RESISTANCE SEAM WELDS IS USUALLY SHORTER THAN THAT OF WELDS MADE BY OTHER SEAM WELDING METHODS

The shorter fatigue life is due to two factors. First, the most common method of cooling the electrode is flood cooling, which quenches the weld. Second, RSEW electrodes move continuously, preventing the application of postweld heat treatment to temper the weld.

TABLE 6 RECOMMENDED MINIMUM RADII OF CURVATURE FOR SEAM WELDED ALUMINUM AND STEEL COMPONENTS

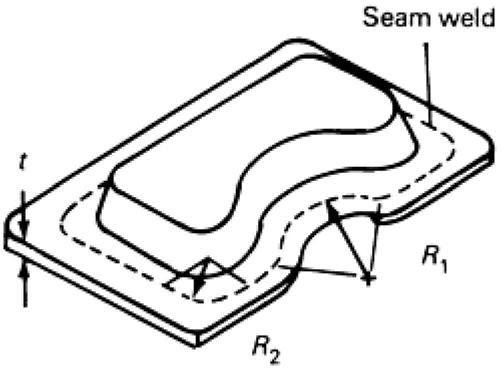

BASE MATERIAL	MINIMUM RADIUS			
	INSIDE RADIUS, R_1		OUTSIDE RADIUS, R_2	
	MM	**IN.**	**MM**	IN.
ALUMINUM	250	10	75	**3**
STEEL	**150**	**6**	**$50^{(A)}$** **$57^{(B)}$** **$64^{(C)}$** **$75^{(D)}$**	**$2^{(A)}$** **$2\frac{1}{4}^{(B)}$** **$2\frac{1}{2}^{(C)}$** **$3^{(D)}$**

Source: Ref 7

(A) 0.8 MM $\le T \le$ 1.3 MM (0.031 IN. $\le T \le$ 0.051 IN.).

(B) $T \le$ 1.6 MM ($T \le$ 0.063 IN.).

(C) $T \le$ 2.0 MM ($T \le$ 0.079 IN.).

(D) $T \le$ 3.0 MM ($T \le$ 0.12 IN.).

Metals Welded. Low-carbon, high-carbon, low-alloy, high-strength low-alloy (HSLA), stainless, and many coated steels can be resistance seam welded satisfactorily. Alloys with carbon levels above about 0.15% may tend to form areas of hard martensite upon cooling.

Reference cited in this section

7. J.G. BRALLA, ED., *HANDBOOK OF PRODUCT DESIGN FOR MANUFACTURING,* MCGRAW-HILL, 1986, P 7-39

Procedure Development and Practice Considerations for Resistance Welding

Projection Welding

Resistance projection welding (RPW) is a welding process in which current flow and heating are localized at a point or points predetermined by the design or shape of one or both of two parts to be welded (Fig. 12). The process is closely related to resistance spot welding (RSW), in which current flow and heating are localized by one or both electrode contact faces, which determine the location, size, and shape of the weld produced.

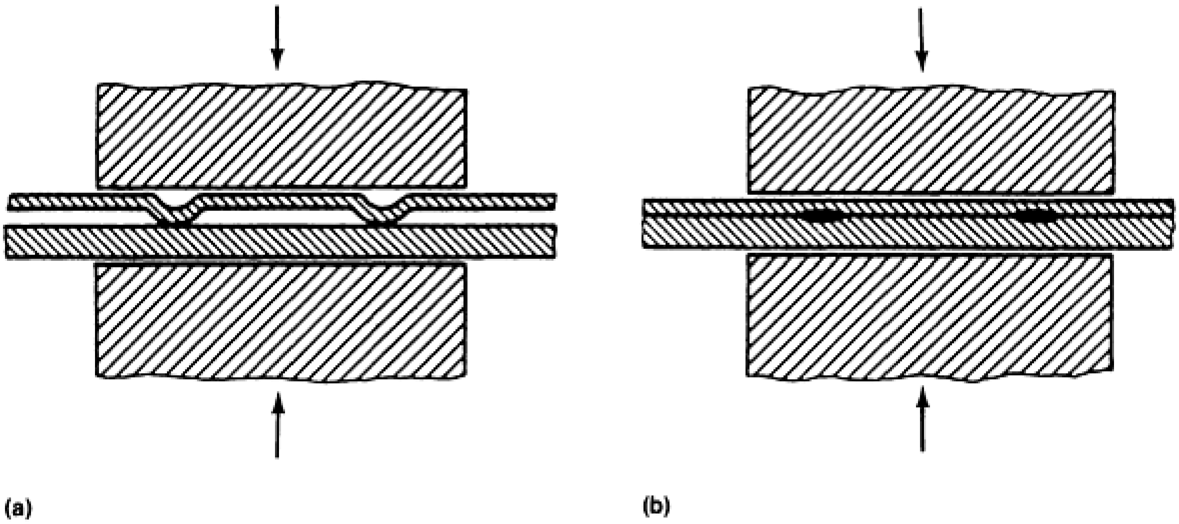

FIG. 12 RELATION OF WELD NUGGETS TO EMBOSSED REGIONS OF METAL SHEETS USED IN PROJECTION WELDING. (A) ELECTRICAL CIRCUIT LOOP IS COMPLETED WHEN CURRENT TRAVELS THROUGH EMBOSSED PROJECTION CONTACT POINTS. (B) ADDITION OF PRESSURE AFTER WELDING CURRENT IS TURNED OFF CAUSES PLASTIC DEFORMATION WHICH FLATTENS THE PROJECTIONS.

Projections may be of any practical shape that can properly concentrate the welding current (Fig. 13). In cross-wire projection welding, the curved surfaces of two intersecting wires perform the function of a projection. The shapes of parts may also take the place of conventional projections in projection welding of special types of joints.

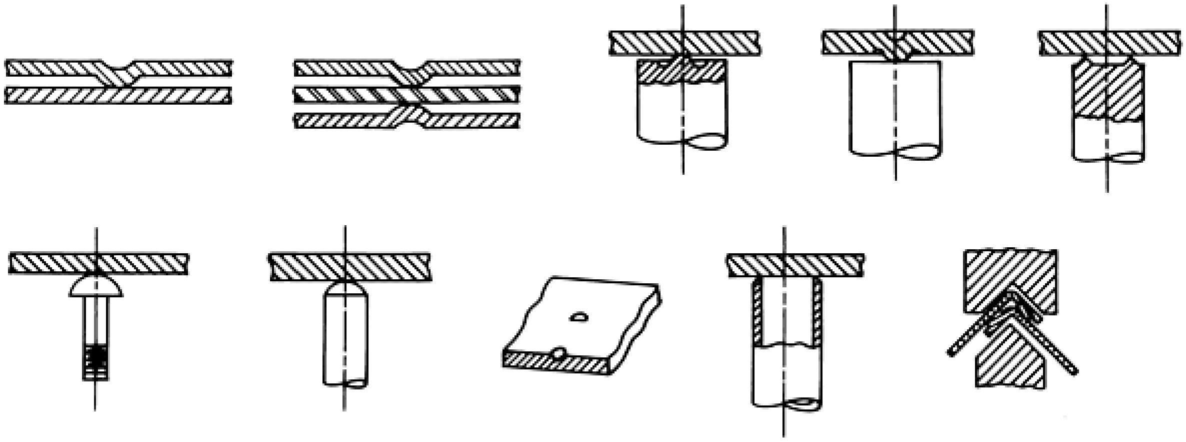

FIG. 13 SCHEMATICS SHOWING SELECTED SETUPS WHICH CAN BE USED IN PROJECTION WELDING OPERATIONS

Additional information is available in the article "Projection Welding" in this Volume.

Formation of the Weld. The formation of a projection weld nugget, which depends on the design of the projection, the selection of welding conditions, and the adequacy of the resistance welding equipment, is shown in Fig. 14. In this application, 2.33 mm (0.092 in.) thick low-carbon steel is projection welded using embossed spherical projections and a weld time of 20 cycles ($\frac{1}{3}$ s).

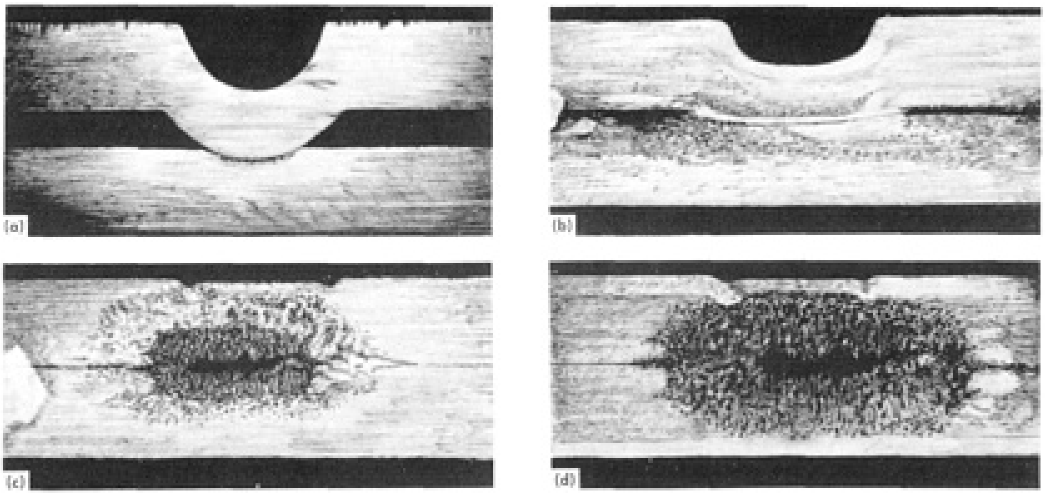

FIG. 14 FORMATION OF A WELD NUGGET DURING PROJECTION WELDING OF AN EMBOSSED SPHERICAL PROJECTION. SEE TEXT FOR EXPLANATION.

In the first stage of fabrication (Fig. 14a), the workpieces are brought together under pressure without application of welding current, and the projection may be slightly compressed and indented into the surface of the mating workpiece. Figures 14(b), 14(c), and 14(d) show the stages of formation of a typical weld at 20, 60 to 70, and 100% of the weld time (or heat time).

At about 20% of weld time (Fig. 14b), collapse of the projection is nearly complete, and a pressure weld is formed. A nugget of fused weld metal does not begin to form until about 50% of the weld time has elapsed. At about 60 to 70% of weld time (Fig. 14c), fusion has progressed a sufficient distance from the interface to produce a well-defined weld nugget that has about half its final thickness (penetration) and diameter. Sheet separation adjacent to the weld nugget has been reduced to zero, and the softened metal above the nugget has been flattened against the face of the upper electrode (compare with Fig. 14b). Nugget diameter, penetration, and shear strength continue to increase as weld time progresses. Figure 14(d) shows the fully developed weld nugget.

Applicability

Projection welding is used where there is a high thickness ratio (for example, nuts welded to sheet), and where the appearance of the sheets must be controlled (for example, appliance surfaces). The principal application of projection welding is the joining of stamped low-carbon, low-alloy, and high-strength low-alloy (HSLA) steel parts. During stamping (punching, drawing, or forming), one of the parts must have a projection which has been formed during the stamping operation to enable projection welding. Projection welding is also used for joining screw-machine parts to stamped parts; the projection is machined or cold formed on the end of the screw-machine part.

Projection welding is most successful in workpieces 0.56 to 3.18 mm (0.022 to 0.125 in.) thick. Stock 0.25 mm (0.010 in.) thick has been projection welded; projection design is critical, however, and machines with low-inertia heads and fast follow-up are needed. Sections less than 0.25 mm (0.010 in.) thick are more adaptable to spot welding.

Advantages. The principal advantages of projection welding are:

- THE NUMBER OF WELDS THAT CAN BE MADE SIMULTANEOUSLY WITH ONE OPERATION OF THE WELDING MACHINE IS LIMITED ONLY BY THE ABILITY OF THE CONTROLS TO REGULATE CURRENT AND FORCE.
- BECAUSE OF GREATER CURRENT CONCENTRATION AT THE WELD, AND THUS LESS CHANCE OF SHUNTING, NARROWER FLANGES CAN BE WELDED, AND WELDS CAN BE SPACED CLOSER TOGETHER BY PROJECTION WELDING THAN BY SPOT WELDING.
- ELECTRODES USED IN PROJECTION WELDING HAVE FACES LARGER THAN THE PROJECTIONS AND LARGER THAN THE FACES OF ELECTRODES USED FOR MAKING SPOT WELDS OF COMPARABLE NUGGET DIAMETER. CONSEQUENTLY, BECAUSE OF LOWER CURRENT DENSITY, ELECTRODES REQUIRE LESS MAINTENANCE THAN DO SPOT WELDING ELECTRODES.
- PROJECTION WELDS CAN BE MADE IN METAL THAT IS TOO THICK TO BE JOINED BY RSW.
- FLEXIBILITY IN SELECTION OF PROJECTION SIZE AND LOCATION ALLOWS WELDING OF WORKPIECES IN THICKNESS RATIOS OF 6 (OR MORE) TO 1. WORK-PIECES IN THICKNESS RATIOS GREATER THAN ABOUT 3 TO 1 SOMETIMES ARE DIFFICULT TO SPOT WELD.
- THE PROCESS CAN BE USED TO MAKE LEAKPROOF JOINTS (FOR EXAMPLE, RING PROJECTIONS).

Limitations of projection welding include:

- FORMING OF ONE OR MORE PROJECTIONS ON ONE OF THE WORKPIECES MAY REQUIRE EXTRA OPERATIONS UNLESS THE PARTS ARE PRESS FORMED TO DESIGN SHAPE.
- WHEN SEVERAL WELDS ARE MADE AT ONCE WITH THE SAME ELECTRODE, ALIGNMENT OF THE WORK AND DIMENSIONS (PARTICULARLY HEIGHT) OF THE PROJECTIONS MUST BE HELD TO CLOSE TOLERANCES TO OBTAIN CONSISTENT WELD QUALITY.
- WHEN MAKING SIMULTANEOUS PROJECTION WELDS, THE PROJECTION LAYOUT WILL BE DICTATED BY THE SHUNT CURRENT PATHS, WHICH MAY NOT COINCIDE WITH THE DESIRED LOCATION.
- BECAUSE TWO PARTS ARE MOVING RELATIVE TO EACH OTHER DURING THE WELDING PROCESS, PROPER GAGING IS DIFFICULT.

Cross-Wire Welding

Resistance welding of crossed wires is a type of projection welding process. In practice, it usually consists of welding a number of parallel wires at right angles to one or more other wires or rods. There are many specific ways to perform the welding operation, depending upon production requirements, but the final finished product is essentially the same regardless of the method used.

Typical crossed-wire products include items such as stove and refrigerator racks, grills of all kinds, lamp shade frames, poultry equipment, wire baskets, fencing, grating and concrete rein-forcing mesh.

Welding Machines

Press-type machines, with either single-phase or three-phase transformers, usually are used for projection welding. The welding head in these machines is guided by bearings or ways and moves in a straight line. Platens with T-slots or tapped holes are used for mounting the welding dies or electrodes. Rocker-arm machines generally are not used for projection welding because the electrode moves in an arc that can cause slippage between the components as the projection collapses.

At the start of the welding stroke (Fig. 15a), equal air pressure is applied to the top of the piston and to the diaphragm. By the time the piston has completed its travel, the diaphragm has been compressed by the retraction of the internal shaft. Retraction of the internal shaft simultaneously compresses the diaphragm and closes the welding-current switch; the control lever for the switch rides in the shaft collar.

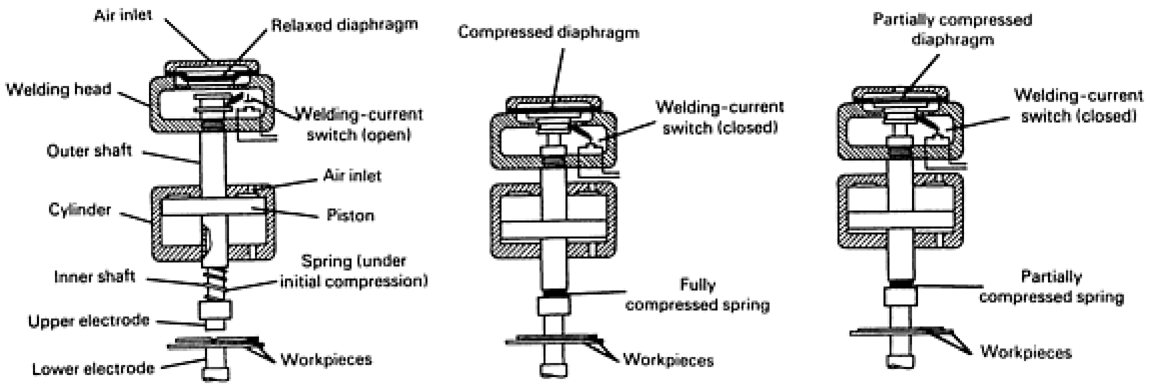
Air inlet
Relaxed diaphragm
Welding head
Welding-current switch (open)
Outer shaft
Air inlet
Cylinder
Piston
Inner shaft
Spring (under initial compression)
Upper electrode
Lower electrode
Workpieces
Compressed diaphragm
Welding-current switch (closed)
Fully compressed spring
Workpieces
Partially compressed diaphragm
Welding-current switch (closed)
Partially compressed spring
Workpieces

(a) Welding head in open position

(b) Welding head in position for squeezing and heating projection

(c) Welding head at instant of projection collapse and start of nugget formation

FIG. 15 SCHEMATIC SHOWING MOTION OF INTERNAL COMPONENTS IN A LOW-INERTIA WELDING HEAD USED FOR PROJECTION WELDING OPERATIONS

When the welding current is initiated (Fig. 15b), the only remaining mass to be moved is the internal shaft and its attached electrode. Air pressure acting on the diaphragm and the force of the compressed spring between the inner and outer shafts easily overcome the low inertia of the system and move the upper electrode downward as the projection collapses. Thus, the workpieces are kept in close contact under partial electrode force until the welding head follows and forges the weld region with the full electrode force (Fig. 15c). This setup, however, is no longer used because diaphragm cylinders are now classified as obsolete technology.

Process Variables

The major variables that affect projection welding are welding current, electrode force, and weld time.

Specifications for projection welding of stainless steels relative to low-carbon steel from 0.36 to 3.18 mm (0.014 to 0.125 in.) thick are given in Table 7. These data are intended to serve as starting points and should be adjusted to suit the specific application and the equipment used.

TABLE 7 RECOMMENDED ELECTRODE FORCE AND WELDING CURRENT SPECIFICATIONS FOR PROJECTION WELDING OF SELECTED STAINLESS STEELS RELATIVE TO 1010 LOW-CARBON STEEL

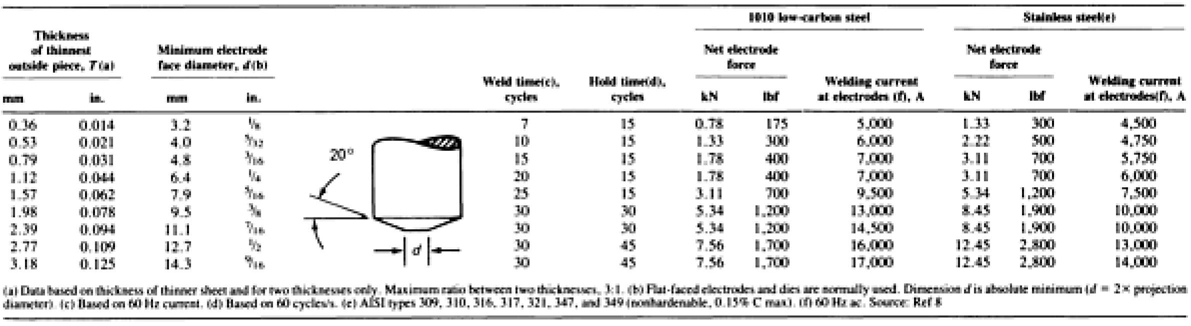

Thickness of thinnest outside piece, *T* (a)		Minimum electrode face diameter, *d* (b)		Weld time(c), cycles	Hold time(d), cycles	1010 low-carbon steel			Stainless steel(e)		
						Net electrode force		Welding current at electrodes (f), A	Net electrode force		Welding current at electrodes(f), A
mm	in.	mm	in.			kN	lbf		kN	lbf	
0.36	0.014	3.2	1/8	7	15	0.78	175	5,000	1.33	300	4,500
0.53	0.021	4.0	5/32	10	15	1.33	300	6,000	2.22	500	4,750
0.79	0.031	4.8	3/16	15	15	1.78	400	7,000	3.11	700	5,750
1.12	0.044	6.4	1/4	20	15	1.78	400	7,000	3.11	700	6,000
1.57	0.062	7.9	5/16	25	15	3.11	700	9,500	5.34	1,200	7,500
1.98	0.078	9.5	3/8	30	30	5.34	1,200	13,000	8.45	1,900	10,000
2.39	0.094	11.1	7/16	30	30	5.34	1,200	14,500	8.45	1,900	10,000
2.77	0.109	12.7	1/2	30	45	7.56	1,700	16,000	12.45	2,800	13,000
3.18	0.125	14.3	9/16	30	45	7.56	1,700	17,000	12.45	2,800	14,000

(a) Data based on thickness of thinner sheet and for two thicknesses only. Maximum ratio between two thicknesses, 3:1. (b) Flat-faced electrodes and dies are normally used. Dimension *d* is absolute minimum ($d = 2\times$ projection diameter). (c) Based on 60 Hz current. (d) Based on 60 cycles/s. (e) AISI types 309, 310, 316, 317, 321, 347, and 349 (nonhardenable, 0.15% C max). (f) 60 Hz ac. Source: Ref 8

Parameters for projection welding of low-carbon steel wire are given in Table 8.

TABLE 8 GUIDELINES FOR PROJECTION WELDING OF LOW-CARBON STEEL WIRE

WIRE DIAMETER		**WELD TIME$^{(B)}$, CYCLES**	**CONDITIONS FOR 15% SETDOWN$^{(A)}$**					**CONDITIONS FOR 30% SETDOWN$^{(A)}$**					CONDITIONS FOR 50% SETDOWN$^{(A)}$				
			ELECTRODE FORCE		**WELDING CURRENT, A**	**WELD STRENGTH**		**ELECTRODE FORCE**		**WELDING CURRENT, A**	**WELD STRENGTH**		**ELECTRODE FORCE**		**WELDING CURRENT, A**	WELD STRENGTH	
MM	**IN.**		**KN**	**LBF**		**KN**	**LBF**	**KN**	**LBF**		**KN**	**LBF**	**KN**	**LBF**		**KN**	LBF
COLD-DRAWN WIRE																	
1.6	$\frac{1}{16}$	5	0.45	100	600	2.0	450	0.67	150	800	2.2	500	0.89	200	1,000	2.4	**550**
3.2	$\frac{1}{8}$	10	0.56	125	1,800	4.3	975	1.2	260	2,650	5.0	1,125	1.6	350	3,400	5.6	**1,250**
4.8	$\frac{3}{16}$	17	1.6	360	3,300	8.9	2,000	2.7	600	5,000	10.7	2,400	3.3	750	6,000	11.1	**2,500**
6.4	$\frac{1}{4}$	23	2.6	580	4,500	16.4	3,700	3.8	850	6,700	18.7	4,200	5.5	1240	8,600	19.6	**4,400**
7.9	$\frac{5}{16}$	30	3.7	825	6,200	22.7	5,100	6.4	1450	9,300	27.1	6,100	8.9	2000	11,400	28.9	**6,500**
9.5	$\frac{3}{8}$	40	4.9	1100	7,400	29.8	6,700	9.2	2060	11,300	37.1	8,350	13.3	3000	14,400	39.1	**8,800**
11	$\frac{7}{16}$	50	6.2	1400	9,300	42.7	9,600	12.9	2900	13,800	50.3	11,300	19.8	4450	17,400	52.9	**11,900**
13	$\frac{1}{2}$	60	7.6	1700	10,300	54.3	12,200	15.1	3400	15,800	60.5	13,600	23.6	5300	21,000	64.9	**14,600**
HOT-DRAWN WIRE																	
1.6	$\frac{1}{16}$	5	0.45	100	600	1.6	350	0.67	150	800	1.8	400	0.89	200	1,000	2.0	**450**
3.2	$\frac{1}{8}$	10	0.56	125	1,850	3.3	750	1.2	260	2,770	3.8	850	1.6	350	3,500	4.0	**900**
4.8	$\frac{3}{16}$	17	1.6	360	3,500	6.7	1,500	2.7	600	5,100	7.6	1,700	3.3	750	6,300	8.0	**1,800**
6.4	$\frac{1}{4}$	23	2.6	580	4,900	12.5	2,800	3.8	850	7,10	13.3	3,000	5.5	1240	9,000	13.8	**3,100**
7.9	$\frac{5}{16}$	30	3.7	825	6,600	20.5	4,600	6.4	1450	9,600	22.2	5,000	8.9	2000	12,000	23.6	**5,300**
9.5	$\frac{3}{8}$	40	4.9	1100	7,700	27.6	6,200	9.2	2060	11,800	30.2	6,800	13.3	3000	14,900	32.0	**7,200**

11	$\frac{7}{16}$	50	6.2	1400	10,000	39.1	8,800	12.9	2900	14,000	42.7	9,600	19.8	4450	18,000	45.4	**10,200**
13	$\frac{1}{2}$	**60**	**7.6**	**1700**	**11,000**	**51.1**	**11,500**	**15.1**	**3400**	**16,500**	**55.2**	**12,400**	**23.6**	**5300**	**22,000**	**57.8**	**13,000**

Source: Ref 9

(A) SETDOWN, % = (DECREASE IN JOINT HEIGHT ÷ DIAMETER OF SMALLER WIRE) × 100.

(B) FOR 15.30, AND 50% SETDOWN.

Welding current required for projection welding, although slightly less per weld than that needed for spot welding, must be high enough to cause fusion before the projection is completely flattened. The recommended current is the highest current that, when used with the current electrode pressure, does not cause excessive expulsion of metal.

Electrode force used in projection welding depends on the work metal, the size and design of the projection, the number of projections in the joint, and the welding machine. Excessive force causes the projections to collapse before the weld area has reached the proper temperature, resulting in the formation of ring welds, in which fusion occurs around the periphery of the projection but is incomplete at the center. For best weld appearance, the electrode force should be such that the projection is flattened completely after the metal has reached welding temperature.

Weld time for a given type and thickness of work metal depends on welding current and rigidity of the projection. Weld time is less important than electrode force in projection welding of low-carbon, low-alloy, and HSLA steel, provided that the time is sufficient to produce a nugget of adequate size at the chosen welding current. A short weld time creates higher production efficiency and less discoloration and distortion of the workpiece. After the proper electrode force and welding current are determined, the weld time is adjusted to make the desired weld.

Electrodes

An electrode designed for RSW can be used for projection welding if the electrode face is large enough to cover the projection being welded, or the pattern of projections being welded simultaneously. To minimize marking and indentation of workpieces, the recommended electrode-face diameter for making a single projection weld usually is two or more times the diameter of the projection. In multiple-projection welding, the electrode face should be large enough to extend beyond the boundaries of the pattern of projections by approximately the diameter of one projection. RWMA class 2 electrode materials generally are preferred as general-purpose materials because they provide the best compromise among electrical and conductivity, strength, hardness, and temperature resistance. Typically, class 10, 11, or 12 materials are used on the bottom side of the weld to resist damage from the high localized force.

Reference cited in this section

9. *RESISTANCE WELDING MANUAL,* 3RD ED., VOL 1, RESISTANCE WELDER MANUFACTURERS ASSOCIATION, 1956, P 88

Procedure Development and Practice Considerations for Resistance Welding

Flash Welding

Flash welding (FW) commonly is used to join sections of metals and alloys in production quantities. It is a resistance/forge welding process in which the items to be welded are securely clamped to electric current-carrying dies, heated by the electric current, and upset (Fig. 16). Clamping ensures good electrical contact between the current-carrying dies and the workpiece and prevents the parts from slipping during the upsetting action. Flash welding equipment must be durable to withstand high clamping force and upset pressures without deflecting. If deflection occurs, misalignment of workpieces may occur during welding. Flash welding is rapid and economical, and when properly executed, welds of uniform high quality are produced. A typical flash welding machine comprises a horizontal press-transformer combination with conducting work-clamping dies mounted on the press platens.

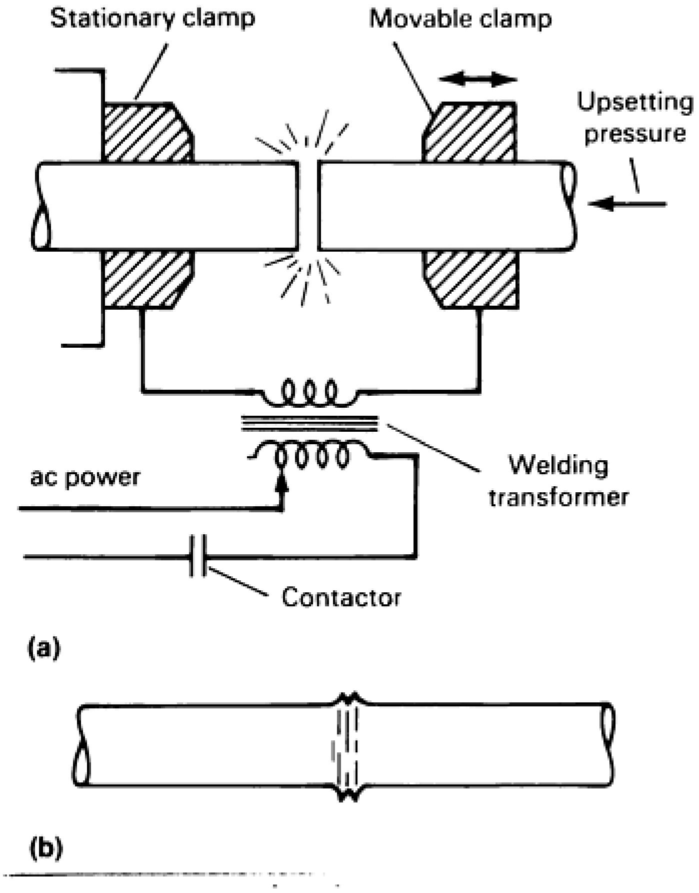

FIG. 16 WELD PRODUCED WHEN USING THE FLASH WELDING PROCESS. (A) WORKPIECES SECURELY CLAMPED IN CURRENT-CARRYING DIES BEFORE UPSETTING OPERATION IS INITIATED. (B) FINISHED WELD PRODUCED AFTER UPSETTING OPERATION

Additional information is available in the article "Flash Welding" in this Volume.

Applications

Flash welding can be used for joining many ferrous and nonferrous alloys and combinations of dissimilar metals. In addition to low-carbon steels, metals that are flash welded on a production basis include low-alloy steels, tool steels, stainless steels, aluminum alloys, magnesium alloys, nickel alloys, and copper alloys. Table 9 gives typical upsetting pressures for numerous classes of alloys.

TABLE 9 RECOMMENDED UPSETTING PRESSURES FOR FLASH WELDING OF SELECTED ALLOYS

STRENGTH CLASSIFICATION	**UPSET PRESSURE**		ALLOYS	
	MPA	**KSI**	**SAE DESIGNATION**	OTHERS
LOW FORGING	69	10	1020, 1112, 1315	**HIGH-STRENGTH LOW-ALLOY (HSLA) STEELS**
MEDIUM FORGING	103	15	1045, 1065, 1335, 3135, 4130, 4140, 8620, 8630	
HIGH FORGING	172	25	4340, 4640, 300M	**TOOL STEELS, TITANIUM, ALUMINUM**

EXTRA-HIGH FORGING	**241**	**35**	**A-286, 19-9 DL**	**NICKEL-BASE ALLOYS, (A)**

Source: Ref 10

(A) MATERIALS EXHIBITING EXTRA-HIGH COMPRESSIVE STRENGTH AT ELEVATED TEMPERATURES.

Welding Parameters

Fundamentally, flash welding involves heating the ends of the pieces to be welded and subsequently forging them together. During the heating phase, a thermal distribution pattern is established along the axial length of the pieces being joined, which is characterized by a steep temperature gradient.

The major difference between the temperature pattern developed in flash welding and that developed in resistance welding is that flash welding produces a much steeper thermal gradient. This steep thermal gradient, combined with the resulting characteristic upset pattern, enables flash welding to accommodate a much greater variety of materials and shapes than can be welded by resistance welding.

Once the proper temperature-distribution pattern has been established, the abutting surfaces are rapidly forced together. The parts must be securely held together during the forging process to prevent slipping. Three distinct peaks are characteristic of flash welds (Fig. 17). The two peaks on either side of the weld line represent the material displaced by the upsetting action; the center peak is the molten metal extruded out of the weld, including oxides or contaminants formed during heating.

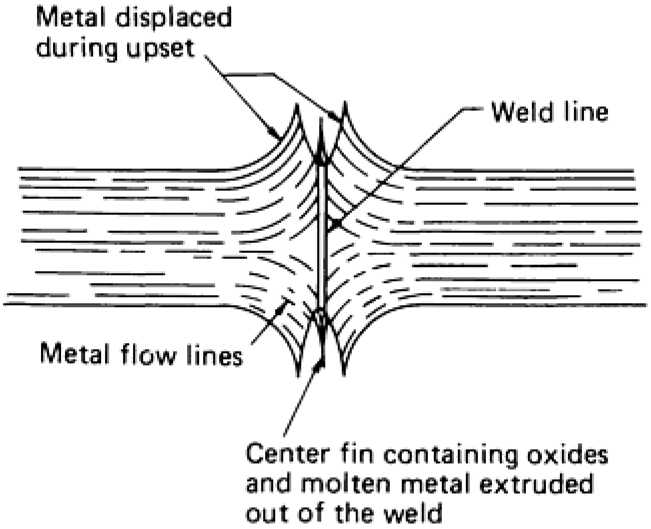
Metal displaced during upset
Weld line
Metal flow lines
Center fin containing oxides and molten metal extruded out of the weld

FIG. 17 SCHEMATIC SHOWING CROSS-SECTIONAL VIEW OF TYPICAL PEAKS AND FLOW LINES GENERATED IN A FLASH WELD

Heat (Energy) Sources

One of the major considerations for flash welding is the electrical power service. Flashing is a term used to describe the major heating process during flash welding. When the ends of the workpiece are brought together under light pressure, an electrical short circuit is established through the material. Because the abutting surfaces are not perfectly matched, the short-circuit current flows across the joint only at a few small contact areas. The large amount of current flowing through a relatively small area causes very rapid heating to the melting point. Heating is so rapid and intense that molten metal is expelled explosively from the joint area. Following this expulsion, a brief period of arcing occurs. Arcing is not sustained due to the low voltages employed. Studies have shown that stable flashing can be maintained with as low as $2\frac{1}{2}$ to 3 V, but typically the flashing voltage for alternating-current machines is from 5 to 10 V.

Following the expulsion of molten metal and subsequent arcing, small craters are formed on the ends of the abutting surfaces. The pieces are steadily advanced toward one another, and other short circuits are formed and additional molten metal is expelled. This process continues as random melting, arcing, and expulsion occur over the entire cross-sectional surface. During flashing, many active areas are in various stages of this sequence (Fig. 18). Flashing surfaces act as heat sources, and the steep thermal profile is established primarily from these heat sources. Temperatures of the flashes are at or above the melting point of the material and are progressively lower as distance progresses from the flashing surface toward the clamp.

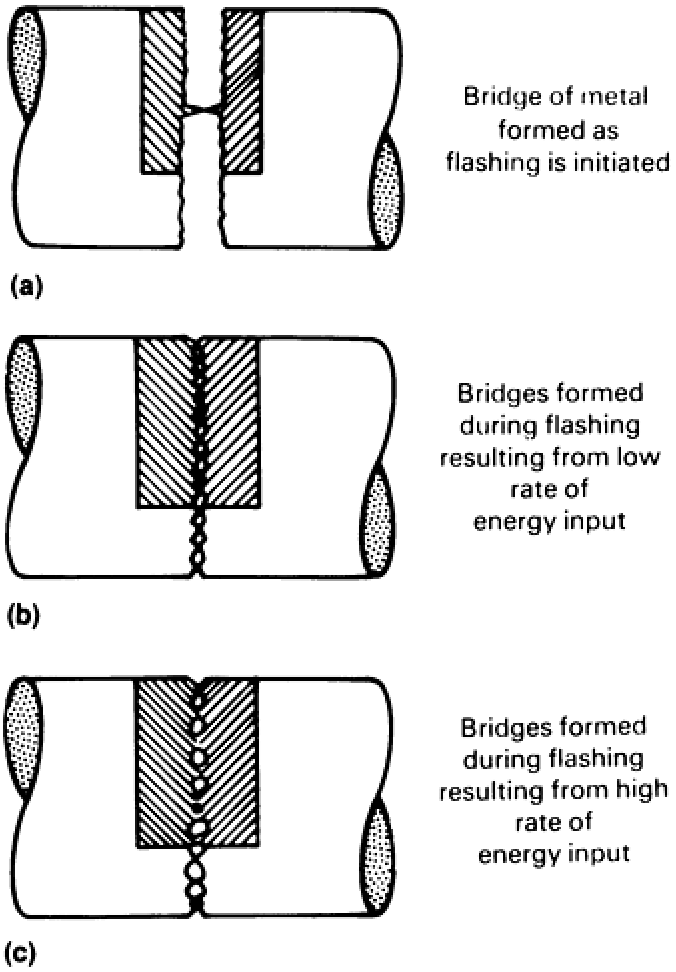
Bridge of metal formed as flashing is initiated
(a)
Bridges formed during flashing resulting from low rate of energy input
(b)
Bridges formed during flashing resulting from high rate of energy input
(c)

FIG. 18 GROWTH OF CONTACT POINTS PRODUCED BY ENERGY INPUT DURING FLASH WELDING

Force

Parts to be welded must be clamped or fixtured securely to provide good electrical contact with current-carrying dies and to transmit the upset forces. Also, a reliable source of force for the upsetting action is required. Systems for generating these forces vary greatly and are determined by the cross sections to be welded. The simplest welding machines derive their power for clamping and upsetting from the operator (that is, the parts are clamped to the dies via screw, lever, cam, or toggle force-multiplying linkages). Upset forces are generated in a similar manner. In larger machines, these forces are generated by pneumatic systems, oil hydraulic systems, combination air/oil hydraulic systems, or motor-driven cam systems.

Machines

Machines used for flash welding generally consist of a mainframe; a low-impedance welding transformer; a stationary platen; a movable platen on which clamping dies, electrodes, and other tools needed to position and hold the work-pieces are mounted; flashing and upsetting mechanisms; and the necessary electrical, air, or hydraulic controls (Fig. 19).

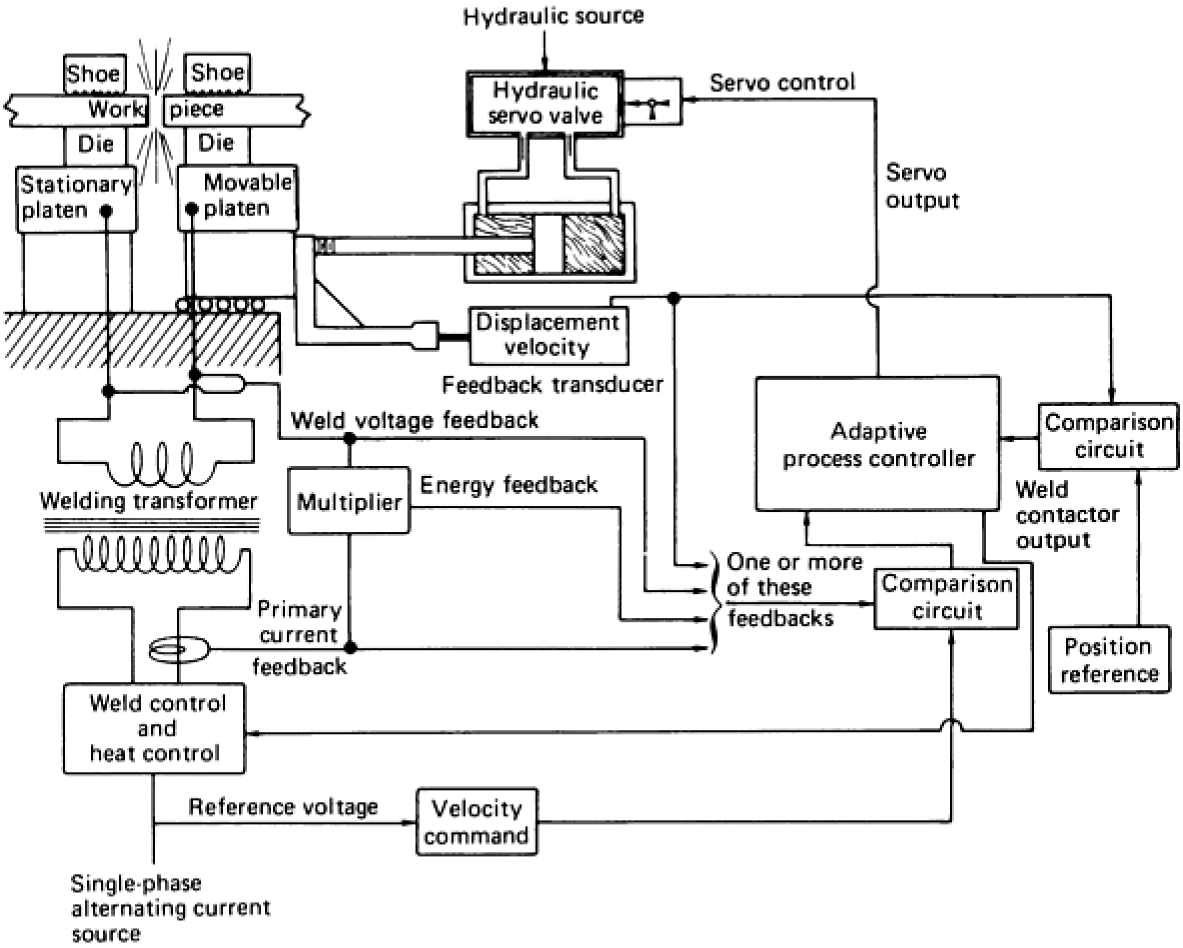
Hydraulic source
Hydraulic servo valve
Servo control
Servo output
Shoe
Shoe
Work
piece
Die
Die
Stationary platen
Movable platen
Displacement velocity
Feedback transducer
Weld voltage feedback
Multiplier
Energy feedback
Welding transformer
Primary current feedback
Adaptive process controller
Comparison circuit
Weld contactor output
One or more of these feedbacks
Comparison circuit
Position reference
Weld control and heat control
Reference voltage
Velocity command
Single-phase alternating current source

FIG. 19 FLOW DIAGRAM SHOWING ADAPTIVE CONTROLS USED TO CONTROL AND MONITOR FLASH WELDING PROCESS

Flash welding machines may be manual, semiautomatic, or fully automatic. With manual operation, the operator controls the speed of the platen from the time flashing is initiated until the upset is completed. In semiautomatic operation, the operator manually initiates flashing and then actuates an automatic cycle that completes the weld. In fully automatic operation, after the operator initiates the welding sequence, a fully automatic cycle can be used through the use of position-indicating devices and timers. Values for the various welding parameters are preselected by the operator. Automatic feedback control is used in some applications. These fully adaptive circuits vary from current to voltage feedback circuits, and data are obtained to control the welding sequences. Figure 19 shows various adaptive controls used by different flash welding unit manufacturers.

Transformers are used to supply electrical energy for the welding operation. The transformer tap switch frequently is a rotary, eight-step, knife-type, fully enclosed locking switch, which provides convenient adjustment of the welding voltage to suit work requirements. Voltages from 4 to 16 V are common.

Clamping Dies and Fixtures

Workpieces must be accurately clamped to maintain alignment, to allow the secondary current to pass into the workpieces, and to apply the upsetting force properly. Generally, the parts of the clamping mechanism that actually grip the workpiece are the electrodes, often called clamping dies. The surface of the material in contact with the dies must be cleaned of scale, rust, paint, or contaminants to ensure proper current flow.

The materials commonly used for clamping dies are RWMA class 3 and hardened tool steels such as H11, L6, and O1. Bronze and other copper-base electrode materials can be used in some applications. The die half (upper or lower) that conducts the current usually is made of a copper-base material. The other half is made of the same copper-base material or of hardened steel.

Cooling of Electrodes. The need for water cooling of flash welding electrodes depends on the size of the electrode, die, or fixture, the magnitude of the flashing current, and the production rate. When the mass of the electrodes is large compared with that of the workpiece, the heat sink generally is large enough to dissipate the heat generated by the resistance to current flow and the heat absorbed from the workpiece. With high flashing currents and high production rates, electrodes generally must be cooled.

Shape and Size of Clamping Dies. Generally, the shape of the clamping-die surface is such that the die encompasses almost the entire workpiece surface. The required area of clamping-die contact depends on the current needed for heating the workpiece and the pressure needed for holding it. Semicircular dies are used where the line contact provided by V-shape dies gives insufficient surface for the current to flow without burning the workpiece, or for holding the workpiece without marking it.

Table 10 gives the minimum lengths of clamping dies for welding rounds, tubing, and other sections of various diameters or minimum dimensions; these data are for welding steels of low or medium forging strength.

TABLE 10 MINIMUM LENGTHS OF CLAMPING DIES, WITH AND WITHOUT BACKUP, FOR FLASH WELDING SELECTED DIAMETERS OF WORKPIECES MADE FROM STEELS OF LOW OR MEDIUM FORGING STRENGTH

WORKPIECE DIAMETER(A)		**MINIMUM LENGTH OF CLAMPING DIE**				**WORKPIECE DIAMETER(A)**		**MINIMUM LENGTH OF CLAMPING DIE WITH BACKUP**		**WORKPIECE DIAMETER(A)**		MINIMUM LENGTH OF CLAMPING DIE WITH BACKUP	
		WITH BACKUP		**WITHOUT BACKUP(B)**									
mm	**in.**	**mm**	**in.**	**mm**	**in.**	**mm**	**in.**	**mm**	**in.**	**mm**	**in.**	**mm**	in.
6.35, 7.92	0.250, 0.312	9.53	0.375	25.4	1.00	50.8	2.00	31.8	1.25	152	6.00	82.3	**3.25**
9.53	0.375	9.53	0.375	38.1	1.50	63.5	2.50	44.5	1.75	165	6.50	88.9	**3.50**
12.70	0.500	9.53	0.375	44.5	1.75	76.2	3.00	50.8	2.00	178	7.00	95.3	**3.75**
19.05	0.750	12.70	0.500	50.8	2.00	88.9	3.50	57.2	2.25	191	7.50	102	**4.00**
25.40	1.000	19.05	0.750	63.5	2.50	102	4.00	63.5	2.50	203	8.00	108	**4.25**
38.10	1.50	25.40	1.000	76.2	3.00	114	4.50	69.9	2.75	216	8.50	114	**4.50**
						127	5.00	69.9	2.75	229	9.00	121	**4.75**
						140	**5.50**	**76.2**	**3.00**	**241**	**9.50**	**127**	**5.00**

(A) DIAMETER OF ROUNDS OR TUBING, OR MINIMUM DIMENSION OF OTHER SECTIONS.

(B) BACKUP IS RECOMMENDED FOR ALL WORKPIECE DIAMETERS OR MINIMUM DIMENSIONS OVER 38.1 MM (1.50 IN.).

Preweld Processing

Parts that are flash welded come in a variety of forms, including forgings; rolled or extruded bar; sheet, strip, or plate, ring preforms; and castings. Each of the above parts or assemblies requires preweld preparation.

Cleaning. As a minimum precaution, the ends of the workpiece that are clamped on the current-conducting die must be free from dirt, scale, surface oxidation, and grease. In addition, the ends of the workpiece that extend into the weld zone must be free from any contamination that could react with the base metal at the high temperatures developed during welding. Cleaning is needed because of the high current densities developed in the workpiece at the current-conducting dies. If insufficient electrical contact is made, weld quality is poor and localized hot spots can develop between the current-conducting dies and the workpiece. These localized hot spots are called "die burns." Many of the alloys welded have tightly adherent, highly resistive oxides that must be removed prior to welding.

In addition, poor fit-up, loose scale, and grease may cause the parts to slip during upsetting. Common cleaning techniques include (1) abrading the surfaces by grinding, grit blasting, or wire brushing; (2) pickling or chemical descaling; and (3) vapor degreasing.

End Preparation. Die burns, upset slippage, and inferior welds may be caused by poor fit-up between the current-conducting dies and the workpiece. In some cases, the ends of the workpiece must be machined or ground to fit the dies, particularly in welding of rough forgings or castings. Also, a chamfer may be required on the ends of the workpiece to initiate flashing action.

Alignment of Workpiece. After the workpieces are clamped in the welding unit, the alignment of the workpieces with respect to each other along the axis of upset must be maintained. After the flashing process and during upsetting, any misalignment may cause the parts to overlap one another or cause a skewed weld. If this occurs, insufficient metal will be extruded out of the weld zone during upsetting and a poor-quality weld will be formed (Fig. 20).

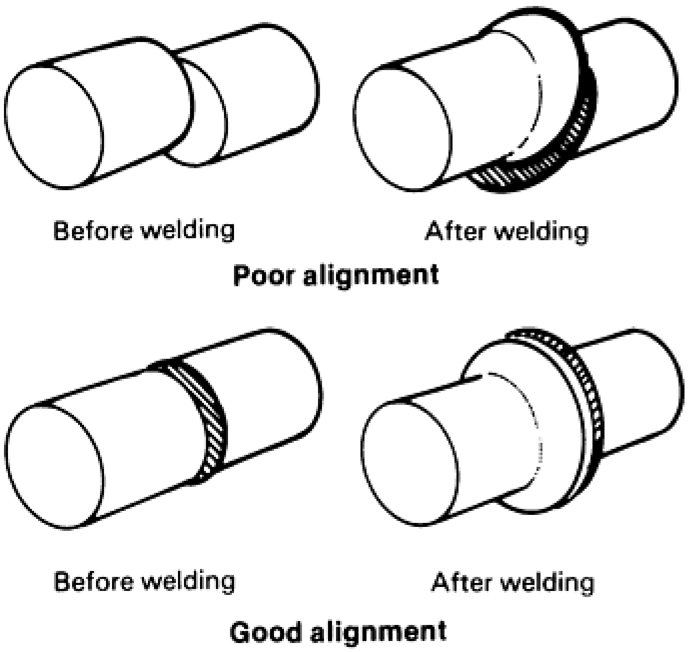

FIG. 20 EFFECT OF WORKPIECE ALIGNMENT ON JOINT QUALITY AND WELD UPSET

Dimensional Tolerances. Some combinations of joint length and stock thickness are difficult or even impossible to weld. Table 11 gives relationships between joint length (sheet width) and sheet thickness that have proven successful in a large number of production jobs. Relationships of tube diameter to wall thickness that have been successfully welded are listed in Table 12.

TABLE 11 LENGTHS OF JOINTS COMMONLY FLASH WELDED IN SELECTED THICKNESSES OF FLAT SHEET

SHEET THICKNESS		JOINT LENGTH	
mm	**in.**	**mm**	in.
0.25	0.010	25	**1**
0.51	0.020	125	**5**
0.76	0.030	255	**10**
1.02	0.040	380	**15**
1.27(A)	0.050(A)	510	**20**
1.52(A)	0.060(A)	635	**25**
2.03(A)	0.080(A)	890	**35**
2.54(A)	0.100(A)	1145	**45**
3.18(A)	0.125(A)	1458	**57**
4.75(A)	**0.187(A)**	**2235**	**88**

(A) LENGTH OF JOINTS IN STOCK THICKNESSES OF 1.27 MM (0.050 IN.) AND GREATER CAN BE UP TO 2540 MM (100 IN.) DEPENDING ON MATERIAL EXTENSION, PLATEN TRAVEL, DIE ALIGNMENT, AND CLAMPING.

TABLE 12 MAXIMUM FLASH-WELDABLE DIAMETERS OF TUBING OF SELECTED WALL THICKNESSES

WALL THICKNESS		MAXIMUM DIAMETER	
mm	**in.**	**mm**	in.
0.51	0.020	13	$\frac{1}{2}$
0.76	0.030	19	$\frac{1}{4}$
1.27	0.050	32	**1** $\frac{1}{4}$
1.57	0.062	38	**1** $\frac{1}{2}$
2.03	0.080	50	**2**
2.54	0.100	75	**3**
3.18	0.125	100	**4**
4.75	0.187	150	**6**
6.35	0.250	230	**9**
9.52	0.375	380	**15**
12.7	**0.500**	**510**	**20**

Burnoff. After the parts are clamped in the welding machine and the welding power is activated, the abutting surfaces are brought together for a brief period of violent flashing. This phase of the welding sequence is called "burnoff." It serves the function of squaring off the abutting surfaces and compensates for inconsistencies in end preparation.

Preheating is a resistive heating phase of the welding process in which the heat is generated by the electrical resistance of the workpieces. It is accomplished by bringing the workpieces together under light pressure, creating a short circuit. The pressure must be great enough to prevent flashing, but not so great that workpieces are prematurely welded.

In practice, preheating is performed in a cyclic manner, workpieces are brought together briefly, then separated for a brief period to allow the heat generated to diffuse into them. This sequence is repeated several times. Fully automatic machines perform this function, thus eliminating operator variables.

Flashing. The primary purpose of flashing is to generate enough heat to produce a plastic zone that permits adequate upsetting. The rate of energy input must be in proper proportion to the travel of the platen or movable die, so that constant flashing is maintained until the appropriate amount of metal is flashed off and the required plastic zone is obtained.

Upsetting (Forging). Bonding takes place during the upsetting action, and some metal must be extruded from the weld zone to remove slag and other inclusions not expelled during flashing. The extruded metal must extend beyond the cross-sectional boundaries of the workpiece to ensure that maximum amounts of slag and inclusions are removed when the weld upset is removed during subsequent trimming. When the weld upset is removed, no evidence of the weld should remain. Porosity near the outer surface on an etched section of the weld, or a crevice around the workpiece after the weld upset has been removed, indicates incomplete bonding because of either insufficient upsetting force or insufficient plasticity during upsetting.

A large upset produced by an excessive upsetting force can be detrimental to weld quality, can waste metal, and indicates that most of the plastic metal has been extruded, requiring bonding to take place in metal where plasticity may not have been sufficient to ensure a good weld.

Reference cited in this section

10. R.L. O'BRIEN, ED., *WELDING HANDBOOK*, 8TH ED., VOL 2, AWS, 1991, P 593

Procedure Development and Practice Considerations for Resistance Welding

Upset Welding (Ref 11)

Upset welding (UW) is a resistance welding process that produces coalescence over the entire area of faying surfaces or progressively along a butt joint by the heat obtained from the resistance to the flow of welding current through the area where those surfaces are in contact. Pressure is applied before heating is started and is maintained throughout the heating period (Fig. 21). The equipment used for upset welding is very similar to that used for flash welding. It can be used only if the parts to be welded are equal in cross-sectional area. Upset welding differs from flash welding in that the parts are clamped in the welding machine and force is applied to bring the parts tightly together. High-amperage current is then passed through the joint, which heats the abutting surfaces. When they have been heated to a suitable forging temperature, an upsetting force is applied and the current is stopped. The high temperature of the work at the abutting surfaces, plus the high pressure, causes coalescence to take place. After cooling, the force is released and the weld is completed. There is no arc or flash in upset welding. The area at the joint is usually enlarged over its original dimension.

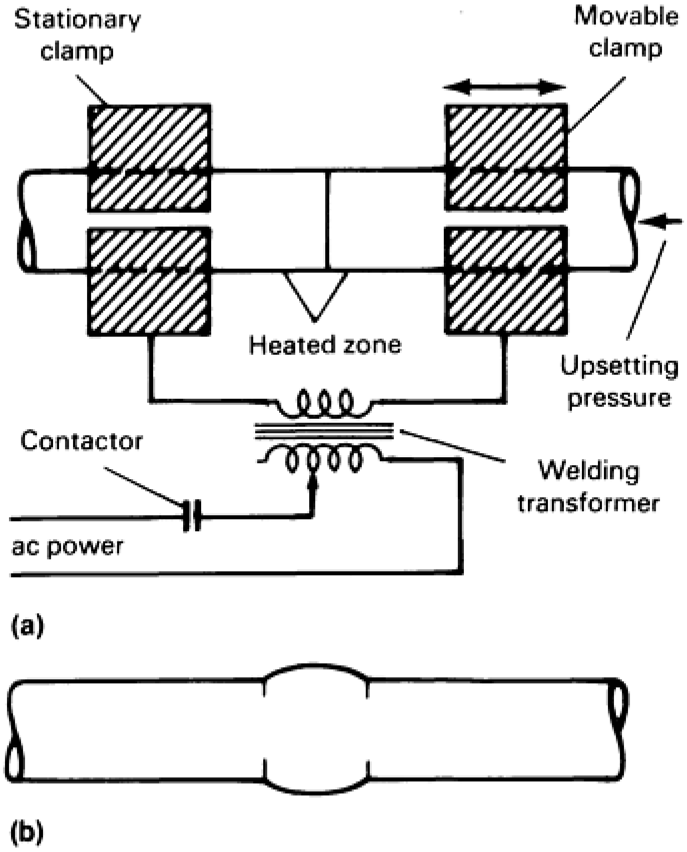

FIG. 21 WELD OBTAINED WHEN USING THE UPSET WELDING PROCESS. (A) WORKPIECES SECURELY CLAMPED IN CURRENT-CARRYING DIES WHILE PRESSURE AND CURRENT ARE APPLIED. (B) FINISHED WELD PRODUCED AFTER WELDING OPERATION IS COMPLETED

This process is used for welding small wires, tubing, piping, rings, and strips where the cross-sectional areas of both pieces are identical. If intimate contact is not obtained because of improper joint preparation, the weld will be defective.

Additional information is available in the article "Upset Welding" in this Volume.

Reference cited in this section

11. H.B. CARY, *MODERN WELDING TECHNOLOGY,* 2ND ED., PRENTICE-HALL, 1989, P 251-252

Procedure Development and Practice Considerations for Resistance Welding

High-Frequency Resistance Welding

High-frequency resistance welding (HFRW) is a resistance welding process that produces coalescence of metals by the heat generated from the resistance of the workpieces to a high-frequency current (10 to 500 kHz) and the rapid application of an upsetting force once heating is completed. The current is applied to the base metal by using electrodes that are held against the surface of the workpiece. The principal applications of HFRW are for making such structural shapes as pipe and tubing from continuous strips of metal. Two process variations of HFRW are continuous seam welding (Fig. 22a, b, c, d, e, and f) and finite-length butt welding (Fig. 22g), which is a variation of induction upset welding (UW-I).

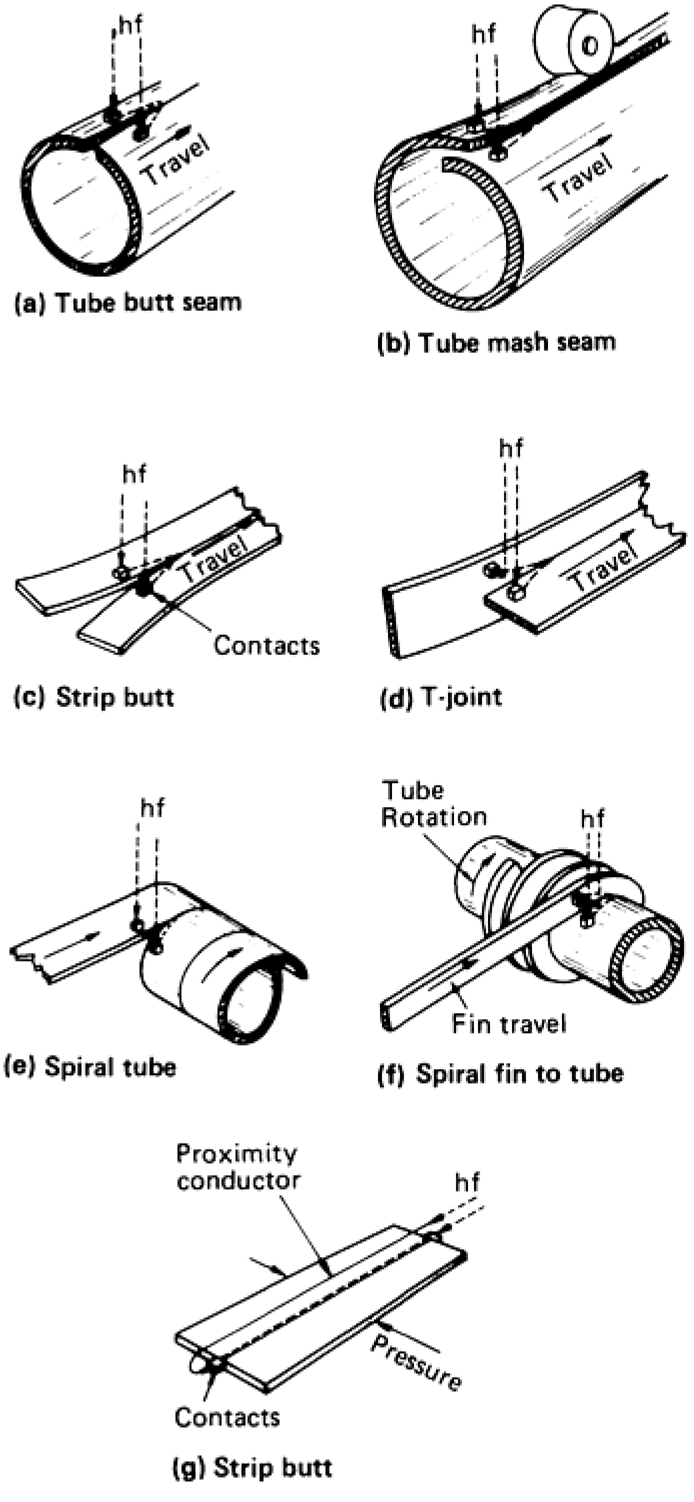
(a) Tube butt seam

(b) Tube mash seam

(c) Strip butt

(d) T-joint

(e) Spiral tube

(f) Spiral fin to tube

(g) Strip butt

hf
Travel
hf
Travel
hf
Travel
Contacts
hf
Travel
hf
Tube
Rotation
hf
Fin travel
Proximity
conductor
hf
Pressure
Contacts

FIG. 22 TYPICAL HIGH-FREQUENCY WELDING APPLICATIONS USED TO JOIN PLATES AND TO FORM TUBES

Detailed information is available in the article "High-Frequency Welding" in this Volume.

Continuous seam welding can be described by its most common application, high-frequency upset welding (UW-HF) of a tube (Fig. 23). A pair of electrodes is placed on either side of an open V formed by the converging edges of the strip being made into tubing. The angle of this V typically varies from 4 to 7°. The edges of the strip converge between the weld pinch rolls, where a certain degree of forging upset is accomplished.

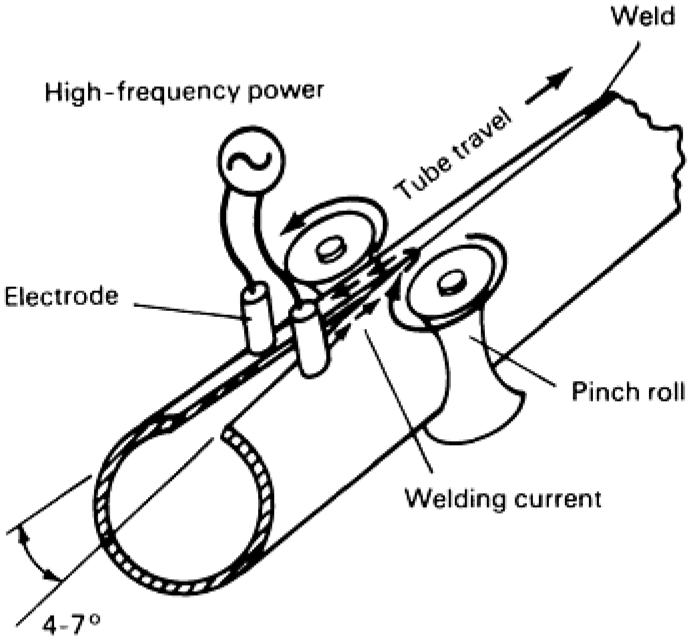

FIG. 23 HIGH-FREQUENCY LONGITUDINAL BUTT SEAM WELDING OF A TUBE

As the material moves through the rolls from the contacts to the point of welding, high-frequency current introduced at the electrodes flows on the opposing faying surfaces of the strip, down one side of the V to the point of welding, and back to the other electrode. Because this process is performed at frequencies of about 400 kHz, the depth of heating is quite shallow and heating occurs rapidly. Power is adjusted to bring the edges to a welding temperature at the time they join, and contaminants on the surfaces, as well as most of the molten material, are squeezed out of the seam toward the inside and outside of the tube when the weld is completed. This upset material can be trimmed off flush to the surfaces of the tube if desired.

The other applications of continuous high-frequency resistance seam welding (Fig. 23) vary from tube welding only in the shape of the workpiece. Contacts introduce current at the open part of a V that is being closed by rolls at the point at which the material has been heated to a high enough temperature to weld.

Finite-Length Welding. The process shown in Fig. 22(g), in which a butt joint of finite length is welded between two parts, is used for applications in coil end joining as well as in the manufacture of strip and sheet blanks for subsequent press forming. In this process, heating is accomplished along a seam that has already been closed under light pressure. The path of the heating current is controlled by the use of a proximity conductor positioned over the seam. The high-frequency heating causes the edges to be brought to welding temperature, after which pressure is increased to weld the joint. The appropriate frequency is selected to achieve heating through the material thickness. For example, 10 kHz can be used up to 6.4 mm (0.25 in.) thickness. Joints made this way can be produced very quickly, typically in 1 s. Compared to flash welding, the amount of upset is small, no shower of sparks and debris is emitted from the weld, and very little metal is consumed in making the joint.

Welding Parameters. Table 13 gives typical HFRW weld speeds as a function of wall thickness for tube and pipe products.

TABLE 13 TYPICAL WELD RATES FOR HIGH-FREQUENCY RESISTANCE WELDING OF TUBE AND PIPE

WALL THICKNESS	SPEED	
	200 KW, 76 MM (3.0 IN.)	600 KW, 305 MM (12 IN.)

		DIAMETER		DIAMETER	
MM	**IN.**	**M/S**	**FT/MIN**	**M/S**	FT/MIN
1.52	0.06	2.91	571	. . .	**. . .**
2.03	0.08	2.17	428	. . .	**. . .**
3.05	0.12	1.45	285	. . .	**. . .**
4.06	0.16	1.08	213	. . .	**. . .**
6.10	0.24	0.67	132	2.00	**393**
8.13	0.32	0.41	81	1.38	**272**
9.91	0.39	. . .	. . .	0.97	**191**
11.94	0.47	. . .	. . .	0.70	**137**
13.97	0.55	. . .	. . .	0.50	**99**
16.00	**0.63**	**. . .**	**. . .**	**0.35**	**69**

Capacitor Discharge Stud Welding

The capacitor discharge process uses direct current produced by the rapid discharge of a capacitor bank to create an electric arc and the heat needed for melting both stud and workpiece. The weld time is very short, ranging from 1 to 6 ms. Because the weld time is so short, no ferrule or fluxing is required. The equipment consists of a stud gun, capacitor discharge power/control unit, and suitable weld cable. Figure 24 illustrates a portable capacitor discharge stud welding system.

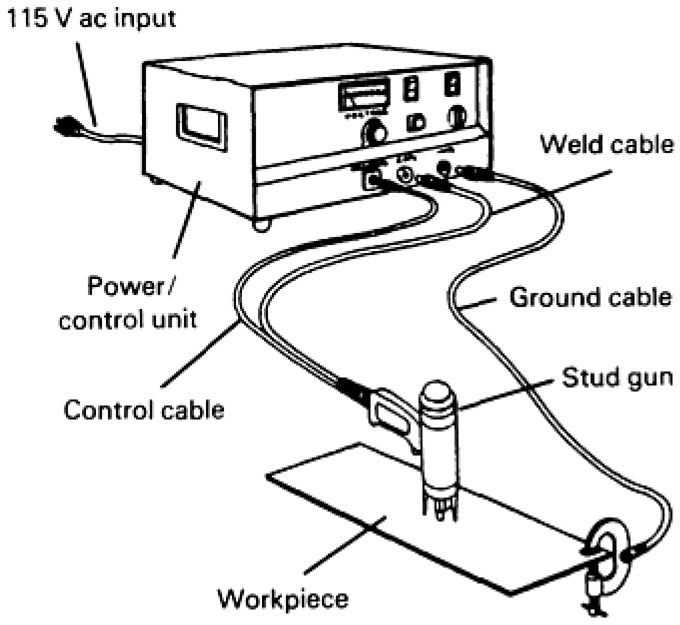

FIG. 24 KEY COMPONENTS OF A PORTABLE CAPACITOR DISCHARGE STUD WELDING SYSTEM

The three different methods of capacitor discharge stud welding include: initial contact, initial gap (Fig. 25), and drawn arc. The difference among these three methods is in the manner of arc initiation. The initial contact and initial gap methods utilize a small, specially engineered projection or tip to obtain arc initiation, whereas the drawn arc method initiates the arc by drawing the stud away from the workpiece.

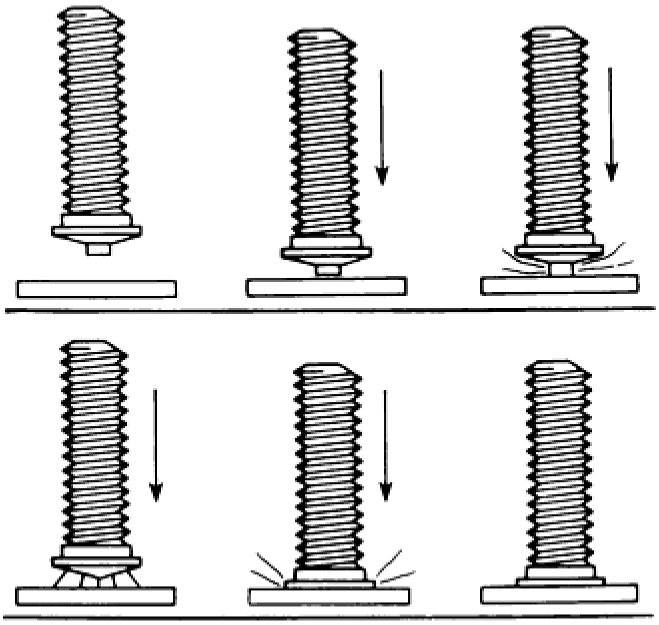

FIG. 25 SEQUENCE OF OPERATIONS REQUIRED TO WELD A STUD TO A WORKPIECE USING THE INITIAL GAP METHOD OF CAPACITOR DISCHARGE STUD WELDING

Studs. Low-carbon steel, stainless steel, aluminum, and brass are the materials used most often for studs welded commercially with the capacitor discharge method. The preferred weld base of a capacitor discharge welding stud is round, and the shank may be of almost any shape--threaded, unthreaded, round, square, rectangular, tapered, or grooved. The maximum weld-base diameter being welded commercially with the capacitor discharge welding process is 13 mm ($\frac{1}{2}$ in.).

Weld Current/Weld Time Relationships. The weld current process ranges from approximately 600 to 25,000 A, depending on stud size and weld method. The weld current range with the drawn arc method is approximately 600 to 3000 A, whereas with the initial contact or gap method the range is approximately 2500 to 25,000 A. Weld time with the drawn arc method is generally in the range of 4 to 6 ms, whereas with the initial contact and gap method it is generally in the range of 1 to 3 ms.

Metals Welded. The capacitor discharge stud welding process is an arc welding process, but differs from other arc welding processes in that the weld is made in a few milliseconds. For this reason, acceptable metallurgical results can be obtained without shielding the arc or using a flux. The welding of aluminum using the drawn arc method of capacitor discharge stud welding does require a shielding gas, however. Because of the speed at which the weld takes place, very little metal is melted. Weld penetration is slight, and very little stud metal intermixes with the plate metal. Therefore, studs may be welded to very thin metals without melt-through, and many dissimilar metals can be welded with acceptable results. Various stud and plate material combinations can be welded (Table 14).

TABLE 14 BASE METAL/STUD METAL COMBINATIONS THAT CAN BE SUCCESSFULLY JOINED USING CAPACITOR DISCHARGE STUD WELDING

BASE METAL	STUD METAL				
	LOW-CARBON STEEL[A]	**STAINLESS STEEL (300 SERIES)**[B]	**LEAD-FREE COPPER ALLOY**	**ALUMINUM ALLOY (1100 SERIES, 5000 SERIES, 6000 SERIES)**	TITANIUM OR TITANIUM ALLOY

LOW-CARBON STEEL (1006-1025)	X	X	X	. . .	...
STAINLESS STEEL (300 SERIES, 400 SERIES)	X	X	X	. . .	...
ALUMINUM ALLOY (1100 SERIES, 3000 SERIES, 5000 SERIES, 6000 SERIES)	. . .	. . .	. . .	X	...
COPPER ALLOY (LEAD-FREE BRASS)	X	X	X	. . .	...
ZINC ALLOY (DIE CAST)	. . .	. . .	. . .	X	...
TITANIUM OR TITANIUM ALLOY	**...**	**...**	**...**	**...**	**X**

(A) COMPOSITION: 0.23% C (MAX), 0.60% MN (MAX), 0.04% P (MAX), 0.05% S (MAX).

(B) WITH EXCEPTION OF FREE-MACHINING TYPE 303 STAINLESS STEEL

Procedure Development and Practice Considerations for Resistance Welding

Resistance Welding of Stainless Steels

Stainless steels are readily resistance welded by spot, seam, projection, flash, upset, high-frequency, and capacitor discharge. Generally, the weld time and welding current are less than those used for welding carbon steel, but the electrode force is usually greater. Austenitic stainless steels of the 300 series are resistance welded more often than any other metal except lowcarbon steel. Ideally, stainless steel to be resistance welded should contain a maximum of 0.08% C (as in types 304, 316, and 347), although steels of higher carbon content (such as types 301, 302, 309, and 310) can be successfully resistance welded.

The martensitic and ferritic types of stainless steel can be welded satisfactorily. The martensitic types are less frequently resistance welded, because joints made in them are hard and brittle in the as-welded condition.

Equipment

Resistance welding of stainless steel requires less transformer capacity than for equivalent thicknesses of other metal, under the same conditions. The welding machines use a singlephase alternating-current power supply, a three-phase rectifier, or a frequency-converter power supply. A weld made with a three-phase power supply usually requires slightly higher current and a longer weld time than a weld made with a single-phase supply.

Factors Affecting Resistance Welding of Stainless Steels

The characteristics of stainless steel that affect resistance welding include electrical resistivity, thermal conductivity, melting temperature, strength at elevated temperatures, coefficient of thermal expansion, and contact resistance.

Electrical Resistivity. Stainless steels have much higher electrical resistance than carbon steels. As a result, more heat is generated in a stainless steel with the same current. Therefore, resistance welding of a stainless steel requires lower currents or shorter weld times than welding of a carbon steel.

Thermal Conductivity. Stainless steel has a lower thermal conductivity than carbon steel and, therefore, heat is conducted away from the weld zone more slowly. This, in conjunction with the electrical resistivity, means that less heat has to be applied to reach the melting temperature.

Melting temperatures of stainless steels have an effect on the amount of heat required to produce fusion for welding. The austenitic stainless steels melt in various ranges between 1370 and 1455 °C (2500 and 2650 °F), and the martensitic and ferritic alloys melt in ranges between 1400 and 1530 °C (2550 and 2790 °F). Plain low-carbon steels melt at temperatures between 1480 and 1540 °C (2700 and 2800 °F).

Coefficient of Thermal Expansion. Austenitic and nitrogen-strengthened austenitic stainless steel expands and contracts with changing temperature almost 50% more than does plain carbon steel. These dimensional changes and the slower heat diffusion in austenitic grades result in greater thermal stress, which leads to warping. The ferritic, martensitic, and precipitation-hardening grades of stainless steel have coefficients of thermal expansion from 6 to 11% lower than that of plain carbon steel.

High strength at room and elevated temperatures of nitrogen-strengthened stainless steels and, to a lesser extent, of straight-chromium grades makes it necessary to use greater electrode force than is required for carbon steel to bring the work-metal surfaces together for the required intimate contact at points of welding.

Contact resistance of stainless steel is higher than that of carbon steel and, therefore, greater electrode pressure is needed to make good resistance welds. In addition, the welding process is more stable because the base material is not as susceptible to surface contamination.

Welding Characteristics of Stainless Steels

Austenitic Stainless Steels. All of the austenitic stainless steels in the solution-annealed condition contain carbon in solution. When the steel is heated to the temperature range of 425 to 815 °C (800 to 1500 °F), as in the weld heat-affected zone (HAZ), carbon combines with chromium, resulting in chromium carbide precipitation at the grain boundaries. Precipitation of chromium carbide (sensitization) makes the material susceptible to intergranular corrosion. Sensitization to intergranular corrosion is influenced in austenitic stainless steel by (1) the carbon content, (2) the solubility of the carbon, (3) the presence of one or more stabilizing elements, (4) the proximity of the actual metal temperature to 1200 °F, and (5) the time period during which the metal is held in that temperature range. The first three factors are controllable only through selection of composition; the latter two are controllable by adjustment of welding conditions such as heat input, size of spot, production rate, and provisions for cooling.

Martensitic stainless steels most often welded are types 403,410,414, and 431. These steels can be resistance welded in the annealed, hardened or hardened-and-tempered condition. Regardless of prior condition, welding produces a hardened martensitic zone adjacent to the weld. The hardness of this zone depends mainly on carbon content, although it can be controlled to a degree by welding procedure. As the hardness of the metal in the HAZ increases, its susceptibility to cracking increases and its toughness decreases. Steels having a maximum carbon content of 0.15%, such as types 403 and 410, often produce satisfactory welds without postweld heat treatment. Steels with higher carbon content, such as types 420 and 440A, generally require postweld heat treatment.

Ferritic grades of stainless steel most often welded are types 405,430,442, and 446. Resistance welds in these steels exhibit reduced ductility at room temperature because of grain growth in the HAZ. Ferritic stainless steels can also become embrittled if heated to approximately 475 °C (885 °F). This phenomenon, known as "885 °F embrittlement," may be encountered in the HAZ of a ferritic stainless steel weld. If properties in the as-welded condition are unsuitable, annealing must follow welding. Postweld annealing, in addition to improving ductility, helps restore normal corrosion resistance.

Precipitation-hardening grades of stainless steel can be resistance welded in the annealed or aged (hardened) condition. If they are welded in the aged condition, the strength of the weld area is significantly reduced because the heat of welding anneals the material subjected to it. Generally, most of the strength may be restored by simply re-aging the weldment. When maximum strength is required in the weld area, it is best to resistance weld these grades in the annealed condition and then age the weldment.

Resistance Welding of Aluminum Alloys

Aluminum alloys, both of the non-heat-treatable and heat-treatable types, either wrought or cast, can be resistance welded. Some of these alloys are welded more readily than others. Gas shielding may be required to maximize joint quality. Characteristics of aluminum alloys include comparatively high thermal and electrical conductivity, a relatively narrow plastic range (about 95 to 205 °C, or 200 to 400 °F, temperature differential between softening and melting), considerable shrinkage during cooling, a troublesome surface oxide, and an affinity for copper electrode materials.

Base-Metal Characteristics

Although aluminum alloys can be resistance spot and seam welded, some alloys or combinations of alloys have higher as-welded properties than others. Table 15 gives melting ranges, electrical and thermal conductivities, and resistance weldability of some wrought alloys and casting alloys.

TABLE 15 MELTING RANGES, ELECTRICAL AND THERMAL CONDUCTIVITIES, AND RESISTANCE WELDABILITY OF COMMON ALUMINUM ALLOYS

Wrought and casting alloys are identified by Aluminum Association designations.

ALLOY AND TEMPER	**MELTING RANGE**		**ELECTRICAL CONDUCTIVITY, %IACS(A)**	**RELATIVE THERMAL CONDUCTIVITY(B),%**	RESISTANCE WELDABILITY(C)
	°C	**°F**			
NON-HEAT-TREATABLE WROUGHT ALUMINUM ALLOYS					
1350-H19	646-657	1195-1215	62	60	**ST**
1060-H18	646-657	1195-1215	61	57	**ST**
1100-H18	643-657	1190-1215	57	55	**RW**
3003-H18	643-654	1190-1210	40	39	**RW**
3004-H38	629-652	1165-1205	42	42	**RW**
5005-H38	632-652	1170-1205	52	51	**RW**
5050-H38	627-652	1160-1205	50	49	**RW**
5052-H38	593-649	1100-1200	35	35	**RW**
5083-H321	574-638	1065-1180	29	30	**RW**
5086-H34	584-640	1084-1184	31	32	**RW**
5154-H38	593-643	1100-1190	32	32	**RW**
5182-O	574-640	1065-1185	31	31	**RW**
5454-H34	602-646	1115-1195	34	34	**RW**
5456-H321	571-638	1060-1180	29	30	**RW**
HEAT-TREATABLE WROUGHT ALUMINUM ALLOYS					
2014-T6	510-638	950-1180	40	39	**ST**
2024-T361	501-638	935-1180	30	31	**ST**
2036-T4	554-649	1030-1200	41	40	**RW**
2219-T37	543-643	1010-1190	28	29	**ST**
6009-T4	560-649	1040-1200	44	43	**RW**
6010-T4	585-	1085-	39	38	**RW**

	649	1200			
6061-T6	593-649	1100-1200	43	43	**RW**
6063-T6	615-654	1140-1210	53	51	**RW**
6101-T6	615-652	1140-1205	57	55	**RW**
7075-T6	477-638	890-1180	33	33	**ST**
ALUMINUM CASTING ALLOYS					
413.0-F	574-582	1065-1080	31	32	**LW**
443.0-F	574-632	1065-1170	37	37	**RW**
308.0-F	521-613	970-1135	37	37	**ST**
238.0-F	507-599	945-1110	25	26	**LW**
513.0-F	579-638	1075-1180	34	34	**ST**
520.0-T4	449-604	840-1120	21	22	**NR**
333.0-T6	516-585	960-1085	29	30	**ST**
C355.0-T61	546-621	1015-1150	39	38	**ST**
356.0-T6	557-613	1035-1135	39	38	**ST**
712-F	**604-643**	**1120-1190**	**40**	**39**	**RW**

(A) INTERNATIONAL ANNEALED COPPER STANDARD. VOLUME BASIS AT 20 °C (68 °F). FOR COMPARISON, COPPER ALLOY 102 (OXYGEN-FREE COPPER) IS 101% AND LOW-CARBON (1010) STEEL ABOUT 14%.

(B) BASED ON COPPER ALLOY 102 AS 100%. WHICH HAS A THERMAL CONDUCTIVITY OF 391 W/M · K (226 BTU/FT · H · °F) AT 20 °C (68 °F). LOW CARBON STEEL HAS A THERMAL CONDUCTIVITY OF ABOUT 13% ON THIS RELATIVE SCALE.

(C) RW, READILY WELDABLE; ST, WELDABLE IN MOST APPLICATIONS BUT MAY REQUIRE SPECIAL TECHNIQUES FOR SPECIFIC APPLICATIONS; LW. LIMITED WELDABILITY AND USUALLY REQUIRES SPECIAL TECHNIQUES; NR. WELDING NOT RECOMMENDED

Alclad Alloys. Resistance welding is also done on Alclad products made by roll cladding some of the alloys listed in Table 15 with a thin layer of aluminum or aluminum alloy. Because this layer is anodic to the core alloy, it provides electrochemical protection for exposed areas of the core. Alclad alloys 2219, 3003, 3004, 6061, and 7075 have a cladding of alloy 7072, which contains 1% Zn; Alclad alloy 2014 has a cladding of alloy 6003 or sometimes alloy 6053, both of which contain about 1.2% Mg; and Alclad alloy 2024 has a cladding of alloy 1230, which contains a minimum of 99.3% Al.

Effects on Weldability. The hardness of an alloy is one variable influencing weldability. Any alloy in the annealed condition (O temper) is more difficult to weld than the same alloy in a harder temper. In general, alloys in the softer tempers are much more susceptible to excessive indentation and sheet separation and to low or inconsistent weld strength. Greater deformation under the welding force causes an increase in the contact area and variations in the distribution of current and pressure. Therefore, welding of aluminum alloys in the annealed condition or in the softer tempers is not recommended without special electromechanical or electronic controls.

High-strength alloys such as 2024 and 7075 are easy to resistance weld, but may require applications of a forge pressure, because they are more susceptible to cracking and porosity than the lower-strength alloys.

Resistance Welding Machines

Aluminum alloys can be resistance welded with single-phase direct-energy, three-phase direct-energy, and stored-energy machines. Best results are obtained by using a machine that has these features:

- ABILITY TO HANDLE HIGH WELDING CURRENTS FOR SHORT WELD TIMES
- SYNCHRONOUS ELECTRONIC CONTROLS FOR WELD TIME AND WELDING CURRENT
- LOW-INERTIA WELDING HEAD FOR RAPID FOLLOW-UP OF ELECTRODE FORCE
- SLOPE CONTROL (FOR SINGLE-PHASE WELDING MACHINES)
- MULTIPLE-ELECTRODE-FORCE SYSTEM TO PERMIT PROPER FORGING OF THE WELD NUGGET AND REDRESSING OF ELECTRODES

Electrodes and Electrode Holders

Selection of electrode material and face shape, maintenance of the face, and cooling of the electrode are important in producing consistent spot and seam welds in aluminum alloys.

Copper Alloy Electrodes. Resistance Welder Manufacturers Association (RWMA) classes 1, 2, and 3 are used for welding aluminum alloys. These electrode materials have high electrical and thermal conductivities, which combined with adequate cooling help keep the temperature of the electrode below the temperature at which aluminum will alloy with copper and cause electrode pickup.

Design of spot welding electrodes suitable for spot welding of aluminum includes both straight and offset electrodes. Construction details of each are shown in Fig. 26. Straight electrodes should be used whenever possible, because deflection and skidding may occur with offset electrodes under similar welding conditions. If offset electrodes are used, the amount of offset should be the minimum permitted by the shape of the assembly being welded.

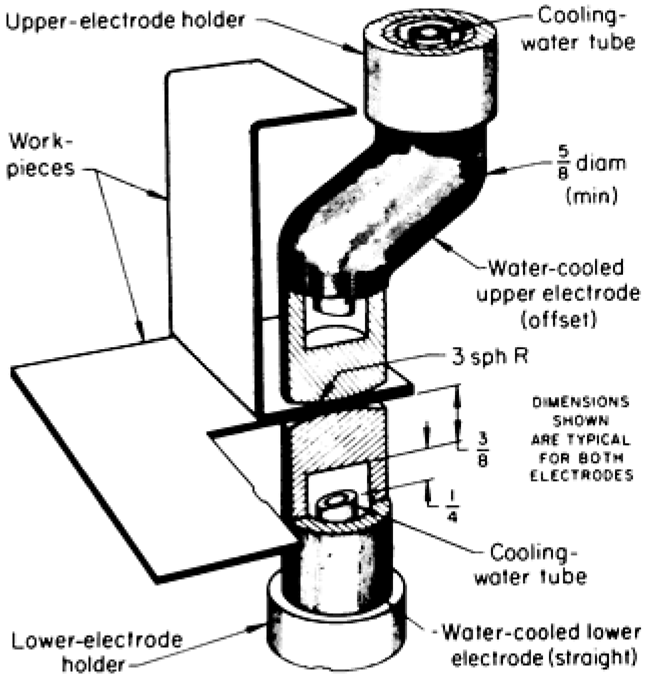
Upper-electrode holder
Cooling-water tube
Work-pieces
$\frac{5}{8}$ diam (min)
Water-cooled upper electrode (offset)
3 sph R
DIMENSIONS SHOWN ARE TYPICAL FOR BOTH ELECTRODES
$\frac{3}{8}$
$\frac{1}{4}$
Cooling-water tube
Lower-electrode holder
Water-cooled lower electrode (straight)

FIG. 26 CUTAWAY VIEW SHOWING INTERNAL COMPONENTS OF STRAIGHT AND OFFSET RADIUS-FACE ELECTRODES USED IN RESISTANCE SPOT WELDING OF ALUMINUM ALLOYS

Only electrodes that have the cooling-water hole within 9.5 mm ($\frac{3}{8}$ in.) of the face surface should be used.

Preweld Surface Preparation

Although welds for some purposes can be made satisfactorily without any preweld surface preparation, welds that are free of cracks, porosity and sheet separation, and that have the most uniform strength and symmetry, are obtained only with correct procedures for cleaning and reduction or removal of oxide film. In addition, adequate surface preparation reduces electrode contamination.

Resistance Welding of Copper and Copper Alloys

Resistance spot welding is widely used for joining copper and copper alloys. Principal applications include welding sections up to about 1.52 mm (0.060 in.) thick, particularly those alloys with low electrical conductivities. Many copper alloys with low conductivities can be seam welded easily. Coppers with a low weldability factor are difficult to seam weld. Projection welding is not recommended for copper or for most brasses. Bronzes can be projection welded with satisfactory results in many applications.

Welding Characteristics

The resistance weldability of any copper alloy is inversely proportional to its electrical and thermal conductivities. Generally, alloys with lower conductivities are easier to weld. Compared with steel, most copper alloys require shorter weld time, lower electrode force, higher current, and different electrode materials that are compatible with the alloy being welded.

Welding Equipment

Single-phase and three-phase direct-energy and electrostatic stored-energy (capacitor-discharge) welding machines are used for resistance welding of copper and copper alloys. The addition of slope control to single-phase direct-energy welding machines is not necessary for spot welding most copper alloys. In welding high-zinc brasses, the use of upslope can result in an increase of as much as 20% in weld strength. Downslope is not recommended for welding of any of the copper alloys.

Welding Machine Controls. Copper alloys are particularly sensitive to variations in welding conditions, and therefore all direct-energy machines used for welding these alloys should be equipped with synchronous electronic controls, especially in applications requiring short weld times. These devices are capable of controlling weld time and welding current for repeated operations with extreme accuracy.

Electrodes

The current used for resistance welding of copper alloys is much higher than that used for welding low-carbon steel, and therefore, the electrode must have high electrical conductivity to minimize heat buildup.

Electrode Materials. The Resistance Welder Manufacturers Association (RWMA) class 1 electrode materials (typically tungsten or molybdenum alloys), containing copper and cadmium, are sometimes used for welding copper and high-conductivity brass and bronze. Class 2 materials, containing copper and chromium, are used on low-conductivity brass and bronze and the copper-nickel alloys. Class 3 materials are used in electrodes for seam welding.

Electrodes must be sufficiently water cooled to minimize sticking to the work metal and to prolong their life. Face contours must be carefully prepared and the electrodes must be properly aligned for welding.

Selection of Process

Weldability of the work metal often determines which process should be used for a given application. Some of the coppers and copper alloys can be spot welded, but not seam welded because of high conductivity, and not projection welded because of low compressive strength of the projections at elevated temperature.

Coppers

Coppers and copper alloys having electrical conductivity higher than about 30% IACS are the least well suited for resistance spot, projection, or seam welding, mainly because of severe electrode pickup.

Thin copper stock can be welded using electrodes faced with RWMA class 13 (tungsten) or class 14 (molybdenum), but surface appearance is poor and frequent electrode maintenance is required. A tinned coating on wire or sheet is helpful in welding copper.

Beryllium Copper

Beryllium copper alloys can be resistance welded most successfully in thin gages. Spot welding produces satisfactory welds; seam welding is less successful. Projection welding is satisfactory, provided that the projections can be formed with the work metal in the annealed condition and without cracking the work metal around the projection. Close control of welding conditions is required for consistent weld size and joint strength.

Oxide films produced by heat treating must be removed to ensure low and consistent contact resistance. Work metals that have not been heated after rolling frequently need only degreasing before welding.

Low electrical conductivity (22% IACS for alloys C17000 and C17200) contributes to the weldability of beryllium copper alloys. However, they are more difficult to resistance weld than low-carbon steel. Alloy C17500 has an electrical conductivity of 45% IACS and is more difficult to resistance weld than higher-strength, lower-conductivity beryllium copper.

Low- and High-Zinc Brasses

The low-zinc brasses are difficult to weld, although easier than copper, and are subject to electrode pickup. Welds made in these brasses may lack strength, principally because of comparatively high electrical conductivity (32 to 56% IACS).

The high-zinc brasses have an electrical conductivity of 27 to 28% IACS and can be both spot and projection welded over a wide range of conditions. Electrode pickup can be a problem unless weld time, welding current, and electrode force are properly selected.

Excessive electrode pickup and blowthrough of the weld may occur when long weld times, high energy input, and low electrode forces are used. Yellow brasses (alloys C26800 and C27000) are less susceptible to electrode pickup than cartridge brass except when long weld times and high energy input are used. Electrode force should be sufficient to prevent arcing or expulsion of molten metal, to which these alloys are subject because of their 30 to 40% Zn content, which boils at about 905 °C (1665 °F). As shown in Table 16, the recommended electrode force, when using electrodes having a face diameter of 4.8 mm ($\frac{3}{16}$ in.), is approximately 1.8 kN (400 lbf).

TABLE 16 GUIDELINES FOR RESISTANCE SPOT WELDING OF SELECTED COPPER ALLOYS

UNS NO.	**ALLOY NAME**	WELDING PARAMETER[A]			
		WELD TIME, CYCLES	**ELECTRODE FORCE**		WELDING CURRENT,A
			KN	**LBF**	
C23000	RED BRASS	6	1.8	400	**25,000**
C24000	LOW BRASS	6	1.8	400	**24,000**
C26000	CARTRIDGE BRASS	4	1.8	400	**25,000**
C26800-C27000	YELLOW BRASS	4	1.8	400	**24,000**
C28000	MUNTZ METAL	4	1.8	400	**21,000**

C51000-C52400	PHOSPHOR BRONZE	6	2.3	510	**19,500**
C62800	ALUMINUM BRONZE	4	2.3	510	**21,000**
C65100-C65500	SILICON BRONZE	6	1.8	400	**16,500**
C66700	MANGANESE BRASS	6	1.8	400	**22,000**
C68700	ALUMINUM BRASS	4	1.8	400	**24,000**
C69200	**SILICON BRASS**	**6**	**2.3**	**510**	**22,000**

(A) FOR SPOT WELDING 0.91 MM (0.036 IN.) THICK SHEET USING RWMA TYPE E ELECTRODES WITH 4.8 MM ($\frac{3}{16}$ IN.) DIAM FACE AND 30 ° CHAMFER AND MADE OF RWMA CLASS 1 MATERIAL

Copper Nickels

The copper-nickel alloys have electrical conductivities of 4.5 to 11% IACS, are readily spot and seam welded with relatively low welding current, and generally do not alloy with the electrode material and cause electrode pickup.

Bronzes

The phosphor bronzes, except alloy C50500, which is not recommended for resistance spot and seam welding because of its high electrical conductivity (48% IACS), have relatively low electrical conductivity (11 to 20% IACS) and am readily spot and seam welded using low welding currents. Electrode pickup can be reduced by use of a type F (radius) electrode face and frequent redressing to keep the face clean and smooth. Hot shortness can be minimized by supporting the workpieces to prevent strain during welding and by using a greater minimum overlap than recommended by the data in Table 17.

TABLE 17 GUIDELINES FOR RESISTANCE SPOT WELDING OF HIGH-ZINC BRASSES

THICKNESS OF THINNEST SHEET		**MINIMUM SPOT SPACING**		**MINIMUM CONTACTING OVERLAP**$^{(A)}$		SHEAR LOAD OF JOINT	
mm	**in.**	**mm**	**in.**	**mm**	**in.**	**kn**	lbf
0.81	0.032	16	$\frac{5}{8}$	13	$\frac{1}{2}$	1.47	**330**
1.27	0.050	16	$\frac{5}{8}$	16	$\frac{5}{8}$	2.28	**512**
1.63	0.064	19	$\frac{3}{4}$	19	$\frac{3}{4}$	3.02	**680**
2.39	0,094	25	1	25	1	5.20	**1168**
3.18	**0.125**	**38**	**1** $\frac{1}{2}$	**32**	**1** $\frac{1}{4}$	**8.33**	**1872**

(A) MINIMUM EDGE DISTANCE IS EQUAL TO ONE HALF THE CONTACTING OVERLAP.

Nickel Silvers

Nickel silvers, which have about the same conductivities (6 to 10.9% IACS) as copper nickels, are spot welded as readily as copper nickels but are more difficult to seam weld. Surface contaminants such as lead and bismuth (which form low-melting eutectics with copper and nickel) or sulfur (which may be introduced in forming) must be removed before resistance welding.

References

1. L.P. CONNER, ED., *WELDING HANDBOOK,* 8TH ED., VOL 1, AWS, 1987, P 16
2. J.F. YOUNG AND R.S. SHANE, ED., *MATERIALS AND PROCESSES, PART B: PROCESSES,* 3RD ED., MARCEL DEKKER, 1985, P 11257-1126
3. J.G. BRALLA, ED., *HANDBOOK OF PRODUCT DESIGN FOR MANUFACTURING,* MCGRAW-HILL, 1986, P 7-35
4. J.F. YOUNG AND R.S. SHANE, ED., *MATERIALS AND PROCESSES, PART B: PROCESSES,* 3RD ED., MARCEL DEKKER, 1985, P 1127
5. J.G. BRALLA, ED., *HANDBOOK OF PRODUCT DESIGN FOR MANUFACTURING,* MCGRAW-HILL, 1986, P 7-38
6. R.L. O'BRIEN, ED., *WELDING HANDBOOK,* 8TH ED., VOL 2, AWS, 1991, P 552
7. J.G. BRALLA, ED., *HANDBOOK OF PRODUCT DESIGN FOR MANUFACTURING,* MCGRAW-HILL, 1986, P 7-39
8. "RECOMMENDED PRACTICES FOR RESISTANCE WELDING," AWS C1.1-66, AWS, 1966
9. *RESISTANCE WELDING MANUAL,* 3RD ED., VOL 1, RESISTANCE WELDER MANUFACTURERS ASSOCIATION, 1956, P 88
10. R.L. O'BRIEN, ED., *WELDING HANDBOOK,* 8TH ED., VOL 2, AWS, 1991, P 593
11. H.B. CARY, *MODERN WELDING TECHNOLOGY,* 2ND ED., PRENTICE-HALL, 1989, P 251-252

Procedure Development and Practice Considerations for Electron-Beam Welding

Introduction

ELECTRON-BEAM WELDING (EBW) can produce deep, narrow, and almost parallel-sided welds with low total heat input and relatively narrow heat-affected zones in a wide variety of common and exotic metals. In many applications, a single pass is sufficient to weld the components. When initially applied to industrial applications in the 1950s, EBW had to be performed totally in a vacuum environment. In this form, EBW was primarily used because of its ability to reliably produce very high quality welds in materials normally considered extremely difficult to join (for example, refractory metals, oxidizable metals, stainless steels, and superalloys). This made it ideally suited for use in aerospace, nuclear, and other similar (high-technology) manufacturing industries.

During the 1960s, the capability of employing the EBW process for medium-vacuum and nonvacuum welding tasks was developed. This increased production capacity helped make EBW more attractive to other industries desiring to take advantage of the numerous joining benefits it provided. However, the use of EBW required that all applications have a demonstrated capacity for fairly high-volume production.

Today, EBW is used to join not only difficult-to-weld materials but also common materials such as steel (≤150 mm, or 6 in., thick), aluminum (≤305 mm, or 12 in., thick), and copper (≤100 mm, or 4 in., thick) (Ref 1). These and other metals (in thicknesses ranging from foil to plate dimensions) can be electron-beam welded. In addition, EBW can be used to join numerous combinations of dissimilar metals. See the Section "Fusion Welding Processes" in this Handbook for a description of EBW systems and equipment.

Reference

1. A. SANDERSON AND K.R. NIGHTINGALE, *WELD. J.,* APRIL 1990, P 45

Procedure Development and Practice Considerations for Electron-Beam Welding

Weld Geometry

Part Configuration. Whether a component is welded in a vacuum or at atmospheric pressure (nonvacuum), part configuration and the corresponding joint designs used are critical parameters in the successful application of EBW. A circular weld in the axial direction, for example, that joins a disk-shaped member (gear or cup) to a shaft (Fig. 1) may experience severe constraint. Such joints are practical in easily welded metals--those that are soft, have low yield points, and exhibit low shrinkage--but not in difficult-to-weld steels that have high strength, hardenability, or machinability due to high sulfur content.

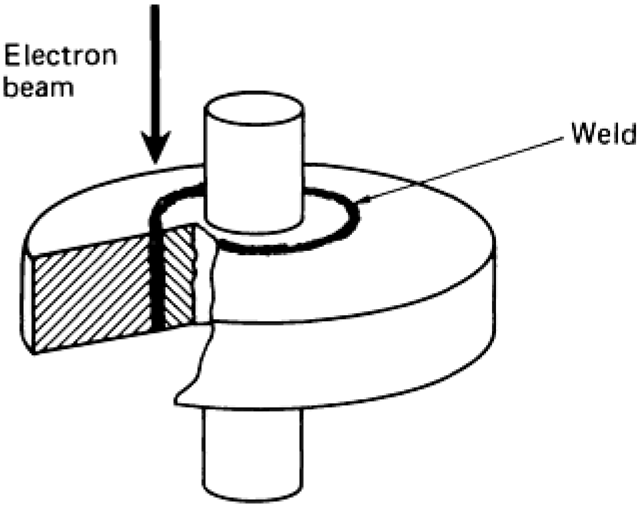

FIG. 1 CUTAWAY VIEW SHOWING LOCATION AND PENETRATION OF A CIRCULAR ELECTRON-BEAM WELD JOINING A DISK-SHAPED COMPONENT TO A SHAFT

The difficulties encountered in welding shapes such as those shown in Fig. 1 are aggravated if the weld zone is broadened by defocusing the electron beam. This is a frequent problem if exact beam position cannot be controlled, either because of beam instability, magnetism of the parts, or excessive tolerances in fixturing on high throughput machines. The need for beam broadening can be eliminated by the use of closed-loop beam position control systems.

Shrinkage stress is minimal in radial welds, as shown in Fig. 2. The volume of molten metal is less constrained by the axial direction if the gear hub is allowed to move by using push fit and no restraints. The weld must also have root clearances--the weld zone must not extend into the shaft. Stresses are lower if the melt zone has parallel side boundaries. Triangular weld zones that are caused by broadened beams, uneven beam power distribution, or incomplete penetration have more complex stresses which may result in cracking.

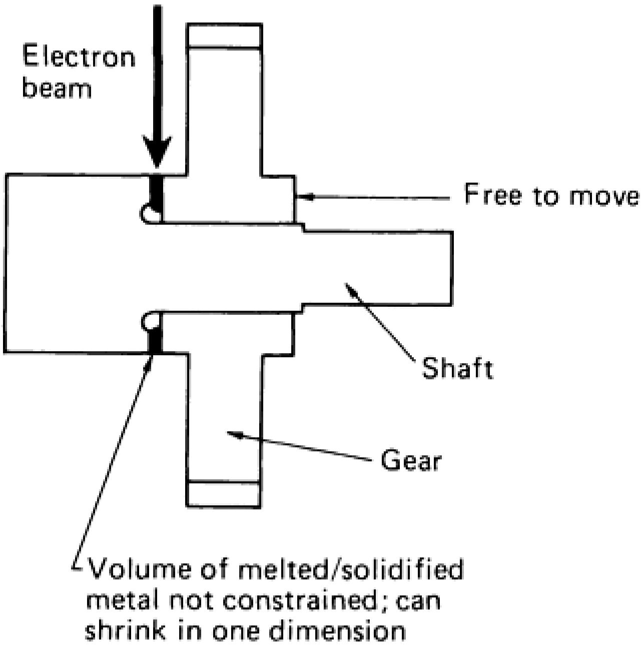
Electron beam
Free to move
Shaft
Gear
Volume of melted/solidified metal not constrained; can shrink in one dimension

FIG. 2 SHRINKAGE STRESS IN A RADIAL WELD

The same principles of joint design discussed above for joints between shafts and disklike members apply equally to T-joints and comer joints between plates. Figure 3 shows recommended and nonrecommended weld configurations. Although true for all welding methods, it is essential in EBW (which is usually performed in a single pass) that constraints on the volume of the melt must be avoided by self-aligning interlocking steps in the joint. In multilayer arc welds, the individual weld bead is less constrained in volume and shrinks from its surface downward. Functionally, it is not useful to make the cross section of the joint any less thick than the plates that are being joined. If two welds are made in succession, the second weld may be constrained by the first (Fig. 3), which may lead to solidification cracking. It is thus advisable to direct the beam at the joint parallel to the faying surfaces and to cause melting over the whole contact area.

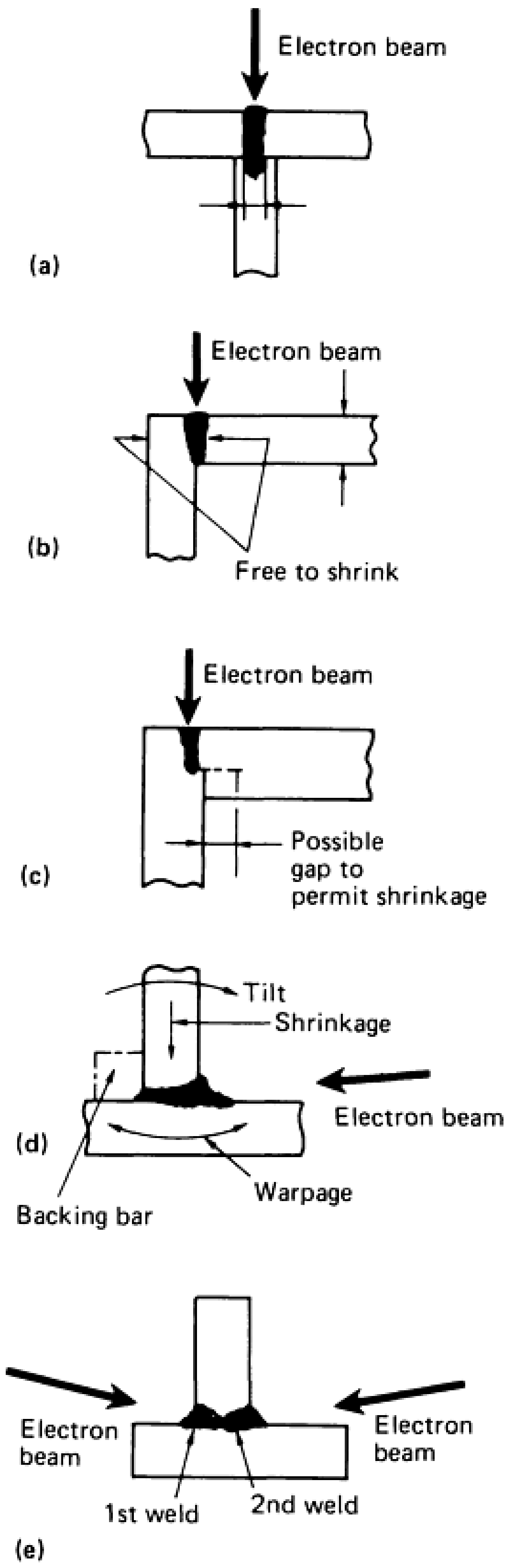
Electron beam
(a)
Electron beam
(b)
Free to shrink
Electron beam
(c)
Possible gap to permit shrinkage
Tilt
Shrinkage
Electron beam
(d)
Warpage
Backing bar
Electron beam
Electron beam
1st weld
2nd weld
(e)

FIG. 3 OPTIMUM VERSUS LEAST DESIRABLE WELD CONFIGURATIONS. (A) NOT RECOMMENDED--MAXIMUM CONFINEMENT OF MOLTEN METAL, MINIMUM JOINING CROSS SECTION (ARROWS); WASTES BEAM ENERGY FOR MELTING, NONFUNCTIONAL METAL. (B) MOST FAVORABLE--VOLUME OF MELT NOT CONFINED; MAXIMUM JOINING CROSS SECTION (ARROWS). (C) NOT RECOMMENDED--MAXIMUM CONFINEMENT OF MELT (UNLESS GAP IS PROVIDED); JOINING CROSS SECTION LESS THAN PLATE CROSS SECTION. (D) MOST FAVORABLE--MINIMUM CONSTRAINT AND CONFINEMENT OF MELT; MINIMUM INTERNAL STRESSES; WARPAGE CAN BE OFFSET BY BENDING PRIOR TO WELDING; TILT CAN BE OFFSET BY LOCATION OF T-ARM AT LESS THAN 90°

TO BASE PRIOR TO WELDING. FILLET OBTAINED BY PLACING WIRE IN RIGHT CORNER AND MELTING IT WITH THE BEAM. (E) NOT RECOMMENDED--TWO SUCCESSIVE WELDS; SECOND WELD IS FULLY CONSTRAINED BY THE FIRST WELD AND SHOWS STRONG TENDENCY TO CRACK.

Surface Geometry. Usually, EBW does not use or need filler wire. Therefore, neither V-grooves nor corners can be filled; in fact, they are detrimental. Surface tension pulls some of the molten metal up the V-shaped surfaces so that it is no longer available to fill the space created by the keyhole that is generated by the moving beam. Voids remain in the actual joint plane (Fig. 4).

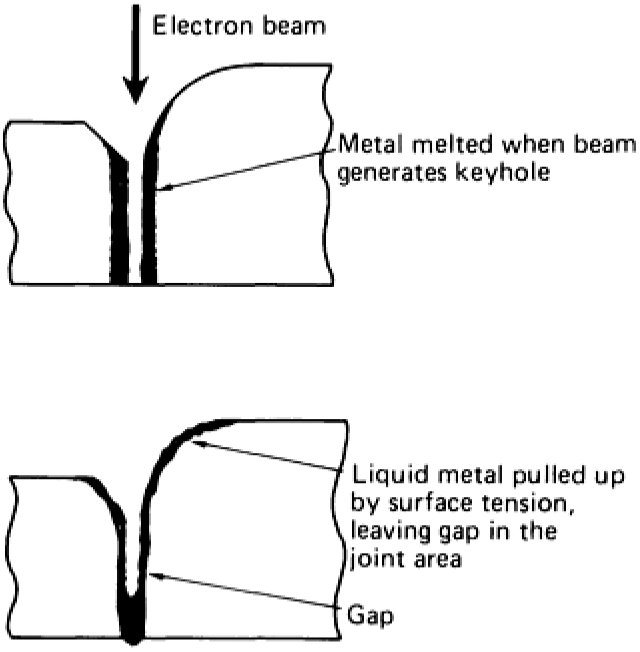
Electron beam
Metal melted when beam generates keyhole
Liquid metal pulled up by surface tension, leaving gap in the joint area
Gap

FIG. 4 GAP IN JOINT PLANE CREATED BY SURFACE TENSION

A step in the surfaces at the joint line is also undesirable. Any small lateral shift of the beam from the low to the high side, and vice versa, changes the penetration to some extent. In the extreme case, this shift may be equal to the height of the step, as shown in Fig. 5. On circular welds, with joint configurations similar to Fig. 5(a), a noncircular weld profile may develop with irregularities that can lead to uneven applied stresses. By contrast, a mere shift in the melt zone, as shown in Fig. 5(b), does not affect weld quality, as long as the entire joint line is covered.

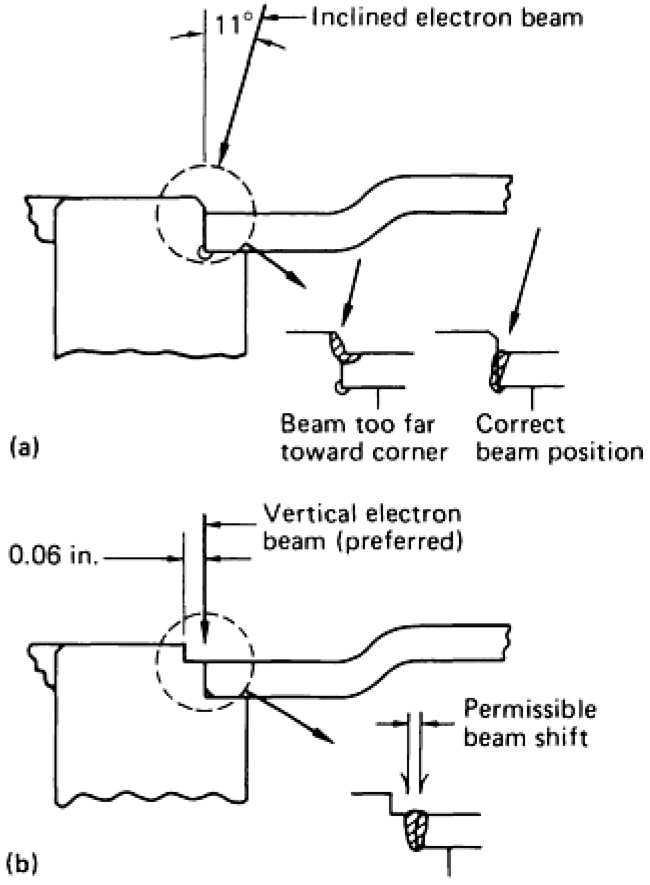
11°
Inclined electron beam
Beam too far toward corner
Correct beam position
(a)
0.06 in.
Vertical electron beam (preferred)
Permissible beam shift
(b)

FIG. 5 EFFECT OF BEAM AND MELT ZONE POSITION ON WELD PENETRATION. (A) INCLINED ELECTRON BEAM. (B) VERTICAL ELECTRON BEAM

Configurations for Wide Welds Bridging a Gap. The high speed and autogenous nature of an electron-beam weld can be advantageous when a wide weld (6.5 to 13 mm, or $\frac{1}{4}$ to $\frac{1}{2}$ in.) is needed without great depth-to-width ratio. The atmospheric electron beam has been used extensively to make such welds in mass-production automotive parts such as catalytic converter cases and die-cast aluminum intake manifolds.

The easiest weld to make with an electron beam is an edge-weld in the vertical position of the beam (weld horizontal and flat), as illustrated by the edge-flange weld in Fig. 6. In vacuum, the beam must be defocused; in air, the standoff must be short enough to prevent substantial beam scattering. The standoff is not critical because energy flow is not influenced by standoff as it is in electric arc welding.

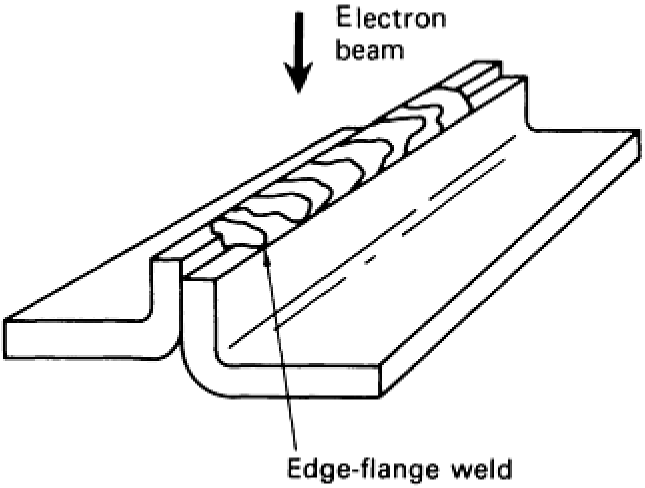

FIG. 6 EDGE-FLANGE WELD TYPICALLY USED TO JOIN THIN PANELS

For edge welds, changes in energy input per unit length and changes in power density affect only the amount of molten metal. Offset and gaps up to the thickness of the sheet to be joined often can be tolerated (Ref 2). In making an edge-flange weld (Fig. 6), the same advantages exist, provided the flange is high enough. For joining thinner sheet that is tightly clamped, edge welds also can be made with the electron beam in the horizontal position.

The nonvacuum EBW (EBW-NV) process can tolerate and fill large gaps if sufficient sacrificial metal is available to be melted by the beam (Fig. 7). If not, surface tension may pull whatever metal is melted to either side of the weld joint.

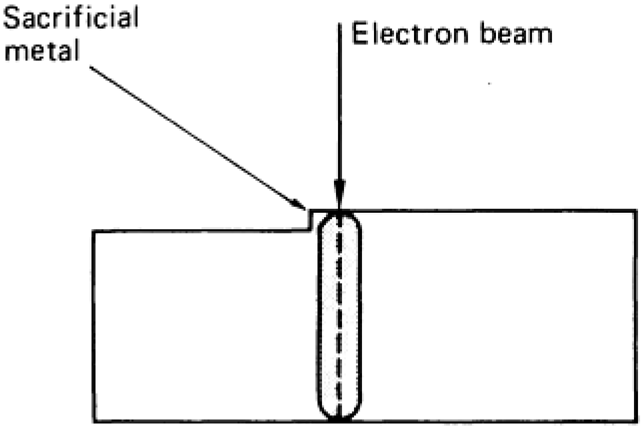

FIG. 7 USE OF SACRIFICIAL METAL TO FILL A LARGE GAP IN A SQUARE-GROOVE WELD

Melt-Zone Configuration. Welds with parallel sides of the melt zone (Fig. 8a) show minimum warpage and only transverse shrinkage when they are unconstrained. Triangular melt zones (Fig. 8b) cause warpage. If there are constraints, the hindered warpage and shrinkage cause internal stresses. A parallel-sided melt zone is preferred in either case; to obtain this configuration, joint design must allow for a full-penetration weld.

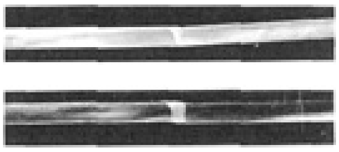

FIG. 8 VARIATION IN BEAM VELOCITY AND SUBSEQUENT WARPAGE OF 0.25 MM (0.10 IN.) THICK TYPE 304 STAINLESS STEEL JOINED BY A BUTT WELD. (A) BEAM TRAVERSE, 7110 MM/MIN (280 IN./MIN); TRIANGULAR MELT ZONE. (B) BEAM TRAVERSE, 4190 MM/MIN (165 IN./MIN); PARALLEL-SIDED MELT ZONE. WELD PARAMETERS: 12 KW BEAM POWER; 11.4 MM (0.45 IN.) STANDOFF; HELIUM SHIELDING GAS

Reference cited in this section

2. J.F. LOWRY, J.H. FINK, AND B.W. SCHUMACHER, A MAJOR ADVANCE IN HIGH-POWER ELECTRON BEAM WELDING IN AIR, *J. APPL. PHYS.*, VOL 47, 1976, P 95-106

Procedure Development and Practice Considerations for Electron-Beam Welding

Joint Design

Figures 9, 10, 11, 12, and 13 show the commonly used joint and weld types for EBW. Different joint preparations, joint designs, and welding positions are used to meet special requirements. For preferred joint configurations and shrinkage stresses encountered in various joint designs, see the section on weld geometry for electron-beam welding in this article.

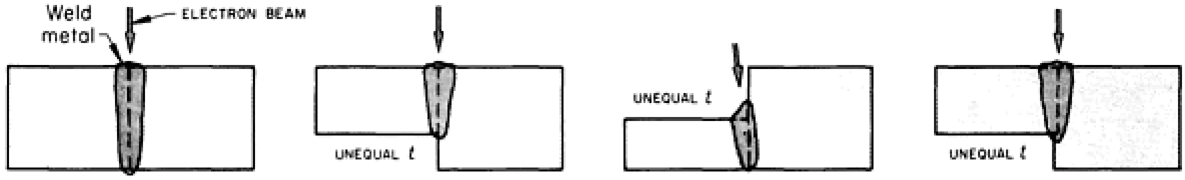

(a) Square-groove weld (b) Square-groove weld; flush joint (c) Square-groove weld; stepped joint (d) Square-groove weld; rabbeted, self-aligning joint

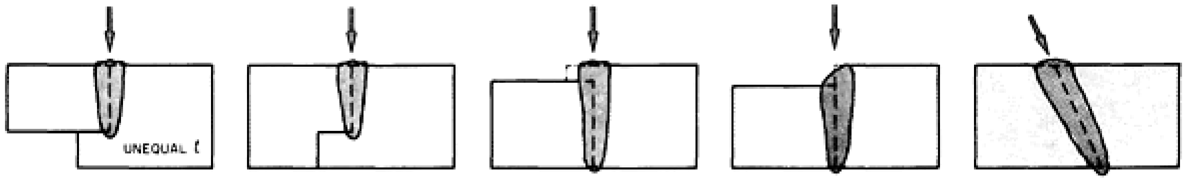

(e) Square-groove weld; rabbeted, self-backing joint (f) Square-groove weld; double-rabbeted, self-backing joint (g) Square-groove weld; integral-filler-metal lip joint (h) Square-groove weld; integral-filler-metal shoulder joint (j) Slant-groove weld; slant-butt joint

FIG. 9 TYPES OF BUTT JOINTS AND WELDS TYPICALLY GENERATED BY EBW

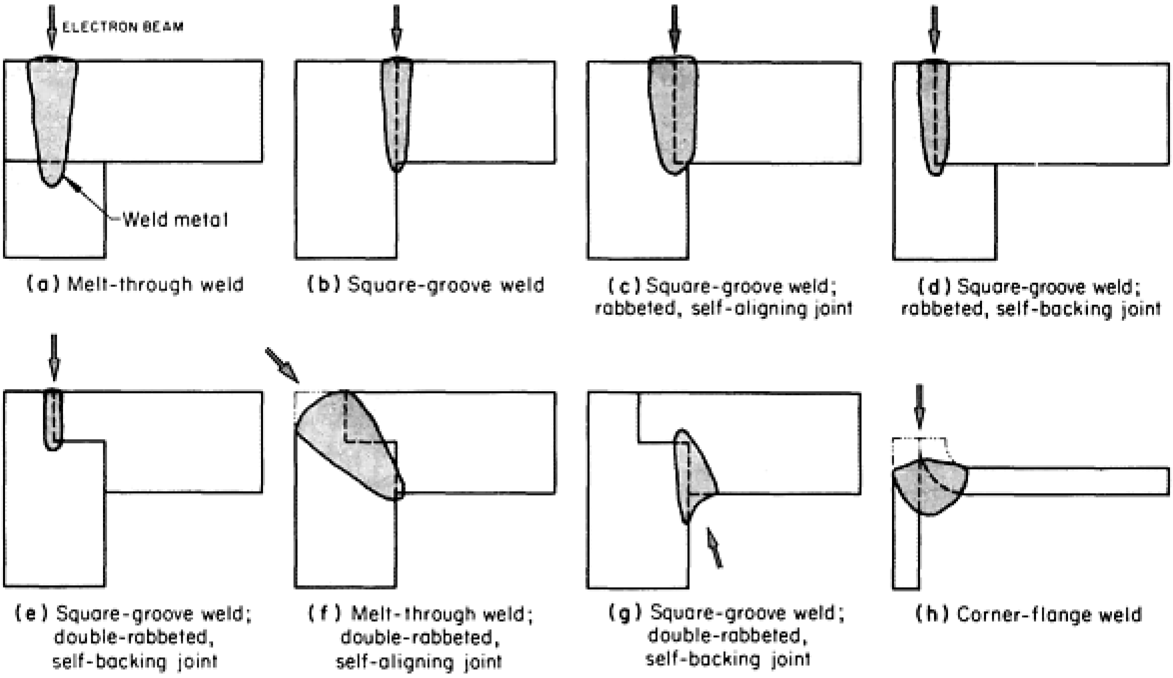
ELECTRON BEAM
Weld metal

(a) Melt-through weld (b) Square-groove weld (c) Square-groove weld; rabbeted, self-aligning joint (d) Square-groove weld; rabbeted, self-backing joint (e) Square-groove weld; double-rabbeted, self-backing joint (f) Melt-through weld; double-rabbeted, self-aligning joint (g) Square-groove weld; double-rabbeted, self-backing joint (h) Corner-flange weld

FIG. 10 TYPES OF CORNER JOINTS AND WELDS TYPICALLY OBTAINED WITH EBW

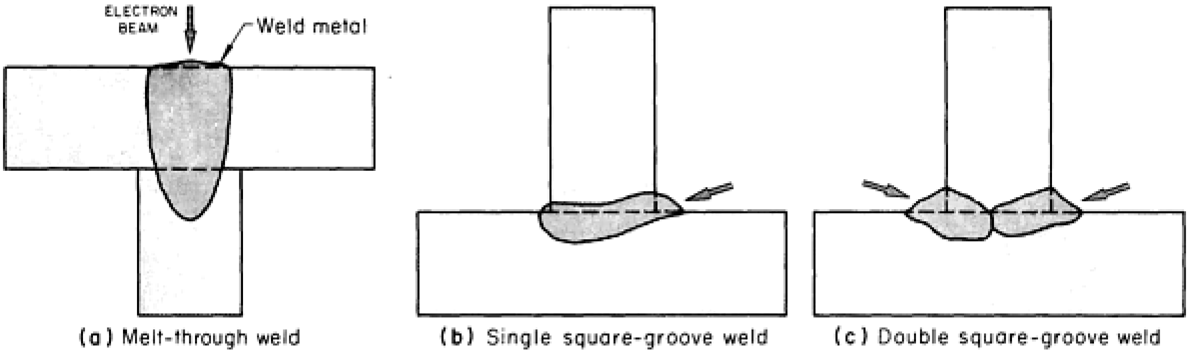
ELECTRON BEAM
Weld metal

(a) Melt-through weld (b) Single square-groove weld (c) Double square-groove weld

FIG. 11 TYPICAL WELDS PRODUCED BY EBW OF T-JOINTS

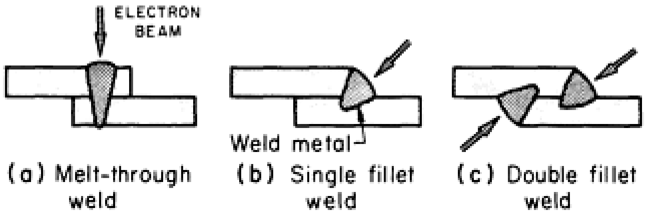
ELECTRON BEAM
Weld metal

(a) Melt-through weld (b) Single fillet weld (c) Double fillet weld

FIG. 12 TYPICAL WELDS OBTAINED BY EBW OF LAP JOINTS

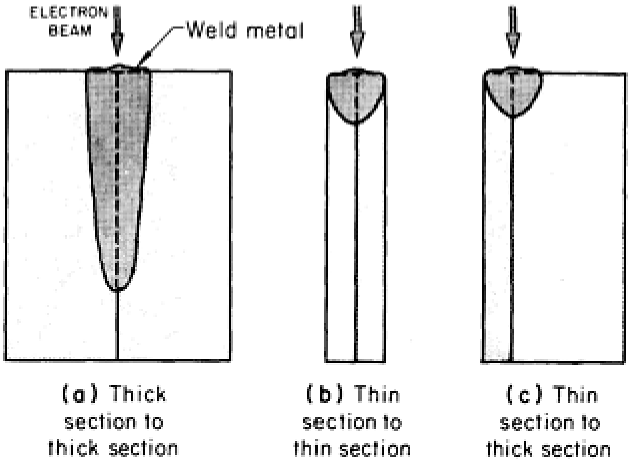
ELECTRON BEAM
Weld metal
(a) Thick section to thick section
(b) Thin section to thin section
(c) Thin section to thick section

FIG. 13 TYPICAL WELDS GENERATED BY EBW OF EDGE JOINTS HAVING COMPONENTS OF EQUAL AND UNEQUAL SECTION THICKNESSES

Welds in Butt Joints

Butt joints are the most common of the five basic joint types used in EBW. The basic square-groove weld, which is shown in Fig. 9(a), 9(b), and 9(c), needs only the simplest and least expensive joint preparation. It is suitable for either partial- or full-penetration welds. Good fit-up and external fixtures are needed. The flush joint (Fig. 9b) is preferred to the stepped joint (Fig. 9c) for joining unequal thicknesses, chiefly because control of conditions for making sound full-penetration welds is less critical than for the stepped joint. A wider beam is used in welding the stepped joint, and the beam angle must he carefully controlled to avoid scarfing the upper edge of the thicker member (angle too small) or missing the bottom of the joint (angle too large), especially if the thinner member is more than about 9.5 mm ($\frac{3}{8}$ in.) thick.

In Fig. 9(d), the joint is rabbeted to make it self-aligning, but the offset is small to avoid leaving an unwelded seam near the root of the weld (compare Fig. 9e). In Fig. 9(d) to (g), the joints are self-aligning and may be self-fixturing in circular, circumferential, and certain other joint arrangements. In Fig. 9(e) and 9(f), the joints are both self-aligning and self-backing; each, however, leaves an unwelded seam near the root of the weld.

In Fig. 9(g) and 9(h), two ways are shown of providing integral and filler metal. The lip of the joint in Fig. 9(g) provides more filler metal than the shoulder of the joint in Fig. 9(h), but because it conceals the joint, a scribed line or other means for beam placement and for scanning must be provided before the welding operation can proceed.

The slant-groove weld (also called angular or scarf) shown in Fig. 9(j) is used in butt joints to facilitate fixturing. It is also used where limitations on the beam location prevent fusion for the entire groove depth if the basic square-groove weld (Fig. 9a) were used. Greater weld depth, as measured parallel to the beam axis, is needed for full penetration of a given metal thickness than with the square-groove weld. This type of weld and variations of it are used both in butt joints and in corner joints.

A slant-butt joint can also be welded with beam alignment set at 90° to the surface of the work. This beam angle permits the production of defect-free welds where fit-up is poor or where the joint opening is larger than can be tolerated when the beam is aligned with the groove. In a related type of weld, the joint has a square-groove preparation, and the beam is at an angle to the groove. The main reason for using this type of weld is difficulty of access with other types of welds.

Welds in Corner Joints

Eight types of electron-beam welds frequently made in corner joints are shown in Fig. 10. Corner joints are second only to butt joints in frequency of use for EBW.

Important differences between butt and corner joints are notch sensitivity and suitability for nondestructive testing. To compare alignment, self-fixturing, and the occurrence of portions of unwelded seam, see Fig. 9 and 10.

The two simplest and most economical welds in corner joints are the melt-through (Fig. 10a) and basic square-groove (Fig. 10b) welds. Neither is self-aligning or self-fixturing; manipulation of the work for weld (b) in Fig. 10 can be simplified by using a horizontal electron beam and corresponding work orientation (rotated 90° from the orientation shown). The melt-through weld (Fig. 10a), unless made with a fusion zone wide enough to eliminate the unwelded seam completely, is weaker than the square-groove weld and is notch sensitive.

The corner-flange weld (Fig. 10h), usually made only on thin stock, requires precision forming of the 90° bend. Welds in corner joints are subject to high stress concentrations.

Welds in T-Joints

Three types of electron-beam welds made in T-joints are shown in Fig. 11. The melt-through or blind weld (Fig. 11a), like melt-through welds in butt or corner joints, leaves an unwelded seam with resulting low strength, notch sensitivity, and corrosion susceptibility. Figure 11(b) is the preferred joint design. The double-square-groove weld (Fig. 11c) is used primarily on sections 25 mm (1 in.) or more thick. For welds similar to Fig. 11(b) and 11(c), filler metal can be added.

Welds in Lap Joints

Three types of electron-beam welds made in edge joints are shown in Fig. 13. Thick sections can be joined by deep, narrow square-groove welds (Fig. 13a). Shallow welds made with a low-power, partially defocused beam are used to join thin sections to each other or to thick sections, as shown in Fig. 13(b) and 13(c). Shallow edge welds can be made at high speed. Welds of this type are particularly useful on hermetically sealed assemblies, which may be designed for planned salvage by removal of the weld region and subsequent rewelding.

Butt Joints Versus Corner and T-Joints

The concentration of bending stresses at the weld in corner and T-joints is sometimes avoided by designing the work for a butt joint on a straight section near the corner or T (flange neck or welding neck construction). This principle of design is applicable also to butt joints between members having different thicknesses, and it can be applied to straight line, circular, and circumferential joints.

In welding such a joint in magnetic metals, if the beam path is too close to the corner, T, or shoulder, the beam may be deflected away from the metal that it passes near on its way to the joint, causing it to enter the work at an angle and possibly to miss the lower portion of the joint. To avoid this, the beam clearance should be 0.8 mm (0.030 in.) or more, depending on the height of the magnetic projection (Fig. 14).

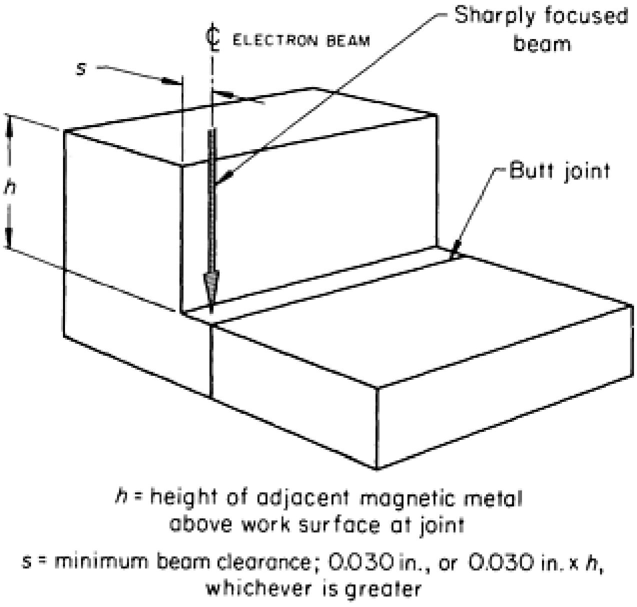

h = height of adjacent magnetic metal above work surface at joint

s = minimum beam clearance; 0.030 in., or 0.030 in. × h, whichever is greater

FIG. 14 JOINT LOCATION AND BEAM CLEARANCE FOR AN EBW BUTT JOINT NEAR A PROJECTING CORNER, T, OR SHOULDER CONSISTING OF MAGNETIC WORK METAL

Special Joints and Welds

Many variations of the basic joint and weld types described in the preceding section and illustrated in Fig. 9, 10, 11, 12, and 13 are used to meet the needs of individual applications. Some of these special joints and welds are discussed below.

Plug and puddle welds are usually made by manually (or automatically) manipulating the work under a fixed beam at low power. Filler metal used for plug welds is often preplaced at the weld site. Puddle welding is used mainly to fuse shallow defects together locally.

Multiple-Pass Welds. Most electron-beam welds produce the desired penetration in a single pass. Tack welds and cosmetic or smoothing passes are not considered penetration passes. Welds several inches deep can be made in most metals in a single pass. Weld depth obtainable in a single pass can be nearly doubled by welding from both the face and the back of the work.

In a variation on straight-through two-pass welding from opposite sides, separate welds that meet or almost meet can be made at an angle of 90° to each other in a rabbeted square-groove joint.

Tangent-tube welds are longitudinal welds joining two parallel tangent tubes (or cylinders). The tubes may differ in size; the beam is perpendicular, or nearly perpendicular, to the common axial plane of the two tubes. Added filler metal may be used for reinforcement. Thinner wall tubes can be joined more easily by EBW than by arc welding methods.

Three-Piece Welds. Welds can be made that join three or more pieces in which there is penetration in a single pass into all of the pieces. Many of the difficulties encountered in welding very thin metal or foil are eliminated by sandwiching it between two thicker sections.

Multiple-tier welds are welds made simultaneously in in-line, separated joints (usually butt joints) in a single pass of the electron beam. A more detailed discussion is given in the section of this article on multiple-tier welding .

Welds using integral filler metal may be of the types shown in Fig. 9(g) and 9(h), in which an overhanging lip or a shoulder provides filler metal to a butt joint, or both members may be made thicker at the joint than elsewhere.

EBW in High Vacuum

Most EBW is done in a vacuum environment where the maximum ambient pressure is less than 0.13 Pa (1 × 10^{-3} torr). Maintenance of this degree of vacuum is important because of the effect that ambient pressure has on both the beam and the weld produced, as discussed below.

Applications. High-precision applications that require welding to be done in a high-purity environment to avoid contamination by oxygen or nitrogen are ideally suited for high-vacuum EBW (EBW-HV). These applications include the nuclear, aircraft, missile, and electronics industries. Typical products include nuclear fuel elements, pressure vessels for rocket propulsion systems, special alloy jet engine components, and hermetically sealed vacuum devices. Welding of reactive metals such as zirconium and titanium in high vacuum is preferred over medium- or nonvacuum EBW because of the affinity of reactive metals to oxygen and nitrogen.

Effect of Pressure on Beam. Under ambient high-vacuum conditions, the frequency of collisions between beam electrons and residual gas molecules is extremely low, and any dispersion (i.e., beam-broadening effect) that would result from such scattering collisions is minimal. However, because the frequency of scattering collisions increases with the density of gas molecules present, this effect becomes greater as the surrounding environment pressure is increased. Only at ambient pressure values below 0.13 Pa (1 × 10^{-3} torr) is the effect of scattering insignificant enough that the beam can be held in sharp focus over distances of several feet (depending on the particular characteristics of the electron gun and electron optics being employed) to achieve maximum effectiveness in producing relatively deep, narrow welds.

Beam scattering from collisions with residual gas molecules begins to occur at ambient pressures above 0.13 Pa (1 × 10^{-3} torr) and initially results in decreased penetration and increased weld width. This scattering effect reduces the maximum working distance that can be employed for producing a similar electron-beam weld as obtained under high-vacuum conditions at ambient pressures greater than 0.13 Pa (1 × 10^{-3} torr).

In EBW-NV, where the ambient pressure is 10 MPa (760 torr) and above, the scattering dispersion of the beam increases so that penetration and working distance capacity are appreciably reduced below values obtainable in either high- or medium-vacuum processes. However, the penetration and working distance capabilities of EBW-NV are appreciably greater than those of conventional arc welding processes. Electron-beam dispersion characteristics and their effect on weld penetration at various ambient pressures are shown in Fig. 15.

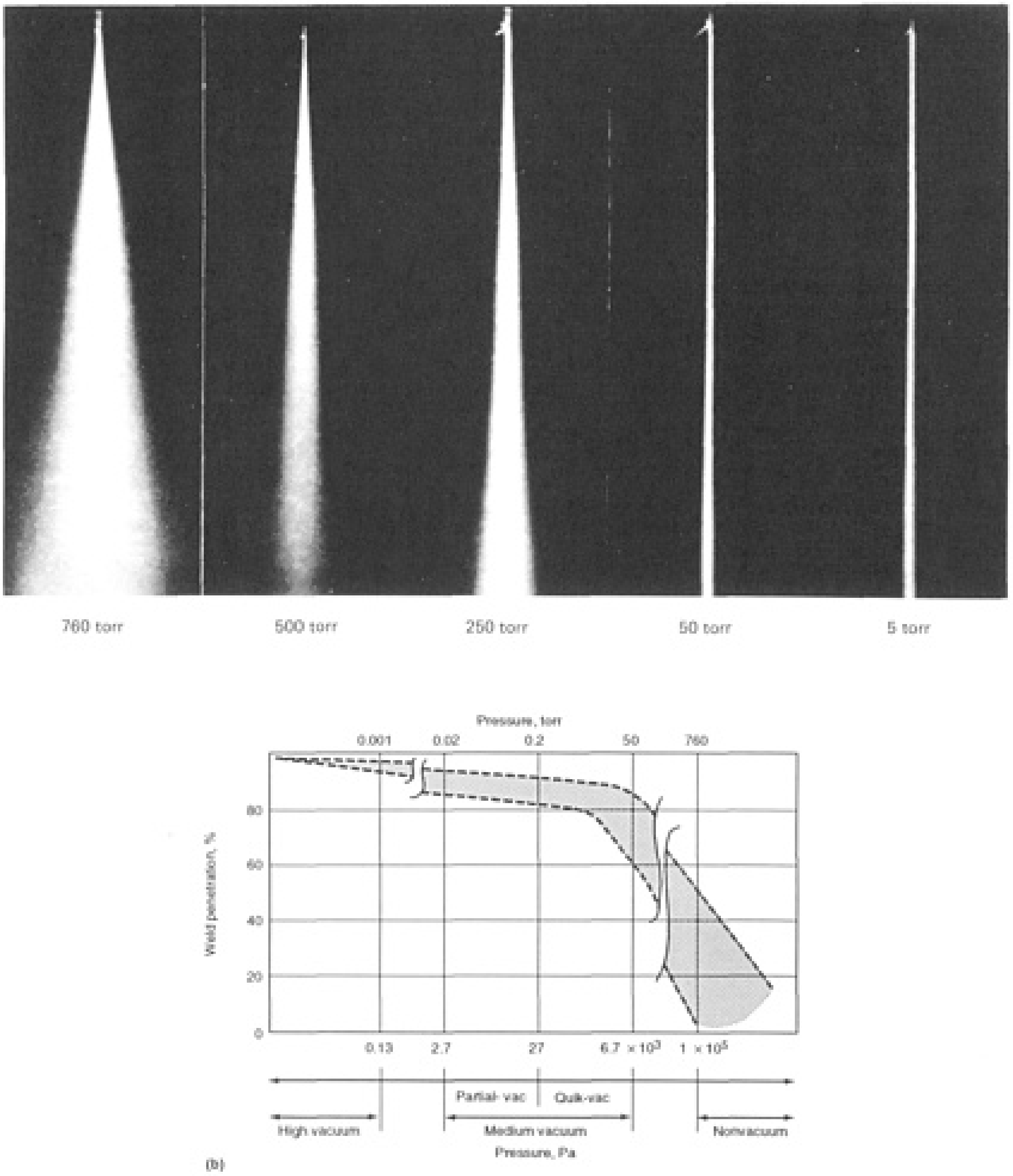

FIG. 15 EFFECT OF PRESSURE ON BEAM INTENSITY AND RESULTING WELD QUALITY. (A) DISPERSION PATTERNS OF ELECTRON BEAM AT SELECTED PRESSURES. (B) WELD PENETRATION AS A FUNCTION OF AMBIENT PRESSURE LEVEL. ADDITIONAL PARAMETERS SUCH AS BEAM VOLTAGE, BEAM TRAVEL DISTANCE, AND TYPE OF AMBIENT GAS USED CAN PRODUCE A WIDE RANGE OF VALUES, THUS ACCOUNTING FOR THE SPREAD IN THE PENETRATION DATA.

Effect of Pressure on Weld and Heat-Affected Zone. The maintenance of a high vacuum in the work chamber protects the weld metal and the HAZ from oxidation and contamination by harmful gases, serving the same function as inert shielding gases do in arc welding, while also degassing the weld. The contamination level as a function of pressure is discussed in the section EBW in Medium Vacuum in this article.

The nearly complete absence of gaseous impurities eliminates the serious contamination difficulties that usually are encountered when attempting to arc weld relative metals such as titanium and zirconium.

Width of Weld and Heat-Affected Zone. Electron-beam welds made in a high vacuum are narrower and have a narrower HAZ than comparable welds made in medium vacuum or at atmospheric pressure, and they are much narrower than the narrowest welds made in production welding by GTAW. The narrow width of the heat-affected zone in hardenable steels and other hardenable alloys permits welding of these metals after heat treatment, in many instances without loss of strength.

Limitations. A major limitation in the use of high vacuum in the work chamber is the effect on unit production time, because of the need to pump the chamber down before each load is welded. Pumpdown time is typically about 3 min for a 0.85 m^3 (30 ft^3) chamber and 10 min for an 8.5 m^3 (300 ft^3) chamber. The above pumpdown times are realistic only for a very clean and well-maintained system. In production, pumpdown may be nearly double the above times.

The effect of this limitation on unit production time is reduced by welding a number of assemblies in each load and by keeping chamber size as small as possible. Small chambers specially designed for use with small workpieces have typical pumpdown times of I min (3.69×10^6 mm^3, or 225 in.^{3}, chamber) and 15 s or less (1.0×10^6 mm^3, or 60 in.^{3}, chamber).

A second major limitation in the use of high vacuum is that work size is limited by chamber dimensions. This limitation is sometimes circumvented by the use of a chamber having special openings and seals that permit oversize work to extend outside the chamber, or by the use of a portable clamp-on chamber.

Welding Conditions. Welding in high vacuum is done with low-voltage, as well as high-voltage, equipment. Beam voltage ranges from 15 to 185 kV; beam current, 2 to 1000 mA; and welding speed, 100 to 5000 mm/min (4 to 200 in./min) for most applications. However, higher welding speeds are used for special thin section applications. The energy input to the work needed to produce a weld of the required depth of penetration and width is the basis for selecting the welding conditions.

Depth of penetration is increased by increasing the voltage or current for greater beam power, or by decreasing the welding speed. Table 1 lists approximate energy inputs per inch of weld length for making narrow single-pass electron-beam welds 6.4 to 75 mm (0.25 to 3 in.) deep in the weldable alloys of copper, iron, nickel, aluminum, and magnesium. These values are intended to serve as guidelines for establishing conditions for welding work for which no previous experience is available. Energy-input requirements for specific applications depend on the alloy composition and special operating conditions such as the use of beam oscillation or a defocused beam.

TABLE 1 ENERGY INPUT AT THE WELD FOR SINGLE-PASS EBW IN HIGH VACUUM FOR VARIOUS DEPTHS OF PENETRATION

DEPTH OF PENETRATION		ENERGY INPUT PER WELD LENGTH FOR INDICATED METAL(A)									
		Cu		**Fe**		**Ni**		**Al**		Mg	
mm	**in.**	**kj/mm**	**kj/in.**	**kj/mm**	**kj/in.**	**kj/mm**	**kj/in.**	**kj/m**	**kj/in.**	**kj/mm**	kj/in.
6.35	0.25	0.3	0.7	0.2	5	0.2	4	0.1	2	0.04	**1**
12.7	0.50	0.6	0.15	0.4	10	0.3	8	0.2	5	0.1	**3**
19.1	0.75	1.0	0.25	0.7	18	0.6	15	0.3	8	0.2	**5**
25.4	1.00	1.5	0.37	1.1	27	0.9	22	0.5	13	0.3	**8**
38.1	1.50	2.4	0.62	1.8	46	1.5	39	0.9	23	0.6	**15**
50.8	2.00	3.4	0.87	2.7	68	2.4	60	1.4	35	0.9	**22**
63.5	2.50	4.4	112	3.5	90	3.1	80	1.9	47	1.1	**29**
76.2	**3.00**	**5.4**	**137**	**4.4**	**112**	**3.9**	**100**	**2.3**	**59**	**1.4**	**36**

(A) VALUES ARE APPROXIMATE AND APPLY TO THE COMMONLY WELDED ALLOYS OF THE METALS. THEY ARE INTENDED TO SERVE AS STARTING POINTS FOR ESTABLISHING CONDITIONS FOR MAKING NARROW WELDS. ENERGY INPUT MAY VARY SUBSTANTIALLY FROM THESE VALUES IN SPECIFIC APPLICATIONS, DEPENDING ON THE COMPOSITION OF THE ALLOY AND SPECIAL OPERATING CONDITIONS. ENERGY INPUT AT WELD = [(BEAM VOLTAGE) (BEAM CURRENT) 0.6]/WELDING SPEED, WHERE ENERGY INPUT IS IN KJ/IN.; BEAM VOLTAGE IS IN KV; BEAM CURRENT IS IN MA; AND WELDING SPEED IS IN IN./MIN.

EBW in Medium Vacuum

Electron-beam welding in a medium vacuum (EBW-MV) usually is performed at a welding (work) chamber pressure of about 10 Pa (7.5×10^{-2} torr), or at a work chamber pressure ranging from 0.4 to 40 Pa (3×10^{-3} to 3×10^{-1} torr)--the total partial-vacuum region. The gun chamber, where the beam is generated, is held at 1.3 to 13 mPa (10^{-5} to 10^{-4} torr), as in high-vacuum welding. A diffusion pump is required to provide this gun chamber vacuum, but only a mechanical-type pump is required to evacuate the work chamber to the final partial vacuum welding pressure desired.

Comparison with High-Vacuum EBW. The chief advantage of welding in medium vacuum, compared with welding in high vacuum, is the short pumpdown time for the welding chamber, which ordinarily does not exceed 40 s for a general-purpose chamber (of about 0.11 m^3, or 4 ft^3) and which is less than 5 s for the specially designed chambers used in welding small parts. Accordingly, EBW-MV permits mass production of parts, using a chamber of minimum volume.

Production rates for welding in medium vacuum depend on part design and other factors; maximum production rate is typically about 60 pieces per hour for general-purpose, manually operated equipment; upwards of 600 pieces per hour for specially tooled semiautomatic machines; and 1500 pieces per hour for fully automatic single-purpose machines. General-purpose chambers and tooling are used for short runs.

Welds on reactive metals and other welds for which the effects of contamination by gases are especially critical are preferably made in high vacuum. The contamination level of air (total concentration of gases present) is proportional to pressure, as follows:

PRESSURE		GASES, PPM
PA	**TORR**	
0.0013	10^{-5}	**0.01**
0.13	10^{-3}	**1.3**
13	10^{-1}	**132**
50	**4×10^{-1}**	**500**

Lower contamination levels are obtainable than those actually observed for arc welding with inert-gas shielding.

Penetration and Weld Shape. Penetration at a given power level is ordinarily 5 to 10% less than for EBW-HV, and welds are wider and slightly more tapered. These changes are caused by the scattering of the electrons in the beam by collision with gas molecules, which makes the beam broader.

The effect of welding chamber pressure on penetration and weld shape is shown in Fig. 16 for welds made on type 304 stainless steel, using a beam length of 405 mm (16 in.) without changing focus. Maximum penetration shown in Fig. 16 is approximately 13 mm ($\frac{1}{2}$ in.). Welding conditions were 150 kV, 30 mA, and 1525 mm/min (60 in./min). Penetration drops off rapidly at pressures of 13 Pa (10^{-1} torr) or more at this beam length and at somewhat lower pressures when the beam path is longer.

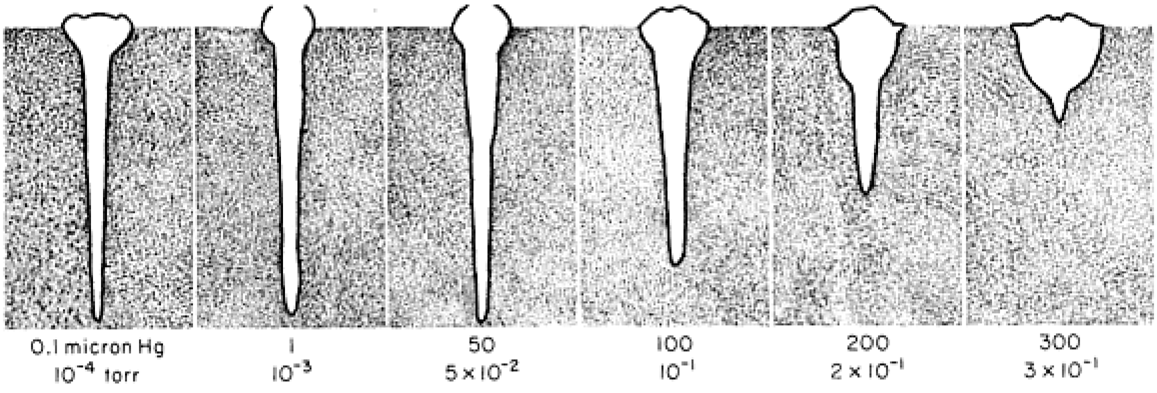

FIG. 16 EFFECT OF WELDING-CHAMBER PRESSURE ON PENETRATION AND WELD SHAPE FOR TYPE 304 AUSTENITIC STAINLESS STEEL

Comparison With Nonvacuum EBW Welding. Disadvantages of medium-vacuum welding, compared with welding at atmospheric pressure, include limitations on work size imposed by the need to enclose the work in a vacuum chamber and the time needed for chamber pumpdown.

Beam deflection and oscillation, which cannot be used in welding at atmospheric pressure, can be used in EBW-MV.

Applications. Most parts that can be welded in high vacuum can also be welded in medium vacuum. Depth of penetration is less than in high-vacuum welding, and depth-to-width ratio is less than about 10 to 1. Reactive metals are less readily welded in medium vacuum than in high vacuum. Medium-vacuum welding can be acceptable, however, for refractory metal welding when absorption of small amounts of oxygen can be tolerated. Medium-vacuum welding has been used for welding mass-produced automobile parts, such as gears and shafts, that are welded to close tolerances and that are not subsequently finished.

Procedure Development and Practice Considerations for Electron-Beam Welding

Nonvacuum (Workpiece Out-of-Vacuum) Welding

In EBW-NV (workpiece out-of-vacuum), which is also referred to as atmospheric EBW, the work to be welded is not enclosed in a vacuum chamber. Instead, it is welded at atmospheric pressure. A radiation-tight enclosure or chamber, similar to that used in high- or medium-vacuum EBW, surrounds the weld area. A vacuum, however, is not provided in this welding enclosure.

A nonvacuum electron beam is generated at high vacuum in the same manner as in high- and medium-vacuum welding. It is focused down through a series of individually pumped stages, each connected to the other by concentrically aligned orifices, decreasing in diameter in the direction of increasing pressure; this differential pumping scheme provides the high vacuum (gun region) to atmospheric (workpiece region) pressure gradient necessary to allow the beam to pass down through the column and exit out into the ambient atmosphere.

Figure 17 depicts a typical EBW-NV gun/column assembly. This unit can be mounted in either a vertical or a horizontal position and is capable of motion during welding. The gun/column assembly depicted in Fig. 17 is capable of beam power levels greater than 35 kW and beam transmission efficiencies in excess of 90%.

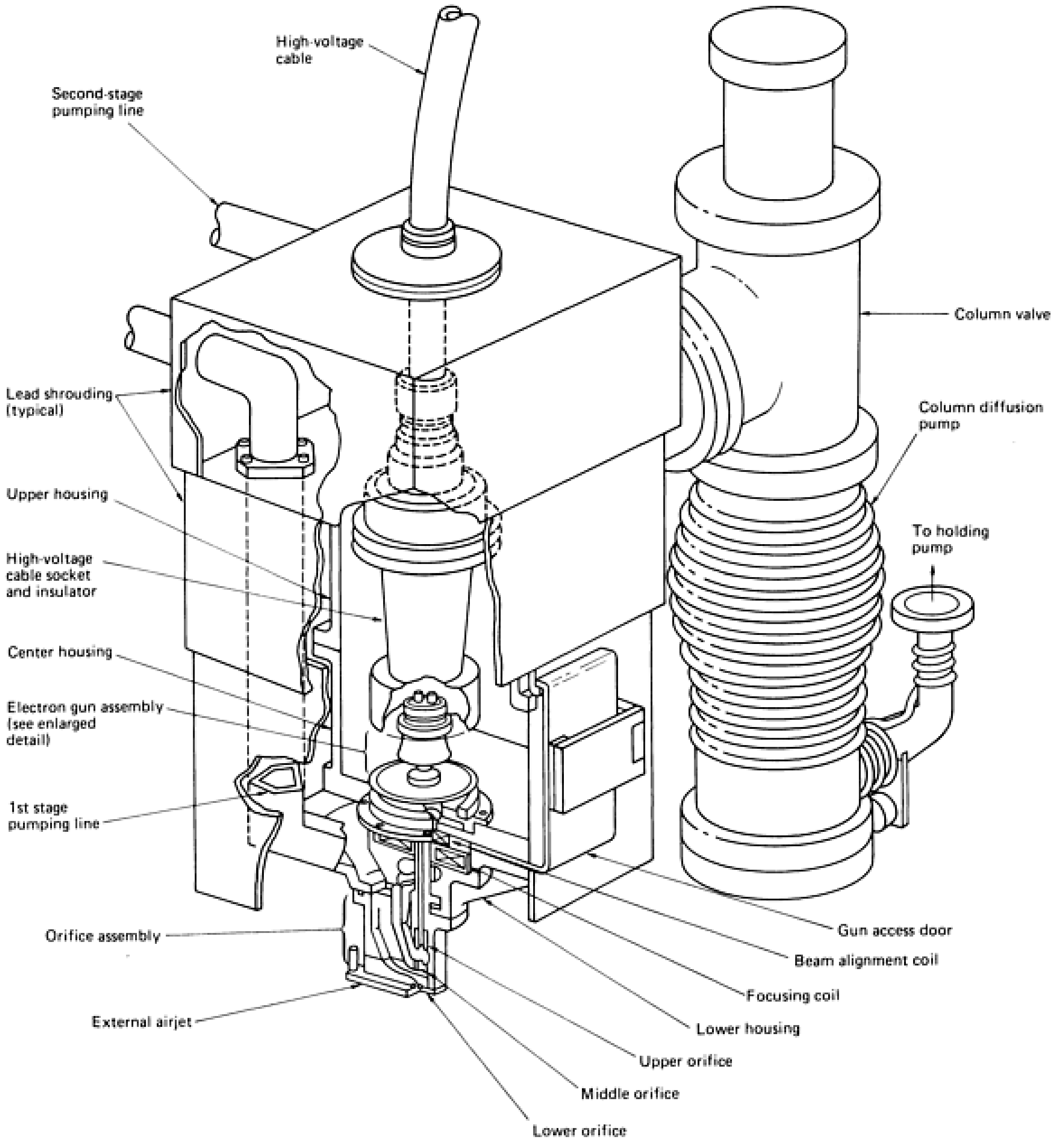
High-voltage cable
Second-stage pumping line
Column valve
Lead shrouding (typical)
Column diffusion pump
Upper housing
To holding pump
High-voltage cable socket and insulator
Center housing
Electron gun assembly (see enlarged detail)
1st stage pumping line
Gun access door
Orifice assembly
Beam alignment coil
Focusing coil
External airjet
Lower housing
Upper orifice
Middle orifice
Lower orifice

FIG. 17 SCHEMATIC WITH CUTAWAY VIEW SHOWING PRIMARY COMPONENTS OF A TYPICAL EBW-NV GUN/COLUMN APPARATUS. COURTESY OF LEYBOLD-HERAEUS VACUUM SYSTEMS INC.

Production rates are generally much higher for EBW-NV than for EBW-HV and EBW-MV, because no work chamber vacuum pumpdown time is required for each chamber load. This results in a lower per part production cost. Because pumpdown time for high- and medium-vacuum welding varies with chamber cleanliness and ambient humidity, elimination of work chamber pumpdown time in EBW-NV produces appreciable time savings. This saving, however, is gained at the expense of a reduction in penetration and working distance capability, thus imposing limits on the thickness and shape of the workpiece. Additionally, most (but not all) out-of-vacuum welds cannot be made as narrow and parallel sided as welds made in a vacuum.

When welding in ambient atmospheric air alone, the practical working distance of the nonvacuum process is normally limited to the range of from 9.5 to 22 mm ($\frac{3}{8}$ to $\frac{7}{8}$ in.); however, when welding in the presence of a lighter ambient gas, such as helium, this practical working distance range is extended to 25 to 50 mm (1 to 2 in.). For this reason, most commercial nonvacuum EBW units are equipped to provide a "helium blowdown," which is an effluent of helium that extends out from the nonvacuum column and along the beam travel path. The effect of this helium effluent is that it reduces the degree of beam scattering incurred using only air, thus allowing the working distance capability of the nonvacuum unit to be approximately doubled.

Most metals that can be arc welded or electron-beam welded in a vacuum also can be welded in nonvacuum. As stated previously, inert-gas shielding of the weld zone can be provided when the molten weld metal (or HAZ) cannot be exposed to air. On readily welded metals, weld properties approach those of welds made in a vacuum; on difficult-to-weld metals, weld properties are usually lower than those of welds made in a vacuum.

Ordinarily, filler metal is not used in EBW-NV. The addition of a suitable filler metal may be helpful, however, in joining metallurgically incompatible metals or in avoiding weld cracking caused by hot shortness. Selection of the filler-metal composition in such applications is critical.

Operating conditions for nonvacuum welding differ substantially from those for welding in high or medium vacuum. Because beam dispersion increases in proportion to pressure and distance traveled, at pressures above the high-vacuum level, the nonvacuum beam becomes dispersed at travel distances greater than several inches in air at atmospheric pressure.

To provide an electron beam of sufficient energy to minimize the scattering effect of collisions with residual gas molecules within the electron-beam transfer column (i.e., the differentially pumped gun/column assembly), the applied gun voltage is held at a constant high value (from 150 to 200 kV). This also helps extend the work distance of the beam after it has exited the column and entered the ambient atmosphere. The practical working distance (standoff distance) for nonvacuum welding, as measured from the bottom of the final exit orifice to the top of the workpiece, ranges from 13 to 50 mm ($\frac{1}{2}$ to 2 in.), although some nonvacuum welding is performed at both shorter and longer standoff distances.

With voltage held constant, beam current, working distance, and welding speed are selected to provide the required penetration and weld shape for a sound weld. Because it is usually desirable to keep working distance as short as possible and welding speed high, beam current is the primary control variable in nonvacuum welding.

The small size of the exit orifice on the electron gun makes it necessary to focus the beam at or very close to the exit orifice. Accordingly, it is not possible to vary weld characteristics to a significant degree by changing focus. Similarly, beam deflection and oscillation are not possible. The absence of these adjustment capabilities in nonvacuum welding is not a handicap, because beam width is ordinarily great enough so that it does not promote undercutting or porosity at the root of the electron-beam weld.

A cosmetic pass is seldom needed on electron-beam welds made at atmospheric pressure, because a smooth surface is produced on welds in most metals. The smoothness of the crown also improves the ease of nondestructive inspection of the as-welded surface; wire brushing is often adequate preparation for radiographic or liquid-penetrant inspection.

The limitation on workpiece shape imposed by the need for a short working distance can be overcome to some degree by the use of electron gun nozzles specially designed to extend into restricted spaces. Filler metal is not ordinarily used except where necessary for weld reinforcement to produce the desired weld properties or to avoid cracking.

A stream of dry filtered air, or an inert shielding gas such as argon or helium, is passed across the weld region in the space between the work and the electron gun, or may be supplied through a special insert nozzle assembly that is part of the exit orifice and is designed to minimize entrance of welding vapors and other contaminants into the gun. Auxiliary inert-gas shielding is supplied where needed for complete protection of the molten weld metal and the HAZ from gaseous contaminants. Welds are sometimes made on carbon and alloy steel and other readily weldable metals without using shielding gas. However, shielding gas is often used; helium is the preferred gas. Weld shape can be changed by varying the flow of helium.

Penetration. Nonvacuum electron-beam welds are seldom deeper than 9.5 mm ($\frac{3}{8}$ in.). Welds having penetration much greater than 25 mm (1 in.) are easily produced, but ordinarily this requires a reduction in welding speed, with a corresponding increase in production cost. Figure 18 provides a comparison of penetration, as a function of weld speed, in steel and aluminum, employing a fixed power level and work distance and using a helium shield to improve power density.

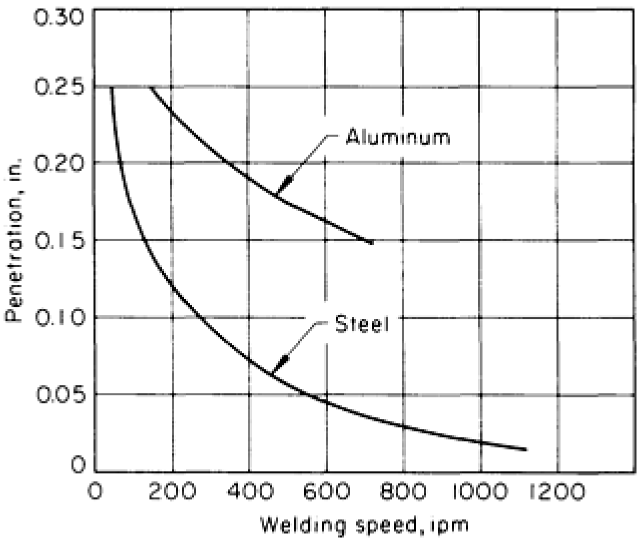

FIG. 18 PLOT OF PENETRATION VERSUS WELDING SPEED AS A FUNCTION OF MATERIAL AT A BEAM POWER OF 9 KW

Studies made of nonvacuum welding of steel pipe showed that, at speeds of 2.5 to 25 m/min (100 to 1000 in./min), full-penetration welds could be made in metal twice as thick if helium shielding gas was used instead of air. Table 2 gives the results of this comparison in terms of welding speed for joining steel pipe 1.3 to 3.8 mm (0.050 to 0.150 in.) thick in air or in helium. Welding speeds were 2.16 to 2.24 times greater in helium as in air. Beam power used was 12 kW.

TABLE 2 SPEEDS FOR NONVACUUM FULL-PENETRATION WELDING OF CARBON AND LOW-ALLOY STEEL PIPE AT A BEAM POWER OF 12 KW

WALL THICKNESS		**WELDING SPEEDS**				RELATIVE WELDING SPEED IN HELIUM(A)
		IN AIR		**IN HELIUM**		
mm	**in.**	**m/min**	**in./min**	**m/min**	**in./min**	
1.27	0.050	10.6	417	22.9	900	**2.16**
2.03	0.80	6.12	241	13.4	528	**2.20**
2.54	0.100	4.75	187	10.5	412	**2.21**
3.81	**0.150**	**3.00**	**118**	**6.68**	**264**	**2.24**

(A) BASED ON WELDING SPEED IN AIR BEING EQUIVALENT TO 1.00

Weld Shape and Heat Input. Nonvacuum electron-beam welds are generally wider and more tapered than high-vacuum welds (medium-vacuum welds differ only slightly in shape from high-vacuum welds).

Weld shrinkage across nonvacuum electron-beam welds is generally less than half that of gas-tungsten arc welds on the same work metal and thickness and is typically about twice that for high-vacuum electron-beam welds.

Tooling generally is designed specifically for welding a particular assembly in mass production. Providing fixtures and equipment that permit welding at high speeds and efficient work handling is the key to obtaining high production rates and low cost.

The work-handling equipment is at atmospheric pressure instead of in a vacuum; hence, general-purpose welding positioners and related fixtures usually are satisfactory. Hold-down clamping devices can be simpler than those often needed for arc welding similar parts, because angular distortion or dihedral warpage is generally low. Locating and alignment mechanisms are simpler than for welding in a vacuum, because of their accessibility and because beams are wider and lower accuracy is required in tracking the joint. Maintenance of work-handling equipment is greatly simplified by its out-of-chamber location.

Applications. Nonvacuum electron-beam welding is used in the production of transmission-train components, steering column jackets, die-cast aluminum intake manifolds, and catalytic converters. In addition, several appliance (and other consumer-related) items also are produced using the EBW-NV process. Figure 19 shows a catalytic converter that was nonvacuum electron-beam welded with the beam in a horizontal position.

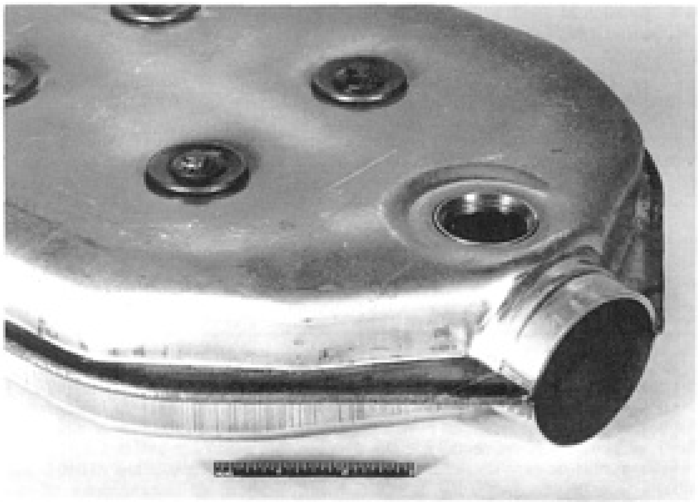

FIG. 19 CATALYTIC CONVERTER ASSEMBLY CONSISTING OF SEAMS JOINED BY EBW-NV PROCESS

Controlling Heat Effects

The extent of heat-affected base metal in EBW is controlled mainly by regulating the welding speed, width, and shape of the weld. In general, the width of the HAZ is proportional to the average width of the weld, which also strongly influences shrinkage and distortion.

Damage to heat-sensitive attachments, inserts, or encapsulated materials close to the weld joint is controlled mainly by regulating the width and shape of the weld bead. Damage to nearby finished functional surfaces on the workpieces or to nearby drilled or tapped holes or other crucial features of the workpieces can often be avoided in EBW because of the narrowness of the welds.

Joint designs that provide constricted sections near the weld to retard heat transfer from the weld, or that provide an integral heat sink, and the use of external heat sinks or chill bars are also helpful in controlling heat effects in EBW.

Narrow, closely controlled electron-beam welds, with low heat input, can be made in close proximity to heat-sensitive attachments, inserts, or encapsulated materials without damaging them, often when no other welding process can meet this requirement.

Welding of Thin Metal

Electron-beam welding can be advantageous for joints in which at least one member is made of thin metal (<2.4 mm, or $\frac{3}{32}$ in., and as thin as 0.025 mm, or 0.001 in.). Applications include instrument parts, instrument enclosures, pressure or hermetic seals, diaphragms, encapsulations, electrical connectors, and electronic devices, of various metals.

The joint area must be heated to melting temperatures in joining thin sections, just as in joining thick sections, but the rate of heat transfer away from the joint is much lower because of the reduced cross section. Hence, local heat buildup is greater, which increases weld width and decreases depth-to-width ratio. In addition, minimum beam spot size has a much greater effect on depth-to-width ratio in joining thin metal than in joining thick metal. The minimum usable beam spot size obtainable generally ranges from about 0.13 to 0.51 mm (0.005 to 0.020 in.) in diameter. Because of these conditions, depth-to-width ratio for full-penetration welds joining two sections thinner than about 2.4 mm ($\frac{3}{32}$ in.) usually does not exceed about 5 to 1 and may be less than 1 to 1.

When voltage, current, welding speed, and other variables have been optimized for a specific joint, use of a heat sink may still be necessary to avoid overheating thin material, depending on the base-metal properties (melting point, specific heat, and thermal conductivity) and metal thickness. Best results usually are obtained by the use of copper chill blocks machined to fit the workpiece closely at the joint area. Fixtures made of other metals may be effective as chill blocks if the contact area is sufficiently large and close to the heat source.

Joining Thin Sections to Thick Sections. In joining thin sections to thick sections by EBW, the joint is usually designed so that the thick member serves as a heat sink for the thin member, and the point of beam impingement is slightly removed from the thin metal. When these precautions are taken, welding behavior and heat dissipation are generally the same as in welding thick sections under similar conditions. Depth of penetration in the thicker section is usually made only slightly greater than the thickness of the thin section.

Even with the aid of the heat-sink capacity of a thicker member, heat input by arc welding often cannot be localized sufficiently and controlled closely enough to avoid damage to metal foil thickness. In some applications, using EBW instead of gas-metal arc welding can eliminate excessive foil melting.

Partial-Penetration Welds. The precise control of heat input obtainable in EBW permits partial-penetration welds having uniform depth to be made in metals as thin or thinner than 0.25 mm (0.010 in.).

Work-metal thickness must be uniform, and workpiece fixture arrangement and dimensions must be selected to provide a constant rate of heat loss adjacent to the path of the weld along its entire length, with special provisions made for weld starting and stopping. Beam power, focus, and workpiece travel speed must be closely regulated to minimize variation in penetration.

For the closest control of penetration, both beam power and focus can be regulated by sensitive, fast-response closed-loop servomechanisms or feedback controllers, and workpiece travel speed can be held within extremely narrow limits by precision, mechanical worktable traverse, or rotary-motion systems. To compensate automatically for the effect of any variation in voltage on focus, the servomechanism regulating the beam focus can be coupled to the voltage supply by solid-state devices that have a response time of less than 17 μs.

Partial-penetration welds in thin metal are not subject to root voids, root porosity, or cold shuts, which are flaws frequently encountered in partial-penetration welds in thick metal, because such welds in thin metal have relatively low depth-to-width ratios and a wide-angle V-shape with a gently radiused bottom.

Data on the relation of depth to width for high-vacuum partial-penetration welds in thin sections are given in Table 3. The ratio of depth to width is lower for the welds in the thinner sections. Penetration was from 50 to 85% of the work-metal thickness, and all welds were made with chill bars in intimate contact with the underside of the work at the weld location.

TABLE 3 RELATION OF DEPTH TO WIDTH (*D*/*W*) FOR PARTIAL-PENETRATION ELECTRON-BEAM WELDS OF THIN SECTIONS IN A HIGH VACUUM

WORK METAL$^{(A)}$	**THICKNESS,**		**WELD DIMENSIONS**$^{(A)}$				*D*/*W* FOR WELD
			DEPTH, *D*		**WIDTH, *W***$^{(B)}$		
	mm	**in.**	**mm**	**in.**	**mm**	**in.**	
ALUMINUM ALLOYS							
6061-T6	0.13	0.005	0.008	0.003	0.38	0.015	**0.2**
5052	0.25	0.010	0.013	0.005	0.30	0.012	**0.4**

	0.51	0.020	0.43	0.017	0.38	0.015	**1.1**
STAINLESS STEELS							
17-4 PH	0.13	0.005	0.10	0.004	0.36	0.014	**0.3**
TYPE 301	**0.25**	**0.010**	**0.15**	**0.006**	**0.28**	**0.011**	**0.5**

(A) CHILL TOOLING WAS IN INTIMATE CONTACT WITH THE UNDERSIDE OF THE WORK AT THE WELD LOCATION.

(B) WIDTH OF THE WELD AT THE CROWN, WHICH WAS THE GREATEST WIDTH FOR THE GENTLY RADIUSED WIDE-ANGLE V-SHAPED WELDS

Depth of penetration generally varies 10 to 15% from a mean value for a work thickness of about 0.51 mm (0.020 in.), and the percentage variation is greater for welds in thinner metal. Accordingly, partial-penetration welds are not ordinarily made in metal about 0.25 mm (0.010 in.). Beam oscillation is sometimes used to smooth out variations in penetration, but such oscillation must be controlled closely in order to be effective.

Procedure Development and Practice Considerations for Electron-Beam Welding

EBW of Thick Metal

For joining thick sections, electron-beam welding in a vacuum has three principal advantages over other welding processes:

- MUCH DEEPER PENETRATION CAN BE OBTAINED IN A SINGLE PASS. EXTREMELY NARROW, ONLY SLIGHTLY TAPERED WELDS AND A SMALL HAZ CAN BE PRODUCED.
- FOR JOINTS ON MOST WELDABLE METALS, NO FILLER METAL IS NEEDED. WHERE A FILLER METAL IS REQUIRED, THE QUANTITY IS USUALLY VERY SMALL.
- JOINTS HAVE CLOSELY FITTED PARALLEL GROOVE FACES REQUIRING NO EDGE PREPARATION, INSTEAD OF V-GROOVES OR U-GROOVES.

The disadvantages of EBW for joining thick metal include: high equipment cost, size limitations imposed by the dimensions of the vacuum chamber in which the work must be placed, and the time needed for evacuating the chamber.

Deep welds made in a vacuum are usually narrow; weld width for penetration deeper than 6.4 mm ($\frac{1}{4}$ in.) is typically about $\frac{1}{5}$ to $\frac{1}{20}$ of the section thickness, except for a somewhat greater width at the crown of the weld. Welds may be either full or partial penetration.

Effects of Pressure. Penetration capability is greatest for EBW in high vacuum. The effect that welding pressure has on penetration of steel is shown in Table 4. A constant weld speed of 890 mm/min (35 in./min) and beam power of 7.5 kW from 150 kV/50 mA for high and medium vacuum, and 175 kV/43 mA for nonvacuum were used to obtain the data in Table 4.

TABLE 4 TYPICAL PRESSURE AND BEAM TRAVEL DISTANCE SPECIFICATIONS REQUIRED TO OBTAIN MAXIMUM WELD PENETRATION OF STEEL

TYPE OF SYSTEM	**AMBIENT WORKPIECE PRESSURE**		**BEAM TRAVEL DISTANCE**		MAXIMUM PENETRATION	
	pa	**torr**	**mm**	**in.**	**mm**	in.
HIGH VACUUM	0.013	1×10^{-4}	455	18	25	**1**
MEDIUM VACUUM	130	1	205	8	16	$\frac{5}{8}$

NONVACUUM	**1.0×10^5**	**760**	**13**	$\frac{1}{2}$	**4.0**	$\frac{5}{32}$

Full-Penetration Welds. In welding carbon steel, single-pass, full-penetration, vacuum (high/medium) electron-beam welds are made commercially in thicknesses up to approximately 100 mm (4 in.); production EBW-HV has been done on aluminum alloy plates 150 mm (6 in.) thick, and 230 mm (9 in.) thick aluminum alloy sections have been welded experimentally under near-optimum conditions.

Two-Pass Full-Penetration Welding. Penetration can be nearly doubled by making two passes, one on each side of the work, but cold shuts, voids, and porosity are encountered at the root of the second weld unless conditions are adjusted for the pass made on the second, or back, side. A broader second-side weld with a larger radius at the root helps to reduce the incidence and severity of these flaws, but they often remain a problem.

Partial-Penetration Welding. Welds that do not penetrate completely through the work metal are satisfactory in many applications in welding thick metal; joint design and product requirements often rule out full penetration.

Problems and Flaws. In making vacuum electron-beam welds 3.2 to 13 mm ($\frac{1}{8}$ to $\frac{1}{2}$ in.) deep in weldable metals, weld quality is ordinarily equal to or better than that of arc welds in the same metal. Special precautions are needed to avoid certain types of discontinuities in welds about 13 mm ($\frac{1}{2}$ in.) deep or deeper. Fusion-zone porosity, gas pockets, and cold shuts are more serious problems than in arc welding, particularly in deep welds.

Reducing the welding speed usually helps to reduce porosity and gas entrapment. Extreme care in cleaning the work metal may be necessary where the problem is severe. When flaws are found near the root of the weld, they can be avoided or minimized by any adjustment of welding conditions that broadens the weld and increases the radius of the weld at the root. Sometimes the joint can be designed so that any root flaws are in a noncritical region, or in integral or separate backup metal that will be machined away after welding.

As in arc welding, cold shuts (incomplete fusion) are sometimes encountered at the root of a deep partial-penetration weld or of a full-penetration weld that has a poorly fitted backing strip. These discontinuities are troublesome because they are difficult to detect. Normally they cannot be detected by radiographic methods. Ultrasonic testing can detect the larger cold shuts. Many cannot be observed on roughly polished macrosections of electron-beams welds, but require a metallographic polish and examination at a magnification of at least 100 diameters for detection. Cold shuts can usually be avoided or minimized by reducing the welding speed or by otherwise changing conditions so as to broaden the weld and increase the radius of the weld at the root.

Poor fit-up or excessive joint gap can cause excessive shrinkage, underfill, undercut, voids, and cold shuts. Joint gap should not exceed 0.25 mm (0.010 in.) for narrow welds deeper than about 13 mm ($\frac{1}{2}$ in.) in most metals, although sound welds have been obtained using joint gaps of 0.76 mm (0.030 in.), by use of a procedure that increased weld width.

Welding with the beam at a slight angle to the joint is helpful in avoiding flaws related to poor fit-up, but may increase weld width excessively and require an unduly low welding speed for deep welds.

Excessive mismatch can cause discontinuities of the same general types as poor fit-up; mismatch limits depend on joint design and dimensions, and on operating conditions. Weld quality in joining thick sections is ordinarily unaffected by changes in surface roughness of the joint faces between about 1.60 to 2.50 μm (63 and 1000 μin.).

A problem that is unpredictable and inconsistent in making electron-beam welds deeper than about 13 mm ($\frac{1}{2}$ in.) is arc-outs, or sudden failures of electron emission during welding. However, problems of arc-out have decreased with the newer machines and experience accumulated in operation.

Work-metal composition and quality also influence arc-outs, which are more frequent when welding materials with low vaporization points, or when welding materials susceptible to outgassing in a vacuum or those that contain nonmetallic inclusions.

Use of Filler Metal

Fusion of closely fitted groove faces generally provides sufficient weld metal, or extra metal is provided where necessary by including extra stock thickness or integral shoulders or lips in the joint preparation. Where product requirements or other circumstances prevent the use of a joint preparation and welding conditions that will provide sufficient weld metal, filler metal is added.

Prevention of Cracking. Filler-metal (shim stock) additions can prevent cracking in electron-beam welds in a crack-susceptible metal or combination of dissimilar metals. It may be needed even when joint design, fit-up, and operating conditions are selected to achieve minimum joint restraint and residual stress. In other applications, the use of preplaced filler metal can provide weld metal that is less brittle than would be obtained by welding a base metal (or metals) alone.

Among the base metals for which added filler metal is often required are heat-treatable aluminum alloys 6061, 6063, and 6066, free-machining steels, and other free-machining alloys.

Many combinations of dissimilar metals that are susceptible to cracking when welded directly, usually because of the formation of brittle intermetallic phases, can be electron-beam welded with the aid of filler metal that has a composition compatible with both metals of the combination.

Preventing Porosity. The EBW of rimmed steel without filler metal ordinarily results in severe porosity in the weld, even when welding speed is slow and other conditions are selected to increase the time during which gas can escape from the molten weld metal. Inserting a filler metal that contains a deoxidizer such as aluminum, manganese, or silicon helps to minimize porosity.

Copper alloys C11000 (tough pitch copper) and other types of copper that do not contain residual deoxidizers also produce porous welds when joined without filler metal, but sound welds can be made with the aid of nickel filler metal or a deoxidizing filler metal.

Preplacement of Filler Metal. The technique most frequently used for the addition of filler metal in EBW, especially for deep welds, is preplacement. A shim of filler metal, usually of foil thickness, can be inserted between the groove faces when the joint is assembled for welding, or a wire or other suitable shape can be preplaced over the joint, being held in position by tack welding, if necessary.

Filler-Wire Feeding. Electrically operated filler-wire feeding systems that are modifications of wire-feed equipment used in GTAW, or specially designed electron-beam wire feeders, are sometimes used in the vacuum chamber. Usually, filler-wire feeding does not help in joining incompatible metals unless they are very thin.

The design of the wire-feeding nozzle and the technique with which it is used are important in guiding the wire so that it intercepts the weld pool. Any filler metal inserted into the beam path absorbs energy and affects penetration.

The wire-feeding nozzle is normally held as close to the weld pool as is possible without damage to the nozzle, and the wire is directed into the forward edge of the pool. The tip of the nozzle should be made of a heat-resistant material and should be coated so as to prevent molten weld metal from adhering to it if inadvertent contact is made. The nozzle is usually pointed in the direction of workpiece travel.

Wire-Feeding Equipment. Several design features are desirable in a wire feeder for EBW. These include means of making positioning adjustments from outside the vacuum chamber during welding, such as changing the proximity of the nozzle to the weld pool, changing the angle of incidence between the workpiece and nozzle, and controlling the movement of the wire relative to the beam to ensure accurate interception.

A simple, variable-resistance, controlled motor drive is ordinarily satisfactory, as a 10% variation in wire speed causes no adverse effect other than a slight variation in weld width. It is desirable to be able to control the timing so that wire feed can be started and stopped independently of the beam and table. Being able to adjust the timing of the wire feeder eliminates sticking of the wire to the workpiece at the end of the weld. Vibrating the nozzle at subsonic rates promotes flow of molten metal into the weld joint and results in less wire sticking at the end of the weld.

Welding Technique When Feeding Filler Wire. The wire-feed rate is usually set at approximately the same rate as the welding speed. Beam power must be sufficient to melt the wire as fast as it is fed. The wire diameter is selected so as to provide 1.25 times the volume of filler metal needed to fill the joint cavity. Where needed, backing strips or rings are used to prevent loss of molten weld metal.

The filler metal can be deposited during the weld pass, using a low welding speed to allow time for the molten metal to fill the joint gap, but for deep welds it is preferably deposited in a separate pass. If cracking is a problem, another method used on thin metal is to deposit filler metal on the joint, and then make the welding pass to distribute the filler metal more effectively. Additional information on the wire-feed process and related equipment can be found in the section Electron Beam Welding as a Repair Method of this article.

Tack Welding

In fixturing work for EBW, clamping often is supplemented or replaced by tack welding, which is usually done with the electron beam. Fixturing time and cost are often substantially reduced by the effective use of tack welds.

In some applications, selection of the tack welding procedure is critical to avoid unacceptable variation in the shape and dimensions of the final weld bead and to avoid root porosity or cold shuts. Reduced power and other special operating conditions are often used.

Techniques. Perhaps the most common technique is to make one or more suitably spaced tack welds at full or reduced beam power before making the welding pass. For circular or circumferential welds, one or more tack welds are usually made. Melt-through (spike) tack welds can be used on straight-line lap joints.

A shallow seal pass at reduced beam power can also be made along the full length of the joint to keep the workpieces aligned, to be followed by the full-power penetration pass. For some applications, it is convenient to leave occasional unwelded gaps as an aid for tracking the joint during the penetration pass.

EBW of Poorly Accessible Joints

One of the advantages of EBW is the capability of reaching into areas lying deep within narrow openings. This is accomplished by virtue of the electron beam having a small diameter, long working distance, and frequently, the capability of being projected at an angle. Many joints that are inaccessible for welding by other processes can be electron-beam welded. This capability is used both in fabrication and in repair work. It is especially useful in salvaging intricate castings.

Workpiece Requirements. For limited-access welding, the workpiece must satisfy three general requirements: (1) the weld area must be on a line-of-sight path from the beam source; (2) there must be sufficient sidewall clearance to avoid beam-fringe interference; and (3) the beam path must be free of magnetic fields.

Beam characteristics that determine applicability of EBW to poorly accessible joints are beam diameter and the working distance or effective beam length available between the exit end of the beam transfer column and the work. These characteristics in turn depend on the beam power used, which is selected to produce the desired penetration, and on beam focal length, which is influenced by the design of the gun, the chamber, and the focusing coil.

Sidewall Clearance. Parts intended for fabrication by EBW can often be designed with sufficient sidewall clearance for the beam power required. Where insufficient sidewall clearance exists in repair welding, lowering the beam power or slightly changing the angle of incidence of the beam often avoids interference. In repair welding, tests should be made on simulated joints, using the same type of work metal and the required beam power, rather than risking damage to difficult-to-replace parts.

Where close sidewall clearances are involved, it is especially important to avoid magnetic fields, which can cause damage to the part by deflecting the beam. A test run on low power is commonly used to detect beam deflection; however, allowance must be made for the increase in deflection at full welding power. Magnetically soft materials with induced magnetism can usually be demagnetized with magnetic-particle inspection equipment or with coils.

Minimum beam clearance for making welds close to magnetic metal that projects above the work surface at the joint is given in Fig. 14.

Procedure Development and Practice Considerations for Electron-Beam Welding

Multiple-Tier Welding

The deep-penetration properties of the electron beam make it possible to weld two or more tiers of joints simultaneously. The joints can be separated by an air space as great as several inches, provided that the space can be evacuated and the joints aligned in the path of the beam. This type of weld cannot be made by any other welding process.

In multiple-tier welding, the electron beam must pierce at least the upper tiers in such a way that the molten metal flows in behind the advancing beam, as in keyholing. The molten weld metal is held in place by the combined forces of capillarity, surface tension, and viscosity. Welding conditions must be carefully selected and controlled.

Difficulties. In-line welds are progressively different in shape and size, because of the scattering and reduction in beam density and other effects on the beam that take place when it penetrates each layer of metal in turn.

Internal weld spatter may cause difficulty if it interferes with service performance and there is no access for cleaning. Other difficulties such as undercut, underfill, and excessive root bead may be correctable. Hidden welds, which are difficult or impossible to inspect, require some acceptable indirect method of controlling joint reliability.

Applications. Multiple-tier welding has been applied to a variety of joints. Two-tier joints separated by as much as 75 mm (3 in.) have been welded in a single pass in joining honeycomb panels to end frames or structural shapes where only one side was accessible.

Procedure Development and Practice Considerations for Electron-Beam Welding

Welding With Extreme Accuracy

The development of feedback servomechanisms for controlling variations in beam power and beam spot size, as well as in work travel and speed, permits EBW within tolerances of a few thousandths of an inch in high-vacuum machines.

With the use of closed-loop feedback control systems, variations in beam voltage and current can be held to approximately 1% for a 10% variation in the supply line, which is well within the tolerances that can be maintained by the remainder of the system.

Although the diameter of the beam spot cannot be measured directly with special techniques, it can be estimated by measuring the width of the weld or by calibrating the controls for the focusing and beam-power settings. Depending on machine characteristics, beam spots effective for most welding purposes can be set up in the range of 0.25 to 0.76 mm (0.010 to 0.030 in.) in diameter without difficulty. Assuming a very small beam variation, an application requiring a beam spot of only 0.25 mm (0.010 in.) diam would require a means of holding the joint-to-beam coincidence with a runout of less than ±0.13 mm (±0.005 in.).

For welding with a 0.25 mm (0.010 in.) diam beam, a pileup of manufacturing tolerances could easily cause the beam to miss the joint. Therefore, for most production purposes, small beam spots are not economically feasible.

Various types of travel mechanisms of rigid construction are available with runout tolerances of approximately 80 μm/m (0.001 in./ft) or with total deviations of 0.08 to 0.12 mm (0.003 to 0.005 in.) for a complex traverse pattern. However, joint runout depends also on the manufacturing tolerances of the part and on accuracy of fit-up and fixturing. In some applications, 20 to 25 possible sources of error, such as magnetic deflection of the beam, and human and systematic errors in beam alignment, can be found.

Procedure Development and Practice Considerations for Electron-Beam Welding

Use of Scanning

Scanning is a method of checking the runout between the beam spot and the joint to be welded. Its purpose is to indicate the final adjustments needed to align the beam with respect to the joint and to expose possible unsuspected discrepancies in beam behavior or workpiece travel that would interfere with welding. Details of the procedure vary with joint design and the type of equipment available; in one widely used method, a low-power beam is used for visual scanning of the joint during a simultaneous welding pass.

Accuracy of Beam Alignment. Beam alignment is generally a matter of workpiece alignment, as even guns of the adjustable type are usually fixed during welding. Where the gun does traverse the joint, the principles of alignment are the same. Most joints are designed for simple path shapes--straight lines and circles--that can be adequately traced with precision-made carriages, cross-feeds, turntables, eccentric tables, and spindles.

Scanning Procedure. The workpiece is fixtured, and the joint is aligned for travel direction and is moved to its approximate location under the gun. The chamber is closed and pumped down to a vacuum sufficient for safe operation of the gun. A low-power beam is turned on and focused sharply on the workpiece surface. The specific power settings will vary with the working distance selected and the power ratings of the machine; however, they should be sufficient only to create a detectable spot without overheating the workpiece.

The joint line is then adjusted to its precise location with respect to the beam spot, the workpiece is traversed through a complete cycle, and the runout is equalized. Maximum allowable tolerance on runout is determined from a consideration of joint design and weld width, which is mainly a function of the spot size and travel speed.

Considerable latitude in required accuracy of alignment can be gained where beam oscillation can be used. Usually, if the runout approaches one-half the effective welding beam spot width (spot size plus oscillation amplitude), the cause must be determined and corrected. If the runout is within tolerances, production parts ordinarily need only periodic scanning as a quality-control measure.

Problems in Scanning. Some of the causes of excessive runout and other difficulties encountered during scanning are as follows:

- POOR JOINT DESIGN, INACCURATE JOINT PREPARATION
- IMPROPER FIT-UP, INADEQUATE FIXTURING
- LACK OF PRECISION IN WORKPIECE TRAVERSING MECHANISM
- OBSTRUCTIONS IN THE BEAM PATH
- INDISTINCT OR UNDETECTABLE JOINT LINE
- BEAM DEFLECTION BY ELECTRIC OR MAGNETIC FIELDS

The corrections required for mechanical discrepancies are self-evident; the effects of electric and magnetic fields may be less obvious.

The negatively charged electrons that comprise the electron beam constitute a negative space charge. When the electrons strike an insulated metal part, negative charges are built up, causing mutual repulsion between the part and the beam. By grounding the part, excess electrons are conducted away. To avoid unintentional deflection of the beam by nearby magnetic fields, workpieces, fixtures, and tooling components made of magnetic materials should be demagnetized before welding.

Scanning techniques vary in detail, depending on the type of equipment used and also on joint design. It is not always necessary to position the beam spot precisely on the joint line. When welding dissimilar metals, dissimilar thicknesses of the same metal, or certain self-locating joints, it is often necessary to position the beam a short distance from the visible joint line. A corner edge or a scribed line can often serve as a reference, or measurement can be made with an optical grid.

Some special-purpose EBW machines are not intended for scanning. Nearly all general-purpose machines, however, make some provision for this operation. On machines with a fixed gun located in an upper section of the vacuum chamber, a built-in internal optical viewing system is generally provided for scanning. Machines having movable guns can be equipped with external telescopes and internal mirror systems.

Optical Scanning. Optical systems usually consist of horizontally mounted monocular or binocular telescopes with internal illumination and reflecting mirrors that provide a line of sight coaxial with the beam path. Magnifying power may be 10 to 40 times, depending partly on working distance. High-precision alignment can be obtained.

There are two common methods of using this equipment. One method makes use of a low-power beam that is sighted telescopically on the workpiece surface. Using a calibrated grid, measurements to 0.025 mm (0.001 in.) can be made. In the other method, the beam is centered on a crosshair reticle embodied in the optical system. The beam is then turned off and the workpiece joint is moved into coincidence with the crosshair, under internal illumination. The advantage of the latter method is that, after the beam is turned off, the chamber need no longer remain under vacuum, and further alignment work can be done in the open, if necessary.

Before scanning, the mirror surfaces of the internal optical system should be cleaned, because vapor deposition takes place during welding. The frequency of cleaning may vary from a few to many hours of operation, depending on the amount of vapor created and the proximity of the weld. Mirror surfaces are usually protected with glass shields that can be removed for cleaning or replacement.

Scanning Without Optics. Scanning can also be done without visual aids where great precision is not required. On machines that are not equipped with viewing optics, scanning is done by direct observation through sight-glass ports. From one to four such ports may be located in the chamber wall, usually spaced 90° apart, to permit checking beam alignment from several angles. The beam spot is generated the same as for optical scanning.

Electronic Scanning. Checking the location of an electron beam with respect to the joint can be done with an electronic device that displays the relationship on an oscilloscope. The equipment makes use of a low-power electron beam that is made to oscillate at 60 Hz transversely across the joint. The amplitude of the oscillation is sufficient to display the workpiece surface for a short distance on either side of the joint, as well as the joint line itself. In addition, each oscillation pulse is interrupted momentarily to permit the beam to assume its normal welding position, causing a relatively large spot to appear on the display. When the spot coincides with the joint line of the display, the beam is accurately aligned for welding.

Figure 20(a) shows the display condition for a beam centered exactly on the joint. The V represents the joint line, with the beam spot centered accurately; the two horizontal arms (of equal length with a centered beam) at the top of the V represent the adjacent workpiece surface. The broken lines above the work (at left in Fig. 20a) show the plane of oscillation; the wavy line along the joint in the same view shows the approximate amplitude of oscillation.

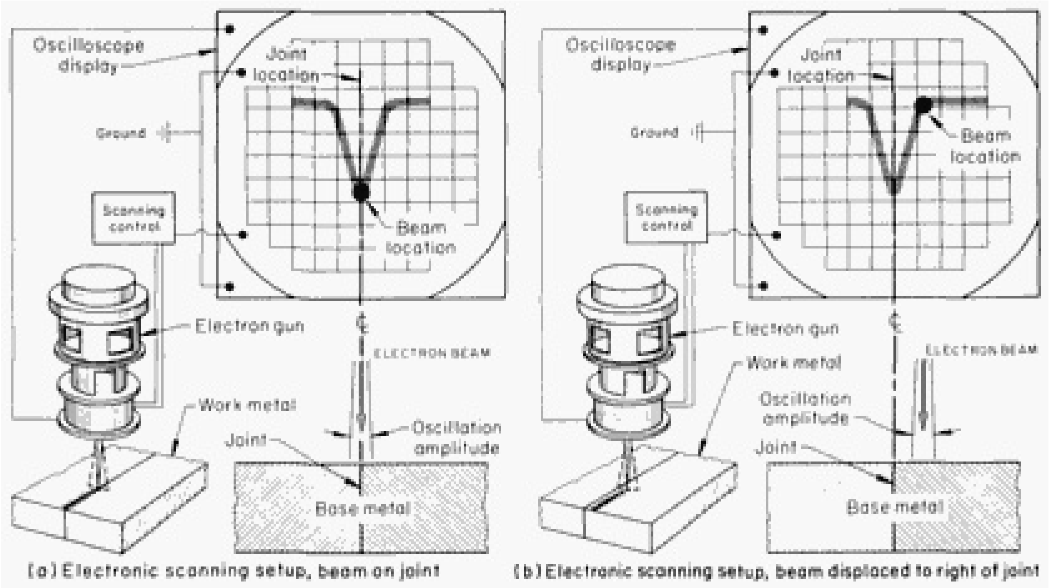

(a) Electronic scanning setup, beam on joint

(b) Electronic scanning setup, beam displaced to right of joint

FIG. 20 ELECTRONIC SCANNING APPARATUS WITH CRT DISPLAY TO SHOW ALIGNMENT OF ELECTRON BEAM RELATIVE TO WELD BEFORE WELDING OPERATION IS CARRIED OUT. (A) ELECTRON BEAM CENTERED ALONG LENGTH OF JOINT. (B) ELECTRON BEAM DISPLACED FROM CENTERLINE OF JOINT

If the beam location were displaced to the right of the joint, the display would look like Fig. 20(b); an opposite-handed display would indicate displacement to the left. By the relative depth of the V, the display also indicates the amount of joint separation from apparent zero to approximately 0.15 mm (0.006 in.). A joint mismatch of a few thousandths of an inch is indicated by the relative height of the horizontal arms. In setting up for electronic scanning (under vacuum), the work is provisionally aligned as for optical scanning, making sure that the joint line is parallel to the direction of travel.

Joint Tracking

Joints that have a path that deviates from a straight line or circle usually require an automatic tracking device for EBW of production quantities. For small production lots, manual tracking, using a coordinate drive having separate controls, sometimes can be used. Whether the method is controlled manually or automatically, the motion required to generate the curve is usually imparted to the workpiece, although limited motion can be imparted to some guns.

Manual Joint Tracking. In manual tracking, the welding operator must closely observe the joint at the point where the weld is being made to anticipate any change in direction. The conditions best suited to this type of operation are: (1) a joint path consisting of a smooth curve, (2) slow travel speed, and (3) a low-power beam with oscillation, to form a weld as wide as possible. The welding operator is capable of regulating two motions of the workpiece (or gun) by manually adjusting the remote controls of the travel mechanism.

For a low heat input, to permit the use of the low welding speeds needed when manual tracking is used, beam pulsation can be combined with beam oscillation.

Automatic Joint Tracking. Mechanical, electromechanical, numerical control, and computerized systems are sometimes used for welding joints that deviate from straight lines or circles. The use of computerized numerical control (CNC) has become highly prevalent in EBW equipment. The capability of a CNC for allowing the weld seam path of a part to be quickly "digitized" and "edited" makes it an ideal tool for programming and tracking the actual weld seam path.

Electromechanical joint tracking systems make use of a stylus, or probe, that rides in the joint. Lateral movements of the stylus (which precedes the beam), resulting from a change in joint position, are converted to electrical signals by a transducer. The electrical signals drive a positioning servomotor that maintains preset alignment of the joint-to-gun focus. These electrical signals define a right error, a left error, and the correct gun position. Precision is usually better, response-delay time is shorter, and other problems associated with mechanical tracking with templates are avoided.

Tape-Controlled Joint Tracking. Workpiece (or gun) motion can be programmed for the complete welding cycle by conventional numerical-control equipment. The actual position of the workpiece with respect to the programmed tape input can be controlled to approximately ±0.025 mm (±0.001 in.). As with other tracking systems in which the generated motion is independent of the joint path, the two paths must coincide. There must be a high degree of assurance that the prior processing of the workpiece, its fixturing, and beam alignment have been done with repeatable accuracy. The tape input information must be sufficiently fine to generate the curve within the desired accuracy.

Beam Oscillations

Through the use of a two-axis deflection coil, either a static (fixed) or an oscillatory (periodic) motion can be imparted to the beam spot through a separate deflection control module. Generally, this control consists of a set of amplifiers for driving the *x*-*y* axes and a function generator for programming the beam spot motion desired. Beam spot motions employed comprise a variety of shapes and may consist of straight, fixed, or periodic deflection, or a combination of both. Square wave, circular, hyperbolic, or parabolic beam spot motions can easily be programmed, employing angular beam deflections of up to 15° and frequencies greater than 5000 Hz. This beam oscillation capability allows wider welds, slower cooling rates, and more uniform weld shapes to be produced, without necessarily using beam defocusing. Also, beam oscillation capability reduces the need for accurate beam-to-seam alignment and makes precise joint tracking less critical. In addition, programmed beam oscillation can be used to control microshuts and porosity in partial-penetration welds, thus reducing undercut and underfill for better bead control and for maximizing penetration capability.

Use of a Pulsed Beam

Beam operation can be changed from continuous to intermittent on machines equipped with pulsing controls. Although available repetition rates (frequencies) and pulse lengths depend on the control unit, frequencies in the range of 0.1 to 3000 Hz with pulses up to about 60 to 70 ms in duration are representative. Beam pulsation reduces the rate of heat input, but is independent of other beam conditions. Therefore, pulsation can be combined with oscillation and deflection, as well as travel speed, to influence weld behavior.

At very low frequencies, such as 1 Hz, with a 5 to 10 ms pulse length, each pulse produces a separate weld, even at low travel speed. By increasing travel speed and adjusting pulse frequency and pulse length, tack welds, spot welds, or intermittent welds can be made at normal production rates. Increasing pulse frequency to approximately 35 to 100 Hz while maintaining short pulse lengths makes it possible to overlap the successive welds to form a continuous weld at speeds suitable for manual tracking.

Pulsed beams generally result in lower peak temperatures, especially in the area adjacent to the weld, as compared to continuous welding, although total heat input may be greater because of the slower welding speed used. At higher pulse frequencies, pulsation has been used to control the solidification pattern of the weld and the microstructure of the HAZ. Pulsation also has been used in rewelding to fuse cold shuts, fill gas pockets, and smooth irregular root areas.

Procedure Development and Practice Considerations for Electron-Beam Welding

Electron Beam Welding as a Repair Method

Restoration of worn or damaged components traditionally has been an important application for welding technology. Some repairs are approached by welding a new detail in the area requiring restoration. However, it is frequently more suitable and cost effective to use a weld buildup process to restore the dimensions and structural integrity of the affected area. Welding processes such as manual or automatic GTAW or GMAW traditionally have been utilized for this type of repair; these processes facilitate the addition of filler wire to accomplish a weld buildup. While these processes enable efficient and relatively fast buildup of filler metal, they have a potential drawback of relatively high heat input, which may cause distortion when section thicknesses are small.

The EBW process has been employed extensively for original equipment manufacture mainly because of its ability to minimize or eliminate distortion through high weld speed, small HAZ, and small bead size. In the past, application of EBW for repair has been relatively limited, confined primarily to butt welding of repair insert details, penetration of parent metal (e.g., to eliminate a crack), or melting of preplaced filler wire. The matching of automatic wire-feed equipment to EBW equipment, however, has led to the development of a system that incorporates the inherent advantages of EBW systems and offers the additional capability of achieving a precise, multiple-pass weld buildup with minimum heat input. The addition of this capability has greatly increased the utilization of EBW as a repair mechanism.

Electron-Beam Wire-Feed Process and Equipment for Repairs

Electron-beam welding units used for weld buildup repairs are high-voltage, high-vacuum machines (150 kV, 13 mPa, or 10^{-4} torr). The availability of a highly controlled welding atmosphere is a distinct advantage for the electron-beam process, particularly when welding reactive materials such as titanium. High-voltage, high-vacuum EBW machines are equipped with signal generators that oscillate and deflect the beam in the *x*-*y* directions. They are also capable of making circles and other geometric figures. Beam deflection frequency can be varied, and the beam can be focused above or below the point of impingement. Beam pulsing and modulation of the pulse are also possible. All these tools can be used in weld buildup repairs. Current electron-beam vacuum chambers can accommodate parts up to 1525 mm (60 in.) in diameter. On the floor of the chamber is an *x*-*y* table; fixed to this table is a variable-speed rotary table. The component to be repaired, typically a cylindrical ring or case, is fixtured to this rotary table. An automatic filler-wire feeder delivers wire to the beam impingement point, enabling a continuous multiple-pass weld buildup when the component is rotated under the beam. Typically, a fused spot is less than 0.51 mm (0.020 in.) in diameter, allowing for an extremely small HAZ.

The wire-feed equipment is an adaptation of standard GTAW wire-feed equipment, and the technique is similar to that employed in cold wire circumferential automatic GTAW. The precision wire feeder delivers wire through a nozzle, which can be positioned in the *x, y,* and *z* directions and which is easily visible in the coaxial optical system of the electron-beam welding machine. Wire entry angle, nozzle position, wire speed, weld speed, accelerating voltage, focal position, deflection, and amperage are critical factors in EBW buildup. Both inside and outside surfaces can be restored, depending on the position of the wire feeder within the chamber.

A relatively small filler-wire diameter (≤ 0.89 mm, or 0.035 in.) typically is employed to attain a precise low-heat input and a weld buildup typically 1.0 to 1.3 mm (0.040 to 0.050 in.) wide and 2.54 to 7.62 mm (0.100 to 0.300 in.) high, although much larger buildups could possibly be attained by multiple applications of the process. Electron-beam weld buildup lends itself readily to automation, thereby opening the process to use by less skilled operators. On some machines, all parameters are preset. The operator merely lines the machine up on the surface to be welded and pushes the start button. The wire feeder automatically activates and adjusts itself to produce a smooth, precise buildup.

Application of Electron-Beam Wire-Feed Process for Repairs

The electron-beam wire-feed process has been applied primarily for restoration of worn or damaged gas turbine aircraft engine components, which are typically cylindrical in shape with diameters ranging from 100 to 1525 mm (4 to 60 in.), lengths from 25 to 915 mm (1 to 36 in.), and section sizes from 1.3 to 12.7 mm (0.050 to 0.500 in.). Repair of gas turbine engine components has become increasingly attractive to commercial and military operators because of the cost and lead time involved in the procurement of spare parts. Repairs are typically conducted at 50 to 70% of the cost of a new component, with three to six times improvement in replacement time.

One of the key gas turbine engine component repairs is the restoration of the rotating labyrinth-type air seals used to maximize engine pressure gradients. As seals wear or are damaged, engine efficiency drops and fuel consumption rises. Hence, restoring minimal seal clearances has become increasingly attractive. Seal design is typically of the straight knife edge or tapered knife edge type (Fig. 21). Materials vary from iron-base alloys (AMS 6508 and A-286) to high-nickel alloys (Incoloy 901 and Waspaloy) to titanium alloys (Ti-6Al-4V, Ti-8Al-1Mo-1V, and Ti-6Al-2Sn-4Zr-2Mo).

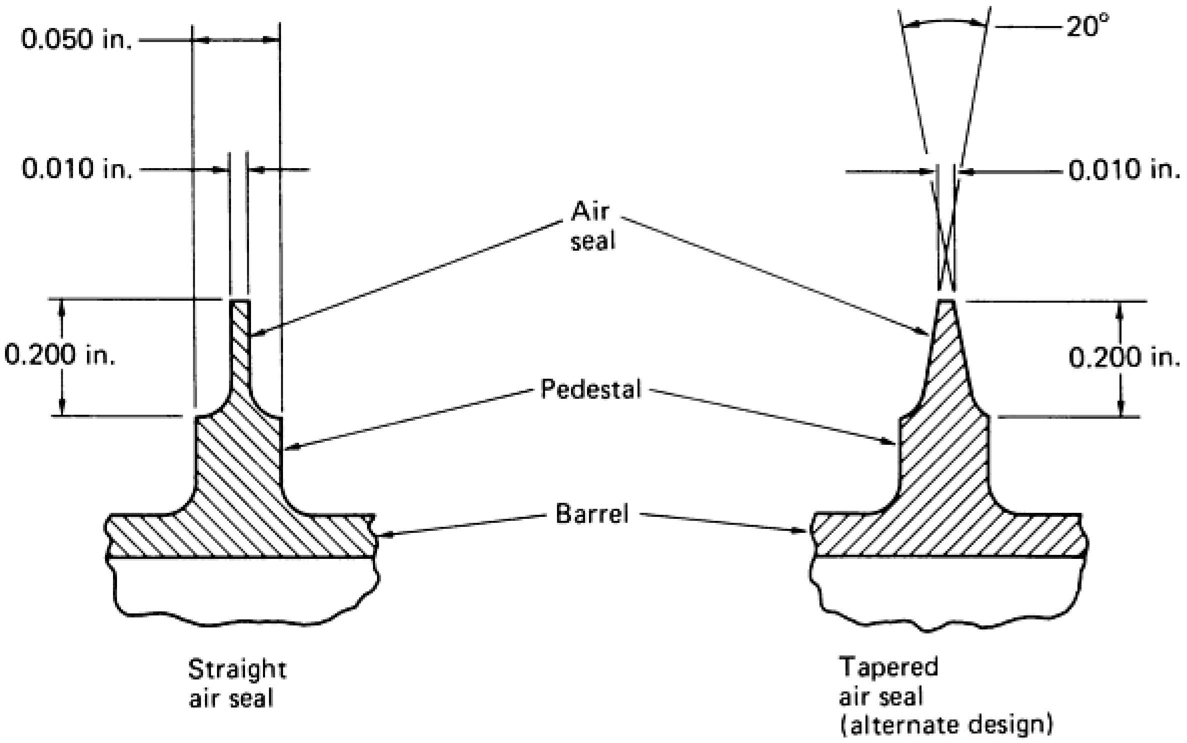

FIG. 21 SCHEMATIC COMPARING CROSS SECTIONS OF STRAIGHT AND TAPERED AIR SEALS

Currently, virtually all air seal restoration is done by welding. Gas-tungsten arc welding, gas-metal arc welding, and electron-beam welding have all been successfully applied depending on requirements of engine manufacturers, component design, material, and experience and equipment of repair stations. Prior to the establishment of the electron-beam wire-feed process, utilization of the standard electron-beam weld process for this type of repair was limited to competitive weld buildup processes (GTAW and GMAW) and was primarily applied where distortion and HAZ size were a primary concern.

Initial application of the standard electron-beam process involved welding a machining ring onto a premachined pedestal on which the worn air seal was removed, as shown in Fig. 22. The knife edge configuration was then machined into the welded ring. Later, a split ring concept was introduced. In this procedure, an edge-rolled ring was assembled into a shallow groove premachined into the pedestal top (Fig. 22). Circumferential spike welds from each side welded the ring in place and the ends of the split ring were radiused to blend the discontinuity in the ring. The wall thickness of the as-rolled ring was the finished thickness of the knife edge, typically about 0.25 mm (0.010 in.), and the outside diameter was finish machined after welding. Disadvantages of this weld buildup system include ring detail cost and the amount of post-repair machining required.

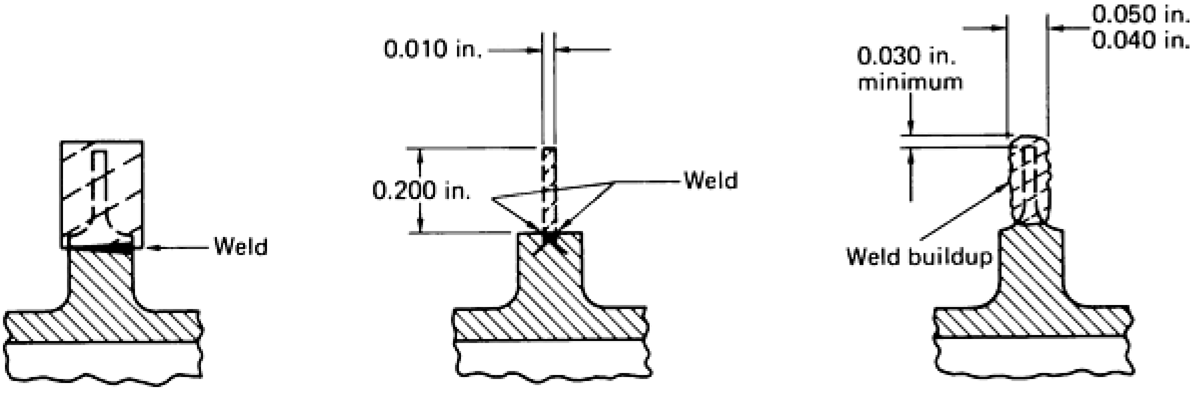

FIG. 22 SEQUENCE OF OPERATIONS REQUIRED FOR EBW REPAIR OF AN AIR SEAL. SEE TEXT FOR DETAILS.

The application of the precision wire-feed electron-beam process (Fig. 22) to gas turbine seal repairs provides both the material application efficiency and cost advantages of weld buildup systems and the inherent low heat input and controlled atmosphere advantages of the EBW process. A typical automatic wire-feed electron-beam air seal repair, applied to a compressor rotor spacer for a gas turbine engine, is shown in Fig. 23. This part is made from a titanium (Ti-6Al-4V) forging. The outside diameter of the knife edges wear in service, and the knife edges must be restored to maintain the proper tip clearance. The knife edge material is machined close to the diameter of the pedestal; a weld buildup is made using Ti-6Al-4V titanium filler wire, and a new knife edge is machined. The multiple-pass technique typically deposits 0.25 to 0.38 mm (0.010 to 0.015 in.) of radial buildup per pass. Penetration of the HAZ below the pedestal top typically does not exceed 0.38 mm (0.015 in.). Because the entire weld buildup operation is performed in a vacuum, there is no weld metal or HAZ oxidation. Cost is minimized by the high speed of the EBW process, which substantially exceeds that of the standard GTAW buildup process.

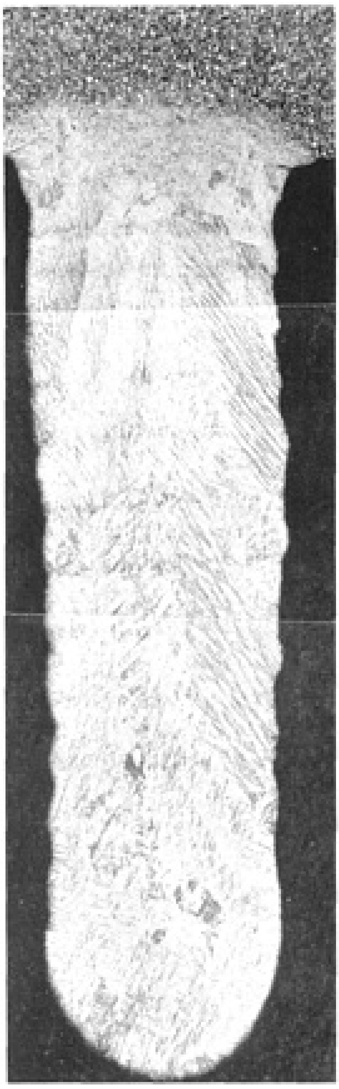

FIG. 23 PHOTOMICROGRAPH OF A TYPICAL AUTOMATIC WIRE-FEED ELECTRON-BEAM REPAIR OF AN AIR SEAL. 15×

Air seal repairs also are routinely performed on steel and nickel alloy parts. Experience with nickel-base superalloys containing relatively high percentages of gamma prime that form hardening elements has shown these alloys to be crack sensitive during welding or postweld heat treatment. Crack-free precision wire-feed EBW buildup repairs, however, are achieved when appropriate process parameters are utilized.

In applying the process to steel alloys, parameters should be optimized to avoid a pronounced swirl pattern at the buildup interface which may be encountered due to uneven mixing of the weld buildup and substrate. Potential swirling is associated with certain steel alloys and may be a potential source of porosity or cracking.

Quality assurance testing is an integral part of all gas turbine engine repairs. The high degree of control inherent in the precision wire-feed electron-beam repair process facilitates in-process control and quality assurance. Because of the high degree of control the electron-beam process offers, process variables can be kept to a minimum. This built-up consistency tends to ensure a long-term high level of quality assurance.

Weld parameters are developed by the analysis of microsections of sample welds and by nondestructive testing. Once an acceptable process has been established for a component, a weld schedule is initiated and weld parameters are maintained. Any changes in equipment or process must be qualified by the analysis of new microsections. Quality of finished air seals is ensured by one or more of the following tests, depending on customer requirements:

- VISUAL
- LIQUID PENETRANT
- FLUORESCENT MAGNETIC PARTICLE
- X-RAY
- ETCH INSPECTION

Future Applications for Repair

Although the most common application of the precision wire-feed electron-beam buildup process has been in the repair of gas turbine engine air seals, the process is not limited to this type of application. Various other wear restoration applications have been developed, including spline and groove buildups. Although the process is most efficiently utilized as a continuous multiple-pass process on cylindrical components, planar buildup can be achieved by utilizing sequential weld pass patterns or oscillation. The process is also potentially applicable for new part manufacture to enable a component to be machined from a smaller, less costly forging with selective electron-beam buildup to provide material for specific structural features (for example, outside diameter bosses on cylindrical cases).

Procedure Development and Practice Considerations for Electron-Beam Welding

Joining of Dissimilar Metals

The joining of dissimilar metals is an important application of EBW. The weldability of two dissimilar metals or alloys depends primarily on their physical properties, such as melting points, thermal conductivities, and coefficients of thermal expansion. Weldability generally can be determined by examination of the phase diagram of the alloys to be welded. If intermetallic compounds are formed by the metals to be welded, the weld metal will be brittle. However, many combinations that are difficult to weld or unweldable by other processes can be welded by the electron-beam process.

Procedure Development and Practice Considerations for Electron-Beam Welding

EBW of Low-Carbon Steel

Low-carbon steels are readily electron-beam welded. Grain size in both the weld and the heat-affected zone (HAZ) is significantly smaller than in arc welds, because of the lower heat input and extremely rapid heating and cooling. Rimmed steel causes excessive gassing. A technique used to weld rimmed steel is to sandwich an aluminum alloy shim about 0.25 mm (0.010 in.) thick in the joint to provide deoxidizing action during welding. Welding at a low travel speed or otherwise adjusting welding conditions to produce a more tapered weld and thus to allow more time for the escape of gases from the molten weld metal also helps to reduce porosity. Semikilled steel is easier to weld than rimmed steel, but some oxygen may remain to cause porosity. Fully killed steel is readily weldable.

Procedure Development and Practice Considerations for Electron-Beam Welding

EBW of Hardenable Steel

The variation in weldability by EBW among hardenable steel follows the same pattern as for arc welding of these steels. Properties of electron-beam welds in a given hardenable steel differ from those of arc welds mainly because of the narrowness of the fusion and HAZ and the extremely rapid heat-melt-cool cycles.

Hardness traverses across the midpoint of electron-beam (high-vacuum) and gas tungsten arc welds in 1.57 mm (0.062 in.) thick heat treated 4340 steel are plotted in Fig. 24. For both welds, maximum hardness was produced uniformly in the weld metal, and hardness dropped off abruptly at the edge of the weld metal.

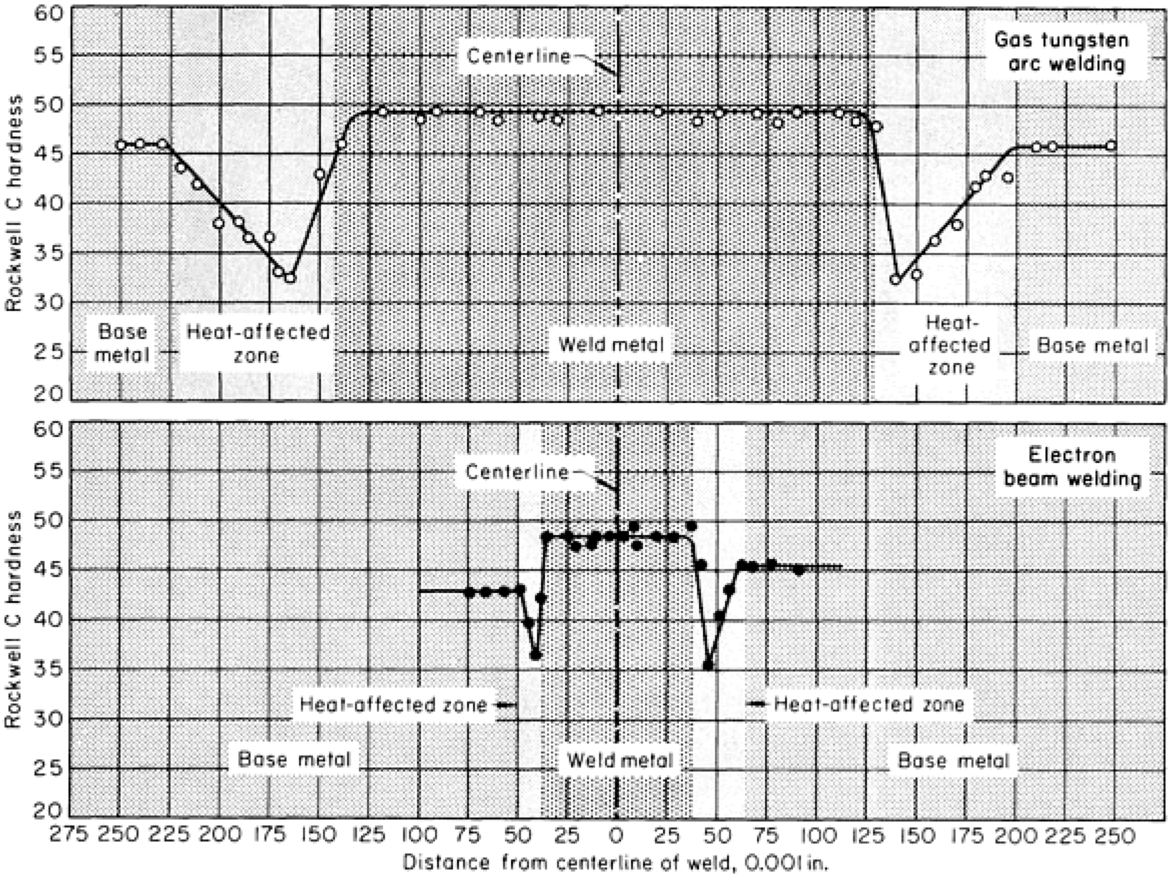

FIG. 24 HARDNESS PROFILE OF HIGH-VACUUM EBW AND GTAW BUTT JOINTS USED TO JOIN 1.57 MM (0.062 IN.) THICK MEDIUM-CARBON ULTRAHIGH-STRENGTH STEEL

Similar hardness traverses for electron-beam welds in 9.65 mm (0.380 in.) thick heat-treated 1024 (modified) are shown in Fig. 25. The hardness profile across this weld was similar to that for the electron-beam welds in 1.57 mm (0.062 in.) thick 4340 steel. Tempering after welding eliminated or minimized the hardness differential in the HAZ (Fig. 25).

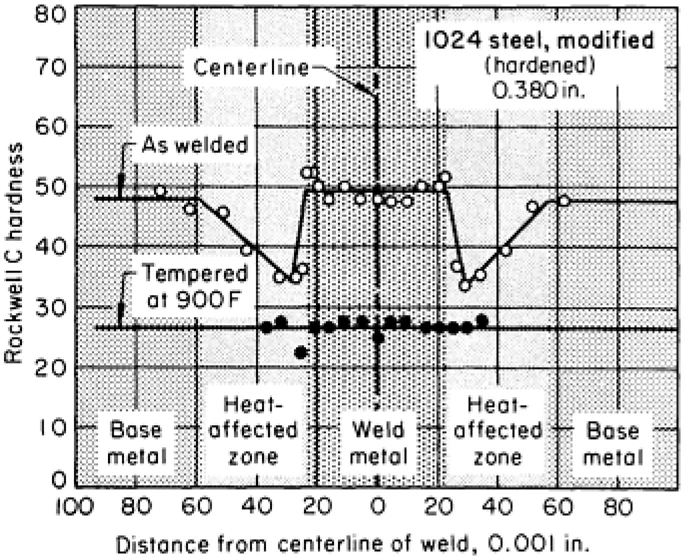

FIG. 25 HARDNESS-PROFILE OF HIGH-VACUUM EBW BUTT JOINT USED TO JOIN 9.65 MM (0.380 IN.) THICK

Prevention of Cracking. Cracking can be a problem in EBW of hardenable steel, just as in arc welding, if the weld is made in a highly restrained joint, especially in welding parts that have been hardened. Deep circular welds in heavy sections are particularly troublesome. Partial-penetration welds are more likely to crack under restraint than full-penetration welds. It is desirable to place the joint in a location that will allow the part to shrink freely as it cools after welding. Welding through carburized or nitrided cases is not recommended.

Cracking of electron-beam welds in hardenable steels can also be minimized by reducing the welding speed to allow more buildup of heat in the base metal, and by preheating, postheating, and allowing the work to cool in the chamber after welding.

Medium-carbon steels except for the free-machining types, are readily electron-beam welded (for hardness traverses, see Fig. 25). Weldability decreases with increasing carbon content.

Low-alloy steels containing less than 0.30% C are usually electron-beam welded without preheating or postheating. When preheating is used--to prevent cracking in highly restrained joints, for example--a temperature of 260 to 315 °C (500 to 600 °F) is usually adequate. If the part has been hardened and tempered prior to welding, the postweld tempering must be done at a temperature slightly below that at which the base metal was originally tempered.

Two of the most frequently electron-beam-welded alloy steels of this type are 8620 and 9310. Components made of these steels often are case hardened, then assembled by EBW, without distortion or heat damage to the case. Before welding, the case should be removed from the immediate vicinity of the joint to prevent microcracks from forming in it.

High-strength alloy steels containing more than 0.30% C are electron-beam welded either in the annealed or the normalized condition or in the quenched and tempered condition, although weldability is better for the annealed or normalized condition. The hardness profile for electron-beam welds in hardened high-strength alloy steel sections no thicker than about 65 mm ($\frac{1}{4}$ in.) is ordinarily like that shown for 1.57 mm (0.062 in.) thick heat-treated 4340 steel in Fig. 24, and joint strength can approach that of the base metal, without preheating or postweld treatment other than stress relief.

When a shielding gas is substituted for air in nonvacuum welding, assuming the same operating conditions, penetration is greater when a gas is used that has a lower molecular weight (or nominal density) than that of air. Penetration is less when a gas that has a higher molecular weight than that of air is used. In tests on welding 4340 steel, using a fixed beam power of 6.4 kW, working distances of 6.4, 13, and 25 mm ($\frac{1}{4}$, $\frac{1}{2}$, and 1 in.), and welding speeds of about 380 to 2030 mm/min (15 to 80 in./min), penetration was increased when helium was substituted for air and was decreased when argon was substituted for air, as shown in Fig. 26. The relative capacities of helium, air, and argon to scatter the electrons of the beam by collisions of electrons with gas molecules are controlled by their molecular weights and relative densities. On the average, when helium was used as a shielding gas, penetration depth was about twice that obtained when welding was done in air, and when argon was used, penetration depth was about half that when welding was done in air. Thus, with helium shielding, penetration averaged about four times that with argon, especially at higher welding speeds. The speed at which a given thickness of 4340 steel could be welded with full penetration, using helium, was two to four times that with air. Using argon, speed was 25 to 50% that with air.

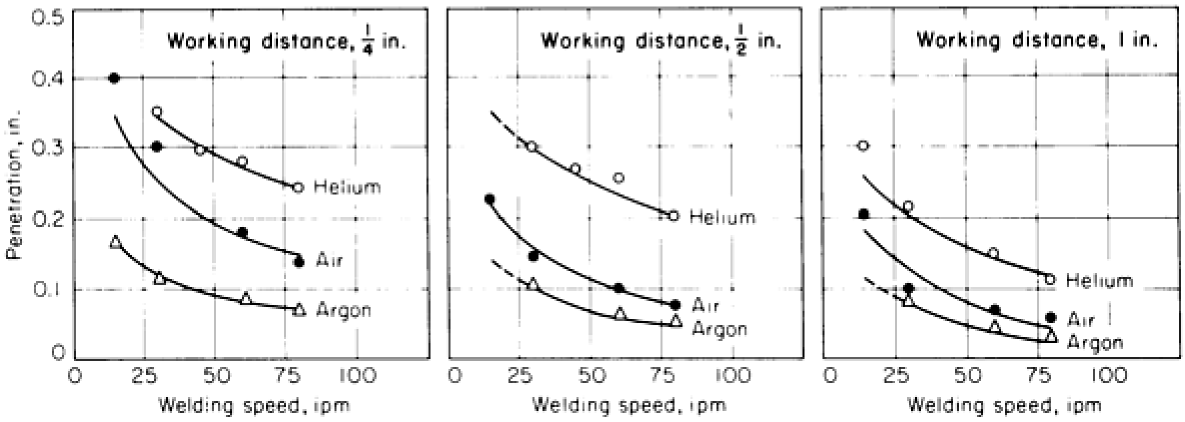

FIG. 26 PLOT OF PENETRATION VERSUS WELDING SPEED AS A FUNCTION OF SHIELDING GASES AND WORKING DISTANCES FOR NONVACUUM EBW OF 4340 MEDIUM-CARBON ULTRAHIGH-STRENGTH STEEL. BEAM POWER WAS 6.4 KW.

In EBW of thicker sections of high-strength alloy steels, preheating or postheating, or both, is usually needed to prevent cracking. In studies on full-penetration, single-pass welding of 15 mm (0.6 in.) thick annealed D-6ac steel, which is used frequently for missile castings, cracking was observed when preheating was not used and was eliminated by preheating at 540 to 565 °C (1000 to 1050 °F). Welding was started with the joint at a temperature of 370 °C (700 °F), and the joint temperature remained above 345 °C (650 °F) for 8 to 10 min after completion of the weld. Hardness profiles for the joints welded with and without preheat are shown in Fig. 27. Sound welds were also obtained in 55.9 mm (2.2 in.) thick annealed D-6ac steel, after preheating to 425 to 455 °C (800 to 850 °F); welding was started when joint temperature was no lower than 260 °C (500 °F).

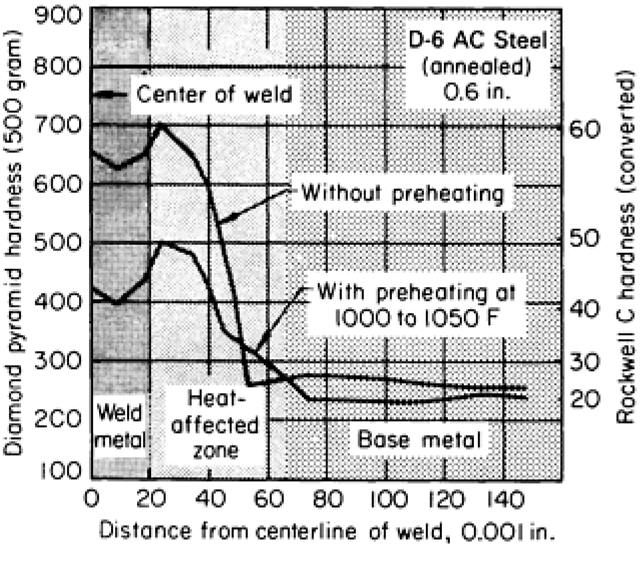

FIG. 27 EFFECT OF PREHEATING ON THE HARDNESS OF 15 MM (0.6 IN.) THICK D-6AC LOW-ALLOY ULTRAHIGH-STRENGTH STEEL

High-Carbon Steels. Because of the low total heat input and the rapid thermal cycling that are characteristic of EBW, weld cracking of steels containing more than 0.50% C is less likely than in arc welding. Even bearing steels such as 52100, which are seldom arc welded, have been electron-beam welded on a high-production basis. The 52100 steel is welded in the spheroidize-annealed condition in applications where joint performance is not critical to the function of the part and where service stresses in the joint are low. Preheating and postheating are not necessary.

EBW of Tool Steels

The chief advantage of EBW over other welding processes in joining tool steels is its ability to produce joints at high speed without annealing or other heat-treating operations. Hardness profiles similar to those shown for low-alloy steels in Fig. 24 and 25 were also obtained for H11 tool steel. For full-penetration butt welds on hardened (50 HRC) sections 5.72 ram (0.225 in.) thick, the hardness of the weld metal and the immediately adjacent base metal was 56 to 57 HRC, and hardness dropped to a minimum of 43 to 46 HRC in the HAZ, as measured after double tempering at 550 °C (1025 °F). The overtempered portion of the HAZ was only 0.13 mm (0.005 in.) wide.

Small dies of D2 tool steel have been electron-beam welded in production. A larger operation is the production of bimetal band saws. Specially designed high-vacuum machines, which are referred to as "air-to-air" machines, are used to produce band-saw blades by welding 1.6 mm ($\frac{1}{16}$ in.) wide M2 high-speed-steel cutting-edge strips to 6150 steel bands. The composite blade is about 0.8 mm ($\frac{1}{32}$ in.) thick and is welded at speeds above 7.6 m/min (300 in./min) in a special vacuum chamber. In "air-to-air" bimetallic strip welding, the backing band and edging strip are continuously fed through slots in a series of differentially pumped chambers and welded in a high vacuum without any interruption of this continuous product flow.

EBW of Hardened and Work-Strengthened Metals

The full properties of hardened and work-strengthened base metals can be retained on functional surfaces very close to the narrow electron-beam weld zone. In addition to confining the decrease in base-metal properties to a very narrow zone, EBW of hardened or work-strengthened metals also produces joint quality superior to that produced by arc welding. Some high-strength steels, however, cannot be electron-beam welded in the hardened condition without cracking or without an unacceptable decrease in strength and hardness. These steels, such as 52100 beating steel, are normally annealed before welding and given a postweld heat treatment for controlled strength and hardness. In Example 1, EBW was chosen over gas-tungsten arc welding (GTAW) because of the narrow fusion zone and HAZ obtained by the electron-beam process.

Example 1: Selection of Electron-Beam Welding to Minimize Tempering in the Heat-Affected Zone.

Fabrication of armor for military aircraft involved the welding of corner joints in 7.9 mm ($\frac{5}{16}$ in.) thick armor plate. The armor plate was made by roll-bonding H 11 tool steel to an alloy steel (AMS 6545) containing 9% Ni, 4% Co, and 0.30% C. Welding was done with the plate in the hardened condition; the H11 steel was hardened to 60 HRC and the alloy steel to 50 HRC. The plate was not difficult to weld. Either GTAW or EBW could be used to obtain sound welds using the joint designs shown in Fig. 28.

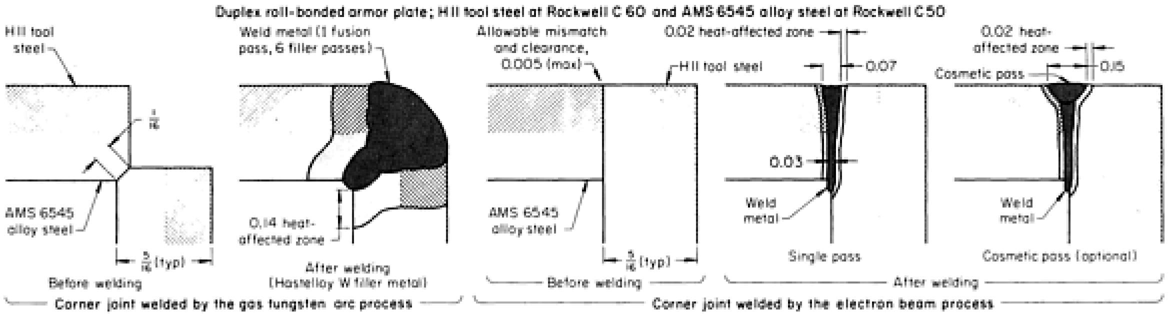
Duplex roll-bonded armor plate; H11 tool steel at Rockwell C 60 and AMS 6545 alloy steel at Rockwell C 50
H11 tool steel
$\frac{1}{16}$
AMS 6545 alloy steel
$\frac{5}{16}$ (typ)
Before welding
Weld metal (1 fusion pass, 6 filler passes)
0.14 heat-affected zone
After welding (Hastelloy W filler metal)
Corner joint welded by the gas tungsten arc process
Allowable mismatch and clearance, 0.005 (max)
H11 tool steel
AMS 6545 alloy steel
$\frac{5}{16}$ (typ)
Before welding
0.02 heat-affected zone
0.07
0.03
Weld metal
Single pass
0.02 heat-affected zone
Cosmetic pass
0.15
Weld metal
Cosmetic pass (optional)
After welding
Corner joint welded by the electron beam process

FIG. 28 SCHEMATIC OF CROSS SECTIONS THROUGH CORNER JOINTS AND WELDS TO SHOW VARIATION IN THE SIZE OF WELDS AND HAZ THAT ARE OBTAINED WITH EBW AND GTAW

The weld and HAZ of the electron-beam process are compared with those of GTAW in Fig. 28. Joint designs differed; if the joint for GTAW had been made the same as the joint for EBW, a V-groove would have been required, and about the same amount of weld metal would have been needed. Gas-tungsten arc welding was done in one fusion pass and six filler-metal passes, using 1.57 mm (0.062 in.) diam Hastelloy W wire.

Electron-beam welding was done in a single pass. Because of the occasional undercutting, a second (cosmetic) pass was sometimes used. Even when the second pass was used, the weld and the HAZ were much narrower than were those created by arc welding, and they were unaffected below the crown of the weld. Because the general effect of the welding heat was to decrease the hardness and strength of the base metal, EBW, which produced less heat, was selected.

There were no unusual problems associated with the EBW procedure. Plates were machined square and then were degreased. Just before welding, surface films were removed from all surfaces to be welded by brushing with a stainless steel wire brush followed by wiping with alcohol. The plates were then fixtured into an assembly roughly resembling a bathtub having sides about 250 mm (10 in.) high and a floor area of about 510 by 1015 mm (20 by 40 in.). A maximum tolerance of 0.13 mm (0.005 in.) was allowed on fit-up for joint mismatch and clearance. Starting and runoff tabs, made of the same duplex plate, were attached to joint ends to contain variations in beam initiation and termination. The fixture was a simple arrangement of toggle clamps and stainless steel blocks that could be set up for flat-position welding on a track-guided carriage in the vacuum chamber.

After beam alignment and vacuum pump-down, each joint was welded in a single pass without tack welding and without beam oscillation. Because undercutting occurred where fit-up approached the maximum tolerance, a second (cosmetic) pass occasionally was used. The effect of the cosmetic pass on the weld cross section is shown in Fig. 28; the machine settings for the defocused beam used to make the cosmetic pass are in the table with Fig. 28.

Procedure Development and Practice Considerations for Electron-Beam Welding

EBW of Stainless Steels

The properties and the welding metallurgy of stainless steels are discussed in the Section "Selection of Stainless Steel" in this Volume.

Austenitic Stainless Steels

The high cooling rates typical of EBW help to inhibit carbide precipitation, because of the short time during which the steel is in the sensitizing-temperature range. Filler metal is used in welding joints that cannot be fitted together closely.

Effect of Air and Helium Atmospheres on Selected Parameters. Figure 29 shows the effect of welding speed on penetration of nonvacuum welds made in austenitic stainless steel and the effect of ambient gas and power on the penetration versus speed curve. Curves identified as helium penetration-versus-speed curves were performed using a helium effluent-style lower orifice design, that is, using a lower (exit) orifice that is capable of blowing helium into the atmosphere with enough directed flow to provide a helium gas column in which the beam can travel.

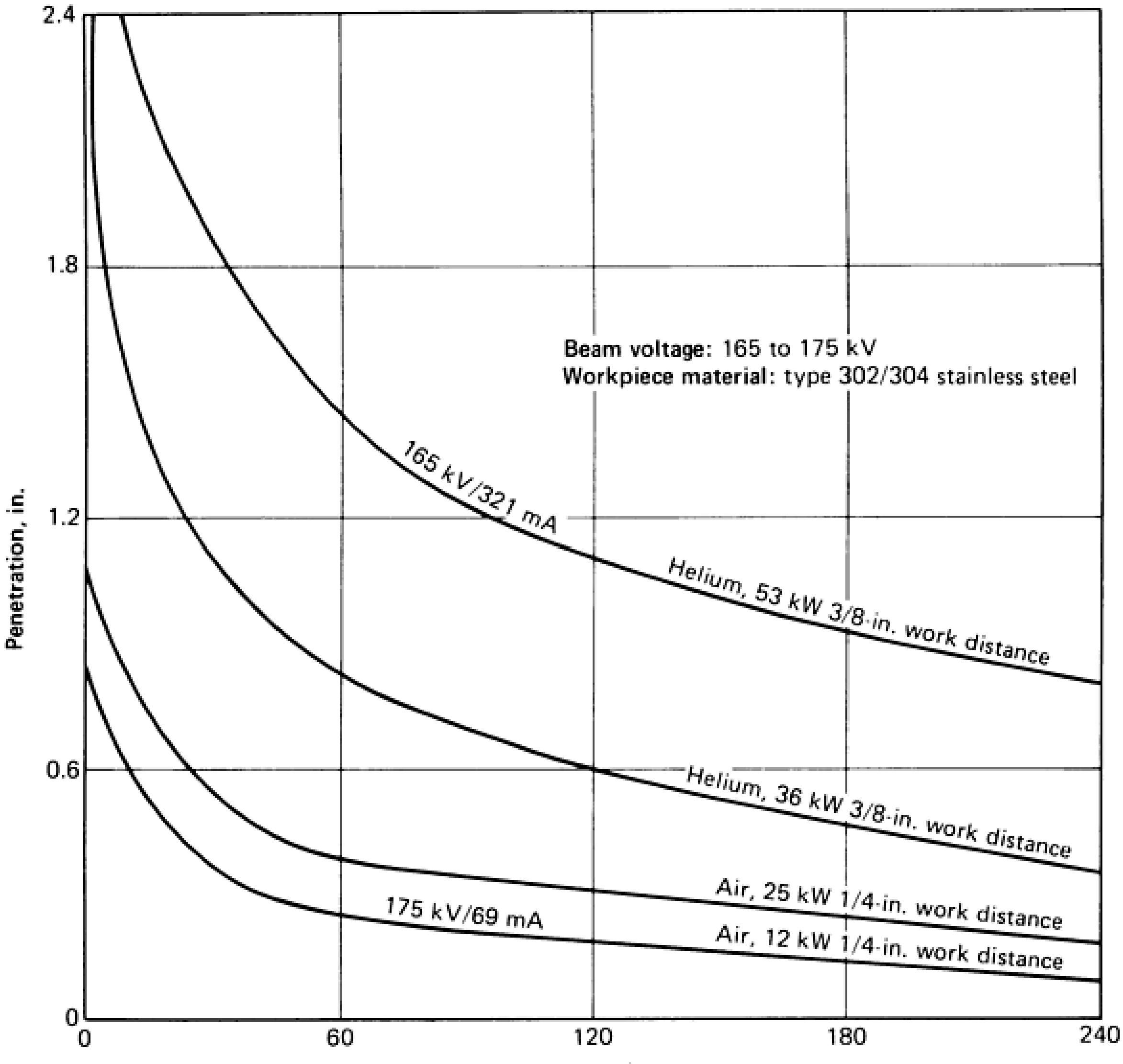

FIG. 29 PLOT OF PENETRATION VERSUS WELDING SPEED AS A FUNCTION OF SELECTED BEAM VOLTAGES, ATMOSPHERES, POWER LEVELS, AND WORK DISTANCES TO WELD TYPE 302 TO TYPE 304 AUSTENITIC STAINLESS STEELS

Example 2: Melt-Through Welding of 0.05 mm (0.002 in.) Thick Foil.

A rupture-disk assembly (Fig. 30) designed for gas-pressure relief consisted of three components: (1) a 50 mm (2 in.) diam nozzle of 2.5 mm (0.100 in.) thick type 304 stainless steel attached to an outer pressure vessel (not shown); (2) a 0.05 mm (0.002 in.) thick cupped rupture disk of type 304 stainless steel; and (3) a 1.02 mm (0.040 in.) thick cap, also of stainless steel, containing an off-center flued-in hole with downward-pointing knife edges.

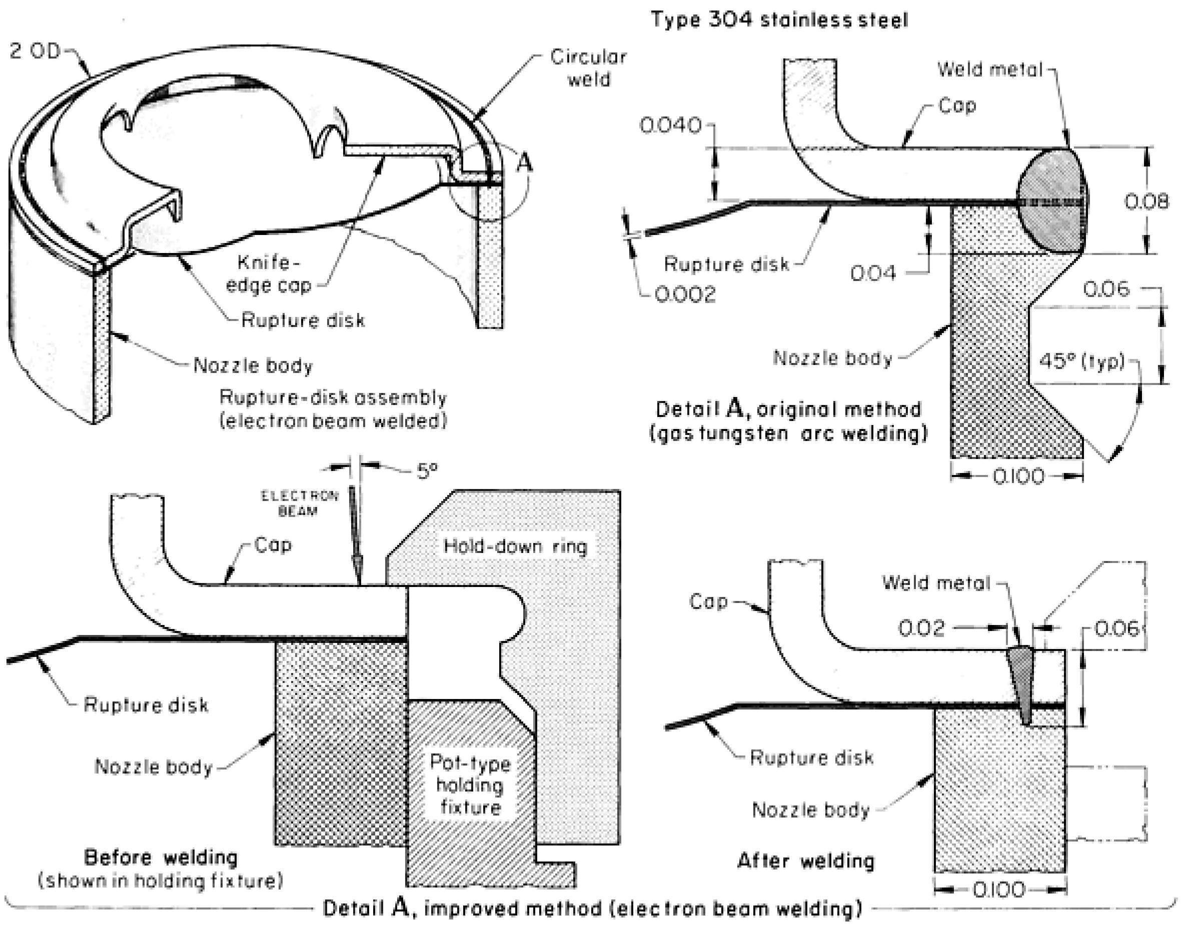
Type 304 stainless steel
2 OD
Circular weld
A
Knife-edge cap
Rupture disk
Nozzle body
Rupture-disk assembly (electron beam welded)
Weld metal
Cap
0.040
0.08
Rupture disk
0.002
0.04
0.06
Nozzle body
45° (typ)
Detail A, original method (gas tungsten arc welding)
0.100
5°
ELECTRON BEAM
Cap
Hold-down ring
Rupture disk
Nozzle body
Pot-type holding fixture
Before welding (shown in holding fixture)
Cap
Weld metal
0.02
0.06
Rupture disk
Nozzle body
After welding
0.100
Detail A, improved method (electron beam welding)

JOINT TYPE	**CIRCULAR, THREE-PIECE COMER**
WELD TYPE	**MELT-THROUGH (SPIKE)**
MACHINE CAPACITY	**50 KV AT 250 MA**
GUN TYPE	**DIODE; FIXED DURING WELDING**
VACUUM CHAMBER	**STEEL: 1370 × 1220 × 1220 MM (54 × 48 × 48 IN.) WITH FULL-WIDTH END DOORS**
FIXTURES	**56-PIECE HOLDING TOOL WITH INDIVIDUAL ROTATION; TABLE WITH *X-Y* MOTION**
WELDING POWER	**26 KV AT 16 MA**
WELDING VACUUM	**0.65 MPA (5 × 10^{-6} TORR)**
PUMPDOWN TIME	**10 MIN**
WORKING DISTANCE	**38 MM (1.5 IN.)**
BEAM FOCAL POINT	**SHARP AT WORK SURFACE**
BEAM SPOT SIZE	**~0.51 MM**

	(~0.020 IN.) DIAM
WELDING SPEED	**2540 MM/MIN (100 IN./MIN)**
NUMBER OF PASSES	**ONE, PLUS 30° DOWNSLOPE**
SETUP TIME	**48 MIN**
PRODUCTION RATE	**44$\frac{1}{2}$ PIECES/H**

FIG. 30 COMPARISON OF EBW AND GTAW WELDS APPLIED TO A TYPE 304 STAINLESS STEEL RUPTURE DISK ASSEMBLY

The three components were joined by a single weld that was required to have a leak rate of less than 10^{-8} cm^3/s (6 × 10^{-10} $in.^3/s$) when tested with a helium mass spectrometer. When pressure in the nozzle exceeded the design limit, the cupped rupture disk reversed its shape from concave to convex, puncturing itself against the knife-edge of the cap. Leaktightness, rather than high tensile strength, was the welding objective. The material was selected for long, maintenance-free service, rather than for resistance to a corrosive environment.

Originally, the components were joined in one pass by GTAW, using the design shown in detail A, original method, Fig. 30. A circular groove was cut in the nozzle wall to help localize the weld bead, which was made without filler metal. Light pressure was applied at both ends of the assembly to hold the parts in alignment for welding.

This procedure resulted in numerous rejects because of leaks. Often, difficulty was encountered in getting the disks to lie absolutely flat on the joint surface. The 0.05 mm (0.002 in.) thick rupture disks, having been blanked and formed in the work-hardened condition, were tough, springy, and sometimes burred. As a result, they occasionally were burned back, instead of being fused into the weld bead, causing leaks. It was usually impossible to salvage this type of defect by rewelding.

To eliminate these difficulties, the operation was converted to EBW, even though initial equipment and tooling costs were high. The EBW machine used for this application had a relatively large vacuum chamber with full-width doors to facilitate loading and unloading of large, multiple-part fixtures. The holding fixture used had 56 receptacles to position and individually rotate 56 assemblies under a stationary gun. The fixture was mounted on a horizontal table equipped with *x*-*y* motion control to permit optical alignment of the beam for welding each part in turn.

Each assembly was fitted into a cylindrical receptacle or pot that was equipped with a hold-down ring that forced the periphery of the disks and caps into close alignment, as shown in detail A, improved method, Fig. 30. A melt-through or spike weld was made, rather than an edge-flange weld at the periphery.

With this setup, the blanking and forming operations on the disks could be allowed the more realistic tolerance of ±0.05 mm (±0.002 in.) on flatness. The electron beam was angled slightly and was positioned close to the hold-down ring so that, in the event of a leak, the joint could be rewelded by running a second weld just inside the first weld.

The weld-localizing groove needed for the original GTAW was eliminated, and the electron-beam melt-through weld was made at a heat input of 0.0098 kJ/mm (0.25 kJ/in.) of weld length, which was much lower than for arc welding. The hold-down ring provided a slight chill effect, but enabled the nozzle body to function effectively as a heat sink.

Welding results from the electron-beam procedure were good. Joint defects of all types dropped to 3%, and most of the rejects could be salvaged by rewelding. A production rate of 44$\frac{1}{2}$ assemblies per hour was a significant improvement over the original arc method.

Martensitic Stainless Steels

Although these steels can be electron-beam welded in almost any heat-treated condition, welding will produce a hardened, martensitic HAZ. Hardness and susceptibility to cracking increase with increasing carbon content and cooling rate.

Precipitation-Hardenable Stainless Steels

Precipitation-hardenable stainless steels can, in general, be electron-beam welded to produce good mechanical properties in the joint. The semiaustenitic types, such as 17-7 PH and PH 14-8 Mo, can be welded as readily as the 18-8 types of austenitic stainless steel. The weld metal becomes austenitic during welding and remains austenitic during cooling. In the more martensitic types, such as 17-4 PH and 15-5 PH, the low-carbon content precludes formation of hard martensite.

Some of the precipitation-hardenable stainless steels have poor weldability because of their high phosphorus content. Steels 17-10 P and HNM are not usually electron-beam welded.

EBW of Heat-Resistant Alloys

Because of the marked differences in composition and weldability among heat-resistant alloys (nickel-base, iron-base, and cobalt-base), generalizations concerning EBW of these alloys are not useful.

Solid-Solution Nickel-Base Alloys. Hastelloy N, Hastelloy X, and Inconel 625 are readily electron-beam welded. Hastelloy B and Inconel 600 can be welded to type 304 stainless steel and to themselves.

Precipitation-hardenable nickel-base alloys that are rated good in weldability by the electron-beam process include Inconel 700, alloy 718, Inconel X-750, and René 41. Alloy 718 can be welded in either the annealed or the aged condition. Inconel X-750 should be welded in the annealed condition, and René 41 should be welded in the solution-treated condition. Alloys of this group that have fair weldability include casting alloys 713C and GMR-235 and wrought Udimet 700 and Waspaloy.

Iron-Nickel-Chromium Base Alloys. This group of alloys is also known as the iron-base heat-resistant alloys. The most readily welded alloy of this group, using EBW, is 19-9 DL, which has excellent weldability and which produces the best results when preheating is used. Alloy N-155 has good weldability, and alloys 16-25-6 and A-286 are rated fair. Alloy A-286 is usually welded in the solution-treated condition; hot cracking may result if welded in the aged condition.

Cobalt-Base Alloys. HS-21 has good weldability in unrestrained joints (and is generally poor in restrained joints). Cast alloy H-31 (X-40) has fair-to-good weldability, and alloy S-816 has fair weldability.

EBW of Refractory Metals

Electron-beam welding is the preferred technique for welding of refractory metals and their alloys. These materials, with melting temperatures in excess of 2200 °C (4000 °F), can be processed more efficiently with the intense heat source of the electron beam, which allows for localized fusion without excessive heating of the adjacent base metal. Weld-zone properties in refractory metals and alloys are particularly sensitive to contamination by interstitial elements such as carbon, hydrogen, nitrogen, and oxygen, which increase the already high ductile-to-brittle transition temperature of the weld zone and increase the notch sensitivity beyond the limits required for practical processing of these materials.

The possibility of atmospheric contamination during welding is essentially eliminated by EBW in a high-vacuum environment. Interstitial impurities already present in the base metal, however, have a significant effect on the weldability of the metal and alloy system. The Group V metals (e.g., niobium and tantalum) are softer, less subject to strain hardening, and less prone to the brittle behavior usually associated with the Group VI metals (molybdenum and tungsten). The very low room-temperature ductility of the Group VI metals gives rise to cracking problems due to cyclic thermal stresses during welding. Preheating above 300 °C (575 °F) reduces the susceptibility to this problem.

Excessive grain growth in the weld HAZ and fusion zones can produce cracking during the welding process or during subsequent application of the weldment. Grain coarsening can be controlled to some extent by specifying a fine-grained starting material and minimizing the total heat input during welding. The morphology of the fusion zone is also important. High welding speeds (>610 mm/min, or 24 in./min) selected to minimize heat input can promote pronounced directionality of grain growth in the weld pool, leading to a plane of weakness at the weld centerline. High-frequency (>1 kHz) beam oscillation can be applied to restore an elliptical weld-pool geometry at high travel speeds and eliminate the pronounced delineation of the weld centerline.

As with any low-ductility weldment, the degree of restraint imposed on the joint during the welding operation should be minimized. Most refractory metals and alloys can be welded in thin (<1.0 mm, or 0.04 in.) sheet. The weldability decreases with increasing section thickness in accordance with the increase in restraint. Weld joints should be designed to reduce the restraint by effective use of flange butt and standing edge designs, and by the incorporation of relief grooves to accommodate elastic/plastic strain in the material adjacent to the weld zone.

In contrast to many other alloy systems, the weldability of refractory metals tends to improve with the addition of alloying elements. Specific alloying elements can improve the low-temperature ductility of refractory metals and thus directly affect the weldability and weld mechanical properties. The overall weldability of tungsten is improved by the addition of substitutional alloying elements such as rhenium and molybdenum. As indicated previously, interstitials also raise the transition temperature; this effect can be reduced by alloying with elements that form stable carbides, oxides, and nitrides. For example, titanium and zirconium in molybdenum "scavenge" the interstitial impurities to improve low-temperature ductility and provide a dispersion-strengthening effect. Examples of successful electron-beam welds made in refractory metals and alloys are given in Table 5.

TABLE 5 TYPICAL APPLICATIONS AND CONDITIONS FOR EBW OF REFRACTORY METALS AND ALLOYS

APPLICATION	**BEAM SETUP**					COMMENTS
	KV	**MA**	**mm/min**	**in./min**	**FOCUS**	
TUNGSTEN HEAT PIPE, FLANGE-BUTT JOINT, 2.54 MM (0.100 IN.) PENETRATION	100	6	760	30	SURFACE	**PREHEAT TO 800 °C (1475 °F)**
TUNGSTEN CRUCIBLES, STANDING EDGE JOINT, 2.03 MM (0.080 IN.)	100	10	1525	60	ABOVE SURFACE	**PREHEAT WITH BEAM TO 800 °C (1475 °F)**
MOLYBDENUM HEAT PIPE, STEP JOINT/PARTIAL-PENETRATION WELD TO ROOT OF STEP, 1.57 MM (0.062 IN.) PENETRATION	125	25	760	30	SURFACE	**HIGH-FREQUENCY (~3 KHZ) DEFLECTION TO PRODUCE ELLIPTICAL POOL**
MO-TZM ALLOY STRAIGHT BUTT WELD IN HEAT SHIELD, 1.52 MM (0.060 IN.) PENETRATION	125	15	760	30	ABOVE SURFACE	**DUCTILITY OF WELD ZONE IMPROVED BY POSTWELD HEAT TREATMENT**
NB-1ZR REACTOR TUBING, BUTT WELD 1.52 MM (0.060 IN.) PENETRATION	100	10	760	30	. . .	**FULL-PENETRATION WELD, EXCELLENT AS-WELDED PROPERTIES**
MO-13 WT% RE TEST REACTOR TUBE 0.81 MM (0.032 IN.) PENETRATION	**100**	**7**	**760**	**30**	**SHARP**	**EXCELLENT WELD APPEARANCE, NO TEST DATA AVAILABLE**

Tungsten is the most difficult refractory metal to weld because of its high melting point (3410 °C, or 6170 °F), sensitivity to thermal shock, and room-temperature brittleness of welds and the HAZ. High-purity tungsten sheet can be electron-beam welded if the weld area is preheated and if external restraint is minimal. Successful welds have been made in sections as thick as 4.1 mm (0.16 in.); however, the resulting as-welded material is brittle and highly notch sensitive. Tungsten metal, manufactured by a chemical vapor deposition process, has also been proven to be weldable in sections <0.5 mm (<0.02 in.).

Molybdenum can be readily welded by EBW if chemical purity, grain size, and texture of the base metal are carefully monitored. Partly because of its lower melting point (2610 °C, or 4730 °F), molybdenum has better thermal shock resistance than tungsten; thus, narrower welds and higher welding speeds can be used. Preheating may or may not be necessary, depending on the starting material and the degree of restraint on the weld joint. When preheating is used, the workpiece should be heated to 205 °C (400 °F) or higher, followed by a postweld stress relief at 870 to 980 °C (1600 to 1800 °F).

Figure 31 shows an electron-beam weld zone in high-purity molybdenum. Penetration is approximately 3.05 mm (0.12 in.) to the root of a step-butt weld joint. This weld was made without preheat but with a high-frequency oscillation to maintain an elliptical weld pool and influence the direction of grain growth in the solidifying metal. The growth direction at the center of the fusion zone is parallel to the welding direction, and the section is therefore an end view of the columnar grains. This type of solidification structure should not be confused with the grain refinement phenomena sometimes claimed to result from stirring of the weld pool. Because solidification in high-purity metals is essentially a columnar-planar growth, refinement most probably cannot be achieved by stirring the weld pool.

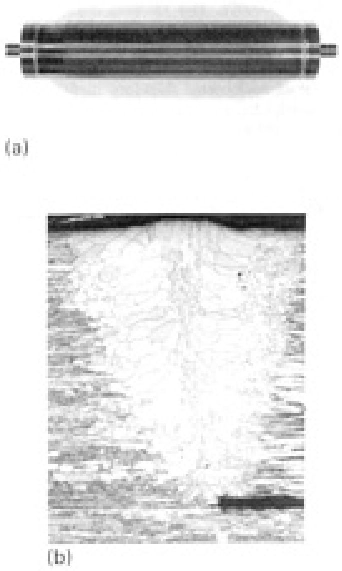

FIG. 31 ELECTRON-BEAM WELD ZONE IN HIGH-PURITY MOLYBDENUM. (A) MOLYBDENUM HEAT PIPE TEST ASSEMBLY SHOWING END CAPS JOINED BY EBW. (B) PHOTOMICROGRAPH SHOWING MORPHOLOGY OF WELD FUSION ZONE. 25×

Titanium and zirconium are the common alloying elements added to molybdenum to scavenge interstitial impurities and improve low-temperature ductility. The transition temperatures of welds in these alloys, however, are still above room temperature. Significant increases in the low-temperature ductility of these alloys can be obtained by solution and aging heat treatments. The greatest improvements in ductility are obtained by solution and aging heat treatments. The greatest improvements in ductility are obtained by the addition of rhenium; a 50 wt% addition of rhenium lowers the transition temperature of recrystallized molybdenum from approximately 52 to 205 °C (125 to 400 °F).

Niobium, which melts at 2468 °C (4474 °F), is easier to weld than molybdenum or tungsten. Gas-tungsten arc welding is often used, but EBW provides better protection, narrower welds, and lower heat input. Preheating is not necessary, but postweld vacuum stress relief is used to restore ductility and toughness, especially in the niobium alloys.

Unalloyed niobium is easily electron-beam welded, but has relatively low strength. More alloys have been developed, mainly for higher strength, than for the other refractory metals. Alloying usually reduces weldability, but most weld joints in alloys retain about 75% efficiency, and good structural properties can be obtained for service temperatures in the range of about 1095 to 1650 °C (2000 to 3000 °F), depending on the alloy. Chill tooling is often used, but copper, nickel, and stainless steel should be avoided to prevent contamination of the weld metal.

Tantalum is the most easily weldable of the four refractory metals. Because of its high melting point (2996 °C, or 5425 °F) and good thermal conductivity, thicknesses of 1.52 mm (0.060 in.) or more must be rapidly heated to high temperatures for welding. Copper chills are used to avoid distortion and weld sagging and to shorten the time the assembly has to remain in the vacuum chamber, thus limiting grain growth.

Weldability of tantalum alloys containing other refractory elements is somewhat lower than that of unalloyed titanium; however, successful welds have been made in alloys Ta-10W, T-111 (Ta-8.0W-2.5Hf-0.003C), and T-222 (Ta-9.64W-2.4Hf-0.01C). Because of high vapor pressure, alloys containing vanadium are better welded by the gas-tungsten arc process.

EBW of Aluminum Alloys

Single-pass, full-penetration welds with depth-to-width ratios of more than 20 to 1 have been made in 150 mm (6 in.) thick aluminum alloy plates. The low heat input, narrow HAZ, and short heat cycle result in minimum decrease in mechanical properties, especially in assemblies of some of the heat-treatable alloys.

Non-Heat-Treatable Alloys. The mechanical properties of electron-beam welds in the non-heat-treatable alloys (1*xxx*, 3*xxx,* and 5*xxx* series) are virtually the same as those obtained by GTAW. If the alloy is in a strain-hardened condition, the weld properties closely approach the annealed properties of the base metal.

Heat-treatable alloys of the 2*xxx,* 6*xxx,* and 7*xxx* series are crack sensitive to varying degrees when welded. Some may also be prone to porosity. The condition of the zone immediately adjacent to the weld is critical in determining weldability.

The effect on penetration of varying working distance (within practical operating limits and with other variables held constant) is shown in Fig. 32 for 6.4 mm ($\frac{1}{4}$ in.) thick aluminum alloy 2219. When maximum working distance was used, penetration was only about 80%, and maximum weld width was about one-third greater than with minimum working distance.

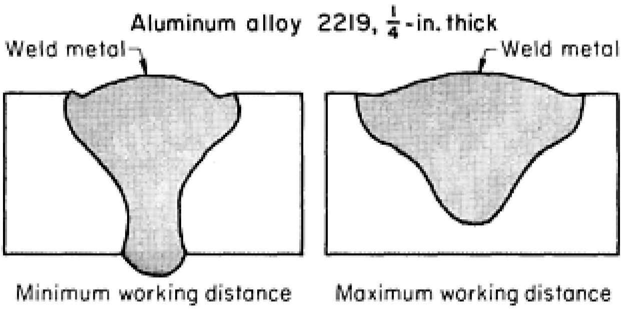
Aluminum alloy 2219, $\frac{1}{4}$-in. thick
Weld metal
Weld metal
Minimum working distance
Maximum working distance

FIG. 32 VARIATION IN PENETRATION OF 6.5 MM ($\frac{1}{4}$ IN.) THICK 2219 ALUMINUM ALLOY AS A FUNCTION OF WORKING DISTANCE WHEN USING NONVACUUM EBW

Typical shapes of full-penetration welds made in 5.69 mm (0.224 in.) thick aluminum alloy 2219 by nonvacuum EBW (EBW-NV) and high-vacuum EBW (EWB-HV) and by GTAW are shown in Fig. 33.

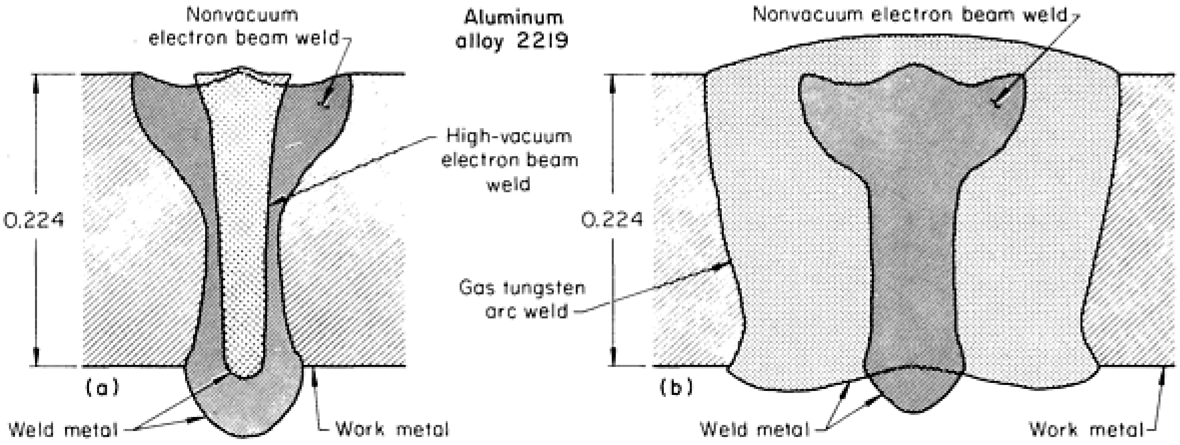
Nonvacuum electron beam weld
Aluminum alloy 2219
Nonvacuum electron beam weld
High-vacuum electron beam weld
0.224
0.224
Gas tungsten arc weld
(a)
(b)
Weld metal
Work metal
Weld metal
Work metal

WELDING CONDITIONS	**EBW-NV VERSUS EBW-HV**		EBW-NV VERSUS GTAW	
	HIGH-VACUUM EBW	**NONVACUUM EBW**	**GTAW**	NONVACUUM EBW
VOLTAGE, KV	150	140	13V	**140**
CURRENT, MA	20	30	270A	**20**
WELDING SPEED, MM/MIN (IN./MIN)	2540 (100)	2540 (100)	510 (20)	**510 (50)**
WELDING ENERGY, KJ/MM (KJ/IN.)	**0.071 (1.8)**	**0.098 (2.5)**	**0.41 (10.5)**	**0.13 (3.36)**

FIG. 33 COMPARISON OF FULL-PENETRATION WELDS OBTAINED BY SELECTED WELDING PROCESSES TO JOIN 2219 ALUMINUM ALLOY. (A) EBW-NV VERSUS EBW-HV. (B) EBW-NV VERSUS GTAW

Comparing the nonvacuum and high-vacuum electron-beam welds in Fig. 33(a), heat input for the nonvacuum weld was about 40% greater, producing a midpoint weld width about 33% greater and a maximum crown width 160% greater than those of the high-vacuum weld. Comparing the nonvacuum electron-beam weld to the gas-tungsten arc weld in Fig. 33(b), heat input for the typical nonvacuum weld was 32% of that for the arc weld, producing a midpoint weld width about 25% and a maximum weld width about 56% of the corresponding widths for the arc weld. Compared to the arc weld, the widest weld produced by EBW-NV under workable operating conditions was about 42% as wide at midpoint and about 70% as wide at the crown.

Studies on welding 6.4 mm ($\frac{1}{4}$ in.) thick aluminum alloy 2219 showed average shrinkage of 0.08 to 0.13 mm (0.003 to 0.005 in.) for high-vacuum welds and shrinkage of 0.36 mm (0.014 in.) for high-heat-input nonvacuum welds, compared with a normal shrinkage of 0.76 mm (0.030 in.) for gas-tungsten arc welds.

The 6*xxx* series alloys are only slightly affected by the heat cycles of EBW. Alloys 2219, 7039, and 7005 appear to be the least affected alloys of their respective series. The techniques used to prevent cracking in EBW of the heat-treatable aluminum alloys are:

- SPECIAL CARE IN PREWELD CLEANING TO AVOID POROSITY AND OXIDE INCLUSIONS IN THE WELD METAL
- PRESTRESSING JOINTS IN COMPRESSION BY USING INTERFERENCE FITS WHEN POSSIBLE
- USE OF A DUCTILE FILLER METAL, USUALLY IN THE FORM OF A THIN STRIP PREPLACED IN THE JOINT, WHICH WILL YIELD UNDER SHRINKAGE STRESS AND FILL JOINT OPENINGS
- SELECTION OF BEAM POWER, BEAM SPOT SIZE, AND WELDING SPEED TO CREATE AS NARROW WELDS AS PRACTICAL DURING SHORT WELDING HEAT CYCLES, THUS AVOIDING EXCESSIVE GRAIN-BOUNDARY MELTING
- USE OF POSTWELD HEAT TREATMENT TO RESTORE THE STRENGTH AND DUCTILITY OF THE WELD AND HAZ
- DESIGNING OR LOCATING JOINTS TO BE FREE OF EXTERNALLY IMPOSED RESTRAINTS, REINFORCING JOINT AREAS, AND, WHEN POSSIBLE, WELDING IN THE SOLUTION-TREATED CONDITION, WITH POSTWELD AGING

Some aluminum alloy welded joints (for example, 38.1 ram, or 1.5 in., thick, welded 7075-T651 plate) exhibit lower mechanical properties than unwelded plate due to overaging in the HAZ. Postweld solution and aging heat treatments improve joint properties. At high travel speeds, porosity can develop in the weld from vaporization of certain elements, such as zinc, in the alloy. Aluminum alloy 7075, which contains 5.6% Zn, is an example of this. Travel speed should be reduced in EBW of high-zinc aluminum alloys so that the vapor has time to escape to the surface before the weld metal solidifies. Zinc-free alloys (the 6*xxx* series alloys, for example) can be welded at higher speeds without the threat of porosity.

EBW of Copper and Copper Alloys

The EBW of copper and its alloys is influenced by the same factors that affect the arc welding of these metals (see the article "Welding of Copper Alloys" in this Volume). The high thermal conductivity of copper causes less difficulty in electron-beam welding than in arc welding.

Molten metal may be expelled from the weld joint during EBW of nondeoxidized coppers (especially alloy C11000, tough pitch copper), causing spatter and uneven weld surfaces. This can usually be remedied by the use of a cosmetic pass. The vacuum environment eliminates possible hydrogen embrittlement; nevertheless, root voids and porosity still can occur.

The presence of low-melting elements ordinarily makes the welding of free-machining copper alloys impractical, and the volatility of zinc prevents the welding of the brasses and other zinc-containing copper alloys. The remaining zinc-free copper alloys can generally be electron-beam welded without any unusual problems.

EBW of Magnesium Alloys

Electron-beam welding is used to a limited extent, chiefly for repair, on commercial wrought and cast magnesium alloys that contain less than 1% Zn. The relative suitability of alloys for EBW is generally the same as for arc welding, as discussed in the article "Welding of Magnesium Alloys" in this Volume.

Techniques. Special techniques and close control of operating variables are needed to prevent voids and porosity at the root of the weld, because of the high vapor pressure of magnesium, which has the lowest boiling point (1107 °C, or 2025 °F) of any commonly welded metal. This difficulty is aggravated by the presence of zinc, which has an even lower boiling point (906 °C, or 1663 °F). It is usually impractical to electron-beam weld magnesium alloys that contain more than 1% Zn.

Circular oscillation of the beam or the use of a slightly defocused beam is helpful in obtaining sound welds. The most satisfactory technique is to use integral or tightly fitted backing of the same alloy and to ensure that welding conditions are held to values that trial welds have shown will either minimize porosity and voids or localize them in an area where they can be tolerated or removed by machining.

EBW of Titanium Alloys

All of the commercial alloys of titanium that can be joined by arc welding can also be joined by EBW. Their relative weldability and response to heat cycling in EBW are generally the same as in arc welding, which is discussed in the article "Welding of Titanium Alloys" in this Volume.

Applicability. The vacuum environment of EBW prevents exposure to the atmospheric contaminants that cause embrittlement of titanium alloys, whereas arc welding processes must use elaborate and costly shielding methods to accomplish this. Cost studies show that direct labor costs for EBW of titanium sections more than 25 mm (1 in.) thick are less than for arc welding, provided a suitably large vacuum chamber is available.

Techniques. Filler metal is not ordinarily used, and the work is not preheated. Tack welding, contrary to experience in GTAW, presents no difficulties in EBW. For optimum results, welding is done in a high vacuum, but medium-vacuum welding is satisfactory for many applications.

Alloy Ti-6Al-4V, the alloy most frequently used in assemblies to be welded, can be electron-beam welded in either the annealed or the solution-treated-and-aged condition. For weldments that will be used at elevated temperatures, a preferred process sequence is anneal, weld, solution treat, and age. For other service conditions, a process sequence of solution treat, age, and weld gives almost the same strength properties and only slightly lower fracture toughness.

Procedure Development and Practice Considerations for Electron-Beam Welding

EBW of Beryllium

Beryllium is well suited for EBW. Narrow fusion zone and low heat input help control cracking. Also, the vacuum atmosphere controls toxic beryllium fumes and prevents oxidation. Weldability primarily is affected by the chemical composition of the metal. Beryllium oxide is the major contaminant in powder source material. Lower beryllium oxide content has been found to improve ductility, as well as weldability (Ref 3). Generally, the beryllium oxide content is limited to 1.5 wt% maximum. Lower aluminum content also is reported to improve weldability by decreasing crack susceptibility. The aluminum tends to produce a hot short condition. This hot short condition is reduced if the aluminum is tied up in a compound, such as $AlFeBe_4$, rather than existing as free aluminum in the grain boundaries (Ref 4).

Autogenous Welds. Cracking can be minimized in autogenous welds (no filler metal added) by reducing joint restraint during cooling. This is accomplished by appropriate fixturing, proper joint design, thin members, and good part fit-up. Most welding must be accomplished at sharp focus to minimize the size of the fusion zone. For deep welds, preheat has been found to improve weldability. For example, welds 1.3 mm (0.050 in.) thick have been made without preheat in low-oxide powder source beryllium (Ref 5). The same metal can be welded up to 2.5 mm (0.1 in.) thick with a preheat of approximately 345 °C (650 °F). In addition to cracking, high oxide levels can cause porosity and excessive cratering in the weld. Typical parameters for autogenous welds are given in Table 6.

TABLE 6 EFFECT OF PREHEAT ON THE WELDABILITY OF BERYLLIUM AND CONDITIONS FOR FULL-PENETRATION WELDS

THICKNESS		VOLTAGE, KV	CURRENT, MA	TRAVEL SPEED		PREHEAT TEMPERATURE TO PREVENT CRACKING	
MM	IN.			MM/MIN	IN./MIN	°C	°F
2.54	0.100	110	7.5	510	20	500	**500**
1.90	0.075	100	7.0	510	20	500	**500**
1.40	**0.055**	**90**	**5.0**	**510**	**20**	**120**	**250**

Braze Welds. The greatest success in EBW of beryllium is obtained by using a ductile filler metal in the weld; generally, aluminum is used. Optimum properties are obtained when the fusion-zone microstructure consists of beryllium dendrites surrounded by a ductile aluminum matrix. A fusion zone with less than 30% Al contains large beryllium dendrites which are susceptible to microcracking (Ref 4). Welds can be made with less than 30% Al in the fusion zone, but caution must be used.

Filler metal may be added by several methods. One method is to preplace aluminum shim stock in the joint. Care must be taken to ensure the shim stays in the joint during the weld. This may be accomplished by tack welding the joint prior to the penetration pass. By using a 0.51 mm (0.020 in.) thick by 2.54 mm (0.100 in.) deep 1100 aluminum shim, welds to 3.05 mm (0.120 in.) penetration have been made without preheat. A typical weld is shown in Fig. 34.

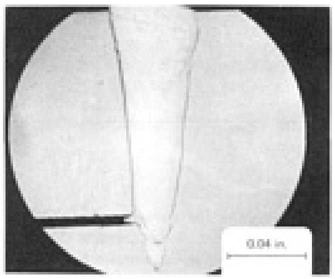

FIG. 34 ELECTRON-BEAM BRAZE WELD IN BERYLLIUM. COURTESY OF ROCKWELL INTERNATIONAL ENERGY SYSTEMS GROUP

An alternative method to brazing welds is to use an electron-beam cold wire-feed process similar to that used for stainless steel (Ref 6). In this process, a narrow, deep groove in the beryllium is filled with aluminum wire fed into the weld pool. A typical joint uses 718 aluminum wire, 0.76 mm (0.030 in.) in diameter, fed into a modified J-groove. The weld joint has a 0.64 mm (0.025 in.) root radius with a 10° 20' included angle. Welds to 2.54 mm (0.100 in.) have been successfully made this way (Fig. 35).

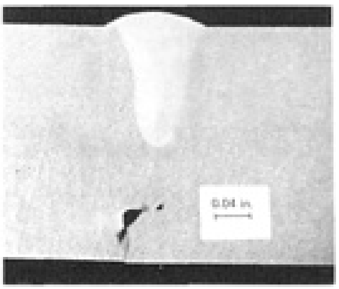

FIG. 35 ELECTRON-BEAM COLD WIRE-FEED WELD IN BERYLLIUM. COURTESY OF ROCKWELL INTERNATIONAL ENERGY SYSTEMS GROUP

References cited in this section

3. D. HAUSER, H.W. MISHLER, R.E. MONROE, AND D.C. MARTIN, *WELD. J.,* DEC 1967, P 525S-540S
4. C.R. HEIPLE AND R.D. DIXON, *WELD. MET. FABR.,* JUNE 1979, P 309-316
5. R.P. CAMPBELL, R.D. DIXON, AND A.L. LIBY, "ELECTRON BEAM FUSION WELDING OF BERYLLIUM," RFP-2621, ROCKWELL INTERNATIONAL, ROCKY FLATS PLANT, GOLDEN, CO, 1978
6. F.K. BENCH AND G.W. ELLISTON, *WELD. J.,* DEC 1974, P 763-766

Procedure Development and Practice Considerations for Electron-Beam Welding

References

1. A. SANDERSON AND K.R. NIGHTINGALE, *WELD. J.,* APRIL 1990, P 45
2. J.F. LOWRY, J.H. FINK, AND B.W. SCHUMACHER, A MAJOR ADVANCE IN HIGH-POWER ELECTRON BEAM WELDING IN AIR, *J. APPL. PHYS.,* VOL 47, 1976, P 95-106
3. D. HAUSER, H.W. MISHLER, R.E. MONROE, AND D.C. MARTIN, *WELD. J.,* DEC 1967, P 525S-540S
4. C.R. HEIPLE AND R.D. DIXON, *WELD. MET. FABR.,* JUNE 1979, P 309-316
5. R.P. CAMPBELL, R.D. DIXON, AND A.L. LIBY, "ELECTRON BEAM FUSION WELDING OF BERYLLIUM," RFP-2621, ROCKWELL INTERNATIONAL, ROCKY FLATS PLANT, GOLDEN, CO, 1978
6. F.K. BENCH AND G.W. ELLISTON, *WELD. J.,* DEC 1974, P 763-766

Procedure Development and Practice Considerations for Laser-Beam Welding

J. Mazumder, University of Illinois at Urbana-Champaign

Introduction

LASER-BEAM WELDING (LBW) is a joining process that produces coalescence of material with the heat obtained from the application of a concentrated coherent light beam impinging upon the surface to be welded. The principles of laser beams and the LBW process are discussed in the article "Laser-Beam Welding" in this Volume. In the present article, the steps that must be considered when selecting the LBW process will be described. Included will be a review of the individual process variables that influence procedure development of the process. In addition, joint design and special practices related to LBW will be discussed.

Procedure Development and Practice Considerations for Laser-Beam Welding

J. Mazumder, University of Illinois at Urbana-Champaign

Process Selection

Laser-beam welding is characterized by its low distortion and low specific energy input. It is an accurate inertialess method capable of high welding speeds for most materials, including many difficult-to-join materials. The process is often selected for high-volume, high-speed applications for joining materials/components that pose problems during arc welding, such as titanium alloys used for pacemakers. Often the selection criteria depends on economics, which is related to joining rate, welding speed, distortion, and postwelding treatment needs, among many other considerations. Another important consideration is whether the higher speeds can offset the higher initial capital costs within a reasonable time period. Low specific energy input is responsible for applications such as pace-makers because it reduces the possibility of thermal damage of the sophisticated electronics contained in such devices. Some manufacturers of cigarette lighters are now using LBW as an alternative to resistance welding because of the lower porosity in laser welds.

Economics combined with the smaller HAZ and higher overall weld toughness associated with LBW makes it a popular joining method for high-volume automotive applications. For example, multiple pieces of different compositions are joined by LBW to form blanks for automobile body components such as side-frame panels. These blanks need considerable weld toughness to withstand the stresses induced during subsequent press forming operations. Weld geometry is another consideration for the selection of the process. Since a smaller volume of material is melted due to the low specific energy input, weld defects such as undercuts are also minimized. This is especially important for applications involving dissimilar workpiece thicknesses, which have a higher probability of undercuts.

Another example where the narrow HAZ combined with high welding speed makes LBW a preferred method is the welding of planetary gears used in automobile powertrains. Planetary gear teeth are so closely spaced that it is difficult to access the weld joint without melting the neighboring gear teeth using conventional (arc) welding. This is one of the oldest LBW applications and it is routinely carried out worldwide.

Metallurgical considerations are sometimes responsible for the selection of the LBW process. The characteristic low specific energy input for this process results in a higher cooling rate. Higher cooling rates are often needed to suppress precipitation of harmful intermetallic compounds (which may promote brittle fracture) during the solidification of the weld pool. This is an important criteria for welding of many superalloys. For titanium alloys, high cooling rate limits the grain growth and oxidation during solidification, making it possible to weld titanium alloys outside a vacuum chamber with modest inert gas shielding. High cooling rates also help to avoid sensitization during welding of stainless steels. On the other hand, high cooling rates may cause problems in many ferrous alloys by promoting martensitic transformation and higher residual stresses; however, this is often balanced by the smaller size of the HAZ.

In many applications where resistance spot welding has been traditionally used, laser welding can offer many advantages, including single-side access and reductions in weight and costs by removing the tab[*] needed in resistance welding. Up to 60 kg (130 lb) of steel can be saved from an automobile by replacing resistance spot welding with laser welding and taking advantage of the design flexibility (Ref 1).

Another key factor in the successful implementation of laser welding is the selection of the right laser for a given application. Characteristics and parameters of commercial lasers available for welding are listed in Table 2 of the article

"Laser-Beam Welding" in this Volume. Carbon dioxide (CO_2) lasers with up to 25 kW power capacities and pulsed Nd:YAG lasers with up to 500 W peak power have been the workhorse for the industry until recently. The majority of the CO_2 laser applications involve lasers up to 5 kW. Recently, continuous Nd:YAG lasers with power capacities up to 2.5 kW have been made available.

In addition to welding speed and depth of penetration, which are the most important LBW attributes, the flexibility and precision of beam delivery are also important criteria for selecting LBW. Multiaxis workstations and robotics are already available in the market for CO_2 laser applications. However, one of the major advantages of the Nd:YAG laser is its capability to deliver the energy through a fiber. Unfortunately, no fibers are yet available for CO_2 lasers with a 10.6 μm wavelength. Most of the high-speed and deep-penetration welding are carried out with CO_2 lasers with power in excess of 1 kW. Continuous Nd:YAG lasers can outperform CO_2 lasers with the same power capacity, but they are limited by the maximum power of 2.5 kW. For reflective materials such as aluminum, Nd:YAG has better coupling, especially for conduction-mode welding, due to their shorter wavelength (1.06 μm). For deep-penetration welding, the coupling advantage is not significant.

Reference cited in this section

1. F.A. DIPIETRO, ROBOTIC LASER WELDING SYSTEMS IN THE AUTOMOTIVE INDUSTRY, *PROCEEDINGS OF LASER SYSTEMS APPLICATIONS IN INDUSTRY,* ATA, TORINO, ITALY, 1990, P 103-119

Note cited in this section

* *TABS CONSIST OF EXTRA MATERIAL PROVIDED DURING THE FABRICATION OF THE SQUARE FRAME IN RESISTANCE WELDING THAT ENABLE THE ELECTRODES TO PUSH THE EXCESS MATERIAL TOGETHER TO COMPLETE THE WELDING PROCESS. ADAPTING JOINT DESIGNS SUCH AS BUTT WELDS TO THE LASER WELDING PROCESS ELIMINATES THE NEED FOR TABS.

Procedure Development and Practice Considerations for Laser-Beam Welding

J. Mazumder, University of Illinois at Urbana-Champaign

Procedure Development

The key factor for procedure development is the selection of optimum independent and dependent process variables. The independent process variables for laser welding include incident laser-beam power, incident laser-beam diameter, traverse speed, absorptivity, shielding gas, depth of focus and focal position, and weld design and gap size. The important dependent variables are depth of penetration, microstructure and mechanical properties of laser-welded joints, and weld pool geometry. Detailed discussions of most of these variables are provided in Ref 2. They are discussed below in view of their role in procedure development.

Power density is defined as the incident power per unit area. In any welding process it is the power density that determines the depth of penetration and joining rate for the process, not the total power of the source. Therefore, beam diameter and spatial distribution of the laser-beam energy play an extremely important role because they determine the area of incidence. Two other variables affecting the ultimate power density driving the welding process are absorptivity and plasma-beam interaction, such as refraction. Position of focus with respect to the substrate and depth of focus of the beam also affects the power density and penetration depth.

Spatial distribution of the energy is generally constant for a given laser whereas absorptivity is a function of the substrate, its surface condition, and the laser-beam wavelength. Therefore, the most common practice for process development with a given laser and material is the selection of focusing optics for optimum power density and working distance by controlling the beam diameter. Thus, a working knowledge of the focusing optics is important for an engineer developing the laser welding process. Important formulae related to the laser as a welding tool will be summarized in following subsections dealing with process variables.

Interaction time is defined as the time a particular point on the substrate spends under the laser beam. In simple terms, it is the ratio of the beam diameter to the traverse speed. Strictly on the basis of heat transfer, one should account for the total time of heating and cooling at the point above the ambient temperature. This is, however, a much more difficult quantity to keep track of compared to the ratio of the traverse speed and beam diameter. Again, it is the interaction time which controls the weld pool profile, not the traverse speed alone. Laser welding procedure development involves the selection of optimum power density and interaction time after the selection of the laser.

Laser-Beam Power. As mentioned in the article "Laser-Beam Welding" in this Volume, laser-beam power is one of the primary variables in determining the penetration depth in laser welding. For process development, the first caution one should take is to calibrate the laser output power with the power delivered at the substrate. There is often a measurable loss due to the reflective and transmissive optics used to deliver the beam from the laser to the workpiece. For systems with fiber delivery, input coupling is often responsible for part of this loss. Fibers also scramble the spatial distribution resulting in poor focusing properties. Almost all commercial lasers are equipped with some kind of power-measuring device based either on a flowing-water calorimeter where the full beam is dumped, or a solid-state detector where a fraction of the laser energy is continuously deposited. Additional calorimetric devices should be maintained to measure power at the workpiece for periodic calibration of output power with the power delivered at the weld pool. For lasers with power in excess of a few hundred watts, water-cooled calorimeters are the most popular. For on-line monitoring many multi-kilowatt systems also use precalibrated solid-state detectors, usually behind one of the cavity mirrors which has a small leak around 1% or less. Several vendors offer both types of power meters. An excellent source of laser-beam device vendors is *Laser Focus World* buyer's guide, which is published annually (Ref 3).

In addition to the absolute power measurement, the beam mode (that is, the spatial distribution of energy) and beam diameter should be monitored since they also strongly affect the power density.

Laser-beam diameter is one of the most important variables since it determines the power density for a given total power (see the article "Laser-Beam Welding" for a more detailed discussion). The spatial distribution of energy in the beam affects the focusing characteristics of the beam. For a Gaussian beam, the diameter is defined as the $1/e^2$ of the central value (Fig. 1) and it contains almost 86% of the total power. The diffraction limited focal spot size, d_{min}, for such a Gaussian beam is given by the following relationship:

$$d_{mim} = \frac{1.27 f l}{D} \qquad \textbf{(EQ 1)}$$

where D is unfocused beam diameter, f is the focal length of the focusing optics, and λ is the wavelength of the laser beam. This is an ideal situation and often the focused beam diameter will be larger due to aberrations and other imperfections in the focusing optics.

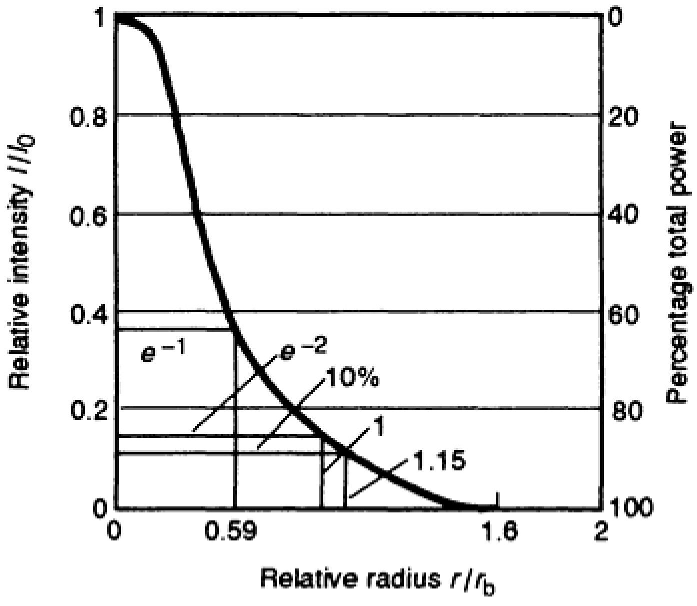

FIG. 1 VARIATION OF RELATIVE INTENSITY AND PERCENTAGE OF TOTAL POWER WITH RADIUS FOR A GAUSSIAN BEAM. SOURCE: REF 4

Diffraction limited spot size for various types of laser beam can be estimated using the relationships given in Eq 2, 3, 4. For a rectangular beam with a plane wave front:

$$d_{mim} = \frac{2 f l}{D} \qquad \textbf{(EQ 2)}$$

For a circular beam with a plane wave front:

$$d_{mim} = \frac{2.44 f l}{D} \qquad \textbf{(EQ 3)}$$

For a multimode beam (or a transverse electromagnetic mode (TEM$_{mnq}$) as described in the following subsection on laser-beam spatial distribution) with a plane wave front:

$$d_{mim} = \frac{2.44 f l}{D}(2m+n+1) \qquad \textbf{(EQ 4)}$$

For example, if a 25 mm (1 in.) TEM$_{01}$CO$_2$ laser beam is focused by a 100 mm (4 in.) focal-length lens, the diffraction limited spot diameter is:

$$d_{mim} = \frac{2.44x100x10.6x10^{-3}}{25}(2.0+1+1) = 0.21mm$$

However, as indicated earlier, these are estimates for the theoretically possible minimum spot size. One of the more popular means for evaluation of beam quality is the recently developed "M^2 concept" (Ref 5), which is defined as:

$$M = \frac{\text{divergence of actual beam}}{\text{divergence of ideal gaussian Beam}} = \frac{q\,Actual}{q\,gaussian}$$

The correlated beam diameter based on the M^2 concept is:

$$d_{mim} = \frac{4M^2 f l}{p D} \qquad \textbf{(EQ 5)}$$

In this case, M has to be measured at two known locations in the beam propagation path. This reinforces to the point that a proper measurement technique for beam diameter is imperative for estimating the actual beam diameter and the power density.

There are several commercial systems now available for beam diameter measurement. Sasnet, *et al.* (Ref 5) describe one system which is specifically designed for beam evaluation based on the M^2 concept. This instrument is capable of measuring the beam width within 0.5 μm accuracy in two mutually perpendicular transverse directions. Knife edges are used to scan the beam and a set of lenses are traversed in the direction of the beam propagation to change the positions of the beam width. The measurement procedure and data acquisition and display are computer controlled.

Another commercial system is based on a hollow rotating needle which samples the beam. This is based on the research carried out by Loosen, *et al.* (Ref 6) and Lim and Steen (Ref 7) which resulted in the development of a laser-beam analysis system based on a rotating needle which reflects part of the beam to a set of detectors.

Although beam quality related to the ideal Gaussian beam is a good starting point, there is evidence that for high-power laser processing, this characterization may not be sufficient (Ref 8). Spatial coherence and temporal and local fluctuations are generally ignored, but they do affect the processing behavior. Many European research groups are working on characterizing the beam in terms of the mode coherence coefficient (MCC) which is based on a set of suitable modes and averaged intensity distribution and which takes into account the partial coherence properties of the beam.

Another complication of high-power laser welding is interaction with the plasma column. Rockstroh, *et al.* (Ref 9) found that free electrons in the plasma column cause absorption and refractions. For welding of aluminum with a CO_2 laser, as much as 30% of the energy is lost in the plasma and the refracted beam diameter is 8% larger. This measurement is based on the electron density measurement using emission spectroscopic technique. This observation is subsequently supported by similar measurements made by Lober, *et al.* (Ref 10, 11). Complete characterization of beam diameter will require both measurement at the focal spot before processing and monitoring of the beam refraction during the process using spectroscopic techniques.

Laser-Beam Spatial Distribution. Every laser resonator cavity, defined by the mirrors of the laser, has certain stable configurations of the electromagnetic field called "modes." A well-defined mode has a definite spatial distribution for the laser beam. As evident from Eq 4, the mode affects the focused beam diameter and leads to higher spot size for a given $f\lambda$ focusing optics. The general convention for indexing a mode is TEM_{mn} (transverse electromagnetic mode with m number of radial zero fields and n number of angular zero fields). TEM_{00} is the ideal Gaussian beam and will provide the smallest focused spot and highest power density. Most of the fast axial-flow lasers offer a TEM_{00} or TEM_{01} beam. Most of the transverse-flow lasers offer TEM_{01} (donut-shaped spatial distribution).

Intensity distribution for different modes can be calculated using the Hermite polynomial for rectangular coordinates and the LaGuerre polynomial for polar coordinates using the expressions for a TEM_{mn} mode beam given in Eq 6, 7, 8. For a rectangular coordinate:

$$\sqrt{\mathrm{I}_{mn}(x,y)} = \{H_m(\sqrt{2.x/R_b})H_n(\sqrt{2.y/R_b}).\exp[-(x^2+y^2)/R_b^2]^2\} \qquad \textbf{(EQ 6)}$$

For a polar coordinate:

$$\mathrm{I}_{mn}(r,q) = [L_m^n(2r^2/R_p^2)]^2(\sqrt{2.r/R_o})^{2n}.\exp[-2r^2/R_b^2]\cos^2 nq \qquad \textbf{(EQ 7)}$$

For a Gaussian beam with TEM_{00} mode:

$$1(\lambda) = \mathrm{EXP}(-2R^2/R_b^2) \qquad \textbf{(EQ 8)}$$

In Eq 6, 7, and 8, x is a variable, r is the radial position, 0 is the angular position of the beam, R_b is the beam radius at the $1/e^2$ point, H_m is the Hermite polynomial of order m, and L_m^n is the generalized LaGuerre polynomial.

A schematic of intensity distributions for different beam modes is shown in Fig. 2. For process development, the lowest-order mode available is the best choice for achieving the highest power density from a given power.

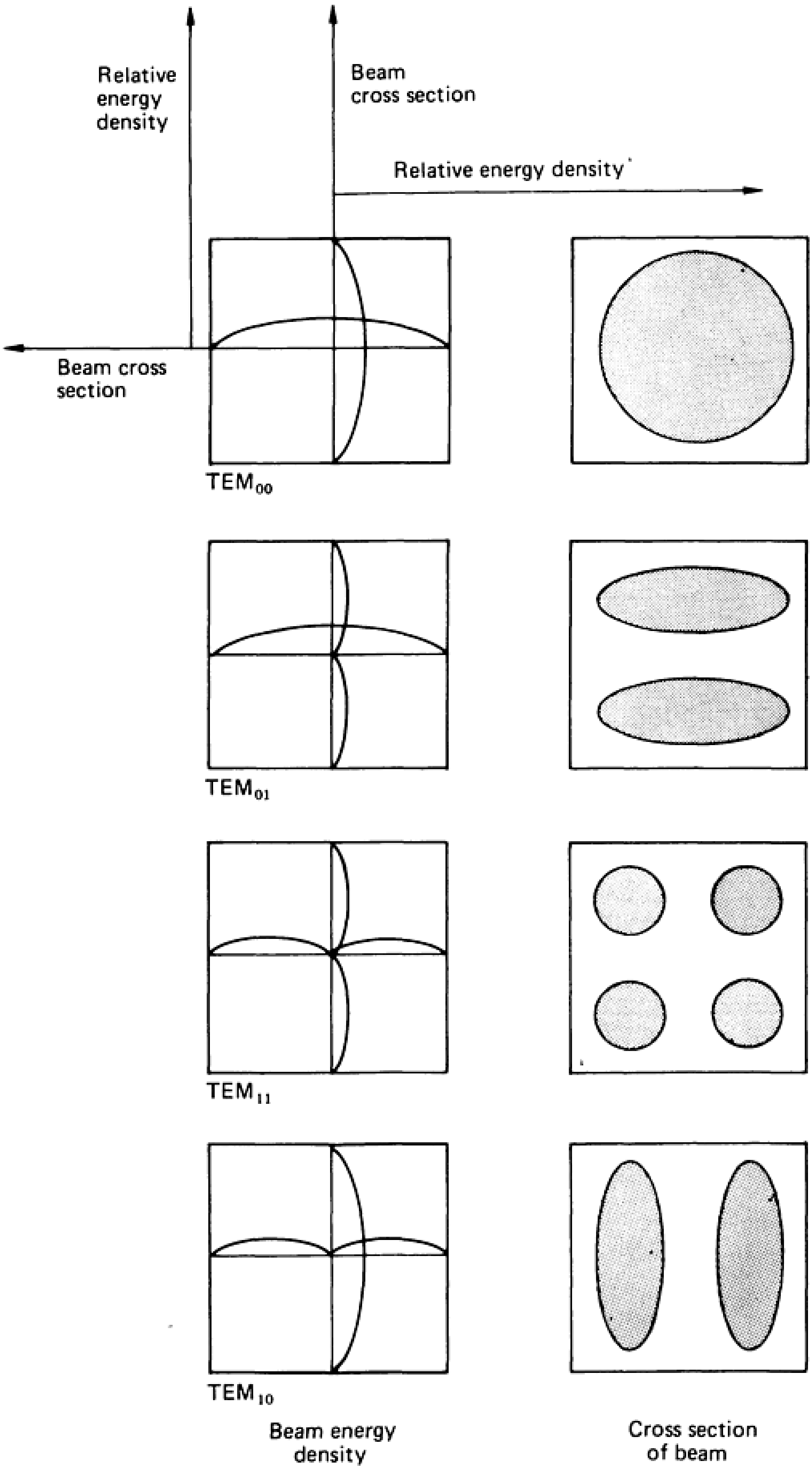
Relative energy density
Beam cross section
Relative energy density
Beam cross section
TEM$_{00}$
TEM$_{01}$
TEM$_{11}$
TEM$_{10}$
Beam energy density
Cross section of beam

FIG. 2 BEAM CROSS SECTIONS FOR FOUR DIFFERENT TEM MODES. SOURCE: REF 12

The position of focus with respect to the substrate surface influences the weld pool profile and penetration during laser welding. The optimum position for the focal point of the laser beam with respect to the substrate was investigated by Wilgoss, *et al.* (Ref 13). Figure 3 shows the transverse profiles generated by moving the focus point in 2.5 mm (0.1 in.) steps perpendicular to the plane of the workpiece. The reported plate thickness for this study was 6 mm (0.25 in.). When the focal point was positioned deep inside the workpiece, a V-shaped weld resulted necessitating more precise alignment than that for a parallel-sided weld of the same cross section. When the beam was focused well above the level of the workpiece surface, a large "nail head" with a consequent loss of penetration was observed.

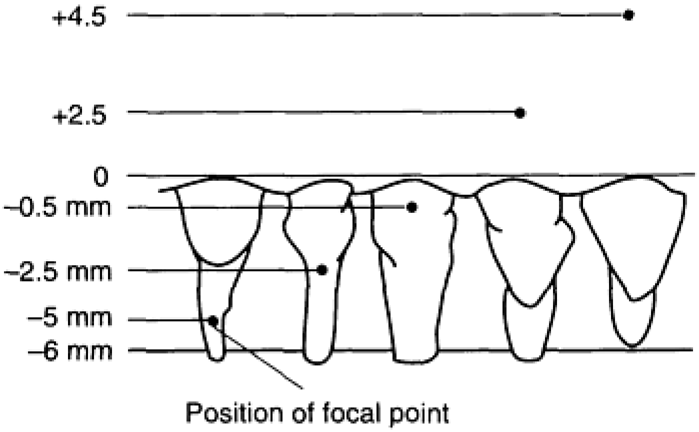

FIG. 3 TRANSVERSE PROFILES AS A FUNCTION OF FOCUS POSITION FOR A LASER-BEAM WELDED TYPE 310 STAINLESS STEEL. NEGATIVE AND POSITIVE NUMBERS INDICATE POSITION OF FOCAL POINT BELOW AND ABOVE, RESPECTIVELY, SURFACE OF PLATE. BEAM POWER, 5 KW. TRAVERSE WELDING SPEED, 16 MM/S (38 IN./MIN). SOURCE: REF 13

Wilgoss, *et al.* (Ref 13), concluded that the optimum focus is approximately 1 mm (0.04 in.) below the level of the workpiece surface. This produced welds with little or no "nail head" and nearly parallel sides.

Engel (Ref 14) reported that for several metals the optimum position of the focus is 1.25 mm to 2.5 mm (0.05 to 0.1 in.) below the surface. Research work by Alexander (Ref 15) using a 10 kW laser and 1.9 mm ($\frac{3}{4}$ in.) thick AISI 1020 steel revealed that the optimum position of focus was 2 mm (0.08 in.) below the surface. When the beam was focused 4 mm (0.16 in.) below the surface, the depth of penetration was almost the same as that for the 2 mm (0.08 in.) focus depth. The general consensus is that the optimum position for the focal point is below the surface, but the exact distance is dependent on the thickness of the workpiece and the laser power used.

Depth of focus is extremely important for welding thin-section materials because for a shallow depth of focus, the substrate comes in and out of the focal position due to thermal distortion. Depth of focus is defined as the distance over which the focus beam has the same approximate intensity. According to the Laser Institute of America, depth of focus is defined as the range over which the focused spot radius is increased by 5%. Based on this definition the depth of focus (Z) is given by:

$$Z = \pm \frac{0.32p R_b^2}{l} \approx \pm \frac{R_b^2}{l} = \frac{2.33(F l)^2}{l} = 1.488 F^2 l \qquad \textbf{(EQ 9)}$$

where F equals $f\lambda$ of the optics.

One can see from the above expression that as the *F* number of the focusing optics increases, depth of focus also increases. However, this is also accompanied by an increased beam diameter and decreased power density. As a result, these variables must be optimized for each particular application.

A common practice in determining the focal position and depth of focus is to carry out a trial bead-on-plate weld on an inclined plate and use the position where the plasma seems to be the visually strongest. This is a quick, but not necessarily the most accurate approach, for establishing the focal position.

Absorptivity. The importance of absorptivity and its role in laser welding is discussed in the aforementioned article "Laser-Beam Welding." For deep-penetration welding with power densities exceeding 10^6 W/cm^2, keyhole formation leads to high absorption by trapping the beam inside the hole by total internal reflection. However, the absorption during the initial transient period is dependent on wavelength and surface conditions. There are several techniques to enhance the initial absorption (Ref 16). They include nonmetallic absorbing coatings, preheating, surface roughening, periodic surface structure or induced grating structure**, and the application of an electric field of the correct sign.

Traverse Speed. The role of traverse speed during laser welding is also discussed in the article "Laser-Beam Welding." Generally, welding speed increases with increasing laser power as shown in Fig. 4. For a given thickness and power, a range of speed can be used to make successful welds. However, the fusion zone size will increase with the decreased speed within that range. For process development, the nugget size or fusion zone size is often prescribed by the design engineer for strength considerations. The laser welding engineer needs to select an optimum combination of speed and power to deliver the required nugget size.

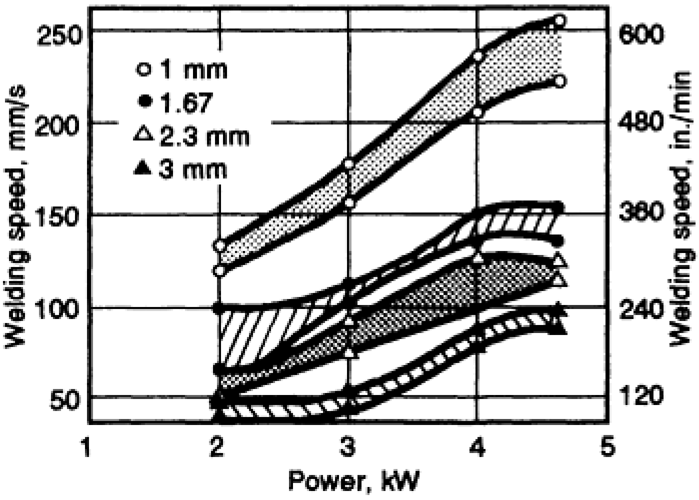

FIG. 4 WELDING SPEED VERSUS LASER POWER FOR A 5 KW CO_2 LASER. THE BASE METAL IS TI-6AL-4V. SOURCE: REF 17

Shielding Gas. The plasma produced during laser welding absorbs and scatters the laser beam. It is necessary, therefore, to remove or suppress plasma. The higher the power, the more clearly the phenomenon can be observed. An example of the way in which the depth of penetration decreases with the amount of beam absorption caused by plasma is shown in Fig. 5. Plasmas of this type can be removed by supplying a shielding gas such as helium. Shielding gas is also required to protect the weld surfaces from oxidation. Both the composition and flow rate of the shielding gas influence the depth of penetration (Ref 19, 20).

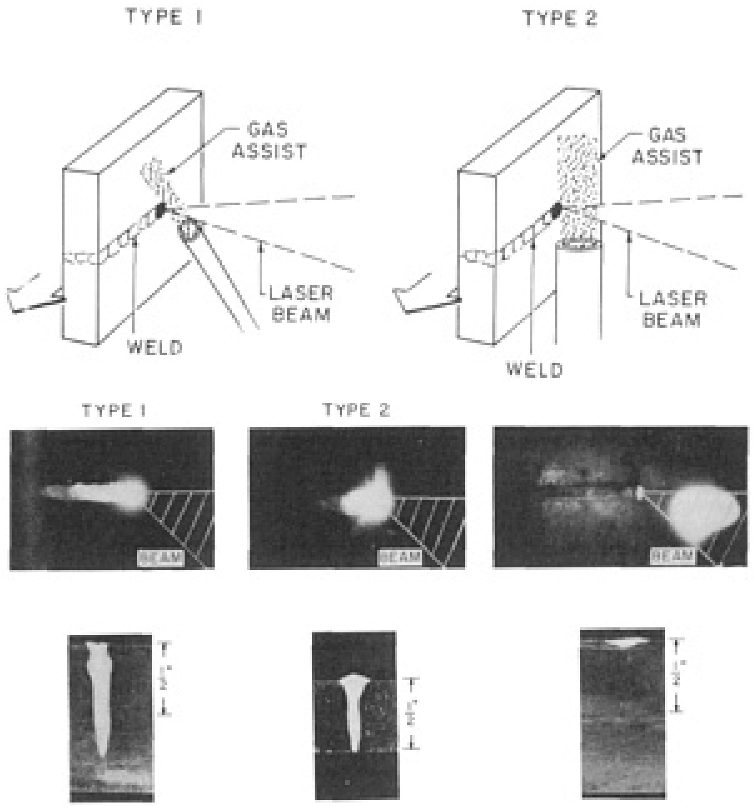

FIG. 5 EFFECT OF GAS ASSIST ON THE CROSS SECTION OF THE WELD. SOURCE: REF 18

Seaman (Ref 19) studied the role of shielding gas in high-power continuous-wave CO_2 laser welding. Weld cross sections made with various shielding gases and gas mixtures show a 60% difference in penetration. However, gases that permit the greatest penetration do not necessarily blanket (displace air from above the weld rapidly) the weld effectively at characteristically high laser welding speeds. This means that compromises are necessary to permit sound, deep-penetration laser welds (Ref 19).

The effect of the composition of the shielding gas on depth of penetration was studied by Seaman (Ref 19) and Rein, *et al.* (Ref 20). Generally, helium is used as the shielding gas for high-power laser welding. As shown in Fig. 6, helium seems to improve beam transmission whereas argon can cause severe beam blockage (Ref 19). This is probably due to the lower ionization potential of argon (15 eV) compared to that of helium (25 eV). The effects of air and CO_2 on beam transmission lie between the extremes represented by argon and helium (Fig. 7). Rein, *et al.* (Ref 20) have shown that the addition of small quantities of hydrogen, sulfur hexafluoride, and CO_2 to the helium enhances penetration at welding speeds below 40 mm/s (95 in./min).

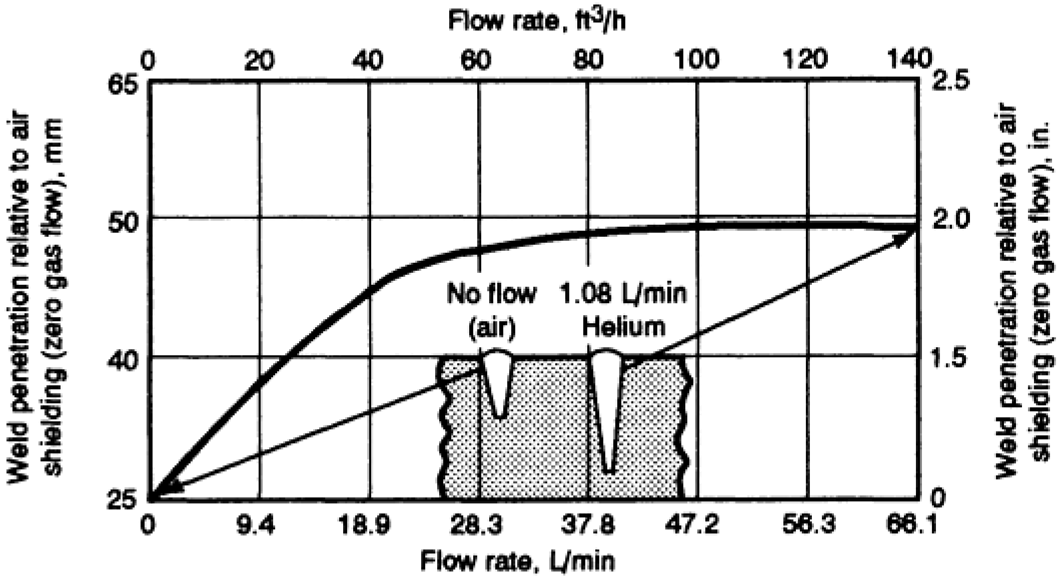

FIG. 6 EFFECT OF SHIELDING GAS ON DEPTH OF PENETRATION DURING LBW OF AN AUSTENITIC STAINLESS STEEL. LASER POWER, 15 KW. TRAVEL SPEED, 25 MM/S (60 IN./MIN). SOURCE: REF 19

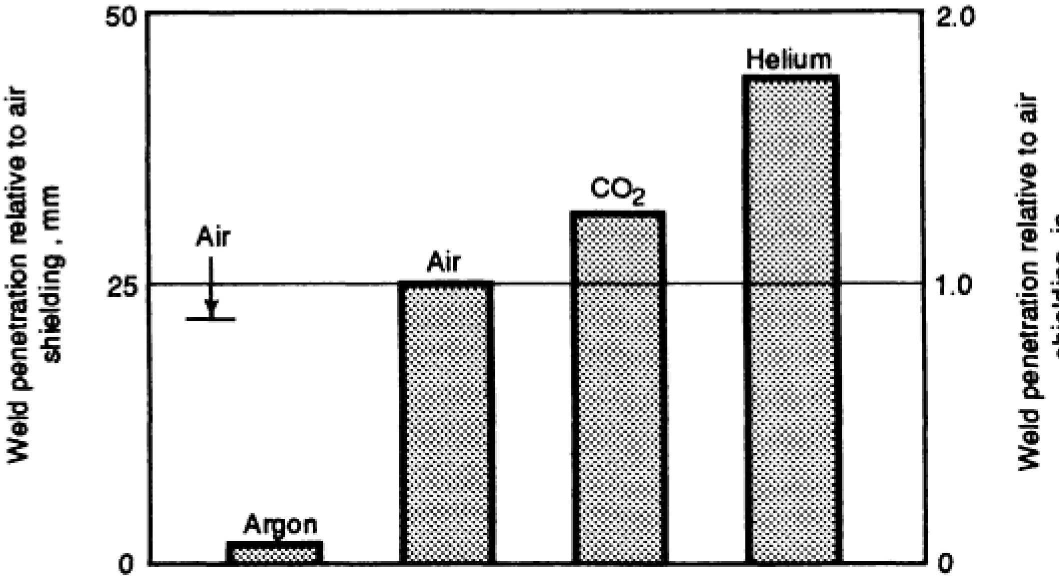

FIG. 7 COMPARISON OF LASER WELD PENETRATION ACHIEVED WITH DIFFERENT SHIELDING GASES. SOURCE: REF 9

The ionization potential of the shielding gas is not the only consideration for laser welding, especially at higher speeds. This is because gases with higher ionization potentials have lower atomic numbers and lower masses. These lighter gases are less effective in displacing air from the laser/material interaction area in the short time available in high-speed welding. Heavier gases are better able to displace air in a short time. Therefore, a mixture of heavier and lighter gases will result in optimum penetration. Figure 8 shows that as speed increases, the improvement resulting from the addition of a small amount of argon to helium (10% Ar and 90% He) becomes more noticeable.

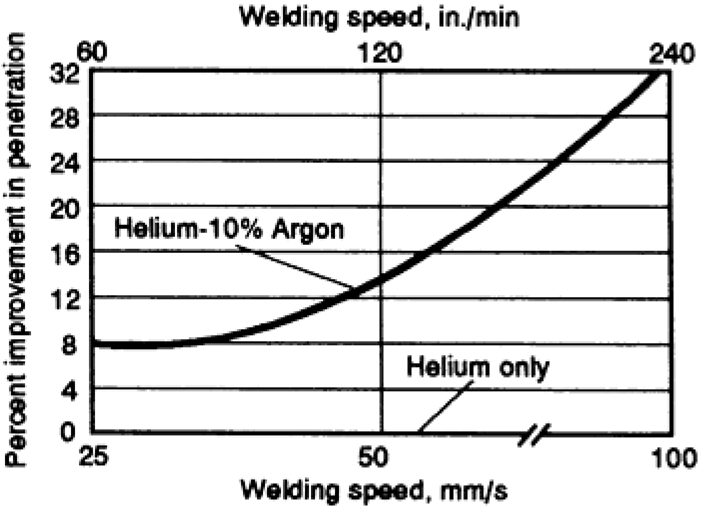

FIG. 8 IMPROVEMENTS IN WELD PENETRATION RESULTING FROM THE BLANKETING EFFECT OF A 10% ARGON ADDITION TO A HELIUM SHIELDING GAS. SOURCE: REF 19

For mass-scale production, the cost of shielding gas is also an important consideration. Recently, Chrysler Corporation replaced helium with CO_2 for plasma suppression during laser welding (Ref 21). The shielding gas is delivered from the side at an angle to suppress the plasma and remove metal vapor from the interaction point. For 14 kW laser power, a welding speed of 20 m/min (65 ft/min), and a shielding gas flow rate of 40 L/min (85 ft^3/h), CO_2 shielding resulted in a wider weld nugget size but mechanical properties were acceptable. The cost for helium gas is estimated to be $403,000 (3 shifts/day, 7 days a week) for a year compared to $80,600 for CO_2 gas; therefore, for process development, there is room for optimization of gas selection. However, one should remember that the use of CO_2 gas for a CO_2 laser may cause problems for the concentric gas shield due to preferential absorption.

Proper implementation of a shielding gas for plasma suppression can also be applied for joining difficult-to-weld materials with volatile alloying elements such as aluminum alloy 5083. Blake and Mazumder (Ref 22) demonstrated that with a properly designed helium suppression jet, magnesium losses and porosity formation could be minimized during laser welding of 5083 aluminum alloy. The resulting weld had a tensile strength equal to its parent material. This demonstrates that for process development, shielding gas can also be used to control the chemistry and weld defects.

References cited in this section

2. J. MAZUMDER, LASER WELDING, *LASER MATERIALS PROCESSING,* M. BASS, ED., NORTH HOLLAND, AMSTERDAM, 1983, P 113-200
3. *LASER FOCUS WORLD,* PENNWELL PUBLISHING COMPANY, TULSA, OK, 1993
4. J.E. HARRY, *INDUSTRIAL APPLICATION OF LASERS,* MCGRAW-HILL, UNITED KINGDOM, 1974
5. M.W. SASNETT AND T.F. JOHNSON, JR., LASER BEAM DIAGNOSTICS, *SPIE,* VOL 1414, 1991, P 21-32
6. P. LOOSEN, A. DRENKER, G. HERZIGER, R. KRAMER, AND U. STURM, DIAGNOSTICS OF HIGH POWER LASER BEAMS, *SPIE PROC.*, VOL 1024, 1988
7. G.C. LIM AND W.M. STEEN, *J. PHYS. E SCI. INSTRUM.,* VOL 17, 1984, P 999-1007
8. G. HERZIGER, LASER PROCESSING RESEARCH IN EUROPE, *LAMP '92,* A. MATSUNAWA AND S. KATAYAMA, ED., JAPAN HIGH TEMPERATURE SOCIETY, 1992, P 23-28
9. T. ROCKSTROH AND J. MAZUMDER, *J. APPL. PHYS.,* VOL 61 (NO. 3), 1987, P 917-923
10. R. LOBER, "SPECTROSCOPIC DIAGNOSTIC OF ARGON-ALUMINUM PLASMA DURING LASER-METAL INTERACTION," M.S. THESIS, UNIVERSITY OF ILLINOIS, 1989

11. R. LOBER AND J. MAZUMDER, SUBMITTED TO *J. APPL. PHYS.*, 1993
12. E.A. METZBOWER *ET AL.*, LASER BEAM WELDING, VOL 6, 9TH ED., *METALS HANDBOOK*, AMERICAN SOCIETY FOR METALS, 1984, P 647-671
13. R.A. WILGOSS, J.H.P.C. MEGAW, AND J.N. CLARK, *WELD. MET. FAB.*, MARCH 1979, P117
14. S.L. ENGEL, *LASER FOCUS*, FEB 1976, P 44
15. J. ALEXANDER, "PENETRATION STUDIES IN LASER AND ARC AUGMENTED LASER WELDING," PH.D. THESIS, LONDON UNIVERSITY, 1982
16. W.W. DULEY, "LASER SURFACE TREATMENT OF METALS," C.W. DRAPER AND P. MAZZOLDI, ED., NATO ASI SERIES E, NO. 115, MARTINUS NIJHOFT PUBLISHERS, THE NETHERLANDS, 1986
17. J. MAZUMDER AND W.M. STEEN, *MET. CONSTR.*, VOL 12 (NO. 9), 1980, P 423
18. E.V. LOCK, E. HOAG, AND R. HELLA, *WELD J.*, VOL 51, 1972, P 2455
19. F. SEAMAN, "ROLE OF SHIELDING GAS IN LASER WELDING," TECHNICAL PAPER NO. MR77-982, SOCIETY OF MANUFACTURING ENGINEERS, DEARBORN, MICHIGAN, 1977
20. R.M. REIN, *ET AL.*, PATENT 1448740, UNITED KINGDOM, 1972
21. M. OGLE, AUTOMOTIVE LASER APPLICATION WORKSHOP '93, INDUSTRIAL DEVELOPMENT DIVISION OF UNIVERSITY OF MICHIGAN, ANN ARBOR, MICHIGAN, 1993
22. A. BLAKE AND J. MAZUMDER, *ASME J. ENG. IND.*, VOL 107 (NO. 3), 1985, P 275-280

Note cited in this section

** *OPTICAL GRATINGS HAVE PERIODIC PEAKS AND VALLEYS, AND THEIR SPACING DETERMINES THE PREFERENTIAL ABSORPTION OF A CERTAIN WAVELENGTH OF LASER LIGHT. SIMILAR SURFACE PERIODIC STRUCTURE CAN BE USED TO ENHANCE ABSORPTION OF THE WAVELENGTH OF INTEREST.

Procedure Development and Practice Considerations for Laser-Beam Welding

J. Mazumder, University of Illinois at Urbana-Champaign

Joint Preparation, Fit-up, and Weld Design

Most weld joint geometries used in conventional fusion welding processes (for example, autogenous automatic gas-tungsten arc welding or electron-beam welding) are suitable for LBW. However, it must be remembered that the laser beam is focused to a spot of a few hundred micrometers in diameter and thus fit-up tolerances and alignment requirements are also of that order of magnitude.

Different types of joint design reported for laser-beam welding were reviewed by Shewell (Ref 23). Some of the important joint geometries are illustrated in Fig. 9 and 10.

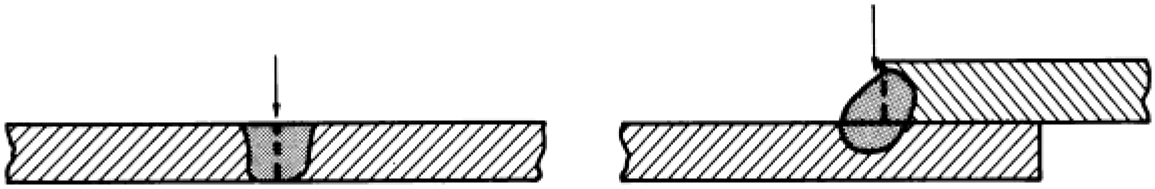

(a) Butt joint

(b) Lap joint

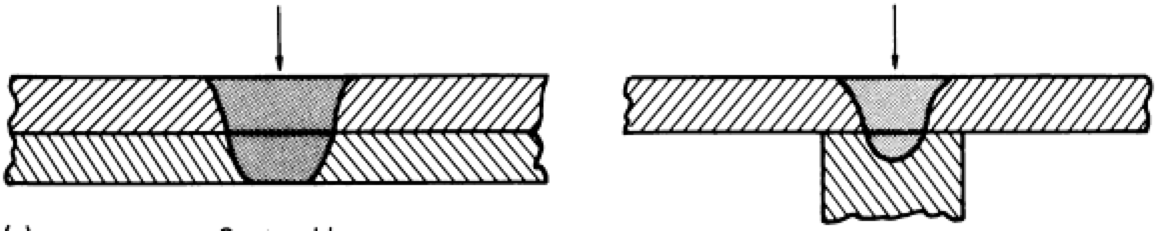

(c) Spot weld

(d) Spike weld or spot weld

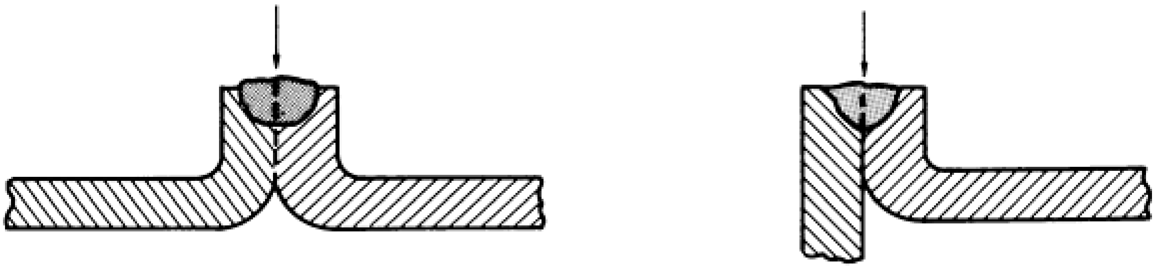

(e) Flange joint

(f) Edge joint

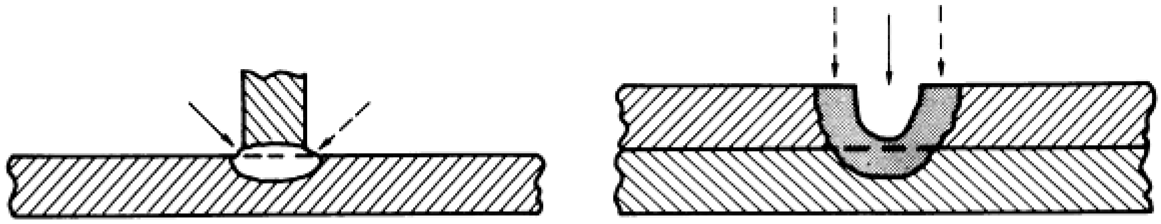

(g) T—joint

(h) Plug joint

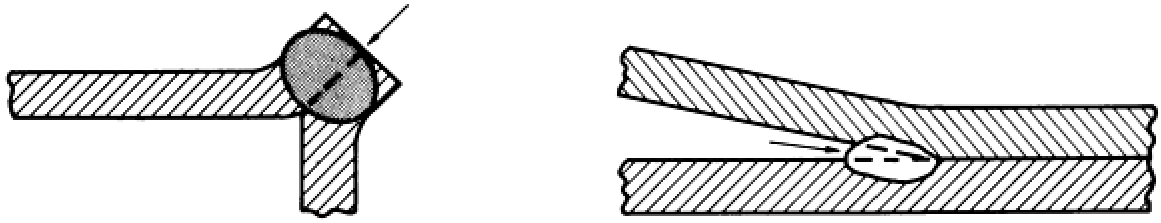

(j) Corner weld

(k) Kissing weld

FIG. 9 JOINT DESIGNS FOR LASER-BEAM WELDS ON SHEET METAL. ARROWS SHOW DIRECTION OF LASER BEAM. SOURCE: REF 23

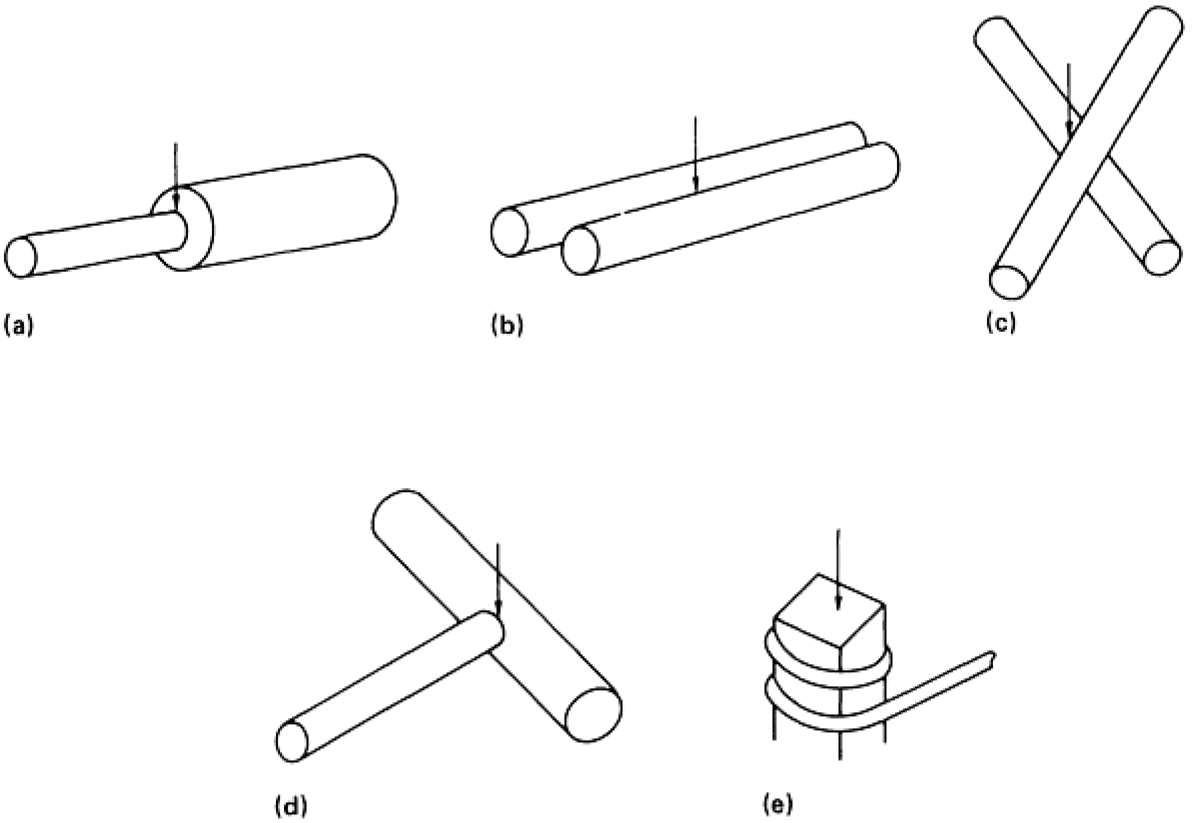

FIG. 10 JOINT DESIGNS FOR LASER-BEAM WELDS ON WIRE. ARROWS SHOW DIRECTION OF LASER BEAM. (A) BUTT WELD. (B) ROUND-TO-ROUND LAP WELD. (C) CROSS-JOINT WELD. (D) SPOT WELD FOR T-JOINT. (E) TERMINAL OR LUG WELD

Butt Joint--Full-Penetration Type. Material to be laser welded does not need beveled edges. Sheared edges are acceptable if square and straight (Ref 24, 25). The fit-up tolerance should be within 15% of the workpiece thickness. Misalignment and out-of-flatness should be less than 25% of material thickness, as shown in Fig. 11(a), to keep the laser beam from wandering. The transverse alignment should be kept within half of the focused beam diameter (Ref 17, 26). Clamping is recommended and buckling is not troublesome except when welding material thinner than 0.25 mm (0.01 in.) (Ref 24).

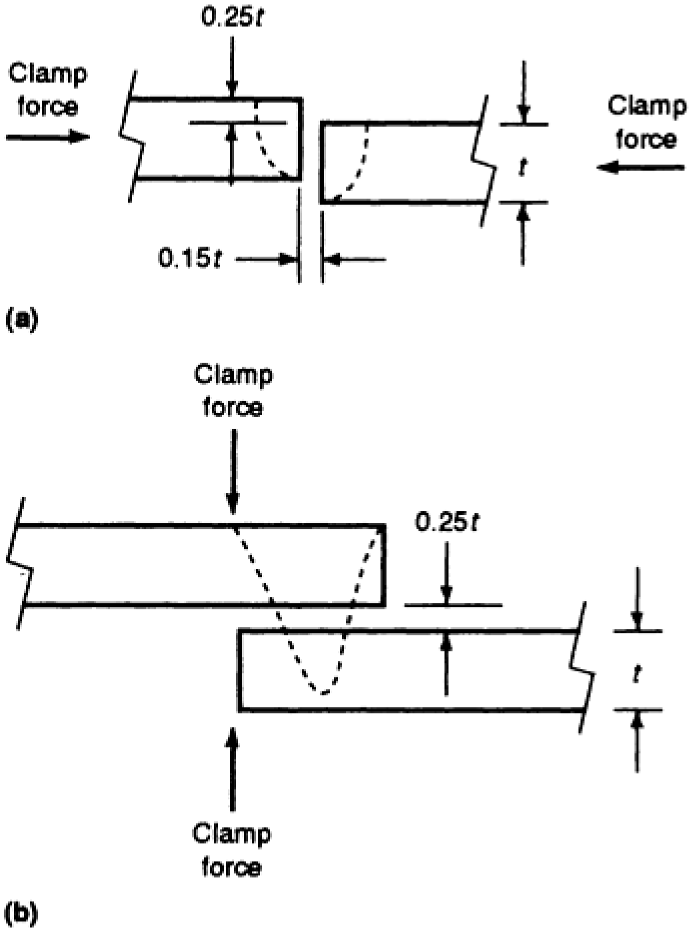

FIG. 11 CLAMPING AND TOLERANCES FOR A BUTT JOINT (A) AND A LAP JOINT (B). DIMENSIONS ARE IN UNITS OF PERCENT OF MATERIAL THICKNESS *T*. SOURCE: REF 25

Lap Joints. Air gaps severely limit penetration and welding speeds (Ref 24). Compressive clamping, as shown in Fig. 11(b), should be used to maintain a separation of less than 25% of the material thickness (Ref 25). For welding of dissimilar thicknesses, the thinner material should be placed on top to be welded to the thicker one (Ref 25).

Pulsed laser spot lap joints demand tighter tolerances than joints produced with continuous power laser welding. The maximum gap for ferrous and nickel alloys is 15% of the material thickness, and the gap must be less for higher conductivity materials (Ref 23). Baranov, *et al.* (Ref 27) studied the influence of gap size on the formation and strength of spot lap joints in laser welding. They reported for laser spot lap joints of 0.4 mm (0.016 in.) thick steel and 0.33 mm (0.013 in.) thick nickel, a gap of up to 0.1 mm (0.004 in.) has little influence on joint strength as there is almost no change in the diameter of the fusion zone. On the other hand, the shape of the weld pool changes with the gap size. It has also been observed that the presence of air in the gap may lead to the formation of porosity and nonfusion (Ref 27). Micropores sometimes form in the fusion zone and the number of pores and their size increase with increasing gap size.

Flange joints, as with butt joints, require straight, square edges, good fit-up, clamping, and precise transverse alignment. Flange joints are suitable for welding high-shrinkage metals such as aluminum (Ref 24).

Kissing Welds. This type of joint is called a kissing weld because the weld pool forms where two pieces just "kiss." This offers a small angle between the two parts that traps most of the energy of the laser beam. Very little (if any) pressure is required during welding, but the facing surfaces must fit well (Ref 23). A gap between the sheets will allow radiation to escape. This joint successfully joins thin foils in situations where spot welding would produce melt through.

Polarization Effect on Kissing Weld. One of the unique advantages of a kissing weld is the ability to take advantage of the polarization behavior of the laser light. If the laser light is polarized perpendicular to the wall of the metal, the beam is preferentially reflected and it delivers all the energy at the joint. This increases the joint efficiency considerably. This idea was originally proposed by Sepold in 1987 (Ref 28). Subsequently, this concept is used for laser welding of pipes.

Wire Joints. The joint configurations for wires shown in Fig. 10 were developed by the electronics industry. For wire-to-wire joints, the two wires must share the incident laser energy. In a cross joint, for example, the laser beam should be directed at the intersection of the wires so that both wires are exposed directly to the beam.

References cited in this section

17. J. MAZUMDER AND W.M. STEEN, *MET. CONSTR.,* VOL 12 (NO. 9), 1980, P 423
23. J.R. SHEWELL, *WELD DES. FABR.,* JUNE 1977, P 106
24. S.L. ENGEL, *WELD DES. FABR.,* JAN 1978, P 62
25. M.M. SCHWARTZ, *METALLURGICAL JOINING MANUAL,* MCGRAW-HILL, 1979
26. J. MAZUMDER, "LASER WELDING OF TITANIUM AND TINPLATE," PH.D. THESIS, LONDON UNIVERSITY, 1978
27. M.S. BARANOV, B.A. VERSHOK, I.N. GEINRIKHS, AND V.I. PRIVEZENSTEV, *WELD. PROD.,* VOL 23 (NO. 5), 1976, P 19
28. G. SEPOLD, R. ROTHE, AND K. TESKE, *LAMP '87,* Y. ARATA, ED., JAPAN HIGH TEMPERATURE SOCIETY, 1987, P 151-156

Procedure Development and Practice Considerations for Laser-Beam Welding

J. Mazumder, University of Illinois at Urbana-Champaign

Use of Consumables and Special Welding Practices

Laser welding is ideal for autogenous welding. The main consumable for laser welding is the shielding gas which was discussed earlier. For thick-section welding, a filler wire can be used to make multipass welds. Sometimes filler metals (or injected powders) are used during laser welding to control the weld bead geometry. The most common use of filler metals, however, is to prevent undercutting. The use of filler materials to control the weld chemistry is also common. For example, insertion of aluminum foil during butt welding of semikilled steel results in a porosity-free sound weld. Vaporization losses during laser welding of some alloys can be compensated for by the addition of filler materials or powder of desired composition. Energy consumption during LBW is lower with powder injection than with wire-fed filler metals. The wire surface reflects considerable amounts of energy and reduces the energy efficiency of the process.

As with arc welding processes, metallurgical considerations such as preweld and postweld heat treatments are also applicable to laser welding. The primary differences between LBW and other welding processes are the narrower HAZ associated with laser welds as well as the high cooling rates that result in higher localized residual stresses. These high cooling rates can be used beneficially for materials with brittle intermetallic phases. High cooling rate leads to extension of solid solubility and avoids brittle precipitates for many alloys.

Procedure Development and Practice Considerations for Laser-Beam Welding

J. Mazumder, University of Illinois at Urbana-Champaign

References

1. F.A. DIPIETRO, ROBOTIC LASER WELDING SYSTEMS IN THE AUTOMOTIVE INDUSTRY, *PROCEEDINGS OF LASER SYSTEMS APPLICATIONS IN INDUSTRY,* ATA, TORINO, ITALY, 1990, P

103-119

2. J. MAZUMDER, LASER WELDING, *LASER MATERIALS PROCESSING,* M. BASS, ED., NORTH HOLLAND, AMSTERDAM, 1983, P 113-200
3. *LASER FOCUS WORLD,* PENNWELL PUBLISHING COMPANY, TULSA, OK, 1993
4. J.E. HARRY, *INDUSTRIAL APPLICATION OF LASERS,* MCGRAW-HILL, UNITED KINGDOM, 1974
5. M.W. SASNETT AND T.F. JOHNSON, JR., LASER BEAM DIAGNOSTICS, *SPIE,* VOL 1414, 1991, P 21-32
6. P. LOOSEN, A. DRENKER, G. HERZIGER, R. KRAMER, AND U. STURM, DIAGNOSTICS OF HIGH POWER LASER BEAMS, *SPIE PROC.*, VOL 1024, 1988
7. G.C. LIM AND W.M. STEEN, *J. PHYS. E SCI. INSTRUM.,* VOL 17, 1984, P 999-1007
8. G. HERZIGER, LASER PROCESSING RESEARCH IN EUROPE, *LAMP '92,* A. MATSUNAWA AND S. KATAYAMA, ED., JAPAN HIGH TEMPERATURE SOCIETY, 1992, P 23-28
9. T. ROCKSTROH AND J. MAZUMDER, *J. APPL. PHYS.,* VOL 61 (NO. 3), 1987, P 917-923
10. R. LOBER, "SPECTROSCOPIC DIAGNOSTIC OF ARGON-ALUMINUM PLASMA DURING LASER-METAL INTERACTION," M.S. THESIS, UNIVERSITY OF ILLINOIS, 1989
11. R. LOBER AND J. MAZUMDER, SUBMITTED TO *J. APPL. PHYS.,* 1993
12. E.A. METZBOWER *ET AL.*, LASER BEAM WELDING, VOL 6, 9TH ED., *METALS HANDBOOK,* AMERICAN SOCIETY FOR METALS, 1984, P 647-671
13. R.A. WILGOSS, J.H.P.C. MEGAW, AND J.N. CLARK, *WELD. MET. FAB.,* MARCH 1979, P117
14. S.L. ENGEL, *LASER FOCUS,* FEB 1976, P 44
15. J. ALEXANDER, "PENETRATION STUDIES IN LASER AND ARC AUGMENTED LASER WELDING," PH.D. THESIS, LONDON UNIVERSITY, 1982
16. W.W. DULEY, "LASER SURFACE TREATMENT OF METALS," C.W. DRAPER AND P. MAZZOLDI, ED., NATO ASI SERIES E, NO. 115, MARTINUS NIJHOFT PUBLISHERS, THE NETHERLANDS, 1986
17. J. MAZUMDER AND W.M. STEEN, *MET. CONSTR.,* VOL 12 (NO. 9), 1980, P 423
18. E.V. LOCK, E. HOAG, AND R. HELLA, *WELD J.,* VOL 51, 1972, P 2455
19. F. SEAMAN, "ROLE OF SHIELDING GAS IN LASER WELDING," TECHNICAL PAPER NO. MR77-982, SOCIETY OF MANUFACTURING ENGINEERS, DEARBORN, MICHIGAN, 1977
20. R.M. REIN, *ET AL.,* PATENT 1448740, UNITED KINGDOM, 1972
21. M. OGLE, AUTOMOTIVE LASER APPLICATION WORKSHOP '93, INDUSTRIAL DEVELOPMENT DIVISION OF UNIVERSITY OF MICHIGAN, ANN ARBOR, MICHIGAN, 1993
22. A. BLAKE AND J. MAZUMDER, *ASME J. ENG. IND.,* VOL 107 (NO. 3), 1985, P 275-280
23. J.R. SHEWELL, *WELD DES. FABR.,* JUNE 1977, P 106
24. S.L. ENGEL, *WELD DES. FABR.,* JAN 1978, P 62
25. M.M. SCHWARTZ, *METALLURGICAL JOINING MANUAL,* MCGRAW-HILL, 1979
26. J. MAZUMDER, "LASER WELDING OF TITANIUM AND TINPLATE," PH.D. THESIS, LONDON UNIVERSITY, 1978
27. M.S. BARANOV, B.A. VERSHOK, I.N. GEINRIKHS, AND V.I. PRIVEZENSTEV, *WELD. PROD.,* VOL 23 (NO. 5), 1976, P 19
28. G. SEPOLD, R. ROTHE, AND K. TESKE, *LAMP '87,* Y. ARATA, ED., JAPAN HIGH TEMPERATURE SOCIETY, 1987, P 151-156

Procedure Development and Practice Considerations for Diffusion Welding

S.B. Dunkerton, The Welding Institute

Introduction

THE JOINING OF MATERIALS IN THE SOLID STATE can be an attractive alternative to fusion-welding processes. In fact, metals that are conventionally difficult to weld and combinations of dissimilar materials can only be welded using a solid-state process. The wide range of techniques that are available for solid-phase joining have various modes of heat and pressure application.

Diffusion welding represents one extreme within this range, in that it involves minimal pressurization, but relatively high temperatures and long periods of time. In its simplest form, the process involves holding premachined and cleaned parts in intimate contact, and then heating them in a protective atmosphere. As its name implies, diffusion has a major role in the joining of mating parts, although other mechanisms, such as microyielding and oxide dissolution, are also critical steps in successful welding.

The process of diffusion welding has been utilized for many years, although it did not attain industrial acceptance until the 1970s or 1980s, and then, only for specialized applications. A comprehensive, although somewhat dated, review of systems that have been successfully joined using diffusion welding is provided in Ref 1.

The specific advantages of the process include its ability to join most metals and some nonmetals in like or dissimilar combinations, the minimal amount of deformation that results, and the fact that the weld area is mostly independent of welding time. These advantages are offset by the need to machine paris to fairly tight tolerances on component fit-up, as well as on surface finish. However, the requirements that are generally specified can readily be achieved using standard machine shop practices.

Reference

1. W.A. OWCZARSKI AND D.F. PAULONIS, APPLICATION OF DIFFUSION WELDING IN THE USA, *WELD. J.,* FEB 1981, P 22-33

Procedure Development and Practice Considerations for Diffusion Welding

S.B. Dunkerton, The Welding Institute

Process Variants

Solid-Phase Process. The more-conventional form of diffusion welding utilizes uniaxial loading, which can be applied by dead weight, mechanical, pneumatic, or hydraulic means. The applied load is such that macrodeformation of the parent material(s) does not have to be reached. Because deformation also depends on welding temperature, temperatures that range from 50 to 75% of the melting point are usually employed. Heat can be applied by radiant, induction, and either direct or indirect resistance heating. Most welding operations are performed in vacuum or inert gas atmospheres, although welding in air has been reported (Ref 2). When welding noble metals, the requirements for a clean atmosphere are less stringent, but protection can prevent parent-metal oxidation while maintaining a clean weld interface.

An examination of the proposed sequence of stages in diffusion welding (Ref 3, 4) emphasizes the importance of the original surface finish. In forming a bond, it is necessary for the two metal surfaces to come into atomic contact (Fig. 1). Hence, microasperities and surface contaminants must be removed from the bonding faces. The applied load first causes the plastic deformation of surface asperities, thereby reducing interfacial voids. Bond development then continues by diffusion-controlled mechanisms, including grain-boundary diffusion and power-law creep.

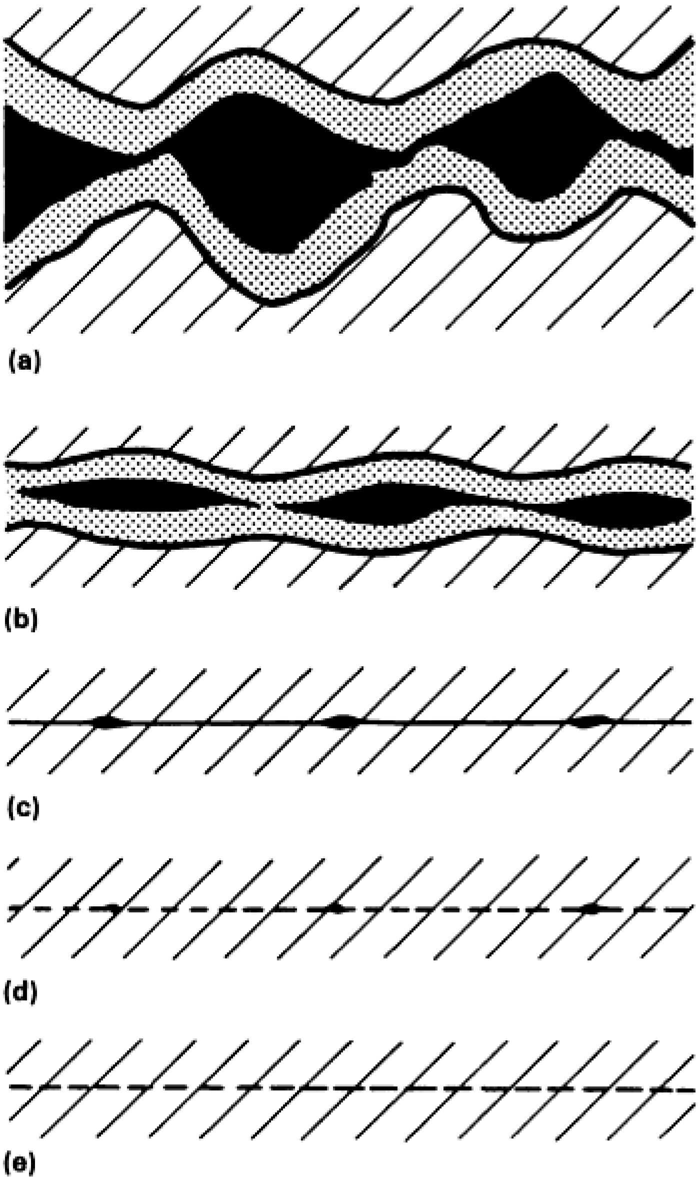

FIG. 1 MECHANISM OF DIFFUSION WELDING. (A) INITIAL "POINT" CONTACT AND OXIDE CONTAMINANT LAYER. (B) AFTER SOME "POINT" YIELDING AND CREEP, A THINNER OXIDE LAYER WITH LARGE VOIDS RESULTS. (C) AFTER FINAL YIELDING AND CREEP, SOME VOIDS REMAIN WITH VERY THIN OXIDE LAYER. (D) CONTINUED VACANCY DIFFUSION ELIMINATES OXIDE LAYER AND LEAVES FEW SMALL VOIDS. (E) COMPLETION OF BONDING

Premachined parts for uniaxial diffusion welding should have a surface finish of greater than 0.4 μm (16 μin.), roughness average (RA). Swabbing with acetone or petroleum ether, which should occur immediately prior to set-up in the welding machine, can suffice as a degreasing measure. Ultrasonic cleaning also has been used. Many researchers have examined the removal of oxide, prior to welding, by ion bombardment or other means. This can be applicable to certain metals if welding is to be carried out in the same vacuum chamber, or if the parts are in-vacuo coated with an oxide-resistant material. However, the additional cost and time involved in this procedure make it impractical for most industrial applications.

The liquid-phase process is applicable only to dissimilar material combinations or to like materials in situations where a dissimilar metal insert is used. Solid-state diffusional processes lead to the formation of a different phase at the weld interface. The liquid-phase welding temperature is based on the temperature at which this phase (usually, eutectic) melts. This thin layer of liquid spreads along the interface to form a joint at a lower temperature than the melting point of either of the parent metals. A reduction in welding temperature leads to the solidification of the melt. At the point where this phase is diffused into the parent metals, by holding at temperature, the process is known as transient liquid-phase diffusion welding. The optimal welding time and temperature can be estimated using the applicable phase diagram and solidification kinetics (Ref 5).

The superplastic forming/diffusion welding (SPF/DW) technique has been developed specifically within the aerospace industry. Its industrial importance is such that it is worthy of separate consideration. The process is used commercially for titanium and its alloys, specifically those that exhibit superplastic properties at elevated temperatures within defined strain-rate conditions (Ref 6, 7, 8). This process also has been successfully utilized for advanced aluminum alloys (Ref 9). Because the temperature and pressure conditions coincide with the conditions required for welding, the two processes have been combined into one manufacturing operation.

Hot Isostatic Pressing (HIP). HIP welding is a solid-phase process that involves the application of a high-temperature, high-pressure gas to components. Isostatic pressurization allows the welding of more-complex geometries than the uniaxial welding process can handle.

A hot isostatic press consists of a furnace within a gas pressure vessel. Its size can range from a diameter of 100 mm (4 in.) to a diameter of 1.1 m (44 in.) and a height of 2.2 m (88 in.). Argon gas is typically used, and there is a need to encapsulate the bond interfaces to prevent the gas from entering the site of the bond. This is most often achieved by locating the specimen in an evacuated and sealed steel can, although various other means can be used, as well.

The pressures associated with HIP welding are significantly higher than those used in uniaxial welding. However, because the pressure is isostatic, the level of deformation is extremely low. In addition, less premachining is required. Typical HIP cycles can last from 6 to 16 h, but because many components can be fitted into one furnace, the unit costs on large production runs can be low. Bryant (Ref 10) has developed a graphical procedure to determine the minimum pressure and temperature required for the HIP welding of several metals and alloys.

References cited in this section

2. O.S. SPANSWICK, DIFFUSION BONDING--A CASE STUDY, REPRINT R269/10/B4, *WELD. INST. RES. BULL.*, THE WELDING INSTITUTE, 1984
3. B. DERBY AND E.R. WALLACH, THEORETICAL MODEL FOR DIFFUSION BONDING, *MET. SCI.*, VOL 16, JAN 1982, P 49-56
4. B. DERBY AND E.R. WALLACH, DIFFUSION BONDING: DEVELOPMENT OF THEORETICAL MODEL, *MET. SCI.*, VOL 18, SEPT 1984, P 427-431
5. I. TUAH-POKU, M. DOLLAR, AND T.B. MASSALSKI, A STUDY OF THE TRANSIENT LIQUID PHASE BONDING PROCESS APPLIED TO AG/CU/AG SANDWICH JOINT, *METALL. TRANS. A*, VOL 19, MARCH 1988, P 675-686
6. J. PILLING, THE KINETICS OF ISOSTATIC DIFFUSION BONDING IN SUPERPLASTIC MATERIALS, *MATER. SCI. ENG.*, VOL 100, 1988, P 137-144
7. J. PILLING, D.W. LIVESEY, J.B. HAWKYARD, AND N. RIDLEY, SOLID STATE BONDING IN SUPERPLASTIC TI-6AL-4V, *MET. SCI.*, VOL 18, MARCH 1984, P 117-122
8. Y. MAEHARA, Y. KOMIZO, AND T.G. LANGDON, PRINCIPLES OF SUPERPLASTIC DIFFUSION BONDING, *MATER. SCI. TECHNOL.*, VOL 4, AUG 1988, P 669-674
9. J. PILLING AND N. RIDLEY, SOLID STATE BONDING OF SUPERPLASTIC AA7475, *MATER. SCI. TECHNOL.*, VOL 3, MAY 1987, P 353-359
10. W.A. BRYANT, A METHOD FOR SPECIFYING HOT ISOSTATIC PRESSURE WELDING PARAMETERS, *WELD. J.*, DEC 1975, P 433-S TO 435-S

Procedure Development and Practice Considerations for Diffusion Welding

S.B. Dunkerton, The Welding Institute

Carbon and Low-Alloy Steels

Although diffusion welding is applicable to this class of steels, industrial applications are limited. The process is used in situations where it is essential to avoid or minimize distortion and postweld machining, and for complicated geometries that involve internal features.

These steels are generally welded without the use of filler material (interlayer). The parts themselves are machined to a good surface finish (<0.4 μm, or 16 μin., RA surface roughness, and flat) and assembled in a heating unit such that force can be applied uniaxially or isostatically.

In terms of uniaxial diffusion welding, a wide range of bonding conditions have been shown to give sound joints, that is, temperatures of 900 to 1200 °C (1650 to 2190 °F), pressures of 5 to 14 MPa (0.7 to 2 ksi) and times of 5 to 120 min. Because the parameters are interdependent, as temperature increases, the load and/or time can be reduced and vice versa. Figure 2 depicts a typical cross section of a joint in 0.4% C steel, in which a layer of ferrite indicates the bond line of the sound bond that has formed.

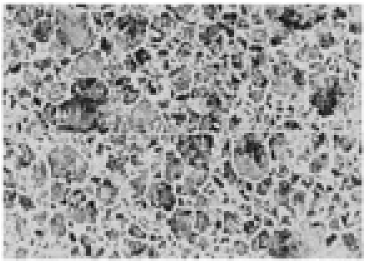

FIG. 2 CROSS SECTION OF DIFFUSION WELD IN 0.4% C STEEL. 50×

Tensile strength develops as temperature/pressure/time increase. Parent-metal tensile strength is readily achieved; failure occurs away from the joint interface, even when welding is not complete. Work has shown that only 80% bonding is necessary to achieve the equivalent parent-metal strength (Ref 11). However, ductility and joint-line toughness require a greater level of bond integrity.

Joints made in 0.4% C steel and a 1Cr-0.25Mo alloy steel have shown low Charpy impact values, when compared with the parent steels (Ref 12, 13). With the 0.4% C steel, it is possible to improve the toughness properties by using a postweld normalizing heat treatment (equivalent to the thermally cycled and normalized parent material). However, although the chromium-molybdenum steel had an improved microstructure after a postweld quench and temper heat treatment, it still had low joint-line toughness. This indicates the greater difficulty in forming a sound joint as alloy content increases.

Therefore, when diffusion welding carbon and low-alloy steels, the final application and service requirements should be considered carefully. The process is useful in low-stress applications, but when impact toughness is required, the fitness of a particular steel for the specific purpose needs to be examined. For higher-carbon-content steels and low-alloy steels, a suitable postweld heat treatment is recommended to return the steel to its original condition. However, it should be noted that some of the high temperatures referred to earlier ([ges]1100 °C, or 2010 °F) can cause significant grain coarsening.

The HIP welding of these steels also can be used when typical conditions involve a temperature of 1100 °C (2010 °F), a pressure of 100 MPa (14.5 ksi), and a bonding time of 4 h. Again, grain coarsening and the property considerations described above need to be reviewed.

References cited in this section

11. C.E. THORNTON AND E.R. WALLACH, IMPACT STRENGTH OF STEEL DIFFUSION BONDS, *MET. CONSTR.,* JULY 1985, P 450R-455R

12. S. ELLIOT, "PROPERTIES OF DIFFUSION BONDS IN A 0.4 C STEEL," REPORT 131/1980, THE WELDING INSTITUTE, 1980
13. S.B. DUNKERTON, "DIFFUSION BONDING OF A 0.4 CARBON CHROMIUM/MOLYBDENUM STEEL," REPORT 178/1982, THE WELDING INSTITUTE, 1982

Procedure Development and Practice Considerations for Diffusion Welding

S.B. Dunkerton, The Welding Institute

High-Strength Steels

Although diffusion welding has not been widely applied to high-strength steels, it is a feasible process, and guidelines similar to those identified for carbon and low-alloy steels will apply. Because diffusion welding relies on microplastic yielding in the early stages of bonding, it is necessary to use parameters at the higher end of the ranges specified previously, that is, temperatures above 1000 °C (1830 °F) and pressures of 7 MPa (1 ksi) and above.

Although good tensile properties can be achieved, the as-bonded toughness will be low. In addition, a postweld heat treatment to restore the parent microstructure is recommended.

Procedure Development and Practice Considerations for Diffusion Welding

S.B. Dunkerton, The Welding Institute

Stainless Steels

Stainless steels can be diffusion welded using conditions similar to those that are used for high-strength steels. These steels are normally covered by a thin adherent oxide (chromium oxide) that must be broken up and/or dissolved during the welding process. This is possible when higher temperatures and/or higher pressures are used. Alternatively, stainless steels can be welded in dry hydrogen, which will reduce the oxide film, or the oxide can be removed prior to welding. The latter measure has been accomplished by anodic cleaning of the steel surfaces, followed immediately by copper plating to protect the surfaces from reoxidation. Again, if a dry hydrogen atmosphere is used, then the copper oxide will be reduced during welding and a sound bond will be attained.

Although few property data are available on the diffusion welding of stainless steels, it is known that tensile strength is readily achieved, whereas toughness is more difficult. If high welding temperatures are employed, then it will generally be necessary to use a postweld heat treatment to restore properties.

Procedure Development and Practice Considerations for Diffusion Welding

S.B. Dunkerton, The Welding Institute

Aluminum-Base Alloys

Diffusion welding of aluminum alloys is difficult, because of the presence of a tenacious surface oxide film. The surface oxide can be disrupted by the use of large deformations and/or high temperatures in, for example, roll bonding. However, for low-deformation diffusion welding, interlayers of similar or dissimilar materials have been used. The specific procedures to be followed for aluminum alloys vary, depending on the alloy system.

For aluminum-magnesium-silicon alloys (specifically, Al-0.6Mg-0.9Si-0.5Mn), diffusion welding without an interlayer has yielded tensile strengths of 145 MPa (21 ksi) for a welding temperature of 550 °C (1020 °F) (Ref 14). However, this does induce deformation of the components. It has been more attractive to use copper or silver interlayers, which can be applied as a coating or a foil. These interlayers can be used to provide a solid-state weld or a transient liquid-phase (eutectic) weld. Care needs to be taken to achieve the latter, to ensure that the heating rate is not so slow as to allow the interlayer to be diffused away before the eutectic temperature is reached. Solid-state diffusion welds made with either copper or silver have yielded joint strengths of 85 MPa (12 ksi) for bonding temperatures as low as 450 °C (840 °F).

More attention has been directed to the diffusion welding of alloys that are of interest to the aerospace industries, that is, 7010 (Ref 15), 7475 (Ref 16, 17) and 8090 (Ref 18, 19, 20, 21). No clear guidelines exist as to the best practice for welding these alloys, because researchers have investigated many techniques and the results point in sometimes conflicting directions. Welding the alloys without an interlayer is possible, if disruption of the oxide can take place during welding. Some success has been achieved by welding the 7475 alloy with a "rough" surface (that is, brushed in air), by surface grinding the alloy in an argon atmosphere prior to welding, and by grit blasting with fine alumina particles followed by chemical cleaning in hot 30 wt% NaOH solution. Long welding times have been used for this material and high strengths have been reported by welding at 515 °C (960 °F) under a load of 5 MPa (0.7 ksi) for 5 h, followed by a postweld heat treatment to T6 condition. In instances where interlayers have been used for this alloy system, the material that has shown the greatest promise is zinc.

Aluminum-lithium alloys are currently attracting great attention for applications in the aerospace industries. This is because lithium significantly increases the elastic modulus of aluminum, by approximately 6% for each wt% addition of lithium, and simultaneously decreases its density by 3%. For aerospace applications, improvements in specific modulus and strength can lead directly to weight savings. In addition, it has been shown that aluminum-lithium alloys can develop superplastic behavior after appropriate thermomechanical processing, which means that advantage can be taken of the SPF/DW procedures already established in this industry.

Superplastic forming after welding restricts the welding temperature that can be used. Ideally, the temperature should be lower than or equal to the superplastic temperature for the alloy. In addition, the welding time must be restricted to avoid grain growth in the parent sheet, which could affect the superplastic-forming properties.

Aluminum-lithium alloys have been bonded with no interlayer and with interlayers of zinc, copper, zinc-copper, pure aluminum, and aluminum alloys. Surfaces have been prepared by a number of techniques, such as grinding with emery paper, shot peening, chemical etching, and degreasing. All techniques have shown some degree of success, but it is difficult to compare results, because the testing methods to achieve shear strength are varied. Generally, all of these welds require some form of postweld heat treatment to restore properties (for example, solution treatment at 535 °C, or 995 °F, followed by quenching). A section through a diffusion weld containing a pure aluminum interlayer is shown in Fig. 3. The position of the original aluminum foil is discernible by the presence of spheroidal precipitates, similar in size and shape to those precipitated in the grain boundaries of the aluminum alloy.

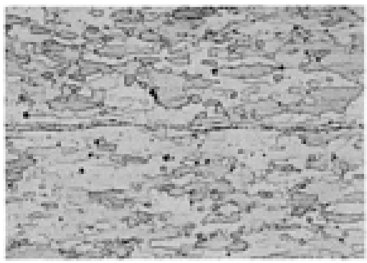

FIG. 3 CROSS SECTION OF DIFFUSION WELD IN ALUMINUM-LITHIUM ALLOY CONTAINING PURE ALUMINUM INTERLAYER. 75×

In order to avoid surface contamination during heating to the welding temperature, greater care needs to be taken with the aluminum-lithium alloys, because of the reactivity of lithium. However, this reactivity is also claimed to make this alloy system somewhat easier to weld than other less-reactive aluminum alloys.

References cited in this section

14. A.E. DRAY AND E.R. WALLACH, "DIFFUSION BONDING OF ALUMINIUM ALLOYS," PAPER PRESENTED AT INTERNATIONAL CONFERENCE ON DIFFUSION BONDING (CRANFIELD, UK), 1987
15. D.V. DUNFORD, C.G. GILMORE, AND P.G. PARTRIDGE, "EFFECT OF SILVER COATING THICKNESS ON THE SHEAR STRENGTH OF DIFFUSION BONDED CLAD AL-ZN-MG (7010) ALLOY," PAPER PRESENTED AT INTERNATIONAL CONFERENCE ON DIFFUSION BONDING (CRANFIELD, UK), 1987
16. N. RIDLEY *ET AL.*, "DIFFUSION BONDING OF SUPERPLASTIC ALUMINIUM ALLOYS," PAPER PRESENTED AT INTERNATIONAL CONFERENCE ON DIFFUSION BONDING (CRANFIELD, UK), 1987
17. D.W. LIVESEY AND N. RIDLEY, DIFFUSION BONDING OF SUPERPLASTIC ALUMINIUM ALLOYS USING A TRANSIENT LIQUID PHASE INTERLAYER (ZINC), *DIFFUSION BONDING 2,* ELSEVIER APPLIED SCIENCE, 1991
18. R.A. RICKS *ET AL.*, DEVELOPMENT OF DIFFUSION BONDING TECHNIQUES FOR AL-LI BASE ALLOY AA8090, *DIFFUSION BONDING 2,* ELSEVIER APPLIED SCIENCE, 1991
19. H.M. TENSI AND M. WITTMAN, INFLUENCE OF SURFACE PREPARATION ON THE DIFFUSION WELDING OF HIGH STRENGTH ALUMINIUM ALLOYS, *DIFFUSION BONDING 2,* ELSEVIER APPLIED SCIENCE, 1991
20. D.V. DUNFORD, P.G. PARTRIDGE, AND C.G. GILMORE, DIFFUSION BONDING OF AL-LI 8090 ALLOY, *DIFFUSION BONDING 2,* ELSEVIER APPLIED SCIENCE, 1991
21. A. URENA AND S.B. DUNKERTON, "DIFFUSION BONDING OF AN ALUMINIUM-LITHIUM ALLOY (AA 8090)," REPORT 403/1989, THE WELDING INSTITUTE, 1989

Procedure Development and Practice Considerations for Diffusion Welding

S.B. Dunkerton, The Welding Institute

Reactive and Refractory Metals

The diffusion welding of titanium alloys is well documented, because of its commercial application and importance to the aerospace industry. Many titanium airframe structures are now manufactured using the SPF/DW process, which provides significant benefits through weight and cost savings. Typical examples include hot-air nozzles, landing-gear doors, intricately shaped housings, keel sections, and heat-exchanger ducts. Most of the detailed procedures for welding are commercially sensitive. The factors that are important are (Ref 22):

- METALLURGICAL CONDITION OF THE STARTING MATERIAL
- CHARACTERIZATION OF THE SUPERPLASTICITY OF THE MATERIAL
- SURFACE CLEANLINESS
- TOOLING MATERIAL AND SURFACE CONDITION
- WELDING ATMOSPHERE
- FORMING AND WELDING PARAMETERS
- PART GEOMETRY

The primary alloy that is used is Ti-6Al-4V. It has been established that commercially available sheet material containing an equiaxed, uniform, small grain size (preferably, <10 μm, or 400 μin.) is suitable for superplastic forming. Conventional cutting processes are used for the preparation of sheet material, which is then subjected to an etching treatment prior to placement in the press. Although tooling is typically made of a 22Cr-4Ni-9Mn stainless steel with appropriate surface-parting compounds, ceramic-faced tooling has also been used. During the application of heat at temperatures ranging from 900 to 930 °C (1650 to 1710 °F), argon gas provides a protective atmosphere that is also utilized, under pressure, to effect forming and diffusion welding. Tool-to-titanium sealing, in order to contain the forming/bonding pressure of 2 to 4 MPa (0.3 to 0.6 ksi), is of paramount importance to leakage prevention.

Multiple-sheet fabrication requires the application of a stop-off material (typically, yttria and a binder) after conventional titanium cleaning. The application of a stop-off material utilizes a silk-screen technique that allows complex patterns to be transferred to the sheet. The multiple-sheet pack is plumbed with stainless steel capillary tubes for internal gas pressurization (between the sheets), as opposed to single-sheet processing, where gas pressure access is achieved directly through tooling parts. The process allows formed titanium shapes that are far more complex than those produced by conventional forming methods. Parts can feature integral end flanges, deep beads, corner beads, and small-bend radii. In addition, compound contour; multiple, variable, hollow sections; and complex sandwich (including circular) structures with varying core configurations can be fabricated.

Zirconium and its alloys also have been joined by diffusion welding, as have the refractory metals of tungsten, tantalum, niobium, and molybdenum (Ref 23).

References cited in this section

22. R.J. REHDER AND D.T. LOVELL, PROCESS DEVELOPMENT FOR DIFFUSION WELDING TI-6AL-4V, *WELD. J.,* MAY 1970, P 213-S TO 218-S
23. A.T. D'ANNESSA, THE SOLID-STATE BONDING OF REFRACTORY METALS, *WELD. J.,* MAY 1964, P 232-S TO 240-S

Other Alloys

Other alloy systems have been diffusion welded, but with few industrial applications. Copper bonds to itself readily, if the surface oxide can be disrupted, and temperatures ranging from 850 to 950 °C (1560 to 1740 °F) have been employed. Alternatively, a liquid-phase procedure that uses silver as an interlayer has been adopted to reduce welding times. The nickel- and cobalt-base alloys are bondable, but some alloy systems exhibit good high-temperature strength that inhibits the microyielding required for diffusion welding. Beryllium alloys also can be bonded at low temperatures (Ref 24).

The noble metals, such as gold, bond readily because of their very thin surface oxide. Gold, in particular, is often used as an interlayer material when welding temperatures are restricted, because bonds can be achieved at temperatures as low as 250 to 350 °C (480 to 660 °F), if the surfaces and atmosphere are clean.

Reference cited in this section

24. J.L. KNOWLES AND T.H. HAZLETT, HIGH-STRENGTH LOW-TEMPERATURE BONDING OF BERYLLIUM AND OTHER METALS, *WELD. J.,* VOL 49, JULY 1970, P 301-S TO 310-S

Dissimilar Metal Combinations

Ferrous-to-Ferrous Combinations. Dissimilar steel joints have not bee extensively investigated in diffusion-welding research, primarily because it was initially believed that the process would mostly be used to join material combinations that were difficult to join by other, more-conventional methods. However, this is not strictly true, and in situations where multiple joints can be formed in one welding operation, the process can be economically viable for dissimilar steel joints.

The literature on steel includes the bonding of high-speed steels to carbon steels for prismatic cutting tools (Ref 25) and the joining of pure iron to a range of carbon steels (Ref 26). In this latter work, the bond tensile strength was greater than 430 MPa (62 ksi), which exceeds that of the weaker parent metal. A slight increase in tensile strength occurred with a decrease in welding temperature from 1000 to 900 °C (1830 to 1650 °F).

Nonferrous-to-Nonferrous Combinations. Many combinations have been studied within this category, but the most common one uses aluminum as one of the parent materials (Ref 27, 28, 29).

The diffusion welding of aluminum to copper and titanium results in the formation of intermetallic compounds at the bond interface. With copper, it has been found that the aluminum/copper intermetallic can reach 15 μm (600 μin.) in thickness before a reduction in the joint strength occurs (Ref 30). Also described in this work is the fact that the parent aluminum strength was not achieved, even before the excessive growth of the intermetallic, which suggests that full bonding can be achieved only after the time necessary for extensive intermetallic formation. However, aluminum-to-titanium diffusion bonds can reach a tensile strength that is equal to that of the parent aluminum, and it is claimed that intermetallic thicknesses of up to 10 μm (400 μin.) can be tolerated (Ref 31). Again, an interlayer of silver can be used to promote bonding between the aluminum and titanium, such that the parent aluminum strength can be achieved in shorter welding times.

The aerospace and nuclear industries have driven the development of a wide variety of special alloy and dissimilar-metal combinations. Examples of materials that have been successfully joined include nickel-stainless steel (Ref 32), uranium-stainless steel (Ref 33), beryllium-nickel (Ref 34), beryllium-stainless steel (Ref 34), thorium dioxide dispersion-strengthened-nickel alloys (Ref 35), zirconium-steel (Ref 36), and copper-Monel (Ref 37).

Ferrous-to-Nonferrous Combinations. The joining of aluminum to steel has received a great deal of interest. This subject has been reviewed by Elliot and Wallach (Ref 38), who recognized that the two major problems associated with this metal combination are:

- THE TENACIOUS OXIDE LAYER ON THE ALUMINUM, WHICH, BEING STOICHIOMETRIC AND HIGHLY STABLE, INHIBITS WELD FORMATION
- THE FORMATION OF BRITTLE INTERMETALLICS, SUCH AS THE $FEAL_3$ AND FE_2AL_5 TYPES, AT THE JOINT INTERFACE

These problems are generally overcome by using interlayers, both to promote welding and to prevent the formation of brittle intermetallic compounds. Silver has been successfully used as an interlayer, either as a foil interlayer or applied by electroplating (Ref 39, 40), vapor deposition (Ref 41, 42), or sputter deposition (Ref 43) on one or both surfaces. Both solid- and liquid-phase joints have been produced, the latter of which involves melting of the aluminum-silver eutectic.

Other interlayer materials that have been examined include nickel and a multilayer system of zinc and silver. However, both of these have disadvantages. The nickel also forms brittle intermetallics with iron that inhibit joint strength. In the multilayer system, the zinc was found to erode the silver, thereby removing the diffusion barrier. In addition, aluminum and steel alloys can be diffusion welded without any interlayer by a careful evaluation of the process parameters and with a higher amount of aluminum deformation.

The diffusion welding of refractory metals to steels has been another major research area. Molybdenum has been welded directly to steel, but the formation of brittle intermetallics limits the strength of the joints. Lison and Stelzer (Ref 25) have investigated the interlayers required to give good joint properties for this metal combination, based on metallurgical compatibility and the coefficients of thermal expansion. A complex multilayer system was devised, in which tantalum, vanadium, nickel, and copper were positioned in the following order: Mo-Ta-Ni-Cu-steel. Initially, the Mo-Ta-V combination was bonded at 1500 °C (2730 °F) for 40 min, whereas the remaining bonds were formed at 850 °C (1560 °F) in 20 min. The mechanical properties of such a joint are not given, but metallographic examination revealed full bond formation.

Metal-Ceramic Joining. Recent developments in gas-turbine and ceramic-engine technology have created an interest in the diffusion welding of ceramics to steels, because the low temperatures used for joining minimize the residual stresses created by thermal mismatch. Silicon nitride and zirconium oxide are two examples of oxides that have been successfully welded to steel using the diffusion welding process (Ref 44, 45).

References cited in this section

25. R. LISON AND J.F. STELZER, DIFFUSION WELDING OF REACTIVE AND REFRACTORY METALS TO STAINLESS STEELS, *WELD. J.,* VOL 58 (NO. 10), 1979, P 306-314
26. V.F. SHATINSKI *ET AL.,* THE DIFFUSION BONDING OF REFRACTORY ALLOYS TO A STEEL WITH A DECARBURIZED SURFACE, *WELD. PROD.,* VOL 24 (NO. 12), 1977
27. E.R. NAIMON, J.H. DOYLE, C.R. RICE, D. VIGIL, AND D.R. WALMSLEY, DIFFUSION WELDING OF ALUMINUM TO STAINLESS STEEL, *WELD. J.,* NOV 1981, P 17-20
28. P.D. CALDERON, D.R. WALMSLEY, AND Z.A. MUNIR, AN INVESTIGATION OF DIFFUSION WELDING OF PURE AND ALLOYED ALUMINUM TO TYPE 316 STAINLESS STEEL, *WELD. J.,* APRIL 1985, P 104-S TO 112-S
29. C.H. CRANE, D.T. LOVELL, W.A. BAGINSKI, AND M.G. OLSEN, DIFFUSION WELDING OF DISSIMILAR METALS, *WELD. J.,* VOL 46, P 23-S TO 31-S
30. T. ENJYO *ET AL.,* DIFFUSION WELDING OF COPPER TO ALUMINIUM, *TRANS. JWRI,* VOL 8, 1979, P 77-83
31. T. ENJYO *ET AL.,* DIFFUSION BONDING OF ALUMINIUM TO TITANIUM, *TRANS. JWRI,* VOL 6 (NO. 1), 1977, P 123-130
32. G.R. KAMAT, SOLID-STATE DIFFUSION WELDING OF NICKEL TO STAINLESS STEEL, *WELD. J.,* JUNE 1988, P 44-46
33. R.S. ROSEN, D.R. WALMSLEY, AND Z.A. MUNIR, THE PROPERTIES OF SILVER-AIDED DIFFUSION WELDS BETWEEN URANIUM AND STAINLESS STEEL, *WELD. J.,* VOL 65, APRIL 1986, P 83-S TO 92-S
34. W. FEDUSKA AND W.L. HORIGAN, DIFFUSION BONDING OF HIGH-TEMPERATURE ALLOYS WITH BERYLLIUM, *WELD. J.,* JAN 1962, P 28-S TO 35-S
35. T.J. MOORE AND K.H. HOLKO, SOLID STATE WELDING OF TD-NICKEL BAR, *WELD. J.,* SEPT 1970, P 395-S TO 409-S
36. M.L. WAYMAN, R.R. SMITH, AND M.G. WRIGHT, THE DIFFUSION BONDING OF ZR-2.5% NB TO STEEL, *METALL. TRANS. A,* VOL 17, MARCH 1986, P 429-433
37. J.T. NIEMANN, R.P. SOPHER, AND P.J. RIEPPEL, DIFFUSION BONDING BELOW 1000 °F, *WELD. J.,* AUG 1958, P 337-S TO 342-S
38. S. ELLIOT AND E.R. WALLACH, JOINING ALUMINIUM TO STEEL, PART 1--DIFFUSION BONDING, *MET. CONSTR.,* VOL 13 (NO. 3), 1981, P 167-171
39. J.W. DINI, USE OF ELECTRODEPOSITION TO PROVIDE COATINGS FOR SOLID-STATE BONDING, *WELD. J.,* NOV 1982, P 33-39
40. J.W. DINI, W.K. KELLEY, W.C. COWDEN, AND E.M. LOPEZ, USE OF ELECTRODEPOSITED SILVER AS AN AID IN DIFFUSION WELDING, *WELD. J.,* JAN 1984, P 26-S TO 34-S
41. C. SCHALANSKY, Z.A. MUNIR, AND D.L. WALMSLEY, AN INVESTIGATION ON THE BONDING OF HOT-HOLLOW CATHODE DEPOSITED SILVER LAYERS TO TYPE 304 STAINLESS STEEL, *J. MATER. SCI.,* VOL 22, 1987, P 745-751
42. P.S. MACLEOD AND G. MAH, THE EFFECT OF SUBSTRATE BIAS VOLTAGE ON THE BONDING OF EVAPORATED SILVER COATINGS, *J. VAC. SCI. TECHNOL. II,* VOL 1, JAN/FEB 1974, P 119-121
43. R.S. ROSEN AND M.E. KASSNER, DIFFUSION WELDING OF SILVER INTERLAYERS COATED ONTO BASE METALS BY PLANAR-MAGNETRON SPUTTERING, *J. VAC. SCI. TECHNOL.,* VOL

A8 (NO. 1), JAN/FEB 1990, P 19-29

44. K. SUGANUMA, T. OKAMOTO, M. KOIZUMI, AND M. SHIMADA, SOLID-STATE BONDING OF PARTIALLY STABILIZED ZIRCONIA TO STEEL WITH TITANIUM INTERLAYER, *J. MATER. SCI. LETT.*, VOL 5, 1986, P 1099-1100
45. K. SUGANUMA, T. OKAMOTO, Y. MIYAMOTO, M. SHIMADA, AND M. KOIZUMI, JOINING SI_3N_4 TO TYPE 405 STEEL WITH SOFT METAL INTERLAYERS, *MATER. SCI. TECHNOL.*, VOL 2, NOV 1986, P 1156-1161

Procedure Development and Practice Considerations for Diffusion Welding

S.B. Dunkerton, The Welding Institute

References

1. W.A. OWCZARSKI AND D.F. PAULONIS, APPLICATION OF DIFFUSION WELDING IN THE USA, *WELD. J.*, FEB 1981, P 22-33
2. O.S. SPANSWICK, DIFFUSION BONDING--A CASE STUDY, REPRINT R269/10/B4, *WELD. INST. RES. BULL.*, THE WELDING INSTITUTE, 1984
3. B. DERBY AND E.R. WALLACH, THEORETICAL MODEL FOR DIFFUSION BONDING, *MET. SCI.*, VOL 16, JAN 1982, P 49-56
4. B. DERBY AND E.R. WALLACH, DIFFUSION BONDING: DEVELOPMENT OF THEORETICAL MODEL, *MET. SCI.*, VOL 18, SEPT 1984, P 427-431
5. I. TUAH-POKU, M. DOLLAR, AND T.B. MASSALSKI, A STUDY OF THE TRANSIENT LIQUID PHASE BONDING PROCESS APPLIED TO AG/CU/AG SANDWICH JOINT, *METALL. TRANS. A*, VOL 19, MARCH 1988, P 675-686
6. J. PILLING, THE KINETICS OF ISOSTATIC DIFFUSION BONDING IN SUPERPLASTIC MATERIALS, *MATER. SCI. ENG.*, VOL 100, 1988, P 137-144
7. J. PILLING, D.W. LIVESEY, J.B. HAWKYARD, AND N. RIDLEY, SOLID STATE BONDING IN SUPERPLASTIC TI-6AL-4V, *MET. SCI.*, VOL 18, MARCH 1984, P 117-122
8. Y. MAEHARA, Y. KOMIZO, AND T.G. LANGDON, PRINCIPLES OF SUPERPLASTIC DIFFUSION BONDING, *MATER. SCI. TECHNOL.*, VOL 4, AUG 1988, P 669-674
9. J. PILLING AND N. RIDLEY, SOLID STATE BONDING OF SUPERPLASTIC AA7475, *MATER. SCI. TECHNOL.*, VOL 3, MAY 1987, P 353-359
10. W.A. BRYANT, A METHOD FOR SPECIFYING HOT ISOSTATIC PRESSURE WELDING PARAMETERS, *WELD. J.*, DEC 1975, P 433-S TO 435-S
11. C.E. THORNTON AND E.R. WALLACH, IMPACT STRENGTH OF STEEL DIFFUSION BONDS, *MET. CONSTR.*, JULY 1985, P 450R-455R
12. S. ELLIOT, "PROPERTIES OF DIFFUSION BONDS IN A 0.4 C STEEL," REPORT 131/1980, THE WELDING INSTITUTE, 1980
13. S.B. DUNKERTON, "DIFFUSION BONDING OF A 0.4 CARBON CHROMIUM/MOLYBDENUM STEEL," REPORT 178/1982, THE WELDING INSTITUTE, 1982
14. A.E. DRAY AND E.R. WALLACH, "DIFFUSION BONDING OF ALUMINIUM ALLOYS," PAPER PRESENTED AT INTERNATIONAL CONFERENCE ON DIFFUSION BONDING (CRANFIELD, UK), 1987
15. D.V. DUNFORD, C.G. GILMORE, AND P.G. PARTRIDGE, "EFFECT OF SILVER COATING THICKNESS ON THE SHEAR STRENGTH OF DIFFUSION BONDED CLAD AL-ZN-MG (7010) ALLOY," PAPER PRESENTED AT INTERNATIONAL CONFERENCE ON DIFFUSION BONDING (CRANFIELD, UK), 1987

16. N. RIDLEY *ET AL.*, "DIFFUSION BONDING OF SUPERPLASTIC ALUMINIUM ALLOYS," PAPER PRESENTED AT INTERNATIONAL CONFERENCE ON DIFFUSION BONDING (CRANFIELD, UK), 1987
17. D.W. LIVESEY AND N. RIDLEY, DIFFUSION BONDING OF SUPERPLASTIC ALUMINIUM ALLOYS USING A TRANSIENT LIQUID PHASE INTERLAYER (ZINC), *DIFFUSION BONDING 2,* ELSEVIER APPLIED SCIENCE, 1991
18. R.A. RICKS *ET AL.,* DEVELOPMENT OF DIFFUSION BONDING TECHNIQUES FOR AL-LI BASE ALLOY AA8090, *DIFFUSION BONDING 2,* ELSEVIER APPLIED SCIENCE, 1991
19. H.M. TENSI AND M. WITTMAN, INFLUENCE OF SURFACE PREPARATION ON THE DIFFUSION WELDING OF HIGH STRENGTH ALUMINIUM ALLOYS, *DIFFUSION BONDING 2,* ELSEVIER APPLIED SCIENCE, 1991
20. D.V. DUNFORD, P.G. PARTRIDGE, AND C.G. GILMORE, DIFFUSION BONDING OF AL-LI 8090 ALLOY, *DIFFUSION BONDING 2,* ELSEVIER APPLIED SCIENCE, 1991
21. A. URENA AND S.B. DUNKERTON, "DIFFUSION BONDING OF AN ALUMINIUM-LITHIUM ALLOY (AA 8090)," REPORT 403/1989, THE WELDING INSTITUTE, 1989
22. R.J. REHDER AND D.T. LOVELL, PROCESS DEVELOPMENT FOR DIFFUSION WELDING TI-6AL-4V, *WELD. J.,* MAY 1970, P 213-S TO 218-S
23. A.T. D'ANNESSA, THE SOLID-STATE BONDING OF REFRACTORY METALS, *WELD. J.,* MAY 1964, P 232-S TO 240-S
24. J.L. KNOWLES AND T.H. HAZLETT, HIGH-STRENGTH LOW-TEMPERATURE BONDING OF BERYLLIUM AND OTHER METALS, *WELD. J.,* VOL 49, JULY 1970, P 301-S TO 310-S
25. R. LISON AND J.F. STELZER, DIFFUSION WELDING OF REACTIVE AND REFRACTORY METALS TO STAINLESS STEELS, *WELD. J.,* VOL 58 (NO. 10), 1979, P 306-314
26. V.F. SHATINSKI *ET AL.,* THE DIFFUSION BONDING OF REFRACTORY ALLOYS TO A STEEL WITH A DECARBURIZED SURFACE, *WELD. PROD.*, VOL 24 (NO. 12), 1977
27. E.R. NAIMON, J.H. DOYLE, C.R. RICE, D. VIGIL, AND D.R. WALMSLEY, DIFFUSION WELDING OF ALUMINUM TO STAINLESS STEEL, *WELD. J.,* NOV 1981, P 17-20
28. P.D. CALDERON, D.R. WALMSLEY, AND Z.A. MUNIR, AN INVESTIGATION OF DIFFUSION WELDING OF PURE AND ALLOYED ALUMINUM TO TYPE 316 STAINLESS STEEL, *WELD. J.,* APRIL 1985, P 104-S TO 112-S
29. C.H. CRANE, D.T. LOVELL, W.A. BAGINSKI, AND M.G. OLSEN, DIFFUSION WELDING OF DISSIMILAR METALS, *WELD. J.,* VOL 46, P 23-S TO 31-S
30. T. ENJYO *ET AL.,* DIFFUSION WELDING OF COPPER TO ALUMINIUM, *TRANS. JWRI,* VOL 8, 1979, P 77-83
31. T. ENJYO *ET AL.,* DIFFUSION BONDING OF ALUMINIUM TO TITANIUM, *TRANS. JWRI,* VOL 6 (NO. 1), 1977, P 123-130
32. G.R. KAMAT, SOLID-STATE DIFFUSION WELDING OF NICKEL TO STAINLESS STEEL, *WELD. J.,* JUNE 1988, P 44-46
33. R.S. ROSEN, D.R. WALMSLEY, AND Z.A. MUNIR, THE PROPERTIES OF SILVER-AIDED DIFFUSION WELDS BETWEEN URANIUM AND STAINLESS STEEL, *WELD. J.,* VOL 65, APRIL 1986, P 83-S TO 92-S
34. W. FEDUSKA AND W.L. HORIGAN, DIFFUSION BONDING OF HIGH-TEMPERATURE ALLOYS WITH BERYLLIUM, *WELD. J.,* JAN 1962, P 28-S TO 35-S
35. T.J. MOORE AND K.H. HOLKO, SOLID STATE WELDING OF TD-NICKEL BAR, *WELD. J.,* SEPT 1970, P 395-S TO 409-S
36. M.L. WAYMAN, R.R. SMITH, AND M.G. WRIGHT, THE DIFFUSION BONDING OF ZR-2.5% NB TO STEEL, *METALL. TRANS. A,* VOL 17, MARCH 1986, P 429-433
37. J.T. NIEMANN, R.P. SOPHER, AND P.J. RIEPPEL, DIFFUSION BONDING BELOW 1000 °F, *WELD. J.,* AUG 1958, P 337-S TO 342-S

38. S. ELLIOT AND E.R. WALLACH, JOINING ALUMINIUM TO STEEL, PART 1--DIFFUSION BONDING, *MET. CONSTR.,* VOL 13 (NO. 3), 1981, P 167-171
39. J.W. DINI, USE OF ELECTRODEPOSITION TO PROVIDE COATINGS FOR SOLID-STATE BONDING, *WELD. J.,* NOV 1982, P 33-39
40. J.W. DINI, W.K. KELLEY, W.C. COWDEN, AND E.M. LOPEZ, USE OF ELECTRODEPOSITED SILVER AS AN AID IN DIFFUSION WELDING, *WELD. J.,* JAN 1984, P 26-S TO 34-S
41. C. SCHALANSKY, Z.A. MUNIR, AND D.L. WALMSLEY, AN INVESTIGATION ON THE BONDING OF HOT-HOLLOW CATHODE DEPOSITED SILVER LAYERS TO TYPE 304 STAINLESS STEEL, *J. MATER. SCI.,* VOL 22, 1987, P 745-751
42. P.S. MACLEOD AND G. MAH, THE EFFECT OF SUBSTRATE BIAS VOLTAGE ON THE BONDING OF EVAPORATED SILVER COATINGS, *J. VAC. SCI. TECHNOL. II,* VOL 1, JAN/FEB 1974, P 119-121
43. R.S. ROSEN AND M.E. KASSNER, DIFFUSION WELDING OF SILVER INTERLAYERS COATED ONTO BASE METALS BY PLANAR-MAGNETRON SPUTTERING, *J. VAC. SCI. TECHNOL.,* VOL A8 (NO. 1), JAN/FEB 1990, P 19-29
44. K. SUGANUMA, T. OKAMOTO, M. KOIZUMI, AND M. SHIMADA, SOLID-STATE BONDING OF PARTIALLY STABILIZED ZIRCONIA TO STEEL WITH TITANIUM INTERLAYER, *J. MATER. SCI. LETT.,* VOL 5, 1986, P 1099-1100
45. K. SUGANUMA, T. OKAMOTO, Y. MIYAMOTO, M. SHIMADA, AND M. KOIZUMI, JOINING SI_3N_4 TO TYPE 405 STEEL WITH SOFT METAL INTERLAYERS, *MATER. SCI. TECHNOL.,* VOL 2, NOV 1986, P 1156-1161

Procedure Development and Practice Considerations for Inertia and Direct-Drive Friction Welding

Tim Stotler, Edison Welding Institute

Introduction

FRICTION WELDING (FRW) is a solid-state welding process that uses the compressive force of the workpieces that are rotating or moving relative to one another, producing heat and plastically displacing material from the faying surfaces, thereby creating a weld. Process variations include inertia, direct-drive, linear, and orbital radial friction welding, as well as friction surfacing. (Additional information is available in the articles "Fundamentals of Friction Welding,", "Friction Welding,", "Radial Friction Welding,", and "Friction Surfacing" in this Volume.) This article provides information about practice considerations for the two most common variations: inertia and direct-drive friction welding.

Inertia welding obtains heat for welding that is supplied by stored rotational kinetic energy. Figure 1 depicts the parameter characteristics of the process. The inertia of the system is changed by either adding or removing flywheels from the rotating spindle that provides rotary motion and/or by changing the spindle speed.

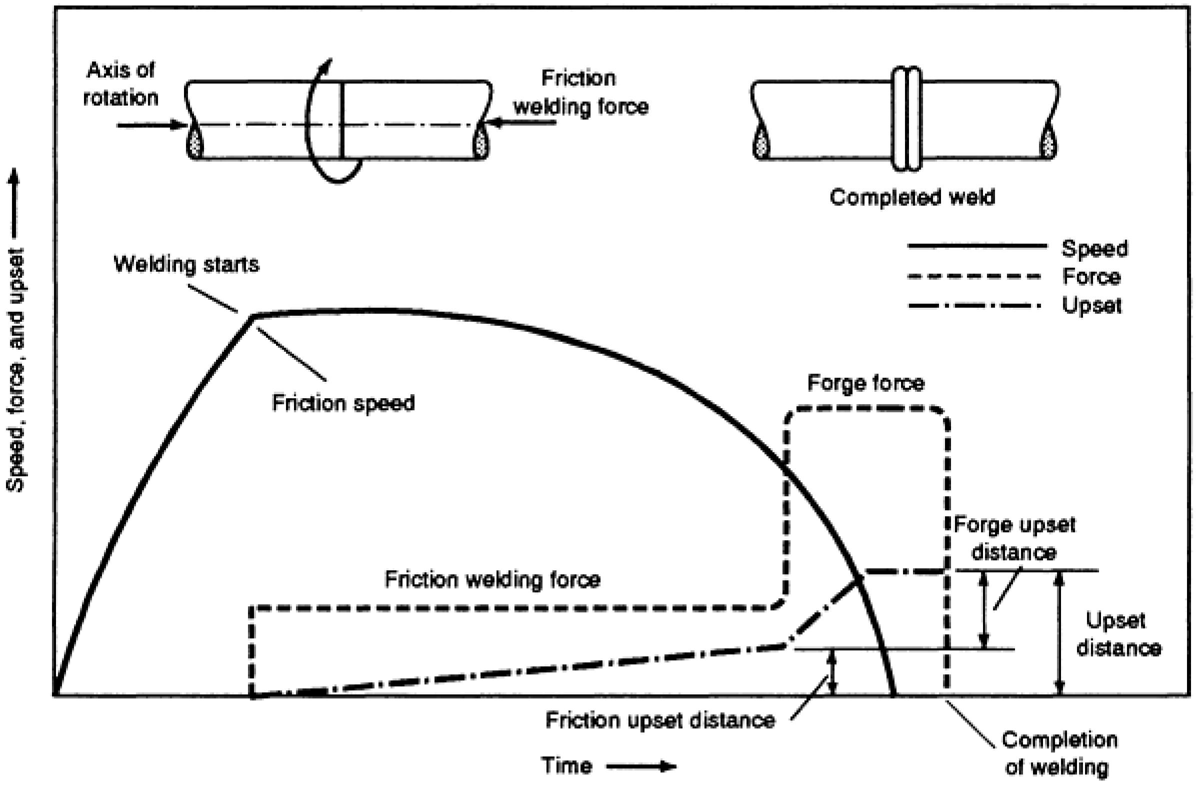

FIG. 1 INERTIA WELDING PARAMETER CHARACTERISTICS. SOURCE: ANSI/AWS C6.1-89

After the inertial mass is set, the parts are loaded in the machine (one in the spindle and one in the fixture/tailstock). The spindle speed and welding forces are set to predetermined levels. The spindle is then accelerated to velocity. After this velocity is attained, the drive system is disengaged, allowing the flywheel (and the part in the spindle) to effectively coast. The stationary part is then brought into contact with the rotating one, under a given applied force, and frictional heating begins. The flywheel is stopped by the transitioning of its kinetic energy into thermal energy (frictional heating) at the welding interface. This energy heats the part and allows forging to occur by displacing material at the interface.

Direct-drive friction welding differs from inertia welding, primarily in how the energy is delivered to the joint. Figure 2 depicts the characteristics of this process. In direct-drive friction welding, like inertia welding, one of the parts is loaded into a rotating spindle while the other is held stationary. However, the spindle is of a given mass and is not changed, as it is in inertia welding. To start the process, the spindle is set into motion at a constant rotational velocity (rev/min). The parts are then brought into contact under a low force (termed first friction force, or preheat force) and allowed to heat for a preset time or distance. In most applications, a higher welding force (termed second friction force, or welding force) is then applied, which increases the rate and amount of material extruded from the joint. The second friction force is controlled through a preset time or distance. After the preset time or distance has been reached for the first two friction stages, the rotating spindle is brought to a stop using some type of braking mechanism. This is typically a caliper brake(s) system and/or an electric-type brake in the spindle drive motor. In most cases, a third increase in the welding force (usually referred to as the upset force) is applied at the start of or after braking, which consolidates the joint. In contrast to inertia welding, where the microstructure at the interface reflects a rotational-type forging, direct-drive friction welding produces an axially forged microstructure in the weld. However, if braking times are increased, allowing the upset force to be applied while the spindle may still be rotating, some rotational-type forging can be obtained.

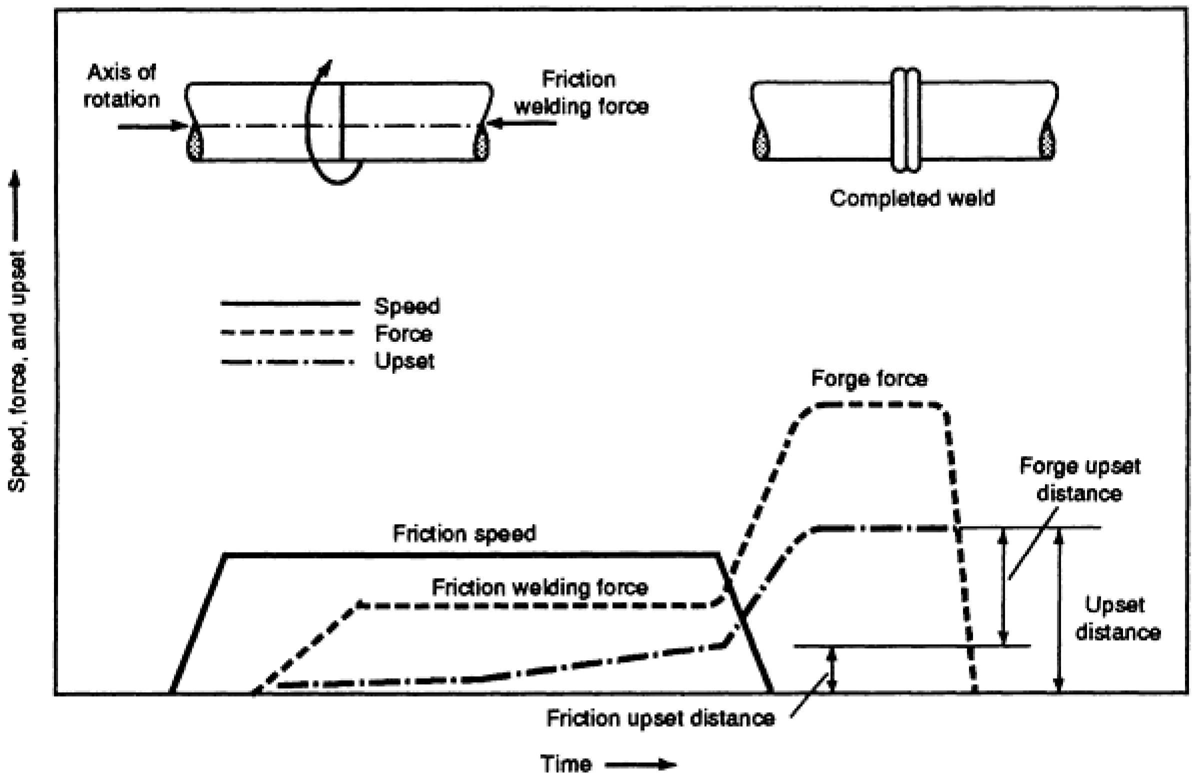
Axis of rotation
Friction welding force
Completed weld
Speed, force, and upset
Speed
Force
Upset
Forge force
Forge upset distance
Friction speed
Friction welding force
Upset distance
Friction upset distance
Time

FIG. 2 DIRECT-DRIVE FRICTION WELDING PARAMETER CHARACTERISTICS. SOURCE:ANSI/AWS C6.1-89

One advantage that direct-drive friction welding possesses is reduced torque peaks, compared with those experienced in inertia welding. The initial torque peak is reduced by applying a low initial welding force (first friction), which reduces the rate at which heating occurs. This practice typically is used to reduce the power requirements at the beginning of the welding cycle. In addition, the spindle rotation in this process is either nearly or completely stopped before the final upset of the parts, thus minimizing the final torque peak. This is advantageous when welding thin-walled tubes, the walls of which could tend to tear under high torque peaks. Although torque peaks in inertia welding can be reduced by using lower welding forces, they will still be larger than those of direct-drive friction welding.

Procedure Development and Practice Considerations for Inertia and Direct-Drive Friction Welding

Tim Stotler, Edison Welding Institute

Inertia Welding Parameter Calculations

Differences in part geometry (bar-to-bar, tube-to-tube, bar-to-plate, and so on) affect welding parameter calculations. The calculation procedures described here are based on carbon steel. For a given alloy system, inertia welding spindle speeds are set within a specified range of surface velocities. The actual surface velocity depends on the size of the part to be joined. Surface velocities for welding carbon steels in all geometries typically range from 90 to 610 m/min (300 to 2000 ft/min). The actual spindle speed (rev/min) is calculated by:

$$rev / mim = \frac{c S}{p OD} \qquad \textbf{(EQ 1)}$$

where χ, the conversion factor, is 12 for the U.S. customary unit and 1000 for the metric SI unit, S is the surface velocity (m/min or ft/min), and OD is the outside diameter of the part (mm or in.).

For solid bars and a fixed welding force, lower surface velocities tend to reduce center heating. Center heating is not a problem with tube-type geometries, which means there is greater flexibility in the surface velocities used. The effect of surface velocity on the heat pattern in a solid bar is shown in Fig. 3.

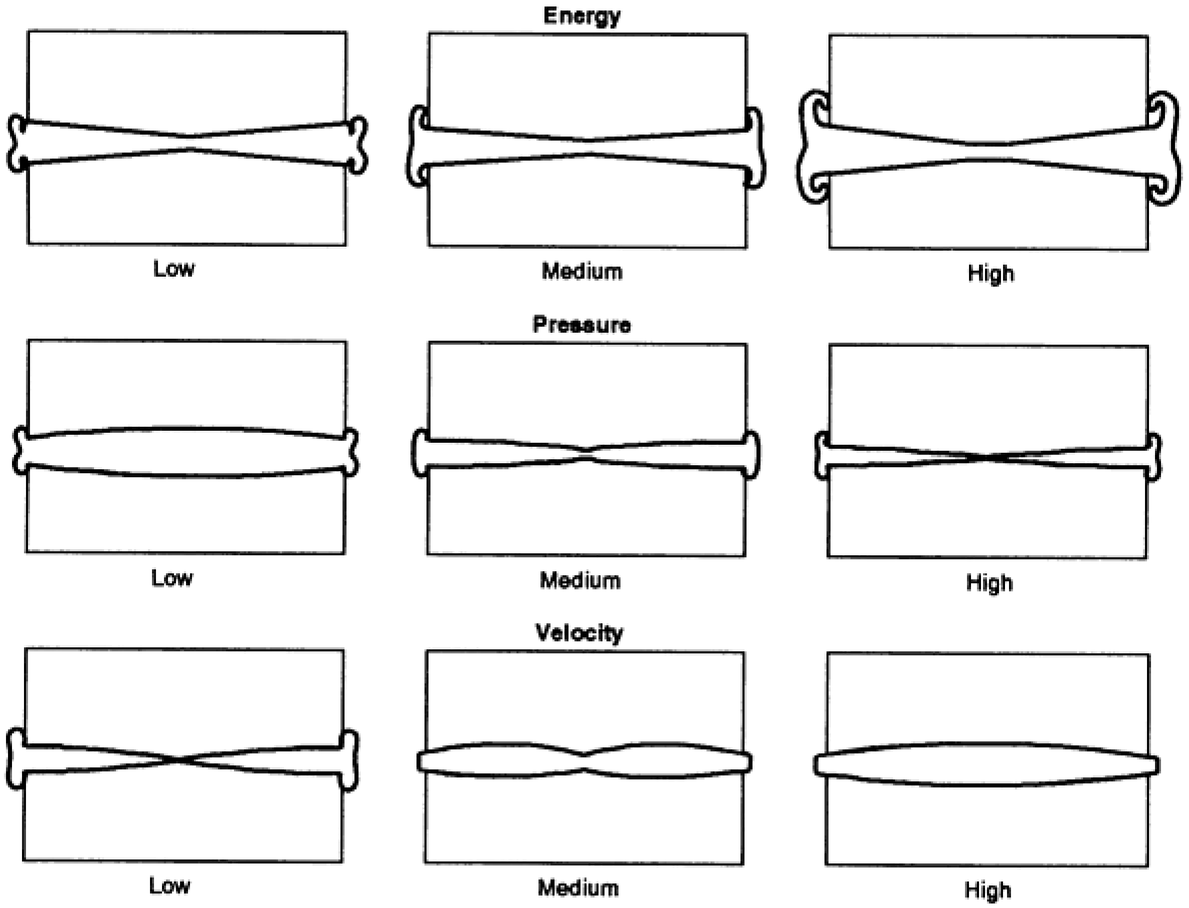

FIG. 3 EFFECT OF WELDING VARIABLES ON THE HEAT PATTERN AT THE INTERFACE AND FLASH FORMATION OF INERTIA WELDS

As mentioned in the introduction of this article, rotating flywheels are the source of energy for joining in the inertia welding process. During the welding-force stage where spindle deceleration occurs (see Fig. 1), the process is characterized by a loss of material from the interface to the flash. Larger inertias prolong the welding cycle. In addition, while maintaining a constant applied energy (combination of spindle speed and mass), rotational forging may be extended using larger flywheels and lower surface velocities.

The calculation of the flywheel energy that is needed to produce a weld depends on the surface velocity, material type, and joint geometry. As shown in Table 1 for a solid bar of low-carbon steel, the moment of inertia (expressed as wK^2 in mass units) is 0.18 kg · m^2 (4.2 lb · ft^2). This is found by using Eq 2 and the approximate flywheel energy (E_T) of 63 N · m/mm^2 (30,000 ft · lbf/in.^{2}) for the inertia welding of low-carbon steels per square inch:

$$wK^2 = \frac{E_T C}{(rev/mim)^2} \qquad \textbf{(EQ 2)}$$

where the value of C is 5873 for the U.S. customary unit and 182.4 for the SI unit, E_T is the total flywheel energy, and the (rev/min)^2 value is that for the spindle, which was calculated in Eq 1.

TABLE 1 TYPICAL INERTIA WELDING CONDITIONS FOR LOW-CARBON STEEL

PARAMETER	VALUE
BAR-TO-BAR	
DIAMETER, MM (IN.)	**25.4 (1)**
WELD AREA, MM^2 ($IN.^2$)	**506 (0.78)**
TOTAL MOMENT OF INERTIA, WK^2, KG · M^2 (LB · FT^2)	**0.18 (4.2)**
SPEED	
ANGULAR VELOCITY, REV/MIN	**5725**
SURFACE VELOCITY, M/MIN (FT/MIN)	**460 (1500)**
TOTAL FORCE, KN (LBF)	**68.9 (15,500)**
UPSET DISTANCE, MM (IN.)	**3.8 (0.150)**
TUBE-TO-TUBE	
OUTSIDE DIAMETER, MM (IN.)	**50.8 (2)**
INSIDE DIAMETER, MM (IN.)	**38.1 (1.5)**
WELD AREA, MM^2 ($IN.^2$)	**886 (1.37)**
TOTAL MOMENT OF INERTIA, WK^2, KG · M^2 (LB · FT^2)	**1.23 (29.3)**
SPEED	
ANGULAR VELOCITY, REV/MIN	**2875**
SURFACE VELOCITY, M/MIN (FT/MIN)	**460 (1500)**
TOTAL FORCE, KN (LBF)	**91.7 (20,600)**
UPSET DISTANCE, MM (IN.)	**5.1 (0.200)**

In the inertia welding process, there is flexibility in how the welding force can be applied, although a single force followed by an upset force is often used. The typical forces for welding carbon steels range from 96.5 to 207 MPa (14 to 30 ksi). The second-stage force (forging force), which can vary from 1.5 to nearly 4 times the welding force, is usually determined through experimentation. The effect of the welding force on the size of the heat-affected zone (HAZ) is inverse to that of surface velocity, that is, welds made at low welding forces have heat-affected areas resembling those produced with high surface velocities (see Fig. 3).

When designing an initial set of inertia welding parameters, the normal order of selection can be summarized as:

- CHOOSE THE DESIRED SPINDLE SPEED (REV/MIN)
- CALCULATE THE TOTAL MOMENT OF INERTIA (WK^2) THAT IS NEEDED FOR THE REQUIRED ENERGY LEVEL, E_T
- CHOOSE THE NEAREST AVAILABLE TOTAL MOMENT OF INERTIA (WK^2) LEVEL ON THE WELDING MACHINE
- RECALCULATE THE NECESSARY REV/MIN WITH THE TOTAL MOMENT OF INERTIA (WK^2) THAT IS AVAILABLE ON THE MACHINE
- CALCULATE THE NECESSARY WELDING FORCE (AND UPSET FORCE, IF USED)

Procedure Development and Practice Considerations for Inertia and Direct-Drive Friction Welding

Tim Stotler, Edison Welding Institute

Direct-Drive Friction Welding Parameter Calculations

Like those of inertia welding, the parameters for direct-drive friction welding were originally developed on carbon steel. Parameter calculations also involve only welding forces and spindle speeds. The same equation (Eq 1) is used for calculating the spindle rev/min value. Surface velocities for the direct-drive friction welding of carbon steels range from

76 to 182 m/min (250 to 600 ft/min). Tubes require slightly higher surface velocities than do bars. Typical welding pressure values for carbon steels range from 83 to 166 MPa (12 to 24 ksi) for the second friction force and from 166 to 332 MPa (24 to 48 ksi) for the upset force. Although first friction forces vary, they are typically 10 to 20% of the second friction force. Examples of welding parameters for bar-to-bar and tube-to-tube direct-drive welds are shown in Table 2.

TABLE 2 TYPICAL DIRECT-DRIVE FRICTION WELDING CONDITIONS FOR LOW-CARBON STEEL

PARAMETER	VALUE
BAR-TO-BAR	
DIAMETER, MM (IN.)	**25.4 (1)**
WELD AREA, MM^2 (IN.^{2})	**506 (0.785)**
SPEED	
ANGULAR VELOCITY, REV/MIN	**1100**
SURFACE VELOCITY, M/MIN (FT/MIN)	**88 (290)**
SECOND FRICTION FORCE, KN (LBF)	**35.6 (8000)**
UPSET FORCE, KN (LBF)	**71.2 (16,000)**
UPSET DISTANCE, MM (IN.)	**4.6 (0.180)**
TUBE-TO-TUBE	
OUTSIDE DIAMETER, MM (IN.)	**50.8 (2)**
INSIDE DIAMETER, MM (IN.)	**38.1 (1.5)**
WELD AREA, MM^2 (IN.^{2})	**886 (1.37)**
SPEED	
ANGULAR VELOCITY, REV/MIN	**900**
SURFACE VELOCITY, M/MIN (FT/MIN)	**143 (470)**
SECOND FRICTION FORCE, KN (LBF)	**48.9 (11,000)**
UPSET FORCE, KN (LBF)	**97.9 (22,000)**
UPSET DISTANCE, MM (IN.)	**3.2 (0.125)**

Procedure Development and Practice Considerations for Inertia and Direct-Drive Friction Welding

Tim Stotler, Edison Welding Institute

Carbon Steels

Generally, most metals that can be hot forged can readily be friction welded with either process variant. Carbon steels, which fall into this category, have been the fundamental basis of welding-parameter calculations for both inertia and direct-drive friction welding. Low- and medium-carbon steels are the easiest to weld and generally require no postweld heat treatments. Higher carbon and alloy steels also are easily joined, but the welding-parameter window is narrower than it is for low-carbon steels. This is primarily due to the reduced forgeability characteristics of the higher-carbon steels. Steels that contain significant amounts of free-machining elements, such as lead, also can be friction welded. However, these free-machining elements can cause undesirable directional properties near the bond line and may also act as dry lubricants that reduce frictional heating. Heat-treatable steels also can be friction welded. The rapid quenching of the weld area can present hardenability problems and a subsequent tempering or annealing procedure may be required. Hardenable steels such as these should be welded with longer heating times in order to reduce the cooling rate.

Procedure Development and Practice Considerations for Inertia and Direct-Drive Friction Welding

Tim Stotler, Edison Welding Institute

Stainless Steels

Stainless steels are commonly joined using either friction welding process. Austenitic and ferritic stainless steels are easily joined to each other and to themselves using either process. Because the heat-treatable stainless steels are sensitive to temperature and pressure, a post-weld thermal treatment may be necessary to achieve the desired weld properties. Precipitation-hardenable stainless steels are often solution heat treated and aged after welding. In some cases, however, heat-treated base material must be stress relieved.

Welding parameters for stainless steels are similar to those used with carbon steels. Surface velocities for both bars and tubes fall within the same range as for carbon steels for both process variants. The flywheel energies, E_T, and welding forces for the inertia welding of stainless steels are typically 30 to 50% higher than those for carbon steels. For direct-drive friction welding, the welding forces are nearly the same as those used for carbon steel. Another factor relevant to the friction welding of stainless steel is the location of the weld when subjected to a corrosive environment. If possible, it should be located outside of the corrosive environment. In some instances, an increased corrosion rate occurs near the weld. This can be a concern, depending on the application.

Procedure Development and Practice Considerations for Inertia and Direct-Drive Friction Welding

Tim Stotler, Edison Welding Institute

Aluminum-Base Alloys

Aluminum alloys are commonly joined by either of the friction welding processes. In inertia welding, rotational forging will form flow lines with radial patterns, as opposed to the straight flow patterns produced with direct-drive friction welding or flash welding. An advantage of direct-drive friction welding is that thin-walled aluminum tubes are somewhat easier to join. This is because the rotation of the spindle can be stopped prior to forging. Because severe hot working of the materials will occur, the heat-affected regions of friction welds are very narrow and have a fine-grained structure. In addition, the severe hot working leads to dynamic recrystallization at the interface.

Parameters for the inertia welding of aluminum alloys differ somewhat from those used for carbon steel. The differences are due to the higher thermal conductivity of these materials. The surface velocities used for welding aluminum alloys are approximately 30 to 50% higher than those used for carbon steel. Inertia welding pressures are significantly lower than they are for carbon steels, ranging from 20 to 50% of those used for a similar-sized carbon steel part. The flywheel energy is from 20 to 50% of that for carbon steel. For direct-drive friction welding, the surface velocities are 10 to 50% higher than they are for carbon steel, depending on the type of aluminum alloy. Welding and upset pressures are typically 30 to 50% of those used for carbon steel.

Procedure Development and Practice Considerations for Inertia and Direct-Drive Friction Welding

Tim Stotler, Edison Welding Institute

Reactive and Refractory Metals

The reactive metals comprise titanium and zirconium, whereas the refractory metals comprise tungsten, tantalum, niobium, and molybdenum. Of all these metals, the titanium alloys have been studied most extensively, primarily because of their application in the aerospace industry. In the inertia welding process, titanium alloys have flywheel energies that are similar to those of steel. Surface velocities are typically at the higher end of the scale, ranging from values that are equivalent to those used for carbon steel to values that are 40% greater. However, welding forces are very low, typically ranging from 10 to 30% of those used for welding carbon steel. This difference is attributed primarily to the lower thermal conductivity of titanium.

Work has been done using direct-drive friction welding for titanium, although to a lesser degree than inertia welding. The welding parameters for the direct-drive friction welding of titanium are similar to those of steel, with a slight reduction in the welding forces. Most friction-welded joints in titanium can be used in the as-welded condition. However, somewhat better tensile and fatigue properties are obtained if stress relieving or solution annealing and aging procedures are used.

Zirconium has properties that are very similar to those of titanium. Therefore, initial welding parameters should be based on the parameters for titanium.

Both friction welding processes can be used to join molybdenum and tungsten. Current applications include the friction welding of molybdenum and tungsten wires in drawing operations for the lighting industry. Although there is not much friction welding currently being performed on niobium and tantalum, it is expected that these alloys could also be successfully welded. Tantalum has a very high melting temperature and very good low-temperature ductility. This suggests that a forge-welding process could be applicable.

Procedure Development and Practice Considerations for Inertia and Direct-Drive Friction Welding

Tim Stotler, Edison Welding Institute

Copper-, Nickel-, and Cobalt-Base Materials

Copper and nickel are commonly friction welded. Copper alloys are typically used in a dissimilar-metal welding application with aluminum. In the inertia welding process, surface velocities are nearly twice those of carbon steel, because of the high thermal conductivity of copper. Flywheel energies and welding forces are approximately 50% of those used for carbon steel.

For direct-drive friction welding, higher surface velocities would be required, given the high thermal conductivity of copper. However, welding forces would be lower, because the strength of copper is lower than that of carbon steel.

Nickel-base alloys also are commonly joined using the inertia welding process. Surface velocities are much lower than those of steel, typically ranging to a level that is no higher than 50% of that for carbon steel. However, because of increased elevated-temperature properties, flywheel energies and welding forces are nearly double those required for carbon steels. After inertia welding, the joint in heat-treatable nickel-base alloys is often subjected to a solution and aging treatment. This is needed to precipitate second-phase particles (that is, γ' and γ'') that have been taken back into solution by the thermal cycle of the weld.

In direct-drive friction welding, the surface velocities are approximately half those used for carbon steel, whereas welding forces are nearly double those used for carbon steel. The welding procedures for cobalt-base alloys are similar to those for nickel-base alloys.

Procedure Development and Practice Considerations for Inertia and Direct-Drive Friction Welding

Tim Stotler, Edison Welding Institute

Dissimilar Metals

In terms of process requirements for the inertia or direct-drive friction welding of dissimilar material combinations, it is typically found that the welding parameters tend to be closer to those used for the higher-strength material in the combination. This results in a significant loss of the lower-strength material during welding and upsetting. As in any other joining process, it must also be remembered that the strength of the bond can be no higher than that of the lower-strength material. The combinations of metals that have been joined using the friction welding processes are shown in Fig. 4. Other combinations may be possible, with additional procedure development.

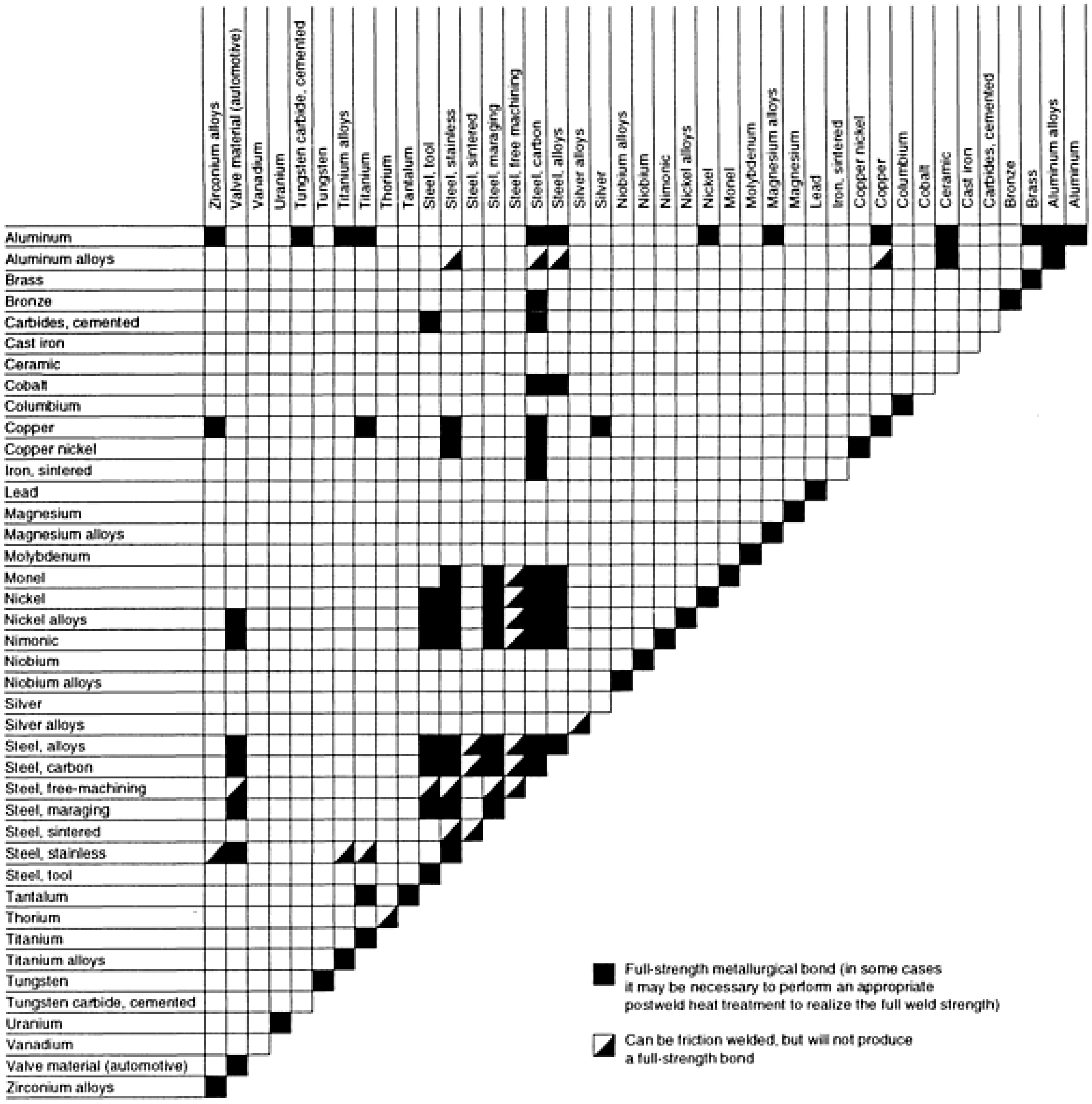

	Zirconium alloys	Valve material (automotive)	Vanadium	Uranium	Tungsten carbide, cemented	Tungsten	Titanium alloys	Titanium	Thorium	Tantalum	Steel, tool	Steel, stainless	Steel, sintered	Steel, maraging	Steel, free machining	Steel, carbon	Steel, alloys	Silver alloys	Silver	Niobium alloys	Niobium	Nimonic	Nickel alloys	Nickel	Monel	Molybdenum	Magnesium alloys	Magnesium	Lead	Iron, sintered	Copper nickel	Copper	Columbium	Cobalt	Ceramic	Cast iron	Carbides, cemented	Bronze	Brass	Aluminum alloys	Aluminum
Aluminum	■				■		■	■								■	■							■			■					■			■				■	■	■
Aluminum alloys												◪				◪	◪															◪			■					■	
Brass																																							■		
Bronze																■																						■			
Carbides, cemented											■					■																									
Cast iron																																									
Ceramic																																									
Cobalt																■	■																								
Columbium																																	■								
Copper	■							■				■				■			■													■									
Copper nickel												■				■															■										
Iron, sintered																■																									
Lead																													■												
Magnesium																												■													
Magnesium alloys																											■														
Molybdenum																										■															
Monel												■		■	◪	■	■								■																
Nickel											■	■		■	◪	■	■							■																	
Nickel alloys		■									■	■		■	◪	■	■						■																		
Nimonic		■									■	■		■	◪	■	■					■																			
Niobium																					■																				
Niobium alloys																				■																					
Silver																																									
Silver alloys																		◪																							
Steel, alloys		■									■	■	◪	■	◪	■	■																								
Steel, carbon		■									■	■	◪	■	◪	■																									
Steel, free-machining		◪									◪	◪		◪	◪																										
Steel, maraging		■									■	■		■																											
Steel, sintered												◪	◪																												
Steel, stainless	◪	■					◪	◪				■																													
Steel, tool											■																														
Tantalum								■		■																															
Thorium									◪																																
Titanium								■																																	
Titanium alloys							■																																		
Tungsten						■																																			
Tungsten carbide, cemented																																									
Uranium				■																																					
Vanadium																																									
Valve material (automotive)		■																																							
Zirconium alloys	■																																								

■ Full-strength metallurgical bond (in some cases it may be necessary to perform an appropriate postweld heat treatment to realize the full weld strength)

◪ Can be friction welded, but will not produce a full-strength bond

FIG. 4 MATERIAL COMBINATIONS THAT ARE WELDABLE BY FRICTION WELDING PROCESSES. LIST OF ALLOYS WAS COMPILED FROM AVAILABLE FRICTION WELDING LITERATURE. SOURCE: ANSI/AWS C6.1-89

Two of the most common dissimilar joints produced using the friction welding processes are:

- COPPER TO ALUMINUM
- ALUMINUM TO STEEL OR STAINLESS STEEL

In such applications, some surface preparation may be required prior to joining, because of oxidation of the aluminum surface. In most cases where aluminum is used in a joint, a dry machining or chemical etching of the aluminum is preferred. As mentioned previously, a majority of the forging (material loss) will occur in the aluminum, especially for the aluminum-to-steel joint.

The joining of aluminum to ceramics has also been performed, primarily for the electrical industry. Steel has also been attached to ceramics using an aluminum interlayer. This is performed by first welding the steel to the aluminum, followed by welding the bimetal joint to the ceramic. Using an interlayer for applications such as this is common when two incompatible materials require joining. The interlayer materials must be selected on the basis of their forgeability and compatibility with the two parent materials.

Procedure Development and Practice Considerations for Inertia and Direct-Drive Friction Welding

Tim Stotler, Edison Welding Institute

Selected References

- *METALS HANDBOOK,* 9TH ED., VOL 6, AMERICAN SOCIETY FOR METALS
- "RECOMMENDED PRACTICES FOR FRICTION WELDING," ANSI/AWS C6.1-89, AMERICAN WELDING SOCIETY
- T. STOTLER AND J.E. GOULD, "INERTIA WELDING OF STEEL TO A STRUCTURAL CERAMIC," 73RD AWS MEETING (CHICAGO, IL), AWS, 1992
- A. SUZUMURA, T. ONZAWA, A. OHMORI, AND Y. ARATA, "FRICTION WELDING OF CERAMICS TO ALUMINUM-CLAD STEEL," DOCUMENT NO. III-890-87, 40TH ANNUAL ASSEMBLY OF INTERNATIONAL INSTITUTE OF WELDING (SOFIA, BULGARIA), IIW, 1987
- V.I. VILL, *FRICTION WELDING OF METALS,* AMERICAN WELDING SOCIETY, 1962
- *WELDING HANDBOOK,* 7TH ED., VOL 3, AMERICAN WELDING SOCIETY
- *WELDING HANDBOOK,* 7TH ED., VOL 4, AMERICAN WELDING SOCIETY
- *WELDING HANDBOOK,* 8TH ED., VOL 2, AMERICAN WELDING SOCIETY

Procedure Development and Practice Considerations for Ultrasonic Welding

Janet Devine, Sonobond Ultrasonics, Inc.

Introduction

ULTRASONIC WELDING (USW) is effectively used to join both similar and dissimilar metals with lap-joint welds. Various metals differ in weldability based on their compositions and properties. The metals that are considered more difficult to weld are those that require either high power or long weld times, or both, and those that incur operational problems, such as tip sticking or short tip life. Most ductile metals can be welded together. Figure 1 identifies some of the monometallic and bimetallic combinations that currently can be ultrasonically welded on a commercial basis.

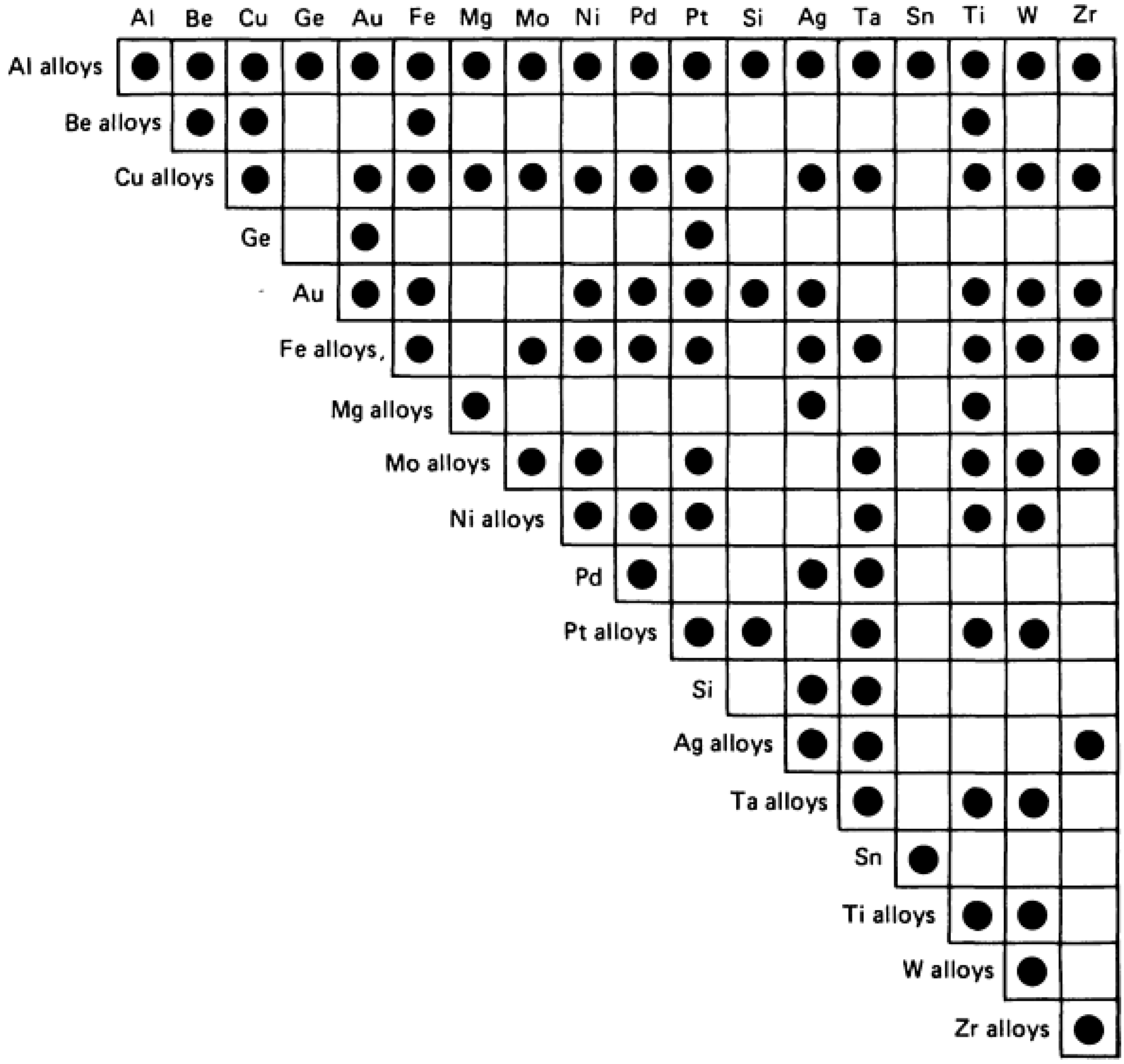

	Al	Be	Cu	Ge	Au	Fe	Mg	Mo	Ni	Pd	Pt	Si	Ag	Ta	Sn	Ti	W	Zr
Al alloys	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Be alloys		●	●			●										●		
Cu alloys			●		●	●	●	●	●	●	●		●	●		●	●	●
Ge					●						●							
Au					●	●			●	●	●	●	●			●	●	●
Fe alloys						●		●	●	●	●		●	●		●	●	●
Mg alloys							●						●			●		
Mo alloys								●	●		●			●		●	●	●
Ni alloys									●	●	●			●		●	●	
Pd										●			●	●				
Pt alloys											●	●		●		●	●	
Si													●	●				
Ag alloys													●	●				●
Ta alloys														●		●	●	
Sn															●			
Ti alloys																●	●	
W alloys																	●	
Zr alloys																		●

FIG. 1 CHART SHOWING ULTRASONICALLY WELDABLE METAL COMBINATIONS. BLANK AREAS IN THE CHART REPRESENT COMBINATIONS THAT HAVE NOT BEEN SUCCESSFULLY JOINED OR IN WHICH WELDING HAS NOT BEEN ATTEMPTED. SOURCE: AMERICAN WELDING SOCIETY

In ultrasonic welding, the temperature at the weld is not raised to the melting point. Therefore, there is no nugget similar to that formed in resistance welding. Weld strength is equal to the strength of the base metal. However, ultrasonic welding is typically restricted to relatively thin materials (foil or extremely thin gage thicknesses). The process, which can also be used for joining plastics, is finding wider applications in that field than in metals joining.

Ultrasonic welding is used extensively in the electronics, aerospace, and instrument industries. It is also used for producing packages and containers and for sealing applications.

This article describes procedure considerations for the ultrasonic welding of specific material types. For descriptions of USW processes and equipment, see the article "Ultrasonic Welding" in this Volume.

Procedure Development and Practice Considerations for Ultrasonic Welding

Janet Devine, Sonobond Ultrasonics, Inc.

Difficult-to-Weld Alloys

Satisfactory welds can be made in certain ferrous alloys. Frequently, however, low tip life and poor peel strength are encountered.

Carbon and Low-Alloy Steels. A weld between a 0.9% C steel and iron is shown in Fig. 2. Because many other processes are available for joining ferrous metals, ultrasonic welding usually is not chosen. Copper flash or copper plating on steel will sometimes improve the quality of the weld.

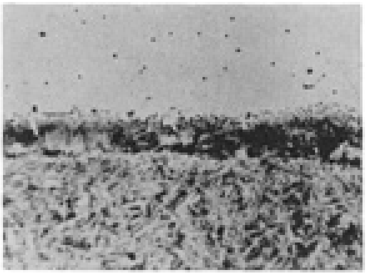

FIG. 2 PHOTOMICROGRAPH SHOWING WELD INTERFACE CROSS SECTION OF 0.81 MM (0.032 IN.) THICK HIGH-CARBON (0.9 WT% C) STEEL ULTRASONICALLY WELDED TO 0.81 MM (0.032 IN.) THICK IRON. 1000×

High-Strength Steels. Ultrasonic welding of high-strength steels is not considered to be viable.

Stainless Steels. A weld in 0.35 mm (0.014 in.) thick half-hard AISI 300-series stainless steel is shown in Fig. 3. Although it is possible to produce welds in stainless steels using ultrasonic welding, it is not the process of choice.

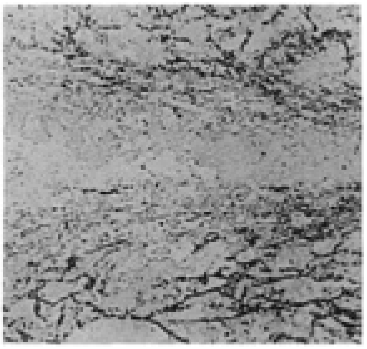

FIG. 3 PHOTOMICROGRAPH SHOWING WELD INTERFACE CROSS SECTION OF TWO 0.36 MM (0.014 IN.) THICK PIECES OF HALF-HARD AISI 300-SERIES STAINLESS STEEL THAT ARE ULTRASONICALLY WELDED TO EACH OTHER. 500×

Procedure Development and Practice Considerations for Ultrasonic Welding

Janet Devine, Sonobond Ultrasonics, Inc.

Weldable Materials

Aluminum Alloys. All combinations of aluminum alloys are ultrasonically weldable. The significant diffusion that is typical of ultrasonic welds in aluminum is shown in Fig. 4. In general, the hardness of work-hardenable aluminum should range from quarter-hard to half-hard. Annealed materials exhibit high deformation when ultrasonically welded.

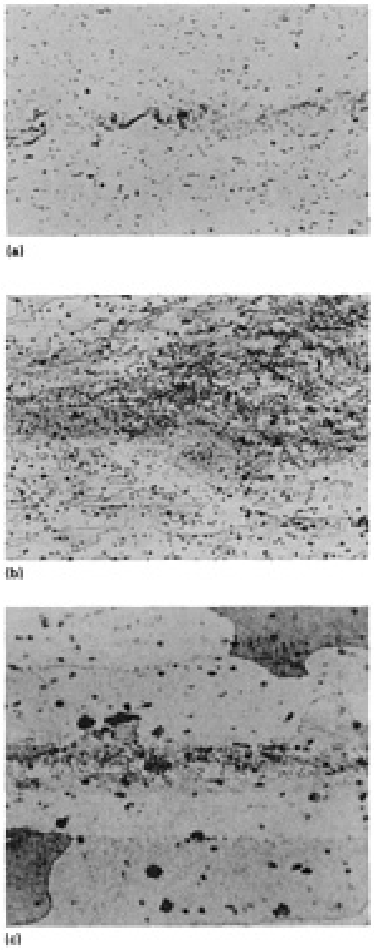

FIG. 4 PHOTOMICROGRAPHS OF WELD INTERFACE CROSS SECTIONS OF SELECTED ULTRASONICALLY WELDED ALUMINUM ALLOYS. (A) TWO PIECES OF 0.30 MM (0.012 IN.) THICK 1100 H14 ALLOY. 500×. (B) TWO PIECES OF 0.81 MM (0.032 IN.) THICK 2024-T3 BARE ALUMINUM ALLOY. 150×. (C) TWO PIECES OF 1.02 MM (0.040 IN.) THICK 2020 ALLOY. 750×

Some aluminum alloys (particularly the 1100 series and 2036 alloy) tend to stick to the welding tip. Tip sticking can sometimes be overcome by using high power and a short welding time. Hardened tool steel tips exhibit the most satisfactory results. Other solutions include using a mechanical stripper where light sticking occurs or placing a thin-gage steel shim between the vibrating tip and the weldment.

Various tip geometries and surface finishes can be used for welding aluminum, including circular radiused, flat cross-hatched, bulls-eye, and serrated types. Tips can be circular, rectangular, line-shaped, and so on, as suits the end use of the workpiece.

The aluminum adjacent to the tip should be less than 2.5 mm (0.1 in.) thick for the highest strength results. The material to which it is welded does not require a thickness limitation. Solid wires with diameters up to 3.18 mm (0.125 in.) or stranded wires with diameters up to 6.4 mm (0.25 in.) also can be welded.

Commercial applications of ultrasonically welded aluminum alloys include:

- SEAM WELDING OF LIGHT-GAGE FOILS
- WELDING OF STRANDED WIRE TO TERMINALS
- JOINING MULTIPLE LAYERS OF ALUMINUM FOIL TO A CAPACITOR TERMINAL
- JOINING ALUMINUM FRAMES TO ALUMINUM GRILLES FOR HEATING DUCTS
- AIRCRAFT PANEL WELDING FOR THE A-10 AIRCRAFT

- ALUMINUM INTERCONNECT WELDING ON PHOTOVOLTAIC PANELS

Reactive and Refractory Base Metals. The photomicrographs in Fig. 5 show ultrasonic welds in molybdenum. This material, as well as other reactive and refractory metal alloys, are difficult, but not impossible, to join using the ultrasonic welding process. In general, weldability appears to depend on the hardness and the ductility of the materials. Softer, more-ductile materials are more likely to weld, whereas hard, brittle materials tend to exhibit stress crocks under ultrasonic activation.

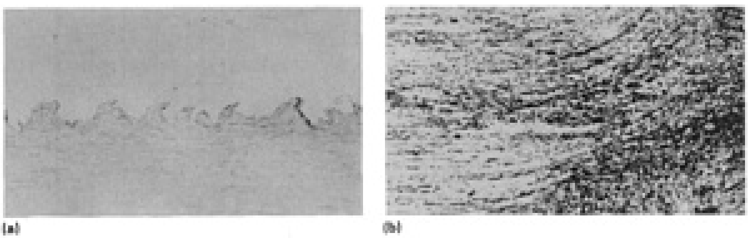

FIG. 5 PHOTOMICROGRAPHS OF WELD INTERFACE CROSS SECTIONS OF SELECTED ULTRASONICALLY WELDED MOLYBDENUM ALLOYS. (A) 0.13 MM (0.005 IN.) THICK NICKEL JOINED TO 0.51 MM (0.02 IN.) THICK MOLYBDENUM. 200×. (B) TWO PIECES OF 0.20 MM (0.008 IN.) THICK ARC CAST MOLYBDENUM WELDED TOGETHER. 140×

The welding of materials in this class is limited to thin sections with thicknesses under 0.4 mm (0.016 in.).

Titanium alloys, which have a very hard oxide layer, should be chemically etched before welding is attempted. Oxide removal should occur not more than 24 h prior to welding, unless the material is sealed and stored in an inert atmosphere.

Copper Alloys. The greatest industrial use of ultrasonic welding is in the joining of copper and brass materials. Copper alloys can be joined together or to other materials, such as aluminum, with great success. Various tip geometries and surface finishes can be used.

Welds in copper sheet that is up to 2 mm (0.08 in.) thick provide excellent joint strength. Stranded wires for wire harness applications represent one common use. Bundles that have a cross-sectional area up to 30 mm^2 (0.05 in.^{2}) are readily weldable.

The high thermal conductivity of copper is not the deterrent to ultrasonic welding that it is with fusion welding. Applications include:

- STRANDED WIRE TO STRANDED WIRE JOINTS FOR AUTOMOTIVE, TRUCK, AND APPLIANCE HARNESSES
- STRANDED WIRES JOINED TO TERMINALS FOR AUTOMOTIVE AND OTHER ELECTRICAL HARNESSES
- JOINING OF STRANDED WIRES TO SOLID WIRES
- SOLID WIRE TO SOLID WIRE JOINTS IN COMMUTATORS FOR AUTOMOTIVE ENGINE STARTER MOTORS
- JOINING OF COPPER STRIP TO COPPER STRIP FOR ELECTRICAL GROUNDING GRIDWORK
- JOINING OF FLAT WIRE TO FLAT WIRE FOR AUTOMOTIVE ENGINE STARTERS AND COILS
- JOINING OF FLAT COPPER SLUG TO ALUMINUM PLATE FOR AUTOMOTIVE ENGINE IGNITION MODULES
- JOINING OF ENAMELED WIRE TO STEEL HOUSING FOR SOLENOID COIL GROUND
- JOINING OF COPPER TUBE CLOSURES FOR HEATING, VENTILATING, AND AIR CONDITIONING (HVAC) INDUSTRY

- JOINING OF BRASS U-SHAPED SECTION TO A BRASS PLATE FOR A BRUSH HOLDER
- JOINING OF COPPER STEEL LAMINATE TO ITSELF FOR A COMMUNICATIONS CABLE SHEATHING

Precious Metals. Most of the precious metals (that is, gold, silver, platinum, and their alloys) are weldable ultrasonically. Materials plated with such metals are also ultrasonically weldable.

Nickel-Base Alloys. Many nickel-base alloys and nickel-plated materials are ultrasonically weldable. Figure 6 is a photomicrograph of nickel ultrasonically welded to gold-plated Kovar.

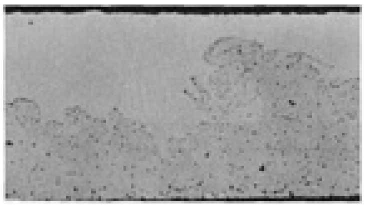

FIG. 6 PHOTOMICROGRAPH SHOWING WELD INTERFACE CROSS SECTION OF 0.08 MM (0.003 IN.) THICK NICKEL ULTRASONICALLY WELDED TO 0.08 MM (0.003 IN.) THICK GOLD-PLATED KOVAR. 300×

Figure 7 shows a weld in the nickel-base alloy Inconel X. A heat-affected zone (HAZ) is produced in nickel. Significant penetration across the interface is exhibited. Nickel sulfamate plating gives better results than electroless nickel plating when joining nickel-plated materials.

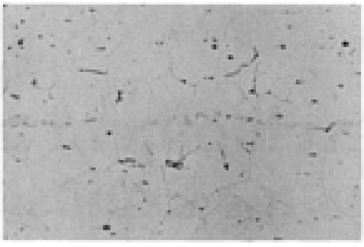

FIG. 7 PHOTOMICROGRAPH SHOWING WELD INTERFACE CROSS SECTION OF TWO 0.30 MM (0.012 IN.) THICK PIECES OF SOLUTION HEAT TREATED AND AGED INCONEL X THAT WERE ULTRASONICALLY WELDED TO EACH OTHER. 150×

Procedure Development and Practice Considerations for Ultrasonic Welding

Janet Devine, Sonobond Ultrasonics, Inc.

Dissimilar Metal (Nonferrous-to-Nonferrous) Combinations

Perhaps one of the most useful characteristics of ultrasonic welding is its ability to weld dissimilar material combinations. Primarily used to join copper and aluminum alloys, this technique promotes the advantages of using aluminum as a backing plate and the solderability of copper for attaching circuitry. [For example, copper slugs are ultrasonically welded to an aluminum backing plate for automotive-engine ignition modules and for antilock braking system (ABS) sensors]. The copper is more suitable for soldering the circuitry and provides a better match to the expansion coefficient of the chip carrier. This is important if the stress of thermal cycling is to be overcome.

When using aluminum for electrical wiring applications, one of the major causes of failure has occurred at the mechanical crimp between the aluminum wire and the connector. Problems have involved corrosion and loosening of the wire at the joint, as a result of thermal stress. In some applications, a copper "pigtail" inserted between the aluminum wire and the connector can eliminate this problem.

Other typical applications include:

- CONSTRUCTION OF A LITHIUM BATTERY, WHICH REQUIRED THE JOINING OF EITHER THIN ALUMINUM, COPPER, OR NICKEL SHEETS TO EXPANDED NICKEL MESH. ALL THREE VARIATIONS PROVIDED ELECTRICALLY AND THERMALLY SOUND JOINTS OF ADEQUATE STRENGTH
- COPPER WIRES WELDED TO COPPER-FLASHED STEEL PLATE TO MAKE A CARBON BRUSH HOLDER FOR A LOCOMOTIVE ELECTRIC TRACTION MOTOR. THE ULTRASONIC WELDED ASSEMBLY REPLACED A COMPLICATED PROCESS THAT REQUIRED RIVETING AND FORMING OF THE "BRIDGE" FOR THE BRUSH
- COPPER WIRES WELDED TO COPPER-PLATED STEEL FOR AN AUTOMOBILE SEAT SWITCH
- COPPER WIRES WELDED TO STAINLESS STEEL AS PART OF A PHOTOVOLTAIC ASSEMBLY
- COPPER WIRES AND STRIPS WELDED TO SILVER-PLATED TERMINALS

Procedure Development and Practice Considerations for Explosion Welding

Vonne D. Linse, Edison Welding Institute

Introduction

EXPLOSION WELDING (EXW), like all other welding or joining processes, has a well-defined set of input parameters or conditions that must fall within certain limits for the desired weld quality to be achieved. Contrary to other joining processes, however, the EXW occurs so rapidly (within a time period of a few microseconds) that there is no time for in-process adaptive control or modification of the welding parameters or conditions. Therefore, a total quality joining approach must be taken to ensure that proper process parameters are applied and controlled to consistently achieve high-quality welds. The quality must be built into the prewelding stages of the preparation of the process materials and setups before welding takes place. It is necessary to be thoroughly familiar with every aspect of the process and the quality level that must be applied in order to ensure the consistent quality of the end-product weld.

This article provides a general overview of the important mechanistic aspects of EXW, the process-material interactions, and the critical aspects or parameters that must be controlled. The procedure for ensuring control is also discussed. This type of approach to setting and controlling process parameters is particularly important when explosion welding dissimilar or metallurgically incompatible metals. Metallurgically incompatible combinations, such as titanium and steel, aluminum and steel, and zirconium and steel, will form brittle intermetallic compounds at the explosion weld interface if excessive energies are used during welding. The intermetallic compounds result in poor-quality welds. Setting and controlling process parameters are also important when welding the newer advanced materials, which have little tolerance for parametric variation (Table 1). Information about the principles underlying this process is available in the article "Fundamentals of Explosion Welding" in this Volume. Information about EXW equipment and applications can be found in the article "Explosion Welding" in this Volume.

TABLE 1 COMMERCIALLY USED METALS AND ALLOYS THAT CAN BE JOINED WITH EXW

METAL NO. 1	METAL NO. 2(A)														
	ZIRCONIUM	**MAGNESIUM**	**COBALT ALLOYS**	**PLATINUM**	**GOLD**	**SILVER**	**NIBIUM**	**TANTALUM**	**TITANIUM**	**NICKEL ALLOYS**	**COPPER ALLOYS**	**ALUMINUM ALLOYS**	**STAINLESS STEELS**	**ALLOY STEELS**	CARBON STEELS
CARBON STEELS	X	X	X	. . .	X	X	X	X	X	X	X	X	X	X	**X**
ALLOY STEELS	X	X	X	. . .	. . .	. . .	X	X	X	X	X	X	X	X	
STAINLESS STEELS	X	. . .	X	. . .	X	X	X	X	X	X	X	X	X		
ALUMINUM ALLOYS	X	X	. . .	. . .	X	X	X	X	X	X	X	X			
COPPER ALLOYS	X	. . .	. . .	. . .	X	X	X	X	X	X	X				
NICKEL ALLOYS	X	X	X	X	X	. . .	X	X	X	X					
TITANIUM	X	X	. . .	. . .	. . .	X	X	X	X						
TANTALUM	X	. . .	. . .	. . .	X	. . .	X	X							
NIOBIUM	. . .	. . .	. . .	X	. . .	. . .	X								
SILVER	. . .	. . .	. . .	. . .	. . .	X									
GOLD	. . .	. . .	. . .	. . .	. . .										
PLATINUM	. . .	. . .	. . .	X											
COBALT ALLOYS	. . .	. . .	X												
MAGNESIUM	X	X													
ZIRCONIUM	**X**														

Source: Adapted from Ref 1

(A) EXES INDICATE EXW IS AN APPLICABLE PROCESS.

Reference

1. *WELDING HANDBOOK*, VOL 2, 8TH ED., R.L. O'BRIEN, ED., AWS, 1991, P 772

Procedure Development and Practice Considerations for Explosion Welding

Vonne D. Linse, Edison Welding Institute

Critical Process Parameters

In order to fully understand EXW and its control, it is helpful to know the general characteristics of the process, particularly the rates, pressures, and energies involved. The following are the general process parameters:

TRAVERSE RATE, M/S (FT/S)	**2000-3000 (6600-9800)**
PRESSURE, GPA (10^6 PSI)	**1.5-6.0 (0.22-0.88)**
PRESSURE TIME DURATION, μS	**5-20**
ENERGY DEPOSITION/UNIT AREA, J/CM^2 (BTU/FT^2)	**100-300 (90-260)**
ENERGY DEPOSITION RATE, MJ/S (BTU/H)	**20-90 (70 × 10^6 - 310 × 10^6)**

By comparison, standard fusion welding processes will typically have traverse rates of a fraction of a meter/second and energy depositions/unit area in the range of 10,000 to 30,000 J/cm^2 (9000 to 26,000 Btu/ft^2). The low-energy input of EXW and the high rate of deposition at the weld interface allow the weld to be made with an absolute minimum of thermal excursion and no resultant beat-affected zone (HAZ).

Procedure Development and Practice Considerations for Explosion Welding

Vonne D. Linse, Edison Welding Institute

Fundamentals of EXW

The overall mechanisms of EXW are well known and are shown in Fig. 1. In general, the energy from an explosive accelerates one of the two components being welded (prime or clad) across a gap or standoff to collide with the second component (base) and form a solid-state metallurgical weld or bond between the two. The various aspects or mechanisms of the process have been investigated and reported over the years; however, the process is still largely controlled by most commercial users on a more or less empirical basis. Important process-input parameters are selected primarily on the basis of past experience and usually are not calculated from basic principles of theory. The following sections ("Explosives and Explosive Detonation" , "Component Acceleration" , and "Component Collision") cover the latter approach to understanding and controlling EXW on a more fundamental basis. This requires breaking down and treating each mechanistic aspect of the process in a step-by-step sequence.

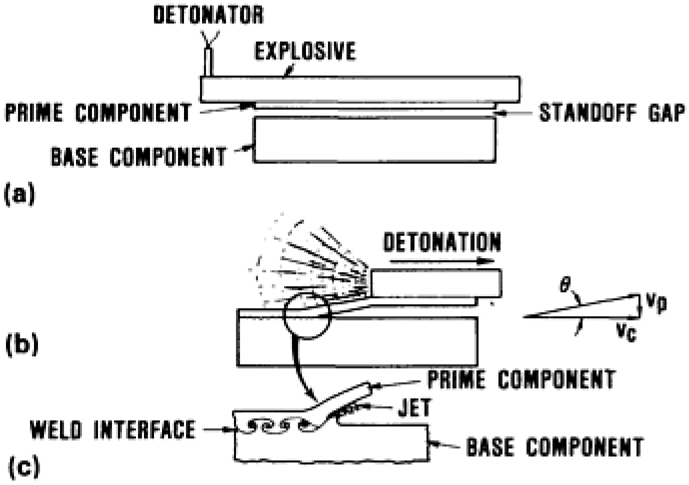

FIG. 1 SCHEMATIC SHOWING MECHANICS OF EXW. (A) ALIGNMENT OF COMPONENTS TO BE JOINED BEFORE DETONATION. (B) MOTION OF COMPONENTS AT DETONATION TO FORM WELD PLUS VECTOR DIAGRAM OF VELOCITY COMPONENTS. (C) CLOSE-UP VIEW OF WAVELIKE WELD INTERFACE CHARACTERISTIC OF EXW PROCESS

Explosives and Explosive Detonation

The explosive is the energy source for the process. It releases its energy in a very specific manner during detonation. It is, therefore, necessary to first understand the manner in which the welding explosive detonates and how its energy is delivered to the prime or cladding component.

In almost all instances, the explosives used for welding can be classified as nonideal. They do not detonate in a manner that can be easily defined by the physics equations-of-state typically used for extremely-high-velocity military or weapons explosives. This results from the fact that these explosives are almost always granular in nature, with the primary ingredient typically being ammonium nitrate (NH_4NO_3). Other ingredients (for example, sensitizers, density control agents, and diluents) are then blended to produce an explosive that yields the desired detonation velocity, pressure, and duration of pressure.

The detonation characteristics of the explosive are controlled by its composition, particle size distribution, pour or packing density, layer thickness, confinement, moisture content, and, in some instances, age. The detonation performance of the explosive should be tailored to match the metals or alloy combination being welded with respect to the density, thickness, and sonic or shock wave velocity of the prime component, as well as the sonic velocity of the base component (if it is different from that of the prime component).

Detonation is a progressive event that starts at the detonator and progresses into the undetonated explosive at a rate known as the detonation velocity. Within the detonation region itself (Fig. 2), the detonation front consists of an initial high-pressure shock wave intense enough to trigger the high-pressure chemical reaction that follows in the reaction zone. In an ideal explosive, the reaction goes almost instantaneously to completion; therefore, the reaction zone is extremely short (usually <1 mm, or 0.04 in.). In a nonideal explosive, however, the reaction zone is much longer (5 to 25 mm, or 0.2 to 1 in.) because the reaction takes longer to complete. While the pressure in the reaction zone of a nonideal explosive is much lower than that in the initial shock wave, it has a much longer duration in which the high-pressure gases in the zone can deliver their impulse to accelerate the prime component. Behind the reaction zone is the third and final region of relatively low-pressure expanding gases, which contribute little, if any, to the impulse delivered to the acceleration of the prime component. Each type of explosive has characteristic detonation velocity, pressure, and duration factors that can be applied to predict the specific velocity, pressure, and duration for a given thickness and density of that explosive. For welding, it is usually desirable for nonideal explosives to detonate in the range of 2000 to 3000 m/s (6600 to 9800 ft/s). If the explosive is manufactured to exacting specifications, it is possible to control the detonation velocity within ±2% of the target terminal velocity. With nonideal explosives, however, such factors as run-up regions near the initiation point and predetonation compression due to the precursor shock wave in the prime component ahead of the EXW collision front can cause nonsteady-state detonation conditions, which in turn can result in nonuniform weld quality.

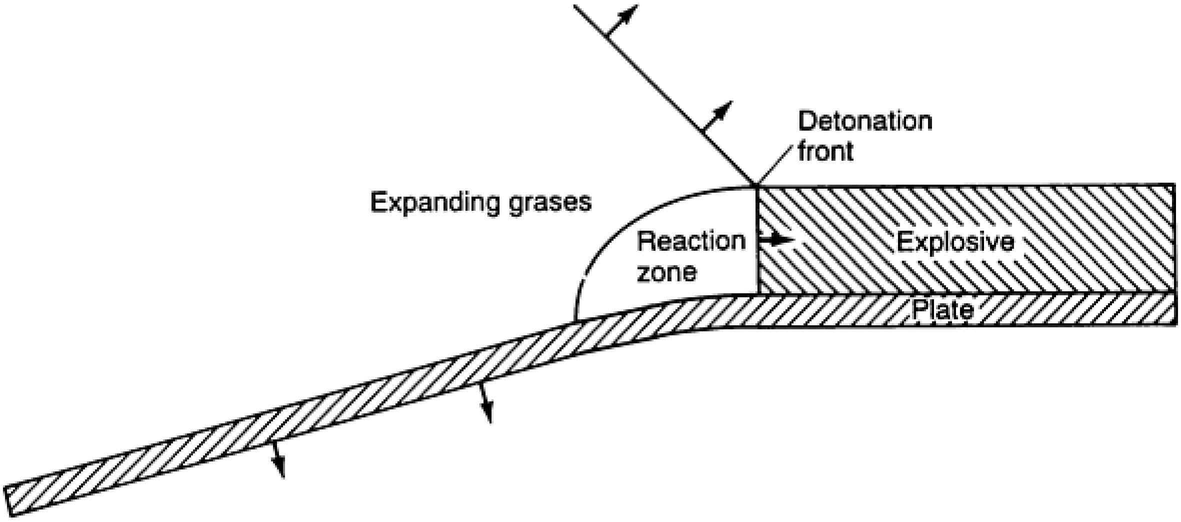

FIG. 2 SCHEMATIC SHOWING DETONATION REGIONS IN A NONIDEAL EXPLOSIVE AND THE ASSOCIATED ACCELERATION OF THE PRIME COMPONENT

Component Acceleration

For welding, the explosive is distributed over the top or exposed surface of the prime component and is detonated such that the detonation front progresses in a grazing or tangential manner over the length of the prime component. In doing so, the explosive delivers a specific level of impulse to the prime component that causes it to be accelerated to a terminal bend angle, θ, and velocity, v_p, as shown in Fig. 1. For explosion welding, the acceleration region before terminal conditions are reached is the important region.

The acceleration of the prime component actually results from the two separate zones of the detonation region in the explosive. First is the contribution of the high-pressure shock front in the explosive. This imparts a high-pressure shock wave into the prime component. The shock wave traverses back and forth, or rattles, between the opposite surfaces of the component, causing it to accelerate quickly in its early stages. The effects of this contribution diminish quickly and cause the component to reach only approximately 80% of the full-potential bend angle (Fig. 3) and velocity. As the effects of the shock component begin to diminish, the expanding high-pressure gas from the trailing reaction zone takes over and continues the acceleration more gradually until the terminal angle and velocity are obtained when the reaction is completed. This contribution (Fig. 3) plus the shock contribution yields the total impulse delivered to the prime component. In an ideal explosive, nearly all of the acceleration comes from the extremely high pressure of the shock wave and very little from the trailing reaction zone; therefore, the primary component reaches terminal conditions much more quickly with an ideal or near-ideal explosive.

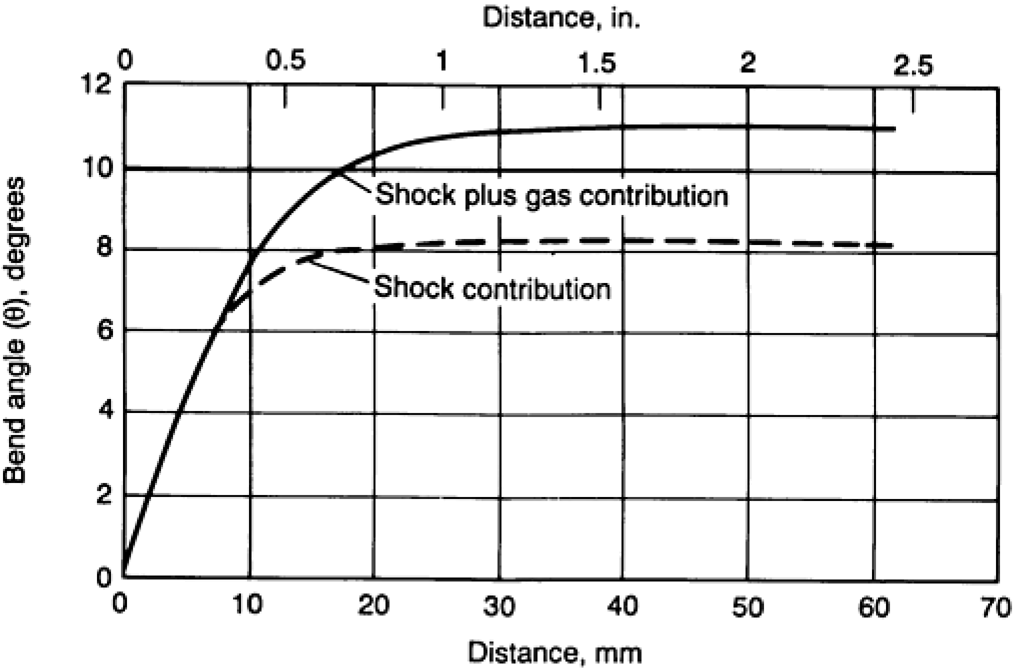

FIG. 3 PLOT OF PRIME PLATE COMPONENT BEND ANGLE VERSUS DISTANCE TO SHOW CONTRIBUTION OF SHOCK WAVE AND GAS PRESSURE TO ACCELERATION

The prime component acceleration depends on the following parameters:

- EXPLOSIVE PRESSURE (SHOCK AND GAS)
- DURATION OF THE EXPLOSIVE PRESSURE PULSE
- SONIC OR SHOCK WAVE VELOCITY OF THE PRIME COMPONENT
- THICKNESS OF THE PRIME COMPONENT
- DENSITY OF THE COMPONENT MATERIAL

With the above variables, the acceleration characteristics of the prime component can be calculated. In addition, techniques exist to directly measure the acceleration of a prime component (Ref 2). Precise knowledge of the acceleration achieved for a specific prime component/explosive combination is extremely important. It allows the exact standoff or separation distance to be set between the prime and base components to obtain the required collision angle and prime component velocity in the next step of EXW (that is, the collision and welding step).

Component Collision

Once the accelerated prime component crosses the standoff gap, it collides with the stationary base component. The exact collision angle and prime component velocity achieved at the time of the collision will, of course, depend on the combination of the acceleration characteristics of the prime component and the standoff distance selected for the system. The standoff distance is usually selected to ensure that the collision occurs while the prime component is still undergoing acceleration, usually in its final stages.

The collision is characterized by:

$$V_p = 2v_c \sin\left(\frac{q}{2}\right) \qquad \textbf{(EQ 1)}$$

where v_p is the velocity of the prime component or the relative velocity of the prime and base components; v_c is the collision point velocity of the welding front, which also equals the detonation velocity, v_d, of the explosive; and θ is the collision angle. If any two of the above variables are known, it is then easy to calculate the third variable. The relative component velocity, v_p, at the time of the collision is very important in that it determines the equation-of-state pressure generated at the collision point. This pressure, which must be sufficient to generate hydrodynamic flow of both materials at the collision point, can be determined from the characteristic pressure versus plate (or particle) velocity for the material or materials being welded. Figure 4 shows a high pressure versus particle velocity plot for a similar material collision (Fig. 5). Initially, the pressure required for dynamic yielding of the two materials being welded is calculated. This pressure is then located on the curve, and a mirror image of the curve is drawn through the selected pressure. The point at which the mirror-image curve intersects the zero pressure line or abscissa is the minimum prime plate velocity necessary to obtain the desired pressure and accompanying dynamic flow and jetting necessary for welding to occur.

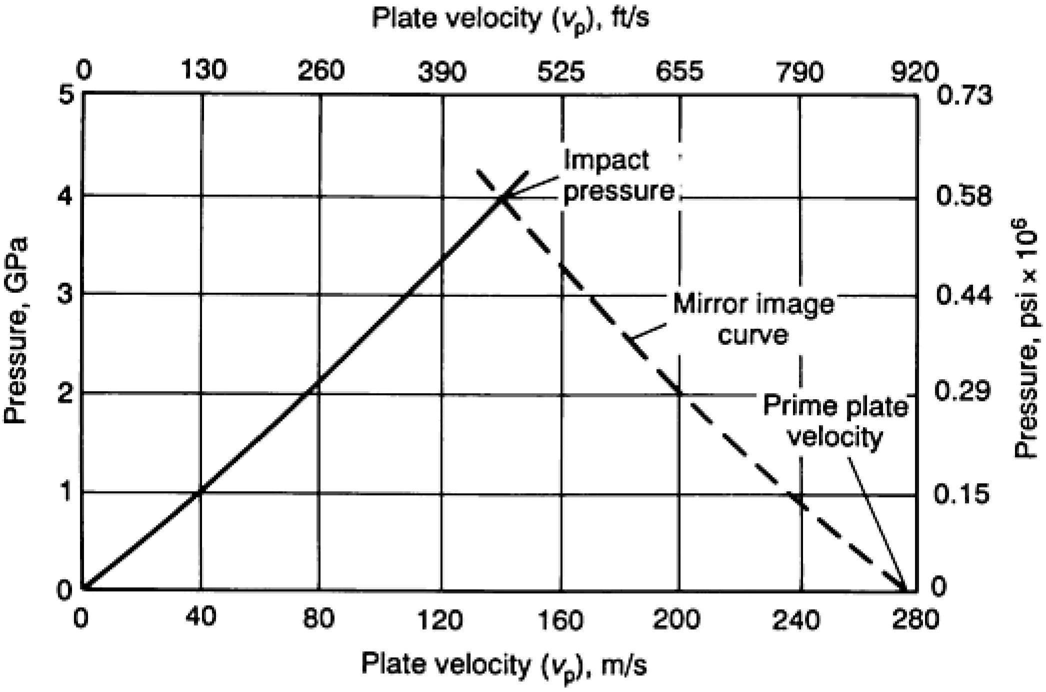

FIG. 4 PLOT OF HIGH PRESSURE VERSUS PARTICLE VELOCITY TYPICALLY USED TO DETERMINE THE PRIME COMPONENT VELOCITY NECESSARY TO ACHIEVE DYNAMIC YIELDING AND WELDING

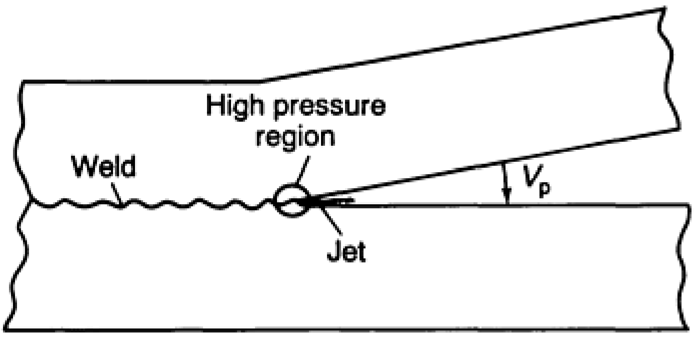

FIG. 5 JET FORMATION AT THE COLLISION POINT TO GENERATE WELD DURING EXW

Reference cited in this section

2. E.G. SMITH, JR. AND V.D. LINSE, THE ACCELERATION CHARACTERISTICS OF EXPLOSIVELY

DRIVEN FLYER PLATES, *PROC. SIXTH INT. CONF. HIGH ENERGY RATE FABRICATION* (ESSEN, WEST GERMANY), 13-16 SEPT 1977

Procedure Development and Practice Considerations for Explosion Welding

Vonne D. Linse, Edison Welding Institute

Jetting and Weld Formation

When the required minimum collision pressure, a minimum duration of the pressure pulse, and a minimum collision angle are obtained, the surfaces of the two materials being joined locally flow in the immediate region of the collision point. Because of the angular geometry of the collision, these surfaces at the collision point are forced to reverse their flow direction with severe shearing flow and to be ejected out ahead of the advancing collision point in the form of a fine fingerlike mist (that is, a jet) (Fig. 5). Under the extremely high hydrodynamic pressures involved, this jet, which consists of a combination of both parent materials, has an extremely high temperature. When the collision conditions of the weld are within an appropriate range (as discussed in the section "Parametric Limits for Welding" in this article), most of the jet, with its attendant heat, is ejected out of the weld interface with little or no effect on the remaining weld zone. Of greatest importance, however, is that the surface-scarfing action of the jet removes the surface layers containing the oxides and other contaminants that typically inhibit solid-state interaction. This allows the two metals to directly weld or bond under the substantial residual pressure of the collision, which continues to act from some period of time on the two metals at the weld interface. This residual pressure on the interface after the collision also counteracts any reflected stress waves from separating the weld while it is forming.

Procedure Development and Practice Considerations for Explosion Welding

Vonne D. Linse, Edison Welding Institute

Weld Characteristics

Of particular importance is the nature or characteristics of the explosion weld itself. When properly made, it is, in essence, a solid-state bond or weld without any HAZ to degrade its strength, as is typical of all other heat-dependent welding or joining processes.

Weld Interface. Morphologically, two types of explosion weld interfaces can be created. First is the flat or straight-line interface, which occurs only when the collision conditions are at or near the lower limits for jetting and welding. This type of condition is typically avoided in most instances because a slight drop in collision conditions or parameters during welding would stop the welding and leave a nonwelded region.

In most cases, the explosion weld interface will have a wave morphology that is exclusively characteristic of the process. A typical, wavy explosion weld interface between copper and niobium is shown in Fig. 6. This wave formation, which has been the subject of considerable investigation (Ref 3, 4), results from the localized dynamic compression of the two metals in the region immediately adjacent to the collision point and weld interface. The compression causes localized bulk flow of the metals, which become unstable and make the weld interface oscillate above and below the theoretical plane of the interface. The size of the wave pattern is the one postwelding characteristic that can be used to determine the collision conditions used to make the weld and to allow for adjustments in subsequent welds of the same material type and thickness. The wavelengths or frequency of the interfacial wave in an explosion weld is directly related to the thickness of the prime or cladding plate and its angle of collision with the base plate (Ref 5, 6), as shown in Fig. 7. For practical purposes, this relationship is relatively independent of the types or combinations of metals or alloys being welded. It also shows that the wave formation ceases at collision angles of 4° or less.

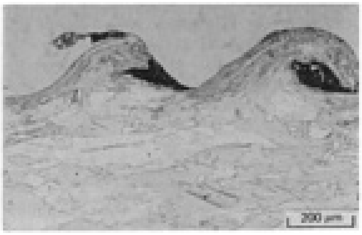

FIG. 6 PHOTOMICROGRAPH SHOWING TYPICAL WAVY WELD INTERFACE OBTAINED WHEN NIOBIUM AND COPPER ARE JOINED BY EXW

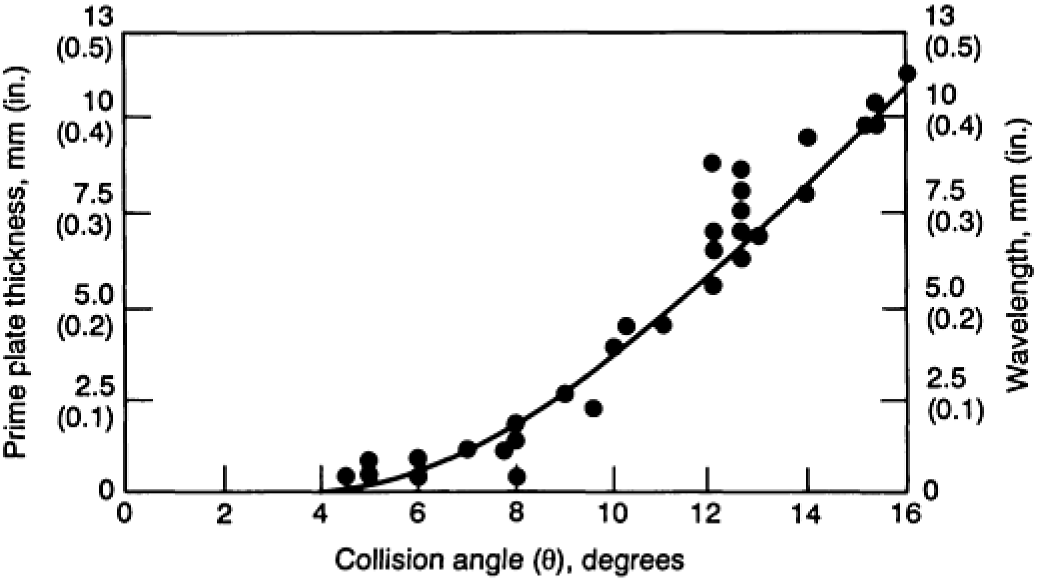

FIG. 7 RELATIONSHIP OF WAVELENGTH TO PLATE THICKNESS AS A FUNCTION OF COLLISION ANGLE

Wave Amplitude. The amplitude of the waves, on the other hand, is more difficult to predict. It depends on several factors, including the relative flow characteristics of the two metals being joined and the collision point velocity of the EXW process.

Control of the wave size, particularly the wavelength, is an important part of controlling several aspects of the explosion weld because:

- WAVE SIZE DETERMINES THE AMOUNT OF LOCALIZED FLOW AND WORK HARDENING IN AND ADJACENT TO THE EXPLOSION WELD INTERFACE.
- EXCESSIVE COLLISION ENERGIES AND WAVE SIZES CAN RESULT IN THE FORMATION OF LARGER MELT POCKETS ON THE FRONT AND BACK SLOPES OF THE WAVES, ALONG WITH RESIDUAL HEAT EFFECTS ON THE REGION IN AND ADJACENT TO THE WELD INTERFACE, WHICH WILL AFFECT THE MECHANICAL PROPERTIES OF THE WELD.
- WAVE SIZE IS AN EXTREMELY CRITICAL FACTOR IN PREVENTING ADIABATIC SHEAR DEFECTS OR CRACKS FROM FORMING ADJACENT TO THE WELD INTERFACE IN SPECIFIC ALLOYS (FOR EXAMPLE, CERTAIN HIGH-STRENGTH NICKEL-BASE AND TITANIUM ALLOYS).

Parametric Limits for Welding. All the above considerations and an understanding of the EXW and its controlling parameters can be assembled to produce a parametric envelope or window of parameters within which a high-quality explosion weld can be consistently produced. The basic concept of the boundaries defining the envelope has been advanced by several researchers (Ref 7, 8, 9).

A typical parametric envelope is defined by four distinct boundary segments (Fig. 8). The lower boundary (1) is controlled by the collision pressure and dwell time of the pressure pulse. The left boundary (2), although somewhat complicated, is primarily controlled by the sonic velocities of the two materials being welded. The right boundary (3) is controlled by the upper sonic-velocity limit at which a jet will form. It is almost never of concern because EXW is typically conducted at collision velocities significantly below this limit. The upper limit (4 or 4A) is primarily controlled by the kinetic-energy release generated at the collision point. In similar or metallurgically compatible combinations of metals or alloys, increasingly higher amounts of energy in this area of the envelope result in increasingly greater amounts of residual heat effects (for example, large melt pockets, interface voids, and a general degradation of the weld quality and properties). Therefore, this boundary (4) cannot be sharply defined in this instance. In metallurgically incompatible systems, however, this upper boundary (4A) becomes very sharply defined because it corresponds to the energy level at which a brittle intermetallic compound begins to form. Welding at or above this limit produces a weld that is typically brittle and can fail catastrophically in service. Typically, this boundary is relatively close to the lower boundary (1), thus leaving a relatively small amount of latitude in collision parameters permissible to obtain a high-quality weld.

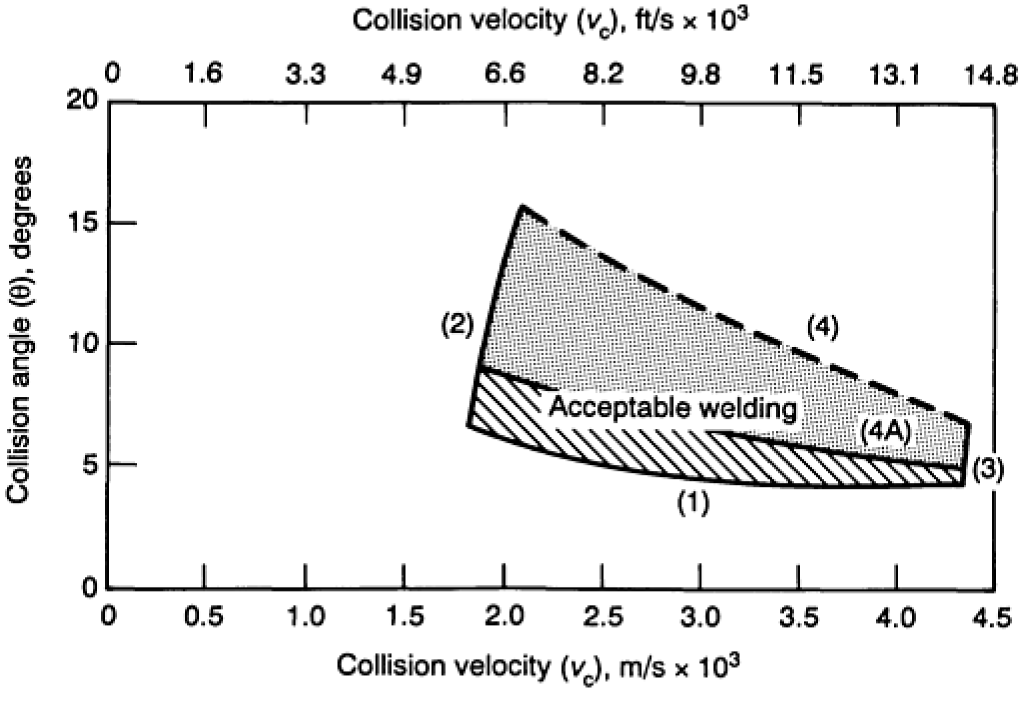

FIG. 8 PLOT OF COLLISION ANGLE VERSUS COLLISION VELOCITY TO OBTAIN TYPICAL EXW PROCESS PARAMETRIC ENVELOPE FOR BOTH SIMILAR AND DISSIMILAR METAL COMBINATIONS. SEE TEXT FOR DETAILS.

Each of the above boundaries can be calculated with relative accuracy for almost all material combinations and thicknesses to be explosion welded. With these limits defined, the collision conditions must then be maintained to fall within the boundaries sufficiently away from the edges to allow for minor variations in the collision conditions during EXW. If this is accomplished, then a reproducible, high-quality explosion weld can always be expected and produced.

References cited in this section

3. A.S. BAHRANNI, T.G. BLACK, AND B. CROSSLAND, THE MECHANICS OF WAVE FORMATION IN EXPLOSIVE WELDING, *PROC. R. SOC.*, VOL 296 (NO. 1445), 1967, P 123-126
4. J.N. HUNT, WAVE FORMATION IN EXPLOSIVE WELDING, *PHILOS. MAG.*, VOL 17 (NO. 148), APRIL 1968, P 669-680

5. A.A. DERIBAS, SIMULATION OF THE PROCESS IN WAVE FORMATION IN EXPLOSIVE WELDING, *COMBUSTION, EXPLOSION AND SHOCK WAVES,* VOL 4 (NO. 1), USSR, 1968
6. UNPUBLISHED DATA, BATTELLE COLUMBUS LABORATORIES
7. K.T. CHRISTENSEN, N.S. EGLEY, AND L. ATLING, EXPLOSIVE WELDING OF TUBES TO TUBE-PLATES, *PROC. FOURTH INT. CONF. CENTER FOR HIGH ENERGY FORMING* (VAIL, CO), 9-13 JULY 1973
8. R.H. WITTMAN, THE INFLUENCE OF COLLISION PARAMETERS ON THE STRENGTH AND MICROSTRUCTURE OF AN EXPLOSION WELDED ALUMINUM ALLOY, *PROC. SECOND INT. SYMP., USE OF EXPLOSIVE ENERGY IN MANUFACTURING METALLIC MATERIALS OF NEW PROPERTIES* (MARIANSKE LAZNE, CZECHOSLOVAKIA), 9-12 OCT 1973
9. A.A. DERIBAS, CLASSIFICATION OF FLOWS APPEARING ON OBLIQUE COLLISIONS OF METALLIC PLATES, *PROC. SECOND INT. SYMP., USE OF EXPLOSIVE ENERGY IN MANUFACTURING METALLIC MATERIALS OF NEW PROPERTIES* (MARIANSKE LAZNE, CZECHOSLOVAKIA), 9-12 OCT 1973

Procedure Development and Practice Considerations for Explosion Welding

Vonne D. Linse, Edison Welding Institute

References

1. *WELDING HANDBOOK,* VOL 2, 8TH ED., R.L. O'BRIEN, ED., AWS, 1991, P 772
2. E.G. SMITH, JR. AND V.D. LINSE, THE ACCELERATION CHARACTERISTICS OF EXPLOSIVELY DRIVEN FLYER PLATES, *PROC. SIXTH INT. CONF. HIGH ENERGY RATE FABRICATION* (ESSEN, WEST GERMANY), 13-16 SEPT 1977
3. A.S. BAHRANNI, T.G. BLACK, AND B. CROSSLAND, THE MECHANICS OF WAVE FORMATION IN EXPLOSIVE WELDING, *PROC. R. SOC.,* VOL 296 (NO. 1445), 1967, P 123-126
4. J.N. HUNT, WAVE FORMATION IN EXPLOSIVE WELDING, *PHILOS. MAG.,* VOL 17 (NO. 148), APRIL 1968, P 669-680
5. A.A. DERIBAS, SIMULATION OF THE PROCESS IN WAVE FORMATION IN EXPLOSIVE WELDING, *COMBUSTION, EXPLOSION AND SHOCK WAVES,* VOL 4 (NO. 1), USSR, 1968
6. UNPUBLISHED DATA, BATTELLE COLUMBUS LABORATORIES
7. K.T. CHRISTENSEN, N.S. EGLEY, AND L. ATLING, EXPLOSIVE WELDING OF TUBES TO TUBE-PLATES, *PROC. FOURTH INT. CONF. CENTER FOR HIGH ENERGY FORMING* (VAIL, CO), 9-13 JULY 1973
8. R.H. WITTMAN, THE INFLUENCE OF COLLISION PARAMETERS ON THE STRENGTH AND MICROSTRUCTURE OF AN EXPLOSION WELDED ALUMINUM ALLOY, *PROC. SECOND INT. SYMP., USE OF EXPLOSIVE ENERGY IN MANUFACTURING METALLIC MATERIALS OF NEW PROPERTIES* (MARIANSKE LAZNE, CZECHOSLOVAKIA), 9-12 OCT 1973
9. A.A. DERIBAS, CLASSIFICATION OF FLOWS APPEARING ON OBLIQUE COLLISIONS OF METALLIC PLATES, *PROC. SECOND INT. SYMP., USE OF EXPLOSIVE ENERGY IN MANUFACTURING METALLIC MATERIALS OF NEW PROPERTIES* (MARIANSKE LAZNE, CZECHOSLOVAKIA), 9-12 OCT 1973

Selection Criteria for Brazing and Soldering Consumables

A. Rabinkin, AlliedSignal Amorphous Metals

Introduction

THE ULTIMATE GOAL of brazing and soldering technologies is to join parts into an assembly through metallurgical bonding (Ref 1, 2, 3, 4, 5). This can be achieved by placing a relatively low-melting-temperature alloy, or filler metal, in the clearance, or gap, between the base materials to be joined, and then heating the assembly until the filler metal has melted and spread throughout the gap. The molten metal that fills the gap reacts with the parts being joined and, after solidification, forms an integral whole. The basic brazing/soldering process is depicted in Fig. 1. Assembly heating can be carried out using a variety of means, including electromagnetic induction, Joule heating, ovens, flames, and other forms.

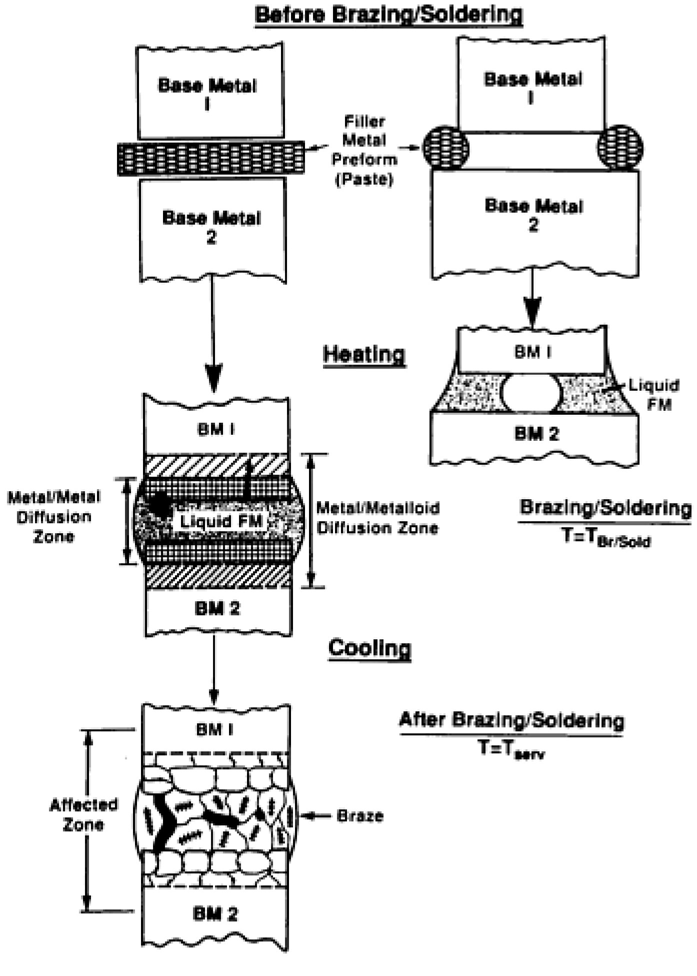
Before Brazing/Soldering
Base Metal I
Filler Metal Preform (Paste)
Base Metal I
Base Metal 2
Base Metal 2
Heating
BM I
Liquid FM
BM 2
BM I
Metal/Metal Diffusion Zone
Liquid FM
Metal/Metalloid Diffusion Zone
Brazing/Soldering
T=T_Br/Sold
BM 2
Cooling
After Brazing/Soldering
T=T_serv
BM I
Affected Zone
Braze
BM 2

FIG. 1 FUNDAMENTALS OF BRAZING/SOLDERING PROCESSES

References

1. *BRAZING HANDBOOK,* 4TH ED., AWS, 1991
2. M.M. SCHWARTZ, *BRAZING,* ASM INTERNATIONAL, 1987
3. R.J. KLEIN-WASSINK, *SOLDERING IN ELECTRONICS,* ELECTROCHEMICAL PUBLICATIONS, 1984
4. M.M. SCHWARTZ, *CERAMIC JOINING,* ASM INTERNATIONAL, 1990
5. A. RABINKIN AND H.H. LIEBERMANN, CHAPTER 21, BRAZING AND SOLDERING WITH

RAPIDLY SOLIDIFIED ALLOYS,*RAPIDLY SOLIDIFIED ALLOYS: PROCESSES, STRUCTURES, AND PROPERTIES, AND APPLICATIONS*, MARCEL DEKKER, 1993, P 691-735

Selection Criteria for Brazing and Soldering Consumables

A. Rabinkin, AlliedSignal Amorphous Metals

Process Stages

There are three major stages in any brazing or soldering process. The first stage occurs during the heating of an assembled workpiece, when the filler metal melts, flows, and completely fills the gap between the parts. Because this gap is usually rather small, ranging from approximately 10 to several hundred μm (0.4 to ~10 mils), the degree to which a molten alloy penetrates into the gap is primarily determined by surface tension effects and, of course, by the chemical-reaction-driven phenomenon that plays an important role in wetting. Because physical properties of the filler metal and base material, such as part dimensions and surface tension forces, have a major role, this first stage can be conditionally called the physical stage.

The second stage, which normally occurs at a given joining temperature, is characterized by an intensive solid-liquid interaction, accompanied by a substantial mass-transfer through the interface with strongly uneven rates. The base material volume that immediately adjoins the liquid filler metal either dissolves or is reacted with the filler metal in this stage. At the same time, a small amount of elements from the liquid phase penetrates into the solid base material. Such mass-transfer unbalance is due to dramatically different diffusion rates in the solid and liquid phases. This redistribution of components in the joint area results in changes to the filler-metal composition, and sometimes, the onset of solidification of the filler metal.

Because all of the processes in this second joining stage have a distinctly metallurgical character, it is called the metallurgical stage. In soldering, these processes proceed to a lesser degree than they do in the brazing process, because of the substantially lower melting temperature of the solder alloys. Here, the dissolution of the base material and mutual diffusion occur to a lesser degree than they do in brazing, but because they can still be significant, protection of the base-material surface is required.

The last stage, which overlaps with the second, is characterized by the formation of the final joint microstructure and progresses during the solidification and cooling of the joint. Here, the crystallization process takes place. Post-solidification cooling causes annealing of the filler metal and partial relief of the thermal stresses induced upon cooling. This stage has very similar features in both brazing and soldering processes.

Cleaning the base-material components being joined and protecting them from oxidation during heating are essential in both brazing and soldering processes. Chemically active substances, called fluxes, are commonly used to provide such cleaning and shielding. The use of nonmetallic agents in fluxes is generally undesirable, because of the increased propensity for their entrapment and the resultant formation of voids in the finished joint. However, the use of certain active elements that can be added to filler-metal alloys to act as a flux can be advantageous. However, this concept usually only works with brazing filler metals. For solders, working temperatures are generally too low for the reduction reactions to take place with adequate speed, although in some cases these reactions do indeed occur.

Selection Criteria for Brazing and Soldering Consumables

A. Rabinkin, AlliedSignal Amorphous Metals

Product Forms and Joint Considerations

Basic forms of both brazing and solder filler metals include solid preforms and powders, which are used mostly in the compound form of paste; plastic-bonded tape (brazing filler metals); and rosin-core wire. Conventional paste forms of joining alloys, which consist of a filler-metal powder, fluxing agent, and a binder/solvent, are applied at externally accessible locations of the clearance between the base-material components. This practice requires substantial filler-metal flow during joining to achieve constituent filler-metal-phase fusion and binder/solvent extraction. In addition, organic binders in a brazing filler metal decompose when used in the high-vacuum brazing of parts intended for critical, high-temperature service. Such binder decomposition can result in the formation of soot in the joint, which can degrade the performance of expensive vacuum equipment.

A basic requirement for any joint is that its strength and ductility be adequate for the service requirements. Generally, the strength of any materialincreases with decreasing grain size, whereas ductility is affected by the presence of brittle phases, which will be discussed later. Therefore, it is important to limit grain size and the amount of brittle phases, in addition to limiting the porosity in joints. Ideally, the maximum size that the grains in a joint should achieve is equal to half the width or even the complete width of the clearance between the base-material components. Thus, the use of a smaller clearance necessarily limits the maximum grain size, promotes higher cooling rates of the filler-metal alloy, and thereby results in a refined joint microstructure (grain size). A smaller clearance also promotes improved retention of base-material properties, because of curtailed base-material erosion achieved by using a smaller volume of filler metal. The size of the grains and the extent of the brittle intermetallic phases in the parent filler metal directly affect the strength of the brazed joint. This consideration is particularly important under transient heating/cooling conditions, such as torch or belt oven brazing, and during automatic solder die bonding, because of the increase in filler-metal grain size and the growth of brittle phases under both liquid and solid-state diffusion processes. The rapid dissolution of constituent filler-metal alloy phases allows the use of higher joining throughput speeds and/or lower joining temperatures. The choice of a specific filler-metal alloy composition depends on the composition of the base material, conditions being compatible with the base material during the joining operation, and the ability of the entire joint to meet the service requirements of the device. Other considerations include a low filler-metal surface tension and, in the case of brazing alloys, the need to reduce surface oxides on the base material. Specific parameters that determine the soundness of a metallic bond include:

- FILLER-METAL MELTING TEMPERATURE SHOULD BE LOWER THAN THAT OF THE CORRESPONDING BASE MATERIAL.
- TO PROVIDE THE DRIVING FORCE NECESSARY FOR JOINING, FILLER-METAL SURFACE TENSION (IN BOTH SOLID AND LIQUID STATES) SHOULD BE LOWER THAN THAT OF THE BASE MATERIAL.
- TO FORM A GOOD METALLIC BOND, BOTH FILLER METAL AND BASE MATERIAL MUST BE SIMILAR RELATIVE TO STRUCTURE, COMPOSITION, AND PHYSICAL PROPERTIES.
- TO REMOVE THE OXIDE FILM PRESENT ON THE BASE MATERIAL, THE FILLER METAL MUST CONTAIN ELEMENTS THAT PROMOTE THE CHEMICAL REDUCTION/DECOMPOSITION OR THE PHYSICAL REMOVAL OF THE OXIDE LAYER.

Brazing Filler Metals

Brazing filler metal alloy compositions can be grouped into four categories. The first and largest group is eutectic-type alloys that have aluminum, nickel, cobalt, or copper as a base, to which silicon/boron (in the case of aluminum- and nickel-base alloys) and phosphorus (in the case of copper- and nickel-base alloys) are added. The presence of one or more of these elements in the alloys tends to impart lower melting temperatures and surface tensions to the filler metal, as well as compatibility with the base material in terms of structure, composition, and properties. Silicon combined with a small amount of magnesium is what lowers the surface tension and melting temperature of aluminum-base alloys, whereas the magnesium addition serves as a fluxing agent. Boron and silicon are used in high-temperature brazing alloys in which the presence of phosphorus (another potential melting-temperature depressant element) could cause unacceptable joint brittleness. Of the alloy additions that promote the self-fluxing of the filler metal during brazing, boron has the greatest "penetrating power." On the other hand, phosphorus is a beneficial fluxing element for use in copper (low-temperature brazing), whereas silicon, for example, would cause unacceptable copper-joint brittleness. The boron- and phosphorus-containing alloys are brittle because various intermetallic phases precipitate when ingot manufacturing technology is used.

This brittle character has limited the available forms of conventionally produced brazing alloys to just powder forms. Fortunately, it was discovered that the presence of silicon, phosphorus, and boron in many conventional filler-metal alloys that have near-eutectic compositions facilitates the conversion of such alloys into ductile, thin, amorphous alloy foil forms when rapid solidification (RS) technology is used. In many cases, a foil filler metal produced by the RS process is very well suited for use in brazing applications. The advantageous utilization of microcrystalline/amorphous RS soldering or brazing foil has been growing since the early 1980s (Ref 5).

The second group of brazing filler metals, which are characterized by a phase diagram that includes a peritectic reaction (for copper-tin alloys) or a minimum in the liquidus curve (for gold-nickel alloys), are used primarily in vacuum-brazing applications and, therefore, require no alloying elements to serve as fluxing agents.

The third group of alloys, which is probably the most widely used, is based on the copper-silver binary eutectic system that is modified by substantial additions of zinc and cadmium (both providing fluxing activity) and minor additions of tin and nickel. However, applicability of cadmium-containing alloys is limited, because of more-stringent Environmental Protection Agency restrictions on cadmium usage.

The fourth group of brazing filler-metal alloys consists of eutectic titanium/zirconium-base alloys to which copper and/or nickel are added. These materials may be used extensively in the brazing of titanium-base alloys in the near future.

The first three groups of conventional brazing filler metals have been classified by the American Welding Society into seven well-defined categories (Table 1).

TABLE 1 MAJOR CLASSES OF BRAZING FILLER METALS

NO.	ALLOY FAMILY AND TYPE	AWS DESIGNATION	FORMS	BASE MATERIALS JOINED	MAJOR APPLICATIONS
1	AI-SI, EUTECTIC	BALSI	PREFORMS, WIRE, RODS, FOIL, POWDER, RS FOIL(A)	ALUMINUM AND ALUMINUM ALLOYS, STEEL TO ALUMINUM AND ALUMINUM TO BERYLLIUM	CAR RADIATORS, HEAT EXCHANGERS, HONEYCOMB AIRCRAFT STRUCTURES, STRUCTURAL PARTS
2	CU-X, SOLID SOLUTION CU-ZN PERITECTIC CU-SN, PERITECTIC	BCU RBCUZN NONE	PREFORMS, WIRE, RODS, FOIL, POWDER, RS FOIL	COPPER AND COPPER ALLOYS, COPPER TO MILD STEEL, COPPER TO STAINLESS STEEL	HEAT EXCHANGERS, STRUCTURAL PARTS, AUTOMOTIVE PARTS
3	CU-P, EUTECTIC	BCUP	PREFORMS, WIRE, RODS, FOIL, POWDER, RS FOIL	COPPER TO COPPER, COPPER TO SILVER/OXIDE POWERED METAL COMPOSITES	ELECTRICAL CONTACTS, BUS BARS, HEAT EXCHANGERS
4	CU-AG, EUTECTIC	BAG	PREFORMS, FOIL, POWDER	MOST FERROUS AND NONFERROUS METALS, EXCEPT ALUMINUM AND MAGNESIUM	MOST WIDELY USED UTILITY FILLER METALS
5	TM-SI-B(B), EUTECTIC (NI/FE + CR)-SI-B	BNI	POWDER, TAPE(C), RS FOIL	AISI 300 AND 400 SERIES STEELS AND NICKEL- AND COBALT-BASE SUPERALLOYS; CARBON STEELS; LOW-ALLOY STEELS; AND COPPER	AIRCRAFT TURBINE COMPONENTS, AUTOMOTIVE PARTS, HEAT EXCHANGERS, HONEYCOMB STRUCTURES
	(NI, PD)-SI-B	NONE	POWDER, TAPE, RS FOIL	AISI 300 SERIES STAINLESS STEELS, CEMENTED CARBIDE, SUPERALLOYS	HONEYCOMB STRUCTURES, CEMENTED CARBIDE/POLYCRYSTALLINE DIAMOND TOOLS, ORTHODONTICS, CATALYTIC CONVERTERS
	(CO,CR)-SI-B	BCO	POWDER, TAPE, RS FOIL	COBALT-BASE HEAT-RESISTANT CORROSION-RESISTANT SUPERALLOYS	AIRCRAFT ENGINES, HONEYCOMB MARINE STRUCTURES

6	AU-NI, SOLID SOLUTION	BAU	PREFORMS, WIRE, RODS, FOIL, TAPE	NICKEL-BASE HEAT-RESISTANT ALLOYS, STEELS	HONEYCOMB STRUCTURES, STRUCTURAL TURBINE PARTS
7	CU-(TI,ZR)-NI EUTECTIC AND PERITECTIC	NONE	CLADDED STRIP, RS FOIL	TITANIUM/ZIRCONIUM-BASE ALLOYS	TITANIUM TUBING, AIRCRAFT ENGINES, HONEYCOMB AIRCRAFT STRUCTURES, AIRCRAFT STRUCTURAL PARTS, CHEMICAL REACTORS

Source: Ref 1

(A) MAY BE PRODUCED AS RAPIDLY SOLIDIFIED, DUCTILE, AMORPHOUS/MICROCRYSTALLINE FOIL.

(B) THIS GROUP INCLUDES ALLOYS BASED ON TRANSITION METALS, SUCH AS NICKEL, IRON, AND COBALT.

(C) BRAZING FILLER METAL IS CARRIED ON A PLASTIC-BONDED TAPE.

References cited in this section

1. *BRAZING HANDBOOK*, 4TH ED., AWS, 1991

5. A. RABINKIN AND H.H. LIEBERMANN, CHAPTER 21, BRAZING AND SOLDERING WITH RAPIDLY SOLIDIFIED ALLOYS,*RAPIDLY SOLIDIFIED ALLOYS: PROCESSES, STRUCTURES, AND PROPERTIES, AND APPLICATIONS*, MARCEL DEKKER, 1993, P 691-735

Selection Criteria for Brazing and Soldering Consumables

A. Rabinkin, AlliedSignal Amorphous Metals

Solders

Despite a long history of development and usage, those solders used today are still lead- or tin-base alloy systems to which a small amount of silver, zinc, antimony, bismuth, indium, or a combination thereof, have been added. The filler-metal and base-material reactions of many solders are characterized by the formation of intermetallic phases in the bulk of the joint as well as at the joint interfaces. Intermetallic phases with low plasticity play a dual role: on one hand, they improve soldered joint strength; on the other hand, they lessen the long-term fatigue life expectancy. The latter is of paramount importance for the reliable service of electronic devices, particularly surface-mount circuit boards and silicon dice that are connected by soldering and continuously experience power and thermal cycling. It is well known that materials with large intermetallic crystals have low resistance to fatigue because of high stress concentration in the vicinity of the crystals. The size of the intermetallic crystals formed in joints upon crystallization with transient soldering processes is determined by the initial size of the crystals in the starting solder filler metal and by the dissolution and precipitation processes initiated during the soldering operation. Here again, RS solders that possess a very uniform microcrystalline structure can have an advantage over conventional solders, because their application results in a strong and uniform joint that has a high resistance to fatigue (Ref 5).

Reference cited in this section

5. A. RABINKIN AND H.H. LIEBERMANN, CHAPTER 21, BRAZING AND SOLDERING WITH RAPIDLY SOLIDIFIED ALLOYS,*RAPIDLY SOLIDIFIED ALLOYS: PROCESSES, STRUCTURES, AND PROPERTIES, AND APPLICATIONS*, MARCEL DEKKER, 1993, P 691-735

Selection Criteria for Brazing and Soldering Consumables

A. Rabinkin, AlliedSignal Amorphous Metals

References

1. *BRAZING HANDBOOK,* 4TH ED., AWS, 1991
2. M.M. SCHWARTZ, *BRAZING,* ASM INTERNATIONAL, 1987
3. R.J. KLEIN-WASSINK, *SOLDERING IN ELECTRONICS,* ELECTROCHEMICAL PUBLICATIONS, 1984
4. M.M. SCHWARTZ, *CERAMIC JOINING,* ASM INTERNATIONAL, 1990
5. A. RABINKIN AND H.H. LIEBERMANN, CHAPTER 21, BRAZING AND SOLDERING WITH RAPIDLY SOLIDIFIED ALLOYS,*RAPIDLY SOLIDIFIED ALLOYS: PROCESSES, STRUCTURES, AND PROPERTIES, AND APPLICATIONS,* MARCEL DEKKER, 1993, P 691-735

Brazing of Cast Irons and Carbon Steels

Ann Severin, Lucas-Milhaupt, Inc.

Introduction

CAST IRONS AND CARBON STEELS are brazeable materials, although the brazeability of cast iron is lower than that of carbon steel. Similar heating methods and brazing procedures are used for both types of materials. However, there are differences between the two that must be considered when making cleaning and filler-metal decisions.

Brazing of Cast Irons and Carbon Steels

Ann Severin, Lucas-Milhaupt, Inc.

Base-Metal Brazeability

Cast irons are lower in brazeability than carbon steels, because of their higher carbon and silicon contents (Table 1), and because of sand inclusions on as-cast surfaces. Carbon, in the form of graphite, inhibits the wettability of a filler metal on a base metal. Silicon promotes graphite formation. Because the different types of cast irons listed in Table 1 have varying carbon and silicon contents, the form of the graphite will vary, giving each type distinct wetting characteristics.

TABLE 1 CHEMICAL COMPOSITION RANGES FOR TYPICAL UNALLOYED CAST IRONS AND PLAIN CARBON STEELS

MATERIAL	COMPOSITION, %				
	C	**SI**	**MN**	**S**	P
GRAY IRON	2.5-4.0	1.00-3.00	0.25-1.00	0.02-0.25	**0.05-1.00**
MALLEABLE IRON	2.00-2.60	1.10-1.60	0.20-1.00	0.04-0.18	**0.18(A)**
DUCTILE IRON	3.00-4.00	1.80-2.80	0.10-1.00	0.03(A)	**0.10(A)**
PLAIN CARBON STEELS	**0.08-1.03**	**0.3(A)**	**0.25-1.00**	**0.050(A)**	**0.040(A)**

Source: Ref 1

(A) MAXIMUM.

Malleable iron is considered to be the most brazeable of the three types of cast iron, because it has low carbon and silicon contents, and because it experiences a heat treatment that causes the graphite to form in irregularly shaped nodules. The nodules can be easily removed by mechanical means, such as shot or grit blasting.

Ductile Iron. The free carbon of ductile iron also is in the form of nodules. These nodules, which are spherical in shape, are formed when a minute amount of magnesium is added to the molten iron. The chemical composition of ductile iron is similar to that of gray iron, but with lower limits on the amounts of sulfur and phosphorus. Ductile iron also can be cleaned for brazing by mechanical means.

Gray iron is the least brazeable of the cast irons, because it forms large graphite flakes that inhibit the wetting of the metal surface. This material is brazeable only in conjunction with special cleaning methods, which are described in the cleaning section of this article.

Carbon steels are easily brazed using a wide variety of heating methods and filler metals, but the heat-treatment requirements and the service conditions need to be considered when choosing the heating method and the filler metal.

Dissimilar metals can be joined to carbon steels and cast irons by means of brazing. This method of joining two or more different materials is successful because the low interaction between the base metals generally results in a ductile joint. Coefficients of thermal expansion (CTE) (Table 2) are the major consideration when designing joints that comprise two different materials. The joints should be designed so that the filler metal is in compression, rather than tension, after brazing. The joint also should be designed so that an adequate gap exists at the brazing temperature to allow the flow of filler metal. The ideal gap, when brazing with the BAg series of materials, is approximately 0.050 to 0.127 mm (0.002 to 0.005 in.).

TABLE 2 COEFFICIENTS OF THERMAL EXPANSION FOR SELECTED MATERIALS

Values represent high and low sides of a range of typical values.

	COEFFICIENT OF THERMAL EXPANSION			
	HIGH		LOW	
	10^{-5}/K	**10^{-6}/°F**	**10^{-5}/K**	10^{-6}/°F
ZINC AND ITS ALLOYS(A)	3.5	19.3	1.9	**10.8**
LEAD AND ITS ALLOYS(A)	2.9	16.3	2.6	**14.4**
MAGNESIUM ALLOYS(B)	2.8	16	2.5	**14**
ALUMINUM AND ITS ALLOYS(A)	2.5	13.7	2.1	**11.7**
TIN AND ITS ALLOYS(A)	2.3	13	. . .	**. . .**
TIN AND ALUMINUM BRASSES(A)	2.1	11.8	1.8	**10.3**
PLAIN AND LEADED BRASSES(A)	2.1	11.6	1.8	**10**
SILVER(A)	2.0	10.9	. . .	**. . .**
CR-NI-FE SUPERALLOYS(C)	1.9	10.5	1.7	**9.2**
HEAT-RESISTANT ALLOYS (CAST)(C)	1.9	10.5	1.1	**6.4**
NODULAR OR DUCTILE IRONS (CAST)(A)	1.9	10.4	1.2	**6.6**
STAINLESS STEELS (CAST)(C)	1.9	10.4	1.1	**6.4**
TIN BRONZES (CAST)(A)	1.8	10.3	1.8	**10**
AUSTENITIC STAINLESS STEELS(A)	1.8	10.2	1.6	**9**
PHOSPHOR SILICON BRONZES(A)	1.8	10.2	1.7	**9.6**
COPPERS(A)	1.8	9.8	. . .	**. . .**
NICKEL-BASE SUPERALLOYS(C)	1.8	9.8	1.4	**7.7**
ALUMINUM BRONZES (CAST)(A)	1.7	9.5	1.6	**9**
COBALT-BASE SUPERALLOYS(D)	1.7	9.4	1.2	**6.8**
BERYLLIUM COPPER(A)	1.7	9.3	. . .	**. . .**
CUPRO-NICKELS AND NICKEL SILVERS(A)	1.7	9.5	1.6	**9**

NICKEL AND ITS ALLOYS(C)	1.7	9.2	1.2	**6.8**
CR-NI-CO-FE SUPERALLOYS(C)	1.6	9.L	1.4	**8**
ALLOY STEELS(C)	1.5	8.6	1.1	**6.3**
CARBON FREE-CUTTING STEELS(C)	1.5	8.4	1.5	**8.1**
ALLOY STEELS (CAST)(C)	1.5	8.3	1.4	**8**
AGE-HARDENABLE STAINLESS STEELS(A)	1.5	8.2	1.0	**5.5**
GOLD(A)	1.4	7.9	. . .	**. . .**
HIGH-TEMPERATURE STEELS(C)	1.4	7.9	1.1	**6.3**
ULTRAHIGH-STRENGTH STEELS(C)	1.4	7.6	1.0	**5.7**
MALLEABLE IRONS(C)	1.3	7.5	1.0	**5.9**
TITANIUM CARBIDE CERMET(C)	1.3	7.5	0.8	**4.3**
WROUGHT IRONS(A)	1.3	7.4	. . .	**. . .**
TITANIUM AND ITS ALLOYS(C)	1.3	7.1	0.9	**4.9**
COBALT(C)	1.2	6.8	. . .	**. . .**
MARTENSITIC STAINLESS STEELS(A)	1.2	6.5	1.0	**5.5**
NITRIDING STEELS(A)	1.2	6.5	. . .	**. . .**
PALLADIUM(A)	1.2	6.5	. . .	**. . .**
BERYLLIUM(B)	1.1	6.4	. . .	**. . .**
CHROMIUM CARBIDE CERMET(A)	1.1	6.3	1.0	**5.8**
THORIUM(B)	1.1	6.2	. . .	**. . .**
FERRITIC STAINLESS STEELS(A)	1.1	6	1.0	**5.8**
GRAY IRONS (CAST)(A)	1.1	6	. . .	**. . .**
BERYLLIUM CARBIDE(C)	1.0	5.8	. . .	**. . .**
LOW-EXPANSION NICKEL ALLOYS(A)	1.0	5.5	0.3	**1.5**
BERYLLIA AND THORIA(D)	0.9	5.3	. . .	**. . .**
ALUMINA CERMETS(C)	0.9	5.2	0.8	**4.7**
MOLYBDENUM DISILICIDE(A)	0.9	5.1	. . .	**. . .**
RUTHENIUM(B)	0.9	5.1	. . .	**. . .**
PLATINUM(A)	0.9	4.9	. . .	**. . .**
VANADIUM(B)	0.9	4.8	. . .	**. . .**
RHODIUM(B)	0.8	4.6	. . .	**. . .**
TANTALUM CARBIDE(C)	0.8	4.6	. . .	**. . .**
BORON NITRIDE(C)	0.8	4.3	. . .	**. . .**
NIOBIUM AND ITS ALLOYS	0.7	4.1	0.68	**3.8**
TITANIUM CARBIDE(C)	0.7	4.1	. . .	**. . .**
STEATITE(A)	0.7	4	0.6	**3.3**
TUNGSTEN CARBIDE CERMET(A)	0.7	3.9	0.4	**2.5**
IRIDIUM(B)	0.7	3.8	. . .	**. . .**
ALUMINA CERAMICS(A)	0.7	3.7	0.6	**3.1**
ZIRCONINM CARBIDE(C)	0.7	3.7	. . .	**. . .**
OSMIUM AND TANTALUM(B)	0.6	3.6	. . .	**. . .**
ZIRCONIUM AND ITS ALLOYS(B)	0.6	3.6	0.55	**3.1**
HAFNIUM(B)	0.6	3.4	. . .	**. . .**
ZIRCONIA(D)	0.6	3.1	. . .	**. . .**
MOLYBDENUM AND ITS ALLOYS	0.6	3.1	0.5	**2.7**
SILICON CARBIDE(D)	0.4	2.4	0.39	**2.2**
TUNGSTEN(B)	0.4	2.2	. . .	**. . .**
ELECTRICAL CERAMICS(A)	0.4	2	. . .	**. . .**
ZIRCON(A)	0.3	1.8	0.2	**1.3**
BORON CARBIDE(D)	0.3	1.7	. . .	**. . .**
CARBON AND GRAPHITE(A)	**0.3**	**1.5**	**0.2**	**1.3**

Source: Ref 2

(A) VALUE FOR A TEMPERATURE RANGE BETWEEN ROOM TEMPERATURE AND 100-390 °C (212-750 °F).

(B) VALUE AT ROOM TEMPERATURE ONLY.

(C) VALUE FOR A TEMPERATURE RANGE BETWEEN ROOM TEMPERATURE AND 540-980 °C (1000-1800 °F).

(D) VALUE FOR A TEMPERATURE RANGE BETWEEN ROOM TEMPERATURE AND 1205-1580 °C (2200-2875 °F).

Figure 1 shows a cross section of the interface of a joint made by the silver brazing of cast iron to carbon steel. The cast iron had been cleaned by the fused-salt method described in the section "Cleaning and Fixturing Procedures." In Fig. 1, note the clean line at the interface of the steel, versus the irregular shape of the cast iron interface. This is caused by the cleaning process, which removes surface impurities and leaves an irregular, but iron-rich, surface. These surface irregularities greatly increase the joint area, so that after the proper fused-salt cleaning, it is possible to produce a brazed bond that exceeds the strength of the cast iron base metal. It should also be noted that the low interaction between the base metal and the filler metal results in a very ductile joint between two materials that would otherwise be difficult to join by other methods.

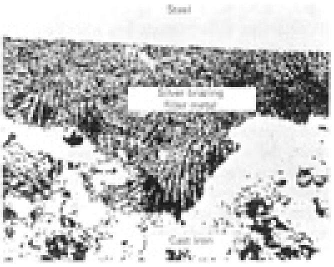

FIG. 1 INTERFACE OF BRAZED CAST IRON AND STEEL AFTER FUSED-SALT CLEANING

References cited in this section

1. C.F. WALTON, ED., *GRAY AND DUCTILE IRON CASTINGS HANDBOOK,* GRAY AND DUCTILE IRON FOUNDERS' SOCIETY, INC, 1971
2. *THE BRAZING BOOK,* HANDY AND HARMAN, 1991

Brazing of Cast Irons and Carbon Steels

Ann Severin, Lucas-Milhaupt, Inc.

Filler-Metal Selection

The most common filler metals used for the brazing of carbon steels and cast irons are those in the American Welding Society (AWS) BAg and BCu series. Filler metals from the RBCuZn series also have been used successfully. However, the BAg series is used because they melt and flow at low temperatures, whereas the BCu series are used because they are inexpensive, when compared with the silver-base filler metals. The trade-off is that the use of BCu materials requires more heat to melt the filler metal. The BCu materials are generally brazed in an atmosphere furnace, which necessitates capital investment in a furnace. The BAg materials, on the other hand, may be brazed by using a torch, an induction heater, a furnace, a resistance heater, or a variety of other means (Ref 2).

The BNi series also can be used in special applications that require chemical or oxidation resistance. The RBCuZn filler metals can be used with a torch, induction heater, or furnace. These alloys can be used either in the typical manner, which relies on capillary flow to make the joint, or in the process known as braze welding, which distributes filler metal by deposition, rather than by capillary flow.

Service Conditions. Temperature and environment are the two factors that need to be evaluated when choosing a filler metal for cast irons and carbon steels. The BAg series of filler metals generally has good long-term strength at temperatures up to 205 °C (400 °F) (Ref 3). Some of the higher-temperature filler metals, such as BAg-13 and BAg-13a, can be used for joints that require high strength up to 425 °C (800 °F) (Ref 4).

The BCu filler metals are generally used for service up to 480 °C (900 °F). The addition of nickel and manganese to copper has created filler-metal alloys with improved high-temperature properties. These materials are not covered by AWS specifications.

Service environment factors that need to be considered when choosing a filler metal include the expected exposure to corrosive environments, water, vibration, and loads. An example of a potential problem is the use of a high-zinc-content filler metal in an application involving contact with flowing water. There have been documented cases of the dezincification of filler-metal joints that eventually failed in service because of the loss of strength in the brazed joint (Ref 5).

A detailed and thorough testing program should be completed for the design and brazement of any critical component, in order to ensure that the finished part will withstand the intended service conditions.

The brazed joint design is another factor in filler-metal selection. Most filler metals are available in either wrought forms (wire, strip, or preformed part) or as a paste (powdered metal in a binder suspension with or without flux, depending on use). Some filler metals are available in limited forms, because of manufacturing difficulties. Ideally, the specific form will allow the brazer to utilize easy assembly and inspection methods and to produce a quality brazed joint. The joint must be designed to accommodate the specific filler-metal form.

In a silver-brazed joint, the greatest strength is achieved when the gap between the two mating parts ranges from 0.0381 to 0.0762 mm (0.0015 to 0.003 in.). Then, the strength of the joint can be equal to or greater than that of the base metal. It is difficult to make a joint that has a narrower gap, because the filler metal would not flow properly between the mating surfaces. When gaps are beyond 0.127 mm (0.005 in.), joint strength drops off, and approaches the tensile strength of the filler metal. If the gap gets too large, then capillary action can be inhibited and the filler metal may not fill the joint, which lowers joint strength even further.

Different silver brazing filler metals have varying degrees of fluidity. Generally, the narrower the melting range (solidus-liquidus), the more fluid the alloy. If the gaps are designed to be tight, then a more fluid filler metal, such as BAg-1 or BAg-7, would be appropriate. If the gaps are larger than 0.102 mm (0.004 in.), then BAg-3 would be a better choice.

Carbon steel and cast iron joints brazed with pure copper should be designed to have a 0.0508 mm (0.002 in.) press fit to a 0.0127 mm (0.0005 in.) gap. Any gap beyond this size will not be filled by pure copper, which is very fluid and has no gap-filling ability. A different filler metal would be required.

If possible, the design should allow visual examination of a joint to ensure that a brazed joint has been made. If a ring can be buried in a fitting/tube joint, then a fillet is evidence that the filler metal has melted and flowed up through the joint to make the fillet.

The heat source is also a factor in determining the type and form of filler metal best suited for a particular brazing cycle. Furnace brazing requires preplaced paste, rings, washers, or other shapes of preforms. Manual brazing can use either the forms described above or bulk wire that is hand fed into the joint when the part reaches a temperature that is higher than the liquidus of the filler metal. Vacuum brazing requires filler metals that will not volatilize in a vacuum at the brazing temperature. Therefore, filler metals with high vapor pressure elements, such as zinc and cadmium, should be avoided. However, these elements are desirable when torch brazing, because they aid wetting and depress melting points.

Heat-treatment requirements will also guide filler-metal selection. The chosen filler metal should allow one of three options when it is used to braze a carbon steel or cast iron that requires a heat treatment.

The first option is that brazing and heat treatment occur at the same temperature. This assumes that the brazing time is short enough to prevent base-metal and filler-metal interactions that can cause erosion or liquid-metal embrittlement. If quenching is required, then it should be conducted after solidification at a temperature that is below that of the filler metal.

The second option is for heat treatment followed by a subsequent brazing operation. In this case, brazing is usually done at the lowest temperature and shortest time possible to minimize the effects on the prior heat treatment. The effect of time at temperatures on heat-treated parts is shown in Table 3.

TABLE 3 EFFECT OF BRAZING TEMPERATURE ON HARDNESS OF SELECTED STEELS

BASE METAL	**CONDITION**	**HARDNESS BEFORE BRAZING**	**HARDNESS AFTER BRAZING SHORT TIME AT 635 °C (1175 °F)**	**HARDNESS AFTER BRAZING SHORT TIME AT 705 °C (1300 °F)**	HARDNESS AFTER BRAZING SHORT TIME AT 760 °C (1400 °F)
LOW-CARBON STEEL	ANNEALED	55-70 HRB	55-70 HRB	55-70 HRB	**55-70 HRB**
	COLD ROLLED	60-90 HRB	55-80 HRB	55-75 HRB	**55-70 HRB**
LOW-ALLOY OR LOW-CARBON STEEL (0.40-0.50 C)	ANNEALED	90-100 HRB	90-100 HRB	90-100 HRB	(A)
	HEAT TREATED TO 1030 MPA (150 KSI)	32 HRC	22-32 HRC	18-25 HRC	(A)
CARBON AND LOW-ALLOY TOOL STEEL	HARDENED AND TEMPERED	50-65 HRC	28-32 HRC	20-25 HRC	(B)
HIGH-SPEED STEEL	**HARDENED**	**65 HRC**	**59-63 HRC**	**46-50 HRC**	**. . .**

Source: Ref 6

(A) MAY HARDEN SLIGHTLY.

(B) MAY HARDEN.

The third option is to braze first and then heat treat. In this case, brazing is done at a temperature above the required heat-treatment temperature. This option usually is used with the BCu filler metals. The subsequent heat treatment is conducted at a low enough temperature to avoid affecting the braze. This is especially useful for parts that need to be case hardened or carburized in situations where brazing after heat treatment would soften or distort the parts.

If high hardness is a requirement and the part must be brazed after hardening, then a different material such as a high-speed tool steel should be considered. Hardened high-speed tool steels and other hot-worked steels exhibiting secondary hardening are only slightly softened by the lower-temperature filler metals in the AWS BAg series.

Before a part is brazed, its physical requirements in finished form must be known in order to select the heat treatment and filler metal that will provide the correct properties. The *ASM Handbook,* Volume 1, *Properties and Selection: Irons, Steels, and High-Performance Alloys,* and Volume 4, *Heat Treating,* offer valuable data for the selection of the appropriate heat treatment.

References cited in this section

2. *THE BRAZING BOOK,* HANDY AND HARMAN, 1991
3. "SILVER BRAZING FOR USE UP TO 400 °F (205 °C)," AMS 2665D, SAE, 1983
4. "SILVER BRAZING FOR USE UP TO 800 °F (425 °C)," AMS 2664D, SAE, 1982
5. *METALS HANDBOOK,* 8TH ED., VOL 10, 1975, P 372
6. "SILVER FILLER METAL BRAZING AND ITS RELATIONSHIP TO HEAT TREATMENT OF THE PARTS JOINED," BRAZING TECHNICAL BULLETIN NO. T-6 (REVISED), HANDY AND HARMAN, 1991

Brazing of Cast Irons and Carbon Steels

Ann Severin, Lucas-Milhaupt, Inc.

Cleaning and Fixturing Procedures

Many processes are used for the brazing of cast irons and carbon steels. All of them involve the same basic considerations, as described below.

Because of the difficulty in wetting cast irons, trials for cleaning the cast irons prior to production should be conducted. Samples should be made that will reflect the cleaning, flux, filler metals, and selected process in order to check the wettability of the base metal. If the samples are not uniformly wetted or if the contact angle is great, then the variables should be changed until uniform wetting and a small contact angle are achieved. Testing is also useful on steel parts, but is not as critical, because they are much easier to braze. It is advisable to run sample parts prior to brazing expensive machined parts, in order to check furnace atmospheres and other processing variables.

Cleaning

The first step required, when making a brazed joint, is to clean all parts, whether they are cast iron or carbon steel. They must be free from grease, dirt, oxides, and other contaminants in order for the filler metal to wet and flow on the surface of the base metal. Therefore, all processing fluids and mill scale must be removed. Generally, solvents or alkaline detergent-type cleaners are used for this purpose. It is important to match the cleaning method to the type of contaminant to be removed (that is, water based for water-based coolants, and solvents or alkaline degreasers for oily residues). Special procedures for specific materials are described below.

Cleaning of Cast Irons. Prior to brazing, cast irons require special cleaning procedures. Ductile and malleable cast iron are usually cleaned by abrasive blasting with steel or nickel shot to remove the graphite, silicon, and sand inclusions on the surface. Other methods for preparing ductile and malleable cast irons include various pickling and other chemical treatments, searing with an oxidizing flame, and heating to temperatures between 870 and 900 °C (1600 and 1650 °F) in a strongly decarburizing atmosphere.

Gray cast irons generally require cleaning in fused salts prior to brazing, because that method has proved to be the most effective. The fused-salt bath generally is maintained at temperatures between 400 and 480 °C (750 and 900 °F). At these temperatures, salt baths exhibit the required high chemical activity that is necessary to remove surface oxides, graphite, and sand from the iron surface. This action is further enhanced by the induction of direct current into the bath. By changing the polarity, the action of the bath can be changed from oxidizing to reducing, and vice versa.

Castings are immersed in the bath, suspended from bus bars or fixtures that are insulated from the tank of molten salt. Generally, the castings are attached to the negative polarity and the tank is made positive, which creates a reducing action in the bath that is capable of removing oxides and sand from the casting surfaces. The polarity is then reversed and the oxidation reaction begins. During this portion of the cycle, the graphite is removed. The current is then reversed again (to a reducing action) to remove the iron oxide that formed in the oxidation reaction. After the castings are rinsed in water and dried, they are ready for brazing. Although it is generally not necessary, ductile and malleable cast irons also can be cleaned by the fused-salt method.

In some applications, the brazing of all types of cast irons has been accomplished without special cleaning by using a black-type flux (AWS FB3-C or AMS 3411) and by silver brazing with a filler metal that contains nickel, such as BAg-3, BAg-4, or BAg-24. Brazing with RBCuZn-type filler metals has also been done without prior precleaning. If joint integrity, repeatability, and consistency are important, then prior precleaning is recommended.

Cleaning of low-carbon steels is generally accomplished by either chemical or mechanical means, the former of which is more popular. Chemical methods include alkaline detergents, solvents, vapor degreasing, and acid pickling. Mechanical methods include dry or wet abrasive blast cleaning, wire brushing, machining, or grinding. The selected cleaning method must provide a chemical match to the contaminant being removed. Some of the newer synthetic lubricants and coolants are easily removed with water-based cleaners, but not by vapor degreasing. Although the part may appear clean, residues of these materials can remain and will inhibit the wetting of the steel.

Chemical Cleaning. Alkaline cleaning, including soak, spray, and barrel cleaning, is widely used for removing oily, semisolid, or solid soils from steel. It is generally satisfactory for removing most cutting and grinding fluids, grinding and polishing abrasives, and some pigmented drawing compounds.

Solvent cleaning is capable of removing oil, grease, loose metal chips, and other contaminants from steel components. Parts are immersed and soaked in a common organic solvent. Spray methods can be employed, as well.

Vapor degreasing utilizes the hot vapors from a boiling chlorinated hydrocarbon solvent to remove surface contaminants such as oils, greases, and waxes. To supplement the vapor, some degreasing units are equipped with facilities for immersing the work in the hot solvent or for spraying with clean solvent. Others are equipped with ultrasonic equipment to remove oils and greases from crevices.

Mechanical Cleaning. Although mechanical methods are less widely used than chemical methods, they are usually preferred for removing heavy scale, and they can be indispensable in removing the more tenacious lubricants, such as a pigmented drawing compound. Mechanical methods are also useful in surface preparation that involves abrading or roughening, which may be required on a very smooth surface to promote wetting and filler-metal flow.

In dry grit blasting, the grits that are used on carbon and low-alloy steels consist of metallic particles of chilled cast iron nickels or of hardened cast steel. Wet blasting employs abrasives of many different types and sizes, suspended in a liquid carrier. Certain ingredients in the liquid carrier, such as rust inhibitors and the minerals in the water, can adversely affect the wetting action in brazing and could necessitate an additional cleaning process to remove all traces of liquid carrier from the work.

Fixturing

Fixturing is necessary to hold parts in alignment during the brazing process. The considerations are the same, whether brazing cast irons or carbon steels. Ideally, the parts would be self-fixturing by virtue of gravity, press fitting (copper filler metal only), tack welding, knurling, staking, expanding, interlocking, or other mechanical means. If tack welding is used, then the area surrounding the tack should be mechanically cleaned to remove any oxides that form during the welding operation. A mechanical method must avoid introducing contaminants, grease, or oils, or else a final cleaning process must be added.

If the part cannot be designed to be self-fixturing, then external fixturing must be designed. A fixture for torch brazing should be lightweight and have as little actual contact as possible with the parts to be brazed, to prevent it from acting as a heat sink.

The fixture material that is used will depend on the temperature and atmosphere of the brazing operation, how many times the fixture is to be used, the base metal, and the brazing process to be used. Fixtures are generally constructed of steel, stainless steel, nickel-base alloy, molybdenum, graphite, or ceramic. Regardless of the process or base metal, the differences in CTE between the base metal and the fixture material must be considered, in order to prevent warpage or damage to the brazed part. CTE differences can also be used advantageously to pull parts into place during the heating process.

Brazing of Cast Irons and Carbon Steels

Ann Severin, Lucas-Milhaupt, Inc.

Filler Metal and Flux/Atmosphere Procedures

Filler metal can be supplied to the joint by hand or by using automatic feeding methods when the base metal reaches the brazing temperature. It also can be preplaced in the joint prior to heating. Hand feeding is the simplest method, however, because it requires no equipment. The choice of wire gage is important, because a wire of large diameter will act as a heat sink, cooling the joint when it touches a part. This requires that the part be heated to a higher temperature, which could cause base-metal degradation. The amount of filler metal that is applied is determined by the operator.

Automatic wire feed machines are used on rotary torch brazing tables. Although this type of machine supplies a measured amount of filler metal to the joint, the wire can act as a heat sink if the wire gage is too large.

Preforms, in the form of wire or foil products, can supply filler metal to the joint in a consistently measured amount. Preforms can be used in any type of brazing operation in which the filler metal can or must be preplaced in the joint prior to heating.

Another common form of filler metal is a paste product, in which a powdered filler metal is mixed with either a binder or a flux/binder system. The choice of paste is determined by the brazing temperature and the atmosphere in which brazing occurs.

The best joint is made by preplacing the filler metal at the base of a joint (Fig. 2) and drawing it up, in order to form a joint. This gives the smallest, most even fillet, and ensures that a joint is formed over the entire shear length of the mating surfaces. However, it is not always practical to create a joint in this manner.

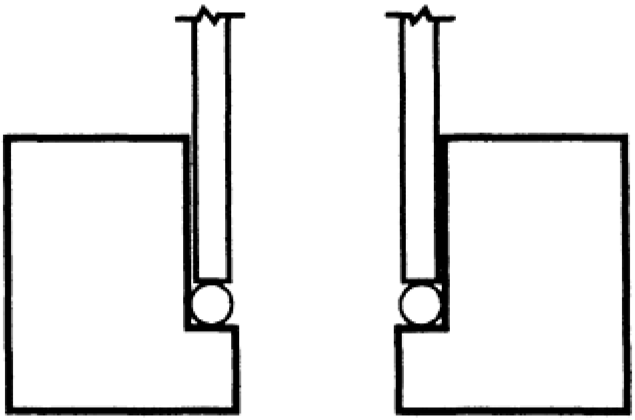

FIG. 2 SCHEMATIC OF A CROSS SECTION OF A PART WITH A BURIED RING OF BRAZING FILLER METAL PRIOR TO BRAZING

Fluxes and/or atmospheres are necessary when making a joint on carbon steels or cast irons. The fluxes used with these materials usually consist of borates and fluorides. Some may have additions of boron to promote wetting, especially on cast irons. Brazing fluxes act to prevent oxidation of both base metal and filler metal and to clean off any residual oxides on the base metal.

Fluxes that conform to AWS FB3A or FB3C can be used when torch brazing with the BAg filler metals. High-temperature torch brazing with RBCuZn filler metals can utilize either a vapor flux that is supplied in the flame or a high-temperature flux that conforms to FB3D. Furnace brazing should utilize a flux that is designed to be active for long periods of time; this type of flux is not covered by AWS designations. The selected flux must be compatible with the combination of base metal and filler metal and with the temperature range in which brazing occurs. Fluxes that are spent by overheating will not allow the filler metal to flow properly.

Brazing atmospheres act in a manner similar to fluxes, by providing a gas atmosphere to protect the part from oxidation and to remove residual oxides. High-temperature brazing using a BCu or a high-temperature copper alloy filler metal can be conducted in hydrogen or dissociated ammonia atmospheres. Low-temperature brazing using BAg filler metals also is possible when hydrogen or dissociated ammonia atmospheres are used. Endothermically generated atmospheres containing from 14 to 16% hydrogen can be used for carbon steels and properly cleaned cast irons. Flux should be used to prevent the vaporization of zinc whenever zinc-containing filler metals are used. Vacuum could be used, if it is available, but it generally is not used, because of the cost.

Brazing of Cast Irons and Carbon Steels

Ann Severin, Lucas-Milhaupt, Inc.

Heating Methods

The heating methods used for the brazing of carbon steel and cast iron are varied. Often, the choice of heating method depends on the process that provides the best results for the lowest cost. When determining the best-suited process and heating method, factors such as part size, planned production and lot sizes, filler-metal and base-metal combinations, and heat-treatment requirements need to be considered.

Torch brazing is commonly used to braze cast irons and carbon steels. This method is particularly suited for cast irons, because the short cycle times preclude the formation of graphite from free carbide or pearlite, which would lower the high-temperature properties of the cast iron. Torch brazing is used for the brazing of single pieces or numerous pieces. Generally, the lower-temperature BAg filler metals are used, although the RBCuZn filler metals can be used, as well. Torch size must be adjusted for the size of part being brazed. Usually, smaller parts or parts that must be brazed in place are the most suitable for torch brazing.

The induction brazing of cast irons and carbon steels is useful on parts that need very fast and localized heating to maintain heat-treatment properties at the maximum level, while ensuring a sound brazed joint. Both cast iron and carbon steel induct heat well, and the only limitations are the size of the part and accessibility to the joint area. Nearly any part shape can be brazed using induction heating, as long as a suitable inductor can be designed to give the desired heating pattern. Induction heating equipment tends to be expensive, but relatively inexpensive fixturing and inductors make it suitable for a wide variety of parts. When making steel parts, induction is useful for simultaneous brazing and hardening operations. When brazing dissimilar metals using an induction method, attention should be given to the differential heating rate of the two metals. The cycle and the coil placement must be such that the two metals are brought to temperature evenly, or else a cold joint can result.

A cold joint occurs when the filler metal melts, but cannot bond to one or both of the base metals because they have not reached the melting temperature of the filler metal. When the molten filler metal hits the colder base metal, it freezes, without making a bond.

Salt-bath brazing is another suitable method for the brazing of cast irons and carbon steels, but it is becoming less common, because of the environmental requirements associated with maintaining a salt bath. Salt-bath brazing is suited to applications where many joints must be made at once, such as the joining of tubing to headers for heat exchangers. This process generally uses the BAg, RBCuZn, or BCu filler metals. The selection of a suitable salt depends on the temperature that is needed for the filler metal to melt and flow. The carburizing of steels can be accomplished by using a carburizing salt. Generally, the higher-temperature copper-zinc filler metals are used in this process. Salt-bath brazing also is suited to applications that would involve daily use of the bath, because of the cost of maintaining and heating the bath.

Furnace brazing of cast irons and carbon steels is generally used when:

- THE PARTS ARE LARGE.
- THERE IS A WIDE VARIETY IN THE GAGES OF THE MATERIALS TO BE BRAZED.
- THERE ARE MANY JOINTS PER PART.
- LARGE PRODUCTION VOLUMES ARE REQUIRED.
- A PROTECTIVE ATMOSPHERE IS REQUIRED.

Furnace brazing is only practical when the filler metal can be preplaced in or at the joint during assembly and held in position during the brazing cycle.

The process can take place in different types of furnaces, such as batch or continuous belt ovens, which can operate in air, hydrogen, nitrogen, argon, or combusted gas atmospheres. Cast irons and carbon steels that are brazed in air need flux, and postcleaning is required, as well. When brazing without flux, the atmosphere must be able to reduce the oxides that are present and prevent the formation of new ones. The reduction of oxides must take place below the solidus temperature of the filler metal, in order to enable the filler metal to wet the base metal and flow into the braze joint.

When furnace brazing, the parts are cleaned and assembled with the filler metal preplaced at or in the joint. Calculations should be made to determine the amount of filler metal needed, and a preformed ring, slug, or shim should be designed to provide the correct amount of filler metal. A paste form of filler metal also can be used.

Fixtures may be necessary to hold parts in the proper alignment during brazing. They should be lightweight and should not hinder the natural expansion or shrinkage of the base metal during heating and cooling. Restriction may cause base-metal distortion and/or brazed-joint cracking. The furnace brazing of cast irons and carbon steels generally employs BAg, BCu, and RBCuZn filler metals.

Regardless of the selected heating method, fluxes must be removed because they are corrosive and may contaminate systems and/or mask leaks. Fluxes are best removed by hot-water rinses or by using proprietary solutions that remove them, along with surface scale and oxidation products.

Brazing of Cast Irons and Carbon Steels

Ann Severin, Lucas-Milhaupt, Inc.

References

1. C.F. WALTON, ED., *GRAY AND DUCTILE IRON CASTINGS HANDBOOK*, GRAY AND DUCTILE IRON FOUNDERS' SOCIETY, INC, 1971
2. *THE BRAZING BOOK*, HANDY AND HARMAN, 1991
3. "SILVER BRAZING FOR USE UP TO 400 °F (205 °C)," AMS 2665D, SAE, 1983
4. "SILVER BRAZING FOR USE UP TO 800 °F (425 °C)," AMS 2664D, SAE, 1982
5. *METALS HANDBOOK*, 8TH ED., VOL 10, 1975, P 372
6. "SILVER FILLER METAL BRAZING AND ITS RELATIONSHIP TO HEAT TREATMENT OF THE PARTS JOINED," BRAZING TECHNICAL BULLETIN NO. T-6 (REVISED), HANDY AND HARMAN, 1991

Brazing of Stainless Steels

Revised by Matthew J. Lucas, Jr., General Electric Aircraft Engines

Introduction

STAINLESS STEEL ALLOYS as a group are not difficult to braze, but those alloys that also contain titanium or aluminum require additional precautions to prevent alloy oxidation during the brazing cycle. Excellent results can be obtained when standard wrought stainless steels are brazed.

The brazeability and weldability of these steels do vary with composition. The quality of brazed joints depends on the selection of the brazing process, process temperature, filler metal, and the type of protective atmosphere or flux that is used. These choices must be compatible with the intended performance of the brazed item.

Acknowledgements

Special thanks are due to Robert L. Peaslee and W. Daniel Kay of Wall Colmonoy Corporation for providing three of the examples cited in this article.

Brazing of Stainless Steels

Revised by Matthew J. Lucas, Jr., General Electric Aircraft Engines

---

Applicability

Brazing is often used to join stainless steels to a variety of dissimilar metals, such as carbon steels, low-alloy steels, and copper alloys, in combinations that would not otherwise weld satisfactorily. This capability is a principal advantage of the brazing process.

A wide variety of brazing filler-metal compositions are available to achieve compatibility, strength, corrosion resistance, and other desirable properties when joining dissimilar metals. When the appropriate heating techniques are employed, brazing provides a means to obtain strong, corrosion-resistant, leak-tight joints in small or thin-walled components with minimal buckling or warpage. Brazing also can produce joints in delicate assemblies and in very thin-gage metals that would be difficult or impossible to obtain if conventional welding techniques were used. Brazing is also suitable for the mass production of small- and medium-sized assemblies on various types of continuous furnace equipment. Finally, brazing enables the production of joints in inaccessible locations by the preplacement of filler metal and the subsequent heating of the workpiece. Often, such joints cannot be made by other joining processes.

Brazing of Stainless Steels

Revised by Matthew J. Lucas, Jr., General Electric Aircraft Engines

---

Brazeability

As a class, stainless steels are no more difficult to braze than carbon and low-alloy steels. The high quantities of chromium that are present in stainless steels cause the chromium oxide films that are on the surfaces of all stainless steels, as well as the films of titanium oxide that form on the surfaces of titanium-stabilized stainless steels, such as type 321. If these oxides, which are both refractory and strongly adherent, are inadequately removed, they will prevent the molten filler metal from wetting the base metal, and thus will prevent a capillary joint from being formed between the metals being joined.

The formation of chromium oxide is accelerated when stainless steels are heated in air. Although the oxide may have been removed from the surface by chemical cleaning at room temperature, a new oxide layer that seriously interferes with wetting will rapidly form when the steel is heated in air to the brazing temperature. The adverse effect of oxides on wetting can be alleviated by:

- CHEMICALLY CLEANING THE SURFACE OXIDE FROM THE STEEL AT ROOM TEMPERATURE, QUICKLY FOLLOWED BY HEATING TO THE BRAZING TEMPERATURE IN A CHEMICALLY INERT GASEOUS ATMOSPHERE, SUCH AS ARGON
- HEATING THE STEEL DIRECTLY TO THE BRAZING TEMPERATURE IN A STRONGLY REDUCING ATMOSPHERE, SUCH AS HYDROGEN, AFTER A LESS-INTENSE CLEANING, WHICH CHEMICALLY REDUCES THE OXIDE AND THEREBY PROMOTES WETTING ACTION
- COATING THE AREA AT THE JOINT WITH A CHEMICALLY ACTIVE FLUX THAT DISSOLVES THE OXIDE DURING HEATING TO THE BRAZING TEMPERATURE
- HEATING IN A VACUUM (AFTER CLEANING), WHICH REDUCES MANY OXIDES, SUCH AS CHROMIUM OXIDE, AND PREVENTS THE GROSS FORMATION OF OTHER OXIDES
- DEGREASING ALONE, IF THE PROPER FLUX OR CONTROLLED ATMOSPHERE IS USED, PRIOR TO BRAZING STAINLESS STEELS WITH CLEAN SURFACES, THAT IS, WITHOUT BLACK OR GREEN OXIDE COATINGS
- SELECTING LOWER-MELTING-POINT BRAZING FILLER METALS TO REDUCE OXIDATION

POTENTIAL

Inclusions and Surface Contaminants. When brazing stainless steel, base-metal inclusions and surface contaminants are even more deleterious than they are when brazing carbon steel. Base-metal inclusions, such as oxides, sulfides, and nitrides, interfere with the flow of filler metal. Flow is also impeded by surface contaminants, which may include lubricants, such as oil, graphite, molybdenum, disulfide, and lead, which are applied during machining, forming, and grinding. Other contaminants are the aluminum oxide particles that are produced either by grit blasting or by grinding with aluminum oxide wheels or belts.

Some filler-metal powders that are used in paste form with stainless steel contain organic binders. Acrylics and other plastics are often used for this purpose. Although some powders form a soot residue, it usually does not interfere with filler-metal flow.

The brazing characteristics of stainless steel can also be seriously impaired by unsuitable fixturing materials, such as graphite, or by a protective atmosphere that contains nitrogen. Carbon in graphite fixtures unites with hydrogen to form methane (CH_4), which carburizes stainless steel and impairs its corrosion resistance. Dissociated ammonia will result in nitriding of the stainless steel, unless the ammonia is sufficiently dry and completely (100%) dissociated. Both carburizing and nitriding interfere with brazing quality.

Brazing Processes. Stainless steels can be brazed by all conventional brazing processes, including furnace, torch, induction, resistance, and salt-bath dip brazing. Furnace brazing is most widely used, because applications generally require brazing in protective atmospheres, including a vacuum. Most of the applications described in this article utilize furnace brazing.

Brazing of Stainless Steels

Revised by Matthew J. Lucas, Jr., General Electric Aircraft Engines

---

Brazing Filler Metal

Most stainless steels can be brazed with any one of several different filler-metal families, including silver, nickel, copper, and gold. In most applications, filler metals are selected on the basis of their mechanical properties, corrosion resistance, service temperature, and compatibility with the base metal. This selection leads to the required brazing temperature, based on service use or manufacturing sequence, and to the required heating method. Table 1 lists the composition requirements of the filler metals that are most commonly used to braze stainless steels. A complete list of brazing filler metals can be found in ANSI/AWS A5.8, which specifies filler metals for the brazing and braze welding processes.

TABLE 1 TYPICAL COMPOSITIONS AND PROPERTIES OF STANDARD BRAZING FILLER METALS FOR BRAZING STAINLESS STEEL

BRAZING FILLER METAL	**COMPOSITION(A), %**									**SOLIDUS TEMPERATURE**		**LIQUIDUS TEMPERATURE**		BRAZING TEMPERATURE RANGE	
	Ag	**Cu**	**Zn**	**Cd**	**Ni**	**Sn**	**Li**	**Mn**	**OTHER ELEMENTS TOTAL**	**°C**	**°F**	**°C**	**°F**	**°C**	°F
SILVER															
BAG-1	44.0-46.0	14.0-16.0	14.0-18.0	23.0-25.0	. . .	. . .	. . .	. . .	0.15	610	1125	620	1145	620-760	**1145-1400**
BAG-	49.	14.	14.	17.	. .	. . .	. . .	. . .	0.15	630	1160	635	1175	635-	**1175-**

1A	0-51.0	5-16.5	5-18.5	0-19.0	.									760	**1400**
BAG-2	34.0-36.0	25.0-27.0	19.0-23.0	17.0-19.0	...	...	...	...	0.15	610	1125	700	1295	700-845	**1295-1550**
BAG-2A	29.0-31.0	26.0-28.0	21.0-25.0	19.0-21.0	...	...	...	...	0.15	610	1125	710	1310	710-845	**1310-1550**
BAG-3	49.0-51.0	14.5-16.5	13.5-17.5	15.0-17.0	2.5-3.5	...	...	...	0.15	630	1170	690	1270	690-815	**1270-1500**
BAG-4	39.0-41.0	29.0-31.0	26.0-30.0	...	1.5-2.5	...	...	...	0.15	670	1240	780	1435	780-900	**1435-1650**
BAG-5	44.0-46.0	29.0-31.0	23.0-27.0	...	...	...	...	...	0.15	680	1250	745	1370	745-845	**1370-1550**
BAG-6	49.0-51.0	33.0-35.0	14.0-18.0	...	...	...	...	...	0.15	690	1270	775	1425	775-870	**1425-1600**
BAG-7	55.0-57.0	21.0-23.0	15.0-19.0	...	...	4.5-5.5	...	...	0.15	620	1145	650	1205	650-760	**1205-1400**
BAG-8	71.0-73.0	BAL	...	...	...	...	...	...	0.15	780	1435	780	1435	780-900	**1435-1650**
BAG-8A	71.0-73.0	BAL	...	...	...	...	0.25-0.50	...	0.15	770	1410	765	1410	765-870	**1410-1600**
BAG-9	64.0-66.0	19.0-21.0	13.0-17.0	...	...	...	...	...	0.15	670	1240	720	1325	720-845	**1325-1550**
BAG-10	69.0-71.0	19.0-21.0	8.0-12.0	...	...	...	...	...	0.15	690	1275	740	1360	740-845	**1360-1550**
BAG-13	53.0-55.0	BAL	4.0-6.0	...	0.5-1.5	...	...	...	0.15	720	1325	860	1575	860-970	**1575-1775**
BAG-13A	55.0-57.0	BAL	...	...	1.5-2.5	...	...	...	0.15	770	1420	895	1640	870-980	**1600-1800**
BAG-18	59.0-61.0	BAL	...	...	...	9.5-10.5	...	...	0.15	600	1115	720	1325	720-845	**1325-1550**
BAG-19	92.0-93.0	BAL	...	...	...	...	0.15-0.30	...	0.15	760	1400	890	1635	875-980	**1610-1800**
BAG-	29.	37.	30.	...	...	...	...	...	0.15	680	1250	765	1410	765-	**1410-**

20	0-31.0	0-39.0	0-34.0		.									870	**1600**
BAG-21	62.0-64.0	27.5-29.5	. . .	. . .	2.0-3.0	5.0-7.0	. . .	. . .	0.15	690	1275	800	1475	800-900	**1475-1600**
BAG-22	48.0-50.0	15.0-17.0	21.0-25.0	. . .	4.0-5.0	. . .	. . .	7.0-8.0	0.15	680	1260	700	1290	700-830	**1290-1525**
BAG-23	84.0-86.0	. . .	. . .	. . .	. . .	. . .	. . .	REMAINDER	0.15	960	1760	970	1780	970-1040	**1780-1900**
BAG-24	49.0-51.0	19.0-21.0	26.0-30.0	. . .	1.5-2.5	. . .	. . .	. . .	0.15	660	1220	705	1305	710-843	**1305-1550**
BAG-25	19.0-21.0	39.0-41.0	33.0-37.0	. . .	. . .	. . .	. . .	4.5-5.5	0.15	740	1360	790	1455	790-846	**1455-1555**
BAG-26	24.0-26.0	37.0-39.0	31.0-35.0	. . .	1.5-2.5	. . .	. . .	1.5-2.5	0.15	710	1305	800	1475	800-870	**1475-1600**
BAG-27	24.0-26.0	34.0-36.0	24.5-28.5	12.5-14.5	. . .	. . .	. . .	. . .	0.15	610	1125	745	1375	745-860	**1375-1575**
BAG-28	39.0-41.0	29.0-31.0	26.0-30.0	. . .	. . .	1.5-2.5	. . .	. . .	0.15	650	1200	710	1310	710-845	**1310-1550**
BAG-33	24.0-26.0	29.0-31.0	26.5-28.5	16.5-18.5	. . .	. . .	. . .	. . .	0.15	610	1125	680	1260	680-760	**1260-1400**
BAG-34	**37.0-39.0**	**31.0-33.0**	**26.0-30.0**	**. . .**	**. . .**	**1.5-2.5**	**. . .**	**. . .**	**0.15**	**650**	**1200**	**720**	**1330**	**720-845**	**1330-1550**

BRAZING FILLER METAL	**COMPOSITION(A), %**					**SOLIDUS TEMPERATURE**		**LIQUIDUS TEMPERATURE**		BRAZING TEMPERATURE RANGE	
	Cu	**P**	**Pb**	**Al**	**OTHER ELEMENTS TOTAL**	**°C**	**°F**	**°C**	**°F**	**°C**	°F
COPPER											
BCU-1	99.90(B)	0.075	0.02	0.01	0.10	1080	1981	1080	1981	1095-1150	**2000-2100**
BCU-1A	99.0(B)	. . .	. . .	. . .	0.30	1080	1981	1080	1981	1095-1150	**2000-2100**
BCU-2	**86.5(B)**	**. . .**	**. . .**	**. . .**	**0.50**	**1080**	**1981**	**1080**	**1981**	**1095-1150**	**2000-2100**

BRAZING FILLER METAL	**COMPOSITION(A), %**														**SOLIDUS TEMPERATURE**		**LIQUIDUS TEMPERATURE**		BRAZING TEMPERATURE RANGE	
	Cr	**B**	**Si**	**Fe**	**C**	**P**	**S**	**Al**	**Ti**	**Mn**	**Cu**	**Zr**	**Ni**	**OTHER ELEM**	**°C**	**°F**	**°C**	**°F**	**°C**	°F

														ENTS TOTAL						
NICKEL$^{(C)}$																				
BNI-1	13.0-15.0	2.75-3.50	4.0-5.0	4.0-5.0	0.6-0.9	0.02	0.02	0.05	0.05	. . .	. . .	0.05	BAL	0.50	975	1790	1035	1900	1065-1205	**1950-2200**
BNI-1A	13.0-15.0	2.75-3.50	4.0-5.0	4.0-5.0	0.06	0.02	0.02	0.05	0.05	. . .	. . .	0.05	BAL	0.50	975	1790	1075	1970	1075-1205	**1970-2200**
BNI-2	6.0-8.0	2.75-3.50	4.0-5.0	2.5-3.5	0.06	0.02	0.02	0.05	0.05	. . .	. . .	0.05	BAL	0.50	990	1780	1000	1830	1010-1175	**1850-2150**
BNI-3	. . .	2.75-3.50	4.0-5.0	0.5	0.06	0.02	0.02	0.05	0.05	. . .	. . .	0.50	BAL	0.50	980	1800	1035	1900	1010-1175	**1850-2150**
BNI-4	. . .	1.5-2.2	3.0-4.0	1.5	0.06	0.02	0.02	0.05	0.05	. . .	. . .	0.05	BAL	0.50	980	1800	1065	1950	1010-1175	**1850-2150**
BNI-5	18.5-19.5	0.03	9.75-10.50	. . .	0.10	0.02	0.02	0.05	0.05	. . .	. . .	0.05	BAL	0.50	1080	1975	1135	2075	1150-1205	**2100-2200**
BNI-6	. . .	. . .	. . .	. . .	0.10	10.0-12.0	0.02	0.05	0.05	. . .	. . .	0.05	BAL	0.50	880	1610	875	1610	925-1095	**1700-2000**
BNI-7	13.0-15.0	0.01	0.10	0.2	0.08	9.7-10.5	0.02	0.05	0.05	0.04	. . .	0.05	BAL	0.50	890	1630	890	1630	925-1095	**1700-2000**
BNI-8	. . .	. . .	6.0-8.0	. . .	0.10	0.02	0.02	0.05	0.05	21.5-24.5	4.0-5.0	0.05	BAL	0.50	980	1800	1010	1850	1010-1095	**1850-2000**
BNI-9	**13.5-16.5**	**3.25-4.0**	**. . .**	**1.5**	**0.06**	**0.02**	**0.02**	**0.05**	**0.05**	**. . .**	**. . .**	**0.05**	**BAL**	**0.50**	**1055**	**1930**	**1055**	**1930**	**1065-1205**	1950-2200

BRAZING FILLER METAL	**COMPOSITION$^{(A)}$, %**					**SOLIDUS TEMPERATURE**		**LIQUIDUS TEMPERATURE**		BRAZING TEMPERATURE RANGE	
	Au	**Cu**	**Pd**	**Ni**	**OTHER ELEMENTS TOTAL**	**°C**	**°F**	**°C**	**°F**	**°C**	°F
PRECIOUS METALS											
BAU-1	37.0-38.0	REMAINDER	. . .	. . .	0.15	990	1815	1015	1860	1015-1095	**1860-2000**
BAU-2	79.5-80.5	REMAINDER	. . .	. . .	0.15	890	1635	890	1635	890-1010	**1635-1850**

BAU-3	34.5-35.5	REMAINDER	. . .	2.5-3.5	0.15	975	1785	1030	1885	1030-1090	**1885-1995**
BAU-4	81.5-82.5	. . .	. . .	REMAINDER	0.15	950	1740	950	1740	950-1005	**1740-1840**
BAU-5	29.5-30.5	. . .	33.5-34.5	35.5-36.5	0.15	1135	2075	1165	2130	1165-1230	**2130-2250**
BAU-6	**69.5-70.5**	**. . .**	**7.5-8.5**	**21.5-22.5**	**0.15**	**1010**	**1845**	**1045**	**1915**	**1045-1120**	**1915-2050**

BRAZING FILLER METAL	**COMPOSITION(A), %**														**SOLIDUS TEMPERATURE**		**LIQUIDUS TEMPERATURE**		BRAZING TEMPERATURE RANGE	
	Cr	**Ni**	**Si**	**W**	**Fe**	**B**	**C**	**P**	**S**	**Al**	**Ti**	**Zr**	**Co**	**OTHER ELEMENTS TOTAL**	**°C**	**°F**	**°C**	**°F**	**°C**	°F
COBALT																				
BCO-1	**18.0-20.0**	**16.0-18.0**	**7.5-8.5**	**3.5-4.5**	**1.0**	**0.7-0.9**	**0.35-0.45**	**0.02**	**0.02**	**0.05**	**0.05**	**0.05**	**REMAINDER**	**0.50**	**1120**	**2050**	**1150**	**2100**	**1150-1230**	**2100-2250**

Source: AWS A5.8-81, "Specification for Brazing Filler Metals"

(A) SINGLE VALUES ARE MAXIMUM PERCENTAGES, UNLESS OTHERWISE INDICATED.
(B) MINIMUM.
(C) ALL BNI ALLOYS HAVE A LIMIT OF 0.10 CO AND 0.005 SE.

Silver brazing filler metals, specifically the BAg group, are the most widely used type of filler metal in the brazing of stainless steels. The BAg-3 filler metal, which contains 3% Ni, is probably selected most frequently, although several other filler metals also can be used successfully. Silver brazing filler metals, especially those that contain zinc and chlorine, are primarily used in torch brazing where flux is applied and should not be considered for furnace brazing.

Joints brazed with silver-base filler metals cannot be used for high-temperature service applications. The recommended maximum service temperature is 370 °C (700 °F) (BAg-3). The recommended joint fitting allowances for silver brazing are relatively loose, that is, the diametral clearance is generally from 0.05 to 0.10 mm (0.002 to 0.004 in.).

Of the silver brazing filler metals shown in Table 1, all except BAg-19 and, possibly, BAg-13, are used at brazing temperatures that fall within the effective range of sensitizing temperatures (540 to 870 °C, or 1000 to 1600 °F) for austenitic stainless steels. Chromium carbide precipitation occurs in the sensitizing temperature range, which impairs the corrosion resistance of the base metal. Carbide precipitation, however, depends on time as well as temperature, and exposure to the sensitizing temperature range for only a few minutes is unlikely to result in a significant amount of precipitate. Nevertheless, the lower melting temperatures of the silver brazing filler metals prohibit the resolution treatment of the base metal after brazing. If corrosion resistance in service is sufficiently critical, an extra-low-carbon, titanium-stabilized or niobium-tantalum-stabilized type should be selected instead of a nonstabilized type of austenitic stainless steel.

Ferritic and martensitic stainless steels that contain little or no nickel are susceptible to interface corrosion in plain water or moist atmospheres, when they are brazed with nickel-free silver brazing filler metals, using a liquid or paste flux. Filler metal that contains nickel helps to prevent interface corrosion. However, for complete protection, special silver brazing filler metals that contain nickel and tin (for example, BAg-21) could be used, and brazing should be done in a protective atmosphere without flux.

Most silver brazing filler metals contain appreciable amounts of copper and zinc, either singly or in combination. Overheating or heating for an excessive period of time may result in extensive penetration of grain boundaries by copper and zinc, thereby embrittling the brazed joint.

Cadmium, which is added to some silver brazing filler metals to lower the melting temperature and improve wetting, also penetrates grain boundaries. The effect is accelerated when parts are brazed under tensile stress. Cadmium-containing fumes are extremely toxic, and operators must take every precaution to avoid inhaling them. Many new alloys that have similar brazing and functional characteristics of the cadmium-bearing alloys, but without their environmental and safety concerns, are being introduced to the market. Filler metals that contain cadmium and/or zinc should not be used in a protective atmosphere, including vacuum, because cadmium gases off readily. Because the loss of zinc or cadmium raises the melting point of the filler metal, brazeability is adversely affected.

Virtually all of the silver brazing filler metals are suitable for brazements that are used in fabricating vacuum chambers and pumps, where pressures as low as 1.3×10^{-3} Pa (10^{-5} torr) are encountered. However, for high-vacuum work (pressures of less than 1.3×10^{-4} Pa, or 10^{-6} torr), the filler metal must not contain cadmium or zinc. Vaporization of these metals interferes with the production of the vacuum and can contaminate the vacuum chamber and pumps. Selection of a suitable silver brazing filler metal for service in vacuum at a pressure below 1.3×10^{-3} Pa (10^{-5} torr) is considered in Example 1.

Nickel brazing filler metals are the next most frequently used filler metals for stainless steels. The BNi group of nickel brazing filler metals provides joints that have excellent corrosion resistance and high-temperature strength. These filler metals are typically supplied in the form of powders mixed with a binder. A limited number of filler metals are also available as sintered rods, preforms, and foils. However, these filler metals may alloy with stainless steel, forming intermetallic phases with two undesirable characteristics. First, the phases are considerably less ductile than either the base metal or the filler metal, even at elevated temperatures, and are therefore a potential source of rupture. Second, the alloys formed with stainless steel are higher-melting-point alloys that are likely to freeze at brazing temperatures, thus blocking further flow into the joint during brazing. To achieve flow in deep joints, diametral clearances as large as 0.10 to 0.20 mm (0.004 to 0.008 in.) are necessary. Knurling of male members sometimes helps in the centering of loosely fitting components. With such large clearances, joints that are brazed with nickel brazing filler metals do not develop their greatest strength. The brazed joint is much stronger with a clearance of 0.025 to 0.076 mm (0.001 to 0.003 in.).

Because of the relatively high brazing temperatures required for the nickel brazing filler metals, their use is generally limited to furnace brazing in a controlled atmosphere (including vacuum), although there are occasional exceptions, as will be seen when Fig. 15 is discussed. Torch brazing with BNi brazing filler metals is used to braze small parts and small quantities.

Copper Brazing Filler Metals. The high brazing temperature and the need for a protective atmosphere generally limits the use of copper filler metals to furnace applications. The BCu brazing filler metals (which are practically pure copper) melt at approximately 1080 °C (1980 °F), and flow freely at 1120 °C (2050 °F).

Copper is not recommended for exposure to certain corrosive substances, such as the sulfur in jet fuel and in sulfur-beating atmospheres. Furthermore, copper brazing filler metals exhibit poor oxidation resistance at elevated temperatures and should not be exposed to service temperatures higher than 430 °C (800 °F). When copper brazing filler metal is used, the recommended diametral allowance on joint fit ranges from a 0.10 mm (0.004 in.) clearance to a 0.05 mm (0.002 in.) interference.

Gold brazing filler metals (the BAu group) are sometimes used for the brazing of stainless steel, although their high cost restricts their use to specialized applications, such as the fabrication of aerospace equipment (Example 4). When gold brazing filler metal is used, them is minimal alloying with the stainless steel base metal. As a result, joints exhibit good ductility, strength, and corrosion resistance. When maximum corrosion resistance is needed, BAu-4 should be used.

Cobalt brazing filler metals (the BCo-1 group) are very rarely used for the brazing of stainless steels. However, this type of filler metal is included in Table 1, and it is available for that purpose.

Brazing of Stainless Steels

Revised by Matthew J. Lucas, Jr., General Electric Aircraft Engines

Fluxes

Flux usually is not required for furnace brazing in strongly reducing or inert atmospheres. However, in some furnace brazing applications, flux is necessary. It is always required for torch brazing and is usually required for induction and resistance brazing, unless atmospheric protection is provided.

Any of the American Welding Society (AWS) types of FB3-A through FB3-J fluxes are suitable for all stainless steel brazing applications where flux is needed. The ANSI/AWS 5.31 specification for fluxes used in brazing and braze welding processes provides more details. There are basically three groups of these fluxes, based on their activity range, and they are further classified by their form (powder, paste, slurry, or liquid). The FB3-A, -F, and -G types of flux contain borates and fluorides and have an effective temperature range from 570 to 870 °C (1050 to 1600 °F). These fluxes are suitable for use with silver brazing filler metals.

Type FB3-C flux contains the same ingredients as type FB3-A, except for the addition of boron, and it has a higher effective temperature range from 570 to 930 °C (1050 to 1700 °F). This flux has an extended heating time capability.

The FB3-D, -I, and -J types of flux also contain borates and fluorides, such that its activity ranges from 760 to 1200 °C (1400 to 2200 °F). These fluxes are often selected for use with silver brazing filler metals if the brazing temperature is above 730 °C (1350 °F). They are well suited for use with copper, nickel, and gold brazing filler metals.

Brazing of Stainless Steels

Revised by Matthew J. Lucas, Jr., General Electric Aircraft Engines

Torch Brazing

For stainless steel, the fundamentals of torch brazing, as well as the advantages and limitations, are basically the same as for carbon steels. However, because of the metallurgical characteristics of stainless steel and its requirements for corrosion resistance, the best results are obtained when special consideration is given to the type of flame at the torch and to the filler-metal composition.

Flame Adjustment. To aid in reducing the oxide that is already present, and to prevent further oxidation of the workpiece surfaces, a flame that ranges from neutral to slightly reducing should be used when torch brazing stainless steel to itself. A reducing flame is also satisfactory for brazing stainless steel to nickel alloys or carbon steels. Although a slightly oxidizing flame is typically best for brazing oxide-containing (tough pitch) copper, a slightly reducing flame is usually best when brazing stainless steel to copper. Use of a reducing flame to braze copper is not recommended because this type of flame will remove oxygen from the copper and can also cause hydrogen embrittlement.

Filler Metals. The silver brazing filler metals that flow at relatively low temperatures are used almost exclusively for the torch brazing of stainless steels. BAg-3 is most often used, because it flows well in the temperature range from 700 to 760 °C (1300 to 1400 °F) and provides joints that have greater resistance to corrosion than those brazed with filler metals such as BAg-1 or BAg-1a (although these filler metals are also used). The use of brazing filler metals that require temperatures higher than approximately 760 °C (1400 °F) results in excessive oxidation, thus making it difficult to obtain adequate wetting of the brazed joint. In special applications, higher-melting-point filler metals must be used (Example 1).

Flux. Type FB3-A flux is most widely used for the torch brazing of stainless steel because it has a working range from 560 to 870 °C (1050 to 1600 °F) and is well suited for use with the lower-melting-point silver brazing filler metals. However, in some applications, other types of FB3 fluxes are preferred.

Example 1: Torch Brazing of Parts for High-Vacuum Service.

The sleeve and tube assembly shown in Fig. 1 is typical of brazements used in vacuum systems. The tubes and the sleeve are type 304L austenitic stainless steel. The BAg-18 (60Ag-10Sn-30Cu) brazing filler metal, which has proved satisfactory for the brazing of assemblies used in high-vacuum service, was used with type FB3-B flux. This silver brazing filler metal does not contain zinc or cadmium. It should be noted that vacuum grades of the silver brazing filler metals, such as BVAg-18, are available, and that they limit the amount of impurities that can be detrimental to components of the vacuum tube industry. Again, ANSI/AWS A5.8 should be consulted for further details.

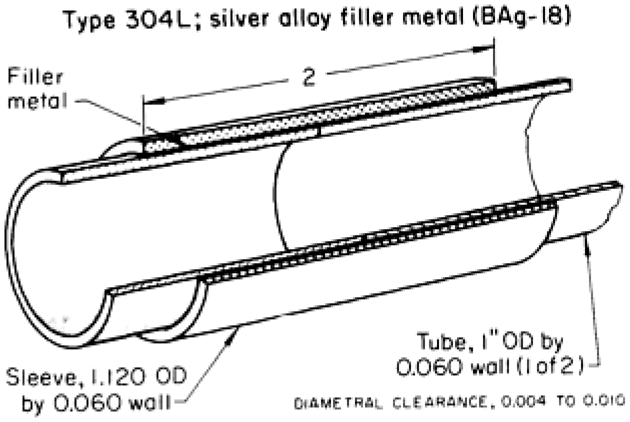
Type 304L; silver alloy filler metal (BAg-18)
Filler metal
2
Tube, 1" OD by 0.060 wall (1 of 2)
Sleeve, 1.120 OD by 0.060 wall
DIAMETRAL CLEARANCE, 0.004 TO 0.010

FIG. 1 ASSEMBLY THAT WAS TORCH BRAZED WITH A SILVER BRAZING FILLER METAL FOR USE IN A HIGH VACUUM ATMOSPHERE SYSTEM

Brazing was conducted with a manually manipulated oxyacetylene torch and a strong reducing flame. The assembly shown in Fig. 1 could have been brazed in a furnace or by induction, but production was small and did not justify the investment for such equipment.

Example 2: Torch Brazing of Stainless Steel to Nickel.

The brazed assembly shown in Fig. 2, which consists of a type 304 stainless steel tube and a pure nickel tube, was resistance heated in service. Requirements for this assembly were:

- TRANSMISSION OF ELECTRICITY WITHOUT DEVELOPING HOT SPOTS
- STRAIGHTNESS AND SMOOTHNESS, BECAUSE THE ASSEMBLY HAD TO SLIDE INTO A LARGER ASSEMBLY
- JOINING AT MINIMUM TEMPERATURE, BECAUSE NUMEROUS SMALL INSULATED WIRES WERE IN THE ASSEMBLY AT THE TIME OF JOINING AND WERE SUBJECT TO DAMAGE IF THE JOINING TEMPERATURE WAS TOO HIGH

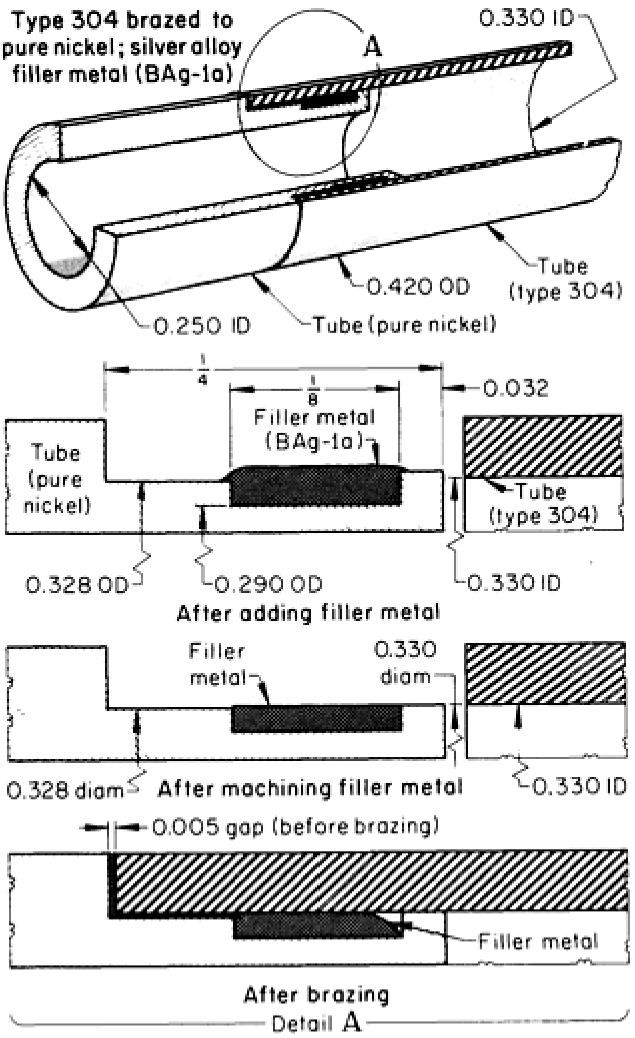
Type 304 brazed to pure nickel; silver alloy filler metal (BAg-1a)
A
0.330 ID
Tube (type 304)
0.420 OD
0.250 ID
Tube (pure nickel)
1/4
1/8
0.032
Filler metal (BAg-1a)
Tube (pure nickel)
Tube (type 304)
0.328 OD
0.290 OD
0.330 ID
After adding filler metal
Filler metal
0.330 diam
0.328 diam
After machining filler metal
0.330 ID
0.005 gap (before brazing)
Filler metal
After brazing
Detail A

FIG. 2 TORCH-BRAZED ASSEMBLY OF A STAINLESS STEEL TUBE AND A PURE NICKEL TUBE

Torch brazing with silver brazing filler metal proved to be a desirable way to make the joint, because of the relatively low brazing temperature, high electrical conductivity of the joint, minimal distortion, ease of removing excess filler metal, and ease of radiographic inspection.

The joint design (Fig. 2) allowed the preplacement of the brazing filler metal (BAg-1a) in a 3.2 mm ($\frac{1}{8}$ in.) wide groove that was machined into the shoulder of the nickel tube. The filler metal was then melted with a torch, and both filler metal and shoulder were machined to match the inside diameter of the stainless tube (zero-clearance fit).

Several heating methods were tried in an attempt to braze the nickel tube to the stainless steel tube, including induction heating in an inert gas and multiple-torch heating. However, the use of a single oxyacetylene torch operated by a skilled technician proved to be the most successful method. The sequence of operations for single-torch brazing was:

- COMPONENTS WERE CLEANED WITH ACETONE
- FLUX PASTE WAS PLACED ON THE NICKEL TUBE IN THE AREA TO BE BRAZED
- THE TWO TUBES WERE ASSEMBLED IN A FIXTURE WITH A 0.13 MM (0.005 IN.) GAP THAT SHOWED SURFACE (DETAIL A IN FIG. 2)
- THE ASSEMBLY WAS HEATED WITH A TORCH UNTIL THE FILLER METAL FLOWED TO THE OUTSIDE SURFACE (FLOW TEMPERATURE OF 640 °C, OR 1175 °F, FOR BAG-1A)

- EXCESS FLUX THAT FLOWED TO THE SURFACE WAS REMOVED MANUALLY

It was possible to inspect the entire joint by making two radiographs. Only scattered porosity was detected in a routine radiographic inspection of the brazed joints.

Brazing of Stainless Steels

Revised by Matthew J. Lucas, Jr., General Electric Aircraft Engines

---

Furnace Brazing

Almost all furnace brazing of stainless steel is done in a protective atmosphere (including vacuum). One exception is the application described in Example 11, in which air was the furnace atmosphere. In this case, a liquid flux was used and the time at brazing temperature was short, which helped to prevent excessive oxidation.

The protective atmospheres most often used in furnace brazing of stainless steel are dry hydrogen and dissociated ammonia. These atmospheres are effective in reducing oxides, protecting the base metal, and promoting the flow of brazing filler metal. The low-cost exothermic atmospheres that are widely used in furnace brazing of low-carbon steel are not suitable for stainless steel. An inert gas, such as argon, or a vacuum environment, can be used to satisfy special requirements and to provide protection in applications for which hydrogen or hydrogen-bearing gases are unsatisfactory.

Selection of the furnace atmosphere depends on the degree of protection that must be given to the base metal or metals, the flow characteristics of the brazing filler metal, the brazing temperature, and cost. Special requirements that arise from the brazing of dissimilar metals are often a major factor in atmosphere selection. The availability of equipment can also be an important consideration.

Furnace Brazing in Dry Hydrogen. A dry hydrogen atmosphere is preferred for many stainless steel brazing applications. Hydrogen, which is the most strongly reducing protective atmosphere, reduces chromium oxide and promotes excellent wetting by many filler metals without the need for flux. The principal disadvantages of hydrogen are its cost, its difficulty in drying sufficiently, the need for special furnace equipment, and the potential danger involved in improper storage and handling. The following examples describe applications in which a specific type of stainless steel was joined to either the same or another type of stainless steel.

Example 3: Selection of BAg-13 Silver Brazing Filler Metal for Brazing at 930 °C (1700 °F).

The type 347 stainless steel retainer assembly shown in Fig. 3 was furnace brazed in dry hydrogen, using BAg-13 silver brazing filler metal, which was selected in preference to lower-melting-point silver brazing filler metals because the upper limit of its brazing temperature range (860 to 970 °C, or 1575 to 1775 °F) permitted brazing at 930 °C (1700 °F). At furnace temperatures above 980 °C (1800 °F), dry hydrogen is strongly reducing, and the use of a brazing flux is not required for satisfactory wetting action. Thus, by judicious selection of the brazing filler metal and the furnace atmosphere, the extra cost of applying the flux and of removing flux residue after brazing was avoided.

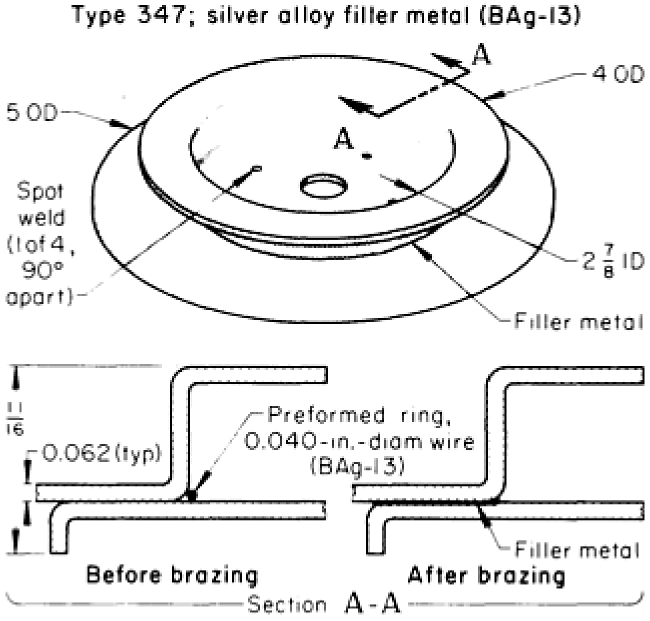
Type 347; silver alloy filler metal (BAg-13)
A
4 OD
5 OD
A
Spot weld (1 of 4, 90° apart)
2 7/8 ID
Filler metal
11/16
0.062 (typ)
Preformed ring, 0.040-in.-diam wire (BAg-13)
Filler metal
Before brazing
After brazing
Section A-A

FURNACE BRAZING IN DRY HYDROGEN	
FURNACE	**CONTINUOUS CONVEYOR(A)**
FIXTURES	**NONE**
FURNACE TEMPERATURE, °C (°F)	**980 ± 5 (1800 ± 10)**
BRAZING TEMPERATURE, °C (°F)	**925 ± 5 (1700 ± 10)**
HYDROGEN DEW POINTS, °C (°F)	**-75 (-100)(B); -60 (-70)(C)**
HYDROGEN FLOW RATE, M^3/H (FT^3/H)	**11 (400)**
FILLER METAL(D)	**BAG- 13**
JOINT POSITION DURING BRAZING	**HORIZONTAL**
CONVEYOR TRAVEL SPEED, M/H (FT/H)	**9 (30)**
TIME AT BRAZING TEMPERATURE, MIN(E)	**5**
PRODUCTION RATE, ASSEMBLIES/H	**120**

(A) ELECTRICITY HEATED (60 KW), CONSTRUCTED WITH HEATING CHAMBER HIGHER THAN ENTRANCE AND DISCHARGE ENDS.

(B) INCOMING.

(C) EXHAUST.

(D) IN FORM OF 1 MM (0.040 IN.) DIAMETER WIRE-RING PREFORMS.

FIG. 3 RETAINER ASSEMBLY FURNACE BRAZED WITH BAG-13 FILLER METAL

The components were vapor degreased, assembled (keeping their outside diameters concentric), and spot welded (to make them self-jigging) at four locations 90° apart (Fig. 3). A ring of brazing filler metal wire in a 1 mm (0.040 in.) diameter was preplaced at the joint, and assemblies were loaded two-across on the mesh belt of a conveyor-type furnace. The heating chamber of the furnace was elevated from the entrance and discharge level to conserve the lighter-than-air hydrogen and to prevent oxygen in the atmosphere from mixing with the hydrogen, which could either raise the dew point or cause an explosion.

Quality standards for brazed assemblies, which were checked by 100% visual inspection, required that the joint exhibit full braze penetration (360° fillets on both sides of the joint) and be pressure-tight.

Example 4: Use of a Gold Brazing Filler Metal for Brazing in an Aerospace Heat Exchanger.

In the fabrication of a high-reliability heat exchanger for manned space flights, 2552 fins of 0.10 mm (0.004 in.) thick type 347 stainless steel were brazed to 0.64 mm (0.025 in.) thick type 347 stainless steel side panels, as shown in Fig. 4. The 5104 fin-to-panel joints had to be strong and corrosion resistant.

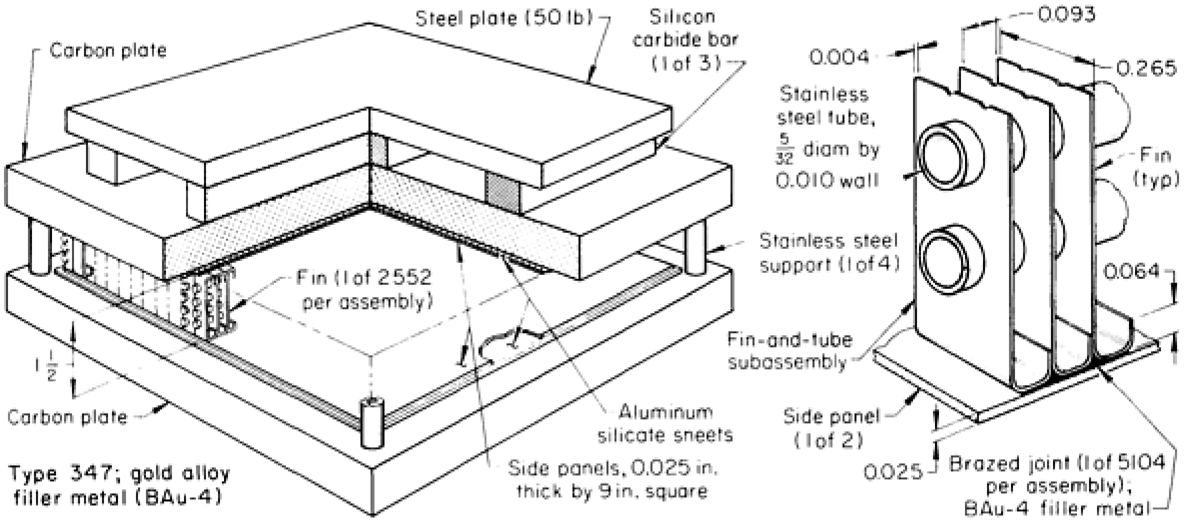
Carbon plate
Steel plate (50 lb)
Silicon carbide bar (1 of 3)
Fin (1 of 2552 per assembly)
$1\frac{1}{2}$
Carbon plate
Type 347; gold alloy filler metal (BAu-4)
Aluminum silicate sheets
Side panels, 0.025 in. thick by 9 in. square
0.004
0.093
0.265
Stainless steel tube, $\frac{5}{32}$ diam by 0.010 wall
Fin (typ)
Stainless steel support (1 of 4)
0.064
Fin-and-tube subassembly
Side panel (1 of 2)
0.025
Brazed joint (1 of 5104 per assembly), BAu-4 filler metal

Fixturing for brazing fin-and-tube subassemblies to side panels

Detail of fin-and-tube subassembly brazed to side panel (typical, both side panels)

FURNACE BRAZING IN DRY HYDROGEN	
FURNACE	**BELL**[(A)]
FIXTURES	**(SEE ILLUSTRATION)**
BRAZING TEMPERATURE, °C (°F)	**1015 (1860)**
HYDROGEN DEW POINT (MAX), °C (°F)	**-60 (-80)**[(B)]
PURGING[(C)]	**5**
FILLER METAL[(D)]	**BAU-4**
NUMBER OF ASSEMBLIES PER LOAD	**1**
PROCESSING TIME PER ASSEMBLY	
CLEAN COMPONENTS, MIN	**45**
PREPLACE FILLER METAL, H	**1.25**
ASSEMBLE COMPONENTS IN FIXTURE, H	**4**
TIME AT BRAZING TEMPERATURE, MIN	**7-10**
TOTAL TIME IN FURNACE[(E)], H	**4**
INSPECT, H	**1**
PRESSURE TEST, H	**40**

(A) ELECTRICALLY HEATED, WITH 36 IN. DIAM RETORT WITH WATER-COOLED RUBBER SEALS.

(B) HYDROGEN WAS PURCHASED AS CYLINDER HYDROGEN, THEN PASSED THROUGH AN ELECTROLYTIC DRIER.

(C) NUMBER OF VOLUME CHANGES IN RETORT.

(D) 200-MESH POWDER SUSPENDED IN AN ORGANIC BINDER.

(E) INCLUDING COOLING TO 150 °C (300 °F) IN RETORT, WHICH WAS PURGED WITH ARGON BEFORE BEING OPENED

FIG. 4 HEAT-EXCHANGER ASSEMBLY IN BRAZING FIXTURE AND DETAIL OF JOINTS BRAZED WITH GOLD BRAZING FILLER METAL

Silver and copper brazing filler metals could not be used, because of their incompatibility with sulfur-bearing rocket fuel. The BNi series of nickel brazing filler metals had the necessary compatibility, but made nonductile joints that were unreliable under tension peel stress. Therefore, gold brazing filler metals were used. The necessary brazing characteristics for the fin-to-panel joints were present in BAu-4 (nominal composition, 82Au-18Ni). The strength and ductility of the resulting brazed joints justified the high cost of this particular brazing filler metal.

The fins and side panels were cleaned by vapor degreasing. The side panels were then pickled, rinsed in clean water, and dried. The brazing filler metal was deposited on the panels in the form of a powder suspended in an organic binder. Multiple lap joints were made between the flat-crown-hairpin ends of the fins and the flat side panels. The assembly was placed in a fixture (Fig. 4), and then the entire assembly and fixture were placed in the retort of a bell-type furnace and sealed. The sealed retort was purged with a volume of hydrogen that was equivalent to five times that of the retort. The retort was then heated to a brazing temperature of 1020 °C (1860 °F) and held for 7 to 10 min. The joint gaps at brazing temperature ranged from 0.000 to 0.254 mm (0.010 in.). After brazing, the retort was purged with argon while being cooled to 150 °C (300 °F) and was not opened until after purging and cooling.

The joints brazed by this procedure were the final brazed joints in the assembly. In a prior brazing operation, tubes had been joined to the fins (Fig. 4) by brazing at 1080 °C (1970 °F) using a higher-melting-point gold brazing filler metal of 70Au-22Ni-8Pd.

The completed assemblies were visually inspected and pressure tested at pressures that far exceeded those of the intended service environment: 1.86 MPa (270 psi) on the outside of the tubes and 11.8 MPa (1710 psi) on the inside. Acceptance pressure test values were 3.7 MPa (540 psi) on the outside and 15.7 MPa (2275 psi) on the inside of the tubes. Selected brazed assemblies were tested to bursting. These assemblies were required to withstand at least three times the service pressures before bursting. The assemblies that were brazed with gold brazing filler metals passed all tests and had three times the bursting strength of the assemblies brazed with undiffused nickel brazing filler metal.

Example 5: Combination Brazing and Solution Heat Treatment of an Assembly of Three Types of Stainless Steel.

Three different stainless steels were selected to make the cover for a hermetically sealed switch. The switching action had to be transmitted through the cover without breaking the seal. This was accomplished by providing a diaphragm through which a shoulder pin was inserted, as shown in Fig. 5. The switch was actuated by depressing the pin, which in turn deflected the diaphragm. The pin (type 303), the diaphragm (PH 15-7 Mo), and the cover (type 305) were assembled as shown in Fig. 3, and then brazed using a silver-base filler metal in a furnace with dry hydrogen.

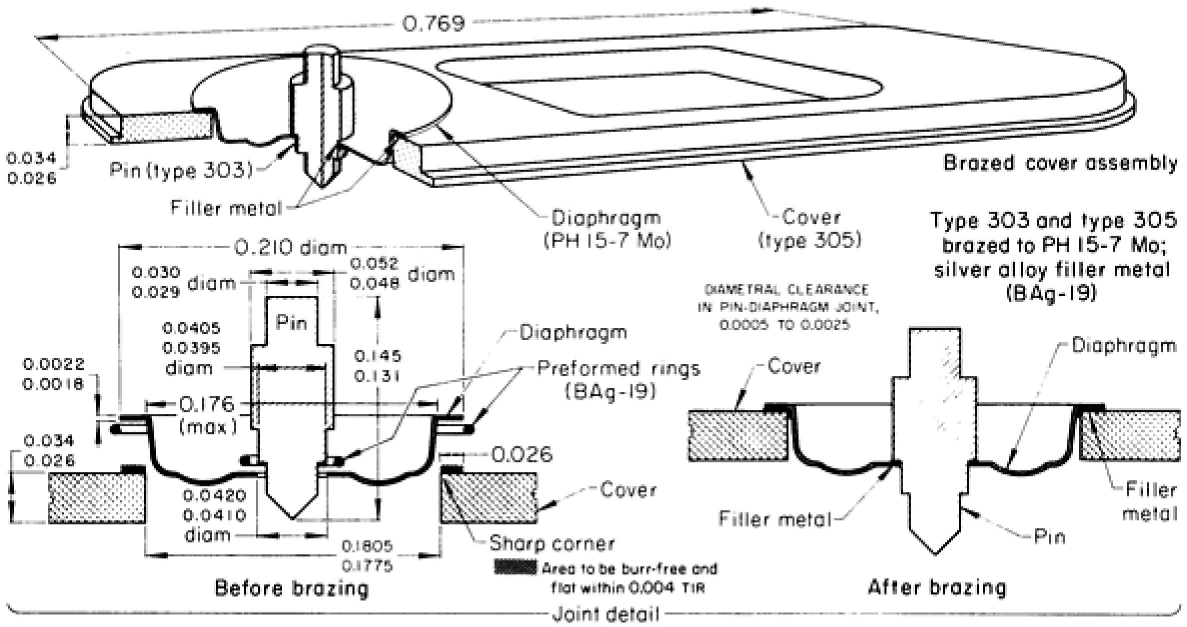
0.769
0.034
0.026
Pin (type 303)
Filler metal
Diaphragm (PH 15-7 Mo)
Cover (type 305)
Brazed cover assembly
Type 303 and type 305 brazed to PH 15-7 Mo; silver alloy filler metal (BAg-19)
0.210 diam
0.030
0.029 diam
0.052
0.048 diam
0.0405
0.0395
diam
Pin
0.145
0.131
0.0022
0.0018
0.176 (max)
Diaphragm
Preformed rings (BAg-19)
0.034
0.026
0.026
0.0420
0.0410
diam
Cover
0.1805
0.1775
Sharp corner
Area to be burr-free and flat within 0.004 TIR
Before brazing
DIAMETRAL CLEARANCE IN PIN-DIAPHRAGM JOINT, 0.0005 TO 0.0025
Cover
Diaphragm
Filler metal
Pin
Filler metal
After brazing
Joint detail

FURNACE BRAZING IN DRY HYDROGEN	
FURNACE(A)	**BATCH-TYPE TUBE**
FIXTURE MATERIAL(B)	**STAINLESS STEEL**
BRAZING TEMPERATURE, °C (°F)	**955 ± 8 (1750 - 15)**
FILLER METAL(C)	**BAG-19**
TIME AT BRAZING TEMPERATURE, MIN	**10**
TIME IN FIRST COOLING ZONE(D), MIN	**5**
TIME IN FINAL COOLING ZONE(D), MIN	**5**

(A) THREE-ZONE FURNACE WITH A HIGH-HEAT ZONE 125 MM (5 IN.) IN DIAMETER BY 460 MM (18 IN.) LONG.

(B) FIXTURE LOCATED AND HELD COMPONENTS OF ASSEMBLY TOGETHER AND WAS PLACED ON A STAINLESS STEEL SLED FOR TRANSPORT THROUGH THE FURNACE.

(C) PREFORMED WIRE RINGS.

(D) AT 540 °C (1000 °F).

(E) WATER-COOLED ZONE, IN WHICH ASSEMBLY WAS COOLED TO ROOM TEMPERATURE. TO COMPLETE HEAT TREATMENT OF THE PH 15-7 MO DIAPHRAGM, ASSEMBLY WAS

FIG. 5 THREE-STEEL SWITCH-COVER ASSEMBLY THAT UTILIZED BRAZING TEMPERATURE AS PART OF SOLUTION HEAT TREATMENT

Silver brazing filler metal BAg-19 was chosen because it flowed at a temperature that coincided with the solution heat-treating temperature for the PH 15-7 Mo diaphragm (950 °C, or 1750 °F). A holding fixture was needed to keep the PH 15-7 Mo diaphragm in position during the brazing cycle. To avoid carburizing the diaphragm, the material selected for the fixture was stainless steel, rather than graphite. The furnace was a batch-type tube furnace with a 120 mm (5 in.) diameter high-heat zone that was 460 mm (18 in.) long. The moisture content of the hydrogen atmosphere was carefully controlled, because the lithium-containing filler metal flowed too freely when the atmosphere was too dry, and it did not seal the joints.

After being cleaned, the components were assembled with two preform rings of BAg-19 wire. Tweezers were used to avoid contaminating the cleaned surfaces. Each assembly was held in a stainless steel fixture, which in turn was placed on a stainless steel furnace sled. The sled was pushed into the high-heat zone of the furnace and held at 950 °C (1750 °F) for 10 min, before being pulled into an intermediate cooling zone at 1830 °C (1000 °F) and held for 5 min. Finally, it was pulled to the water-cooled zone, where it cooled to room temperature. The brazing of the two joints and the solution treating of the PH 15-7 Mo diaphragm were accomplished simultaneously at the brazing temperature of 950 °C (1750 °F). To complete the heat-treating process, the assembly was cooled to -70 °C (-100 °F), held for 8 h, and then aged at 510 °C (950 °F) for 1 h.

A 25 mm (1 in.) square piece of PH 15-7 Mo was processed with each batch of cover assemblies and used as a hardness test specimen to verify that the diaphragms had been correctly heat treated. Brazed assemblies were inspected by the brazing operator. The joints were required to be fully sealed and to not have any voids. The pins were required to be perpendicular within 4°. Perpendicularity was measured on a comparator. Randomly selected samples were given a push-out test, in which joints had to withstand a push of 60 N (14 lbf). All assemblies were given 100% visual inspection at high magnification.

Example 6: Simultaneous Brazing of a Heat-Exchanger Assembly.

An air-to-air heat-exchanger assembly, shown in Fig. 6, consisted of 185 thin-walled (0.20 mm or 0.008 in.) tubes and two 1.6 mm ($\frac{1}{16}$ in.) thick headers. All components were made of type 347 stainless steel. The tubes were assembled with the headers by flaring the tube ends to lock them in place and provide metal-to-metal contact for the brazing filler metal. All 370 joints were brazed during a single pass through a continuous conveyor-type electric furnace.

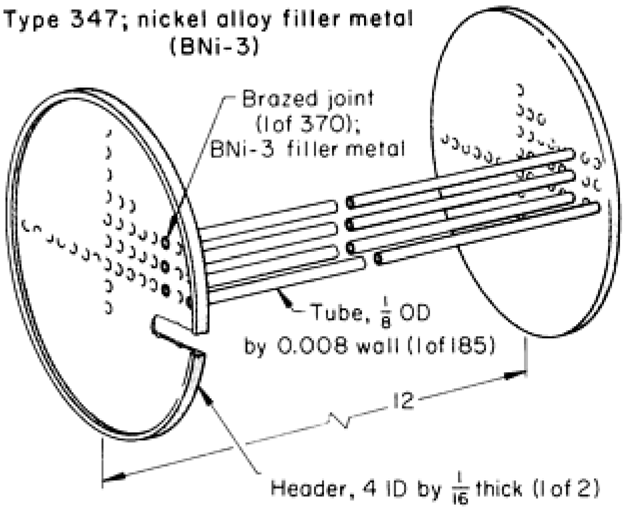
Type 347; nickel alloy filler metal (BNi-3)
Brazed joint (1of 370); BNi-3 filler metal
Tube, $\frac{1}{8}$ OD by 0.008 wall (1of 185)
12
Header, 4 ID by $\frac{1}{16}$ thick (1 of 2)

FURNACE BRAZING IN DRY HYDROGEN	
FURNACE(A)	**CONTINUOUS CONVEYOR**
FIXTURE MATERIAL(B)	**TYPE 347 STAINLESS STEEL**
FURNACE TEMPERATURE, °C (°F)	**1120 ± 5 (2050 ± 10)**
BRAZING TEMPERATURE, °C (°F)	**1065 ± 5 (1950 ± 10)**
HYDROGEN DEW POINTS, °C (°F)	**-75 (-100)(C); -60 (-70)(D)**
HYDROGEN FLOW RATE, M^3/H (FT^3/H)	**17 (600)**
FILLER METAL(E)	**BNI-3 POWDER**
CONVEYOR TRAVEL SPEED, M^3/H (FT^3/H)	**9 (30)**
TIME AT BRAZING TEMPERATURE, MIN	**5**
COOLING	**IN HYDROGEN ATMOSPHERE**
ASSEMBLY PRODUCTION RATE/H	**15**

(A) ELECTRICALLY HEATED (60 KW), CONSTRUCTED WITH HEATING CHAMBER HIGHER THAN ENTRANCE AND DISCHARGE ENDS.

(B) HOLDING FIXTURE FABRICATED FROM 3.2 MM ($\frac{1}{8}$ IN.) THICK SHEET.

(C) INCOMING.

(D) EXHAUST.

(E) MIXED TO A SLURRY WITH ACRYLIC RESIN AND XYLENE THINNER; POWDER-TO-VEHICLE RATIO, 70/30

FIG. 6 HEAT-EXCHANGER ASSEMBLY WITH TUBE-TO-HEADER JOINTS BRAZED IN ONE PASS THROUGH A FURNACE

Although a nickel brazing filler metal was preferred for this high-temperature application, because of the resistance to heat and corrosion that it provides, the selection of a specific nickel brazing filler metal presented a problem. Higher-melting-point, boron-containing nickel brazing filler metals, such as BNi-1 and BNi-3, will react with the base metal, partially dissolve, and therefore are likely to erode thin materials. Fortunately, the extent of erosion can be modified by controlling the brazing temperature and time, as well as the amount of filler metal.

Although there are nickel brazing filler metals that contain silicon in place of boron, they generally require much higher brazing temperatures, which can result in grain coarsening in the base metal. Therefore, after numerous tests, BNi-3 filler metal was selected on the basis of its brazing temperature and excellent fluidity. The problem of applying the correct amount of filler metal to avoid erosion was solved by preparing a slurry from an accurately controlled mixture of filler-metal powder, acrylic-resin binder, and xylene thinner.

Before the filler metal was applied, the heat-exchanger assembly was cleaned ultrasonically in acetone and carefully weighed to determine the proportionate weight of filler metal that would be required. Half of the total amount of filler metal was then applied to one end of the assembly by spraying. The assembly was reweighed, and the remaining filler metal was applied to the opposite end. At all stages of processing, the assembly was handled by operators wearing clean, lint-free cotton gloves.

The assembly was placed on a holding fixture made of stainless steel sheet, on which a stop-off compound had been applied to prevent the assembly from brazing to the fixture if the brazing filler metal flowed excessively. Assemblies were placed 300 mm (12 in.) apart on the conveyor, as they traveled through the furnace at 9 m/h (30 ft/h) under the protection of dry hydrogen.

After brazing, each side was subjected to 100% visual inspection to detect the presence of fillets, and the assemblies were pressure tested in accordance with customer requirements. Because of the thin-walled (0.20 mm, or 0.008 in.) tubing, this assembly was brazed more consistently and at a lower cost than could have been achieved by other joining processes.

Example 7: Combined Brazing and Hardening of a Shaft Assembly.

The shaft assembly shown in Fig. 7 consists of three bars or screw machine products (a shaft, a drive pin, and a guide pin) and two stampings (upper and lower mounting plates), all made from type 410 stainless steel and furnace brazed together using four joints. By brazing with copper filler metal at 1120 °C (2050 °F), it was possible to austenitize and harden the assembly to the required minimum hardness value of 40 HRC during the brazing and cooling operations, thereby avoiding separate hardening operations after brazing.

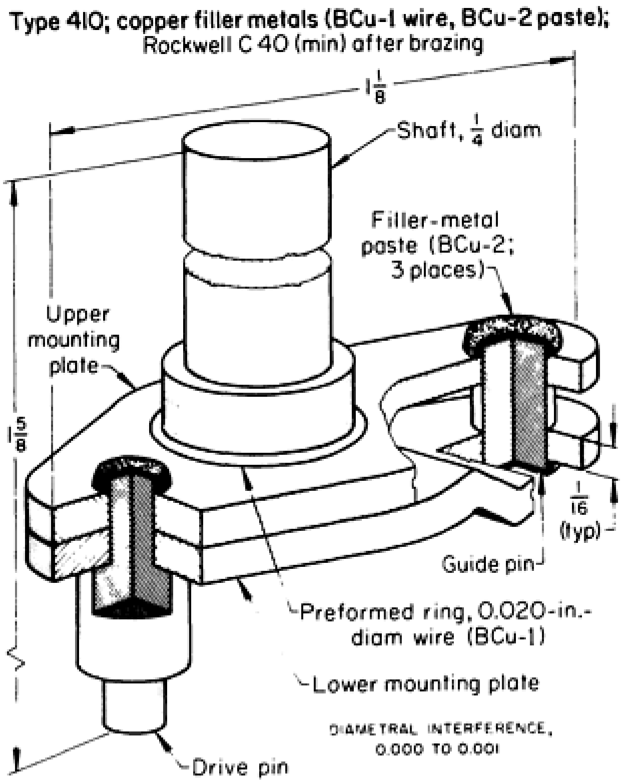
Type 410; copper filler metals (BCu-1 wire, BCu-2 paste);
Rockwell C 40 (min) after brazing
1 1/8
Shaft, 1/4 diam
Filler-metal paste (BCu-2; 3 places)
Upper mounting plate
1 5/8
1/16 (typ)
Guide pin
Preformed ring, 0.020-in.-diam wire (BCu-1)
Lower mounting plate
DIAMETRAL INTERFERENCE, 0.000 TO 0.001
Drive pin

FURNACE BRAZING IN DRY HYDROGEN	
FURNACE(A)	**CONTINUOUS CONVEYOR**
FIXTURES(B)	**NONE**
FURNACE TEMPERATURE, °C (°F)	**1175 ± 5 (2150 ± 10)**
BRAZING TEMPERATURE, °C (°F)	**1120 ± 5 (2050 ± 10)**
HYDROGEN DEW POINTS, °C (°F)	**-75 (-100)(C); -60 (-75)(D)**
HYDROGEN FLOW RATE, M^3/H (FT^3/H)	**11 (400)**
FILLER METAL(E)	**BCU-1 WIRE, BCU-2 PASTE**
CONVEYOR TRAVEL SPEED, M/H (FT/H)	**6 (20)**
TIME AT BRAZING TEMPERATURE (F), MIN	**8**
COOLING	**IN HYDROGEN ATMOSPHERE**
ASSEMBLY PRODUCTION RATE/H	**800**

(A) ELECTRICALLY HEATED (60 KW), CONSTRUCTED WITH HEATING CHAMBER HIGHER THAN ENTRANCE AND DISCHARGE ENDS.

(B) COMPONENTS WERE STAKED, FOR SELF-FIXTURING. ASSEMBLIES, SUPPORTED BY CERAMIC SPACERS TO KEEP SHAFT END UP, WERE BRAZED ON TRAYS.

(C) INCOMING.

(D) EXHAUST.

ONE END OF DRIVE PIN, BOTH ENDS OF GUIDE PIN.
(F) ASSEMBLIES WERE IN HIGH HEAT ZONE FOR ABOUT 10 MIN.

FIG. 7 FOUR-JOINT SHAFT ASSEMBLY THAT WAS SIMULTANEOUSLY FURNACE BRAZED AND HEATED FOR HARDENING

Because the joints were all relatively short, an interference fit of 0.000 to 0.025 mm (0.001 in.) was satisfactory. Typically, a clearance fit between mating parts is required with longer joints in stainless steel. The automatic staking of components was used to make the assembly self-fixturing.

As shown in Fig. 7, a full ring of 0.50 mm (0.020 in.) diameter BCu-1 copper wire was preplaced around the 13 mm ($\frac{1}{2}$ in.) diameter shaft to braze the shaft to the upper and lower mounting plates. A small amount of BCu-2 copper paste was applied at one end of the drive pin to braze it to the two mounting plates. Because of the separation between the two plates on the guide-pin side, a small amount of BCu-2 copper paste was manually applied on each end of the guide pin. The assemblies were placed in brazing trays, with the shaft in a vertical position, and were supported in this position by ceramic spacers.

The brazing trays were then placed on the mesh belt of a continuous-type conveyor furnace containing a dry hydrogen atmosphere. They were transported up an incline to the horizontal preheat and high-heat chambers at a speed of 6 m/h (20 ft/h). Because the assemblies were small, they became heated to the brazing temperature in about 2 min. After 8 min at the brazing temperature, the assemblies were conveyed into water-jacketed cooling chambers, where they cooled rapidly in the hydrogen atmosphere to room temperature. Brazed assemblies that were bright and oxidation-free emerged from the exit end of the furnace.

The brazed assemblies were 100% visually inspected for complete joint coverage. Hardness tests on a sampling basis were used to determine whether the assemblies had responded properly to hardening. Tempering to the desired final hardness followed the simultaneous brazing and hardening operation.

Example 8: Medical Device Brazed, Rather than Welded, in Hydrogen.

Because of the need for strong, corrosion-resistant, and leak-proof joints in a stainless steel blood-cell washer (Fig. 8), the process that was selected was hydrogen furnace brazing with BNi-7 brazing filler metal. The devices are used to expedite and standardize cell-washing procedures in blood banks and hematology laboratories. Therefore, neither voids nor cracks could be tolerated, because the possibility of breeding bacteria in the devices had to be avoided.

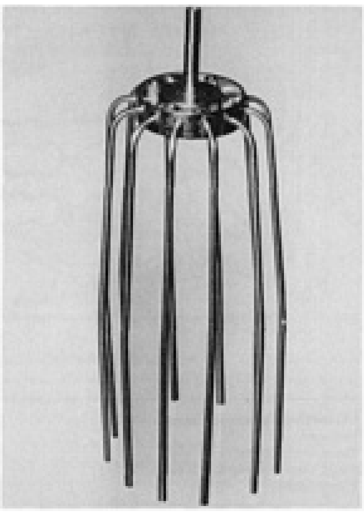

FIG. 8 MANIFOLD AND CANNULAS TUBE ASSEMBLY OF A BLOOD-CELL WASHER THAT WAS HYDROGEN FURNACE BRAZED WITH BNI-7 BRAZING FILLER METAL AT 1040 °C (1900 °F). COURTESY OF WALL

Brazing, rather than welding, was chosen to join the manifold and the delicate tube parts of the washer, a rake-like component with twelve prong-like cannulas tubes that extend approximately 150 mm (6 in.) from a cylindrical manifold.

The manifold assembly was brazed using BNi-7 nickel brazing filler metal in a hydrogen atmosphere at 1040 °C (1900 °F). After assembling the tubes to the manifold, the brazing filler metal was applied to the joints and the assembly was placed in the furnace. Components were wired to stainless steel fixtures to maintain uniform tube spacing.

The brazing permitted the filler metal to flow completely around the thin-walled tubes, which was not possible with welding, leaving smooth fillets without voids that could trap harmful bacterial particles. Additional advantages of brazing were that it: minimized the amount of assembly distortion, eliminated flux hazards, simplified inspection procedures, and prevented the oxidation.

Brazing of Stainless Steels

Revised by Matthew J. Lucas, Jr., General Electric Aircraft Engines

---

Furnace Brazing in Dissociated Ammonia

When ammonia is free of moisture and is 100% dissociated, it becomes a suitable atmosphere for the brazing of stainless steel using selected brazing filler metals without requiring a flux. Although dissociated ammonia is strongly reducing, it is less so than pure, dry hydrogen. Consequently, it will promote wetting action by reducing chromium oxide on the surface of stainless steel, but it may not be sufficiently reducing to promote the flow of some brazing filler metals, such as copper oxide powders. Because of its high (75%) hydrogen content, dissociated ammonia forms explosive mixtures with air and must be handled with the same precautions as those required for the handling of hydrogen.

A dissociated-ammonia atmosphere is prepared by heating anhydrous liquid ammonia in the presence of an iron or nickel catalyst. The decomposition of ammonia to form hydrogen and nitrogen begins at 315 °C (600 °F), and the rate of decomposition increases with temperature. Unless the atmosphere used in brazing stainless steel is completely decomposed, that is, 100% dissociated, even minute amounts of raw ammonia (NH_3) in the atmosphere will cause the nitriding of stainless steel, especially steels containing little or no nickel. In addition, because of the solubility of ammonia in water, the atmosphere that comes from the dissociator must be extremely dry (preferably having a dew point of -60 °C, or -80 °F, or lower). To ensure a very low dew point, the atmosphere that comes from the dissociator is commonly processed by being passed through a molecular-sieve dryer. To avoid the oxidation of base metal and brazing filler metal, the atmosphere must be kept pure and dry while it is inside the furnace. In the following examples of production practices, dissociated ammonia was used successfully in the furnace brazing of austenitic and precipitation-hardening (PH) stainless steels.

Example 9: Brazing in Dissociated Ammonia Without Flux.

The pressure gage subassembly shown in Fig. 9 comprises five diaphragms of 17-7 PH stainless steel, a deep-drawn cup and a top fitting of type 304 stainless steel, and a connector of copper alloy 145 (tellurium-bearing copper). Originally, these subassemblies were furnace brazed with a silver brazing filler metal that required a flux. Because applying flux and assembling the fluxed components with gloved hands was time consuming, the decision was made to change to fluxless brazing in an atmosphere of dissociated ammonia. Although this necessitated using a more-expensive brazing filler metal (BAg-19), the higher cost was offset by the greater productivity of each operator. In addition, subassemblies brazed with BAg-19 in dissociated ammonia exhibited fewer leaks and had improved corrosion resistance and a better appearance than those brazed with the original filler metal and a flux.

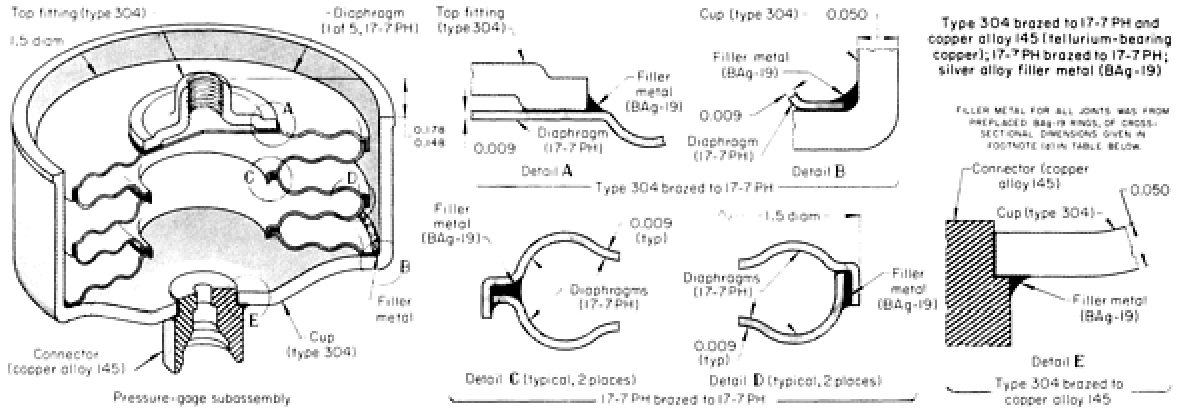
Top fitting (type 304)
1.5 diam
Diaphragm (1 of 5, 17-7 PH)
A
C
D
B
E
0.178
0.148
Filler metal
Cup (type 304)
Connector (copper alloy 145)
Pressure-gage subassembly
Top fitting (type 304)
Filler metal (BAg-19)
0.009
Diaphragm (17-7 PH)
Detail A
Cup (type 304)
0.050
Filler metal (BAg-19)
0.009
Diaphragm (17-7 PH)
Detail B
Type 304 brazed to 17-7 PH
Filler metal (BAg-19)
0.009 (typ)
Diaphragms (17-7 PH)
Detail C (typical, 2 places)
1.5 diam
Diaphragms (17-7 PH)
Filler metal (BAg-19)
0.009 (typ)
Detail D (typical, 2 places)
17-7 PH brazed to 17-7 PH
Connector (copper alloy 145)
0.050
Cup (type 304)
Filler metal (BAg-19)
Detail E
Type 304 brazed to copper alloy 145

Type 304 brazed to 17-7 PH and copper alloy 145 (tellurium-bearing copper); 17-7 PH brazed to 17-7 PH; silver alloy filler metal (BAg-19)

FILLER METAL FOR ALL JOINTS WAS FROM PREPLACED BAg-19 RINGS, OF CROSS-SECTIONAL DIMENSIONS GIVEN IN FOOTNOTE (d) IN TABLE BELOW.

FURNACE BRAZING IN DISSOCIATED AMMONIA	
FURNACE(A)	**CHAIN-BELT CONVEYOR**
FURNACE TEMPERATURE(B), °C (°F)	**980 (1800)**
DISSOCIATED-AMMONIA DEW POINT(C), °C (°F)	**-60 (-80)**
FILLER METAL(D)	**BAG-19**
FLUX(E)	**NONE**
FURNACE BELT SPEED, MM/MIN (IN./MIN)	**255 (10)**
HEATING TIME, MIN	**5**
COOLING-CHAMBER TEMPERATURE(F), °C (°F)	**15 (60)**
PRECIPITATION-HARDENING TEMPERATURE(G), °C (°F)	**510 (950)**

(A) ELECTRICALLY HEATED, WITH ELEVATED HIGH-BEAT ZONE.

(B) FOR BRAZING THE SUBASSEMBLY AND SIMULTANEOUSLY SOLUTION HEAT TREATING THE 17-7 PH DIAPHRAGMS.

(C) ACHIEVED BY RUNNING THE DISSOCIATED AMMONIA THROUGH A MOLECULAR-SIEVE DRYER AFTER CRACKING.

(D) CROSS-SECTIONAL DIMENSIONS (AND PRODUCT FORMS) OF PREPLACED RINGS WERE: FOR JOINT BETWEEN DIAPHRAGM AND TOP FITTING (DETAIL A), 1.3 MM (0.050 IN.) WIDE BY O.10 TO 0.13 MM (0.004 TO 0.005 IN.) THICK STAMPING; FOR OUTSIDE JOINTS BETWEEN DIAPHRAGM SEGMENTS (DETAIL D) AND JOINT BETWEEN DIAPHRAGM AND CUP (DETAIL B), THICK (RIBBON); FOR INSIDE JOINTS BETWEEN DIAPHRAGM SEGMENTS (DETAIL C), 0.76 MM (0.030 IN.) WIDE BY 0.13 MM (0.005 IN.) THICK (RIBBON); AND FOR JOINT BETWEEN CUP AND CONNECTOR (DETAIL E), 1.52 BY 0.254 MM (0.060 BY 0.010 IN.) (WIRE).

(E) USE OF BAG-19 ELIMINATED THE NEED FOR FLUX, WHICH HAD BEEN REQUIRED WITH THE SILVER FILLER METAL ORIGINALLY USED.

(F) TO COOL RAPIDLY FROM 980 °C (1800 °F) AND ENSURE SOLUTION TREATMENT OF THE 17-7 PH DIAPHRAGMS.

(G) IN DRY DISSOCIATED AMMONIA, AFTER SUBASSEMBLY HAD BEEN COOLED TO -40 °C (-

FIG. 9 PRESSURE-GAGE SUBASSEMBLY THAT COMBINED FURNACE BRAZING WITH SOLUTION HEAT TREATMENT

Prior to brazing, the deep-drawn type 304 cups were fully annealed at 1090 °C (2000 °F). Annealing served to avoid the erosive penetration of brazing filler metal in zones of high residual stress. All components were chemically cleaned and then assembled by hand, along with seven preplaced rings of brazing filler metal. The assemblies were placed on holding fixtures, which were loaded on the belt of a conveyor furnace heated to 980 °C (1800 °F). The cooling chamber of the furnace was cooled to below 15 °C (60 °F) to ensure rapid cooling of the 17-7 PH diaphragms from the solution-treating temperature, thereby combining solution treating with the brazing operation.

After brazing, the assemblies were cooled to -40 °C (-40 °F), dried, and then heated to 510 °C (950 °F) in dry, dissociated ammonia to harden the diaphragms to 44 to 48 HRC. Brazed assemblies were pressure tested in a bellows halogen leak detector by applying Freon at 520 kPa (75 psi) and then adding compressed air to bring the total pressure up to 2.1 MPa (300 psi). Leakage of Freon in the gas-air mixture would have been detected by the halogen leak detector. The requirement was for no leakage at the most sensitive setting of the leak detector.

The rejection rate for leakage, based on the 750,000 bellows that were produced, dropped from 2.8%, with the original silver brazing filler metal, to 1.0%, with the BAg-19 brazing filler metal. Field corrosion returns dropped 96%. By eliminating the stains caused by flux, it was no longer necessary to paint the assemblies.

Brazing of Stainless Steels

Revised by Matthew J. Lucas, Jr., General Electric Aircraft Engines

---

Furnace Brazing in Argon

Argon is occasionally used as a furnace atmosphere when brazing stainless steels to other stainless steels or to reactive metals such as titanium (Example 10). Argon has the advantage of being chemically inert in relation to all metals. Therefore, it is a useful protective atmosphere for metals that can combine with or absorb reactive atmospheres such as hydrogen. Because an argon atmosphere has the disadvantage of being unable to reduce oxides, the surface of stainless steel components must be exceptionally clean and oxide-free when brazed in argon.

Example 10: Brazing in an Argon Atmosphere.

A manufacturer of jet engines designed a gear-reduction box of commercially pure titanium. This complicated fabrication was made from assemblies of stampings and machined forgings, most of which were joined by gas-tungsten arc welding (GTAW) in argon-filled welding chambers. However, brazing was more appropriate for the joining of some assemblies.

A typical assembly that was furnace brazed in argon is shown in Fig. 10. This assembly consisted of a machined forging of commercially pure titanium (per Aerospace Material Specification 4921) and a length of seamless type 347 stainless steel tubing that was flared or expanded for a distance of approximately 8 mm ($\frac{5}{16}$ in.) to accept the titanium forging.

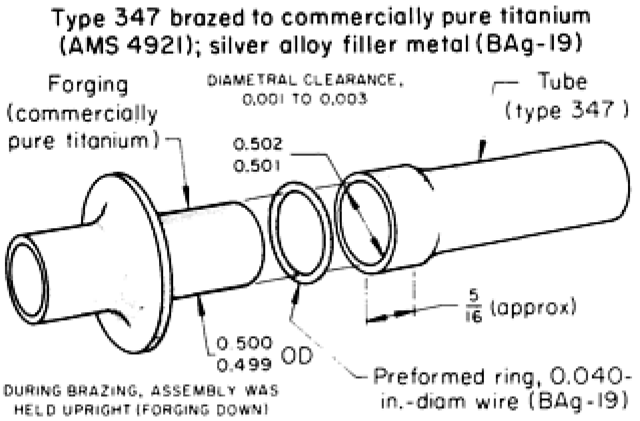
Type 347 brazed to commercially pure titanium
(AMS 4921); silver alloy filler metal (BAg-19)
Forging (commercially pure titanium)
DIAMETRAL CLEARANCE, 0.001 TO 0.003
Tube (type 347)
0.502
0.501
5/16 (approx)
0.500
0.499 OD
Preformed ring, 0.040-in.-diam wire (BAg-19)
DURING BRAZING, ASSEMBLY WAS HELD UPRIGHT (FORGING DOWN)

FURNACE BRAZING IN ARGON	
FURNACE(A)	**PIT**
RETORT(B)	**INCONEL**
FIXTURE MATERIAL(C)	**TITANIUM SHEET**
FURNACE TEMPERATURE, °C (°F)	**925 ± 5 (1700 ± 10)**
BRAZING TEMPERATURE, °C (°F)	**900 ± 5 (1650 ± 10)**
ARGON DEW POINTS, °C (°F)	**-65 (-85)(D); -60 (-70)(E)**
ARGON FLOW RATE (PURGING TO COOLING), M^3/H (FT^3/H)	**4 (150)**
FILLER METAL(F)	**BAG-19**
TIME AT BRAZING TEMPERATURE, MIN	**<1**
COOLING	**IN ARGON, TO ROOM TEMPERATURE**
ASSEMBLY PRODUCTION LOT PER CYCLE	**1-20**

(A) GAS FIRED, 1.8 M (72 IN.) DIAMETER, 1.8 M (72 IN.) DEEP.
(B) 610 MM (24 IN.) DIAMETER AND LENGTH.
(C) HOLDING FIXTURE, TO KEEP ASSEMBLY UPRIGHT (FORGING DOWN) DURING BRAZING.
(D) INCOMING.
(E) EXHAUST.
(F) 1 MM (0.040 IN.) DIAMETER WIRE

FIG. 10 STAINLESS STEEL AND TITANIUM ASSEMBLY THAT WAS FURNACE BRAZED IN AN ARGON ATMOSPHERE

The outside diameter of the titanium forging was held to 12.70 mm (0.500 in.), +0.000 and -0.025 mm (-0.001 in.). The inside diameter of the stainless steel tube was held to 12.73 mm (0.501 in.), +0.025 mm (+0.001 in.) and -0.000 mm or in. This allowed for a diametral clearance of 0.025 to 0.076 (0.001 to 0.003 in.) between components at room temperature. From 0 to 900 °C (32 to 1650 °F), the mean coefficient of thermal expansion (CTE) of commercially pure titanium is 10.3^{-6}/K. From 0 to 870 °C (32 to 1600 °F), the mean CTE of type 347 stainless steel is 20×10^{-6}/K. Calculation of the expansion that would occur when both components were heated to 900 °C (1650 °F) indicated a 0.102 mm (0.004 in.) diametral clearance between the titanium and the stainless steel. Adding the diametral clearance at room temperature (0.025 to 0.076 mm, or 0.001 to 0.003 in.) to the 0.102 mm (0.004 in.) clearance gave a total diametral clearance at brazing temperature of 0.127 to 0.178 mm (0.005 to 0.007 in.), which is within a range that will result in a successful joint.

The selected brazing filler metal was the BAg-19 silver, because it has high fluidity in an argon atmosphere and a brazing temperature that is lower than that of pure silver. Most alloy elements in silver brazing filler metals form brittle intermetallic compounds with titanium, which result in unreliable joints. With the exception of a minute amount of lithium, the only alloying element contained in BAg-19 is 7.5% Cu. By limiting the time at brazing temperature, sound ductile joints were made, and the formation of the titanium copper intermetallic phase was minimized.

Prior to assembly, the titanium forging was degreased and cleaned in a solution that contained 40% nitric acid plus 2% hydrofluoric acid. The stainless steel tubing and the brazing filler metal (preformed rings of 0.102 mm, or 0.004 in., diameter BAg-19 wire) were cleaned by washing in acetone. Operators wore clean, lint-free, white cotton gloves for all subsequent handling of the components during assembly, and kept them on until after brazing was completed.

The titanium forging was inserted into the expanded end of the stainless steel tube until it was completely seated. A brazing filler metal ring was placed around the outside diameter of the titanium tube at the joint intersection. The assembly was placed upright (forging down) in a titanium sheet-metal-holding fixture and was loaded into an Inconel retort. The retort was designed for displacement purging with an inlet and exit manifold for the argon atmosphere. After loading, the retort cover was seal welded to its base, using the GTAW process.

The retort was purged for 30 min with pure, dry argon at 4.2 m^3/h (150 ft^3/h) and then placed in a gas-fired pit-type furnace. The retort was heated to 315 °C (600 °F), and it was held at that temperature for an additional 30 min, or until the dew point of the exiting argon was -60 °C (-70 °F), as recorded on an electrolytic water analyzer. The furnace temperature was then raised until the assembly temperature reached 900 °C (1650 °F), as indicated by a Chromel-Alumel thermocouple attached to the titanium holding fixture within the retort. As soon as this temperature was reached, the retort was removed from the furnace and fan-cooled to room temperature.

The retort cover was opened by grinding away the seal weld, and the assemblies were removed. The titanium and stainless steel components emerged bright and clean, with evidence of excellent brazing filler metal flow. Radiographic inspection showed over 95% joint coverage. All joints were visually inspected on both sides.

Brazing of Stainless Steels

Revised by Matthew J. Lucas, Jr., General Electric Aircraft Engines

---

Furnace Brazing in an Air Atmosphere

The principal advantages of furnace brazing are high production rates and a means for using controlled protective atmospheres at controlled dew points, which often precludes the use of a flux to obtain satisfactory wetting action. In most furnace-brazing applications, both of these advantages are exploited. Occasionally, however, furnace brazing is selected solely on the basis of production rate, and brazing is performed without a protective atmosphere, but with a suitable flux. Under these conditions, the lower-melting-point brazing filler metals are generally selected, as in the following application.

Example 11: Substitution of Furnace Brazing in Air Atmosphere for Torch Brazing.

The gas-valve bobbin assembly shown in Fig. 11 was satisfactorily brazed by both the torch and furnace brazing processes. The choice of process primarily depended on the required production rate. Cost data proved that furnace brazing increased the production rate per hour, reduced the direct labor rate per hour, and reduced the direct labor cost per assembly.

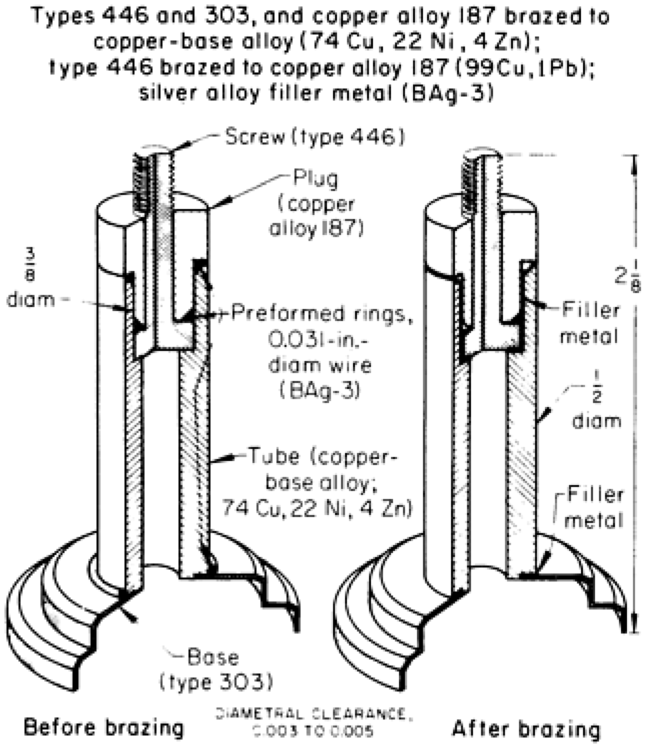
Types 446 and 303, and copper alloy 187 brazed to copper-base alloy (74 Cu, 22 Ni, 4 Zn); type 446 brazed to copper alloy 187 (99Cu, 1Pb); silver alloy filler metal (BAg-3)
Screw (type 446)
Plug (copper alloy 187)
3/8 diam
Preformed rings, 0.031-in.-diam wire (BAg-3)
Tube (copper-base alloy; 74 Cu, 22 Ni, 4 Zn)
Base (type 303)
Before brazing
DIAMETRAL CLEARANCE, 0.003 TO 0.005
After brazing
2 1/8
Filler metal
1/2 diam
Filler metal

FURNACE BRAZING IN AIR ATMOSPHERE	
FURNACE(A)	**CONTINUOUS BELT**
FURNACE TEMPERATURE, °C (°F)	**745-790 (1370-1450)**
FILLER METAL(B)	**BAG-3**
FLUX	**AMS 3410D** **(AWS TYPE 3A)**
TIME AT BRAZING TEMPERATURE, MIN	**1 (MAX)**
ASSEMBLY PRODUCTION RATE/H	**140**

(A) AIR ATMOSPHERE.

(B) PREFORMED RINGS, PREPLACED AS SHOWN IN ILLUSTRATION, 0.8 MM (0.031 IN.)

FIG. 11 STAINLESS STEEL AND COPPER GAS-VALVE BOBBIN THAT WAS FURNACE BRAZED IN AN AIR ATMOSPHERE

As Fig. 11 shows, the bobbin assembly consisted of four parts: a screw made of type 446 stainless steel, a base made of type 303 stainless steel, a tube made of a copper alloy closely related to nickel-silver (74Cu-22Ni-4Zn), and a plug made of copper alloy 187 (99Cu-1Pb), which held the screw in place and blocked gas passage through the tube.

All components were thoroughly vapor degreased before brazing. They were assembled with two preformed rings of 0.787 mm (0.031 in.) diameter BAg-3 brazing filler metal wire. The diametral clearance on the joints was 0.076 to 0.127 mm (0.003 to 0.005 in.). One preform, with a 9.5 mm ($\frac{3}{8}$ in.) internal diameter, was placed over the neck of the plug. The other, with a 12.7 mm (0.5 in.) internal diameter, was placed over the tube adjacent to the base joint. The joint areas were coated with type FB3-A brazing flux, and the assemblies were brazed in a continuous-belt conveyor furnace. Brazing filler metal BAg-3 was chosen, in preference to BAg-1 or BAg-1a, in order to avoid the risk of interface corrosion.

Brazing of Stainless Steels

Revised by Matthew J. Lucas, Jr., General Electric Aircraft Engines

---

Furnace Brazing in a Vacuum Atmosphere

The majority of vacuum brazing is performed in two types of equipment utilizing either hot- or cold-walled furnaces. The hot-walled structure utilizes a retort that is evacuated and placed into a furnace, which provides the heat source. The retort can be single-pumped, providing a vacuum to temperatures up to 980 °C (1800 °F), or double-pumped for temperatures above 980 °C (1800 °F) to prevent the retort from collapsing. Limitations of the hot-walled furnace (retort) include longer cycle times, because the retort is heated externally; a 1200 °C (2200 °F) temperature limit; and slower cooling rates. Some advantages of hot-walled furnaces are lower initial capital expenditures, reduced contamination from the retort, and easy upkeep and maintenance.

The cold-walled vacuum furnace is typically designed with a water-cooled outer jacket that is protected by radiation shielding adjacent to the inner wall. The heating elements are exposed directly to the workload. A braze cycle may have temperatures that exceed 2200 °C (4000 °F), depending on the heating element material and the load-support structure. Heating and cooling rates for cold-walled vacuum furnaces are substantially less than they are for the hot-walled vacuum retort. Higher braze temperatures, in excess of 1260 °C (2300 °F), are obtainable in vacuum furnaces and are frequently employed.

Effect of Filler-Metal Composition. Vacuum brazing of many structural configurations made of austenitic stainless steels offers excellent heat and corrosion resistance for high-temperature service applications. Brazing filler metals, such as gold, gold-palladium, and nickel, offer greater high-temperature strength and oxidation resistance. Problems can occur when brazing the 300 series of stainless steels, because of the carbide precipitation and loss of corrosion resistance that results when brazing in the temperature range from 480 to 815 °C (900 to 1500 °F). Brazing at temperatures in excess of 815 °C (1500 °F), using brazing filler metals with melting points that are higher than this temperature, followed by rapid cooling, will reduce carbide precipitation and improve corrosion properties.

Occasionally, wetting does not occur with a particular lot of stainless steel. Generally, alloying elements such as titanium and aluminum contribute to poor wettability.

Martensitic stainless steels, such as the 400 series, can be brazed successfully in vacuum. The use of a suitable brazing filler metal can result in austenitizing and brazing at the same time, followed by rapid cooling to harden the stainless steel.

Care must be exercised when selecting brazing filler metals for use in vacuum. Silver brazing filler metals, such as BAg-1, BAg-1a, and BAg-3, contain alloying elements of zinc and cadmium, which have very high vapor pressures. These elements vaporize if the furnace pressure is too low or the brazing temperature is excessive, or if a combination of these conditions exists. Copper and silver also vaporize under low-pressure conditions at higher brazing temperatures. Therefore, if brazing alloys containing copper or silver are to be used in vacuum, the furnace chamber must be backfilled with an appropriate atmosphere to 300 to 500 μm (12 to 20 mils) until melting occurs to prevent the loss of these elements. Caution must be exercised with the furnace exit gases because these gases are extremely toxic. Two materials that should never be put in a vacuum furnace under any condition are zinc and cadmium. It is recommended that specific alloys (for example, BAg-8 or BAg-13a) or other alloys free of zinc or cadmium be used when brazing in a vacuum furnace. A partial pressure may still be required for many of the alloys to prevent the vaporization of the silver.

Example 12: Vacuum Brazing to Improve Beverage Can Filling Nozzles.

Because of a continuous problem in obtaining uniform and void-free fillets when using silver brazing filler metals and induction brazing methods, a switch was made to vacuum furnace brazing using a nickel brazing filler metal. The beverage filling nozzle shown in Fig. 12 is an assembly consisting of 16 short tubes brazed to a cast body section that connects to the supply piping. The tubes are pressed into the holes that are drilled at angles through the body, and then brazed to form a tight joint. Both the body and tubes are made from type 304 stainless steel.

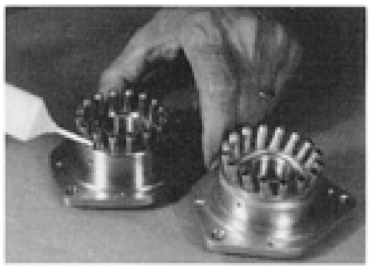

FIG. 12 STAINLESS STEEL TYPE 304 BEVERAGE CAN FILLING NOZZLE. TUBES ARE VACUUM BRAZED WITH A NICKEL BRAZING FILLER METAL AT 1120 °C (2050 °F). LEFT, LOCATION WHERE PASTE ALLOY IS PLACED AROUND TUBE. RIGHT, COMPLETED NOZZLE, SHOWING SMOOTH, VOID-FREE FILLETS. COURTESY OF WALL COLMONOY CORPORATION

The nozzles require a smooth surface that will not harbor bacteria. This means that the brazed joints must have uniform, smooth fillets and full-length void-free filler metal penetration. In addition, the entire assembly must be resistant to chemical attack by the beverages being handled, by steam, or by cleaning compounds.

The prior method of brazing using silver-base filler metals was not very successful in obtaining void-free joints or consistent results. In addition, many of the beverages being handled did attack the silver brazing filler metal, causing corrosion and discoloration in the joint area.

The nickel brazing filler metal selected for this application was BNi-9 (81.5Ni-15Cr-3.5B), because of its excellent capillary flow characteristics, low base-metal erosion, and self-fluxing properties. In addition, the alloy would be unaffected by subsequent processing operations, which include passivation and electropolishing.

The stainless parts were degreased and a small bead of the nickel brazing filler metal was placed around each tube. The brazing was done in a vacuum furnace at 1135 °C (2075 °F), where the part was heated and cooled at a controlled rate to ensure thermal stability. Careful control of the brazing cycle is the key element in obtaining the uniform fillets and void-free joints required for the nozzles.

Example 13: Cryogenic Valve Nickel Brazed in Vacuum.

Cryogenic valves that are used to handle liquid nitrogen and liquid hydrogen require construction that provides high strength and impact resistance at extremely low temperatures, as well as light weight and good corrosion resistance (Fig. 13). Valves that are part of the propellant loading systems for ballistic missiles were used at temperatures of -250 °C (-425 °F), and at pressures of nearly 41.4 MPa (6 ksi). The body, flanges, seat, and bonnet were brazed with BNi-2 brazing filler metal in a vacuum furnace.

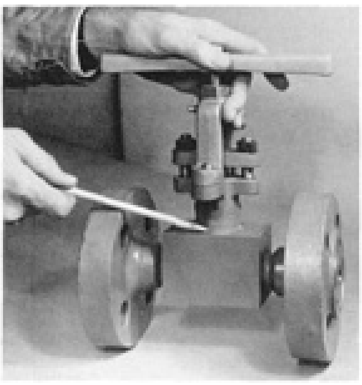

FIG. 13 CRYOGENIC VALVE THAT WAS VACUUM BRAZED WITH BNI-2 BRAZING FILLER METAL. ALL JOINTS ON THE BODY, FLANGES, SEAT, AND BONNETS WERE BRAZED SIMULTANEOUSLY. COURTESY OF WALL

The valve seat and disc were cast from 17-7 or 17-4 PH steel or from stellite. They were brazed in place at the same time the assembly was brazed. Brazing replaced the welding of the seat, which in this case had trouble in terms of weld deposit cracking. All parts, except the valve seat, were finish-machined before brazing. By eliminating the need for machining after brazing, costs were reduced enough to pay for the brazing operation.

Brazing of Stainless Steels

Revised by Matthew J. Lucas, Jr., General Electric Aircraft Engines

Induction Brazing

The behavior of stainless steels, in terms of heating via electrical induction, depends on the metallurgical and physical properties of the particular material. This behavior can be considerably different from that of carbon and low-alloy steels, as well as from that of the more widely used nonferrous metals. In addition, the response of the steel to induction heating varies considerably, depending on whether a stainless steel is magnetic or nonmagnetic at room temperature. Differences in specific heat and electrical conductivity markedly affect the response to heating by induction.

Ferritic and martensitic (SAE 400 series) stainless steels are ferromagnetic at all temperatures, up to the Curie temperature. Thus, given the same power input, these steels generally heat faster than austenitic stainless steels, which are nonmagnetic in the annealed condition. Although cold working may induce slight magnetism in the austenitic chromium steels, the 400 series of chromium-nickel steels are strongly magnetic. The rate of heating to the temperature at which the filler metal flows usually affects induction-coil design and coupling. It may also influence the selection of power output frequency and other processing variables.

Stainless steels can be induction brazed in an air atmosphere, using a suitable flux. However, for critical applications, induction brazing is sometimes done in a protective or a vacuum atmosphere (refer to Example 14). In other applications, an inert gas, such as argon, can be used as a protective atmosphere to either minimize or prevent oxidation.

Example 14: Brazing a Tube to an End Blank.

The assembly shown in Fig. 14, which is part of a solenoid, consisted of a type 321 stainless steel brazed to a type 416 end blank. The former material is a nonmagnetic austenitic steel, whereas the latter is martensitic and ferromagnetic. Consequently, although both metals were easily brazed, the achievement of a proper joint clearance between the two components was complicated by the marked differences in the CTEs of the two steels. Thus, calculations were needed to determine the room-temperature clearance required to provide a suitable clearance at the brazing temperature.

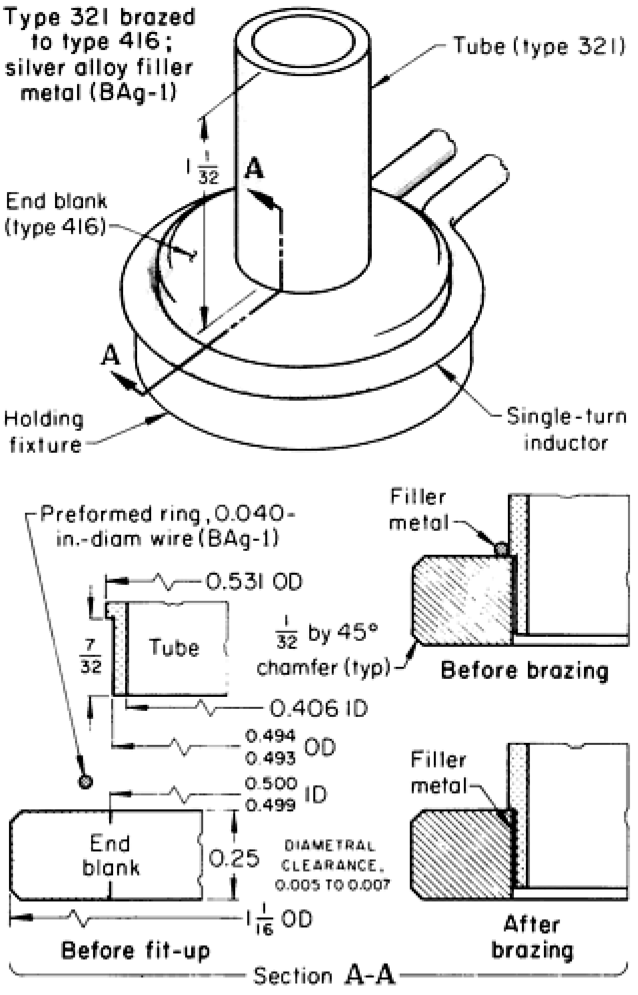
Type 321 brazed to type 416; silver alloy filler metal (BAg-1)
Tube (type 321)
1 1/32
A
End blank (type 416)
A
Holding fixture
Single-turn inductor
Preformed ring, 0.040-in.-diam wire (BAg-1)
0.531 OD
7/32
Tube
0.406 ID
0.494 0.493 OD
0.500 0.499 ID
End blank
0.25
DIAMETRAL CLEARANCE, 0.005 TO 0.007
1 1/16 OD
Before fit-up
1/32 by 45° chamfer (typ)
Filler metal
Before brazing
Filler metal
After brazing
Section A-A

INDUCTION BRAZING	
POWER SUPPLY[(A)], KW; KHZ	**20; 450**
INDUCTOR	**SINGLE-TURN, COPPER TUBE**
BRAZING TEMPERATURE, °C (°F)	**650 (1200)**
FILLER METAL[(B)]	**BAG-1**
FLUX	**AWS TYPE FB3A**
TIME AT BRAZING TEMPERATURE, S	**10**
COOLING TIME IN FIXTURE, S	**10**
ASSEMBLY PRODUCTION RATE/H	**140**

(A) VACUUM TUBE.

FIG. 14 INDUCTION-BRAZED ASSEMBLY

Because the assembly was not intended for high-temperature service, the selection of the low-melting-point silver brazing filler metal (BAg-1) and a relatively low brazing temperature of 650 °C (1200 °F) were utilized to minimize heating and oxidation of the stainless steel components.

For brazing that is conducted at this temperature, calculations based on CTE values showed that the following dimensions and tolerances in the joining area would be satisfactory: for the tube diameter, 12.4 mm (0.494 in.) +0.000, -0.025 mm (-

0.001 in); and for the inside diameter of the end blank, 12.7 mm (0.500 in.) +0.000, -0.025 mm (-0.001 in.). Thus, the diametral clearance at room temperature ranged from 0.127 to 0.178 mm (0.005 to 0.007 in.).

The shape of the assembly and the low melting point and brazing temperatures favored brazing by induction. The end blank was in the hardened and tempered condition prior to brazing, and the short induction heating cycle (10 s) did not reduce the hardness to less than the required minimum.

Prior to brazing, the components were vapor degreased. The end of the tube was dipped in flux and inserted in the end blank. A preformed ring of filler-metal wire was slipped over the tube and positioned at the top of the joint. Then, the end blank was placed on a holding fixture, positioned in a single-turn inductor (Fig. 14), and heated for 10 s. After brazing, the assembly was cooled in air for 10 s before being removed from the holding fixture. The assembly was then washed in hot water to remove the flux residue.

Example 15: Induction Brazing in a Vacuum Atmosphere.

A distinctive advantage of induction brazing, as applied to stainless steel, is its suitability for simple setups that permit brazing in vacuum. Closed, nonmetallic containers with reasonably good strength and dielectric properties can provide an enclosure for the assembly to be brazed and can be evacuated prior to brazing. Because the inductor can be placed outside the container, it can heat the assembly efficiently without being part of the vacuum system.

Stainless steel collar-and-tube assemblies (Fig. 15) were brazed in a simple setup that combined induction heating and the protection afforded by heating in vacuum. The vacuum container consisted of a high-silica, low-expansion glass tube with copper end fittings connected to a vacuum system. The collar-and-tube components, with preformed BNi-7 filler-metal rings pressed into place on the shoulder of each collar, were positioned and held inside the glass tube by means of a simple holding fixture. The tube was sealed and evacuated with a multiple-turn inductor, outside the tube, in position to heat one of the collars. When the vacuum reached 0.133 Pa (10^{-3} torr), heating was started. The collar was heated slowly to 970 °C (1775 °F). After 4 min, the power was shut off. The tube was then repositioned to bring the second collar into the field of the inductor, and the heating sequence was repeated.

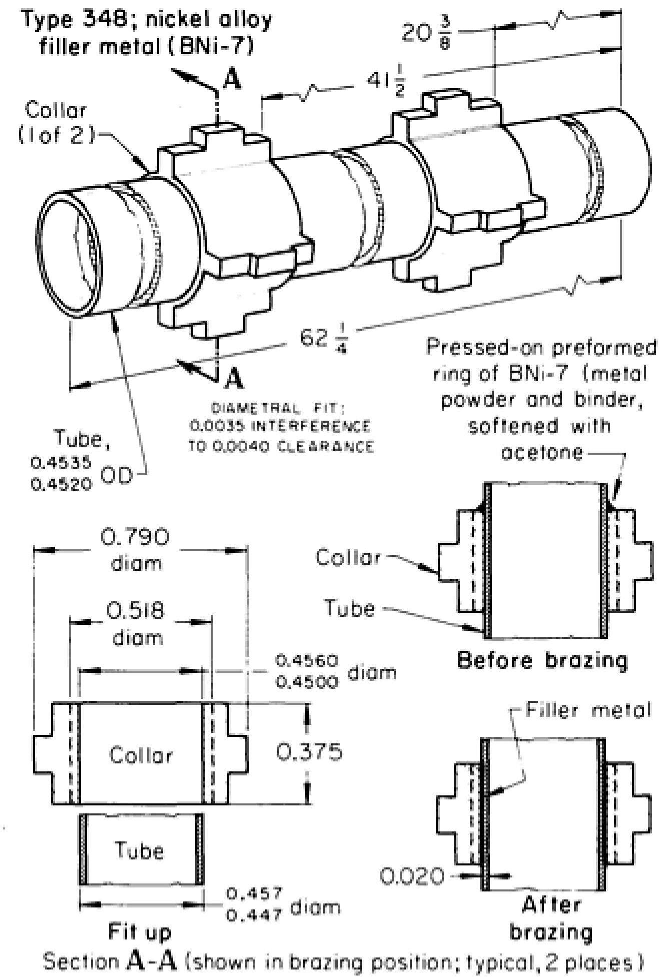
Type 348; nickel alloy filler metal (BNi-7)
A
Collar (1 of 2)
20 3/8
41 1/2
62 1/4
A
DIAMETRAL FIT: 0.0035 INTERFERENCE TO 0.0040 CLEARANCE
Pressed-on preformed ring of BNi-7 (metal powder and binder, softened with acetone
Tube, 0.4535 0.4520 OD
0.790 diam
0.518 diam
0.4560 0.4500 diam
Collar
Tube
Before brazing
Filler metal
Collar
0.375
Tube
0.020
0.457 0.447 diam
Fit up
After brazing
Section A-A (shown in brazing position; typical, 2 places)

FIG. 15 COLLAR AND TUBE ASSEMBLY THAT WAS INDUCTION BRAZED IN VACUUM

When the second collar had cooled to the point at which no glow was visible in normal light, the tube was backfilled for 5 min with argon. Brazed joints were inspected visually and metallographically. They were found to be sound and acceptable in all respects. The induction heating source was an 8 kVA spark-gap converter with an operating frequency of 175 to 200 kHz. The water-cooled external inductor coil was made of 6.4 mm ($\frac{1}{4}$ in.) diameter copper tubing. The production rate was 22 assemblies per day.

Brazing of Stainless Steels

Revised by Matthew J. Lucas, Jr., General Electric Aircraft Engines

---

Dip Brazing in a Salt Bath

The brazing of stainless steel by immersing either all or a portion of the assembly in molten salt offers essentially the same advantages and limitations that would apply to the brazing of similar assemblies made of carbon steel.

Example 16: Change From Torch or Induction Brazing to Dip Brazing.

The television wave-guide assembly shown in Fig. 16 consisted of a type 304 stainless steel flange brazed to a tube of copper alloy 230 (red brass, 85Cu-15Zn). Satisfactory end use depended on minimal distortion. When the assembly was brazed by torch or induction brazing, the rejection rate sometimes reached 70%, because of distortion caused by uneven heating. When dip brazing was adopted, the rejection rate dropped to nearly zero.

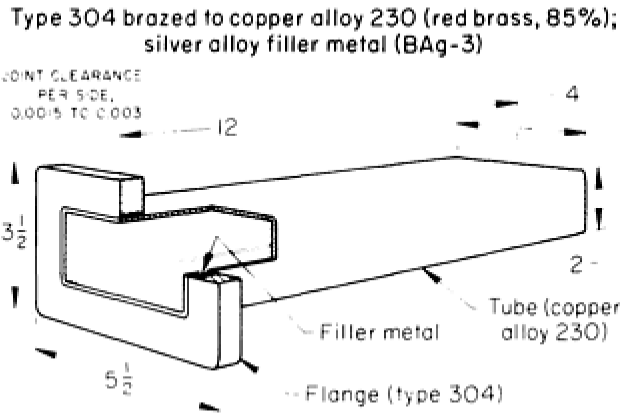
Type 304 brazed to copper alloy 230 (red brass, 85%);
silver alloy filler metal (BAg-3)
JOINT CLEARANCE PER SIDE, 0.0015 TO 0.003
12
4
$3\frac{1}{2}$
2
Tube (copper alloy 230)
Filler metal
$5\frac{1}{2}$
Flange (type 304)

FIG. 16 TELEVISION WAVE-GUIDE ASSEMBLY JOINED BY SALT-BATH DIP BRAZING

Prior to brazing, the stainless steel flange was degreased and pickled, and the brass tube was degreased and bright dipped. Then, the flange was placed on the tube, the tube end was flared outward slightly, a preform of the BAg-3 filler metal was placed over the tube adjacent to the flange, and FB3-A flux (AMS 3410D) flux was applied to the joint. The assembly was suspended flange-down over an electrically heated salt bath to preheat the flange and dry the flux. Next, the assembly was lowered slowly into the molten bath, which was maintained at 730 °C (1350 °F) for a distance of approximately 25 mm (1 in.) above the flange. After being held in the bath for 0.5 min, the assembly was removed and air cooled. The flux residue was removed by rinsing the assembly in 60 °C (140 °F) water. The production rate was 30 assemblies per hour.

Brazing of Stainless Steels

Revised by Matthew J. Lucas, Jr., General Electric Aircraft Engines

---

High-Energy-Beam Brazing

High-energy-beam brazing techniques, such as those based on electron or laser beams, have been used to a limited extent. Both electron- and laser-beam brazing are performed in a manner similar to electron- and laser-beam welding, except that the beam is defocused to provide a larger beam and to reduce the power density to prevent the base metal from melting. Generally, the speed at which the beam is swept is increased so that a larger area of the part is heated and more uniform heating of the part occurs.

In electron-beam brazing, the high vacuum used in the work chamber (0.0133 to 0.00133 Pa, or 10^{-4} to 10^{-5} torr) permits the adequate flow of brazing filler metal on properly cleaned joints without the use of a reducing atmosphere or flux. Thus, flux entrapment does not occur, and the work does not require cleaning after brazing. The high-vacuum atmosphere and the absence of flux provide a brazing environment that avoids the problems associated with prepared atmospheres, which are encountered when brazing some stainless steels, as well as the more-reactive metals (such as titanium).

In laser-beam brazing, the paris are normally protected with a shielding gas to prevent the occurrence of oxidation. If necessary, the beam spot diameter can be enlarged substantially, depending on the type of equipment, while providing an adequate amount of heat input for brazing. A work movement technique can be used if an area substantially larger than the beam spot size is to be heated, and the work can be rotated or indexed under the beam for uniform heating. In high-energy brazing, the brazing temperatures are quickly attained, and heat can be localized to minimize grain growth, the softening of cold-worked metal, and, in austenitic stainless steels, the sensitizing of the material by carbide precipitation.

Applications. Electron-beam brazing is a convenient method for brazing small assemblies, such as instrument packages. It combines the versatility and close controllability of electron-beam heating with the advantages of vacuum brazing. Packaged devices can be encapsulated with an internal vacuum without damaging the basic package.

Tube-to-header joints in small heat-transfer equipment made of heat-resistant alloys and refractory metals are sometimes brazed using an electron beam. In one technique, the tube-to-header joint is electron-beam welded on the top side of the header. The heat of the beam causes the brazing filler metal that is preplaced on the reverse side of the header at the joint to melt and flow. Small-diameter, thin-walled stainless steel tubes are readily joined by electron-beam brazing, as in the following example.

Example 17: Use of Defocused Beam for the Electron-Beam Brazing of Small Tubes.

Capillary and other small-diameter tubes used in instrument packages required that leak-tight joints be made without overheating the other portions of the assembly. The avoidance of flux was also necessary, because entrapped flux would be either difficult or impossible to remove. These conditions were satisfied by electron-beam brazing.

Figure 17 shows a typical joint in type 304 tubing with a 2.55 mm (0.100 in.) outside diameter and a 0.254 mm (0.010 in.) thick wall that was brazed by the electron-beam process. The joint design was based on the use of a 19 mm ($\frac{3}{4}$ in.) long socket coupling that was counterbored with a diametral clearance of 0.076 to 0.127 mm (0.003 to 0.005 in.) over the tube diameter and to a depth of 6.4 mm ($\frac{1}{4}$ in.). The average joint clearance (per side) was therefore 0.050 mm (0.002 in.). Tubes and socket couplings were deburred and solvent cleaned. They were then assembled with two wire-ring preforms of BCu-1a brazing filler metal, as shown in Fig. 17. The tubes were held in position with a small clamping fixture, and the assembly was mounted in a fixed position on a table in the vacuum chamber.

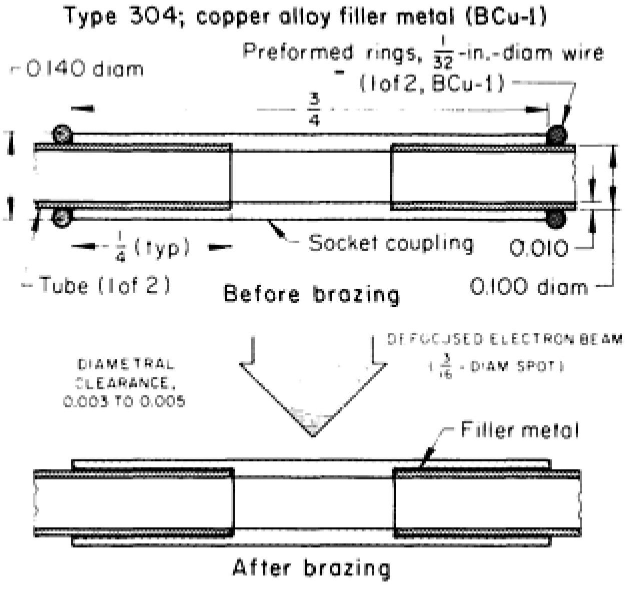
Type 304; copper alloy filler metal (BCu-1)
0.140 diam
Preformed rings, 1/32-in.-diam wire (1 of 2, BCu-1)
3/4
1/4 (typ)
Socket coupling
0.010
Tube (1 of 2)
Before brazing
0.100 diam
DIAMETRAL CLEARANCE, 0.003 TO 0.005
DEFOCUSED ELECTRON BEAM (3/16 - DIAM SPOT)
Filler metal
After brazing

ELECTRON-BEAM BRAZING	
JOINT TYPE	**CYLINDRICAL SLEEVE**
FILLER METAL$^{(A)}$	**BCU-1**
MACHINE CAPACITY, KW	**3**
GUN TYPE	**FIXED DIODE**
VACUUM CHAMBER DIAMETER, MM (IN.)	**610 (24)**
MAXIMUM VACUUM, PA (TORR)	**0.00133 (1 × 10^{-5})**
FIXTURE	**HOLDING JIG**

PUMPDOWN TIME, MIN	**30**
BRAZING POWER, KV; MA	**18; 20-30**
BEAM SPOT SIZE DIAMETER, MM (IN.)	**4.76 ($\frac{3}{16}$)**
BRAZING VACUUM, PA (TORR)	**0.00133 (1×10^{-5})**
BRAZING TIME, S	**30**

(A) WIRE FORM IN 0.03 MM ($\frac{1}{32}$ IN.) DIAMETER

FIG. 17 JOINT BETWEEN TWO CAPILLARY TUBES OF AN INSTRUMENT PACKAGE, MADE BY ELECTRON-BEAM BRAZING USING A LOW-POWER DEFOCUSED BEAM IN A HIGH-VACUUM ATMOSPHERE

After pumpdown, the joint was brought to the brazing temperature by moving the table back and forth under the defocused electron beam, which caused the heat of the 4.76 mm ($\frac{3}{16}$ in.) diameter beam spot to be applied primarily to the central portion of the coupling. After being heated by conduction at a relatively low beam power, the brazing filler metal melted and flowed through the joint in approximately 30 s.

About ten assemblies were brazed, at a rate of one per pumpdown, using the machine settings and other brazing conditions specified in Fig. 17. Sensitizing the austenitic stainless steel was not a problem in this application, because the service environment was not significantly corrosive. The relatively short-time brazing cycle minimized grain growth and the dilution of the thin-walled tubing with copper brazing filler metal.

Brazing of Heat-Resistant Alloys, Low-Alloy Steels, and Tool Steels

Revised by Darrell Manente, Vac-Aero International Inc.

Introduction

PROCEDURES for brazing heat-resistant alloys, low-alloy steels, and tool steels share much in common. This article focuses primarily on brazing of heat-resistant alloys; some information about particulars involved in the brazing of low-alloy steels and tool steels is also provided.

Brazing of Heat-Resistant Alloys, Low-Alloy Steels, and Tool Steels

Revised by Darrell Manente, Vac-Aero International Inc.

Brazing of Heat-Resistant Alloys

Heat-resistant alloys are frequently referred to as superalloys, because of their strength, oxidation resistance, and corrosion resistance at elevated service temperatures (650 to 1205 °C, or 1200 to 2200 °F). This article discusses primarily nickel- and cobalt-base alloys. Superalloys can also be subdivided (according to manufacturing technology) into two categories: conventional cast and wrought alloys and powder metallurgy (P/M) products. Powder metallurgy products may be produced in conventional alloy compositions and as oxide-dispersion-strengthened (ODS) alloys. Almost any metal, as well as nonmetallics, can be brazed to these heat-resistant alloys, if it can withstand the heat of brazing. For additional information on these alloys, see Volume 1 of the *ASM Handbook.*

Brazing Filler Metals

The American Welding Society (AWS) has classified several gold-, nickel-, and cobalt-base brazing filler metals which can be used for elevated-temperature service (Table 1). These brazing filler metals are suitable for high-temperature service; however, if the application is for temperatures above 980 °C (1800 °F) or in severe environments, the required brazing filler metal may not be in Table 1. It should be noted that for lower service temperatures, copper (BCu) (below 480 °C, or 900 °F) and silver (BAg) (below 425 and 200 °C, or 800 and 400 °F) brazing filler metals have been used for many successful applications.

TABLE 1 AWS BRAZING ALLOYS FOR ELEVATED-TEMPERATURE SERVICE

AWS CLASSIFICATION	**COMPOSITION, WT%**														**SOLIDUS**		**LIQUIDUS**		BRAZING RANGE	
	Cr	**B**	**Si**	**Fe**	**C**	**P**	**S**	**Al**	**Ti**	**Mn**	**Cu**	**Zr**	**Ni**	**OTHER ELEMENTS TOTAL**	**°C**	**°F**	**°C**	**°F**	**°C**	°F
NICKEL-BASE ALLOY FILLER METALS(A)																				
BNI-1	13.0-15.0	2.75-3.50	4.0-5.0	4.0-5.0	0.6-0.9	0.02	0.02	0.05	0.05	. . .	. . .	0.05	BAL	0.50	975	1790	975	1900	1065-1205	**1950-2200**
BNI-1A	13.0-15.0	2.75-3.50	4.0-5.0	4.0-5.0	0.06	0.02	0.02	0.05	0.05	. . .	. . .	0.05	BAL	0.50	975	1790	1075	1970	1075-1205	**1970-2200**
BNI-2	6.0-8.0	2.75-3.50	4.0-5.0	2.5-3.5	0.06	0.02	0.02	0.05	0.05	. . .	. . .	0.05	BAL	0.50	970	1780	1000	1830	1010-1175	**1850-2150**
BNI-3	. . .	2.75-3.50	4.0-5.0	0.5	0.06	0.02	0.02	0.05	0.05	. . .	. . .	0.05	BAL	0.50	980	1800	1040	1900	1010-1175	**1850-2150**
BNI-4	. . .	1.5-2.2	3.0-4.0	1.5	0.06	0.02	0.02	0.05	0.05	. . .	. . .	0.05	BAL	0.50	980	1800	1065	1950	1010-1175	**1850-2150**
BNI-5	18.5-19.5	0.03	9.75-10.50	. . .	0.10	0.02	0.02	0.05	0.05	. . .	. . .	0.05	BAL	0.50	1080	1975	1135	2075	1150-1205	**2100-2200**
BNI-6	. . .	. . .	. . .	. . .	0.10	10.0-12.0	0.02	0.05	0.05	. . .	. . .	0.05	BAL	0.50	875	1610	875	1610	925-1095	**1700-2000**
BNI-7	13.0-15.0	0.01	0.10	0.2	0.08	9.7-10.5	0.02	0.05	0.05	0.04	. . .	0.05	BAL	0.50	890	1630	890	1630	925-1095	**1700-2000**
BNI-8	**. . .**	**. . .**	**6.0-8.0**	**. . .**	**0.10**	**0.02**	**0.02**	**0.05**	**0.05**	**21.5-24.5**	**4.0-5.0**	**0.05**	**BAL**	**0.50**	**980**	**1800**	**1010**	**1850**	**1010-1095**	**1850-2000**

AWS CLASSIFICATION	**COMPOSITION, WT%**					**SOLIDUS**		**LIQUIDUS**		BRAZING RANGE	
	AU	**CU**	**PD**	**NI**	**OTHER ELEMENTS TOTAL**	**°C**	**°F**	**°C**	**°F**	**°C**	°F
PRECIOUS METALS											
BAU-1	37.0-38.0	BAL	. . .	. . .	0.15	990	1815	1015	1860	1015-1095	**1860-2000**
BAU-2	79.5-80.5	BAL	. . .	. . .	0.15	890	1635	890	1635	890-1010	**1635-1850**
BAU-3	34.5-35.5	BAL	. . .	2.5-3.5	0.15	975	1785	1030	1885	1030-1090	**1885-1995**
BAU-4	81.5-82.5	. . .	. . .	BAL	0.15	950	1740	950	1740	950-1005	**1740-1840**
BAU-5	**29.5-30.5**	**. . .**	**33.5-34.5**	**35.5-36.5**	**0.15**	**1135**	**2075**	**1165**	**2130**	**1165-1230**	**2130-2250**

AWS CLASSIFICATION	**COMPOSITION, WT%**														**SOLIDUS**		**LIQUIDUS**		BRAZING RANGE	
	Cr	**Ni**	**Si**	**W**	**Fe**	**B**	**C**	**P**	**S**	**Al**	**Ti**	**Zr**	**Co**	**OTHER ELEMENTS TOTAL**	**°C**	**°F**	**°C**	**°F**	**°C**	°F
COBALT-BASE ALLOY FILLER METALS																				

BCO-1	**18.0-20.0**	**16.0-18.0**	**7.5-8.5**	**3.5-4.5**	**1.0**	**0.7-0.9**	**0.35-0.45**	**0.02**	**0.02**	**0.05**	**0.05**	**0.05**	**BAL**	**0.50**	**1120**	**2050**	**1150**	**2100**	**1150-1230**	**2100-2250**

Source: AWS 5.8

(A) IF DETERMINED, COBALT IS 0.1% MAXIMUM UNLESS OTHERWISE SPECIFIED.

Many nickel-palladium-base brazing filler metals exhibiting good wetting and flow are not classified by AWS but are also available. These filler metals have been developed to replace gold-containing (BAu) brazing filler metals, which are more expensive.

Another group of brazing filler metals not classified by AWS are used for repair and overhaul of nickel- and cobalt-base superalloy aerospace and industrial turbine engine components. Many of these brazing filler metals are proprietary compositions of the engine manufacturer for which they were developed. However, a good variety of these filler metals are commercially available. These "diffusion braze" filler metals do not contain silicon. However, they do contain boron, which acts as a melting point suppressant. Both boron and silicon react with metals (such as chromium and molybdenum) to form intermetallic phases. These intermetallic phases are very hard and brittle and can seriously weaken a brazed joint. It is possible to perform annealing heat treatments on diffusion brazed joints that result in the diffusion of the boron out of the brazed joint and into the base metal. The result is a brazed joint in which intermetallic phases are partially dissolved or eliminated completely (Fig. 1). Diffusion heat treatment increases the strength and ductility of the brazed joint.

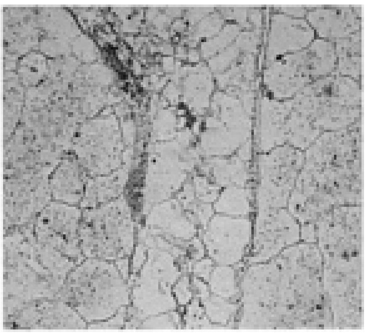

FIG. 1 PHOTOMICROGRAPH OF THE CRACK REPAIR OF A DIFFUSION BRAZED JOINT BY HEAT TREATING OF THE IN-792 BASE METAL. SAMPLE WAS ETCHED WITH KALLING'S REAGENT. 76×

Silicon atoms do not diffuse as easily as boron atoms (because silicon atoms are larger and sit on a regular lattice site, whereas boron atoms sit on interstitial sites, providing much faster diffusion paths). Therefore, similar heat treatments do not tend to dissolve or eliminate silicon intermetallic phases.

These boron-containing filler metals (used on their own or mixed with base metal powder to act as a gap filler) are used to fill cracks and eroded areas on turbine engine components. Some type of cleaning operation (such as hydrogen fluoride cleaning or mechanical cleaning) is employed before the brazing operation in order to remove oxides on the components and thus facilitate brazing filler metal wetting.

A second group of filler metals not classified by AWS are used for wide gaps (1.52 mm, or 0.060 in., and greater) and eroded areas. These filler metals are generally nickel- or cobalt-base superalloys that have additions of a melting point suppressant. These filler metals do not tend to flow away from the area in which they are placed. The microstructure of the joint formed using these filler metals is homogeneous and more closely approximates the base metal microstructure (Fig. 2).

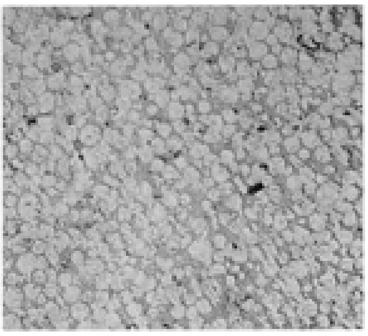

FIG. 2 MICROSTRUCTURE OF A WIDE GAP FILLER METAL BRAZE REPAIR. SAMPLE WAS ETCHED WITH KALLING'S REAGENT. 40×

Generally, heat-resistant alloys are brazed with nickel- or cobalt-base alloys containing boron and/or silicon, which serve as melting-point depressants. In many commercial brazing filler metals, the levels are 2 to 3.5% B and 3 to 10% Si. Phosphorus is another effective melting-point depressant for nickel and is used in filler metals from 0.02 to 10%. It is also used where good flow is important in applications of low stress, where service temperatures do not exceed 760 °C (1400 °F).

In addition to boron, silicon, and phosphorus, chromium is often present to provide oxidation and corrosion resistance. The amount may be as high as 20%, depending on the service conditions. Higher amounts, however, tend to lower brazement strength.

Cobalt-base filler metals are used mainly for brazing cobalt-base components, such as first-stage turbine vanes for jet engines. Most cobalt-base filler metals are proprietary. In addition to containing boron and silicon, these alloys usually contain chromium, nickel, and tungsten to provide corrosion and oxidation resistance and to improve strength.

Product Forms. Available forms of AWS classified and proprietary brazing filler metals include wire, foil, tape, paste, and powder. The form used can be dictated by the application or by the composition of the filler metal. If the filler metal required for a specific application is only available as a dry powder, then brazing aids such as cements and pastes are available to help position the brazing filler metal.

Brazing filler metal powders usually are gas atomized and sold in a range of specified particle sizes, which ensures uniform heating and melting of the brazing filler metal during the brazing cycle. These powders can be mixed with plasticizers or organic cements to facilitate positioning. If the mixture must support its own weight until the brazing cycle begins, an organic binder or cement is required. These binders burn off in atmosphere brazing and little or no residue results. When the brazing filler metal is supplied as a paste, it is simply a premixed powder and binder.

Brazing Tapes and Foils. Brazing filler metals in the form of tapes and amorphous foils appear similar, but the foils are usually made by rapid solidification during melt spinning operations and tend to exhibit homogeneous amorphous structures. Brazing tapes are made of powder that is held together by a binder and formed into a rather fragile sheet. Most amorphous foils have a high metalloid (phosphorus, silicon, boron) content, while tapes can be made from brazing filler metals that have no metalloid content. The metalloids usually are melting-point depressants and frequently form brittle phases. In some cases, where the composition is workable, such as BAu, foils can be made by cold rolling. Foil products can also be produced by rolling an alloy of suitable composition into a foil before adding the metalloids. Tapes and foils are best suited for applications requiring a large area joint, good fit-up, or where brazing flow and wetting may be a problem.

Brazing wires of nonfabricable alloys usually are made by P/M processes from gas atomized powder, which is held together by a binder or by extruding powder into wire and sintering. This form of brazing filler metal is better able to support itself than are pastes and powders of filler metal, but is not used to replace tapes or amorphous foils where precision is needed in preplacing the filler metal, such as joint gaps less than 0.13 mm (0.005 in.).

Surface Cleaning and Preparation

Cleaning of all surfaces that are involved in the formation of the desired brazed joint is necessary to achieve successful and repeatable brazed joints. All obstruction to wetting, flow, and diffusion of the molten brazing filler metal must be removed from both surfaces to be brazed prior to assembly. The presence of contaminants on one or both surfaces to be brazed may result in void formation, restricted or misdirected filler-metal flow, and contaminants included within the solidified brazed area, which reduces the mechanical properties of the resulting brazed joint. Common contaminants are oils, greases, residual liquid penetrant fluids, pigmented markings, residual casting or coring materials, and oxides formed either through previous thermal exposures or by exposures to contaminating environments or engine service.

Chemical cleaning methods are most widely used. As part of any chemical cleaning procedure for preprocessing assemblies for brazing, a degreasing solvent should be used to remove all oils and greases and to ensure wettability of the chemicals used for cleaning. Oils and greases form a very thin film on metals, which prevents wettability by both the subsequent chemical cleaning and/or the molten filler metals. Oils and grease removers that are widely used include degreasing solutions such as stabilized perchloroethylene or stabilized trichloroethylenes. These may be used as simple manual soaks, sprays, or by suspending the parts in a hot vapor of the aforementioned chemicals, commonly referred to as a vapor degreasing process. In conjunction with these processes, anodic and cathodic electrolytic cleaning can be used.

Other cleaning methods include alkaline-based cleaners and emulsion-type cleaners. Alkaline-based cleaners use mixtures of phosphates, carbonates, silicates, soaps, detergents, and hydroxides. Emulsion-type cleaners contain wetting agents, surface actuators, fatty acids, and hydrocarbons.

A chemical cleaning procedure can be a simple single-step process or may involve multistep operations. If the surfaces of the brazed joint are in the machined condition, vapor degreasing may be sufficient to remove machining oils, handling oils, and liquid penetrant fluids to yield a sound, clean surface for brazing. If, on the other hand, one or both of the surfaces to be brazed is not a machined surface, then additional chemical cleanings should be employed. Once vapor degreasing is accomplished, care must be taken to maintain the surface cleanliness of the brazed components by handling in an environmentally clean atmosphere. Additional methods of chemical cleaning to remove oxides and other adherent metallic contaminants include immersion in phosphate acid cleaners or acid pickling solutions, which are comprised of nitric, hydrochloric, or sulfuric acids or combinations of these.

Care must be taken in time of exposure for both acid cleaning and acid pickling of heat-resistant base metals. Overexposure during chemical cleaning can lead to excessive metal loss, grain-boundary attack, and selective attack on microstructural phases. As the last step in chemical cleaning, an ultrasonic cleaning in alcohol or clean hot water is recommended to ensure removal of all traces of previous cleaning solutions.

If no subsequent mechanical cleaning is used, the components to be brazed should be stored and transported to the braze preparation areas in dry, clean containers, such as plastic bags. The time between cleaning and braze application to the assembled joints should be kept as short as manufacturing processes allow.

Mechanical cleaning usually is confined to those metals with heavy tenacious oxide films or to repair brazing on components exposed to service. (Nickel- or cobalt-base superalloy components that require repair brazing can be processed in hydrogen fluoride atmospheres in order to remove oxide films.) Mechanical methods are standard machining processes--abrasive grinding, grit blasting, filing, or wire brushing (stainless steel bristles must be used). These are used not only to remove surface contaminants, but to slightly roughen or fray the surfaces to be brazed.

Care must be taken that the surfaces are not burnished and that mechanical cleaning materials are not embedded in the metals to be brazed. In grit blasting, choice of the blasting medium is critical. Wet and dry grit blasting commonly are used, but wet mediums are subject to additional cleaning requirements to remove the embedded grit. The mediums used are iron grits, silicon carbide grits, and grits composed of brazing filler metals. Grit sizes as coarse as No. 30 (0.589 mm, or 0.0232 in.) are recommended for cleaning forgings and castings. Finer grits (No. 90 and No. 100, 0.165 and 0.150 mm [0.0065 and 0.0059 in.], respectively) are used for general blasting. All grit mediums should be changed frequently, as extensive reuse of the same medium results in loss of sharp angles or facets. Once these configurations are lost or markedly reduced in the grit medium, burnishing rather than cleaning occurs. Overused medium results in the entrapment of oxides in the metal. If possible, the angle of grit blasting should be less than 90° to the surfaces to be cleaned. This also reduces the chances of embedding the oxides or medium in the surfaces to be brazed.

Care must be taken to remove all blasting medium from the surface after mechanical cleaning, as these mediums will contaminate the braze. Iron grit may impart an iron film which oxidizes as a rust. Aluminum oxides, if not removed, prevent wetting and flow of the brazing filler metal; thus, use of aluminum oxides are not recommended. Silicon carbide is extremely hard and has sharp facets. Consequently, it becomes embedded if an improper blasting angle is used. Blasting with a nickel-base brazing filler metal or similar alloy gives the best results; stainless steel blasting medium is also acceptable.

After mechanical cleaning, air blasting or ultrasonic cleaning should be used to remove all traces of loosened oxides or cleaning medium. Care must be taken to ensure maintenance of the clean surfaces and components; once cleaned, they should be assembled and brazed as soon as possible.

Hydrogen Fluoride Cleaning. The successful repair of superalloy components requires that oxides formed during service, on external surfaces and within cracks, be removed prior to brazing to ensure proper wetting and flow of the braze material. For cobalt-base superalloy parts, heating in a hydrogen-rich atmosphere (hydrogen reduction) is generally successful in this regard. However, hydrogen reduction is not effective in reducing oxides of aluminum and titanium that are formed on nickel-base superalloy parts in service because these oxides exhibit a high degree of thermodynamic stability. Hydrogen fluoride reduction, on the other hand, has been shown capable of reducing these oxides from the surfaces of nickel-base parts.

Oxides of aluminum and titanium are changed to their respective metallic fluorides by the fluoride ions created in the hydrogen fluoride process. Oxides of chromium are also changed to chromium fluoride by the fluoride ions. In addition, workpiece surfaces are depleted of aluminum and titanium so that oxides of these metals do not tend to reform during subsequent operations, such as braze repair (Fig. 3).

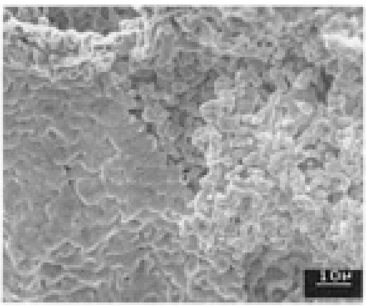

FIG. 3 SEM PHOTOMICROGRAPH OF A RENÉ 125 AEROSPACE ENGINE COMPONENT THAT WAS CLEANED WITH HYDROGEN FLUORIDE TO DEPLETE THE ELEMENTS AT THE SURFACE

Nickel Flashing. Certain heat-resistant alloys that are used as base metals in brazed assemblies--particularly nickel-base alloys containing high percentages of aluminum and titanium (such as Inconel 718)--may require a surface pretreatment to ensure maintenance of the cleaned surfaces. This surface pretreatment after cleaning is generally an electroplate of nickel, commonly referred to as nickel flashing. Thickness of the plate flashing is kept under 0.015 mm (0.0006 in.) for alloys with less than 4% (Ti + Al) and 0.020 to 0.030 mm (0.0008 to 0.0012 in.) for alloys with greater than 4% (Ti + Al). This promotes wettability in the braze joint without seriously affecting the braze strength and other mechanical properties of the braze. The thickness of nickel plating may have to be increased as the brazing temperature is increased and as the time above 980 °C (1800 °F) is increased. Titanium and aluminum will diffuse to the surface of the nickel plating upon heating.

Fixturing

One prerequisite to successful brazing that is often neglected is proper fixturing. Two types of fixtures are used: "cold" fixtures and "hot" fixtures.

Cold fixtures are used primarily for assembly purposes. In most cases, cold fixtures are made of hot or cold rolled iron, stainless alloys, nonferrous alloys, or nonmetals, such as phenolics and laminates. These fixtures are used for assembly and subsequent tack welding of parts to be brazed. They need not be massive or heavy, but should be sturdy enough to assemble components as required by design.

Hot fixtures (fixtures used in the furnace for brazing) must have good stability at elevated temperatures and the ability to cool rapidly. Metals can be used but generally are not stable enough to maintain tolerances during the brazing cycle. Therefore, ceramic, carbon, or graphite assemblies often are used for hot fixturing. Ceramics, due to their high processing cost, are used for small fixtures and for spacer blocks to maintain gaps during brazing of small components. Graphite has been found to be the most suitable material for maintaining flatness in a high vacuum or argon atmosphere, and it provides faster cooling because of its high thermal conductivity, which is partially due to its porosity. Graphite should be coated with an Al_2O_3 slurry to prevent carburization of parts during the brazing cycle. It should not be used in a pure dry hydrogen atmosphere as it will cause carburization of the base metals by gaseous transfer. Molybdenum and tungsten may be used, but they are generally avoided because of their cost.

Controlled Atmospheres

Controlled atmospheres (including vacuum) are used to prevent the formation of oxides during brazing and to reduce the oxides present so that the brazing filler metal can wet and flow on clean base metal. Controlled-atmosphere brazing is widely used for the production of high-quality joints. Large tonnages of assemblies of a wide variety of base metals are mass produced by this process.

Controlled atmospheres are not intended to perform the primary cleaning operation for the removal of oxides, coatings, grease, oil, dirt, or other foreign materials from the parts to be brazed. All parts for brazing must be subjected to appropriate prebraze cleaning operations as dictated by the particular metals. Controlled atmospheres commonly are employed in furnace brazing; however, they may also be used with brazing processes that utilize induction, resistance, infrared, laser, and electron-beam heat sources. In applications where a controlled atmosphere is used, postbraze cleaning is generally not necessary. In special cases, flux may be used with a controlled atmosphere (1) to prevent the formation of oxides of titanium and aluminum when brazing in a gaseous atmosphere, (2) to extend the useful life of the flux, and (3) to minimize postbraze cleaning. Fluxes should not be used in a vacuum environment. Some types of equipment, such as metallic muffle furnaces and vacuum systems, may be damaged or contaminated by the use of flux.

The use of controlled atmospheres inhibits the formation of oxides and scale over the whole part and permits finish machining to be done before brazing in many applications. In some applications, such as the manufacture of electronic tubes, eliminating flux is tremendously important.

Pure dry hydrogen is used as a protective atmosphere because it dissociates the oxides of many elements. Hydrogen with a dew point of -51 °C (-60 °F) dissociates the oxides of most elements found in heat-resistant alloys, with the notable exception of aluminum and titanium, which are found in most of the high-strength heat-resistant base metals.

Inert gases, such as helium and argon, do not form compounds with metals. In equipment designed for brazing at ambient pressure, inert gases reduce the evaporation rate of volatile elements, in contrast to brazing in a vacuum. Inert gases permit the use of weaker retorts than required for vacuum brazing. Elements such as zinc and cadmium, however, vaporize in pure dry inert atmospheres.

Vacuum. An increasing amount of brazing of heat-resistant alloys, particularly precipitation-hardenable alloys that contain titanium and aluminum, is done in a vacuum. Vacuum brazing in the range of 13 mPa (10^{-4} torr) has proved adequate for brazing most of the nickel-base superalloys. By removal of gases to a suitably low pressure, including gases that are evolved during heating to brazing temperature, very clean surfaces are obtainable. A vacuum is particularly useful in the manufacture of parts for the aerospace, electronic tube, and nuclear fields, and where metals that react chemically with a hydrogen atmosphere are used or where entrapped fluxes or gases are intolerable. The maximum tolerable pressure for successful brazing depends on a number of factors that are primarily determined by the composition of the base metals, the brazing filler metal, and the gas that remains in the evacuated chamber.

Vacuum brazing is economical for fluxless brazing of many similar and dissimilar basemetal combinations. Vacuums are especially suited for brazing very large, continuous areas where (1) solid or liquid fluxes cannot be removed adequately from the interfaces after brazing, and (2) gaseous atmospheres are not completely efficient because of their inability to purge occluded gases evolved from close-fitting brazing interfaces. It is interesting to note that a vacuum system evacuated to 1.3 mPa (10^{-5} torr) contains only 0.00000132% residual gases based on a starting pressure of 100 kPa (760 torr).

Vacuum brazing has the following advantages and disadvantages compared with brazing that is carried out under other high-purity brazing atmospheres:

- VACUUM REMOVES ESSENTIALLY ALL GASES FROM THE BRAZING AREA, THEREBY ELIMINATING THE NECESSITY FOR PURIFYING THE SUPPLIED ATMOSPHERE. COMMERCIAL VACUUM BRAZING GENERALLY IS DONE AT PRESSURE VARYING FROM

0.0013 TO 13 PA (10^{-5} TO 10^{-1} TORR), DEPENDING ON THE MATERIALS BRAZED, THE FILLER METALS BEING USED, THE AREA OF THE BRAZING INTERFACES, AND THE DEGREE TO WHICH GASES ARE EXPELLED BY THE BASE METALS DURING THE BRAZING CYCLE.

- CERTAIN OXIDES THAT FORM ON BASE METALS (PARTICULARY ON COPPER-BASE METALS) DISSOCIATE IN VACUUM AT BRAZING TEMPERATURES.
- DIFFICULTIES SOMETIMES EXPERIENCED WITH CONTAMINATION OF BRAZING INTERFACES, DUE TO EXPULSION OF GASES FROM THE PARTS TO BE BRAZED, ARE NEGLIGIBLE IN VACUUM BRAZING.
- LOW PRESSURE EXISTING AROUND THE BASE AND FILLER METALS AT ELEVATED TEMPERATURES REMOVES VOLATILE IMPURITIES AND GASES FROM THE METALS. FREQUENTLY, THE PROPERTIES OF BASE METALS ARE IMPROVED. THIS CHARACTERISTIC IS ALSO A DISADVANTAGE WHEN ELEMENTS IN THE FILLER METAL OR BASE METALS VOLATILIZE AT BRAZING TEMPERATURES, THUS CHANGING THE MELTING POINT OF THE FILLER METAL OR PROPERTIES OF THE BASE METAL. THIS TENDENCY MAY, HOWEVER, BE CORRECTED BY EMPLOYING PARTIAL-PRESSURE VACUUM BRAZING TECHNIQUES.

Nickel-Base Alloys

In the selection of a brazing process for a nickel-base alloy, the characteristics of the alloy must be carefully considered. The nickel-base alloy family includes alloys that differ significantly in physical metallurgy (such as the mechanism of strengthening) and in process history (cast versus wrought). These characteristics can have a profound effect on their brazeability.

Precipitation-hardening alloys present several difficulties not normally encountered with solid-solution alloys. Precipitation-hardening alloys often contain appreciable (greater than 1%) quantities of aluminum and titanium for improved strength and corrosion resistance. The oxides of these elements are almost impossible to reduce in a controlled (vacuum or hydrogen) atmosphere. Therefore, nickel plating or the use of a flux is necessary to obtain a surface that allows wetting by the filler metal.

Because these alloys are hardened at temperatures of 540 to 815 °C (1000 to 1500 °F), brazing at or above these temperatures may alter the alloy properties. This frequently occurs when using silver-copper filler metals, which occasionally are used on heat-resistant alloys.

Liquid metal embrittlement is another difficulty encountered in brazing of precipitation-hardening alloys. Many nickel-, iron-, and cobalt-base alloys crack when stressed parts are exposed to molten metals. This is usually confined to the silver-copper filler metals. If precipitation-hardening alloys are brazed in the hardened condition, residual stresses am often high enough to initiate cracking.

Cleanliness, as in all metallurgical joining operations, is important when brazing nickel and nickel-base alloys. Cleanliness of base metal, filler metal, flux (when used for torch or induction brazing), and purity of atmosphere should be as high as practical to achieve the required joint integrity. Elements that cause surface contamination or interfere with braze wetting or flow should be avoided in prebraze processing. All forms of surface contamination such as oils, chemical residues, scale, or other oxide products should be removed by using suitable cleaning procedures. The use of nickel-base filler metals can offer some cost effectiveness in this regard because certain nickel-base brazing filler metals containing boron and/or silicon are known to be self-fluxing and thus more forgiving to slight imperfections in cleanliness.

Attempting to braze over the refractory oxides of titanium and aluminum that may be present on precipitation-hardenable nickel-base alloys must be avoided. Procedures to prevent or inhibit the formation of these oxides before and/or during brazing include special treatments of the surfaces to be joined or brazing in a highly controlled atmosphere. Surface treatments include electrolytic nickel plating and reducing the oxides to metallic form. As stated earlier in this section, a typical practice is to nickel plate the joint surface of any alloy that contains aluminum and/or titanium. For vacuum brazing of alloys containing aluminum and titanium in trace amounts, use of 0.0025 to 0.0075 mm (0.0001 to 0.0003 in.) thick plate is considered optimal. Alloys with up to 4% AI and/or titanium require 0.010 to 0.015 mm (0.0004 to 0.0006 in.) thick plate, and alloys with aluminum and/or titanium contents greater than 4% require 0.020 to 0.030 mm (0.0008 to

0.0012 in.) thick plate. When brazing in a pure dry hydrogen atmosphere, thicker plating (0.025 to 0.038 mm, or 0.001 to 0.0015 in.) is desirable for alloys with high (>4%) aluminum and/or titanium contents.

Atmospheres. Dry, oxygen-free atmospheres that are frequently used include inert gases, reducing gases, and vacuum. The brazing atmosphere, whether gaseous or vacuum, should be free from harmful constituents such as sulfur, oxygen, and water vapor. When brazing in a gaseous atmosphere, monitoring of the water vapor content of the atmosphere (that is, the dew point) is the common practice. A dew point of -50 °C (-60 °F) results in average brazeability; average; -60 °C (-80 °F) or below produces a better quality braze.

Stresses. During brazing, residual or applied tensile stress should be eliminated or minimized as much as possible. Also, inherent stresses present in the precipitation-hardening alloys may lead to stress-corrosion cracking. Therefore, stress relieving or annealing prior to brazing is also recommended for all furnace, induction, or torch brazing operations. Brazing filler metals that melt below the annealing temperature are likely to cause stress-corrosion cracking of the base metal.

Thermal Cycles. Consideration must be given to the effect of the brazing thermal cycle on the base metal. Filler metals that are suitable for brazing nickel-base alloys may require relatively high temperatures. This is particularly true for the filler-metal alloy systems most frequently used in brazing of nickel-base alloys--the nickel-chromium-silicon or nickel-chromium-boron systems.

Solid-solution-strengthened nickel-base alloys such as Inconel 600 may not be adversely affected by brazing temperatures of 1010 to 1230 °C (1850 to 2250 °F). Precipitation-strengthened alloys such as Inconel 718 may, however, display adverse property effects when exposed to brazing cycles with maximum temperatures that are higher than their normal solution heat treatment temperatures. Inconel 718, for example, is solution heat treated at 955 °C (1750 °F) for optimum stress-rupture life and ductility. Braze temperatures of 1010 °C (1850 °F) or above result in grain growth and a corresponding decrease in stress-rupture properties, which cannot be recovered by subsequent heat treatment.

Consideration of base-metal property requirements for service enables selection of an appropriate braze alloy. Lower melting temperature (below 1040 °C, or 1900 °F) braze filler metals are available within the nickel-base alloy family and within other brazing filler-metal systems (see Table 1).

Inconel 718 is often used in the fabrication of air diffusers for aerospace turbine engines. One manufacturer found that vacuum brazing of diffuser components at 10^{-4} torr in a cold-walled vacuum furnace provided the best results. Prior to brazing, all joint surfaces were nickel plated to 0.005 to 0.015 mm (0.0002 to 0.0006 in.) thicknesses. Plating was done in accordance with specification AMS 2424 or equivalent. Prior to assembly, application of BNi-2 braze filler metal tape (approximately 0.11 mm, or 0.0045 in., thick) was preplaced between all joint surfaces. After assembly, an additional braze slurry of BNi-2 filler metal was applied to all joints to ensure joint soundness.

Brazing of ODS Alloys

Oxide dispersion-strengthened alloys are P/-M alloys that contain stable oxides evenly distributed throughout the matrix. The oxide does not go into solution in the alloy even at the liquidus temperature of the matrix. However, the oxide is usually rejected from the matrix upon melting of the matrix, which occurs during fusion welding, and cannot be redistributed in the matrix on solidification; therefore, these alloys are usually joined by brazing. There are two commercial classes of ODS alloys: the dispersion-strengthened nickel, and mechanically alloyed Inconel MA 754, Inconel MA 6000, and Incoloy MA 956.

Inconel MA 754, dispersion-strengthened nickel alloys, and dispersion-strengthened nickel-chromium alloys are the easiest to braze of the family of ODS alloys. Vacuum, hydrogen, or inert atmospheres can be used for brazing. Prebraze cleaning consists of grinding or machining the faying surfaces and washing with a solvent that evaporates without leaving a residue. Generally, brazing temperatures should not exceed 1315 °C (2400 °F) unless demanded by a specific application that has been well examined and tested. The brazing filler metals for use with these ODS alloys usually are not classified by AWS. In most cases, the brazing filler metals used with these alloys have brazing temperatures in excess of 1230 °C (2250 °F). These include proprietary alloys that are based on nickel, cobalt, gold, or palladium.

Brazements made of ODS alloys to be used at elevated temperature must be tested at elevated temperature to prove fitness-for-purpose. In the case of stress-rupture testing, AWS specification C3.2 may be used as a guide because it gives the actual joint configuration. The configuration shown in Fig. 4 is preferred by some as a test model, although any test configuration without stress raisers is adequate. The elevated-temperature brazement properties for Inconel alloy MA 754 should meet the following requirements:

SHEAR STRESS		**TEMPERATURE**		SERVICE LIFE, H
MPA	**KSI**	**°C**	**°F**	
26	3.8	980	1800	>**1000**
9.0	**1.3**	**1095**	**2000**	>**1000**

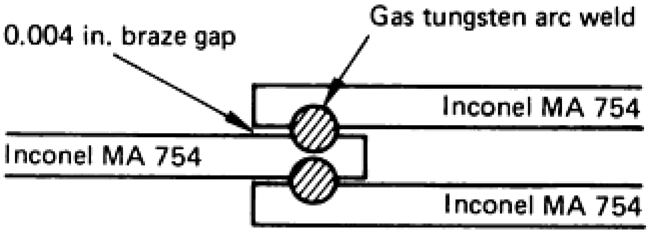

FIG. 4 BRAZING SPECIMEN ASSEMBLY USED TO EVALUATE STRESS-RUPTURE PROPERTIES OF BRAZED INCONEL MA 754

Inconel MA 6000 is a nickel-base ODS alloy that is also γ' strengthened. The amount of alloying elements in the base metal matrix plus the γ' precipitations cause an interesting problem in the brazing of this alloy. Inconel MA 6000 has a solidus temperature of 1300 °C (2372 °F); therefore, the brazing temperature should be no higher than 1250 °C (2282 °F). Additionally, because 1230 °C (2250 °F) is the γ' solution treatment temperature, it becomes important to carefully select the brazing filler metal and to heat treat the assembly after brazing. The AWS-specified BNi and BCo filler metals, as well as specially formulated filler metals, have been used for this alloy.

Inconel MA 6000 is used for its high-temperature strength and corrosion resistance; unfortunately, the passive oxide scale that provides good corrosion resistance also prevents wetting and flow of brazing filler metal. Therefore, correct cleaning procedures are very important. Surfaces to be brazed should be mechanically cleaned with a water-cooled, low-speed belt or wheel of approximately 320-grit and stored in a solvent, such as methanol, until immediately before the beginning of the brazing cycle.

Cobalt-Base Alloys

The brazing of cobalt-base alloys is readily accomplished with the same techniques used for nickel-base alloys. Because most of the popular cobalt-base alloys do not contain substantial amounts of aluminum or titanium, brazing atmosphere requirements are less stringent. Table 2 gives typical compositions of several cobalt-base alloys. These materials can be brazed in either a hydrogen atmosphere or a vacuum. Filler metals are usually nickel- or cobalt-base alloys or gold-palladium compositions. Silver- or copper-base brazing filler metals may not have sufficient strength and oxidation resistance in many high-temperature applications. Although cobalt-base alloys do not contain appreciable amounts of aluminum or titanium, an electroplate or flash of nickel is often used to promote better wetting of the brazing filler metal.

TABLE 2 TYPICAL COMPOSITIONS OF SELECTED COBALT-BASE ALLOYS

ALLOY(A)	**NOMINAL COMPOSITION, WT%**															CHARACTERISTICS AND TYPICAL APPLICATIONS
	C	**Mn**	**Si**	**Cr**	**Ni**	**Co**	**Mo**	**W**	**Nb**	**Ti**	**Al**	**B**	**Zr**	**Fe**	**OTHER**	
AIRESIST 13(B)	0.45	0.5(C)	. . .	21	1.0(C)	BAL	. . .	11	2.0	. . .	3.5	. . .	. . .	2.5(C)	0.1 Y	**HIGH-TEMPERATURE PARTS**
AIRESIST 213	0.18	. . .	. . .	19	. . .	BAL	. . .	4.7	. . .	. . .	3.5	. . .	0.15	. . .	6.5 TA, 0.1 Y	**SHEETS, TUBING; RESISTANT TO HOT CORROSION**
AIRESIST 215(B)	0.35	. . .	. . .	19	. . .	BAL	. . .	4.5	. . .	. . .	4.3	. . .	0.13	. . .	7.5 TA, 0.17 Y	**NOZZLE VANES; RESISTANT TO HOT CORROSION**
ELGILOY	0.15	2.0	. . .	20.0	15.0	40.0	7.0	. . .	. . .	. . .	. . .	. . .	. . .	BAL	0.04 BE	**SPRINGS; CORROSION RESISTANT, HIGH STRENGTH**
FSX-414(B)	0.25	1.0(C)	1.0(C)	29.5	10.5	BAL	. . .	7.0	. . .	. . .	. . .	0.012	. . .	2.0(C)	. . .	**GAS TURBINE VANES**
FSX-418(B)	0.25	1.0(C)	1.0(C)	29.5	10.5	BAL	. . .	7.0	. . .	. . .	. . .	0.012	. . .	2.0(C)	0.15 Y	**GAS TURBINE VANES; IMPROVED OXIDATION RESISTANCE**
FSX-430(B)	0.40	. . .	. . .	29.5	10.0	BAL	. . .	7.5	. . .	. . .	. . .	0.027	0.9	. . .	0.5 Y	**GAS TURBINE VANES; IMPROVED STRENGTH AND DUCTILITY**
X-40(B)	0.50	0.50	0.50	25	10	BAL	. . .	7.5	. . .	. . .	. . .	. . .	. . .	1.5	. . .	**GAS TURBINE PARTS, NOZZLE VANES**
HAYNES 150	0.08	0.65	0.35	28	3.0(B)	BAL	1.5(B)	. . .	. . .	. . .	. . .	. . .	. . .	20.0	. . .	**RESISTS THERMAL SHOCK, HIGH-TEMPERATURE CORROSION (AIR AND AIR-SO_2)**
HAYNES 188 (SHEET)	0.10	1.25(B)	0.3	22	22	BAL	. . .	14	. . .	. . .	. . .	. . .	. . .	3.0(B)	0.04 LA	**BETTER OXIDATION RESISTANCE THAN HASTELLOY X**

MAR-M302(B)	0.85	0.10	0.20	21.5	. . .	BAL	. . .	10.0	. . .	. . .	. . .	0.005	0.15	. . .	9.0 TA	**JET ENGINE BLADES, VANES**
MAR-M322(B)	1.00	0.10	0.10	21.5	. . .	BAL	. . .	9.0	. . .	0.75	. . .	. . .	2.25	. . .	4.5 TA	**JET ENGINE BLADES, VANES**
MAR-M509(B)	0.60	0.10(C)	0.10(C)	21.5	10	BAL	. . .	7.0	. . .	0.2	. . .	0.010(C)	0.50	1.0	3.5 TA	**JET ENGINE BLADES, VANES**
MAR-M918	0.05	0.2(C)	0.2(C)	20	20	BAL	. . .	. . .	. . .	. . .	. . .	. . .	0.16	0.5(C)	7.5 TA	**HIGH-TEMPERATURE SHEET**
MP35N	. . .	. . .	. . .	20.0	35.0	35.0	10.0	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	**STRESS-CORROSION RESISTANT, HIGH-STRENGTH FASTENERS**
NASA CO-W-RE(B)	0.40	. . .	. . .	3	. . .	BAL	. . .	25	. . .	1.0	. . .	. . .	1.0	. . .	2.0 RE	**HIGH-TEMPERATURE SPACE APPLICATIONS**
S-816	0.38	1.20	0.40	20	20	BAL	4.0	4.0	4.0	. . .	. . .	. . .	. . .	4	. . .	**GAS TURBINE BLADES, BOLTS, SPRINGS**
V-36	0.27	1.00	0.40	25	20	BAL	4.0	2.0	2.0	. . .	. . .	. . .	. . .	3	. . .	**HIGH-TEMPERATURE SHEET**
WF-11, L605, HAYNES 25	0.10	1.50	0.50	20	10	BAL	. . .	15	. . .	. . .	. . .	. . .	. . .	3.0(B)	. . .	**JET ENGINE PARTS, SHEET**
WF-31	0.15	1.42	0.42	20	L0	BAL	2.6	10.7	. . .	1.0	. . .	. . .	. . .	. . .	. . .	**HIGH-TEMPERATURE SHEETS**
WI-52(B)	0.45	0.50(C)	0.50(C)	21	1.0(C)	BAL	. . .	11	2.0	. . .	. . .	. . .	. . .	2.0	. . .	**GAS TURBINE PARTS, NOZZLE VANES**
X-45(B)	**0.25**	**1.0(C)**	**. . .**	**25.5**	**10.5**	**BAL**	**. . .**	**7.0**	**. . .**	**. . .**	**. . .**	**0.010**	**. . .**	**2.0(C)**	**. . .**	**NOZZLE VANES**

(A) SOME SUPERALLOYS ARE MADE BY MORE THAN ONE MANUFACTURER. THE PROPRIETARY DESIGNATIONS FOR SUCH ALLOYS HAS BEEN USED IN THIS COMPILATION.

(B) CAST ALLOY.

(C) MAXIMUM COMPOSITION

Nickel-base brazing alloys such as AWS BNi-3 have been used successfully on Haynes 25 for honeycomb structures. After brazing, a diffusion cycle was reportedly used to raise the braze joint remelt temperature to 1260 to 1315 °C (2300 to 2400 °F). Table 3 presents the effects of a high-temperature braze (1225 °C, or 2240 °F, for 15 rain) on the mechanical properties of Haynes 25. One cobalt-base brazing filler metal (AWS BCo-1, see Table 1) appears to offer a good combination of strength, oxidation resistance, and remelt temperature for use on Haynes 25 foil.

TABLE 3 EFFECT OF BRAZING ON MECHANICAL PROPERTIES OF HAYNES 25

CONDITION	**TEST TEMPERATURE(A)**		**ULTIMATE STRENGTH**		**0.2% OFFSET YIELD STRENGTH**		ELONGATION, %
	°C	**°F**	**MPA**	**KSI**	**MPA**	**KSI**	
TENSILE TESTING							
MILL ANNEAL	RT	RT	1020	147.9	477	69.2	**56**
AFTER BRAZE CYCLE	RT	RT	747	108.3	477	69.2	**12**
MILL ANNEAL	815	1500	396	57.4	210	30.5	**17**
AFTER BRAZE CYCLE	815	1500	393	57.0	229	33.2	**24**
MILL ANNEAL	980	1800	145	21.0	125	18.1	**35**
AFTER BRAZE CYCLE	**980**	**1800**	**148**	**21.5**	**130**	**18.8**	**35**

CONDITION	**TEST TEMPERATURE**		**STRESS**		TIME TO RUPTURE, H
	°C	**°F**	**MPA**	**KSI**	
STRESS-RUPTURE TESTING					
MILL ANNEAL	815	1500	169	24.5	**82**
AFTER BRAZE CYCLE	815	1500	169	24.5	**72.3**
MILL ANNEAL	900	1650	103	15.0	**56**
AFTER BRAZE CYCLE	900	1650	103	15.0	**36**
MILL ANNEAL	980	1800	45	6.5	**110**
AFTER BRAZE CYCLE	**980**	**1800**	**45**	**6.5**	**120**

(A) RT, ROOM TEMPERATURE

Cobalt-base alloys, much like nickel-base alloys, can be subject to liquid metal embrittlement or stress-corrosion cracking when brazed under residual or dynamic stresses. This frequently is observed when using silver- or silver-copper-base (BAg) filler metals. Liquid metal embrittlement of cobalt-base alloys by copper-base (BCu) filler metals occurs with or without the application of stress; therefore, BCu filler metals should be avoided when brazing cobaltbase alloys.

Brazing of Heat-Resistant Alloys, Low-Alloy Steels, and Tool Steels

Revised by Darrell Manente, Vac-Aero International Inc.

Brazing of Low-Alloy Steels and Tool Steels

Low-alloy steels constitute a category of ferrous materials that exhibit mechanical properties superior to plain carbon steels as the result of additions of such alloying elements as manganese, nickel, chromium, and molybdenum. Total alloy content can range from 2.07% up to levels just below that of stainless steels, which contain a minimum of 10% Cr. For many low-alloy steels, the primary function of the alloying elements is to increase hardenability in order to optimize mechanical properties and toughness after heat treatment. In some cases, however, alloy additions are used to reduce environmental degradation under certain specific service conditions. For compositions of specific low-alloy steels, see the article "Classification and Designation of Carbon and Low-Alloy Steels" in Volume 1 of the *ASM Handbook.*

Tool steels are any steels used to make tools for cutting, forming, or otherwise shaping a material into a part or component adapted to a definite use. Tool steels could be classified as low-alloy steels, but they are generally considered to be a separate group. Table 4 lists selected tool steels and their typical uses.

TABLE 4 PRINCIPAL TYPES OF TOOL STEELS

AISI/SAE DESIGNATION	**DESCRIPTION**	TYPICAL USES
W1, W2	WATER-HARDENING GRADES	**COLD-HEADING DIES, WOODWORKING TOOLS**
S1, S5	SHOCK-RESISTING TOOL STEELS	**CHISELS, HAMMERS, RIVET SETS**
O1, O2	OIL-HARDENING COLD-WORK TOOL STEELS	**SHORT-RUN COLD-FORMING DIES, CUTTING TOOLS**
A2, A5	AIR-HARDENING MEDIUM-ALLOY COLD-WORK TOOL STEELS	**THREAD ROLLING AND SLITTING DIES, INTRICATE DIE SHAPES**
D2, D3, D4	HIGH-CARBON HIGH-CHROMIUM COLD-WORK STEELS	**USES UNDER 480 °C (900 °F), GAGES, LONG-RUN FORMING AND BLAKING DIES**
H12, H13, H16	CHROMIUM HOT-WORK STEELS	**ALUMINUM OR MAGNESIUM EXTRUSION DIES, DIE-CASTING DIES, MANDRELS, HOT SHEARS, FORGING DIES**
H21, H23	TUNGSTEN HOT-WORK STEELS	**HOT EXTRUSION DIES FOR BRASS, NICKEL. AND STEEL; HOT-FORGING DIES**
T1, T15(A)	TUNGSTEN HIGH-SPEED STEELS	**ORIGINAL HIGH-SPEED CUTTING STEEL**
M1. M2. M3, M10, M15(A)	**MOLYBDENUM HIGH-SPEED STEELS**	**85% OF ALL CUTTING TOOLS IN UNITED STATES MADE FROM THIS GROUP**

(A) MOST WEAR-RESISTANT GRADE

Brazing Filler Metals

The AWS BAg, BCu, and BNi brazing filler metals can be used effectively to braze low-alloy steels or tool steels. The specific type of brazing filler metal used depends on the application, the available brazing equipment, and the strength requirements of the braze joint. Many BAg brazing filler metals are used for torch or induction brazing. BCu and BNi brazing filler metals are used for furnace brazing. High-solidus-point brazing filler metals can be chosen so that a component can be brazed and then heat treated. For example, a plate assembly made from 4340 material can be brazed with copper filler metal at 1120 °C (2050 °F) for 10 min. The plates are then kept at a temperature of 845 °C (1550 °F) for 1 h in order to austenitize the base metal and subsequently quenched in oil. The plates are tempered at 510 °C (950 °F) for 2 h in order to produce a hardness of about 40 HRC. The copper braze filler is not affected by the heat treatment due to its relatively high melting point.

Precleaning

The surface condition of the material will determine the cleaning method that is chosen. For example, a newly machined component would normally require only a vapor degreasing operation or a solvent wipe in order to remove machining lubricants. A component with a light oxide layer on the surface might be grit blasted with Al_2O_3 and then vapor degreased. This component could also be cleaned with emery paper or a stainless steel wire brush and wiped with a solvent.

A component with a heavy oxide layer on the surface would require a combination of mechanical and chemical cleaning in order to ensure that the oxide layer was removed. Another option available for removing oxides is to furnace clean components in an atmosphere of vacuum or hydrogen. While this method will reduce oxides of iron and nickel, other oxides may not be reduced (Fig. 4).

Fixturing

Fixturing methods for heat-resistant alloys are also suitable for low-alloy steels and tool steels. However, greater use can be made of metal fixturing because many of these brazements are made at lower temperatures.

Atmospheres and Fluxes

Low-alloy steels and tool steels can be brazed in air, vacuum, hydrogen, nitrogen or endothermic or exothermic atmospheres. Components can also be torch or induction brazed in air using fluxes containing fluorides, borates, or chlorides.

Components brazed with AWS BNi brazing filler metals are usually brazed in a vacuum. Vacuum nitrogen, or hydrogen atmospheres are generally used for brazing components with BCu or BAg brazing filler metals. Care must be taken to ensure that the chosen atmosphere does not cause carburization or decarburization of the component.

Brazing of Copper, Copper Alloys, and Precious Metals

Roy E. Beal, Amalgamated Technologies, Inc. Rodger E. Cook, The Wilkinson Company

Introduction

COPPER, COPPER ALLOYS, AND PRECIOUS METALS are probably the most easily brazed metals available. A wide range of brazing filler metals is used to join the many different coppers, copper alloys, and precious metals that are manufactured. The selection of the brazing process and the filler metal depends on the alloy or material composition, the shape and dimension of the parts to be joined, and the intended application. Gold, silver, and the platinum group metals are well known for their resistance to oxidation at high temperatures. Not only does this characteristic make the precious metals desirable in industrial and medical applications, it helps enhance the brazeability of components made from them.

Brazing of Copper, Copper Alloys, and Precious Metals

Roy E. Beal, Amalgamated Technologies, Inc. Rodger E. Cook, The Wilkinson Company

Copper and Copper Alloys

Because the brazeability of most copper alloys is very good, the material considerations are generally not as difficult as those of some other metals. However, the specific metallurgy of the individual copper or copper alloy is an important factor and should be considered when selecting a manufacturing method. Most brazing operations will result in recrystallization of the copper alloy being joined. Fine grain sizes can be eliminated and the cold-working step removed, which may or may not be desirable. Often, electrical conductivity is an important factor in a copper alloy brazed joint when overall resistance across the joint must be controlled. Therefore, the finished metallurgical structures, grain sizes, and mechanical properties of the specific copper alloy must be considered when brazing is utilized.

The metallurgical structures of brazed joints are largely those of the brazing filler metal, enriched with copper from the parent material. Because some surface melting of the copper alloys being joined occurs, the interfaces usually show primary dendrites of a copper-rich phase growing from the copper alloy and braze metal interface. The brazing filler metal becomes more copper-rich, producing additional dendritic growth with increased brazing temperature and time.

The widely used oxygen-containing coppers must be carefully handled when heated. Brazing temperatures are normally above 480 °C (900 °F), which allows the dispersed copper oxides in the material to react with a hydrogen atmosphere, creating a high-pressure steam within the solid metal. Tough-pitch coppers cannot be satisfactorily brazed by furnace techniques that utilize hydrogen as the atmosphere, either in the form of exothermic or endothermic gases or in the form of dissociated ammonia. Because the torch brazing of tough-pitch coppers also can cause problems with embrittlement, it generally should be avoided.

Copper metals that have been deoxidized by phosphorus or other elements and oxygen-free coppers can be readily brazed in hydrogen-containing atmospheres with no embrittlement. Brazing can also be accomplished without flux, if a copper-phosphorus brazing filler metal is used. When other copper materials, such as those containing lead, tellurium, silver, and other elements, are brazed, a copper-phosphorus filler metal can be used without flux, although flux is sometimes recommended.

The brazing characteristics of most copper filler metals and alloys are improved when a flux is used. In this case, phosphorus-deoxidized and oxygen-free coppers can be joined with a brazing filler metal from the silver series, which may contain zinc, cadmium, or nickel.

High-copper-content or copper alloys that contain metals that are more refractory, such as chromium or zirconium, will form oxide films that slow or prevent the flow of the brazing filler metal. Under these conditions, cleaning operations are often needed to remove the surface oxides prior to brazing. The brazing operation is then carried out in either a protective atmosphere or one that has essentially reducing conditions. Fluxes can be used to assist in the surface oxide removal or to prevent their reformation during the brazing process.

Some of the high-copper-content materials, such as chromium copper, are effectively hardened by the presence of the chromium element. In these cases, it is important to heat treat the material. This can sometimes be carried out in conjunction with the brazing operation when an appropriate brazing filler metal is selected.

Brass represents the largest group of copper alloys. Brass compositions usually contain from 5 to 40% zinc. The higher zinc contents require the use of fluxes in most brazing applications, especially when the zinc level is above 15%. Other elements are added to generate special properties, such as when lead is added to improve machining. This addition makes brazing very difficult. At a lead level of 8%, brazing is essentially impossible. Tin is added to produce a range of materials, including admiralty and naval brass. Modifications may contain other elements, such as manganese and nickel. Brasses with some aluminum content are more difficult to join than the other brasses mentioned and require specially made fluxes.

The next-largest group is bronze alloys, the most important of which are phosphor-bronze, silicon-bronze, and aluminum-bronze. Although phosphor-bronze alloys can be joined by most methods, this material has a weakness for hot cracking in cold-worked conditions. Therefore, parts should be stress relieved at temperatures from 290 to 345 °C (550 to 650 °F) before brazing. Similar problems, such as hot shortness, can be encountered with silicon-bronze alloys. Aluminum-bronze or silicon-bronze alloys that contain aluminum require special attention, because of the tenacity of the surface oxides. Generally, a special American Welding Society (AWS) type 4 flux is used in conjunction with low-melting-point, high-silver-percentage filler metals.

Copper-nickel alloys are widely used for corrosion protection and for durability in marine environments. Filler metals that contain phosphorus are not acceptable because of brittle nickel-phosphide formation. Most copper alloys can be joined to other coppers and copper alloys. Compatibility is important and can require compromise to suit the individual characteristics of the two metals involved. For example, when copper alloys that contain aluminum need to be joined to any other copper material, the AWS type FB4-A flux is required for good results.

Thus, the first step in the successful brazing of copper and copper alloys is to understand the metallurgy of the alloy involved and the effect of alloy additions on the properties of the surface to be brazed. The finished joints between the same or dissimilar metals must have a sufficiently matched composition in order to avoid joint brittleness, adverse diffusion, and large nonuniform phases. Ductility values also must be satisfactory.

Filler Metals

More filler metals are available for copper and copper alloys than for any other metal system. They vary widely in composition and cost. Standard alloys are listed by AWS reference code in Table 1. Many of the alloys contain high percentages of silver and copper, because these elements form a eutectic at a 77% silver level. Because of the high cost of silver, alloys containing phosphorus, zinc, cadmium, and other elemental additions were developed.

TABLE 1 COMPOSITIONS AND SELECTED PROPERTIES OF FILLER METALS USED IN BRAZING OF COPPER AND COPPER ALLOYS

AWS FILLER METAL DESIGNATION	**NOMINAL COMPOSITION, %**							**SOLIDUS TEMPERATURE**		**LIQUIDUS TEMPERATURE**		**CONDUCTIVITY(A), %IACS**	TYPICAL DIAMETRAL JOINT CLEARANCE	
	Ag	**Cu**	**P**	**Zn**	**Cd**	**Ni**	**OTHER**	**°C**	**°F**	**°C**	**°F**		**MM**	IN.
RBCUZN-A	. . .	59.25	. . .	40	. . .	. . .	0.75 SN	890	1630	900	1650	26	0.05-0.13	**0.002-0.005**
BCUZN-D	. . .	48	. . .	42	. . .	10	. . .	920	1690	935	1715	. . .	. . .	**. . .**
BCUP-1	. . .	95	5	. . .	. . .	. . .	. . .	710	1310	900	1650	. . .	0.05-0.13	**0.002-0.005**
BCUP-2	. . .	92.75	7.25	. . .	. . .	. . .	. . .	710	1310	795	1460	. . .	0.025-0.075	**0.011-0.003**
BCUP-4	6	86.75	7.25	. . .	. . .	. . .	. . .	645	1190	725	1335	. . .	0.025-0.075	**0.001-0.003**
BCUP-5	15	80	5	. . .	. . .	. . .	. . .	645	1190	800	1475	10	0.025-0.13	**0.001-0.005**
BAG-1	45	15	. . .	16	24	. . .	. . .	605	1125	620	1145	28	0.05-0.13	**0.002-0.005**
BAG-1A	50	15.5	. . .	16.5	18	. . .	. . .	625	1160	635	1175	24	0.05-0.13	**0.002-0.005**
BAG-2	35	26	. . .	21	18	. . .	. . .	605	1125	700	1295	29	0.05-0.13	**0.002-0.005**
BAG-3	50	15.5	. . .	15.5	16	3	. . .	630	1170	690	1270	18	0.05-0.13	**0.002-0.005**
BAG-5	45	30	. . .	25	. . .	. . .	. . .	675	1250	745	1370	19	0.05-0.13	**0.002-0,005**
(B)	75	22	. . .	3	. . .	. . .	. . .	. . .	. . .	740	1365	. . .	0.05-0.13	**0.002-0.005**
BAG-8	77	23	. . .	. . .	. . .	. . .	. . .	780	1435	780	1435	. . .	0.05-0.13	**0.002-0.005**
BAG-SA	72	27.8	. . .	. . .	. . .	. . .	0.2 LI	765	1410	765	1410	89(C)	0.05-0.13	**0.002-0.005**
BAG-19	92.5	7.3	. . .	. . .	. . .	. . .	0.2 LI	780	1435	890	1635	88(C)	0.05-0.03	**0.002-0.005**
BAU-4	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**18**	**82 AU**	**950**	**1740**	**950**	**1740**	**6**	**0.05-0.13**	**0.002-0.005**

(A) RATIO OF THE RESISTIVITY OF THE MATERIAL AT 20 °C (68 °F) TO THAT OF IACS, EXPRESSED AS A PERCENTAGE AND CALCULATED ON A VOLUME BASIS.

(B) SPECIAL FILLER METAL USED IN BRAZING NICKEL SILVER KNIFE HANDLES.

(C) CONDUCTIVITY OF FILLER METAL AFTER VOLATILIZATION OF LITHIUM IN BRAZING

The most widely used silver alloy is BAg-1, which has a solidus temperature of 607 °C (1125 °F) and a liquidus temperature of 618 °C (1145 °F). It is generally recognized as being the lowest-melting-point silver brazing alloy.

Other widely used filler metals are the RB-CuZn, BCuP, and BAu series. The copper-zinc (RBCuZn) filler metals have relatively high liquidus temperatures and are anodic to copper, leading to lower corrosion resistance. Zinc-containing filler metals sometimes lose this element or form voids if overheating is allowed to occur.

The copper-phosphorus (BCuP) filler metals are self-fluxing when used to braze copper. Fluxes may be necessary when joining other copper alloys. The high-phosphorus filler metals are not resistant to sulfur-containing atmospheres. Copper-phosphorus and copper-silver-phosphorus brazing filler metals are recommended only for copper and copper alloys without nickel, because brittle phosphide can form when nickel is present.

Silver alloy (BAg) filler metals can join all brazeable coppers and copper alloys, as well as many dissimilar combinations. The lower-melting-point alloys are good for joining brass. There is a trend toward avoiding filler metals that contain cadmium, because of its toxicity. Cadmium-bearing filler metals must not be used for food equipment. Lithium-containing filler metals that are joined in protective atmospheres are self-fluxing.

Gold (BAu) alloys are high-cost filler metals that are used in specialized applications, such as vacuum and electronic products. Although these filler metals can join copper alloys, they are generally not used because of their cost. The main application is in situations where low vapor pressure is important.

Brazing Fluxes

Four primary categories of fluxes are used to braze coppers and copper alloys. There are no standard compositions and fluxes are classified by type only.

Type FB3-A is the general-purpose, low-temperature flux used with all coppers and copper alloys, except those with high levels of aluminum.

Types FB3-C and FB3-D have higher active temperature ranges, respectively, up to 1205 °C (2200 °F). These fluxes can be used with any filler metal that is appropriate from Table 1. Other fluxes are classified by liquid, powder, and slurry forms, according to application.

Type FB4-A flux is specifically designed for aluminum-containing copper alloys and has a wide range of activity from 595 to 879 °C (1100 to 1600 °F).

The AWS brazing flux classifications for coppers and copper alloys are given in Table 2. Major ingredients, filler metal type for which the flux is suitable, activity temperature range, application process, and recommended base metals are also given in Table 2.

TABLE 2 BRAZING FLUXES

CLASSIFICATION	**FORM**	**FILLER METAL TYPE**	**TYPICAL INGREDIENTS**	**APPLICATION**	**ACTIVITY TEMPERATURE RANGE**		RECOMMENDED BASE METALS
					°C	**°F**	
FB3-A	PASTE	BAG AND BCUP	BORATES, FLUORIDES	GENERAL-PURPOSE FLUX FOR MOST FERROUS AND NONFERROUS ALLOYS. (NOTABLE EXCEPTION ALUMINUM BRONZE, ETC. SEE FLUX 4-A)	565-870	1050-1600	**ALL BRAZEABLE FERROUS AND NONFERROUS METAL, EXCEPT THOSE WITH ALUMINUM OR MAGNESIUM AS A CONSTITUENT. ALSO USED TO BRAZE CARBIDES**
FB3-C	PASTE	BAG AND BCUP	BORATES, FLUORIDES, BORON	SIMILAR TO 3-A, BUT WITH CAPABILITY FOR EXTENDED HEATING TIMES OR TEMPERATURE THROUGH USE OF A DEOXIDIZING ADDITIVE	565-925	1050-1700	**ALL BRAZEABLE FERROUS AND NONFERROUS METAL, EXCEPT THOSE WITH ALUMINUM OR MAGNESIUM AS A CONSTITUENT. ALSO USED TO BRAZE CARBIDES**
FB3-D	PASTE	BAG, BCU, BNI, BAU, AND RBCUZN	BORATES, FLUORIDES	SIMILAR TO 3-C, BUT WITH A HIGHER ACTIVE-TEMPERATURE RANGE	760-1205	1400-2200	**ALL BRAZEABLE FERROUS AND NONFERROUS METAL, EXCEPT THOSE WITH ALUMINUM OR MAGNESIUM AS A CONSTITUENT. ALSO USED TO BRAZE CARBIDES**
FB3-E	LIQUID	BAG AND BCUP	BORATES, FLUORIDES	LOW-ACTIVITY LIQUID FLUX USED IN BRAZING JEWELRY OR TO AUGMENT FURNACE BRAZING ATMOSPHERES	565-870	1050-1600	**ALL BRAZEABLE FERROUS AND NONFERROUS METAL, EXCEPT THOSE WITH ALUMINUM OR MAGNESIUM AS A CONSTITUENT. ALSO USED TO BRAZE CARBIDES**
FB3-F	POWDER	BAG AND BCUP	BORATES, FLUORIDES	SIMILAR TO 3-A IN A POWDER FORM	650-870	1200-1600	**ALL BRAZEABLE FERROUS AND NONFERROUS METAL, EXCEPT THOSE WITH ALUMINUM OR MAGNESIUM AS A CONSTITUENT. ALSO USED TO BRAZE CARBIDES**
FB3-G	SLURRY	BAG AND BCUP	BORATES, FLUORIDES	SIMILAR TO 3-A IN A SLURRY FORM	565-870	1050-1600	**ALL BRAZEABLE FERROUS AND NONFERROUS METAL, EXCEPT THOSE WITH ALUMINUM OR MAGNESIUM AS A CONSTITUENT. ALSO USED TO BRAZE CARBIDES**
FB3-H	SLURRY	BAG AND BCUP	BORATES, FLUORIDES, BORON	SIMILAR TO 3-C IN A SLURRY FORM	565-925	1050-1700	**ALL BRAZEABLE FERROUS AND NONFERROUS METAL, EXCEPT THOSE WITH ALUMINUM OR MAGNESIUM AS A CONSTITUENT. ALSO USED TO BRAZE CARBIDES**
FB3-I	SLURRY	BAG, BCU,	BORATES,	SIMILAR TO 3-D IN A	760-	1400-	**ALL BRAZEABLE FERROUS AND**

		BNI, BAU, AND RBCUZN	FLUORIDES	SLURRY FORM	1205	2200	**NONFERROUS METAL, EXCEPT THOSE WITH ALUMINUM OR MAGNESIUM AS A CONSTITUENT. ALSO USED TO BRAZE CARBIDES**
FB3-J	POWDER	BAG, BCU, BNI, BAU, AND RBCUZN	BORATES, FLUORIDES	SIMILAR TO 3-D IN A SLURRY FORM	760-1205	1400-2200	**ALL BRAZEABLE FERROUS AND NONFERROUS METAL, EXCEPT THOSE WITH ALUMINUM OR MAGNESIUM AS A CONSTITUENT. ALSO USED TO BRAZE CARBIDES**
FB3-K	LIQUID	BAG, BCUP, AND RBCUZN	BORATES	EXCLUSIVELY USED IN TORCH BRAZING BY PASSING FUEL GAS THROUGH A CONTAINER OF FLUX. FLUX APPLIED BY THE FLAME	760-1205	1400-2200	**ALL BRAZEABLE FERROUS AND NONFERROUS METAL, EXCEPT THOSE WITH ALUMINUM OR MAGNESIUM AS A CONSTITUENT. ALSO USED TO BRAZE CARBIDES**
FB4-A	**PASTE**	**BAG AND BCUP**	**CHLORIDES, FLUORIDES, BORATES**	**GENERAL-PURPOSE FLUX FOR MANY ALLOYS CONTAINING METALS THAT FORM REFRACTORY OXIDES**	**595-870**	**1100-1600**	**BRAZEABLE BASE METALS CONTAINING UP TO 9% AL (ALUMINUM BRASS, ALUMINUM BRONZE, MONEL K500). MAY ALSO HAVE APPLICATION WHEN MINOR AMOUNTS OF TITANIUM OR OTHER METALS ARE PRESENT,WHICH FORM REFRACTORY OXIDES**

Note: The selection of a flux designation for a specific type of work may be based on the filler metal type and the description above, but the information here is generally not adequate for flux selection.

The highest-temperature fluxes are types FB3-I, FB3-J, and FB3-K, which work at temperatures from 760 to 1205 °C (1400 to 2200 °F). These are generally used with the RBCuZn filler metals, with long and active conditions at temperature, although they can be used with all other brazing filler metals for copper alloys.

Joint Clearance and Design

Joint clearances are primarily related to the brazing filler metals being used and their capillary attraction. If the gap is too small, then the brazing filler metal will not flow into the joint. Typical joint clearances are identified in Table 1.

Clearance can also influence the mechanical strength of the brazed joints. Allowances should be made in joint design for relative expansion of the parts at the brazing temperature, especially where dissimilar metals are involved. Penetration of the brazing filler metal through a joint to the opposite side of placement can be an effective measure of brazing filler metal flow, and a preferable joint design allows this observation to be made. Generally, braze joints should be designed for an effective joint area of 80%, in order to allow for voids.

In the case of a lap joint, the joint length should be, at a minimum, three times the thickness of the thinnest component being joined.

Brazing Processes

The selection of the brazing process is as important as the selection of the brazing filler metal. In fact, one often depends on the other. When a specific process is required, the range of allowable filler metals may be restricted. The number of parts to be made and the production rates desired also influence brazing process selection. Another consideration is the materials to be joined, because brazing may affect their mechanical properties. All parts for brazing must be cleaned prior to the brazing operation.

Brazing can be carried out with a flux or in an atmosphere. Several different atmospheres are appropriate for copper and copper-alloy brazing. AWS atmosphere types, as well as approximate compositions, filler metals, and braze metal combinations, are described in Table 3. Note that some base metals require a low dew point, in addition to the protective atmospheres.

TABLE 3 ATMOSPHERES FOR BRAZING COPPER AND COPPER ALLOYS

AWS BRAZING ATMOSPHERE TYPE NO.	**SOURCE**	**MAXIMUM DEW POINT INCOMING GAS**	**COMPOSITION OF ATMOSPHERE, %**				**FILLER METALS**	**BASE METALS**	REMARKS
			H_2	**N_2**	**CO**	**CO_2**			
1	COMBUSTED FUEL GAS (LOW HYDROGEN)	ROOM TEMPERATURE	5-1	87	5-1	11-12	BAG(A), BCUP, RBCUZN(A)	COPPER, BRASS(A)	**REFERRED TO COMMONLY AS EXOTHERMIC GENERATED ATMOSPHERES**
2	COMBUSTED FUEL GAS (DECARBURIZING)	ROOM TEMPERATURE	14-15	70-71	9-10	5-6	BCU, BAG(A), RBCUZN, BCUP	COPPER(B), BRASS(A), LOW NICKEL, MONEL, MEDIUM-CARBON STEEL(C)	**DECARBURIZES REFERRED TO COMMONLY AS ENDOTHERMIC GENERATED ATMOSPHERES**
3	COMBUSTED FUEL GAS,DRIED	-40 °C (-40 °F)	15-16	73-75	10-11		SAME AS 2	SAME AS 2 PLUS MEDIUM- AND HIGH-CARBON STEELS, MONEL, NICKEL ALLOYS	**REFERRED TO COMMONLY AS ENDOTHERMIC GENERATED ATMOSPHERES**
4	COMBUSTED FUEL GAS, DRIED (DECARBURIZING)	-40 °C (-40 °F)	38-40	41-45	17-19		SAME AS 2	SAME AS 2 PLUS MEDIUM- AND HIGH-CARBON STEELS	**CARBURIZES**
5	DISSOCIATED AMMONIA	-54 °C (-65 °F)	75	25	. . .	. . .	BAG(A), BCUP, RBCUZN(A), BCU, BNI	SAME AS FOR 1, 2, 3, 4 PLUS ALLOYS CONTAINING CHROMIUM	
6A	CRYOGENIC OR PURIFIED N_2 + H_2	-68 °C (-90 °F)	1-30	70-99	. . .	. . .	SAME AS 5	SAME AS 3	
6B	CRYOGENIC OR PURIFIED N_2 + H_2 + CO	-29 °C (-20 °F)	2-20	70-99	1-10	. . .	SAME AS 5	SAME AS 4	

6C	CRYOGENIC OR PURIFIED N_2	-68 °C (-90 °F)	. . .	100	. . .	. . .	SAME AS 5	SAME AS 3	
7	DEOXYGENATED AND DRIED HYDROGEN	-59 °C (-75 °F)	100	. . .	. . .	. . .	SAME AS 5	SAME AS 5 PLUS COBALT, CHROMIUM, TUNGSTEN ALLOYS AND CARBIDES(D)	
8	HEATED VOLATILE MATERIALS	INORGANIC VAPORS(ZINC, VOLATILE FLUORIDES)	. . .	. . .	. . .	. . .	BAG	BRASSES	**SPECIAL PURPOSE. MAY BE USED IN CONJUNCTION WITH 1 THROUGH 5 TO AVOID USE OF FLUX**
9	PURIFIED INERT GAS	INERT GAS (HELIUM,ARGON, ETC.)	. . .	. . .	. . .	. . .	SAME AS 5	SAME AS 5 PLUS TITANIUM,ZIRCONIUM, HAFNIUM	**SPECIAL PURPOSE. PARTS MUST BE VERY CLEAN AND ATMOSPHERE MUST BE PURE**
9A	**PURIFIED INERT GAS + H_2**	**INERT GAS (HELIUM, ARGON, ETC.)**	**1-10**	**. . .**	**. . .**	**. . .**			

(A) FLUX REQUIRED IN ADDITION TO ATMOSPHERE WHEN ALLOYS CONTAINING VOLATILE COMPONENTS ARE USED.

(B) COPPER SHOULD BE FULLY DEOXIDIZED OR OXYGEN-FREE.

(C) HEATING TIME SHOULD BE MINIMIZED TO AVOID OBJECTIONABLE DECARBURIZATION.

(D) FLUX MUST BE USED IN ADDITION TO THE ATMOSPHERE IF APPRECIABLE QUANTITIES OF ALUMINUM, TITANIUM, SILICON. OR BERYLLIUM ARE PRESENT.

Furnace brazing has, as its primary advantage, the ability to handle mass production. However, a small furnace can be used for a wide variety of parts, materials, and filler metals. Furnaces can operate on either a batch or continuous basis. Furnaces can be mesh belt conveyor, batch, or pusher (where parts are loaded on trays) systems. The objective is to achieve minimum-cost operation for the particular application. Another advantage is that a number of joints can be brazed simultaneously in a furnace.

Copper and copper alloys are often furnace brazed on a conveyor or mesh belt system. This allows the use of protective atmospheres that can be either exothermic or endothermic. The conveyor furnace also allows for controlled heating and cooling rates. Separate belts are sometimes used for each zone to control these rates. The metallurgy of the alloys to be joined must be considered.

Brazing furnaces can be expensive, especially for mass-production operations with protective atmospheres, as well as preheating and cooling zones. Although electrical furnaces cost more to run, gas furnaces usually require muffles or linings that must be periodically replaced, which raises the overall expense and narrows the cost differential.

The brazing furnace must be designed to handle the desired temperature and mass load. Furnace linings are used when fluxes are necessary, even in electrically heated units, because flux can corrode the heating elements. Parts that can be self-fluxed and joined in a protective atmosphere avoid this potential flux problem.

Copper and copper alloys must not be exposed to gases containing sulfur. Leaks into the lining from gas firing can be a problem. Brasses are not generally furnace brazed.

The operation of a brazing furnace depends on the filler metal, the melting temperature, and the characteristics of the copper alloys to be joined. It is usually desirable to braze at approximately 28 °C (50 °F) higher than the liquidus of the filler metal. Most brazing filler metals can be used, except those that have large amounts of zinc and are subject to volatilization.

A dry hydrogen atmosphere can be used for some brazing filler-metal and base-metal combinations. However, as noted earlier, the hydrogen atmosphere must be completely avoided when oxygen-containing copper metal is used, because it would embrittle the copper.

An example of a furnace braze application without a protective atmosphere is the contact arm assembly shown in Fig. 1, which was joined with BCuP-5 filler metal.

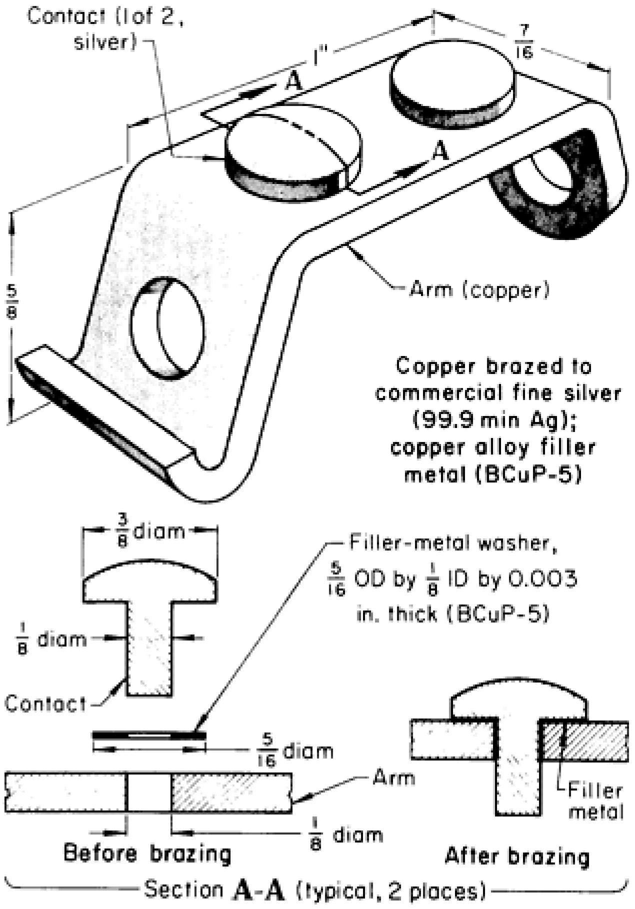
Contact (1 of 2, silver)
1"
7/16
A
A
5/8
Arm (copper)
Copper brazed to commercial fine silver (99.9 min Ag); copper alloy filler metal (BCuP-5)
3/8 diam
1/8 diam
Contact
Filler-metal washer, 5/16 OD by 1/8 ID by 0.003 in. thick (BCuP-5)
5/16 diam
Arm
1/8 diam
Filler metal
Before brazing
After brazing
Section A-A (typical, 2 places)

FIG. 1 FURNACE BRAZED CONTACT ARM ASSEMBLY, BRAZED WITHOUT PROTECTIVE ATMOSPHERE, 42 °C (75 °F) BELOW THE LIQUIDUS OF THE FILLER METAL

Components to be furnace brazed are preassembled with the filler metal preplaced for brazing. Preferably, the parts should be self-jigging. External jigs and fixtures can be used when absolutely necessary, but add extra capital and operating expense, because the parts have to be assembled in the jigs and the jigs themselves have a finite life. Fully enclosed parts must be vented to permit the escape of trapped air. Failure to do so can be dangerous.

Torch brazing of copper and copper alloys is used on many components and in numerous industrial applications. Examples are radiators, condensers, refrigerators, construction piping, motors, electrical equipment, and others. Equipment costs are low. The braze operator must be skilled in adjusting the heating flame and directing energy to the part to be joined. Torches can be extremely small, for electronic applications, or very large, for heavy fabrications and castings. Production capacity varies from one part to mass-production lots.

Tough-pitch coppers require reducing atmospheres because of hydrogen embrittlement. Brass can lose zinc by volatilization and must not be overheated or held too long at brazing temperature. Alloys with refractory elements require flux and reducing flames. Some copper alloys, such as phosphor bronze, crack if heated too rapidly.

Filler metals for torch brazing were described in Table 1. The filler material can be in wire, powder, or paste form. The application can be manual or automatic.

Examples of torch brazing applications are return bends in heat exchangers, which are usually automatically torch brazed; fin assemblies, which are usually manually torch brazed; instrumentation tubes; oil coolers; and large electrical parts.

A common example is a return bend joint 0made with alloy C12000, which is a phosphorus-deoxidized copper. A 9.5 mm ($\frac{3}{8}$ in.) diameter tube with a 0.5 mm (0.02 in.) wall thickness is typical. Tube ends of the heat exchanger are expanded to allow insertion of the return bend. A clearance of 0.10 to 0.23 mm (0.004 to 0.009 in.) is used to provide a pressure-tight joint. Each leg of the rerum bend is fitted with a 0.75 mm (0.03 in.) diameter BCuP-5 filler wire. The joint is self-fluxing and multiple joints are passed through a series of torch flames to quickly make the joint in approximately 5 s.

Another example is an eight-station turntable for the automatic torch brazing of instrument assemblies (Fig. 2). These components are usually small and are joined at multiple stations by indexing the table. This system uses unskilled operators and can achieve high production rates. The brazing filler metal can be in flux-paste form or a wire feeding unit can be used in conjunction with separate fluxing.

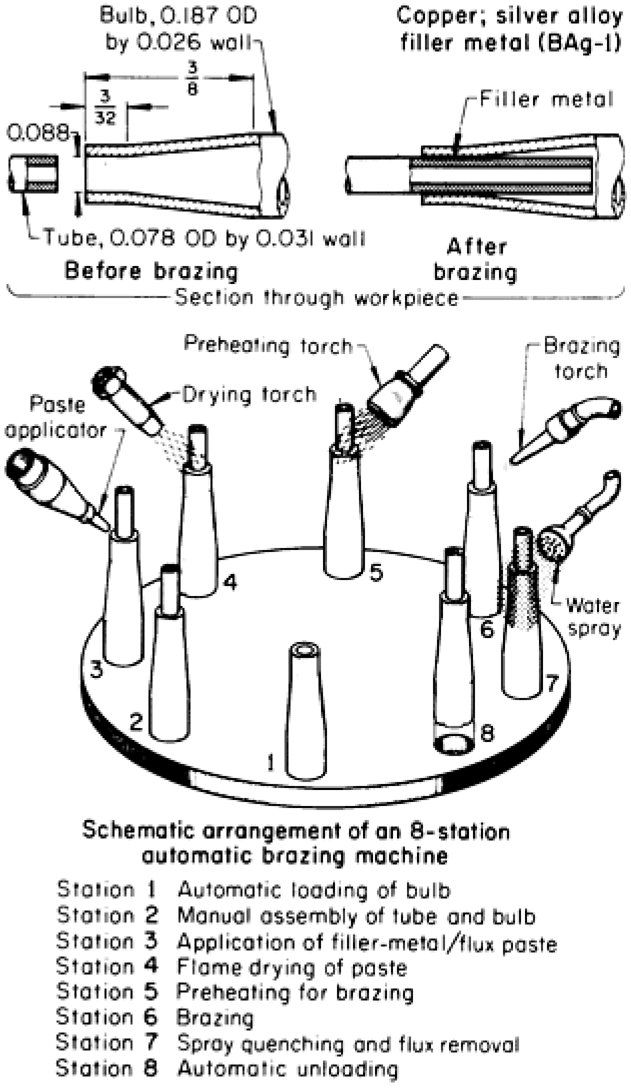

Bulb, 0.187 OD by 0.026 wall
Copper; silver alloy filler metal (BAg-1)
3/8
3/32
0.088
Filler metal
Tube, 0.078 OD by 0.031 wall
Before brazing
After brazing
Section through workpiece
Preheating torch
Brazing torch
Drying torch
Paste applicator
Water spray
1
2
3
4
5
6
7
8
Schematic arrangement of an 8-station automatic brazing machine
Station 1 Automatic loading of bulb
Station 2 Manual assembly of tube and bulb
Station 3 Application of filler-metal/flux paste
Station 4 Flame drying of paste
Station 5 Preheating for brazing
Station 6 Brazing
Station 7 Spray quenching and flux removal
Station 8 Automatic unloading

FIG. 2 BULB AND TUBE ASSEMBLY, BRAZED BY MECHANICALLY HELD TORCHES ON AN EIGHT-STATION TURNTABLE

Induction brazing is less efficient with copper alloys than with other metals. However, this process is widely used because of its high-production capability, especially in situations where flames cannot be tolerated and atmosphere protection is an alternative to fluxing. Induction brazing usually minimizes the possible warping of parts. However, the initial costs of the induction heating equipment is high. The low efficiency of the process also demands relatively high power requirements.

Inductor design is similar to that needed for other metals. The objective is controlled heat input and distribution. Single- or multiple-turn inductors are utilized, sometimes in conjunction with concentrators, to focus heat in desirable locations.

Brazing filler metals for induction heating are usually the silver-containing alloys and the copper-phosphorus and copper-phosphorus-silver alloys.

An advantage of induction brazing is exemplified by the chromium-copper (C18200) post shown in Fig. 3. The heating step in the brazing process is a critical factor in obtaining optimum mechanical properties, because precipitation hardening can occur with improper heating. The brazing of a tungsten-silver contact required the use of a nickel-containing silver alloy, BAg3, to join with chromium-copper. Induction brazing is the best option for retaining material properties.

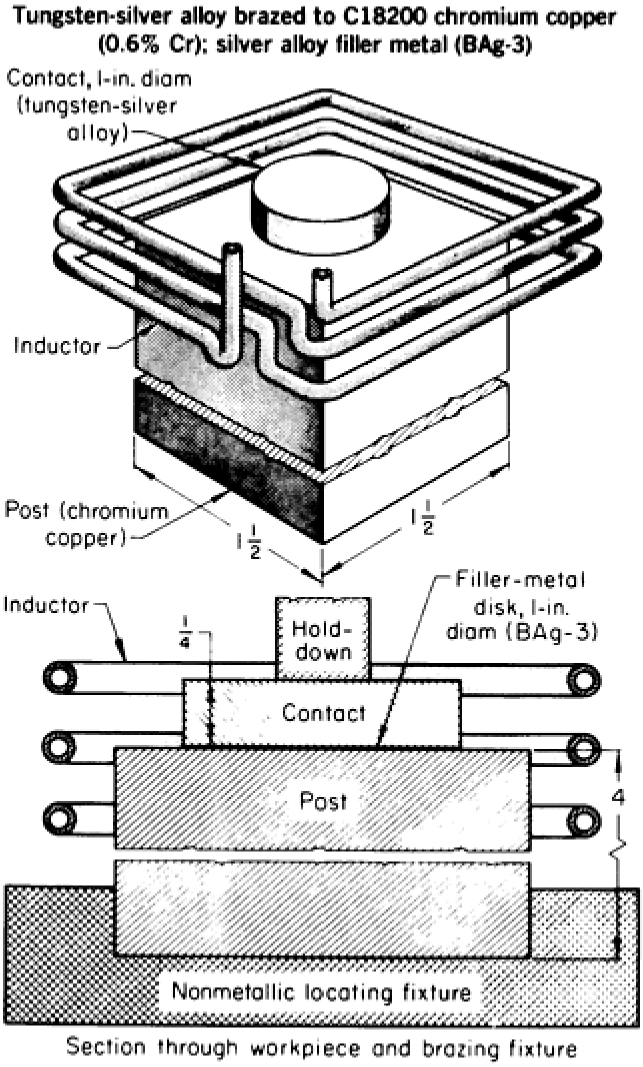
Tungsten-silver alloy brazed to C18200 chromium copper (0.6% Cr); silver alloy filler metal (BAg-3)
Contact, 1-in. diam (tungsten-silver alloy)
Inductor
Post (chromium copper)
$1\frac{1}{2}$
$1\frac{1}{2}$
Inductor
$\frac{1}{4}$
Hold-down
Filler-metal disk, 1-in. diam (BAg-3)
Contact
Post
4
Nonmetallic locating fixture
Section through workpiece and brazing fixture

FIG. 3 TUNGSTEN-SILVER ALLOY CONTACT IN POSITION FOR INDUCTION BRAZING TO C18200 CHROMIUM-COPPER (0.6% CR) POST; BAG-3 SILVER ALLOY FILLER METAL

Another example is the use of multiple-station inductors in the brazing of armatures for motors. A preformed BCuP-5 filler metal is placed at the desired joint area and the inductor is allowed to bring the assembly to temperature quickly without flux.

Resistance brazing is used for small parts, such as terminals, fasteners, and attachments. The filler metal, which is usually preplaced, is often BCuP-5, because it can be used without flux. However, most brazing filler metals that contain silver have been used. Heat generation occurs by resistance through the assembly, using a standard resistance welding machine. A second heating method is to use graphite blocks as the contact, which spreads the heat over a larger area. This method is widely used in the electronics industry and for joining high-conductivity coppers and alloys. Brazing times are very short at 10 to 20 Hz with very high current inputs. A small quantity of hydrogen can be used for oxidation protection, except when joining tough-pitch coppers.

Portable resistance welding machines are used in the manufacture of electrical equipment, transformers, bus bars, and other components. Large areas can be joined by making a series of spot brazes.

Salt-bath dip brazing of copper and copper alloys is a less-popular brazing process. The procedure uses a molten neutral salt at the brazing temperature. The brazing filler metal is preplaced and the molten salt is used to activate joining. The procedure is useful for unusually shaped parts or those with multiple joints, such as heat exchangers. Applications include waveguides, which allow simultaneous brazing of all joints in a single immersion. A cast bronze alloy (C22000), for example, can be joined to a thin-walled tube to close tolerances without warpage. Self-jigging of parts is preferable during the assembly process, before the parts are mounted into the external jig used in the dip braze. External jigs must allow for expansion of the parts during brazing. A disadvantage of this process is the need to remove all residual salts after brazing.

Brazing of Copper, Copper Alloys, and Precious Metals

Roy E. Beal, Amalgamated Technologies, Inc. Rodger E. Cook, The Wilkinson Company

Precious Metals

Throughout history, precious metals have been used as adornments and jewelry. Even today, the use of gold in jewelry is the major source of demand for fabricated gold. Brazing is routinely employed in the manufacture of jewelry to produce joints that match in color and surface condition to the base materials. Biocompatibility and relative ease of workability have also historically placed precious metals in the fields of dentistry and medicine.

Like copper, precious metals exhibit good thermal and electrical conductivity, in addition to superior corrosion resistance--required traits for materials used for high-reliability electrical contacts.

Material Composition

Jewelry golds are usually yellow or white in color. The term "karat" refers to the fineness or percentage of gold in the alloy, with 24 karat (k) being pure gold. Diluting gold to 18 k would then mean that the resulting alloy is 75% Au (18 parts gold, plus six parts of other metals). Yellow golds are generally alloys containing gold, copper, silver, and zinc. White golds are lightened by the substitution of nickel for some of the copper in the alloy.

Dental materials may contain any of the precious metals in almost any combination. It is not unusual to find a dental casting material that is composed of gold, silver, platinum, palladium, iridium, and/or other precious or nonprecious metals to achieve special performance characteristics (for example, matched coefficient of thermal expansion with dental porcelain).

Precious-metal-base electrical and electronic materials are found in wide variety. Many of the original precious metal contact materials were derived from existing dental and jewelry alloys. Silver graphite, silver nickel, and silver metal oxides (for example, Ag-CdO) are common switch and relay materials fabricated by powder metallurgy or the oxidation of a wrought alloy. In many cases, these silver contact materials are provided with a silver backing to facilitate brazing. Parts that are not silver backed must be treated to provide a suitable surface for brazing. Silver graphite contacts can be air fired to remove surface carbon. Silver metal oxide parts can be chemically treated to remove surface nonmetallics. Additional surface treatment, plating, or coining can then be employed. Another very common silver contact alloy is coin silver (90Ag-10Cu).

Sterling silver (92.SAg-7.5Cu) is well-known for its application in jewelry, flatware, and decorative household serving utensils.

The platinum-group metals are platinum, palladium, iridium, rhodium, ruthenium, and osmium. With the exception of osmium, pure and alloyed platinum-group metals have a wide range of applications including jewelry, thermocouples, glass melting vessels and tools, labware, medical devices, and catalysts.

Precious metal plated or clad materials can be brazed, but they require process controls sufficient to prevent excessive diffusion or blistering of the plating or cladding.

Filler Metals

The fields of jewelry and dentistry each have selections of brazing filler metals that satisfy the specific demands of their disciplines. An extremely wide variety of gold jewelry filler metals is available. One important consideration in the selection of filler metals for gold jewelry is that the filler metal does not reduce the gold content of the finished brazement. Table 4 gives a partial list of gold filler metals for jewelry applications.

TABLE 4 SELECTED FILLER METALS USED IN GOLD JEWELRY BRAZING APPLICATIONS

COMPOSITION, %							**SOLIDUS**		LIQUIDUS	
Au	**Ag**	**Cu**	**Zn**	**Cd**	**Sn**	**Ni**	**°C**	**°F**	**°C**	°F
80.0(A)	. . .	. . .	8.0	. . .	. . .	12.0	782	1440	871	**1600**
75.0	12.0	8.0	. . .	5.0	. . .	. . .	826	1519	887	**1629**
75.0	9.0	6.0	. . .	10.0	. . .	. . .	776	1429	843	**1549**
75.0	9.0	6.0	10.0	. . .	. . .	. . .	730	1346	783	**1441**
75.0	2.8	11.2	9.0	2.0	. . .	. . .	747	1377	788	**1450**
75.0	. . .	15.0	1.8	8.2	. . .	. . .	793	1459	822	**1512**
66.6(A)	10.0	6.4	12.0	. . .	. . .	5.0	718	1324	810	**1490**
66.6	15.0	15.0	3.4	. . .	. . .	. . .	796	1465	826	**1519**
58.5	25.0	12.5	. . .	4.0	. . .	. . .	788	1450	840	**1544**
58.5	10.3	24.2	. . .	7.0	. . .	. . .	792	1458	831	**1528**
58.5	8.8	22.7	. . .	10.0	. . .	. . .	751	1384	780	**1436**
58.5	11.8	25.7	4.0	. . .	. . .	. . .	816	1501	854	**1569**
58.5	25.7	11.8	4.0	. . .	. . .	. . .	786	1447	818	**1504**
58.5	24.2	10.3	7.0	. . .	. . .	. . .	765	1409	808	**1486**
58.5	4.9	25.6	2.0	9.0	. . .	. . .	738	1360	760	**1400**
58.5	8.0	22.0	2.1	9.4	. . .	. . .	744	1371	776	**1429**
58.3	20.8	19.0	1.9	. . .	. . .	. . .	793	1459	830	**1526**
58.3	18.0	12.0	11.7	. . .	. . .	. . .	720	1328	754	**1389**
58.3(A)	15.0	5.7	15.0	. . .	. . .	6.0	. . .	. . .	. . .	**. . .**
50.0(A)	25.0	10.0	9.0	. . .	. . .	6.0	. . .	. . .	. . .	**. . .**
50.0	30.5	17.5	2.0	. . .	. . .	. . .	775	1427	806	**1483**
41.7	32.0	16.3	10.0	. . .	. . .	. . .	724	1335	749	**1380**
41.7(B)	24.0	16.3	9.0	9.0	. . .	. . .	. . .	. . .	. . .	**. . .**
41.7	35.0	21.9	1.4	. . .	. . .	. . .	. . .	. . .	. . .	**. . .**
33.3	30.0	16.7	. . .	0.0	. . .	. . .	635	1175	709	**1303**
33.3	30.0	16.7	20.0	. . .	. . .	. . .	695	1283	704	**1299**
33.3	40.5	17.0	6.6	2.6	. . .	. . .	722	1332	749	**1380**
33.3	1.8	49.4	2.3	10.2	3.0	. . .	689	1272	776	**1429**
33.3	31.0	28.0	7.7	. . .	. . .	. . .	737	1359	808	**1486**
33.3(A)	42.0	10.0	9.7	. . .	. . .	5.0	738	1360	807	**1485**
25.0	35.0	20.0	10.0	10.0	. . .	. . .	. . .	. . .	. . .	**. . .**
25.0(A)	**58.0**	**. . .**	**17.0**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**

Source: Ref 1

(A)	WHITE GOLD.
(B)	GREEN GOLD.

Platinum jewelry can be brazed with binary alloys of gold-platinum, platinum-silver, or palladium-silver, or multiple precious metal alloys (some of which contain copper).

Silver jewelry is generally brazed with silver-base filler metals that contain copper and zinc (and in some cases, include additions of tin, nickel, or cadmium). However, the elimination of cadmium is becoming more prevalent due to stringent Occupational Safety and Health Administration (OSHA) regulations.

There are some proprietary variations in the compositions of brazing filler metals for dentistry, although they are usually composed of gold, silver, copper, zinc, and tin. Table 5 gives typical dental gold filler metals with physical properties.

TABLE 5 COMPOSITION AND PHYSICAL PROPERTIES OF SELECT GOLD-BASE BRAZING FILLER METALS USED IN DENTAL APPLICATIONS

BRAZING DESIGNATION	**COMPOSITION, %**					**MELTING RANGE**		**HEAT TREATMENT**	**PROPORTIONAL LIMIT**		**TENSILE STRENGTH**		LONGATION, %
	AU	**AG**	**CU**	**ZN**	**SN**	**°C**	**°F**		**MPA**	**KSI**	**MPA**	**KSI**	
A	65.4	15.4	12.4	3.9	3.1	745-785	1375-1445	SOFTENED	186	27	293	42.5	**14**
								HARDENED	379	55	434	63	**1**
B	66.1	12.4	16.4	3.4	2.0	750-805	1385-1480	SOFTENED	203	29.5	307	44.5	**12**
								HARDENED	534	77.5	576	83.5	**<1**
C	65.0	16.3	13.1	3.9	1.7	765-800	1410-1470	SOFTENED	207	30	303	44	**9**
								HARDENED	531	77	634	92	< **1**
D	72.9	12.1	10.0	3.0	2.0	755-835	1390-1535	SOFTENED	165	24	248	36	**7**
								HARDENED	424	61.5	483	70	< **1**
E(A)	**80.9**	**8.1**	**6.8**	**2.1**	**2.0**	**745-870**	**1375-1595**	**SOFTENED**	**141**	**20.5**	**259**	**37.5**	**18**

Source: Ref 2

(A) NOT APPRECIABLY AFFECTED BY AGE HARDENING.

Filler metals for industrial applications of precious metals usually come from the American Welding Society (AWS) BAu and BAg groups. Pure gold and gold-base alloys can be brazed with materials from the BAu or BAg metals. Pure silver and silver-base alloys can be brazed with the BAg and, in some cases, the BCuP (see Fig. 1) filler metals. Platinum and palladium are brazeable with pure gold (which melts at 1063 °C, or 1945 °F) as a filler or by using one of the platinum jewelry filler metals.

Fluxes and Atmospheres

Generally, fluxes and atmospheres suitable for copper brazing will be sufficient for applications joining precious metals because precious metals are oxidation resistant. The user can choose a flux or atmosphere that corresponds to the type of filler metal used, or in the case of dissimilar metal brazing to precious metals, a flux or atmosphere that will protect the nonprecious base material.

Joint Clearance and Configuration

The brazing of precious metals may be achieved by following the criteria established for joint clearance and design for copper brazing (see the section "Copper and Copper Alloys" in this article).

Brazing Processes

Precious metals can be joined using furnace brazing, torch brazing, induction brazing, and resistance brazing processes.

Furnace Brazing. In the production of crown and bridge work, a dental laboratory technician may have to join a part that has been partially porcelain coated. The technician may in that case employ the use of a vacuum porcelain firing furnace to perform the brazing operation on the exposed metal surfaces. High-volume industrial applications (for example, electrical contact manufacture) may call for the use of furnace brazing.

Torch Brazing. In the dental laboratory and at the jeweler's bench, the most widely used process for brazing is manual torch brazing. Dental lab technicians frequently torch braze one-of-a-kind configurations that have been fixtured by embedding the parts into high-temperature-resistant plaster. Automated torch brazing has been applied to high-volume jewelry applications, such as chainmaking. Industrial precious metal users employ both manual and automated torch brazing.

Induction Brazing. Precious metal parts brazed to dissimilar metals (for example, the brazing of a contact to a holder), sometimes require the localized heating attainable by induction heating (Fig. 3). The same type of efficiency deficit encountered when induction brazing is used to braze copper is also present when the process is applied to precious metals (see the section "Copper and Copper Alloys" in this article).

Resistance Brazing. Because copper and precious metals are highly conductive, the process of resistance brazing of precious metals is applicable, but also limited by short times and high current outputs.

References cited in this section

1. *PRECIOUS METALS SCIENCE AND TECHNOLOGY,* INTERNATIONAL PRECIOUS METALS INSTITUTE, 1991, P 653
2. R.W. PHILLIPS, *SKINNER'S SCIENCE OF DENTAL MATERIALS,* 8TH ED., W.B. SAUNDERS COMPANY, 1982

Roy E. Beal, Amalgamated Technologies, Inc. Rodger E. Cook, The Wilkinson Company

---

References

1. *PRECIOUS METALS SCIENCE AND TECHNOLOGY,* INTERNATIONAL PRECIOUS METALS INSTITUTE, 1991, P 653
2. R.W. PHILLIPS, *SKINNER'S SCIENCE OF DENTAL MATERIALS,* 8TH ED., W.B. SAUNDERS COMPANY, 1982

Brazing of Aluminum Alloys

E.B. Gempler

Introduction

ALUMINUM is a commonly used base material for brazing. It is easily fabricated by most manufacturing methods, such as machining, forming, and stamping. Although most brazing processes are applicable, the use of aluminum requires more attention. It is different from most metals in two ways: it quickly reforms its tenacious oxide coating whenever a reactive form of oxygen is present, such as air or water, and it is always brazed at temperatures close to its melting point. The latter factor compromises strength (in terms of holding shape during brazing) and tends to increase the rapidity in which the brazing alloy reacts with the base metal. The net result is that cycle times are critical and fixturing forces must be controlled.

Brazing of Aluminum Alloys

E.B. Gempler

Base Metals

The non-heat-treatable wrought alloys typically used as base metals are listed in Table 1. Those that are most successfully brazed are the l*xxx* and 3*xxx* series and the low-magnesium members of the 5*xxx* series.

TABLE 1 MELTING RANGES AND BRAZEABILITY OF SOME COMMON ALUMINUM ALLOYS

ALLOY	**MELTING RANGE**		BRAZEABILITY[(A)]
	°C	**°F**	
NON-HEAT-TREATABLE WROUGHT ALLOYS			
1350	646-657	1195-1215	**A**
1100	643-657	1190-1215	**A**
3003[(B)]	643-654	1190-1210	**A**
3004	629-652	1165-1205	**B**
5005	632-652	1170-1205	**B**
5050	627-652	1160-1205	**B**
5052	593-649	1100-1200	**C**
HEAT-TREATABLE WROUGHT ALLOYS			
6053	593-652	1100-1205	**A**
6061	593-649	1100-1200	**A**
6063	616-654	1140-1210	**A**
6951[(C)]	616-654	1140-1210	**A**
7005	607-649	1125-1200	**B**
CASTING ALLOYS[(D)]			
443.0	574-632	1065-1170	**B**
356.0	557-613	1035-1135	**B**

710.0	596-646	1105-1195	**B**
711.0	**604-643**	**1120-1190**	**A**

(A) A, GENERALLY BRAZEABLE BY ALL COMMERCIAL PROCEDURES; B, BRAZEABLE WITH SPECIAL TECHNIQUES OR IN SPECIFIC APPLICATIONS THAT JUSTIFY PRELIMINARY TRIALS OR TESTING TO DEVELOP THE PROCEDURE AND TO CHECK THE PERFORMANCE OF BRAZED JOINTS; C, LIMITED BRAZEABILITY.
(B) USED BOTH PLAIN AND AS THE CORE OF BRAZING SHEET.
(C) USED ONLY AS THE CORE OF BRAZING SHEET.
(D) SAND AND PERMANENT MOLD CASTINGS ONLY

The commonly brazed heat-treatable wrought alloys are the 6*xxx* series. Because the 2*xxx* and 7*xxx* series of aluminum alloys have melting points that are too low, they are not normally brazeable. Exceptions are the 7072 alloy, which is used for cladding material only, and 7005 alloy. Alloys with a solidus temperature above 595 °C (1100 °F) are readily brazed using the aluminum-silicon filler metals.

There are several brazeable cast aluminum alloys, which are listed in Table 1 along with their melting range and degree of brazeability. The most readily brazeable are those with the higher solidus temperatures, such as the 710 and 711 alloys.

Brazing of Aluminum Alloys

E.B. Gempler

Filler Metals

Alloys used in flux brazing usually contain between 7 and 12% silicon-balanced aluminum, and tramp metals are controlled to levels below 0.2%.

Alloys employed in fluxless brazing use higher percentages of silicon (>9%) and have varying additions of magnesium to enhance oxide film modification to promote wetting, as well as reduce the partial pressures of oxygen beating gases in the chamber. These alloys are primarily found in clad form. Some are semiproprietary to processes used in the automotive heat-exchanger industry. The vacuum (fluxless) alloys, BAlSi-6 through -11, are identified in Table 2. The alloys BAlSi-3 through -5 also can be used with the fluxless process if modifications related to magnesium are made, either in the base metal or as an addition to the furnace.

TABLE 2 COMPOSITIONS AND SOLIDUS, LIQUIDUS, AND BRAZING TEMPERATURE RANGES OF BRAZING FILLER METALS FOR USE ON ALUMINUM ALLOYS

AWS CLASSIFICATION	**COMPOSITION(A), %**						TEMPERATURE					
							SOLIDUS		**LIQUIDUS**		BRAZING	
	Si	**Cu**	**Mg**	**Zn**	**Mn**	**Fe**	**°C**	**°F**	**°C**	**°F**	**°C**	°F
BALSI-2	6.8-8.2	0.25	. . .	0.20	0.10	0.8	577	1070	613	1135	599-621	**1110-1150**
BALSI-3(B)	9.3-10.7	3.3-4.7	0.15	0.20	0.15	0.8	521	970	585	1085	571-604	**1060-1120**
BALSI-4	11.0-13.0	0.30	0.10	0.20	0.15	0.8	577	1070	582	1080	582-604	**1080-1120**
BALSI-5(C)	9.0-11.0	0.30	0.05	0.10	0.05	0.8	577	1070	591	1095	588-604	**1090-1120**
BALSI-6(D)	6.8-8.2	0.25	2.0-3.0	0.20	0.10	0.8	559	1038	607	1125	599-621	**1110-1150**

BALSI-7$^{(D)}$	9.0-11.0	0.25	1.0-2.0	0.20	0.10	0.8	559	1038	596	1105	588-604	**1090-1120**
BALSI-8$^{(D)}$	11.0-13.0	0.25	1.0-2.0	0.20	0.10	0.8	559	1038	579	1075	582-604	**1080-1120**
BALSI-9$^{(D)}$	11.0-13.0	0.25	0.10-0.5	0.20	0.10	0.8	562	1044	582	1080	582-604	**1080-1120**
BALSI-10$^{(D)}$	10.0-12.0	0.25	2.0-3.0	0.20	0.10	0.8	559	1038	582	1080	582-604	**1080-1200**
BALSI-11$^{(D)(E)}$	**9.0-11.0**	**0.25**	**1.0-2.0**	**0.20**	**0.10**	**0.8**	**559**	**1038**	**596**	**1105**	**582-604**	**1080-1120**

(A) PRINCIPAL ALLOYING ELEMENTS.
(B) CONTAINS 0.15% CR.
(C) CONTAINS 0.20% TI.
(D) SOLIDUS AND LIQUIDUS TEMPERATURE RANGES VARY WHEN USED IN VACUUM.
(E) CONTAINS 0.02-0.20% BI

The ranges of brazing temperatures of both aluminum alloy base and filler metals are shown in Fig. 1.

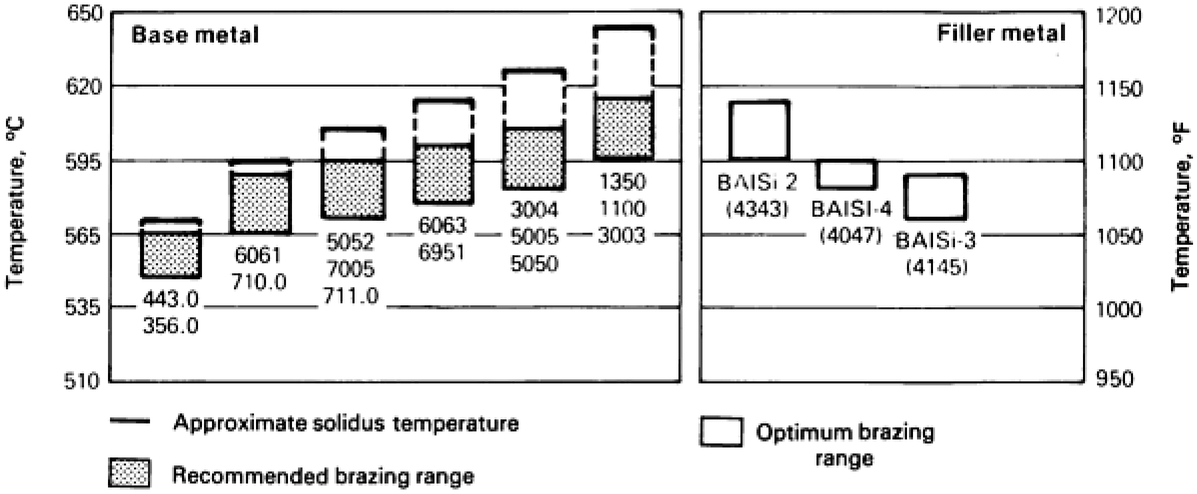

FIG. 1 COMPARISON OF BRAZING TEMPERATURE RANGES OF ALUMINUM ALLOY BASE METALS AND ALUMINUM ALLOY BRAZING FILLER METALS

Non-clad forms of brazing alloys are usually shim or wire stock and are limited to AA4047. Some paste forms are also available from powdered alloy and flux or from powdered alloy and binder for either flux or fluxless processes. These materials are available in the United States from only a few vendors.

Brazing of Aluminum Alloys

E.B. Gempler

Brazing Sheet

Brazing sheet is usually made by roll bonding the filler metal to the base metal. It can be single clad (on one side only) or double clad, and is an extremely useful way to apply filler metal.

Brazing sheet is advantageously used in the assembly of high-production items, such as heat exchangers and other complex assemblies, because the filler metal is already in place, thus avoiding an extra step that would slow down the assembly operation. Brazing sheet is an excellent aid in assemblies where many joints need to be brazed simultaneously.

Brazing of Aluminum Alloys

E.B. Gempler

Fluxes

Brazing that is performed in air or other oxygen-containing atmospheres requires the use of a chemical flux. These fluxes, whether used in torch, furnace, or dip brazing, or in braze pastes, usually contain chlorides and/or fluorides, the fumes of which are hazardous to the operator if inhaled. Flux fumes must be exhausted, not only for health reasons, but because they are highly corrosive to most metals.

Fluxes act as deoxidizers and prevent reoxidation by coating the surfaces. They are used to aid the wetting and flow of the filler metal on deoxidized and clean aluminum.

There are four generic types of fluxes in two basic formulas: chloride active and fluoride active. Chloride-active fluxes consist of either a straight chloride flux or a chloride flux modified with zinc. Fluoride-active fluxes consist of either a straight fluoride or a modified fluoride that produces an inert film over the brazed parts after brazing. In both cases, the modification is intended to reduce corrosion in postbrazed products.

In torch brazing, fluxes can be applied dry or mixed with either alcohol or water and painted, or sprayed, or they can be applied by dipping the joining parts. A suitable flux for torch or furnace brazing must:

- BEGIN TO MELT AT A TEMPERATURE LOW ENOUGH TO MINIMIZE OXIDATION OF THE PARTS
- BE ESSENTIALLY MOLTEN AT THE TIME THE FILLER METAL MELTS
- FLOW OVER THE JOINT AND FILLER METAL TO SHIELD THEM FROM OXIDIZING GASES
- PENETRATE OXIDE FILMS
- LOWER THE SURFACE TENSION TO ENCOURAGE WETTING
- REMAIN LIQUID UNTIL THE FILLER METAL HAS MELTED
- BE RELATIVELY EASY TO REMOVE AFTER BRAZING IS COMPLETE

Fluxes for dip brazing are less active than those used for torch or furnace brazing, because parts to be brazed are totally immersed in the molten flux. Therefore, oxygen cannot reach the surface of the parts to reform oxides. In dip brazing, the flux acts as a heat transfer medium, in addition to its fluxing action, to bring the aluminum parts to final temperature for brazing.

Powdered fluxes are mixed with liquids and powders to form pastes. Pastes can be mixed in the brazing operation or they can be bought premixed from several suppliers.

Brazing of Aluminum Alloys

E.B. Gempler

Joint Designs

Assemblies to be brazed can be designed with many types of joints. When joints are designed to permit the application of flux, provisions must be made to ensure the displacement of flux and accompanying gases and to allow the flow of filler metal into the joint.

Lap joints can be used to achieve joint strength that is equal to the strength of the base metal. Lap joints that require a long flow path for the filler metal should be designed for flow in one direction only, to prevent the entrapment of potentially corrosive flux. Scarf and butt joints usually are not as strong as the base metal, but they can be designed to give satisfactory performance. (It should be noted that scarf and butt joints are weaker than lap joints because their mating area is less than that of lap joints.) For fluxless brazing, line contact is desirable and zero clearance is preferred. Certain brazing alloys (for example, the BNi series) can approach the strength of the parent metal if homogenized for a long period of time to provide grain size and microstructure characteristics similar to those of the parent metal.

In flux brazing, joint clearances must be sufficient to prevent premature solidification of the filler metal in small capillary spaces that are caused by the mutual solubility of the base metal and filler metal. These spaces can entrap flux and cause porosity. Clearances that range from 0.05 to 0.10 mm (0.002 to 0.004 in.) can suffice for narrow laps when preplaced filler metal is used. Wider laps may require clearances up to 0.25 mm (0.010 in.). With brazing sheet, line contact is desirable. Clearances should not exceed 0.076 mm (0.003 in.).

Brazing of Aluminum Alloys

E.B. Gempler

Preblaze Cleaning

Non-heat-treatable alloys usually can be cleaned by degreasing. However, this cleaning method is being replaced by solvent cleaning using special water-base products, because of environmental requirements. Chemical deoxidation may be required for components that have been severely worked, such as in spinning, drawing, or extrusion processes.

For heat-treatable alloys, chemical cleaning is necessary to reduce the tenacious oxide films, which tend to be thicker, because of thermal processing methods such as solution heat treatment. Aluminum-silicon alloys often require a special nitric-hydrofluoric acid mix to remove the excess silicon smut that is not removed by alkaline (5% NaOH) or nitric-acid solutions alone. If necessary, stainless steel wire brushes, steel wool, or abrasive cloth can be used to achieve abrasive cleaning.

Thin, high-silicon-content braze foil should only be degreased to prevent the loss of available filler metal that is needed at the braze joint.

Brazing of Aluminum Alloys

E.B. Gempler

Assembly

In flux brazing, self-fixturing is an excellent assembly method. It may be accomplished through the use of tabs, rivets, screws, or tack welds. Self-fixturing is cost-effective, quick, and lends itself to many different types of assemblies. Care must be exercised when choosing fastener or tack weld alloys with solidus temperatures that are higher than the brazing process temperature.

Spring-loaded fixtures can be useful, especially for high-production and complicated assemblies. These fixtures are excellent for stacked assemblies, such as plate-fin heat exchangers. When adjusted properly, spring-loaded fixtures can prevent the crushing of assemblies that often is due to thermal expansion and aluminum softening during heating.

Both coil springs and strap springs are used. Strap springs are less costly to use, but typically do not have the inherent resiliency of coil springs. Coil springs are more expensive than strap springs, but coil springs offer variations in coil diameter, coil thickness, stiffness, initial pressure, and the number of coils used that cannot be obtained with strap springs. Strap springs vary in width and thickness and are tightly bound around the fixture. When heated, strap springs tend to stretch and do not return to their original setting, thus losing stiffness and producing an effective loss in resiliency.

Weighted fixtures are often used on low-production runs and in cases where variations in shape and size preclude the use of more-expensive fixture types.

It is often necessary to apply stop-offs to parts of the assembly to be brazed in order to prevent filler metal from flowing onto certain areas. Stop-offs can consist of refractory oxides, graphites, or any number of commercial compounds readily available from welding and brazing suppliers. Furnace and dip brazing operations often use stop-offs to prevent fixtures and jigs from being brazed to the work. Stop-offs for fluxless brazing are usually refractory oxides sprayed on the jigs or fixtures. Stop-offs also can be formed on the fixtures by heating at a high temperature in an atmosphere.

Brazing of Aluminum Alloys

E.B. Gempler

Dip Brazing

Dip brazing is done by immersing the brazing assembly and fixture into a molten bath of flux. The flux acts as both a heating mechanism to bring the parts to temperature, as well as a deoxidizer. Prior to immersion, the part and fixture must be preheated to remove moisture and prevent solidification of the flux during immersion. Preheating is done in a hot-air furnace at approximately 540 °C (1000 °F) for about 0.5 h. Duration of the preheat cycle varies with the complexity of the brazement, its size, and the size of the fixture. Nevertheless, even with small waveguide components and thin box brazements that require a decreased preheat time, using a preheat cycle time of 0.5 h is still recommended. Moisture removal is necessary to prevent steam explosions when the part is lowered into the bath. Steam explosions can be very hazardous to the operator.

Initial temperatures are determined by attaching a thermocouple to the aluminum assembly. Time and temperature readings are taken and used for subsequent parts as a mechanism to control the braze quality.

Adequate salt drainage is necessary for several reasons. Salt must be removed to prevent trapping and the possible corrosion of the parts. Adequate drainage is necessary to allow filler metal solidification. Retention of excess salt in the assembly depletes the level of the molten salt in the dip pot, necessitating more-frequent additions of dry flux to the bath and resulting in higher costs.

The bulk of the salt is removed by immersing the fixture and brazement in a bath of boiling water. Then, the brazement is separated from the fixture and put through a series of water baths to remove more salt. Final baths, which use distilled water, are changed at intervals until chemical tests show an acceptable limit of salt in the clear runoff water. Concurrently, the fixtures are further boiled and then blasted with water to remove any residual salt.

The molten flux bath is regularly dehydrated by immersing clean sheets of 1100 or 3003 series aluminum to remove water, oxides, and other contaminants for an amount of time that varies from bath to bath. The sheets must be cleaned and dried to maintain their effectiveness. They can be used repeatedly.

Brazing of Aluminum Alloys

E.B. Gempler

Furnace Brazing

Standard fluxes must be used in furnace brazing because of the oxide atmospheres. Flux can be applied by brushing, dipping, or spraying. Furnace brazing is used for large loads that require good temperature control. Stop-offs are often used on parts and fixtures to inhibit the flow of molten filler metal to certain areas and to prevent brazing to the fixtures.

Special fluxes can be used with special furnaces that have reduced-oxygen atmospheres. One such flux is NOCOLOK, which has modest cleaning and fluxing ability and requires well-cleaned parts. During the brazing process, the chamber is filled with flowing dry nitrogen. At the completion of brazing, the parts are coated with an inert residue from the flux. The parts do not require postbraze cleaning to remove the flux residue.

The temperature accuracy requirement for these aluminum brazing furnaces is usually ±5.5 °C (±10 °F). However, ±3 °C (±5 °F) is preferred, especially where critical components are being brazed.

Brazing of Aluminum Alloys

E.B. Gempler

Torch Brazing

Torch brazing is used in either manual or automatic fabrication, as well as in repair operations. Torch brazing ranges from simple tube-to-tube joints to more-complex and mechanized assemblies. Torch brazing is similar to oxyfuel gas welding in that the heat to effect the joint is applied locally. Oxyacetylene, oxyhydrogen, and oxy-natural gas are used in commercial torch brazing applications. Close temperature control is needed when using the typical BAlSi-3 and BAlSi-4 filler metals.

Because aluminum shows no change in color when hot, it is difficult to know when parts approach the brazing temperature. Fluxes that melt at a slightly lower temperature than the filler metal can serve as a temperature indicator. Joint areas are pointed with a slurry of brazing flux. A soft, slightly reducing flame is played over the entire area. Filler metal can be either preplaced or force fed. After brazing, residual flux chloride must be removed.

Brazing of Aluminum Alloys

E.B. Gempler

Flux Removal Techniques

Whether a flux is used in torch brazing, furnace brazing, or dip brazing processes, it must be removed to prevent entrapment and the possibility of subsequent corrosion effects in the presence of moisture. Flux removal from dip-brazed assemblies was discussed in the dip brazing section of this article.

All fluoride fluxes are not safe to leave on the workpiece. If the flux residue remains in the inert form (for example, cryolite in the NOCOLOK process), it is considered to be an acceptable condition. However, under no circumstances should fluoride that can ionize remain on the surface because it can cause subsequent corrosion problems.

As much flux as possible should be removed by immersing torch-brazed and furnace-brazed parts in a bath of boiling water just after the filler metal has solidified. Scrubbing with a fiber brush in boiling water removes most of the flux. Pressure spray washers are often effective in opening passages that are plugged by flux. Agitation and turbulence, as well as ultrasound, help to remove flux from tight spaces. Acids (HNO_3, HNO_3-HF), as well as chromates, are often used in the final rinse water. Because of disposal problems, many companies avoid using chromates as corrosion inhibitors.

Testing for complete flux removal is done by monitoring the presence of chlorides in the final rinse water. An acidified solution of 5% silver nitrate is used to check the final rinse water for clarity. If white chloride residues cloud the water, then salt is still present. After several tests on subsequent rinses, the salt is considered removed if the water remains clear. Although this is a simple test, it is quite accurate.

Fluxless Vacuum Brazing

The main advantage of fluxless vacuum brazing is that no flux removal is necessary. If assemblies have been made using non-heat-treatable alloys, then they are ready for the next operation. If heat-treatable alloys are brazed, then a subsequent heat treatment will be necessary to reach the strengths required.

The required equipment is a vacuum furnace that can reach a vacuum of 133×10^{-5} Pa (10^{-5} torr) or lower, is leak tight, and is able to obtain temperature accuracy of ±3 °C (±5 °F) on the load being brazed.

Fluxless brazing is ideal for braze sheets where no flux is required and assemblies can easily be stacked and fixtured. Magnesium in some form is usually required to expedite the break-up of surface oxides and to facilitate the wetting mechanism. Magnesium is supplied by several techniques. It can be an integral part of the filler metal or base metal or it can be a part of other alloys used in the assemblies, such as the 5*xxx* series of aluminum-magnesium alloys. It can also be supplied in the form of chips or prealloyed aluminum-magnesium ingots.

Vacuum furnaces must be cleaned periodically to remove magnesium and magnesium oxide that condenses on the cold walls of the furnace and other appendages to those walls. These residues may, in time, short some of the electrodes and tend to be hygroscopic to water. Water is a contaminant that is very difficult to remove in vacuum.

Postbraze Heat Treatment

Dip-brazed heat-treatable assemblies can be spray quenched or immersion quenched after most of the salt has been drained and while the parts still remain at about 480 °C (900 °F). This puts the parts into a solution-heat-treated stage for subsequent precipitation hardening.

Heat-treatable assemblies usually are in the annealed or soft condition after furnace, torch, or fluxless brazing. Clean assemblies are usually heated at temperatures ranging from 480 to 515 °C (900 to 960 °F) for a period of time that allows the magnesium silicates to go into solid solution prior to quenching. The quenched parts are then aged for a period of time that allows precipitation hardening to take effect. The amount of time depends on the alloy system and on whether the parts are made from wrought or cast alloys. A typical precipitation heat treatment for 6061 shapes and tube is 160 °C (320 °F) for 16 to 20 h.

Safety Precautions

Exhaust fumes must be removed from the immediate vicinity of the operators. Suitable masks should be worn to prevent inhalation of these fumes.

Precautions are necessary when dip brazing. Operators must wear suitable burn-and heat-resistant coverings to protect themselves from splashes or burns and from inhaling flux fumes. As with all flux-type operations, fume exhaust is a necessity.

Lowering and removal of the parts and fixtures must be done at a rate that minimizes spatter from the molten bath. Parts to be brazed (including fixtures) must be preheated to prevent steam explosions as the part is lowered into the bath. Preheating also prevents the lower-temperature fixture and aluminum assembly from freezing out the molten salt as it is immersed.

The temperature of aluminum can be very deceptive, because this material does not turn red or glow when it is hot, like steel does. Precautions must be taken to prevent burns.

The cleaning of vacuum furnaces has already been discussed. When removing magnesium and magnesium oxides, nonsparking scraping tools should be used to prevent sparks that can ignite magnesium. Dust should be swept frequently into a metal container and removed from the vicinity of the furnace to prevent fires. The cleaning operators must work with adequate ventilation. Masks must also be used by the operators. The cleaning must be done by more than one operator as a precaution against accidents. Several buckets of appropriate materials, such as sand, as well as fire extinguishers, should be available at all times to douse potential fires.

Brazing of Refractory and Reactive Metals

Roy I. Batista, TRW, Inc.

Introduction

REFRACTORY METALS are metals that have melting points that are either higher than those of iron, nickel, and cobalt or exceed 2000 °C (3630 °F). Molybdenum, tungsten, niobium, and tantalum are the refractory metals discussed in this article. Reactive metals, which include titanium, zirconium, and beryllium, have in common the ability to form stable oxides at elevated temperatures. These diverse groups of metals have a wide range of mechanical and physical properties (Table 1).

TABLE 1 SELECTED PHYSICAL PROPERTIES OF REACTIVE AND REFRACTORY METALS

METAL	**LIQUIDUS TEMPERATURE**		**SOLIDUS TEMPERATURE**		**ANNEALING TEMPERATURE**		**COEFFICIENT OF THERMAL EXPANSION, 10^{-6}/K**	CRITICAL TEMPERATURE	
	°C	**°F**	**°C**	**°F**	**°C**	**°F**		**°C**	°F
MOLYBDENUM	2610	4730	2610	4730	1200	2192	4.9	RT(A)	**RT(A)**
MO-0.5TI	2610	4730	2610	4730	1450	2642	4.9	. . .	**. . .**
MO-0.5TI-0.008ZR	2620	4748	. . .	. . .	1550	2822	4.9	. . .	**. . .**
MO-50RE	2550	4622	2480	4496	1500	2732	6.3	. . .	**. . .**
TUNGSTEN	3410	6170	3410	6170	1300	2372	4.3	300(A)	**572(A)**
W-25RE	3100	5612	. . .	. . .	1500	2732	6.7	RT(A)	**RT(A)**
NIOBIUM	2468	4474	2468	4474	1150	2102	7.2	<-100(A)	**<-148(A)**
NB-10HF	2350	4262	. . .	. . .	1300	2372	8.1	. . .	**. . .**
NB-28TA-10W-1ZR	2590	4694	. . .	. . .	1370	2498	9	. . .	**. . .**
NB-10TA-10W	2600	4712	. . .	. . .	1400	2552	14	. . .	**. . .**
TANTALUM	2996	5425	. . .	. . .	1300	2372	6.6	<-200(A)	**<-328(A)**
TA-10W	3035	5495	. . .	. . .	1550	2822	6.7	. . .	**. . .**
TA-8W-2HF	3000	5432	. . .	. . .	1650	3002	6.5	. . .	**. . .**
TITANIUM 99+ TI(B)	1671	3040	1671	3040	700	1292	9.2	950(C)	**1742(C)**
TI-5AL-2.5SN (ALPHA)	1677	3050	. . .	. . .	820	1508	9.5	1040(C)	**1904(C)**
TI-8AL-1MO-IV	1650	3002	. . .	. . .	790	1454	9	1040(C)	**1904(C)**

(NEAR ALPHA)									
TI-6AL-2NB-1TA (NEAR ALPHA)	1650	3002	. . .	. . .	900	1652	9	1110(C)	**2030(C)**
TI-6AL-4V (ALPHA-BETA)	1660	3020	1615	2939	790	1454	9.1	990(C)	**1814(C)**
TI-6AL-6V-2SN (ALPHA-BETA)	1649	3000	1627	2961	800	1472	9	946(C)	**1735(C)**
TI-3AL-8V-6CR-4ZR-4MO (BETA)	1650	3002	1555	2831	800	1472	8.4	795(C)	**1463(C)**
ZIRCONIUM	1845	3353	1845	3353	. . .	. . .	5.8	865(C)	**1589(C)**
ZR-1.5SN-0.2FE-0.1CR	1845	3353	1845	3353	. . .	. . .	5.5	>865(C)	**>1589(C)**
BERYLLIUM	1285	2345	1285	2345	730	1346	11.4	. . .	**. . .**
BE-38AL	**660**	**1220**	**. . .**	**. . .**	**. . .**	**. . .**	**13.9**	**. . .**	**. . .**

RT, room temperature.

(A) DBTT OF ANNEALED MATERIAL.

(B) SIMILAR ALLOYS WITH VARYING INTERSTITIAL COMPOSITIONS.

(C) BETA TRANSUS

A vacuum atmosphere is suitable for brazing all of the refractory and reactive metals. Tungsten and molybdenum also can be brazed in reducing atmospheres and, in some applications, in air using a flux. The use of flux in the brazing of beryllium has been routine, but the preferred vacuum techniques are being developed. Inert gases with properly controlled purity are also useful atmospheres for the brazing of both refractory and reactive metals.

This article describes brazing filler metal selection and brazing procedures for each of these metals. The importance of fabricating samples to design requirements and then testing to the degree necessary to validate brazing procedures cannot be overemphasized. These materials do not have the data base that exists for alloys based on copper, iron, or nickel, and contradictions exist in the literature. Because specific details on brazing equipment, such as vacuum levels, type of vacuum equipment, quality of inert gases, and brazing cycles, are not reported in most of the development work on the refractory and reactive metals, verification is needed for specific braze designs.

The selection of filler metals presents a problem, because many recommendations are for materials that are not readily available. However, an attempt was made in this article to provide selections from commercially available compositions. Although these commercial alloys may not be optimal in all applications, they should be evaluated first to assess their suitability for specific designs. Consult the selected references at the end of this article for specific information on nonstandard brazing filler metals.

Brazing of Refractory and Reactive Metals

Roy I. Batista, TRW, Inc.

Refractory Metals

The four refractory metals that have the most commercial significance are molybdenum, tungsten, niobium, and tantalum. The brazing of these alloys is a complex subject, because their high melting points and specific mechanical, physical, and chemical properties have highly diverse applications, and the brazing procedures required by each application can be quite distinct. Molybdenum is used in high-temperature structures and electronic devices. Tungsten is most commonly associated with filaments, heating elements, and welding electrodes. Niobium is widely used in spacecraft propulsion systems and applications that require its corrosion resistance. The largest use of tantalum is for capacitors, followed by chemical processing equipment.

Ductile-to-Brittle Transition Temperature (DBTT)

DBTT is a metallurgical consideration that greatly influences the selection of brazing filler metals and the brazing environment. Tantalum and niobium are Group V metals and have a high solubility of interstitial elements and a low DBTT. Conversely, tungsten and molybdenum are Group VI metals and have a lower solubility for interstitial elements, but a high DBTT. Tungsten and molybdenum are specifically processed and alloyed to lower the DBTT. For both tungsten and molybdenum, a warm-worked structure has a lower DBTT than a recrystallized structure, and alloys have been developed to retard annealing. Annealed molybdenum has a DBTT that is near room temperature, whereas annealed tungsten may be brittle at temperatures below 315 °C (600 °F). Tantalum does not appear to have a ductile-to-brittle transition, and the DBTT of niobium can be as low as -200 °C (-330 °F).

Gas Reactions

The refractory metals are all sensitive to the absorption of, or reaction to, oxygen. Niobium and tantalum begin oxidizing at 200 and 400 °C (390 and 750 °F), respectively. Below 1370 °C (2500 °F), the oxides are nonvolatile, but the ductility of the parent metals is reduced, because of the absorption of oxygen.

Molybdenum starts to oxidize at about 500 °C (930 °F). At temperatures over 778 °C (1430 °F), which is the eutectic temperature of MoO_2-MoO, oxides become particularly volatile and the oxidation rate accelerates rapidly. Tungsten oxidizes rapidly at temperatures over 500 °C (930 °F), but not as rapidly as molybdenum. In general, oxygen absorption is not as much of a problem with tungsten and molybdenum as it is with tantalum and niobium.

Hydrogen exposure presents no problems to the use of molybdenum and tungsten. However, both tantalum and niobium will absorb hydrogen and form hydrides at moderate temperatures. The results can range from a loss of ductility at low concentrations to complete destruction, depending on time and hydrogen concentration. If it is not heavily contaminated, then hydrogen can be removed from tantalum and niobium by a high-temperature vacuum anneal.

Alloy Availability

Although there has been considerable development of the refractory metal alloys, their commercial availability in small quantities can be erratic. The alloys that are most common are identified in Table 1. Generally, tantalum, niobium, and their alloys are usually procured and used in the annealed condition. Tungsten and molybdenum are often used in the stress-relieved condition, because of higher strength and a lower DBTT. TZM is the most common alloy of molybdenum, and both molybdenum and tungsten are alloyed with rhenium to increase strength and ductility and to lower the DBTT. Molybdenum and tungsten may both contain dopants, such as potassium, to reduce grain growth at high temperatures, particularly in filament applications.

Brazing Filler Metal Selection

The wide range of applications and the high melting points of the refractory metals mean that a wide range of brazing alloys are available, but this also makes the selection process more complex. Alloys have been specifically developed for high-temperature, low-temperature, high-strength, or corrosion-resistant applications.

The higher the brazing temperature and the service temperature, the more difficult it becomes to select a brazing filler alloy. As temperature increases, adverse metallurgical reactions, erosion, and fixturing concerns are amplified. For example, nickel-containing alloys have resulted in brittle joints in both tungsten and molybdenum brazing, because of intermetallic formation. This brittleness can be minimized by short brazing cycles.

The refractory metals are not difficult to join. The objective of the selection process is to select an alloy that will provide the right combination of properties in service, such as high temperature, mechanical properties, or corrosion resistance.

Filler Metals for Brazing of Molybdenum. The commercially available alloys based on silver, gold, or palladium have all been evaluated for refractory metal brazing. In Ref 1, the 19 filler metals considered for molybdenum and the 32 considered for TZM are ranked on a wettability index. It was found that TZM was less-readily brazed than molybdenum, because of titanium oxide formation on its surface. As expected, higher brazing temperatures promoted wettability. Table 2 lists commercially available brazing filler alloys, as well as those alloys that may have special applications, but may not be commercially available.

TABLE 2 BRAZING FILLER METALS FOR REFRACTORY METAL BRAZING

BRAZING ALLOYS	**LIQUIDUS TEMPERATURE**		**SOLIDUS TEMPERATURE**		COMMENTS
	°C	**°F**	**°C**	**°F**	
COMMON BASE METALS					
25CR-20NI-FE	1454	2650	. . .	. . .	**25-20 STAINLESS STEEL**
55CO-20NI-15W-10NI	1427	2600	. . .	. . .	**L-605 COBALT ALLOY**
18CR-8NI-FE	1427	2600	. . .	. . .	**STAINLESS STEEL**
54CO-27CR-6MO-5NI	1400	2550	. . .	. . .	**ALLOY 21**
80NI-14CR-6FE	1393	2539	. . .	. . .	**INCONEL**
67NI-33CU	1349	2460	. . .	. . .	**MONEL**
COMMERCIALLY AVAILABLE SILVER-BASE BRAZING ALLOYS					
75AG-20PD-5MN	1072	1962	1008	1846	
AG	961	1762	961	1762	
54AG-21.3CU-24.7PD	950	1742	900	1652	**VARIOUS AG-CU-PD ALLOYS ARE AVAILABLE WITH DECREASING MELTING POINTS**
72AG-28CU	780	1436	780	1436	
82AG-9GA-9PD	880	1616	845	1553	
71.5AG-28.1CU-0.75NI	795	1463	780	1436	
54AG-42CU-2NI	893	1639	771	1420	
68AG-27CU-10SN	760	1400	743	1369	
COMMERCIALLY AVAILABLE GOLD-BASE BRAZING ALLOYS					
92AU-8PD	1240	2264	1200	2192	
50AU-25PD-25NI	1121	2050	1102	2016	**NUMEROUS AU-PD-NI COMPOSITIONS ARE AVAILABLE WITH DECREASING MELTING POINTS**
73.8AU-26.2NI	1010	1850	990	1814	
50AU-50CU	970	1778	955	1751	
82AU-18NI	950	1742	950	1742	
PURE METAL BRAZING FILLER METALS					
RE	3186	5767	. . .	. . .	**TUNGSTEN BRAZING, REMELT 2830 °C (5126 °F)WITHOUT DIFFUSION TREATMENT**
TA	3020	5468	. . .	. . .	**TUNGSTEN APPLICATIONS**
NB	2471	4480	. . .	. . .	**TUNGSTEN APPLICATIONS**
RU	2254	4089	. . .	. . .	**TUNGSTEN APPLICATIONS**
RH	1963	3565	. . .	. . .	**TUNGSTEN APPLICATIONS**
PT	1768	3214	. . .	. . .	**MOLYBDENUM AND TUNGSTEN APPLICATIONS**
PD	1552	2826	. . .	. . .	**MOLYBDENUM AND TUNGSTEN APPLICATIONS**
OTHER COMMERCIALLY AVAILABLE BRAZING ALLOYS					
80MO-20RU	2275	4127	2125	3857	**MOLYBDENUM AND TUNGSTEN APPLICATIONS**
60PT-20RH	1990	3614	1950	3542	**MOLYBDENUM AND TUNGSTEN APPLICATIONS**
75PT-20PD-5AU	1695	3083	1645	2993	**MOLYBDENUM AND TUNGSTEN APPLICATIONS**
65PD-35CO	1220	2228	1220	2228	**MOLYBDENUM AND TUNGSTEN APPLICATIONS**
OTHER BRAZING ALLOYS (NOT COMMERCIALLY AVAILABLE, SPECIAL ORDER REQUIRED)					
W-50MO-3RE	2900(A)	5252(A)	. . .	. . .	
MO-50S	2600(A)	4712(A)	. . .	. . .	
W-250S	2600(A)	4712(A)	. . .	. . .	
V-50MO	2257(A)	4095(A)	. . .	. . .	
TI-25CR-21V	1482(A)	2700(A)	. . .	. . .	**REMELT 1921 °C (3490 °F)**
84NI-16TI	1288(A)	2350(A)	. . .	. . .	**HIGH STRENGTH AT 1000 °C (1830 °F)**
FE-15MO-4C-1B	1200(A)	2192(A)	. . .	. . .	**CORROSION RESISTANT IN LIQUID BISMUTH AND SODIUM**
52NB-48NI	1190(A)	2174(A)	. . .	. . .	**HIGH STRENGTH AT 1000 °C (1830 °F)**
AG-30PD	1148(A)	2098(A)	. . .	. . .	**MINIMUM EROSION OF MOLYBDENUM**
TI-25CR-3BE	1110	2030	993	1819	**REMELT 1600 °C (2912 °F) AFTER DIFFUSION**

					TREATMENT
PT-B	**1100$^{(A)}$**	**2012$^{(A)}$**	**. . .**	**. . .**	**BORON CONTENT 1-4.5%; REMELT TO 2100 °C (3812 °F) AFTER DIFFUSION**

(A) BRAZING RANGE

Examples of special-purpose brazing include the joining of thin foils, such as the honeycomb structure in which the alloy Ag-30Pd alloy was used. At a brazing temperature of 1160 °C (2120 °F), 0.025 mm (0.001 in.) molybdenum foils were joined without excess erosion. In a high-temperature application, a brazed Ti-25Cr-21V alloy (solidus temperature of 1160 °C, or 2120 °F) had a remelt temperature of 1600 °C (2910 °F) after a 16 h postbraze diffusion cycle at the solidus temperature. The other recommended alloy systems for high-temperature performance include Ti-48Zr-4Be (brazing temperature of 986 °C, or 1810 °F), Ti-25Cr-21V (brazing temperature of 1482 °C, or 2700 °F), and Ti-28Zr-8Ge (brazing temperature of 1327 °C, or 2420 °F). The TiCrV composition has. a high remelt temperature of 1920 °C (3490 °F) (Ref 2). The high remelt temperatures of the chromium-containing alloys may be a result of the volatilization of the chromium during brazing. Therefore, a vacuum environment may be required to achieve these remelt temperatures.

Other high-temperature brazing filler alloys are Ti-65V (brazing temperature of 1700 °C, or 3090 °F) and V-50Mo (brazing temperature of 2257 °C, or 4095 °F). After postbraze diffusion treatment, shear test specimens failed in the parent metal when tested at 1400 °C (2550 °F). Additions of base-metal powder to the brazing filler alloy were necessary to prevent Kirkendall void formation (Ref 3).

A Mo-20Ru brazing filler was used to braze a doped Mo-3Re filament to a tungsten post. Because of a requirement to minimize any recrystallization of the filament, a laser welder working in an argon atmosphere was used to heat the assembly. Excellent wetting of both the molybdenum and tungsten was achieved without recrystallization of the filament adjacent to the braze (Ref 4).

When using brazing alloy powders, the addition of elemental powders to act as a "diffusion sink" is an effective method of raising remelt temperatures. The diffusion sink material will either shorten the diffusion time to achieve a high remelt temperature or it may form higher-melting-point compounds within the brazed joint. An example of a diffusion sink combination is Ti-25Cr-13Ni alloy combined with molybdenum powder that brazes at 1260 °C (2300 °F), but remelts at 1710 °C (3110 °F).

Filler Metals for Brazing of Tungsten. The brazing filler alloys that are appropriate for use with molybdenum are generally well suited for use with tungsten. In addition, there has been development work for specific tungsten applications. Platinum-boron alloys containing from 1 to 4.5% boron were brazed at about 1100 °C (2010 °F), both with and without tungsten powder additions. Remelt temperatures of 2100 °C (3810 °F) were obtained. The use of a diffusion sink (tungsten powder) was not necessary when gaps were below 0.038 mm (0.0015 in.). The addition of a halide salt, NaF or KHF_2, reduced the possibility of W_2B compound formation, thus raising the remelt temperature to about 2300 °C (4170 °F) (Ref 5).

Three alloys were evaluated for use at temperatures up to 2500 °C (4530 °F): W-25Os (melting temperature of 2600 °C, or 4710 °F), W-50Mo (melting temperature of 2900 °C, or 5250 °F), and Mo-5Os (melting temperature of 2600 °C, or 4710 °F). At 2200 °C (3990 °F), base-metal strength values were reported (Ref 6).

The number of alloys evaluated for use with both tungsten and molybdenum is extensive. A review of the selected references is recommended if nonstandard alloys are indicated for a specific application.

Filler Metals for Brazing of Niobium. Niobium may be selected for its corrosion resistance in chemical and nuclear applications or for its low density and high melting point in aerospace structural and propulsion applications. The welding of niobium does not present the brittleness problems associated with tungsten and molybdenum. However, brazing must be considered when welding is inappropriate.

Although commercial brazing alloys based on silver, gold, and nickel will wet niobium, brittleness is often noted with their use. In addition, their low melting points make them unusable for many applications. The active-metal modifications (titanium additions) of the silver-copper and gold-nickel alloys may be useful for niobium joining, particularly if residual oxide on the surface presents a wetting problem.

A series of Ta-V-Nb and Ta-V-Ti alloys is compatible with niobium and will resist liquid sodium. They can be brazed at temperatures ranging from 1760 to 1925 °C (3200 to 3500 °F), and have shear strength values that range from 140 to 205 MPa (20 to 30 ksi). The alloy Nb-1Zr, when brazed with Ti-28V-4Be, was resistant to potassium vapor and liquid exposed for 1000 h at 815 °C (1500 °F). Honeycomb structures of Nb alloys were evaluated using Pd-30Cu and Ti-11Cr-13V-3Al brazing filler alloys. There was some erosion with the palladium alloy, but the titanium-base alloy produced joints with high ductility and minimum base-metal erosion.

Two diffusion-enhanced approaches have been explored to produce higher-temperature properties in niobium brazes. One method adds base-metal powder to the brazing alloy, that is, niobium powder added to a Ti-33Cr alloy brazed at 1485 °C (2705 °F). Another approach, which depends on diffusion between the braze and parent material, is niobium brazed with a Ti-30V-4Be alloy that was brazed at 1315 °C (2400 °F). Postdiffusion of the Ti-33Cr alloy raised the remelt temperature by over 300 °C (540 °F), with corresponding increases to high-temperature strength. The Ti-Cr-V alloy had about the same remelt temperature, but its strength was lower.

A 316L stainless steel sleeve was brazed to the inner diameter of a Nb-10Hf rocket engine thrust chamber using 82Au-18Ni filler metal. The joint failed shortly after removal from the vacuum furnace, because of residual stresses resulting from the difference in thermal expansion. The joint was redesigned so that the 316L material was the outer member. This configuration was successfully brazed and the engine was flight qualified (Ref 4). This example emphasizes the need to consider the thermal expansion of the brazement pieceparts to prevent failure after brazing. It was fortunate that the assembly failed shortly after brazing. The residual stresses in a mismatched joint could just as easily cause failure weeks or months after brazing.

Filler Metals for Brazing of Tantalum. The brazing of tantalum is not as well documented as the brazing of tungsten and molybdenum. This situation may exist because, like niobium, it is an easily weldable material without the brittleness problem encountered with welded tungsten and molybdenum. Many of the brazing filler alloys presented for tungsten and molybdenum, as well as all of the recommendations made for niobium, can be considered for use with tantalum.

The diffusion-sink titanium powder and a Ti-30V brazing filler alloy (braze temperature of 1760 °C, or 3200 °F) produced a remelt temperature of 2100 °C (3810 °F). A similar approach with a 33Zr-34Ti-33V (brazing temperature of 1427 °C, or 2600 °F) resulted in a remelt temperature of 1650 °C (3000 °F). Other approaches using hafnium-base alloys have been considered. These include Hf-7Mo, Hf-40Ta, and Hf-19Ta-2.5Mo.

References cited in this section

1. M.M. MCDONALD, *ET AL.*, THE WETTABILITY OF BRAZING ALLOYS ON MOLYBDENUM AND TZM, *WELD. J. SUPPL.*, OCT 1988
2. G.M. SLAUGHTER *ET AL.*, WELDING AND BRAZING OF ADVANCED REFRACTORY ALLOYS, *SPACE SHUTTLE MATERIALS PROCEEDINGS*, SAMPE, OCT 1971
3. L.B. LUNBERG *ET AL.*, BRAZING OF MOLYBDENUM AND TUNGSTEN FOR HIGH-TEMPERATURE SERVICE, *WELD. RES. SUPPL.*, OCT 1978
4. R.I. BATISTA, TRW, INC., UNPUBLISHED NOTES
5. C.W. HAYNES, "DEVELOPMENT OF LOW TEMPERATURE BRAZING OF TUNGSTEN FOR HIGH TEMPERATURE SERVICE," REPORT R>D>R 1249-1/DTIC AD 260580, DEFENSE TECHNOLOGY INFORMATION CENTER, MAY 1962
6. T.J. MOORE *ET AL.*, "BRAZING TUNGSTEN FOR HIGH TEMPERATURE NUCLEAR SERVICE IN HYDROGEN," REPORT TN D-3944, NASA, MAY 1967

Brazing of Refractory and Reactive Metals

Roy I. Batista, TRW, Inc.

Reactive Metals

Of the reactive metals, titanium and its alloys are the most likely to be fabricated outside of a specialized facility. It is a structural material that finds varied uses, ranging from aircraft to body implants. The use of zirconium is most common in nuclear applications. Beryllium has many specific aerospace applications, because of its low density and high modulus values, but is most commonly utilized in its pure form, in x-ray window applications. Only beryllium and titanium have useful properties after exposure to elevated temperatures in an oxidizing atmosphere.

Titanium Alloys

Titanium is the most commercially available material among both reactive and refractory metals, in terms of its number of alloys, available product forms, and metallurgical complexity. These factors all impact brazing recommendations.

Because of the cost of titanium, its use is usually confined to those applications that take advantage of its specific properties, such as high strength-to-weight ratio, corrosion resistance, and high-temperature strength. Aerospace applications include airframe and engine components for both aircraft and spacecraft. Applications that require its corrosion resistance include chemical/petrochemical processing, marine exposure, and medical uses, such as instruments and implants. There are also emerging applications in the automotive and consumer products industries.

Titanium alloys are available in commercially pure and alloy compositions. They are generally categorized into four groups, depending on the principal phases present in the microstructure:

- COMMERCIALLY PURE (CP) ALLOYS (HEXAGONAL CLOSE-PACKED, OR HCP)
- ALPHA ALLOYS (HCP)
- ALPHA-BETA ALLOYS
- BETA ALLOYS (BODY-CENTERED CUBIC, OR BCC)

In pure titanium, the transformation (beta transus) occurs at 888 °C (1630 °F). By alloying, the alpha phase can be stabilized to increase the beta transus. Other alloy compositions may stabilize the beta phase or result in a two-phase alpha-beta structure. These various structures have a marked effect on mechanical properties and their response to heat treatment.

Commercially Pure and Alpha Alloys. The alpha-stable compositions are the CP grades and those alloyed with alpha stabilizers. The variations in strength of the CP grades depend on purity, particularly in terms of oxygen, nitrogen, hydrogen, silicon, and iron. The control of oxygen and iron is specified in order to control strength. The other interstitials are minimized to avoid embrittlement. As a class, the CP alloys are usually specified for applications requiring corrosion resistance, but they have mechanical properties that are useful for structural applications. The Ti-0.2Pd and Ti-0.3Mo-0.8Ni grades are also considered to be CP grades and are specified for more-corrosion-resistant applications. These alloys are generally used in the annealed condition and cannot be strengthened by heat treatment.

Alpha alloys are stabilized with aluminum and tin/zirconium. These alloys have both high-temperature and cryogenic applications. Their creep strengths are higher than those of the alpha-beta and beta compositions. The extra-low-interstitial (ELI) grades are specified for severe applications. The alpha alloys also have good weldability, but poor forging properties and an inability to be strengthened by heat treatment.

Near-beta alloys, which contain small amounts of beta stabilizers, are included with the alpha alloys. They have enhanced creep strength and elevated-temperature properties.

Alpha-beta alloys contain both alpha and beta stabilizers. The beta stabilizers are generally vanadium, molybdenum, and manganese. Titanium 6Al-4V, a classic example of an alpha-beta alloy, represents over 40% of the titanium produced.

Although these alloys can be strengthened by heat treatment, some of them, including Ti-6Al-4V, are used in the annealed condition. The heat treatment consists of quenching from the solution heat-treating temperature, followed by aging (solution-treated and aged, or STA, condition). The phases that are present after quenching vary with the alloy, the solution temperature, the cooling rate, and the cross-section size, but a beta phase will always be retained. Aging decomposes the retained beta as a fine alpha structure in the retained beta matrix and, accordingly, strengthens the alloy. Aging temperatures are between 425 and 650 °C (795 and 1200 °F). The equiaxed annealed structure has higher ductility and better low-cycle fatigue properties than the STA condition. An acicular structure results from the STA heat treatment and has higher creep resistance, higher fracture toughness, and higher stress-corrosion resistance than the annealed condition.

The beta alloys have high concentrations of beta stabilizers that slow the transformation from the beta to the alpha phase. After normal processing, the structure is nearly all beta. The beta phase partially transforms into a finely dispersed alpha structure while aging at temperatures from 450 to 650 °C (840 to 1200 °F). When these alloys are in the beta phase and are solution heat treated, their ductility is excellent and their abilities to be forged and formed are superior to those of the alpha-beta alloys. After aging, the beta alloys are generally stronger and have higher fracture toughness values than the alpha-beta alloys, but they have lower creep resistance and ductility.

Applications that utilize the high formability and strength of beta alloys include complex aerospace structures and fasteners. The compositions and physical properties of selected titanium alloys are shown in Table 1.

Restrictive Titanium Brazing Attributes. Because of the variety of titanium alloys and corollary heat treatments, there are brazing limitations that can avoid possible compromise of the mechanical properties after brazing. To maintain properties, the brazing alloys must be selected to ensure that the brazing temperature will be within the recommended heat treatments of the alloys being brazed. Commercially pure and alpha alloys should be brazed below the beta transus to prevent a coarse structure that reduces ductility. Although alpha-beta alloys are usually solution treated below the beta transus, the beta transus can be exceeded in order to obtain specific high-temperature properties. Brazing temperatures higher than those specified for beta alloys can result in grain coarsening and a loss in ductility.

For any titanium brazement, it is recommended that the design requirements for strength and ductility be reviewed to confirm the suitability of the selected alloy and of using a brazed approach. Certain combinations that require high-temperature brazing and/or quenching may not be acceptable. For example, the brazing of honeycomb sandwich, or any thin-sheet metal structure, is not practical with an alloy that requires quenching for the required properties. When possible, the brazing filler alloy should be selected to coincide with a standard heat treatment, such as an annealing or a solution heat-treating temperature.

Brazing filler metal selection should occur within the restraints noted above and with consideration to the environment to which the brazement will be exposed. Of particular concern is the possibility of galvanic corrosion. Although it is a reactive metal, oxidized titanium is noble to many structural alloys. In critical applications, braze samples should be exposed to the end-use environments to ensure that there is not preferential corrosion of the brazing filler alloy.

A review of titanium brazing literature reveals a considerable body of work that includes many compositions that would not be commercially available. Considerable effort was expended on silver, both pure and alloyed, with various combinations of lithium, copper, manganese, cadmium, and tin. Pure silver appears preferable to most of the silver alloys. The Ag-Cu-Li alloys were successful if the brazing cycles were short to prevent excessive erosion of the base metal and to minimize the formation of the brittle Cu-Ti intermetallic.

Aluminum alloys, including 3003 and 4043, are also available, if their lower melting points are acceptable. Pure aluminum (99.9%) forms a thicker intermetallic zone than the preferred alloys, because the alloy additions, particularly silicon, reduce intermetallic formation.

A 82Al-9Pd-9Ga alloy is commercially available. It easily wets titanium and has good strength and corrosion resistance. Base-metal erosion is minimal when short brazing cycles and minimum brazing temperatures are used.

A Ga-Pd-Ag brazing filler alloy was used on two structural joints on the Mars Viking landing instrument. One application was a triangular truss structure consisting of a socket comer fitting and three tubular legs. A cold-wall vacuum furnace was used in combination with a short brazing cycle, about 5 min at temperature. After brazing, the part did not meet dimensional tolerances because of an improperly machined fixture. The brazement was separated by immersion in HNO_3 until the brazing filler metal was dissolved. There was essentially no attack of the titanium by the molten filler and the assembly was rebrazed on modified fixtures. Because of the required close tolerances of the part, the fixtures were made of titanium to avoid potential problems with thermal expansion differences.

The second Viking application was a titanium thermal standoff that was brazed to 304L stainless steel. The stainless steel was electrolytically plated with nickel and the part was brazed in a cold-wall vacuum furnace. Although wetting of the stainless steel by the Ga-Pd-Ag filler was not satisfactory, the nickel plating was easily wet by the filler.

The Pd-Cu-Ag series of alloys is also reported to have superior wetting characteristics on titanium, but its use would be limited by a high copper content. Ti-Ni-Cu alloys are available in several compositions and are increasingly employed for titanium brazing. Commercially available titanium brazing alloys are identified in Table 3.

TABLE 3 BRAZING FILLER METALS FOR REACTIVE METAL BRAZING

BRAZING ALLOYS	**LIQUIDUS TEMPERATURE**		**SOLIDUS TEMPERATURE**		APPLICATIONS
	°C	**°F**	**°C**	**°F**	
AL-0.12CU-1.2MN (3003)	649	1200	638	1180	**BERYLLIUM, LOW-STRENGTH TITANIUM APPLICATIONS**
AL-5.2SI (4043)	632	1170	574	1065	**BERYLLIUM, LOW-STRENGTH TITANIUM APPLICATIONS**
AG-30CU-10SN	718	1324	602	1116	**BERYLLIUM**
AG-28CU-0.2LI	779	1434	779	1434	**BERYLLIUM, HIGH-TEMPERATURE APPLICATIONS**
AG-5AL	810	1490	780	1436	**TITANIUM ZIRCONIUM**
TI-20ZR-20CU-20NI	848	1558	842	1548	**TITANIUM ZIRCONIUM**
TI-15NI-15CU	850	1562	830	1526	**TITANIUM ZIRCONIUM**
AG-26.7CU-4.5TI	850	1562	830	1526	**TITANIUM ZIRCONIUM**
AG-9PD-9GA	880	1616	845	1553	**TITANIUM ZIRCONIUM**
TI-15CU-15NI	932	1710	902	1656	**TITANIUM ZIRCONIUM; ALTERNATE CHEMISTRY WITH LOWER MELTING POINTS AVAILABLE**
AG-21.3CU-24.7PD	950	1742	900	1652	**TITANIUM, ZIRCONIUM; ALTERNATE CHEMISTRY WITH LOWER MELTING POINTS AVAILABLE**
ZR-8CR-8NI	950(A)	1742(A)	. . .	. . .	**ZIRCONIUM, NOT A COMMERCIAL ALLOY**
ZR-5BE	990(A)	1814(A)	. . .	. . .	**ZIRCONIUM, NOT A COMMERCIAL ALLOY**
TI-48ZR-4BE	**985(A)**	**1805(A)**	**. . .**	**. . .**	**TITANIUM, ZIRCONIUM, NOT A COMMERCIAL ALLOY**

(A) BRAZING TEMPERATURE

Of the noncommercial alloys, there has been success with Ti-Zr-Be and Ag-Al-Mn alloys. The beryllium additions have potential health considerations and the section of this article that addresses beryllium should be reviewed before its use is considered.

For all but the most-benign applications, the selection of titanium filler alloys and brazing cycles should be supported by braze samples that simulate the fabrication and design requirements.

Zirconium Alloys

Zirconium is available in most common mill products, such as sheet, strip, bar, plate, and billet. It can be forged and extruded. Commercially pure material and alloys that contain tin or niobium are available. The commercial grades may contain up to 4.5% hafnium, a neutron absorber, which prohibits their use in nuclear applications. The successful use of zirconium as a nuclear material is due to its low thermal neutron cross section, good mechanical properties, and excellent hot-water and steam corrosion resistance. Its high corrosion resistance to both organic and inorganic acids and its immunity to attack by alkalies has led to applications in the petrochemical and food processing industries.

Pure zirconium is hexagonal close-packed (alpha phase) at room temperature. It transforms to body-centered cubic at about 870 °C (1600 °F). The commercially pure grades and the tin-containing alloys are alpha structures at room temperature. Niobium is a beta stabilizer, and its addition to zirconium results in an alpha-beta structure at room temperature with a beta transus at 854 °C (1570 °F).

During primary fabrication, zirconium alloys are worked in the beta region. This is followed by a solution heat treatment in the beta region, at approximately 1065 °C (1950 °F), and a water quench. Further processing is performed in the alpha region to preserve the interstitial distribution established by solution heat treatment. The retention of fine disbursement of interstitials is critical to the corrosion resistance of the zirconium. Subsequent heating near or over the beta transus will redistribute the intermetallics, with a resulting loss in corrosion resistance and mechanical properties.

The composition and physical properties of zirconium alloys are shown in Table 1.

Brazing Filler Metal Selection. The narrow commercial use of zirconium has limited the development of brazing alloys. Basic joints have utilized alloys of copper, silver, and phosphorus, but such a joint would not have the corrosion resistance of the parent material. The major effort in the area of zirconium brazing has centered on nuclear applications that require high strength, high temperature, and hot-water corrosion resistance.

The wetting of zirconium alloys is not a particular problem, but the brazing filler metal selection is critical, because of the demanding corrosion resistance of most zirconium applications, and testing in the proposed use environment is compulsory.

Brazing filler alloys that have demonstrated acceptable pressurized hot-water corrosion properties after 1450 h at 360 °C (680 °F) include Zr-5Be, Cu-20Pd-3In, and Ni-20Pd-10Si. Zircaloy brazed with Zr-50Ag or Zr-29Mn filler alloy has shown acceptable corrosion rates and mechanical properties after testing in pressurized hot water at 315 °C (600 °F) for nearly 3000 h.

Brazed zirconium components are predominantly used in nuclear applications, particularly fuel rods. One example is the brazing of Zr-1.5Sn alloy tube cap assemblies using Zr-5Be alloy. Another example was the brazing of Zr-1.5Sn alloy to molybdenum using a 48Ti-48Zr-4Be brazing filler metal. The 48Ti-48Zr-4Be alloy has also been effective in joining zirconium to alumina. Note that the beryllium content of these fillers limits their use to facilities that have the proper beryllium control procedures.

The filler metals shown in Table 3 for use with titanium are acceptable for use with zirconium.

Beryllium Alloys

Virtually all beryllium mill forms are based on powder metallurgy techniques. Ingot sheet from cast material has limited availability. Large blocks of beryllium that approach 1270 mm (50 in.) in diameter and 2995 kg (6600 lb) are produced by the vacuum hot pressing of beryllium powder. Because the beryllium is very active, the powder is oxidized. The resulting pressing contains a significant oxide content. It is partially through the control of this oxide that the final properties of the beryllium are achieved. Them are structural, optical, and instrument grades of beryllium to permit design optimization. Other forms of beryllium sheet, plate, or extrusions are fabricated from the block material.

Beryllium has a low density, 2.7 g/cm^3, and a high modulus of elasticity, 295 GPa (43×10^6 psi). It is this high specific stiffness that compels the use of beryllium for aerospace applications. Nuclear applications provide another large market for beryllium.

The design of beryllium hardware is not without problems. The high stiffness and low ductility values of the material may result in failures if proper design requirements are not followed. The elongation values of beryllium mill products range from 1 to 3% for block forms, 5 to 10% for plate and extrusions, and 20% for sheet. These elongations are for the longitudinal direction; wrought beryllium is very anisotropic. Transverse properties are significantly lower. Thus, the uniaxial tensile properties will not indicate how beryllium will perform under complex stress conditions.

The aluminum-beryllium alloy system has had considerable attention in an effort to increase ductility with some loss in stiffness. Lockalloy, a 38% aluminum alloy originally developed around 1961, has a ductility and weldability advantage, compared with unalloyed beryllium, but its applications were limited because of its lower stiffness. The Al-Be system is again becoming available in several compositions. Because the two metals are mutually immiscible, the alloys are processed to produce a fine structure of aluminum and beryllium. Beryllium alloy compositions and physical properties are shown in Table 1.

Although beryllium has a hexagonal close-packed structure, it is unlike titanium and zirconium in that it does not undergo a phase change at elevated temperatures. Wrought products must be worked hot, at temperatures from 800 to 1000 °C (1470 to 1830 °F). The temperatures used during processing are adjusted to prevent recrystallization, and the strength of these products rises significantly with increasing warm work.

The annealing temperature varies with the amount of work, but is reported to be between 725 and 900 °C (1340 and 1650 °F). The mechanical properties are reduced by recrystallization and grain coarsening if the material is heated too much, but there are no thermal strengthening treatments for beryllium.

The brazing of beryllium is well established, partly because of the difficulties of other joining methods. Because of a brittle, coarse-grained structure, the properties of welded joints have low strength and ductility, and an increase in the DBTT. Although mechanical fasteners, such as bolts and rivets, are used, careful design is required to prevent cracking during assembly and use. The Al-Be alloys are ductile and can be welded.

Brazing Filler Metal Selection. Brazing filler alloys for beryllium are primarily selected from the zinc, aluminum, and silver alloys. The advantages of zinc are that no reactions with the beryllium occur, the brazing temperatures are low to prevent recrystallization, and adequate strengths are obtained. Strength decreases rapidly with elevated temperatures. Aluminum systems are the most widely used brazing filler alloys. They have the same advantages as zinc, but higher strength values at both room temperature and elevated temperatures. The silver systems provide higher joint strengths at elevated temperatures, but at some sacrifice to parent metal strength because of recrystallization.

The wettability of beryllium with any brazing filler metal is difficult, and the preplacement of aluminum fillers is recommended. With close-tolerance joints, capillary flow is possible with the silver alloys, but not always predictable. Although fluxless brazing is possible, the use of proprietary fluxes is common. With flux, joint tolerances are less critical, and wetting of the base metal is more predictable. However, the potential corrosion problems from residual flux must be considered in the design and postbraze cleaning of the brazed component.

Aluminum-silicon alloys are the most common aluminum-base filler metals, although pure aluminum is also recommended. Shear strength values that exceed 62 MPa (9 ksi) have been reported, and useful strengths are possible at 175 °C (345 °F). Silver brazing was an early development in the beryllium joining process and is still recommended for joints that require high-temperature strength. The 60Ag-30Cu-10Sn, Ag-7Cu-0.2Li and pure silver have both been used with success. The shear strength of silver-brazed beryllium strongly depends on the braze cycle, and strength values between 140 and 275 MPa (20 and 40 ksi) have been reported. Lower brazing temperatures and short brazing cycles favor higher strengths. Copper, silver, and copper/silver filler metals were evaluated using low pressure (2.1 MPa, or 300 psi) and extended braze cycles (up to 8 h) to evaluate joints for high-temperature operation. With this braze/ diffusion approach, shear-strength values of 140 MPa (20 ksi) were obtained in the temperature range from 540 to 650 °C (1000 to 1200 °F).

The brazing of aluminum-beryllium alloys would be similar to that of beryllium, but the brazing filler selection must be based on the melting point of the aluminum-base material and would be limited to the zinc and aluminum-silicon alloys.

Health and safety rules must be thoroughly understood before conducting brazing operations that involve beryllium. Breathing of beryllium oxide has caused a loss of lung capacity in a manner similar to coal dust. Beryllium particles that enter open wounds have caused chronic ulcerations. Any beryllium process that can generate dust must be controlled. In machining operations, this is usually accomplished by using vacuum cleaners with output filtration that precludes the exhaust of any beryllium particles. Any etching, fluxing, or other treatments that may leave a beryllium dust residue after drying also must be controlled. Approved procedures must be used to dispose of chips, scrap material, and solutions.

Beryllium handling guidelines can be found in the Code of Federal Regulations, Title 29, Part 1910. The Office of Safety and Health Administration (OSHA) has established limits of atmospheric contamination at 2 μg/m^3 averaged over an 8 h day. The Environmental Protection Agency (EPA) has established 10 g/day as the maximum beryllium emission into the atmosphere. The time and cost of meeting safety requirements may be prohibitive for random beryllium brazing operations.

The handling of bulk beryllium poses minimal safety issues. Elevated-temperature heating below 780 °C (1435 °F) does not produce volatile or nonadherent oxides. Brazing can be accomplished with standard equipment if the surfaces to be brazed are plated to permit wetting of the filler metals and if no flux is used. Either vacuum or inert gas atmospheres would be suitable for fluxless brazing, and no postprocess cleaning would be required. Although the preceding procedure should safely permit beryllium use under normal conditions, appropriate approvals must be obtained.

Brazing of Refractory and Reactive Metals

Roy I. Batista, TRW, Inc.

Brazing Procedures

The joint designs that are valid for brazing the more-common metals, such as stainless steel or copper, are equally suitable for the reactive and refractory metals. At brazing temperatures, joint clearances should be between 0.013 and 0.050 mm (0.0005 and 0.002 in.) when using fillers that have good capillary flow. As happens with other metals, increased porosity and lower strength result from excessive joint clearances. Filler metals that have minimal wettability with the base metal should be preplaced in the braze joint and, when possible, a small load should be applied to the joint to accommodate some of the gap at the brazing temperature.

Because of the low thermal expansion of this group of metals, the calculation of the gap at the brazing temperature is of particular concern when brazing dissimilar metals that have great differences in thermal expansion. In tubular joints, the high-expansion material should be on the outside, so that the joint is in compression upon cooling. If these joints must be thermally cycled, then thermal fatigue failure can be a likely result. Therefore, such applications should not be brazed. For example, shear specimens of Mo-50Re brazed to L-605 with 82Au-18Ni failed after 20 to 30 cycles after cycling to 700 °C (1290 °F).

Fixturing. Whenever possible, the design should incorporate provisions for self-fixturing. Fixturing becomes more difficult with increasing braze temperatures. Two techniques that could avoid the use of hard tooling are alignment pins and tack welding, if satisfactory equipment is available in the latter case.

The need to fixture refractory metal brazements is required, just as it is for the brazing of any metal. However, it is complicated for refractory metals, because of their low thermal expansion values. The expense of fabricating fixtures from refractory metals is probably prohibitive for small-volume jobs. At lower temperatures, steel tools with springs or flexures that are made from stainless steel, superalloys, or molybdenum can be used to control the position of the brazements. Higher-temperature brazing requires that the tools be made from refractory materials, metals, or ceramics.

Because very clean atmospheres are usually recommended for brazing the reactive and refractory metals, the fixture is likely to join to the brazement if stop-off materials are not provided. Oxide slurries are suitable, as are more-permanent solutions, such as ceramic pads or flame-sprayed coatings. Because of the reactivity of this group of metals, any stop-off materials should be high in purity and nonreactive with the brazement. Zirconium oxide may not be suitable in contact with zirconium alloys at high temperature.

Surface preparation techniques include prebraze cleaning, surface plating, or other measures that improve the brazeability of the design or accommodate less-than-ideal brazing conditions.

The direct brazing of reactive and refractory metal surfaces, as discussed in this article, is based on the assumption that the filler metals would wet these surfaces. Another approach is to plate the surfaces with a metal, such as nickel, and then to consider the braze as a nickel joint, rather than a refractory or reactive metal. Although the use of platings has value, it also has disadvantages, because platings can peel or blister and because they are not suitable for refractory metal brazed assemblies that must operate at high temperatures.

Beryllium forms the most refractory oxide in this group of metals and is the most difficult to braze. Techniques such as electroplating or vacuum depositing coatings such as nickel, copper, or titanium have been used to enhance wettability. When powdered brazing filler alloys are used, additions of titanium hydride have been made. During brazing, the hydride decomposes and plates the beryllium surfaces with titanium.

Because of the reactivity of the reactive and refractory metals, prebraze cleaning is more important than it is when iron-, nickel-, or copper-base alloys are used. The sequence below is one of many that can be used:

- *ALKALINE CLEAN,* WHICH CAN BE PRECEDED BY SOLVENT CLEANING, AS REQUIRED BY THE SURFACE CONTAMINANTS
- *WATER RINSE,* PREFERABLY WITH DEIONIZED WATER
- *ACID PICKLE.* FOR TITANIUM, ZIRCONIUM, AND BERYLLIUM, USE 30 TO 40% HNO_3 WITH 1 TO 5% HYDROFLUORIC ACID, FOR 1 TO 5 RAIN AT 60 °C (140 °F). FOR NIOBIUM, TANTALUM, TUNGSTEN, AND MOLYBDENUM, USE 40 TO 70% HNO_3 WITH 10 TO 30% HYDROFLUORIC ACID. THE HIGHER CONCENTRATIONS ARE MORE SUITABLE FOR TANTALUM AND NIOBIUM; USE AT 55 TO 65 °C (130 TO 150 °F).
- *HOT-WATER RINSE,* PREFERABLY WITH DEIONIZED WATER
- *DRY AND PACKAGE* TO PREVENT CONTAMINATION BEFORE BRAZING

Parts should be brazed as soon as possible after brazing to minimize oxide formation. All cleaning of titanium must exclude any chlorinated solutions and menthol alcohol. Titanium is prone to stress-corrosion cracking when exposed to these solutions.

Brazing of Refractory and Reactive Metals

Roy I. Batista, TRW, Inc.

Brazing Equipment

Strict control of the brazing atmosphere is mandatory for the successful brazing of the reactive and refractory metals. The use of marginal atmospheres will, at a minimum, reduce the quality of reactive metal brazements and can potentially destroy the refractory metals, because of oxidation.

A cold-wall, high-vacuum furnace is highly recommended for the brazing of reactive and refractory metals. At brazing temperatures, a vacuum of 0.013 Pa (10^{-4} torr) can be considered a minimum level, and the leak rate of the furnace should be less than 1 μm/h (40 μin./h). Refractory metal heating elements are preferred to those of carbon in order to minimize potential contamination. The use of a cryogenic cold trap in diffusion-pumped equipment is also highly desirable.

The potential contamination problems associated with the use of hot-wall vacuum furnaces make their use questionable for high-quality work.

Induction brazing is a preferred process when the size and design of the hardware permits its use. However, the control of the brazing atmosphere is still critical. Vacuum induction, at a level of 0.013 Pa (10^{-4} torr), is acceptable, as are inert gases such as argon. Bottled inert gases are often not as clean as labeled, and a gettering system is recommended to minimize residual water and oxygen. A stainless steel tube filled with zirconium or titanium turnings and operating at about 800 °C (1470 °F) will reduce water to a few parts per million. A liquid-argon source minimizes gas contamination problems.

Tungsten and molybdenum can also be furnace or induction brazed in a hydrogen atmosphere, if the filler metal is compatible with hydrogen. The hydrogen dew point should be below -50 °C (-60 °F).

It is recommended, when furnace or induction brazing, that the equipment be cycled to a temperature higher than the braze temperature before actual brazements are made. This technique will completely heat the equipment and minimize outgassing problems during the subsequent brazing operation. If the recommended brazing equipment is not available, then successful brazing may be possible by protecting the brazement with metal foil, such as tantalum, which reacts with water vapor and oxygen and thereby offers some protection to the brazement. To check this approach or to check any brazing equipment, a piece of tantalum foil can be processed through the braze cycle and then checked for foil ductility. A more-sensitive test would be to metallographically examine the foil for contamination or to run a carbon and gas analysis on the sample and compare it to the starting material.

Although torch brazing with fluxes has been used to join reactive and refractory metals, the likely contamination of the base metal and future corrosion problems associated with residual flux should discourage this approach. Heliarc torch brazing has been performed within a dry box using inert gas atmospheres that are purified, as for induction brazing. The heating of components in an electron-beam welder also can be useful when permitted by the component configuration.

The use of flux, with one exception, is not recommended. The exception is the brazing of beryllium, especially with the zinc- and aluminum-base filler metals. The beryllium oxide will prohibit wetting by these fillers without the use of flux.

Whatever braze equipment is used, the lowest temperature for the shortest amount of time that will achieve the desired joint properties should be used as a technique to reduce contamination and erosion in the base metal.

References

1. M.M. MCDONALD, *ET AL.*, THE WETTABILITY OF BRAZING ALLOYS ON MOLYBDENUM AND TZM, *WELD. J. SUPPL.,* OCT 1988
2. G.M. SLAUGHTER *ET AL.*, WELDING AND BRAZING OF ADVANCED REFRACTORY ALLOYS, *SPACE SHUTTLE MATERIALS PROCEEDINGS,* SAMPE, OCT 1971
3. L.B. LUNBERG *ET AL.*, BRAZING OF MOLYBDENUM AND TUNGSTEN FOR HIGH-TEMPERATURE SERVICE, *WELD. RES. SUPPL.,* OCT 1978
4. R.I. BATISTA, TRW, INC., UNPUBLISHED NOTES
5. C.W. HAYNES, "DEVELOPMENT OF LOW TEMPERATURE BRAZING OF TUNGSTEN FOR HIGH TEMPERATURE SERVICE," REPORT R>D>R 1249-1/DTIC AD 260580, DEFENSE TECHNOLOGY INFORMATION CENTER, MAY 1962
6. T.J. MOORE *ET AL.*, "BRAZING TUNGSTEN FOR HIGH TEMPERATURE NUCLEAR SERVICE IN HYDROGEN," REPORT TN D-3944, NASA, MAY 1967

Selected References

- N.C. COLE AND G.M. SLAUGHTER, DEVELOPMENT OF BRAZING FILLER METALS FOR MOLYBDENUM, *NUCL. TECHNOL.,* VOL 26, JUNE 1975
- G.D. CREMER, J.R. WOODWARD, AND L.A. GRANT, BERYLLIUM BRAZING TECHNOLOGY, *SAE TRANS.,* VOL 76, SECT. 24, SAE REPRINT 670805, REPRINTED 1985
- A.H. DONOVAN, "BRAZE ALLOY DEVELOPMENT FOR ZIRCALOY," REPORT WHC-SA-0317-FP, WESTINGHOUSE HANFORD COMPANY, MAY 1988
- A.H. FREEDMAN *ET AL.*, "TANTALUM AND MOLYBDENUM BRAZING TECHNIQUES," TECHNICAL REPORT ML-TDR-64-270 PART II, AIR FORCE MATERIALS LABORATORY, SEPT 1964
- M. GARIN AND L.A. GRANT, "DESIGN AND FABRICATION OF BRAZED BERYLLIUM ASSEMBLIES," PAPER 83-0868-CP, 24TH ANNUAL AIAA SDM CONFERENCE, MAY 1983
- W.E. LUNBLAD, "INVESTIGATION OF WELDING AND BRAZING OF MOLYBDENUM AND TZM ALLOY TUBES," REPORT SRI-MME-91-097-7244, SOUTHERN RESEARCH INSTITUTE, FEB 1991

- A. ONZAWA *ET AL.*, BRAZING OF TITANIUM USING LOW MELTING POINT TI-BASED FILLER METALS, *WELD. RES. SUPPL.*, DEC 1990
- L.H. STONE *ET AL.*, RECRYSTALLIZATION BEHAVIOR AND BRAZING OF THE TZM MOLYBDENUM ALLOY, *WELD. RES. SUPPL.*, JULY 1967
- Y. SUEZAWA, BRAZING CHARACTERISTICS OF ZIRCONIUM AND SILVER FILLER METALS TO 304L STAINLESS SINTERED STEEL, *BRAZING SOLDERING*, NO. 6, SPRING 1984
- T. TAKEMOTO *ET AL.*, STRENGTH OF TITANIUM JOINTS BRAZED WITH ALUMINUM FILLER METALS, *TRANS. JWRI*, VOL 19, WELDING RESEARCH INSTITUTE OF OSAKA UNIVERSITY, NOV 1990
- T. TAKEMOTO AND I. OKAMOTO, INTERMETALLIC COMPOUNDS FORMED DURING BRAZING OF TITANIUM WITH ALUMINUM FILLER METALS, *J. MATER. SCI.*, APRIL 1988
- M.L. TORTI AND R.W. DOUGLASS, "BRAZING OF TANTALUM AND ITS ALLOYS," REFRACTORY METALS AND ALLOYS III APPLIED PROCESS CONFERENCE, AIME, 1963
- W.R. YOUNG, "ALLOY SYSTEMS FOR BRAZING OF COLUMBIUM AND TUNGSTEN," REFRACTORY METALS AND ALLOYS III APPLIED PROCESS CONFERENCE, AIME, 1963

Brazing of Ceramic and Ceramic-to-Metal Joints

A.J. Moorhead and W.H. Elliott, Jr., Oak Ridge National Laboratory H.-E. Kim, Seoul National University

---

Introduction

THIS ARTICLE is intended to assist in the development of procedures for the brazing of ceramic-to-ceramic or ceramic-to-metal joints for service under one or more of the following conditions: elevated temperatures, mechanical or thermal stresses, or corrosive atmospheres. Also described in this article are those factors that must be considered when preparing a procedure for the brazing of graphitic materials. However, because the interest in joining graphite is not nearly as intense at this time as the interest in joining ceramics, the former topic will be given less attention.

There is worldwide interest in using ceramic materials for applications not previously considered, that is, engineering applications. In these proposed or actual applications, ceramic-to-ceramic or ceramic-to-metal joints must be compatible with elevated temperatures and environments that are either oxidizing or corrosive, and they must have significant mechanical properties, including strength and fracture toughness. In addition, the joints must have demonstrable thermal stability at the temperatures of interest.

Material scientists and engineers now have an increased awareness of the critical nature of joints, their microstructural complexity, and the possibilities for joining a given set of components for a specific application. It should be noted that the uses for ceramic materials are changing rapidly, with a major emphasis on structural applications. Furthermore, these materials are becoming more complex, in terms of being strengthened and toughened by transformation processes, the addition of other ceramic or metallic materials, or the development of performance-enhancing microstructures made possible by processing. The successful use of advanced ceramic materials requires the development of equally advanced joining materials and technology.

Although advanced or fine ceramics also include the electronic ceramics, such as those used as substrates and encapsulants, ionic conductors, piezoelectric devices, and high-critical-temperature superconductors, this category of materials and the related joining technology is outside the scope of this article.

A primary impetus behind the emphasis on structural ceramics is the desire for improved efficiency and increased component life in various internal-combustion engines. For example, recent reviews of current research on ceramic joining technology in Europe, the United States, and Japan (Ref 1, 2, 3) strongly emphasize the research and development efforts that are aimed at utilizing structural ceramics in advanced heat engines.

Ceramic components are either being considered or actually used in a wide range of applications in experimental and commercial automotive applications. These applications include:

- VARIOUS WEAR-REDUCING PARTS IN CONVENTIONAL GASOLINE OR DIESEL ENGINES (REF 4)
- IN-CYLINDER COMPONENTS FOR LOW-HEAT-REJECTION DIESEL ENGINES (REF 1)
- SMALL CERAMIC TURBINE ROTORS THAT CAN REACH 900 °C (1650 °F) AND ARE SUBJECT TO A COMBINED CENTRIFUGAL AND THERMAL STRESS OF 200 MPA (29 KSI) IN TURBOCHARGERS FOR CONVENTIONAL GASOLINE ENGINES (REF 5) OR LARGER ROTORS THAT ARE EXPOSED TO EVEN MORE SEVERE TEMPERATURE AND STRESS CONDITIONS IN RELATIVELY SMALL, HIGH-EFFICIENCY, ADVANCED GAS TURBINE ENGINES (REF 6, 7, 8)

Many of these applications require ceramic-to-ceramic or ceramic-to-metal joints.

A second major impetus, at least in the United States, is the development of ceramic heat exchangers to recover the waste heat that is traditionally lost in the high-temperature, corrosive exhausts of industrial furnaces (Ref 9, 10). It should be obvious that much of the present interest in using ceramics for high-temperature structural applications is based on the worldwide desire to reduce the use of fossil fuels through energy conservation.

Carbon and graphite are widely used as electrodes in metallurgical applications and as moderator materials in nuclear applications. Specialized uses include rocket nozzles, guide vanes, nose cones, electric motor brushes and switches, bushings and bearings, high-temperature heat exchangers, and plumbing, as well as heart valves, synthetic teeth posts, air-frame composites, and high-performance brake linings. An antenna used for heating the plasma in a fusion reactor experiment, a recent example of a successful application, is discussed later in this article.

It is important to note that in the very early stages of brazing procedure development, there must be a clear understanding of the type of ceramic or graphite involved, how it differs from the more-familiar metallic materials, and whether the ceramic or graphite is to be attached to the same material, a similar material, or a metal. Such an understanding will facilitate the selection of a brazing filler material, the selection and use of a transition material in ceramic-to-metal joints, and the selection of a thermal cycle that is compatible with the heat-treatment response of the materials to be joined. Therefore, the following two sections describe ceramics and graphites, respectively, in terms of how they are made and how they differ from metals. These sections are followed by a discussion of procedure development and the available brazing processes, as well as a brief description of factors that are important to ceramic-to-metal joints.

References

1. D.A. PARKER, TECHNOLOGICAL REQUIREMENTS FOR DESIGNING INTERFACES FOR MECHANICAL AND THERMAL APPLICATIONS, *DESIGNING INTERFACES FOR TECHNOLOGICAL APPLICATIONS: CERAMIC-CERAMIC, CERAMIC-METAL JOINING,* S.D. PETEVES, ED., ELSEVIER, 1989, P 3-32
2. R.E. LOEHMAN, CERAMIC JOINING IN THE U.S., *DESIGNING INTERFACES FOR TECHNOLOGICAL APPLICATIONS: CERAMIC-CERAMIC, CERAMIC-METAL JOINING,* S.D PETEVES, ED., ELSEVIER, 1989, P 235-245
3. T. SUGA, CURRENT RESEARCH AND FUTURE OUTLOOK IN JAPAN, *DESIGNING INTERFACES FOR TECHNOLOGICAL APPLICATIONS: CERAMIC-CERAMIC, CERAMIC-METAL JOINING,* S.D. PETEVES, ED., ELSEVIER, 1989, P 247-263
4. L.M. SHEPPARD, AUTOMOTIVE PERFORMANCE ACCELERATES WITH CERAMICS, *AM. CERAM. SOC. BULL.,* VOL 69 (NO. 6), 1990, P 1012-1021
5. K. MATOBA, K. KATAYAMA, M. KAWAMURA, AND T. MIZUNO, THE DEVELOPMENT OF SECOND GENERATION CERAMIC TURBOCHARGER ROTOR--FURTHER IMPROVEMENTS IN RELIABILITY, *AUTOMOTIVE CERAMICS-RECENT DEVELOPMENTS,* SAE P-207, SAE, 1988, P 37-46
6. S. ISHIWATA, AUTOMOTIVE CERAMIC GAS TURBINE R & D IN JAPAN, *PROC. 27TH AUTOMOTIVE TECHNOLOGY DEVELOPMENT CONTRACTORS COORDINATION MEETING,* SAE P-

230, SAE, 1990, P 287-292

7. H.E. HELMS, P.J. HALEY, L.E. GROSECLOSE, S.J. HILPISCH, AND A.H. BELL, ADVANCED TURBINE TECHNOLOGY APPLICATIONS PROJECT, *PROC. 27TH AUTOMOTIVE TECHNOLOGY DEVELOPMENT CONTRACTORS COORDINATION MEETING,* SAE P-230, SAE, 1990, P 293-301
8. J.R. KIDWELL, L.J. LINDBERG, AND R.E. MOREY, ATTAP/AGT101--YEARS PROGRESS IN CERAMIC TECHNOLOGY DEVELOPMENT, *PROC. 27TH AUTOMOTIVE TECHNOLOGY DEVELOPMENT CONTRACTORS COORDINATION MEETING,* SAE P-230, SAE, 1990, P 327-334
9. B.D. FOSTER AND J.B. PATTON, ED., *CERAMICS IN HEAT EXCHANGERS,* VOL 14, *PROC. THREE SYMPOSIA ON HEAT EXCHANGERS,* AMERICAN CERAMIC SOCIETY, 1984-1985
10. A.J. HAYES, W.W. LIANG, S.L. RICHLEN, AND E.S. TABB, ED., *INDUSTRIAL HEAT EXCHANGERS, PROC. 1985 EXPOSITION AND SYMPOSIUM ON INDUSTRIAL HEAT EXCHANGER TECHNOLOGY,* AMERICAN SOCIETY FOR METALS, 1985

Brazing of Ceramic and Ceramic-to-Metal Joints

A.J. Moorhead and W.H. Elliott, Jr., Oak Ridge National Laboratory H.-E. Kim, Seoul National University

Ceramic Materials

Structural ceramics can be categorized as either monolithic materials or ceramic-matrix composites. The monolithic materials include the aluminas, mullites, silicon carbides, silicon nitrides, and Sialons. The ceramic-matrix composites can be subcategorized as those containing a random dispersion of particles or whiskers that am intended to improve fracture toughness and those containing semicontinuous fibers that are intended to carry the load during matrix-microcracking in a service-overload condition (Ref 11). Examples of oxide-matrix composites toughened by the dispersion of a second phase or material include the transformation-toughened zirconias (TTZ) (Ref 12), zirconia-toughened aluminas (Ref 13), and silicon carbide whisker-reinforced alumina and mullite (Ref 14). Examples of continuous fiber-reinforced composites include silicon carbide fiber-reinforced glass-ceramic composites (Ref 15); glass or glass-ceramic matrices reinforced with silicon carbide, alumina, or other fibers (Ref 16); and continuous and discontinuous fiber-reinforced oxide-matrix composites in which the matrix is formed from the hydrothermal decomposition of metal-organic precursors (Ref 17).

Materials that develop a microstructure that mimics the geometry and function of second-phase additions in ceramic-matrix composites represent a recent innovation. A principal example of these so-called self-reinforced ceramics is silicon nitride, in which a beneficial acicular microstructure is developed through the combined effects of chemical composition and thermal treatment (Ref 18, 19).

Generally, structural ceramics have lower coefficients of thermal expansion (CTE) than those of metals. For example, the CTE of silicon carbide ranges from 4 to 5 × 10^{-6}/K and that of silicon nitride ranges from 3 to 3.5 × $l0^{-6}$/K. Of the metals, tungsten has the lowest CTE of 4.5 × 10^{-6}/K, but it is generally not considered to be a structural metal, because it is brittle and subject to oxidation at low temperatures. Although differences in CTE values can introduce damaging thermal stresses, even in ceramic-to-ceramic joints brazed with metals or glasses, there are significantly greater potential risks associated with ceramic-to-metal joints. Although a thin metal or glass layer at the joint may be able to yield or even fail in a noncatastrophic manner, thereby reducing damaging residual stresses, such reductions may be precluded by the material, size, and geometry of a metallic component.

The general statement that structural ceramics have lower CTEs than structural metals does not apply to the TTZ ceramics, which have CTE values that range from approximately 8 to 11 × 10^{-6}/K. Such relatively high thermal-expansion behavior is similar to that of structural metals, including some cast irons and ferritic and martensitic stainless steels, as well as low-expansion superalloys, such as Incoloy 909. A potentially close CTE match improves the likelihood of success when zirconia ceramics are attached to metals. The damaging residual stresses associated with other ceramic-metal combinations that have significant CTE differences can be avoided by either designing the joint to always be subjected to compressive stresses or introducing one or more interlayers or graded joints between the ceramic and the metal. Such techniques for accommodating CTE differences will be discussed below.

Another important consideration when dealing with ceramics is that they are brittle and, therefore, less tolerant of tensile loads, stress intensifiers, and thermal shock than are most metals. For example, even advanced ceramic-matrix composites have fracture toughness values ranging from only 8 MPa $\sqrt{m}$ (7.3 ksi $\sqrt{in}$) for silicon carbide whisker-reinforced alumina to 14 MPa $\sqrt{m}$ (12.7 ksi $\sqrt{in}$) for some TTZ. These values are less than those of cast aluminum alloys or cast irons. Finally, because the structural ceramics tend to have greater thermo-dynamic stability than metals, it is difficult to form in ceramic joints strong chemical bonds that enhance adhesion. As will be seen in the section "Processing" in this article, this stability can be overcome by adding active elements to brazing filler metals, or by coating the ceramic surface with an active element prior to brazing with a conventional filler method.

Processing. In general, three processing routes prevail in the fabrication of high-performance ceramics. In the sintering process, powder particles in a "green" body coalesce in order to effect densification. Sintering can be accomplished in some materials either with or without pressure. However, a totally solid-state process that does not use sintering aids is unusual. The role of the sintering aids varies from material to material. Generally, these aids promote densification by modifying the diffusional process. In some systems, the sintering aid remains as a solid throughout the process. However, in the case of the silicon nitrides and Sialons, material is transported through a liquid phase (not by surface diffusion) that is formed by the reaction of metal-oxide sintering aids and the silicon dioxide that coats each of the silicon nitride grains. In most cases, at least some residue of the sintering aid remains at the grain boundaries after the ceramic densities. This residue, which in many cases is a glass, can significantly affect the joining process.

Reaction sintering and reaction bonding are two special processes that have been developed for fabricating silicon carbide or silicon nitride ceramics without the large volume decrease associated with conventional sintering. Reaction-bonded silicon nitride is fabricated by the nitridation of a compact of silicon powder that has been formed into the desired shape by a process such as isostatic pressing, slip casting, or injection molding. Reaction-sintered silicon carbide is fabricated by exposing a compact of silicon carbide powder and carbon to molten or vapor-phase silicon, which reacts with the carbon to form *in situ* silicon carbide that bonds the original silicon carbide particles together. Reaction-sintered silicon nitride has superior creep resistance, because it is densified without sintering aids. However, it retains about 15% open porosity, which can wick a brazing filler material away from the joint area. Reaction-sintered silicon carbide can have significant amounts of unreacted metallic silicon that can affect joining by reacting with the active element (such as titanium) in the filler metal. Such reactions reduce the amount of active element available to react with the silicon carbide, thus reducing wetting tendencies and resulting in the formation of excessive amounts of embrittling titanium silicides in the joint.

Table 1 identities some structural ceramic materials and the properties that are important to the joining process. It should he noted that most of these materials are still in the development stage, having been manufactured only since the late 1970s. Any tabulation of properties serves only as a general guideline and not as engineering data. The mechanical properties of ceramics heavily depend on specimen preparation and size, as well as the type of test being conducted. These factors can also explain why there is so much variation among data compilations. It is important to note that the type of test and the specimen geometry used to characterize the strength or toughness of a brazement also can have a very significant effect on the measured property values. These effects are discussed more extensively in Ref 28.

TABLE 1 SELECTED PROPERTIES OF VARIOUS CLASSES OF STRUCTURAL CERAMICS

	FLEXURAL STRENGTH		**FRACTURE TOUGHNESS**		**COEFFICIENT OF THERMAL EXPANSION, 10^{-6}/K**	**THERMAL CONDUCTIVITY**		REF
	MPA	**KSI**	**MPA** $\sqrt{m}$	**KSI** $\sqrt{in}$		**W/M · K**	**BTU/FT · H · °F**	
ALUMINA (AL_2O_3)	200-600	30-90	3-4	2.7-3.6	7-9	15-40	8.7-23.2	**20, 21, 22, 23, 24**
PSZ (CAO, MGO)	600-1000	90-145	8-16	7.3-14.5	10	2.2	1.3	**25**
TZP (Y_2O_3, CEO_2)	800-2000	115-290	7-8	6.4-7.3	10	2.2	1.3	**21, 22, 23, 25**
ALUMINA-ZIRCONIA ($AL_2O_3 \cdot ZRO_2$)	1500-2000	220-290	8-13	7.3-11.8	. . .	. . .	. . .	**21, 22, 25, 26**
$3AL_2O_3 \cdot 2SIO_2$	200	30	2	1.8	5.3	3.5	2.0	**20, 21, 23, 27**
$2MGO_2 \cdot 2AL_2O_3 \cdot 5SIO_2$	150	20	1-2	0.9-	2	3	1.7	**20, 23**

				1.8				
SILICON CARBIDE WHISKER-REINFORCED ALUMINA $(AL_2O_3 \cdot SIC_W)$	500-800	70-115	6-10	5.5-9.1	6	30(A)	17(A)	**22, 89**
SILICON CARBIDE (SIC)	300-600	45-90	3-5	2.7-4.6	4.5	85(A)	49(A)	**22, 90**
SILICON NITRIDE (SI_3N)	**200-1200**	**30-175**	**4-10**	**3.6-9.1**	**3.6**	**17(A)**	**10(A)**	**22, 90**

(A) AT ROOM TEMPERATURE

It is also important to note that alumina, silicon carbide, and other materials that are often considered to be single-component materials are, in fact, families of materials that may respond very differently to the conditions imposed in a joining procedure. For example, there are a wide variety of high-alumina ceramics ranging in aluminum oxide content from 79 wt% (alumina-mullite) (Ref 29) to 99.9 wt% (Ref 30, 31). Because these materials readily densify without pressure, they are almost invariably fabricated by pressureless sintering. The aluminas not only vary widely in chemical composition (and, thus, in the amounts and composition of grain-boundary phases), but also in grain size and porosity. The various high-alumina ceramics can, therefore, be expected to respond very differently during joining or service. Differences during joining might include the formation of different interfacial reaction products, depending on chemical composition of the ceramic, or the wicking of filler metal away from a brazed interface, in the case of a porous ceramic or graphite body.

The materials that are sometimes referred to individually as silicon nitride or silicon carbide also are families of materials, in actuality. Again, the compatibility of these materials with a filler metal or glass and the responses of these materials to the brazing process may vary widely, because of differences in sintering aids, processing techniques, degree of densification, and microstructural features (such as the acicular grains in the self-reinforced silicon nitrides). Thus, as a brazing procedure is being developed, it is very important that the actual ceramic be used in all wetting, adherence, and strength or toughness studies.

References cited in this section

11. M.H. LEWIS, CERAMICS TO BE JOINED 10 YEARS FROM NOW, *DESIGNING INTERFACES FOR TECHNOLOGICAL APPLICATIONS: CERAMIC-CERAMIC, CERAMIC-METAL JOINING,* S.D. PETEVES, ED., ELSEVIER, 1989, P 271-290
12. P.F. BECHER AND M.K. FERBER, MECHANICAL BEHAVIOUR OF MGO-PARTIALLY STABILIZED ZRO_2 CERAMICS AT ELEVATED TEMPERATURES, *J. MATER. SCI.,* VOL 22, 1987, P 973-980
13. M. RÜHLE AND N. CLAUSSEN, TRANSFORMATION AND MICROCRACK TOUGHENING AS COMPLEMENTARY PROCESSES IN ZRO_2-TOUGHENED AL_2O_3, *J. AM. CERAM. SOC.,* VOL 69 (NO. 3), 1986, P 195-197
14. P.F. BECHER AND T.N. TIEGS, TEMPERATURE DEPENDENCE OF STRENGTHENING BY WHISKER REINFORCEMENT: SIC WHISKER-REINFORCED ALUMINA IN AIR, *ADV. CERAM. MATER.,* VOL 3 (NO. 2), 1988, P 148-153
15. K.M. PREWO, TENSION AND FLEXURAL STRENGTH OF SILICON CARBIDE FIBRE-REINFORCED GLASS CERAMICS, *J. MATER. SCI.,* VOL 21, 1986, P 3590-3600
16. J.J. BRENNAN, GLASS AND GLASS-CERAMIC MATRIX COMPOSITES, *FIBER REINFORCED CERAMIC COMPOSITES,* K.S. MAZDIYASNI, ED., NOYES PUBLICATIONS, 1990, P 222-259
17. T. MAH *ET AL.,* CERAMIC FIBER REINFORCED METAL-ORGANIC PRECURSOR MATRIX COMPOSITES, *FIBER REINFORCED CERAMIC COMPOSITES,* K.S. MAZDIYASNI, ED., NOYES PUBLICATIONS, 1990, P 278-310
18. C.W. LI AND J. YAMANIS, SUPER-TOUGH SILICON NITRIDE WITH *R*-CURVE BEHAVIOR, *CERAM. ENG. SCI. PROC.,* VOL 10 (NO. 7-8), 1989, P 632-645
19. T. KAWASHIMA, H. OKAMOTO, H. YAMAMOTO, AND A. KITAMURA, GRAIN SIZE DEPENDENCE OF THE FRACTURE TOUGHNESS OF SILICON NITRIDE CERAMICS, *J. CERAM.*

SOC. JAPAN, INT. ED., VOL 99 (NO. 4), 1991, P 1-4

20. W.D. KINGERY, H.K. BOWEN, AND D.R. UHLMANN, *INTRODUCTION TO CERAMICS,* WILEY, 1976
21. R.W. RICE, CERAMIC COMPOSITES: FUTURE NEEDS AND OPPORTUNITIES, *FIBER REINFORCED CERAMIC COMPOSITES,* K.S. MAZDIYASNI, ED., NOYES PUBLICATIONS, 1990, P 451-495
22. P.F. BECHER, MICROSTRUCTURAL DESIGN OF TOUGHENED CERAMICS, *J. AM. CERAM. SOC.,* VOL 74 (NO. 2), 1991, P 255-269
23. "MATERIAL PROPERTIES DATA SHEET," COORS CERAMICS COMPANY
24. P. CHANTIDUL, S.J. BENNISON, AND B.R. LAWN, ROLE OF GRAIN SIZE IN THE STRENGTH AND *R*-CURVE PROPERTIES OF ALUMINA, *J. AM. CERAM. SOC.,* VOL 78 (NO. 8), 1991, P 2419-2427
25. S. SOMIYA, N. YAMAMOTO, AND H.H. YANAGIDA, ED., *SCIENCE AND TECHNOLOGY OF ZIRCONIA III,* AMERICAN CERAMIC SOCIETY, 1988
26. D. SHIN, K.K. ORR, AND H. SCHUBERT, MICROSTRUCTURE-MECHANICAL PROPERTY RELATIONSHIPS IN HOT ISOSTATICALLY PRESSED ALUMINA AND ZIRCONIA-TOUGHENED ALUMINA, *J. AM. CERAM. SOC.,* VOL 73 (NO. 5), 1990, P 1181-1188
27. H. SCHNEIDER AND E. EBERHARD, THERMAL EXPANSION OF MULLITE, *J. AM. CERAM. SOC.,* VOL 73 (NO. 7), 1990, P 2073-2076
28. A.J. MOORHEAD AND H.-E. KIM, JOINING OXIDE CERAMICS, *CERAMICS AND GLASSES,* VOL 4, *ENGINEERED MATERIALS HANDBOOK,* ASM INTERNATIONAL, 1991, P 511-531
29. J.B. WACHTMAN, JR. AND R.A. HABER, ADVANCED CERAMICS INVOLVING ALUMINA, *ALUMINA CHEMICALS: SCIENCE AND TECHNOLOGY HANDBOOK,* L.D. HART, ED., AMERICAN CERAMIC SOCIETY, 1990, P 329-335
30. "MECHANICAL AND INDUSTRIAL CERAMICS," CAT/2T8505TKU/3000E, KYOCERA CORPORATION
31. "COORS CERAMICS--MATERIALS FOR TOUGH JOBS," BULLETIN NO. 953, COORS PORCELAIN COMPANY
89. P.H. MCCLUSKEY, R.K. WILLIAMS, R.S. GRAVES, AND T.N. TIEGS, THERMAL DIFFUSIVITY AND CONDUCTIVITY OF ALUMINA-SILICON CARBIDE COMPOSITES, *J. AM. CERAM. SOC.,* VOL 73 (NO. 2), 1990, P 461-464
90. J.B. WACHTMAN, JR., ED., *STRUCTURAL CERAMICS, TREATISE ON MATERIALS SCIENCE AND TECHNOLOGY,* VOL 29, ACADEMIC PRESS, 1989

Brazing of Ceramic and Ceramic-to-Metal Joints

A.J. Moorhead and W.H. Elliott, Jr., Oak Ridge National Laboratory H.-E. Kim, Seoul National University

Graphitic Materials

Carbon, graphite, and composites based on these materials are somewhat similar to structural ceramics and ceramic-matrix composites, but they are different enough to require a separate discussion. Although both material types have high melting points and relatively low CTE values, structural ceramics are compatible with oxidizing environments, whereas the carbon-based materials typically begin oxidizing at temperatures ranging from 400 to 650 °C (750 to 1200 °F), depending on the perfection of the carbon structure and its purity.

In addition, many structural ceramics have little or no open porosity, whereas many carbon materials have significant amounts of open porosity. Such porosity complicates the brazing process by wicking filler materials away from a joint. On the other hand, carbon materials have higher thermal conductivity values and, in tandem with their low CTE values, are much less sensitive to thermal shock. Finally, the carbon materials are also more creep resistant than structural ceramics.

Carbon materials vary widely in degree of crystallinity and crystal orientation and in the size, quantity, and distribution of porosity in the microstructure. These factors strongly depend on the precursor materials and on processing and, in turn, govern the physical and mechanical properties of these products. This potential for variation in properties can be advantageous, because properties that are significant from a brazing standpoint (such as CTE) can be readily modified by processing. For example, CTE values can be varied from essentially zero in the fiber direction in fiber-reinforced composites to 8×10^{-6}/K for isostatically molded isotropic graphites.

Carbons and graphites can be manufactured by several processes that yield materials with a wide range of crystalline perfection and properties. In the most widely used process, polycrystalline graphites are made from natural pitch sources or from cokes produced from the residue that remains from petroleum after all volatile materials have been removed. The coke product is broken up and then calcined at temperatures ranging from 900 to 1400 °C (1650 to 2550 °F) to further reduce the volatile content and create a filler coke that will not shrink excessively during subsequent heat treatments.

The calcined coke is crushed, milled, and sized through screens into various fractions. The size of the particles in the coarse fraction may exceed 1 mm (0.04 in.) in large bodies and may be as small as 0.1 mm (0.004 in.) in smaller blocks, depending on the desired properties. The shape and properties of the crushed coke particles primarily depend on the coke source. Some cokes naturally break up into highly anisotropic particles with platelike or needlelike shapes, whereas others produce rounder isotropic particles.

The size fractions are mixed according to the properties that are desired in the final material, and then blended with a hot coal-tar pitch. The pitch creates a plastic mix that can be shaped by extrusion or molding. The shape of particles is important, because forming processes tend to preferentially align the anisotropic particles. Particles that are aligned in this manner as the result of extrusion or molding processes produce bodies with highly anisotropic properties.

The formed body is typically baked in large gas-fired floor furnaces to pyrolize the binder pitch at temperatures ranging from 800 to 900 °C (1470 to 1650 °F). During baking, the binder experiences a weight loss of approximately 50 to 60% and a volume loss that is even greater. The effect is a reduction in overall density and subsequently increased porosity. Density can be increased by rebaking after impregnation with low-melting-point, high-viscosity pitches. Special impregnants, such as thermosetting resins or mixtures of resins and pitches, can be used to control porosity. Final graphitization of carbon bodies is achieved at temperatures ranging from 2300 to 3000 °C (4170 to 5430 °F) in an electric furnace similar to those used to produce silicon-carbide grains.

Carbon-carbon composites consist of carbon or graphite fibers in a carbon or graphite matrix. The fibers have been made into yarns that are then woven into either two-dimensional fabrics or three-dimensional bodies of the desired shape. These preforms are then combined with a matrix to form a composite by either of two processes: liquid impregnation or chemical-vapor infiltration (CVI) (Ref 32). In the former, a preform is impregnated with liquid phenolic resin or pitch, baked to reduce porosity, and graphitized. The process of impregnation under pressure, carbonization, and high-temperature heat treatment is typically repeated five or more times in order to produce a composite that has a higher density and a stronger matrix.

In the CVI process, gaseous hydrocarbons, such as methane, are used to deposit an internal carbon matrix within a carbon-fiber preform. Several specialized procedures have been developed to overcome a major problem with the CVI process, which involves achieving a uniform deposition of the carbon matrix throughout a thick preform.

The carbon-carbon composite bodies that result from either process generally have very high tensile strength values, as well as low CTE values, which provide excellent resistance to thermal shock or stresses. However, these materials tend to be anisotropic, and their mechanical strength is much lower in directions that are not reinforced with fibers, such as between plies in two-dimensional bodies. In addition, for applications in which oxidation resistance of the composite is a concern, the materials are given a primary oxygen barrier consisting of an external coating of a ceramic, such as silicon carbide or silicon nitride, and internal inhibitors, such as compounds of boron, titanium, and silicon. At high temperatures, the latter materials form glassy oxide phases within the composite, which seal any CTE mismatch-induced cracks that form in the coating and restrict passage of oxygen through the porous matrix.

Of course, such oxidation-protection methods can significantly affect the brazing of these materials. One consideration, for example, would be whether the external coating is applied before or after brazing, because the coatings are applied at temperatures from 1000 to 1400 °C (1830 to 2550 °F), depending on the process. The glass-forming inhibitors may also affect wetting and/or adherence of a brazing filler metal.

Reference cited in this section

32. R.F. DIEFENDORF, CONTINUOUS CARBON FIBER REINFORCED CARBON MATRIX COMPOSITES, *COMPOSITES,* VOL 1, *ENGINEERED MATERIALS HANDBOOK,* ASM INTERNATIONAL, 1987, P 911-914

Brazing of Ceramic and Ceramic-to-Metal Joints

A.J. Moorhead and W.H. Elliott, Jr., Oak Ridge National Laboratory H.-E. Kim, Seoul National University

---

Procedure Development

Although ceramic joining technology has been highly developed since the 1940s, most of the effort has been directed toward the development of materials and techniques for applications involving "low" temperatures and low structural stress requirements. The technology of joining ceramics for use at elevated temperatures, at high stress levels, and in dirty environments is less well developed. By the careful selection of location or through judicious design, a joint in such applications may not be required to withstand the full intensity of the various service conditions. However, it is highly desirable that a ceramic-to-ceramic or ceramic-to-metal joint be capable of operation at near-peak conditions.

Wetting and adherence are the principal requirements for brazing. However, most ceramic and graphitic materials are inherently difficult to wet using the more-common filler metals. Most filler metals simply ball up when melted in contact with a structural ceramic or graphite. This problem can be overcome by either coating the ceramic surface with a suitable metal layer prior to brazing (indirect brazing) or using specially formulated filler metals that wet and adhere directly to an untreated ceramic surface (direct brazing).

Indirect Brazing. In this technique, the ceramic is first coated in the joint area with a material that can be wetted by a filler metal that will not wet the untreated surface. Coating techniques include sputtering (Ref 33, 34, 35, 36), vapor plating (Ref 37, 38, 39), and thermal decomposition of a metal-containing compound (Ref 40). The coating is usually, but not always, a metal. For instance, Hammond and Slaughter (Ref 41) re-acted chromium vapor with the surface of a graphite body to form a coating of chromium carbide. For graphite coated in this manner and then brazed with pure copper, the shear strength at room temperature and at 700 °C (1290 °F) was 140 MPa (5.8 ksi). Failure occurred at the chromium carbide and carbon interface.

The most widely used brazing technique for joining alumina ceramics is the molybdenum-manganese (Mo-Mn) process, in which a slurry consisting of powders of molybdenum and/or molybdenum trioxide (or tungsten/tungsten tri-oxide), manganese or manganese dioxide, and various glass-forming compounds is applied to the ceramic in the form of a paint. The coated ceramic is fired in a wet hydrogen atmosphere at a temperature of approximately 1500 °C (2730 °F). During firing, glassy material from the ceramic migrates into the molybdenum (or tungsten) powder layer, bonding it to the ceramic surface. Because the process depends on this glass migration out of the ceramic, it is also dependent on such factors as amount and viscosity of the glassy phases, ratio of grain size in the ceramic to the pore size in the molybdenum powder, and processing temperature and atmosphere. The molybdenum-manganese process is a well-developed commercial process that was thoroughly reviewed in a series of papers by Twentyman (Ref 42, 43, 44).

An example of using physical vapor deposition for coating is described by Santella (Ref 39), who coated cold isostatic pressed and sintered silicon nitride with a 1.0 μm (0.04 μin.) thick layer of titanium using vacuum evaporation, and then brazed the silicon nitride to itself and metals with standard (nonactive) filler metals. The room-temperature flexural strengths of some of the silicon nitride-silicon nitride brazements were comparable to that of the silicon nitride itself (Fig. 1). By depositing the titanium using vapor plating, Santella produced strong joints and overcame a problem that arose in earlier work.

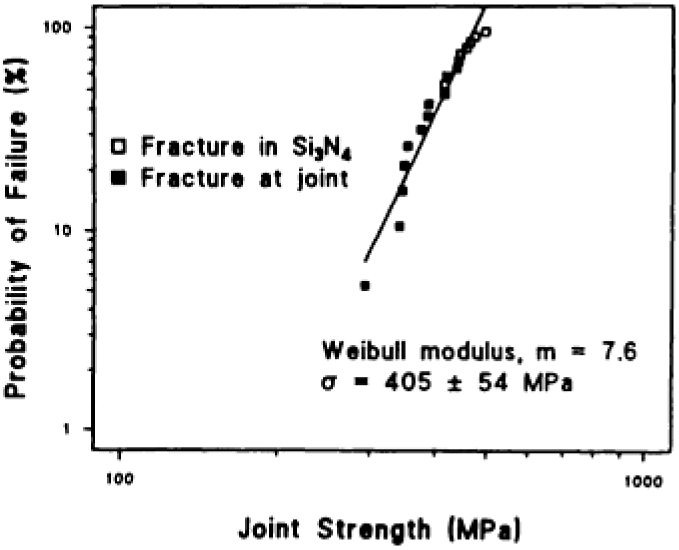

FIG. 1 FOUR-POINT FLEXURAL STRENGTH OF TITANIUM-VAPOR-COATED SILICON NITRIDE BRAZED TO ITSELF AT 1130 °C (2065 °F) WITH A AU-25NI-25PD, WT%, FILLER METAL. SOURCE: REF 39

In that earlier work (Ref 34), partially stabilized zirconia (PSZ) was coated by sputtering titanium and then brazed at 750 °C (1380 °F), either to nodular cast iron or to titanium with a commercial silver-copper-tin filler metal. Room-temperature shear strengths of both the PSZ/nodular cast iron and PSZ/titanium joints were about 140 MPa (20 ksi). However, at 400 °C (750 °F), the shear strength of the PSZ/ nodular cast iron brazement was only 28 MPa (4 ksi). Santella found that titanium coatings deposited by sputtering contain microporosity (apparently from trapped plasma gas) that coagulates during brazing and adversely affects high-temperature mechanical properties. Physical deposition of titanium by evaporation in a vacuum solved this problem.

Kohno *et al.* (Ref 45) coated silicon nitride by brazing at 1200 °C (2190 °F) with iron-nickel-chromium "paints" that contained silicon and carbon as melting-point depressants. They determined the adherence of the coatings by brazing the ceramic to itself at 1200 °C (2190 °F) with the paints themselves, and they achieved four-point bend strengths ranging from 125 to 200 MPa (18 to 29 ksi), depending on paint composition. However, the strength of similarly coated silicon nitride brazed to Kovar at 830 °C (1525 °F) with a silver-copper filler metal was only 50 MPa (7 ksi), apparently the result of damaging residual stresses.

Dunn *et al.* (Ref 46) studied the effects on mechanical properties of three silicon nitrides of various coatings applied by electron-beam evaporation. Polished bars were coated with hafnium, titanium, tantalum, or zirconium prior to being subjected to a thermal cycle simulating a brazing process (10 min at 980 °C, or 1795 °F, in vacuum) and then tested in four-point bending. The investigators found that zirconium and hafnium coatings did minimal damage to the silicon nitride. Titanium was found to be the coating most damaging to SNW-1000 (Si_3N_4-$13Y_2O_3$-$2Al_2O_3$), and it was even more severely damaging to AY6 (Si_3N_4-$6Y_2O_3$-$2Al_2O_3$). Interestingly, the effect of any of the coatings on PY6 material, which contains 6 wt% yttrium oxide, but no aluminum oxide, was minimal, which suggests a role for aluminum oxide in the property degradation of silicon nitride by active metal coatings. When silicon carbide was coated, heated, and tested in the same way, it was found that the hafnium coating was the most damaging.

Direct Brazing of Ceramics With Metallic Filler Metals. As mentioned above, most brazing filler metals will not wet structural ceramics and graphitic materials. Standard filler metals used to braze metals simply form a ball when melted in contact with a structural ceramic or graphite. This is due to the fact that the ceramic surface has very low surface energy, which means that there is no thermodynamic driving force for interface formation (wetting), as would be the case if a high surface energy surface were being wetted. However, for a number of years, it has been known that special alloying additions to conventional filler metals, such as the silver-copper eutectic composition, can promote wetting of and adherence to structural ceramics and graphite. The wetting of an untreated ceramic by a molten metal and the adhesion in such a system increases with the growing affinity of the metal constituents for the elements constituting the solid phase (Ref 47). Thus, the addition of chemically oxygen-active metals, such as titanium, zirconium, hafnium, aluminum, silicon, manganese, or lithium, should enhance both the wetting of and adherence to oxide ceramics, without the need for coating the ceramic surface. In the same way, the wetting of and adherence to nonoxide ceramics, such as silicon carbide and silicon nitride, are aided by filler metals containing elements that strongly interact with silicon, carbon, or nitrogen. Whether the ceramic is an oxide, carbide, or nitride, the active metal reacts with the ceramic surface, forming an interfacial layer that can be wetted by the bulk of the filler metal.

Titanium is the most extensively studied and widely used active element added to filler metals formulated to directly braze structural ceramics and graphite (Ref 48, 49, 50, 51, 52). In the case of oxide ceramics brazed with titanium-containing filler metals, the critical interfacial reaction product is either titanium monoxide or dititanium trioxide, with appreciably higher adhesion resulting when titanium monoxide is formed (Ref 53). Titanium additions to filler metals have also been shown to enhance wetting of and adherence to silicon carbide (Ref 54), silicon nitride (Ref 55), and Sialon (Ref 56). The reaction products of the wetting process in the latter cases are titanium silicides, titanium carbides, and titanium nitride.

In recent years, there has been significant research in the use of another active element, aluminum, as a brazing material in joining silicon nitride and silicon carbide. In the case of silicon nitride, the reaction product after brazing with aluminum at 800 °C (1470 °F) was identified as a 500 nm (20 μin.) thick dual layer consisting of a fine-particle layer of β'-Sialon attached to the silicon nitride and an amorphous aluminum oxide layer (Ref 57, 58). The interfacial reaction product observed when silicon carbide is brazed with aluminum filler metal has been identified as aluminum carbide (Ref 59).

Some of the commercially available filler metals that have been developed for direct brazing of structural ceramics are shown in Table 2. Note that numerous active filler metals have been developed and, in many cases, patented, but only relatively few are commercially available. Generally, these filler metals consist of two types: those that have a relatively high active element content and will wet and flow on structural ceramics and graphite, and those that represent a new generation of materials that have a lower active element content and will wet and strongly bond to these ceramics, but only flow onto the ceramic surface very slowly, if at all (Ref 61). Therefore, filler metals in the latter category are preplaced between the mating components, rather than relying on capillary attraction to draw them into the joint. The filler metals with a high active element content are manufactured in the form of composites (such as a titanium core inside a sheath of silver-copper eutectic), because such high titanium contents form brittle and, therefore, unfabricable structures.

TABLE 2 SELECTED COMMERCIALLY AVAILABLE ACTIVE BRAZING FILLER METALS

ALLOY SYSTEM	**CHEMICAL COMPOSITION, WT%**	**SOLIDUS TEMPERATURE**		**LIQUIDUS TEMPERATURE**		**BRAZING TEMPERATURE RANGE**		MANUFACTURER
		°C	**°F**	**°C**	**°F**	**°C**	**°F**	
AUNITI	96.4AU-3NI-06TI	1003	1837	1030	1886	1025-1030	1880-1886	(A)
CUALSITI	92.75CU-2AL-3SI-2.25TI	958	1756	1024	1875	1025-1050	1880-1920	(A)
AGCUALTI	92.75AG-5CU-1AL-1.25TI	860	1580	912	1674	900-950	1650-1740	(A)
AGCUTI	63AG-35.2CU-1.75TI	780	1436	815	1499	830-850	1525-1560	(A)
AGCUSNTI	63AG-34.2CU- 1SN-1.75TI	775	1427	806	1483	810-860	1490-1580	(A)
AGCUINTI	59AG-27.2CU-12.5IN-1.25TI	605	1121	715	1320	700-750	1290-1380	(A)
TICUNI	70TI-15CU-15NI	910	1670	960	1760	960-1000	1760-1830	(A)
TICUAG	68.8AG-26.7CU-4.5TI	830	1526	850	1560	810-900	1490-1650	(A)
AGTI	96AG-4TI	970	1778	970	1778	1000-1050	1830-1920	(B)
AGCUTI	64AG-34.5CU-1.5TI	770	1418	810	1490	850-950	1560-1740	(B)
AGCUTI	70.5AG-26.5CU-3TI	780	1436	805	1481	850-950	1560-1740	(B)

AGCULNTI	72.5AG-19.5CU-5IN-3TI	730	1346	760	1400	850-950	1560-1740	(B)
PBINTI	92PB-4IN-4TI	320	608	325	617	850-950	1560-1740	(B)
SNAGTI	86SN-10AG-4TI	221	430	300	572	850-950	1560-1740	(B)
AGCUINTI	61.5AG-24CU-14.5IN + (TI)	620	1148	710	1310	~845	~1555	(C)
AGCUNITI	71.5AG-28CU-0.5NI + (TI)	780	1436	795	1463	~927	~1700	(C)
AGCUNITI	**56AG-42CU-2IN + (TI)**	**770**	**1418**	**895**	**1643**	**954**	**~1750**	(C)

Source: Ref 60

(A)	GTE WESGO.
(B)	DEGUSSA AG.
(C)	LUCAS-MILHAUPT.

Direct Brazing With Nonmetallic Glasses. Numerous studies have been conducted to develop materials and techniques for joining structural ceramics, including alumina (Ref 62), PSZ (Ref 63), silicon nitride (Ref 64, 65, 66, 67, 68), and Sialon (Ref 69, 70) by brazing with nonmetallic glasses used as filler materials. The rationale for this approach is that many ceramics have amorphous phases remaining at the grain boundaries after the sintering process. Therefore, they would be expected to be wet by and compatible with similar glassy materials introduced at the joint. Glass viscosity and flow properties can be controlled over a wide range of temperatures. However, glasses have much lower toughness values than even monolithic ceramics, as well as low Young's modulus values, and are susceptible to stress corrosion (slow crack growth). As will be described below, some of the glasses, even those formulated to match the grain-boundary composition, have significantly different CTE values from that of the ceramic, which causes high residual stresses and cracking in the joint.

Generally, powdered glass in a slurry with a binder is applied to one or both of the mating components of a joint that is heated until the glass melts and reacts with the ceramic. In a variation of the process (Ref 66), the surface of one of the components was first coated with molten glass, and this glazed surface was ground to precisely control the thickness of the joining material. In another variation (Ref 64), aluminum foil was placed in a silicon nitride joint, along with sources of alumina and silicon dioxide, such as powders or oxidation layers on the aluminum and silicon nitride. The joints so produced had three-point bending strengths ranging from 35 to 210 MPa (5 to 30 ksi) at both room temperature and 1200 °C (2190 °F). In contrast to diffusion welding with high pressures applied to the joint area, brazing with glasses and metals uses relatively low pressures ranging from that provided by the weight of one of the parts being joined to 14 MPa (2 ksi) (Ref 63). Typical atmospheres are air (when joining oxides) and nitrogen (when joining silicon nitride).

The most extensive studies on brazing with glasses have been those on silicon nitride. The glasses include those similar to that which is found at the grain boundaries of a hot-pressed silicon nitride ($55SiO_2$-$35MgO$-$10Al_2O_3$) (Ref 66, 68, 69) and a very different CaO-TiO_2-SiO_2 (CTS) glass ($35CaO$-$25TiO_2$-$40SiO_2$) (Ref 67). Baik (Ref 65) stated that there are three potential problems when joining silicon nitride by brazing with glasses:

- A REACTION BETWEEN THE GLASS AND SILICON NITRIDE FORMS ACICULAR SILICON OXYNITRIDE CRYSTALS THAT PREVENT THE FORMATION OF A MICROSTRUCTURE TYPICAL OF A GRAIN BOUNDARY
- CTE MISMATCHES CAUSE CRACKING IN THE GLASS
- POROSITY FORMS IN THE GLASS LAYER DURING BRAZING OR UPON SUBSEQUENT HEATING TO ELEVATED TEMPERATURES

Such behavior was observed by Johnson and Rowcliff (Ref 66), who reported excellent strength (maximum of 400 MPa, or 58 ksi, within the optimum bonding conditions of 30 to 60 min and 1575 to 1650 °C, or 2870 to 3000 °F) in joints brazed with the composition similar to that found in the grain boundaries. However, significant cracking and porosity occurred in the joints. The cracking was attributed to differences in CTE, whereas the porosity was thought to result either from the vaporization of silicon monoxide and magnesium or from the volume change associated with the formation of silicon oxynitride crystals.

Iwamoto (Ref 67) minimized the cracking problem by using a CTS glass that had a CTE value less than that of the silicon nitride and by mixing silicon nitride powder with the glass powder. At room temperature, the resulting joints had four-point bending strengths ranging from 235 MPa (34 ksi) with no silicon nitride powder in the glass to 272 MPa (44 ksi) with glass containing 30 wt% silicon nitride powder.

Baik (Ref 65) alleviated the porosity problem that he attributed to the decomposition of the silicon nitride to silicon and nitrogen by adding elemental silicon to a magnesium-silicon-oxynitride glass to increase the silicon activity and, thereby, retard the formation of nitrogen gas.

Neilson (Ref 69) and Coon (Ref 70) used glasses in the yttrium-silicon-aluminum-oxynitride system to bond a sintered Sialon at 1600 and 1725 °C (2910 and 3140 °F) under an applied load of 0.63 N (0.14 lbf) and nitrogen gas pressures of 138 and 207 MPa (20 and 30 ksi), respectively. Joint fracture toughness values (approximately 4.5 MPa$\sqrt{m}$, or 4 ksi$\sqrt{\text{in.}}$) and four-point flexure strengths (approximately 450 MPa, or 65 ksi) at temperatures to 1000 °C (1830 °F) were equal to that of the Sialon. At temperatures greater than 1000 °C (1830 °F), the strength of the joints made at 1600 °C (2910 °F), which contained significant amounts of glass (glass-filled joint), decreased rapidly until at 1100 °C (2010 °F) it was only 25 MPa (4 ksi). However, the joints produced at 1725 °C (3140 °F) (closed joint with little retained glass) had strengths of 200 MPa (29 ksi) at 1250 °C (2280 °F).

In addition to the above-mentioned study on the joining of silicon nitride, a CTS glass was also investigated for joining a magnesia-PSZ ceramic (Ref 63). The use of this type of glass was chosen because the system has several eutectic compositions with melting temperatures below the 1400 to 1500 °C (2550 to 2730 °F) annealing range of the ceramic. Therefore, the brazing step could be combined with the annealing treatment of this material. An applied pressure of 14 MPa (2 ksi) was used to distribute the molten glass through the joint. Unfortunately, the differences in the CTE values of the materials resulted in extensive cracking through the glass and interface when the joint area was large (20 cm^2, or 3 in.^{2}). Based on the work of Iwamoto (Ref 67), the CTE of the interlayer was made closer to that of the PSZ by adding 30 wt% zirconium oxide to the CTS interlayer powder. Using a thinner layer of the modified powder and a longer pressing time at 1420 °C (2590 °F), joints that had average four-point bending strengths of 66 MPa (10 ksi) were produced.

Several glass compositions were investigated for joining alumina ceramics (Ref 62). This work differed from that of the preceding studies in that one of the goals of these researchers was to initiate and control crystallization of the glass in the joint to increase fracture toughness and resistance to stress corrosion. Zirconium oxide powder was added to a magnesium aluminum silicate glass as a nucleating agent. Unfortunately, spontaneous crystallization occurred at the glass-alumina interface, limiting the desired degree of control over crystallization and resulting in large voids at the interface. Accordingly, the fracture toughness of the joint was only 1.8 MPa$\sqrt{m}$ (1.6 ksi$\sqrt{in}$) versus a value of about 4 MPa$\sqrt{m}$ (3.6 ksi$\sqrt{in}$) for the alumina (no joint strength data reported). The most successful joints (fracture toughness of 2.3 MPa$\sqrt{m}$, or 2.1 ksi$\sqrt{in}$) resulted when the alumina was joined at 1500 °C (2730 °F) with a mixture of 50 wt% talc (hydrated magnesium silicate) and 50% aluminum oxide. However, these joints contained a substantial number of pores that degraded the mechanical integrity of the joints. The authors suggested that the pores were caused by the evolution of adsorbed gases from the alumina body and, therefore, would be difficult to eliminate.

Alternate Brazing Techniques. Although brazing is typically accomplished through melting of a metallic alloy or glass-forming compound placed in the joint area, there are some other viable techniques for generating a liquid at an interface to effect a brazement. The most common of these, generally referred to as "eutectic brazing," relies on the reaction at the interface of an insert material or of the metallic member of a ceramic-metal joint with either the ceramic, another metal, or even a gaseous component of the atmosphere. Such reaction is based on the presence of a low-melting region (typically a eutectic reaction) of the respective phase diagram, and hence comes the name of the process.

The most widely studied of the eutectic brazing processes, known as gas-metal eutectic brazing, is based on the formation of a liquid skin on the metallic member as a result of reaction with the atmosphere (Ref 71, 72, 73). According to Burgess *et al.* (Ref 71), the idea is to form a liquid skin on a metal surface, with the requirement that the liquid wets both the ceramic and metal components being joined and forms a strong bond on cooling. Ideal for use as the liquid skin material are eutectics formed by the reaction of the brazing atmosphere with the metal component itself. Such a eutectic exists between copper and oxygen at 0.39 wt% oxygen. The melting point of this eutectic, 1065 °C (1950 °F), is sufficiently below the melting point of pure copper, 1083 °C (1980 °F), so that by careful temperature control an oxide ceramic can be brazed to a copper component using this liquid. According to Yoshino (Ref 72) this technique is used in the electronics industry to manufacture copper-bonded alumina substrates for power semiconductor applications. Generally the copper is preoxidized prior to the joining operation, but Okamoto and coworkers (Ref 73)joined alumina to mild steel by simply using a copper filler metal foil in a slightly oxidizing atmosphere at 1200 °C (2190 °F). In addition to the copper-oxygen system, Burgess suggests that other beneficial metal-gas eutectics for joining might be iron-oxygen, nickel-oxygen, and cobalt-oxygen and even silver-phosphorus, nickel-phosphorus, copper-phosphorus, or molybdenum-silicon, and aluminum-silicon. The latter two types would be achieved by having small amounts of a phosphorus or silicon bearing compound (such as phosphine or silane) in the atmosphere.

A metal-metal eutectic was used by Naka and Okamoto (Ref 74) to braze an alumina (99.6 wt% Al_2O_3) to type 304 stainless steel. Titanium foil, which melts at 1670 °C (3040 °F), was placed between the two base materials and heated to 1400 °C (2550 °F) in vacuum of 0.13 mPa (1×10^{-6} mm Hg). The investigators suggest that the bond was made by the eutectic in the titanium-iron system at 68 at.% Ti at 1085 °C (1985 °F).

A recent and possibly very significant variation of the eutectic brazing process is one termed "partial transient liquid-phase" joining (Ref 75, 76). This process is based on the well-known transient liquid-phase bonding of nickel-base and cobalt-base superalloys (Ref 77). In this earlier work, a melting point depressant such as boron was added to a brazing filler metal and subsequently diffused into the base metals during brazing. Removal of the boron from the braze joint rids the brazement of embrittling compounds and results in a higher remelt temperature.

In the recent modifications of this process for joining of ceramics, three layers of metal are preplaced in the joint, with the outer two layers much thicker than the center layer. When the composite is heated, a thin film of low-melting-point liquid forms at the interfaces to make the braze, but the center layer is not totally consumed. Further holding at temperature results in diffusion of the liquid layer into the center layer, resulting in a more heat-resistant joint. The new process is termed "partial" transient liquid-phase joining because the transient liquid phase is formed only near the interface and the metal layer is not fully homogenized. One example of this process, cited by Iino (Ref 75), uses a thin layer of titanium on either side of a much thicker layer of nickel. According to the nickel-titanium phase diagram, when a nickel-titanium bilayer is heated at 1050 °C (1920 °F), a liquid phase occurs in the composition range of 60 to 75 wt% Ti. Because the titanium layer is much thinner than the nickel layer the liquid layer forms, the braze is made and then the liquid disappears as the titanium diffuses into the nickel and the local concentration of titanium falls below the liquidus line (below 60 wt%). If, by further diffusion, the titanium content falls below 30 wt%, the remelt temperature would be 1304 °C (2380 °F).

Shalz and coworkers (Ref 76) are considering a number of other possible material combinations for joining of ceramics such as alumina. Their work to date has included tri-layer compositions of copper/platinum/copper, copper/nickel/copper, copper/niobium/copper, and tin/ niobium/tin for joining alumina to alumina. The results of high-temperature postbonding anneals suggest that copper/platinum/copper and copper/ niobium/copper bonded alumina assemblies may be suitable for use at elevated temperatures in oxidizing and nonoxidizing environments, respectively.

Surface Preparation. Whether the filler material is a metal or glass, one of the initial decisions that must be made in brazing procedure development is the technique for preparing the ceramic surface. Although an as-fired surface is sometimes suitable for brazing, at least some machining is usually required to provide the contact surfaces necessary for joining. Because the machining of ceramic surfaces must generally be done by grinding with diamond-abrasive wheels, which is an inherently slow and costly process, surface preparation should be minimized. A surface preparation technique that is economically feasible, provides adequate surface finish and tolerances, and causes minimal damage to the ceramic is optimal. Unfortunately, ceramic machining itself is currently experiencing significant development, and it is not possible to recommend a set of grinding conditions (wheel grit, feeds, speeds, and such) that efficiently removes ceramic material while minimizing surface damage. However, enough information is available to state that the effect of a given preparation procedure on the ceramic is very material specific, even within a given family of ceramics.

For example, in working with an alumina ceramic containing 97.6% Al_2O_3, Mizuhara and Mally (Ref 50) found that grinding damage could be negated by retiring the ground parts at 1650 °C (3000 °F). Such a firing step, which apparently healed surface flaws induced during grinding, resulted in much higher peel strengths in Kovar/refired alumina joints brazed with silver-40 at.% copper plus titanium than in similar joints between Kovar and as-fired, ground, or ground/lapped alumina. However, a heat treatment of a higher-purity alumina (containing 99.5% Al_2O_3) after grinding did not have the same effect. Joints made with resintered samples of this alumina were much stronger than those made with ground material, but not as strong as those made with unground or ground/lapped alumina. It is assumed that the amount of glassy phase in the 99.5% alumina was insufficient to heal or blunt many of the grinding defects under the heat-treatment conditions used.

In the experience of the authors, a very different response to surface preparation has been observed among materials. For example, as shown in Fig. 2, it was found that the three-point bend strength of high-purity alumina (McDanel AD-998) ceramic-to-ceramic joints brazed in vacuum for 5 min at 800 °C (1470 °F) with an experimental filler metal were essentially the same, whether the joining surfaces were ground with a 180-grit diamond abrasive wheel or ground and then polished with 6 μm (240 μin.) and 3 μm (120 μin.) grit diamond pastes.

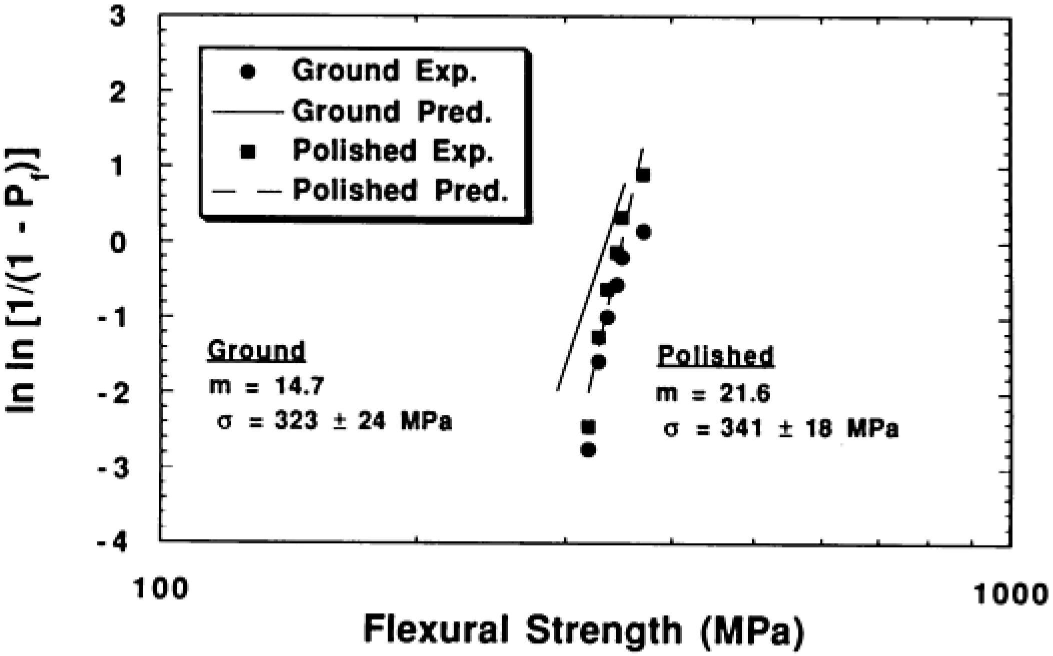

FIG. 2 EFFECT OF POLISHING INTERFACIAL AREA ON STRENGTH OF CU-41AG-3.5SN-7TI, AT.%, BRAZEMENT IN MCDANEL AD-998 ALUMINA

On the other hand, polishing the mating surfaces of coupons of silicon carbide whisker-rein-forced alumina prior to brazing led to significantly higher joint strengths, as shown in Fig. 3.

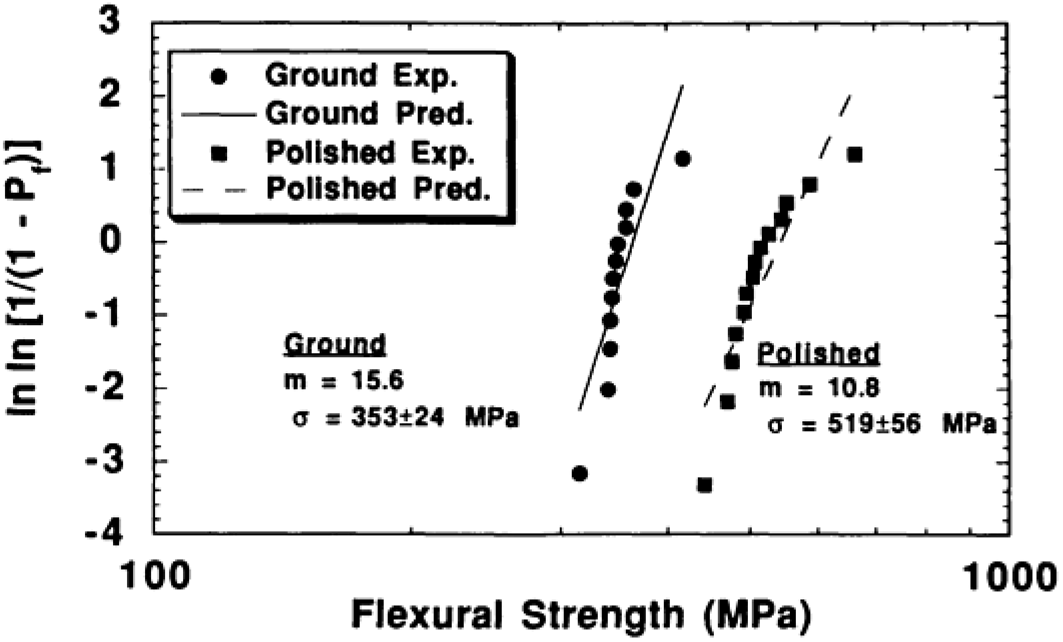

FIG. 3 EFFECT OF POLISHING INTERFACIAL AREA ON STRENGTH OF CU-41AG-3.5SN-7TI, AT.%, BRAZEMENT

Examination of typically prepared coupons of both materials by scanning electron microscopy revealed that the morphologies of the ground surfaces were very different. As shown in Fig. 4(a), the ground alumina surface exhibited predominantly intergranular fracture, although some transgranular fracture was also evident, such as the grain just left of center. However, the most notable features in the alumina microstructure were the many cracks that appear to be both intergranular and transgranular. The microstructure of the ground composite (Fig. 5a) exhibited the general striations typically of grinding, with linear regions of predominantly transgranular fracture separated by areas with a "smeared" appearance suggesting plastic flow. Although the concept of plastic flow in an inherently brittle ceramic may seem incongruous, this behavior has been widely observed in machined ceramic surfaces. Despite the fact that the grain size in the alumina was much larger than that of the silicon carbide whisker-reinforced alumina material, the polished microstructures were somewhat similar (Fig. 4b, 5b), with both materials showing evidence of porosity and grain pullouts.

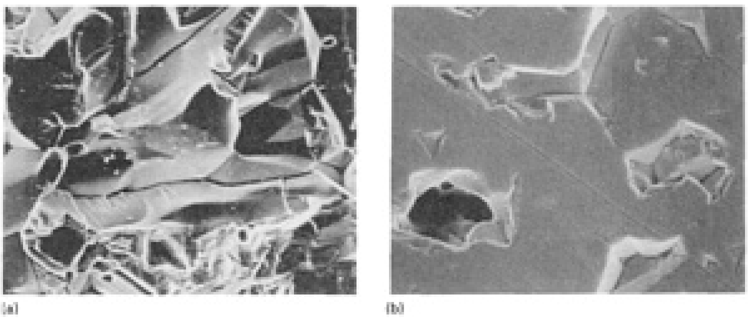

FIG. 4 SCANNING ELECTRON MICROGRAPHS OF SURFACE OF MCDANEL AD-998 ALUMINA AS PREPARED FOR JOINING. (A) GROUND, ROOT MEAN SQUARE (RMS) ROUGHNESS, R_Q, OF 0.63 μM (25 μIN.). (B) GROUND/POLISHED, R_Q OF 0.15 μM (6 μIN.). 3000×

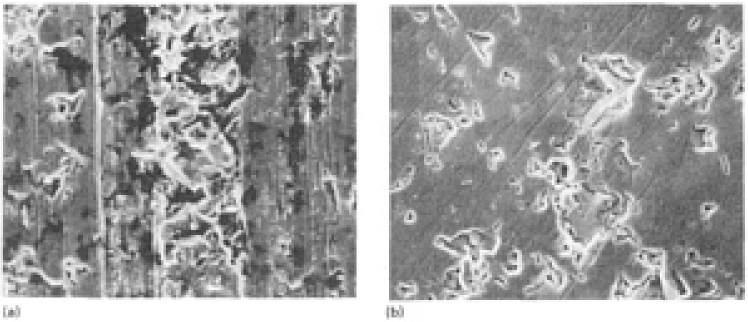

FIG. 5 TOPOGRAPHY OF SILICON CARBIDE WHISKER-REINFORCED ALUMINA SURFACES TO BE JOINED. (A) GROUND, R_Q OF 0.35 μM (14 μIN.). (B) GROUND/POLISHED, R_Q OF 0.04 μM (1.6 μIN.). 3000×

However, when the edges of the coupons were examined by scanning acoustic microscopy at 50 MHz before and after polishing, very significant differences were observed. Although the polishing of the composite material reduced the defect population, the procedure actually induced new subsurface defects in the alumina. A different set of grinding/polishing conditions might not produce the same results, but the importance of surface preparation on brazement strength should be obvious.

The brazing parameters themselves, that is, temperature, time at temperature (actually, the entire thermal cycle) and atmosphere, play a major role in the properties of a brazement and, therefore, are of critical importance in procedure development. Because strong bonds depend on chemical reactions at the ceramic/filler metal interface, one might assume that the strongest joints result from the highest temperatures and longest times. In many cases, it has been experimentally found that increasing a given parameter does, in fact, increase joint strength. However, this is true only up to a point. The authors and other researchers have found that joint strengths tend to reach a maximum as temperature and/or time are increased and then decrease with further increases, as illustrated in Fig. 6.

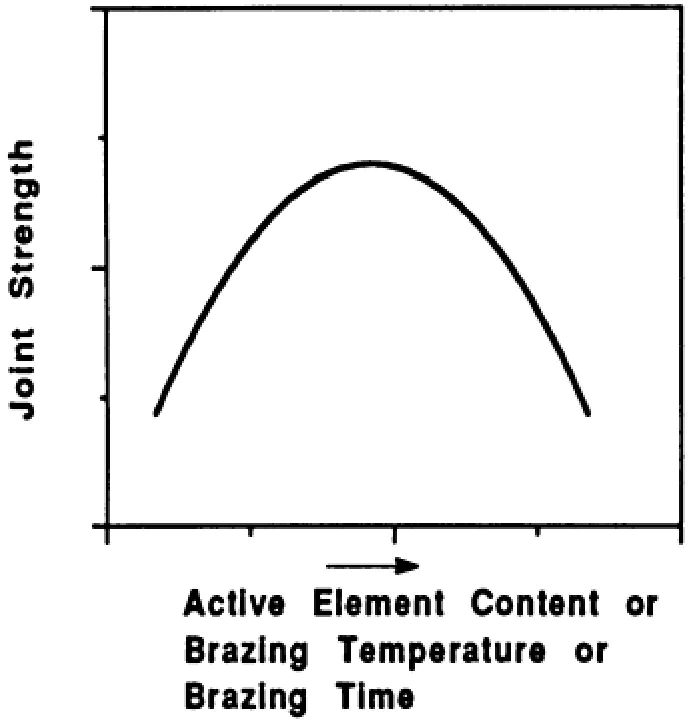

FIG. 6 RELATIONSHIP BETWEEN STRENGTH OF ACTIVE METAL BRAZEMENTS AND PRIMARY VARIABLES

For example, Kim and coworkers (Ref 78) observed this behavior when brazing a pressureless sintered silicon nitride with a series of copper-titanium filler metals. They found that the shear strengths of silicon nitride-silicon nitride joints were more affected by the morphology and thickness of the interfacial reaction layer than by wettability. As the titanium content of the binary filler alloys was increased, the equilibrium contact angle continually decreased, indicating improved wetting. As shown in Fig. 7, shear strength rapidly increased up to a relatively low titanium content of 5 wt% and then much more slowly afterwards. However, as shown in Fig. 8, the shear strengths initially increased and then decreased with further increases as temperature and time evidently exceeded optimum conditions. They attributed this behavior to excessive growth of the reaction layer, accompanied by the development of flaws in that layer.

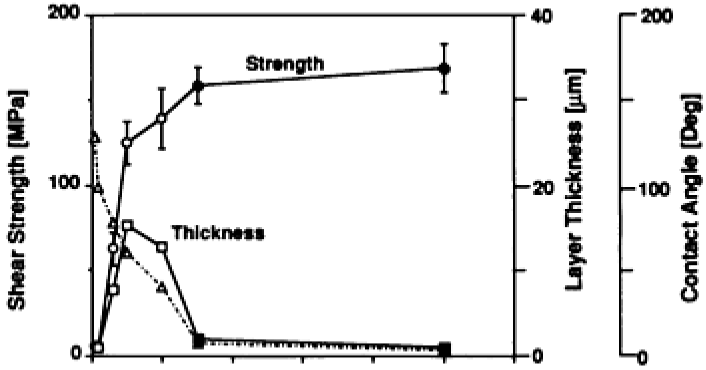

FIG. 7 EFFECT OF ACTIVE METAL CONTENT ON SHEAR STRENGTH, INTERFACIAL REACTION LAYER THICKNESS, AND WETTING ANGLE WHEN SILICON NITRIDE IS BRAZED WITH COPPER-TITANIUM FILLER METALS. OPEN DATA POINTS MEASURED AT 1100 °C (2010 °F), 1800 S; SOLID DATA POINTS MEASURED AT

1020 °C (1870 °F), 1800 S. SOURCE: REF 78

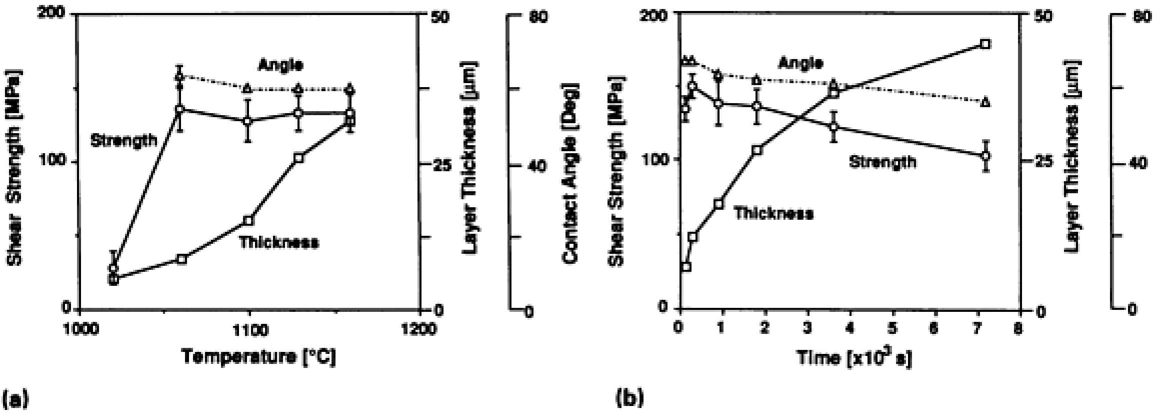

FIG. 8 EFFECT OF (A) BRAZING TEMPERATURE AND (B) HOLDING TIME AT TEMPERATURE ON SHEAR STRENGTH, CONTACT ANGLE, AND REACTION LAYER THICKNESS OF SILICON NITRIDE-SILICON NITRIDE JOINTS BRAZED WITH 95CU-5TI FILLER METAL. THE HOLDING TIME FOR THE BRAZEMENT OF (A) WAS 1800 S; THE SAMPLES FOR (B) WERE BRAZED AT 1130 °C (2070 °F). SOURCE: REF 78

The atmosphere used when brazing with active filler metals is also an important variable. Generally, a vacuum in the range of approximately 6 mPa (5×10^{-5} mm Hg) is used, but the literature of some manufacturers states that these filler metals can also be brazed in inert atmospheres, but of unspecified purity. Published reports have compared the strength of joints brazed in vacuum versus those brazed in inert gas, but little has been published on the effect of the actual magnitude of oxygen partial pressure on brazement wetting or adherence.

To facilitate an understanding of the effect of oxygen in the brazing atmosphere, the authors recently conducted a study that included PO_2 as one of the variables. A statistically designed experiment was conducted in order to objectively evaluate the role of certain parameters in the brazing of ceramic materials. An L9 orthogonal array and the Taguchi methodology were employed to examine four parameters at three levels on five different ceramic materials. The use of statistics in designing the experiment reduced the necessary number of brazing runs (for the five materials) from 320 to 45, while producing the same results. Each brazing run produced from six to twelve bend bars for testing. Therefore, the results were based on the average bend bar strength for each material and condition, or approximately 400 data points.

The materials were two aluminas (Coors AD-85, 85% Al_2O_3, and McDanel AD-998, 99.8% Al_2O_3), a silicon carbide whisker-reinforced alumina composite (Cercom PAD-AS30W), a sintered α-silicon carbide (Carborundum Hexoloy SA), and a sintered silicon nitride (GTE SNW-1000) that contained 13 wt% yttrium oxide (Y_2O_3) and 3 wt% Al_2O_3 as sintering aids. Material from a single lot of a commercially available brazing foil (GTE Wesgo's Cusil ABA, Ag-35Cu-2.0Ti, wt%) was used to join coupons of each ceramic material. The manufacturer recommends that this filler metal be brazed at temperatures ranging from 830 to 850 °C (1525 to 1560 °F), with a time at temperature of 3 to 10 min.

The mating edge of each coupon to be joined was prepared by one of three techniques: (1) Coarse grinding with a 180-grit diamond-impregnated grinding wheel; (2) Coarse grinding as in technique 1, followed by fine grinding with a 320-grit diamond wheel; or (3) Polishing the edges of some fine-ground coupons on a vibratory polisher using 6 μm (240 μin.) and 3 μm (120 μin.) grit diamond pastes. All of the grinding parameters, feeds, and speeds were held constant for coupons.

The brazing environments were flowing gases consisting of approximately 10, 100, and 1000 parts per million of oxygen in argon. The purity of the bottled argon was reduced to 10 ppm or less for each run by passing the gas through a commercial oxygen absorber. For the 100 and 1000 ppm runs, the P O_2 in the argon was controlled by mixing the purified argon with oxygen, and then measured with an oxygen analyzer with sensors located at the inlet and exit of the furnace tube.

The coupons were joined in a butt configuration in a silicon carbide fixture shown schematically in Fig. 9. The filler-metal foil was placed in the joint along with small, 38 μm (1.5 mil) thick tantalum shims on either end to prevent the joint from closing completely under the slight end loading provided by 250 μm (0.01 in.) thick molybdenum sheet "springs." Both the shims and filler metal were held in place during assembly by small drops of a Nicrobraz cement.

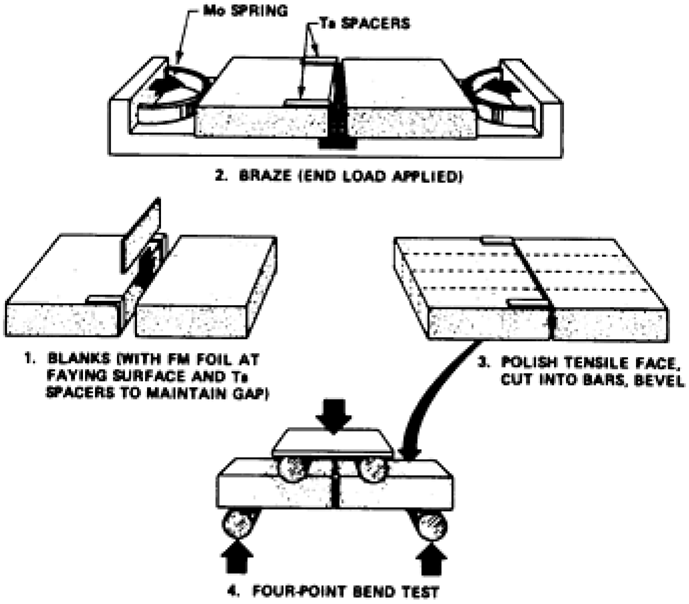
Mo SPRING
Ta SPACERS
2. BRAZE (END LOAD APPLIED)
1. BLANKS (WITH FM FOIL AT FAYING SURFACE AND Ta SPACERS TO MAINTAIN GAP)
3. POLISH TENSILE FACE, CUT INTO BARS, BEVEL
4. FOUR-POINT BEND TEST

FIG. 9 SEQUENCE OF OPERATIONS FOR FABRICATING BRAZED FLEXURAL BARS FOR THREE- OR FOUR-POINT BENDING TESTS

It is important to note that because the filler metal was confined between the ceramic coupons, the authors do not know what the PO_2 actually was within that space. The filler metal that exuded from the joint during brazing was oxidized gray or black, but beneath the oxide layer, the filler metal was still bright and shiny. This observation was not surprising, because in a previous study on oxidation of titanium-containing brazing filler metals (Ref 79), the authors found that the oxidation followed a simple parabolic rate law. Such behavior is indicative of the formation of a nonporous, coherent oxide film acting as a physical barrier to restrict further oxidation. Although an arrangement that has the filler metal placed within the confines of the mating surfaces may not be typical for all brazements, the authors think it is typical for active-metal brazements made with filler metals containing relatively low amounts of the active species. The ability of these materials to flow through a capillary joint is restricted by their low titanium content, which requires that they be preplaced within the joint.

Brazing of the nine sets of coupons of each material was done in a resistance-heated, alumina-tube furnace. The brazing cycle consisted of heating to a brazing temperature of 800 °C (1470 °F), 850 °C (1560 °F), or 900 °C (1650 °F) in about 15 min, holding at that temperature for 6, 12, or 18 min, and then cooling to room temperature in about 90 min. After brazing, the coupons were attached to steel backing bars and ground on both faces using a 180-grit diamond-abrasive wheel with the grinding direction perpendicular to the braze joint. The ground blanks were polished (on what would become the tensile surface of the flexure bars) with 6 μm (240 μin.) and 3 μm (120 μin.) diamond pastes on a vibratory polishing machine. The polished brazed coupons were then sliced in a series of parallel cuts (Fig. 10) to produce 3 by 4 by 25 mm (0.12 by 0.16 by 1 in.) flexure bars with the brazed joint across the center of the bars. The long edges of the tensile face of each sample were beveled on a 6 μm (240 μin.) grit diamond lap. Each brazing run produced from six to twelve bend bars for testing. Testing was in four-point bending, with spans of 6.35 and 19.05 mm (0.25 and 0.75 in.) and a crosshead speed of 0.08 mm/s (3 mils/s).

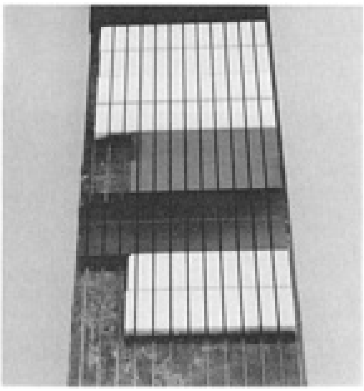

FIG. 10 POLISHED BRAZED COUPONS AFTER BEING SLICED IN A SERIES OF PARALLEL CUTS TO PRODUCE BARS FOR FLEXURAL STRENGTH TESTING

The results were very interesting and somewhat surprising. For example, very weak joints were produced in the sintered α-silicon carbide under many of the experimental conditions. In fact, the flexure bars from four of the nine brazements in this material failed after slicing. Because we have been able to prepare and test a number of bars with strengths of 20 MPa (3 ksi), we may assume that the failed bars had strengths less than that. As far as the effect of atmosphere is concerned, it was found that strong brazements could be produced in any of the materials, even when the oxygen partial pressure during brazing was 1000 ppm.

Figure 11 shows the ratios of the joint strength to ceramic strength for four of the materials. This figure illustrates that strong joints, with almost 50% of the bending strength of the respective ceramic, can be produced in each material, even in a relatively impure atmosphere. In order to make the comparisons valid, the strengths of the joints made with coarse-ground coupons brazed at 900 °C (1650 °F) for 12 min in an argon-1000 ppm O_2 atmosphere were used to calculate the ratios, not the maximum joint strength achieved for a given material. These were the conditions that produced the strongest brazement in the silicon carbide material, 256 ± 42 MPa (37 ± 6 ksi). The set of parameters, out of the nine sets of experimental conditions evaluated, that yielded the highest average joint strength for each material are given in Table 3. Interestingly, although these results may have been skewed by the failure of half of the silicon carbide brazements during bar fabrication, the recommended conditions (after statistical analysis of the results) that were predicted to produce optimum strength in the silicon carbide brazements were: 900 °C (1650 °F), 12 min, a coarse-ground finish, and 1000 ppm oxygen. In fact, two brazements in this material that were made in a vacuum of 4×10^{-3} Pa (3×10^{-5} mm Hg), which would be considered the "normal" atmosphere for active-metal brazing, for 6 min at 800 and 850 °C (1470 and 1560 °F), respectively, failed during flexure bar preparation.

TABLE 3 CONDITIONS FOUND TO PRODUCE THE HIGHEST AVERAGE JOINT STRENGTH WHEN BRAZING EACH OF FIVE STRUCTURAL CERAMICS WITH CUSIL-ABA IN AR-O_2 ATMOSPHERES

	TEMPERATURE, °C (°F)			**TIME, MIN**			**AR-*P* O_2, PPM**			JOINT FINISH
	800 (1470)	**850 (1560)**	**900 (1650)**	**6**	**12**	**18**	**10**	**100**	**1000**	
AD-85 ALUMINA (228 ± 42 MPA, OR 33 ± 6 KSI)	...	...	X	X	...	...	...	X	...	**POLISHED**
AD-998 ALUMINA (317 ± 16 MPA, OR 46 ± 2.5 KSI)	X	...	...	...	...	X	...	...	X	**POLISHED**
SILICON CARBIDE WHISKER-REINFORCED ALUMINA (445 ± 37 MPA, OR 65 ± 5.5 KSI)	...	...	X	X	...	...	...	X	...	**POLISHED**
α-SILICON CARBIDE (256 ± 42 MPA, OR 37 ± 6 KSI)	...	...	X	...	X	...	...	...	X	**COARSE**
SILICON NITRIDE (512 ± 74 MPA, OR 74 ± 11 KSI)	**X**	**...**	**...**	**...**	**...**	**X**	**...**	**...**	**X**	**POLISHED**

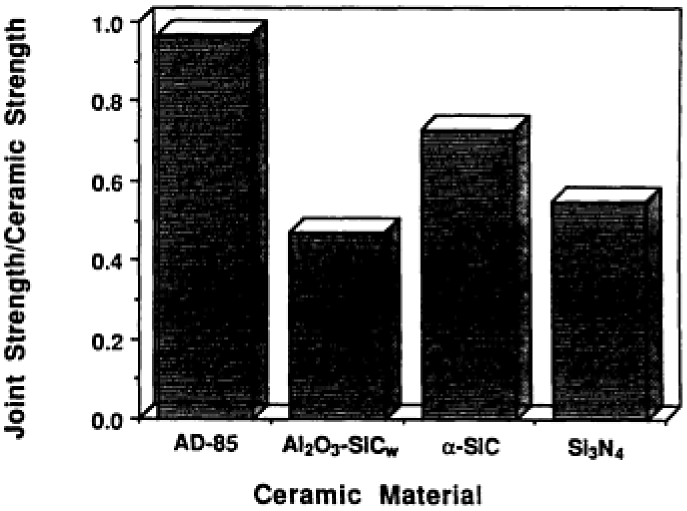

FIG. 11 RATIO OF JOINT STRENGTH TO CERAMIC STRENGTH FOR FOUR CERAMICS BRAZED WITH AG-35CU-2TI, WT%, FILLER METAL FOR 12 MIN AT 900 °C (1650 °F) IN AN AR-1000 PPM O_2 ATMOSPHERE

The statistical analysis also recommended the highest level of oxygen included in the experimental plan (1000 ppm), rather than 10 ppm, as one would anticipate, in the parameters *projected* to produce the highest strengths in brazements with this filler metal (Ag-35Cu-2.0Ti, wt%) in the AD-998 alumina and silicon nitride materials. The only material for which 10 ppm oxygen was recommended was the silicon carbide whisker-reinforced alumina composite. The necessary confirmatory braze runs have not been conducted, as of mid-1993, to determine whether even higher strengths are produced in the silicon carbide whisker-reinforced alumina and silicon nitride materials. The strengths of the alumina materials have already been exceeded by the strengths of the brazements made under several conditions.

However, the authors have conducted an interesting side study (outside the test matrix) that should provide some guidance to someone preparing a brazing procedure at this time. Coupons of the SNW-1000 silicon nitride ceramic were brazed as had been done before, but in a vacuum of 3 mPa (2×10^{-5} mm Hg), to determine whether that atmosphere had a significant effect on brazement strength. Very little differences in brazement strengths were observed. For example, the average strengths of the joints made with brazing conditions of 850 °C (1560 °F), 6 min, and a fine-ground joint surface were 505 ± 74 MPa (73.2 ± 10.5 ksi) and 496 ± 59 MPa (71.9 ± 8.5 ksi) for an atmosphere of vacuum and argon-1000 ppm O_2, respectively. These results suggest that atmosphere purity is not as critical as one might assume when brazing silicon nitride with an active filler metal (at least when the filler metal is preplaced in the joints).

References cited in this section

33. R.L. BRONNES, R.C. HUGHES, AND R.C. SWEET, CERAMIC-TO-METAL BONDING WITH SPUTTERING AS A METALLIZATION TECHNIQUE, *PHILIPS TECH. REV.,* VOL 35, 1975, P 209-211
34. J.P. HAMMOND, S.A. DAVID, AND M.L. SANTELLA, BRAZING CERAMIC OXIDES TO METALS AT LOW TEMPERATURES, *WELD. J.,* VOL 7, 1988, P 227-S TO 232-S
35. P.O. JARVINEN, "NOVEL CERAMIC RECEIVER FOR SOLAR BRAYTON SYSTEMS," CONF-790305-7, MIT LINCOLN LABORATORY, 1979
36. F. HEIM AND H.L. MCDOWELL, "METHOD OF METALLIZING A CERAMIC SUBSTRATE," U.S. PATENT 4,342,632, AUGUST 3, 1982
37. S. WEISS AND C.M. ADAMS, JR., THE PROMOTION OF WETTING AND BRAZING, *WELD. J. SUPPL.,* VOL 46, 1967, P 49S-57S
38. E.F. BRUSH AND C.M. ADAMS, VAPOR COATED SURFACES FOR BRAZING CERAMICS, *WELD. J.,* VOL 47 (NO. 3), 1968, P 106-S TO 114-S

39. M.L. SANTELLA, BRAZING OF TITANIUM-VAPOR-COATED SILICON NITRIDE, *ADV. CERAM. MATER.,* VOL 3 (NO. 5), 1988, P 457-462
40. M.J. RAMSEY AND M.H. LEWIS, INTERFACIAL REACTION MECHANISMS IN SYALON CERAMIC BONDING, *MATER. SCI. ENG.,* VOL 71, 1985, P 113-122
41. J.P. HAMMOND AND G.M. SLAUGHTER, BONDING GRAPHITE TO METALS WITH TRANSITION PIECES, *WELD. J.,* VOL 50 (NO. 1), 1971, P 33-40
42. M.E. TWENTYMAN, PART 1. THE MECHANISM OF GLASS MIGRATION IN THE PRODUCTION OF METAL-CERAMIC SEALS, *J. MATER. SCI.,* VOL 10, 1975, P 765-776
43. M.E. TWENTYMAN AND P. POPPER, PART 2. THE EFFECT OF EXPERIMENTAL VARIABLES ON THE STRUCTURE OF SEALS TO DEBASED ALUMINAS, *J. MATER. SCI.,* VOL 10, 1975, P 777-790
44. M.E. TWENTYMAN AND P. POPPER, PART 3. THE USE OF METALLIZING PAINTS CONTAINING GLASS OR OTHER INORGANIC BONDING AGENTS, *J. MATER. SCI.,* VOL 10, 1975, P 791-798
45. A. KOHNO *ET AL.,* METALLIZING OF NON-OXIDE CERAMICS, *SPECIAL CERAMICS 8, BR. CERAM. PROC.,* NO. 37, S.P. HOWLETT AND D. TAYLOR, ED., OCT 1986, P 125-130
46. E.M. DUNN *ET AL.,* ANALYTICAL AND EXPERIMENTAL EVALUATION OF JOINING SILICON NITRIDE TO METAL AND SILICON CARBIDE TO METAL, *PROC. 26TH AUTOMOTIVE TECHNOLOGY DEVELOPMENT CONTRACTORS COORDINATION MEETING,* SAE P-219, SAE, 1988, P 165-171
47. YU.V. NAIDICH, THE WETTABILITY OF SOLIDS BY LIQUID METALS, *PROGRESS IN SURFACE AND MEMBRANE SCIENCE,* VOL 14, INSTITUTE OF MATERIAL PROBLEMS, ACADEMY OF SCIENCES, UKRAINIAN S.S.R., 1981, P 333-484
48. M.G. NICHOLAS, ACTIVE METAL BRAZING, *BR. CERAM. TRANS. J.,* VOL 85, 1986, P 144-146
49. M.G. NICHOLAS, T.M. VALENTINE, AND M.J. WAITE, THE WETTING OF ALUMINA BY COPPER ALLOYED WITH TITANIUM AND OTHER ELEMENTS, *J. MATER. SCI.,* VOL 15, 1980, P 2197-2220
50. H. MIZUHARA AND K. MALLY, CERAMIC-TO-METAL JOINING WITH ACTIVE BRAZING FILLER METAL, *WELD. J.,* VOL 64 (NO. 10), 1985, P 27-32
51. A.J. MOORHEAD AND H. KEATING, DIRECT BRAZING OF CERAMICS FOR ADVANCED HEAVY-DUTY DIESELS, *WELD. J.,* VOL 65 (NO. 10), 1986, P 17-31
52. M.G. NICHOLAS, PROMOTING THE WETTING AND BRAZING OF CERAMICS, *STRUCT. CERAM. JOINING II,* CERAMIC TRANSACTIONS, VOL 35, A.J. MOORHEAD, R.E. LOEHMAN, AND S.M. JOHNSON, ED., AMERICAN CERAMIC SOCIETY, 1993
53. YU.V. NAIDICH AND V.S. ZHURAVLEV, ADHESION, WETTABILITY AND REACTION OF TITANIUM-CONTAINING MELTS WITH HIGH-MELTING OXIDES, TRANSLATED FROM *OGNEUPORY,* VOL 1, 1974, P 50-55
54. T. ISEKI, H. MATSUZAKI, AND J.K. BOADI, BRAZING OF SILICON CARBIDE TO STAINLESS STEEL, *AM. CERAM. SOC. BULL.,* VOL 64 (NO. 2), 1985, P 322-324
55. R.E. LOEHMAN AND A.P. TOMSIA, *AM. CERAM. SOC. BULL.,* VOL 67 (NO. 2), 1988, P 375
56. M.J. RAMSEY AND M.H. LEWIS, *MATER. SCI. ENG.,* VOL 71, 1985, P 113-122
57. T. ISEKI, H. MATSUZAKI, AND J.K. BOADI, BRAZING OF SILICON CARBIDE TO STAINLESS STEEL, *AM. CERAM. SOC. BULL.,* VOL 64 (NO. 2), 1985, P 322-324
58. M. NAKA *ET AL.,* JOINING OF SILICON NITRIDE WITH AL-CU ALLOYS, *J. MATER. SCI.,* VOL 22, 1987, P 4417-4421
59. T. ISEKI *ET AL., CERAMIC SURFACES AND SURFACE TREATMENTS, BR. CERAM. PROC.,* NO. 34, R. MORRELL AND M.G. NICHOLAS, ED., AUGUST 1984, P 241-248
60. H. MIZUHARA AND T. OYAMA, CERAMIC-METAL SEALS, *CERAMICS AND GLASSES,* VOL 4, *ENGINEERED MATERIALS HANDBOOK,* ASM INTERNATIONAL, 1991, P 502-510
61. M.G. NICHOLAS, PROMOTING THE WETTING AND BRAZING OF CERAMICS, *STRUCT. CERAM.*

JOINING II, CERAMIC TRANSACTIONS, VOL 35, A.J. MOORHEAD, R.E. LOEHMAN, AND S.M. JOHNSON, ED., AMERICAN CERAMIC SOCIETY, 1993

62. W.A. ZDANIEWSKI *ET AL.,* CRYSTALLIZATION TOUGHENING OF CERAMIC ADHESIVES FOR JOINING ALUMINA, *ADV. CERAM. MATER.,* VOL 2 (NO. 3A), 1987, P 204-208
63. S.L. SWARTZ *ET AL.,* JOINING OF ZIRCONIA CERAMICS WITH A CAO-TIO_2-SIO_2 INTERLAYER, *PROC. 26TH AUTOMOTIVE TECHNOLOGY DEVELOPMENT CONTRACTORS COORDINATION MEETING,* P-219, SAE, 1988, P 149-154
64. M.E. MILBERG *ET AL.,* THE NATURE OF SIALON JOINTS BETWEEN SILICON NITRIDE BASED BODIES, *J. MATER. SCI.,* VOL 22, 1987, P 2560-2568
65. S. BAIK AND R. RAJ, LIQUID-PHASE BONDING OF SILICON NITRIDE CERAMICS, *J. AM. CERAM. SOC.,* VOL 70 (NO. 5), 1987, P C105-C107
66. S.M. JOHNSON AND D.J. ROWCLIFFE, MECHANICAL PROPERTIES OF JOINED SILICON NITRIDE, *J. AM. CERAM. SOC.,* VOL 68 (NO. 9), 1985, P 468-472
67. N. IWAMOTO *ET AL.,* JOINING OF SILICON NITRIDE CERAMICS WITH CAO-TIO_2-SIO_2 GLASS SOLDER, *CERAMIC MATERIALS AND COMPONENTS FOR ENGINES, PROC. 2ND INT. SYMPOSIUM,* W. BUNK AND H. HAUSNER, ED., DEUTSCHE KERAMISCHE GESELLSCHAFT E.V., 1986, P 467-474
68. M.L. MECARTNEY *ET AL.,* SILICON NITRIDE JOINING, *J. AM. CERAM. SOC.,* VOL 68 (NO. 9), 1985, P 472-478
69. M. NEILSON, JR. AND D.N. COON, STRENGTH OF SILICON NITRIDE-SILICON NITRIDE JOINTS BONDED WITH OXYNITRIDE GLASS, *CERAM. ENG. SCI. PROC.,* VOL 10 (NO. 11-12), 1989, P 1893-1907
70. D.N. COON, R.L. TALLMAN, AND R.M. NEILSON, JR., HOT ISOSTATICALLY PRESSED SI_3N_4-SI_3N_4 JOINTS BONDED WITH OXYNITRIDE GLASS, *ADV. CERAM. MATER.,* VOL 3 (NO. 2), 1988, P 154-158
71. J.F. BURGESS, C.A. NEUGEBAUER, AND G. FLANAGAN, THE DIRECT BONDING OF METALS TO CERAMICS BY THE GAS-METAL EUTECTIC METHOD, *J. ELECTROCHEM. SOC.,* VOL 122 (NO. 5), 1975, P 688-690
72. Y. YOSHINO, ROLE OF OXYGEN IN BONDING COPPER TO ALUMINA, *J. AM. CERAM. SOC.,* VOL 72 (NO. 8), 1989, P 1322-1327
73. I. OKAMOTO, M. NAKA, K. ASAMI, AND K. HASHIMOTO, FORMATION OF HERCYNITE ($FEAL_2O_4$) AT INTERFACE OF AL_2O_3/STEEL JOINT, *TRANS. JWRI,* VOL 11 (NO. 1), 1982, P 131-133
74. M. NAKA AND I. OKAMOTO, JOINING OF AL_2O_3 TO SUS 304 WITH TI FOIL, *TRANS. JWRI,* VOL 15 (NO. 1), 1986, P 157-158
75. Y. IINO, PARTIAL TRANSIENT LIQUID-PHASE METALS LAYER TECHNIQUE OF CERAMIC-METAL BONDING, *J. MATER. SCI. LETT.,* VOL 10 (NO. 2), 1991, P 104-106
76. M.L. SHALZ, B.J. DALGLEISH, A.P. TOMSIA, AND A.M. GLAESER, NEW APPROACHES TO JOINING CERAMICS FOR HIGH-TEMPERATURE APPLICATIONS, *STRUCTURAL CERAM. JOINING II,* CERAMIC TRANSACTIONS, VOL 35, A.J. MOORHEAD, R.E. LOEHMAN, AND S.M. JOHNSON, ED., AMERICAN CERAMIC SOCIETY, 1993
77. D.S. DUVALL, W.A. OWCZARSKI, AND D.F. PAULONIS, TLP BONDING: A NEW METHOD FOR JOINING HEAT RESISTANT ALLOYS, *WELD. J.,* VOL 53 (NO. 4), 1974, P 203-214
78. D.H. KIM, S.H. HWANG, AND S.S. CHUN, THE WETTING, REACTION AND BONDING OF SILICON NITRIDE BY CU-TI ALLOYS, *J. MATER. SCI.,* VOL 26 (NO. 12), JUNE 1991, P 3223-3234
79. A.J. MOORHEAD AND H.-E. KIM, OXIDATION BEHAVIOUR OF TITANIUM-CONTAINING BRAZING FILLER METALS, *J. MATER. SCI.,* VOL 26, 1991, P 4067-4075

Ceramic-to-Metal Joints

In addition to the factors considered above for ceramic-to-ceramic joints, there are specific factors that must be considered when developing a procedure for brazing ceramic-to-metal joints. One is that a thermal cycle must be chosen that does not degrade the properties of the materials being joined. Thermal exposure may be a problem for some structural ceramics, such as the magnesia-PSZ, the strength and toughness of which tend to be altered by exposure to temperatures of 1000 °C (1830 °F) and higher for only a few hours (Ref 80). However, most structural ceramics are thermally stable at temperatures typical of those used for brazing with active filler metals (see Table 2) or glasses. The problem, which has been deemed "thermal compatibility" by Pejryd (Ref 81), is generally more severe for the metallic member of a ceramic-to-metal joint. For example, he has shown (Fig. 12) that even a very modest thermal cycle can have a major effect on the mechanical properties of low-expansion superalloy Incoloy 909, which is often considered in the literature as a likely candidate metal in ceramic-to-metal joints. Such thermal behavior tendencies must be taken into account both during material selection and procedure development.

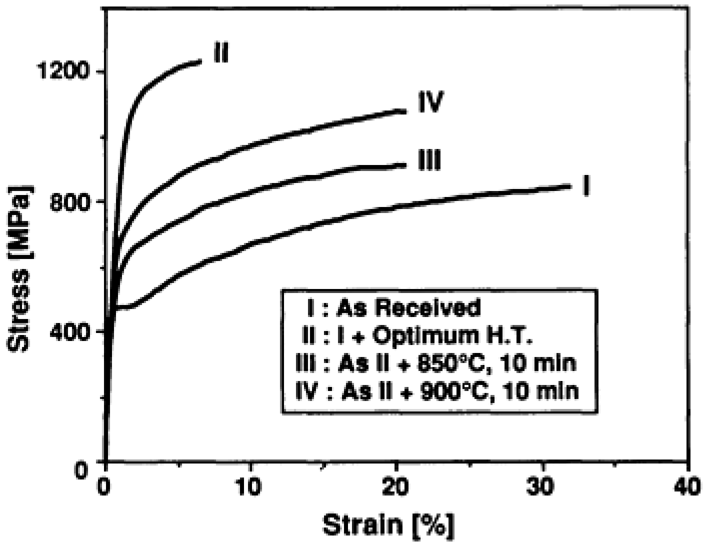

FIG. 12 TENSILE TEST CURVES FOR INCOLOY 909 SUBJECTED TO DIFFERENT BRAZING HEAT-TREATMENT CYCLES. SOURCE: REF 81

Another critical factor is the CTE mismatch of the materials. Differences in expansion behavior can generate significant thermal residual stresses in a ceramic-to-metal brazement, resulting in distortion and even failure during cooldown from the brazing temperature, as shown in Fig. 13.

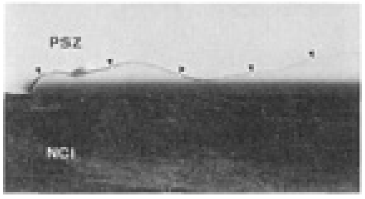

FIG. 13 FAILURE IN THE CERAMIC SIDE OF A CERAMIC-TO METAL BRAZEMENT RESULTING FROM A MISMATCH IN COEFFICIENTS OF THERMAL EXPANSION BETWEEN THE CERAMIC (MAGNESIA-PSZ) AND THE

METAL (NODULAR CAST IRON)

In this example, 2.5 by 4.5 by 26 mm (0.1 by 0.2 by 1 in.) coupons of nodular cast iron and magnesia-PSZ were joined along their long edges using an active filler metal. Upon cooling from the brazing temperature, failure occurred in the ceramic parallel to the interface. These materials are considered to have similar CTE values and had been successfully joined before. However, when thermal expansion data were taken for this particular piece of ceramic, an anomalous and large hysteresis loop was found, indicating that the material had been improperly heat treated during its manufacture, resulting in a different phase composition (and expansion behavior) than expected.

The obvious solution to the problem of CTE mismatch among the members of a ceramic-to-metal joint and the resulting damaging residual stresses is to select an alternate material for either the ceramic or the metallic member. When this cannot be done because of service requirements, or when the thermal expansion is still too great, there are several other alternatives. One is to design the joint such that the ceramic member and joint interfacial region are always in compression. Another is to use ductile metal, which elastically and plastically deforms, to relieve the residual stresses generated during cooldown (or thermal cycling during service), thereby avoiding joint failure. Both of these techniques are widely used by the commercial ceramic/metal or glass/metal seal industry (Ref 60, 82). A third option is to introduce one or more materials between the metal and ceramic components. These materials will serve as a transition between the properties of the two structural materials. The transition material can be one or more discrete layers of monolithic materials (Ref 83, 84, 85) or composite layers that can comprise varying proportions of the materials being joined (Ref 41, 86, 87).

The concept of monolithic transition layers, their compositions (and, thus, properties such as thermal expansion, modulus of elasticity, yield strength, and others), and the optimum thicknesses required are complex subjects. Readers are referred to the references cited above for the details of this approach.

The use of a series of composite layers (with a stepwise transition from one material to the next or a linear gradation in material composition) that are subsequently inserted in the braze joint area is not new. It has been applied successfully, as evidenced by the work of Hammond and Slaughter (Ref 41), who used a series of tungsten-nickel-iron inserts to braze graphite to a nickel-base alloy. In one example, seven transition inserts were used with compositions that varied from 97.5 wt% W and 2.5 wt% of 7Ni/3Fe, adjacent to the graphite, to 40 wt% W and 60 wt% of 7Ni/3Fe, in contact with the nickel alloy. The surface of the graphite was metallized with chromium carbide, and the entire assembly was brazed with pure copper as the filler material.

However, there are numerous cases in which this approach has led to failure. The recent work of Rabin and Heaps (Ref 80) provides a new understanding of how the geometry and micro-structure of the composite transition layer(s) affects the stress distributions in the materials. Using finite-element modeling of alumina-nickel functionally gradient materials, followed by experimental studies, they showed that it is desirable to have a continuous metal phase (high plasticity, low elastic modulus) in each of the transition pieces to best relieve the thermal residual stresses. From a practical standpoint, they were able to accomplish this by using different particle sizes of the alumina and nickel in the compacts. Thus, for the alumina, alumina-20% nickel, and alumina-40% nickel layers, they used 0.5 μm (20 μin.) doped aluminum oxide and 15 to 20 μm (600 to 800 μin.) nickel powders. For the nickel, 60% nickel, and 80% nickel layers, they used 22 μm (880 μin.) doped aluminum oxide and 3 to 7 μm (120 and 280 μin.) nickel powders. The "dopants" in this study were small amounts of manganese and titanium used as sintering aids.

Direct Brazing of Graphitic Materials. Most of the factors that are a concern when brazing ceramics to themselves or to metals are also valid when brazing the various graphites. Two major differences should be considered when these materials are to be joined. The first, which may be overcome through careful material selection, is the fact that the graphites generally tend either to be more porous than the structural ceramics or, if of similar bulk porosity, to have a much finer network of porosity. The effect of such fine porosity can be to draw the filler metal away from the interface, leaving insufficient metal to join the bodies (Ref 88). This effect is readily seen in the micrograph of Fig. 14(a), which shows extensive penetration of a porous graphite when brazed with GTE Wesgo's Ticuni (Ti-15Cu-15Ni, wt%) filler metal. Penetration of the graphite ranged in depth from 200 to 890 μm (0.008 to 0.035 in.) across the metallographic specimen. The solution to this problem was to use a graphite of much greater POCO AXF-5Q density (Fig. 14b), thereby minimizing the loss of filler metal into the component, resulting in the sound braze joint shown in Fig. 14(b).

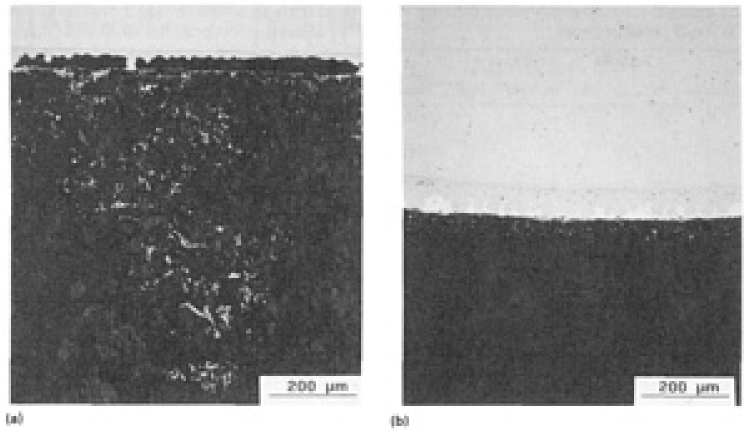

FIG. 14 EFFECT OF POROSITY ON ACTIVE METAL BRAZEMENTS OF GRAPHITE. (A) POROUS GRAPHITE HAS DRAWN THE MOLTEN FILLER METAL FROM THE JOINT, LEAVING A LARGE VOID. (B) MORE HIGHLY DENSE GRAPHITE HAS LEFT MOST OF THE FILLER METAL IN THE JOINT, RESULTING IN A SOUND BRAZEMENT. SOURCE: REF 88

The second major concern when brazing graphites is the stronger tendency for impurities, such as oxygen or water, to be either adsorbed on the surface or absorbed in the bulk of the material. Moisture retention always occurs to some extent, with levels as high as 0.25 wt%. This problem was overcome in one case (Ref 81) by cleaning and baking the graphite components in vacuum for 2 h at 1000 °C (1830 °F) prior to assembly, and then baking the subassemblies themselves (650 °C, or 1200 °F, for 10 min, followed by an overnight evacuation) as part of the brazing cycle.

References cited in this section

41. J.P. HAMMOND AND G.M. SLAUGHTER, BONDING GRAPHITE TO METALS WITH TRANSITION PIECES, *WELD. J.,* VOL 50 (NO. 1), 1971, P 33-40
60. H. MIZUHARA AND T. OYAMA, CERAMIC-METAL SEALS, *CERAMICS AND GLASSES,* VOL 4, *ENGINEERED MATERIALS HANDBOOK,* ASM INTERNATIONAL, 1991, P 502-510
80. D.J. GREEN, R.H.J. HANNINK, AND M.V. SWAIN, *TRANSFORMATION TOUGHENING OF CERAMICS,* CRC PRESS, 1989, P 115-121
81. L. PEJRYD, METAL CERAMIC JOINING FOR HIGH TEMPERATURE APPLICATIONS, *4TH INT. SYMPOSIUM ON CERAMIC MATERIALS AND COMPONENTS FOR ENGINES,* R. CARLSSON, T. JOHANSSON, AND L. KAHLMAN, ED., ELSEVIER APPLIED SCIENCE, 1992, P 50-66
82. E.K. BEAUCHAMP AND S.N. BURCHETT, TECHNIQUES OF SEAL DESIGN, *CERAMICS AND GLASSES,* VOL 4, *ENGINEERED MATERIALS HANDBOOK,* ASM INTERNATIONAL, 1991, P 532-541
83. M.G. NICHOLAS AND R.M. CRISPIN, DIFFUSION BONDING STAINLESS STEEL TO ALUMINA USING ALUMINUM INTERLAYERS, *J. MATER. SCI.,* VOL 17, 1982, P 3347-3360
84. K. SUGANUMA, T. OKAMOTO, AND M. KOIZUMI, EFFECT OF INTERLAYERS IN CERAMIC-METAL JOINTS WITH THERMAL EXPANSION MISMATCHES, *J. AM. CERAM. SOC.,* VOL 67 (NO. 12), 1984, P C256-C257
85. A.-P. XIAN AND Z.-Y. SI, INTERLAYER DESIGN FOR JOINING PRESSURELESS SINTERED SIALON CERAMIC AND 40CR STEEL BRAZING WITH $AG_{57}CU_{38}TI_5$ FILLER METAL, *J. MATER. SCI.,* VOL 27, 1992, P 1560-1566
86. K. SUGANUMA, T. OKAMOTO, M. SHIMADA, AND M. KOIZUMI, NEW METHOD FOR SOLID-STATE BONDING BETWEEN CERAMICS AND METALS, *J. AM. CERAM. SOC.,* 1983, P C117-C118
87. B.H. RABIN AND R.J. HEAPS, POWDER PROCESSING OF $NI\text{-}AL_2O_3$ FGM, TO BE PUBLISHED IN *CERAMIC TRANSACTIONS, PROC. 2ND INT. SYMPOSIUM ON FUNCTIONALLY GRADIENT*

MATERIALS, AMERICAN CERAMIC SOCIETY, 1993

88. J.F. KING *ET AL.*, MATERIALS TESTS AND ANALYSES OF FARADAY SHIELD TUBES FOR ICRF ANTENNAS, *FUSION TECHNOL.*, VOL 15 (NO. 2B), 1989, P 1093-1096

Brazing of Ceramic and Ceramic-to-Metal Joints

A.J. Moorhead and W.H. Elliott, Jr., Oak Ridge National Laboratory H.-E. Kim, Seoul National University

---

References

1. D.A. PARKER, TECHNOLOGICAL REQUIREMENTS FOR DESIGNING INTERFACES FOR MECHANICAL AND THERMAL APPLICATIONS, *DESIGNING INTERFACES FOR TECHNOLOGICAL APPLICATIONS: CERAMIC-CERAMIC, CERAMIC-METAL JOINING,* S.D. PETEVES, ED., ELSEVIER, 1989, P 3-32
2. R.E. LOEHMAN, CERAMIC JOINING IN THE U.S., *DESIGNING INTERFACES FOR TECHNOLOGICAL APPLICATIONS: CERAMIC-CERAMIC, CERAMIC-METAL JOINING,* S.D PETEVES, ED., ELSEVIER, 1989, P 235-245
3. T. SUGA, CURRENT RESEARCH AND FUTURE OUTLOOK IN JAPAN, *DESIGNING INTERFACES FOR TECHNOLOGICAL APPLICATIONS: CERAMIC-CERAMIC, CERAMIC-METAL JOINING,* S.D. PETEVES, ED., ELSEVIER, 1989, P 247-263
4. L.M. SHEPPARD, AUTOMOTIVE PERFORMANCE ACCELERATES WITH CERAMICS, *AM. CERAM. SOC. BULL.,* VOL 69 (NO. 6), 1990, P 1012-1021
5. K. MATOBA, K. KATAYAMA, M. KAWAMURA, AND T. MIZUNO, THE DEVELOPMENT OF SECOND GENERATION CERAMIC TURBOCHARGER ROTOR--FURTHER IMPROVEMENTS IN RELIABILITY, *AUTOMOTIVE CERAMICS-RECENT DEVELOPMENTS,* SAE P-207, SAE, 1988, P 37-46
6. S. ISHIWATA, AUTOMOTIVE CERAMIC GAS TURBINE R & D IN JAPAN, *PROC. 27TH AUTOMOTIVE TECHNOLOGY DEVELOPMENT CONTRACTORS COORDINATION MEETING,* SAE P-230, SAE, 1990, P 287-292
7. H.E. HELMS, P.J. HALEY, L.E. GROSECLOSE, S.J. HILPISCH, AND A.H. BELL, ADVANCED TURBINE TECHNOLOGY APPLICATIONS PROJECT, *PROC. 27TH AUTOMOTIVE TECHNOLOGY DEVELOPMENT CONTRACTORS COORDINATION MEETING,* SAE P-230, SAE, 1990, P 293-301
8. J.R. KIDWELL, L.J. LINDBERG, AND R.E. MOREY, ATTAP/AGT101--YEARS PROGRESS IN CERAMIC TECHNOLOGY DEVELOPMENT, *PROC. 27TH AUTOMOTIVE TECHNOLOGY DEVELOPMENT CONTRACTORS COORDINATION MEETING,* SAE P-230, SAE, 1990, P 327-334
9. B.D. FOSTER AND J.B. PATTON, ED., *CERAMICS IN HEAT EXCHANGERS,* VOL 14, *PROC. THREE SYMPOSIA ON HEAT EXCHANGERS,* AMERICAN CERAMIC SOCIETY, 1984-1985
10. A.J. HAYES, W.W. LIANG, S.L. RICHLEN, AND E.S. TABB, ED., *INDUSTRIAL HEAT EXCHANGERS, PROC. 1985 EXPOSITION AND SYMPOSIUM ON INDUSTRIAL HEAT EXCHANGER TECHNOLOGY,* AMERICAN SOCIETY FOR METALS, 1985
11. M.H. LEWIS, CERAMICS TO BE JOINED 10 YEARS FROM NOW, *DESIGNING INTERFACES FOR TECHNOLOGICAL APPLICATIONS: CERAMIC-CERAMIC, CERAMIC-METAL JOINING,* S.D. PETEVES, ED., ELSEVIER, 1989, P 271-290
12. P.F. BECHER AND M.K. FERBER, MECHANICAL BEHAVIOUR OF MGO-PARTIALLY STABILIZED ZRO_2 CERAMICS AT ELEVATED TEMPERATURES, *J. MATER. SCI.,* VOL 22, 1987, P 973-980
13. M. RÜHLE AND N. CLAUSSEN, TRANSFORMATION AND MICROCRACK TOUGHENING AS COMPLEMENTARY PROCESSES IN ZRO_2-TOUGHENED AL_2O_3, *J. AM. CERAM. SOC.,* VOL 69 (NO. 3), 1986, P 195-197

14. P.F. BECHER AND T.N. TIEGS, TEMPERATURE DEPENDENCE OF STRENGTHENING BY WHISKER REINFORCEMENT: SIC WHISKER-REINFORCED ALUMINA IN AIR, *ADV. CERAM. MATER.,* VOL 3 (NO. 2), 1988, P 148-153
15. K.M. PREWO, TENSION AND FLEXURAL STRENGTH OF SILICON CARBIDE FIBRE-REINFORCED GLASS CERAMICS, *J. MATER. SCI.,* VOL 21, 1986, P 3590-3600
16. J.J. BRENNAN, GLASS AND GLASS-CERAMIC MATRIX COMPOSITES, *FIBER REINFORCED CERAMIC COMPOSITES,* K.S. MAZDIYASNI, ED., NOYES PUBLICATIONS, 1990, P 222-259
17. T. MAH *ET AL.,* CERAMIC FIBER REINFORCED METAL-ORGANIC PRECURSOR MATRIX COMPOSITES, *FIBER REINFORCED CERAMIC COMPOSITES,* K.S. MAZDIYASNI, ED., NOYES PUBLICATIONS, 1990, P 278-310
18. C.W. LI AND J. YAMANIS, SUPER-TOUGH SILICON NITRIDE WITH *R*-CURVE BEHAVIOR, *CERAM. ENG. SCI. PROC.,* VOL 10 (NO. 7-8), 1989, P 632-645
19. T. KAWASHIMA, H. OKAMOTO, H. YAMAMOTO, AND A. KITAMURA, GRAIN SIZE DEPENDENCE OF THE FRACTURE TOUGHNESS OF SILICON NITRIDE CERAMICS, *J. CERAM. SOC. JAPAN,* INT. ED., VOL 99 (NO. 4), 1991, P 1-4
20. W.D. KINGERY, H.K. BOWEN, AND D.R. UHLMANN, *INTRODUCTION TO CERAMICS,* WILEY, 1976
21. R.W. RICE, CERAMIC COMPOSITES: FUTURE NEEDS AND OPPORTUNITIES, *FIBER REINFORCED CERAMIC COMPOSITES,* K.S. MAZDIYASNI, ED., NOYES PUBLICATIONS, 1990, P 451-495
22. P.F. BECHER, MICROSTRUCTURAL DESIGN OF TOUGHENED CERAMICS, *J. AM. CERAM. SOC.,* VOL 74 (NO. 2), 1991, P 255-269
23. "MATERIAL PROPERTIES DATA SHEET," COORS CERAMICS COMPANY
24. P. CHANTIDUL, S.J. BENNISON, AND B.R. LAWN, ROLE OF GRAIN SIZE IN THE STRENGTH AND *R*-CURVE PROPERTIES OF ALUMINA, *J. AM. CERAM. SOC.,* VOL 78 (NO. 8), 1991, P 2419-2427
25. S. SOMIYA, N. YAMAMOTO, AND H.H. YANAGIDA, ED., *SCIENCE AND TECHNOLOGY OF ZIRCONIA III,* AMERICAN CERAMIC SOCIETY, 1988
26. D. SHIN, K.K. ORR, AND H. SCHUBERT, MICROSTRUCTURE-MECHANICAL PROPERTY RELATIONSHIPS IN HOT ISOSTATICALLY PRESSED ALUMINA AND ZIRCONIA-TOUGHENED ALUMINA, *J. AM. CERAM. SOC.,* VOL 73 (NO. 5), 1990, P 1181-1188
27. H. SCHNEIDER AND E. EBERHARD, THERMAL EXPANSION OF MULLITE, *J. AM. CERAM. SOC.,* VOL 73 (NO. 7), 1990, P 2073-2076
28. A.J. MOORHEAD AND H.-E. KIM, JOINING OXIDE CERAMICS, *CERAMICS AND GLASSES,* VOL 4, *ENGINEERED MATERIALS HANDBOOK,* ASM INTERNATIONAL, 1991, P 511-531
29. J.B. WACHTMAN, JR. AND R.A. HABER, ADVANCED CERAMICS INVOLVING ALUMINA, *ALUMINA CHEMICALS: SCIENCE AND TECHNOLOGY HANDBOOK,* L.D. HART, ED., AMERICAN CERAMIC SOCIETY, 1990, P 329-335
30. "MECHANICAL AND INDUSTRIAL CERAMICS," CAT/2T8505TKU/3000E, KYOCERA CORPORATION
31. "COORS CERAMICS--MATERIALS FOR TOUGH JOBS," BULLETIN NO. 953, COORS PORCELAIN COMPANY
32. R.F. DIEFENDORF, CONTINUOUS CARBON FIBER REINFORCED CARBON MATRIX COMPOSITES, *COMPOSITES,* VOL 1, *ENGINEERED MATERIALS HANDBOOK,* ASM INTERNATIONAL, 1987, P 911-914
33. R.L. BRONNES, R.C. HUGHES, AND R.C. SWEET, CERAMIC-TO-METAL BONDING WITH SPUTTERING AS A METALLIZATION TECHNIQUE, *PHILIPS TECH. REV.,* VOL 35, 1975, P 209-211
34. J.P. HAMMOND, S.A. DAVID, AND M.L. SANTELLA, BRAZING CERAMIC OXIDES TO METALS AT LOW TEMPERATURES, *WELD. J.,* VOL 7, 1988, P 227-S TO 232-S

35. P.O. JARVINEN, "NOVEL CERAMIC RECEIVER FOR SOLAR BRAYTON SYSTEMS," CONF-790305-7, MIT LINCOLN LABORATORY, 1979
36. F. HEIM AND H.L. MCDOWELL, "METHOD OF METALLIZING A CERAMIC SUBSTRATE," U.S. PATENT 4,342,632, AUGUST 3, 1982
37. S. WEISS AND C.M. ADAMS, JR., THE PROMOTION OF WETTING AND BRAZING, *WELD. J. SUPPL.,* VOL 46, 1967, P 49S-57S
38. E.F. BRUSH AND C.M. ADAMS, VAPOR COATED SURFACES FOR BRAZING CERAMICS, *WELD. J.,* VOL 47 (NO. 3), 1968, P 106-S TO 114-S
39. M.L. SANTELLA, BRAZING OF TITANIUM-VAPOR-COATED SILICON NITRIDE, *ADV. CERAM. MATER.,* VOL 3 (NO. 5), 1988, P 457-462
40. M.J. RAMSEY AND M.H. LEWIS, INTERFACIAL REACTION MECHANISMS IN SYALON CERAMIC BONDING, *MATER. SCI. ENG.,* VOL 71, 1985, P 113-122
41. J.P. HAMMOND AND G.M. SLAUGHTER, BONDING GRAPHITE TO METALS WITH TRANSITION PIECES, *WELD. J.,* VOL 50 (NO. 1), 1971, P 33-40
42. M.E. TWENTYMAN, PART 1. THE MECHANISM OF GLASS MIGRATION IN THE PRODUCTION OF METAL-CERAMIC SEALS, *J. MATER. SCI.,* VOL 10, 1975, P 765-776
43. M.E. TWENTYMAN AND P. POPPER, PART 2. THE EFFECT OF EXPERIMENTAL VARIABLES ON THE STRUCTURE OF SEALS TO DEBASED ALUMINAS, *J. MATER. SCI.,* VOL 10, 1975, P 777-790
44. M.E. TWENTYMAN AND P. POPPER, PART 3. THE USE OF METALLIZING PAINTS CONTAINING GLASS OR OTHER INORGANIC BONDING AGENTS, *J. MATER. SCI.,* VOL 10, 1975, P 791-798
45. A. KOHNO *ET AL.,* METALLIZING OF NON-OXIDE CERAMICS, *SPECIAL CERAMICS 8, BR. CERAM. PROC.,* NO. 37, S.P. HOWLETT AND D. TAYLOR, ED., OCT 1986, P 125-130
46. E.M. DUNN *ET AL.,* ANALYTICAL AND EXPERIMENTAL EVALUATION OF JOINING SILICON NITRIDE TO METAL AND SILICON CARBIDE TO METAL, *PROC. 26TH AUTOMOTIVE TECHNOLOGY DEVELOPMENT CONTRACTORS COORDINATION MEETING,* SAE P-219, SAE, 1988, P 165-171
47. YU.V. NAIDICH, THE WETTABILITY OF SOLIDS BY LIQUID METALS, *PROGRESS IN SURFACE AND MEMBRANE SCIENCE,* VOL 14, INSTITUTE OF MATERIAL PROBLEMS, ACADEMY OF SCIENCES, UKRAINIAN S.S.R., 1981, P 333-484
48. M.G. NICHOLAS, ACTIVE METAL BRAZING, *BR. CERAM. TRANS. J.,* VOL 85, 1986, P 144-146
49. M.G. NICHOLAS, T.M. VALENTINE, AND M.J. WAITE, THE WETTING OF ALUMINA BY COPPER ALLOYED WITH TITANIUM AND OTHER ELEMENTS, *J. MATER. SCI.,* VOL 15, 1980, P 2197-2220
50. H. MIZUHARA AND K. MALLY, CERAMIC-TO-METAL JOINING WITH ACTIVE BRAZING FILLER METAL, *WELD. J.,* VOL 64 (NO. 10), 1985, P 27-32
51. A.J. MOORHEAD AND H. KEATING, DIRECT BRAZING OF CERAMICS FOR ADVANCED HEAVY-DUTY DIESELS, *WELD. J.,* VOL 65 (NO. 10), 1986, P 17-31
52. M.G. NICHOLAS, PROMOTING THE WETTING AND BRAZING OF CERAMICS, *STRUCT. CERAM. JOINING II,* CERAMIC TRANSACTIONS, VOL 35, A.J. MOORHEAD, R.E. LOEHMAN, AND S.M. JOHNSON, ED., AMERICAN CERAMIC SOCIETY, 1993
53. YU.V. NAIDICH AND V.S. ZHURAVLEV, ADHESION, WETTABILITY AND REACTION OF TITANIUM-CONTAINING MELTS WITH HIGH-MELTING OXIDES, TRANSLATED FROM *OGNEUPORY,* VOL 1, 1974, P 50-55
54. T. ISEKI, H. MATSUZAKI, AND J.K. BOADI, BRAZING OF SILICON CARBIDE TO STAINLESS STEEL, *AM. CERAM. SOC. BULL.,* VOL 64 (NO. 2), 1985, P 322-324
55. R.E. LOEHMAN AND A.P. TOMSIA, *AM. CERAM. SOC. BULL.,* VOL 67 (NO. 2), 1988, P 375
56. M.J. RAMSEY AND M.H. LEWIS, *MATER. SCI. ENG.,* VOL 71, 1985, P 113-122
57. T. ISEKI, H. MATSUZAKI, AND J.K. BOADI, BRAZING OF SILICON CARBIDE TO STAINLESS

STEEL, *AM. CERAM. SOC. BULL.*, VOL 64 (NO. 2), 1985, P 322-324

58. M. NAKA *ET AL.*, JOINING OF SILICON NITRIDE WITH AL-CU ALLOYS, *J. MATER. SCI.*, VOL 22, 1987, P 4417-4421
59. T. ISEKI *ET AL.*, *CERAMIC SURFACES AND SURFACE TREATMENTS, BR. CERAM. PROC.*, NO. 34, R. MORRELL AND M.G. NICHOLAS, ED., AUGUST 1984, P 241-248
60. H. MIZUHARA AND T. OYAMA, CERAMIC-METAL SEALS, *CERAMICS AND GLASSES,* VOL 4, *ENGINEERED MATERIALS HANDBOOK,* ASM INTERNATIONAL, 1991, P 502-510
61. M.G. NICHOLAS, PROMOTING THE WETTING AND BRAZING OF CERAMICS, *STRUCT. CERAM. JOINING II,* CERAMIC TRANSACTIONS, VOL 35, A.J. MOORHEAD, R.E. LOEHMAN, AND S.M. JOHNSON, ED., AMERICAN CERAMIC SOCIETY, 1993
62. W.A. ZDANIEWSKI *ET AL.*, CRYSTALLIZATION TOUGHENING OF CERAMIC ADHESIVES FOR JOINING ALUMINA, *ADV. CERAM. MATER.*, VOL 2 (NO. 3A), 1987, P 204-208
63. S.L. SWARTZ *ET AL.*, JOINING OF ZIRCONIA CERAMICS WITH A CAO-TIO$_2$-SIO$_2$ INTERLAYER, *PROC. 26TH AUTOMOTIVE TECHNOLOGY DEVELOPMENT CONTRACTORS COORDINATION MEETING,* P-219, SAE, 1988, P 149-154
64. M.E. MILBERG *ET AL.*, THE NATURE OF SIALON JOINTS BETWEEN SILICON NITRIDE BASED BODIES, *J. MATER. SCI.*, VOL 22, 1987, P 2560-2568
65. S. BAIK AND R. RAJ, LIQUID-PHASE BONDING OF SILICON NITRIDE CERAMICS, *J. AM. CERAM. SOC.*, VOL 70 (NO. 5), 1987, P C105-C107
66. S.M. JOHNSON AND D.J. ROWCLIFFE, MECHANICAL PROPERTIES OF JOINED SILICON NITRIDE, *J. AM. CERAM. SOC.*, VOL 68 (NO. 9), 1985, P 468-472
67. N. IWAMOTO *ET AL.*, JOINING OF SILICON NITRIDE CERAMICS WITH CAO-TIO$_2$-SIO$_2$ GLASS SOLDER, *CERAMIC MATERIALS AND COMPONENTS FOR ENGINES, PROC. 2ND INT. SYMPOSIUM,* W. BUNK AND H. HAUSNER, ED., DEUTSCHE KERAMISCHE GESELLSCHAFT E.V., 1986, P 467-474
68. M.L. MECARTNEY *ET AL.*, SILICON NITRIDE JOINING, *J. AM. CERAM. SOC.*, VOL 68 (NO. 9), 1985, P 472-478
69. M. NEILSON, JR. AND D.N. COON, STRENGTH OF SILICON NITRIDE-SILICON NITRIDE JOINTS BONDED WITH OXYNITRIDE GLASS, *CERAM. ENG. SCI. PROC.*, VOL 10 (NO. 11-12), 1989, P 1893-1907
70. D.N. COON, R.L. TALLMAN, AND R.M. NEILSON, JR., HOT ISOSTATICALLY PRESSED SI$_3$N$_4$-SI$_3$N$_4$ JOINTS BONDED WITH OXYNITRIDE GLASS, *ADV. CERAM. MATER.*, VOL 3 (NO. 2), 1988, P 154-158
71. J.F. BURGESS, C.A. NEUGEBAUER, AND G. FLANAGAN, THE DIRECT BONDING OF METALS TO CERAMICS BY THE GAS-METAL EUTECTIC METHOD, *J. ELECTROCHEM. SOC.*, VOL 122 (NO. 5), 1975, P 688-690
72. Y. YOSHINO, ROLE OF OXYGEN IN BONDING COPPER TO ALUMINA, *J. AM. CERAM. SOC.*, VOL 72 (NO. 8), 1989, P 1322-1327
73. I. OKAMOTO, M. NAKA, K. ASAMI, AND K. HASHIMOTO, FORMATION OF HERCYNITE (FEAL$_2$O$_4$) AT INTERFACE OF AL$_2$O$_3$/STEEL JOINT, *TRANS. JWRI,* VOL 11 (NO. 1), 1982, P 131-133
74. M. NAKA AND I. OKAMOTO, JOINING OF AL$_2$O$_3$ TO SUS 304 WITH TI FOIL, *TRANS. JWRI,* VOL 15 (NO. 1), 1986, P 157-158
75. Y. IINO, PARTIAL TRANSIENT LIQUID-PHASE METALS LAYER TECHNIQUE OF CERAMIC-METAL BONDING, *J. MATER. SCI. LETT.*, VOL 10 (NO. 2), 1991, P 104-106
76. M.L. SHALZ, B.J. DALGLEISH, A.P. TOMSIA, AND A.M. GLAESER, NEW APPROACHES TO JOINING CERAMICS FOR HIGH-TEMPERATURE APPLICATIONS, *STRUCTURAL CERAM. JOINING II,* CERAMIC TRANSACTIONS, VOL 35, A.J. MOORHEAD, R.E. LOEHMAN, AND S.M. JOHNSON, ED., AMERICAN CERAMIC SOCIETY, 1993
77. D.S. DUVALL, W.A. OWCZARSKI, AND D.F. PAULONIS, TLP BONDING: A NEW METHOD FOR JOINING HEAT RESISTANT ALLOYS, *WELD. J.*, VOL 53 (NO. 4), 1974, P 203-214

78. D.H. KIM, S.H. HWANG, AND S.S. CHUN, THE WETTING, REACTION AND BONDING OF SILICON NITRIDE BY CU-TI ALLOYS, *J. MATER. SCI.,* VOL 26 (NO. 12), JUNE 1991, P 3223-3234
79. A.J. MOORHEAD AND H.-E. KIM, OXIDATION BEHAVIOUR OF TITANIUM-CONTAINING BRAZING FILLER METALS, *J. MATER. SCI.,* VOL 26, 1991, P 4067-4075
80. D.J. GREEN, R.H.J. HANNINK, AND M.V. SWAIN, *TRANSFORMATION TOUGHENING OF CERAMICS,* CRC PRESS, 1989, P 115-121
81. L. PEJRYD, METAL CERAMIC JOINING FOR HIGH TEMPERATURE APPLICATIONS, *4TH INT. SYMPOSIUM ON CERAMIC MATERIALS AND COMPONENTS FOR ENGINES,* R. CARLSSON, T. JOHANSSON, AND L. KAHLMAN, ED., ELSEVIER APPLIED SCIENCE, 1992, P 50-66
82. E.K. BEAUCHAMP AND S.N. BURCHETT, TECHNIQUES OF SEAL DESIGN, *CERAMICS AND GLASSES,* VOL 4, *ENGINEERED MATERIALS HANDBOOK,* ASM INTERNATIONAL, 1991, P 532-541
83. M.G. NICHOLAS AND R.M. CRISPIN, DIFFUSION BONDING STAINLESS STEEL TO ALUMINA USING ALUMINUM INTERLAYERS, *J. MATER. SCI.,* VOL 17, 1982, P 3347-3360
84. K. SUGANUMA, T. OKAMOTO, AND M. KOIZUMI, EFFECT OF INTERLAYERS IN CERAMIC-METAL JOINTS WITH THERMAL EXPANSION MISMATCHES, *J. AM. CERAM. SOC.,* VOL 67 (NO. 12), 1984, P C256-C257
85. A.-P. XIAN AND Z.-Y. SI, INTERLAYER DESIGN FOR JOINING PRESSURELESS SINTERED SIALON CERAMIC AND 40CR STEEL BRAZING WITH $AG_{57}CU_{38}TI_5$ FILLER METAL, *J. MATER. SCI.,* VOL 27, 1992, P 1560-1566
86. K. SUGANUMA, T. OKAMOTO, M. SHIMADA, AND M. KOIZUMI, NEW METHOD FOR SOLID-STATE BONDING BETWEEN CERAMICS AND METALS, *J. AM. CERAM. SOC.,* 1983, P C117-C118
87. B.H. RABIN AND R.J. HEAPS, POWDER PROCESSING OF $NI\text{-}AL_2O_3$ FGM, TO BE PUBLISHED IN *CERAMIC TRANSACTIONS, PROC. 2ND INT. SYMPOSIUM ON FUNCTIONALLY GRADIENT MATERIALS,* AMERICAN CERAMIC SOCIETY, 1993
88. J.F. KING *ET AL.,* MATERIALS TESTS AND ANALYSES OF FARADAY SHIELD TUBES FOR ICRF ANTENNAS, *FUSION TECHNOL.,* VOL 15 (NO. 2B), 1989, P 1093-1096
89. P.H. MCCLUSKEY, R.K. WILLIAMS, R.S. GRAVES, AND T.N. TIEGS, THERMAL DIFFUSIVITY AND CONDUCTIVITY OF ALUMINA-SILICON CARBIDE COMPOSITES, *J. AM. CERAM. SOC.,* VOL 73 (NO. 2), 1990, P 461-464
90. J.B. WACHTMAN, JR., ED., *STRUCTURAL CERAMICS, TREATISE ON MATERIALS SCIENCE AND TECHNOLOGY,* VOL 29, ACADEMIC PRESS, 1989

Brazing of Ceramic and Ceramic-to-Metal Joints

A.J. Moorhead and W.H. Elliott, Jr., Oak Ridge National Laboratory H.-E. Kim, Seoul National University

Selected References

- *CERAMICS AND GLASSES,* VOL 4, *ENGINEERED MATERIALS HANDBOOK,* ASM INTERNATIONAL, 1991

Application of Clad Brazing Materials

Michael Karavolis, Sunil Jha, and James Forster, Texas Instruments, Inc.; Kris Meekins, Long Manufacturing Ltd.

Introduction

BRAZING, which is one of the most versatile joining techniques, is the preferred method for making a large number of joints in an assembly. The manufacture of heat exchangers, for example, requires the creation of leak-free joints between specific points and along contact surfaces and seams. The fabrication of high-density heat exchangers, which have a large heat transfer area and a compact size, is best accomplished by brazing (Fig. 1).

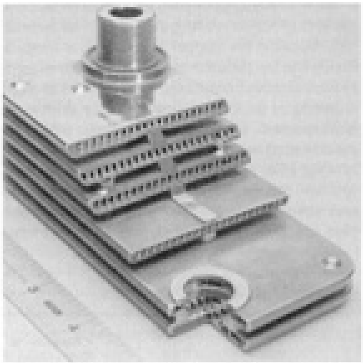

FIG. 1 AUTOMOTIVE ENGINE OIL COOLER MANUFACTURED WITH CLAD BRAZING MATERIALS

In a typical heat-exchanger assembly operation, the components are formed, cleaned, assembled, and fixtured prior to furnace brazing. During the assembly process, the brazing filler metal, which is often in the form of a wire, foil, paste, or powder sheet, is inserted at the appropriate locations. To simplify the assembly operation, the brazing filler metal can be clad to the materials to be brazed, as in aluminum-clad brazing sheet, where the brazing alloy is hot roll bonded to the aluminum alloy sheet (Ref 1). When the brazing filler metal is an integral part of the sheet or strip, the material is called a clad brazing material, because it can be brazed without the need for additional filler metal; that is, the brazing filler metal is already attached to the base material.

Although hot roll bonding is cited in the aluminum-clad brazing sheet example, many combinations of clad brazing materials are manufactured by cold roll bonding, a process that does not involve the explicit heating of the materials prior to the bonding operation. This material system is an economical solution for the fabrication of many high-volume, furnace-brazed products.

Additional information about the process involved in brazing with these materials is available in the article "Brazing With Clad Brazing Materials" in this Volume.

Acknowledgements

The authors thank the many people at Texas Instruments, Inc. who have been involved in the development and production of clad metals. In particular, we thank Jim Matthews, Jerry Johnson, Dick Delagi, and Russ Miller for their continued enthusiastic support and encouragement.

Reference

1. R. BROWN, ALUMINUM ALLOY LAMINATES: AL-CLAD AND CLAD ALUMINUM ALLOY PRODUCTS, *COMPOSITE ENGINEERING LAMINATES,* A.G.H. DIETZ, ED., MIT PRESS, 1969, P 227-239

Application of Clad Brazing Materials

Michael Karavolis, Sunil Jha, and James Forster, Texas Instruments, Inc.; Kris Meekins, Long Manufacturing Ltd.

Material Selection

The design of heat exchangers for automotive applications varies widely, as does the manufacturing method. The selection of the base metal and brazing filler metal used in a specific heat exchanger depends on several factors, the most important of which are the performance and cooling requirements of the unit. Typical design parameters include space allotted for the heat exchanger, maximum weight of the unit, burst-pressure resistance, cooling capacity, pressure drop, life expectancy, and cost. These factors, among others, influence the size, type, and number of brazed joints in the heat exchanger, and all must be considered.

The materials chosen for a particular heat-exchanger application can vary greatly. In terms of the performance requirements that typically represent the starting point in base metal and brazing filler metal selection, the major concerns are durability and corrosion resistance. Although the need for compatibility with hot and cold fluids is obvious, consideration must also be given to the other materials that the heat exchanger may contact, in order to avoid premature failure that is due to corrosion by galvanic coupling. Some under-the-hood automotive heat exchangers may be exposed to the environment, whereas those that are contained within the engine block are exposed to a totally different environment and temperature range.

Other design considerations include:

- TOTAL NUMBER AND TYPES OF JOINTS
- APPLICATION OF ADDITIONAL CONNECTORS ON THE EXTERIOR OF THE UNIT
- APPLICATION OF A GREATER AMOUNT OF BRAZING FILLER METAL IN CERTAIN AREAS THAN IN THE REST OF THE LOCATIONS TO BE BRAZED
- FIXTURING METHODS NEEDED TO ENSURE PROPER JOINT CONTACT AT THE BRAZING TEMPERATURE DURING THE BRAZING CYCLE

Many engine oil coolers are currently made from stainless steel and use copper as the brazing filler metal. This base metal and filler metal combination is also available as a clad brazing material, where a copper brazing filler metal is clad to a stainless steel base material. Although the AISI 300 and 400 series of stainless steels have been the most widely used materials, low-carbon steels are being used in applications with less-strict corrosion-resistance and strength requirements.

Precautions that must be exercised during the brazing cycle, when using clad brazing materials, are identical to those that must be used when brazing the monolithic metals. For example, low-carbon versions of the 300 series of stainless steels are recommended to reduce the potential for sensitization during cooling from the brazing temperature. An excessively slow cooling rate should be avoided, because it can result in carbide precipitation in the grain boundaries of the stainless steel base material.

For a similar reason, a stabilized grade of 400 stainless steel is recommended to avoid the embrittlement phenomenon that also can result from improper cooling down from the brazing temperature. Typically, low levels of titanium, niobium, or a combination of the two, provide sufficient protection against embrittlement by minimizing the migration of carbon and any dissolved gases, without seriously reducing the wettability of the stainless steel. The niobium grades should be preferred because they are easy to braze, when compared to titanium-stabilized grades.

The copper brazing filler metal that is typically used in these applications conforms to the American Welding Society (AWS) filler-metal specification AWS A5.B, Class BCu-1. Past processing experience has shown that this type of filler metal performs very well in brazed heat exchangers that utilize both AISI 300 and 400 series stainless steels as base materials.

Application of Clad Brazing Materials

Michael Karavolis, Sunil Jha, and James Forster, Texas Instruments, Inc.; Kris Meekins, Long Manufacturing Ltd.

Fabrication of Clad Brazing Materials

The clad brazing materials described here are metals that are metallurgically joined by cold roll bonding. The technology and science of roll bonding and solid-phase welding have been described in detail elsewhere (Ref 2, 3, 4), but are summarized below.

As shown in Fig. 2, the process involves three basic steps:

- CLEANING OF THE MATING SURFACES BY CHEMICAL AND/OR MECHANICAL MEANS TO REMOVE DIRT, LUBRICANTS, SURFACE OXIDES, AND ANY OTHER CONTAMINANTS
- JOINING OF THE MATERIALS IN A BONDING MILL BY ROLLING THEM TOGETHER WITH A THICKNESS REDUCTION THAT RANGES FROM 50 TO 80% IN A SINGLE PASS, IMMEDIATELY AFTER WHICH, THE MATERIALS HAVE AN INCIPIENT, OR GREEN, BOND CREATED BY THE MASSIVE COLD REDUCTION
- SINTERING, A HEAT TREATMENT DURING WHICH THE BOND AT THE INTERFACE IS COMPLETED. DIFFUSION OCCURS AT THE ATOMIC LEVEL ALONG THE INTERFACE AND RESULTS IN A METALLURGICAL BOND THAT IS DUE TO A SHARING OF ATOMS BETWEEN THE MATERIALS. THE RESULTING BOND CAN EXCEED THE STRENGTH OF EITHER OF THE PARENT MATERIALS

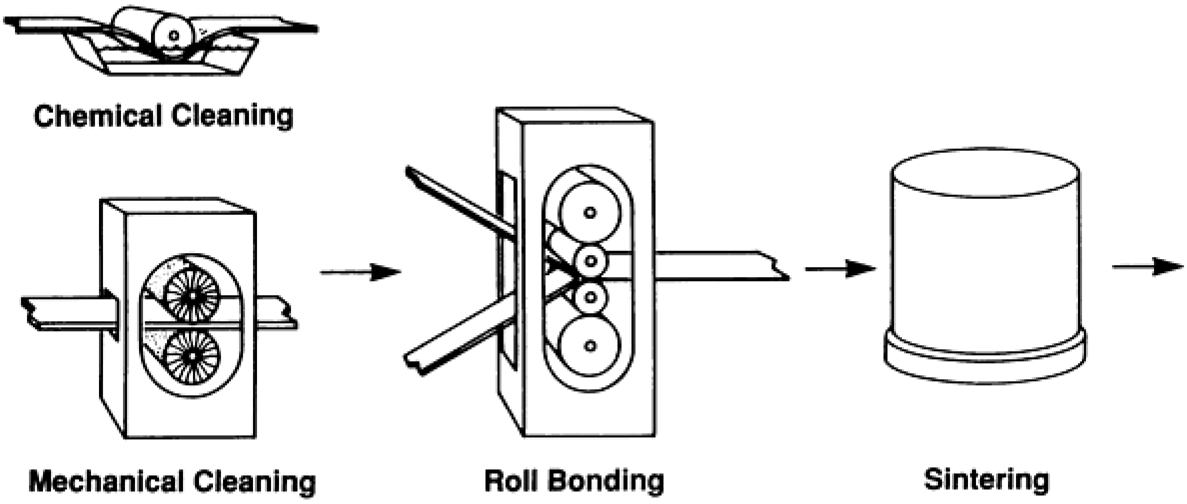

FIG. 2 PROCESS STEPS IN COLD ROLL BONDING

Upon completion of this three-step process, the resultant clad material can be treated in the same way as any other conventional monolithic metal. The clad material can be worked by any of the traditional processing methods for strip metals. Rolling, annealing, pickling, and slitting are typically performed to produce the finished strip to specific customer requirements, so that the material can be roll formed, stamped, or drawn into the required part.

Clad metals are defined by the materials that are present and the volume or thickness ratio of each layer. For example, a copper-clad stainless steel whose total thickness of 0.51 mm (0.020 in.) comprises 0.025 mm (0.001 in.) of copper, 0.46 mm (0.018 in.) of 301 stainless steel, and 0.025 mm (0.001 in.) of copper would be described as 5Cu/90/301SS/5Cu.

The thickness ratio can be determined by a number of techniques, including metallographic examination. A weigh/strip/weigh procedure can be used, where the material is weighed, the cladding is stripped by dissolving it in an appropriate solution such as an acid, and then the sample is weighed again. Another procedure utilizes the magnetic properties of the clad material. The volume ratio that is created during the roll-bonding operation remains through all subsequent finishing operations, including rolling. If appropriate starting material thicknesses and processing conditions are selected, then clad metals can be produced in an infinite number of layer ratios. It should be noted that it is not necessary that the cladding thickness be uniform on both sides of the core material, nor that the core material be clad on both sides. Nonsymmetrical material systems, as well as two-layer material systems, such as thermostat hi-metal, can also be produced. Figure 3 shows the cross section of a Cu/1008AK steel/Cu clad brazing material with a ratio of 5/90/5.

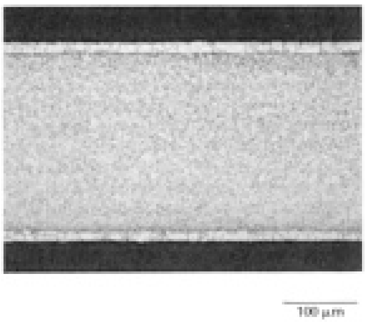

FIG. 3 PHOTOMICROGRAPH OF CU122/1008AK STEEL/CU122 CLAD BRAZING MATERIAL IN 5/90/5 RATIO, NO. 4 TEMPER

The amount of copper cladding is determined by the specific application and by the unit to be brazed. In areas where several hundred brazed joints can be made simultaneously, such as when a plate is joined to turbulator stock, a thicker copper layer (0.05 to 0.1 mm, or 0.002 to 0.004 in.) may be needed. In areas where only one or several joints are to be made, such as a can-to-lid or a plate-to-plate joint, a thinner copper layer (0.025 to 0.05 mm, or 0.001 to 0.002 in.) can be used. For applications in which a different number or different types of joints must be made on either side of a plate, different amounts of copper brazing filler metal can be clad on either side of the base metal. Nonsymmetrical layer ratio materials, such as 5/85/10 or 5/75/20, can be produced when needed.

References cited in this section

2. G. DURST, A NEW DEVELOPMENT IN METAL CLADDING, *J. MET.*, MARCH 1956
3. R.F. TYLECOTE, *THE SOLID PHASE WELDING OF METALS*, ST. MARTINS PRESS, 1968
4. R. DELAGI, DESIGNING WITH CLAD METALS, *MACH. DES.*, NOV 1980

Application of Clad Brazing Materials

Michael Karavolis, Sunil Jha, and James Forster, Texas Instruments, Inc.; Kris Meekins, Long Manufacturing Ltd.

Stamping, Cleaning, and Assembly

Clad metals should be considered to have the same forming characteristics as the core material. For example, a copper-clad 304L stainless steel system will have approximately the same ductility as the unclad 304L stainless steel. In some instances, the formability of the clad is better than that of the monolithic metal core because the copper on the surface of the stainless steel acts as a lubricant in the forming tool. The typical tensile properties of several copper-clad stainless steel clad brazing materials are shown in Tables 1 and 2.

TABLE 1 TWO-LAYER COPPER-CLAD BRAZING SYSTEMS

MATERIAL SYSTEM	**LAYER THICKNESS RATIO**	**TENSILE STRENGTH**		**0.2% YIELD STRENGTH**		ELONGATION IN 50 MM (2 IN.), %
		MPA	**KSI**	**MPA**	**KSI**	
C122/304LSS	6/94	590	86	255	37	**55**
	13.5/86.5	650	94	300	43	**55**
C122/409SS	**15/85**	**400**	**58**	**215**	**31**	**36**

TABLE 2 THREE-LAYER COPPER-CLAD BRAZING SYSTEMS

MATERIAL SYSTEM	**VOLUME RATIO**	**TENSILE STRENGTH**		**0.2% YIELD STRENGTH**		ELONGATION IN 50 MM (2 IN.),%
		MPA	**KSI**	**MPA**	**KSI**	
C122/304LSS/C122	10/80/10	600	87	310	45	**55**
	13/74/13	575	83	290	42	**53**
	32/34/32	380	55	170	25	**48**
C122/409SS/C122	10/80/10	385	56	205	30	**37**
	15/80/5	385	56	205	30	**37**
C122/1008AK/C122	**10/80/10**	**295**	**43**	**215**	**31**	**39**

Notes: All properties shown are typical values for annealed materials, except for Cu122/1008AK/Cu122, which is in the No. 4 temper. Recommended brazing temperature for CDA-122 (AWS A5.8, Class B Cu-1) copper brazing filler metal is 1095 to 1150 °C (2000 to 2100 °F). Brazing atmosphere should be appropriate for the unclad core material.

The tooling that is used for stamping parts from clad metals can be the same as that used to make the monolithic metal core. Because the brazing filler metal is an integral part of the core material, additional stamping and forming operations associated with the production of the shim material are eliminated.

The parts must be cleaned after stamping, prior to assembly. The procedure for the clad metal parts should be similar to that used for cleaning the brazing filler metal shims or preforms. Copper-clad materials should be cleaned by a procedure that is appropriate for copper pans. Because the copper cladding is metallurgically bonded to the stainless steel core, wetting by the copper is ensured, as shown in Fig. 4, and the cleaning of any stainless steel parts is eliminated.

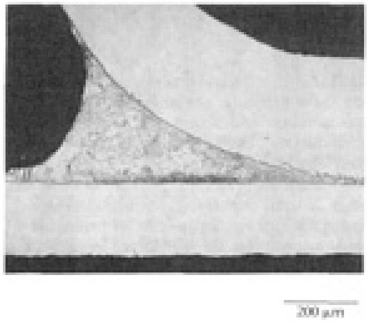

FIG. 4 PHOTOMICROGRAPH OF A TYPICAL BRAZED JOINT THAT WAS FORMED USING A CLAD BRAZING MATERIAL

As with any copper part, care must be taken to prevent oxidation during assembly prior to brazing. Because the copper brazing filler metal is bonded to the stainless steel, the number of pans to be assembled is significantly reduced and any handling of the thin, delicate brazing shim stock is eliminated. Unlike preform shims, which are usually used as flat, two-dimensional inserts, the brazing filler metal on the clad brazing material follows the complex surface topography that was stamped into the base material. Two examples are shown in Fig. 5 and 6. The former depicts a typical plate-type heat exchanger made from a clad brazing material, whereas the latter shows typical contoured joints that can be made easily with clad brazing materials.

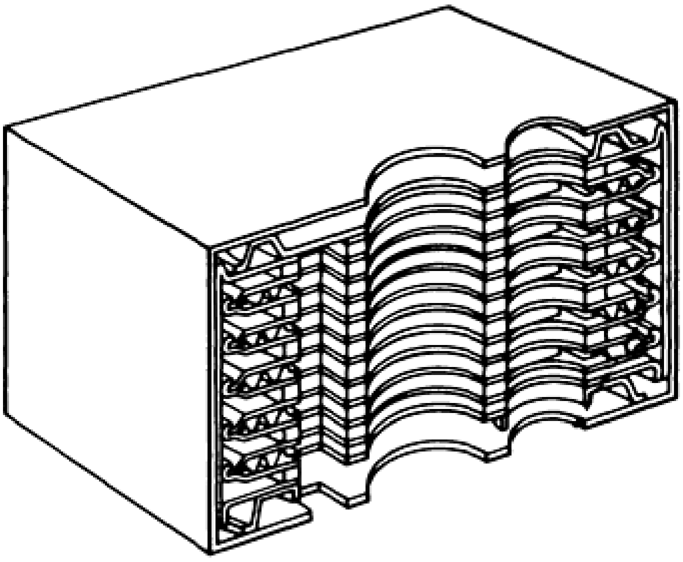

FIG. 5 CUTAWAY VIEW OF A TYPICAL PLATE-TYPE HEAT EXCHANGER

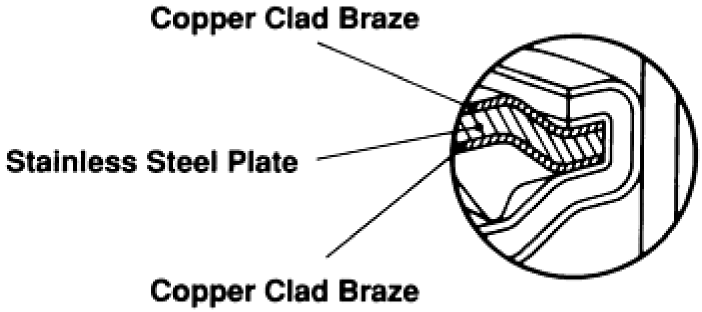

FIG. 6 CONTOURED JOINTS THAT CAN BE MADE EASILY WITH CLAD BRAZING MATERIALS

One design benefit of using clad brazing materials is a reduction of the clearance between the parts, because no allowance needs to be made for the insertion of thick shims, wires, or preforms.

Another benefit is that less copper brazing filler metal can be used, because the concerns associated with the handling of shim material have been eliminated. The thinner copper coating thickness, which typically ranges from 0.025 to 0.075 mm (0.001 to 0.003 in.), results in less shrinkage during brazing and is less likely to cause large brazing fillets. Consequently, the possibility that excess brazing filler metal will clog small passageways is greatly reduced. The typical joint shown in Fig. 7 illustrates the small-geometry passages that can be achieved and the capillary action of the brazing filler metal as it forms the fillets at the joints.

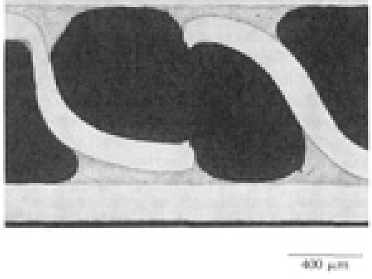

FIG. 7 PHOTOMICROGRAPH OF A PASSAGEWAY IN A HEAT EXCHANGER FABRICATED USING COPPER-CLAD STAINLESS STEEL CLAD BRAZING MATERIAL

Yet another benefit of using clad metals is that once the parts are assembled and fixtured, the brazing filler metal will not shift, move, or fall out as a result of the movement of fixtured assemblies prior to or during the brazing cycle.

Brazing Parameters

Brazing in a vacuum furnace is recommended for clad brazing materials, although the use of a protective gas atmosphere is also possible. Because stainless steel is typically used as the core material of heat exchangers, and because dissociation of the chromium oxide on the surface is essential, vacuum is the preferred brazing atmosphere. In many ways, vacuum furnace brazing with clad metals is similar to the vacuum brazing of stainless steel plates with copper shims. However, there are key differences. Furnace pressures from 0.0015 to 13.5 Pa (10^{-5} to 10^{-1} torr) are typical of the vacuum requirements for brazing with a stainless steel cored clad brazing material (Ref 5). The specific stainless steel alloy will determine the vacuum pressure that is required. No fluxes are needed, and the brazing operation can be considered to be a "clean" process.

Because the brazing filler metal is a pure, deoxidized copper, no additional time is needed to allow any binders to burn off or to outgas during the preheat stage. The heating rate is limited by the need to ensure a uniform temperature throughout the furnace load. Because the clad filler metal is metallurgically bonded to the core material, the wetting and rapid flow of the filler material is ensured. For these reasons, production costs can be lowered by reducing the total cycle time. Temperature control of ±5 °C (± 10 °F) is desired to ensure the proper flow of the brazing filler metal and to limit any filler metal-base metal penetration.

A fast cooling rate will help minimize any filler metal-base metal reactions, as well as production costs. Fast cooling rates are preferred for 300 series stainless steels, whereas slow cooling rates are preferred for the 400 series. The use of low-carbon 300 series and stabilized 400 series stainless steels, as described earlier, should result in a uniform, nonsensitized, non-embrittled core material.

Reference cited in this section

5. *BRAZING MANUAL,* 3RD ED., AWS, 1976, P 66

References

1. R. BROWN, ALUMINUM ALLOY LAMINATES: AL-CLAD AND CLAD ALUMINUM ALLOY PRODUCTS, *COMPOSITE ENGINEERING LAMINATES,* A.G.H. DIETZ, ED., MIT PRESS, 1969, P 227-239
2. G. DURST, A NEW DEVELOPMENT IN METAL CLADDING, *J. MET.,* MARCH 1956
3. R.F. TYLECOTE, *THE SOLID PHASE WELDING OF METALS,* ST. MARTINS PRESS, 1968

4. R. DELAGI, DESIGNING WITH CLAD METALS, *MACH. DES.*, NOV 1980
5. *BRAZING MANUAL*, 3RD ED., AWS, 1976, P 66

General Soldering

Paul T. Vianco, Sandia National Laboratories

Introduction

SOLDERING TECHNOLOGY has been used in applications that range from the packaging of integrated circuit chips to the fabrication of industrial heat exchangers. Consequently, soldering applications are informally categorized as being either structural or electronic in nature. Regardless of these categories, soldering technology depends on certain underlying fundamental parameters and the issues associated with each. This article provides information about these topics:

- *SOLDER ALLOY.* A VARIETY OF ALLOYS AND THEIR PROPERTIES ARE DESCRIBED IN TERMS OF THE SELECTION PROCESS.
- *SUBSTRATE MATERIAL.* ALTHOUGH THE SELECTION OF THE SUBSTRATE BASE MATERIAL IS GENERALLY DETERMINED BY CRITERIA OTHER THAN THE JOINING TECHNIQUE, COATINGS CAN BE USED TO ENHANCE SOLDERABILITY.
- *FLUX.* FLUX SELECTION IS PRIMARILY BASED ON PROMOTING ADEQUATE WETTABILITY BY THE SOLDER. HOWEVER, THE NEED FOR POSTASSEMBLY CLEANING PROCEDURES MUST BE ADDRESSED WHEN THE CORROSIVITY OF THE FLUX RESIDUES JEOPARDIZES THE RELIABILITY REQUIRED OF THE WORKPIECE DURING SERVICE.
- *COMBINED SUBSTRATE, SOLDER, AND FLUX PROPERTIES.* THE COMBINATION OF PROPERTIES MUST BE CONSIDERED IN ORDER TO ESTABLISH THE JOINT DESIGN AND MANUFACTURING PROCEDURES THAT ARE NECESSARY TO SUCCESSFULLY CONSTRUCT THE WORKPIECE.

Each of these parameters is fully explored in this article, using examples of both structural and electronic applications. The article "Soldering in Electronic Applications" explores the latter topic in further detail.

Acknowledgements

This work was performed at Sandia National Laboratories and was supported by the U.S. Department of Energy under contract DE-AC04-76DP00789.

General Soldering

Paul T. Vianco, Sandia National Laboratories

Solder (Filler-Metal) Alloys

The primary criterion used to select a solder is its melting properties. Alloy selection is also determined by the heat sensitivity of the substrate material, the temperature conditions that are expected during service, and the pasty range of the alloy. The melting characteristics of solders are expressed as the solidus and the liquidus temperatures for the alloy. Solders composed of one metallic element, such as tin or indium, have a single melting temperature. Alloys with a eutectic composition have liquidus and solidus temperatures, but they are coincident, and the temperature value is referred to as the "eutectic temperature." The temperature spread between the solidus and liquidus temperatures define the pasty range of the alloy. Solders are commercially available with liquidus temperatures that are as low as 11 °C (51 °F), for the ternary gallium-indium-tin alloy, to as high as 425 °C (795 °F), for the germanium-aluminum solder. Metal alloys with melting temperatures that exceed 450 °C (812 °F) are categorized as brazing filler metals. The pasty region can range from only a few degrees to nearly 100 °C (180 °F).

Other attributes of the solder include mechanical properties, such as strength and creep resistance; physical properties, such as thermal expansion or thermal conductivity; corrosion resistance; and cost. Because the solder provides a means of fastening two substrates together, its mechanical properties are of great importance. Besides their monotonic strength (tensile and shear) to resist overload stresses, solders must possess sufficient creep strength under sustained loads. This factor is particularly critical, because typical service environments (including room temperature) are actually high-temperature applications, relative to the low melting temperatures of the alloy. The resistance of the solder to fatigue deformation is important to both structural and electronic applications. The coefficient of thermal expansion (CTE) mismatch between the solder and substrate(s) can directly affect the long-term integrity of the solder joint. Thermal conductivity is important for joints that must meet heat-transfer specifications. Corrosion must be considered at all stages of solder joint design, because solder joints have dissimilar metals in intimate contact. Many solders are microstructurally heterogeneous and have other corrosion concerns.

Specifications and standards have established requirements for certain solder alloys in terms of composition limits, impurity levels, and a nomenclature for referencing particular compositions. In the United States, the specifications of solder alloy compositions are found in ASTM B 32 (Ref 1) and Federal Specification QQ-S-571E (Ref 2). Users in the European community prefer the International Organization for Standardization (ISO) specification (draft form) ISO/DIS 9453 (Ref 3). Candidate alloys must be submitted to the particular governing organization for consideration and acceptance as a "specified composition." It should be noted that the commercial availability of a solder material does not necessarily imply that it is included in these and other federal, military, or industrial standards and specifications. In the case of nonstandardized compositions, the user should insist that the vendors certify the composition (including impurities) of their products. Many government and military projects contractually require that the supplier or manufacturer used specified solder compositions in their products.

Although the U.S. standards are generally recognized by foreign industries, the expansion of trade with overseas companies will require that U.S. producers and manufacturers gain an increased familiarity with international standards.

The physical and mechanical properties of bulk solders and solder joints are discussed below. When designing a product, solder joint data are preferred to bulk solder properties. Generally, the use of bulk strength data will underestimate the loading capacity of the joint; that is, calculations will typically be conservative. Although they are not always available, the solder joint properties should include the joint thickness and the testing rate used to obtain the values. Solder selection is also guided by cost. The 1992 prices of metals commonly used in solder alloys are shown in Table 1.

TABLE 1 PRICES OF METALS COMMONLY USED IN SOLDERS

METAL	**$/KG**	$/LB
IN	242.50	**110.00**
AG	132.30	**60.00**
BI	6.60	**3.00**
SN	4.85	**2.20**
CU	2.90	**1.30**
SB	2.00	**0.90**
PB	**0.80**	**0.35**

Prices shown are from 1992.

Source: Ref 4

Tin-lead, tin-lead-antimony, tin-lead-silver, and lead-silver solder compositions, per ASTM B 32, are listed in Table 2. The values are similar in the QQ-S-571E specification. The ISO/DIS 9453 guidelines are given in Table 3. Antimony may be found in solders for any one of three reasons: as an impurity, that is, not intentionally added to the material; as a minor addition of 0.25 wt%, minimum, to try to prevent the formation of the low-temperature allotrope of tin (α of tin at 13.2 °C, or 55.8 °F); and as a principal component, representing 1 to 5 wt%, to improve monotonic and creep strength of the solder. The levels of added antimony should be less than 7 wt% of the tin component of the solder (the solubility limit of antimony in tin) in order to prevent the formation of tin-antimony intermetallic compounds, which can severely impact the fluidity of the molten solder or reduce the ductility of the solidified joint. The ASTM designation for antimony levels in the solders are: Class A (alloys with tin content greater than 35 wt%), maximum of 0.12 wt%; Class B, 0.2 to 0.5 wt%; and Class C (tin content, 20 to 40 wt%), maximum of 6 wt% (Ref 5).

TABLE 2 ASTM B 32 SPECIFICATION FOR TIN-LEAD, TIN-LEAD-ANTIMONY, TIN-LEAD-SILVER, AND LEAD-SILVER SOLDERS

ALLOY GRADE	**COMPOSITION, %(A)**											MELTING RANGE(B)			
												SOLIDUS		LIQUIDUS	
	Sn	**Pb**	**Sb**	**Ag**	**Cu**	**Cd**	**Al**	**Bi**	**As**	**Fe**	**Zn**	**°C**	**°F**	**°C**	°F
SN70	69.5-71.5	BAL	0.50	0.015	0.08	0.001	0.005	0.25	0.03	0.02	0.005	183	361	193	**377**
SN63	62.5-63.5	BAL	0.50	0.015	0.08	0.001	0.005	0.25	0.03	0.02	0.005	183	361	183	**361**
SN62	61.5-62.5	BAL	0.50	1.75-2.25	0.08	0.001	0.005	0.25	0.03	0.02	0.005	179	354	189	**372**
SN60	59.5-61.5	BAL	0.50	0.015	0.08	0.001	0.005	0.25	0.03	0.02	0.005	183	361	190	**374**
SN50	49.5-51.5	BAL	0.50	0.015	0.08	0.001	0.005	0.25	0.025	0.02	0.005	183	361	216	**421**
SN45	44.5-46.5	BAL	0.50	0.015	0.08	0.001	0.005	0.25	0.025	0.02	0.005	183	361	227	**441**
SN40A	39.5-41.5	BAL	0.50	0.015	0.08	0.001	0.005	0.25	0.02	0.02	0.005	183	361	238	**460**
SN40B	39.5-41.5	BAL	1.8-2.4	0.015	0.08	0.001	0.005	0.25	0.02	0.02	0.005	185	365	231	**448**
SN35A	34.5-36.5	BAL	0.50	0.015	0.08	0.001	0.005	0.25	0.02	0.02	0.005	183	361	247	**447**
SN35B	34.5-36.5	BAL	1.6-2.0	0.015	0.08	0.001	0.005	0.25	0.02	0.02	0.005	185	365	243	**470**
SN30A	29.5-31.5	BAL	0.50	0.015	0.08	0.001	0.005	0.25	0.02	0.02	0.005	183	361	255	**491**
SN30B	29.5-31.5	BAL	1.4-1.8	0.015	0.08	0.001	0.005	0.25	0.02	0.02	0.005	185	365	250	**482**
SN25A	24.5-26.5	BAL	0.50	0.015	0.08	0.001	0.005	0.25	0.02	0.02	0.005	183	361	266	**511**
SN25B	24.5-26.5	BAL	1.1-1.5	0.015	0.08	0.001	0.005	0.25	0.02	0.02	0.005	185	365	263	**504**
SN20A	19.5-21.5	BAL	0.50	0.015	0.08	0.001	0.005	0.25	0.02	0.02	0.005	183	361	277	**531**
SN20B	19.5-21.5	BAL	0.8-1.2	0.015	0.08	0.001	0.005	0.25	0.02	0.02	0.005	184	363	270	**517**
SN15	14.5-16.5	BAL	0.50	0.015	0.08	0.001	0.005	0.25	0.02	0.02	0.005	225	437	290	**554**
SN10A	9.0-11.0	BAL	0.50	0.015	0.08	0.001	0.005	0.25	0.02	0.02	0.005	268	514	302	**576**
SN10B	9.0-11.0	BAL	0.20	1.7-2.4	0.08	0.001	0.005	0.03	0.02	0.02	0.005	268	514	299	**570**
SN5	4.5-5.5	BAL	0.50	0.015	0.08	0.001	0.005	0.25	0.02	0.02	0.005	308	586	312	**594**
SN2	1.5-2.5	BAL	0.50	0.015	0.08	0.001	0.005	0.25	0.02	0.02	0.005	316	601	322	**611**
AG1.5	0.75-1.25	BAL	0.40	1.3-1.7	0.30	0.001	0.005	0.25	0.02	0.02	0.005	309	588	309	**588**
AG2.5	0.25	BAL	0.40	2.3-2.7	0.30	0.001	0.005	0.25	0.02	0.02	0.005	304	580	304	**580**
AG5.5	**0.25**	**BAL**	**0.40**	**5.0-6.0**	**0.30**	**0.001**	**0.005**	**0.25**	**0.02**	**0.02**	**0.005**	**304**	**580**	**380**	**716**

(A) LIMITS ARE MAXIMUM PERCENTAGES, UNLESS SHOWN AS A RANGE OR STATED OTHERWISE. FOR PURPOSES OF DETERMINING CONFORMANCE TO THESE LIMITS, AN OBSERVED VALUE OR CALCULATED VALUE OBTAINED FROM ANALYSIS SHALL BE ROUNDED TO THE NEAREST UNIT IN THE LAST RIGHT-HAND PLACE OF FIGURES USED IN EXPRESSING THE SPECIFIED LIMIT, IN ACCORDANCE WITH THE ROUNDING METHOD OF ASTM RECOMMENDED PRACTICE E 29.

(B) TEMPERATURES GIVEN ARE APPROXIMATIONS AND ARE FOR INFORMATION ONLY.

TABLE 3 ISO/DIS 9453 SPECIFICATION FOR TIN-LEAD, TIN-LEAD-ANTIMONY, TIN-LEAD-SILVER, AND LEAD-SILVER SOLDERS

ALLOY NO.	**ALLOY DESIGNATION**	**MELTING OR SOLIDUS/LIQUIDUS TEMPERATURE (A)**		**CHEMICAL COMPOSITION, %(B)**											SUM OF ALL IMPURITIES EXCEPT SB, BI, AND CU
		°C	**°F**	**Sn**	**Pb**	**Sb**	**Cd**	**Zn**	**Al**	**Bi**	**As**	**Fe**	**Cu**	**Ag**	
TIN-LEAD ALLOYS															
1	S-SN63PB37	183	360	62.5-63.5	BAL	0.12	0.002	0.001	0.001	0.10	0.03	0.02	0.05	...	**0.08**
1A	E-SN63PB37	183	360	62.5-63.5	BAL	0.05	0.002	0.001	0.001	0.05	0.03	0.02	0.05	...	**0.08**
2	S-SN60PB40	183-190	360-375	59.5-60.5	BAL	0.12	0.002	0.001	0.001	0.10	0.03	0.01	0.05	...	**0.08**
2A	E-SN60PB40	183-190	360-375	59.5-60.5	BAL	0.05	0.002	0.001	0.001	0.05	0.03	0.02	0.05	...	**0.08**
3	S-PB50SN50	183-215	360-420	49.5-50.5	BAL	0.12	0.002	0.001	0.001	0.10	0.03	0.02	0.05	...	**0.08**
3A	E-PB50SN50	183-215	360-420	49.5-50.5	BAL	0.05	0.002	0.001	0.001	0.05	0.03	0.02	0.05	...	**0.08**
4	S-PB55SN45	183-226	360-440	44.5-45.5	BAL	0.50	0.005	0.001	0.001	0.25	0.03	0.02	0.08	...	**0.08**
5	S-PB60SN40	183-235	360-455	39.5-40.5	BAL	0.50	0.005	0.001	0.001	0.25	0.03	0.02	0.08	...	**0.08**
6	S-PB65SN35	183-245	360-470	34.5-35.5	BAL	0.50	0.005	0.001	0.001	0.25	0.03	0.02	0.08	...	**0.08**
7	S-PB70SN30	183-255	360-490	29.5-30.5	BAL	0.50	0.005	0.001	0.001	0.25	0.03	0.03	0.08	...	**0.08**
8	S-PB90SN10	268-302	515-575	9.5-10.5	BAL	0.50	0.005	0.001	0.001	0.25	0.03	0.02	0.08	...	**0.08**
9	S-PB92SN8	280-305	535-580	7.5-8.5	BAL	0.50	0.005	0.001	0.001	0.25	0.03	0.02	0.08	...	**0.08**
10	S-PB98SN2	320-325	610-620	1.5-2.5	BAL	0.12	0.002	0.001	0.001	0.10	0.03	0.02	0.05	...	**0.08**
TIN-LEAD ALLOYS WITH ANTIMONY															
11	S-SN63PB37SB	183	360	62.5-63.	BAL	0.12-0.5	0.002	0.001	0.001	0.10	0.03	0.02	0.05	...	**0.08**

				5		0								
12	S-SN60PB40SB	183-190	360-375	59.5-60.5	BAL	0.12-0.50	0.002	0.001	0.001	0.10	0.03	0.02	0.05	...	**0.08**
13	S-PB50SN50SB	183-216	360-420	49.5-50.5	BAL	0.12-0.50	0.002	0.001	0.001	0.10	0.03	0.02	0.05	...	**0.08**
14	S-PB58SN40SB2	185-231	365-450	39.5-40.5	BAL	2.0-2.4	0.005	0.001	0.001	0.25	0.03	0.02	0.08	...	**0.08**
15	S-PB69SN30SB1	185-250	360-480	29.5-30.5	BAL	0.5-1.8	0.005	0.001	0.001	0.25	0.03	0.02	0.08	...	**0.08**
16	S-PB74SN25SB1	185-263	360-505	24.5-25.5	BAL	0.5-2.0	0.005	0.001	0.001	0.25	0.03	0.02	0.08	...	**0.08**
17	S-PB78SN20SB2	185-270	365-520	19.5-20.5	BAL	0.5-3.0	0.005	0.001	0.001	0.25	0.03	0.02	0.08	...	**0.08**
LEAD-SILVER AND LEAD-TIN-SILVER ALLOYS														
32	S-PB98AG2	304-305	580-581	0.25	BAL	0.10	0.002	0.001	0.001	0.10	0.03	0.02	0.05	2.0-3.0	**0.2**
33	S-PB95AG5	304-365	580-690	0.25	BAL	0.10	0.002	0.001	0.001	0.10	0.03	0.02	0.05	4.5-6.0	**0.2**
34	**S-PB93SN5AG2**	**296-301**	**560-570**	**4.8-5.2**	**BAL**	**0.10**	**0.002**	**0.001**	**0.001**	**0.10**	**0.03**	**0.02**	**0.05**	**1.2-1.8**	0.2

Source: Ref 3

(A) TEMPERATURES GIVEN ARE FOR INFORMATION PURPOSES AND ARE NOT SPECIFIED REQUIREMENTS FOR THE ALLOYS.

(B) ALL SINGLE FIGURE LIMITS ARE MAXIMUM PERCENTAGES.

The thick, tenacious oxide of lead and the corrosion product layers that form on tin limit the extent of general corrosion to tin-lead solders. However, the lead-rich phase is particularly susceptible to corrosion in the presence of chloride ions, water, or carbon dioxide (from the atmosphere). The reaction by-product from the latter two is a white powder, identified as lead carbonate (Ref 6). The poor adhesion of lead carbonate to the lead causes it to drop away (known as "spalling"), exposing new metal to corrosion, until the solder film becomes a skeleton of tin-rich crystals.

The high-lead solders, which have a 5 to 20 wt% Sn content, are used extensively in micro-electronics packaging (flipchip or C^4 technology) or in high-temperature applications. Other applications include radiator assembly and automobile body repair, which utilize torch heating techniques. A drop in the liquidus temperature is obtained when tin content increases from 20 to 30 wt%. The 5 to 20 wt% Sn alloys have a wider pasty range that provides greater workability, whereas the higher tin content improves the wetting and flow of the solder. On the other hand, increased concentrations of tin also raise the price of the solder.

The most widely used solders are the eutectic 63Sn-37Pb and the near-eutectic alloys, 60Sn-40Pb and 50Sn-50Pb. These solders are used in electronics and structural applications, such as the assembly of conduit for nonpotable water, industrial fluids, compressed gas products, and vacuum assemblies. The narrow pasty range of these alloys limits their workability prior to solidification. The largest use of the eutectic and near-eutectic solder is in the assembly of circuit boards for electronic products.

The alloys with higher tin contents of 65 to 100 wt% have more limited applications, because the melting temperatures and pasty range needed for structural applications can be provided by the less-expensive lead-rich compositions. However, the tin-rich alloys have improved wettability for hard-to-solder substrates.

The physical and mechanical properties of bulk tin-lead solders are given in Tables 4, 5, 6, 7, 8, and 9. The room-temperature physical properties of the tin-lead solders, ranked by tin content, are listed in Table 4. The CTE decreases with increasing tin content, whereas both thermal and electrical conductivities rise with tin content. Surface tension and viscosity data, which are indicative of capillary filling in holes and gaps by the solders, are listed in Table 5. The addition of flux will tend to *lower* the surface tension of the molten solder.

TABLE 4 PHYSICAL PROPERTIES OF TIN-LEAD SOLDERS

TIN CONTENT, WT%	**DENSITY, G/CM^3**	**ELECTRICAL CONDUCTIVITY OF COPPER, %IACS$^{(A)}$**	**THERMAL CONDUCTIVITY**		COEFFICIENT OF LINEAR THERMAL EXPANSION, 10 $^{-6}$/K
			W/M · K	**BTU · IN./S · FT^2 · °F**	
0	11.34	7.9	34.8	0.067	**29.3**
5	10.80	8.1	35.3	0.068	**28.4**
10	10.50	8.2	35.8	0.069	**27.9**
20	10.04	8.7	37.4	0.072	**26.5**
30	9.66	9.3	40.5	0.078	**25.6**
40	9.28	10.1	43.6	0.084	**24.7**
50	8.90	10.9	46.7	0.090	**23.6**
60	8.52	11.5	49.8	0.096	**21.6**
63	8.34	11.8	50.9	0.098	**21.2**
70	8.17	12.5	. . .	. . .	**20.7**
100	**7.29**	**15.6**	**73.2**	**0.141**	**23.5**

(A) IACS, INTERNATIONAL ANNEALED COPPER STANDARD

TABLE 5 SURFACE TENSION AND VISCOSITY OF TIN-LEAD SOLDERS

TIN CONTENT, WT%	**TEMPERATURE**		**SURFACE TENSION**		VISCOSITY	
	°C	**°F**	**10^3 N/M**	**LBF/FT**	**PA · S**	LB · S/FT^2
100	260	500	548	0.0378	42.3$^{(A)}$	**0.884$^{(A)}$**
80	280	536	514	0.0354	44.4	**0.928**
63	280	536	490	0.0338	45.6	**0.952**
50	280	536	476	0.0328	50.7	**1.058**
42	280	536	474	0.0327	53.0	**1.107**
30	280	536	470	0.0324	56.7	**1.184**
20	280	536	467	0.0322	62.9	**1.314**
0	**390**	**734**	**439**	**0.0302**	**56.5**	**1.179**

Source: Ref 7

(A) COMPUTED ESTIMATE AT 260 °C (500 °F) FROM DATA AT 232 °C (450 °F) AND 351 °C (665 °F).

TABLE 6 MECHANICAL PROPERTIES OF TIN-LEAD SOLDERS

TIN CONTENT, WT%	**TENSILE STRENGTH**		**SHEAR STRENGTH**		**ELONGATION, %**	**MODULUS OF ELASTICITY**		**HARDNESS, HB**	**IZOD IMPACT STRENGTH**		STRESS TO PRODUCE A CREEP STRAIN RATE OF 10^{-4}/DAY	
	MPA	**KSI**	**MPA**	**KSI**		**GPA**	**10^6 PSI**		**J**	**FT · LBF**	**MPA**	KSI
0	12	1.8	12	1.8	55	18	2.61	4	8.1	6	1.7	**0.250**
5	28	4.0	14	2.1	45			8	9.5	7	1.4	**0.200**
10	30	4.4	17	2.4	30	19	2.76	I0	11	8	. . .	**. . .**
20	33	4.8	21	3.0	20	20	2.90	11	15	11	. . .	**. . .**
30	34	5.0	28	4.0	18	21	3.05	12	16	12	0.80	**0.115**
40	37	5.4	32	4.6	25	23	3.34	12	19	14	. . .	**. . .**
50	41	6.0	36	5.2	35	. . .	. . .	14	20	15	0.86	**0.125**
60	52	7.6	39	5.6	40	30	4.35	16	20	15	. . .	**. . .**
63	54	7.8	37	5.4	37	. . .	. . .	17	20	15	2.3	**0.335**
70	**54**	**7.8**	**36**	**5.2**	**30**	**35**	**5.08**	**17**	**19**	**14**	**. . .**	**. . .**

TABLE 7 MECHANICAL PROPERTIES OF TIN-LEAD SOLDERS AT CRYOGENIC TEMPERATURES

TIN CONTENT, WT%	**TENSILE STRENGTH**		**SHEAR STRENGTH**		ELONGATION, %
	MPA	**KSI**	**MPA**	**KSI**	
TESTED AT -73 °C (-100 °F)					
10	43	5.9	31	4.5	**34**
20	48	7.0	37	5.3	**32**
40	48	6.9	40	5.8	**43**
60	59	8.5	54	7.9	**48**
TESTED AT -196 °C (-320 °F)					
10	59	8.6	43	6.3	**27**
20	85	12.0	58	8.4	**30**
40	87	13.0	77	11.0	**30**
60	**130**	**19.9**	**110**	**16.0**	**10**

Source: Ref 9

TABLE 8 MECHANICAL PROPERTIES OF TIN-LEAD-ANTIMONY SOLDERS

COMPOSITION, WT%			**TENSILE STRENGTH**		**SHEAR STRENGTH**		**IZOD IMPACT STRENGTH**		**ELONGATION, %**	STRESS FOR CREEP STRAIN RATE OF 10^{-3}/DAY	
SN	**PB**	**SB**	**MPA**	**KSI**	**MPA**	**KSI**	**J**	**FT · LBF**		**MPA**	KSI
30	69	1.0	46	6.6	30	4.4	15.3	11.3	21	2.0	**0.30**
40	57.5	2.5	49	7.1	37	5.3	14.1	10.4	34	2.9	**0.42**
50	47	3.0	52	7.5	42	6.1	15.0	11.1	29	3.3	**0.48**
60	**36.4**	**3.6**	**61**	**8.8**	**42**	**6.1**	**16.0**	**11.5**	**18**	**3.3**	**0.48**

TABLE 9 CREEP LIFE DATA FOR BULK TIN-LEAD AND TIN-LEAD-ANTIMONY SOLDERS

COMPOSITION, WT%			STRESS AT INDICATED TEMPERATURE					
			AT 20 °C (68 °F)		**AT 100 °C (212 °F)**		AT 149 °C (300 °F)	
SN	**PB**	**SB**	**MPA**	**KSI**	**MPA**	**KSI**	**MPA**	KSI
FOR CREEP LIFE OF 1000 H								
10	90	. . .	35.0	5.08	1.10	0.159	. . .	**. . .**
40	58	2	49.0	7.11	5.90	0.856	. . .	**. . .**
40	60	. . .	21.0	3.05	4.20	0.609	. . .	**. . .**
60	40	. . .	29.0	4.21	4.50	0.653	. . .	**. . .**
FOR CREEP LIFE OF 10 YEARS								
5	95	. . .	3.4	0.50	1.7	0.250	1.0	**0.150**
10	90	. . .	3.2	0.47	1.4	0.20	0.69	**0.100**
20	80	. . .	2.5	0.36	0.83	0.12	0.34	**0.05**
30	70	. . .	2.1	0.30	0.62	0.09	0.21	**0.03**
40	60	. . .	1.8	0.26	0.52	0.075	0.21	**0.03**
50	50	. . .	1.7	0.25	0.52	0.075	0.21	**0.03**
60	**40**	**. . .**	**1.7**	**0.25**	**0.52**	**0.075**	**0.21**	**0.03**

The mechanical properties of the tin-lead alloys are listed in Table 6. The tensile strength values were obtained at a strain rate of 0.5/min. The tensile and shear strength values of solders are very sensitive to the rate of testing, and they typically decrease with decreasing test rate. In fact, the percent decrease in strength is significantly *larger* at the elevated temperatures.

The room-temperature strength properties are a function of the time period between alloy solidification and testing. Annealing of the test sample (at room or higher temperatures) should be performed to stabilize the microstructure to ensure consistent, reproducible data. Tensile strength data of 63Sn-37Pb suggest that stability is reached by room-temperature aging for 60 days (Ref 8). Recommended high-temperature-aging treatments to accelerate stabilization have not been specified. The use of high-temperature annealing should be accompanied by microstructural evaluations to ensure that the aging mechanism is similar to that which exists at room temperature.

Strength data for the tin-lead alloys at cryogenic temperatures of -73 °C (-100 °F) and -196 °C (-320 °F) are included in Table 7. It is important to note that alloy ductility drops sharply at the lower temperatures. The Brinell hardness values of the tin-lead solders as a function of composition are given in Ref 9.

The mechanical properties of the tin-lead-antimony solders are listed in Table 8. Antimony additions of 2 to 4 wt% increase both the tensile and shear strengths and the creep resistance of the tin-lead alloys. Similar behavior is recorded with the addition of 1 to 3 wt% Ag, as occurs with the Sn62 (ASTM B 32) alloy. Impact strength is not significantly altered by antimony additions. Ductility varies with antimony content, reaching a maximum at approximately 2.5 wt% Sb.

Testing temperatures affect mechanical properties significantly. General trends can be obtained from the tensile strength and ductility data shown in Fig. 1 for 60Sn-40Pb and 90Pb-10Sn solders. Strength decreases and ductility increases as the test temperature increases. The tensile strength trend is not monotonic for the 60Sn-40Pb and 90Pb-10Sn solders. At approximately -129 °C (-200 °F), strength reaches a maximum.

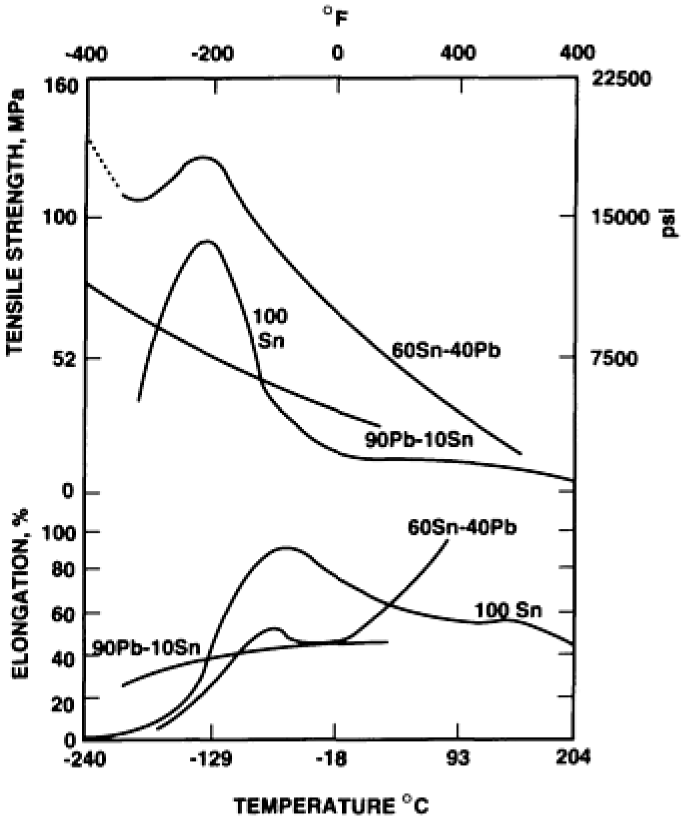

FIG. 1 TENSILE PROPERTIES OF BULK SOLDERS AS A FUNCTION OF TESTING TEMPERATURE

Table 9 gives the maximum creep stress values required to yield a life of 1000 h at temperatures of 20 °C (68 °F) and 100 °C (212 °F), as well as the maximum stress levels for the 10-year life of bulk tin-lead solders, as a function of tin content.

These data indicate that for a constant creep life, the stress value must decrease as the tin content increases. The addition of antimony improves the creep strength.

The fatigue life of bulk 60Sn-40Pb solder is shown in Fig. 2, whereas that of bulk 60Sn-40Pb solder as a function of testing rate is depicted in Fig. 3.

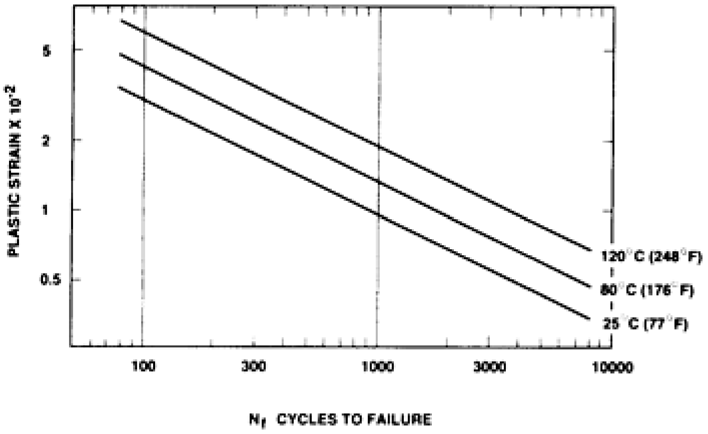

FIG. 2 ISOTHERMAL FATIGUE LIFE (PLASTIC STRAIN DEPENDENCE) OF BULK 60SN-40PB SOLDER. SOURCE: REF 10

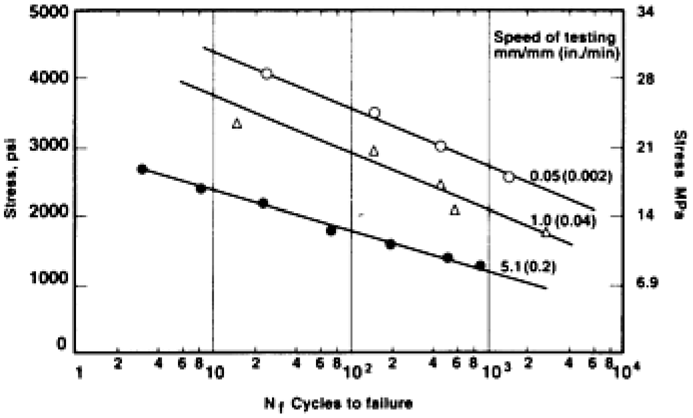

FIG. 3 ROOM-TEMPERATURE FATIGUE LIFE (STRESS DEPENDENCE) OF BULK 60SN-40PB SOLDER

The tin-lead-antimony solders demonstrate room-temperature aging of the microstructure and corresponding changes in the mechanical properties (Ref 11). Tensile strength data for the 60Sn-40Pb alloy following room-temperature aging are shown in Fig. 4. Strength and hardness do not decrease monotonically with time for alloys that have >3 wt% Sb. As shown in Fig. 5, a maximum in hardness is observed in the data for 47Sn-47Pb-6Sb.

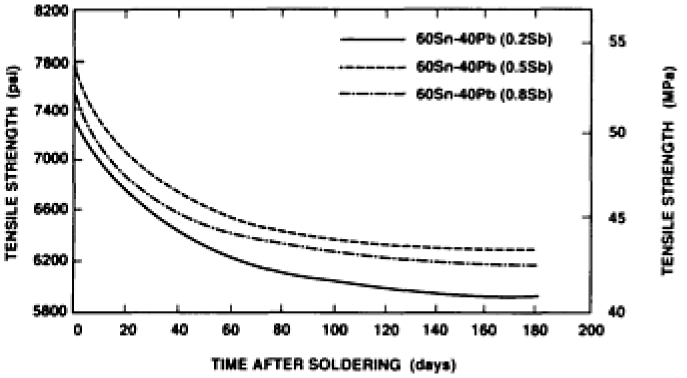

FIG. 4 TENSILE STRENGTH OF 60SN-40PB SOLDER AFTER ROOM-TEMPERATURE AGING. SOURCE: REF 8

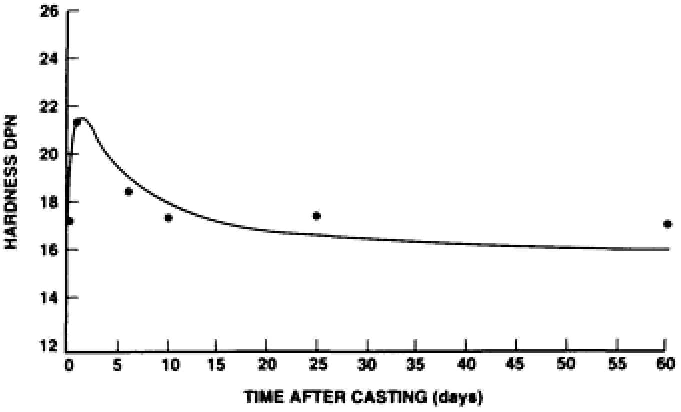

FIG. 5 HARDNESS OF 47SN-47PB-6SB SOLDER AFTER ROOM-TEMPERATURE AGING. SOURCE: REF 8

The effect of joint thickness on shear strength of copper substrates and 56Sn-44Pb solder is shown in Fig. 6. Typically, the maximum strength of solder joints is achieved with a thickness that ranges from 0.076 to 0.18 mm (3 to 7 mils).

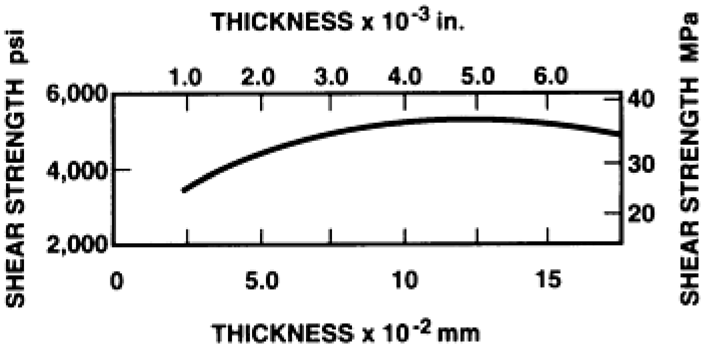

FIG. 6 SHEAR STRENGTH OF COPPER/56SN-44PB/COPPER SOLDER JOINTS AS A FUNCTION OF JOINT THICKNESS. SOURCE: REF 12

The primary sample geometries for the mechanical testing of solder joints are shown in Fig. 7. The appropriate ASTM test specifications for each configuration are D 1002, for the single-lap shear test; D 3528, for the double-lap shear test; F 1044, for the ring-in-plug test; D 897, for the tensile strength of joints; and D 1876, for the T-peel test. The sample geometries test the joints in shear, the predominant loading mode for soldered joints. These test methods, along with those for measuring the creep and fatigue of adhesive joints, are listed in Table 10. The ring-in-plug test can be used for shear testing by tension (pulling the plug out of the ring) and by torsion (twisting).

TABLE 10 TEST METHODS FOR ADHESIVE JOINT MECHANICAL PROPERTIES

ASTM DESIGNATION(A)	DESCRIPTION
D 897-78	**STANDARD TEST METHOD FOR TENSILE PROPERTIES OF ADHESIVE BONDS**
D 1002-72	**STANDARD TEST METHOD FOR STRENGTH PROPERTIES OF ADHESIVES IN SHEAR BY TENSION LOADING (SINGLE-LAP SHEAR)**
D 1876-72	**STANDARD TEST METHOD FOR PEEL RESISTANCE OF ADHESIVES T-PEEL TEST**
D 3528-76	**STANDARD TEST METHOD FOR STRENGTH PROPERTIES OF DOUBLE-LAP SHEAR ADHESIVE JOINTS BY TENSION LOADING**
F 1044-87	**STANDARD TEST METHOD FOR SHEAR TESTING OF POROUS METAL COATING (RING-IN-PLUG)**
D 2294-69	**STANDARD TEST METHOD FOR CREEP PROPERTIES OF ADHESIVES IN SHEAR BY TENSION LOADING**
D 3166-73	**STANDARD TEST METHOD FOR FATIGUE PROPERTIES OF ADHESIVES IN SHEAR BY TENSION LOADING**
E 8-89	**STANDARD TEST METHOD FOR TENSION TESTING OF METALLIC MATERIALS (BULK MATERIAL)**
E 143-87	**STANDARD TEST METHOD FOR SHEAR MODULUS AT ROOM TEMPERATURE**

(A) PER VOL 15.06, *ANNUAL BOOK OF ASTM STANDARDS,* 1990

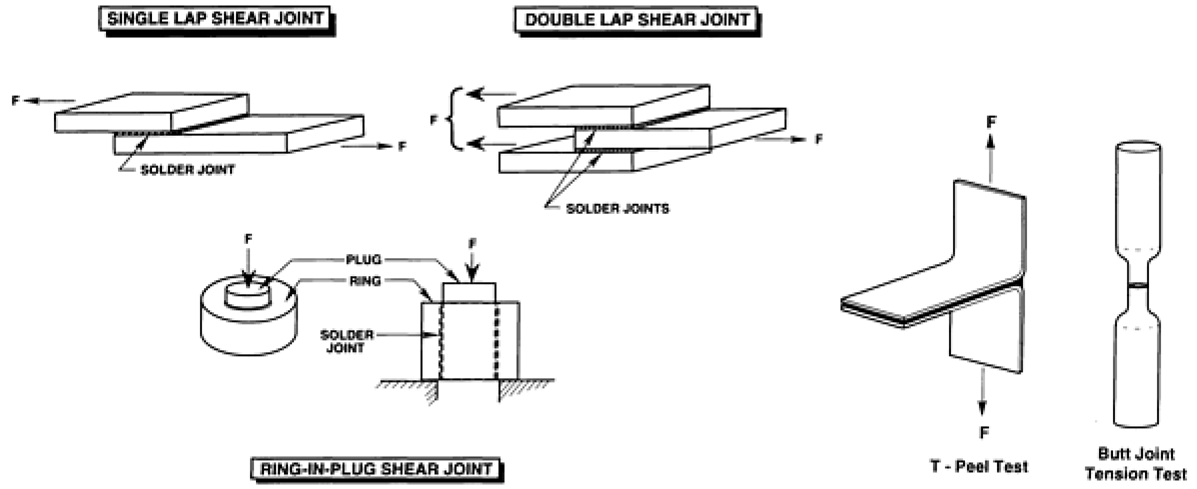

FIG. 7 MECHANICAL TEST SAMPLE GEOMETRIES

Additional properties that are specific to the eutectic 63Sn-37Pb and near-eutectic 60Sn-40Pb solders are provided below. The properties of either group of these solders can be approximated with those of the other. The physical properties of 60Sn-40Pb are listed in Table 11, whereas creep-rupture data are given in Table 12.

TABLE 11 PHYSICAL PROPERTIES OF 60SN-40PB SOLDER

PROPERTY	VALUE
SOLIDUS TEMPERATURE, °C (°F)	**183 (361)**
LIQUIDUS TEMPERATURE, °C (°F)	**188 (370)**
DENSITY, G/CM^3	**8.52**

TEMPERATURE DEPENDENCE OF DENSITY, °C/(G/CM^3)	**9.079 ± (9.708 × 10^4)**
SPECIFIC HEAT CAPACITY, J/KG · K	**150**
LATENT HEAT OF FUSION, J/KG × 10^3 (BTU/LB)	**37 (16)**
POISSON'S RATIO AT 25 °C (77 °F)$^{(A)}$	**0.4**

Source: Ref 13

(A) FOR 63SN-37PB

TABLE 12 CREEP-RUPTURE STRENGTH OF 60SN-40PB SOLDER

TEMPERATURE		**STRESS**		LIFE, H
°C	**°F**	**MPA**	**KSI**	
20	68	2.0	0.29	**2750**
		2.9	0.42	**670**
		3.9	0.57	**270**
		4.9	0.71	**180**
		5.9	0.86	**60**
		6.9	1.0	**70**
		8.8	1.3	**30**
100	212	0.2	0.029	**2860 ± 1710**
		0.8	0.120	**210 ± 40**
		1.0	0.150	**100 ± 50**
		1.2	**0.170**	**70 ± 70**

Source: Ref 13

The temperature dependence of the tensile strength, yield strength, and ductility of bulk 63Sn-37Pb solder at a test rate of 1.3 mm/min (0.05 in./min) is shown in Fig. 8; the temperature dependence of the tensile modulus and Poisson's ratio is shown in Fig. 9. These properties are particularly useful for finite-element modeling of solder joint residual stresses.

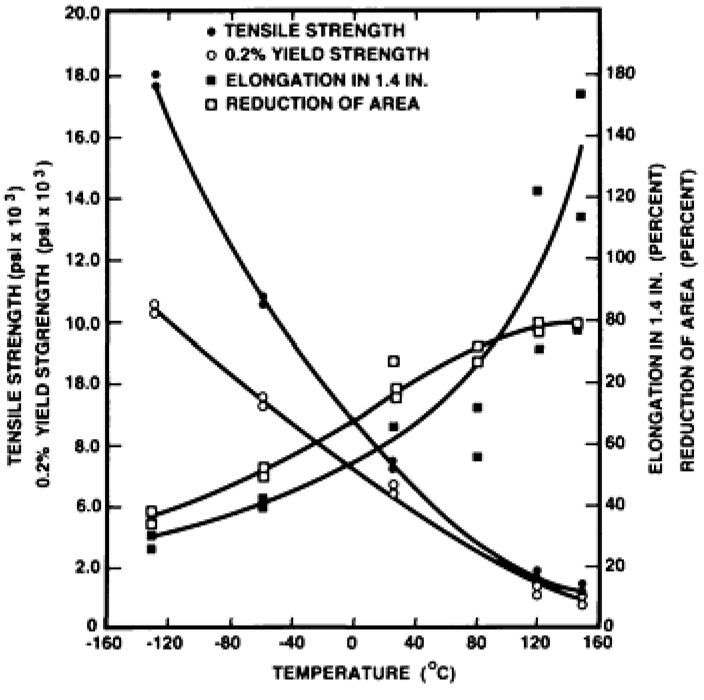

FIG. 8 TEMPERATURE DEPENDENCE OF TENSILE STRENGTH, YIELD STRENGTH, AND DUCTILITY OF 63SN-37PB

SOLDER. SOURCE: REF 14

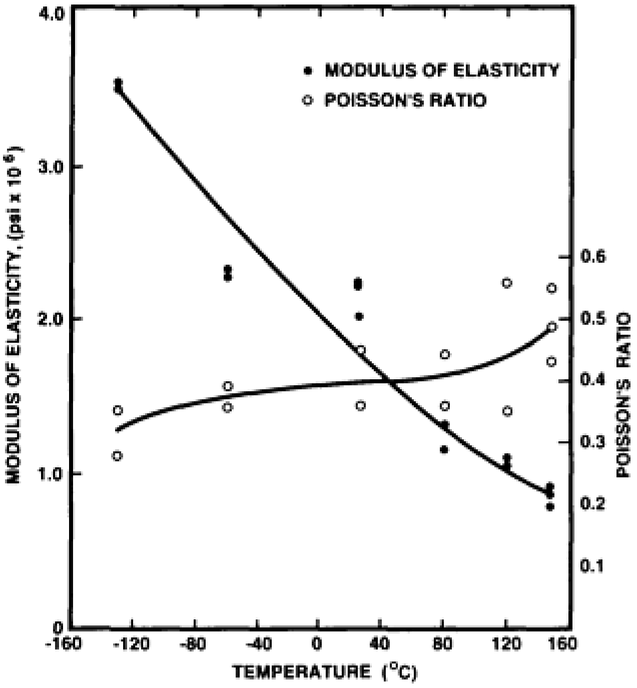

FIG. 9 TEMPERATURE DEPENDENCE OF TENSILE MODULUS AND POISSON'S RATIO OF 63SN-37PB SOLDER. SOURCE: REF 14

The room-temperature shear strength (copper ring-in-plug and single-overlap shear) of 60Sn-40Pb solder joints, as well as data from tests performed at elevated temperatures, are shown in Table 13. The single-lap joint gap thickness was 0.13 mm (0.005 in.). Thermal aging of the solder joints decreases the joint shear strength, and the effect on room-temperature data becomes much greater at higher strain rates (Ref 13).

TABLE 13 ROOM-TEMPERATURE AND ELEVATED-TEMPERATURE SHEAR STRENGTHS OF 60SN-40PB SOLDER

TEST SPEED		SHEAR STRENGTH					
		AT 20 °C (68 °F)		**AT 100 °C (212 °F)**		AT 150 °C (302 °F)	
mm/min	**in./min**	**MPA**	**KSI**	**MPA**	**KSI**	**MPA**	KSI
RING-IN-PLUG TEST							
50	2.0	49.7 ± 7.4	7.21 ± 1.07	33.3	4.83	. . .	**. . .**
20	0.8	47.8 ± 5.3	6.94 ± 0.77	31.2	4.53	23.4 ± 0.2	**3.40 ± 0.03**
5	0.2	41.9 ± 1.3	6.08 ± 0.19	25.5	3.70	18.1	**2.63**
1	0.04	33.6 ± 5.1	4.88 ± 0.74	21.6	3.13	9.7 ± 0.9	**1.41 ± 0.13**
0.2	0.008	26.1 ± 0.5	3.79 ± 0.07	15.7	2.28	4.7	**0.068**
0.05	0.002	21.3 ± 1.4	3.09 ± 0.20	12.2	1.77	4.2 ± 0.3	**0.061 ± 0.04**
SINGLE-LAP TEST$^{(A)}$							
50	2.0	35.4	5.14	. . .	. . .	. . .	**. . .**
20	0.8	37.1	5.38	. . .	. . .	. . .	**. . .**
5	0.2	32.2	4.67	. . .	. . .	. . .	**. . .**
1	0.04	30.1	4.37	. . .	. . .	. . .	**. . .**
0.2	0.08	26.2	3.80	. . .	. . .	. . .	**. . .**
0.05	**0.002**	**23.2**	**3.37**	**. . .**	**. . .**	**. . .**	**. . .**

Source: Ref 13

(A) OVERALL AREA. 48-51 MM^2 (0.07-0.08 IN.^{2}).

Fatigue life data expressed as cycles to failure for a given load stress are presented in Fig. 3. Fatigue resistance can also be expressed as a function of the applied strain. Lap shear fatigue data from a variety of applied strain levels are shown in Fig. 10. Increasing the temperature above 20 °C (68 °F) generally decreases the number of cycles to failure. Isothermal fatigue life is also reduced by decreasing the frequency of cycles. The frequency effect is negligible for strains that exceed 4% and frequencies that exceed 0.01 Hz (Ref 16).

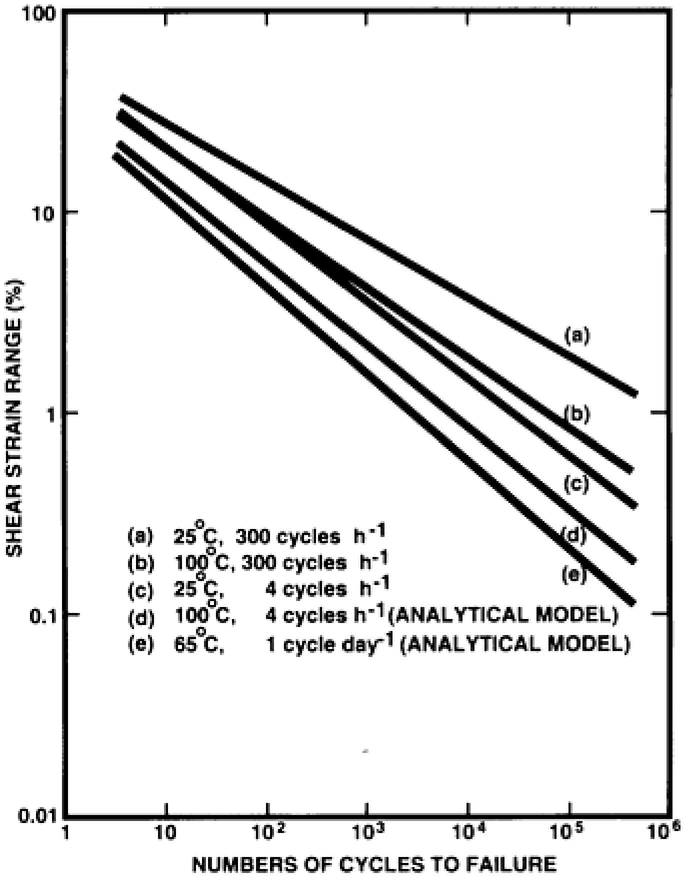

FIG. 10 FATIGUE LIFE OF 60SN-40PB SOLDER AS FUNCTION OF APPLIED STRAIN AMPLITUDE. SOURCE: REF 15

Resistance to creep diminishes as the temperature increases. Creep-rupture data at three temperatures are shown in Table 14.

TABLE 14 CREEP-RUPTURE DATA FOR 60SN-40PB COPPER RING-IN-PLUG JOINTS

STRESS		LIFE, H
MPA	**KSI**	
AT 20 °C (68 °F)		
3.4	0.49	**1050**
3.9	0.57	**610 ± 320**
4.9	0.71	**480**
5.9	0.86	**220**

7.8	1.13	**70**
AT 100 °C (212 °F)		
0.9	0.13	**2060**
1.1	0.16	**490**
1.3	0.19	**270**
1.5	0.22	**270**
1.7	0.25	**60**
AT 150 °C (302 °F)		
0.6	0.09	**130 ± 70**
1.0	0.15	**11 ± 8**
1.5	0.22	**5**
2.4	**0.35**	**0.4**

Source: Ref 13

Tin-Antimony and Tin-Antimony-Silver Solders. Tin-antimony solder compositions, as defined by three standards, are given in Tables 15, 16, and 17. The 95Sn-5Sb solder is used for most applications and is the only composition included in the tables.

TABLE 15 TIN-ANTIMONY SOLDER COMPOSITIONS PER ASTM B 32

ALLOY GRADE	**COMPOSITION, %(A)**											MELTING RANGE(B)			
												SOLIDUS		LIQUIDUS	
	Sn	**Pb**	**Sb**	**Ag**	**Cu**	**Cd**	**Al**	**Bi**	**As**	**Fe**	**Zn**	**°C**	**°F**	**°C**	°F
SB5	94.0 MIN	0.20	4.5-5.5	0.015	0.08	0.03	0.005	0.15	0.05	0.04	0.005	233	450	240	**464**
HA(C)	**BAL**	**0.10**	**0.5-4.0**	**0.1-3.0**	**0.1-2.0**	**0.005**	**0.005**	**0.15**	**0.05**	**0.02**	**0.5-4.0**	**216**	**420**	**227**	**440**

(A) LIMITS ARE % MAX, UNLESS SHOWN AS A RANGE OR STATED OTHERWISE. FOR PURPOSES OF DETERMINING CONFORMANCE TO THESE LIMITS, AN OBSERVED VALUE OR CALCULATED VALUE OBTAINED FROM ANALYSIS SHALL BE ROUNDED TO THE NEAREST UNIT IN THE LAST RIGHT-HAND PLACE OF FIGURES USED IN EXPRESSING THE SPECIFIED LIMIT, IN ACCORDANCE WITH THE ROUNDING METHOD OF PRACTICE E 29.

(B) TEMPERATURES GIVEN ARE APPROXIMATIONS AND FOR INFORMATION ONLY.

(C) DESIGNATION FOR SN-SB-AG-CU ALLOY COMPOSITION.

TABLE 16 TIN-ANTIMONY SOLDER COMPOSITION PER FEDERAL SPECIFICATION QQ-S-571E

ALLOY	**COMPOSITION, %**												APPROXIMATE MELTING RANGE(A)			
													SOLIDUS		LIQUIDUS	
	Sn	**Pb**	**Sb**	**Bi max**	**Ag**	**Cu max**	**Fe max**	**Zn max**	**Al max**	**As max**	**Cd max**	**TOTAL OF ALL OTHERS MAX**	**°C**	**°F**	**°C**	°F
SB5	**94.0 MIN**	**0.20 MAX**	**4.0-6.0**	**. . .**	**. . .**	**0.08**	**0.08**	**0.03**	**0.03**	**0.05**	**0.03**	**0.03**	**235**	**455**	**240**	**465**

(A) FOR INFORMATION ONLY

TABLE 17 TIN-ANTIMONY SOLDER COMPOSITION PER ISO/DIS 9453

ALL	**ALLOY**	**MELTING**	CHEMICAL COMPOSITION, %

OY NO.	**DESIGNATION**	**OR SOLIDUS/ LIQUIDUS TEMPERATURE**														
		°C	**°F**	**Sn**	**Pb**	**Sb**	**Bi**	**Cd**	**Cu**	**In**	**Ag**	**Al**	**As**	**Fe**	**Zn**	SUM OF ALL IMPURITIES
18	**S-SN95SB5**	**230-240**	**445-465**	**BAL**	**0.10**	**4.5-5.5**	**0.10**	**0.002**	**0.10**	**0.05**	**0.05**	**0.001**	**0.03**	**0.02**	**0.001**	**0.2**

Selected properties of the 95Sn-5Sb alloy are:

DENSITY, G/CM^3	**7.25**
HARDNESS, HB	**15**
ELECTRICAL CONDUCTIVITY, %IACS(A)	**12**
YOUNG'S MODULUS, GPA (10^6 PSI)	**49.99 (7.255)**

Source: Ref 8

(A) IACS, INTERNATIONAL ANNEALED COPPER STANDARD.

Alloys with 1 to 5 wt% Sb are commercially available. Antimony remains in solid solution in tin at levels up to 6 or 7 wt%. Higher amounts cause the precipitation of tin-antimony intermetallic crystals.

The tin-antimony-silver solder is certified in the ASTM specification. The silver improves the spreading of the solder, lowers the solidus and liquidus temperature, and extends the alloy pasty range. Tin-antimony solders exhibit excellent monotonic and creep strength levels for applications with high structural loads and elevated-temperature service, as well as in food handling equipment and potable water conduit and containers. Antimony additions above 0.3 wt% will deteriorate wetting and spreading. Solders with more than 0.2 to 0.5 wt% Sb can become embrittled when joined to aluminum, zinc, or zinc-coated parts because of the formation of aluminum-antimony and zinc-antimony intermetallic phases.

The room-temperature physical and mechanical properties of 95Sn-5Sb alloy are given in Table 18. The sensitivity of the room-temperature bulk tensile strength and copper ring-in-plug shear strength to testing rate (Table 19) is less than that for the tin-lead alloys. The tensile strength and elongation of tin-antimony solders also exhibit a smaller sensitivity to temperature (Fig. 11).

TABLE 18 PHYSICAL AND MECHANICAL PROPERTIES OF BULK 95SN-5SB SOLDER

PROPERTY	VALUE
DENSITY, G/CM^3	**7.25**
ELASTIC MODULUS, GPA (10^6 PSI)	**50 (7.26)**
ELONGATION, %	**38**
HARDNESS, HB	**15**
ELECTRICAL CONDUCTIVITY, %IACS	**11.9**
COEFFICIENT OF THERMAL EXPANSION, 10^{-6}/K	**27**

TABLE 19 TENSILE STRENGTH OF BULK 95SN-5SB COPPER RING-IN-PLUG JOINTS VERSUS TEST

SPEED

TENSILE STRENGTH(A)				**TEST SPEED**		SHEAR STRENGTH(B)			
AT 20 °C (68 °F)		**AT 100 ° (212 °F)**				**AT 20 °C (68 °F)**		AT 100 °C (212 °F)	
MPA	**KSI**	**MPA**	**KSI**	**MM/MIN**	**IN./MIN**	**MPA**	**KSI**	**MPA**	KSI
47.5	6.89	30.4	4.41	50.00	2.0	53.9	7.82	28.3	**4.11**
46.5	6.75	28.6	4.15	20.00	0.8	48.0	6.97	29.0	**4.21**
42.6	6.18	25.7	3.73	5.00	0.2	44.6	6.47	23.5	**3.41**
36.3	5.27	22.8 ± 2.6(C)	3.31 ± 0.38(C)	1.00	0.04	37.2	5.40	21.1	**3.06**
26.0	3.77	24.2	3.51	0.20	0.008	31.4	4.56	18.8	**2.73**
30.0	**4.35**	**19.9**	**2.89**	**0.05**	**0.002**	**27.4**	**3.98**	**13.2**	**1.92**

Source: Ref 13

(A) OF BULK MATERIAL.

(B) OF COPPER RING-IN-PLUG.

(C) PLUS OR MINUS ONE STANDARD DEVIATION, $N = 2$.

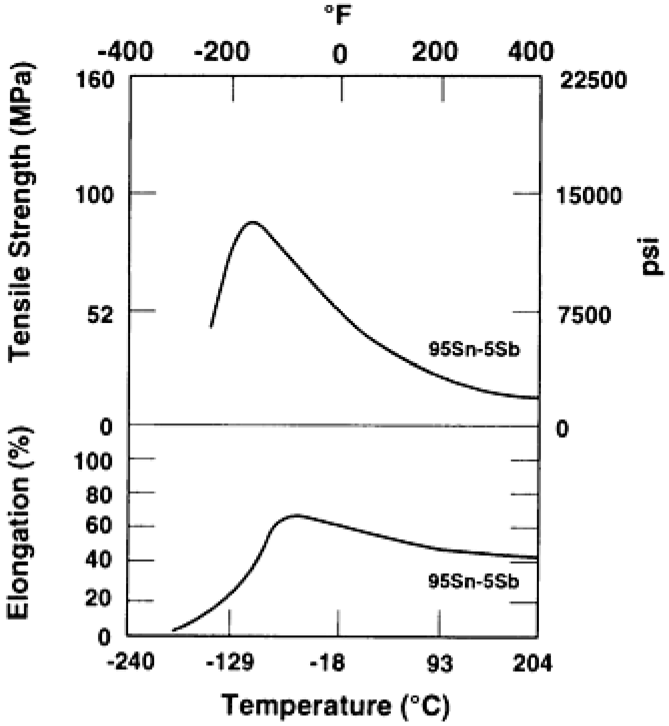

FIG. 11 TEMPERATURE SENSITIVITY OF TENSILE STRENGTH AND ELONGATION OF 95SN-5SB SOLDER

Bulk creep properties are illustrated in stress-rupture data in Table 20. Copper ring-in-plug creep-rupture data are shown in Table 21. Fatigue resistance is demonstrated by the results in Table 22 for cyclic compression/tension (zero mean stress). Both creep and fatigue properties of 95Sn-5Sb are superior to those of 60Sn-40Pb solder. The improved fatigue resistance is noted by the stress required to cause failure at 1000 cycles at a testing rate of 0.2 mm/min (0.008 in./min). The failure stresses at 20 °C (68 °F) are 21 MPa (3.05 ksi) for 95Sn-5Sb and 17 MPa (2.47 ksi) for 60Sn-40Pb.

TABLE 20 CREEP-RUPTURE DATA FOR BULK 95SN-5SB SOLDER

STRESS		LIFE, H
MPA	**KSI**	
AT 20 °C (68 °F)		

17.0	2.47	**497**
19.6	2.84	**10574**
20.6	2.99	**493**
20.7	3.00	>**4000**
21.6	3.13	**183**
22.5	3.27	**0.1**
AT 100 °C (212 °F)		
0.6	0.03	>**2770**
1.1	0.16	>**2090**
2.3	0.33	>**2030**
3.4	0.49	>**2020**
4.5	0.65	>**1870**
6.8	0.99	**4750**
9.0	**1.31**	**1050 ± 380(A)**

Source: Ref 13

(A) PLUS OR MINUS ONE STANDARD DEVIATION, 2 SAMPLES.

TABLE 21 CREEP STRENGTH OF 95SN-5SB COPPER RING-IN-PLUG JOINTS

STRESS		LIFE, H
MPA	**KSI**	
AT 20 °C (68 °F)		
8.8	1.28	**2130**
9.8	1.42	>**2300**
11.8	1.71	**580**
12.7	1.84	**22 ± 19**
13.7	1.99	**15**
14.7	2.13	**18 ± 12**
AT 100 °C (212 °F)		
2.9	0.42	>**3000**
3.9	0.57	**350**
4.9	0.71	**210**
6.4	0.93	**55**
7.3	1.06	**32**
8.0	**1.16**	**15**

Source: Ref 13

TABLE 22 FATIGUE STRENGTH VERSUS TESTING RATE OF 95SN-5SB COPPER RING-IN-PLUG JOINTS

STRESS		LIFE, CYCLES TO FAILURE(A)
MPA	**KSI**	
AT 20 °C (68 °F)		
20.6	2.99	**1200**
22.5	3.27	**43**
24.5	3.56	**1**
AT 100 °C (212 °F)		
13.7	1.99	**1800**
14.2	2.06	**15 ± 16(B)**

15.7	**2.28**	**1**

Source: Ref 13

(A) CYCLIC COMPRESSION/TENSION; MEAN STRESS = 0; SPEED, 0.2 MM/MIN (0.008 IN./MIN).

(B) PLUS OR MINUS ONE STANDARD DEVIATION, 2 SAMPLES.

Tin-silver solders include the tin-silver and tin-copper-silver compositions. The alloy compositions are listed in Tables 23, 24, and 25. ASTM B 32 recognizes the Sn96 solder as that composition which has a silver content of 3.4 to 3.8 wt%. The "same" alloy specified by QQ-S-571E has a composition of 3.6 to 4.4 wt% Ag. The ASTM document includes specifications for two additional tin-silver solders: Sn95, with 4.4 to 4.8 wt% Ag, and Sn94, with 5.4 to 5.8 wt% Ag. The tin-copper-silver alloy "E" included in ASTM B 32 contains 0.25 to 0.75 wt% Ag and 3.0 to 5.0 wt% Cu. Silver improves strength and spreading and lowers the melting temperature, when compared with pure tin. The addition of copper in the "E" alloy further improves its strength. Both the tin-silver (primarily, the eutectic 96.5Sn-3.5Ag) and the tin-copper-silver solders are used extensively in plumbing applications for potable water and in food handling equipment, where the use of lead-containing joints is restricted. Tin-silver alloys also form the high-temperature step of step-soldering processes with tin-lead alloys.

TABLE 23 TIN-SILVER SOLDER COMPOSITION PER ASTM B 32

ALLOY GRADE	**COMPOSITION, %$^{(A)}$**											MELTING RANGE$^{(B)}$			
												SOLIDUS		LIQUIDUS	
	Sn	**Pb**	**Sb**	**Ag**	**Cu**	**Cd**	**Al**	**Bi**	**As**	**Fe**	**Zn**	**°C**	**°F**	**°C**	°F
SN96	BAL	0.10	0.12 MAX	3.4-3.8	0.08	0.005	0.005	0.15	0.01 MAX	0.02	0.005	221	430	221	**430**
SN95	BAL	0.10	0.12	4.4-4.8	0.08	0.005	0.005	0.15	0.01	0.02	0.005	221	430	245	**473**
SN94	BAL	0.10	0.12	5.4-5.8	0.08	0.005	0.005	0.15	0.01	0.02	0.005	221	430	280	**536**
E$^{(C)}$	**BAL**	**0.10**	**0.05**	**0.25-0.75**	**3.0-5.0**	**0.005**	**0.005**	**0.02**	**0.05**	**0.02**	**0.005**	**225**	**440**	**349**	**660**

Note: Data in table represent solder alloys containing >0.2% Pb.

(A) APPLICABLE ONLY TO COMPOSITION 60-40 IN THE FORM OF FLUX-CORE WIRE OR SOLDER PASTE.

(B) APPLICABLE ONLY TO COMPOSITION 60-40 IN THE FORM OF FLUX-CORE WIRE.

(C) APPLICABLE ONLY TO FLUX-CORE WIRE AND SOLDER PASTE.

TABLE 24 TIN-SILVER SOLDER COMPOSITION PER FEDERAL SPECIFICATION QQ-S-571E

ALLOY	**COMPOSITION, %**							APPROXIMATE MELTING RANGE$^{(A)}$			
								SOLIDUS		LIQUIDUS	
	Sn	**Pb max**	**Ag**	**Cu max**	**Zn max**	**As max**	**Cd max**	**°C**	**°F**	**°C**	°F
SN96	**BAL**	**0.10**	**3.6-4.4**	**0.20**	**0.005**	**0.05**	**0.005**	**221**	**430**	**221**	**430**

(A) FOR INFORMATION ONLY

TABLE 25 TIN-SILVER SOLDER COMPOSITION PER ISO/DIS 9453

ALLOY NO.	**ALLOY DESIGNATION**	**MELTING OR SOLIDUS/ LIQUIDUS TEMPERATURE**		**CHEMICAL COMPOSITION, %**												SUM OF ALL IMPURITIES
		°C	**°F**	**Sn**	**Pb**	**Sb**	**Bi**	**Cd**	**Cn**	**In**	**Ag**	**Al**	**As**	**Fe**	**Zn**	
28	S-SN96AG4	221	430	BAL	0.10	0.10	0.10	0.002	0.05	0.05	3.5-	0.001	0.03	0.02	0.001	**0.2**

											4.0					
29	**S-SN97AG3**	**221-230**	**430-445**	**BAL**	**0.10**	**0.10**	**0.10**	**0.005**	**0.10**	**0.05**	**3.0-3.5**	**0.001**	**0.03**	**0.02**	**0.001**	**0.2**

Selected physical properties of the 96.5Sn-3.5Ag alloy are:

DENSITY, G/CM^3	**7.36**
HARDNESS, HB	**14.8**
ELECTRICAL CONDUCTIVITY, %IACS	**14**

IACS, International Annealed Copper Standard

The bulk tensile strength and copper ring-in-plug joint shear strength of this alloy, as a function of testing rate and temperature, are shown in Table 26.

TABLE 26 BULK TENSILE STRENGTH AND COPPER RING-IN-PLUG SHEAR STRENGTH OF 96.5SN-3.5AG SOLDER

TENSILE STRENGTH[(A)]				**TEST SPEED**		SHEAR STRENGTH[(B)]			
AT 20 °C (68 °F)		**AT 100 °C (212 °F)**				**AT 20 °C (68 °F)**		AT 100 °C (212 °F)	
MPA	**KSI**	**MPA**	**KSI**	**MM/MIN**	**IN./MIN**	**MPA**	**KSI**	**MPA**	KSI
51.2	7.43	36.7	5.33	50.00	2.0	55.9	8.11	31.8	**4.62**
56.8	8.24	31.4	4.56	20.00	0.8	50.0	7.26	29.9	**4.34**
43.7	6.34	30.4	4.41	5.00	0.2	44.6	6.47	. . .	**. . .**
44.1	6.40	28.0	4.06	1.00	0.04	37.7	5.47	22.5	**3.27**
41.9	6.08	26.6	3.86	0.20	0.008	28.9	4.19	18.1	**2.63**
36.3	**5.27**	**24.3**	**3.51**	**0.05**	**0.002**	**28.4**	**4.12**	**17.6**	**2.55**

Source: Ref 13

(A) OF BULK MATERIAL.

(B) OF COPPER RING-IN-PLUG.

Creep-rupture stress data on the bulk solder and copper ring-in-plug shear tests are shown in Fig. 12 and 13. Room-temperature fatigue life data (copper ring-in-plug), as a function of test rate, are shown in Table 27 for zero mean stress tension-compression loading.

TABLE 27 COPPER RING-IN-PLUG SHEAR JOINT FATIGUE TESTS FOR 96.5SN-3.5AG SOLDER

STRESS		LIFE, CYCLES TO FAILURE
MPA	**KSI**	
TEST RATE, 0.5 MM/MIN (0.02 IN./MIN)		
21.6	3.13	**3024**
25.4	3.69	**273**
29.4	4.27	**17**
TEST RATE, 2.0 MM/MIN (0.08 IN./MIN)		
19.6	2.84	**2600**

23.5	3.41	**262**
27.4	**3.98**	**119**

Zero mean stress data.

Source: Ref 13

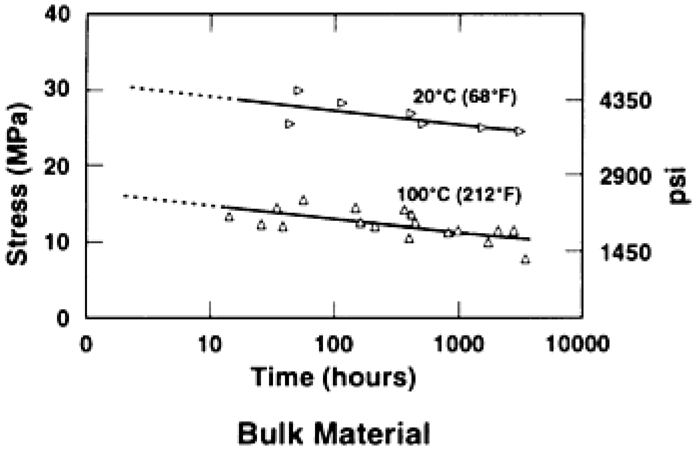

FIG. 12 BULK CREEP-RUPTURE TESTS OF 96.5S-3.5AG. SOURCE: REF 13

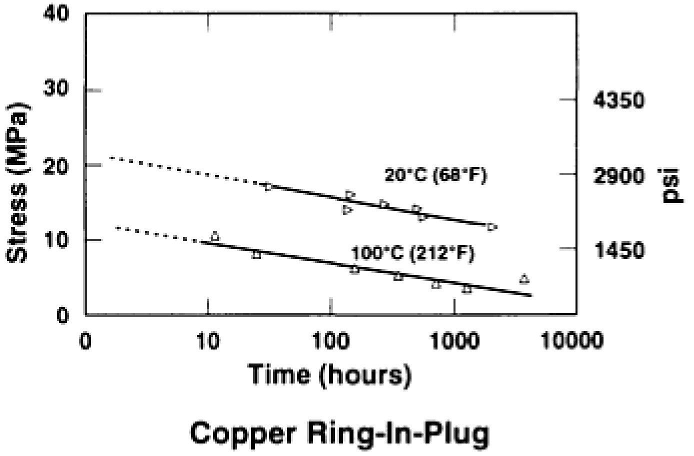

FIG. 13 COPPER RING-IN-PLUG SHEAR CREEP TESTS OF 96.5SN-3.5AG SOLDER. SOURCE: REF 13

Tin-zinc and zinc-aluminum solders are used primarily in joints composed of aluminum-base metals in order to limit galvanic corrosion (Table 28). The eutectic tin-zinc composition, 91Sn-9Zn, has a melting temperature of 199 °C (390 °F), which makes it suitable for low-temperature applications that traditionally use the tin-lead solders. However, oxidation of the zinc constituent of the solder causes excessive dross formation and sluggish spreading when processed in air. As a precaution, the potential for galvanic corrosion between the zinc constituent of the solder and the base metal(s) should be thoroughly investigated prior to its use.

TABLE 28 PHYSICAL PROPERTIES OF TIN-ZINC AND ZINC-ALUMINUM SOLDERS

ALLOY	**DENSITY, G/CM^3**	EUTECTIC TEMPERATURE	
		°C	°F
91SN-9ZN	7.27	199	**390**

95ZN-5AL	**6.60**	**382**	**720**

Zinc-aluminum solders were also developed for aluminum soldering. The eutectic 95Zn-5Al has a melting temperature of 382 °C (720 °F) and is used in tube joining for heat-exchanger applications. Solder joint mechanical properties have been measured in ultrasonic soldering studies (Ref 17). These alloys are not included in the ASTM, ISO, or QQ-S-571E specifications.

Indium-containing solders are used for joining temperature-sensitive substrates or as the low-temperature step in multiple ("step") soldering processes to prevent reflow of the pre-existing solder joints (typically, tin-lead eutectic). Good plasticity at cryogenic temperatures make these alloys suitable for applications such as space vehicles and satellites. Selected properties of these solders are listed in Table 29.

TABLE 29 PROPERTIES OF INDIUM-CONTAINING SOLDERS

ALLOY	**TEMPERATURE**				**HARDNESS, HB**	**TENSILE STRENGTH**		**ELONGATION, %**	**ELECTRICAL CONDUCTIVITY %IACS**	COEFFICIENT OF THERMAL EXPANSION, 10^{-6}/K
	SOLIDUS		**LIQUIDUS**							
	°C	**°F**	**°C**	**°F**		**MPA**	**KSI**			
44IN-42SN-14CD	58	136	58	136	. . .	. . .	. . .	. . .	. . .	**24**
50IN-50SN	117	243	125	257	4.94	11.8	1.72	83	11.7	**20**
52IN-48SN	118	244	118	244	. . .	11.8	1.72	. . .	11.7	**20**
90IN-10AG	141	286	238	460	2.68	11.4	1.65	61	22.1	**. . .**
97IN-3AG	143	289	143	289	. . .	5.5	0.80	. . .	23.0	**22**
100IN	157	315	157	315	(A)	3.5	0.52	41	24.0	**29**
50IN-50PB	**180**	**356**	**209**	**408**	**9.60**	**32.1**	**0.47**	**55**	**6.0**	**27**

Source: Indium Corporation of America

(A) TOO SOFT TO MEASURE.

Indium is frequently added to tin-lead solders as a ternary addition in order to depress the melting temperature (for example, 40Sn-40Pb-20In and 70Sn-18Pb-12In). Indium also improves ductility and oxidation resistance. Indium-containing alloys have poor corrosion resistance in the presence of halide ions, such as those used for some activators in flux chemistries. The alloys 52In-48Sn and 97In-3Ag easily wet glass, fused silica, and other ceramics for glass-to-metal or glass-to-glass joints. These alloys also have very low vapor pressure, which makes them ideal for vacuum applications. The 501n-50Pb alloy is used in electronic applications to limit the scavenging of precious-metal substrates or coatings, which occurs very rapidly with tin-base solders. The indium-containing solder compositions are not certified by ASTM B 32 or QQ-S-571E. However, the latter document does give nominal compositions and melting temperatures. ISO/DIS 9453 does certify the composition of the 50In-50Sn alloy. The high cost of indium has limited the general use of these solders.

Low-melting fusible alloy solders, which contain bismuth, are used for soldering processes that involve temperature-sensitive substrates or devices, as well as the low-temperature part of step soldering procedures. Their ability to melt at low temperatures makes these materials ideal for thermal fuses. Compositions, mechanical, and selected physical properties are listed in Table 30.

TABLE 30 PHYSICAL AND MECHANICAL PROPERTIES OF FUSIBLE ALLOYS

ALLOY(A)	**TEMPERATURE**				**TENSILE STRENGTH**		COEFFICIENT OF THERMAL EXPANSION, 10^{-6}/K
	SOLIDUS		**LIQUIDUS**				
	°C	**°F**	**°C**	**°F**	**MPA**	**KSI**	
44.7BI-22.6PB-8.3SN-5.3CD-19.1IN	47	117	47	117	. . .	. . .	**. . .**
49BI-21IN-18PB-12SN	58	136	58	136	43.4	6.30	**12.8**
50BI-25PB-12.5SN-12.5CD	68	154	73	163	27.8	4.04	**. . .**
55.5BI-44.5PB	124	255	124	255	. . .	. . .	**. . .**
58BI-42SN	**138**	**281**	**138**	**281**	**54.9**	**7.98**	**13.8**

Source: Ref 13

(A) EXTENDED LIST AVAILABLE IN REF 6 (P 110).

Excessive oxidation of bismuth promotes dross formation and can limit spreading for soldering processes conducted in air. The eutectic 58Bi-42Sn alloy has found some applications in electronic assemblies. Bulk properties and copper ring-in-plug shear strength, creep, and fatigue data are listed in Table 31.

TABLE 31 MECHANICAL TEST DATA FOR 58BI-42SN SOLDER

SHEAR STRENGTH				TEST SPEED	
AT 20 °C (68 °F)		**AT 100 °C (212 °F)**			
MPA	**KSI**	**MPA**	**KSI**	**MM/MIN**	IN./MIN
57.4	8.33	18.5	2.69	20.0	**0.8**
63.6	9.23	17.1	2.48	10.0	**0.4**
50.0	7.26	19.5	2.83	1.0	**0.04**
36.0	**5.22**	**11.3**	**1.64**	**0.1**	**0.004**

RUPTURE STRESS		LIFE, H
MPA	**KSI**	
AT 20 °C (68 °F)		
3.0	0.44	**1110**
5.0	0.73	**450**
7.0	1.02	**150**
AT 100 °C (212 °F)		
1.0	0.15	**360 ± 20(A)**
1.5	0.22	**90 ± 40(B)**
3.0	**0.44**	**30**

MAXIMUM STRESS		FATIGUE LIFE, CYCLES
MPA	**KSI**	
AT 20 °C (68 °F)		
9.8	1.42	**5200**
15.0	2.18	**32**
25.0	3.63	**1580**
AT 100 °C (212 °F)		
6.0	0.87	**15,400**
10.0	1.45	**420**
12.0	**1.74**	**7**

Source: Ref 13

(A) SAMPLE SIZE, *N* = 3.

(B) SAMPLE SIZE, *N* = 5.

The ISO/DIS 9453 specification sets guidelines on the composition of 60Sn-38Pb-2Bi, 49Pb-48Sn-3Bi, and 57Bi-43Sn alloys. Alloys with greater than 47 wt% Bi exhibit varying degrees of expansion upon cooling after solidification, because of changes to the lattice structure of the alloys (Ref 18). Dimensional changes to the alloy also take place during room-temperature aging. Whether the dimensional change is due to contraction, expansion, or both, and the magnitude of this change, depends on the composition. These alloys can be used to eliminate solder-joint cracking during the cooling of two substrates that have vastly different degrees of thermal contraction.

Precious-metal solders include the gold-tin, gold-germanium, and gold-silicon compositions. Selected physical and mechanical properties are listed in Table 32. The high concentration of gold causes the alloys to be very expensive.

TABLE 32 PHYSICAL AND MECHANICAL PROPERTIES OF SELECTED GOLD-BASE SOLDERS

ALLOY	**EUTECTIC MELTING POINT**		**YOUNG'S MODULUS**				**ULTIMATE TENSILE STRENGTH**				COEFFICIENT OF THERMAL EXPANSION, 10^{-6}/K
			AT 23 °C (73 °F)		**AT 150 °C (302 °F)**		**AT 23 °C (73 °F)**		**AT 150 °C (302 °F)**		
	°C	**°F**	**GPA**	**10^6 PSI**	**GPA**	**10^6 PSI**	**MPA**	**KSI**	**MPA**	**KSI**	
80AU-20SN	280	536	59.2	8.6	35.8	5.2	275	39.9	205	29.8	**15.9**
80AU-12GE	356	673	69.3	10.1	62.7	9.1	185	26.9	175	25.4	**13.4**
97AU-3SI	**363**	**685**	**83.0**	**12.0**	**83.0**	**12.0**	**255**	**37.0**	**225**	**32.7**	**12.3**

Source: Ref 19

These solders are used for attaching integrated circuit chips to packages. They are also used in the construction and hermetic sealing of packages that house delicate sensors and instrumentation. Their high melting temperature also allows them to be used in the step-soldering process with conventional tin-lead and lead-indium solders.

An attribute of their precious-metal content is that the gold-tin, gold-germanium, and gold-silicon alloys exhibit excellent solderability without the need for a flux (including short-duration processes in air) when soldering gold-coated parts. However, spreading is limited by the high surface tension of the alloys, particularly in the absence of the flux. Mechanical pressure is frequently used to assist spreading of the molten solder in the joint. Gold, tin, germanium, and silicon can be thermally evaporated, making them suitable for thin-film soldering processes, such as the assembly of microsensors (Ref 20). Ultimate tensile strength and yield strength data are given in Fig. 14, which shows that the values far exceed those of the tin-lead alloys. Conversely, these solders show very low ductility (4 to 5%) at room temperature. The composition of the precious-metal solders are not described by the ASTM, QQ-S-571E, or ISO specifications.

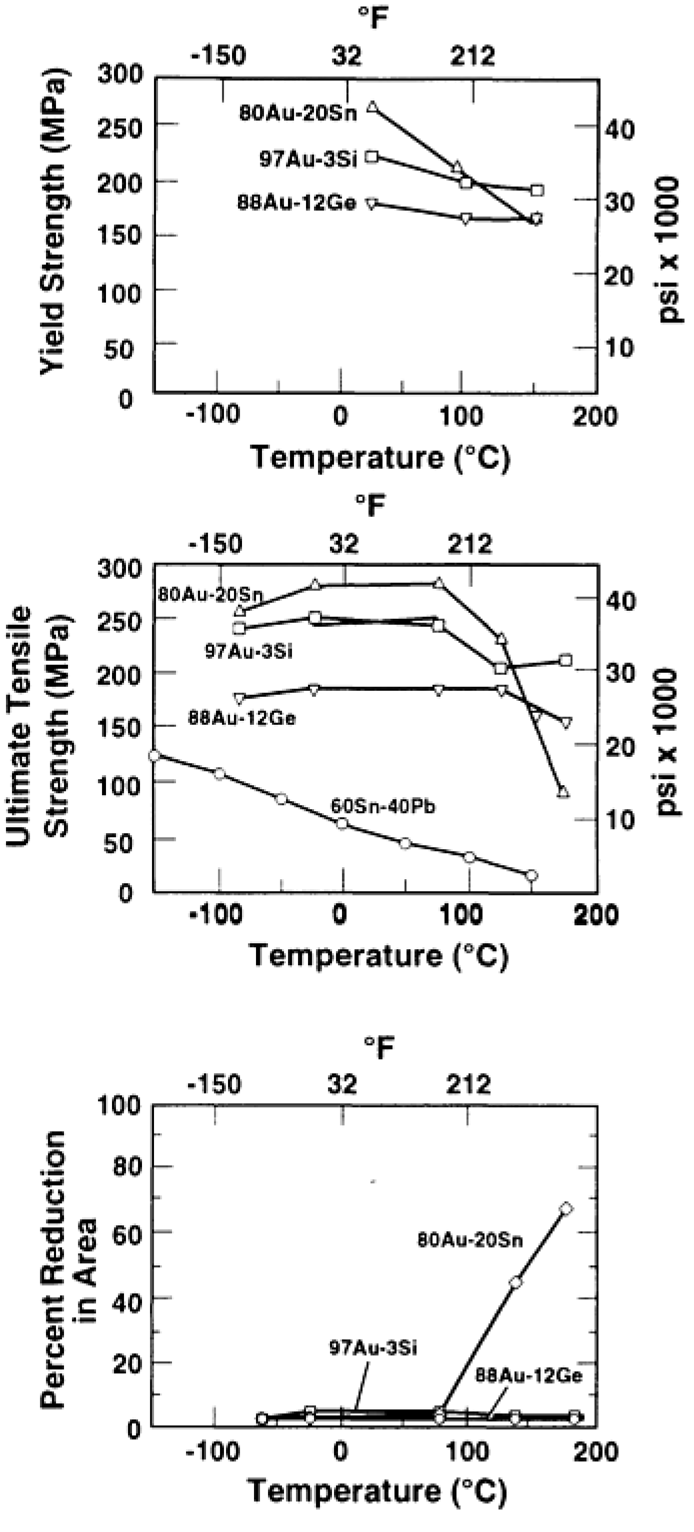

FIG. 14 TENSILE STRENGTH AND YIELD STRENGTH OF THE HIGH-MELTING-POINT PRECIOUS-METAL SOLDERS. SOURCE: REF 19

Cadmium-containing solders include special low-temperature alloys, such as 44In-42Sn-14Cd, which has a eutectic temperature of 93 °C (200 °F), as well as higher-temperature solders, such as the alloy 82.6Cd-17.4Zn, which has a eutectic temperature of 266 °C (511 °F). The corrosion protection offered by the cadmium permits these alloys to be used on aluminum or in corrosive service environments. The relatively high vapor pressure of cadmium precludes the use of cadmium-bearing solders in vacuum equipment. Generally, the use of cadmium-bearing solders is declining because of the health hazards associated with cadmium, particularly in terms of the vapors released during the soldering process. In fact, cadmium-bearing solders are now prohibited in numerous applications.

Available Solder-Metal Forms. In addition to alloy chemistry, the form of solder stock is also described by the ASTM, ISO, and QQ-S-571E specifications. Based on ASTM B 32, these forms include pig, 9.1 to 45 kg (20 to 100 lb); ingot, 1.4 to 4.5 kg (3 to 10 lb); bar, 0.22 to 0.91 kg (0.5 to 2.0 lb); and foil, sheet, or ribbon, which have various widths and thicknesses that are typically greater than 0.025 mm (0.001 in.). Solders that are contained in creams or pastes are specified according to solder powder size, metal content, flux, and viscosity. Solder wire without flux is specified by the wire diameter, which ranges from 0.25 to 6.35 mm (0.010 to 0.250 in.). Flux-cored wire must include flux type and quantity (percentage). Solder sheet may be cut or punched into particular shapes, and the stock is then referred to as a preform. The standard specifications also designate testing procedures to certify the composition and properties (flux activity of cored wire, viscosity of pastes, dimensions of wire and sheet, and so on) for the particular solder or solder-containing material.

References cited in this section

1. "STANDARD SPECIFICATION FOR SOLDER METAL," B 32, *ANNUAL BOOK OF ASTM STANDARDS,* ASTM
2. "SOLDER, TIN ALLOY: TIN-LEAD ALLOY; AND LEAD ALLOY," FEDERAL SPECIFICATION QQ-S-571E, SUPERINTENDENT OF DOCUMENTS, U.S. GOVERNMENT PRINTING OFFICE, WASHINGTON, DC
3. "SOFT SOLDER ALLOYS--CHEMICAL COMPOSITIONS AND FORMS," ISO/DIS 9453, INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, THE HAGUE, NETHERLANDS
4. AMERICAN METALS MARKET, FAIRCHILD PUBLICATIONS, 17 JAN 1992
5. *SOLDERING MANUAL,* AWS, 1977, P 5
6. H. MANKO, *SOLDERS AND SOLDERING,* MCGRAW-HILL, 1979, P 14
7. H. MANKO, *SOLDERS AND SOLDERING,* MCGRAW-HILL, 1979, P 82
8. B. LAMPE, ROOM TEMPERATURE AGING PROPERTIES OF SOME SOLDER ALLOYS, *WELD. J., RES. SUPPL.,* OCT 1976, P 330S
9. SOLDERING MANUAL, AWS, 1977, P 5
10. R. KLEIN-WASSINK, *SOLDERING IN ELECTRONICS,* ELECTROCHEMICAL PUB. LTD., AYR, SCOTLAND, 1989, P 189
11. H. MANKO, *SOLDERS AND SOLDERING,* MCGRAW-HILL, 1979, P 89
12. S. NIGHTINGALE *ET AL.,* TIN SOLDERS, BRITISH NONFERROUS METALS RESEARCH ASSOCIATION, VOL 1, 1942
13. "SOLDER ALLOY DATA, MECHANICAL PROPERTIES OF SOLDERS AND SOLDERED JOINTS," PUBLICATION 656, INTERNATIONAL TIN RESEARCH INSTITUTE, UNITED KINGDOM, 1986
14. *DEVELOPMENT OF HIGHLY RELIABLE SOLDER JOINTS FOR PRINTED CIRCUIT BOARDS,* WESTINGHOUSE DEFENSE AND SPACE CENTER, 1968, P 4-55 TO 4-57
15. R. WILD, FATIGUE PROPERTIES OF SOLDER JOINTS, *WELD. RES. J.,* VOL 15, 1972, P 521S
16. H. SOLOMON, "FATIGUE OF 60/40 SOLDER," REPORT 86CRD024, GENERAL ELECTRIC CO., MAY 1986
17. J. JONES AND J. THOMAS, "ULTRASONIC SOLDERING OF ALUMINUM," RESEARCH REPORT 55-24, FRANKFORD ARSENAL, DEPARTMENT OF THE ARMY, PHILADELPHIA, FEB 1955
18. H. MANKO, *SOLDERS AND SOLDERING,* MCGRAW-HILL, 1979, P 112-115
19. D. OLSEN AND H. BERG, PROPERTIES OF DIE BOND ALLOYS RELATING TO THERMAL FATIGUE, *IEEE TRANS.,* CHMT-2, 1979, P 257
20. P. VIANCO AND J. REJENT, "SOLDER BOND APPLICATIONS IN A PIEZOELECTRIC SENSOR ASSEMBLY," *45TH ANNUAL FREQ. CONTROL SYMP.* (LOS ANGELES, CA), MAY 1991, P 266

General Soldering

Paul T. Vianco, Sandia National Laboratories

Substrate Materials

A substrate is solderable if a metallurgical bond can be formed between the molten solder alloy and the substrate surface. Solderability deteriorates with the presence of organic contaminants and surface oxide layers that restrict the metallurgical reaction. Organic films and heavy oxide layers must be removed by precleaning procedures prior to soldering. Most fluxes are capable of removing only relatively light oxide (or tarnish) films. The tenacity of an oxide film is determined by its intrinsic chemical nature (including adhesion to the base metal) and its thickness.

Table 33 gives a qualitative ordering of solderability for several categories of materials. Generally, the noble metals (for example, gold and silver) are readily solderable, because of limited oxide formation and strong reactivity with the solder. The materials that are most difficult to solder are ceramics and refractory metals, which require metal coatings to promote solder wetting.

TABLE 33 SOLDERABILITY ASSESSMENT OF SELECTED MATERIALS

GOOD	**FAIR**	**MODERATE**	**DIFFICULT**	PRACTICALLY IMPOSSIBLE
GOLD	BRONZE	KOVAR	ALUMINUM-BRONZE	**CHROMIUM**
TIN-LEAD	BRASS	NICKEL-IRON	ALLOYED STEEL	**MAGNESIUM**
TIN	MONEL	NICKEL	ALUMINUM	**MOLYBDENUM**
SILVER	NICKEL-SILVER	STEEL		**TUNGSTEN**
PALLADIUM		ZINC		**BERYLLIUM**
COPPER				

Source: Ref 21

The solderability of marginal materials (nickel, beryllium-copper, steels, or heavily tarnished copper) can be improved by the use of stronger fluxes. However, the highly corrosive nature of the residues typically require postsolder process cleaning steps.

Preassembly Cleaning Procedures. Organic contaminants need to be removed by organic solvents or by semiaqueous or aqueous cleaning solutions. The list of frequently used organic solvents has changed, because the use of chlorofluorocarbon-base materials has been restricted by environmental laws and codes. Substitute materials include acetone, isopropyl alcohol, terpenes, alkaline detergent solutions (for example, 1 to 3% trisodium phosphate with surfactants), and newer semiaqueous compounds. Contamination removal by solvents can be assisted by vapor degreasing, although toxicity to workers and the potential fire hazard generated by fumes or aerosols must be adequately assessed.

Agitated baths and ultrasonics also can improve the cleaning capabilities of organic, semiaqueous, and aqueous cleaning material. However, possible damage to parts by sonic energy must always be assessed before using ultrasonic cleaning processes. The use of soft or distilled water is recommended for aqueous and semiaqueous solutions in order to limit residue formation. Regardless of the cleaning system, all cleaning solution residues must be thoroughly removed from the workpiece prior to storage or processing in order to prevent the formation of stubborn surface films.

Degreasing requirements can be significantly reduced through preventive measures, such as the use of gloves by workers handling parts, enclosing parts in bags or containers, and providing storage facilities that do not contain detrimental materials or airborne particulates. Solderable surfaces should not contact silicon-base materials, because their residues (particularly silicone oils) are extremely difficult to remove and nearly always cause a significant deterioration in substrate solderability.

The removal of heavy oxide layers from metal surfaces is performed by chemical (or acid) cleaning when the chemical action of the flux is incapable of removing the oxide layer during the soldering process. Prior to acid cleaning, the surface must be completely cleaned of organic contaminants, to ensure the effectiveness of the cleaning process. Water-base solutions typically do not penetrate organic films. Surface layers other than oxides, such as sulfides, hydrides, or chlorides, may require specialized cleaning solutions. For increasingly thicker layers, whether an oxide film or another type of surface contaminant, longer cleaning periods or more aggressive chemicals are required for the removal process. However, either approach increases the chance of damage to the substrate.

Table 34 identifies some general solutions used to clean metal surfaces. Inhibitors are typically added to commercially available acid solutions to prevent pitting. Distilled or deionized water should be used in solutions to prevent unwanted deposits and residual films after cleaning. Electropolishing processes can also be used to remove metal-oxide layers (Ref 23). In such procedures, the substrate forms the anode.

TABLE 34 CLEANING SOLUTIONS FOR SUBSTRATE MATERIALS

MATERIAL	**SOLUTION**	COMMENTS
IRON, STEELS, ALUMINUM	HCL (25%):$3H_2O$ TO HCL (25%):$9H_2O$	**30 TO 38 °C (85 TO 100 °F), 10 TO 45 MIN; TIME DEPENDS ON SCALE THICKNESS**
STEELS	H_2SO_2 (77%):$19H_2O$ TO H_2SO_2 (77%):$9H_2O$	**70-82 °C (160 TO 180 °F), 30 TO 120 S FOR LIGHT TARNISH, 15 MIN FOR HEAVY SCALE; BLACK SMUT IS WATER RINSED FOR COPPER ALLOYS, ADD 1 WT% SODIUM DICHROMATE OR 2 VOL% NITRIC ACID**
NICKEL ALLOYS	0.5HF:$4HNO_3$ (70%):$8H_20$	
STAINLESS STEELS	1. H_3PO_4 (100%):$9H_2O$ TO $2H_3PO_4$ (100%):$3H_20$	**CAUTION: HYDROFLUORIC ACID IS EXTREMELY CORROSIVE; AVOID ALL SKIN CONTACT**
	2. $2H_2SO_4$ (77%):HCL (25%):$8H_2O$	
	3. $2HNO_3$ (70%):3HF:$5H_2O$	
COPPER, COPPER ALLOYS	(SEE "STEELS" ABOVE)	**25 °C (75 °F); 2 TO 5 MIN FOR MODERATE OXIDES**
	$2HNO_3$ (70%):$17H_2O$ TO $2HNO_3$ (70%):$4H_2O$	
BRASS	0.015 HCL (25%):$4HNO_3$ (70%):$8H_2SO_4$ (77%):H_2O	
BERYLLIUM-COPPER (THREE STEPS)	FIRST STEP--H_2SO_4 (77%):$3H_2O$	**FIRST STEP, 75 °C (165 °F)**
	SECOND STEP--H_2O:RINSE	**THIRD STEP, 25 °C (75 °F)**
	THIRD STEP--HNO_3 (70%):$2H_2O$	
ALUMINUM, HIGH-SILICON ALLOYS, CAST IRON	**HF:H_2SO_4 (77%):$18H_2O$**	**25 °C (75 °F); 2 TO 5 MIN FOR MODERATE OXIDES**

Source: Ref 22

Whether a chemical etch or an electropolishing procedure is used, the corrosive solutions must be thoroughly rinsed from the surfaces to prevent staining or latent corrosion. These artifacts may also deteriorate solderability in follow-up processing steps. Rinses with distilled water, followed by an alcohol rinse to remove the water, is a typical sequence. Workpieces should not be dried with unfiltered or undried "house" compressed air, because oil and water droplets from the air line will quickly contaminate the surface. Once cleaned, the surface should be soldered as soon as possible to limit reoxidation, or measures should be taken to protect surfaces from further contamination.

Mechanical abrasion can be used to remove excessively thick oxide films. Techniques include using sandpaper, steel wool, or metal files, or blasting with particulates. However, solderability can be quickly degraded by abrasive media that become embedded in the substrate surface and are not readily wetted by the molten solder. Procedures that use abrasive particles (blasting or sandpaper) or steel wool should be followed by a chemical etching treatment to remove the layer of surface material containing the foreign particles. Metal files are used to prepare the surfaces of larger, more-rugged substrates, such as pipes and fittings, and they should be cleaned of contaminant materials prior to use.

Coatings. Metals and alloys that are difficult to solder (aluminum, chromium-nickel steels, cast irons, and others) may lose their solderability too quickly after precleaning and become unprocessible. In these instances, coatings can be applied to protect the base metal prior to final assembly. Methods of application include electroplating, electroless plating, evaporation, chemical vapor deposition, or dipping in a molten metal bath. Table 35 identifies 100Sn and tin-lead solder coatings that are used to protect base-metal solderability. These protective coatings can be electroplated on the surface or they can be applied by dipping the substrate into a hot solder or tin bath. Substrates with electroplated films can be heated above the melting temperature of the coating to "fuse" it, which provides an improved seal against air and contaminants, similar to the protection obtained with hot-dipped tin or solder coatings. Electroless tin processes, including "immersion" platings, are also used.

TABLE 35 TIN AND SOLDER COATINGS TO PRESERVE BASE-METAL SOLDERABILITY

COATING	**THICKNESS**		COMMENTS
	µM	**µIN.**	
ELECTROPLATED 100SN	7.6-13	300-500	**CONCERN FOR WHISKER GROWTH WITH MOISTURE AND RESIDUAL STRESSES; RECOMMENDED FUSING**
ELECTROPLATED 100SN, FUSED	2.5-13	100-500	**MELTING POINT OF TIN, 232 °C (450 °F)**
ELECTROPLATED SN-PB	7.6-23	300-900	**AVAILABLE COMPOSITIONS: 63SN-37PB, 95PB-5SN, ETC.; PLATING COMPOSITION MAY DEVIATE FROM NOMINAL VALUES**
ELECTROPLATED SN-PB, FUSED	2.5-13	100-500	**MELTING POINTS: 63SN-37PB, 183 °C (361 °F); 95PB-5SN, 314 °C (597 °F)**
HOT-DIPPED 100SN OR SN-PB (63SN-37PB)	>5.1	>200	**UNIFORM COVERAGE REQUIRED**
ELECTROLESS ("IMMERSION") 100SN	**~1.5**	**~60**	**PROTECTION SENSITIVE TO DEPOSITION PROCESS DETAILS; QUALIFICATION TESTS RECOMMENDED**

Source: Ref 24

For metals such as aluminum and magnesium, which simply cannot be kept solderable, or for ceramics that are intrinsically unsolderable, a metal layer is deposited on the substrate. The solder more readily wets this layer, which is often referred as to the "solderable" layer. These films must be adequately thick to avoid being consumed by the metallurgical reaction with the solder, and they must exhibit adequate adhesion to the base material. In some instances, an "adhesive" layer may be deposited between the base material and the solderable layer. A "protective" layer can also be deposited onto the solderable coating in order to prevent excessive oxidation or contamination. This protective layer is then consumed by the solder. Table 36 lists the coating materials and processing steps used for substrates that are commonly involved in solder processing.

TABLE 36 SOLDERABLE AND PROTECTIVE FINISHES FOR COMMON SUBSTRATE MATERIALS

SUBSTRATE	**SOLDERABLE LAYER**	**SOLDERABLE LAYER THICKNESS**		**PROTECTIVE LAYER**	**PROTECTIVE LAYER THICKNESS**		COMMENTS
		µM	**µIN.**		**µM**	**µIN.**	
COPPER	ELECTROPLATED NICKEL(A)	1.5-3.8	60-150	ELECTROPLATED GOLD(B)	1.3-2.5	50-100	**. . .**
COPPER ALLOYS WITH ZINC	ELECTROPLATED COPPER	2.5(C)	100(C)	. . .	. . .	. . .	**12 µM (5000 µIN.) FOR ELEVATED-TEMPERATURE SERVICE**
NICKEL AND NICKEL ALLOYS	ELECTROPLATED NICKEL(A)	1.5-3.8	60-150	ELECTROPLATED GOLD(B)	1.3-2.5	50-100	**. . .**
	ELECTROPLATED SILVER	3.8-8.9	150-350	. . .	. . .	. . .	**SMALL SOLDERABILITY LOSS THAT IS DUE TO SILVER TARNISH**
	ELECTROPLATED COPPER	3.8-7.6	150-300	. . .	. . .	. . .	**. . .**
ALUMINUM AND ALUMINUM ALLOYS(D)	ELECTROPLATED COPPER (ZINCATE UNDERCOAT)	3.8-25	300-1000	. . .	. . .	. . .	**ANODIC FINISH (CD OR ZN) RECOMMENDED TO PREVENT CORROSION BETWEEN ALUMINUM AND COPPER OR NICKEL**
	ELECTROPLATED NICKEL (ZINCATE UNDERCOAT)(A)	1.3-3.8	50-150	ELECTROPLATED GOLD(B)	1.3-2.5	50-100	**. . .**
IRON,IRON-BASE ALLOYS (ALLOY 42, ALLOY 52, KOVAR), LOW-CARBON STEELS, AND STAINLESS STEELS	ELECTROPLATED NICKEL(A)	1.3-3.8	50-150	ELECTROPLATED GOLD(B)	1.3-2.5	50-100	**FOR COATED STEELS, USE RECOMMENDED SOLDERABLE AND PROTECTION FINISHES FOR THE PARTICULAR COATING MATERIAL**
	ELECTROPLATED SILVER	3.8-8.9	150-350	. . .	. . .	. . .	**. . .**
	ELECTROPLATED PALLADIUM-NICKEL	1.3-3.5	50-150	. . .	. . .	. . .	**. . .**
CAST IRON(D)	ELECTROPLATED IRON	1.3-3.8	50-150	. . .	. . .	. . .	**EXTENDED THICKNESS TO 2.5 TO 7.6 µM (100 TO 300 µIN.) FOR IMPROVED COVERAGE**
CERAMICS (A1$_2$O$_3$, BEO, SIO$_2$, SI$_3$N$_4$, ETC.)(D)	PRECIOUS METAL THICK FILMS	10-20	100-800	ELECTROPLATED GOLD(B)	2.0-2.5	80-100	**CERAMIC SURFACE ROUGHNESS AND THICK-FILM POROSITY MAY REQUIRE THICKER GOLD DEPOSITS**
	REFRACTORY THICK FILMS	10-20	100-800	. . .	. . .	. . .	**. . .**
	MAY INCLUDE NICKEL	**1.0-9.0**	**40-350**	**. . .**	**. . .**	**. . .**	**. . .**

	(SOLDERABLE OR BARRIER LAYER BARRIER LAYER)						

(A) THICKER LAYERS OF 2.5 TO 7.6 μM (100 TO 300 μIN.) PERMITTED FOR IMPROVED COVERAGE; ELECTROLESS NICKEL, 1.3 TO 3.0 μM (50 TO 120 μIN.) OR 2.5 TO 5.1 μM (100 TO 200 μIN.) PERMITTED ON NONFLEXIBLE SUBSTRATES; LAYER ADHESION AND SOLDERABILITY REQUIRE CLOSE MONITORING.

(B) THICKER LAYERS OF 2.5 TO 7.5 μM (100 TO 300 μIN.) PERMITTED FOR SOLDERABLE LAYER PROTECTION; HOWEVER, THINNER LAYERS ARE RECOMMENDED AND SOLDER JOINTS SHOULD BE MONITORED FOR EMBRITTLEMENT.

(C) MINIMUM.

(D) NOT CONTAINED IN MIL-STD-1276D

Gold plating layers should be removed by hot-solder dipping twice in a bath of molten solder if the calculated gold content of the solder joint will exceed 3 to 4 wt%, in order to prevent solder-joint embrittlement by gold-tin intermetallic formation. Electroplated gold layers should be pure gold, preferably type III, 99.9% pure, and grade A, according to MIL-G-45204C. Gold coatings alloyed with cobalt, nickel, or both (termed "hard gold"), which are used for wear resistance, are difficult to solder, because of oxidation of the cobalt or nickel component. Only matte finishes should be specified for soldering applications. "Bright" platings require organic additives in the bath, which then become entrapped in the gold coating and subsequently volatilize during soldering. This creates voids in the joints and poor wetting properties.

References cited in this section

21. R. KLEIN-WASSINK, *SOLDERING IN ELECTRONICS*, ELECTROCHEMICAL PUB. LTD., AYR, SCOTLAND, 1989, P 189
22. *SOLDERING MANUAL*, AWS, 1977, P 35-39
23. P. VIANCO *ET AL.*, SOLDERABILITY TESTING OF KOVAR WITH 60SN-40PB SOLDER AND ORGANIC FLUXES, *WELD. J. RES. SUPPL.*, JUNE 1990, P 230S
24. *BRAZING HANDBOOK*, AWS, 1991, P 267-276

General Soldering

Paul T. Vianco, Sandia National Laboratories

Fluxes

The major role of the flux is the removal of thin tarnish layers during the initial stages of the soldering process, thereby permitting the molten solder to react with the substrate and to spread. Fluxes do not effectively displace organic contaminants, so the substrate underneath such films will remain untouched by the flux and will therefore be unsolderable. A degreasing step should precede flux application. Even the strongest fluxes cannot remove thick oxides or heavy scales, so a mechanical or chemical precleaning procedure may be required.

The flux has two additional functions. One is that it lowers the surface tension of the solder, allowing it to more readily fill gaps and holes by capillary action. The other function is that the flux coating protects the metal surface from reoxidation during the heating steps just prior to soldering.

Fluxes contain three principal ingredients: an active chemical compound, such as a halide, for oxide removal; wetting agents to improve surface coverage; and a vehicle to dilute and mix the cleaning compound and wetting agents together. The vehicle, which is removed by evaporation during the soldering process, is typically water, isopropyl alcohol, glycerin, glycol (for liquid fluxes), or petroleum jelly (for flux pastes or creams). Fluxes are characterized by their cleaning agent and are assigned to one of these categories of increasing activity: rosin-base fluxes, organic-acid fluxes (also called "intermediate" or "water-soluble" fluxes), and inorganic-acid fluxes.

Flux specifications are most numerous for the rosin-base materials, because of the criticality of their properties in the assembly of electronic products. Sample specifications include military standards MIL-F-14256E and MIL-STD-2000; federal specification QQ-S-571 E; and the industry standards ASTM D 509, IPC-S-815A, and IPC-TM-650 (test methods). Fluxes are tested for bulk corrosivity (copper mirror test), concentrations of Cl^- and F^- (halide content), and solids content. Tests of residues left by the flux on the substrate (typically a printed wiring board) include surface insulation resistance, and ionic residues. The ISO/DIS 9454-1 document includes organic- and inorganic-acid flux specifications and test procedures (ISO/DIS 9455-X, where X = 1, 2, . . ., 14), in addition to rosin-base flux specifications.

The removal of flux residues depends on the specific application, as well as the potential corrosivity of the residues left behind. If residues are specified for removal, then the workpiece should be cleaned as soon as possible after processing. Rosin residues become particularly difficult to remove with time.

Rosin-base fluxes contain "water-white" rosin, a distillation product from pine tree sap. When used alone, rosin-base fluxes are referred to as a "nonactivated," or type R, grade. Rosin is solid, noncorrosive, and insulating at room temperature. Heating causes it to liquefy and become slightly chemically active, because of its abietic and pimaric acid components. The rosin is typically dissolved in an organic vehicle, such as isopropyl alcohol, to form the liquid solution.

The addition of an activator to rosin fluxes increases their chemical activity. Activators can be organic halogenated compounds, such as amine hydrohalides that contain chloride, fluoride, or bromide ion groups or "halide-free" activators, such as oleic, stearic, or lactic acids. Halide-free fluxes are recommended for materials that are sensitive to stress-corrosion cracking. The concentration of activators, which defines the corrosivity of the flux, determines the flux category as being one of the following: rosin-base, mildly activated (RMA), fully activated (RA), and superactivated (SA). Generally, rosin-base fluxes have poor high-temperature stability; that is, the fluxes degrade rapidly after excessive exposure to elevated temperatures (for example, >25 s at 260 °C, or 500 °F).

R-type fluxes are used on electronic assemblies, where the benign residues are left on the soldered assembly. However, the substrate(s) must have excellent solderability, because the weak activity of these fluxes offers a very small window of variability for oxide thicknesses. R-type fluxes are effective on lightly tarnished copper surfaces and on precious-metal substrates, as well as on tin- and solder-coated base metals. Residues can be removed by nonpolar solvents, organic solvents (isopropyl alcohol, 1,1,1-trichloroethane), semi-aqueous solutions (terpenes), and aqueous baths or detergent solutions.

The residues left by R-type fluxes can be reduced by dilution with an organic solvent (for example, isopropyl alcohol). Such fluxes are termed "low solids" because the solids, which are responsible for the residues, are significantly reduced so that cleaning steps are not required. These fluxes are also referred to as "no-clean;" however, the term "no-clean" covers a much wider range of flux formulations that must simply comply with residue corrosivity, irrespective of the solids content.

The increased activity of the RMA fluxes improves solder wetting on more heavily tarnished copper substrates. Residues are relatively benign and only require removal when the fluxes are used on high-reliability electronic systems (for example, in military applications) or when the fluxes contain potentially harmful halide activators. Residue cleaning is a two-step process. The first step requires the use of organic solvents to remove the nonpolar rosin, and in the second step, a polar solvent (water or polar alcohols) is used to rinse away activator residues. Aqueous and semi-aqueous products can perform both functions.

The greater corrosivity of RA and SA fluxes is due, primarily, to higher levels of the activators that are used in RMA fluxes. These fluxes are used on base metals such as nickel (and nickel plate), lightly tarnished low-carbon steels, copper, iron-base alloys, and copper-base alloys (brasses, bronzes, and beryllium copper). Residues of the RA and SA fluxes must be removed to prevent the corrosion of the substrate later in service. Cleaning procedures similar to those identified for RMA fluxes are used. Cleanliness testing for ionic residues should be implemented to verify the adequate removal of flux deposits, particularly for the stronger fluxes.

Organic-acid fluxes have a chemical activity range that varies from the levels associated with RMA fluxes to those that exceed the activity of RA fluxes. These fluxes contain one or more organic acids, such as lactic, oleic, or stearic acid.

Chemical activity can be enhanced by adding organic halogen compounds (amine hydrohalide, which may contain chloride and bromide derivatives) or nonhalogenated substances, such as one of the amines or amides (urea or thylene diamine). Typical vehicles include water, isopropyl alcohol, polyglycols, or petroleum jelly for pastes.

The organic-acid fluxes are used in many electronic applications involving machine processes and hand assembly, as well as the hot tin or solder dipping of nickel and iron-base alloy leads and devices. These fluxes are also used on structural applications with copper and copper alloy workpieces that have light to moderate tarnishes.

Whether the fluxes are used in structural or electronic applications, their residues should be removed. The water-soluble residues are typically removed by rinsing with water or a polar organic solvent. The organic acid fluxes have better high-temperature stability, when compared with the rosin-base materials, which makes them good candidates for solders with melting temperatures that exceed 200 °C (363 °F).

Inorganic-acid fluxes have the highest levels of chemical activity. There are two categories of these fluxes:

- PURE ACIDS, SUCH AS HYDROCHLORIC, HYDROFLUORIC, OR PHOSPHORIC ACIDS, WHICH HAVE SURFACTANTS ADDED TO ENHANCE COVERAGE.
- INORGANIC SALT MIXTURES OR SOLUTIONS, WHICH MAY ALSO CONTAIN SURFACTANTS.

The pure acids are very strong and are capable of removing heavy oxide layers and scales. However, they offer the clean surface limited protection from reoxidation during the soldering process. Binary inorganic salt "alloys" are formed from the combination of zinc chloride ($ZnCl_2$), ammonium chloride (NH_4Cl), stannous chloride ($SnCl_2$), or sodium chloride (NaCl) components. The inorganic salt combinations form simple eutectics with minimum temperatures near the melting point of solders. (The melting temperature of the single salts are too high for them to flow at typical soldering temperatures.) For example, the NH_4Cl:3 ($ZnCl_2$) eutectic composition has a eutectic temperature (T_e) of 177 °C (351 °F). Other eutectic salt combinations include: $SnCl_2$-CuCl, T_e = 170 °C (338 °F); AgCl-CuCl, T_e = 260 °C (500 °F); and CuCl-$ZnCl_2$, T_e = 230 °C (446 °F).

Proper flux selection is based on the salt mixture having a melting temperature that is less than that of the solder. The molten salt reduces the surface oxides and coats the base metal to prevent reoxidation during the soldering process. Flux activity is further increased by dissolving the flux in a water vehicle. The salts break down to release Cl^- ions, which combine with water to form a hydrochloric acid (HCl) "activator." The fluxes, particularly the water solutions, may contain surfactants to assist the complete coverage of the joint area. A paste form of the flux uses petroleum jelly as the vehicle.

The inorganic-acid fluxes are limited to structural applications, such as plumbing or mechanical assemblies. Their corrosive activity is unacceptable for electronic devices or substrates, or for their assembly. These fluxes are effective on nickel and nickel alloys, stainless steels, chromium, and heavily tarnished copper and copper alloys. The inorganic fluxes are very stable at high temperatures. The flux residues and the fluxes themselves are extremely corrosive and must be thoroughly removed after processing.

It should be noted that chloride-containing residues and small amounts of water are particularly damaging to aluminum and aluminum alloys, because of the potential for stress-corrosion cracking. Residues are removed with hot-water rinses and follow-up polar-solvent rinses. Neutralizing solutions (mild caustics) can be used on nonsensitive substrates. A list of selected inorganic flux solutions for particular base metals is provided in Table 37.

TABLE 37 TYPICAL INORGANIC FLUXES FOR SELECTED BASE METALS

FOR STAINLESS STEEL AND GALVANIZED IRON	
ZINC CHLORIDE, ML (OZ)	**2510 (85)**
AMMONIUM CHLORIDE, ML (OZ)	**190 (6.5)**
STANNOUS CHLORIDE, ML (OZ)	**270 (9)**
HYDROCHLORIC ACID, ML (OZ)	**60 (2)**
WETTING AGENT (OPTIONAL), WT%	**0.1**
WATER$^{(A)}$	
FOR STAINLESS STEEL	
ZINC CHLORIDE, ML (OZ)	**1420 (48)**
AMMONIUM CHLORIDE, ML (OZ)	**150 (5)**
HYDROCHLORIC ACID, ML (OZ)	**90 (3)**
WETTING AGENT (OPTIONAL), WT%	**0.1**
WATER$^{(A)}$	
FOR MONEL	
ZINC CHLORIDE, ML (OZ)	**470 (16)**
AMMONIUM CHLORIDE, ML (OZ)	**470 (16)**
GLYCERIN, ML (OZ)	**470 (16)**
WATER, L (GAL)	**0.5 (0.125)**
FOR HIGH-TENSILE-STRENGTH MANGANESE, BRONZE, COPPER, OR BRASS	
ORTHOPHOSPHORIC ACID (85%), ML (OZ)	**1000 (34)**
WATER, ML (OZ)	**470 (16)**
FOR CAST IRON	
ZINC CHLORIDE, ML (OZ)	**950 (32)**
AMMONIUM CHLORIDE, ML (OZ)	**120 (4)**
SODIUM CHLORIDE, ML (OZ)	**240 (8)**
HYDROCHLORIC ACID, ML (OZ)	**240 (8)**
WATER$^{(A)}$	
FOR CAST IRON	
ZINC CHLORIDE, ML (OZ)	**1180 (40)**
AMMONIUM CHLORIDE, ML (OZ)	**120 (4)**
HYDROFLUORIC ACID, ML (OZ)	**40 (1.25)**
WATER$^{(A)}$	
PASTE FLUX FOR SOLDERING ALUMINUM	
STANNOUS CHLORIDE, ML (OZ)	**2450 (83)**
ZINC DIHYDRAZINIUM CHLORIDE, ML (OZ)	**210 (7)**
HYDRAZINE HYDROBROMIDE, ML (OZ)	**295 (10)**
WATER, ML (OZ)	**295 (10)**
FOR SOLDERING ALUMINUM	
CADMIUM FLUOBORIDE, ML (OZ)	**150 (5)**
ZINC FLUOBORIDE, ML (OZ)	**150 (5)**
FLUOBORIC ACID, ML (OZ)	**180 (6)**
DIETHANOL AMINE, ML (OZ)	**590 (20)**
DIETHANOL DIAMINE, ML (OZ)	**120 (4)**
DIETHANOL TRIAMINE, ML (OZ)	**295 (10)**
POTASSIUM CHLORIDE, ML (OZ)	**1330 (45)**
SODIUM CHLORIDE, ML (OZ)	**890 (30)**
LITHIUM CHLORIDE, ML (OZ)	**440 (15)**
POTASSIUM FLUORIDE, ML (OZ)	**210 (7)**

SODIUM PYROPHOSPHATE, ML (OZ)	**90 (3)**
TRIETHANOLAMINE, ML (OZ)	**740 (25)**
FLUOBORIC ACID, ML (OZ)	**90 (3)**
CADMIUM FLUOBORATE, ML (OZ)	**60 (2)**
STANNOUS CHLORIDE, ML (OZ)	**1300 (44)**
AMMONIUM CHLORIDE, ML (OZ)	**150 (5)**
SODIUM FLUORIDE, ML (OZ)	**30 (1)**
ZINC CHLORIDE, ML (OZ)	**1300 (44)**
AMMONIUM CHLORIDE, ML (OZ)	**150 (5)**
SODIUM FLUORIDE, ML (OZ)	**30 (1)**

(A) TO MAKE 3.8 L (1 GAL)

General Soldering

Paul T. Vianco, Sandia National Laboratories

Solder Joint Assembly

The configuration of a solder joint must have two basic attributes: It must meet the service requirements of the assembly, and it must be manufacturable to those requirements. Specifications and guidelines for the design of structural solder joints (for example, conduit) are limited, when compared with those used in electronic applications. Some British and European standards have been established, but few exist in the United States. Because standard joint design guidelines are not available, performance data are not well documented. However, the brazing community has several sets of standards that pertain to structural brazed joints (Ref 24). These specifications, which were established by the American Society of Mechanical Engineers (ASME), American National Standards Institute (ANSI), American Welding Society (AWS), and the Society of Automotive Engineers (SAE), cover joint construction, strength tests, and safety issues. Although these specific documents were developed for brazements, they can provide general guidelines for applying similar tests and criteria on solder joints. Under the circumstances, conservative design practices and inspection and test protocols that certify process quality have substituted for insufficient data. Additional information is available in the article "Evaluation and Quality Control of Soldered Joints" in this Volume.

Design for Service Requirements. Solder joint service conditions are typically represented by either electronic interconnect applications or structural applications. The former are described in the article "Soldering in Electronic Applications." Structural applications, which are described below, range from the assembly of small housings and containers to plumbing for fluid or gas conduit. Typical service requirements for structural joints include the support of mechanical loads (both high-temperature and cryogenic conditions), hermetic sealing for vacuum applications, and leak-tight joints for positive-pressure conduits and containers.

Several common structural solder joint configurations are shown in Fig. 15. It is recommended that solder joints be loaded in shear. Tensile loads should be avoided, because a slight deviation in the loading direction can place the joint in a peel-type fracture mode, which typically causes premature failure of the assembly. Lap-shear joints should have an overlap of approximately one-third of the base-metal thickness (Ref 25). Lap joints retain their strength over a range of gap thicknesses. However, the gap thicknesses of a single joint should be uniform for consistent capillary filling by the molten solder.

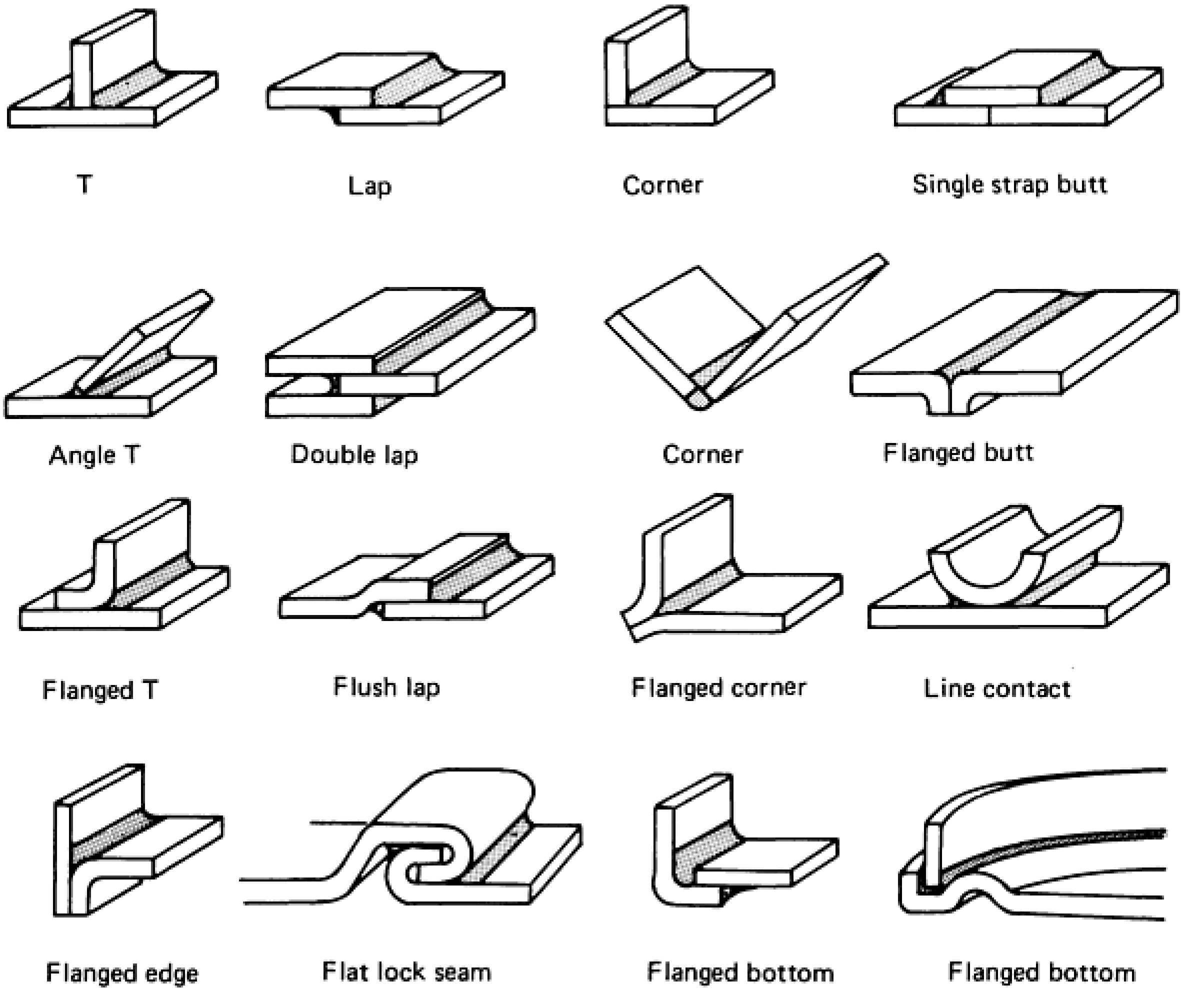

FIG. 15 COMMON STRUCTURAL SOLDER JOINTS

One of the first of the design criteria that pertains to the selection of a suitable solder is melting temperature, specifically the solidus and liquidus points. The solidus temperature is determined directly by the service conditions of the final assembly. Of course, the service temperature should never exceed the solidus point. However, how close the service temperature should be allowed to approach the solidus temperature depends on the level of monotonic or creep deformation and the fatigue life acceptable for the particular application. Tables of appropriate data either should be referred to or the necessary experiments performed.

A suitable liquidus temperature is primarily determined by the temperature sensitivity of the substrate materials during the soldering process. The molten solder working temperature should be approximately 25 to 50 °C (45 to 90 °F) greater than the liquidus temperature. A secondary factor in the selection of the liquidus and solidus temperatures is the desired pasty range for solder workability prior to solidification. Finally, multiple soldering processes require the solidus temperature of the first (high-temperature) soldering step to exceed the liquidus point by 30 to 50 °C (55 to 90 °F) of the second (low-temperature) step.

The mechanical properties of the solder must be considered in terms of the particular application. For example, the monotonic and creep strengths (shear or tensile) must be appropriate for the particular loading conditions and temperature environments to which the joint is subjected. Fatigue resistance is important for joints subjected to time-varying loads. Although such loads may be well below the ultimate monotonic or creep strength limits of the solder, they can result in significant cracking of the solder layer in a joint.

The size of the substrates and the thermal conductivity of the material from which they are constructed affect the joining process. A disparity of mass between the two substrates will result in the larger part heating up more slowly than the smaller one. Insufficient heat can result in a cold joint (that is, an absence of capillary flow and/or poor fillet formation caused by premature solder solidification). Conversely, overheating the smaller workpiece may cause undesirable changes to the mechanical and physical properties of the base material. Similarly, two substrates with vastly different thermal conductivities (for example, stainless steel and aluminum) are difficult to uniformly heat, particularly in localized heating processes, such as soldering with an iron or torch.

When designing joints of dissimilar materials, the coefficient of thermal expansion (CTE) differences must be considered because sufficient stresses can build within the joint to cause catastrophic failure, particularly if the substrates are brittle. Time variations in environmental temperatures cause fatigue deformation to accumulate in the solder, thereby weakening the mechanical strength of the joint. An example of the method used to estimate the fatigue life of a 60Sn-40Pb solder joint can be made by calculations based on empirical data, as shown below. Unfortunately, data to support similar computations with other solders have not been compiled.

Laboratory tests of the isothermal fatigue life of 60Sn-40Pb over the temperature range from -50 to 125 °C (-58 to 257 °F) were used to develop the following equation for the fatigue life, N_f (Ref 16):

$$N_f = \left(\frac{1.14}{\Delta g_p} \right)^{1.96}$$

where N_f is the number of cycles to failure (50% load drop for equivalent strain) and $\Delta\gamma_p$ is the plastic strain level in the joint. A conservative estimate of $\Delta\gamma_p$ is made by using the total strain, $\Delta\gamma_t$, determined by:

$$\Delta g_t = \frac{(a_1 - a_2)\Delta TL}{t}$$

where α_i is the CTE of the two substrates ($i = 1, 2$), $\Delta T$is the temperature difference, L is the maximum dimension of the joint, and t is the joint thickness. The parameters and joint configuration are illustrated in Fig. 16. Solder fatigue cracks may be visible in the solder fillets, typically where L is a maximum.

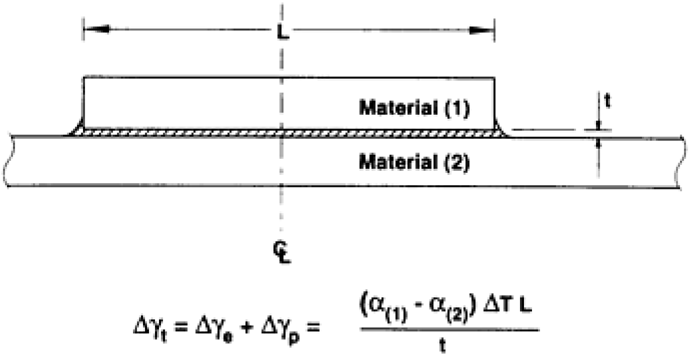
L
Material (1)
Material (2)
t
℄
$\Delta\gamma_t = \Delta\gamma_e + \Delta\gamma_p = \frac{(\alpha_{(1)} - \alpha_{(2)})\,\Delta T\,L}{t}$

FIG. 16 PARAMETERS FOR THE CALCULATION OF TOTAL THERMAL MISMATCH STRAIN IN A GENERIC SOLDER JOINT CONFIGURATION. *T,* JOINT THICKNESS; *L,* MAXIMUM JOINT DIMENSION BETWEEN THE TWO SUBSTRATES

Thermal expansion mismatch between workpieces affects joint clearances during the soldering process. Consider the enclosed geometry tube-in-socket joint in Fig. 17. When the CTE of the outer part is larger than that of the inner part, the gap will increase with heating. An excessive gap will result in voids in the joint that are caused by an insufficient solder supply and/or poor capillary action. A sample calculation follows:

- INNER PART AT 20 °C (68 °F), COPPER; CTE (α_1), 17.0 × 10^{-6}/K; OUTER DIAMETER, 74.7 MM (2.94 IN.)
- OUTER PART AT 20 °C (68 °F), ALUMINUM; CTE (α_0), 23.5 × 10^{-6}/K; INNER DIAMETER, 75.0

MM (2.95 IN.)

- GAP AT 20 °C (68 °F), 0.15 MM (0.006 IN.)

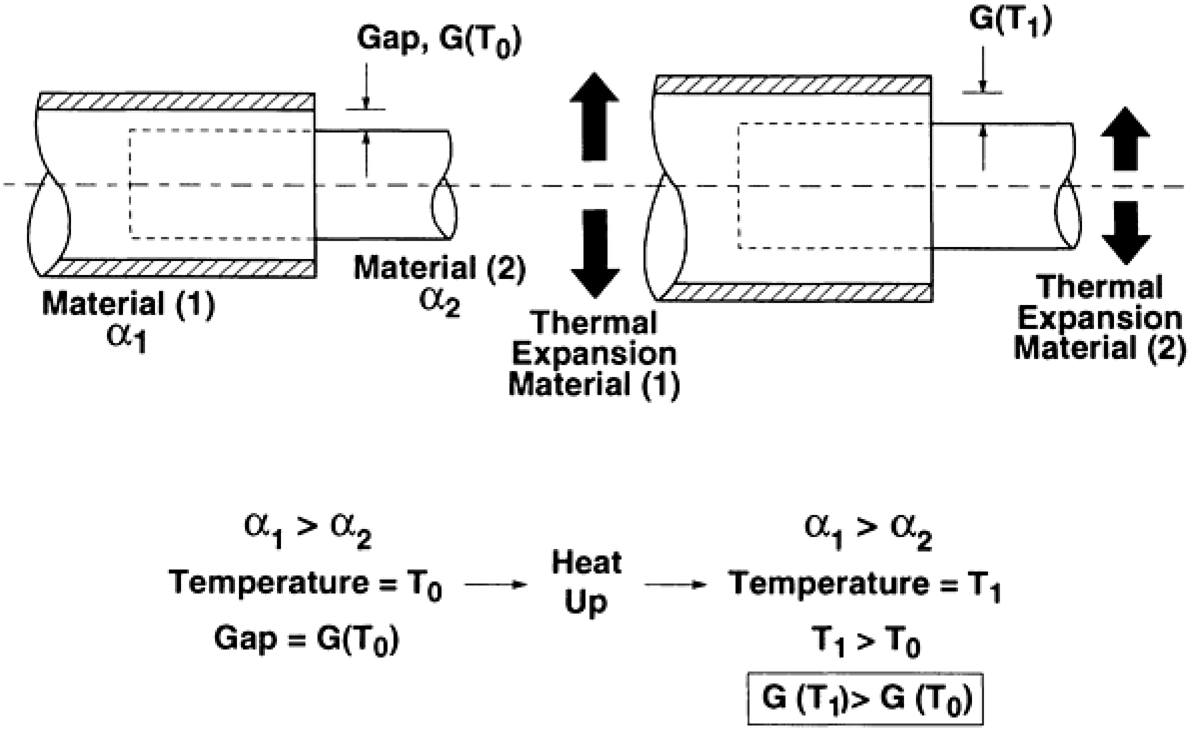

FIG. 17 EFFECTS OF HEATING ON THE JOINT GAP BETWEEN DISSIMILAR METALS

At 250 °C (482 °F), the new dimension, L_f, of the parts (the outer diameter of the aluminum and the inner diameter of the copper) are calculated by:

$$L_F = L_O + L_O \alpha \Delta T$$

where L_o is the dimension at 20 °C (68 °F), $\alpha$is the CTE of the material, and ΔT is the temperature difference (T_f - T_o).

- T_F = 250 °C (482 °F), SO THAT ΔT = 230 °C (414 °F)
- INNER PART AT 250 °C (482 °F), COPPER; OUTER DIAMETER, 75.0 MM (2.95 IN.)
- OUTER PART AT 250 °C (482 °F), ALUMINUM; INNER DIAMETER, 75.4 MM (2.97 IN.)
- GAP AT 250 °C (482 °F), 0.2 MM (0.008 IN.)

In this example, the solder joint gap grew by 0.05 mm (0.002 in.). This computation can be applied to any geometry in which L_o is the dimension of the respective parts.

Conversely, an inner part with a CTE that is greater than that of the outer section will cause the gap to decrease with heating, causing voids and poor solder filling resulting from the inability of flux volatiles to escape from the joint.

Finally, consideration must be given to material compatibility. The first concern is that the substrate materials are not damaged by chemicals or the mechanical procedures used to preclean surfaces. Damage includes excessive metal removal, stress-corrosion cracking, or the deposition of hard-to-remove residues. Secondly, the corrosion potential of joining dissimilar metals, including the two base metals as well as the solder/base metal couples, must be adequately assessed. In the case of conduit for liquids or gases, the role of the carried medium, as an electrolyte in the corrosion of the joint, must also be considered. In many cases, references to the electromotive series or fundamental chemistry textbooks (Ref 26) can provide corrosion information for a preliminary assessment. Finally, the assembled product must be compatible with the chemicals and processes used to remove flux residues after soldering.

Design for Processing Requirements. It is assumed that the substrates have adequate solderability for the selected solder alloy. Two processes must occur during product assembly to ensure an adequate joint: The solder must fill the gap area and flux volatiles and liquids must be ejected from the joint. Some general guidelines at the design stage will facilitate these processes.

First, the joint gaps should be in the range from 0.076 to 0.18 mm (0.003 to 0.007 in.). Workpiece design must also account for changes in gap dimensions caused by the heating/cooling of dissimilar substrates, surface roughness, plate flatness, eccentricity of circular-section workpieces, and consumption of the base metal or substrate coatings by the molten solder. Smaller gaps restrict solder flow and hinder the removal of flux volatiles, whereas larger gaps diminish the capillary effect required by the molten solder to fill the joint.

Second, the preferred gap geometry is a straight path opened to the atmosphere at two locations. One opening is used to supply flux and solder to the joint, while the other serves to vent flux volatiles and gases as the molten solder fills the gap. "Blind" gaps (or blind holes), which do not allow a vent path, should be avoided. A vent hole can be added and later filled in with a lower-melting-temperature solder if the design permits. Another rule of thumb is that solder has difficulty in "turning corners." Therefore, straight paths are preferred over sharp comers in the joint.

Third, ready access to the joint by a reservoir of solder containing an adequate quantity to fill the gap and to form the fillet must be provided. Inadequately filled gaps can be caused by an insufficient supply of solder to the joint. Poorly formed solder fillets reduce the monotonic strength, fatigue life, and creep strength of the joints, as well as limit the solderability assessment.

Fourth, solder joints that use preforms are always susceptible to void formation. Because the joint does not fill by capillary action, volatiles and gases have an increased likelihood of being trapped within the gap. Such joints must also be carefully assessed for corrosion potential by retained flux residues in those voids.

Fifth, the metallurgical reaction between the substrate and the solder during wetting necessarily consumes some of the substrate to form the reaction layer (typically, an intermetallic compound) at the interface. The dissolution of substrate material must be accounted for when soldering to very thin foils or coatings. The rate of dissolution will depend on substrate geometry (round or cylindrical sections dissolve slightly faster than flat plates), substrate composition, and the composition and volume of solder (flowing baths will remove more substrate metal than static baths). The temperature of the solder and the time of immersion (an increase of both raise the rate of substrate dissolution), are also factors. Dissolution data from unspecified geometries are determined from Fig. 18 and 19.

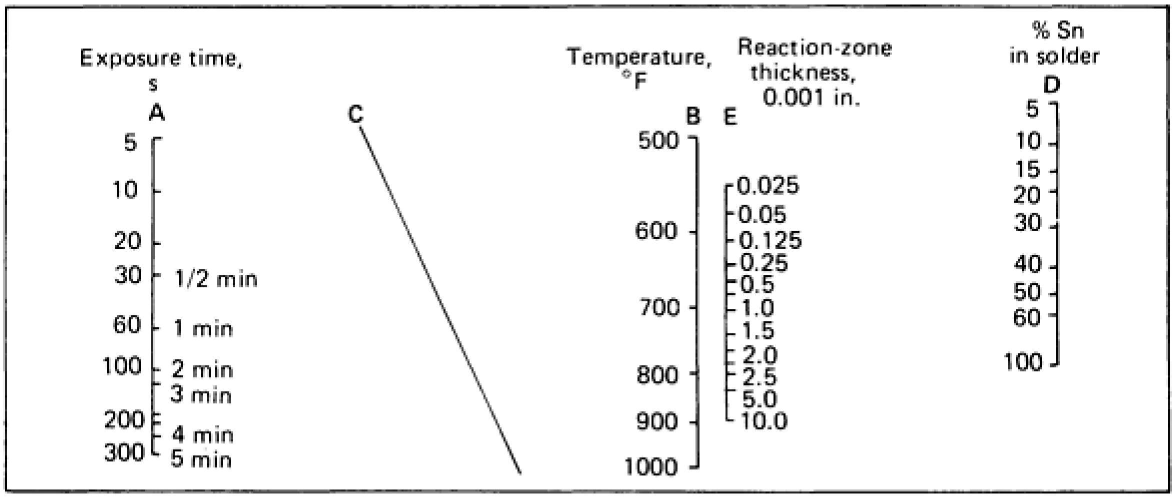

FIG. 18 NOMOGRAPH TO DETERMINE THE TIN-LEAD SOLDER REACTION ZONE THICKNESS FOR THE COPPER ALLOYS C10200 AND C26000 ALLOYS UNDER STATIC INFINITE-VOLUME CONDITIONS. SELECT THE EXPOSURE CONDITIONS ON SCALES A AND B. WITH A STRAIGHT EDGE BETWEEN THE CONDITIONS, FIND THE INTERSECTION WITH BASE LINE C. USING A STRAIGHT EDGE, CONNECT THIS POINT ON LINE C WITH THE

TIN CONTENT OF THE SOLDER ON SCALE D. READ THE REACTION-ZONE THICKNESS ON SCALE E. THE NOMOGRAPH CAN ALSO BE USED IN REVERSED SEQUENCE TO SELECT OPERATING CONDITIONS, GIVEN A PERMISSIBLE REACTION ZONE THICKNESS AND SOLDER COMPOSITION.

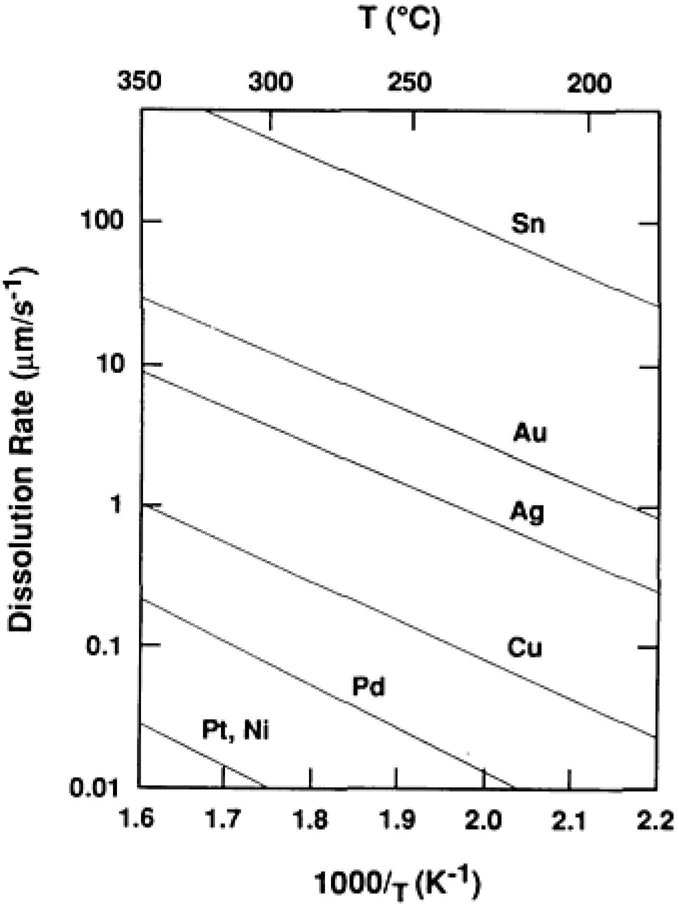

FIG. 19 METAL DISSOLUTION RATES FOR ELEMENTAL METALS AS A FUNCTION OF TEMPERATURE. SOURCE: REF 27

Intermetallic compounds represent the metallurgical reaction product between the solder and the substrate material. The compound formulations for common base metals and coatings with tin-bearing solders are:

- *COPPER:* CU_6SN_5 AND CU_3SN
- *NICKEL:* NI_3SN_4
- *GOLD:* $AUSN_4$ AND $AUSN_2$
- *IRON:* $FESN_2$
- *SILVER:* AG_3SN

These layers form when the solder is molten, as well as after solidification, by means of solid-state diffusion during storage or service at elevated temperature. In the solid state, the thickness is a function of the temperature and the time. Although intermetallic compounds are a necessary part of solder wetting, excessively thick layers can be detrimental to joint integrity or follow-up processing. Specifically, the brittle nature of intermetallic compounds can jeopardize the mechanical and fatigue strength of the joint, particularly at high loading rates, such as those that occur under impact or vibration conditions. Another detriment is that the solid-state growth of the layers consumes both the solder film and the base metal. Complete consumption of the tin or solder coating exposes the intermetallic surface to oxidation, making it very difficult to solder coat, even when using the strongest fluxes. The solid-state growth kinetics of several solders that were hot dip coated on copper are shown in Fig. 20. The typical as-soldered thickness of intermetallic compounds of

60Sn-40Pb on copper and nickel are 1 to 2 μm (39 μin.) and 0.5 μm (20 μin.), respectively. The layers grow very slowly at room temperature.

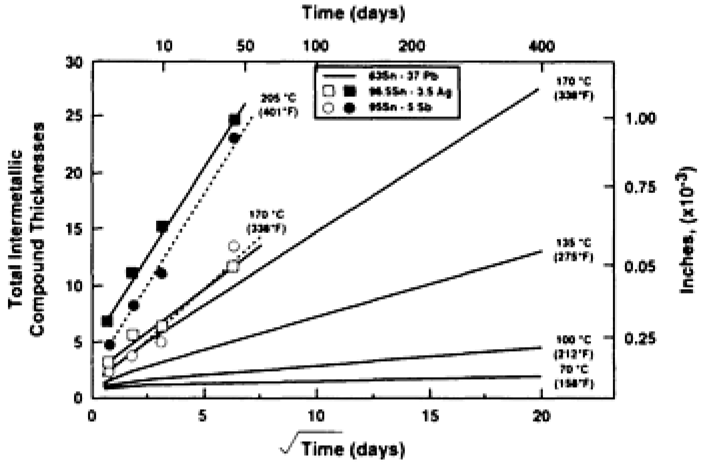

FIG. 20 SOLID-STATE INTERMETALLIC GROWTH KINETICS FOR HOT-DIPPED SOLDER COATINGS ON COPPER. SOURCE: SANDIA NATIONAL LABORATORIES

Joint design must include postassembly cleaning requirements. Materials compatibility between the cleaning agents and the workpiece, as well as the geometry of the workpiece, are important factors. Reentrant comers or hidden gaps and passages can entrap flux residues and prevent the flow of clean solvents into the areas.

The use of the high-temperature, precious-metal solders 80Au-20Sn or 88Au-12Ge on precious-metal substrates and coatings does not require a flux to remove the very limited oxide layers. The melting points of these solders preclude the use of rosin-base fluxes, because of significant high-temperature degradation. Other fluxes may also experience some degradation. The absence of a flux causes the surface tension of the solder to remain high, impeding capillary flow by the molten alloy. Therefore, the solders are typically introduced into the joint as preforms. Mechanical pressure on the joint is used to assist the spreading of the solder throughout the gap. Small quantities of oxide particles in the solder arise from the non-precious-metal component. Random spreading of the solder causes the joint to have a higher number of voids than is observed in a joint filled by capillary action.

Maintaining the position of the substrates during solder wetting is necessary for successful joint formation. Special fixturing can be designed to hold the workpieces for critical applications. Another approach is to design the joint itself to be self-jigging, in order to maintain the proper gaps and clearances during soldering (Fig. 21).

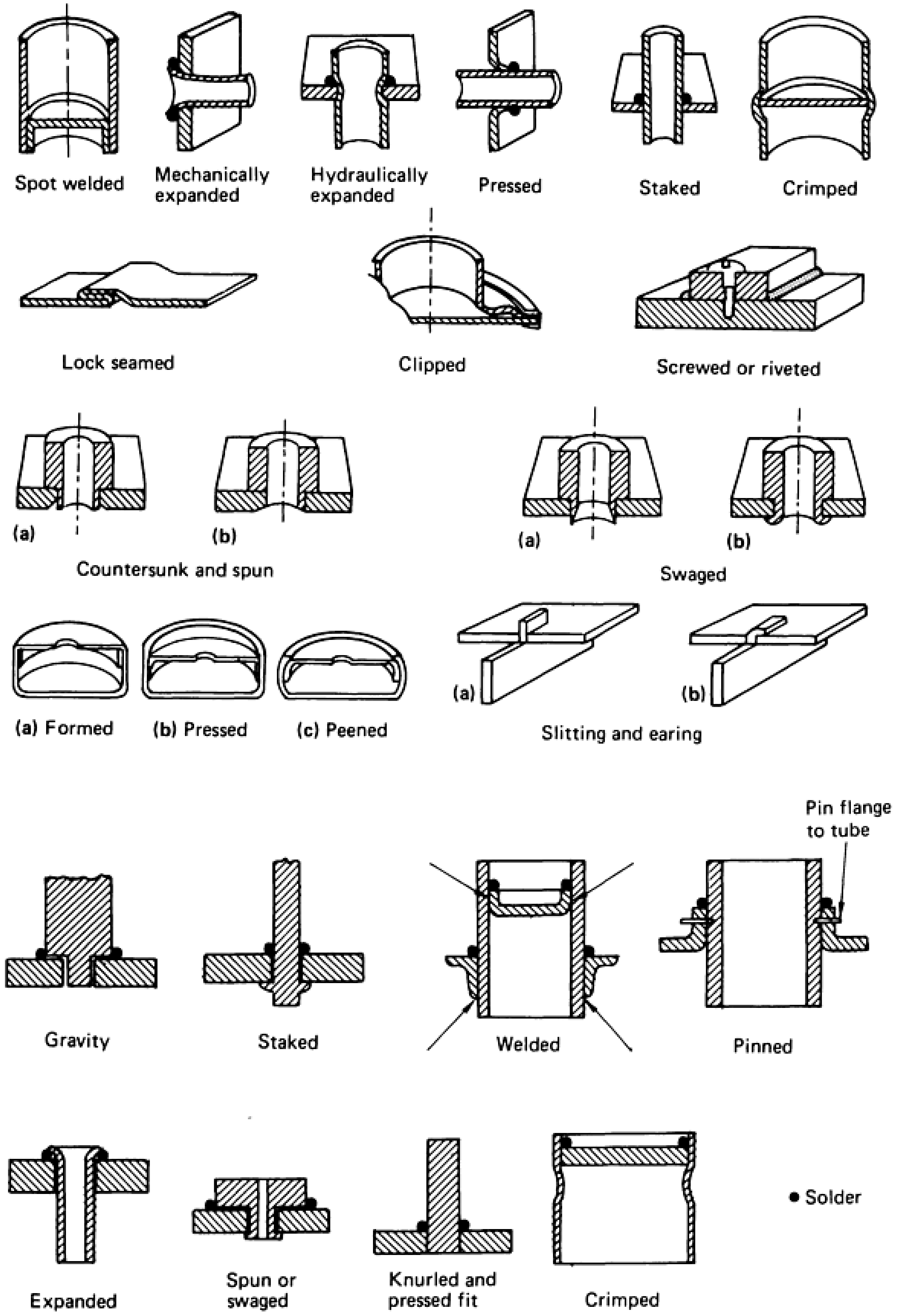
Spot welded
Mechanically expanded
Hydraulically expanded
Pressed
Staked
Crimped
Lock seamed
Clipped
Screwed or riveted
(a)
(b)
Countersunk and spun
(a)
(b)
Swaged
(a) Formed
(b) Pressed
(c) Peened
(a)
(b)
Slitting and earing
Pin flange to tube
Gravity
Staked
Welded
Pinned
Expanded
Spun or swaged
Knurled and pressed fit
Crimped
● Solder

FIG. 21 SELF-JIGGING JOINT CONFIGURATIONS

Design for Inspection Requirements. The inspection techniques selected to qualify a solder joint will have particular requirements in terms of joint configuration. Visual inspection is typically used to determine whether the solder has flowed within the gap and on the external surfaces (fillets). For blind holes or gaps, witness holes can be drilled into the substrate(s). During soldering, the solder fills the witness hole to the outside surface, indicating to the operator that the solder has indeed filled the gap. Although they are not always practical, unobstructed views of the solder joint gap at the external surface vastly improve the visual inspection of the joint.

References cited in this section

16. H. SOLOMON, "FATIGUE OF 60/40 SOLDER," REPORT 86CRD024, GENERAL ELECTRIC CO., MAY 1986
24. *BRAZING HANDBOOK,* AWS, 1991, P 267-276
25. C. THWAITES, *CAPILLARY JOINING,* WILEY AND SONS, UNITED KINGDOM, 1982, P 52
26. J. SCULLY, *THE FUNDAMENTALS OF CORROSION,* 2ND ED., PERGAMON PRESS, INC., 1975
27. C. LEA, A SCIENTIFIC GUIDE TO *SURFACE-MOUNT TECHNOLOGY,* ELECTROCHEM. PUBL., LTD., 1988

General Soldering

Paul T. Vianco, Sandia National Laboratories

Soldering Techniques

The numerous soldering processes have the common goal of providing heat energy to the joint area in the presence of the flux and solder alloy. The heat energy (and subsequent temperature rise) must support activation of the flux cleaning activities, heating of the substrate, and melting of the solder alloy to permit wetting and joint formation. Too little heat limits solder wetting, whereas excessive heat chemically degrades the flux (which loses activity and deposits hard-to-remove residues), reoxidizes the substrate surface prior to completion of solder wetting, or causes heat damage to the substrate.

Heating techniques can be categorized as global processes, in which the entire substrate reaches the soldering temperature, and local techniques, in which only the immediate joint area is heated. Examples of global techniques include vapor-phase (condensation) reflow, furnace (infrared or convection) reflow, wave soldering, ultrasonic soldering (bath), as well as dip (including hot-air leveling) and drag soldering. Local heating techniques include the soldering iron (heat and heat plus ultrasonics), torch, hot gas reflow, focused infrared soldering, induction soldering, and laser soldering. Operator interaction depends on the particular process. For example, the entire process can be fully automated, as in furnace systems or specialized equipment used in electronic circuit board assembly. Nevertheless, operators must establish the proper zone temperatures, conveyor belt speeds, and heating rates. Thermocouple profiling of a prototype workpiece is essential to verifying that the proper thermal conditions will be achieved. Manual soldering, such as with the iron or torch, requires greater operator interaction and skill.

Control of the full sequence of heating, fluxing, and applying the filler metal is particularly critical in manual solder processes (whether iron, induction, torch, or hot-gas processes), because they are usually performed in air. Competing factors that occur during substrate heating are:

- NEED FOR AN INCREASE IN SUBSTRATE TEMPERATURE TO MAINTAIN THE MOLTEN STATE OF THE SOLDER
- ACTIVATION OF THE FLUX OXIDE REMOVAL PROCESS (OR PROCESSES)
- ACCELERATED OXIDATION OF THE SUBSTRATE SURFACE
- DEGRADATION OF SOME FLUX COATINGS

Circuit board soldering and structural applications are described below as general examples.

Circuit Board Soldering. First, flux is applied to the leads and circuit board, unless flux-cored soldering wire is used, in which case this step is omitted. The coated area should extend beyond the immediate joint to ensure adequate wetting by the solder. Next, the soldering iron is made to contact the lead (never the circuit board land), and the solder wire is fed so that it contacts the opposite side of the lead. Melting of the solder indicates that the lead has reached temperature. The wire, followed by the iron tip, is then removed from the joint area upon formation of the joint fillet.

Structural Applications. The larger thermal mass of typical workpieces requires preheating in order to bring the joint area to the working temperature. Then, the flux is applied to fully cover the hot surface. The joint is further heated and the filler metal is applied to the joint surfaces. Most inorganic and organic fluxes can tolerate the high temperatures required to supply heat to larger members. The rosin-base fluxes can quickly degrade, as indicated by the formation of thick, blackened residues on the surface. Heat and solder are removed when an adequate fillet forms at the joint opening, in order to ensure complete filling of the joint gap.

The hot dip coating of parts can be performed by ultrasonic activation, without using a flux. Ultrasonic energy is coupled to the sample through the solder bath (Fig. 22). Optimum coupling is a function of the sample geometry, power level, and workpiece-horn separation. Oxide removal does not occur through simple line-of-sight erosion by the horn (Ref 28). Rather, the oxide is disrupted by the ultrasonic energy that is transferred from the horn into the substrate. Therefore, hidden surfaces can also be coated by the solder. Hot-dipped coatings can be applied to large workpieces or to leads on small electronic devices. In the latter case, care should be taken to ensure that internal connections are not damaged by the ultrasonic energy. This technology has been used to assemble tube-socket joints on heat-exchanger equipment (Ref 29).

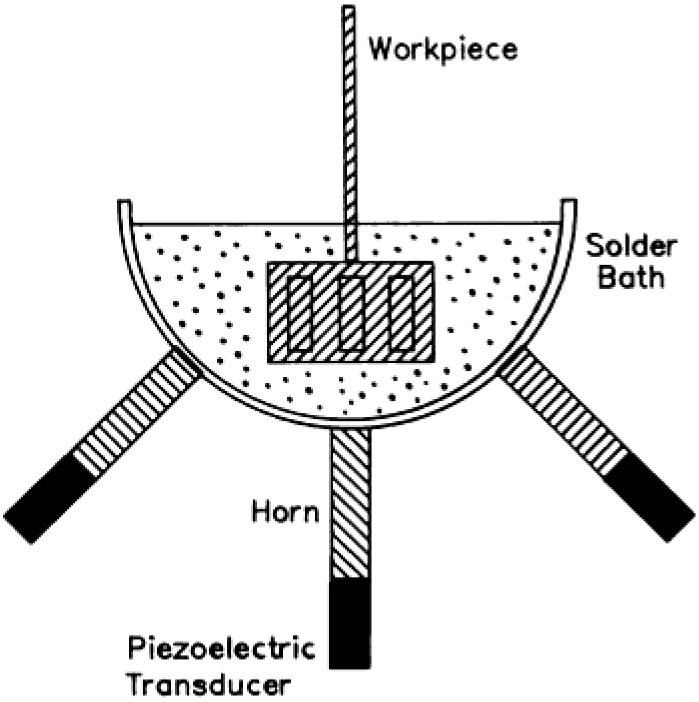

FIG. 22 ULTRASONIC SOLDERING BATH EQUIPMENT

An important component of the solder assembly process is the fixturing that is used to support the workpieces being assembled. Fixturing details are most critical for global heating processes in which the fixture will also be raised to the soldering temperature. The purpose of fixturing is to either anchor the two substrates to prevent movement while the solder is in the molten state or permit the controlled displacement of the substrates to establish specific joint gaps not achieved with the solder preform in the solid state. Lubricants include high-temperature petroleum products or solid lubricants (MoS_2 and graphite). Preloads on the workpieces are provided by springs (coil or Belleville configurations). The joint gap can be established by spacers between the workpieces or particles within the joint itself (Fig. 23). The spacer materials should not cause undesirable reactions with the solder or jeopardize the strength capacity of the joint.

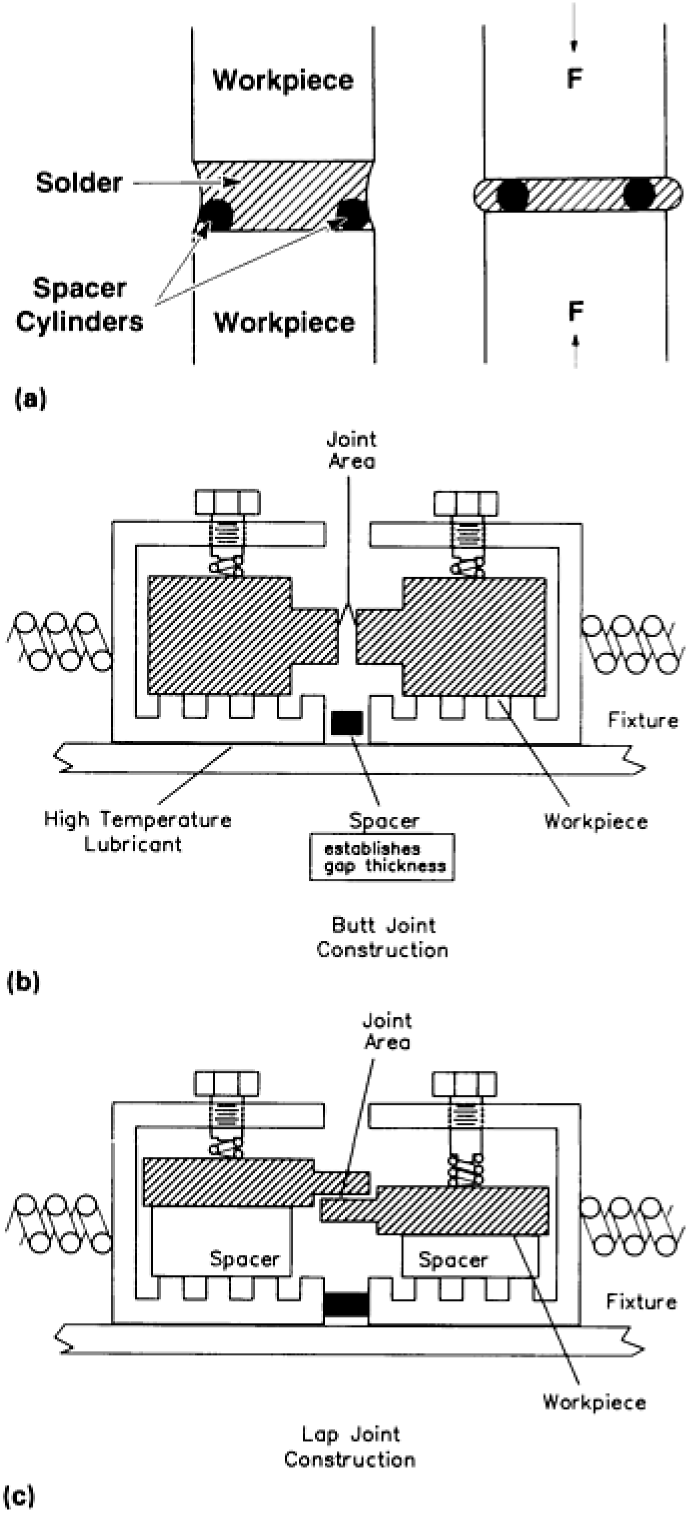
Workpiece
Solder
Spacer Cylinders
Workpiece
F
F
(a)
Joint Area
High Temperature Lubricant
Spacer
establishes gap thickness
Workpiece
Fixture
Butt Joint Construction
(b)
Joint Area
Spacer
Spacer
Fixture
Workpiece
Lap Joint Construction
(c)

FIG. 23 TECHNIQUES TO MAINTAIN JOINT GAPS. (A) BUTT JOINT WITH SPACERS IN THE SOLDER. (B) BUTT JOINT WITH SPACER IN THE FIXTURE. (C) LAP JOINT CONSTRUCTION

Fixture materials should be carefully assessed. Although high vapor pressures generated by fixture metals is not of concern at the relatively low temperatures used in soldering, factors such as thermal expansion and fixture size should be considered. For example, a CTE mismatch between the workpiece and the fixture materials, with improperly designed dimensions, can cause the joint gap to grow or close at soldering temperatures. Fixtures with large thermal mass will lengthen the heating and cooling times for the workpiece, leading to possible flux degradation, base metal or coating erosion, and longer production time. Preferred fixturing materials include oxide ceramics (alumina, beryllia, or mullite), refractory metals (molybdenum, tungsten, or tantalum), and well-oxidized steels (low-carbon and stainless steels). These materials will generally not be wetted by solder (and flux) spillage and can accommodate the elevated temperatures. Fixtures made of machined metals, such as aluminum, copper alloys, or steels, should be annealed in order to relieve any residual stresses that arise from the stock material or machining operations. These internal stresses may cause misalignment of the workpieces or the binding of moving parts, caused by warpage generated at soldering temperatures. Preoxidized surfaces will prevent inadvertent wetting by the solder.

Fixtures should be cleaned of organic residues to limit their outgassing in vacuum furnaces, prevent deposition of the contaminants onto the solderable surfaces, and limit spatter by their volatilization at soldering temperatures. Materials that contain or are coated with cadmium or zinc should not be used, because their high vapor pressures (even at soldering temperatures) may cause them to contaminate the joint area, as well as to poison vacuum and inert-atmosphere furnace systems.

The workpiece temperature should be monitored, at least in the prototype stages of process development, in order to document temperature conditions. In furnace operations, it is preferred that the controlling thermocouple contact the substrates to confirm the desired temperature profile at the joint area. The relatively low temperatures of soldering allow the use of inexpensive thermocouples, such as types J, T, and K, for many furnace cycles. The thermocouples should not interfere with the fixtures or the filling of the joint.

An undesirable spread of the molten solder on the substrate surface can be prevented by a number of "solder-stop" products. High-temperature tapes, which are popular in the electronics industry to prevent solder wetting of areas on printed circuit boards, are suited for many structural applications. Another product is a slurry made from powders that are mixed with water or alcohol and painted on the surface areas that are to be free of solder. The vehicle evaporates, leaving a coating that prevents spreading by the solder.

Detailed descriptions of the individual soldering processes, equipment, and applications are provided in the Section "Solid-State Welding, Brazing, and Soldering Processes" in this Volume.

References cited in this section

28. P. VIANCO AND F. HOSKING, *ANALYSIS OF ULTRASONIC TINNING,* NEPCON WEST, FEB 1992, P 1718
29. J. SCHUSTER AND R. CHILKO, ULTRASONIC SOLDERING OF ALUMINUM HEAT EXCHANGERS, *WELD. J.,* OCT 1975, P 711

Paul T. Vianco, Sandia National Laboratories

Postassembly Cleaning Procedures

After the soldering operation, the workpiece is cleaned, primarily to remove flux residues that can cause corrosion of the part while in storage or during service. Other cleaning procedures include the removal of solder-stop materials, as well as stray solder particles, which can interfere with mechanical or electrical performance of the assembly. Flux residues should be removed as soon as possible after the soldering process, because their ability to be removed decreases with time, whereas their tenacity and potential for corrosive damage increase with time.

Cleaning fluids and solvents to be used are determined by the particular flux residues. Guidelines are presented in the section "Fluxes" in this article. The selection of organic solvents is rapidly changing as chlorofluorocarbon materials are phased out by environmental regulations and new materials (and processes) become qualified as replacements.

Some general practices must be considered when establishing cleaning procedures. First, assess the compatibility of substrate materials and filler metals with the cleaning solutions. Organic solvents are benign toward metal surfaces. Alkaline solutions used to neutralize strongly acidic fluxes can affect some base-metal finishes (for example, copper alloys, iron alloys, and some steels). When in doubt, samples of the substrate should be exposed to the cleaning agent prior to use on final assemblies.

Second, determine whether the cleaning solutions leave undesirable residues. Mineral deposits from tap water can corrode or stain substrate surfaces (especially in the continued presence of water vapor). Tap-water rinses should be followed with rinses in either deionized or distilled water. All traces of water can be removed by a final alcohol rinse. Isopropyl alcohol is generally used. Denatured alcohol and acetone should be avoided, because they also can leave residues.

Third, the drying of solvent or other cleaning-solution residues should utilize dry, clean gas, such as bottled or cryogenic nitrogen gas. Compressed "house" air may contain water particles or compressor oils that can quickly recontaminate the workpiece.

Fourth, limit the contact of the assembly with oily rags and fingerprints. Many instances of cosmetic staining or pitting of the workpiece surface have been traced to fingerprints.

Fifth, test for the effectiveness of postcleaning operations. Unlike the well-specified procedures used by the electronics industry, such procedures are not well standardized for structural applications. Temperature-humidity chambers can be used to assess the propensity for corrosion on the workpiece (a destructive test).

Cleaning effectiveness can be enhanced by thermal and mechanical assistance. Cleaning solvents and solutions have higher solubility for residues at elevated temperatures. Caution should be observed when heating solutions because of the generation of vapors that can result in health or fire hazards. The use of solvent vapors at their boiling point in vapor degreasers can remove residues from remote locations on the workpiece. However, the solvents that have been popular for vapor degreasing are being restricted from use by environmental statutes.

Mechanical agitation of cleaning solutions is obtained by ultrasonic activation, high-energy sprays, and manual scouring procedures. Ultrasonics are very effective for loosening residues, particularly in hidden locations. Although generally safe for the cleaning of structural members, care must be exercised when using ultrasonics on electronic assemblies because of the possible damage to internal connections. The use of sprays or jets to force the cleaning solution into crevices and hidden areas of the workpiece can increase cleaning efficiency. Batch and in-line equipment based on spray and jet technology is currently available. Because the cleaning material passes through a jet, aerosols and mists are generated, which may create an explosion hazard.

Manual scouring can remove residues on exposed surfaces only. Cleaning with sandpaper or vapor blasting metal surfaces with abrasive particles should be avoided in the postprocess cleaning steps for three reasons: First, the base-metal oxide layer protects the surface from corrosion or excessive oxidation later in service. For example, stainless steels are particularly susceptible to corrosion attack after abrasive treatments, particularly those that use steel wool or a steel brush. Second, solders are generally much softer than the base metals. Therefore, inadvertent damage can be easily done to the joint fillets, possibly jeopardizing monotonic strength and fatigue resistance. Third, abrasive grit particles can become embedded in the substrate and, particularly, in the softer solder. Dislodged particles can damage mechanical actuators that are part of the soldered assembly. The abrasive particles can also deteriorate surface solderability of the workpiece during subsequent assembly, repair, or rework procedures. Damage occurs readily to circuit boards (solder masks, coatings, and the laminate itself) by abrasive particles.

Finally, storage of the parts must be considered. The extent of storage control depends on factors such as the type of assembly and its service requirements, the cost of rework, repair or scrapping of damaged parts, and the environment of the factory. Acute contamination or corrosion of finished parts can be prevented by their enclosure in bags. Popular containers are polyethylene plastic bags, which can provide short-term storage (<1 year) when properly sealed. Solderability may deteriorate for longer storage times because of the outgassing of polymer components contained in the plastic. Only new bags should be considered for the storage of solderable parts because contamination of the bag interior from previous usage may deteriorate solderability. Plastic bag material should be a minimum 0.1 mm (0.004 in.) thick. The bags can be filled with an inert gas (nitrogen) to further limit oxidation of solderable surfaces. Paper bags should be considered only for the transport of devices because their inability to be sealed from the environment, as well as contamination of solderable surfaces by chemicals in the paper, limit their appropriateness as storage containers. Metal foils can be used as storage media, although greater effort is required to seal the material. The use of desiccants that comply with MIL-D-3464E, "Desiccants, Activated, Bagged, Packaging Use and Static Dehumidification," is acceptable.

More elaborate and expensive storage measures include the construction of special facilities that control temperature, humidity, and air-particulate counts. The prevention of mechanical damage that results from the movement of inventory must also be addressed. Storage conditions are of particular concern for electronic devices and circuit boards in which solderability loss can severely hinder subsequent automated assembly processes.

General Soldering

Paul T. Vianco, Sandia National Laboratories

Postassembly Cleaning Procedures

After the soldering operation, the workpiece is cleaned, primarily to remove flux residues that can cause corrosion of the part while in storage or during service. Other cleaning procedures include the removal of solder-stop materials, as well as stray solder particles, which can interfere with mechanical or electrical performance of the assembly. Flux residues should be removed as soon as possible after the soldering process, because their ability to be removed decreases with time, whereas their tenacity and potential for corrosive damage increase with time.

Cleaning fluids and solvents to be used are determined by the particular flux residues. Guidelines are presented in the section "Fluxes" in this article. The selection of organic solvents is rapidly changing as chlorofluorocarbon materials are phased out by environmental regulations and new materials (and processes) become qualified as replacements.

Some general practices must be considered when establishing cleaning procedures. First, assess the compatibility of substrate materials and filler metals with the cleaning solutions. Organic solvents are benign toward metal surfaces. Alkaline solutions used to neutralize strongly acidic fluxes can affect some base-metal finishes (for example, copper alloys, iron alloys, and some steels). When in doubt, samples of the substrate should be exposed to the cleaning agent prior to use on final assemblies.

Second, determine whether the cleaning solutions leave undesirable residues. Mineral deposits from tap water can corrode or stain substrate surfaces (especially in the continued presence of water vapor). Tap-water rinses should be followed with rinses in either deionized or distilled water. All traces of water can be removed by a final alcohol rinse. Isopropyl alcohol is generally used. Denatured alcohol and acetone should be avoided, because they also can leave residues.

Third, the drying of solvent or other cleaning-solution residues should utilize dry, clean gas, such as bottled or cryogenic nitrogen gas. Compressed "house" air may contain water particles or compressor oils that can quickly recontaminate the workpiece.

Fourth, limit the contact of the assembly with oily rags and fingerprints. Many instances of cosmetic staining or pitting of the workpiece surface have been traced to fingerprints.

Fifth, test for the effectiveness of postcleaning operations. Unlike the well-specified procedures used by the electronics industry, such procedures are not well standardized for structural applications. Temperature-humidity chambers can be used to assess the propensity for corrosion on the workpiece (a destructive test).

Cleaning effectiveness can be enhanced by thermal and mechanical assistance. Cleaning solvents and solutions have higher solubility for residues at elevated temperatures. Caution should be observed when heating solutions because of the generation of vapors that can result in health or fire hazards. The use of solvent vapors at their boiling point in vapor degreasers can remove residues from remote locations on the workpiece. However, the solvents that have been popular for vapor degreasing are being restricted from use by environmental statutes.

Mechanical agitation of cleaning solutions is obtained by ultrasonic activation, high-energy sprays, and manual scouring procedures. Ultrasonics are very effective for loosening residues, particularly in hidden locations. Although generally safe for the cleaning of structural members, care must be exercised when using ultrasonics on electronic assemblies because of the possible damage to internal connections. The use of sprays or jets to force the cleaning solution into crevices and hidden areas of the workpiece can increase cleaning efficiency. Batch and in-line equipment based on spray and jet technology is currently available. Because the cleaning material passes through a jet, aerosols and mists are generated, which may create an explosion hazard.

Manual scouring can remove residues on exposed surfaces only. Cleaning with sandpaper or vapor blasting metal surfaces with abrasive particles should be avoided in the postprocess cleaning steps for three reasons: First, the base-metal oxide layer protects the surface from corrosion or excessive oxidation later in service. For example, stainless steels are particularly susceptible to corrosion attack after abrasive treatments, particularly those that use steel wool or a steel brush. Second, solders are generally much softer than the base metals. Therefore, inadvertent damage can be easily done to the joint fillets, possibly jeopardizing monotonic strength and fatigue resistance. Third, abrasive grit particles can become embedded in the substrate and, particularly, in the softer solder. Dislodged particles can damage mechanical actuators that are part of the soldered assembly. The abrasive particles can also deteriorate surface solderability of the workpiece during subsequent assembly, repair, or rework procedures. Damage occurs readily to circuit boards (solder masks, coatings, and the laminate itself) by abrasive particles.

Finally, storage of the parts must be considered. The extent of storage control depends on factors such as the type of assembly and its service requirements, the cost of rework, repair or scrapping of damaged parts, and the environment of the factory. Acute contamination or corrosion of finished parts can be prevented by their enclosure in bags. Popular containers are polyethylene plastic bags, which can provide short-term storage (<1 year) when properly sealed. Solderability may deteriorate for longer storage times because of the outgassing of polymer components contained in the plastic. Only new bags should be considered for the storage of solderable parts because contamination of the bag interior from previous usage may deteriorate solderability. Plastic bag material should be a minimum 0.1 mm (0.004 in.) thick. The bags can be filled with an inert gas (nitrogen) to further limit oxidation of solderable surfaces. Paper bags should be considered only for the transport of devices because their inability to be sealed from the environment, as well as contamination of solderable surfaces by chemicals in the paper, limit their appropriateness as storage containers. Metal foils can be used as storage media, although greater effort is required to seal the material. The use of desiccants that comply with MIL-D-3464E, "Desiccants, Activated, Bagged, Packaging Use and Static Dehumidification," is acceptable.

More elaborate and expensive storage measures include the construction of special facilities that control temperature, humidity, and air-particulate counts. The prevention of mechanical damage that results from the movement of inventory must also be addressed. Storage conditions are of particular concern for electronic devices and circuit boards in which solderability loss can severely hinder subsequent automated assembly processes.

General Soldering

Paul T. Vianco, Sandia National Laboratories

Inspection

Because inspection necessarily implies a determination of defects in the workpiece, the type and number of defects that jeopardize the service requirements of the product should be documented. The search for, and rework of, noncritical defects (such as unnecessary cosmetic attributes) simply increases product cost, the amount of scrapped product, as well as the likelihood of heat damage to the substrates. Defects and their acceptance limits should be defined in conjunction with the design engineer, who understands service requirements and codes, and the manufacturing engineer, who understands the process limitations. Several defect types and testing procedures are summarized below. The defects include:

- *INCOMPLETE FILLING OF THE JOINT*, CAUSED BY POOR SOLDERABILITY (NONWETTING OR DEWETTING), AN INADEQUATE QUANTITY OF FILLER METAL, LOW REFLOW TEMPERATURE, FLUX ENTRAPMENT (HOLES AND VOIDS), OR POOR JOINT DESIGN (REENTRANT CORNERS OR BLIND HOLES THAT PREVENT THE ESCAPE OF GASES)
- *POOR FILLET GEOMETRY*, CAUSED BY POOR SOLDERABILITY OF THE BASE METALS, AN INADEQUATE SUPPLY OF FILLER METAL FOR THE GIVEN JOINT GEOMETRY, OR INSUFFICIENT COVERAGE OF THE SURFACES BY THE FLUX COATING
- *CRACKS IN THE SOLDER FILLET*, WHICH SIGNIFY PART MOVEMENT DURING THE SOLIDIFICATION PROCESS, SOLIDIFICATION SHRINKAGE OF THE SOLDER METAL, OR, IN THE WORST CASE, A MAJOR FLAW IN THE JOINT DESIGN IN WHICH RESIDUAL STRESSES (FOR EXAMPLE, CTE MISMATCH) OVERLOAD THE SOLDER JOINT. THE LATTER CONDITION CAN CAUSE CRACKS IN THE SUBSTRATES, PARTICULARLY IN

BRITTLE MATERIALS, SUCH AS CERAMICS OR REFRACTORY METALS

- *GRAINY OR VERY DULL FILLET SURFACES,* WHICH INDICATE EXCESSIVE SOLDER CONTAMINATION, SUCH AS THAT ARISING FROM THE DISSOLUTION OF THE BASE-METAL MATERIALS AND SURFACE COATINGS, INADEQUATE HEATING OF THE SUBSTRATES, OR SUBSTRATE MOVEMENT DURING SOLIDIFICATION. (HOWEVER, SOME SOLDERS NORMALLY EXHIBIT A DULL FINISH UPON SOLIDIFICATION. SLOW COOLING, ESPECIALLY IN VAPOR-PHASE SOLDERING, CREATES DULL JOINTS)
- *FLUX RESIDUES,* WHICH SUGGEST INADEQUATE CLEANING PROCEDURES OR THE OVERHEATING OF THE FLUX TO PRODUCE VERY TENACIOUS RESIDUES. LARGER PARTICLES MAY BE EMBEDDED IN THE SOLDER FILLET.
- *DISCOLORATION IN THE BASE METAL,* WHICH SIGNIFIES EXCESSIVE HEAT EXPOSURE

Nondestructive inspection techniques range from visual assessment to elaborate tests involving complex equipment and procedures. The principal nondestructive testing methods are discussed below.

Visual inspection is the most widely used technique. Pictorial guides help the inspector qualify the particular joint configuration. Because the inspector can only infer joint quality from external observations, some destructive assessments should accompany the visual evaluations on early prototypes to correlate with subsequent inspector observations on the actual product. Low-power microscopes (<70×) are used.

Radiographic (x-ray) and ultrasonic inspection are used to visualize the interior sections of the joint. X-ray radiographs are used to detect mass differences in the joint, which indicate areas that are absent of solder, such as gas voids or entrapped flux. Ultrasonic techniques detect discontinuities in the path between the receiver and the transmitter (transmission mode) or in the reflected path (reflection mode). Besides unfilled sections of the gaps, this process also detects discontinuities such as poor bonding or cracks.

Infrared (thermal transfer) imaging is based on the transfer of heat across the joint gap. Voids and, to a lesser extent, cracks, have a lower thermal conductivity than the continuous base metal-solder-base metal path. The decreased thermal conduction of these defects is imaged by an infrared-sensitive camera that views the opposite side of a heated substrate. This technique has relatively poor resolution because of lateral heat conduction within the materials.

Proof testing involves subjecting the solder joint to a mechanical load that exceeds the service design load. The joint is then inspected for damage using nondestructive techniques. In the prototype phase of process development, this procedure may be considered "destructive."

Pressure and vacuum testing represent other nondestructive techniques. In pressure testing, which is detailed in the ASME Boiler and Pressure Vessel Code, a positive pressure is applied to conduit joints or soldered vessels. The joint is either coated with a soapy solution or submerged in a water bath before being pressurized. Leaks are identified by the formation of bubbles on the workpiece surface. Air can be replaced by helium or a halogen gas, and special detection equipment can be used to locate leaks in the joints. Vacuum checks of conduit also require special equipment to detect the leakage of trace gases, such as helium, into the conduit or vessel. Mass spectrometers can be used to detect and identify the passage of other gas species into the vacuum. The detector monitors the inside of the conduit (vacuum environment) while the gas is passed along the exterior walls.

Fluorescent dye penetration involves the introduction of a fluorescent dye to one side of the joint. The opposite side is then inspected for dye leakage, which would indicate a continuous path of cracks or voids.

Destructive testing techniques include all types of mechanical testing techniques as well as metallographic sectioning for microstructural examination.

Mechanical testing involves numerous techniques that have been standardized to examine tensile, shear, peel, impact, and torsion strengths, as well as the fatigue life of adhesive joints (Table 10). The various strength measurements can be extended to quantify defects in the joints, such as voids or cracks that cannot support a load, or a microstructural modification to the solder metal, which changes its strength. Inspection of the fracture surface provides critical information on the failure mechanism, as well as data on void content and nonwetting that are caused by poor solderability.

Additional information is available in the article "Evaluation and Quality Control of Soldered Joints" in this Volume.

General Soldering

Paul T. Vianco, Sandia National Laboratories

Environmental, Safety, and Health Issues

Both the materials and processes used in soldering present some general, as well as unique, environmental, safety, and health considerations. There is an increased awareness of the environmental damage caused by manufacturing processes. Applicable rules and regulations can be set by local, state, and federal legislation, as well as regulations developed through the Environmental Protection Agency (EPA). The environmental concerns that relate to soldering operations include:

- RELEASE OF CLEANING SOLVENTS AND SOLUTIONS INTO THE LIQUID WASTE STREAM
- VENTING OF SOLVENT FUMES INTO THE ATMOSPHERE
- DISPOSAL OF HEAVY-METAL SOLID WASTES, SUCH AS SOLDER DROSS AND SCRAP MATERIAL

Rules and regulations that govern worker safety are established by local, state, and federal employment statutes. The federal agency responsible for worker safety is the Occupational Safety and Health Administration (OSHA). Industrial organizations, such as the National Fire Protection Agency, the Underwriters Laboratory, and many insurance companies, have developed excellent safety guidelines for industrial practices. Some safety concerns in soldering operations include:

- BURNS FROM MOLTEN METAL AND ELEVATED-TEMPERATURE PROCESSING EQUIPMENT
- BURNS FROM THE IGNITION OF FLAMMABLE SOLVENTS USED IN CLEANING PROCESSES OR AS FLUX VEHICLES
- CHEMICAL BURNS FROM CORROSIVE ACIDS IN CLEANING SOLUTIONS AND FLUXES, THE LATTER OF WHICH INCLUDES SPATTER DURING SOME SOLDER PROCESSES
- CHEMICAL BURNS FROM ALKALINE MATERIALS USED IN POSTASSEMBLY CLEANING

An extensive effort has been made to identify and control the worker health risks of industrial soldering operations. Unlike safety issues, which tend to address acute dangers, health concerns can manifest themselves in accumulated toxicity to the worker which may not be identified for months or years. Besides the legislative bodies, regulations are also set by agencies such as OSHA, EPA, and the U.S. Mine Safety and Health Administration (MSHA). In addition, various guidelines are provided by organizations such as the Food and Drug Administration (FDA) and the National Institutes of Health--National Cancer Institute. Some of the health issues associated with soldering operations are:

- EXPOSURE TO ORGANIC SOLVENT FUMES (ACETONE, ALCOHOLS) IN FLUXES AND CLEANING COMPOUNDS OR THEIR HEAT-GENERATED BY-PRODUCTS (FOR EXAMPLE, ALDEHYDES FROM ROSIN-BASE FLUXES)
- ILLNESS FROM THE INHALATION OF ACID OR ALKALINE FUMES USED IN CLEANING SOLUTIONS
- HEAVY-METAL TOXICITY FROM MOLTEN SOLDER FUMES, BASE-METAL COATINGS, AND FROM THE BASE METALS THEMSELVES (AIRBORNE PARTICLES OF BERYLLIUM, CHROMIUM STEELS, AND SUCH)

Exposure to toxins occurs through skin absorption, ingestion, or inhalation. The use of protective clothing and face protection, as well as the practice of good hygiene habits, will prevent the absorption and ingestion, respectively, of toxic materials. Because the most likely source of accidental exposure is inhalation, some guidelines and exposure limits are described below for materials and processes used in soldering operations. Air-sampling procedures have been established and should be used to certify compliance with these specifications. The exposure limits provide some insight into the relative hazard presented by each of the listed materials. It is important to note that regulations and exposure limits can change frequently as data are accumulated by the appropriate agencies. Therefore, it is recommended that the reader consult the necessary documents or agency representative to ensure compliance with the most recent guidelines.

The exposure limits are expressed as the time-weighted average (TWA), which is equivalent to the threshold limit value (TLV) established by the American Conference of Governmental Industrial Hygienists (ACGIH). The TWA limits were also accepted by OSHA in 1972 as the legal permissible exposure limits (PEL). The TWA value represents the time-averaged airborne concentration of the material over an 8-h period, which is assumed to be the exposure period per day.

Fluxes and Cleaning Agents. The primary concern when using rosin-base fluxes are the thermal breakdown products generated upon their exposure to the molten solder temperature. These reaction products are primarily the aliphatic aldehydes, typically measured by the equivalent formaldehyde concentration (Ref 30). The adopted TWAs for these substances, together with values for organic solvents used as flux vehicles or cleaning agents, are listed in Table 38. No limits are available for the typical constituents of organic acid fluxes (lactic acid, benzoic acid, or glutamic acid) or any suspected thermal breakdown products. The exposure levels of fumes from the inorganic acids, whether used as the flux or precleaning material, and from sodium hydroxide (neutralizing agent) are listed in Table 39. New regulations and exposure limits change as new data become available. The reader should remain aware of updates to these guidelines for current practices.

TABLE 38 TWA LIMITS FOR ORGANIC SUBSTANCES USED IN SOLDER PROCESSING

MATERIAL	**TWA STANDARD, MG/M^3**	COMMENTS
FORMALDEHYDE (AS A ROSIN PYROLYSIS PRODUCT)	0.1	**REPRESENTATIVE OF ROSIN FLUX BY-PRODUCTS FROM HEAT EXPOSURE**
LSOPROPYL ALCOHOL	980	**SKIN EXPOSURE$^{(A)}$**
ETHYL ALCOHOL	1900	
METHYL ALCOHOL	260	**SKIN EXPOSURE**
METHYL CHLORIDE	105	**LEVEL OF INTENDED CHANGE AS OF 1981**
METHYLENE CHLORIDE (DICHLOROMETHANE)	360	**LEVEL OF INTENDED CHANGE AS OF 1981**
CARBON TETRACHLORIDE	30	**SKIN EXPOSURE; SUSPECTED CARCINOGEN. LEVEL OF INTENDED CHANGE AS OF 1981**
BENZENE	30	
ACETONE	1780	**LEVEL OF INTENDED CHANGE AS OF 1981**
ETHYLENE GLYCOL	**10$^{(B)}$**	**PARTICULATE**

Source: Ref 30

(A) INDICATES THAT A CUTANEOUS ROUTE (MUCOUS MEMBRANES, EYES, ETC.) CAN ALSO CONTRIBUTE TO OVERALL EXPOSURE.

(B) VAPOR, 125 MG/M^3.

TABLE 39 TWA LIMITS FOR INORGANIC ACIDS AND ALKALINES

MATERIAL	TWA STANDARD, MG/MM^3
HYDROCHLORIC ACID (HCL)	**7**
SULFURIC ACID (H_2SO_4)	**1**
NITRIC ACID (HNO_3)	**5**
HYDROFLUORIC ACID (HF)	**2.5**
PHOSPHORIC ACID (H_3PO_4)	**1**
SODIUM HYDROXIDE (NAOH)	**2**

ZINC CHLORIDE ($ZNCL_2$)	**1**
AMMONIUM CHLORIDE (NH_4CL)	**10**

Source: Ref 30

Solder and Base-Metal Constituents. The acute and chronic side effects of human exposure to metals are becoming better understood by the medical community. Toxicity hazards associated with the so-called "heavy" metals (lead, cadmium, antimony, tin, and others) used in the soldering processes are of particular concern to the health and regulatory agencies. Two mechanisms are responsible for airborne particulates: evaporation of metals by heat energy, which is applied to the molten solder or substrate, and mechanical agitation, which releases dust or abrasive particles.

Molten solder is generally used at temperatures well below the boiling point of either the solder or the substrate material. Exceptions are workpieces constructed with high vapor pressure metals or coatings, such as cadmium or zinc, or solders that contain these elements. Molten solders, like other liquids, release greater amounts of metal vapors as their temperature increases toward their vaporization point. Manual soldering processes, which exhibit the least degree of temperature control, are most susceptible to overheating of the solder and are therefore most likely to increase levels of toxic fumes in the atmosphere. Unfortunately, such processes also require the operator to be in close proximity to the workpiece. Therefore, proper ventilation must be maintained in the work area to minimize worker exposure.

Mechanical agitation of the molten solder for prolonged processing times, during which the solder is liquid, can accelerate the release of metal fumes. Airborne metal fumes and particulates are also caused by the agitation of solder dross and through abrasion of the solid metals. The dross that forms on the surface of solder baths can be a source of airborne particulates of tin, lead, or antimony, for example, when it is removed for disposal. Base-metal particles are generated by the use of abrasive precleaning techniques, such as grit or sand-blasting techniques. Base-metal particles and solder particles can also be generated when similar techniques are used to remove residues from the finished joints.

Table 40 gives the TWA values of metals that are used in solders or that compose substrate materials. Particulates of lead, cadmium, indium, platinum, silver, and chromium VI compounds are particularly hazardous, whereas those of zinc, copper, tin, and aluminum are relatively benign. Data are not available for bismuth and gold, a commonly used finish layer.

TABLE 40 TWA LIMITS FOR METAL PARTICULATES

MATERIAL	**TWA STANDARD, MG/M^3**	COMMENTS
LEAD	0.05	**FUMES AND DUST; ACTION LEVEL, 0.03**
TIN	10	**TIN OXIDE AS TIN, NUISANCE; FOR PARTICULATE, AUTOMATIC LEVEL SET AT 10 MG/M$^{3(A)}$**
ANTIMONY	0.5	**INCLUDES ALL COMPOUNDS**
SILVER	0.1	**METAL DUST AND FUMES**(B)
INDIUM	0.1	**INDIUM AND COMPOUNDS**
BISMUTH	NONE AVAILABLE	**. . .**
ZINC	5	**ZINC OXIDE FUMES**
CADMIUM	0.05	**SUSPECTED CARCINOGEN**
GOLD	NONE AVAILABLE	**. . .**
PLATINUM	0.002	**. . .**
COPPER	1	**DUST AND MIST**
ALUMINUM	10	**OXIDE, NUISANCE PARTICULATE**
NICKEL	1(C)	**METAL**
IRON	5	**IRON OXIDE FUMES**

CHROMIUM	**0.5**	**SOLUBLE CHROMIC OR CHROMOUS SALTS, METALS**$^{(D)}$

Source: Ref 30

(A) LEVEL OF INTENDED CHANGE AS OF 1981, 2 MG/M^3.
(B) SOLUBLE COMPOUNDS, 0.01,
(C) SOLUBLE COMPOUNDS, 0.1.
(D) CR VI COMPOUNDS, CHROMATES, 0.05 MG/M^3.

Lead regulations governing airborne lead concentrations in the workplace are very complex. The maximum airborne concentration (TWA) is 0.05 mg/m^3. However, an "action level" was established at 0.03 mg/m^3, above which a list of measures identified by OSHA regulations must be carried out. Below the action level, employees need only be monitored for lead exposure and receive adequate job training to minimize potential hazards.

Data were accumulated in the late 1970s that measured the blood lead levels of workers performing various soldering operations, as well as the airborne lead concentration of a number of soldering processes. Table 41 shows the blood lead levels of eight persons from a group of 37 who performed one of the tasks under the "job description." No significant difference was observed between these values (mean of the entire group of 37 was 11.7 μg/100 mL) and the blood lead level in the normal population with no occupational exposure to lead (10 to 30 μg/100 mL).

TABLE 41 BLOOD LEAD LEVELS IN WORKERS PERFORMING VARIOUS SOLDERING OPERATIONS

EMPLOYEE	**BLOOD LEAD LEVEL, μG/100 ML**	JOB DESCRIPTION
1	15	**MASS SOLDERING**
2	14	**HAND IRON SOLDERING**
3	14	**HAND IRON SOLDERING**
4	14	**HAND IRON SOLDERING**
5	13	**OCCASIONAL HAND SOLDERING**
6	11	**LAYOUT OPERATOR**
7	17	**MASS SOLDERING**
37	**19**	**MASS SOLDERING**

Note: Mean blood level for 37 employees: 11.7 μg/100 mL.

(A) VALUES CITED ARE TYPICAL FOR ALL 37 EMPLOYEES.

Airborne levels for several soldering operations and industries are shown in Table 42. None of the examples matched the maximum level of 0.05 mg/m^3. However, radiator manufacturing and repair typically exceeded the "action" level of 0.03 mg/m^3. This study emphasized that adequate ventilation of the work area was the most effective means to minimize airborne concentrations.

TABLE 42 AIRBORNE LEAD CONCENTRATIONS FOR SEVERAL SOLDERING OPERATIONS

TYPE OF SOLDERING	TYPICAL AIRBORNE LEAD CONCENTRATIONS$^{(A)}$, MG/M^3
AUTOMATIC WAVE	
ELECTRONICS ASSEMBLY	**0.010**
MANUAL GUN AND IRON	

ELECTRONICS AND ELECTROMECHANICAL ASSEMBLY	**0.010**
RADIATOR MANUFACTURE AND REPAIR	**0.028**
OPEN DIP POT TINNING	
ELECTRONICS, ELECTROMECHANICAL ASSEMBLY, AND TOOL AND DIE REPAIR	**0.004**
RADIATOR MANUFACTURE AND REPAIR	**0.046**
OPEN-FLAME JOINING	
ARTS AND CRAFTS	**0.015**
RADIATOR MANUFACTURE AND REPAIR	**0.038**

(A) MEAN

Reference cited in this section

30. *PRUDENT PRACTICES FOR HANDLING HAZARDOUS CHEMICALS IN LABORATORIES,* NATIONAL ACADEMY PRESS, 1981, P 257-276

General Soldering

Paul T. Vianco, Sandia National Laboratories

References

1. "STANDARD SPECIFICATION FOR SOLDER METAL," B 32, *ANNUAL BOOK OF ASTM STANDARDS,* ASTM
2. "SOLDER, TIN ALLOY: TIN-LEAD ALLOY; AND LEAD ALLOY," FEDERAL SPECIFICATION QQ-S-571E, SUPERINTENDENT OF DOCUMENTS, U.S. GOVERNMENT PRINTING OFFICE, WASHINGTON, DC
3. "SOFT SOLDER ALLOYS--CHEMICAL COMPOSITIONS AND FORMS," ISO/DIS 9453, INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, THE HAGUE, NETHERLANDS
4. AMERICAN METALS MARKET, FAIRCHILD PUBLICATIONS, 17 JAN 1992
5. *SOLDERING MANUAL,* AWS, 1977, P 5
6. H. MANKO, *SOLDERS AND SOLDERING,* MCGRAW-HILL, 1979, P 14
7. H. MANKO, *SOLDERS AND SOLDERING,* MCGRAW-HILL, 1979, P 82
8. B. LAMPE, ROOM TEMPERATURE AGING PROPERTIES OF SOME SOLDER ALLOYS, *WELD. J., RES. SUPPL.,* OCT 1976, P 330S
9. SOLDERING MANUAL, AWS, 1977, P 5
10. R. KLEIN-WASSINK, *SOLDERING IN ELECTRONICS,* ELECTROCHEMICAL PUB. LTD., AYR, SCOTLAND, 1989, P 189
11. H. MANKO, *SOLDERS AND SOLDERING,* MCGRAW-HILL, 1979, P 89
12. S. NIGHTINGALE *ET AL.,* TIN SOLDERS, BRITISH NONFERROUS METALS RESEARCH ASSOCIATION, VOL 1, 1942
13. "SOLDER ALLOY DATA, MECHANICAL PROPERTIES OF SOLDERS AND SOLDERED JOINTS," PUBLICATION 656, INTERNATIONAL TIN RESEARCH INSTITUTE, UNITED KINGDOM, 1986
14. *DEVELOPMENT OF HIGHLY RELIABLE SOLDER JOINTS FOR PRINTED CIRCUIT BOARDS,* WESTINGHOUSE DEFENSE AND SPACE CENTER, 1968, P 4-55 TO 4-57
15. R. WILD, FATIGUE PROPERTIES OF SOLDER JOINTS, *WELD. RES. J.,* VOL 15, 1972, P 521S
16. H. SOLOMON, "FATIGUE OF 60/40 SOLDER," REPORT 86CRD024, GENERAL ELECTRIC CO.,

MAY 1986

17. J. JONES AND J. THOMAS, "ULTRASONIC SOLDERING OF ALUMINUM," RESEARCH REPORT 55-24, FRANKFORD ARSENAL, DEPARTMENT OF THE ARMY, PHILADELPHIA, FEB 1955
18. H. MANKO, *SOLDERS AND SOLDERING,* MCGRAW-HILL, 1979, P 112-115
19. D. OLSEN AND H. BERG, PROPERTIES OF DIE BOND ALLOYS RELATING TO THERMAL FATIGUE, *IEEE TRANS.,* CHMT-2, 1979, P 257
20. P. VIANCO AND J. REJENT, "SOLDER BOND APPLICATIONS IN A PIEZOELECTRIC SENSOR ASSEMBLY," *45TH ANNUAL FREQ. CONTROL SYMP.* (LOS ANGELES, CA), MAY 1991, P 266
21. R. KLEIN-WASSINK, *SOLDERING IN ELECTRONICS,* ELECTROCHEMICAL PUB. LTD., AYR, SCOTLAND, 1989, P 189
22. *SOLDERING MANUAL,* AWS, 1977, P 35-39
23. P. VIANCO *ET AL.,* SOLDERABILITY TESTING OF KOVAR WITH 60SN-40PB SOLDER AND ORGANIC FLUXES, *WELD. J. RES. SUPPL.,* JUNE 1990, P 230S
24. *BRAZING HANDBOOK,* AWS, 1991, P 267-276
25. C. THWAITES, *CAPILLARY JOINING,* WILEY AND SONS, UNITED KINGDOM, 1982, P 52
26. J. SCULLY, *THE FUNDAMENTALS OF CORROSION,* 2ND ED., PERGAMON PRESS, INC., 1975
27. C. LEA, A SCIENTIFIC GUIDE TO *SURFACE-MOUNT TECHNOLOGY,* ELECTROCHEM. PUBL., LTD., 1988
28. P. VIANCO AND F. HOSKING, *ANALYSIS OF ULTRASONIC TINNING,* NEPCON WEST, FEB 1992, P 1718
29. J. SCHUSTER AND R. CHILKO, ULTRASONIC SOLDERING OF ALUMINUM HEAT EXCHANGERS, *WELD. J.,* OCT 1975, P 711
30. *PRUDENT PRACTICES FOR HANDLING HAZARDOUS CHEMICALS IN LABORATORIES,* NATIONAL ACADEMY PRESS, 1981, P 257-276

Soldering in Electronic Applications

Paul T. Vianco, Sandia National Laboratories

Introduction

SOLDERING represents the primary method of attaching electronic components, such as resistors, capacitors, or packaged integrated circuits, to either printed wiring boards (PWBs) or the ceramic substrates used for hybrid microcircuits.

Electronic miniaturization has fostered change in soldering technology as applied to electronic products. The increased functionality of electronic devices has placed a greater emphasis on solder joint integrity, not only in terms of increased reliability, but also with respect to limiting the cost of scrap product, inspection, or the repair of defective solder joints (especially on fine-pitch packages). The minimization of solder joint defects in electronic products depends on proper PWB design (for example, bonding pad geometries, conductor line layout, and laminate material), device packages (lead pitches and configurations or termination materials and finishes), and board assembly (flux and solder selection, process parameters and control, and cleaning).

This article discusses the categories that are most important to successful electronic soldering:

- SOLDERS AND FLUXES
- BASE MATERIALS AND FINISHES
- SOLDER JOINT DESIGN
- SOLDERABILITY TESTING

Additional information about the soldering processes discussed in this article can be found in the Section "Solid-State Welding, Brazing, and Soldering Processes" in this Volume.

Soldering in Electronic Applications

Paul T. Vianco, Sandia National Laboratories

Solders and Fluxes

Solders. The most commonly used solder alloys in electronics manufacturing are summarized in Table 1. Typical soldering temperatures in manufacturing operations are from 30 to 50 °C (55 to 90 °F) above the liquidus temperature to ensure adequate flow of the solder and proper heating of the substrate. The solder can be delivered to the joint area using numerous techniques. For example, the entire substrate can be placed in contact with the surface of a large molten solder bath, permitting the solder to either wet lands or fill holes. Solder wire can be directly heated at the joint by hand processes that use hot irons, gas jets, and other instruments. Preforms or solder paste can be placed at the joint area, and the substrate is either directly or indirectly heated, such as by an infrared (IR) furnace, causing the solder to melt and form the joint. Although solder pastes are often used in manual assembly or rework operations, their properties have been developed primarily for full-scale assembly processes.

TABLE 1 COMMONLY USED SOLDERS IN ELECTRONIC ASSEMBLIES

SOLDER ALLOY, WT%	SOLIDUS/LIQUIDUS TEMPERATURES	
	°C	°F
521N-48SN	118/118	244/244
50IN-50SN	118/125	244/257

58BI-42SN	138/138	281/281
43SN-43PB-14BI	143/163	289/325
97IN-3AG	143/143	289/289
70SN-18PB-12IN	162/162	324/324
63SN-37PB	183/183	361/361
60SN-40PB	183/190	361/374
62.5SN-36.1PB-1.4AG	170/179	354/354
50PB-501N	180/209	356/408
60PB-40IN	195/225	383/437
96.5SN-3.5AG	221/221	430/430
95SN-5PB	233/240	451/464
97.5PB-1.5AG-1.0SN	304/309	588/588
90PB-10SN	268/299	514/570
95PB-5SN	270/312	518/594

Tin-Base Tin-Lead Solders. Pure tin, which has a melting temperature of 232 °C (450 °F), readily wets the substrate materials typically found in electronics applications. In fact, it is the tin component of electronic solders that forms the metallurgical bond with those materials. However, pure tin is not used for PWB assembly, because of its higher melting point, increased cost, and reduced flow, when compared with tin-lead alloys. Tin plate is sometimes used as a protective finish on device leads and terminations, although a solder dip coating is the preferred finish.

Tin-base tin-lead solders represent the most widely used solders for electronic assembly: eutectic 63Sn-37Pb, near-eutectic 60Sn-40Pb, and the eutectic 62.5Sn-36.1Pb-l.4Ag alloys. The solidus-to-liquidus (mushy zone) temperature range of 60Sn-40Pb makes this alloy preferable to the eutectic composition when used with some leaded surface-mount devices in order to prevent solder wicking (Ref 1) or to facilitate the filling of large holes and vias on PWBs. The silver content of 62.5Sn-36.1Pb-1.4Ag reduces the dissolution of silver coatings, such as electroplated finishes on base metals, or the silver conductor used in thick-film terminations on either surface-mount packages or in hybrid microcircuit thick-film networks. The tin-lead with silver composition will improve creep-rupture strengths and will slightly reduce reactions when soldering to silver or copper. These solders typically contain 0.25% Sb to prevent the formation of the low-temperature allotrope of tin ("tin pest") at 13.2 °C (56 °F).

Lead-base tin-lead alloys are not used in general PWB assembly, because they have poor flow and their melting temperatures are incompatible with PWB materials. These alloys can be applied to surfaces such as electroplated finishes, evaporated thin films, or hot dip coatings. They have excellent fatigue and creep properties. The alloys 95Pb-5Sn and 90Pb-10Sn are used in the controlled collapse chip component (C^4) attachment of silicon chips to package frames (Ref 2). Their elevated melting temperatures prevent reflow of the C^4 joints during subsequent joining of the unit to the PWB with eutectic tin-lead solder (step soldering).

Tin-Silver Alloys. The eutectic composition, 96.5Sn-3.5Ag, is generally used, although alloy compositions are available with silver contents of up to 5%. These alloys are used at the high-temperature attachment stage of step soldering processes with eutectic tin-lead solders. The tin-silver solders exhibit excellent wetting, strength, and fatigue properties that are generally superior to those of the tin-lead eutectic solder (Ref 3). No significant silver migration has been observed in solder joints fabricated with 96Sn-4Ag alloy (Ref 4). The combination of tin and silver forms Sn-Ag intermetallic compounds that are stable and are insensitive to electromigration reactions.

Tin-Antimony Alloys. The 95Sn-5Sb alloy is used at the high-temperature attachment portion of step soldering processes with tin-lead eutectic alloy. This solder type has excellent creep strength and wettability, comparable to the tin-silver solders. Localized heating methods (soldering iron, laser, and others) are preferred for applications that utilize this alloy because the high heat input required to melt the solder may damage some PWB substrates or heat-sensitive devices.

Indium-Tin Alloys. The eutectic and near-eutectic indium-tin solders (52In-48Sn and 50In-50Sn, respectively) are low-melting-point alloys used in the solder attachment of heat-sensitive devices or in the soldering steps that follow the use of tin-lead alloys. Because most electronic fluxes are unable to activate at lower working temperatures, there can be some wetting difficulties when indium-tin alloys are used. Indium-base alloys can wet several ceramic materials, including quartz, and some glasses. However, they can be susceptible to corrosion damage, and have poor mechanical properties, especially in creep. Although the relatively high cost of indium metal has discouraged more-extensive use of these solders, the cost per device is not excessive, and indium-base alloys are finding increased usage where their properties are advantageous, such as in cryogenic applications.

Bismuth-containing alloys are also referred to as "fusible" solders. They are used on heat-sensitive devices or in step soldering processes. Upon solidification, bismuth-containing alloys exhibit either limited contraction or expansion, as in the case of solders with more than 47 wt% Bi (Ref 5). These alloys are prone to moderate oxidation in both the liquid and solid states, and they exhibit some difficulty in wetting because of limited flux activity at the lower working temperatures. These alloys have good mechanical properties and are used very successfully in a number of electronic applications.

Lead-indium alloys are used when soldering to precious-metal (gold, silver, platinum, palladium, or their alloys) substrates or coatings and when attaching components to hybrid thick-film networks. These solders have a lower rate of metal dissolution (and subsequent intermetallic compound formation), when compared with tin-containing alloys. Wettability is slightly poorer than that of the tin-lead solders. The lead-indium solders are also used in C^4 technology (Ref 6). Because of the relatively high cost of indium, these alloys are used on a limited basis. They are also susceptible to corrosion and interface voiding reactions (with high thermal gradients). However, they have excellent fatigue properties.

Solder Impurities. Solder alloys are available in reclaimed or virgin grades, and their purity is established by government, military, and industrial specifications. Excessive contaminant levels are detrimental to the quality of solder joints made by full-scale assembly equipment. Such processes, which use large baths as the solder source (wave, drag, or dip soldering), are particularly prone to impurity buildup from the dissolution of PWB and lead finishes (for example, copper, gold, or cadmium) or metals from the fixturing and soldering processes (aluminum or iron).

Table 2 provides contaminant limits for tin-lead solder material procurement (per federal specification QQ-S-571-E) and for tin-lead solder baths (per IPC-S-815A) that are for hot-dip coatings or PWB assembly. Amendment 6 to the QQ-571-E specification indicates a reduction in the maximum antimony content of powdered material to 0.120% in the upcoming QQ-571-F edition. The complete list of impurities is found in Amendment 6 of the QQ-571-E specification.

TABLE 2 RECOMMENDED IMPURITY LIMITS OF SOLDERS AT TIME OF PROCUREMENT AND IN OPERATIONAL TIN-LEAD BATHS

CONTAMINANT	**MAXIMUM CONCENTRATION IN VIRGIN SOLDER, %(A)**	MAXIMUM CONCENTRATION IN SOLDER BATH, %(B)
COPPER	0.08	**0.75/0.30**
GOLD	. . .	**0.50/0.20**
CADMIUM	. . .	**0.10/0.005**
ZINC	0.005	**0.008/0.005**
ALUMINUM	0.005	**0.008/0.006**
ANTIMONY	0.2-0.5	**0.50/0.50**
IRON	0.02	**0.02/0.02**
ARSENIC	0.03	**0.03/0.03**
BISMUTH	0.25	**0.25/0.25**
SILVER	. . .	**0.75/0.10**
NICKEL	**. . .**	**0.025/0.01**

(A) PER QQ-S-571-E.

(B) PER IPC-S-815A

To prevent contamination, the recommended materials for solder pots are stainless steel or cast iron. Although solder dross in solder baths is not an impurity, excessive levels can impede the proper formation of solder joints. Organic contaminants from fluxes, soldering oils, and PWB residues can be picked up by solder joints on the printed boards. These materials polymerize under prolonged exposure to the solder temperatures and become difficult to remove using conventional cleaning measures. Ionic species remaining from flux residues can cause corrosion on assembled solder joints or leakage currents between conductors.

Solder paste is used primarily in PWB assembly based on surface-mount technology. Processes such as vapor-phase (condensation) and infrared (IR) furnace reflow are well suited to solder paste use. Solder paste is a mixture of solder particles with a binder composed of flux, rheological components (such as thickeners), and a solvent material. Solder particle shape, size distribution, and concentration, along with binder properties, determine the flow properties of the paste.

As shown in Fig. 1, the paste is deposited by screen printing, stencil printing, or bulk dispensing techniques. Paste properties (whether applied using screen or stencil) affect the number and kinds of defects observed on the finished solder joint. Printed circuit boards (PCBs) with 1.25 mm (50 mil) pitch require 0.20 to 0.25 mm (0.008 to 0.010 in.) of paste thickness, whereas fine-pitch PCBs (630 μm, or 25 mils, or less) typically use 0.08 to 0.20 mm (0.003 to 0.008 in.) of paste.

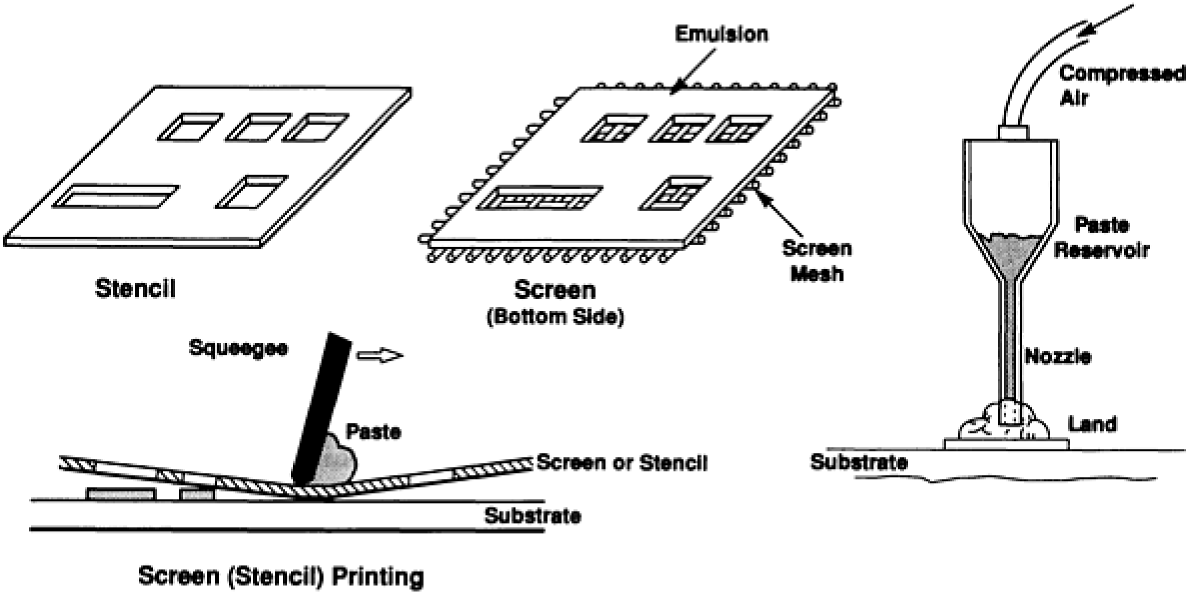

FIG. 1 STENCIL PRINTING, SCREEN PRINTING, AND BULK DISPENSING OF SOLDER PASTE

Printability can be affected by the distribution of the solder particles, metal content, slump, and viscosity of the paste. For example, smaller solder particles pass more easily through stencil or screen openings. However, the higher oxide content of the smaller particles (larger surface-to-volume ratio) creates a greater number of post-assembly solder balls, which are small spheres of solder that fail to coagulate with the larger solder mass in the joint (Fig. 2). Solder balls that are not removed by the cleaning process can cause short circuits during device electrical operation.

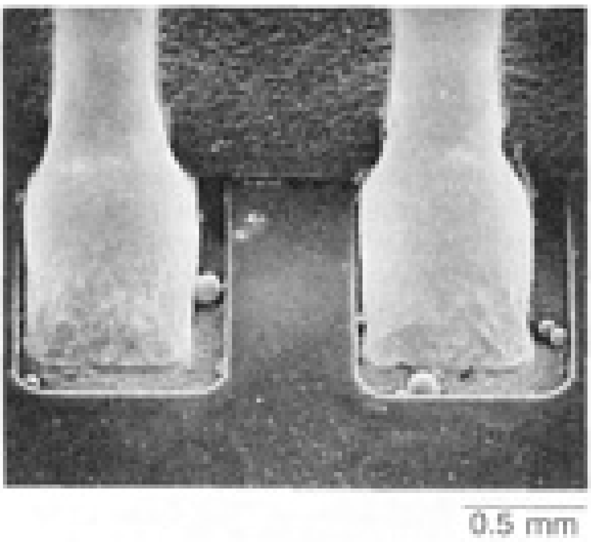

FIG. 2 SCANNING ELECTRON MICROGRAPH OF SOLDER BALLS ON A LEADLESS CERAMIC CHIP CARRIER SOLDER JOINT. COURTESY OF SANDIA NATIONAL LABORATORIES

Metal content refers to the mass percentage of the paste, that is, the solder component. Slump is a measure of the tendency of the paste to spread away from the deposited configuration prior to reflow. Excessive slump causes the deposits to contact one another (Fig. 3a), resulting in solder joint bridges (Fig. 3d), which can short circuit electrical signals (Ref 7).

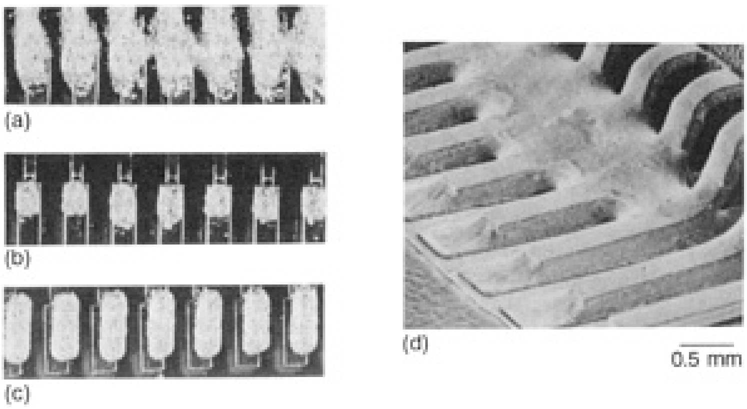

FIG. 3 DEFECTS OF SOLDER PASTE DEPOSITS. (A) APPEARANCE OF PASTE WITH POOR SLUMP (OPTICAL MICROSCOPY). (B) SKIPS IN PASTE DEPOSITS (OPTICAL MICROSCOPY). (C) PASTE DEPOSIT/LAND MISREGISTRATION (OPTICAL MICROSCOPY). (D) SOLDER BRIDGES (SEM). SOURCE: REF 1, 7

Viscosity determines the fluidity of the paste. When viscosity is too high, the paste will not pass through the screen or stencil, and will cause skips (Fig. 3b). When viscosity is too low, the paste will run out under the screen or stencil and will deteriorate the slump properties. Tackiness refers to the ability of the paste to hold devices onto the board prior to the reflow cycle. Paste with poor tackiness causes components to become displaced or to fall from the board during handling prior to reflow, resulting in part misalignment after assembly.

Test procedures that have been developed to quantify solder ball formation, slump, tackiness, and viscosity are identified in IPC-SP-819, "General Requirements and Test Methods for Electronic Grade Solder Paste."

Solder pastes deteriorate with storage time, which increases viscosity. The addition of thinners will reduce viscosity. Solder particles can further oxidize, leading to increased solder ball formation during reflow. Pastes that contain indium deteriorate more quickly than those based on other alloys. The shelf life of solder pastes can be extended by storage under refrigerated (not freezing) conditions, where the paste surface is sealed from the atmosphere. Manufacturer recommendations should be strictly followed.

Process factors that affect the quality of the printed deposit are:

- STENCIL OR SCREEN THICKNESS
- DAMAGE AND WEAR TO THE STENCIL OR SCREEN
- SQUEEGEE STIFFNESS AND PRESSURE
- SNAP-OFF DISTANCE
- DRYING CONDITIONS FOLLOWING DISPENSING
- STENCIL (SCREEN) SUBSTRATE REGISTRATION (FIG. 3C)
- PRESSURE AND ORIFICE SIZE ON BULK DISPENSING EQUIPMENT

Table 3 provides typical solder paste properties for the noted delivery techniques (Ref 8, 9).

TABLE 3 GENERAL SOLDER PASTE PARAMETERS FOR DISPENSING TECHNIQUES

DISPENSING TECHNIQUE	**PARTICLE DIAMETER**		**METAL CONTENT, %**	VISCOSITY, PA · S(A)
	µM	**µIN.**		
SCREEN PRINTING (-80 MESH SCREEN)	45-75	1800-3000	85	**250-550**
STENCIL PRINTING	45-75	1800-3000	85-90	**400-800**
BULK	**30-75**	**1200-1800**	**80-85**	**100-400**

(A) 1 PA · S = 1 KCPS

Fluxes. The role of fluxes in electronic soldering is to reduce thin tarnish layers on the substrates and solder, lower the surface tension of the solder to improve capillary flow and optimize fillet geometries, and protect the substrate surfaces from reoxidation during preheat prior to reflow. Electronic-grade fluxes for assembly processes can effectively remove only light oxide layers. Precleaning steps with chemical etchants should be used to remove heavy oxide films and scales prior to assembly. Surface oxide removal by the flux is inhibited by organic films that can be present on the surface. Therefore, surfaces must be degreased prior to application of the flux coating. In electronics applications, the least chemically active flux that satisfies production yields is preferred in order to minimize potential corrosion from unremoved residues.

Fluxes are applied to the PWB by spraying, immersing the board into a bath of flux, passing it through a flux wave, or coating it with a flux foam. The condition of the flux, which changes with use because of solvent evaporation and contamination buildup, is typically monitored by tracking its density. Density can also be used to indicate the viscosity of fluxes used in foaming applications, which in turn determines the ability of the flux to form a foam. Manufacturer recommendations should always be strictly followed because formulation variations are often proprietary.

Safety and health concerns should be considered in flux usage. Alcohol is a widely used solvent in fluxes, which makes the baths and vapors flammable. Fluxes, like corrosives, can cause acute and chronic irritation to the respiratory tract and skin. Because rosin-base fluxes break down at soldering temperatures to produce fumes of compounds from the aldehyde family, adequate ventilation of the work area is required.

Flux use must address three materials compatibility issues. First, the potential for damage to device packages (header material, glass-to-metal seals, and such), base materials, and plated surfaces prior to assembly will increase with more-active fluxes. Second, postassembly corrosion may arise from flux residues with high ionic contents, necessitating their removal by post-assembly cleaning measures. Third, the cleaning agents and processes used to remove flux residues must be compatible with the devices, substrates, and environmental regulations. Organic solvents, such as 1,1,1-trichloroethane and Freon, are very effective in removing rosin-base flux residues. However, these hazardous materials are being restricted from use due to potential environmental damage. Although aqueous and semiaqueous cleaning systems based on water, polar alcohols, and terpenes can be very effective, their compatibility with all electronic materials has not been fully characterized. It should be noted that ultrasonic cleaning processes may damage active and passive devices. Therefore, the compatibility of this technique with assemblies must be verified prior to its implementation. Designers and process engineers must address these concerns at the early stages of process development to prevent costly production errors.

The most critical aspect of PWB cleanliness is ionic residues from the fluxes. When combined with atmospheric moisture, these residues can generate "structural" corrosion between the dissimilar metals found in the solder joint. The residues can also create metal whiskers by the electromigration of metal ions between lands or conductor lines at different voltage potentials, causing leakage currents that disrupt electrical performance. Ionic flux residues on PWBs are measured by the extraction method, in which the residues are rinsed into a bath, the resistivity of which is converted into the equivalent amount of ionics removed from the board (expressed as μg NaCl/cm^2 of board surface).

The potential for leakage currents, which are a function of substrate material, moisture, and the ionic content of the flux residues, is assessed by the surface insulation resistance (SIR) test. A conductive comb pattern is constructed on the test board. A voltage difference is applied between adjacent legs and the resistance breakdown is monitored under different temperature-humidity-time conditions. Ionic residue and SIR tests are recommended in the process development stages of soldered assemblies. Several industrial, military, and federal specifications that quantify flux activity and PWB cleaning procedures and criteria for electronic applications have been established (Table 4).

TABLE 4 FLUX ACTIVITY TESTS AND WIRING BOARD CLEANLINESS GUIDELINES

SPECIFICATION	SUBJECT
IPC-TM-650, TEST METHODS MANUAL	**TESTS FOR CORROSIVITY OF FLUXES AND FLUX RESIDUES:** **COPPER MIRROR TEST** **SILVER CHROMATIC TEST** **FLUX SOLID DETERMINATION** **HALIDE CONTENT TEST** **CORROSION TEST** **FLUX CORROSIVITY ASSESSMENT** **MOISTURE TEST** **INSULATION RESISTANCE TEST** **DETERMINES LEAKAGE CURRENTS ON SUBSTRATE USING IPC-B-25 TEST VEHICLE**
IPC-S-815A, GENERAL REQUIREMENTS FOR SOLDERING ELECTRONIC INTERCONNECTIONS	**IONIC CONTAMINATION RESIDUE TEST (MEASURES IONIC RESIDUES ON THE ASSEMBLY; USED WITH IPC-TP-207 IF AUTOMATED TEST EQUIPMENT IS SPECIFIED)**
	ROSIN FLUX REMOVAL TEST (MEASURES THE EFFECTIVENESS OF CLEANING PROCESS TO REMOVE FLUX RESIDUES)
MIL-F-14256 E, FLUX, SOLDERING, LIQUID (ROSIN-BASE)	**TEST OF ROSIN-BASE FLUX CORROSIVITY AND SURFACE INSULATION RESISTANCE**
MIL-STD-2000 A, STANDARD REQUIREMENTS FOR SOLDERED ELECTRICAL AND ELECTRONIC ASSEMBLIES	**DETAILS IONIC CONTAMINATION (EXTRACTION) TEST PROCEDURES AND ACCEPTABILITY CRITERIA**

Rosin-base fluxes, which are based on the distilled products of pine saps, are the most widely used fluxes for electronics assemblies. Pure rosin (termed "water-white" rosin, after the test method for purity) is designated by the letter R. Because the R flux is a very weak acid, its residues are not corrosive in most applications. The activity of the rosin-base fluxes is strengthened by the addition of activators. Such fluxes are designated as mildly activated (RMA), fully activated (RA), and superactivated (RSA) (Ref 10).

RMA flux residues must be removed from most products designated for high-reliability applications. RA flux residues should be removed from all PWBs, except those used in low-end commercial electronic products. RSA fluxes are extremely corrosive, and their residues must by thoroughly removed from all products.

Activators are typically halide ions (Cl^-, F^-, and Br^-), which increase the activity of the rosin-base flux. However, "halide-free" activated fluxes are available and will lessen the corrosion potential of flux residues. Visible residues are tan or white in appearance (Ref 11). Black residues indicate that the flux has been exposed to excessive heat, and they are consequentially very difficult to remove. Ionic residue determination and SIR tests are strongly recommended for flux activities that exceed those of RMA materials.

No-Clean or Low-Solids Fluxes. Flux formulations for which the residues do not pose a corrosion concern after soldering--and thus do not need to be removed--are called "no-clean" fluxes. The fluxing activity of these materials can approach that of traditional RMA fluxes (Ref 12). A second approach to eliminating the need for postprocess cleaning is to use a "low-solids" flux, that is, a flux with limited solids content. (Solids form the residue after soldering.) However, low-solids fluxes have reduced oxide removal potential. Therefore, substrate tarnishes and contamination, flux density, and process conditions (inert atmospheres may be necessary) must be more tightly controlled to ensure consistent solderability.

Organic acid (OA) fluxes , which are also called water-soluble or intermediate fluxes, are water- or alcohol-base and have chemical activities greater than the rosin-base materials. OA fluxes improve the solderability of metals such as iron- or nickel-base alloys. Residues must be removed from the product using aqueous or semiaqueous cleaning methods. Flux activity can be provided by halide-containing compounds. The fluxes are also available in halide-free forms. Although they are more heat resistant than the rosin-base fluxes, the OA fluxes will char, producing hard-to-remove brown or black deposits. Ionic residue determination and SIR tests are strongly recommended for prototype development and lot sampling in assembly production to ensure adequate flux residue removal.

Inorganic acid (IA) fluxes either contain extremely corrosive acids (hydrochloric acid, phosphoric acid, and others) or are composed of metal chloride salts that form hydrochloric acid in the presence of water. Surfactants or wetting agents are also added. The residues are highly corrosive. These fluxes can be used for hot solder dipping of iron-base alloy or nickel leads, provided that adequate cleaning measures are taken to remove residues. These fluxes are not used in PWB assembly operations.

References cited in this section

1. R. PRASAD, *SURFACE MOUNT TECHNOLOGY, PRINCIPLE AND PRACTICE,* VAN NOSTRAND-RHEINHOLD, 1989, P 363
2. V. MARCOTTE, C^4 FLIP CHIP JOINING, *THE METAL SCIENCE OF JOINING,* M. CIESLAK *ET AL.*, ED., TMS, 1992, P 315
3. C. LEA, A SCIENTIFIC GUIDE TO SURFACE MOUNT TECHNOLOGY, ELECTROCHEM. PUB., LTD, 1988, P 169
4. R. GEHMAN, DENDRITIC GROWTH EVALUATION OF SOLDERED THICK FILMS, *INT. J. HYBRID MICROELEC.,* VOL 6, 1983, P 239
5. H. MANKO, *SOLDERS AND SOLDERING,* MCGRAW-HILL, 1979, P 112
6. L. GOLDMANN *ET AL.*, LEAD-INDIUM FOR CONTROLLED-COLLAPSE CHIP JOINING, *PROC. 27TH ELECTRON. COMP. CONF.,* 1977, P 25
7. R. KLEIN-WASSINK, *SOLDERING IN ELECTRONICS,* 2ND ED., ELECTROCHEM. PUB., LTD, 1989, P 548
8. R. KLEIN-WASSINK, *SOLDERING IN ELECTRONICS,* 2ND ED., ELECTROCHEM. PUB., LTD, 1989, P 557
9. J. LEE, SOLDER PASTE DISPENSING MATERIALS AND REQUIREMENTS, *ELECTRON. PACK. AND PROD.,* NOV 1990, P 22
10. H. MANKO, *SOLDERING HANDBOOK FOR PRINTED CIRCUITS AND SURFACE MOUNTING,* VAN NOSTRAND-RHEINHOLD, 1986, P 286

11. D. LOVERING, ROSIN ACIDS REACT TO FORM TAN RESIDUES, *ELECTRON. PACK. AND PROD.*, FEB 1985, P 232
12. D. KOCKA, NO-CLEAN FLUXES ARE A VIABLE ALTERNATIVE TO CFC CLEANING, *ELECTRON. PACK. AND PROD.*, JUNE 1990, P 95

Soldering in Electronic Applications

Paul T. Vianco, Sandia National Laboratories

Base Materials, Finishes, and Storage/Corrosion Issues

The substrate can be characterized by the nature of the base material and the surface finish. The base materials of device leads are typically copper or one of the iron-base, low-expansion alloys. Leadless surface-mount devices have ceramic oxide base materials. Copper foil is used as the base material of PWB conductor lines and lands (bonding pads).

The surface layer of a substrate comprises the so-called *solderable layer* to which the molten solder metallurgically reacts and, frequently, an additional coating or *protective layer* to prevent the formation of excessive oxidation and/or contamination by organic films on the solderable layer. This solderable surface can be either the base material surface itself or a coating deposited by electroplating, electroless plating, evaporation, pretinning (solder dip coating), sputtering, or chemical vapor deposition (CVD). The protective layer is typically an electroplated film that is entirely consumed by the solder during wetting. It is primarily the condition of the surface of the solderable layer (base material or a separate coating) on the component lead, termination, and PWB lands that accounts for joint solderability during assembly. Base material bulk properties, such as thermal conductivity or heat capacity, indirectly affect solderability, particularly the time-dependent wetting performance.

Coatings that serve as the solderable layer must be sufficiently thick to ensure that:

- THE COATING COMPLETELY COVERS THE BASE MATERIAL SURFACE
- THE LAYER IS NOT DISSOLVED BY THE LIQUID SOLDER
- THE LAYER IS NOT CONSUMED BY SOLID-STATE GROWTH OF INTERMETALLIC COMPOUNDS WITH THE SOLDER

On the other hand, excessive thickness can lead to residual stresses that cause delamination of the coating and the entrapment of organic plating compounds and gases that in turn cause the deterioration of solderability. Coatings used as solderable surfaces are typically the elemental metals, that is, nickel or copper. Multi-elemental thick-film layers are used as solderable surfaces on ceramic substrates (for example, conductor networks for hybrid microcircuits or terminations on discrete leadless ceramic devices or chip carriers).

The protective layer must be of sufficient thickness to protect the wettability of the surface of the solderable layer. Nickel that has been coated by a protective coating can be wet by the molten solder using the rosin-base fluxes. An unprotected nickel film requires much more active fluxes to promote solder wetting. Because the protective coating is absorbed into the solder, its thickness must be limited to prevent excessive contamination of the solder, thereby affecting its physical and mechanical properties.

Protective layers are often made from precious metals, the most popular of which is gold. It is imperative that gold coatings be removed by hot solder dipping the leads twice in flowing or nonflowing solder baths prior to assembly in order to prevent solder joint embrittlement (MIL-STD-1276D). However, *three* immersions in solder at 250 °C (480 °F) are recommended to ensure the complete removal of all thicknesses of gold from beam-leaded, surface-mount devices (Ref 13). Additional information is provided in the section "Precious Metal and Alloy Base Materials" of this article.

Other protective layers include electroplated tin (the production of which is termed tin plating) and electroplated tin-lead solder (for example, 60Sn-40Pb). The tin and tin-lead solder platings can be heated above their respective melting temperatures of 232 °C (450 °F) and 183 °C (361 °F) to remove pores or gaps. These layers are then referred to as *fused* tin or *fused* tin-lead coatings, respectively. A protective layer of tin or tin-lead solder can also be added by applying flux to the solderable surface and immersing it into a molten bath of tin or tin-lead solder. These finishes are referred to as hot-dipped tin or hot-dipped solder layers, respectively.

Organic coatings, such as benzotriazole and imidazole, are used as protective finishes on solderable surfaces. They are popular for protecting bare copper surfaces on PWBs during storage prior to soldering.

Coating materials and their preferred thicknesses for solderable and protective finishes are specified by MIL-STD-1276D and MIL-M-3851OH. Examples are described below according to the base materials used most frequently in electronics applications. In general, properly prepared and protected solderable surfaces (base material or deposited layer) can be readily wetted by the solders listed in Table 1 using standard electronic fluxes and assembly practices.

Copper Alloys. Copper is used in electronic applications as wire leads (individual or as frames) on through-hole components, electroless/electroplated layers in PWB holes, or thin foil in the construction of lines and lands on PWB surfaces. In some PWB fabrication processes, copper foil surfaces can be built up with electroless and electroplated copper, as well. The thickness of copper layers that coat PWB through-holes is approximately 0.025 to 0.076 mm (0.001 to 0.003 in.). Typical foil thicknesses for PWB surface patterns range from 0.018 to 0.071 mm (0.0007 to 0.0028 in.). The most common thicknesses are termed 0.5 oz copper (0.018 mm, or 0.0007 in.), 1 oz copper (0.036 mm, or 0.0014 in.), and 2 oz copper (0.071 mm, or 0.0028 in.).

Copper is readily wetted by tin-lead solder using rosin-base fluxes, provided that the surface is not heavily oxidized. Oxide removal is performed by immersion in dilute hydrochloric acid or sulfuric acid solutions or another type of mixture. Compositions of acid solutions used for oxide removal are given in the article "General Soldering" in this Volume. Protective finishes for copper and copper alloys include layers of electroplated silver of a thickness from 3.8 to 8.9 μm (150 to 350 μin.), electroplated tin (7.6 to 13 μm, or 300 to 500 μin.), and electroplated tin-lead (7.6 to 23 μm, or 300 to 900 μin.). The latter two layers can be fused; the preferred thicknesses range from 2.5 to 13 μm (100 to 500 μin.). The silver coating is usually covered by a tarnish layer caused by a reaction with sulfur in the air. The resulting sulfide film causes the solderability to deteriorate with time of exposure. Silver coatings need to be removed just like the gold coatings to prevent embrittlement of the solder joint by silver-tin intermetallic formation.

A hot-dipped tin-lead finish (5.1 μm, or 200 μin., minimum) is most often specified for copper-base metal. A solid-state reaction takes place between copper and tin (or the tin component of solders) to form a layer of intermetallic compounds, Cu_3Sn and Cu_6Sn_5, at the solder-copper interface. The thickness of the intermetallic layer increases as exposure temperature and time increase. The total intermetallic layer thickness ($Cu_3Sn + Cu_6Sn_5$) as a function of time and temperature is shown in Fig. 4(a) for electroplated tin, Fig. 4(b) for electroplated tin-lead coatings, and Fig. 4(c) for hot-dipped tin-lead coatings. Growth kinetics at room temperature are illustrated in Fig. 4(d) (Ref 14, 15).

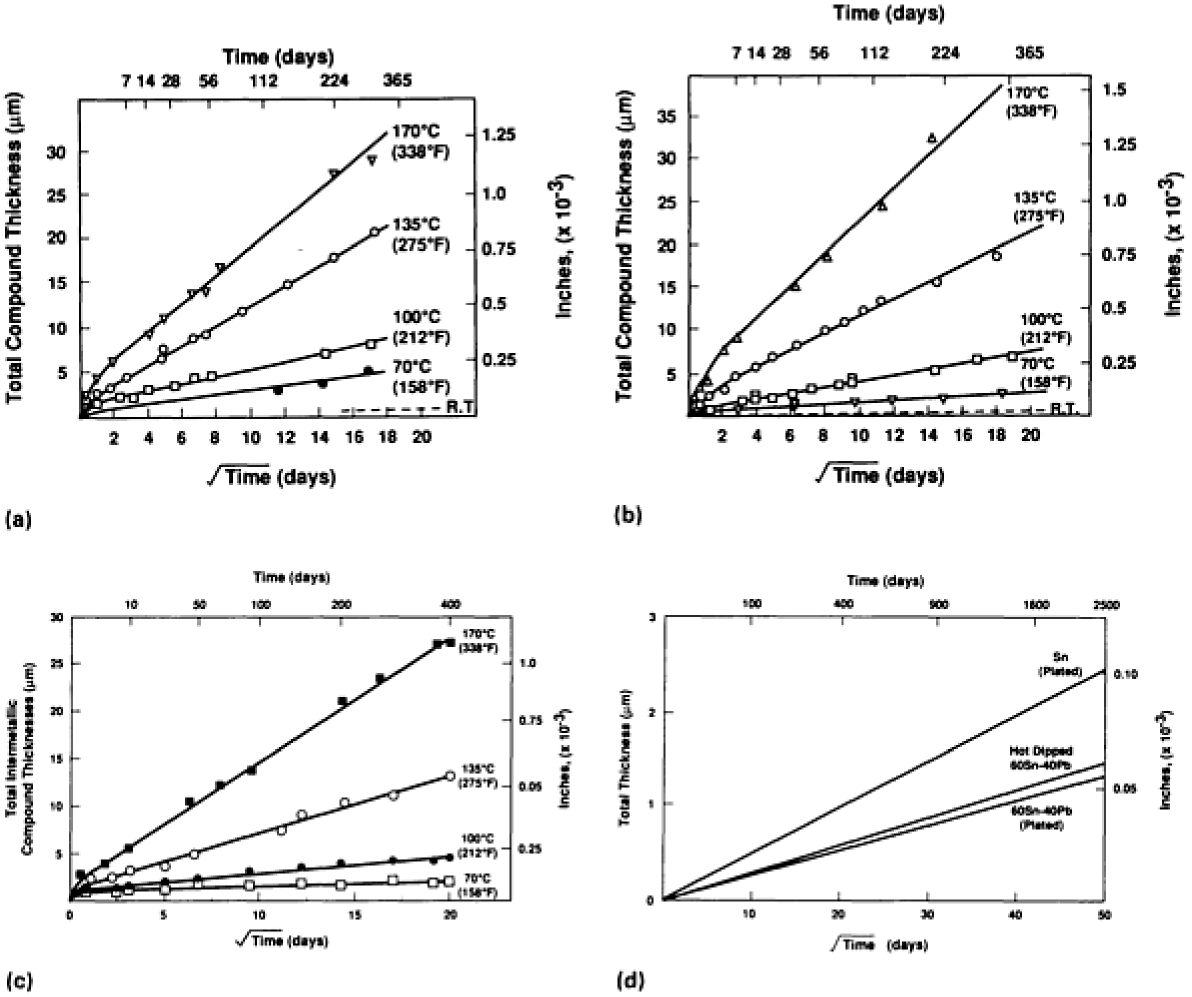

FIG. 4 COPPER-TIN INTERMETALLIC LAYER (CU$_6$SN$_5$ + CU$_3$SN) GROWTH KINETICS. (A) FOR ELECTROPLATED TIN COATING. (B) FOR ELECTROPLATED 60SN-40PB COATING. (C) FOR HOT-DIPPED 63SN-37PB COATING. (D) FOR TIN-LEAD COATINGS AT ROOM TEMPERATURE. SOURCE: INTERNATIONAL TIN RESEARCH INSTITUTE AND SANDIA NATIONAL LABORATORIES

Intermetallic layers have low ductility and, depending on their thickness, can affect the mechanical integrity of the solder joint (Fig. 5). In addition, the solid-state growth of the intermetallic film may consume very thin layers of tin or tin-lead alloy. The exposed intermetallic layer of Cu_6Sn_5 can readily oxidize and is difficult to wet with molten solder. Parts stored for over 1 or 2 years or those that will experience elevated temperatures (as in testing requirements) should be covered by a solderable layer of electroplated nickel (1.5 to 3.8 µm, or 60 to 150 µin.), followed by one of the above protective finishes. This coating structure will prevent the excessive formation of copper-tin intermetallic compounds at the solder-substrate interface, which may consume the tin or tin-lead protective finish and cause subsequent solderability to deteriorate.

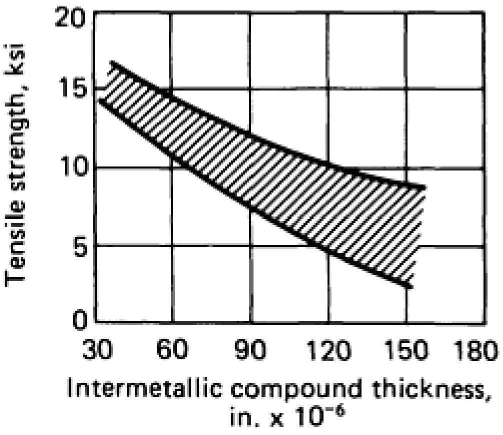

FIG. 5 EFFECT OF INTERMETALLIC COMPOUND THICKNESS ON ROOM-TEMPERATURE TENSILE STRENGTH OF SOLDER JOINTS (COPPER/COPPER BASE METALS)

Copper lines and lands on PWBs are protected with a hot-dipped solder coating that is generally applied by the hot-air leveling technique. The bare copper can also be protected with an organic inhibitor. Alloys of copper, including brasses, bronzes, copper-iron-zinc, copper-iron-tin, and copper-zinc-aluminum-cobalt, should be plated with a solderable layer (and protective coating) to overcome difficult-to-remove oxides or to act as a barrier against the diffusion of base metal constituents (for example, zinc) into the solder. Finishes include either a minimum 2.5 μm (100 μin.) of copper or a nickel layer with one of the protective coatings (described above).

Beryllium copper forms a tenacious oxide. Solder wetting with the use of activated rosin-base fluxes requires that the surface oxide be reduced by etching with strong acids. The soldering operation should immediately follow surface preparation steps to prevent reoxidation. Beryllium copper with very thin oxides can be wetted by solder if inorganic acid fluxes are used. However, a thorough cleaning of residues is required. The inorganic acid fluxes are used only to apply a protective layer of hot-dipped tin-lead solder. They are not recommended for PWB assembly.

Nickel and nickel-base alloys are used as the base metals on package leads. However, the most common electronics application is electroplated pure nickel used as the solderable coating on nickel-and iron-base lead materials that cannot be subjected to the cleaning measures and aggressive fluxes required to promote solderability. Nickel and its alloys are difficult to wet, because of the formation of a thin, tenacious oxide layer. Substrates must be chemically etched with aggressive solutions. The soldering operation or the application of a protective layer must immediately follow the etching treatment in order to minimize reoxidation of the surface. Freshly deoxidized leads made with nickel or nickel-containing alloys, or those that are electroplated with a solderable layer of nickel (1.3 to 3.8 μm, or 50 to 150 μin.), require one of the following protective layers prior to assembly on the PWB:

- ELECTROPLATED (MATTE) TIN (7.6 TO 13 μM, OR 300 TO 500 μIN.)
- ELECTROPLATED TIN-LEAD COATING (7.6 TO 23 μM, OR 300 TO 900 μIN.)
- HOT-DIPPED SOLDER COATING (5.1 μM, OR 200 μIN.)

The electroplated layers can be fused. The hot-dipped solder coating step may require the use of more-aggressive fluxes (organic acid or inorganic acid fluxes) to achieve a satisfactory finish. These fluxes must be compatible with the device package construction and their residues must be thoroughly removed after coating. It is very important that hot-dipped solder (or tin) layers on device leads and terminations contain neither "lumps" nor icicles, because these defects can interfere with the automated placement of components on the PWB, particularly with the use of fine-pitch devices. However, a properly applied hot solder dipped lead finish is superior to most tin or tin-lead plated surfaces.

The nickel solderable coating is protected by a finish of electroplated gold, particularly if the base metal is subject to temperatures that exceed the melting points of tin or tin-lead finishes prior to assembly. The MIL-G-45204C thickness recommendation for solderability is from 1.3 to 2.5 μm (50 to 100 μin.). However, general specification of gold thickness (MIL-STD-1276 D) can be as high as 2.5 to 7.6 μm (100 to 300 μin.) so that complete removal for soldering should be verified if the thinner layers do not offer adequate protection. The previously described procedure regarding the removal of gold coatings pertains. Details are provided in the section "Precious Metal and Alloy Base Materials" in this article.

A less-expensive silver coating (3.9 to 8.9 μm, or 150 to 350 μin.) can also be used. However, silver tarnishes upon atmospheric exposure, causing a deterioration to solderability. As already mentioned, silver coatings need to be removed, just like gold coatings are.

Electroplated copper (3.8 to 7.6 μm, or 150 to 300 μin.), followed by one of the tin or tin-lead protective finishes, can also provide excellent solderability to nickel and nickel-alloy base metals. Although nickel forms an intermetallic compound layer with tin (primarily, Ni_3Sn_4), its growth rate is much slower than that of copper and tin. Therefore, it does not significantly affect the mechanical properties of solder joints in electronic assemblies. An ultrasonically activated solder pot can be used to apply a hot-dipped solder (or tin) coating to the lead without the use of aggressive fluxes. The ultrasonic energy disrupts the oxide layer, thereby allowing the solder to wet without the use of aggressive fluxes. Ultrasonics should not be used on devices that can be damaged by the ultrasonic energy.

Aluminum and its alloys are used as connector housings that attach cables to PWBs or other cable assemblies. A "strap" or lead is soldered to the housing to provide an electrical ground.

Aluminum alloys are difficult to wet, because of a tenacious surface oxide. Although aggressive fluxes, such as the inorganic acids, are required to promote solderability, they leave behind extremely corrosive residues that must be thoroughly removed to prevent corrosion in service.

Another issue when solder joining aluminum is galvanic corrosion between aluminum and a dissimilar substrate (for example, copper or nickel) or between aluminum and solder constituents (for example, tin). Solder alloys, such as 80Sn-20Zn (T_1 = 270 °C, or 518 °F; T_s = 198 °C, or 388 °F) and 95Zn-5Al (382 °C, or 720 °F), were developed for compatibility with aluminum. However, their high melting points require localized heating techniques and prohibit joining to organic materials or hybrid thick-film networks. The high thermal conductivity of aluminum alloys requires large heat input into the substrate to promote wetting.

Although solderable and protective finishes were not cited by the specifications noted above, adequate solderability can be obtained with coatings of electroplated nickel or copper to thicknesses of 1.3 to 3.8 μm (50 to 150 μin.) and 7.6 to 25 μm (300 to 1000 μin.), respectively. A zincate coating is applied to the aluminum prior to the nickel or copper finishes to promote adhesion. Once the soldering operation has been performed, it is recommended that the exposed nickel or copper layer be coated with an anodic finish (for example, cadmium or zinc) to reduce the potential for a corrosion couple with aluminum. Ultrasonic activation of the solder pot provides a means of depositing a hot solder or tin-dipped finish to aluminum substrates without a flux.

Iron-base alloys constitute the base materials for leads on a number of through-hole and surface-mount device packages. These alloys include Kovar, 29Ni-17Co-0.2Mn-balance Fe; alloy 52, 0.5Mn-0.25Si-50.5ONi-balance Fe; and alloy 42, 0.5Mn-O.25Si-5.5Cr-42Ni-balance Fe. Solder wetting of the iron-base alloys requires the removal of a thick, tenacious surface oxide by such procedures as the use of etchants and chemical brighteners or aggressive electropolishing treatments (Ref 16).

The hot-dipped solder coating may require the use of more-aggressive fluxes (organic acid or inorganic acid fluxes) to achieve a satisfactory finish. These fluxes must be compatible with the device package materials and their residues must be thoroughly removed after coating. Surfaces must be protected from reoxidation after surface preparation, typically by a hot-dipped tin or solder coating.

However, the most common technique to achieve solder wetting of these materials is the use of combined solderable and protective layers. The most frequent approach is to electroplate device leads with a solderable layer of nickel (1.3 to 3.8 μm, or 50 to 150 μin.), followed by a protective layer of electroplated (matte) tin (7.6 to 13 μm, or 300 to 500 μin.), electroplated tin-lead coating (7.6 to 23 μm, or 300 to 900 μin.), or hot-dipped solder coating (5.1 μm, or 200 μin., minimum). The electroplated layers can be fused. Silver (3.9 to 8.9 μm, or 150 to 350 μin.) can also be used as a protective finish. As previously mentioned, hot-dipped solder (or tin) finishes on package leads should not contain "lumps" or icicles, because they can interfere with the automated placement of components on the PWB.

The nickel solderable coating can be protected by a finish of electroplated gold, particularly if the base metal is subject to temperatures that exceed the melting points of tin or tin-lead prior to assembly. The recommended thickness for soldering operations ranges from 1.3 to 2.5 μm (50 to 100 μin.), per MIL-G-45204C. Gold coatings must be removed in the manner previously described. Details are provided in the section "Precious Metal and Alloy Base Materials" in this article.

Alternative solderable coatings include electroless nickel (1.3 to 3.0 μm, or 50 to 120 μin.), palladium-nickel (1.3 to 2.5 μm, or 50 to 100 μin.), or copper (3.8 to 7.6 μm, or 150 to 300 μin.). Electroless nickel is not recommended for flexible or semirigid leads or substrates exposed to high-temperature testing procedures prior to PWB assembly. Phosphorus from the plating bath may become entrapped in the layer, resulting in film embrittlement or poor solderability because of its diffusion to the nickel surface at elevated temperatures. A palladium-nickel coating does not require a protective finish because the palladium component imparts oxidation resistance to the alloy coating. The electroless nickel and electroplated copper coatings require one of the protective finishes noted above.

Precious Metal and Alloy Base Materials. Excellent solderability of precious metals (gold, silver, palladium, platinum, and others) is achieved with tin-and indium-base solders, because of a strong metallurgical reaction at the solder/base metal interface. This metallurgical reaction results in the formation of intermetallic compounds at the interface. Their growth can take place when the solder is in either the liquid or solid state. In either case, intermetallic compound growth can quickly consume wires, leads, or the entire thick-film layer.

The dissolution rates of several precious metals in molten 60Sn-40Pb solder are shown in Fig. 6 (Ref 17). The excessive concentration of intermetallic compounds severely embrittles bulk tin-lead solder. In addition, thick intermetallic layers at the solder/base metal interface can drastically decrease solder joint ductility. Gold contents of 6 wt% decrease the ductility (reduction-in-area) by 50% (Ref 18); these data indicate that gold embrittlement is generally avoided in tin-lead solders by maintaining gold contamination to the joint at a level less than 4 wt%. (Some segments of the electronics industry have specified maximum gold contents of 1 wt%.)

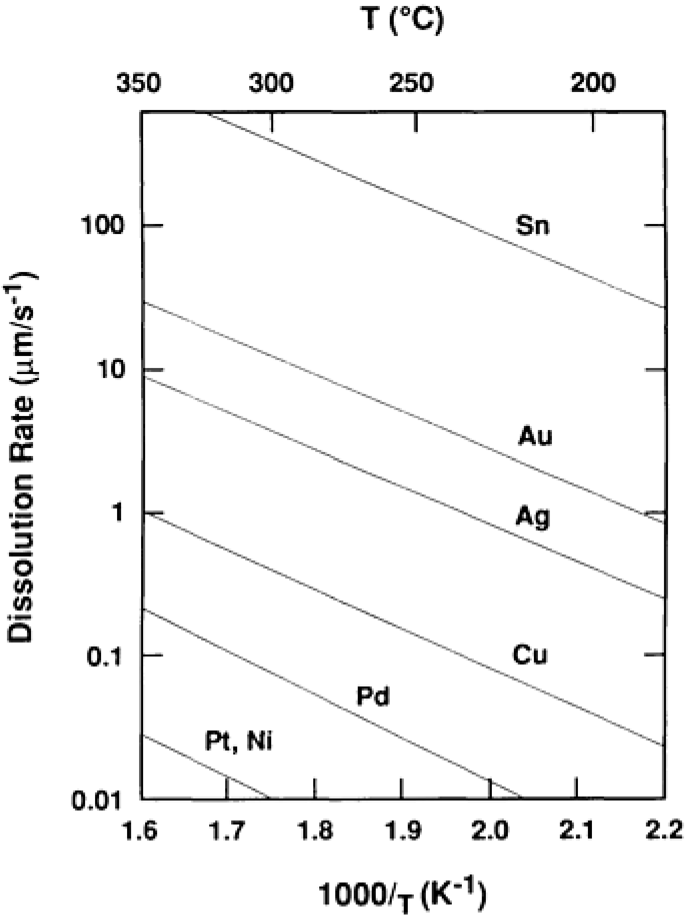

FIG. 6 DISSOLUTION RATES OF VARIOUS METALS IN MOLTEN 60SN-40PB SOLDER. SOURCE: REF 3

Silver wires or silver coatings on other metals have been used in place of gold to reduce cost. Silver has a lower dissolution rate in tin-lead solders than gold. The rate can be further diminished by adding from 1 to 2 wt% silver to the solder. However, silver rapidly tarnishes from airborne sulfur pollution, causing the surface to tarnish rapidly and lose solderability. The dissolution of precious metals is significantly less when lead-indium solders are used.

Wetting of precious metal base materials is assisted by the limited oxide layer that forms on them, which sometimes allows wetting by solders without the use of a flux. Wires or ribbons are used as jumper leads on PWBs by virtue of their low resistance. Pastes that contain precious metals or their alloys as the conductive component are used as thick-film networks for hybrid microcircuitry.

Ceramic materials used for electronic packages and substrates include alumina (Al_2O_3), beryllia (BeO), and silica (SiO_2), as well as several of the engineered ceramics, such as silicon nitride (Si_3N_4). Solders do not readily wet ceramic-base materials, because of the low surface energy of the substrate and the absence of a solder-ceramic metallurgical reaction.

Solderable surfaces are provided by thick-film metallization inks. These inks contain a powder of electrically conductive metals or alloys (for example, copper, nickel, gold-palladium-platinum, platinum-silver, platinum-gold, and others), an oxide powder and glass binder that bond the conductive component to the substrate by a high-temperature firing process, and an organic carrier comprising resin and solvents to give the ink "body" for printability on the substrate. The thick films have excellent solderability, because of their noble metal content. However, the 10 to 20 μm (100 to 800 μin.) thick film can be consumed by intermetallic formation with the tin component of the liquid solder, as well as through solid-state reaction at the solder-film interface of the solidified joint. The use of lead-indium solders greatly reduces the growth of intermetallic compounds. Copper and nickel thick films are not as quickly consumed by the tin-base solders and are less costly. However, these coatings must be protected from excessive oxidation during firing, solder processing, and contaminated atmospheres during storage.

Other thick-film systems are based on moly-manganese and refractory metals (Ref 19). The moly-manganese process uses an ink of molybdenum and manganese that is printed onto the selected location(s) and fired to bond the film to the ceramic. Next, a layer of copper or nickel is electroplated onto the moly-manganese film to form the solderable coating. Then, a protective layer of electroplated gold is added. Alternatively, some refractory metal thick-film systems use molybdenum or tungsten as the ceramic binder. Gold is electroplated onto the refractory metal binder. The resulting metallized layer is heated to partially consume the gold as gold-molybdenum or gold-tungsten alloy at the interface. The alloys form the solderable surface while the remaining gold serves as the protective finish. Electroplated nickel can be used to create the solderable surface in place of gold-molybdenum or gold-tungsten.

Storage. The base material or surface finish solderability is strongly affected by storage conditions (time, temperature, humidity, gases, and packing materials) and the integrity of the protective coating. Recommended electroplated and hot-dipped solder finishes that pass specified solderability tests are considered to preserve solderability for 1 to 2 years under typical industrial environments (Ref 20, 21). Longer periods may require special storage conditions, such as inert atmospheres in containers, controlled temperature and humidity levels, or additional solderability testing prior to assembly. Plasticizers and silicon-base mold releases in plastic containers, as well as sulfur compounds in paper products and factory atmospheres, can degrade solderability. Inventory control practices, including "first in, first out" and "just-in-time," can prevent solderability deterioration caused by excessive storage periods.

Corrosion. Solder alloys have necessarily dissimilar compositions, when compared with the base materials that they join together. Therefore, solder joints have a potential for corrosion activity. Water vapor from the atmosphere can provide the electrolyte medium, particularly given that most PWB laminates (and ceramic materials) are hygroscopic to varying degrees.

Sources of ionic species to form the electrolyte medium are flux residues, processing chemicals that remain in the PWB (laminate), or external contamination by handling or service environments. The electromotive series provides a preliminary estimate of corrosion "potential" for the metal systems in the joint. However, the series cannot describe the kinetics (rate) of corrosion.

Assembly processes for high-reliability electronics or those systems destined for harsh service conditions (for example, saltwater spray or high-humidity locations) must be thoroughly assessed for corrosion potential. Issues include: the compatibility of the solder and base material, the activity of the flux and the need for its residue removal after assembly, and the use of either conformal (organic) coatings to exclude atmospheric moisture or coatings with metal finishes (zinc or cadmium electroplating) to act as sacrificial anodes.

References cited in this section

3. C. LEA, A SCIENTIFIC GUIDE TO SURFACE MOUNT TECHNOLOGY, ELECTROCHEM. PUB., LTD, 1988, P 169
13. P. VIANCO AND J. DAL PORTO, "EMBRITTLEMENT OF SURFACE MOUNT TRANSISTOR SOLDER JOINTS INVOLVING PRETINNED LEADS," INTERNATIONAL BRAZING AND SOLDERING CONFERENCE (DETROIT, MI), AWS, 1991
14. D. UNSWORTH AND C. MACKAY, A PRELIMINARY REPORT ON GROWTH OF COMPOUND LAYERS ON VARIOUS METAL BASES PLATED WITH TIN AND ITS ALLOYS, *TRANS. INST. MET.,* VOL 51, 1973, P 85
15. P. KAY AND C. MACKAY, THE GROWTH OF INTERMETALLIC COMPOUNDS ON COMMON BASIS MATERIALS COATED WITH TIN AND TIN-LEAD ALLOYS, *TRANS. INST. MET.,* VOL 54, 1976, P 68
16. P. VIANCO *ET AL.,* SOLDERABILITY TESTING OF KOVAR WITH 60SN-40PB SOLDER AND ORGANIC FLUXES, *WELD. J.,* JUNE 1990, P 230-S
17. C. LEA, *A SCIENTIFIC GUIDE TO SURFACE MOUNT TECHNOLOGY,* ELECTROCHEM. PUB., LTD, 1988, P 167
18. M. BESTER, METALLURGICAL ASPECTS OF SOLDERING GOLD OR GOLD PLATING, *PROC. INTERNEPCON,* 1968, P 221
19. P. HOLMES AND R. LOASBY, *HANDBOOK OF THICK FILM TECHNOLOGY,* ELECTROCHEM. PUB., LTD., 1976, P 192
20. R. WILD, COMPONENT LEAD SOLDERABILITY VS. ARTIFICIAL AGEING II, *PROC. 1 L TH ELECT. MFG. SEMINAR* (CHINA LAKE, CA), NAVAL WEAPONS CENTER, 1987, P 289
21. R. EDINGTON AND L. CONRAD-LOWANE, AGING ENVIRONMENTS AND THEIR EFFECTS ON SOLDERABILITY, *PROC. 12TH ELECTRON. MFG. SEMINAR* (CHINA LAKE, CA), NAVAL WEAPONS CENTER, 1988, P 31

Soldering in Electronic Applications

Paul T. Vianco, Sandia National Laboratories

Solder Joint Design

Solder joint performance depends on the package size of a device, the materials and layout of the PWB (or hybrid substrate), and the manufacturing processes used to assemble the PWB. Hardware package configurations and PWB layout guidelines have been established by several industrial standards organizations, such as the Electronics Industry Association (EIA), Institute for Interconnection and Packaging Electronic Circuits (IPC), American National Standards Institute (ANSI), and the Joint Electronic Devices Engineering Council (JEDEC). General solder joint configuration and performance recommendations are discussed below for through-hole technology, surface-mount technology, mixed technology (through-hole plus surface mount), and connector technology. Table 5 lists the numerous guidelines that provide details on these topics. (A list of industrial specifications can be obtained from IPC, 7380 N. Lincoln Ave., Lincolnwood, IL, 60646-1705; federal and military specifications are available from the Standardization Documents Order Desk, Building 4D, 700 Robbins Ave., Philadelphia, PA, 19111-5094.)

TABLE 5 GUIDELINES AND SPECIFICATIONS FOR SOLDER JOINT DESIGN

IPC-D-300G, PRINTED BOARD DIMENSIONS AND TOLERANCES
IPC-SM-782, SURFACE MOUNT LAND PATTERNS (CONFIGURATIONS AND DESIGN RULES)
IPC-S-815A, GENERAL REQUIREMENTS FOR SOLDERING ELECTRONIC INTERCONNECTIONS
IPC-SM-780, COMPONENT PACKAGING AND INTERCONNECTING WITH EMPHASIS ON SURFACE MOUNTING
IPC-D-275, DESIGN STANDARD FOR RIGID PRINTED BOARDS AND RIGID PRINTED BOARD ASSEMBLIES
IPC-D-322, GUIDELINES FOR SELECTING PRINTED WIRING BOARD SIZES USING STANDARD PANELS
IPC-MC-324, PERFORMANCE SPECIFICATION FOR METAL CORE BOARDS
IPC-D-330, DESIGN GUIDE
IPC-PD-325, ELECTRONIC PACKAGING HANDBOOK
IPC-CM-770, PRINTED BOARD COMPONENT MOUNTING
IPC-D-279, DESIGN GUIDELINES FOR RELIABLE SURFACE MOUNT TECHNOLOGY
IPC-SM-785, GUIDELINES FOR ACCELERATED RELIABILITY TESTING OF SURFACE MOUNT SOLDER ATTACHMENTS
IPC-S-816, SMT PROCESS GUIDELINE AND CHECKLIST
J-STD-001, REQUIREMENTS FOR SOLDERED ELECTRICAL AND ELECTRONIC ASSEMBLIES
J-STD-002, SOLDERABILITY TESTS FOR COMPONENT LEADS, TERMINATIONS, LUGS, TERMINALS AND WIRES
J-STD-003, SOLDERABILITY TESTS FOR PRINTED BOARDS
MIL-C-55302, CONNECTORS, PRINTED CIRCUIT SUBASSEMBLY, AND ACCESSORIES

An assessment of solder joint performance requires knowledge of several physical and mechanical properties of the solder alloy, package, and substrate materials. Strength calculations based on bulk solder properties typically provide a conservative estimate of solder joint performance (Ref 22); more accurate predictions are made from solder joint properties. The properties of those materials commonly used in electronics packaging and substrates are summarized in Table 6.

TABLE 6 SELECTED PROPERTIES FOR ELECTRONIC ASSEMBLY MATERIALS

MATERIAL	**GLASS TRANSITION TEMPERATURE**		**PLANAR COEFFICIENT OF THERMAL EXPANSION$^{(A)}$, 10^{-6}/K**	**THERMAL CONDUCTIVITY**		PLANAR TENSILE MODULUS	
	°C	**°F**		**W/M · K**	**10^{-4} BTU/FT · S · °F**	**GPA**	10^{6} PSI
PRINTED BOARD SUBSTRATES							
EPOXY FIBERGLASS	125	257	13-18	0.16	0.26	17	**2.5**
POLYIMIDE FIBERGLASS	250	482	12-16	0.35	0.56	19	**2.8**
EPOXY-ARAMID FIBER	125	257	6-8	0.12	0.19	30	**4.4**
POLYIMIDE-ARAMID FIBER	250	482	3-7	0.15	0.24	28	**4.0**
POLYIMIDE QUARTZ	250	482	6-8	0.30	0.48	. . .	**. . .**
FIBERGLASS-TEFLON	75	167	20	0.26	0.42	1.4	**0.2**
CERAMICS							
ALUMINA, 86%	. . .	. . .	6.0	20	32	300	**44**
BERYLLIA, 99.5%	. . .	. . .	6.4	200	320	350	**51**
SILICON	. . .	. . .	2.3	150	240	115	**17**
METALS							
COPPER	. . .	. . .	17.0	370	590	130	**19**

NICKEL	. . .	. . .	13.0	89	140	220	**32**
ALUMINUM	. . .	. . .	24.0	230	370	70	**10**
BRASS, 33 WT% ZN	. . .	. . .	20.0	120	190	112	**16**
BRONZE, 5 WT% SN	. . .	. . .	18.0	75	120	121	**18**
LOW-CARBON STEEL	. . .	. . .	12.0	58	93	210	**30**
STAINLESS STEEL, 18-8	. . .	. . .	15.0	16	26	210	**30**
KOVAR	. . .	. . .	5.2	17	27	210	**30**
60SN-40PB SOLDER	**. . .**	**. . .**	**25.0**	**51**	**82**	**32**	**4.6**

Source: Ref 1

(A) THERMAL EXPANSION OF CIRCUIT BOARD SUBSTRATES IN THE *Z* (THICKNESS) DIMENSION CAN BE AN ORDER OF MAGNITUDE GREATER THAN IN THE *XY* (PLANAR) DIMENSIONS.

Organic materials used in PWB laminates soften significantly at temperatures that exceed their glass transition point. Time periods at temperatures greater than the transition value should be minimized to prevent permanent warpage or twist to the PWB, which can cause debonded lands and conductor lines or fractured joints and packages due to misfit in cabinet frames.

Coefficients of thermal expansion (CTE) are required to assess the thermal fatigue resistance of solder joints, a principal source of failure for leadless surface-mount components. The CTEs of PWB laminates are specified for the in-plane dimensions. It is important to note that the through-thickness CTEs of PWB laminates can be as much as an order of magnitude larger than the in-plane values. Because PWB CTEs are very sensitive to the resin content, resin type, and fiber weave of the laminate, precise values should be obtained by measuring them directly, rather than relying on generic information from the manufacturer.

Thermal conductivity is important in solder joint design for high-power applications.

Excessive moisture retention by the PWB (organic laminate or ceramic) can lead to voids in solder joints that are due to water vapor formed at soldering temperatures or it can cause movement of surface-mount parts during reflow that results in misaligned packages on the assembly.

Through-hole technology typically refers to the use of leaded device packages. The leads are inserted into holes in the PWB and are soldered in place. Several package/lead configurations are shown in Fig. 7. Device leads are typically copper or one of the iron-base alloys (typically coated with a solderable layer). These lead-base materials have a protective finish of tin-lead alloy (plated, plated and fused, or hot dipped) to facilitate solder wetting during assembly on the PWB. Although the coatings can accommodate lead-forming operations, such practices should not be performed on leads with an electroless nickel solderable layer.

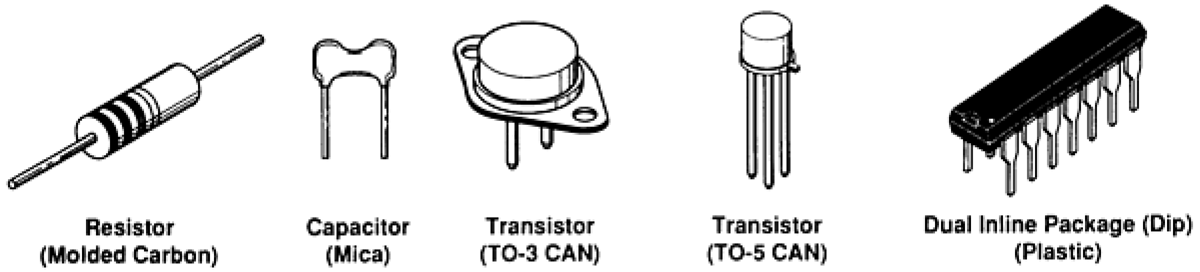

FIG. 7 PACKAGE AND LEAD CONFIGURATIONS FOR THROUGH-HOLE DEVICES

Substrates for through-hole technology include organic laminates, metal-clad substrates, and ceramics for hybrid microcircuits. Ceramic substrates are discussed below in the section on surface-mount technology. Through-hole PWBs are categorized as:

- SINGLE-SIDED, WITH CIRCUITRY ON ONE SURFACE ONLY
- DOUBLE-SIDED, WITH CIRCUITRY PLACED ON BOTH SURFACES
- MULTILAYER, WITH CIRCUIT LAYERS ON BOTH EXTERNAL SURFACES AS WELL AS CONDUCTIVE PATHS WITHIN THE LAMINATE

The solder joint configuration for each case is shown in Fig. 8. Lands (and conductor lines) are constructed of copper foil bonded to the laminate; the foil thicknesses are typically 0.017 mm (0.0007 in.), 0.035 mm (0.0014 in.), or 0.071 mm (0.0028 in.) and are designated as 0.5 oz, 1 oz, and 2 oz, respectively. The surfaces of lands (and lines) can be built up with a combination of electroless and electroplated copper to a thickness that ranges from 0.013 to 0.076 mm (0.0005 to 0.0030 in.) to satisfy final thickness requirements and enhance solderability.

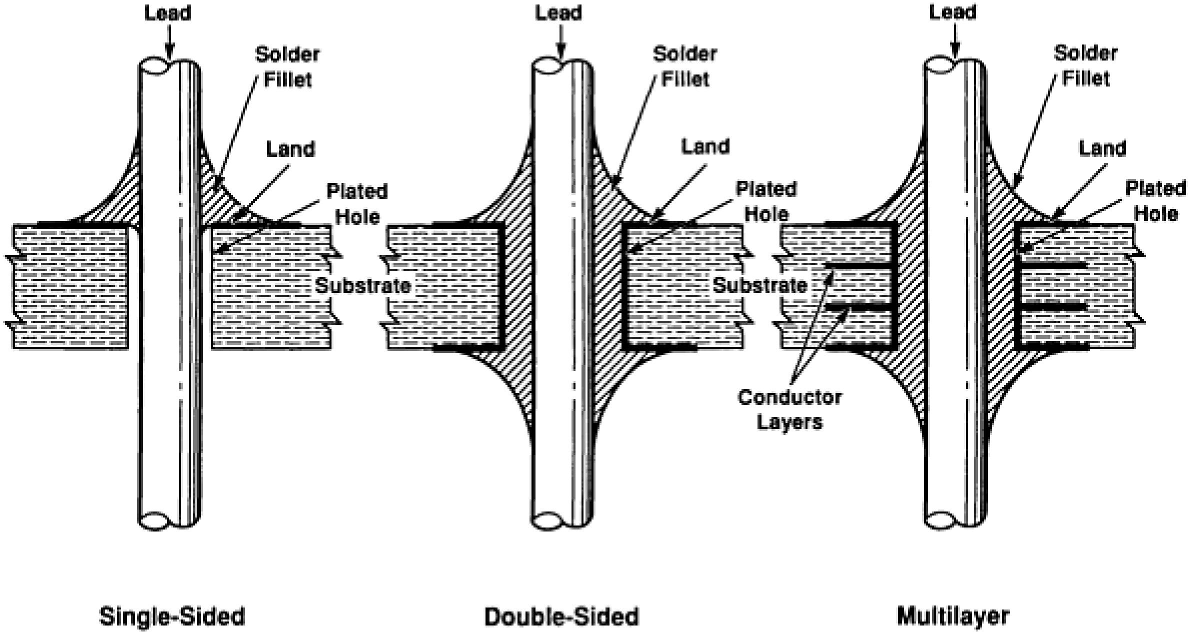

Single-Sided

Double-Sided

Multilayer

FIG. 8 SINGLE-SIDED, DOUBLE-SIDED, AND MULTILAYER SOLDER JOINT CONFIGURATIONS FOR THROUGH-HOLE TECHNOLOGY

Although the role of holes on single-sided circuitry is to mechanically secure the electronic devices, holes on double-sided and multilayer boards have an expanded role of providing electrical signal conduction between the two surfaces and internal layers. Holes used for interlayer signal transmission, but without a device lead, are termed "vias." The hole walls are deposited with copper to establish electrical conductivity and promote solder flow. The walls are coated with from 0.0003 to 0.0025 mm (0.0001 to 0.0001 in.) of electroless copper, followed by 0.025 to 0.076 mm (0.001 to 0.003 in.) of electroplated copper (Ref 23).

General guidelines for hole design are provided in Fig. 9. The hole size specified for a lead or wire should permit a gap of 0.15 to 0.20 mm (0.006 to 0.008 in.) to optimize the strength and capillary flow of the solder. Smaller gaps cause interference between the lead and hole, which can either damage the plating layer during insertion or generate voids caused by incomplete venting of flux gases produced during soldering. The recommended lead protrusion from the board underside is from 0.8 to 2.0 mm (0.03 to 0.07 in.). Values are kept small in order to minimize solder drainage from the fillet. The hole bonding land, or pad, should be round, and the outer diameter should be approximately three times the wire diameter. The annular width of the land should allow a lead height-to-pad width ratio of 1:1 for adequate fillet formation.

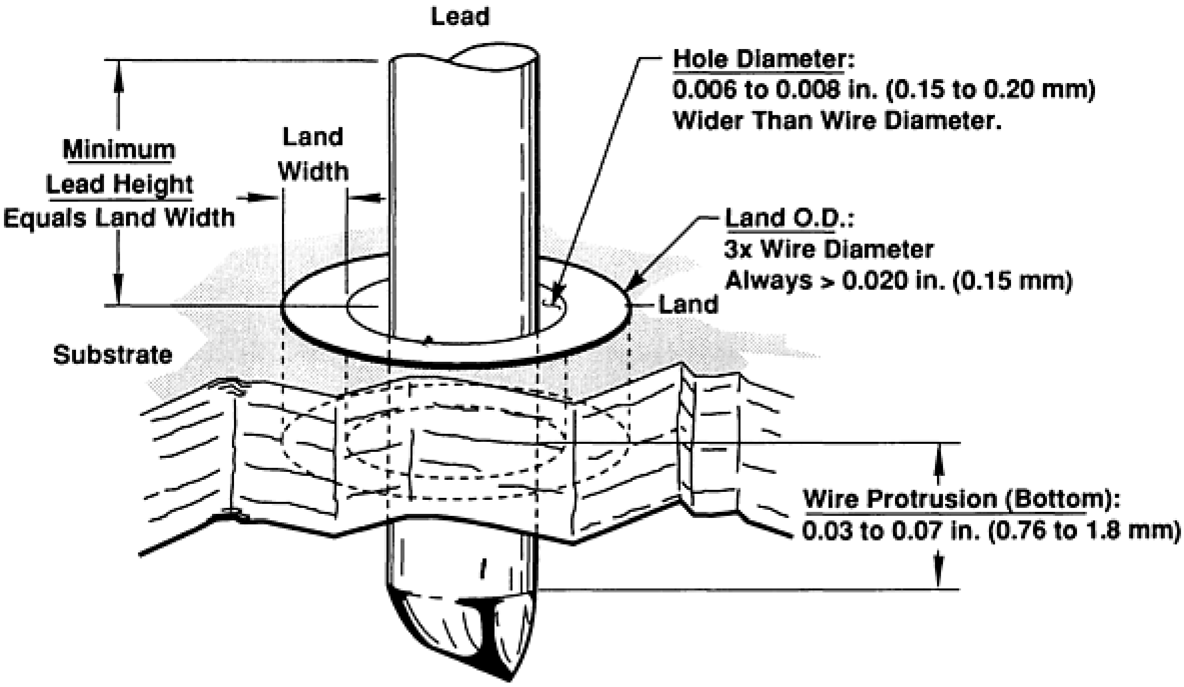
Lead
Hole Diameter:
0.006 to 0.008 in. (0.15 to 0.20 mm)
Wider Than Wire Diameter.
Minimum
Lead Height
Equals Land Width
Land
Width
Land O.D.:
3x Wire Diameter
Always > 0.020 in. (0.15 mm)
Land
Substrate
Wire Protrusion (Bottom):
0.03 to 0.07 in. (0.76 to 1.8 mm)

FIG. 9 GENERAL GUIDELINES FOR LEAD AND SUBSTRATE GEOMETRIES USED IN THROUGH-HOLE TECHNOLOGY

Loss of electrical continuity in the through-hole can arise from cracking (termed barrel cracking) of the hole plating layer because of the large CTE in the thickness dimension of the laminate (nearly a factor of 10 higher than in the laminate plane). This phenomena is particularly severe for via holes with large aspect ratios (the board thickness/hole diameter). Aspect ratios should be kept to a value smaller than 3 to ensure adequate reliability by typical assembly processes (Ref 24). The use of holes with larger aspect ratios requires additional attention to reliability.

Calculations have shown that the mechanical strength of the through-hole joint is limited primarily by PWB properties, specifically, the peel strength of the copper foil/laminate bond and the adhesion between the plating layer of the hole and the laminate (Ref 25). Similarly, the bond strength between the thick-film network and the ceramic substrate limits the strength of solder joints on hybrid microcircuits. Mechanical integrity can be lost through thermal fatigue. However, CTE mismatch between the device package and the PWB is not a principal source of thermal fatigue damage to through-hole solder joints, because the leads take up much of the displacement difference. Rather, thermal fatigue arises primarily from the difference between the lead CTE (for example, copper, at 9.4×10^{-6}/K, or 5.2×10^{-6}/°F) and the through-thickness thermal expansion value of PWB laminate (as much as 175×10^{-6}/K, or 98×10^{-6}/°F). Fatigue damage (Fig. 10) in through-hole solder joints is lessened by the CTE of the solder (25×10^{-6}/K, or 14×10^{-6}/°F), which is between those of the lead and PWB materials, as well as through proper assembly practices such as the incorporation of strain relief bends or loops placed in leads.

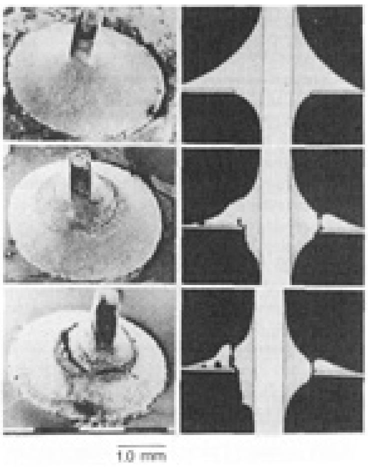

FIG. 10 PROGRESSION OF FATIGUE DAMAGE IN A THROUGH-HOLE SOLDER JOINT (OPTICAL MICROSCOPY AND SEM). SOURCE: REF 7

Plated-through holes are soldered from one side. Multilayer boards are more difficult to solder, because the interior conductor lines act as additional heat sinks that inhibit board heating. Solder assembly techniques for through-hole PWBs include hand soldering and larger-scale, automated processes such as dip, drag, or wave soldering. Solder pot working temperatures for the latter processes range from 240 to 260 °C (464 to 500 °F) for 63Sn-37Pb and 60Sn-40Pb solders. A good rule of thumb is to limit contact between the assembly and the molten solder to 3 to 10 s to minimize damage to the laminate and flux charring.

Heat damage to the attached device, neighboring components, and the substrate material must be considered when reworking through-hole solder joints. When a part is intended to be removed after assembly, it can be attached by hand soldering techniques with one of the lower melting temperature solders (Table 1). This technique is generally reserved for prototype assemblies; military and most commercial products specifications do not allow this operation.

Conformal coatings must be removed prior to reheating a solder joint. Clinched leads, that is, leads that are bent slightly on the opposite surface of the board to provide mechanical rigidity during assembly, hinder the removal of components. Generally, heat inputs will be greater when reworking a joint because the solder, lead, and land must be heated, which increases the likelihood of damage to the land/substrate bond or to the device. Cleanliness specifications must be followed after reworking to prevent functional failures caused by corrosion from flux residues.

The reliability of double-sided and multilayer circuit boards depends strongly on the plating quality of the hole walls. Water vapor and contaminants from the plating solutions can become entrapped in the deposited film. Plating layers should be free of embrittling contaminants that make them unable to adapt to board thermal expansion. The coating must have adequate thickness (and uniformity) so as not to be totally consumed by the metallurgical reaction with the solder. Contaminant materials vaporize at the soldering temperatures, causing cracks in the plating layer. Although plating cracks and voids, or delamination, do not necessarily jeopardize joint strength, they can interrupt electrical continuity between layers and surfaces.

A plated-through hole that is coated with a hot-dipped or plated-and-fused solder layer will have a very thin film of solder around its edge upon reflow. This phenomenon is a consequence of the surface physics of molten liquids. The limited protection offered by the thin solder film can cause nonwetting of the hole edge during assembly, in which case the joint is said to have a "weak knee." The effects of a weak knee on joint reliability are largely cosmetic for plated-through holes.

The traditional solders used in through-hole technology have been the tin-lead alloys. The solders with higher melting temperatures, such as tin-silver, tin-antimony, and lead-base tin-lead solders, have also been used in through-hole PWB systems designed for elevated-temperature service applications or step soldering processes. However, these solders are limited to specialized soldering assembly or those processes in which heat application is localized to the joint (for example, hot-air, hot-bar, fiber-optic infrared, or laser techniques), away from the electronic component. The higher heat inputs require that heat sinks be used on heat-sensitive devices. In addition, the higher reflow temperatures may decompose the typical rosin-base fluxes. Synthetic-activated (SA) fluxes, which have activities comparable to those of RMA fluxes and water-soluble fluxes, can withstand the higher-temperature operations.

Surface-mount technology utilizes several types of assembly techniques. The technology generally refers to products that use conventional PWB assemblies, which include organic laminates, clad-metal substrates, as well as ceramic substrates for hybrid microcircuits. Although the architecture differs between organic laminates and hybrid substrates, the package configurations for electronic devices is similar. Surface-mount technology also includes chip (silicon)-on-board technologies (such as tape automated bonding, TAB), controlled collapse chip connection (C^4), or flip chip.

PWB Assemblies (Organic Laminate and Ceramic). Denser board populations and adaptability to fully automated assembly processes have increased the popularity of surface-mount technology. Figure 11 shows several leaded and leadless surface-mount packages. Lands on surface-mount PWBs are constructed with the same conduction layer material specifications used on boards with through-hole components:

- LAND THICKNESSES (COPPER) ARE BETWEEN 0.018 AND 0.071 MM (0.0007 AND 0.0028 IN.).
- PROTECTIVE FINISHES ON PWBS INCLUDE HOT-DIPPED SOLDER WITH HOT-AIR LEVELING OR ORGANIC COATINGS (INHIBITORS) ON BARE COPPER.
- THE CONTROL OF SOLDER THICKNESS ON HOT-AIR LEVELED LANDS IS CRITICAL WITH SURFACE-MOUNT PWBS.
- EXCESSIVE SOLDER QUANTITIES WILL INTERFERE WITH PART ALIGNMENT DURING PLACEMENT, DAMAGE FINE LEADS DURING AUTOMATED SETUP OF COMPONENTS ON THE BOARD, AND INCREASE THE LIKELIHOOD OF NONSYMMETRIC FILLETS ON DEVICES.
- NONSYMMETRIC FILLETS CAN LEAD TO MISALIGNMENT ("TOMBSTONING") OF DEVICES DURING AUTOMATED SOLDERING PROCESSES.

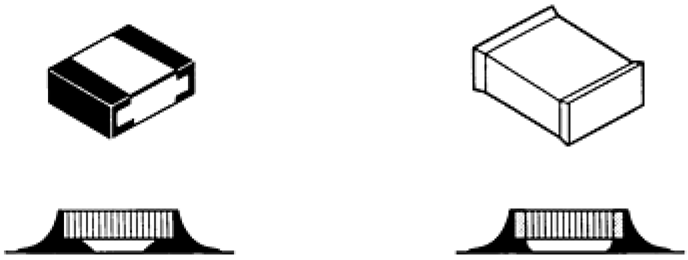

Leadless Ceramic Resistor

Leadless Ceramic Capacitor

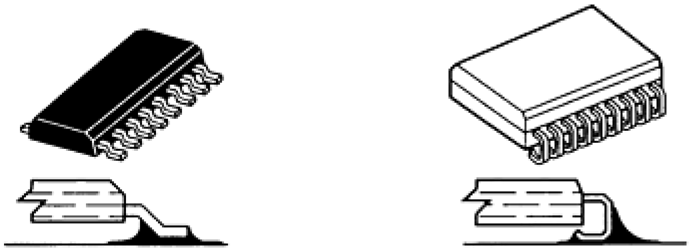

Small Outline Chip Carrier (Gullwing Lead)

Small Outline Chip Carrier (J- Lead)

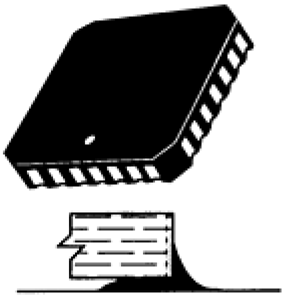

Leadless Ceramic Chip Carrier

FIG. 11 PACKAGE AND LEAD CONFIGURATION OF SURFACE-MOUNT DEVICES, INCLUDING SOLDER JOINT PROFILES

A variety of land geometries are used in order to accommodate the range of package input/output (I/O) configurations. Holes, or vias, may be required on the PWB to permit signal transmission between the two external surfaces or between internal conductors of multilayer boards. In some cases, through-hole components are also found on the same PWBs. These assemblies are known as mixed technology. The quality of the plating layer in the hole is critical to system reliability, whether in vias for signal transmission or component holes for mixed technology applications.

Conductor lines and lands on ceramic boards are composed of one of the thick-film inks. The pattern is screen printed into the circuit configuration and fired to promote adhesion. Lines are usually formed from a single printed layer, with a thickness of approximately 0.01 mm (400 μin.). Lands to which soldering will take place may receive a second printing (double-printed), which improves solder joint reliability by reducing porosity in the film and limiting the dissolution of the conductive layer that is due to the formation of intermetallics between the solder and the precious metal. Double-printed layers are typically from 0.02 to 0.025 mm (0.0008 to 0.0010 in.) thick, depending on the porosity of the first layer.

Leadless surface-mount resistors and capacitors use terminations to form the solder interconnect. Terminations are regions on the package surface that are coated with a solderable layer to promote wetting by the solder. Leadless chip resistors and capacitors have a silver-bearing thick-film ink that is fired onto areas of the ceramic substrate; these areas are connected to the active electrical elements. Dissolution, or scavenging, of the thick film is reduced by adding more than 35% Pd to its composition (Ref 26). More commonly, the thick film is coated with a solderable nickel or copper finish to prevent the scavenging of silver by the solder. Candidate protective layers for the solderable finishes include electroplated tin or tin-lead and hot-dipped tin-lead solder.

Leadless ceramic packages for integrated circuits have multiple I/O terminations composed of metallized grooves on the package sides (castellation) and a metallized bonding area on the bottom of the package (blind lap joint). The metallized surfaces have a refractory layer (tungsten or molybdenum) that provides the electrical feedthrough to the interior. The refractory layer is then coated with a nickel solderable layer (1.3 to 8.9 μm, or 50 to 150 μin.). A protective layer of gold (1.3 to 2.5 μm, or 50 to 100 μin.) coats the nickel. The gold finish is removed by a hot-dipped solder coating step to prevent gold embrittlement of the solder joints. Gold embrittlement is of particular concern in surface-mount solder joints, because the limited amount of solder required to form the small joints can easily exceed the 4 wt% maximum Au content observed to degrade solder properties. Lower potential gold contents are preferred as a safety factor.

Leaded packages for surface-mount assembly have conductors fabricated from copper or one of the iron-base alloys and formed into the appropriate configuration (J-lead, gullwing, and others). A solderable layer of electroplated or electroless nickel, followed by a protective coating of electroplated gold, is deposited onto the iron-base alloy leads, per the recommended thicknesses (see the section "Base Materials, Finishes, and Storage/Corrosion Issues" in this article). The leads are hot dipped in tin-lead solder to remove the gold coating prior to assembly or to protect the copper solderable finish; however, tin or tin-lead electroplated coatings are preferred with fine-pitched leads and packages to eliminate the "crown" of solder on the lead. This crown causes the lead to deform and slide off of the similar crown of solder on the land when the part is placed on the PWB, causing component misalignment, solder joint distortion, or electrical failure.

The predominant solders used in surface-mount PWBs have been the tin-lead alloys. The tin-lead-silver solder (62.5Sn-36.1Pb-1.4Ag) has improved flow and a slightly improved isothermal fatigue life (for service conditions with limited temperature variations), when compared with the eutectic tin-lead alloy. The silver component of this solder improves creep strength. Silver-bearing solders also restrict the scavenging of silver from silver-bearing thick-film terminations on leadless ceramic devices that do not have a solderable (barrier) coating. A noneutectic tin-lead-silver alloy formed by the addition of 2% Ag, or the use of 60Sn-40Pb solder, will reduce the occurrence of solder wicking on leaded devices. Solder wicking is a phenomenon on leaded devices whereby the entire quantity of molten solder is drawn away from the pad to the lead because the lead heats up faster than the pad. This problem is particular to the vapor-phase reflow process and is compounded by leads that are noncoplanar with one another.

The noneutectic solders melt more slowly than eutectic alloys because they melt over a temperature range. This allows the lead and bonding pad surfaces to reach the same temperature, so that solder will simultaneously wet both surfaces. A longer melting sequence of the noneutectic alloys also decreases the incidence of "drawbridging," or "tombstoning," on leadless chip components (Fig. 12). A cause of drawbridging is the nonsimultaneous melting of the solder at the two terminations, causing the component to be drawn toward the molten joint by the surface tension of the solder. Other factors in drawbridging are paste tackiness, paste quantity, part location, and land design.

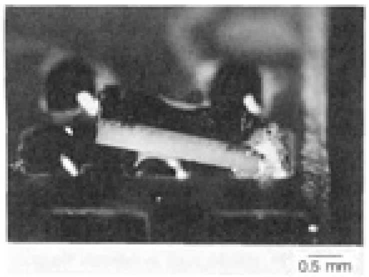

FIG. 12 DRAWBRIDGING OF A LEADLESS CERAMIC CHIP RESISTOR (OPTICAL MACROSCOPY). COURTESY OF SANDIA NATIONAL LABORATORIES

Other solders used for surface mounting include lead-indium alloys, which are used in hybrid microcircuit systems. The lead-indium solders limit the leaching of precious metals from thick-film networks and finishes on devices, and reduce the formation of brittle intermetallic compounds that jeopardize solder joint integrity. In addition, these alloys are more ductile than the tin-lead solders, thereby limiting the loads on the more-fragile thick-film/substrate ceramic bond. The low-temperature ductility of these solders provides a niche for them in cryogenic applications (for example, space vehicle electronics).

The low-melting-temperature tin-bismuth solders are used on PWBs with heat-sensitive devices or with those components that have to be removed and replaced on the board one or more times to satisfy testing or operational modes.

The high tin-silver solders are well suited for surface-mount assemblies because of their improved fatigue properties as compared to those of the tin-lead alloys. Their higher melting temperatures (96.5Sn-3.5Ag eutectic, 221 °C, or 430 °F) increase the service temperature window of assemblies, as well. It is critical that the finishes on substrate lands and device leads or terminations be compatible with the solder alloy. The mixture of different metals, which is due to nonsimilar pastes and finishes, can produce lower-melting-temperature phases that can deteriorate the physical and mechanical properties of the solder joints. For example, mixing bismuth-containing solders and lead finishes with tin-lead solder causes a low-melting-temperature tin-lead-bismuth phase that melts at 96 °C (205 °F).

The use of alternative solders implies solder joint cosmetics that are different from those of the tin-lead alloys. Cosmetics should not be relied upon to judge service performance. Typically, the tin-bismuth, lead-indium, and tin-silver solders have a grainy solder fillet. The fillets of lead-indium and tin-silver will be less concave than those of tin-lead, because of their higher surface tensions.

Typical fluxes used for surface-mount assembly are the rosin-base materials, usually the RMA forms. Water-soluble fluxes (organic acids) are being used increasingly in electronics, because of the variety of aqueous cleaning processes available to remove their residues. Although flux residue removal is determined by the reliability required of the assembly and flux activity, surface-mount technology requires additional attention to cleaning requirements and procedures. The high density of devices places conductors closer together, so that ionic residues are more likely to cause electrical shorts by the electromigration mechanism. In addition, flux residues can be easily entrapped in the 0.05 to 0.013 mm (0.002 to 0.005 in.) gap between the package and the substrate.

Surface-mount PWBs are manufactured by mass production techniques, such as vapor-phase reflow, heating in IR furnaces, or wave soldering. These processes, and numerous others, are described in separate articles in the Section "Joining Processes" in this Volume. Briefly, a typical assembly sequence for reflow techniques (that is, using vapor-phase or furnace processes) involves:

- DEPOSITING SOLDER PASTE ON THE SUBSTRATE LANDS BY EITHER SCREEN PRINTING, STENCIL PRINTING, OR BULK DISPENSING TECHNIQUES
- BAKING THE PASTE TO DRIVE OFF VOLATILES
- PLACING THE ELECTRONIC COMPONENTS ON THE PASTE DEPOSITS, OVER THE LANDS (MANUALLY OR WITH ROBOTICS)
- REFLOWING THE SOLDER
- CLEANING THE BOARDS OF FLUX RESIDUES, IF NECESSARY

The use of alternate solder alloys depends on their availability in paste form (with the optimum flux and flow properties). Preforms or wires of nontypical solders are more readily available. For vapor-phase reflow, working fluids that would accommodate solders with higher melting temperatures than tin-lead alloys may have limited availability. Infrared heating is more versatile in this regard. Damage to heat-sensitive components must be considered when higher-melting-temperature solders are used. Wave soldering utilizes a solder bath, which requires that an ingot form of the solder be available in order to fill the pot. Surface-mount parts are typically secured to the PWB with epoxies prior to being passed through the wave.

Some surface-mount PWBs are hand soldered. This technique is reserved for larger-pitch assemblies and runs of small quantities or for very high reliability electronic systems. It is also used as a means of repair and rework. For component placement, the conventional soldering iron tip has been replaced by hot bars or rectangles that conform to the various I/O configuration of surface-mount packages in order to simultaneously melt the solder in all joints. Heating rates should be minimized on chip capacitors and resistors to prevent cracking of the ceramic chip or thick-film termination. Total heat input from the iron must not damage the land/substrate bond. An intermetallic layer formed between the copper land and molten tin-lead solder generally is not sufficient to restrict rework operations. However, solder removal that exposes the intermetallic layer to atmospheric oxidation can deteriorate solderability. Hand soldering processes may improve solder joint fatigue life by resulting in a finer solder microstructure, because of the fast cooling rate of joints after reflow (Ref 27). However, the fine microstructure coarsens with time and temperature, causing the strength advantage to be lost.

Solder joints for surface-mount technology fulfill electrical and mechanical attachment requirements. Adequate joint configurations have more than sufficient bonding strength to secure the device to the board. Shear tests (load application parallel to the board surface) are typically performed on leadless chip resistors and capacitors or on leadless ceramic packages to quantify the joint strength. Shear loads of chip resistors on polyimide-quartz PWBs are shown in Table 7 for as-fabricated, thermally cycled, and thermally shocked units (Ref 28). Leaded packages are more often tested in tension, that is, the load is applied perpendicular to the PWB surface.

TABLE 7 SHEAR STRENGTH OF SURFACE-MOUNT LEADLESS CERAMIC CHIP RESISTOR

TEST IDENTIFICATION	SHEAR STRENGTH	
	N	LBF
AS FABRICATED	84 ± 8	**19.0 ± 1.7**
300 THERMAL CYCLES	72 ± 6	**16.2 + 1.3**
100 THERMAL SHOCK CYCLES	**97 ± 4**	**21.9 ± 1.0**

Resistor dimensions, 2.67 × 1.27 × 0.457 mm (0.105 × 0.050 × 0.0018 in.); termination width, 0.25 mm (0.010 in.); thermal cycle, -55 to 125 °C (-67 to 257 °F), 120 min hold period, 6 °C/min (11 °F/min) ramp; thermal shock, -55 to 125 °C (-67 to 257 °F), 10 min hold period, liquid-to-liquid transfer; displacement rate, l0 mm/min (0.41 in./min).

Source: Sandia National Laboratories

A particular concern with surface-mount solder joints is thermal fatigue damage caused by the coefficient of thermal expansion (CTE) mismatch between the PWB laminate and the device package. Figure 13 shows scanning electron micrographs (SEM) of fatigue cracks on several different device packages. Thermal fatigue damage is of greater concern on leadless packages than on leaded configurations, because the latter can accommodate some of the displacement mismatch through the leads. For leadless packages (resistors, capacitors, or chip carriers), cracks begin under the package, as shown in Fig. 13(e), and grow toward the outer surface. Thus, cracks are well established prior to being visually detected on the fillet surface.

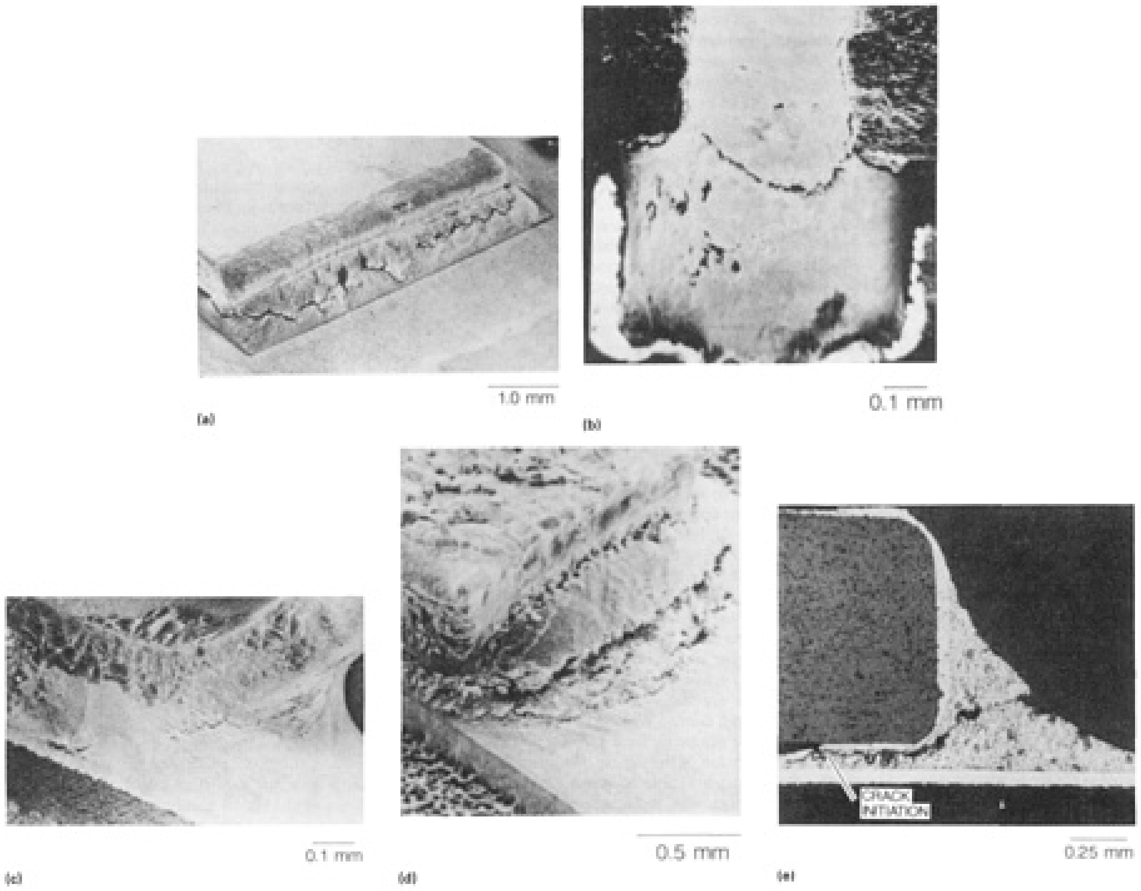

FIG. 13 FATIGUE CRACKS IN SURFACE-MOUNT SOLDER JOINTS. (A) LEADLESS CAPACITOR. (B) LEADLESS CERAMIC CHIP CARRIER. (C) J-LEADED PACKAGE. (D) GULLWING LEADED PACKAGE (SEM). (E) CRACKING IN A LEADLESS CERAMIC CHIP RESISTOR SOLDER FILLET. SOURCE: REF 7

Fatigue deformation that occurs under variable temperature conditions greatly complicates the prediction of solder joint failures. However, the thermal fatigue life of surface-mount solder joints can be estimated from isothermal fatigue test data when the microstructural damage is not strongly temperature dependent. Isothermal fatigue tests on surface-mount and leadless ceramic chip carrier 60Sn-40Pb solder joints at 35 °C (95 °F) and 125 °C (257 °F) show similar behavior. Tests at -55 °C (-67 °F) suggest a longer fatigue life, when compared with data from the higher-temperature tests (Ref 29). Therefore, isothermal fatigue test data on 60Sn-40Pb, taken at 35 °C (95 °F), can be used to estimate thermal fatigue performance over the temperature range from -25 to 125 °C (-77 to 257 °F).

The thermal fatigue life for eutectic and near-eutectic tin-lead solders can be predicted with the aid of Fig. 14. The configuration of the solder joint being tested (leadless ceramic chip resistor) is shown in Fig. 14(a). Let α_1 and α_2 be the CTEs of the package and PWB, respectively. ΔT is the absolute temperature range. The extension (or contraction) difference between the package and PWBs, Δx, is given as:

$$\Delta X = (\alpha_1 - \alpha_2)\,\Delta T L \qquad \textbf{(EQ 1)}$$

where L is the maximum separation between the solder joints, given the conservative assumption that all damage occurs to one joint.

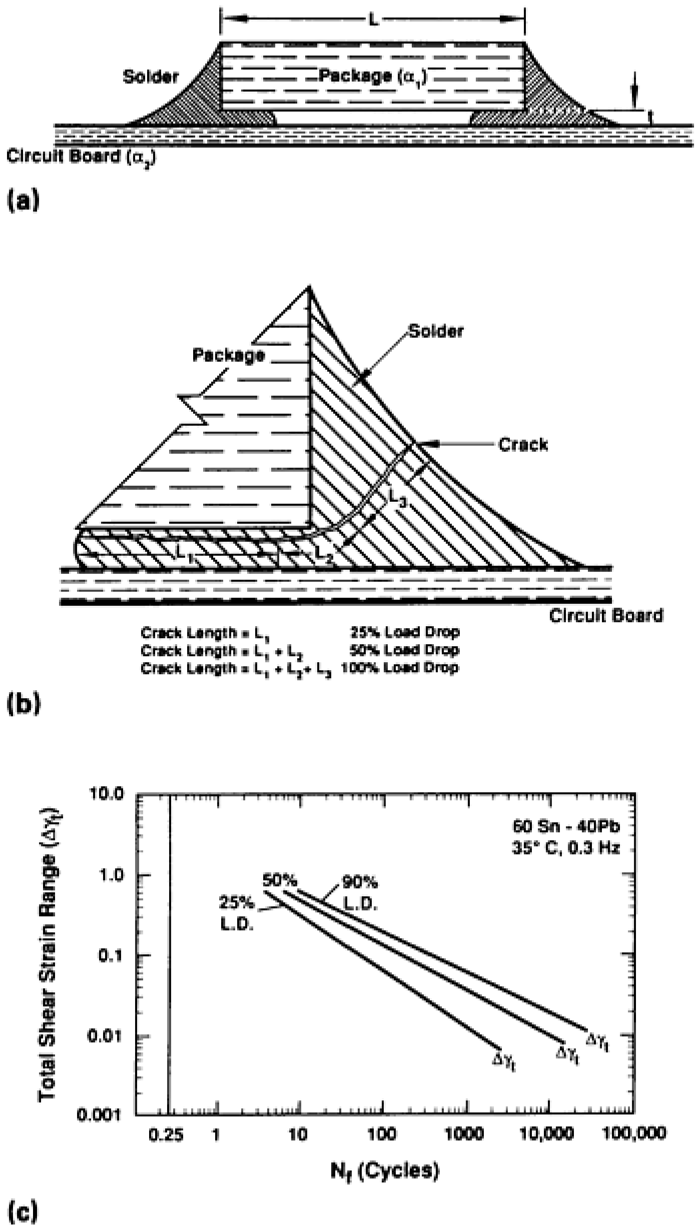
L
Solder
Package (α_1)
t
Circuit Board (α_2)
(a)
Package
Solder
Crack
L_3
L_1
L_2
Circuit Board
Crack Length = L_1 25% Load Drop
Crack Length = $L_1 + L_2$ 50% Load Drop
Crack Length = $L_1 + L_2 + L_3$ 100% Load Drop
(b)
Total Shear Strain Range ($\Delta\gamma_t$)
10.0
1.0
0.1
0.01
0.001
60 Sn - 40Pb
35° C, 0.3 Hz
50%
90% L.D.
25% L.D.
$\Delta\gamma_t$
$\Delta\gamma_t$
$\Delta\gamma_t$
0.25
1
10
100
1000
10,000
100,000
N_f (Cycles)
(c)

FIG. 14 ISOTHERMAL FATIGUE LIFE CALCULATION FOR THERMAL FATIGUE RESISTANCE APPROXIMATION. (A) PACKAGE/SOLDER JOINT GEOMETRY. L IS THE PACKAGE SIZE (DISTANCE BETWEEN SOLDER JOINTS), α_l IS THE THERMAL EXPANSION COEFFICIENT, AND T IS THE SOLDER JOINT GAP. (B) FATIGUE LIFE VERSUS TOTAL SHEAR STRAIN RANGE PER LOAD DROP CRITERIA. (C) FATIGUE CRACK GROWTH CORRESPONDING TO PERCENT LOAD DROP CRITERIA FOR FATIGUE LIFE. SOURCE: REF 29, 30, AND 31

Equipartition of strain between the two joints requires the use of (*L*/2) in place of *L*. The total shear strain rate, $\Delta\gamma_t$, is:

$$\Delta g_t = \frac{\Delta x}{t} \qquad \textbf{(EQ 2)}$$

where *t* is the solder joint gap. For the crack extension (and associated load drop) depicted in Fig. 14(b), the value of $\Delta\gamma_t$ is found on the ordinate of Fig. 14(c), and the number of cycles to failure, N_f, is determined for each of the three failure criteria. The computed fatigue life is an approximation, that is, its accuracy is an order of magnitude. An in-depth treatment of the solder joint reliability issue is provided in Ref 32.

The data presented in Fig. 14(c) pertain to a cycle period of 3.3 s and are suitable for calculations of cycle times up to 1000 s (Ref 31). Periods in excess of 10,000 s/cycle cause the analysis described above to *overestimate* the fatigue life by nearly an order of magnitude.

Thermal fatigue life can be lengthened by matching the CTE of the substrate to that of the package material (note Table 6). Fatigue resistance is improved by increasing the standoff distance between the component and the substrate (that is, increase the value of *t*). Other techniques include the use of nonmelting balls in the solder paste or protrusions on the bonding surface of the package, both of which act as spacers to control the gap. In addition, it has been shown that fatigue life increases as the solder joint cooling rate is increased (Ref 33). The cooling rate can be controlled on automated assembly equipment such as IR furnaces and some vapor-phase units. However, thermal shock to the electronic devices and the substrate must be considered when increasing the cooling rate. Fast cooling rates are an inherent feature of localized heating techniques, such as hand soldering with an iron, hot gas or hot bar techniques, and laser soldering. However, room-temperature aging of the solder may counteract any fatigue life improvement.

Tape automated bonding (TAB) refers to bonding the silicon integrated circuit (IC) chip to a metal lead frame that, in turn, is soldered to the PWB. This technique is shown in Fig. 15. The Au_xSn_{1-x} solder inner lead bond, which attaches the silicon chip to the lead frame, is formed by the intermixing of gold and tin diffusing from corresponding metallizations deposited on the chip and lead, respectively, under elevated temperature and pressure. The outer lead bond, which attaches the silicon chip plus lead frame to the circuit substrate, is a tin-lead solder joint formed by one of the conventional assembly processes. A lead exists for each of the I/Os on the silicon chip. The thermal fatigue of solder joints is reduced by the lead, which takes up the thermal strain mismatch between the silicon IC and the substrate. Additional compliance is added to the lead by jogs or bends in the geometry.

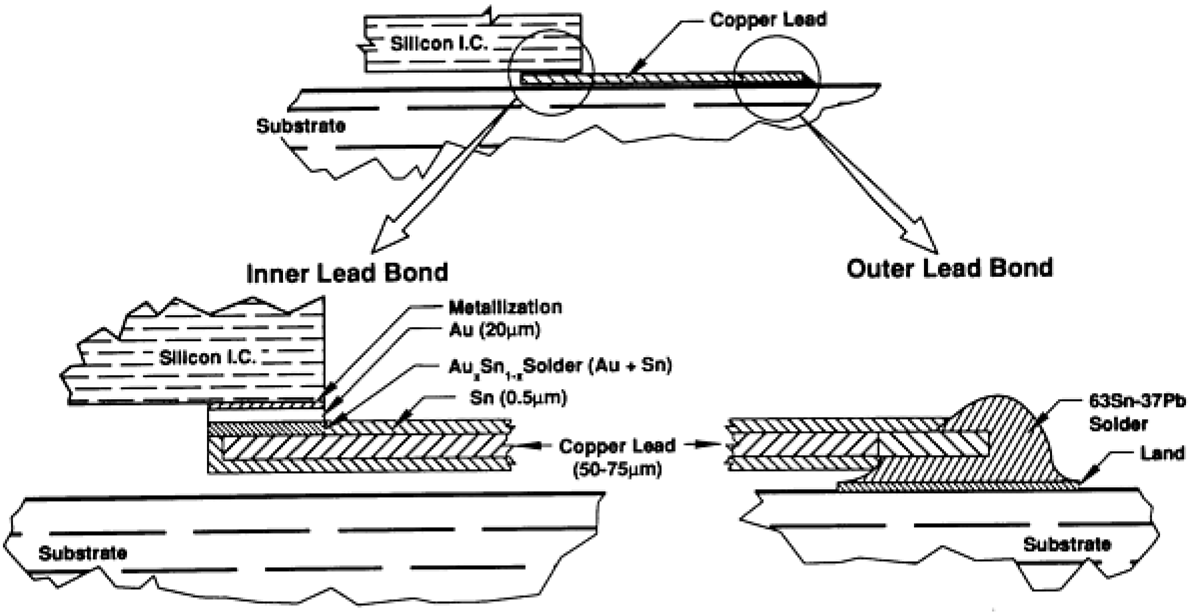

FIG. 15 TAPE AUTOMATED BONDING OF A SILICON CHIP TO THE SUBSTRATE

Controlled collapse chip connection (C^4) describes the process whereby the silicon IC chip is bonded directly to a ceramic substrate by small solder bumps that also serve as the signal I/O path (Fig. 16). This technique is widely used in computer assembly technology. The solderable surface on the silicon is a vacuum-deposited copper layer. A vacuum-deposited chromium film promotes adhesion of the copper to the SiO_2 layer. A gold layer (also vacuum-deposited) is the final protective finish. The ceramic substrate can be single-layered or multilayered.

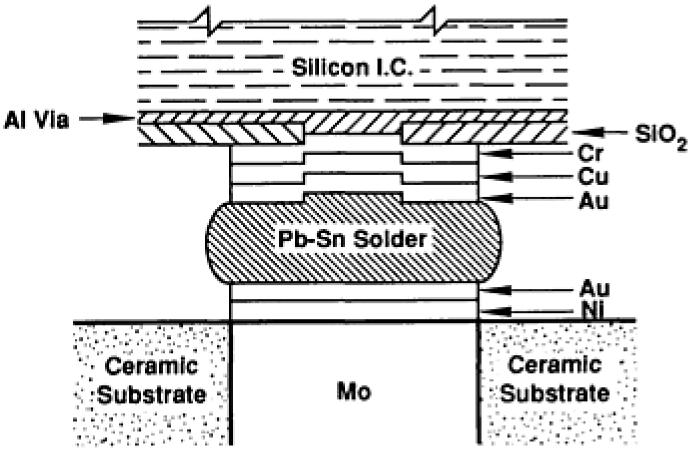

FIG. 16 SOLDER "BUMP" JOINT USED IN CONTROLLED COLLAPSE CHIP CONNECTION

When working with solder bumps to molybdenum vias, the molybdenum is coated with vacuum-deposited nickel and gold films, where nickel is the solderable surface and gold is the protective finish. The solder can be one of the lead-rich, tin-lead alloys (90Pb-10Sn or 95Pb-5Sn) or a lead-indium alloy (50In-50Pb). The lead-tin alloys are suited for assemblies that will later be processed with eutectic tin-lead soldering. The solder films are vacuum deposited onto the silicon chip in thicknesses ranging from 0.100 to 0.123 mm (0.004 to 0.005 in.). The units are passed through a belt furnace under a protective atmosphere to melt the solder.

The solder joints of C^4 interconnects must accommodate the thermal strain mismatch between the silicon IC and the substrate. Fatigue damage is reduced by maximizing and controlling uniformity of the silicon IC chip standoff distance above the substrate (hence the term, controlled collapse), limitation of the chip size, and the use of substrates with CTEs similar to that of silicon.

Connector Technology. In the signal analysis of an electronic product, the backplane forms the hub through which PWBs "talk" with one another and receive electrical power to operate devices mounted on them. As a result, the physical requirements placed on the backplane construction cause it to be a unique PWB. The backplane must mechanically support numerous PWBs and connector-cable assemblies that are frequently removed and reinserted in addition to heavier devices such as large capacitors or transformers. Therefore, the backplane is typically much thicker (2.4 to 3.2 mm, or 0.093 to 0.125 in.) than conventional PWBs (Ref 34). As the hub of electrical signal transmission, the backplane is often multilayered with thick conductor lines (0.071 to 2.8 mm, or 0.0028 to 0.11 in.), depending on the power transmission requirements of the system. Connector pins may be soldered on the substrate using preforms and vapor-phase assembly or manual processes. The thicker laminate and conductors as well as the multilayer construction can be a formidable heat sink to soldering processes in addition to being susceptible to warpage by overheating or inadequate support. Connector pins may also be press fit onto the substrate.

Signal transmission between devices that are not electrically connected via the PWB, between PWBs themselves, or between PWBs and the backplane, require the use of wires, cables, and connectors. Connectors are manufactured using attachment modes of soldering or a combination of soldering and mechanical fastening techniques. Examples of various connector techniques and design guidelines are shown in Fig. 17. The solder joints in connectors must be mechanically robust to accommodate loads caused by rigid, unyielding cables, as well as their frequent disconnection and reinstallation.

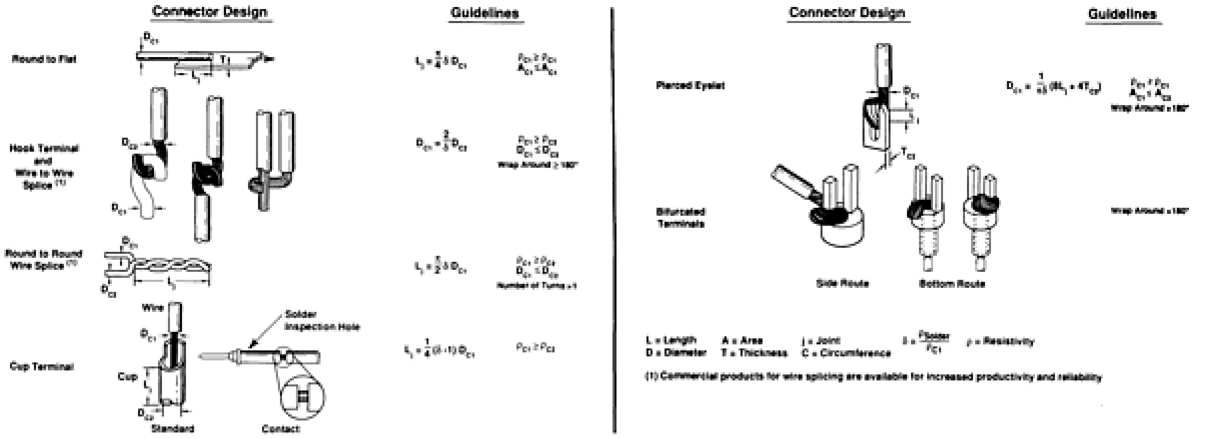
Connector Design
Guidelines
Round to Flat
Hook Terminal and Wire to Wire Splice (1)
Round to Round Wire Splice (1)
Wire
Solder Inspection Hole
Cup Terminal
Cup
Standard
Contact
Wrap Around ≥ 180°
Number of Turns ≥ 1
Pierced Eyelet
Bifurcated Terminals
Side Route
Bottom Route
L = Length
D = Diameter
A = Area
T = Thickness
j = Joint
C = Circumference
ρ = Resistivity
(1) Commercial products for wire splicing are available for increased productivity and reliability

FIG. 17 CONNECTOR METHODOLOGIES

Base materials for connectors, wires, and cable conductors include copper alloys (for example, oxygen-free, high-conductivity copper and electrolytic pitch-grade copper, brass, beryllium copper, and others), iron-base alloys (Kovar, stainless steel, and others), and aluminum alloys. Preparation of the base metal and selection of solderable and protective finishes for solder attachment are described in the section "Base Materials, Finishes, and Storage/Corrosion Issues" in this article. Because connector assemblies typically use larger quantities of solder to form the joint, the electroplated gold that serves as the protective layer may not have to be removed by a hot-dipped solder coating prior to assembly. The solder joint may have a frosty appearance, because of the dissolved gold in the solder. However, a hot-dipped coating should be applied if the thickness of gold electroplating and the estimated quantity of solder used in the joint indicate a dissolved gold content in the solder that will potentially exceed the 4 wt% limit, which causes embrittlement in tin-lead solder joints.

Pins, which are part of a plug-and-socket connection, can be gold plated with so-called "hard coatings," that is, the gold electroplate contains approximately 0.3 wt% cobalt or nickel to improve abrasion resistance during removal and installation cycles of the connector. However, the cobalt or nickel additions can cause the solderability of the pin to be poor.

Silver coatings are preferred on some connector parts in order to lower contact resistance. Unfortunately, silver is tarnished by the formation of sulfides from airborne sulfur compounds, which also deteriorate solderability. For grounding applications, soldering processes to aluminum connector housings should be followed with an electroplated coating of zinc or cadmium to guarantee adequate corrosion resistance.

Solders used in connector assemblies are generally tin-lead alloys. Additional strength, particularly for products exposed to elevated temperatures or mechanical loads, is obtained with the tin-silver and tin-antimony solders. The higher melting temperatures, coupled with the generally massive structure of connectors (when compared with electrical components), requires greater heat input into the assembly for proper joint fabrication. Joining can be facilitated by first hot dipping the mating substrates in the particular solder prior to final assembly. The use of preforms can facilitate the introduction of solder into joints, particularly those of very small assemblies.

The preferred fluxes for connectors are the rosin-base materials. Residues from pure rosin need not be removed. Mildly activated (RMA) flux residues are removed for high-reliability applications. Residues of more-active fluxes (RA, SA, or water soluble) must be removed. Hand soldering with low-solids fluxes in air can give inconsistent solderability because only very thin layers of tarnish can be removed by the flux. The higher heat inputs required for connector structures can degrade rosin-base fluxes, necessitating the use of water-soluble and synthetic fluxes that will tolerate the higher working temperatures. In either case, the residues must be removed, because they are corrosive.

Connectors are typically assembled by localized heating, such as by soldering iron, hot gas jet, or hot bar techniques. The use of automated assembly processes that exploit vapor-phase reflow or IR furnaces is prohibited in order to prevent the exposure of heat-sensitive structural materials in the connector to elevated temperatures. Moreover, the generally cumbersome assemblies are not well suited for batch or continuous processes.

References cited in this section

1. R. PRASAD, *SURFACE MOUNT TECHNOLOGY, PRINCIPLE AND PRACTICE,* VAN NOSTRAND-RHEINHOLD, 1989, P 363
7. R. KLEIN-WASSINK, *SOLDERING IN ELECTRONICS,* 2ND ED., ELECTROCHEM. PUB., LTD, 1989, P 548
22. H. MANKO, *SOLDERS AND SOLDERING,* MCGRAW-HILL, 1979, P 96
23. R. BAHN, PLATING, *PRINTED CIRCUIT HANDBOOK,* C. COOMBS, ED., MCGRAW-HILL, 1979, P 7-1
24. R. PRASAD, *SURFACE MOUNT TECHNOLOGY, PRINCIPLE AND PRACTICE,* VAN NOSTRAND-RHEINHOLD, 1989, P 158
25. H. MANKO, *SOLDERING HANDBOOK FOR PRINTED CIRCUITS AND SURFACE MOUNTING,* VAN NOSTRAND-RHEINHOLD, 1986, P 23
26. R. KLEIN-WASSINK, NOTES ON THE EFFECTS OF METALLIZATION OF SURFACE MOUNTED COMPONENTS ON SOLDERING, *HYBRID MICROCIR.,* MAY 1987, P 89
27. R. WILD, SOME FACTORS AFFECTING LEADLESS CHIP CARRIER SOLDER JOINT FATIGUE LIFE, *PROC. 12TH ELECTRON. MFG. SEMINAR* (CHINA LAKE, CA), NAVAL WEAPONS CENTER, 1988, P 275
28. P. VIANCO AND J. DAL PORTO, AN EVALUATION OF THE BLIND LAP JOINT FOR THE SURFACE MOUNT ATTACHMENT OF CHIP COMPONENTS, *PROC. SURF. MOUNT INTER.,* TO BE PUBLISHED, 1992
29. H. SOLOMON, "INFLUENCE OF TEMPERATURE ON THE LOW CYCLE FATIGUE OF SURFACE MOUNTED CHIP CARRIER/PRINTED WIRING BOARD JOINTS," TECHNICAL INFORMATION SERIES, GENERAL ELECTRIC CRD, 1987
30. H. SOLOMON *ET AL.,* "PREDICTION OF SOLDER JOINT FATIGUE LIFE," TECHNICAL INFORMATION SERIES, GENERAL ELECTRIC CRD, 1988
31. H. SOLOMON, "FATIGUE OF 60/40 SOLDER," TECHNICAL INFORMATION SERIES, GENERAL ELECTRIC CRD, 1986
32. W. ENGELMAIER, PERFORMANCE CONSIDERATIONS: THERMAL-MECHANICAL EFFECTS, *PACKAGING,* VOL 1, *ELECTRONIC MATERIALS HANDBOOK,* ASM, 1989
33. D. KINSER *ET AL.,* RELIABILITY OF SOLDERED JOINTS IN THERMAL CYCLING ENVIRONMENTS, *PROC. NEPCON,* MAY 1976, P 61
34. P. VIANCO, AN OVERVIEW OF THE MENISCOMETER/WETTING BALANCE TECHNIQUE FOR WETTABILITY MEASUREMENTS, *THE METAL SCIENCE OF JOINING,* M. CIESLAK *ET AL.,* ED., TMS, 1992, P 265

Soldering in Electronic Applications

Paul T. Vianco, Sandia National Laboratories

Solderability Testing in Electronics Applications

The formation of a solder joint that provides the necessary mechanical, electrical, and thermal properties for the particular device requires that all joint surfaces be wettable by the molten solder and that, upon solidification, a uniform solder film is retained. Solderability must be tightly controlled. Increased device functionality has placed greater demands on the performance of *each* solder joint. Full-scale manufacturing processes of PWBs, which can contain thousands of solder joints, make inspection and rework economically unfeasible.

The purpose of solderability testing is to quantitatively assess the wetting performance of component leads or terminations, as well as lands and holes of substrates, to ensure proper wetting performance by the solder at the time of PWB assembly. The goal is to prevent the expensive repair of solder joint defects on the assembled hardware and the operational failure of the device or entire electronic system. Solderability testing is typically performed on a lot sample with acceptance or rejection criteria that are based on the tested attribute(s).

Solderability tests, as well as lot sampling procedures and acceptance/rejection statistics, are specified by military and industrial specifications. Those specifications that control test procedures and acceptability criteria are listed in Table 8. The two predominant solderability tests are the dip-and-look evaluation and the wetting balance test.

TABLE 8 GUIDELINES AND SPECIFICATIONS FOR SOLDERABILITY TESTING OF ELECTRONICS APPLICATIONS

DOCUMENT	SUMMARY DESCRIPTION
MIL-STD-202F, METHOD 208F, SOLDERABILITY	**STRAND AND SOLID WIRES, LUGS, TABS, HOOK LEADS, AND SO ON, EVALUATED DIP-AND-LOOK TEST; INCLUDES STEAM AGING PROCEDURES**
MIL-STD-750C, METHOD 2026.5, SOLDERABILITY	**ELECTRONIC DEVICES EVALUATED BY DIP-AND-LOOK TEST PER MIL-STD-202F**
MIL-STD-883C, METHOD 2022.2, WETTING BALANCE SOLDERABILITY	**RIBBON LEADS EVALUATED BY THE WETTING BALANCE TEST; INCLUDES STEAM AGING PROCEDURES**
MIL-P-55110, GENERAL SPECIFICATION FOR PRINTED WIRING BOARDS	**INCLUDES SOLDERABILITY EVALUATION FOR RIGID PRINTED WIRING BOARDS**
MIL-P-50884C, PRINTED WIRING, FLEXIBLE AND RIGID-FLEX	**SOLDERABILITY EVALUATION OF FLEXIBLE AND RIGID-FLEX PRINTED WIRING BOARDS**
MIL-STD-883C, METHOD 5005.11, QUALIFICATION AND QUALITY CONFORMANCE PROCEDURES	**QUALITY CONFORMANCE AND CRITERIA FOR INSPECTION OF MICROELECTRONICS COMPONENTS USING INSPECTION TESTS (INCLUDING SOLDERABILITY TESTS)**
M1L-STD-150D, SAMPLING PROCEDURES AND TABLES FOR INSPECTION BY ATTRIBUTES	**SAMPLING PROCEDURES AND STATISTICS FOR ACCEPTANCE/REJECTION OF PARTS INSPECTED BY ATTRIBUTES (FOR EXAMPLE, SOLDERABILITY TESTING)**
MIL-STD-2000A, STANDARD REQUIREMENTS FOR SOLDERED ELECTRICAL AND ELECTRONIC ASSEMBLIES	**SPECIFIES SOLDERABILITY TESTS PER MIL-STD-202 AND CRITERIA FOR MAINTAINING SOLDERABILITY**
IPC-TM-650, TEST METHODS MANUAL	**SOLDERABILITY OF PRINTED WIRING BOARDS AND DIP-AND-LOOK PROCEDURE, AS WELL AS WAVE SOLDERING EVALUATION**
IPC-S-804, SOLDERABILITY TEST METHODS FOR PRINTED WIRING BOARDS	**SOLDERABILITY OF PRINTED WIRING BOARDS; EDGE DIP TEST ROTARY DIP TEST, WAVE SOLDERING TEST, THROUGH-HOLE SOLDER RISE TIME TEST; DIP-AND-LOOK TESTS**
IPC-S-805, SOLDERABILITY TESTS FOR COMPONENT LEADS AND TERMINATIONS	**SOLDERABILITY TESTS: DIP-AND-LOOK, WETTING BALANCE, AND GLOBULE TEST FOR COMPONENT LEADS AND TERMINATIONS; INCLUDES STEAM AGING PROCEDURES**
IPC-AJ-820, ASSEMBLY-JOINING HANDBOOK	**NO. 8.2.3, PROCEDURES FOR ACCELERATED AGING (IN CONJUNCTION WITH SOLDERABILITY TESTING)**
IPC-A-600C, GUIDELINES FOR	**INSPECTION CRITERIA FOR QUALITY ASSURANCE**

The dip-and-look test (for example, MIL-STD-883D, Method 2003.7) assesses the uniformity of the solder film that remains on a lead or termination after immersion in a solder pot. Shown in Fig. 18 are optical micrographs of samples demonstrating the three primary categories of wettability.

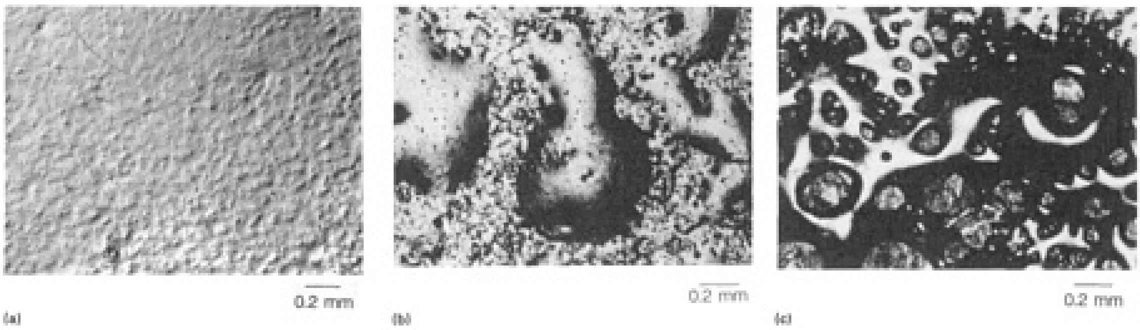

FIG. 18 OPTICAL MICROGRAPHS OF 60SN-40PB SOLDER ON SAMPLE SURFACES. (A) WETTING. (B) DEWETTING. (C) NONWETTING. SOURCE: SANDIA NATIONAL LABORATORIES

A less-subjective test is the wetting balance test (for example, MIL-STD-883C, Method 2022.2), in which a representative sample is attached to a force balance. The meniscus formed on the lead or termination causes a weight increase, which is detected by the balance. The typical wetting curve (force versus time) is shown in Fig. 19. The acceptability criteria of this test (by MIL-STD-883c), are the time for the trace to cross the zero line after the initial dip and the time required for the wetting force to reach two-thirds of its maximum. These parameters are also shown in Fig. 19. Variations to the wetting balance test procedure have been implemented to accommodate the wetting of terminations on leadless surface-mount devices. The wetting balance test should be accompanied by the dip-and-look procedure to verify the quality of the solder film. The wetting balance will *not* always detect dewetting of a solder film (Ref 34).

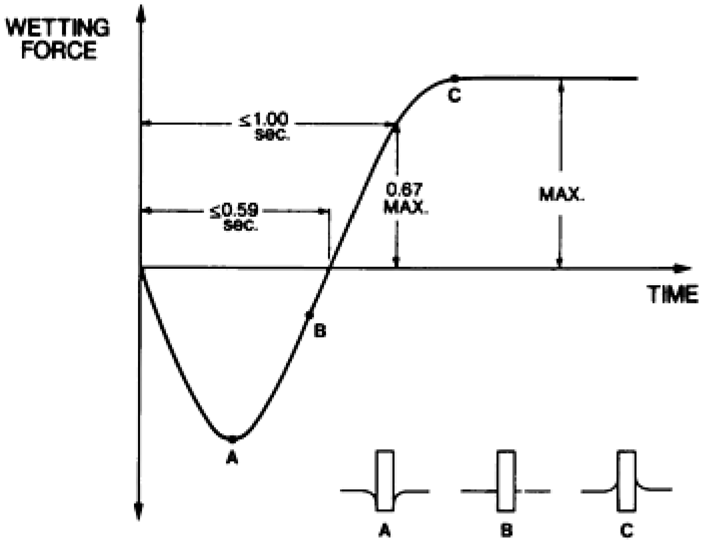

FIG. 19 WETTING BALANCE TRACE FOR SOLDERABILITY TESTS AND ACCEPTANCE CRITERIA PER MIL-STD-883C

Solderability tests for PWBs are typically dip-and-look tests, such as those specified in IPC-S-804 (J-STD-003) or the military specifications MIL-P-55110 and MIL-P-50804. Another test is the globule test, which measures the time for a solder globule to surround a horizontal wire or fill a hole (IPC-S-805).

The solderability of devices (and PWBs) deteriorates with time in storage. The tests noted above ensure adequate solderability within 120 days of the evaluation (MIL-STD-2000A). For longer periods, the steam aging treatment was instituted for the purpose of screening, on a lot sample basis, the leads, terminations, and other elements that would not be likely to retain adequate solderability under typical storage conditions prior to assembly. Steam aging (8.0 ± 0.5 h), followed by a successful solderability test, will indicate surfaces that can maintain adequate solderability for two years of "typical" storage conditions prior to use at the assembly level (MIL-STD-2000). Preferred storage conditions (see MIL-STD-2000; atmosphere, temperature, and humidity) can extend the storage time to 3 y. Parts that have exceeded the recommended storage life are considered to have questionable solderability and should be retested prior to use in an assembled product.

Reference cited in this section

34. P. VIANCO, AN OVERVIEW OF THE MENISCOMETER/WETTING BALANCE TECHNIQUE FOR WETTABILITY MEASUREMENTS, *THE METAL SCIENCE OF JOINING*, M. CIESLAK *ET AL.*, ED., TMS, 1992, P 265

Soldering in Electronic Applications

Paul T. Vianco, Sandia National Laboratories

References

1. R. PRASAD, *SURFACE MOUNT TECHNOLOGY, PRINCIPLE AND PRACTICE,* VAN NOSTRAND-RHEINHOLD, 1989, P 363
2. V. MARCOTTE, C^4 FLIP CHIP JOINING, *THE METAL SCIENCE OF JOINING*, M. CIESLAK *ET AL.*, ED., TMS, 1992, P 315
3. C. LEA, A SCIENTIFIC GUIDE TO SURFACE MOUNT TECHNOLOGY, ELECTROCHEM. PUB., LTD, 1988, P 169
4. R. GEHMAN, DENDRITIC GROWTH EVALUATION OF SOLDERED THICK FILMS, *INT. J. HYBRID MICROELEC.,* VOL 6, 1983, P 239
5. H. MANKO, *SOLDERS AND SOLDERING,* MCGRAW-HILL, 1979, P 112
6. L. GOLDMANN *ET AL.*, LEAD-INDIUM FOR CONTROLLED-COLLAPSE CHIP JOINING, *PROC. 27TH ELECTRON. COMP. CONF.*, 1977, P 25
7. R. KLEIN-WASSINK, *SOLDERING IN ELECTRONICS,* 2ND ED., ELECTROCHEM. PUB., LTD, 1989, P 548
8. R. KLEIN-WASSINK, *SOLDERING IN ELECTRONICS,* 2ND ED., ELECTROCHEM. PUB., LTD, 1989, P 557
9. J. LEE, SOLDER PASTE DISPENSING MATERIALS AND REQUIREMENTS, *ELECTRON. PACK. AND PROD.,* NOV 1990, P 22
10. H. MANKO, *SOLDERING HANDBOOK FOR PRINTED CIRCUITS AND SURFACE MOUNTING,* VAN NOSTRAND-RHEINHOLD, 1986, P 286
11. D. LOVERING, ROSIN ACIDS REACT TO FORM TAN RESIDUES, *ELECTRON. PACK. AND PROD.*, FEB 1985, P 232
12. D. KOCKA, NO-CLEAN FLUXES ARE A VIABLE ALTERNATIVE TO CFC CLEANING, *ELECTRON. PACK. AND PROD.,* JUNE 1990, P 95

13. P. VIANCO AND J. DAL PORTO, "EMBRITTLEMENT OF SURFACE MOUNT TRANSISTOR SOLDER JOINTS INVOLVING PRETINNED LEADS," INTERNATIONAL BRAZING AND SOLDERING CONFERENCE (DETROIT, MI), AWS, 1991
14. D. UNSWORTH AND C. MACKAY, A PRELIMINARY REPORT ON GROWTH OF COMPOUND LAYERS ON VARIOUS METAL BASES PLATED WITH TIN AND ITS ALLOYS, *TRANS. INST. MET.,* VOL 51, 1973, P 85
15. P. KAY AND C. MACKAY, THE GROWTH OF INTERMETALLIC COMPOUNDS ON COMMON BASIS MATERIALS COATED WITH TIN AND TIN-LEAD ALLOYS, *TRANS. INST. MET.,* VOL 54, 1976, P 68
16. P. VIANCO *ET AL.,* SOLDERABILITY TESTING OF KOVAR WITH 60SN-40PB SOLDER AND ORGANIC FLUXES, *WELD. J.,* JUNE 1990, P 230-S
17. C. LEA, *A SCIENTIFIC GUIDE TO SURFACE MOUNT TECHNOLOGY,* ELECTROCHEM. PUB., LTD, 1988, P 167
18. M. BESTER, METALLURGICAL ASPECTS OF SOLDERING GOLD OR GOLD PLATING, *PROC. INTERNEPCON,* 1968, P 221
19. P. HOLMES AND R. LOASBY, *HANDBOOK OF THICK FILM TECHNOLOGY,* ELECTROCHEM. PUB., LTD., 1976, P 192
20. R. WILD, COMPONENT LEAD SOLDERABILITY VS. ARTIFICIAL AGEING II, *PROC. 1 L TH ELECT. MFG. SEMINAR* (CHINA LAKE, CA), NAVAL WEAPONS CENTER, 1987, P 289
21. R. EDINGTON AND L. CONRAD-LOWANE, AGING ENVIRONMENTS AND THEIR EFFECTS ON SOLDERABILITY, *PROC. 12TH ELECTRON. MFG. SEMINAR* (CHINA LAKE, CA), NAVAL WEAPONS CENTER, 1988, P 31
22. H. MANKO, *SOLDERS AND SOLDERING,* MCGRAW-HILL, 1979, P 96
23. R. BAHN, PLATING, *PRINTED CIRCUIT HANDBOOK,* C. COOMBS, ED., MCGRAW-HILL, 1979, P 7-1
24. R. PRASAD, *SURFACE MOUNT TECHNOLOGY, PRINCIPLE AND PRACTICE,* VAN NOSTRAND-RHEINHOLD, 1989, P 158
25. H. MANKO, *SOLDERING HANDBOOK FOR PRINTED CIRCUITS AND SURFACE MOUNTING,* VAN NOSTRAND-RHEINHOLD, 1986, P 23
26. R. KLEIN-WASSINK, NOTES ON THE EFFECTS OF METALLIZATION OF SURFACE MOUNTED COMPONENTS ON SOLDERING, *HYBRID MICROCIR.,* MAY 1987, P 89
27. R. WILD, SOME FACTORS AFFECTING LEADLESS CHIP CARRIER SOLDER JOINT FATIGUE LIFE, *PROC. 12TH ELECTRON. MFG. SEMINAR* (CHINA LAKE, CA), NAVAL WEAPONS CENTER, 1988, P 275
28. P. VIANCO AND J. DAL PORTO, AN EVALUATION OF THE BLIND LAP JOINT FOR THE SURFACE MOUNT ATTACHMENT OF CHIP COMPONENTS, *PROC. SURF. MOUNT INTER.,* TO BE PUBLISHED, 1992
29. H. SOLOMON, "INFLUENCE OF TEMPERATURE ON THE LOW CYCLE FATIGUE OF SURFACE MOUNTED CHIP CARRIER/PRINTED WIRING BOARD JOINTS," TECHNICAL INFORMATION SERIES, GENERAL ELECTRIC CRD, 1987
30. H. SOLOMON *ET AL.,* "PREDICTION OF SOLDER JOINT FATIGUE LIFE," TECHNICAL INFORMATION SERIES, GENERAL ELECTRIC CRD, 1988
31. H. SOLOMON, "FATIGUE OF 60/40 SOLDER," TECHNICAL INFORMATION SERIES, GENERAL ELECTRIC CRD, 1986
32. W. ENGELMAIER, PERFORMANCE CONSIDERATIONS: THERMAL-MECHANICAL EFFECTS, *PACKAGING,* VOL 1, *ELECTRONIC MATERIALS HANDBOOK,* ASM, 1989
33. D. KINSER *ET AL.,* RELIABILITY OF SOLDERED JOINTS IN THERMAL CYCLING ENVIRONMENTS, *PROC. NEPCON,* MAY 1976, P 61
34. P. VIANCO, AN OVERVIEW OF THE MENISCOMETER/WETTING BALANCE TECHNIQUE FOR WETTABILITY MEASUREMENTS, *THE METAL SCIENCE OF JOINING,* M. CIESLAK *ET AL.,* ED.,

TMS, 1992, P 265

35. F. KEAR, *PRINTED CIRCUIT ASSEMBLY MANUFACTURING,* MARCEL DEKKER, 1987, P 8-22

Introduction to Special Welding and Joining Topics

Hendrikus H. Vanderveldt, American Welding Society

Introduction

WELDING, BRAZING, AND SOLDERING are pervasive in our industrialized society, and it should come as no surprise that there are many unique aspects related to these.joining technologies that are not covered within the more traditional topic organization in the remainder of this Volume. It is the purpose of this "catchall" Section to provide coverage of some important topics that do not fit neatly in other sections of the book. However, no handbook--no matter how comprehensive--can exhaustively cover all subjects related to joining technology. This Section seeks to provide information about some of the more common, albeit unique, areas of interest and concern to the joining community. The articles that make up this Section provide overviews of these areas, point the reader towards more extensive literature, and help to identify additional areas of investigation.

The Section opens with an article on thermal spray coatings. The application of thermal spray aluminum and zinc coatings has seen a significant increase in the United States during recent years. The corrosion protection provided by these coatings can be as much as an order of magnitude greater than that obtained with traditional coating approaches. In addition, thermal spray coatings can be used to increase wear and abrasion resistance of in-service equipment and to repair and upgrade industrial components.

Extreme Welding Environments

The next few articles in the Section provide information about extreme welding environments and service conditions. The construction of ships and offshore structures makes use of many of the same joining technologies used in shore-based construction industries, but also has its own set of unique requirements related to construction materials, fabrication environment, joining techniques, and so on. The article on underwater welding in this Section details some of the requirements involved during the original construction phase or later in-place repairs of large maritime structures; in addition, reference material is provided to point the reader to other sources of information.

Application of materials in cryogenic service (where temperatures can be as low as 4 K) provides opportunities to exploit special material properties associated with such low temperatures (superconductivity, for example). However, these opportunities must be weighed against some potential high penalties, such as the loss of structural properties exhibited at cryogenic operating temperatures. These problems can be significantly amplified in the joints produced for components intended for cryogenic service. The article "Welding for Cryogenic Service" discusses the impact of extremely low temperatures on the performance of welded structures.

The existence of long-term manned facilities in space is coming closer and closer to reality. Hence, the need to build structures or do repairs in space must be addressed. To date, very limited amount of work has been done to prepare ourselves for that day. The article "Welding in Space and Low-Gravity Environments" seeks to address what has been done and provide some guidance for future development efforts.

This Section also addresses the joining of nonmetallic and special-purpose materials. Although metals are generally the structural materials of choice, it is evident that increasing use is being made of plastics and composites in applications that include automotive parts and aerospace components, among many others. These materials are used because they exhibit high strength to weight and high stiffness, are producible as to properties, and have demonstrated reliability.

Experience related to joining of metals does not directly translate to joining of composite materials. The article "Joining of Organic-Matrix Composites" in this Section provides an extensive overview of this important topic as well as supplemental references to other sources of information. "Composite-to-Metal Joining" addresses the key factors involved in joining these dissimilar materials to one another.

The joining of plastics is generally accomplished by adhesive bonding or mechanical fastening. However, thermoplastic resins can be molded through the application of heat, and they can also be fusion welded. Various means of accomplishing such welding are discussed in the article "Welding of Plastics" in this Section.

Oxide-dispersion-strengthened materials provide high creep strength at very high temperatures and thus are used for applications such as gas turbine engines, high-temperature processing components, and so on. However, significant difficulties can be encountered in the processing of the oxide dispersions of these alloys. These can result in subsequent welding problems, such as porosity in the weldment and potentially in the adjacent heat-affected zone. The current state of the art in this area is high-lighted in the article "Joining of Oxide-Dispersion-Strengthened Materials" references are provided for further evaluation of this technology.

Computer technology has advanced to the stage where its application to welding and joining can make a major impact on the technical quality and cost of weldments, and this technology has the potential to revolutionize the entire joining manufacturing industry. The advent of personal desktop computer systems with high computational speed; the ability to model very complex parametric interrelationships without the need for exact theoretical understanding of those relationships; and the availability of comprehensive data base systems, computer-based engineering worksheets, and so on, all allow the modern manufacturing engineer to accomplish complex welding and joining tasks using minimal time and effort. The article "Intelligent Automation for Joining Technology" provides a general overview of this subject and detailed information about one particular development effort in this area. Related information is contained in the Section "Modeling of Joining Processes" in this Volume.

This Section concludes with an article on corrosion of weldments. Corrosion is a particularly difficult fabrication problem because it often appears only after significant in-service time has elapsed. A thorough understanding of corrosion phenomena can provide the designer of weldments with insight towards its avoidance. Additional information on this topic is available in *Corrosion,* Volume 13 of the *ASM Handbook.*

Thermal Spray Coatings

Herbert Herman, State University of New York, Robert A. Sulit, Sulit Engineering

Introduction

THERMAL SPRAY COATINGS (TSCs) are surface coatings engineered to provide wear-, erosion-, abrasion-, and corrosion-resistant coatings for original equipment manufacture and for the repair and upgrading of in-service equipment. In general, TSCs can be applied to a range of substrate materials, including metals, ceramics, plastics, and polymer composites. Thermal spray technology evolved and developed within the welding field because of similarities in the electrical and combustion equipment and the feedstock materials used in thermal spray with those used in welding and brazing. Accordingly, thermal spraying is classified as an allied welding process.

This article presents a general overview of the five thermal spray processes and the specific flame and arc spray processes used to preserve large steel components and structures. Also included are a TSC selection guide and an industrial process procedure guide for applying aluminum and zinc TSCs onto steel. Additional information on thermal spray processes is available in the article "Hardfacing, Weld Cladding, and Dissimilar Metal Joining" in this Volume.

Thermal spraying processes deposit finely divided metallic or nonmetallic feedstock surfing materials (wire, cored wires, ceramic rods, and powders) are heated to their plastic or molten state by combustion (flame, high-velocity oxygen fuel, and detonation) and by electric (arc and plasma) processes. The materials in their plastic or molten state are accelerated toward the substrate. The particles or droplets strike the surface, flatten, and form thin platelets (splats) that conform, adhere, and interlock with the roughened surface irregularities and with each other. As the sprayed particles impinge on the substrate, they cool and build up, particle by particle, into a lamellar-structured coating on the substrate. In general, the substrate temperature can be kept at 200 °C (390 °F) or below to prevent metallurgical changes in the substrate material.

The properties of the TSC depend on such factors as its porosity, the cohesion between particles, adhesion to the substrate (including interface integrity), and the chemistry of the coating material. The particles bond to the substrate mechanically (the primary mechanism), metallurgically, or chemically. Particle impact velocity, particle size, substrate roughness, particle temperature, and substrate temperature influence bond strength. Figure 1 shows the lamellar structure of particle splats, oxide inclusions, and unmelted particles in a typical TSC cross section.

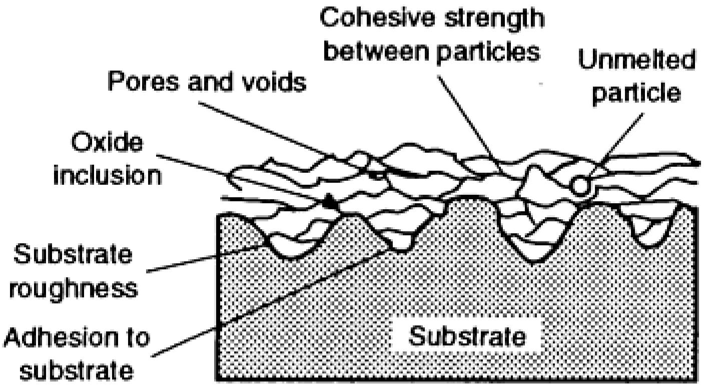

FIG. 1 SCHEMATIC SHOWING CROSS-SECTIONAL VIEW OF TSC LAMELLAR COMPONENTS THAT AFFECT COATING INTEGRITY

Aluminum and zinc TSCs have a long history of corrosion protection in structural steel work:

- BUILDINGS
- BRIDGES
- TOWERS
- RADIO AND TV ANTENNA MASTS
- STEEL GANTRY STRUCTURES
- HIGH-POWER SEARCH RADAR AERIALS
- OVERHEAD WALKWAYS
- RAILROAD OVERHEAD LINE SUPPORT COLUMNS
- ELECTRIFICATION MASTS
- TOWER CRANES
- TRAFFIC ISLAND POSTS
- STREET AND BRIDGE RAILINGS

On a smaller scale, aluminum-zinc TSCs have been successfully used to protect lawn furniture in corrosive seacoast environments.

Thermal Spray Coatings

Herbert Herman, State University of New York, Robert A. Sulit, Sulit Engineering

---

Thermal Spray Coating Processes

Thermal spray processes differ in the feedstock materials used, the method of heating applied, and the method of propelling the material toward the substrate. The primary TSC feedstock materials can be classified as powder, wire, cored wire, and ceramic rod. Thermal spray processes can be categorized into two basic groups (combustion or electrical) according to the method of heat generation (Table 1).

TABLE 1 CLASSIFICATION OF THERMAL SPRAY PROCESSES RELATIVE TO METHOD OF HEAT GENERATION AND FEEDSTOCK FORM

FEEDSTOCK FORM	THERMAL SPRAY PROCESS(A)					
	COMBUSTION			ELECTRICAL		
				ARC	PLASMA	
	FLAME	**HVOF(B)**	**DETONATION**		**DIRECT CURRENT**	COUPLED INDUCTION
POWDER	X	X	X	. . .	X	**X**
WIRE	X	. . .	. . .	X	. . .	**. . .**
CORED WIRE	X	. . .	. . .	X	. . .	**. . .**
CERAMIC ROD	**X**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**

(A) X, APPLICABLE.
(B) HIGH-VELOCITY OXYFUEL

Table 2 gives typical characteristics of these thermal spray processes. The combustion flame and the electric arc processes are the main ones used for preserving steel. Recommended feedstock requirements, area coverage, and spray rates for flame and arc spraying are given in the section "Typical Process Parameters" in this article.

TABLE 2 TYPICAL PARAMETERS OF SELECTED TSC PROCESSES

PROCESS	**FEEDSTOCK**	**OPERATION**	**GAS FLOW**		**FLAME OR ARC TEMPERATURE**		**PARTICLE IMPACT VELOCITY**		SPRAY RATES	
			M^3/H	**FT^3/H**	**°C**	**°F**	**M/S**	**FT/S**	**KG/H**	LB/H
GROUP I (COMBUSTION)										
SUBSONIC	FLAME POWDER	POWDER FEEDSTOCK FED INTO OXYFUEL FLAME	10	350	2200	3990	30	100	$\frac{1}{2}$ - 10	**1.1-22**
	FLAME WIRE, ROD, AND CORD	FEEDSTOCK FED INTO OXYFUEL FLAME	70	2500	2800	5070	180	590	$\frac{1}{2}$ - 10	**1.1-22**
SUPERSONIC	HIGH-VELOCITY OXYGEN FUEL POWDER	POWDER FED INTO SUPERSONIC VELOCITY COMBUSTION FLAME	110	3900	2600	4710	900	2950	2-15	**4.4-33**
	DETONATION GUN	POWDER AND EXPLOSIVE GAS CYCLICALLY FED EXPLOSION	. . .	. . .	2600	4710	900	2950	. . .	**. . .**

		CHAMBER								
GROUP II (ELECTRICAL)										
ARC	WIRE	TWO CONSUMABLE ELECTRODES FED INTO ATOMIZING AIR/GAS STREAM	70	2500	5000	9300	240	790	2(A)	**4.4(B)**
PLASMA	**POWDER**	**POWDER FED INTO PLASMA TORCH WITH TEMPERATURES TO 8300 °F (15,000 °F)**	**4**	**140**	**5000**	**9300**	**240**	**790**	**0.1(A)**	**0.22(B)**

(A)	KG/H/KW.
(B)	LB/H/KW

Flame spraying is the simplest and most versatile process. All five feedstock forms can be sprayed with subsonic flame sp and most versatile process. All five feedstock forms can be sprayed with subsonic flame spray. The feed material is melted by an oxyfuel gas flame. In flame powder spraying (Fig. 2), powder particles are heated to 2200 °C (3990 °F), and impact the substrate at 30 m/s (100 ft/s). In flame wire spraying (Fig. 3), melted and vaporized wire particles are heated to 2800 °C (5070 °F), and impact the substrate at 180 m/s (590 ft/s). The higher gas flow used in flame wire spraying gives the droplets greater impact energy. The impact particle velocity in flame powder spraying is the lowest, resulting in reduced adhesion and increased porosity relative to the other thermal spray processes. Spray rates are usually in the range of 0.5 to 10 kg/h (1.1 to 22 lb/h) for both flame wire and flame powder spraying.

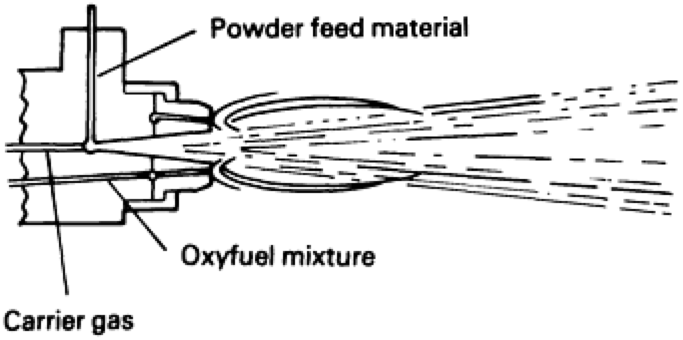

FIG. 2 CROSS-SECTIONAL VIEW OF A FLAME POWDER SPRAYING SYSTEM SHOWING POWDER FEED MATERIAL BEING TRANSPORTED BY THE CARRIER GAS AND THEN MELTED BY THE OXYFUEL MIXTURE

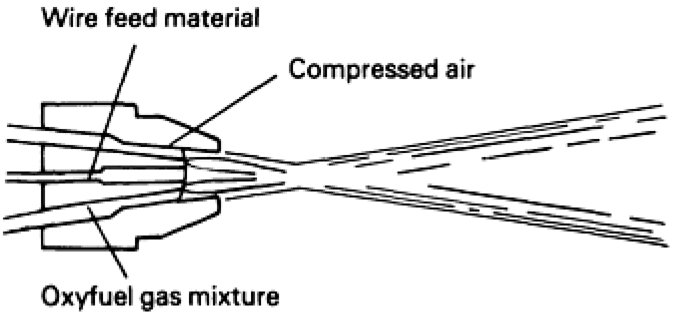

FIG. 3 ATOMIZATION OF WIRE FEEDSTOCK FROM THE NOZZLE OF A FLAME WIRE SPRAYING GUN

Detonation Gun. In the detonation gun (D-gun) process, oxygen and fuel gases are mixed with powder and cyclically injected into a constricted shotgunlike barrel and ignited. A spark ignites the mixture, producing a supersonic detonation wave that heats the powder particles to 2600 °C (4710 °F) and propels them at 900 m/s (2950 ft/s) onto the substrate.

The HVOF gun (Fig. 4) operates continuously and uses higher oxidant and fuel gas pressures, resulting in higher particle impact velocities, thus producing extremely dense, well-bonded TSCs.

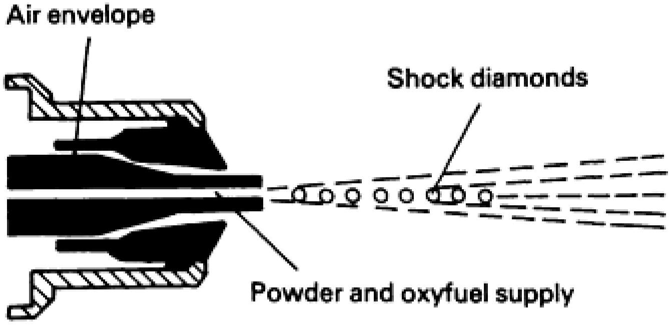

FIG. 4 CROSS-SECTIONAL VIEW SHOWING THE OUTPUT OF AN HVOF SPRAYING SYSTEM

Detonation Gun Versus HVOF Gun. High-velocity oxyfuel equipment is portable and can be transported to the job site, while the detonation gun equipment must be operated in fixed facilities. The HVOF gun can spray powder feedstock materials at the rate of 2 to 15 kg/h (4.4 to 33 lb/h).

Arc Spraying. Two oppositely charged consumable wires are fed into the gun (Fig. 5). Where they emerge from the nozzle, at their point of intersection, an arc is formed, melting the wire at 5000 °C (9030 °F). The molten metal is atomized and propelled by an air or gas stream, impacting the substrate at 240 m/s (785 ft/s). Metal and alloy wires and powders in metallic sheaths (cored wire) can be sprayed at rates of over 25 kg/h (55 lb/h).

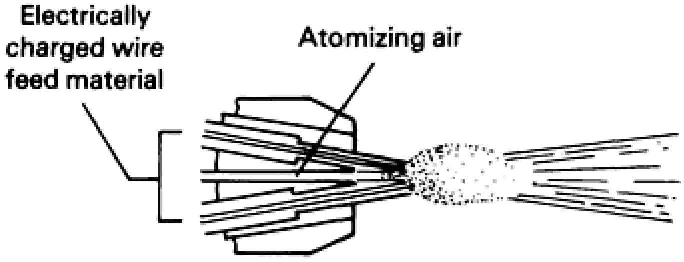

FIG. 5 SPRAY PATTERN GENERATED BY TWO ELECTRICALLY CHARGED WIRES MELTED AT THE NOZZLE OF AN ARC SPRAYING GUN

Plasma spraying uses a nontransferred arc gun to create the plasma jet that melts and propels the powder to the substrate (Fig. 6). The nontransferred plasma arc is contained in the gun. The arc is maintained between a thoriated tungsten cathode and a constricting nozzle anode (copper). An inert or reducing gas under pressure enters the annular space between the anode and the cathode, where it becomes partially ionized. The gas emerges from the nozzle as a high-temperature (up to 17,000 °C, or 30,600 °F) high-velocity jet. Powder feedstock material is injected into the hot gas stream, where it melts or softens, then impacts the substrate at 250 m/s (790 ft/s). Spray rates are 0.1 kg/h/kW (0.2 lb/h/kW). The usual power supplies range from 15 to 80 kW. Higher-power plasma guns, operating up to 200 kW can, in principle, spray ceramics at 25 to 50 kg/h (55 to 110 lb/h) and metals at an even higher rate.

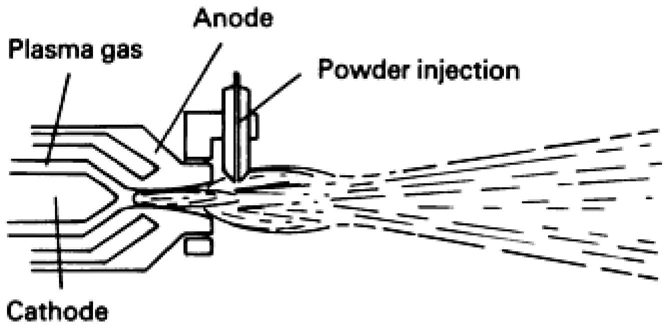

FIG. 6 SCHEMATIC OF A PLASMA ARC POWDER SPRAY SYSTEM SHOWING ROUTING OF PLASMA GAS AND POWDER MATERIAL AT THE OUTPUT NOZZLE

Thermal Spray Coatings

Herbert Herman, State University of New York, Robert A. Sulit, Sulit Engineering

Selection of TSCs to Preserve Integrity of Steels

Thermal spray coatings are extensively used for the corrosion protection of steel and iron in a wide range of environments. The long-term effectiveness (over 20 years) in rural, industrial, and marine environments is well documented. The corrosion tests carried out by the American Welding Society (Ref 1) and the 34-year marine-atmosphere performance report of the LaQue Center for Corrosion Technology (Ref 2) confirm the effectiveness of flame-sprayed aluminum and zinc coatings over long periods of time in a wide range of hostile environments. The British Standards Institution code of practice for the corrosion protection of steel specifies that only TSCs are able to give more than 20 years of protection before maintenance is necessary in the 19 industrial and marine environments considered. Only sealed, sprayed aluminum or zinc gives such protection in seawater immersion or splash zones (Ref 3). The Canadian Standards Association tabulates the life expectancy at more than 40 years for aluminum and zinc TSCs in various exposure environments (Ref 4).

The selection of a TSC depends on the service environment, the desired service life, operating duty cycle, and the maintenance and repair support to be provided during the life cycle. Tables 3 and 4 give current service life information for aluminum and zinc TSCs in various service environments. Figures 7 and 8 plot the required thickness specifications given in Tables 3 and 4, respectively. The service life estimates for 85/15 wt% Zn/Al alloy and the 90/10 vol% aluminum metal-matrix composites (MMCs) TSCs introduced in the late 1970s and 1980s, respectively, are based on accelerated laboratory tests and service applications through 1992. In a marine environment, powder TSC with higher aluminum content exhibits improved corrosion resistance (Ref 2). Where resistance to wear, abrasion, or both, is required in addition to corrosion protection, 90/10-MMC TSCs should be considered. The 90/10 MMC wire is composed of 90 vol% Al + 10 vol% Al_2O_3 (8 to 10 μm, or 320 to 400 μin., diameter).

TABLE 3 SERVICE LIFE ESTIMATES OF 85/15 ZN/AL TSCS IN SELECTED CORROSIVE ENVIRONMENTS

TYPE OF EXPOSURE	COATING THICKNESS REQUIRED FOR INDICATED SERVICE LIFE							
	5-10 YEARS		**10-20 YEARS**		**20-40 YEARS**		>40 YEARS	
	μm	**in.**	**μm**	**in.**	**μm**	**in.**	**μm**	in.
RURAL ATMOSPHERE	. . .	. . .	75-125	0.003-0.005	125-175	0.005-0.007	250-300	**0.010-0.012**
INDUSTRIAL ATMOSPHERE	. . .	. . .	150-200	0.006-0.008	300-375	0.012-0.015	350-400	**0.014-0.016**

MARINE ATMOSPHERE	. . .	. . .	250-300	0.010-0.012	300-375	0.012-0.015	350-400	**0.014-0.016**
FRESHWATER IMMERSION	150-200	0.006-0.008	250-350	0.010-0.014	300-375	0.012-0.015	. . .	**. . .**
SALTWATER IMMERSION	**250-300**	**0.010-0.012**	**350-400**	**0.014-0.016**	**. . .**	**. . .**	**. . .**	**. . .**

TABLE 4 SERVICE LIFE ESTIMATES OF ALUMINUM AND 90/10 ALUMINUM METAL-MATRIX COMPOSITE TSCS

TYPE OF EXPOSURE	COATING THICKNESS REQUIRED FOR INDICATED SERVICE LIFE(A)							
	5-10 YEARS		**10-20 YEARS**		**20-40 YEARS**		>40 YEARS	
	µm	**in.**	**µm**	**in.**	**µm**	**in.**	**µm**	in.
RURAL ATMOSPHERE	. . .	. . .	. . .	. . .	150-200	0.006-0.008	. . .	**. . .**
INDUSTRIAL ATMOSPHERE	. . .	. . .	150-200	0.006-0.008	250-300	0.010-0.012	250-375	**0.010-0.015**
MARINE ATMOSPHERE	150-200	0.006-0.008	200-250	0.008-0.010	250-300	0.010-0.012	250-375	**0.010-0.015**
FRESHWATER IMMERSION	150-200	0.006-0.008	200-250	0.008-0.010	250-300	0.010-0.012	. . .	**. . .**
SALTWATER IMMERSION	200-250	0.008-0.010	250-300	0.010-0.012	300-350	0.012-0.014	. . .	**. . .**
HIGH-TEMPERATURE (100-540 °C, OR 210-1000 °F)	150-200	0.006-0.008	200-250	0.008-0.010	250-300	0.010-0.012	. . .	**. . .**
WEAR, ABRASION, EROSION, AND IMPACT (90/10 MMC PREFERRED)	**150-200**	**0.006-0.008**	**250-300**	**0.010-0.012**	**. . .**	**. . .**	**. . .**	**. . .**

(A) WITH EXCEPTION OF WEAR, ABRASION, EROSION, AND IMPACT PROPERTIES,DATA FOR ALUMINUM AND 90/10 ALUMINUM MMC TSC ARE IDENTICAL.

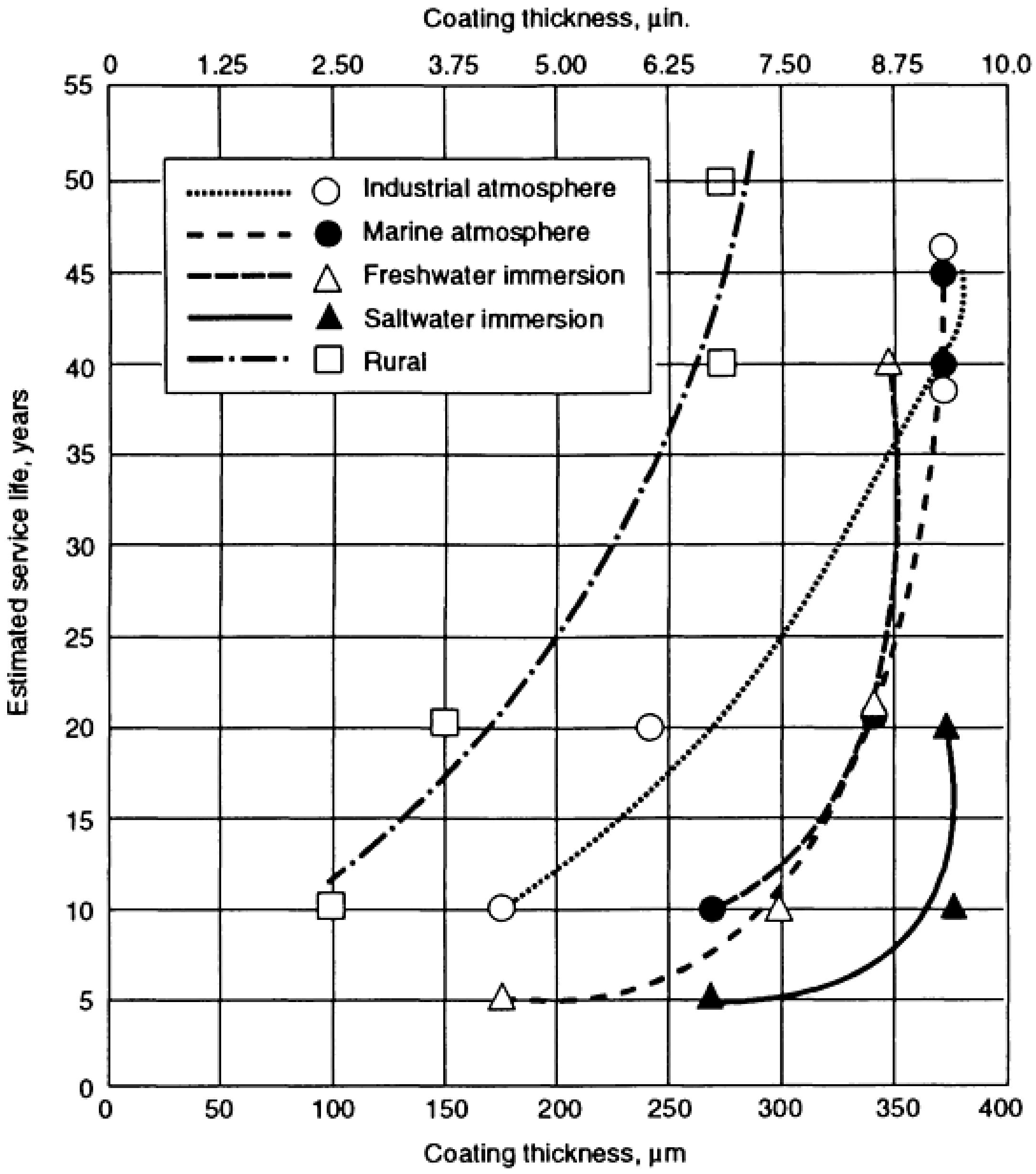

FIG. 7 PLOT OF SERVICE LIFE VERSUS COATING THICKNESS AS A FUNCTION OF SELECTED ENVIRONMENTS FOR AN 85/15 ZN/AL ALLOY WITH A TSC

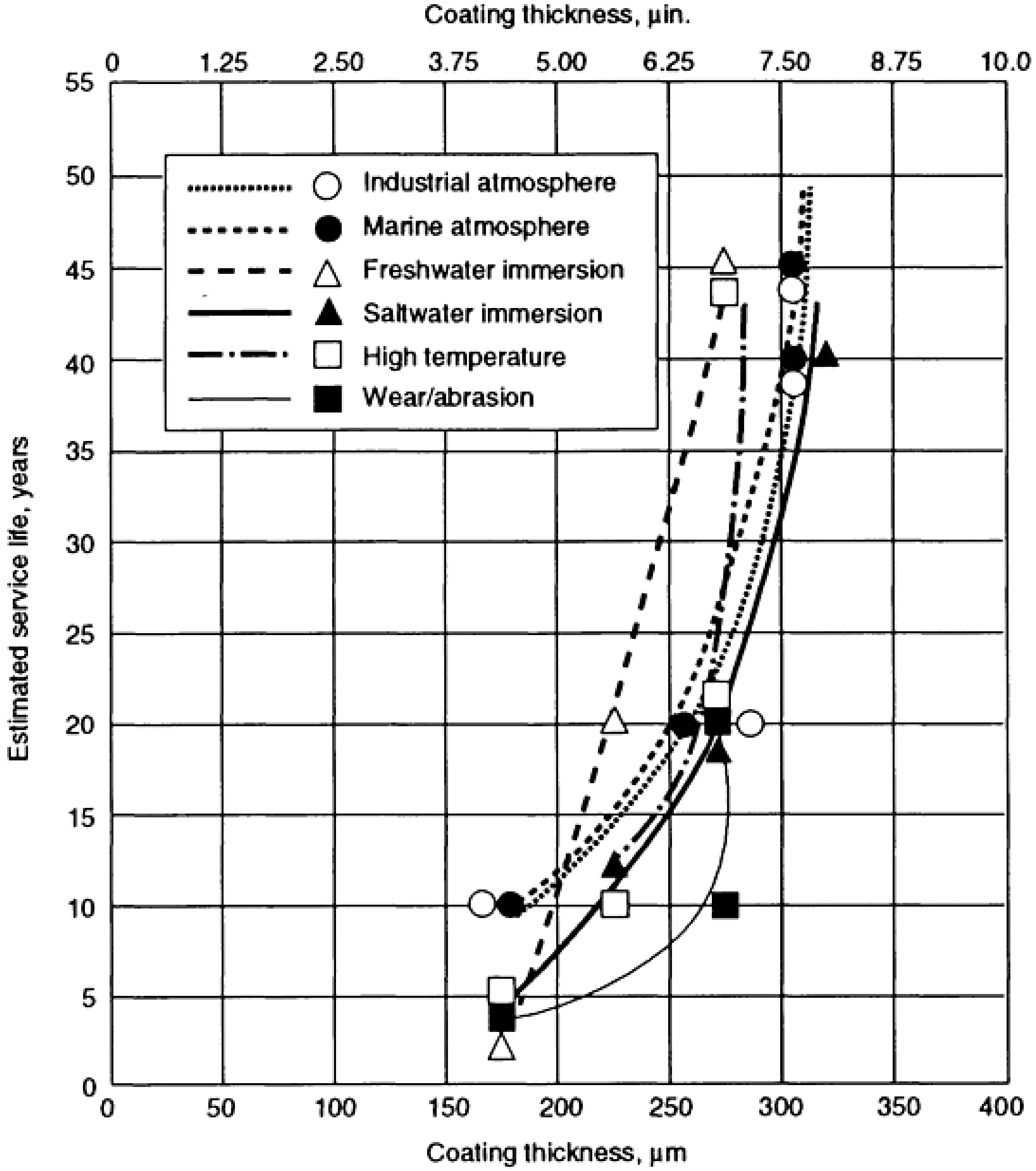

FIG. 8 PLOT OF SERVICE LIFE VERSUS COATING THICKNESS AS A FUNCTION OF SELECTED ENVIRONMENTS FOR A 90/10 ALUMINUM METAL-MATRIX COMPOSITE WITH A TSC

Galvanic Properties. Aluminum and zinc and their alloys and composites provide broad atmospheric protection. Aluminum and zinc are anodic to steel and protect the ferrous substrate in electrolytic solutions. When applied in a nonthrough porosity thickness, they provide both barrier and cathodic protection. When cut through to the steel or when applied too thinly to role out through porosity, these TSCs provide galvanic protection to the steel. Aluminum corrodes less rapidly than zinc in highly acidic conditions, while zinc performs better in alkaline conditions. Aluminum TSCs immediately oxidize to form a loosely adherent Al_2O_3 protective film (as compared to the tightly adherent protective film on stainless steels) that prevent further oxidation. Thus, there is no advantage to applying aluminum TSCs greater than the non-through-porosity thickness of 150 to 200 µm (0.006 to 0.008 in.) unless wear or abrasion resistance is also required in addition to corrosion resistance. Aluminum TSCs have greater wear, abrasion, and erosion resistance than zinc TSCs.

Zinc alloyed with aluminum forms an effective corrosion-resistant coating, combining the attributes of both elements. The greater electrochemical activity of zinc provides greater galvanic protection than aluminum. Aluminum, with its lower electrochemical activity and a loosely adherent aluminum oxide film, provides long-term protection even through porous coatings and gives better wear, abrasion, and erosion resistance than zinc TSCs.

Aluminum and aluminum composite TSCs can be used where the temperature is greater than 200 °C (390 °F). Aluminum composite TSCs are used when wear, abrasion, and erosion resistance are required beyond that provided by aluminum and zinc TSCs.

Thermite Sparking Hazard (Ref 5, 6). Thermite sparking is caused by the reaction of rusted steel and aluminum (in the form of a finely divided smear) when this combustible mix is ignited by an impact. Aluminum smears may be generated on rusting steel by striking or dragging steel components and tools over bare aluminum surfaces or vice versa. Bare aluminum and bare aluminum TSCs should be avoided whenever there is a thermite sparking hazard. Situations where bare aluminum or bare aluminum TSCs and rusted steel will be near a combustion or explosion source can occur on a regular basis and should be avoided. There appears to be little risk of thermite sparking as a result of the impact of rusted steel on a sealed or painted aluminum, aluminum alloy, or aluminum composite TSC surface.

Porosity is an inherent feature of TSCs. When applied to nonthrough porosity thickness (nominally about 175 μm, or 0.007 in.) for flame-sprayed aluminum and 90/10 MMC, these TSCs will protect the substrate steel. When applied to a less than nonthrough porosity thickness, the TSC will retard substrate corrosion because of galvanic protection. The porosity of the TSC is a function of the feedstock material, the application method, and the spray parameters.

Role of Sealants and Topcoats. Thermal spray coatings are applied in a single application. They emit no volatile organic compounds, need no drying time, and provide galvanic protection to exposed steel substrates. The natural surface texture of TSCs provides an excellent base for paint and thermoplastic or thermosetting powder sealers as well as paint and powder-coating topcoats.

It is preferable to seal TSCs. The sealer fills the pores of the TSC, smooths the sprayed surface, and improves appearance and service life. Sealers also simplify maintenance, which is generally only the reapplication of the sealer. Sealers should be applied immediately after applying the TSC.

TSC paint sealers should have low viscosity to flow and should be absorbed into the natural pores of the TSC. The pigment particle size of colored paint sealers must be small enough to freely flow into the pores of the TSC, nominally a 5 fineness grind per ASTM D 1210. Thermoplastic and thermosetting powder sealers have fine particles but should be electrostatically applied to maximize penetration into the pores of the TSC. Furthermore, thermal sprayed components must be sufficiently heated to maximize outgassing of the pores prior to electrostatic application of the powder coating.

Thermal spray coatings should be topcoated only when:

- THE ENVIRONMENT IS VERY ACID OR VERY ALKALINE (THAT IS, WHEN THE PH IS OUTSIDE THE RANGE OF 5 TO 12 FOR ZINC TSCS, OR 4 TO 9 FOR ALUMINUM AND 90/10 MMC TSCS).
- THE METAL IS SUBJECT TO DIRECT ATTACK BY SPECIFIC CHEMICALS.
- THE REQUIRED DECORATIVE FINISH CAN BE OBTAINED ONLY WITH PAINT.
- ADDITIONAL WEAR, ABRASION, AND EROSION RESISTANCE ARE REQUIRED.

One or two coats of paint may be sufficient except in abnormally aggressive environments. Sealed TSCs are preferable to painted TSCs. Paints usually have a longer life on TSCs than on bare steel, because the TSC reduces or prevents rusting and pitting of the steel.

Zinc TSCs complement hot-dip galvanizing and should be considered when fabricated goods are excessively large or cannot be hot-dip galvanized. Zinc TSCs should also be considered for repairing galvanized coatings damaged during fabrication (for example, welding and cutting) and for maintenance recoating.

Zinc spraying can restore corrosion protection to areas where galvanizing is inadvertently removed during fabrication and assembly. This can occur on the threaded end of electrical conduits or along the welded seams inside galvanized barrels and drums. Zinc spray is particularly advantageous in such cases because it ensures the uniformity and reproducibility of the galvanized coating thickness.

Wellhead valve assemblies for offshore use have been coated for salt atmosphere protection since the 1950s. Aluminum TSCs have been used for flare stacks. Aluminum and zinc TSCs have been used for the external protection of oil and propane gas storage tanks. The interiors of steel coal rail cars have been sprayed with aluminum for sulfuric acid corrosion protection and with aluminum composite for both corrosion and abrasion protection. The exteriors have been sprayed with zinc for atmospheric-corrosion protection. Thermal spray coatings have been used to protect pipelines against many environments. Pile couplings, valves, manhole covers, industrial gas bottles, and other small industrial items are candidates for protection by TSCs.

Zinc TSCs are used to protect potable water pipelines and storage tanks (Ref 7). Aluminum and zinc TSCs are used in sluice and canal lock gates in irrigation systems and in shipping canal lock gates. Once coated, these components require virtually no maintenance for decades.

Sealed zinc coatings improve the resistance of steel bridgework and railings to marine corrosion and deicing salts. Rebars in reinforced concrete can be zinc sprayed to retard corrosion. Reinforced concrete bridges and highways, especially those in marine and freezing environments where deicing salts are used, commonly suffer from chloride intrusion into the concrete, leading to rebar corrosion and concrete spalling. Zinc TSCs are sometimes used for rebar protection before concrete is poured. Zinc TSCs are sprayed directly onto bridge concrete substructures to provide a cathodic-protection coating or to serve as secondary anodes connected to an impressed-current cathodic-protection system.

In marine applications, the structural areas and components of ships are preserved with aluminum and zinc TSCs. The U.S. Navy routinely uses aluminum TSCs in new ship construction, in the maintenance of ship structures, and for a wide range of shipboard components, especially those in topside and wet spaces. The British, Australian, and New Zealand Navies use a duplex zinc (base) and aluminum (top) TSC system. Commercial ships and barges have used TSCs to preserve ship superstructures and various topside and interior components.

Sealed TSCs have a longer service life than unsealed TSCs, are easier to clean and maintain, and provide both barrier and galvanic protection. Sealer paints and powder coatings must be chemically compatible with the TSC material, the service environment, and the intermediate/topcoat paint. Sealers must be suitably thinned and have a fine pigment grind to effectively penetrate the TSC. As indicated above, sealed TSCs are preferable to painted TSCs. Descriptions and specifications of sealer and intermediate/topcoat paints are given in Ref 3 and 8.

Future Outlook. The suitability of a particular paint or thermoplastic/thermosetting powder (applied electrostatically or by thermal spray) must be evaluated when the coating is designed and specified. Such an evaluation is necessary in order to get the sealer/topcoat paint or powder system that best meets the service and life cycle cost (LCC) requirements compliant with the evolving environmental and safety regulations. Specific paint or powder systems can be ranked and validated with accelerated exposure tests simulating the service environment and duty cycle. The penetrability of the sealer into the TSC can be validated by metallographic analysis and the service life ranked by accelerated weathering tests.

Thermal spray coatings are supplementing and replacing paint coating systems because of their predictable service life, increased effectiveness, and lower life cycle costs. The LCC includes the initial installation cost plus the repair and maintenance costs incurred during the service life of the coating system. The installation cost of TSC systems may range up to one and one-half times that of high-performance paint systems, but the LCCs could be less than three-fourths the paint system LCC. Comparative TSC and painting system LCC may be estimated with the economic models in Ref 3 and 8.

References cited in this section

1. "CORROSION TESTS OF FLAME-SPRAYED COATED STEEL, 19-YEAR REPORT," C2.14-74, AMERICAN WELDING SOCIETY, 1974
2. R.M. KAIN AND E.A. BAKER, MARINE ATMOSPHERIC CORROSION MUSEUM REPORT ON THE PERFORMANCE OF THERMAL SPRAY COATINGS ON STEEL, *TESTING OF METALLIC AND INORGANIC COATINGS,* STP 947, AMERICAN SOCIETY FOR TESTING AND MATERIALS, 1987
3. "CODE OF PRACTICE FOR PROTECTIVE COATING OF IRON AND STEEL STRUCTURES AGAINST CORROSION," BS 5493, BRITISH STANDARDS INSTITUTION, 1977; AVAILABLE FROM THE AMERICAN NATIONAL STANDARDS INSTITUTE
4. "SPRAYED METAL COATINGS FOR ATMOSPHERIC CORROSION PROTECTION," G 189-1966,

CANADIAN STANDARDS ASSOCIATION, 1966; REAFFIRMED 1980

5. ROSENBURGH *ET AL., ALUMINUM OFFSHORE--SAFETY ANALYSIS,* PUBLICATION 85-3208, VERITEC, 1985
6. THERMITE SPARKING, *ALUMINUM DESIGN GUIDE,* VOL 2, *DESIGN GUIDE ON THE USE OF ALUMINUM IN OFFSHORE STRUCTURES,* WIMPEY OFFSHORE, LONDON, SECTION 4.5
7. "PAINTING WATER STORAGE TANKS," ANSI/AWS D102-78, AMERICAN WELDING SOCIETY, 1978
8. *STEEL STRUCTURES PAINTING MANUAL,* STEEL STRUCTURES PAINTING COUNCIL

Thermal Spray Coatings

Herbert Herman, State University of New York, Robert A. Sulit, Sulit Engineering

---

Industrial Process Instruction for Applying TSCs on Steel

An industrial process instruction is an engineering procedure detailing the necessary equipment, consumables, method (step-by-step procedure), trained/certified personnel, and records required to accomplish the objective of that process with the requisite quality control, safety, and environmental compliance in a timely and cost-effective manner. The scope of an industrial process instruction for applying TSCs on steel should include (Ref 9):

- *DEFINITION* OF TERMS UNIQUE AND CRITICAL TO THE SAFE AND SUCCESSFUL ACCOMPLISHMENT OF THE PROCESS
- *SAFETY AND ENVIRONMENTAL COMPLIANCE* REQUIRED FOR OPERATOR, OBSERVERS, AND PASSERSBY. FOR THERMAL SPRAYING, AIRBORNE ALUMINUM AND ZINC DUST COULD BE AN EXPLOSIVE HAZARD IF NOT PROPERLY HANDLED. ALUMINUM AND ZINC FUMES PRODUCED IN THERMAL SPRAYING AND IN BLASTING DEBRIS FROM SURFACE PREPARATION, ESPECIALLY IF THERE WAS OLD LEAD-BEATING PAINT, REQUIRE PERSONNEL PROTECTION, CONTAINMENT, AND THE REQUISITE WASTE DISPOSAL
- *JOB AND CONTRACT DESCRIPTION* DETAILING THE FOLLOWING: THE THERMAL SPRAY WORK AREA BOUNDARY; WHAT IS AND WHAT IS NOT TO BE THERMAL SPRAYED; THE THERMAL SPRAY FEEDSTOCK AND TSC THICKNESSES; SEALING/TOPCOATING MATERIALS; AND THE INSPECTION AND ACCEPTANCE REQUIREMENTS
- *MATERIALS* REQUIRED, INCLUDING THERMAL SPRAY FEEDSTOCK (BASED ON THE NOMINAL AREA COVERAGE AND SPRAY RATES FOR THE SPECIFIED PROCESS EQUIPMENT), ABRASIVE BLASTING MEDIA FOR ROUGH AND FINAL ANCHOR-TOOTH BLASTING, SEALER AND TOPCOATING PAINTS OR THERMOPLASTIC/THERMOSETTING POWDER COATINGS, QUALITY CONTROL CONSUMABLES, AND ENERGY REQUIREMENTS (FUEL GASES, INERT GASES, AND OXYGEN)
- *EQUIPMENT FOR SURFACE PREPARATION AND THERMAL SPRAYING* TO INCLUDE SPRAY GUN QUALIFICATIONS, AIR COMPRESSOR, AND AIR DRYERS
- *QUALITY CONTROL EQUIPMENT* TO DETERMINE THE ACCEPTABILITY OF SURFACE PREPARATION (WHITE AND NEAR-WHITE METAL FINISH, BLASTING MEDIA CLEANLINESS, AND ANCHOR-TOOTH DEPTH) AND TSC APPLICATION (SUBSTRATE TEMPERATURE, AIR TEMPERATURE, DEW POINT AND HUMIDITY, TSC THICKNESS, AND TSC BEND TEST)
- *APPLICATION PROCESS:* SURFACE PREPARATION (INCLUDES NEW STEEL, OLD STEEL, MASKING, DEGREASING, AND SURFACE PROFILE WITH QUALITY CONTROL CHECK POINTS; POST BLASTING SUBSTRATE CONDITION AND THERMAL SPRAYING PERIOD; THERMAL SPRAYING (INCLUDES EQUIPMENT SETUP AND BEND TEST; PLANNING AND EXECUTING THERMAL SPRAY SEQUENCE AND CONTROLLING OVERSPRAY WITH TWO QUALITY CONTROL POINTS)

- *PROCEDURES* FOR SPRAYING EDGES OF MATERIALS OR INTO CAVITIES (THAT IS, HOLES)
- *SEALING AND TOPCOATING* WITH QUALITY CONTROL CONTROL CHECK POINTS
- *MAINTENANCE AND REPAIR OF TSCS* BASED ON DEGREE OF DAMAGE AND WEAR AS RELATED TO EXPOSURE OF THE UNDERLYING COATING OR SUBSTRATE STEEL AND THE SIZE OF THE DEGRADED AREA
- *DEBRIS* CONTAINMENT AND CONTROL
- *RECORDS,* INCLUDING A JOB CONTROL RECORD SUMMARIZING THE TSC JOB REQUIREMENTS, OPERATOR QUALIFICATIONS, AND A CHECK LIST OF THE NINE QUALITY CONTROL CHECK POINTS
- *OPERATOR TEST AND QUALIFICATIONS* PER ANSI/AWS C2.16-92 ("GUIDE FOR THERMAL-SPRAY OPERATOR QUALIFICATION")

Reference cited in this section

9. "GUIDE FOR THE PROTECTION OF STEEL WITH THERMAL SPRAY COATINGS OF ALUMINUM AND ZINC AND THEIR ALLOYS AND COMPOSITES," ANSI/AWS C2.18-X, AMERICAN WELDING SOCIETY, IN PREPARATION

Thermal Spray Coatings

Herbert Herman, State University of New York, Robert A. Sulit, Sulit Engineering

---

Typical Process Parameters (Ref 9)

Table 5 gives the nominal thermal spray deposit efficiency and feedstock material required per unit area or unit thickness for representative materials sprayed onto a flat plate. Deposit efficiency is the percentage by weight of sprayed materials adhering to a large flat plate. Table 6 gives the typical spray rates and area coverage for flame and arc spraying on a flat plate. Additional feedstock will be required to spray complex geometrical shapes. Specific values vary with the shop or field job site, thermal spray equipment, spray parameters, process used, and the geometrical shape of the components being sprayed. The data in Tables 5 and 6 are technical data supplied by various equipment manufacturers and are given for comparison purposes and relative ranking; they are not intended for cost estimating.

TABLE 5 DEPOSITION RATES FOR SELECTED FEEDSTOCK USED IN FLAME AND ARC TSC APPLICATIONS

PARAMETER	FEEDSTOCK MATERIAL					
	WIRE				POWDER	
	Al	**Zn**	**85/15 Zn/Al**	**90/10 AL MMC**	**Al**	Zn
FLAME SPRAY						
DEPOSIT EFFICIENCY$^{(A)}$, %	80-85	65-70	80-85	80-85	90	**85**
MATERIAL REQUIRED, KG/M^2/μM	0.0024-0.0026	0.0098-0.0105	0.0066-0.0070	0.0024-0.0026	0.0027	**0.0076**
ARC SPRAY						
DEPOSIT EFFICIENCY$^{(A)}$, %	75-80	65-70	72-78	75-80	$^{(B)}$	$^{(B)}$
MATERIAL REQUIRED, KG/M^2/μM	**0.0029-0.0031**	**0.009-0.010**	**0.0089-0.0096**	**0.0029-0.0031**	$^{(B)}$	$^{(B)}$

(A) DATA FOR FLAT PLATE.

(B) NOT APPLICABLE

TABLE 6 TYPICAL DEPOSITION RATES (SPRAY RATE AND AREA COVERAGE) FOR FLAME SPRAYING AND ARC SPRAYING WITH SELECTED FEEDSTOCK ON A FLAT PLATE

APPLICATION METHOD	DEPOSITION RATES PER INDICATED FEEDSTOCK							
	AL		**ZN**		**85/15 ZN/AL**		90/10 AL MMC	
	SPRAY RATE, KG/H	**COVERAGE, M^2/H/μM**	**SPRAY RATE, KG/H**	**COVERAGE, M^2/H/μM**	**SPRAY RATE, KG/H**	**COVERAGE, M^2/H/μM**	**SPRAY RATE, KG/H**	COVERAGE, M^2/H/μM
FLAME SPRAYING								
WITH WIRE DIAMETER: 2.38 MM (0.0937 IN.)	2.5	1.44	9.1	2.05	8.2	1.96	2.5	**1.44**
WITH WIRE DIAMETER: 3.18 MM (0.125 IN.)	5.4	3.18	20	3.33	18	3.33	5.4	**3.18**
WITH WIRE DIAMETER: 4.76 MM (0.187 IN.)	7.3	4.22	30	4.81	26	4.81	7.3	**4.22**
WITH POWDER	6.8	100	14	80	. . .	. . .	. . .	**. . .**
ARC SPRAYING (PER 100 A)	**2.7**	**1.39**	**12**	**2.85**	**10**	**2.66**	**2.7**	**1.39**

Thermal spray coating applicators should develop their own shop and field deposit-efficiency data and production planning factors for their specific equipment and their method of operation. Past production records should provide a good information base for job cost estimating, production scheduling, and cost control.

Reference cited in this section

9. "GUIDE FOR THE PROTECTION OF STEEL WITH THERMAL SPRAY COATINGS OF ALUMINUM AND ZINC AND THEIR ALLOYS AND COMPOSITES," ANSI/AWS C2.18-X, AMERICAN WELDING SOCIETY, IN PREPARATION

Thermal Spray Coatings

Herbert Herman, State University of New York, Robert A. Sulit, Sulit Engineering

---

References

1. "CORROSION TESTS OF FLAME-SPRAYED COATED STEEL, 19-YEAR REPORT," C2.14-74, AMERICAN WELDING SOCIETY, 1974
2. R.M. KAIN AND E.A. BAKER, MARINE ATMOSPHERIC CORROSION MUSEUM REPORT ON THE PERFORMANCE OF THERMAL SPRAY COATINGS ON STEEL, *TESTING OF METALLIC AND INORGANIC COATINGS,* STP 947, AMERICAN SOCIETY FOR TESTING AND MATERIALS, 1987
3. "CODE OF PRACTICE FOR PROTECTIVE COATING OF IRON AND STEEL STRUCTURES AGAINST CORROSION," BS 5493, BRITISH STANDARDS INSTITUTION, 1977; AVAILABLE FROM THE AMERICAN NATIONAL STANDARDS INSTITUTE

4. "SPRAYED METAL COATINGS FOR ATMOSPHERIC CORROSION PROTECTION," G 189-1966, CANADIAN STANDARDS ASSOCIATION, 1966; REAFFIRMED 1980
5. ROSENBURGH *ET AL., ALUMINUM OFFSHORE--SAFETY ANALYSIS,* PUBLICATION 85-3208, VERITEC, 1985
6. THERMITE SPARKING, *ALUMINUM DESIGN GUIDE,* VOL 2, *DESIGN GUIDE ON THE USE OF ALUMINUM IN OFFSHORE STRUCTURES,* WIMPEY OFFSHORE, LONDON, SECTION 4.5
7. "PAINTING WATER STORAGE TANKS," ANSI/AWS D102-78, AMERICAN WELDING SOCIETY, 1978
8. *STEEL STRUCTURES PAINTING MANUAL,* STEEL STRUCTURES PAINTING COUNCIL
9. "GUIDE FOR THE PROTECTION OF STEEL WITH THERMAL SPRAY COATINGS OF ALUMINUM AND ZINC AND THEIR ALLOYS AND COMPOSITES," ANSI/AWS C2.18-X, AMERICAN WELDING SOCIETY, IN PREPARATION

Underwater Welding

S. Liu and D.L. Olson, Colorado School of Mines, S. Ibarra, Amoco Corporation Research

---

Introduction

UNDERWATER ARC WELDING is performed in either the wet environment or in a protective habitat (Ref 1). Although both environments experience increased pressure with depth (0.1 MPa, or 1 bar, for each 10 m, or 33 ft, increase in depth), the wet environment also significantly alters the cooling rate during welding, which affects the nature of the weld-metal phase transformation. Whereas shielded metal arc welding (SMAW) for surface welding has a time to cool from 800 to 500 °C (1470 to 930 °F), $\Delta t_{8/5}$, in the range of 8 to 16 s, a typical wet-welding procedure reportedly has $\Delta t_{8/5}$ values between 1 and 6 s, depending on heat input (0.8 to 3.6 kJ/mm) and plate thickness (Ref 2, 3, 4). The greater cooling rate produces significant amounts of heat-affected zone (HAZ) martensite in nearly all low-carbon steels. As the carbon equivalent (CE) of these steels approaches a value of 0.40 wt%, the fusion-line hardness usually exceeds 400 on the Vickers scale (98 N, or 10 kgf), where

$$CE = \%C + \frac{\%Mn}{6} + \frac{\%(Cr+Mo+V)}{5} + \frac{\%(Ni+Cu)}{15} \qquad \textbf{(EQ 1)}$$

As the martensite content increases in the coarse-grained HAZ, the susceptibility to hydrogen cracking becomes a measurable concern.

When underwater arc welding is conducted directly in the water, the SMAW process is usually used, although more recently flux-cored arc welding (FCAW) has been used, as well. Wet underwater welding has been demonstrated to be an acceptable repair technique at depths of 100 m (330 ft). However, wet welds have been made on carbon steel structures at depths as low as 200 m (660 ft).

Welding that is performed in a dry, but high-pressure (hyperbaric) environment requires the construction of a chamber around the weld zone to create the dry habitat. The cost of hyperbaric welding is therefore much higher than that of wet welding. The setting up and tearing down of equipment can be very expensive in a deep-sea environment. Additionally, hyperbaric welding exhibits low mobility. Each fabrication situation often requires a customized setup. The gas-metal arc welding (GMAW) and gas-tungsten arc welding (GTAW) processes can both be used in hyperbaric welding, unlike wet welding.

Despite differences in process characteristics, the influences of water depth (pressure) on arc behavior are similar for both wet welding and hyperbaric welding (Ref 5, 6, 7). The specific technological features of each of these processes are provided in Ref 1. This article discusses the metallurgical aspects of the underwater welds.

References

1. C.L. TSAI AND K. MASUBUCHI, INTERPRETIVE REPORT ON UNDERWATER WELDING, *WELD. RES. COUNC. BULL.*, VOL 224, 1977, P 1-37
2. N. CHRISTENSEN, THE METALLURGY OF UNDERWATER WELDING, *UNDERWATER WELDING*, PERGAMON PRESS, 1983, P 71-79
3. A. HAUSI AND Y. SUGA, ON COOLING OF UNDERWATER WELDS, *TRANS. JPN. WELD. SOC.*, VOL 2 (NO. 1), 1980
4. C.L. TSAI AND K. MASUBUCHI, MECHANISMS OF RAPID COOLING IN UNDERWATER WELDING, *APPL. OCEAN RES.*, VOL 1 (NO. 2), 1979, P 99-110
5. C.E. GRUBBS AND O.W. SETH, "MULTIPASS ALL POSITION WET WELDING, A NEW UNDERWATER TOOL," OFFSHORE TECHNOLOGY CONFERENCE, OTC 1620 (HOUSTON), 1972
6. K. MASUBUCHI, UNDERWATER FACTORS AFFECTING WELDING METALLURGY, *PROC. CONF. UNDERWATER WELDING OF OFFSHORE PLATFORMS AND PIPELINES*, AWS, 1980, P 81-98
7. N.M. MADATOR, INFLUENCE OF THE PARAMETERS OF THE UNDERWATER WELDING PROCESS ON THE INTENSITY OF METALLURGICAL REACTIONS, *WELD. RES. ABROAD*, VOL 3, 1972, P 63

Underwater Welding

S. Liu and D.L. Olson, Colorado School of Mines, S. Ibarra, Amoco Corporation Research

Underwater Welding Pyrometallurgy

The pyrochemical behavior of underwater welds (both hyperbaric and wet) is reported (Ref 2, 8, 9, 10, 11) to be a function of depth (pressure) down to 300 m (990 ft). The variation of weld-metal manganese content as a function of depth is shown in Fig. 1. There is a sharp decrease in manganese content from 0.6 wt% for the surface weld to 0.25 wt% for the weld made at 30 m (90 ft). This decrease in weld-metal manganese content can be directly correlated to the rapid increase in weld-metal oxygen content for the same range of depth, as seen in Fig. 2. This observation suggests that manganese recovery in the weld pool is controlled by oxidation. Similar behavior was found for weld-metal silicon (Fig. 1). It is expected that all oxide-forming elements would react in this manner to an increase in oxygen.

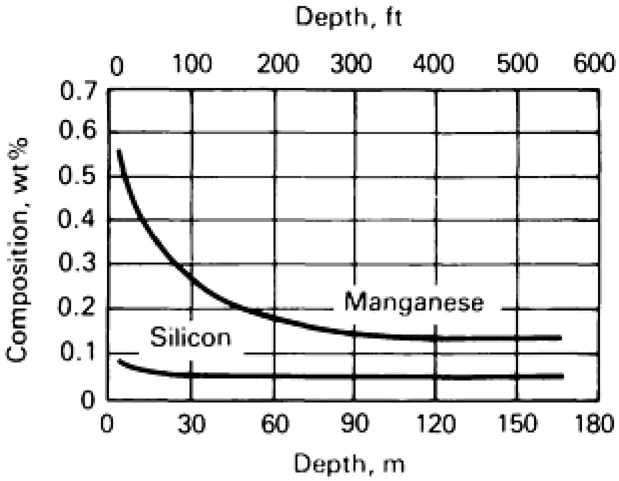

FIG. 1 WET UNDERWATER WELD-METAL MANGANESE AND SILICON AS A FUNCTION OF DEPTH. SOURCE: REF 11

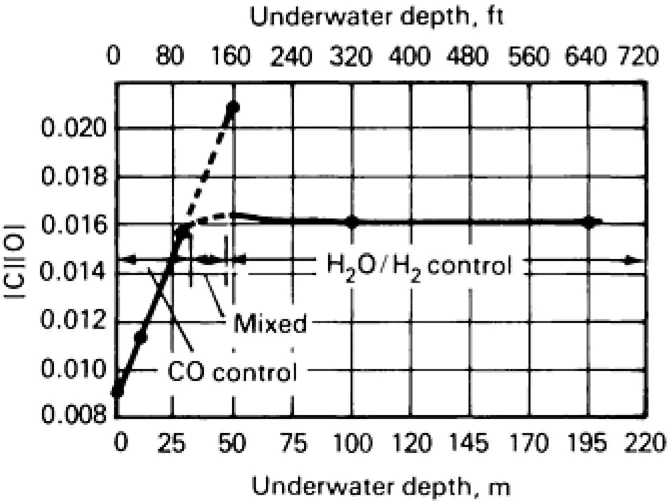

FIG. 2 PRODUCT OF WELD-METAL CARBON AND OXYGEN CONTENT AS A FUNCTION OF UNDERWATER DEPTH FOR WELD METAL PRODUCED WITH TREATED E6013 SMAW ELECTRODE. SOURCE: REF 11

The weld-metal carbon content, on the other hand, was found to increase significantly from the weld made at the surface to the welds made at 50 m (165 ft), as shown in Fig. 1. A linear relationship between the [C][O] product and the total pressure, *P*, or depth, has been found for both wet and hyperbaric SMAW processes, which suggests that the carbon monoxide reaction is a controlling mechanism. This behavior was observed for the hyperbaric SMAW process at depths equivalent to 300 m (990 ft). The carbon monoxide is a product of the decomposition of the calcium or magnesium carbonate that exists in the SMAW electrode coating to provide a working and protective welding plasma and cover. The resulting increases in weld-metal oxygen and carbon contents cause concern as to the use of a flux coating that contains carbonates for hyperbaric welding. Carbon pickup in the weld metal results in an increase in cracking susceptibility. Ferrosilicon additions to the flux did reduce weld-metal oxygen, but not sufficiently to alleviate this problem.

The carbon monoxide reaction was also found (Ref 10, 11) to be a controlling mechanism in wet underwater welding. The product of carbon and oxygen content, [C][O], for wet underwater welds was also plotted as a function of depth (pressure) in Fig. 2, which demonstrated an excellent linear correlation for the weld-metal carbon and oxygen product results down to 50 m (165 ft). This observation suggests that the carbon monoxide reaction controls the oxygen content down to 50 m (165 ft), and that the oxygencontent, in turn, controls the oxidation of manganese and silicon and, thus, the weld-metal contents of these elements. The actual depth to which the carbon monoxide reaction is a controlling factor is a function of the specific electrode composition.

In Fig. 2, the weld-metal oxygen and carbon contents become fairly constant between 50 and 100 m (165 and 330 ft). The weld-metal manganese and silicon contents are also fairly constant within this same range. These observations suggest that oxygen content still controls the weld-metal manganese and silicon contents, but that carbon monoxide formation is not the controlling reaction at depths greater than 50 m (165 ft).

The compositional results for depths greater than 50 m (165 ft) suggest that the H_2O decomposition reaction may be the controlling mechanism. At temperatures above 1000 °C (1830 °F), water vapor begins to dissociate into hydrogen and oxygen and dynamic equilibrium takes place (Ref 10, 11). For this reaction the weld-metal oxygen content is not a function of pressure (depth) and the [C][O] product will also remain constant. The diameter of the arc plasma column decreases with increasing pressure (depth). Thus, at depths greater than 50 m (165 ft), the water reaction becomes more dominant and the [C][O] product becomes less important and insensitive to depth, as seen in Fig. 2.

Another indication that hydrogen is more abundant in and involved with the rate-control chemistry is evident when considering arc stability, which decreases in extreme depth. In Fig. 3, the operating parameter space for underwater welding is shown to decrease with depth. This behavior can be explained by the high ionization potential for hydrogen, which makes it more difficult to sustain the welding arc.

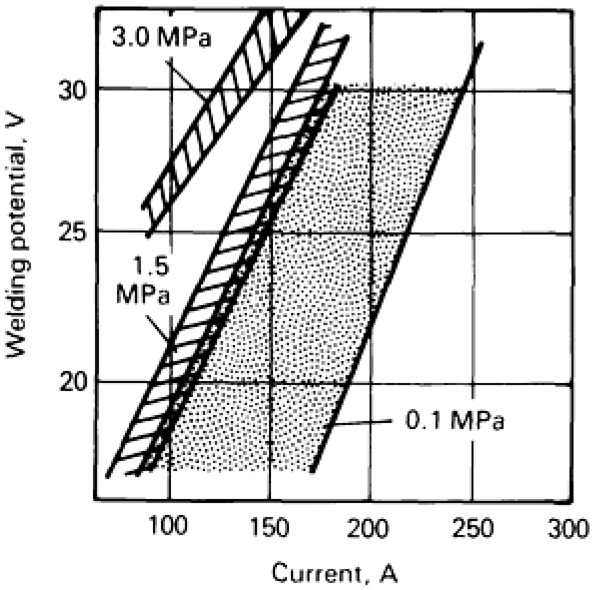

FIG. 3 REGIONS IN WELDING PARAMETER SPACE FOR SUCCESSFUL HIGH-PRESSURE WELDING AS A FUNCTION OF PRESSURE, OR DEPTH. NOTE THAT THE ACCEPTABLE WELDING PARAMETER SPACE REDUCES WITH PRESSURE, OR DEPTH. SOURCE: REF 11

References cited in this section

2. N. CHRISTENSEN, THE METALLURGY OF UNDERWATER WELDING, *UNDERWATER WELDING,* PERGAMON PRESS, 1983, P 71-79
8. O. GRONG, D.L. OLSON, AND N. CHRISTENSEN, ON THE CARBON OXIDATION IN HYPERBARIC MMA WELDING, *MET. CONSTR.,* VOL 17 (NO. 12), 1985, P 810R-814R
9. D.L. OLSON AND S. IBARRA, UNDERWATER WELDING METALLURGY, *PROC. FIRST OMAE SPECIALTY SYMPOSIUM ON OFFSHORE AND ARCTIC FRONTIERS,* ASME, 1986, P 439-447
10. S. IBARRA, C.E. GRUBBS, AND D.L. OLSON, FUNDAMENTAL APPROACHES TO UNDERWATER WELDING METALLURGY, *J. MET.,* VOL 40 (NO.12), 1988, P 8-10
11. S. IBARRA AND D.L. OLSON, WET UNDERWATER STEEL WELDING, *FERROUS ALLOY WELDMENTS,* VOL 69-70, *KEY ENG. MATERIALS,* TRANS TECH PUBLICATIONS, ZURICH, 1992, P 329-378

Underwater Welding

S. Liu and D.L. Olson, Colorado School of Mines, S. Ibarra, Amoco Corporation Research

---

Microstructural Development of Underwater Welds

Three types of ferrite are associated with low-carbon steel weld metal: grain-boundary ferrite (GBF), Widmanstätten, or sideplate, ferrite (SPF), and acicular ferrite (AF). Prior austenite grain boundaries can promote the formation of GBF and represent the sites for the nucleation of SPF, which protrudes from the grain boundaries into the prior austenite grains. Acicular ferrite is intragranular and has a much finer basket-weave structure. Other microconstituents, such as pearlite, cementite, and martensite, can also result. At faster cooling rates, the formation of bainite or ferrite with aligned carbide (AC) and martensite is possible. It is recognized that acicular ferrite is the microstructural constituent that gives a high resistance to cleavage fracture.

A typical wet (underwater) weld-metal microstructure, as a function of depth, is illustrated in Fig. 4. At shallow depths, the weld metal is primarily grain-boundary ferrite, with 10 to 20% aligned carbide. With increasing depths, the relative amount of grain-boundary ferrite decreases to about 50%, with increasing amounts of aligned carbide and sideplate ferrite. The major changes in microstructure occur in the first 50 m (165 ft) of depth. At depths greater than 50 m (165 ft), the weld-metal composition and microstructure remain fairly constant. This observation is also consistent with the constant weld-metal oxygen content described previously, which suggests a constant oxidation rate.

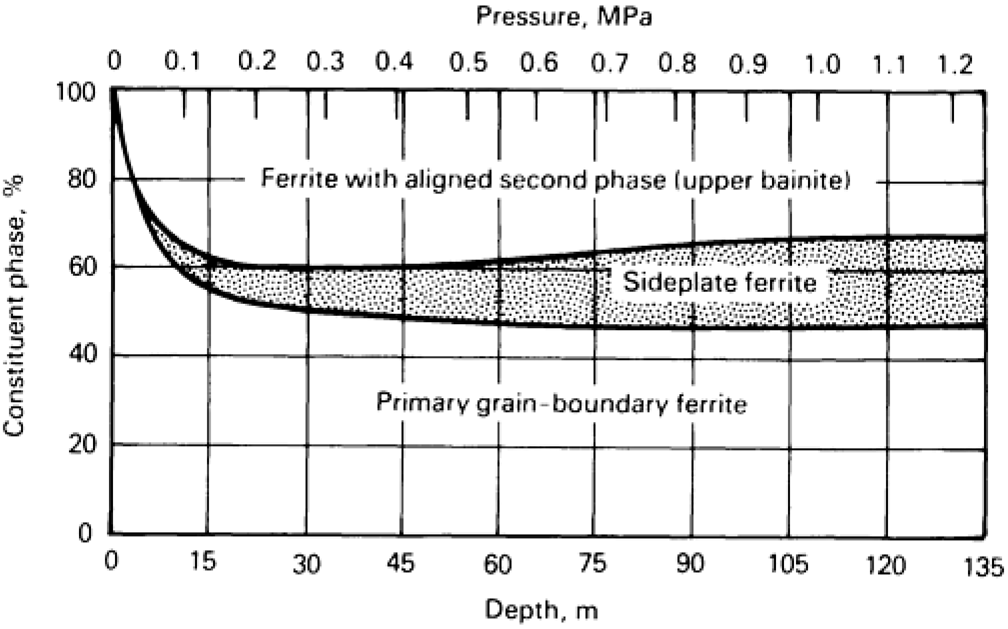

FIG. 4 PERCENTAGE OF WELD-METAL MICROSTRUCTURAL CONSTITUENTS FOR WET UNDERWATER WELDS AS A FUNCTION OF WATER DEPTH

The formation of acicular ferrite, instead of the significant sideplate ferrite content, is highly desirable in order to improve toughness. Acicular ferrite formation in underwater welds is possible, but only with specific weld-metal alloying additions, such as titanium or boron and with the proper weld-metal oxygen and manganese contents.

A diagram of weld-metal oxygen content versus effective weld-metal manganese content (Mn + 6C) has been used (Ref 8, 9, 10) to conceptually assist in selecting compositional modifications for the consumables of wet underwater welding, as shown in Fig. 5. Figure 5 represents a particular case in which deposits of an underwater wet E6013 electrode were constructed on the basis of microstructural and compositional analysis. This particular diagram represents the cooling rate typically found in wet welding.

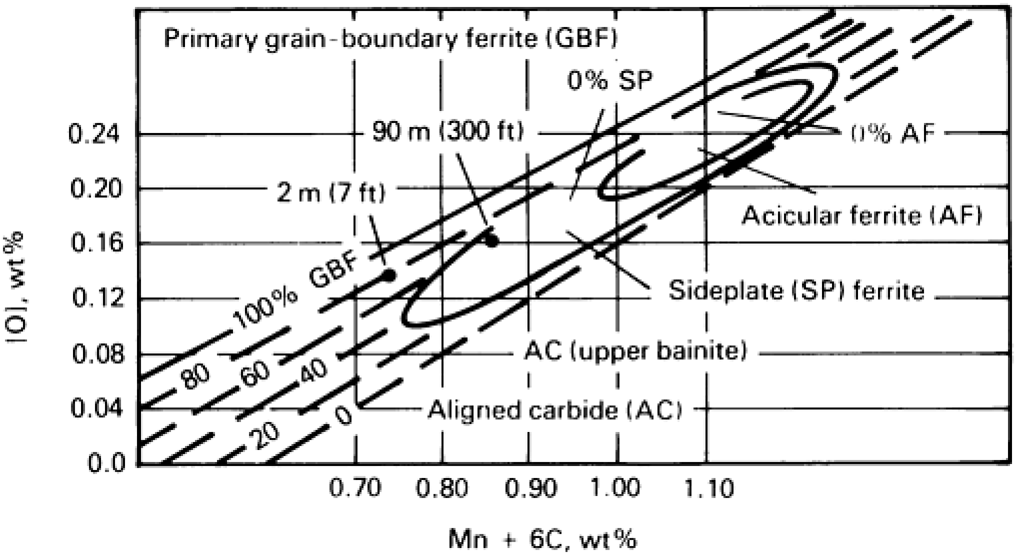

FIG. 5 SUGGESTED COMPOSITIONAL DIAGRAM FOR PREDICTION OF WELD-METAL MICROSTRUCTURE FOR WET UNDERWATER WELDS

References cited in this section

8. O. GRONG, D.L. OLSON, AND N. CHRISTENSEN, ON THE CARBON OXIDATION IN HYPERBARIC MMA WELDING, *MET. CONSTR.*, VOL 17 (NO. 12), 1985, P 810R-814R
9. D.L. OLSON AND S. IBARRA, UNDERWATER WELDING METALLURGY, *PROC. FIRST OMAE SPECIALTY SYMPOSIUM ON OFFSHORE AND ARCTIC FRONTIERS,* ASME, 1986, P 439-447
10. S. IBARRA, C.E. GRUBBS, AND D.L. OLSON, FUNDAMENTAL APPROACHES TO UNDERWATER WELDING METALLURGY, *J. MET.*, VOL 40 (NO.12), 1988, P 8-10

Underwater Welding

S. Liu and D.L. Olson, Colorado School of Mines, S. Ibarra, Amoco Corporation Research

---

Hydrogen Mitigation

Because moisture drastically increases the availability of hydrogen, the wet underwater welding of higher-strength steels is very susceptible to hydrogen cracking (Ref 12). Even during hyperbaric welding, after water has been removed from the chamber, the partial pressure of moisture is still high, causing the same effects as in wet welding.

The HAZ cracking that results from the wet welding of higher-strength steels is due to three necessary factors: hydrogen from the weld pool, microstructures that develop in the HAZ, and stress levels that develop in the weld joint. Hydrogen cracking can be reduced or managed by:

- USING CONSUMABLES THAT CAN HOLD A HIGH CONCENTRATION OF HYDROGEN IN MOLTEN WELD POOLS, AS WELL AS IN THE SOLIDIFIED WELD BEAD
- ALTERING THE FLUX COATINGS TO INTRODUCE ALTERNATE GASES IN THE WELDING PLASMA, THEREBY REDUCING THE HYDROGEN CONTENT
- SELECTING THE WELDING PARAMETERS THAT MINIMIZE WELD POOL HYDROGEN PICKUP

- DESIGNING THE WELD FOR MINIMUM STRESS

The use of ferritic-based weld deposits will cause the rejection of large amounts of hydrogen from the weld solidification. The hydrogen will go into the fully martensitic HAZ and will promote underbead cracking. A method that partially alleviates hydrogen susceptibility is to use an austenitic weld deposit that has a much larger solubility for hydrogen and, thus, a smaller tendency to transport hydrogen to the HAZ (Ref 13).

However, the use of an austenitic stainless steel consumable results in a deposit with a high coefficient of thermal expansion (CTE), relative to ferritic-based metal, as well as higher residual stresses and an increasing tendency for cracking. A solution to this problem, which is a compromise between high solubility for hydrogen and CTE mismatch, is to use high-nickel-content weld deposits, because nickel-base alloys have approximately the same CTE as ferritic steel and are capable of managing a large hydrogen content.

Although the nickel weld-metal deposit is virtually immune to the restraint cracking that was noted in the austenitic stainless steel welds, the use of nickel-base electrodes is restricted by their depth sensitivity. Although these electrodes are successful in preventing hydrogen cracking in the HAZ, the resulting welds can be porous and, in some cases, susceptible to embrittlement. Nickel-base welds at a depth of 30 m (100 ft) exhibited excessive porosity that was due to a lack of necessary heat input at this depth. These nickel-base wet-weld deposits have been successful only down to depths of 10 m (35 ft) when the SMAW process is used. Welding at greater depths, when using nickel-base electrodes, requires a higher level of heat input.

To reduce the influence of the plasma hydrogen content, other gases need to be introduced into the arc. Increasing the carbonate content of welding flux has been shown to increase the carbon monoxide content in the arc and, thus, to reduce the hydrogen content (Ref 14, 15). Carbonates (Ref 15) also increase the degree of oxidation of the deposited metal, which decreases weld-metal hydrogen content. The decomposition of carbonates results in higher oxygen and lower hydrogen contents.

The use of carbonates to reduce the plasma hydrogen content may not be as attractive when considering all the effects, because carbonates also promote weld-metal carbon and oxygen pickup with increasing pressure. Carbon has also been reported to reduce the hydrogen solubility in ferrite and to increase the hydrogen solubility in austenite. Thus, increasing carbon content will cause greater potential hydrogen supersaturation with the decomposition of austenite, thus reducing the favorable hydrogen reduction effects in the arc. This problem could limit the use of increased carbonate content to the shallower depths.

The influence of welding parameters on the underwater welding process and on the associated metallurgical reactions has been studied. The diffusible hydrogen content increases with increasing voltage, but decreases with increasing welding current. These results suggest that welding with lower potential and higher current will promote a significant reduction in the weld-metal diffusible hydrogen content (Ref 16). The diffusible hydrogen content in wet underwater welds decreases with increasing heat input for both SMAW and FCAW processes (Ref 17, 18).

The third factor that influences hydrogen cracking is the applied and residual stress states. The applied stress can be reduced by improving the fit-up on the weld joint to be used in the repair. In the underwater wet welding repair of platforms, scalloped sleeves are frequently fillet welded over the damaged area. Careful fabrication of the scalloped sleeve will result in improved fit-up and, thus, reduced stresses. The applied stress on a weldment can also be improved by using a larger weld-metal deposit. This practice will result in a larger HAZ, with a smaller applied stress.

The residual stresses are more difficult to manage in an environment that is not easily accessible to postweld heat treatment. Welding practices that can reduce residual stresses include:

- THE USE OF SMALL WELD DEPOSITS
- THE USE OF CONSUMABLES WITH CTES THAT ARE COMPATIBLE TO THAT OF THE BASE MATERIALS
- THE SELECTION OF EDGE PREPARATIONS THAT REDUCE THE SIZE OF THE TOTAL WELD DEPOSIT

The cross-sectional area of the total weld deposit is directly related to a shrinkage tendency. In designing an underwater weld joint, the concerns for applied stress and for residual stress must be balanced, because a different weld size is required in each case.

Temper Bead Practice. To further reduce stress and, thus, hydrogen cracking, the thermal experience(s) of the wet welding practice must be considered (Ref 8, 19). Although it is difficult to perform any significant preheat treatment procedures on the base material in wet, heavy-section welding, a temper bead practice that reduces the near-fusion-line hardness and cracking tendency has been perfected. A typical temper bead practice involves a weld deposition that is laid down over a previous weld deposit that has a less-susceptible cracking tendency than that of the base metal (Ref 20). As shown in Fig. 6, the second deposition, or temper bead, must be carefully located relative to the previous bead fusion line, such that its thermal experience tempers the near-fusion-line HAZ of the base material, which has a cracking-susceptible carbon equivalent (Ref 20).

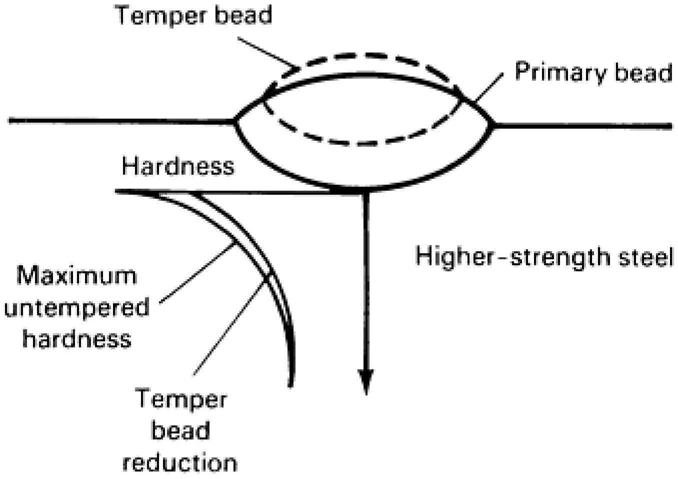

FIG. 6 USE OF TEMPER BEAD PRACTICE TO REDUCE HAZ HARDNESS AND SUSCEPTIBILITY FOR UNDERBEAD CRACKING. SOURCE: REF 11

There is evidence that temper weld beads can influence the properties of wet underwater welds. In the weld metal, Charpy V-notch toughness improved with an increasing number of weld passes (Ref 17, 18). This technique will reduce residual stresses, as well as hardness, and represents the optimum practice to wet weld the higher-CE steels. Another advantage to the temper bead practice is that while the temper bead is in the high-temperature austenitic condition, it is a favorable reservoir for hydrogen and, thus, it extracts some of the hydrogen from the more-crack-susceptible HAZ.

Welding consumables that generate chemical heat have been investigated as an additional heat source during underwater welding (Ref 8). Samples from a platform steel with a CE of 0.40 wt% were welded in a 10 m (35 ft) tank. A heat-generating thermit electrode weld was used after each pass to postweld heat treat the HAZ. Thermit-type reactions during arc welding have been used to improve the underwater temper bead practice in order to reduce the hydrogen cracking susceptibility of the HAZ.

References cited in this section

8. O. GRONG, D.L. OLSON, AND N. CHRISTENSEN, ON THE CARBON OXIDATION IN HYPERBARIC MMA WELDING, *MET. CONSTR.,* VOL 17 (NO. 12), 1985, P 810R-814R
11. S. IBARRA AND D.L. OLSON, WET UNDERWATER STEEL WELDING, *FERROUS ALLOY WELDMENTS,* VOL 69-70, *KEY ENG. MATERIALS,* TRANS TECH PUBLICATIONS, ZURICH, 1992, P 329-378
12. O. OZAKI, J. NAIMAN, AND K. MASUBUCHI, A STUDY OF HYDROGEN CRACKING IN UNDERWATER STEEL WELDS, *WELD. J.,* VOL 56 (NO. 8), 1977, P 231S-237S

13. W.A. STALKER, P.H.M. HART, AND G.R. SALTER, "AN ASSESSMENT OF SHIELDED METAL ARC ELECTRODES FOR THE UNDERWATER WELDING OF CARBON MANGANESE STRUCTURAL STEELS," OFFSHORE TECHNOLOGY CONFERENCE, OTC REPORT 2301 (HOUSTON), 1975
14. B. CHEW, PREDICTION OF WELD METAL HYDROGEN LEVELS OBTAINED UNDER TEST CONDITIONS, *WELD. J.,* VOL 52, 1973, P 386S-391S
15. L.I. SOROKIN AND Z.A. SIDLIN, THE EFFECT OF ALLOYING ELEMENTS AND OF MARBLE IN AN ELECTRODE COATING ON THE SUSCEPTIBILITY OF A DEPOSITED NICKEL CHROME METAL TO PORE FORMATION, *SVAR. PROIZVOD.,* NO. 11, 1974, P 7-9
16. V.Y. KONONENKO, EFFECT OF WATER SALINITY AND MECHANIZED UNDERWATER WELDING PARAMETERS ON HYDROGEN AND OXYGEN CONTENT OF WELD METAL, *WELDING UNDER EXTREME CONDITIONS,* PERGAMON PRESS, 1989, P 113-118
17. H. HOFFMEISTER AND K. KUSTER, PROCESS VARIABLES AND PROPERTIES OF UNDERWATER WET SHIELDED METAL ARC LABORATORY WELDS, *UNDERWATER WELDING,* PERGAMON PRESS, 1983, P 115-125
18. K. HOFFMEISTER AND K. KUSTER, PROCESS VARIABLES AND PROPERTIES OF WET UNDERWATER GAS METAL ARC LABORATORY AND SEA WELDS OF MEDIUM STRENGTH STEELS, *UNDERWATER WELDING,* PERGAMON PRESS, 1983, P 239-246
19. A. MATSUNAWA, H. TAKEMATA, AND I. OKAMOTO, ANALYSIS OF TEMPERATURE FIELD WITH EQUIRADIAL COOLING BOUNDARY AROUND MOVING HEAT SOURCES--HEAT CONDUCTION ANALYSIS FOR ESTIMATION OF THERMAL HYSTERESIS DURING UNDERWATER WELDING, *TRANS. JWRI,* VOL 9 (NO. 1), 1980, P 11-18
20. K. OLSEN, D.L. OLSON, AND N. CHRISTENSEN, WELD BEAD TEMPERING OF HEAT AFFECTED ZONE, *SCAND. J. METALL.,* VOL 11, 1982, P 163-168

Underwater Welding

S. Liu and D.L. Olson, Colorado School of Mines, S. Ibarra, Amoco Corporation Research

---

Weld-Metal Porosity

Weld-metal porosity is a major concern in underwater welds. The amount of porosity determines whether a wet weld will meet the type A, B, or C weld specification (ANSI/AWS D3.6-83 specification). Pore formation results from entrapment, supersaturation of dissolved gas, or gas-producing chemical reactions. The nature and amount of weld-metal porosity involves at least four competing time-dependent processes (Ref 21): the nucleation, growth, transport, and coalescence of pores. The physical requirement for pore formation is that the sum of the partial pressure, P_g, of soluble gases must exceed the sum of the following pressure terms:

$$P_G > P_A + P_H + P_B \qquad \textbf{(EQ 2)}$$

where P_a is the atmosphere pressure, P_h is the hydrostatic pressure, and P_b is the pressure increase that is due to the curvature of the pore. For underwater welding, P_h is the controlling term, because it is directly related to depth.

The effect of water pressure on the formation of porosity in underwater gravity welding with electrodes of the ilmenite type, high titanium oxide type, and iron powder/iron oxide type and with SM41 steel base metals has been reported (Ref 22). The amount of porosity increased with water pressure (depth) (Fig. 7). The pore shape changed from spherical to long and narrow with increasing pressure, which indicates a change of the pore-formation mechanism. In wet underwater welds, the pores contained approximately 96 vol% H_2, 0.4 vol% CO, and 0.06 vol% CO_2. Other bubble compositions (Ref 6, 23) of 62 to 82% H_2, 11 to 24% CO, and 4 to 6% CO_2 have been reported (Ref 24). More-specific data on a rutile iron powder (E7014) electrode are 45% H_2, 43% CO, 8% CO_2, and 4% other. The variation in bubble composition indicates the broad nature of the chemical processes associated with wet underwater welding and is most likely due to variations in electrode covering compositions, energy input, and water depth.

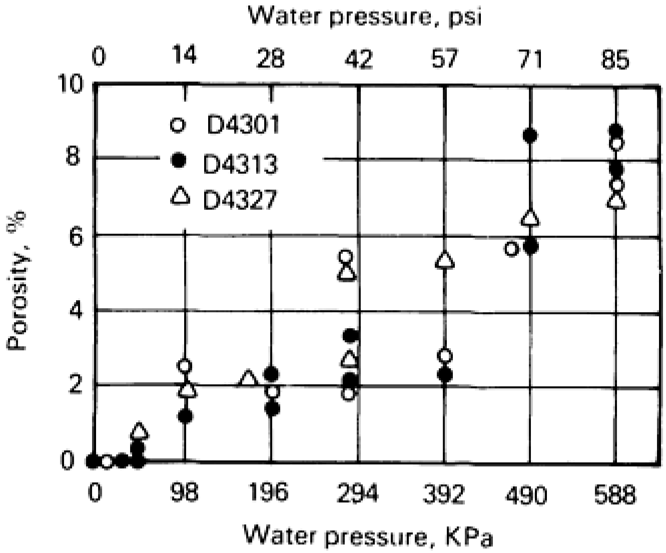

FIG. 7 EFFECT OF WATER PRESSURE ON POROSITY. SOURCE: REF 22

Two possible methods of reducing weld-metal porosity with depth, in the case of wet welding, are to add to the weld pool specific elements that getter the hydrogen by forming hydrides and to alter the welding parameters to reduce pore formation. At increasing travel speeds, the number of pores per volume of weld deposit goes through a maximum. A relationship between the product of travel speed (*S*) and bead area (*A*) and the amount of gas absorbed per unit volume of weld metal exists (Ref 25). This relationship suggests that for decreasing values of the product (*SA*), the gas absorbed per unit volume of weld metal increases for the same gas partial pressure. With small beads and a slow welding speed, the amount of gas absorbed per unit volume of weld metal is high. The faster freezing rate of the smaller, short-circuiting arc weld allows less time for gas desorption and causes more gas bubbles to be trapped before floating to the surface of the weld pool. This relationship holds significance for many of the practices used in wet underwater welding, which produce small weld deposits. In underwater welding, the relationship between porosity and welding current is strongly influenced by the type of moisture resistance coating on the electrode. The amount of gas pickup increases as the arc was lengthened (Ref 21). In general, the hydrogen absorption and, therefore, porosity levels in welding can be minimized by using a low current with direct current electrode positive (DCEP), a high current with direct current electrode negative (DCEN), a short arc, and a fast travel speed.

References cited in this section

6. K. MASUBUCHI, UNDERWATER FACTORS AFFECTING WELDING METALLURGY, *PROC. CONF. UNDERWATER WELDING OF OFFSHORE PLATFORMS AND PIPELINES,* AWS, 1980, P 81-98
21. R.E. TREVISAN, D.D. SCHWEMMER, AND D.L. OLSON, THE FUNDAMENTALS OF WELD METAL PORE FORMATION, *WELDING: THEORY AND PRACTICE,* D.L. OLSON, R.D. DIXON, AND A.L. LIBY, ED., ELSEVIER SCIENCE B.V., 1990
22. Y. SUGA AND H. ATSUSHI, "ON THE FORMATION OF POROSITY IN UNDERWATER WELD METAL (THE 1ST REPORT)--EFFECT OF WATER PRESSURE ON FORMATION OF POROSITY," IIW DOC. IX-1388-86, AMERICAN COUNCIL, AWS, 1986
23. N.M. MALATOR, THE PROPERTIES OF THE BUBBLES OF STEAM AND GAS AROUND THE ARC IN UNDERWATER WELDING, *AUTOMAT. WELD.,* VOL 18 (NO. 12), 1965, P 25-29
24. E.A. SILVA, GAS PRODUCTION AND TURBIDITY DURING UNDERWATER SHIELDED METAL-ARC WELDING WITH IRON POWDER ELECTRODES, *NAV. ENG. J.,* VOL 12, 1971
25. G.R. SALTER AND D.R. MILNER, GAS METAL REACTIONS IN ARC WELDING, *BRIT. WELD. J.,* VOL 33, 1965

Fatigue as a Function of Porosity

The designers of offshore structures are usually concerned with the initiation and growth of fatigue cracks in fracture-critical members, because of noted decreases in fracture toughness, as well as the presence of varying amounts of porosity. Cyclic fatigue stresses will develop with any natural sea movement and will increase during storms.

Drilling the tip of a fatigue crack is a common field technique that is used to prevent further crack growth. In the same manner, the presence of porosity in weldments can be beneficial, because the pores act as pinning sites for fatigue cracks and can either retard or stop their growth. In particular, the presence of some amount of porosity retarded fatigue crack growth at low stress in surface habitat and wet underwater welds (Ref 26) (Fig. 8). The low-porosity dry habitat weld was shown to contain approximately 3% porosity on the fatigue fracture surface, whereas the high-porosity wet underwater weld contained approximately 12% porosity on the fatigue fracture surface.

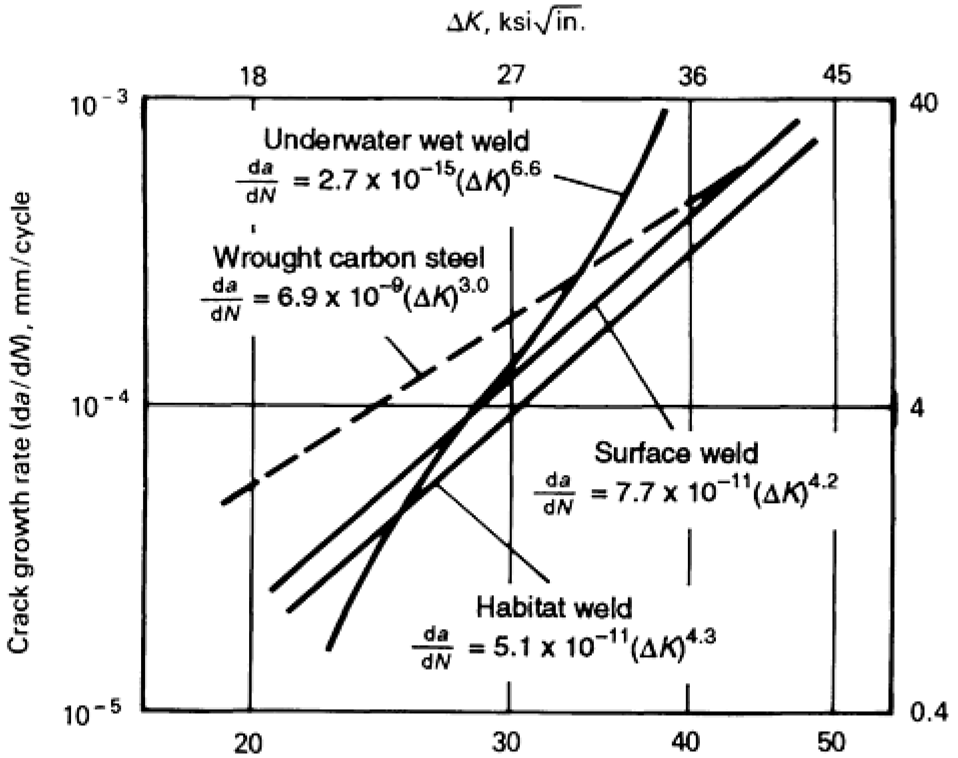

FIG. 8 COMPARISON OF EFFECTS OF POROSITY AND FREQUENCY ON THE FATIGUE CRACK GROWTH BEHAVIOR OF WELDS TESTED IN SEAWATER. SOURCE: REF 26

In comparison to normal wrought ferrite-pearlite steels, all experimental underwater welds exhibited lower growth rates for low values of ΔK, where ΔK is the cyclic variation in the stress intensity factor, K. However, the wet welds tended to have a higher fatigue crack growth rate than the wrought alloys at high values of ΔK, particularly in wet welds.

At low ΔK values, the wet underwater welds have the most fatigue crack growth resistance. The large number of pores act as pinning sites for the advancing fatigue crack front. Increased porosity in a weldment yielded lower fatigue crack growth at low ΔK values.

At high ΔK values, the crack growth rate is much greater for the higher-porosity content. At higher stress levels, the mechanical behavior of the sample is more like that experienced in a tensile test. The high numbers of pores act to reduce the cross-sectional area and, thus, allow the cracks to propagate at a high rate. The use of wet weld repairs in areas of high cyclic stress should be accompanied by caution and a load-reducing joint design.

Reference cited in this section

26. D.K. MATLOCK, G.R. EDWARDS, D.L. OLSON, AND J. IBARRA, AN EVALUATION OF THE FATIGUE BEHAVIOR IN SURFACE, HABITAT, AND UNDERWATER WET WELDS, *UNDERWATER WELDING,* PERGAMON PRESS, 1983, P 303-310

Underwater Welding

S. Liu and D.L. Olson, Colorado School of Mines, S. Ibarra, Amoco Corporation Research

---

Heat Sources

With increasing depth and, thus, pressure, the welding parameter space for sustaining a welding arc becomes more restricted. This effect necessitates a more careful selection of the welding voltage (gap) and current for a given size and type of welding electrode. The decreasing welding parameter space also reflects a shift of current to lower values and voltage to higher values (see Fig. 3), which changes the nature of the heat input. Electrode coating thickness can also influence the arc characteristics (Ref 16). Figure 9 illustrates that increasing coating thickness also raises the arc characteristic curves.

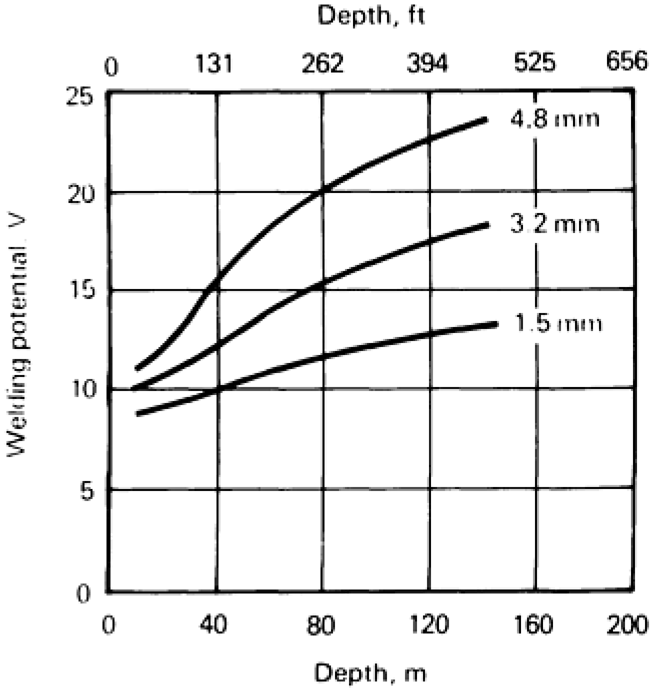

FIG. 9 EFFECTS OF DEPTH ON ARC VOLTAGE FOR THREE DIFFERENT GTAW ELECTRODE DIAMETERS. SOURCE: REF 29

With an increase in the welding depth, the duration and frequency of short circuits and arc extinctions were reported to grow significantly (Ref 28), and the process stability deteriorated. The observed rise of welding current is associated with the increase in the number of short circuits. The addition of rare earth metals into the charge of rutile-base flux-cored wire provides considerable improvement in arc stability and reduces arc downtime (arc extinctions and short circuits) by nearly two times. The more-stable arc burning that occurs when welding with the rare earth metal-containing wire allows its use in any spatial position and within the depth ranges investigated.

An approach that separates the wet conditions from the weld pool has been demonstrated, in which a welding torch that produces a local gas cavity by controlling the fluid flow in the environment of the weld is utilized (Ref 29). This method hinders the water invasion by the momentum of the fluid jet and by the smooth exhausting of cavity gases by flow entrapment. The process allows near-surface welding conditions to exist and permits the flexibility of wet welding practices while achieving weld quality that is similar to dry-environment underwater welding.

References cited in this section

16. V.Y. KONONENKO, EFFECT OF WATER SALINITY AND MECHANIZED UNDERWATER WELDING PARAMETERS ON HYDROGEN AND OXYGEN CONTENT OF WELD METAL, *WELDING UNDER EXTREME CONDITIONS,* PERGAMON PRESS, 1989, P 113-118
28. I.K. POKHODNYA, V.N. GORPENYUK, V.YA. KONONENKO, V.E. PONOMAREV, AND S.YU. MAKSIMOV, SOME PECULIARITIES OF ARC BURNING AND METAL TRANSFER IN WET UNDERWATER SELF-SHIELDING FLUX-CORED WIRE WELDING, *WELDING UNDER EXTREME CONDITIONS,* PERGAMON PRESS, 1989, P 151
29. S. FUKUSHIMA, T. FUKUSHIMA, AND J. KINUGAWA, UNDERWATER WET PLASMA WELDING IN PRESSURIZED WATER, *TRANS. MET. INST. MET.*, VOL 19 (NO. 3), 1977, P 133-151

Underwater Welding

S. Liu and D.L. Olson, Colorado School of Mines, S. Ibarra, Amoco Corporation Research

---

Practical Applications for Underwater Welding

Underwater welding techniques are used primarily for the repair of offshore platforms, particularly after the normal 5 year inspection period. However, these techniques have also been very useful during the installation of new offshore structures and undersea pipelines, the installation of hot taps, the repair of dock and harbor facilities, the modification of and addition to underwater structures, and the repair of nuclear facilities.

Examples of maintenance and repair applications are described below.

Offshore Structures. Underwater welding techniques are used to repair damages caused by corrosion or by fatigue cracking, and to repair members that have been damaged by ship impact. The repair or replacement of structural members damaged during installation by objects that have fallen overboard or by other accidents have also been reported. The underwater wet welding repair of these structures requires that the carbon equivalent of the steel be below 0.40 wt% (as calculated using the International Institute of Welding formula) to prevent hydrogen cracking. For steels with high carbon equivalents, the use of multiple temper bead techniques is required in order to reduce the hardness in the HAZ and eliminate the presence of hydrogen.

Undersea Pipelines. The repair or replacement of undersea pipelines is usually conducted in a hyperbaric dry chamber. Although sleeve repairs using wet welds have been tested successfully, their use has not been reported to date. Hyperbaric chambers are used regularly for underwater tie-ins and for repairs of defects.

Harbor Facilities. Corrosion and collision damage to sheet and pilings have been repaired. The repair and replacement of tubular dock braces, dock supports, and tanker moorings have also been reported.

Floating Vessels. Permanent and temporary repairs to holes in ship and barge hulls have been conducted. The hulls and pontoons of semisubmersible drill ships have also been repaired.

Nuclear Power Plants. The wet welding of nuclear power plant components is usually conducted with austenitic stainless steel electrodes in order to match the austenitic base metal used in the construction of these power plants. Damaged reactor internal components and leaks in pool liners are the usual types of repairs.

Specifications. The first underwater welding specification, known as ANSI/AWS D3.6, "Specification for Underwater Welding," was published in 1983 by the American Welding Society. This specification provides comprehensive information that enables engineers to select the welding method (wet or dry) that meets the established fitness-for-purpose requirements. The specification also provides a means for the engineer to make critical decisions in structural fabrication when underwater welding techniques are used. It is also a performance specification that is intended to assist the purchaser in specifying and obtaining the desired mechanical properties when using underwater welding techniques. The specification also allows contractors to prepare estimates and quotations for each project with a clear understanding of what is required. The original specification has been revised and reissued several times.

The AWS D3.6 specification defines four weld types, or classes. Each weld type has a specific set of quality requirements that must be verified during procedure qualification. The type of weld selected for a specific application is determined by the user.

Type A welds are represented by a set of requirements that produces an underwater weld comparable to a surface weld. Underwater welds that meet these requirements are usually hyperbaric (dry) welds.

Type B welds are usually reserved for wet underwater welds. They are intended for less-critical applications where reduced ductility and increased porosity can be tolerated. Typical applications include scallop sleeves on offshore structures.

Type C welds satisfy lesser requirements than type B welds and are intended for applications where load-beating function is not a primary consideration.

Type O welds usually meet the requirements of surface dry welds, as well as those of other codes and specifications.

Other underwater welding documents are developed by agencies such as the International Institute of Welding (IIW Doc SCUW 124-90), the American Society of Mechanical Engineers (ASME Section X/IWA 4550 for Underwater Welding of Nuclear Vessels), and the Norwegian agency Det Norske Veritas (DNV) (Document RP3604, Recommended Practices for Underwater Welding).

Underwater Welding

S. Liu and D.L. Olson, Colorado School of Mines, S. Ibarra, Amoco Corporation Research

---

References

1. C.L. TSAI AND K. MASUBUCHI, INTERPRETIVE REPORT ON UNDERWATER WELDING, *WELD. RES. COUNC. BULL.*, VOL 224, 1977, P 1-37
2. N. CHRISTENSEN, THE METALLURGY OF UNDERWATER WELDING, *UNDERWATER WELDING,* PERGAMON PRESS, 1983, P 71-79
3. A. HAUSI AND Y. SUGA, ON COOLING OF UNDERWATER WELDS, *TRANS. JPN. WELD. SOC.*, VOL 2 (NO. 1), 1980
4. C.L. TSAI AND K. MASUBUCHI, MECHANISMS OF RAPID COOLING IN UNDERWATER WELDING, *APPL. OCEAN RES.*, VOL 1 (NO. 2), 1979, P 99-110
5. C.E. GRUBBS AND O.W. SETH, "MULTIPASS ALL POSITION WET WELDING, A NEW UNDERWATER TOOL," OFFSHORE TECHNOLOGY CONFERENCE, OTC 1620 (HOUSTON), 1972
6. K. MASUBUCHI, UNDERWATER FACTORS AFFECTING WELDING METALLURGY, *PROC. CONF. UNDERWATER WELDING OF OFFSHORE PLATFORMS AND PIPELINES,* AWS, 1980, P 81-98

7. N.M. MADATOR, INFLUENCE OF THE PARAMETERS OF THE UNDERWATER WELDING PROCESS ON THE INTENSITY OF METALLURGICAL REACTIONS, *WELD. RES. ABROAD,* VOL 3, 1972, P 63
8. O. GRONG, D.L. OLSON, AND N. CHRISTENSEN, ON THE CARBON OXIDATION IN HYPERBARIC MMA WELDING, *MET. CONSTR.,* VOL 17 (NO. 12), 1985, P 810R-814R
9. D.L. OLSON AND S. IBARRA, UNDERWATER WELDING METALLURGY, *PROC. FIRST OMAE SPECIALTY SYMPOSIUM ON OFFSHORE AND ARCTIC FRONTIERS,* ASME, 1986, P 439-447
10. S. IBARRA, C.E. GRUBBS, AND D.L. OLSON, FUNDAMENTAL APPROACHES TO UNDERWATER WELDING METALLURGY, *J. MET.,* VOL 40 (NO.12), 1988, P 8-10
11. S. IBARRA AND D.L. OLSON, WET UNDERWATER STEEL WELDING, *FERROUS ALLOY WELDMENTS,* VOL 69-70, *KEY ENG. MATERIALS,* TRANS TECH PUBLICATIONS, ZURICH, 1992, P 329-378
12. O. OZAKI, J. NAIMAN, AND K. MASUBUCHI, A STUDY OF HYDROGEN CRACKING IN UNDERWATER STEEL WELDS, *WELD. J.,* VOL 56 (NO. 8), 1977, P 231S-237S
13. W.A. STALKER, P.H.M. HART, AND G.R. SALTER, "AN ASSESSMENT OF SHIELDED METAL ARC ELECTRODES FOR THE UNDERWATER WELDING OF CARBON MANGANESE STRUCTURAL STEELS," OFFSHORE TECHNOLOGY CONFERENCE, OTC REPORT 2301 (HOUSTON), 1975
14. B. CHEW, PREDICTION OF WELD METAL HYDROGEN LEVELS OBTAINED UNDER TEST CONDITIONS, *WELD. J.,* VOL 52, 1973, P 386S-391S
15. L.I. SOROKIN AND Z.A. SIDLIN, THE EFFECT OF ALLOYING ELEMENTS AND OF MARBLE IN AN ELECTRODE COATING ON THE SUSCEPTIBILITY OF A DEPOSITED NICKEL CHROME METAL TO PORE FORMATION, *SVAR. PROIZVOD.,* NO. 11, 1974, P 7-9
16. V.Y. KONONENKO, EFFECT OF WATER SALINITY AND MECHANIZED UNDERWATER WELDING PARAMETERS ON HYDROGEN AND OXYGEN CONTENT OF WELD METAL, *WELDING UNDER EXTREME CONDITIONS,* PERGAMON PRESS, 1989, P 113-118
17. H. HOFFMEISTER AND K. KUSTER, PROCESS VARIABLES AND PROPERTIES OF UNDERWATER WET SHIELDED METAL ARC LABORATORY WELDS, *UNDERWATER WELDING,* PERGAMON PRESS, 1983, P 115-125
18. K. HOFFMEISTER AND K. KUSTER, PROCESS VARIABLES AND PROPERTIES OF WET UNDERWATER GAS METAL ARC LABORATORY AND SEA WELDS OF MEDIUM STRENGTH STEELS, *UNDERWATER WELDING,* PERGAMON PRESS, 1983, P 239-246
19. A. MATSUNAWA, H. TAKEMATA, AND I. OKAMOTO, ANALYSIS OF TEMPERATURE FIELD WITH EQUIRADIAL COOLING BOUNDARY AROUND MOVING HEAT SOURCES--HEAT CONDUCTION ANALYSIS FOR ESTIMATION OF THERMAL HYSTERESIS DURING UNDERWATER WELDING, *TRANS. JWRI,* VOL 9 (NO. 1), 1980, P 11-18
20. K. OLSEN, D.L. OLSON, AND N. CHRISTENSEN, WELD BEAD TEMPERING OF HEAT AFFECTED ZONE, *SCAND. J. METALL.,* VOL 11, 1982, P 163-168
21. R.E. TREVISAN, D.D. SCHWEMMER, AND D.L. OLSON, THE FUNDAMENTALS OF WELD METAL PORE FORMATION, *WELDING: THEORY AND PRACTICE,* D.L. OLSON, R.D. DIXON, AND A.L. LIBY, ED., ELSEVIER SCIENCE B.V., 1990
22. Y. SUGA AND H. ATSUSHI, "ON THE FORMATION OF POROSITY IN UNDERWATER WELD METAL (THE 1ST REPORT)--EFFECT OF WATER PRESSURE ON FORMATION OF POROSITY," IIW DOC. IX-1388-86, AMERICAN COUNCIL, AWS, 1986
23. N.M. MALATOR, THE PROPERTIES OF THE BUBBLES OF STEAM AND GAS AROUND THE ARC IN UNDERWATER WELDING, *AUTOMAT. WELD.,* VOL 18 (NO. 12), 1965, P 25-29
24. E.A. SILVA, GAS PRODUCTION AND TURBIDITY DURING UNDERWATER SHIELDED METAL-ARC WELDING WITH IRON POWDER ELECTRODES, *NAV. ENG. J.,* VOL 12, 1971
25. G.R. SALTER AND D.R. MILNER, GAS METAL REACTIONS IN ARC WELDING, *BRIT. WELD. J.,* VOL 33, 1965

26. D.K. MATLOCK, G.R. EDWARDS, D.L. OLSON, AND J. IBARRA, AN EVALUATION OF THE FATIGUE BEHAVIOR IN SURFACE, HABITAT, AND UNDERWATER WET WELDS, *UNDERWATER WELDING,* PERGAMON PRESS, 1983, P 303-310
27. G. EDWARDS, C.J. ALLUM, B.E. PINFOLD, AND J.H. NIXON, THE EFFECT OF PRESSURE ON THE TUNGSTEN ARGON WELDING ARC, *PROC. CONF. PHYSICS AND WELD POOL BEHAVIOR,* WELDING INSTITUTE, 1979, P 101-107
28. I.K. POKHODNYA, V.N. GORPENYUK, V.YA. KONONENKO, V.E. PONOMAREV, AND S.YU. MAKSIMOV, SOME PECULIARITIES OF ARC BURNING AND METAL TRANSFER IN WET UNDERWATER SELF-SHIELDING FLUX-CORED WIRE WELDING, *WELDING UNDER EXTREME CONDITIONS,* PERGAMON PRESS, 1989, P 151
29. S. FUKUSHIMA, T. FUKUSHIMA, AND J. KINUGAWA, UNDERWATER WET PLASMA WELDING IN PRESSURIZED WATER, *TRANS. MET. INST. MET.*, VOL 19 (NO. 3), 1977, P 133-151

Welding for Cryogenic Service

T.A. Siewert and C.N. McCowan, National Institute of Standards and Technology

Introduction

CRYOGENIC TEMPERATURES cause many structural alloys to become brittle, which is an unacceptable condition in most structural applications. Therefore, structures built for service at low temperatures are typically made from alloys that maintain some ductility at the service temperatures. Cryogenic alloys include 9Ni steels, austenitic stainless steels, manganese stainless steels, maraging steels, titanium, aluminum, and nickel alloys. The choice of weld-metal alloy may depend solely on the strength of the alloy at a given temperature or on a combination of strength, toughness, fatigue resistance, thermal conductivity, magnetic permeability, and other considerations.

Wrought or cast alloy structural members are usually joined using filler materials with compositions that are similar, but not identical, to the parent material. The weld composition is optimized to compensate for the inherent differences in the properties of weld and parent material that are due to grain structure, inclusion content, and cooling rate differences. The goal is to match, as closely as possible, the mechanical and physical properties of the weld and parent material. When nonmatching weld compositions are most appropriate, differences between the welds and parent material in terms of thermal contraction, corrosion, and other factors must be considered. This article discusses these differences and explains how they affect the choice of the weld filler metal.

Effects of Cryogenic Service on Properties

Strength. As the temperature drops below room temperature, some cryogenic alloys strengthen appreciably. Unless some other property degrades and offsets this effect, such strengthening is very beneficial in reducing the required weight and thickness of structural members that are designed on the basis of load. Figure 1 shows this strengthening effect for several alloy systems, using data that are representative of both base and weld metal. Unfortunately, not all alloys strengthen at the same rate. Therefore, a comparison of the strengths at one temperature cannot be used to predict the relative strengths at a different temperature. Most literature data are for certain convenient test temperatures (4 K, 77 K, and 298 K). Data for intermediate temperatures indicate that strength is a smoothly varying function of temperature, whereas data for the three common test temperatures can be linked with curved lines. Because the behavior is very predictable, this discussion concentrates on the extreme temperature, 4 K, where the effects are most pronounced. Behavior at 77 K can be estimated by interpolation.

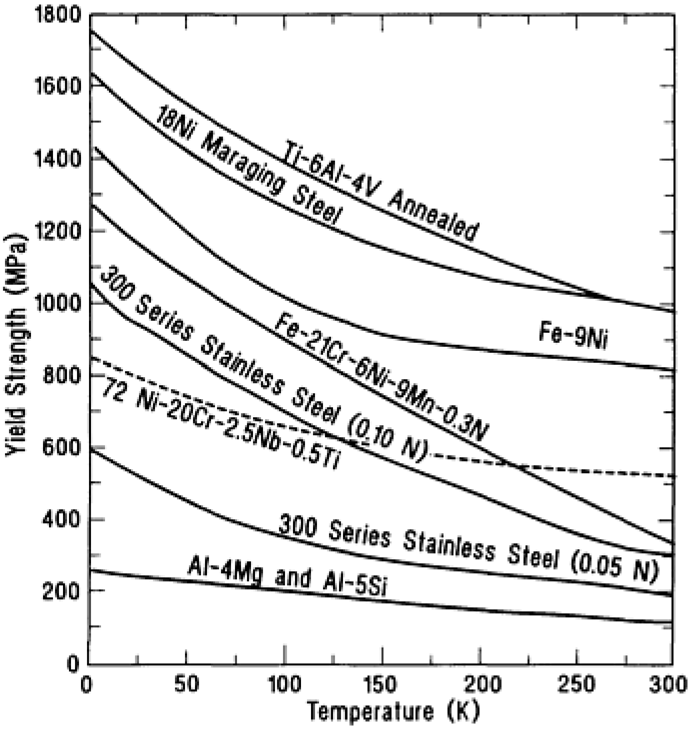

FIG. 1 YIELD STRENGTH VERSUS TEMPERATURE (4 TO 300 K) FOR VARIOUS BASE METALS AND WELD METALS THAT HAVE BEEN STUDIED FOR CRYOGENIC APPLICATIONS

As shown in Fig. 1, some alloys can reach very high yield strengths (well above 1000 MPa, or 140 ksi) at 4 K. Others, such as aluminum alloys, start from moderate or low strengths and respond mildly to a lowering of temperature (Ref 1, 2). Nine-percent nickel (Fe-9Ni) starts with a strength that is high at room temperature and rises from that point (Ref 1, 2). High-manganese steels, such as Fe-22Mn, and 21Cr-6Ni-9Mn stainless steel at 4 K approach the ultra-high-strength levels of 9Ni steel, maraging steel, and titanium (Ref 3, 4).

The austenitic (300 series) stainless steels offer medium strength at room temperature, but strengthen with decreasing temperature at a rate dependent on the nitrogen content of the alloy. As shown in Fig. 1, the strengthening of these alloys is a strong function of their nitrogen content: the trend line for the alloys with a higher nitrogen content shows much greater strengthening than the trend line for the alloys with a lower nitrogen content (Ref 5).

The maximum strength that can be reached is a function of the solubility limit of nitrogen in the stainless steel. This limit, in turn, is a function of composition and other factors. To control strength values, the nitrogen content of the weld must be carefully controlled, or else two effects can occur: pores may form in the weld during solidification (as the supersaturated liquid weld metal rejects nitrogen) or the weld pool may lose nitrogen through evaporation, weakening the weld. It should be noted that the strength of 21Cr-6Ni-9Mn is much higher than that of other austenitic steels because its nitrogen content is near 0.3 wt%. This higher level of strength is made possible by the improved nitrogen solubility in an alloy with a higher manganese content.

Equation 1 was developed to predict the strength of welds in austenitic stainless steel at a temperature of 4 K:

$$\sigma_Y(\text{MPA}) = 180 + 3200\ \text{N} + 33\ \text{Mo} + 32\ \text{Mn} + 13\ \text{Ni} \qquad \textbf{(EQ 1)}$$

where the elemental symbols represent their concentration in wt% (Ref 5). Equation 1 shows which elements have the greatest effect on strength and is very similar, in terms of elements and their coefficients, to an equation developed to predict the strength of austenitic stainless steel wrought material (Ref 6). Thus, it can be concluded that:

- AUSTENITIC WELD STRENGTH CAN BE ADJUSTED BY CHANGING THE COMPOSITION TO MEET THE NEEDS OF THE APPLICATION.
- WELD STRENGTH IS DETERMINED BY THE SAME FACTORS AS THOSE THAT DETERMINE THE BASE METAL STRENGTH.
- MATCHING THE COMPOSITION OF THE WELD TO THAT OF THE WROUGHT MATERIAL WILL RESULT IN NEARLY EQUIVALENT STRENGTHS.

Fracture Toughness at 4 K. Because many cryogenic structures are subject to fracture and other failure mechanisms, most designs must consider toughness, as well as strength. Unfortunately, toughness often degrades as the strength is increased or as the temperature decreases.

Figure 2 shows data for strength and fracture toughness at 4 K for some alloys that have been either proposed or actually used in cryogenic applications. The plotting of yield strength versus fracture toughness, K_{Ic}, provides an easy way to evaluate alloys, because these mechanical properties are common design criteria. A trend band or border was drawn around the data when warranted by the quantity of data. For some alloys, the data were so sparse that individual data points were plotted. Although most of these data are for base-metal alloys, data for selected welds are included to show that base-metal properties for important cryogenic alloys can be met or exceeded for high-toughness, fracture-critical applications.

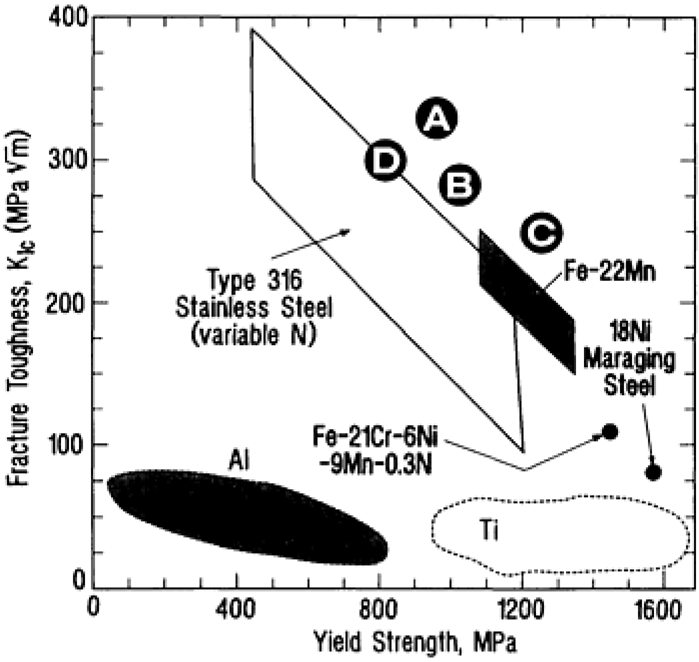

FIG. 2 YIELD STRENGTH VERSUS FRACTURE TOUGHNESS FOR VARIOUS BASE METAL ALLOYS AND SELECTED WELDS AT 4 K

Data for a Ti-6Al-4V alloy and an electron-beam (EB) weld showed an abrupt decrease in toughness at test temperatures below 125 K; toughness was lowest at the boundary between the heat-affected zone (HAZ) and the fusion zone of the weld (Ref 7). These data are grouped in the lower right corner of Fig. 2. Generally, titanium alloys and other ultrahigh-strength alloys, such as 18Ni maraging steel or 9Ni steel, exhibit low toughness at 4 K (as might be expected of an alloy with very high strength). Therefore, these alloys are most commonly used for compression loading or other applications where their limited ductility is acceptable (Ref 8, 9, 10). The 18Ni maraging steel and the titanium alloy do not show a transition from ductile to brittle fracture at low temperatures, as do 9Ni and other lower-nickel-content steels (Ref 1, 2, 9). No data for weld toughness were found for 18Ni maraging steels at 4 K.

Data for aluminum alloys and some welds are also included in Fig. 2 (Ref 11, 12). These data represent traditional aluminum alloys, as well as some of the new aluminum-lithium alloys. The zones indicated in Fig. 2 for the aluminum and titanium alloys are intended to loosely identify mechanical property regions for the alloys, rather than to define limits for design purposes.

At a temperature of 4 K, considerable data exist for type 316L stainless steel wrought alloy (a common cryogenic structural material). Therefore, Fig. 2 includes a trend band for yield strength versus toughness (Ref 6). This alloy has a good balance of strength and toughness, and has found application as the structural case for large superconducting magnets. As its strength increases (through additions of nitrogen), its toughness decreases. This variation in strength with nitrogen content was shown in Fig. 1 as curves for two different nitrogen contents, whereas Fig. 2 shows the effect in terms of toughness as a function of strength at constant temperature. This trend shows the interrelationships between properties that must be considered by structural designers.

Although the trend line for the matching-composition electrodes is not included in Fig. 2, it extends over the same strength range, but at only 50 to 70% of the toughness of the base metal (Ref 4). This lower toughness for welding electrodes of matching composition has driven the search for electrodes of different compositions that can match the properties of these base metals. Two fully austenitic stainless steel weld compositions (shown as A and B in Fig. 2 and described in more detail in Table 1) have been proposed for joining AISI 316LN stainless steel (Ref 13, 14, 15).

TABLE 1 COMPOSITIONS FOR THE FOUR WELDS (A THROUGH D) INCLUDED IN FIG. 2

WELD	**COMPOSITION, %**								PROCESS
	FE	**CR**	**NI**	**MN**	**N**	**C**	**S**	**P**	
A	BAL	13	5	22	0.21	0.04	0.004	0.013	**EBW**
B	BAL	20	25	1.6	0.16	0.01	0.001	0.010	**GMAW**
C	BAL	18.1	20.4	5.4	0.16	0.03	0.007	0.006	**GMAW**
D	**1**	**20**	**BAL**	**3**	**. . .**	**0.02**	**. . .**	**. . .**	**GTAW**

Two other iron-base alloys, Fe-22Mn and 21Cr-6Ni-9Mn, have been used in structures that require high strength and toughness at 4 K. Figure 2 includes a trend line for a high-manganese alloy, Fe-22Mn, which is one of several iron-manganese alloys that have been developed since the early 1980s (Ref 3, 16). Its particular advantage is an ability to attain a strength higher than that of type 316 stainless steel. A matching-composition weld, shown as C in Fig. 2, was produced by electron-beam welding (EBW) (Ref 17). The strength of the weld was approximately 5% lower than that of the parent plate material, presumably because of the loss of nitrogen in the welding process. However, the fracture toughness of the weld exceeded the toughness of the plate material. Matching-composition welds have also been made using shielded metal arc welding (SMAW), gas-metal arc welding (GMAW), gas-tungsten arc welding (GTAW), and submerged arc welding (SAW) processes. The toughness of the welds produced by these other processes is lower than that of the EB weld, because their inclusion contents are higher (Ref 16).

Another common cryogenic alloy is 21Cr-6Ni-9Mn, for which one value is included in Fig. 2 (Ref 2). A trend band has not been developed, but the data would be expected to have a similar trend to those of Fe-22Mn and 316LN (but extending to a higher strength than that of 316LN, because the 21Cr-6Ni-9Mn can contain up to 0.4 wt% N). A nickel-base weld composition (weld D in Fig. 2) has been proposed for 21Cr-6Ni-9Mn. Although this nickel-base weld undermatches the strength, it overmatches the toughness. Thus, welds can be made in locations of lower stress, which permits the use of this high-strength alloy in the construction of structures.

Table 1 lists the compositions of the four welds included in Fig. 2. It shows that these welds (other than the Fe-22Mn composition) have more nickel than the iron-base alloys that they join. These weld composition data represent the best combinations of strength and toughness reported for the alloys. Because the matching-composition welds tend to have lower toughness than the parent plate, these nonmatching compositions are being considered for fracture-critical welds. Subsequent sections in this article describe how thermal expansion and other characteristics of the nonmatching compositions might affect the suitability of these alloys in particular applications.

Another common cryogenic temperature is 77 K, the normal boiling temperature of liquid nitrogen. Because toughness is higher at 77 K than at 4 K, higher-strength alloys can be considered for fracture-critical applications. Like other iron-base alloys that show a significant improvement in toughness with increasing temperature, Fe-21Cr-6Ni-9Mn is more suited for use at 77 K, where its high strength is an attractive option, when compared with that of 316LN alloy. For example, Fe-

21Cr-6Ni-9Mn base metal drops in strength from a level near 1450 MPa (210 ksi) at 4 K to near 1150 MPa (165 ksi) at 77 K, which is still a very respectable level of strength.

Charpy V-Notch Impact Energy at 77 K. Although not as useful in fitness-for-service calculations as K_{Ic} fracture toughness, the Charpy V-notch (CVN) test is a low-cost way to measure the resistance of an alloy to fracture. A literature search found that most data on fracture resistance at 77 K were reported as CVN data. In Fig. 3, CVN absorbed-energy data at 77 K are plotted versus the yield strength for a number of weld metals. The trends reiterate the effects of strength and nickel content on CVN toughness. Because many structural designs specify a minimum of 30 or 40 J (22 to 30 ft · lbf), numerous weld compositions can be considered at this temperature, including compositions that match that of the base plate.

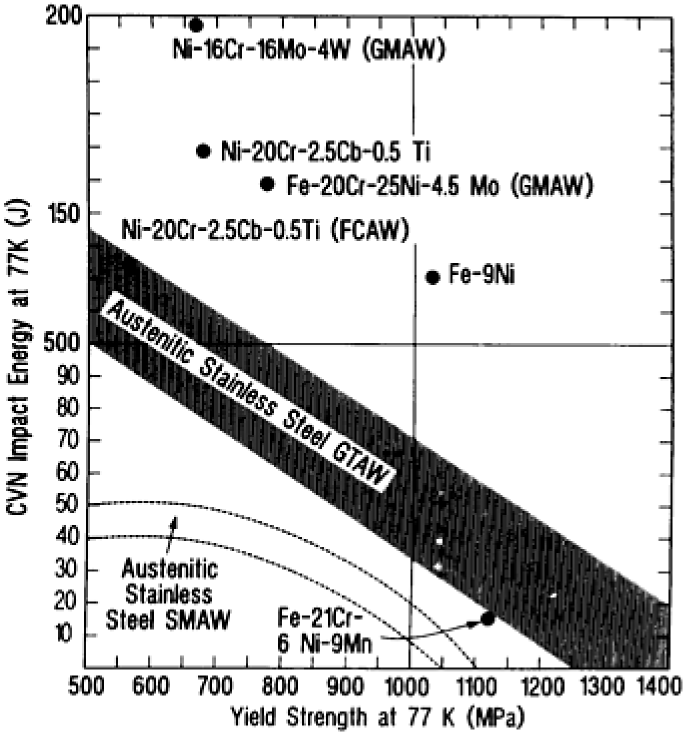

FIG. 3 CHARPY V-NOTCH ABSORBED ENERGY VERSUS STRENGTH FOR VARIOUS WELD METALS AT 77 K

In the lower right quadrant of Fig. 3 is a data point for a matching-composition weld for the Fe-21Cr-6Ni-9Mn steel (Ref 18). The low absorbedenergy indicates that the nickel-base undermatching weld compositions should be considered for applications with this composition at 77 K (Ref 19). Nickel-base weld compositions are also used for welding the Fe-9Ni alloy included in Fig. 3 (Ref 10).

For the austenitic (300 series) stainless steels, there are two trend bands. The lower band is for welds produced by the SMAW process, whereas the upper is for welds produced by the GTAW process. The difference is attributed to weld inclusions, as discussed in the next section of this article. Above the GTAW trend band are the four special weld compositions identified in Table 1. These higher absorbed energy levels are comparable to the level that type 316LN plate would develop at this temperature.

Equation 2 was developed to predict the absorbed energy at 77 K for the welds in austenitic stainless steel:

$$\mathrm{CVN(J)} = 19 - 1.4\ \mathrm{FN} - 890\ \mathrm{C}^2 + 1.4\ \mathrm{NI} \qquad \textbf{(EQ 2)}$$

where the elemental symbols represent the composition in wt%, and FN is a measure of the delta ferrite content (calculated by the Schaeffler diagram coefficients and allowed to be negative) in the weld (Ref 20).

This equation was not developed for weld compositions strengthened by nitrogen. Therefore, carbon represents the principal strengthening element. If carbon is considered as a generic strengthening term (combined effects of carbon, nitrogen, manganese, molybdenum, and others), then Eq 2 shows that increased strength and ferrite content lower the CVN toughness, and that nickel increases toughness. The effects of these factors on toughness are generally accepted to be true for stainless steel weld and base material. Because an increase in nickel content also decreases the ferrite content, nickel additions improve the toughness via two terms in Eq 2. This information on the effect of nickel confirms the reasons for the better toughness of the four nonmatching compositions (Ref 13, 14, 15, 17).

Inclusion Content. In austenitic stainless steels and manganese stainless steels, the fracture modes of both base metals and welds are usually quite ductile, and fracture originates at inclusions in the microstructure. A minor difference between the fracture surfaces of the base metal and weld metal is that those of the welds have a finer ductile dimple spacing. This finer spacing has been attributed to the higher inclusion content of the welds and explains why the welds are less tough than those of the matching-composition wrought material (Ref 21, 22, 23).

Figure 4 shows normalized toughness data versus inclusion spacings for both wrought materials and welds (Ref 21, 24). The vertical axis is known as a quality index, and it allows alloys with different strengths and nickel contents to be compared on a scale that has fracture toughness in the numerator (higher values are better). The similarity of the trend line for both sets of data supports the need to reduce the inclusion content of welds that require a higher fracture toughness. Most of the welds in Fig. 2 were produced using welding processes and procedures that minimize the inclusion content. The exceptions here are the nickel-alloy welds and the very high nickel content (greater than 20 wt%) stainless steel welds. Preliminary data indicate that the toughness of these compositions is less sensitive to inclusion content. Therefore, these materials could be more appropriate for welding at higher deposition rates.

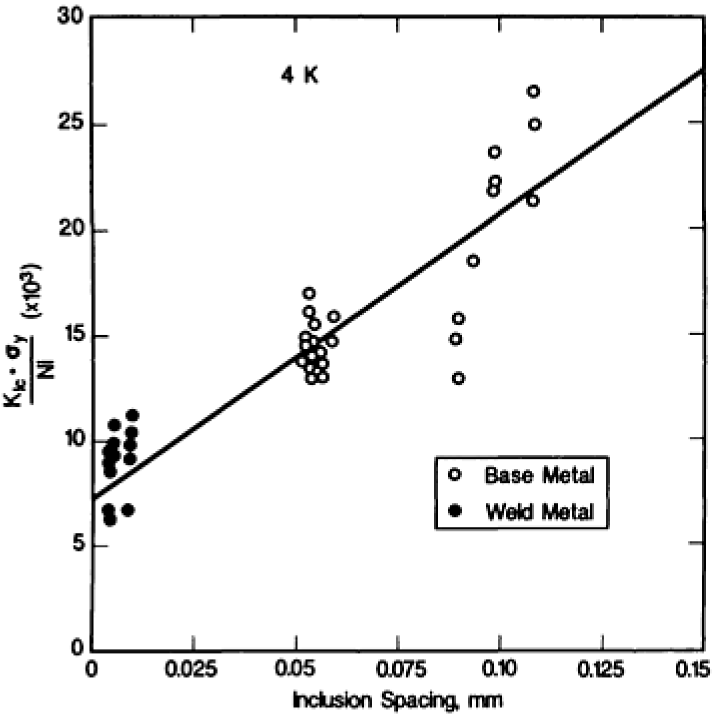

FIG. 4 NORMALIZED TOUGHNESS DATA (QUALITY INDEX) VERSUS INCLUSION SPACING FOR TYPE 316LN BASE METALS AND WELDS, SHOWING HOW THE WELD DATA FOLLOW THE SAME TREND AS THE BASE METAL, BUT A SMALLER INCLUSION SPACING RESULTS IN LOWER TOUGHNESS FOR THE WELDS

Fatigue Strength. Few data are available for the fatigue strength of cryogenic welds. However, the data that are available suggest that fatigue strength could limit the structural life in high-cycle applications. Table 2 lists data from a study that used the GTAW process to join a number of different materials (Ref 25). The data show that fatigue strength does not necessarily increase as temperatures decrease from 295 to 4 K. Expected yield strength data, extracted from Fig. 1, have been added for comparison.

TABLE 2 FATIGUE STRENGTH FOR GAS-TUNGSTEN ARC WELDS AT 4×10^4 CYCLES

SPECIMEN	**ESTIMATED FATIGUE STRENGTH**				YIELD STRENGTH AT 4 K	
	AT 295 K		**AT 4 K**			
	MPA	**KSI**	**MPA**	**KSI**	**MPA**	KSI
21CR-6NI-9MN TO 304L	380	55	360	52	1050	**150**
21CR-6NI-9MN TO 316L	420	60	390	57	1100	**160**
21CR-6NI-9MN BASE METAL	**370**	**54**	**290**	**42**	**1200**	**175**

The coefficient of thermal expansion (CTE) can have an important bearing on structural integrity. If the weld metal has a CTE that is very different from that of the base metal, cooling from room temperature to cryogenic temperature can introduce substantial stresses. These stresses can add to the residual stresses that exist around any weld and can permit the plastic deformation of the structure. Repeated thermal cycles could lead to work hardening of the microstructure or to fatigue failure, in the most severe cases. Other physical properties that are often important in cryogenic structures are phase stability and magnetic permeability.

References cited in this section

1. MECHANICAL, THERMAL, ELECTRICAL, AND MAGNETIC PROPERTIES OF STRUCTURAL MATERIALS, *HANDBOOK ON MATERIALS FOR SUPERCONDUCTING MACHINERY,* 1ST ED., MCIC-HB-04, METALS AND CERAMICS INFORMATION CENTER, BATTELLE COLUMBUS LABORATORIES, JAN 1977
2. "MATERIALS FOR SUPERCONDUCTING MAGNET SYSTEMS: MECHANICAL AND PHYSICAL PROPERTIES DATA AT LOW TEMPERATURES," BELFOUR STULEN DIVISION, TRAVERSE CITY, MI, 1979
3. T. HORIUCHI, R. OGAWA, AND M. SHIMADA, CRYOGENIC FE-MN AUSTENITIC STEELS, *ADV. CRYOGENIC ENG. (MATER.),* VOL 32, 1985
4. R.L. TOBLER, T.A. SIEWERT, AND H.I. MCHENRY, STRENGTH-TOUGHNESS RELATIONSHIP OF AUSTENITIC STAINLESS STEEL WELDS AT 4 K, *CRYOGENICS,* VOL 26, 1986
5. C.N. MCCOWAN AND T.A. SIEWERT, INFLUENCE OF MOLYBDENUM ON THE STRENGTH AND TOUGHNESS OF STAINLESS STEEL WELDS FOR CRYOGENIC SERVICE, "MATERIALS STUDIES FOR MAGNETIC FUSION ENERGY APPLICATIONS AT LOW TEMPERATURES--X," NBSIR 87-3067, NATIONAL BUREAU OF STANDARDS, 1987
6. N.J. SIMON AND R.P. REED, STRENGTH AND TOUGHNESS OF AISI 304 AND 316 AT 4 K, "MATERIALS STUDIES FOR MAGNETIC FUSION ENERGY APPLICATIONS AT LOW TEMPERATURES--X," NBSIR 86-3050, NATIONAL BUREAU OF STANDARDS, 1986
7. R.L. TOBLER, *LOW TEMPERATURE FRACTURE BEHAVIOR OF A TI-6AL-4V ALLOY AND ITS ELECTRON BEAM WELDS,* STP 651, ASTM, 1978
8. R.L. TOBLER, R.P. REED, AND R.E. SCHRAMM, CRYOGENIC TENSILE, FATIGUE, AND FRACTURE PARAMETERS FOR A SOLUTION-ANNEALED 18 PERCENT NICKEL MARAGING STEEL, *J. ENG. MATER. TECHNOL.,* VOL 100, APRIL 1978, P 189-194
9. J.A. WAGNER, CORRELATION OF MECHANICAL PROPERTIES WITH METALLURGICAL STRUCTURE FOR 18NI 200 GRADE MARAGING STEEL AT ROOM AND CRYOGENIC TEMPERATURES, *CRYOGENICS,* VOL 31, SEPT 1991
10. R.D. STOUT AND S.J. WIERSMA, FRACTURE TOUGHNESS OF MODERN 9% NICKEL CRYOGENIC STEELS, *ADV. CRYOGENIC ENG. (MATER.),* VOL 32, 1985
11. J. GLAZER, S.L. VERASCONI, E.N.C. DALDER, W. YU, R.A. EMIGH, R.O. RICHIE, AND J.W. MORRIS, JR., CRYOGENIC MECHANICAL PROPERTIES OF AL-CU-LI-ZR ALLOY 2090, *ADV. CRYOGENIC ENG. (MATER.),* VOL 32, 1985

12. R.P. REED, P.T. PURTSCHER, N.J. SIMON, J.D. MCCOLSKEY, R.P. WALSH, J.R. BERGER, E.S. DREXLER, AND R.L. SANTOYO, "ALUMINUM ALLOYS FOR ALS CRYOGENIC TANKS: COMPARATIVE MEASUREMENTS OF CRYOGENIC MECHANICAL PROPERTIES OF AL-LI ALLOYS AND ALLOY 2219," REPORT PL-TR--91-3073, PHILLIPS LABORATORY, PROPULSION DIRECTORATE, AIR FORCE SYSTEMS COMMAND, EDWARDS AIR FORCE BASE, 1991
13. C.N. MCCOWAN, T.A. SIEWERT, AND R.L. TOBLER, TENSILE AND FRACTURE PROPERTIES OF AN FE-18CR-20NI-5MN-0.16N FULLY AUSTENITIC WELD METAL AT 4 K, *J. ENG. MATER. TECHNOL.,* VOL 108, 1986
14. T.A. SIEWERT AND C.N. MCCOWAN, THE FRACTURE TOUGHNESS OF 25CR-22NI-4MN-2MO STAINLESS STEEL WELDS AT 4 K, "MATERIALS STUDIES FOR MAGNETIC FUSION ENERGY APPLICATIONS AT LOW TEMPERATURES--XII," NISTIR 3931, NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, 1990
15. A.O. KLUKEN, C.N. MCCOWAN, AND T.A. SIEWERT, EFFECT OF INCLUSION VOLUME FRACTION AND SIZE DISTRIBUTION ON THE CRYOGENIC TOUGHNESS OF AUSTENITIC STAINLESS STEEL WELD METALS, SUBMITTED FOR PUBLICATION IN *MICROSTRUCTURAL SCIENCE,* VOL 19, ASM INTERNATIONAL/INTERNATIONAL METALLOGRAPHIC SOCIETY, 1993
16. Y.W. CHENG, H.I. MCHENRY, P.N. LI, T. INOUE, AND T. OGAWA, FRACTURE TOUGHNESS OF 25MN AUSTENITIC STEEL WELDMENTS AT 4 K, *ADV. CRYOGENIC ENG. (MATER.),* VOL 30, 1983
17. S. TONE, M. HIROMATSU, J. NUMATA, T. HORIUCHI, H. NAKAJIMA, AND S. SHIMAMOTO, CRYOGENIC PROPERTIES OF ELECTRON-BEAM WELDED JOINTS IN A 22MN-13CR-5NI-0.22N AUSTENITIC STAINLESS STEEL, *ADV. CRYOGENIC ENG. (MATER.),* VOL 32, 1985
18. R.H. ESPY, WELDABILITY OF 21-6-9 STAINLESS STEEL (ARMCO NITRONIC STAINLESS STEEL), "MATERIALS STUDIES FOR MAGNETIC FUSION ENERGY APPLICATIONS AT LOW TEMPERATURES--I, " NBSIR 78-884, NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, 1978
19. D.J. ALEXANDER AND G.M. GOODWIN, THICK-SECTION WELDMENTS IN 21-6-9 AND 316LN STAINLESS STEEL FOR FUSION ENERGY APPLICATIONS, SUBMITTED TO *J. NUCL. MATER.,* 1992
20. T.A. SIEWERT, PREDICTING THE TOUGHNESS OF SMA AUSTENITIC STAINLESS STEEL WELDS AT 77 K, *WELD. J.,* VOL 65, 1986
21. C.N. MCCOWAN AND T.A. SIEWERT, FRACTURE TOUGHNESS OF 316L STAINLESS STEEL WELDS WITH VARYING INCLUSION CONTENTS AT 4 K, *ADV. CRYOGENIC ENG. (MATER.),* VOL 34, 1989
22. J.H. KIM, B.W. OH, J.G. YOUN, G.-W. BAHNG, AND H.-M. LEE, EFFECT OF OXYGEN CONTENT ON CRYOGENIC TOUGHNESS OF AUSTENITIC STAINLESS STEEL WELD METAL, *ADV. CRYOGENIC ENG. (MATER.),* VOL 34, 1989
23. N. YAMAGAMI, Y. KOHSAKA, AND C. OUCHI, MECHANICAL PROPERTIES OF WELDED JOINTS IN MN-CR AND NI-CR STAINLESS STEELS AT 4 K, *ADV. CRYOGENIC ENG. (MATER.),* VOL 34, 1989
24. C.N. MCCOWAN AND T.A. SIEWERT, INCLUSIONS AND FRACTURE TOUGHNESS IN STAINLESS STEEL WELDS AT 4 K, "MATERIALS STUDIES FOR MAGNETIC FUSION ENERGY APPLICATIONS AT LOW TEMPERATURES--XI," NBSIR 88-3082, NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, 1988
25. T.A. SIEWERT, C.N. MCCOWAN, AND D.P. VIGLIOTTI, CRYOGENIC MATERIAL PROPERTIES OF STAINLESS STEEL TUBE-TO-FLANGE WELDS, *CRYOGENICS,* VOL 30, 1990

References

1. MECHANICAL, THERMAL, ELECTRICAL, AND MAGNETIC PROPERTIES OF STRUCTURAL MATERIALS, *HANDBOOK ON MATERIALS FOR SUPERCONDUCTING MACHINERY,* 1ST ED., MCIC-HB-04, METALS AND CERAMICS INFORMATION CENTER, BATTELLE COLUMBUS LABORATORIES, JAN 1977
2. "MATERIALS FOR SUPERCONDUCTING MAGNET SYSTEMS: MECHANICAL AND PHYSICAL PROPERTIES DATA AT LOW TEMPERATURES," BELFOUR STULEN DIVISION, TRAVERSE CITY, MI, 1979
3. T. HORIUCHI, R. OGAWA, AND M. SHIMADA, CRYOGENIC FE-MN AUSTENITIC STEELS, *ADV. CRYOGENIC ENG. (MATER.),* VOL 32, 1985
4. R.L. TOBLER, T.A. SIEWERT, AND H.I. MCHENRY, STRENGTH-TOUGHNESS RELATIONSHIP OF AUSTENITIC STAINLESS STEEL WELDS AT 4 K, *CRYOGENICS,* VOL 26, 1986
5. C.N. MCCOWAN AND T.A. SIEWERT, INFLUENCE OF MOLYBDENUM ON THE STRENGTH AND TOUGHNESS OF STAINLESS STEEL WELDS FOR CRYOGENIC SERVICE, "MATERIALS STUDIES FOR MAGNETIC FUSION ENERGY APPLICATIONS AT LOW TEMPERATURES--X," NBSIR 87-3067, NATIONAL BUREAU OF STANDARDS, 1987
6. N.J. SIMON AND R.P. REED, STRENGTH AND TOUGHNESS OF AISI 304 AND 316 AT 4 K, "MATERIALS STUDIES FOR MAGNETIC FUSION ENERGY APPLICATIONS AT LOW TEMPERATURES--X," NBSIR 86-3050, NATIONAL BUREAU OF STANDARDS, 1986
7. R.L. TOBLER, *LOW TEMPERATURE FRACTURE BEHAVIOR OF A TI-6AL-4V ALLOY AND ITS ELECTRON BEAM WELDS,* STP 651, ASTM, 1978
8. R.L. TOBLER, R.P. REED, AND R.E. SCHRAMM, CRYOGENIC TENSILE, FATIGUE, AND FRACTURE PARAMETERS FOR A SOLUTION-ANNEALED 18 PERCENT NICKEL MARAGING STEEL, *J. ENG. MATER. TECHNOL.,* VOL 100, APRIL 1978, P 189-194
9. J.A. WAGNER, CORRELATION OF MECHANICAL PROPERTIES WITH METALLURGICAL STRUCTURE FOR 18NI 200 GRADE MARAGING STEEL AT ROOM AND CRYOGENIC TEMPERATURES, *CRYOGENICS,* VOL 31, SEPT 1991
10. R.D. STOUT AND S.J. WIERSMA, FRACTURE TOUGHNESS OF MODERN 9% NICKEL CRYOGENIC STEELS, *ADV. CRYOGENIC ENG. (MATER.),* VOL 32, 1985
11. J. GLAZER, S.L. VERASCONI, E.N.C. DALDER, W. YU, R.A. EMIGH, R.O. RICHIE, AND J.W. MORRIS, JR., CRYOGENIC MECHANICAL PROPERTIES OF AL-CU-LI-ZR ALLOY 2090, *ADV. CRYOGENIC ENG. (MATER.),* VOL 32, 1985
12. R.P. REED, P.T. PURTSCHER, N.J. SIMON, J.D. MCCOLSKEY, R.P. WALSH, J.R. BERGER, E.S. DREXLER, AND R.L. SANTOYO, "ALUMINUM ALLOYS FOR ALS CRYOGENIC TANKS: COMPARATIVE MEASUREMENTS OF CRYOGENIC MECHANICAL PROPERTIES OF AL-LI ALLOYS AND ALLOY 2219," REPORT PL-TR--91-3073, PHILLIPS LABORATORY, PROPULSION DIRECTORATE, AIR FORCE SYSTEMS COMMAND, EDWARDS AIR FORCE BASE, 1991
13. C.N. MCCOWAN, T.A. SIEWERT, AND R.L. TOBLER, TENSILE AND FRACTURE PROPERTIES OF AN FE-18CR-20NI-5MN-0.16N FULLY AUSTENITIC WELD METAL AT 4 K, *J. ENG. MATER. TECHNOL.,* VOL 108, 1986
14. T.A. SIEWERT AND C.N. MCCOWAN, THE FRACTURE TOUGHNESS OF 25CR-22NI-4MN-2MO STAINLESS STEEL WELDS AT 4 K, "MATERIALS STUDIES FOR MAGNETIC FUSION ENERGY APPLICATIONS AT LOW TEMPERATURES--XII," NISTIR 3931, NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, 1990
15. A.O. KLUKEN, C.N. MCCOWAN, AND T.A. SIEWERT, EFFECT OF INCLUSION VOLUME FRACTION AND SIZE DISTRIBUTION ON THE CRYOGENIC TOUGHNESS OF AUSTENITIC

STAINLESS STEEL WELD METALS, SUBMITTED FOR PUBLICATION IN *MICROSTRUCTURAL SCIENCE,* VOL 19, ASM INTERNATIONAL/INTERNATIONAL METALLOGRAPHIC SOCIETY, 1993

16. Y.W. CHENG, H.I. MCHENRY, P.N. LI, T. INOUE, AND T. OGAWA, FRACTURE TOUGHNESS OF 25MN AUSTENITIC STEEL WELDMENTS AT 4 K, *ADV. CRYOGENIC ENG. (MATER.),* VOL 30, 1983
17. S. TONE, M. HIROMATSU, J. NUMATA, T. HORIUCHI, H. NAKAJIMA, AND S. SHIMAMOTO, CRYOGENIC PROPERTIES OF ELECTRON-BEAM WELDED JOINTS IN A 22MN-13CR-5NI-0.22N AUSTENITIC STAINLESS STEEL, *ADV. CRYOGENIC ENG. (MATER.),* VOL 32, 1985
18. R.H. ESPY, WELDABILITY OF 21-6-9 STAINLESS STEEL (ARMCO NITRONIC STAINLESS STEEL), "MATERIALS STUDIES FOR MAGNETIC FUSION ENERGY APPLICATIONS AT LOW TEMPERATURES--I, " NBSIR 78-884, NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, 1978
19. D.J. ALEXANDER AND G.M. GOODWIN, THICK-SECTION WELDMENTS IN 21-6-9 AND 316LN STAINLESS STEEL FOR FUSION ENERGY APPLICATIONS, SUBMITTED TO *J. NUCL. MATER.*, 1992
20. T.A. SIEWERT, PREDICTING THE TOUGHNESS OF SMA AUSTENITIC STAINLESS STEEL WELDS AT 77 K, *WELD. J.,* VOL 65, 1986
21. C.N. MCCOWAN AND T.A. SIEWERT, FRACTURE TOUGHNESS OF 316L STAINLESS STEEL WELDS WITH VARYING INCLUSION CONTENTS AT 4 K, *ADV. CRYOGENIC ENG. (MATER.),* VOL 34, 1989
22. J.H. KIM, B.W. OH, J.G. YOUN, G.-W. BAHNG, AND H.-M. LEE, EFFECT OF OXYGEN CONTENT ON CRYOGENIC TOUGHNESS OF AUSTENITIC STAINLESS STEEL WELD METAL, *ADV. CRYOGENIC ENG. (MATER.),* VOL 34, 1989
23. N. YAMAGAMI, Y. KOHSAKA, AND C. OUCHI, MECHANICAL PROPERTIES OF WELDED JOINTS IN MN-CR AND NI-CR STAINLESS STEELS AT 4 K, *ADV. CRYOGENIC ENG. (MATER.),* VOL 34, 1989
24. C.N. MCCOWAN AND T.A. SIEWERT, INCLUSIONS AND FRACTURE TOUGHNESS IN STAINLESS STEEL WELDS AT 4 K, "MATERIALS STUDIES FOR MAGNETIC FUSION ENERGY APPLICATIONS AT LOW TEMPERATURES--XI," NBSIR 88-3082, NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, 1988
25. T.A. SIEWERT, C.N. MCCOWAN, AND D.P. VIGLIOTTI, CRYOGENIC MATERIAL PROPERTIES OF STAINLESS STEEL TUBE-TO-FLANGE WELDS, *CRYOGENICS,* VOL 30, 1990

Welding for Cryogenic Service

T.A. Siewert and C.N. McCowan, National Institute of Standards and Technology

---

Selected References

- T. OGAWA AND T. KOSEKI, WELDABILITY OF NEWLY DEVELOPED AUSTENITIC ALLOYS FOR CRYOGENIC SERVICE: PART 2--HIGH-NITROGEN STAINLESS WELD METAL, *WELD. J.,* NOV 1987
- R.P. REED, AUSTENITIC STAINLESS STEELS WITH EMPHASIS ON STRENGTHENING AT LOW TEMPERATURES, *ALLOYING,* J.L. WALTER, M.R. JACKSON, AND C.T. SIMS, ED., ASM INTERNATIONAL, 1988
- R.L. TOBLER, R.P. MIKESELL, AND R.P. REED, *CRYOGENIC EFFECTS ON THE FRACTURE MECHANICS PARAMETERS OF FERRITIC NICKEL ALLOY STEELS,* STP 677, ASTM, 1979
- K.A. YUSHCHENKO, PROGRESS IN CRYOSTRUCTURAL MATERIALS AND THEIR WELDING IN

THE USSR, *ADV. CRYOGENIC ENG. (MATER.)*, VOL 30, 1983

Welding in Space and Low-Gravity Environments

Milo Nance, Martin Marietta Corporation; Jerald E. Jones, Colorado School of Mines

Introduction

WELDING AS AN ASSEMBLY PROCESS has become increasingly more attractive to designers of space structures for the same reasons welding is popular on earth. Welding applications in space are similar to those on earth in that they both involve the need to assemble structural components that will exhibit sufficient strength, endurance, and reliability during their service lives. In addition, ease of repair of welded structures is another consideration that makes various welding methods viable for space applications. Typical space welding applications generally fall into three categories:

- ON-ORBIT CONSTRUCTION OF LARGE STRUCTURES
- REPAIR WELDING ON-ORBIT AND FOR EXPLORATION FLIGHTS BEYOND THE VICINITY OF THE EARTH
- WELDING AND REPAIR OF STRUCTURES ON A LUNAR BASE

This article will review a variety of applications for welding in space and low-gravity environments, describe the unique aspects of the space environment, compare applicable welding processes, and examine the metallurgy of low-gravity welds. In addition, steps that must be taken to ensure the continued development of welding technology in space are discussed.

Welding in Space and Low-Gravity Environments

Milo Nance, Martin Marietta Corporation; Jerald E. Jones, Colorado School of Mines

Examples of Application

Welded joints have the potential to reduce the cost of fabrication and repair on-orbit. For example, large mechanical joints are planned for use on the U.S. Space Station Freedom. Each of these joints weighs approximately 1.36 kg (3 lb). When one considers that the cost for low earth orbit missions is several thousand dollars per pound of payload, it is easy to see how the replacement of these joints with a welded design would save significant amounts of weight and offer the potential to save large amounts of capital.

There are other applications for welding technology in space. These involve the modification and repair of structures and components in orbit. It has been suggested that structural material could be scavenged for reuse in orbit. An example of this is the Space Shuttle external tank (Fig. 1). On each shuttle mission the external tank is brought nearly into orbit. With minimal loss of payload capability, these tanks could be brought into orbit allowing the use of the tank structure for experimental platforms or for raw stock material. Cutting and joining technologies could enable the reuse of this valuable resource.

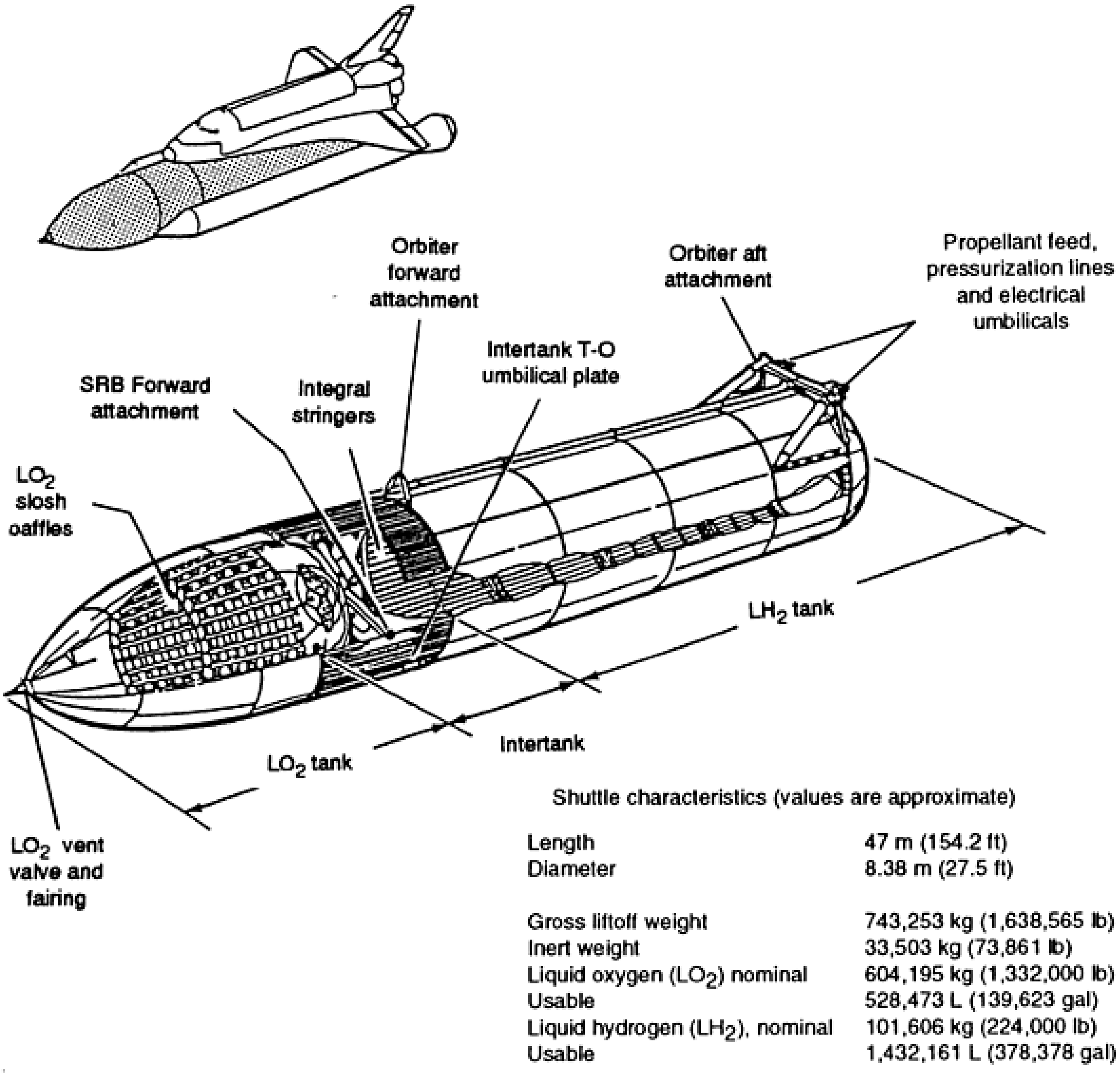
Orbiter forward attachment
Orbiter aft attachment
Propellant feed, pressurization lines and electrical umbilicals
Intertank T-O umbilical plate
SRB Forward attachment
Integral stringers
LO_2 slosh oaffles
LH_2 tank
Intertank
LO_2 tank
LO_2 vent valve and fairing

Shuttle characteristics (values are approximate)

Length	47 m (154.2 ft)
Diameter	8.38 m (27.5 ft)
Gross liftoff weight	743,253 kg (1,638,565 lb)
Inert weight	33,503 kg (73,861 lb)
Liquid oxygen (LO_2) nominal	604,195 kg (1,332,000 lb)
Usable	528,473 L (139,623 gal)
Liquid hydrogen (LH_2), nominal	101,606 kg (224,000 lb)
Usable	1,432,161 L (378,378 gal)

FIG. 1 SPACE SHUTTLE EXTERNAL FUEL TANK

Another application is the repair of major structures. Damage to space structures can result from many causes, including accidental misuse. However, one common cause of damage is "space debris," or miscellaneous small and large pieces of disintegrated satellites, material jettisoned from spacecraft, and other miscellaneous material. This debris is travelling at very high speeds, and may not be in the same orbital path and direction as a spacecraft. A resulting high-speed collision, even with a small piece of material, can cause significant damage. The amount of space debris in low earth orbit is increasing. Figure 2 shows the increase in concentration of debris in orbit from data collected by the Long Duration Exposure Facility (LDEF). The Space Shuttle has already been hit by at least one particle of debris, resulting in a damaged windshield, and there is speculation of impact damage to satellites.

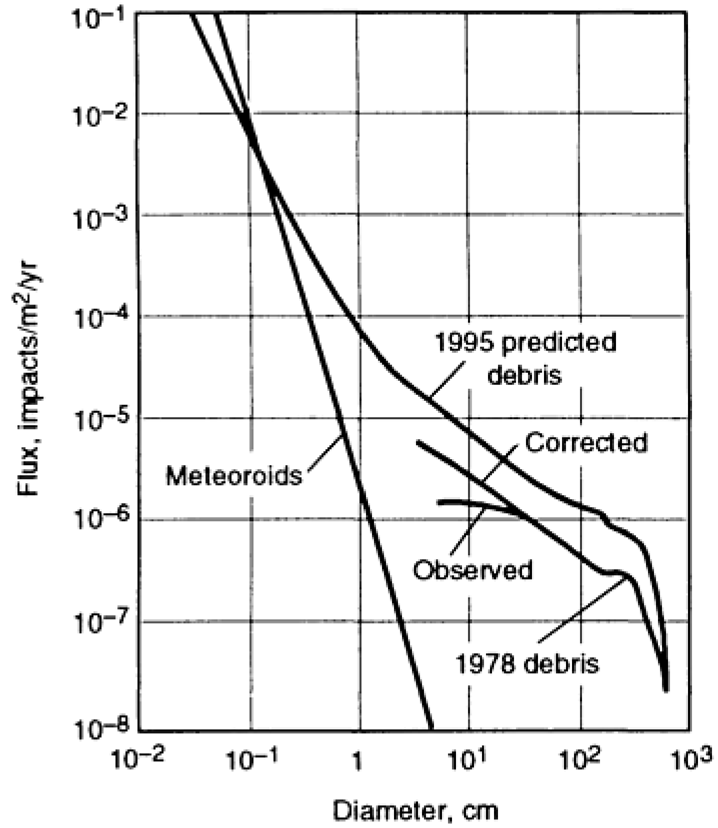

FIG. 2 DIAGRAM SHOWING THE INCREASE IN THE CONCENTRATION OF "SPACE DEBRIS" ENCOUNTERED IN ORBIT. COURTESY OF D.J. KESSLER, NASA JOHNSON SPACE CENTER

On-orbit repair of high-value, long-duration space-based assets has proven to be more effective than to return them to earth. On-orbit welding is a logical technology to assist in the repair of systems such as a space station. In other cases, such as a manned mission to Mars or the vicinity of the moon, it is impractical or impossible to return damaged structures to earth. The Soviet space program already has used welding to repair Soyuz space station tubing.

If a lunar base is established, structures will need to be economically constructed. Welding offers the potential for fast and cost-effective fabrication. In addition, repair of structures and equipment damaged by use and/or meteor collision will be necessary. Lunar dust contains many abrasive components, and wear of equipment will be inevitable. The cost of equipment replacement will probably be prohibitive, while repair by weld hardfacing and surfacing is a very cost-effective technology that is employed in earth-based applications subjected to wear. See the article "Hardfacing, Weld Cladding, and Dissimilar Metal Joining" in this Volume.

Both the United States and the former Soviet Union have carried out welding experiments in space. The United States used Skylab to weld 2219-T87 aluminum alloy, type 304 stainless steel, and pure tantalum samples. The Soviet program used the Soyuz to perform its experiments and still has its space welding technology available. Both efforts involved electron-beam welding. The findings from these experiments are discussed later.

Space Welding Environment

Space presents a unique welding environment and many technological challenges. The most obvious and dramatic factors are microgravity and high vacuum. Microgravity in earth orbit occurs due to the centrifugal force of the orbiting craft that counters the effects of gravity, making the spacecraft and all orbiting objects, including those within the spacecraft, in continual free fall until a force acts upon them. Typical gravitation accelerations are on the order of 10^{-6} G (acceleration due to gravity).

High vacuum is present around the spacecraft. Typical levels in low earth orbit are 1×10^{-4} to 1×10^{-16} Pa. This vacuum aids some welding processes, such as electron beam. Since virtually no atmosphere is present, there is no attenuation caused by scatter of the electron beam. In addition, for all welding processes there is no need to protect the metal from oxidation or other atmospheric contaminants with a vacuum chamber, flux, or an inert gas atmosphere.

The space environment also offers several technological challenges. One of the problems is the lack of a high-capacity electrical power supply. Electrical power generation is currently produced by one of three methods: batteries, fuel cells (power generation from the combination of oxygen and hydrogen to form water), and direct solar conversion. All three of these methods produce a relative small amount of power compared to the requirements of the typical earth-based welding system. Typically, power is limited to 4.5 kW for a period of a few minutes on board the Space Shuttle. This power constraint will be present for the foreseeable future. Larger-capacity generation systems based on thermal-nuclear generators are being developed, but have not yet been used in service with manned spacecraft.

A second problem is the lack of nondestructive testing methods qualified for space use. Many of the traditional methods of inspection will not operate well in a space environment. For example, there is no method of dye penetrant inspection developed to date that can be applied in a vacuum. Methods such as radiography and eddy current inspection may be feasible, but no space-qualified equipment has yet been developed. Even if such equipment was available, the competition for weight and volume onboard a spacecraft might prevent its usage.

The application of intelligent automation techniques that allow the production of high-quality welds with a high degree of reliability appears to be the most attractive alternative with regard to the available space and weight limitations (see the discussion of the future of space welding later in this article). In order to produce such welds, a clear and precise knowledge of the mechanisms involved in space welding must be developed as well as the equipment necessary to carry out such welds. There have been several advances made in the necessary sensor and control technologies for earth-based welding that can be used in space welding.

Milo Nance, Martin Marietta Corporation; Jerald E. Jones, Colorado School of Mines

---

Welding Processes

The space environment has constraints that limit the application of certain conventional welding processes. For example, submerged arc welding (SAW) generally depends on gravity to feed the granulated flux and hold it in place on the weld; this is not possible in low-gravity environments. The space environment also affects some of the basic physical processes involved in welding, making welding processes operate differently in space than in earth-based environments. In addition, the weight limitations of orbital payloads on spacecraft will impose constraints on the processes that can be feasibly used in space.

The three processes that appear to have good potential for space application are electron-beam welding (EBW), laser-beam welding (LBW), and gas-tungsten arc welding (GTAW). All three processes are often used for autogenous welding, thereby eliminating the need for transport of high-weight consumables. These processes are also amenable to welding in vacuum (with certain modifications to the GTAW process as discussed below). Also, all of these processes have relatively low power requirements when used for thin-section materials typical of spacecraft and space structures, making these processes most useful in space, where all utilities, including electricity, are limited.

Electron-Beam Welding (EBW)

The EBW process consists of an electron-beam gun that generates a high-energy stream of electrons. When directed through a series of focusing electromagnets, this high-energy stream has sufficient power density to melt metals. The

EBW process is described in detail in the articles "Electron-Beam Welding" and "Procedure Development and Practice Considerations for Electron-Beam Welding" in this Volume. In earth-based environments, EBW generally requires a vacuum since the electrons scatter in a gas atmosphere. In space, however, the natural vacuum environment is ideally suited for the EBW process.

Both the United States and the former Soviet Union have experimented with EBW and melting in space. The M512 metals melting experiments using an electron-beam gun, which were conducted by the United States in the Skylab in 1972, are described later in this article. In this case, the operations were performed in a vacuum chamber that was inside the spacecraft, but vented to the outside. At the time of the experiments, there was some concern as to the quality of the vacuum achievable. Often, certain gases and liquids are jettisoned from spacecraft that remain in the environment in the local vicinity of the vehicle and can produce contamination of the vacuum.

Advantages. The electron-beam gun is, perhaps, the most efficient of the processes proposed for on-orbit welding. Due to the power limitations of spacecraft, this process is the most capable for welding of thicker-section materials. For autogenous welding, no consumables are required, thereby eliminating the cost of including them in the spacecraft payload.

Limitations. The impingement of electrons onto a metal surface causes the emission of x-rays. The former Soviet Union experimented with a manual electron-beam welding gun for space application. However, x-ray shielding capacity is proportional to the mass of the shield; therefore, shielding of workers may be difficult due to the high cost of payload weight on spacecraft. Automated or robotic welding may be recommended due to the radiation exposure to workers, particularly for extended periods of welding activity.

Safety Considerations. Grounding of the part to be welded is very important. Without proper grounding, the part being welded will acquire a net negative charge from the absorbed electrons. This will cause the beam to be deflected and the weld will not form properly. In addition, if the part becomes charged, then it will discharge to anything that it comes in contact with. Such a discharge could be very dangerous for a worker, or could damage electronic equipment such as computers or communications equipment onboard the spacecraft. Although the inherent repulsion of the like-charged electrons will cause beam spread, the beam will remain a fairly concentrated heat source for some distance from the electron-beam gun. Care will be required, both for manual as well as automatic or robotic welds, to ensure that the gun is not activated except when directed at properly prepared components to be welded. Significant damage could occur to workers, to component parts, or to the spacecraft itself if the gun were activated when directed improperly.

Laser-Beam Welding

The LBW process consists of a laser and a series of lenses or mirrors that direct and focus the beam at the component to be welded. The process is described in detail in the articles "Laser-Beam Welding" and "Procedure Development and Practice Considerations for Laser-Beam Welding" in this Volume. While the LBW process, unlike the EBW process, does not require a vacuum, the inherent vacuum environment in space will help to prevent oxidation and other contamination of the weld. There are two types of lasers used for welding, gas (CO_2) and solid state (Nd:YAG). While the solid-state lasers are relatively low power, this power level is adequate for thin-section components that are likely to be welded in space.

Advantages. As with EBW, the LBW process provides a concentrated heat source that can quickly melt through thin materials without heating large sections of the workpiece. Thus, the necessary power is less than that required for many other processes. However, the laser system is inherently inefficient in the conversion of energy so the input power requirements will be much greater than for an equivalent output heating capacity when compared to EBW. Also, for autogenous welding, no consumables are necessary for LBW, so the payload cost will be low. In addition, the LBW process, unlike EBW, can be used to weld electrically nonconducting materials, including ceramics, glasses, and polymers.

Limitations. The inefficiency of the laser system is due to a significant amount of heat generated in the lasing process itself. This heat will need to be dissipated through the spacecraft radiators or using separate radiators for the laser system. Because heat radiation from low-temperature systems is fairly slow, this may require large radiators to be installed for the laser system; thus the cost and payload weight of the system is increased. Secondly, due to the power input and heat dissipation requirements, it is difficult to produce a laser system for portable application. Recent work with Nd:YAG lasers and fiber optics has resulted in optical fiber delivery of the laser beam at a location remote to the laser system.

However, the laser system generally should be approximately 10 to 30 m (30 to 100 ft) from the LBW torch. Also, great care must be taken to avoid any damage to the cable containing the optical fiber.

Safety Considerations. A laser beam will not spread and dissipate with distance as will an electron beam. Consequently, a misdirected laser beam has the potential to cause damage, even over large distances to workers, components, and to the spacecraft. The laser light itself is dangerous to workers if reflected into the eyes. Proper shielding for persons working around laser welding equipment is necessary.

Gas-Tungsten Arc Welding

The GTAW process consists of a tungsten electrode connected to one lead of a power source, the other lead of which is connected to the workpiece. When the power supply is activated, current flows to the tungsten electrode through an arc plasma to the workpiece, and returns to the power supply to complete the circuit. The arc plasma is the primary source of heat as electrons move from the metal surface, are accelerated through the plasma column, and impinge on the metal surface at the opposite side of the arc. This welding process is described in the article "Gas-Tungsten Arc Welding" in this Volume.

The arc plasma consists of a superheated gas of ions and electrons. To shield the weld from atmospheric contaminants and to provide a stable plasma, inert gas is used. Typically argon, helium, or a mixture of these is chosen for GTAW. Because of the vacuum in the space welding environment, it is not necessary to shield the weld from contaminants; however, a small amount of gaseous material is necessary to sustain a stable arc plasma. The GTAW torch to be used for space welding will be modified from the torch used for earth-based welding. Through the use of a hollow tungsten electrode, the small amount of gas needed to sustain the arc plasma can be provided. A non-reactive gas or gas mixture will be necessary to avoid contamination of the weld. This modified GTAW process is similar to the earth-based PAW process.

Advantages. The equipment used for GTAW is simpler and much less expensive to produce and maintain in comparison to EBW and LBW. Secondly, the arc plasma only exists as a part of the overall welding power circuit, therefore it is nearly impossible to direct the GTAW arc toward a component, worker, or spacecraft that is not connected to this power circuit. Thirdly, as the length of the plasma increases, the voltage required to initiate and maintain the arc increases. By limiting the available voltage in the welding power circuit, the distance that the arc can strike is limited, and the possible damage resulting from a misdirected arc is limited to near-vicinity (25 to 50 mm, or 1 to 2 in.) objects.

Limitations. The GTAW arc plasma requires a constant flow of gas. This gas supply, in the form of fairly heavy compressed gas cylinders, adds to the cost and weight of the spacecraft payload. In addition, this supply will need to be constantly replenished as welding progresses. If resupply is not available, the process will not be available for use. With both LBW and EBW, the welding process would always be available regardless of a resupplied source of welding gas.

The arc plasma is a much less concentrated heat source in comparison to EBW and LBW. This results in the need for a higher power feed from the spacecraft to produce a weld in the same thickness material.

Safety Considerations. The primary concern with arc welding is grounding and insulation for workers. Should the worker become part of the welding circuit, the power supply will generally be capable of delivering several kilowatts through the worker at a current level that can cause permanent disability and/or death. Secondly, in order to ionize the inert gas plasma and initiate the arc, a high voltage is usually applied. This is accomplished through the use of a high voltage pulse, or a superimposed high-frequency high-peak voltage signal. Either of these generates a significant amount of electromagnetic interference (EMI) radiation, which can result in damage or shutdown of such equipment as computers and other advanced electronics onboard the spacecraft.

Other Joining Processes

With a few exceptions (for example, SAW), virtually all welding processes could be used in space. The EBW, LBW, and GTAW generally have been based on weight limitations or on the inherent aspects of the space environment (for example, vacuum for EBW). The materials likely to be found in space that would be welded are typically welded with EBW, LBW, or GTAW in the earth-based environment. These include aluminum and titanium and their alloys. In addition, these processes are well understood, commonly used, and very controllable for thin-section aerospace applications.

Another process that takes advantage of the space environment is solar heat welding. Solar radiation in space is unshielded by the atmosphere. A spacecraft in low earth orbit typically receives 1.4 kW/m^2 (0.13 kW/ft^2) in daylight conditions. This energy can be taken advantage of by focusing the solar radiation on the workpiece through the use of a large collecting mirror and focusing lenses. Because the optics of the mirror need not be perfect, the cost would be reasonable and the system would not be a drain on the limited power capacity of the spacecraft.

Other joining processes that could be utilized include mechanical fastening, solid-state bonding, brazing, and adhesive bonding. However, each of these has inherent limitations that tend to make the three welding processes discussed above more attractive. For example, it may be difficult to use adhesives because they require curing by the application of heat and/or the use of hardening (curing) agents. As a result, adhesives may outgas and contaminate the local environment with volatile components. Mechanical fastening has been used extensively for some joining applications in space. However, mechanical fastening is generally not capable of producing pressurized seals, and mechanically fastened joints are more susceptible to external mechanical damage than are welds.

Welding in Space and Low-Gravity Environments

Milo Nance, Martin Marietta Corporation; Jerald E. Jones, Colorado School of Mines

---

Metallurgy of Low-Gravity Welds

Influences of Microgravity. The microgravity environment that exists in space applications can have significant effect on the metallurgy of welds. Two important effects on the microstructure of a weld are cooling and fluid flow within the weld pool. In addition, certain other effects can be evident, including the deformation of the weld pool that occurs due to absence of gravity in the space environment. In the latter case, large weld pools in through-penetration or out-of-position welds will not droop or run, but instead their shape will be determined by surface energy and other forces.

Both the cooling rate and the liquid pool stirring behavior are influenced by gravity in earth-based welds. In space, the cooling rate will be different because convective flow is completely suppressed, and this affects the stirring forces in the pool. The principal heat loss from a weld is conduction into the surrounding solid metal. Because there is a significant temperature difference between the weld pool and its surroundings during the time the weld pool is at an elevated temperature, radiative cooling is also important. Generally, convective cooling from the weld surface is only a secondary cooling mechanism until the weld has dropped to a temperature at which little other cooling is occurring and further metallurgical transformation is unlikely to occur.

Both the conduction of heat into the surrounding metal and the radiation of heat from the weld pool are directly affected by the local surface temperatures of the weld pool. This includes the exposed surfaces as well as the interface between the liquid pool and the solid base metal. The conductive heat flow is governed by:

$$\boldsymbol{Q} = K \nabla \boldsymbol{T}$$

where $\nabla \boldsymbol{T}$ is the temperature difference vector between the liquid metal and the solid metal at the interface and $\boldsymbol{q}$ is the heat flux vector. This same equation governs the loss of heat by convection from the exposed surface of the weld. However, the value of k is significantly less for heat flow across an interface into a gaseous medium than into a solid metal medium, and the k value for the convective heat transfer in a gas is much less still.

The equation for radiative heat flow is:

$$\boldsymbol{Q} = E\sigma(\nabla \boldsymbol{T})^4$$

where $\boldsymbol{e}$ is the emissivity of the emitting surface and σ is the Steffan-Boltzmann constant. Because the $\nabla \boldsymbol{T}$ term is raised to the fourth power when the temperature difference vector is large, radiation heat transfer is a significant heat loss mechanism for the weld. It should be noted that the radiation of heat from the weld as well as the weld heat source may cause surrounding surfaces to heat, particularly if the weld is made in a vacuum such as outside the spacecraft environment. The result might be damage to exposed components that are heat sensitive if appropriate shielding is not provided.

If convective pool stirring is suppressed, the temperature difference at the liquid-solid interface decreases. The liquid at that surface cools by conduction and the flow of hot liquid from the center of the pool is reduced. The result is a reduced cooling rate for the weld. This reduced cooling rate should tend to cause the grain size in the weld to increase. The opposite effect has, however, been observed and reported by investigators of space-based weld experiments.

The absence of gravitational forces also suppresses any buoyancy effects. Because of buoyancy forces, small particles of contaminants in earth-based welds will tend to either float to the weld surface or sink to the weld bottom depending on their density relative to the weld pool liquid. Surface energy effects may cause these particles to accumulate near surfaces. In low-gravity environments, it would be expected that these particles would remain suspended in the liquid pool as it solidifies. If these particles are good nucleants for metallurgical transformations in the weld metal, those transformations would tend to be enhanced. The result would be a difference in the microstructure between earth-based and space-based welds.

Differences between the distribution of inclusion particles in earth-based versus space-based welds has been observed. The effects of buoyancy and of cooling rate on actual welds were examined based on experiments conducted on Skylab in June of 1972. Some of those effects are discussed below.

Russian Welding Program. Space welding has long been a priority for the space program of the former Soviet Union. In 1969 Soyuz-6 performed the first welds in space on AMG6 and DM-20 aluminum alloys using the Vulkan apparatus. This work was carried on by the E.O. Paton Welding Institute of the Ukraine SSR Academy of Sciences. This work was conducted to develop methods of space repair and construction. Soon thereafter, the United States performed space welding experiments onboard the first manned Skylab Mission in June of 1973. Russian welding efforts have continued, and a portable EBW system has been developed. This hand-held EBW system is currently available for use on aircraft on the Mir space station.

United States M512 Experiments. The M551 EBW experiment was conducted in 1973 by astronaut Pete Conrad inside the multiple docking adapter of Skylab. Figures 3 and 4 show the Skylab spacecraft and the M512 material melting apparatus, respectively. The objectives of the M551 experiment, as described by McKannan and Poorman (Ref 1), were to demonstrate the feasibility of EBW in space and to investigate the differences caused by microgravity. The three materials tested included 2219T87 aluminum alloy, 304 type stainless steel, and tantalum. The aluminum and stainless steel alloys were chosen due to their commonality of use for aerospace applications; the tantalum was chosen as a nearly pure material with no complex phase transformations that would provide information on grain size and other physical and mechanical influences by microgravity. The density, melting point, thermal conductivity, and composition of these base metals are listed in Table 1. In this experiment, an electron beam was used to cut and also produce a weld bead on a 165 mm (6.5 in.) diam rotating disk that varied in thickness from 0.64 to 6.4 mm (0.025 to 0.250 in.) for the stainless steel and aluminum, and from 0.43 to 1.57 mm (0.017 to 0.062 in.) for the tantalum. Finally, the beam was directed at a stationary point on the disk to make a single dwell weld and tO produce a large molten pool. The electron-beam gun was held a constant distance of 38 mm (1.5 in.) with beam adjusted to 80 mA (70 mA for the tantalum sample) and 20 kV. These samples were then compared to samples welded with a duplicate earthbound apparatus using the same welding parameters. All the samples from this experiment were examined in depth at the Battelle Memorial Institute (Ref 2) and at NASA laboratories. In addition, several other organizations received samples for testing and evaluation.

TABLE 1 PROPERTIES AND COMPOSITIONS OF BASE METALS STUDIED IN THE M512 MELTING EXPERIMENTS

ALLOY	**DENSITY, G/CM^3**	**MELTING POINT**		**THERMAL CONDUCTIVITY**		COMPOSITION, WT%															
		°C	**°F**	**W/M · K**	**CAL/CM · S · °C**	**Al**	**C**	**Nb**	**Cr**	**Cu**	**Fe**	**Mg**	**Mn**	**Mo**	**Ni**	**P**	**S**	**Si**	**Ta**	**Ti**	Zn
2219-T87 ALUMINUM	2.83	638	1180	251	0.6	BAL	. . .	. . .	. . .	6.2	0.25	0.02	0.3	. . .	. . .	. . .	. . .	0.15	. . .	0.15	**0.10**
TYPE 304 STAINLESS STEEL	8.03	1442	2628	159	0.38	. . .	0.05	. . .	19.0	0.5	BAL	. . .	1.5	0.5	8.5	0.04	0.03	0.75	. . .	. . .	**. . .**
TANTALUM	**16.6**	**2996**	**5425**	**54.4**	**0.13**	**. . .**	**. . .**	**0.1**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**. . .**	**BAL**	**. . .**	**. . .**

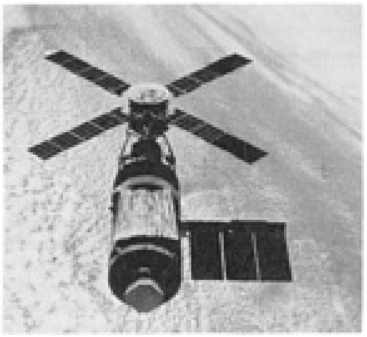

FIG. 3 PHOTOGRAPH OF SKYLAB, WHICH HOUSED THE M512 METALS MELTING EXPERIMENT

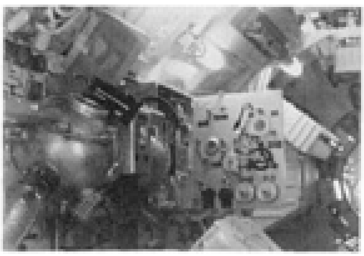

FIG. 4 PHOTOGRAPH OF THE M512 MELTING APPARATUS TAKEN AT THE AIR AND SPACE MUSEUM IN WASHINGTON D.C.

The metallurgical study concluded that electron-beam space welding was feasible. However, distinct changes in weld morphology and metallurgy were found. These are reported in Table 2.

TABLE 2 DIFFERENCES IN SPACED-BASED (SKYLAB) AND EARTH-BASED WELD SAMPLES

MATERIAL	CHARACTERISTICS OF SKYLAB SAMPLE
2219-T87 ALUMINUM	**INCREASED OXIDE ON BACK SIDE OF WELD** **GRAIN REFINEMENT IN WELD** **REDUCTION IN SURFACE RIPPLE, BEADING, AND SPATTER** **INCREASED NUGGET SYMMETRY**
304 STAINLESS STEEL	**GRAIN AND SUBGRAIN REFINEMENT** **DECREASED BANDING** **DECREASED WELD POOL SAG** **REDUCTION IN SURFACE RIPPLE, BEADING, AND SPATTER** **INCREASED NUGGET SYMMETRY**
TANTALUM	**GRAIN REFINEMENT** **DECREASED WELD POOL SAG** **INCREASED NUGGET SYMMETRY**

A subsequent investigation by Nance and Jones at the Colorado School of Mines has shown that both the grain size and the inclusion distribution were affected by the microgravity environment. This study concentrated on the tantalum material. It had been previously hypothesized that the grain size of the weld made in space would be larger due to the reduced cooling rate of the liquid pool. This hypothesis was contradicted by the observation of a 23% reduction in grain size observed in the Skylab tantalum welds in comparison to the earth-based welds. Figures 5(a) and 5(b) compare the grain size of the earth-based and Skylab tantalum welds. A similar grain size reduction was also noted in the aluminum and stainless steel alloys. The mechanism for grain size refinement in these alloys is complex, and it is difficult to draw direct conclusions. For the pure tantalum samples, there were no solid-state transformations; however, the grain size change may be attributable to an inoculant nucleation mechanism.

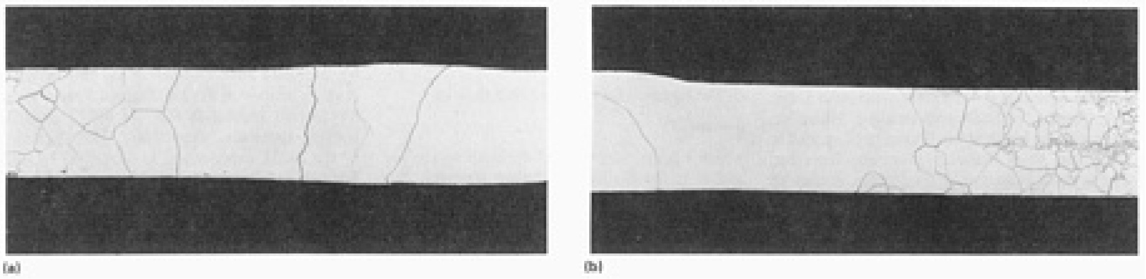

FIG. 5 COMPARISON OF (A) GROUND-BASED AND (B) SKYLAB-BASED WELDS IN PURE TANTALUM SAMPLES. NOTE THE LARGER GRAIN SIZE IN THE GROUND-BASED WELD. SOURCE: REF 2

Utilizing scanning electron microscopy (SEM) with energy dispersive x-ray analysis and scanning Auger spectroscopy (SAS), it was determined that oxide inclusions were well distributed throughout the Skylab welds. In the earth-based welds, these oxide particles, possibly due to the influence of gravity-induced buoyant forces, had floated to the top surface of the weld pool. Figure 6(a) is a SEM micrograph of the Skylab tantalum weld. There is a large carbide particle near the center of the SEM photomicrograph, as confirmed by the SAS carbon map shown in Fig. 6(b). Figure 6(c) is the SAS oxygen map, showing only a small accumulation of oxide particles near the surface of the weld, with a well-distributed oxide particle dispersion in the center of the weld. The earth-based weld showed virtually no oxide particles in the middle of the weld and a very heavy accumulation of oxides at the weld surface.

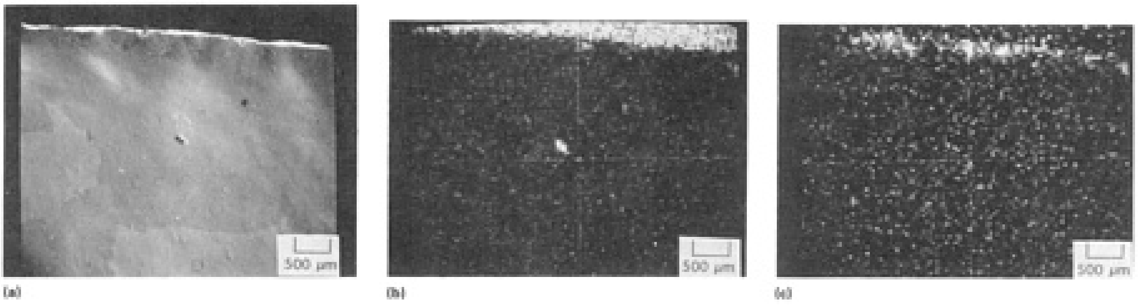

FIG. 6 MATERIALS CHARACTERIZATION TECHNIQUES FOR EXAMINING A TANTALUM WELD CARRIED OUT IN SKYLAB. (A) SEM MICROGRAPH SHOWING A LARGE CARBIDE PARTICLE NEAR THE CENTER OF THE SAMPLE. (B) SCANNING AUGER SPECTROSCOPIC CARBON MAP SHOWING A LARGE CARBIDE PARTICLE NEAR THE CENTER OF THE SAMPLE. (C) SCANNING AUGER SPECTROSCOPIC OXYGEN MAP SHOWING ONLY A SMALL ACCUMULATION OF OXIDE PARTICLES NEAR THE SURFACE OF THE WELD WITH A WELL-DISTRIBUTED OXIDE PARTICLE DISPERSION IN THE CENTER OF THE WELD

References cited in this section

1. E.C. MCKANNAN AND R.M. POORMAN, PAPER 35812, GEORGE C. MARSHALL SPACE FLIGHT CENTER, ALABAMA, 1974, P 20
2. R.E. MONROE, "CHARACTERIZATION OF METALS MELTING RISKS: SKYLAB EXPERIMENT M551," NASA CONTRACTOR REPORT, BATTELLE MEMORIAL INSTITUTE, COLUMBUS, OHIO, 4 DEC 1973, P 109

Welding in Space and Low-Gravity Environments

Milo Nance, Martin Marietta Corporation; Jerald E. Jones, Colorado School of Mines

Future of Space Welding

The most important space-based welding application for long-duration manned spacecraft is repair welding. This technology offers the possibility of decreasing assembly time and decreasing the weight and cost of space structures. However, before welding technology can become commonplace in space, several steps must be taken. These involve the development of processes and controls that allow highly reliable welds to be performed in space. These steps include:

- DEVELOPMENT OF EQUIPMENT THAT HAS SUFFICIENT SENSOR TECHNOLOGY TO REPORT IN REAL TIME THE QUALITY OF WELDS AND TO FACILITATE REAL-TIME PROCESS CONTROL AND CORRECTION OF WELDS IN PROGRESS
- CHARACTERIZATION OF THE DIFFERENCES IN SPACE-BASED WELDS AND EQUIVALENT EARTH-BASED WELDS TO PROVIDE A BASIS FOR MODELING AND UNDERSTANDING OF THE DIFFERENCES IN THESE WELDS
- PROCESS DEVELOPMENT FOR WELDS CONDUCTED IN SPACE

If these three actions are taken, welding can potentially become the preferred method of joining metallic components in space and an important method for the long-term maintenance of structures, including large space-borne structures that are independent of earth.

The Need for Automation. Earth-based welding is a difficult task and requires both significant skill as well as time-consuming and tedious effort. If performed manually, space welding will be, at best, a very difficult task. The problems with space welding are compounded if the welding is to be done outside the spacecraft environment. Extra-vehicular activity (EVA) requires large numbers of ground personnel and considerable time and preparation by the mission specialist. In order for spacecraft crews to have the skill level required to make critical welds, considerable time and additional effort would have to be expended. Such training would be very expensive. Computer intelligent autonomous or semi-autonomous welding systems will probably be the most cost-effective means for space welding, particularly for EVA welding.

Intelligent Engineering/Planning Systems. Recently developed systems have demonstrated the possibility of coupling extensive databases with artificial-intelligence-based software to produce "intelligent" engineering and planning software that can make many routine engineering decisions for welding applications and guide a human user to optimum decisions for welding. These systems utilize expert system and neural network technology to provide accurate weld process models and engineering decision-making capability. This subject is addressed for earth-based welding applications in the article "Intelligent Automation for Joining Technology" in this Volume.

Intelligent engineering and planning workstations that are electronically coupled to automatic welding machines will substantially reduce the need for time-consuming work by mission specialists. The planning and development of welding schedules and procedures can be accomplished in a matter of a few minutes on a computer terminal rather than hours of work by hand. By having a complete set of databases and most of the routine engineering completed by the computer, the human user will need only to answer a few questions about the application, and the system will produce a complete welding plan and weld schedule that can be downloaded directly to the automated welding system. Such systems will offer considerable cost savings and manhour reduction while at the same time providing a carefully optimized welding plan and procedure.

Welding in Space and Low-Gravity Environments

Milo Nance, Martin Marietta Corporation; Jerald E. Jones, Colorado School of Mines

---

References

1. E.C. MCKANNAN AND R.M. POORMAN, PAPER 35812, GEORGE C. MARSHALL SPACE FLIGHT CENTER, ALABAMA, 1974, P 20
2. R.E. MONROE, "CHARACTERIZATION OF METALS MELTING RISKS: SKYLAB EXPERIMENT M551," NASA CONTRACTOR REPORT, BATTELLE MEMORIAL INSTITUTE, COLUMBUS, OHIO, 4 DEC 1973, P 109

Joining of Organic-Matrix Composites

L.J. Hart-Smith, McDonnell Douglas Aerospace

Introduction

SUCCESSFUL ADHESIVE BONDING of organic-matrix composites is dependent on the nature of the adherend surfaces. Factors that influence the integrity and durability of composite-to-composite bonds include wetting of the adherend, cleanliness of the substrate, chemical and physical properties of the adherends, and joint design. In this article, major emphasis is placed on the critical importance of proper surface preparation; both thermoset and thermoplastic composites are covered. The techniques used to analyze and design adhesively bonded composite joints are described in a companion article (Ref 1). Additional information can also be found in *Adhesives and Sealants,* Volume 3 of the *Engineered Materials Handbook* published by ASM International.

Reference

1. L.J. HART-SMITH, JOINTS, *COMPOSITES,* VOL 1, *ENGINEERED MATERIALS HANDBOOK,* T.J. REINHART, ED., ASM INTERNATIONAL, 1987, P 479-495

Joining of Organic-Matrix Composites

L.J. Hart-Smith, McDonnell Douglas Aerospace

Common Problems Encountered

The Importance of Proper Processing. It is well known that the adhesive bonding of metal structures requires strict adherence to the process specifications if the glue is to remain stuck for a long service life. It is also apparent that not all approved specifications are adequate. The Primary Adhesively Bonded Structures Technology (PABST) program (Ref 2) was conducted to demonstrate the tremendous benefits to metallic aircraft structures from good adhesive bonding, to associate previous problems in service with processing techniques that should be discontinued, and to validate the use of phosphoric acid anodizing with a corrosion-inhibiting primer as a reliable technique--as durable as the British chromic acid anodizing with Redux adhesive, and easier to use.

Only when the adhesive is stuck securely to the adherends is it possible to develop a bond strength that can be relied upon. Only then is it possible to use rational design procedures in which the joint can be designed for the bond to never fail, or if it does fail, for the failure to be characterized by cohesive failure of the adhesive layer itself. The interfacial failures associated with improper or inadequate processing are both unpredictable and variable--so variable, in fact, that neither the strength nor fatigue life can be established. This poor processing is the source of the so-called "weak bonds" that are discovered only after the structure has fallen apart.

Industrial experience with bonding of both metals and composites has shown that it is very easy to fabricate defective bonds that are strong enough to pass all initial inspections, but that have only a small fraction of the strength to be expected from properly processed bonds. Techniques for creating suitable surface preparations must be followed to achieve the normal situation in which the bonds are so much stronger than the thin members they join that all failures will occur outside the joint. The necessary steps are not difficult, but are often omitted because of a failure to appreciate the need for them.

Unfortunately, the importance of surface preparation for the adhesive bonding of fibrous composites is not widely acknowledged. Robert Schliekelmann, the pioneer of Redux bonding at Fokker, was sufficiently concerned about this oversight to make a plea for more attention to this issue (Ref 3). Hart-Smith and co-workers at Douglas Aircraft Company prepared a document on surface preparation of fibrous composites to help the airlines until repair manuals were completed (Ref 4). When this document was reprinted in the Canadair house journal (Ref 5), every experiment had to be repeated to create new photographs and every phenomenon was duplicated in order to verify that the problems discussed in the Douglas Aircraft article really existed.

A similar concern was expressed in England, where Parker and Waghorn (Ref 6) reported on a far more comprehensive test program on the effects of surface preparations on adhesive bond strength for carbon-epoxy laminates. They also concluded that composite surfaces must be abraded to achieve strong adhesive-bonded joints. Pocius and Wenz (Ref 7) advocated the use of Scotchbrite pads with embedded abrasive particles as an effective and reliable surface preparation technique for achieving good composite bonds (Ref 7). The form of Scotchbrite giving the best results is of the consistency now used widely in floor-scrubbing pads, which is a little finer than that used in rotary discs for stripping rust from automobiles before repainting.

Sources of Defective Bonds. A major source of defective bonding of composite adherends is the omission of a thorough mechanical abrasion of the surface, either by a low-powered grit-blast machine or thorough sanding, following the removal of a peel ply from the surface of the laminate. Grit blasting is shown to lead to the best thermoset bonds, which are enhanced by the use of an adhesive film cured on the surfaces of the part as it is made. Although peel plies can protect against gross contamination, they do not create a surface that is easily wetted by adhesives. Despite the cautions about the many wrong ways to make bonded joints, correct processing is neither difficult nor expensive and results in strong durable structures.

Most, if not all, peel-ply surfaces are inadequate for adhesive bonding. This is a contentious issue because many bonded composite structures are so thin that the laminate or honey-comb core fails first during testing, even when the bond is less than 10% effective. However, that is often not the case in service. It has sometimes been found that when a part damaged in service is disassembled to provide access to abrade the surfaces for a bonded or co-cured repair, it is relatively easy to break the bonds apart without delaminating much of the composite. In such a case, all of the original bond area that was neither detected as defective during initial inspection, nor delaminated as the doublers and stiffeners were removed from the skin, should be regarded as weak (defective) bonds that could not be detected by ultrasonic inspection. This fraction of the bond area may be as high as 90% of the total if the original bonding had not been processed correctly. This condition is usually associated with virtually all of the adhesive on one surface or the other, when assessed locally, although the surface to which the adhesive adhered may vary throughout the total bond area. The local areas in which there was no adhesion are typically quite large and cannot be mistaken for porosity.

It is extremely rare to see half the bond completely covering each surface after disassembly; the cohesive strength of the adhesive layer, if it has been processed correctly, is usually so much greater than that of the adherends that trying to pry good bonds apart will result in delaminating the composite, not the adhesive layer. As a point of reference for the tremendous peel strength of most modern adhesives, attempts to salvage machined metal fittings from completed test panels for reuse on later panels in the PABST bonded metal fuselage program were largely fruitless, even when the structure was struck with sledge hammers and cold chisels after chilling the panels with liquid nitrogen. (Some older adhesives have so little peel strength at low temperature that it used to be possible to disassemble bonded metal structures by giving them a sharp rap after chilling them with dry ice.) It would seem to make more sense to heat bonded composite structures above the glass transition temperature for the adhesive, but below that of the parent laminate to facilitate safe disassembly. The disadvantage here would be that the working environment of some 120 °C (250 °F) would be rather unpleasant, unless it could be confined to the immediate vicinity of the area of prying by using a jet of hot air.

There is no history of bonds that were truly good at the time of manufacture deteriorating in service. There have, however, been numerous instances in which the interface of a metal bonded structure has corroded in service because it had not been properly prepared in the first place and was just sound enough to pass the initial inspection. Although poor composite bonds deteriorate by mechanisms other than corrosion, they share with poor metal bonding the characteristic that their short-term static strength is sufficient to conceal their deficiencies. Weak bonds have been notoriously difficult to detect prior to failure when only the parts have been examined; however, they are easy to identify after failure and can be recognized reliably from a straightforward reading of the process monitoring records and an on-site survey of the procedures followed (and ignored) during manufacture. The process monitoring records should always take precedence over successful test coupon results, even when it means rejecting and rebuilding (not reworking) complete assemblies.

The reason for the difficulty in bonding of polymer composites is that it is difficult to create strong chemical bonds between the adhesive and a cured (and therefore largely fully crosslinked) polymer matrix. The cured surface needs to be activated and, even then, chemical bonds may be degraded by absorbed water (Ref 8). The most durable bond strengths rely upon mechanical interlocking of the adhesive and the thermoset composite surface, just as with the primer and porous anodized surface on metal adherends. If the surface is not adequately roughened, the glue will not stick, even if the surface is not contaminated. The most common contaminants are industrial-grade cleaning solvents and coupling fluids for ultrasonic inspections. These are particularly difficult to remove from delaminations and areas of porosity. Failure to remove them completely results in a bond with inadequate adhesion to the adherend, just as is the case with a metal-bond surface that had been contaminated before the primer was applied.

Effects of Absorbed Moisture. Water is the most widespread and troublesome source of defective bonded joints and repairs in composite structures. It may be absorbed by the basic composite material before it is ever cured. If it is not removed by drying prior to curing or by venting during cure, it will result in porosity as well as harmful changes in the chemical structure of the polymer. Water may also be absorbed by uncured adhesive films that have been improperly stored or that have not thawed out before being unwrapped after removal from the freezer or that have been subjected to an extended layup time in a humid work environment. Adhesives vary in susceptibility to this problem, but most have a far greater affinity for water than do epoxy prepregs. The consequences are much the same, and just as detrimental. Water may also be absorbed within cured laminates that are to be secondarily bonded either during initial manufacture or as part of an in-service repair. If this water is not removed by slow drying prior to bonding, the water will react adversely with the adhesive and result in defective bonds. Cured epoxies have very low diffusion coefficients, so water is not absorbed very rapidly. However, they cannot be dried rapidly for the same reason.

A systematic series of tests to differentiate between the effects of different kinds of prebond moisture has not been carried out to date. However, a weak, porous adhesive that crumbles easily into powder when the fractured surface is handled after the adhesive layer has been exposed, but that has stuck to the adherends, can be associated with too much moisture absorbed by the adhesive before cure. A perfectly sound solid adhesive film that simply did not stick is more likely to be caused by precured laminates to be bonded that were not dried or by condensate on the surface of an otherwise dry adhesive film that was unwrapped too soon after removal from storage in the freezer. However, some film adhesives left out for several days in a humid layup area can absorb so much water that they will not stick even if they are not visibly covered by condensate during layup. Even then, there is a chance that the strength of the fillets at the edges of the bonds will be sufficient to disguise the weakness by preventing the parts from falling apart after completion of the cure.

Labor and Myhre (Ref 9) found that if they did not dry the parent laminate first, they could not attach even a single 0.13 mm (0.005 in.) thick carbon-epoxy tape without serrating the ends with pinking shears. With square-cut ends, the single layer would rapidly delaminate under fatigue loads of only about 25% of design limit load. Obviously, however, such weak bonds would have passed all nondestructive inspections at the time the repair was executed. The only way of knowing that the bond would have failed had it been placed in service was recognition that water definitely was present during the cure of the adhesive--water that was added deliberately to make the coupons representative of structures in service that were being quickly repaired without first being dried out slowly and carefully. In extreme cases, with a rapid temperature rise during bonding, the parent laminate may even split apart because of the steam generated by the water that was unable to escape fast enough. Figure 1 shows the dramatic reduction in the hot-wet strength of adhesively bonded joints caused by as little as 0.25% absorbed moisture in the composite laminates before bonding. Less than 20% of the room-temperature strengths were retained. Figure 2 shows equally dramatic losses in the fatigue life under slow-cycle testing of bonded aluminum joints as a function of different levels of prebond moisture in the adhesive film prior to bonding.

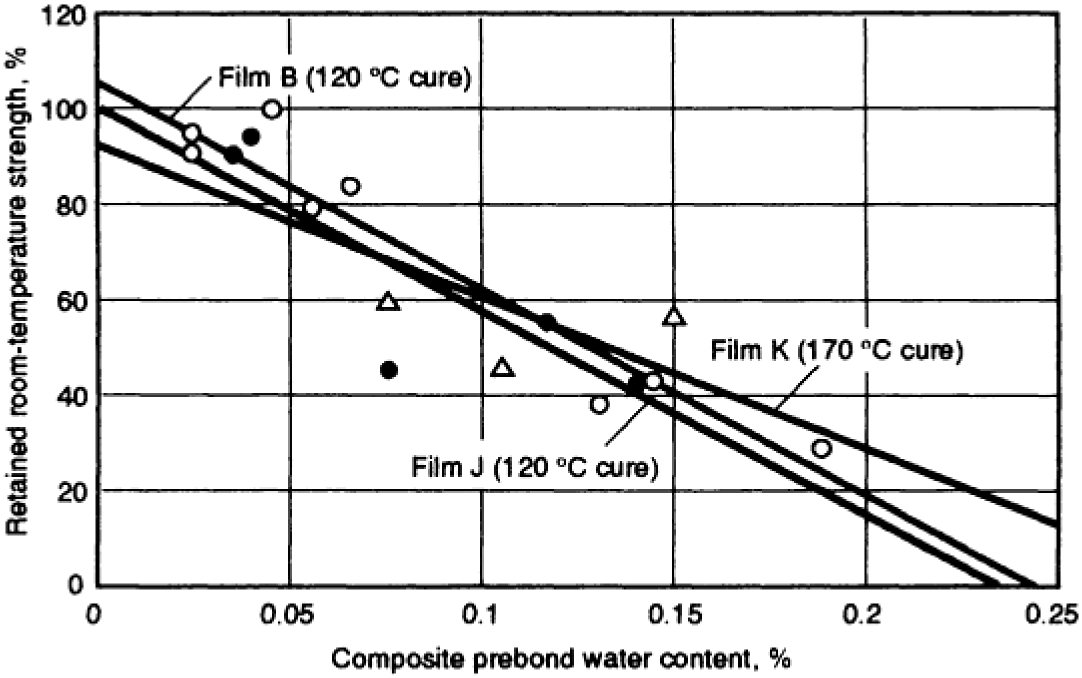

FIG. 1 EFFECT OF MOISTURE CONTENT ON THE HOT-WET BOND STRENGTH OF CARBON-EPOXY COMPOSITES. SOURCE: REF 10

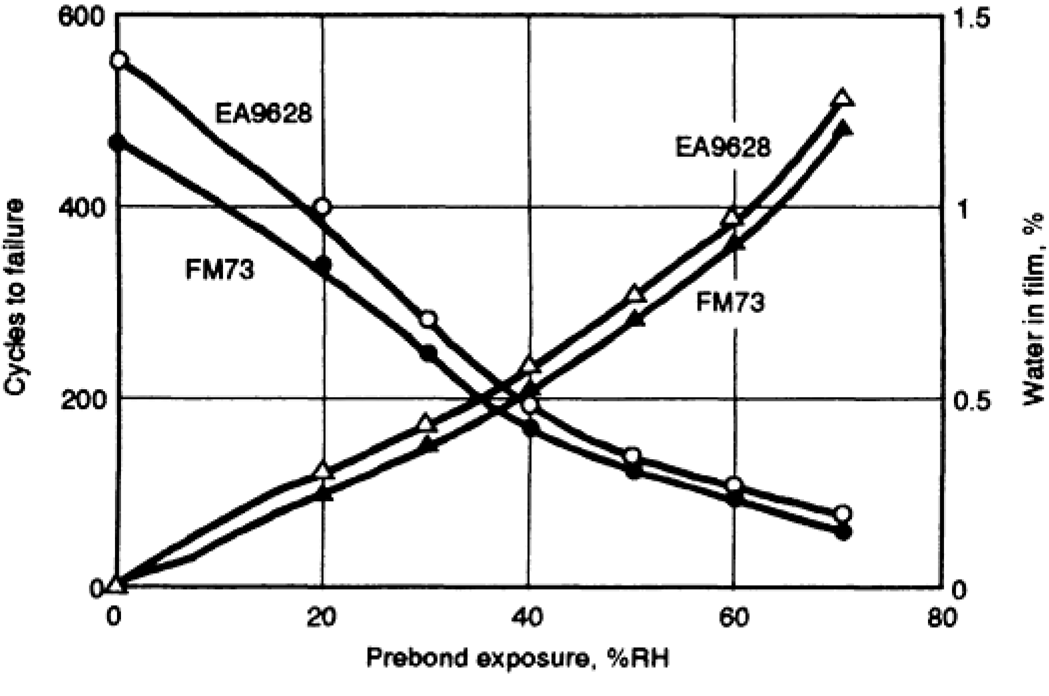

FIG. 2 EFFECT OF PREBOND MOISTURE ON FATIGUE STRENGTH OF ADHESIVELY BONDED ALUMINUM JOINTS. FILMS WERE PRECONDITIONED FOR 10 DAYS AT ROOM TEMPERATURE. RAAB SPECIFICATIONS, 10.3 MPA (1500 PSI), 4 CYCLES/H, 60 °C (140 °F), 100% RELATIVE HUMIDITY (RH). SOURCE: REF 11

Although the evidence is not yet absolutely conclusive, the epoxy resins may be more tolerant of absorbed moisture than is generally acknowledged if moisture can easily be driven off early in the cure cycle. However, if there are no vents to allow the steam (or solvents) to escape, as is often the case with no-bleed cure cycles, moisture cannot be driven off. There is conclusive evidence that the combination of the presence of moisture and a closed cure environment is disastrous. However, there have been some instances with metal-bonded structures when water was known to be present in the uncured adhesive, yet the bonds appear to have been good (per service). In each case, the bonding tool permitted easy egress of the steam generated during cure. Companion panels manufactured with a different bonding tool that trapped the moisture in the adhesive from the very same rolls, passed all ultrasonic inspections but were found to be unbonded during assembly.

A similar problem arose with the two-stage bonding of a composite spoiler. Although water was present in some form during both bonds, there was adequate venting for the steam to escape while the first skin was bonded to an oversize block of honeycomb core, to stabilize it for subsequent machining of the aerodynamic profile (it is not known if the water was absorbed in the uncured adhesive film, if the Nomex honeycomb core had not been dried properly, or if the precured laminates had not been dried properly). This first bond was so strong as to give no warning of the problems to follow. When the second skin was bonded to the other side of the core, the water was trapped and the second bond had only a small fraction of the strength exhibited by the first bond. When tested, the second bond failed prematurely; it looked porous and weak, with poor filleting at the cell walls, while the first bond showed good filleting, as if the water had never been present.

The above examples are included not as an invitation to disregard the concerns about moisture present during the cure, but as encouragement to pursue this subject further to see if there are any consistently reliable alternatives to careful drying prior to bonding, which is obviously time consuming and costly. It is also intended to emphasize the need for venting during cure. There are other known instances in which simple thin laminates have suffered locally from very wavy fibers caused by the lack of an escape path for bubbles of gas created from the solvent used during the process of impregnating the fibers with resin. The bubbles lifted the fibers off the tool surface, pulling the ends in toward the bubbles. Eventually, the bubbles burst, causing the fibers to wrinkle as they were flattened out. There was no mechanism to push the fibers back to where they had been.

Difficulties in Detecting Defective Bonds. The most difficult aspect of defective bonds caused by prebond moisture or inadequate surface preparation is that the condition cannot be detected by ultrasonic inspection. It must be avoided by adherence to the correct procedures--which may differ from those specified--during processing because the weak bonds resulting from such errors cannot be repaired, even if they are detected before they fail. The parts must be disassembled, the surfaces cleaned and prepared again and the entire bonding cycle repeated. In general, it is easy to rework improperly manufactured mechanically fastened metallic structures and to restore 100% of the design strength. That is not true for either adhesively bonded metal or composite structures; the rework is difficult and the strength recovered far less than complete. This is why it is so important to do the job properly in the first place.

The cost of accepting defective parts is compounded as the original panels are embedded ever more deeply in the structural assemblies. It is bad enough when the defects are unrecognized because the end-item inspections are ineffective; inspection can exclude a few very bad parts, but cannot ensure the integrity of the great majority of assemblies. It is worse when parts are recognized as defective from a visual inspection, but are accepted anyway on the basis of coupon tests or ultrasonic inspections.

Bonding of Thermoplastic Composites. Thermoplastics are far more difficult to bond than are thermoset composites, but good results have been achieved by chemically etching the surface prior to bonding. The most promising approach for joining thermoplastic composites appears to be fusion bonding, using a compatible surface layer of plastic that melts at a lower temperature than that needed to melt the parent matrix. The ability of typical modified epoxy adhesives to bond chemically to the thermoplastic matrix is even less than for thermoset composites. Also, the thermoplastic surface is so resilient that it is very difficult to abrade sufficiently without breaking too many surface fibers in the process of creating the roughened bond surface. The best bonds made with layers of thermoset adhesives between thermoplastic adherends seem to have a strength of only about 80% of that of the best bonds using the same adhesive between thermoset adherends. One must conclude, therefore, that this loss of strength results from an interfacial bond strength that is weaker than the intrinsic cohesive strength of the adhesive layer. It follows, therefore, that the strength to be attained would be highly variable, as would the service life. Far better success has been achieved by chemically changing the surface to be bonded by oxidation using hot gases, by cold gas plasma treatment, or by corona discharge. The surface energy is increased by each of these techniques, permitting better wetting by the adhesive. However, these processes are rather slow if a large area has to be covered.

Success in bonding thermoplastics has also been achieved by fusion, or welding. The best success has been achieved when the original thermoplastic laminates have been fused together with a special surface layer of a different polymer that can be melted at a lower temperature than is needed to melt the base laminates. Direct fusion of thermoplastic laminates often results in considerable distortion, even if a structurally sound bond is achieved. Even making thermoplastic composites using focused local heat to fuse each ply to the substrate has proved to be a slow process that is not well suited to large panels.

Similarly, it is common practice in making thermoset composite aircraft panels to include a surfacing layer of film adhesive co-cured with the parent matrix to both facilitate getting the paint to adhere and to provide a tough gel coat to compensate for the excessively dry surface plies in normal composite laminates. (The weight of this surfacing layer is less than that of the glazing that would otherwise need to be applied, by hand, to avoid pinhole porosity and an unacceptable paint finish.) Curiously, less use is made of this concept for preparing surfaces to be bonded, although an Australian firm has reported great success with it (Ref 12).

References cited in this section

2. E.W. THRALL, JR. AND R.W. SHANNON, ED., *ADHESIVE BONDING OF ALUMINUM ALLOYS,* MARCEL DEKKER, 1985
3. R.J. SCHLIEKELMANN, ADHESIVE BONDING AND COMPOSITES, *PROGRESS IN SCIENCE AND ENGINEERING OF COMPOSITES,* VOL 1, PROC. 4TH INT. CONF. COMPOSITE MATERIALS (TOKYO), 22-28 OCT 1982, JAPAN SOCIETY FOR COMPOSITE MATERIALS AND THE METALLURGICAL SOCIETY (TMS) OF AIME, T. HAYASHI, K. KAWATA, AND S. UMEKAWA, ED., NORTH-HOLLAND, AMSTERDAM, 1982, P 53-78
4. L.J. HART-SMITH, R.W. OCHSNER, AND R.L. RADECKY, SURFACE PREPARATION OF FIBROUS COMPOSITES FOR ADHESIVE BONDING OR PAINTING, *DOUGLAS SERVICE MAGAZINE,* FIRST QUARTER 1984, P 12-22
5. L.J. HART-SMITH, R.W. OCHSNER, AND R.L. RADECKY, SURFACE PREPARATION OF FIBROUS COMPOSITES FOR ADHESIVE BONDING OR PAINTING, *CANADAIR SERVICE NEWS,* VOL 14 (NO. 2), SUMMER 1985, P 2-8
6. B.M. PARKER AND R.M. WAGHORNE, SURFACE PRETREATMENT OF CARBON FIBRE-REINFORCED COMPOSITES FOR ADHESIVE BONDING, *COMPOSITES,* VOL 13, JULY 1982, P 280-288
7. A.V. POCIUS AND R.P. WENZ, MECHANICAL SURFACE PREPARATION OF GRAPHITE-EPOXY COMPOSITE FOR ADHESIVE BONDING, *SAMPE J.,* SEPT/OCT 1985, P 50-58
8. A.J. KINLOCH AND G.K.A. KODOKIAN, "THE ADHESIVE BONDING OF THERMOPLASTIC COMPOSITES," EUROPEAN RESEARCH OFFICE OF THE U.S. ARMY, CONTRACT DAJA 45-86-C-0037, FOURTH PERIODIC REPORT, IMPERIAL COLLEGE, LONDON, AUG-DEC, 1987
9. J.D. LABOR AND S.H. MYHRE, "LARGE AREA COMPOSITE STRUCTURE REPAIR," NORTHROP AIRCRAFT CORPORATION, USAF TECHNICAL REPORT AFFDL-TR-79-3040, MARCH 1979
10. B.M. PARKER, THE EFFECT OF COMPOSITE PREBOND MOISTURE ON ADHESIVE-BONDED CFRP-CFRP JOINTS, *COMPOSITES,* VOL 14, JULY 1983, P 226-232
11. E.W. THRALL, JR., PROSPECTS FOR BONDING PRIMARY AIRCRAFT STRUCTURES IN THE 80'S, *25TH NATIONAL SAMPE SYMPOSIUM,* VOL 25, SOCIETY FOR THE ADVANCEMENT OF MATERIAL AND PROCESS ENGINEERING, 6-8 MAY 1980, P 716-727
12. I. MCARTHUR, HAWKER DEHAVILLAND, BANKSTOWN, AUSTRALIA, PRIVATE COMMUNICATION, 1992

Joining of Organic-Matrix Composites

L.J. Hart-Smith, McDonnell Douglas Aerospace

Examples of Good and Bad Adhesive Bonds

Properly Processed Bonds. Figure 3 shows the magnified fractured surface of a properly processed adhesive bond, with continuous coverage of the adherend. With composite adherends, this condition can occur only in test coupons deliberately designed with so little bond area that the bond will fail by shear. It is also necessary to suppress the generation of peel stresses at the ends of the overlap, by keeping the tip thickness of the adherends low. Properly designed structural joints will always fail outside the bond area, and while a coupon test based on such a joint can confirm the adequacy of the design, it cannot confirm that the processing was done properly because there is no fractured bond line to inspect. In properly designed bonded joints, the integrated shear strength of the bond should exceed the load needed to fail the adherends. It is almost impossible not to satisfy this condition for thin adherends, even for improperly processed bonded joints that will fail in service, but are strong enough to pass all the initial inspections. However, for thick adherends, it becomes necessary to use stepped-lap joints to prevent the bond from becoming a weak-link fuse.

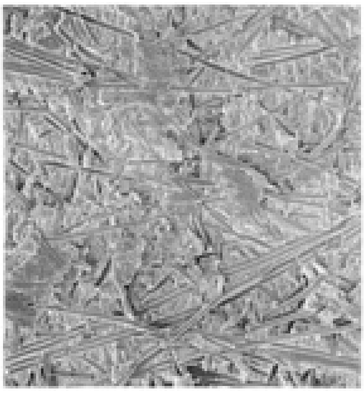

FIG. 3 COHESIVE FRACTURE SURFACE IN A PROPERLY PROCESSED COMPOSITE-TO-COMPOSITE ADHESIVE BOND. VISIBLE FIBERS ARE THE RANDOM MAT CARRIER IN THE ADHESIVE FILM. 68×

Adhesion Failures. * Figure 4 shows a typical adhesion failure in which the adhesive almost totally failed to adhere to one of the adherends because the surface had not been adequately roughened. Although this photomicrograph is of the surface of the adhesive, not the composite, it clearly shows the imprint of the peel-ply woven surface to which it had not bonded over most of the surface. The comparison between the two modes of failure is dramatic. This failure of the adhesive to bond to a surface arises frequently when an attempt is made to bond directly to the surface that remains after removal of a peel ply, but it also occurs when one tries to bond directly to the other surface of the laminate, which is usually contaminated by release agents used to prevent the part from sticking permanently to the layup tool.

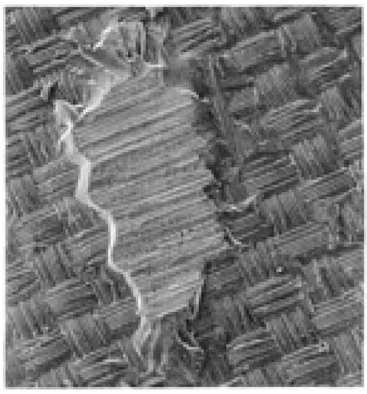

FIG. 4 ADHESION FAILURE SHOWING IMPRINT OF THE PEEL PLY AND FAILURE TO BOND OVER MOST OF THE SURFACE. A SMALL PIECE OF RESIN ADHERED AND IS IDENTIFIED BY THE IMPRINT OF THE FIBERS. 27×

In a different kind of adhesion failure, the adhesive does not remain in one continuous disbonded layer, but remains attached to one of the adherend surfaces in discrete full-thickness fragments. The underlying shiny peel-ply surface is clearly evident in the holes in the adhesive on both surfaces. Roughly half the adhesive is attached to each surface, but only weakly. This kind of fracture has been observed in uncontaminated bond surfaces that had not been mechanically abraded to make the glue stick. It is the author's belief that this condition, in contrast to that shown in Fig. 4, can occur only in conjunction with small amounts of moisture trapped in the adhesive as it cures. Given that the interfacial bond is uniformly poor, it seems inconceivable that it could have been strong enough to tear a properly cured adhesive film from side to side. However, it is well known that a weak powdery bond results if the adhesive is cured in the presence of moisture. This moisture can either be absorbed by the adhesive film, through improper storage and handling, or by the laminate during service (with even more absorbed moisture, some adhesives will not stick at all). Actually, even the time involved between detail manufacturing and assembly can be sufficient to permit the absorption of sufficient moisture to create a weak powdery adhesive bond, even if the bond surfaces had been properly prepared by thorough abrasion. The laminates need to be stored in sealed containers until bonded or in a dehumidified air-conditioned room.

Inspection Methods. None of the structural deficiencies associated with weak or improperly cured adhesive bonds are detected by conventional ultrasonic nondestructive inspection (NDI), and it is most improbable that they could be detected prior to a gap being opened up by mechanical failure in service. Although the fillets at the ends of the bond overlap will be visibly porous for adhesives cured with trapped prebond moisture, it is usually only the hot-wet lap-shear tests that fail to meet the specified strengths. Consequently, the customary room-temperature tests that follow would not have caught this condition had it been on a real part. Figure 5 shows why visual inspections of adhesive bonds can be more informative and reliable than conventional ultrasonic NDI. More reliable results can be obtained from a visual inspection of the fillet than from either ultrasonic inspections or a defined "adequate" lap-shear test result on a coupon with abnormal fillets. Ironically, the ultrasonic signals are unreliable in the immediate vicinity of the edges of panels--which are the only locations where any significant load is transferred through bonded joints.

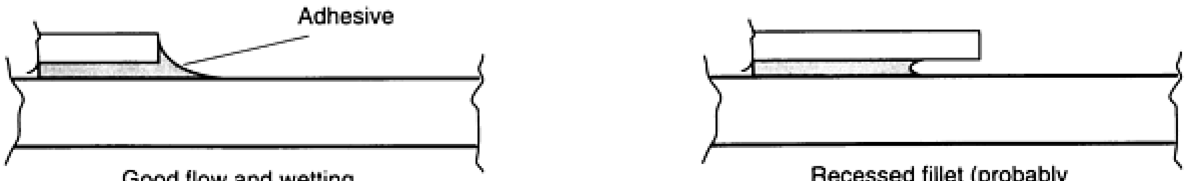

Good flow and wetting

Recessed fillet (probably due to poor fit or lack of pressure)

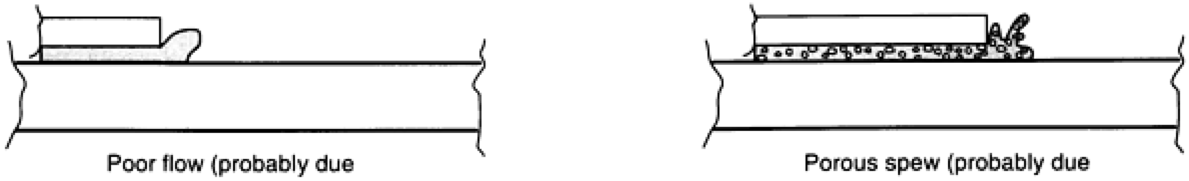

Poor flow (probably due to slow heat-up rate)

Porous spew (probably due to moisture in adhesive or excessive heat-up rate)

FIG. 5 EXAMPLES OF GOOD AND BAD ADHESIVE FILLETS

Note cited in this section

* *IT IS A COMMON MALPRACTICE TO USE THE WORDS "ADHESIVE FAILURES" EXCLUSIVELY TO DESCRIBE ONLY THOSE SITUATIONS IN WHICH THE ADHESIVE ITSELF HAS *NOT* FAILED, BECAUSE ADHESION IS A NOUN AND NOT AN ADJECTIVE. ACTUAL FRACTURE OF THE ADHESIVE ITSELF IS DESCRIBED AS A COHESIVE FAILURE. THE FORMER FAILURE OCCURS WHEN THE GLUE HAS NOT ADHERED TO ONE OR BOTH ADHERENDS. THE CURRENT MEANING ASCRIBED TO ADHESIVE FAILURES IS THUS VERY MISLEADING, EVEN IF IT IS GRAMMATICALLY CORRECT. IN THIS ARTICLE, THE WORD ADHESION IN PLACE OF ADHESIVE WILL BE USED IN THIS CONTEXT.

L.J. Hart-Smith, McDonnell Douglas Aerospace

Surface Preparation for Thermoset Composite Adherends

While not always used in production, light grit blasting, thorough sanding, or some other form of mechanical abrasion are the best surface preparations for both adhesive bonding and painting of thermoset composites. This is confirmed by the test results shown in Fig. 6. As shown in this figure, the best example of sanding was as good as the grit-blasted coupons, but all other treatments were markedly inferior.

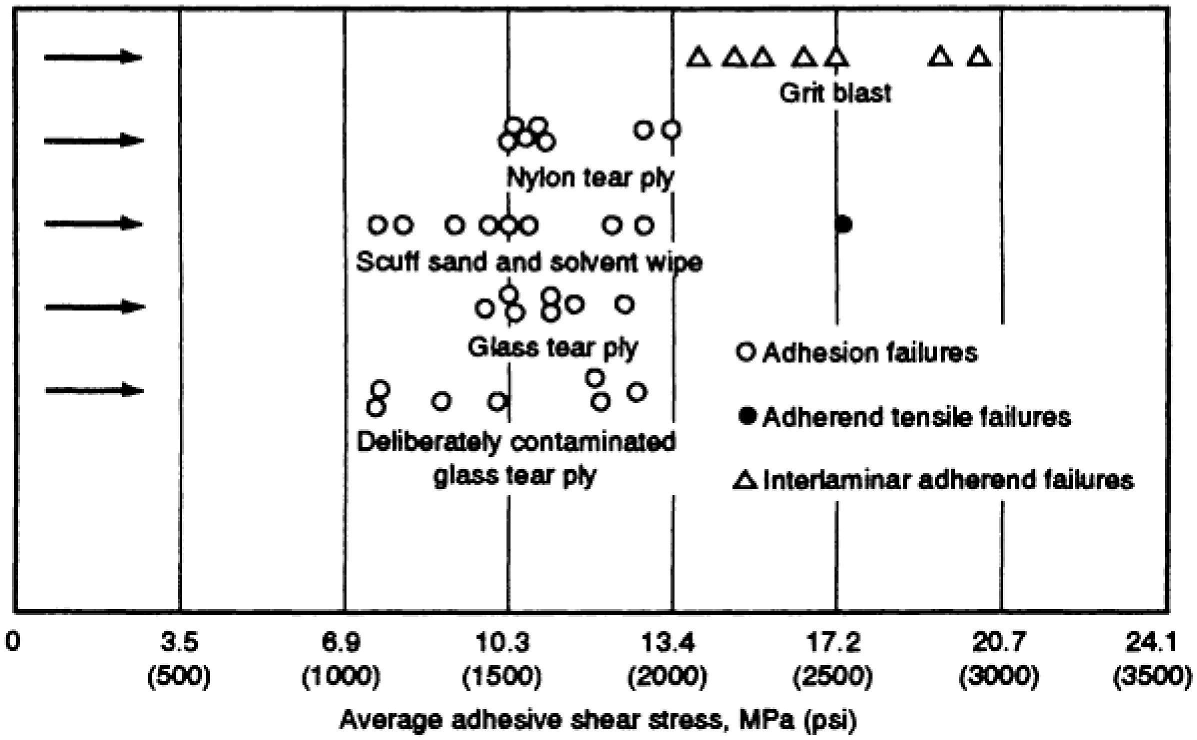

FIG. 6 EFFECT OF SURFACE PREPARATION ON THE STRENGTH OF DOUBLE-LAP GRAPHITE/EPOXY-TO-ALUMINUM ADHESIVE JOINTS (13 MM, OR $\frac{1}{2}$ IN., OVERLAP)

Peel-Ply Surfaces. While it is extremely difficult to ensure that a peel-ply surface will be satisfactory before the peel ply is removed, it is quite easy to evaluate the surface after removal in time to prevent bonding to an unsatisfactory surface. Figure 7 shows a magnification of the imprint of a nylon peel ply that had supposedly been corona treated to enhance its adhesion to the resin matrix to create a fresh fractured surface over the entire bond. Yet it is quite clear that, at the microscopic level, the surface is absolutely slick. The groove left by every fiber in the peel ply is clearly shown. The matrix is fractured over only an extremely small fraction of the surface, as shown in the higher magnification view of the same surface in Fig. 8. The adhesive will stick only to the fractured surface, not the obviously ultraslick surface so clearly evident in most of Fig. 8. The adhesive could bond to barely 10% of this surface and will not stick anywhere else. A reliable peel ply would leave an imprint so rough at the microscopic level that the weave of the peel ply so prominent in Fig. 7 would never be seen.

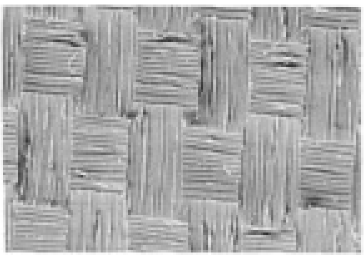

FIG. 7 IMPRINT OF A NYLON PEEL PLY LEFT AFTER REMOVAL FROM COMPOSITE LAMINATE. 58×

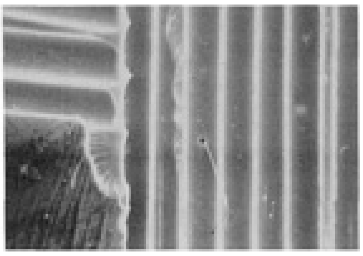

FIG. 8 IMPRINT OF A PEEL PLY SHOWING EXTREMELY SMALL FRACTION OF SURFACE COVERED BY FRACTURED RESIN MATRIX TO WHICH THE ADHESIVE COULD BOND. 580×

Figure 9 shows the imprint left by a different peel ply, that was believed to be producing satisfactory bonds, or at least much stronger bonds than had been achieved using another peel ply that left a surface like that shown in Fig. 7. There is clearly far less imprint visible in Fig. 9 than in Fig. 7, and far more fractured surface resin. Even so, one would be relying on good adhesion to no more than about 30% of the total area. Also, the surface in Fig. 9 is so rough as to be likely to trap pockets of air between the adhesive and the substrate. Fortunately, unless they coalesced, these air pockets would prevent contact only between the glue and those parts of the laminate to which it would not stick anyway.

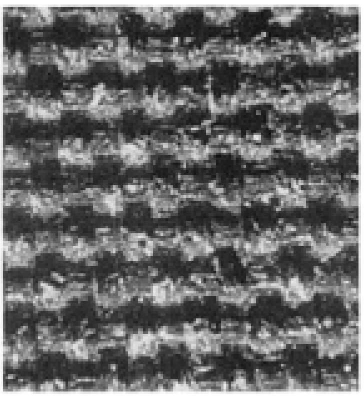

FIG. 9 IMPRINT OF A BETTER PEEL PLY ON A COMPOSITE SURFACE SHOWING A FAR GREATER FRACTION OF FRACTURED RESIN OF THE SURFACE. 34×

The real purpose of peel plies is to protect against gross contamination, which is removed as the peel ply is lifted off. Its removal must be followed by grit blasting or by sanding so thorough that all traces of the imprint from the peel ply have been removed. Figure 10 shows a scanning electron micrograph of a carbon-epoxy composite thoroughly sanded with 120 grit abrasive. There is no similarity with either Fig. 7 or Fig. 9. However, if the abrasion is so severe as to expose bare fibers, from which the surface treatment (coating) has been removed everywhere by sanding, the bond will be even weaker. The solution to this problem is a resin-rich surface layer, like a conventional gel coat applied to simple fiberglass composite components before they are painted.

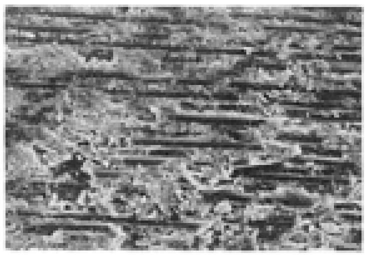

FIG. 10 SURFACE OF A CARBON-EPOXY COMPOSITE AFTER 120 GRIT ABRASIVE SURFACE PREPARATION. NOTE EXPOSED FIBERS AND SMALL AMOUNT OF DAMAGED FIBERS. 415×

Peel plies are either coated with release agents so that they will not delaminate the composite they are meant to protect or are so slick that the surface is like a mirror at the microscopic level. In either case, the glue cannot stick, except at the small nodules of fractured laminating resin in between the bundles of fibers in the weave of the peel ply. This increment of strength is often sufficient to bond successfully honeycomb cores in sandwich construction. Providing that all the remainder of the processing is done properly, even this limited bond strength is often enough to break the core away from the faces. In addition, this bond strength seems always to be enough to pass any ultrasonic NDI tests, even for faying-surface bonds. Another reason for thorough abrasion after removal of a peel ply is that the polymer molecules formed adjacent to the peel ply (or released tool surface) are oriented in such a way as to make them very difficult to be wetted by the adhesive. A stronger bond is possible in the more randomly oriented interior molecules.

Alternatives to Peel-Ply Surfaces. An alternative technique used to some extent in both Europe and North America is to not use a peel ply as such, but to add an extra layer of preimpregnated fine-weave aramid cloth to both surfaces of the composite laminate, using the same resin as for the primary laminate. This achieves two things. First, it avoids the acutely resin-starved surfaces left as the conventional dry peel plies absorb resin from the base laminate; second, it ensures a completely roughened surface to which to bond, without any trace of release agent from the caul plate.

Tests using fiberglass tear plies that were not intended to separate cleanly, but to remove a little resin from the entire surface, failed because the glass fibers were too brittle and broke repeatedly. The surface was contaminated, and possibly damaged as well, whenever razor blades were used to strip off the surface ply. Obviously, if the tear ply sticks too well, it will not come off alone, but will probably take one or two top plies from the laminate as well. This has been known to occur during the removal of metallic fittings from composite structures that had been installed with a layer of rubber-base sealant between the faying surfaces to prevent galvanic corrosion.

Grit Blasting. The objections to grit blasting are easier to recognize, even though they are not all valid. Grit blasting can create dust--black dust from carbon-epoxy laminates is perceived as a greater health hazard than white dust from fiberglass-polyester laminates. Also, it relies to some extent on operator skill, and if the only available grit-blasting machine must be borrowed from a nearby foundry where it is normally used to clean sand castings, it will be so powerful that it is easy to burn right through a laminate in a matter of seconds. Both of these objections can be overcome, as they were on the Lear Fan all-composite aircraft, by buying an inexpensive low-pressure machine in which the nozzle is surrounded by a ring of vacuum ports, which are themselves surrounded by a circular comb of soft bristles like those on domestic vacuum cleaners. It took only two to three days to properly train someone with no prior mechanical skills to operate such a machine. The bristles held the nozzle a fixed distance away from the work. An adequately blasted surface had a dull matt texture, which contrasted strongly against the shiny black unblasted areas nearby. The pressure was only 140 kPa (20 psi) instead of 700 kPa (100 psi), so it would take such a long time of holding the nozzle in one place to damage the composite surface that it would be obvious that something was wrong long before any damage actually occurred. The grit-blast machine was used in a room that was depressurized slightly with respect to the rest of the factory, but there was no need for protective clothing or respiratory masks. A fixed machine with many nozzles under which the parts pass on a roller table has also been developed. The nozzles may someday be computer controlled.

Currently, there are many forms of grit available, some of which are recyclable. Aluminum oxide (Al_2O_3) has been used traditionally, but is sufficiently expensive that recirculation of the used grit must be considered. It would seem that the use of recirculated abrasives can be permitted on an otherwise uncontaminated peel-ply surface. However, so much release agent is transferred from caul plates to the adherend surface that no recirculation can be permitted on that kind of surface. Examples of inexpensive media are plastic beads, which can be separated from the debris by floatation in a water tank, biodegradable corn husks, and a number of powders, such as sodium bicarbonate, which can be recovered from aqueous solutions. Since chemical paint stripping is being outlawed, there is considerable pressure to come up with safe replacements. Grit blasting is certainly preferable to the use of jets of hot air, which have been observed to burn as many as two surface plies to the bare fibers in several places on thin-skinned fibrous composite parts.

Perhaps the most compelling argument in favor of using grit blasting is that none of the more widely used techniques for initial production can ever be used for repairs, whether by an airline during service or by the original manufacturer because of damage during manufacture. For example, a peel-ply surface cannot be recreated on a previously cured laminate--one is forced to abrade it.

Joining of Organic-Matrix Composites

L.J. Hart-Smith, McDonnell Douglas Aerospace

The Water-Break Test

The water-break test has long been a standard technique for verifying the adequacy of surface preparation for metal bonding. It can also be used for composite surfaces, as described in Ref 4. This is clear from Fig. 11, which compares the even wetting of a properly grit-blasted surface with the repelling of water on an untreated surface contaminated by the release agent on the lay-up tool. Although the water is on the surface for only a small time, and very little is absorbed into the laminate, some adhesives are so moisture-sensitive that it is necessary to dry even that moisture out before bonding. It would seem to make more sense to prime the composite surface after grit blasting so that it could be recleaned using solvents, than to use the water-break test on a surface known to have been prepared properly immediately beforehand. The benefit from the water-break test comes from confirming that a previously prepared panel has not been contaminated during storage.

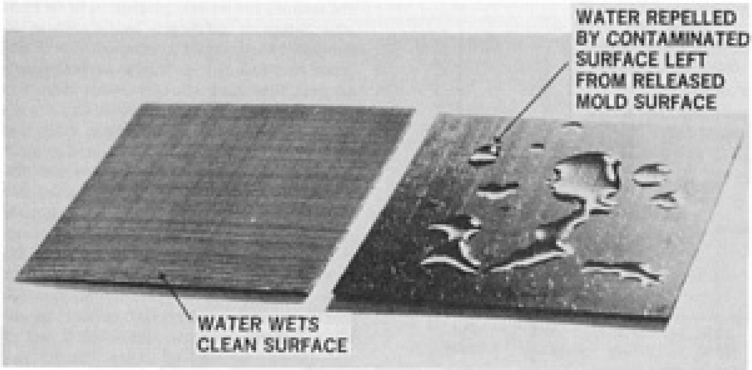
WATER REPELLED BY CONTAMINATED SURFACE LEFT FROM RELEASED MOLD SURFACE
WATER WETS CLEAN SURFACE

FIG. 11 COMPARATIVE WATER-BREAK TEST RESULTS. LEFT: PROPER PREPARATION. RIGHT: IMPROPER PREPARATION. COURTESY OF CANADAIR

Experience gained since the preparation of Ref 4 indicates that water-break tests are inappropriate for surfaces that have not been abraded. If water is applied to the surface left by removing a peel ply, the water will bridge between the small areas of fractured resin matrix and totally mask the unwettability of the majority of the surface. The test is ineffective in this case. Worse, it falsely creates the illusion that the surface has been prepared adequately for bonding, as well as the opportunity for someone to shift the blame for a poor bond from failure to dry out water absorbed from other sources to the microscopic traces of water left by the water-break test. Such traces are easily removed by a moderately hot air gun because the water is not on the surface long enough to permit significant absorption.

Matienzo *et al.* (Ref 13) indicate that lightly sanding the surface created by curing a laminate directly against a tool surface coated with mold release can be good enough to pass the water-break test, yet such surfaces cause bonds weaker than those achieved by untreated peel-ply surfaces. This indicates that even more emphasis must be placed on the thorough sanding or abrasion of the surface to be bonded and less on any subsequent verification. Unfortunately, their test data do not include baseline tests on anodized aluminum adherends against which the composite bonds could be assessed. One of their prime conclusions was that the simple removal of peel plies gave stronger bonds than were achieved with light sanding. However, it is significant that according to their test data, even higher average strengths were attained by deliberately contaminating one of the two peel plies by spraying it with mold release before the panel was cured. It should be pointed out that mold-release agents will stick like glue if they are not baked before a part is cured on the surface. This "contaminated" peel ply gave such better results than the uncontaminated peel ply that it is clear that the untreated peel-ply surfaces were capable of considerable improvement. Even a peel ply that showed no sign of silicon transfer gave slightly weaker average results than the spray-contaminated peel ply, although the highest individual strength was attained with the silicon-free peel ply. This peel ply was also effective in isolating the laminate from the silicon introduced from the mold release. However, as noted in their work, the use of peel plies on both sides of the part would not be applicable to compound-curved panels or to tool surfaces very much larger than available pieces of peel ply, which could not be lapped without unacceptable mark-off on the outside of the finished part. While even the highest strengths reported in Ref 13 are only about half what one might expect from a grit-blasted surface, the very low strengths they show from either gross contamination or from failure to make any attempt to remove the mold release indicate an irrefutable need for proper surface preparation for bonding polymer composites.

The change of surface texture between thoroughly abraded and bare surfaces is so great that, even though the water-break test works reliably, it is not always necessary. A visual inspection, which is capable of distinguishing between thorough and light sanding with the use of a magnifying glass, should suffice. This justification for omitting the water-break test is particularly relevant whenever the adhesive bonding follows immediately after the surface preparation and there is no opportunity for contamination. On the other hand, a drying cycle to remove water applied during the water-break test would also remove absorbed water from near the surface of the part. The omission of a compulsory drying cycle would demand great care in ensuring that the parts to be bonded had not absorbed water since their manufacture, even if they had been protected against the more widely recognized contaminants.

The water-break test would still seem to be very valuable for repairs to structures contaminated by chemicals absorbed during service or by coupling fluid for ultrasonic NDI of a porous or delaminated panel. It has already been shown to be effective in proving that most cleaning solvents applied to an unprimed surface actually contaminate the surface and need to be removed by water-soluble solvents such as alcohol. This issue is discussed in Ref 4, which includes photographs of the beach marks left as the solvent evaporates.

Experience with using the water-break test has revealed that the application of the least amount of water possible is preferable because a thick film of water can bridge a small local area of contamination and conceal a defect such as a thumb print. One needs to watch the film as it evaporates. A slow uniform drying out simultaneously over the entire surface is an indication of satisfactory surface preparation. If, on the other hand, a discrete dry area suddenly appears at some point during the evaporation, that is an indication of contamination.

References cited in this section

4. L.J. HART-SMITH, R.W. OCHSNER, AND R.L. RADECKY, SURFACE PREPARATION OF FIBROUS COMPOSITES FOR ADHESIVE BONDING OR PAINTING, *DOUGLAS SERVICE MAGAZINE*, FIRST QUARTER 1984, P 12-22
13. L.J. MATIENZO, J.D. VENABLES, J.D. FUDGE, AND J.J. VELTEN, SURFACE PREPARATION OF BONDING ADVANCED COMPOSITES, PART 1; EFFECT OF PEEL PLY MATERIALS AND MOLD RELEASE AGENTS ON BOND STRENGTH, *30TH NATIONAL SAMPE SYMPOSIUM*, SOCIETY FOR THE ADVANCEMENT OF MATERIAL AND PROCESS ENGINEERING, 19-21 MARCH 1985, P 302-314

Joining of Organic-Matrix Composites

L.J. Hart-Smith, McDonnell Douglas Aerospace

Control of the Bonding Process for Thermoset Adhesives

The chemistry of creating a properly cross-linked adhesive bond is well understood by a large number of materials and process engineers. However, far fewer people who actually control autoclaves are knowledgeable of adhesive chemistry or have real-time access to such information. With the help of a polymer chemistry expert, an experienced process engineer can identify the cause of a problem and modify a cure cycle in progress to save a part that would otherwise need to be scrapped. Some of the more important process variables that need to be understood and controlled are outlined below.

Effects of Temperature Variations on Bond Strength. Some aspects of the process specification are almost so self-evident that their significance is underestimated. It is obvious that if an adhesive is heated up too slowly, it will not wet the surface to be bonded. The adhesive may eventually cure, but there will be a dry bond. Therefore, the process specifications include minimum heat-up rates. These vary with the adhesive, but are on the order of 3 °C (5 °F) per minute. Likewise, an adhesive can be made to boil if heated up very much faster, so there must be maximum heat-up rates and maximum temperatures as well. These limits present no problem when making a small, uniformly thick test panel. The problems arise when a large, complex assembly of variable thickness is either cured or bonded on a large tool having a variable heat sink. It then becomes very easy for some parts of the structure to be heated up too slowly while other parts are on the verge of being burned. For all but the simplest of parts, a proof-of-tool demonstrator should be carried out to ensure that the assembly can be properly cured the first time. It makes sense to save on the cost of this operation by using cheaper fibers with the same resin.

One solution for the problem of a large range in temperature within the composite part or assembly, once a cure is already in progress, is to put constant-temperature dwells in the cure cycle to give the various segments of the part time to equilibrate. This is best done early in the cure cycle, before any of the adhesive (or resin matrix) has gelled. Thus, in addition to maximum and minimum heat-up rates, there is a need for a maximum variation between high and low temperatures--measured on the part, not on the tool, even if that takes more time than it does to hook up a permanently wired thermocouple. In some cases, multiple dwells may be needed. There is a limit, however, since an excessive number of dwells will cause the adhesive to gel before it has a chance to wet the surface needing to be bonded.

Once such a problem is detected, blankets can be added for subsequently made parts to shield the hot spots and local heaters added to overcome the colder areas. It is not acceptable to simply disconnect the thermocouples giving the highest and lowest readings in subsequent cure cycles. It is better still to thermally isolate large heat sinks from the rest of the tool if this can be done without causing distortion of the part. A simple but effective solution to some such problems is to thermally isolate the tool from the bed of the autoclave and, wherever possible, to thermally isolate the tool surface contacting the part or assembly from the supporting structure. Other than the possibility of tool distortion, which may or may not adversely affect the product, no harm will come to the composite part or bonded assembly if it is heated correctly and only the tool support structure lags behind.

Most tools are made from metal because hand layed-up plastic tools tend to have internal voids that collapse and need patching. However, the metal tools expand and contract more than composite parts. A solution to this problem is to use autoclave-cured composite members to hold composite parts in the work station during a bond cycle. The composite holding frame is supported by the basic metal tool, but it is held at only one point and permitted to slide freely as the size of the metal tool changes. This enables linear dimensions to be held and avoids the problem of crushing the composite parts as the tool contracts during cool down. It also thermally isolates composite parts from the metal tools.

The problem of tools having excessive heat sinks arises because tool designers are used to extremely rigid tools rather than tools that are barely able to maintain their shape, but so simple and compact as to permit many of them to be loaded into a single autoclave cycle. For thin, flexible skins, it makes more sense to rack several caul plates in a single holding fixture than to make one massive lay-up and cure tool per part. The productivity of the autoclave is enhanced this way as well. While the caul plates may distort out-of-plane, they will retain their in-plane dimensions, and only light pressure will be needed to restore the profile of the parts as they are bonded into assemblies. For example, 9 m (30 ft) long Lear Fan unstiffened fuselage skin panels, which were mildly compound curved, were rolled up like giant cigars to stiffen them for safer as well as easier handling as they were transported from the autoclave to the assembly room. The more common flat panels flex even more easily. Obviously, the entire shape must be well defined for the final bonding operation since the assembly will be too rigid to match the surrounding structure if it is made to the wrong shape. However, the bond cure cycle is typically 25 to 40 °C (50 to 100 °F) lower in peak temperature than the cure cycle for detail composite parts, so the problems of variable temperatures are diminished.

Effects of Prebond Moisture on the Strength of Bonded Joints. Another widespread problem is the need for the removal of volatiles before the start of cure and throughout its duration. Some manufacturers of composite and bonded structures want materials that are so tolerant of abusive cure cycles that it is permissible to: (1) apply full pressure from the start of the cure, flipping one switch for the heaters (with no automatic control system); and (2) to come back once during the cure at a preestablished time to cut back on the heat when the peak temperature has been attained and once more to switch the heat off when the cure is complete, preferably quickly enough not to need the services of a second shift of workers. Material development engineers have failed to create such forgiving materials, even though some process specifications permit parts to be made this way, so there are opportunities for using closed-loop control systems. Such systems, however, require experts to anticipate what should happen, what might happen, and what must not be permitted to happen--along with corrective responses.

The reliability of manufacture will be increased by demanding that all materials and parts be dried thoroughly immediately before each cure cycle. This refers also to Nomex honeycomb cores, which are known to have a great affinity for water and to absorb it rapidly if left in a humid environment. Wet cores result in weak glue fillets at honeycomb cell walls (Fig. 12). Such a goal of thorough drying exceeds the current practice at many composite manufacturing facilities today, whether or not the up-front savings exceed the consequently higher down-stream costs. But even if thorough drying were the norm, bagging procedures that permitted small amounts of moisture to be removed harmlessly during cure would also be desirable. The initial manufacturer also needs to be aware of the very high costs (to the consumer and/or the manufacturer) whenever a composite part or assembly is manufactured with sufficient trapped moisture to create an inferior part of unreliable service strength, but sufficient initial strength to be accepted after normal ultrasonic inspection. Under the prevailing circumstances, the original manufacturer is exposed to the consequences of improper processing only when there is sufficient residual strength in a rejected bonded panel to make it virtually impossible to disassemble the pieces for rebonding without destroying them by delaminations in the process.

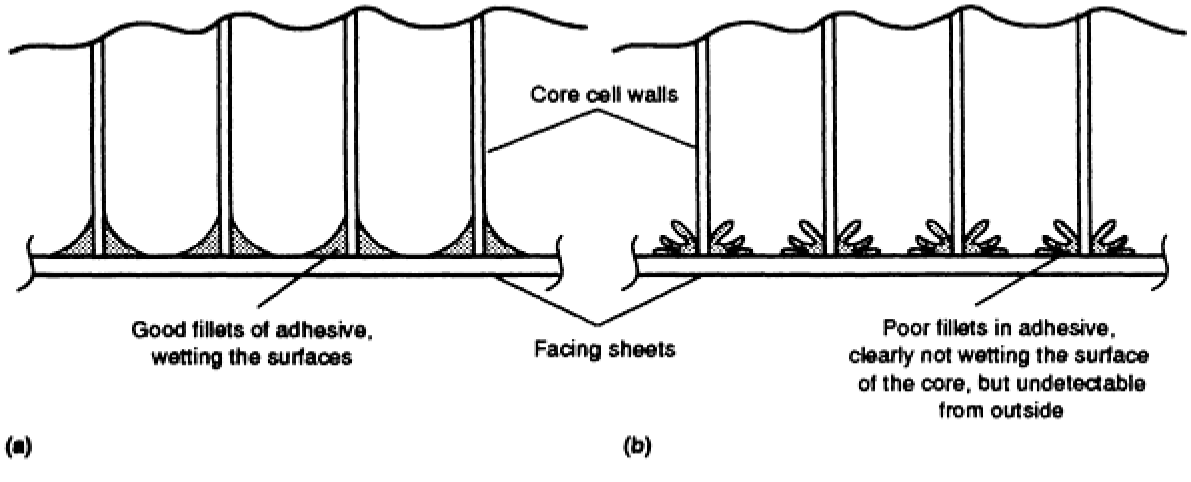
Core cell walls
Good fillets of adhesive, wetting the surfaces
Facing sheets
Poor fillets in adhesive, clearly not wetting the surface of the core, but undetectable from outside
(a)
(b)

FIG. 12 EFFECT OF NOT PREDRYING ORGANIC HONEYCOMB CORES. (A) DRIED CORE. (B) UNDRIED CORE

Adhesives vary in their susceptibility to problems caused by moisture absorption. The first generation of modified epoxy adhesives introduced in the 1960s were discredited because of corrosion problems with bonded metal structures. These adhesives had a high permeability for moisture absorption after cure. This led to corrosion of the metal surfaces, which were usually only etched, not anodized, and were not protected by a corrosion-inhibiting primer. Ironically, these same adhesives have been used successfully with composite adherends that cannot corrode, but they have suffered limited application because of their bad reputation earned in a totally unrelated context. Actually, it was proved more than 20 years ago that the primary cause of such problems was inadequate surface preparation. However, other contemporary adhesives, cured at higher temperatures, lasted longer on the same poor surfaces and were then used in their place.

The second generation of modified epoxy adhesives, introduced in the late 1970s, were specially formulated to impede moisture migration and were used in combination with a phenolic-based primer baked on an aluminum alloy surface anodized with phosphoric acid. Their durability was established during the PABST program, and they have performed well in metal bonded structures. However, this class of adhesives has a great affinity for moisture in the uncured state. One of the first developed was so sensitive that it was withdrawn from sale. Other adhesives in this class are now associated with numerous problems with bonding composite laminates together and with the co-cure and bonding of composite patches. The problem, which was exposed and evaluated during the PABST program, occurs only when the adhesive is allowed to absorb moisture before bonding or when making sandwich panels with wet foam or Nomex honeycomb cores. The adhesive film can be restored by drying before use, but the best solution is to avoid the problem by limiting the exposure to humid environments or by regulating the atmosphere in the lay-up room.

It is vital that adhesives be stored in relatively small rolls so that the supply of material for an excessive lay-up period is not exposed to the atmosphere. They should be removed from frozen storage the day before they are to be used and allowed to thaw out overnight while still sealed, to prevent condensate from forming on the adhesive. If they do absorb moisture, they should be dried before further use. It is better to close down production for a day than to take a roll of adhesive straight out of the freezer, unwrap it, and start using it immediately while it is covered with condensate.

Laboratory experiments performed to prove that trapped moisture could actually prevent the adhesive from sticking, not merely degrade its properties, required fairly large blister panels to replicate the effect. It was shown that 150 mm (6 in.) square panels that were sealed around the edges would show the effect, but that smaller ones would not. It follows, therefore, that the conventional half-square-inch lap-shear test coupons used as travellers while parts are made cannot possibly expose this condition. If the adhesive has been exposed to high humidity and temperature for even the normal duration of a lay-up operation, or covered with condensate because it was unsealed before it had thawed out, the weak bonds will almost certainly be in the structure. If left out for an entire week, them is a good chance that the pieces will fall apart as they are removed from the bonding tool.

What is far worse than the above scenario is a degraded adhesive bond with sufficient residual strength that it does fall apart. There is then no way of characterizing the strength or the time taken for the bonds to fail in service. They will pass all ultrasonic inspections, unless there was so much trapped water that large blisters occurred during cure.

Better composite parts can be made more cheaply if the manufacturer starts with a slow vacuum-consolidation cycle, usually overnight, not in an autoclave, but possibly in a low-temperature oven or under an insulating styrofoam shelter. This enables trapped air bubbles to escape, solvents to be driven off, and some moisture to be expelled. Failure to do this, or the equivalent step of thoroughly drying detail parts to be bonded, as well as the adhesive itself if it is suspected of absorbing moisture, guarantees that all of these ingredients will be trapped inside the part once the autoclave pressure is applied. Full vacuum would be applied inside the bag until slightly short of the minimum viscosity point for the resin matrix, or until no further volatiles could be detected in the vent, if that occurred sooner. Most resins will boil if cured under high vacuum, so the bag needs to be vented to the atmosphere once the vent is no longer useful. The cure can then proceed, with the autoclave gage pressure holding the laminate down and the resin in the bag being subjected to the atmospheric pressure. After gelation, the pressure felt by the rosin will increase progressively toward that experienced inside the autoclave.

Partial-Vacuum Cures. A minor variation of the scheme above can eliminate the need for autoclaves, when curing simple parts that need little pressure to hold the laminates against the tool surface. After an unusually thorough consolidation cycle, the cure would start with a full vacuum inside the bag, using only an oven as a heat source. Once the point of minimum viscosity was reached, a regulated pressure of between one-third and two-thirds of an atmosphere would be maintained inside the bag, with the remaining differential pressure holding the laminate onto the tool. Many, but not all, resins should be capable of being cured this way. Phenolics, which generate steam during cure, are the most obvious exception.

There are new adhesives and resins being developed that will tolerate curing under full vacuum throughout the entire cycle, but the approach outlined above would produce better laminates and bonds even with such materials. It would also eliminate the need for an autoclave for many more resins and adhesives than a requirement for full vacuum would permit. It is a technique worthy of further study, particularly for repairs to thin composite sections. In-house assessments of both new and old adhesives at Douglas Aircraft have shown that the partial-vacuum cure achieves bond strengths almost as great as those produced by full autoclave pressure, and far better than those achieved under a full-vacuum cure of the same adhesive. This is true of adhesives specially developed for vacuum-bag curing as well as those formulated for autoclave cures and known to be totally unsatisfactory if cured under full vacuum. It is likely that the need for more heat to maintain the heat-up rate on thicker panels, or with large tools that have not been thermally isolated from the work, would favor the use of an autoclave because the thermal conductivity of air increases as it is compressed.

Susman (Ref 14) has described the benefits of the dwell cycle for autoclave cures. Maintaining full vacuum while holding the temperature for about an hour at some 55 °C (100 °F) below the cure temperature permitted such a thorough removal of volatiles that the interlaminar shear strengths were doubled with respect to traditional cures. The worst laminates were made by the so-called straight-up cure, in which both full heat and full autoclave pressure were applied at the start of cure to minimize the cycle time.

Reference cited in this section

14. S.E. SUSMAN, THE DWELL CYCLE: A COST SAVINGS ROUTE TO QUALITY COMPOSITES, *25TH NATIONAL SAMPE SYMPOSIUM*, SOCIETY FOR THE ADVANCEMENT OF MATERIAL AND PROCESS ENGINEERING, 6-8 MAY 1980, P 251-258

Joining of Organic-Matrix Composites

L.J. Hart-Smith, McDonnell Douglas Aerospace

Bonding of Thermoplastic Composite Panels

Three different approaches have been evaluated for bonding thermoplastic composite components together. Adhesive bonding has been used, much as for thermoset composites, but needing different surface preparations. Fusing the panels together with or without an interlayer has also been used. This technique is of interest to engineers because it suggests that repairs could be performed far more easily than is possible with thermoset composites. However, there is the obvious disadvantage that it is difficult to melt such a localized area of resin without distorting the rest of the structure. Perhaps the most promising approach is that of coating the surfaces to be bonded with a different thermoplastic, which melts at a lower temperature than that in the basic laminate. Fusion bonding is then possible at a temperature low enough not to distort the part itself.

Bonding with Thermoset Adhesives. The reluctance of liquid polymers to wet solid thermoplastics is even greater than for thermosets. Having the surfaces free from contamination is but one of many criteria that must be met for one material to stick to another. Having the right surface energies is equally critical. An uncured resin or adhesive will typically not wet a cured surface having a far lower surface tension, as is achieved with release coatings. It is much more likely to adhere to a surface of far higher surface energy, as can be created by oxidation via corona treatment. Figure 13 shows how little corona energy is needed to almost double the surface energy of polyether ether ketone (PEEK). However, testing also showed that far more corona energy was needed to ensure a cohesive rather than interfacial failure (Ref 8). This supports the assessment that even after surface contaminants have been removed, there are still polymer layers of low surface energy between the bulk matrix and the surface created by curing against an inert tool surface. Much the same problem exists when trying to epoxy bond injection-molded thermoplastic inserts into honey-comb sandwich panels. The glue will not stick unless the initially inert plastic surface has been activated to form a chemical bond or etched to permit a mechanical bond.

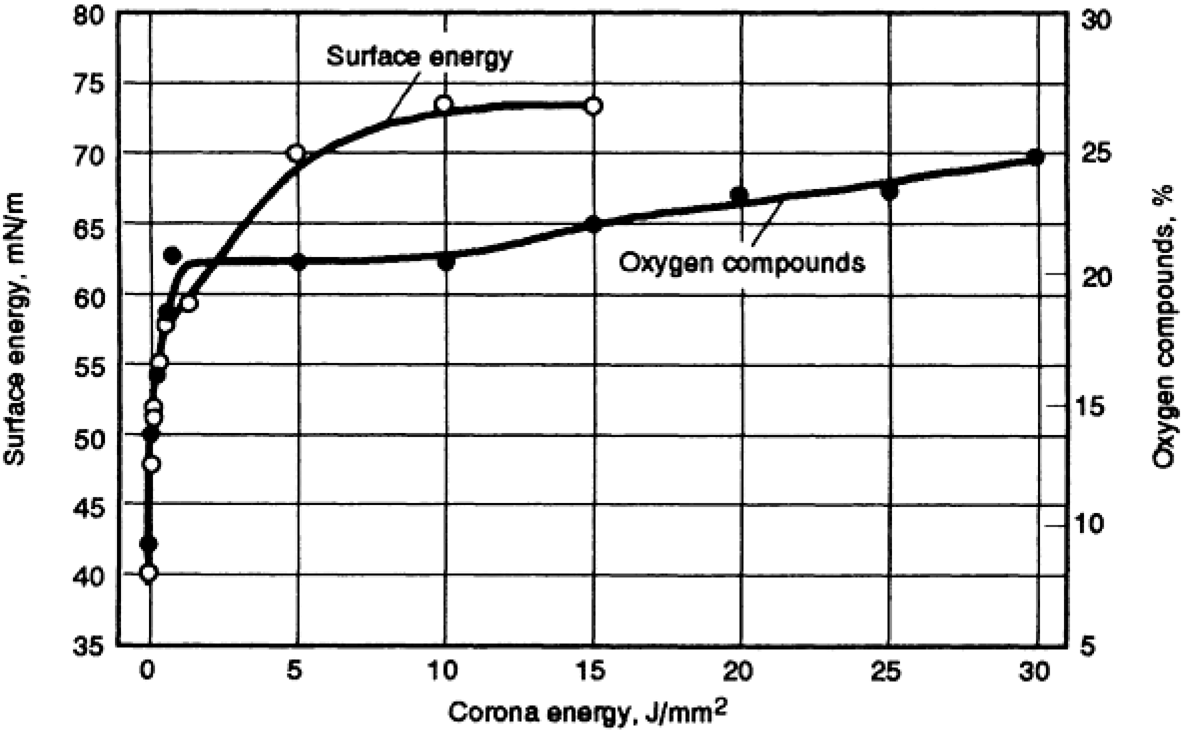

FIG. 13 EFFECT OF CORONA TREATMENT ON THE SURFACE ENERGY OF PEEK. SOURCE: REF 8

Wu *et al.* (Ref 15) describe the adhesive bonding of thermoplastic composites using several adhesives and a variety of surface treatments. Figure 14 shows results from their study. It is clear from Fig. 14 that weak interfacial failures are the inevitable result of failure to appreciate the need for proper surface treatment, for both the hot-bond epoxy FM-300 film adhesive and the BMI film adhesive EA-9673. Cohesive failures within the bond layer were achieved only for the chromic-sulfuric acid etch and plasma treatments. Significantly, these are the highest strengths obtained. While the highest average shear strengths are slightly less than one would expect on anodized aluminum coupons, they are credible for fibrous composite adherends having a quasi-isotropic fiber pattern and a slightly lower modulus. Higher strengths should have been obtained on unidirectional test coupons.

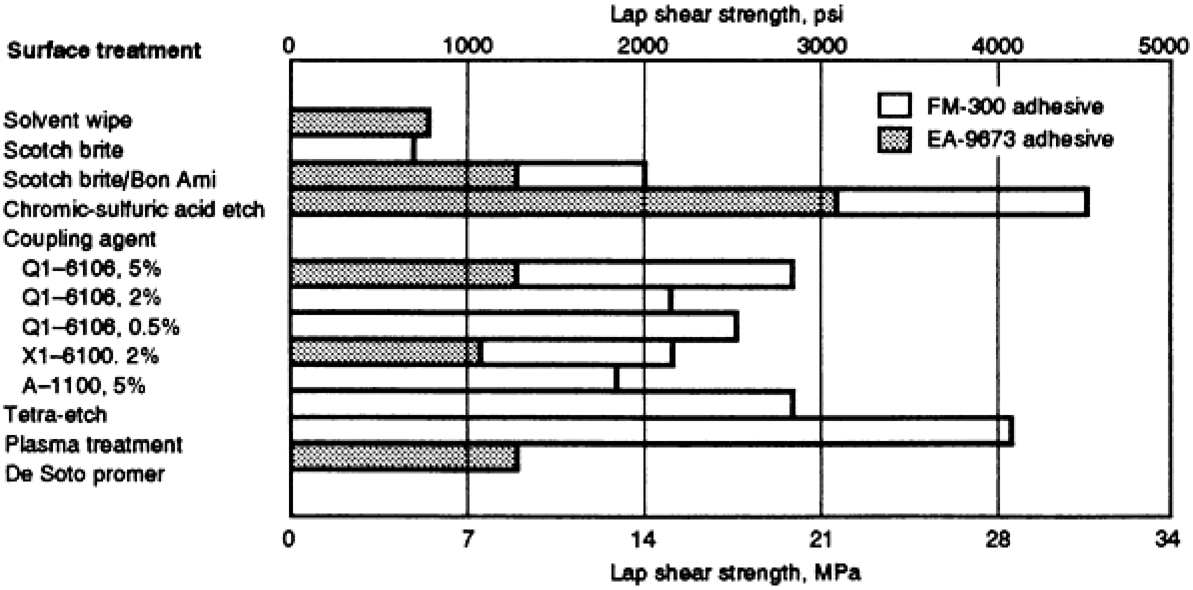

FIG. 14 EFFECT OF SURFACE TREATMENT ON THE LAP SHEAR STRENGTH OF AS4-APC-2 LAMINATE. SOURCE:

REF 15

Other researchers have also reported success with plasma treatment, with various gaseous environments. While the technique is effective, and probably quite practical for small parts, the need to enclose the composite part in an evacuated chamber during the surface preparation would seem to render it impractical for very large components. On the other hand, the reliability of this process in comparison with others is of paramount concern. The acid etching would require a large tank, in any event, and the corona-discharge technique would require electrodes tailored for each part to achieve uniform treatment over the area if the thickness varied or if the part were contoured. The capital cost of the equipment for carrying out the process correctly is a small fraction of the cost of scrapping one or two large composite assemblies because the surfaces had not been prepared properly while there was access to the individual panels.

Figure 15 makes it clear that mere abrasive cleaning of thermoplastic adherends is unsatisfactory, no matter what thermoset adhesive is used, while the chromic-sulfuric acid etch is effective with all the adhesives tested. In other tests, Wu and co-workers found that extended etching did not increase the joint strength; worse, it pitted and degraded the surface. They also established by preliminary testing that them is a need to consider the environmental durability of bonds to thermoplastic adherends, just as them is with thermoset adherends.

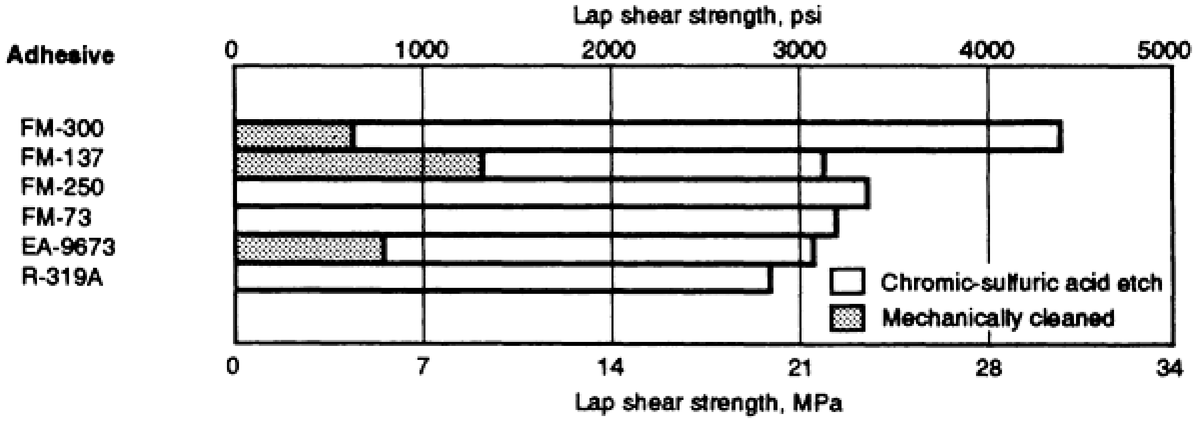

FIG. 15 EFFECT OF ABRASIVE CLEANING AND ACID ETCHING ON THE LAP SHEAR STRENGTH OF AS4/APC-2 LAMINATES. SOURCE: REF 15

Fusion Bonding of Thermoplastic Composites. Wu and co-workers also included some welded thermoplastic joints in their study (Ref 15). Lap-shear coupons that were bonded by ultrasonic welding exhibited good strengths for this process, whether the surface had been acid etched or merely abraded (Fig. 16). However, the procedure was not considered practical for large structures. Wu and co-workers examined such techniques as resistance welding, induction welding, laser welding, hot plate welding, and ultrasonic welding. It was determined that the high carbon-fiber content in the thermoplastic matrix and the high thermal and electrical conductivities of the fibers impose a tremendous amount of difficulty on fusion bonding processing. Uneven heating throughout the bonded area and delamination and distortion of the composite materials must be overcome with fusion bonding methods. In addition, such welding methods will change the crystalline structure of the resin because the remelting would be followed by cooling at a different rate from that experienced when the laminate was originally made.

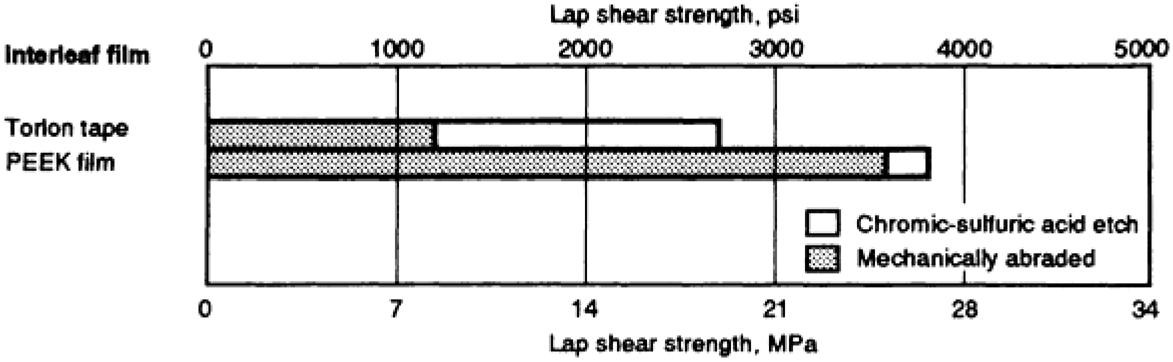

FIG. 16 EFFECT OF ABRASIVE CLEANING AND ACID ETCHING ON FUSION-BONDED AS4-APC-2 LAMINATES. TEST COUPONS WERE JOINED BY ULTRASONIC WELDING. SOURCE: REF 15

Fusion Bonding Using the Dual Resin Approach. The problems cited above can be avoided by using an interlayer that melts at a lower temperature than the parent thermoplastic matrix, but which is miscible with it. Several studies have been made using polyetherimide (PEI) to join PEEK laminates. Zelenak and Radford (Ref 16) cite success with a process in which a PEI film was melted onto the surface of APC2/PEEK tubes prior to joining. The PEI flowed into the PEEK parent, becoming a part of the parent material. This process occurred at the 320 °C (610 °F) processing temperature for the PEEK. To form the actual joint, another film of PEI was placed between the pieces to be joined as a filler. The joint was made under light pressure at a temperature of 300 °C (570 °F). This temperature is substantially above the glass transition temperature of PEI, which is 210 °C (410 °F), but below the melting temperature of PEEK (334 °C, or 633 °F). At this temperature the joint was formed by the PEI film bonding with the PEI in the parent material, which was intimately mixed with the PEEK material.

Unfortunately, the testing was marred by the limited strength of the structure outside the joint which, while it may have been able to carry the design loads, was not strong enough to fail either the laminate adjacent to the joint or the bond itself, except for two premature failures. Consequently, the absolute strengths of the joints were not established. However, in other regards, the use of PEI/PEEK as a miscible combination for bonding of structural PEEK matrix composites gives excellent joint performance with minimal susceptibility to surface contamination during bonding. While there remains the question of resistance to the environment, various chemicals, and age, this method is most promising. It would obviously be feasible to use the same technique to either repair damage in service or to add additional bonds as requirements changed, provided that the entire surface of the parts had been coated with PEI when the details were initially made.

Yoon and McGrath (Ref 17) have reported on bonding PEEK laminates with polyimide thermoplastic adhesive films. Their work is similar to that above in terms of processing at temperatures insufficient to melt the PEEK, but it differed significantly in that they bonded directly to the PEEK laminate. The absence of a compatible surface layer caused them to investigate the effectiveness of different surface treatments. They were able to make very good bonds without adhesive at 360 °C (680 °F), developing shear strengths of 47 MPa (6800 psi), which are not surprising given that their adherends were 24-ply thick unidirectional carbon-fiber reinforced laminates. However, they concluded that this temperature could not be used due to severe deformation. Their joints that fused together at 340 and 350 °C (645 and 660 °F) attained less than 20% of the highest strengths reached with adhesive films fused to treated surfaces that had been promptly primed with a solution of the adhesive to protect the surface. It was concluded that the highest practical temperature for the PEEK-graphite composite adhesion was around 340 °C (645 °F); however, high pressure was also required. Simple fusion of thermoplastic laminates does not seem to be a productive avenue at the present time.

It should be pointed out, however, that the work of Yoon and McGrath in activating the surface prior to bonding with thermoplastic adhesive closely parallels what was reported earlier for bonding with thermoset adhesives. Grit blasting showed a distinct improvement but clearly failed to enable the true material strengths to be exploited. The various gas plasma treatments were effective in producing cohesive failures, and the combinations of grit blasting and gas plasma treatments were better still, reaching nearly 41 MPa (6000 psi) and coming close to the strengths attained by fusion at 360 °C (680 °F).

Of the successful techniques described above, only one is applicable for joining large components. Fusion welding at a reduced temperature, made possible because the thermoplastic parts have been coated during fabrication with a miscible plastic that melts at a lower temperature than does the base matrix, should enable full use to be made of the strength of the parent laminates. The plasma treatment and corona-discharge techniques are reliable for small parts. The other techniques for developing high strengths are nevertheless invaluable as a warning of how weak the structures will be if no attention is paid to the need for proper surface preparation. They establish the target that should be obtained by any method used in practice.

Solvent Bonding of Amorphous Thermoplastics. The difficulties cited above, when bonding aircraft-grade fiber-reinforced thermoplastic composites, apply specifically to crystalline thermoplastics. Amorphous thermoplastics have not been used widely for such structures because of their lower resistance to common solvents. However, this greater susceptibility to solvents can become an advantage for adhesive bonding. The extreme case of assembling polystyrene plastic model aircraft kits by local application of styrene monomer cement or methyl ethyl ketone solvent is well known. Solvent bonding of nonstructural thermoplastics is quite common in other industries. The same effect can also be used to enhance the bonding of amorphous thermoplastics by epoxy-type adhesives. Such adhesives do not normally bond well to these plastics. However, wiping the surfaces with a suitable solvent and bonding the parts together before the thermoplastic has time to revert to an inert state can produce adhesive bonds strong enough to tear honeycomb cores apart; that is, strength similar to good metal-bond structures. This has been demonstrated by the Ten Cate organization for their continuous-fiber-reinforced PEI thermoplastic bonded with 120 °C (250 °F) cured modified epoxy adhesives. They even advocate the direct use of solvent bonding for cosmetic repairs of local damage to such structures. This technique is no panacea, however, because some thermoplastics will craze after being subjected to the wrong solvent. Nevertheless, under suitable circumstances, solvent bonding or solvent-enhanced bonding can be a very useful technique.

References cited in this section

8. A.J. KINLOCH AND G.K.A. KODOKIAN, "THE ADHESIVE BONDING OF THERMOPLASTIC COMPOSITES," EUROPEAN RESEARCH OFFICE OF THE U.S. ARMY, CONTRACT DAJA 45-86-C-0037, FOURTH PERIODIC REPORT, IMPERIAL COLLEGE, LONDON, AUG-DEC, 1987
15. S-I.Y. WU, A.M. SCHULER, AND D.V. KEANE, ADHESIVE BONDING OF THERMOPLASTIC COMPOSITES, 1. THE EFFECT OF SURFACE TREATMENT ON ADHESIVE BONDING, *19TH INT. SAMPE TECH. CONF.,* SOCIETY FOR THE ADVANCEMENT OF MATERIAL AND PROCESS ENGINEERING, 13-15 OCT 1987, P 277-290
16. S. ZELENAK, D.W. RADFORD AND M.W. DEAN, THE PERFORMANCE OF CARBON FIBER REINFORCED PEEK SUBASSEMBLIES JOINED USING A DUAL RESIN BONDING APPROACH, *37TH INTERNATIONAL SAMPE SYMPOSIUM,* SOCIETY FOR THE ADVANCEMENT OF MATERIAL AND PROCESS ENGINEERING, 9-12 MARCH 1992, P 1346-1356
17. T-H. YOON AND J.E. MCGRATH, ADHESION STUDY OF PEEK/GRAPHITE COMPOSITES, *36TH INTERNATIONAL SAMPE SYMPOSIUM,* SOCIETY FOR THE ADVANCEMENT OF MATERIAL AND PROCESS ENGINEERING, 15-18 APRIL 1991, P 428-436

Joining of Organic-Matrix Composites

L.J. Hart-Smith, McDonnell Douglas Aerospace

References

1. L.J. HART-SMITH, JOINTS, *COMPOSITES,* VOL 1, *ENGINEERED MATERIALS HANDBOOK,* T.J. REINHART, ED., ASM INTERNATIONAL, 1987, P 479-495
2. E.W. THRALL, JR. AND R.W. SHANNON, ED., *ADHESIVE BONDING OF ALUMINUM ALLOYS,* MARCEL DEKKER, 1985
3. R.J. SCHLIEKELMANN, ADHESIVE BONDING AND COMPOSITES, *PROGRESS IN SCIENCE AND ENGINEERING OF COMPOSITES,* VOL 1, PROC. 4TH INT. CONF. COMPOSITE MATERIALS (TOKYO), 22-28 OCT 1982, JAPAN SOCIETY FOR COMPOSITE MATERIALS AND THE

METALLURGICAL SOCIETY (TMS) OF AIME, T. HAYASHI, K. KAWATA, AND S. UMEKAWA, ED., NORTH-HOLLAND, AMSTERDAM, 1982, P 53-78

4. L.J. HART-SMITH, R.W. OCHSNER, AND R.L. RADECKY, SURFACE PREPARATION OF FIBROUS COMPOSITES FOR ADHESIVE BONDING OR PAINTING, *DOUGLAS SERVICE MAGAZINE,* FIRST QUARTER 1984, P 12-22
5. L.J. HART-SMITH, R.W. OCHSNER, AND R.L. RADECKY, SURFACE PREPARATION OF FIBROUS COMPOSITES FOR ADHESIVE BONDING OR PAINTING, *CANADAIR SERVICE NEWS,* VOL 14 (NO. 2), SUMMER 1985, P 2-8
6. B.M. PARKER AND R.M. WAGHORNE, SURFACE PRETREATMENT OF CARBON FIBRE-REINFORCED COMPOSITES FOR ADHESIVE BONDING, *COMPOSITES,* VOL 13, JULY 1982, P 280-288
7. A.V. POCIUS AND R.P. WENZ, MECHANICAL SURFACE PREPARATION OF GRAPHITE-EPOXY COMPOSITE FOR ADHESIVE BONDING, *SAMPE J.,* SEPT/OCT 1985, P 50-58
8. A.J. KINLOCH AND G.K.A. KODOKIAN, "THE ADHESIVE BONDING OF THERMOPLASTIC COMPOSITES," EUROPEAN RESEARCH OFFICE OF THE U.S. ARMY, CONTRACT DAJA 45-86-C-0037, FOURTH PERIODIC REPORT, IMPERIAL COLLEGE, LONDON, AUG-DEC, 1987
9. J.D. LABOR AND S.H. MYHRE, "LARGE AREA COMPOSITE STRUCTURE REPAIR," NORTHROP AIRCRAFT CORPORATION, USAF TECHNICAL REPORT AFFDL-TR-79-3040, MARCH 1979
10. B.M. PARKER, THE EFFECT OF COMPOSITE PREBOND MOISTURE ON ADHESIVE-BONDED CFRP-CFRP JOINTS, *COMPOSITES,* VOL 14, JULY 1983, P 226-232
11. E.W. THRALL, JR., PROSPECTS FOR BONDING PRIMARY AIRCRAFT STRUCTURES IN THE 80'S, *25TH NATIONAL SAMPE SYMPOSIUM,* VOL 25, SOCIETY FOR THE ADVANCEMENT OF MATERIAL AND PROCESS ENGINEERING, 6-8 MAY 1980, P 716-727
12. I. MCARTHUR, HAWKER DEHAVILLAND, BANKSTOWN, AUSTRALIA, PRIVATE COMMUNICATION, 1992
13. L.J. MATIENZO, J.D. VENABLES, J.D. FUDGE, AND J.J. VELTEN, SURFACE PREPARATION OF BONDING ADVANCED COMPOSITES, PART 1; EFFECT OF PEEL PLY MATERIALS AND MOLD RELEASE AGENTS ON BOND STRENGTH, *30TH NATIONAL SAMPE SYMPOSIUM,* SOCIETY FOR THE ADVANCEMENT OF MATERIAL AND PROCESS ENGINEERING, 19-21 MARCH 1985, P 302-314
14. S.E. SUSMAN, THE DWELL CYCLE: A COST SAVINGS ROUTE TO QUALITY COMPOSITES, *25TH NATIONAL SAMPE SYMPOSIUM,* SOCIETY FOR THE ADVANCEMENT OF MATERIAL AND PROCESS ENGINEERING, 6-8 MAY 1980, P 251-258
15. S-I.Y. WU, A.M. SCHULER, AND D.V. KEANE, ADHESIVE BONDING OF THERMOPLASTIC COMPOSITES, 1. THE EFFECT OF SURFACE TREATMENT ON ADHESIVE BONDING, *19TH INT. SAMPE TECH. CONF.,* SOCIETY FOR THE ADVANCEMENT OF MATERIAL AND PROCESS ENGINEERING, 13-15 OCT 1987, P 277-290
16. S. ZELENAK, D.W. RADFORD AND M.W. DEAN, THE PERFORMANCE OF CARBON FIBER REINFORCED PEEK SUBASSEMBLIES JOINED USING A DUAL RESIN BONDING APPROACH, *37TH INTERNATIONAL SAMPE SYMPOSIUM,* SOCIETY FOR THE ADVANCEMENT OF MATERIAL AND PROCESS ENGINEERING, 9-12 MARCH 1992, P 1346-1356
17. T-H. YOON AND J.E. MCGRATH, ADHESION STUDY OF PEEK/GRAPHITE COMPOSITES, *36TH INTERNATIONAL SAMPE SYMPOSIUM,* SOCIETY FOR THE ADVANCEMENT OF MATERIAL AND PROCESS ENGINEERING, 15-18 APRIL 1991, P 428-436

Joining of Oxide-Dispersion-Strengthened Materials

David O'Donnell, Inco Alloys International Inc.

Introduction

OXIDE - DISPERSION - STRENGTHENED (ODS) MATERIALS utilize an extremely fine oxide dispersion for strengthening, such as yttria (Y_2O_3) or alumina (Al_2O_3). The majority of ODS alloys are produced by mechanical means, often referred to as mechanical alloying (MA), where elemental or alloyed powders are ball milled, or attrited. This mechanical working of different powders between large shot will cold weld particles and break them apart until a fine interlayered structure is produced. The powder is typically consolidated and homogenized by hot isostatic pressing or hot extrusion. Further consolidation is typically performed, producing a final product that is more than 99% dense (Ref 1, 2). The compositions of selected ODS alloys are given in Table 1.

TABLE 1 COMPOSITION OF SELECTED ODS ALLOYS

ODS ALLOY	**COMPOSITIONS, %**								TYPICAL APPLICATIONS
	Cr	**Ni**	**Fe**	**Ti**	**Al**	**C**	**Y_2O_3**	**OTHER**	
MA 956	20.0	. . .	74.0	0.5	4.5	. . .	0.5	. . .	**COMBUSTOR COMPONENTS**
PM 2000	20.0	. . .	73.0	0.5	5.5	. . .	0.5	. . .	**. . .**
MA 754	20.0	78.0	1.0	0.5	0.3	0.05	0.6	. . .	**TURBINE VANES, NOZZLES**
PM 1000	20.0	79.0	. . .	0.5	0.3	. . .	0.6	. . .	**. . .**
MA 758	30.0	68.0	1.0	0.5	0.3	. . .	0.6	. . .	**GLASS EQUIPMENT**
MA 6000	15.0	69.0	. . .	2.5	4.5	0.05	1.1	4.0 W; 2.0 MO; 2.0 TA; 0.01 B; 0.15 ZR	**TURBINE BLADES**
PM 3030	17.0	68.2	. . .	. . .	6.0	<0.05	1.1	3.5 W; 2.0 MO; 2.0 TA; 0.01 B; 0.15 ZR	**. . .**
AL-905XL	. . .	. . .	. . .	. . .	93.2	1.1	. . .	4.0 MG; 1.3 LI; 0.4 O_2	**AIRFRAME COMPONENTS**
AL-9052	**. . .**	**. . .**	**. . .**	**. . .**	**94.5**	**1.1**	**. . .**	**4.0 MG; 0.4 O_2**	**MARINE CORROSION**

MA, mechanically alloyed; PM, powdered metal

A typical dispersoid diameter of Incoloy alloy MA 956, for example, is on the order of 30 nm (1.2×10^{-6} in.), which is too small for observation using optical metallographic techniques (Ref 3). The dispersion of stable oxides provides exceptional creep-rupture strength at temperatures above 980 °C (1800 °F) because carbides and many secondary phases used for strengthening conventional wrought alloys go into solution. The level of creep strength achieved at very high temperatures in ODS alloys has not been matched by similar wrought or cast alloys.

The processing techniques employed in the production of ODS alloys produce some entrapped gases, which tend to create porosity during fusion welding. Porosity can form either within the bulk of the fusion zone or near the fusion boundary, as a very fine porosity interface. This tendency is generally most severe in aluminum alloys and least evident in nickel-base alloys. The porosity tendency in iron-base alloys is between that of aluminum- and nickel-base alloys. The level of porosity that results from the fusion welding of aluminum ODS alloys is such that they are not generally considered to be fusion weldable.

Most ODS applications involve high temperatures and, possibly, corrosive environments. The first commercial applications were aircraft gas-turbine engines, followed by industrial engines. Vane airfoils are typically furnace brazed to platforms, both of which can be made from ODS alloys. Combustor parts can be riveted, resistance welded, or brazed. Some ODS alloys are being used to a significant extent in various components of molten glass handling equipment. Heat-treatment baskets, trays, shields, and furnace skid rails are also becoming common applications.

References

1. L.R. CURWICK, THE MECHANICAL ALLOYING PROCESS: POWDER TO MILL PRODUCT, *FRONTIERS OF HIGH TEMPERATURE ALLOYS I, PROC. 1ST INT. CONF. ON OXIDE DISPERSION STRENGTHENED ALLOYS BY MECHANICAL ALLOYING*, J.S. BENJAMIN, ED., INCO ALLOYS INTERNATIONAL, 1981, P 3-10

2. G.A. HACK, FUNDAMENTALS OF MECHANICAL ALLOYING, *FRONTIERS OF HIGH TEMPERATURE ALLOYS II, PROC. 2ND INT. CONF. ON OXIDE DISPERSION STRENGTHENED ALLOYS BY MECHANICAL ALLOYING,* J.S. BENJAMIN AND R.C. BENN, ED., INCO ALLOYS INTERNATIONAL, 1983, P 3-18
3. R. PETKOVIC-LUTON, D.J. SROLOVITZ, AND M.J. LUTON, MICROSTRUCTURAL ASPECTS OF CREEP IN OXIDE DISPERSION STRENGTHENED ALLOYS, *FRONTIERS OF HIGH TEMPERATURE ALLOYS II, PROC. 2ND INT. CONF. ON OXIDE DISPERSION STRENGTHENED ALLOYS BY MECHANICAL ALLOYING,* J.S. BENJAMIN AND R.C. BENN, ED., INCO ALLOYS INTERNATIONAL, 1983, P 73-97

Joining of Oxide-Dispersion-Strengthened Materials

David O'Donnell, Inco Alloys International Inc.

Design Strategy

The high-temperature creep strength obtained from a cast weld structure, using ODS alloys, is generally comparable to that obtained with conventional wrought alloys and does not exhibit the same exceptional high-temperature creep strength of unwelded ODS materials. The stable oxides, or dispersoids, that are used in the ODS alloy systems have essentially no equilibrium solubility in the liquid phase. If melting occurs, then the dispersoids rapidly agglomerate, and buoyancy floats the oxide to an upper surface, if possible. In fusion welds, this results in the loss of the primary means of high-temperature strengthening. Agglomerated semicontinuous dispersoids in the partially melted zones of fusion welds are also suspected as contributing to early weld-service failures.

A number of successful strategies have been employed in joining ODS alloys. Conventional fusion welds or brazes are often placed in cooler locations in order to retain acceptable strength levels. A design is often altered so that conventional fusion welds are not the principal load-carrying members, but are attachments instead. Solid-state joining techniques can often be used, but the resulting grain structure must also be considered.

Grain Structure. At very high temperatures, grain boundaries become the weak link, so to speak, and fracture becomes intergranular. In order to minimize such a grain boundary strength limitation, ODS alloys are produced with a macro grain size, where the grain boundary area normal to the principal stresses are minimized without incurring the difficulties inherent in producing single crystals. This elongated grain structure is often referred to as a high grain aspect ratio (GAR). Welds or joints that significantly interrupt this large interlocked grain structure will fail prematurely at extreme temperatures, even if they are produced by solid-state means. A friction weld in a rod, where there is a straight grain boundary interface through the thickness and normal to the principal stresses, typically will not perform significantly better than a conventional fusion weld in which a similar structure is produced. The large, interlocking grain structure in such materials is a key attribute for good performance at extreme temperatures.

Because of grain-structure considerations, it is often preferable to use resistance welds, braze welds, or both in a lap joint, rather than a butt joint, where there is invariably a through-thickness noninterlocked grain structure normal to the principal stresses across the joint. When lap welds are used, double-lap joints outperform single-lap joints, which fail by peeling at lower stresses.

Postweld recrystallization anneals have been used with all types of welds to grow grains across the welds to the greatest extent possible. The best structural and strength results are generally obtained by joining base metals in a fine, unrecrystallized condition and by performing a grain-coarsening recrystallization anneal after welding. However, a postweld recrystallization anneal typically improves properties only in narrow joints, such as those commonly produced by electron-beam, diffusion, resistance, and friction welding techniques. Creep-rupture properties from wide-gap joints, such as those produced in gas-metal or gas-tungsten-arc welds can be reduced by such an anneal. This anneal is typically 1316 °C (2400 °F) for 1 h, when MA 754 and MA 956 alloys are used. The fine-grain approach, if adopted, requires coordination with materials suppliers, because most ODS alloys are typically supplied in the coarse-grained, recrystallized condition. However, some measure of improved properties may be obtained in narrow joints through postweld recrystallization, even where the base metal was joined in the coarse-grained condition. Table 2 provides transverse property data on selected alloys used in each of the welding processes discussed later in this article.

TABLE 2 WELD TRANSVERSE PROPERTIES OF SELECTED ODS ALLOYS

ODS ALLOYS	**CONDITION(A)**	**WELDING PROCESS(B)**	**JOINT TYPE**	**FILLER METAL**	**POSTWELD HEAT TREATMENT(C)**		**TEST TEMPERATURE**		**STRESS(D)**		**LIFE(E), H**	**LONGITUDINAL BASE-METAL STRENGTH, APPROXIMATE %**	COMMENT
					°C	**°F**	**°C**	**°F**	**MPA**	**KSI**			
MA 956	FG	FRW	BUTT	. . .	1315	2400	982	1800	77.9	11.3	HTT	75	**. . .**
	CG	FRW	BUTT	. . .	1315	2400	982	1800	6.9	1.0	24.0	8	**. . .**
	FG	FRW	BUTT	. . .	1315	2400	982	1800	6.9/13.8	1.0/2.0	13.0	17	**. . .**
MA 758	CG	GMAW	BUTT	52	. . .	. . .	1093	2000	13.8	2.0	40.4	11	**. . .**
	CG	GMAW	BUTT	52	. . .	. . .	1093	2000	6.9	1.0	1601	8	**. . .**
	CG	GMAW	BUTT	52	1315	2400	1093	2000	13.8	2.0	16.5	10	**. . .**
	CG	GMAW	BUTT	52	1315	2400	1093	2000	6.9	1.0	259.9	7	**. . .**
MA 956	CG	EBW	BUTT	. . .	. . .	. . .	982	1800	13.8	2.0	5.4	15	**90° BUTT**
	CG	EBW	BUTT	. . .	. . .	. . .	982	1800	27.6	4.0	38	33	**5° BUTT**
	CG	EBW	SINGLE LAP	. . .	. . .	. . .	1093	2000	13.8	2.0	710	25	**DOUBLE WELD**
	CG	EXW	SINGLE LAP	. . .	1100(F)	2012(F)	1100	2012	18/28/38	2.6/4.1/5.5	14.0	58	**STRESS IS BASE METAL**
	FG	EXW	SINGLE LAP	. . .	1315	2400	1100	2012	18/28/38	2.6/4.1/5.5	8.0	56	**STRESS IS BASE METAL**
	CG	EXW	SINGLE LAP	. . .	. . .	. . .	1100	2012	18	2.6	2.0	25	**STRESS IS BASE METAL**
	FG	RSW	DOUBLE LAP	. . .	(G)	(G)	1093	2000	37.9	5.5	HTT	42	**SINGLE-SPOT STRESS**
	FG	**RSW**	**DOUBLE LAP**	**. . .**	**(G)**	**(G)**	**982**	**1800**	**44.1(H)**	**6.4(H)**	**HTT**	**43**	**SINGLE-SPOT STRESS**

(A) FG, FINE GRAIN; CG, COARSE GRAIN.

(B) FRW, FRICTION WELDING; GMAW, GAS-METAL ARC WELDING; EBW, ELECTRON-BEAM WELDING; EXW, EXPLOSION WELDING; RSW, RESISTANCE SPOT WELDING.

(C) 1 H, EXCEPT WHERE NOTED.

(D) STEP LOADING OF STRESS DENOTED BY "/." STRESS WAS INCREASED EVERY 24 H UNTIL FAILURE.

(E) HTT, HIGH-TEMPERATURE TENSILE TEST.

(F) 10 H.

(G) 1204 °C (2200 °F) FOR 24 H PLUS 1315 °C (2400 °F) FOR 1 H.

(H) ATTACHMENT PIN SHEARED THROUGH SHEET

Other attributes of ODS alloys, relative to welding, are typical of their respective alloy groups. For example, ferritic ODS alloys will exhibit a ductile-to-brittle transition temperature (DBTT), and, because of their high alloy content, will have a marked sensitivity to delayed hydrogen cracking.

David O'Donnell, Inco Alloys International Inc.

Iron-Base ODS Alloys

The alloy MA 956 utilizes a coarse, high-aspect grain structure as part of its high-temperature strengthening ability. This structure is achieved through processing and a final recrystallization anneal of 1316 °C (2400 °F) before service. The alloy has a high melting point of 1480 °C (2700 °F), and is often used in service at 1093 °C (2000 °F) and above. The alloy is quite ductile above its DBTT, which is near room temperature. All strain-forming operations, such as bend tests, friction welding, and explosion welding, should be performed after heating to 100 °C (212 °F) or higher. Below the DBTT, the alloy can be notch brittle. The MA 956 alloy undergoes no solid-state phase transformations between melting and cryogenic temperatures, and it cannot be quench hardened.

Iron-base ODS alloys are generally very sensitive to delayed weld zone and heat-affected zone (HAZ) hydrogen cracking. Good low-hydrogen welding practice should be followed for all fusion-welding processes, and only low-hydrogen fusion welding processes are recommended. For this reason, shielded metal arc welding or other flux-bearing welding processes are generally not recommended. Delayed cracking has been observed, even in heavy resistance welds, when no precautions were taken.

Fusion welding should utilize low-hydrogen practices, on the assumption that enough hydrogen will be present to cause cracking, and should incorporate preheating, minimum interpass temperatures, and a postweld hydrogen diffusion treatment prior to cooling in order to prevent cracking. Preheating and use of minimum interpass temperatures ranging from 149 to 204 °C (300 to 400 °F) have generally prevented cracking during welding. A postweld diffusion treatment before cooling at 300 °C (572 °F) for 10 h, or 400 °C (752 °F) for 6 h, has been effective in removing hydrogen in section thicknesses less than 11 mm (0.4 in.) thick. More information on hydrogen diffusion treatments for various thicknesses and temperatures is provided in Ref 4. The unique aspects of the different welding processes are described in the section "Welding Processes" in this article.

Reference cited in this section

4. F.R. COE, *WELDING STEELS WITHOUT HYDROGEN CRACKING,* THE WELDING INSTITUTE, 1973

David O'Donnell, Inco Alloys International Inc.

Nickel-Base ODS Alloys

Most of the nickel-base ODS alloys use a high GAR, but some are also highly textured to provide the best resistance to thermal fatigue. This structure is achieved in MA 754 through processing and a final recrystallization anneal of 1316 °C (2400 °F) before service. In MA 6000, zone annealing is used for an optimum longitudinal GAR.

The alloy MA 758 has been used increasingly in handling equipment for molten glass; such equipment is exposed to high temperatures, stress, and corrosive glasses. Welds have been made, primarily for attachment purposes, using the gas-metal arc, gas-tungsten arc, and shielded metal arc welding processes. Filler consumables that contain 30% Cr, such as Inconel filler metal 52 or Inconel welding electrode 152, are needed to provide corrosion resistance to most molten glasses.

The alloy MA 6000, an advanced gas-turbine blade material that utilizes both oxide-dispersion and γ' strengthening, is not typically fusion welded. Its aluminum content may make it susceptible to strain-age cracking, which can occur in fast precipitation-hardening alloys where significant residual stresses are present. These stresses can be caused by cold working or welding. If the piece is not heated rapidly to annealing temperatures after welding, then it may age fast enough for the locked-in stresses to exceed the rupture strength, at which point cracking will occur. Furnace brazing and diffusion welding should be performed after plating the joint area with a thin layer of nickel to prevent the oxidation of aluminum, which could cause incomplete bonding.

The alloy MA 754 is primarily a turbine vane material that is used for its exceptional combination of creep-rupture strength and thermal-fatigue resistance. Most applications involve vacuum furnace brazing to similar and dissimilar alloys, in situations where the joint is located in relatively cool locations or is subjected to reduced stresses (Fig. 1).

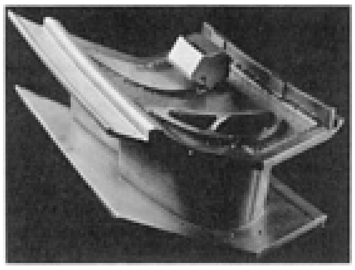

FIG. 1 FURNACE BRAZED MA 754 GAS TURBINE VANE

Joining of Oxide-Dispersion-Strengthened Materials

David O'Donnell, Inco Alloys International Inc.

Welding Processes

Gas-tungsten arc welding of ODS alloys is typically carried out using minimal welding current and techniques that are designed to minimize melting of the alloys. When welding alloys with a significant aluminum content, such as MA 956, there is a high propensity for an aluminum-rich light gray oxide film to form on the weld-bead surface. The dispersoid from the melted base metal also floats to the weld-bead surface. These oxides should be removed by light abrasive techniques before subsequent weld beads are deposited, in order to prevent incomplete-fusion defects or oxide inclusions. Balanced-wave alternating current is generally not recommended, because the reverse current (electrode positive) part of the cycle attempts to "cathodically clean" the dispersoid oxides from the bulk of the base metal by arcing into the base metal. Power supplies for which the reverse-polarity cycle time can be minimized have been used with some success, producing only a weld-bead surface cleaning. Figures 2 and 3 show MA 956 products that have been gas-tungsten arc welded using different filler metals.

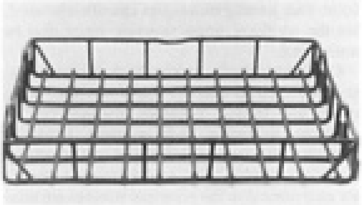

FIG. 2 GAS-TUNGSTEN ARC WELDED MA 956 HEAT TREAT BASKET. INCONEL 617 FILLER METAL WAS USED.

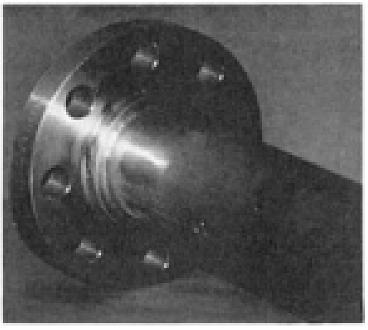

FIG. 3 MA 956 PIPE GAS-TUNGSTEN ARC WELDED TO A 304 STAINLESS STEEL FLANGE. 308 STAINLESS STEEL FILLER METAL WAS USED.

Because joint properties are representative of conventional wrought alloys, attempts should be made to either design welds as attachments that are subjected to reduced stresses or place the welds in cooler locations where they can withstand the stress. Table 3 lists typical filler-metal alloys for the ODS alloys discussed in this article.

TABLE 3 SELECTED WELDING CONSUMABLES FOR ODS ALLOYS

ODS ALLOY	**FILLER METAL**	**SHIELDED METAL ARC WELDING ELECTRODE**	BRAZING ALLOY
MA 956	BRIGHTRAY ALLOY ICA	. . .	**B93[(A)], AM 788[(B)], TD6[(C)]**
MA 754	INCONEL 617	. . .	**B93, AM 788, TD6**
MA 758	**INCONEL 52**	**INCONEL WELDING ELECTRODE 152**	

(A) GENERAL ELECTRIC ALLOY B50TF108, AVAILABLE IN POWDER FORM FROM ALLOY METALS, INC. COMPOSITION: 14% CR, 9.5% CO, 4.9% TI, 4% W, 4% MO, 3% AL, 4.5% SI, 0.7% B, BALANCE NI.

(B) AVAILABLE IN POWDER FORM FROM ALLOY METALS, INC. COMPOSITION: 22% CR, 21% NI, 14% W, 2% B, 0.03% LA, BALANCE CO.

(C) GENERAL ELECTRIC ALLOY B5DTF83, AVAILABLE IN POWDER FORM. COMPOSITION: 16% CR, 17% MO, 5% W, 5% SI, 1% AL, BALANCE NI

Gas-metal arc welding of ODS alloys is typically implemented with pulsed or synergic power sources and low-to-average current settings to minimize the melting of the base metals. This typically amounts to an average welding current of 60 to 180 A, depending on the section thickness of the joint.

Electron-beam and laser-beam welding have been used to join MA 956, and the best results have been obtained in lap joints where the beam is oriented as close to parallel to the joint as possible. In butt joints, the best results can be obtained by inclining the joint and beam as much as possible from 90°. This will reduce the load per unit area on the weld fusion-boundary and centerline grain structure (Ref 5, 6). At the same time, melting should be minimized and the beam should be projected vertically to help prevent incomplete-fusion defects from dispersoid entrapment. Because MA 956 is magnetic, it should be demagnetized before electron-beam welding to prevent unwanted beam deflection.

Resistance welding of ODS alloys is typically implemented with parameters designed to minimize the thickness of the weld nugget and solidification time (high current, tip pressure, and minimum weld time). The typical lap joint produces a favorable grain orientation relative to the weld and principal stress, and the process is fast and repeatable. The overlap and number of spots are controlled to develop the required joint properties. Delayed hydrogen cracking can occur in iron-base

ODS alloys, particularly in thicker restrained pieces if preheat, minimum interpass temperatures, and postweld diffusion treatments are not employed.

Dedicated efforts to minimize melting or the thickness of the weld nugget have succeeded in producing solid-state, or very-near-solid-state, resistance welds. The commercial practicality and benefits of these techniques still need to be better defined.

Furnace brazing of ODS alloys has been extensively used by the aircraft industry. Iron- and nickel-base ODS alloys containing significant aluminum or titanium are typically plated with a very thin barrier of nickel to prevent oxidation of these elements during the brazing cycle. If oxidation occurs, then the flow of the brazing filler metals is inhibited. Brazing filler metals that are rich in silicon and lean in boron are preferred to avoid erosion of the ODS base metals during the brazing cycle. If significant erosion occurs, then dispersoids can form oxide inclusions or incomplete-fusion defects within the joint. For best results, joint gaps should be as narrow as possible without hampering filler-metal flow. Typical properties are shown in Table 4. Attempts to improve properties have utilized special techniques, such as diffusion brazing, transient liquid-phase bonding, and pressure brazing.

TABLE 4 FURNACE-BRAZED WELD PROPERTIES

ODS ALLOY	JOINT TYPE	FILLER METAL	TEST TEMPERATURE		STRESS[A]		LIFE, H	LONGITUDINAL BASE-METAL STRENGTH, APPROXIMATE %	COMMENT
			°C	°F	MPA	KSI			
MA 754	DOUBLE LAP	B93	982	1800	110	16	392	85	**SHEARED BRAZE**
	DOUBLE LAP	AM 788	982	1800	110	16	168	84	**SHEARED BRAZE AND BASE METAL**
	DOUBLE LAP	AM 788	1093	2000	41.4	6	178	40	**SHEARED BRAZE**
	DOUBLE LAP	TD6	1093	2000	56.2	8	105.5	53	**SHEARED BRAZE**
MA 956	SINGLE LAP (8*T*)	B93	982	1800	41.4	6	89.4	30	**SHEARED BRAZE**
	SINGLE LAP (8*T*)	**AM 788**	**982**	**1800**	**48.3**	**7**	**23.5**	**58**	**SHEARED BRAZE**

(A) STRESSES ARE BELIEVED TO BE BASED ON BASE METAL.

Although it has not yet been confirmed, the aluminum-base ODS alloys AL-905XL and AL-9052 should vacuum furnace braze well, by virtue of their magnesium content.

Friction welding is capable of creating sound solid-state joints in ODS alloys. The plastic deformation and flow at the bond interface typically produce a relatively straight grain-boundary interface through-thickness after annealing. This relatively flat grain-boundary interface normal to the principal stresses is undesirable, and it produces properties that are no better than most fusion processes. The best results have been obtained by welding fine-grained unrecrystallized rod and then performing a postweld recrystallization anneal.

Explosion welding generally creates a favorable grain structure by virtue of the lap joints used. In a limited study in which MA 956 was joined, the best results were obtained by bonding coarse-grained material and giving it a postweld diffusion treatment of 1100 °C (2012 °F) for 10 h.

References cited in this section

5. T.J. KELLY, JOINING MECHANICAL ALLOYS FOR FABRICATION, *FRONTIERS OF HIGH TEMPERATURE ALLOYS II, PROC. 2ND INT. CONF. ON OXIDE DISPERSION STRENGTHENED ALLOYS BY MECHANICAL ALLOYING,* J.S. BENJAMIN AND R.C. BENN, ED., INCO ALLOYS

INTERNATIONAL, 1983, P 129-148

6. H.D. HEDRICH, H.G. MAYER, G. HAUFLER, M. KOPH, AND N. REHEIS, "JOINING OF ODS-SUPERALLOYS," ASME 91-GT-411, THE INTERNATIONAL GAS TURBINE AND AEROENGINE CONGRESS AND EXPOSITION (ORLANDO, FL), AMERICAN SOCIETY OF MECHANICAL ENGINEERS, JUNE 1991

Joining of Oxide-Dispersion-Strengthened Materials

David O'Donnell, Inco Alloys International Inc.

References

1. L.R. CURWICK, THE MECHANICAL ALLOYING PROCESS: POWDER TO MILL PRODUCT, *FRONTIERS OF HIGH TEMPERATURE ALLOYS I, PROC. 1ST INT. CONF. ON OXIDE DISPERSION STRENGTHENED ALLOYS BY MECHANICAL ALLOYING*, J.S. BENJAMIN, ED., INCO ALLOYS INTERNATIONAL, 1981, P 3-10
2. G.A. HACK, FUNDAMENTALS OF MECHANICAL ALLOYING, *FRONTIERS OF HIGH TEMPERATURE ALLOYS II, PROC. 2ND INT. CONF. ON OXIDE DISPERSION STRENGTHENED ALLOYS BY MECHANICAL ALLOYING,* J.S. BENJAMIN AND R.C. BENN, ED., INCO ALLOYS INTERNATIONAL, 1983, P 3-18
3. R. PETKOVIC-LUTON, D.J. SROLOVITZ, AND M.J. LUTON, MICROSTRUCTURAL ASPECTS OF CREEP IN OXIDE DISPERSION STRENGTHENED ALLOYS, *FRONTIERS OF HIGH TEMPERATURE ALLOYS II, PROC. 2ND INT. CONF. ON OXIDE DISPERSION STRENGTHENED ALLOYS BY MECHANICAL ALLOYING,* J.S. BENJAMIN AND R.C. BENN, ED., INCO ALLOYS INTERNATIONAL, 1983, P 73-97
4. F.R. COE, *WELDING STEELS WITHOUT HYDROGEN CRACKING,* THE WELDING INSTITUTE, 1973
5. T.J. KELLY, JOINING MECHANICAL ALLOYS FOR FABRICATION, *FRONTIERS OF HIGH TEMPERATURE ALLOYS II, PROC. 2ND INT. CONF. ON OXIDE DISPERSION STRENGTHENED ALLOYS BY MECHANICAL ALLOYING,* J.S. BENJAMIN AND R.C. BENN, ED., INCO ALLOYS INTERNATIONAL, 1983, P 129-148
6. H.D. HEDRICH, H.G. MAYER, G. HAUFLER, M. KOPH, AND N. REHEIS, "JOINING OF ODS-SUPERALLOYS," ASME 91-GT-411, THE INTERNATIONAL GAS TURBINE AND AEROENGINE CONGRESS AND EXPOSITION (ORLANDO, FL), AMERICAN SOCIETY OF MECHANICAL ENGINEERS, JUNE 1991

Composite-to-Metal Joining

Raymond E. Bohlmann, McDonnell Douglas Aerospace-East

Introduction

SELECTING THE BEST METHODS and practices for joining composites to metals is highly dependent on the particular combination of composite and metal materials being joined and the design details and/or requirements of the joint. For example, if a carbon-epoxy laminate must be joined to an all-metal fitting through an adhesive bondline that should perform well in a hot-wet environment, aluminum would be a poor choice. To develop the hot-wet strength, the adhesive must be cured at elevated temperature. Because of the high coefficient of thermal expansion (CTE) of the aluminum and a low CTE for the carbon-epoxy composite, significant fabrication thermal stresses can occur in the adhesive bondline at room temperature after the elevated-temperature cure. These fabrication thermal stresses, which are at a maximum at the ends of the joint, can be sufficient to cause the adhesive to fail before any mechanical loads are applied. Also, the carbon-epoxy and aluminum combination is susceptible to galvanic corrosion. Titanium would be a better material selection for the metal adherend--both for minimizing thermal stresses (CTE of titanium is less than one-half that of the aluminum) and for minimizing corrosion.

This article will discuss material combinations, design details, and fabrication processes that should be considered when joining resin-matrix composites to metals. Resin-matrix composites, also referred to as organic-matrix composites, can be divided into thermoset composites (resin chemical reaction during 120 to 175 °C, or 250 to 350 °F cure) and thermoplastic composites (resin melt fuse between 315 and 400 °C, or 600 and 750 °F). Thermoset composites can be joined to metals by adhesive bonding, whereas thermoplastic composites can be joined by adhesive bonding or through the use of a melt-fuse interface (amorphous bond). Themoplastics can be further divided into semicrystalline and amorphous materials. A semicrystalline thermoplastic material has a low moisture saturation level (0.25%) as compared to an amorphous thermoplastic (0.6%) and a typical thermoset (1.6%). Composite hot-wet properties and bonded-joint shear strengths (adhesive or amorphous bond) are reduced by the degree of absorbed moisture and the joining materials should be selected to minimize this effect.

Composite-to-Metal Joining

Raymond E. Bohlmann, McDonnell Douglas Aerospace-East

Bonded Joints

Design Issues. The analysis of adhesively bonded joints has advanced to the stage where it can legitimately be called a science. Hart-Smith has added nonlinear adhesive behavior to the analysis and design of adhesive bonded joints in the form of an elastic-plastic adhesive model (Ref 1, 2, 3). In addition, a series of digital computer programs has been developed for the analysis of bolted and bonded joints of various geometries (Ref 4, 5). As was stated earlier, selecting the best methods and practices for joining composites to metals depends on the particular combination of materials being joined and the design details/requirements of the joint.

The effect of several important design variables will be illustrated by application of the A4EI STEPLAP analysis computer program (Ref 5). The capabilities of the A4EI program include:

- ELASTIC OR ELASTIC-PLASTIC STRESS DISTRIBUTIONS
- STEPPED-LAP ADHESIVELY BONDED SPLICES OR DOUBLERS
- CHECKS FOR ADHEREND OR ADHESIVE FAILURE
- ACCOUNTS FOR ADHEREND STIFFNESS IMBALANCE, ADHEREND THERMAL MISMATCH, VARIABLE ADHEREND THICKNESS AND STIFFNESS, NONUNIFORM BONDLINE THICKNESS AND ADHESIVE PROPERTIES, AND TENSION, COMPRESSION, AND SHEAR LOADING

The STEPLAP computer procedure is based on a closed-form analysis, which, through an iterative scheme, provides the complete internal stress distribution of a stepped-lap joint (Fig. 1). Two program-detectable failure modes are adherend axial failure and adhesive shear failure. The elastic-plastic methodology represents a major improvement in stepped-lap joint analysis because it accounts for all significant variables in one procedure. In particular, the analysis accounts for adherend stiffness imbalance, adherend thermal mismatch, step thicknesses, number of steps, and adhesive ductility (in terms of an elastic-plastic model). This analytic capability allows the user to quickly identify the critical or limiting detail of a stepped-lap joint configuration and make changes in the joint design to maximize the stepped-lap joint strength. To apply these analyses procedures to a design requires the availability of the adhesive shear stress-strain data (Fig. 2), which are obtained by conducting a thick adherend test on the adhesive (Fig. 3). For design-limit loads (operating loads), the recommended procedure is to limit the bondline shear strain to the elastic knee of the shear stress-strain curve to prevent adhesive failure during repeated fatigue loading.

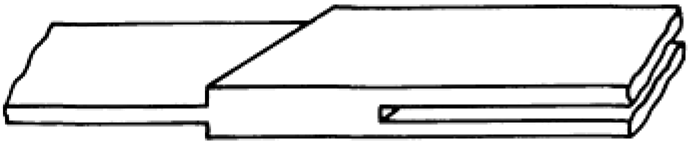

Double lap

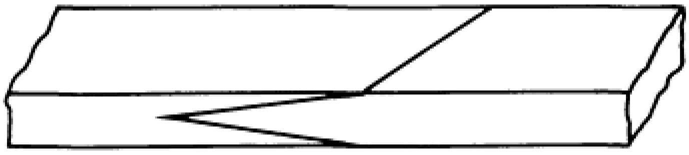

Double scarf

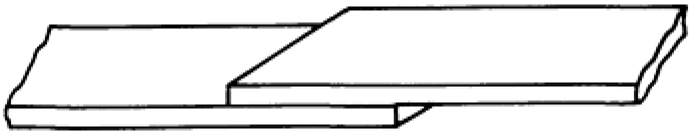

Single lap

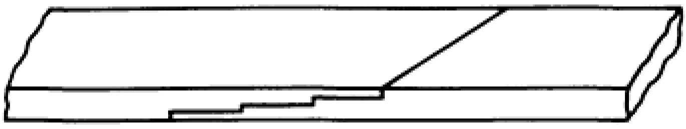

Stepped lap

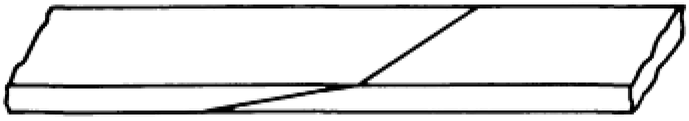

Single scarf

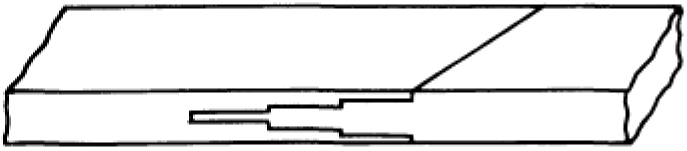

Double-stepped lap

FIG. 1 JOINT DESIGNS USED FOR ADHESIVE BONDING

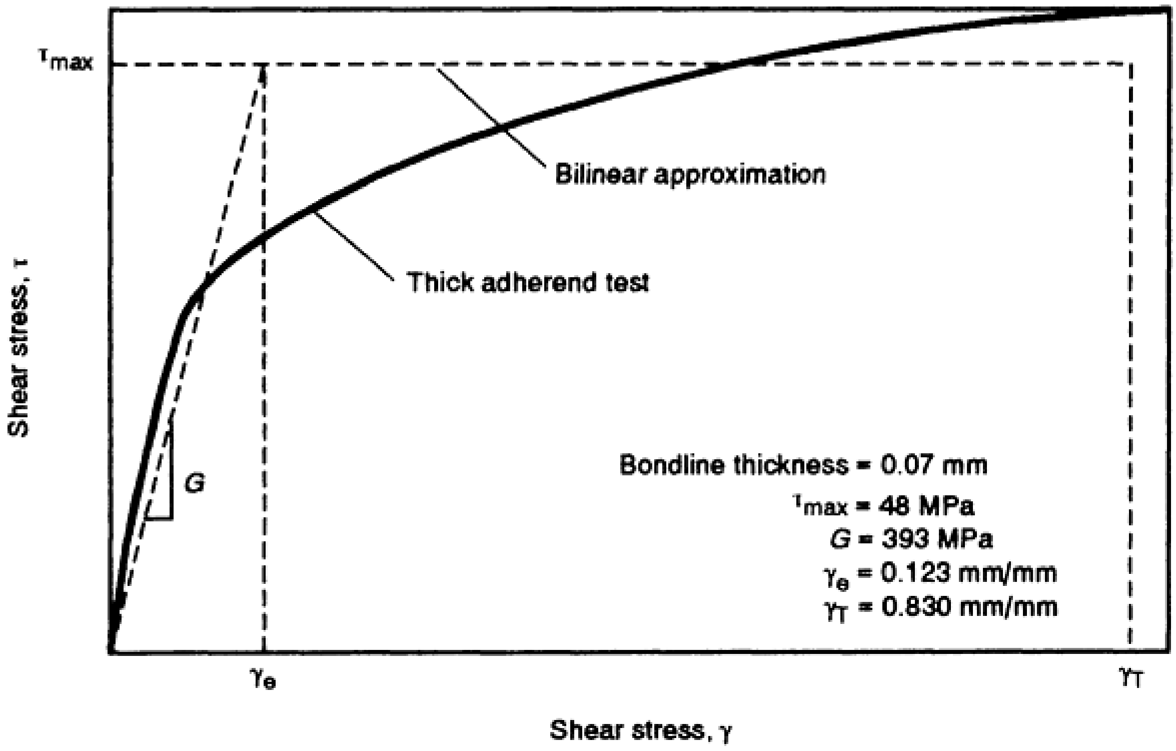

FIG. 2 STRESS-STRAIN CURVE DEVELOPED FROM A THICK ADHEREND (ROOM-TEMPERATURE/DRY) TEST USING FM300K ADHESIVE. *G,* SHEAR MODULUS; γ_E, ELASTIC STRAIN; γ_T, TOTAL STRAIN

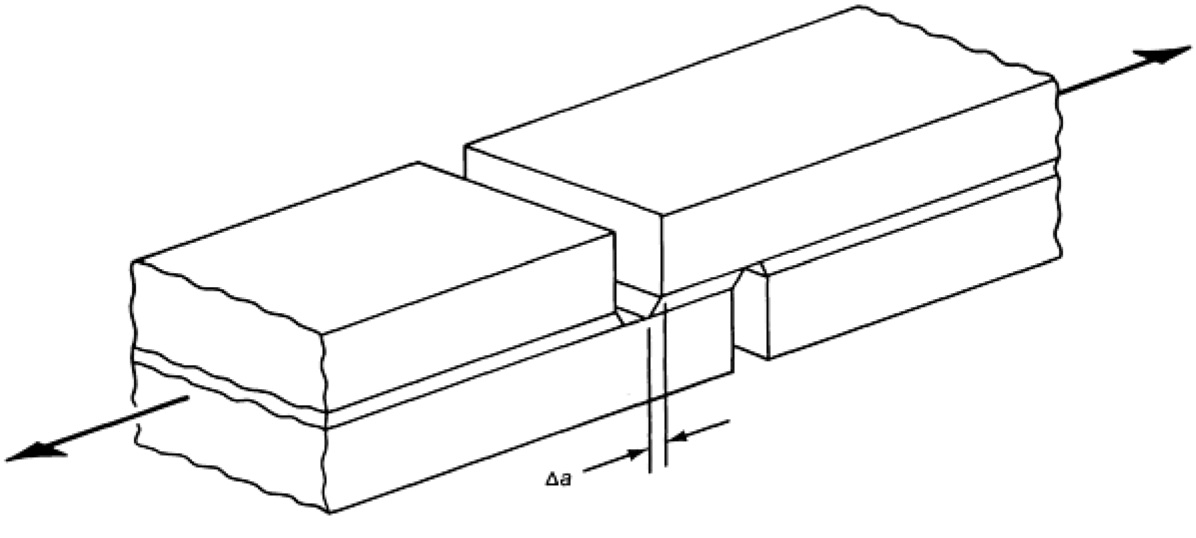

FIG. 3 THICK ADHEREND LAP-SHEAR SPECIMEN

The effect of adhesive bondline thickness on joint load capability is shown in Fig. 4 for a constant thickness composite-to-metal splice joint. The shear stress distributions along the bond length for two different bondline thicknesses were predicted by the computer program. The joint load capability is equivalent to the area under the curve. The curve for a bondline thickness of 0.13 mm (0.005 in.) has 60% more area under the curve and therefore can carry 60% more load than a 0.05 mm (0.002 in.) thickness. For a joint with an elastic distribution, the joint load-carrying capability increase is proportional to the square root of the bondline thickness increase. In this example, the joint load increased by (0.005 in./0.002 in.)$^{0.5}$, which is a 60% increase.

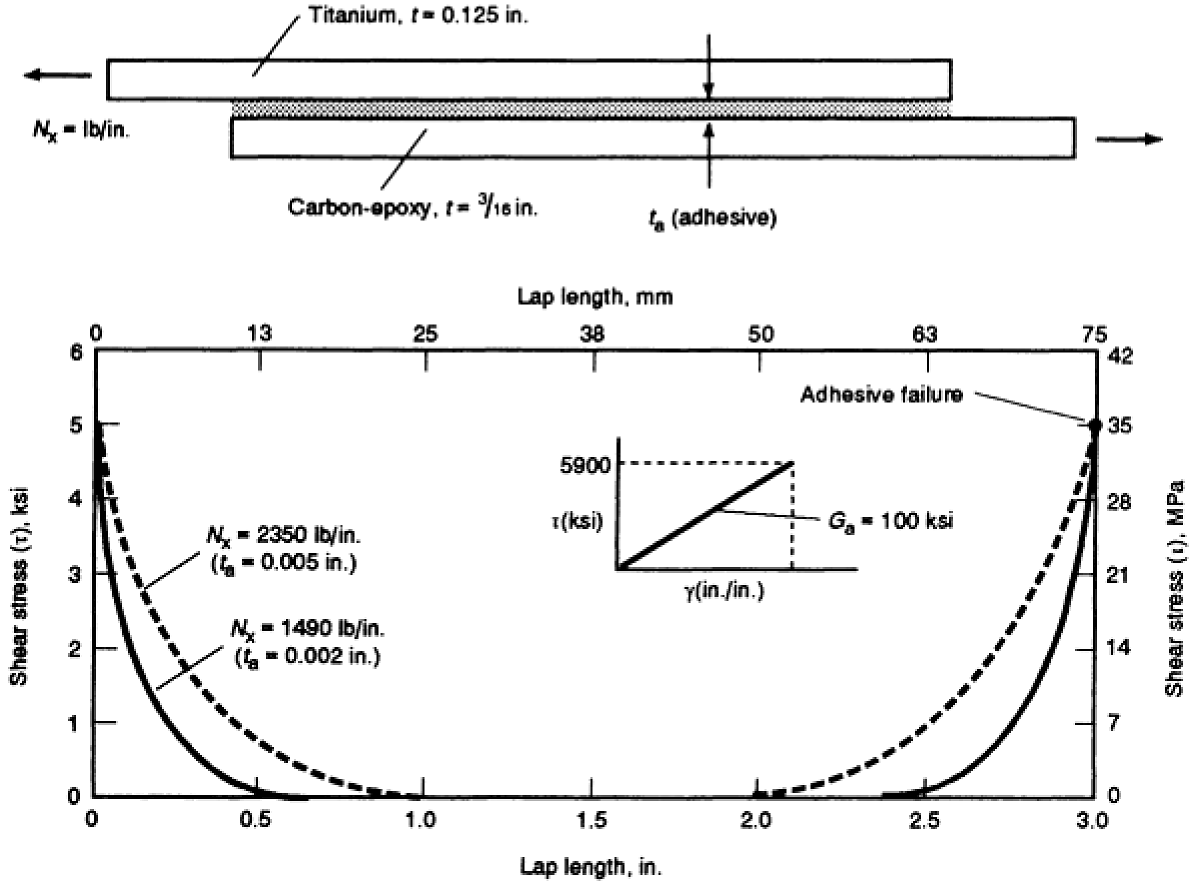

FIG. 4 THE EFFECT OF BONDLINE THICKNESS ON JOINT LOAD-CARRYING CAPABILITY. N_X, JOINT LOAD CAPABILITY; G_A, ADHESIVE SHEAR MODULUS; T_A, ADHESIVE THICKNESS; γ^{SU}, ADHESIVE PLASTIC SHEAR STRAIN (SU, SHEAR ULTIMATE). ORIGINAL SPECIMEN DIMENSIONS GIVEN IN INCHES

The benefits of tapering the adherend thickness are shown in Fig. 5. In the composite (carbon-epoxy)-to-titanium splice joint, two cases are compared using an elastic analysis. In the first case, both adherends are a constant thickness; in the second case, the titanium adherend is tapered as if the titanium is a repair patch that could be tapered prior to bonding. For the first case (adherend constant thickness), the shear stress in the adhesive is maximum on the right end of the joint. For the second case, tapering the thickness of the titanium patch on the right end of the joint reduces the maximum shear stresses in the adhesive, forcing the failure in the adhesive at the left end of the joint and increasing the joint load-carrying capability by 16%. The joint load capability can be further increased by tapering the carbon-epoxy laminate at the left end of the joint, which will further reduce the maximum shear stress in the adhesive. An additional benefit of tapering the ends is to reduce the peel stresses in the adhesive and the interlaminar tension in the laminate.

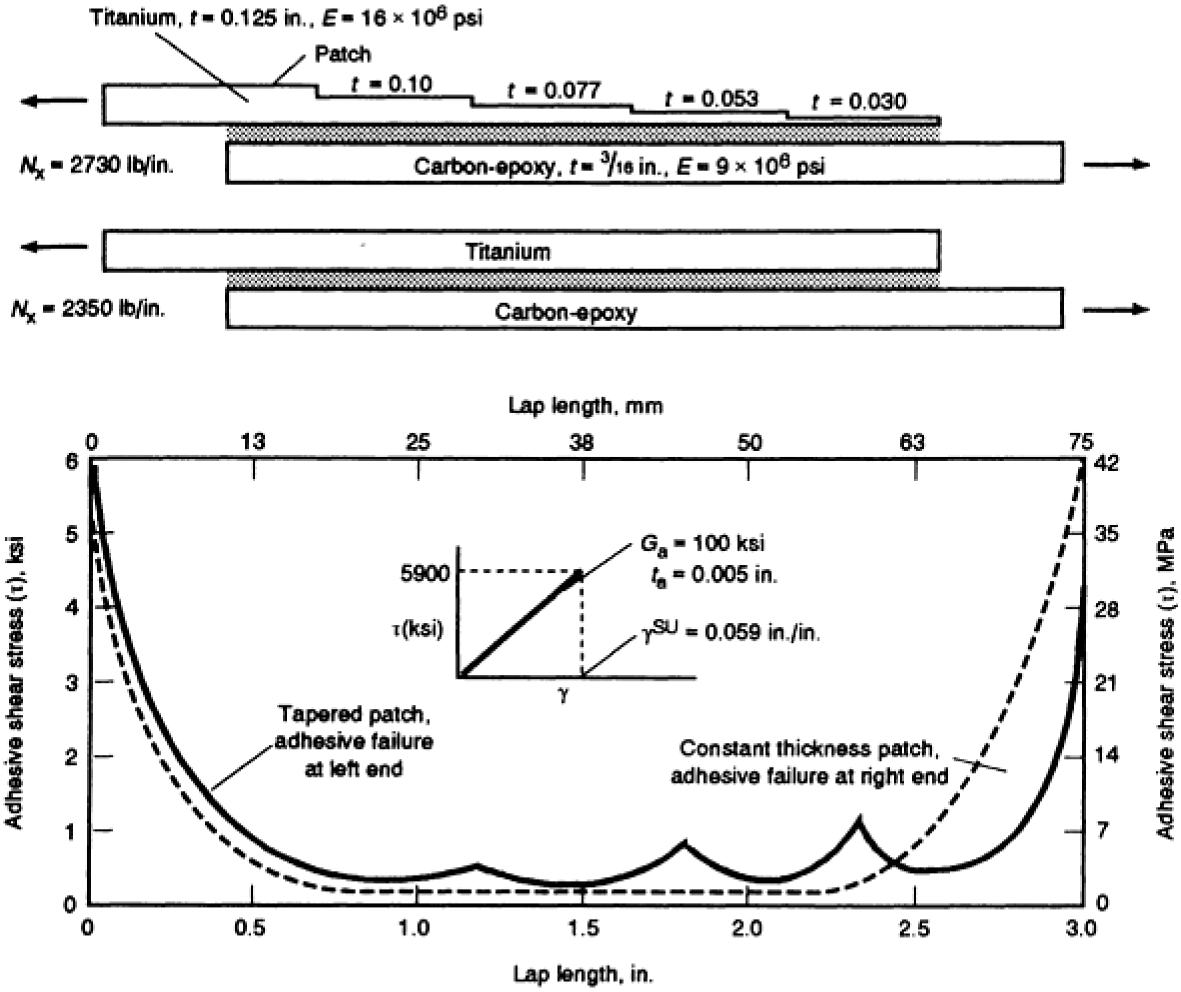

FIG. 5 THE EFFECT OF TAPERING THE ADHEREND THICKNESS ON JOINT LOAD-CARRYING CAPABILITY. SEE FIG. 4 FOR EXPLANATION OF SYMBOLS. ORIGINAL SPECIMEN DIMENSIONS GIVEN IN INCHES

Taking advantage of the adhesive plasticity will provide a significant increase in the area under the shear stress versus lap length curve and the resultant joint load capability (Fig. 6). The example shown in this figure is the same as the tapered-thickness titanium patch shown in Fig. 5 with the adhesive plasticity condition added. For the elastic case, the joint load capability is 49.5 kg/mm (2730 lb/in.). Increasing the adhesive shearing strain allowable from the elastic value of 0.059 mm/mm (in./in.) to the plastic value of 0.20 mm/mm (in./in.) increases the joint load to 118 kg/mm (6520 lb/in.) (a factor of 2.4). Increasing the adhesive plastic shear strain to 0.30 mm/mm (in./in.) further increases the joint load capability to 148 kg/mm (8140 lb/in.), which is a three-fold increase over the elastic condition. Adhesive plasticity will generally provide the added load-carrying capability needed between the operating loads (design limit) and the design ultimate loads when the operating loads (fatigue loading) are limited to the elastic knee in the adhesive shear stress-strain curve.

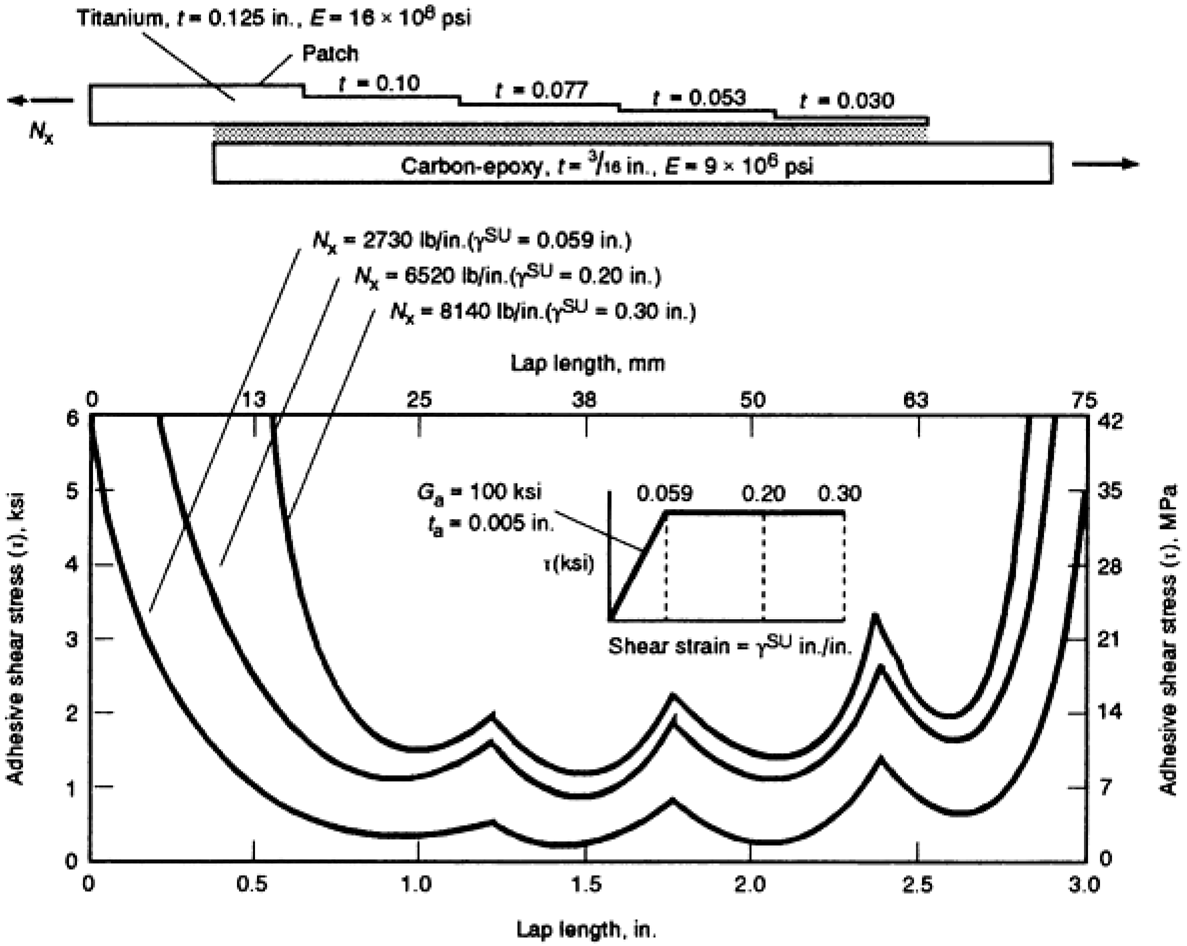

FIG. 6 THE EFFECT OF ADHESIVE PLASTICITY ON JOINT LOAD-CARRYING CAPABILITY. SEE FIG. 4 FOR EXPLANATION OF SYMBOLS. ORIGINAL SPECIMEN DIMENSIONS GIVEN IN INCHES

Thermoset Composites. (Carbon-epoxy)-to-titanium adhesively bonded splice joints are being utilized in aircraft primary structures. Joining carbon-epoxy to titanium using adhesive bonding has provided substantial weight savings. For these co-cured bonded joints, adhesive comparison testing led to the selection of FM300K, which is a film adhesive that contains an open-knit polyester carrier cloth to control bondline thickness. Figure 7 is a plot of the STEPLAP program output for typical joint analysis at room temperature. Joint configuration, ultimate loading, and adhesive strain are included. The adhesive strain plot illustrates the elastic trough on each step and the spike at each step end. Shear strains at the joint ends exceed the elastic strain of the bilinear FM300K stress-strain design curve (Fig. 4), and the shear stress plateaus at 39 MPa (5700 psi). For the room-temperature analysis, only a small length of the end steps is strained in the plastic range.

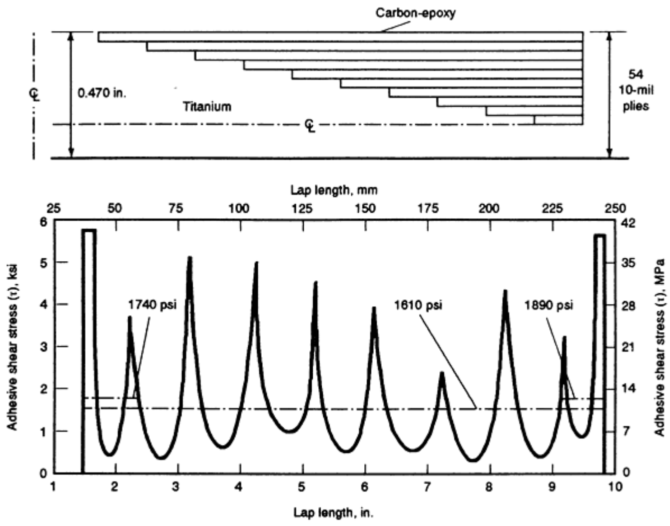

FIG. 7 ADHESIVE SHEAR STRESS DISTRIBUTION FOR AN ADHESIVELY BONDED STEPPED-LAP JOINT GENERATED USING THE STEPLAP COMPUTER PROGRAM

Thermoplastic Composites. AS4/APC-2 is one of the most frequently used and well characterized thermoplastic materials available. AS4/APC-2 is a semicrystalline material with excellent resistance to fluids (except fuel) and solvents and has low moisture absorption (AS4/APC-2 absorbs 0.25% moisture at 95% relative humidity). The two current methods used for bonding thermoplastics to metals or other composites are amorphous bonding and adhesive bonding. The primary joining technique used for thermoplastic semicrystalline adherends (such as AS4/APC-2) is to melt-fuse an amorphous thermoplastic film (Ultem) at the interface. This joining process (amorphous bonding) is accomplished by first adding a lower melt-fuse temperature resin onto the surface of the adherend to be subsequently joined (Fig. 8). Since the melting point of the amorphous film on the adherend surface and at the interface is about 300 °C (570 °F), the bonding does not affect the integrity of the AS4/APC-2 laminates, which are processed at 380 °C (720 °F). Amorphous bonded joints are structurally efficient and can produce room-temperature shear strengths in the 28 to 34 MPa (4000 to 5000 psi) range (Fig. 9). One limitation of amorphous joining is that the amorphous thermoplastic resin absorbs more water (1.3%) than semicrystalline thermoplastic composite laminates and the service temperature must be kept below 200 °C (390 °F) to prevent steam-pressure delamination damage of the joint (Fig. 10). A detailed explanation of the steam-pressure delamination phenomenon is described in Ref 6.

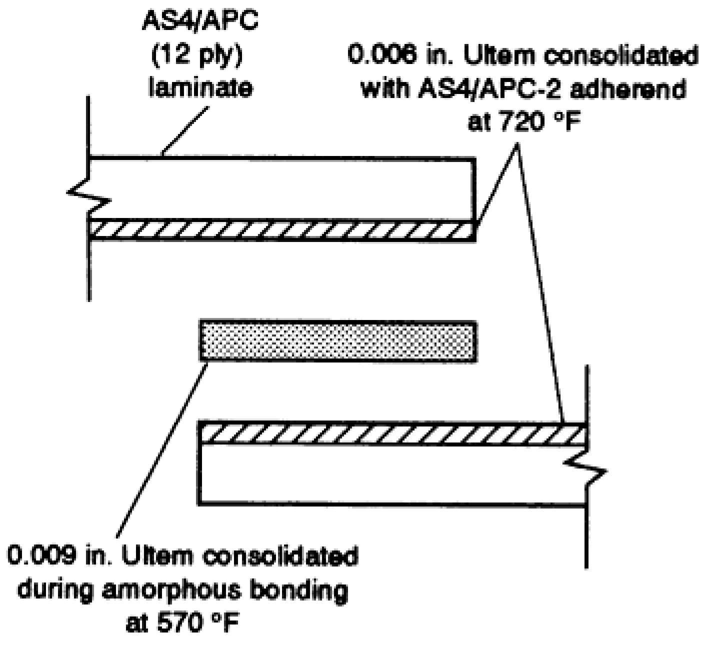

FIG. 8 SCHEMATIC OF THE AMORPHOUS BONDING PROCESS. ORIGINAL SPECIMEN DIMENSIONS GIVEN IN INCHES

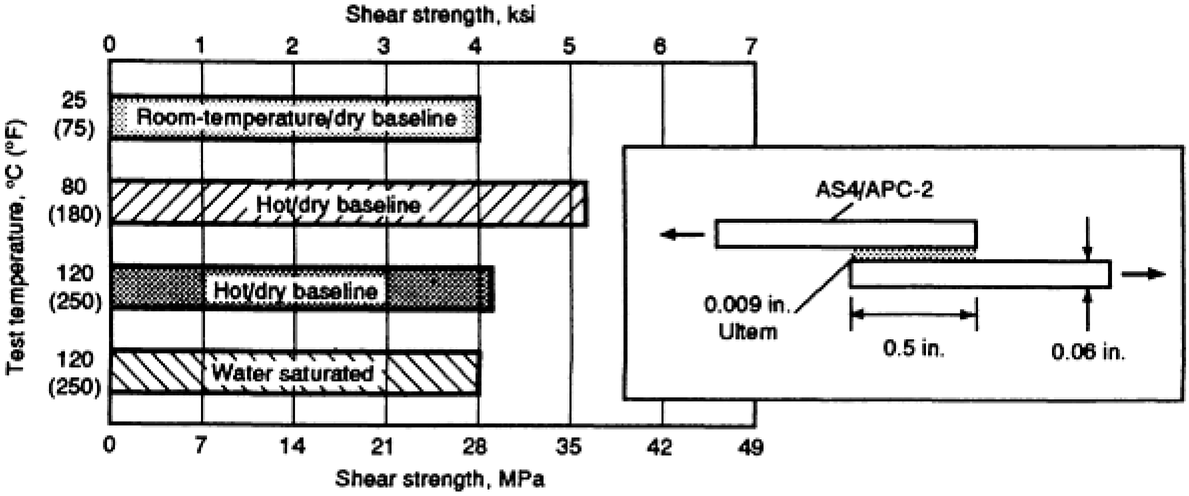

FIG. 9 SHEAR STRENGTHS OF SINGLE-LAP JOINTS PRODUCED BY AMORPHOUS BONDING

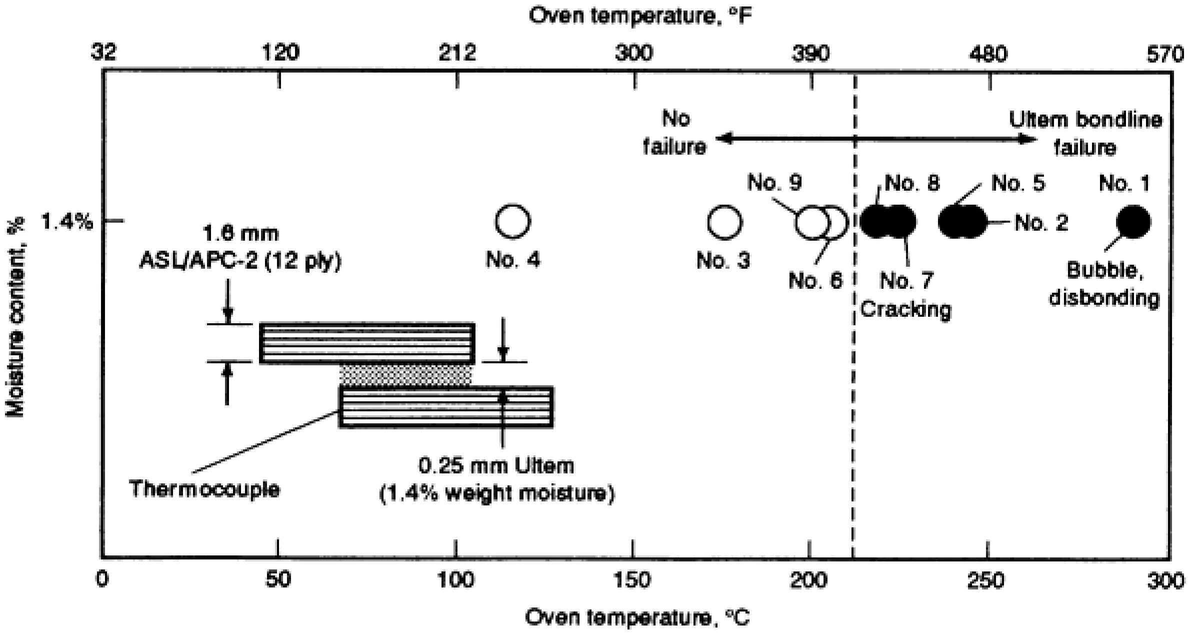

FIG. 10 THERMAL SHOCK (STEAM DELAMINATION) PROPERTIES FOR MOISTURE-SATURATED THERMOPLASTIC ADHERENDS JOINED BY AMORPHOUS BONDING WITH ULTEM. EACH CIRCLE REPRESENTS TESTING RESULTS OF ONE SPECIMEN. NUMBER INDICATES TESTING SEQUENCE.

Amorphous bonding of thermoplastics that requires an autoclave process (290 °C at ~700 kPa, or 550 °F at ~100 psi) is difficult to accomplish in a field-repair environment where minimal equipment is available. Therefore, adhesive bonding is a viable approach for attaching a metal or composite repair patch because the adhesive bond can be cured with a heating blanket (up to 175 °C, or 350 °F) and a vacuum bag. Epoxy adhesive systems do not bond as well to semicrystalline thermoplastics as to amorphous thermoplastics or thermoset composites. A test program was conducted to evaluate the benefit of adhesive bonding of an AS4/APC-2 thermoplastic with or without a layer of Ultem film. Two different adhesives were included in the evaluation --FM300 and FM73. The room-temperature dry test results indicated that the optimum combination was with the co-consolidated Ultem film and the FM73 adhesive. This combination produced the highest lap shear strength (34 MPa, or 5000 psi), minimum test scatter, and a desirable cohesive failure mode. The quasi-isotropic AS4/APC-2 adherends used in the double-lap shear specimens (Fig. 1) were 2 mm (0.08 in.) thick. All adherends were surface treated by grit blasting followed by an isopropyl alcohol wipe and were bonded under vacuum bag pressure. The FM300 adhesive was processed at 150 °C (300 °F) for 4 h, and the FM73 adhesive was processed at 120 °C (250 °F) for 2 h. Prior to testing, all specimens were heated in a 95 °C (200 °F) oven until dry, stabilized weights were achieved.

Since previous shear strength tests of amorphous bonded thermoplastic joints showed that Ultem was sensitive to absorbed moisture, double-lap shear tests were conducted with and without Ultem for the FM73 adhesive at 25 °C (77 °F) and 60 °C (140 °F) in the dry condition and in the wet condition (95% relative humidity) at 60 °C (140 °F). The results summarized in Fig. 11 show that the moisture reduction factor is 28% at room temperature and 51% at 60 °C (140 °F) for the specimens with Ultem.

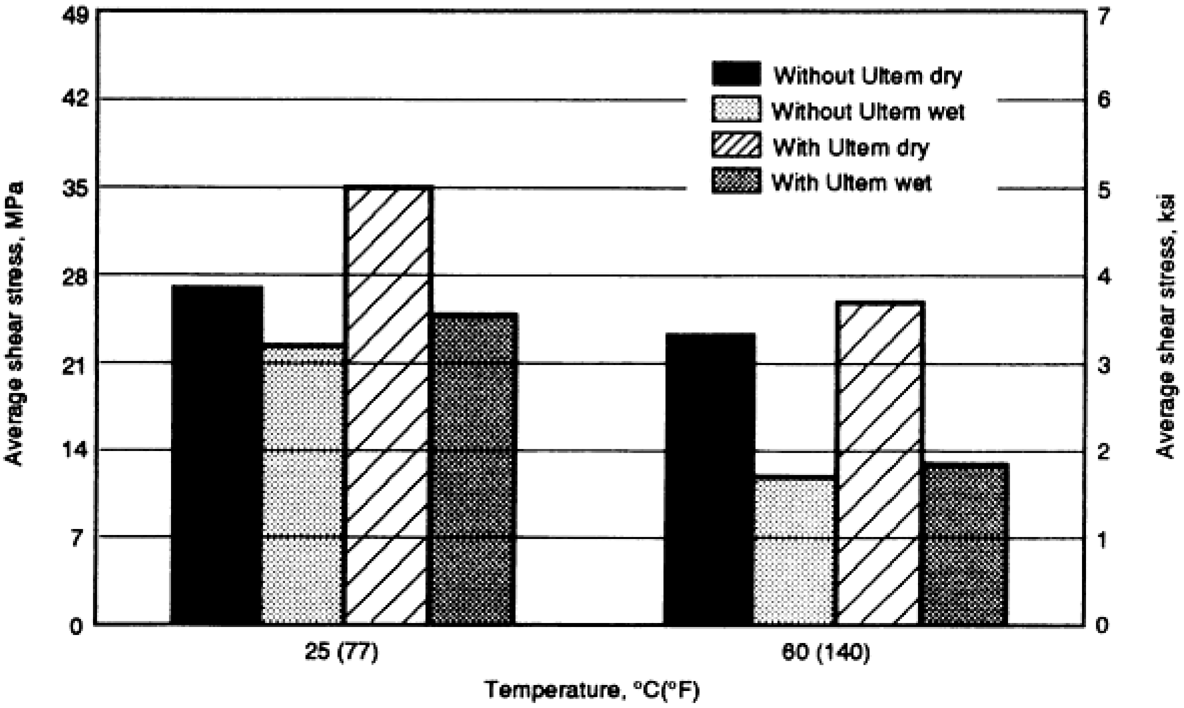

FIG. 11 EFFECTS OF MOISTURE CONTENT ON THE BOND STRENGTH OF AS4/APC-2 ADHERENDS JOINED BY AMORPHOUS BONDING WITH ULTEM

References cited in this section

1. L.J. HART-SMITH, "ANALYSIS AND DESIGN OF ADVANCED COMPOSITE BONDED JOINTS," DOUGLAS AIRCRAFT COMPANY, NASA CR-2218, JAN 1973, REPRINTED AUG 1974
2. L.J. HART-SMITH, DESIGN AND ANALYSIS OF ADHESIVE-BONDED JOINTS, PROC. *1ST AIR FORCE CONF. FIBROUS COMPOSITES IN FLIGHT VEHICLE DESIGN,* AFFDL-TR-72-130, AIR FORCE FLIGHT DYNAMICS LABORATORY, 1972, P 813-856
3. L.J. HART-SMITH, ADVANCES IN THE ANALYSIS AND DESIGN OF ADHESIVE-BONDED JOINTS IN COMPOSITE AEROSPACE STRUCTURES, *14TH NATIONAL SAMPE SYMP. AND EXHIBITION,* SOCIETY FOR THE ADVANCEMENT OF MATERIAL AND PROCESS ENGINEERING, APRIL 1974, P 722-737
4. L.J. HART-SMITH, BONDED-BOLTED COMPOSITE JOINTS, *J. AIRCRAFT,* VOL 22, 1985, P 993-1000
5. L.J. HART-SMITH, ADHESIVELY BONDED JOINTS FOR FIBROUS COMPOSITE STRUCTURES, *JOINING FIBRE-REINFORCED PLASTICS,* F.L. MATTHEWS, ED., ELSEVIER, 1987, P 271-311
6. H.J. KIM AND R.E. BOHLMANN, "THERMAL SHOCK TESTING OF WET THERMOPLASTIC

LAMINATES," *37TH INT. SAMPE SYMP. EXHIBITION,* MARCH 1992

Composite-to-Metal Joining

Raymond E. Bohlmann, McDonnell Douglas Aerospace-East

Bolted Joints

A composite-to-metal bolted joint was studied as an alternative to the high-load transfer bonded stepped-lap joint discussed earlier (Ref 7). The titanium seven-fastener design shown in Fig. 12 is 380 mm (15 in.) in length (about twice the length of the bonded design) and is 30% heavier than the baseline adhesively bonded joint design. A fastener load-distribution curve is shown in Fig. 13. The test specimen was designed for a zero margin at 150% of design limit load (DLL); failure occurred at 173% DLL. The specimen failed as predicted at the end fastener hole (fastener No. 1) in the high bypass strain area of the carbon-epoxy laminate.

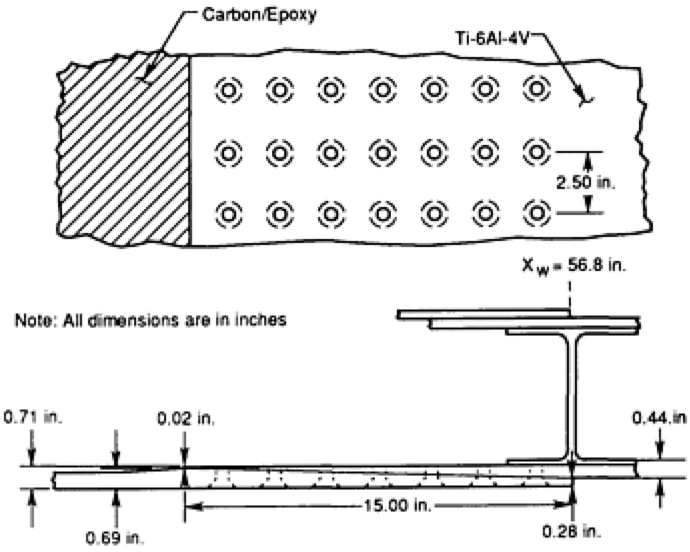

Note: All dimensions are in inches

FIG. 12 COMPOSITE-TO-METAL BOLTED JOINT CONCEPT. ORIGINAL SPECIMEN DIMENSIONS GIVEN IN INCHES

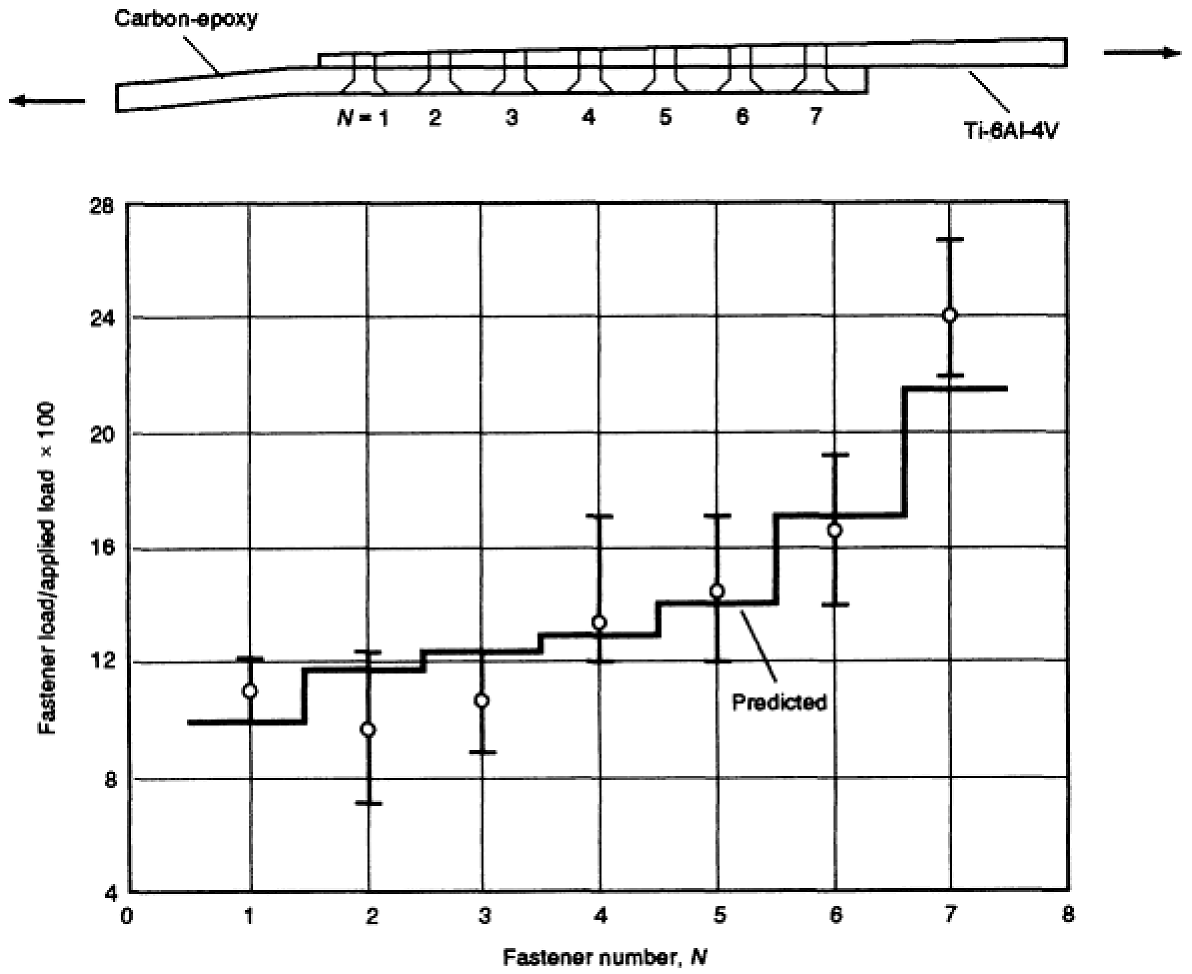

FIG. 13 CORRELATION BETWEEN PREDICTED RESPONSE AND ACTUAL TEST VALUES FOR BOLTED TITANIUM-TO-COMPOSITE (CARBON-EPOXY) JOINTS

Reference cited in this section

7. D.L. BUCHANAN AND S.P. GARBO, "DESIGN OF HIGHLY LOADED COMPOSITE JOINTS AND ATTACHMENTS FOR WING STRUCTURES," REPORT NADC-81194-80, NAVAL AIR DEVELOPMENT CENTER, AUGUST 1981

Composite-to-Metal Joining

Raymond E. Bohlmann, McDonnell Douglas Aerospace-East

Submersible Thick Composite-to-Metal Joint

The optimum composite-to-metal joining method for a thick (>25 mm, or 1.0 in.) composite compression-loaded submersible is a block compression joint (Fig. 14). The type of loading for a submersible structure is primarily hydrostatic pressure, which results in biaxial compression loading of 2:1 for cylinders and equal biaxial loading for spheres. Bolted joints are not recommended for submersible structures because they provide a path for a potential water leak. Submersibles made from composites for significant depth operation require thick composite walls that cannot be readily joined by adhesive bonding. Bonded joints are best suited for joining thin structures (<2.5 mm, or 0.10 in.) where peeling stresses are acceptable. Resin-matrix composites have good in-plane properties (fibers are in-plane) but low interlaminar properties and are weak in resisting peel-type loads.

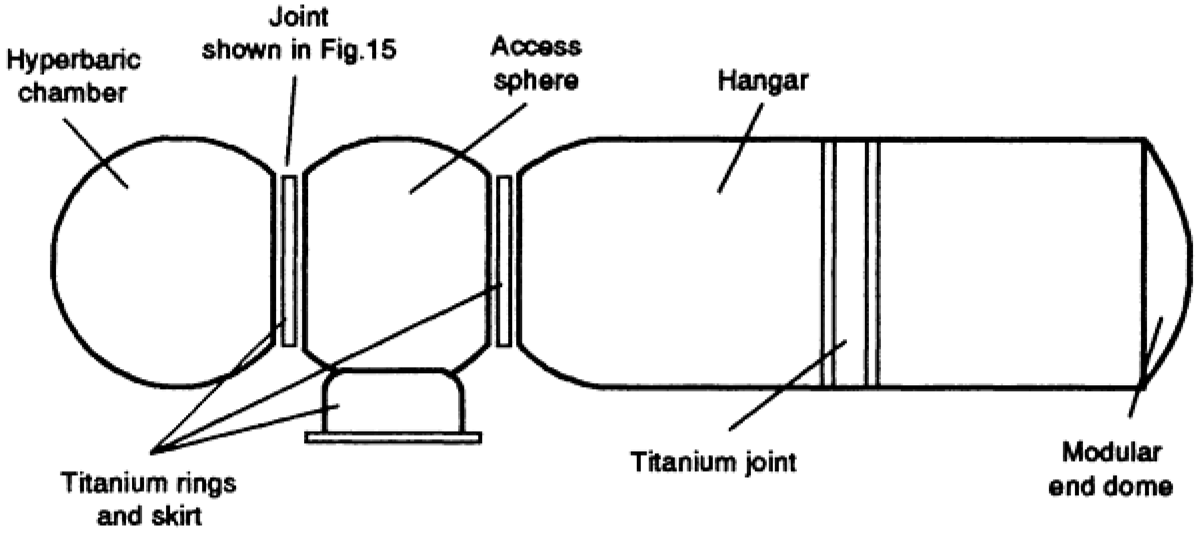
Hyperbaric chamber
Joint shown in Fig.15
Access sphere
Hangar
Titanium rings and skirt
Titanium joint
Modular end dome

FIG. 14 SCHEMATIC OF THE MAN-RATED DEMONSTRATION ARTICLE (MRDA). SEE ALSO FIG. 15.

A Defense Advanced Research Projects Agency (DARPA) Man-Rated Demonstration Article (MRDA) is being developed (Ref 8). The MRDA is composed of a spherical hyperbaric chamber, a spherical access chamber, a cylindrical hangar section, and a modular end dome (Fig. 14). These components are joined to form a single structure; the joint assemblies used to join the composite spherical elements are made from titanium. The stress state in the regions where the titanium and composite material are in contact are very complex. Analysis of the joint region of the composite spheres indicates that small changes in manufacturing tolerances create significant strain concentrations in the composite spherical hyperbaric chamber. In this region, strains have been shown to be particularly high in the sphere at the point where it contacts the titanium joint. The joint region of concern is shown in Fig. 15.

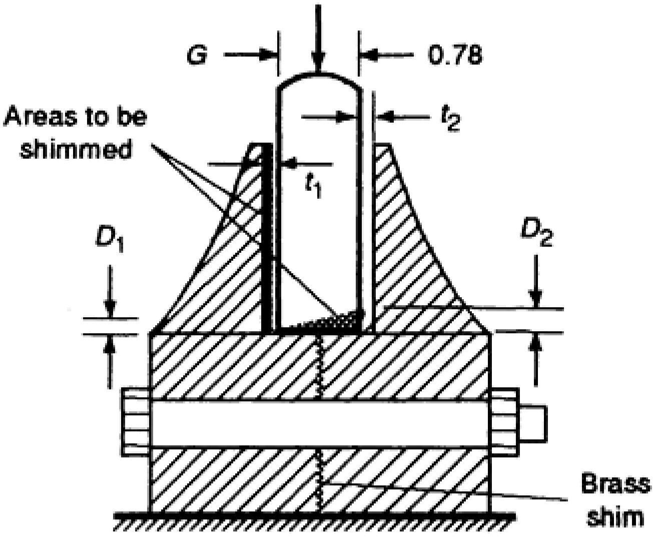
G
0.78
Areas to be shimmed
t_2
t_1
D_1
D_2
Brass shim

FIG. 15 CROSS-SECTIONAL VIEW OF THE MODULAR COMPOSITE SPHERE JOINT IN THE MRDA SHOWN IN FIG. 14

The tolerance variables considered in this investigation were the thickness of the fiberglass tabs and the perpendicularity between the composite sphere tabs and the composite sphere ends, as shown in the figure accompanying Table 1. Since the fiberglass tabs were to serve as sacrificial material to be machined for fit-up between the sphere and joint, two possible machining conditions were considered. The first condition was to have equal amounts of fiberglass (1.5 mm, or 0.060 in.) on each side of the composite sphere. The second condition was to have all machining performed on the inside of the sphere and resulting in full tab thickness on the outside of the sphere (~3 mm, or 0.12 in.) and essentially no tab material on the inside.

TABLE 1 TEST COUPON TOLERANCES USED TO STUDY COMPOSITE-TO-METAL JOINTS IN THE MRDA SHOWN IN FIG. 14

Original dimensions given in inches (1 in. = 25 mm)

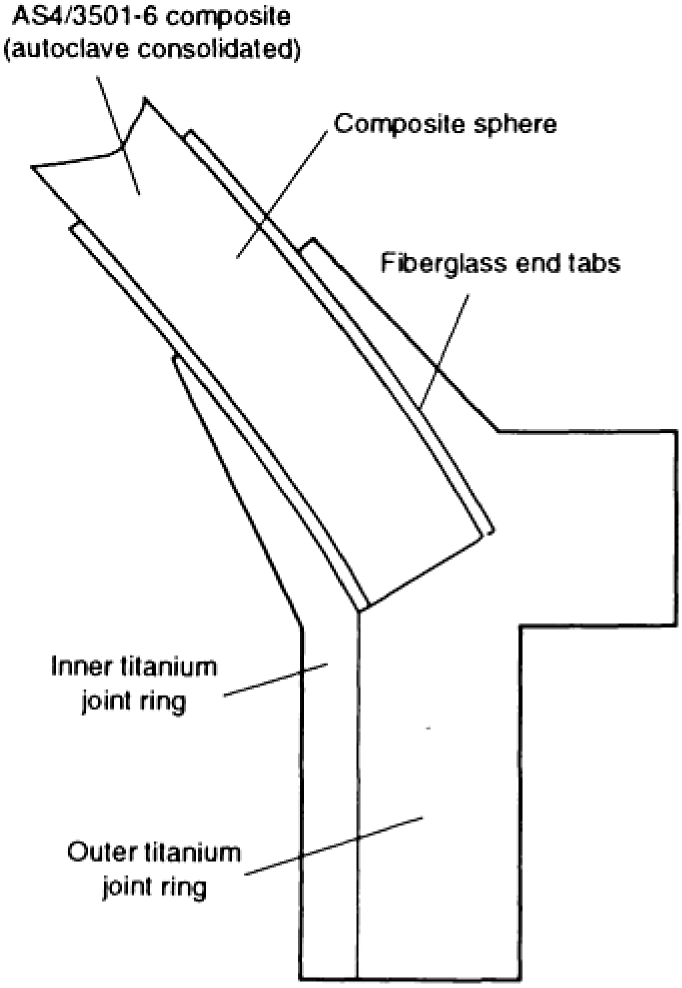

TEST CASE	EVALUATION OBJECTIVE	LATERAL GAP (*G*), IN.	LEFT-SIDE GAP (D_1), IN.	RIGHT-SIDE GAP (D_2), IN.	END TAB THICKNESS, IN.		END SHIM
					T_1	T_2	
1A,B	BASELINE, NOMINAL CASE	0	0	0	0.06	0.06	**NO**
2	MAXIMUM END TAB THICKNESS	0	0	0	0.02	0.12	**NO**
3	EFFECT OF END TAPER	0	0	0.020	0.06	0.06	**NO**
4	EFFECT OF SEVERE END TAPER	0	0	0.040	0.06	0.06	**NO**
5	LATERAL GAP/LIQUID SHIM	0.02	0	0	0.06	0.06	**NO**
6	LATERAL GAP/TAPER LIQUID SHIM	0.02	0	0.020	0.06	0.06	**NO**
7	END TAPER AND LIQUID SHIM	0	0.06	0.080	0.06	0.06	**YES**
8	END TAPER AND LIQUID SHIM	0	0	0.020	0.06	0.06	**YES**
9	**BASELINE, TEST-TO-FAILURE**	**0**	**0**	**0**	**0.06**	**0.06**	**NO**

For the second tolerance condition, out-of-perpendicularity between the sphere end and the joint, which may result from machining or fitup, occurs where the end of the sphere sits in the titanium joint. For the best condition, the entire end of the sphere would contact the joint with no clearance in the fit. Considering the spherical geometry and size of both components, there is the possibility of a clearance in this region rather than intimate contact. To simulate both tolerance conditions, a series of rectilinear coupons to be loaded in uniaxial compression were machined. Table 1 summarizes the test coupon tolerances.

The quasi-isotropic AS4/3501-6 unidirectional tape thermoset compression specimens 20 mm (0.78 in.) thick were autoclave consolidated and then tested. The results of the complete test program are summarized in Ref 8. Only two of the nine specimens tested will be discussed here. These include test case No. 3 (see Table 1), which featured a 0.5 mm (0.02 in.) end taper and test case No. 9, which is a no-gap baseline specimen. The design goal of these tests was to evaluate the assembly up to a gage section strain of 6700 microstrain.

The results of test case No. 3 with the 0.5 mm (0.02 in.) showed significant strain variation at the load introduction end and only minor strain differences in the gage region. After unloading the specimen, no damage was observed. The results of test case No. 9, the baseline with no end taper, showed that the gage section strains were 13,500 microstrain when failure occurred in the gage region. A failed test specimen from test case No. 9 is shown in Fig. 16. An encouraging result of this test is that the 13,500 microstrain is the same level obtained in testing thin-section composites, indicating that thick-section compression properties are not reduced.

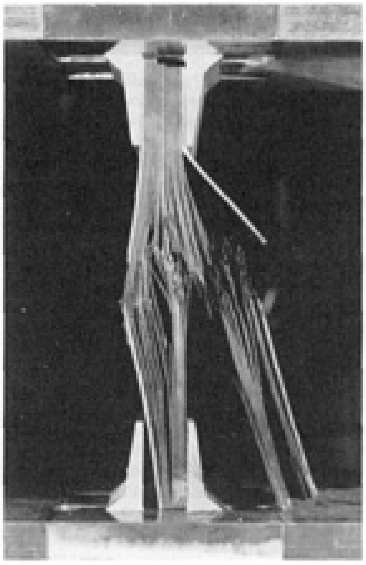

FIG. 16 PHOTOGRAPH OF A FAILED SPECIMEN FROM TEST CASE NO. 9 (SEE TABLE 1)

Reference cited in this section

8. E.T. CAMPONESCHI, R.E. BOHLMANN, J. HALL, AND T.T. CARR, "EFFECT OF ASSEMBLY ANOMALIES ON THE STRAIN RESPONSE OF COMPOSITES IN THE SPHERE JOINT REGION OF THE DARPA MAN RATED DEMONSTRATION ARTICLE," CDNSWC-SME-92/22, DEFENSE ADVANCED RESEARCH PROJECT AGENCY, ARLINGTON, VA, 4 MARCH 1992

Composite-to-Metal Joining

Raymond E. Bohlmann, McDonnell Douglas Aerospace-East

References

1. L.J. HART-SMITH, "ANALYSIS AND DESIGN OF ADVANCED COMPOSITE BONDED JOINTS,"

DOUGLAS AIRCRAFT COMPANY, NASA CR-2218, JAN 1973, REPRINTED AUG 1974

2. L.J. HART-SMITH, DESIGN AND ANALYSIS OF ADHESIVE-BONDED JOINTS, PROC. *1ST AIR FORCE CONF. FIBROUS COMPOSITES IN FLIGHT VEHICLE DESIGN,* AFFDL-TR-72-130, AIR FORCE FLIGHT DYNAMICS LABORATORY, 1972, P 813-856
3. L.J. HART-SMITH, ADVANCES IN THE ANALYSIS AND DESIGN OF ADHESIVE-BONDED JOINTS IN COMPOSITE AEROSPACE STRUCTURES, *14TH NATIONAL SAMPE SYMP. AND EXHIBITION,* SOCIETY FOR THE ADVANCEMENT OF MATERIAL AND PROCESS ENGINEERING, APRIL 1974, P 722-737
4. L.J. HART-SMITH, BONDED-BOLTED COMPOSITE JOINTS, *J. AIRCRAFT,* VOL 22, 1985, P 993-1000
5. L.J. HART-SMITH, ADHESIVELY BONDED JOINTS FOR FIBROUS COMPOSITE STRUCTURES, *JOINING FIBRE-REINFORCED PLASTICS,* F.L. MATTHEWS, ED., ELSEVIER, 1987, P 271-311
6. H.J. KIM AND R.E. BOHLMANN, "THERMAL SHOCK TESTING OF WET THERMOPLASTIC LAMINATES," *37TH INT. SAMPE SYMP. EXHIBITION,* MARCH 1992
7. D.L. BUCHANAN AND S.P. GARBO, "DESIGN OF HIGHLY LOADED COMPOSITE JOINTS AND ATTACHMENTS FOR WING STRUCTURES," REPORT NADC-81194-80, NAVAL AIR DEVELOPMENT CENTER, AUGUST 1981
8. E.T. CAMPONESCHI, R.E. BOHLMANN, J. HALL, AND T.T. CARR, "EFFECT OF ASSEMBLY ANOMALIES ON THE STRAIN RESPONSE OF COMPOSITES IN THE SPHERE JOINT REGION OF THE DARPA MAN RATED DEMONSTRATION ARTICLE," CDNSWC-SME-92/22, DEFENSE ADVANCED RESEARCH PROJECT AGENCY, ARLINGTON, VA, 4 MARCH 1992

Welding of Plastics

Thomas H. North and Geetha Ramarathnam, University of Toronto

Introduction

POLYMERS AND POLYMERIC COMPOSITES are attractive because of their high strength-to-weight ratio, chemical inertness, and ability to be molded into complex shapes at relatively low cost. Polymers can be categorized as thermosets or thermoplastics.

In the case of thermoset resins, a chemical reaction occurs during processing and curing, that is, as a result of irreversible cross-linking reactions in the mold. Both molded thermoset and vulcanized elastomer components cannot be reshaped by means of heating, because degradation occurs. It follows that thermoset and vulcanized robber components can only be joined using adhesive bonding or mechanical fastening methods.

Thermoplastic resins, on the other hand, can be softened, as a result of the weakening of secondary van der Waals or hydrogen bonding forces between adjacent polymer chains. Therefore, thermoplastics can be remolded by the application of heat, and they can be fusion welded successfully.

Thermoplastics can be broadly divided into amorphous and crystalline resins, based on morphology, or structure. These materials are more attractive than thermosets because they:

- ARE CHEAPER
- ARE RECYCLABLE
- CAN BE JOINED USING FUSION WELDING
- HAVE GREATER DAMAGE TOLERANCE
- ARE EASY TO PROCESS
- HAVE GREATER IMPACT-RESISTANCE PROPERTIES

Table 1 lists the major thermoplastic resins. Filled thermoplastics are being increasingly used in semistructural applications. Fillers can reduce material cost, enhance mechanical properties, improve thermal properties, provide flame retardation, and so on.

TABLE 1 MAJOR CATEGORIES OF THERMOPLASTICS AND COMPOSITES

COMMODITY PLASTICS
POLYETHYLENE
POLYPROPYLENE
POLYVINYL CHLORIDE
POLYSTYRENE
ENGINEERING PLASTICS
ACETAL
ACRYLONITRILE-BUTADIENE-STYRENE
POLYMETHYL METHACRYLATE
POLYTETRAFLUOROETHYLENE
NYLON
POLYETHYLENE TEREPHTHALATE
POLYBUTYLENE TEREPHTHALATE
POLYPHENYLENE OXIDE
POLYCARBONATE
ADVANCED ENGINEERING PLASTICS
POLYETHERETHER KETONE
POLYETHERIMIDE
POLYETHER KETONE
POLYPHENYLENE SULFIDE
LIQUID-CRYSTAL POLYMER
POLYSULFONE
REINFORCED AND ADVANCED THERMOPLASTIC COMPOSITES
THERMOPLASTIC RESINS FILLED WITH SHORT, LONG, OR CONTINUOUS FIBERS OF GLASS, CARBON, OR ARAMID

Joining is generally the final step in any fabrication cycle. Some important review articles are represented by Ref 1, 2, 3, 4, 5, and 6. The effectiveness of the joining operation can have a large influence on the application of any polymer or composite material. A variety of polymer joining techniques are available. Figure 1 provides a classification of these different methods (Ref 4).

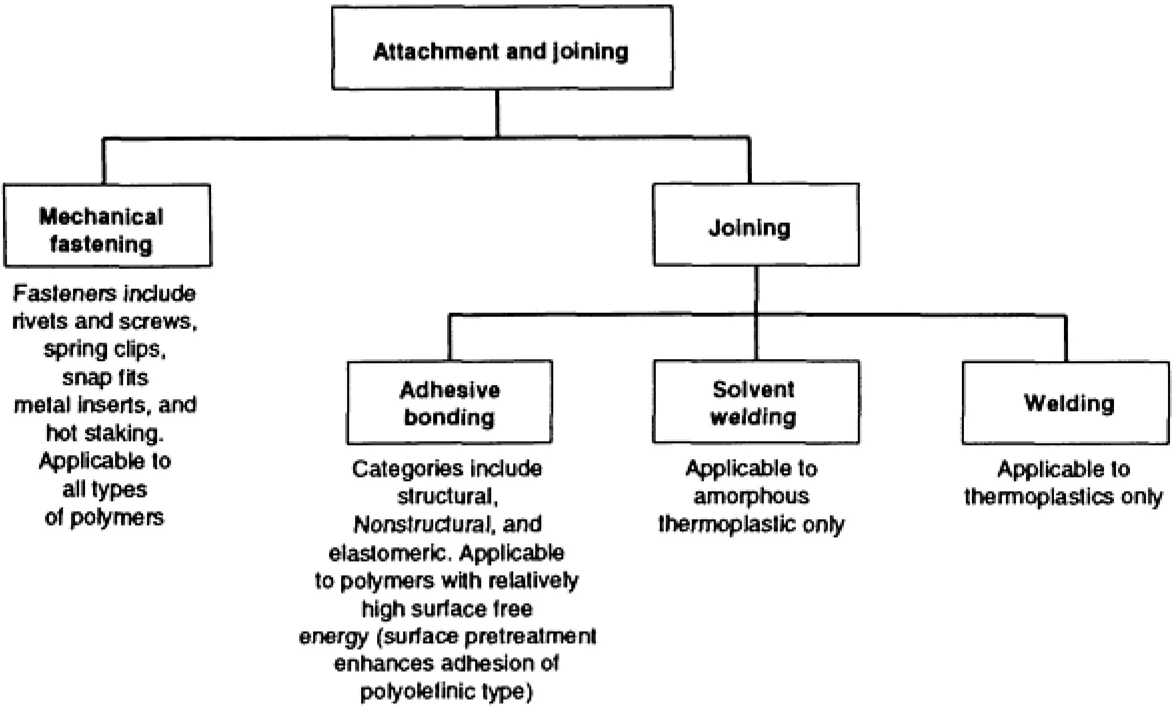

Attachment and joining
Mechanical fastening
Fasteners include rivets and screws, spring clips, snap fits metal inserts, and hot staking. Applicable to all types of polymers
Joining
Adhesive bonding
Categories include structural, Nonstructural, and elastomeric. Applicable to polymers with relatively high surface free energy (surface pretreatment enhances adhesion of polyolefinic type)
Solvent welding
Applicable to amorphous thermoplastic only
Welding
Applicable to thermoplastics only

FIG. 1 CLASSIFICATION OF DIFFERENT JOINING METHODS. SOURCE: REF 4

Polymers are low-energy substrates, with surface free energies of less than approximately 50 mJ/m^2 (0.003 ft · lbf/ft^2) (Ref 7). The creation of a successful joint depends on four factors: the chemical nature of the polymer, the surface free energy, the surface topography, and contamination of the polymer surface by dust, oil, and grease. These factors markedly affect the effectiveness of the adhesive and solvent bonding methods. Fusion welding, however, is much more tolerant of aspects such as surface contamination and material variations from sample to sample.

Adhesives are thermoset-type polymers that are classified as structural, nonstructural, or elastomeric. Reference 7 provides additional details. Structural adhesives are generally used in load-bearing applications, in joints that have high strength-to-weight ratios. They are also used to improve resistance to both component fatigue and corrosion resistance. The principal disadvantages of adhesive bonding are that surface preparation is required prior to bonding, curing times can be long, the components cannot be disassembled following the joining operation, and health and/or safety hazards may be involved in their use. In spite of these problems, adhesive bonding is used extensively in numerous industries.

Mechanical fastening and adhesive bonding can be used to join both similar and dissimilar materials. For example, mechanical fastening is commonly used when joining a plastic to a metal (Ref 8), producing either permanent joints or connections that can be opened and sealed again. The advantages of this approach are that no surface treatment is required and disassembly of the components for inspection and repair is straightforward. The main limitations of this approach are increased weight, the presence of large stress concentrations around the fastener holes, and subsequent in-service corrosion problems (Ref 8). The typical applications of mechanical fastening are in the aerospace, automotive, and construction industry.

Polymeric materials that possess similar solubility parameters can be joined using solvent or fusion welding. Interdiffusion of polymer chains plays a major role in achieving intrinsic adhesion (Ref 9) and in promoting chain diffusion, either by applying a suitable solvent or by heating the polymer sample.

As mentioned previously, only thermoplastics can be joined using the fusion-welding process. The glass transition temperature, T_g, in amorphous polymers, and the melting temperature, T_m, in crystalline polymers must be exceeded so that the polymer chains can acquire sufficient mobility to interdiffuse. A variety of methods exist for welding thermoplastics and thermoplastic composites (Fig. 2). Thermal energy can be delivered externally via conduction, convection, and/or radiation methods, or internally via molecular friction caused by mechanical motion at the joint interface. In the case of external heating, the heat source is removed prior to the application of pressure, and longer welding times are balanced by the greater tolerance to variations in material characteristics. Internal heating methods depend markedly on the material properties (Ref 10). Heating and pressure are applied simultaneously, and shorter welding times are generally involved during the joining process.

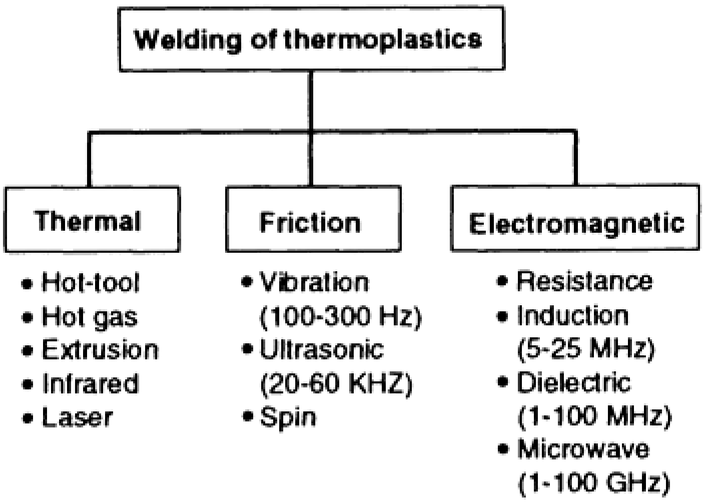

FIG. 2 CLASSIFICATION OF DIFFERENT WELDING METHODS FOR THERMOPLASTICS

Welding is accomplished in many stages. During the initial stage, the polymer-chain molecules become mobile and surface rearrangement occurs. This is followed by wetting and the diffusion of polymer chains across the interface. The final stage involves cooling and solidification. For linear random-coil chains, the mechanical energy required to separate the welded substrates, G, is given by the relation (Ref 11):

$$G = W(T,T,P,M) \qquad \textbf{(EQ 1)}$$

where W is the welding function to be determined, t is the time of contact, T is the temperature of welding, P is the pressure, and M is the molecular weight of the polymer.

In solvent welding, the application of a solvent at the bond line induces sufficient mobility for the polymer chains to interdiffuse (Ref 12, 13). Because the solvent must strongly plasticize the polymer surface, this joining technique is primarily applied to glassy amorphous thermoplastics, such as polycarbonate, acrylic, and polystyrene resins.

References

1. G. GEHARDSSON, THE WELDING OF PLASTICS, *WELD. REV.*, FEB 1983, P 17-22
2. M.N. WATSON, R.M. RIVETT, AND K.I. JOHNSON, "PLASTICS--AN INDUSTRIAL AND LITERATURE SURVEY OF JOINING TECHNIQUES," REPORT NO. 7846.01/85/471.3, THE WELDING INSTITUTE, ABINGTON, UK, 1986
3. H. POTENTE AND P. MICHEL, THE STATE OF THE ART DEVELOPMENT TRENDS IN THE WELDING OF PLASTICS, *PROC. 5TH ANNUAL NORTH AMERICAN WELDING RESEARCH CONFERENCE,* EWI AND AWS, 1989
4. V.K. STOKES, JOINING METHODS FOR PLASTICS AND PLASTIC COMPOSITES: AN OVERVIEW, *POLYM. ENG. SCI.,* VOL 29 (NO. 19), 1989, P 1310-1324
5. R.A. GRIMM, FUSION WELDING TECHNIQUES FOR PLASTICS, *WELD. J.,* MARCH 1990, P 23-28
6. G. MENGES, THE JOINING OF PLASTICS AND THEIR COMPOSITES, *PROC. LNT. CONF. ADVANCES IN JOINING NEWER STRUCTURAL MATERIALS,* IIW (MONTREAL), P 33-63

7. A.J. KINLOCH, *ADHESION AND ADHESIVES,* CHAPMAN AND HALL, 1987, P 101
8. D. CHANT, JOINING TECHNOLOGY FOR THERMOPLASTIC COMPOSITE STRUCTURES IN AEROSPACE APPLICATIONS, *PROC. INT. CONF. ADVANCES IN JOINING PLASTICS AND COMPOSITES,* THE WELDING INSTITUTE, CAMBRIDGE, 1991
9. S.S. VOYUTSKII, AUTOHESION AND ADHESION OF HIGH POLYMERS, WILEY INTERSCIENCE, 1963
10. A. BENATAR, MATERIAL CHARACTERISTICS FOR WELDING, *PROC. 5TH ANNUAL NORTH AMERICAN WELDING RESEARCH CONFERENCE,* EWI AND AWS, 1989
11. R.P. WOOL, B.-L. YUAN, AND O.J. MCGAREL, *POLYM. ENG. SCI.,* VOL 29 (NO. 19), 1989, P 1340-1367
12. W.V. TITOW, CHAPTER 12, *SOLVENT WELDING OF PLASTICS,* APPLIED SCIENCE PUBLISHERS, LONDON, P 181-196
13. M. LICATA AND E. HAAG, SOLVENT WELDING WITH POLYCARBONATE, *PROC. SPE 44TH ANTEC MEETING,* SOCIETY OF PLASTICS ENGINEERS, 1986, P 1092-1094

Welding of Plastics

Thomas H. North and Geetha Ramarathnam, University of Toronto

Fusion-Welding Techniques

The commonly available fusion-welding techniques will be described in terms of the basis of the joining method, the key joining parameters involved in the process operation, and the application areas of each joining method.

Hot-Tool Welding

Hot-tool welding occupies a central position among the different thermal fusion-welding techniques. This method can provide joint strength that is equal to that of the base material. The surfaces to be joined are brought to the melting or softening temperature by direct contact with a heated tool. Once the desired molten film thickness is produced, the heated tool is removed, the melted surfaces are brought together, and the joint is allowed to cool and consolidate under pressure. Typical welding times range from 10 s to 60 min, depending on the size of the component being joined.

The surfaces of the heated tool are usually coated with polytetrafluoroethylene (PTFE) to prevent the polymers from sticking to the platen. However, the PTFE coating generally restricts the maximum tool surface temperature to 260 °C (500 °F). Some research on heated-tool welding has involved hot-plate temperatures in excess of 350 °C (660 °F) for very short heating times (4 to 6 s). This thermal cycle has been made possible through the use of high-temperature aluminum-bronze heating elements (Ref 1).

The hot-tool welding process can be described by four phases (Fig. 3). During phase I, the surfaces of the material are brought in contact with the heated tool and are held under pressure. This pressure is maintained until a molten film appears. In phase II, the contact pressure between the heated tool and the substrate is reduced to increase the molten-film thickness. The rate of increase of melted-film thickness depends on the thermal conductivity of the polymer. Phase III is the change-over (removal of the hot tool) time, whereas phase IV is the joining and cooling under pressure. The amount of melted-polymer displacement from the weld zone is controllable. For example, computer-controlled machines will allow preselection of the pressure or displacement values that are applied during joining (Ref 15). In this connection, detailed mathematical analyses of the hot-tool welding process have already been carried out (Ref 15, 17).

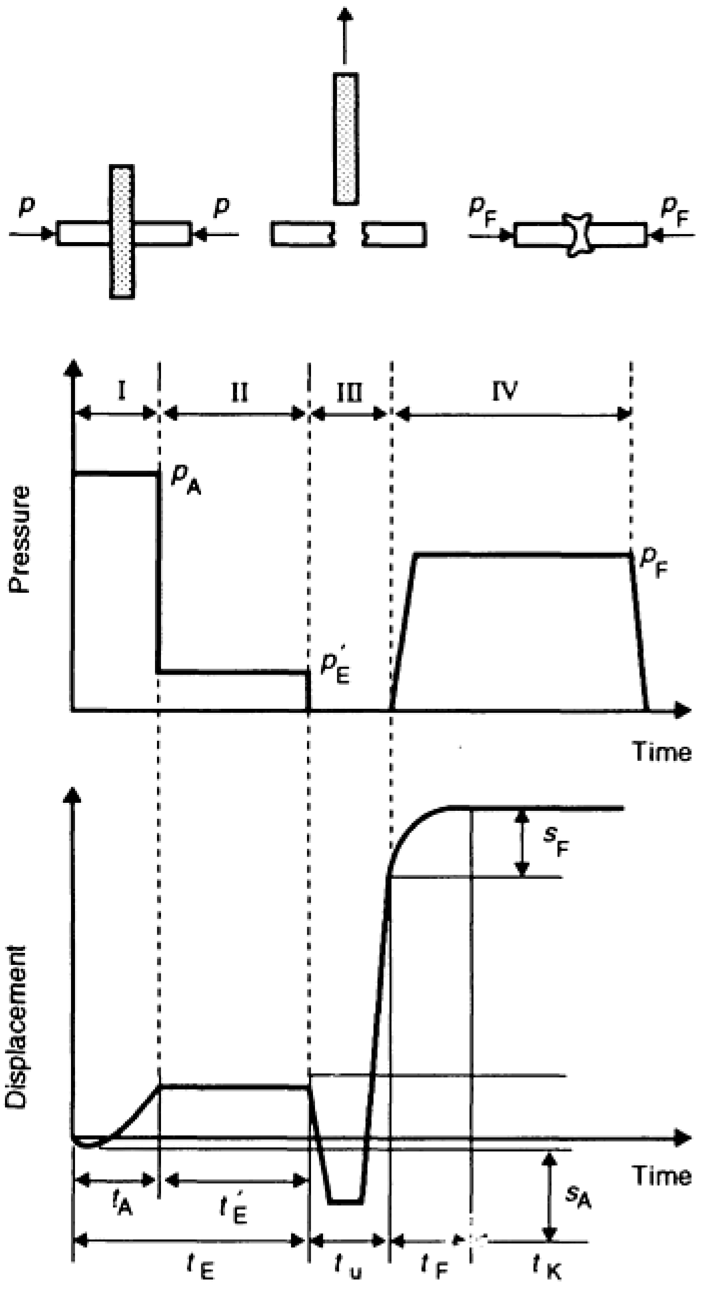

FIG. 3 PRESSURE-TIME AND DISPLACEMENT-TIME GRAPH SHOWING DIFFERENT PHASES OF HOT-TOOL WELDING PROCESS. P_A, P_E, AND P_F ARE MATCHING, HEATING, AND JOINING PRESSURES, RESPECTIVELY. T_E, T_U, T_F, AND T_K REPRESENT HEATING, CHANGE-OVER, JOINING, AND COOLING TIMES, RESPECTIVELY. S_F, REDUCTION IN LENGTH OF THE PART BEING JOINED; S_A, DISPLACEMENT PATH PRODUCED DURING TIME T_E. SOURCE: REF 16

The key joining parameters of hot-tool welding are:

- THE HEATED-TOOL TEMPERATURE, WHICH DEPENDS ON THE POLYMER BEING WELDED
- THE PRESSURE APPLIED AND THE DURATION OF PHASE I. (HOWEVER, HIGH PRESSURE DURING PHASE IV WILL PRODUCE TOO MUCH LATERAL FLOW AND POLYMER-CHAIN ORIENTATION IN THE COMPLETED JOINT, RESULTING IN ADVERSE MECHANICAL PROPERTY EFFECTS, AS DESCRIBED IN REF 15 AND 16)
- THE THERMAL CONDUCTIVITY AND SPECIFIC HEAT OF THE POLYMER AND THE SIZE

AND WALL THICKNESS OF THE PART BEING JOINED, WHICH ARE THE PARAMETERS THAT DETERMINE THE DURATION OF PHASE II

- THE TOOL TRANSFER, OR CHANGE-OVER, TIME, WHICH MUST BE MINIMIZED SO THAT COOLING OF THE MOLTEN LAYER DOES NOT OCCUR (TO AVOID INHIBITING POLYMER-CHAIN INTERDIFFUSION)

Applications. All thermoplastic materials can be joined using hot-tool welding. Components that have large, flat surface areas are commonly butt welded using this technique. It is also possible to join dissimilar materials through the use of two heated platens that are at different temperatures (Ref 14). The weldability factor, in this case, is the degree of compatibility.

Hot-tool welding can be readily automated. Portable equipment is primarily used for on-site weld repairs. Small amounts of joint misalignment prior to joining have negligible effects on the weld quality. Hot-tool welding is used in a variety of industrial applications. The automotive sector, for example, uses it to weld polypropylene (PP) copolymer cases for batteries and to weld rear-light casings of acrylonitrile-butadiene-styrene (ABS) joints to polymethyl methacrylate (PMMA) or polycarbonate (PC) lenses. Hot-tool welding is also employed when welding thermoplastic tanks and when joining large-diameter polyethylene (PE) pipeline to transport gas, water, and sewage wastes.

Hot-Gas Welding

A stream of hot air or gas (nitrogen, air, carbon dioxide, hydrogen, or oxygen) is directed toward the filler and the joint area using a torch (Ref 18). A filler rod or tape (of a similar composition to the polymer being joined) is gently pushed into the gap between the substrates (Fig. 4). A variety of nozzles are available for different applications, and either automated or manual welding can be carried out. During welding, the gas temperature can range from 200 to 600 °C (390 to 1110 °F), depending on the polymer being joined.

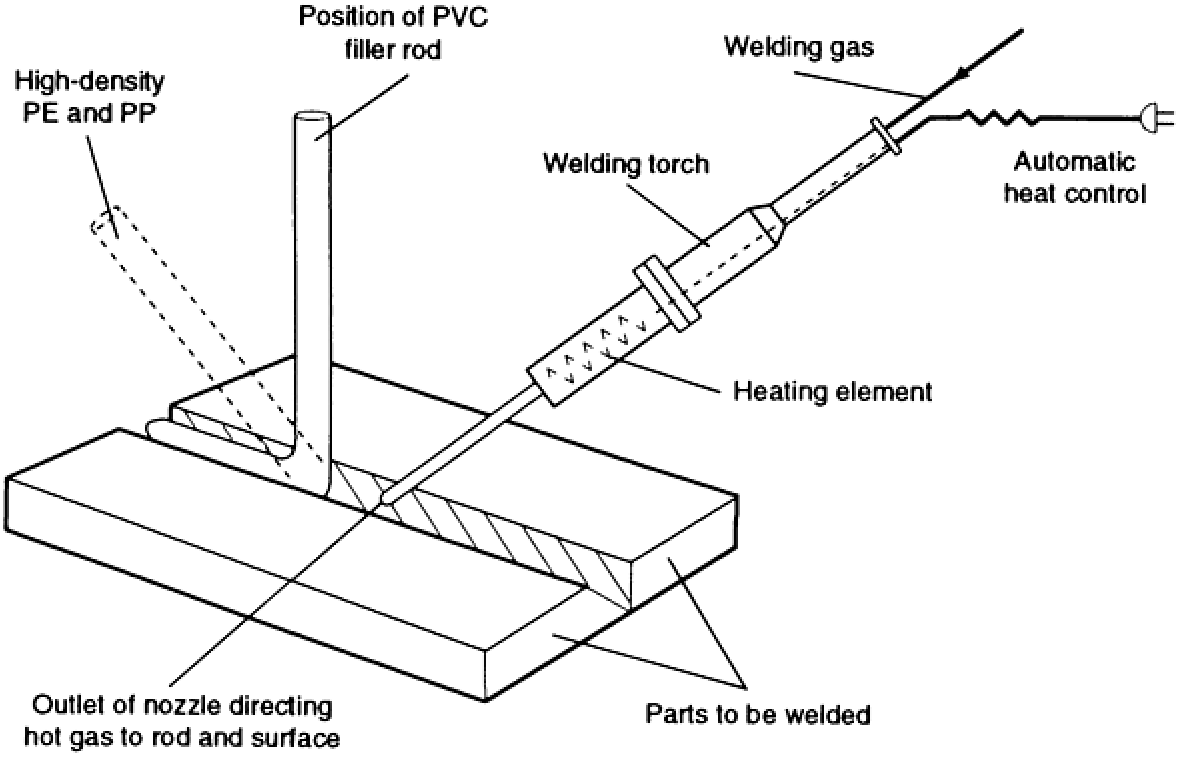

FIG. 4 SCHEMATIC OF HOT-GAS WELDING, SHOWING THE CORRECT POSITION OF TORCH AND FILLER ROD FOR DIFFERENT THERMOPLASTICS. SOURCE: REF 19

The key joining parameters of the hot-gas welding process are:

- GAS TEMPERATURE, WHICH DEPENDS ON THE TYPE OF POLYMER BEING JOINED, AND WHICH DETERMINES THE HEATING ELEMENT, NOZZLE DIMENSIONS, AND GAS/AIR FLOW RATES THAT ARE USED
- WELDING SPEED AND DOWNWARD PRESSURE OF THE WELDING ROD (MANUAL JOINING OPERATIONS)

Applications. The principal application areas involve the continuous welding of polyolefin tanks and containers, the welding of polyvinyl chloride (PVC), ABS, PE, and PP pipe sections, the sealing of packaging materials, and the field repair of PVC and other thermoplastic resins that are used in the construction and automotive industries. Hot-gas welding has a disadvantage in that the temperature of the hot gas/air is much higher than the melting point of the polymer being joined. Therefore, the process has poor energy efficiency, and degradation of the polymer substrate is possible unless care is taken. Polymers that oxidize at temperatures close to their melting points cannot be welded by this technique.

Extrusion Welding

Extrusion welding is similar to hot-gas welding. The weld area is heated using hot air, and plasticized filler material is extruded into V-shaped or lap seam joints under pressure by means of a welding shoe (Ref 20, 21). The preheating and welding-shoe assemblies are connected to the welding head, and the welding speed is kept constant by an automatic traversing unit that has adjustable motor speeds (Fig. 5).

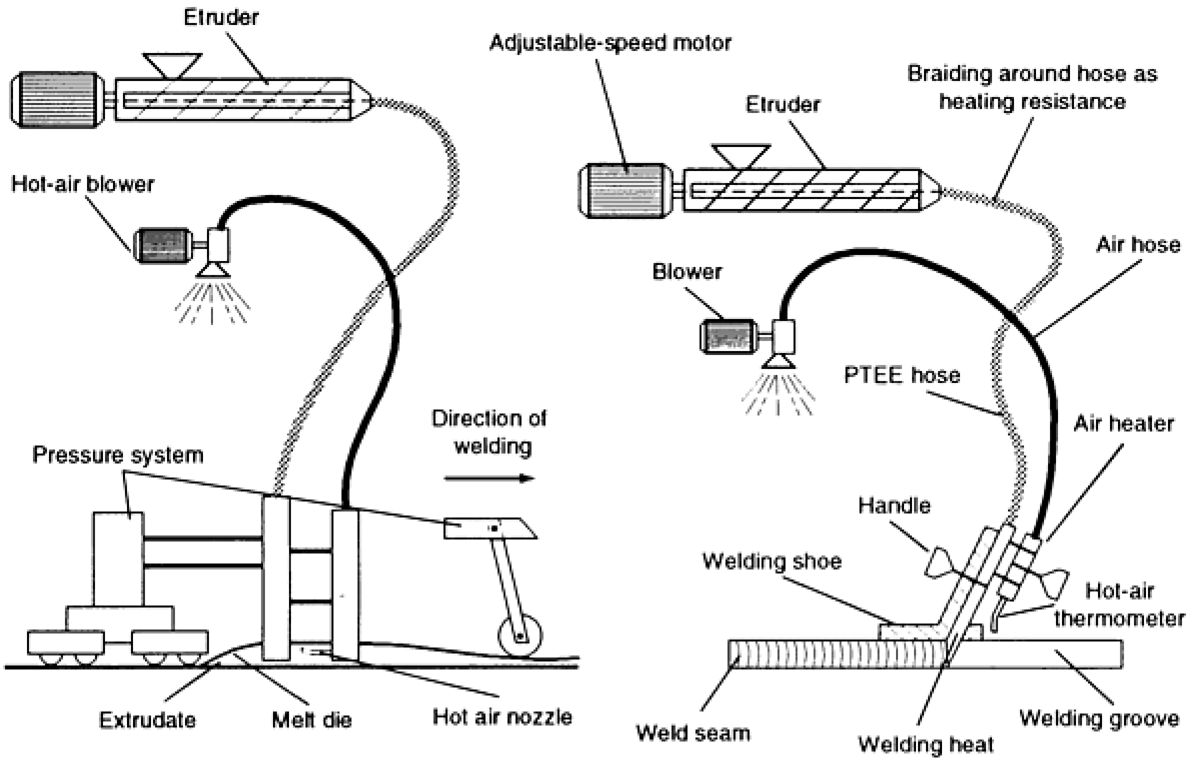
Etruder
Hot-air blower
Pressure system
Direction of welding
Extrudate
Melt die
Hot air nozzle
Adjustable-speed motor
Etruder
Braiding around hose as heating resistance
Air hose
Blower
PTEE hose
Air heater
Handle
Welding shoe
Hot-air thermometer
Weld seam
Welding heat
Welding groove

FIG. 5 SCHEMATIC OF VARIANTS OF EXTRUSION WELDING PROCESS. SOURCE: REF 20

The key joining parameters of this process are the:

- FEED ANGLE AND TEMPERATURE OF THE HOT GAS
- TEMPERATURE AND PRESSURE OF THE EXTRUDATE
- RELATION BETWEEN THE EXTRUDATE DELIVERY RATE AND THE WELDING SPEED
- HORIZONTAL AND VERTICAL FORCES APPLIED AT THE WELDING SHOE

- WELD LENGTH AND TIME REQUIRED TO COMPLETETHE JOINT

Applications. Extrusion welding is primarily used for producing long seams in thick-section polyolefin components. This joining technique is particularly useful when welding large container sections.

Focused Infrared Welding

This joining technique uses a quartz lamp and focuses the infrared (IR) radiation using highly polished, parabolic, elliptical reflectors. The focused IR method directs a precise, high-intensity reciprocating IR beam of a width that ranges from 1.5 to 3.0 mm (0.06 to 0.12 in.) onto the joint interface (Ref 22). A robotic fixture scans the IR beam back and forth over the two adherend surfaces, and when the joint line reaches the desired temperature, the heat source is removed and the substrates are forged together in a press. Sensors are mounted adjacent to the lamp fixture and are calibrated so that they selectively measure the adherend interface temperature only (Fig. 6).

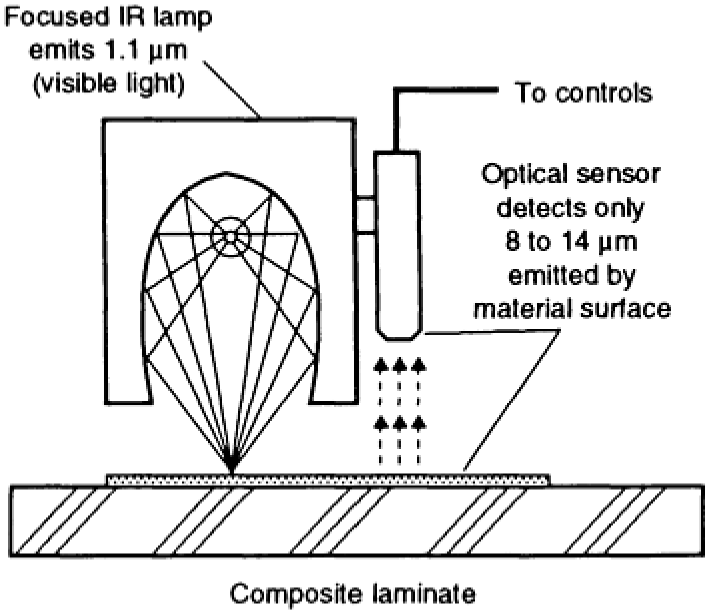

FIG. 6 SCHEMATIC OF FOCUSED IR LAMP AND OPTICAL SENSOR. SOURCE: REF 22

The welding of polyphenylene sulfide, using this IR technique, has been reported (Ref 23). Focused IR welding can be automated. Because this is a relatively new joining process, a detailed analysis of process operation has yet to be conducted.

The key joining parameters of this welding process are the:

- RADIATION DENSITY AND TIME OF IRRADIATION
- RADIATION ABSORPTION, REFLECTION, AND TRANSMISSION PROPERTIES OF THE POLYMER, WHICH DETERMINE THE RATE OF THE TEMPERATURE RISE IN THE IRRADIATED MATERIAL
- FLOW TEMPERATURE AND THERMAL CONDUCTIVITY OF THE IRRADIATED POLYMER
- RECIPROCATION RATE OF THE RADIATOR, WHICH DEPENDSON THE TYPE OF POLYMER BEING JOINED, AND ON THE PART SIZE
- RETRACTION RATE OF THE RADIATIVE HEAT SOURCE AND CLOSING SPEED OF THE PRESS (IN ORDER TO AVOID COOLING THE MELTED MATERIAL AT THE JOINT INTERFACE)

Applications. This technique can be used to join both simple and complex joint configurations when a noncontacting method of heating is essential. In addition, reinforced-fiber disruption can be minimized when advanced thermoplastic composites are joined.

Laser Welding

Limited information is available on laser welding. Carbon dioxide (CO_2) lasers induce excitation of the vibrational modes and, hence, heating of the irradiated organic material (Ref 24, 25). At low power input levels, satisfactory weld penetration can be achieved. The technique also provides high welding speeds and produces very small heat-affected zones (HAZ). Typical joining conditions involve an energy input of 100 W using a wavelength of 10.6 μm. The absorption of laser radiation depends on the relation:

$$I(Z) = I_0 \text{ EXP } (-\alpha Z) \qquad \textbf{(EQ 2)}$$

where z is the distance into the sample at which the laser intensity I (W/cm^2) is measured, I_0 is the laser intensity at the surface of the polymer sample (at z = 0), and α (1/cm) is the absorption coefficient. The laser power, P (watts) required for any given material to depth a (meters) can be derived using the relation (Ref 25):

$$p = \frac{vWa(r\,C_p T_m + \Delta H_m)}{0.484} \qquad \textbf{(EQ 3)}$$

where v is the welding speed (m/s), W is the weld width (m), ρ is the density of polymer (ks/m^3), C_p is the heat capacity (J/kg · °C), T_m is the melting temperature (°C), and ΔH_m is the latent heat of melting (J/m^3).

There is not much that has been published concerning the commercial application of this joining method.

References cited in this section

1. G. GEHARDSSON, THE WELDING OF PLASTICS, *WELD. REV.,* FEB 1983, P 17-22
14. K. GABLER AND H. POTENTE, WELDABILITY OF DISSIMILAR THERMOPLASTICS--EXPERIMENTS IN HEATED TOOL WELDING, *J. ADHES.,* VOL 11, 1980, P 145-163
15. H. POTENTE AND P. TAPPE, SCALE-UP LAWS IN HEATED TOOL BUTT WELDING OF HDPE AND PP, *POLYM. ENG. SCI.,* VOL 29 (NO. 23), 1989, P 1642-1648
16. H. POTENTE AND J. NATROP, COMPUTER AIDED OPTIMIZATION OF THE PARAMETERS OF HEATED TOOL BUTT WELDING, *POLYM. ENG. SCI.,* VOL 29 (NO. 23), 1989, P 1649-1654
17. A.J. POSLINSKI AND V.K. STOKES, ANALYSIS OF THE HOT-TOOL WELDING PROCESS, *PROC. SPE 50TH ANTEC MEETING,* 1992, SOCIETY OF PLASTICS ENGINEERS, P 1228-1233
18. H. GUMBLETON, HOT GAS WELDING OF THERMOPLASTICS--AN INTRODUCTION, *JOIN. MAT.,* VOL 5, 1989, P 215-218
19. "RECOMMENDED PRACTICES FOR JOINING PLASTIC PIPING," DOCUMENT XVI-322-78-E, IIW
20. P. MICHEL, "AN ANALYSIS OF THE EXTRUSION WELDING PROCESS," *POLYM. ENG. SCI.,* VOL 29 (NO. 19), 1989, P 1376-1382
21. M. GEHDE AND G.W. EHRENSTEIN, STRUCTURAL AND MECHANICAL PROPERTIES OF OPTIMIZED EXTRUSION WELDS, *POLYM. ENG. SCI.,* VOL 31 (NO. 7), 1991, P 495-501
22. H. SWARTZ AND J.L. SWARTZ, FOCUSSED INFRARED--A NEW JOINING TECHNOLOGY FOR HIGH PERFORMANCE THERMOPLASTICS AND COMPOSITE PARTS, *PROC. 5TH ANNUAL NORTH AMERICAN WELDING RESEARCH CONFERENCE,* EWI AND AWS, 1989
23. H. POTENTE, P. MICHEL, AND M. HEIL, INFRARED RADIATION WELDING: A METHOD FOR WELDING HIGH TEMPERATURE RESISTANT THERMOPLASTICS, *PROC. SPE 49TH ANTEC MEETING,* SOCIETY OF PLASTICS ENGINEERS, 1991, P 2502-2504

24. W.W. DULEY, IN *LASER PROCESSING AND ANALYSIS OF MATERIALS,* PLENUM PRESS, 1983

25. W.W. DULEY AND R.E. MUELLER, CO_2 LASER WELDING OF POLYMERS, *POLYM. ENG. SCI.*, VOL 32 (NO. 9), 1992, P 582-585

Welding of Plastics

Thomas H. North and Geetha Ramarathnam, University of Toronto

Friction Welding

Frictional heating is important in the ultrasonic, vibration, and spin-welding methods. In ultrasonic welding, heat generation depends on the storage and loss moduli of the polymer. In the cases of vibration and spin welding, heat generation depends on the frictional coefficient of the polymer material, and on the storage and loss moduli values (Ref 10). As would be expected, joining processes that depend on direct heat generation at the bond line are more efficient in terms of energy utilization, and the heating and pressure cycles are applied simultaneously.

Vibration Welding

Samples are clamped together and vibrated using an oscillating motion under pressure (Fig. 7). Frictional heating is due to one solid being rubbed against the other, and to the shear heating of the melt. Vibratory motion is applied until the polymer melts at the bond line. Then, the motion is terminated and melted polymer at the joint interface is cooled under pressure.

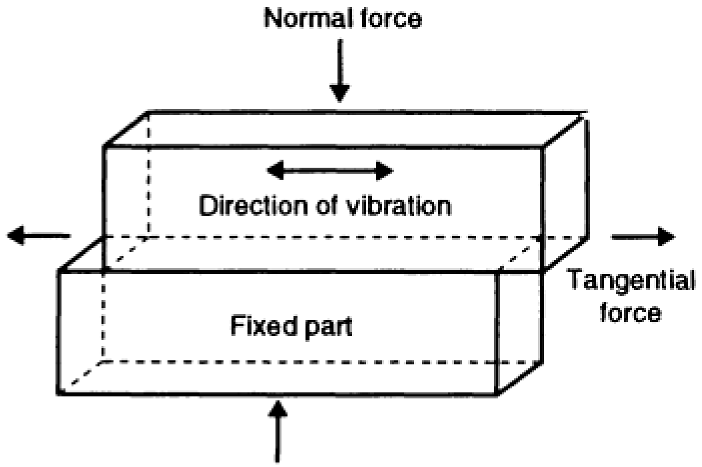

FIG. 7 SCHEMATIC OF VIBRATION WELDING PROCESS

Two types of vibration welding machines are available: linear machines, in which the center of motion lies outside of the molded part; and angular machines, in which the center of motion lies within the molded part (Ref 3). Linear vibration welding is more popular, because of its ability to weld long and narrow molded parts.

During welding, the samples are fixed to upper and lower platens, and the lower platen is kept fixed while the upper platen is vibrated using an electromagnetic or hydraulic drive system at frequencies ranging from 100 to 500 Hz. The vibrational amplitude ranges from 0.1 to 5.0 mm (0.004 to 0.2 in.), depending on the frequency used (Ref 26, 27).

The detailed features of the vibrational welding process have been studied, along with welding parameter optimization, by a number of investigators (Ref 28, 29, 30). It has been shown that the vibration welding process involves four distinct phases (Fig. 8). Phase I comprises the initial heating of the interface to the melting point as a result of coulomb friction. Phase II involves melting and unsteady material flow in the lateral direction. In phase III, melting and flow are at a steady state, and weld penetration (the decrease in the melted zone width that is due to lateral flow) increases linearly with time. When the vibratory motion is terminated, weld penetration continues to increase until the material solidifies, which represents phase IV.

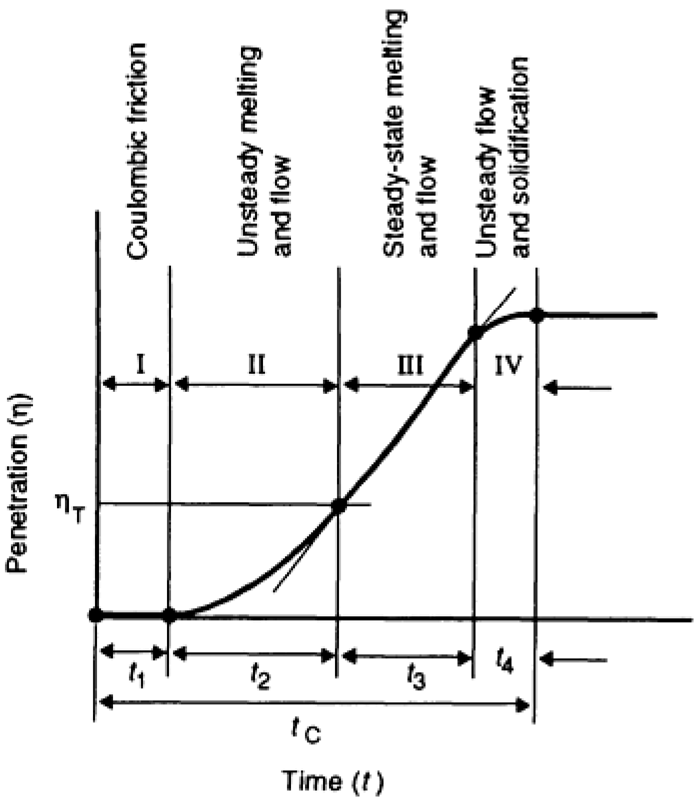

FIG. 8 SCHEMATIC OF PENETRATION-TIME GRAPH SHOWING THE FOUR PHASES OF VIBRATION WELDING PROCESS. SOURCE: REF 26

The key joining parameters of this process are the:

- AMPLITUDE, FREQUENCY, AND TIME OF VIBRATION APPLIED DURING WELDING
- PRESSURE APPLIED DURING THE JOINING OPERATION

It has been reported that the use of a process-controlled pressure profile (high pressures during phase I and lower applied pressures during the subsequent phases) will improve final joint mechanical properties and decrease the welding time, when compared with the time used in a constant weld pressure cycle (Ref 29).

Applications. The success of vibration welding lies in the fact that it can be used to join large parts and is limited by machine size. The joining process, which can be readily automated, can be applied to a variety of plastics, particularly semicrystalline polymers such as PP, PE, acetal, and nylon, which are difficult to weld using ultrasonic welding. Butt weld joints are the principal type of joint geometry that is used. The primary application areas are in the automotive, aeronautical, and household appliance sectors.

Spin Welding

Spin welding is ideally suited for joining cylindrical or circular parts. One half of the component is fixed, and the other half rotates at a prescribed angular velocity while pressure is applied (Fig. 9). Once a desired melted polymer thickness is achieved, the rotational motion is terminated and the parts are pressed together and cooled.

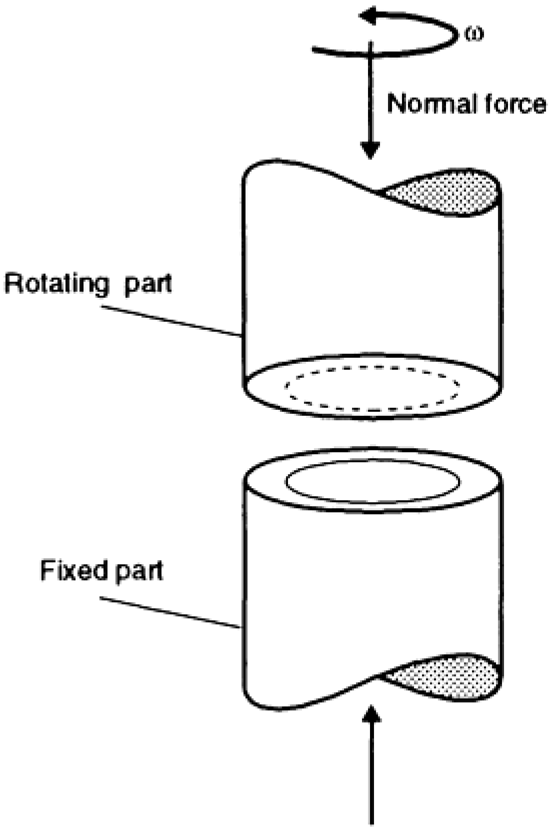

FIG. 9 SCHEMATIC OF SPIN WELDING PROCESS

Spin-welding machines can contain a range of sensors that indicate the rotational speed, axial pressure, weld penetration, velocity, and torque during the joining operation (Ref 31). The joining process can be described as having five distinct phases (Fig. 10). Solid (coulombic) friction, followed by friction and generation of wear particles, occurs during phases I and II. Phase III is the melting transition. Shearing of the melt produces further heat generation, and the weld penetration increases linearly with time in phase IV. Phase V involves cooling and consolidation of the joint under pressure.

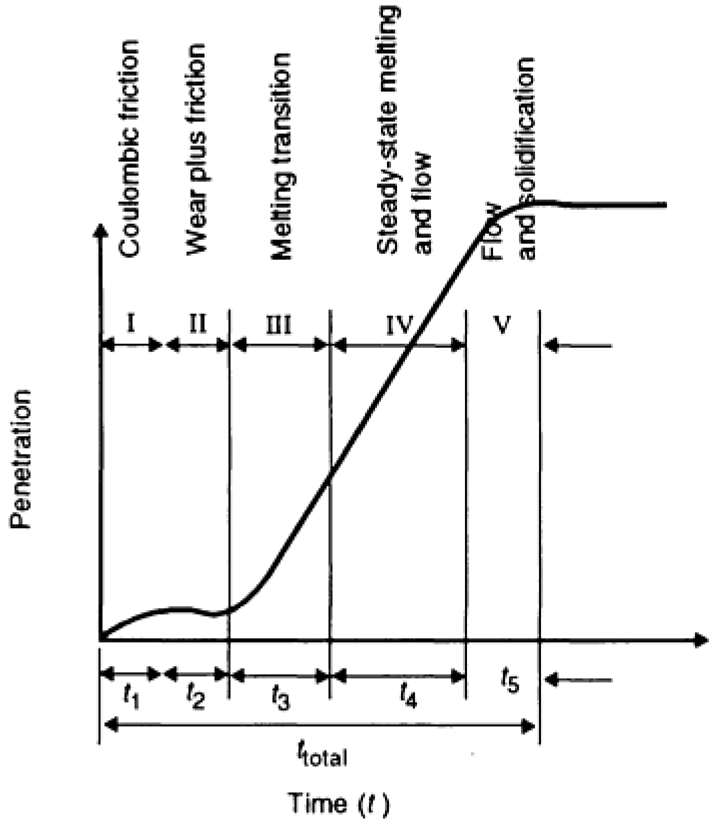

FIG. 10 PENETRATION-TIME GRAPH OF THE SPIN-WELDING PROCESS. SOURCE: REF 32

The key welding parameters are the:

- AXIAL FORCE APPLIED DURING THE WELDING OPERATION, WHICH DETERMINES THE ORIENTATION OF POLYMER CHAINS THAT DEVELOP IN THE COMPLETED JOINT AND MARKEDLY AFFECTS JOINT MECHANICAL PROPERTIES (REF 33)
- TIME REQUIRED TO COMPLETE THE WELDING OPERATION
- ROTATION SPEED EMPLOYED DURING SPIN WELDING

Applications. This inherently simple joining technique is capable of joining large-diameter sections, although it is applicable only to rotationally symmetric components. In the automotive sector, a typical application involves the joining of small, circular parts to large, blow-molded or injection-molded sections. Other automotive-related examples are the joining of ventilation fittings to blow-molded fuel tanks and the welding of gasoline filters and carburetor housings. Spin welding is also used in the manufacture of floats and aerosol bottles in the packaging industry.

Ultrasonic Welding

Ultrasonic welding is the most widely used polymer joining process. Numerous published papers describe the basic features of this joining technique. In this welding process, the sonotrode delivers low-amplitude (15 to 60 μm), high-frequency sinusoidal vibration in a direction perpendicular to the bond line. Although ultrasonic welding machine frequencies range from 10 to 50 kHz, the commonly available devices apply frequencies that are between 20 and 40 kHz.

Ultrasonic-welding machines generally comprise an actuator that contains the converter, booster, horn, and other pneumatic controls (Fig. 11). The ultrasonic vibrations are generated by exciting piezoelectric or magnetostrictive transducer crystals using electrical energy from a power supply. The transducer assembly is rigidly attached to a booster rod that increases or decreases the vibrational amplitude. Finally, the booster is connected to the sonotrode, which transfers the vibratory motion to the workpiece. The converter, booster, and horn must be resonant at the operating frequency.

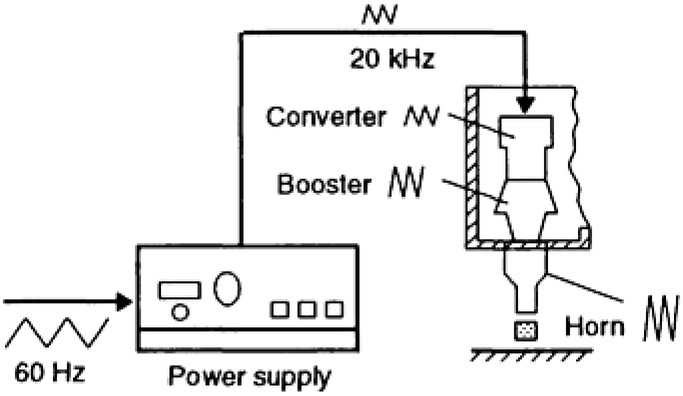

FIG. 11 SCHEMATIC OF ULTRASONIC-WELDING COMPONENTS

The weldability of any polymer depends on the damping capacity, or attenuation characteristics, and size of the component being joined. Depending on the distance between the sonotrode and the bond line, ultrasonic welding is classified as being either near-field (for distances under 6 mm, or 0.24 in.) or far-field (for distances over 6 mm, or 0.24 in.) welding.

It follows that the component geometry has a crucial influence on energy transmission, heat generation, and polymer melting at the joint interface. The transmission of ultrasonic vibrations through the workpiece sets up standing-wave patterns and regions of maximum and minimum displacements (Ref 34). The material fuses in regions of maximum strain or stress.

Amorphous thermoplastics are more readily weldable when an energy director is used. A triangular projection is molded onto one of the specimen surfaces, usually the one that contacts the sonotrode (Fig. 12a). Ultrasonic welds in semicrystalline resins are commonly produced using an interference, or shear joint, design (Fig. 12b). A novel technique that utilizes a tielayer at the interface has been developed recently (Fig. 12c) for use during the ultrasonic welding of flat contacting surfaces.

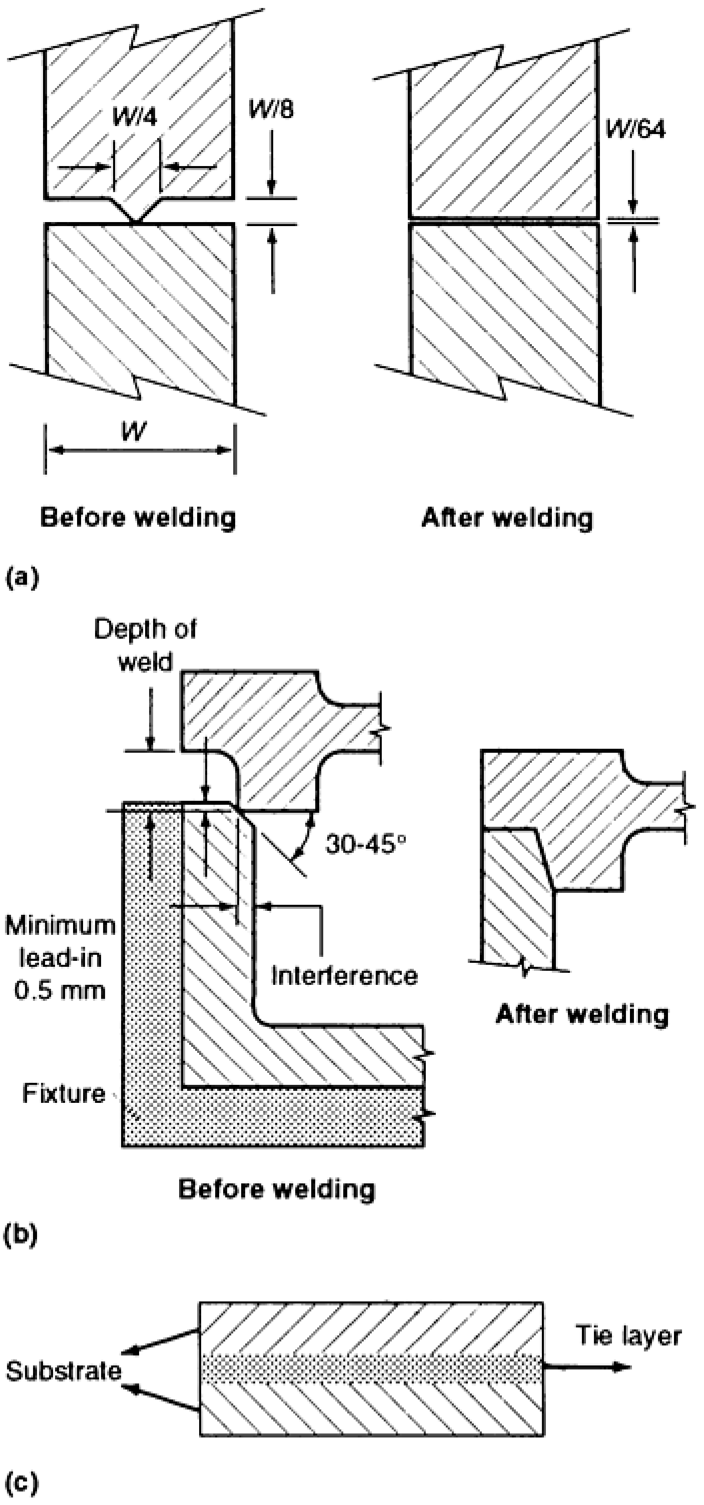
W/4
W/8
W
Before welding
W/64
After welding
(a)
Depth of weld
30-45°
Minimum lead-in 0.5 mm
Interference
Fixture
Before welding
After welding
(b)
Substrate
Tie layer
(c)

FIG. 12 SCHEMATIC OF ULTRASONIC CONFIGURATION. (A) ENERGY DIRECTOR DESIGN. (B) SHEAR JOINT DESIGN. (C) TIE-LAYER DESIGN. SOURCE: REF 35 AND 36

Energy generation and welding-parameter optimization during the ultrasonic welding of polystyrene (PS), a blend of PC and polybutylene terephthalate, and a composite of polyetherether ketone (PEEK) and graphite APC-2 have been analyzed in detail by different authors (Ref 37, 38, 39). The joining process comprises four distinct phases when tie layers are employed (Fig. 13). The rapid increase in temperature up to the polymer melting point results from frictional heating at surface asperities and, to a much lesser extent, viscoelastic dissipation. In phase II of the ultrasonic-welding process, bulk melting of the interface region predominantly results from viscoelastic energy dissipation. Further heating is due to viscous dissipation during phase III. In phase IV, the joint cools under pressure.

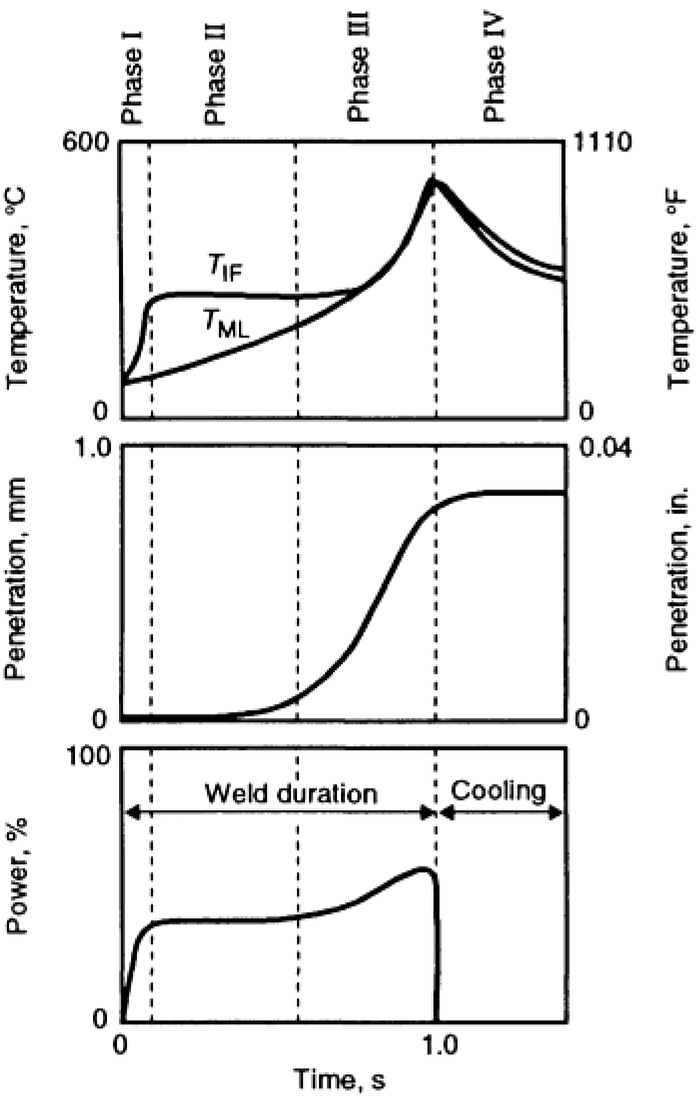

FIG. 13 TYPICAL PROCESS DATA TRACES-TIME GRAPH OF ULTRASONIC-WELDING PROCESS WHEN USING TIE LAYERS AT THE INTERFACE. SOURCE: REF 40

The weldability of a thermoplastic can be defined using the following relationships (Ref 3):

$$f = \int_{T1}^{T2} \frac{\rho c}{\left\{ E' \dfrac{\tan \delta}{2} + \dfrac{0.25}{\rho} \left[1 - \left(\dfrac{P}{P_x} \right)^{0,7} \right] v \right\}} dt \qquad \textbf{(EQ 4)}$$

$$\wedge = \pi \mathrm{TAN}\ \delta \qquad \textbf{(EQ 5)}$$

where f is the energy index, T_1 is the ambient temperature, T_2 is the softening point or crystalline melting point of the polymer, ρ is the density, c is the specific heat, E' is the real component of the complex elastic modulus, tan δ is the mechanical loss factor, P and P_x are the joining and critical joining pressure values, respectively, v is the coefficient of friction, and $\wedge$ is the logarithmic damping decrement. When the f value increases, more energy is required during the ultrasonic-welding process. When the logarithmic decrement factor $\wedge$ has a low value, more energy arrives at the joint interface. The f and $\wedge$ indexes are plotted in Fig. 14.

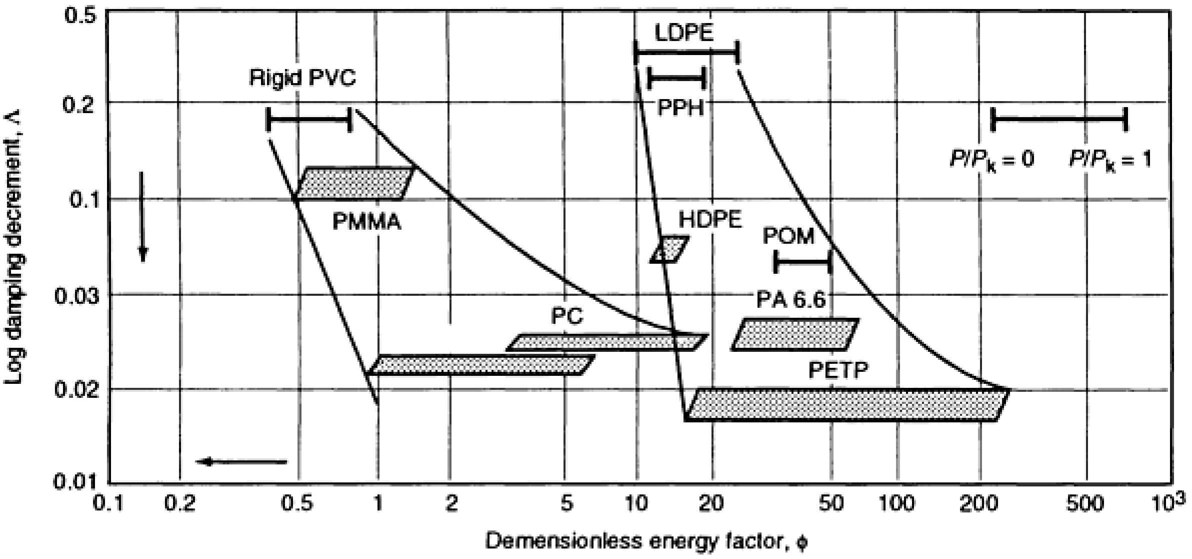

FIG. 14 DIAGRAM FOR ASSESSMENT OF WELDING CAPACITY OF THERMOPLASTICS. ARROW DIRECTION SHOWS INCREASE IN WELDABILITY BY THE INDIRECT METHOD. LDPE, LOW-DENSITY POLYETHYLENE; HDPE, HIGH-DENSITY POLYETHYLENE; PA, POLYAMIDE; POM, POLYOXYMETHYLENE. SOURCE: REF 34

The key joining parameters of ultrasonic welding are the:

- AMPLITUDE OF VIBRATION DURING JOINING
- PRESSURE APPLIED DURING THE WELDING OPERATION
- JOINT GEOMETRY WHEN WELDING AMORPHOUS AND SEMICRYSTALLINE POLYMERS
- WELDING TIME AND ENERGY APPLIED DURING THE WELDING PROCESS
- TRIGGER PRESSURE/FORCE RELATIONSHIP (THE PRESSURE AT WHICH ULTRASONIC VIBRATIONS ARE INITIATED)

Applications. The success of ultrasonic welding lies in the fact that the process is very fast (welding times are approximately 1 s), the equipment is compact, the process can be readily automated, and consistently acceptable weld quality can be produced. Ultrasonic welding is commonly used to join components in the automotive, toy, domestic appliance, and packaging industries. It is also the method of choice when welding many medical plastic components. The only limitation is that large parts cannot be welded using this method.

References cited in this section

3. H. POTENTE AND P. MICHEL, THE STATE OF THE ART DEVELOPMENT TRENDS IN THE WELDING OF PLASTICS, *PROC. 5TH ANNUAL NORTH AMERICAN WELDING RESEARCH CONFERENCE,* EWI AND AWS, 1989
10. A. BENATAR, MATERIAL CHARACTERISTICS FOR WELDING, *PROC. 5TH ANNUAL NORTH AMERICAN WELDING RESEARCH CONFERENCE,* EWI AND AWS, 1989
26. V.K. STOKES, VIBRATION WELDING OF THERMOPLASTICS, PART I, *POLYM. ENG. SCI.,* VOL 28, 1988, P 718-727
27. E. PECHA, D. CALINSKI, AND H. DIETER, PRODUCTION MACHINES AND TECHNICAL APPLICATION FOR HOT PLATE AND VIBRATION WELDING, *PROC. LNT. CONF. ADVANCES IN JOINING PLASTICS AND COMPOSITES,* THE WELDING INSTITUTE, CAMBRIDGE, 1991
28. V.K. STOKES, VIBRATION WELDING OF THERMOPLASTICS, PART II, *POLYM. ENG. SCI.,* VOL

28, 1988, P 727-739

29. H. POTENTE AND H. KAISER, PROCESS VARIANT OF VIBRATION WELDING WITH VARIABLE WELDING PRESSURE, *PROC. SPE 48TH ANTEC MEETING,* SOCIETY OF PLASTICS ENGINEERS, 1990, P 1762-1765
30. H. POTENTE AND M. UEBBING, COMPUTER AIDED LAYOUT OF THE VIBRATION WELDING PROCESS, *PROC. SPE 50TH ANTEC MEETING,* 1992, SOCIETY OF PLASTICS ENGINEERS, P 888
31. M. CAKMAK AND K. KEUCHEL, U.S. PATENT 4,998,663, MARCH 1991
32. K. KEUCHEL AND M. CAKMAK, SPIN WELDING OF POLYPROPYLENE: CHARACTERIZATION OF HEAT AFFECTED ZONE BY MICRO-BEAM WAXS TECHNIQUE, *PROC. SPE 49TH ANTEC MEETING,* SOCIETY OF PLASTICS ENGINEERS, 1991, P 2477-2481
33. H. RAJARAMAN AND M. CAKMAK, THE EFFECT OF GLASS FIBER FILLERS ON THE WELDING BEHAVIOR OF POLY P-PHENYLENE SULFIDE, *PROC. SPE 50TH ANTEC MEETING,* SOCIETY OF PLASTICS ENGINEERS, 1992, P 896-899
34. H. POTENTE, ULTRASONIC WELDING--PRINCIPLES AND THEORY, *MAT. DES.,* VOL 5, 1984
35. T.B. ZACH, J. LEW, T.H. NORTH, AND R.T. WOODHAMS, "JOINING OF HIGH STRENGTH ORIENTED POLYPROPYLENE USING ELECTROMAGNETIC INDUCTION BONDING AND ULTRASONIC WELDING, *MAT. SCI. TECH.,* VOL 5, 1989, P 281-287
36. J.S. WOLCOTT, DESIGNING PARTS FOR ULTRASONIC ASSEMBLY, *PROC. SPE 48TH ANTEC MEETING,* SOCIETY OF PLASTICS ENGINEERS, 1990, P 1829-1833
37. M.N. TOLUNAY, P.R. DAWSON, AND K.K. WANG, HEATING AND BONDING MECHANISMS IN ULTRASONIC WELDING OF THERMOPLASTICS, *POLYM. ENG. SCI.,* VOL 23, 1983, P 726-733
38. G. HABERNICHT AND J. RITTER, ENERGY CONVERSION IN THE ULTRASONIC WELDING OF THERMOPLASTICS, *KUNSTST. GER. PLAST.,* VOL 78, 1988, P 49-66
39. A. BENATAR AND T.G. GUTOWSKI, ULTRASONIC WELDING OF PEEK GRAPHITE APC-2 COMPOSITES, *POLYM. ENG. SCI.,* VOL 29 (NO. 23), 1989, P 1705-1721
40. N. TATEISHI, T.H. NORTH, AND R.T. WOODHAMS, ULTRASONIC WELDING USING TIE-LAYERS, PART I: ANALYSIS OF PROCESS OPERATION, *POLYM. ENG. SCI.,* VOL 32, 1992, P 600-611

Welding of Plastics

Thomas H. North and Geetha Ramarathnam, University of Toronto

Electromagnetic Welding

The joining techniques described in this section include resistance, induction, dielectric, and microwave welding.

Resistance or Implant Welding

In this technique, a conducting wire or mesh is inserted at the interface between the parts being welded. This insert is resistively heated by the passage of electric current (Fig. 15a), which melts the thermoplastic surrounding the insert, forming a weld. It should be noted that the implant remains in the completed joint. Electrofusion is a technique whereby the two ends of the pipe to be welded are inserted into a fitting/collar/coupler, the inside of which constitutes the heating element (Ref 41, 42). The electric current that passes through the heating element promotes melting on the inside surface of the fitting. Figure 15(b) shows a typical rectangular cross section of an electrofusion weld.

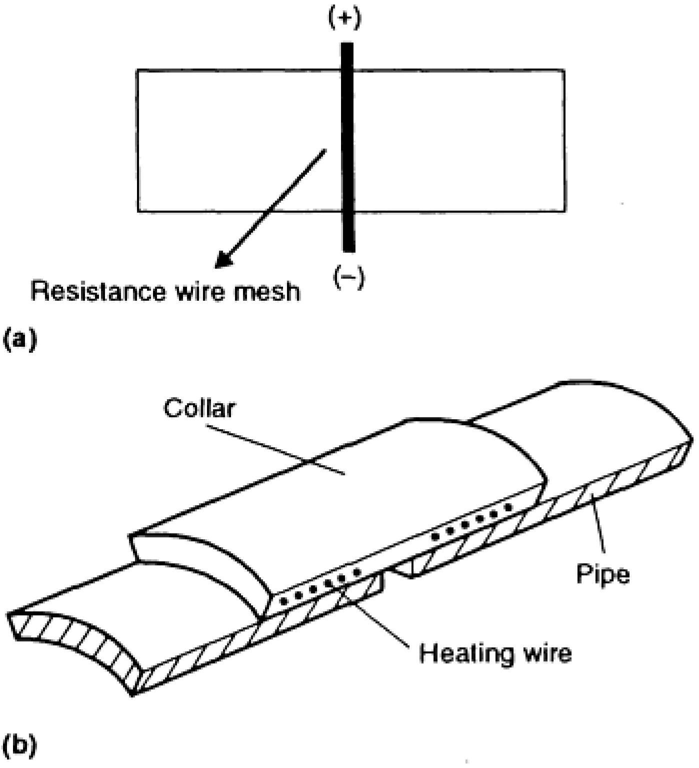

FIG. 15 (A) ELECTROFUSION (RESISTANCE) WELDING PROCESS. (B) ELECTROFUSION WELDING PROCESS. SOURCE: REF 42

The key joining parameters are the:

- APPLIED VOLTAGE AND APPLIED CURRENT, WHICH CAN BE AS HIGH AS 150 A
- CLAMPING PRESSURE DURING WELDING
- HEATING TIME
- AVOIDANCE OF CONTACTING SURFACE CONTAMINATION. BECAUSE MELT FLOW IS RESTRICTED DURING WELDING, IMPURITIES CAN BE TRAPPED AT THE JOINT INTERFACE, REDUCING THE STRENGTH OF THE WELD

Applications. The principal advantage of this joining method is that the equipment is relatively simple and portable. Therefore, it can be used in on-site repair situations. The primary application area for electrofusion welding is the joining of PE pipes with diameters less than 180 mm (7 in.), which are used for water transportation and natural gas distribution (Ref 42).

Induction Welding

In this joining method, a layer of electromagnetic material in the form of a tape or a thin sheet is placed at the joint interface. This insert material is heated using a high-frequency (2 to 10 MHz) inductive supply (Ref 43, 44). The electromagnetic tapes contain conductive and ferromagnetic particles. The particle size, type, and concentration depend on the application desired. The welding equipment comprises a high-frequency induction generator, a water-cooled copper working coil, and press fixtures, which hold the polymer component so that it can be heated and loaded.

Heating results from eddy currents that are induced in the thermoplastic tape containing the ferromagnetic particles (Fig. 16), and energy generation is proportional to I^2R. Heat generation during electromagnetic welding has been modeled, particularly for cross-ply carbon fiber thermoplastic composites (Ref 45, 46).

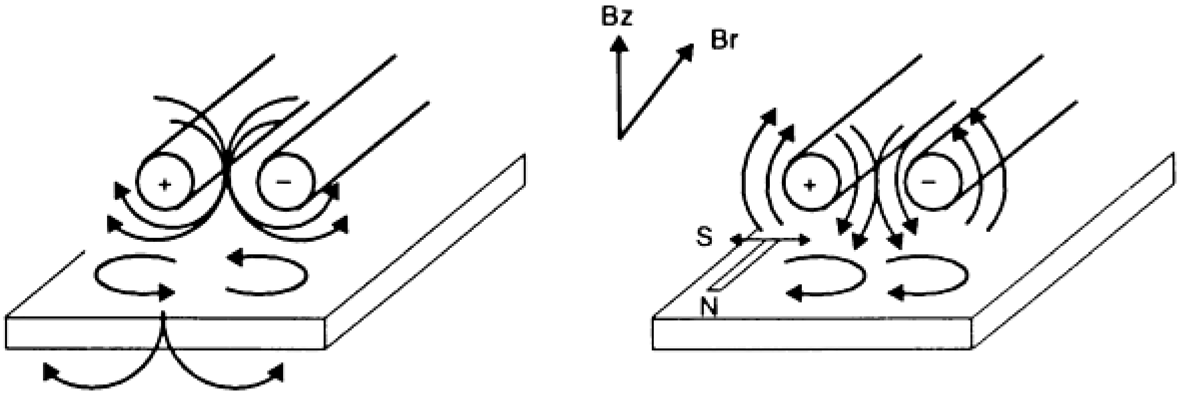

Eddy-current heating

Permeable material,
eddy-current heating,
dipole heating

FIG. 16 INDUCTION-HEATING MECHANISM. SOURCE: REF 5

The key joining parameters of this technique are the:

- COIL DESIGN, WHICH MUST BE CAREFULLY TAILORED TO SUIT THE END USE AND TO PRODUCE A UNIFORM RADIO-FREQUENCY (RF) FIELD. A RANGE OF COIL DESIGNS (SINGLE TURN, MULTITURN, HELICAL, BENT HAIRPIN, AND SO ON) ARE ALL POSSIBLE
- FREQUENCY AND POWER OUTPUT OF THE INDUCTION GENERATOR
- PARTICLE SIZE AND TYPE OF FILLER USED IN ELECTROMAGNETIC TAPE MANUFACTURE
- DISTANCE BETWEEN THE INDUCTOR COILS AND THE BOND LINE

Applications. The principal advantage of this joining method is that it can handle components of complex geometry. The process is also suitable for joining filled thermoplastic materials. The primary applications are in the automotive, packaging, and appliance industries. In the medical field, electromagnetic induction welding has been employed for joining PC blood oxygenators and arterial filter components.

The principal disadvantages of this joining process are the high cost of the equipment, the recurring costs associated with the purchase of ferromagnetic-filled tapes, and the care that is required in the design and construction of the inductive heating coils.

Dielectric Welding

In dielectric welding, the joint interface is subjected to much higher RF electrical energy. The process utilizes these major components:

- A GENERATOR THAT PRODUCES A FREQUENCY RANGING FROM 13 TO 100 MHZ (27.12 MHZ IS TYPICAL) OF ELECTRICAL ENERGY AT AN OUTPUT OF 1 TO 25 KW
- A PRESS THAT CLAMPS THE SECTIONS TOGETHER DURING THE HEATING AND SUBSEQUENT COOLING CYCLES
- A BRASS ELECTRODE OR DIE THAT IS MOUNTED ON AN ALUMINUM PLATE, WHICH APPLIES THE ELECTRICAL ENERGY DIRECTLY TO THE PLASTIC (REF 47)

The key joining parameters of this process are the:

- DIELECTRIC LOSS FACTOR OF THE POLYMER OR THE PRESENCE OF A DIPOLE, BECAUSE A HIGH DIELECTRIC LOSS FACTOR INCREASES THE EFFECTIVENESS OF THE JOINING OPERATION
- THICKNESS OF THE MATERIAL BEING JOINED
- OUTPUT POWER AND FREQUENCY OF THE GENERATOR EMPLOYED DURING WELDING

Applications. Dielectric welding is a fast, efficient joining method for sealing thin polymer films and sheet materials. The equipment can be readily automated. However, the process can only be used on materials that have a high dielectric loss factor, such as rigid or flexible PVC, ABS, and thermoplastic polyurethane materials. The principal application area is in the packaging industry, where the process is used to join medical components, such as blood-pressure cuffs, medical bags, rainwear, pool liners, and so on.

The primary disadvantage of this joining technique is that the initial equipment cost is high.

Microwave Joining

This joining technique depends on the heat produced by microwave-frequency radiation. The parts being joined are held together under a slight pressure of 35 to 70 kPa (5 to 10 psi), and then the interface is irradiated by electromagnetic (EM) radiation using well-focused horn antennas. In order to absorb EM energy, the polymer must contain polar groups. To attain viable heating, either a nonconducting or a low-conductivity polymer may have to be doped with conducting polymer or, as an alternative, an adhesive film that contains polar and/or electron-withdrawing groups can be placed at the joint interface (Ref 48).

This joining method is relatively new and is still in the experimental stage. The experimental equipment (Ref 48) consists of a magnetron tube that generates the microwave energy, a ferrite isolator that shields the magnetron and promotes the efficient transmission of power with very little attenuation, a forward/reflected-power indicator that monitors the whole system while it is in operation, a tuner that can be adjusted to minimize the reflected power and aid in the maximum transfer of power from the source to the workpiece, a variable-coupling iris that optimizes and adjusts the power coupled to the load, and, finally, an applicator that concentrates the microwave power at the joint region.

The key joining parameters of this process are the:

- MICROWAVE PROPERTIES OF THE MATERIAL, SUCH AS PERMITTIVITY AND PERMEABILITY
- FREQUENCY AND ENERGY OF THE MICROWAVE SUPPLY

Applications. This welding technique is simple, fast, clean, and very efficient in terms of energy conversion. Consequently, short welding times are involved. Dissimilar materials can be joined, and the use of a microwave-absorbing adhesive will permit the repair of thermoset- or thermoplastic-based composite materials. This joining method can be very useful for in-situ repairs of structures, especially on space vehicles and other aircraft.

References cited in this section

5. R.A. GRIMM, FUSION WELDING TECHNIQUES FOR PLASTICS, *WELD. J.,* MARCH 1990, P 23-28
41. R.G. WILLIAMS, DEVELOPMENT OF A NOVEL ELECTROFUSION SYSTEM," *GAS ENGINEERING AND MANAGEMENT,* VOL 26, 1986, P 43-46
42. J.R. ATKINSON AND I.M. WARD, THE JOINING OF BIAXIALLY ORIENTED POLYETHYLENE PIPES, *POLYM. ENG. SCI.,* VOL 29 (NO. 23), 1989, P 1638-1641
43. SM. CHOOKAZIAN, BONDING PLASTICS BY INDUCTION HEATING, *SPE J.,* VOL 26, 1970, P 49-53
44. A. LEATHERMAN, INDUCTION BONDING FINDS A NICHE IN AN EVOLVING PLASTICS

INDUSTRY, *PLAST. ENG.,* VOL 4, 1981, P 27-29

45. S.P. MOLNAR, CHARACTERIZATION AND CONTROL OF INDUCTION FUSION BONDING OF THERMOPLASTIC COMPOSITE,*PROC. SPE 50TH ANTEC MEETING,* SOCIETY OF PLASTICS ENGINEERS, 1992, P 2102-2105
46. B.K. FINK, R.L. MCCULLOUGH, AND J.W. GILLESPIE JR., A LOCAL THEORY OF HEATING IN CROSS-PLY CARBON FIBER THERMOPLASTIC COMPOSITES BY MAGNETIC INDUCTION, *POLYM. ENG. SCI.,* VOL 32, 1992, P 357-369
47. C.J. GIANNANDREA, RADIO FREQUENCY SEALING, *MOD. PLAST.,* ENCYCLOPEDIA ISSUE, OCT 1990, P 386-388
48. V.K. VARADAN AND V.V. VARADAN, MICROWAVE JOINING AND REPAIR OF COMPOSITE MATERIALS, *POLYM. ENG. SCI.,* VOL 31, 1992, P 470-486

Welding of Plastics

Thomas H. North and Geetha Ramarathnam, University of Toronto

Evaluation of Welds

Because polymers are being used increasingly for semistructural and load-bearing applications, the testing of the welded joints is important. The mechanical properties of welded plastics can be evaluated by performing standard destructive tests, such as tensile, peel, wedge, and four-point bending. Of particular interest is the development of short-term tests (Ref 49), where a hole is notched in the weld area (to create multiaxial stress). These short-term tests are for quality-assurance purposes only.

For long-term applications, creep tests are much more favored. The measurement of thermal and residual stress has been evaluated using the Moire interferometric method (Ref 50).

The microstructure of the weld zone can be examined using light microscopy, scanning electron microscopy (SEM), or x-ray techniques (Ref 32, 51, 52). The microstructural examination of bonded regions is important because it allows:

- MEASUREMENT OF THE WELD ZONE AND HEAT-AFFECTED ZONE DIMENSIONS AND THEIR UNIFORMITY ALONG THE WELD LINE
- EVALUATION OF THE DEVELOPMENT OF CRYSTALLINITY, ESPECIALLY THE TRANSCRYSTALLINE REGION
- EVALUATION OF THE CHAIN ORIENTATION, ESPECIALLY PARALLEL TO THE BOND-LINE REGION
- ESTIMATION OF THE FAILURE MODES OF MECHANICALLY TESTED WELD SECTIONS
- OPTIMIZATION OF WELDING CONDITIONS IN ORDER TO CONSISTENTLY PRODUCE JOINTS OF OPTIMUM STRENGTH

Nondestructive evaluation (NDE) techniques for the detection of flaws in metal joints are well established. However, in the case of plastic joints, much work is still required to optimize the application of NDE techniques (Ref 53).

References cited in this section

32. K. KEUCHEL AND M. CAKMAK, SPIN WELDING OF POLYPROPYLENE: CHARACTERIZATION OF HEAT AFFECTED ZONE BY MICRO-BEAM WAXS TECHNIQUE, *PROC. SPE 49TH ANTEC MEETING,* SOCIETY OF PLASTICS ENGINEERS, 1991, P 2477-2481
49. G. MENGES, E. SCHMACHTENBERG, AND R. KRAUSENBERGER, EIN SCHNELLER WEG ZUR

ERMITTLUNG DES LANGZEITVERHALTENS VON SCHWEIßNAHTEN, *DVS--BER.,* VOL 84, P 62-65

50. J.B. PARK AND A. BENATAR, MOIRE INTERFEROMETRY MEASUREMENT OF RESIDUAL STRAINS IN IMPLANT RESISTANCE WELDING, *PROC. SPE 50TH ANTEC MEETING,* SOCIETY OF PLASTICS ENGINEERS, 1992, P 353-358
51. L.C. SAWYER AND D.T. GRUBB, IN *POLYMER MICROSCOPY,* CHAPMAN AND HALL, 1987
52. M.J. DAY AND M.F. GITTOS, "THE APPLICATION OF LIGHT MICROSCOPY TO WELDED JOINTS IN THERMOPLASTICS," REPORT 390, THE WELDING INSTITUTE, 1989
53. G.R. EDWARDS, NONDESTRUCTIVE EVALUATION OF WELDS IN PLASTICS, *PROC. 5TH ANNUAL NORTH AMERICAN WELDING RESEARCH CONFERENCE,* EWI AND AWS, 1989

Welding of Plastics

Thomas H. North and Geetha Ramarathnam, University of Toronto

---

References

General Review Articles on Joining and Welding of Polymers:

1. G. GEHARDSSON, THE WELDING OF PLASTICS, *WELD. REV.,* FEB 1983, P 17-22
2. M.N. WATSON, R.M. RIVETT, AND K.I. JOHNSON, "PLASTICS--AN INDUSTRIAL AND LITERATURE SURVEY OF JOINING TECHNIQUES," REPORT NO. 7846.01/85/471.3, THE WELDING INSTITUTE, ABINGTON, UK, 1986
3. H. POTENTE AND P. MICHEL, THE STATE OF THE ART DEVELOPMENT TRENDS IN THE WELDING OF PLASTICS, *PROC. 5TH ANNUAL NORTH AMERICAN WELDING RESEARCH CONFERENCE,* EWI AND AWS, 1989
4. V.K. STOKES, JOINING METHODS FOR PLASTICS AND PLASTIC COMPOSITES: AN OVERVIEW, *POLYM. ENG. SCI.,* VOL 29 (NO. 19), 1989, P 1310-1324
5. R.A. GRIMM, FUSION WELDING TECHNIQUES FOR PLASTICS, *WELD. J.,* MARCH 1990, P 23-28
6. G. MENGES, THE JOINING OF PLASTICS AND THEIR COMPOSITES, *PROC. LNT. CONF. ADVANCES IN JOINING NEWER STRUCTURAL MATERIALS,* IIW (MONTREAL), P 33-63

Adhesive Bonding:

7. A.J. KINLOCH, *ADHESION AND ADHESIVES,* CHAPMAN AND HALL, 1987, P 101

Mechanical Fastening:

8. D. CHANT, JOINING TECHNOLOGY FOR THERMOPLASTIC COMPOSITE STRUCTURES IN AEROSPACE APPLICATIONS, *PROC. INT. CONF. ADVANCES IN JOINING PLASTICS AND COMPOSITES,* THE WELDING INSTITUTE, CAMBRIDGE, 1991

Welding Mechanisms and Weldability of Thermoplastics:

9. S.S. VOYUTSKII, AUTOHESION AND ADHESION OF HIGH POLYMERS, WILEY INTERSCIENCE, 1963
10. A. BENATAR, MATERIAL CHARACTERISTICS FOR WELDING, *PROC. 5TH ANNUAL NORTH AMERICAN WELDING RESEARCH CONFERENCE,* EWI AND AWS, 1989
11. R.P. WOOL, B.-L. YUAN, AND O.J. MCGAREL, *POLYM. ENG. SCI.,* VOL 29 (NO. 19), 1989, P 1340-1367

Solvent Welding:

12. W.V. TITOW, CHAPTER 12, *SOLVENT WELDING OF PLASTICS,* APPLIED SCIENCE PUBLISHERS, LONDON, P 181-196

13. M. LICATA AND E. HAAG, SOLVENT WELDING WITH POLYCARBONATE, *PROC. SPE 44TH ANTEC MEETING,* SOCIETY OF PLASTICS ENGINEERS, 1986, P 1092-1094

Heated-Tool Welding:

14. K. GABLER AND H. POTENTE, WELDABILITY OF DISSIMILAR THERMOPLASTICS--EXPERIMENTS IN HEATED TOOL WELDING, *J. ADHES.,* VOL 11, 1980, P 145-163
15. H. POTENTE AND P. TAPPE, SCALE-UP LAWS IN HEATED TOOL BUTT WELDING OF HDPE AND PP, *POLYM. ENG. SCI.,* VOL 29 (NO. 23), 1989, P 1642-1648
16. H. POTENTE AND J. NATROP, COMPUTER AIDED OPTIMIZATION OF THE PARAMETERS OF HEATED TOOL BUTT WELDING, *POLYM. ENG. SCI.,* VOL 29 (NO. 23), 1989, P 1649-1654
17. A.J. POSLINSKI AND V.K. STOKES, ANALYSIS OF THE HOT-TOOL WELDING PROCESS, *PROC. SPE 50TH ANTEC MEETING,* 1992, SOCIETY OF PLASTICS ENGINEERS, P 1228-1233

Hot-Gas Welding:

18. H. GUMBLETON, HOT GAS WELDING OF THERMOPLASTICS--AN INTRODUCTION, *JOIN. MAT.,* VOL 5, 1989, P 215-218
19. "RECOMMENDED PRACTICES FOR JOINING PLASTIC PIPING," DOCUMENT XVI-322-78-E, IIW

Extrusion Welding:

20. P. MICHEL, "AN ANALYSIS OF THE EXTRUSION WELDING PROCESS," *POLYM. ENG. SCI.,* VOL 29 (NO. 19), 1989, P 1376-1382
21. M. GEHDE AND G.W. EHRENSTEIN, STRUCTURAL AND MECHANICAL PROPERTIES OF OPTIMIZED EXTRUSION WELDS, *POLYM. ENG. SCI.,* VOL 31 (NO. 7), 1991, P 495-501

Infrared Welding:

22. H. SWARTZ AND J.L. SWARTZ, FOCUSSED INFRARED--A NEW JOINING TECHNOLOGY FOR HIGH PERFORMANCE THERMOPLASTICS AND COMPOSITE PARTS, *PROC. 5TH ANNUAL NORTH AMERICAN WELDING RESEARCH CONFERENCE,* EWI AND AWS, 1989
23. H. POTENTE, P. MICHEL, AND M. HEIL, INFRARED RADIATION WELDING: A METHOD FOR WELDING HIGH TEMPERATURE RESISTANT THERMOPLASTICS, *PROC. SPE 49TH ANTEC MEETING,* SOCIETY OF PLASTICS ENGINEERS, 1991, P 2502-2504

Laser Welding:

24. W.W. DULEY, IN *LASER PROCESSING AND ANALYSIS OF MATERIALS,* PLENUM PRESS, 1983
25. W.W. DULEY AND R.E. MUELLER, CO_2 LASER WELDING OF POLYMERS, *POLYM. ENG. SCI.,* VOL 32 (NO. 9), 1992, P 582-585

Vibration Welding:

26. V.K. STOKES, VIBRATION WELDING OF THERMOPLASTICS, PART I, *POLYM. ENG. SCI.,* VOL 28, 1988, P 718-727
27. E. PECHA, D. CALINSKI, AND H. DIETER, PRODUCTION MACHINES AND TECHNICAL APPLICATION FOR HOT PLATE AND VIBRATION WELDING, *PROC. LNT. CONF. ADVANCES IN JOINING PLASTICS AND COMPOSITES,* THE WELDING INSTITUTE, CAMBRIDGE, 1991
28. V.K. STOKES, VIBRATION WELDING OF THERMOPLASTICS, PART II, *POLYM. ENG. SCI.,* VOL 28, 1988, P 727-739
29. H. POTENTE AND H. KAISER, PROCESS VARIANT OF VIBRATION WELDING WITH VARIABLE WELDING PRESSURE, *PROC. SPE 48TH ANTEC MEETING,* SOCIETY OF PLASTICS ENGINEERS, 1990, P 1762-1765
30. H. POTENTE AND M. UEBBING, COMPUTER AIDED LAYOUT OF THE VIBRATION WELDING PROCESS, *PROC. SPE 50TH ANTEC MEETING,* 1992, SOCIETY OF PLASTICS ENGINEERS, P 888

Spin Welding:

31. M. CAKMAK AND K. KEUCHEL, U.S. PATENT 4,998,663, MARCH 1991
32. K. KEUCHEL AND M. CAKMAK, SPIN WELDING OF POLYPROPYLENE: CHARACTERIZATION OF HEAT AFFECTED ZONE BY MICRO-BEAM WAXS TECHNIQUE, *PROC. SPE 49TH ANTEC MEETING,* SOCIETY OF PLASTICS ENGINEERS, 1991, P 2477-2481
33. H. RAJARAMAN AND M. CAKMAK, THE EFFECT OF GLASS FIBER FILLERS ON THE WELDING BEHAVIOR OF POLY P-PHENYLENE SULFIDE, *PROC. SPE 50TH ANTEC MEETING,* SOCIETY OF PLASTICS ENGINEERS, 1992, P 896-899

Ultrasonic Welding:

34. H. POTENTE, ULTRASONIC WELDING--PRINCIPLES AND THEORY, *MAT. DES.,* VOL 5, 1984
35. T.B. ZACH, J. LEW, T.H. NORTH, AND R.T. WOODHAMS, "JOINING OF HIGH STRENGTH ORIENTED POLYPROPYLENE USING ELECTROMAGNETIC INDUCTION BONDING AND ULTRASONIC WELDING, *MAT. SCI. TECH.,* VOL 5, 1989, P 281-287
36. J.S. WOLCOTT, DESIGNING PARTS FOR ULTRASONIC ASSEMBLY, *PROC. SPE 48TH ANTEC MEETING,* SOCIETY OF PLASTICS ENGINEERS, 1990, P 1829-1833
37. M.N. TOLUNAY, P.R. DAWSON, AND K.K. WANG, HEATING AND BONDING MECHANISMS IN ULTRASONIC WELDING OF THERMOPLASTICS, *POLYM. ENG. SCI.,* VOL 23, 1983, P 726-733
38. G. HABERNICHT AND J. RITTER, ENERGY CONVERSION IN THE ULTRASONIC WELDING OF THERMOPLASTICS, *KUNSTST. GER. PLAST.,* VOL 78, 1988, P 49-66
39. A. BENATAR AND T.G. GUTOWSKI, ULTRASONIC WELDING OF PEEK GRAPHITE APC-2 COMPOSITES, *POLYM. ENG. SCI.,* VOL 29 (NO. 23), 1989, P 1705-1721
40. N. TATEISHI, T.H. NORTH, AND R.T. WOODHAMS, ULTRASONIC WELDING USING TIE-LAYERS, PART I: ANALYSIS OF PROCESS OPERATION, *POLYM. ENG. SCI.,* VOL 32, 1992, P 600-611

Electrofusion (Resistance) Welding:

41. R.G. WILLIAMS, DEVELOPMENT OF A NOVEL ELECTROFUSION SYSTEM," *GAS ENGINEERING AND MANAGEMENT,* VOL 26, 1986, P 43-46
42. J.R. ATKINSON AND I.M. WARD, THE JOINING OF BIAXIALLY ORIENTED POLYETHYLENE PIPES, *POLYM. ENG. SCI.,* VOL 29 (NO. 23), 1989, P 1638-1641

Electromagnetic Induction Welding:

43. SM. CHOOKAZIAN, BONDING PLASTICS BY INDUCTION HEATING, *SPE J.,* VOL 26, 1970, P 49-53
44. A. LEATHERMAN, INDUCTION BONDING FINDS A NICHE IN AN EVOLVING PLASTICS INDUSTRY, *PLAST. ENG.,* VOL 4, 1981, P 27-29
45. S.P. MOLNAR, CHARACTERIZATION AND CONTROL OF INDUCTION FUSION BONDING OF THERMOPLASTIC COMPOSITE,*PROC. SPE 50TH ANTEC MEETING,* SOCIETY OF PLASTICS ENGINEERS, 1992, P 2102-2105
46. B.K. FINK, R.L. MCCULLOUGH, AND J.W. GILLESPIE JR., A LOCAL THEORY OF HEATING IN CROSS-PLY CARBON FIBER THERMOPLASTIC COMPOSITES BY MAGNETIC INDUCTION, *POLYM. ENG. SCI.,* VOL 32, 1992, P 357-369

Dielectric or Radio-Frequency Welding:

47. C.J. GIANNANDREA, RADIO FREQUENCY SEALING, *MOD. PLAST.,* ENCYCLOPEDIA ISSUE, OCT 1990, P 386-388

Microwave Welding:

48. V.K. VARADAN AND V.V. VARADAN, MICROWAVE JOINING AND REPAIR OF COMPOSITE MATERIALS, *POLYM. ENG. SCI.,* VOL 31, 1992, P 470-486

Evaluation of Plastic Welds:

49. G. MENGES, E. SCHMACHTENBERG, AND R. KRAUSENBERGER, EIN SCHNELLER WEG ZUR

ERMITTLUNG DES LANGZEITVERHALTENS VON SCHWEIßNAHTEN, *DVS--BER.*, VOL 84, P 62-65

50. J.B. PARK AND A. BENATAR, MOIRE INTERFEROMETRY MEASUREMENT OF RESIDUAL STRAINS IN IMPLANT RESISTANCE WELDING, *PROC. SPE 50TH ANTEC MEETING,* SOCIETY OF PLASTICS ENGINEERS, 1992, P 353-358
51. L.C. SAWYER AND D.T. GRUBB, IN *POLYMER MICROSCOPY,* CHAPMAN AND HALL, 1987
52. M.J. DAY AND M.F. GITTOS, "THE APPLICATION OF LIGHT MICROSCOPY TO WELDED JOINTS IN THERMOPLASTICS," REPORT 390, THE WELDING INSTITUTE, 1989
53. G.R. EDWARDS, NONDESTRUCTIVE EVALUATION OF WELDS IN PLASTICS, *PROC. 5TH ANNUAL NORTH AMERICAN WELDING RESEARCH CONFERENCE,* EWI AND AWS, 1989

Intelligent Automation for Joining Technology

Xiaoshu Xu and Jerald E. Jones, American Welding Institute

Introduction

AS JOINING TECHNOLOGY IMPROVES, industry is incorporating more types of automatic joining equipment and is expending significant efforts to make automated equipment more efficient. One of the most important approaches being used is called "intelligent automation." This approach combines automatic joining equipment, the knowledge of human experts in terms of joining, and a modem computer technique known as artificial intelligence (AI).

Artificial intelligence traditionally covers several application areas. AI techniques that can be applied to the intelligent automation of joining include expert systems, intelligent data-base systems, image processing, signal analysis, fuzzy-logic systems, and artificial neural networks.

An intelligent automation system usually includes an off-line planning system and a real-time adaptive control system that are connected through a computer communications interface. In the off-line planning system, the user develops a plan, with computer assistance, for the specific joining application, such as designing the joint, determining the process control parameters, or planning the robot path. The plan is then sent to the real-time control system through the interface. The real-time controller, using this down-loaded plan, controls the operation of the equipment. During execution, several sensors monitor the process and feed back information to the real-time controller. Using the feedback information, the real-time adaptive controller will adjust the robot trajectory, process parameters, and other controllable variables to correct any run-time errors.

Intelligent automation systems that vary in complexity are being developed by several organizations, as of 1993. Less-complex systems should be released in the near future by welding robot vendors. Currently, an intelligent automation system for arc welding, called WELDEXCELL, is being developed for the U.S. Navy at the American Welding Institute (AWI).

A graphical overview of the WELDEXCELL system is shown in Fig. 1. This system, which is relatively complex, utilizes most of the intelligent system techniques. For this reason, and because the authors are very familiar with it, WELDEXCELL is used for the examples described in this article. Although the discussion concentrates on the application of intelligent automation to arc welding processes, the technique also can be used for other joining processes.

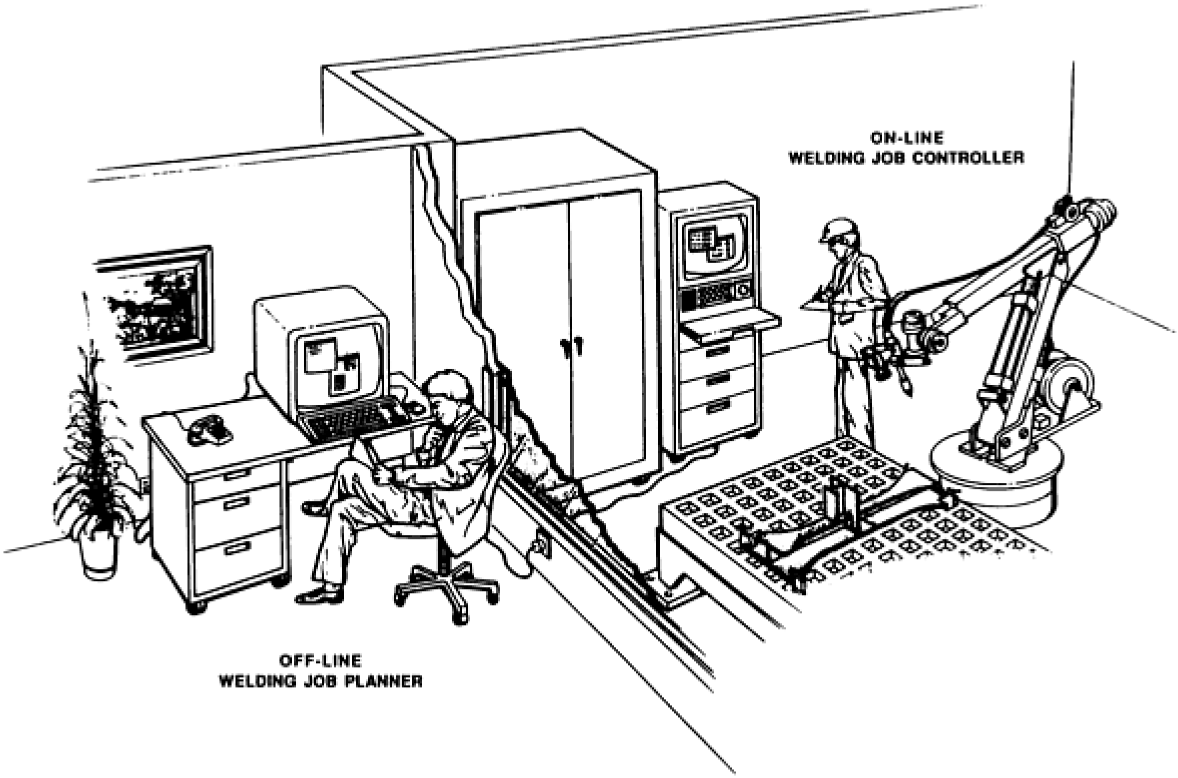
ON-LINE
WELDING JOB CONTROLLER
OFF-LINE
WELDING JOB PLANNER

FIG. 1 INTELLIGENT AUTOMATED WELDING SYSTEM

Intelligent Automation for Joining Technology

Xiaoshu Xu and Jerald E. Jones, American Welding Institute

Off-Line Planning System

An intelligent off-line planning system comprises several modules. WELDEXCELL, for example, contains four: the parts designer, path planner, welding schedule developer, and system integrator (Fig. 2).

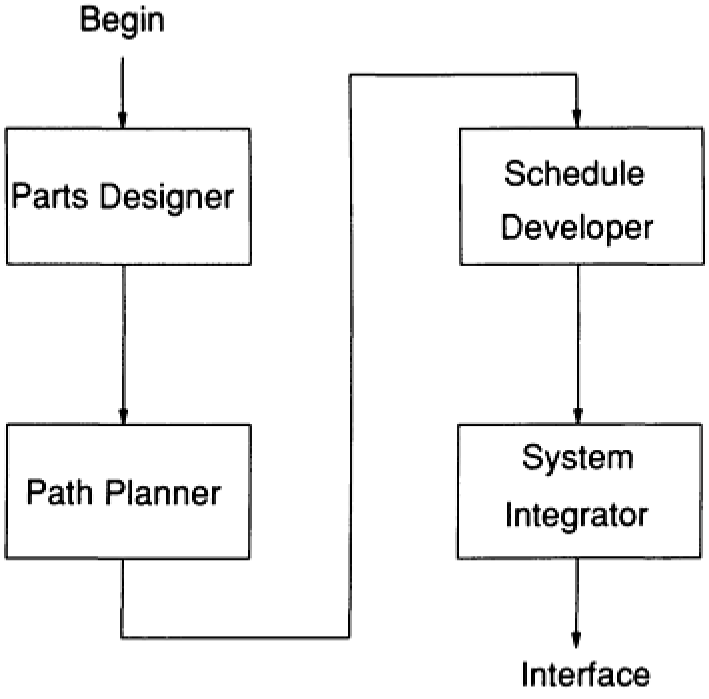

FIG. 2 SEQUENTIAL OPERATION OF AN OFF-LINE PLANNING SYSTEM

Parts Designer. The parts to be joined should first be drawn in a three-dimensional (3-D) computer-aided design (CAD) system. Many suitable commercial CAD systems are available. For a CAD system to be used in an off-line planning system, the required functions include:

- DRAWING FUNCTIONS THAT DRAW EITHER 3-D OR PSEUDO 3-D OBJECTS, INCLUDING LINES, CIRCLES, ARCS, POLYGONS, AND ELLIPSES
- EDIT FUNCTIONS THAT ERASE, MOVE, OR COPY DRAWN OBJECTS
- A FUNCTION THAT CHANGES THE COORDINATE SYSTEM, INCLUDING THE ORIGIN AND THE VIEW POINT (IMPORTANT FOR THE USER INTERFACE, AS WELL AS FOR LATER ROBOT GUIDANCE)
- OUTPUT DRAWING FILES IN ONE OF THE STANDARD VECTOR GRAPHICS REPRESENTATION FORMATS

In the WELDEXCELL system, the user superimposes the weld segments on a separate layer of the CAD part representation. A weld segment is a portion of a weld, and a series of one or more weld segments will produce a single weld on the part. It can be a single- or a multiple-pass weld. However, it has to satisfy the following two restrictions. First, the welding segment must contain a continuous trajectory, which can be a straight line, polyline, circle, arc, or spline in 3-D space. Second, the welding process, position, and all other non-real-time adjustable parameters, such as filler-metal diameter or type of shielding gas, must be kept constant during each weld pass when welding the defined segment.

In the parts designer module, the user is able to set any required welding restrictions for any welding segment. After the parts are designed and the weld segments are defined, they are stored by the CAD system in a vector-based format. This allows the later retrieval of any points that are located on the parts, especially the weld segments and associated restrictions.

- *THE PATH PLANNER* IS AN INTELLIGENT, OFF-LINE, WELD PLANNING SYSTEM COMPRISING TWO PARTS: PATH GENERATION AND ROBOT SIMULATION. FIRST, THE INITIAL ROBOT WELDING PATH IS DETERMINED. THIS PATH INCLUDES THE

TRAJECTORY OF THE ROBOT, AS WELL AS THE MOVEMENT OF A POSITIONING TABLE (IF APPLIED). THE PATH CONTAINS THREE MAJOR SECTIONS.

The start section represents the path from the home (or start) position of the robot to the beginning of the first weld segment.

The welding section, which is the most important section of the path plan, includes the path for each weld segment and the transition path between segments and between welding passes for each segment (in the case of a multiple-pass weld). Because the number of passes and the offset of each pass from the joint centerline for each segment are not known until the welding schedules are later developed, the initial path plan does not take multiple passes into consideration.

The end section represents the path from the end of the last weld segment back to the robot home position.

After the initial path is generated, it can then be sent electronically to the robot simulation model for execution. The simulation environment on the computer should include an animated robot and its environment. Objects that the robot could collide with, such as the welding torch, positioning table, parts to be welded, fixtures to secure the part, and others, should be included in the robot simulation environment (Fig. 3). The simulation will run through the designated path to ensure that no collision occurs between the robot arm and any environmental object and that the path is smooth and satisfies the requirements of the user. The simulation should provide a full 3-D or pseudo 3-D representation and should allow viewing from various positions. Them may be a system inside the simulation that can automatically correct or even optimize the robot trajectory to avoid any possible collision.

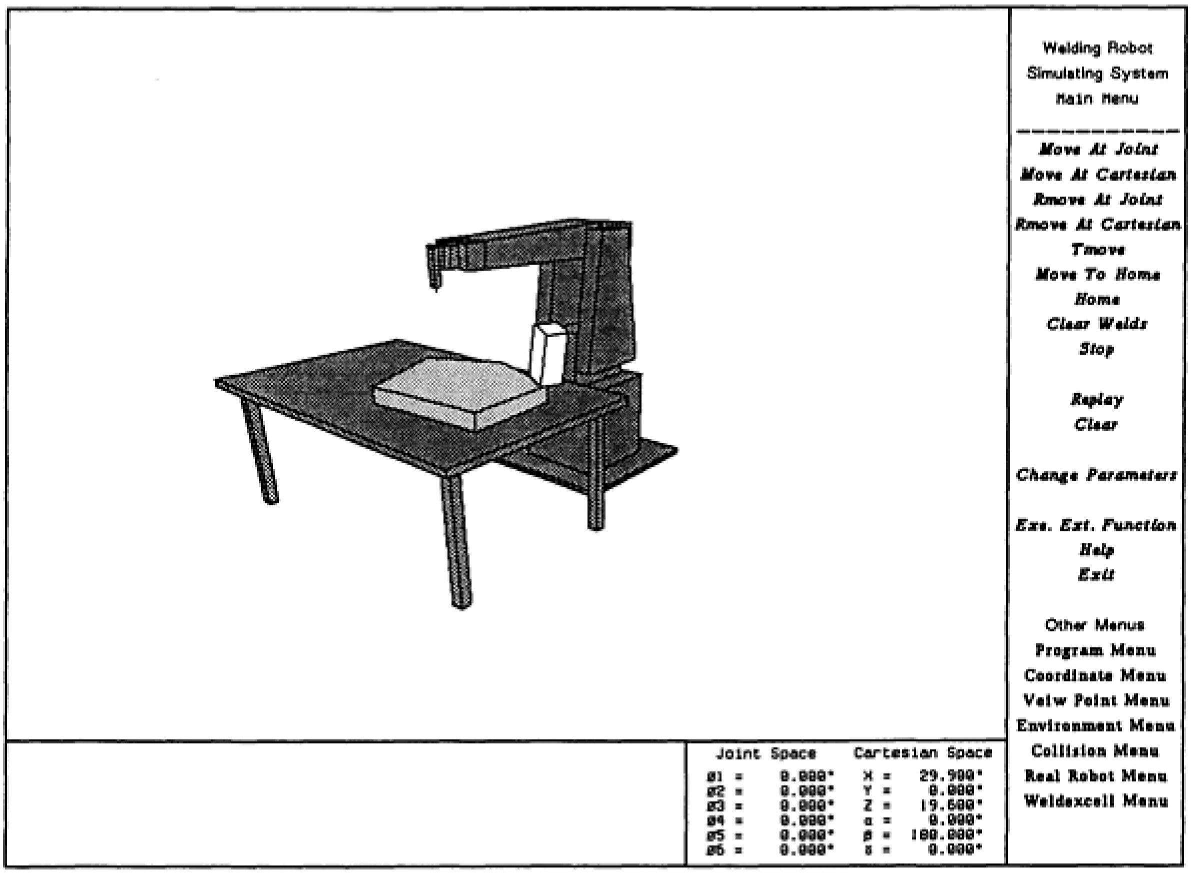
Welding Robot
Simulating System
Main Menu
Move At Joint
Move At Cartesian
Rmove At Joint
Rmove At Cartesian
Tmove
Move To Home
Home
Clear Welds
Stop
Replay
Clear
Change Parameters
Exe. Ext. Function
Help
Exit
Other Menus
Program Menu
Coordinate Menu
Veiw Point Menu
Environment Menu
Collision Menu
Real Robot Menu
Weldexcell Menu

```
  Joint Space         Cartesian Space
Ø1 =      0.000°    X =    29.900°
Ø2 =      0.000°    Y =     0.000°
Ø3 =      0.000°    Z =    19.600°
Ø4 =      0.000°    α =     0.000°
Ø5 =      0.000°    β =   180.000°
Ø6 =      0.000°    γ =     0.000°
```

FIG. 3 PATH PLANNER AND ROBOT SIMULATOR IN THE WELDEXCELL SYSTEM

Based on the results of this simulation, either the system or the user can modify the path by using any extra degrees-of-freedom of the robot and/or by rotating the positioning table to adjust the trajectory. If necessary, the welding position can be changed for some weld segments. If the welding position has been restricted by the user, then input from the user is needed to modify that portion of the path plan. This modified path can then be sent to the robot simulation to run again. Several iterations of this cycle may be required before a satisfactory path plan can be attained. Note that the path generated here is not complete. Multiple-pass welding segments have not been planned, nor has the velocity been set. Final path planning will occur in the system integrator module.

Schedule Developer. A welding schedule is developed for each individual welding segment. All welding variables, except for welding position (determined in the path planner module), are determined by the schedule developer (in combination with user input). These variables include welding process, number of welding passes, consumable type and size, welding current, voltage, wire feed rate, travel speed, shielding gas type and flow rate, required heat treatment, and others.

Several modules in the WELDEXCELL system are used to develop a welding schedule for each segment. These modules, called knowledge sources, will determine the individual variables. The knowledge sources include expert systems, neural networks, and CAD systems. A distributed data base system that contains several data bases is required to support these knowledge sources.

Data Base Systems. Several data bases should be available to the off-line planner. Data can be directly used by either the human user, to develop the welding schedule, or the intelligent off-line planner, in which case the data are available to the other knowledge sources. The data base system should include:

- A BASE-MATERIAL DATA BASE, WHICH STORES THE DESIGNATION AND THE CHEMICAL, THERMOPHYSICAL, AND MECHANICAL PROPERTIES OF THE BASE MATERIALS
- A FILLER-MATERIAL DATA BASE, WHICH STORES THE DESIGNATIONS, CHEMICAL PROPERTIES, AND MECHANICAL PROPERTIES OF THE FILLER MATERIALS
- A JOINT DATA BASE, WHICH STORES JOINT DESIGNS THAT ARE STANDARD (SUCH AS THOSE FROM THE AWS/ANSI STRUCTURAL WELDING CODE D 1.1) AND NONSTANDARD (USER), AS WELL AS DIMENSION INFORMATION, INCLUDING GRAPHICAL (CAD) REPRESENTATIONS
- A WELDING PROCEDURE SPECIFICATION (WPS) DATA BASE (REFER TO REF 1 FOR AN EXPLANATION OF WPS)
- A PROCEDURE QUALIFICATION RECORD (PQR) DATA BASE (REFER TO REF 1)
- A WELD-SCHEDULE DATA BASE, WHICH MAINTAINS THE WELD SCHEDULES ONCE THEY ARE DEVELOPED

An expert system consists of a "knowledge base," which stores information on a subject from a human expert, and an "inference engine," which contains a set of logic processes that is useful in applying the information in the knowledge base. The separation of the knowledge from the logic processes is one way in which expert systems differ from conventional software.

In the field of artificial intelligence, several methodologies have been developed for the representation of knowledge. Some of these have been used to develop computerized applications. Two of the methodologies, specifically "frames" and "production rules," have been successfully used to develop expert systems.

In this type of expert system, the knowledge of the human expert is stored in the production rules. These rules are written in the form of: if (A), then (C); where (A) is known as the antecedent and (C) is known as the consequent. For organizational purposes, and to aid in the application of the logic processes, the rules are then grouped together in a frame structure that is hierarchical in nature.

The expert system operates by selectively choosing rules that generate conclusions or formulate parts of answers to questions. This process is accomplished by the inference engine as the antecedent of a rule is shown to be true by either user input or data derived from a data base for a specific situation.

Sometimes, these rules are also associated with certainties. This gives the final result, or a set of results from an expert system operation, a certainty value. A detailed discussion of expert systems can be found in Ref 2 and 3. As the logic system applies specific rules to generate conclusions, those rules are "chained" together. If the system is designed to form a conclusion, then "backward" chaining is used (that is, working backward from a possible conclusion to show whether it is true or false). Forward chaining is applied if the purpose is to synthesize information, such as to generate a welding procedure.

In a schedule developer, the expert system usually includes such tasks as: select suitable filler-metal type and size, determine preheat and postweld heat-treatment schedule, determine welding parameters, and others.

The WELDEXCELL system, for example, contains several expert systems in the area of arc welding, which can be used in the off-line planning activity. Examples of these include:

- WELDSELECTOR: AN EXPERT SYSTEM TO HELP SELECT AN ELECTRODE (FILLER MATERIAL) AND ELECTRODE DIAMETER. THIS SYSTEM HAS A CERTAINTY-FACTOR-BASED, BACKWARD CHAINING, INFERENCE ENGINE AND GIVES A RANKED LIST OF ELECTRODES FOR A GIVEN APPLICATION. THIS WAS THE FIRST PERSONAL-COMPUTER-BASED EXPERT SYSTEM COMMERCIALLY AVAILABLE IN THE UNITED STATES.
- FERRITEPREDICTOR: AN EXPERT SYSTEM THAT PERFORMS THE FERRITE ANALYSIS FOR STAINLESS STEEL WELDING. FOUR CONSTITUTION DIAGRAMS CAN BE CONSULTED: SCHAEFFLER, DELONG, AND THE 1988 AND 1992 WELDING RESEARCH COUNCIL (WRC) DIAGRAMS. FIGURE 4(A) SHOWS THE GENERAL USER INTERFACE, WHEREAS FIG. 4(B) SHOWS THE WRC DIAGRAM ANALYSIS.
- WELDHEAT: AN EXPERT SYSTEM THAT DETERMINES THE PREHEAT, INTERPASS TEMPERATURE, AND POSTWELD HEAT TREATMENT. THE WELDHEAT EXPERT SYSTEM USES FOUR DIFFERENT METHODS: INTERNATIONAL INSTITUTE OF WELDING (IIW) CARBON EQUIVALENT, WRC PREHEAT AND POSTWELD HEAT TREATMENT TABLE (FROM REF 4), AND A MATHEMATICAL MODEL PREDICTION BASED ON THE WORK DONE BY N. YURIOKA AND M. OKUMURA (REF 5, 6, 7).
- WELDPROSPEC: A SYSTEM THAT INTELLIGENTLY SEARCHES THE PQR DATA BASE, BASED ON A SPECIFIED WELDING CODE, SUCH AS THE AWS/ANSI STRUCTURAL WELDING CODE D1.1. THE WELDPROSPEC SEARCHES THE PQR(S) IN THE PQR DATA BASE THAT CAN QUALIFY THE CURRENT WELD SCHEDULE. IF NECESSARY, WELDPROSPEC CAN CREATE A WPS FROM SELECTED PQR(S) TO SUPPORT THE CURRENT WELD SCHEDULE.

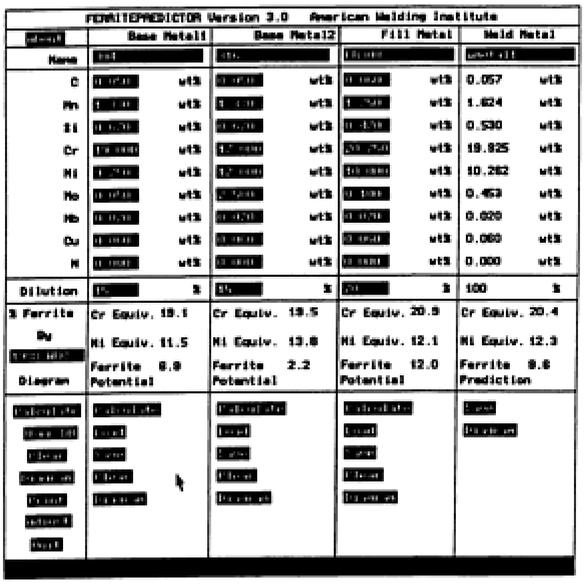

FERRITEPREDICTOR Version 3.0 American Welding Institute

[illegible]	Base Metal1	Base Metal2	Fill Metal	Weld Metal
Name	[illegible]	[illegible]	[illegible]	[illegible]
C	[illegible] wt%	[illegible] wt%	[illegible] wt%	0.057 wt%
Mn	[illegible] wt%	[illegible] wt%	[illegible] wt%	1.624 wt%
Si	[illegible] wt%	[illegible] wt%	[illegible] wt%	0.530 wt%
Cr	[illegible] wt%	[illegible] wt%	[illegible] wt%	18.925 wt%
Ni	[illegible] wt%	[illegible] wt%	[illegible] wt%	10.282 wt%
Mo	[illegible] wt%	[illegible] wt%	[illegible] wt%	0.453 wt%
Nb	[illegible] wt%	[illegible] wt%	[illegible] wt%	0.020 wt%
Cu	[illegible] wt%	[illegible] wt%	[illegible] wt%	0.080 wt%
N	[illegible] wt%	[illegible] wt%	[illegible] wt%	0.000 wt%
Dilution	[illegible] %	[illegible] %	[illegible] %	100 %
% Ferrite By [illegible] Diagram	Cr Equiv. 19.1 Ni Equiv. 11.5 Ferrite Potential 8.9	Cr Equiv. 19.5 Ni Equiv. 13.8 Ferrite Potential 2.2	Cr Equiv. 20.9 Ni Equiv. 12.1 Ferrite Potential 12.0	Cr Equiv. 20.4 Ni Equiv. 12.3 Ferrite Prediction 8.8

(a)

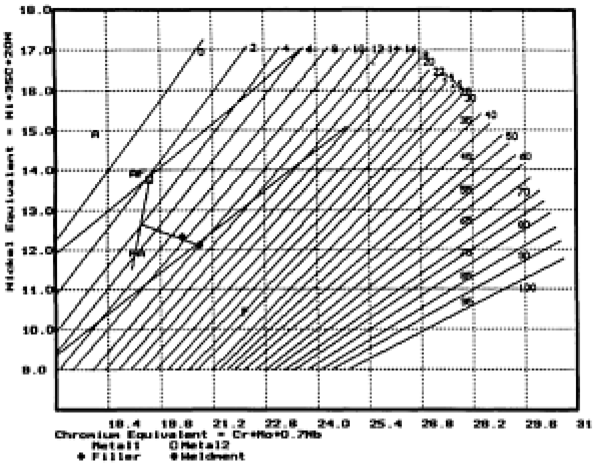

(b)

FIG. 4 EXAMPLES OF EXPERT SYSTEM SCREEN DISPLAYS. (A) USER INTERFACE OF THE FERRITEPREDICTOR EXPERT SYSTEM FOR STAINLESS STEEL FERRITE ANALYSIS. (B) A FERRITE ANALYSIS OF A DISSIMILAR METAL STAINLESS STEEL WELD (BASE METALS, 304 AND 316; FILLER METAL, ER308) USING THE 1988 WRC DIAGRAM IN THE FERRITEPREDICTOR EXPERT SYSTEM

A *neural network system* (NNS) is a computer hardware and software technique that simulates an interconnected network of biological neurons, similar to a portion of the human brain. A neural network can be trained with given patterns, called training sets. After the NNS "learns" from the training set, the network can retrieve these patterns and can make generalizations for patterns that are not included in the training set. Neural networks can be used in associative memory, object or pattern classification, process modeling and control, and signal processing. A detailed discussion of neural networks can be found in Ref 8 and 9.

The important NNSs contained in the WELDEXCELL off-line planner are WELDBEAD and HEATFLOW. WELDBEAD, a neural network model of the welding process, predicts bead geometry, such as the bead width, height, and penetration, for a given set of welding parameters and a specific joint configuration. Figure 5 shows the user interface for the WELDBEAD software. A more detailed discussion of the WELDBEAD system can be found in Ref 10.

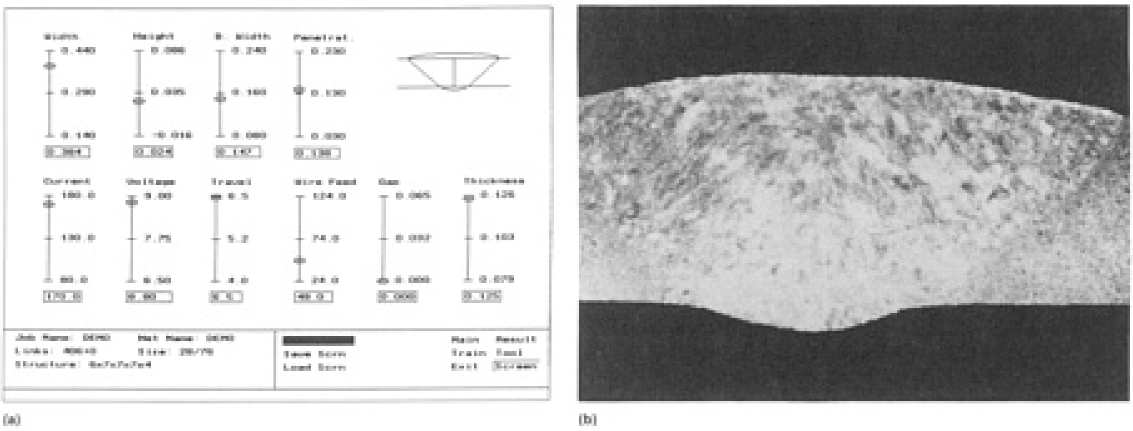

(a) (b)

FIG. 5 NEURAL NETWORK SYSTEM FOR AUTOMATED WELDING. (A) USER INTERFACE OF WELDBEAD NEURAL NETWORK WELD MODEL FOR A GTAW WELD IN THIN-SECTION STAINLESS STEEL. (B) MACROGRAPH OF THE RESULTING TEST WELD CORRESPONDING TO THE PREDICTION BY THE WELDBEAD NEURAL NETWORK, SHOWN IN (A)

The HEATFLOW system is used to predict the temperature distribution in the weld. This neural network is a model of the heat flow during welding. When given the voltage, current, wire speed, and travel speed, it can predict the size and shape of the liquid weld pool, as well as provide a map of isothermal lines. This system produces information that is helpful in selecting the proper preheat and postweld heat treatment schedule. Figure 6 shows the WELDHEAT user interface. Reference 11 provides more information.

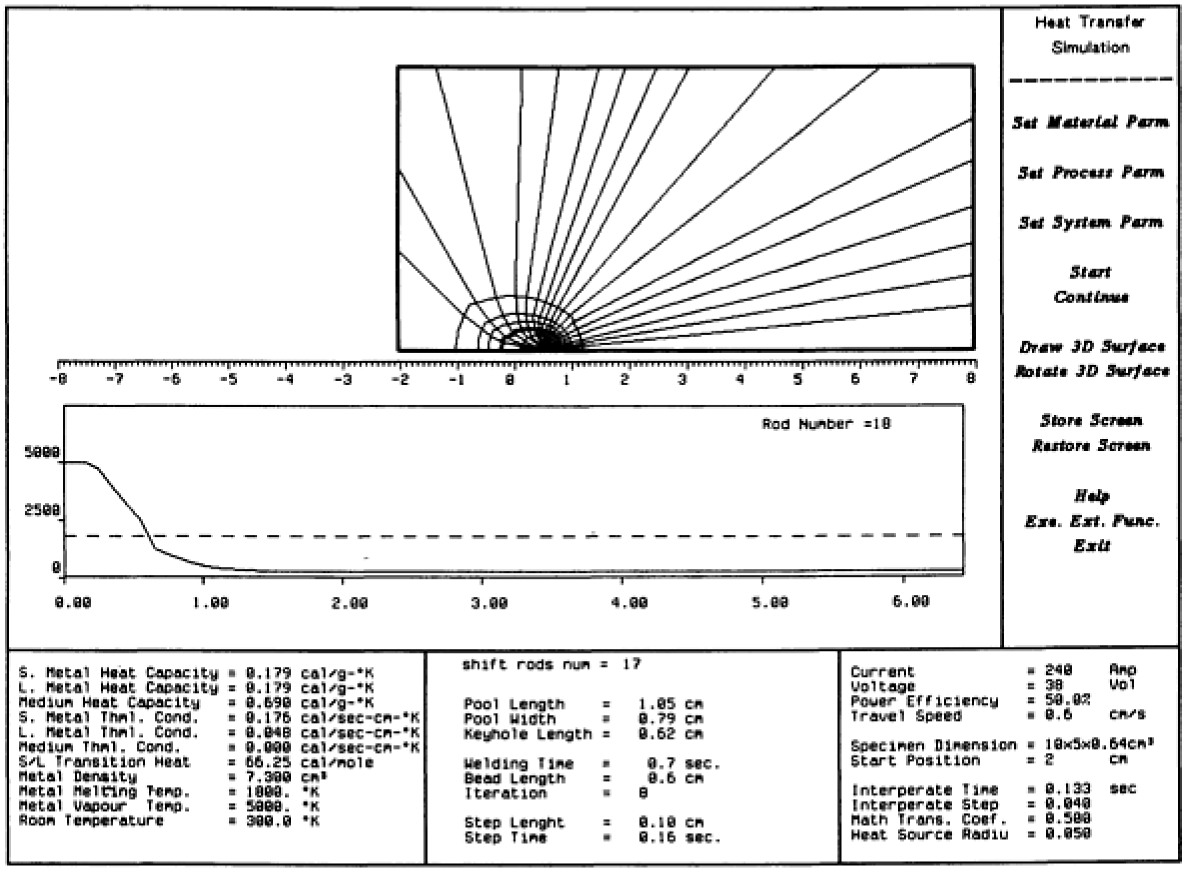
Heat Transfer
Simulation
Set Material Parm
Set Process Parm
Set System Parm
Start
Continue
Draw 3D Surface
Rotate 3D Surface
Store Screen
Restore Screen
Help
Exe. Ext. Func.
Exit
-8
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8
Rod Number =10
5000
2500
0
0.00
1.00
2.00
3.00
4.00
5.00
6.00

```
S. Metal Heat Capacity = 0.179 cal/g-*K
L. Metal Heat Capacity = 0.179 cal/g-*K
Medium Heat Capacity   = 0.690 cal/g-*K
S. Metal Thml. Cond.   = 0.176 cal/sec-cm-*K
L. Metal Thml. Cond.   = 0.040 cal/sec-cm-*K
Medium Thml. Cond.     = 0.000 cal/sec-cm-*K
S/L Transition Heat    = 66.25 cal/mole
Metal Density          = 7.300 cm³
Metal Melting Temp.    = 1800. *K
Metal Vapour  Temp.    = 5000. *K
Room Temperature       = 300.0 *K
```

```
shift rods num =  17

Pool Length     =    1.05 cm
Pool Width      =    0.79 cm
Keyhole Length =     0.62 cm

Welding Time    =     0.7 sec.
Bead Length     =     0.6 cm
Iteration       =    0

Step Lenght     =    0.10 cm
Step Time       =    0.16 sec.
```

```
Current             = 240    Amp
Voltage             = 30     Vol
Power Efficiency    = 50.0%
Travel Speed        = 0.6    cm/s

Specimen Dimension  = 10x5x0.64cm³
Start Position      = 2      cm

Interperate Time    = 0.133  sec
Interperate Step    = 0.040
Math Trans. Coef.   = 0.500
Heat Source Radiu   = 0.050
```

FIG. 6 WELDHEAT PROGRAM USER INTERFACE

In a complex computer software system, such as intelligent welding off-line planning, supervisory software is needed to coordinate the system. One type of supervisory system, known as a "blackboard," was first developed in the 1980s (Ref 12, 13). The computer blackboard architecture has proven to be an efficient system for controlling multiple-knowledge sources, including expert systems, data base systems, neural network systems, and human users. The blackboard integrates and organizes the actions of the various software systems. Often, a blackboard will have an expert system as the principal control element. As an example, the structure of the WELDEXCELL schedule developer is shown in Fig. 7 with the blackboard as the central controller.

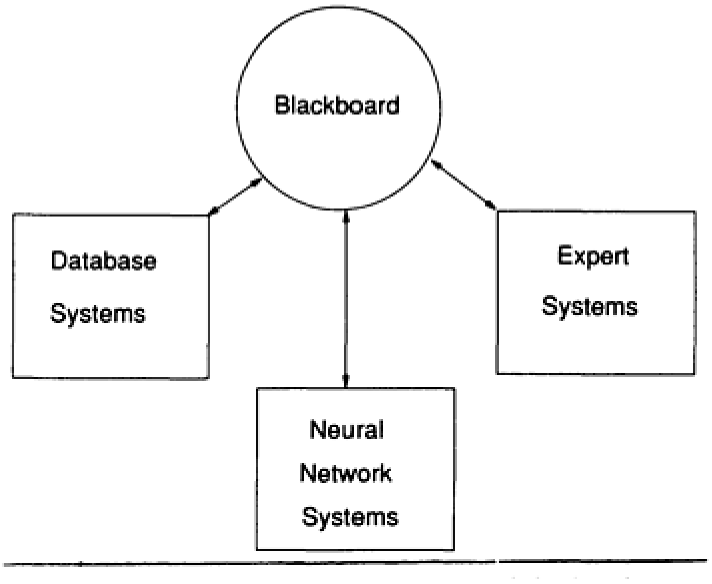

FIG. 7 OPERATION OF THE WELDING SCHEDULE DEVELOPER FROM THE WELDEXCELL SYSTEM

System Integrator. Once an initial weld path has been established and the welding schedule determined, the two separate sets of information must be integrated. Then, the automated welding system will be able to control the process and the torch motion simultaneously to produce a weld.

The first task of this module is to develop a complete robot path. Using the previously developed initial path, this module expands the welding portion to include all passes of a multiple-pass weld. Torch velocity is also assigned to each segment based on the travel speed specified in the schedule developer model. The final welding instruction (path and process control information) can then be sent to the robot simulation, where final modifications can be made by the human user, based on the performance of the simulation execution.

The other task of this module is to combine all of the information, such as pass number(s) and welding parameters, into a set of interface files. These interface files can then be sent to the real-time control system for execution.

References cited in this section

1. CHAPTER 14, *WELDING HANDBOOK,* 8TH ED., VOL 1, AWS, 1987
2. M.Y. DEMDERI, "EXPERT SYSTEM APPLICATION IN MATERIAL PROCESSING AND MANUFACTURING," FORD MOTOR COMPANY, OCT 1989
3. J. JONES, D. WHITE, X. XU, *ET AL.,* DEVELOPMENT OF AN OFF-LINE WELD PLANNING SYSTEM, *TEXAS INSTRUMENTS TECH. J.,* WINTER 1987, P 47-53
4. R.D. STOUT, *WELDABILITY OF STEELS,* 4TH ED., WELDING RESEARCH COUNCIL, 1987
5. H. SUZUKI, N. YURIOKA, AND M. OKUMURA, A NEW CRACKING PARAMETER FOR WELDED STEELS CONSIDERING LOCAL ACCUMULATION OF HYDROGEN, *WELD. WORLD,* VOL 21, 1983, P 2-15
6. H. SUZUKI AND N. YURIOKA, PREVENTION AGAINST COLD CRACKING BY THE HYDROGEN ACCUMULATION CRACKING PARAMETER PHA, *TRANS. JPN. WELD. SOC.,* VOL 14 (NO. 1), APRIL 1983, P 2-14
7. N. YURIOKA, M. OKUMURA, T. KASUYA, AND S. OHSHITA, *WELDING NOTE,* 2ND ED., NIPPON STEEL R&D LABORATORIES
8. D.E. RUMELHART, J.L. MCCLELLAND, AND THE PDP RESEARCH GROUP, *PARALLEL DISTRIBUTED PROCESSING,* MIT PRESS, 1988

9. C. LAU, ED., *NEURAL NETWORKS,* IEEE PRESS, 1992
10. A. ROCK, X. XU, AND J. JONES, "TYPICAL NEURAL NETWORK APPLICATIONS IN SIGNAL PROCESSING AND PROCESS MODELLING," PAPER NO. 256, CORROSION 92, NACE
11. X. XU AND J. JONES, "A COMPUTER MODEL OF HEAT FLOW IN ARC WELDING," PAPER PRESENTED AT 1991 AWS CONFERENCE (DETROIT, MI), AWS
12. H.P. NII, BLACKBOARD SYSTEMS: THE BLACKBOARD MODEL OF PROBLEMS SOLVING AND THE EVOLUTION OF BLACKBOARD ARCHITECTURES, *THE AI MAGAZINE,* SUMMER 1986, P 28-53
13. X. XU, J. JONES, *ET AL.,* INVESTIGATION OF A BLACKBOARD ARTIFICIAL INTELLIGENCE COMPUTER ARCHITECTURE FOR WELDING PROCEDURE SPECIFICATION AND STRUCTURE INTEGRITY ANALYSIS, PROC. *2ND INT. CONF. ON TRENDS IN WELDING RESEARCH,* ASM INTERNATIONAL 1990, P 985-989

Intelligent Automation for Joining Technology

Xiaoshu Xu and Jerald E. Jones, American Welding Institute

Interface Between Off-Line Planners and Real-Time Control Systems

The interface is the communication link between the off-line planner and the real-time control system. This interface is based on computer files, which include the:

- *JOB DESCRIPTION FILE* (*JDF*). THIS FILE CONTAINS THE NAMES OF THE OTHER FILES THAT ARE ASSOCIATED WITH A SPECIFIC WELD JOB.
- *PART SEGMENT FILE* (*PSF*). THIS FILE PROVIDES A DESCRIPTION OF EACH PART SEGMENT TO WELD.
- *JOB TASK FILE* (*JTF*). THIS FILE CONTAINS A SEQUENCE OF TASKS DESCRIBING THE EXECUTION SEQUENCE OF A WELD JOB.
- *WELD SET FILE* (*WSF*). THIS FILE CONTAINS ALL OF THE WELD SETS TO BE USED FOR THE JOB. (WELD SETS ARE DESCRIBED BELOW.)
- *WELD PASS FILE* (*WPF*). THIS FILE PROVIDES INFORMATION FOR CONTROLLING THE MOVEMENTS OF THE ROBOT ALONG EACH PASS.
- *ROB TRAJECTORY FILE* (*RTF*). THIS FILE DEFINES THE ROBOT TRAJECTORY AND POSITIONER MOVEMENT.
- *PART SETUP FILE* (*PSF*).THIS FILE CONTAINS INFORMATION FOR POSITIONING THE PART TO BE WELDED ON THE PART POSITION TABLE, AND FOR LOCATING THE PART RELATIVE TO THE WORLD COORDINATE FRAME OPERATOR.
- *JOB PICTURE FILE* (*JPF*). THIS FILE CONTAINS PIXEL-MAP IMAGES SHOWING THE CROSS-SECTIONAL VIEW OF ALL WELDED JOINTS.
- *JOB NOTES FILE* (*JNF*). THE JNF CONTAINS NOTES OR COMMENTS ABOUT THE WELDING PROCEDURE AND PROVIDES FOR COMMUNICATION BETWEEN THE ENGINEER AND THE WELDING SYSTEM OPERATOR.
- *JOB INTERMEDIATE FILE* (*JIF*). THIS FILE REFLECTS THE JOB STATE AND CONTAINS ANY CHANGES DETERMINED BY THE OPERATOR IN THE SETUP PROCEDURE.
- *JOB RESULTS FILE* (*JRF*). THIS FILE REFLECTS THE FINAL JOB STATE, AS IT WAS ACTUALLY EXECUTED.

This set of files has been proposed to the American Welding Society (AWS) A9 Committee as part of a standard (A9.3) for such interfaces.

The JDF, JTF, PSF, WSF, WPF, RTF, PSF, JPF, and JNF are generated by the off-line planner and sent to the real-time controller. The JIF and JRF store information retrieved during weld job execution from the real-time controller. These files are then sent back to the off-line planner in order to update it and to provide additional training examples for the neural network systems. Many automated welding systems that are computer controlled utilize a "weld set."

- THE FOLLOWING INFORMATION SHOULD HELP TO CLARIFY THE DESCRIPTION OF THE INTERFACE:
- *WELD SET.* A WELD SET IS A COLLECTION OF WELDING PARAMETERS THAT INCLUDES CURRENT, VOLTAGE, TRAVEL SPEED, WIRE FEED SPEED, AND WELDING TORCH ANGLE. WELD SETS ARE IDENTIFIED BY A DESIGNATOR SUCH AS A WELD SET NUMBER (WSN).
- *TRAJECTORY.* A TRAJECTORY IS A CONTINUOUS MOVEMENT OF THE ROBOT WITH THE FOLLOWING RESTRICTIONS: FIRST, A TRAJECTORY HAS EITHER ONE WELD PATH ASSOCIATED WITH IT OR NO PATH, AS IN THE CASE OF A TRANSITION FROM ONE WELD SEGMENT TO AN ADJACENT WELD SEGMENT. SECOND, A TRAJECTORY HAS ONE MOVE PATTERN. THE MOVE PATTERN CAN BE A 3-D STRAIGHT LINE, ARC, SPLINE, OR NONE. NONE, IN THIS CONTEXT, MEANS THAT NO SPECIFIC MOVE PATTERN IS REQUIRED. FOR ROBOTS OF THE PRESENT TIME, THE MOVING VELOCITY HAS TO BE CONSTANT. IT IS LIKELY THAT THIS WILL NOT BE TRUE OF FUTURE ROBOT SYSTEMS. A TRAJECTORY IS IDENTIFIED BY AN IDENTIFIER, SUCH AS A TRAJECTORY NUMBER. BASED ON THIS DEFINITION, EACH PASS IS ASSOCIATED WITH ONE UNIQUE TRAJECTORY.
- *WELD JOB.* A WELD JOB CAN HAVE SEVERAL WELD SCHEDULES. THIS MEANS THAT IN THE EXECUTION OF A WELD JOB, THE WELDING PARAMETERS IDENTIFIED ABOVE CAN BE VARIED. PARAMETERS THAT CANNOT BE ADJUSTED DURING WELDING, SUCH AS FILLER-METAL TYPE AND SIZE OR GAS TYPE AND FLOW RATE, REMAIN CONSTANT FOR ANY CONTINUOUS OPERATION. HOWEVER, THESE PARAMETERS MAY CHANGE FROM ONE PASS TO ANOTHER.

Intelligent Automation for Joining Technology

Xiaoshu Xu and Jerald E. Jones, American Welding Institute

Real-lime Intelligent Control System

The real-time controller for an intelligent automated welding system has several functions. First, it must control the motion of the welding torch. Second, the control of the welding process equipment must be coordinated with the location of the torch in the weld joint. Third, software and hardware are needed to process and interpret the signal data from one or more sensors. Fourth, the controller must have an adaptive process planning and control modification system for making adjustments to the control parameters based on the interpreted sensor data and process models. Finally, for process control and planning improvement, the real-time controller is a platform for the collection of data about the weld that can be uploaded electronically to the off-line planning system.

Standards

In order for the real-time controller to communicate with the welding process systems and with the off-line planner, it is important that standards be established for the storage and communication of computerized welding information. The AWS A9 and D16 Committees are developing such standards, and the reader is referred to AWS/ANSI A9.1-92 and A9.2-92, as well as the soon to be published standards, such as AWS/ANSI A9.3 and A9.4, and ASTM E49 Committee standards.

Real-Time Control System

The real-time control system comprises six major modules:

- THE INTERFACE TO A JOB DATA FILE OR AN OFF-LINE PLANNER
- THE JOB INTERPRETER
- THE JOB EDITOR AND USER INTERFACE
- THE HARDWARE CONTROLLER OF THE WELDING CELL
- THE HARDWARE OF THE WELDING CELL AND THE ASSOCIATED SENSORS
- THE SENSOR DATA PROCESSOR

Figure 8 shows the structure of a typical real-time intelligent welding control system. Such a system generally consists of a stand-alone computer system and video display device that are built into a shop floor "hardened" cabinet. The cabinet also includes the necessary computing hardware (that is, backplane, processor boards, and other special-purpose electronics) for the real-time control of the welding cell hardware and sensor data input.

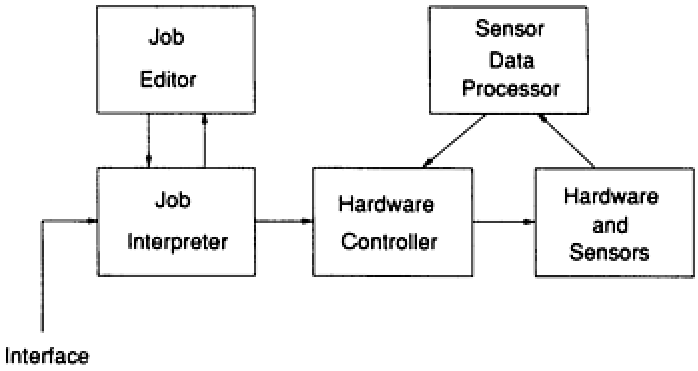

FIG. 8 TYPICAL REAL-TIME CONTROL SYSTEM

The interface module is capable of bidirectional data transfer between the real-time control system and external job data. The job data can reside in a simple data base, for example, on a central data storage device in a network. The interface may only need to download this data. Alternatively, the job data can be obtained by direct communication with the off-line planning system.

The job interpreter reads the interface files and generates the control information. The files that are sent from the interface are used to develop all of the "low-level" control-system parameter values, such as the trajectory for the robot controller and welding parameters for the welding power supply controller. This information is then sent electronically to the welding cell hardware controller subsystem. In addition, the control information is presented to a user interface video screen that the welding cell operator can access. The job interpreter translates the control parameters into data that can be understood by the operator.

The job editor provides an interface for the operator to modify the set points of the welding parameters and robot trajectory prior to the execution. An "intelligent" system will provide software that ensures that these modifications cannot exceed the given minimum and maximum limitations, as specified by the welding engineer or the applicable codes and specifications.

Hardware Controller. A controller is provided for each piece of controllable hardware, such as the robot, welding power supply, wire feed system, gas solenoid and pressure regulators, and others. These hardware controllers typically consist of a digital signal processing integrated circuit that is fast enough to provide control signals to the cell hardware components. These controllers receive information from a software system that interprets sensor data, combines the data with information from the job interpreter, and sends the corresponding commands to the hardware controllers to perform the proper task. Such software systems can be simple set-point or proportional-integral-differential (PID) control systems, or they can be very complex, adaptive, model-based control systems or fuzzy-logic-based control systems.

Basic software control systems attempt to control specific variables based on predefined set-point values. The simplest is an on/off type that turns either on or off, depending on the current value of a control parameter. For example, if the voltage is less than the set-point value, then an on/off type controller would step up the voltage by the difference between the current value and the set-point value. A more complex controller, the PID type, would make an adjustment to the voltage, but would try not to "overshoot" the value by making small changes to the voltage that were "proportional" to the difference between the current value and the set-point value.

Intelligent adaptive software control systems change the values of one or more control parameters simultaneously, according to the desired "result" of the weld. That is, if a sensor indicated that the weld was not penetrating completely through a root pass, then an intelligent adaptive system would simultaneously change the values of the voltage, current, and travel speed to achieve the end result of increasing the weld penetration. These systems represent a significant improvement in the control capability of the welding process and will, in the future, offer the ability to improve quality and productivity while reducing costs and postweld inspection requirements.

Intelligent adaptive control depends on accurate models of the welding process and, in turn, good theoretical research and empirical results from adequate experimentation. One system design that appears to offer a good methodology for solving the intelligent adaptive control problem is the "hierarchical" architecture. In a hierarchical system, one or more control loops exercise supervisory control over the bottom-level process parameter control system (generally, a PID-type controller).

For example, there could be two supervisory control loops. A neural-network-based inner loop could make direct modifications to the weld control parameter set points for the PID controller. An expert-system-based outer loop would make longer-range control decisions, such as overall weld size for adaptive fill, or overall heat-input changes to account for large dimensional changes to the material being welded. The inner loop would take input from sensors, such as a penetration sensor, and make adjustments quickly to account for penetration variations. The outer loop would accept input from sensors, such as a laser scanner joint profiler, and make changes to the weld deposit size and shape inputs to the inner loop.

Welding Cell Hardware and Sensors. The welding cell includes a set of equipment designed to produce automated welds. This equipment consists of a welding torch manipulation system (for example, a robot), a fixturing and positioning system, a parts transport system, a welding power supply, welding consumable supply systems (wire feeder, flux feeder, gas control equipment, and others), control hardware as described above, fume/consumable removal systems (flux vacuum, fume extraction, and others), sensor systems, and other cell-servicing systems (for example, an automated torch cleaning system).

Robot System. The torch manipulation system must be capable of positioning the torch correctly for the welding operation, moving the torch along the weld in a smooth and controllable manner, operating at useful speeds for the welding operations, and positioning the torch for service (for automatic cleaning, for example). Because not all robot systems can move in continuous paths at the appropriate speeds for welding, the choice of robot hardware is important when setting up an automated welding manufacturing cell. Although a robot system is capable of traversing continuous paths at the appropriate speeds, the robot control software and programming language may not be amenable to welding. The software should be designed to easily produce continuous smooth-path motions, including automatic straight line, arc, and spline interpolation and speed control.

The range of motion of the robot should be considered. Its required motion can be very dependent on the motion capability of the welding part fixture and positioning system. The "operating window" of the robot can easily be obtained from its manufacturer and should be carefully examined to determine the suitability of the robot for the parts to be produced. Because robots can have extended lifetimes, all of the anticipated uses of the robot during its expected life should be considered.

The weight-carrying capacity of the robot and the influence of weight on robot accuracy and repeatability should also be considered. The manipulator must be able to lift the weight of the welding torch; the gas, water, and electrode supply tubes; the sensor systems; the flux supply system; and other equipment. As the weight on the robot end effector increases, the accuracy and repeatability of the robot may degrade.

Robot accuracy is a measure of the ability of the robot hardware and associated control system to position the robot at a specific location in 3-D space, given the coordinates of that position. Robot repeatability is the measure of the ability of the robot to return to the same position in space at a later time. Both of these values will affect the ability of the robot to repeatedly position the welding electrode and sensors in the correct positions for welding during the life of the system. These values can be obtained from the robot vendor. Often, these values are also affected by the speed of the robot, and either the vendor or an independent testing laboratory can test the degradation of accuracy and repeatability under dynamic conditions. For welding purposes, it is important to obtain information about the dynamic response of the system, because robot accuracy and repeatability will change as the speed increases. Furthermore, during the welding process, the robot will always be moving (except in the case of spot resistance welds).

Fixturing and Positioning System. The parts to be welded must be held in position for welding. The robot is capable of being programmed to travel along a complex path and to position the welding torch in that preplanned path. However, the accuracy of that positioning not only depends on the robot system and its controller, but on the accuracy of the part location, as well. The positioning system must accurately locate the weld joint so that the robot can "find" the joint. It is difficult and often costly to have a sensor system (for example, a camera) that can locate the part for the robot. Therefore, it is advantageous to utilize a positioning system that can locate the part with sufficient accuracy and repeatability for the weld being made.

Often, it is advantageous to have a positioning system that can move the part to the robot operating window. Such positioners may have several "indexed" positions. An indexed positioner could have four positions: one for loading the part, one for inspecting the part to be welded, one for the welding operation, and one for postweld inspection.

Finally, moving positioners are capable of positioning the weld dynamically during the welding process. In this way, the weld can be maintained in the same welding position (for example, flat) throughout the welding operation. Flat welding often will provide greater weld quality and consistency. Consequently, positioners of this type represent a useful addition to an automated welding cell. Such a system, however, is much more complex than a simple indexed positioner. The motion of the robot and of the positioner must be coordinated. Generally, the positioner has a completely separate controller. Therefore, either the robot controller or a separate computer system must integrate the actions of the positioner with the robot.

Parts Transport System. The parts to be welded must be transported to and from the automated welding cell. Because of the expense of the cell, it should weld as many parts as possible per unit time. To accomplish this, parts need to be transported to and from the weld cell at rates and times that provide the most efficient operation. Thus, if an automated transport system is used, it should be electronically or mechanically coordinated to operate in conjunction with the robot cell positioner and welding operation cycle time.

Welding Power Supply. The choice of welding power supply is important to the success of an intelligent automated welding system. Because the adaptive control system will need to both change the welding parameters during the welding process and control the power supply, a communications link is essential. The power supply must be controllable by an external computer system through this communications link, and must allow modification of voltage and/or current set-point values during the welding process.

Most robots are operated by electric motors with sensitive controls. If the power supply emits high-frequency electromagnetic noise (for example, a high-frequency start or a pulsed-voltage start for gas-tungsten arc welding) or other types of electromagnetic noise, then the robot system may not be able to operate. This type of electronic noise can cause the robot computer to "crash." More serious damage to the robot or the cell controller can also result.

Welding Consumable Supply Systems. The intelligent automated welding cell requires equipment that can supply welding consumables to the welding operation. This includes devices such as a flux feeder, gas supplies, wire/electrode feed system, and others. Each of these components should have a digital (computer) or analog control communications link that can be accessed by the control system. These can be as simple as turning the gas flow on and off or as complex as fully computer controlled linear mass flow metering for accurate gas mixing and flow control. The systems selected should reflect the type of welding that needs to be performed and the necessary control communication links.

Control hardware, as described above, needs to be in a shop floor "hardened" cabinet that may require external cooling. The cabinet also should be shielded from any electromagnetic noise that is generated by the weld or any associated equipment.

Fume/Consumable Removal Systems. Many welding fume components are now recognized as being hazardous or noxious. Therefore, the welding fumes should be removed from the weld at a point that is as close to the weld as possible in order to minimize the volume of air that requires filtering. Fume extraction systems should be obtained and fitted so that they minimally interfere with the operation and motion of the robot and welding torch. For submerged arc welding, flux vacuum systems can provide flux recovery, and the reuse of used flux can provide some cost savings. In some cases, the flux can easily be screened and reused. Because some specifications put limits on the number of times a flux can be recycled, a used-flux management system may need to be developed for the recovery system.

Sensor Systems. A wide variety of sensor systems can be employed in an intelligent automated welding cell. Types of sensors are described below.

Laser scanning sensors use reflected light information from a scanned laser stripe or region to provide data for weld seam guidance, dynamic calculation of the amount of fill required, and location of the part and weld joint for purposes of robot orientation.

A charge-coupled device (CCD) video camera provides information that can be used to determine weld pool size and shape, welding joint guidance (Fig. 9), and collision detection. The camera can be placed inside of a gas-tungsten arc welding (GTAW) torch in order to decrease the sensor package size and to obtain weld pool data. It can be combined with a laser stripe to provide joint geometry and fill information. In addition, a camera has been used to "look" at the back side of a weld and to determine, through measurements of total light intensity, the penetration or lack thereof by the liquid weld pool.

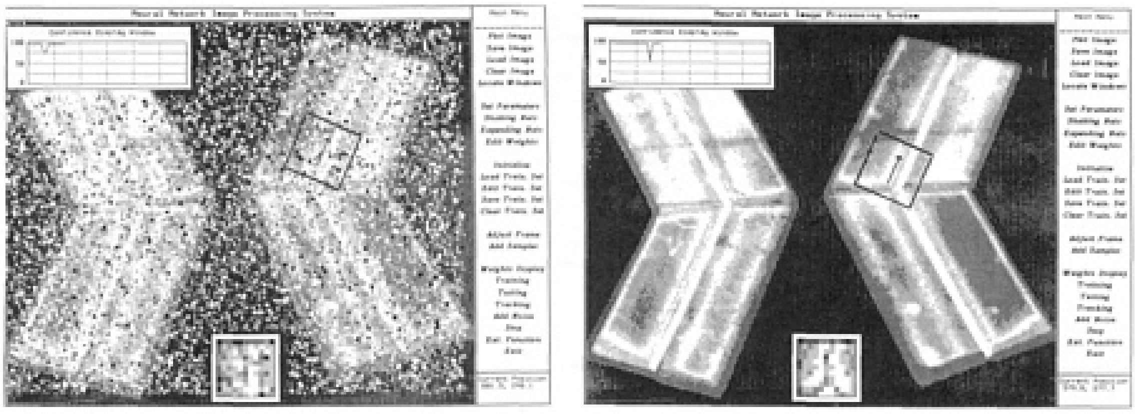

FIG. 9 NEURAL-NETWORK-BASED SEAM-TRACKING SYSTEM. (A) SYSTEM UTILIZING A CCD ARRAY CAMERA TRACKING A CLEAR IMAGE. (B) SYSTEM UTILIZING A CCD ARRAY CAMERA TRACKING AN IMAGE WITH SIMULATED SMOKE AND SPATTER FROM A FCAW PROCESS

Ultrasonic sensors, of both the contacting (piezoelectric crystal) and noncontacting (electromagnetic acoustic transducer) varieties, have been used in the laboratory to evaluate the penetration characteristics of the weld. Other uses also are being investigated.

Acoustic emission sensor data have been used in the laboratory and in limited production applications to determine the formation of welding discontinuities. This sensor has been reported to be able to detect cracking, incomplete fusion, inadequate penetration, and porosity formation.

Through-the-arc voltage/current sensors can detect voltage changes. In a weaving-type weld, the arc gap decreases as the sides of the joint are approached, causing a detectable change in voltage. This voltage change can be used to "guide" the robot along a weld seam in a prepared joint groove. This type of sensor can also determine the transfer modes for gas-metal arc welding and flux-cored arc welding (FCAW), as demonstrated in the laboratory. In addition, it can detect inadequate penetration defect formation through the use of advanced signal processing technology.

Airborne acoustics information (that is, the sound that the weld makes) has been used to diagnose the formation of defects and to determine metal-transfer modes. This type of sensor has been demonstrated in the laboratory. Through the use of multiple sensors and a triangulation algorithm, airborne acoustic data also have been used to locate the position of the welding torch in laboratory demonstrations.

An emission spectroscopic sensor has been developed by the U.S. Army and is patented for use in detecting hydrogen contamination of the welding arc plasma. This sensor can detect the presence of excessive amounts of hydrogen that cause cracking in higher-strength welds.

Infrared thermal area-type cameras have been used in the laboratory and in limited production applications to determine the shape and size of the weld pool and to provide some cooling rate data. A point-type camera also has been used, in conjunction with an AI-based heat-flow model, to determine thermal information and to control the cooling rate of welds.

A low-voltage "touch" sensor operates by energizing the electrode with a very low voltage, which enables the reasonably accurate detection of the electrode when it is brought in very close proximity to the part. This type of sensor also has been used to allow a GTAW electrode to be withdrawn an accurate and small distance from the part and then to energize the arc plasma. This eliminates the need for a high-frequency or high-voltage pulse starter for the automated GTAW process. Experimentally, such sensors are being developed to locate the joint or other locating positions on a part for robot spatial orientation.

An eddy current seam locator (probe) has been used in the laboratory and in limited production applications to scan the surface of a weld joint just ahead of the weld torch in order to "find" the weld seam and track the automated welding system along the joint. This is particularly useful for tracking square butt joints in shiny materials (for example, polished aluminum) that are tightly closed and not visible in CCD camera images.

Other cell services may be required in order to keep the welding equipment operational. For example, automated torch cleaning systems utilize rotating wire brushes or other devices to clean the spatter from GMAW and FCAW welding torches.

Automated welding cell utility services provide adequate power for the welding power supply and for the operation of other equipment; water as a coolant; welding gas; compressed air (for pneumatic devices on the robot or positioner); and network communication links to the computer-integrated manufacturing system.

Intelligent Automation for Joining Technology

Xiaoshu Xu and Jerald E. Jones, American Welding Institute

References

1. CHAPTER 14, *WELDING HANDBOOK,* 8TH ED., VOL 1, AWS, 1987
2. M.Y. DEMDERI, "EXPERT SYSTEM APPLICATION IN MATERIAL PROCESSING AND MANUFACTURING," FORD MOTOR COMPANY, OCT 1989
3. J. JONES, D. WHITE, X. XU, *ET AL.,* DEVELOPMENT OF AN OFF-LINE WELD PLANNING SYSTEM, *TEXAS INSTRUMENTS TECH. J.,* WINTER 1987, P 47-53
4. R.D. STOUT, *WELDABILITY OF STEELS,* 4TH ED., WELDING RESEARCH COUNCIL, 1987
5. H. SUZUKI, N. YURIOKA, AND M. OKUMURA, A NEW CRACKING PARAMETER FOR WELDED STEELS CONSIDERING LOCAL ACCUMULATION OF HYDROGEN, *WELD. WORLD,* VOL 21, 1983, P 2-15
6. H. SUZUKI AND N. YURIOKA, PREVENTION AGAINST COLD CRACKING BY THE HYDROGEN ACCUMULATION CRACKING PARAMETER PHA, *TRANS. JPN. WELD. SOC.,* VOL 14 (NO. 1), APRIL 1983, P 2-14
7. N. YURIOKA, M. OKUMURA, T. KASUYA, AND S. OHSHITA, *WELDING NOTE,* 2ND ED., NIPPON STEEL R&D LABORATORIES
8. D.E. RUMELHART, J.L. MCCLELLAND, AND THE PDP RESEARCH GROUP, *PARALLEL DISTRIBUTED PROCESSING,* MIT PRESS, 1988
9. C. LAU, ED., *NEURAL NETWORKS,* IEEE PRESS, 1992
10. A. ROCK, X. XU, AND J. JONES, "TYPICAL NEURAL NETWORK APPLICATIONS IN SIGNAL

PROCESSING AND PROCESS MODELLING," PAPER NO. 256, CORROSION 92, NACE

11. X. XU AND J. JONES, "A COMPUTER MODEL OF HEAT FLOW IN ARC WELDING," PAPER PRESENTED AT 1991 AWS CONFERENCE (DETROIT, MI), AWS
12. H.P. NII, BLACKBOARD SYSTEMS: THE BLACKBOARD MODEL OF PROBLEMS SOLVING AND THE EVOLUTION OF BLACKBOARD ARCHITECTURES, *THE AI MAGAZINE,* SUMMER 1986, P 28-53
13. X. XU, J. JONES, *ET AL.,* INVESTIGATION OF A BLACKBOARD ARTIFICIAL INTELLIGENCE COMPUTER ARCHITECTURE FOR WELDING PROCEDURE SPECIFICATION AND STRUCTURE INTEGRITY ANALYSIS, PROC. *2ND INT. CONF. ON TRENDS IN WELDING RESEARCH*, ASM INTERNATIONAL 1990, P 985-989

Corrosion of Weldments

A. Wahid, D.L. Olson, and D.K. Matlock, Colorado School of Mines; C.E. Cross, Martin Marietta Astronautics Group

Introduction

WELDMENTS exhibit special microstructural features that need to be recognized and understood in order to predict acceptable corrosion service life of welded structures (Ref 1). This article describes some of the general characteristics associated with the corrosion of weldments. The role of macrocompositional and microcompositional variations, a feature common to weldments, is emphasized in this article to bring out differences that need to be realized in comparing corrosion of weldments to that of wrought materials. A more extensive presentation, with data for specific alloys, is given in Volume 13 of the *ASM Handbook* (Ref 2).

Weldments inherently possess compositional and microstructural heterogeneities, which can be classified by dimensional scale. On the largest scale, a weldment consists of a transition from wrought base metal through a heat-affected zone and into solidified weld metal and includes five microstructurally distinct regions normally identified (Ref 3) as the fusion zone, the unmixed region, the partially melted region, the heat-affected zone, and the base metal. This microstructural transition is illustrated in Fig. 1. The unmixed region is part of the fusion zone, and the partially melted region is part of the heat-affected zone, as described below. Not all five zones are present in any given weldment. For example, autogenous (that is, no filler metal) welds do not have an unmixed zone.

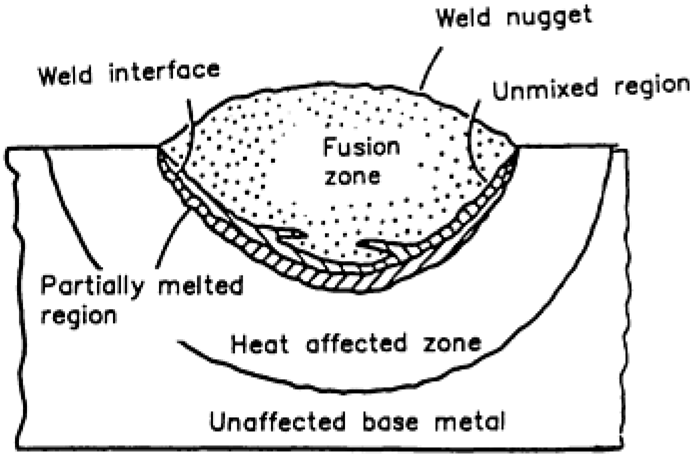

FIG. 1 SCHEMATIC SHOWING THE REGIONS OF A HETEROGENEOUS WELD. SOURCE: REF 3

The fusion zone is the result of melting which fuses the base metal and filler metal to produce a zone with a composition that is most often different from that of the base metal. This compositional difference produces a galvanic couple, which can influence the corrosion process in the vicinity of the weld. This dissimilar-metal couple can produce macroscopic galvanic corrosion.

The fusion zone itself offers a microscopic galvanic effect due to microstructural segregation resulting from solidification (Ref 4). The fusion zone also has a thin region adjacent to the fusion line, known as the unmixed (chilled) region, where the base metal is melted and then quickly solidified to produce a composition similar to the base metal (Ref 5). For example, when type 304 stainless steel is welded using a filler metal with high chromium-nickel content, steep concentration gradients of chromium and nickel are found in the fusion zone, whereas the unmixed zone has a composition similar to the base metal (Fig. 2).

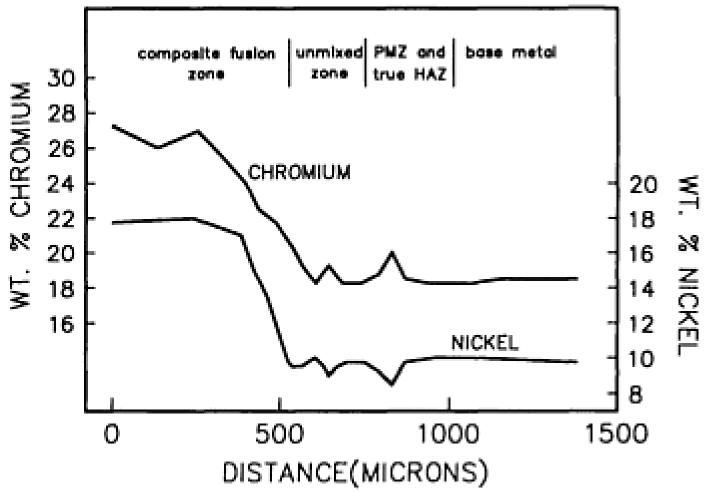

FIG. 2 CONCENTRATION PROFILE OF CHROMIUM AND NICKEL ACROSS THE WELD FUSION BOUNDARY REGION OF TYPE 304 STAINLESS STEEL. SOURCE: REF 5

Heal-Affected Zone. Every position in the heat-affected zone relative to the fusion line experiences a unique thermal experience during welding, in terms of both maximum temperature and cooling rate. Thus, each position has its own microstructural features and corrosion susceptibility.

The partially melted region is usually one or two grains into the heat-affected zone relative to the fusion line. It is characterized by grain boundary liquation, which may result in liquation cracking. These cracks, which are found in the grain boundaries one or two grains below the fusion line, have been identified as potential initiation sites for hydrogen-promoted underbead cracking in high-strength steel.

Microstructural Gradients. On a fine scale, microstructural gradients exist within the heat-affected zone due to different time-temperature cycles experienced by each element of material. Gradients on a similar scale exist within solidified multi-pass weld metal due to bead-to-bead variations in thermal experience. Compositional gradients on the scale of a few microns, referred to as microsegregation, exist within individual weld beads due to segregation of major and trace elements during solidification (Ref 4).

References

1. O.I. STEKLOV *ET AL.*, METHOD OF EVALUATING THE INFLUENCE OF NON-UNIFORMITY OF WELDED JOINT PROPERTIES ON CORROSION, *SVAR. PROIZ.*, VOL 19 (NO. 9), 1972, P 34-36
2. K.F. KRYSIAK *ET AL.*, *METALS HANDBOOK*, VOL 13, 9TH ED., ASM, 1987, P 344-367
3. W.F. SAVAGE, NEW INSIGHT INTO WELD CRACKING AND A NEW WAY OF LOOKING AT WELDS, *WELD. DES. ENG.*, DEC 1969
4. R.G. BUCHHEIT, JR., J.P. MORAN, AND G.E. STONER, LOCALIZED CORROSION BEHAVIOR OF ALLOY 2090--THE ROLE OF MICROSTRUCTURAL HETEROGENEITY, *CORROSION*, VOL 46 (NO. 8), 1990, P 610-617
5. W.A. BAESLACK III, J.C. LIPPOLD, AND W.F. SAVAGE, UNMIXED ZONE FORMATION IN AUSTENITIC STAINLESS STEEL WELDMENTS, *WELD. J.*, VOL 58 (NO. 6), 1979, P 168S-176S

Corrosion of Weldments

A. Wahid, D.L. Olson, and D.K. Matlock, Colorado School of Mines; C.E. Cross, Martin Marietta Astronautics Group

---

Forms of Weld Corrosion

Weldments can experience all the classical forms of corrosion, but they are particularly susceptible to those affected by variations in microstructure and composition. Specifically, galvanic corrosion, pitting, stress corrosion, intergranular corrosion, and hydrogen cracking must be considered when designing welded structures.

Galvanic Couples. Although some alloys can be autogenously welded, filler metals are more commonly used. The use of filler metals with compositions different from the base material may produce an electrochemical potential difference that makes some regions of the weldment more active. For example, Fig. 3 depicts weld metal deposits that have different corrosion behavior from the base metal in three aluminum alloys (Ref 6).

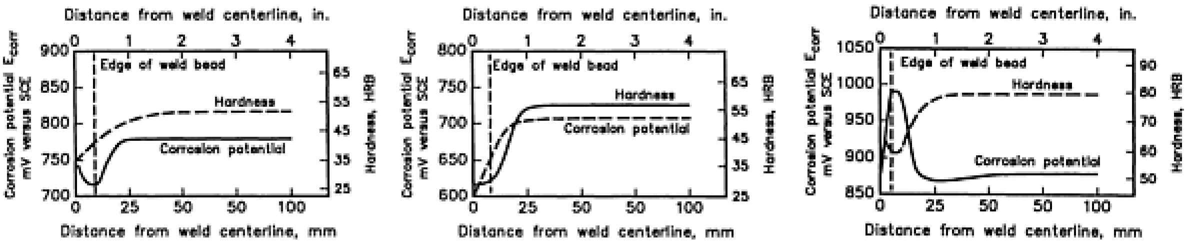

(a) (b) (c)

FIG. 3 EFFECT OF WELDING HEAT ON MICROSTRUCTURE, HARDNESS, AND CORROSION POTENTIAL OF THREE ALUMINUM ALLOY WELDED ASSEMBLIES. (A) ALLOY 5456-H321 BASE METAL WITH ALLOY 5556 FILLER. (B) ALLOY 2219-T87 BASE METAL WITH ALLOY 2319 FILLER. (C) ALLOY 7039-T651 BASE METAL WITH ALLOY 5183 FILLER. SOURCE; REF 6

For the majority of aluminum alloys, the weld metal and the heat-affected zone become more noble relative to the base metal, as demonstrated in Fig. 3(a) and 3(b) for a saltwater environment (Ref 6). Certain aluminum alloys, however, form narrow anodic regions in the heat-affected zone and are prone to localized attack. Alloys 7005 and 7039 are particularly susceptible to this problem (Fig. 3c).

There are a number of other common weld deposit/base metal combinations that are known to form galvanic couples. It is common practice to use austenitic stainless steel welding consumables for field repair of heavy machinery, particularly those fabricated from high-strength low-alloy steel. This practice leaves a cathodic stainless steel weld deposit in electrical contact with the steel. In the presence of corrosive environments, hydrogen is generated at the austenitic weld metal cathode, which is capable of maintaining a high hydrogen content without cracking. However, the cathodic behavior of the austenitic weld deposit may increase the susceptibility for stress-corrosion cracking in the heat-affected zone of the high-strength steel. A 40% thermal expansion mismatch between the austenitic stainless steel and ferritic base metal produces a significant residual stress field in the weldment; this residual stress field also contributes to cracking susceptibility. A similar, but more localized, behavior may explain the correlation between stress-corrosion cracking susceptibility and the presence of retained austenite in high-strength steel weld deposits.

Another common dissimilar metal combination involves the use of high-nickel alloys for weld repair of cast iron. Fe-55Ni welding electrodes are used to make weld deposits that can hold in solid solution many of the alloying elements common to cast iron. Furthermore, weld deposits made with Fe-55Ni welding consumables have an acceptable thermal expansion match to the cast iron. Because cast iron is anodic to the high-nickel weld deposit, corrosive attack occurs in the cast iron adjacent to the weld deposit. It is suggested that cast iron welds made with high-nickel deposits be coated (painted) to reduce the susceptibility to selective corrosion attack.

Plain carbon steel weldments can also exhibit galvanic attack. For example, the E6013 welding electrode is known to be highly anodic to A285 base metal in a seawater environment (Ref 7). It is important to select a suitable filler metal when an application involves a harsh environment.

Weld Decay of Stainless Steel. During welding of stainless steels, local sensitized zones (i.e., regions susceptible to corrosion) often develop. Sensitization is due to the formation of chromium carbide along grain boundaries, resulting in depletion of chromium in the region adjacent to the grain boundary (Ref 8, 9, 10, 11, 12, 13, 14). This chromium depletion produces very localized galvanic cells (Fig. 4). If this depletion drops the chromium content below the necessary 12 wt% that is required to maintain a protective passive film, the region will become sensitized to corrosion, resulting in intergranular attack. This type of corrosion most often occurs in the heat-affected zone. Intergranular corrosion causes a loss of metal in a region that parallels the weld deposit (Fig. 5). This corrosion behavior is called weld decay (Ref 13).

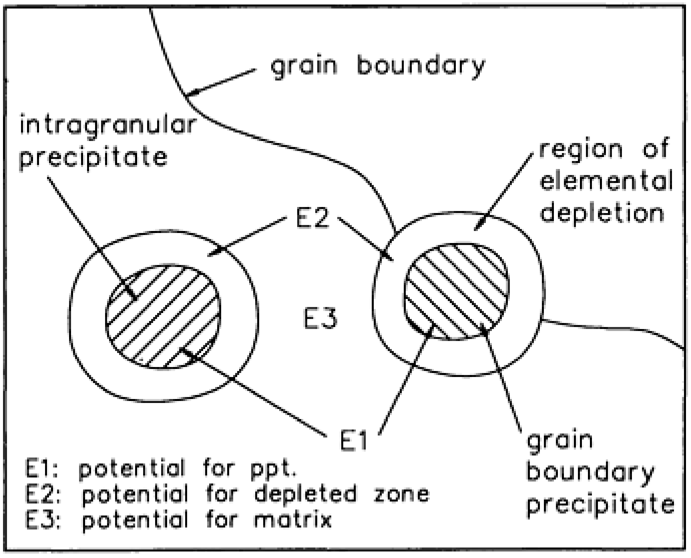
grain boundary
intragranular precipitate
region of elemental depletion
E2
E3
E1
grain boundary precipitate
E1: potential for ppt.
E2: potential for depleted zone
E3: potential for matrix

FIG. 4 DEPLETED REGIONS ADJACENT TO PRECIPITATES. THESE REGIONS CAUSE AN ELECTROCHEMICAL POTENTIAL (E) DIFFERENCE THAT CAN PROMOTE LOCALIZED CORROSION AT THE MICROSTRUCTURAL LEVEL.

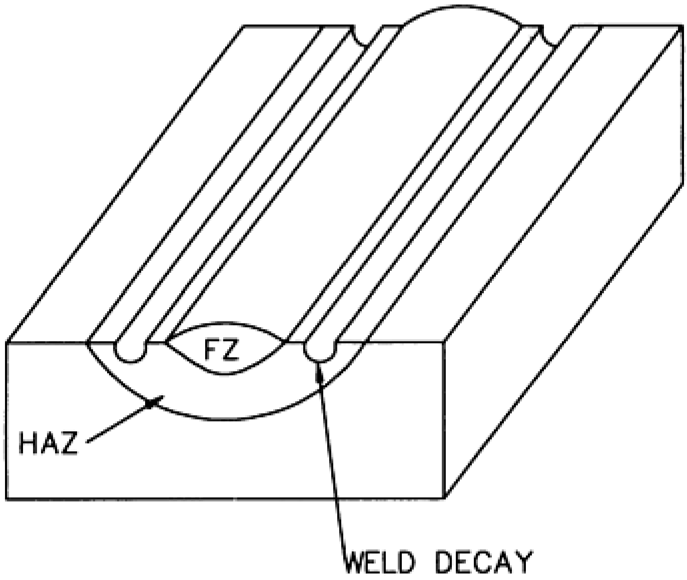
FZ
HAZ
WELD DECAY

FIG. 5 INTERGRANULAR CORROSION (WELD DECAY) OF STAINLESS STEEL WELDMENTS. FZ, FUSION ZONE

The formation of sufficient chromium carbide to cause sensitization can be described by the C-shaped curves on the continuous cooling diagram illustrated in Fig. 6. The figure shows susceptibility to sensitization as a function of temperature, time, and carbon content (Ref 15). If the cooling rate is sufficiently great (curve A in Fig. 6), the cooling curve will not intersect the given C-shaped curve for chromium carbide and the stainless steel will not be sensitized. By decreasing the cooling rate, the cooling curve (curve B) eventually intersects the C-shape nucleation curve, indicating that sensitization may occur. At very low cooling rates, the formation of chromium carbide occurs at higher temperature and allows for more nucleation and growth, resulting in a more extensive chromium-depleted region.

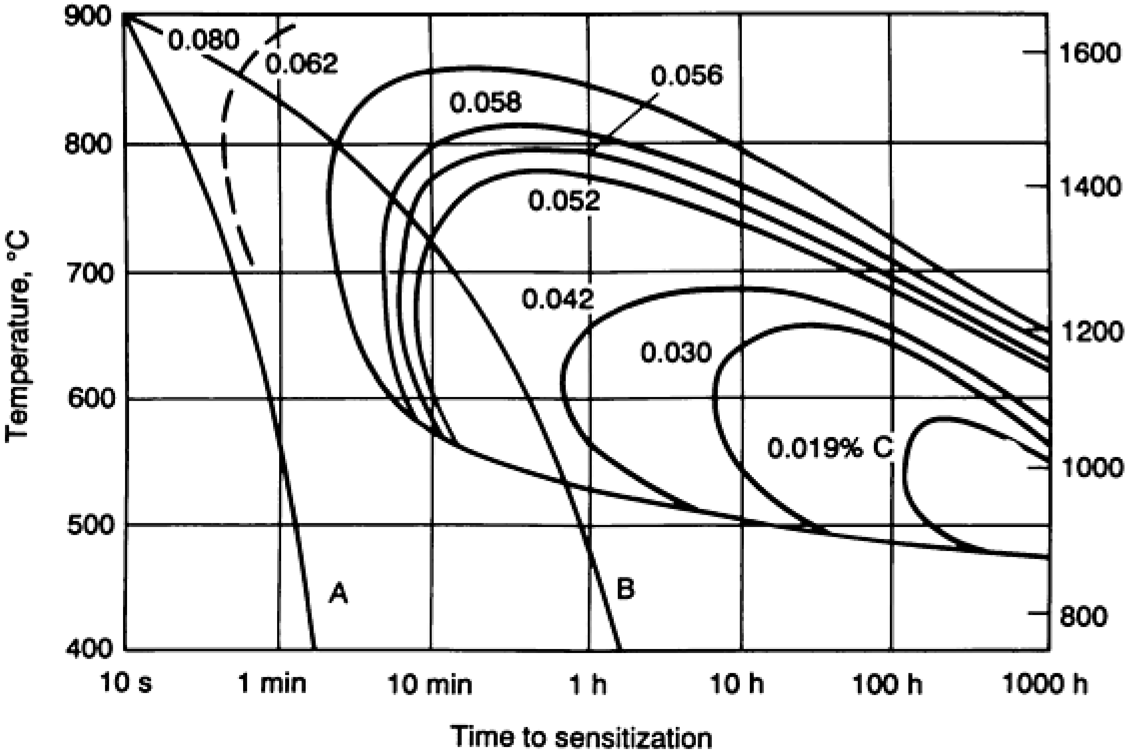

FIG. 6 TIME-TEMPERATURE-SENSITIZATION CURVES FOR TYPE 304 STAINLESS STEEL IN A MIXTURE OF $CUSO_4$ AND HSO_4 CONTAINING COPPER. SOURCE: REF 15. CURVES A AND B INDICATE HIGH AND MEDIUM COOLING RATES, RESPECTIVELY.

The minimum time required for sensitization as a function of carbon content in a typical stainless steel alloy is depicted in Fig. 7. Because the normal welding thermal cycle is completed in approximately two minutes, for this example the carbon content must not exceed 0.07 wt% to avoid sensitization. Notice that the carbide nucleation curves of Fig. 6 move down and to longer times with decreasing carbon content, making it more difficult to form carbides for a given cooling rate.

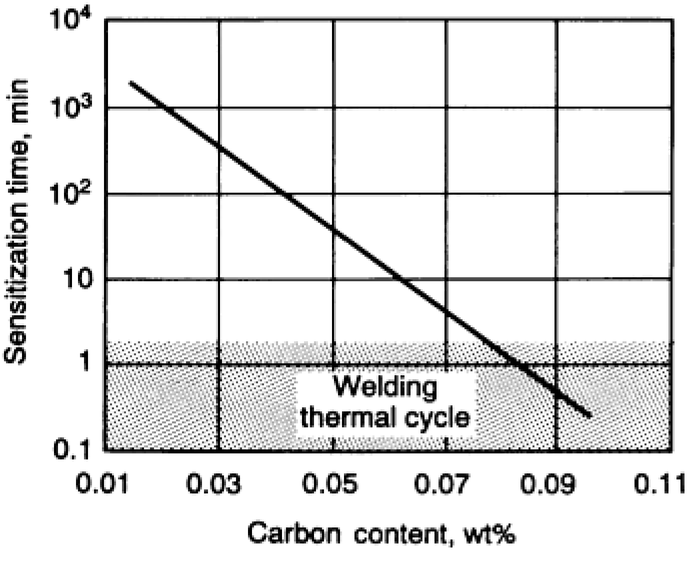

FIG. 7 MINIMUM SENSITIZATION TIME FROM A TIME-TEMPERATURE-SENSITIZATION DIAGRAM AS A FUNCTION OF CARBON CONTENT FOR A TYPICAL 300-SERIES STAINLESS STEEL ALLOY. SOURCE: REF 15

The control of stainless steel sensitization may be achieved by using:

- A POSTWELD HIGH-TEMPERATURE ANNEAL AND QUENCH TO REDISSOLVE THE CHROMIUM AT GRAIN BOUNDARIES, AND HINDER CHROMIUM CARBIDE FORMATION ON COOLING
- A LOW-CARBON GRADE OF STAINLESS STEEL (E.G., 304L OR 316L) TO AVOID CARBIDE FORMATION
- A STABILIZED GRADE OF STAINLESS STEEL CONTAINING TITANIUM (ALLOY 321) OR NIOBIUM (ALLOY 327), WHICH PREFERENTIALLY FORM CARBIDES AND LEAVE CHROMIUM IN SOLUTION. (THERE IS THE POSSIBILITY OF KNIFE-LINE ATTACK IN STABILIZED GRADES OF STAINLESS STEEL.)
- A HIGH-CHROMIUM ALLOY (E.G., ALLOY 310)

Role of Delta Ferrite in Stainless Steel Weld Deposits. Austenitic weld deposits are frequently used to join various ferrous alloys. It has been well established that it is necessary to have austenitic weld deposits solidify as primary ferrite, also known as a δ ferrite, if hot cracking is to be minimized (Ref 16, 17). The amount and form of ferrite in the weld metal can be controlled by selecting a filler metal with the appropriate chromium and nickel equivalent. A high chromium/nickel ratio favors primary ferrite formation, whereas a low ratio promotes primary austenite (Fig. 8). An optimum condition can be attained for ferrite contents between 3 and 8 vol% in the weld deposit. Ferrite contents above 3 vol% usually guarantee primary ferrite formation and thus reduce hot cracking susceptibility. However, ferrite above 10 vol% can degrade mechanical properties at low- or high-temperature service. At low temperatures, excess ferrite can promote crack paths when the temperature is below the ductile-brittle transition temperature. At high temperatures, continuous brittle sigma phase may form at the interface between the austenite and the ferrite. The ferrite content can be confirmed using magnetic measuring equipment (Ref 16, 17).

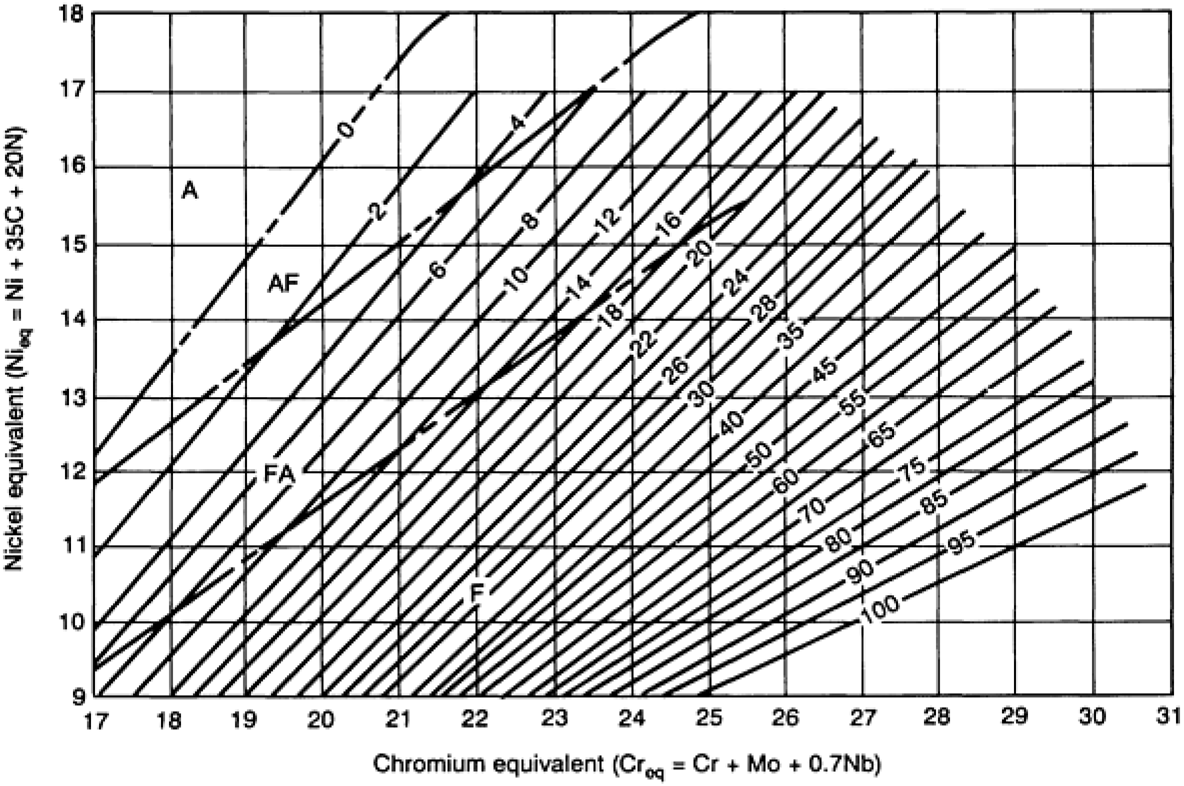

FIG. 8 WELDING RESEARCH COUNCIL (WRC-1988) DIAGRAM USED TO PREDICT WELD METAL FERRITE CONTENT. SOURCE: REF 18

Figure 8 can be used to predict the type of ferrite (primary or eutectic) and the ferrite content when a difference exists between the stainless steels being joined, such as when welding type 304 to type 310 stainless steel (Ref 18). This diagram shows the compositional range for the desirable primary solidification mode. The dotted lines on the diagram indicate the various transitions in the primary solidification phase. Because not all ferrite is primary ferrite (i.e., some is a phase component of a ferrite-austenite eutectic), this diagram can be used to ensure that ferrite is the first solid (primary) phase to form. This condition occurs when the weld deposit has a composition in the range labeled FA in Fig. 8. Because primary ferrite is the preferable microstructure, use of this diagram should reduce problems of hot cracking during welding. Also, the corrosion behavior of stainless steel weld deposits and castings is measurably different depending on whether the stainless steel has a microstructure generated with primary ferrite or primary austenite (Ref 19, 20, 21, 22, 23, 24, 25). Thus, knowledge of the weld metal ferrite content and form is necessary in order to be able to properly characterize and predict corrosion behavior.

Pitting is a form of localized attack caused by a breakdown in the thin passive oxide film that protects material from the corrosion process. Pits are commonly the result of a concentration cell established by a variation in solution composition that is in contact with the alloy material. Such compositional variations result when the solution at a surface irregularity is different from that of the bulk solution composition. Once a pit has formed, it acts as an anode supported by relatively large cathodic regions. Pitting has a delay time prior to nucleation and growth, and nucleation is very site-selective and microstructure-dependent. Pits are often initiated at specific microstructural features in the weld deposit (Ref 26). Pitting occurs when the material/solution combination achieves a potential that exceeds a critical value, known as the pitting potential. The tendency for a given alloy/solution combination to pit can often be characterized by critical potentials for pitting and repassivation determined by a cyclic potentiodynamic polarization technique.

Pits develop more readily in metallurgically heterogeneous materials. For example, when austenitic stainless steel is heated to temperatures where sensitization takes place (Ref 26, 27), the resulting chromium-depleted region is subject to pitting. Pits may also initiate at the austenite-ferrite interfaces in stainless steel weld metal.

Although weld metal has a higher probability of being locally attacked because of microsegregation in the dendritic structure, filler metals are now available that have better pitting resistance than their respective base metals; information about these filler metals can be obtained from consumable suppliers. However, even when the proper filler metal is used, pitting may still occur in the unmixed zone.

Duplex stainless steels, with ferrite contents in the range of 40 to 50 vol%, are often used to decrease the tendency of stress-corrosion cracking in chromium-nickel high-alloy steels. The welding practice for duplex stainless steels must be given special attention (Ref 20, 21, 22, 24, 25, 27) to avoid reduction in corrosion resistance. The combination of a low carbon content and a carefully specified nitrogen addition have been reported to improve resistance to pitting corrosion, stress-corrosion cracking, and intergranular corrosion in the as-welded condition. The low carbon content helps avoid sensitization, while the addition of nitrogen slows the precipitation kinetics associated with the segregation of chromium and molybdenum during the welding process (Ref 1). On rapid cooling from high temperature, nitrogen also has been reported to form deleterious precipitates (for example, Cr_2N) in the ferrite, thus reducing the corrosion resistance (Ref 28). Nitrogen also increases the formation of austenite in the heat-affected zone and weld metal during cooling. A minimum pitting corrosion rate is achieved at a ferrite content of about 50 vol%.

Stress-Corrosion Cracking. Weldments can be susceptible to stress-corrosion cracking under specific environmental conditions. This cracking requires the proper combination of corrosive media, susceptible microstructure, and tensile stress. Welds are often loaded in tension (due to residual stress) to a level approaching the yield strength of the base metal. A weld, with its various heterogeneous microstructural features, thus becomes an excellent candidate for stress-corrosion cracking.

Stress-corrosion cracks have an anodic crack tip and often leave apparent corrosion products along the fracture. Cracking is often characterized by crack branching and usually has a delay time prior to crack initiation, with initiation occurring at corrosion pits. Increasing the ferrite content in stainless steel weld metal reduces stress-corrosion cracking susceptibility. Approximately 50 vol% ferrite gives optimum stress-corrosion cracking resistance.

Welding parameters influence the amount and distribution of residual stress, because the extent of the stressed region and the amount of distortion are directly proportional to the size of the weld deposit; this deposit is directly related to the heat input. The thermal experience of welding is often very localized, resulting in strains that can cause distortion and residual stress. These residual stresses can be important in the initiation and propagation of environmentally assisted cracking. The use of small weld deposits reduces the stress and thus reduces the susceptibility of environmentally enhanced cracking.

It is known that postweld heat treatment can reduce stress-corrosion cracking by redistributing the localized load and by reducing the magnitude of the residual tensile stress available to induce corrosion cracking (Ref 29). In a recent study on a cast austenitic stainless steel, post-solidification heat treatments were also shown to modify the local composition gradients, significantly altering the susceptibility of the solidified microstructure to stress-corrosion cracking (Ref 30).

Hydrogen damage results from the combined actions of hydrogen and stress (Ref 31); the stress can be residual or applied. The weld pool in the liquid state has a much higher solubility for hydrogen than solid metal, and during solidification the hydrogen content becomes supersaturated in the weld pool as hydrogen attempts to transport from the weld deposit. While some of the hydrogen leaves the weldment, significant amounts transport through the fusion line and into the heat-affected zone.

The fusion line region also has the highest cooling rate, which may lead in some alloys to fracture-sensitive microstructures (e.g., martensite). The combination of high hydrogen content and brittle microstructures can produce cracking. The result is underbead cracking, characterized by delayed slow propagating cracks that travel adjacent to the fusion line in the heat-affected zone. The typical hydrogen crack travels one or two grains from the fusion line in the heat-affected zone.

Weld metal hydrogen cracking often takes the form of chevron cracks (Ref 32, 33, 34). Chevron crack formation and propagation are influenced by the combination of stress and orientation of the columnar grain boundaries. Chevron cracking occurs at high hydrogen content levels and appears to be inactive at low levels. This type of cracking appears to be sensitive to reheating during multi-pass deposition of welding beads. Chevron cracking is most likely found with flux-related welding processes.

Microbiologically Influenced Corrosion (MIC) is a phenomenon in which microorganisms play a role in the corrosion of metals. This role may be to initiate or accelerate the corrosion process. For example, water and some organic media may contain certain microorganisms that can produce a biofilm when exposed to a metal surface. The resulting nonuniform coverage may lead to a concentration cell and eventually initiate corrosion. In addition, the metabolic process of the microorganism can produce a localized acid environment that changes the corrosion behavior of the exposed metal by, for example, altering anodic and cathodic reactions, destroying protective films, or creating corrosive deposits (Ref 35).

In austenitic stainless steel weldments, the effects of MIC are usually observed as pitting on or adjacent to welds (Ref 36, 37). MIC attacks either γ or α phases, and chlorides am sometimes found in a pit, even when the water has extremely low chloride content. Pits are found in regions of the heat-affected zone at the fusion line, and in the base metal near the weld for reasons not well understood. Them is some evidence that MIC takes place along with stress-corrosion cracking in weldments of austenitic stainless steel. Welding design and plant operation can minimize MIC attack, mainly by preventing an acceptable environment for microorganisms.

Heat-Tint Oxide Formation. The welding process, especially with poor gas shielding, can produce a variation in the thickness of the passivating oxide. The variation in oxidation will result in a gradient in the degree of chromium depletion adjacent to a stainless steel weld. This behavior will cause some tendency for localized corrosion (Ref 1). An indication of this problem can be seen by the heat-tint oxide formation (Ref 38).

References cited in this section

1. O.I. STEKLOV *ET AL.*, METHOD OF EVALUATING THE INFLUENCE OF NON-UNIFORMITY OF WELDED JOINT PROPERTIES ON CORROSION, *SVAR. PROIZ.*, VOL 19 (NO. 9), 1972, P 34-36
6. J.E. HATCH, ED., *ALUMINUM: PROPERTIES AND PHYSICAL METALLURGY*, ASM, 1984, P 283
7. C.A. ARNOLD, "GALVANIC CORROSION MEASUREMENT OF WELDMENTS," PRESENTED AT NACE CORROSION/80, CHICAGO, MARCH 1980
8. M.W. MARSHALL, CORROSION CHARACTERISTICS OF TYPES 304 AND 304L WELDMENTS, *WELD. J.*, VOL 38 (NO. 6), 1959, P 247S-250S
9. M.E. CARRUTHERS, WELD CORROSION IN TYPE 316 AND 316L STAINLESS STEEL AND RELATED PROBLEMS, *WELD. J.*, VOL 38 (NO. 6), 1959, P 259S-267S
10. K.E. PINNOW AND A. MOSKOWITZ, CORROSION RESISTANCE OF STAINLESS STEEL

WELDMENTS, *WELD. J.,* VOL 49 (NO. 6), 1970, P 278-284S

11. T.J. MOORE, TIME-TEMPERATURE PARAMETERS AFFECTING CORROSION OF 18CR-8NI WELD METALS, *WELD. J.,* VOL 39 (NO. 5), 1960, P 199S-204S
12. J. HONEYCOMBE AND T.G. GOOCH, INTERGRANULAR ATTACK IN WELDED STRESS-CORROSION RESISTANT STAINLESS STEEL, *WELD. J.,* VOL 56 (NO. 11), 1977, P 339S-353S
13. T.G. GOOCH AND D.C. WILLINGHAM, WELD DECAY IN AISI 304 STAINLESS STEEL, *MET. CONST. BRIT. WELD. J.,* VOL 3 (NO. 10), 1971, P 366
14. T.G. GOOCH, CORROSION OF AISI TYPE 304 AUSTENITIC STAINLESS STEEL, *BRIT. WELD. J.,* VOL 15 (NO. 7), 1968, P 345
15. R.M. DAVIDSON, T. DEBOLD, AND M.J. JOHNSON, *METALS HANDBOOK,* VOL 13, 9TH ED., ASM INTERNATIONAL, 1987, P 551
16. D.J. KOTECKI, UNDERSTANDING DELTA FERRITE, *WELD. DES. FABR.,* VOL 63 (NO. 12), 1990, P 33-36
17. D.J. KOTECKI, FERRITE CONTROL IN DUPLEX STAINLESS STEEL WELD METAL, *WELD. J.,* VOL 65 (NO. 10), 1987, P 273S-278S
18. T.A. SIEWERT, C.N. MCCOWAN, AND D.L. OLSON, FERRITE NUMBER PREDICTION TO 100FN IN STAINLESS STEEL WELD METAL, *WELD. J.,* VOL 67 (NO. 12), 1988, P 289S-298S
19. W.A. BAESLACK III, D.J. DUQUETTE, AND W.F. SAVAGE, THE EFFECT OF FERRITE CONTENT ON STRESS CORROSION CRACKING IN DUPLEX STAINLESS STEEL WELD METALS AT ROOM TEMPERATURE, *CORROSION,* VOL 35 (NO. 2), 1979, P 45-54
20. W.A. BAESLACK III, D.J. DUQUETTE, AND W.F. SAVAGE, TECHNICAL NOTE: STRESS CORROSION CRACKING IN DUPLEX STAINLESS STEEL WELDMENTS, *WELD. J.,* VOL 57 (NO. 6), 1978, P 175S-177S
21. D.H. SHERMAN, D.J. DUQUETTE, AND W.F. SAVAGE, STRESS CORROSION CRACKING BEHAVIOR OF DUPLEX STAINLESS STEEL WELDMENTS IN BOILING $MGCL_2$, *CORROSION,* VOL 31 (NO. 10), 1975, P 376-380
22. T.G. GOOCH, WELDABILITY OF DUPLEX FERRITIC-AUSTENITIC STAINLESS STEELS, *DUPLEX STAINLESS STEELS,* ASM, 1983, P 573-602
23. A. BACKMAN AND B. LUNDQVIST, PROPERTIES OF A FULLY AUSTENITIC STAINLESS STEEL WELD METAL FOR SEVERE CORROSION ENVIRONMENTS, *WELD. J.,* VOL 56, 1977, P 23S-28S
24. H. MENENDEZ AND T.M. DEVINE, THE INFLUENCE OF MICROSTRUCTURE ON THE SENSITIZATION BEHAVIOR OF DUPLEX STAINLESS STEEL WELDS, *CORROSION,* VOL 46 (NO. 5), 1990, P 410-417
25. T. OGAWA AND T. KOSEKI, EFFECT OF COMPOSITION PROFILES ON METALLURGY AND CORROSION BEHAVIOR OF DUPLEX STAINLESS STEEL WELD METALS, *WELD. J.,* VOL 68 (NO. 5), 1989, P 181S-191S
26. A. GARNER, THE EFFECT OF AUTOGENEOUS WELDING ON CHLORIDE PITTING CORROSION IN AUSTENITIC STAINLESS STEELS, *CORROSION,* VOL 35 (NO. 3), 1979, P 108-114
27. T.A. DEBOLD, J.W. MARTIN, AND J.C. TVERBERG, DUPLEX STAINLESS OFFERS STRENGTH AND CORROSION RESISTANCE, *DUPLEX STAINLESS STEEL,* ASM, 1983, P 169-189
28. N. SRIDHAR AND J. KOLTS, EFFECTS OF NITROGEN ON THE SELECTIVE DISSOLUTION OF A DUPLEX STAINLESS STEEL, *CORROSION,* VOL 43 (NO. 11), 1987, P 646-651
29. K.N. KRISHNAN AND K.P. RAO, ROOM-TEMPERATURE STRESS CORROSION CRACKING RESISTANCE OF POST-WELD HEAT-TREATED AUSTENITIC WELD METALS, *CORROSION,* VOL 46 (NO. 9), 1990, P 734-742
30. G.L. BERRY, JR., D.L. OLSON, AND D.K. MATLOCK, INFLUENCE OF MICROCOMPOSITIONAL GRADIENTS ON STRESS CORROSION CRACK PROPAGATION, *MAT. ENG. SCI.,* VOL A148, 1991, P 1-6
31. B. CRAIG, HYDROGEN DAMAGE, *METALS HANDBOOK,* VOL 13, 9TH ED., ASM, 1987, P 163-164

32. R.A. FARRAR, THE NATURE OF CHEVRON CRACKING IN SUBMERGED ARC WELD METALS, *WELD. RES. LNT.*, VOL 7 (NO. 2), 1987, P 85-89
33. B.R. KEVILLE, AN INVESTIGATION TO DETERMINE THE MECHANISMS INVOLVED IN THE FORMATION AND PROPAGATION OF CHEVRON CRACKS IN SUBMERGED ARC WELDMENTS, *WELD. RES. LNT.*, VOL 6 (NO. 6), 1976, P 47-66
34. D.J. ALLEN, B. CHEW, AND P. HARRIS, THE FORMATION OF CHEVRON CRACKS IN SUBMERGED ARC WELD METAL, *WELD. J.*, VOL 61 (NO. 7), 1982, P 212S-221S
35. S.C. DEXTER, LOCALIZED BIOLOGICAL CORROSION, *METALS HANDBOOK*, VOL 13, 9TH ED., ASM, 1987, P 114-120
36. G.J. LICINA, *SOURCEBOOK OF MICROBIOLOGICALLY INFLUENCED CORROSION IN NUCLEAR POWER PLANTS*, NP5580, ELECTRIC POWER RESEARCH INSTITUTE, 1988
37. S.W. BORENSTEIN, "MICROBIOLOGICALLY INFLUENCED CORROSION FAILURES OF AUSTENITIC STAINLESS STEEL WELDS," PAPER 78, *CORROSION '88*, NACE, 1988
38. S. TURNER AND F.P.A. ROBINSON, THE EFFECT OF THE SURFACE OXIDES PRODUCED DURING WELDING ON THE CORROSION RESISTANCE OF STAINLESS STEELS, *CORROSION*, VOL 45 (NO. 9), 1989, P 710-716

Corrosion of Weldments

A. Wahid, D.L. Olson, and D.K. Matlock, Colorado School of Mines; C.E. Cross, Martin Marietta Astronautics Group

---

Welding Practice to Minimize Corrosion

Several methods are available to minimize corrosion in weldments (Ref 39). The most important of these are discussed below.

Material and Welding Consumable Selection. Careful selection of materials and welding consumables can reduce the macro- and micro- compositional differences across the weldment and thus reduce the galvanic effects.

Surface Preparation. A properly selected cleaning process can reduce defects that are often sites for corrosive attack in aggressive environments. However, the cleaning process can also be a source of trouble. For example, any mechanically cleaned surface (i.e., cleaned by sand blasting or grinding) can leave impurities on the surface. The type of wire brush used can also be an important consideration (Ref 39). Stainless steel brushes are generally preferred because they do not form corrosion products capable of holding moisture.

Welding design should promote deposits that have relatively flat beads with low profiles and have minimal slag entrapment. A poor design can generate crevices that trap stagnant solutions, leading to pitting and crevice corrosion. Irregular weld deposit shapes can promote turbulent flow in a tubular product and result in erosion corrosion.

Welding Practice. Complete penetration is preferred to avoid underbead gaps. Slag should be removed after each pass with a power grinder or power chipping tool. If the welding method uses flux, the geometry of the joint must permit thorough flux removal, because many flux residues are hydrophilic and corrosive.

Weld Surface Finishing. The weld deposit should be inspected visually immediately after welding. Maximum corrosion resistance usually demands a smooth uniformly oxidized surface that is free from foreign particles and irregularities (Ref 37). Deposits normally vary in roughness and in degree of weld spatter, a concern that can be minimized by grinding. For smooth weld deposits, wire brushing may be sufficient. For stainless steel, however, brushing disturbs the existing passive film and may aggravate corrosive attack.

Surface Coating. When a variation in composition across the weld metal can cause localized attack, it may be desirable to use protective coatings. The coating needs to cover both the weldment and the parent metal and often requires special surface preparation.

Postweld Heat Treatment. A postweld heat treatment can be an effective way to reduce corrosion susceptibility. This improved corrosion resistance is accomplished through a reduction in residual stress gradients that influence stress-corrosion cracking growth. Postweld heat treatment can assist in the transport of hydrogen from the weldment and reduce susceptibility to hydrogen cracking. The treatment can also reduce compositional gradients (i.e., microsegregation) and corresponding microgalvanic cells.

Preheat and Interpass Temperature. The selection and use of proper preheat treatment and interpass temperature may prevent hydrogen cracking in carbon and low-alloy steel.

Passivation Treatment. A passivation treatment may increase the corrosion resistance of stainless steel welded components.

Avoidance of Forming Crevices. Slag that is still adhering to the weld deposit and defects such as lack of penetration and microfissures can result in crevices that can promote a localized concentration cell, resulting in crevice corrosion. Proper selection of welding consumables, proper welding practice, and thorough slag removal can alleviate this form of corrosion damage.

Removing Sources of Hydrogen. Through proper selection of welding consumables (that is, low-hydrogen shielded metal arc welding electrodes), proper drying of flux, and welding clean surfaces, the hydrogen pickup can be drastically reduced.

References cited in this section

37. S.W. BORENSTEIN, "MICROBIOLOGICALLY INFLUENCED CORROSION FAILURES OF AUSTENITIC STAINLESS STEEL WELDS," PAPER 78, *CORROSION '88*, NACE, 1988
39. F.C. BRAUTIGAM, WELDING PRACTICES THAT MINIMIZE CORROSION, *CHEM. ENG.,* 17 JAN 1977, P 145-147

Corrosion of Weldments

A. Wahid, D.L. Olson, and D.K. Matlock, Colorado School of Mines; C.E. Cross, Martin Marietta Astronautics Group

References

1. O.I. STEKLOV *ET AL.,* METHOD OF EVALUATING THE INFLUENCE OF NON-UNIFORMITY OF WELDED JOINT PROPERTIES ON CORROSION, *SVAR. PROIZ.,* VOL 19 (NO. 9), 1972, P 34-36
2. K.F. KRYSIAK *ET AL., METALS HANDBOOK,* VOL 13, 9TH ED., ASM, 1987, P 344-367
3. W.F. SAVAGE, NEW INSIGHT INTO WELD CRACKING AND A NEW WAY OF LOOKING AT WELDS, *WELD. DES. ENG.,* DEC 1969
4. R.G. BUCHHEIT, JR., J.P. MORAN, AND G.E. STONER, LOCALIZED CORROSION BEHAVIOR OF ALLOY 2090--THE ROLE OF MICROSTRUCTURAL HETEROGENEITY, *CORROSION*, VOL 46 (NO. 8), 1990, P 610-617
5. W.A. BAESLACK III, J.C. LIPPOLD, AND W.F. SAVAGE, UNMIXED ZONE FORMATION IN AUSTENITIC STAINLESS STEEL WELDMENTS, *WELD. J.,* VOL 58 (NO. 6), 1979, P 168S-176S
6. J.E. HATCH, ED., *ALUMINUM: PROPERTIES AND PHYSICAL METALLURGY,* ASM, 1984, P 283
7. C.A. ARNOLD, "GALVANIC CORROSION MEASUREMENT OF WELDMENTS," PRESENTED AT NACE CORROSION/80, CHICAGO, MARCH 1980
8. M.W. MARSHALL, CORROSION CHARACTERISTICS OF TYPES 304 AND 304L WELDMENTS, *WELD. J.,* VOL 38 (NO. 6), 1959, P 247S-250S
9. M.E. CARRUTHERS, WELD CORROSION IN TYPE 316 AND 316L STAINLESS STEEL AND

RELATED PROBLEMS, *WELD. J.,* VOL 38 (NO. 6), 1959, P 259S-267S

10. K.E. PINNOW AND A. MOSKOWITZ, CORROSION RESISTANCE OF STAINLESS STEEL WELDMENTS, *WELD. J.,* VOL 49 (NO. 6), 1970, P 278-284S
11. T.J. MOORE, TIME-TEMPERATURE PARAMETERS AFFECTING CORROSION OF 18CR-8NI WELD METALS, *WELD. J.,* VOL 39 (NO. 5), 1960, P 199S-204S
12. J. HONEYCOMBE AND T.G. GOOCH, INTERGRANULAR ATTACK IN WELDED STRESS-CORROSION RESISTANT STAINLESS STEEL, *WELD. J.,* VOL 56 (NO. 11), 1977, P 339S-353S
13. T.G. GOOCH AND D.C. WILLINGHAM, WELD DECAY IN AISI 304 STAINLESS STEEL, *MET. CONST. BRIT. WELD. J.,* VOL 3 (NO. 10), 1971, P 366
14. T.G. GOOCH, CORROSION OF AISI TYPE 304 AUSTENITIC STAINLESS STEEL, *BRIT. WELD. J.,* VOL 15 (NO. 7), 1968, P 345
15. R.M. DAVIDSON, T. DEBOLD, AND M.J. JOHNSON, *METALS HANDBOOK,* VOL 13, 9TH ED., ASM INTERNATIONAL, 1987, P 551
16. D.J. KOTECKI, UNDERSTANDING DELTA FERRITE, *WELD. DES. FABR.,* VOL 63 (NO. 12), 1990, P 33-36
17. D.J. KOTECKI, FERRITE CONTROL IN DUPLEX STAINLESS STEEL WELD METAL, *WELD. J.,* VOL 65 (NO. 10), 1987, P 273S-278S
18. T.A. SIEWERT, C.N. MCCOWAN, AND D.L. OLSON, FERRITE NUMBER PREDICTION TO 100FN IN STAINLESS STEEL WELD METAL, *WELD. J.,* VOL 67 (NO. 12), 1988, P 289S-298S
19. W.A. BAESLACK III, D.J. DUQUETTE, AND W.F. SAVAGE, THE EFFECT OF FERRITE CONTENT ON STRESS CORROSION CRACKING IN DUPLEX STAINLESS STEEL WELD METALS AT ROOM TEMPERATURE, *CORROSION,* VOL 35 (NO. 2), 1979, P 45-54
20. W.A. BAESLACK III, D.J. DUQUETTE, AND W.F. SAVAGE, TECHNICAL NOTE: STRESS CORROSION CRACKING IN DUPLEX STAINLESS STEEL WELDMENTS, *WELD. J.,* VOL 57 (NO. 6), 1978, P 175S-177S
21. D.H. SHERMAN, D.J. DUQUETTE, AND W.F. SAVAGE, STRESS CORROSION CRACKING BEHAVIOR OF DUPLEX STAINLESS STEEL WELDMENTS IN BOILING $MGCL_2$, *CORROSION,* VOL 31 (NO. 10), 1975, P 376-380
22. T.G. GOOCH, WELDABILITY OF DUPLEX FERRITIC-AUSTENITIC STAINLESS STEELS, *DUPLEX STAINLESS STEELS,* ASM, 1983, P 573-602
23. A. BACKMAN AND B. LUNDQVIST, PROPERTIES OF A FULLY AUSTENITIC STAINLESS STEEL WELD METAL FOR SEVERE CORROSION ENVIRONMENTS, *WELD. J.,* VOL 56, 1977, P 23S-28S
24. H. MENENDEZ AND T.M. DEVINE, THE INFLUENCE OF MICROSTRUCTURE ON THE SENSITIZATION BEHAVIOR OF DUPLEX STAINLESS STEEL WELDS, *CORROSION,* VOL 46 (NO. 5), 1990, P 410-417
25. T. OGAWA AND T. KOSEKI, EFFECT OF COMPOSITION PROFILES ON METALLURGY AND CORROSION BEHAVIOR OF DUPLEX STAINLESS STEEL WELD METALS, *WELD. J.,* VOL 68 (NO. 5), 1989, P 181S-191S
26. A. GARNER, THE EFFECT OF AUTOGENEOUS WELDING ON CHLORIDE PITTING CORROSION IN AUSTENITIC STAINLESS STEELS, *CORROSION,* VOL 35 (NO. 3), 1979, P 108-114
27. T.A. DEBOLD, J.W. MARTIN, AND J.C. TVERBERG, DUPLEX STAINLESS OFFERS STRENGTH AND CORROSION RESISTANCE, *DUPLEX STAINLESS STEEL,* ASM, 1983, P 169-189
28. N. SRIDHAR AND J. KOLTS, EFFECTS OF NITROGEN ON THE SELECTIVE DISSOLUTION OF A DUPLEX STAINLESS STEEL, *CORROSION,* VOL 43 (NO. 11), 1987, P 646-651
29. K.N. KRISHNAN AND K.P. RAO, ROOM-TEMPERATURE STRESS CORROSION CRACKING RESISTANCE OF POST-WELD HEAT-TREATED AUSTENITIC WELD METALS, *CORROSION,* VOL 46 (NO. 9), 1990, P 734-742
30. G.L. BERRY, JR., D.L. OLSON, AND D.K. MATLOCK, INFLUENCE OF MICROCOMPOSITIONAL GRADIENTS ON STRESS CORROSION CRACK PROPAGATION, *MAT. ENG. SCI.,* VOL A148, 1991, P 1-6

31. B. CRAIG, HYDROGEN DAMAGE, *METALS HANDBOOK,* VOL 13, 9TH ED., ASM, 1987, P 163-164
32. R.A. FARRAR, THE NATURE OF CHEVRON CRACKING IN SUBMERGED ARC WELD METALS, *WELD. RES. LNT.,* VOL 7 (NO. 2), 1987, P 85-89
33. B.R. KEVILLE, AN INVESTIGATION TO DETERMINE THE MECHANISMS INVOLVED IN THE FORMATION AND PROPAGATION OF CHEVRON CRACKS IN SUBMERGED ARC WELDMENTS, *WELD. RES. LNT.,* VOL 6 (NO. 6), 1976, P 47-66
34. D.J. ALLEN, B. CHEW, AND P. HARRIS, THE FORMATION OF CHEVRON CRACKS IN SUBMERGED ARC WELD METAL, *WELD. J.,* VOL 61 (NO. 7), 1982, P 212S-221S
35. S.C. DEXTER, LOCALIZED BIOLOGICAL CORROSION, *METALS HANDBOOK,* VOL 13, 9TH ED., ASM, 1987, P 114-120
36. G.J. LICINA, *SOURCEBOOK OF MICROBIOLOGICALLY INFLUENCED CORROSION IN NUCLEAR POWER PLANTS,* NP5580, ELECTRIC POWER RESEARCH INSTITUTE, 1988
37. S.W. BORENSTEIN, "MICROBIOLOGICALLY INFLUENCED CORROSION FAILURES OF AUSTENITIC STAINLESS STEEL WELDS," PAPER 78, *CORROSION '88,* NACE, 1988
38. S. TURNER AND F.P.A. ROBINSON, THE EFFECT OF THE SURFACE OXIDES PRODUCED DURING WELDING ON THE CORROSION RESISTANCE OF STAINLESS STEELS, *CORROSION,* VOL 45 (NO. 9), 1989, P 710-716
39. F.C. BRAUTIGAM, WELDING PRACTICES THAT MINIMIZE CORROSION, *CHEM. ENG.,* 17 JAN 1977, P 145-147

Overview of Weld Discontinuities

Introduction

DISCONTINUITIES are interruptions in the desirable physical structure of a weld. A discontinuity constituting a danger to the fitness-for-service of a weld is a defect. By definition, a defect is a condition that must be removed or corrected (Ref 1). The word "defect" should therefore be carefully used, because it implies that a weld is defective and requires corrective measures or rejection. Thus, repairs may be made unnecessarily and solely by implication, without a critical engineering assessment. Consequently, the engineering community now tends to use the word "discontinuity" or "flaw" instead of "defect."

The significance of a weld discontinuity should be viewed in the context of the fitness-for-service of the welded construction. Fitness-for-service is a concept of weld evaluation that seeks a balance among quality, reliability, and economy of welding procedure. Fitness-for-service is not a constant. It varies depending on the service requirements of a particular welded structure, as well as on the properties of the material involved. (See the article "Fitness-for-Service Assessment of Welded Structures" in this Section for more information.)

Neither construction materials nor engineered structures are free from imperfections. Welds and weld repairs are not exceptions. Weld acceptance standards are used when a discontinuity has been clearly located, identified, sized, its orientation determined, and its structural significance questioned. Critical engineering assessments of weld discontinuities are performed to define acceptable, harmless discontinuities in a structure that will not sacrifice weldment reliability. One of the major reasons for understanding the engineering meaning of weld discontinuities is to decrease the cost of welded structures by avoiding unnecessary repairs of harmless weld discontinuities (Ref 2). Welders, of course, must constantly be encouraged to make sound (perfect) welds independent of prevailing acceptance standards.

This article describes the types of weld discontinuities that are characteristic of the principal welding processes. In addition, information is provided about the common inspection methods used to detect these discontinuities. More detailed descriptions of these inspection methods can be found in the article "Inspection of Welded Joints" in this Section of the Handbook.

References

1. ASME BOILER AND PRESSURE VESSEL CODE, APPENDIX IV, DRAFT 1, 21 OCT 1981

2. L.W. SANDOR, THE MEANING OF WELD DISCONTINUITIES IN SHIPBUILDING, *PROC. 1ST NAT. CONF. FITNESS-FOR-SERVICE IN SHIPBUILDING* (BOULDER, CO), OCT 1980

Classification of Weld Discontinuities

Discontinuities may be divided into three broad classifications: design related, welding process related, and metallurgical. Design-related discontinuities include problems with design or structural details, choice of the wrong type of weld joint for a given application, or undesirable changes in cross section.

Process-related discontinuities include:

- *UNDERCUT:* A GROOVE MELTED INTO THE BASE METAL ADJACENT TO THE TOE OR ROOT OF A WELD AND LEFT UNFILLED BY WELD METAL
- *SLAG INCLUSIONS:* NONMETALLIC SOLID MATERIAL ENTRAPPED IN WELD METAL OR BETWEEN WELD METAL AND BASE METAL
- *POROSITY:* CAVITY-TYPE DISCONTINUITIES FORMED BY GAS ENTRAPMENT DURING SOLIDIFICATION
- *OVERLAP:* THE PROTRUSION OF WELD METAL BEYOND THE TOE, FACE, OR ROOT OF THE WELD
- *TUNGSTEN INCLUSIONS:* PARTICLES FROM TUNGSTEN ELECTRODES THAT RESULT FROM IMPROPER GAS-TUNGSTEN ARC WELDING PROCEDURES
- *BACKING PIECE LEFT ON:* FAILURE TO REMOVE MATERIAL PLACED AT THE ROOT OF A WELD JOINT TO SUPPORT MOLTEN WELD METAL
- *SHRINKAGE VOIDS:* CAVITY-TYPE DISCONTINUITIES NORMALLY FORMED BY SHRINKAGE DURING SOLIDIFICATION
- *OXIDE INCLUSIONS:* PARTICLES OF SURFACE OXIDES THAT HAVE NOT MELTED AND ARE MIXED INTO THE WELD METAL
- *LACK OF FUSION (LOF):* A CONDITION IN WHICH FUSION IS LESS THAN COMPLETE
- *LACK OF PENETRATION (LOP):* A CONDITION IN WHICH JOINT PENETRATION IS LESS THAN THAT SPECIFIED
- *CRATERS:* DEPRESSIONS AT THE TERMINATION OF A WELD BEAD OR IN THE MOLTEN WELD POOL
- *MELT-THROUGH:* A CONDITION RESULTING WHEN THE ARC MELTS THROUGH THE BOTTOM OF A JOINT WELDED FROM ONE SIDE
- *SPATTER:* METAL PARTICLES EXPELLED DURING WELDING THAT DO NOT FORM A PART OF THE WELD
- *ARC STRIKES (ARC BURNS):* DISCONTINUITIES CONSISTING OF ANY LOCALIZED REMELTED METAL, HEAT-AFFECTED METAL, OR CHANGE IN THE SURFACE PROFILE OF ANY PART OF A WELD OR BASE METAL RESULTING FROM AN ARC
- *UNDERFILL:* A DEPRESSION ON THE FACE OF THE WELD OR ROOT SURFACE EXTENDING BELOW THE SURFACE OF THE ADJACENT BASE METAL

Metallurgical discontinuities include:

- *CRACKS:* FRACTURE-TYPE DISCONTINUITIES CHARACTERIZED BY A SHARP TIP AND HIGH RATIO OF LENGTH AND WIDTH TO OPENING DISPLACEMENT
- *FISSURES:* SMALL CRACKLIKE DISCONTINUITIES WITH ONLY A SLIGHT SEPARATION (OPENING DISPLACEMENT) OF THE FRACTURE SURFACES
- *FISHEYE:* A DISCONTINUITY FOUND ON THE FRACTURE SURFACE OF A WELD IN STEEL THAT CONSISTS OF A SMALL PORE OR INCLUSION SURROUNDED BY A BRIGHT, ROUND AREA

- *SEGREGATION:* THE NONUNIFORM DISTRIBUTION OR CONCENTRATION OF IMPURITIES OR ALLOYING ELEMENTS THAT ARISES DURING THE SOLIDIFICATION OF THE WELD
- *LAMELLAR TEARING:* A TYPE OF CRACKING THAT OCCURS IN THE BASE METAL OR HEAT-AFFECTED ZONE (HAZ) OF RESTRAINED WELD JOINTS THAT IS THE RESULT OF INADEQUATE DUCTILITY IN THE THROUGH-THICKNESS DIRECTION OF STEEL PLATE

The observed occurrence of discontinuities and their relative amounts depend largely on the welding process used, the inspection method applied, the type of weld made, the joint design and fit-up obtained, the material utilized, and the working and environmental conditions. The most frequent weld discontinuities found during manufacture, ranked in order of decreasing occurrence on the basis of arc-welding processes are:

SHIELDED METAL ARE WELDING (SMAW)
SLAG INCLUSIONS
POROSITY
LOF/LOP
UNDERCUT
SUBMERGED ARE WELDING (SAW)
LOF/LOP
SLAG INCLUSIONS
POROSITY
FLUX-CORED ARC WELDING (FCAW)
SLAG INCLUSIONS
POROSITY
LOF/LOP
GAS-METAL ARC WELDING (GMAW)
POROSITY
LOF/LOP
GAS-TUNGSTEN ARC WELDING (GTAW)
POROSITY
TUNGSTEN INCLUSIONS

The commonly encountered inclusions-as well as cracking, the most serious of weld defects--will be discussed in this Section.

Gas porosity can occur on or just below the surface of a weld. Pores are characterized by a rounded or elongated teardrop shape with or without a sharp point. Pores can be uniformly distributed throughout the weld or isolated in small groups; they can also be concentrated at the root or toe of the weld. Porosity in welds is caused by gas entrapment in the molten metal, by too much moisture on the base or filler metal, or by improper cleaning of the joint during preparation for welding.

The type of porosity within a weld is usually designated by the amount and distribution of the pores. Some of the types are classified as follows:

- *UNIFORMLY SCATTERED POROSITY:* CHARACTERIZED BY PORES SCATTERED UNIFORMLY THROUGHOUT THE WELD (FIG. 1)
- *CLUSTER POROSITY:* CHARACTERIZED BY CLUSTERS OF PORES SEPARATED BY POROSITY-FREE AREAS (FIG. 1B)
- *LINEAR POROSITY:* CHARACTERIZED BY PORES THAT ARE LINEARLY DISTRIBUTED (FIG. 1C). LINEAR POROSITY GENERALLY OCCURS IN THE ROOT PASS AND IS ASSOCIATED

WITH INCOMPLETE JOINT PENETRATION

- *ELONGATED POROSITY:* CHARACTERIZED BY HIGHLY ELONGATED PORES INCLINED TO THE DIRECTION OF WELDING. ELONGATED POROSITY OCCURS IN A HERRINGBONE PATTERN (FIG. 1D)
- *WORMHOLE POROSITY:* CHARACTERIZED BY ELONGATED VOIDS WITH A DEFINITE WORMLIKE SHAPE AND TEXTURE (FIG. 2)

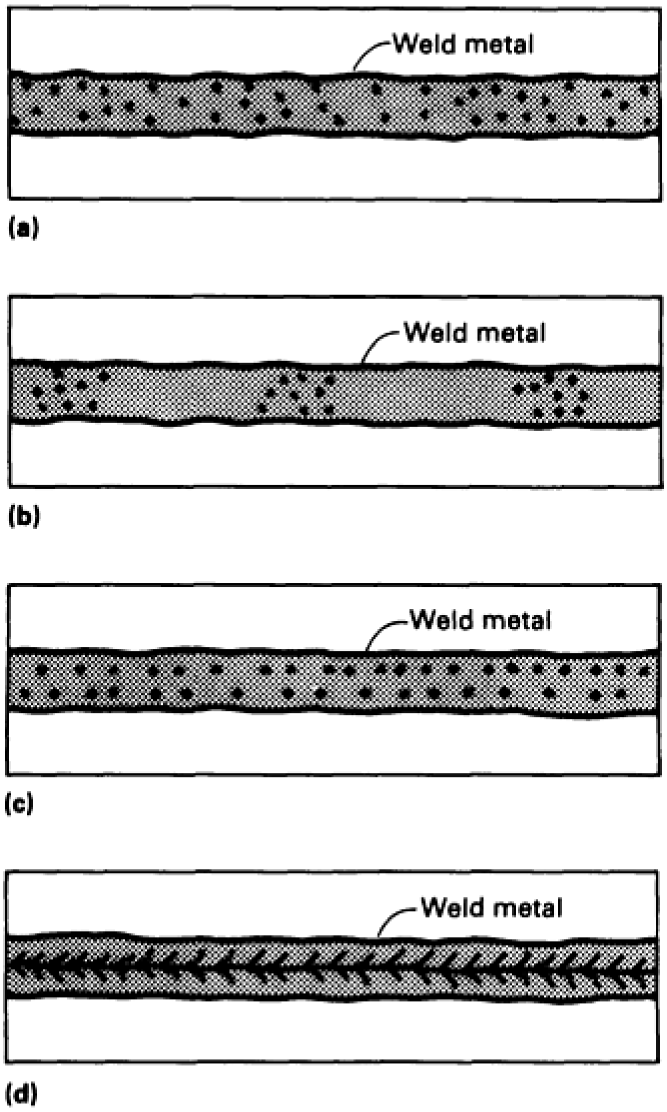

FIG. 1 TYPES OF GAS POROSITY COMMONLY FOUND IN WELD METAL. (A) UNIFORMLY SCATTERED POROSITY. (B) CLUSTER POROSITY. (C) LINEAR POROSITY. (D) ELONGATED POROSITY

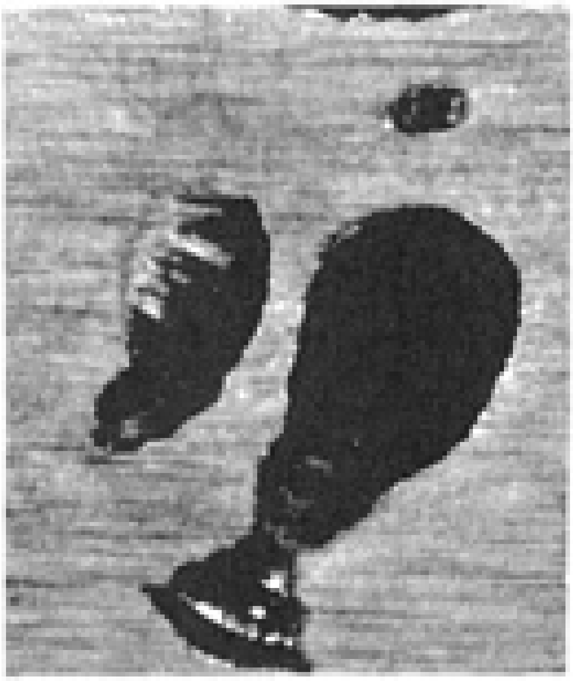

FIG. 2 WORMHOLE POROSITY IN A WELD BEAD. LONGITUDINAL CUT. 20×

Radiography is the most widely used nondestructive method for detecting subsurface gas porosity in weldments. The radiographic image of round porosity appears as round or oval spots with smooth edges, and elongated porosity appears as oval spots with the major axis sometimes several times longer than the minor axis. The radiographic image of wormhole porosity depends largely on the orientation of the elongated cavity with respect to the incident x-ray beam. The presence of top-surface or root reinforcement affects the sensitivity of inspection, and the presence of foreign material, such as loose scale, flux, or weld spatter, may interfere with the interpretation of results.

Ultrasonic inspection is capable of detecting subsurface porosity. However, it is not extensively used for this purpose except to inspect thick sections or inaccessible areas where radiographic sensitivity is limited. Surface finish and grain size affect the validity of the inspection results.

Eddy current inspection, like ultrasonic inspection, can be used for detecting subsurface porosity. Normally, eddy current inspection is confined to use on thin-wall welded pipe and tubing because eddy currents are relatively insensitive to flaws that do not extend to the surface or into the near-surface layer.

Magnetic particle inspection and liquid penetrant inspection are not suitable for detecting subsurface gas porosity. These methods are restricted to the detection of only those pores that are open to the surface.

Slag inclusions may occur when using welding processes that employ a slag covering for shielding purposes. (With other processes, the oxide present on the metal surface before welding may also become entrapped.) Slag inclusions can be found near the surface and in the root of a weld (Fig. 3a), between weld beads in multiple-pass welds (Fig. 3b), and at the side of a weld near the root (Fig. 3c).

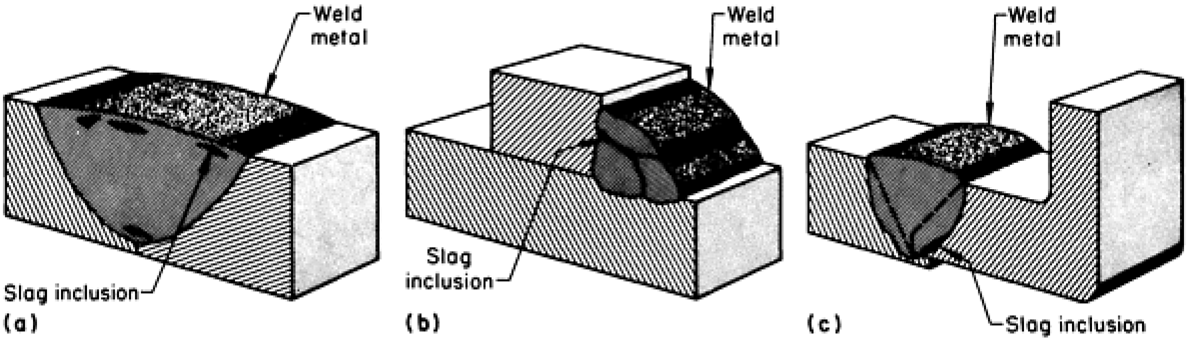

FIG. 3 SECTIONS SHOWING LOCATIONS OF SLAG INCLUSIONS IN WELD METAL. (A) NEAR THE SURFACE AND IN THE ROOT OF A SINGLE-PASS WELD. (B) BETWEEN WELD BEADS IN A MULTIPLE-PASS WELD. (C) AT THE SIDE OF A WELD NEAR THE ROOT

During welding, slag may spill ahead of the arc and subsequently be covered by the weld pool because of poor joint fit-up, incorrect electrode manipulation, or forward arc blow. Slag trapped in this manner is generally located near the root. Radical motions of the electrode, such as wide weaving, may also cause slag entrapment on the sides or near the top of the weld after the slag spills into a portion of the joint that has not been filled by the molten pool. Incomplete removal of the slag from the previous pass in multiple-pass welding is another common cause of entrapment. In multiple-pass welds, slag may be entrapped any number of places in the weld between passes. Slag inclusions are generally oriented along the direction of welding.

Three methods used for the detection of slag below the surface of single-pass or multiple-pass welds are magnetic particle, radiographic, and ultrasonic inspection. Depending on their size, shape, orientation, and proximity to the surface, slag inclusions can be detected by magnetic particle inspection with a dc power source, provided the material is ferromagnetic. Radiography can be used for any material, but is the most expensive of the three methods. Ultrasonic inspection can also be used for any material and is the most reliable and least expensive method. If the weld is machined to a flush contour, flaws as close as 0.8 mm ($\frac{1}{32}$ in.) to the surface can be detected with the straight-beam technique of ultrasonic inspection, provided the instrument has sufficient sensitivity and resolution. A 5 or 10 MHz dual-element transducer is normally used in this application. If the weld cannot be machined, near-surface sensitivity will be low because the initial pulse is excessively broadened by the rough as-welded surface. Unmachined welds can be readily inspected by direct-beam and reflected-beam techniques, using an angle-beam (shear-wave) transducer.

Tungsten inclusions are particles found in the weld metal from the nonconsumable tungsten electrode used in GTAW. These inclusions are the result of:

- EXCEEDING THE MAXIMUM CURRENT FOR A GIVEN ELECTRODE SIZE OR TYPE
- LETTING THE TIP OF THE ELECTRODE MAKE CONTACT WITH THE HOT TIP OF THE ELECTRODE
- USING AN EXCESSIVE ELECTRODE EXTENSION
- INADEQUATE GAS SHIELDING OR EXCESSIVE WIND DRAFTS, WHICH RESULT IN OXIDATION
- USING IMPROPER SHIELDING GASES SUCH AS ARGON-OXYGEN OR ARGON-CO2 MIXTURES, WHICH ARE USED FOR GMAW

Tungsten inclusions, which are not acceptable for high-quality work, can only be found by internal inspection techniques, particularly radiographic testing.

Lack of fusion and lack of penetration result from improper electrode manipulation and the use of incorrect welding conditions. Fusion refers to the degree to which the original base metal surfaces to be welded have been fused to the filler metal; penetration refers to the degree to which the base metal has been melted and resolidified to result in a deeper throat that was present in the joint before welding. In effect, a joint can be completely fused but have incomplete root penetration to obtain the throat size specified. Based on these definitions, LOF discontinuities are located on the sidewalls of a joint, and LOP discontinuities are located near the root (Fig. 4). With some joint configurations, such as butt joints, the two terms can be used interchangeably. The causes of LOF include excessive travel speed, bridging, excessive electrode size, insufficient current, poor joint preparation, overly acute joint angle, improper electrode manipulation, and excessive arc blow. Lack of penetration may be the result of low welding current, excessive travel speed, improper electrode manipulation, or surface contaminants such as oxide, oil, or dirt that prevent full melting of the underlying metal.

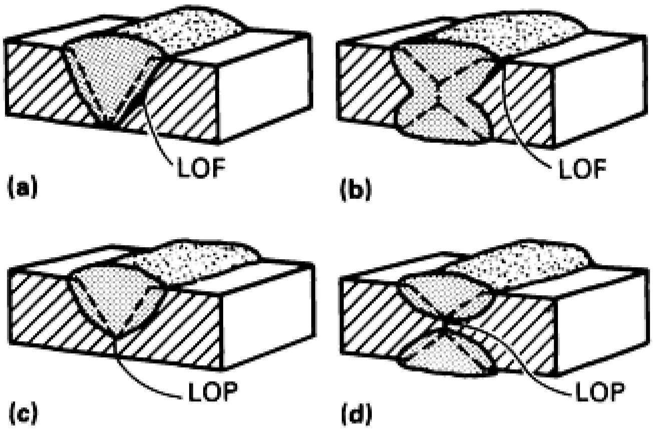

FIG. 4 LACK OF FUSION IN (A) A SINGLE-V-GROOVE WELD AND (B) DOUBLE-V-GROOVE WELD. LACK OF PENETRATION IN (C) A SINGLE-V-GROOVE WELD AND (D) A DOUBLE-V-GROOVE WELD

Radiographic methods may be unable to detect these discontinuities in certain cases, because of the small effect they have on x-ray absorption. As will be described later, however, lack of sidewall fusion is readily detected by radiography. Ultrasonically, both types of discontinuities often appear as severe, almost continuous, linear porosity because of the nature of the unbonded areas of the joint. Except in thin sheet or plate, these discontinuities may be too deep-lying to be detected by magnetic particle inspection.

Geometric weld discontinuities are those associated with imperfect shape or unacceptable weld contour. Undercut, underfill, overlap, excessive reinforcement, fillet shape, and melt-through, all of which were defined earlier, are included in this grouping. Geometric discontinuities are shown schematically in Fig. 5. Radiography is used most often to detect these flaws.

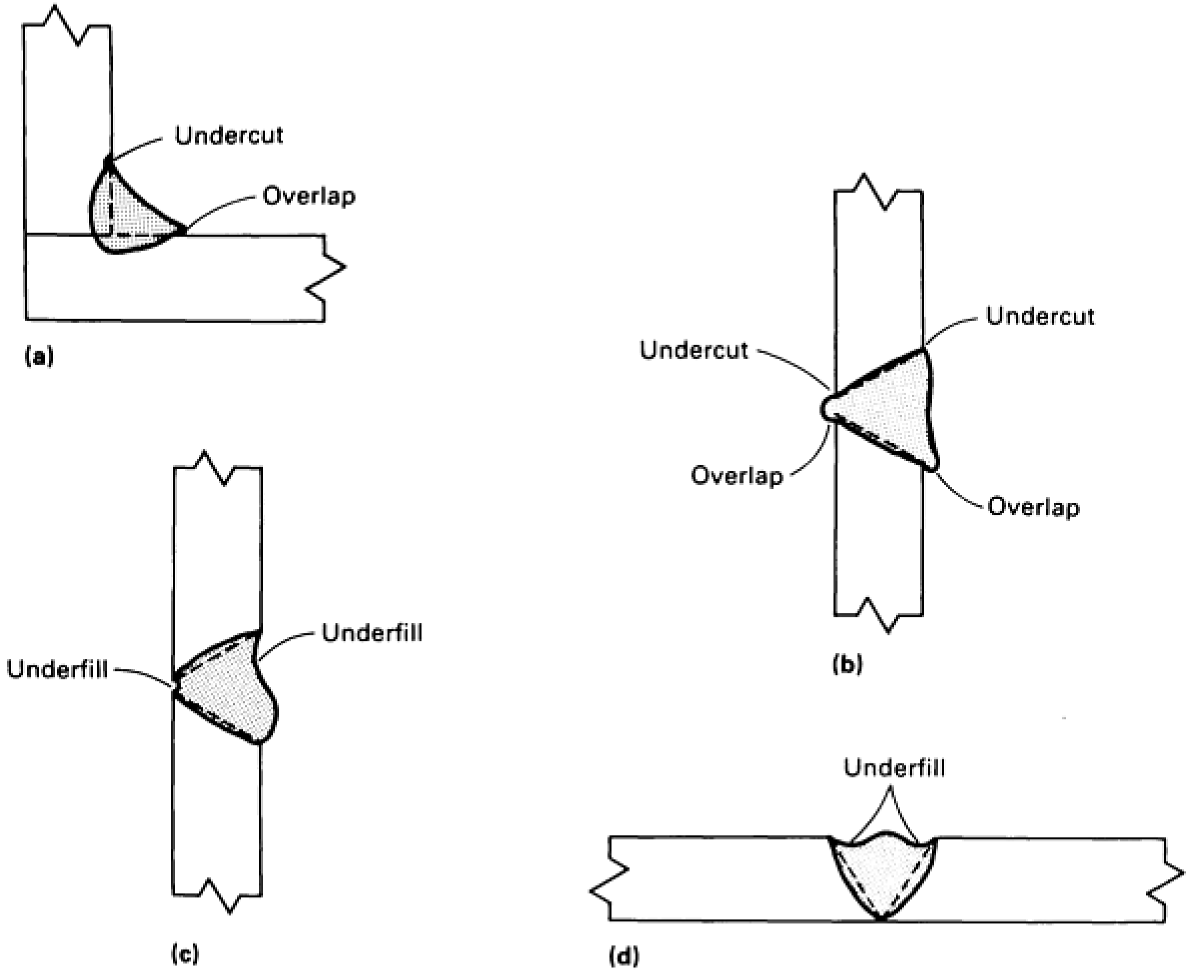
Undercut
Overlap
(a)
Undercut
Undercut
Overlap
Overlap
(b)
Underfill
Underfill
(c)
Underfill
(d)

FIG. 5 WELD DISCONTINUITIES AFFECTING WELD SHAPE AND CONTOUR. (A) UNDERCUT AND OVERLAPPING IN A FILLET WELD. (B) UNDERCUT AND OVERLAPPING IN A GROOVE WELD. (C) AND (D) UNDERFILL IN GROOVE WELDS

Cracks can occur in a wide variety of shapes and types and can be located in numerous positions in and around a welded joint (Fig. 6). Cracks associated with welding can be categorized according to whether they originate in the weld itself or in the base metal. Four types commonly occur in the weld metal: transverse, longitudinal, crater, and hat cracks. Base metal cracks can be divided into seven categories: transverse cracks, underbead cracks, toe cracks, root cracks, lamellar tearing, delaminations, and fusion-line cracks.

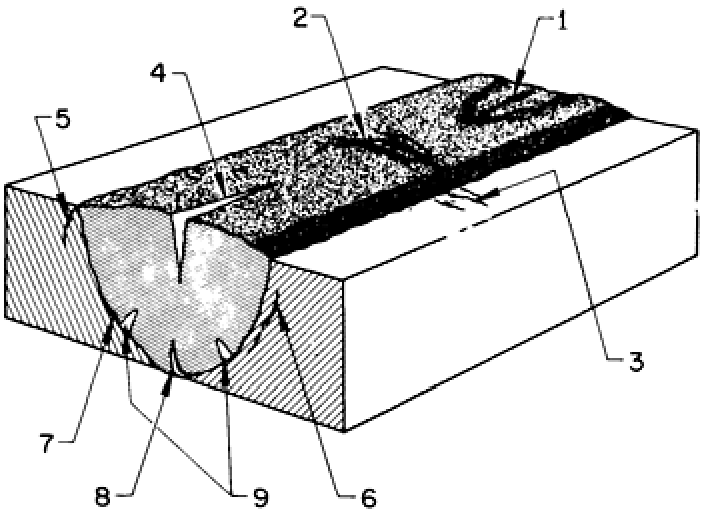

FIG. 6 IDENTIFICATION OF CRACKS ACCORDING TO LOCATION IN WELD AND BASE METAL. 1, CRATER CRACK IN WELD METAL; 2, TRANSVERSE CRACK IN WELD METAL; 3, TRANSVERSE CRACK IN HAZ;, 4, LONGITUDINAL CRACK IN WELD METAL; 5, TOE CRACK IN BASE METAL; 6, UNDERBEAD CRACK IN BASE METAL; 7, FUSION-LINE CRACK; 8, ROOT CRACK IN WELD METAL; 9, HAT CRACKS IN WELD METAL

Weld metal cracks and base metal cracks that extend to the surface can be detected by liquid penetrant and magnetic particle inspection. Magnetic particle inspection can also detect subsurface cracks, depending on their size, shape, and proximity to the surface. Although the orientation of a crack with respect to the direction of the radiation beam is the dominant factor in determining the ability of radiography to detect the crack, differences in composition between the base metal and the weld metal may create shadows to hide a crack that otherwise might be visible. Ultrasonic inspection is generally effective in detecting most cracks in the weld zone.

Transverse cracks in weld metal (No. 2, Fig. 6) are formed when the predominant contraction stresses are in the direction of the weld axis. They can be hot cracks, which separate intergranularly as the result of hot shortness or localized planar shrinkage, or they can be transgranular separations produced by stresses exceeding the strength of the material. Transverse cracks lie in a plane normal to the axis of the weld and are usually open to the surface. They usually extend across the entire face of the weld and sometimes propagate into the base metal.

Transverse cracks in base metal (No. 3, Fig. 6) occur on the surface in or near the HAZ. They are the result of the high residual stresses induced by thermal cycling during welding. High hardness, excessive restraint, and the presence of hydrogen promote their formation. Such cracks propagate into the weld or beyond the HAZ into the base metal as far as is needed to relieve the residual stresses.

Underbead cracks (No. 6, Fig. 6) are similar to transverse cracks in that they form in the HAZ because of high hardness, excessive restraint, and the presence of hydrogen. Their orientation follows the contour of the HAZ.

Longitudinal cracks (No. 4, Fig. 6) may exist in three forms, depending on their positions in the weld. Check cracks are open to the surface and extend only partway through the weld. Root cracks extend from the root to some point within the weld. Full centerline cracks may extend from the root to the face of the weld metal.

Check cracks are caused either by high contraction stresses in the final passes applied to a weld joint or by a hot-cracking mechanism.

Root cracks are the most common form of longitudinal weld metal crack because of the relatively small size of the root pass. If such cracks are not removed, they can propagate through the weld as subsequent passes are applied. This is the usual mechanism by which full centerline cracks are formed.

Centerline cracks may occur at either high or low temperatures. At low temperatures, cracking is generally the result of poor fit-up, overly rigid fit-up, or a small ratio of weld metal to base metal.

All three types of longitudinal cracks are usually oriented perpendicular to the weld face and run along the plane that bisects the welded joint. Seldom are they open at the edge of the joint face, because this requires a fillet weld with an extremely convex bead.

Crater cracks (No. 1, Fig. 6) are related to centerline cracks. As the name implies, crater cracks occur in the weld crater formed at the end of a welding pass. Generally, this type of crack is caused by failure to fill the crater before breaking the arc. When this happens, the outer edges of the crater cool rapidly, producing stresses sufficient to crack the interior of the crater. This type of crack may be oriented longitudinally or transversely or may occur as a number of intersecting cracks forming the shape of a star. Longitudinal crater cracks can propagate along the axis of the weld to form a centerline crack. In addition, such cracks may propagate upward through the weld if they are not removed before subsequent passes are applied.

Hat cracks (No. 9, Fig. 6) derive their name from the shape of the weld cross section with which they are usually associated. This type of weld flares out near the weld face, resembling an inverted top hat. Hat cracks are the result of excessive voltage or welding speed. The cracks are located about halfway up through the weld and extend into the weld metal from the fusion line of the joint.

Toe and root cracks (No. 5 and 8, Fig. 6) can occur at the notches present at notch locations in the weld when high residual stresses are present. Both toe and root cracks propagate through the brittle HAZ before they are arrested in more ductile regions of the base metal. Characteristically, they are oriented almost perpendicular to the base metal surface and run parallel to the weld axis.

Lamellar tearing is the phenomenon that occurs in T-joints that are fillet welded on both sides. This condition, which occurs in the base metal or HAZ of restrained weld joints, is characterized by a steplike crack parallel to the rolling plane. The crack originates internally because of tensile strains produced by the contraction of the weld metal and the surrounding HAZ during cooling. Figure 7 shows a typical condition.

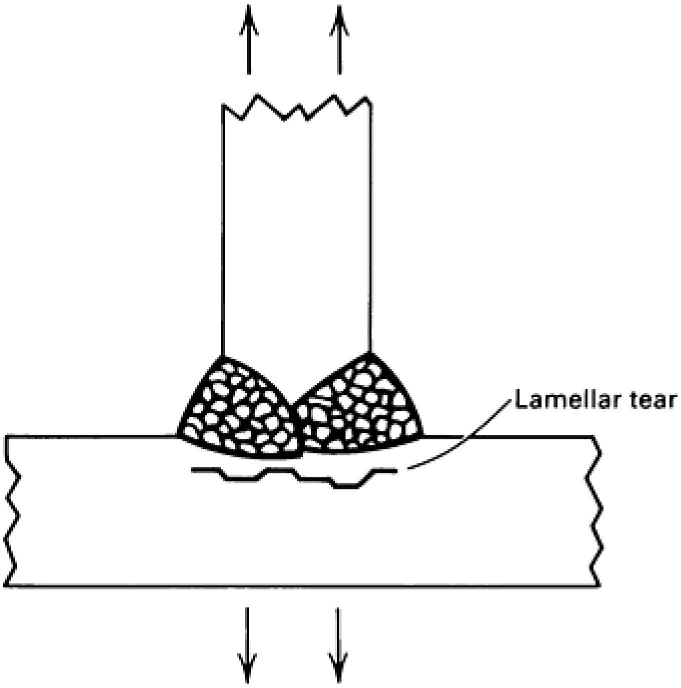

FIG. 7 LAMELLAR TEAR CAUSED BY THERMAL CONTRACTION STRAIN

Fusion-line cracks (No. 7, Fig. 6) can be classified as either weld metal cracks or base metal cracks because they occur along the fusion line between the two. There are no limitations as to where along the fusion line these cracks can occur or how far around the weld they can extend.

Discontinuities Associated With Specialized Welding Processes

The preceding section has dealt mainly with the discontinuities common to conventional arc-welding processes. In addition, there are certain more specialized welding methods that may have discontinuities unique to them. These methods include electron-beam, plasma arc, electroslag, friction, and resistance welding. In general, the types of discontinuities associated with these processes are the same as those associated with conventional arc welding; however, because of the nature of the processes and the joint configurations involved, such discontinuities may be oriented differently from those previously described, or they may present particular problems of location and evaluation.

Electron-Beam Welding

In electron-beam welding, as in all other welding processes, weld discontinuities can be divided into two major categories:

- THOSE THAT OCCUR AT, OR ARE OPEN TO, THE SURFACE
- THOSE THAT OCCUR BELOW THE SURFACE

Surface flaws include undercut, mismatch, underfill, reinforcement, cracks, missed seams, and LOP. Subsurface flaws include porosity, massive voids, bursts, cracks, missed seams, and LOP. Figure 8 shows poor welds containing these flaws, and a good weld with none of them.

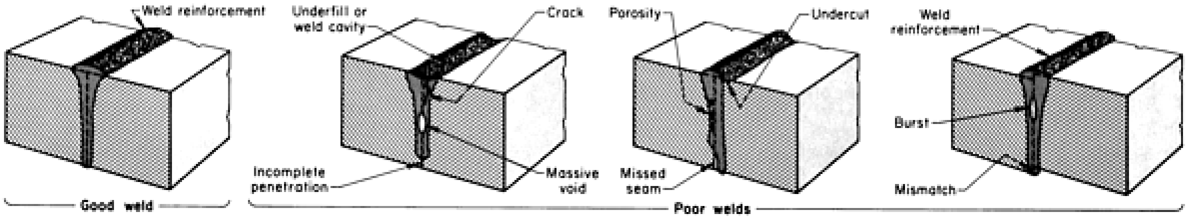

FIG. 8 ELECTRON-BEAM WELDS SHOWING FLAWS THAT CAN OCCUR IN POOR WELDS AND THE ABSENCE OF FLAWS IN A GOOD WELD WITH REINFORCEMENT

Surface discontinuities such as undercut, mismatch, reinforcement, and underfill are macroscopic discontinuities related to the contour of the weld bead or the joint. As such, they are readily detected visually or dimensionally. Surface discontinuities such as cracks are usually detected visually using liquid penetrant inspection or using magnetic particle inspection if the material is ferromagnetic.

When liquid penetrants are used to inspect a weld for surface discontinuities such as cracks, missed seams, and LOP, the surface to be inspected must be clean and the layers of metal smeared from machining or peened from grit- or sandblasting must be removed. Generally, some type of etching or pickling treatment works well, but the possibility of hydrogen pickup from the treatment must be considered.

Occasionally, special inspection procedures must be employed to detect some types of surface discontinuities in electron-beam welds. Missed seams and LOP are often difficult to detect because they are frequently associated with complex weld joints that prevent direct viewing of the affected surface.

Because of this difficulty, missed seams are often detected using a visual witness-line procedure in which equally spaced parallel lines are scribed on both sides of the unwelded joint at the crown and root surfaces. Missed seams, which result from misalignment of the electron beam with the joint such that the fusion zone fails to encompass the entire joint, are detected by observing the number of witness lines remaining on either side of the weld bead. By establishing the relationship between the width of the fusion zone and the spacing of the witness lines, reasonably accurate criteria for determining whether the joint has been contained within the weld path (and therefore whether missed seams are present) can be developed.

Lack-of-penetration discontinuities occur when the fusion zone fails to penetrate through the entire joint thickness, resulting in an unbonded area near the root of the joint. These discontinuities are best detected by etching the root surface and observing the macroscopic shape and width of the fusion zone for full and even penetration.

An alternative method for inspecting complete weld penetration is that of immersion pulse-echo ultrasonic testing. The planet gear carrier assembly shown in Fig. 9(a) consists of three decks (plates) and eight curved spacer sections. The assembly, which is made from SAE 15B22M boron-treated structural steel, is held together by 16 welds. Weld integrity is monitored by ultrasonic methods.

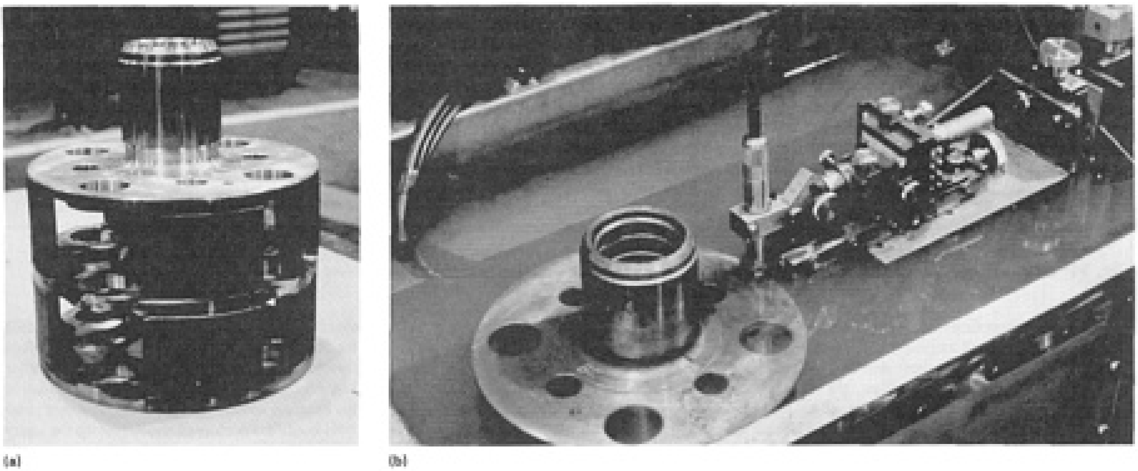

FIG. 9 PLANET GEAR CARRIER ASSEMBLY (A) SHOWING THE FOUR WELDS IN ONE STACK THAT CONNECT THE THREE DECKS. THE TOP AND TOP-CENTER WELDS ARE TESTED BY THE TOP ULTRASONIC TRANSDUCER, AND THE BOTTOM AND BOTTOM-CENTER WELDS ARE INSPECTED BY THE BOTTOM TRANSDUCER. (B) CLOSE-UP OF UPPER TRANSDUCER IN POSITION TO TEST THE WELDS. COURTESY OF CATERPILLAR, INC.

The parts are mounted on a turntable on a locating fixture so that the welds along the outside diameter are accessible. Two transducers, located above and below the assembly, are used. Figure 9(b) shows the upper transducer in position to test the welds. An overall view of the tank, turntable, controls, and reject/accept light panel is shown in Fig. 10. The use of such a system involves little downtime and enables a high quality level to be maintained. More information on ultrasonic test methods is presented later in this article.

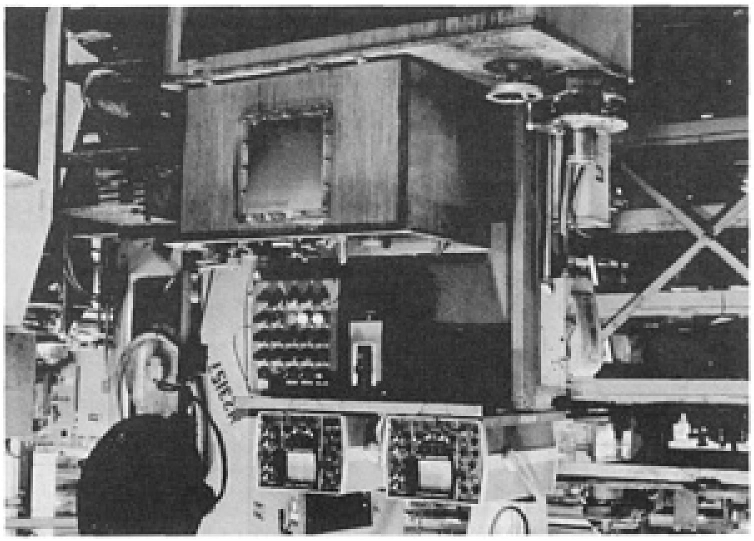

FIG. 10 OVERALL VIEW OF THE ULTRASONIC UNIT USED TO TEST THE ELECTRON-BEAM WELDED ASSEMBLY

SHOWN IN FIG. 9. COURTESY OF CATERPILLAR, INC.

Subsurface discontinuities in electron-beam welds are generally considerably more difficult to detect than surface discontinuities because observation is indirect. The two most reliable and widely used nondestructive inspection methods are radiography and ultrasonics.

Volume-type discontinuities such as porosity, voids, and bursts are detected by radiographic inspection, provided their cross sections presented to the radiating beam exceed 1 to 2% of the beam path in the metal. Discontinuities that present extremely thin cross sections to the beam path, such as cracks, missed seams, and LOP, are detectable with x-rays only if they are viewed from the end along their planar dimensions.

Ultrasonic inspection can detect most volume discontinuities as well as planar discontinuities. Planar discontinuities are best detected normal to the plane of the discontinuities, but missed seams and LOP often appear as continuous porosity when viewed looking down from the crown to the root of the weld in the plane of the discontinuity.

Because of the inherent dependence of both radiographic and ultrasonic inspection on the shape and orientation of flaws because each of the two methods can generally detect those flaws that the other misses, it is most advisable to complement one method with the other. Furthermore, to increase the likelihood of properly viewing a flaw, one of the methods should be employed in at least two (preferably perpendicular) directions.

Plasma Arc Welding

Discontinuities that occur in plasma arc welds include both surface and subsurface types, as shown in Fig. 11.

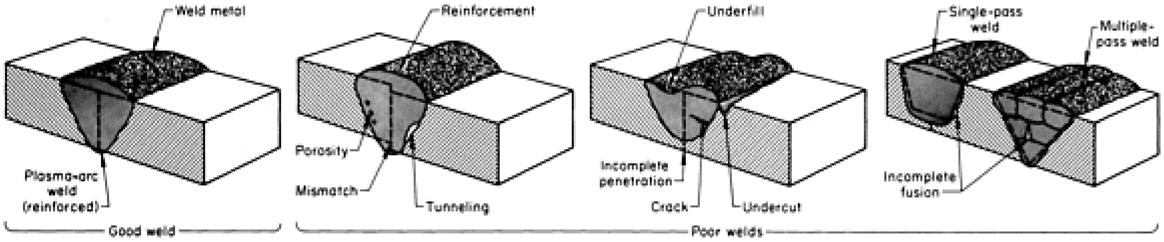

FIG. 11 PLASMA ARC WELDS SHOWING FLAWS THAT CAN OCCUR IN POOR WELDS AND THE ABSENCE OF FLAWS IN GOOD REINFORCED WELD

Surface discontinuities such as irregular reinforcement, underfill, undercut, and mismatch that are associated with weld bead contour and joint alignment are easily detected visually or dimensionally. Lack of penetration is also detected visually through the absence of a root bead. Weld cracks that are open to the surface are detected with liquid penetrants. Surface contamination, which results from insufficient shielding-gas coverage, is detected by the severe discoloration of the weld bead or adjacent HAZ.

Subsurface discontinuities are generally more prevalent in manual than in automatic plasma arc welding and are detected primarily by radiographic or ultrasonic inspection.

Porosity is by far the most commonly encountered discontinuity. Radiographic inspection is limited to detecting pores greater than approximately 1 to 2% of the joint thickness. Visibility is greater if both the crown and root beads are machined flush. Ultrasonic inspection can detect porosity if the joint is machined flush and joint thickness exceeds approximately 1.3 mm (0.050 in.).

Tunneling, as shown in Fig. 11, is a severe void along the boundary of the fusion zone and the HAZ. This discontinuity results from a combination of torch alignment and welding conditions (particularly travel speed). Tunneling is readily detectable by radiographic inspection.

Lack-of-fusion discontinuities occur in either single-pass or multiple-pass repair welds (Fig. 11). These discontinuities result from insufficient heat input to permit complete fusion of a particular weld bead to the part. Incomplete fusion can be detected by radiographic or ultrasonic inspection. Depending on the orientation of the discontinuity, one method may have an advantage over the other, so both should be used for optimum inspection.

Subsurface weld cracks, regardless of their cause, are detectable by radiographic and ultrasonic inspection.

Subsurface contamination in plasma arc welding results when copper from the torch nozzle is expelled into the weld. This is caused by excessive heat, usually produced in manual repair welding when the torch nozzle is placed too close to the weld, particularly in a groove. The resulting contamination, which may be detrimental, is undetectable by conventional nondestructive inspection methods. The only way of detecting copper contamination is by alerting the operator to watch for copper expulsion, which then must be machined off.

Electroslag Welding

Electroslag welding involves the use of copper dams over the open surfaces of a butt joint to hold the molten metal and the slag layer as the joint is built up vertically. Wire is fed into the slag layer continuously and is melted by the heat generated as current passes through the highly resistant slag layer.

Generally, electroslag welds are inspected with the same nondestructive examination methods as other heavy-section welds. With the exception of procedure qualification, all testing is nondestructive because of the sizes used. Techniques such as radiography and ultrasonic inspection are most often used, while visual, magnetic particle, and liquid penetrant testing are used also. Internal defects are generally more serious. Radiography and ultrasonic tests are the best methods for locating internal discontinuities.

Because of the nature of the process, LOF is rare. If fusion is achieved on external material edges, then fusion is generally complete throughout. Cracking may occur either in the weld or the HAZ. Porosity may either take the form of a rounded or a piped shape; the latter is often called wormhole porosity. Ultrasonic inspection is probably the quickest single method for inspecting any large weldment. If defects should occur, they appear as porosity or centerline cracking. Ultrasonic inspection is effective for locating either type of detect; however, only well-qualified personnel should set up the equipment and interpret the test results.

Electroslag welding results in large dendritic grain sizes because of the slow cooling rate. Inexperienced personnel often use high sensitivity and actually pick up the large coarse grains; when such welds are sectioned, usually no defects are present. Inspectors must learn to use low sensitivity to obtain good results when inspecting electroslag welds. Magnetic particle inspection is not a particularly good inspection method, because the areas examined by this technique are primarily surface or near surface. This is only a small percentage of the total weld; the only useful information is either checking the ends for craters, cracks, or centerline cracking or possibly for lack of edge fusion on the weld faces. Usually, a visual examination gives the same result unless the defect is subsurface. Visual examinations are only effective for surface defects, which are not common in this process.

Friction Welding

If impurities are properly dispelled during upsetting, friction or inertia welds are generally free of voids and inclusions. Incomplete center fusion can occur when flywheel speed is too low, when the amount of upset is insufficient, or when mating surfaces are concave. Tearing in the HAZ can be caused by low flywheel speed or excessive flywheel size. Cracks can occur when materials that are prone to hot shortness are joined. The penetration of a split between the extruded flash into the workpiece cross section is most prevalent during the welding of thin-wall tubing using improper conditions that do not allow for sufficient material upset.

The area where LOF generally occurs is at or near the center of the weld cross section. Because this is a subsurface discontinuity, detection is limited to radiographic or ultrasonic inspection; ultrasonic inspection is more practical. The longitudinal wave test (either manual-contact or immersion method) with beam propagation perpendicular to the area of LOF gives the most reliable results. This test can be performed as long as one end of the workpiece is accessible to the transducer.

Penetration of the split between the extruded flash on the outer surface of a tube is readily detected by liquid penetrant or magnetic particle inspection after the flash has been removed by machining. A split between the weld flash on the inner surface of the tube can be detected by ultrasonic inspection using the angle-beam technique with manual contact of the transducer to the outside surface of the tube. The transducer contacts the tube so that the sound propagates along the longitudinal axis of the tube through the weldment.

Resistance Welding

Resistance welding encompasses spot, seam, and projection welding, each of which involves the joining of metals by passing current from one side of the joint to the other. The types of discontinuities found in resistance welds include porosity, LOF, and cracks. Porosity will generally be found on the centerline of the weld nugget. Lack of fusion may also be manifested as a centerline cavity. Either of these can be caused by overheating, inadequate pressure, premature release of pressure, or late application of pressure. Cracks may be induced by overheating, removal of pressure before weld quenching is completed, improper loading, poor joint fit-up, or expulsion of excess metal from the weld.

Weld Appearance. On the surface of a resistance spot welded assembly, the weld spot should be uniform in shape and relatively smooth, and it should be free of surface fusion, deep electrode indentations, electrode deposits, pits, cracks, sheet separation, abnormal discoloration around the weld, or other conditions indicating improper maintenance of electrodes or functioning of equipment. However, surface appearance is not always a good indicator of spot weld quality, because shunting and other causes of insufficient heating or incomplete penetration usually leave no visible effects on the workpiece.

The common practice for monitoring spot weld quality in manufacturing operations is the teardown method augmented with pry testing and visual inspection (Ref 3). In visual inspection, the operator uses the physical features of the weld surface, such as coloration, indentation, and smoothness, for assessing the quality of the weld. In pry testing, a wedge-shaped tool is inserted between the metal sheets next to the accessible welds, and a prying action is performed to see if the sheets will separate in the weld zone. The teardown method consists of physically tearing apart the welded members with hammers and chisels to determine the presence and size adequacy of fused metal nuggets at the spot weld site. The specifications require that the parent metal be torn and that the weld nugget remain intact. The destructiveness and/or inadequacy of these common inspection methods has long been recognized, and as a result, nondestructive methods have been extensively studied. The pulse-echo ultrasonic inspection of spot welds is now feasible.

Diffusion Welding

Diffusion welding (also known as diffusion bonding) is a metal joining process that requires the application of controlled pressures at elevated temperatures and usually a protective atmosphere to prevent oxidation. No melting and only limited macroscopic deformation or relative motion between the faying surfaces of the parts occur during bonding. As such, the principal mechanism for joint formation is solid-state diffusion. A diffusion aid (filler metal) may or may not be used. Diffusion welding has the advantage of producing a product finished to size, with joint efficiencies approaching 100%.

Discontinuities in Diffusion Welds. In the case of fusion welds, the detection of discontinuities less than 1 mm (0.04 in.) in size is not generally expected. In diffusion welding, in which no major lack of bonding occurs, individual discontinuities may be only micrometers in size. To understand how discontinuities form in diffusion-welded structures, it is first necessary to consider the principles of the process.

As illustrated in Fig. 12, metal surfaces have several general characteristics:

- ROUGHNESS
- AN OXIDIZED OR OTHERWISE CHEMICALLY REACTED AND ADHERENT LAYER
- OTHER RANDOMLY DISTRIBUTED SOLID OR LIQUID PRODUCTS SUCH AS OIL, GREASE, AND DIRT
- ADSORBED GAS, MOISTURE, OR BOTH

Because of these characteristics, two necessary conditions that must be met before a satisfactory diffusion weld can be made are:

- MECHANICAL INTIMACY OF METAL-TO-METAL CONTACT MUST BE ACHIEVED
- INTERFERING SURFACE CONTAMINANTS MUST BE DISRUPTED AND DISPERSED TO

PERMIT METALLIC BONDING TO OCCUR (SOLVENT CLEANING AND INERT GAS ATMOSPHERES CAN REDUCE OR ELIMINATE PROBLEMS ASSOCIATED WITH SURFACE CONTAMINATION AND OXIDE FORMATION, RESPECTIVELY)

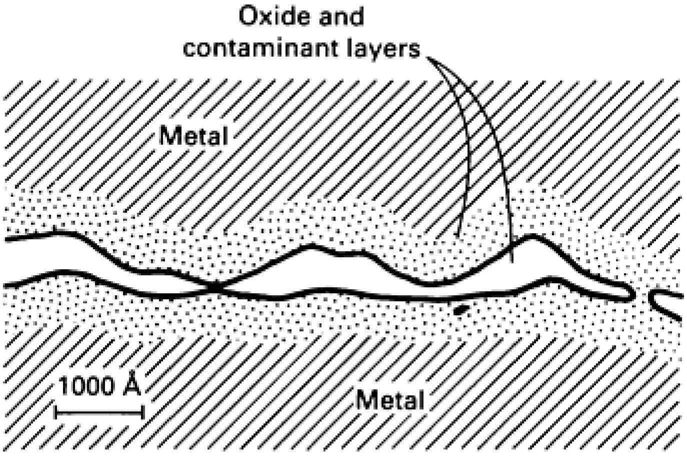

FIG. 12 CHARACTERISTICS OF A METAL SURFACE SHOWING ROUGHNESS AND CONTAMINANTS PRESENT. SOURCE: REF 7

For a given set of processing parameters, surface roughness is probably the most important variable influencing the quality of diffusion-welded joints. The size of the discontinuities (voids) is principally determined by the scale of roughness of the surfaces being bonded. The degree of surface roughness is dependent on the material and fabrication/machining technique used. It has been shown that bonding becomes easier with finer surface roughness prior to bonding. Figure 13 compares the surface roughness values produced by a variety of fabrication methods. More detailed information on surface roughness can be found in the article "Surface Finish and Surface Integrity" in *Machining,* Volume 16 of the *ASM Handbook.*

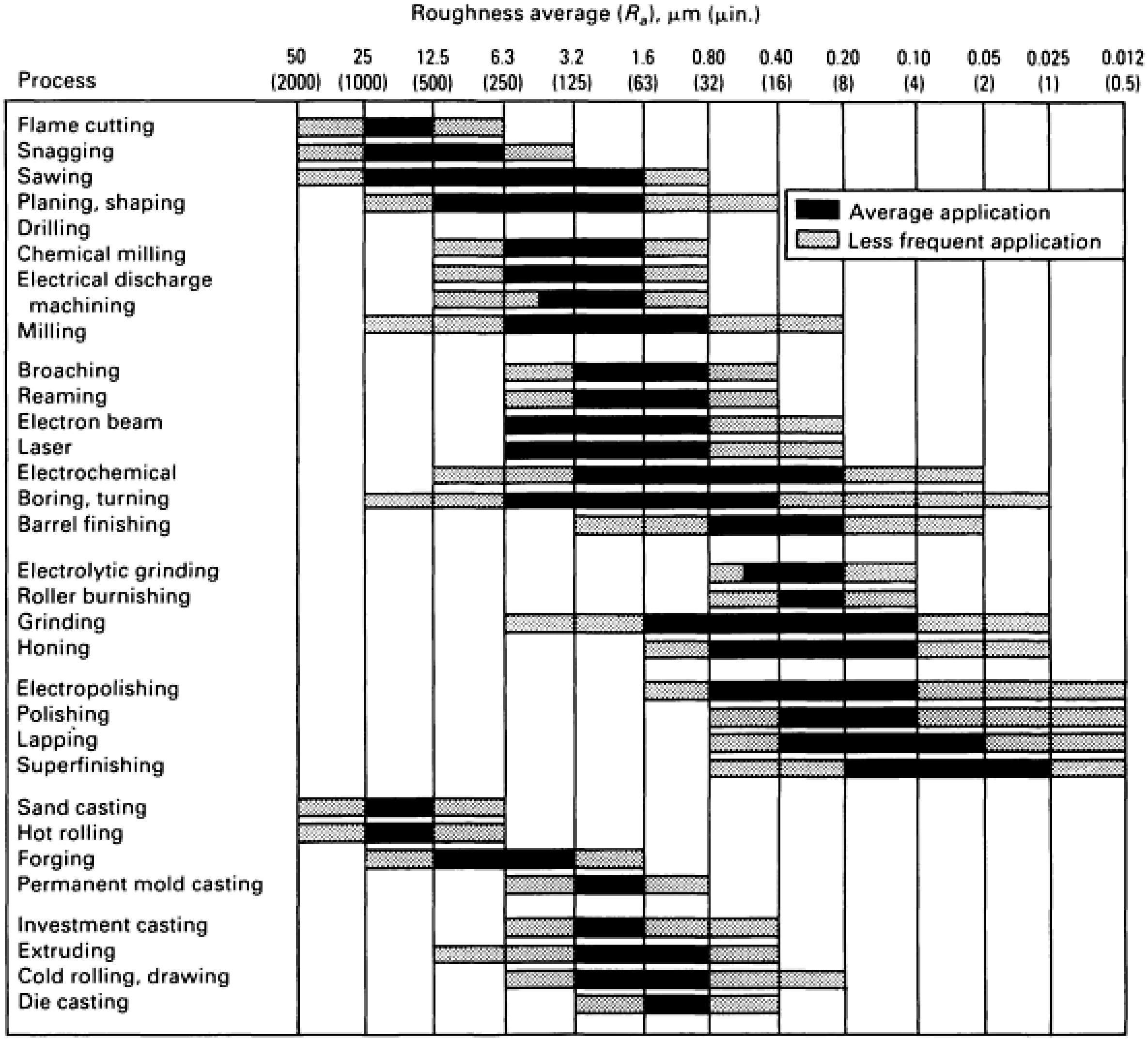

FIG. 13 SURFACE ROUGHNESS PRODUCED BY COMMON PRODUCTION METHODS. THE RANGES SHOWN ARE TYPICAL OF THE PROCESSES LISTED. HIGHER OR LOWER VALUES CAN BE OBTAINED UNDER SPECIAL CONDITIONS.

The mechanism of bond formation in diffusion welding is believed to be the deformation of the surface roughness in order to cause metal-to-metal contact at asperities, followed by the removal of interfacial voids and cracks by diffusional and creep processes. For conventional diffusion bonding without a diffusion aid, the three-stage mechanistic model shown in Fig. 14 describes bond formation. In the first stage, deformation of the contacting asperities occurs primarily by yielding and by creep deformation mechanisms to produce intimate contact over a large fraction of the interfacial area. At the end of this stage, the joint is essentially a grain boundary at the areas of contact with voids between these areas. During the second stage, diffusion becomes more important than deformation, and many of the voids disappear as the grain-boundary diffusion of atoms continues. Simultaneously, the interfacial grain boundary migrates to an equilibrium configuration away from the original plane of the joint, leaving many of the remaining voids within the grains. In the third stage, the remaining voids are eliminated by the volume diffusion of atoms to the void surface (equivalent to diffusion of vacancies away from the void). Successful completion of stage three is dependent on proper surface processing and joint processing.

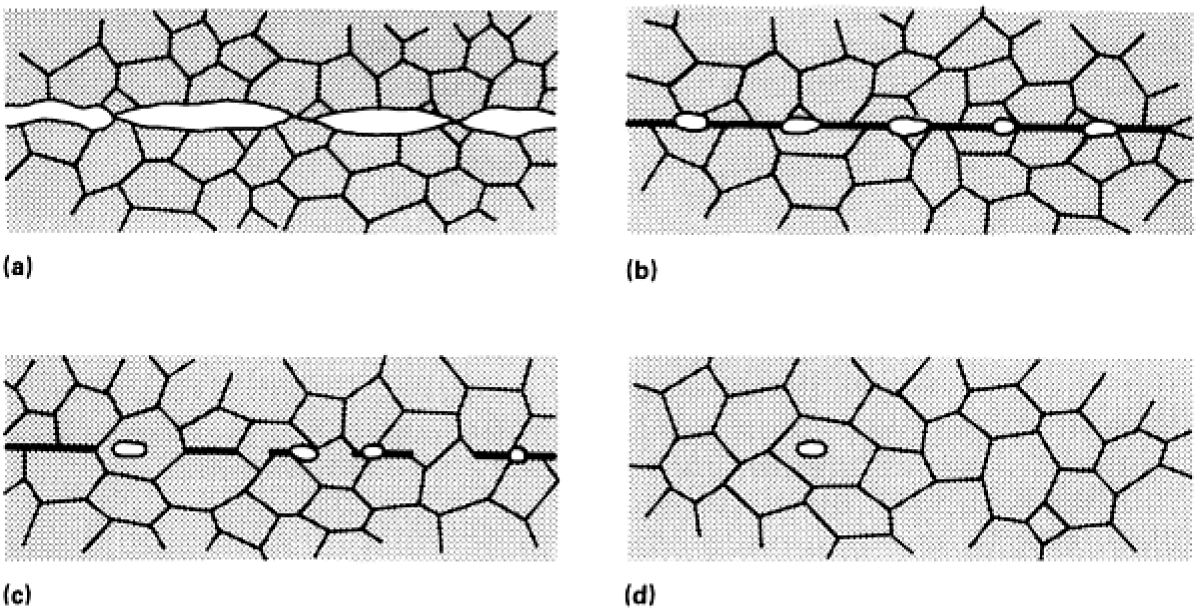

FIG. 14 THREE-STAGE MECHANISTIC MODEL OF DIFFUSION WELDING. (A) INITIAL ASPERITY CONTACT. (B) FIRST-STAGE DEFORMATION AND INTERFACIAL BOUNDARY FORMATION. (C) SECOND-STAGE GRAIN BOUNDARY MIGRATION AND PORE ELIMINATION. (D) THIRD-STAGE VOLUME DIFFUSION AND PORE ELIMINATION. SOURCE: REF 7

Successful nondestructive evaluation of diffusion-welded joints requires that the maximum size and distribution of discontinuities be determined. However, many conventional NDE methods and equipment are not adequate for discontinuity determination. Fluorescent penetrants provide excellent detection capability for lack of bonding as long as the interfacial crack breaks the surface. They are completely ineffective for defects that have no path to the surface.

Conventional film radiography is not suitable for detecting the extremely small defects involved in diffusion welding, but the use of x-ray microfocus techniques coupled with digital image enhancement offers an improvement in resolution. Discontinuities as small as 50 μm (0.002 in.) have been detected.

Conventional ultrasonic testing has some applications, although only significant lack of bonding or clusters of smaller defects can be reliably detected. High-resolution flaw detection involving frequencies approaching 100 MHz appears to have distinct advantages over conventional testing. Scanning acoustic microscopy also appears to offer excellent possibilities in diffusion bond inspection. Eddy current and thermal methods are relatively unsatisfactory for most applications.

References cited in this section

3. T.M. MANSOUR, ULTRASONIC INSPECTION OF SPOT WELDS IN THIN-GAGE STEEL, *MATER. EVAL.*, VOL 46 (NO. 5), APRIL 1988, P 650-658
7. DIFFUSION WELDING AND BRAZING, *WELDING HANDBOOK*, 7TH ED., AMERICAN WELDING SOCIETY, 1980, P 311-335

References

1. ASME BOILER AND PRESSURE VESSEL CODE, APPENDIX IV, DRAFT 1, 21 OCT 1981
2. L.W. SANDOR, THE MEANING OF WELD DISCONTINUITIES IN SHIPBUILDING, *PROC. 1ST NAT.*

CONF. FITNESS-FOR-SERVICE IN SHIPBUILDING (BOULDER, CO), OCT 1980

3. T.M. MANSOUR, ULTRASONIC INSPECTION OF SPOT WELDS IN THIN-GAGE STEEL, *MATER. EVAL.,* VOL 46 (NO. 5), APRIL 1988, P 650-658
4. P. KAPRANOS AND R. PRIESTNER, NDE OF DIFFUSION BONDS, *MET. MATER.,* VOL 3 (NO. 4), APRIL 1987, P 194-198
5. P.G. PARTRIDGE, "DIFFUSION BONDING OF METALS," ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT (NATO), AUG 1987, P 5.1-5.23
6. G. TOBER AND S. ELZE, "ULTRASONIC TESTING TECHNIQUES FOR DIFFUSION-BONDED TITANIUM COMPONENTS," ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT (NATO), JULY 1986, P 11.1-11.10
7. DIFFUSION WELDING AND BRAZING, *WELDING HANDBOOK,* 7TH ED., AMERICAN WELDING SOCIETY, 1980, P 311-335

Inspection of Welded Joints

G.R. Edwards, The Welding Institute

Introduction

WELDED JOINTS in any component or structure require thorough inspection. The role of nondestructive evaluation (NDE) in the inspection of welds is very important, and the technology has become highly developed as a result. This article describes the applications, methods, and limitations of NDE, and provides guidance as to method selection.

Nondestructive evaluation comprises a range of test methods for detecting discontinuities in a material, component, or structure, without causing damage. Therefore, the principal advantages of nondestructive tests, as opposed to alternative destructive proof tests, are that a 100% inspection can be performed at manufacture and that monitoring of the structure or component can continue while it is in service. The principal disadvantage of NDE is that the measurements obtained are only indirectly related to the presence and severity of the flaws, and much subjective interpretation is necessary.

Any discussion of NDE requires a clear definition of the term "defect." A defect is a discontinuity that creates a substantial risk of failure in a component or structure during its service life. Although the aim of NDE is to detect defects, it provides evidence of flaws, as well. The significance of these flaws and whether or not they are actually defects must then be determined. The development of NDE procedures often includes a collaboration between the NDE, engineering, and design functions to investigate the significance of defects.

There are five principal test methods: penetrant testing, magnetic-particle testing, eddy-current testing, radiographic testing, and ultrasonic testing. Although visual inspection is not strictly a test method, it does have an important role in NDE procedures and is often the only inspection method used. Acoustic emission (AE) testing and leak testing are described separately, because they are somewhat different in character and are not used as widely. There are other NDE methods that have been omitted, because they are very specialized.

Applications

A useful distinction can be made between those NDE methods that are applied to welds during manufacturing and those that are applied to welds as part of plant and machinery maintenance. Although the same methods and techniques can be applied in both areas, there are important differences in inspection goals and, therefore, in test procedures.

In manufacturing, NDE is one of a series of quality-control techniques. The integration of these techniques into the manufacturing process is an important management subject. The first step is to decide at what stage or stages the NDE is to be carried out. Components that are brought in-house can be inspected prior to welding. An ultrasonic examination of a parent plate, for example, will show areas that contain laminations. Although these laminations may not be defects in themselves, lamellar tears could result where they occur near the weld in a specific orientation.

During the welding of thick sections, the magnetic-particle testing (MPT) of the root-pass and hot pass, in an attempt to find cracks before filling the weld groove, could save on repair costs. Using dry powder techniques, MPT can be carried out while surfaces are still hot.

Attempts have been made to incorporate NDE techniques with continuous monitoring of the weld process. Eddy-current testing, for example, has been used on small-bore induction-welded tube to provide feedback to control the welding process. However, total automation with real-time evaluation of test results is still not generally available, except in very simple applications, where results can be evaluated with simple go/no-go acceptance criteria.

Nondestructive evaluation is usually conducted at the end of a manufacturing process. In that capacity, it functions as part of the specification requirements and thus provides a safeguard against sending defective components or structures to customers. Unfortunately, conducting NDE after welding has been completed is not satisfactory. The necessity of repairs has given NDE a negative image in the manufacturing process. The early implementation of NDE in the process and the consequent actions to reduce repair rates is a more positive use.

NDE techniques have been used increasingly in plant and machinery maintenance, as a result of numerous industrial trends. The trend toward improved safety of operation, particularly in aircraft and nuclear power plants, is an obvious example. In times of economic recession, the operating life of a plant tends to be extended, and NDE has been utilized with increasing frequency. Environmental issues also have provided a stimulus to NDE activity. The inspection of storage tanks at refineries, where corrosion needs to be detected before leaks develop, is a good example.

As already noted, the performance capability required of NDE during the manufacturing process can be quite different from that needed during plant maintenance. Generally, the sizing accuracy is more important in maintenance, whereas the reliability of detection is more important in manufacturing. Because of the inherent unreliability of many techniques, a wide tolerance between critical defect size and flaw detection threshold is needed in manufacturing. However, the incorporation of fitness-for-purpose philosophies into quality control is having the effect of reducing this tolerance and placing greater emphasis on NDE sizing accuracy in all applications.

Inspection of Welded Joints

G.R. Edwards, The Welding Institute

NDE Methods

A variety of NDE methods are briefly described below, along with some of the operating problems that can be encountered. The limitations of NDE methods are further described in the section "NDE Performance" in this article.

Visual inspection, although not itself a test method, is the most common NDE method and, arguably, the most important. Indeed, a thorough visual examination is a prerequisite of successful NDE. Much can be inferred from the surface of the weld, including information about its internal condition, although a good weld surface does not necessarily indicate a defect-free weld.

Mechanical gages are available for measuring such dimensions as the height of the weld bead, the depth of the undercut, and the weld size (Fig. 1). When surfaces are not accessible, it may be possible to use boroscopes and flexible endoscopes to view the weld. Their use does require some familiarization, because they give a very flat field of view in which important features can be lost unless the surface is illuminated carefully.

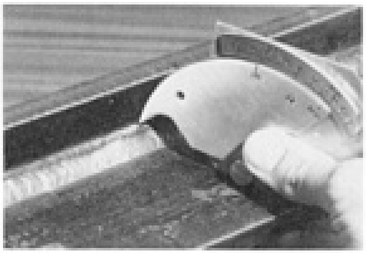

FIG. 1 WELD GAGE

Penetrant Testing. Penetrants provide a method of enhancing the visibility of flaws. A strongly colored, or fluorescent, liquid is applied liberally to the test surface and left to infiltrate surface-breaking cavities and cracks. The surface contact time is typically about 15 min. The excess liquid is then wiped from the surface, using some solvent, and a fine white developer powder is applied, usually as an aerosol in a volatile solvent. As the solvent evaporates, the penetrant liquid that is trapped in the flaws is drawn into the dry developer powder by a reverse-capillary action. For a cavity, this process may take no more than 1 s, but for a fine hairline crack, it may be necessary to wait more than 1 h for a visible indication.

The several penetrant types that exist can be distinguished on the basis of color, method of removal, and the form of the developer. Red penetrants, which are the most commonly used, require good daylight viewing conditions and a white background to improve contrast. Fluorescent penetrants are used when high sensitivity to fine flaws on relatively smooth surfaces is needed. They require black-light viewing conditions.

Solvent-removable red penerrants are the most convenient for site use. They can be provided in three aerosol cans of penetrant, cleaning fluid, and developer. Unfortunately, solvents are a health hazard and should be applied only in well-ventilated areas. A safer alternative is the water-washable penetrant.

When the geometry of the test piece is too complicated for an even coating to be applied using a spray, the developer can be applied directly as a dry powder. A dry powder is more suited to fluorescent penetrants, because it does not provide a white background, and the thickness of the coating reduces the blotting action and sensitivity, but improves image definition.

The compatibility of the penetrant with the test material must be considered. Residues of penetrant can be difficult to remove after the test. Traces of chlorine will corrode stainless steels and hydroxides will attack aluminum. The trace element content is provided by the manufacturer of the penetrant.

Although it is the most common NDE method, penetrant testing is widely misused. The test surfaces are often inadequately cleaned, the penetrant contact time with the surface is too short, or the excess penetrant is removed carelessly from flaws, as well as from the surface. In addition, because the method only detects flaws that are open to the surface, abrasive cleaning methods may peen over fine, surface-breaking cracks. This problem can arise when penetrants are used to chase fine hairline cracks or incomplete fusion with grinding.

Magnetic particle testing is the preferred method for detecting surface-breaking and, under certain circumstances, near-surface flaws in situations where a test material can be magnetized.

This method depends on the flaw disrupting the magnetic flux generated along the surface of the weld by a permanent magnet or electromagnets, or by electric-current-carrying electrodes and cables. Flux leakage from the surface creates magnetic poles, which attract the magnetic particles, creating a clearly visible image of the flaw. The method is particularly sensitive to the detection of cracks, where, like penetrants, the particles identify flaws with well-defined features. However, the flux leakage may be absent altogether, if the crack runs parallel with the magnetic field. It is therefore vital that the inspection surface be magnetized in two directions at right angles to each other.

Important considerations associated with MPT are the method of magnetization, the detecting medium, and the ways in which the field direction and strength can be measured. A very convenient method of magnetization is to use permanent magnets, which do not need a power source. Because the magnetic field is in a constant direction and does not alternate, it penetrates the surface to give some sensitivity to subsurface flaws. However, permanent magnets are difficult to manipulate in confined spaces. If proper contact between the feet of the magnet and the test surface is not achieved, then the density of magnetic flux in the test piece is weakened, and the test sensitivity is reduced.

Electromagnets are easier to use and can generate magnetic fields with alternating current (ac) (Fig. 2). The skin effect caused by ac concentrates the flux at the surface, improving the sensitivity to surface-breaking flaws, particularly on contoured surfaces.

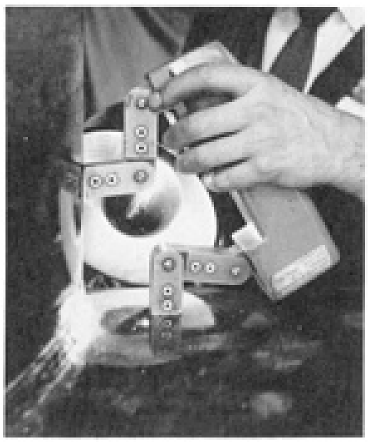

FIG. 2 MAGNETIC-PARTICLE TESTING OF A WELD WITH AN ELECTROMAGNET

Prod magnetization, in which electrodes carrying a high-amperage, low-voltage current are pressed onto the test surface to induce a distorted, circular magnetic field, provides the most controlled method of magnetization. Although the electrodes can be manipulated more easily into tight corners, they do have two major disadvantages, which can preclude their usefulness. First, the arc strikes that can occur at the prod contact points can damage the test surface. Second, the electrodes may require a two-man operation.

Cable magnetization is common in diver inspections of underwater node welds on oil platforms. The cables are wrapped around the pipe adjacent to the weld. The induced field, which is directed across the weld, is sensitive to longitudinal cracks only.

The fine particles that are applied to the magnetized surface to detect the flux-leakage fields must be highly magnetic. Materials commonly used are black iron and red or yellow iron oxides. The iron particles can be coated with a fluorescent material and viewed in black light. Before black particles are applied, the surface can be coated with a thin, white lacquer to improve contrast. Although the particles are typically sprayed on the surface as a suspension in water or solvent, they also can be applied as a dry powder. Dry powders can be used on hot surfaces or in cases where contamination is an issue.

To ensure that the magnetic flux is adequate, indicators such as the Berthold gage can be placed on the test surface. This type of indicator contains fine graticules of a highly magnetic material set in a nonmagnetic base. The graticules are made visible when the particles are applied and the magnetic field is of sufficient strength in a right-angled direction to the graticule. However, the indicators are influenced more by the tangential field across the test surface than by the magnetic field within the test piece. Therefore, as an indicator of test sensitivity, they should be treated with circumspection. It is possible to see defects, even when the indicator has been placed on a nonmagnetic material, where there is obviously no defect sensitivity at all.

The MPT process can leave residual fields that interfere with welding processes, Demagnetization can be achieved by slowly wiping the surface with an ac yoke.

Eddy-current testing techniques have been used widely in automated tube-testing systems and material sorting, as well as in the inspection of aircraft components and structures. For the testing of welds, there is a growing interest in using eddy-current testing as an alternative to MPT and penetrant testing, because the method is able to detect cracks beneath a thick layer of paint.

Like MPT, eddy-current testing relies on the flaw disrupting the energizing field, which is electrical, circular, and induced by a small coil carrying high-frequency alternating current. Therefore, the orientation of vertical planar flaws is not important, but the coil has to be scanned intensively over the test surface to ensure coverage. Laminar flaws are not detectable.

The method relies on detecting changes in the impedance of a coil that are caused by deflections in the eddy-current field around a flaw. The eddy-current field is affected by changes in material properties, surface geometry, edges, and many other factors that are unrelated to flaws. For this reason, the analysis of eddy-current test results can be difficult. The important components of the test equipment are the instrument, the probe containing the coil, and the calibration block containing slots or other artificial flaws.

Meter-reading instruments may be adequate for simple tests, but for more complex applications, such as weld testing, a cathode-ray tube with a vector point display is needed for analyzing the induced coil current (Fig. 3). The vector point provides information about the impedance of the coil. This is a complex quantity, which measures the amplitude and phase of the coil current. Certain directional movements in the vector point on the screen representing phase changes between coil current and voltage are associated with well-defined effects, such as lift-off effect (the up and down movement of the coil) and the edge effect (coil approaching an edge of the test piece). These can be distinguished from vector movements in other directions that are caused by cracks.

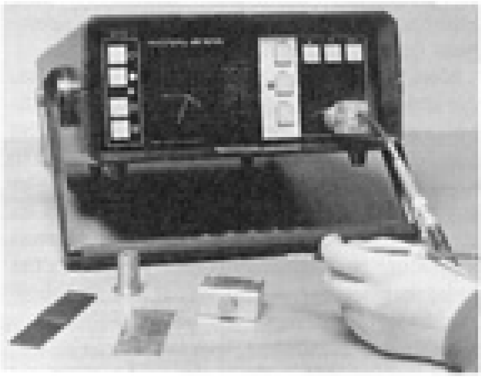

FIG. 3 EDDY-CURRENT TEST INSTRUMENT WITH VECTOR POINT DISPLAY

For weld testing, the coils are housed in pencil-like probes, which can be scanned along the weld cap. Coil arrangements can be varied to improve performance. Simple absolute coils, for example, are generally too noisy for testing welds. Alternative differential coil arrangements have been developed, which almost remove the lift-off noise completely as the probe scans the rough surface of the weld cap.

A measurement of crack depth is achieved by comparing the crack signal with the signal from a slot of known depth in a calibration block. This can be regarded only as a very approximate measure. It does not consider the crack width and inclination to the surface.

On the vector point display, the phase separation between lift-off and crack signals depends on the frequency of the coil current, as well as the coil design. The test frequency also determines the depth of penetration of the eddy-current field and, therefore, the sensitivity to subsurface flaws. At test frequencies below 1 kHz, the field can penetrate aluminum to a depth of several millimeters. However, for ferromagnetic materials, the overriding factor determining depth of penetration is the magnetic permeability, which restricts the eddy-current field to the surface unless it is removed by saturating the material with magnetism.

Radiographic testing, in the past, has formed the basis of weld inspection for internal flaws. However, it is an expensive method, and increasing restrictions on the use of ionizing radiation has led to a shift toward alternative ultrasonic testing techniques.

The radiographic technique uses radiation, either as x-rays or gamma rays, to penetrate the weld to create a latent image on radiographic film. The test piece absorbs radiation, but when flaws are present, less is absorbed than the amount absorbed by the parent material, which produces a localized darkening of the film. If the radiation source is sufficiently far from the object, then the image of the flaw will be sharp. Cavities with a through thickness of the order of only 1% of wall thickness can be discerned.

Most radiography is carried out with x-rays from tubes that operate in a voltage range between 120 and 400 kV. Higher voltages can be achieved using special equipment. As the thickness of the test object increases, the amount of kilovolts also must increase to keep the exposure time down to 4 to 6 min, typically, to avoid shortening the tube service life. However, high voltages reduce the image quality.

Gamma rays are produced by the radioactive decay of isotopes and do not need an electric power supply. Therefore, they are used extensively in site work. Iridium 192 is the usual isotope choice. However, for thick sections (72 in.), Cesium 137 and Cobalt 60 provide higher energy, whereas for thin sections, Ytterbium 169 is suitable. Generally, the quality of gamma radiographs is not as good as that of x-radiographs.

The test results are viewed in a radiograph. The radiographic films are chosen to achieve optimum definition and contrast within practicable exposure times. Faster films are needed where sections are thick, but at the sacrifice of image definition. Speed can be increased by replacing lead-intensifying screens that are in contact with the film with fluorescent ones, but this will result in a loss of definition.

The radiographs are viewed as negatives. The highest contrast is attained at image densities of the order of 3.0 on a log scale and, therefore, the radiographs can only be viewed in high-intensity transmitted light. If the image can be viewed with the film held up in front of normal light bulbs, then the density is much too low.

As an alternative to film radiography, where it can take 20 min to develop a radiograph that can be viewed, radioscopy offers images in real time. The images are created on x-ray-sensitive screens that can be intensified and scanned with a camera for visual display. However, screen resolution and contrast are not as high as with film radiography.

Obviously, taking a conventional radiograph necessitates access to both sides of the object. Thus, there is a variety of techniques for inspecting pipe welds. For pipe diameters that exceed 750 mm (30 in.) and allow access to the center, panoramic techniques are the most efficient (Fig. 4). For smaller pipes that do not allow internal access, both film and source must be on the outside. Exposure times are increased because the radiation must penetrate two walls. Inspection rates are reduced further, because the pipe has to be covered in multiple exposures.

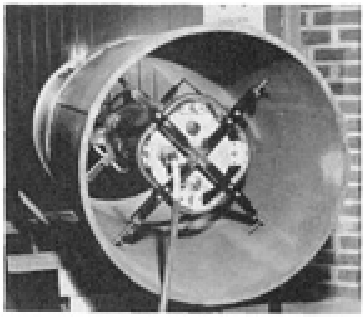

FIG. 4 RADIOGRAPHIC TESTING OF A PIPE WITH A PANORAMIC TECHNIQUE

If the pipe diameter is less than 75 mm (3 in.), then the number of exposures can be halved by putting the source at a sufficient distance away from the pipe to provide images of both the near and the far sides of the weld. The so-called double-wall, double-image technique is often used on welds in steam-generator pipes and condenser tubes.

Radiography can be used on welds in more complex geometries, such as nodes and T-joints, but only for detecting large volumetric flaws. Film cassettes that are double loaded with one fast and one slow film are used to accommodate the large variation in penetrated thickness across the weld.

The sensitivity of the radiograph is assessed using image quality indicators (IQIs), which can either be fine wires placed across the weld or plaques with small drilled holes placed adjacent to the weld. The dimension of the smallest visible hole or wire in the radiograph is a measure of its sensitivity. These are described in greater detail in the section "Sensitivity" in this article.

Ultrasonic testing receives more attention in the literature than any other NDE method, because of its use in critical applications, such as in the nuclear industry, and because it requires the highest level of skill. Like sonar, the method relies on propagating pulses of sound through a medium and then picking up reflected echoes. However, because the sound is of an ultrasonic frequency, the pulses are propagated along a narrow beam in a fixed direction, typically at refraction angles of 0, 45, 60, or 70° from the probe placed on the surface. To achieve coverage, the beam must therefore be scanned in a tight raster fashion over the test surface.

The surface of the flaw will have an important effect on the echo. Planar specular reflectors will only reflect the sound back to the probe if they lie in a plane perpendicular to the beam. Therefore, the detection of incomplete fusion in the sidewall of groove welds will require the judicious selection of probe beam angles. On the other hand, the echo amplitude is independent of beam direction for spherical reflectors, such as pores.

In conventional ultrasonic testing, the information collected by manually scanning the probe up to the weld and, where the cap has been removed, across the weld, is displayed on a cathoderay oscilloscope in an A-scan display (Fig. 5). This provides information about the amplitude of the echo and the distance of the reflector from the probe. Interpretation of the A-scan requires a high level of skill. It does not provide a permanent record of the test, which is, perhaps, the most important disadvantage of ultrasonic testing. On the other hand, the A-scan does provide information about the nature, size, and location of the flaw, which is further described in the section "NDE Performance" in this article.

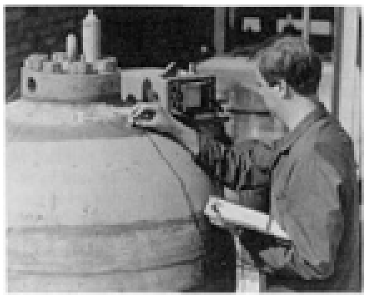

FIG. 5 ULTRASONIC EXAMINATION OF A NOZZLE WELD

The A-scan is viewed with the ultrasonic flaw-detector instrument, which has controls to adjust range, delay, and gain. Recently, instruments with digitally created A-scans have started to replace those based on analog. They are capable of making print-outs of the screen, but with some loss in resolution.

In addition to the beam angle mentioned above, probes can be selected according to ultrasound frequency, transducer diameter, and wave mode (shear or compression). Lower frequencies, such as 1 to 2.5 MHz, are used on thick sections.

An important aspect of ultrasonic testing is calibration. The time base of the A-scan must be calibrated for distance in millimeters or inches, and the amplitude of signals must be calibrated in dB, relative to some reference reflector. This is achieved using standard calibration blocks that are made from the same material as the test piece and contain reflectors at known distances and of known dimensions. To calibrate the amplitude of the A-scan, side-drilled holes are typically used in a range of distances. By observing the fall-off in signal amplitude with distance, the important distance-amplitude correction (DAC) can be applied to the A-scan.

Besides the flaw detector, probes, and calibration blocks, thick liquid couplants are needed to transmit the ultrasound from the probe into the test piece. The test surface has an important bearing on coupling efficiency. It should be smooth and free from loose debris. Although ultrasonic tests can be carried out on painted surfaces, the instrument gain should be increased to account for transfer losses.

One of the most common problems with ultrasonic testing is insufficient access for probe scanning. The width of scanning surfaces should be considered carefully at the weld design stage.

Other NDE methods also exist. Acoustic emission (AE) and leak testing are worth mentioning here because of their use in weld testing. Acoustic emissions are stress waves generated from crack growth and plastic deformation. Therefore, the weld has to be loaded. The emissions range in frequency from 20 to 1200 kHz and are picked up with piezoelectric transducers fixed to the structure in arrays that are from 1 to 6 m (3.3 to 19.7 ft) apart. The emissions generally occur in bursts and are of very low amplitude. The instruments are able to amplify the signals and filter out the high levels of associated noise, producing a record of emissions over time. The emissions are difficult to define, because of their wide amplitude and frequency range. Moreover, the response of the transducers has an overriding effect on the signals. In most conventional systems, the analysis of emissions is simplified to include the number of counts (emissions exceeding a threshold level), their amplitude, and the rise time of pulses.

Acoustic emission testing has three important applications in the welding field: the in-process monitoring of resistance welds, the detection of leaking welds in storage tanks, and the proof testing of welded pressure vessels.

In the first of these applications, the AE is produced by the thermal stresses created as the weld cools and solidifies. Each weld process has a particular AE signature, which can be monitored to indicate suboptimal performance.

Leaks that produce turbulent flow in the escaping fluid also produce stress waves. Therefore, AE provides a noninvasive method of detecting leaks in storage tanks. By taking triangulation measurements from points around the tank, the source can be located.

Perhaps the most effective use of AE is in the proof testing of pressure vessels. The AE process was originally developed for use with pressure tests of vessels in glass fiber reinforced plastic. In effect, it sensitizes the proof test, so that subcritical defects, which do not lead to failure in the test, but which may lead to subsequent failure, are revealed.

Most AE problems are associated with noise. Site testing requires quiet environmental conditions, without rain or wind. In addition, great care has to be taken when loading structures to avoid flaw propagation to critical levels. The past loading history of the structure affects its AE signature. For this reason, effective AE tests of pressure vessels can only take place with the initial pressure test. After the first loading, some information about flaws will be lost.

As already mentioned, leak testing can be carried out using AE methods. However, there are simpler methods that can be used in the field for inspecting welds. For example, penetrant testing can be adapted by applying the liquid to one surface and observing the development of flaws on the other.

The most common method of leak testing is the bubble test, in which a bubble-forming solution is sprayed onto the surface. To create the pressure difference across the weld, which is necessary to form the bubbles, the test area is either encased in a transparent vacuum box or a positive pressure is applied to the opposite side of the component.

For very small leaks, it is necessary to use tracer gases with specialized detection equipment and probes to locate the position of the leak. Gases of small molecular size, such as helium, are used for this purpose.

Inspection of Welded Joints

G.R. Edwards, The Welding Institute

Test Procedures

A very important aspect of NDE is the development of test procedures. These are control documents, which state how the test is to be performed and how the results are to be evaluated and recorded. Written by an NDE specialist, the procedure attempts to provide the test operator with information to ensure that the defects being sought are detected in a repeatable manner.

NDE procedures can be classified as either routine or special. The routine procedure is written in accordance with a specification or code. The special procedure is written to meet a particular requirement, which is not covered by a specification. Special procedures are more prevalent during in-service monitoring and may require validation trials on specimens that possess simulated defects.

A good procedure will be clear and concise, and it will provide adequate coverage within the constraints set by cost and time for inspection. The procedure will start with a title page that includes a reference number as part of a quality-assurance system. At times, the procedures may be amended, and the system must be able to cope with this. In addition to the name of the person who wrote the procedure, someone in authority, who must also print his or her name and sign and date the document, must approve the procedure. The next section of the procedure will list the precise details of the equipment to be used, listing alternatives, if possible.

The level of competence required of the test operators must also be defined. In situations where nationally accepted schemes, such as the American Society of Nondestructive Testing (ASNT), are mentioned, finding an equivalent scheme can be a problem. Attempts are being made to formalize agreements among national certification schemes.

The section of the procedure that is known as the technique comprises equipment setup and test operation. This is at the core of the procedure document and requires the most careful preparation. It is not uncommon to list several techniques within one procedure document, where each refers to a different type of weld in a structure, but is covered by the same equipment, personnel, calibration, evaluation, and recording requirements as the other techniques.

The methods used to evaluate the test results must be described carefully. They should consider the way in which the results are presented, that is, images or signals. It may be appropriate to evaluate the porosity visible in a radiographic image in terms of density as a percentage of area. However, with ultrasonic testing, the signals on the A-scan cannot be interpreted in this way.

The final section of the document deals with the recording of results. Written evidence is always required to document how the test was carried out, so that it can be repeated. Report forms must be designed carefully to include serial numbers of equipment, for example. Radiography is the only routine method that provides a permanent record of the test results. Although photographs can be taken of flaws found by penetrant and magnetic-particle testing, the results of eddy-current and ultrasonic tests are more difficult to record. The information about signal amplitude and position can be logged in a table and plotted onto diagrams of the weld, but it is the skill of the test operator that ultimately determines the quality of the report.

Inspection of Welded Joints

G.R. Edwards, The Welding Institute

Test-Operator Training

The training and certification of NDE operators receives a great deal of attention. The approach established by the American Society of Nondestructive Testing (ASNT) has been adopted by many national schemes, but with important differences. Broadly speaking, there are three levels of competence for each method. At Level I, the test operators have progressed from trainees to a stage at which they can conduct tests under supervision. At Level II, they are able to conduct tests independently, write test procedures, and generally provide the base manpower for NDE operations. At Level III are supervisors and managers responsible for the NDE function. They have a more general knowledge of NDE and quality control and are in a position to approve NDE procedure documents, for example. At each level, the technician will have passed a written examination and demonstrated practical proficiency in conducting tests.

The situation is complicated by the need to qualify test operators not only for the method, but for the application. Personnel involved in the inspection of groove welds in plate, for example, may not have received training for the inspection of groove welds in T-joints. This led the ASNT to establish schemes that were company specific. Each company establishes its own training and certification scheme according to the document SNT-TC-1A, "Recommended Practice for Qualification and Certification in NDT." This is subject to external audit.

European certification schemes such as PCN (United Kingdom), COFREND (France), and NORDTEST (Scandinavia) also have three levels of competence, but the approvals are conducted at independent test centers.

Inspection of Welded Joints

G.R. Edwards, The Welding Institute

NDE Performance

NDE detects and evaluates flaws indirectly by the way in which the flaw interrupts a magnetic field or reflects ultrasound, for example. Measuring the performance of NDE in carrying out this task is difficult, because of ambiguities in the interpretation of test results. Nondestructive evaluation is not an exact science.

The detection capability of the test is expressed in terms of its sensitivity, but the test may also provide information about the nature, size, and location of the flaw. These are all important attributes in determining whether or not a flaw is significant.

Nature of the Flaw. NDE methods can be segregated in terms of those that provide images of the flaw (visual inspection and penetrant, magnetic particle, and radiographic testing) and those that provide signals from the flaw (ultrasonic and eddy-current testing).

The nature of the flaw can be more easily assessed from images than from signals. However, interpreting images is not always straight-forward. Penetrant tests, for example, indicate flaws that are a result of a blotting action. This can have the effect of forming a line of rounded flaws from a linear flaw. Cleaning away the developer and closely examining the purported flaw with a magnifying glass is usually sufficient to assess ambiguity.

With MPT, the sharp change in contour that occurs at the weld toe produces a flux-leakage field that attracts magnetic particles. This indicates a flaw that can be mistaken for a toe crack. Generally, however, the flaw is weak and can be blown away from the surface. Grinding the weld toe to a favorable profile can improve accuracy.

Because radiographic testing provides a two-dimensional image of a three-dimensional object, images from the surface can be confused with those from internal flaws. Visual inspection may be necessary to distinguish an undercut from a slag inclusion, for example.

The identification of flaws from signals is severely limited. With eddy-current testing, an observation of vector displacements on the cathode-ray tube may distinguish surface from subsurface cracks, for example, but only after careful selection of the test frequency.

The signals displayed in the A-scan of an ultrasonic flaw detector provide more information, including distance and amplitude information collected as the probe scans the surface. These so-called echo dynamic patterns make it easier to characterize a flaw using an ultrasonic test than it is with an eddy-current test. Flaw characterization based on ultrasonic tests is currently a matter of much interest.

Flaw Size. After a flaw has been detected and its nature determined, the next step is to measure its dimensions. Measuring the length of a flaw is straightforward. However, the blotting effects of penetrants must be taken into account. Other problems can arise in radiography, where the full extent of a fine linear flaw can be seriously underestimated if its plane moves out of parallel with the radiation beam.

In an eddy-current or ultrasonic test, the length of probe movement between the indication end points is what is being measured. The definition of the end points can be a problem. The most common method employed in ultrasonic tests is the 6 dB drop sizing method, in which the probe is moved off the edge of the flaw to the point where the signal has dropped in amplitude by 6 dB or half its original amplitude. The axis of the sound beam is now half on and half off the flaw and can be taken as delineating the flaw edge. This can be done to an accuracy of ±5 mm (0.2 in.).

The width of a flaw also can be measured in an image, but it is generally too small to be measured by eddy-current or ultrasonic test methods.

After length, the next most important dimension in assessing the significance of a flaw is its height in the through-wall direction. Here, the performance of imaging methods is severely limited. Only qualitative assessments can be made from the density of a defect indicated by penetrant, magnetic particle, or radiographic testing. On the other hand, eddy-current and ultrasonic tests do provide quantifiable information.

Comparisons with slots of known depth in a reference block serve to calibrate the vector displacements on an eddy-current instrument. Cracks to a depth of 3 mm (0.12 in.) can be measured with an accuracy of 0.5 mm (0.02 in.) in this way.

Ultrasonic tests can provide a measure of the through height of an internal flaw. There are four methods, the simplest of which involves comparing the amplitude of the echo signal from the flaw with the echo signal from a side-drilled or flat-bottomed hole. This can only be applied if the dimension is smaller than the width of the ultrasound beam (typically, 5 to 10 mm, or 0.2 to 0.4 in.). More importantly, when dealing with planar flaws, the orientation must be taken into account. If the plane of the flaw is more than 20° off perpendicular to the axis of the sound beam, then the echo will not be reflected back to the probe and it even may go undetected.

For small planar flaws with a height of more than 3 mm (0.12 in.), the preferred ultrasonic sizing method is the 20 dB drop method. This involves scanning the beam across the flaw and finding the 20 dB drop points. The beam path and probe positions at these points are plotted on a diagram of the weld using a cursor or overlay with the 20 dB beam profile to give the flaw edges. An accuracy of ±3 mm (0.12 in.) is achievable.

For planar flaws with large through-wall dimensions, such as cracks, the maximum-amplitude method involves plotting the facets of a flaw to indicate its orientation and dimension.

The location of the flaw in the through dimension of the weld is important in determining its significance. Flaws that are near a surface are more significant than flaws in the center of a weld. In a radiograph, this can only be inferred. Root flaws may appear in the center of the weld image, but may be confused with flaws that are near the weld cap. Only ultrasonic testing can provide definitive information about the through-wall location of a flaw. Indeed, it is often necessary to supplement radiographic tests with ultrasonic tests for this purpose.

Flaw location is, perhaps, the most reliable function of ultrasonic tests. To locate an echo signal, the time base of the A-scan must be accurately calibrated for distance. Variations in the velocity of ultrasound are not uncommon, but generally an accuracy to within 1 mm (0.04 in.) can be expected. The angle of the ultrasound beam and the probe index point must also be checked regularly with a calibration block. Shoe wear can lead to inaccuracies.

Finally, the position of the probe on the test surface at the point of maximum signal amplitude must be measured from the weld centerline, which should always be marked clearly on the test surface. However, the reliability of flaw location with ultrasonic testing depends on the skill of the operator. Carelessness can lead to errors, for example, in plotting the flaw on the wrong side of the weld.

Sensitivity. When dealing with images, sensitivity is expressed in terms of contrast and definition. When dealing with signals, it is the separation of signals from the background noise or signal/noise ratio.

The contrast and definition of penetrant testing and MPT can be enhanced by using fluorescent materials. Variations in contrast can be recognized more easily than with colored materials, and finer details in the flaw can be observed. However, on rough surfaces, background color can lead to a loss in sensitivity.

The contrast in a radiographic image is improved by decreasing the energy of the radiation and increasing the density of the image. In gamma radiography, the energy is fixed by the type of isotope, whereas in x-radiography, it can be varied with the tube voltage. The radiation from Iridium 192 has an x-ray equivalent in the 500 kV to 1 MeV range and, therefore, image quality is not as good as it is with x-rays in the usual 150 to 400 kV range. Ytterbium 169 has a much lower energy level, producing nearly x-ray image quality, but it suffers from a short half life and low output.

Although radiographic definition is also reduced by using higher-energy radiation, the most important factor that affects definition is the unsharpness, or penumbra, effect. This is caused by having a source, or x-ray focal spot, of finite size, typically 3 to 4 mm (0.12 to 0.16 in.). It is reduced by moving the object away to a distance of 0.5 to 1 m (1.6 to 3.3 ft) from the source, depending on object thickness, or by using special minifocused or microfocused x-ray tubes.

The contrast and definition of a radiograph can be assessed using a penetrameter or an image quality indicator (IQI). The size and type of penetrameter are specified by standards such as ASTM 747, "Method for Controlling Quality of

Radiographic Examination Using Wire Parameter." This instrument is placed near the weld or, in the case of a wire-type IQI, across the weld, on its source side at the time the radiograph is taken. The IQI comprises carefully dimensioned wires or steps with holes. By counting the number of wires or stepped holes that are visible on the radiograph, the sensitivity can be calculated as a percentage of wall thickness, or ***T*** value. If calculated as a percentage, then a value of 1 to 2% is normally acceptable. A higher percentage indicates a lower sensitivity.

Despite a high level of sensitivity, fine planar flaws can be missed if they are not oriented almost parallel with the radiation beam. Radiography is not effective at detecting incomplete side-wall fusion in metal-inert gas welds with a 30° chamfer angle, for example, unless the radiographic source is aligned with the angle, which requires multiple exposures.

In eddy-current testing, improving the signal/noise ratio is based on selecting appropriate coil configurations, test frequencies, and *y*/*x* potential ratios on the vector point display. A phase separation approaching 90° between the impedance vector caused by noise as the probe moves along the weld surface (lift-off) and the impedance vector from cracks can be achieved. Further reductions in noise can be made using signal filters, but the probe then has to be scanned at a constant speed.

The test frequency also affects the signal-to-noise separation in ultrasonic testing. As the frequency increases, the ultrasound attenuation rises, and the reduced ultrasound wavelength is scattered to produce more noise. Generally, the optimal frequencies for ultrasonic testing of welds in ferritic steels range from 4 to 5 MHz. The frequencies are higher for welds that are less than 10 mm (0.4 in.) thick, and lower for welds that are more than 50 mm (2 in.) thick. It should be noted, however, that some test procedures call for lower frequencies, because their flaw evaluation is based on echo amplitude only. At lower frequencies, beam spread is greater and, therefore, signal amplitude is less dependent on flaw orientation.

Inspection of Welded Joints

G.R. Edwards, The Welding Institute

Evaluation of Test Results

The evaluation of test results and the setting of accept/reject criteria for weld flaws are the subjects of much current debate. This is because of the empirical way in which the criteria were derived in the past. As new developments in defect assessment, such as fracture mechanics, set the trend toward more sophisticated methods of evaluating NDE results, the discrepancies in current practice become more apparent. This is particularly true of ultrasonic testing.

In general, the accept/reject criteria used in national standards draw a distinction between rounded and elongated flaws, as well as volumetric and planar ones. With NDE methods that produce an image of the flaw, this is a simple task. An elongated flaw is usually defined as one in which the length is more than three times the width.

Flaws rarely occur in isolation, and their distribution and proximity to each other are important characteristics. Porosity may be isolated, clustered, or aligned. If it is aligned, then it is more significant than if it were random, because there is likely to be incomplete fusion.

The NDE methods that produce signals from the flaw are the most difficult to evaluate. In ultrasonic tests, the simplest method of evaluation is to use echo amplitude and length criteria only. This can lead to planar flaws being undersized, if they are poorly oriented in the ultrasound beam. It is therefore important, in an ultrasonic test, to first characterize the flaw and, in particular, to distinguish planar flaws from volumetric ones. The identification of planar flaws such as cracks and incomplete fusion primarily depends on the skill of the test operator, who considers the signal location, shape, response to probe movement, and relative amplitude from different sides of the weld and with different probe beam angles.

Volumetric flaws, on the other hand, can be evaluated in terms of signal amplitude and length. In some specifications, for example, if the signal amplitude exceeds the DAC set-off 3 mm (0.12 in.) diameter holes, and the dB-drop length of the flaw exceeds the wall thickness of the weld, then the flaw is rejected as a defect.

The accept/reject criteria are usually set by a fabrication or manufacturing standard. In the past, these standards have been application specific. For example, ASME Section VIII has governed the manufacture of unfired pressure vessels. In the new European standards, however, the move is toward blanket quality levels, where each quality level defines the NDE procedures and defect accept/reject criteria and is selected according to the application.

Inspection of Welded Joints

G.R. Edwards, The Welding Institute

Selection of the NDE Method

The selection of the appropriate NDE method is influenced by a number of factors related to the test conditions. Those described here are not exclusive. There may be other constraints on selection, such as availability of trained personnel or costs. The methods shown in Tables 1, 2, and 3 are rated according to their applicability. This rating is not easily quantifiable, because it must account for the ease of test operation, its efficacy in finding flaws common to the application, and whether or not it is the only applicable NDE method.

TABLE 1 SELECTION OF NDE TECHNIQUE ON BASIS OF MATERIAL

A, most applicable; B, applicable; C, least applicable

MATERIAL	VISUAL	PENETRANTS	MPT	EDDY CURRENT	RADIOGRAPHIC TESTING	ULTRASONIC TESTING
BULK PLASTIC	A	B	. . .	. . .	A	**C**
GLASS FIBER REINFORCED PLASTIC	A	B	. . .	. . .	A	**C**
CARBON FIBER REINFORCED PLASTIC	B	C	. . .	C	A	**A**
LOW-CARBON STEEL	B	B	A	C	B	**A**
HIGH-ALLOY STEEL	B	B	C	B	B	**B**
ALUMINUM ALLOY	B	B	. . .	B	A	**B**
TITANIUM ALLOYS	B	B	. . .	B	A	**B**
DUPLEX STAINLESS STEEL	B	B	C	C	A	**B**
COPPER ALLOYS	**B**	**A**	**. . .**	**C**	**B**	**C**

TABLE 2 SELECTION OF NDE TECHNIQUE ON BASIS OF WELD GEOMETRY

A, most applicable; B, applicable; C, least applicable

GEOMETRY	VISUAL	PENETRANTS	MPT	EDDY CURRENT	RADIOGRAPHIC TESTING	ULTRASONIC TESTING
GROOVE PLATE/PIPE	B	B	B	C	A	**A**
T GROOVE	B	B	B	C	C	**A**
BRANCH GROOVE	B	B	C	C	C	**A**
NODE	B	B	B	B	C	**B**
T FILLET	A	B	A	C	C	**C**
BRANCH FILLET	A	B	A	C	C	**C**
LAP	A	B	A	C	B	**C**
<6 MM (<0.2 IN.)	A	B	A	B	B	**C**
6-15 MM (0.2-0.6 IN.)	B	B	B	C	A	**B**
16-50 MM (0.6-2 IN.)	C	C	C	C	B	**A**
>50 MM (>2 IN.)	C	C	C	C	B	**A**

ONE-SIDE ACCESS	C	C	C	C	. . .	**A**
BOTH-SIDE ACCESS	B	B	B	C	A	**A**
SINGLE V	C	C	C	C	B	**A**
DOUBLE V	**B**	**B**	**B**	**C**	**A**	**A**

TABLE 3 SELECTION OF NDE TECHNIQUE ON BASIS OF FLAW TYPE

A, most applicable; B, applicable; C, least applicable

FLAW	**VISUAL**	**PENETRANTS**	**MPT**	**EDDY CURRENT**	**RADIOGRAPHIC TESTING**	ULTRASONIC TESTING
MICROCRACKS	A	C	C	C	C	**C**
LONGITUDINAL CRACKS	C	B	B	B	A	**A**
TRANSVERSE CRACKS	C	B	B	B	A	**B**
RADIATING CRACKS	C	A	B	C	A	**B**
CRATER CRACKS	B	B	B	C	A	**C**
GROUP DISCONTINUOUS CRACKS	B	B	B	C	A	**C**
BRANCHING CRACKS	B	B	B	B	A	**C**
UNIFORMLY DISTRIBUTED POROSITY	C	B	C	C	A	**C**
LINEAR POROSITY	C	B	C	C	A	**C**
ELONGATED CAVITY	C	C	C	C	A	**B**
WORM HOLE	C	C	C	C	A	**C**
SURFACE PORE	B	A	B	C	B	**C**
SHRINKAGE CAVITY	A	B	B	C	B	**C**
INTERDENDRITIC SHRINKAGE	A	C	C	C	C	**C**
MICROSHRINKAGE	A	C	C	C	C	**C**
INTERDENDRITIC MICROSHRINKAGE	A	C	C	C	C	**C**
CRATER PIPE	B	A	B	C	B	**C**
SOLID INCLUSION	. . .	. . .	. . .	C	A	**B**
SLAG INCLUSION	. . .	. . .	. . .	C	A	**B**
FLUX INCLUSION	. . .	. . .	. . .	C	A	**B**
OXIDE INCLUSION	C	. . .	. . .	C	A	**B**
PUCKERING	. . .	. . .	. . .	. . .	A	**C**
METALLIC INCLUSION	. . .	. . .	. . .	. . .	A	**B**
INCOMPLETE SIDEWALL FUSION	C	B	B	C	B	**A**
INCOMPLETE INTERRUN FUSION	C	B	B	C	B	**A**
INCOMPLETE ROOT FUSION	C	B	B	C	A	**B**
INCOMPLETE PENETRATION	B	B	B	C	A	**B**
UNDERCUT	A	B	B	C	B	**C**
SHRINKAGE GROOVE	A	. . .	. . .	. . .	B	**C**
EXCESSIVE REINFORCEMENT	A	. . .	. . .	. . .	B	**C**

EXCESSIVE CONVEXITY	A	. . .	. . .	. . .	B	**C**
EXCESSIVE PENETRATION	A	. . .	. . .	. . .	B	**C**
BAD REINFORCEMENT ANGLE	A	. . .	. . .	. . .	. . .	**. . .**
OVERLAP	B	A	B	C	C	**C**
MISALIGNMENT	A	. . .	. . .	. . .	. . .	**C**
BURN THROUGH	A	B	C	C	B	**B**
INCOMPLETELY FILLED GROOVE	A	. . .	. . .	. . .	B	**C**
ASYMMETRICAL FILLET	A	. . .	. . .	. . .	. . .	**. . .**
IRREGULAR BEAD	A	. . .	. . .	. . .	B	**. . .**
ROOT CONCAVITY	A	. . .	. . .	. . .	B	**C**
POOR RESTART	A	. . .	. . .	. . .	B	**. . .**
MISCELLANEOUS DEFECTS (SPATTER, ETC.)	**A**	**C**	**C**	**. . .**	**B**	**. . .**

Source: ISO 6520

There are various techniques for applying each NDE method, some of which are more appropriate than others and might therefore alter the rating. It is often the case that two methods are complementary to each other. For example, radiographic testing will detect volumetric flaws and ultrasonic testing will detect planar ones. The method of selection discussed below is based on the material, weld geometry, and type of flaw.

Material. For NDE purposes, the material being tested can be categorized as shown in Table 1. First, a distinction is drawn between metal and nonmetal.

NDE methods are being increasingly applied to welds in thermoplastics, which can take the form of bulk plastics or composites with reinforcing fibers. Electrical and magnetic testing methods cannot be applied to bulk plastics, of course. Eddy-current techniques have been used on carbon fiber reinforced plastics, because of the conductive properties of the fibers. Penetrants also can be used, but care must be taken to ensure that solvents do not attack the plastic.

Radiographic techniques can be applied using low-kilovolt techniques, which are necessary because of the low absorption coefficient for x-rays of plastics. Typical values are between 10 and 25 kV. This can introduce problems such as cassette pick-up and air absorption.

Plastics are highly attenuative of ultrasound, and they will only propagate compressional waves to any distance. However, ultrasonic techniques have been used with some success for the inspection of hot-plate groove fusion welds in polyethylene pipe.

Metals can be subcategorized as being either magnetic or nonmagnetic, dense or light, and isotropic or anisotropic. Obviously, MPT cannot be used on nonferrous metals, but there are also many high-alloy steels that are too weakly magnetic to be magnetized. Other ferrous steels that are often used in the aerospace industry are ferromagnetic, but require stronger magnetizing fields to create adequate flux leakage from flaws. These also tend to have high magnetic retentivity and are difficult to demagnetize after the inspection.

The density and anisotropy in metals have the greatest effect on test sensitivity in radiographic and ultrasonic tests.

In general, radiation is absorbed more by dense materials, such as austenitic steel, than by less-dense materials, such as aluminum. Because higher-energy radiation is needed to penetrate the same thickness of material, it has the effect of decreasing image contrast and, therefore, test sensitivity. Anisotropy that is due to a large grain structure also results in radiation scattering and, sometimes, the appearance of diffraction mottling in the radiographs.

The velocity and attenuation of ultrasound is affected by changes in material composition or structure. This can be particularly apparent in aluminum welds and duplex stainless steel. In addition, large grain size can distort ultrasonic signals.

Eddy-current tests are particularly susceptible to interference from quite minor changes in the material. For example, the change in magnetic permeability at the heat-affected zone of a weld in low-carbon steel can create noise that obscures signals from toe cracks.

Weld geometry also has an important effect on test method selection. The factors that need to be considered include joint type and thickness, whether or not there is access to both sides, and, in the case of groove welds, the inclination of the weld groove faces (Table 2).

Visual inspection is applicable to all weld geometries, but is particularly applicable to fillet welds and welds in thin sections. Penetrant testing and MPT serve to enhance visual inspection.

Eddy-current testing is the least applicable of the test methods for weld inspection, but it is being used increasingly in underwater node inspection as an alternative to MPT. For thin-section welds in nonmagnetic materials, it has sufficient penetrating power to offer an alternative to ultrasonic testing.

Radiographic testing is of limited use in testing welds that have a complex geometry, as well as thick-section welds. Access to both sides of the weld is essential.

Ultrasonic testing does not suffer from as many geometric restrictions as radiographic testing, but the method is difficult to use on small fillet welds and thin sections, where A-scan interpretation is complicated by the numerous geometric echo signals.

Type of Flaw. Visual inspection is not applicable to internal flaws, but it is the only method that can be used with most geometric flaws (Table 3). Microflaws will only be visible with the aid of magnification.

Penetrant testing and MPT enhance visual inspection, but are particularly sensitive to surface-breaking cracks and incomplete fusion. Undercut and other shallow and open flaws usually give weak images, whereas overlap can be difficult to distinguish from the sharp change in contour at the weld toe. Eddy-current tests are very sensitive to planar flaws at the surface, but not to volumetric flaws.

There are very few flaws that cannot be detected by radiographic testing. Planar flaws must be oriented favorably in the radiation beam, and it can be difficult to distinguish surface from subsurface flaws, because of the two-dimensional nature of the image. The angular "cusp" shapes of an undercut can, for example, be indistinguishable from inclusions. In addition to magnification by way of microfocus radiographic techniques, microflaws need very high levels of contrast and definition in the radiograph, in order to be detected.

Ultrasonic testing is often regarded as being complementary to radiographic testing. Planar flaws need to be oriented in a perpendicular plane to the ultrasound beam, but because a range of beam angles can be utilized, ultrasonic testing is generally regarded as more sensitive to cracks and incomplete fusion than radiography. However, the characterization of different types of flaw, both planar and volumetric, is extremely difficult. Porosity produces low-amplitude echo signals that barely exceed the recording threshold levels in most procedures.

Advanced Methods

As with all technologies, advances in NDE are being dominated by the introduction of computers. The emphasis has been on the acquisition and recording of data to aid the test operator and interpreter. In the future, it is likely that computers may complete the analysis and evaluation of test results, although the problems in interpreting the results of even the simplest NDE procedure should not be underestimated.

In radiographic testing, the replacement of films by real-time imaging systems has been aided by the introduction of image processing software that is able to improve the image quality to almost film-like levels.

In eddy-current testing, the characterization of different types of flaws by their effect on the impedance at mixed coil frequencies is a complex problem that is being tackled with neural network software.

The greatest advances have been made in ultrasonic testing. Digitally controlled flaw detectors make calibration more consistent and can provide print-outs of the A-scan display. With probe scanners, ultrasonic images can be created to provide a hard copy of test results. Digitized signals can be processed at much higher gain settings to reveal diffracted echoes in accurate flaw sizing.

Weld Procedure Qualification

Harvey R. Castner, Edison Welding Institute

Introduction

WELDING CODES AND STANDARDS usually require the qualification of welding procedures prior to being used in production to ensure that welds will meet the minimum quality and mechanical property requirements for the application. This article provides an overview of welding procedure qualification guidelines and test methods. The reader is referred to codes, standards, and specifications that govern the design and fabrication of welded structures for the procedure qualification details that are appropriate for the given application.

Some of the most widely recognized codes, standards, and specifications that cover welding procedure qualification include:

- SECTION IX OF THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS (ASME) BOILER AND PRESSURE VESSEL CODE
- AMERICAN NATIONAL STANDARDS INSTITUTE/AMERICAN WELDING SOCIETY (ANSI/AWS) B2.1, STANDARD FOR WELDING PROCEDURE AND PERFORMANCE QUALIFICATION
- ANSI/AWS D1.1, STRUCTURAL WELDING CODE FOR STEEL
- ANSI/AWS D1.2, STRUCTURAL WELDING CODE FOR ALUMINUM
- ANSI/AWS D1.3, STRUCTURAL WELDING CODE FOR SHEET STEEL
- ANSI/AWS D1.5, BRIDGE WELDING CODE
- ANSI/AWS D1.4, STRUCTURAL WELDING CODE FOR REINFORCING STEEL
- MILITARY STANDARD 248, WELDING AND BRAZING PROCEDURE AND PERFORMANCE QUALIFICATION
- AMERICAN PETROLEUM INSTITUTE (API) 1104, STANDARD FOR WELDING PIPELINES AND RELATED FACILITIES
- INDUSTRIAL SPECIFICATIONS, SUCH AS ANSI/AWS D14.1, SPECIFICATION FOR WELDING INDUSTRIAL AND MILL CRANES AND OTHER MATERIAL HANDLING EQUIPMENT; ANSI/AWS D14.3, SPECIFICATION FOR WELDING EARTHMOVING AND CONSTRUCTION EQUIPMENT; ANSI/AWS D14.5, SPECIFICATION FOR WELDING OF PRESSES AND PRESS COMPONENTS; ANSI/AWS D14.6, SPECIFICATION FOR WELDING ROTATING ELEMENTS OF EQUIPMENT; AND ANSI/AWS D15.1, RAILROAD WELDING SPECIFICATION

Additional information about standards and specifications for particular welding applications is available in the Section "Material Requirements for Service Conditions" in this Volume.

Procedure qualification is an important part of any total weld quality management program, because it confirms that the welding procedure will meet the mechanical design requirements. Therefore, some form of welding procedure qualification should be performed for all critical welding operations, regardless of whether the application is covered by a code or standard.

Weld Procedure Qualification

Harvey R. Castner, Edison Welding Institute

Purpose of Qualification and Responsibility for the Task

Fabricators, manufacturers, and contractors are responsible for the selection, definition, and control of the welding procedures they use. This control normally includes the preparation of detailed written instructions known as welding procedure specifications (WPS), and may include the qualification of both welding procedures and personnel. The purpose of qualifying a welding procedure is to demonstrate that the resulting welds will meet prescribed quality standards and have the requisite mechanical properties for the intended application. A sample WPS form is shown in the article "Welding of Carbon Steels" in this Volume.

Some codes allow the use of prequalified welding procedures, if they meet all of the requirements contained in the code. The use of prequalified procedures does not lessen the responsibility of the fabricator or contractor to ensure that the welding procedures are satisfactory for the application.

The qualification of welders, welding operators, and tack welders is another important part of weld quality control. The qualification of personnel demonstrates the ability to produce satisfactory welds when the prescribed welding procedure is used.

Procedure qualification involves examining the welds according to the specified welding procedure, testing specimens taken from these welds, and documenting the results on a procedure qualification record. The fabricator, manufacturer, or contractor must supervise and control the preparation of the qualification test welds. Many codes permit another organization to assist in the removal and assessment of test specimens. However, the contractor or fabricator is ultimately responsible for the procedure qualification tests, the results of these tests, and the control and application of the procedures in production. A single welding procedure qualification test may support a number of WPSs, or a single WPS may be supported by several procedure qualifications.

Weld Procedure Qualification

Harvey R. Castner, Edison Welding Institute

Types of Tests

Welding procedure qualification tests can be categorized as either standard or special tests. Standard tests include visual, metallurgical, and nondestructive examinations; chemical analysis; crack susceptibility tests; and tests to determine mechanical properties. The mechanical property tests can include tension, bend, and hardness tests, as well as tests to determine notch toughness.

Although the use of standard tests to qualify a WPS demonstrates the ability of the procedure to produce sound welds that meet the minimum required properties, these tests may not ensure that the welding procedure is suitable for all applications. Standard test plates are relatively small and are normally welded under laboratory, rather than production, conditions.

The fabricator, designer/owner, or operator of the welded structure or component may determine that additional special qualification tests are required in order to establish the acceptability of a welding procedure for the application and to ensure that the welded structure will meet the mechanical requirements. Special tests may involve welding mock-up or prototype structures and then subjecting these structures to static or dynamic tests, corrosion tests, high-temperature tests, macrosectioning, or other tests.

Standard Tests

Standard procedure qualification test weldments represent the type of weld and weld joint that are to be used in production. Groove welds in butt joints in plate and pipe are shown in Fig. 1 and 2, respectively. Two examples of fillet welds in T-joints are shown in Fig. 3. Fillet welds can also be deposited on lap, corner, or edge joints. A partial penetration weld in a T-joint is shown in Fig. 4, whereas a plug weld test specimen is shown in Fig. 5. The latter could be a slot weld as well, and a similar weldment could be used for spot welds.

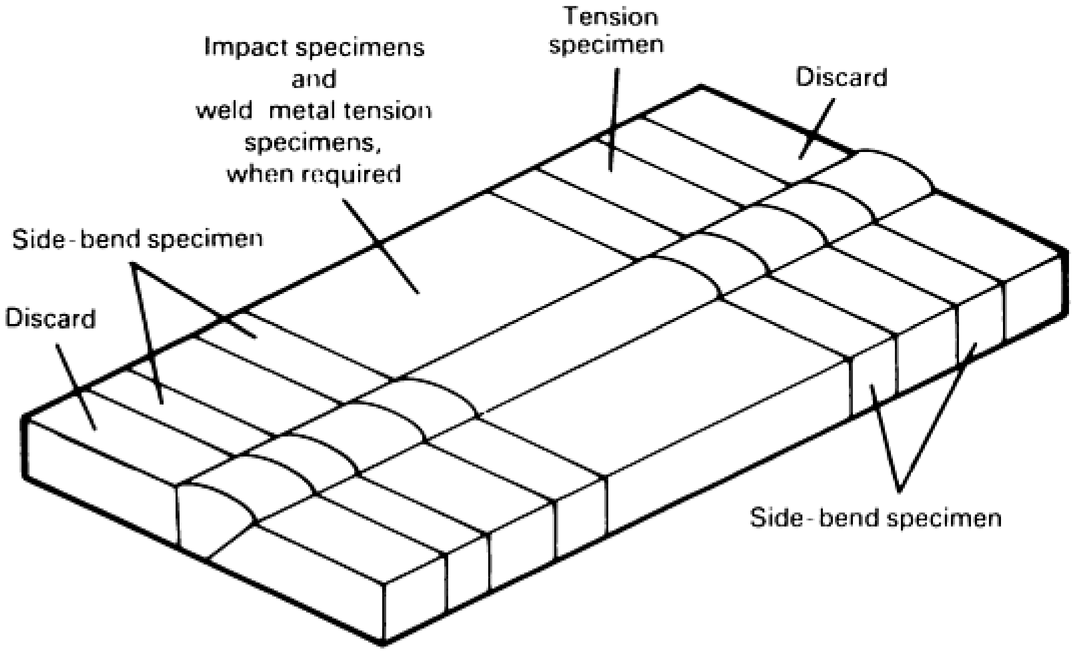
Tension specimen
Impact specimens and weld metal tension specimens, when required
Discard
Side-bend specimen
Discard
Side-bend specimen

FIG. 1 TYPICAL LOCATION OF TEST SPECIMENS FROM GROOVE WELDS IN BUTT-JOINT TEST PLATE

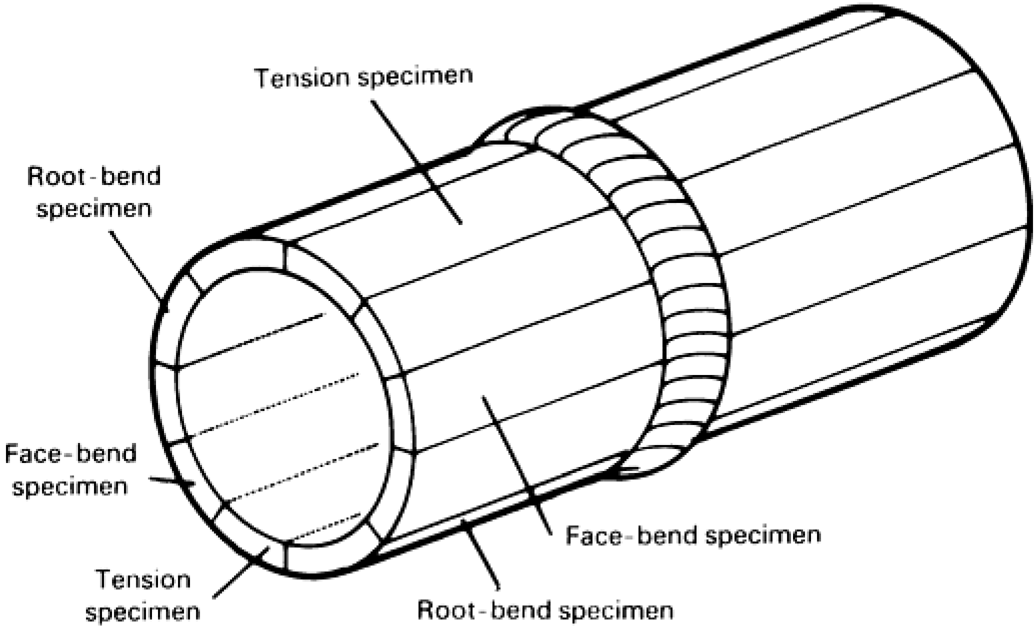
Tension specimen
Root-bend specimen
Face-bend specimen
Tension specimen
Face-bend specimen
Root-bend specimen

FIG. 2 TYPICAL LOCATION OF TEST SPECIMENS FROM GROOVE WELDS IN BUTT-JOINT PIPE TEST

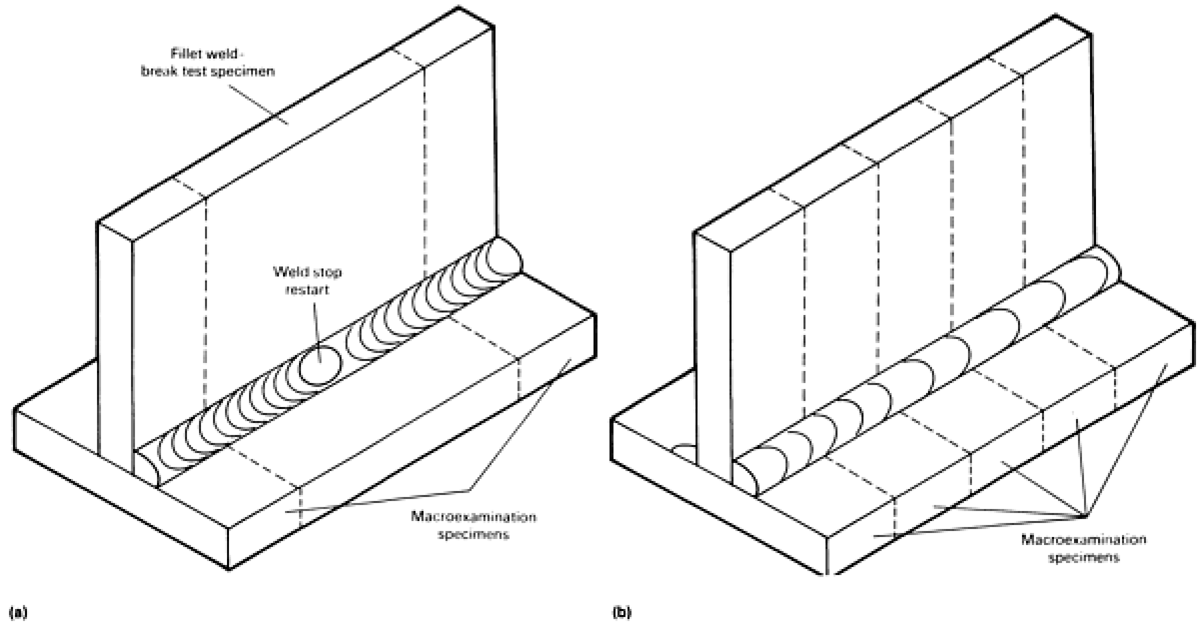
Fillet weld - break test specimen
Weld stop restart
Macroexamination specimens
(a)
Macroexamination specimens
(b)

FIG. 3 TWO EXAMPLES OF T-JOINT WELDMENTS FOR FILLET WELD QUALIFICATION. (A) FILLET WELD ON ONE SIDE OF T-JOINT. (B) FILLET WELD ON BOTH SIDES OF T-JOINT

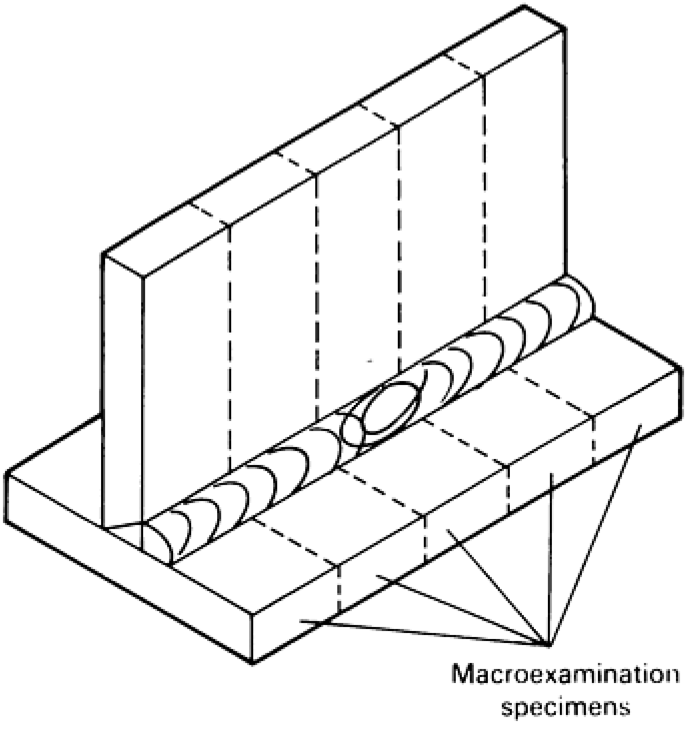
Macroexamination specimens

FIG. 4 PARTIAL-PENETRATION WELD SPECIMEN

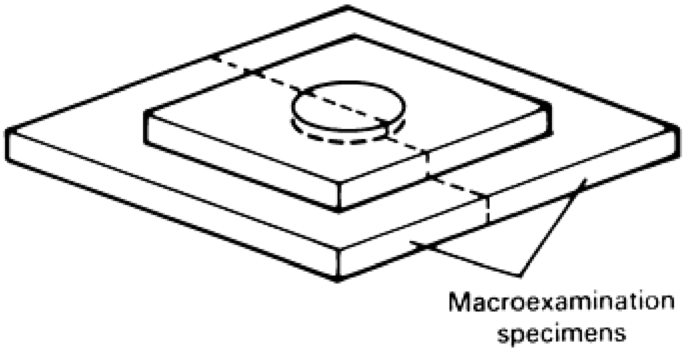

FIG. 5 PLUG WELDS QUALIFIED BY GROOVE WELD TEST

Resistance spot and seam welds are typically made on lap joints. Other weldments are appropriate for friction, laser-beam, electron-beam, and stud welding, or for surfacing weld deposits. The thickness of test weldments should be selected after considering the thickness of materials used in actual fabrications and the thickness limitations of the code or standard. The length and width of the test weldment should be sufficient to permit removal of the necessary test specimens.

The test weldment is prepared and welded using the methods that are detailed in the WPS. Standard test requirements vary, depending on the code or standard being used. However, it is common to examine the completed weldment by visual and nondestructive examination methods. It is also common to section the weldment and remove specimens for mechanical property tests or other examinations.

Tests for full-penetration groove welds include tension, bend, and toughness tests. It is also common to subject weldments to chemical analysis, hardness tests, and metallurgical examinations. Figures 1 and 2 indicate the location of tension and bend specimens in plate and pipe test weldments, respectively. Figure 6 shows the configuration of tension and bend specimens. The details of specimen size, orientation, removal, testing, and acceptance are provided in the welding codes and standards listed above, in ASTM test standards, or in ANSI/AWS B4.0.

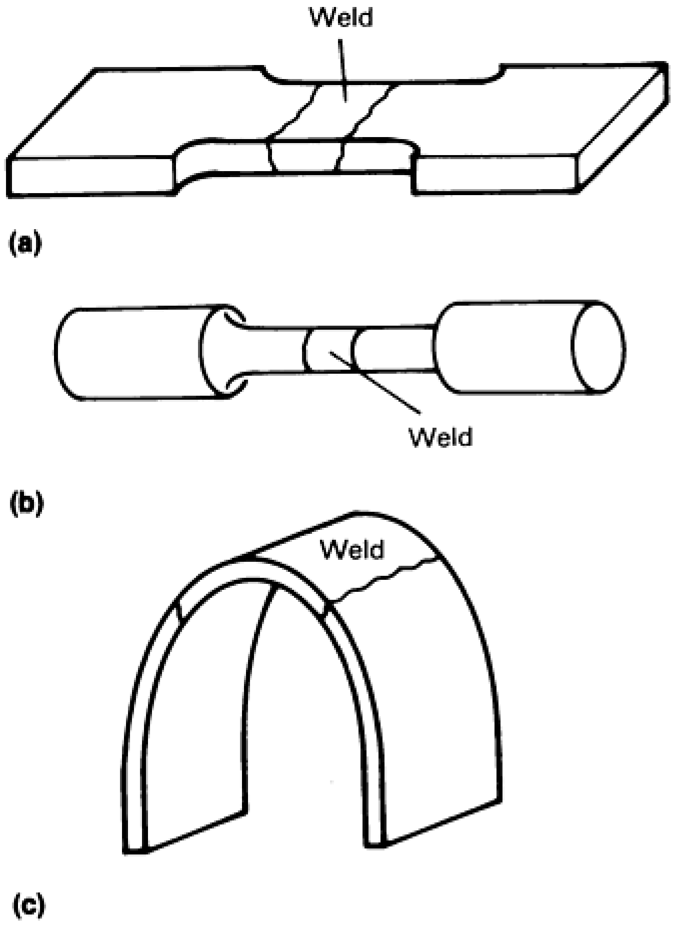

FIG. 6 TENSION AND BEND SPECIMENS FOR TESTING OF FULL-PENETRATION GROOVE WELDS. (A) REDUCED-SECTION TENSION SPECIMEN. (B) ROUND TENSION SPECIMEN. (C) BEND SPECIMEN

Transverse tension tests are performed on full-penetration groove welds to determine the tensile strength of the weld region. The specimens can be one of several configurations: full-section specimens, reduced-section specimens, or round specimens (see Fig. 6). Multiple tension specimens can be used for plates that are more than 25 mm (1 in.), as shown in Fig. 7, in order to test the entire thickness. The acceptance of the tension test results normally is based on a tensile strength that is either equal to or greater than the minimum specified tensile strength of the base metal being welded. If the weld metal has a lower specified strength than the base metal, then the tension test must exceed the specified minimum tensile strength of the weld metal. Some codes permit the strength of the tension specimen to be 5% below the specified minimum tensile strength of the base metal if the specimen breaks outside of the weld or fusion line. If all-weld-metal tension tests are required, then round specimens are typically removed from the center of the weld and oriented parallel to the weld axis.

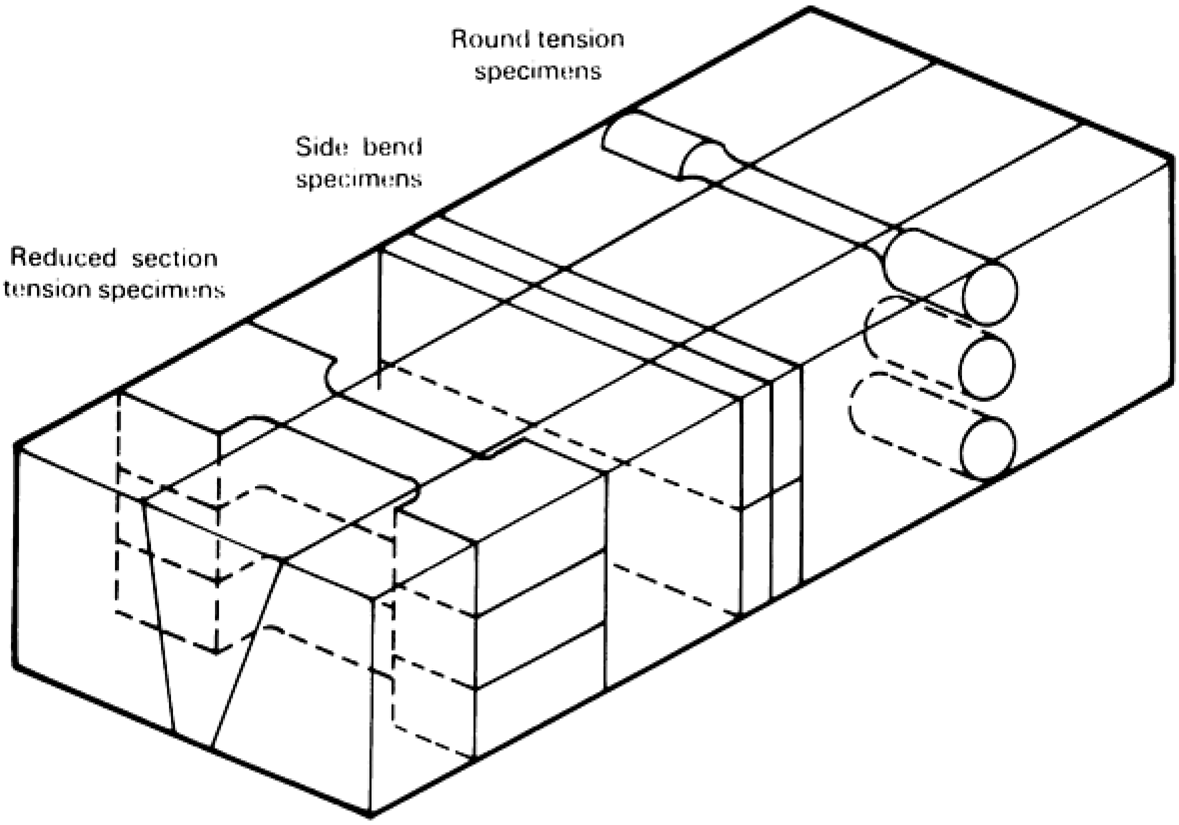

FIG. 7 REMOVAL OF MULTIPLE BEND AND TENSION SPECIMENS FROM THICK TEST PLATES

Bend tests are performed to verify the soundness and ductility of the weld and the heat-affected zone (HAZ), as shown in Fig. 6. Figures 1, 2 and 7 show bend specimens that are typically oriented transverse to the weld axis, with the center of the weld in the center of the specimen. Under some circumstances, longitudinal bends are removed with their length parallel to the axis of the weld. Longitudinal bend specimens are used for welds of dissimilar metal, where there is a large difference in bend properties between the two base metals or between the weld metal and the base metal.

The face of the weld (face-bend), the root of the weld (root-bend), or the cross section of the weld (side-bend) can be placed on the convex face of the bent specimen. Face-bend and root-bend specimens are typically specified for base-metal thicknesses of less than 9.5 mm (0.375 in.). Side-bend specimens are required for thicker test plates in order to test the entire weld cross section. Plates that are more than 38 mm (1.5 in.) thick may require multiple side-bend specimens to ensure that the entire weld thickness is tested (see Fig. 7).

Guided bend specimens are forced into a specified curvature (Fig. 6) by either applying a load to a plunger or forming the specimen around a fixed mandrel. The width and thickness of the specimen and the radius of the plunger or mandrel are designed to produce a minimum specified elongation (approximately 20%) in the outer surface of the specimen. It is important to follow the recommendations for specimen dimensions, because width and thickness influence test performance. The most common bend specimen is 9.5 mm (0.375 in.) thick and is formed around a 19.1 mm (0.75 in.) radius. However, codes specify larger radii for low-ductility base metals and for weld metals that are not capable of being bent to a small radius. Acceptable bend specimens must not fracture or have cracks or other open defects that exceed a specified maximum size (usually, 3.2 mm, or 0.125 in.) on their convex surface.

Notch Toughness Tests. Charpy V-notch impact tests are used to measure the toughness of weld metals and HAZs. Charpy V-notch specimens are usually removed from the weld metal and HAZ, with the notch oriented normal to the surface. One face of the V-notch specimen is located within 1.6 mm ($\frac{1}{16}$ in.) of the surface of the material. Thick materials require additional specimens at subsurface locations. The number of specimens, their location and orientation, and the test temperature are determined by the design requirements or material specifications. Acceptance criteria may be a minimum specified absorbed energy in joules (ft · lbf), a minimum percentage of shear fracture appearance, or a minimum specified lateral expansion opposite the notch.

Drop-weight notch toughness tests are sometimes used to determine the minimum nil-ductility transition temperature for ferritic steel weld metal and HAZs. The assessment of toughness, in terms of the stress-intensity factor, K, the J integral, or crack tip opening displacement, may be required for some welding procedures. Codes and specifications describe how specimens can be removed from the weld metal and HAZs to permit the measurement of each of these fracture-toughness properties.

Tests for Fillet Welds, Partial-Penetration Groove Welds, and Spot and Plug Welds. The procedure qualification of fillet welds typically involves T-joint test weldments, like those shown in Fig. 3. Some codes and specifications permit lap joints, comer joints, or other joint configurations that represent the actual joint used in production.

Procedures for partial-penetration groove welds can be qualified by making test weldments that represent the joint used for the application. Figure 4 shows a T-joint test weldment for a partial-penetration groove weld. Figure 5 shows the configuration of the test weldment for the qualification of plug or slot welds.

Completed fillet, partial-penetration, and plug or slot weldments are usually examined visually and by nondestructive methods to ensure that they meet the acceptance requirements of the weld quality specification. These welds do not normally lend themselves to mechanical property testing, although tension-shear tests and weld-break tests can be conducted on them. The required mechanical properties must be determined from a standard groove weld test. Macrosections are removed to verify root penetration.

Figures 3 and 4 show how macroexamination specimens are removed from fillet or partial-penetration welds. Figure 5 illustrates how plug or slot welds are cut through their center to produce macroetch specimens. The faces of macroetch specimens are ground smooth, etched, and then examined. Acceptable welds must have:

- THE DESIGNATED WELD SIZE AND PROFILE
- FUSION TO THE ROOT OF THE JOINT
- NO CRACKS
- COMPLETE FUSION BETWEEN ADJACENT LAYERS OF WELD METAL, AS WELL AS BETWEEN WELD METAL AND BASE METAL

Shear tests can also be used to measure the strength of fillet, partial-penetration, plug, or slot welds. Usually, an acceptable shear strength is at least 60% of the minimum specified tensile strength of the base metal.

Weld-break or fracture tests may also be required for fillet and partial-penetration welds. These specimens are removed from the weld, as shown in Fig. 3, and loaded so that the root of the weld is in tension until the specimen fractures or bends flat upon itself. Acceptance of a fractured specimen is determined by examining the fractured surface for complete fusion, the required depth of penetration, and porosity and inclusion levels that do not exceed a specified maximum size. The specimen is also acceptable if it bends flat upon itself without fracturing.

Procedure qualification tests for arc and resistance spot welds include tension-shear, torsion, peel, and hardness tests, as well as macroexaminations. Tension, torsion, and peel test specimens are examined to determine the location of failure and the size and soundness of the weld nugget.

Tests For Stud Welds. The qualification of stud welding procedures can involve the macroexamination of cross sections, bend tests, tensile tests, or torsion tests. Sectioning and macroexamination can determine whether or not the welds and HAZs have the proper fusion and are free of cracks. Bend tests for stud welds involve bending the stud through a minimum angle, usually 15°, and then bending it back to the original position. An alternative test is to bend the studs by hammering them flat onto the test plate. The bent studs must be free of visible separation or fracture. Stud welds also are tested by applying torque or tension loads.

Tests For Weld Cladding and Surfacing. The qualification of weld cladding procedures involves chemical analysis of the deposit, macroexamination of cross sections, and bend tests. The qualification for hard-surfacing welding procedures can include hardness tests, in addition to chemical analysis of the deposit and macroexamination of cross sections.

Special Tests

Special tests for welding procedure qualification often utilize prototype structures or mock-up components that are fabricated per the specified welding procedure. These structures or components can then be subjected to loads that simulate the actual conditions encountered in service. Tests can include static or dynamic tensile loads, bending loads, fatigue loads, or impact loading. After load testing, the structure can also be tested destructively; welds can be sectioned and the cross sections examined for proper penetration, bead placement, profile, and so on. An example of this type of testing, for roll-over protection structures, is described in ANSI/AWS D14.3.

Weld Procedure Qualification

Harvey R. Castner, Edison Welding Institute

Limitations on Procedure Qualification

A new welding procedure must be qualified if it differs significantly from another procedure that has already been qualified. The application range of a welding procedure qualification is determined by placing limits on the welding variables specified in a welding procedure. Codes define these limiting variables as "essential variables," because the WPS must be requalified if the values are changed so that they exceed the specified limits.

Supplementary essential variables are the variables that affect notch toughness. Supplementary variables become essential variables when notch toughness tests are required.

Nonessential variables are the remaining variables that are contained in the WPS, but that can be changed without requiring requalification. An essential variable for one process may be nonessential for another process.

Although there are differences in how codes and standards define essential variables, the variables described below are considered to be essential by most codes and standards.

Welding Process. A change of welding process will require requalification, because the process has a significant effect on the mechanical properties and metallurgical structure of the weld.

Base Metal. Changing from one type of base metal to another can affect weldability, as well as the metallurgical and mechanical properties of the weld. For certain welding processes, changes in base metals also can cause differences in penetration and weld fusion characteristics. Therefore, significant changes in base-metal type, composition, heat treatment, or processing will require requalification.

Some codes require separate welding procedure qualifications for pipe and plate. Although a procedure qualification on pipe may also cover plate, a procedure qualification on plate might be restricted to the welding of large-diameter pipes. Base-metal thickness influences cooling rate and, therefore, metallurgical and mechanical properties. The composition of a metal may also change with thickness. Some codes require qualification on the exact thickness of metal that is to be used for production welds. Other codes require qualification on the thinnest and thickest metals that are used in production. Most codes permit qualification on one thickness of test specimen to cover a range of base-metal thicknesses.

Codes often group together those base metals that have similar weldability characteristics, metallurgical properties, chemical compositions, or mechanical properties. The grouping of base metals reduces the number of required qualifications by allowing the qualification of one of the materials in the group to qualify the other base metals in that group.

Filler Metal. Codes group filler metals and welding electrodes with similar chemical compositions and operating characteristics together. Grouping reduces the number of tests by allowing the qualification of one of the filler metals in the group to also qualify the others in that group. Ferrous filler metals can be grouped by both chemical composition and usability. The addition or deletion of a consumable insert is an essential variable.

Position is not an essential variable for all codes and specifications, which may allow qualification in one position to cover qualification in another position. For example, qualification in the horizontal position may also qualify for the flat position.

Joint design is an essential variable for some codes, but not for others. Weld joint design should be an essential variable for processes in which joint changes affect metallurgical structures and weld properties. These processes include stud, laser-beam, electron-beam, and friction welding.

Electrical Characteristics. Variations in the type of welding current (alternating or direct, for arc welds) or in polarity normally require requalification. A small change in welding current or voltage does not typically require requalification. Many codes limit changes to ± 10% from the mean value, without necessitating requalification. The use of current pulsation and the pulse parameters may be essential variables. For example, beam current or beam power are essential variables for both laser- and electron-beam welding.

Shielding. The use and the composition of a shielding gas are essential variables for arc welding, friction welding, and laser-beam welding processes. An essential variable for electron-beam welding is the use of a vacuum.

Preheat and Interpass Temperatures. Significant changes in preheat or interpass temperatures will affect properties and, therefore, will require requalification. Codes and standards differ on the size of the change in temperature that is permitted. Some codes limit changes to a maximum of 14 °C (25 °F), whereas others permit an increase or decrease of up t 56 °C (100 °F) before requalification is required.

Postweld Heat Treatment. The use of postweld heat treatment, as well as changes in the postweld heat-treatment temperature or time at temperature, will require requalification.

Other variables that affect the soundness, metallurgical structure, or mechanical properties of the weld are essential variables. These variables depend on the welding process being used. Examples of other essential variables include:

- CHANGES IN MEAN TRAVEL SPEED OF MORE THAN 10%
- CHANGING FROM A STRINGER TO A WEAVE BEAD, OR CHANGING THE DIRECTION OF WELDING FOR VERTICAL ARC WELDING PROCEDURES
- CHANGES IN ELECTRODE COMPOSITION OR CONFIGURATION FOR RESISTANCE WELDING
- CHANGES IN FILAMENT TYPE, SIZE, OR SHAPE, AND IN GUN-TO-WORK DISTANCE FOR ELECTRON-BEAM WELDING
- CHANGES IN FOCAL LENGTH, FOCUS POSITION, OR SPOT SIZE FOR LASER-BEAM WELDING
- CHANGES IN ENERGY, APPLIED LOADS, OR THE AMOUNT OF UPSET FOR FRICTION WELDING

Weld Procedure Qualification

Harvey R. Castner, Edison Welding Institute

Limitations on Procedure Qualification

A new welding procedure must be qualified if it differs significantly from another procedure that has already been qualified. The application range of a welding procedure qualification is determined by placing limits on the welding variables specified in a welding procedure. Codes define these limiting variables as "essential variables," because the WPS must be requalified if the values are changed so that they exceed the specified limits.

Supplementary essential variables are the variables that affect notch toughness. Supplementary variables become essential variables when notch toughness tests are required.

Nonessential variables are the remaining variables that are contained in the WPS, but that can be changed without requiring requalification. An essential variable for one process may be nonessential for another process.

Although there are differences in how codes and standards define essential variables, the variables described below are considered to be essential by most codes and standards.

Welding Process. A change of welding process will require requalification, because the process has a significant effect on the mechanical properties and metallurgical structure of the weld.

Base Metal. Changing from one type of base metal to another can affect weldability, as well as the metallurgical and mechanical properties of the weld. For certain welding processes, changes in base metals also can cause differences in penetration and weld fusion characteristics. Therefore, significant changes in base-metal type, composition, heat treatment, or processing will require requalification.

Some codes require separate welding procedure qualifications for pipe and plate. Although a procedure qualification on pipe may also cover plate, a procedure qualification on plate might be restricted to the welding of large-diameter pipes. Base-metal thickness influences cooling rate and, therefore, metallurgical and mechanical properties. The composition of a metal may also change with thickness. Some codes require qualification on the exact thickness of metal that is to be used for production welds. Other codes require qualification on the thinnest and thickest metals that are used in production. Most codes permit qualification on one thickness of test specimen to cover a range of base-metal thicknesses.

Codes often group together those base metals that have similar weldability characteristics, metallurgical properties, chemical compositions, or mechanical properties. The grouping of base metals reduces the number of required qualifications by allowing the qualification of one of the materials in the group to qualify the other base metals in that group.

Filler Metal. Codes group filler metals and welding electrodes with similar chemical compositions and operating characteristics together. Grouping reduces the number of tests by allowing the qualification of one of the filler metals in the group to also qualify the others in that group. Ferrous filler metals can be grouped by both chemical composition and usability. The addition or deletion of a consumable insert is an essential variable.

Position is not an essential variable for all codes and specifications, which may allow qualification in one position to cover qualification in another position. For example, qualification in the horizontal position may also qualify for the flat position.

Joint design is an essential variable for some codes, but not for others. Weld joint design should be an essential variable for processes in which joint changes affect metallurgical structures and weld properties. These processes include stud, laser-beam, electron-beam, and friction welding.

Electrical Characteristics. Variations in the type of welding current (alternating or direct, for arc welds) or in polarity normally require requalification. A small change in welding current or voltage does not typically require requalification. Many codes limit changes to ± 10% from the mean value, without necessitating requalification. The use of current pulsation and the pulse parameters may be essential variables. For example, beam current or beam power are essential variables for both laser- and electron-beam welding.

Shielding. The use and the composition of a shielding gas are essential variables for arc welding, friction welding, and laser-beam welding processes. An essential variable for electron-beam welding is the use of a vacuum.

Preheat and Interpass Temperatures. Significant changes in preheat or interpass temperatures will affect properties and, therefore, will require requalification. Codes and standards differ on the size of the change in temperature that is permitted. Some codes limit changes to a maximum of 14 °C (25 °F), whereas others permit an increase or decrease of up t 56 °C (100 °F) before requalification is required.

Postweld Heat Treatment. The use of postweld heat treatment, as well as changes in the postweld heat-treatment temperature or time at temperature, will require requalification.

Other variables that affect the soundness, metallurgical structure, or mechanical properties of the weld are essential variables. These variables depend on the welding process being used. Examples of other essential variables include:

- CHANGES IN MEAN TRAVEL SPEED OF MORE THAN 10%
- CHANGING FROM A STRINGER TO A WEAVE BEAD, OR CHANGING THE DIRECTION OF WELDING FOR VERTICAL ARC WELDING PROCEDURES
- CHANGES IN ELECTRODE COMPOSITION OR CONFIGURATION FOR RESISTANCE WELDING
- CHANGES IN FILAMENT TYPE, SIZE, OR SHAPE, AND IN GUN-TO-WORK DISTANCE FOR ELECTRON-BEAM WELDING
- CHANGES IN FOCAL LENGTH, FOCUS POSITION, OR SPOT SIZE FOR LASER-BEAM WELDING
- CHANGES IN ENERGY, APPLIED LOADS, OR THE AMOUNT OF UPSET FOR FRICTION WELDING

Weld Procedure Qualification

Harvey R. Castner, Edison Welding Institute

Documentation of Procedure Qualification

Record keeping is an essential part of welding procedure qualification. The WPS defines the variables and methods to be used for welding. As noted earlier, a procedure qualification record is used to document the actual welding variables that are used for the procedure qualification test weldment, as well as the results of the tests. The manufacturer or contractor should have these records available for review by customers or authorized inspectors.

Weld Procedure Qualification

Harvey R. Castner, Edison Welding Institute

Selected References

- "STANDARD FOR WELDING PROCEDURE AND PERFORMANCE QUALIFICATION," ANSI/AWS B2.1, AWS
- "STANDARD METHODS FOR MECHANICAL TESTING OF WELDS," ANSI/AWS B4.0, AWS
- "STRUCTURAL WELDING CODE, STEEL," ANSI/ AWS D1.1, AWS
- "WELDING AND BRAZING PROCEDURE AND PERFORMANCE QUALIFICATION," MIL-STD 248, U.S. DEPARTMENT OF THE NAVY
- WELDING AND BRAZING QUALIFICATIONS, "BOILER AND PRESSURE VESSEL CODE," ANSI/ASME BPV-IX, ASME
- *WELDING HANDBOOK,* 8TH ED., VOL 1, AWS, 1987

Residual Stresses and Distortion

Koichi Masubuchi, Massachusetts Institute of Technology

Introduction

COMPLEX THERMAL STRESSES occur in parts during welding due to the localized application of heat. Residual stresses and distortion remain after welding is completed. Transient thermal stresses, residual stresses, and distortion sometimes cause cracking and mismatching of joints. High tensile residual stresses in areas near the weld can cause premature failures of welded structures under certain conditions. Distortion, especially out-of-plane distortion, and compressive residual stresses in the base plate can reduce the buckling strength of a structural member subjected to compressive loading. Correction of unacceptable distortion is costly and, in some cases, impossible. The subjects related to residual stresses, distortion, and their consequences are too broad to be discussed in depth in this article. Those who need more detailed information are advised to refer to the references, especially Ref 1, 2, 3.

References

1. K. MASUBUCHI, *ANALYSIS OF WELDED STRUCTURES--RESIDUAL STRESSES, DISTORTION, AND THEIR CONSEQUENCES,* PERGAMON PRESS, 1980
2. K. MASUBUCHI, RESIDUAL STRESSES AND DISTORTION, *METALS HANDBOOK,* 9TH ED., VOL 6, *WELDING, BRAZING, AND SOLDERING,* ASM, 1983, P 856-895
3. K. MASUBUCHI, O.W. BLODGETT, S. MATSUI, E.P. ROSS, AND C.L. TSAI, RESIDUAL STRESSES AND DISTORTION, *WELDING HANDBOOK,* 8TH ED., VOL 1, *WELDING TECHNOLOGY,* AWS, 1987

Residual Stresses and Distortion

Koichi Masubuchi, Massachusetts Institute of Technology

Formation of Residual Stresses and Distortion

Residual stresses, also referred to as internal stresses, initial stresses, inherent stresses, and locked-in stresses, are stresses that exist in a body after all external loads are removed. Residual stresses also occur when a body is subjected to nonuniform temperature changes; these stresses are called thermal stresses. The intensity of stress is usually expressed in load or force per unit area, such as newtons per square meter (N/m^2) or pascals (Pa), kilograms force per square millimeter (kg/mm^2), or pounds per square inch (psi).

Residual stresses in metal structures occur for many reasons during various manufacturing stages, including casting, rolling, bending, flame cutting, forging, machining, and grinding. Heat treatments at various stages also influence residual stresses. For example, quenching treatments produce residual stresses, while stress-relieving heat treatments may reduce such stresses. Areas in which residual stresses exist vary greatly--from microscopic areas to large sections of the metal structure. This article focuses primarily on macroscopic residual stresses. Residual stresses can be classified into two groups according to the mechanisms that produce them:

- STRESSES PRODUCED BY STRUCTURAL MISMATCH
- STRESSES PRODUCED BY AN UNEVEN DISTRIBUTION OF NONELASTIC STRAINS, INCLUDING PLASTIC AND THERMAL STRAINS

Thermal Stresses and Metal Movement During Welding

Because a weldment is locally heated by the welding heat source, the temperature distribution is not uniform and changes as welding progresses. During the welding thermal cycle, complex transient thermal stresses are produced in the weldment and the surrounding joint. The weldment also undergoes shrinkage and deformation during solidification and cooling.

Thermal Stresses During Welding. Figure 1 schematically shows changes in temperature and resulting stresses that occur during welding by examining a bead-on-plate weld of a thin plate made along the *x*-axis. The welding arc, which is moving at a speed *v*, is presently located at the origin, *O*, as shown in Fig. 1(a). The area where plastic deformation occurs during the welding thermal cycle is shown by the shaded area, *M-M* ', in Fig. 1(a). The region where the metal is molten is indicated by the ellipse near *O.* The region outside the shaded area remains elastic throughout the entire welding thermal cycle.

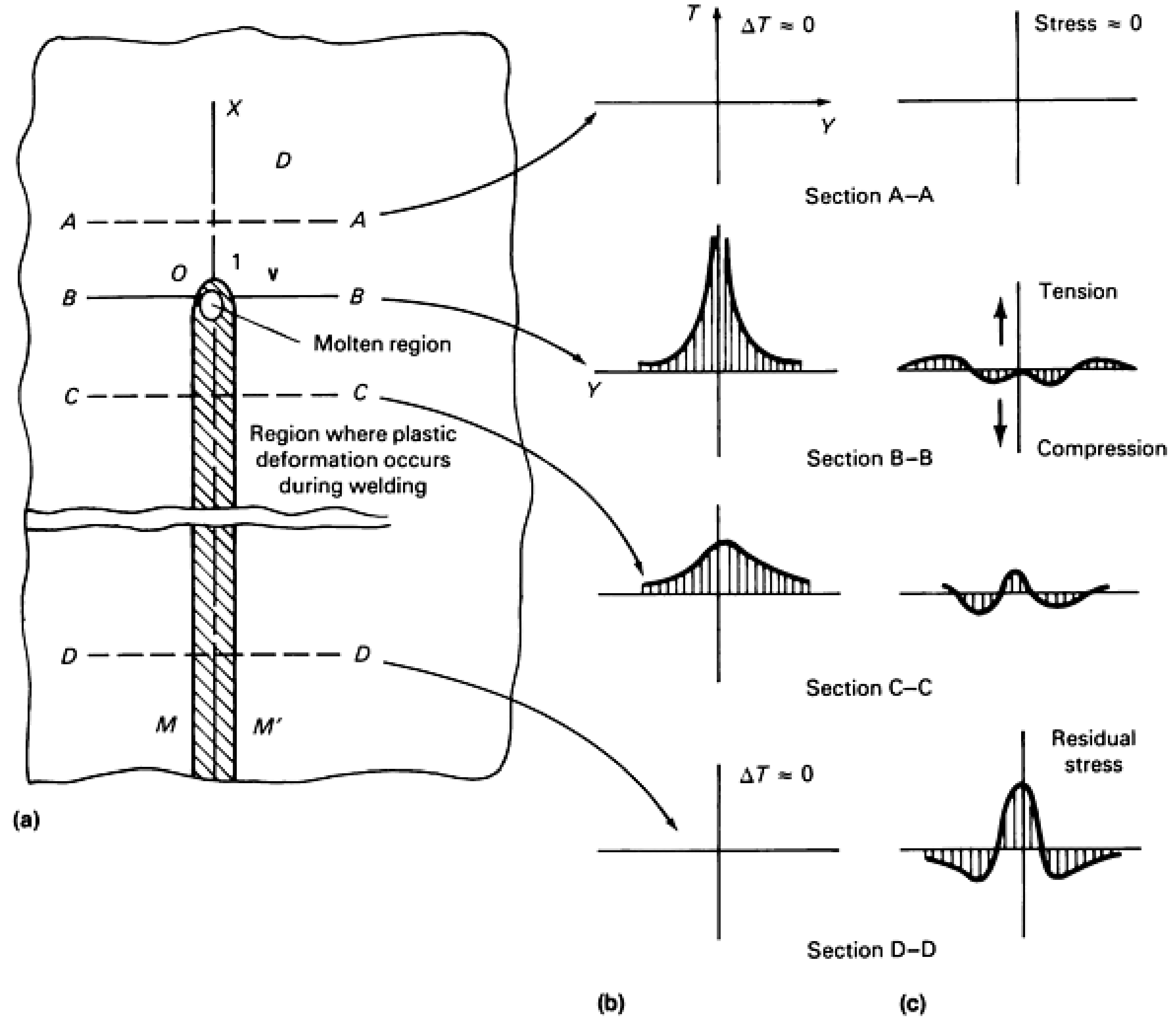

FIG. 1 SCHEMATIC REPRESENTATION OF CHANGES IN TEMPERATURE AND STRESSES DURING WELDING. (A) WELD. (B) TEMPERATURE CHANGE. (C) STRESS, σ_X. STRESS DISTRIBUTION IS SHOWN IN THE PLANE STRESS CONDITION. THEREFORE, STRESSES ARE REGARDED AS BEING UNIFORM IN THE THICKNESS DIRECTION. SOURCE: WELDING RESEARCH COUNCIL

Temperature gradients along several cross sections through the weld bead path are indicated in Fig. 1 (b). In the base metal ahead of the welding arc (denoted as section A-A), the slope of the temperature gradient due to welding ($\Delta T/\Delta Y$) is almost zero. However, along section B-B, which crosses the welding arc, the slope becomes very steep. Along section C-C, somewhat behind the welding arc, the slope becomes less steep. The slope of the temperature gradient due to welding once again approaches zero along section D-D, which is some distance behind the welding arc. Figure 1(c) shows the distribution of normal stress in the *x*-direction (σ_x) along the cross sections. Normal stress in the *y*-direction (σ_y) and shearing stress (t_{xy}) also exist in a two-dimensional stress field. In a three-dimensional stress field, six stress components exist: σ_x, σ_y, σ_z, t_{xy}, t_{yz}, and t_{zx}.

Figure 1(c) shows the distribution (along the *y*-axis) of normal stress in the *x*-direction (σ_x) due to welding (that is, thermal stress). Along section A-A, the stresses are almost zero. The stress distribution along section B-B is complicated. Beneath the welding arc, stresses are close to zero, because molten metal does not support shear loading. Moving away from the arc, stresses become compressive, because expansion of the metal surrounding the weld pool is restrained by the base metal. Because the temperatures of these areas are quite high, the yield stress of the material becomes quite low. In other words, a situation occurs in which stresses in these areas are as high as the yield stress of the base metal at corresponding temperatures. The magnitude of compressive stress passes through a maximum with increasing distance from the weld or with decreasing temperature. However, stresses occurring in regions farther away from the welding arc are tensile in nature and balance with compressive stresses in areas near the weld pool.

Figure 1(b) shows that along section C-C the weld-metal and base-metal regions have cooled. As they shrink, tensile stresses are caused in regions in and adjacent to the weld (Fig. 1c). As the distance from the weld increases, stresses become compressive. High tensile stresses exist along section D-D in and adjacent to the weld. Compressive stresses are produced in areas away from the weld.

Equilibrium Condition of Residual Stresses. Because residual stresses exist without external forces, the resultant force and the resultant moment produced by the residual stresses must vanish on any plane section:

$$\int \sigma \cdot DA = 0 \qquad \textbf{(EQ 1)}$$

$$\int DM = 0 \qquad \textbf{(EQ 2)}$$

where *dA* is area and *dM* is the resultant moment. Residual stress data must be checked in any experiment to ensure that they satisfy the above condition.

Metal Movement During Welding. During welding, the weldment undergoes shrinkage and deformation. The transient deformation, or metal movement, is most evident when the weld line is away from the neutral axis of the weldment, causing a large amount of bending moment. Figure 2 schematically shows how a rectangular plate distorts when its longitudinal edge is heated by a welding arc or an oxyacetylene torch (for cutting, welding, or flame heating). Areas near the heat source, or the upper regions of the rectangular plate, are heated to higher temperatures and thus expand more than areas away from the heat source, or the lower regions of the plate. Therefore, the plate first deforms as shown by curve *AB.*

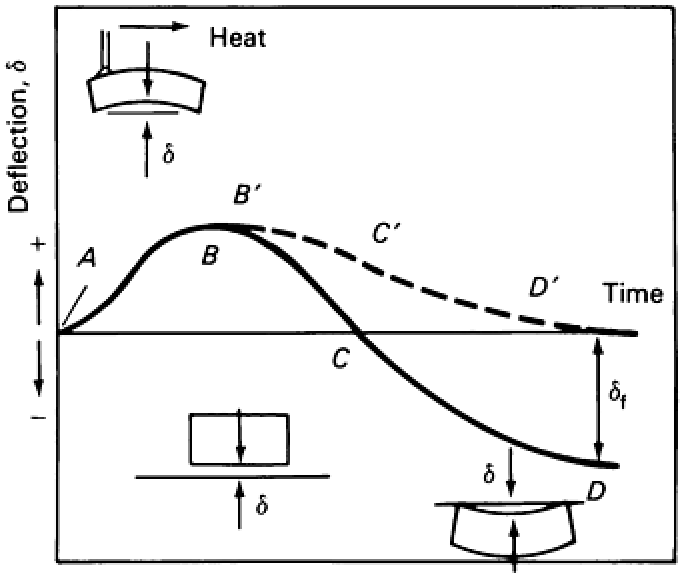

FIG. 2 CHANGES OF DEFLECTION AT THE CENTER OF THE LOWER EDGE OF A RECTANGULAR PLATE DUE TO HEATING BY A HEAT SOURCE MOVING ALONG THE UPPER EDGE AND SUBSEQUENT COOLING

If all the parts of the material remained completely elastic during the entire thermal cycle, thermal stresses produced during the heating and cooling cycle would disappear when the temperature returned to the initial temperature. Deformation of the plate during the welding process would be indicated by curve *AB'C'D'*, resulting in no deformation after the thermal cycle. However, in most practical materials, plastic strains are produced in areas heated to elevated temperatures, causing plastic upsetting. The thermal stresses do not disappear when the temperature returns to the initial temperature, causing residual stresses. Transient deformation of the plate during heating and cooling is shown by curve *ABCD.* After the plate cools to the initial temperature, final deformation in the amount δ_f remains, which also is called distortion. The metal movement (that is, transient distortion or transient deformation) during welding and the distortion after welding is completed occur in opposite directions, generally with the same order of magnitude.

In summary, some basic characteristics of residual stresses and distortion are (Ref 4):

- THE METAL MOVEMENT DURING WELDING (EXPANSION) IS OPPOSITE TO THE DISTORTION THAT REMAINS AFTER WELDING.
- ALTHOUGH WELDING IS COMPLETED IN A SHORT PERIOD (IN A FEW MINUTES, FOR EXAMPLE), RESIDUAL STRESSES AND DISTORTION TAKE A RELATIVELY LONG TIME TO DEVELOP COMPLETELY (30 MIN TO 1H)
- BECAUSE MOST OF THE NONELASTIC STRAINS THAT CAUSE RESIDUAL STRESSES AND DISTORTIONS ARE PRODUCED DURING WELDING, NECESSARY ACTIONS MUST BE PERFORMED DURING WELDING TO EFFECTIVELY CONTROL AND REDUCE RESIDUAL STRESSES AND DISTORTION.

Analyses of Residual Stresses and Distortion in Weldments

Residual stresses and distortion in welded structures have been recognized and studied since the 1930s. Analyses of these subjects require complex computation; therefore, most early studies were primarily empirical or limited to the analysis of simple cases. With the advancement of modern computers and computational techniques (for example, the finite-element and finite-difference method), a renewed effort has been made in recent years to study residual stresses and related phenomena. Therefore, it is now possible using computer programs to simulate the transient thermal stresses and metal movement during welding as well as the residual stresses and distortion that remain after welding is completed (Ref 5, 6, 7, 8, 9). However, because these analysis methods are very complex, most are not practical for analyzing and controlling residual stresses and distortion in weldments.

References cited in this section

4. K. MASUBUCHI, RESEARCH ACTIVITIES EXAMINE RESIDUAL STRESSES AND DISTORTION IN WELDED STRUCTURES, *WELD. J.,* VOL 70 (NO. 12), 1991, P 41-47
5. L. TALL, RESIDUAL STRESSES IN WELDED PLATES--A THEORETICAL STUDY, *WELD. J.,* VOL 43 (NO. 1), 1991, P 10S-23S
6. H.D. HIBBIT AND P.V. MARCEL, "A NUMERICAL THERMOMECHANICAL MODEL FOR THE WELDING AND SUBSEQUENT LOADING OF A FABRICATED STRUCTURE," NSRDC CONTRACT NO. N00014-67-A-019-0006, TECH. REP. NO. 2, DEPARTMENT OF THE NAVY, MARCH 1972
7. T. MURAKI, J.J. BRYAN, AND K. MASUBUCHI, ANALYSIS OF THERMAL STRESSES AND METAL MOVEMENT DURING WELDING--PART I: ANALYTICAL STUDY; PART II: COMPARISON OF EXPERIMENTAL DATA AND ANALYTICAL RESULTS, *J. ENG. MATER. TECHNOL. (TRANS. ASME),* 1975, P 81-84, 85-91
8. E.F. RYBICKI *ET AL.,* A FINITE-ELEMENT MODEL FOR RESIDUAL STRESSES AND DEFLECTION IN GIRTH-BUTT WELDED PIPES, *J. PRESSURE VESSEL TECHNOL. (TRANS. ASME),* VOL 100, 1978, P 256-262
9. P.O. DEXTER AND D. PONT, "EVALUATION OF AVAILABLE WELDING SIMULATION SOFTWARE," PREPARED FOR EPRI UNDER RESEARCH PROJECT C102-3 BY SOUTHWEST RESEARCH INSTITUTE, MAY 1991

Residual Stresses and Distortion

Koichi Masubuchi, Massachusetts Institute of Technology

Techniques for Measuring Residual Stresses

Many techniques have been used to measure residual stresses in metals (Ref 1, 10, 11). Table 1 lists most of the presently available measurement techniques and classifies them as follows:

- STRESS-RELAXATION TECHNIQUES
- X-RAY DIFFRACTION TECHNIQUES
- TECHNIQUES USING STRESS-SENSITIVE PROPERTIES
- CRACKING TECHNIQUES

TABLE 1 CLASSIFICATION OF TECHNIQUES FOR MEASURING RESIDUAL STRESS

A-1 STRESS RELAXATION TECHNIQUES USING ELECTRIC AND MECHANICAL STRAIN GAGES
PLATE
SECTIONING TECHNIQUE USING ELECTRIC RESISTANCE STRAIN GAGES
GUNNERT TECHNIQUE
MATHAR-SOETE DRILLING TECHNIQUE
STÄBLEIN SUCCESSIVE MILLING TECHNIQUE
SOLID CYLINDERS AND TUBES
HEYN-BAUER SUCCESSIVE MACHINING TECHNIQUE
MESNAGER-SACHS BORING-OUT TECHNIQUE
THREE-DIMENSIONAL SOLIDS
GUNNERT DRILLING TECHNIQUE
ROSENTHAL-NORTON SECTIONING TECHNIQUE
A-2 STRESS RELAXATION TECHNIQUES USING APPARATUS OTHER THAN ELECTRIC AND MECHANICAL STRAIN GAGES
GRID SYSTEM-DIVIDING TECHNIQUE
BRITTLE COATING-DRILLING TECHNIQUE
PHOTOELASTIC COATING-DRILLING TECHNIQUE
B X-RAY DIFFRACTION TECHNIQUES
X-RAY FILM TECHNIQUE
X-RAY DIFFRACTOMETER TECHNIQUE
C TECHNIQUES USING STRESS-SENSITIVE PROPERTIES
ULTRASONIC TECHNIQUES
POLARIZED ULTRASONIC WAVE TECHNIQUE
ULTRASONIC ATTENUATION TECHNIQUE
HARDNESS TECHNIQUES
D CRACKING TECHNIQUES
HYDROGEN-INDUCED CRACKING TECHNIQUE
STRESS-CORROSION CRACKING TECHNIQUE

Source: Ref 12

In the stress-relaxation technique, residual stress is determined by measuring the elastic-strain release that takes place when a specimen is cut into pieces or has a piece removed. In most cases, electric or mechanical gages are used for measuring this strain release (group A-1 in Table 1). A variety of such techniques exist (that is, there are many ways to section a specimen to determine residual stresses). Certain techniques are used to study stresses in a plate, while others are used for cylinders, tubes, or three-dimensional solids. Strain release caused by stress relaxation can also be determined by using grid systems, brittle coatings, or photoelastic coatings (group A-2 in Table 1). An inherent disadvantage of stress-relaxation techniques is that they are destructive (that is, the specimen must be partially or entirely sectioned). Nevertheless, the stress-relaxation techniques provide reliable quantitative data and are the most widely used method for measuring residual stresses in weldments.

In metals with crystalline structures, elastic strains can also be determined by using x-ray diffraction to measure the lattice parameter. Because the lattice parameter of a metal in the unstressed state is known or can be determined separately, elastic strain in the metal can be measured nondestructively without machining or drilling. Two methods are presently available: the x-ray film technique and the x-ray diffractometer technique. With x-ray diffraction, surface strains can be determined within a small area--for example, to a depth of 0.025 mm (0.001 in.) and a diameter of 0.0025 mm (0.0001 in.). Modern x-ray diffraction stress measurement instrumentation can provide one complete measurement in time scales as short as a fraction of a second, and typically less than 1 min. However, x-ray diffraction techniques have some disadvantages. They are time consuming, and they are not accurate, especially in situations where high temperatures have distorted the atomic structure of the material. Studies have recently been made using diffraction of neutrons, which can penetrate deeper into the metal than x-rays (Ref 13).

Attempts have been made to determine the residual stresses in metals by measuring their stress-sensitive properties. The following stress-measurement techniques have been proposed:

- ULTRASONIC TECHNIQUES, SUCH AS THE POLARIZED ULTRASONIC WAVE TECHNIQUE (WHICH MAKES USE OF STRESS-INDUCED CHANGES IN THE ANGLE OF POLARIZATION OF POLARIZED ULTRASONIC WAVES) AND THE ULTRASONIC ATTENUATION TECHNIQUE (WHICH MAKES USE OF STRESS-INDUCED CHANGES IN THE RATE OF ABSORPTION OF ULTRASONIC WAVES)
- HARDNESS TECHNIQUES, WHICH MAKE USE OF STRESS-INDUCED CHANGES IN HARDNESS

However, none of these techniques has been developed beyond the laboratory stage, and they have not been used successfully for the measurement of residual stresses in weldments.

Another technique developed for the study of residual stresses involves the close observation of cracks induced in the specimen by hydrogen embrittlement or stress-corrosion cracking (SCC). Cracking techniques are useful when studying residual stresses in complex structural models in which the distribution of residual stresses is complicated, but these techniques provide only qualitative, not quantitative, data.

References cited in this section

1. K. MASUBUCHI, *ANALYSIS OF WELDED STRUCTURES--RESIDUAL STRESSES, DISTORTION, AND THEIR CONSEQUENCES,* PERGAMON PRESS, 1980
10. R.G. TREUTING, J.J. LYNCH, H.B. WISHART, AND D.G. RICHARDS, *RESIDUAL STRESS MEASUREMENTS,* ASM, 1952
11. K. MASUBUCHI, "NONDESTRUCTIVE MEASUREMENT OF RESIDUAL STRESSES IN METALS AND METAL STRUCTURES," RSIC-410, REDSTONE SCIENTIFIC INFORMATION CENTER, REDSTONE ARSENAL, 1965
12. K. MASUBUCHI, RESIDUAL STRESSES AND DISTORTION, *WELDING HANDBOOK,* VOL 1, 7TH ED., AWS, 1976
13. R.R. HOSBONS, THE USE OF NEUTRON DIFFRACTION TO DETERMINE NONDESTRUCTIVELY THE RESIDUAL STRAIN AND TEXTURE IN WELDS, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* ASM INTERNATIONAL, 1989, P 103-106

Koichi Masubuchi, Massachusetts Institute of Technology

Magnitude and Distribution of Residual Stresses in Weldments

A number of investigators have studied distributions of residual stresses in various types of weldments. Most of the data presented in this section relates to single welds as opposed to multipass welds. References 1, 2, and 3 present detailed information.

Residual Welding Stresses, Reaction Stresses, and Stress Distributions in a Groove Weld. As stated earlier, residual stresses can be classified as (1) stresses produced by structural mismatch or displacement and (2) stresses produced by uneven distribution of nonelastic strains. The same method of classification is applicable to residual stresses in weldments: (1) reaction stresses caused when the weldment is restrained externally and (2) residual stresses produced in an unrestrained weldment.

Figure 3 shows a typical distribution of residual stresses in a groove weld. The significant stresses are those parallel to the weld direction, designated σ_x, which are usually called the longitudinal stresses, and those transverse to σ_x, designated σ_y, which are called the transverse stresses. The distribution of the longitudinal stress is shown in Fig. 3(b). Near the weld, tensile stresses of high magnitude are produced. These taper off quickly and become compressive at a distance equal to several times the width of the weld metal. The stress distribution is characterized by two parameters:

- THE MAXIMUM STRESS AT THE WELD REGION, σ_M
- THE WIDTH OF THE TENSION ZONE OF RESIDUAL STRESS, B

The distribution of longitudinal residual stress, σ_x, can be approximated by:

$$s_x(y) = s_m \left[1 - \left(\frac{y}{b}\right)^2\right] e^{-[1/2(y/b)^2]} \qquad \textbf{(EQ 3)}$$

Equation 3 satisfies the equilibrium conditions given in Eq 1 and 2.

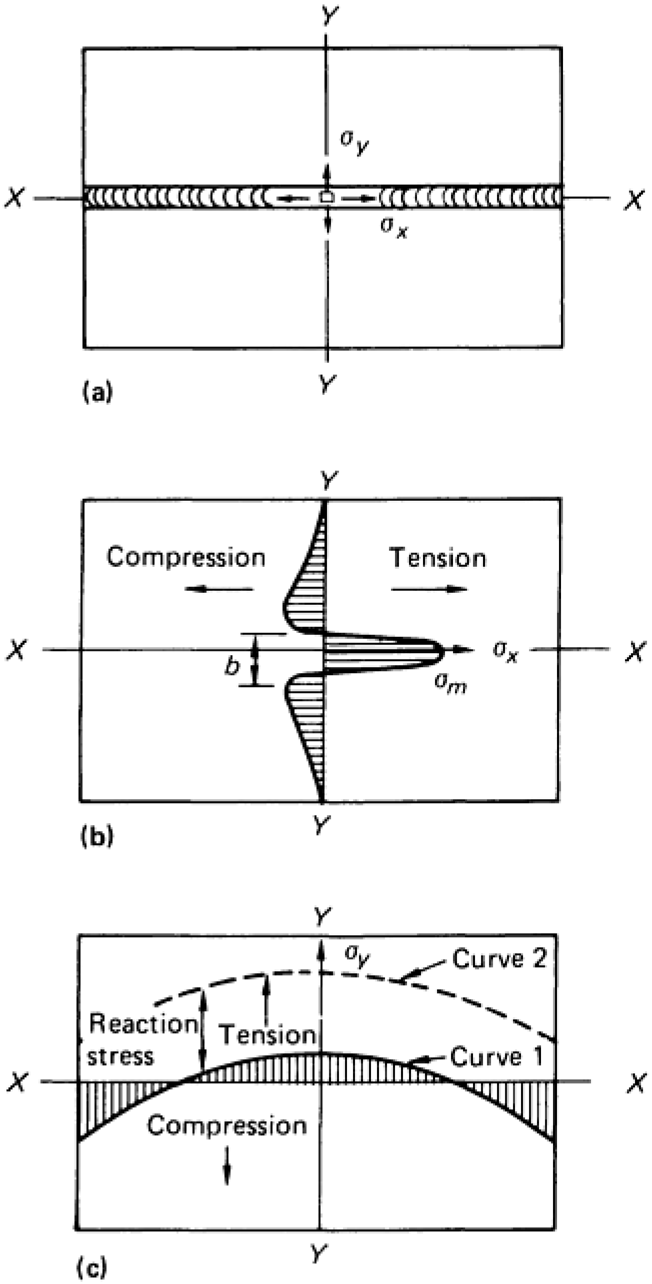

FIG. 3 TYPICAL DISTRIBUTIONS OF RESIDUAL STRESSES IN A GROOVE WELD. (A) GROOVE WELD. (B) DISTRIBUTION OF σ_X ALONG *YY.* (C) DISTRIBUTION OF σ_Y ALONG *XX.* SOURCE: WELDING RESEARCH COUNCIL

Curve 1 of Fig. 3(c) shows the distribution of transverse stress,σ_y, along the length of the weld. In the center of the joint, tensile stresses of relatively low magnitude are produced. Compressive stresses are produced at the ends of the joint. If an external constraint is used to restrain the lateral contraction of the joint, approximately uniform tensile stresses along the weld are added, resulting in curve 2 of Fig. 3(c). However, an external constraint has little influence on the distribution of residual stress (σ_x).

Residual Stresses in a Plug Weld. Figure 4 shows the distribution of both the theoretical and experimental residual stresses in a circular plug weld (Ref 14). Tensile stresses, as high as the yield stress of the material, are produced in both radial and tangential directions in the weld and adjacent areas. In areas away from the weld, radial stresses (σ_r) are tensile, and tangential stresses ($\sigma\theta$) are compressive; both stresses decrease with increasing distance from the weld (*r*).

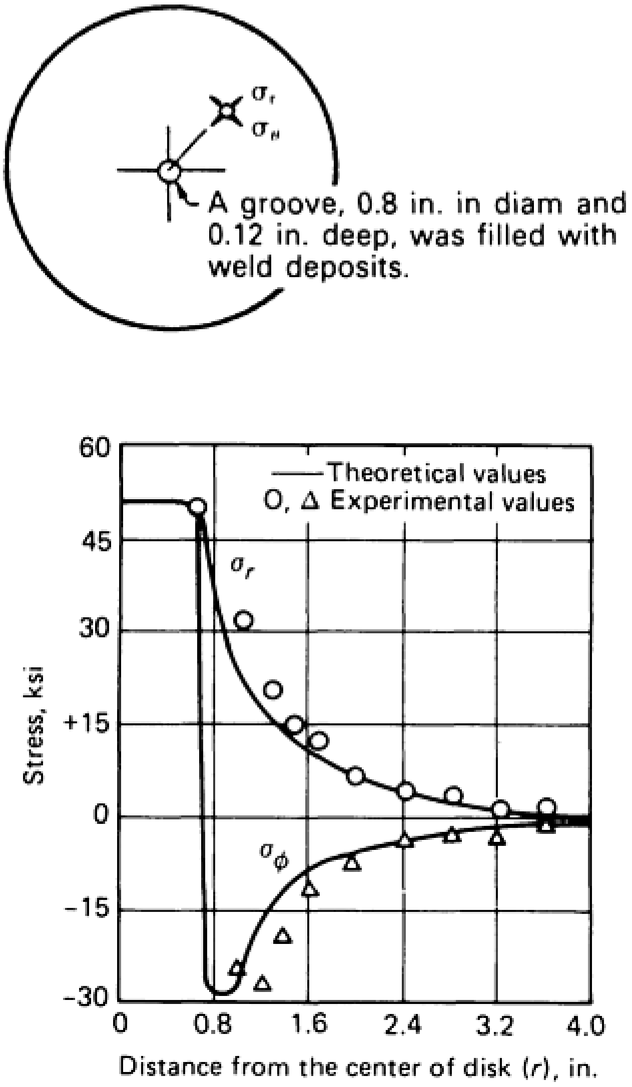

A groove, 0.8 in. in diam and 0.12 in. deep, was filled with weld deposits.

FIG. 4 THEORETICAL AND EXPERIMENTAL DISTRIBUTIONS OF RESIDUAL STRESSES IN A PLUG WELD

Residual Stresses in a Welded Shape and a Column. Figure 5 shows typical distributions of residual stresses and distortion in various welded shapes (Ref 15). As shown in section *xx* of Fig. 5(a), high tensile residual stresses are produced in areas near the weld, in the direction parallel to the weld line (axis). In the flange, longitudinal stresses are compressive (-) in areas away from the weld and tensile (+) in areas near the weld. Tensile stresses in areas near the upper edge of the web are caused by longitudinal bending distortion of the T-shape, which are caused by the longitudinal shrinkage of the weld. Angular distortion also is produced.

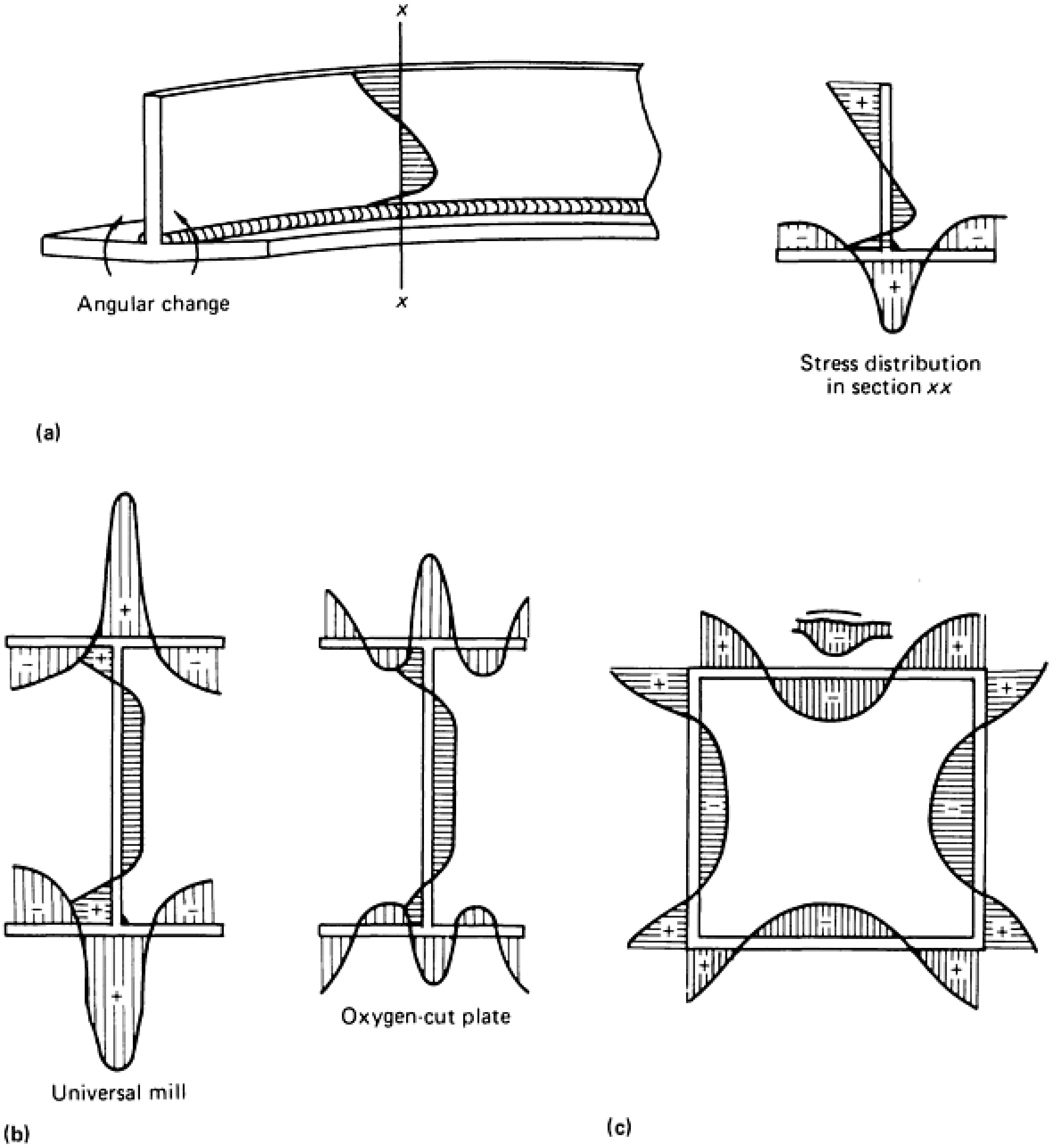

FIG. 5 TYPICAL RESIDUAL STRESSES IN WELDED SHAPES. (A) T-SHAPES. (B) H-SHAPES. (C) BOX SHAPES. COMPRESSIVE STRESS, -; TENSILE STRENGTH, +

Figures 5(b) and 5(c) show typical distributions of residual stress in H-shapes (I-beams) and in a welded box shape, respectively. Residual stresses are shown parallel to the weld line (axis); they are compressive in areas away from the welds and tensile in areas near the welds.

Residual Stresses in Weldments of Selected Materials. Distributions of residual stresses in welds in various metals are similar to those in low-carbon steel welds, discussed above. In most cases, the maximum residual stresses are as high as the yield stress of the material. One exception is ultrahigh-strength steels, such as 4130, D-6a, and H13 (Ref 1). For such steels, peak values of residual stresses in some areas can be as high as the yield stress (or even higher in the case of triaxiality); however, average values of residual stresses in the weld metal are considerably lower than the yield stress.

References cited in this section

1. K. MASUBUCHI, *ANALYSIS OF WELDED STRUCTURES--RESIDUAL STRESSES, DISTORTION, AND THEIR CONSEQUENCES,* PERGAMON PRESS, 1980
2. K. MASUBUCHI, RESIDUAL STRESSES AND DISTORTION, *METALS HANDBOOK,* 9TH ED., VOL 6, *WELDING, BRAZING, AND SOLDERING,* ASM, 1983, P 856-895
3. K. MASUBUCHI, O.W. BLODGETT, S. MATSUI, E.P. ROSS, AND C.L. TSAI, RESIDUAL STRESSES AND DISTORTION, *WELDING HANDBOOK,* 8TH ED., VOL 1, *WELDING TECHNOLOGY,* AWS, 1987
14. H. KIHARA, M. WATANABE, K. MASUBUCHI, AND K. SATOH, *RESEARCHES ON WELDING STRESSES AND SHRINKAGE DISTORTION IN JAPAN,* VOL 4, 60TH ANNIVERSARY SERIES OF THE SOCIETY OF NAVAL ARCHITECTS OF JAPAN, TOKYO, 1959
15. N.R. NAGARAJA RAO, F.R. ESATUAR, AND K. TALL, RESIDUAL STRESSES IN WELDED SHAPES, *WELD. J.,* VOL 39 (NO. 3), 1964, P 295-306S

Residual Stresses and Distortion

Koichi Masubuchi, Massachusetts Institute of Technology

Distortion in Weldments

Distortions found in fabricating structures are caused by three fundamental dimensional changes:

- TRANSVERSE SHRINKAGE
- LONGITUDINAL SHRINKAGE PARALLEL TO THE WELD LINE
- ANGULAR DISTORTION AROUND THE WELD LINE

These dimensional changes are shown in Fig. 6.

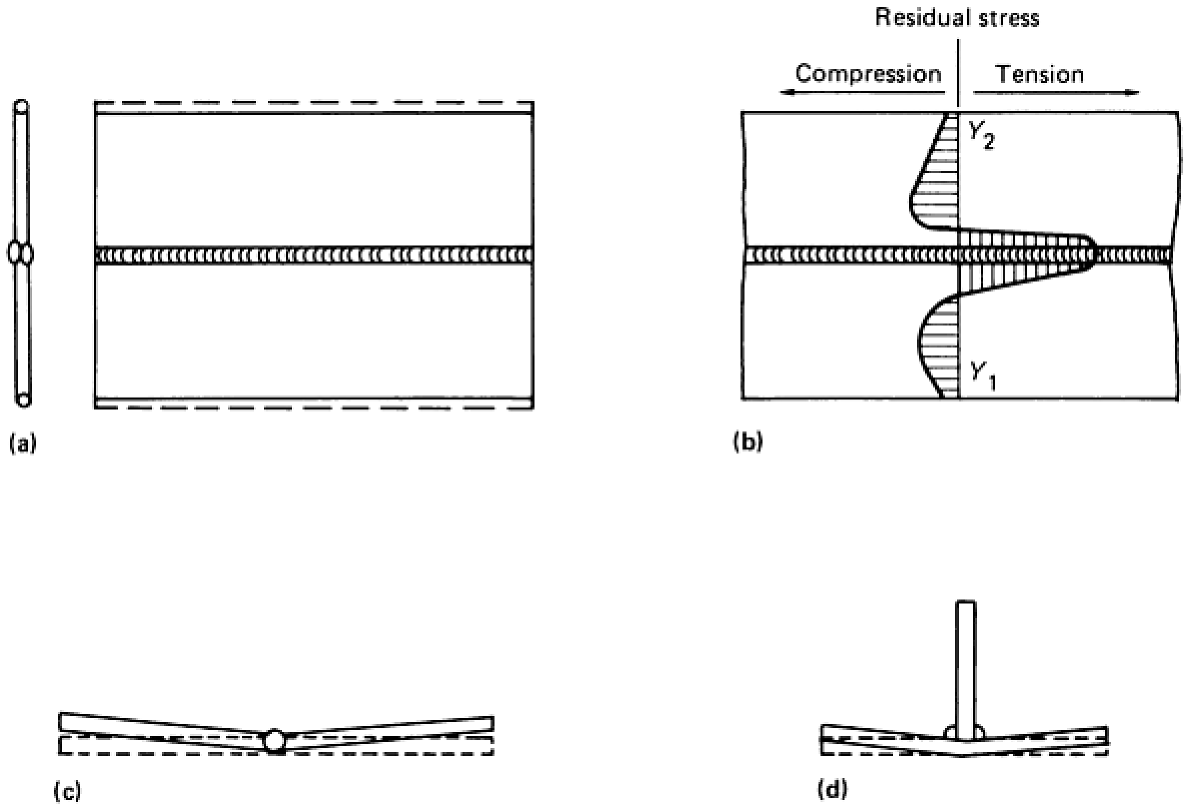

FIG. 6 DIMENSIONAL CHANGES OCCURRING IN WELDMENTS. (A) TRANSVERSE SHRINKAGE IN A GROOVE WELD. (B) LONGITUDINAL SHRINKAGE IN A GROOVE WELD. DISTRIBUTION OF LONGITUDINAL RESIDUAL STRESS, σ_x, IS ALSO SHOWN. (C) ANGULAR CHANGE IN A GROOVE WELD. (D) ANGULAR CHANGE IN A FILLET WELD. SOURCE: WELDING RESEARCH COUNCIL

The shrinkage and distortion that occur during fabrication of actual structures are far more complex than those shown in Fig. 6. For example, when longitudinal shrinkage occurs in a fillet joint, the joint bends longitudinally unless the weld line is located along the neutral axis of the joint. A common structural element is a stiffened panel structure in which longitudinal and transverse stiffeners are fillet welded to a flat plate. Angular changes that occur in the fillet welds cause out-of-plane distortion in the plate, resulting in a shape which looks like a "hungry horse." Presented below are brief discussions of:

- TRANSVERSE SHRINKAGE OF GROOVE WELDS
- ANGULAR DISTORTION OF GROOVE WELDS
- OUT-OF-PLANE DISTORTION CAUSED BY ANGULAR DISTORTION OF FILLET WELDS
- LONGITUDINAL DISTORTION OF BUILT-UP BEAMS
- BUCKLING DISTORTION

Transverse Shrinkage of Square Groove Welds. In Fig. 6(a), shrinkage is uniform along the weld; however, transverse shrinkage that occurs in a groove weld, especially in a long groove weld, is not uniform and is typically much more complex. The major factors causing nonuniform transverse shrinkage in a groove weld are rotational distortion and joint restraint. When two separate planes are joined progressively from one end to the other, the plate edges of the unwelded portion of the joint move, causing a rotational distortion. This can be understood from Fig. 2. The rotational distortion is affected by welding conditions and the location of tack welds. In many practical joints, the degree of joint restraint is not uniform along the weld. Because the amount of transverse shrinkage is less when the joint is more severely restrained, nonuniform distribution of the magnitude of restraint results in an uneven distribution of transverse shrinkage.

Mechanisms of Transverse Shrinkage. The most important finding of mathematical analyses of mechanisms of transverse shrinkage is that the major portion of the transverse shrinkage of a groove joint is caused by contraction of the base plate adjacent to the weld. The base plate near the weld expands during welding. When the weld metal solidifies, the expanded base metal must shrink; this shrinkage accounts for the major part of transverse shrinkage. Shrinkage of the weld metal accounts for about 10% of the actual total shrinkage. Therefore, most formulas developed for computing transverse shrinkage of a square groove weld are composed of two parts:

- SHRINKAGE OF THE ADJACENT BASE PLATE
- SHRINKAGE OF THE WELD METAL

Transverse Shrinkage During Multipass Welding. Welding of a square groove joint in a thick plate commonly requires a number of passes. Figure 7(a) shows how transverse shrinkage increases during multipass welding of a groove joint in 19 mm ($\frac{3}{4}$ in.) thick low-carbon steel plates. Shrinkage is pronounced during the early weld passes, but diminishes during later passes because of the resistance to shrinkage that increases as the weld metal is increased. The effects of transverse shrinkage on various factors, including root opening, joint design, electrode diameter, and degree of restraint, have been studied extensively (Ref 1).

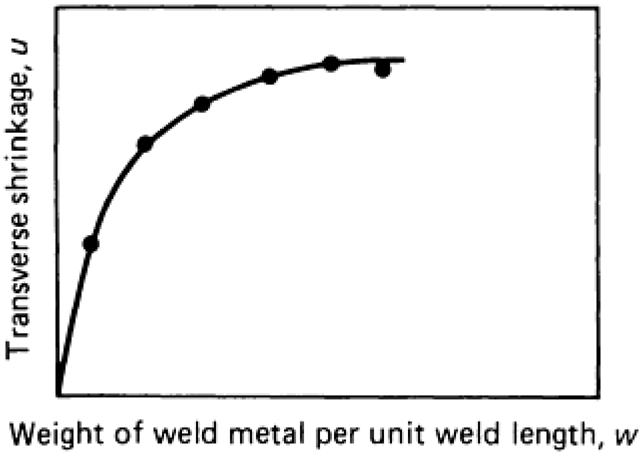

FIG. 7 INCREASE OF TRANSVERSE SHRINKAGE DURING MULTIPASS WELDING OF A GROOVE JOINT. SPECIMENS WERE 19 MM ($\frac{3}{4}$ IN.) THICK LOW-CARBON STEEL PLATES.

Transverse Shrinkage in a Square Groove Weld in Aluminum. As stated earlier, the major portion of transverse shrinkage of a square groove weld is caused by contraction of the base plate adjacent to the weld. Compared with steel, aluminum has a much higher heat conductivity and coefficient of thermal expansion; therefore, transverse shrinkage of an aluminum groove weld is approximately three times greater than that of a steel groove weld of similar dimensions (Ref 1). Because the value of thermal conductivity of aluminum is much greater than that of steel, heat generated by the welding arc dissipates faster in aluminum than in steel, resulting in a broader HAZ (Ref 2).

Angular Distortion of Groove Welds. In a groove weld, when transverse shrinkage is not uniform in the thickness direction, angular distortion often occurs. The angular change of a weld in a thick plate can be minimized by properly selecting the ratio of the size of the weld groove on the front surface to that of the back surface (Ref 1).

Out-of-Plane Distortion Caused by Angular Distortion in Fillet Welds. A stiffened panel composed of a flat plate with longitudinal and transverse stiffeners fillet welded to the plate is widely used for many structures. Angular changes that occur at fillet welds cause out-of-plane distortion of the flat plate (Ref 1). Figure 8 compares values of distortion at the panel center (δ_m) for steel and aluminum welded structures on the basis of the same fillet size (D_f) for selected plate thicknesses. Distortion in aluminum structures is less than in steel structures because the temperature distribution in the thickness direction is more uniform in an aluminum weld than in a steel weld (Ref 2).

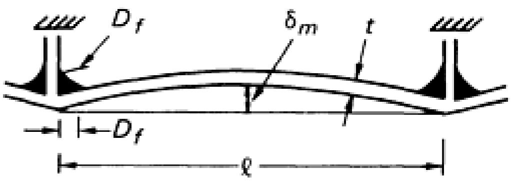
D_f
δ_m
t
D_f
ℓ

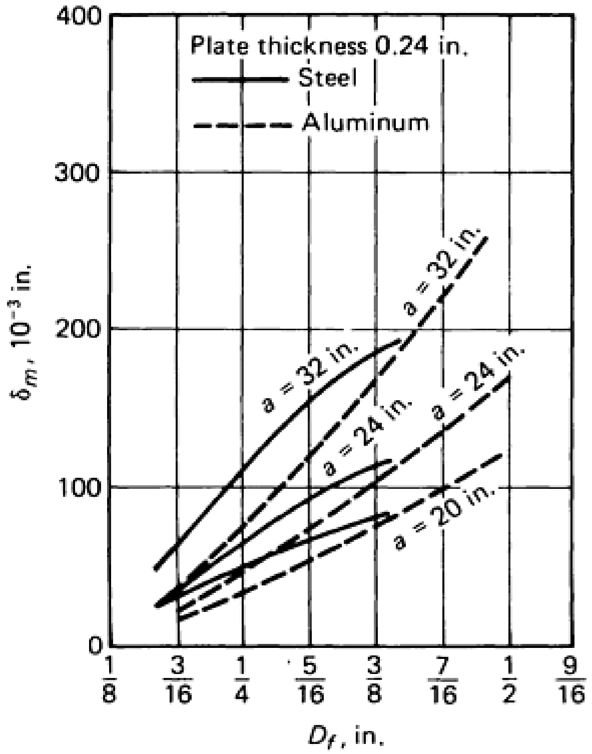
Plate thickness 0.24 in.
Steel
Aluminum
a = 32 in.
a = 24 in.
a = 32 in.
a = 24 in.
a = 20 in.
400
300
200
100
0
δ_m, 10^{-3} in.
$\frac{1}{8}$ $\frac{3}{16}$ $\frac{1}{4}$ $\frac{5}{16}$ $\frac{3}{8}$ $\frac{7}{16}$ $\frac{1}{2}$ $\frac{9}{16}$
D_f, in.

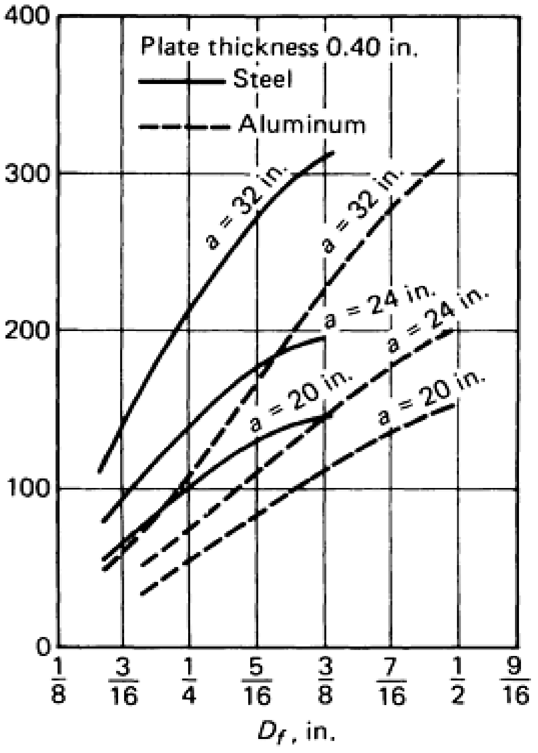
Plate thickness 0.40 in.
Steel
Aluminum
a = 32 in.
a = 32 in.
a = 24 in.
a = 24 in.
a = 20 in.
a = 20 in.
400
300
200
100
0
$\frac{1}{8}$ $\frac{3}{16}$ $\frac{1}{4}$ $\frac{5}{16}$ $\frac{3}{8}$ $\frac{7}{16}$ $\frac{1}{2}$ $\frac{9}{16}$
D_f, in.

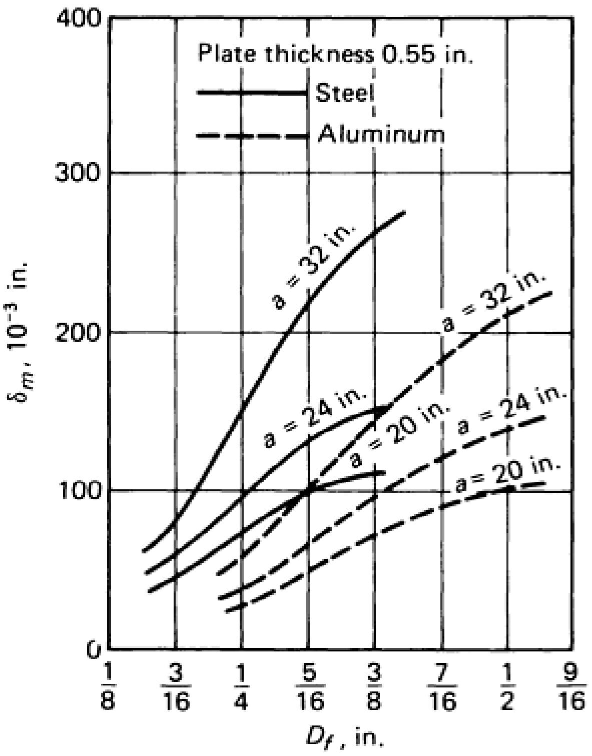
Plate thickness 0.55 in.
Steel
Aluminum
a = 32 in.
a = 32 in.
a = 24 in.
a = 20 in.
a = 24 in.
a = 20 in.
400
300
200
100
0
δ_m, 10^{-3} in.
$\frac{1}{8}$ $\frac{3}{16}$ $\frac{1}{4}$ $\frac{5}{16}$ $\frac{3}{8}$ $\frac{7}{16}$ $\frac{1}{2}$ $\frac{9}{16}$
D_f, in.

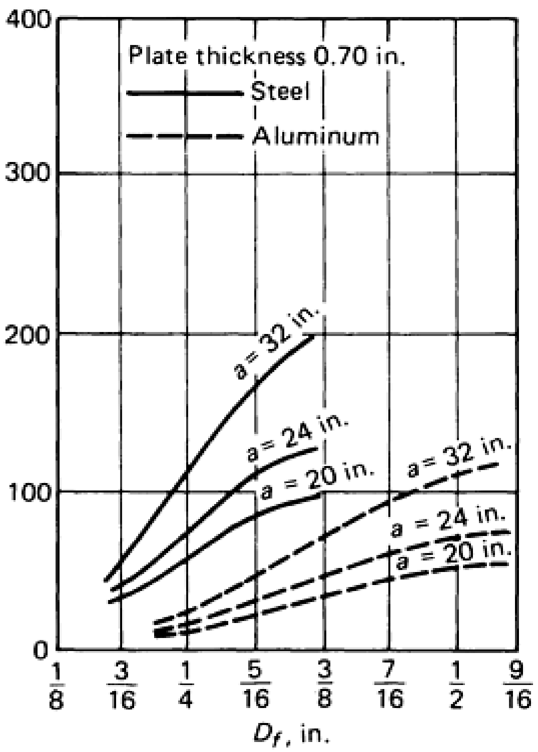
Plate thickness 0.70 in.
Steel
Aluminum
a = 32 in.
a = 24 in.
a = 20 in.
a = 32 in.
a = 24 in.
a = 20 in.
400
300
200
100
0
$\frac{1}{8}$ $\frac{3}{16}$ $\frac{1}{4}$ $\frac{5}{16}$ $\frac{3}{8}$ $\frac{7}{16}$ $\frac{1}{2}$ $\frac{9}{16}$
D_f, in.

FIG. 8 OUT-OF-PLANE DISTORTION (δ_M) AS A FUNCTION OF PLATE THICKNESS (*T*), SPAN LENGTH (*L*), AND THE SIZE OF THE FILLET WELD (D_F) FOR STEEL AND ALUMINUM. RELATION OF *L* TO *A* IN GRAPHS:*A* = *L*

Longitudinal Distortion of Built-up Beams. Figure 9 compares values of longitudinal distortion expressed in terms of the radius of curvature of built-up beams in steel and aluminum. In this instance, aluminum welds distorted less than steel welds, perhaps because the temperature distribution in the *z*-direction is more uniform in an aluminum weld than in a steel weld.

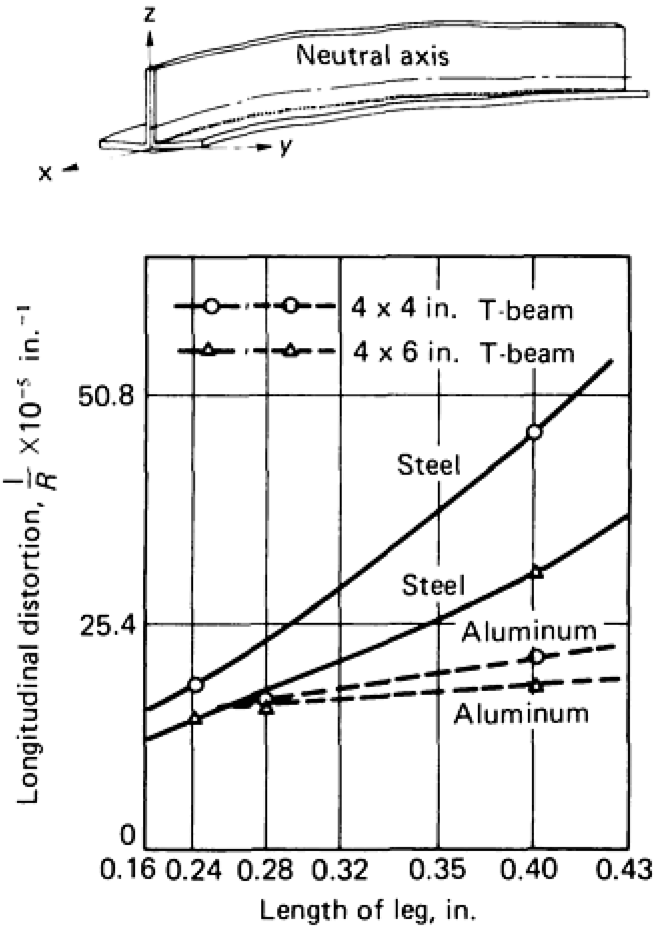

FIG. 9 RELATIONSHIP BETWEEN LENGTH OF LEG AND CURVATURE OF LONGITUDINAL DEFLECTION IN T-SECTION BEAM. SOURCE: REF 1, 16

Buckling Distortion. In any weldment, longitudinal residual stresses are compressive in areas away from the weld, as shown in Fig. 3. When the plate is thin, it may buckle due to these compressive residual stresses. Buckling distortion differs from bending distortion in that the amount of distortion in buckling is much greater. Also, there may be more than one stable deformed shape in buckling distortion. The most effective method for dealing with buckling distortion is prevention by proper selection of plate thickness, length of free span, and welding heat input.

When stiffeners are fillet welded to thin plate, the plate may buckle because of the compressive residual stresses that occur in the plate. Figure 10 shows relationships between distortion at the center of the panel and heat input for square panels 500 mm (20 in.) long made with plates 4.6 to 10 mm (0.18 to 0.4 in.) thick. For example, in the case of a 500 × 500 mm (20 × 20 in.) panel with a plate 6 mm (0.24 in.) thick, the plate deflection increases suddenly when the heat input exceeds approximately 3700 cal/cm^2. This indicates that the critical heat input is approximately 3700 cal/cm^2. When the plate is 4.6 mm (0.18 in.) thick, this critical heat input is only approximately 2000 cal/cm^2. The following formula can be used to determine the critical heat input, H_{cr}, above which excessive buckling distortion occurs:

$$H_{cr} = \frac{Q}{t^3}\ b \geq \sim 4\text{x}10^5 cal/cm^3$$ **(EQ 4)**

where *b* is the length of the panel and *t* is the plate thickness, both in millimeters. The above parameter, H_{cr}, is also called the critical heat input index.

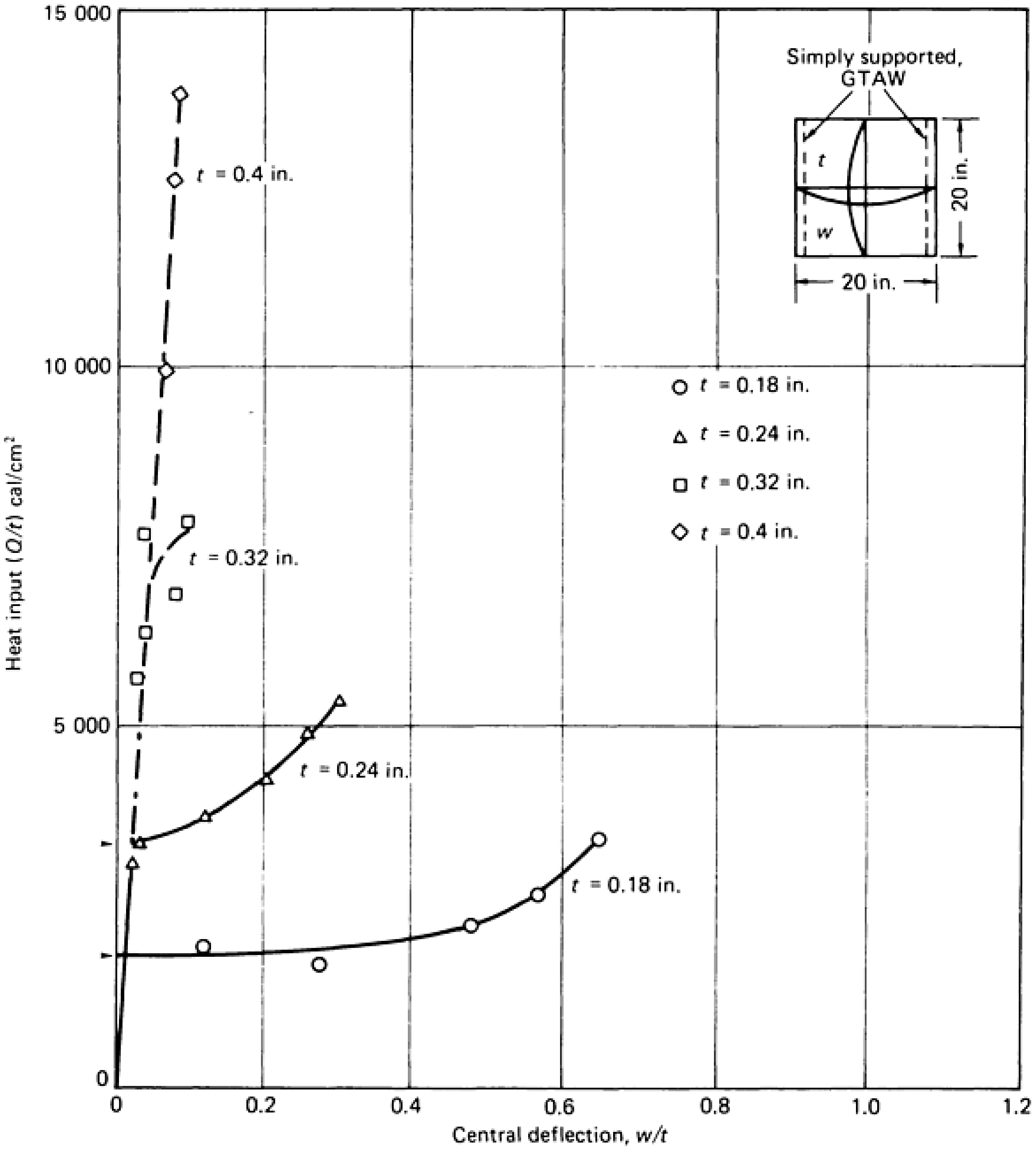

FIG. 10 PLOT OF HEAT INPUT VERSUS DEFLECTION FOR 500 × 500 MM (20 × 20 IN.) LOW-CARBON STEEL PANELS AS A FUNCTION OF PANEL THICKNESS

References cited in this section

1. K. MASUBUCHI, *ANALYSIS OF WELDED STRUCTURES--RESIDUAL STRESSES, DISTORTION, AND THEIR CONSEQUENCES,* PERGAMON PRESS, 1980
2. K. MASUBUCHI, RESIDUAL STRESSES AND DISTORTION, *METALS HANDBOOK,* 9TH ED., VOL

6, *WELDING, BRAZING, AND SOLDERING,* ASM, 1983, P 856-895

16. G. YAMAMOTO, "STUDY OF LONGITUDINAL DISTORTION OF WELDED BEAMS," M.S. THESIS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, MAY 1975

Residual Stresses and Distortion

Koichi Masubuchi, Massachusetts Institute of Technology

Thermal Treatments of Weldments

Thermal treatments are often necessary to maintain or restore the properties of base metal affected by the heat of welding. Thermal treatment may also affect the properties of the weld metal. The extent of changes in the properties of the base metal, weld metal, and heat-affected zone (HAZ) are determined by several factors, including the soaking temperature, time, cooling rate, and material thickness. Further detailed information is available in Ref 2 and 3.

Preheat is the most common thermal treatment applied to weldments. Proper use of preheat can minimize residual stresses and distortion that would normally occur during welding as a result of lower thermal gradients around the weld. Preheat also has the beneficial effect in steels of reducing the tendency for the formation of a HAZ and weld metal cracking.

Postweld Thermal Treatments. A properly executed postweld heat treatment results in uniform mechanical properties and reduced residual stresses. The effects of time at temperature and the stress-relieving temperature on residual stresses are shown in Fig. 14(a) and 14(b), respectively (Ref 3). When thick weldments require a postweld machining operation, a stress-relief treatment is usually necessary to achieve normal machining tolerance.

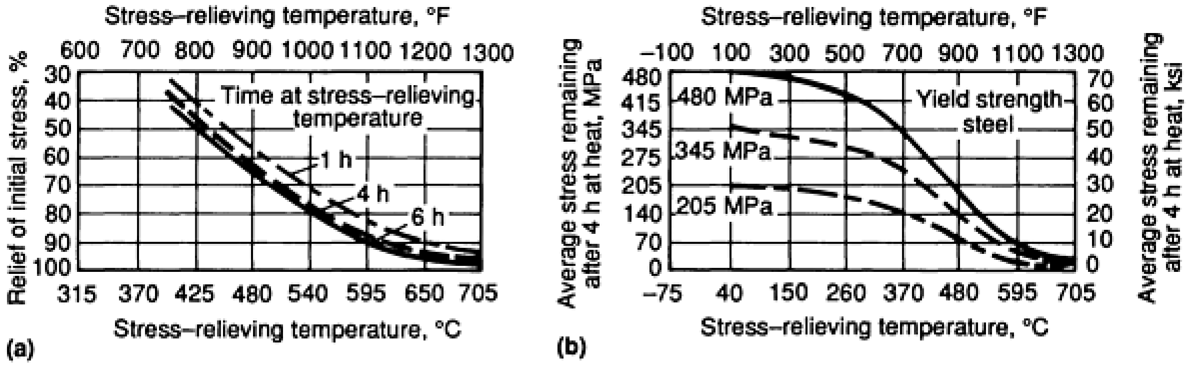

FIG. 14 EFFECT OF TEMPERATURE AND TIME ON STRESS RELIEF. TIME AND TEMPERATURE, 4 H

References cited in this section

2. K. MASUBUCHI, RESIDUAL STRESSES AND DISTORTION, *METALS HANDBOOK,* 9TH ED., VOL 6, *WELDING, BRAZING, AND SOLDERING,* ASM, 1983, P 856-895
3. K. MASUBUCHI, O.W. BLODGETT, S. MATSUI, E.P. ROSS, AND C.L. TSAI, RESIDUAL STRESSES AND DISTORTION, *WELDING HANDBOOK,* 8TH ED., VOL 1, *WELDING TECHNOLOGY,* AWS, 1987

Residual Stresses and Distortion

Koichi Masubuchi, Massachusetts Institute of Technology

References

1. K. MASUBUCHI, *ANALYSIS OF WELDED STRUCTURES--RESIDUAL STRESSES, DISTORTION, AND THEIR CONSEQUENCES,* PERGAMON PRESS, 1980
2. K. MASUBUCHI, RESIDUAL STRESSES AND DISTORTION, *METALS HANDBOOK,* 9TH ED., VOL 6, *WELDING, BRAZING, AND SOLDERING,* ASM, 1983, P 856-895
3. K. MASUBUCHI, O.W. BLODGETT, S. MATSUI, E.P. ROSS, AND C.L. TSAI, RESIDUAL STRESSES AND DISTORTION, *WELDING HANDBOOK,* 8TH ED., VOL 1, *WELDING TECHNOLOGY,* AWS, 1987
4. K. MASUBUCHI, RESEARCH ACTIVITIES EXAMINE RESIDUAL STRESSES AND DISTORTION IN WELDED STRUCTURES, *WELD. J.,* VOL 70 (NO. 12), 1991, P 41-47
5. L. TALL, RESIDUAL STRESSES IN WELDED PLATES--A THEORETICAL STUDY, *WELD. J.,* VOL 43 (NO. 1), 1991, P 10S-23S
6. H.D. HIBBIT AND P.V. MARCEL, "A NUMERICAL THERMOMECHANICAL MODEL FOR THE WELDING AND SUBSEQUENT LOADING OF A FABRICATED STRUCTURE," NSRDC CONTRACT NO. N00014-67-A-019-0006, TECH. REP. NO. 2, DEPARTMENT OF THE NAVY, MARCH 1972
7. T. MURAKI, J.J. BRYAN, AND K. MASUBUCHI, ANALYSIS OF THERMAL STRESSES AND METAL MOVEMENT DURING WELDING--PART I: ANALYTICAL STUDY; PART II: COMPARISON OF EXPERIMENTAL DATA AND ANALYTICAL RESULTS, *J. ENG. MATER. TECHNOL. (TRANS. ASME),* 1975, P 81-84, 85-91
8. E.F. RYBICKI *ET AL.,* A FINITE-ELEMENT MODEL FOR RESIDUAL STRESSES AND DEFLECTION IN GIRTH-BUTT WELDED PIPES, *J. PRESSURE VESSEL TECHNOL. (TRANS. ASME),* VOL 100, 1978, P 256-262
9. P.O. DEXTER AND D. PONT, "EVALUATION OF AVAILABLE WELDING SIMULATION SOFTWARE," PREPARED FOR EPRI UNDER RESEARCH PROJECT C102-3 BY SOUTHWEST RESEARCH INSTITUTE, MAY 1991
10. R.G. TREUTING, J.J. LYNCH, H.B. WISHART, AND D.G. RICHARDS, *RESIDUAL STRESS MEASUREMENTS,* ASM, 1952
11. K. MASUBUCHI, "NONDESTRUCTIVE MEASUREMENT OF RESIDUAL STRESSES IN METALS AND METAL STRUCTURES," RSIC-410, REDSTONE SCIENTIFIC INFORMATION CENTER, REDSTONE ARSENAL, 1965
12. K. MASUBUCHI, RESIDUAL STRESSES AND DISTORTION, *WELDING HANDBOOK,* VOL 1, 7TH ED., AWS, 1976
13. R.R. HOSBONS, THE USE OF NEUTRON DIFFRACTION TO DETERMINE NONDESTRUCTIVELY THE RESIDUAL STRAIN AND TEXTURE IN WELDS, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* ASM INTERNATIONAL, 1989, P 103-106
14. H. KIHARA, M. WATANABE, K. MASUBUCHI, AND K. SATOH, *RESEARCHES ON WELDING STRESSES AND SHRINKAGE DISTORTION IN JAPAN,* VOL 4, 60TH ANNIVERSARY SERIES OF THE SOCIETY OF NAVAL ARCHITECTS OF JAPAN, TOKYO, 1959
15. N.R. NAGARAJA RAO, F.R. ESATUAR, AND K. TALL, RESIDUAL STRESSES IN WELDED SHAPES, *WELD. J.,* VOL 39 (NO. 3), 1964, P 295-306S
16. G. YAMAMOTO, "STUDY OF LONGITUDINAL DISTORTION OF WELDED BEAMS," M.S. THESIS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, MAY 1975
17. H. KIHARA AND K. MASUBUCHI, EFFECT OF RESIDUAL STRESS ON BRITTLE FRACTURE, *WELD. J.,* VOL 38 (NO. 4), 1959, P 159S-168S
18. W.J. HALL, H. KIHARA, W. SOETE, AND A.A. WELLS, *BRITTLE FRACTURE OF WELDED PLATES,* PRENTICE-HALL, 1967
19. H. KIHARA, K. MASUBUCHI, K. IIDA, AND H. OHBA, "EFFECT OF STRESS RELIEVING ON BRITTLE FRACTURE STRENGTH OF WELDED STEEL PLATE," DOCUMENT X-218-59,

INTERNATIONAL INSTITUTE OF WELDING, LONDON, 1959 (DISTRIBUTED IN U.S. BY AWS)
20. W.H. MUNSE, *FATIGUE OF WELDED STEEL STRUCTURES,* WELDING RESEARCH COUNCIL, 1964
21. T.R. GURNEY, *FATIGUE OF WELDED STRUCTURES,* 2ND ED., CAMBRIDGE UNIVERSITY PRESS, LONDON, 1979
22. *RESIDUAL STRESS EFFECTS IN FATIGUE,* STP 776, ASTM, 1982
23. A.S. TETELMAN AND A.J. MCEVILY, JR., *FRACTURE OF STRUCTURAL MATERIALS,* JOHN WILEY & SONS, 1967

Residual Stresses and Distortion

Koichi Masubuchi, Massachusetts Institute of Technology

Selected References

- O.W. BLODGETT, TYPES AND CAUSES OF DISTORTION IN WELDED STEEL AND CORRECTIVE MEASURES, *WELD. J.,* JULY 1960
- W.R. OSGOOD, *RESIDUAL STRESSES IN METALS AND METAL CONSTRUCTION,* REINHOLD PUBLISHING, 1954

Repair Welding

Vern Sutter, American Welding Institute; Robert J. Dybas, General Electric Company

---

Introduction

REPAIR AND MAINTENANCE of parts and components is a multibillion dollar industry. Repair welding can be carried out as a logical procedure that ensures the production of a usable and safe component or it can be approached haphazardly. The latter approach results in poor-quality workmanship and can lead to failed parts, large warranty claims, and dissatisfied customers.

It is to the advantage of the individual welder, job shop owner, end user, and others who depend on the weld repair industry to approach each repair with a thorough knowledge of component history in terms of:

- COMPONENT FUNCTION
- MATERIAL COMPOSITION
- COMPONENT SURFACE OR THROUGH HARDNESS
- PART ORIGINALLY CAST, FABRICATED, OR WELDED

Repair welding can fall into one of three general categories: repair of weld defects, repair of failed parts, and repair of worn parts. This article describes the repair of weld defects and structural failures. Welding buildup of worn parts is described in the article "Cladding, Hardfacing, and Dissimilar Metal Joining" in this Volume.

Requirements and repair techniques discussed in this article apply to arc and oxyfuel welding processes. Materials can be any that are usually joined successfully using arc or oxyfuel welding processes. Generally, the repair welding procedures will apply whether the structure is built under code construction specifications or not. Many construction codes require that a written procedure be prepared prior to any repair welding. Code or contractual requirements supersede any statements made in this article.

If the component to be repaired was originally a welded fabrication, then data and information on the original process are important to a successful repair. If access to this information is not practical, then an analysis of the base material, including previous weld deposits, becomes mandatory. If dimensions require close tolerances or if flatness is critical, then benchmarks that will aid the repair without causing excessive and expensive damage to the workpiece must be established.

Understanding the basics of each group of metals covered in this article helps to ensure a successful repair. All references and sources of information should be exhausted before any job commences. Excellent sources include filler-metal manufacturers and manufacturers of bar, plate, forgings, and castings. Either of these producers can supply key information on the weldability of the workpiece alloys.

Depending on the specific application, all of the common welding processes can be used for repair welding:

- SHIELDED METAL ARC WELDING (SMAW)
- GAS-METAL ARC WELDING (GMAW)
- GAS-TUNGSTEN ARC WELDING (GTAW)
- SUBMERGED ARC WELDING (SAW)
- PLASMA ARC WELDING (PAW)

For the highest-quality welds, the GTAW and PAW processes find the widest application. For long runs or when a large amount of weld metal must be deposited and mechanization is feasible, the SAW process or, to a lesser extent, the GMAW process, is utilized. For general repairs, the SMAW process still enjoys the widest range of applications for out-of-position welding and for short runs, especially when time is critical and when readily portable equipment is utilized. Electrodes are easily transported in sealed 4.5 kg (10 lb) containers.

In addition to the SMAW process, the GMAW process has been gaining wider support and utilization since the introduction of the flux-cored arc welding (FCAW) process, a core wire variation of solid-wire welding. The cored wires are manufactured in smaller diameters to make the wires suitable for out-of-position welding and, especially, for many field and construction applications.

Repair Welding

Vern Sutter, American Welding Institute; Robert J. Dybas, General Electric Company

Preliminary Assessment

Before attempting a repair, three factors must first be considered: material weldability, nature of the failure that prompted the repair, and involvement of any code requirements.

Base Metal Weldability

If the item to be repaired was not welded previously, then a special investigation may be necessary to determine weldability. Original documents and drawings are helpful in determining the specifications or description of the base metal. Some tests that can determine degree of weldability include the spark test, chemical analysis test, and simulated weld tests.

The spark test determines the approximate base metal chemistry for carbon steels. Full chemical analysis using drillings can accurately determine chemistry, but not necessarily the heat treatment history of the base metal. Simulated weld tests can determine some practical approaches to repair welding, but seldom give any absolute requirements. These tests require samples of some of the base metal from the workpiece to be repaired.

Nature of Failure

When designing and producing a successful repair, it is often necessary to assess the reasons that a repair is required. Failures can be grouped into four major categories:

- BASE-METAL DEFECTS
- BASE-METAL STRENGTH DEFICIENCIES
- FAILURES INTRODUCED DURING FABRICATION
- DEFECTS CAUSED BY ABRASION, CAVITATION, CORROSION, OR EROSION

Base-Metal Defects. During or subsequent to fabrication, defects in the base metal may be detected. Defects can include pits, stringers, slivers, and a variety of internal discontinuities, the latter of which may be discovered when sawing through the base metal or subjecting the base metal to nondestructive testing (NDT).

Base-Metal Strength Deficiencies. Weld repair alone may not adequately repair a structure that has been overloaded. Reduction of load and additional reinforcing may be necessary. The loading and member arrangement should be thoroughly investigated. Repair of a crack or other structural discontinuity may involve repairing the crack as a weld repair and adding reinforcing plates or members to the workpiece by welding.

Weld Failure if Fabricated. Repair may be required on weld joints that have failed structurally or have undergone nondestructive evaluation (NDE), but do not meet soundness criteria. In these cases, it is usually prudent to prepare the joint for rewelding with procedures similar to the original welding operations.

Defects Caused by Abrasion, Cavitation, Corrosion, and Erosion. Buildup of the workpiece to repair surface damage that is due to abrasion, cavitation, corrosion, or erosion can be carried out as a weld repair. These conditions may or may not be associated with an original weld. The same weldability problems associated with structural repairs are encountered with the defects induced by these conditions. The article "Cladding, Hardfacing, and Dissimilar Metal Joining" in this Volume addresses the conditions that produce these defects in much more detail.

Codes and Standards

Repairs may be governed by welding or construction standards. Some codes address repair welding requirements as a specific subject and others simply require welding qualification that would apply to both new and repair welding. Organizations such as the American Society of Mechanical Engineers (ASME), American Welding Society (AWS), American Petroleum Institute (API), and military standards (MIL-STD) publish their own codes. Owner or engineer approval of repair welding procedures may be required by contract specifications.

Welding Process Selection

When selecting a welding process for a specific application, several factors that affect productivity and weld quality must be balanced. Selection can be difficult because each process has a number of conflicting advantages and limitations in specific situations.

Productivity is usually not an important consideration on most repair jobs. Each process can be ranked in terms of its deposition rate in kilograms (pounds) of weld metal deposited per hour. However, there are other factors that must be considered, as a minimum, before selecting a welding process:

- BASE-METAL TYPE
- JOINT DESIGN AND THICKNESS
- WELDING POSITION
- ENVIRONMENTAL CONDITIONS

- EQUIPMENT AVAILABILITY

Base-Metal Type. Some processes are better suited for use with certain base materials. The maximum heat input must be limited on certain types of materials, such as heat-treated materials that have been quenched, tempered, and age hardened, as well as other heat-treated or cold-worked materials. Processes that derive their advantage from high productivity and high heat input may be unsuitable for these applications.

Joint Design and Thickness. As section thickness increases, welding productivity becomes more important, which should be reflected in the process selection. The length of the weld must also be considered, because a higher-productivity process may not realize this advantage, particularly if the operator must frequently stop the process in order to set up the next pass, as would be necessary with a small repair. Because some processes require more access to the joint root to avoid fusion defects, the selection of some processes may also require a joint design change.

Welding Position. The weld joint position is very important in terms of process selection, because many processes are limited to only a few positions. Whenever possible, the joint should be in the flat (1G) position, because the highest productivity and weld quality are attained when welding is accomplished in this position. Because most field repair work is done on large weldments that cannot be repositioned and because access to the joint is limited, the use of high-productivity processes and filler materials is also limited.

Environmental Conditions. Wind and rain are the two conditions that typically affect field welding. It takes very little wind to disturb the gas shield that is critical for high-quality GMAW and GTAW processes. This restricts their use outside of sheltered containments. The SMAW and FCAW processes can also be affected by wind, but to a lesser degree. Because no process ever tolerates direct exposure to rain, the proper placement of tarps, dams, or other temporary containments is in order.

Availability of Equipment. Most large welding shops have access to the welding equipment required for the processes discussed. However, there are occasions when new equipment must be evaluated to determine if its increased productivity or versatility would offset the capital expense of its initial cost and training.

A general comparison of the most commonly used processes relative to some of the factors discussed above is provided in Table 1. Although some of the ratings are very subjective, they do provide a general guideline for many applications.

TABLE 1 RATING OF SELECTED WELDING PROCESSES AS A FUNCTION OF WELD PARAMETERS AND CHARACTERISTICS

PARAMETER OR CHARACTERISTIC	PROCESS				
	SMAW	**GTAW**	**GMAW**	**FCAW**	SAW
WELD QUALITY	GOOD	EXCELLENT	EXCELLENT	GOOD	**EXCELLENT**
WELD DEPOSITION RATE	FAIR	POOR	GOOD	GOOD	**EXCELLENT**
FIELD WORK	EXCELLENT	POOR	FAIR	EXCELLENT	**POOR**
EQUIPMENT MAINTENANCE	LOW	LOW	MEDIUM	MEDIUM	**MEDIUM**
SMOKE/FUME EMISSION	HIGH	LOW	MEDIUM	HIGH	**VERY LOW**
HEAT INPUT CONTROL	EXCELLENT	POOR	GOOD	GOOD	**SATISFACTORY**
ARC VISIBILITY AND FILLER-METAL PLACEMENT	GOOD	EXCELLENT	SATISFACTORY	SATISFACTORY	**POOR**
VARIETY OF METALS WELDABLE	**EXCELLENT**	**EXCELLENT**	**GOOD**	**GOOD**	**FAIR**

Repair Welding

Vern Sutter, American Welding Institute; Robert J. Dybas, General Electric Company

---

Base-Metal Preparation

Methods of preparing base metal for repair welding may vary, depending on the specific metal to be welded. Generally, all coatings in the vicinity of the repair weld should be removed. Coatings can cause defective welds and can become surface contaminants when heated to the temperatures required for welding.

Where cracks are to be excavated and the part rewelded, a welding groove should be prepared. Most base metals can be cut or grooved using air-carbon arc cutting (CAC-A). Table 2 shows some parameters for carbon arc cutting, whereas Fig. 1 shows a typical joint preparation using the CAC technique. It should be noted that the groove must be prepared in such a way as to allow adequate access to the root of the joint. This preparation is sometimes described as "canoe-shaped." Additional information on this cutting process is provided in the article "Air-Carbon Arc Cutting" in this Volume.

TABLE 2 RECOMMENDED CURRENT RANGE FOR SELECTED SIZES OF AC AND DC ELECTRODES USED FOR CARBON ARC GOUGING AND CUTTING IN AIR

ELECTRODE DIAMETER		CURRENT, A
MM	**IN.**	
DIRECT-CURRENT ELECTRODES (DCEP POLARITY)		
3.2	$\frac{1}{8}$	**30-60**
4.0	$\frac{5}{32}$	**90-150**
4.8	$\frac{3}{16}$	**200-250**
6.4	$\frac{1}{4}$	**300-400**
7.9	$\frac{5}{16}$	**350-450**
9.5	$\frac{3}{8}$	**450-600**
$9.5^{(A)}$	$\frac{3}{8}$ (A)	**250-450**
13	$\frac{1}{2}$	**800-1000**
16	$\frac{5}{8}$	**1000-1250**
$16^{(A)}$	$\frac{5}{8}$ (A)	**300-500**
19	$\frac{3}{4}$	**1250-1600**
25	1	**1600-2200**
ALTERNATING-CURRENT ELECTRODES		
4.8	$\frac{3}{16}$	**200-250**
6.4	$\frac{1}{4}$	**300-400**
7.9	$\frac{5}{16}$	**325-425**
9.5	$\frac{3}{8}$	**350-450**
13	$\frac{1}{2}$	**500-600**

(A) FLAT

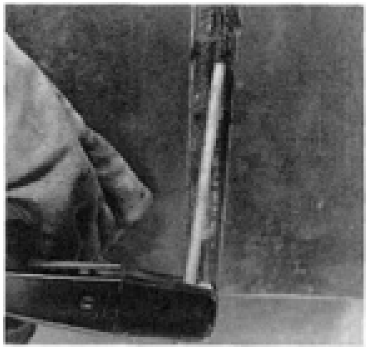

FIG. 1 USE OF AIR-CARBON ARC CUTTING TO PRODUCE A GROOVE IN A PLATE. ELECTRODE HOLDER IS EQUIPPED WITH A JET OF COMPRESSED AIR IN LINE WITH THE ELECTRODE TO BLOW AWAY THE MOLTEN METAL.

Grooves can also be prepared by grinding. However, grinding particles left on a groove surface can cause problems, especially in nonferrous materials. The proper grinding wheels must be carefully selected. Carbon arc cutting usually leaves some undesirable by-product on the cut surfaces. A grinding operation following CAC is necessary to remove this layer.

Repair Welding

Vern Sutter, American Welding Institute; Robert J. Dybas, General Electric Company

---

Guidelines for Various Base Metals

Each group of base metals has specific general repair guidelines that must be followed to ensure successful weld repairs of both ferrous (carbon steels, cast irons, and stainless steels) and nonferrous (for example, titanium) base metals.

Low-Carbon Steel

The most commonly welded group of materials is that which has a carbon-manganese-iron base and is known as mild, or low-carbon, steel. They constitute the bulk of construction materials used today and are the common choice for equipment and fabricated component assemblies, because of their weldability.

Mild steels (up to 0.25% C) are readily weldable and, in cross sections of less than 50 mm (2 in.), usually do not require any preheat or postweld heat treatment. Hardened heat-affected zones (HAZ) are not a problem, and cracking is nonexistent when electrodes are used from fresh containers or electrode storage ovens and are dry. American Welding Society specification A5.1, "Carbon Steel Electrodes for Shielded Metal Arc Welding," covers the filler metals of choice for greater than 50% of all welding fabrication and repair work performed in the world. With the greater demand for higher-production deposition rates, continuous welding wire has grown in usage and application areas.

AWS specifications A5.17 ("Carbon Steel Electrodes and Fluxes for Submerged Arc Welding"), A5.18 ("Carbon Steel Filler Metals for Gas-Shielded Arc Welding"), and A5.20 ("Carbon Steel Electrodes for Flux-Cored Arc Welding") have gained wide support for automatic, semiautomatic, and mechanized welding processes.

Electrodes that are not suitably protected nor kept dry in an electrode storage oven can absorb moisture, which is released as hydrogen gas during the welding process. This hydrogen gas is entrapped in the solidifying weld metal, resulting in immediate or delayed cracking and expensive repairs. When abrasion, friction, or sliding wear is the reason for equipment downtime, either AWS specification A5.13 ("Solid Surfacing Welding Rods and Electrodes") or A5.21 ("Composite Surfacing Welding Rods and Electrodes") is the first choice in the selection process to deposit and build up a resistant surface capable of withstanding the original wear phenomena in the most severe working environment. In addition, there are numerous electrodes of proprietary composition on the market that perform satisfactorily in service. These electrodes require evaluation and testing before being used.

Base-metal preparation usually consists of removing paint, mill scale, rust, or grease, because they contribute to porous welds. When making field repairs, especially in high-humidity areas, it is always a good practice to apply a preheat in the range from 120 to 150 °C (250 to 300 °F). Postweld heat treatment, though normally not required on mild steels, is sometimes used to ensure dimensional stability where close, flat tolerances are required.

Medium-Carbon Steels

Steels containing from 0.25 to 0.60% C and 0.25 to 1.65% Mn are classified as medium-carbon steels. Based on their base-metal mechanical properties, the electrodes in AWS specification A5.1 usually suffice in most applications.

The use of preheat, interpass temperature, and postweld heat treatment is always required to ensure fusion of the base metal and the weld metal, to prevent distortion, and to ensure dimensional stability of the finished product. Depending on chemical composition, the preheat temperature ranges from 120 to 205 °C (250 to 400 °F). Postweld heat treatment for this group of materials ranges from 595 to 620 °C ± 14 °C (1100 to 1150 °F ± 25 °F). The AWS electrode specifications listed for mild steels are suitable for repairing medium-carbon steels.

High-Carbon Steels

High-carbon steels have a carbon content ranging from 0.60 to 1.00% C. These metals require preheats in the range from 205 to 260 °C (400 to 500 °F), interpass temperatures from 205 to 290 °C (400 to 550 °F), and postweld heat treatments from 620 to 650 °C (1150 to 1200 °F).

Because of the higher carbon content of these steels, greater care must be exercised when welding and during subsequent heat treatment. The HAZ has a higher hardness and cooling must be controlled to preclude any chilling effects that could result in fusion line cracking.

For steels in the annealed or normalized condition, the welding electrodes listed for mild steels will suffice, based on mechanical properties. In the heat-treated condition (that is, higher strength levels), the repair may require stronger filler metal as provided by AWS specifications A5.5 ("Low Alloy Steel Covered Arc Welding Electrodes"), A5.23 (Low Alloy Steel Electrodes and Fluxes for Submerged Arc Welding"), A5.28 ("Low Alloy Steel Filler Metals for Gas-Shielded Arc Welding"), and A5.29 ("Low Alloy Steel Electrodes for Flux Cored Arc Welding").

Cast Irons

Cast irons are a family of high-carbon iron-base alloys containing from 2 to 4% C, 0.50 to 2% Si, and varying amounts of added manganese and chromium strengtheners, in addition to other elements that control graphite formation.

The most common members of this family are the gray cast irons (flake graphite) described in ASTM specification A 48. They are primarily carbon-manganese-silicon in content and no other elements are deliberately added. These materials have tensile strengths ranging from 140 to 415 MPa (20 to 60 ksi) with nil yield ductility and elongations less than 0.5%. Because of this characteristic, they are most difficult to weld and usually suffer from fusion line cracking, because of the presence of carbides formed by the flake graphite. If the broken area (cracked zone) can be placed in compression, then the chance of a successful repair is significantly improved.

Ductile cast irons, clearly identified by the presence of nodular graphite in their microstructure, are sometimes called nodular irons or spheroidal-graphite cast irons. ASTM A 536 grades 60-40-18 through 120-90-02 (designations indicate tensile strength in units of ksi, yield strength in ksi, and elongation in percent, respectively) have better ductility, which makes them more readily weldable by the higher-strength filler metals.

Malleable irons, described in specification ASTM A 47 (ferritic malleable iron castings), A 220 (pearlitic malleable irons), and A 602 (automotive malleable iron castings), are generally weldable and do not suffer the same problems that gray irons have.

The welding processes suitable for repair work are identified in AWS specification D11.2 ("Guide for Welding Iron Castings"):

- SHIELDED METAL ARC WELDING
- GAS-METAL ARC WELDING
- FLUX-CORED ARC WELDING
- GAS-TUNGSTEN ARC WELDING
- SUBMERGED ARC WELDING

The gas processes used for the repair of iron castings are oxyfuel welding (OFW) and braze welding, which utilizes specific filler metals of AWS specification A5.8 ("Filler Metals for Brazing"). The group of welding electrodes used for the arc welding of cast irons is described in AWS specification A5.15 ("Welding Electrodes and Rods for Cast Iron").

Stainless Steels

Stainless steel is a generic term encompassing three distinct families of iron-base alloys:

- AUSTENITIC STAINLESS STEELS
- FERRITIC STAINLESS STEELS
- MARTENSITIC STAINLESS STEELS

The rules for the repair welding of stainless steels will vary with the composition and physical characteristics of the material.

Austenitic stainless steels can be welded by the SMAW process, using electrodes listed in AWS specification A5.4 ("Stainless Steel Electrodes for Shielded Metal Arc Welding"). The GMAW, GTAW, FCAW, SAW, and PAW processes use electrodes listed in AWS specification A5.9 ("Bare Stainless Steel Welding Electrodes and Welding Rods"). SAW process fluxes of proprietary composition are available from manufacturers.

Austenitic stainless steels are usually not preheated and are readily weldable, but they must be clean and free from contamination to ensure weld-metal quality. The free-machining sulfurized grades and those containing selenium or phosphorus are generally not weldable. The titanium or niobium-plus-tantalum stabilized grades of types 318, 321, 347, and 348 have altered chemistries to prevent intergranular precipitation upon heating to a temperature-sensitizing range, such as occurs in welding. The selection of a suitable electrode is critical, and the steel manufacturer or electrode manufacturer should be consulted for the latest technology. Because of their lower thermal conductivity and higher electrical resistivity, stainless steel welds require from 20 to 30% less heat input to prevent weld cracking, loss of corrosion resistance, warping, and undesirable changes in mechanical properties.

The GMAW and GTAW processes use argon gas for a cool, stable arc and helium for a hotter, deeper-penetrating, but somewhat unstable arc. Table 3 lists various combinations of base metal and filler metal that are suitable for a variety of service conditions. This table is an excellent starting point for any repair.

TABLE 3 ELECTRODES OR WELDING RODS AS FILLER METALS FOR USE IN ARC WELDING OF STAINLESS STEELS

TYPE OF STEEL WELDED	**CONDITION OF WELDMENT FOR SERVICE**(A)	ELECTRODE OR WELDING ROD(B)
AUSTENITIC STEELS		
301, 302, 304, 305, 308(C)	1 OR 2	**308**
302B(D)	1	**309**
304L	1 OR 4	**347, 308L**
303, 303SE(E)	1 OR 2	**312**

309, 309S	1	**309**
310, 310S	1	**310**
316(F)	1 OR 2	**316**
316L(F)	1 OR 4	**318, 316L**
317(F)	1 OR 2	**317**
317L(F)	1 OR 4	**317CB**
318, 316CB(F)	1 OR 5	**318**
321(G)	1 OR 5	**347**
347(H)	1 OR 5	**347**
348(I)	1 OR 5	**347**
MARTENSITIC STEELS		
403, 410, 416, 416SE(J)	2 OR 3	**410**
403, 410(K)	1	**308, 309, 310**
416, 416SE(K)	1	**308, 309, 312**
420(L)	2 OR 3	**420**
431(L)	2 OR 3	**410**
431(M)	1	**308, 309, 310**
FERRITIC STEELS		
405(N)	2	**405CB, 430**
405,430(K)	1	**308, 309, 310**
430F, 430FSE(K)	1	**308, 309, 312**
430, 430F, 430FSE(O)	2	**430**
446	2	**446**
446(P)	**1**	**308, 309, 310**

Source: Ref 1

(A) 1, AS WELDED; 2, ANNEALED; 3, HARDENED AND STRESS RELIEVED; 4, STRESS RELIEVED; 5, STABILIZED AND STRESS RELIEVED.

(B) PREFIX E OR ER OMITTED.

(C) TYPE 308 WELD METAL IS ALSO REFERRED TO AS 18-8 AND 19-9 COMPOSITION. ACTUAL WELD ANALYSIS REQUIREMENTS ARE: 0.08% C MAX, 19.0% CR MIN, AND 9.0% NI MIN.

(D) TYPE 310 (1.50% SI MAX) MAY BE USED AS FILLER METAL, BUT THE PICKUP OF SILICON FROM THE BASE METAL MAY RESULT IN WELD HOT CRACKING.

(E) FREE-MACHINING BASE METAL WILL INCREASE THE PROBABILITY OF HOT CRACKING OF THE WELD METAL. TYPE 312 FILLER METAL PROVIDES WELD DEPOSITS THAT CONTAIN A LARGE AMOUNT OF FERRITE TO PREVENT HOT CRACKING.

(F) WELDS MADE WITH TYPES 316, 316L, 317, AND 317CB ELECTRODES OR WELDING RODS MAY OCCASIONALLY DISPLAY POOR CORROSION RESISTANCE IN THE AS-WELDED CONDITION. IN SUCH CASES, CORROSION RESISTANCE OF THE WELD METAL MAY BE RESTORED BY THE FOLLOWING HEAT TREATMENTS: FOR TYPES 316 ANE 317 BASE METAL, FULL ANNEAL AT 1065 TO 1120 °C (1950 TO 2050 °F); FOR TYPES 316L AND 317L BASE METAL, 870 °C (1600 °F) STRESS RELIEF; FOR TYPE 518 BASE METAL, 870 TO 900 °C (1600 TO 1650 °F) STABILIZING TREATMENT. WHERE POSTWELD HEAT TREATMENT IS NOT POSSIBLE, OTHER FILLER METALS MAY BE SPECIALLY SELECTED TO MEET THE REQUIREMENTS OF THE APPLICATION FOR CORROSION RESISTANCE.

(G) TYPE 321 COVERED ELECTRODES ARE NOT REGULARLY MANUFACTURED, BECAUSE TITANIUM IS NOT READILY RECOVERED DURING DEPOSITION.

(H) CAUTION IS NEEDED IN WELDING THICK SECTIONS, BECAUSE OF CRACKING PROBLEMS IN HEAT-AFFECTED ZONES.

(I) IN BASE METAL AND WELD METAL, FOR NUCLEAR SERVICE, TANTALUM IS RESTRICTED TO 0.10% MAX AND COBALT TO 0.20% MAX.

(J) ANNEALING SOFTENS AND IMPARTS DUCTILITY TO HEAT-AFFECTED ZONES AND WELD. WELD METAL RESPONDS TO HEAT TREATMENT IN A MANNER SIMILAR TO THE BASE METAL.

(K) THESE AUSTENITIC WELD METALS ARE SOFT AND DUCTILE IN AS-WELDED CONDITION, BUT THE HEAT-AFFECTED ZONE WILL HAVE LIMITED DUCTILITY.

(L) REQUIRES CAREFUL PREHEATING AND POSTWELD HEAT TREATMENT TO AVOID CRACKING.

(M) REQUIRES CAREFUL PREHEATING. SERVICE IN AS-WELDED CONDITION REQUIRES CONSIDERATION OF HARDENED HEAT-AFFECTED ZONES.

(N) ANNEALING INCREASES DUCTILITY OF HEAT-AFFECTED ZONES AND WELD METAL. TYPE 405 WELD METAL CONTAINS NIOBIUM, RATHER THAN ALUMINUM, TO REDUCE HARDENING.

(O) ANNEALING IS EMPLOYED TO INCREASE DUCTILITY OF THE WELDED JOINT.

(P) TYPE 308 FILLER METAL WILL NOT DISPLAY SCALING RESISTANCE EQUAL TO THAT OF THE BASE METAL. CONSIDERATION MUST BE GIVEN TO DIFFERENCES IN THE COEFFICIENTS OF THERMAL EXPANSION OF THE BASE METAL AND THE WELD METAL.

Ferritic stainless steels (types 446, 430, 405, and 409) are generally less weldable than the austenitic grades. Preheat temperatures ranging from 150 to 230 °C (300 to 450 °F) are necessary to prevent weld cracking, warping, and other serious repair problems. Postweld heat treatment ranging from 790 to 815 °C (1450 to 1500 °F) is necessary to transform a mixed structure to a wholly ferritic structure and to restore the mechanical properties and corrosion resistance that were seriously affected by the heat of welding. The filler metals in Table 3 apply to these grades of stainless steels, as well.

Martensitic stainless steels (types 403, 410, 414, 420, and the 440 series) are weldable with adequate preheat and interpass temperature control ranging from 205 to 315 °C (400 to 600 °F). To restore mechanical properties and reduce the hardened HAZ, a postweld heat treatment ranging from 650 to 675 °C (1200 to 1250 °F) is commonly used to restore properties and reduce internal welding stresses. Austenitic electrodes are used for minor repairs or when properties are not the prime consideration (see Table 3).

Precipitation-hardening stainless steels are divided into three groups on the dual basis of characteristic alloying additions, particularly the elements added to promote precipitation hardening (PH), and the matrix structures of the steels in the solution-annealed and aged condition. Differences among the steels have a direct bearing on the behavior of the steels in welding.

Martensitic Precipitation-Hardening Steels. These steels have a predominantly austenitic structure at the solution-annealing temperature of approximately 1040 to 1065 °C (1900 to 1950 °F), but they undergo an austenite-to-martensite transformation when cooled to room temperature. These steels can be readily welded. The welding procedures resemble those ordinarily used for the 300-series stainless steels, despite differences in composition and structure between the two classes.

The selection of filler metal depends on the properties required for the welded joint (Table 4). If strength comparable to that of the base metal is not required at the welded joint, then a tough 300-series stainless steel filler metal may be adequate. When a weld having mechanical properties comparable to those of the hardened base metal is desired, the filler metal must be of comparable composition, although slight modifications are permissible to obtain better weldability.

TABLE 4 RECOMMENDED FILLER METALS FOR WELDING PRECIPITATION-HARDENING STAINLESS STEELS

TYPE	**UNS DESIGNATION**	**COVERED ELECTRODES**	**BARE WELDING WIRE**	DISSIMILAR PH STAINLESS STEELS
MARTENSITIC				
17-4 PH	S17400	AMS 5827B (17-4 PH) OR E308	AMS 5826 (17-4 PH) OR ER308	**E309 OR ER309, E309CB OR ER309CB**
15-5 PH	S15500	AMS 5827B (17-4 PH) OR E308	AMS 5826 (17-4 PH) OR ER308	**E309 OR ER309, E309CB OR ER309CB**
SEMIAUSTENITIC				
17-7 PH	S17700	AMS 5827B (17-4 PH), E308, OR E309	AMS 5824A (17-7 PH)	**E310 OR ER310, ENICRFE-2, OR ERNICR-3**

PH 15-7 MO	S15700	E308 OR E309	AMS 5812C (PH 15-7 MO)	**E309 OR ER309, E310 OR ER310**
AM350	S35000	AMS 5775A (AM350)	AMS 5774B (AM 350)	**E308 OR ER308, E309 OR ER309**
AM355	**S35500**	**AMS 5781A (AM355)**	**AMS 5780A (AM 355)**	**E308 OR ER308, E309 OR ER309**

Source: Ref 2

Semiaustenitic Precipitation-Hardening Steels. Unlike martensitic PH steels, semiaustenitic PH steels are soft enough in the annealed condition to permit cold working. When cooled rapidly from the annealing temperature to room temperature, they retain their austenitic structure, which displays good toughness and ductility in cold-forming operations.

The semiaustenitic PH steels are normally welded in the annealed condition. The tough austenitic structure imparts welding characteristics similar to those of 300-series stainless steels. The semiaustenitic PH steels are not susceptible to cracking when welded after transformation to martensite, because the low-carbon martensite developed is not of high hardness or low ductility. Also, cold cracking does not occur in the base metal adjacent to the weld because the HAZ is austenitized during welding and remains substantially austenitic as the joint cools to room temperature.

The choice of filler metal depends largely on the weld properties desired. The filler metal can be an alloy of precipitation-hardening composition that is capable of developing mechanical properties comparable to those of the base metal (Table 4). If high strength is not required, then the filler metal can be a 300-series austenitic stainless steel. When these steels are welded in the annealed condition, certain microstructural relations are generally obtained as a result of relatively rapid heating and cooling at the joint:

- WELD METAL CONTAINS SMALL AMOUNTS OF FERRITE IN AN ESSENTIALLY AUSTENITIC MATRIX; HARDNESS IS APPROXIMATELY 90 HRB
- BASE METAL IMMEDIATELY ADJACENT TO THE WELD DISPLAYS HIGH-TEMPERATURE ANNEALED (AUSTENITIC) STRUCTURE; HARDNESS IS APPROXIMATELY 90 HRB
- BASE METAL IN THE NARROW ZONE JUST BEYOND THE ANNEALED ZONE NEXT TO THE WELD IS HARDENED SLIGHTLY; HARDNESS IS APPROXIMATELY 90 TO 98 HRB

Austenitic Precipitation-Hardened Steels. The alloy content of these steels is high enough to maintain an austenitic structure after annealing and after any aging or hardening treatment. Although the austenitic PH steels remain austenitic during all phases of forming, welding, and heat treatment, some contain alloying elements (for precipitation-hardening purposes) that greatly affect behavior in welding.

The austenitic PH stainless steels can be welded using the arc welding techniques applied when welding the austenitic stainless steels. The major difference is that these steels are usually heat treated after welding to achieve the required mechanical properties, which is usually unnecessary with austenitic PH stainless steels. Austenitic PH stainless steels can be welded with matched or dissimilar filler metals or without filler metals, as in the case of most stainless steels. There is a wide variety of hardenable filler metals available for these PH grades through the manufacturer of consumables. The most commonly used grade is the AWS A5.4 Class E630/AWS A5.9 Class ER630 alloy, the only one currently included in the AWS specifications. Its composition is shown in Table 5.

TABLE 5 COMPOSITION OF WELDING CONSUMABLES FOR PRECIPITATION-HARDENING STAINLESS STEELS

All data are maximum percentages.

AWS CLASSIFICATION	COMPOSITION, %									
	C	**Cr**	**Ni**	**Mo**	Nb + Ta	**Mn**	**Si**	**P**	**S**	Cu
E630(A), ER630(B)	**0.05**	**16.0-16.75**	**4.5-5.0**	**0.75**	**0.15-0.30**	**0.25-0.75**	**0.75**	**0.04**	**0.03**	**3.25-4.00**

Source: AWS A5.4, A5.9

(A) UNDILUTED WELD METAL COMPOSITION.

(B) CONSUMABLE COMPOSITION.

Duplex Stainless Steels (Ref 3, 4, 5). The physical and mechanical properties of duplex (ferritic-austenitic) stainless steels affect the welding process. Because of their better stress-corrosion cracking resistance and appreciably higher yield and tensile strengths, these steels are currently used as direct substitutes for austenitic stainless steels when service above about 260 to 315 °C (500 to 600 °F) is not required.

Because duplex stainless steels generally have a higher thermal conductivity and a lower coefficient of thermal expansion than austenitic stainless steels, they generally exhibit less distortion during welding than do austenitic stainless steels.

The general welding characteristics of the duplex alloy steels are very similar to those of austenitic stainless steels. They can be welded by any of the conventional arc welding processes (SMAW, GTAW, GMAW, PAW, and SAW). In addition, electron-beam and laser-beam welding are used, as well as resistance welding. Heat input should be low enough to minimize intergranular carbide precipitation. Surface cleanliness is a must when welding duplex stainless steels. It is necessary to eliminate any source of hydrogen in the welding operation. For the gas-shielded processes, particularly on pipe, argon-helium purge gas should be used.

Filler metal selection is critical. A filler metal with matching composition may result in a higher ferrite content than that of the base metal. Gooch (Ref 4) and others have shown that filler metals of duplex stainless steel with higher nickel and/or nitrogen are preferred. Weld metal cracking in duplex steels has rarely been identified. Because extensive austenite-to-ferfite and ferrite-to-austenite transformations occur in the HAZ, along with grain growth, the welding procedure must be carefully controlled. The extent of the above transformations greatly depends on the composition and the precise weld thermal cycle experienced by the HAZ during welding. Because of their high chromium level, these steels are prone to σphase and 475 °C (885 °F) embrittlement. Though the weld thermal cycle is too short for σ-phase formation of 475 °C (885 °F) embrittlement to occur, care must be exercised in welding heavy-section steels of this type.

Titanium

Titanium is a reactive metal that requires special considerations before and during the welding cycle. The metal must be cleaned at least 50 mm (2 in.) beyond any gas shield. Abrasive blasting is an acceptable cleaning method. Chlorinated solvents (for example, trichloroethylene), should not be used for degreasing titanium alloys because the chlorine residues can result in intergranular cracking during subsequent heating operations.

Titanium alloys are most successfully welded in sealed chambers, although welding in air can be accomplished with a suitable trailing gas shield. Argon gas is the preferred gas for welding and is also used for welding in argon atmosphere chambers to protect the molten metal from oxidation.

The most commonly used process for repairing titanium components is GTAW, using AWS specification A5.12 Class EWTh2 and EWTh1 tungsten electrodes. Helium and argon-helium gas mixtures are also used when higher heat inputs are necessary for deeper penetration and when welding thicker sections. The GMAW process is also effectively applied on thicker parts (6.35 mm, or 0.25 in.) and large parts or when significant amounts of weld metal is to be deposited. Thirteen filler metals of pure and alloyed titanium are classified in AWS specification A5.16 ("Titanium and Titanium Alloy Welding Electrodes and Rods").

All grades of unalloyed titanium are welded in the annealed condition because annealing offsets the effect of cold working with its attendant loss of strength. Metastable beta alloys are welded in the annealed or solution-treated condition, where the welds are low in strength, but ductile. To obtain full strength, the weld is cold worked by peening or planishing and the weldment is solution treated and aged. Additional information can be found in the article "Welding of Titanium" in this Volume.

References cited in this section

1. G.E. LINNERT, WELDING CHARACTERISTICS OF STAINLESS STEELS, *MET. ENG. QUART.,* NOV 1967
2. *WELDING HANDBOOK,* VOL 4, *METALS AND THEIR WELDABILITY,* 7TH ED., AWS, 1982
3. H.B. CARY, *MODERN WELDING TECHNOLOGY,* 2ND ED., PRENTICE-HALL, 1989
4. T.G. GOOCH, WELDABILITY OF DUPLEX FERRITIC-AUSTENITIC STAINLESS STEELS, *DUPLEX STAINLESS STEELS,* R.A. LULA, ED., PROC. CONF. ON DUPLEX AUSTENITIC-FERRITIC STAINLESS STEELS, ASM, OCT 1982
5. S.A. DAVID, WELDING OF STAINLESS STEELS, *ENCYCLOPEDIA OF MATERIALS SCIENCE AND ENGINEERING,* VOL 7, M.B. BEVER, ED., THE MIT PRESS, P 5316-5320

Repair Welding

Vern Sutter, American Welding Institute; Robert J. Dybas, General Electric Company

References

1. G.E. LINNERT, WELDING CHARACTERISTICS OF STAINLESS STEELS, *MET. ENG. QUART.,* NOV 1967
2. *WELDING HANDBOOK,* VOL 4, *METALS AND THEIR WELDABILITY,* 7TH ED., AWS, 1982
3. H.B. CARY, *MODERN WELDING TECHNOLOGY,* 2ND ED., PRENTICE-HALL, 1989
4. T.G. GOOCH, WELDABILITY OF DUPLEX FERRITIC-AUSTENITIC STAINLESS STEELS, *DUPLEX STAINLESS STEELS,* R.A. LULA, ED., PROC. CONF. ON DUPLEX AUSTENITIC-FERRITIC STAINLESS STEELS, ASM, OCT 1982
5. S.A. DAVID, WELDING OF STAINLESS STEELS, *ENCYCLOPEDIA OF MATERIALS SCIENCE AND ENGINEERING,* VOL 7, M.B. BEVER, ED., THE MIT PRESS, P 5316-5320

Repair Welding

Vern Sutter, American Welding Institute; Robert J. Dybas, General Electric Company

Selected References

- M.J. DONACHIE, ED., *TITANIUM: A TECHNICAL GUIDE,* ASM INTERNATIONAL, 1988
- "RECOMMENDED PRACTICES FOR AIR CARBON ARC GOUGING AND CUTTING," ANSI/AWS C5.3-91, AMERICAN WELDING SOCIETY
- R.D. STOUT, *WELDABILITY OF STEELS,* 4TH ED., WELDING RESEARCH COUNCIL, 1987
- *WELD SURFACING AND HARDFACING,* THE WELDING INSTITUTE, CAMBRIDGE, U.K., 1980

Fitness-for-Service Assessment of Welded Structures

J.R. Gordon, Edison Welding Institute

Introduction

IT IS GENERALLY ACCEPTED THAT all welded structures enter service containing flaws that can range from volumetric discontinuities, such as porosity or slag inclusions, to planar defects, such as a lack of side-wall fusion or hydrogen cracks. For this reason, most welding codes include clauses that cover the inspection of production welds and set acceptance standards for the discontinuities that may be found. Any flaws that are not acceptable are repaired.

The acceptance standards included in the majority of codes are based on good workmanship practices. Although this is a very useful basis, in some cases it can result in the repair of relatively harmless flaws, with the possibility that the repair weld (often made under more-demanding conditions than the production weld) will contain a more-serious flaw. Since the early 1980s, a number of fracture mechanics-based assessment procedures have been developed that enable the significance of weld discontinuities to be assessed on a "fitness-for-service" basis. Using this concept, a structure is considered to be fit for service if it can be operated safely throughout its design life. The adoption of fitness-for-service concepts in several codes has resulted in the development of more-rational flaw-acceptance criteria.

Fitness-for-service assessment procedures also can be used to assess the integrity, or remaining life, of components in service. Indeed, one of the most widespread uses of fitness-for-service concepts is the justification of life extension, which is a term that has often been misunderstood. The purpose of life-extension activities is not to continue the operation of a plant beyond its useful life but merely to ensure full utilization up to its useful life. The idea is to avoid premature retirement of plants and plant components, on the basis of the so-called design life, because actual useful lives can often exceed the design life.

Overview of Failure Modes. Depending on the operating environment and the nature of the applied loading, a structure can fail by a number of different modes, including:

- BRITTLE FRACTURE
- DUCTILE FRACTURE
- PLASTIC COLLAPSE
- FATIGUE
- CREEP
- CORROSION
- BUCKLING

These failure modes can be broken down into the following broad categories:

- FRACTURE
- FATIGUE
- ENVIRONMENTAL CRACKING
- HIGH-TEMPERATURE CREEP

Fracture includes the conventional static failure modes, such as brittle fracture, ductile fracture, plastic collapse, buckling, as well as dynamic fracture. Fatigue failure is caused by cyclic loading. Environmentally assisted cracking includes corrosion, stress-corrosion cracking, hydrogen-induced cracking, and other environmentally assisted failure modes. High-temperature creep includes creep rupture and creep crack growth. This article discusses each of these categories, as well as the benefits of a fitness-for-service approach.

Fracture

Of all the conventional failure modes, brittle fracture is undoubtedly the most spectacular and dangerous mode of fracture, frequently occurring with no prior warning and with the potential to produce devastating consequences. The notched-bar impact test was one of the earliest tests developed to study the fracture behavior of steel. The most widely used impact test, as of the early 1990s, is the Charpy impact test, the results of which are most useful when presented as a plot of absorbed energy or percentage shear fracture against test temperature (Fig. 1).

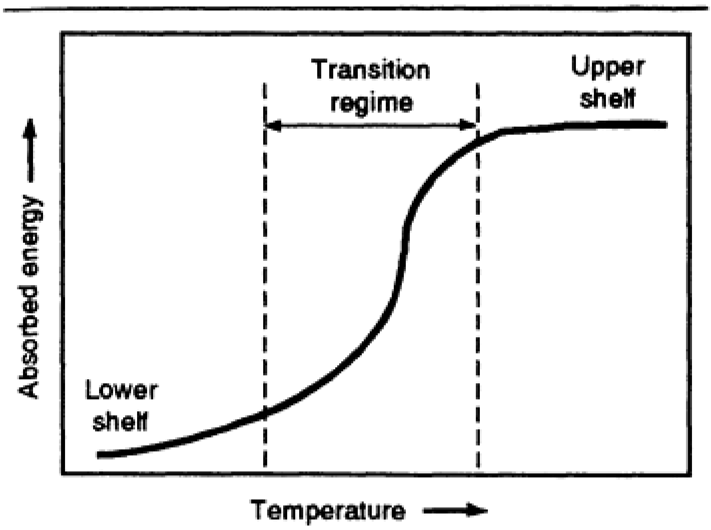

FIG. 1 TYPICAL CHARPY IMPACT ENERGY TRANSITION CURVE

In the case of ferritic steels, the subsequent results show fracture behavior that ranges from low-energy brittle fracture (lower shelf) to high-energy ductile fracture (upper shelf). Between these two modes of fracture, there is a transition regime, in which the fracture mode changes from brittle to ductile fracture as the temperature increases. In design, Charpy impact data are used to ensure that the material exhibits good toughness properties (that is, upper-transition or upper-shelf behavior) at the minimum operating temperature. Most design codes include Charpy impact toughness requirements to minimize the risk of structures failing by unstable fracture.

Although the Charpy impact test has received widespread acceptance, it has a number of drawbacks. First, unlike a tensile test, the results from notched-bar impact tests are highly dependent on specimen size. Larger specimens translate the transition curve to higher temperatures because of increased specimen constraint and sampling size.

Second, toughness is also a function of loading rate. In most structural applications, this rate is relatively low, whereas in a Charpy test, the specimen is subjected to impact loading. If Charpy specimens were tested under loading rates that were more representative of structural applications, then the transition curve could be translated to a lower temperature.

Third, Charpy specimens contain 2 mm (0.8 in.) deep V notches, which are relatively blunt. In comparison, structures can contain much sharper cracks, such as fatigue cracks. Increasing the notch acuity in a Charpy specimen would cause the transition curve to shift to a higher temperature.

Fracture comprises two phases: initiation and propagation. Initiation is defined as the first extension of an existing defect, whereas propagation is the continued extension thereafter. Which of these phases is the most important depends on the material and the structural application. Unfortunately, the Charpy impact test does not distinguish between these two phases; that is, it just provides an indication of the energy necessary to initiate and propagate a crack.

Thus, although the Charpy impact test is widely accepted by engineers and designers because it can indicate the "quality" of a material, its most-significant shortcoming is that the results cannot be used in any quantitative sense to assess the significance of flaws in a structure. Fracture mechanics can, to a certain extent, bridge the gap between the measurement of material properties and their application to design.

Fracture Mechanics Assessment Procedures. As shown in Fig. 2, fracture mechanics seeks to relate the three parameters that combine to control the fracture process. These parameters are: the size of the discontinuity, material toughness, and the applied stress.

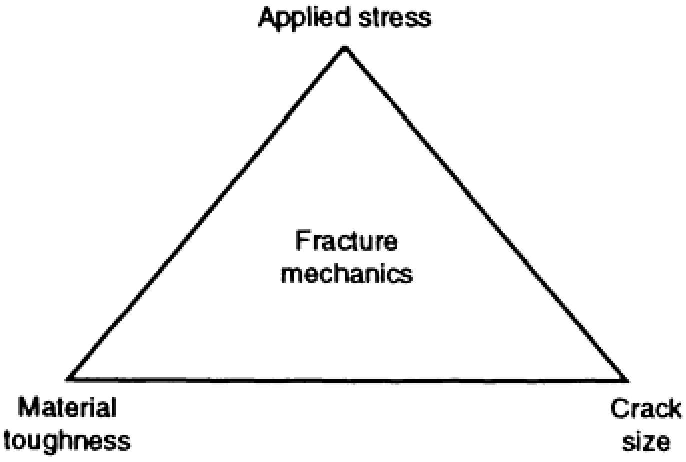
Applied stress
Fracture mechanics
Material toughness
Crack size

FIG. 2 ROLE OF FRACTURE MECHANICS

If any two of these three parameters are known, then it is possible, using the principles of fracture mechanics, to estimate the value of the third parameter that will elicit the failure of the structure. Alternatively, if all three parameters are known, then it is possible to predict whether or not the structure is fit for service, and, if it is, estimate the margin of safety.

Figure 3 summarizes the different loading paths that can result in the failure of a structure (assuming that fatigue, creep, and corrosion effects are not significant). The loading paths range from brittle fracture under nominally elastic loading (applied stresses well below yield) to plastic collapse (overload of remaining ligament). In cases where brittle fracture occurs at low applied stresses, the concept of linear-elastic fracture mechanics (LEFM) can be applied, that is, a stress-intensity factor, *K*, approach. At the other extreme, when the failure mechanism is plastic overload, then assessments should be performed using limit load or plastic-collapse analyses.

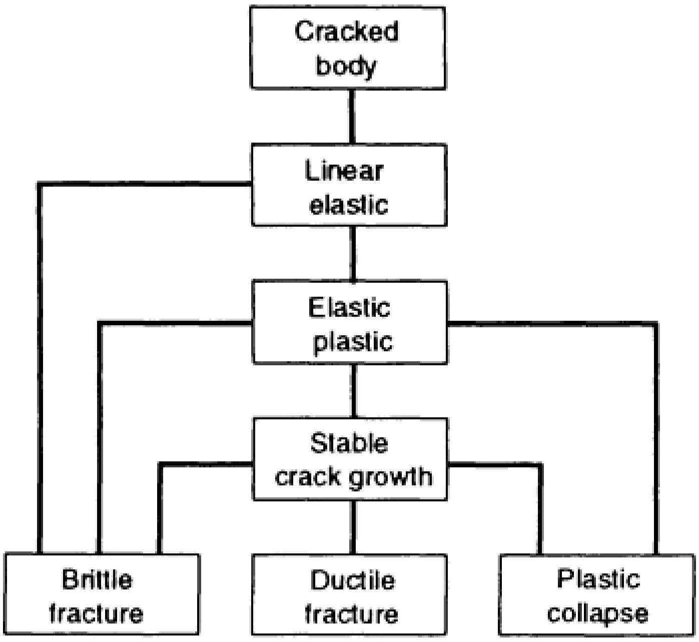
Cracked body
Linear elastic
Elastic plastic
Stable crack growth
Brittle fracture
Ductile fracture
Plastic collapse

FIG. 3 SUMMARY OF LOADING PATHS TO FAILURE

Between these two extremes, elastic-plastic fracture mechanics (EPFM) methods can be applied to assess the integrity of structures. The two most common EPFM fracture-characterizing parameters are the crack-tip opening displacement (CTOD) and the J-integral, *J*. Fitness-for-service methodologies have been developed using both of these parameters.

A number of different fitness-for-service assessment methodologies for calculating allowable or critical flaw sizes are currently in use throughout the world, such as those cited in Ref 1, 2, 3, 4, 5.

Failure-Assessment Diagrams. The concept of performing fitness-for-service assessments using a failure-assessment diagram (FAD) was first proposed in 1975. These diagrams provide a convenient method of assessing the risk of failure by brittle fracture, ductile fracture, or plastic collapse, that is, all the loading/failure paths presented in Fig. 3. As a result, most of the widely used fracture assessment methodologies are now based on FADs. Although the exact form of the FADs in the different assessment procedures can vary, the basic concept of the FAD is identical; that is, the *y*-axis (K_r axis) of a FAD provides an indication of the resistance of the structure to brittle fracture, whereas the *x*-axis (S_r axis) assesses resistance to plastic collapse. The failure-assessment curve interpolates between these two limiting failure modes, as shown in Fig. 4.

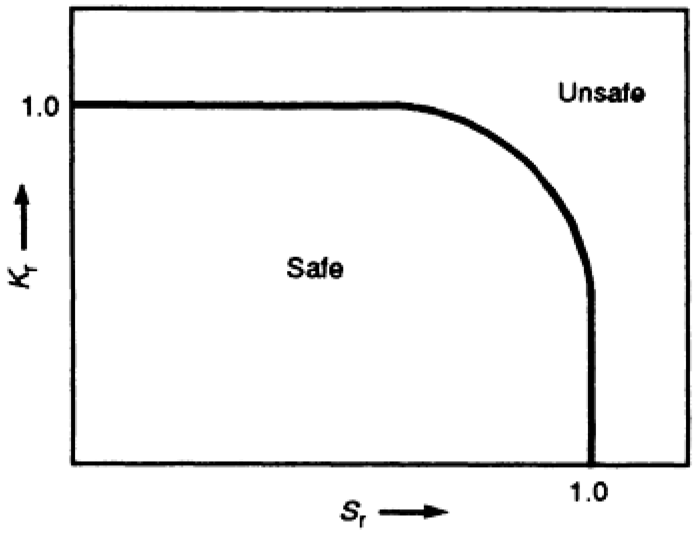

FIG. 4 BASIC CONCEPT OF FAILURE-ASSESSMENT DIAGRAM

To perform a structural integrity assessment, the user plots assessment points on the FAD. The location of each point is a function of the applied loading conditions, flaw size, material properties, and other factors.

Although the exact procedures for calculating assessment points can vary from procedure to procedure, the general procedure to determine an assessment point is as follows:

$$K_t = \frac{k_l}{k_{\mathrm{ma}\,t}} \qquad \textbf{(EQ 1)}$$

$$S_t = \frac{s}{s_{pc}} \qquad \textbf{(EQ 2)}$$

where K_l is the applied stress intensity for crack under consideration subjected to the primary and secondary loading conditions of interest; K_{mat} is the material toughness (for example, K_{Ic}); σis the applied primary stress; and σ_{pc} is the plastic collapse stress for the component under consideration.

It should be noted from the above equations that secondary stresses (for example, welding residual stresses or thermal stresses) are incorporated in the calculation of the K_r coordinate but are not included in the calculation of S_r because, by definition, secondary stresses are self-equilibrating and, as a result, cannot contribute to plastic collapse.

Although fracture assessment procedures have improved significantly in recent years due to the advances in fracture mechanics, there are still a number of issues, particularly with regard to performing fracture assessments of welded structures, that have not been fully resolved. It should be noted that the current fracture assessment procedures are generally unsafe. The unresolved issues tend to make current predictions conservative. A better understanding of these complexities would merely make analyses of critical conditions more accurate.

The major unresolved issues with regard to applying fracture mechanics assessment procedures to welded structures are undoubtedly weld metal mismatch and welding residual stresses. When designing a welded structure, it is normal practice to specify welding consumables with a higher yield strength than the surrounding parent material to avoid the possibility of localized plastic straining in the weld metal. Conversely, there are a number of materials that can only be welded successfully with consumables that will give rise to undermatching welds. This is becoming more common with the increasing use of high-heat-affected zone (HAZ), and parent material cannot be taken into account in the currently available fracture assessment models, it is normal practice to assume the tensile properties of the material in which the defect is located (that is, if a crack is located in the weld metal, the assessment is performed assuming the tensile properties of the weld metal). In reality, the difference in tensile properties between the different weld regions will influence how plasticity develops at defects and, hence, the relationship between crack driving force and applied loading. This can lead to significant errors in a fracture assessment.

At present, the only method of performing an accurate fracture assessment of under or overmatched welded joints is to perform a detailed finite-element analysis to compute the applied crack driving force. This level of sophistication is clearly excessive in situations where assessments of simple welded components are required on a regular basis. Research is required to investigate the effect of weld metal mismatch on the accuracy of fracture prediction models. Results from such research will hopefully lead to the proposal of modifications to fracture assessment models to account for weld metal mismatch.

Another unresolved issue, which, again, is particularly relevant to welded structures, concerns the estimation (prediction) and incorporation of residual stresses (or secondary stresses) in fracture assessment models. The major problem is that there is no quick or convenient method of measuring or predicting the through-thickness variation of residual stresses in a welded joint. Consequently, in the absence of definitive measurements, it is common practice to assume that residual stresses in an as-welded structure are both tensile and yield strength in magnitude. Moreover, it is common practice to assume that the residual stresses are uniform across the section under consideration, despite the fact that the assumed residual stress distribution is not self-equilibrating. This, undoubtedly, leads to extremely conservative assessments, particularly when it is considered that the residual stress components frequently play a major role in an assessment. The other area of concern with regard to residual stresses is the relaxation of residual stresses with increasing primarily load (that is, mechanical stress relief). Although the latest version of PD 6493 includes a procedure for calculating the relaxation of secondary stresses with increasing primary load, there is concern that the method may still overestimate the subsequent residual stress levels.

If the assessment points lie within the area bounded by the axes of the failure-assessment diagram and the failure-assessment curve, then the structure is deemed to be safe. Alternatively, if the assessment points lie outside the FAD, then the structure is considered to be unsafe. This approach also can be used to determine the limiting flaw size by plotting a series of assessment points for different crack sizes. The limiting flaw size corresponds to the flaw size that results in the assessment locus intersecting the FAD, as shown in Fig. 5.

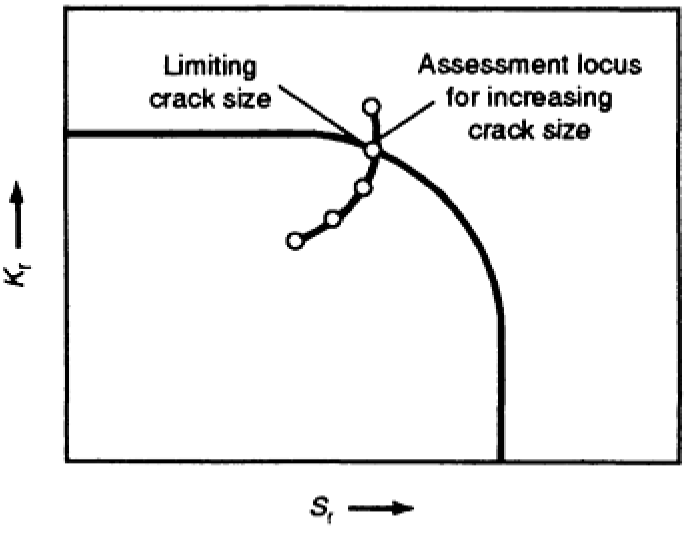

FIG. 5 ASSESSMENT PROCEDURE TO DETERMINE LIMITING CRACK SIZE

When performing fracture assessments, it is essential that the fracture-toughness values used in the assessment are representative of the material in which the crack is located (for example, parent material, weld metal, or heat-affected zone). Furthermore, the fracture-toughness values should be representative of the condition of the material or the welded joints in the structure. For example, if the material has suffered degradation, such as temper or hydrogen embrittlement, then tests should be performed on material/welded joints in the same condition. Finally, to ensure conservative assessments, the fracture-toughness tests should be performed at the minimum operating temperature of the structure.

Application of Fracture-Assessment Procedures. Fitness-for-service design philosophies can be applied to welded structures in several different ways. The most widely used approach is to develop relaxed flaw-acceptance criteria. These are typically derived by assuming a minimum toughness level that is incorporated in a weld procedure qualification. These relaxed flaw-acceptance criteria can result in significant economic savings and can accelerate construction schedules through a dramatic reduction in the number of repairs. Nevertheless, it should be emphasized that the adoption of fitness-for-service concepts should never be used to condone poor workmanship. The general quality of fabrication should remain constant, regardless of whether fitness-for-service concepts are applied.

Fitness-for-service concepts also can be used to develop inspection criteria, including sensitivity requirements. These concepts also can be used to demonstrate that flaws that may escape detection will not impair overall structural integrity, even under accident or overload conditions.

References cited in this section

1. "GUIDANCE ON SOME METHODS FOR THE DERIVATION OF ACCEPTANCE LEVELS FOR DEFECTS IN FUSION WELDED JOINTS," PD 6493, BRITISH STANDARDS INSTITUTION, 1991
2. F.M. BURDEKIN, "FINAL REPORT ON QUESTIONNAIRE ON THE USE OF FRACTURE MECHANICS METHODS FOR THE ASSESSMENT OF THE SIGNIFICANCE OF WELD DEFECTS," DOCUMENT X-1076-84, IIW, 1984
3. I. MILNE, R.A. AINSWORTH, A.R. DOWLING, AND A.T. STEWART, "ASSESSMENT OF THE INTEGRITY OF STRUCTURES CONTAINING DEFECTS," CEGB R/H/R6, REVISION 3, MAY 1986
4. V. KUMAR, M.D. GERMAN, AND C.F. HIH, "AN ENGINEERING APPROACH FOR ELASTIC-PLASTIC FRACTURE ANALYSIS," REPORT EPRI NP 1931, RESEARCH PROJECT 1237-1, GENERAL ELECTRIC COMPANY, JULY 1981
5. J.M. BLOOM, *VALIDATION OF A DEFORMATION PLASTICITY FAILURE ASSESSMENT DIAGRAM APPROACH TO FLAW EVALUATION,* STP 803, VOL II, ASTM, 1983, P 206

Fitness-for-Service Assessment of Welded Structures

J.R. Gordon, Edison Welding Institute

Fatigue Design

The fatigue life of a welded structure is almost invariably limited by the fatigue life of the weld details. As a result, the design stresses in welded structures that are subjected to cyclic loading are limited by the fatigue strength of the weld details. For this reason, extensive research has been undertaken in recent years to study the fatigue behavior of welded joints. This has culminated in a number of proposed methodologies to assess the fatigue behavior of different welded joint geometries. The most widely used procedure for assessing the fatigue performance of welded joints is the *S-N* curve approach.

In general, the fatigue life of a component is comprised of initiation and propagation phases. For smooth components, the crack-initiation period represents the bulk of the total fatigue life. This is particularly noticeable with high-fatigue applications, where the fatigue crack-initiation period may exceed 95% of the fatigue life. In the case of machined components, once a fatigue crack has grown to a detectable size, the component is virtually at the end of its useful life and would normally be withdrawn from service. In the case of welded joints, there is practically no crack-initiation period, because of the presence of weld-toe discontinuities, which behave as preexisting cracks. As a result, the bulk of the fatigue life of a welded joint can be attributed to fatigue-crack propagation.

The fatigue strength of unwelded components increases with material tensile strength. In the case of welded joints, however, the fatigue strength is relatively unaffected by material tensile strength, as shown in Fig. 6. This is because the bulk of the fatigue life of a welded joint is spent in the propagation phase and, although crack-propagation rates can change from one material to another, there is no consistent trend with regard to tensile strength. This indicates that the improved fatigue performance produced by increasing material tensile strength is limited primarily to the crack-initiation phase. As a result, increasing the material tensile strength will not improve the fatigue performance of welded joints.

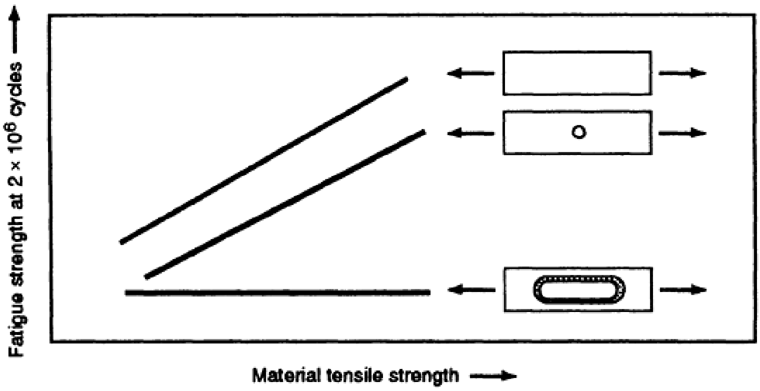

FIG. 6 EFFECT OF MATERIAL TENSILE STRENGTH OR THE FATIGUE STRENGTH OF WELDED AND UNWELDED SPECIMENS

The potential initiation sites at which fatigue cracks may develop in welded joints depends on the geometry of the joint details. Nevertheless, the most common initiation site, by far, is the weld toe. This is a particularly severe location, because of the geometric discontinuity that gives rise to a local stress concentration and the possible presence of undercutting and weld-toe intrusions (Ref 6, 7, 8, 9). The weld-toe intrusions that form as a result of the welding process are generally angular in nature and protrude below the plate surface. The precise origin of these discontinuities is not fully understood, although it has been suggested that they may be caused by slag or impurities in the steel.

Limited studies (Ref 6, 7, 8, 9) have indicated that the weld-toe discontinuities range in depth from 0.1 to 1.0 mm (0.004 to 0.04 in.). The depth of the weld-toe intrusions depends on a number of variables, including welding process, welding consumables, flux, and welding parameters. These studies have indicated that the average root radius of intrusions is from 2 to 3 μm (80 to 120 μin.), although root radii of less than 1 μm (40 μin.) have been recorded. In addition to the local stress concentration and the presence of sharp weld-toe discontinuities, the weld-toe region is also subjected to tensile residual stresses that form as a result of the thermal cycle associated with welding. Thus, the weld toe is a location of high local stress with preexisting crack-like discontinuities and large tensile residual stresses, which means that it is a prime source for fatigue cracks.

For this reason, the fatigue performance of welded joints is generally much lower than the fatigue performance of unwelded components, even if they contain stress concentrations (Fig. 7).

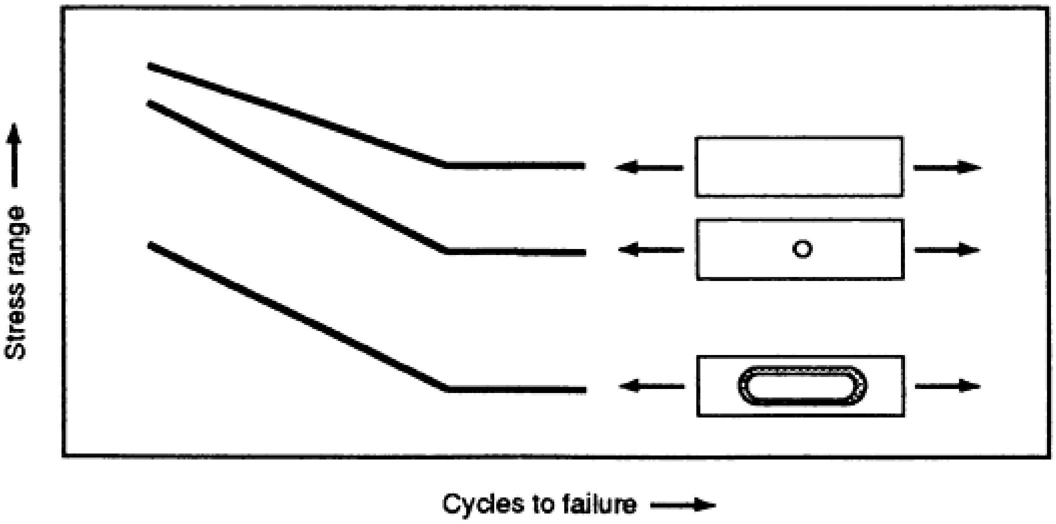

FIG. 7 RELATIVE FATIGUE BEHAVIOR OF WELDED JOINTS AND UNWELDED COMPONENT (WITH AND WITHOUT STRESS CONCENTRATORS)

The *S-N* curve approach is probably the most widely used method of assessing the fatigue performance of welded structures. The fatigue design rules proposed by Gurney in 1976 (Ref 10) provide a framework to assess the fatigue behavior of welded joints. Based on the results of an extensive welded joint fatigue test program, a series of *S-N* curves that incorporate a wide range of welded joint details was developed. Similar design approaches have now been included in many national and international codes and standards (Ref 11, 12, 13, 14). In a typical set of fatigue design curves (Fig. 8), the different classifications correspond to different welded joint geometries.

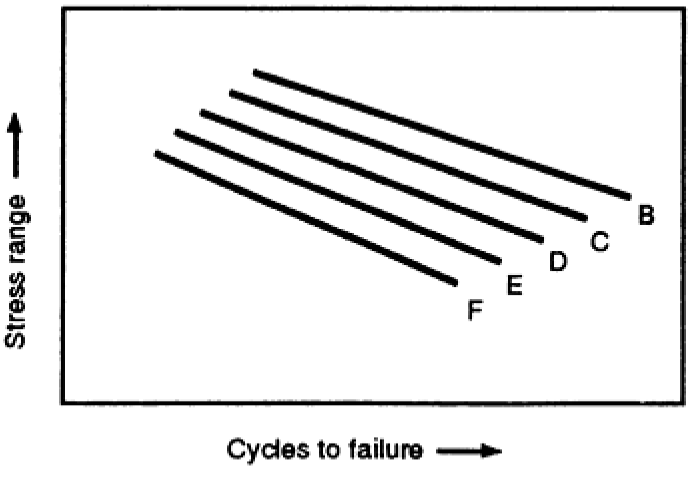

FIG. 8 SCHEMATIC OVERVIEW OF TYPICAL FATIGUE DESIGN RULES

Although the exact classification system can vary from code to code, the general approach is consistent. Most codes adopt a weld joint classification system that uses letters to denote different classes. In the original fatigue design rules proposed by Gurney, the following nine weld joint classifications were adopted: A, B, C, D, E, F, F2, G, and W; where class A is the parent material *S-N* curve obtained from uniform, machined, and polished specimens; class W is the welded joint *S-N* curve based on nominal shear stress on the minimum weld throat area; and classes B through G are the welded joint *S-N* curves for joints of decreasing fatigue performance.

S-N curves are used in design either to determine the maximum allowable cyclic stress for a given fatigue design life or to estimate design life for a given set of operating conditions.

Fracture-Mechanics Assessment Procedures. As already mentioned, the majority of the fatigue life of a welded joint is spent in the crack-propagation phase. Consequently, the analysis of fatigue cracking in welded joints is well suited to fracture mechanics, which can be used to predict the fatigue strength of nominally sound welds by considering the propagation of a fatigue crack from the inherent discontinuities that exist at the weld toe and the weld root. Fracture mechanics also can be used to predict the life of a joint with a weld discontinuity. The most widely accepted fatigue crack growth relationship is the Paris law (Ref 15), which can be expressed as:

$$\frac{da}{dN} = C(\Delta K)^m \qquad \textbf{(EQ 3)}$$

where ΔK is the applied stress-intensity factor range, a is the crack size, N is the number of cycles, and C and m are constants that depend on the material and environment.

Maddox (Ref 16) has carried out a series of fatigue crack propagation tests on various parent steels, weld metals, and heat-affected zone (HAZ) microstructures. He found that there was no significant variation in crack-propagation rates among any of the materials tested. The upper-bound crack-growth rate constants for ferritic steels operating in nonaggressive environments are given by:

$$C = 3.0 \times 10^{-13} \text{ AND } M = 3.0 \qquad \textbf{(EQ 4)}$$

For structural ferritic steels operating in marine environments up to 20 °C (68 °F), the following crack-growth constants can be used:

$$C = 2.3 \times 10^{-13} \text{ AND } M = 3.0 \qquad \textbf{(EQ 5)}$$

For design purposes, it is sometimes necessary to estimate the limiting flaw size that will not extend by fatigue during service. The limiting crack size, below which fatigue crack growth will not occur, can be calculated using fatigue-threshold concepts. The following threshold stress-intensity factor ranges have been proposed for carbon and carbon-manganese steels operating in air or seawater environments (Ref 1, 2):

$$\begin{aligned} &\Delta K_0 = 5.44 - 6.85R \text{ MPA } \sqrt{m} \text{ FOR } 0 \leq \text{R} \leq 0.5 \\ &K_0 = 63 \text{ N} \sqrt{mm} \text{ FOR } 0.5 \leq \text{R} \end{aligned} \qquad \textbf{(EQ 6)}$$

where R is the stress ratio (that is, minimum stress divided by maximum stress). It should be noted that welding residual stresses must be taken into account when calculating the stress ratio. Using the above expressions, it is possible to calculate the limiting flaw size that will not extend by fatigue during service.

Fracture mechanics-based fatigue-assessment procedures can also be used to define inspection criteria and, in particular, inspection intervals. This is achieved by assuming the maximum flaw size that may escape detection and by specifying a critical crack size that should be detected to avoid unnecessary damage to the structure or the risk of failure. Using fracture mechanics-based fatigue crack growth procedures, the user can determine the number of cycles (or time) needed for the crack to extend from the initial size to the critical size.

The time for the crack to extend from its initial size to the critical size represents the maximum inspection interval (that is, if an inspection is not performed during this period, the structure could conceivably fail without any obvious warning). To guard against premature failure, it is normal practice to set inspection intervals that are considerably less than the maximum inspection interval. This will enable the operator to determine if cracks have initiated or extended during service before they have the potential of reaching a critical size.

After inspecting the component and sizing any defects/cracks that are present, the operator can determine a new maximum inspection interval for the component by calculating the time needed for the current cracks to extend to a critical size. Using this approach, an operator may specify decreasing inspection intervals for a component that is experiencing in-service cracking. Another reason for specifying inspection intervals that are considerably less than the maximum inspection interval is that it provides the operator with an opportunity to evaluate the validity of the assessment and the initial assumptions.

References cited in this section

1. "GUIDANCE ON SOME METHODS FOR THE DERIVATION OF ACCEPTANCE LEVELS FOR DEFECTS IN FUSION WELDED JOINTS," PD 6493, BRITISH STANDARDS INSTITUTION, 1991
2. F.M. BURDEKIN, "FINAL REPORT ON QUESTIONNAIRE ON THE USE OF FRACTURE MECHANICS METHODS FOR THE ASSESSMENT OF THE SIGNIFICANCE OF WELD DEFECTS," DOCUMENT X-1076-84, IIW, 1984
6. E.J. SIGNES, R.G. BAKER, J.D. HARRISON, AND F.M. BURDEKIN, FACTORS AFFECTING THE FATIGUE STRENGTH OF WELDED HIGH STRENGTH STEELS, *BR. WELD. J.*, VOL 14 (NO. 3), 1967, P 108-116
7. S. BERGE, O.I. EIDE, AND E.T. MOE, FATIGUE CRACK INITIATION IN WELDMENTS OF A CARBON-MANGANESE STEEL, *PROC. EUROPEAN OFFSHORE STEELS RESEARCH SEMINAR*, (CAMBRIDGE, U.K.), PAPER 6, 1978
8. F.C. SMITH AND R.A. SMITH, DEFECTS AND CRACK SHAPE DEVELOPMENT IN FILLET WELDED JOINTS, *FATIGUE FRACT. ENG. MATER. STRUCT.*, VOL 5 (NO. 2), 1982, P 151-165
9. F. WATKINSON, P.H. BODGER, AND J.D. HARRISON, THE FATIGUE STRENGTH OF WELDED JOINTS IN HIGH STRENGTH STEELS AND METHODS FOR ITS IMPROVEMENT, *PROC. LNT. CONF. FATIGUE OF WELDED STRUCTURES*, THE WELDING INSTITUTE, 1971
10. T.R. GURNEY, FATIGUE DESIGN RULES FOR WELDED JOINTS, *WELD. INST. RES. BULL.*, VOL 17, MAY 1976, P 115-124
11. "SPECIFICATION FOR STEEL, CONCRETE AND COMPOSITE BRIDGES--CODE OF PRACTICE FOR FATIGUE," B5 5400, PART 10, BRITISH STANDARDS INSTITUTION, 1980
12. "SPECIFICATION FOR UNFIRED FUSION WELDED PRESSURE VESSELS--ENQUIRY CASE NO. 79," BS 5500 1988, BRITISH STANDARDS INSTITUTION
13. AWS STRUCTURAL WELDING CODE D1-1, 1992
14. AASHTO STANDARD SPECIFICATION FOR HIGHWAY BRIDGES, 13TH ED., 1983
15. P.C. PARIS AND F. ERDOGAN, A CRITICAL ANALYSIS OF CRACK PROPAGATION LAWS, *J. BASIS ENG.*, VOL 85, 1960, P 528-534
16. S.J. MADDOX, *WELD. RES. LNT.*, VOL 4 (NO. 1), 1974, P 36-60

Characterization and Modeling of the Heat Source

S.S. Glickstein and E. Friedman, Westinghouse Electric Corporation

Introduction

THE HEAT THAT IS SUPPLIED TO THE WORKPIECE, which then is transferred within the workpiece to produce melting, forms the basis of every welding process. The heat-transfer process, or thermal cycle, in the weldment has many consequences, including the complex metallurgical changes that take place in the fusion zone, where the metal is liquified and subsequently solidified, and in the adjacent heat-affected zone, where material is heated to temperatures that are below the melting point, but are sufficiently high to produce changes in the microstructure and in mechanical properties.

The description of the input-energy source is basic to any numerical modeling formulation designed to predict the outcome of the welding process. Both the magnitude and distribution of the source are fundamental and unique to each joining process, and the resultant output of any numerical model is therefore affected by the initial description of the heat source. An understanding of both the physics and the mathematical simulation of these sources is essential for characterizing the heat source. This article briefly reviews the physical phenomena that influence the input-energy distribution and discusses several simplified and detailed heat-source models that have been used in the modeling of arc welding, high-energy-density welding, and resistance welding processes.

Characterization and Modeling of the Heat Source

S.S. Glickstein and E. Friedman, Westinghouse Electric Corporation

Simplified Modeling of the Heat Source

Analytical modeling of the welding heat source is generally complex, because of the nature of energy transfer to the workpiece, whether the source of that energy is an arc, a high-energy-density beam of electrons or laser light, or joule heating. For numerical modeling purposes, heat input to the weldment is usually applied as a distribution of surface flux, a distribution of heat generated internally, or a combination of both. Many analytical treatments, however, have sought to simplify the characterization of the heat source by assuming that the effective thermal energy supplied by the heat source is deposited in such a narrow band of material that it may be idealized mathematically as a point or a line source, depending on the geometry of the weldment (Ref 1). Heat input idealizations of this sort lend themselves to the derivation of closed-form welding temperature solutions and the avoidance of developing numerical finite-element or finite-difference models to calculate temperatures. These solutions are valid only for simple geometries and in regions removed from the fusion and heat-affected zones, where details of the distribution of heat input from the source and accurate representations of the thermal energy transferred from the weld bead to the rest of the weld joint are not important.

The fundamental simplified heat-source model is developed for a flat plate of infinite extent bounded by the planes $z = 0$ and $z = h$. Heat is input at a point that is either stationary or is moving at uniform speed, v, in the x direction on the surface $z = 0$, so that at any time, t, the point source is located at $x = vt$ (Fig. 1). If the end effects that result from the initiation or termination of the heat source or the finite dimensions of the weldment are neglected, then the resulting temperature distribution associated with the moving source is stationary with respect to a moving coordinate system, the origin of which coincides with the point of application of the heat source. This class of temperature response is termed "quasi-stationary."

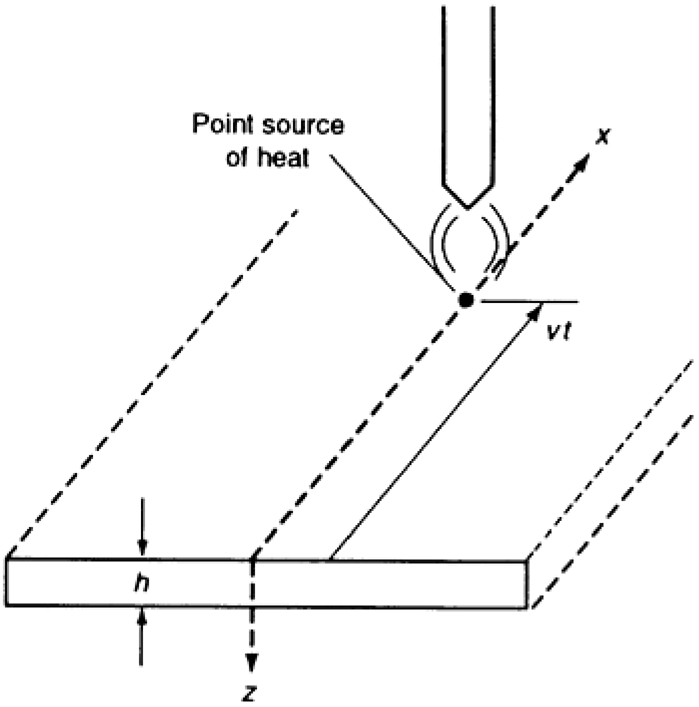

FIG. 1 MOVING POINT SOURCE IN INFINITE PLATE

In general, for either a stationary or a moving point source, heat is conducted through the plate without hindrance until the insulating effect of the ideally adiabatic surface at $z = h$ is felt. For a sufficiently thick plate, the temperature rise at $z = h$ is so small that the solution for an infinitely thick plate, which is associated with a single source of heat, is applicable. For a moderately thick plate, the temperature rise at $z = h$ from the infinitely thick plate point source solution is large enough to result in nonzero heat flow at $z = h$. An image source of the same strength applied at $z = 2h$ ensures that the surface $z = h$ is adiabatic. However, the image source produces nonzero heat flow at $z = 0$, and another image source applied at $z = -2h$ is now required to satisfy conditions at $z = 0$.

Carrying this imposition of image point sources of heat along ad infinitum (Fig. 2), an infinite distribution of image sources superposed with the original source at $z = 0$ yields the desired adiabatic conditions at both the $z = 0$ and $z = h$ surfaces. The temperature solution for this series of image sources is in the form of an infinite series (Ref 2), which converges more rapidly for thicker plates. The solution for the moderately thick plate applies to thin plates as well, but convergence of the infinite series would be extremely slow. As an alternative, the heat source can be applied as a line source distributed uniformly through the thickness (Ref 3). This approach eliminates any variation of the temperature distribution through the thickness, and is often used to model high-energy-density welding processes that result in the formation of a keyhole, as will be discussed later.

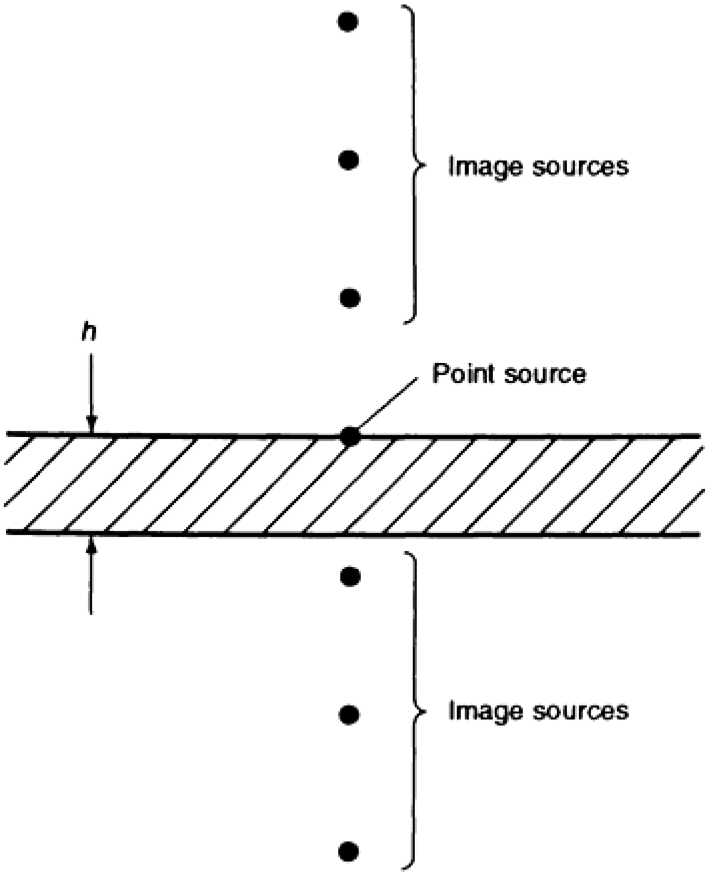

FIG. 2 SUPERPOSITION OF IMAGE SOURCES IN FINITE-THICKNESS PLATE

Offshoots of these cases also have been developed. These include considerations of plates of finite width, in which case sets of image sources are needed to ensure adiabatic conditions at the edges (Ref 4); line sources of heat traveling on a circular path (Ref 5); and conditions that simulate either initiation or termination of the heat source (Ref 6). Variations of the moving point source approach have been employed (Ref 7) to calculate the temperatures needed for weld-induced residual stress calculations in piping.

References cited in this section

1. P.S. MYERS, O.A. UYEHARA, AND G.L. BORMAN, FUNDAMENTALS OF HEAT FLOW IN WELDING, *WELD. RES. COUNC. BULL.,* NO. 123, JULY 1967
2. H.S. CARSLAW AND J.C. JAEGER, *CONDUCTION OF HEAT IN SOLIDS,* CLARENDON PRESS, LONDON, 1959
3. D. ROSENTHAL AND R. SCHMERBER, THERMAL STUDY OF ARC WELDING, *WELD J. RES. SUPPL.,* VOL 17 (NO. 4), 1938, P 2S-8S
4. D. ROSENTHAL, THE THEORY OF MOVING SOURCES OF HEAT AND ITS APPLICATION TO METAL TREATMENTS, *TRANS. ASME.,* VOL 68, 1946, P 849-866
5. W. SOEDEL AND R. COHEN, ARC-WELDING TEMPERATURES IN A CIRCULAR DISK STRUCTURE, *WELD. J. RES. SUPPL.,* VOL 49 (NO. 7), 1970, P 337S-340S
6. T. NAKA AND K. MASUBUCHI, TEMPERATURE DISTRIBUTION OF WELDED PLATES, *J. JPN. WELD. SOC.,* VOL 16 (NO. 7 AND 12), 1947, P 281-290 (PART 1) AND P 374-378 (PART II)
7. E.F. RYBICKI *ET AL.,* A FINITE-ELEMENT MODEL FOR RESIDUAL STRESSES AND DEFLECTIONS IN GIRTH-BUTT WELDED PIPES, *J. PRESSURE VESSEL TECHNOL. (TRANS. ASME),*

VOL 100, AUG 1978, P 256-262

Characterization and Modeling of the Heat Source

S.S. Glickstein and E. Friedman, Westinghouse Electric Corporation

Arc Welding

Gas-tungsten arc welding (GTAW) is the most frequently modeled arc-welding process in which the heat source is a nonconsumable electrode. In the direct current electrode negative GTAW process, the pieces of material are joined together by energy that is transferred to the workpiece by four primary mechanisms (Ref 8):

- KINETIC ENERGY OF THE ELECTRONS THAT CONSTITUTE THE ARC CURRENT
- HEAT OF CONDENSATION OF THE ELECTRONS (WORK FUNCTION) PENETRATING THE SOLID WORK SURFACE
- RADIATION FROM THE ARC
- THERMAL CONDUCTION FROM THE ARC PLASMA TO THE WORKPIECE

The first two mechanisms constitute the major source of energy to the weldment (Ref 9).

Because of the complicated nature in which energy is transferred from the arc, heat input to the weldment can be modeled by one of the point- or line-source approaches discussed in the preceding section of this article. If none of these treatments is appropriate for the particular application, then a more realistic approach is to input the energy by a distribution of surface flux, a distribution of heat generated internally, or a combination of the two. If an internal heat distribution is confined to a thin layer of material adjacent to the heated surface, then the choice of heat-input model is immaterial (Ref 10). The magnitude of heat input to the workpiece is expressed simply as the product of what has traditionally been defined as the arc power (that is, the product of the voltage drop across the arc and the arc current) and a factor called the arc efficiency, which accounts for energy losses from the arc.

Estimates of arc efficiency can be made by correlating computed temperatures with thermocouple readings from test welds, making calorimeter measurements for a specific set of welding conditions, or utilizing some other observable response to the welding thermal cycle. Because the voltage used is usually determined from measurements at some point within the power supply, estimates need to be made to account for losses within the electrode and other parts of the system (Ref 11). Changes in welding process variables, such as shielding gas, electrode configuration, arc gap, and minor element additions to the arc, all affect arc efficiency, as does the material that is being welded.

The magnitude of heat applied to the weldment surface is usually not known with great accuracy, and the distribution of the heat input is even more uncertain. Input-energy distribution depends on factors related to the electrode and the physics of the arc, as well as on the interaction of the arc with the molten weld pool. For example, the arc characteristics and the depth-to-width ratio of the weld puddle will change as the vertex angle of the conical tip of the electrode is altered (Fig. 3). As the arc becomes more constricted, the heat source becomes more concentrated, thus increasing weld penetration. Changes in heat-source distribution (that is, input current) will affect the electromagnetic forces in both the arc and the weld puddle, resulting in changes in weld-puddle motion, which can promote alterations in convection heat transfer within the weld puddle. Shielding gas composition and minor alloying elements also can influence the distribution of input energy.

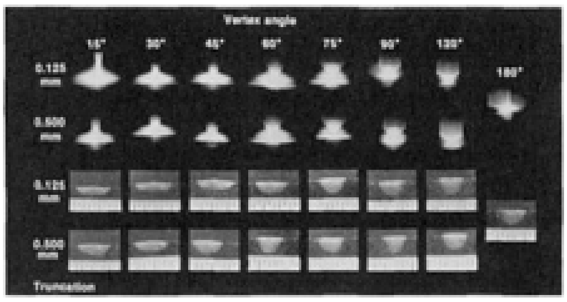

FIG. 3 ARC SHAPE AND WELD-BEAD GEOMETRY AS A FUNCTION OF ELECTRODE TIP ANGLE IN A PURE ARGON SHIELD FOR 2.38 MM (0.10 IN.) DIAMETER ELECTRODES TRUNCATED TO 0.125 AND 0.500 MM (0.005 AND 0.2 IN.); ARC GAP, 1 MM (0.04 IN.). SOURCE: REF 12

When welding at low currents and high voltages, the perturbation of the weld-puddle surface may be considered to have a small effect on the heat-input distribution, and a model that considers heat to be applied as a surface flux is usually satisfactory. However, as the current increases (particularly when the arc voltage is small, indicating a short electrode-to-work surface distance), the arc jet can depress the weld-puddle surface and affect the configuration of the distribution on the surface of the weldment (Fig. 4). The effects of depressing the weld puddle can be accommodated empirically by postulating some or all of the heat from the arc to be deposited internally. However, the specification of an internal heat-generation distribution is more complex than that of a surface flux.

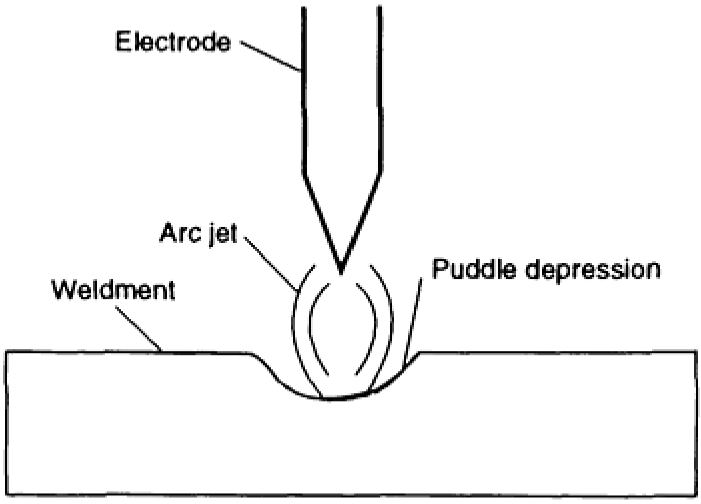

FIG. 4 EFFECT OF ARC JET ON DEPRESSION OF WELD POOL

For models in which heat from the welding arc is postulated to be deposited on the surface of the weldment, the input energy is often assumed to be a radially symmetric, normally distributed surface flux at any instant in time (Ref 13). The incident heat flux, *q*, can be expressed conveniently as an exponential, or Gaussian, function such that:

$$Q\ (R) = (3Q/\pi R'^2)\ \mathrm{EXP}[-\ 3(R/R')^2] \qquad \textbf{(EQ 1)}$$

where Q is the magnitude of the heat input per unit time, r is the distance from the center of the heat source to the surface, and r' is a characteristic radial dimensional distribution parameter that defines the region in which 95% of the heat flux is deposited. In the limit as $r' \rightarrow 0$, the heat input becomes a point source of strength Q on the surface. The applicability of point sources to characterize energy input for laser-beam welding processes will be explored later in this article.

Equation 1 describes the heat-input distribution from a stationary arc. Only the magnitude of the heat input, the distribution parameter, and the duration of heating are necessary to fully characterize the input energy. Figure 5 shows typical variations of the weld bead depth and width for stationary spot welds on 6.5 mm (0.25 in.) thick plates with both the distribution parameter, *r*', and the duration of heating, *t**.

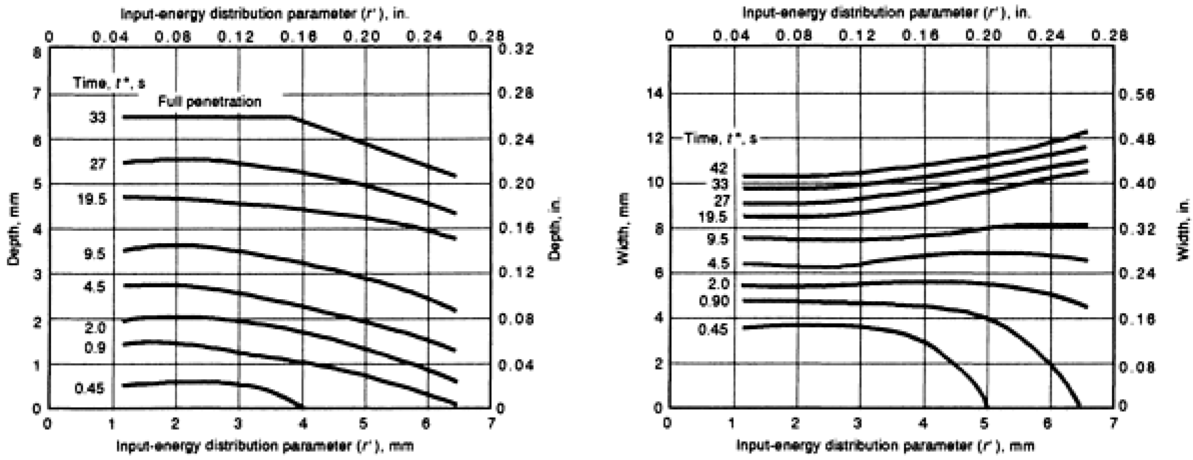

FIG. 5 WELD-BEAD DIMENSIONS FOR DIFFERENT DURATIONS OF HEATING, *T, AND INPUT-ENERGY DISTRIBUTION PARAMETER, *R*'. Q_O = 1060 W. SOURCE: REF 10**

Other forms of the heat-input distribution also can be used. These include ramp, or triangular, distributions; uniform heat-input distributions over finite areas (for example, the area of the weld puddle surface); or combinations of the two (that is, trapezoidal distributions).

The axisymmetric characterization of heat input for stationary arcs enables temperatures in the weldment to be calculated using axisymmetric thermal analysis methods, if the weldment geometry justifies such an analysis. On the other hand, if the heat source is moving, then the resultant flow of heat is three-dimensional, that is, temperature gradients develop through the thickness of the weldment, as well as in directions parallel to and transverse to the welding direction. If the weldment geometry is regular (for example, flat plate), then the heat source from the arc is moving at a constant, sufficiently high speed, *v*, and if end effects such as the starting and stopping of the arc are neglected, then transient two-dimensional thermal analyses at cross sections normal to the welding direction may be satisfactory (Ref 14). The heat input at a section transverse to the direction of arc travel is given by:

$$Q\,(X,T) = (3Q/\pi R'^2)\ \mathrm{EXP}[-3(X/R')^2]\ \mathrm{EXP}[-3(VT/R')^2] \qquad \textbf{(EQ 2)}$$

where x is the distance from the centerline of the weld and t is the time measured from the moment the center of the heat source passes over the weld section being analyzed. The three heat-input parameters characterizing a moving arc (Q,r',v) are embodied in this formulation. The inverse of the speed of a moving arc is akin to the duration, t^*, of a stationary arc, because a moving arc supplies a heat input, Q/v, per unit length of weld, whereas a stationary arc provides a total amount of heat, Qt^*, to the weldment.

For models in which heat is deposited internally, rather than as a surface flux, the input energy can be assumed to be deposited within a spheroidal or ellipsoidal region that is similar in shape to the weld puddle. As described by Ref 15, the heat input can be taken to vary exponentially as a Gaussian distribution in all directions within the ellipsoid. Thus, the energy-input distribution is completely specified by the dimensions of the ellipsoid. More elaborate models can be formulated by prescribing either double-ellipsoid heat-input regions that better match weld-puddle shape or conical regions of internal energy deposition. Regardless of the methodology chosen to characterize the distribution of the heat input, the geometric parameters of the distribution should be established, at this stage of computer-modeling development, by correlating calculated thermal-response characteristics, such as the weld bead and heat-affected zone dimensions, with measured values.

Gas-metal arc welding (GMAW) modeling is more complex than that of the GTAW process, because of the need to consider thermal energy that is transferred to the workpiece not only by mechanisms similar to those associated with the GTAW process, but, more importantly, by heat contained in the molten metal drops that are transferred as filler metal from the consumable electrode wire to the workpiece. Refer to Ref 16 for additional details. Because the GMAW process utilizes direct current electrode positive, the heat associated with the kinetic energy of the electrons and the heat of condensation is now input to the electrode. The energy previously input to the weldment is replaced by the energy of positively charged ions, which contribute a much smaller amount of heat to the weldment.

In addition, the radiation and conduction via the arc plasma make minor contributions to the heat input to the weldment. The more-important source of heat is derived from the mass of molten material and is related directly to both the temperature and the melting rate of the electrode wire. The melting rate of the electrode is governed primarily by joule heating and by arc heating similar to that of the GTAW process. The transfer of material to the workpiece can be globular, such that the metal is deposited in droplets, the diameters of which are greater than that of the electrode, or in a spray, in which mass is transferred to the weld puddle as droplets with diameters that are smaller than that of the electrode.

The thermal energy applied to the workpiece therefore consists of two sources: the arc-energy distribution generated at the cathode area (workpiece) and the thermal energy of the metal drops transferred from the filler wire. The arc input energy has a significant effect on the width of the weld puddle and, thus, of the solidified weld bead, whereas the energy in the molten metal contributes significantly to the melting of the workpiece (Ref 17). Furthermore, the impingement of the metal drops on the puddle significantly affects the depth of penetration (Ref 18), because the impact of each drop causes a distinct indentation in the weld puddle, especially at high currents.

Modeling of the heat input from the GMAW process should generally include the effects of heat input from the filler-metal droplets, the addition of new weld material from the filler wire, and, in many cases, the significant indentation of the weld-puddle surface, as well as a spatially distributed heat flux on the surface to represent radiation energy from the arc and positive ion impingement. Accurate representations of these phenomena are so complex that alternate treatments are essential. In Ref 19, the heat input from the GMAW process was modeled as a Gaussian distribution and accounted for the energy necessary to melt the electrode wire. In Ref 20, a model in which heat is input partly in the weld puddle and partly as a surface flux was developed. A more common approach is to model the addition of filler metal by "creating" new metal in an analytical model.

In finite-element treatments, for example, new metal is simulated by adding sets of elements to the model. For three-dimensional models, these elements can be introduced in distinct increments of time, whereas for a two-dimensional analysis of a cross section normal to the welding direction, a set of elements corresponding to a weld bead cross section can be created. More recently, the contour outlining the weld bead has been determined experimentally and then used as the starting boundary condition for the thermal modeling of welds. As an example, the boundary condition can be a time-dependent temperature that may reach or exceed the melting temperature of the material.

Other Arc-Welding Processes. Shielded metal arc welding (SMAW), submerged arc welding (SAW), and flux-cored arc welding (FCAW) are other processes in which energy is transferred to the workpiece both from the arc and from an overheated consumable electrode. The previously outlined semiempirical methodology for simulating the heat input for the GMAW process also appears to be well suited for the flux welding processes, although the introduction of a slag layer, whether it develops from the melting and solidification of electrode-covering material or from a flux feeding tube, presents another factor to be considered when calculating weld-induced temperatures.

References cited in this section

8. P.A. SCHOECK, AN INVESTIGATION OF THE ANODE ENERGY BALANCE OF HIGH INTENSITY ARCS IN ARGON, *MODERN DEVELOPMENTS IN HEAT TRANSFER,* ACADEMIC PRESS, 1963, P 353-400
9. M.B.C. QUIGLEY, P.H. RICHARDS, D.T. SWIFT-HOOK, AND A.E.F. GICK, HEAT FLOW TO THE WORKPIECE FROM A TIG WELDING ARC, *J. PHYS. D: APPL. PHYS.*, VOL 6, DEC 1973, P 2250-2259
10. E. FRIEDMAN AND S.S. GLICKSTEIN, AN INVESTIGATION OF THE THERMAL RESPONSE OF STATIONARY GAS TUNGSTEN-ARC WELDS, *WELD. J. RES. SUPPL.*, VOL 55 (NO. 12), DEC 1976, P 408S-420S

11. R.A. CHIHOSKI, THE RATIONING OF POWER BETWEEN THE GAS TUNGSTEN ARC AND ELECTRODE, *WELD. J. RES. SUPPL.*, VOL 49 (NO. 2), FEB 1970, P 69S-82S
12. J.F. KEY, ANODE/CATHODE GEOMETRY AND SHIELDING GAS INTERRELATIONSHIPS IN GTAW, *WELD. J. RES. SUPPL.*, DEC 1980, P 364-S
13. V. PAVELIC, R. TANBAKUCHI, O.A. UYEHARA, AND P.S. MYERS, EXPERIMENTAL AND COMPUTED TEMPERATURE HISTORIES IN GAS TUNGSTEN-ARC WELDING OF THIN PLATES, *WELD. J. RES. SUPPL.*, VOL 48 (NO. 7), 1969, P 295S-305S
14. E. FRIEDMAN, THERMOMECHANICAL ANALYSIS OF THE WELDING PROCESS USING THE FINITE ELEMENT METHOD, *J. PRESSURE VESSEL TECHNOL. (TRANS. ASME)*, VOL 97, AUG 1975, P 206-213
15. J. GOLDAK *ET AL.*, COMPUTER MODELING OF HEAT FLOW IN WELDS, *METALL. TRANS. B*, VOL 17, SEPT 1986, P 587-600
16. J.H. WASZINK AND G.J.P.M. VAN DEN HEUVAL, HEAT GENERATION AND HEAT FLOW IN THE FILLER METAL IN GMA WELDING, *WELD. J. RES. SUPPL.*, VOL 61, AUG 1982, P 269S-282S
17. W.G. ESSERS AND R. WALTER, HEAT TRANSFER AND PENETRATION MECHANISMS WITH GMA AND PLASMA-GMA WELDING, *WELD. J. RES. SUPPL.*, VOL 60, FEB 1981, P 37S-42S
18. K.C. TSAO AND C.S. WU, FLUID FLOW AND HEAT TRANSFER IN GMA WELD POOLS, *WELD. J. RES. SUPPL.*, VOL 67, MAR 1988, P 70S-75S
19. P. TEKRIWAL AND J. MAZUMDER, FINITE ELEMENT ANALYSIS OF THREE-DIMENSIONAL TRANSIENT HEAT TRANSFER IN GMA WELDING, *WELD. J. RES. SUPPL.*, VOL 67 (NO. 7), 1988, P 150S-156S
20. B.L. JOSEFSON, RESIDUAL STRESSES AND THEIR REDISTRIBUTION DURING ANNEALING OF A GIRTH-BUTT WELDED THIN-WALLED PIPE, *J. PRESSURE VESSEL TECHNOL. (TRANS. ASME)*, VOL 104, AUG 1982, P 245-250

Characterization and Modeling of the Heat Source

S.S. Glickstein and E. Friedman, Westinghouse Electric Corporation

High-Energy-Density Welding

In modeling the heat source for high-energy-density welding, one first needs to determine the type of welding process being formulated. Depending on the weld parameters, high-energy-density welding can simulate either a conduction-mode weld process or a keyhole weld process. If a strict conduction-mode weld process is being modeled, then the energy density will be low enough to prevent intense vaporization of the material. All of the energy of the beam is deposited directly on the surface of the weldment as a heat flux, similar to that discussed in terms of the GTAW process. The major difference between this mode of welding and that of arc welding is in the magnitude and distribution of the heat source. In laser-beam welding, the input-energy distribution is very dependent on the operational mode of the laser system.

Figure 6 shows several shapes of the source distribution that can be present at any one time. The fundamental transverse electromagnetic mode, TEM_{00} (a Gaussian distribution), is often selected as the heat-input description for conduction welding. It is also assumed to be the source distribution for electron-beam welding. The exact determination of the Gaussian width parameter, r', depends on the optics of the laser-or electron-beam welding system. In many cases, a point source on the surface is an excellent approximation. In laser-beam welding, if the energy is delivered via a fiber-optic system, then the output source distribution is often described as a truncated Gaussian source distribution of specified width. However, the beam exiting a fiber-optic system has been examined (Ref 21), and the output-energy distribution is reported to depend on the incident angle of the beam into the fiber. A variety of output configurations can be obtained as a function of the angle.

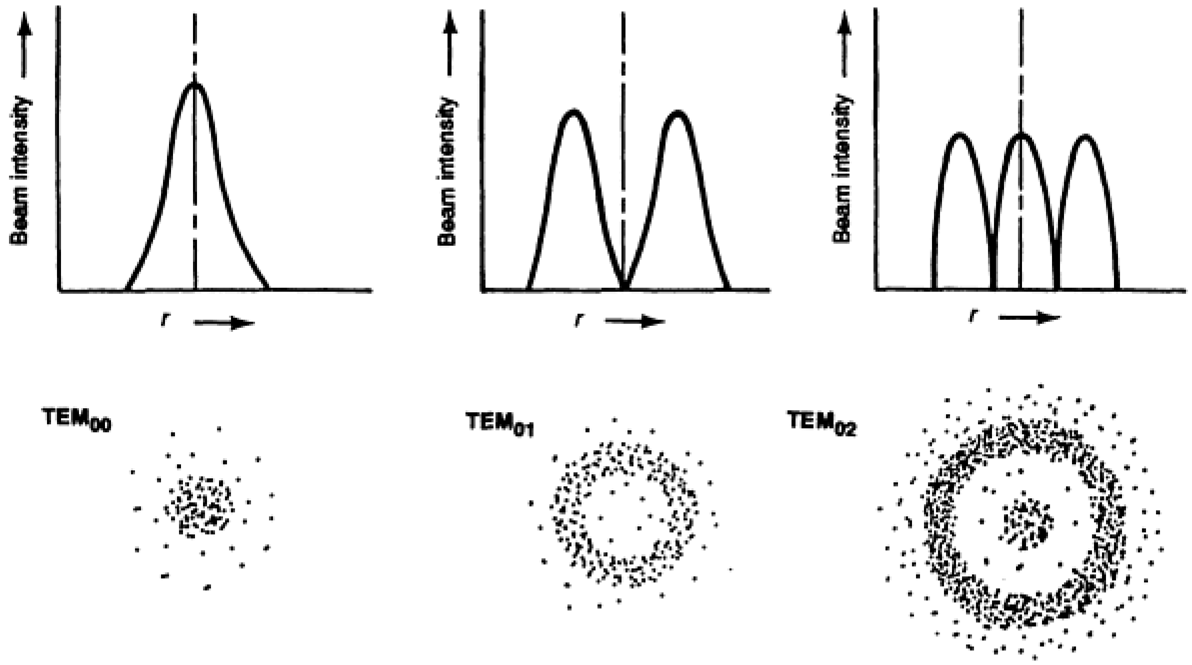

FIG. 6 APPEARANCE OF LASER BEAM FOR THREE DIFFERENT TRANSVERSE ELECTROMAGNETIC MODES (TEM). HORIZONTAL AXIS SHOWS INCREASING DISTANCE FROM THE CENTER OF THE HEAT SOURCE TO THE SURFACE.

Some laser systems, such as the multikilowatt carbon-dioxide laser, operate with the TEM_{01} as the dominant mode. This donut shape (Fig. 6) has been used very effectively for specific welding operations. Other laser systems may display a combination of these modes (for example, TEM_{02}), as well as higher-order modes. Beam analyzers are currently being employed to measure the precise distribution of the laser input energy to the weldment.

If a keyhole mode of welding is being formulated, then modeling of the input-energy distribution is complicated. Several different formulations have been discussed in the literature. In the keyhole mode, the energy density is sufficiently high to cause material ablation near the center of the beam, resulting in beam penetration further into the weldment. A hole created in the maternal is maintained by equilibrium between vapor pressure in the keyhole, surface tension, and hydrodynamic pressure in the surrounding melt. The exact distribution of energy deposited in the hole is complicated and depends on the scattering and absorption of the beam through the weldment, as well as on the focal parameters of the system. Several different models have been used to simulate these conditions.

The simplest model for the keyhole mode of welding is that of a moving line source through the weldment (Ref 22). The appropriate solution of the heat equation was obtained by Rosenthal (Ref 23). The model does not describe the typically observed semicircular part of the weld cross section at the top, because it corresponds to a liquid region, the size of which does not vary through the thickness of the workpiece. This simple model has been found to be very useful, however, in providing relations between such quantities as the power absorbed and the width of the weld.

A thermal analysis of the laser process has been derived for a Gaussian source moving at a constant velocity over a large range of conditions, from simple heat treating to deep-penetration welding (Ref 24). To obtain a more realistic weld shape, the simple line-source approach was modified (Ref 25) by adding a point source close to the surface of the workpiece to the line source distributed uniformly through the thickness. The line-source models uniform absorption of heat by the workpiece with depth, whereas the point source corresponds to a much more concentrated region of absorption in the vicinity of the laser focus, which contrasts with the absorption from multiple reflections that occur in the main part of the keyhole. The temperature distribution from this combined-source model yields a vertical weld seam shape that reflects the general shape of that found from experimental investigations (Fig. 7).

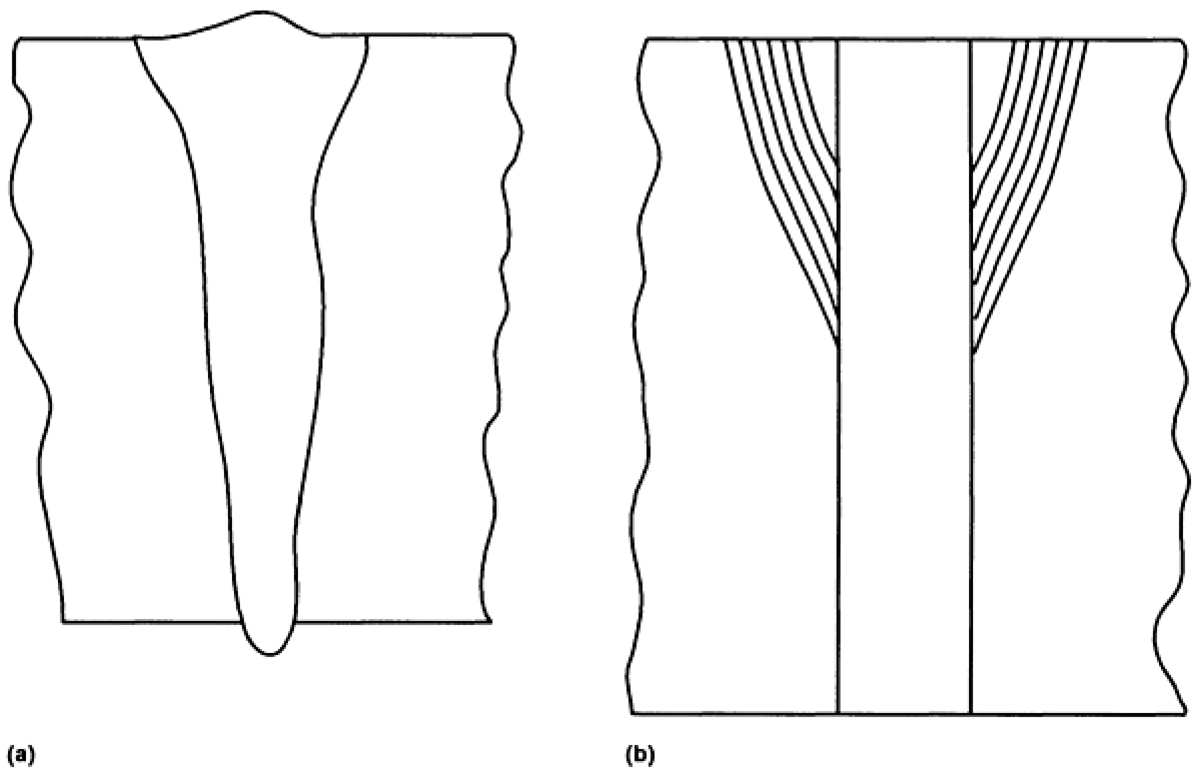

FIG. 7 OUTLINES OF CROSS SECTIONS OF (A) EXPERIMENTAL AND (B) CALCULATED KEYHOLE WELDS. CALCULATED RESULTS SHOWN FOR SEVERAL DIFFERENT POINT SOURCES ADDED TO THE LINE SOURCE. SOURCE: REF 25

A simple description of the input energy avoids the problem of defining the exact nature of the energy-transfer process in the keyhole itself. A more sophisticated model is necessary to characterize this effect. The attenuation of the laser beam in the vertical direction has been accounted for (Ref 26) by calculating the temperature at a depth, *d*, below the surface. If the temperature at depth, *d*, exceeds the boiling point of the material, then point *d* is deemed to be transparent. The incident power is then applied at a point below it, after suffering some absorption. This sequence continues through the substrate. The transparent grid points keep their high temperatures as if the keyhole were filled with a hot plasma.

This brief overview of several of the approaches used to model the high-energy-density welding process points out the difficulty in defining a unique methodology. Experimental data must be used for specific welding conditions to establish the overall approach, as well as the magnitude and distribution of the input source energy.

References cited in this section

21. Y. LIU AND K.H. LEONG, LASER BEAM DIAGNOSTICS FOR KILOWATT POWER PULSED YAG LASER, *PROC. LASER MATERIALS PROCESSING,* INTERNATIONAL CONGRESS ON APPLICATION OF LASERS AND ELECTRO-OPTICS '92, VOL 72, LASER INSTITUTE OF AMERICA, 1992
22. D.T. SWIFT-HOOK AND A.E.F. GICK, PENETRATION WELDING WITH LASERS--ANALYTICAL STUDY INDICATES THAT PRESENT LASER BEAM WELDING CAPABILITIES MAY BE EXTENDED TENFOLD, *WELD. J. RES. SUPPL.,* VOL 52, 1973, P 492S-499S
23. D. ROSENTHAL, MATHEMATICAL THEORY OF HEAT DISTRIBUTION DURING WELDING AND CUTTING, *WELD. J. RES. SUPPL.,* VOL 20, 1941, P 220S-234S
24. H.E. CLINE AND T.R. ANTHONY, HEAT TREATING AND MELTING MATERIAL WITH A SCANNING LASER OR ELECTRON BEAM, *J. APPL. PHYS.,* VOL 48 (NO. 9), 1977, P 3895

25. W.M. STEEN, J. DOWDEN, M. DAVIS, AND P. KAPADIA, A POINT AND LINE SOURCE MODEL OF LASER KEYHOLE WELDING, *J. PHYS. D: APPL. PHYS.*, VOL 21, 1988, P 1255-1260
26. J. MAZUMDER AND W.M. STEEN, HEAT TRANSFER MODEL FOR CW LASER MATERIAL PROCESSING, *J. APPL. PHYS.*, VOL 51 (NO. 2), 1980, P 941

Characterization and Modeling of the Heat Source

S.S. Glickstein and E. Friedman, Westinghouse Electric Corporation

Resistance Spot Welding

The electrical resistance spot welding process for joining two materials at their common interface involves a complicated interaction of electrical, thermal, mechanical, metallurgical, and surface phenomena. Figure 8, a simplified representation of the process, shows some essential features for producing a weld.

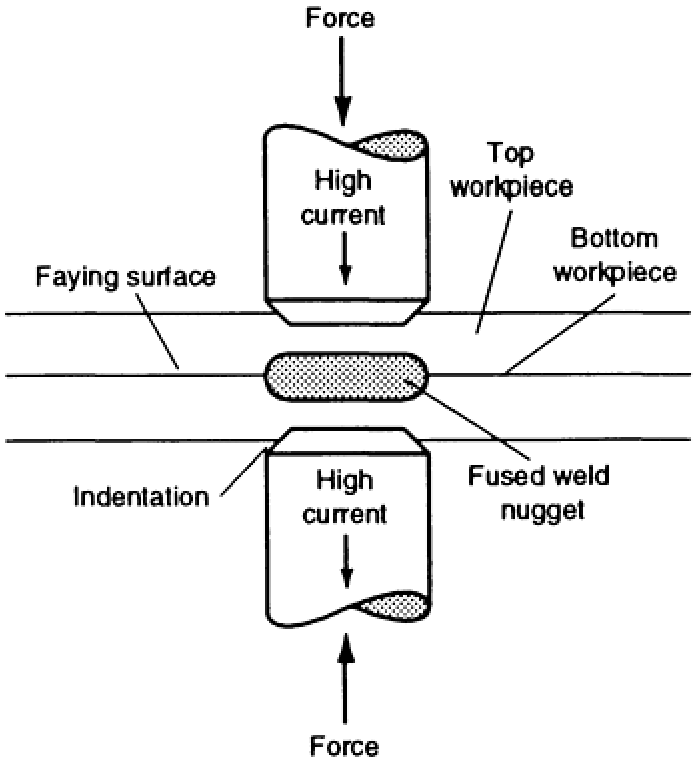

FIG. 8 RESISTANCE SPOT WELDING PROCESS

In this welding process, the faying surfaces of two or more workpieces are fused and joined at a spot by electric current flowing through the weldment (joule heating). The weldment is actually held together by the compressive force exerted by the electrodes. Because the input energy in this weld process is due to joule heating, the spatial distribution of the current and the electrical resistance within the workpiece must be determined when modeling the heat source. Most importantly, the contact resistance at the surfaces of both the electrode and the work-piece, as well as the interfaces of the work-pieces, must be established. The electrical contact resistance depends on the surface condition of the material, its hardness, the degree of oxidation, the amount and kind of impurities, as well as the apparent pressure at the surfaces and the temperature. The fact that these quantities are time dependent is another complication, because the joule heat source must be considered as a time-dependent parameter.

A detailed description of the process is presented in the article "Resistance Spot Welding" in this Volume. Literature reviews of the resistance spot welding process simulation are provided in Ref 27 and 28. The brief overview provided below of some important findings in the literature on resistance welding highlights several points that need to be considered when developing a model of the heat source.

In 1958, experimental and analytical investigations of the resistance spot welding process were performed (Ref 29). The researchers reported current-density singularities at the outer rim of the contact area from theory and correlated this phenomenon with experimental results that showed heat concentration at the periphery. They concluded that the bulk of the material near the contact region was not heated appreciably by the flow of current through it, but was heated indirectly by conduction from the peripheral region of the contact area. However, further study indicated that the contact area plays a major role only in the early stages of heat production and becomes less influential in later stages of weld-nugget formation. This fact reinforces the earlier remarks that the modeling of the heat source must consider not only the spatial distribution of the heat source, but its time dependence.

Further reinforcing this idea, it was found (Ref 30) that in the beginning of the welding process, most of the voltage drop occurred at the interface between the electrode and the weldment and in the interface of the two workpieces (21% and 66%, respectively). The researchers determined that after six cycles from the instant that weldment fusion begins, most of the voltage drop takes place in the base metal (87%). This occurs because the interfacial resistance decreases as the temperature rises, which is due to melting at the interface, and because the resistivity of the base metal increases with increased temperature.

Complicating the modeling process is the unknown, but very important, contact diameter at the faying surface. This parameter depends on the geometry of the electrodes, the surface conditions, and, most importantly, the mechanical load applied by the electrode. Recent work (Ref 28) in modeling this process uses the diameter of the electrode face as the contact diameter between the electrode and the workpiece, but a slightly larger diameter is used to represent the contact diameter at the interface of the work-pieces. Figures 9 and 10 show the current density and joule heating as functions of the nondimensional radius R/R_c, where R_c is the assumed contact radius.

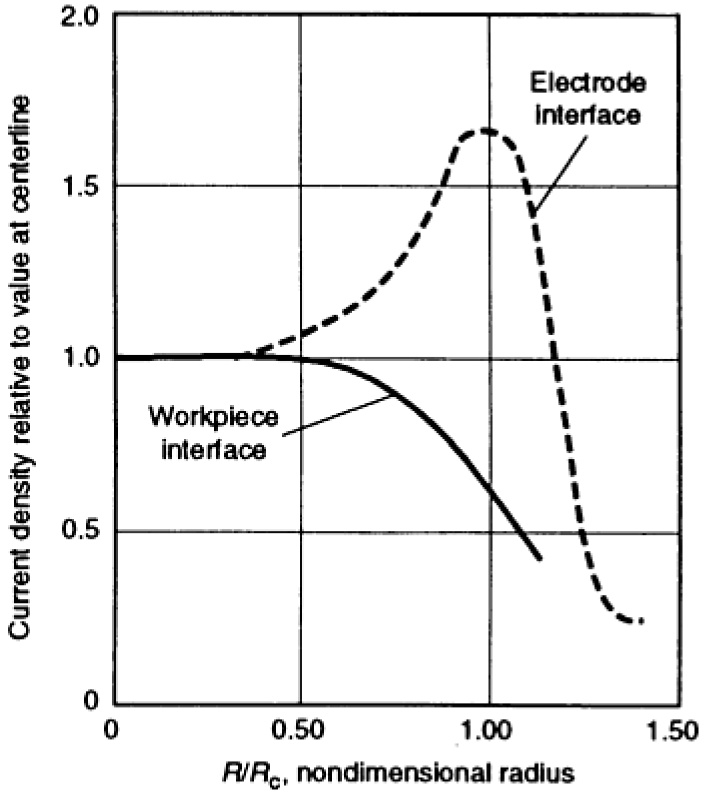

FIG. 9 CURRENT DENSITY RATIO VERSUS NONDIMENSIONAL RADIUS. TIME, 2 CYCLES; R_c, RADIUS OF CONTACT. SOURCE: REF 27

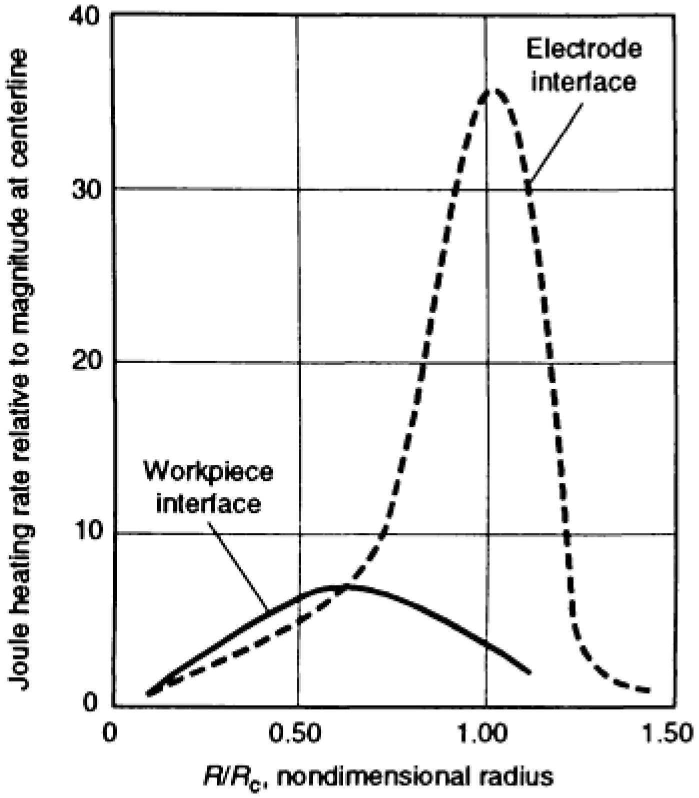

FIG. 10 JOULE HEATING RATE RATIO VERSUS NONDIMENSIONAL RADIUS. TIME, 2 CYCLES; R_C, RADIUS OF CONTACT. SOURCE: REF 27

The authors of both Ref 27 and 28 stress the importance of developing a thermomechanical coupling model that will account for changes in resistance and current distributions during the welding process. This article has tried to point out that the input description of the heat source for resistance welding is extremely complicated. It depends on the electrode geometry, the weldment material, and the weldment surface condition, as well as the weld parameters (including the mechanical forces on the electrode) that must be considered as being time dependent.

References cited in this section

27. H.A. NIED, THE FINITE ELEMENT MODELING OF THE RESISTANCE SPOT WELDING PROCESS, *WELD. J. RES. SUPPL.*, APRIL 1984, P 123-S
28. C.L. TSAI, O.A. JAMMAL, J.C. PAPRITAN, AND D.W. DICKINSON, MODELING OF RESISTANCE SPOT WELD NUGGET GROWTH, *WELD. J. RES. SUPPL.*, FEB 1992, P 47-S
29. J.A. GREENWOOD AND J.B.P. WILLIAMSON, ELECTRICAL CONDUCTION IN SOLIDS--II. THEORY OF TEMPERATURE-DEPENDENT CONDUCTORS, *PROC. ROYAL SOC. LONDON*, VOL 246, 1958, P 13-31
30. H.S. CHO AND Y.J. CHO, A STUDY OF THE THERMAL BEHAVIOR IN RESISTANCE SPOT WELDING, *WELD J. RES. SUPPL.*, JUNE 1989, P 236-S TO 244-S

Characterization and Modeling of the Heat Source

S.S. Glickstein and E. Friedman, Westinghouse Electric Corporation

References

1. P.S. MYERS, O.A. UYEHARA, AND G.L. BORMAN, FUNDAMENTALS OF HEAT FLOW IN WELDING, *WELD. RES. COUNC. BULL.*, NO. 123, JULY 1967
2. H.S. CARSLAW AND J.C. JAEGER, *CONDUCTION OF HEAT IN SOLIDS*, CLARENDON PRESS, LONDON, 1959
3. D. ROSENTHAL AND R. SCHMERBER, THERMAL STUDY OF ARC WELDING, *WELD J. RES. SUPPL.*, VOL 17 (NO. 4), 1938, P 2S-8S
4. D. ROSENTHAL, THE THEORY OF MOVING SOURCES OF HEAT AND ITS APPLICATION TO METAL TREATMENTS, *TRANS. ASME.*, VOL 68, 1946, P 849-866
5. W. SOEDEL AND R. COHEN, ARC-WELDING TEMPERATURES IN A CIRCULAR DISK STRUCTURE, *WELD. J. RES. SUPPL.*, VOL 49 (NO. 7), 1970, P 337S-340S
6. T. NAKA AND K. MASUBUCHI, TEMPERATURE DISTRIBUTION OF WELDED PLATES, *J. JPN. WELD. SOC.*, VOL 16 (NO. 7 AND 12), 1947, P 281-290 (PART 1) AND P 374-378 (PART II)
7. E.F. RYBICKI *ET AL.*, A FINITE-ELEMENT MODEL FOR RESIDUAL STRESSES AND DEFLECTIONS IN GIRTH-BUTT WELDED PIPES, *J. PRESSURE VESSEL TECHNOL. (TRANS. ASME)*, VOL 100, AUG 1978, P 256-262
8. P.A. SCHOECK, AN INVESTIGATION OF THE ANODE ENERGY BALANCE OF HIGH INTENSITY ARCS IN ARGON, *MODERN DEVELOPMENTS IN HEAT TRANSFER*, ACADEMIC PRESS, 1963, P 353-400
9. M.B.C. QUIGLEY, P.H. RICHARDS, D.T. SWIFT-HOOK, AND A.E.F. GICK, HEAT FLOW TO THE WORKPIECE FROM A TIG WELDING ARC, *J. PHYS. D: APPL. PHYS.*, VOL 6, DEC 1973, P 2250-2259
10. E. FRIEDMAN AND S.S. GLICKSTEIN, AN INVESTIGATION OF THE THERMAL RESPONSE OF STATIONARY GAS TUNGSTEN-ARC WELDS, *WELD. J. RES. SUPPL.*, VOL 55 (NO. 12), DEC 1976, P 408S-420S
11. R.A. CHIHOSKI, THE RATIONING OF POWER BETWEEN THE GAS TUNGSTEN ARC AND ELECTRODE, *WELD. J. RES. SUPPL.*, VOL 49 (NO. 2), FEB 1970, P 69S-82S
12. J.F. KEY, ANODE/CATHODE GEOMETRY AND SHIELDING GAS INTERRELATIONSHIPS IN GTAW, *WELD. J. RES. SUPPL.*, DEC 1980, P 364-S
13. V. PAVELIC, R. TANBAKUCHI, O.A. UYEHARA, AND P.S. MYERS, EXPERIMENTAL AND COMPUTED TEMPERATURE HISTORIES IN GAS TUNGSTEN-ARC WELDING OF THIN PLATES, *WELD. J. RES. SUPPL.*, VOL 48 (NO. 7), 1969, P 295S-305S
14. E. FRIEDMAN, THERMOMECHANICAL ANALYSIS OF THE WELDING PROCESS USING THE FINITE ELEMENT METHOD, *J. PRESSURE VESSEL TECHNOL. (TRANS. ASME)*, VOL 97, AUG 1975, P 206-213
15. J. GOLDAK *ET AL.*, COMPUTER MODELING OF HEAT FLOW IN WELDS, *METALL. TRANS. B*, VOL 17, SEPT 1986, P 587-600
16. J.H. WASZINK AND G.J.P.M. VAN DEN HEUVAL, HEAT GENERATION AND HEAT FLOW IN THE FILLER METAL IN GMA WELDING, *WELD. J. RES. SUPPL.*, VOL 61, AUG 1982, P 269S-282S
17. W.G. ESSERS AND R. WALTER, HEAT TRANSFER AND PENETRATION MECHANISMS WITH GMA AND PLASMA-GMA WELDING, *WELD. J. RES. SUPPL.*, VOL 60, FEB 1981, P 37S-42S
18. K.C. TSAO AND C.S. WU, FLUID FLOW AND HEAT TRANSFER IN GMA WELD POOLS, *WELD. J. RES. SUPPL.*, VOL 67, MAR 1988, P 70S-75S
19. P. TEKRIWAL AND J. MAZUMDER, FINITE ELEMENT ANALYSIS OF THREE-DIMENSIONAL TRANSIENT HEAT TRANSFER IN GMA WELDING, *WELD. J. RES. SUPPL.*, VOL 67 (NO. 7), 1988, P 150S-156S
20. B.L. JOSEFSON, RESIDUAL STRESSES AND THEIR REDISTRIBUTION DURING ANNEALING OF A GIRTH-BUTT WELDED THIN-WALLED PIPE, *J. PRESSURE VESSEL TECHNOL. (TRANS. ASME)*, VOL 104, AUG 1982, P 245-250
21. Y. LIU AND K.H. LEONG, LASER BEAM DIAGNOSTICS FOR KILOWATT POWER PULSED YAG LASER, *PROC. LASER MATERIALS PROCESSING*, INTERNATIONAL CONGRESS ON

APPLICATION OF LASERS AND ELECTRO-OPTICS '92, VOL 72, LASER INSTITUTE OF AMERICA, 1992

22. D.T. SWIFT-HOOK AND A.E.F. GICK, PENETRATION WELDING WITH LASERS--ANALYTICAL STUDY INDICATES THAT PRESENT LASER BEAM WELDING CAPABILITIES MAY BE EXTENDED TENFOLD, *WELD. J. RES. SUPPL.,* VOL 52, 1973, P 492S-499S
23. D. ROSENTHAL, MATHEMATICAL THEORY OF HEAT DISTRIBUTION DURING WELDING AND CUTTING, *WELD. J. RES. SUPPL.,* VOL 20, 1941, P 220S-234S
24. H.E. CLINE AND T.R. ANTHONY, HEAT TREATING AND MELTING MATERIAL WITH A SCANNING LASER OR ELECTRON BEAM, *J. APPL. PHYS.,* VOL 48 (NO. 9), 1977, P 3895
25. W.M. STEEN, J. DOWDEN, M. DAVIS, AND P. KAPADIA, A POINT AND LINE SOURCE MODEL OF LASER KEYHOLE WELDING, *J. PHYS. D: APPL. PHYS.,* VOL 21, 1988, P 1255-1260
26. J. MAZUMDER AND W.M. STEEN, HEAT TRANSFER MODEL FOR CW LASER MATERIAL PROCESSING, *J. APPL. PHYS.,* VOL 51 (NO. 2), 1980, P 941
27. H.A. NIED, THE FINITE ELEMENT MODELING OF THE RESISTANCE SPOT WELDING PROCESS, *WELD. J. RES. SUPPL.,* APRIL 1984, P 123-S
28. C.L. TSAI, O.A. JAMMAL, J.C. PAPRITAN, AND D.W. DICKINSON, MODELING OF RESISTANCE SPOT WELD NUGGET GROWTH, *WELD. J. RES. SUPPL.,* FEB 1992, P 47-S
29. J.A. GREENWOOD AND J.B.P. WILLIAMSON, ELECTRICAL CONDUCTION IN SOLIDS--II. THEORY OF TEMPERATURE-DEPENDENT CONDUCTORS, *PROC. ROYAL SOC. LONDON,* VOL 246, 1958, P 13-31
30. H.S. CHO AND Y.J. CHO, A STUDY OF THE THERMAL BEHAVIOR IN RESISTANCE SPOT WELDING, *WELD J. RES. SUPPL.,* JUNE 1989, P 236-S TO 244-S

Validation Strategies for Heat-Affected Zone and Fluid-Flow Calculations

J. Mazumder, University of Illinois at Urbana-Champaign

Introduction

FUSION WELDING PROCESSES involve all four phase changes: solid-solid state, solid-liquid, liquid-vapor, and vapor-plasma. Attempts to model the associated heat transport and related phenomena have been ongoing for more than half a century. The temperature history and the peak temperature determine the nature and extent of the phase change. One of the earlier analytical models for heat conduction in welding developed by Rosenthal (Ref 1) is still being used by many welding practitioners to estimate the heat-affected zone (HAZ). Although the existence of convection and its effect on molten weld pools were first reported in mid-1950s (Ref 2, 3), major modeling efforts on convection in welding started after the work reported by Heiple and Roper (Ref 4) on the effect of surface-tension-driven flow in the weld pool. During the past decade, several research groups developed two- and three-dimensional models for convection in the weld pool (Ref 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15). Of course, these convection models also predict the thermal history of the weld. Another effect of thermal history is the generation of residual stress and strain due to thermal expansion and solid-state phase transformation. Due to the importance of residual stress on the mechanical performance of welded joints, several groups also became involved in modeling the stress/strain history of welds (Ref 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35).

Each model has its own sets of assumptions. Validation of a model requires careful experiments to satisfy the underlying assumptions for that particular model. For example, Rosenthal's two-dimensional model is only valid for thin-plate welding and temperature prediction (Ref 1). Prediction for the center of the heat source for his three-dimensional model is erroneous because the assumption of a point heat source will lead to infinitely high temperatures, which is physically impossible. Also, lack of physics such as latent heat of fusion, radiative and convective heat loss, convection in the weld pool, and vaporization heat loss will lead to significant error in the melt pool prediction. Therefore, the only reasonable strategy for validation of such an analytical model will be temperature measurement at the HAZ, preferably away from the liquid-solid interface.

In designing experiments for validation of any analytical or numerical models for thermal, momentum, or stress history, one should clearly be aware of the model assumptions and boundary conditions. In this article, some important techniques to validate temperature, momentum, and stress history observed in welded materials will be discussed.

References

1. D. ROSENTHAL, *WELD. J.,* VOL 20, 1941, P 220S-225S
2. P.T. HOULDCROFT, *BR. WELD. J.,* VOL 1, 1954, P 468
3. W.J. PUMPHREY, *BR. WELD. J.,* VOL 2, 1955, P 92
4. C.R. HEIPLE AND J.R. ROPER, *WELD. J.,* VOL 61, 1982, P 973-1025
5. J. MAZUMDER AND A. KAR, *THERMOMECHANICAL ASPECTS OF MANUFACTURING AND MATERIALS PROCESSING*, R.K. SHAH *ET AL.,* ED., HEMISPHERE PUBLISHING, P 283-304
6. J. MAZUMDER, *OPT. ENG.,* VOL 30 (NO. 8), 1992, P 1208-1219
7. A. PAUL AND T. DEBROY, *METALL. TRANS. B,* VOL 19, 1988, P 851-858
8. N. PIRCH, E.W. KREUTZ, L. MOLLER, A. GASSER, AND K. WISSENBACH, *PROC. 3RD EUROPEAN CONF. ON LASER TREATMENT OF MATERIALS, ECLAT '90,* VOL 1, ERLANGEN, GERMANY, 17-19 SEPT 1990, SPRECHSAAL PUBLISHING GROUP
9. R.L. ZEHR, "THERMOCAPILLARY CONVECTION IN LASER MELTED POOL DURING MATERIALS PROCESSING," PH.D. THESIS, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, 1991
10. C.L. CHAN, R. ZEHR, J. MAZUMDER, AND M.M. CHEN, *PROC. 3RD ENGINEERING FOUNDATION CONF. ON MODELING AND CONTROL CASTING AND WELDING PROCESSES,* S. KOU AND R. MEHRABIAN, ED., TMS-AIME, 1986, P 229-246
11. M.C. TSAI AND S. KOU, *INT. J. NUMER. METHODS FLUIDS,* VOL 9 (NO. 12), DEC 1989, P 1503-1516
12. M.C. TSAI AND S. KOU, *NUMER. HEAT TRANSFER A: APPLICATIONS,* VOL 17 (NO. 1), 1990, P 73-89
13. T. ZACHARIA, A.H. ERASLAN, AND D.K. AIDUN, *WELD. J.,* VOL 67 (NO. 3), MARCH 1988, P 53S-62S
14. T. ZACHARIA, A.H. ERASLAN, AND D.K. AIDUN, *WELD. J.,* VOL 67 (NO. 1), JAN 1988, P 18S-27S
15. R.T.C. CHOO, J. SZEKELY, AND R.C. WESTHOFF, *WELD. J.,* VOL 69 (NO. 9), SEPT 1990, P 346S-361S
16. L. KARLSSON, CHAPTER 5, *THERMAL STRESSES I,* R.B. HETNARSKI, ED., NORTH-HOLLAND, 1986, P 299-389
17. J. GOLDAK, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID AND J.M. VITEK, ED., ASM INTERNATIONAL, 1989, P 71-82
18. H.D. HIBBIT AND P.V. MARCAL, *COMPUTERS AND STRUCTURES,* VOL 3 (NO. 5), SEPT 1973, P 1145-1174
19. E. FRIEDMAN, *J. PRESSURE VESSEL TECHNOL. (TRANS. ASME),* VOL 97 (NO. 3), AUG 1975, P 206-213
20. B.A.B. ANDERSSON, *ENG. MATER. TECHNOL. (TRANS ASME),* VOL 100, OCT 1978, P 345-362
21. E.F. RYBICKI, D.W. SCHMUESER, R.W. STONESIFER, J.J. GROOM, AND H.W. MISHLER, *J. PRESSURE VESSEL TECHNOL. (TRANS ASME),* VOL 100, AUG 1978, P 256-262
22. E.F. RYBICKI AND R.B. STONESIFER, *J. PRESSURE VESSEL TECHNOL. (TRANS. ASME),* VOL 101, MAY 1979, P 149-154
23. Y. UEDA AND K. NAKACHO, *TRANS. JWRI,* VOL 9 (NO. 1), 1980, P 107-114
24. D.H. ALLEN AND W.E. HAISLER, *COMPUTERS AND STRUCTURES,* VOL 13, 1981, P 129-135
25. J.H. ARGYRIS, J. SZIMMAT, AND K.J. WILLIAM, *COMPUTER METHODS IN APPLIED*

MECHANICS AND ENGINEERING, VOL 33, 1982, P 635-666

26. B. ANDERSSON AND L. KARLSSON, *J. THERM. STRESSES,* VOL 4, 1981, P 491-500
27. B. PATEL, "THERMO-ELASTO-PLASTIC FINITE ELEMENT FORMULATION FOR DEFORMATION AND RESIDUAL STRESSES DUE TO WELDS," PH.D. THESIS, CARLETON UNIVERSITY, OTTAWA, ONTARIO, CANADA, APRIL 1985
28. K.W. MAHIN, S. MACEWEN, W.S. WINTERS, W. MASON, M. KANOUFF, AND E.A. FUCHS, *MODELING AND CONTROL OF CASTING AND WELDING PROCESSES IV,* PROC. INT. CONF. ON MODELING OF CASTING AND WELDING PROCESSES IV, A.F. GIAMEI AND G.J. ABBASCHIAN, ED., ENGINEERING FOUNDATION AND TMS/AIME, APRIL 1988, P 339-350
29. J. GOLDAK, A. ODDY, M. MCDILL, A. CHAKRAVARTI, M. BIBBY, AND R. HOUSE, *ADVANCES IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID, ED., ASM INTERNATIONAL, 1986, P 523-527
30. P. TEKRIWAL AND J. MAZUMDER, *ADVANCES IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID, ED., ASM INTERNATIONAL, 1986, P 71-80
31. R.I. KARLSSON AND B.L. JOSEFSON, *J. PRESSURE VESSEL TECHNOL. (TRANS. ASME),* VOL 112, FEB 1990, P 76-84
32. C.K. LEUNG, "STUDY OF FINITE ELEMENT MODELING OF MULTIPASS WELDS," PH.D. THESIS, UNIVERSITY OF WATERLOO, WATERLOO, ONTARIO, CANADA, 1986
33. S.Z. NEWMAN, "FEM MODEL OF 3D TRANSIENT TEMPERATURE AND STRESS FIELDS IN WELDED PLATES," PH.D. THESIS, CARNEGIE MELLON UNIVERSITY, PITTSBURGH, PA, APRIL 1986
34. P. TEKRIWAL, "THREE-DIMENSIONAL TRANSIENT THERMO-ELASTO-PLASTIC MODELING OF GAS METAL ARC WELDING USING THE FINITE ELEMENT METHOD," PH.D. THESIS, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, URBANA, IL, AUGUST 1989
35. P. TEKRIWAL AND J. MAZUMDER, *TRANS. ASME,* VOL 113, JULY 1991, P 336-343

Validation Strategies for Heat-Affected Zone and Fluid-Flow Calculations

J. Mazumder, University of Illinois at Urbana-Champaign

Validation of Fluid-Flow Calculation

Convection is the single most important factor influencing the geometry of the weld pool including pool shape, aspect ratio, and surface ripples, and can result in defects such as variable penetration, porosity, and lack of fusion. Convection is also primarily responsible for mixing and, as a result, also affects the composition of the weld pool.

Recently, the sophistication of fluid-flow modeling has reached a stage where most of the important physics are considered. Buoyancy, surface tension, and electromagnetic forces are the three driving forces for convection. Electromagnetic force is important for arc welding. Initial work by Orepar, Eager, and Szekely (Ref 36) showed that even for arc welding, surface-tension-driven flow is dominant until the current level is very high (exceeding 250 A). Recently, several authors published models for thermocapillary flow (surface-tension-driven flow) for welding (Ref 6, 7, 8, 9, 10, 11, 12, 13, 14, 15). For validation of fluid-flow models, free surface models are the best candidates since they incorporate the most physics of the welding process. However, until recently, measurement of free surface deformation was difficult to perform. A new technique utilizing the specular reflection of a 488 nm argon-ion laser beam from the melt pool is employed to measure the free surface deformation (Ref 37, 38). Another challenge for validation of the fluid flow model is the measurement of liquid-metal velocity during welding. Temperature measurement techniques for the weld pool were recently reviewed by Kraus (Ref 39). Important techniques for the measurement of free surface deformation, fluid velocity, and pool temperature are discussed below.

References cited in this section

6. J. MAZUMDER, *OPT. ENG.*, VOL 30 (NO. 8), 1992, P 1208-1219
7. A. PAUL AND T. DEBROY, *METALL. TRANS. B,* VOL 19, 1988, P 851-858
8. N. PIRCH, E.W. KREUTZ, L. MOLLER, A. GASSER, AND K. WISSENBACH, *PROC. 3RD EUROPEAN CONF. ON LASER TREATMENT OF MATERIALS, ECLAT '90,* VOL 1, ERLANGEN, GERMANY, 17-19 SEPT 1990, SPRECHSAAL PUBLISHING GROUP
9. R.L. ZEHR, "THERMOCAPILLARY CONVECTION IN LASER MELTED POOL DURING MATERIALS PROCESSING," PH.D. THESIS, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, 1991
10. C.L. CHAN, R. ZEHR, J. MAZUMDER, AND M.M. CHEN, *PROC. 3RD ENGINEERING FOUNDATION CONF. ON MODELING AND CONTROL CASTING AND WELDING PROCESSES,* S. KOU AND R. MEHRABIAN, ED., TMS-AIME, 1986, P 229-246
11. M.C. TSAI AND S. KOU, *INT. J. NUMER. METHODS FLUIDS,* VOL 9 (NO. 12), DEC 1989, P 1503-1516
12. M.C. TSAI AND S. KOU, *NUMER. HEAT TRANSFER A: APPLICATIONS,* VOL 17 (NO. 1), 1990, P 73-89
13. T. ZACHARIA, A.H. ERASLAN, AND D.K. AIDUN, *WELD. J.,* VOL 67 (NO. 3), MARCH 1988, P 53S-62S
14. T. ZACHARIA, A.H. ERASLAN, AND D.K. AIDUN, *WELD. J.,* VOL 67 (NO. 1), JAN 1988, P 18S-27S
15. R.T.C. CHOO, J. SZEKELY, AND R.C. WESTHOFF, *WELD. J.,* VOL 69 (NO. 9), SEPT 1990, P 346S-361S
36. G.M. OREPAR, T.W. EAGER, AND J. SZEKELY, *WELD. J.,* VOL 63, 1983, P 307S
37. J. MAZUMDER AND D. VOELKEL, *PROCEEDINGS OF LAMP '92,* A. MATSUNAWA AND KATAYAMA, ED., HIGH TEMPERATURE SOCIETY OF JAPAN, JUNE 1992, P 373-380
38. J. MAZUMDER AND D. VOELKEL, U.S. PATENT APPLICATION 08/004763, JAN 1993
39. H.G. KRAUS, *MODELING AND CONTROL OF CASTING AND WELDING PROCESSES IV,* PROC. CONF., A.F. GIAMEI AND G.J. ABBASCHIAN, ED., TMS-AIME, 1988, P 205-212

Validation Strategies for Heat-Affected Zone and Fluid-Flow Calculations

J. Mazumder, University of Illinois at Urbana-Champaign

Measurement of Free Surface Deformation

Measurement of free surface deformation during welding is extremely complicated due to melt pool characteristics such as the specular pool surface, the wide range of deformation slopes possible, vaporization of metal from the pool surface, shielding gas, radiation of many wavelengths from the hot surface and plasma, small pool size, high temperature, and the incoming processing beam. Various optical techniques have been considered, but were found inadequate for characterizing the aforementioned characteristics (Ref 40). Since none of the classic methods were appropriate, a new technique was developed based on the specularity and steepness of the surface of the weld pool.

The fundamental step of this method is to reflect light off of a specular surface and measure the slope of the surface where the reflection is observed. This slope is measured indirectly by measuring the difference between the angles of the incident light when it reflects into a camera at a location on a flat surface and the corresponding location of the melt pool. If the angle of incidence is varied over a large enough range, each portion of the pool will at some time reflect the incident light into the camera, thereby revealing its slope. This will result in a known slope for every portion of the melt pool surface. If the slope between the areas of interest and the edge of the pool is known, it can be integrated to obtain the surface elevation. This is the basic idea behind the melt pool surface contour measurement method (Ref 38, 40). Figure 1 shows a schematic of the surface contour measurement system.

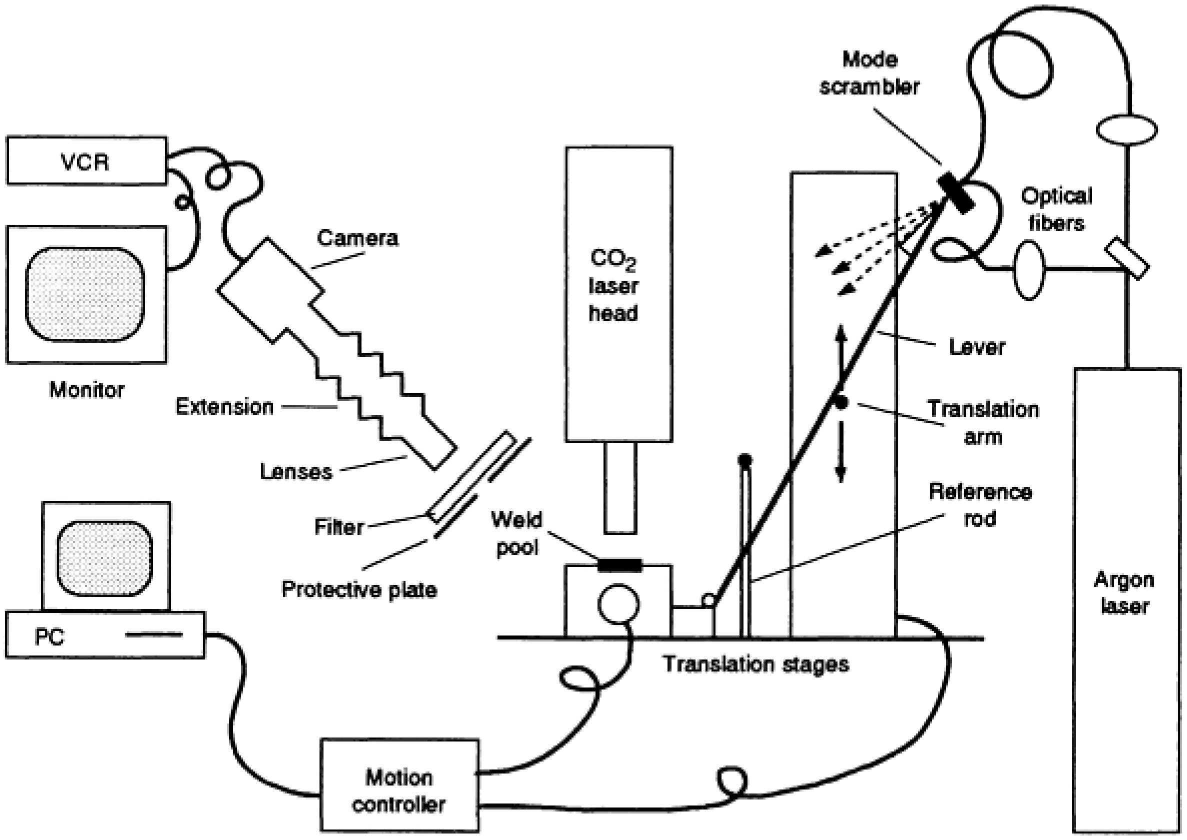
Mode scrambler
VCR
Optical fibers
Camera
CO$_2$ laser head
Monitor
Lever
Extension
Translation arm
Lenses
Reference rod
Filter
Weld pool
Protective plate
Argon laser
PC
Translation stages
Motion controller

FIG. 1 SCHEMATIC OF THE SPECULAR SURFACE CONTOUR MEASUREMENT SYSTEM

Figure 2 shows a typical laser weld pool free surface measured by the "reflective topography" technique described above. This technique has shown the capability of measuring all types of free surface shapes including convex, concave (Ref 37), and complicated shapes with a lip as shown in Fig. 2.

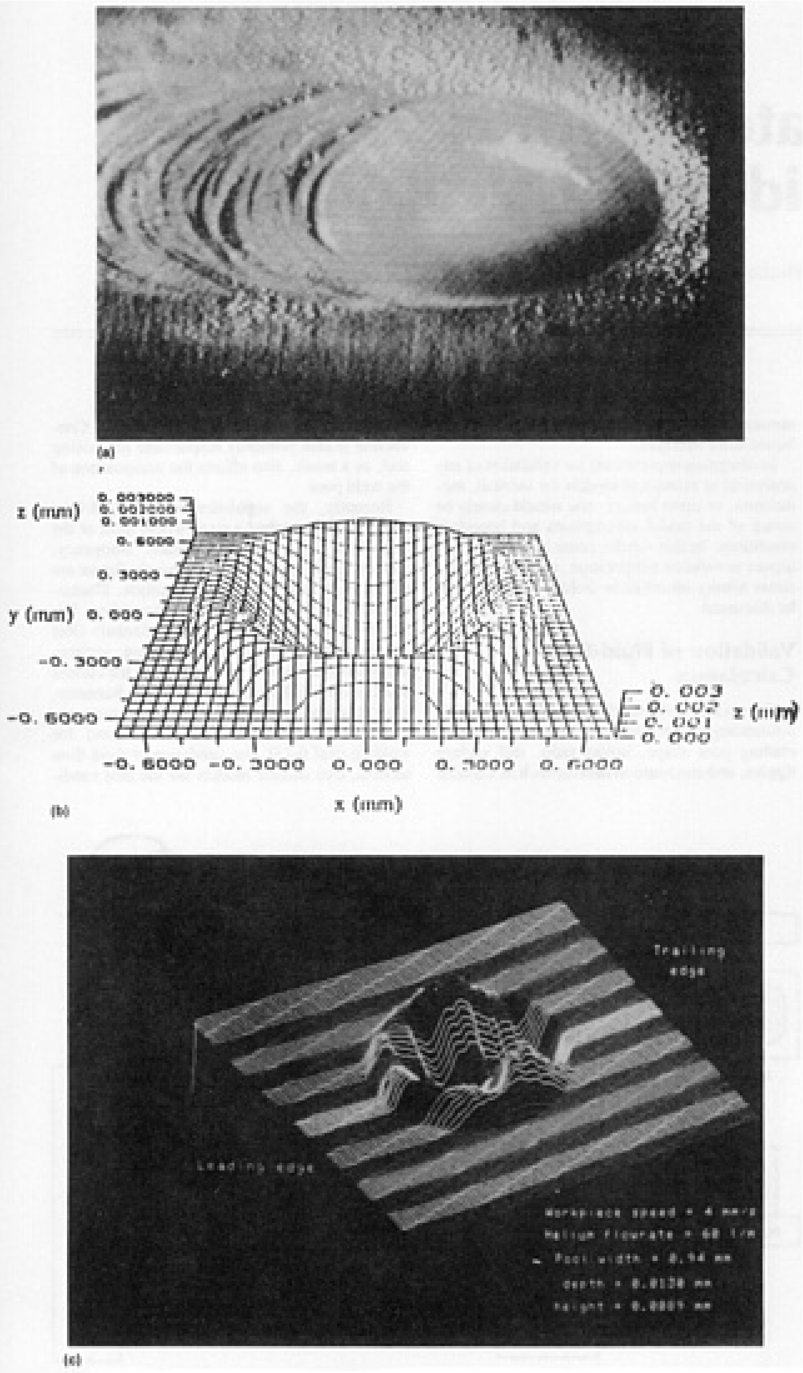

FIG. 2 LASER WELD POOL FREE SURFACE MEASURED BY THE REFLECTIVE TOPOGRAPHY TECHNIQUE. (A) VISUALIZED PICTURE. (B) PREDICTED MODEL. (C) REFLECTIVE TOPOGRAPHY IMAGE

There are several uncertainties in the modeling of free surface deformation of the weld pool. Most common for weld modeling is the paucity of temperature-dependent thermophysical property data. For the free surface deformation problem in particular, the nature of the surface tension gradient as a function of temperature and composition is poorly understood, and very little data are available. Presently, a linear dependence on temperature is assumed by most researchers. Recently, experimental measurements carried out by Sahoo *et al.* (Ref 41) and McNallan *et al.* (Ref 42), indicate a nonlinear relationship which may affect the free surface contour significantly. Also, there is controversy over the role of plasma, which is often present in welding, on the surface tension characteristics. The effect of vaporization on the pressure boundary condition also needs attention. Although significant progress has been made in modeling the melt pool free surface, there is considerable controversy over the basic assumptions for process phenomena and the physics of the process.

Various researchers have used metallographic techniques to measure the weld pool aspect ratio for the solidified weld and compared that with modeling data for model validation. This technique provides a qualitative validation of the overall model. However, comparison of the surface contour measurement of the solidified weld pool using a stylus profilometer with that of the liquid metal free surface measured by the reflective topographic technique reveal that the assumption of similar liquid and resolidified surface contours, which is sometimes suggested because the pool is expected to resolidify quickly, is not an accurate assumption for model validation (Ref 37). Therefore, for more accurate validation of the physics of the model, reflective topography accompanied with velocity measurement by x-ray radiography is more desirable.

References cited in this section

37. J. MAZUMDER AND D. VOELKEL, *PROCEEDINGS OF LAMP '92*, A. MATSUNAWA AND KATAYAMA, ED., HIGH TEMPERATURE SOCIETY OF JAPAN, JUNE 1992, P 373-380
38. J. MAZUMDER AND D. VOELKEL, U.S. PATENT APPLICATION 08/004763, JAN 1993
40. D. VOELKEL, "LASER WELD POOL OPTICAL DIAGNOSTICS: TOPOGRAPHY AND VISUALIZATION USED FOR SURFACE CONTOUR ANALYSIS AND PROCESS MONITORING," PH.D. THESIS, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, URBANA, IL, JAN 1992
41. P. SAHOO, T. DEBROY, AND M.J. MCNALLAN, *METALL. TRANS. B,* VOL 19, 1988, P 484-491
42. M.J. MCNALLAN AND T. DEBROY, *METALL. TRANS. B,* VOL 22, 1991, P 557-560

Validation Strategies for Heat-Affected Zone and Fluid-Flow Calculations

J. Mazumder, University of Illinois at Urbana-Champaign

Velocity Measurement

Heiple and Roper (Ref 4) used a high-speed movie to track the flow of Al_2O_3 particles on the weld pool surface to measure surface velocity. To validate the nature and velocity of the recirculating flow at the weld pool, x-ray radiography must be utilized. A schematic diagram of the experimental arrangement used by Mazumder *et al.* (Ref 43) is shown in Fig. 3. The substrate material used for this arc welding experiment is aluminum. Tungsten particles are used as tracers to provide contrast. A microfocused x-ray beam is directed through the test sections and an image intensifier is located on the back of the sample to receive the image. This image is processed and stored on videotape. Because of the density difference between aluminum and tungsten, the tungsten particles can be traced as they travel with the moving molten aluminum. Because the frame speed of the television camera is known, the relative position of the tungsten particles will provide the velocity of the weld pool, assuming the particles flow with the same velocity as the liquid metal (Ref 43).

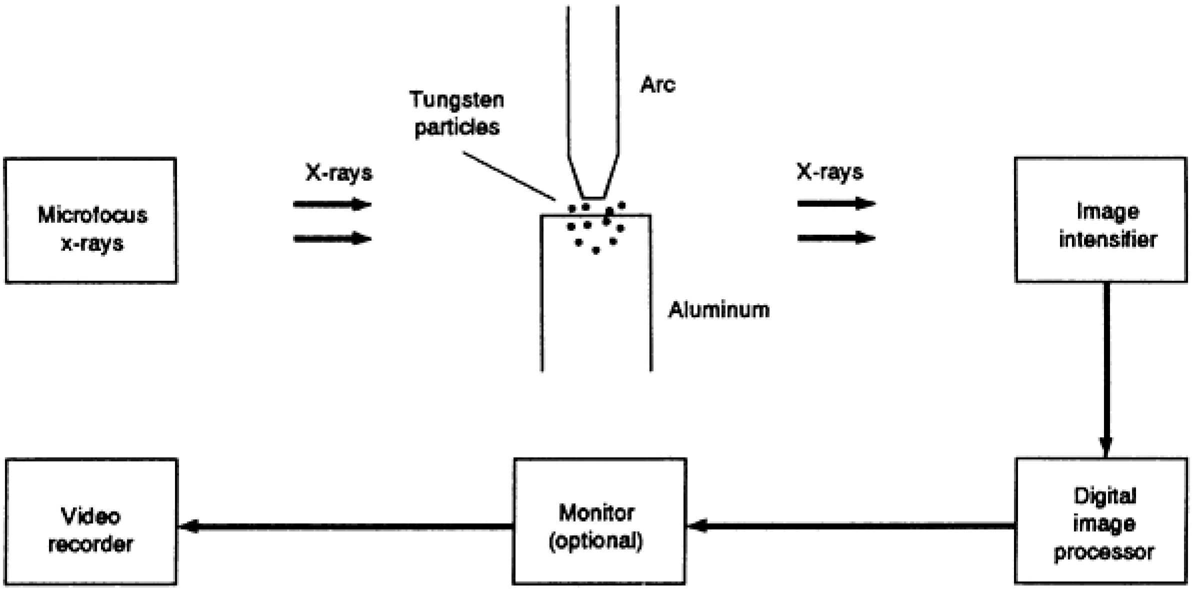

FIG. 3 SCHEMATIC OF THE X-RAY RADIOGRAPHY SYSTEM USED FOR MEASURING WELD POOL SURFACE VELOCITY

References cited in this section

4. C.R. HEIPLE AND J.R. ROPER, *WELD. J.,* VOL 61, 1982, P 973-1025

43. J. MAZUMDER, M.M. CHEN, C.L. CHAN, R. ZEHR, AND D. VOELKEL, *PROC. ECLAT '90,* H. BERGMANN, ED., SPRECHSAAL PUBLISHING GROUP, COBURG, GERMANY, 1990, P 37-53

Validation Strategies for Heat-Affected Zone and Fluid-Flow Calculations

J. Mazumder, University of Illinois at Urbana-Champaign

Temperature Measurement

Thermocouples. Temperature measurement for model validation is probably the most mature arena for model validation. An excellent review of the field has been published by Kraus (Ref 39). Most of the earlier work for temperature measurement for welding was with thermocouples. Sundell *et al.* (Ref 44) used tungsten-rhenium thermocouples to measure the temperature for low-alloy steel, type 316L stainless steel, and Inconel 718 weld pools created during gas-tungsten] arc welding (GTAW). Thermocouples were inserted from the underside of the work-piece in holes drilled to within 0.38 mm (0.015 in.) from the surface in order to measure weld pool surface temperatures. Some thermocouples were found to protrude beyond the free surface due to the weld pool surface deformation. Draw-backs for thermocouple measurements are the several-millisecond time constant for temperature measurement and weld pool temperature and flow field perturbations. However, the response time constant can be reduced by 3 to 4 times using the procedure reported by Dantzig (Ref 45), which entails calibration of the thermocouple and signal conditioning.

Thermography. Infrared (IR) thermographic cameras are now commercially available. A number of research groups have used them for weld pool temperature measurement (Ref 46, 47, 48, 49, 50, 51). Thermographic cameras can be used for high-resolution measurements of relative emissive power of the weld pools, but lack of emissivity data prevents their use for accurate absolute temperature measurement (Ref 39). Lukens and Morris (Ref 49) correlated the IR emission intensity measurements with temperatures measured by thermocouples embedded 2 mm (0.08 in.) below the surface. Although this provides some idea about the average temperature of the weld pool, it fails to validate the surface temperature observed by the thermographic camera. However, this technique can be applied for weld control and monitoring of weld cooling rate (Ref 49).

Narrow-band IR pyrometry was used for weld pool temperature measurement by Schaver *et al.* (Ref 52) for translating electron-beam welds, and Giedt *et al.* (Ref 53) for stationary welds. They used a 1.4 μm wavelength, and spectral emissivity was measured at the gas-tungsten arc at the melting point of the metals welded.

Hunter, Allemand, and Eager (Ref 54, 55) developed techniques for temperature measurement using multiwavelength pyrometry to alleviate the emissivity variation problem. However, so far the only proof of the concept was demonstrated using a heated platinum substrate. This could be a valuable tool for surface temperature measurement. Spatial resolution of 100 μm with 5% accuracy in temperature measurements above 1100 K is possible in this technique. They reported measurement of spectral radiance at approximately 200 wavelengths in a range of 0.6 to 0.8 μm. In real life, measurement of 5 or 6 wavelengths is adequate.

Optical Spectral Radiometric/Laser Reflectance (OSRLR). Kraus (Ref 56, 57) developed the OSRLR technique for high-resolution weld pool surface temperature measurements. This technique combines high-speed-film radiographic measurements of weld pool spectral directional thermal emissive power with spectral directional emissivity measurements to enable temperature calculation up to a point (Ref 39). However, this method requires knowledge of the weld pool free surface topology as well as an oxide-free weld pool surface because the surface has to reflect a focused laser beam. High resolution with spot sizes up to 60 μm have been achieved using this technique. Figure 4 shows the experimental setup for measurement of spectral directional emissivity in the weld pool. The details of the application of this technique for weld pool surface temperature measurements using OSRLR for gas-tungsten arc welds in types 304 and 316 stainless steels and AISI 8630 alloy steel are provided in Ref 58. Similar temperature measurements based on laser line reflection were also applied by Fuchs *et al.* (Ref 51) for weld pool surface temperature measurement of gas-metal arc welds.

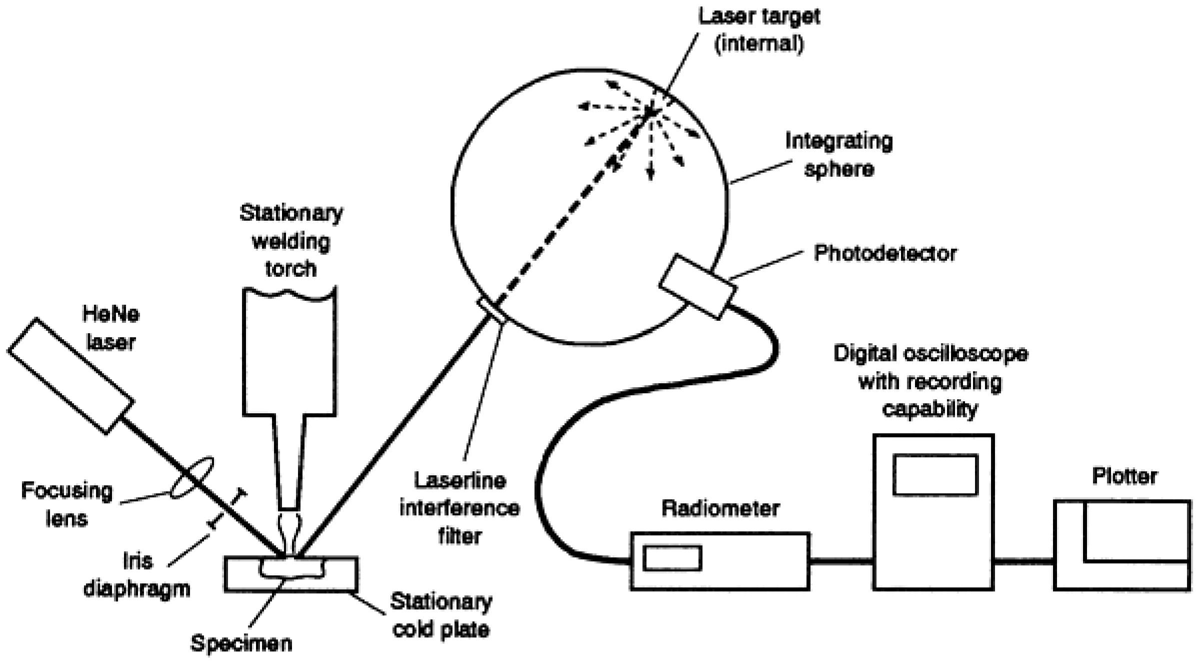

FIG. 4 SCHEMATIC OF THE SETUP USED FOR MEASURING SPECTRAL DIRECTIONAL EMISSIVITY IN THE WELD POOL

References cited in this section

39. H.G. KRAUS, *MODELING AND CONTROL OF CASTING AND WELDING PROCESSES IV,* PROC. CONF., A.F. GIAMEI AND G.J. ABBASCHIAN, ED., TMS-AIME, 1988, P 205-212
44. R.E. SUNDELL *ET AL.*, REPORT NO. 86SRD013 SUBMITTED TO THE NATIONAL SCIENCE FOUNDATION, GENERAL ELECTRIC COMPANY, SCHENECTADY, NY, 1986
45. J.A. DANTZIG, *REV. SCI. INSTRUM.,* VOL 56 (NO. 5), 1984, P 723-725
46. R.E. SAMPSON, G. SUITS, AND C.B. ARNOLD, REPORT NO. 176200-1-F, ENVIRONMENTAL

RESEARCH INSTITUTE OF MICHIGAN, MARCH 1985

47. B.A. CHIN, J.S. GOODLING, AND N.H. MADSEN, *ROBOTICS TODAY,* VOL 5 (NO. 1), 1983, P 85-87
48. M.A. KHAN, N.H. MADSEN, J.S. GOODLING, AND B.A. CHIN, *OPT. ENG.,* VOL 25 (NO. 6), 1986, P 799-805
49. W.E. LUKENS AND R.A. MORRIS, *WELD. J.,* VOL 61 (NO. 1), 1982, P 27-33
50. J.P. BOILLOT, P. CIELO, G. BEGIN, C. MICHEL, M. LESSARD, P. FAFARD, AND D. VILLEMURE, *WELD. J.,* VOL 64 (NO. 7), 1985, P 209S-217S
51. E.A. FUCHS, K.W. MAHIN, A.R. ORTEGA, L.A. BERTRAM, D.R. WILLIAMS, AND A.R. POMPLUN, *THERMOSENSE XIII,* SOCIETY OF PHOTO-OPTICAL INSTRUMENTATION ENGINEERS (SPIE), VOL 1467, 1991, P 136-149
52. D.A. SCHAUER, W.H. GIEDT, AND S.M. SHINTAKU, *WELD. J.,* VOL 57 (NO. 5), 1978, P 127S-133S
53. W.H. GEIDT, X.-C. WEI, AND S.-R. WEI, *WELD. J.,* VOL 63 (NO. 12), 1984, P 376S-383S
54. G.B. HUNTER, C.D. ALLEMAND, AND T.W. EAGER, *OPT. ENG.,* VOL 24 (NO. 6), 1985, P 1081-1085
55. G.B. HUNTER, C.D. ALLEMAND, AND T.W. EAGER, *OPT. ENG.,* VOL 25 (NO. 11), 1986, P 1222-1231
56. H.G. KRAUS, *OPT. LETT.,* VOL 11 (NO. 12), 1986, P 773-775
57. H.G. KRAUS, *OPT. ENG.,* VOL 26 (NO. 12), 1987, P 1183-1190
58. H.G. KRAUS, *WELD. J.,* VOL 68, JULY 1989, P 269S-279S

Validation Strategies for Heat-Affected Zone and Fluid-Flow Calculations

J. Mazumder, University of Illinois at Urbana-Champaign

Model Validation and Boundary Condition

The importance of matching the boundary condition during the experimental measurement for model validation cannot be overemphasized. One of the concepts employs various measurement techniques at different parts of the weld pool to define the boundary conditions as tightly as possible (Ref 51). Another approach uses a cold plate, as shown in Fig. 5, which was specially designed to hold the specimens in order to achieve symmetry of temperature and electromagnetic boundary conditions.

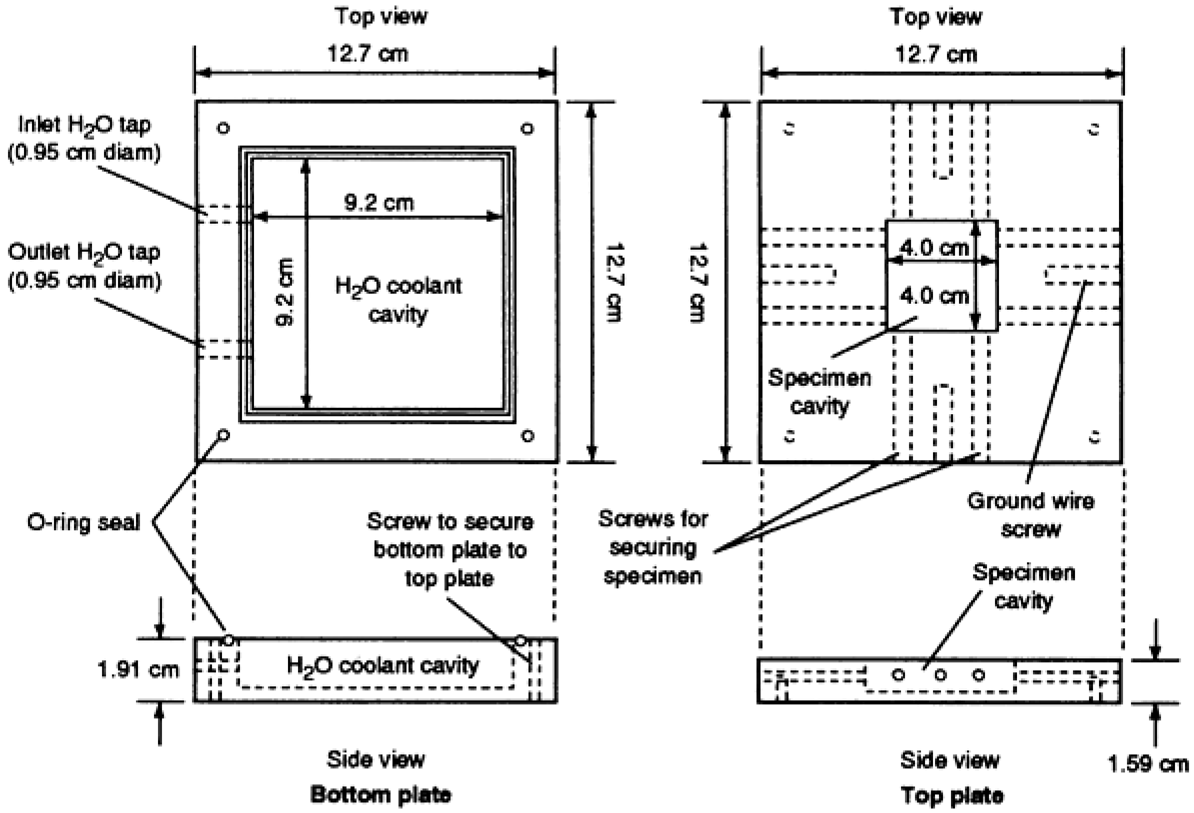
Top view
12.7 cm
Inlet H_2O tap (0.95 cm diam)
Outlet H_2O tap (0.95 cm diam)
9.2 cm
9.2 cm
H_2O coolant cavity
12.7 cm
O-ring seal
Screw to secure bottom plate to top plate
1.91 cm
H_2O coolant cavity
Side view
Bottom plate
Top view
12.7 cm
12.7 cm
4.0 cm
4.0 cm
Specimen cavity
Screws for securing specimen
Ground wire screw
Specimen cavity
Side view
Top plate
1.59 cm

FIG. 5 SCHEMATIC OF THE COLD PLATE USED FOR SECURING SPECIMENS IN ORDER TO ACHIEVE SYMMETRY OF TEMPERATURE AND ELECTROMAGNETIC BOUNDARY CONDITIONS DURING MODEL VALIDATION

Reference cited in this section

51. E.A. FUCHS, K.W. MAHIN, A.R. ORTEGA, L.A. BERTRAM, D.R. WILLIAMS, AND A.R. POMPLUN, *THERMOSENSE XIII*, SOCIETY OF PHOTO-OPTICAL INSTRUMENTATION ENGINEERS (SPIE), VOL 1467, 1991, P 136-149

Validation Strategies for Heat-Affected Zone and Fluid-Flow Calculations

J. Mazumder, University of Illinois at Urbana-Champaign

Validation of Residual Strain Calculation

Residual strains for weld HAZs are often measured by strain gages (Ref 35). Nonobtrusive techniques such as interferometry are also employed by some researchers (Ref 59, 60). For postweld residual strain measurements, x-ray diffraction is the most popular. Recently, neutron scattering techniques have been employed by some groups (Ref 61, 62). From the strain measurement data, stress data can be calculated provided modulus data for the particular material at the temperature range of interest are available.

Strain gages are essentially Wheatstone bridges that use the change of electrical resistance of a wire under strain to measure pressure. Strain gages generally consist of an array of four metal thin-film resistors. When they are deformed due to strain, the difference of output voltage is the measure of strain. Output voltages are generally very small; for example, 3 mv per volt of excitation for full-scale deformation with accuracies ranging from 1% to 0.1% (Ref 63). Therefore, one needs an amplifier to either plot the data on a chart recorder or to send it to the input/output board of a computer. Strain gages are generally mounted on the weld specimen with bonding agents (adhesives). Because both bonding agents and strain gages are very sensitive to high temperatures, strain measurement is limited to a region substantially outside the immediate vicinity of the HAZ.

Interferometry. All interferometric techniques are based on the interference of two coherent light wave trains. Masubuchi *et al.* (Ref 59) measured out-of-plane displacement using laser interferometer. Chavez *et al.* (Ref 60) used Moire interferometry to obtain full-field strain data across welds in stainless steel plates with Ferralium or nickel filler. Moire interferometry provides full-field surface displacement data with displacement resolution on the order of 1 μm. Holographic interferometric techniques can provide a higher resolution, whereas the Moire technique can be used in situations involving large strain variations such as occur during welding. Norvell (Ref 64) found strain variation for a 1 °C (1.8 °F) change in temperature can be measured by holographic interferometry; thus this method is too sensitive for welding. On the other hand, Moire interferometry is much more forgiving. Another interferometric technique called shearography seems to be more promising because the interference pattern is created simply by shearing the image by placing an optical wedge in front of the camera (Ref 65). Although Fourier transformation is required to deduce the strain data, unlike other interferometric techniques, shearography provides a strain field directly and minimizes data reduction.

X-Ray and Neutron Diffraction Methods. Both x-ray and neutron diffraction follow Bragg's law:

$$\lambda = 2D_{HKL} \text{ SIN } \theta_{HKL}$$

where λ is the wavelength, d_{hkl} is interplanar spacing of the plane *hkl,* and θ_{hkl} is the scattering angle.

An advantage of neutron diffraction is that it can penetrate to much greater depths. For example, a penetration depth in excess of 16 mm (0.63 in.) was observed for steel (Ref 61). For an x-ray, one needs to section off a specimen layer by layer, assuming that machining or chemical etching will not introduce any additional strain. This increased penetration depth is due to the extremely small wavelength (~0.165 nm) for neutrons. Root *et al.* (Ref 62) have determined the longitudinal, normal, and transverse strain components at 1.5 mm (0.06 in.) from inside the bottom surface of a multipass ferritic weldment. However, the major draw-back is that one needs a nuclear reactor for the neutron source. Penetration depth provides a unique opportunity to compose the prediction from three-dimensional models.

References cited in this section

35. P. TEKRIWAL AND J. MAZUMDER, *TRANS. ASME,* VOL 113, JULY 1991, P 336-343
59. K. MASUBUCHI, A. DEBICCARI, AND W.J.C. COOK, *ADVANCES IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID, ED., ASM INTERNATIONAL, 1986, P 529-532
60. S.A. CHAVEZ, V.A. DEASON, AND J.S. EPSTEIN, *ADVANCES IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID, ED., ASM INTERNATIONAL, 1986, P 533-537
61. K.W. MAHIN, W.S. WINTERS, T.M. HOLDEN, AND R.R. HOSBONS, *MODELING OF CASTING, WELDING AND ADVANCED SOLIDIFICATION PROCESSES V,* M. RAPPAZ, M.R. ÖZGÜ, AND K.W. MAHIN, ED., TMS-AIME, 1990, P 265-272
62. J.H. ROOT, T.M. HOLDEN, J. SCHÖDER, S. SPOONER, C.A. HUBBARD, T.A. DODSON, AND S.A. DAVID, *ADVANCES IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID AND J. VITEK, ED., ASM INTERNATIONAL, 1992, P 99-104
63. P. HOROWITZ AND W. HILL, *THE ART OF ELECTRONICS,* CAMBRIDGE UNIVERSITY PRESS, CAMBRIDGE, U.K., 1980
64. W. NORVELL, "THE STUDY OF MECHANICAL AND THERMAL DEFORMATION USING MULTI-DIMENSIONAL OPTICAL NON-DESTRUCTIVE TESTING METHOD," MASTERS THESIS, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, 1987
65. Y.Y. HUNG, *OPT. ENG.,* VOL 21 (NO. 3), 1982, P 391-395

Validation Strategies for Heat-Affected Zone and Fluid-Flow Calculations

J. Mazumder, University of Illinois at Urbana-Champaign

References

1. D. ROSENTHAL, *WELD. J.,* VOL 20, 1941, P 220S-225S
2. P.T. HOULDCROFT, *BR. WELD. J.,* VOL 1, 1954, P 468
3. W.J. PUMPHREY, *BR. WELD. J.,* VOL 2, 1955, P 92
4. C.R. HEIPLE AND J.R. ROPER, *WELD. J.,* VOL 61, 1982, P 973-1025
5. J. MAZUMDER AND A. KAR, *THERMOMECHANICAL ASPECTS OF MANUFACTURING AND MATERIALS PROCESSING,* R.K. SHAH *ET AL.,* ED., HEMISPHERE PUBLISHING, P 283-304
6. J. MAZUMDER, *OPT. ENG.,* VOL 30 (NO. 8), 1992, P 1208-1219
7. A. PAUL AND T. DEBROY, *METALL. TRANS. B,* VOL 19, 1988, P 851-858
8. N. PIRCH, E.W. KREUTZ, L. MOLLER, A. GASSER, AND K. WISSENBACH, *PROC. 3RD EUROPEAN CONF. ON LASER TREATMENT OF MATERIALS, ECLAT '90,* VOL 1, ERLANGEN, GERMANY, 17-19 SEPT 1990, SPRECHSAAL PUBLISHING GROUP
9. R.L. ZEHR, "THERMOCAPILLARY CONVECTION IN LASER MELTED POOL DURING MATERIALS PROCESSING," PH.D. THESIS, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, 1991
10. C.L. CHAN, R. ZEHR, J. MAZUMDER, AND M.M. CHEN, *PROC. 3RD ENGINEERING FOUNDATION CONF. ON MODELING AND CONTROL CASTING AND WELDING PROCESSES,* S. KOU AND R. MEHRABIAN, ED., TMS-AIME, 1986, P 229-246
11. M.C. TSAI AND S. KOU, *INT. J. NUMER. METHODS FLUIDS,* VOL 9 (NO. 12), DEC 1989, P 1503-1516
12. M.C. TSAI AND S. KOU, *NUMER. HEAT TRANSFER A: APPLICATIONS,* VOL 17 (NO. 1), 1990, P 73-89
13. T. ZACHARIA, A.H. ERASLAN, AND D.K. AIDUN, *WELD. J.,* VOL 67 (NO. 3), MARCH 1988, P 53S-62S
14. T. ZACHARIA, A.H. ERASLAN, AND D.K. AIDUN, *WELD. J.,* VOL 67 (NO. 1), JAN 1988, P 18S-27S
15. R.T.C. CHOO, J. SZEKELY, AND R.C. WESTHOFF, *WELD. J.,* VOL 69 (NO. 9), SEPT 1990, P 346S-361S
16. L. KARLSSON, CHAPTER 5, *THERMAL STRESSES I,* R.B. HETNARSKI, ED., NORTH-HOLLAND, 1986, P 299-389
17. J. GOLDAK, *RECENT TRENDS IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID AND J.M. VITEK, ED., ASM INTERNATIONAL, 1989, P 71-82
18. H.D. HIBBIT AND P.V. MARCAL, *COMPUTERS AND STRUCTURES,* VOL 3 (NO. 5), SEPT 1973, P 1145-1174
19. E. FRIEDMAN, *J. PRESSURE VESSEL TECHNOL. (TRANS. ASME),* VOL 97 (NO. 3), AUG 1975, P 206-213
20. B.A.B. ANDERSSON, *ENG. MATER. TECHNOL. (TRANS ASME),* VOL 100, OCT 1978, P 345-362
21. E.F. RYBICKI, D.W. SCHMUESER, R.W. STONESIFER, J.J. GROOM, AND H.W. MISHLER, *J. PRESSURE VESSEL TECHNOL. (TRANS ASME),* VOL 100, AUG 1978, P 256-262
22. E.F. RYBICKI AND R.B. STONESIFER, *J. PRESSURE VESSEL TECHNOL. (TRANS. ASME),* VOL 101, MAY 1979, P 149-154
23. Y. UEDA AND K. NAKACHO, *TRANS. JWRI,* VOL 9 (NO. 1), 1980, P 107-114
24. D.H. ALLEN AND W.E. HAISLER, *COMPUTERS AND STRUCTURES,* VOL 13, 1981, P 129-135
25. J.H. ARGYRIS, J. SZIMMAT, AND K.J. WILLIAM, *COMPUTER METHODS IN APPLIED MECHANICS AND ENGINEERING,* VOL 33, 1982, P 635-666

26. B. ANDERSSON AND L. KARLSSON, *J. THERM. STRESSES,* VOL 4, 1981, P 491-500
27. B. PATEL, "THERMO-ELASTO-PLASTIC FINITE ELEMENT FORMULATION FOR DEFORMATION AND RESIDUAL STRESSES DUE TO WELDS," PH.D. THESIS, CARLETON UNIVERSITY, OTTAWA, ONTARIO, CANADA, APRIL 1985
28. K.W. MAHIN, S. MACEWEN, W.S. WINTERS, W. MASON, M. KANOUFF, AND E.A. FUCHS, *MODELING AND CONTROL OF CASTING AND WELDING PROCESSES IV,* PROC. INT. CONF. ON MODELING OF CASTING AND WELDING PROCESSES IV, A.F. GIAMEI AND G.J. ABBASCHIAN, ED., ENGINEERING FOUNDATION AND TMS/AIME, APRIL 1988, P 339-350
29. J. GOLDAK, A. ODDY, M. MCDILL, A. CHAKRAVARTI, M. BIBBY, AND R. HOUSE, *ADVANCES IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID, ED., ASM INTERNATIONAL, 1986, P 523-527
30. P. TEKRIWAL AND J. MAZUMDER, *ADVANCES IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID, ED., ASM INTERNATIONAL, 1986, P 71-80
31. R.I. KARLSSON AND B.L. JOSEFSON, *J. PRESSURE VESSEL TECHNOL. (TRANS. ASME),* VOL 112, FEB 1990, P 76-84
32. C.K. LEUNG, "STUDY OF FINITE ELEMENT MODELING OF MULTIPASS WELDS," PH.D. THESIS, UNIVERSITY OF WATERLOO, WATERLOO, ONTARIO, CANADA, 1986
33. S.Z. NEWMAN, "FEM MODEL OF 3D TRANSIENT TEMPERATURE AND STRESS FIELDS IN WELDED PLATES," PH.D. THESIS, CARNEGIE MELLON UNIVERSITY, PITTSBURGH, PA, APRIL 1986
34. P. TEKRIWAL, "THREE-DIMENSIONAL TRANSIENT THERMO-ELASTO-PLASTIC MODELING OF GAS METAL ARC WELDING USING THE FINITE ELEMENT METHOD," PH.D. THESIS, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, URBANA, IL, AUGUST 1989
35. P. TEKRIWAL AND J. MAZUMDER, *TRANS. ASME,* VOL 113, JULY 1991, P 336-343
36. G.M. OREPAR, T.W. EAGER, AND J. SZEKELY, *WELD. J.,* VOL 63, 1983, P 307S
37. J. MAZUMDER AND D. VOELKEL, *PROCEEDINGS OF LAMP '92,* A. MATSUNAWA AND KATAYAMA, ED., HIGH TEMPERATURE SOCIETY OF JAPAN, JUNE 1992, P 373-380
38. J. MAZUMDER AND D. VOELKEL, U.S. PATENT APPLICATION 08/004763, JAN 1993
39. H.G. KRAUS, *MODELING AND CONTROL OF CASTING AND WELDING PROCESSES IV,* PROC. CONF., A.F. GIAMEI AND G.J. ABBASCHIAN, ED., TMS-AIME, 1988, P 205-212
40. D. VOELKEL, "LASER WELD POOL OPTICAL DIAGNOSTICS: TOPOGRAPHY AND VISUALIZATION USED FOR SURFACE CONTOUR ANALYSIS AND PROCESS MONITORING," PH.D. THESIS, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, URBANA, IL, JAN 1992
41. P. SAHOO, T. DEBROY, AND M.J. MCNALLAN, *METALL. TRANS. B,* VOL 19, 1988, P 484-491
42. M.J. MCNALLAN AND T. DEBROY, *METALL. TRANS. B,* VOL 22, 1991, P 557-560
43. J. MAZUMDER, M.M. CHEN, C.L. CHAN, R. ZEHR, AND D. VOELKEL, *PROC. ECLAT '90,* H. BERGMANN, ED., SPRECHSAAL PUBLISHING GROUP, COBURG, GERMANY, 1990, P 37-53
44. R.E. SUNDELL *ET AL.,* REPORT NO. 86SRD013 SUBMITTED TO THE NATIONAL SCIENCE FOUNDATION, GENERAL ELECTRIC COMPANY, SCHENECTADY, NY, 1986
45. J.A. DANTZIG, *REV. SCI. INSTRUM.,* VOL 56 (NO. 5), 1984, P 723-725
46. R.E. SAMPSON, G. SUITS, AND C.B. ARNOLD, REPORT NO. 176200-1-F, ENVIRONMENTAL RESEARCH INSTITUTE OF MICHIGAN, MARCH 1985
47. B.A. CHIN, J.S. GOODLING, AND N.H. MADSEN, *ROBOTICS TODAY,* VOL 5 (NO. 1), 1983, P 85-87
48. M.A. KHAN, N.H. MADSEN, J.S. GOODLING, AND B.A. CHIN, *OPT. ENG.,* VOL 25 (NO. 6), 1986, P 799-805
49. W.E. LUKENS AND R.A. MORRIS, *WELD. J.,* VOL 61 (NO. 1), 1982, P 27-33
50. J.P. BOILLOT, P. CIELO, G. BEGIN, C. MICHEL, M. LESSARD, P. FAFARD, AND D. VILLEMURE, *WELD. J.,* VOL 64 (NO. 7), 1985, P 209S-217S
51. E.A. FUCHS, K.W. MAHIN, A.R. ORTEGA, L.A. BERTRAM, D.R. WILLIAMS, AND A.R.

POMPLUN, *THERMOSENSE XIII,* SOCIETY OF PHOTO-OPTICAL INSTRUMENTATION ENGINEERS (SPIE), VOL 1467, 1991, P 136-149

52. D.A. SCHAUER, W.H. GIEDT, AND S.M. SHINTAKU, *WELD. J.,* VOL 57 (NO. 5), 1978, P 127S-133S
53. W.H. GEIDT, X.-C. WEI, AND S.-R. WEI, *WELD. J.,* VOL 63 (NO. 12), 1984, P 376S-383S
54. G.B. HUNTER, C.D. ALLEMAND, AND T.W. EAGER, *OPT. ENG.,* VOL 24 (NO. 6), 1985, P 1081-1085
55. G.B. HUNTER, C.D. ALLEMAND, AND T.W. EAGER, *OPT. ENG.,* VOL 25 (NO. 11), 1986, P 1222-1231
56. H.G. KRAUS, *OPT. LETT.,* VOL 11 (NO. 12), 1986, P 773-775
57. H.G. KRAUS, *OPT. ENG.,* VOL 26 (NO. 12), 1987, P 1183-1190
58. H.G. KRAUS, *WELD. J.,* VOL 68, JULY 1989, P 269S-279S
59. K. MASUBUCHI, A. DEBICCARI, AND W.J.C. COOK, *ADVANCES IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID, ED., ASM INTERNATIONAL, 1986, P 529-532
60. S.A. CHAVEZ, V.A. DEASON, AND J.S. EPSTEIN, *ADVANCES IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID, ED., ASM INTERNATIONAL, 1986, P 533-537
61. K.W. MAHIN, W.S. WINTERS, T.M. HOLDEN, AND R.R. HOSBONS, *MODELING OF CASTING, WELDING AND ADVANCED SOLIDIFICATION PROCESSES V,* M. RAPPAZ, M.R. ÖZGÜ, AND K.W. MAHIN, ED., TMS-AIME, 1990, P 265-272
62. J.H. ROOT, T.M. HOLDEN, J. SCHÖDER, S. SPOONER, C.A. HUBBARD, T.A. DODSON, AND S.A. DAVID, *ADVANCES IN WELDING SCIENCE AND TECHNOLOGY,* S.A. DAVID AND J. VITEK, ED., ASM INTERNATIONAL, 1992, P 99-104
63. P. HOROWITZ AND W. HILL, *THE ART OF ELECTRONICS,* CAMBRIDGE UNIVERSITY PRESS, CAMBRIDGE, U.K., 1980
64. W. NORVELL, "THE STUDY OF MECHANICAL AND THERMAL DEFORMATION USING MULTI-DIMENSIONAL OPTICAL NON-DESTRUCTIVE TESTING METHOD," MASTERS THESIS, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, 1987
65. Y.Y. HUNG, *OPT. ENG.,* VOL 21 (NO. 3), 1982, P 391-395

Oxyfuel Gas Cutting

Introduction

OXYFUEL GAS CUTTING (OFC) includes a group of cutting processes that use controlled chemical reactions to remove preheated metal by rapid oxidation in a stream of pure oxygen. A fuel gas/oxygen flame heats the workpiece to ignition temperature, and a stream of pure oxygen feeds the cutting (oxidizing) action. The OFC process, which is also referred to as burning or flame cutting, can cut carbon and low-alloy plate of virtually any thickness. Castings more than 760 mm (30 in.) thick commonly are cut by OFC processes. With oxidation-resistant materials, such as stainless steels, either a chemical flux or metal powder is added to the oxygen stream to promote the exothermic reaction.

The simplest oxyfuel gas cutting equipment consists of two cylinders (one for oxygen and one for the fuel gas), gas flow regulators and gages, gas supply hoses, and a cutting torch with a set of exchangeable cutting tips. Such manually operated equipment is portable and inexpensive. Cutting machines, employing one or several cutting torches guided by solid template pantographs, optical line tracers, numerical controls, or computers, improve production rates and provide superior cut quality. Machine cutting is important for profile cutting--the cutting of regular and irregular shapes from flat stock.

Acknowledgements

This article was adapted from "Thermal Cutting" in *Welding, Brazing, and Soldering,* Volume 6 of the 9th edition of *Metals Handbook.* The author of the original article is Rosalie Brosilow of Penton Publishing. Special thanks are due to Lance Soisson of Welding Consultants Inc. for his help in adapting the article for this volume.

Oxyfuel Gas Cutting

Principles of Operation

Oxyfuel gas cutting begins by heating a small area on the surface of the metal to the ignition temperature of 760 to 870 °C (1400 to 1600 °F) with an oxyfuel gas flame. Upon reaching this temperature, the surface of the metal will appear bright red. A cutting oxygen stream is then directed at the preheated spot, causing rapid oxidation of the heated metal and generating large amounts of heat. This heat supports continued oxidation of the metal as the cut progresses. Combusted gas and the pressurized oxygen jet flush the molten oxide away, exposing fresh surfaces for cutting. The metal in the path of the oxygen jet burns. The cut progresses, making a narrow slot, or kerf, through the metal.

To start a cut at the edge of a plate, the edge of the preheat flame is placed just over the edge to heat the material. When the plate heats to red, the cutting oxygen is turned on, and the torch moves over the plate to start the cut.

During cutting, oxygen and fuel gas flow through separate lines to the cutting torch at pressures controlled by pressure regulators, adjusted by the operator. The cutting torch contains ducts, a mixing chamber, and valves to supply an oxyfuel gas mixture of the proper ratio for preheat and a pure oxygen stream for cutting to the torch tip. By adjusting the control valves on the torch handle or at the cutting machine controller, the operator sets the precise oxyfuel gas mixture desired. Depressing the cutting oxygen lever on the torch during manual operation initiates the cutting oxygen flow. For machine cutting, oxygen is normally controlled by the operator at a remote station or by numerical control. Cutting tips have a single cutting oxygen orifice centered within a ring of smaller oxyfuel gas exit ports. The operator changes the cutting capacity of the torch by changing the cutting tip size and by resetting pressure regulators and control valves. Because different fuel gases have different combustion and flow characteristics, the construction of cutting tips, and sometimes of mixing chambers, varies according to the type of gas.

Oxyfuel gas flames initiate the oxidation action and sustain the reaction by continuously heating the metal at the line of the cut. The flame also removes scale and dirt that may impede or distort the cut.

The rate of heat transfer in the workpiece influences the heat balance for cutting. As the thickness of the metal to be cut increases, more heat is needed to keep the metal at its ignition temperature. Increasing the preheat gas flow and reducing the cutting speed maintains the necessary heat balance.

Oxygen flow also must increase as the thickness of the metal to be cut increases. The jet of cutting oxygen must have sufficient volume and velocity to penetrate the depth of the cut and still maintain its shape and effective oxygen content.

Quality of Cut. Oxyfuel gas cutting operations combine more than 20 variables. Suppliers of cutting equipment provide tables that give approximate gas pressures for various sizes and styles of cutting torches and tips and recommended cutting speeds; these variables are operator controlled. Other variables include type and condition (scale, oil, dirt, flatness) of material, thickness of cut, type of fuel gas, and quality and angle of cut.

Where dimensional accuracy and squareness of the cut edge are important, the operator must adjust the process to minimize the kerf, the width of metal removed by cutting, and to increase smoothness of the cut edge. Careful balancing of all cutting variables helps attain a narrow kerf and smooth edge. The thicker the work material, the greater the oxygen volume required and therefore, the wider the cutting nozzle and kerf.

Oxyfuel Gas Cutting

Process Capabilities

Oxyfuel gas cutting processes are used primarily for severing carbon and low-alloy steels. Other iron-base alloys and some nonferrous metals can be oxyfuel gas cut, although process modification may be required, and cut quality may not be as high as is obtained in cutting the more widely used grades of steel. High-alloy steels, stainless steels, cast iron, and nickel alloys do not readily oxidize and therefore do not provide enough heat for a continuous reaction. As the carbon and alloy contents of the steel to be cut increase, preheating or postheating, or both, often are necessary to overcome the effect of the heat cycle, particularly the quench effect of cooling.

Some of the high-alloy steels, such as stainless steel, and cast iron can be cut successfully by injecting metal powder (usually iron) or a chemical additive into the oxygen jet. The metal powder supplies combustion heat and breaks up oxide films. Chemical additives combine with oxides to form lower temperature melting products that flush away.

Applications. Large-scale applications of oxyfuel cutting are found in shipbuilding, structural fabrication, manufacture of earthmoving equipment, machinery construction, and in the fabrication of pressure vessels and storage tanks. Many machine structures, originally made from forgings and castings, can be made at less cost by redesigning them for OFC and welding with the advantages of quick delivery of plate material from steel suppliers, low cost of oxyfuel gas cutting equipment, and flexibility of design.

Structural shapes, pipe, rod, and similar materials can be cut to length for construction or cut up in scrap and salvage operations. In steel mills and foundries, projections such as caps, gates, and risers can be severed from billets and castings. Mechanical fasteners can be quickly cut for disassembly using OFC. Holes can be made in steel components by piercing and cutting. Machine OFC is used to cut steel plate to size, to cut various shapes from plate, and to prepare plate edges (bevel cutting) for welding.

Gears, sprockets, handwheels, clevises and frames, and tools such as wrenches can be cut out by oxyfuel gas torches. Often, these oxyfuel cut products can be used without further finishing. However, when cutting medium- or high-carbon steel or other metal that hardens by rapid cooling, the hardening effect must be considered, especially if the workpiece is to be subsequently machined.

Thickness Limits. Gas can cut steels less than 3.2 mm ($\frac{1}{8}$ in.) thick to over 1525 mm (60 in.) thick, though some sacrifice in quality occurs near both ends of this range. With very thin material, operators may have some difficulty in keeping heat input low to avoid melting the kerf edges and to minimize distortion. Steel under 6.4 mm ($\frac{1}{4}$ in.) thick often is stacked for cutting of several parts in a single torch pass.

There are a number of advantages and disadvantages when OFC is compared to other cutting operations such as arc cutting, milling, shearing, or sawing. The advantages of OFC are:

- METAL CAN BE CUT FASTER BY OFC. SETUP IS GENERALLY SIMPLER AND FASTER THAN FOR MACHINING AND ABOUT EQUAL TO THAT OF MECHANICAL SEVERING (SAWING AND SHEARING).
- OXYFUEL GAS CUTTING PATTERNS ARE NOT CONFINED TO STRAIGHT LINES AS IN SAWING AND SHEARING, OR TO FIXED PATTERNS AS IN DIE CUTTING PROCESSES. CUTTING DIRECTION CAN BE CHANGED RAPIDLY ON A SMALL RADIUS DURING OPERATION.
- MANUAL OFC EQUIPMENT COSTS ARE LOW COMPARED TO MACHINE TOOLS. SUCH EQUIPMENT IS PORTABLE AND SELF-CONTAINED, REQUIRING NO OUTSIDE POWER AND WELL SUITED FOR FIELD USE.
- WHEN PROPERTIES AND DIMENSIONAL ACCURACY OF GAS CUT PLATE ARE ACCEPTABLE, OFC CAN REPLACE COSTLY MACHINING OPERATIONS. IT OFFERS REDUCED LABOR AND OVERHEAD COSTS, REDUCED MATERIAL COSTS, REDUCED TOOLING COSTS, AND FASTER DELIVERY.
- WITH ADVANCED MACHINERY, OFC LENDS ITSELF TO HIGH-VOLUME PARTS PRODUCTION.
- LARGE PLATES CAN BE CUT QUICKLY IN PLACE BY MOVING THE GAS TORCH RATHER THAN THE PLATE.
- TWO OR MORE PIECES CAN BE CUT SIMULTANEOUSLY USING STACK CUTTING METHODS AND MULTIPLE-TORCH CUTTING MACHINES.

The limitations of the OFC process include:

- DIMENSIONAL TOLERANCES ARE POORER THAN FOR MACHINING AND SHEARING.
- BECAUSE OFC RELIES ON OXIDATION OF IRON, IT IS LIMITED TO CUTTING STEELS AND CAST IRON.
- HEAT GENERATED BY OFC CAN DEGRADE THE METALLURGICAL PROPERTIES OF THE WORK MATERIAL ADJACENT TO THE CUT EDGES. HARDENABLE STEELS MAY REQUIRE PREHEAT AND/OR POSTHEAT TO CONTROL MICROSTRUCTURE AND MECHANICAL PROPERTIES.
- PREHEAT FLAMES AND THE EXPELLED RED HOT SLAG POSE A FIRE HAZARD TO PLANT AND PERSONNEL.

Oxygen consumption and flow rates vary depending on whether economy, speed, or accuracy of cut is desired. For average straight-line cutting of low-carbon steel, consumption of cutting oxygen per pound of metal removed varies with thickness of the metal and is lowest at a thickness of 100 to 125 mm (4 to 5 in.).

By assuming that for every unit mass of iron oxidized an equal mass of iron melts, one can calculate the amount of heat generated by the cutting reaction--heat emitted is 6680 kJ/kg (2870 Btu/lb) of iron oxidized. Melting of 0.5 kg (1 lb) of iron takes 715 kJ (680 Btu), based on a melting point of 1540 °C (2800 °F), 0.8 kJ/kg · K (0.2 Btu/lb · °F) as the specific heat, and 272 kJ/kg (117 Btu/lb) as the heat of fusion. Only a small amount of the heat melts the iron; most of it, about 4900 kJ/kg (2100 Btu/lb), goes into the reaction. Some of this superheats the molten metal, some soaks into the workpiece, and some leaves by radiation and convection. Most of it leaves with the slag and hot exhaust gases.

As cutting oxygen flows down through the cut, the quantity available for reaction decreases. If the flow of oxygen is large and well collimated, the rate of cutting through the depth of the cut is approximately constant. The cutting face remains vertical if the oxygen is in excess and if cutting speed is not excessive.

If oxygen flow is insufficient, or cutting speed too high, the lower portions of the cut react more slowly, and the cutting face curves behind the torch. The horizontal distance between point of entry and exit is called drag (see Fig. 1).

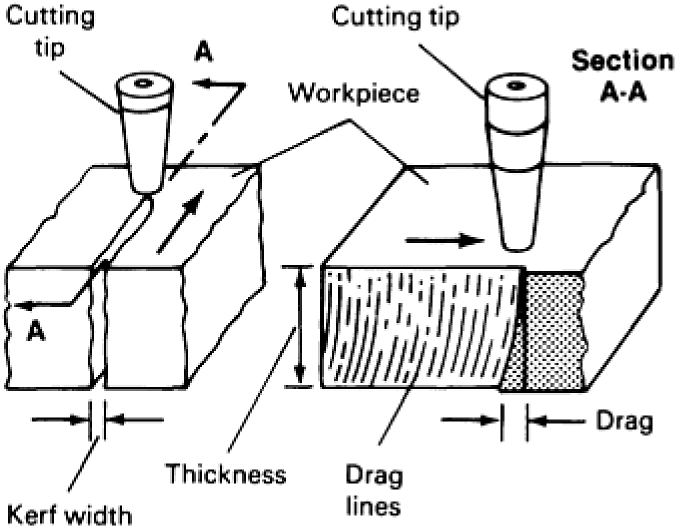

FIG. 1 SCHEMATIC CROSS SECTION OF WORK METAL DURING OFC SHOWING DRAG ON CUTTING FACE

Drag influences edge quality. Optimum edge quality results from zero drag--the oxygen stream enters and leaves the cut in a straight line along the cutting tip axis. This is called a "drop-cut," designating a clean, fully severed edge. Increasing cutting speed or reducing oxygen flow makes less oxygen available at the bottom of the cut, causing the bottom of the cut to "drag" behind the top of the cut. A drag of 20% means that the bottom of the cut edge lags the top surface by 20% of the material thickness. Drag lines appear as curved ripples on the cut edge. For fast, rough cuts, some drag is acceptable.

Drag is a rough measure of cut quality and of economy in oxygen consumption. In metal thicknesses up to 50 to 75 mm (2 or 3 in.), 10 to 15% drag indicates good quality of cut and economy. The key to quality cuts is control of heat input. Higher quality demands less drag; more drag indicates poorer quality and low oxygen consumption. Excessive drag may lead to incomplete cutting.

In very thin sections, drag has little significance. In very thick sections, the goal is to avoid excessive drag.

Preheating may consist of merely warming a cold workpiece with a torch or may require furnace heating of the work beyond 540 °C (1000 °F). For some alloy steels, preheat temperatures are 205 to 315 °C (400 to 600 °F). Carbon steel billets and other sections occasionally are cut at 870 °C (1600 °F) and higher.

In OFC, preheating is accomplished by means of the oxyfuel gas flame, which surrounds the cutting oxygen stream. At cut initiation, the preheat flame, the result of oxygen and fuel gas combustion, brings a small amount of material to ignition temperature so that combustion can proceed. After cutting begins, the preheat flame merely adds heat to compensate for heat lost by convection and radiation or through gas exhausted during cutting. The flame also helps to remove or burn off scale and dirt on the plate surface; the hot combusted gases protect the stream of cutting oxygen from the atmosphere.

Preheating may also be applied over a broader area of the work. It may include soaking the entire workpiece in a furnace to bring it up to 95 to 205 °C (200 to 400 °F), or a simple overall warm-up with a torch to bring cold plate to room temperature. A preheat improves cutting speed significantly, allowing faster torch travel for greater productivity and reduced consumption of fuel gas. Broader preheat smooths the temperature gradient between the base metal and the cut edge, possibly reducing thermal stress and minimizing hardening effects in some steels.

Properties of Fuel Gases

Each cutting job entails a different type or volume of work to be completed. Consequently, the best gas for all cutting in a fabricating plant is found through experimentation. Evaluating a gas for a single job requires a test run that monitors fuel gas and oxygen flow rate, labor costs, overhead, and the amount of work performed. If plant production varies from week to week, gas performance should be measured over a long enough period to achieve an accurate cost analysis. Any of the fuel gases may perform well over a range of flow rates. When comparing gases, performance should be rated at the lowest flow rate that gives acceptable results for each gas.

Gas manufacturers provide comprehensive data on flame temperature, heat of combustion, oxygen-to-gas consumption ratio, heat transfer, and heat distribution in the flame. Gas properties can be rated on a weight or volume basis. Flow rates, for example, commonly are given in terms of volume. Gas may be sold, however, by weight or by volume, and specifications may state properties in terms of one or the other to make the product seem as attractive as possible. For example, acetylene occupies more volume per pound than does propane. Consequently, cost based on weight makes propane liquid seem inexpensive. Cost based on volume makes the lighter fuels, such as acetylene, seem less costly. The user must understand these ratings and comparisons.

Heat Distribution in the Flame. Combustion of fuel gas produces the preheat flame, which initiates cutting action and helps sustain the operation. These flames heat the surface of the work to ignition temperature to (1) initiate cutting, (2) descale and clean the work surface, (3) supply heat to the work and the cutting oxygen stream to maintain the heat needed for continuous cutting, and (4) shield the cutting oxygen stream from surrounding air. Preheat gases consist of hydrocarbons, which produce water vapor and carbon dioxide as products of chain reactions. These reactions occur in cones within the flame, usually visible as an inner and outer flame. A gas whose inner flame has a high temperature and high heat release provides the most concentrated heat. These gases are superior for fast starts in flame cutting of high-alloy steels that are difficult to cut.

When low heat release is accompanied by high flame temperature, even though it is above the melting point of steel, heat is diffused and gives slow starts in flame cutting. Gases, such as natural gas, that release most of their heat in the outer flame are well suited to heating and heavy cutting. Heat distribution is a good indication of the potential performance of a particular gas.

Acetylene

Acetylene is the most widely used cutting gas. It burns hotter than any of the other common fuel gases, making it indispensable for certain jobs. The volume of oxygen consumed per volume of acetylene is $2\frac{1}{2}$. Combustion occurs in stages. In the small inner cone at the tip of the torch, acetylene burns with feed-line oxygen. This reaction gives off a blistering amount of heat; the tip of the inner cone is the hottest part. Burning continues in the outer envelope of the flame, in a cooler, blue-colored region.

Most oxygen for this reaction comes from the air surrounding the flame. Of the $2\frac{1}{2}$ parts of oxygen needed for acetylene to burn completely, about $1\frac{1}{2}$ comes from the air and 1 part from line oxygen. A neutral flame, recommended for manual cutting, consumes equal volumes of line oxygen and acetylene. In practice, operators may raise the line oxygen ratio to $1\frac{1}{2}$, a proportion that produces an oxidizing flame. The inner cone contracts and becomes sharply defined. This is the hottest flame attainable with acetylene or with any other raw fuel gas. The triple bond in acetylene makes it the hottest burning gas. In cutting of steel, operators should adjust the acetylene-to-oxygen ratio to produce a neutral flame, even though it is below the highest possible temperature. A neutral flame minimizes oxidation and carburization of the base metal. Users sometimes use the hotter oxidizing flame for piercing at the start of a cut.

Acetylene is unstable at room temperature and moderate pressure and can explode under a blow even without the presence of air or oxygen. For safe storage and handling, it is dissolved in acetone. A typical cylinder is filled with heavy porous clay, which holds the acetone. Acetylene, forced into the cylinder, saturates the acetone. When the cylinder valve opens, acetylene vaporizes above the acetone solution. The regulator limits outlet pressure to 105 kPa gage pressure (15 psig) or less.

Cylinders for acetylene should be handled with care. Safety requires them to be placed upright to prevent acetone solvent from entering the regulator. The ratio of line oxygen consumption to acetylene consumption is lower than that for the other fuel gases. Acetylene uses less oxygen at its maximum flame temperature, so that handling of oxygen cylinders may be reduced, thus offsetting the inconvenience of the acetylene cylinder.

Cost of acetylene as a fuel gas changes with consumption. Large-volume users of acetylene often install bulk trailers that hook up to manifold piping systems. When volume justifies, users may install acetylene-generating plants, thus reducing handling costs.

Heat content of acetylene (kJ/m^3, or Btu/ft^3) is lower than all gases except natural gas. Although acetylene burns at a high temperature, it has to burn at higher flow rates than other gases to deliver the same amount of heat. Acetylene releases heat rapidly in a small, concentrated area. Acetylene is a poor choice for cutting a large block of metal because of high fuel cost and the need for large volumes of the fuel to obtain required total heat. To concentrate heat on a limited area, as in cutting thin plate, acetylene is a good choice. It burns faster than other gases, and it burns close to the torch tip. Heat energy concentrates at the tip of the tiny cone.

Acetylene gives a narrow heat-affected zone (HAZ) and possibly less distortion (depending on the volume and shape of the work) than a wider spreading heat. It is particularly well suited to fast cutting of plate under 13 mm ($\frac{1}{2}$ in.) thick. The hot flame cuts through heavy mill scale or rust and can make a bevel quickly with good edge quality. Acetylene facilitates short, stop-and-start cutting jobs, such as cutting structural members and reinforcing bars, because of its short preheat time. For stack cutting and for cutting heavy sections--operations requiring large heat input into a deep kerf--acetylene does not perform as well as other gases. This type of work calls for a fuel gas with high heat output in the secondary flame, such as propane or methylacetylene-propadiene-stabilized (MPS or MAPP) gas.

However, manual cutting is cooler with acetylene than with other gases; less heat rises into the operator's face. It is one of the most versatile and convenient fuels for shops that perform a wide range of fuel gas operations and for shops that cannot justify handling more than one gas. Data for both machine and manual oxyacetylene cutting (OFC-A) of carbon steel plate of varying thickness are given in Tables 1 and 2.

TABLE 1 RECOMMENDED PARAMETERS FOR MANUAL OFC-A OF CARBON STEELS

PLATE THICKNESS		DIAMETER OF CUTTING ORIFICE(A)		OXYGEN PRESSURE(A)		CUTTING SPEED(A)(B)		GAS CONSUMPTION					
								OXYGEN			ACETYLENE(C)		
mm	**in.**	**mm**	**in.**	**kpa**	**psi**	**mm/min**	**in./min**	**m^3/h**	**ft^3/h**	**ft^3/linear ft**	**m^3/h**	**ft^3/h**	**ft^3/linear ft**
3.2	$\frac{1}{8}$	0.965-1.016	0.0380-0.0400	103-159	15-23	510-760	20-30	1.3-1.6	45-55	0.37-0.45	0.20-0.25	7-9	0.06-0.07
6.4	$\frac{1}{4}$	0.965-1.511	0.0380-0.0595	76-96	11-20	405-660	16-26	1.4-2.6	50-93	0.63-0.72	0.25-0.31	9-11	0.08-0.11
9.5	$\frac{3}{8}$	0.965-1.511	0.0380-0.0595	117-173	17-25	380-610	15-24	1.7-3.26	60-115	0.80-0.96	0.28-0.34	10-12	0.10-0.13
13	$\frac{1}{2}$	1.181-1.511	0.0465-0.0595	138-207	20-30	305-560	12-22	1.9-3.54	66-125	1.10-1.14	0.28-0.37	10-13	0.12-0.17
19	$\frac{3}{4}$	1.181-1.511	0.0465-0.0595	165-241	24-35	305-510	12-20	3.31-4.04	117-143	1.43-1.95	0.34-0.42	12-15	0.15-0.20
25	1	1.181-1.511	0.0465-0.0595	193-276	28-40	230-460	9-18	3.7-4.5	130-160	1.78-2.89	0.37-0.45	13-16	0.18-0.29
38	$1\frac{1}{2}$	1.511-2.057	0.0595-0.0810	241-331	35-48	150-355	6-14	4.05-5.04	143-178	1.96-3.18	0.42-0.51	15-18	0.21-0.33
50	2	1.702-2.057	0.0670-0.0810	152-345	22-50	150-330	6-13	5.23-6.54	185-231	3.55-6.16	0.45-0.57	16-20	0.31-0.53
75	3	1.702-2.057	0.0670-0.0810	228-379	33-55	100-255	4-10	6.8-8.2	240-290	5.80-12.00	0.54-0.65	19-23	0.46-0.95
100	4	2.057-2.184	0.0810-0.0860	290-414	42-60	100-205	4-8	8.29-10.9	293-388	9.70-14.64	0.59-0.74	21-26	0.65-1.05
125	5	2.057-2.184	0.0810-0.0860	365-483	53-70	90-165	3.5-6.4	9.82-12.4	347-437	13.66-19.83	0.68-0.82	24-29	0.91-1.37
150	6	2.489-2.577	0.0980-0.0995	310-552	45-80	76-137	3.0-5.4	11.3-16.1	400-567	21.00-26.70	0.76-0.91	27-32	1.19-1.80
205	8	2.527	0.0995	414-531	60-77	66-107	2.6-4.2	14.3-17.4	505-615	29.30-38.84	0.892-1.09	31.5-38.5	1.83-2.42
255	10	2.527	0.0995	517-662	75-96	48-81	1.9-3.2	17.3-21.2	610-750	46.90-64.20	1.04-1.28	36.9-45.1	2.57-3.84
305	12	3.048	0.1200	476-593	69-86	36-66	1.4-2.6	20.4-24.9	720-880	67.70-103.00	1.20-1.46	42.3-51.7	3.98-6.05

(A) VALUES DO NOT NECESSARILY VARY IN EXACT PROPORTION TO PLATE THICKNESS, BECAUSE STRAIGHT-LINE RELATIONS DO NOT EXIST AMONG PRESSURE, SPEED, AND ORIFICE SIZES.

(B) LOWEST SPEEDS AND HIGHEST GAS CONSUMPTIONS ARE FOR INEXPERIENCED OPERATORS, SHORT CUTS, DIRTY OR NONUNIFORM MATERIAL. HIGHEST SPEEDS AND LOWEST GAS CONSUMPTIONS ARE FOR EXPERIENCED OPERATORS, LONG CUTS, CLEAN AND UNIFORM MATERIAL.

(C) PRESSURE OF ACETYLENE FOR THE PREHEATING FLAMES IS MORE A FUNCTION OF TORCH DESIGN THAN OF THE THICKNESS OF THE PART BEING CUT. FOR ACETYLENE PRESSURE DATA, SEE CHARTS FROM MANUFACTURERS OF APPARATUS.

TABLE 2 RECOMMENDED PARAMETERS FOR MACHINE OFC-A OF CARBON STEELS

PLATE THICKNESS		**DIAMETER OF CUTTING ORIFICE(A)**		**OXYGEN PRESSURE(A)**		**CUTTING SPEED(B)**		**GAS CONSUMPTION**					
								OXYGEN			**ACETYLENE(C)**		
mm	**in.**	**mm**	**in.**	**kpa**	**psi**	**mm/min**	**in./min**	**m^3/h**	**ft^3/h**	**ft^3/linear ft**	**m^3/h**	**ft^3/h**	**ft^3/linear ft**
3.2	$\frac{1}{8}$	0.635-1.016	0.0250-0.0400	103-159	15-23	560-815	22-32	1.13-1.56	40-55	0.34-0.36	0.20-0.25	7-9	0.054).06
6.4	$\frac{1}{4}$	0.787-1.511	0.0310-0.0595	76-241	11-35	510-710	20-28	1.27-2.63	45-93	0.34-0.66	0.23-0.31	8-11	0.07-0.08
9.5	$\frac{3}{8}$	0.787-1.511	0.0310-0.0595	117-276	17-40	485-660	19-26	2.32-3.26	82-115	0.86-0.89	0.25-0.34	9-12	0.08-0.09
13	$\frac{1}{2}$	0.787-1.511	0.0310-0.0595	138-379	20-55	430-610	17-24	2.97-3.54	105-125	1.04-1.24	0.28-0.37	10-13	0.11-0.12
19	$\frac{3}{4}$	0.965-1.511	0.0380-0.0595	165-345	24-50	380-560	15-22	3.31-4.50	117-159	1.45-1.56	0.34-0.42	12-15	0.14-0.16
25	1	1.181-1.511	0.0465-0.0595	193-379	28-55	355-485	14-19	3.68-4.93	130-174	1.83-1.86	0.37-0.45	13-16	0.17-0.19
38	$1\frac{1}{2}$	1.702-2.057	0.0670-0.0810	172-379	25-55	305-380	12-15	5.24-6.79	185-240	3.20	0.40-0.51	14-18	0.23-0.24
50	2	1.702-2.057	0.0670-0.0810	152-415	22-60	255-355	10-14	5.24-7.36	185-260	3.70-3.72	0.45-0.57	16-20	0.29-0.32
75	3	2.057-2.184	0.0810-0.0860	228-345	33-50	205-280	8-11	6.79-9.40	240-332	6.00-6.04	0.51-0.65	18-23	0.42-0.45
100	4	2.057-2.184	0.0810-0.0860	290-415	42-60	165-230	6.5-9	8.29-10.9	293-384	8.53-9.02	0.59-0.74	21-26	0.58-0.65
125	5	2.057-2.184	0.0810-0.0860	365-448	53-65	140-190	5.5-7.5	9.82-11.6	347-411	10.97-12.62	0.65-0.82	23-29	0.77-0.84
150	6	2.489-2.527	0.0980-0.0995	310-448	45-65	115-165	4.5-6.5	11.3-13.9	400-490	15.10-17.78	0.74-0.91	26-32	0.98-1.16
205	8	2.489-2.527	0.0980-0.0995	415-620	60-90	94-124	3.7-4.9	14.3-17.7	505-625	25.52-27.30	0.88-1.10	31-39	1.59-1.68
255	10	2.527-2.794	0.0995-0.1100	517-620	75-90	74-102	2.9-4.0	17.3-21.2	610-750	37.50-42.10	1.05-1.27	37-45	2.25-2.55
305	12	2.794-3.048	0.1100-0.1200	476-724	69-105	61-89	2.4-3.5	20.4-24.9	720-880	49.70-60.00	1.19-1.47	42-52	2.97-3.50

(A) VALUES DO NOT NECESSARILY VARY IN EXACT PROPORTION TO PLATE THICKNESS, BECAUSE STRAIGHT LINE RELATIONS DO NOT EXIST AMONG PRESSURE, SPEED, AND ORIFICE SIZES.

(B) LOWEST SPEEDS AND HIGHEST GAS CONSUMPTIONS ARE FOR INEXPERIENCED OPERATORS, SHORT CUTS, DIRTY OR NONUNIFORM MATERIAL. HIGHEST SPEEDS AND LOWEST GAS CONSUMPTIONS ARE FOR EXPERIENCED OPERATORS, LONG CUTS, CLEAN AND UNIFORM MATERIAL.

(C) PRESSURE OF ACETYLENE FOR THE PREHEATING FLAMES IS MORE A FUNCTION OF TORCH DESIGN THAN OF THE THICKNESS OF THE PART BEING CUT. FOR ACETYLENE PRESSURE DATA. SEE CHARTS FROM MANUFACTURERS OF APPARATUS.

Natural Gas

The composition of natural gas varies, depending on its composition at the well, but its main component is methane (CH_4). One volume of methane requires two volumes of oxygen. The combustion ratio is 2 to 1.

Although available in pressurized cylinders, natural gas usually is supplied through low-pressure lines from a local utility. Torchmakers overcome low line pressure through torch design. A siphon stream in the mixing chamber pulls the gas in at the required rate for complete combustion. Torches operate at pressures as low as 2 kPa (4 ozf/in.^{2}) gage line pressure.

A natural gas flame is more diffuse than an acetylene flame--heat intensity is lower, and adjustment for reducing (carburizing), neutral, and oxidizing flames is less clearly defined. Preheating time is longer.

Because heat content is low and heat emission is diffuse, natural gas cannot be used to weld steel. Consequently, if cutting and welding must be performed with a single fuel gas, natural gas cannot be used.

Despite these disadvantages, use of natural gas for cutting has increased. It is the lowest cost commercial fuel gas and, with careful torch adjustment, produces excellent cuts in light- to heavy-gage material. Data for manual oxynatural gas straight-line cutting of carbon steel plate of varying thickness are given in Table 3. Table 4 gives data for oxynatural gas shape cutting of carbon steel plate. Table 5 gives data for machine oxynatural gas drop cutting of shapes from thick carbon steel plate.

TABLE 3 RECOMMENDED PARAMETERS FOR MANUAL OXYNATURAL GAS STRAIGHT-LINE CUTTING OF CARBON STEELS

PLATE THICKNESS		DIAMETER OF CUTTING ORIFICE$^{(A)}$		OXYGEN PRESSURE		MINIMUM NATURAL GAS PRESSURE		CUTTING SPEED$^{(B)}$		APPROXIMATE GAS CONSUMPTION			
										OXYGEN		NATURAL GAS	
mm	in.	mm	in.	kpa	psi	kpa	psi	mm/min	in./min	m^3/h	ft^3/h	m^3/h	ft^3/h
3.2	$\frac{1}{8}$	1.17	0,046	103	15	21	3	510-710	20-28	0.71	25	0.23	8
6.4	$\frac{1}{4}$	1.17	0.046	124	18	21	3	455-710	18-28	0.99	35	0.28	10
9.5	$\frac{3}{8}$	1.17	0.046	138	20	21	3	405-510	16-20	1.27	45	0.40	14
13	$\frac{1}{2}$	1.50	0.059	207	30	21	3	330-430	13-17	1.42	50	0.51	18
19	$\frac{3}{4}$	1.50	0.059	241	35	21	3	255-380	10-15	2.12	75	0.57	20
25	1	1.50	0.059	275	40	21	3	230-330	9-13	2.83	100	0.68	24
38	$1\frac{1}{2}$	1.70	0.067	275	40	21	3	180-305	7-12	4.10	145	0.74	26
50	2	1.70	0.067	310	45	21	3	150-255	6-10	5.38	190	0.79	28
64	$2\frac{1}{2}$	1.70	0.067	345	50	21	3	150-230	6-9	6.94	245	0.85	30
75	3	2.36	0.093	345	50	21	3	125-205	5-8	7.64	270	0.91	32
100	4	2.36	0.093	379	55	21	3	125-180	5-7	9.06	320	1.02	36
125	5	2.36	0.093	414	60	21	3	100-	4-6	11.3	400	1.13	40

								150					
150	6	2.79	0.110	414	60	21	3	100-150	4-6	13.3	470	1.36	48
180	7	2.79	0.110	483	70	21	3	75-125	3-5	14.7	520	1.47	52
205	8	2.79	0.110	552	80	21	3	75-100	3-4	16.4	580	1.59	56
255	10	2.79	0.110	620	90	21	3	75-100	3-4	24.1	850	1.70	60
305	12	2.79	0.110	690	100	21	3	50-75	2-3	28.3	1000	1.81	64

(A) USING INJECTOR-TYPE TORCH AND TWO-PIECE TIPS.

(B) VARIATIONS IN CUTTING SPEEDS MAY BE CAUSED BY MILL SCALE ON PLATE, VARIATION IN OXYGEN PURITY, FLAME ADJUSTMENT, CONDITION OF EQUIPMENT, IMPURITIES IN STEEL, AND VARIATION IN HEAT CONTENT OF NATURAL GAS.

TABLE 4 RECOMMENDED PARAMETERS FOR MACHINE OXYNATURAL GAS SHAPE CUTTING OF CARBON STEELS

PLATE THICKNESS		DIAMETER OF CUTTING ORIFICE(A)		HIGH PREHEAT				LOW PREHEAT				NATURAL GAS		CUTTING SPEED(B)		WIDTH OF KERF (APPROX)	
				OXYGEN		NATURAL GAS		OXYGEN		NATURAL GAS							
mm	in.	mm	in.	kpa	psi	kpa	psi	kpa	psi	pa	ozf/in.^{2}	kpa	psi	mm/min	in./min	mm	in.
6.4	$\frac{1}{4}$	0.91	0.036	193	28	21	3	48-83	7-12	430	1	482-517	70-75	455-710	18-28	2.0	0.08
9.5	$\frac{3}{8}$	0.94	0.037	193	28	21	3	48-83	7-12	430	1	482-551	70-80	455-660	18-26	2.0	0.08
13	$\frac{1}{2}$	0.99	0.039	234	34	21	3	48-83	7-12	430	1	517-551	75-80	405-610	16-24	2.3	0.09
16	$\frac{5}{8}$	0.99	0.039	234	34	21	3	48-83	7-12	430	1	517-551	75-80	405-585	16-23	2.3	0.09
19	$\frac{3}{4}$	1.17	0.046	317	46	21	3	55-83	8-12	430	1	551-586	80-85	405-560	16-22	2.5	0.10
25	1	1.17	0.046	331	48	21	3	55-83	8-12	430	1	551-586	80-85	355-510	14-20	2.5	0.10
32	$1\frac{1}{4}$	1.37	0.054	303	44	21	3	55-83	8-12	430	1	551-586	80-85	355-560	14-18	3.0	0.12
38	$1\frac{1}{2}$	1.37	0.054	303	44	21	3	55-83	8-12	430	1	620-655	90-95	330-455	13-18	3.0	0.12
44	$1\frac{3}{4}$	1.37	0.054	310	45	21	3	55-83	8-12	430	1	620-655	90-95	305-430	12-17	3.0	0.12
50	2	1.37	0.054	310	45	21	3	55-83	8-12	430	1	690	100	255-380	10-15	3.0	0.12
57	$2\frac{1}{4}$	1.40	0.055	317	46	21	3	55-83	8-12	430	1	690	100	230-380	9-15	3.3	0.13
64	$2\frac{1}{2}$	1.40	0.055	317	46	21	3	55-83	8-12	430	1	690	100	205-355	8-14	3.3	0.13
70	$2\frac{3}{4}$	1.70	0.067	317	46	21	3	55-83	8-12	430	1	690	100	205-330	8-13	3.6	0.14
75	3	1.70	0.067	317	46	21	3	55-83	8-12	430	1	690	100	180-330	7-13	3.6	0.14
89	$3\frac{1}{2}$	1.85	0.073	345	50	21	3	69-97	10-14	430	1	724	105	150-305	6-12	3.8	0.15
100	4	1.85	0.073	345	50	21	3	69-	10-	430	1	758	110	150-280	6-11	3.8	0.15

								97	14								
114	$4\frac{1}{2}$	2.08	0.082	345	50	21	3	69-97	10-14	430	1	758	110	125-255	5-10	4.3	0.17
125	5	2.08	0.082	345	50	21	3	69-97	10-14	430	1	793	115	125-255	5-10	4.3	0.17
140	$5\frac{1}{2}$	2.44	0.096	345	50	21	3	69-97	10-14	430	1	793	115	125-230	5-9	4.6	0.18
150	6	2.44	0.096	345	50	21	3	69-97	10-14	430	1	827	120	125-230	5-9	4.6	0.18
165	$6\frac{1}{2}$	2.44	0.096	345	50	21	3	69-97	10-14	430	1	827	120	100-205	4-8	4.6	0.18
190	$7\frac{1}{2}$	2.44	0.096	345	50	21	3	69-97	10-14	430	1	827	120	100-205	4-8	4.6	0.18
205	8	2.44	0.096	345	50	21	3	69-97	10-14	430	1	827	120	75-180	3-7	4.6	0.18

(A) TWO-PIECE TIPS FOR HIGH-SPEED MACHINE CUTTING.

(B) VARIATIONS IN CUTTING SPEED MAY BE CAUSED BY MILL SCALE ON PLATE, VARIATION IN OXYGEN PURITY, FLAME ADJUSTMENT, CONDITION OF EQUIPMENT, IMPURITIES IN STEEL, AND VARIATION IN HEAT CONTENT OF NATURAL GAS.

TABLE 5 TYPICAL PARAMETERS FOR MACHINE OXYNATURAL GAS DROP CUTTING OF SHAPES FROM LOW-CARBON STEEL PLATE 255 TO 545 MM (10 TO $21\frac{1}{2}$ IN.) THICK

Torch constructed with two-piece recessed tips that have milled preheat flutes (heavy preheat) and straight-bore cutting-oxygen orifices

PLATE THICKNESS		CUTTING ORIFICE DIAMETER		PREHEAT				CUTTING OXYGEN$^{(A)}$		CUTTING SPEED		CUTTING OXYGEN	
				OXYGEN		NATURAL GAS							
mm	in.	mm	in.	kpa	psi	kpa	psi	kpa	psi	mm/min	in./min	m^3/h	ft^3/h
255	10	6.35	0.250	240-310	35-45	34	5	186	27	82.6	3.25	39.6	1400
325	$12\frac{3}{4}$	7.14	0.281	240-310	35-45	34	5	193	28	82.6	3.25	49.5	1750
395	$15\frac{1}{2}$	7.14	0.281	240-310	35-45	34	5	207	30	76.2	3.0	53.8	1900
455	18	7.92	0.312	240-310	35-45	34	5	193	28	108	4.25	63.6	2245

535	21	7.92	0.312	240-310	35-45	34	5	193	28	76.2	3.0	63.6	2245
545	21 $\frac{1}{2}$	7.92	0.312	240-310	35-45	34	5	207	30	95.3	3.75	67.0	2365

(A) PRESSURE MEASURED AT TORCH INLET

Oxyfuel Gas Cutting

Methylacetylene-Propadiene-Stabilized Gas

The common trade name of methylacetylene-propadiene-stabilized gas is MAPP gas. Methylacetylene-propadiene-stabilized mixtures are by-products of the manufacture of chemicals such as ethylene. These mixtures combine the qualities of an acetylene flame with a more even heat distribution in a fuel that is less prone to explosion and less costly than acetylene. It is supplied as a liquid, in large tanks or in portable cylinders.

Methylacetylene-propadiene-stabilized gases contain a mixture of several hydrocarbons, including propadiene (allene), propane, butane, butadiene, and methylacetylene. Methylacetylene, like acetylene, is a high-energy triple-bond compound. It is unstable, but other compounds in the mixture dilute it sufficiently to enable safe handling.

Compositions of MPS mixtures are proprietary and may vary; consequently, an exact combustion equation cannot be specified. The mixture bums hotter than propane or propylene. It also affords a high release of heat energy in the primary flame cone, characteristic of acetylene. The outer flame gives relatively high heat release, similar to propane and propylene. The overall heat distribution in the flame is the most even of any of the other gases. The inner cone releases 19.3 MJ/m^3 (517 Btu/ft^3), and the outer flame releases 70.37 MJ/m^3 (1889 Btu/ft^3). The coupling distance is therefore less exacting than for acetylene. The best coupling distance for MPS fuel places the outer cone on the plate; however, a shorter coupling distance also delivers considerable heat.

The neutral MPS gas-oxygen flame generates 89.62 MJ/m^3 (2406 Btu/ft^3) with a 2925 °C (5300 °F) flame with a ratio of 3.5 to 4 parts oxygen to 1 part fuel. Values vary with the composition of the gas. The carburizing flame, 2.2 parts or less oxygen to 1 part fuel, can weld alloys, such as aluminum, which oxidize readily. A neutral flame, with a ratio of 2.3 parts line oxygen to 1 part fuel gas, can weld steel. At 2.8 line oxygen ratio, the flame becomes oxidizing, unsuitable for welding. Methylacetylene-propadiene-stabilized gas produces its hottest flame at 3.3 line oxygen-to-gas ratio.

In comparing the cost of MPS gas with the cost of acetylene, differences in cylinder yield and consumption rate must be considered. A 54 kg (120 lb) cylinder of MPS gas yields 17.5 m^3 (620 ft^3) of gas; a 115 kg (240 lb) cylinder of acetylene yields only 7.4 m^3 (260 ft^3) of gas. In addition, acetylene bums faster. Thus, storage, transportation, and time and labor for changing cylinders become important cost factors.

Methylacetylene-propadiene-stabilized gas competes with acetylene for almost every job that uses fuel gas. Its most unusual use perhaps is in deep water cutting. Because acetylene outlet pressure is limited to 105 kPa gage pressure (15 psig), it cannot be used below 9 m (30 ft) of water. The Navy and several shipyards use MPS gas for underwater work. Table 6 provides data for oxy/MPS gas cutting of carbon steel plate.

TABLE 6 RECOMMENDED PARAMETERS FOR OXY/MPS GAS CUTTING OF CARBON STEELS

PLATE THICKNESS		CUTTING TIP NO.(A)	CUTTING SPEED		OXYGEN								MPS GAS					
					CUTTING PRESSURE(B)		CUTTING RATE OF FLOW		PREHEAT PRESSURE		PREHEAT RATE OF FLOW		CUTTING PRESSURE		CUTTING RATE OF FLOW		KERF WIDTH	
mm	in.		mm/min	in./min	kpa	psi	m^3/h	ft^3/h	kpa	psi	m^3/h	ft^3/h	kpa	psi	m^3/h	ft^3/h	mm	in.
CUTTING WITH STANDARD-PRESSURE TIPS																		
3.2	$\frac{1}{8}$	75	760-915	30-36	275 345	40-50	0.344).42	12-15	35-70	5-10	0.20-0.71	7-25	14-70	2-10	0.06-0,28	2-10	0,64	0.025
4.7	$\frac{3}{16}$	72	660-815	26-32	275-345	40-50	0.57-0.85	20-30	35-70	5-10	0.20-0.71	7-25	14-70	2-10	0.06-0.28	2-10	0.76	0.03
6.4	$\frac{1}{4}$	68	610-760	24-30	275-345	40-50	0.85-1.13	30-40	35-70	5-10	0.20-0.71	7-25	14-70	2-10	0.06-0.28	2-10	1.02	0.04
13	$\frac{1}{2}$	61	560-710	22-28	275-345	40-50	1.56-1.84	55-65	35-70	5-10	0.34-0,71	12-25	14-70	2-10	0.14-0.28	5-10	1.27	0.05
19	$\frac{3}{4}$	56	405-560	16-22	275-345	40-50	1.70-2.12	60-75	35-70	5-10	0.34-0.71	12-25	14-70	2-10	0.14-0.28	5-10	1.52	0.06
25	1	56	355-510	14-20	275-345	40-50	1.70-2.12	60-75	35-70	5-10	0.34-0.71	12-25	14-70	2-10	0.14-0.28	5-10	1.52	0.06
32	$1\frac{1}{4}$	54	330-430	13-17	345-415	50-60	2.97-3.40	105-120	70-140	10-20	0.57-0.99	20-35	14-70	2-10	0.23-0.42	8-15	2.03	0.08
38	$1\frac{1}{2}$	54	305-405	12-16	345-415	50-60	2.97-3.40	105-120	70-140	10-20	0.57-0.99	20-35	14-70	2-10	0.23-0.42	8-15	2.03	0.08
50	2	52	255-355	10-14	345-415	50-60	4.10-5.38	145-190	70-140	10-20	0.57-0.99	20-35	14-70	2-10	0.23-0.42	8-15	2.29	0.09
64	$2\frac{1}{2}$	48	230-330	9-13	345-415	50-60	5.95-7.50	210-265	70-205	10-30	0.57-1.42	20-50	41-70	6-10	0.23-0.57	8-20	2.54	0.10
75	3	48	205-330	8-13	345-415	50-60	5.95-7.50	210-265	70-205	10-30	0.57-1.42	20-50	41-70	6-10	0.23-0.57	8-20	2.54	0.10
100	4	46	180-305	7-12	415-485	60-70	8.21-9.34	290-330	70-205	10-30	0.71-1.42	25-50	41-70	6-10	0.28-0.57	10-20	3.81	0.15
125	5	46	150-255	6-10	485-550	70-80	9.34-11.46	330-405	70-205	10-30	0.71-1.42	25-50	41-70	6-10	0.28-0.57	10-20	3.81	0.15
150	6	42	125-205	5-8	415-485	60-70	10.62-13.31	375-470	70-205	10-30	0.71-1.42	25-50	41-105	6-15	0.28-0.57	10-20	4.06	0.16
205	8	35	100-	4-7	415-	60-	13.73-	485-	205-	30-	1.13-	40-	70-	10-	0.57-	20-	4.83	0.19

			180		485	70	16.70	590	345	50	2.83	100	105	15	1.27	45		
255	10	30	75-150	3-6	275-485	40-70	14.16-17.69	500-625	205-345	30--50	1.13-2.83	40-100	70-105	10-15	0.57-1.27	20-45	5.08	0.20
305	12	30	75-125	3-5	345-585	50-85	18.26-24.49	645-865	205-345	30-50	1.70-4.25	60-150	70-105	10-15	0.85-1.70	30-60	5.33	0.21
CUTTING WITH HIGH-SPEED TIPS																		
3.2	$\frac{1}{8}$	75	815-965	32-38	415-485	60-70	0.57-0.71	20-25	35-70	5-10	0.20-0.71	7-25	14-70	2-10	0.08-0.28	3-10	0.64	0.025
4.7	$\frac{1}{16}$	72	710-815	28-32	485-550	70-80	0.85-1.13	30-40	35-70	5-10	0.20-0.71	7-25	14-70	2-10	0.08-0.28	3-10	0.76	0.03
6.4	$\frac{1}{4}$	68	660-815	26-32	485-550	70-80	1.56-1.84	55-65	35-70	5-10	0.20-0.71	7-25	14-70	2-10	0.08-0.28	3-10	1.27	0.05
13	$\frac{1}{2}$	61	610-760	24-30	550-620	80-90	2.12-2.70	75-95	35-70	5-10	0.34-0.71	12-25	14-70	2-10	0.14-0.28	5-10	1.52	0.06
19	$\frac{3}{4}$	56	510-660	20-26	550-620	80-90	3.26-3.68	115-130	35-70	5-10	0.34-0.71	12-25	14-70	2-10	0.14-0.28	5-10	1.78	0.07
25	1	56	455-610	18-24	550-620	80-90	3.26-3.68	115-130	35-70	5-10	0.34-0.71	12-25	14-70	2-10	0.14-0.28	5-10	1.78	0.07
32	$1\frac{1}{4}$	54	405-510	16-20	485-550	70-80	4.39-4.81	155-170	70-140	10-20	0.57-0.99	20-35	14-70	2-10	0.23-0.42	8-15	2.03	0.08
38	$1\frac{1}{2}$	54	380-485	15-19	550-620	80-90	4.81-5.10	170-180	70-140	10-20	0.57-0.99	20-35	14-70	2-10	0.23-0.42	8-15	2.03	0.08
50	2	52	355-455	14-18	550-620	80-90	6.09-7.22	215-255	70-140	10-20	0.57-0.99	20-35	14-70	2-10	0.23-0.42	8-15	2.29	0.09
64	$2\frac{1}{2}$	52	305-430	12-17	550-620	80-90	6.09-7.22	215-255	70-140	10-20	0.57-0.99	20-35	14-70	2-10	0.23-0.42	8-15	2.29	0.09
75	3	48	255-380	10-15	550-620	80-90	9.48-11.32	335-400	70-140	10-20	0.57-0.99	20-35	41-70	6-10	0.28-0.42	10-15	2.54	0.10
100	4	46	230-355	9-14	550-620	80-90	10.61-12.03	375-425	70-140	10-20	0.57-0.99	20-35	41-70	6-10	0.28-0.42	10-15	3.05	0.12

Note: All recommendations are for straight-line cutting with a three-hose torch perpendicular to work.

(A) ALL TIPS ARE OF DESIGN RECOMMENDED BY THE SUPPLIER.

(B) PRESSURE OF CUTTING OXYGEN MEASURED AT THE TORCH.

Effect of Oxyfuel Cutting on Base Metal

During cutting of steel, the temperature of a narrow zone adjacent to the cut face is raised considerably above the transformation range. As the cut progresses, the steel cools through this range. Cooling rate depends on the heat conductivity and the mass of the surrounding material, on loss of heat by radiation and convection, and on speed of cutting. When steel is at room temperature, the rate of cooling at the cut is sufficient to produce a quenching effect on the cut edges, particularly in heavier cuts in large masses of cold metal. Depending on the amount of carbon and alloying elements present, and on the rate of cooling, pearlitic steel transforms into structures ranging from spheroidized carbides in ferrite to harder constituents. The HAZ may be 0.8 to 6.4 mm ($\frac{1}{32}$ to $\frac{1}{4}$ in.) deep for steels 9.5 to 150 mm ($\frac{3}{8}$ to 6 in.) thick. Approximate depths of the HAZ in oxyfuel gas cut carbon steels are given in Table 7. Some increase in hardness usually occurs at the outer margin of the HAZ of nearly all steels.

TABLE 7 APPROXIMATE DEPTHS OF HAZ IN OXYGEN CUT CARBON STEELS

PLATE THICKNESS		HAZ DEPTH	
MM	IN.	MM	IN.
LOW-CARBON STEELS			
<13	$<\frac{1}{2}$	<0.8	$<\frac{1}{32}$
13	$\frac{1}{2}$	0.8	$\frac{1}{32}$
150	6	3.2	$\frac{1}{8}$
HIGH-CARBON STEELS			
<13	$<\frac{1}{2}$	<0.8	$<\frac{1}{32}$
13	$\frac{1}{2}$	0.8-1.6	$\frac{1}{32}-\frac{1}{16}$
150	6	3.2-6.4	$\frac{1}{8}-\frac{1}{4}$

Note: The depth of the fully hardened zone is considerably less than the depth of the HAZ. For most applications of gas cutting, the affected metal does not have to be removed.

Low-Carbon Steel. For steels containing 0.25% C or less cut at room temperature, the hardening effect usually is negligible, though at the upper carbon limit it may be significant if subsequent machining is required. Short of preheating or annealing the workpiece, hardening may be lessened by ensuring that (1) the cutting flame is neutral to slightly oxidizing, (2) the flame is burning cleanly, and (3) that the inner cones of the flame are at the correct height. By increasing the machining allowance slightly, the first cut usually can be made deep enough to penetrate below the hardened zone in most steels. Mechanical properties of low-carbon steels generally are not adversely affected by OFC. Typical data for OFC of low-carbon steel plate are given in Tables 1, 2, 3, 4, 5, and 6.

Medium-Carbon Steels. Steels having carbon contents of 0.25 to 0.45% are affected only slightly by hardening caused by OFC. Up to 0.30% C, steels with very low alloy content show some hardening of the cut edges, but generally not enough to cause cracking. Over 0.35% C, preheating to 260 to 315 °C (500 to 600 °F) is needed to avoid cracking. All medium-carbon steels should be preheated if the gas cut edges are to be machined.

High-Carbon and Alloy Steels. Oxyfuel gas cutting of steels containing over 0.45% C and of hardenable alloy steels at room temperature may produce a thin layer of hard, brittle material on the cut surface that may crack from the stress of cooling. Preheating and annealing may alleviate hardening and the formation of residual stress.

Preheating to 260 to 315 °C (500 to 600 °F) is sufficient for high-carbon steels; alloy steels may require preheating as high as 540 °C (1000 °F). Preheat temperature should be maintained during cutting. Thick preheated sections should be cut as soon as possible after the piece has been withdrawn from the furnace.

Local preheating heats the volume of the workpiece enclosing the HAZ of the cut. If the volume of material to be heated is small, the flame of a cutting. torch can be used for preheating. When the workpiece is thick and broad, a special heating torch may be necessary. Workpieces must be heated uniformly through the section to be cut, without excessive temperature gradient.

Distortion, which is the result of heating by the gas flame, can cause considerable damage during (1) cutting of thin plate (less than 7.9 mm, or $\frac{5}{16}$ in., thick), (2) cutting of long narrow widths, (3) close-tolerance profile cutting, and (4) cutting of plates that contain high residual stresses. The heat may release some of the restraint to locked-in stress, or may add new stress. In either case, deformation (warpage) may occur, thus causing inaccurate finished cuts. Plates in the annealed condition have little or no residual stress.

Deformation. In cuts made from large plates, the cutting thermal cycle changes the shape of narrow sections and leaves residual stress in the large section (see Fig. 2). The temperature gradient near the cut is steep, ranging from melting point at the cut to room temperature a short distance from it. Plate does not return to its original shape unless the entire plate is uniformly heated and cooled.

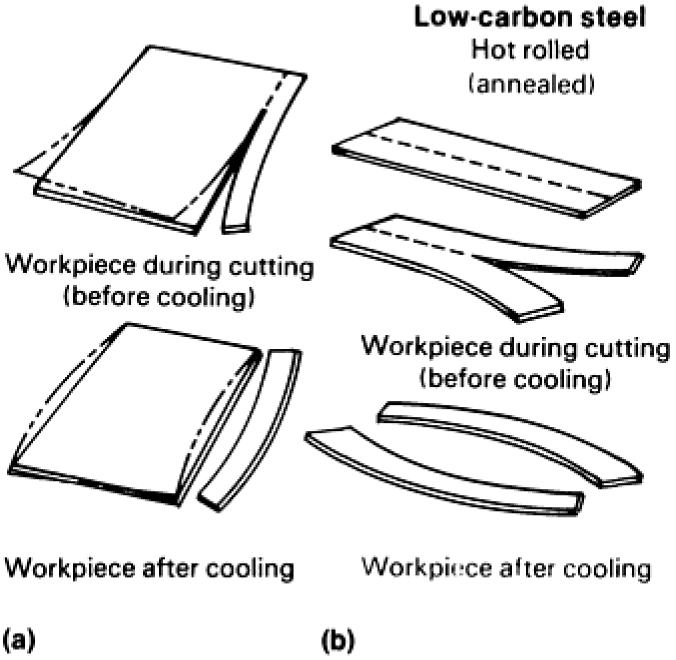

FIG. 2 EFFECTS OF OFC THERMAL CYCLE ON SHAPE OF SECTIONS. (A) PLATE WITH LARGE RESTRAINT ON ONE SIDE OF KERF, LITTLE RESTRAINT ON THE OTHER SIDE. PHANTOM LINES INDICATE DIRECTION OF RESIDUAL STRESS THAT WOULD CAUSE DEFORMATION EXCEPT FOR RESTRAINT. (B) PLATE WITH LITTLE RESTRAINT ON EITHER SIDE

As the metal heats, it expands and its yield strength decreases; the weakened heated material is compressed by the surrounding cooler, stronger metal. The hotter metal continues to expand elastically in all directions until its compressive yield strength is reached, at which point it yields plastically in directions not under restraint. The portion of this upset metal above 870 °C (1600 °F) is virtually stress-free; the remainder is under compressive stress that is equal to its yield strength. Metal that expands but does not upset is under compressive strength below yield. The net stress on the heated side of the neutral axis causes bowing of a narrow plate during cutting, as shown in Fig. 2.

As the heated metal begins to cool, it contracts, and its strength increases. First, the contraction reduces the compressire stress in the still-expanded metal. When the compressive stress reaches zero and the plate regains its original shape, previously upset metal also has regained strength. This metal is now in tension as it cools, and its tensile yield strength increases. Tension increases until the metal reaches room temperature. Residual tensile stress in the cooling side of the neutral axis causes the bowing of narrow plates after cooling (see Fig. 2). Controlled upsetting is the basis of flame straightening.

Control of Distortion. Preheating the workpiece can reduce distortion by reducing differential expansion, thereby decreasing stress gradients. Careful planning of the cutting sequence also may help. For example, when trimming opposite sides of a plate, both sides should be cut in the same direction at the same time. When cutting rings, the inside diameter should be cut first; the remaining plate restrains the material for the outside-diameter cut. In general, the larger portion of material should be used to retain a shape for as long as possible; the cutting sequence should be balanced to maintain even heat input and resultant residual stresses about the neutral axis of plate or part.

Oxyfuel Gas Cutting

Equipment

Commercial gases usually are stored in high-pressure cylinders. Natural gas--primarily methane--is supplied by pipeline from gas wells. The user taps into local gas lines. Acetylene, dissolved in acetone, is available in clay-filled cylinders. High-volume users often have acetylene generators on site. For heavy consumption or where many welding and cutting stations use fuel gas, banks of gas cylinders are maintained at a central location in the plant, and the gas is manifolded and piped to the point of use.

Manual gas cutting equipment consists of gas regulators, gas hoses, cutting torches, cutting tips, and multipurpose wrenches. Auxiliary equipment may include a hand truck, tip cleaners, torch ignitors, and protective goggles. Machine cutting equipment varies from simple rail-mounted "bug" carriages to large bridge-mounted torches that are driven by computer-directed drives.

Gas regulators reduce gas pressure and moderate gas flow rate between the source of gas and its entry into the cutting torch to deliver gas to the cutting apparatus at the required operating pressure. Gas enters the regulating device at a wide range of pressures. Gas flows through the regulator and is delivered to the hose-torch-tip system at the operating pressure, which is preset by manual adjustment at the regulator and at the torch. When pressure at the regulator drops below the preset pressure, regulator valves open to restore pressure to the required level. During cutting, the regulator maintains pressure within a narrow range of the pressure setting.

Regulators should be selected for use with specific types of gas and for specific pressure ranges. Portable oxyacetylene equipment requires an oxygen regulator on the oxygen cylinder and an acetylene regulator on the acetylene cylinder, which are not interchangeable.

High-low regulators conserve preheat oxygen when natural gas or liquefied petroleum gas (LPG) is the preheat fuel used in OFC. These gases require a longer time to start a cut than acetylene or MPS. High-low regulators reduce preheat flow to a predetermined level when the flow of cutting oxygen is initiated. When the regulator switches from high to low, preheat cut-back may range from 75 to 25% as plate thicknesses increase from 9.5 to 205 mm ($\frac{3}{8}$ to 8 in.). High-low regulators are used for manual and automatic cutting with natural gas and with LPG.

Hose. Flexible hose, usually 3.2 to 13 mm ($\frac{1}{8}$ to $\frac{1}{2}$ in.) in diameter, rated at 1.4 MPa gage pressure (200 psig) maximum, carries gas from the regulator to the cutting torch. Oxygen hoses are green; the fittings have right-hand threads. Fuel gas hoses are red; the fittings have left-hand threads and a groove cut around the fittings. For heavy cutting, two oxygen hoses may be necessary--one for preheat and one for cutting oxygen. Multi-torch cutting machines often have three-hose torches.

Cutting torches, such as the one shown in Fig. 3, control the mixture and flow of preheat oxygen and fuel gas and the flow of cutting oxygen. The curing torch discharges these gases through a cutting tip at the proper velocity and flow rate. Pressure of the gases at the torch inlets, as well as size and design of the cutting tip, limits these functions, which are operator controlled.

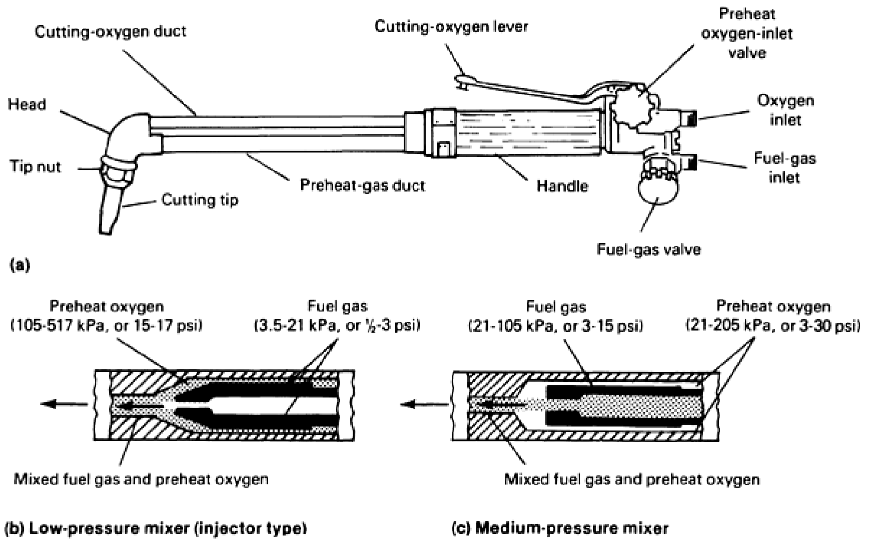
Cutting-oxygen duct
Cutting-oxygen lever
Preheat oxygen-inlet valve
Head
Oxygen inlet
Tip nut
Fuel-gas inlet
Cutting tip
Preheat-gas duct
Handle
Fuel-gas valve
(a)
Preheat oxygen (105-517 kPa, or 15-17 psi)
Fuel gas (3.5-21 kPa, or ½-3 psi)
Fuel gas (21-105 kPa, or 3-15 psi)
Preheat oxygen (21-205 kPa, or 3-30 psi)
Mixed fuel gas and preheat oxygen
Mixed fuel gas and preheat oxygen
(b) Low-pressure mixer (injector type)
(c) Medium-pressure mixer

FIG. 3 TYPICAL MANUAL CUTTING TORCH IN WHICH PREHEAT GASES ARE MIXED BEFORE ENTERING TORCH HEAD. (B) AND (C) SECTIONS THROUGH PREHEAT GAS DUCT SHOWING TWO TYPES OF MIXERS COMMONLY USED WITH THE TORCH SHOWN. AFTER THE WORKPIECE IS SUFFICIENTLY PREHEATED, THE OPERATOR DEPRESSES THE LEVER TO START THE FLOW OF CUTTING OXYGEN. VALVES CONTROL THE FLOW OF OXYGEN AND FUEL GAS TO ACHIEVE REQUIRED FLOW AND MIXTURE AT THE CUTTING TIP.

Oxygen inlet control valves and fuel gas inlet control valves permit operator adjustment of gas flow. Fuel gas flows through a duct and mixes with the preheat oxygen; the mixed gases then flow to the preheating flame orifices in the cutting tip. The oxygen flow is divided--a portion of the flow mixes with the fuel gas, and the remainder flows through the cutting oxygen orifice in the cutting tip. A lever-actuated valve on the manual torch starts the flow of cutting oxygen; machine cutting starts the oxygen from a panel control.

Fuel gases supplied at low pressure, such as natural gas tapped from a city line, require an injector-mixer (Fig. 3b) to increase fuel gas flow above normal operating pressures. Optimum torch performance relies on proper matching of the mixer to the available fuel gas pressure.

Cutting tips are precision-machined nozzles, produced in a range of sizes and types. Figure 4(a) shows a single-piece acetylene cutting tip. A two-piece tip used for natural gas (methane) or LPG is shown in Fig. 4(b). A tip nut holds the tip in the torch. For a given type of cutting tip, the diameters of the central hole, the cutting oxygen orifice, and the preheat ports increase with the thickness of the metal to be cut. Cutting tip selection should match the fuel gas: hole diameters must be balanced to ensure an adequate preheat-to-cutting oxygen ratio. Preheat gas flows through ports that surround the cutting oxygen orifice. Smoothness of bore and accuracy of size and shape of the oxygen orifice are important to efficiency. Worn, dirty bores reduce cut quality by causing turbulence in the cutting oxygen stream.

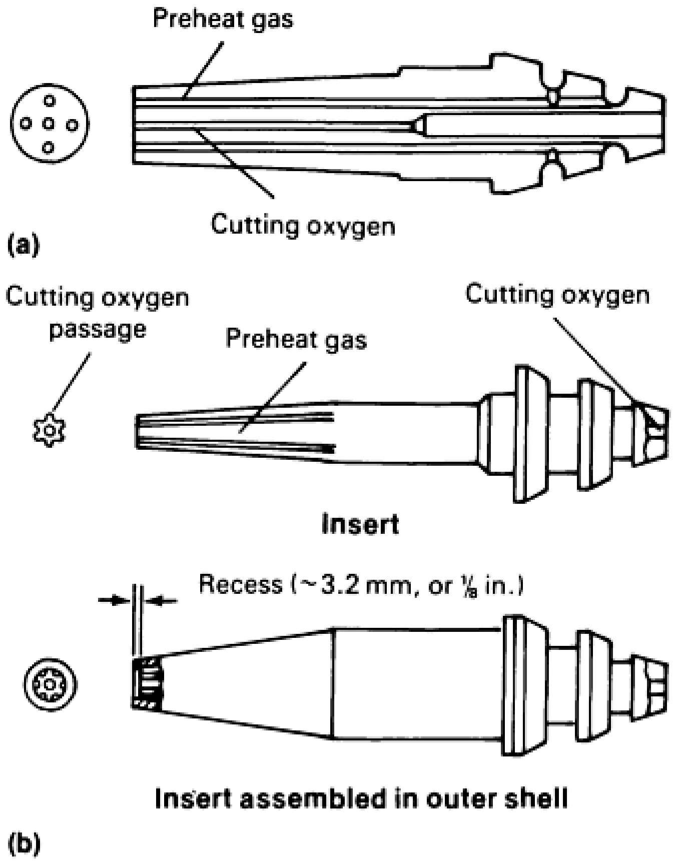

FIG. 4 TYPES OF CUTTING TIPS. (A) SINGLE-PIECE ACETYLENE CUTTING TIP. (B) TWO-PIECE TIP FOR NATURAL GAS OR LPG. FUEL GAS AND PREHEAT OXYGEN MIX IN TIP. RECESSED BORE HELPS PROMOTE LAMINAR FLOW OF GAS.

The size of the cutting tip orifice determines the rate of flow and velocity of the preheat gases and cutting oxygen. Flow to the cutting tip can be varied by adjustment at the torch inlet valve or at the regulator, or both.

Increasing cutting oxygen flow solely by increasing the oxygen pressure results in turbulence and reduces cutting efficiency. Turbulence in the cutting oxygen causes wide kerfs, slows cutting, increases oxygen consumption, and lowers quality of cut. Consequently, larger cutting tips are required for making heavier cuts.

Standard tips, as shown in Fig. 5(a), have a straight-bore oxygen port. Oxygen pressures range from 205 to 415 kPa (30 to 60 psi) and are used for manual cutting. High-speed tips, or divergent cutting tips (Fig. 5b), use a converging, diverging orifice to achieve high gas velocities. The oxygen orifice flares outward. High-speed tips operate at cutting oxygen pressures of about 670 kPa (100 psi) and provide cutting jets of supersonic velocity. These tips are precision made and are more costly than straight-drilled tips, but they produce superior results--improved edge quality and cutting speeds 20% higher than standard tips. Best suited to machine cutting, high-speed tips produce superior cuts in plate up to about 150 mm (6 in.) thick. Above this thickness, advantages of their use decrease; they are not recommended for cutting metal more than 255 mm (10 in.) thick.

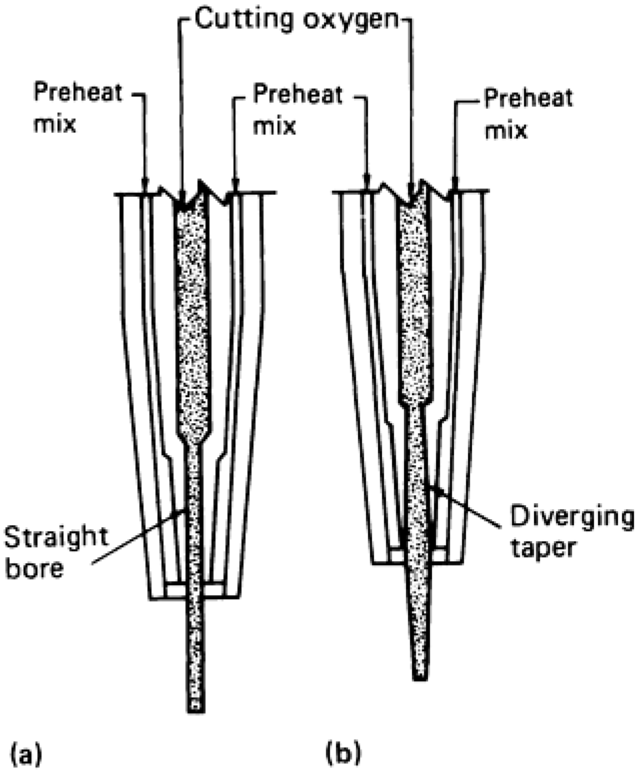

FIG. 5 OXYFUEL CUTTING TIPS. (A) STANDARD CUTTING TIP WITH STRAIGHT-BORE OXYGEN ORIFICE. (B) HIGH-SPEED CUTTING TIP WITH DIVERGENT-BORE OXYGEN ORIFICE

Equipment Selection Factors. Natural gas and liquefied petroleum gases operate most efficiently with high-low gas regulators, injector-type cutting torches, and two-piece, divergent, recessed cutting tips. Acetylene cutting is most efficient with divergent single-piece tips. If acetylene is supplied by low-pressure generators, an injector-type torch is best suited to most cutting applications.

Two-piece, divergent cutting tips are best suited for use with MPS gas; the tip recess should be less than for use of natural gas or propane. Injector-type torches and high-low regulators are not required with MPS gas.

Starting the Cut

To start a cut at the outer edge of the workpiece, the operator should place the cutting torch so that the ends of the inner cones in the preheating flame just clear the work metal. When a spot of metal at the top of the edge is heated to bright red, the stream of cutting oxygen is turned on. The reactions form a slot in the plate edge, and the torch can move along the desired line of cut. Piercing starts are necessary to cut holes, slots, and shapes.

In manual piercing, the spot should be heated to bright red. Then the torch should be raised to 13 mm ($\frac{1}{2}$ in.) above the normal cutting distance, and cutting oxygen should be turned on slowly. As soon as the metal is penetrated, the torch is lowered to cutting height and cutting action is started.

Machine torch piercing is similar to manual operations, except that torch travel begins after the cutting oxygen is slowly turned on and the flame starts to eject molten slag. Initial piercing action does not completely penetrate the plate, so that the completed pierce covers a short distance (depending on plate thickness) and forms a sloping trough with molten slag ejected from the torch. Length of the lead-in should be increased for heavy plate.

Light Cutting

Light cutting (material less than 9.5 mm, or $\frac{3}{8}$ in., thick) requires extra care in tip selection, tip cleanliness, and control of gas pressure. Quality of cut when oxyfuel gas cutting steel thinner than 3.2 mm ($\frac{1}{8}$ in.) is virtually impossible to control. Wide, uneven kerf, ragged edge, and distortion result from heat buildup.

For straight or large-radius cuts in thicknesses 3.2 mm ($\frac{1}{8}$ in.) and over, angling the cutting torch forward increases the distance through the cut and may help to deflect some of the heat input. When cutting shapes, however, the torch must be vertical.

One of the most important techniques for cutting of thin sections is stack cutting (see the section of this article on stack cutting for more detailed information on this process). This technique avoids the difficulties inherent in light cutting and also increases production. Sections stacked for cutting must be flat and tightly clamped together.

Oxyfuel Gas Cutting

Medium Cutting

In cutting plate 9.5 to 255 mm ($\frac{3}{8}$ to 10 in.) thick, OFC achieves its highest efficiency and produces the most satisfactory cuts. Conditions for medium cutting of carbon steel plate using a variety of oxyfuel combinations are presented in Tables 1, 2, 3, 4, 5, and 6. Commercial cuts must meet these requirements:

- DRAG IS SHORT, MARKINGS ON THE FACE OF THE CUT APPROACH VERTICAL.
- SIDE OF THE CUT IS SMOOTH, NOT FLUTED, GROOVED, OR RAGGED.
- SLAG SHOULD NOT ADHERE TO THE BOTTOM OF THE CUT.
- UPPER AND LOWER EDGES SHOULD BE SHARP.
- COST IS MODERATE.

The following variables require adjustment to obtain satisfactory commercial cuts:

- SUITABLE CUTTING TIP, WITH CUTTING ORIFICE OF CORRECT TYPE AND SIZE, AND PROPER DEGREE OF PREHEAT
- SUITABLE OXYGEN AND FUEL GAS PRESSURES
- CORRECT CUTTING SPEED
- UNIFORM TORCH MOVEMENT
- CLEAN, SMOOTH-BORE CUTTING ORIFICE AND PREHEAT HOLES
- HIGH-PURITY OXYGEN
- PROPER ANGLE OF THE CUTTING JET IN RELATION TO THE UPPER EDGE OF THE CUT

Surface finish for gas cut production parts is compared with sample cuts. One such group of machine cut samples is shown in Fig. 6. Sample A represents high quality, with drag varying about 2 to 7%. Samples B, C, D, and E show typical defects resulting from improper cutting speed or torch settings. The American Welding Society (AWS) provides standard plastic samples of cut edges.

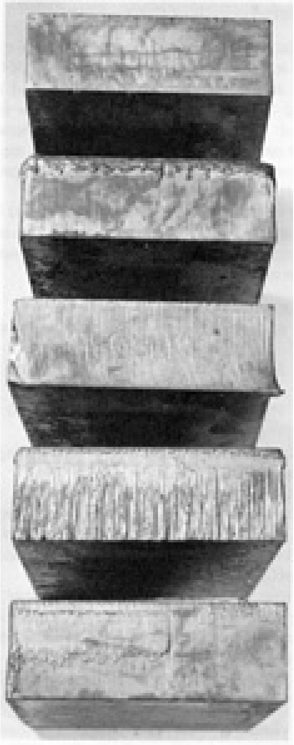

TORCH SETTINGS FOR BAR A[(A)]	
PREHEAT OXYGEN PRESSURE:	
HIGH (EDGE STARTING)	310 KPA (45 PSI)
LOW (CUTTING)	55-83 KPA (8-12 PSI)
PREHEAT NATURAL GAS PRESSURE:	
HIGH (EDGE CUTTING)	21 KPA (3 PSI)
LOW (CUTTING)	UNDER 430 PA (1 OZ/IN.^{2})
CUTTING OXYGEN PRESSURE	670 KPA (100 PSI)
CUTTING SPEED	255-380 MM/MIN (10-15 IN./MIN)

(A) SETTINGS FOR BARS B, C, D, AND E WERE VARIED AS NOTED IN THE LEGEND ABOVE. ALL BARS WERE CUT WITH TWO-PIECE DIVERGENT-NOZZLE (HIGH-SPEED) TIPS WITH 1.37 MM (0.054 IN.) DIAM CUTTING OXYGEN ORIFICE. ALL CUTTING WAS DONE WITH LOW PREHEAT SETTINGS, EXCEPT FOR BAR B, WHICH WAS CUT WITH BOTH PREHEAT

FIG. 6 COMPARISON OF SURFACE FINISH OF GAS CUT SPECIMENS. (A) PROPER SPEED, PREHEAT, AND CUTTING-OXYGEN PRESSURE. NOTE CLEAN FACE AND NEARLY STRAIGHT DRAG LINES. (B) PROPER SPEED AND CUTTING-OXYGEN PRESSURE, TOO MUCH PREHEAT. NOTE EXCESSIVE SLAG AND ROUNDING OF TOP EDGE. (C) PROPER PREHEAT AND CUTTING-OXYGEN PRESSURE, BUT TOO MUCH SPEED. NOTE INCREASE IN DRAG AND UNCUT CORNER AT LOWER LEFT. (D) PROPER SPEED AND PREHEAT, BUT TOO MUCH CUTTING-OXYGEN PRESSURE. SURFACE IS ROUGH AND TOP EDGE IS MELTED. (E) PROPER PREHEAT AND CUTTING-OXYGEN PRESSURE, BUT SPEED TOO LOW. BURNED SLAG ADHERES TO CUT SURFACE.

Tip Design. General-purpose cylindrical-bore tips operate best with cutting oxygen pressures of 205 to 415 kPa (30 to 60 psi); a pressure of 275 kPa (40 psi) is a good initial setting. Divergent-orifice tips operating with cutting oxygen pressures of about 690 kPa (100 psi), with variations of ± 140 kPa (±20 psi) may provide better results. Because tip size controls kerf width, a smaller tip should be used for a narrower kerf; the part may not drop free, and slower speed must be used. Larger-than-normal tips may be used when the metal is covered with thick scale.

Cutting Speed. Within recommended speed ranges, slow cutting produces the best results; higher speeds are used for maximum output. Optimum cutting speed may be determined by observing the material that is expelled from the underside of the workpiece. When speed is too slow, the cutting stream forms slag drippings. At excessive speed, a stream trails at a sharp angle to the bottom of the workpiece. Highest quality cutting occurs when a single, continuous stream is expelled from the underside for about 6.4 mm ($\frac{1}{4}$ in.), then breaks into a uniform spray. This condition is accompanied by a sound similar to that of canvas ripping.

Preheat. Less preheat is needed to continue a cut than to start it. When optimum quality is desired, the amount of preheat should be reduced after starting the cut. Commercially available devices automatically reduce preheat after cutting begins.

In thicknesses up to 50 mm (2 in.), excessive preheat may cause the top corner to melt and roll over. Insufficient preheat results in loss of cut. Correct preheat provides a continuous reaction with a sharp, clean top edge of the cut.

Kerf compensation, or the allowance of half the kerf width on the outline of the template, is the amount added for outside cuts and subtracted for inside cuts. In straight-line cutting or in following a layout, the operator may compensate for this measurement; in machine cutting, kerf compensation should be incorporated in the template. Many automatic machines incorporate kerf compensation in their controllers.

Machine Accuracy. When a machine is properly maintained, the guidance system controls accuracy. Mechanical followers are very accurate when properly adjusted. Compensation for the diameter of the follower wheel or spindle must be made in the same manner as for kerf width when making a template.

Kerf angle, the small deviation of the kerf wall from a right angle, may be caused by failure to set the torch perpendicular to the work. It may also be the result of widening of the cutting oxygen stream. In the latter case, the width of the kerf increases from the top to the bottom of the cut; this may be corrected in straight cutting by angling the torch slightly to make one wall of the kerf perpendicular. In shape cutting, where this cannot be done, kerf angle can be corrected by slightly reducing cutting oxygen pressure.

Plate movement may be vertical, caused by warp, or lateral, caused by planar expansion and contraction of the part. Vertical movement requires either manual or automatic control of torch height. A very slight allowance can be made in the initial setting of the tip-to-work distance. Lateral movement can be controlled by fixing the part to be cut on the plate support, wedging the cut within the stock plate, or allowing only a small amount of scrap trim around the part so that the scrap will move instead of the workpiece.

Oxyfuel Gas Cutting

Heavy Cutting

Cutting of metal 255 mm (10 in.) or more in thickness requires attention to gas flow, preheat flame setting, drag, and starting technique. Standard cutting torches may be used to cut steel up to 455 to 510 mm (18 to 20 in.) thick; heavy-duty torches are necessary for sections up to 1525 mm (60 in.) thick.

Gas Flow Requirements. Heavy cutting requires a uniform supply of high-purity oxygen and fuel gas at constant pressure, and an adequate volume of cutting oxygen. Torches should have three hoses--one for the cutting oxygen, one for the preheat oxygen, and one for the fuel gas. This allows setting of proper pressures independently for preheating and cutting. Cutting oxygen hose should be at least 13 mm ($\frac{1}{2}$ in.) in diameter to ensure an adequate volume of oxygen at low pressure settings. Regulators should be used that are capable of providing the large volumes of gas required. Cutting oxygen flow requirements (ft^3/h) vary from 80 to 120 times ($80t$ to $120t$) the thickness to be cut measured in inches. Cutting oxygen consumption for shape cutting approaches $120t$ or higher; for straight-line cutting, it is between $80t$ and $100t$. Pressures are 140 to 415 kPa (20 to 60 psi) at the torch.

Preheating. In heavy cutting, the preheat flame must extend almost to the bottom of the thickness to be cut to avoid excessive drag. If the cutting torch cannot deliver enough heat, an extra torch may be needed, or the workpiece may be furnace heated. A temperature boost of 95 to 150 °C (200 to 300 °F) assists cutting action considerably. Tip height above the work is important; if the tip is too close, excessive melting and rounding of the top edge occur. If too high, preheat will be insufficient and cutting will be slowed. Generally, 25 to 50 mm (1 to 2 in.) of separation is satisfactory.

The starting edge should be thoroughly preheated, with the preheated zone extending far down the face as shown in Fig. 7(a). As soon as the metal begins to melt, the flow of cutting oxygen and the cutting motion should proceed simultaneously, cutting at normal speed. The cut should progress down the face at a constant rate until it breaks through the bottom. Drag will be long at first, but it will shorten as soon as the cut is confined. Too slow a speed causes a shelf part way through the cut edge (Fig. 7e). Too high a speed results in incomplete penetration or in extremely long drag, as shown in Fig. 7(f).

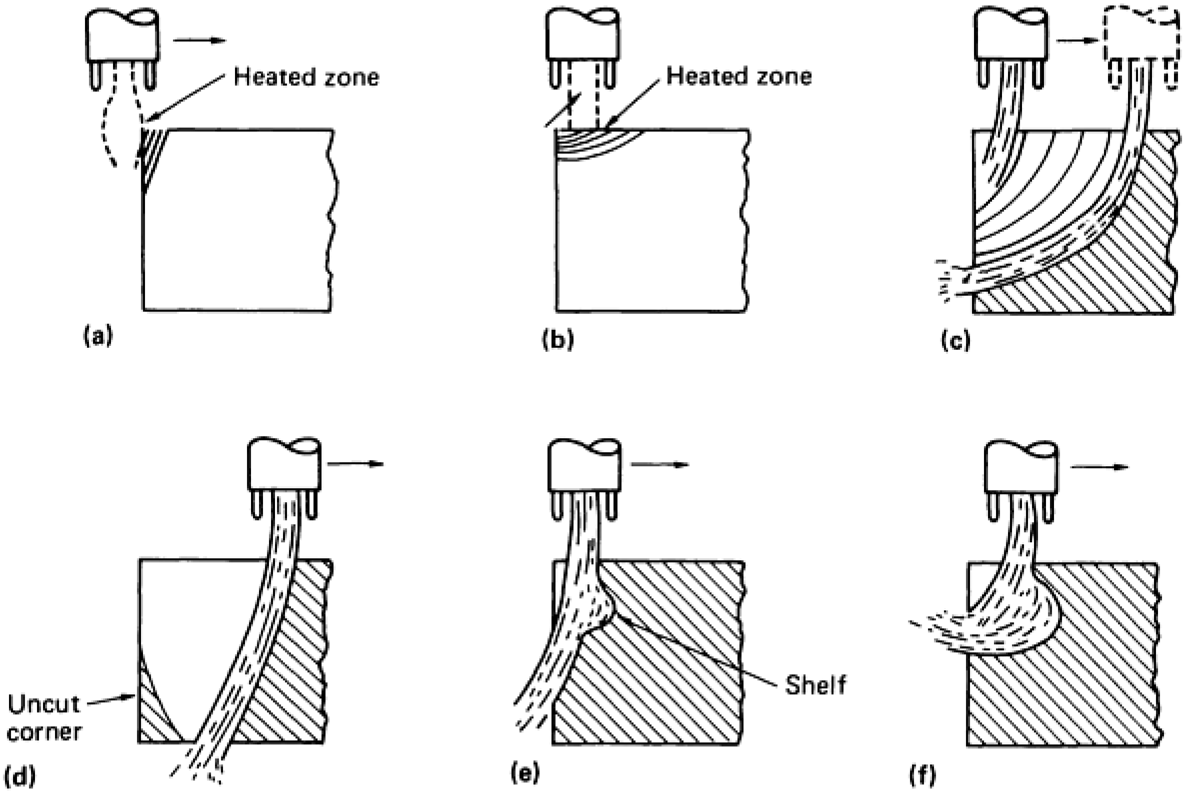

FIG. 7 PROPER AND IMPROPER TECHNIQUES FOR HEAVY CUTTING. (A) PROPER TORCH POSITION; PREHEAT IS PRIMARILY ON STARTING FACE. (B) IMPROPER START; OXYGEN STREAM IS TOO FAR ONTO WORK, WHICH RESULTS IN ACTION OF CUT AS SHOWN IN (C) AND IN UNCUT CORNER AS SHOWN IN (D). (E) EXCESSIVE OXYGEN PRESSURE OR ACTION OF CUT AT TOO LOW A SPEED. (F) INSUFFICIENT OXYGEN PRESSURE OR ACTION OF CUT AT TOO HIGH A SPEED

Drag. To successfully complete the cut (drop cutting), minimum drag conditions must exist, as shown in Fig. 8(a). Too much drag leaves an uncut corner at the end of the cut (Fig. 8b and c). Angled cutting may be used in straight-line work to counter drag (Fig. 8d and e), but for shape cutting, the torch must be perpendicular to the work.

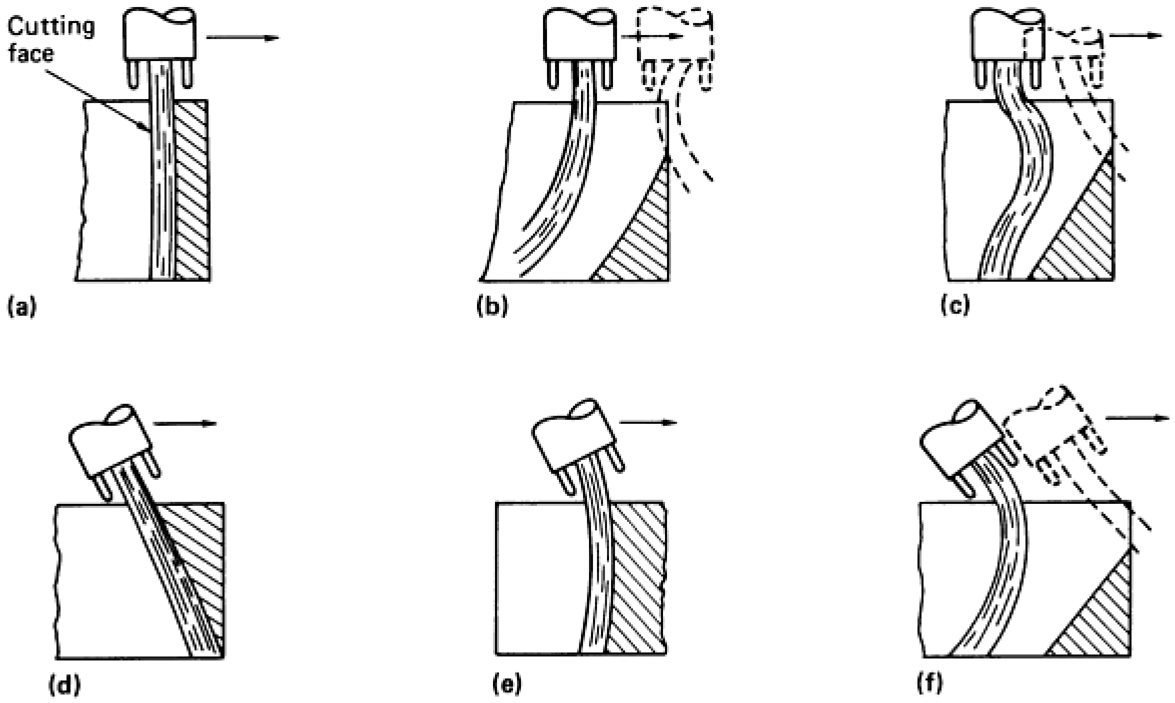

FIG. 8 PROPER AND IMPROPER TECHNIQUES FOR A HEAVY CUT. (A) MINIMUM DRAG (TYPICAL AT BALANCED CONDITIONS) PERMITS FLAME TO BREAK THROUGH CUTTING FACE UNIFORMLY AT ALL POINTS. (B) EXCESSIVE DRAG, TYPICALLY CAUSED BY INSUFFICIENT OXYGEN OR EXCESSIVE SPEED, RESULTS IN UNDERCUT. (C) FORWARD DRAG RESULTING FROM EXCESSIVE OXYGEN PRESSURE OR TOO LITTLE SPEED. (D) AND (E) FORWARD ANGLING OF TORCH TO MINIMIZE DRAG AND THUS AVOID AN UNCUT CORNER. (F) EXCESSIVE FORWARD ANGLING OF TORCH, RESULTING IN UNDERCUT

Starting. Extra care is needed in starting a heavy cut to avoid leaving an uncut corner or pocketing the flame in the lower portion of the cut. Starting sometimes is facilitated by first undercutting the forward edge of the material with a hand torch.

Cutting of Bars and Structural Shapes

To cut round steel bars, gas pressure should be adjusted for the maximum thickness (diameter), and the cut should be started at the outside. As the cut progresses, the torch should be raised and lowered to follow the circumference. The bar should be nicked with a chisel to make a burr at the point where the cut is to begin. Oxyfuel gas cutting of round bars is usually manual, so that the cut surface is rough. The chief advantage of OFC over sawing is that the torch can be brought to the work. The same advantage applies to OFC of various structural shapes.

Close-Tolerance Cutting

Shapes can be oxyfuel gas cut to tolerances of 0.8 mm ($\frac{1}{32}$ in.) in plate up to 64 mm ($2\frac{1}{2}$ in.) thick with multi-torch pantograph or coordinate-drive machines. Conditions must be optimum:

- OPERATORS MUST BE WELL TRAINED AND EXPERIENCED.
- MACHINES MUST BE PROPERLY ADJUSTED SO THAT THEY CAN TRACE AND RETRACE A PATTERN WITH ACCURACY OF ±0.13 MM (±0.005 IN.) IN A GIVEN DIRECTION.
- WORKPIECE MUST BE FLAT AND FREE OF DIRT, SCALE, GREASE, AND OIL.

- WORKPIECE SHOULD BE SUPPORTED OR RESTRAINED TO PREVENT MOVEMENT OF THE SHAPE DURING CUTTING.
- CUTTING TIP MUST BE CLEAN (SELECTED FOR HIGH-QUALITY, NARROW-KERF CUTTING).
- CUTTING SPEED AND GAS PRESSURES SHOULD BE TUNED TO SPECIFIC PRODUCTION CONDITIONS.

Optical tracing devices can follow an ink line 0.64 mm (0.025 in.) wide on paper. If cutting speed does not exceed 760 mm/min (30 in./ min), a template with an outside corner radius of 2.4 mm ($\frac{3}{32}$ in.) will produce a sharp corner. A 2.4 mm ($\frac{3}{32}$ in.) inside radius on the template produces a somewhat larger radius in the workpiece, depending on thickness and cutting tip. Numerical control tracing, whether punched tape or computer assisted, offers the greatest accuracy. Inaccuracies introduced by distortion of templates are avoided, and electronic directions are more exacting than those generated by mechanical cutting machines, which set limits on cutting speed.

Oxyfuel Gas Cutting

Safety*

The hazards of combustion or possible explosion associated with the gases used in OFC, as well as the presence of toxic gases and dust, make it necessary for the user to follow established safety precautions. The three areas of most concern, dealt with in this section, are protective clothing, handling and storage of gas cylinders, and the working environment.

Protective Clothing. Size, nature, and location of the work to be cut dictate the necessary protective clothing. The operator may require some or all of the following:

- TINTED GOGGLES OR FACE SHIELD:

- LIGHT CUTTING, UP TO 25 MM (1 IN.)--SHADE NO. 3 OR 4
- MEDIUM CUTTING, 25 TO 150 MM (1 TO 6 IN.)--SHADE NO. 4 OR 5
- HEAVY CUTTING, THICKER THAN 150 MM (6 IN.)--SHADE NO. 5 OR 6

- HARD HAT AND SOMETIMES A CLOSE-FITTING HAT BENEATH
- SAFETY GLASSES
- FLAME-RESISTANT JACKET, COAT, HOOD, APRON; WEAR WOOLEN CLOTHING, NOT COTTON OR SYNTHETIC FIBER; KEEP SLEEVES AND COLLARS BUTTONED, NO CUFFS; BUTTON-DOWN POCKETS
- PROTECTIVE GLOVES, WRISTLETS, LEGGINGS, AND SPATS. WHEN CUTTING PLATED OR COATED STOCK, RESPIRATORY PROTECTION SHOULD BE USED IF NECESSARY
- CLOTHING SHOULD BE FREE OF GREASE AND OIL, AND FREE OF RAGGED EDGES

Cylinders. Fuel gas cylinders should be placed in a location where they cannot be knocked over. Cylinders should be kept upright with valve ends up to avoid hazardous liquid withdrawal; the protective caps should never be removed except when the cylinders are connected for use. Cylinders should never be exposed to high heat or open flame, which could cause them to explode.

Cylinders should not be placed where they might complete an electric circuit, such as against a welding bench. They should not be used as an electrical ground or as a surface on which to strike an arc.

Damage to cylinders by impact should be avoided. Regulators should be removed before cylinders are moved, except on portable outfits. Slings or magnets should not be used as supports or rollers.

Damaged or defective cylinders should be removed from service immediately, tagged with a statement of the problem, and returned to the supplier. Leaking cylinders should be transported to an isolated location outdoors, the supplier notified, and his instructions for disposal followed.

No attempt should be made to fill a small cylinder from a large one. Filling requires special equipment and training.

Before storing, valves should be closed and caps replaced on empty or unused cylinders. Empty cylinders should be marked accordingly and kept separate from filled ones. Unused oxygen or fuel gas should not be drained from cylinders, and oxygen cylinders should be kept away from fuel gas cylinders.

Working Environment. Torches should not be used around chlorinated solvent vapors; heat forms phosgene and other corrosive, toxic products from them. Adequate ventilation must be provided, especially when alloys containing lead, cadmium, zinc, mercury, beryllium, or other toxic elements are being cut. A flame should not be used in a closed vessel or pipeline that has held flammable or explosive material, such as gasoline. Adequate ventilation should also be provided when work is being performed in a confined area. Under such conditions, the person doing the cutting should have an assistant in case trouble occurs.

Fire prevention warnings in "Safety in Welding and Cutting" (ANSI/AWS 249.1), which have legal status as Occupational Safety and Health Administration (OSHA) regulations, should be heeded. Workers should know the locations and operating procedures of the nearest fire extinguishers.

If cutting is to be done over concrete, it should be protected with metal or another suitable material; concrete spalls explosively when overheated. Hoses should be protected from sparks, hot slag, hot objects, sharp edges, and open flame.

Additional information is available in the article "Safe Practices" in this volume.

Note cited in this section

* *FROM WELDING ENGINEERING DATA SHEET NO. 491, "SAFETY CHECKLIST FOR OXYFUEL GAS CUTTING."

Plasma Arc Cutting

Chuck Landry, Thermal Dynamics Corporation

Introduction

PLASMA ARC CUTTING (PAC) is an erosion process that utilizes a constricted arc in the form of a high-velocity jet of ionized (electrically conductive) gas to melt and sever metal in a narrow, localized area.

Process Description. In the PAC process, a cool and inert gas, such as compressed air, is forced under pressure through a small orifice in the front of the cutting torch. This torch is connected by leads to a direct-current (dc) power supply. In the torch, a portion of the inert gas is changed into a plasma (ionized gas) by heat created by the discharge of a high-voltage arc from the power supply. This arc is created between an electrode (dc negative) in the torch and the tip (nozzle) of the torch through which the gas flows.

When a small amount of dc from the power supply is imposed on this high-voltage arc, it achieves a pilot arc that extends approximately 13 mm ($\frac{1}{2}$ in.), as a plasma jet from the tip orifice. This pilot arc acts as a path that establishes the main cutting arc to the workpiece. When the pilot arc is brought in contact with the metal workpiece, which is connected to the positive side of the dc power supply, the transferred (main) arc occurs and the pilot arc is shut off. The cutting torch is then usually operated with a standoff, that is, the distance between torch and workpiece, of approximately 3 to 10 mm ($\frac{1}{8}$ to $\frac{3}{8}$ in.).

However, in some applications, the torch tip can actually be dragged along the workpiece. The transferred plasma arc heats the metal by the combined effects of the electric arc from the power supply, the high-temperature plasma created by constriction of the arc by the orifice, and the reassociation of the gas molecules at the workpiece. The molten metal is blown out of the cut, or kerf, by the high-velocity, well-columnated plasma jet, resulting in fast, clean cuts.

Plasma Arc Cutting

Chuck Landry, Thermal Dynamics Corporation

Equipment

The required equipment for the PAC process comprises:

- A TORCH DESIGNED FOR EITHER MANUAL OR MECHANIZED OPERATION
- TORCH LEADS THAT CARRY THE ELECTRIC POWER, GAS, AND COOLANT TO THE TORCH
- A COOLANT SYSTEM FOR THE TORCH
- A POWER SUPPLY THAT PROVIDES THE PROPER CURRENT AND VOLTAGE FOR THE PLASMA ARC
- A MANIFOLD SYSTEM FOR COMBINING POWER, GAS, AND COOLANT FOR CONNECTION TO THE TORCH
- A CONTROL SYSTEM THAT CONSISTS OF SWITCHES, GAGES, DIALS, METERS, AND OTHER CONTROLS FOR OPERATING THE PAC SYSTEM AND THE GAS SUPPLY

In its simplest form, a PAC system includes a gas-cooled torch with leads that connect to a power supply.

All PAC torches possess the same basic components (Fig. 1). The tip, or nozzle, of the torch contains the orifice that constricts the arc. The diameter of the orifice, which ranges from 1 to 6.4 mm (0.038 to 0.250 in.), determines the maximum cutting current that can be used with that tip. The larger the orifice, the more current that can be forced through it.

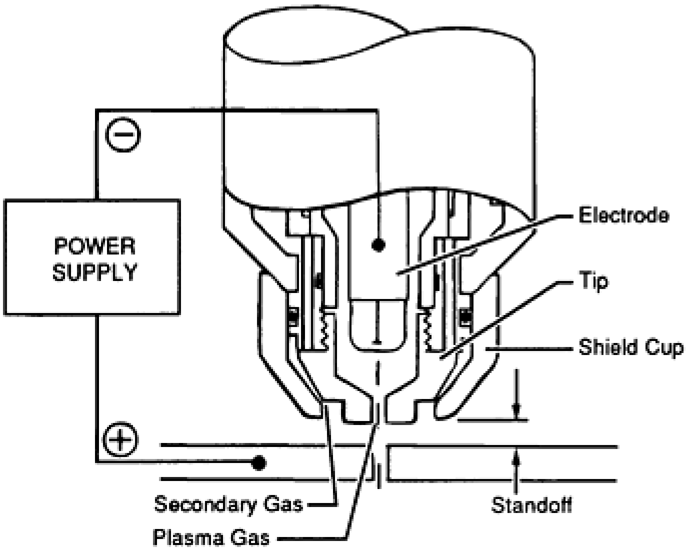

FIG. 1 TYPICAL PAC TORCH

Tips are designed so that the orifice provides maximum arc constriction. If a tip orifice is overpowered, then it will become gouged or washed out, resulting in a change of arc characteristics and in the deterioration of cut quality and cutting speed. To attain optimal cutting conditions, tips are used at or very near their maximum cutting rating. Torch tips are either water cooled, through internal passages connected to a closed-loop coolant system, or indirectly cooled, by good thermal contact with other torch parts that are either water or gas cooled.

The electrode, or cathode, in a PAC torch is the surface from which the plasma arc is generated. This electrode is in electrical contact with the negative potential in the power supply. Both electrode material and design are quite varied. For most applications that use nitrogen or argon-hydrogen as the plasma gas, a 2% thoriated tungsten is used as the electrode. Hafnium and zirconium are commonly used as electrode materials for both air and oxygen plasma cutting.

In terms of its construction, the electrode can be solid, such as a rod of tungsten or a tungsten element held in a solid-copper holder, or it can be hollow and have two pieces (copper and the electrode element) that allow direct water or gas cooling. Because proper cooling will promote both tip and electrode life, PAC torches designed for cutting above 150 A are usually water cooled.

In the area between the electrode and the tip orifice, which is called the plenum, the initial ionization or plasma formation takes place. The plasma gas enters the plenum in either a straight (laminar) flow pattern or a tangentially swirling pattern. Although torch design depends on several factors, pointed electrodes usually utilize laminar flow patterns, whereas flat or blunt-ended electrodes utilize a swirling gas to keep the arc attachment point moving around the center of the electrode. The swirl is most commonly imparted on the gas by a gas distributor, or swirl ring, that is located near the end of the electrode.

Other parts that are common to most PAC torches include internal ceramic or high-temperature plastic insulators for electrical insulation, collets or other mechanisms for retaining the electrode, and shield cups to insulate the front end of the torch from the workpiece.

Leads. The PAC torch is connected to the other system components by means of a leads package that consists of hoses, fittings, and wires. The leads supply the electric power, gases, and cooling medium to the torch. Lead lengths vary from 3 to 68 m (10 to 225 ft), depending on the application. Simple extension kits can be used to join leads packages together for increased length.

Cooling of PAC torches that operate at less than 150 A can be simply accomplished by properly channeling the gases used in torch operation. High-powered systems require a water-based coolant system, which consists of a reservoir, pump, and heat-exchanger assembly. Because this type of coolant is often in contact with both negative and positive electric potentials within the torch, it is important to use a deionized (nonconductive) coolant.

The power supply for a PAC system must have a dc output with the negative side connected to the electrode in the torch and the positive side connected to the workpiece. Systems that include a pilot arc also have a positive connection, limited in current through a resistor, to the tip in the torch. Power supplies are usually of a constant-current, dropping volt-amp curve design, and they have higher open-circuit voltages (up to 400 V) than those found in common welding power supplies, in order to accommodate the high operating voltage (90 to 200 V) of the torches. The amperage output of power supplies can be designed to have one fixed level, or several switchable fixed settings, or an infinitely adjustable output by means of a control potentiometer.

The manifold assembly combines the dc power, gas, coolant, and high-voltage starting circuitry and provides it to the cutting torch via the leads package.

Controls for starting and stopping the cutting operation, selecting the proper amperage output, adjusting gas flows, and regulating other system functions can be located on the manifold box, on the power supply, or in a dedicated control cabinet for convenient mounting near an operator station.

All-in-one packaged PAC systems that combine the coolant system, power supply, manifold assembly, and controls in one unit are available. These systems offer improved portability, simplicity of operation, and reduced costs.

Optional equipment for manual PAC systems includes drag-type shield cups with protrusions, or feet, that enable direct contact with the workpiece to allow the operator to maintain a constant standoff. Wheels that can be attached to the torch for manual guidance, as well as circle-cutting attachments that include an adjustable radius rod to allow the cutting of various-sized holes, are also available.

Optional equipment for mechanized PAC systems includes torch-standoff controls that automatically find and maintain the proper standoff for the cutting operation. Water-shrouded systems are also available. When used in conjunction with water tables, these systems provide a blanket of water around the front of the torch during the cut, in order to minimize smoke, fumes, noise, and arc glare.

Plasma Arc Cutting

Chuck Landry, Thermal Dynamics Corporation

Gases

The plasma or secondary shielding gases that are used in PAC systems are inert, and they are usually supplied from high-pressure cylinders, bulk tanks, or air compressors. Pressure regulators are adjusted to provide from 140 to 560 kPa (20 to 80 psi) of gas, depending on system requirements.

If water is used in place of a secondary gas, then it is usually supplied from a filtered, in-plant water system at a flow rate of 0.63 to 1.9 L/min (10 to 30 gal/h), again, depending on system requirements.

The purity of plasma gases is usually of a welding grade or higher. The plasma gas can be compressed air, which is inexpensive and provides fast, clean, high-quality cuts on carbon steels, but a relatively short electrode life, because of the rapid oxidation that occurs at high temperatures. If purity from an in-plant compressor is a problem, then either charcoal filters that remove oil vapors and moisture traps or dryers can be added into the line. As an alternative, clean, dry, compressed air can be obtained in cylinder form.

Another choice is nitrogen, which is relatively inexpensive and provides good-quality cuts on stainless steels and aluminum. It also provides good electrode life.

Argon-hydrogen, which is usually supplied in a 65 or 70% Ar and a 35 or 30% H mixture, operates at a higher temperature. It can reduce the quantity of smoke and fumes created during the cutting operation and provide a higher-quality surface finish on aluminum and stainless steels that are more than 12 mm (0.5 in.) thick. However, argon-hydrogen is usually from two to three times more expensive than nitrogen. In addition, it is not as readily available and requires more-critical adjustment of the operating parameters.

Oxygen plasma provides the same benefits as air plasma when cutting carbon steels. Because of the absence of nitrogen, it can minimize the amount of nitrides in the cut surface, thereby enabling better-quality welds on "as-cut" pieces that have not undergone additional finishing operations. Like air plasma, oxygen causes more-rapid electrode erosion than does nitrogen or argon-hydrogen plasma.

Plasma Arc Cutting

Chuck Landry, Thermal Dynamics Corporation

Operating Sequence

A routine check of the PAC system before power is applied involves examining the torch to ensure proper assembly, verifying adequate coolant and gas supplies, and inspecting all components of the system to verify that they are in order.

The disconnect switch can then be turned on, which applies power to the system. Gas supplies should be turned on and adjusted according to the recommended pressure or flow. The plasma gas should be allowed to purge for an additional 2 to 3 min in order to eliminate any traces of moisture that accumulated within the plasma lines. This moisture is usually the result of condensation that forms when the system is off for prolonged periods of time.

The "start-cut" signal can be supplied by either an operator who simply pushes a start button on a mechanized control station to energize the switch on a manually operated torch or by a computer-generated signal on a fully automated system. The start-cut signal will cause the main contactor in the power supply to energize and initiate a gas prepurge. The prepurge ensures that gas is at the torch before an arc is struck and can last from 2 to 10 s if extra-long torch leads are used. At the end of the prepurge, the high-frequency arc and pilot arc circuits are energized, establishing the pilot arc at the torch. The pilot arc duration can be controlled by a timing circuit, limiting it to approximately 5 s, or it can be continuous, thereby shutting off only after an actual cut is established. Some PAC systems include circuitry for an automatic restart pilot, which enables automatic cycling between the pilot and cut modes as long as the start signal is present. This type of operation is valuable when cutting expanded metal, grating, or other material with an interrupted cut line.

The cutting arc will begin as soon as the torch and pilot arc are brought close enough to the workpiece (usually, 13 to 25 mm, or $\frac{1}{2}$ to 1 in.) to allow main arc transfer. However, some systems do not use a pilot arc to start the torch. Rather, they only use high frequency, and require that the torch be placed within 1 mm (0.038 in.) to establish a cutting arc. The cutting action will continue as long as there is metal to which the arc can transfer. If the cutting speed is too fast, then the arc will not fully penetrate the metal, and an incomplete cut or gouge will occur. If the cutting speed is too slow, then the kerf will widen as available metal is blown away and, eventually, the arc will be extinguished.

When the "start signal" is removed, or when a stop signal is provided, the cutting arc or pilot arc will shut off and the system will enter a postpurge mode, which is a continuation of plasma gas flow for 2 to 40 s to ensure that the electrode cools in a good atmosphere and to remove any residual heat within the torch.

The system is then ready for the next start signal to begin the cutting sequence again. Some systems include circuitry that allows immediate starting of the pilot or cutting sequence at any time during the postpurge.

For manual operation, the torch is usually held with two hands to provide optimum comfort and speed control. An exception to this rule occurs when small, lightweight torches are used in conjunction with templates or a straight edge that requires one hand to be used for positioning while the other hand operates the torch.

There are two means of starting the cut in manual PAC systems: edge starts and pierce starts. For edge starting, the torch tip is usually placed directly over the edge of the workpiece. Resting the torch at the start position ensures that the cut will begin at the proper location. The torch switch is pressed and the torch is moved along the cut line as soon as the arc is established.

Piercing is best accomplished by slightly angling the torch over the starting point so that the molten metal is directed away from the torch. After starling the arc, the torch is moved to the vertical position and the cut proceeds along the cut line. This technique prevents molten metal from blowing back at the torch and causing damage. Thicker metals may require that the torch be left stationary over the start point for a few seconds to ensure full arc penetration. Because a pierce will sometimes result in a blow hole that is larger than the normal kerf width, it is good practice to start the pierce off the cut line in the scrap piece to eliminate discontinuities in the cut surface of a good piece.

In mechanized applications, edge starts are accomplished by lining the torch up with the cut line. Pierce starts cannot be done in the same way as they are in manual cutting, because the torch cannot be conveniently angled and then brought back to a vertical position. Best results are achieved with a slow-moving pierce, in which the torch motion begins when the arc is struck. This allows the molten metal to flow back along the cut line until full penetration is achieved. Some allowance should be made for the start of a moving pierce to ensure that the entire piece is cut out.

A third means of starting the arc in mechanized applications is the running-edge start. In this mode, the torch is positioned off the workpiece, the pilot arc is established, and then the torch is put in motion. The cutting arc is established as soon as the pilot hits the edge of the work.

For either manual or mechanized operation, the best cut quality is usually achieved with a torch standoff of 3 to 10 mm ($\frac{1}{8}$ to $\frac{1}{2}$ in.). Thinner metals, under 3 mm ($\frac{1}{8}$ in.) thick, require a closer standoff. At low power levels (below 40 A), good-quality cuts can be obtained by dragging the torch tip in contact with the work.

The proper cutting speed for achieving the best quality is determined by the angle that the arc makes with a vertical plane as it exits the workpiece (Fig. 2). The optimal speed for air cutting is usually achieved when the arc is vertical (0°). Nitrogen and argon-hydrogen cuts are optimized by adjusting speed to obtain a 5 to 10° trailing arc. The best oxygen cuts are made using a speed that provides a slight (1 to 2°) leading arc. The PAC torch should always be kept square with the workpiece to provide maximum arc penetration and to obtain consistent quality when contour cutting.

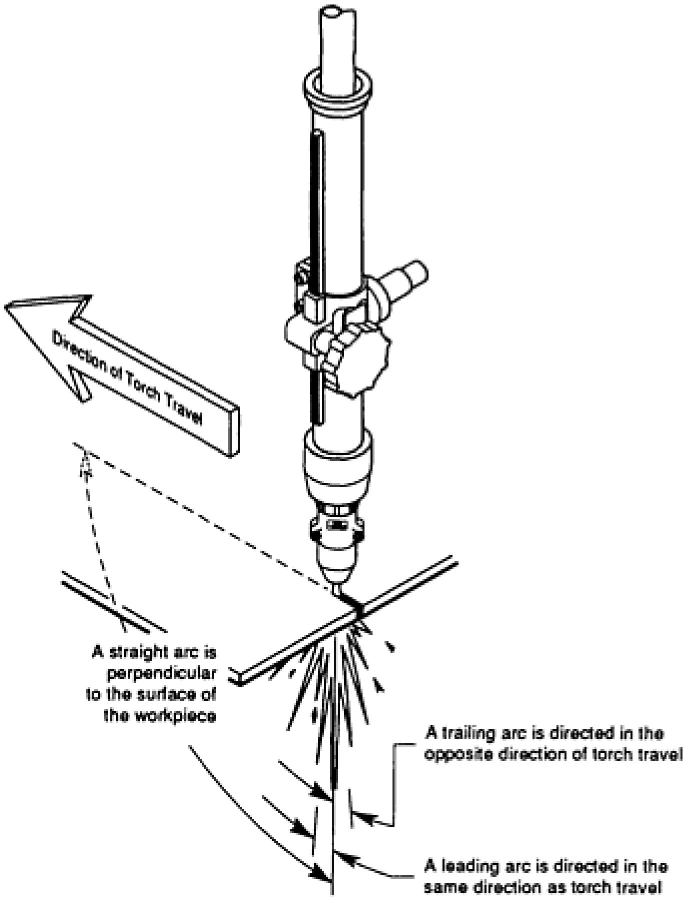
Direction of Torch Travel
A straight arc is perpendicular to the surface of the workpiece
A trailing arc is directed in the opposite direction of torch travel
A leading arc is directed in the same direction as torch travel

FIG. 2 ARC CHARACTERISTICS OF PAC TORCH

Plasma Arc Cutting

Chuck Landry, Thermal Dynamics Corporation

Process Considerations

Process Variations. The major difference between a conventional PAC process, which utilizes the basic system components described earlier, and other process variations is that the conventional PAC torch uses only a single gas, and the orifice in its tip represents the only gas exit. Because these torches do not incorporate any secondary gas, they must be liquid cooled.

Gas-shielded plasma arc cutting utilizes the traditional plasma gases plus a secondary gas that forms a shield around the plasma arc. The torch design provides special passages for this secondary gas. The functions of the secondary gas are to:

- ASSIST THE PLASMA IN BLOWING AWAY MOLTEN METAL
- ENABLE FASTER AND CLEANER CUTS
- PROVIDE BETTER COOLING OF THE TORCH FRONT END
- REDUCE ACCUMULATION OF METAL SPATTER ON THE TORCH
- REDUCE TOP-EDGE ROUNDING OF THE WORKPIECE
- PERMIT EASIER PIERCING
- MINIMIZE DOUBLE ARCING

In gas-cooled, rather than liquid-cooled, torches, it is the secondary gas that provides the cooling action. The choice of secondary gas can also affect cut quality, cutting speed, and the amount of smoke and fumes produced during the PAC process.

Water-shielded plasma arc cutting is a variation of the gas-shielded process that is only used in mechanized applications. The water, which can be provided from standard plant water systems at a rate of 0.63 to 0.95 L/min (10 to 15 gal/h), is substituted for the secondary gas in torches designed for this type of operation. Using a water shield, rather than a gas shield, is less expensive, because water is virtually free. It also results in reduced top-edge rounding, a reduced amount of smoke and fumes, and a clean, shiny, cut surface.

Water-injection plasma arc cutting is similar to the water-shielded process, except that the water is forced into the plasma stream, at approximately 1.26 L/min (20 gal/h), just after the arc exits the tip orifice and just before it leaves the ceramic piece on the front end of the torch (Fig. 3). The impinging water increases the constriction and density of the plasma arc, resulting in improved cut squareness and increased cutting speed. Because of the entry of water at the tip orifice in a water-injection torch, the quality of water is more critical than that used in water-shielded cutting. A plant water system that has high levels of dissolved minerals may require water softeners in conjunction with the system in order to avoid rapid deterioration of part lifetimes. Water-injection cutting is only used in mechanized setups.

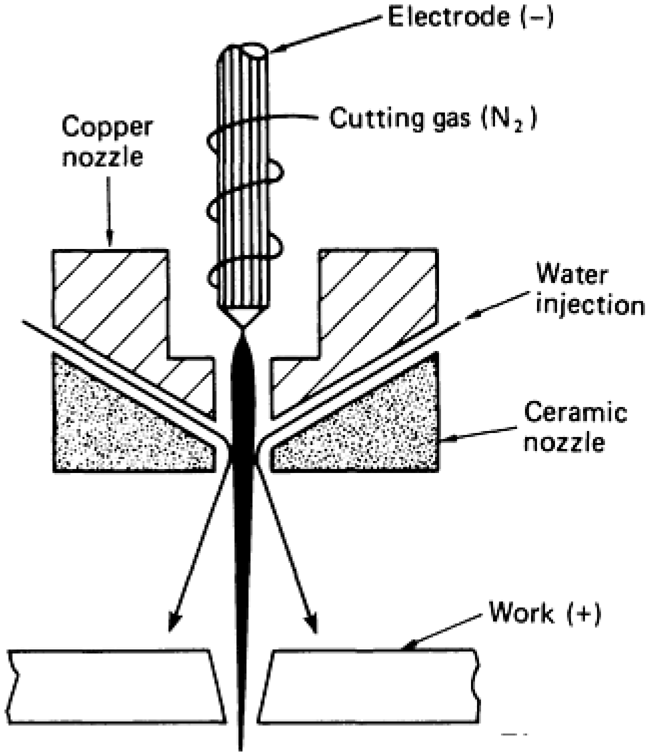

FIG. 3 WATER-INJECTION PAC TORCH

Process Capabilities. Plasma arc cutting is unique in its ability to make fast, clean cuts on both ferrous and nonferrous metals. When compared with the oxyfuel cutting of ferrous metals, the major advantage of this process it its speed, which can be from 4 to 5 times greater than the speed of oxyfuel cutting. For example, when cutting 6 mm ($\frac{1}{4}$ in.) carbon steel with an oxyfuel process, a speed of 0.68 m/min (27 in./min) is used, but when using a 150 A PAC system with air plasma, the speed goes up to 3.81 m/min (150 in./min) (Fig. 4). Because labor and overhead represent the primary costs of operating a mechanized cutting system, reducing the time to cut a part will, in turn, reduce the cost of the part itself. The five-fold increase in cutting speed of the PAC process, versus that of the oxyfuel process on carbon steel, will translate to a part cost of only one-fifth that achieved with the oxyfuel process.

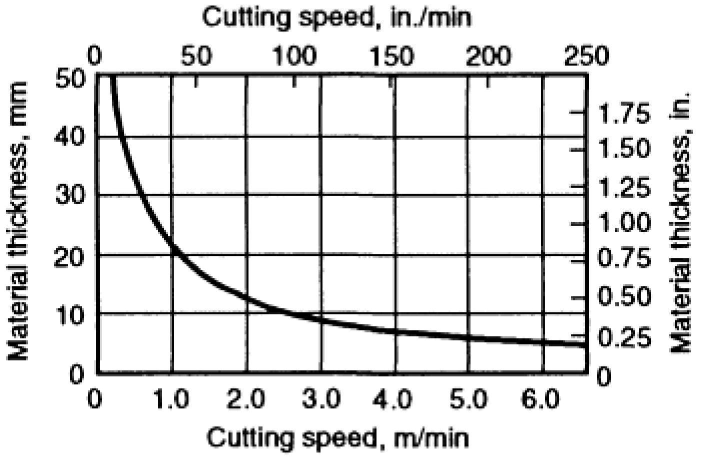

FIG. 4 CUTTING SPEEDS VERSUS MILD STEEL THICKNESS (150 A, AIR PLASMA)

Because the speed of the PAC process, like that of the oxyfuel process, is dependent on metal thickness, increasing thickness will diminish the speed advantage. Therefore, other factors that favor the PAC process have to be determined when comparing it with the oxyfuel process for cutting carbon steel that is more than 50 mm (2 in.) thick.

Another advantage of the PAC process, versus the oxyfuel process, is its ability to cut carbon steel with virtually no distortion, because of the fast travel speed and narrow, localized heating of the metal. This same high-speed localized heating results in a much narrower heat-affected zone (HAZ), that is, the depth of metallurgical change from the cut surface, than that of the oxyfuel process. Typical HAZs of up to 1.78 mm (0.070 in.) from oxyfuel cuts on carbon steel have been measured, whereas the PAC process produces HAZs of only 0.51 mm (0.020 in.). The HAZ will have an effect on the overall strength, corrosion resistance, and susceptibility to surface cracking of the cut part. If the application is such that the cut part must undergo machining to remove the HAZ, then the depth of machining required is much less with the PAC process.

Normally, PAC surfaces do not require any additional finishing operations before they are welded, although this usually depends on the final application and the codes to which the cut part being welded must conform.

The PAC process has the ability to make both edge starts and pierce starts without a preheat, as is required for oxyfuel cutting. This advantage becomes quite dramatic when the application calls for interrupted cuts, such as on grating or expanded metal.

These same capabilities apply to the cutting of stainless steel and aluminum. No preheat is required, and there is no distortion of the cut part. For plasma arc cuts on stainless steel and aluminum, respectively, the HAZ has been measured at 0.25 and 2.03 mm (0.010 and 0.080 in.).

Process Mechanization. Plasma arc cutting can be mechanized in several ways, the simplest of which takes the form of portable carriages or tractors. These devices can travel along a track to achieve straight-line cutting or along an arc guided either by a radius rod or by hand to cut curves. Although tractors are primarily used for simple cut-off, squaring, and trimming operations, they are also used as parts of large fixtures for cutting cast and rolled metals.

Pattern/follower wheel-shaped cutting machines or pantographs are used for the plasma arc cutting of relatively small, simple shapes. A scale model of the desired piece is held and traced by a follower that is connected, through a series of arms, to the torch, which cuts either an identical or a larger-scale part.

Optical tracing or electric-eye machines are used to trace part shapes that are in the form of templates or, more commonly, in the form of line drawings that are the same scale as the desired part. A mechanized PAC setup that uses optical tracing is shown in Fig. 5. The speed of an optical tracing machine is limited by the capability of the electric eye to follow a line. Low-speed machines that operate at speeds up to 1.52 m/min (60 in./min), which were originally designed for use with oxyfuel torches, are compatible with lower-amperage PAC systems. New high-speed machines, with speeds of up to 12.7 m/min (500 in./min), have been developed for operation with high-amperage plasma cutting systems.

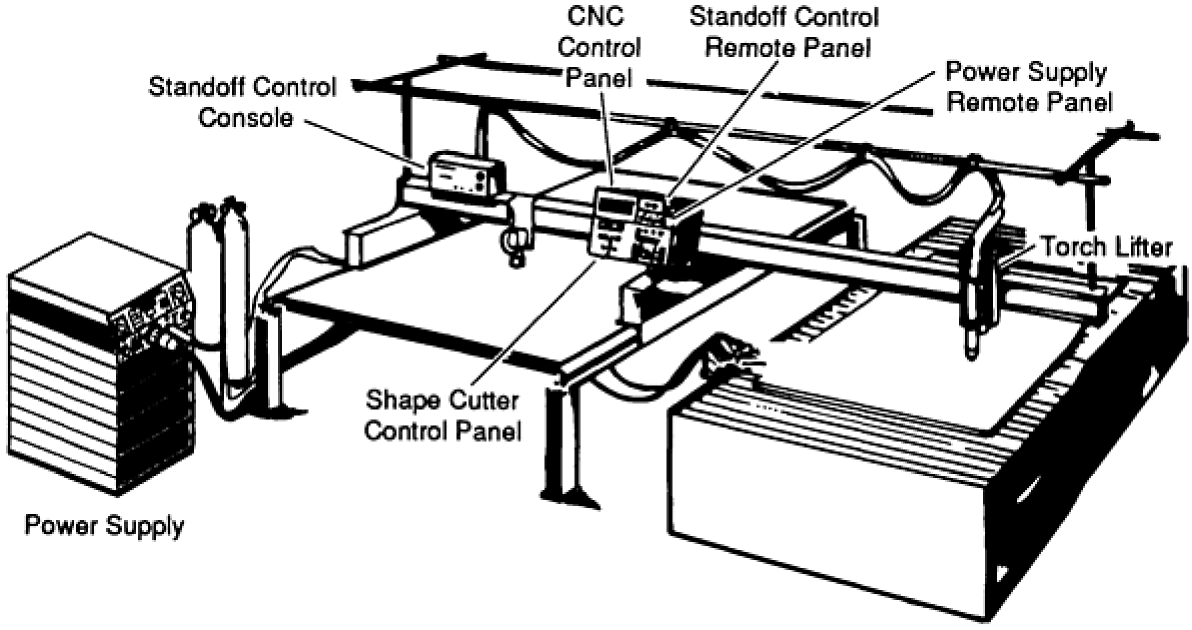
CNC Control Panel
Standoff Control Remote Panel
Standoff Control Console
Power Supply Remote Panel
Torch Lifter
Shape Cutter Control Panel
Power Supply

FIG. 5 MECHANIZED PAC SYSTEM THAT USES OPTICAL TRACING OF PART SHAPES

Computer numerical control (CNC) shape-cutting machines with speeds of up to 25.4 m/min (1000 in./min) represent the state of the art in automated cutting. CNC machines are much more reliable, have greater capabilities, and are lower priced than their numerical (tape) control machine predecessors, which are now considered obsolete. CNC machines can be connected directly into the central processor (direct numerical control) of a plant for complete integration of product design, material requirements, work in process, and other functions. Because the evolution of CNC machines paralleled that of mechanized PAC systems, their combination represents the highest level of productivity available in metal cutting today.

Automated material handling systems are sometimes used to keep up with the speed of CNC PAC systems. Typical systems utilize a raw material loading station, a cutting station, and a finished parts station. While the cutting operation is taking place, the operator is loading raw material and unloading finished parts. Upon completion of the cutting sequence, a conveyor moves the finished parts off the cutting area as it moves raw material onto the cut station. The operator signals the start-cut sequence and the cycle is repeated.

Robots are being used in increasing numbers for PAC operations involving the contour cutting of pipe and vessels, the removal of sprues and risers from castings, and the cutting of shapes in various planes of cast or formed parts. Material handling is also a great concern in robotics applications. Multiple work stations are the rule, because they allow an operator or another robot to load unfinished work and remove finished parts while the cutting operation is ongoing.

Cut Quality. Besides the HAZ, which was discussed above, several other factors (Fig. 6) that also must be considered in determining the quality of a plasma arc cut are the:

- CUT ANGLE
- DROSS
- SURFACE FINISH
- TOP-EDGE SQUARENESS
- KERF WIDTH

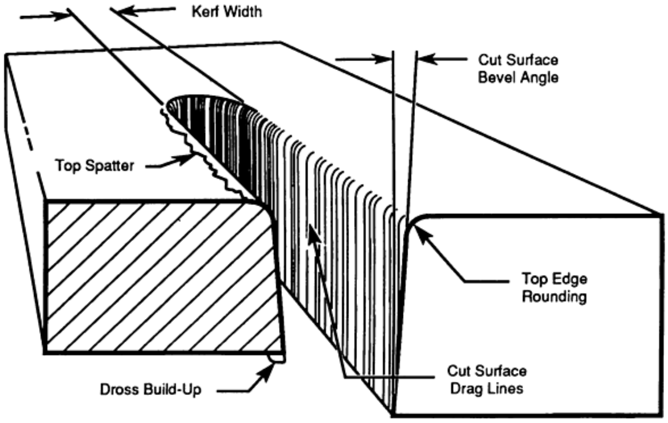

FIG. 6 CHARACTERISTICS OF A PLASMA ARC CUT

Cut Angle. Plasma arc cutting will usually result in an angle on the cut surface of approximately 1 to 3° on the "good" side and 3 to 8° on the "bad" side, when using torches that swirl the plasma. With laminar-flow torches, the angle on both sides is usually about 4 to 8°. These angles are most noticeable in mechanized applications where the torch is square to the workpiece. The good and bad sides of the cut are determined by torch travel direction and plasma gas swirl. With a clockwise gas swirl, the good side of the cut will be on the right with respect to torch motion. Therefore, in the cutting of a ring or flange in which the minimum angle is desired on both the inside and outside diameters, the outside would be cut in the clockwise direction, while the inside would be cut counterclockwise (Fig. 7). This change in angle from good side to bad side of the cut is the result of the arc attachment point on the cut surface and the release of energy on the good side before contact with the bad side occurs.

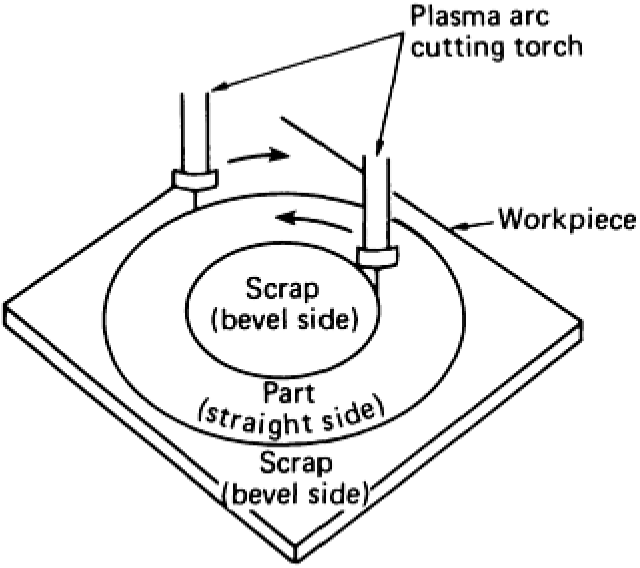

FIG. 7 PAC TORCH DIRECTION THAT ACHIEVES BEST QUALITY

Changing an internal torch part, such as the electrode or gas distributor, can enable those plasma cutting torches that create a gas swirl to change from a clockwise to a counterclockwise swirl direction. This then enables mirror-image cutting when using two torches in a mechanized setup, such as in strip-cutting applications.

Dross is the resolidified metal that adheres to the bottom edge of the plasma cut. The concentration of dross will be heavier on the bad side of the cut. The amount of dross that forms is a result of the type of metal being cut, the cutting speed, and the arc current. Dross can be formed from either too high or too low a cutting speed, but there is a "window" between these two extremes in which dross-free cuts can usually be achieved. The dross-free range is greater on stainless steel and aluminum than it is on carbon steel and copper alloys. If dross-free cuts cannot be achieved, then a minimum amount of low-speed dross is more desirable, because it is more easily removed than the high-speed variety.

Surface Finish. The cut surface from the plasma arc process is normally rougher than that achieved by oxyfuel cutting on carbon steels, and it is definitely more rough than most machining processes. In most metals, this roughness usually appears as a ripple along the cut surface. This is partly due to the output waveform of the dc power supply (the smoother the output, the smoother the cut), but it is also determined to an extent by the gases used and the torch design. Water-shielded and water-injection plasma arc cutting provide much smoother cuts than do gas-shielded or conventional plasma techniques. The use of argon-hydrogen as a plasma or a shielding gas will result in smoother cut surfaces on most metals than when using nitrogen plasma or any other gas shield.

Top-Edge Squareness. Most metals experience some top-edge rounding when the PAC process is applied. Top-edge rounding is more pronounced on the thinner metals. This rounding is due to a higher heat concentration at the top of the cut, and can be minimized by using a gas-shielded PAC process. It can be even better controlled when water-shielded or water-injection plasma arc cutting is used.

The kerf width obtained by plasma arc cutting will be greater than that achieved by oxyfuel cutting on carbon steel, but not as great as that obtained by other processes, such as abrasive cutting or arc gouging. The rule of thumb for estimating the kerf in plasma arc cutting is that its width will be approximately 1.5 to 2 times the tip orifice diameter.

Plasma Arc Cutting

Chuck Landry, Thermal Dynamics Corporation

Applications

Plasma arc cutting is used in a variety of industries. End users of smaller systems with output capabilities of 100 A or less include:

- SMALL JOB SHOPS THAT MUST FABRICATE SPECIALTY AND SUBCONTRACTED ITEMS FROM STAINLESS STEEL, ALUMINUM, BRASS, COPPER, CARBON STEEL, GALVANIZED METAL, AND OTHER METALS
- MANUFACTURERS AND INSTALLERS OF DUCTWORK FOR HEATING, VENTILATING, AND AIR CONDITIONING, WHO MUST CUT GALVANIZED METALS
- MANUFACTURERS OF FOOD PROCESSING AND INSTITUTIONAL KITCHEN EQUIPMENT, WHO MUST CUT STAINLESS STEELS
- AUTOMOBILE MANUFACTURERS, AUTO BODY REPAIR SHOPS, AND DISMANTLERS, WHO MUST CUT HIGH-STRENGTH STEELS USED IN UNIBODY CONSTRUCTION
- MANUFACTURERS AND REBUILDERS OF OVER-THE-ROAD LIQUID TRANSPORT TRAILERS USED TO HAUL PETROLEUM, CHEMICALS, MILK, AND FERTILIZERS, WHO MUST CUT STAINLESS STEELS AND ALUMINUM
- CONTRACTORS AND MAINTENANCE DEPARTMENTS THAT ERECT AND MAINTAIN ALL KINDS OF PLANTS, INCLUDING FOOD PROCESSING, PETROLEUM, CHEMICAL, PULP AND PAPER MILLS, POWER PLANTS, AND MINING OPERATIONS

The end users of PAC systems with an output greater than 100 A include:

- LARGE SHOPS THAT FABRICATE SPECIALTY AND SUBCONTRACTED PARTS FROM VIRTUALLY ANY METAL
- STEEL WAREHOUSES AND SERVICE CENTERS THAT SIZE AND SHAPE-CUT BOTH FERROUS AND NONFERROUS METALS, AS WELL AS THOSE THAT SIZE SCRAP

- PIPE FABRICATORS AND PIPING CONTRACTORS, WHO MUST SIZE, BEVEL, AND MAKE INTERSECTION CUT-OUTS ON CARBON AND STAINLESS STEEL PIPE
- STRUCTURAL FABRICATORS, WHO MUST PERFORM HIGH-SPEED SLITTING OF H AND I BEAMS AND THE SIZING AND SLITTING OF PLATE
- MANUFACTURERS OF SHIPS, BARGES, AND OFF-SHORE RIG PLATFORMS, WHO MUST PERFORM PLATE SIZING, PIPE CUTOFF, BULKHEAD FABRICATION, AND PROPELLER TRIMMING
- ALUMINUM, STEEL, AND COPPER MILLS THAT SIZE AND STRIP ROLLED PLATE AND SHEET, OR THAT SIZE SCRAP, OR THAT PERFORM PLANT EQUIPMENT MAINTENANCE
- SCRAP YARDS THAT PERFORM HIGH-SPEED SIZING OF STAINLESS STEEL, COPPER, BRASS, ALUMINUM, AND OTHER METALS

Plasma Arc Cutting

Chuck Landry, Thermal Dynamics Corporation

Safety

Operators of PAC equipment should be aware of seven areas of potential hazards that are associated with the process. The proper operation of the equipment and adherence to certain precautions will minimize the risks involved.

Gases and Fumes. Potentially hazardous gases associated with the PAC process include ozone, nitrogen dioxide, and phosgene (caused by the breakdown of chlorinated solvents in ultraviolet light). Metal fumes vary with the composition of the metal being cut. Fumes of beryllium, cadmium, cobalt, copper, lead, mercury, silver, vanadium, and several other elements may be hazardous to individuals in the PAC area. Various limits imposed on the levels of gases and fumes in the work area are listed in a handbook published by the Occupational Safety and Health Administration (OSHA) and in the American Welding Society document A6.3-69, "Recommended Safe Practices for Plasma Arc Cutting."

Gases and fumes created by the PAC process can be effectively controlled by properly installed ventilation systems in the work area. Water tables that are used either alone, for surface or underwater cutting, or in conjunction with water-shrouded systems can effectively reduce gases and fumes. Respirators can be used when ventilation systems and water tables are impractical. Chlorinated solvents and other sources of phosgene gas should be removed from the PAC work area.

Noise. The high level of noise associated with some PAC operations (usually, high power) can lead to hearing damage. Maximum exposure times to various noise levels are printed in the OSHA handbook. The operator and any other personnel within the vicinity of the PAC operation should wear proper ear protection. Underwater cutting and the use of water-shrouded systems are effective means for reducing noise levels.

Radiant Energy. The ultraviolet infrared and visible light from the plasma arc can lead to injury of the eyes and skin. Operators and anyone else in the vicinity of the PAC operation should be protected by proper clothing that covers all skin and by welding helmets or shields that contain the proper shaded lens. For high-powered operations, the use of a number 10 (or greater) shaded lens may be required by the operator. If the intensity of the PAC process causes eye discomfort to the operator, a lens with a darker shade should be utilized.

Electric Shock. Both the input power to and the output from a PAC system is sufficient to be hazardous if an individual receives a shock from the system. Output voltages in the PAC process usually range from 100 to 200 V dc; typical welding processes run at 20 to 40 V dc. All equipment should be installed and maintained according to the USA Standard National Electrical Code and any local standards. All components of a PAC system should be properly earth grounded. System power should be shut off at a wall disconnect before the torch or any other system component is serviced. Any cracked insulators, power leads, or exposed live parts should be corrected immediately. Operators of PAC systems should be insulated from the workpiece and ground, and those who work in wet or damp areas should take precautions to ensure adequate protection from electric shock.

Fire. The sparks and the hot molten metal that is blown out of the cut can cause fires. Sparks that result from piercing can travel a considerable distance. All combustibles should be removed from the work area, and a fire extinguisher should be located in the vicinity. Because the pilot arc from a PAC torch is very hot, extra care should be taken to prevent it from contacting clothing and other flammable materials.

Compressed Gas Cylinders. High-pressure gas cylinders, which are often used to supply PAC systems, can be hazardous if they are not treated properly. Proper handling requires that valve covers always be in place when transporting cylinders, and that cylinders be secured to stationary objects by chains or lashing straps to prevent them from falling over.

Explosions. When hydrogen is used as a plasma or secondary gas, in conjunction with argon, extra care should be taken, because it is highly explosive. In some instances, explosions of hydrogen gas can occur even if hydrogen is not supplied as a portion of the plasma or secondary gas. This is because hydrogen can form when some metals, particularly aluminum, are being cut over a water table, by a reaction between the metal and the water, and by the dissociation of water that is due to the plasma arc. During the cutting operation, the plasma arc can ignite this hydrogen, which will explode if it is confined. Adequate ventilation of these hydrogen pockets is necessary for the safe operation of PAC systems. The water in the water table should not be any deeper than 50.8 mm (2 in.) below the bottom of the metal workpiece support bars. A blower system should be used to flush out all hydrogen from beneath the workpiece.

Air-Carbon Arc Cutting

Robert L. Strohl, Tweco Arcair

Introduction

AIR-CARBON ARC CUTTING (CAC-A) removes metal physically, rather than chemically, as in the oxyfuel cutting process. Gouging, or cutting, occurs when the intense heat of the arc between the carbon electrode and the workpiece melts part of the workpiece. Air simultaneously passes through the arc quickly enough to blow the molten material away. Because this process does not require oxidation to maintain the cut, it can gouge metals that the oxyfuel cutting process cannot. The metal re- moval rate depends on the melting rate and efficiency of the air jet in removing the molten metal. The air must lift the molten metal clear of the arc before the metal solidifies.

Applications. Industry has enthusiastically adopted CAC-A gouging for numerous applications, such as metal fabrication and casting finishing, chemical and petroleum technology, construction, mining, and general repair and maintenance.

The process is used throughout the world to create consistent grooves requiring little or no additional cleaning, on square-butt plate seams to prepare them for welding. The process can then be used to backgouge the seam to sound metal to ensure 100% penetration of the welded joint. If a problem arises during the welding process and an area of the weld does not meet specifications, then the process can be used to remove the defective area without damaging or detrimentally affecting the base metal.

The CAC-A process is also used in foundries to remove fins and risers from castings. It can then be used to smooth areas that contact the surface to prepare the casting for shipment. The process is flexible, efficient, and cost effective to use on numerous metals, such as carbon steel, stainless steel, and other ferrous alloys; gray, malleable, and ductile iron; aluminum; nickel; copper alloys; and other nonferrous metals.

Air-Carbon Arc Cutting

Robert L. Strohl, Tweco Arcair

Principles of Operation

Like electric arc welding, the CAC-A process requires an arc of intense heat to develop a molten pool on the workpiece. Compressed air then blows away this molten metal. The process requires a welding power source, an air compressor, a carbon electrode, and a gouging torch. Figure 1 shows how the components are typically arranged.

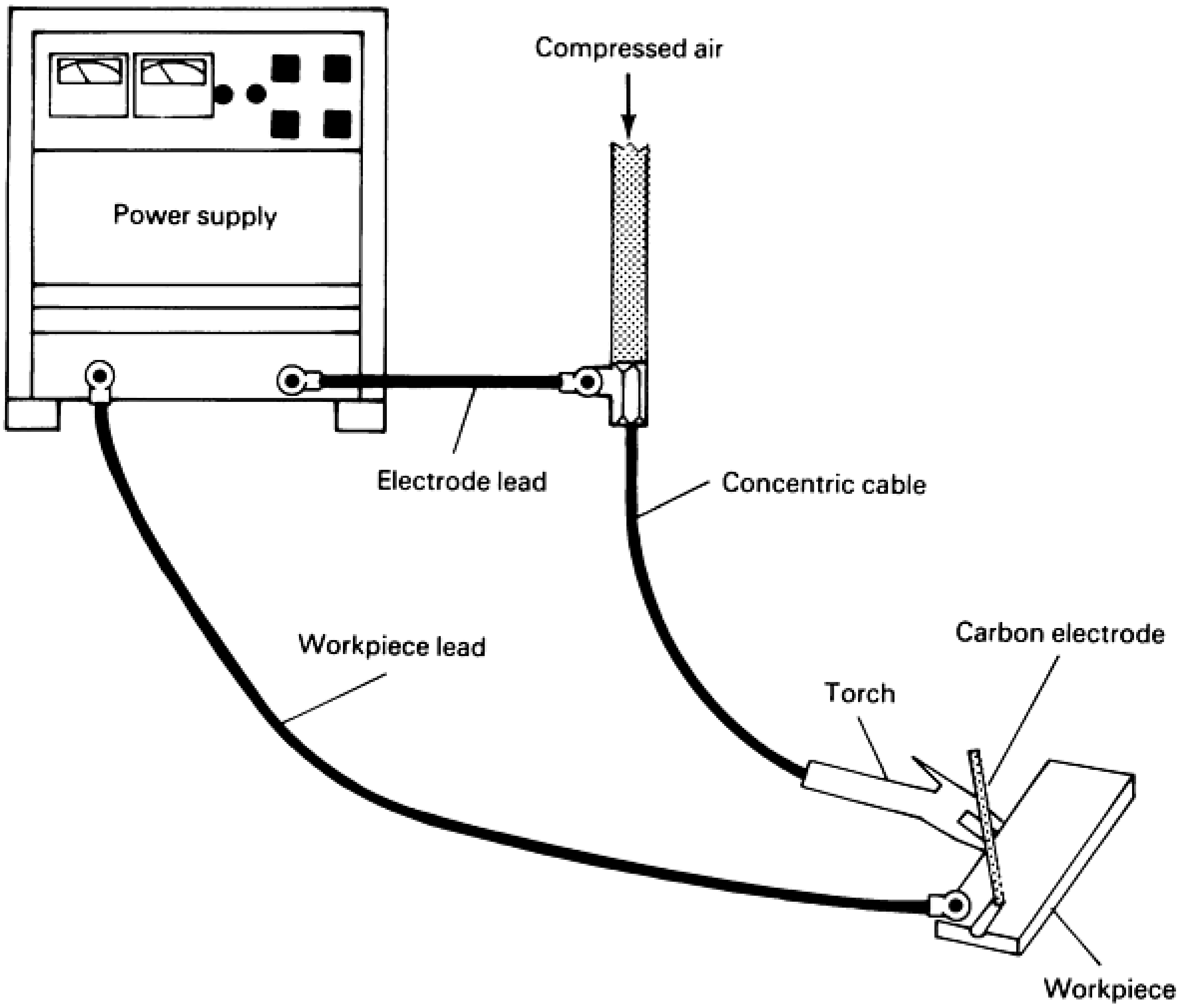
Compressed air
Power supply
Electrode lead
Concentric cable
Workpiece lead
Carbon electrode
Torch
Workpiece

FIG. 1 ARRANGEMENT OF COMPONENTS USED IN THE CAC-A PROCESS

Cuts or gouges should be made only in the direction of the air flow. The electrode angle will vary, depending on the application. The cutter should maintain the correct arc length to enable air to remove molten metal (Fig. 2).

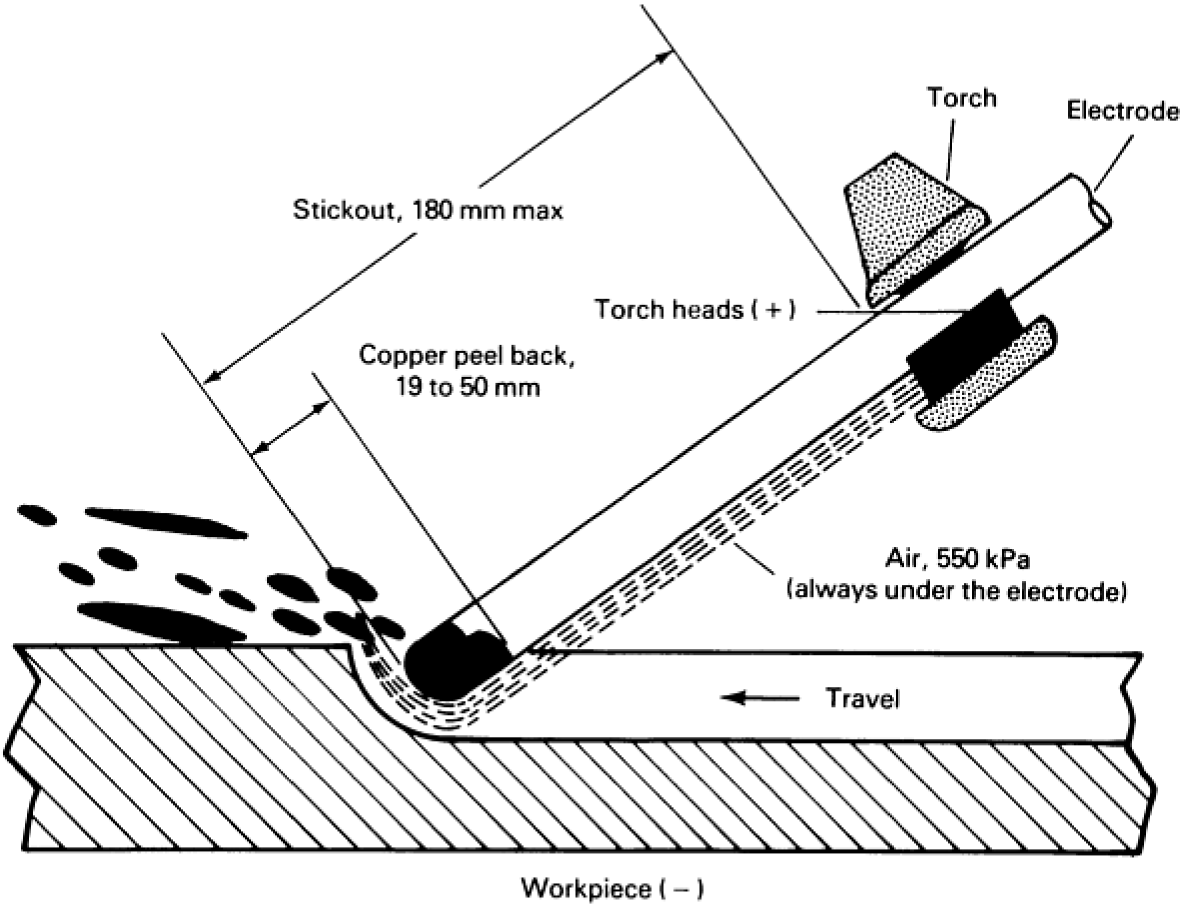

FIG. 2 PRINCIPLES OF CAC-A CUTTING

Power Sources. Single-phase machines with low open-circuit voltage may not work for CAC-A gouging. However, any three-phase welding power source of sufficient capacity can be used. The open-circuit voltage should be higher than the required arc voltage to allow for voltage drop in the circuit. Because the arc voltage typically ranges from a low of 35 to a high of 56 V, the open-circuit voltage should be at least 60 V. The actual arc voltage used in the process is governed by arc length and the type of gouging.

Direct current, electrode positive (reverse polarity, DCEP) is used with the process, except in special applications discussed later. The electrode should extend, at most, 178 mm (7 in.) from the gouging torch, with the air jet between the electrode and workpiece. A minimum extension of 50 mm (2 in.) should be used. Damage of the torch parts will occur if the electrode sticks out less than 50 mm (2 in.).

Compressed Air. Ordinary compressed air should be used for CAC-A gouging. Normal pressures range from 550 to 690 kPa (80 to 100 psi) at the torch. Higher pressures can be used, but they do not remove metal more efficiently. A pressure of 415 kPa (60 psi) should be used with the light-duty manual torch. This low level of pressure should not be used with general-duty torches.

Regardless of the pressure used with manual torches, the air hose supplying the cable assembly that is connected to the torch body should have an inside diameter (ID) of at least 6.4 mm ($\frac{3}{8}$ in.). Mechanized torches with automatic arc length control should have an air-supply hose with minimum ID of 12 mm ($\frac{1}{2}$ in.). Attention should be given to the field connections to ensure that this minimum is maintained without severely restricting the air-flow volume. Table 1 gives the consumption rates of compressed air for manual and mechanized torches, as well as compressor power rating for intermittent and continuous use. The compressor receiver tank must be large enough to accommodate the compressor rating.

TABLE 1 RECOMMENDED MINIMUM AIR SUPPLY REQUIREMENTS FOR AIR-CARBON ARC CUTTING

TYPE OF TORCH	AIR PRESSURE(A)		AIR CONSUMPTION		RECOMMENDED COMPRESSOR RATING				ASME RECEIVER SIZE	
					INTERMITTENT USE		CONTINUOUS USE			
	kpa	psi	l/min	ft^3/min	kw	hp	kw	hp	l	gal
LIGHT DUTY(B)	280	40	227	8	0.4	0.5	1.1	1.5	227	60
GENERAL DUTY(B)	550	80	708	25	3.7	5	5.6	7.5	303	80
MULTIPURPOSE(C)	550	80	934	33	5.6	7.5	7.5	10	303	80
AUTOMATIC(D)	414	60	1303	46	. . .	. . .	11.2	15	303	80

(A) PRESSURE WHILE TORCH IS IN OPERATION.
(B) ACCOMMODATES FLAT ELECTRODES.
(C) GENERALLY CONSIDERED TO BE A FOUNDRY TORCH.
(D) REQUIRES SOME KIND OF MECHANICAL MANIPULATION

Electrodes. The basic types of electrodes are described below.

Direct-current copper-coated electrodes are the most widely used type, because of their comparatively long electrode life, stable arc characteristics, and groove uniformity. These electrodes are made by mixing carbon and graphite with a binder. When this mixture is baked, it produces dense, homogeneous graphite electrodes of low electrical resistance. These electrodes are then coated with a controlled thickness of copper. The available diameters of these electrodes are 3.2 mm ($\frac{1}{8}$ in.), 4.0 mm ($\frac{5}{32}$ in.), 4.8 mm ($\frac{3}{16}$ in.), 6.4 mm ($\frac{1}{4}$ in.), 7.9 mm ($\frac{5}{16}$ in.), 9.5 mm ($\frac{3}{8}$ in.), 12.7 mm ($\frac{1}{2}$ in.), 15.9 mm ($\frac{5}{8}$ in.), and 19.1 mm ($\frac{3}{4}$ in.).

Jointed electrodes work without stub loss. They are furnished with a female socket and a matching male tang, and are available in diameters of 7.9 mm ($\frac{5}{16}$ in.), 9.5 mm ($\frac{3}{8}$ in.), 12.7 mm ($\frac{1}{2}$ in.), 15.9 mm ($\frac{5}{8}$ in.), 19.1 mm ($\frac{3}{4}$ in.), and 25.4 mm (1 in.). Jointed carbons are always used with the male tang pointing up (that is, away from the gouge). This ensures proper joint performance.

Flat, or rectangular, coated electrodes are available in dimensions of 4.0 × 9.5 mm ($\frac{5}{32} \times \frac{3}{8}$ in.) and 4.8 × 15.9 mm ($\frac{3}{16} \times \frac{5}{8}$ in.). These electrodes are used to make rectangular grooves and to remove weld reinforcements.

Direct-current electrodes are generally restricted to diameters of less than 9.5 mm ($\frac{3}{8}$ in.), and are therefore limited in use. During cutting, these electrodes erode more than the coated electrodes. They are manufactured in the same way as the coated electrodes, but without the copper coating. Plain electrodes are available in diameters of 4.0 mm ($\frac{5}{32}$ in.), 4.8 mm ($\frac{3}{16}$ in.), 6.4 mm ($\frac{1}{4}$ in.), 7.9 mm ($\frac{5}{16}$ in.), and 9.5 mm ($\frac{3}{8}$ in.).

Alternating-current (ac) coated electrodes are made from a mixture of carbon, graphite, and a special binder. Rare-earth materials are added to ensure arc stabilization when using alternating current. These electrodes, coated with copper, are available in diameters of 4.8 mm ($\frac{3}{16}$ in.), 6.4 mm ($\frac{1}{4}$ in.), and 9.5 mm ($\frac{3}{8}$ in.).

A manual gouging torch is shown in Fig. 3. The electrode is held in a swivel head that contains one or more air holes, so that the air jet stays aligned with the electrode, regardless of the angle of the electrode to the gouging torch. Torches with two heads (that is, the air jet is on two sides of the electrode) or with a fixed angle between the electrode and the holder are better for certain applications, such as removing pads and risers from large castings (pad washing).

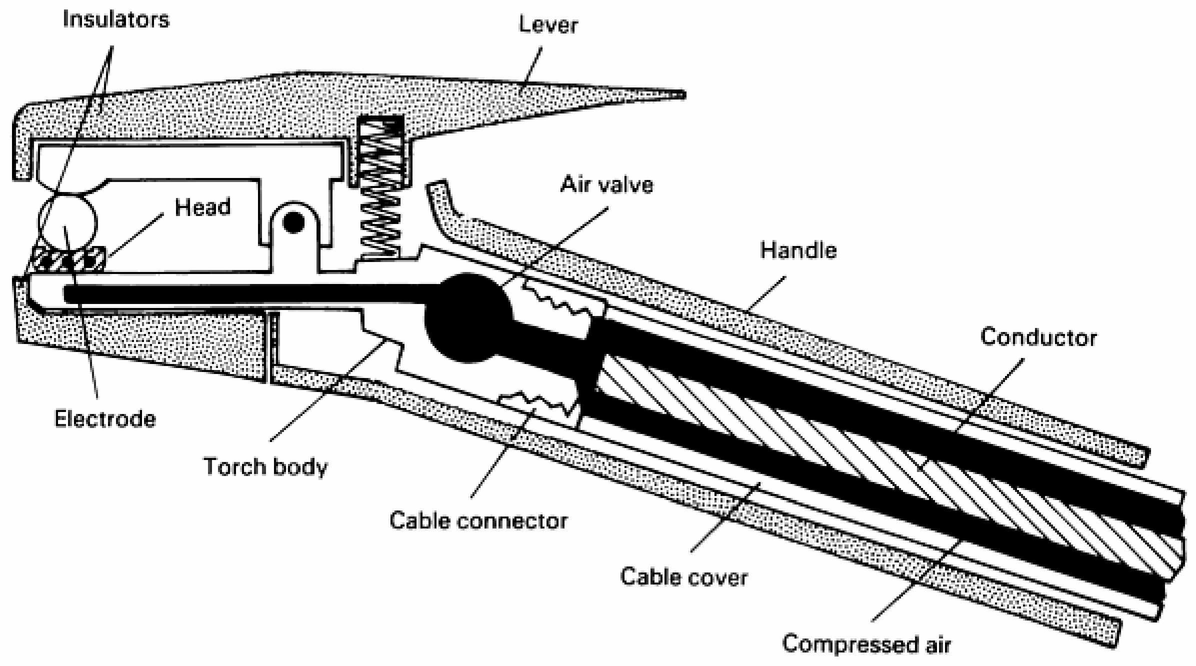

FIG. 3 CUTAWAY VIEW OF A MANUAL GOUGING TORCH WITH AIR FLOW ON BOTTOM OF ELECTRODE

Manual torches are usually air cooled. In high-current applications, water-cooled cable assemblies can be used with heavy-duty torches.

Control of Automatic Gouging Torches. There are two methods of controlling automatic CAC-A gouging torches. Either method can make grooves of consistent depth to a tolerance of ±0.64 mm (±0.025 in.). These automatic units are used to achieve high-quality gouges and to increase production (Fig. 4).

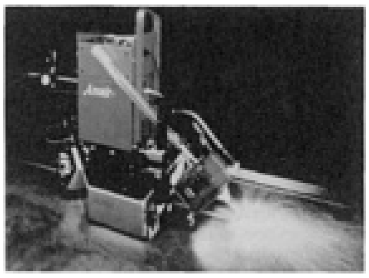

FIG. 4 AUTOMATIC GOUGING UNIT

The amperage-controlled method maintains the arc current by amperage signals through solid-state controls. This method controls the electrode feed speed, which maintains the preset amperage. It is run with constant-voltage power sources only.

A voltage-controlled method maintains arc length by voltage signals through solid-state electronic controls. This method controls the electrode feed speed, which maintains the preset voltage. It can run with constant-current power sources.

Air-Carbon Arc Cutting

Robert L. Strohl, Tweco Arcair

Operating Techniques

Gouging with Manual Torches. As shown in Fig. 2, the electrode should be gripped so that a maximum of 178 mm (7 in.) extends from the torch. For aluminum, this extension should be 76.5 mm (3 in.). Table 2 shows suggested currents for the different electrode types and sizes.

TABLE 2 SUGGESTED CURRENT RANGES FOR VARIOUS ELECTRODE SIZES

ELECTRODE SIZE		MINIMUM AMPERAGE	MAXIMUM AMPERAGE
mm	in.		
3	$\frac{1}{8}$	60	90
4	$\frac{5}{32}$	90	150
5	$\frac{3}{16}$	200	250
6	$\frac{1}{4}$	300	400
8	$\frac{5}{16}$	350	450
10	$\frac{3}{8}$	450	600
13	$\frac{1}{2}$	800	1000
16	$\frac{5}{8}$	1000	1250
19	$\frac{3}{4}$	1250	1600
25	1	1600	2200

The air jet should be turned on before striking the arc, and the torch should be held as shown in Fig. 5. The electrode should slope back from the direction of travel, with the air jet behind the electrode. During gouging, the air jet sweeps beneath the electrode end and removes all molten metal. The arc can be struck by lightly touching the electrode to the workpiece. The electrode should not be drawn back once the arc is struck.

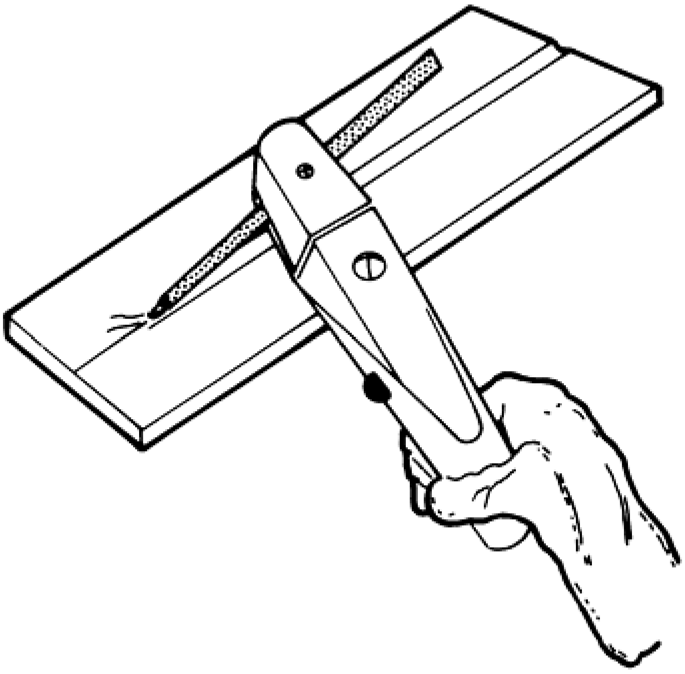

FIG. 5 FLAT GOUGING POSITION

Gouging is different from arc welding in that metal is removed, rather than deposited. A short arc should be maintained in the direction of cut by working fast enough to keep up with metal removal. Steadiness of movement controls the smoothness of the resulting cut.

Vertical gouging should be conducted downhill, permitting gravity to help remove the molten metal. Although vertical gouging can be accomplished uphill, it is difficult. Horizontal gouging can be done either to the right or to the left, but always with forehand gouging.

The air jet should always be positioned under the electrode. When gouging in the overhead position, the electrode and torch should be held to prevent molten metal from dripping on the glove of the operator. The depth of the groove is controlled by the travel speed. Grooves up to 25 mm (1 in.) deep can be made. However, the deeper the groove, the more experienced the operator needs to be. Slow travel speeds will produce a deep groove. Fast speeds will produce a shallow groove. The width of the groove is determined by the size of the electrode used. The groove is usually about 3.2 mm ($\frac{1}{8}$ in.) wider than the diameter of the electrode. A wider groove can be made with a small electrode that is oscillated in either a circular or weave motion. When gouging, a push angle that is 35° from the surface of the workpiece should be used for most applications. A steady rest will ensure a smoothly gouged surface, especially in the over-head position. Proper travel speed depends on electrode size, the base metal being used, amperage, and air pressure. The proper speed, which produces a smooth hissing sound, will result in a good gouge.

The severing technique is like gouging, except that the operator holds the electrode at a steeper angle that is between 70 and 80° to the workpiece surface. When cutting thick, nonferrousmetals, the electrode should be held perpendicular to the workpiece, with the air jet favoring the desired side. With the electrode in this position, the operator can sever the metal by moving the arc up and down using a sawing motion.

Washing. When using the CAC-A process to remove metal from large areas (for example, surfacing metal or riser pads on castings), the operator should position the electrode as shown in Fig. 6. The electrode should weave from side to side while it is being pushed forward at the depth desired. In the pad-washing operation, an angle that is 15 to 70° to the workpiece should be used. The 15° angle is used for light finishing passes, whereas steeper angles allow deeper rough gouging with greater ease.

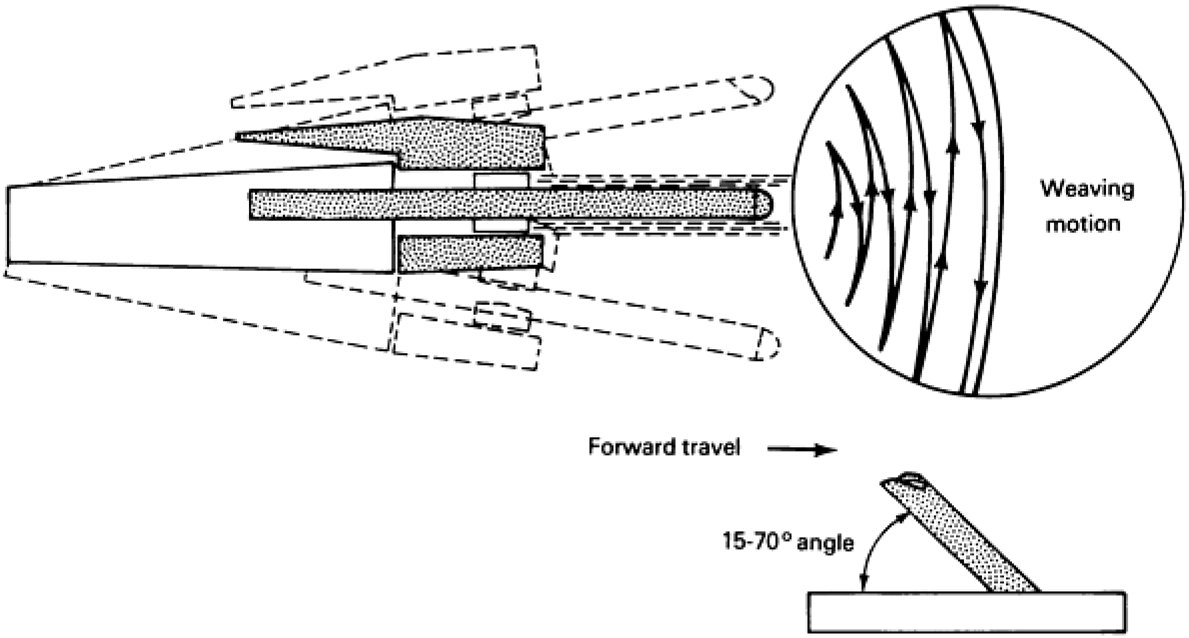

FIG. 6 PAD-WASHING TECHNIQUE. THE TORCH REMAINS PARALLEL TO THE WORKPIECE, AND A WEAVING MOTION IS USED FORWARD ACROSS THE WIDTH OF THE AREA BEING CLEANED. THE ELECTRODE-TO-WORKPIECE ANGLE SHOULD BE FROM 15 TO 70°. THE MORE SHALLOW THE ANGLE, THE SMOOTHER THE FINISH; THE 70° ANGLE IS USED MAINLY ON CAST IRON.

Gouging torches with fixed-angle heads that hold the electrode at the correct angle are well-suited for this application. When other types of torches are used, the air jet should be kept behind the electrode. Operator steadiness determines the smoothness of the surface produced.

Beveling. In one beveling method that is used for thick plates, the electrode should be held with a travel angle of 90° and a work angle that is equal to the bevel angle. The air jet should be positioned between the electrode and the workpiece. A second method, usually used for thin plates, requires that the torch be held parallel to the edge being beveled. Its position should equal the bevel angle. The air jet should be positioned between the electrode and the workpiece surface.

Air-Carbon Arc Cutting

Robert L. Strohl, Tweco Arcair

Equipment Selection

Gouging torches are selected for the specific application. They range from light-duty sizes appropriate for farm and body shop applications to extra-heavy-duty torches that are best suited to foundry applications.

General-purpose torches are available in four duty levels. Light-duty torches accept round electrodes that range from 3.2 to 6.5 mm ($\frac{1}{8}$ to $\frac{1}{4}$ in.) and 9.5 mm ($\frac{3}{8}$ in.) flat electrodes. Their maximum current level is 450 A. Medium-duty torches accept round electrodes that range from 3.97 to 9.5 mm ($\frac{5}{32}$ to $\frac{3}{8}$ in.) and 9.5 mm ($\frac{3}{8}$ in.) flat electrodes. Their maximum current level is 600 A. Heavy-duty torches accept round electrodes that range from 3.97 to 12.7 mm ($\frac{5}{32}$ to $\frac{1}{2}$ in.) and flat electrodes of 9.5 or 15.9 mm ($\frac{3}{8}$ or $\frac{5}{8}$ in.). Their maximum current level is 1000 A. Extra-heavy-duty torches accept round electrodes that range from 3.97 to 15.9 mm ($\frac{5}{32}$ to $\frac{5}{8}$ in.) and flat electrodes of 9.5 or 15.9 mm ($\frac{3}{8}$ or $\frac{5}{8}$ in.). Their maximum current level is 1250 A.

Foundry heavy-duty torches are appropriate for general foundry work and heavy-duty fabrication. They are limited to 1600 A when used with air-cooled cables and 2000 A when used with water-cooled cables.

Automated gouging torches are used for edge preparations and backgouging, as well as high-quality and high-productivity applications. Jointed carbon electrodes ranging from 7.9 to 19.1 mm ($\frac{5}{16}$ to $\frac{3}{4}$ in.) are accepted.

Power Sources. Any three-phase welding power source with enough capacity can be used for the CAC-A gouging process if the open-circuit voltage is high enough to allow for a voltage drop in the circuit. Some constant-voltage power sources with drooping characteristics require a very high open-circuit voltage to run CAC-A gouging equipment. Mechanized gouging and other applications that require maximum arc time should utilize a power source with a 100% duty cycle for the required amperage.

A dc constant-current power source motor generator, rectifier, or resistor grid unit is preferred for use with all electrode sizes. A dc constant-voltage power source motor generator or rectifier is usable only for 7.9 mm ($\frac{5}{16}$ in.) diameter and larger electrodes. It may cause carbon deposits when used with small electrodes, and it is not suitable for automatic torches with voltage control only. An ac constant-current power source (transformer) is recommended for ac electrodes only. With ac/dc transformer rectifiers, the direct current should be supplied by a three-phase source, because single-phase sources give unsatisfactory arc characteristics. Alternating current output from ac/dc units is satisfactory, if ac electrodes are used.

Automatic systems are often used in modern fabrication facilities. These systems offer a high-quality, high-productivity alternative to manual gouging. Two types of systems can be considered. Both operate on a signal from the arc to control the gouging.

Dual-Signal System. Either constant-current or constant-voltage power supplies can be used with this type of automatic system. When the system uses constant current, the arc length is maintained through a voltage-signal system. A predetermined voltage setting is set on the system controller, which then advances or retracts the electrode through a stepping motor to maintain the arc length. With a constant-voltage power source, amperage sensing controls the feeding or retracting of the electrode to maintain the right arc current.

The single-signal system, like the dual-signal system, maintains arc length through voltage signal. However, it will not operate with an amperage signal. This type of system only operates with a constant-current power source.

Advantages. Automatic CAC-A systems ensure better productivity and quality. The systems can perform out-of-position gouging. They perform well when making long gouges in flat workpieces using a moving gouging apparatus and when making circular gouges in pipes and tanks using stationary gouging apparatus. They produce an even U-shaped groove and can control the depth of the groove to within ±0.64 mm (±0.025 in.). Table 3 shows typical operating information for U-shaped grooves.

TABLE 3 AUTOMATIC AIR-CARBON ARC U-SHAPED GROOVE OPERATING DATA$^{(A)}$

ELECTRODE DIAMETER		DESIRE DEPTH		DC CURRENT, A	TRAVEL SPEED	
mm	in.	mm	in.		mm/min	in./min
7.9	$\frac{5}{16}$	3.2	$\frac{1}{8}$	400	1651	65
7.9	$\frac{5}{16}$	4.8	$\frac{3}{16}$	400	1143	45

7.9	$\frac{5}{16}$	6.4	$\frac{1}{4}$	450	914	36
7.9	$\frac{5}{16}$	7.9	$\frac{5}{16}$	450	838	33
7.9	$\frac{5}{16}$	11.1	$\frac{7}{16}$	450	572	22.5
9.5	$\frac{3}{8}$	3.2	$\frac{1}{8}$	500	1778	70
9.5	$\frac{3}{8}$	4.8	$\frac{3}{16}$	500	1118	44
9.5	$\frac{3}{8}$	6.4	$\frac{1}{4}$	500	889	35
9.5	$\frac{3}{8}$	9.5	$\frac{3}{8}$	500	508	20
9.5	$\frac{3}{8}$	14.3	$\frac{9}{16}$	500	445	17.5
12.7	$\frac{1}{2}$	3.2	$\frac{1}{8}$	850	2438	96
12.7	$\frac{1}{2}$	6.4	$\frac{1}{4}$	850	1448	57
12.7	$\frac{1}{2}$	9.5	$\frac{3}{8}$	850	889	35
12.7	$\frac{1}{2}$	12.7	$\frac{1}{2}$	850	610	24
12.7	$\frac{1}{2}$	19.1	$\frac{3}{4}$	850	445	17.5
15.9	$\frac{5}{8}$	6.4	$\frac{1}{4}$	1250	1829	72
15.9	$\frac{5}{8}$	9.5	$\frac{3}{8}$	1250	1219	48
15.9	$\frac{5}{8}$	12.7	$\frac{1}{2}$	1250	940	37
15.9	$\frac{5}{8}$	15.9	$\frac{5}{8}$	1250	762	30
15.9	$\frac{5}{8}$	23.8	$\frac{15}{16}$	1250	495	19.5
19.1	$\frac{3}{4}$	6.4	$\frac{1}{4}$	1400	1829	72
19.1	$\frac{3}{4}$	9.5	$\frac{3}{8}$	1400	1068	42
19.1	$\frac{3}{4}$	12.7	$\frac{1}{2}$	1400	865	34
19.1	$\frac{3}{4}$	15.9	$\frac{5}{8}$	1400	687	27
19.1	$\frac{3}{4}$	19.1	$\frac{3}{4}$	1400	560	22
19.1	$\frac{3}{4}$	28.6	$1\frac{1}{8}$	1400	330	13

Note: If a groove depth greater than 1.5 times the diameter of the electrode being used is desired, then the groove should be made in two or more passes.

(A) BASED ON LABORATORY CONDITIONS; INFORMATION SHOULD BE USED AS A GUIDE

AND ADJUSTED FOR FIELD VARIANCE AS REQUIRED.

Vacuum gouging represents a new variation of gouging that was developed in the late 1980s. This process replaces the air jet that is used to evacuate the molten slag from the groove area with a high-volume vacuum that not only captures the slag and fume associated with the gouging process, but also reduces the noise level created when performing the gouging process.

The vacuum gouging process utilizes a specially designed nozzle that attaches to an automatic gouging head. By water cooling the nozzle and the attached vacuum hose, the slag and fume from the gouging process are pulled by the vacuum into a capture dram. The slag is knocked into the bottom of the catch tank, which is partially filled with water, and the fume is caught in a filter before the air is exhausted from the vacuum. Currently, this new vacuum system is limited to automated gouging operations on flat plate or pressure vessel circumferential seams.

Air-Carbon Arc Cutting

Robert L. Strohl, Tweco Arcair

Important Process Variables

Like any thermal-cutting process, the CAC-A process is sensitive to variables in operation. These variables can cause changes in the finished gouge that range from being undetectable to being unacceptable.

Electrode diameter and type are variables that determine the groove size, as well as productivity, groove quality, and metal-removal rates. The width of the groove will be about 3.2 mm ($\frac{1}{8}$ in.) wider than the diameter of the electrode. The choice of a proper electrode can be based on the size of the desired groove. Available power will dictate the outer limit. For example, a 13 mm ($\frac{1}{2}$ in.) wide, 6.4 mm ($\frac{1}{4}$ in.) deep groove that is 254 mm (10 in.) long could be made manually in two passes using a 6.4 mm ($\frac{1}{4}$ in.) diameter electrode, or in one pass using a 9.5 mm ($\frac{3}{8}$ in.) diameter electrode. In the first case, the best gouging rate would be 254 mm/min (10 in./min), divided by 2, or 125 mm/min(5 in./min). The travel speed of the second case (one pass) is 430 mm/min (17 in./min).

The 9.5 mm ($\frac{3}{8}$ in.) diameter electrode increases the gouging rate by 200%, which could offset the additional cost of the electrode. Automatic systems increase the productivity rate even further, through finite control of the arc voltage.

The gouging amperage is the melting force of the process. It is affected by electrode size. If the amperage were set too low for the electrode size, then the melting rate of the base metal would be inadequate and free-carbon deposits would occur. Although the base metal would melt, a setting that is too high would rapidly deteriorate the electrode while reducing the metal removed per electrode. Too high a setting can also substantially reduce torch life.

Voltage is the pressure, or arc force, that enables the current to flow across the arc gap. The CAC-A process often requires a higher voltage than do most welding processes. To ensure proper operation, a power source with a high enough open-circuit voltage to maintain a 28 V operating minimum should be used. Inadequate voltage can create a sputtering arc or it can prevent arc establishment, resulting in uneven grooves and, probably, free-carbon deposits that would require excessive grinding to remove.

Adequate air pressure and flow rate are required for the adequate removal of molten metal by the air jet. Air volume is as important as air pressure. Pressure is the speed of the air that moves the molten metal from the groove. If there is an inadequate volume of air to lift the molten material out of the groove, then it cannot be removed by the pressure or velocity. The result is excessive slag adhesion and unnecessary grinding to clean up the groove.

Travel speed affects the depth of the gouge and the quality of the groove. The faster the travel speed of an electrode, the shallower the gouge. A smaller electrode can be used if the travel speed is too fast for the comfort of the operator, or automatic gouging can be used. A groove that is too deep for the diameter of the electrode results in a poor-quality groove that requires much grinding.

The electrode push angle can vary some-what. When gouging manually, a steeper angle tends to give a more V-shaped groove. When gouging automatically, a steeper angle gives a slightly deeper groove at the same travel speed.

Base Metals. The gouging procedures that are recommended for specific base metals are:

- FOR CARBON STEEL AND LOW-ALLOY STEEL, SUCH AS ASTM A 514 AND A 517, USE DC ELECTRODES WITH DCEP. ALTHOUGH AC ELECTRODES WITH AN AC TRANSFORMER CAN BE USED, THEIR EFFICIENCY IS HALF THAT OF DC ELECTRODES.
- FOR STAINLESS STEEL, THE SAME GUIDELINES APPLY AS FOR CARBON STEEL.
- FOR CAST IRON, INCLUDING MALLEABLE AND DUCTILE (NODULAR) FORMS, USE 13 MM ($\frac{1}{2}$ IN.) DIAMETER OR LARGER ELECTRODES AT THE HIGHEST RATED AMPERAGE. SPECIAL TECHNIQUES ARE NEEDED WHEN GOUGING THESE METALS. THE PUSH ANGLE SHOULD BE AT LEAST 70° OFF THE WORKPIECE, AND THE DEPTH OF THE CUT SHOULD NOT EXCEED 13 MM ($\frac{1}{2}$ IN.) PER PASS.
- FOR COPPER ALLOYS (COPPER CONTENT $\leq 60\%$), USE DC ELECTRODES WITH DIRECT CURRENT ELECTRODE NEGATIVE (DCEN) AT THE MAXIMUM AMPERAGE RATING OF THE ELECTRODE.
- FOR COPPER ALLOYS (COPPER CONTENT >60% OR LARGE WORKPIECE), USE DC ELECTRODES WITH DCEN AT THE MAXIMUM AMPERAGE RATING OF THE ELECTRODE, OR USE AC ELECTRODES WITH ALTERNATING CURRENT.
- FOR ALUMINUM BRONZE AND ALUMINUM-NICKEL-BRONZE (SPECIAL NAVAL PROPELLER ALLOYS), USE DC ELECTRODES WITH DCEN.
- FOR NICKEL ALLOYS (NICKEL CONTENT >80% OF MASS), USE AC ELECTRODES WITH ALTERNATING CURRENT.
- FOR NICKEL ALLOYS (NICKEL CONTENT <80% OF MASS), USE DC ELECTRODES WITH DCEP.
- FOR MAGNESIUM ALLOYS, USE DC ELECTRODES WITH DCEP AND WIRE-BRUSH THE GROOVE BEFORE WELDING.
- FOR ALUMINUM ALLOYS, USE DC ELECTRODES WITH DCEP AND WIRE-BRASH THE GROOVE WITH STAINLESS BRUSHES BEFORE WELDING. ELECTRODE EXTENSION (LENGTH OF ELECTRODE BETWEEN TORCH AND WORK) SHOULD NOT EXCEED 76.2 MM (3 IN.). DIRECT-CURRENT ELECTRODES WITH DCEN CAN ALSO BE USED.
- FOR TITANIUM, ZIRCONIUM, HAFNIUM, AND THEIR ALLOYS, THERE ARE NO CURRENT PROCEDURES. IF THESE METALS ARE PREPARED FOR WELDING USING THE CAC-A PROCESS, THEN THE GROOVE SHOULD BE CLEANED BEFORE WELDING. THESE METALS CAN BE CUT FOR REMELTING. IN CASES WHERE PREHEAT IS REQUIRED FOR WELDING, A SIMILAR PREHEAT SHOULD BE USED FOR GOUGING.

Effects of the Cutting Process on Base Metals. Metallurgically, the base metal is not affected by the carbon particles being carried toward it from the electrode, because it remains solid and the atoms are very tightly packed. However, because the material being removed is molten, the atoms are loosely spread. This material, usually referred to as slag, easily absorbs the carbon particles.

If the CAC-A process is properly used, no surface carburization will occur. Due to the rapid quenching effect of the air used to remove the metal, the heat-affected zone (HAZ) will undergo hardening. The heat-affected zone for CAC-A is typically 2.5 mm (0.10 in.) or less in thickness. This HAZ hardening must be considered when a finish machining operation follows the gouging process, especially when working on high-carbon steels. Some materials (for example, nickel) will undergo surface cracking when gouged. These conditions should be considered when choosing the parameters for any required machining operations.

When the CAC-A process is used under improper conditions, this molten carburized metal may be left on the workpiece surface. Its color is a dull gray-black, in contrast to the bright blue of the properly made groove. Inadequate air flow may leave small pools of carburized metal in the bottom of the groove. Irregular electrode travel, especially in a manual operation, may cause ripples in the groove wall that trap the carburized metal. Finally, an improper electrode push angle may cause small beads of carburized metal to remain on the edge of the groove.

Copper from copper-coated electrodes does not transfer to the cut surface in base metal, unless the process is improperly used.

Carburized metal can be removed from a cut surface by grinding, but the necessity to do so can be prevented by gouging properly in the right condition.

When compared with oxyfuel cutting, the CAC-A process requires less heat input. Therefore, a workpiece that is gouged or cut by the CAC-A process is less distorted. The machining of low-carbon and nonhardenable steel is not affected by the CAC-A process. However, when used on cast iron and high-carbon steels, this process may cause enough hardening to make the cut surface tough to machine. Still, because the hardened zone is shallow (approximately 0.15 mm, or 0.006 in.), a cutting tool can penetrate it to remove the hardened surface.

Air-Carbon Arc Cutting

Robert L. Strohl, Tweco Arcair

Safety and Health

Safe practices in welding and cutting processes are described in ANSI Z49.1, "Safety in Welding and Cutting," and ANSI Z49.2, "Fire Prevention in Use of Welding and Cutting Processes." The operators of the CAC-A process and their supervisors should adhere to the practices discussed in these documents.

Electrodes. An electrode that carries electric current will arc if it touches the workpiece or any grounded metal object. Proper handling of the torch and electrode will prevent accidental arcing. Electrodes should always be used within their proper amperage range in order to effectively remove metal and to avoid damage to the torch. Carbon electrodes should be kept dry. Damp electrodes should be baked for 10 h at 176 °C (300 °F). Wet electrodes may shatter.

Torches. Oxygen should never be used in an air carbon-arc torch. Neither the torch nor the electrode should ever be immersed in water. All electrical connections should be wrench tight before the torches are used.

Electrical Power. The CAC-A process uses electric energy from a welding power source. The same safety precautions should be used as when welding. An operator should never stand in water while using CAC-A equipment, and the torches and electrodes should never be cooled by water. To prevent accidental arcing, partially used electrodes should be removed from torches when work is interrupted.

Personal Protective Equipment and Clothing. A No. 12 shade-filter lens should be used for eye protection against arc radiation. When conducting heavy metal-removal operations with large electrodes, a No. 14 shade-filter lens or an equivalent combination of filter lenses provides the best protection. Flash goggles should be used, especially when two or more welders are working in the same area. Proper protective clothing must be worn to provide enough protection from the infrared and ultraviolet radiations of the arc. Leather aprons, sleeves, leggings, and so on should be used for out-of-position gouging or for heavy metal-removal operations with large electrodes. Ear protection is recommended when noise exceeds permissible levels.

Fire and Burn Hazards. The CAC-A process may cause fire. All flammable materials should be removed from the work area. Booths, metal screens, and other deflectors should be placed to catch the hot metal and particle spray ejected by the compressed air stream. Protective clothing should be used to prevent burns caused by contact with hot metal or electrodes. Out-of-position or heavy metal-removal operations require that many precautions be taken.

Ventilation Hazards. The fumes from the CAC-A process can be harmful. Ventilation and/or an exhaust around the arc is necessary. When using gouging torches, the operator should keep his head away from the fumes. Protective-breathing or air-circulation gear may be needed.

Additional information is available in the article "Safe Practices" in this Volume.

Mechanical Cutting for Weld Preparation

ASM Committee on Mechanical Cutting for Welding Preparation*

Introduction

OXYFUEL AND PLASMA CUTTING are the processes most commonly associated with preparing metals for welding; however, mechanical cutting methods are also widely used by the metal fabrication industry. The purpose of this article is to introduce the welding fabricator to some of the mechanical equipment used to shape or prepare metals for welding. The most prevalent equipment used for mechanical cutting includes shears, iron workers, nibblers, and band saws. Each has specific applications for which it is efficient and cost effective.

Acknowledgements

Some of the information in this article was drawn from the articles "Shearing of Plate and Flat Sheet" and "Shearing of Bars and Bar Sections" in *Forming and Forging,* Volume 14 of the *ASM Handbook,* and "Sawing" in *Machining,* Volume 16 of the *ASM Handbook.* Special thanks are due to the contributors to those articles.

Note

* LANCE R. SOISSON, *CHAIRMAN,* WELDING CONSULTANTS, INC.; CHRIS CABLE, FEIN POWER TOOLS, INC.; RICHARD S. CREMISIO, RESCORP INTERNATIONAL, INC.; CHUCK DVORAK AND JIM DVORAK, UNI-HYDRO, INC.; GERHARD LAEPPLE, C & E FEIN GMBH; AND JOHN R. WHALEN, CONTOUR SAWS, INC.

Shears

Shears are normally used to prepare large, relatively thin materials ($\leq$25 mm, or 1 in., thick) by cutting them to final dimensions. When maintained and adjusted properly, shears provide a clean, straight edge preparation. For thin sections, the square butt edge preparation provided by shearing may be all that is required for the welding operation. As a role, material with a thickness of 10 mm ($\frac{3}{8}$ in.) or more will require further preparation; the straightness of the sheared edge is an excellent reference point for setting up other equipment. The two principal shearing variations are straight-knife shearing and rotary shearing. Because straight-knife shearing is more widely used for welding preparation, it will be the focus of this section.

In straight-knife shearing, the work metal is placed between a stationary lower knife and a movable upper knife. As the upper knife is forced down, the work metal is penetrated to a specific portion of its thickness. The unpenetrated portion then fractures, and the work metal separates (Fig. 1). The amount of penetration depends largely on the ductility and thickness of the work metal. The knife will penetrate 30 to 60% of the work metal thickness for low-carbon steel, depending on thickness (see the section "Capacity" in this article). The penetration will be greater for a more ductile metal such as copper. Conversely, the penetration will be less for metals that are harder than low-carbon steel.

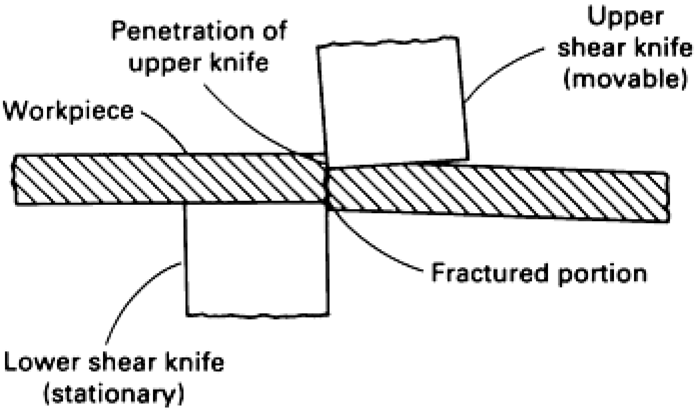
Penetration of upper knife
Upper shear knife (movable)
Workpiece
Fractured portion
Lower shear knife (stationary)

FIG. 1 SCHEMATIC SHOWING THE MECHANISM OF STRAIGHT-KNIFE SHEARING

A sheared edge is characterized by the smoothness of the penetrated portion and the relative roughness of the fractured portion. Sheared edges cannot compete with machined edges, but when knives are kept sharp and in proper adjustment, it is possible to obtain sheared edges that are acceptable for a wide range of applications. The quality of sheared edges generally improves as work metal thickness decreases.

Applicability. Straight-knife shearing is the most economical method of cutting straightsided blanks from stock no more than 50 mm (2 in.) thick. The process is also widely used for cutting sheet into blanks that will subsequently be formed or drawn. Because shear gaging can be set within ±0.13 mm (±0.005 in.), the shearing process is generally limited to ±0.4 mm ($\pm\frac{1}{64}$ in.) tolerances in 16 gage material. The tolerance range increases with thickness.

Straight-knife shearing is seldom used for shearing metal harder than about 30 HRC. When extremely soft, ductile metal (especially thin sheet) is sheared, the edges of the metal roll and large burrs result. As the hardness of the work metal increases, knife life decreases for shearing a given thickness of metal.

In general, it is practical to shear flat stock up to 38 mm ($1\frac{1}{2}$ in.) thick in a squaring shear. Squaring shears up to 9 m (30 ft) long are available (even longer shears have been built), and some types are equipped with a gap that permits shearing of work metal longer than the shear knife.

Machines for Straight-Knife Shearing

Punch presses and press brakes are sometimes used for shearing a few pieces or are used temporarily when more efficient equipment is not available. Production shearing, however, is usually done in machines that are designed for this operation.

Squaring shears are usually used for trimming and cutting sheet or plate to specific size (Fig. 2). Squaring shears (also called resquaring or guillotine shears) are available in a wide range of sizes and designs. Some types permit slitting by moving the work metal a predetermined amount in a direction parallel with the cutting edge of the knife after each stroke of the shear.

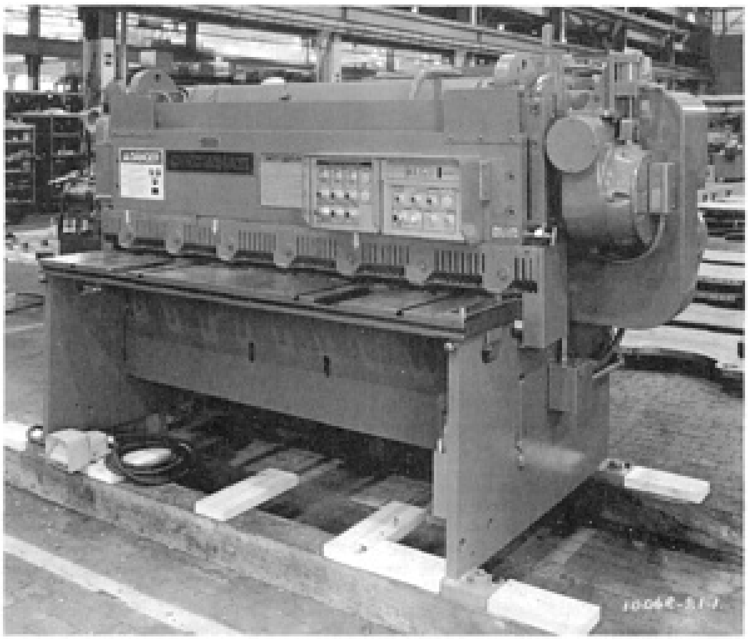

FIG. 2 TYPICAL SQUARING SHEAR. COURTESY OF CINCINNATI INC.

The sheet or plate is held rigid by hold-down devices while the upper knife moves down past the lower knife. Most sheet or plate is sheared by setting the upper knife at an angle. The position of one knife can be adjusted to maintain optimal clearance between the knives. Squaring shears can be actuated mechanically, hydraulically, or pneumatically.

Mechanical Shears. The power train of a mechanical shear consists of a motor, the flywheel, a worm shaft that is gear driven by a flywheel, a clutch that connects the worm gear drive to the driveshaft, and a ram actuated by the driveshaft through eccentrics and connecting links. Under most operating conditions, a mechanical shear can deliver more strokes per minute (spm) than a hydraulic shear. Some mechanical shears cycle as fast as 100 spm.

Another advantage of the mechanical shear is that, because of the energy stored in the flywheel, a smaller motor can be used for intermittent shearing. For example, a mechanical shear with a no-cutting or free-running speed of 65 spm can make approximately six full shearing strokes (maximum thickness and length of cut) per minute with a standard motor. However, when the same shear is cutting at full capacity in a rapid shear mode, a much larger motor is required. For such rapid cutting, there is not enough time between cuts for the smaller motor to restore the speed of the flywheel.

An additional advantage of the mechanical shear is that its moving knife travels faster than the moving knife of a hydraulic shear. In some cases, greater knife speed can decrease work metal twist, bow, and camber.

Hydraulic shears are actuated by a motordriven pump that forces oil into a cylinder against a piston; the movement of the piston energizes the ram holding the upper knife. A hydraulic shear can make longer strokes than a mechanical shear. In general, long shears and shears with low-carbon steel capacities above 12.7 mm ($\frac{1}{2}$ in.) are almost all hydraulic.

Hydraulic shears are designed with a fixed load capacity. This prevents the operator from shearing material that exceeds capacity and therefore saves costly damage to the machine structure. This is a basic advantage of hydraulic shears.

The total load that is experienced during the cut is related to the rake angle, sharpness of the knives, mechanical properties of the material, type of material, knife clearance, and the depth of the back piece. It is possible to stall the machine on a rated-capacity cut if the clearance is incorrect, the knife is dull, or the back piece is excessively deep. In this way, the hydraulic shear is protected from damage caused by overloading. A mechanical shear would not be constrained by an overload prevention system and would continue to cut under nearly all conditions.

Most mechanical shears are provided with enough horsepower to build up the flywheel speed after each cut, but not enough to allow the operator to run full-capacity cuts in a high-speed manner. Mechanical shears are rated in strokes per minute, not cuts per minute. Most shearing applications do not require high-speed cutting.

Hold-down pressure must be greater than the forces generated in cutting the work material. These forces depend on the knife clearance, rake angle, and depth of material back piece.

Greater knife clearance must be used to prevent the hydraulic shears from stalling when shearing large-area work, especially at or near maximum thickness capacity. For example, in cutting a 3 × 3 m (10 × 10 ft) plate into two equal parts, greater shearing force is required than for trimming a narrow strip from a 3 m (10 ft) long plate.

Capacity. Most shearing machines are rated according to the section size of low-carbon steel they can cut. The tensile strength of low-carbon steel sheet and plate is generally no higher than 520 MPa (75 ksi); the yield strength, no greater than 350 MPa (51 ksi). Shears frequently are rated in terms of their ability to cut low-carbon steel with a tensile strength of 414 MPa (60 ksi) and yield strength of 276 MPa (40 ksi). An allowance for normal over-tolerance material thickness is included in the capacity rating of the machine. The use of a machine for shearing other metals is primarily based on the relationship of the tensile strength and ductility of low-carbon steel to that of the metal to be sheared. Metals with a tensile strength greater than that of low-carbon steel almost always reduce the capacity of the machine. For example, the machine capacity for shearing high-strength low-alloy steels is reduced to about two-thirds to three-quarters of the rated capacity for low-carbon steel. Conversely, for shearing aluminum alloys, machine capacity can range from $1\frac{1}{4}$ to $1\frac{1}{2}$ times the rated capacity for low-carbon steel.

Table 1 compares the shearing capacities of various metals with those of low-carbon steel. The metal thicknesses given in Table 1 are based on the thickness of low-carbon steel that can be sheared with the same shearing capacity. For example, a specific force is required to shear 6.4 mm ($\frac{1}{4}$ in.) thick low-carbon steel. Table 1 shows that the same force can shear only a 4.8 mm ($\frac{3}{16}$ in.) thickness of type 302 stainless steel, but can shear a 9.5 mm ($\frac{3}{8}$ in.) thickness of aluminum.

TABLE 1 SHEARING CAPACITIES FOR VARIOUS METALS COMPARED TO THOSE FOR LOW-CARBON STEEL

THICKNESS OF LOW-CARBON STEEL(A)		THICKNESS THAN CAN SHEARED WITH SAME FORCE AS FOR LOW-CARBON STEEL					
		AISI TYPE 203 STAINLESS STEEL (B)		FULL-HARD STEEL STRIP		ALUMINUM ALLOYS	
mm	**in.**	**mm**	**in.**	**mm**	**in.**	**mm**	**in.**
1.52	0.060	0.91	0.036	1.22	0.048	1.90	0.075
1.90	0.075	1.22	0.048	1.52	0.060	3.05	0.120
3.05	0.120	1.52	0.060	1.90	0.075	3.40	0.134
3.40	0.134	1.90	0.075	2.67	0.105	4.8	$\frac{3}{16}$
4.8	$\frac{3}{16}$	3.40	0.134	3.9	$\frac{5}{32}$	5.6	$\frac{7}{32}$
6.4	$\frac{1}{4}$	4.8	$\frac{3}{16}$	4.8	$\frac{3}{16}$	6.4	$\frac{1}{4}$
7.9	$\frac{5}{16}$	5.6	$\frac{7}{32}$	5.6	$\frac{7}{32}$	9.5	$\frac{3}{8}$
9.5	$\frac{3}{8}$	6.4	$\frac{1}{4}$	6.4	$\frac{1}{4}$	11.1	$\frac{7}{16}$
11.1	$\frac{7}{16}$	7.9	$\frac{5}{16}$	7.9	$\frac{5}{16}$	12.7	$\frac{1}{2}$
12.7	$\frac{1}{2}$	9.5	$\frac{3}{8}$	9.5	$\frac{3}{8}$	15.9	$\frac{5}{8}$

15.9	$\frac{5}{8}$	11.1	$\frac{7}{16}$	11.1	$\frac{7}{16}$	19.0	$\frac{3}{4}$
19.0	$\frac{3}{4}$	12.7	$\frac{1}{2}$	12.7	$\frac{1}{2}$	25.4	1
22.2	$\frac{7}{8}$	15.9	$\frac{5}{8}$	15.9	$\frac{5}{8}$	31.8	$1\frac{1}{4}$
25.4	1	19.0	$\frac{3}{4}$	19.0	$\frac{3}{4}$	38.1	$1\frac{1}{2}$
31.8	$1\frac{1}{4}$	25.4	1	25.4	1	50.8	2

(A) ALSO APPLICABLE TO SOFT TO HALF-HARD STRIP STEEL, ALCLAD STEEL, AND COPPER AND COPPER ALLOYS.

(B) ALSO APPLIES TO MOST OTHER AUSTENITIC STAINLESS STEELS, NORMALIZED ALLOY STEELS SUCH AS 4130 OR 8630, ANNEALED HIGH-CARBON STEELS, AND ANNEALED TOOL STEELS

Ductility, measured by the elongation of the work metal, can also affect machine capacity. For example, annealed copper, because of its high elongation, requires as much shearing effort as low-carbon steel, even though copper has considerably lower tensile strength. Similarly, carbon steel with very low carbon (<0.1% C) and higher-than-normal elongation will reduce the capacity of a machine.

Accessory Equipment for Straight-Knife Shearing

Certain pieces of accessory equipment have been incorporated into most shear designs and are required for efficient and accurate straight-knife shearing.

Hold-downs (Fig. 2) are mechanical or hydraulic devices that hold the work metal firmly in position to prevent movement during shearing. The most efficient hold-down system is a series of independent units that securely clamps stock of varying thicknesses automatically and without adjustment.

The force on each hold-down foot must be substantial and can range from several hundred pounds on a machine for shearing sheet to several tons for shearing plate. Hold-downs must be timed automatically with the stroke of the ram so that they clamp the work metal securely before the knife makes contact and release their hold instantly after shearing is completed.

Back gages are adjustable stops that permit reproducibility of dimensions of sheared work-pieces in a production run. Most gages are controlled electrically. Push-button control provides a selection of fast traverse speeds and slow locating movements for accurate final positioning. The addition of a microcomputer (Fig. 3) permits gage positions to be quickly entered. An LED display enables the operator to confirm immediately the gage position entered, the current gage position, or the final gage location after positioning. Accurate gage screws, compensating nuts, precision slides and guides, and decimal indicators permit repeatable gage settings to an accuracy of 0.025 to 0.05 mm (0.001 to 0.002 in.).

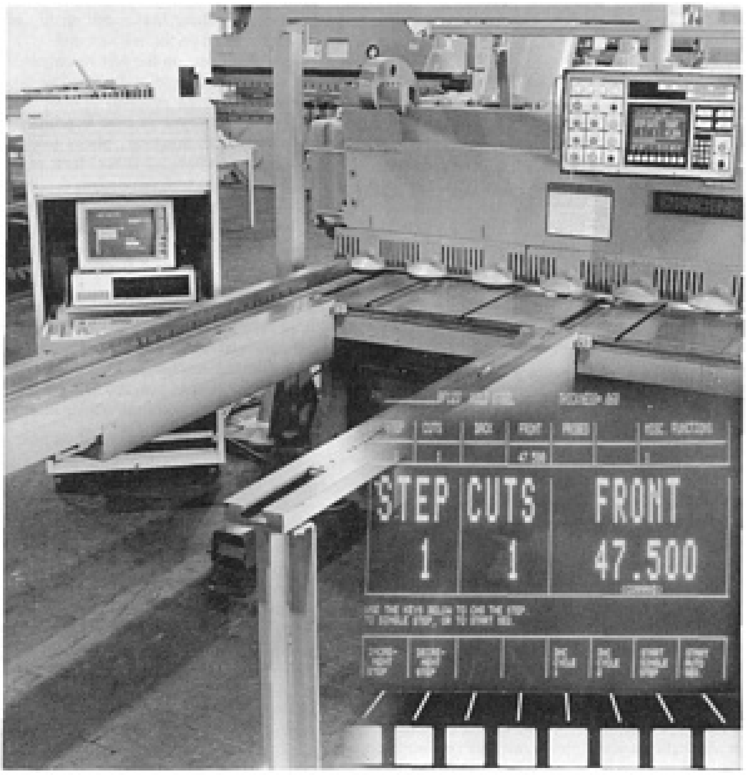

FIG. 3 MICROPROCESSOR-CONTROLLED SHEARING MACHINE WITH INSET (LOWER RIGHT) SHOWING LED DISPLAY. COURTESY OF CINCINNATI INC.

For thin sheet, magnetic overhead rollers eliminate sag and support the sheet for accurate gaging to a depth of 1.2 m (4 ft) into the shear. For rapid and accurate cutting, back gages are equipped with electronic sensors that automatically trip the shear only when the sheet is accurately positioned.

Pneumatic sheet supports are used to support ferrous and nonferrous thin sheet. Sheet support arms are designed to elevate into a horizontal position, flush with the shear table, permitting material to be supported in the correct position against the backgage stop. Blank inaccuracies due to unsupported and poorly positioned sheets are virtually eliminated.

Back gages are also equipped with retractable stops for shearing mill plate. With the stops out of the way, mill plate of almost any length can be fed into the shear and cut to the desired length. When stops are not used, the workpiece can be notched or scribed to indicate the cutoff position.

Front Gages. When gaging from the front of the machine, the operator locates the work metal by means of stops secured in the table or in the front support arms. Power operation of the front support arms allows the blank dimensions to be entered digitally using a microcomputer-based control. Front gaging is often done by means of a squaring arm.

Squaring arms (Fig. 4) are extensions attached to the entrance side of a shearing machine that are used to locate long sections of work metal in the proper position for shearing. Each arm is provided with a linear scale and with stops for accurate, consistent positioning of the work metal. Squaring arms are reversible to allow use of the shear at either end and to distribute the wear on the shear knives. Power operation can be added to the squaring arm for increasing blank accuracy and reducing setup time. This type of squaring arm prevents the arm from moving from side to side.

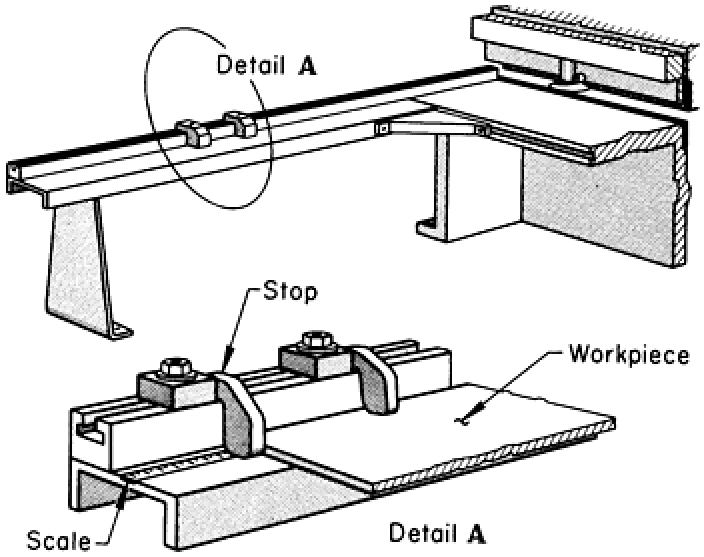

FIG. 4 SQUARING ARM ATTACHMENT FOR POSITIONING LONG PIECES IN A SHEARING MACHINE

Knife Rake

Rake is the slope of the angle formed by the cutting edges of the upper and lower knives (Fig. 2). It is usually expressed as the ratio of the amount of rise to a given linear measurement. For example, a rake of 21 mm/m ($\frac{1}{4}$ in./ft) means that the upper knife rises 21 mm for each meter ($\frac{1}{4}$ in. for each foot) of linear distance along the knives. Rakes below 21 mm/m ($\frac{1}{4}$ in./ft) are rarely used; a rake of 42 mm/m ($\frac{1}{2}$ in./ft) or higher is typical of many plate shears.

Rake is used to permit progressive shearing of the work metal along the length of the knife. This reduces the amount of force required and allows the use of a smaller machine than would be necessary if the cutting edges of the knives were parallel.

It is not possible, however, to calculate the required force for different work metal thicknesses based solely on change of rake, because knife penetration varies for different thicknesses. Even for low-carbon steel, the amount of knife penetration before fracture occurs can be as great as 60% of the work metal thickness for 3.43 mm (0.135 in.) thick stock and as little as 30% for 19 mm ($\frac{3}{4}$ in.) thick stock.

The primary disadvantage of using a high rake angle is that it increases the distortion of the work. Large rake angles can also cause slippage and therefore require high hold-down forces.

Knife Clearance

Excessive knife clearance causes the work metal to be wiped down between the knives during cutting and results in heavy burring or flanging of the work metal. The burrs and deformed metal are objectionable because of their interference with subsequent processing. A more serious consequence of excessive clearance is that it can cause the workpiece to be pulled between the knives, and this in turn causes overloading of the machine and may result in failure of machine components or shear knives.

When soft metals are sheared, insufficient clearance causes double (secondary) shearing, which appears as a burnished area at the top and bottom of a sheared edge with a rough area between the burnished edges. Because of the difference in action, greater tolerances are usually required when shearing plate than when shearing sheet--all other conditions being equal. Some mechanical shears are constructed to operate with a fixed clearance, and no adjustments are made for variations in work metal composition or thickness. The knife clearance is set for the thinnest material to be sheared, and double shear is avoided if the range of thicknesses sheared is not too large.

Because of deflection, clearance at the center of the knife is usually set less than that at the ends. Knife clearance (except on machines using a fixed clearance) is generally increased as work metal thickness increases. For example, in one plant, a clearance of approximately 0.076 mm (0.003 in.) is used for squaring shears with a capacity of up to 6.4 mm ($\frac{1}{4}$ in.) thick low-carbon steel. To minimize knife deflection, clearance at the center of the knife is reduced to 0.051 mm (0.002 in.). Similarly, the clearance for shearing 6.4 to 25 mm ($\frac{1}{4}$ to 1 in.) thick low-carbon steel is 0.36 mm (0.014 in.) at each end of the knives and 0.30 mm (0.012 in.) at the center of 3 to 3.7 m (10 to 12 ft) long knives used in a mechanical shear.

Ram Speed in Straight-Knife Shearing

The speed of the ram (and, in turn, that of the knife) has a slight effect on results in the shearing of flat stock. Low linear speed produces a rough sheared surface. As speed is increased, a cleaner sheared surface is obtained. In general, speeds to 21 to 24 m/min (70 to 80 ft/min) can be used without difficulty when shearing annealed metals. Regardless of the speed used, adequate hold-down force is mandatory.

Accuracy in Straight-Knife Shearing

The dimensional accuracy obtained in shearing is influenced by the capacity and condition of the machine, condition of the knives, knife clearance, work metal thickness, and work metal condition. Expressed as total tolerance, sheets no thicker than approximately 3.43 mm (0.135 in.) can be cut to size within 0.25 mm (0.010 in.) and strips can be sheared to width at the same total tolerance. These tolerances apply to stock up to 3.7 m (12 ft) long that is essentially free from stress and is flat within commercial limits. Sheets that are not flat or that have residual stress, or both, cannot be sheared with the same accuracy.

Greater tolerances are required in shearing plate. A total tolerance of 0.5 to 1.0 mm (0.020 to 0.040 in.) can be maintained when plate is sheared in squaring shears. Dimensions can be held to tolerance of about ±1.6 mm ($\pm\frac{1}{16}$ in.) when shearing in alligator shears.

Safety

Shearing machines must be equipped with devices for protecting personnel from the hazards of shear knives, flywheels, gears, and other moving parts. The guards and safety devices used must be rigid enough to withstand damage from operating personnel moving heavy material into position.

The squaring shears used for sheet metal should have guards on all moving parts, including flywheels, hold-downs, and knives. The treadle, whether mechanical or electrical, should have a lock for supervisory control. Knife and hold-down guard openings should be large enough to provide visibility but small enough to keep the operator's fingers out of the danger area. Proper opening dimensions are outlined in ANSI standard B11.4 (latest edition).

The shears used for shearing plate are more difficult to safeguard because of the greater clearances needed under the hold-downs and upper knife to permit entry of the plate (especially when it is bowed or buckled). Guards on shears for plate should be of the type that raises only when the plate is inserted and then rests on the surface of the plate. When there is no workpiece in the machine, the guards rest within 6.4 mm ($\frac{1}{4}$ in.) of the surface of the table.

Shears should comply with the construction requirements of the Occupational Safety & Health Act and National Safety Standards such as ANSI B11.4. Additional safety information can be obtained from the loss prevention group of major insurance carders for Workman's Compensation and the National Safety Council. Safety regulations also cover the maximum noise level permitted from a shearing operation to prevent permanent impairment of hearing.

Mechanical Cutting for Weld Preparation

Iron Workers

Iron workers (heavy-duty shears) are used primarily for preparing structural shapes such as bars, angles, and T-bars. Like the shears used for plate and flat sheet, iron workers leave a square butt edge that may require further preparation for heavier sections before welding. The production shearing of bars and bar sections is usually done in machines with a throat opening designed for large, bulky workpieces. Guillotine and multipurpose (combination) machines are widely used for welding preparation. The multipurpose machines feature interchangeable punches and dies for shearing, punching, and coping. Squaring shears, normally used for sheet and plate, can also be used for cutting bar stock to length. Punch presses and press brakes can be provided with appropriate tooling for cutting operations.

Combination Machines

Combination machines are multipurpose machines used primarily in metal-fabricating shops where there is a constant need for shearing small quantities (often two or three pieces) of a variety of shapes and sizes from bars or bar sections. Some combination machines can also be used for punching, slotting, and notching.

Many combination machines incorporate several devices within the frame for performing different operations; therefore, a new setup is not required for each. A holder for an interchangeable punch and die is located in an area with a deep throat. This facilitates the punching of holes, slots, or notches in plates and bars, and in webs or legs of structural members.

A slide moving at 45° from vertical carries a blade for shearing angles. The support bed is on a swivel so that the ends of the angle section can be varied as desired from 45 to 90°. In two strokes of the machine, angles can be sheared to produce miter joints for subsequent welding (see the section "Cutting of Angles" in this article).

Combination machines can be used for cutting square and rectangular notches in the leg of an angle. These machines can be set up to cut a 90° V-shaped notch in angles that subsequently will be bent into frames. Other shapes, such as beams and channels, can be notched in a similar manner if the machine can accommodate the vertical height between the upper and lower legs of the workpiece. Provision is also made for shearing bars with guillotine-type blades or special blades (see the section "Conforming Blades" in this article).

Combination machines are available in capacities ranging from 110 to 890 kN (12 to 100 tonf). The 110 kN (12 tonf) machine can punch a 14 mm ($\frac{9}{16}$ in.) diam hole through a 6.4 mm ($\frac{1}{4}$ in.) thick section and can shear 75 × 75 × 6.4 mm (3 × 3 × $\frac{1}{4}$ in.) angles, 22 mm ($\frac{7}{8}$ in.) diam rounds, 19 mm ($\frac{3}{4}$ in.) squares, and 102 × 6.4 mm (4 × $\frac{1}{4}$ in.) flats. The 890 kN (100 tonf) machine can shear 152 × 152 × 16 mm (6 × 6 × $\frac{5}{8}$ in.) rounds, 50 mm (2 in.) squares, and 203 × 19 mm (8 × $\frac{3}{4}$ in.) flats.

Guillotine Machines

Guillotine iron-working machines are designed for cutting bars and bar sections to desired lengths from mill stock. They are extensively used throughout the fabricating industry. Two general types are available: open end (Fig. 5) and closed end. The machine illustrated in Fig. 5 is termed an open-end machine because it has a C-frame construction with one end open and unsupported. Open-end machines are either single end, for one operator, or double end, for two operators. On double-end machines, both ends can be right-hand or left-hand, or one end can be right-hand and the other end left-hand, depending on the type of cutting to be done. A closed-end guillotine machine, on the other hand, is basically the same as the one shown in Fig. 5 except that it has frame supports on both sides.

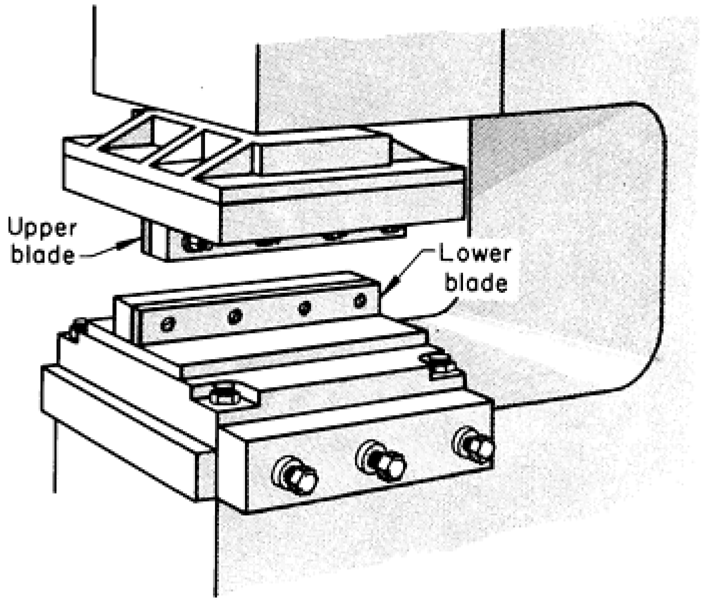

FIG. 5 OPEN-END GUILLOTINE IRON-WORKING MACHINE

An open-end machine has the advantage of giving the operator a clear view of the blades. However, because one end is open, a heavier frame and more floor space are required than for a closed-end machine of equal capacity.

Guillotine iron workers for bars and angles are available in capacities to 2700 kN (300 tonf). Either intermittent or continuous operation is possible. Guillotine iron workers can be equipped with simple straight blades (Fig. 5) or with two or more short blades having specific shapes.

Guillotine iron workers are actuated mechanically, hydraulically, or pneumatically. Hydraulic and pneumatic machines are lighter in weight for a specific shearing power, are more economical, and operate with less vibration than mechanical iron workers. Therefore, hydraulic or pneumatic machines up to 89 kN (10 tonf) capacity are completely portable, and units of 1800 to 2700 kN (200 to 300 tonf) capacity are semiportable (need not be solidly mounted in concrete).

Punching and Shearing Machines

Punching and shearing machines are deep-throat C-frame machines for the punching, shearing, notching, or coping of plates, bars, and structural sections. Shoes, into which punches, dies, and shear blades can readily be inserted, are mounted on the bed and ram.

The shoes either rest on the bed or are overhung; the overhanging type is designed so that structural shapes can be punched in both web and flange. The plain type of shoe is primarily used for plate work; however, plates can be worked with the overhanging shoes. Both types of die blocks are fitted with die sockets that hold dies of different inside diameters. The punch holder is adjustable to suit the location of the dies.

Fixturing

Fixturing is an important consideration in the shearing of bars. For safety and the proper functioning of open-end machines (Fig. 5) and many other iron-working machines, hold-down fixtures are essential. Guide pins are also helpful, especially when cutting with conforming blades.

Blade Design and Production Practice

Straight-edge blades can cut almost any bar or shape that is within the capacity of the machine. However, unacceptable distortion may result in some shapes of workpieces when they are sheared with blades that are not designed for cutting specific edges.

Conforming Blades. One method of minimizing distortion in cut bars employs two hardened blades mounted face-to-face, with identical holes through each blade. The holes should conform to the shape of the work metal and should be large enough to allow easy passage through the blades (Fig. 6a). One blade is movable vertically and one is stationary. Relatively little movement of the machine is required when blades of this type are used. In addition, because the blades completely encircle the work metal, hold-downs are not needed. However, these blades are usually limited for use on specially built or combination machines.

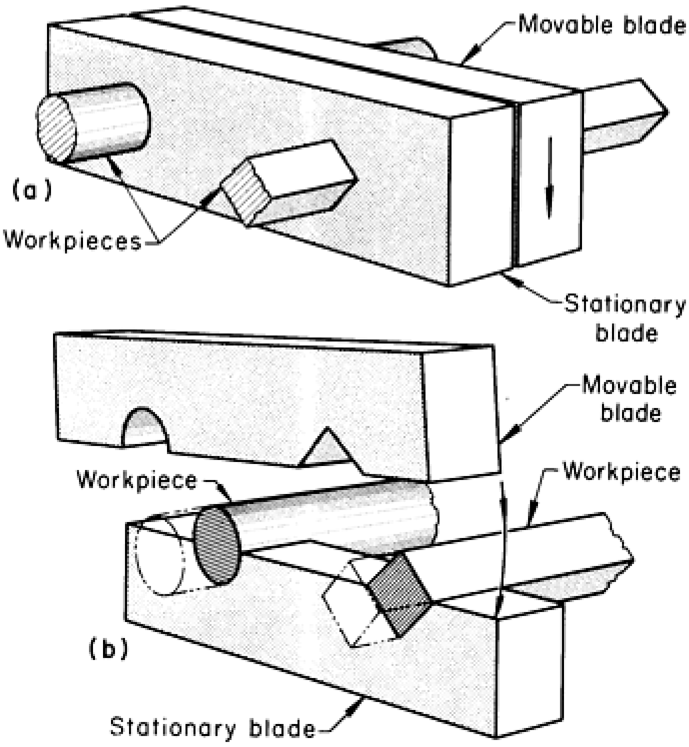

FIG. 6 TWO TYPES OF BLADES FOR THE CUTTING OF BARS

The cutting of round and square bars is more frequently done with the open-type blades illustrated in Fig. 6(b). Each blade is contoured to accommodate one-half the cross section of the work metal. The upper blade moves in a vertical direction, while the lower blade remains stationary. When using this technique, some type of hold-down is needed. Because of the stiffness of the work metal, the hold-down for bars should permit slight movement of the work metal in the axial direction to avoid double shearing. The hold-down can be a simple set screw (to permit adjustment) in a bracket or can be a more elaborate unit, such as a handwheel assembly utilizing an Acme thread.

For cutting square bars with any type of blade, the work metal should be placed so that the movement of the blade is across the diagonal of the square. With this technique the cutting force is applied to four sides instead of two, resulting in a smoother sheared surface. Cutting across the diagonal provides support on two sides of the square shape, which minimizes distortion and permits more than one size of bar stock to be cut in a given hole.

Best practice for cutting round bars is to use blades with holes for each size of stock to be cut. Blade holes appreciably larger than the stock size cause excessive distortion of the workpiece.

Cutting of angles is done either in a combination machine or by double cutting. In a combination machine--the more common method--two blades such as those shown in Fig. 7(a) are used. One blade, usually the one that is stationary, is L-shaped and is positioned as shown. The movable blade is square or rectangular and is mounted with its two cutting edges parallel to those of the stationary blade. Figure 7(a) also shows that the space between the blades in the loading position is the same shape as the workpiece.

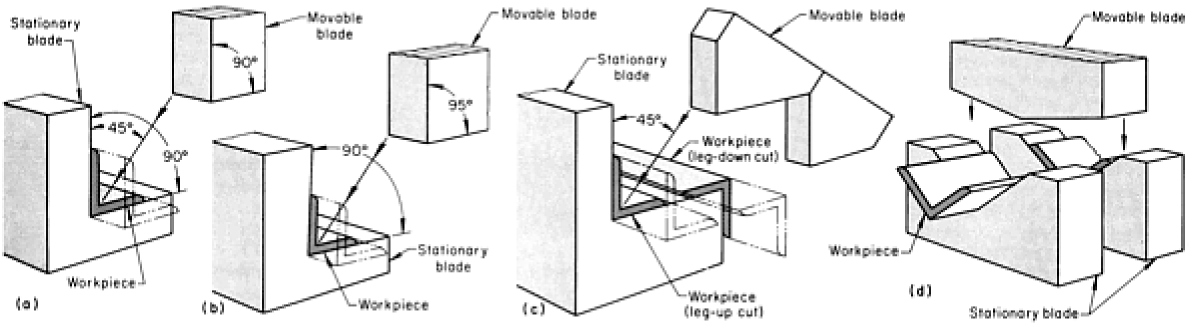

FIG. 7 FOUR TYPES OF BLADES FOR THE CUTTING OF ANGLE SECTIONS. (A) TO (C) CUTTING IN A COMBINATION MACHINE. (D) DOUBLE-CUTTING METHOD

The movable blade travels at 45° toward the stationary blade, and both blades contact the work metal uniformly. Shearing by this technique is essentially a blanking cut, and distortion of the workpiece is minimal. One disadvantage of the method is that all cutting occurs at once, resulting in a high shear load. This condition is not important when small angle sections are cut; however, for work metal larger than 102 × 102 × 13 mm (4 × 4 × $\frac{1}{2}$ in.), the movable blade should be provided with rake to prevent excessive loading.

To provide a rake angle between the movable and stationary blades, the included angle between the cutting edges of the upper or movable blade is increased to 95°, as shown in Fig. 7(b). Shearing begins at the extremity of each leg and progresses toward the root of the angle. The increase in the included angle of the movable blade results in some distortion of the drop-cut piece; the amount of distortion is about equal to the difference in angle between the movable and stationary blades (5° is normal). The part remaining on the table or stationary blade is not distorted.

Most combination cutting machines use a more versatile blade arrangement than those shown in Fig. 7(a) and 7(b). The setup shown in Fig. 7(c) is used to shear angle sections in both the leg-up and the leg-down positions. A swiveling table locates and holds the workpiece during shearing. With the swiveling table and two positions for the workpiece, the flanges can be easily mitered to any specific angle. For example, when cutting angle sections for a frame having the leg on the inside, the table would be set and locked at 45°. One end is mitered by placing the section in a leg-down position on the table and shearing off enough to make a clean cut. The other end is mitered by placing the section leg-up on the table and cutting to the proper length. The opposite positions are applicable when angle sections for a frame having the leg on the outside are being cut.

Cutting at a 45° angle reduces the capacity of the machine because a greater length of metal is cut at one time when a 90° cut is made. For example, a machine with a capacity of 203 × 203 × 32 mm (8 × 8 × $1\frac{1}{4}$ in.) when making a 90° cut has a capacity of only 203 × 203 × 25 mm (8 × 8 × 1 in.) when cutting at 45°.

Double cutting of angle sections, also called slugging, is used less frequently than cutting in a combination-type machine. This technique uses two stationary blades, spaced 13 mm ($\frac{1}{2}$ in.) apart, and one movable V-shaped blade arranged as shown in Fig. 7(d). The movable blade has a shallow V-shape that does not conform to the shape of the workpiece. Shearing starts at the extremity of each leg and progresses to the root of the angle, producing a 13 mm ($\frac{1}{2}$ in.) wide slug that is pushed out the bottom between the stationary blades.

In the double-cutting method, distortion occurs only in the slug, because the work metal is supported by the two stationary blades. There are two disadvantages of using the double-cutting method. First, increased power is required for making two cuts at the same time, and second, some metal is lost in the slug. The two stationary blades must be firmly supported to prevent their spreading during the cutting operation.

A similar tool can be used for cutting a channel section. The stationary blades should closely fit the contour of the channel section. Double cutting is adaptable to a guillotine machine, a combination machine, or a press.

Multiple Setups. Fabricating shops often must cut small quantities to a variety of shapes. To handle such work, many shops use a machine with a multiple setup. Without changing blades, the user can perform operations such as double cutting of angle sections, straight-blade cutting, cutting of round and square bars, and single cutting of L-sections. In this type of setup, all of the movable blades are attached to a single ram.

Mechanical Cutting for Weld Preparation

ASM Committee on Mechanical Cutting for Welding Preparation*

Iron Workers

Iron workers (heavy-duty shears) are used primarily for preparing structural shapes such as bars, angles, and T-bars. Like the shears used for plate and flat sheet, iron workers leave a square butt edge that may require further preparation for heavier sections before welding. The production shearing of bars and bar sections is usually done in machines with a throat opening designed for large, bulky workpieces. Guillotine and multipurpose (combination) machines are widely used for welding preparation. The multipurpose machines feature interchangeable punches and dies for shearing, punching, and coping. Squaring shears, normally used for sheet and plate, can also be used for cutting bar stock to length. Punch presses and press brakes can be provided with appropriate tooling for cutting operations.

Combination Machines

Combination machines are multipurpose machines used primarily in metal-fabricating shops where there is a constant need for shearing small quantities (often two or three pieces) of a variety of shapes and sizes from bars or bar sections. Some combination machines can also be used for punching, slotting, and notching.

Many combination machines incorporate several devices within the frame for performing different operations; therefore, a new setup is not required for each. A holder for an interchangeable punch and die is located in an area with a deep throat. This facilitates the punching of holes, slots, or notches in plates and bars, and in webs or legs of structural members.

A slide moving at 45° from vertical carries a blade for shearing angles. The support bed is on a swivel so that the ends of the angle section can be varied as desired from 45 to 90°. In two strokes of the machine, angles can be sheared to produce miter joints for subsequent welding (see the section "Cutting of Angles" in this article).

Combination machines can be used for cutting square and rectangular notches in the leg of an angle. These machines can be set up to cut a 90° V-shaped notch in angles that subsequently will be bent into frames. Other shapes, such as beams and channels, can be notched in a similar manner if the machine can accommodate the vertical height between the upper and lower legs of the workpiece. Provision is also made for shearing bars with guillotine-type blades or special blades (see the section "Conforming Blades" in this article).

Combination machines are available in capacities ranging from 110 to 890 kN (12 to 100 tonf). The 110 kN (12 tonf) machine can punch a 14 mm ($\frac{9}{16}$ in.) diam hole through a 6.4 mm ($\frac{1}{4}$ in.) thick section and can shear 75 × 75 × 6.4 mm (3 × 3 × $\frac{1}{4}$ in.) angles, 22 mm ($\frac{7}{8}$ in.) diam rounds, 19 mm ($\frac{3}{4}$ in.) squares, and 102 × 6.4 mm (4 × $\frac{1}{4}$ in.) flats. The 890 kN (100 tonf) machine can shear 152 × 152 × 16 mm (6 × 6 × $\frac{5}{8}$ in.) rounds, 50 mm (2 in.) squares, and 203 × 19 mm (8 × $\frac{3}{4}$ in.) flats.

Guillotine Machines

Guillotine iron-working machines are designed for cutting bars and bar sections to desired lengths from mill stock. They are extensively used throughout the fabricating industry. Two general types are available: open end (Fig. 5) and closed end. The machine illustrated in Fig. 5 is termed an open-end machine because it has a C-frame construction with one end open and unsupported. Open-end machines are either single end, for one operator, or double end, for two operators. On double-end machines, both ends can be right-hand or left-hand, or one end can be right-hand and the other end left-hand, depending on the type of cutting to be done. A closed-end guillotine machine, on the other hand, is basically the same as the one shown in Fig. 5 except that it has frame supports on both sides.

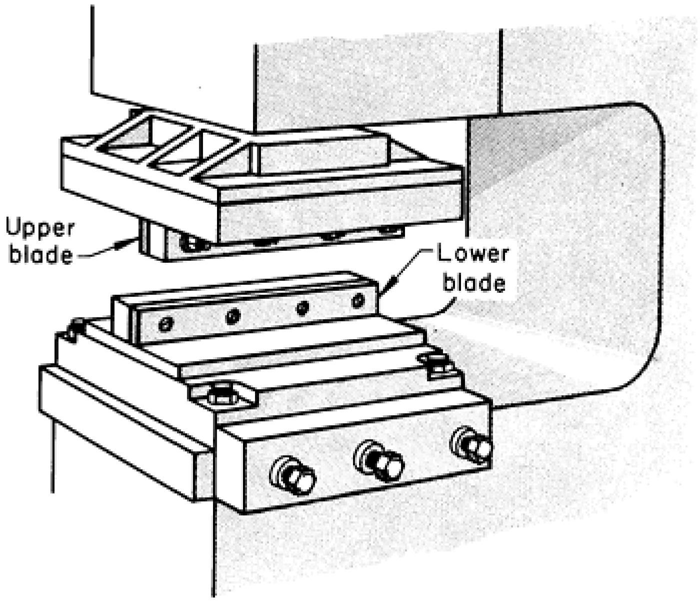

FIG. 5 OPEN-END GUILLOTINE IRON-WORKING MACHINE

An open-end machine has the advantage of giving the operator a clear view of the blades. However, because one end is open, a heavier frame and more floor space are required than for a closed-end machine of equal capacity.

Guillotine iron workers for bars and angles are available in capacities to 2700 kN (300 tonf). Either intermittent or continuous operation is possible. Guillotine iron workers can be equipped with simple straight blades (Fig. 5) or with two or more short blades having specific shapes.

Guillotine iron workers are actuated mechanically, hydraulically, or pneumatically. Hydraulic and pneumatic machines are lighter in weight for a specific shearing power, are more economical, and operate with less vibration than mechanical iron workers. Therefore, hydraulic or pneumatic machines up to 89 kN (10 tonf) capacity are completely portable, and units of 1800 to 2700 kN (200 to 300 tonf) capacity are semiportable (need not be solidly mounted in concrete).

Punching and Shearing Machines

Punching and shearing machines are deep-throat C-frame machines for the punching, shearing, notching, or coping of plates, bars, and structural sections. Shoes, into which punches, dies, and shear blades can readily be inserted, are mounted on the bed and ram.

The shoes either rest on the bed or are overhung; the overhanging type is designed so that structural shapes can be punched in both web and flange. The plain type of shoe is primarily used for plate work; however, plates can be worked with the overhanging shoes. Both types of die blocks are fitted with die sockets that hold dies of different inside diameters. The punch holder is adjustable to suit the location of the dies.

Fixturing

Fixturing is an important consideration in the shearing of bars. For safety and the proper functioning of open-end machines (Fig. 5) and many other iron-working machines, hold-down fixtures are essential. Guide pins are also helpful, especially when cutting with conforming blades.

Blade Design and Production Practice

Straight-edge blades can cut almost any bar or shape that is within the capacity of the machine. However, unacceptable distortion may result in some shapes of workpieces when they are sheared with blades that are not designed for cutting specific edges.

Conforming Blades. One method of minimizing distortion in cut bars employs two hardened blades mounted face-to-face, with identical holes through each blade. The holes should conform to the shape of the work metal and should be large enough to allow easy passage through the blades (Fig. 6a). One blade is movable vertically and one is stationary. Relatively little movement of the machine is required when blades of this type are used. In addition, because the blades completely encircle the work metal, hold-downs are not needed. However, these blades are usually limited for use on specially built or combination machines.

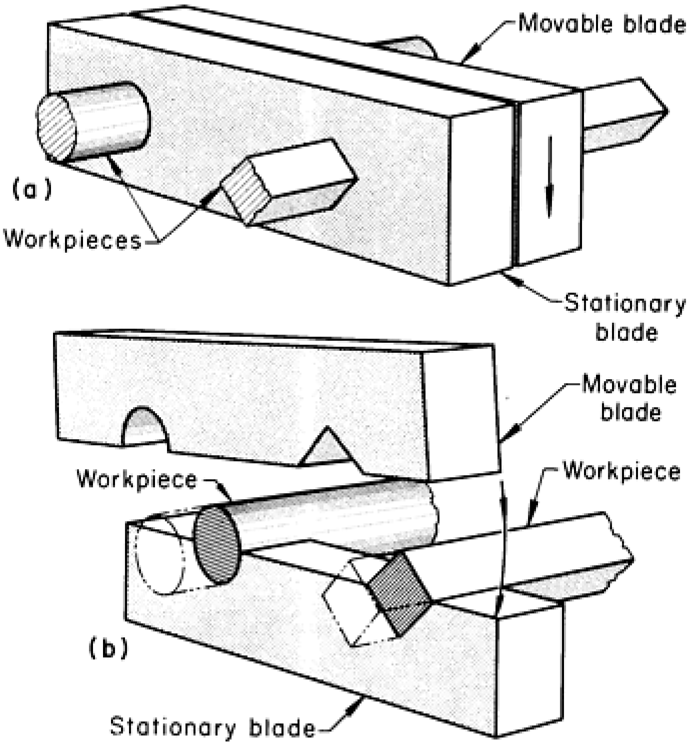

FIG. 6 TWO TYPES OF BLADES FOR THE CUTTING OF BARS

The cutting of round and square bars is more frequently done with the open-type blades illustrated in Fig. 6(b). Each blade is contoured to accommodate one-half the cross section of the work metal. The upper blade moves in a vertical direction, while the lower blade remains stationary. When using this technique, some type of hold-down is needed. Because of the stiffness of the work metal, the hold-down for bars should permit slight movement of the work metal in the axial direction to avoid double shearing. The hold-down can be a simple set screw (to permit adjustment) in a bracket or can be a more elaborate unit, such as a handwheel assembly utilizing an Acme thread.

For cutting square bars with any type of blade, the work metal should be placed so that the movement of the blade is across the diagonal of the square. With this technique the cutting force is applied to four sides instead of two, resulting in a smoother sheared surface. Cutting across the diagonal provides support on two sides of the square shape, which minimizes distortion and permits more than one size of bar stock to be cut in a given hole.

Best practice for cutting round bars is to use blades with holes for each size of stock to be cut. Blade holes appreciably larger than the stock size cause excessive distortion of the workpiece.

Cutting of angles is done either in a combination machine or by double cutting. In a combination machine--the more common method--two blades such as those shown in Fig. 7(a) are used. One blade, usually the one that is stationary, is L-shaped and is positioned as shown. The movable blade is square or rectangular and is mounted with its two cutting edges parallel to those of the stationary blade. Figure 7(a) also shows that the space between the blades in the loading position is the same shape as the workpiece.

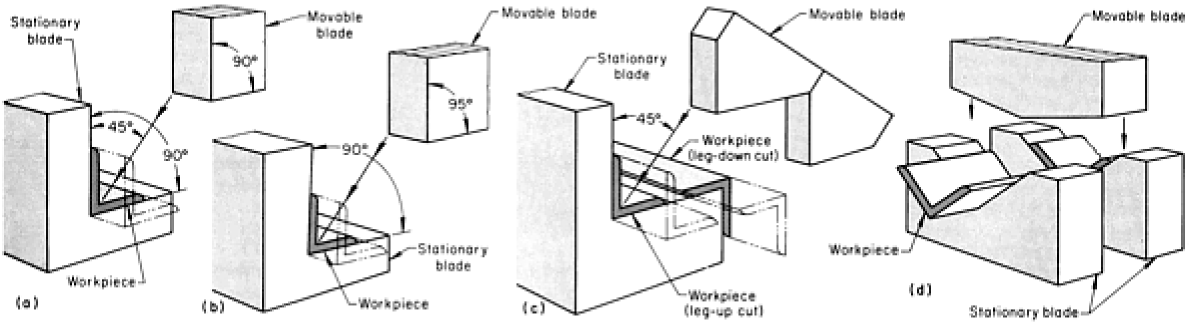

FIG. 7 FOUR TYPES OF BLADES FOR THE CUTTING OF ANGLE SECTIONS. (A) TO (C) CUTTING IN A COMBINATION MACHINE. (D) DOUBLE-CUTTING METHOD

The movable blade travels at 45° toward the stationary blade, and both blades contact the work metal uniformly. Shearing by this technique is essentially a blanking cut, and distortion of the workpiece is minimal. One disadvantage of the method is that all cutting occurs at once, resulting in a high shear load. This condition is not important when small angle sections are cut; however, for work metal larger than 102 × 102 × 13 mm (4 × 4 × $\frac{1}{2}$ in.), the movable blade should be provided with rake to prevent excessive loading.

To provide a rake angle between the movable and stationary blades, the included angle between the cutting edges of the upper or movable blade is increased to 95°, as shown in Fig. 7(b). Shearing begins at the extremity of each leg and progresses toward the root of the angle. The increase in the included angle of the movable blade results in some distortion of the drop-cut piece; the amount of distortion is about equal to the difference in angle between the movable and stationary blades (5° is normal). The part remaining on the table or stationary blade is not distorted.

Most combination cutting machines use a more versatile blade arrangement than those shown in Fig. 7(a) and 7(b). The setup shown in Fig. 7(c) is used to shear angle sections in both the leg-up and the leg-down positions. A swiveling table locates and holds the workpiece during shearing. With the swiveling table and two positions for the workpiece, the flanges can be easily mitered to any specific angle. For example, when cutting angle sections for a frame having the leg on the inside, the table would be set and locked at 45°. One end is mitered by placing the section in a leg-down position on the table and shearing off enough to make a clean cut. The other end is mitered by placing the section leg-up on the table and cutting to the proper length. The opposite positions are applicable when angle sections for a frame having the leg on the outside are being cut.

Cutting at a 45° angle reduces the capacity of the machine because a greater length of metal is cut at one time when a 90° cut is made. For example, a machine with a capacity of 203 × 203 × 32 mm (8 × 8 × $1\frac{1}{4}$ in.) when making a 90° cut has a capacity of only 203 × 203 × 25 mm (8 × 8 × 1 in.) when cutting at 45°.

Double cutting of angle sections, also called slugging, is used less frequently than cutting in a combination-type machine. This technique uses two stationary blades, spaced 13 mm ($\frac{1}{2}$ in.) apart, and one movable V-shaped blade arranged as shown in Fig. 7(d). The movable blade has a shallow V-shape that does not conform to the shape of the workpiece. Shearing starts at the extremity of each leg and progresses to the root of the angle, producing a 13 mm ($\frac{1}{2}$ in.) wide slug that is pushed out the bottom between the stationary blades.

In the double-cutting method, distortion occurs only in the slug, because the work metal is supported by the two stationary blades. There are two disadvantages of using the double-cutting method. First, increased power is required for making two cuts at the same time, and second, some metal is lost in the slug. The two stationary blades must be firmly supported to prevent their spreading during the cutting operation.

A similar tool can be used for cutting a channel section. The stationary blades should closely fit the contour of the channel section. Double cutting is adaptable to a guillotine machine, a combination machine, or a press.

Multiple Setups. Fabricating shops often must cut small quantities to a variety of shapes. To handle such work, many shops use a machine with a multiple setup. Without changing blades, the user can perform operations such as double cutting of angle sections, straight-blade cutting, cutting of round and square bars, and single cutting of L-sections. In this type of setup, all of the movable blades are attached to a single ram.

Mechanical Cutting for Weld Preparation

ASM Committee on Mechanical Cutting for Welding Preparation*

Nibblers

Nibblers are power tools that "nibble" small pieces away from the material being worked. They range in size from hand-held machines to large floor-mounted equipment. Because the hand-held machines are portable, they are suitable for preparation of large parts on site, where thermal cutting may be impractical.

Nibblers can cut flat sheet of both ferrous and nonferrous metals, as well as of fiberglass and plastic. Corrugated and trapezoidal metal sheets can also be cut. Nibblers are frequently used to make edge preparations for welding of stainless steels and nonferrous metals. Unlike shears, which can curl or distort the separated pieces, nibblers remove material, creating a kerf or cutting slot; therefore, distortion of the sheet is eliminated.

Nibblers use a reciprocating punch and stationary die to accomplish the cutting process. During operation, the punch approaches and makes contact with the material being worked. The stroke of the punch compresses the material; the increased compression causes the material to deflect. As the shear point of the material is exceeded, the material is penetrated by the punch. The material then breaks (Fig. 8), and the break continues to widen until a cone-shaped chip is ejected. The punch is then withdrawn.

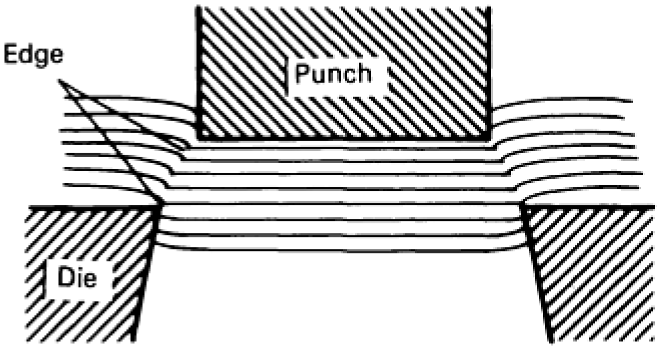

FIG. 8 SCHEMATIC OF NIBBLER CUTTING ACTION

The resulting edges are usually free of burrs. If burrs do occur, the gap between the punch and die is too wide (Fig. 9). On some equipment this gap can be varied.

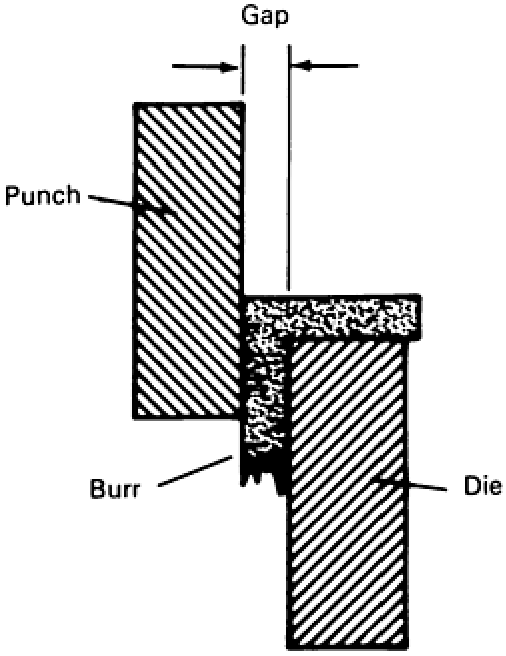

FIG. 9 SCHEMATIC OF NIBBLER SETUP WITH EXCESSIVE GAP BETWEEN THE PUNCH AND DIE, RESULTING IN THE FORMATION OF BURRS ON THE WORKPIECE

Because of their unique cutting action, nibblers do not need to be forced into the material. This so-called "natural feed" requires only light operator pressure. Also, the weight of the nibbler does not need to be supported solely by the operator: Once the nibbler has started cutting, the weight of the tool is supported by the sheet being worked. Nibblers are considered to be very safe to operate.

Some equipment is available with guide fences (similar to those used with circular saws), circle-cutting attachments, and template guides. Punches and dies on some nibblers can be resharpened; however, in many cases these parts are disposable. For optimum punch and die life, cutting oil should be used to lubricate the cutting process.

Mechanical Cutting for Weld Preparation

ASM Committee on Mechanical Cutting for Welding Preparation[*]

Band Saws

Band sawing is one method of cutting a workpiece to prepare it for further processing such as welding. The band saw machine, in its simplest form, consists of two wheels--one driven and one idle. The cutting tool is a welded continuous loop of flexible saw blade carried in one direction around the wheels and through the workpiece. This configuration permits full utilization of a cutting tool with hundreds of cutting edges, or teeth (Fig. 10). The band saw blade is thinner than other cutoff tools (typically 1.6 mm, or $\frac{1}{16}$ in. thick), which allows it to use less power and waste less material.

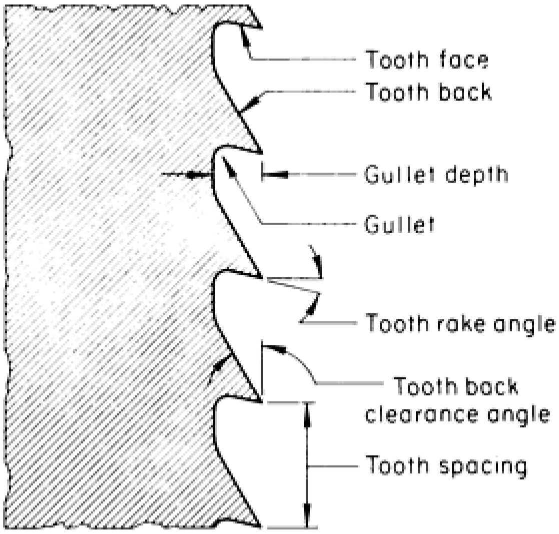

FIG. 10 SCHEMATIC SHOWING THE KEY ELEMENTS OF A BAND SAW BLADE CUTTING EDGE

Types of Machines

Band saw machines are either horizontal or vertical types; some are available as combination horizontal-vertical machines.

Contour band saws are vertical types that allow an operator to cut various contoured shapes by guiding the blade along a layout line. These machines can use very narrow bands for cutting small radii as well as wider bands for larger radii and straight cuts.

Cutoff band saws can be either horizontal or vertical. They cannot cut contours, but can make straight cuts very efficiently. Some machines can cut angles and compound miters in solids as well as structural shapes.

Band Saw Blade Selection

The two most widely used types of band saw blades available are carbon and bimetal.

Carbon steel blades are made from very high carbon tool steel and have either flexible backs or hard backs. The hard-back type has a heat-treated back for better strength and straighter cuts. Both types have hardened tooth tips for good wear resistance.

Carbon steel blades are most widely used for cutting of nonmetallic materials. However, they also do a satisfactory job when cutting low-carbon and low-alloy steels and perform well when cutting nonferrous alloys (for example, aluminum).

Bimetal blades are made with high-speed steel that is electron-beam welded to a lower-alloy backing steel. The result is a blade with the cutting properties of high-speed steel at the tooth tips and the flexibility and strength of a heat-treated alloy on the backing or carrier band. Bimetal blades should be used when cutting difficult-to-machine materials, and for high-production sawing. Bimetal blades will last longer than carbon blades, and they typically cut at ten times faster rates than those achieved with carbon blades.

Bimetal blades can be made with different grades of high-speed steels electron-beam welded to the tooth tips. The most popular grades used are M42 and Matrix II. Matrix II edges are best for cutting the lower-alloy materials and for cutting structural shapes or materials with varying cross sections. The higher hot hardness property of M42 is necessary when cutting difficult-to-machine alloys (for example, stainless steels and nickel alloys).

Table 2 gives the composition and physical properties of some welded-edge high-speed steel band saws.

TABLE 2 COMPOSITION AND PHYSICAL PROPERTIES OF SELECTED WELDED-EDGE HIGH-SPEED STEEL BAND SAW BLADES

PRODUCT	COMPOSITION, WT%								HEAT RESISTANCE		HARDNESS, HRC	
	C	Si	Mn	Cr	V	W	M	Co	°C	°F	TEETH	BODY
M2 HIGH-SPEED WELDED-EDGE BAND SAW	0.79-0.86	0.25	0.35	4.25	1.95	6.50	5.00	. . .	540	1000	64-66	40-47
MATRIX BAND SAW	0.70-0.78	0.30	0.25	4.25	1.00	1.00	5.00	8.00	590	1100	65-67	40-47
COBALT M42 BAND SAW	1.05-1.10	0.25	0.25	3.75	1.15	1.50	9.50	8.00	700	1300	67-69	40-47

Saw Width. The width of the saw blade along with its thickness determines the beam strength and thus its ability to cut straight and at high rates. Operators should always select the widest blade available in order to obtain the straightest cuts. For contour sawing, the widest blade that will cut the smallest radius should be selected.

The pitch of a saw blade is the number of teeth per 25 mm (1 in.). Selection of the correct pitch depends on the thickness of the material being cut. Thin materials require fine pitches, and thicker materials require coarser pitches. Blades are available in pitches as fine as 32 pitch and as coarse as 0.75 pitch. At least two teeth--and preferably three or more--should always be in contact with the piece being cut. It is especially important to observe this practice when cutting thin materials. A pitch that is too coarse will cause the teeth to rip out.

The selection of pitch often determines the success or failure of a particular cutting job. A pitch that is too coarse can cause the teeth to strip. A pitch that is too fine makes excessive feed pressure necessary, can cause binding in the cut, and will result in slow cutting or stripped teeth.

Variable-pitch blades are shapes designed to dampen resonance vibrations, especially when cutting structural shapes and workpieces with varying cross sections. The tooth spacing varies from tooth to tooth along an established length. This change in spacing cuts down on noise and vibration, resulting in quieter operation and smoother cuts.

Tooth Form. The choice of tooth form varies with each band saw blade manufacturer. Two basic forms are now most widely used: regular tooth and hook tooth (Fig. 11).

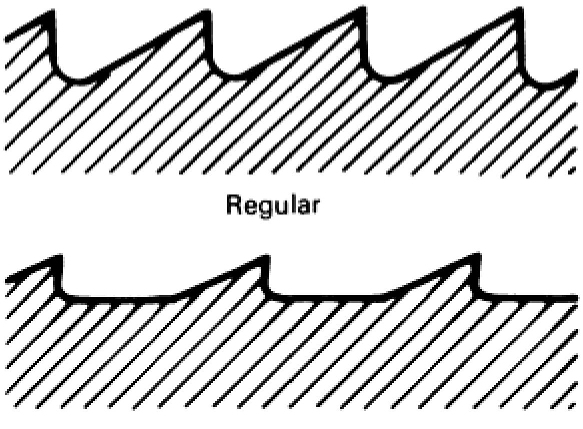

FIG. 11 BASIC TOOTH FORMS USED IN BAND SAW BLADES

The regular tooth form has deep gullets with smooth radii. The face angles vary from 0° rake to 15° positive rake. The neutral-rake blades work best for cutting structural shapes and thin materials, and for general-purpose low-volume cutting. The positive-rake blades use less feed pressure and cut more freely. This results in faster cutting rates and longer blade life, especially on heavy-duty band saws and at high production rates.

The hook tooth form is also a positive-rake blade, but the gullets are more shallow than those on the regular tooth form. The back angle is smaller than that on most regular tooth form blades. These factors combine to make a strong tooth that is capable of fast cutting speeds. The hook tooth form is especially suited for cutting of wide cross sections and for cutting of nonferrous materials such as aluminum.

Selection of Speeds and Cutting Rates

Table 3 lists the recommended speeds and cutting rates for many commonly sawed materials. The numbers are based on sawing of solid materials in the fully annealed condition. The data given should only be considered as a starting point; actual speeds and cutting rates will vary due to material variations and the type and condition of the saw that is used.

TABLE 3 SPEED AND CUTTING RATES FOR CUTOFF BAND SAWING WITH BIMETAL BLADES OF SELECTED FERROUS AND NONFERROUS METALS

WORK METAL	BAND SPEED(A)		CUTTING RATE(A)	
	m/min	sfm	$mm^2 \times 10^3$/min	in.^{2}/min
FERROUS METALS				
CARBON AND STEEL LOW-ALLOY STEELS				
1008-1013	100-84	325-275	9.0-6.4	14-10
1015-1035	106-90	350-300	9.7-7.1	15-11
1036-1064	68-58	225-190	5.8-4.5	9-7
1065-1095	52-44	170-145	5.2-3.9	8-6
1108-1132	106-84	350-275	9.7-7.7	15-12
1137-1151	80-68	260-225	6.4-5.2	10-8
1212-1213	106-90	350-300	9.7-7.7	15-12
1330-1345	65-58	210-190	5.2-3.9	8-6
4023-4047	80-70	260-230	5.2-3.9	8-6
4130-4140	75-67	250-220	5.8-4.5	9-7
4320-4340	70-55	230-180	4.5-3.2	7-5
4815-4820	58-53	190-175	3.9-2.9	6-4.5
5046	75-67	250-220	5.8-4.5	9-7
5140-5160	70-60	230-200	4.2-3.2	6.5-5
50100-52100	50-37	170-120	3.9-2.6	6-4
6118-6150	68-45	225-150	4.8-2.6	7.5-4
8615-8645	70-53	230-175	4.5-3.2	7-5
8720-8740	68-53	225-175	4.5-3.2	7-5
9310	53-45	175-150	2.6-1.9	4-3
TOOL STEELS				
W1	67-55	220-180	3.9-3.2	6-5
S2, S5	45-33	150-110	2.6-1.9	4-3
O1, O2	65-55	210-180	3.9-2.6	6-4
A2	60-52	200-170	2.6-1.9	4-3
D2, D3	37-27	120-90	1.9-1.3	3-2
D7	27-18	90-60	1.3-0.6	2-1
H12, H13, H21	58-49	190-160	3.2-2.6	5-4
T1, T2	40-30	130-100	2.3-1.3	3.5-2
T6, T8	30-21	100-70	1.6-0.6	2.5-1

T15	23-15	75-50	1.3-0.6	2-1
M1	45-37	150-120	3.2-1.9	5-3
M2, M3	33-24	110-80	2.6-1.3	4-2
M4, M10, M15	27-18	90-60	1.6-0.6	2.5-1
L6	55-49	180-160	3.9-2.6	6-4
STAINLESS STEELS				
201,202, 302, 304	37-24	120-80	2.6-1.3	4-2
303,303F	40-27	130-90	3.2-1.3	5-2
308, 309, 310, 330	24-18	80-60	1.3-0.6	2-1
314, 316, 317	23-15	75-50	1.3-0.6	2-1
321,347	37-27	120-90	2.6-1.3	4-2
410, 420, 420F	43-30	140-100	2.6-1.3	4-2
416, 430F	55-43	180-140	4.5-3.2	7-5
430, 446	27-18	90-60	2.6-1.9	4-3
440A, B, C	33-21	110-70	2.6-1.3	4-2
44OF, 443	40-30	130-100	2.6-1.3	4-2
17-7 PH, 17-4 PH	27-15	90-50	2.6-1.3	4-2
NONFERROUS METALS				
COPPER ALLOYS				
170, BERYLLIUM COPPER	84-60	275-200	5.2-3.9	8-6
	68-53	225-175	3.9-2.6	6-4
	43-27	140-90	1.9-1.3	3-2
510 PHOSPHOR BRONZE 5% A	90-75	300-250	6.4-5.2	10-8
	53-38	175-125	3.2-1.9	5-3
614, ALUMINUM BRONZE D	106-90	350-300	9.0-6.4	14-10
	53-38	175-125	3.2-1.9	5-3
656, HIGH-SILICON BRONZE	100-384	325-275	9.7-7.7	15-12
	53-38	175-125	3.9-1.9	6-3
675, MANGANESE BRONZE A	100-584	325-275	9.7-7.7	15-12
	60-45	200-150	3.9-2.6	6-4
NICKEL ALLOYS				
INCONEL	30-18	100-60	1.9-1.3	3-2
INCONEL X-750	24-18	80-60	1.0-0.3	1.5-0.5
MONEL 400	30-18	100-60	1.9-0.6	3-1
MONEL R-405	45-23	150-75	2.6-1.3	4-2
MONEL K-500	24-18	80-60	1.3-0.3	2-0.5
MONEL 501	30-18	100-60	1.9-0.6	3-1
HASTELLOY A	37-23	120-75	1.9-1.0	3-1.5
HASTELLOY B	30-23	100-75	1.6-0.6	2.5-1
HASTELLOY C	27-18	90-60	1.0-0.45	1.5-0.7
TITANIUM ALLOYS				
TI; TI-1.5FE-2.5CR	27-18	90-60	0.6-0.2	1-0.3
TI-4AL-4MN; TI-6AL-4V	33-21	110-70	1.3-3.9	2-6
TI-2FE-2CR-2MO	27-18	90-60	1.0-0.3	1.5-0.5

(A) BASED ON THE USE OF A 25 MM (1 IN.) WIDE HIGH-SPEED STEEL BAND, REGULAR TOOTH FORM (EXCEPT HOOK TOOTH FORM FOR METAL THICKER THAN ABOUT 250 MM, OR 10 IN.), RAKER SET, TO CUT SCALE-FREE, SOLID BAR STOCK UP TO 460 MM (18 IN.) THICK; BASED ON THE USE OF A CUTTING FLUID, EXCEPT FOR D2, D3, AND D7 TOOL STEELS, WHICH ARE CUT DRY

Table 3 only applies when using bimetal blades. Carbon steel blades cannot be run at the band speeds shown because the heat generated at these speeds will cause them to soften and fail immediately.

When cutting pipes, tubes, or structural shapes, it is necessary to compute slower cutting rates than those given in Table 3. The cutting rates should be modified as follows:

PART SIZE		**PERCENTAGE OF SPEED SHOWN IN TABLE 3**
MM	**IN.**	
<4.8	$<\frac{3}{16}$	40
4.8-9.5	$\frac{3}{16}-\frac{3}{8}$	50
9.5-16	$\frac{3}{8}-\frac{5}{8}$	60
>16	$>\frac{5}{8}$	70

Proper break-in is important to ensure optimum blade performance. When using a new blade, cutting at the full rate will cause fracturing of the very sharp tooth edges. Breaking in a new blade removes the ultrasharp edges and allows the blade to retain its cutting ability longer. A new blade can be broken in by running at the recommended band speed, but reducing the feed pressure for the initial cuts. When cutting work-hardening materials (for example, nickel alloys), it is important to use enough feed pressure to generate a chip and to prevent work hardening of the cut surface.

Knowledge of chip formation is helpful in finding the optimum speeds and cutting rates to use when band sawing. When a blade cuts, each tooth tip penetrates the workpiece and shears off a chip of material. Parameters such as band speed, feed, lubrication, and blade tip design all affect chip formation.

Visual examination of the chips can be used to determine proper speeds and cutting rates (Table 4). In general, the optimum conditions are present when the chips are curled and silver in color, indicating they were properly formed without excessive heat generation.

TABLE 4 VISUAL EXAMINATION OF CHIPS TO DETERMINE ADJUSTMENTS IN BAND SAW PARAMETERS THAT ARE REQUIRED TO OPTIMIZE MACHINABILITY

Source: "Operator's Guide and Application Guide," DoALL Co., 1990

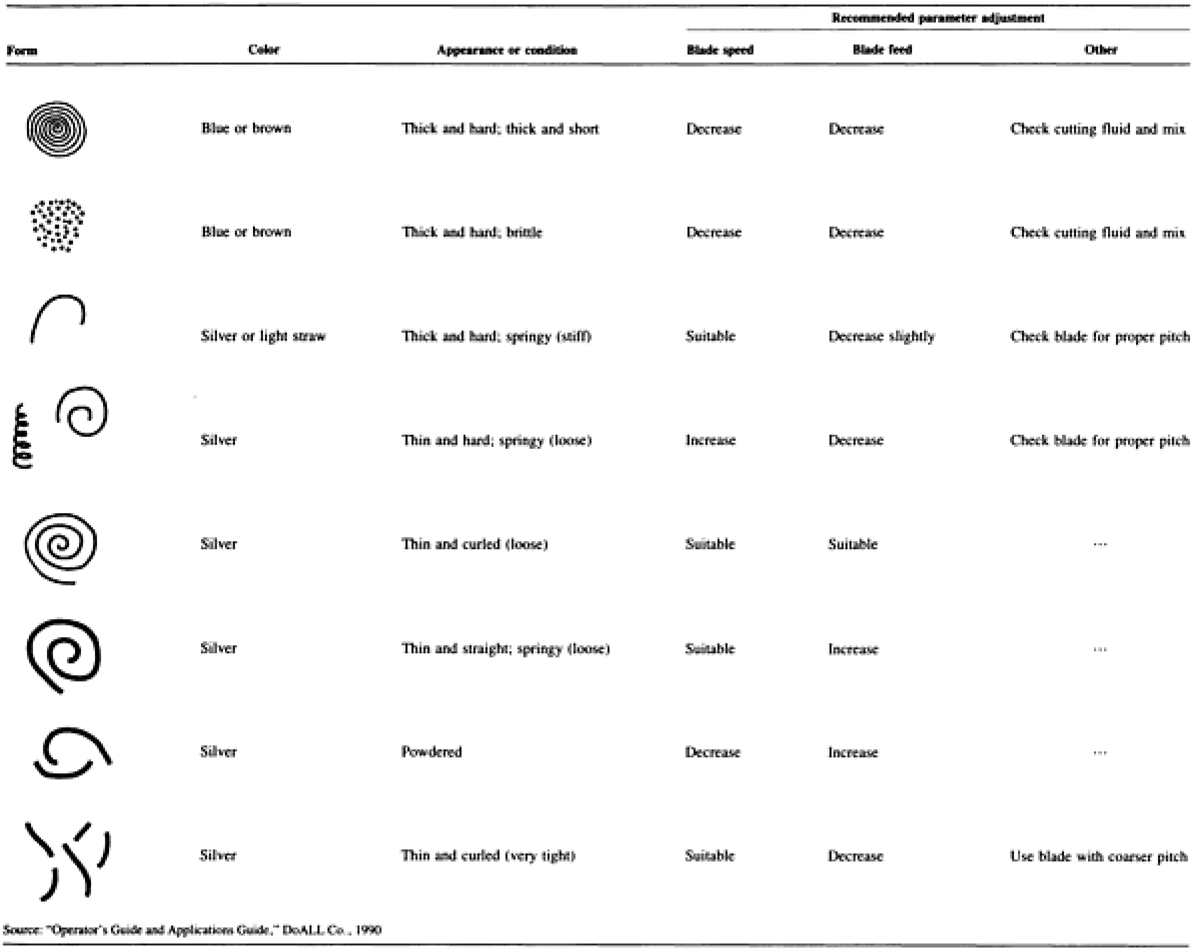

Form	Color	Appearance or condition	Recommended parameter adjustment		
			Blade speed	Blade feed	Other
	Blue or brown	Thick and hard; thick and short	Decrease	Decrease	Check cutting fluid and mix
	Blue or brown	Thick and hard; brittle	Decrease	Decrease	Check cutting fluid and mix
	Silver or light straw	Thick and hard; springy (stiff)	Suitable	Decrease slightly	Check blade for proper pitch
	Silver	Thin and hard; springy (loose)	Increase	Decrease	Check blade for proper pitch
	Silver	Thin and curled (loose)	Suitable	Suitable	...
	Silver	Thin and straight; springy (loose)	Suitable	Increase	...
	Silver	Powdered	Decrease	Increase	...
	Silver	Thin and curled (very tight)	Suitable	Decrease	Use blade with coarser pitch

Source: "Operator's Guide and Applications Guide," DoALL Co., 1990

Safe Practices*

American Welding Society

Introduction

THIS ARTICLE covers the basic elements of safety general to all welding, cutting, and related processes. It includes safety procedures common to a variety of applications. However, it does not cover all safety aspects of every welding process; especially not those involving sophisticated technology. For this reason, those chapters in this Volume devoted to specific processes should also be referenced for additional important safety information.

Note

* THIS ARTICLE IS REPRINTED BY PERMISSION OF THE AMERICAN WELDING SOCIETY FROM THE *WELDING HANDBOOK,* 8TH EDITION, VOLUME 1, AMERICAN WELDING SOCIETY, 1987. IT WAS PREPARED BY THE AWS WELDING HANDBOOK SAFE PRACTICES CHAPTER COMMITTEE CONSISTING OF G.R. SPIES, *CHAIRMAN,* THE B.O.C. GROUP, LNC.; G.C. BARNES, ALLOY RODS COMPANY; K.L. BROWN, THE LINCOLN ELECTRIC COMPANY; W. BEISNER, HOBART BROTHERS COMPANY; O.J. FISHER, BABCOCK AND WILCOX COMPANY; W.S. HOWES, NEMA; C. PHILP, HANDY AND HARMAN COMPANY; UNDER THE DIRECTION OF AWS WELDING HANDBOOK COMMITTEE MEMBER A.F. MANZ, A.F. MANZ ASSOCIATES, AND AWS WELDING HANDBOOK COMMITTEE CHAIRMEN J.H. HANNAHS, MIDMARK CORPORATION, AND M.J. TOMSIC, PLASTRONICS INCORPORATED (AFFILIATIONS CIRCA 1987).

Safe Practices*

American Welding Society

General Welding Safety

Safety is an important consideration in all welding, cutting, and related work. No activity is satisfactorily completed if someone is injured. The hazards that may be encountered and the practices that will minimize personal injury and property damage are discussed here.

Management Support

The most important component of an effective safety and health program is management support and direction. Management must clearly state objectives and demonstrate its commitment to safety and health by consistent execution of safe practices.

Management must designate approved areas where welding and cutting operations may be carried on safely. When these operations must be done in other than designated areas, management must ensure that proper procedures to protect personnel and property are established and followed.

Management must be certain that only approved welding, cutting, and allied equipment is used. Such equipment includes torches, regulators, welding machines, electrode holders, and personal protective devices. Adequate supervision must be provided to ensure that all equipment is properly used and maintained.

Training

Thorough and effective training is a key aspect of a safety program. Adequate training is mandated under provisions of the U.S. Occupational Safety and Health Administration (OSHA), especially those of the Hazard Communication Standard (29 CFR 1910.1200). Welders and other equipment operators perform most safely when they are properly trained in the subject. (The term *welder* as used in this article is intended to include all welding and cutting personnel, brazers, and solderers.) Proper training includes instruction in the safe use of equipment and processes, and the safety rules that must be followed. Personnel need to know and understand the rules and the consequences of disobeying them. For example, welders must be trained to position themselves while welding or cutting so that their heads are not in the gases or fume plume. (Fume plume is the smokelike cloud containing minute solid particles arising directly from the area of melting metal. In distinction to a gas, fumes are metallic vapors that have condensed to solid and are often associated with a chemical reaction, such as oxidation.)

Before work begins, users must always read and understand the manufacturer's instructions on safe practices for the materials and equipment, and also the material safety data sheets.

Certain AWS specifications call for precautionary labels on consumables and equipment. These labels concerning the safe use of the products should be read and followed. A typical label is illustrated in Fig. 1.

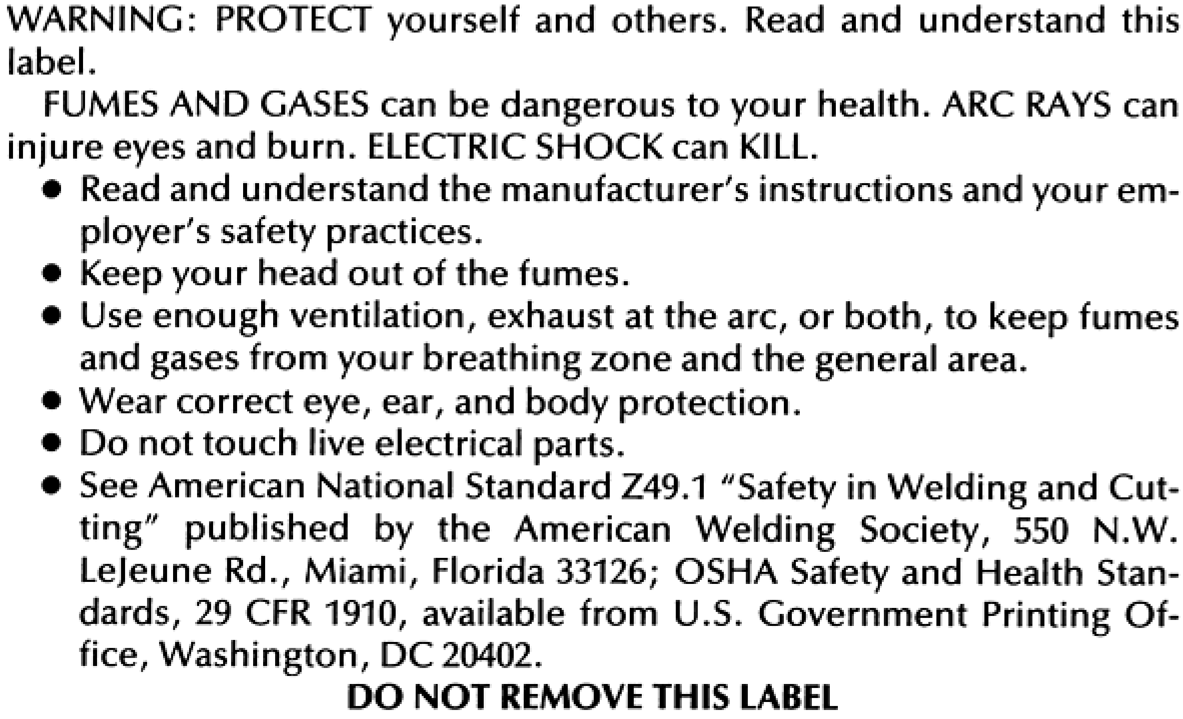

WARNING: PROTECT yourself and others. Read and understand this label.

FUMES AND GASES can be dangerous to your health. ARC RAYS can injure eyes and burn. ELECTRIC SHOCK can KILL.

- Read and understand the manufacturer's instructions and your employer's safety practices.
- Keep your head out of the fumes.
- Use enough ventilation, exhaust at the arc, or both, to keep fumes and gases from your breathing zone and the general area.
- Wear correct eye, ear, and body protection.
- Do not touch live electrical parts.
- See American National Standard Z49.1 "Safety in Welding and Cutting" published by the American Welding Society, 550 N.W. LeJeune Rd., Miami, Florida 33126; OSHA Safety and Health Standards, 29 CFR 1910, available from U.S. Government Printing Office, Washington, DC 20402.

DO NOT REMOVE THIS LABEL

FIG. 1 MINIMUM WARNING FOR ARC WELDING PROCESSES AND EQUIPMENT

Manufacturers of welding consumables must provide on request a Material Safety Data Sheet (MSDS) that identifies those materials present in their products that have hazardous physical or health properties. The MSDS provides the OSHA permissible exposure limit, called the Threshold Limit Value (TLV), and any other exposure limit used or recommended by the manufacturer.

Employers that use consumables must make the applicable MSDS readily available to their employees, as well as train them to read and understand the contents. The MSDS contain important information about the ingredients contained in welding electrodes, rods, and fluxes, the composition of fumes that may be generated in their use, and means to be followed to protect the welder and others from hazards which might be involved.

Under OSHA Hazard Communication Standard 29 CFR 1910.1200, employers are responsible for the training of employees with respect to hazardous materials used in their workplace. Many welding consumables are included in the definition of hazardous materials according to this standard. Welding employers must comply with the communication and training requirements of this standard.

The proper use and maintenance of the equipment must also be taught. For example, defective or worn electrical insulation cannot be tolerated in arc welding or cutting, nor can defective or worn hoses be used in oxyfuel gas welding and cutting, brazing, or soldering. Proper training in equipment operation is fundamental to safe operation.

Persons must be trained to recognize safety hazards. If they are to work in an unfamiliar situation or environment, they must be thoroughly briefed on the potential hazards involved. For example, consider a person who must work in a confined space. If the ventilation is poor and an air-supplied helmet is required, the need and instructions for its proper use must be thoroughly explained to the employee. The consequences of improperly using the equipment must be covered. When employees believe that the safety precautions for a given task are not adequate, or not understood, they should question their supervisors before proceeding.

General Housekeeping

Good housekeeping is essential to avoid injuries. A welder's vision is often restricted by necessary eye protection. Persons passing a welding station must shield their eyes from the flame or arc radiation. The limited vision of the welder and passers-by make them vulnerable to tripping over objects on the floor. Therefore, welders and supervisors must always make sure that the area is clear of tripping hazards. Management must lay out the production area so that gas hoses, cables, mechanical assemblies, and other equipment do not cross walkways or interfere with routine tasks.

When work is above ground or floor level, safety rails or lines must be provided to prevent falls as a result of restricted vision from eye protection devices. Safety lines and harnesses can be helpful to restrict workers to safe areas, and to catch them in case of a fall.

Unexpected events, such as fire and explosions, do occur in industrial environments. All escape routes must be identified and kept clear so that orderly, rapid, and safe evacuation of an area can take place. Storage of goods and equipment in evacuation routes should be avoided. If an evacuation route must be temporarily blocked, employees who would normally use that route must be trained to use an alternate route.

Protection in the General Area

Equipment, machines, cables, hoses, and other apparatus should always be placed so that they do not present a hazard to personnel in passageways, on ladders, or on stairways. Warning signs should be posted to designate welding areas and to specify that eye protection must be worn.

Protective Screens. Persons in areas adjacent to welding and cutting must be protected from radiant energy and hot spatter by (1) flame-resistant screens or shields, or (2) suitable eye and face protection and protective clothing. Appropriate radiation-protective, semitransparent materials are permissible.

Where operations permit, work stations should be separated by noncombustible screens or shields (Fig. 2). Booths and screens should permit circulation of air at floor level as well as above the screen.

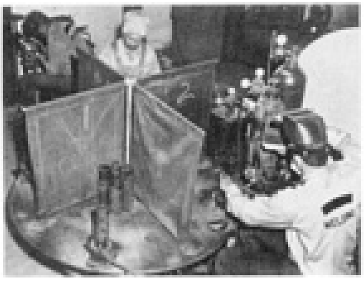

FIG. 2 PROTECTIVE SCREENS BETWEEN WORK STATIONS

Wall Reflectivity. Where arc welding or cutting is regularly carried on adjacent to painted walls, the walls should be painted with a finish having a low reflectivity of ultraviolet radiation. Finishes formulated with certain pigments, such as titanium dioxide or zinc oxide, have low reflectivity to ultraviolet radiation. Color pigments may be added if they do not increase reflectivity. Pigments based on powdered or flaked metals are not recommended because they reflect ultraviolet radiation. Additional information is available in *Ultraviolet Reflectance of Paint,* published by the American Welding Society.

Public Demonstrations

Persons putting on public demonstrations involving observations of arc or oxyfuel gas welding or cutting processes are responsible for the safety of observers and the general public. Observers are not likely to have the necessary protective equipment to let them observe demonstrations safely. For exhibits involving observation of arc or oxyfuel gas welding and cutting processes, appropriate eye protection for both observers and passers-by is mandatory.

Fume exposure must be controlled by appropriate ventilation. Electric cables and hoses must be routed to avoid audience exposure to possible electric shock or tripping hazards. Protection must be provided against fires from fuels, combustibles, and overheated apparatus and wiring.

Fire extinguishers must be on hand in case of fire. Combustible materials must be removed from the area or shielded from flames, sparks, and molten metal.

Safety precautions at public events should be passive types, that is, they should not require the audience to take action to protect itself. For example, a protective, moveable transparent screen allows an audience to observe a welding operation with the screen in place. After welding is completed, the screen can be moved to allow the audience to observe the completed weld. Additional information is given in ANSI/AWS Z49.1, *Safety in Welding and Cutting.*

Fire

In most welding, cutting, and allied processes, a high-temperature heat source is present. Open flames, electric arcs, hot metal, sparks, and spatter are ready sources of ignition. Many fires are started by sparks, which can travel horizontally up to 10 m (35 ft) from their source and fall much greater distances. Sparks can pass through or lodge in cracks, holes, and other small openings in floors and walls.

The risk of fire is increased by combustibles in the work area, or by welding or cutting too close to combustibles that have not been shielded. Materials most commonly ignited are combustible floors, roofs, partitions, and building contents including trash, wood, paper, textiles, plastics, chemicals, and flammable liquids and gases. Outdoors, the most common combustibles are dry grass and brush.

The best protection against fire is to perform welding and cutting in specially designated areas or enclosures of noncombustible construction that are kept free of combustibles. Combustibles should always be removed from the work area or shielded from the operation.

Fuel for engine-driven equipment should be stored and used with care. Equipment manufacturer instructions should be followed because fuels and their vapors are combustible and can be explosive under some conditions.

Fuel gases, such as acetylene or propane, are other common flammables often found in cutting and welding areas. Special attention should be given to fuel gas cylinders, hoses, and apparatus to prevent gas leakage.

Combustibles that cannot be removed from the area should be covered with tight fitting, flame-resistant material. These include combustible walls and ceilings. Floors should be free of combustible materials for a radius of 10 m (35 ft) around the work area. All doorways, windows, cracks, and other openings should be covered with a flame-resistant material or, if possible, the work area should be enclosed with portable flame-resistant screens.

If welding or cutting is to be performed on or adjacent to a metal wall, ceiling, or partition, combustibles on the other side must be moved to a safe location. If this cannot be done, a fire watcher should be stationed where the combustibles are located. Heat from welding can be conducted through metal partitions and ignite combustibles on the opposite side. A thorough examination for evidence of fire should be made before leaving the work area. Fire inspection should be continued for at least 30 min after the operation is completed.

Welding or cutting should not be performed on material having a combustible coating or internal structure, as in walls or ceilings. Hot scrap or slag must not be placed in containers holding combustible materials. Suitable fire extinguishers should always be available nearby.

Welding, brazing, or cutting should not be performed on combustible floors or platforms that may readily be ignited by heat from the operation. Welders must be alert for traveling vapors of flammable liquids. Vapors are often heavier than air and can travel along floors and in depressions for distances of several hundred feet from where the flammable liquid is stored. Light vapors can travel along ceilings to adjacent rooms.

Hot Work Permit System

When welding, cutting, or similar hot working operations are to be performed in areas not normally assigned for such operations, a hot work permit system should be used. The purpose of the hot work permit system is to alert area supervisors to an extraordinary danger of fire that will exist at a particular time. The permit system should include a check list of safety precautions that includes an inspection for fire extinguishers, establishment of fire watches if necessary, search for flammable materials, and safety instructions for personnel in the area who are not involved in the hot work.

Explosion

Flammable gases, vapors, and dusts, when mixed with air or oxygen in certain proportions, present danger of explosion as well as fire. To prevent explosions, operators must avoid all sources of ignition. Welding, brazing, soldering, cutting, or operating equipment that can produce heat or sparks must not be done in atmospheres containing flammable gases, vapors, or dusts. Such flammables must be kept in leak-fight containers or be well removed from the work area. Heat or sparks may cause otherwise low-volatile materials to produce flammable vapors.

Hollow containers must be vented before applying heat. Heat must not be applied to a container that has held an unknown material, a combustible substance, or a substance that may form flammable vapors on heating. Additional information is given in AWS F4.1, *Recommended Safe Practices for the Preparation for Welding and Cutting of Containers and Piping That Have Held Hazardous Substances*, American Welding Society, latest edition. The container must first be thoroughly cleaned or filled with an inert gas. Heat should never be applied to a workpiece that is covered by an unknown substance, nor to a substance that may form flammable or toxic vapors on heating.

Adequate eye and body protection must be worn when operations involve a risk of explosion.

Burns

Burns of the eye or body are serious hazards of welding, brazing, soldering, and cutting. Eye, face, and body protection for the operator and others in the work area are required to prevent burns from ultraviolet and infrared radiation, sparks, and spatter.

Eye and Face Protection

Arc Welding and Cutting. Welding helmets or handshields containing appropriate filter lenses and cover plates must be used by welders and welding operators and nearby personnel when viewing an arc. Suggested shade numbers of filter plates for various welding, brazing, soldering, and thermal cutting operations are given in Table 1.

TABLE 1 SUGGESTED VIEWING FILTER PLATES

OPERATION	PLATE THICKNESS		WELDING CURRENT, A	LOWEST SHADE NO.	COMFORT SHADE NO.(A)
	mm	in.			
SHIELDED METAL ARC WELDING	. . .	. . .	<60	7	. . .
			60-160	7	10
			160-250	10	12
			250-550	11	14
GAS-METAL ARC AND FLUX-CORED ARC WELDING	. . .	. . .	<60	7	. . .
			60-160	10	11
			160-250	10	12
			250-500	10	14
GAS-TUNGSTEN ARC WELDING	. . .	. . .	<50	8	10
			50-150	8	12
			150-500	10	14
PLASMA ARC WELDING	. . .	. . .	<20	6	6-8
			20-100	8	10
			100-400	10	12
			400-800	11	14
OXYFUEL GAS WELDING (STEEL)(B)	3.2	$<\frac{1}{8}$	. . .	. . .	4, 5
	3.2-12.7	$\frac{1}{8}-\frac{1}{2}$	. . .	. . .	5, 6

	12.7	$> \frac{1}{2}$	. . .	. . .	6, 8
PLASMA ARC CUTTING(C)	. . .	. . .	<300	8	9
			300-400	9	12
			400-800	10	14
AIR-CARBON ARC CUTTING	. . .	. . .	<500	10	12
			500-1000	11	14
OXYFUEL GAS CUTTING (STEEL)(B)	25	<1	. . .	. . .	3, 4
	25-100	1-6	. . .	. . .	4, 5
	150	>6	. . .	. . .	5, 6
TORCH BRAZING	. . .	. . .	. . .	. . .	3, 4
TORCH SOLDERING	. . .	. . .	. . .	. . .	2

(A) TO SELECT THE BEST SHADE FOR THE APPLICATION, FIRST SELECT A DARK SHADE. IF IT IS DIFFICULT TO SEE THE OPERATION PROPERLY, SELECT SUCCESSFULLY LIGHTER SHADES UNTIL THE OPERATION IS SUFFICIENTLY VISIBLE FOR GOOD CONTROL. HOWEVER, DO NOT GO BELOW THE LOWEST RECOMMENDED NUMBER, WHERE GIVEN.

(B) WITH OXYFUEL GAS WELDING OR CUTTING, THE FLAME EMITS STRONG YELLOW LIGHT. A FILTER PLATE THAT ABSORBS YELLOW OR SODIUM WAVE LENGTHS OF VISIBLE LIGHT SHOULD BE USED FOR GOOD VISIBILITY.

(C) THE SUGGESTED FILTERS ARE FOR APPLICATIONS WHERE THE ARC IS CLEARLY VISIBLE. LIGHTER SHADES MAY BE USED WHERE THE ARC IS HIDDEN BY THE WORK OR SUBMERGED IN WATER.

Safety spectacles, goggles, or other suitable eye protection must also be worn during welding and cutting operations. Such devices must have full conforming side shields when there is danger of exposure to injurious rays or to flying particles from grinding or chipping operations. Spectacles or goggles may have clear or colored lenses, depending on the intensity of the radiation that may come from adjacent welding or cutting operations when the welding helmet is raised or removed. Number 2 filter lenses are recommended for general purpose protection.

Oxyfuel Gas Welding and Cutting, Submerged Arc Welding. Safety goggles with filter lenses (Table 1) and full conforming side shields must be worn while performing oxyfuel gas welding and cutting. During submerged arc welding, the arc is covered by flux and not readily visible; hence, an arc welding helmet is not needed. However, because the arc occasionally flashes through the flux burden, the operator should wear tinted safety glasses.

Torch Brazing and Soldering. Safety spectacles with or without side shields and with appropriate filter lenses are recommended for torch brazing and soldering. As with oxyfuel gas welding and cutting, a bright yellow flame may be visible during torch brazing. A filter similar to that used with those processes should be used for torch brazing.

Resistance, Induction, Salt-Bath, Dip, Infrared Welding, and Brazing. Operators and helpers engaged in these processes must wear safety spectacles, goggles, and a face shield to protect their eyes and face from spatter. Filter lenses are not necessary but may be used for comfort.

Standards for welding helmets, handshields, face shields, goggles, and spectacles are given in ANSI Z87.1, *Practices for Occupational and Educational Eye and Face Protection,* American National Standards Institute (latest edition).

Protective Clothing

Sturdy shoes or boots, and heavy clothing similar to that in Fig. 3 should be worn to protect the whole body from flying sparks, spatter, and radiation burns. Woolen clothing is preferable to cotton because it is not so readily ignited. Cotton clothing, if used, should be chemically treated to reduce its combustibility. Clothing treated with nondurable flame retardants must be retreated after each washing or cleaning. Clothing or shoes of synthetic or plastic materials which can melt and cause severe burns should not be worn. Outer clothing should be kept free of oil and grease, especially in an oxygen-rich atmosphere.

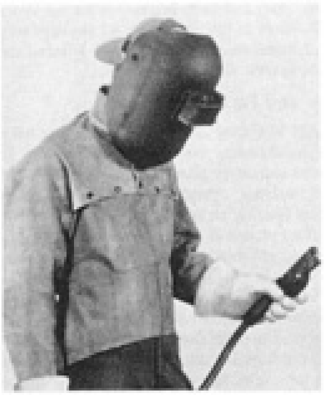

FIG. 3 TYPICAL PROTECTIVE CLOTHING FOR ARC WELDING

Cuffless pants and covered pockets are recommended to avoid spatter or spark entrapment. Pockets should be emptied of flammable or readily ignitable materials before welding because they may be ignited by sparks or weld spatter and result in severe burns. Pants should be worn outside of shoes. Protection of the hair with a cap is recommended, especially if a hair-piece is worn. Flammable hair preparations should not be used.

Durable gloves of leather or other suitable material should always be worn. Gloves not only protect the hands from burns and abrasion, but also provide insulation from electrical shock. A variety of special protective clothing is also available for welders. Aprons, leggings, suits, capes, sleeves, and caps, all of durable materials, should be worn when welding overhead or when special circumstances warrant additional protection of the body.

Sparks or hot spatter in the ears can be particularly painful and serious. Properly fitted, flame-resistant ear plugs should be worn whenever operations pose such risks.

Noise

Excessive noise, particularly continuous noise at high levels, can damage hearing. It may cause either temporary or permanent hearing loss. U.S. Department of Labor Occupational Safety and Health Administration regulations describe allowable noise exposure levels. Requirements of these regulations can be found in General Industry Standards, 29 CFR 1910.95.

In welding, cutting, and allied operations, noise may be generated by the process or the equipment, or both. Processes that tend to have high noise levels are air carbon arc and plasma arc cutting. Engine-driven generators sometimes emit a high noise level, as do some high-frequency and induction welding power sources.

Additional information is presented in *Arc Welding and Cutting Noise,* American Welding Society, 1979.

Machinery Guards

Welders and other workers must be protected from injury by machinery and equipment that they are operating or by other machinery operating in the work area. Moving components and drive belts must be covered by guards to prevent physical contact.

Because welding helmets and dark filter lenses restrict the visibility of welders, they may be even more susceptible than ordinary welders to injury from unseen, unguarded machinery. Therefore, special attention is required to this hazard.

When repairing machinery by welding or brazing, the power to the machinery must be disconnected, locked out, and tagged to prevent inadvertent operation and injury. Welders assigned to work on equipment with safety devices removed should fully understand the hazards involved and the steps to be taken to avoid injury. When the work is completed, the safety devices must be replaced. Rotating and automatic welding machines, fixtures, and welding robots must be equipped with appropriate guards or sensing devices to prevent operation when someone is in the danger area.

Pinch points on welding and other mechanical equipment can also result in serious injury. Examples of such equipment are resistance welding machines, robots, automatic arc welding machines, jigs, and fixtures. To avoid injury with such equipment, the machine should be equipped so that both of the operator's hands must be at safe locations when the machine is actuated. Otherwise, the pinch points must be suitably guarded mechanically. Metalworking equipment should not be located where a welder could accidently fall into or against it while welding. During maintenance of the equipment, pinch points should be blocked to prevent them from closing in case of equipment failure. In very hazardous situations, an observer should be stationed to prevent someone from turning the power on until the repair is completed.

Safe Practices*

American Welding Society

Fumes and Gases

Welders, welding operators, and other persons in the work area must be protected from overexposure to fumes and gases produced during welding, brazing, soldering, and cutting. Overexposure is exposure that is hazardous to health and exceeds the permissible limits specified by a government agency, such as the U.S. Department of Labor, Occupational Safety and Health Administration Regulations 29 CFR 1910.1000, or other recognized authority, such as the American Conference of Governmental Industrial Hygienists in its publications *Threshold Limit Values for Chemical Substances* and *Physical Agents in the Workroom Environment.* Persons with special health problems may have unusual sensitivity that requires even more stringent protection.

Fumes and gases are usually a greater concern in arc welding than in oxyfuel gas welding, cutting, or brazing because a welding arc may generate a larger volume of fume and gas, and greater varieties of materials are usually involved.

Protection from excess exposure is usually accomplished by ventilation. Where exposure would exceed permissible limits with available ventilation, respiratory protection must be used. Protection must be provided not only for the welding and cutting personnel but also for other persons in the area.

Arc Welding

Nature and Sources. Fumes and gases from arc welding and cutting cannot be classified simply. Their composition and quantity depend upon the base metal composition; the process and consumables used; coatings on the work, such as paint, galvanizing, or plating; contaminants in the atmosphere, such as halogenated hydrocarbon vapors from cleaning and degreasing activities; and other factors.

In welding and cutting, the composition of the fume usually differs from the composition of the electrode or consumables. Reasonably expected fume constituents from normal operations include products of volatilization, reaction, or oxidation of consumables, base metals, coatings, and atmospheric contaminants. Reasonably expected gaseous products include carbon monoxide, carbon dioxide, fluorides, nitrogen oxides, and ozone.

The quantity and chemical composition of air contaminants change substantially with the process and with a wide range of variables inherent in each process. During arc welding, the arc energy and temperature depend on the process as well as the welding variables. Therefore, fumes and gases are generated in varying degrees in different welding operations.

Welding fume is a product of vaporization, oxidation, and condensation of components in the consumable and, to some degree, the base metal. The electrode, rather than the base metal, is usually the major source of fume. However, significant fume constituents can originate from the base metal if it contains alloying elements or a coating that is volatile at elevated temperatures.

Various gases are generated during welding. Some are a product of the decomposition of fluxes and electrode coatings. Others are formed by the action of arc heat or ultraviolet radiation emitted by the arc on atmospheric constituents and contaminants. Potentially hazardous gases include carbon monoxide, oxides of nitrogen, ozone, and phosgene or other decomposition products of chlorinated hydrocarbons, such as phosgene.

Helium and argon, although chemically inert and nontoxic, are simple asphyxiants and could dilute the atmospheric oxygen concentration to potentially harmful low levels. Carbon dioxide (CO_2) and nitrogen can also cause asphyxiation.

Ozone may be generated by ultraviolet radiation from welding arcs. This is particularly true with gas shielded arcs, especially when argon is used. Photochemical reactions between ultraviolet radiation and chlorinated hydrocarbons result in the production of phosgene and other decomposition products.

The arc heat is responsible for the formation of oxides of nitrogen from atmospheric nitrogen. Hence, nitrogen oxides may be produced by a welding arc or other high-temperature heat sources. Thermal decomposition of carbon dioxide and inorganic carbonate compounds by an arc results in the formation of carbon monoxide. Levels can be significant when CO_2 is used as the shielding gas.

Reliable estimates of fume and gas composition cannot be made without considering the nature of the welding process and system being examined. For example, aluminum and titanium are normally welded in an atmosphere of argon or helium, or mixtures of the two gases. The arc creates relatively little fume but may emit an intense radiation that can produce ozone. Inert gas shielded arc welding of steels also creates a relatively low fume level.

Arc welding of steel in oxidizing environments, however, generates considerable fume and can produce carbon monoxide and oxides of nitrogen. The fumes generally consist of discrete particles of amorphous slags containing iron, manganese, silicon, and other metallic constituents, depending on the alloy system involved. Chromium and nickel compounds are found in fumes when stainless steels are arc welded.

Some covered and flux-cored electrodes are formulated with fluorides. The fumes associated with those electrodes can contain significantly more fluorides than oxides.

Factors Affecting Generation Rates. The rate of generation of fumes and gases during arc welding of steel depends on numerous variables. Among these are:

- WELDING CURRENT
- ARC VOLTAGE (ARC LENGTH)
- TYPE OF METAL TRANSFER OR WELDING PROCESS
- SHIELDING GAS

These variables are interdependent and can have a substantial effect on total fume generation.

Welding Current. In general, fume generation rate increases with increased welding current. The increase, however, varies with the process and electrode type. Certain covered, flux-cored, and solid-wire electrodes exhibit a nonproportional increase in fume generation rate with increasing current.

Studies have shown that fume generation rates with covered electrodes are proportional to the welding current raised to a power. The exponent is 2.24 for E6010 electrodes and 1.54 for E7018 electrodes. Similar trends were reported in other studies. Additional information is available in *Fumes and Gases in the Welding Environment,* American Welding Society, 1979.

Flux-cored and solid electrode fume generation rates are more complexly related to welding current. Welding current levels affect the type of metal droplet transfer. As a result, the fume generation rate can decrease with increasing current until some minimum is reached. Then, it will increase in a relatively proportional fashion.

An increase in current can increase ultraviolet radiation from the arc. Therefore, the generation of gases formed photochemically by this radiation, such as ozone, can be expected to increase as welding current is increased. Measurements of ozone concentration during gas-metal arc and gas-tungsten arc welding have shown such behavior.

Arc voltage and arc length are directly related. For a given arc length, there is a corresponding arc voltage, mostly dependent upon the type of electrode, welding process, and power supply. In general, increasing arc voltage (arc length) increases the fume generation rate for all open arc welding processes. The levels differ somewhat for each process or electrode type.

Type of Metal Transfer. When steel is joined with gas-metal arc welding using a solid-wire electrode, the mode of metal transfer depends upon the current and voltage. At low welding current and voltage, short-circuiting transfer takes place; that is, droplets are deposited during short circuits between the electrode and molten weld pool. As the current or voltage is increased, metal transfer changes to globular type, where large globules of metal are projected across the arc into the weld pool. At high currents, transfer changes to a spray mode, where fine metal droplets are propelled rapidly across the arc. Fume generation rate appears to follow a transition also. It is relatively high during short-circuiting transfer because of arc turbulence. As the transition current is approached in an argon-rich shielding gas, the fume rate decreases and then increases again as spray transfer is achieved. In the spray region, the rate of fume generation is proportional to welding current.

For other welding processes, the type of metal transfer does not vary substantially with current or voltage. In these cases, fume generation follows the relationship for welding current changes.

Shielding Gas. When gas-metal arc welding or flux-cored arc welding with certain electrodes, shielding gas must be used. The type of shielding gas affects both the composition of the fume and its rate of generation. It also affects the kind of gases found in the welding environment. For example, the fume generation rate is higher with CO_2 shielding than with argon-rich shielding. The rate of fume formation with argon-oxygen or argon-CO_2 mixtures increases with the oxidizing potential of the mixture.

For welding processes where inert gas shielding is used, such as gas-tungsten arc or plasma arc welding, the fume generation rate varies with the type of gas or gas mixture. For example, there can be more fume with helium than with argon shielding.

By-product gases also vary with shielding gas composition. The rate of formation of ozone depends upon the wave lengths and intensity of the ultraviolet rays generated in the arc; ozone is more commonly found with argon-rich gases than with carbon dioxide. Oxides of nitrogen are present in the vicinity of any open arc process. Carbon monoxide is commonly found around CO_2 shielded arcs.

Welding Process. Studies of the relative fume generation rates of consumable electrode processes for welding on mild steel have shown definite trends. Considering the ratio of the weight of fumes generated per weight of metal deposited, covered electrodes and self-shielded fluxcored electrodes produce the most fume. Gas-shielded flux-cored electrodes produce less fume, and solid-wire electrodes produce an even smaller amount. The submerged arc welding process consistently produces the least fumes because the fume is captured in the flux and slag cover.

Consumables. With a specific process, the fume rate depends on the composition of the consumables. Some components of covered and flux-cored electrodes are designed to decompose and form protective gases during welding. Hence, they will generate relatively high fume levels.

Because the constituents of many covered and flux-cored electrodes are proprietary, the fume generation rates of electrodes of the same AWS classification produced by different manufacturers can differ substantially. The only reliable method for comparing filler metals is actual product testing to determine specific fume generation characteristics.

Oxyfuel Gas Welding and Cutting

The temperatures encountered in oxyfuel gas welding and cutting are lower than those found in electric arc processes. Consequently, the quantity of fumes generated is normally lower. The gases formed are reaction products of fuel gas combustion and of chemical reactions between the gases and the other materials present. The fumes generated are the reaction products of the base metals, coatings, filler metals, fluxes, and the gases being used. In oxyfuel gas cutting of steel, the fumes produced are largely oxides of iron.

Fume constituents of greater hazard may be expected when coatings such as galvanizing, paint primers, or cadmium plating are present. Gases of greatest concern include oxides of nitrogen, carbon monoxide, and carbon dioxide. Oxides of nitrogen may be present in especially large amounts when oxyfuel gas cutting stainless steels using either the chemical flux or the iron powder process.

Exposure Factors

Position of the Head. The single most important factor influencing exposure to fume is the position of the welder's head with respect to the fume plume. When the head is in such a position that the fume envelops the face or helmet, exposure levels can be very high. Therefore, welders must be trained to keep their heads to one side of the fume plume. In some cases, the work can be positioned so the fume plume rises to one side.

Type of Ventilation. Ventilation has a significant influence on the amount of fumes in the work area and hence the welder's exposure. Ventilation may be local, where the fumes are extracted near the point of welding, or general, where the shop air is changed or filtered. The appropriate type will depend on the welding process, the material being welded, and other shop conditions. Adequate ventilation is necessary to keep the welder's exposure to fumes and gases within safe limits.

Shop Size. The size of the welding or cutting enclosure is important. It affects the background fume level. Fume exposure inside a tank, pressure vessel, or other confined space will certainly be higher than in a high-bay fabrication area.

The background fume level depends on the number and type of welding stations and the duty cycle for each station.

Design of Welding Helmet. The extent to which the helmet curves under the chin towards the chest affects the amount of fume exposure. Close-fitting helmets can be effective in reducing exposure.

Base Metal and Surface Condition. The type of base metal being welded influences both the constituents and the amount of fume generated. Surface contaminants or coatings may contribute significantly to the hazard potential of the fume. Paints containing lead and platings containing cadmium generate dangerous fumes during welding and cutting. Galvanized material evolves zinc fume.

Ventilation

The bulk of fume generated during welding and cutting consists of small particles that remain suspended in the atmosphere for a considerable time. As a result, fume concentration in a closed area can build up over time, as can the concentration of any gases evolved or used in the process. The particles eventually settle on the walls and floor, but the settling rate is low compared to the generation rate of the welding or cutting processes. Therefore, fume concentration must be controlled by ventilation.

Adequate ventilation is the key to control of fumes and gases in the welding environment. Natural, mechanical, or respirator ventilation must be provided for all welding, cutting, brazing, and related operations. The ventilation must ensure that concentrations of hazardous airborne contaminants are maintained below recommended levels. These levels must be no higher than the allowable levels specified by the U.S. Occupational Safety and Health Administration or other applicable authorities.

Many ventilation methods are available. They range from natural convection to localized devices, such as air-ventilated welding helmets. Examples of ventilation include:

- NATURAL VENTILATION
- GENERAL AREA MECHANICAL VENTILATION
- OVERHEAD EXHAUST HOODS
- PORTABLE LOCAL EXHAUST DEVICES
- DOWNDRAFT TABLES
- CROSSDRAFT TABLES
- EXTRACTORS BUILT INTO THE WELDING EQUIPMENT
- AIR-VENTILATED HELMETS

General Ventilation. In most cases, general ventilation is more effective in protecting personnel in adjacent areas than in protecting the welders themselves. Such ventilation may occur naturally outdoors, or when the shop doors and windows are open. It is acceptable when precautions are taken by the welder to keep his or her breathing zone away from the fume plume and when sampling of the atmosphere shows that concentrations of contaminants do not exceed permissible limits. Natural ventilation often meets these criteria when all of the following conditions are present:

- SPACE OF MORE THAN 283 M^3 (10,000 FT^3) PER WELDER IS PROVIDED.
- CEILING HEIGHT IS MORE THAN 5 M (16 FT).
- WELDING IS NOT DONE IN A CONFINED AREA.
- THE GENERAL WELDING AREA IS FREE OF PARTITIONS, BALCONIES, OR OTHER STRUCTURAL BARRIERS THAT SIGNIFICANTLY OBSTRUCT CROSS VENTILATION. (GENERAL AREA REFERS TO A BUILDING OR A ROOM IN A BUILDING, NOT A WELDING BOOTH OR SCREENED AREA THAT IS USED TO PROVIDE PROTECTION FROM WELDING RADIATION.)
- TOXIC MATERIALS WITH LOW PERMISSIBLE EXPOSURE LIMITS ARE NOT PRESENT AS DELIBERATE CONSTITUENTS. (THE EMPLOYER SHOULD REFER TO THE MATERIAL SAFETY DATA SHEET PROVIDED BY THE MATERIAL SUPPLIER.) WHEN NATURAL VENTILATION IS NOT SUFFICIENT, FANS MAY BE USED TO FORCE AND DIRECT THE REQUIRED AMOUNT OF AIR THROUGH A BUILDING OR WORK ROOM.

The effectiveness of general ventilation, natural or forced, depends on the design of the system. Work areas where fresh air is introduced and contaminated air is exhausted must be arranged so that the welding fumes and gases are carried away, not concentrated in dead zones. In some cases, the fresh air supply may be located so that incoming air provides the required protection for the welders as well as for personnel in the general area.

Air movement should always be from either side across the welder. This makes it easier for the welder to keep out of the plume and to keep fumes and gases from entering the welding helmet. Air should not blow toward the face or back of the welder because it may force the fume behind the helmet.

General mechanical ventilation may be a necessary supplement to local ventilation to maintain the background level of airborne contaminants within acceptable limits. Local ventilation is usually necessary to provide satisfactory health hazard control at the individual welding station.

Local Ventilation. General ventilation may control contamination levels in the broad area, but in many cases, it will not provide the local control needed to protect the welder. Local exhaust ventilation is usually the most effective way of providing protection at the local work station. It provides efficient and economical fume control and may be applied by one of the methods listed below.

A fixed open or enclosing hood consisting of a top and at least two sides can be placed to surround the welding or cutting operation. It must have sufficient air flow and velocity to keep contaminant levels at or below permissible limits. Additional information is available in *Industrial Ventilation: A Manual of Recommended Practice,* American Conference of Governmental Industrial Hygienists (latest edition).

A movable hood with a flexible duct can be positioned by the welder as close as practicable to the point of welding (Fig. 4). It should have sufficient air flow to produce a velocity of 30 m/min (100 ft/min) in the zone of welding.

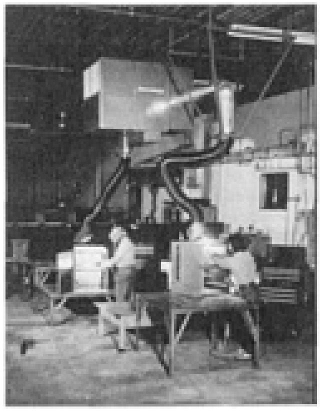

FIG. 4 MOVABLE HOOD POSITIONED NEAR THE WELDING ARC

Air flow requirements range from 4 m^3/min (150 ft^3/min) when the hood is positioned at 100 to 150 mm (4 to 6 in.) from the weld, to 17 m^3/min (600 ft^3/min) at 250 to 300 mm (10 to 12 in.) from the weld. This is particularly applicable for bench work but may be used for any location, provided the hood is moved as required. An air velocity of 31 m/min (100 ft/min) will not disturb the torch gas shield during gas shielded arc welding, provided adequate gas flow rates are used. Higher air velocities may disturb the gas shield and render it less effective.

Crossdraft or Downdraft Table. A crossdraft table is a welding bench with the exhaust hood placed to draw air horizontally across the table. The welder should face in a direction perpendicular to the air flow so that the air flow is across his or her body. A downdraft table has a grill to support the work above an exhaust hood that draws the air downward and away from the welder's head.

Gun-mounted fume removal equipment extracts the fumes at the point of welding and results in an almost smokeless environment. The exhaust rate must be set so that it does not interfere with the shielding gas pattern provided by the welding process. The flux-cored arc welding process produces large quantities of fume, and virtually all of the fume can be collected using a gun-mounted fume removal device (Fig. 5).

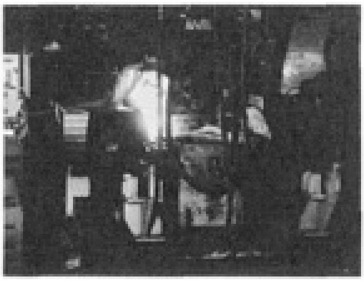

FIG. 5 REMOVAL OF FUME WITH A GUN-MOUNTED DEVICE

Where permissible, air cleaners that have high efficiencies in the collection of submicron particles may be used to recirculate a portion of ventilated air that would otherwise be exhausted. Some air cleaners do not remove gases. Therefore, the filtered air must be monitored to ensure that harmful gas concentrations do not exceed safe limits.

Water Table Used for Oxyfuel Gas and Plasma Arc Cutting Operations. This is a cutting table filled with water near or in contact with the bottom surface of the work. Much fume emerging from the cut is captured in the water.

Respiratory Protective Equipment. Where natural or mechanical ventilation is not adequate or where very toxic materials require a supplement to ventilation, respiratory protective equipment must be used. Airline respirators or face masks that give protection against all contaminants are generally preferred. Air-supplied welding helmets are also available commercially. Filter-type respirators, approved by the U.S. Bureau of Mines for metal fume, give adequate protection against particulate contaminants that are less toxic than lead, provided they are used and maintained correctly. Their general use is not recommended, however, because of the difficulty in ensuring proper use and maintenance. They will not protect against mercury vapor, carbon monoxide, or nitrogen dioxide. For these hazards an airline respirator, hose mask, or gas mask is required.

For additional information, refer to ANSI Z88.2, *Practices for Respiratory Protection,* American National Standards Institute (latest edition).

Special Ventilation Situations

Welding in Confined Spaces. Special consideration must be given to the safety and health of the welders and other workers in confined spaces. Gas cylinders must be located outside of the confined space to avoid possible contamination of the space with leaking gases or volatiles. Welding power sources should also be located outside to reduce danger of engine exhaust and electric shock.

A means for removing persons quickly in case of emergency must be provided. Safety belts and lifelines, when used, should be attached to the worker's body in a manner that avoids the possibility of the person becoming jammed in the exit. A trained helper should be stationed outside the confined space with a preplanned rescue procedure to be put into effect in case of emergency.

In addition to keeping airborne contaminants in breathing atmospheres at or below recommended limits, ventilation in confined spaces must also (1) ensure adequate oxygen for life support (at least 19.5% by volume), (2) prevent accumulation of an oxygen-enriched atmosphere (i.e., not over 23.5% by volume), and (3) prevent accumulation of flammable mixtures. Asphyxiation can quickly result in unconsciousness and death without warning if oxygen is not present in sufficient concentration to support life. Air contains approximately 21% O by volume. A confined space must not be entered unless it is well ventilated, or the welder is properly trained to work in such spaces and is wearing an approved air-supplied breathing apparatus. (Approved air-supplied respirators or hose masks are those accepted by the U.S. Bureau of Mines or other recognized agency.) A similarly equipped second person must be present.

Confined spaces should be tested before entering for (1) toxic or flammable gases and vapors and (2) adequate or excess oxygen. The tests should be made with instruments approved by the U.S. Bureau of Mines. Heavier-than-air gases, such as argon, methylacetylene-propadiene, propane, and carbon dioxide, may accumulate in pits, tank bottoms, low areas, and near floors. Lighter-than-air gases, such as helium and hydrogen, may accumulate in tank tops, high areas, and near ceilings. The precautions for confined spaces also apply to those areas. If practical, a continuous monitoring system with audible alarms should be used for work in a confined space.

Oxygen-enriched atmospheres pose great danger to occupants of confined areas. They are especially hazardous at oxygen concentrations above 25%. Materials that burn normally in air may flare up violently in an oxygen-enriched atmosphere. Clothing may burn fiercely; oil or grease-soaked clothing or rags may catch fire spontaneously; paper may flare into flame. Very severe and fatal burns can result.

Protection in confined spaces must be provided not only for welders but also for other persons in the enclosure. Only clean, respirable air must be used for ventilation. Oxygen, other gases, or mixtures of gases must never be used for ventilation.

When welding, cutting, or related processes are performed in confined areas where adequate and proper ventilation cannot be provided and there is immediate danger to life and health, positive pressure self-contained breathing apparatus must be used. It must have an emergency air supply of at least five minutes duration in the event that the main source fails.

For further precautions, see ANSI Z117.1, *Safety Requirements for Working in Tanks and Other Confined Spaces* (latest edition).

Welding of Containers. Welding or cutting on the outside or inside of containers or vessels that have held dangerous substances presents special hazards. Flammable or toxic vapors may be present, or may be generated by the applied heat. The immediate area outside and inside the container should be cleared of all obstacles and hazardous materials. For more complete procedures, refer to AWS F4.1, *Recommended Safe Practices for the Preparation for Welding and Cutting of Containers and Piping That Have Held Hazardous Substances*, American Welding Society, latest edition. When repairing a container in place, entry of hazardous substances released from the floor or the soil beneath the container must be prevented. The required personal and fire protection equipment must be available, serviceable, and in position for immediate use.

When welding or cutting inside of vessels that have held dangerous materials, the precautions for confined spaces must also be observed.

Gases generated during welding should be discharged in a safe and environmentally acceptable manner in accordance with Government rules and regulations. Provisions must be made to prevent pressure buildup inside containers. Testing for gases, fumes, and vapors should be conducted periodically to ensure that recommended limits are maintained during welding.

An alternative method of providing safe welding of containers is to fill them with an inert medium such as water, inert gas, or sand. When using water, the level should be kept to within a few inches of the point where the welding is to be done. The space above the water should be vented to allow the heated air to escape. With inert gas, the responsible individual needs to know the percentage of inert gas that must be present in the tank to prevent fire or explosion, and how to safely produce and maintain a safe atmosphere during welding.

Highly Toxic Materials. Certain materials, which are sometimes present in consumables, base metals, coatings, or atmospheres for welding or cutting operations, have permissible exposure limits of 1.0 mg/m^3 (4.4×10^{-4} gr/ft^3) or less. Among such materials are the following metals and their compounds:

- ANTIMONY
- ARSENIC
- BARIUM
- BERYLLIUM
- CADMIUM
- CHROMIUM
- COBALT
- COPPER
- LEAD
- MANGANESE
- MERCURY
- NICKEL
- SELENIUM
- SILVER
- VANADIUM

Base metals and filler metals that may release some of these materials as fume during welding or cutting are given in Table 2.

TABLE 2 POSSIBLE TOXIC MATERIALS EVOLVED DURING WELDING OR THERMAL CUTTING

BASE OR FILLER METAL	EVOLVED METALS OR THEIR COMPOUNDS
CARBON AND LOW-ALLOY STEELS	CHROMIUM, MANGANESE, VANADIUM
STAINLESS STEELS	CHROMIUM, MANGANESE, NICKEL
MANGANESE STEELS AND HARDFACING MATERIALS	CHROMIUM, COBALT, MANGANESE, NICKEL, VANADIUM
HIGH-COPPER ALLOYS	BERYLLIUM, CHROMIUM, COPPER, LEAD, NICKEL
COATED OR PLATED STEEL OR COPPER	CADMIUM(A), CHROMIUM, COPPER, LEAD, NICKEL, SILVER

(A) WHEN CADMIUM IS A CONSTITUENT IN A FILLER METAL, A WARNING LABEL MUST BE AFFIXED TO THE CONTAINER OR COIL. REFER TO ANSI/ASC Z49.1, *SAFETY IN WELDING AND CUTTING,* AMERICAN NATIONAL STANDARDS INSTITUTE (LATEST EDITION).

Manufacturer's material safety data sheets should be consulted to determine if any of these highly toxic materials are present in welding filler metals and fluxes being used. Material safety data sheets should be requested from suppliers. However, welding filler metals and fluxes are not the only source of these materials. They may also be present in base metals, coatings, or other sources in the work area. Radioactive materials under Nuclear Regulatory Commission jurisdiction require special considerations.

When toxic materials are encountered as designated constituents in welding, brazing, or cutting operations, special ventilation precautions must be taken to ensure that the levels of these contaminants in the atmosphere are at or below the limits allowed for human exposure. All persons in the immediate vicinity of welding or cutting operations involving toxic materials must be similarly protected. Unless atmospheric tests under the most adverse conditions establish that exposure is within acceptable concentrations, the following precautions must be observed.

Confined Spaces. Whenever any toxic materials are encountered in confined space operations, local exhaust ventilation and respiratory protection must be used.

Indoors. When any toxic materials are encountered in indoor operations, local exhaust mechanical ventilation must be used. When beryllium is encountered, respiratory protection in addition to local exhaust ventilation is essential.

Outdoors. Whenever any toxic materials are encountered in outdoor operations, respiratory protection approved by the Mine Safety and Health Association (MSHA), the National Institute of Occupational Safety and Health (NIOSH), or other approving authority may be required.

Persons should not consume food in areas where fumes that contain materials with very low allowable exposure limits may be generated. They should also practice good personal hygiene, such as washing hands before touching food, to prevent ingestion of toxic contaminants.

Fluorine Compounds. Fumes and gases from fluorine compounds can be dangerous to health and can burn the eyes and skin on contact. Local mechanical ventilation or respiratory protection must be provided when welding, brazing, cutting, or soldering in confined spaces involving fluxes, coatings, or other material containing fluorine compounds.

When such processes are employed in open spaces, the need for local exhaust ventilation or respiratory protection will depend upon the circumstances. Such protection is not necessary when air samples taken in breathing zones indicate that all fluorides are within allowable limits. However, local exhaust ventilation is always desirable for fixed-location production welding and for all production welding of stainless steels when filler metals or fluxes containing fluorides are used.

Zinc. Fumes containing zinc compounds may produce symptoms of nausea, dizziness, or fever (sometimes called metal fume fever). Welding or cutting where zinc may be present in consumables, base metals, or coatings should be done as described for fluorine compounds.

Cleaning compounds often require special ventilation precautions because of their possible toxic or flammable properties. Manufacturers' instructions should be followed before welding or cutting is done with cleaned consumables or on cleaned base metal.

Chlorinated Hydrocarbons. Degreasing or cleaning operations involving chlorinated hydrocarbons must be so located that vapors from such operations do not enter the atmosphere surrounding molten weld metal or the welding arc. A reaction product having an objectionable, irritating odor, and containing highly toxic phosgene gas is produced when such vapors enter the atmosphere of arc welding operations. Low levels of exposure can produce feelings of nausea, dizziness, and weakness. High exposures may produce serious health impairment.

Cutting of stainless steel by oxyfuel gas, gas shielded arc, or plasma arc cutting should be done using local mechanical ventilation to remove the fumes generated. Fume from plasma arc cutting done under water is mostly captured in the water.

Measurement of Exposure

The American Conference of Governmental Industrial Hygienists (ACGIH) and the U.S. Department of Labor, Occupational Safety and Health Administration (OSHA) have established allowable limits of airborne contaminants. They are called threshold limit values (TLV) or permissible exposure limits (PEL).

The TLV is the concentration of an airborne substance to which most workers may be repeatedly exposed day after day, without adverse effect. In adapting these to the working environment, a threshold limit value-time weighted average (TLV-TWA) quantity is defined. TLV-TWA is the time-weighted average concentration for a normal 8 h workday or 40 h workweek to which nearly all workers may be repeatedly exposed without adverse effect. TLV-TWA values should be used as guides in the control of health hazards and should not be interpreted as sharp lines between safe and dangerous concentrations.

TLVs are revised annually as necessary. They may or may not correspond to OSHA permissible limits (PEL) for the same materials. In many cases, current ACGIH values for welding materials are more stringent than OSHA levels.

The only way to ensure that airborne contaminant levels are within the allowable limits is to take air samples at the breathing zones of the personnel involved. An operator's actual on-the-job exposure to welding fume should be measured following the guidelines provided in ANSI/AWS F1.1, *Method for Sampling Airborne Particulates Generated by Welding and Allied Processes.* This document describes how to obtain an accurate breathing zone sample of welding fume for a particular welding operation. Both the amount of fume and the composition of the fume can be determined in a single test using this method. Multiple samples are recommended for increased accuracy. When a helmet is worn, the sample should be collected inside the helmet in the welder's breathing zone.

Safe Practices*

American Welding Society

Handling of Compressed Gases

Gas Cylinders and Containers

Gases used in welding and cutting operations are packaged in containers called cylinders. Only cylinders designed and maintained in accordance with U.S. Department of Transportation (DOT) specifications may be used in the United States. The use of other cylinders may be extremely dangerous and is illegal. Cylinders requiring periodic retest under DOT regulations may not be filled unless the retest is current.

Filling. Cylinders may be filled only with the permission of the owner. They should only be filled by recognized gas suppliers or those with the proper training and facilities to do so. Filling one cylinder from another is dangerous and should not be attempted by anyone not qualified to do so. Combustible or incompatible combinations of gases must never be mixed in cylinders.

Usage and Storage. Welding must not be performed on gas cylinders. Cylinders must not be allowed to become part of an electrical circuit because arcing may result. Cylinders containing shielding gases used in conjunction with arc welding must not be grounded. Electrode holders, welding torches, cables, hoses, and tools should not be stored on gas cylinders, which might cause arcing or interference with valve operation. Are-damaged gas cylinders may rupture and result in injury or death.

Cylinders must not be used as work rests or rollers. They should be protected from bumps, falls, falling objects, and weather, and they should not be dropped. Cylinders should not be kept in passageways where they might be struck by vehicles. They should be kept in areas where temperatures do not fall below -30 °C (-20 °F) nor exceed 55 °C (130 °F). Any of these exposures, misuses, or abuses could damage them to the extent that they might fail, with serious consequences.

Cylinders must not be hoisted using ordinary slings or chains. A proper cradle or cradle sling that securely retains the cylinder should be used. Electromagnets should not be used to handle cylinders.

Cylinders must always be secured by the user against falling during either use or storage. Acetylene and liquefied gas cylinders should always be stored and used in the upright position. Other cylinders are preferably stored and used in the upright position, but this is not essential in all circumstances.

Before using gas from a cylinder, the contents should be identified by the label thereon. Contents must never be identified by any other means such as cylinder color, banding, or shape. These may vary among manufacturers, geographical areas, or product lines and could be completely misleading. The label on the cylinder is the only proper notice of the contents. If a label is not on a cylinder, the contents should not be used and the cylinder should be returned to the supplier.

A valve protection cap is provided on many cylinders to protect the safety device and the cylinder valve. This cap should always be in place except when the cylinder is in use. The cylinder should never be lifted manually or by hoist by the valve protection cap. The threads that secure these valve protection caps are intended only for that purpose and may not be capable of supporting full cylinder weight. The caps should always be threaded completely onto the cylinders and hand tightened.

Gas cylinders and other containers must be stored in accordance with all state and local regulations and the appropriate standards of OSHA and the National Fire Protection Association (NFPA). Safe handling and storage procedures are discussed in the Compressed Gas Association's *Handbook of Compressed Gases.*

Withdrawal of Gas. Many gases in high-pressure cylinders are filled to pressures of 14 MPa (2000 psig) or more. Unless the equipment to be used with a gas is designed to operate at full cylinder pressure, an approved pressure-reducing regulator must be used to withdraw gas from a cylinder or manifold. Simple needle valves should never be used. A pressure-relief or safety valve, rated to function at less than the maximum allowable pressure of the welding equipment, should also be employed. The valve function is to prevent failure of the equipment at pressures in excess of working limits if the regulator should fail in service.

Valves on cylinders containing high-pressure gas, particularly oxygen, should always be opened slowly to avoid the high temperature of adiabatic recompression, which can occur if the valves are opened rapidly. In the case of oxygen, the heat can ignite the valve seat, which, in turn, may cause the metal to melt or burn. The cylinder valve outlet should point away from the operator and other persons when the valve is being opened to avoid injury should a fire occur.

Prior to the connection of a gas cylinder to a pressure regulator or a manifold, the valve outlet should be cleaned of dirt, moisture, and other foreign matter by first being wiped with a clean, oil-free cloth. Then the valve should be opened momentarily and closed immediately. This is known as "cracking the cylinder valve." Fuel gas cylinders must never be cracked near sources of ignition (i.e., sparks and flames), while the operator is smoking, or in confined spaces.

A regulator should be drained of gas pressure prior to being connected to a cylinder and also after the cylinder valve is closed upon shutdown of operation. The outlet threads on cylinder valves are standardized for specific gases so that only regulators or manifolds with similar threads can be attached. (Refer to ANSI/CGA V-l, *Compressed Gas Cylinder Valve Outlet and Inlet Connections,* Compressed Gas Association.) Standard valve thread connections for gases normally used for welding, brazing, and allied processes are given in Table 3.

TABLE 3 COMPRESSED GAS ASSOCIATION STANDARD AND ALTERNATE VALVE THREAD CONNECTIONS FOR COMPRESSED GAS CYLINDERS

GAS	**CONNECTION NO.**
ACETYLENE	510 OR 300
ARGON	580
BUTANE	510
CARBON DIOXIDE (CO_2)	320
HELIUM	580
HYDROGEN	350
METHYLACETYLENE-PROPADIENE (MPS)	510
NITROGEN	580 OR 555
OXYGEN	540
PROPANE	510
PROPYLENE	510

It is preferable not to open valves on low-pressure fuel gas cylinders more than one turn. This usually provides adequate flow and permits rapid closure of the cylinder valve in an emergency. High-pressure cylinder valves, on the other hand, usually must be opened fully to backseat the packing and prevent packing leaks during use.

The cylinder valve should be closed after each use of a cylinder and when an empty cylinder is to be returned to the supplier. This prevents loss of product through leaks that might develop and go undetected while the cylinder is unattended and also avoids hazards that might be caused by leaks. It also prevents backflow of contaminants into the cylinder. It is advisable to return cylinders to the supplier with about 172 kPa (25 psi) of contents remaining. This prevents possible contamination by the atmosphere during shipment.

Pressure-Relief Devices. Only trained personnel should be allowed to adjust pressure-relief devices on cylinders. These devices are intended to provide protection in the event the cylinder is subjected to a hostile environment, usually fire or other source of heat. Such environments may raise the pressure within cylinders. To prevent cylinder pressures from exceeding safe limits, the safety devices are designed to relieve the contents.

Cryogenic Cylinders and Tanks

Cryogenic cylinders and tanks are used to store at very low temperatures those liquids that change to gases at normal conditions of temperature and pressure. So-called cryogenic liquids for commercial purposes include oxygen, nitrogen, and argon, although other gases may be handled as cryogenic liquids.

Cylinders and tanks for storing cryogenic liquids are usually double-walled vessels that are evacuated and insulated between the walls. They are designed to contain liquids at very low temperatures with a minimum of heat gain. Heat gain from the atmosphere causes evaporation of the product. Liquid containers hold a greater amount of product for a given volume than high-pressure gas cylinders. For safety, containers of liquid must be handled carefully. They must always be maintained in an upright position. Whenever they are moved, a cylinder handling truck designed for this purpose must be used. They should not be rolled on a bottom edge, as is often done with high-pressure cylinders.

Failure to properly handle these cylinders can result in rupture of either the inner or outer cylinder wall with consequent loss of vacuum and rapid rise of internal pressure. This will result in activation of the cylinder's protective devices and loss of the cylinder contents. The cylinder protection devices must never be tampered with nor deactivated. Overpressurization could result in explosive failure.

Damage to the internal walls or fittings of a cryogenic-liquid container is often evidenced by visible frosting on the exterior of the container. Whenever frosting appears on the exterior of a cryogenic-liquid container, personnel should be kept clear and the gas supplier notified. In general, people should stay clear of the container until the frost disappears. When the frost disappears it means that the contents have evaporated and any dangerous internal pressure has been relieved.

Although the contents of cryogenic-liquid containers are predominantly in liquid form at very low temperatures, the product withdrawn from these containers should be a gas at room temperature. The conversion takes place within a vaporizer system that evaporates the liquid and warms the gas to atmospheric temperature.

In some cases, the user may want to withdraw liquid from the container. When liquid is withdrawn, protective clothing should be worn to prevent bodily contact with the cold product. Loose-fitting insulated gloves and an adequate face shield are essential. Contact of liquid with the skin will cause burns similar to heat burns, and prolonged contact will cause severe frostbite. Injuries of this nature should be treated in the same manner as injuries from exposure to *low* temperatures.

Compared to their normal properties, the properties of many materials are drastically different at cryogenic liquid temperatures. Many metals, including carbon steel, and most elastomers, such as rubber, become extremely brittle. When cryogenic liquids are to be withdrawn from cylinders, all materials in the transfer line must have satisfactory properties at the low temperatures.

Liquid oxygen may react with explosive violence on contact with asphalt or similar bituminous materials. Therefore, liquid oxygen tanks must never be mounted on such surfaces and liquid oxygen must not be allowed to contact them. Liquid oxygen tanks must always be installed on concrete pads.

Regulators

A pressure-reducing regulator should always be used when withdrawing gas from gas cylinders for welding or cutting operations. Gas regulators should meet the requirements of E-4, *Standard for Gas Regulators for Welding and Cutting*, Compressed Gas Association, and other code regulations.

Pressure-reducing regulators must be used only for the gas and pressure given on the label. They should not be used with other gases, or at other pressures even though the cylinder valve outlet threads may be the same. The threaded connections to the regulator must not be forced. Improper fit of threads between a gas cylinder and regulator or between the regulator and hose indicates that an improper combination of devices is being used.

Use of adapters to change the cylinder connection thread is not recommended because of the danger of using an incorrect regulator or of contaminating the regulator. For example, gases that are oil-contaminated can deposit an oily film on the internal parts of the regulator. This film can contaminate oil-free gas or result in fire or explosion in the case of oxygen.

The threads and connection glands of regulators should be inspected before use for dirt or damage. If a hose or cylinder connection leaks, it should not be forced with excessive torque. Damaged regulators and components should be repaired by properly trained mechanics or returned to the manufacturer for repair.

A suitable valve or flowmeter should be used to control gas flow from a regulator. The internal pressure in a regulator should be drained before it is connected to or removed from a gas cylinder or manifold.

Manifolds

A manifold is used when gas is needed without interruption or at a higher delivery rate than can be supplied from a single cylinder. A manifold must be designed for the specific gas and operating pressure, and be leak tight. The manifold components should be approved for such purpose and used only for the gas and pressure for which they are approved. Oxygen and fuel gas manifolds must meet specific design and safety requirements. (Refer to ANSI/NFPA 51, *Oxygen-Fuel Gas Systems for Welding, Cutting, and Allied Processes,* National Fire Protection Association, latest edition, for information on manifold and piping systems.)

Piping and fittings for acetylene and methyl-acetylene-propadiene (MPS) manifolds must not be unalloyed copper or alloys containing 70% or more copper. These fuel gases react with copper under certain conditions to form unstable copper acetylide. This compound may detonate under shock or heat.

Manifolded piping systems must contain an appropriate over-pressure relief valve. Each fuel gas cylinder lead should incorporate a backflow check valve and a flash arrester. Backflow check valves must also be installed in each line at each station outlet where both fuel gas and oxygen are provided for a welding, cutting, or preheating torch.

Piping Systems. Unless it is known that a piping system is specifically designed and constructed to withstand the full cylinder pressure or tank pressure of the compressed gas source supplying it, the piping system must always be protected with safety pressure-relief devices sufficient to prevent development of pressure in the system beyond the capacity of the weakest element.

Such pressure-relief devices may be relief valves or bursting disks. A pressure-reducing regulator must never be solely relied upon to prevent overpressurization of the system. A pressure-relief device must be located in every section of the system which could be exposed to the full source supply pressure while isolated from other protective relief devices (such as by a closed valve).

Some pressure regulators have integral safety relief valves. These valves are designed for the protection of the regulator only and should not be relied upon to protect the downstream system.

In cryogenic piping systems, relief devices should be located in every section of the system where liquified gas may become trapped. Upon warming, such liquids vaporize to gas, and in a confined space, the gas pressure can increase dramatically.

Pressure-relief devices protecting fuel gas piping systems or other hazardous gas systems should be vented to safe locations.

Oxygen

Oxygen is nonflammable, but it supports the combustion of flammable materials. It can initiate combustion and vigorously accelerate it. Therefore, oxygen cylinders and liquid oxygen containers should not be stored in the vicinity of combustibles nor with cylinders of fuel gas. Oxygen should never be used as a substitute for compressed air. Pure oxygen supports combustion more vigorously than air, which contains only 20% O. Therefore, the identification of oxygen and air should be differentiated.

Oil, grease, and combustible dusts may spontaneously ignite on contact with oxygen. Hence, all systems and apparatus for oxygen service must be kept free of any combustibles. Valves, piping, or system components that have not been expressly manufactured for oxygen service must be cleaned and approved for this service before use. (Refer to G4.1, *Cleaning Equipment for Oxygen Service,* Compressed Gas Association.)

Apparatus that has been manufactured expressly for oxygen service, and is usually so labeled, must be kept in the clean condition as originally received. Oxygen valves, regulators, and apparatus should never be lubricated with oil. If lubrication is required, the type of lubricant and the method of applying the lubricant should be specified in the manufacturer's literature. If it is not, then the device should be returned to the manufacturer or authorized representative for service.

Oxygen must never be used to power compressed air tools. These are almost always lubricated with oil. Similarly, oxygen must not be used to blow dirt from work and clothing because they are often contaminated with oil, or grease, or combustible dust.

Only clean clothing should be worn when working with oxygen systems. Oxygen must not be used to ventilate confined spaces. Severe bums may result from ignition of clothing or the hair in an oxygen-rich atmosphere.

Fuel Gases

Fuel gases commonly used in oxyfuel gas welding (OFW) and cutting (OFC) are acetylene, methylacetylene-propadiene (MPS), natural gas, propane, and propylene. Hydrogen is used in a few applications. Gasoline is sometimes used as a fuel for oxygen cutting. It vaporizes in the torch. These gases should always be referred to by name.

Acetylene in cylinders is dissolved in a solvent so that it can be safely stored under pressure. In the free state, acetylene should never be used at pressures over 15 psig (100 kPa) because it can dissociate with explosive violence at higher pressures.

Acetylene and MPS should never be used in contact with silver, mercury, or alloys containing 70% or more copper. These gases react with these metals to form unstable compounds that may detonate under shock or heat. Valves on fuel gas cylinders should never be opened to clean the valve outlet, especially not near possible sources of flame ignition or in confined spaces.

When fuel gases are used for a brazing furnace atmosphere, they must be burned or vented to a safe location. Prior to filling a furnace or retort with fuel gas, the equipment must first be purged with a nonflammable gas, such as nitrogen or argon, to prevent formation of an explosive air-fuel mixture.

Special attention must be given when using hydrogen. Flames of hydrogen may be difficult to see and parts of the body, clothes, or combustibles may come in contact with hydrogen flames.

Fuel Gas Fires. The best procedures for avoiding fire from a fuel gas or liquid is to keep it contained within the system, that is, to prevent leaks. All fuel systems should be checked carefully for leaks upon assembly and at frequent intervals thereafter. Fuel gas cylinders should be examined for leaks, especially at fuse plugs, safety devices, and valve packing. One common source of fire in welding and cutting is ignition of leaking fuel by flying sparks or spatter.

In the event of a fuel fire, one of the most effective means for controlling the fire is to shut off the fuel valve, if accessible. A fuel gas valve should not be opened beyond the point necessary to provide adequate flow. In this way it can be shut off quickly in an emergency. In most cases, this is less than one turn of the handle. If the immediate valve controlling the burning gas is inaccessible, another upstream valve may cut off the flow of gas.

Most fuel gases in cylinders are in liquid form or dissolved in liquids. Therefore, the cylinders should always be used in the upright position to prevent liquid surges into the system.

A fuel gas cylinder can develop a leak and sometimes result in a fire. In case of fire, the fire alarm should be sounded, and trained fire personnel should be summoned immediately. A small fire in the vicinity of a cylinder valve or a safety device should be extinguished, if possible, by closing the valve or by the use of water, wet cloths, or fire extinguishers. If the leak cannot be stopped, the cylinder should be removed by trained fire personnel to a safe outdoor location, and the supplier notified. A warning sign should be posted, and no smoking or other ignition sources should be permitted in the area.

In the case of a large fire at a fuel gas cylinder, the fire alarm should be actuated, and all personnel should be evacuated from the area. The cylinder should be kept wet by fire personnel with a heavy stream of water to keep it cool. It is usually better to allow the fire to continue to burn and consume the issuing gas rather than attempt to extinguish the flame. If the fire is extinguished, there is danger that the escaping gas may reignite with explosive violence.

Shielding Gases

Argon, helium, nitrogen, and carbon dioxide (CO_2) are used for shielding with some welding processes. All, except carbon dioxide, are used as brazing atmospheres. They are odorless and colorless and can displace air needed for breathing.

Confined spaces filled with these gases must be well ventilated before personnel enter them. If there is any question, the space should be checked first for adequate oxygen concentration with an oxygen analyzer. If an analyzer is not available, an air-supplied respirator should be worn by anyone entering the space. Containers of these gases should not be placed in confined spaces, as discussed previously.

Safe Practices*

American Welding Society

Electrical Safety

Electric Shock

Electric shock can cause sudden death. Injuries and fatalities from electric shock in welding and cutting operations can occur if proper precautionary measures are not followed. Most welding and cutting operations employ some type of electrical equipment. For example, automatic oxyfuel gas cutting machines use electric motor drives, controls, and systems.

Some electrical accidents may not be avoidable, such as those caused by lightning. However, the majority are avoidable, including those caused by lack of proper training.

Shock Mechanism. Electric shock occurs when an electric current of sufficient magnitude to create an adverse effect passes through the body. The severity of the shock depends mainly on the amount of current, the duration of flow, the path of flow, and the state of health of the person. The current is caused to flow by the applied voltage. The amount of current depends upon the applied voltage and the resistance of the body path. The frequency of the current may also be a factor when alternating current is involved.

Shock currents greater than about 6 milliamperes (mA) are considered primary because they are capable of causing direct physiological harm. Steady state currents between 0.5 and 6 mA are considered secondary shock currents. Secondary shock currents are defined as those capable of causing involuntary muscular reactions without normally causing direct physiological harm. The 0.5 mA level is called the perception threshold because it is the point at which most people just begin to feel the tingle from the current. The level of current sensation varies with the weight of the individual and to some extent between men and women.

Shock Sources. Most electrical equipment can be shock hazards if improperly installed, used, or maintained. Shock can occur from lightning-induced voltage surges in power distribution systems. Even earth grounds can attain high potential relative to true ground during severe transient phenomena. Such circumstances, however, are rare.

In welding and cutting work, most electrical equipment is powered from ac sources of between 115 and 575 V, or by engine-driven generators. Most welding is done with less than 100 arc volts. (Fatalities have resulted with equipment operating at less than 80 volts.) Some arc cutting methods operate at over 400 V, and electron-beam welding machines at up to about 150 kV. Most electric shock in the welding industry occurs as the result of accidental contact with bare or poorly insulated conductors operating at such voltages. Therefore, welders must take precautions against contacting bare elements in the welding circuit, as well as those in the primary circuits.

Electrical resistance is usually reduced in the presence of water or moisture. Electrical hazards are often more severe under such circumstances. When arc welding or cutting is to be done under damp or wet conditions including heavy perspiration, the welder must wear dry gloves and clothing in good condition to prevent electric shock. The welder should be protected from electrically conductive surfaces, including the earth, by rubber-soled shoes as a minimum and preferably by an insulating layer such as a rubber mat or dry wooden board. Similar precautions against accidental contact with bare conducting surfaces must be taken when the welder is required to work in a cramped kneeling, sitting, or lying position. Rings and jewelry should be removed before welding to decrease the possibility of electric shock.

Wearers of Pacemakers

The technology of heart pacemakers and the extent to which they are influenced by other electrical devices is constantly changing. It is impossible to make a general statement concerning the possible effects of welding operations on such devices. Wearers of pacemakers or other electronic equipment vital to life should check with the device manufacturer or their doctor to determine whether any hazard exists.

Equipment Selection

Electric shock hazards are minimized by proper equipment installation and maintenance, good operator practice, proper operator clothing and body protection, and the use of equipment designed for the job and situation. Equipment should meet applicable NEMA or ANSI standards, such as American National Standards Institute's ANSI/UL 551, *Safety Standard for Transformer Type Arc Welding Machines,* latest edition.

If a significant amount of welding and cutting work is to be done under electrically hazardous conditions, automatic machine controls that reduce the no-load (open circuit) voltage to a safe level are recommended. When special welding and cutting processes require open-circuit voltages higher than those specified in ANSI/NEMA publication EW-1, *Electrical Arc Welding Apparatus,* insulation and operating procedures that are adequate to protect the welder from these higher voltages must be provided.

Personnel Training

A good safety training program is essential. Employees must be fully instructed in electrical safety by a competent person before being allowed to commence operations. As a minimum, this training should include the points covered in ANSI/ASC Z49.1, *Safety in Welding and Cutting* (published by the American Welding Society). Persons should not be allowed to operate electrical equipment until they have been properly trained.

Equipment Installation

Equipment should be installed in a clean, dry area. When this is not possible, it should be adequately guarded from dirt and moisture. Installation must be done to the requirements of ANSI/NFPA 70, *National Electric Code,* and local codes. This includes necessary disconnects, fusing, and type of incoming power lines.

Terminals for welding leads and power cables must be shielded from accidental contact by personnel or by metal objects, such as vehicles and cranes. Connections between welding leads and power supplies may be guarded using (1) dead front construction and receptacles for plug connections, (2) terminals located in a recessed opening or under a non-removable hinged cover, (3) insulating sleeves, or (4) other equivalent mechanical means.

Grounding

The workpiece being welded and the frame or chassis of all electrically powered machines must be connected to a good electrical ground. Grounding can be done by locating the workpiece or machine on a grounded metal floor or platen, or by connecting it to a properly grounded building frame or other satisfactory ground. Chains, wire ropes, cranes, hoists, and elevators must not be used as grounding connectors nor to carry welding current.

The work lead is not the grounding lead. The work lead connects the work terminal on the power source to the workpiece. A separate lead is required to ground the workpiece or power source work terminal.

Care should be taken when connecting the grounding circuit. Otherwise, the welding current may flow through a connection intended only for grounding and may be of higher magnitude than the grounding conductor can safely carry. Special radio-frequency grounding may be necessary for arc welding machines equipped with high-frequency arc initiating devices. (See EW-1, *Electric Arc Welding Power Sources,* Section 10.5.6, National Electrical Manufacturers Association.)

Connections for portable control devices, such as push buttons to be carried by the operator, must not be connected to circuits with operating voltages above about 120 V. Exposed metal parts of portable control devices operating on circuits above 50 V must be grounded by a grounding conductor in the control cable. Controls using intrinsically safe voltages below 30 V are recommended.

Cables and Connections

Electrical connections must be tight and checked periodically for tightness. Magnetic work clamps must be free of adherent metal particles and spatter on contact surfaces. Coiled welding leads should be spread out before use to avoid overheating and damage to the insulation. Jobs alternately requiring long and short leads should be equipped with insulated cable connectors so that idle lengths can be disconnected when not needed.

Equipment, cables, fuses, plugs, and receptacles must be used within their current carrying and duty cycle capacities. Operation of apparatus above the current rating or the duty cycle results in overheating and rapid deterioration of insulation and other parts. Actual welding current may be higher than that shown by indicators on the welding machine if welding is done with short leads or low voltage, or both. High currents are likely with general-purpose welding machines when they are used with processes that use low arc voltage, such as gas-tungsten arc welding.

Welding leads should be the flexible type of cable designed especially for the rigors of welding service. Insulation on cables used with high voltages or high-frequency oscillators must provide adequate protection. The recommendations and precautions of the cable manufacturer should always be followed. Cable insulation must be kept in good condition, and cables repaired or replaced promptly when necessary.

Operations

Welders should not allow the metal parts of electrodes, electrode holders, or torches to touch their bare skin or any wet covering of the body. Dry gloves in good condition must always be worn. The insulation on electrode holders must be kept in good repair. Electrode holders should not be cooled by immersion in water. If water-cooled welding guns or holders are used, they should be free of water leaks and condensation that would adversely affect the welder's safety. Welders should not drape nor coil the welding leads around their bodies.

A welding circuit must be de-energized to avoid electric shock while the electrode, torch, or gun is being changed or adjusted. One exception concerns covered electrodes with shielded metal arc welding. When the circuit is energized, covered electrodes must be changed with dry welding gloves, not with bare hands. In any case, de-energization of the circuit is desirable for optimum safety even with covered electrodes.

When a welder has completed the work or has occasion to leave the work station for an appreciable time, the welding machine should be turned off. Similarly, when the machine is to be moved, the input power supply should be electrically disconnected at the source. When equipment is not in use, exposed electrodes should be removed from the holder to eliminate the danger of accidental electrical contact with persons or conducting objects. Also, welding guns of semi-automatic welding equipment should be placed so that the gun switch cannot be operated accidently.

Multiple Arc Welding Operations

There can be increased danger of electrical shock when several welders are welding on a large metal structure, such as a building frame or ship, and the structure is part of the return welding circuits. Proper electrical contact must exist at all joints in the structure. Sparking or heating at any point in the structure makes it unsuitable as a return circuit.

Where two or more welders are working on the same structure and one is likely to touch simultaneously the exposed parts of more than one electrode holder, the welding machines must be connected to minimize shock hazard. It is preferable that all dc welding machines be connected with the same polarity. A test lamp or voltmeter can be used to determine whether the polarities are matched. It is preferable to connect all single-phase ac welding machines to the same phase of the supply circuit with the same instantaneous polarity. These precautions minimize the potential difference between electrode holders.

In some cases, the preferable connections may not be possible. Welding may require the use of both dc polarities, or supply circuit limitations may require distribution of ac welding machines among the phases of the supply circuit. In such cases, the no-load voltage between electrode holders or welding guns may be twice the normal voltage. Because of the increased voltage, the welders and other area personnel must be instructed to avoid simultaneous contact with more than one electrode holder, welding gun, or installed electrode.

Modification and Maintenance

Only qualified personnel should perform equipment modification and maintenance. Commutators on rotating welding machines should be kept clean to prevent excessive arcing. Rectifier-type welding machines should be inspected frequently for accumulations of dust or lint that interfere with ventilation. Louvers and internal electrical coil ventilating ducts should be similarly inspected. It is good practice to occasionally blow out the welding machine with clean, dry compressed air at low pressure. Adequate safety precautions such as proper eye protection should be taken. Air filters in the ventilating systems of electrical components are not recommended unless provided by the manufacturer of the welding machine. The filters should be inspected as recommended by the manufacturer. The reduction of air flow resulting from the accumulation of dust on the air filter can subject internal components to an overheating condition and subsequent failure. Machines that have become wet should be thoroughly dried and properly retested before being used.

Prevention of Fires

Fires resulting from electric welding equipment are generally caused by overheating of electrical components, flying sparks or spatter from the welding or cutting operation, or mishandling of fuel in engine-driven equipment. Most precautions against electrical shock are also applicable to the prevention of fires caused by overheating of the equipment. Avoidance of fire from sparks and spatter was covered previously in this article.

The fuel systems of engine-driven equipment must be in good condition. Leaks must be repaired promptly. Engine-driven machines must be turned off before refueling, and any fuel spills should be wiped up and fumes allowed to dissipate before the engine is restarted. Otherwise, the ignition system, electrical controls, spark producing components, or engine heat may start a fire.

Safe Practices*

American Welding Society

Processes

The broad areas of welding safety that are applicable to most welding, cutting, brazing, or soldering processes have been addressed previously in this article. The precautions and procedures that are unique to the particular processes are discussed here. All applicable information must be considered to provide complete safety precautions and procedures for each process.

Oxyfuel Gas Welding and Cutting

Torches. Only approved welding and cutting torches should be used. Oxyfuel gas torches should meet the requirements of E-5, *Torch Standard for Welding and Cutting,* Compressed Gas Association, and appropriate government regulations. They should be kept in good working order and should be serviced at regular intervals by the manufacturer or qualified technicians. A torch must be used only with the fuel gas for which it is designed. The fuel gas and oxygen pressures should be those recommended by the torch manufacturer.

The procedures recommended by the manufacturer should be followed when lighting and extinguishing the torch. The torch should be lighted only with a friction lighter, pilot light, or similar ignition source. Matches or cigarette lighters should not be used.

Hoses. Hoses used should be only those specified for oxyfuel gas welding and cutting systems. Generally, these hoses are manufactured in accordance with *Specification 1P-7 for Rubber Welding Hose,* published by the Compressed Gas Association and the Rubber Manufacturer's Association. Fuel gas hose is usually red with left-hand threaded fittings. Green hose with right-hand threaded fittings is generally used for oxygen. Hose connections must comply with E-l, *Standard Connections for Regulators, Torches and Fitted Hose for Welding and Cutting Equipment*, Compressed Gas Association. Hoses should be free of oil and grease, and in good condition. When parallel lengths are strapped together for convenience, no more than 100 mm (4 in.) of any 300 mm (12 in.) section of hose should be covered.

Only proper ferrules and clamps should be used to secure hose to fittings. Long runs of hose should be avoided. Excess hose should be coiled to prevent kinks and tangles, but it should not be wrapped around cylinders or cylinder carts while in use.

Backfire and Flashback. A backfire during welding and cutting is a momentary retrogression of the flame back into the tip. It usually results in a momentary flame-out followed by reignition of the normal tip flame and is accompanied by a pop or bang, depending upon the size of the tip. In severe cases, the hot combustion products within the tip may be forced back into the torch and even the hoses. Occasionally (but especially with oxygen), such backfires ignite the inner liner of the hose and result in burn-through of the hose wall. Such backfires can cause injury. Furthermore, when the hose ruptures, the flow of gases into the area is continuous until the valve at the tank is closed.

A flashback is an occurrence initiated by a backfire where the flame continues to burn inside the equipment instead of being re-established at the tip. Flashbacks result in very rapid internal heating of the equipment and can quickly destroy it. A flashback is usually recognized by a whistling or squealing sound. The equipment will heat up rapidly and sparks may issue from the tip. The flashback should be extinguished by turning off the torch valves as quickly as possible. Different manufacturers may recommend shutting off either the fuel or oxygen first, but the most important concern is to get both valves closed quickly.

Backfires and flashbacks are not ordinarily a concern if the apparatus is operated in accordance with the manufacturer's instructions. Generally, they occur from allowing a tip to become overheated by flame backwash, forcing the tip into the work, or providing insufficient gas flow for the size of the tip. If frequent backfiring or flashbacks are experienced, the cause should be investigated. There is probably something wrong with the equipment or operation.

To prevent backfires and flashbacks, hose lines should be purged before lighting oxyfuel gas equipment. Purging flushes out any combustible oxygen-fuel or air-fuel gas mixtures in the hoses. It is done by first opening either the fuel or oxygen valve on the torch and allowing that gas to flow for several seconds to clear that hose of any possible gas mixtures. That valve is then closed, and the other valve is opened to allow the other gas to flow for a similar period of time. Purging should always be done before any welding or cutting tip is lighted. The purge stream must not be directed toward any flame or source of ignition. Torches should not be purged in confined spaces because of possible explosion of accumulated gases.

Hose Line Safety Devices. Reverse-flow check valves and flashback arresters for hose line service are available. These devices can prevent backflow of gases and flashbacks into hoses provided they are operating properly. They must be used strictly in accordance with the manufacturer's instructions and maintained regularly in accordance with the manufacturer's recommendations.

Shutdown Procedures. When oxyfuel gas operations are completed, the equipment should always be completely shut down and the gas pressures drained from the system. Cylinder supply valves must be closed. The equipment should not be left unattended until the shutdown has been completed.

Ventilated Storage. Oxyfuel gas cylinders or equipment connected to cylinders must always be stored in well-ventilated spaces and should not be stored in confined areas or unventilated cabinets. Even small gas leaks in confined spaces can result in explosive mixtures that might be ignited with disastrous results. For the same reason, oxygen gas cylinders should never be transported in enclosed vehicles, particularly not in closed vans or the trunks of automobiles.

Arc Welding and Cutting

The potential hazards of arc welding are fumes and gases, electric shock, infrared and ultraviolet radiation, burns, fire, explosion, and sometimes noise. These hazards have been described previously. Precautions to avoid injury or death from these hazards must be followed. The safety precautions for arc welding also apply to arc cutting.

Noise during arc cutting operations can be high. Prolonged exposure to high noise levels can lead to hearing damage. Where necessary, ear protection must be provided for the operator and others in the area.

Plasma arc cutting is a particularly noisy and a high fume generating process. Two common accessories are available for mechanized plasma arc cutting of plate to aid in fume and noise control. One accessory is a water table, which is simply a cutting table filled with water to the bottom surface of the plate or above the plate. In the latter case, cutting is done under water using a special torch to minimize noise and reduce radiation. The high-speed gases emerging from the plasma jet produce turbulence in the water. This action traps almost all of the fume particles in the water.

Another accessory is a water muffler to reduce noise. The muffler is a nozzle attached to a special torch body that produces a curtain of water around the front of the torch. It is always used in conjunction with a water table. The combination of a water curtain at the top of the plate and a water table contacting the bottom of the plate encloses the arc in a noise-reducing shield. The noise output is reduced by roughly 20 dB. This equipment should not be confused with cutting variations using water injection or water shielding.

Resistance Welding

The main hazards of resistance welding processes and equipment are:

- ELECTRIC SHOCK FROM CONTACT WITH HIGH-VOLTAGE TERMINALS OR COMPONENTS
- EJECTION OF SMALL PARTICLES OF MOLTEN METAL FROM THE WELD
- CRUSHING OF SOME PART OF THE BODY BETWEEN THE ELECTRODES OR OTHER MOVING COMPONENTS OF THE MACHINE

Guarding. Initiating devices on resistance welding equipment, such as push buttons and switches, must be positioned or guarded to prevent the operator from inadvertently activating them.

With some machines, the operator's hands can be expected to pass under the point of operation during loading and unloading. These machines must be effectively guarded by proximity-sensing devices, latches, blocks, barriers, dual hand controls, or similar accessories that prevent (1) the hands from passing under the point of operations or (2) the ram from moving while the hands are under the point of operation.

Static Safety Devices. Press, flash, and upset welding machines should have static safety devices to prevent movement of the platen or head during maintenance or setup for welding. Pins, blocks, and latches are examples of such devices.

Portable Welding Machines. The support system of suspended portable welding gun equipment, with the exception of the gun assembly, must be capable of withstanding the total mechanical shock load in the event of failure of any component of the system. Devices such as cables, chains, or clamps are considered satisfactory.

Guarding should be provided around the mounting and actuating mechanism of the movable arm of a portable welding gun if it can cause injury to the operator's hands. If suitable guarding cannot be achieved, the gun should have two handles and two operating switches that must be actuated to energize the machine.

Stop Buttons. One or more emergency stop buttons should be provided on all welding machines, with a minimum of one at each operator position.

Guards. Eye protection against expelled metal particles must be provided by a guard of suitable fire-resistant material or by the use of approved personal protective eye wear. For flash welding equipment, flash guards of suitable fire-resistant material must be provided to control flying sparks and molten metal.

Electrical Considerations. All external welds initiating control circuits should operate at a maximum of about 120 V for stationary equipment and about 36 V for portable equipment.

Resistance welding equipment containing high-voltage capacitors must have adequate electrical insulation and must be completely enclosed. All enclosure doors must be provided with suitable interlocks that are wired into the control circuit. The interlocks must effectively interrupt power and discharge all high-voltage capacitors when the door or panel is open. In addition, a manually operated switch or suitable positive device should be provided to ensure complete discharge of all high-voltage capacitors. The doors or panels must be kept locked except during maintenance.

The welding transformer secondary should be grounded by one of the following methods:

- PERMANENT GROUNDING OF THE WELDING SECONDARY CIRCUIT
- CONNECTION OF A GROUNDING REACTOR ACROSS THE SECONDARY WINDING WITH A REACTOR TAP TO GROUND

As an alternative on stationary machines, an isolation contactor may be used to open all of the primary lines.

The grounding of one side of the secondary windings on multiple spot welding machines can cause undesirable transient currents to flow between the transformers when they are connected to different primary phases or have different secondary voltages, or both. A similar condition can also exist with portable spot welding guns when several units are used to weld the same assembly or another one that is nearby. Such situations require use of a grounding reactor or isolation contactor.

Installation. All equipment should be installed in conformance with the ANSI/NFPA 70, *National Electric Code* (latest edition). The equipment should be installed by qualified personnel under the direction of a competent technical supervisor. Prior to production use, the equipment should be inspected by competent safety personnel to ensure that it is safe to operate.

Brazing and Soldering

Hazards encountered with brazing and soldering operations are similar to those associated with welding and cutting processes. Brazing and soldering operations may be done at temperatures where some elements in the filler metal will vaporize. Personnel and property must be protected against hot materials, gases, fumes, electrical shock, radiation, and chemicals.

It is essential that adequate ventilation be provided so that personnel do not inhale gases and fumes generated during brazing or soldering. Some filler metals and base metals contain toxic materials such as cadmium, beryllium, zinc, mercury, or lead that vaporize during brazing. Fluxes contain chemical compounds of fluorine, chlorine, and boron that are harmful if they are inhaled or contact the eyes or skin. Suitable ventilation to avoid these hazards was described previously.

Brazing Atmospheres. Flammable gases are sometimes used as atmospheres for furnace brazing operations. These include combusted fuel gas, hydrogen, and dissociated ammonia. Prior to introducing such atmospheres, the furnace or retort must be purged of air by safe procedures recommended by the furnace manufacturer.

Adequate area ventilation must be provided to exhaust and discharge to a safe place explosive or toxic gases that may emanate from furnace purging and brazing operations. Local environmental regulations should be consulted when designing the exhaust system.

Dip Brazing and Soldering. In dip brazing and soldering, the parts to be immersed in the bath must be completely dry. The presence of moisture on the parts will cause an instantaneous generation of steam that may expel the contents of the dip pot with explosive force and create a serious burn hazard. Predrying of the parts will prevent this problem. If supplementary flux is necessary, it must be adequately dried to remove not only moisture but also water of hydration to avoid explosion hazards.

Solder Flux. Some fluxes, such as the rosin, petrolatum, and reaction types, give off considerable smoke, the amount depending on the soldering temperature and the duration of heating. The American Conference of Governmental Industrial Hygienists has established a safe threshold limit value for decomposition products of rosin core solder of 0.1

mg/m^3 (4×10^{-5} gr/ft^3) aliphatic aldehydes, measured as formaldehyde. Suitable ventilation must be provided for soldering operations to meet this requirement.

Other fluxes give off fumes that are harmful if breathed in any but small quantities. Prolonged inhalation of halides and some of the newer organic fluxes should be avoided. The aniline type fluxes and some of the amines also evolve fumes that are harmful and can cause dermatitis. Fluorine in flux can be dangerous to health, cause burns, and be fatal if ingested. Thus, proper ventilation must be provided in the work area to remove fume from soldering operations.

High-Frequency Welding

High-frequency generators are electrical devices and require all usual safety precautions in handling and repairing such equipment. Voltages are in the range from 400 to 20,000 V and are lethal. These voltages may be either low or high frequency. Proper care and safety precautions should be taken while working on high-frequency generators and their control systems. Units must be equipped with safety interlocks on access doors and with automatic safety grounding devices to prevent operation of the equipment when access doors are open. The equipment should not be operated with panels or high-voltage covers removed nor with interlocks and grounding devices blocked. This equipment should not be confused with high-frequency arc stabilization equipment used in GTAW.

The output high-frequency primary leads should be encased in metal ducting and should not be operated in the open. Induction coils and contact systems should always be properly grounded for operator protection. High-frequency currents are more difficult to ground than low-frequency currents, and grounding lines must be kept short and direct to minimize inductive impedance. The magnetic field from the output system must not induce heat in adjacent metallic sections and cause fires or burns.

The injuries from high-frequency power, especially at the upper range of welding frequencies, tend to produce severe local surface tissue damage. However, they are not likely to be fatal because current flow is shallow.

High-frequency welding stations often emit a loud steady whine that can cause permanent hearing loss. Ear protection is essential under these circumstances.

Electron-Beam Welding

The primary hazards associated with electron-beam welding equipment are electric shock, x-radiation, fumes and gases, and damaging visible radiation. Precautionary measures must be taken at all times to ensure that proper protective procedures are always observed. AWS F2.1, *Recommended Safe Practices for Electron Beam Welding and Cutting* and ANSI/ASC Z49.1, *Safety in Welding and Cutting* (latest edition) give the general safety requirements that should be strictly adhered to at all times.

Electric Shock. Every electron-beam welding machine operates at voltages above 20 kV. These voltages can cause fatal injury, regardless of whether the machine is referred to as a low-voltage or a high-voltage machine. The manufacturers of electron-beam welding equipment, in meeting various regulatory requirements, produce machines that are well-insulated against high voltage. However, precautions should be exercised with all systems when high voltage is present. The manufacturer's instructions should be followed for operation and maintenance of the equipment.

X-rays. The x-rays generated by an electron-beam welding machine are produced when electrons, traveling at high velocity, collide with matter. The majority of x-rays are produced when the electron beam impinges upon the workpiece. Substantial amounts are also produced when the beam strikes gas molecules or metal vapor in the gun column and work chamber. Underwriters Laboratories and OSHA regulations have established firm rules for permissible x-ray exposure levels, and producers and users of equipment must observe these rules.

Generally, the steel walls of the chamber are adequate protection in systems up to 60 kV, assuming proper design. High-voltage machines utilize lead lining to block x-ray emission beyond the chamber walls. Leaded glass windows are employed in both high- and low-voltage electron-beam systems. Generally, the shielded vacuum chamber walls provide adequate protection for the operator.

In the case of nonvacuum systems, a radiation enclosure must be provided to ensure the safety of the operator and other persons in the area. Thick walls of high-density concrete or other similar material may be employed in place of lead, especially for large radiation enclosures on nonvacuum installations. In addition, special safety precautions should be imposed to prevent personnel from accidently entering or being trapped inside the enclosure when equipment is in operation.

A complete x-ray radiation survey of the electron-beam equipment should always be made at the time of installation and at regular intervals thereafter. This should be done by personnel trained to make a proper radiation survey to ensure initial and continued compliance with all radiation regulations and standards applicable to the site where the equipment is installed.

Fumes and Gases. It is unlikely that the very small amount of air in a high-vacuum electron-beam chamber would be sufficient to produce ozone and oxides of nitrogen in harmful concentrations. However, nonvacuum and medium-vacuum electron-beam systems are capable of producing these by-products, as well as other types of airborne contaminants in concentrations above acceptable levels.

Adequate area ventilation must be employed to reduce concentrations of airborne contaminants around the equipment within permissible exposure limits. Proper exhausting techniques should be employed to maintain residual concentrations in the enclosure within those limits.

Visible Radiation. Direct viewing of intense radiation emitted by molten weld metal can be harmful to eyesight. Viewing of the welding operation should be done through a filter lens commonly used for arc welding.

Laser-Beam Welding and Cutting

The basic hazards associated with laser operation are:

- EYE DAMAGE FROM THE BEAM INCLUDING BURNS OF THE CORNEA OR RETINA, OR BOTH
- SKIN BURNS FROM THE BEAM
- RESPIRATORY SYSTEM DAMAGE FROM HAZARDOUS MATERIALS EVOLVED DURING OPERATION
- ELECTRICAL SHOCK
- CHEMICAL HAZARDS
- CONTACT WITH CRYOGENIC COOLANTS

Laser manufacturers are required to qualify their equipment with the U.S. Bureau of Radiological Health (BRH). Electrical components should be in compliance with NEMA standards. User action is governed by OSHA requirements. In all cases, American National Standard Z136.1, *Safe Use of Lasers* (latest edition), should be followed.

Eye Protection. Eye injury is readily caused by laser beams. With laser beams operating at visible or near infrared wavelengths, even a 5 mW beam can inflict retinal damage. Safety glasses are available that are substantially transparent to visible light but are opaque to specific laser beam outputs. Selective filters for ruby, Nd:YAG, and other laser systems are available. Glasses appropriate to the specific laser system must be used. At longer infrared wavelengths (such as that of a CO_2 laser), ordinarily transparent materials such as glass are opaque. Clear safety glasses with side shields may be used with these systems, and the only light reaching the eye will be from incandescence of the workpiece. Nevertheless, extreme brilliance can result if a plasma is generated at high power and, filter lenses should then be used for viewing the operation.

Burns. Laser burns can be deep and very slow to heal. Exposure must be avoided by appropriate enclosure of the beam or by methods that prevent beam operation unless the beam path is unobstructed. This is particularly important for nonvisible beams that provide no external evidence of their existence unless intercepted by a solid.

Electric Shock. High voltages as well as large capacitor storage devices are associated with lasers. Therefore, the possibility for lethal electrical shock is always present. Electrical system enclosures should have appropriate interlocks on all access doors and provisions for discharging capacitor banks before entry. The equipment should be appropriately grounded.

Fumes and Gases. Hazardous products may be generated from interaction of the beam and the workpiece. For example, plastic materials used for "burn patterns" to identify beam shape and distribution in high-power CO_2 laser systems can generate highly toxic vapors if irradiated in an oxygen-lean atmosphere.

In deep-penetration welding, fine metal fume can arise from the joint. Also, intense plasma generation can produce ozone. Consequently, adequate ventilation and exhaust provisions for laser work areas are necessary.

Friction Welding

Friction welding machines are similar to machine tool lathes in that one workpiece is rotated by a drive system. They are also similar to hydraulic presses in that one workpiece is forced against the other. Therefore, safe practices for lathes and power presses should be used as guides for the design and operation of friction welding machines.

Machines should be equipped with appropriate mechanical guards and shields as well as two-hand operating switches and electrical interlocks. These devices should be designed to prevent operation of the machine when the work area, rotating drive, and force system are accessible to the operator or others.

Operating personnel should wear appropriate eye protection and safety apparel commonly used with machine tool operations. In any case, applicable Occupational Safety and Health Administration (OSHA) standards should be strictly observed.

Explosion Welding

Explosives and explosive devices are a part of explosion welding. Such materials and devices are inherently dangerous, but there are safe methods for handling them. However, if the materials are misused, they can kill or injure, and destroy or damage property.

Explosive materials should be handled and used only by trained personnel who are experienced in that field. Handling and safety procedures must comply with all applicable federal, state, and local regulations. Federal jurisdiction on the sale, transport, storage, and use of explosives is through the U.S. Bureau of Alcohol, Tobacco, and Firearms; the Hazardous Materials Regulation Board of the U.S. Department of Transportation; the Occupational Safety and Health Administration; and the Environmental Protection Agency. Many states and local governments require a blasting license or permit, and some cities have special explosive requirements.

The Institute of Makers of Explosives provides educational publications to promote the safe handling, storage, and use of explosives. The National Fire Protection Association provides recommendations for safe manufacture, storage, handling, and use of explosives. (See ANSI/NFPA 495, *Manufacture, Transportation, Storage, and Use of Explosive Materials,* American National Standards Institute, latest edition.)

Ultrasonic Welding

With high-power ultrasonic equipment, high voltages are present in the frequency converter, the welding head, and the coaxial cable connecting these components. Consequently, the equipment should not be operated with the panel doors open or housing covers removed. Door interlocks are usually installed to prevent introduction of power to the equipment when the high-voltage circuitry is exposed. The cables are shielded fully and present no hazard when properly connected and maintained.

Because of hazards associated with application of clamping force, the operator should not place hands or arms in the vicinity of the welding tip when the equipment is energized. For manual operation, the equipment should be activated by dual palm buttons that meet the requirements of OSHA. Both buttons must be pressed simultaneously to actuate a weld cycle, and both must be released before the next cycle is initiated. For automated systems in which the weld cycle is sequenced with other operations, guards should be installed for operator protection. Such hazards can be further minimized by setting the welding stroke to the minimum that is compatible with workpiece clearance.

Thermite Welding

Thermite mix, in the crucible or on the workpieces, can lead to rapid formation of steam when the chemical reaction for thermite welding takes place. This may cause ejection of molten metal from the crucible. Therefore, the thermite mix should be stored in a dry place, the crucible should be dry, and moisture should not be allowed to enter the system before or during welding.

The work area should be free of combustible materials that may be ignited by sparks or small particles of molten metal. The area should be well ventilated to avoid the buildup of fumes and gases from the reaction. Starting powders and rods should be protected against accidental ignition.

Personnel should wear appropriate protection against hot particles or sparks. This includes full face shields with filter lenses for eye protection and headgear. Safety boots are recommended to protect the feet from hot sparks. Clothing should not have pockets or cuffs that might catch hot particles.

Preheating should be done using the safety precautions applicable to oxyfuel gas equipment and operations.

Thermal Spraying

The potential hazards to the health and safety of personnel involved in thermal spraying operations and to persons in the immediate vicinity are:

- ELECTRICAL SHOCK
- FIRE
- FUMES AND GASES
- DUST
- ARC RADIATION
- NOISE

These hazards are not unique to thermal spraying methods. For example, flame spraying has hazards similar to those associated with the oxyfuel gas welding and cutting processes. Likewise, arc spraying and plasma spraying are similar in many respects to gas metal arc and plasma arc welding, respectively. Safe practices for these processes should be followed when thermal spraying with similar equipment. However, thermal spraying does generate dust and fumes to a greater degree. Additional information may be found in *Thermal Spraying: Practice, Theory, and Application,* American Welding Society, 1985.

Fire Prevention. Airborne finely divided solids, especially metal dusts, must be treated as explosives. To minimize danger from dust explosions, spray booths must have adequate ventilation.

A wet collector of the water-wash type is recommended to collect the spray dust. Bag or filter-type collectors are not recommended. Good housekeeping in the work area should be maintained to avoid accumulation of metal dusts, particularly on rafters, tops of booths, and in floor cracks.

Paper, wood, oily rags, and other combustibles in the spraying area can cause a fire and should be removed before the equipment is operated.

Protection of Personnel. The general requirements for the protection of thermal spray operators are the same as for welders.

Eye Protection. Helmets, hand shields, face shields, or goggles should be used to protect the eyes, face, and neck during all thermal spraying operations. Safety goggles should be worn at all times. Table 4 is a guide for the selection of the proper filter shade number for viewing a specific spraying operation.

TABLE 4 RECOMMENDED EYE FILTER PLATES FOR THERMAL SPRAYING OPERATIONS

OPERATION	FILTER SHADE NUMBERS
WIRE FLAME SPRAYING (EXCEPT MOLYBDENUM)	2 TO 4

WIRE FLAME SPRAYING OF MOLYBDENUM	3 TO 6
FLAME SPRAYING OF METAL POWDER	3 TO 6
FLAME SPRAYING OF EXOTHERMICS OR CERAMICS	4 TO 8
PLASMA AND ARC SPRAYING	9 TO 12
FUSING OPERATIONS	4 TO 6

Respiratory Protection. Most thermal spraying operations require that respiratory protective devices be used by the operator. The nature, type, and magnitude of the fume and gas exposure determine which respiratory protective device should be used. All devices selected for use should be of a type approved by the U.S. Bureau of Mines, National Institute for Occupational Safety and Health, or other approving authority for the purpose intended.

Ear Protection. Ear protectors or properly fitted soft rubber ear plugs should be worn to protect the operator from the high-intensity noise from the process. Federal, state, and local codes should be checked for noise protection requirements.

Protective Clothing. Appropriate protective clothing required for a thermal spraying operation will vary with the size, nature, and location of the work to be performed. When working in confined spaces, flame-resistant clothing and gauntlets should be worn. Clothing should be fastened tightly around the wrists and ankles to keep dusts from contacting the skin.

The intense ultraviolet radiation of plasma and electric arc spraying can cause skin burns through normal clothing. Protection against radiation during arc spraying is practically the same as that for normal arc welding at equivalent current levels.

Adhesive Bonding

Adequate safety precautions must be observed with adhesives. Corrosive materials, flammable liquids, and toxic substances are commonly used in adhesive bonding. Therefore, manufacturing operations should be carefully supervised to ensure that proper safety procedures, protective devices, and protective clothing are being used. All federal, state, and local regulations should be complied with, including OSHA Regulation 29CRF 1900.1000, *Air Contaminants.*

Flammable Materials. All flammable materials, such as solvents, should be stored in tightly sealed drums and issued in suitably labeled safety cans to prevent fires during storage and use. Solvents and flammable liquids should not be used in poorly ventilated, confined areas. When solvents are used in trays, safety lids should be provided. Flames, sparks, or spark-producing equipment must not be permitted in the area where flammable materials are being handled. Fire extinguishers should be readily available.

Toxic Materials. Severe allergic reactions can result from direct contact, inhalation, or ingestion of phenolics and epoxies as well as most catalysts and accelerators. The eyes or skin may become sensitized over a long period of time even though no signs of irritation are visible. Once workers are sensitized to a particular type of adhesive, they may no longer be able to work near it because of allergic reactions. Careless handling of adhesives by production workers may expose others to toxic materials if proper safety rules are not observed. For example, co-workers may touch tools, door knobs, light switches, or other objects contaminated by careless workers.

For the normal individual, proper handling methods that eliminate skin contact with an adhesive should be sufficient. It is mandatory that protective equipment, barrier creams, or both be used to avoid skin contact with certain types of formulations.

Factors to be considered in determining the extent of precautionary measures to be taken include:

- THE FREQUENCY AND DURATION OF EXPOSURE
- THE DEGREE OF HAZARD ASSOCIATED WITH A SPECIFIC ADHESIVE
- THE SOLVENT OR CURING AGENT USED
- THE TEMPERATURE AT WHICH THE OPERATIONS ARE PERFORMED
- THE POTENTIAL EVAPORATION SURFACE AREA EXPOSED AT THE WORK STATION

All these elements should be evaluated in terms of the individual operation.

Precautionary Procedures. A number of measures are recommended in the handling and use of adhesives and auxiliary materials.

Personal Hygiene. Personnel should be instructed in proper procedures to prevent skin contact with solvents, curing agents, and uncured base adhesives. Showers, wash bowls, mild soaps, clean towels, refatting creams, and protective equipment should be provided.

Curing agents should be removed from the hands with soap and water. Resins should be removed with soap and water, alcohol, or a suitable solvent. Any solvent should be used sparingly and be followed by washing with soap and water. In case of allergic reaction or burning, prompt medical aid should be obtained.

Work Area. Areas in which adhesives are handled should be separated from other operations. These areas should contain the following facilities in addition to the proper fire equipment:

- SINK WITH RUNNING WATER
- EYE SHOWER OR RINSE FOUNTAIN
- FIRST AID KIT
- VENTILATING FACILITIES

Ovens, presses, and other curing equipment should be individually vented to remove fume. Vent hoods should be provided at mixing and application stations.

Protective Devices. Plastic or rubber gloves should be worn at all times when working with potentially toxic adhesives. Contaminated gloves must not contact objects that others may touch with their bare hands. Those gloves should be discarded or cleaned using procedures that remove the particular adhesive. Cleaning may require solvents, soap and water, or both. Hands, arms, face, and neck should be coated with a commercial barrier ointment or cream. This type of material may provide short-term protection and facilitate removal of adhesive components by washing.

Full face shields should be worn for eye protection whenever the possibility of splashing exists, otherwise glasses or goggles should be worn. In case of irritation, the eyes should be immediately flushed with water and then promptly treated by a physician.

Protective clothing should be worn at all times by those who work with the adhesives. Shop coats, aprons, or coveralls may be suitable, and they should be cleaned before reuse.

Safe Practices*

American Welding Society

Selected References

- *ARC WELDING AND CUTTING NOISE,* AWS, 1979
- *ARC WELDING SAFELY,* AWS, 1988
- *ASM HANDBOOK,* VOL 4, *HEAT TREATING,* ASM INTERNATIONAL, 1991
- N.C. BALCHIN, *HEALTH AND SAFETY IN WELDING AND ALLIED PROCESSES,* 3RD ED., THE WELDING INSTITUTE, 1983
- L.O. BARTHOLD *ET AL.,* ELECTROSTATIC EFFECTS OF OVERHEAD TRANSMISSION LINES, PART I: HAZARDS AND EFFECTS, *IEEE TRANS. POWER APP. SYST.,* VOL PAS-91, 1972, P 422-444
- *CHARACTERIZATION OF ARC WELDING FUMES,* AWS, 1983

- *CUTTING AND WELDING PROCESSES,* ANSI/NFPA SIB-1977, AMERICAN NATIONAL STANDARDS INSTITUTE/NATIONAL FIRE PROTECTION ASSOCIATION, 1977
- C.F. DALZIEL, EFFECTS OF ELECTRIC CURRENT ON MAN, *ASEE J.,* JUNE 1973, P 18-23
- *EFFECTS OF WELDING ON HEALTH, I, II, III, AND IV,* AWS, 1979, 1981, 1983
- *THE FACTS ABOUT FUME,* THE WELDING INSTITUTE, 1976
- *FUMES AND GASES IN THE WELDING ENVIRONMENT,* AWS, 1979
- *HANDBOOK OF COMPRESSED GASES,* 2ND ED., COMPRESSED GAS ASSOCIATION, INC./VAN NOSTRAND REINHOLD, 1981
- *HANDLING ACETYLENE CYLINDERS IN FIRE SITUATIONS,* SB-4, COMPRESSED GAS ASSOCIATION, 1974
- *OXYFUEL GAS WELDING, CUTTING, AND HEATING SAFELY,* AWS
- *OXYGEN-FUEL GAS SYSTEMS FOR WELDING, CUTTING, AND APPLIED PROCESSES,* ANSI/NFPA S1-1983, AMERICAN NATIONAL STANDARDS INSTITUTE/NATIONAL FIRE PROTECTION ASSOCIATION, 1983
- *RECOMMENDED SAFE PRACTICES FOR ELECTRON BEAM WELDING AND CUTTING,* AWS F2.1-78, AWS, 1978
- *RECOMMENDED SAFE PRACTICES FOR THE PREPARATION FOR WELDING AND CUTTING OF CONTAINERS THAT HAVE HELD HAZARDOUS SUBSTANCES,* AWS F4.1-80, AWS, 1980
- *RECOMMENDED SAFE PRACTICES FOR THERMAL SPRAYING,* AWS C2.1-73, AWS, 1973
- *SAFE HANDLING OF COMPRESSED GASES IN CONTAINERS,* P-1, COMPRESSED GAS ASSOCIATION, 1974
- *SAFETY IN WELDING AND CUTTING,* ANSI/ASC Z49.1, AWS
- *ULTRAVIOLET REFLECTANCE OF PAINT,* AWS, 1976
- *THE WELDING ENVIRONMENT,* AWS, 1973
- *WELDING FUME CONTROL: A DEMONSTRATION PROJECT,* AWS, 1982
- *WELDING FUME CONTROL WITH MECHANICAL VENTILATION,* 2ND ED., FIREMAN'S FUND INSURANCE COMPANIES, 1981

Glossary of Terms*

American Welding Society Committee on Definitions and Symbols

THE COMMITTEE on Definitions and Symbols was formed by the American Welding Society to establish standard terms and symbols to aid in the communication of welding information. *Standard Welding Terms and Definitions*, ANSI/AWS A3.0, is the major product of work done by the Subcommittee on Definitions in support of this purpose. This glossary is a selection of terms and definitions from ANSI/AWS A3.0; the reader is referred to the latest edition of ANSI/AWS A3.0 for the most current and complete welding vocabulary, including illustrations and commentary.

The standard terms and definitions published here are those that should be used in both the oral and the written language of welding; however, their use is particularly important in the writing of standards (codes, specifications, recommended practices, methods, classifications, and guides) and all other documents concerning welding. Because ANSI/AWS A3.0, from which this glossary is extracted, is intended to be a comprehensive compilation of welding terminology, nonstandard terms used in the welding industry are also included. All terms are either standard or nonstandard, and nonstandard terms are labeled as such.

To make this glossary most useful, the terms are arranged in the conventional dictionary letter-by-letter alphabetical sequence. It is the policy of the American Welding Society to use only generic terms and definitions in this glossary. The numerous proprietary brand and trademark names commonly used to describe welding processes, equipment, and filler metals are not included.

It should be noted that the *ASM Handbook* editorial style for the hyphenation of certain terms deviates from that dictated by ANSI/AWS A3.0. The style shown in this glossary is the standard AWS usage.

- A
- ACTIVATED ROSIN FLUX
- A rosin-base flux containing an additive that increases wetting by the solder.
- ACTUAL THROAT
- The shortest distance between the weld root and the face of a fillet weld.
- ADHESIVE BONDING
- A materials joining process in which an adhesive is placed between the faying surfaces. The adhesive solidifies to produce an adhesive bond.
- AIR CARBON ARC CUTTING (AC-A)
- An arc cutting process that melts base metals by the heat of a carbon arc and removes the molten metal by a blast of air.
- ARC BLOW
- The deflection of an electric arc from its normal path because of magnetic forces.
- ARC BRAZING (AB)
- A brazing process that uses an arc to provide the heat. See carbon arc brazing .
- ARC CUTTING (AC)
- A group of cutting processes that melt the base metal with the heat of an arc between an electrode and the base metal.
- AIR CUTTING GUN (GAS METAL ARC CUTTING)
- A device used in semiautomatic, machine, and automatic arc cutting to transfer current, guide the consumable electrode, and direct the shielding gas.
- ARC FORCE
- The axial force developed by an arc plasma.
- ARC GOUGING
- An arc cutting process variation used to form a bevel or groove.
- ARC OXYGEN CUTTING
- A nonstandard term for oxygen arc cutting .
- ARC SEAM WELD
- A seam weld made by an arc welding process.
- ARC SPOT WELD
- A spot weld made by an arc welding process.
- ARC SPRAYING (ASP)
- A thermal spraying process using an arc between two consumable electrodes of surfacing materials as a heat source and a compressed gas to atomize and propel the surfacing material to the substrate.
- ARC STRIKE
- A discontinuity consisting of any localized remelted metal, heat-affected metal, or change in the surface profile of any part of a weld or base metal resulting from an arc.
- ARC WELDING (AW)
- A group of welding processes that produces coalescence of metals by heating them with an arc, with or without the application of pressure, and with or without the use of filler metal.
- ARC WELDING ELECTRODE
- A component of the welding circuit through which current is conducted and which terminates at the arc.
- ARC WELDING GUN
- A device used in semiautomatic, machine, and automatic arc welding to transfer current, guide the consumable electrode, and direct the shielding gas.
- AS-WELDED
- The condition of weld metal, welded joints, and weldments after welding, but prior to any subsequent thermal, mechanical, or chemical treatments.
- AUTOGENOUS WELD
- A fusion weld made without the addition of filler metal.
- AUTOMATIC WELDING
- Welding with equipment that performs the welding operation without adjustment of the controls by a welding operator. The equipment may or may not load and unload the workpieces. See also machine welding .

- B
- BACK BEAD
- A weld bead resulting from a back weld pass.
- BACKFIRE
- The momentary recession of the flame into the welding tip or cutting tip followed by immediate reappearance or complete extinction of the flame.
- BACK GOUGING
- The removal of weld metal and base metal from the other side of a partially welded joint to facilitate complete fusion and complete joint penetration upon subsequent welding from that side.
- BACKHAND WELDING
- A welding technique in which the welding torch or gun is directed opposite to the progress of welding.
- BACKING
- A material or device placed against the back side of the joint, or at both sides of a weld in electroslag and electrogas welding, to support and retain molten weld metal. The material may be partially fused or remain unfused during welding and may be either metal or nonmetal.
- BACKING BEAD
- A weld bead resulting from a backing pass.
- BACKING FILLER METAL
- A nonstandard term for consumable insert .
- BACKING PASS
- A weld pass made for a backing weld.
- BACKING RING
- Backing in the form of a ring, generally used in the welding of pipe.
- BACKING SHOE
- A nonconsumable backing device used in electroslag and electrogas welding.
- BACKING WELD
- Backing in the form of a weld.
- BACKSTEP SEQUENCE
- A longitudinal sequence in which weld passes are made in the direction opposite to the progress of welding.
- BACK WELD
- A weld made at the back of a single-groove weld.
- BALLING UP
- The formation of globules of molten brazing filler metal or flux due to lack of wetting of the base metal.
- BASE MATERIAL
- The material to be welded, brazed, soldered, or cut. See also base metal and substrate .
- BASE METAL
- The metal to be welded, brazed, soldered, or cut. See also base material and substrate .
- BEAD WELD
- A nonstandard term for surfacing weld .
- BEVEL
- An angular edge preparation.
- BEVEL ANGLE
- The angle formed between the prepared edge of a member and a plane perpendicular to the surface of the member.
- BEVEL GROOVE WELD
- A type of groove weld.
- BIT
- That part of the soldering iron, usually made of copper, that actually transfers heat (and sometimes solder) to the joint.
- BLACKSMITH WELDING
- A nonstandard term for forge welding .
- BLOCK SEQUENCE
- A combined longitudinal and cross-sectional sequence for a continuous multiple-pass weld in which separated increments are completely or partially welded before intervening increments are welded.
- BLOWHOLE
- A nonstandard term for porosity .

- BOND
- See bonding force , covalent bond , mechanical bond , and metallic bond .
- BOND COAT (THERMAL SPRAYING)
- A preliminary (or prime) coat of material that improves adherence of the subsequent spray deposit.
- BONDING FORCE
- The force that holds two atoms together. It results from a decrease in energy as two atoms are brought closer to one another.
- BOND LINE
- The cross section of the interface between thermal spray deposits and substrate, or between adhesive and adherend in an adhesive bonded joint.
- BOTTLE
- A nonstandard term for gas cylinder .
- BOXING
- The continuation of a fillet weld around a corner of a member as an extension of the principal weld.
- BRAZE
- A weld produced by heating an assembly to the brazing temperature using a filler metal having a liquidus above 450 °C (840 °F) and below the solidus of the base metal. The filler metal is distributed between the closely fitted faying surfaces of the joint by capillary action.
- BRAZE INTERFACE
- The interface between filler metal and base metal in a brazed joint.
- BRAZEMENT
- An assembly whose component parts are joined by brazing.
- BRAZER
- One who performs a manual or semiautomatic brazing operation.
- BRAZE WELDING
- A welding process variation in which a filler metal, having a liquidus above 450 °C (840 °F) and below the solidus of the base metal, is used. Unlike brazing, in braze welding the filler metal is not distributed in the joint by capillary action.
- BRAZING (B)
- A group of welding processes that produce coalescence of materials by heating them to the brazing temperature in the presence of a filler metal having a liquidus above 450 °C (840 °F) and below the solidus of the base metal. The filler metal is distributed between the closely fitted faying surfaces of the joint by capillary action.
- BRAZING ALLOY
- A nonstandard term for brazing filler metal .
- BRAZING FILLER METAL
- The metal that fills the capillary joint clearance and has a liquidus above 450 °C (840 °F) but below the solidus of the base metals.
- BRAZING OPERATOR
- One who operates machine or automatic brazing equipment.
- BRITTLE NUGGET
- A nonstandard term used to describe a faying plane failure in a resistance weld peel test.
- BRONZE WELDING
- A nonstandard term for braze welding .
- BUILDUP
- A surfacing variation in which surfacing metal is deposited to achieve the required dimensions. See also buttering .
- BURNER
- A nonstandard term for oxygen cutter .
- BURNING
- A nonstandard term for oxygen cutting .
- BURN THROUGH
- A nonstandard term for excessive melt through or a hole.
- BURN THROUGH WELD
- A nonstandard term for a seam weld or spot weld .
- BUTTERING

- A surfacing variation that deposits surfacing metal on one or more surfaces to provide metallurgically compatible weld metal for the subsequent completion of the weld. See also buildup .
- BUTT JOINT
- A joint between two members aligned approximately in the same plane.
- BUTTON
- That part of a weld, including all or part of the nugget, that tears out in the destructive testing of spot, seam, or projection welded specimens.
- BUTT WELD
- A nonstandard term for a weld in a butt joint.
- C
- CARBON ARC BRAZING (CAB)
- A brazing process that produces coalescence of metals by heating them with an electric arc between two carbon electrodes. The filler metal is distributed in the joint by capillary action.
- CARBON ARC CUTTING (CAC)
- An arc cutting process that severs base metals by melting them with the heat of an arc between a carbon electrode and the base metal.
- CARBON ARC WELDING (CAW)
- An arc welding process that produces coalescence of metals by heating them with an arc between a carbon electrode and the base metal. No shielding is used. Pressure and filler metal may or may not be used.
- CARBONIZING FLAME
- A nonstandard term for reducing flame .
- CAULK WELD
- A nonstandard term for seal weld .
- CHAIN INTERMITTENT WELD
- An intermittent weld on both sides of a joint in which the weld increments on one side are approximately opposite those on the other side.
- CHAMFER
- A nonstandard term for bevel .
- CHEMICAL FLUX CUTTING (FOC)
- An oxygen cutting process that severs base metals using a chemical flux to facilitate cutting.
- CHILL RING
- A nonstandard term for backing ring .
- CLAD BRAZING SHEET
- A metal sheet on which one or both sides are clad with brazing filler metal.
- COALESCENCE
- The growing together or growth into one body of the materials being welded.
- COATED ELECTRODE
- A nonstandard term for covered electrode .
- COATING DENSITY
- A nonstandard term for spray deposit density ratio .
- COEXTRUSION WELDING (CEW)
- A solid-state welding process that produces coalescence of the faying surfaces by heating and forcing base metals through an extrusion die.
- COLD CRACK
- A crack that develops after solidification is complete.
- COLD SOLDERED JOINT
- A joint with incomplete coalescence caused by insufficient application of heat to the base metal during soldering.
- COLD WELDING (CW)
- A solid-state welding process in which pressure is used at room temperature to produce coalescence of metals with substantial deformation at the weld. See also diffusion welding , forge welding , and hot pressure welding .
- COMPLETE FUSION
- Fusion which has occurred over the entire base metal surface intended for welding and between all adjoining weld beads.
- COMPLETE JOINT PENETRATION

- A penetration by weld metal for the full thickness of the base metal in a joint with a groove weld.
- COMPLETE PENETRATION
- A nonstandard term for complete joint penetration .
- CONCAVITY
- The maximum distance from the face of a concave fillet weld perpendicular to a line joining the weld toes.
- CONE
- The conical part of an oxyfuel gas flame next to the orifice of the tip.
- CONSTRICTED ARC (PLASMA ARC WELDING AND CUTTING)
- A plasma arc column that is shaped by a constricting nozzle orifice.
- CONSUMABLE INSERT
- Preplaced filler metal that is completely fused into the joint root and becomes part of the weld.
- CONTACT RESISTANCE (RESISTANCE WELDING)
- Resistance to the flow of electric current between two workpieces or an electrode and a workpiece.
- CONTACT TUBE
- A device that transfers current to a continuous electrode.
- CONVEXITY
- The maximum distance from the face of a convex fillet weld perpendicular to a line joining the weld toes.
- COPPER BRAZING
- A nonstandard term for brazing with a copper filler metal.
- CORED SOLDER
- A solder wire or bar containing flux as a core.
- CORNER-FLANGE WELD
- A flange weld with only one member flanged at the joint.
- CORNER JOINT
- A joint between two members located approximately at right angles to each other.
- CORONA (RESISTANCE WELDING)
- The area sometimes surrounding the nugget of a spot weld at the faying surface which provides a degree of solid-state welding.
- CO_2 WELDING
- A nonstandard term for gas metal arc welding .
- COVALENT BOND
- A primary bond arising from the reduction in energy associated with overlapping half-filled orbitals of two atoms.
- COVERED ELECTRODE
- A composite filler metal electrode consisting of a core of a bare electrode or metal-cored electrode to which a covering sufficient to provide a slag layer on the weld metal has been applied. The covering may contain materials providing such functions as shielding from the atmosphere, deoxidation, and arc stabilization and can serve as a source of metallic additions to the weld.
- CRACK
- A fracture type discontinuity characterized by a sharp tip and high ratio of length and width to opening displacement.
- CRATER
- A depression at the termination of a weld bead.
- CUTTING ATTACHMENT
- A device for converting an oxyfuel gas welding torch into an oxygen cutting torch.
- CUTTING BLOWPIPE
- A nonstandard term for cutting torch.
- CUTTING NOZZLE
- A nonstandard term for cutting tip .
- CUTTING TIP
- The part of an oxygen cutting torch from which the gases issue.
- CUTTING TORCH (ARC)
- A device used in air carbon arc cutting, gas tungsten arc cutting, and plasma arc cutting to control the position of the electrode, to transfer current, and to control the flow of gases.
- CUTTING TORCH (OXYFUEL GAS)

- A device used for directing the preheating flame produced by the controlled combustion of fuel gases and to direct and control the cutting oxygen.
- CYLINDER MANIFOLD
- A multiple header for interconnection of gas or fluid sources with distribution points.
- D
- DEFECT
- A discontinuity or discontinuities that by nature or accumulated effect (for example, total crack length) render a part or product unable to meet minimum applicable acceptance standards or specifications. This term designates rejectability. See also discontinuity and flaw .
- DEPOSIT (THERMAL SPRAYING)
- A nonstandard term for spray deposit .
- DEPOSITED METAL
- Filler metal that has been added during welding.
- DEPOSITION EFFICIENCY (ARC WELDING)
- The ratio of the weight of deposited metal to the net weight of filler metal consumed, exclusive of stubs.
- DEPOSITION EFFICIENCY (THERMAL SPRAYING)
- The ratio of the weight of spray deposit to the weight of the surfacing material sprayed, usually expressed in percent.
- DEPOSITION SEQUENCE
- A nonstandard term for weld pass sequence .
- DEPTH OF FUSION
- The distance that fusion extends into the base metal or previous pass from the surface melted during welding.
- DIFFUSION BONDING
- A nonstandard term for diffusion brazing and diffusion welding .
- DIFFUSION BRAZING (DFB)
- A brazing process that produces coalescence of metals by heating them to brazing temperature and by using a filler metal or an *in situ* liquid phase. The filler metal may be distributed by capillary action or may be placed or formed at the faying surfaces. The filler metal is diffused with the base metal to the extent that the joint properties have been changed to approach those of the base metal. Pressure may or may not be applied.
- DIFFUSION WELDING (DFW)
- A solid-state welding process that produces coalescence of the faying surfaces by the application of pressure at elevated temperature. The process does not involve macroscopic deformation, melting, or relative motion of the workpieces. A solid filler metal may or may not be inserted between the faying surfaces. See also cold welding , forge welding , and hot pressure welding .
- DILUTION
- The change in chemical composition of a welding filler metal caused by the admixture of the base metal or previous weld metal in the weld bead. It is measured by the percentage of base metal or previous weld metal in the weld bead.
- DIP BRAZING (DB)
- A brazing process using the heat furnished by a molten chemical or metal bath. When a molten chemical bath is used, the bath may act as a flux. When a molten metal bath is used, the bath provides the filler metal.
- DIP SOLDERING (DS)
- A soldering process using the heat furnished by a molten metal bath which provides the solder filler metal.
- DIRECT CURRENT ELECTRODE NEGATIVE (DCEN)
- The arrangement of direct current arc welding leads in which the workpiece is the positive pole and the electrode is the negative pole of the welding arc.
- DIRECT CURRENT ELECTRODE POSITIVE (DCEP)
- The arrangement of direct current arc welding leads in which the workpiece is the negative pole and the electrode is the positive pole of the welding arc.
- DIRECT CURRENT REVERSE POLARITY
- A nonstandard term for direct current electrode positive .
- DIRECT CURRENT STRAIGHT POLARITY
- A nonstandard term for direct current electrode negative .
- DISCONTINUITY

- An interruption of the typical structure of a weldment, such as a lack of homogeneity in the mechanical, metallurgical, or physical characteristics of the material or weldment. A discontinuity is not necessarily a defect. See also defect and flaw .
- DOUBLE-BEVEL-GROOVE WELD
- A type of groove weld.
- DOUBLE-FLARE-BEVEL-GROOVE WELD
- A weld in grooves formed by a member with a curved surface in contact with a planar member.
- DOUBLE-FLARE-V-GROOVE WELD
- A weld in grooves formed by two members with curved surfaces.
- DOUBLE-J-GROOVE WELD
- A type of groove weld.
- DOUBLE-SQUARE-GROOVE WELD
- A type of groove weld.
- DOUBLE-U-GROOVE WELD
- A type of groove weld.
- DOUBLE-V-GROOVE WELD
- A type of groove weld.
- DOUBLE-WELDED JOINT
- A fusion-welded joint that is welded from both sides.
- DOWNHAND
- A nonstandard term for flat position .
- DRAG (THERMAL CUTTING)
- The offset distance between the actual and straight line exit points of the gas stream or cutting beam measured on the exit surface of the material.
- E
- EDGE-FLANGE WELD
- A flange weld with two members flanged at the location of welding.
- EDGE JOINT
- A joint between the edges of two or more parallel or nearly parallel members.
- EDGE WELD
- A weld in an edge joint.
- EDGE WELD SIZE
- The weld metal thickness measured at the weld root.
- EFFECTIVE THROAT
- The minimum distance minus any convexity between the weld root and the face of a fillet weld.
- ELECTRIC ARC SPRAYING
- A nonstandard term for arc spraying .
- ELECTRIC BONDING
- A nonstandard term for surfacing by thermal spraying.
- ELECTRIC BRAZING
- A nonstandard term for arc brazing and resistance brazing .
- ELECTRODE
- See welding electrode .
- ELECTRODE EXTENSION
- For gas metal arc welding, flux cored arc welding, and submerged arc welding, the length of unmelted electrode extending beyond the end of the contact tube.
- ELECTRODE FORCE
- The force between the electrodes in making spot, seam, or projection welds by resistance welding.
- ELECTRODE HOLDER
- A device used for mechanically holding the electrode while conducting current to it.
- ELECTRODE INDENTATION (RESISTANCE WELDING)
- The depression formed on the surface of workpieces by electrodes.
- ELECTRODE LEAD
- The electrical conductor between the source of arc welding current and the electrode holder.
- ELECTROGAS WELDING (EGW)
- An arc welding process that produces coalescence of metals by heating them with an arc between a continuous filler metal electrode and the work. Molding shoes are used to confine the molten weld metal

for vertical position welding. The electrodes may be either flux cored or solid. Shielding may or may not be obtained from an externally supplied gas or mixture.

- ELECTRON BEAM CUTTING (EBC)
- A cutting process that uses the heat obtained from a concentrated beam composed primarily of high-velocity electrons which impinge upon the workpieces; it may or may not use an externally supplied gas.
- ELECTRON BEAM GUN
- A device for producing and accelerating electrons. Typical components include the emitter (also called the filament or cathode), which is heated to produce electrons via thermionic emission; a cup (also called the grid or grid cup); and the anode.
- ELECTRON BEAM WELDING (EBW)
- A welding process that produces coalescence of metals with the heat obtained from a concentrated beam composed primarily of high-velocity electrons impinging on the joint.
- ELECTROSLAG WELDING (ESW)
- A welding process that produces coalescence of metals with molten slag that melts the filler metal and the surfaces of the workpieces. The weld pool is shielded by this slag, which moves along the full cross section of the joint as welding progresses. The process is initiated by an arc that heats the slag. The arc is then extinguished by the conductive slag, which is kept molten by its resistance to electric current passing between the electrode and the workpieces.
- END RETURN
- A nonstandard term for boxing .
- EROSION (BRAZING)
- A condition caused by dissolution of the base metal by molten filler metal resulting in a reduction in the thickness of the base metal.
- EXPLOSION WELDING (EXW)
- A solid-state welding process that affects coalescence by high-velocity movement together with the workpieces produced by a controlled detonation.
- F
- FACE REINFORCEMENT
- Weld reinforcement at the side of the joint from which welding was done. See also root reinforcement .
- FAYING SURFACE
- That mating surface of a member that is in contact with or in close proximity to another member to which it is to be joined.
- FERRITE NUMBER
- An arbitrary, standardized value designating the ferrite content of an austenitic stainless steel weld metal. It should be used in place of percent ferrite or volume percent ferrite on a direct replacement basis.
- FILLER METAL
- The metal to be added in making a welded, brazed, or soldered joint. See also brazing filler metal , consumable insert , solder , welding electrode , welding rod , and welding wire .
- FILLER WIRE
- A nonstandard term for welding wire .
- FILLET WELD
- A weld of approximately triangular cross section joining two surfaces approximately at right angles to each other in a lap joint, T-joint, or corner joint.
- FILLET WELD BREAK TEST
- A test in which the specimen is loaded so that the weld root is in tension.
- FILLET WELD LEG
- The distance from the joint root to the toe of the fillet weld.
- FILLET WELD SIZE
- For equal leg fillet welds, the leg lengths of the largest isosceles right triangle that can be inscribed within the fillet weld cross section. For unequal leg fillets, the leg lengths of the largest right triangle that can be inscribed within the fillet weld cross section.
- FILLET WELD THROAT
- See actual throat , effective throat , and theoretical throat .
- FIRECRACKER WELDING
- A variation of the shielded metal arc welding process in which a length of covered electrode is placed along the joint in contact with the workpieces. During the welding operation, the stationary electrode is consumed as the arc travels the length of the electrode.
- FISHEYE

- A discontinuity found on the fracture surface of a weld in steel that consists of a small pore or inclusion surrounded by an approximately round, bright area.
- FLAME CUTTING
- A nonstandard term for oxygen cutting .
- FLAME PROPAGATION RATE
- The speed at which a flame travels through a mixture of gases.
- FLAME SPRAYING (FLSP)
- A thermal spraying process in which an oxyfuel gas flame is the source of heat for melting the surfacing material. Compressed gas may or may not be used for atomizing and propelling the surfacing material to the substrate.
- FLANGE WELD
- A weld made on the edges of two or more members to be joined, usually light gage metal, at least one of the members being flanged.
- FLANGE WELD SIZE
- The weld metal thickness measured at the weld root.
- FLARE-BEVEL-GROOVE WELD
- A weld in a groove formed by a member with a curved surface in contact with a planar member.
- FLARE-V-GROOVE WELD
- A weld in a groove formed by two members with curved surfaces.
- FLASH
- Material that is expelled from a flash weld prior to the upset portion of the welding cycle.
- FLASH BUTT WELDING
- A nonstandard term for flash welding .
- FLASH COAT
- A thin coating usually less than 0.05 mm (0.002 in.) in thickness.
- FLASH WELDING (FW)
- A resistance welding process that produces coalescence at the faying surfaces of a butt joint by a flashing action and by the application of pressure after heating is substantially completed. The flashing action, caused by the very high current densities at small contacts between the parts, forcibly expels the material from the joint as the parts are slowly moved together. The weld is completed by a rapid upsetting of the workpieces.
- FLAT POSITION
- The welding position used to weld from the upper side of the joint; the face of the weld is approximately horizontal.
- FLAW
- A near synonym for discontinuity but with an undesirable connotation. See also defect and discontinuity .
- FLOW BRIGHTENING (SOLDERING)
- Fusion of a metallic coating on a base metal.
- FLUX
- Material used to prevent, dissolve, or facilitate removal of oxides and other undesirable surface substances.
- FLUX CORED ARC WELDING (FCAW)
- An arc welding process that produces coalescence of metals by heating them with an arc between a continuous filler metal electrode and the work. Shielding is provided by a flux contained within the tubular electrode. Additional shielding may or may not be obtained from an externally supplied gas or gas mixture. See also flux cored electrode .
- FLUX CORED ELECTRODE
- A composite filler metal electrode consisting of a metal tube or other hollow configuration containing ingredients to provide such functions as shielding atmosphere, deoxidation, arc stabilization, and slag formation. Minor amounts of alloying materials may be included in the core. External shielding may or may not be used.
- FLUX COVER (METAL BATH DIP BRAZING AND DIP SOLDERING)
- A layer of molten flux over the molten filler metal bath.
- FLUX OXYGEN CUTTING
- A nonstandard term for chemical flux cutting .
- FOREHAND WELDING
- A welding technique in which the welding torch or gun is directed toward the progress of welding.

- FORGE WELDING (FOW)
- A solid-state welding process that produces coalescence of metals by heating them in air in a forge and by applying pressure or blows sufficient to cause permanent deformation at the interface. See also cold welding , diffusion welding , hot pressure welding , and roll welding .
- FRICTION WELDING (FRW)
- A solid-state welding process that produces coalescence of materials under compressive force contact of workpieces rotating or moving relative to one another to produce heat and plastically displace material from the faying surfaces.
- FURNACE BRAZING (FB)
- A brazing process in which the workpieces are placed in a furnace and heated to the brazing temperature.
- FURNACE SOLDERING (FS)
- A soldering process in which the workpieces are placed in a furnace and heated to the soldering temperature.
- FUSED SPRAY DEPOSIT (THERMAL SPRAYING)
- A self-fluxing spray deposit that is subsequently heated to coalescence within itself and with the substrate.
- FUSION
- The melting together of filler metal and base metal (substrate), or of base metal only, which results in coalescence. See also depth of fusion .
- FUSION FACE
- A surface of the base metal that will be melted during welding.
- FUSION WELDING
- Any welding process that uses fusion of the base metal to make the weld.
- FUSION ZONE
- The area of base metal melted as determined on the cross section of a weld.
- G
- GAP
- A nonstandard term for joint clearance and root opening .
- GAS BRAZING
- A nonstandard term for torch brazing .
- GAS CUTTER
- A nonstandard term for oxygen cutter .
- GAS CUTTING
- A nonstandard term for oxygen cutting .
- GAS CYLINDER
- A portable container used for transportation and storage of a compressed gas.
- GAS GOUGING
- A nonstandard term for oxygen gouging .
- GAS LASER
- A laser in which the lasing medium is a gas.
- GAS METAL ARC CUTTING (GMAC)
- An arc cutting process in which metals are severed by melting them with the heat of an arc between a continuous filler metal electrode and the workpiece. Shielding is obtained entirely from an externally supplied gas.
- GAS METAL ARC WELDING (GMAW)
- An arc welding process that produces coalescence of metals by heating them with an arc between a continuous filler metal electrode and the workpieces. Shielding is obtained entirely from an externally supplied gas.
- GAS POCKET
- A nonstandard term for porosity .
- GAS REGULATOR
- A device for controlling the delivery of gas at some substantially constant pressure.
- GAS SHIELDED ARC WELDING
- A general term used to describe flux cored arc welding (when gas shielding is employed), gas metal arc welding , and gas tungsten arc welding .
- GAS TORCH
- A nonstandard term for cutting torch and welding torch.

- GAS TUNGSTEN ARC CUTTING (GTAC)
- An arc cutting process in which metals are severed by melting them with an arc between a single tungsten electrode and the workpiece. Shielding is obtained from a gas.
- GAS TUNGSTEN ARC WELDING (GTAW)
- An arc welding process that produces coalescence of metals by heating them with an arc between a tungsten electrode (nonconsumable) and the workpieces. Shielding is obtained from a gas. Pressure may or may not be used, and filler metal may or may not be used.
- GAS WELDING
- A nonstandard term for oxyfuel gas welding .
- GLOBULAR TRANSFER (ARC WELDING)
- The transfer of molten metal in large drops from a consumable electrode across the arc. See also short circuiting transfer and spray transfer .
- GOUGING
- The forming of a bevel or groove by material removal. See also arc gouging , back gouging , and oxygen gouging .
- GROOVE ANGLE
- The total included angle of the groove between the workpieces.
- GROOVE FACE
- That surface of a joint member included in the groove.
- GROOVE RADIUS
- The radius used to form the shape of a J- or U-groove weld.
- GROOVE WELD
- A weld made in a groove between the workpieces.
- GROOVE WELD SIZE
- The joint penetration of a groove weld.
- GROOVE WELD THROAT
- A nonstandard term for groove weld size .
- GROUND CONNECTION
- An electrical connection of the welding machine frame to the earth for safety. See also workpiece connection and workpiece lead .
- GROUND LEAD
- A nonstandard term for workpiece lead .
- GUN
- See air cutting gun , arc welding gun , electron beam gun , resistance welding gun , soldering gun , and thermal spraying gun .
- H
- HAMMER WELDING
- A nonstandard term for cold welding and forge welding .
- HARD SOLDER
- A nonstandard term for silver-base brazing filler metals.
- HEAT AFFECTED ZONE
- That portion of the base metal that has not been melted, but whose mechanical properties or microstructure have been altered by the heat of welding, brazing, soldering, or cutting.
- HIGH FREQUENCY RESISTANCE WELDING
- A group of resistance welding process variations that uses high frequency welding current to concentrate the welding heat at the desired location.
- HORIZONTAL FIXED POSITION (PIPE WELDING)
- The position of a pipe joint in which the axis of the pipe is approximately horizontal, and the pipe is not rotated during welding.
- HORIZONTAL POSITION (FILLET WELD)
- The position in which welding is performed on the upper side of an approximately horizontal surface and against an approximately vertical surface.
- HORIZONTAL POSITION (GROOVE WELD)
- The position of welding in which the weld axis lies in an approximately horizontal plane and the weld face lies in an approximately vertical plane.
- HORIZONTAL ROLLED POSITION (PIPE WELDING)

- The position of a pipe joint in which the axis of the pipe is approximately horizontal, and welding is performed in the flat position by rotating the pipe.
- HOT CRACK
- A crack that develops during solidification.
- HOT PRESSURE WELDING (HPW)
- A solid-state welding process that produces coalescence of metals with heat and application of pressure sufficient to produce macrodeformation of the base metal. Vacuum or other shielding media may be used. See also diffusion welding and forge welding .
- HYDROGEN BRAZING
- A nonstandard term for any brazing process that takes place in a hydrogen atmosphere.
- I
- IMPULSE (RESISTANCE WELDING)
- A group of pulses occurring on a regular frequency separated only by an interpulse time.
- INCLINED POSITION
- The position of a pipe joint in which the axis of the pipe is at an angle of approximately 45° to the horizontal, and the pipe is not rotated during welding.
- INCLINED POSITION (WITH RESTRICTION RING)
- The position of a pipe joint in which the axis of the pipe is at an angle of approximately 45° to the horizontal, and a restriction ring is located near the joint. The pipe is not rotated during welding.
- INCLUDED ANGLE
- A nonstandard term for groove angle .
- INDUCTION BRAZING (IB)
- A brazing process in which the heat required is obtained from the resistance of the workpieces to induced electric current.
- INDUCTION SOLDERING (IS)
- A soldering process in which the heat required is obtained from the resistance of the workpieces to induced electric current.
- INDUCTION WELDING (IW)
- A welding process that produces coalescence of metals by the heat obtained from the resistance of the workpieces to the flow of induced high frequency welding current with or without the application of pressure. The effect of the high frequency welding current is to concentrate the welding heat at the desired location.
- INERT GAS
- A gas that normally does not combine chemically with the base metal or filler metal. See also protective atmosphere .
- INERT GAS METAL ARC WELDING
- A nonstandard term for gas metal arc welding .
- INERT GAS TUNGSTEN ARC WELDING
- A nonstandard term for gas tungsten arc welding .
- INFRARED BRAZING (IRB)
- A brazing process in which the heat required is furnished by infrared radiation.
- INFRARED SOLDERING (IRS)
- A soldering process in which the heat required is furnished by infrared radiation.
- INTERGRANULAR PENETRATION
- The penetration of a filler metal along the grain boundaries of a base metal.
- INTERPASS TEMPERATURE
- In a multipass weld, the temperature of the weld metal before the next pass is started.
- IRON SOLDERING (INS)
- A soldering process in which the heat required is obtained from a soldering iron.
- J
- J-GROOVE WELD
- A type of groove weld.
- JOINT
- The junction of members or the edges of members which are to be joined or have been joined.
- JOINT CLEARANCE
- The distance between the faying surfaces of a joint. In brazing, this distance is referred to as that which is present before brazing, at the brazing temperature, or after brazing is completed.

- JOINT EFFICIENCY
- The ratio of the strength of a joint to the strength of the base metal, expressed in percent.
- JOINT PENETRATION
- The depth a weld extends from its face into a joint, exclusive of reinforcement.
- JOINT ROOT
- That portion of a joint to be welded where the members approach closest to each other. In cross section, the joint root may be either a point, a line, or an area.
- JOINT TYPE
- A weld joint classification based on the five basic arrangements of the component parts such as butt joint, corner joint, edge joint, lap joint, and T-joint.
- K
- KERF
- The width of the cut produced during a cutting process.
- L
- LAMELLAR TEAR
- A terracelike fracture in the base metal with a basic orientation parallel to the wrought surface. It is caused by the high stress in the thickness direction that results from welding.
- LAND
- A nonstandard term for root face .
- LAP JOINT
- A joint between two overlapping members in parallel planes.
- LASER
- A device that produces a concentrated coherent light beam by stimulating electronic or molecular transitions to lower energy levels. Laser is an acronym for light amplification by stimulated emission of radiation.
- LASER BEAM CUTTING (LBC)
- A thermal cutting process that severs materials by melting or vaporizing them with the heat obtained from a laser beam, with or without the application of gas jets to augment the removal of material.
- LASER BEAM WELDING (LBW)
- A welding process that produces coalescence of materials with the heat obtained from the application of a concentrated coherent light beam impinging upon the joint.
- LEAD BURNING
- A nonstandard term for the welding of lead.
- LIQUATION
- The separation of a low melting constituent of an alloy from the remaining constituents, usually apparent in alloys having a wide melting range.
- LOCKED-UP STRESS
- A nonstandard term for residual stress .
- LONGITUDINAL CRACK
- A crack with its major axis orientation approximately parallel to the weld axis.
- M
- MACHINE WELDING
- Welding with equipment that performs the welding operation under the constant observation and control of a welding operator. The equipment may or may not load and unload the workpieces. See also automatic welding .
- MACROETCH TEST
- A test in which the specimen is prepared with a fine finish and etched to give a clear definition of the weld.
- MANUAL WELDING
- A welding operation performed and controlled completely by hand. See also automatic welding , machine welding , and semiautomatic welding .
- MASK (THERMAL SPRAYING)
- A device for protecting a substrate surface from the effects of blasting or adherence of a spray deposit.
- MECHANICAL BOND (THERMAL SPRAYING)
- The adherence of a spray deposit to a roughened surface by the mechanism of particle interlocking.
- MELT-THROUGH
- Visible root reinforcement produced in a joint welded from one side.

- METAL ARC CUTTING (MAC)
- Any of a group of arc cutting processes that serves metals by melting them with the heat of an arc between a metal electrode and the base metal. See also gas metal arc cutting and shielded metal arc cutting .
- METAL CORED ELECTRODE
- A composite filler metal electrode consisting of a metal tube or other hollow configuration containing alloying materials. Minor amounts of ingredients and fluxing of oxides may be included. External shielding gas may or may not be used.
- METAL ELECTRODE
- A filler or nonfiller metal electrode used in arc welding or cutting that consists of a metal wire or rod that has been manufactured by any method and that is either bare or covered.
- METALLIC BOND
- The principal bond that holds metals together and is formed between base metals and filler metals in all welding processes. This is a primary bond arising from the increased spatial extension of the valence electron wave functions when an aggregate of metal atoms is brought close together. See also bonding force and covalent bond .
- METALLIZING
- A nonstandard term for thermal spraying .
- METALLURGICAL BOND
- A nonstandard term for metallic bond .
- METAL POWDER CUTTING (POC)
- An oxygen cutting process that severs metals through the use of powder, such as iron, to facilitate cutting.
- MIG WELDING
- A nonstandard term for flux cored arc welding and gas metal arc welding .
- MIXING CHAMBER
- That part of a welding or cutting torch in which a fuel gas and oxygen are mixed.
- MOLTEN WELD POOL
- A nonstandard term for weld pool .
- MULTIPORT NOZZLE (PLASMA ARC WELDING AND CUTTING)
- A constricting nozzle containing two or more orifices located in a configuration to achieve a degree of control over the arc shape.
- N
- NEUTRAL FLAME
- An oxyfuel gas flame in which the portion used is neither oxidizing nor reducing. See also oxidizing flame and reducing flame .
- NONTRANSFERRED ARC (PLASMA ARC WELDING AND CUTTING, AND PLASMA SPRAYING)
- An arc established between the electrode and the constricting nozzle. The workpiece is not in the electrical circuit. See also transferred arc .
- NOZZLE
- A device that directs shielding media.
- NUGGET
- The weld metal joining the workpieces in spot, roll spot, seam, or projection welds.
- NUGGET SIZE (RESISTANCE WELDING)
- The diameter of a spot or projection weld or width of a seam weld measured in the plane of the faying surfaces.
- O
- ORIFICE GAS (PLASMA ARC WELDING AND CUTTING)
- The gas that is directed into the torch to surround the electrode. It becomes ionized in the arc to form the plasma and issues from the orifice in the torch nozzle as the plasma jet.
- OVEN SOLDERING
- A nonstandard term for furnace soldering .
- OVERHEAD POSITION
- The position in which welding is performed from the underside of the joint.
- OVERLAP
- The protrusion of weld metal beyond the weld toes or weld root.
- OVERLAP (RESISTANCE SEAM WELDING)

- The portion of the preceding weld nugget remelted by the succeeding weld.
- OVERLAYING
- A nonstandard term for surfacing
- OXIDIZING FLAME
- An oxyfuel gas flame having an oxidizing effect due to excess oxygen. See also neutral flame and reducing flame .
- OXYACETYLENE WELDING (OAW)
- An oxyfuel gas welding process that produces coalescence of metals by heating them with a gas flame or flames obtained from the combustion of acetylene with oxygen. The process may be used with or without the application of pressure and with or without the use of filler metal.
- OXYFUEL GAS CUTTING (OFC)
- A group of cutting processes used to sever metals by means of the chemical reaction of oxygen with the base metal at elevated temperatures. The necessary temperature is maintained by means of gas flames obtained from the combustion of a specified fuel gas and oxygen. See also oxygen cutting .
- OXYFUEL GAS SPRAYING
- A nonstandard term for flame spraying .
- OXYFUEL GAS WELDING (OFW)
- A group of welding processes that produces coalescence by heating materials with an oxyfuel gas flame or flames, with or without the application of pressure, and with or without the use of filler metal.
- OXYGAS CUTTING
- A nonstandard term for oxygen cutting .
- OXYGEN ARC CUTTING (AOC)
- An oxygen cutting process used to sever metals by means of the chemical reaction of oxygen with the base metal at elevated temperatures. The necessary temperature is maintained by an arc between a consumable tubular electrode and the base metal.
- OXYGEN CUTTER
- One who performs a manual oxygen cutting operation.
- OXYGEN CUTTING (OC)
- A group of cutting processes used to sever or remove metals by means of the chemical reaction between oxygen and the base metal at elevated temperatures. In the case of oxidation-resistant metals, the reaction is facilitated by the use of a chemical flux or metal powder. See also chemical flux cutting , metal powder cutting , oxyfuel gas cutting , oxygen arc cutting and oxygen lance cutting .
- OXYGEN CUTTING OPERATOR
- One who operates machine or automatic oxygen cutting equipment.
- OXYGEN GOUGING
- An application of oxygen cutting in which a bevel or groove is formed.
- OXYGEN GROOVING
- A nonstandard term for oxygen gouging .
- OXYGEN LANCE
- A length of pipe used to convey oxygen to the point of cutting in oxygen lance cutting.
- OXYGEN LANCE CUTTING (LOC)
- An oxygen cutting process used to sever metals with oxygen supplied through a consumable lance. The preheat to start the cutting is obtained by other means.
- OXYGEN LANCING
- A nonstandard term for oxygen lance cutting .
- OXYHYDROGEN WELDING (OHW)
- An oxyfuel gas welding process that produces coalescence of materials by heating them with a gas flame or flames obtained from the combustion of hydrogen with oxygen, without the application of pressure and with or without the use of filler metal.
- P
- PARALLEL WELDING
- A resistance welding secondary circuit variation in which the secondary current is divided and conducted through the workpieces and electrodes in parallel electrical paths to simultaneously form multiple resistance spot, seam, or projection welds.
- PARENT METAL
- A nonstandard term for base metal .
- PARTIAL JOINT PENETRATION

- Joint penetration that is intentionally less than complete.
- PENETRATION
- A nonstandard term for joint penetration and root penetration .
- PERCUSSION WELDING (PEW)
- A welding process that produces coalescence at the faying surface using the heat from an arc produced by a rapid discharge of electrical energy. Pressure is applied percussively during or immediately following the electrical discharge.
- PILOT ARC (PLASMA ARC WELDING)
- A low current continuous arc between the electrode and the constricting nozzle to ionize the gas and facilitate the start of the welding arc.
- PLASMA ARC CUTTING (PAC)
- An arc cutting process that severs metal by melting a localized area with a constricted arc and removing the molten material with a high velocity jet of hot, ionized gas issuing from the constricting orifice.
- PLASMA ARC WELDING (PAW)
- An arc welding process that produces coalescence of metals by heating them with a constricted arc between an electrode and the workpiece (transferred arc) or the electrode and the constricting nozzle (nontransferred arc). Shielding is obtained from the hot, ionized gas issuing from the torch which may be supplemented by an auxiliary source of shielding gas. Shielding gas may be an inert gas or a mixture of gases. Pressure may or may not be used, and filler metal may or may not be supplied.
- PLASMA METALLIZING
- A nonstandard term for plasma spraying .
- PLASMA SPRAYING (PSP)
- A thermal spraying process in which a nontransferred arc of a plasma torch is utilized to create a gas plasma that acts as the source of heat for melting and propelling the surfacing material to the substrate.
- PLENUM CHAMBER (PLASMA ARC WELDING AND CUTTING, AND PLASMA SPRAYING)
- The space between the inside wall of the constricting nozzle and the electrode.
- PLUG WELD
- A weld made in a circular hole in one member of a joint, fusing that member to another member. A fillet-welded hole is not to be construed as conforming to this definition.
- POLARITY
- See direct current electrode negative and direct current electrode positive .
- POROSITY
- Cavity type discontinuities formed by gas entrapment during solidification.
- POSTHEATING
- The application of heat to an assembly after welding, brazing, soldering, thermal spraying, or thermal cutting. See also postweld heat treatment .
- POSTWELD HEAT TREATMENT
- Any heat treatment after welding.
- POWDER CUTTING
- A nonstandard term for chemical flux cutting and metal powder cutting .
- PRECOATING
- Coating the base metal in the joint by dipping, electroplating, or other applicable means prior to soldering or brazing.
- PREFORM
- Brazing or soldering filler metal fabricated in a shape or form for a specific application.
- PREHEAT
- A nonstandard term for preheat temperature .
- PREHEAT CURRENT (RESISTANCE WELDING)
- An impulse or series of impulses that occur prior to and are separated from the welding current.
- PREHEAT TEMPERATURE
- A specified temperature that the base metal must attain in the welding, brazing, soldering, thermal spraying, or cutting area immediately before these operations are performed.
- PRESSURE-CONTROLLED WELDING
- A resistance welding process variation in which a number of spot or projection welds are made with several electrodes functioning progressively under the control of a pressure-sequencing device.
- PRESSURE GAS WELDING (PGW)

- An oxyfuel gas-welding process that produces coalescence simultaneously over the entire area of faying surfaces by heating them with gas flames obtained from the combustion of a fuel gas and oxygen and by the application of pressure, without the use of filler metal.
- PRETINNING
- A nonstandard term for precoating .
- PROCEDURE QUALIFICATION
- The demonstration that welds made by a specific procedure can meet prescribed standards.
- PROCEDURE QUALIFICATION RECORD (PQR)
- A document providing the actual welding variables used to produce an acceptable test weld and the results of tests conducted on the weld to qualify a welding procedure specification.
- PROCESS
- A grouping of basic operational elements used in welding, cutting, adhesive bonding, or thermal spraying.
- PROJECTION WELDING (PW)
- A resistance welding process that produces coalescence by the heat obtained from the resistance to the flow of the welding current. The resulting welds are localized at predetermined points by projections, embossments, or intersections.
- PROTECTIVE ATMOSPHERE
- A gas or vacuum envelope surrounding the workpieces used to prevent or facilitate removal of oxides and other detrimental surface substances.
- PUDDLE
- A nonstandard term for weld pool .
- PULL GUN TECHNIQUE
- A nonstandard term for backhand welding .
- PULSE (RESISTANCE WELDING)
- A current of controlled duration of either polarity through the welding circuit.
- R
- RANDOM INTERMITTENT WELDS
- Intermittent welds on one or both sides of a joint in which the weld increments are made without regard to spacing.
- REACTION FLUX (SOLDERING)
- A flux composition in which one or more of the ingredients reacts with a base metal upon heating to deposit one or more metals.
- REACTION STRESS
- A stress that cannot exist in a member if the member is isolated as a free body without connection to other parts of the structure.
- REDUCING ATMOSPHERE
- A chemically active protective atmosphere which at elevated temperature will reduce metal oxides to their metallic state.
- REDUCING FLAME
- A gas flame having a reducing effect due to excess fuel gas. See also neutral flame and oxidizing flame .
- REFLOWING
- A nonstandard term for flow brightening .
- REFLOW SOLDERING
- A nonstandard term for a soldering process variation in which preplaced solder is melted to produce a soldered joint or coated surface.
- RESIDUAL STRESS
- Stress present in a member that is free of external forces or thermal gradients.
- RESISTANCE BRAZING (RB)
- A brazing process in which the heat required is obtained from the resistance to electric current flow in a circuit of which the workpiece is a part.
- RESISTANCE BUTT WELDING
- A nonstandard term for flash welding and upset welding .
- RESISTANCE SEAM WELDING (RSW)
- A resistance welding process that produces coalescence at the faying surfaces of overlapped parts progressively along a length of a joint. The weld may be made with overlapping weld nuggets, a

continuous weld nugget, or by forging the joint as it is heated to the welding temperature by resistance to the flow of the welding current.

- RESISTANCE SOLDERING (RS)
- A soldering process in which the heat required is obtained from the resistance to electric current flow in a circuit of which the workpiece is a part.
- RESISTANCE SPOT WELDING (RSW)
- A resistance welding process that produces coalescence at the faying surfaces of a joint by the heat obtained from resistance to the flow of welding current through the workpieces from electrodes that serve to concentrate the welding current and pressure at the weld area.
- RESISTANCE WELDING (RW)
- A group of welding processes that produces coalescence of the faying surfaces with the heat obtained from resistance of the work to the flow of the welding current in a circuit of which the work is a part, and by the application of pressure.
- RESISTANCE WELDING ELECTRODE
- The part(s) of a resistance welding machine through which the welding current and, in most cases, force are applied directly to the work. The electrode may be in the form of a rotating wheel, rotating bar, cylinder, plate, clamp, chuck, or modification thereof.
- RESISTANCE WELDING GUN
- A manipulatable device to transfer current and provide electrode force to the weld area (usually in reference to a portable gun).
- REVERSE POLARITY
- A nonstandard term for direct current electrode positive .
- ROLL WELDING (ROW)
- A solid-state welding process that produces coalescence of metals by heating and by applying sufficient pressure with rolls to cause deformation at the faying surfaces. See also forge welding .
- ROOT
- A nonstandard term for joint root and weld root .
- ROOT BEAD
- A weld that extends into or includes part or all of the joint root.
- ROOT EDGE
- A root face of zero width. See also root face .
- ROOT FACE
- That portion of the groove face adjacent to the joint root.
- ROOT GAP
- A nonstandard term for root opening .
- ROOT OPENING
- The separation at the joint root between the workpieces.
- ROOT PENETRATION
- The depth that a weld extends into the joint root.
- ROOT RADIUS
- A nonstandard term for groove radius .
- ROOT REINFORCEMENT
- Weld reinforcement opposite the side from which welding was done.
- ROOT SURFACE
- The exposed surface of a weld opposite the side from which welding was done.
- S
- SCARF JOINT
- A form of butt joint.
- SEAL COAT
- Material applied to infiltrate the pores of a thermal spray deposit.
- SEAL WELD
- Any weld designed primarily to provide a specific degree of tightness against leakage.
- SEAM WELD
- A continuous weld made between or upon overlapping members, in which coalescence may start and occur on the faying surfaces, or may have proceeded from the outer surface of one member. The continuous weld may consist of a single weld bead or a series of overlapping spot welds. See also arc seam weld and resistance seam welding .

- SECONDARY CIRCUIT
- That portion of a welding machine that conducts the secondary current between the secondary terminals of the welding transformer and the electrodes, or electrode and workpiece.
- SELF-FLUXING ALLOYS (THERMAL SPRAYING)
- Surfacing materials that wet the substrate and coalesce when heated to their melting point, without the addition of a flux.
- SEMIAUTOMATIC ARC WELDING
- Arc welding with equipment that controls only the filler metal feed. The advance of the welding is manually controlled.
- SERIES WELDING
- A resistance welding secondary circuit variation in which the secondary current is conducted through the workpieces and electrodes or wheels in a series electrical path to simultaneously form multiple resistance spot, seam, or projection welds.
- SET DOWN
- A nonstandard term for upset .
- SHADOW MASK
- A thermal spraying process variation in which an area is partially shielded during thermal spraying, thus permitting some overspray to produce a feathering at the coating edge.
- SHEET SEPARATION (RESISTANCE WELDING)
- The gap surrounding the weld between faying surfaces, after the joint has been welded in spot, seam, or projection welding.
- SHIELDED METAL ARC CUTTING (SMAC)
- A metal arc cutting process in which metals are severed by melting them with the heat of an arc between a covered metal electrode and the base metal.
- SHIELDED METAL ARC WELDING (SMAW)
- An arc welding process that produces coalescence of metals by heating them with an arc between a covered metal electrode and the workpieces. Shielding is obtained from decomposition of the electrode covering. Pressure is not used, and filler metal is obtained from the electrode.
- SHIELDING GAS
- Protective gas used to prevent atmospheric contamination.
- SHORT CIRCUITING TRANSFER (ARC WELDING)
- Metal transfer in which molten metal from a consumable electrode is deposited during repeated short circuits. See also globular transfer and spray transfer .
- SHOULDER
- A nonstandard term for root face .
- SHRINKAGE STRESS
- A nonstandard term for residual stress .
- SHRINKAGE VOID
- A cavity type discontinuity normally formed by shrinkage during solidification.
- SILVER ALLOY BRAZING
- A nonstandard term for brazing with a silver-base filler metal.
- SILVER SOLDERING
- A nonstandard term for brazing with a silver-base filler metal.
- SINGLE-BEVEL-GROOVE WELD
- A type of groove weld.
- SINGLE-FLARE-BEVEL-GROOVE WELD
- A weld in a groove formed by a member with a curved surface in contact with a planar member.
- SINGLE-FLARE-V-GROOVE WELD
- A weld in a groove formed by two members with curved surfaces.
- SINGLE IMPULSE WELDING
- A resistance welding process variation in which spot, projection, or upset welds are made with a single impulse.
- SINGLE-J-GROOVE WELD
- A type of groove weld.
- SINGLE-PORT NOZZLE
- A constricting nozzle containing one orifice, located below and concentric with the electrode.
- SINGLE-SQUARE-GROOVE WELD

- A type of groove weld.
- SINGLE-U-GROOVE WELD
- A type of groove weld.
- SINGLE-V-GROOVE WELD
- A type of groove weld.
- SINGLE-WELDED JOINT
- A fusion welded joint that is welded from one side only.
- SKULL
- The unmelted residue from a liquated filler metal.
- SLAG INCLUSION
- Nonmetallic solid material entrapped in weld metal or between weld metal and base metal.
- SLOT WELD
- A weld made in an elongated hole in one member of a joint fusing that member to another member. The hole may be open at one end. A fillet weld slot is not to be construed as conforming to this definition.
- SLUGGING
- The act of adding a separate piece or pieces of material in a joint before or during welding that results in a welded joint not complying with design, drawing, or specification requirements.
- SOFT SOLDER
- A nonstandard term for solder .
- SOLDER
- A filler metal used in soldering that has a liquidus not exceeding 450 °C (840 °F).
- SOLDERING (S)
- A group of welding processes that produces coalescence of materials by heating them to the soldering temperature and by using a filler metal having a liquidus not exceeding 450 °C (840 °F) and below the solidus of the base metals. The filler metal is distributed between the closely fitted faying surfaces of the joint by capillary action.
- SOLDERING GUN
- An electrical soldering iron with a pistol grip and a quick heating, relatively small bit.
- SOLDERING IRON
- A soldering tool having an internally or externally heated metal bit usually made of copper.
- SOLDER INTERFACE
- The interface between filler metal and base metal in a soldered joint.
- SOLID-STATE WELDING (SSW)
- A group of welding processes that produces coalescence at temperatures essentially below the melting point of the base metal without the addition of a brazing filler metal. Pressure may or may not be used.
- SPACER STRIP
- A metal strip or bar prepared for a groove weld and inserted in the joint root to serve as a backing and to maintain the root opening during welding. It can also bridge an exceptionally wide root opening due to poor fit.
- SPIT
- A nonstandard term for flash .
- SPLIT PIPE BACKING
- Backing in the form of a pipe segment used for welding round bars.
- SPOOL
- A filler metal package consisting of a continuous length of welding wire in coil form wound on a cylinder (called a barrel) which is flanged at both ends. The flange contains a spindle hole of smaller diameter than the inside diameter of the barrel.
- SPOT WELD
- A weld made between or upon overlapping members in which coalescence may start and occur on the faying surfaces or may proceed from the outer surface of one member. The weld cross section (plan view) is approximately circular. See also arc spot weld and resistance spot welding .
- SPRAY DEPOSIT
- The coating or layer of surfacing material applied by a thermal spraying process.
- SPRAY DEPOSIT DENSITY RATIO (THERMAL SPRAYING)
- The ratio of the density of the spray deposit to the theoretical density of a surfacing material, usually expressed as a percent of theoretical density.
- SPRAY TRANSFER (ARC WELDING)

- Metal transfer in which molten metal from a consumable electrode is propelled axially across the arc in small droplets. See also globular transfer and short circuiting transfer .
- SQUARE-GROOVE WELD
- A type of groove weld.
- STACK CUTTING
- Thermal cutting of stacked metal plates arranged so that all the plates are severed by a single cut.
- STAGGERED INTERMITTENT WELD
- An intermittent weld on both sides of a joint in which the weld increments on one side are alternated with respect to those on the other side.
- STANDOFF DISTANCE
- The distance between a nozzle and the workpiece.
- STICK ELECTRODE
- A nonstandard term for covered electrode .
- STICK ELECTRODE WELDING
- A nonstandard term for shielded metal arc welding .
- STICKOUT
- A nonstandard term for electrode extension .
- STOPOFF
- A material used on the surfaces adjacent to the joint to limit the spread of soldering or brazing filler metal.
- STRAIGHT POLARITY
- A nonstandard term for direct current electrode negative .
- STRANDED ELECTRODE
- A composite filler metal electrode consisting of stranded wires that may mechanically enclose materials to improve properties, stabilize the arc, or provide shielding.
- STRESS RELIEF CRACKING
- Intergranular cracking in the heat affected zone or weld metal that occurs during the exposure of weldments to elevated temperatures during postweld heat treatment or high temperature service.
- STRESS RELIEF HEAT TREATMENT
- Uniform heating of a structure or a portion thereof to a sufficient temperature to relieve the major portion of the residual stresses, followed by uniform cooling.
- STRINGER BEAD
- A type of weld bead made without appreciable weaving motion. See also weave bead .
- STUB
- The short length of welding rod or consumable electrode that remains after its use for welding.
- STUD ARC WELDING (SW)
- An arc welding process that produces coalescence of metals by heating them with an arc between a metal stud, or similar part, and the other workpiece. When the surfaces to be joined are properly heated, they are brought together under pressure. Partial shielding may be obtained by the use of a ceramic ferrule surrounding the stud. Shielding gas or flux may or may not be used.
- STUD WELDING
- A general term for joining a metal stud or similar part to a workpiece. Welding may be accommodated by arc, resistance, friction, or other processes with or without external gas shielding.
- SUBMERGED ARC WELDING (SAW)
- An arc welding process that produces coalescence of metals by heating them with an arc or arcs between a bare metal electrode or electrodes and the workpieces. The arc and molten metal are shielded by a blanket of granular, fusible material on the workpieces. Pressure is not used, and filler metal is obtained from the electrode and sometimes from a supplemental source (welding rod, flux, or metal granules).
- SUBSTRATE
- Any material to which a thermal spray deposit is applied.
- SURFACE EXPULSION
- Expulsion occurring at an electrode-to-workpiece contact rather than at the faying surface.
- SURFACING
- The application by welding, brazing, or thermal spraying of a layer or layers of material to a surface to obtain desired properties or dimensions, as opposed to making a joint.
- SURFACING MATERIAL
- The material that is applied to a base metal or substrate during surfacing.

- SURFACING METAL
- The metal that is applied to a base metal or substrate during surfacing. See also surfacing material .
- SURFACING WELD
- A weld applied to a surface, as opposed to making a joint, to obtain desired properties or dimensions.
- SWEAT SOLDERING
- A soldering process variation in which two or more parts that have been precoated with solder are reheated and assembled into a joint without the use of additional solder.
- SYNCHRONOUS TIMING (RESISTANCE WELDING)
- The initiation of each half cycle of welding transformer primary current on an accurately timed delay with respect to the polarity reversal of the power supply.
- T
- TACKER
- A nonstandard term for a tack welder.
- TACK WELD
- A weld made to hold parts of a weldment in proper alignment until the final welds are made.
- THEORETICAL THROAT
- The distance from the beginning of the joint root perpendicular to the hypotenuse of the largest right triangle that can be inscribed within the cross section of a fillet weld. This dimension is based on the assumption that the root opening is equal to zero.
- THERMAL CUTTING (TC)
- A group of cutting processes that melts the base metal. See also arc cutting , electron beam cutting , laser beam cutting , and oxygen cutting .
- THERMAL SPRAYING (THSP)
- A group of processes in which finely divided metallic or non-metallic surfacing materials are deposited in a molten or semimolten condition on a substrate to form a spray deposit. The surfacing material may be in the form of powder, rod, or wire. See also arc spraying , flame spraying , and plasma spraying .
- THERMAL SPRAYING GUN
- A device for heating, feeding, and directing the flow of a surfacing material.
- THERMAL STRESS
- Stress resulting from nonuniform temperature distribution.
- THERMIT CRUCIBLE
- The vessel in which the thermit reaction takes place.
- THERMIT MIXTURE
- A mixture of metal oxide and finely divided aluminum with the addition of alloying metals as required.
- THERMIT MOLD
- A mold formed around the workpieces to receive the molten metal.
- THERMIT REACTION
- The chemical reaction between metal oxide and aluminum that produces superheated molten metal and a slag containing aluminum oxide.
- THERMIT WELDING (TW)
- A welding process that produces coalescence of metals by heating them with superheated liquid metal from a chemical reaction between a metal oxide and aluminum, with or without the application of pressure. Filler metal is obtained from the liquid metal. (*Note*: The *ASM Handbook* uses the spelling "thermit*e* welding.")
- THERMOCOMPRESSION BONDING
- A nonstandard term for hot pressure welding .
- THROAT OF A FILLET WELD
- See actual throat , effective throat , and theoretical throat .
- THROAT OF A GROOVE WELD
- A nonstandard term for groove weld size .
- TIG WELDING
- A nonstandard term for gas tungsten arc welding .
- TINNING
- A nonstandard term for precoating .
- T-JOINT
- A joint between two members located approximately at right angles to each other in the form of a T.
- TOE CRACK

- A crack in the base metal at the toe of a weld.
- TORCH BRAZING (TB)
- A brazing process in which the heat required is furnished by a fuel gas flame.
- TORCH SOLDERING (TS)
- A soldering process in which the heat required is furnished by a fuel gas flame.
- TORCH TIP
- See cutting tip and welding tip .
- TRANSFERRED ARC (PLASMA ARC WELDING)
- A plasma arc established between the electrode and the workpiece.
- TRANSVERSE CRACK
- A crack with its major axis oriented approximately perpendicular to the weld axis.
- TWIN CARBON ARC BRAZING
- A nonstandard term for carbon arc brazing .
- U
- U-GROOVE WELD
- A type of groove weld.
- ULTRASONIC COUPLER (ULTRASONIC SOLDERING AND ULTRASONIC WELDING)
- Elements through which ultrasonic vibration is transmitted from the transducer to the tip.
- ULTRASONIC SOLDERING
- A soldering process variation in which high frequency vibratory energy is transmitted through molten solder to remove undesirable surface films and thereby promote wetting of the base metal. This operation is usually accomplished without a flux.
- ULTRASONIC WELDING (USW)
- A solid-state welding process that produces coalescence of materials by the local application of high frequency vibratory energy as the workpieces are held together under pressure.
- UNDERBEAD CRACK
- A crack in the heat affected zone generally not extending to the surface of the base metal.
- UNDERCUT
- A groove melted into the base metal adjacent to the weld toe or weld root and left unfilled by weld metal.
- UNDERFILL
- A depression on the weld face or root surface extending below the adjacent surface of the base metal.
- UPSET
- Bulk deformation resulting from the application of pressure in welding. The upset may be measured as a percent increase in interface area, a reduction in length, a percent reduction in lap joint thickness, or a reduction in cross wire weld stack height.
- UPSET BUTT WELDING
- A nonstandard term for upset welding .
- UPSET DISTANCE
- The total loss of axial length of the workpieces from the initial contact to the completion of the weld. In flash welding, the upset distance is equal to the platen movement from the end of flash time to the end of upset.
- UPSET WELDING (UW)
- A resistance welding process that produces coalescence over the entire area of faying surfaces or progressively along a butt joint by the heat obtained from the resistance to the flow of welding current through the area where those surfaces are in contact. Pressure is used to complete the weld.
- V
- VACUUM BRAZING
- A nonstandard term for various brazing processes that take place in a chamber or retort below atmospheric pressure.
- VERTICAL POSITION
- The position of welding in which the weld axis is approximately vertical.
- VERTICAL POSITION (PIPE WELDING)
- The position of a pipe joint in which welding is performed in the horizontal position and the pipe may or may not be rotated.
- V-GROOVE WELD
- A type of groove weld.
- W

- WAVE SOLDERING (WS)
- An automatic soldering process where workpieces are passed through a wave of molten solder. See also dip soldering .
- WAX PATTERN (THERMIT WELDING)
- Wax molded around the workpieces to the form desired for the completed weld.
- WEAVE BEAD
- A type of weld bead made with transverse oscillation.
- WELD
- A localized coalescence of metals or non-metals produced either by heating the materials to the welding temperature, with or without the application of pressure, or by the application of pressure alone, with or without the use of filler metal.
- WELDABILITY
- The capacity of a material to be welded under the imposed fabrication conditions into a specific, suitably designed structure and to perform satisfactorily in the intended service.
- WELD AXIS
- A line through the length of the weld, perpendicular to and at the geometric center of its cross section.
- WELD BEAD
- A weld resulting from a pass. See also stringer bead and weave bead .
- WELD BONDING
- A resistance spot welding process variation in which the spot weld strength is augmented by adhesive at the faying surfaces.
- WELD BRAZING
- A joining method that combines resistance welding with brazing.
- WELD CRACK
- A crack located in the weld metal or heat affected zone.
- WELDER
- One who performs manual or semiautomatic welding operation.
- WELDER PERFORMANCE QUALIFICATION
- The demonstration of a welder's ability to produce welds meeting prescribed standards.
- WELD FACE
- The exposed surface of a weld on the side from which welding was done.
- WELDING
- A materials joining process used in making welds.
- WELDING BLOWPIPE
- A nonstandard term for welding torch.
- WELDING CURRENT
- The current in the welding circuit during the making of a weld.
- WELDING CYCLE
- The complete series of events involved in the making of a weld.
- WELDING ELECTRODE
- A component of the welding circuit through which current is conducted and that terminates at the arc, molten conductive slag, or base metal. See also arc welding electrode , flux cored electrode , metal cored electrode , metal electrode , resistance welding electrode , and stranded electrode .
- WELDING GROUND
- A nonstandard term for workpiece connection .
- WELDING LEADS
- The workpiece lead and electrode lead of an arc welding circuit.
- WELDING MACHINE
- Equipment used to perform the welding operation; for example, spot welding machine, arc welding machine, and seam welding machine.
- WELDING OPERATOR
- One who operates machine or automatic welding equipment.
- WELDING POSITION
- See flat position , horizontal fixed position , horizontal position, horizontal rolled position , inclined position , overhead position , and vertical position .
- WELDING PROCEDURE

- The detailed methods and practices involved in the production of a weldment. See also welding procedure specification .
- WELDING PROCEDURE SPECIFICATION (WPS)
- A document providing in detail the required variables for a specific application to ensure repeatability by properly trained welders and welding operators.
- WELDING ROD
- A form of welding filler metal, normally packaged in straight lengths, that does not conduct electrical current.
- WELDING SEQUENCE
- The order of making the welds in a weldment.
- WELDING TIP
- That part of an oxyfuel gas welding torch from which the gases issue.
- WELDING TORCH (ARC)
- A device used in the gas tungsten and plasma arc welding processes to control the position of the electrode, to transfer current to the arc, and to direct the flow of shielding and plasma gas.
- WELDING TORCH (OXYFUEL GAS)
- A device used in oxyfuel gas welding, torch brazing, and torch soldering for directing the heating flame produced by the controlled combustion of fuel gases.
- WELDING WHEEL
- A nonstandard term for resistance welding electrode .
- WELDING WIRE
- A form of welding filler metal, normally packaged as coils or spools, that may or may not conduct electrical current depending upon the welding process with which it is used. See also welding electrode and welding rod .
- WELD INTERFACE
- The interface between weld metal and base metal in a fusion weld, between base metals in a solid-state weld without filler metal or between filler metal and base metal in a solid-state weld with filler metal, and in a braze.
- WELD LINE
- A nonstandard term for weld interface .
- WELDMENT
- An assembly whose component parts are joined by welding.
- WELD METAL
- That portion of a weld that has been melted during welding.
- WELDOR
- A nonstandard term for welder .
- WELD PASS
- A single progression of welding or surfacing along a joint or substrate. The result of a pass is a weld bead, layer, or spray deposit.
- WELD PASS SEQUENCE
- The order in which the weld passes are made.
- WELD PENETRATION
- A nonstandard term for joint penetration and root penetration .
- WELD POOL
- The localized volume of molten metal in a weld prior to its solidification as weld metal.
- WELD PUDDLE
- A nonstandard term for weld pool .
- WELD REINFORCEMENT
- Weld metal in excess of the quantity required to fill a joint. See also face reinforcement and root reinforcement .
- WELD ROOT
- The points, as shown in cross section, at which the back of the weld intersects the base metal surfaces.
- WELD SIZE
- See edge weld size , fillet weld size , flange weld size , and groove weld size .
- WELD TAB
- Additional material on which the weld may be initiated or terminated.
- WELD THROAT

- See actual throat , effective throat , and theoretical throat .
- WELD TOE
- The junction of the weld face and the base metal.
- WETTING
- The phenomenon whereby a liquid filler metal or flux spreads and adheres in a thin continuous layer on a solid base metal.
- WIPED JOINT
- A joint made with solder having a wide melting range and with the heat supplied by the molten solder poured onto the joint. The solder is manipulated with a hand-held cloth or paddle so as to obtain the required size and contour.
- WORK CONNECTION
- A nonstandard term for workpiece connection .
- WORK LEAD
- A nonstandard term for workpiece lead .
- WORKPIECE
- The part being welded, brazed, soldered, or cut.
- WORKPIECE CONNECTION
- The connection of the workpiece lead to the workpiece.
- WORKPIECE LEAD
- The electrical conductor between the arc welding current source and the workpiece connection.

Note

* THIS GLOSSARY IS REPRINTED BY PERMISSION OF THE AMERICAN WELDING SOCIETY FROM THE *WELDING HANDBOOK*, 8TH EDITION, VOLUME 1, AMERICAN WELDING SOCIETY, 1987. IT WAS PREPARED BY THE AWS COMMITTEE ON DEFINITIONS AND SYMBOLS CONSISTING OF G.E. METGER, *CHAIRMAN*, AFWAL/MLLS; J.T. BISKUP, CANADIAN WELDING BUREAU; W.F. BROWN, WESTINGHOUSE HANFORD; C.D. BURNHAM, CONSULTANT; R.J. CHRISTOFFEL, GENERAL ELECTRIC COMPANY; G.B. COATES, GENERAL ELECTRIC COMPANY; MICHAEL D. COOPER, HOBART SCHOOL OF WELDING TECHNOLOGY; W.L. GREEN, OHIO STATE UNIVERSITY; J.E. GREER, GENERAL ELECTRIC RAILROAD; M.J. GRYCO, PACKER ENGINEERING ASSOCIATES; D.E. HAMILTON, TRINITY ENGINEERING TESTING CORPORATION; E.A. HARWART, CONSULTANT; A. RAY HOLLINS, JR., DUKE POWER COMPANY; M.J. HOULE, NATIONAL BOARD OF BOILER AND PRESSURE VESSEL INSPECTORS; STEPHEN R. MORSE, DEERE AND COMPANY; L.C. NORTHARD, TENNESSEE VALLEY AUTHORITY; D.H. ORTS, ARMCO, INCORPORATED; E.J. SEEL, AMERICAN WELDING SOCIETY; J.J. STANCZAK, STEEL DETAILERS & DESIGNERS; J.J. VAGI, BABCOCK AND WILCOX; J.L. WILK, COAST TO COAST CONSTRUCTION COMPANY; UNDER THE DIRECTION OF AWS WELDING HANDBOOK COMMITTEE CHAIRMEN J.H. HANNAHS, MIDMARK CORPORATION, AND M.J. TOMSIC, PLASTRONICS INCORPORATED (AFFILIATIONS CIRCA 1987).

Abbreviations, Symbols, and Tradenames

- ABBREVIATIONS AND SYMBOLS
- A
- austenite; ampere
- Å
- angstrom
- AA
- Aluminum Association
- AAR

- Association of American Railroads
- AASHTO
- American Association of State Highway and Transportation Officials
- ABS
- American Bureau of Shipping
- AC
- alternating current
- AC
- air cool; aligned carbide
- $\mathrm{AC_{CM}}$
- in hypereutectoid steel, temperature at which cementite completes solution in austenite
- $\mathrm{AC_1}$
- temperature at which austenite begins to form during heating
- $\mathrm{AC_3}$
- temperature at which transformation of ferrite to austenite is completed during heating
- $\mathrm{AE_{CM}}$, $\mathrm{AE_1}$, $\mathrm{AE_3}$
- equilibrium transformation temperatures in steel
- AES
- Auger electron spectroscopy; acoustic emission spectroscopy
- AF
- acicular ferrite
- AFM
- atomic force microscopy
- AI
- artificial intelligence
- AISC
- American Institute of Steel Construction
- AISI
- American Iron and Steel Institute
- AMS
- Aerospace Material Specification
- ANSI
- American National Standards Institute
- AOD
- argon-oxygen decarburization
- API
- American Petroleum Institute
- $\mathrm{AR_{CM}}$
- temperature at which cementite begins to precipitate from austenite upon cooling
- $\mathrm{AR_1}$
- temperature at which transformation to ferrite or to ferrite plus cementite is completed on cooling
- $\mathrm{AR_3}$
- temperature at which transformation of austenite to ferrite begins on cooling
- AREA
- American Railway Engineers Association
- ASM
- American Society for Metals (now ASM International)
- ASME
- American Society of Mechanical Engineers
- ASTM
- American Society for Testing and Materials
- AT.%
- atomic percent
- ATM
- atmospheres (pressure)
- AWS
- American Welding Society

- B
- bainite
- BAL
- balance
- BCC
- body-centered cubic
- BCT
- body-centered tetragonal
- C
- edge length in crystal structure; speed of light; specific heat; constant
- C
- cementite; coulomb; heat capacity
- CAC
- carbon arc cutting
- CAC-A
- air carbon arc cutting
- CAD/CAM
- computer-aided design/computer-aided manufacturing
- CAFE
- Corporate Average Fuel Economy (standard)
- CAW
- carbon arc welding
- CAW-G
- gas carbon arc welding
- CAW-S
- shielded carbon arc welding
- CBN
- cubic boron nitride
- CCD
- charge-coupled device
- CCT
- continuous-cooling transformation
- CD
- candela
- CE
- carbon equivalent
- CGS
- centimeter-gram-second (system of units)
- CM
- centimeter
- CMC
- ceramic-matrix composite
- CNC
- computer numerical control
- COD
- crack-opening displacement
- C_P
- specific heat at constant pressure
- CP
- centipoise
- CP
- commercially pure
- CPT
- critical pitting temperature
- CR
- corrosion resistant
- CRT

- cathode-ray tube
- CST
- centistokes
- CTE
- coefficient of thermal expansion
- CTOD
- crack tip opening displacement
- CVD
- chemical vapor deposition
- CVN
- Charpy V-notch (impact test or specimen)
- *D*
- diameter
- *DA*/*DN*
- fatigue crack growth rate
- DARPA
- Defense Advanced Research Projects Agency
- DB
- decibel
- DB
- diffusion bonding; dip brazing
- DBTT
- ductile-to-brittle transition temperature
- DC
- direct current
- DCEN
- direct current, electrode negative
- DCEP
- direct current, electrode positive
- DFB
- diffusion brazing
- DFW
- diffusion welding
- DHC
- delayed hydride cracking
- DIAM
- diameter
- DIN
- Deutsche Industrie-Normen (German Industrial Standards)
- DPH
- diamond pyramid hardness
- DRA
- discontinuously reinforced aluminum
- DS
- dip soldering
- DTA
- differential thermal analysis
- *E*
- natural log base, 2.71828; charge of an electron
- EB
- electron beam
- EBC
- electron-beam cutting
- EBW
- electron-beam welding
- EDM
- electrical discharge machining

- EDXA
- energy-dispersive x-ray analysis
- EELS
- electron energy loss spectroscopy
- EERC
- Earthquake Engineering Research Center
- EGW
- electrogas welding
- ELI
- extra-low interstitial
- EMF
- electromotive force
- EPA
- Environmental Protection Agency
- EPFM
- elastic plastic fracture mechanics
- EPRI
- Electric Power Research Institute
- EQ
- equation
- ESW
- electroslag welding
- *ET AL.*
- and others
- ETP
- electrolytic tough pitch
- EV
- electron volt
- EXB
- exothermic brazing
- EXP
- base of the natural logarithm
- EXW
- explosion welding
- *F*
- frequency
- F
- ferrite
- FAA
- Federal Aviation Administration
- FAD
- failure-assessment diagram; fracture-assessment diagram
- FB
- furnace brazing
- FCAW
- flux-cored arc welding
- FCC
- face-centered cubic
- FCM
- fracture critical tension member
- FCT
- face-centered tetragonal
- FEA
- finite-element analysis
- FEM
- finite-element method
- FHA

- Federal Highway Administration
- FIG.
- figure
- FM
- frequency modulation
- FMVSS
- Federal Motor Vehicle Safety Standards
- FN
- ferrite number
- FRW
- friction welding
- FS
- furnace soldering
- FT
- foot
- FW
- flash welding
- G
- gram; acceleration due to gravity
- G
- graphite; gauss
- GAL
- gallon
- GBF
- grain-boundary ferrite
- GF
- gram force
- GMA
- gas-metal arc
- GMAC
- gas-metal arc cutting
- GMAW
- gas-metal arc welding
- GPA
- gigapascal
- GTA
- gas-tungsten arc
- GTAC
- gas-tungsten arc cutting
- GTAW
- gas-tungsten arc welding
- H
- hour; Planck's constant (6.626×10^{-27} erg · s)
- H
- Henry
- *H*
- enthalpy; hardness; magnetic field
- HAZ
- heat-affected zone
- HB
- Brinell hardness
- HCP
- hexagonal close-packed
- HHC
- hot-hollow cathode
- HIC
- hydrogen-induced cracking

- HIP
- hot isostatic pressing
- HK
- Knoop hardness
- HP
- horsepower
- HR
- Rockwell hardness (requires scale designation, such as HRC for Rockwell C hardness)
- HREM
- high-resolution electron microscopy
- HSLA
- high-strength low-alloy (steel)
- HSS
- high-speed steel(s)
- HTLA
- heat-treatable low-alloy (steel)
- HV
- Vickers hardness
- HZ
- hertz
- *I*
- intensity; electrical current
- IA
- inorganic acid
- ID
- inside diameter
- IGA
- intergranular attack
- IGSCC
- intergranular stress-corrosion cracking
- IIW
- International Institute of Welding
- I/M
- ingot metallurgy
- IMO
- International Maritime Organization
- IN.
- inch
- IPTS
- International Practical Temperature Scale
- IR
- infrared
- IRS
- infrared soldering
- IS
- induction soldering
- ISCC
- intergranular stress-corrosion cracking
- ISO
- International Organization for Standardization
- J
- joule
- K
- equilibrium distribution coefficient; thermal conductivity; Boltzmann constant
- K
- Kelvin
- KG

- kilogram
- KGF
- kilogram force
- K_{IC}
- plane-strain fracture toughness
- K_{ISCC}
- threshold stress intensity to produce stress-corrosion cracking
- KM
- kilometer
- KN
- kilonewton
- KPA
- kilopascal
- KSI
- kips (1000 lbf) per square inch
- KV
- kilovolt
- KW
- kilowatt
- L
- longitudinal, liter
- LB
- pound
- LBF
- pound force
- LBW
- laser-beam welding
- LCL
- lower control limit
- LEED
- low-energy electron diffraction
- LEFM
- linear elastic fracture mechanics
- LME
- liquid-metal embrittlement
- LMP
- Larson-Miller parameter
- IN
- natural logarithm (base *e*)
- LNG
- liquefied natural gas
- LOG
- common logarithm (base 10)
- LPG
- liquefied petroleum gas
- LSI
- large-scale integrated (circuit)
- LT
- long transverse (direction)
- LTE
- local thermodynamic equilibrium
- LTS
- low-temperature sensitization
- LVDT
- linear variable differential transformer
- M
- meter

- MA
- milliampere
- MC
- metal carbide
- MCL
- maximum crack length
- MEV
- megaelectronvolt
- M_F
- temperature at which martensite formation finishes during cooling
- MG
- milligram
- MG
- megagram (metric tonne, or kg × 10^3)
- MIC
- microbiologically influenced corrosion
- MIN
- minute; minimum
- ML
- milliliter
- MM
- millimeter
- MO
- multiple operator
- MPA
- millipascal
- MPA
- megapascal
- MPG
- miles per gallon
- MPH
- miles per hour
- MRDA
- man-rated demonstration vehicle
- MRR
- material removal rate
- MS
- millisecond
- M_S
- temperature at which martensite starts to form from austenite on cooling
- MS
- megasiemens
- MT
- millitesla
- MV
- millivolt
- MV
- megavolt
- MVC
- most vulnerable component
- N
- newton
- N
- number of cycles; normal solution; normal force
- NACE
- National Association of Corrosion Engineers
- NASA

- National Aeronautics and Space Administration
- NBS
- National Bureau of Standards (former name of NIST)
- NDE
- nondestructive examination
- NDE
- nondestructive evaluation
- NEC
- National Electric Code
- NEMA
- National Electrical Manufacturers Association
- NIST
- National Institute of Standards and Technology
- NM
- nanometer
- NMMA
- National Marine Manufacturers Association
- NNS
- neural network system
- NO.
- number
- NS
- nanosecond
- OA
- organic acid
- OAW
- oxyacetylene welding
- OCV
- open-circuit voltage
- OE
- oersted
- OFG
- oxyfuel gas
- OFW
- oxyfuel gas welding
- ORNL
- Oak Ridge National Laboratory
- OSHA
- Occupational Safety and Health Administration
- OZ
- ounce
- P
- page
- P
- pearlite
- PA
- pascal
- PAW
- plasma arc welding
- PEEM
- photoemission electron microscopy
- PH
- negative logarithm of hydrogen-ion activity
- PH
- precipitation hardenable; precipitation hardening
- PID
- proportional-integral-differential

- PM
- planar-magnetron
- PPB
- parts per billion
- PPM
- parts per million
- PPT
- parts per trillion
- PRE
- pitting resistance equivalent
- PRR
- pulse repetition rates
- PSI
- pounds per square inch
- PSIA
- pounds per square inch absolute
- PSIG
- gage pressure (pressure relative to ambient pressure) in pounds per square inch
- PSTM
- photon scanning tunneling microscopy
- PW
- projection welding
- PWAA
- postweld artificial aging
- PWB
- printed wiring board
- PWHT
- postweld heat treatment
- PWM
- pulse width modulation
- Q&T
- quenched and tempered
- R
- roentgen
- R_A
- surface roughness in terms of arithmetic average
- RA
- rosin fully activated; roughness average
- *RA*
- reduction in area
- RE
- rare earth
- REF
- reference
- REF
- relative erosion factor
- RF
- radio frequency
- RH
- relative humidity
- RMA
- rosin mildly activated
- RMS
- root mean square
- RPF
- relative potency factor
- RPM

- revolutions per minute
- R_Q
- rms (root mean square) roughness
- RS
- resistance soldering
- RSA
- rosin superactivated
- RSEW
- seam welding
- RSW
- resistance spot welding
- RT
- room temperature
- RW
- resistance welding
- RWMA
- Resistance Welder Manufacturers Association
- R_Y
- maximum peak-to-valley height
- R_Z
- ten-point height
- S
- second
- S
- siemens
- SA
- synthetic activated
- SACP
- selected-area channeling patterns
- SAE
- Society of Automotive Engineers
- SAM
- scanning acoustic microscope/microscopy
- SAW
- submerged arc welding
- SCC
- stress-corrosion cracking
- SCPM
- scanning chemical potential microscopy
- SCR
- silicon-controlled rectifier
- SEM
- scanning electron microscope/microscopy
- SFM
- surface feet per minute
- SHE
- standard hydrogen electrode
- SI
- Système International d'Unités
- SIR
- surface insulation resistance
- SLAM
- scanning laser acoustic microscope/microscopy
- SLEEM
- scanning low-energy electron microscopy
- SMA
- shielded metal arc

- SMAW
- shielded metal arc welding
- SPA-LEED
- spot profile analysis-low-energy electron diffraction
- SPC
- statistical process control
- SPF
- sideplate ferrite
- SQC
- statistical quality control
- SRA
- stress-relief annealing
- SSW
- solid-state welding
- STA
- solution-treated and aged
- STEM
- scanning transmission electron microscope/microscopy
- STM/S
- scanning tunneling microscopy/spectroscopy
- SUS
- Saybolt Universal Seconds (viscosity)
- SV
- sievert
- SW
- stud arc welding
- T
- thickness; time
- T
- tesla
- *T*
- temperature
- TB
- torch brazing
- TCP
- topologically close packed
- TEM
- transmission electron microscope/microscopy
- TGA
- thermogravimetric analysis
- TGSCC
- transgranular stress-corrosion cracking
- TIG
- tungsten inert gas (welding)
- TMCP
- thermomechanically controlled process
- TS
- torch soldering
- TTT
- time-temperature transformation
- TW
- thermit welding; thermite welding
- TZP
- tetragonal zirconia polycrystals
- UCL
- upper control limit
- UNS

- Unified Numbering System
- USW
- ultrasonic welding
- UTS
- ultimate tensile strength
- UW
- upset welding
- V
- velocity
- V
- volt
- VIM-VAR
- vacuum induction melted-vacuum arc remelted
- VOL
- volume
- VOL%
- volume percent
- VPPA
- variable polarity plasma arc
- W
- watt
- WI
- wettability index
- WRC
- Welding Research Council
- WS
- wave soldering
- WSN
- weld set number
- WT%
- weight percent
- XPS
- x-ray photoelectron spectroscopy
- YR
- year
- *Z*
- atomic number
- °
- angular measure; degree
- °C
- degree Celsius (centigrade)
- °F
- degree Fahrenheit
- **€**
- direction of reaction
- ÷
- divided by
- =
- equals
- ≅
- approximately equals
- ≠
- not equal to
- ≡
- identical with
- >
- greater than

- **?**
- much greater than
- [GES]
- greater than or equal to
- ∞
- infinity
- $\propto$
- is proportional to; varies as
- $\int$
- integral of
- $<$
- less than
- =
- much less than
- $\leq$
- less than or equal to
- $\pm$
- maximum deviation
- -
- minus; negative ion charge
- $\times$
- diameters (magnification); multiplied by
- $\cdot$
- multiplied by
- /
- per
- %
- percent
- $+$
- plus; positive ion charge
- $\sqrt{}$
- square root of
- $\sim$
- approximately; similar to
- ∂
- partial derivative
- α
- Stefan-Boltzmann constant
- γ
- shear strain rate
- γ
- $_E$
- elastic strain
- γ_T
- total strain
- Δ
- change in quantity; an increment; a range
- ε
- strain
- $\dot{e}$
- strain rate
- θ
- angle
- λ
- wavelength
- μ
- friction coefficient; magnetic permeability

- μ_S
- static coefficient of friction
- μ F
- microfarads
- μ IN.
- microinch
- μ M
- micrometer (micron)
- μS
- microsecond
- *n*
- Poisson's ratio
- π
- pi (3.141592)
-
- density
- σ
- stress; standard deviation
- Σ
- summation of
- *t*
- applied stress
- Ω
- angular velocity; frequency
- Ω
- ohm

- GREEK ALPHABET

- **GREEK ALPHABET**
- **A, A**
  - ALPHA
- **B, B**
  - BETA
- Γ, **Γ**
  - GAMMA
- Δ, **Δ**
  - DELTA
- **E, E**
  - EPSILON
- **Z,** *z*
  - ZETA
- **H, H**
  - ETA
- Θ, **Θ**
  - THETA
- **I,** *i*
  - IOTA
- **K, K**
  - KAPPA
- Λ, **LAMBDA;**
  - LAMBDA
- **M, M**
  - MU
- **N,** *u*
  - NU

- Ξ, *x*
    - XI
- **Ο, Ο**
    - OMICRON
- Π, **Π**
    - PI
- **Ρ, Ρ**
    - RHO
- Σ, **Σ**
    - SIGMA
- **Τ,** *t*
    - TAU
- ϒ, *u*
    - UPSILON
- Φ, *f*
    - PHI
- **Χ, Χ**
    - CHI
- **Ψ,**
    - PSI
- **Ω, Ω**
    - OMEGA

- TRADENAMES
- ACTIVATED DIFFUSION BONDING
- is a registered trademark of General Electric Company.
- ACTIVATED DIFFUSION HEALING
- is a trademark of General Electric Company.
- AD-85
- is a trademark of Coors.
- AD-988
- is a tradename of McDanel.
- ALLOY 52
- is a trademark of Carpenter Technology Corporation.
- ALLOY 42
- is a trademark of Carpenter Technology Corporation.
- BETA C
- is a tradename of RMI, Inc.
- CORONA 5
- is a tradename of Crucible, Inc.
- CUSIL ABA
- is a tradename of GTE Wesgo.
- FM73
- is a tradename of American Cyanimide Company.
- FM300K
- is a tradename of American Cyanimide Company.
- HAYNES AND HASTELLOY
- are registered trademarks of Haynes International, Inc.
- HEXOLOY SA
- is a tradename of Carborundum.
- 230 AND HR160
- are trademarks of Haynes International, Inc.
- INCOLOY 909
- is a tradename of Inco Alloys International, Inc.
- INCONEL AND NIMONIC

- registered trademarks of Inco Family of Companies.
- KOVAR
- is a trademark of Carpenter Technology Corporation.
- MAR-M
- is a trademark of Martin Marietta.
- MP35N AND MP159
- are registered trademarks of SPS Technologies.
- MULTIPHASE
- is a registered trademark of SPS Technologies.
- NICHROME
- is a registered trademark of Harris Co., Inc.
- NICROBRAZ
- is a tradename of Wall Colmonoy Corporation.
- NOCOLOK
- is a tradename of Alcan Aluminum.
- PAD-AS30W
- is a tradename of Cercom.
- POCO AXF-5Q
- is a tradename of Unocal Chemicals Div./Nitrogen Group.
- PYROMET
- is a registered trademark of Carpenter Technology.
- RA 333
- is a registered trademark of Rolled Alloys, Inc.
- RENÉ
- is a registered trademark of Teledyne Industries, Inc.
- SNW-1000
- is a registered trademark of GTE.
- THRESHOLD LIMIT VALUE
- is a registered trademark of the American Conference of Governmental and Industrial Hygienists.
- TICUNI
- is a registered tradename of GTE Wesgo.
- TRANSIENT LIQUID-PHASE BONDING
- is a trademark of Pratt & Whitney.
- UDIMET
- is a registered trademark of Special Metals, Inc.
- ULTEM
- is a tradename of General Electric Company.
- ULTIMET
- is a registered trademark of Haynes International, Inc.
- WASPALOY
- is a trademark of United Technologies.
- WELDALITE
- is a registered trademark of Martin Marietta Corporation.

WELDING

Skills, Processes and Practices for Entry-Level Welders

BOOK 2

- Gas Metal Arc Welding
- Flux Cored Arc Welding

LARRY JEFFUS • LAWRENCE BOWER

Welding

Skills, Processes and Practices for Entry-Level Welders

Book 2

- Gas Metal Arc Welding
- Flux Cored Arc Welding

First Edition

Larry Jeffus
Lawrence Bower

DELMAR
CENGAGE Learning™

Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

DELMAR
CENGAGE Learning™

Welding: Skills, Processes and Practices for Entry-Level Welders: Book Two
Larry Jeffus/Lawrence Bower

Vice President, Editorial: Dave Garza
Director of Learning Solutions: Sandy Clark
Executive Editor: David Boelio
Managing Editor: Larry Main
Senior Product Manager: Sharon Chambliss
Editorial Assistant: Lauren Stone
Vice President, Marketing: Jennifer McAvey
Executive Marketing Manger: Deborah S. Yarnell
Senior Marketing Manager: Jimmy Stephens
Marketing Specialist: Mark Pierro
Production Director: Wendy Troeger
Production Manager: Mark Bernard
Content Project Manager: Cheri Plasse
Art Director: Benj Gleeksman
Technology Project Manager: Christopher Catalina
Production Technology Analyst: Thomas Stover

© 2010 Delmar, Cengage Learning

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored, or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, Web distribution, information networks, or information storage and retrieval systems, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the publisher.

For product information and technology assistance, contact us at
Professional & Career Group Customer Support,
1-800-648-7450

For permission to use material from this text or product,
submit a request online at **cengage.com/permissions.**
Further permissions questions can be e-mailed to
permissionrequest@cengage.com.

Library of Congress Control Number: 2008911231

ISBN-13: 978-1-4354-2790-7
ISBN-10: 1-4354-2790-4

Delmar
5 Maxwell Drive
Clifton Park, NY, 12065-2919
USA

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

For your lifelong learning solutions, visit **delmar.cengage.com**
Visit our corporate website at **cengage.com.**

Notice to the Reader
Publisher does not warrant or guarantee any of the products described herein or perform any independent analysis in connection with any of the product information contained herein. Publisher does not assume, and expressly disclaims, any obligation to obtain and include information other than that provided to it by the manufacturer. The reader is expressly warned to consider and adopt all safety precautions that might be indicated by the activities described herein and to avoid all potential hazards. By following the instructions contained herein, the reader willingly assumes all risks in connection with such instructions. The publisher makes no representations or warranties of any kind, including but not limited to, the warranties of fitness for particular purpose or merchantability, nor are any such representations implied with respect to the material set forth herein, and the publisher takes no responsibility with respect to such material. The publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or part, from the readers' use of, or reliance upon, this material.

Printed in the United States of America
1 2 3 4 5 XX 11 10 09

Brief Contents

Preface viii

AWS Acknowledgment xi

Acknowledgments xii

About the Authors xiii

Chapter 1 Gas Metal Arc Welding Equipment, Setup, and Operation 1

Chapter 2 Gas Metal Arc Welding 55

Chapter 3 Flux Cored Arc Welding Equipment, Setup, and Operation 113

Chapter 4 Flux Cored Arc Welding 143

Appendix 197

Glossary 199

Index 203

This page intentionally left blank

Contents

Preface viii

AWS Acknowledgment xi

Acknowledgments xii

About the Authors xiii

Chapter 1 Gas Metal Arc Welding Equipment, Setup, and Operation 1

Introduction 2
Metal Transfer 4
  Axial Spray Metal Transfer 5
  Globular Transfer 7
  Pulsed-Spray Transfer: GMAW-P 7
  Shielding Gases for Spray or Pulsed-spray Transfer 9
  Buried-arc Transfer 9
  Short-circuiting Transfer: GMAW-S 11
Filler Metal Specifications 13
  Filler Metal Selection 14
  Stainless Steel Electrodes 15
Aluminum and Aluminum Alloys 15
  Aluminum Bare Welding Rods and Electrodes. . . . 20
Wire Melting and Deposition Rates. 20
Welding Power Supplies 21
  Speed of the Wire Electrode 23
  Power Supplies for Short-circuiting Transfer 23
  Inductance 25
Molten Weld Pool Control. 25
  Shielding Gas 26
  Power Settings 29
  Gun Manipulation. 30
  Travel Speed. 30
  Electrode Extension. 30
  Gun Angle 31
Equipment 33
  Power Source 34
  Electrode (Wire) Feed Unit. 37
  Push-type Feed System 37

- Pull-type Feed System 38
- Push-pull-type Feed System 39
- Linear Electrode Feed System 39
- Spool Gun 40
- Electrode Conduit 41
- Welding Gun 41

Spot Welding 42

Face, Eye, and Ear Protection 44
- Face and Eye Protection. 44
- Ear Protection. 48

General Work Clothing. 49

Special Protective Clothing 51
- Hand Protection 51
- Body Protection 52
- Waist and Lap Protection 52
- Arm Protection 52
- Leg and Foot Protection. 52

Chapter 2 Gas Metal Arc Welding 55

Introduction 56

Setup. 57

Gas Density and Flow Rates 65

Arc-voltage and Amperage Characteristics 68

Electrode Extension 70

Welding Gun Angle 73

Effect of Shielding Gas on Welding 74

Practices. 77

Metal Preparation 78

Flat Position, 1G and 1F Positions 79

Vertical Up 3G and 3F Positions 84

Vertical Down 3G and 3F Positions 86

Horizontal 2G and 2F Positions 89

Overhead 4G and 4F Positions. 91

Axial Spray Transfer 94

Chapter 3 Flux Cored Arc Welding Equipment, Setup, and Operation 113

Introduction 114

Principles of Operation. 114

Equipment 117
- Power Supply 117
- Guns 117
- Smoke Extraction Nozzles. 119
- Electrode Feed 119

Advantages 119
High Deposition Rate. 119
Minimum Electrode Waste 119
Narrow Groove Angle 120
Minimum Precleaning 120
All-position Welding 120
Flexibility 120
High Quality. 120
Excellent Control. 121
Limitations 121
Electrodes 121
Methods of Manufacturing 121
Electrode Cast and Helix 124
Flux. 125
Flux Cored Steel Electrode Identification 128
Metal Cored Steel Electrode Identification 130
Care of Flux Core Electrodes. 130
Shielding Gas 131
Welding Techniques 132
Gun Angle 132
Forehand/Perpendicular/Backhand Techniques. . 134
Travel Speed. 136
Mode of Metal Transfer 137
Electrode Extension 138
Porosity 138

Chapter 4 Flux Cored Arc Welding. 143

Introduction 144
Practices. 145
Flat-position Welds 151
Square-groove Welds 152
V-Groove and Bevel-Groove Welds 154
Root Pass 155
Filler Pass 156
Cover Pass 157
Fillet Welds. 160
Vertical Welds 164
Horizontal Welds. 170
Overhead-position Welds 175
Thin-gauge Welding. 180

Appendix 197

Glossary 199

Index. 203

Preface

ABOUT THE SERIES

Welding: Skills, Processes and Practices for Entry-Level Welders is an exciting new series that has been designed specifically to support the American Welding Society's (AWS) SENSE EG2.0 training guidelines. Offered in three volumes, these books are carefully crafted learning tools consisting of theory-based texts that are accompanied by companion lab manuals, and extensive instructor support materials. With a logical organization that closely follows the modular structure of the AWS guidelines, the series will guide readers through the process of acquiring and practicing welding knowledge and skills. For schools already in the SENSE program, for those planning to join, or for schools interested in obtaining certifiable outcomes based on nationally recognized industry standards order to comply with the latest Carl D. Perkins Career and Technical Education in requirements, *Welding: Skills, Processes and Practices for Entry-Level Welders* offers a turnkey solution of high quality teaching and learning aids.

Career and technical education instructors at the high school level are often called upon to be multi-disciplinary educators, teaching welding as only one of as many as five technical disciplines in any given semester. The *Welding: Skills, Processes and Practices for Entry-Level Welders* package provides these educators with a process-based, structured approach and the tools they need to be prepared to deliver high level training on processes and materials with which they may have limited familiarity or experience. Student learning, satisfaction and retention are the target of the logically planned practices, supplements and full color textbook illustrations. While the AWS standards for entry level welders are covered, students are also introduced to the latest in high technology welding equipment such as pulsed gas metal arc welding (GMAW-P). Career pathways and career clusters may be enhanced by the relevant mathematics applied to real world activities as well as oral and written communication skills linked to student interaction and reporting.

Book 1, the core volume, introduces students to the welding concepts covered in AWS SENSE Modules 1, 2, 3, 8 and 9 (Occupational Orientation, Safety and Health of Welders, Drawing and Welding Symbol Interpretation, Thermal Cutting, and Weld Inspection Testing and Codes). Book 1 contains all the material needed for a SENSE program that prepares students for qualification in Thermal Cutting processes. The optional Books 2 and 3 cover other important welding processes and are grouped in logical combinations. Book 2 corresponds to AWS SENSE Modules 5 and 6 (GMAW, FCAW), and Book 3 corresponds to AWS SENSE Modules 4 and 7 (SMAW, GTAW).

The texts feature hundreds of four-color figures, diagrams and tight shots of actual welds to speed beginners to an understanding of the most widely used welding processes.

FEATURES

- Produced in close collaboration with experienced instructors from established SENSE programs to maximize the alignment of the content with SENSE guidelines and to ensure 100% coverage of Level I-Entry Welder Key Indicators.
- Chapter introductions contain general performance objectives, key terms used, and the AWS SENSE EG2.0 Key Indicators addressed in the chapter.
- Coverage of Key Indicators is indicated in the margin by a torch symbol and a numerical reference.
- Contains scores of fully illustrated Practices, which are guided exercises designed to help students master processes and materials. Where applicable, the Practices reproduce and reference actual AWS technical drawings in order to help students create acceptable workmanship samples.
- Each section introduces students to the materials, equipment, setup procedures and critical safety information they need in order to weld successfully.
- Hundreds of four-color figures, diagrams and tight shots of actual welds to speed beginners to an understanding of the most widely used welding processes.
- End of chapter review questions develop critical thinking skills and help students to understand “why” as well as “how.”

SUPPLEMENTS

Each book in the Welding Skills series is accompanied by a ***Lab Manual*** that has been designed to provide hands-on practice and reinforce the student’s understanding of the concepts presented in the text. Each chapter contains practice exercises to reinforce the primary objectives of the lesson, including creation of workmanship samples (where applicable), and a quiz to test knowledge of the material. Artwork and safety precautions are included throughout the manuals.

Instructor Resources (on CD-ROM), designed to support Books 1–3 and the accompanying Lab Manuals, provide a wealth of time-saving tools, including:

- An Instructor’s Guide with answers to end-of-chapter Review Questions in the texts and Lab Manual quizzes.
- Modifiable model Lesson Plans that aid in the design of a course of study that meets local or state standards and also maps to the SENSE guidelines.
- An extensive ExamView Computerized Test Bank that offers assessments in true/false, multiple choice, sentence completion and short answer formats. Test questions have been designed to expose students to the types of questions they’ll encounter on the SENSE Level 1 Exams.
- PowerPoint Presentations with selected illustrations that provide a springboard for lectures and reinforce skills and processes covered in the texts. The PowerPoint Presentations can be modified or expanded as instructors desire, and can be augmented with additional illustrations from the Image Library.
- The Image Library contains nearly all (well over 1000!) photographs and line art from the texts, most in four-color.
- A SENSE Correlation Chart that shows the close alignment of the *Welding* series to the SENSE Entry Level 1 training guidelines. Each Key Indicator within each SENSE Module is mapped to the relevant text and lab manual page or pages.

TITLES IN THE SERIES

Welding: Skills, Processes and Practices for Entry-Level Welders: Book 1, Occupational Orientation, Safety and Health of Welders, Drawing and Welding Symbol Interpretation, Thermal Cutting, Weld Inspection Testing and Codes
(Order #: 1-4354-2788-2)
Lab Manual, Book One (Order #: 1-4354-2789-0)

Welding: Skills, Processes and Practices for Entry-Level Welders: Book 2, Gas Metal Arc Welding, Flux Cored Arc Welding (Order #:1-4354-2790-4)
Lab Manual, Book Two (Order #: 1-4354-2795-5)

Welding: Skills, Processes and Practices for Entry-Level Welders: Book 3, Shielded Metal Arc Welding, Gas Tungsten Arc Welding (Order #:1-4354-2796-3)
Lab Manual, Book Three (Order #: 1-4354-2797-1)

AWS Acknowledgment

The Authors and Publisher gratefully acknowledge the support provided by the American Welding Society in the development and publication of this textbook series. “American Welding Society,” the AWS logo and the SENSE logo are the trade and service marks of the American Welding Society and are used with permission.

For more information on the American Welding Society and the SENSE program, visit **http://www.aws.org/education/sense/** or contact AWS at (800) 443-9353 ext. 455 or by email: **education@aws.org**.

Acknowledgments

The authors and publisher would like to thank the following individuals for their contributions to this series:

Garey Bish, *Gwinnett Technical College, Lawrenceville, GA*
Julius Blair, *Greenup County Area Technology Center, Greenup, KY*
Rick Brandon, *Pemiscot County Career & Technical Center, Hayti, MO*
Stephen Brandow, *University of Alaska Southeast, Ketchikan, Ketchikan, AK*
Francis X Brieden, *Career Technology Center of Lackawanna County, Scranton, PA*
John Cavenaugh, *Community College of Southern Nevada, Las Vegas, NV*
Clay Corey, *Washington-Saratoga BOCES, Fort Edward, NY*
Keith Cusey, *Institute for Construction Education, Decatur, IL*
Craig Donnell, *Whitmer Career Technology Center, Toledo, OH*
Steve Farnsworth, *Iowa Lakes Community College, Emmetsburg, IA*
Ed Harrell, *Traviss Career Center, Lakeland, FL*
Robert Hoting, *Northeast Iowa Community College, Sheldon, IA*
Steve Kistler, *Moberly Area Technical Center, Moberly, MO*
David Lynn, *Lebanon Technology & Career Center, Lebanon, MO*
Frank Miller, *Gadsden State Community College, Gadsden, AL*
Chris Overfelt, *Arnold R Burton Tech Center, Salem, VA*
Kenric Sorenson, *Western Technical College, LaCrosse, WI*
Pete Stracener, *South Plains College, Levelland, TX*
Bill Troutman, *Akron Public Schools, Akron, OH*
Norman Verbeck, *Columbia/Montour AVTS, Bloomsburg, PA*

About The Authors

Larry Jeffus is a dedicated teacher and author with over twenty years experience in the classroom and several Delmar Cengage Learning welding publications to his credit. He has been nominated by several colleges for the Innovator of the Year award for setting up nontraditional technical training programs. He was also selected as the Outstanding Post-Secondary Technical Educator in the State of Texas by the Texas Technical Society. Now retired from teaching, he remains very active in the welding community, especially in the field of education.

Lawrence Bower is a welding instructor at Blackhawk Technical College, an AWS SENSE School, in Janesville, Wisconsin. Mr. Bower is an AWS-certified Welding Inspector and Welding Educator. In helping to create *Welding: Skills, Processes and Practices for Entry-Level Welders*, he has brought to bear an excellent mix of training experience and manufacturing know-how from his work in industry, including fourteen years at United Airlines, and six years in the US Navy as an aerospace welder.

This page intentionally left blank

CHAPTER

1

Gas Metal Arc Welding Equipment, Setup, and Operation

OBJECTIVES

After completing this chapter, the student should be able to

- demonstrate the proper use of personal protective equipment (PPE) for gas metal arc welding (GMAW)
- list the various terms used to describe gas metal arc welding
- describe methods of metal transfer including the axial spray metal transfer process, globular transfer, pulsed-arc metal transfer (GMAW-P), and short-circuiting transfer (GMAW-S)
- list four common shielding gases or gas mixtures used for short-circuiting, spray, and pulsed-spray transfer on plain carbon steel
- locate the GMA welding filler metal on a welding procedure specification (WPS)
- define *deposition efficiency,* and tell how a welder can control the deposition rate
- define *voltage, electrical potential, amperage,* and *electrical current* as related to GMA welding
- tell how wire-feed speed is determined and demonstrate its relationship to welding current
- list five ways the GMAW molten weld pool can be controlled by varying the shielding gas, power settings, travel speed, electrode extension, and gun angle
- describe and demonstrate the backhand and forehand welding techniques and their relationship to weld bead profile and penetration in the short-circuiting transfer mode
- list and describe the basic GMAW equipment
- use a chart to select the correct eye and face protective devices for working and welding in a shop
- describe what type of general work clothing should be worn in a welding shop
- describe special protective clothing worn by welders to protect hands, arms, body, waist, legs, and feet

KEY TERMS

axial spray metal transfer

electrode extension (stickout)

globular transfer

pinch effect

pulsed-arc metal transfer

short-circuiting transfer

slope

synergic systems

transition current

welding helmet

AWS SENSE EG2.0

Key Indicators Addressed in this Chapter:

Module 2: Safety and Health of Welders

Key Indicator 1: Demonstrates proper use and inspection of personal protection equipment (PPE)

Module 5: Gas Metal Arc Welding (GMAW-S, GMAW spray transfer)

Key Indicator 1: Performs safety inspections of GMAW equipment and accessories

Key Indicator 3: Sets up for GMAW-S operations on carbon steel

Key Indicator 8: Sets up for GMAW (spray) operations on carbon steel

INTRODUCTION

In the 1920s, a metal arc welding process using an unshielded wire was being used to assemble the rear axle housings for automobiles. The introduction of the shielded metal arc welding electrode rapidly replaced the bare wire. The shielded metal arc welding electrode made a much higher-quality weld. In 1948, the first inert gas metal arc welding (GMAW) process, as it is known today, was developed and became commercially available, Figure 1.1. In the beginning, the GMAW process was used to weld aluminum using argon (Ar) gas for shielding. As a result, the process was known as MIG, which stands for metal inert gas welding. The later introduction of CO_2 and O_2 to the shielding gas has resulted in the American Welding Society's preferred term *gas metal*

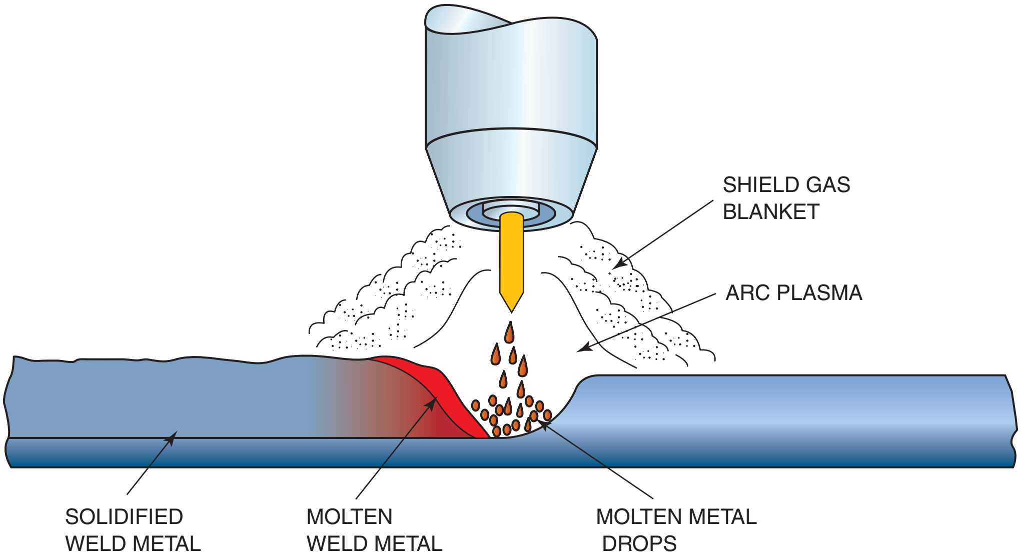

Figure 1.1 **Gas-shielded metal arc welding (GMAW)**

Table 1.1 Methods of Performing Welding Processes

Function	Manual (MA) (Example: SMAW)	Semiautomatic (SA) (Example: GMAW)	Machine (ME) (Example: GMAW)	Automatic (AU) (Example: GMAW)
Maintain the arc	Welder	Machine	Machine	Machine
Feed the filler metal	Welder	Machine	Machine	Machine
Provide the joint travel	Welder	Welder	Machine	Machine
Provide the joint guidance	Welder	Welder	Welder	Machine

arc welding (GMAW). Although the American Welding Society uses the term *gas metal arc welding* to describe this process, it is known in the field by several other terms, such as

- MIG, which is short for metal inert gas welding
- MAG, which is short for metal active gas welding
- *wire welding*, which describes the electrode used

The GMAW process may be performed as semiautomatic (SA), machine (ME), or automatic (AU) welding, Table 1.1. The GMA welding process is commonly performed as a semiautomatic process and is often mistakenly referred to as "semiautomatic welding." Equipment is available to perform most of the wire-feed processes semiautomatically, and the GMAW process can be fully automated. Robotic arc welding often uses GMAW because of the adaptability of the process in any position, Figure 1.2.

The rising use of all the various types of consumable wire welding processes has resulted in the increased sales of wire. At one time, wire made up less than 1% of the total market of filler metal. The total tonnage of filler metals used has grown and so has the percentage of wire. Today, wire exceeds 50% of the total tonnage of filler metals produced and used.

Much of the increase in the use of the wire welding processes is due to the increases in the quality of the welds produced. This improvement is due to an increased reliability of the wire-feed systems, improvements in the filler metal, smaller wire sizes, faster welding speed, higher weld deposition rates, less expensive shielding gases, and improved welding techniques. Table 1.2 shows the typical weld deposition rates using the GMA welding process. The increased usage has led to a reduction in the cost of equipment. GMA welding equipment is now found even in small shops.

In this chapter, the semiautomatic GMA welding process will be covered. The skill required to set up and operate this process is basic to the understanding and operation of other wire-feed processes. The reaction of the weld to changes in voltage, amperage, feed speed, stickout, and gas is similar to that of most wire-feed processes.

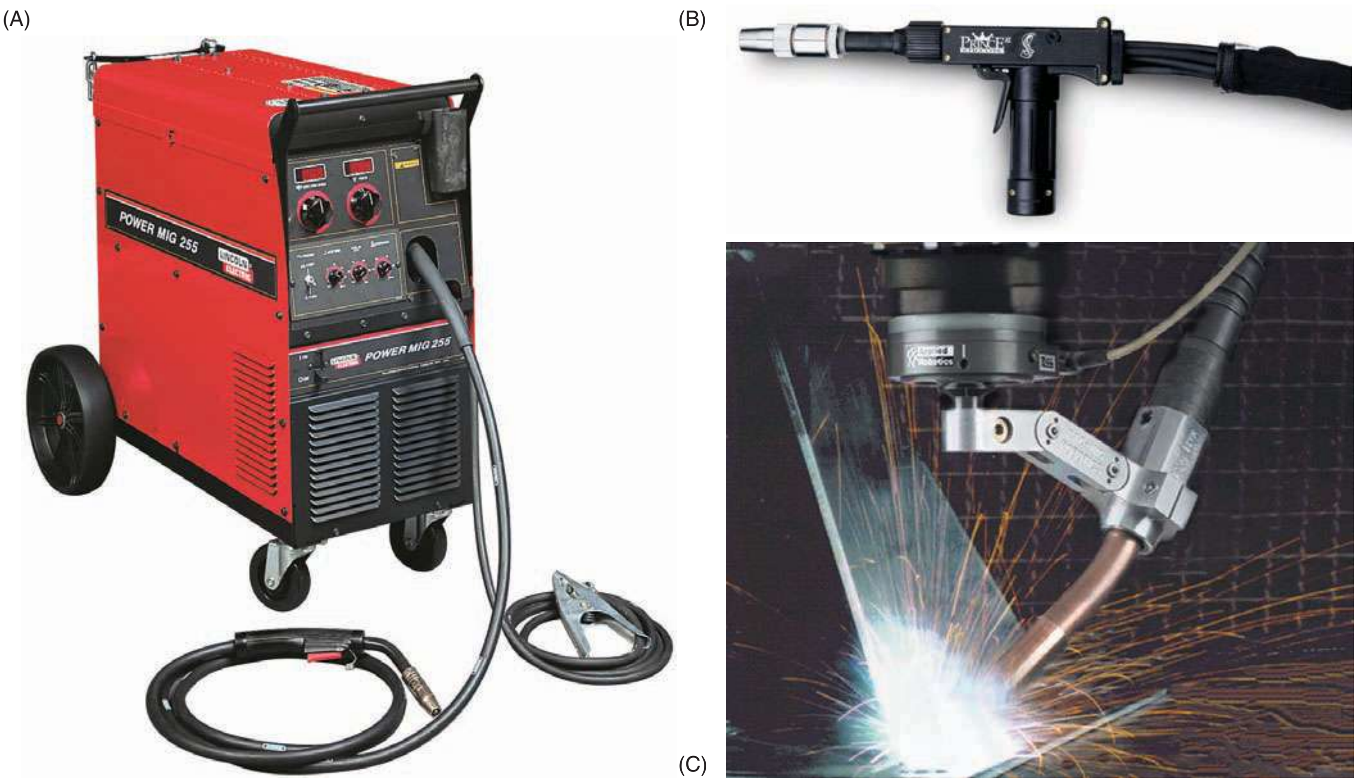

Figure 1.2 **GMAW equipment**
(A) Semiautomatic GMA welding setup. (B) Machine GMA welding gun with a friction drive, which provides both uniform nozzle-to-work distance and travel speed. (C) Automatic GMA welding.
Source: Courtesy of Lincoln Electric Company

Table 1.2 **GMA Weld Deposition Rates**

Electrode Diameter	Pounds per Hour		
Amperage	**0.35**	**0.45**	**0.63**
50	2.0	–	–
100	4.8	4.2	–
150	7.5	6.7	5.1
200	–	8.7	7.8
250	–	12.7	11.1
300	–	–	14.4

METAL TRANSFER

When first introduced, the GMA welding process was used with argon as a shielding gas to weld aluminum. Even though argon (Ar) was then expensive, the process was accepted immediately because it was much more productive than the gas tungsten arc (GTA) process and because it produced higher-quality welds than the shielded metal arc (SMA) process. This new arc welding process required very little postweld cleanup because it was slag and spatter free.

Axial Spray Metal Transfer

The freedom from spatter associated with the argon-shielded GMAW process results from a unique mode of metal transfer called **axial spray metal transfer**, Figure 1.3. This process is identified by the pointing of the wire tip from which droplets smaller than the diameter of the electrode wire are projected axially across the arc gap to the molten weld pool. There are hundreds of drops per second crossing from the wire to the base metal. These drops are propelled by arc forces at high velocity in the direction the wire is pointing. In the case of plain carbon steel and stainless steels, the molten weld pool may be too large and too hot to be controlled in vertical or overhead positions. Because the drops are very small and directed at the molten weld pool, the process is spatter free.

The spray transfer mode for carbon steels requires three conditions: argon-rich shielding gas mixtures, direct current electrode positive (DCEP) polarity (also called direct current reverse polarity, DCRP), and a current level above a critical amount called the **transition current**. The shielding gas is usually a mixture of 95% to 98% argon and 5% to 2% oxygen, or 80% to 90% argon and 20% to 10% carbon dioxide. The added percentage of active gases allows greater weld penetration. Figure 1.4 illustrates how the rate of drops transferred changes in relationship to the welding current. At low currents, the drops are large and are transferred at rates below 10 per second. These drops move slowly, falling from the electrode tip as gravity pulls them down. They tend to bridge the gap between the electrode

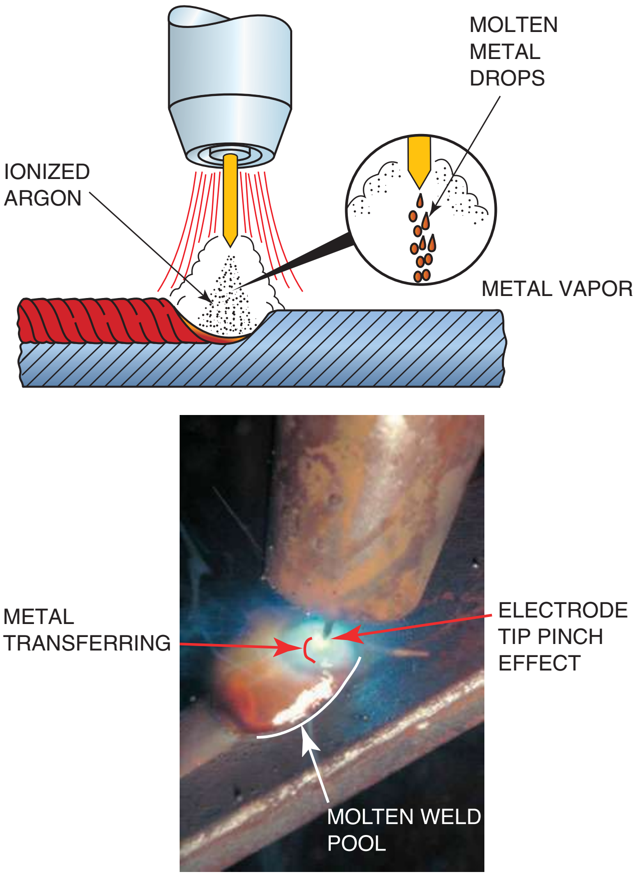

Figure 1.3 Axial spray metal transfer
Note the pinch effect of filler wire and the symmetrical metal transfer column.
Source: Courtesy of Larry Jeffus

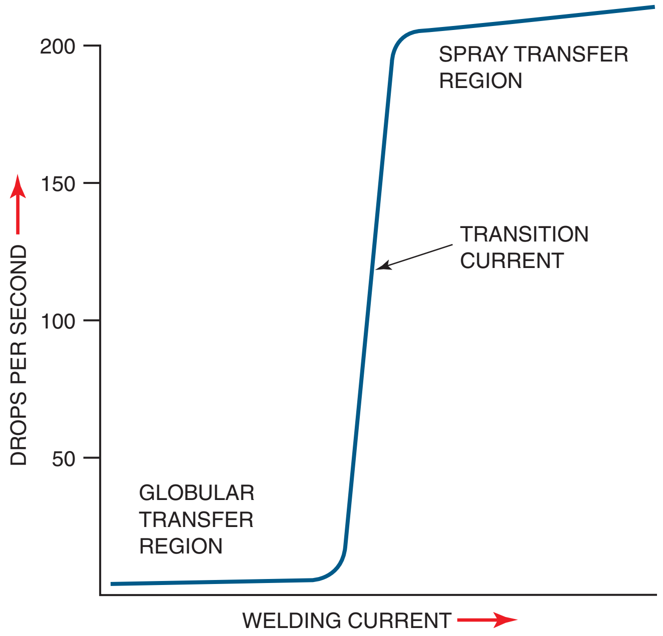

Figure 1.4 Desirable spray transfer shown schematically

tip end and the molten weld pool. This produces a momentary short circuit that throws off spatter. However, the mode of transfer changes abruptly above the critical current, producing the desirable spray.

The transition current depends on the alloy being welded. It is also proportional to the wire diameter, meaning that higher currents are needed with larger diameter wires. The need for high current density imposes some restrictions on the process. The high current hinders welding of sheet metal because the high heat cuts through sheet metal. High current also limits its use to the flat and horizontal welding positions. Weld control in the vertical or overhead position is very difficult to impossible to achieve. Table 1.3 lists the welding parameters for a variety of gases, wire sizes, and metal thicknesses for GMA welding of mild steel.

Table 1.3 GMA Welding Parameters for Mild Steel

Mild Steel	Wire-feed Speed, in./min		Voltage, V				
Base-material Thickness, in.	0.035-in.	0.045-in.	CO_2	75 Ar-25 CO_2	Ar	98 Ar-$2O_2$	Current A
0.036	105–115	–	18	16	–	–	50–60
0.048	140–160	70	19	17	–	–	70–80
0.060	180–220	90–110	20	17.7	–	–	90–110
0.075	240–260	120–130	20.7	18	20	–	120–130
1/8	280–300	140–150	21.5	18.5	20.5	–	140–150
3/16	320–340	160–175	22	19	21.5	23.5	160–170
1/4	360–380	185–195	22.7	19.5	22.5	24.5	180–190
5/16	400–420	210–220	23.5	20.5	23.5	25	200–210
3/8	420–520	220–270	25	22	25	26.5	220–250
1/2 and up	–	375	28	26	29	31	300

Globular Transfer

The **globular transfer process** is rarely used by itself because it transfers the molten metal across the arc in much larger droplets. It is used in combination with pulsed-spray transfer.

Globular transfer can be used on thin materials and at a very low current range. It can be used with higher current but is not as effective as other welding modes of metal transfer.

Caution

The heat produced during pulsed-spray transfer welding using large-diameter wire or high current may be intense enough to cause the filter lens in a welding helmet to shatter. Be sure the helmet is equipped with a clear plastic lens on the inside of the filter lens. Avoid getting your face too close to the intense heat.

Pulsed-Spray Transfer: GMAW-P

Because the change from spray arc to globular transfer occurs within a very narrow current range and the globular transfer occurs at the rate of only a few drops per second, a controlled spray transfer at significantly lower average currents is achievable. Gas metal arc welding with pulsed-spray transfer (GMAW-P), or **pulsed-arc metal transfer** involves pulsing the current from levels below the transition current to those above it. The time interval below the transition current is short enough to prevent a drop from developing. About 0.1 second is needed to form a globule, so no globule can form at the electrode tip if the time interval at the low base current is about 0.01 second. Actually, the energy produced during this time is very low—just enough to keep the arc alive.

The arc's real work occurs during those intervals when the current pulses to levels above the transition current. The time of that pulse is controlled to allow a single drop of metal to transfer. This is typical of the drops normally associated with spray transfer. In fact, with many power supplies, a few small drops could transfer during the pulse interval. As with conventional spray arc, the drops are propelled across the arc gap, allowing metal transfer in all positions.

The average current can be reduced sufficiently to reduce penetration enough to weld sheet metal or reduce deposition rates enough to control the molten weld pool in all positions. This level controlling the weld heat input and rate of weld metal deposit is achieved by changing the following variables, graphed in Figure 1.5:

- Frequency—The number of times the current is raised and lowered to form a single pulse; frequency is measured in pulses per second.
- Amplitude—The amperage or current level of the power at the peak or maximum, expressed in amperage.
- Width of the pulses—The amount of time the peak amperage is allowed to stay on.

Figure 1.6 shows a typical pulsed-arc welding system. Although developed in the mid-1960s, this technology did not receive much attention until solid state electronics were developed to handle the high power required of welding power supplies. Solid state electronics provided a better, simpler, and more economical way to control the pulsing process. The newest generation of pulsed-arc systems interlocks the power supply and wire feeder so that the proper settings of the wire-feed and power supply are obtained for any given job by adjusting a single knob. Such systems have been termed **synergic systems**. The greatest benefit to synergic GMAW-P is that the power supply reacts to changes in the

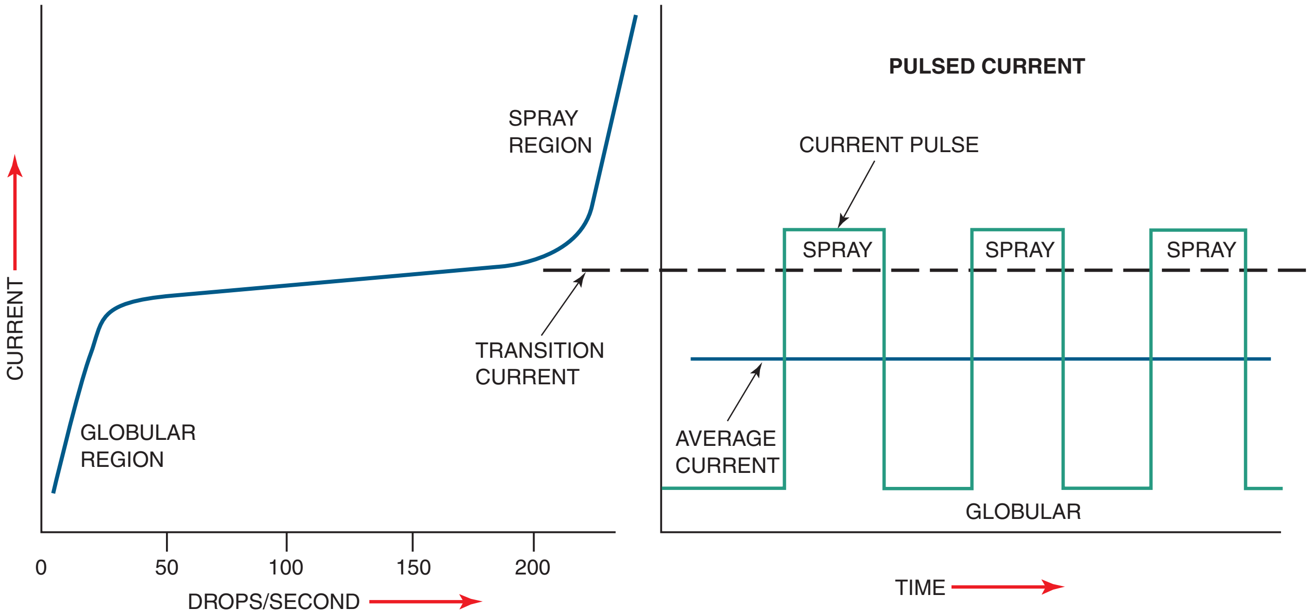

Figure 1.5 Mechanism of pulsed-arc spray transfer at a low average current

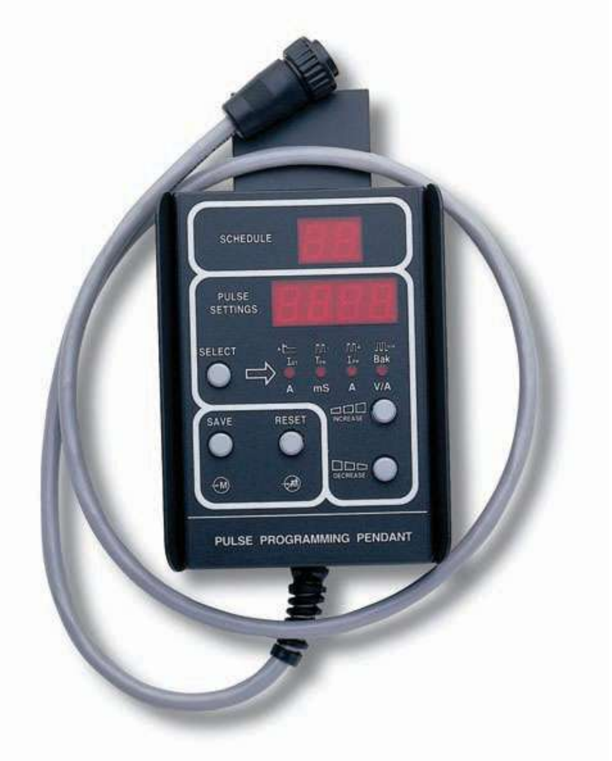
SCHEDULE
PULSE SETTINGS
SELECT
Bak
A
mS
A
V/A
SAVE
RESET
INCREASE
DECREASE
PULSE PROGRAMMING PENDANT

Figure 1.6 GMA pulsed-arc welding system controller
Source: Courtesy of Thermal Arc, a Thermadyne Company

work conditions such as inconsistent fitups or operator changes in stickout. For example, with a traditional GMA weld, amperage is reduced as the operator increases the stickout; but with synergic GMAW-P, the power supply senses the change in stickout and adjusts the current to maintain a more consistent weld.

In some respects, synergic GMAW-P systems are more complex to use because the correct relationships between the wire-feed speeds and power supply settings must be programmed into the equipment, and each wire composition, wire size, and shielding gas requires a special program. The manufacturer generally programs the most common combinations, allowing space in the computer for additional user input. In addition, welding power supplies that produce synergic pulsation must be capable of both constant-current (CC) and constant-voltage (CV) power output.

Shielding Gases for Spray or Pulsed-spray Transfer

Axial spray transfer is impossible without shielding gases rich in argon. Pure argon is used with all metals except steels. As much as 80% helium can be added to the argon to increase the heat in the arc without affecting the desirable qualities of the spray mode. With more helium, the transfer becomes progressively more globular, forcing the use of a different welding mode, to be described later. Since these gases are inert, they do not react chemically with any metals. This factor makes the GMAW process the only productive manual or semiautomatic method for welding metals sensitive to oxygen (essentially all metals except iron or nickel). The cathodic cleaning action which helps to remove the thin layer oxides that form on metals is associated with argon at DCEP (DCRP) and is also very important for fabricating metals such as aluminum, which quickly develop these undesirable surface oxides when exposed to air.

This same cleaning action causes problems with steels. Iron oxide in and on the steel surface is a good emitter of electrons that attracts the arc. But these oxides are not uniformly distributed, resulting in very irregular arc movement and in turn irregular weld deposits. This problem was solved by adding small amounts of an active gas such as oxygen or carbon dioxide to the argon. The reaction produces a uniform film of iron oxide on the weld pool and provides a stable site for the arc. This discovery enabled uniform welds in ferrous alloys and expanded the use of GMAW to welding those materials.

The amount of oxygen needed to stabilize arcs in steel varies with the alloy. Generally, 2% is sufficient for carbon and low-alloy steels. In the case of stainless steels, about 0.5% should prevent a refractory scale of chromium oxide, which can be a starting point for corrosion in stainless steels. Carbon dioxide can substitute for oxygen. Between 8% and 10% is optimum for low-alloy steels. In many applications, carbon dioxide is the preferred addition because the weld bead has a better contour and the arc appears to be more stable. Gas mixes of 98% argon–2% oxygen as well as 98% argon–2% carbon dioxide are commonly used for spray transfer GMA welding of stainless steels.

Buried-arc Transfer

Carbon dioxide was one of the first gases studied during the development of the GMAW process. It was abandoned temporarily because of excessive spatter and porosity in the weld. After argon was accepted for shielding, further work with carbon dioxide demonstrated that the spatter was associated with globular metal transfer. The large drops are partially supported by arc forces, Figure 1.7. As they become heavy enough to overcome those forces and drop into the pool, they bridge the gap between the wire and the weld pool, producing explosive short circuits and spatter.

Additional work showed that the arc in carbon dioxide was very forceful. Because of this, the wire tip could be driven below the surface of the molten weld pool. With the shorter arcs, the drop size is reduced, and any spatter produced as the result of short circuits is trapped in the cavity produced by the arc—hence the name **buried-arc transfer**, Figure 1.8. The resultant welds tend to be more highly crowned than those produced with open arcs, but they are relatively free of spatter and offer a decided advantage of welding speed. These characteristics make the buried-arc process useful for high-speed mechanized welding of thin sections, such as that found in compressor domes for hermetic air-conditioning and refrigeration equipment or for automotive components.

Because carbon dioxide is an oxidizing gas, its applications to welding carbon steels are restricted. It cannot be used to fabricate most nonferrous materials. Neither should it be used to weld stainless steels, because carbon corrodes the weld metal.

Carbon dioxide and helium are similar in that metal transfer in both gases is globular. Helium, too, can be used with the buried-arc technique. It has the advantage of being inert, potentially making it useful for the same types of applications as carbon dioxide but in nonferrous alloys.

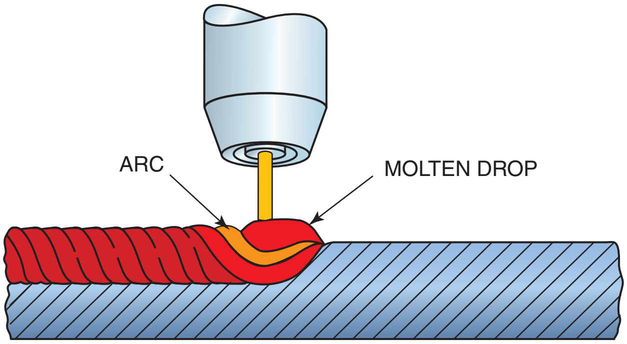

Figure 1.7 Globular metal transfer
Large drop is supported by arc forces.

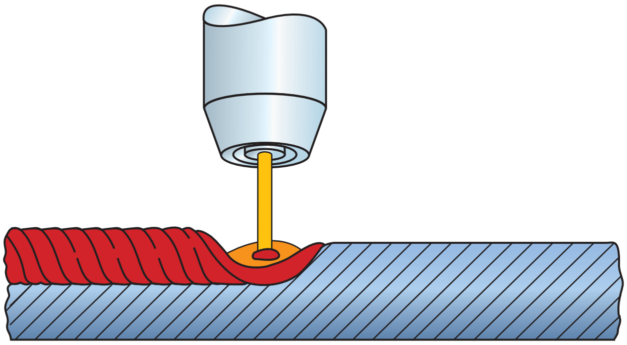

Figure 1.8 Buried-arc transfer
Wire tip is within the weld crater, so spatter is trapped.

Short-circuiting Transfer: GMAW-S

Low currents allow the liquid metal at the electrode tip to be transferred by direct contact with the molten weld pool. This process requires close interaction between the wire feeder and the power supply. This technique is called **short-circuiting transfer**, or gas metal arc welding with short-circuiting transfer (GMAW-S).

The short-circuiting mode of transfer is the most common process used with GMA welding

- on thin or properly prepared thick sections of material
- on a combination of thick to thin materials
- with a wide range of electrode diameters
- with a wide range of shielding gases
- in all positions

The 0.023, 0.030, 0.035, and 0.045 wire electrodes are the most common diameters for the short-circuiting mode. The most popular shielding gases used on carbon steel are 100% carbon dioxide (CO_2) or a combination of 75% argon (Ar) and 25% CO_2. The amperage range may be as low as 35 for materials of 24 gauge or as high as 225 for materials up to 1/8 inch in thickness on square groove weld joints. Thicker base metals can be welded if the edges are beveled to accept complete joint weld penetration.

The transfer mechanisms in this process are quite simple and straightforward, as shown schematically in Figure 1.9. To start, the wire is in direct contact with the molten weld pool, Figure 1.9A. Once the electrode touches the molten weld pool, the arc and its resistance are removed. Without the arc resistance, the welding amperage quickly rises as it begins to flow freely through the tip of the wire into the molten weld pool. The resistance to current flow is highest at the point where the electrode touches the molten weld pool. The resistance is high because both the electrode tip and weld pool are very hot. The higher the temperature,

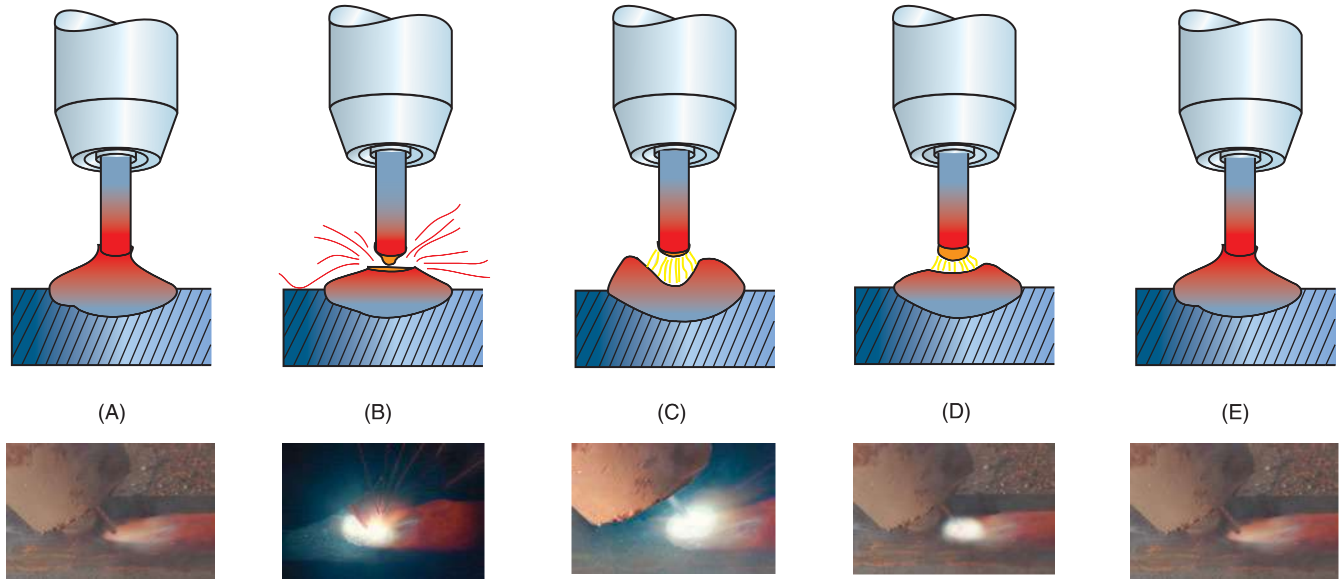

Figure 1.9 Schematic of short-circuiting transfer
Source: Courtesy of Larry Jeffus

the higher the resistance to current flow. A combination of high current flow and high resistance causes a rapid rise in the temperature of the electrode tip.

As the current flow increases, the interface between the wire and molten weld pool is heated until it explodes into a vapor (Figure 1.9B), establishing an arc. This small explosion produces sufficient force to depress the molten weld pool. A gap between the electrode tip and the molten weld pool (Figure 1.9C) immediately opens. With the resistance of the arc reestablished, the voltage increases as the current decreases.

The low current flow is insufficient to continue melting the electrode tip off as fast as it is being fed into the arc. As a result, the arc length rapidly decreases (Figure 1.9D) until the electrode tip contacts the molten weld pool (Figure 1.9E). The liquid formed at the wire tip during the arc-on interval is transferred by surface tension to the molten weld pool, and the cycle begins again with another short circuit.

If the system is properly tuned, the rate of short circuiting can be repeated from approximately 20 to 200 times per second, causing a characteristic buzzing sound. The spatter is low and the process easy to use. The low heat produced by GMAW-S makes the system easy to use in all positions on sheet metal, low-carbon steel, low-alloy steel, and stainless steel ranging in thickness from 25 gauge (0.02 in.; 0.5 mm) to 12 gauge (0.1 in.; 2.6 mm). The short-circuiting process does not produce enough heat to make quality welds in sections much thicker than 1/4 in. (6 mm) unless it is used for the root pass on a grooved weld or to fill gaps in joints. Although this technique is highly effective, lack-of-fusion defects can occur unless the process is perfectly tuned and the welder is highly skilled, especially on thicker metal. For this reason the American Welding Society D1.1 code for structural steel does not list short-circuiting GMAW as a prequalified process. In this code all short-circuiting procedures must be qualified by extensive testing.

Carbon dioxide works well with this short-circuiting process because it produces the forceful, high-energy arc needed during the arc-on interval to displace the weld pool. Helium can be used as well. Pure argon is not as effective because its arc tends to be sluggish and not very fluid. However, a mixture of 25% carbon dioxide and 75% argon produces a less harsh arc and a flatter, more fluid, and desirable weld profile. Although more costly, this gas mixture is preferred. A gas mixture of 98% argon and 2% oxygen may also be used on thinner carbon steels and sheet metal. This mixture produces lower-energy short-circuiting transfer and can be an advantage on thin-gauge metals, producing minimal burn-through at voltages as low as 13 volts.

New technology in wire manufacturing has allowed smaller wire diameters to be produced. These smaller diameters have become the preferred size even though they are more expensive due to the cost of drawing wires down to these desirable sizes. The short-circuiting process works better with a short electrode stickout.

The power supply is most critical. It must have a constant voltage, otherwise known as constant-potential output, and sufficient inductance to slow the time rate of current increase during the short-circuit interval. Too little inductance causes spatter due to high-current surges. Too much inductance causes the system to become sluggish. The short-circuiting rate decreases enough to make the process difficult to use. Also, the power supply must sustain an arc long enough to premelt the electrode tip in anticipation of the transfer at recontact with the weld pool.

FILLER METAL SPECIFICATIONS

GMA welding filler metals are available for a variety of base metals, Table 1.4. The most frequently used filler metals are AWS specification A5.18 for carbon steel and AWS specification A5.9 for stainless steel. Wire electrodes are produced in diameters of 0.023, 0.030, 0.035, 0.045, and 0.062 inch. Other, larger diameters are available for production work and can include wire diameter sizes such as 1/16, 5/64, and 7/64 inch. Table 1.5 lists the most common sizes and the amperage ranges for these electrodes. The amperage will vary depending on the method of metal transfer, type of shielding gas, and base metal thickness. Some steel wire electrodes have a thin copper coating. This coating provides some protection to the electrode from rusting and improves the electrical contact between the wire electrode and the contact tube. It also acts as a lubricant to help the wire move more smoothly through the liner and contact tube. These electrodes may look like copper wire because of the very thin copper cladding. The amount of copper is so small that it

Table 1.4 **AWS Filler Metal Specifications for Different Base Metals**

Base Metal Type	AWS Filler Metal Specification
Aluminum and aluminum alloys	A5.10
Copper and copper alloys	A5.6
Magnesium alloys	A5.19
Nickel and nickel alloys	A5.14
Stainless steel (austenitic)	A5.9
Steel (carbon)	A5.18
Titanium and titanium alloys	A5.16

Table 1.5 **Filler Metal Diameters and Amperage Ranges**

	Electrode Diameter		Amperage
Base Metal	Inch	Millimeter	Range
Carbon Steel	0.023	0.6	35–190
	0.030	0.8	40–220
	0.035	0.9	60–280
	0.045	1.2	125–380
	1/16	1.6	275–450
Stainless Steel	0.023	0.6	40–150
	0.030	0.8	60–160
	0.035	0.9	70–120
	0.045	1.2	140–310
	1/16	1.6	280–450

either burns off or is diluted into the weld pool with no significant effect on the weld deposit.

Filler Metal Selection

Solid Wire

The AWS specification for carbon steel filler metals for gas-shielded welding wire is A5.18. Filler metal classified within this specification can be used for GMAW, gas tungsten arc welding (GTAW), and plasma arc welding (PAW) processes. Because in GTAW and PAW the wire does not carry the welding current, the letters *ER* are used as a prefix. The *ER* is followed by two numbers to indicate the minimum tensile strength of a good weld. The actual strength is obtained by adding three zeroes to the right of the number given. For example, ER70S-x is 70,000 psi.

The *S* located to the right of the tensile strength indicates that this is a solid wire. The last number—2, 3, 4, 5, 6, or 7—or the letter *G* is used to indicate the filler metal composition and the weld's mechanical properties, Figure 1.10.

ER70S-2

This is a deoxidized mild steel filler wire. The deoxidizers allow this wire to be used on metal that has light coverings of rust or oxides. There may be a slight reduction in the weld's physical properties if the weld is made on rust or oxides, but this reduction is only slight, and the weld will usually still pass the classification test standards. This is a general-purpose filler that can be used on killed, semikilled, and rimmed steels. Argon-oxygen, argon-CO_2, and CO_2 can be used as shielding gases. Welds can be made in all positions.

ER70S-3

This is a popular filler wire. It can be used in single or multiple-pass welds in all positions. ER70S-3 does not have the deoxidizers required to weld over rust, over oxides, or on rimmed steels. It produces high-quality welds on killed and semikilled steels. Argon-oxygen, argon-CO_2, and CO_2 can be used as shielding gases. This is the low-carbon steel filler most commonly used to weld galvanized steel.

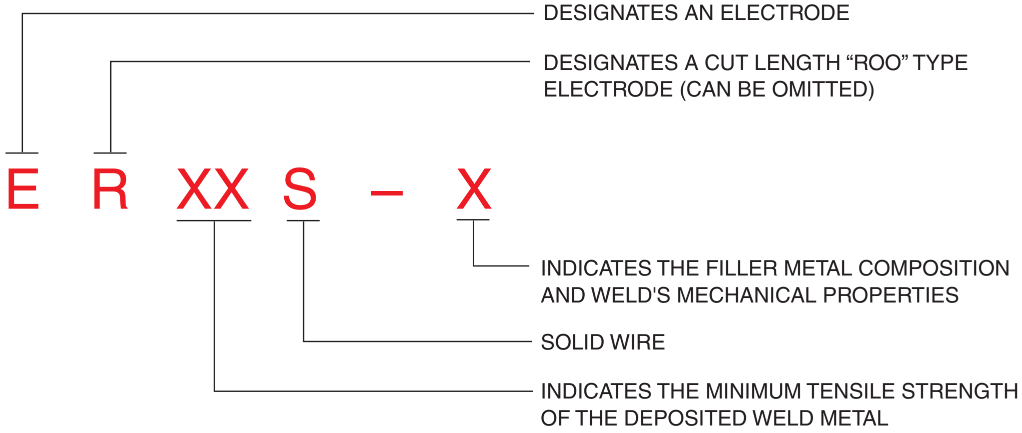
DESIGNATES AN ELECTRODE
DESIGNATES A CUT LENGTH "ROO" TYPE ELECTRODE (CAN BE OMITTED)
E R XX S – X
INDICATES THE FILLER METAL COMPOSITION AND WELD'S MECHANICAL PROPERTIES
SOLID WIRE
INDICATES THE MINIMUM TENSILE STRENGTH OF THE DEPOSITED WELD METAL

Figure 1.10 AWS numbering system for carbon steel filler metal for GMAW

ER70S-6

This is a good general-purpose filler wire. It has the highest levels of manganese and silicon. The wire can be used to make smooth welds on sheet metal or thicker sections. Welds over rust, oxides, and other surface impurities will lower the mechanical properties, but not normally below the specifications of this classification. Argon-oxygen, argon-CO_2, and CO_2 can be used as shielding gases. Welds can be made in all positions. This filler wire is not suggested for galvanized steel as its higher silicon content may induce cracking.

Stainless Steel Electrodes

The AWS specification for stainless steel–covered arc electrodes is A5.4 and for stainless steel bare, cored, and stranded electrodes and welding rods is A5.9. Filler metal classified within A5.4 uses the letter E as its prefix, and the filler metal within A5.9 uses the letters ER as its prefix.

Following the prefix, the three-digit stainless steel number from the American Iron and Steel Institute (AISI) is used. This number indicates the type of stainless steel in the filler metal, Table 1.6.

ALUMINUM AND ALUMINUM ALLOYS

The International Alloy Designation System is the most widely accepted naming scheme for aluminum alloys. Each alloy has a four-digit number, in which the first digit indicates the major alloying elements. Heat-treatable alloys can be strengthened by furnace operations after they are produced, while non-heat-treatable alloys are work hardened by plastic deformation such as bending and rolling operations. Heat treatment of aluminum alloys is sometimes referred to as precipitation hardening, and work hardening is sometimes referred to as strain hardening.

- 1000 series are essentially pure aluminum with a minimum 99% aluminum content by weight, and can be work hardened.
- 2000 series are alloyed with copper, and are heat treatable.
- 3000 series are alloyed with manganese, and can be work hardened.
- 4000 series are alloyed with silicon and can be work hardened.
- 5000 series are alloyed with magnesium, derive most of their strength from heat treatment, but can also be work hardened.
- 6000 series are alloyed with magnesium and silicon, are easy to machine, and are heat treatable, but not to the high strengths that 2000, 5000, and 7000 can reach.
- 7000 series are alloyed with zinc, and can be heat treated to the highest strengths of any aluminum alloy.
- 8000 series are a miscellaneous category where other elements not listed above may be used.

The AWS specifications for aluminum and aluminum alloy filler metals are A5.3 for covered arc welding electrodes and A5.10 for bare welding rods and electrodes. Filler metal classified within A5.3 uses the atomic symbol Al, and in A5.10 the prefix ER is used with the Aluminum Association number for the alloy, Table 1.7.

Aluminum Bare Welding Rods and Electrodes

ER1100

1100 aluminum has the lowest percentage of alloy agents of all the aluminum alloys, and it melts at 1215°F. The filler wire is also relatively pure. ER1100 produces welds that have good corrosion resistance and high ductility, with tensile strengths ranging from 11,000 to 17,000 psi. The weld deposit has a high resistance to cracking during welding. This wire can be used with oxyfuel gas welding (OFW), GTAW, and GMAW. Preheating to 300°F to 350°F is required for GTA welding on plate or pipe 3/8 in. and thicker to ensure good fusion. Flux is required for OFW. 1100 aluminum is commonly used for items such as food containers, food-processing equipment, storage tanks, and heat exchangers. ER1100 can be used to weld 1100 and 3003 grade aluminum.

ER4043

ER4043 is a general-purpose welding filler metal. It has 4.5% to 6.0% silicon added, which lowers its melting temperature to 1155°F. The lower melting temperature helps promote a free-flowing molten weld pool. The welds have high ductility and a high resistance to cracking during welding. This wire can be used with OFW, GTAW, and GMAW. Preheating to 300°F to 350°F is required for GTA welding on plate or pipe 3/8 in. and thicker to ensure good fusion. Flux is required for OFW. ER4043 can be used to weld on 2014, 3003, 3004, 4043, 5052, 6061, 6062, and 6063 and cast alloys 43, 355, 356, and 214.

ER5356

ER5356 has 4.5% to 5.5% magnesium added to improve the tensile strength. The weld has high ductility but only an average resistance to cracking during welding. This wire can be used for GTAW and GMAW. Preheating to 300°F to 350°F is required for GTA welding on plate or pipe 3/8 in. and thicker to ensure good fusion. ER5356 can be used to weld on 5050, 5052, 5056, 5083, 5086, 5154, 5356, 5454, and 5456.

ER5556

ER5556 has 4.7% to 5.5% magnesium and 0.5% to 1.0% manganese added to produce a weld with high strength. The weld has high ductility and only average resistance to cracking during welding. This wire can be used for GTAW and GMAW. Preheating to 300°F to 350°F is required for GTA welding on plate or pipe 3/8 in. and thicker to ensure good fusion. ER5556 can be used to weld on 5052, 5083, 5356, 5454, and 5456.

WIRE MELTING AND DEPOSITION RATES

The wire melting rates, deposition rates, and wire-feed speeds of the consumable wire welding processes are affected by the same variables. Before discussing them, however, these terms need to be defined. The wire melting rate, measured in inches per minute (in/min) or pounds per hour (lb/hr), is the rate at which the arc consumes the wire. The deposition rate, the measure of weld metal deposited, is nearly always less than the melting rate because not all of the wire is converted to weld metal. Some is lost as slag, spatter, or fume. The amount of weld metal deposited in ratio to the wire used is called the deposition efficiency.

Deposition efficiencies depend on the process, on the gas used, and even on how the welder sets welding conditions. With efficiencies of approximately 98%, solid wires with argon-rich shield gas mixes are best. Some of the self-shielded cored wires are poorest, with efficiencies as low as 80%.

Welders can control the deposition rate by changing the current, electrode extension, and diameter of the wire. To obtain higher melting rates, they can increase the current or wire extension or decrease the wire diameter. Knowing the precise constants is unimportant. However, it is important to know that current greatly affects melting rate and that the electrode extension must be controlled if results are to be reproducible.

WELDING POWER SUPPLIES

To better understand the terms used to describe the different welding power supplies, you need to know the following electrical terms:

- Voltage, or volts (V), is a measurement of electrical pressure and is the force that causes the current (amperage) to flow, in the same way that pounds per square inch is a measurement of water pressure.
- Electrical potential means the same thing as voltage and is usually expressed by using the term *potential (P)*. The terms *voltage*, *volts*, and *potential* can all be interchanged when referring to electrical pressure.
- Amperage, or amps (A), is the measurement of the total number of electrons flowing, in the same way that gallons are a measurement of the amount of water flowing.
- Electrical current means the same thing as amperage and is usually expressed by using the term *current (C)*. The terms *amperage*, *amps*, and *current* can all be interchanged when referring to electrical flow.

GMAW power supplies are constant-voltage, constant-potential (CV, CP) machines, unlike shielded metal arc welding (SMAW) power supplies, which are constant-current (CC) machines and are sometimes called drooping arc voltage (DAV). It is impossible to make acceptable welds using the wrong type of power supply. GMAW power supplies are available as transformer rectifiers, motor generators, or inverters, Figure 1.11. Some newer machines use electronics, enabling them to supply both types of power at the flip of a switch; they may be referred to as CC/CV power supplies.

The relationships between current and voltage with different combinations of arc length or wire-feed speeds are called volt-ampere characteristics. The volt-ampere characteristics of arcs in argon with constant arc lengths or constant wire-feed speeds are shown in Figure 1.12. To maintain a constant arc length while increasing current, it is necessary to increase voltage. For example, with a 1/8-in. (3-mm) arc length, increasing current from 150 to 300 amperes requires increasing the

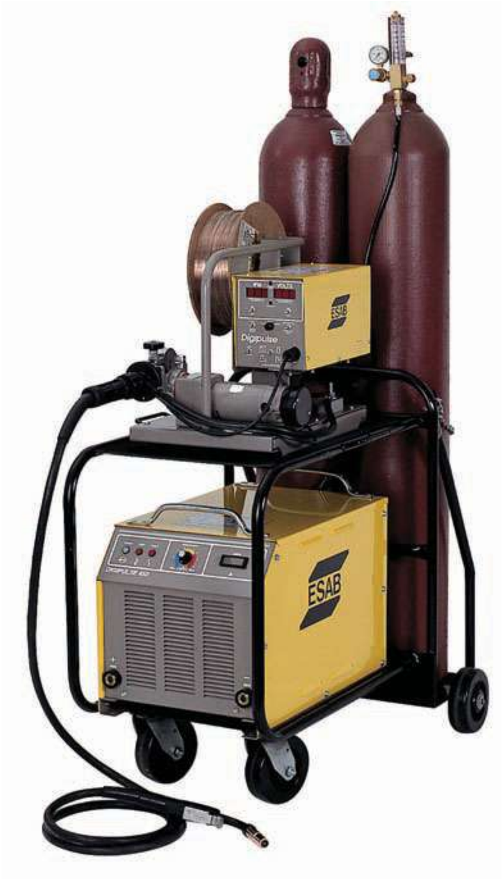

Figure 1.11 **Transformer rectifier welding power supply**
Source: Courtesy of ESAB Welding & Cutting Products

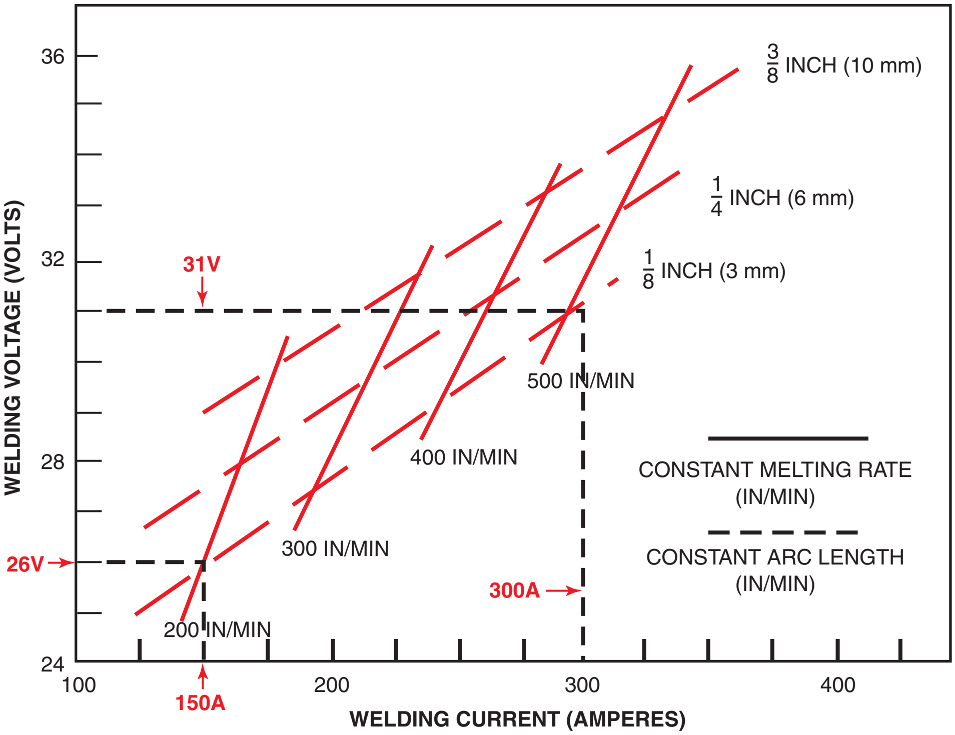

Figure 1.12 **Volt-ampere characteristics of arcs with argon**
The arc length and arc voltage are affected by the welding current and wire-feed speed (0.045-in. [1.43-mm] wire; 1-in. [25-mm] electrode extension).

voltage from about 26 to 31 volts. The current increase illustrated here results from increasing the wire-feed speed from 200 to 500 inches per minute.

Speed of the Wire Electrode

The wire-feed speed is generally recommended by the electrode manufacturer and is selected in inches per minute (ipm), or how fast the wire exits the contact tube. The welder uses a wire-speed control dial on the wire-feed unit to control ipm. It can be advanced or slowed to control the burn-off rate, or how fast the electrode transfers into the weld pool, to meet the welder's skill in controlling the weld pool. Note the direct relationship between current (amps) and wire-feed speed (wfs); as wire-feed speed increases, the amperage increases. If wire-feed speed is reduced, amperage will decrease, Table 1.8.

To accurately measure wire-feed ipm, snip off the wire at the contact tube. Wearing safety glasses and pointing the contact tube away from your face, squeeze the trigger for ten seconds; release and snip off the wire electrode. Measure the number of inches of wire that was fed out in the ten seconds. Now using basic shop math, multiply its total length in inches by six. The result is how many inches of wire were fed per minute.

Power Supplies for Short-circuiting Transfer

Although the GMA power source is said to have a constant voltage (CV) or constant potential (CP), it is not perfectly constant. The graph in Figure 1.13 shows that there is a slight decrease in voltage as the amperage increases within the working range. The rate of decrease is known as **slope**. It is expressed as the voltage decrease per 100-ampere increase—for example, 10 V/100 A. For short-circuiting welding, some welding power supplies are equipped to allow changes in the slope by steps or continuous adjustment.

The slope, which is called the *volt-ampere curve*, is often drawn as a straight line because it is fairly straight within the working range of the machine. Whether it is drawn as a curve or a straight line, the slope can be found by finding two points. The first point is the set voltage as read from the voltmeter when the gun switch is activated but no welding is

Table 1.8 Typical Amperages for Carbon Steel

	Wire Diameter Amperages			
Wire-feed Speed* in./min (m/min)	.030 in. (0.8 mm)	.035 in. (0.9 mm)	.045 in. (1.2 mm)	.062 in. (1.6 mm)
100 (2.5)	40	65	120	190
200 (5.0)	80	120	200	330
300 (7.6)	130	170	260	425
400 (10.2)	160	210	320	490
500 (12.7)	180	245	365	–
600 (15.2)	200	265	400	–
700 (17.8)	215	280	430	–

*To check feed speed, run out wire for one minute and then measure its length.

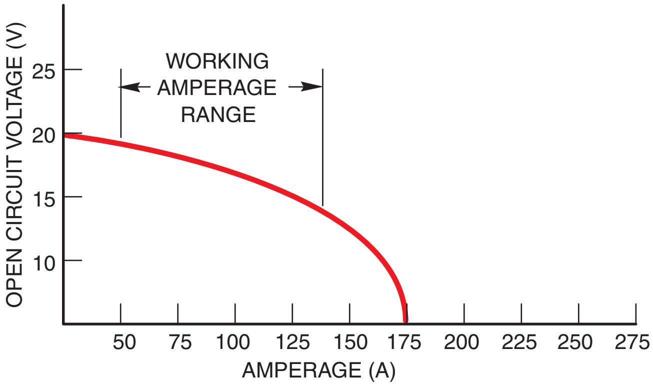

Figure 1.13 Constant potential welder slope

being done. This is referred to as the open circuit voltage. The second point is the voltage and amperage as read during a weld. The voltage control is not adjusted during the test but the amperage can vary. The slope is the voltage difference between the first and second readings. The difference can be found by subtracting the second voltage from the first voltage. Therefore, for settings over 100 amperes, it is easier to calculate the slope by adjusting the wire feed so that you are welding with 100 amperes, 200 amperes, 300 amperes, and so on. In other words, the voltage difference can be simply divided by 1 for 100 amperes, 2 for 200 amperes, and so forth.

The machine slope is affected by circuit resistance. Circuit resistance may result from a number of factors, including poor connections, long leads, or a dirty contact tube. A higher resistance means a steeper slope. In short-circuiting machines, increasing the inductance increases the slope. This increase slows the current's rate of change during short circuiting and the arcing intervals, Figure 1.14. Therefore, slope and inductance become synonymous in this discussion. As the slope increases, both the short-circuit current and the **pinch effect** are reduced. A flat slope has both an increased short-circuit current and a greater pinch effect.

The machine slope affects the short-circuiting metal transfer mode more than it does the other modes. Too much current and pinch effect from a flat slope cause a violent short and arc restart cycle, which results in increased spatter. Too little current and pinch effect from a steep slope result in the short circuit not being cleared as the wire freezes in the molten pool and piles up on the work, Figure 1.15.

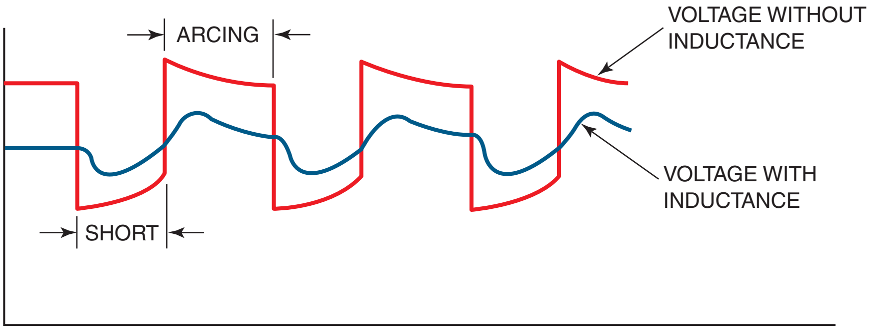

Figure 1.14 Voltage pattern with and without inductance

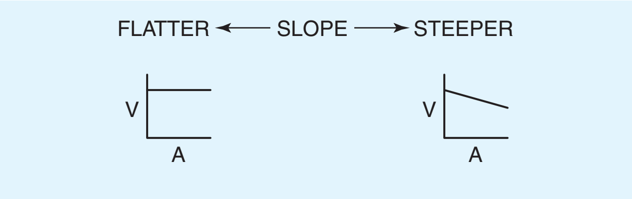

Buildup	Decreases	Increases
Depth of fusion	Increases	Decreases
Spatter	Increases	Decreases
Shorting	Violently cleared	May not clear

Figure 1.15 **Effect of slope**

The slope should be adjusted so that a proper spatter-free metal transfer occurs. On machines that have adjustable slopes, this is easily set. Experiment 2-2 describes a method of adjusting the circuit resistance to change the slope on machines that have a fixed slope. This is done by varying the contact tube-to-work distance. The GMA filler wire is much too small to carry the welding current and heats up due to its resistance to the current flow. The greater the tube-to-work distance, the greater the circuit resistance and the steeper the slope. By increasing or decreasing this distance, a proper slope can be obtained so that the short circuiting is smoother with less spatter.

Inductance

Many newer GMA welding power supplies are supplied with an inductance control. Inductance controls are used primarily in the GMA short-circuiting transfer mode, especially when open root joints are involved. Inductance controls the rise up to peak current during GMAW short-circuiting transfer. A low inductance control setting will provide a greater short-circuiting frequency, which may be beneficial in welding thin material where burn-through is an issue. A higher setting will reduce the frequency of short circuits, creating slightly longer arc periods, which allows greater current to go to the work. The greater current can be used to increase penetration in thicker sections or where complete penetration is required on open root joints welded from one side.

In the case of the GMA welding arc, the rate of change in the amperage relative to the arc voltage determines how the metal droplet detaches from the end of the electrode. If the rate of change is too rapid (inductance too low), the droplet detaches violently and produces excessive spatter. If the rate of change is too slow (inductance too high) the metal droplet doesn’t detach cleanly and the arc is unstable. A midrange setting on the inductance control can be used for general-purpose short-circuit GMA welding and adjustments made for thick, thin, or open root conditions.

MOLTEN WELD POOL CONTROL

The GMAW molten weld pool can be controlled by varying the following factors: shielding gas, power settings, gun manipulation, travel speed, electrode extension, and gun angle.

Shielding Gas

The shielding gas selected for a weld has a definite effect on the weld produced. The properties that can be affected include the method of metal transfer, welding speed, weld contour, arc cleaning effect, and fluidity of the molten weld pool.

In addition to the effects on the weld itself, the metal to be welded must be considered in selecting a shielding gas. Some metals must be welded with an inert gas such as argon or helium or mixtures of argon and helium. Other metals weld more favorably with reactive gases such as carbon dioxide or with mixtures of inert gases and reactive gases such as argon and oxygen or argon and carbon dioxide, Table 1.9. The most commonly used shielding gases are 75% argon + 25% CO_2, argon + 1% to 5% oxygen, and carbon dioxide, Figure 1.16.

- **Argon:** The atomic symbol for argon is *Ar*, and it is an inert gas. *Inert gases* do not react with any other substance and are insoluble in molten metal. One hundred percent argon is used on nonferrous metals such as aluminum, copper, magnesium, nickel, and their alloys; but 100% argon is not normally used for making welds on ferrous metals.

  Because argon is denser than air, it effectively shields welds by pushing the lighter air away. Argon is relatively easy to ionize. Easily ionized gases can carry long arcs at lower voltages. This makes it less sensitive to changes in arc length.

  Argon gas is naturally found in all air and is collected in air separation plants. There are two methods of separating air to extract nitrogen, oxygen, and argon. In the cryogenic process, air is super-cooled to temperatures that cause it to liquefy and the gases are separated. In the noncryogenic process, molecular sieves (strainers with very small holes) separate the various gases, much like using a screen to separate sand from gravel.
- **Argon gas blends:** Oxygen, carbon dioxide, helium, and nitrogen can be blended with argon to change argon's welding characteristics. Adding reactive gases (oxidizing), such as oxygen or carbon dioxide, to argon tends to stabilize the arc, promote favorable metal transfer, and minimize spatter. As a result, the penetration pattern is improved, and undercutting is reduced or eliminated. Adding helium or nitrogen gases (nonreactive or inert) increases the arc heat for deeper penetration.

  The amount of the reactive gases, oxygen or carbon dioxide, required to produce the desired effects is quite small. As little as a half percent change in the amount of oxygen will produce a noticeable effect on the weld. Most of the time, blends containing 1% to 5% of oxygen are used. Carbon dioxide may be added to argon in the range of 10% to 30%. Blends of argon with less than 10% carbon dioxide may not have enough arc voltage to give the desired results. The most commonly used blend for short-circuiting transfer is 75% argon and 25% CO_2.

  When using oxidizing shielding gases with oxygen or carbon dioxide added, a suitable filler wire containing deoxidizers should be used to prevent porosity in the weld. The presence of oxygen in the shielding gas can also cause some loss of certain alloying

Table 1.9 (A) Metals Matched with GMAW Shielding Gases and Gas Blends; (B) GMAW Shielding Gases and Gas Blends Matched with Metals

GMAW Shielding Gas, Gas Blends, Metals, and Welding Process							
Gases/ Blend	**Gas Reaction**	**Application**	**Remarks**	**Gas/ Blend**	**Gas Reaction**	**Application**	**Remarks**
Ar + 5% CO_2	Oxidizing	Low-alloy steel	Pulse spray and short-circuit transfer in out-of-position welds	Ar + CO_2 + O_2	Oxidizing	Low-alloy steel and some stainless steels	All metal transfer for automatic and robotic applications
Ar + 10% CO_2	Oxidizing	Low-alloy steel	Same as above with a wider, more fluid weld pool	Ar + CO_2 + N	Almost inert	Stainless steel	All metal transfer, excellent for thin gauge material
Ar + 25% CO_2	Oxidizing	Mild, low-alloy steels and stainless steel	Smooth weld surface, reduces penetration with short-circuiting transfer	He + 7.5% Ar + 2.5% CO_2	Almost inert	Stainless steel and some low-alloy steels	Excellent toughness, arc stability, wetting characteristics, and bead contour, little spatter with short-circuiting transfer

(B)

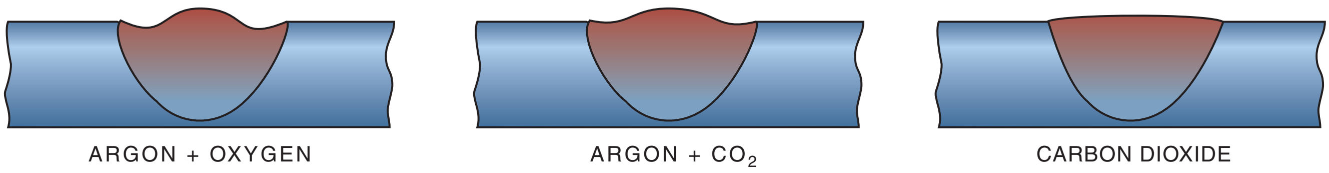

Figure 1.16 Effect of shielding gas on weld bead shape

elements, such as chromium, vanadium, aluminum, titanium, manganese, and silicon.

- **Helium:** The atomic symbol for helium is *He*, and it is an inert gas that is a product of the natural gas industry. It is removed from natural gas as the gas undergoes separation (fractionation) for purification or refinement.

  Helium is lighter than air; thus its flow rates must be about twice as high as argon's for acceptable stiffness in the gas stream to be able to push air away from the weld. Proper protection is difficult in drafts unless high flow rates are used. It requires a higher voltage to ionize, which produces a much hotter arc. There is a noticeable increase in both the heat and temperature of a helium arc. This hotter arc makes it easier to make welds on thick sections of aluminum and magnesium.

  Small quantities of helium are blended with heavier gases. These blends take advantage of the heat produced by the lightweight helium and weld coverage by the other heavier gas. Thus, each gas is contributing its primary advantage to the blended gas.
- **Carbon dioxide:** Carbon dioxide is a compound made up of one carbon atom (C) and two oxygen atoms (O_2), and its molecular formula is CO_2. One hundred percent carbon dioxide is widely used as a shielding gas for GMA welding of steels. In the short-circuiting transfer mode it allows higher welding speed, better penetration, good mechanical properties, and costs less than the inert gas mixes. The chief drawback in the use of carbon dioxide is the less-steady arc characteristics and a considerable increase in weld spatter. The spatter can be kept to a minimum by maintaining a very short, uniform arc length and strict attention to amperage and voltage parameters. CO_2 can produce sound, spatter-free welds of the highest quality, provided established procedures are followed and a filler wire having the proper deoxidizing additives is selected.
- **Nitrogen:** The atomic symbol for nitrogen is *N*. It is not an inert gas but is relatively nonreactive to the molten weld pool. It is often used in blended gases to increase the arc's heat and temperature. One hundred percent nitrogen can be used to weld copper and copper alloys and is an economical choice for gas purging of some austenitic stainless steel pipe welds.

Power Settings

As the power settings, voltage, and amperage are adjusted, the weld bead is affected. Making an acceptable weld requires a balancing of the voltage and amperage. If either or both are set too high or too low, the

Module 5
Key Indicator 3, 8

weld penetration can decrease. A GMA welding machine has no direct amperage settings. Instead, the amperage at the arc is adjusted by changing the wire-feed speed. As a result of the welding machine's maintaining a constant voltage when the wire-feed speed increases, more amperage flows across the arc. This higher amperage is required to melt the wire so that the same arc voltage can be maintained. The higher amperage is used to melt the filler wire and does not increase the penetration. In fact, the weld penetration may decrease significantly.

Increasing and decreasing the voltage changes the arc length; however, it may not put more heat into the weld. Like changes in the amperage, these voltage changes may decrease weld penetration.

Gun Manipulation

The GMA welding process is greatly affected by the location of the electrode tip and molten weld pool. During the short-circuiting process if the arc is directed to the base metal and outside the molten weld pool, the welding process may stop. Without the resistance of the hot molten metal, high-amperage surges occur each time the electrode tip touches the base metal, resulting in a loud pop and a shower of sparks. It is something that occurs each time a new weld is started. So when making a weave pattern, which is a gun manipulation technique of moving side to side in order to produce a wider weld bead, you must keep the arc and electrode tip directed into the molten weld pool. Other than the sensitivity to arc location, most of the SMAW weave patterns that keep the electrode wire at or near the leading edge of the weld pool can be used for short-circuiting GMA welds.

Travel Speed

Because the location of the arc inside the molten weld pool is important, the welding travel speed cannot exceed the ability of the arc to melt the base metal. Too high a travel speed can result in overrunning of the weld pool and an uncontrollable arc. Fusion between the base metal and filler metal can completely stop if the travel rate is too fast. If the travel rate is too slow and the weld pool size increases excessively, it can also restrict fusion to the base plate.

Electrode Extension

Module 5
Key Indicator 3, 8

The **electrode extension (stickout)** is the distance from the contact tube to the arc measured along the wire. Adjustments in this distance cause a change in the wire resistance and the resulting weld bead, Figure 1.17.

GMA welding currents are relatively high for the wire sizes, even for the low current values used in short-circuiting arc metal transfer, Figure 1.18. As the length of wire extending from the contact tube to the work increases, the voltage, too, should increase. Since this change is impossible with a constant-voltage power supply, the system compensates by reducing the current. In other words, by increasing the electrode extension and maintaining the same wire-feed speed, the current has to change to provide the same resistance drop. This situation leads to a reduction in weld heat, penetration, and fusion, and an increase in buildup. On the

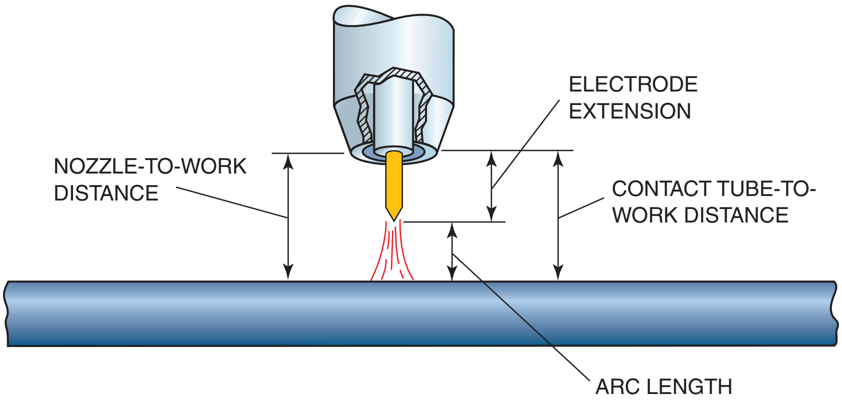

Figure 1.17 **Electrode-to-work distances**

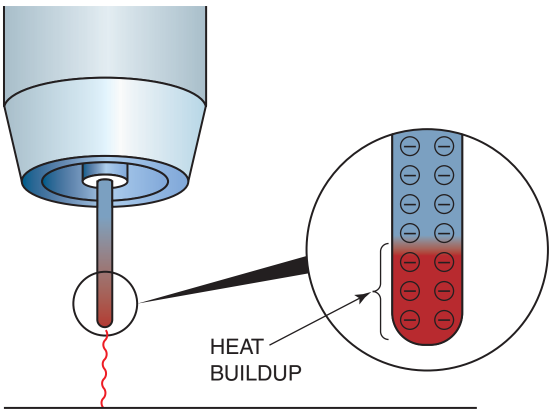

Figure 1.18 **Heat buildup due to the extremely high current for the small conductor (electrode)**

other hand, as the electrode extension distance is shortened, the weld heats up, penetrates more, and builds up less, Figure 1.19.

Experiment 2-3 explains the technique of using varying extension lengths to change the weld characteristics. Using this technique, a welder can make acceptable welds on metal ranging in thickness from 16 gauge to 1/4 in. (6 mm) or more without changing the machine settings. When using this technique, the nozzle-to-work distance should be kept the same so that enough shielding gas coverage is provided. Some nozzles can be extended to provide coverage. Others must be exchanged with the correct-length nozzle, Figure 1.20.

Gun Angle

Module 5
Key Indicator 3, 8

The GMA welding gun may be held so that the relative angle between the gun, work, and welding bead being made is either vertical or has a drag angle or a push angle. Changes in this angle will affect the weld bead. The effect is most noticeable during the short-circuiting arc and globular transfer modes.

Backhand welding is the welding technique that uses a drag angle, Figure 1.21. The welding technique that uses a push angle is known as forehand welding, Figure 1.22.

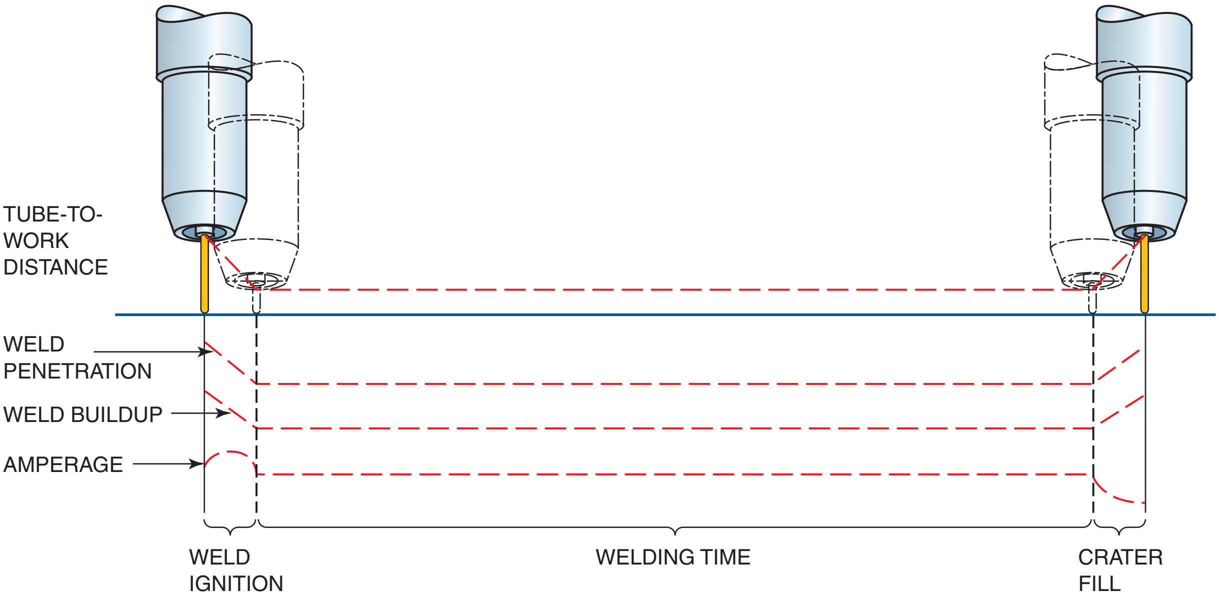
TUBE-TO-
WORK
DISTANCE
WELD
PENETRATION
WELD BUILDUP
AMPERAGE
WELD
IGNITION
WELDING TIME
CRATER
FILL

Figure 1.19 Using the changing tube-to-work distance to improve both the starting and stopping points of a weld

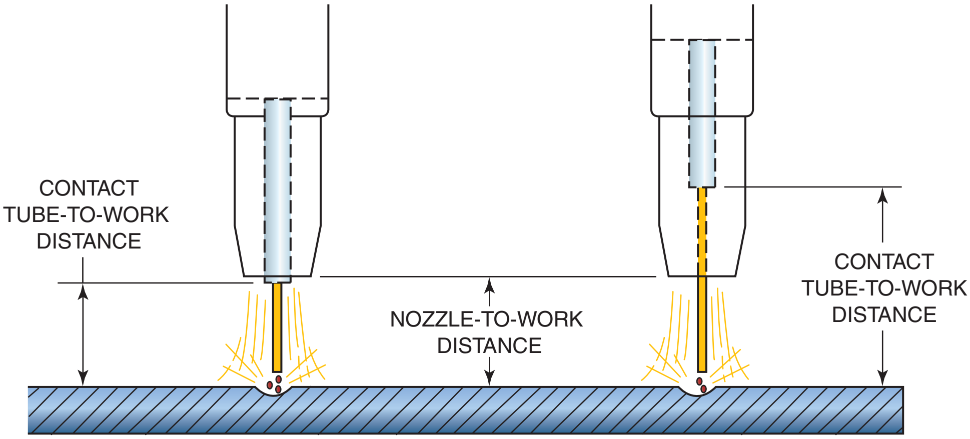
CONTACT
TUBE-TO-WORK
DISTANCE
NOZZLE-TO-WORK
DISTANCE
CONTACT
TUBE-TO-WORK
DISTANCE

Figure 1.20 Nozzle-to-work distance can differ from the contact tube-to-work distance

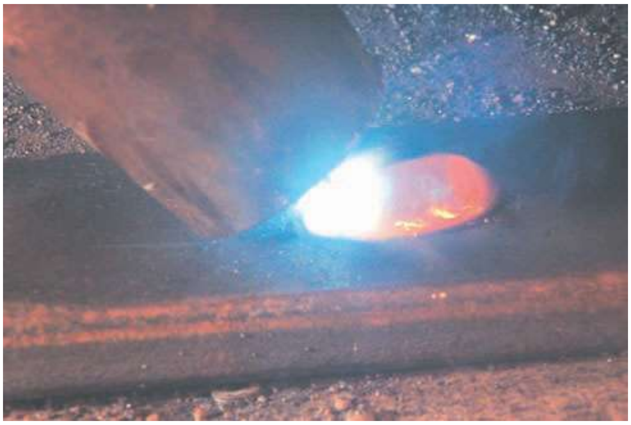

Figure 1.21 Backhand welding, or drag angle
Source: Courtesy of Larry Jeffus

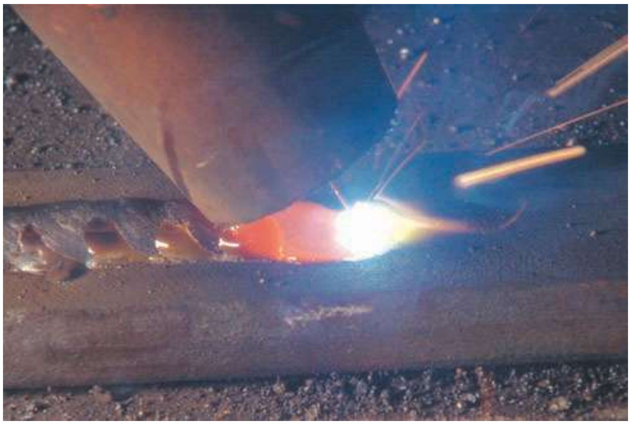

Figure 1.22 Forehand welding, or push angle
Source: Courtesy of Larry Jeffus

Backhand Welding

Module 5
Key Indicator 3, 8

A dragging angle, or backhand, welding technique directs the arc force into the molten weld pool of metal. This action, in turn, forces the molten metal back onto the trailing edge of the molten weld pool and exposes more of the unmelted base metal, Figure 1.23. The digging action pushes the penetration deeper into the base metal while building up the weld head. If the weld is sectioned, the profile of the bead is narrow and deeply penetrated, with high buildup.

Forehand Welding

Module 5
Key Indicator 3, 8

In a push angle, or forehand, welding technique, the arc force pushes the weld metal forward and out of the molten weld pool onto the cooler metal ahead of the weld, Figure 1.24. The heat and metal are spread out over a wider area. The sectional profile of the bead is wide, showing shallow penetration with little buildup.

The greater the angle, the more defined is the effect on the weld. As the angle approaches vertical, the effect is reduced. This allows the welder to change the weld bead as effectively as the changes resulting from adjusting the machine current settings.

EQUIPMENT

Module 5
Key Indicator 3, 8

The basic GMAW equipment consists of the gun, electrode (wire) feed unit, electrode (wire) supply, power source, shielding gas supply with flowmeter/regulator, control circuit, and related hoses, liners, and cables, Figure 1.25 and Figure 1.26. Larger, more complex systems may have water for cooling, solenoids for controlling gas flow, and carriages for moving the work or the gun or both, Figure 1.27. The system may be

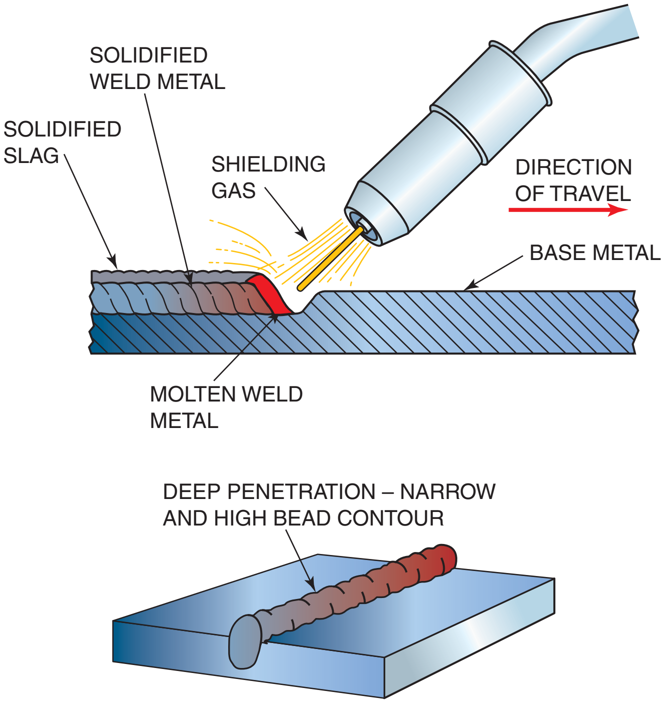

Figure 1.23 Backhand welding, or dragging angle

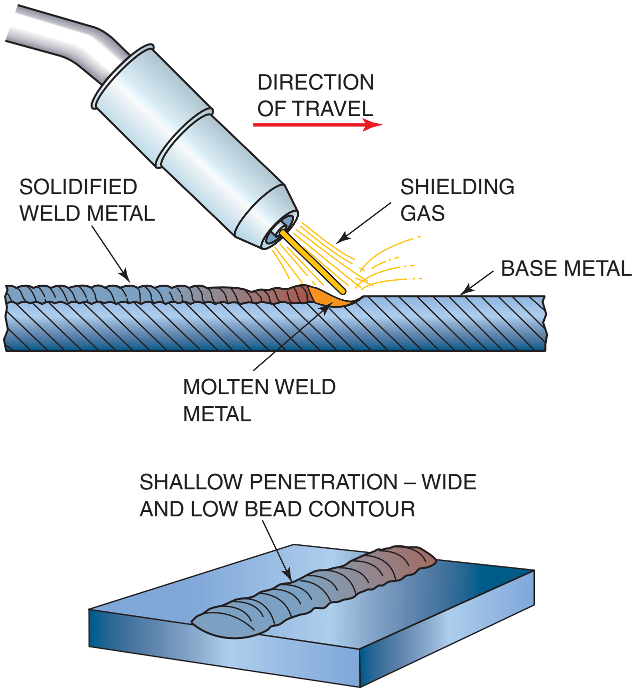
DIRECTION
OF TRAVEL
SOLIDIFIED
WELD METAL
SHIELDING
GAS
BASE METAL
MOLTEN WELD
METAL
SHALLOW PENETRATION – WIDE
AND LOW BEAD CONTOUR

Figure 1.24 Forehand welding, or pushing angle

stationary or portable, Figure 1.28. In most cases, the system is meant to be used for only one process. Some manufacturers, however, do make power sources that can be switched over for other uses.

Power Source

The power source may be either a transformer rectifier, inverter, or generator type. The transformers are stationary and commonly require a three-phase power source. The inverter power sources are smaller, lighter, and may be designed to accept a variety of different electrical inputs, from 208 volts to 440 volts, single or three-phase. Engine generators are ideal for portable use or where sufficient power is not available.

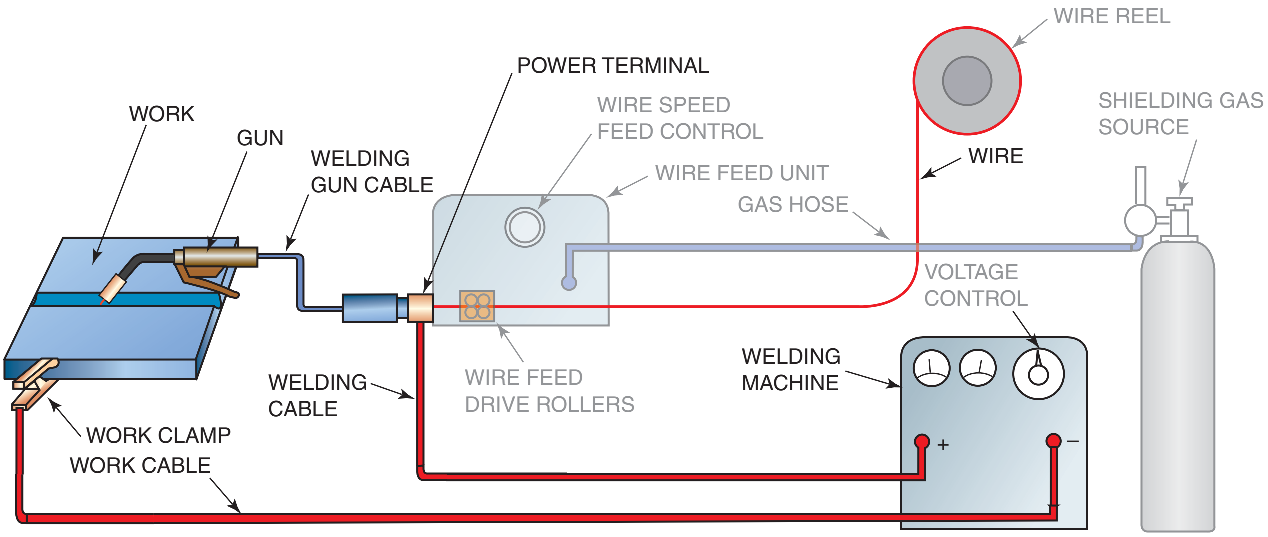
WIRE REEL
POWER TERMINAL
WORK
GUN
WIRE SPEED
FEED CONTROL
SHIELDING GAS
SOURCE
WELDING
GUN CABLE
WIRE
WIRE FEED UNIT
GAS HOSE
VOLTAGE
CONTROL
WELDING
MACHINE
WELDING
CABLE
WIRE FEED
DRIVE ROLLERS
WORK CLAMP
WORK CABLE
+
–

Figure 1.25 Schematic of equipment setup for GMA welding
Source: Courtesy of Hobart Brothers Company

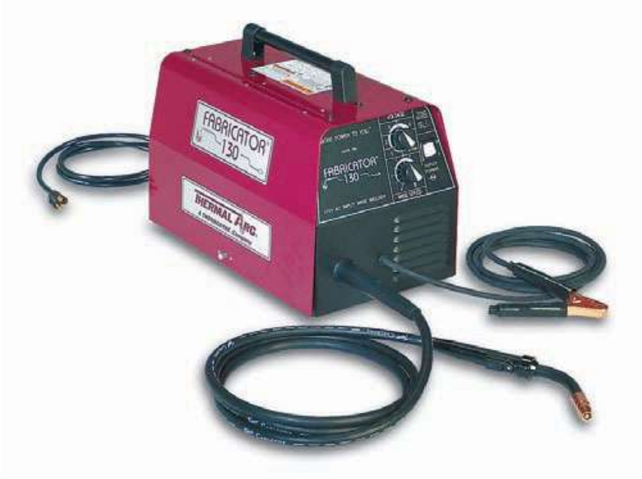

Figure 1.26 **Small 110-V GMA welder**
Source: Courtesy of Thermal Arc, a Thermadyne Company

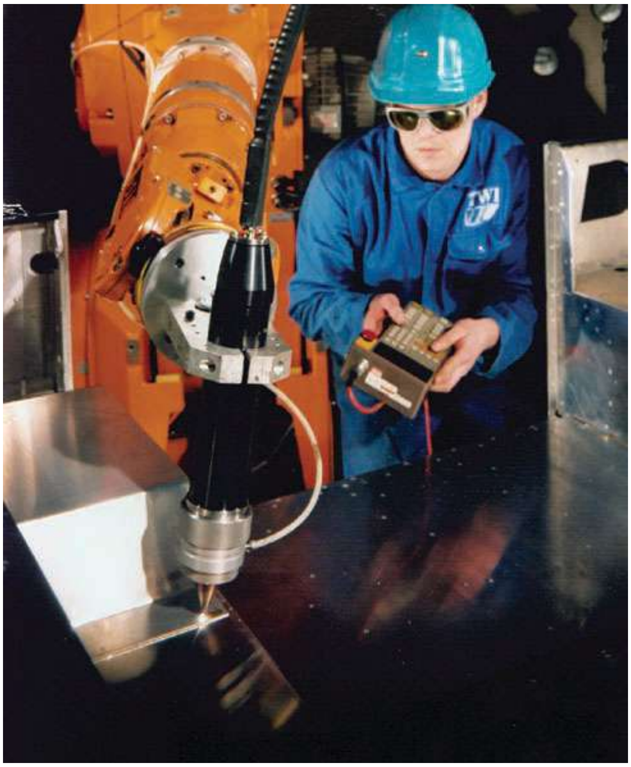

Figure 1.27 **Robot welding**
Source: Courtesy of ESAB Welding & Cutting Products

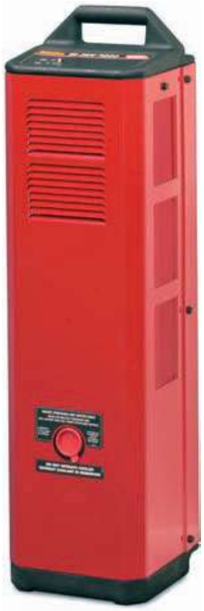

Figure 1.28 **Portable water cooler for GMA welding equipment**
Source: Courtesy of Lincoln Electric Company

Typical GMA welding machines produce a DC welding current ranging from 40 amperes to 600 amperes with 10 volts to 40 volts, depending upon the machine. In the past, some GMA processes used AC welding current, but DCRP is used almost exclusively now. Typical power supplies are shown in Figure 1.29.

Because many GMAW power supplies are used in automation and require long periods of continuous use, it is not unusual for GMA welding machines to have a 100% duty cycle. This allows the machine to be run continuously at its highest-rated output without damage.

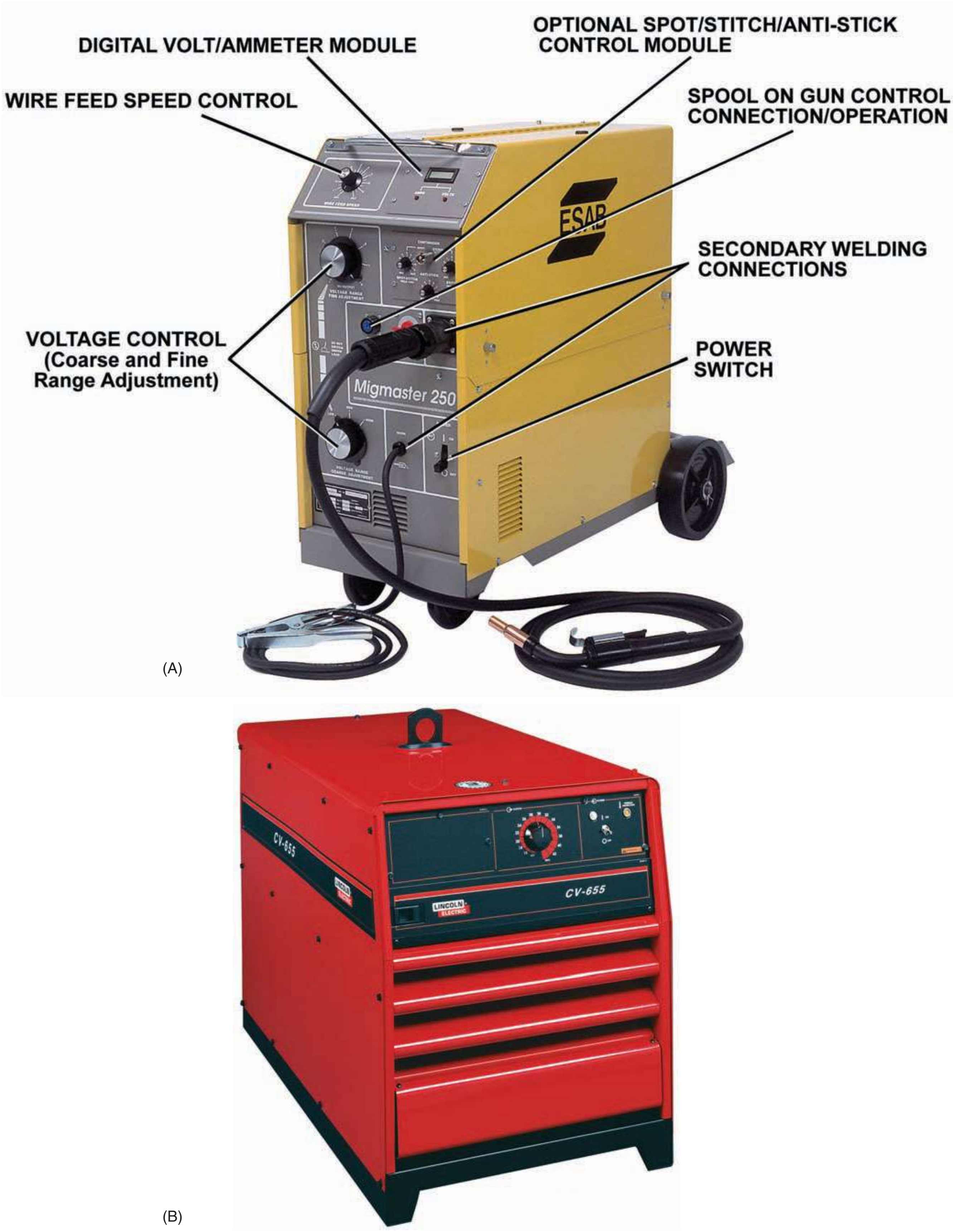
DIGITAL VOLT/AMMETER MODULE
OPTIONAL SPOT/STITCH/ANTI-STICK CONTROL MODULE
WIRE FEED SPEED CONTROL
SPOOL ON GUN CONTROL CONNECTION/OPERATION
SECONDARY WELDING CONNECTIONS
VOLTAGE CONTROL (Coarse and Fine Range Adjustment)
POWER SWITCH
ESAB
Migmaster 250
(A)
CV-655
LINCOLN ELECTRIC
(B)

Figure 1.29 **Two power supplies for multipurpose GMAW applications**
(A) An expensive 200-ampere constant-voltage power supply, and (B) a 650-ampere constant-voltage and constant-current power supply.
Source: (A) Courtesy of ESAB Welding & Cutting Products (B) Courtesy of Lincoln Electric Company

Electrode (Wire) Feed Unit

The purpose of the electrode feeder is to provide a steady and reliable supply of wire to the weld. Slight changes in the rate at which the wire is fed have distinct effects on the weld.

The motor used in a feed unit is usually a DC type that can be continuously adjusted over the desired range. Figure 1.30 and Figure 1.31 show typical wire-feed units and accessories.

Push-type Feed System

The wire rollers are clamped securely against the wire to provide the necessary friction to push the wire through the conduit to the gun. The pressure applied on the wire can be adjusted. A groove is provided in the roller to aid in alignment and to lessen the chance of slippage. Most manufacturers provide rollers with smooth or knurled U-shaped or V-shaped grooves, Figure 1.32. Knurling (a series of ridges cut into the groove) helps grip larger-diameter wires so that they can be pushed along more easily. Soft wires, such as aluminum, are easy to damage if knurled rollers are used. Soft wires are best used with U-grooved rollers. Even V-grooved rollers can distort the surface of the wire, causing problems. V-grooved rollers are best suited for hard wires, such as mild steel and stainless steel. It is also important to use the correct-size grooves in the rollers.

Variations of the push-type electrode wire feeder include the pull type and push-pull type. The difference is in the size and location of the drive rollers. In the push-type system, the electrode must have enough strength to be pushed through the conduit without kinking. Mild steel and stainless steel can be readily pushed 15 to 20 ft (4 to 6 m), but aluminum is much harder to push more than 10 ft (3 m).

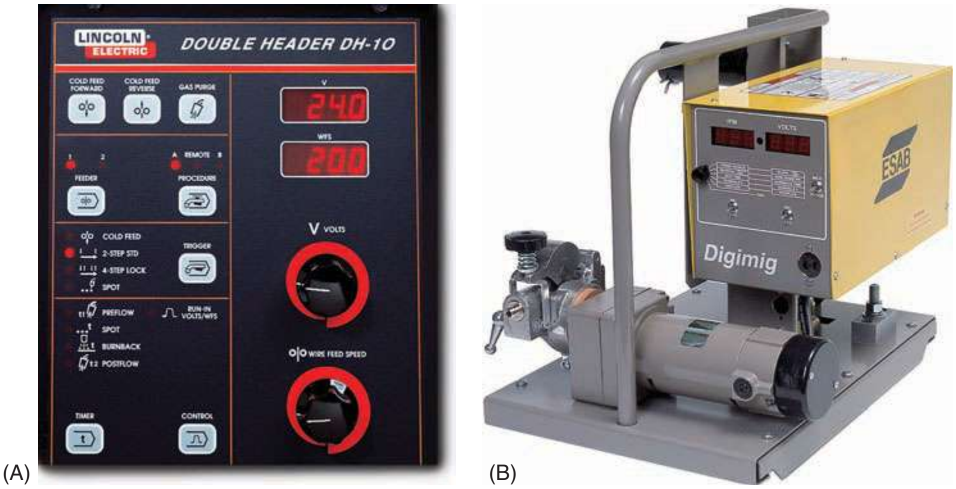
(A)
LINCOLN ELECTRIC
DOUBLE HEADER DH-10
COLD FEED FORWARD
COLD FEED REVERSE
GAS PURGE
V
240
WFS
200
FEEDER
PROCEDURE
COLD FEED
2-STEP STD
4-STEP LOCK
SPOT
TRIGGER
V VOLTS
PREFLOW
SPOT
BURNBACK
POSTFLOW
RUN-IN VOLTS/WFS
WIRE FEED SPEED
TIMER
CONTROL
(B)
ESAB
Digimig

Figure 1.30 Examples of wire feeders
(A) A 90-ampere power supply and wire feeder for welding sheet steel with carbon dioxide shielding. (B) Modern wire feeder with digital preset and readout of wire-feed speed and closed-loop control.
Source: (A) Courtesy of Lincoln Electric Company (B) Courtesy of ESAB Welding & Cutting Products

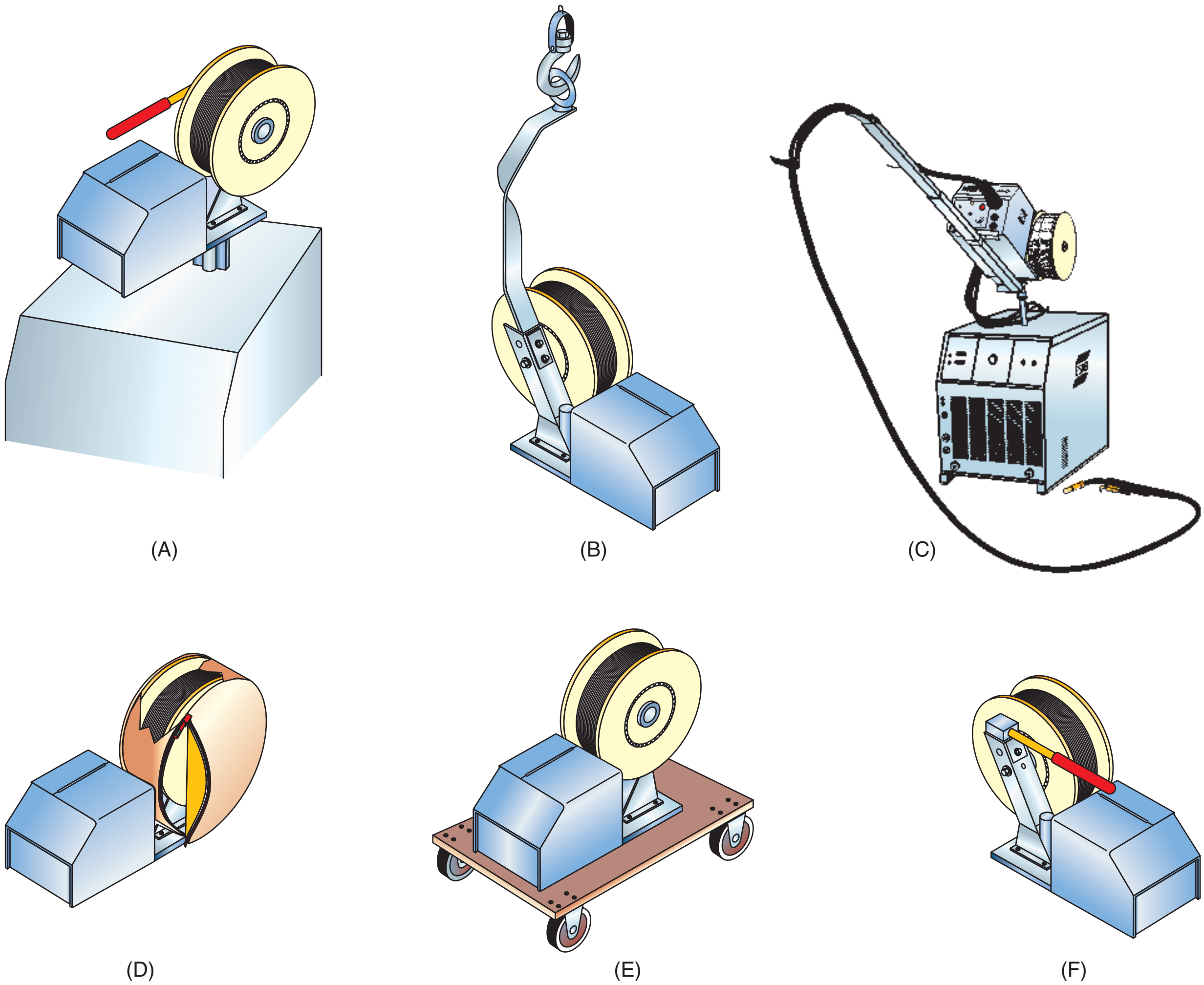

Figure 1.31 **A variety of accessories are available for most electrode feed systems.** (A) Swivel post, (B) boom hanging bracket, (C) counterbalance mini-boom, (D) spool cover, (E) wire feeder wheel cart, and (F) carrying handle.

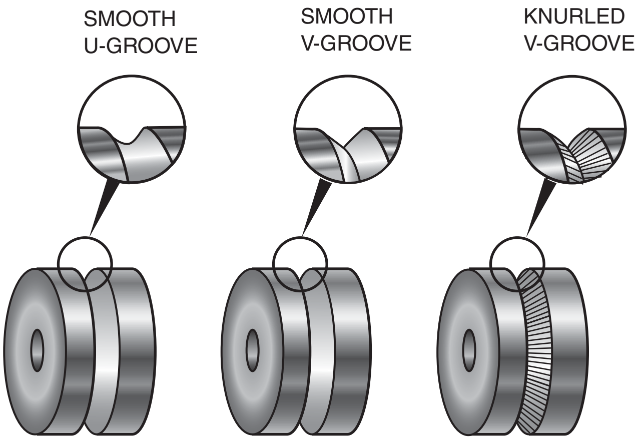
SMOOTH U-GROOVE
SMOOTH V-GROOVE
KNURLED V-GROOVE

Figure 1.32 **Feed rollers**

Pull-type Feed System

In pull-type systems, a smaller but higher-speed motor is located in the gun to pull the wire through the conduit. Using this system, it is possible to move even soft wire over great distances. The disadvantages are that the gun is heavier and more difficult to use, rethreading the wire takes more time, and the operating life of the motor is shorter.

Push-pull-type Feed System

Push-pull-type feed systems use a synchronized system with feed motors located at both ends of the electrode conduit, Figure 1.33. This system can be used to move any type of wire over long distances by periodically installing a feed roller into the electrode conduit. Compared to the pull-type system, the advantages of this system include the ability to move wire over longer distances, faster rethreading, and increased motor life due to the reduced load. A disadvantage is that the system is more expensive.

Linear Electrode Feed System

Linear electrode feed systems use a different method to move the wire and change the feed speed. Standard systems use rollers that pinch the wire between the rollers. A system of gears is used between the motor and rollers to provide roller speed within the desired range. The linear feed system does not have gears or conventional-type rollers.

The linear feed system uses a small motor with a hollow armature shaft through which the wire is fed. The rollers are attached so that they move around the wire. Changing the roller pitch (angle) changes the speed at which the wire is moved without changing the motor speed.

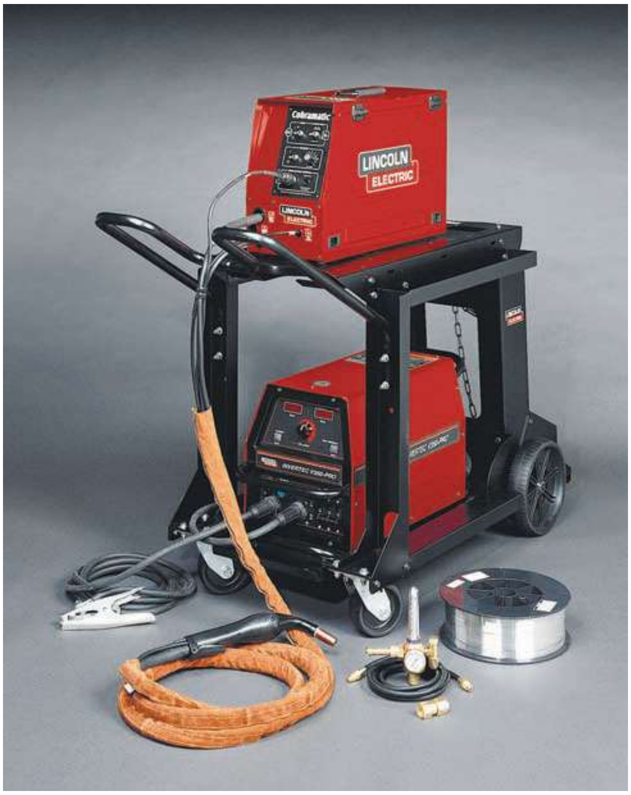

Figure 1.33 **Wire-feed system that enables the wire to be moved through a longer cable**
Source: Courtesy of Lincoln Electric Company

This system works in the same way that changing the pitch on a screw, either coarse threads or fine threads, affects the rate that the screw will move through a spinning nut.

The advantage of a linear system is that the bulky system of gears is eliminated, thus reducing weight, size, and wasted power. The motor operates at a constant high speed where it is more efficient. The reduced size allows the system to be housed in the gun or within an enclosure in the cable. Several linear wire feeders can be synchronized to provide an extended operating range. The disadvantage of a linear system is that the wire may become twisted as it is moved through the feeder.

Spool Gun

A spool gun is a compact, self-contained system consisting of a small drive system and a wire supply, Figure 1.34A. This system allows the welder to move freely around a job with only a power lead and shielding gas hose to manage. The major control system is usually mounted on the welder. The feed rollers and motor are found in the gun just behind the

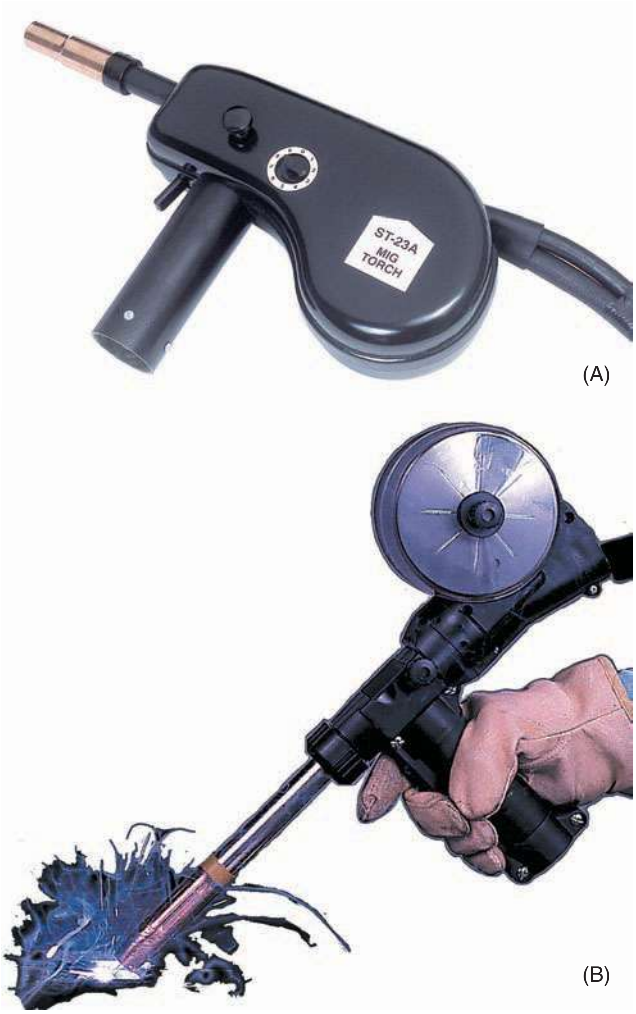

Figure 1.34 **Feeder/guns for GMA welding**
Source: Courtesy of ESAB Welding & Cutting Products

nozzle and contact tube, Figure 1.34B. Because of the short distance the wire must be moved, very soft wires (aluminum) can be used. A small spool of welding wire is located just behind the feed rollers. The small spools of wire required in these guns are often very expensive. Although the guns are small, they feel heavy when being used.

Electrode Conduit

The electrode conduit or liner guides the welding wire from the feed rollers to the gun. It may be encased in a lead that contains the shielding gas.

Power cable and gun switch circuit wires are contained in a conduit that is made of a tightly wound coil having the needed flexibility and strength. The steel conduit may have a nylon or Teflon liner to protect soft, easily scratched metals, such as aluminum, as they are fed.

If the conduit is not an integral part of the lead, it must be firmly attached to both ends of the lead. Failure to attach the conduit can result in misalignment, which causes additional drag or makes the wire jam completely. If the conduit does not extend through the lead casing to make a connection, it can be drawn out by tightly coiling the lead, Figure 1.35. Coiling will force the conduit out so that it can be connected. If the conduit is too long for the lead, it should be cut off and filed smooth. Too long a lead will bend and twist inside the conduit, which may cause feed problems.

Welding Gun

The welding gun attaches to the end of the power cable, electrode conduit, and shielding gas hose, Figure 1.36. It is used by the welder to produce the weld. A trigger switch is used to start and stop the weld cycle. The gun also has a contact tube, which is used to transfer welding current to the electrode moving through the gun, and a gas nozzle, which directs the shielding gas onto the weld, Figure 1.37.

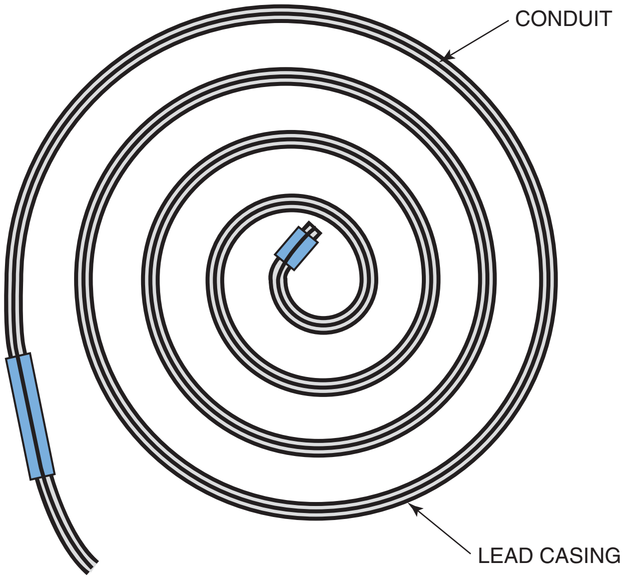

Figure 1.35 **Tightly coiled lead casing will force the liner out of the gun**

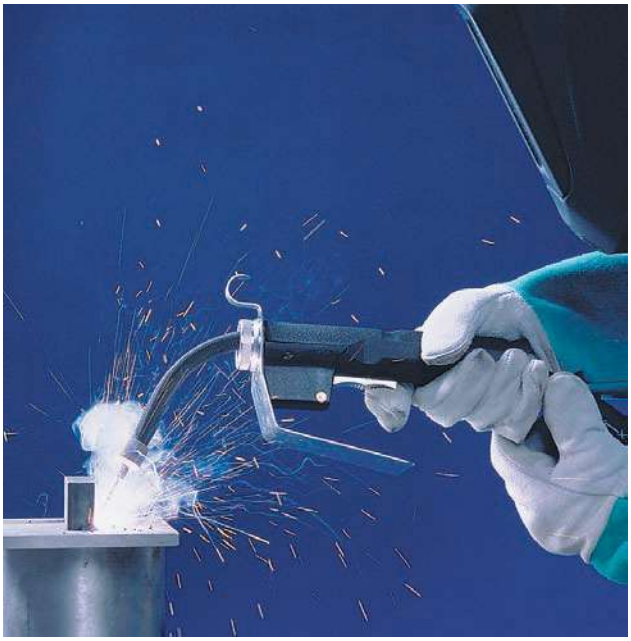

Figure 1.36 **A typical GMA welding gun**
Guns like this are used for most welding processes with a heat shield attached to protect the welder's gloved hand from intense heat generated when welding with high amperages.
Source: Courtesy of Tweco, a Thermadyne Company

SPOT WELDING

GMA can be used to make high-quality arc spot welds. Welds can be made using standard or specialized equipment, Figure 1.38. The arc spot weld produced by GMAW differs from electric resistance spot welding. The GMAW spot weld starts on one surface of one member and burns through to the other member, Figure 1.39. Fusion between the members occurs, and a small nugget is left on the metal surface.

GMA spot welding has some advantages such as the following: (1) welds can be made in thin-to-thick materials; (2) the weld can be made when only one side of the materials to be welded is accessible; and (3) the weld can be made when there is paint on the interfacing surfaces. The arc spot weld can also be used to assemble parts for welding to be done at a later time.

Thin metal can be attached to thicker sections using an arc spot weld. If a thin-to-thick butt, lap, or tee joint is to be welded with complete joint penetration, often the thin material will burn back, leaving a hole, or there will not be enough heat to melt the thick section. With an arc spot weld, the burning back of the thin material allows the thicker metal to be melted. As more metal is added to the weld, the burn-through is filled, Figure 1.39.

The GMA spot weld is produced from only one side. Therefore, it can be used on awkward shapes and in cases where the other side of the surface being welded should not be damaged. This makes it an excellent process for auto body repair. In addition, because the metals are melted

Caution

Safety glasses and/or flash glasses must be worn to protect the eyes from flying sparks.

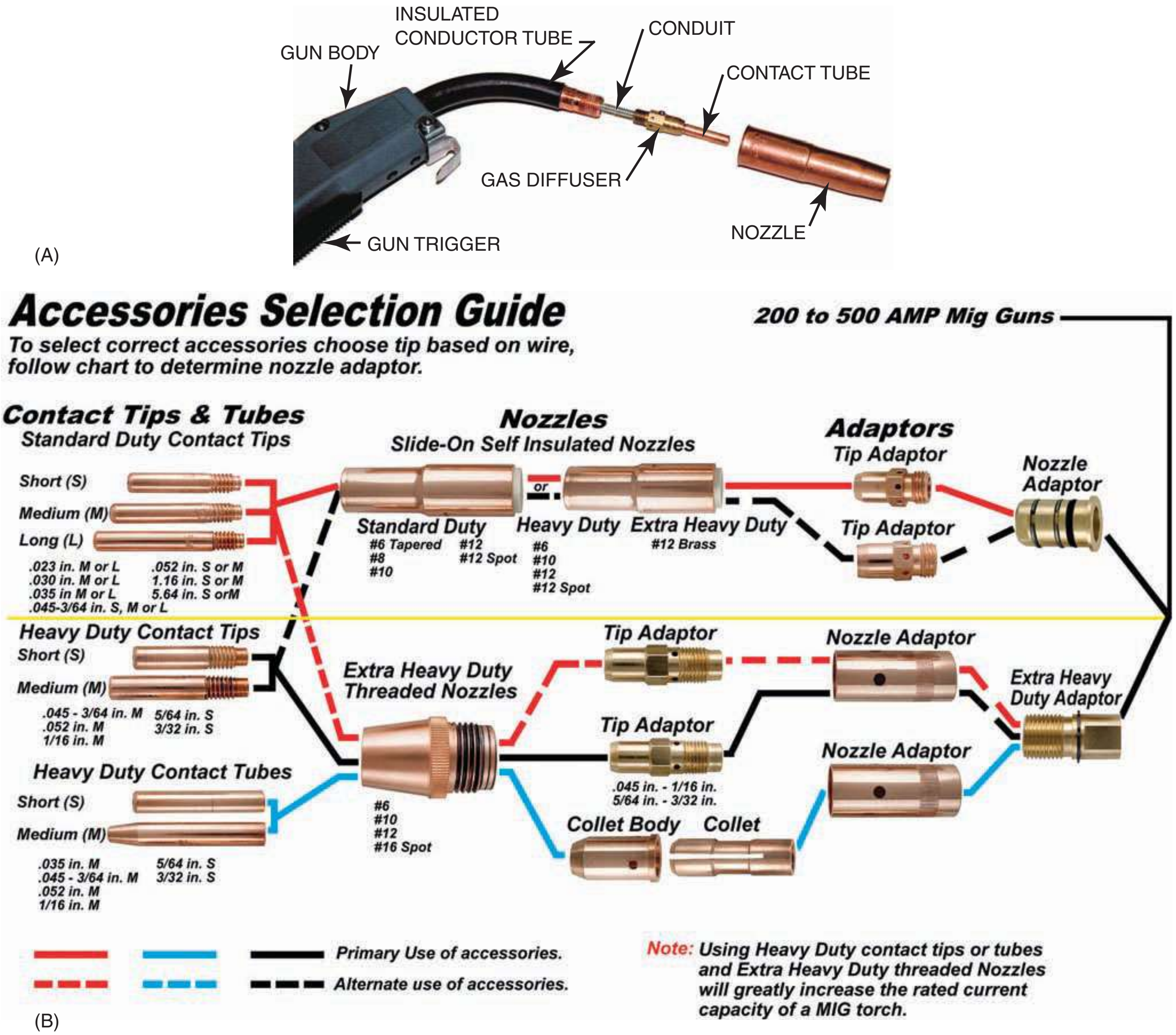
(A)
GUN BODY
INSULATED CONDUCTOR TUBE
CONDUIT
CONTACT TUBE
GAS DIFFUSER
NOZZLE
GUN TRIGGER
Accessories Selection Guide
To select correct accessories choose tip based on wire, follow chart to determine nozzle adaptor.
200 to 500 AMP Mig Guns
Contact Tips & Tubes
Standard Duty Contact Tips
Short (S)
Medium (M)
Long (L)
.023 in. M or L
.030 in. M or L
.035 in M or L
.045-3/64 in. S, M or L
.052 in. S or M
1.16 in. S or M
5.64 in. S orM
Nozzles
Slide-On Self Insulated Nozzles
or
Standard Duty
#6 Tapered
#8
#10
#12
#12 Spot
Heavy Duty
#6
#10
#12
#12 Spot
Extra Heavy Duty
#12 Brass
Adaptors
Tip Adaptor
Tip Adaptor
Nozzle Adaptor
Heavy Duty Contact Tips
Short (S)
Medium (M)
.045 - 3/64 in. M
.052 in. M
1/16 in. M
5/64 in. S
3/32 in. S
Extra Heavy Duty Threaded Nozzles
#6
#10
#12
#16 Spot
Tip Adaptor
Tip Adaptor
.045 in. - 1/16 in.
5/64 in. - 3/32 in.
Nozzle Adaptor
Nozzle Adaptor
Extra Heavy Duty Adaptor
Collet Body
Collet
Heavy Duty Contact Tubes
Short (S)
Medium (M)
.035 in. M
.045 - 3/64 in. M
.052 in. M
1/16 in. M
5/64 in. S
3/32 in. S
Primary Use of accessories.
Alternate use of accessories.
Note: Using Heavy Duty contact tips or tubes and Extra Heavy Duty threaded Nozzles will greatly increase the rated current capacity of a MIG torch.
(B)

Figure 1.37 **GMA welding gun parts**
(A) Typical replaceable parts of a GMA welding gun. (B) Accessories and parts selection guide for a GMA welding gun.
Source: Courtesy of ESAB Welding & Cutting Products

and the molten weld pool is agitated, thin films of paint between the members being joined need not be removed. This is an added benefit for auto body repair work.

Specially designed nozzles provide flash protection, part alignment, and arc alignment, Figure 1.40. As a result, for some small jobs it is possible to perform the weld with only safety glasses. The optional control timer provides weld time and burn-back time. To make a weld, the amperage, voltage, and length of welding time must be set correctly. The burn-back time is a short period at the end of the weld when the wire feed stops but the current does not. This allows the wire to be burned back so it does not stick in the weld, Figure 1.39D.

Caution

It is not advisable for any spot welding work requiring more than just a few spot welds to be done without full welder's safety gear. Prolonged exposure to the reflected ultraviolet light will cause skin burns.

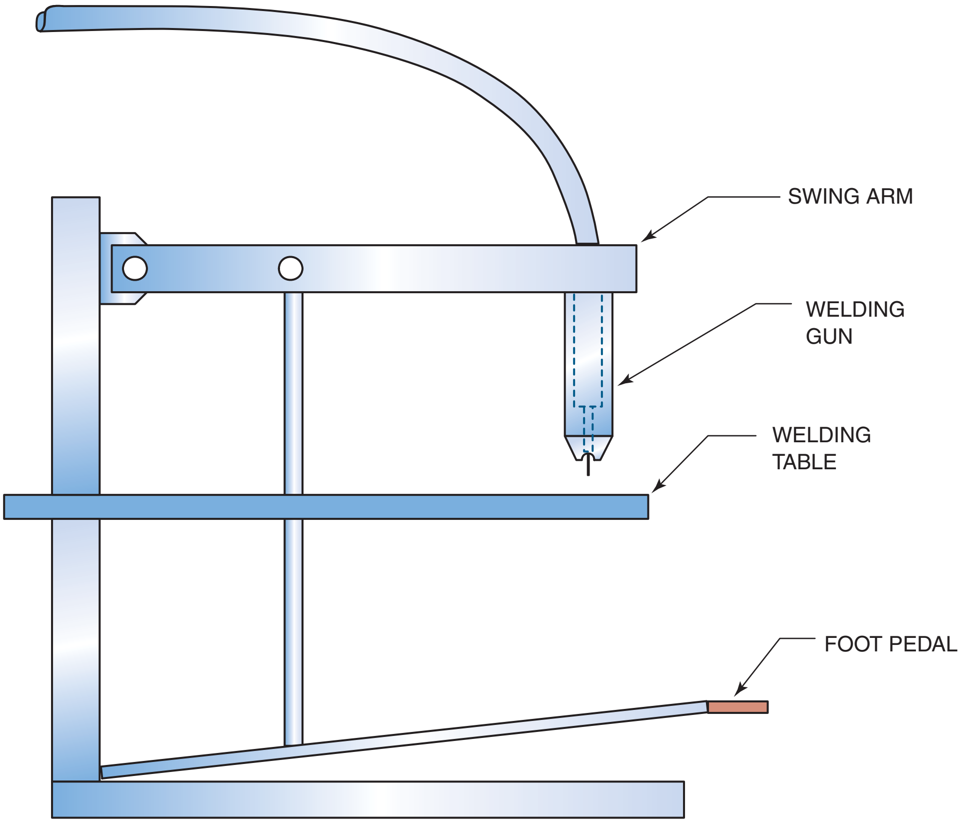

Figure 1.38 **GMA spot welding machine**

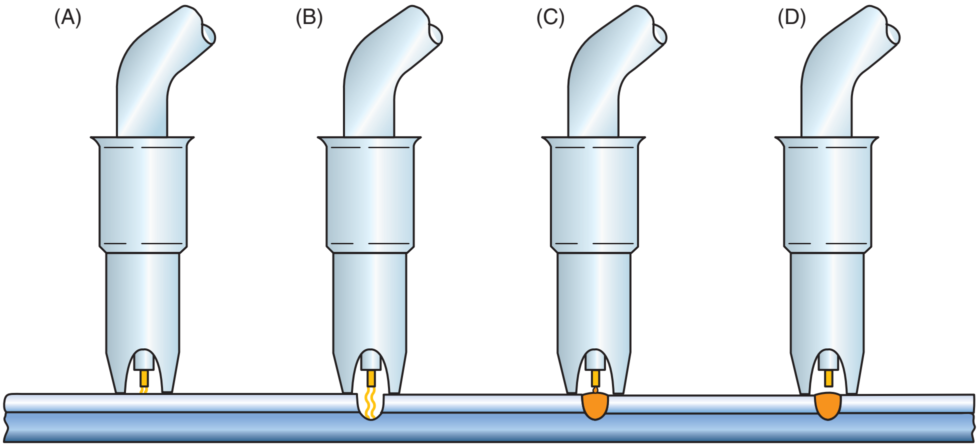

Figure 1.39 **A GMA spot weld**
(A) The arc starts, (B) a hole is burned through the first plate, (C) the hole is filled with weld metal, and (D) the wire feed stops and the arc burns the electrode back.

FACE, EYE, AND EAR PROTECTION

Module 2
Key Indicator 1

Module 5
Key Indicator 1

Face and Eye Protection

Eye protection must be worn in the shop at all times. Eye protection can be safety glasses with side shields, Figure 1.41, goggles, or a full face shield. To give better protection when working in brightly lit areas or outdoors, some welders wear flash glasses, which are special, lightly tinted, safety glasses. These safety glasses provide protection from both flying debris and reflected light.

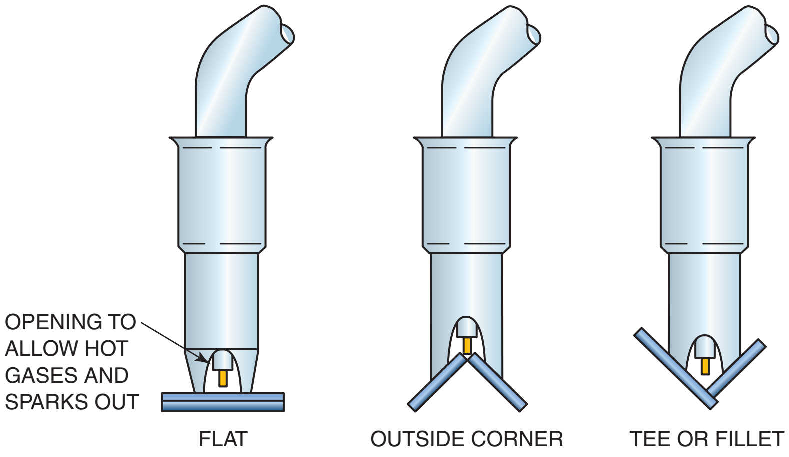

Figure 1.40 Specialized nozzles for GMA spot welding

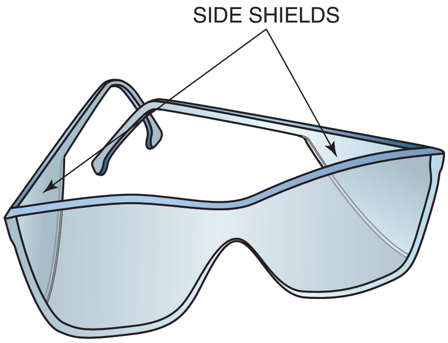

Figure 1.41 Safety glasses with side shields

Suitable eye protection is important because eye damage caused by excessive exposure to arc light is not noticed. Welding light damage often occurs without warning, like a sunburn's effect that is felt the following day. Therefore, welders must take appropriate precautions in selecting filters or goggles that are suitable for the process being used, Figure 1.42. Selecting the correct shade lens is also important, because both extremes of too light or too dark can cause eye strain. New welders often select too dark a lens, assuming it will give them better protection, but this results in eye strain in the same manner as if they were trying to read in a poorly lit room. In reality, any approved arc welding lenses will filter out the harmful ultraviolet light. Select a lens that lets you see comfortably. At the very least, the welder's eyes must not be strained by excessive glare from the arc.

Ultraviolet light can burn the eye in two ways. This light can injure either the white of the eye or the retina, which is the back of the eye, Figure 1.43. Burns on the retina are not painful but may cause some loss of eyesight. The whites of the eyes are very sensitive, and burns are very painful. The eyes are easily infected because, as with any burn, many cells are killed. These dead cells in the moist environment of the eyes will promote the growth of bacteria that cause infection. When the eye is burned, it feels as though there is something in the eye. Without a professional examination, however, it is impossible to tell if there is something in the eye. Because there may actually be something in the eye and because of the high risk of infection, home remedies or other medicines should never be used for eye burns. Anytime you receive an eye injury you should see a doctor.

Welding Helmets

Even with quality **welding helmets**, like those shown in Figure 1.44, the welder must check for potential problems that may occur from accidents or daily use. Small, undetectable leaks of ultraviolet light in an arc welding helmet can cause a welder's eyes to itch or feel sore after a day of welding. To prevent these leaks, make sure the lens gasket is installed correctly, Figure 1.45. The outer and inner clear lenses must be plastic. As shown in Figure 1.46, a lens can be checked for cracks by twisting it between your fingers. Worn or cracked spots on a helmet must be repaired.

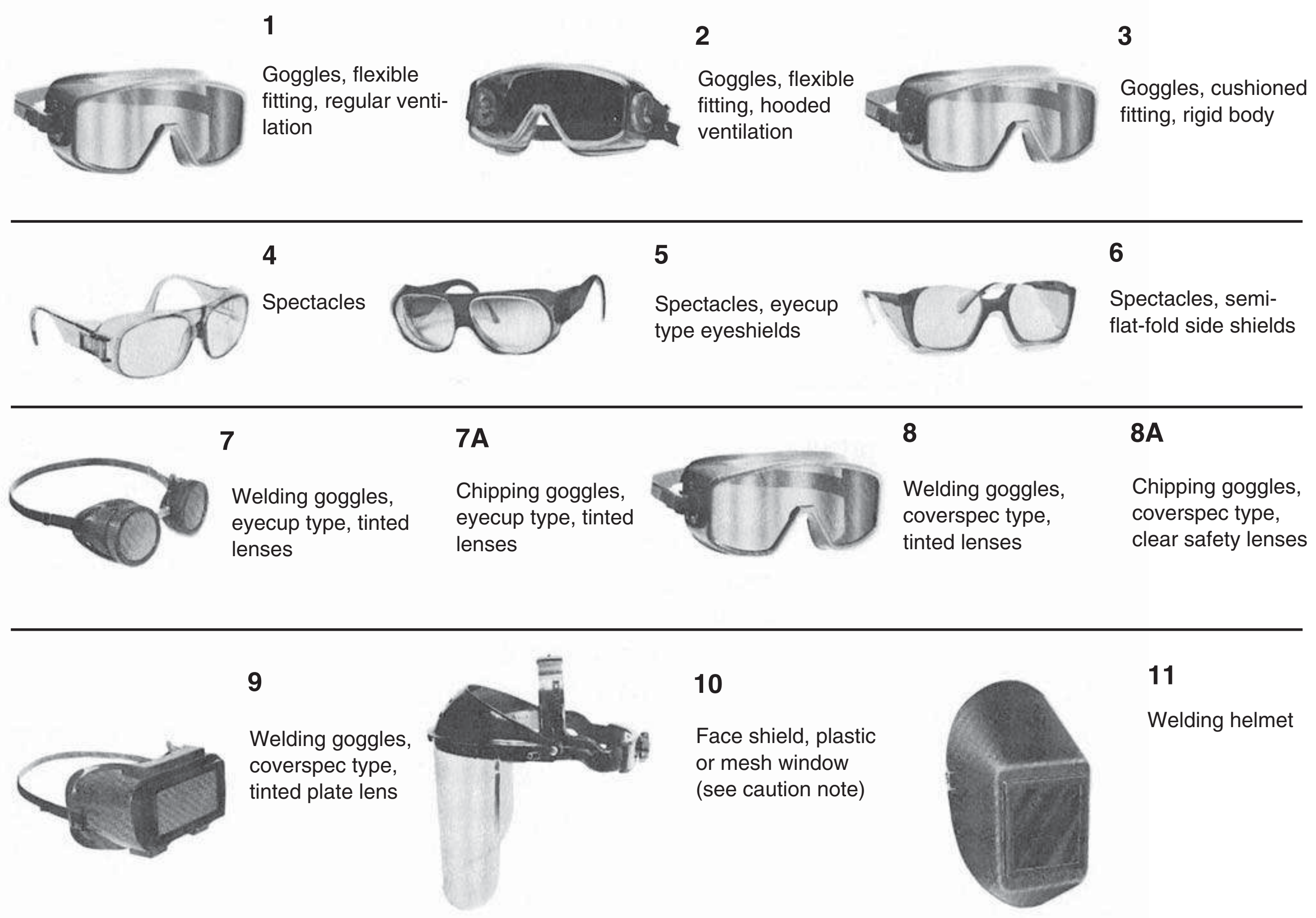
1
Goggles, flexible fitting, regular ventilation
2
Goggles, flexible fitting, hooded ventilation
3
Goggles, cushioned fitting, rigid body
4
Spectacles
5
Spectacles, eyecup type eyeshields
6
Spectacles, semi-flat-fold side shields
7
Welding goggles, eyecup type, tinted lenses
7A
Chipping goggles, eyecup type, tinted lenses
8
Welding goggles, coverspec type, tinted lenses
8A
Chipping goggles, coverspec type, clear safety lenses
9
Welding goggles, coverspec type, tinted plate lens
10
Face shield, plastic or mesh window (see caution note)
11
Welding helmet

**Non-side-shield spectacles are available for limited hazard use requiring only frontal protection.*

Applications

Operation	Hazards	Protectors
Acetylene-burning Acetylene-cutting Acetylene-welding	Sparks, harmful rays, molten metal, flying particles	7,8,9
Chemical handling	Splash, acid burns, fumes	2 (for severe exposure add 10)
Chipping	Flying particles	1,2,4,5,6,7A,8A
Electric (arc) welding	Sparks, intense rays, molten metal	11 (in combination with 4,5,6 in tinted lenses advisable)
Furnace operations	Glare, heat, molten metal	7,8,9 (for severe exposure add 10)
Grinding—light	Flying particles	1,3,5,6 (for severe exposure add 10)
Grinding—heavy	Flying particles	1,3,7A,8A (for severe exposure add 10)
Laboratory	Chemical splash, glass breakage	2 (10 when in combination with 5,6)
Machining	Flying particles	1,3,5,6 (for severe exposure add 10)
Molten metals	Heat, glare, sparks, splash	7,8 (10 in combination with 5,6 in tinted lenses)
Spot welding	Flying particles, sparks	1,3,4,5,6 (tinted lenses advisable, for severe exposure add 10)

CAUTION:

Face shields alone do not provide adequate protection. Plastic lenses are advised for protection against molten metal splash. Contact lenses, of themselves, do not provide eye protection in the industrial sense and shall not be worn in a hazardous environment without appropriate covering safety eyewear.

Figure 1.42 **Huntsman selector chart**

Source: Courtesy of Kedman Co., Huntsman Product Division

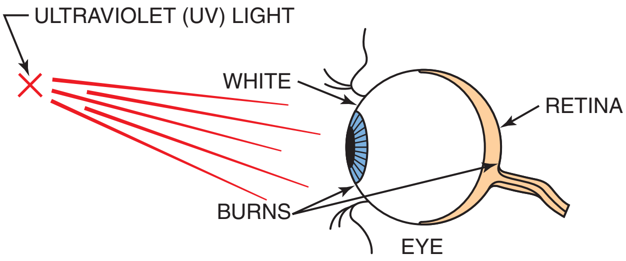

Figure 1.43 **The eye can be burned on the white or on the retina by ultraviolet light**

Figure 1.44 **Typical arc welding helmets for eye and face protection during welding**
Source: Courtesy of Hornell, Inc.

Safety Glasses

Safety glasses with side shields are adequate for general use, but if heavy grinding, chipping, or overhead work is being done, goggles or a full face shield should be worn in addition to safety glasses, Figure 1.47. Safety glasses are best for general protection. They must be worn under an arc welding helmet at all times.

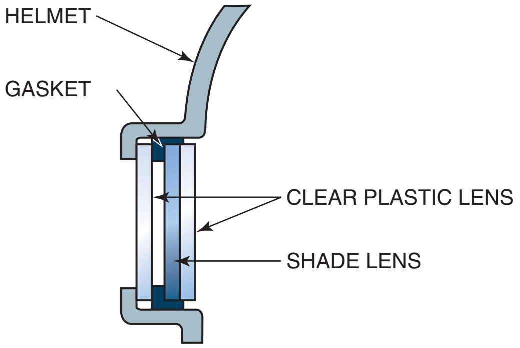

Figure 1.45 Correct placement of the gasket around the shade lens
Correct gasket placement around the shade lens of a welding helmet is important because it can stop ultraviolet light from bouncing around the lens assembly.

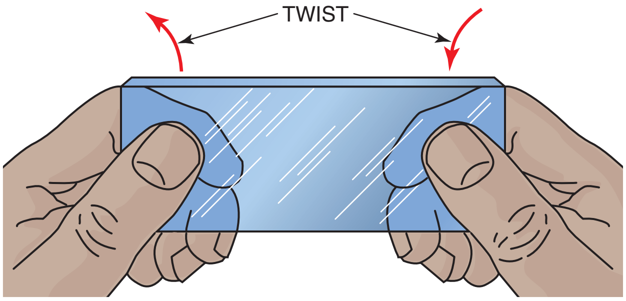

Figure 1.46 **To check the shade lens for possible cracks, gently twist it**

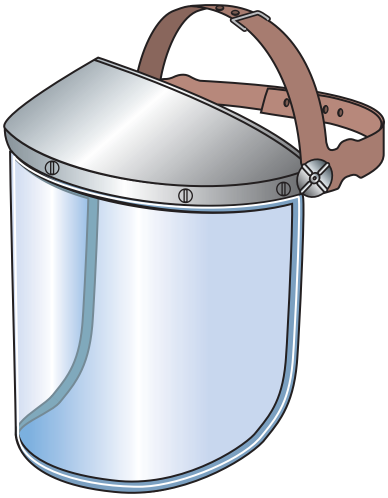

Figure 1.47 **Full face shield**

Ear Protection

The welding environment can be very noisy. The sound level is at times high enough to cause pain and some loss of hearing if the welder's ears are unprotected. Hot sparks can also drop into an open ear, causing severe burns.

Ear protection is available in several forms. One form of protection is earmuffs that cover the outer ear completely, Figure 1.48. Another form

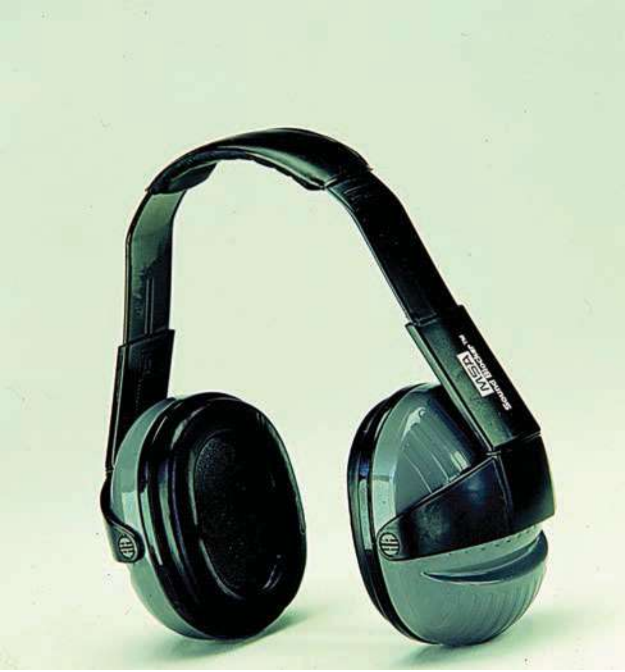

Figure 1.48 **Earmuffs**
Earmuffs provide complete ear protection and can be worn under a welding helmet.
Source: Courtesy of Mine Safety Appliances Company

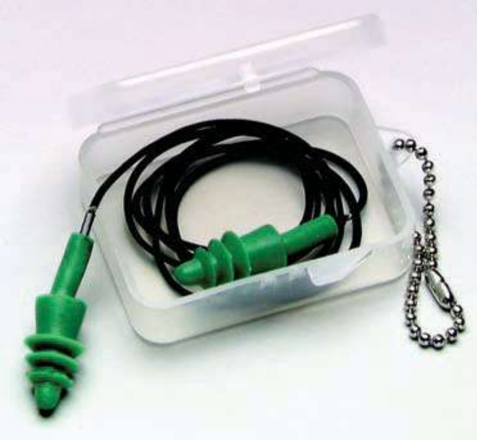

Figure 1.49 **Earplugs**
Earplugs are used as protection from noise only.
Source: Courtesy of Mine Safety Appliances Company

of protection is earplugs that fit into the ear canal, Figure 1.49. Both of these protect a person's hearing, but only the earmuffs protect the outer ear from burns.

GENERAL WORK CLOTHING

Special protective clothing cannot be worn at all times. It is, therefore, important to choose general work clothing that will minimize the possibility of getting burned because of the high temperature and amount of hot sparks, metal, and slag produced during welding, cutting, or brazing.

Caution

Damage to your hearing caused by high sound levels may not be detected until later in life, and the resulting loss in hearing is permanent. Your hearing will not improve with time, and each exposure to high levels of sound will further damage your hearing.

Work clothing must also stop ultraviolet light from passing through it. This is accomplished if the material chosen is a dark color, thick, and tightly woven. The best choice is 100% wool but it is difficult to find. Another good choice is 100% cotton clothing, the most popular fabric used.

You must avoid wearing synthetic materials including nylon, rayon, and polyester. They can easily melt or catch fire. Some synthetics produce a hot, sticky residue that can make burns more severe. Others may produce poisonous gases.

The following are some guidelines for selecting work clothing:

- Shirts must be long-sleeved to protect the arms, have a high-buttoned collar to protect the neck, Figure 1.50, be long enough to tuck into the pants to protect the waist, and have flaps on the pockets to keep sparks out (or have no pockets).
- Pants must have legs long enough to cover the tops of the boots and must be without cuffs that would catch sparks.
- Boots must have high tops to keep out sparks, have steel toes to prevent crushed toes, Figure 1.51, and have smooth tops to prevent sparks from being trapped in seams.
- Caps should be thick enough to prevent sparks from burning the top of a welder's head.

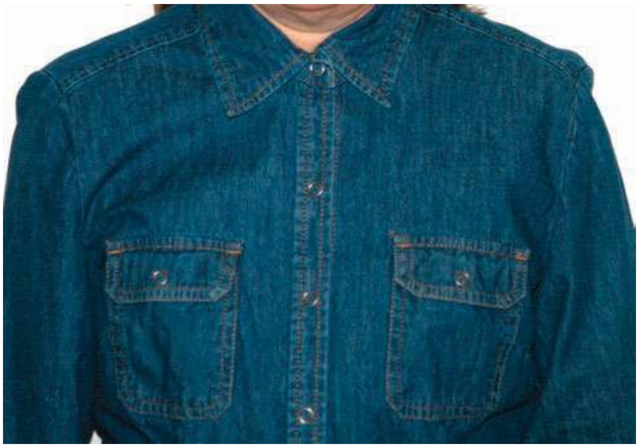

Figure 1.50 Neck protection
The top button of the shirt worn by the welder should always be buttoned in order to avoid severe neck burns.
Source: Courtesy of Larry Jeffus

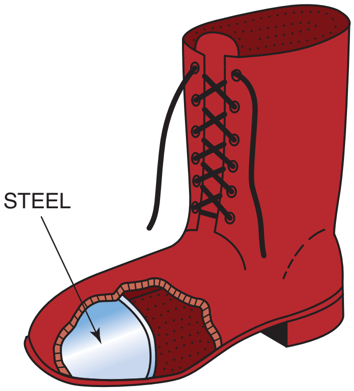

Figure 1.51 Safety boots with steel toes are required by many welding shops

All clothing must be free of frayed edges and holes. The clothing must be relatively tight-fitting in order to prevent excessive folds or wrinkles that might trap sparks.

Some welding clothes have pockets on the inside to prevent the pockets from collecting sparks. However, it is not safe to carry butane lighters or matches in these or any pockets while welding. Lighters and matches can easily catch on fire or explode if they are subjected to the heat and sparks of welding.

Caution

There is no safe place to carry butane lighters or matches while welding or cutting. They can catch fire or explode if subjected to welding heat or sparks. Butane lighters may explode with the force of a quarter of a stick of dynamite. Matches can erupt into a ball of fire. Both butane lighters and matches must always be removed from the welder's pockets and placed a safe distance away before any work is started.

SPECIAL PROTECTIVE CLOTHING

General work clothing is worn by each person in the shop. In addition to this clothing, extra protection is needed for each person who is in direct contact with hot materials. Leather is often the best material to use, as it is lightweight, is flexible, resists burning, and is readily available. Synthetic insulating materials are also available. Ready-to-wear leather protection includes capes, jackets, aprons, sleeves, gloves, caps, pants, knee pads, and spats, among other items.

Hand Protection

All-leather, gauntlet-type gloves should be worn when doing any welding, Figure 1.52. Gauntlet gloves that have a cloth liner for insulation are best for hot work. Noninsulated gloves will give greater flexibility for fine work. Some leather gloves are available with a canvas gauntlet top, which should be used for light work only.

When a great deal of manual dexterity is required for gas tungsten arc welding, brazing, soldering, oxyfuel gas welding, and other delicate processes, soft leather gloves may be used, Figure 1.53. All-cotton gloves are sometimes used when doing very light welding.

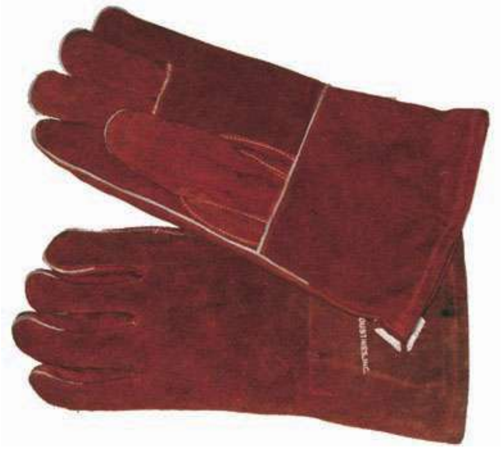

Figure 1.52 **All-leather, gauntlet-type welding gloves**
Source: Courtesy of Larry Jeffus

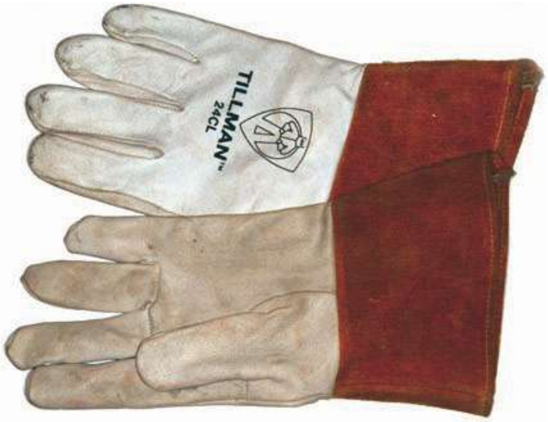

Figure 1.53 **Soft leather gloves**
For welding that requires a great deal of manual dexterity, soft leather gloves can be worn.
Source: Courtesy of Larry Jeffus

Body Protection

Full leather jackets and capes will protect a welder's shoulders, arms, and chest, Figure 1.54. A jacket, unlike the cape, protects a welder's back and complete chest. A cape is open and much cooler but offers less protection. The cape can be used with a bib apron to provide some additional protection while leaving the back cooler. Either the full jacket or the cape with a bib apron should be worn for any out-of-position work.

Waist and Lap Protection

Bib aprons or full aprons will protect a welder's lap. Welders will especially need to protect their laps if they squat or sit while working and when they bend over or lean against a table.

Arm Protection

For some vertical welding, a full or half sleeve can protect a person's arm, Figure 1.55. The sleeves work best if the work level is not above the welder's chest. Work levels higher than this usually require a jacket or cape to keep sparks off the welder's shoulders.

Leg and Foot Protection

When heavy cutting or welding is being done and a large number of sparks are falling, leather pants and spats should be used to protect the welder's legs and feet. If the weather is hot and full leather pants are uncomfortable, leather aprons with leggings are available. Leggings can be strapped to the legs, leaving the back open. Spats will prevent sparks from burning through the front of lace-up boots.

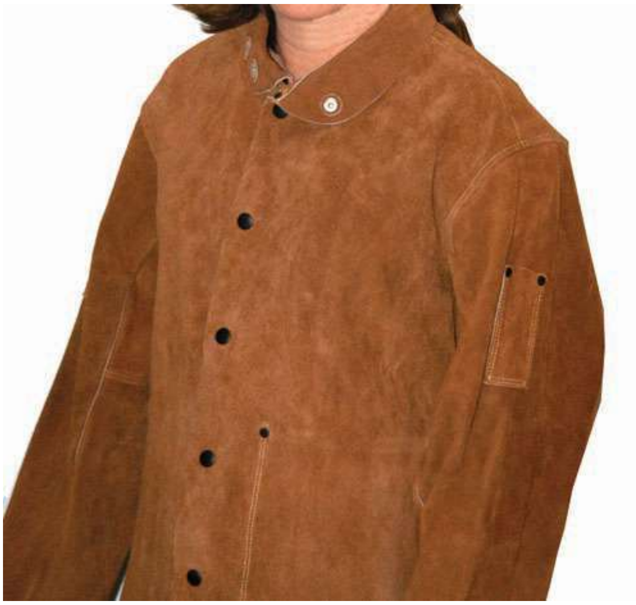

Figure 1.54 **Full leather jacket**
Source: Courtesy of Larry Jeffus

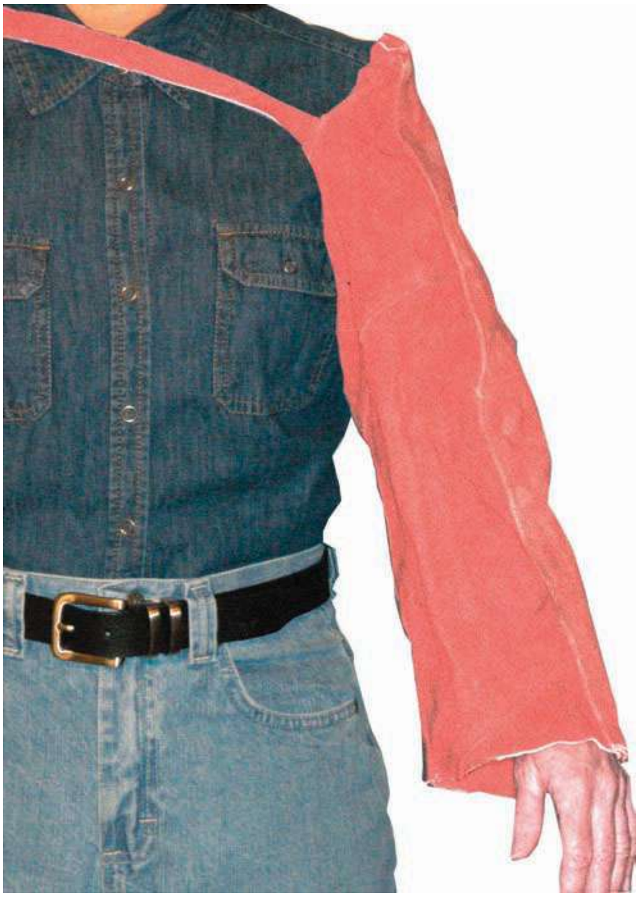

Figure 1.55 **Full leather sleeve**
Source: Courtesy of Larry Jeffus

SUMMARY

The keys to producing quality GMA welds are equipment, setup, and adjustments. Once you have mastered these skills the only remaining obstacle to your producing consistent, uniform, high-quality welds is your ability to follow, or track, the joint consistently. Some welders find that lightly dragging their glove along the metal surface or edge of the fabrication can aid them in controlling the weld consistency. One of the advantages of the GMA welding process is its ability to produce long, uninterrupted welds. However, this often leads to welder fatigue. Finding a comfortable welding position that you can maintain for several minutes at a time will both improve your weld quality and reduce your fatigue.

Selecting the proper method of metal transfer—short arc, globular, or spray transfer—is normally done by the welding shop foreman or supervisor. He or she makes these selections based on the material being welded, the welding position, and other factors, including welding procedure specifications and applicable codes. A welder must be proficient with each of the various methods of metal transfer. It is therefore important that you spend time practicing and developing your skills with each of these processes.

REVIEW

1. Why is usage of the term *GMAW* preferable to *MIG* for gas metal arc welding?
2. Using Table 1.1, answer the following:
   a. What maintains the arc in machine welding?
   b. What feeds the filler metal in manual welding?
   c. What provides the joint travel in automatic welding?
   d. What provides the joint guidance in semiautomatic welding?
3. What factors have led to increased usage of the GMAW process?
4. In what form is metal transferred across the arc in the axial spray metal transfer method of GMA welding?
5. What three conditions are required for the spray transfer process to occur?
6. Using Table 1.3, answer the following:
   a. What should the wire-feed speed and voltage ranges be to weld 1/8-in. metal with 0.035-in. wire using argon shielding gas?
   b. What should the amperage and voltage ranges be when using 98% Ar + 2% O_2 to weld 1/4-in. metal with 0.045-in. wire?
7. What ranges does the pulsed-arc metal transfer shift between?
8. How do frequency, amplitude, and width of the pulses affect the GMA pulse welding process?
9. How have electronics helped the pulsed-arc process?
10. Why is helium added to argon when making some spray or pulsed-spray transfer welds?
11. Why does DCEP help with welds on metals such as aluminum?
12. Why is CO_2 added to argon when making GMA spray transfer welds?
13. Why should CO_2 not be used to weld stainless steel?
14. How is the metal transferred from the electrode to the plate during the GMAW-S process?
15. Using Figure 1.12, what should be the approximate voltage at 175 amps at 200 in./min when using 0.035-in. ER70S-6 electrode wire?
16. Using Table 1.8, what would the amperage be for 0.035-in. (0.9-mm) wire at 200 in./min (5 m/min)?
17. Using Table 1.9, what shielding gas should be used for welding on copper?
18. What may happen if the GMA welding electrode is allowed to strike the base metal outside the molten weld pool?
19. What effect does shortening the electrode extension have on weld penetration?
20. Describe the weld produced by a backhand welding angle.
21. Describe the weld produced by a forehand welding angle.
22. What components make up a GMA welding system?
23. Why must GMA welders have a 100% duty cycle?
24. What can happen if rollers of the wrong shape are used on aluminum wire?
25. Where is the drive motor located in a pull-type wire-feed system?
26. How is the wire-feed speed changed with a linear feed system?
27. What type of liner should be used for aluminum wire?
28. What parts of a typical GMA welding gun can be replaced?
29. Describe the spot welding process using a GMA welder.

CHAPTER 2

Gas Metal Arc Welding

OBJECTIVES

After completing this chapter, the student should be able to

- set up a GMA welding work station
- thread the electrode wire on a GMAW machine
- set the shielding gas flow rate on a GMAW machine
- use various settings on a GMA welding machine, and compare the effects on a weld
- demonstrate the effect of changing the electrode extension on a weld
- describe the effects of changing the welding gun angle on the weld bead
- list six variables to consider when selecting the shielding gas for a particular application
- evaluate weld beads made with various shielding gas mixtures
- explain why hot-rolled steel should be cleaned to bright metal before welding
- make GMA welds in butt joints, lap joints, and tee joints in all positions that will pass a specified standard's visual or destructive examination criteria

KEY TERMS

bird-nesting
cast
conduit liner
contact tube
feed rollers
flow rate
spool drag
wire-feed speed

AWS SENSE EG2.0

Key Indicators Addressed in this Chapter:

Module 1: Occupational Orientation

Key Indicator 1: Prepares time or job cards, reports or records
Key Indicator 2: Performs housekeeping duties
Key Indicator 3: Follows verbal instructions to complete work assignments
Key Indicator 4: Follows written details to complete work assignments

Module 2: Safety and Health of Welders

Key Indicator 1: Demonstrates proper use and inspection of personal protection equipment (PPE)
Key Indicator 2: Demonstrates proper work area operation

Key Indicator 3: Demonstrates proper use and inspection of ventilation equipment

Key Indicator 4: Demonstrates proper hot zone operation

Key Indicator 7: Demonstrates proper inspection and operation of equipment for each welding or thermal cutting process used (This is best done as part of the process module/unit for each of the required welding and thermal cutting processes.)

Module 3: Drawing and Welding Symbol Interpretation

Key Indicator 1: Interprets basic elements of a drawing or sketch

Key Indicator 2: Interprets welding symbol information

Key Indicator 3: Fabricates parts from a drawing or sketch

Module 5: Gas Metal Arc Welding [GMAW-S, GMAW (spray)]

Key Indicator 1: Performs safety inspections of GMAW equipment and accessories

Key Indicator 2: Makes minor external repairs to GMAW equipment and accessories

Short-circuit Transfer

Key Indicator 3: Sets up for GMAW-S operations on carbon steel

Key Indicator 4: Operates GMAW-S equipment on carbon steel

Key Indicator 5: Makes GMAW-S fillet welds, in all positions, on carbon steel

Key Indicator 6: Makes GMAW-S groove welds, in all positions, on carbon steel

Key Indicator 7: Passes GMAW-S welder performance qualification test (workmanship sample) on carbon steel

Spray Transfer

Key Indicator 8: Sets up for GMAW (spray) operations on carbon steel

Key Indicator 9: Operates GMAW (spray) equipment on carbon steel

Key Indicator 10: Makes GMAW (spray) fillet welds, in the 1F and 2F positions, on carbon steel

Key Indicator 11: Makes GMAW (spray) groove welds, in the 1G position, on carbon steel

Key Indicator 12: Passes GMAW (spray) welder performance qualification test (workmanship sample) on carbon steel

Module 9: Welding Inspection and Testing Principles

Key Indicator 1: Examines cut surfaces and edges of prepared base metal parts.

Key Indicator 2: Examines tacks, root passes, intermediate layers, and completed welds

INTRODUCTION

Performing a satisfactory GMA weld requires more than just manipulative skill. The setup, voltage, amperage, electrode extension, and welding angle, as well as other factors, can dramatically affect the weld produced. The very best welding conditions are those that will allow a welder to produce the largest quantity of successful welds in the shortest period of time with the highest productivity. Because these are semiautomatic or automatic processes, increased productivity may require only that the welder increase the travel speed and current. This does not mean that the welder will work harder but, rather, that the welder will work more productively, resulting in a greater cost efficiency.

The more cost efficient welders can be, the more competitive they and their companies become. This can make the difference between being awarded a job or losing work.

SETUP

The same equipment may be used for semiautomatic GMAW, flux cored arc welding (FCAW), and submerged arc welding (SAW). Often, FCAW and SAW equipment have a higher amperage range. In addition, equipment for FCAW and SAW is more likely to be automated than that for GMAW. However, GMA welding equipment can easily be automated.

The basic GMAW installation consists of the following: welding gun, gun switch circuit, electrode conduit or liner, welding contractor control, electrode feed unit, electrode supply, power source, shielding gas supply, shielding gas flowmeter regulator, shielding gas hoses, and both power and work cables. Typical water-cooled and air-cooled guns are shown in Figure 2.1. The equipment setup in this chapter is similar to equipment built by other manufacturers, which means that any skills developed can be transferred easily to other equipment.

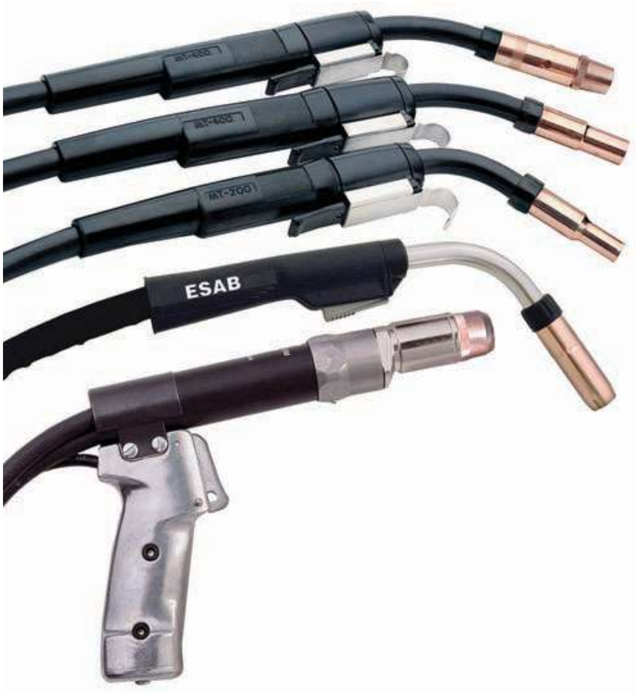

Figure 2.1 **GMA welding guns are available in a variety of sizes and shapes**
Source: Courtesy of ESAB Welding & Cutting Products

PRACTICE 2-1

GMAW Equipment Setup

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Short-circuit Transfer
Key Indicator 3
Spray Transfer
Key Indicator 8

For this practice, you will need a GMAW power source, a welding gun, an electrode feed unit, an electrode supply, a shielding gas supply, a shielding gas flowmeter regulator, electrode conduit, power and work leads, shielding gas hoses, assorted hand tools, spare parts, and any other required materials. In this practice, you will properly set up a GMA welding installation.

If the shielding gas supply is a cylinder, it must be chained securely in place before the valve protection cap is removed, Figure 2.2. Standing to one side of the cylinder and making sure no bystanders are in line with the valve, quickly crack the valve to blow out any dirt in the valve before the flowmeter regulator is attached, Figure 2.3. With the flowmeter regu-

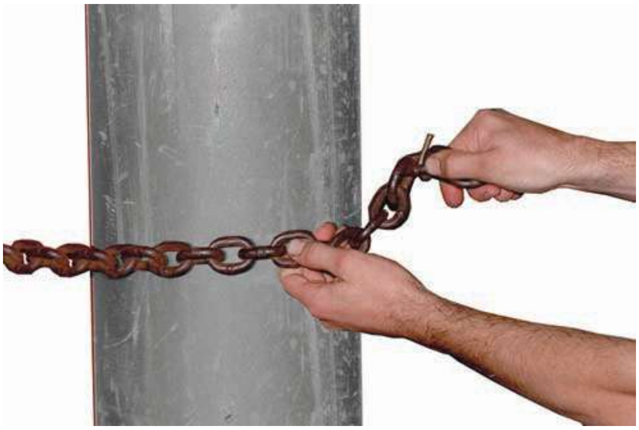

Figure 2.2 **Make sure the gas cylinder is chained securely in place before removing the safety cap**
Source: Courtesy of Larry Jeffus

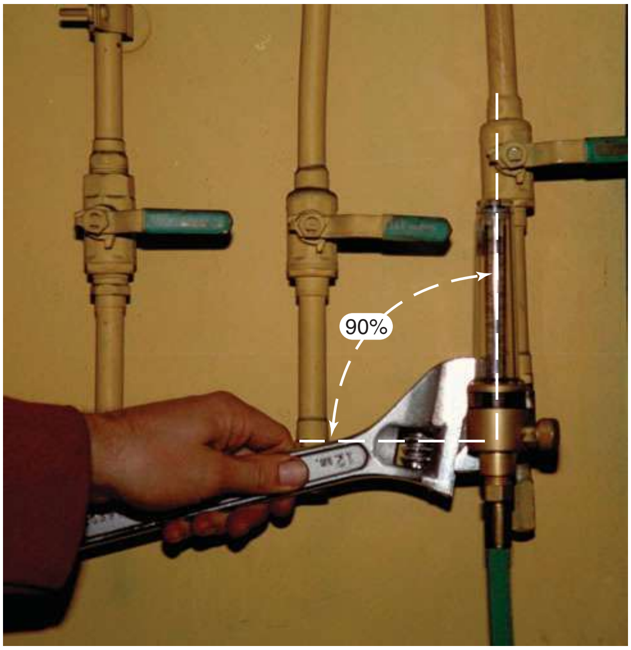

Figure 2.3 **Attaching the flowmeter regulator**
Be sure the tube is vertical.
Source: Courtesy of Larry Jeffus

lator attached securely to the cylinder valve, attach the correct hose from the flowmeter to the “gas-in” connection on the electrode feed unit or machine.

Install the reel of electrode (welding wire) on the holder and secure it, Figure 2.4. Check the feed roller size to ensure that it matches the wire size, Figure 2.5. The conduit liner size should be checked to be sure that it is compatible with the wire size. Connect the conduit to the feed unit. The conduit or an extension should be aligned with the groove in the roller and set as close to the roller as possible without touching, Figure 2.6. Misalignment at this point can contribute to a bird’s nest, Figure 2.7. **Bird-nesting** of the electrode wire results when the feed roller pushes the wire into a tangled ball, like a bird’s nest, because the wire would not go through the outfeed side conduit.

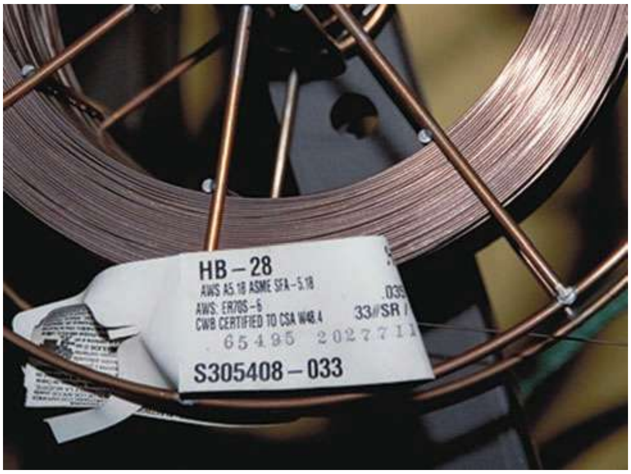
HB – 28
AWS A5.18 ASME SFA-5.18
AWS: ER70S-6
CWB CERTIFIED TO CSA W48.4
.035
33#SR
65495 2027711
S305408 – 033

Figure 2.4 Wire label
When installing the spool of wire, check the label to be sure that the wire is the correct type and size.
Source: Courtesy of Larry Jeffus

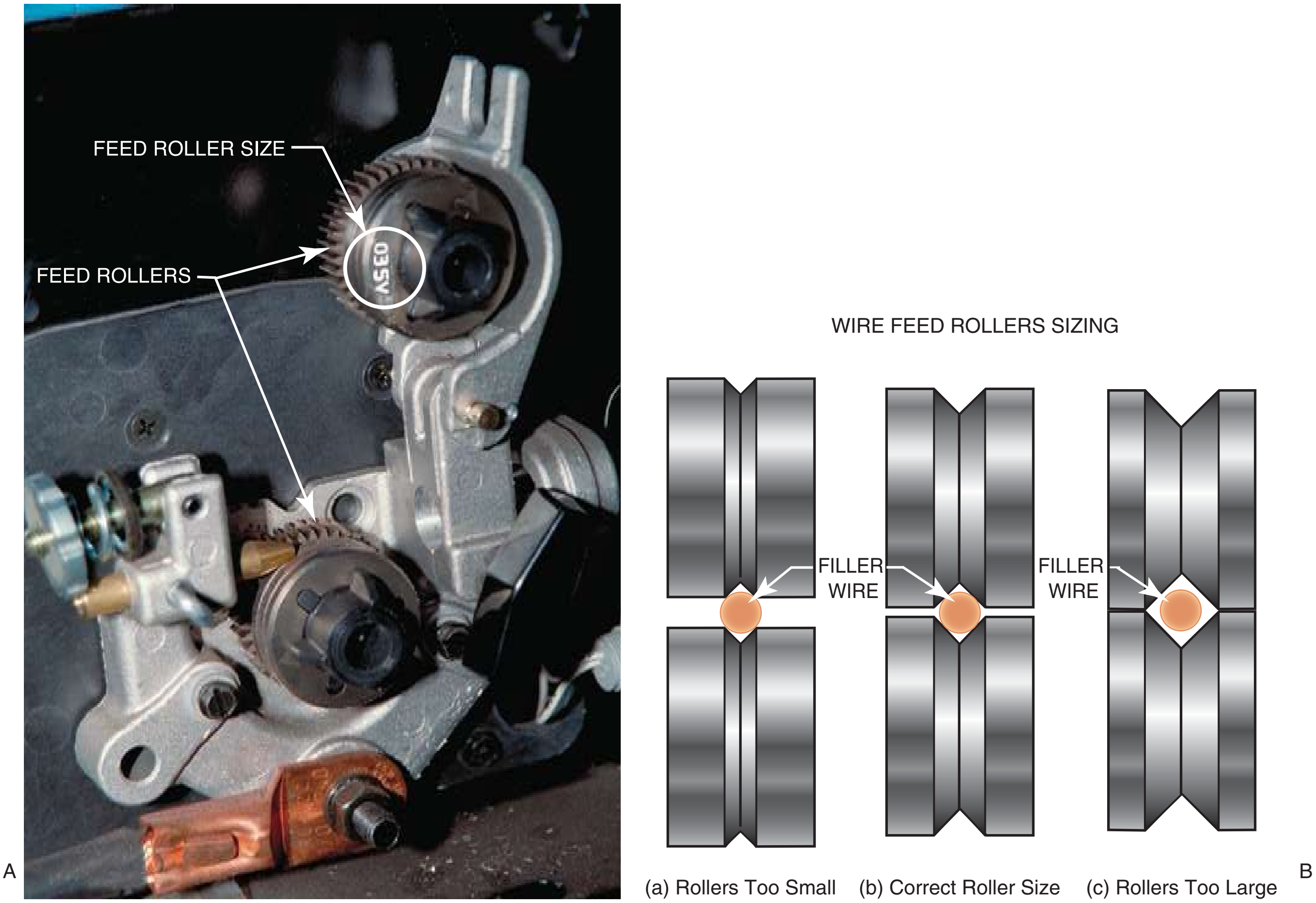
FEED ROLLER SIZE
FEED ROLLERS
035V
A
WIRE FEED ROLLERS SIZING
FILLER WIRE
FILLER WIRE
B

(a) Rollers Too Small (b) Correct Roller Size (c) Rollers Too Large

Figure 2.5 Checking feed roller size
(A) Check to be certain that the feed rollers are the correct size for the wire being used. (B) If the wirefeed rollers are too small, the welding wire could be damaged. If the wirefeed rollers are too large, the rollers will not grip the wire.
Source: (A) Courtesy of Larry Jeffus

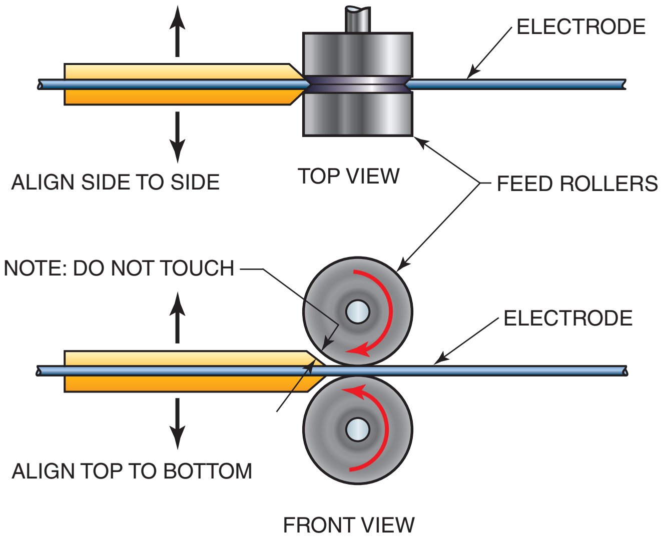
ELECTRODE
ALIGN SIDE TO SIDE
TOP VIEW
FEED ROLLERS
NOTE: DO NOT TOUCH
ELECTRODE
ALIGN TOP TO BOTTOM
FRONT VIEW

Figure 2.6 Feed

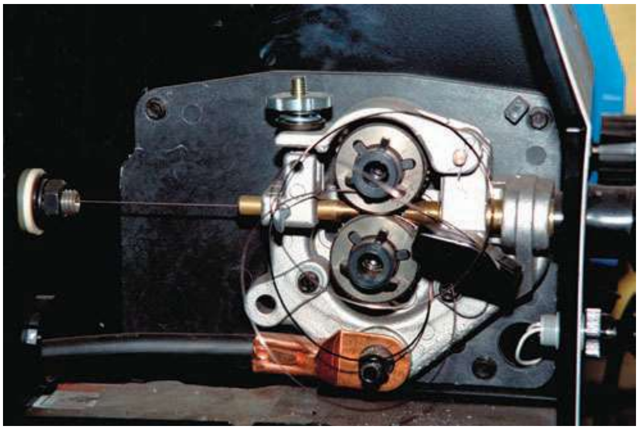

Figure 2.7 "Bird's nest" in the filler wire at the feed rollers
Source: Courtesy of Larry Jeffus

Be sure the power is off before attaching the welding cables. The electrode and work leads should be attached to the proper terminals. The electrode lead should be attached to the terminal marked electrode or positive (+). If necessary, it is also attached to the power cable part of the gun lead. The work lead should be attached to the terminal marked work or negative (–).

The shielding "gas-out" side of the solenoid is then also attached to the gun lead. If a separate splice is required from the gun switch circuit to the feed unit, it should be connected at this time, Figure 2.8. Check to see that the welding contractor circuit is connected from the feed unit to the power source.

The welding gun should be securely attached to the main lead cable and conduit, Figure 2.9. There should be a gas diffuser attached to the end of the conduit liner to ensure proper alignment. A **contact tube** (tip) of the correct size to match the electrode wire size being used should be installed, Figure 2.10. A shielding gas nozzle is attached to complete the assembly.

Recheck all fittings and connections for tightness. Loose fittings can leak; loose connections can cause added resistance, reducing the welding

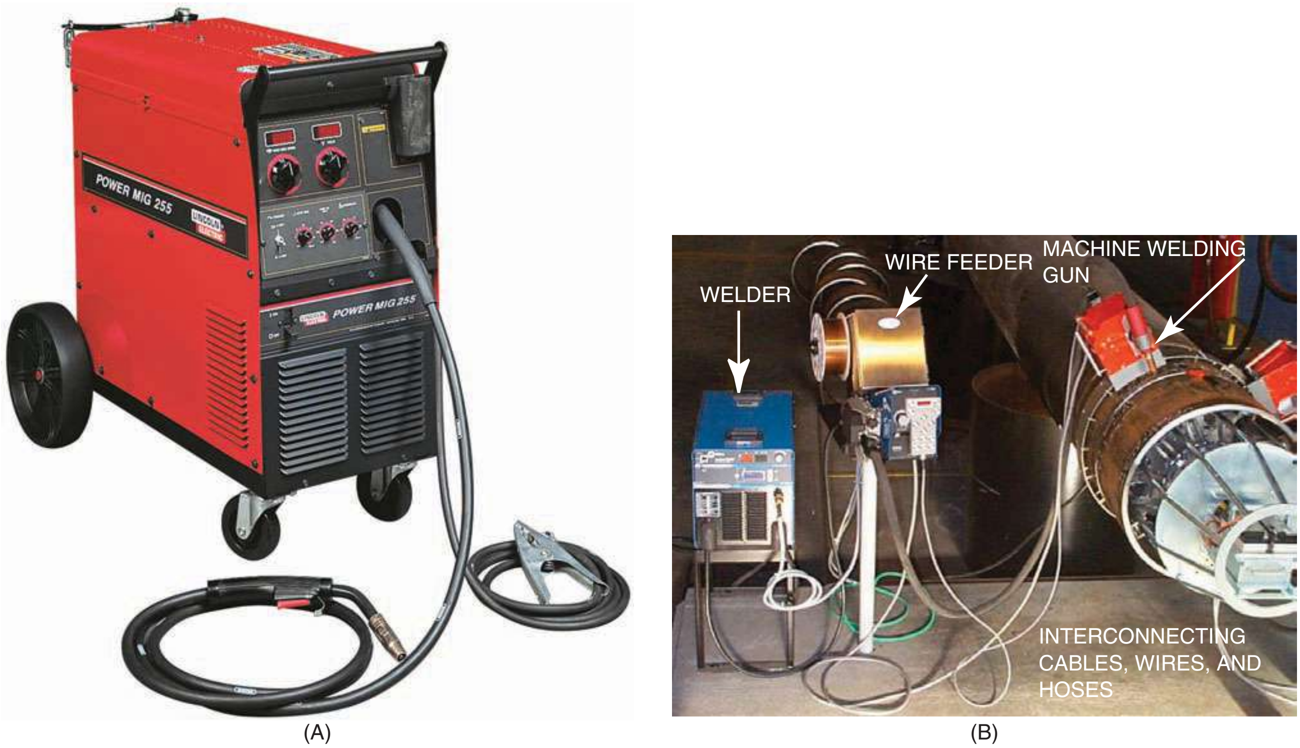
POWER MIG 255
WELDER
WIRE FEEDER
MACHINE WELDING GUN
INTERCONNECTING CABLES, WIRES, AND HOSES
(A)
(B)

Figure 2.8 **GMAW station setup**
(A) Typical GMA welding machine. (B) Typical interconnecting cables and wires for a semiautomatic GMA welding station.
Source: (A) Courtesy of Lincoln Electric Company (B) Courtesy of Dynatorque

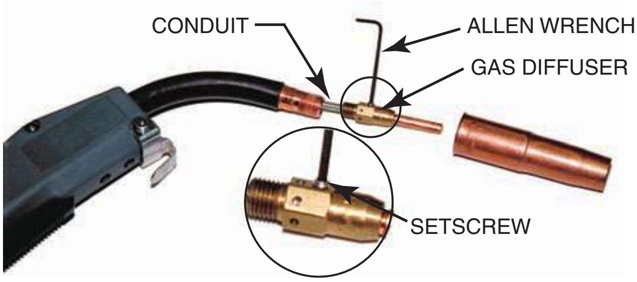
CONDUIT
ALLEN WRENCH
GAS DIFFUSER
SETSCREW

Figure 2.9 **GMA welding gun assembly**
Source: Courtesy of Larry Jeffus

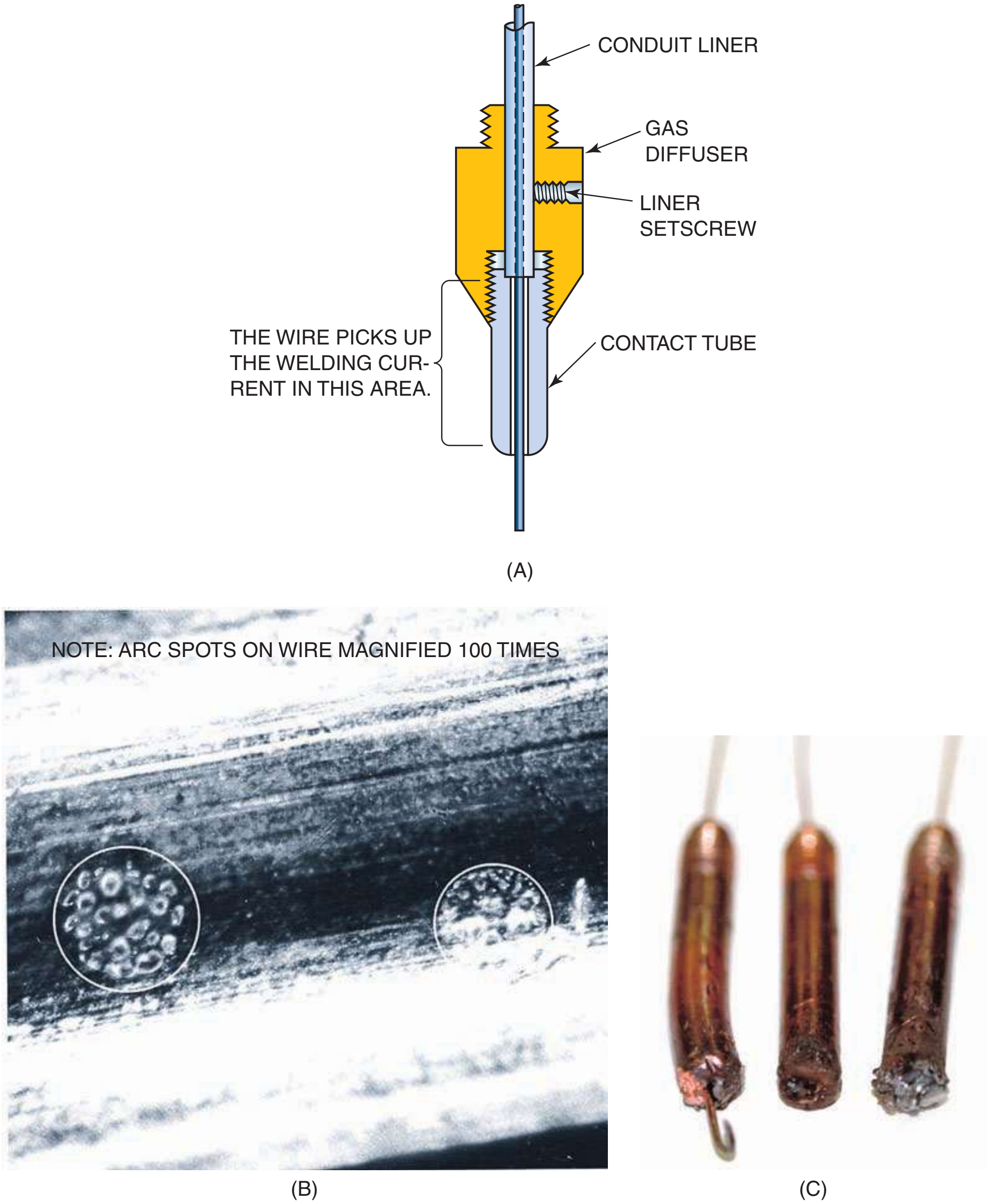
CONDUIT LINER
GAS DIFFUSER
LINER SETSCREW
THE WIRE PICKS UP THE WELDING CURRENT IN THIS AREA.
CONTACT TUBE
(A)
NOTE: ARC SPOTS ON WIRE MAGNIFIED 100 TIMES
(B)
(C)

Figure 2.10 The contact tube must be the correct size
(A) Too small a contact tube will cause the wire to stick. (B) Too large a contact tube can cause arcing to occur between the wire and tube. (C) Heat from the arcing can damage the tube.
Source: (B) Courtesy of Brett V. Hahn (C) Courtesy of Larry Jeffus

efficiency. Some manufacturers include detailed setup instructions with their equipment, Figure 2.11.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Short-circuit Transfer
Key Indicator 3
Spray Transfer
Key Indicator 8

PRACTICE 2-2

Threading GMAW Wire

Using the GMAW machine that was properly assembled in Practice 2-1, you will turn the machine on and thread the electrode wire through the system.

Open the side cover.

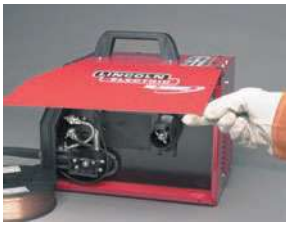

Remove the empty wire spool.

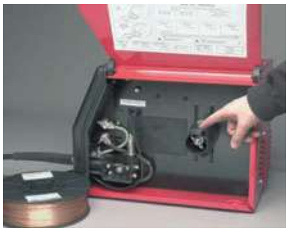

Release upper feed roller.

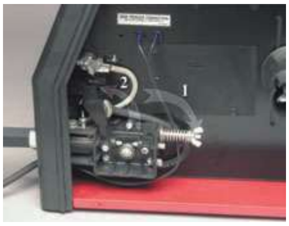

Reload the wire spool with the free end unreeling from the bottom.

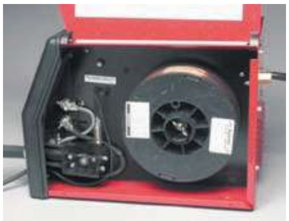

Thread wire through guide between rollers and into wire cable.

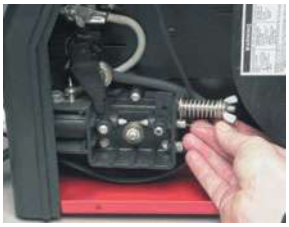

Set the polarity as DCEP from GMA welding.

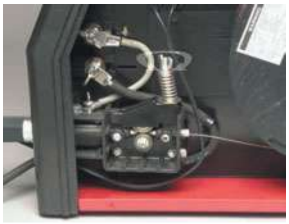

Turn the input switch on.

With the gun trigger pressed, adjust the feed roller tension.

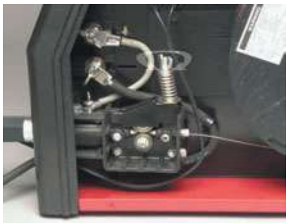

Check the setting guide inside the machine door.

Set the voltage and wire feed for the metal you are going to be welding.

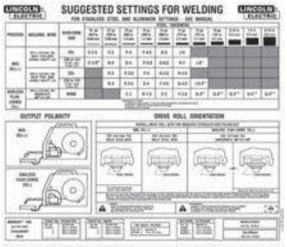

LINCOLN ELECTRIC SUGGESTED SETTINGS FOR WELDING LINCOLN ELECTRIC

OUTPUT POLARITY

DRIVE ROLL ORIENTATION

Attach work cable clamp to work to be welded.

Connect gas to coupling at rear of case and turn on shielding gas.

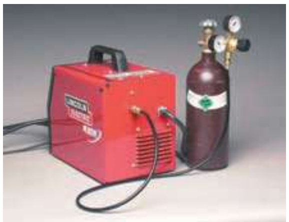

ALWAYS WEAR PROPER SAFETY EQUIPMENT. Pull trigger and weld.

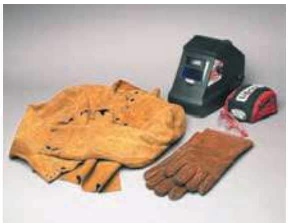

Figure 2.11 Example of manufacturer's setup instructions
Source: Courtesy of Lincoln Electric Company

Check to see that the unit is assembled correctly according to the manufacturer's specifications. Switch on the power and check the gun switch circuit by depressing the switch. The power source relays, feed relays, gas solenoid, and feed motor should all activate.

Cut the end of the electrode wire free. Hold it tightly so that it does not unwind. The wire has a natural curve that is known as its **cast**. The cast is measured by the diameter of the circle that the wire would make if it were loosely laid on a flat surface, Figure 2.12. The cast helps the wire make a good electrical contact as it passes through the contact tube. However, the cast can be a problem when threading the system. To make threading easier, straighten about 12 in. (305 mm) of the end of the wire and cut any kinks off.

Separate the wire **feed rollers** and push the wire first through the guides, Figure 2.13, then between the rollers, and finally into the **conduit liner**. Reset the rollers so there is a slight amount of compression on the wire, Figure 2.14. Set the **wire-feed speed** control to a slow speed. Hold the welding gun so that the electrode conduit and cable are as straight as possible.

Caution

If the wire stops feeding before it reaches the end of the contact tube, stop and check the system. If no obvious problem can be found, mark the wire with tape and remove it from the gun. It then can be held next to the system to determine the location of the problem.

With safety glasses on and the gun pointed away from the welder's face, press the gun switch, or the cold feed switch if your wire feeder is equipped with one. The cold feed switch feeds wire without delivering current to the gun. The wire should start feeding into the liner. Watch to make certain that the wire feeds smoothly and release the gun switch as soon as the end comes through the contact tube.

With the wire feed running, adjust the feed roller compression so that the wire reel can be stopped easily by a slight pressure. Too light a roller pressure will cause the wire to feed erratically. Too high a pressure can turn a minor problem into a major disaster. If the wire jams at a high roller pressure, the feed rollers keep feeding the wire, causing it to bird-nest and possibly short out. With a light pressure, the wire can stop, preventing bird-nesting. This is very important with soft wires. The other

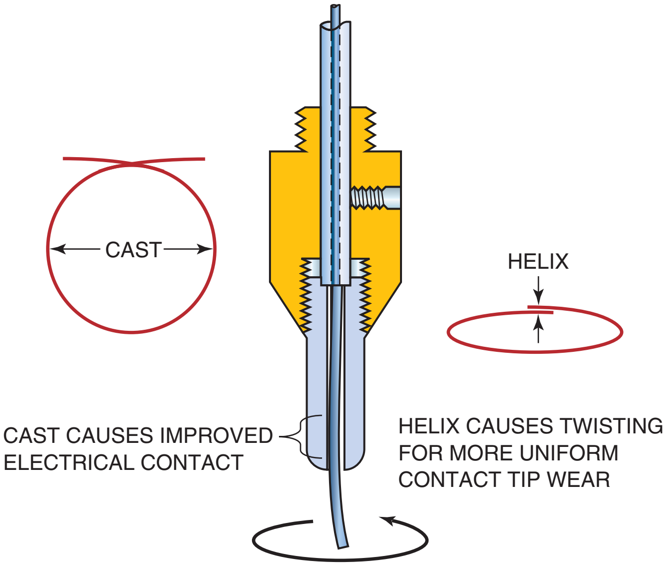
CAST
HELIX
CAST CAUSES IMPROVED ELECTRICAL CONTACT
HELIX CAUSES TWISTING FOR MORE UNIFORM CONTACT TIP WEAR

Figure 2.12 **Cast of a welding wire**
The cast of the welding wire causes it to rub firmly inside the contact tube for good electrical contact. The helix causes the electrode to twist inside the contact tube so that the tube is worn uniformly.

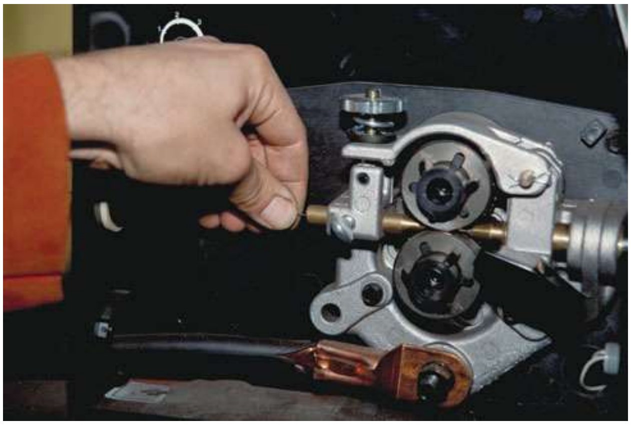

Figure 2.13 **Push the wire through the guides by hand**
Source: Courtesy of Larry Jeffus

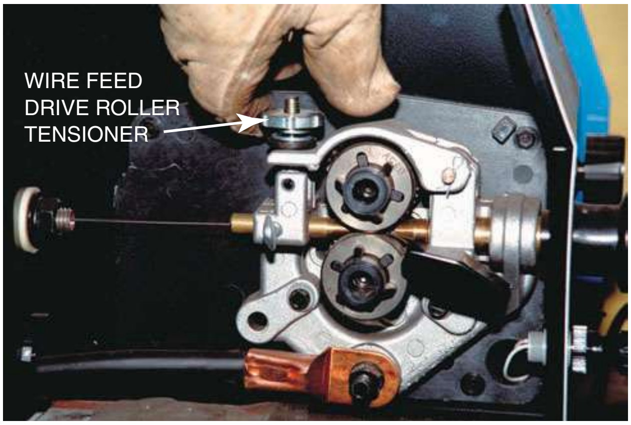

Figure 2.14 **Adjust the wire-feed tensioner**
Source: Courtesy of Larry Jeffus

advantage of a light pressure is that the feed will stop if something like clothing or a gas hose is caught in the reel.

With the feed running, adjust the **spool drag** so that the reel stops when the feed stops. The reel should not coast to a stop because the wire can be snagged easily. Also, when the feed restarts, a jolt occurs when the slack in the wire is taken up. This jolt can be enough to momentarily stop the wire, possibly causing a discontinuity in the weld.

When the test runs are completed, the wire can either be rewound or cut off. Some wire-feed units have a retract button. This allows the feed driver to reverse and retract the wire automatically. To rewind the wire on units without this retraction feature, release the rollers and turn them backward by hand. If the machine will not allow the feed rollers to be released without upsetting the tension, you must cut the wire.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

Caution

Do not discard pieces of wire on the floor. They present a hazard to safe movement around the machine. In addition, a small piece of wire can work its way into a filter screen on the welding power source. If the piece of wire shorts out inside the machine, it could become charged with high voltage, which could cause injury or death. Always wind the wire tightly into a ball or cut it into short lengths before discarding it in the proper waste container.

GAS DENSITY AND FLOW RATES

Density is the chief determinant of how effective a gas is for arc shielding. The lower the density of a gas, the higher will be the flow rate required for equal arc protection. Flow rates, however, are not in proportion to the densities. Helium, with about one-tenth the density of argon, requires about twice the flow for equal protection.

The correct flow rate can be set by checking welding guides that are available from the welding equipment and filler metal manufacturers. These welding guides list the gas flow required for various nozzle sizes and welding amperage settings. Some welders feel that a higher gas flow will provide better weld coverage, but that is not always the case. High gas flow rates waste shielding gases and may lead to contamination. The contamination comes from turbulence in the gas at high flow rates. Air is drawn into the gas envelope by the venturi effect around the edge of the nozzle. Also, the air can be drawn in under the nozzle if the torch is held at too sharp an angle to the metal, Figure 2.15.

NOTE

If you need more shielding gas coverage in a windy or drafty area, use both a larger-diameter gas nozzle and a higher gas flow rate. The larger the nozzle size, the higher the permissible flow rate without causing turbulence. Larger nozzle sizes may restrict your view of the weld. You might also consider setting up a wind barrier to protect your welding from the wind, Figure 2.16.

EXPERIMENT 2-1

Setting Gas Flow Rate

Module 1
Key Indicator 1, 2, 4

Module 5
Short-circuit Transfer
Key Indicator 3
Spray Transfer
Key Indicator 8

Using the equipment setup as described in Practice 2-1, and the threaded machine as described in Practice 2-2, you will set the shielding gas flow rate.

The exact **flow rate** required for a certain job will vary depending upon welding conditions. This experiment will help you determine how those conditions affect the flow rate. You will start by setting the shielding gas flow rate at 35 cubic feet per hour (cfh) (16 L/min).

Standing to one side, turn on the shielding gas supply valve. If the supply is a cylinder, the valve is opened all the way. With the machine power on and the welding gun switch depressed, you are ready to set

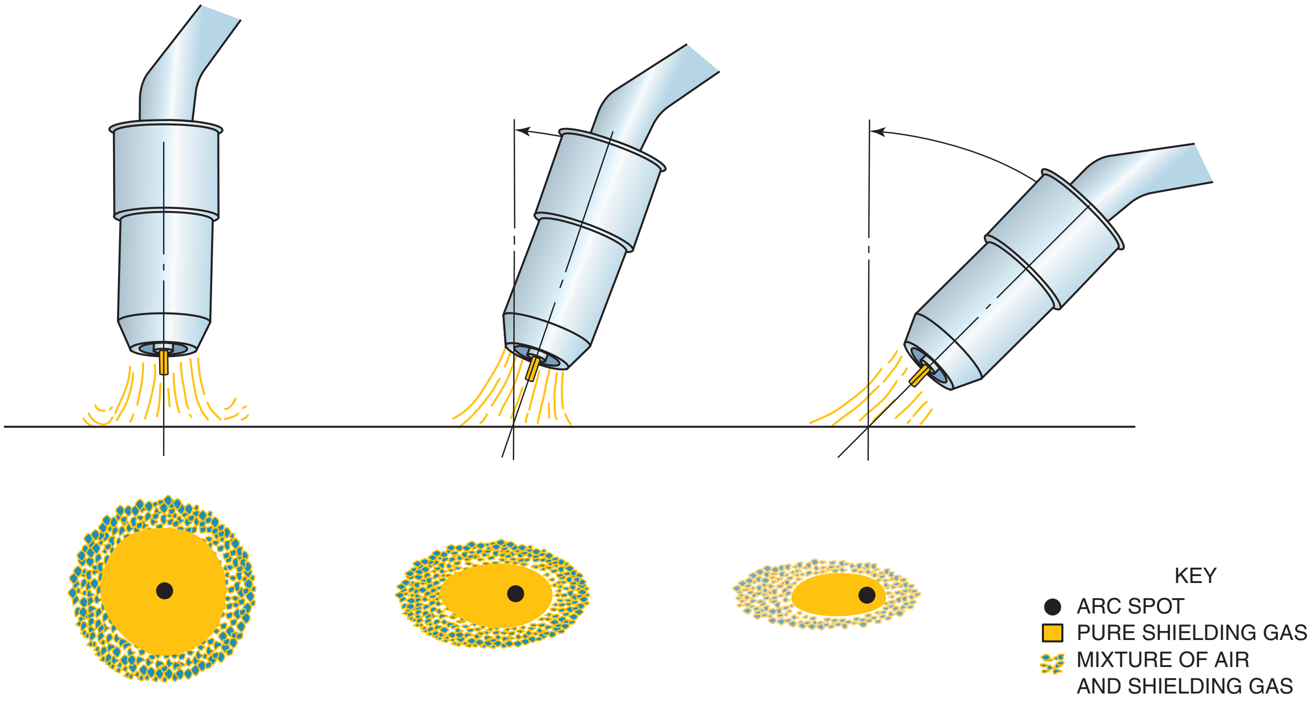

Figure 2.15 The welding gun angle affects the shielding gas coverage for the molten weld pool.

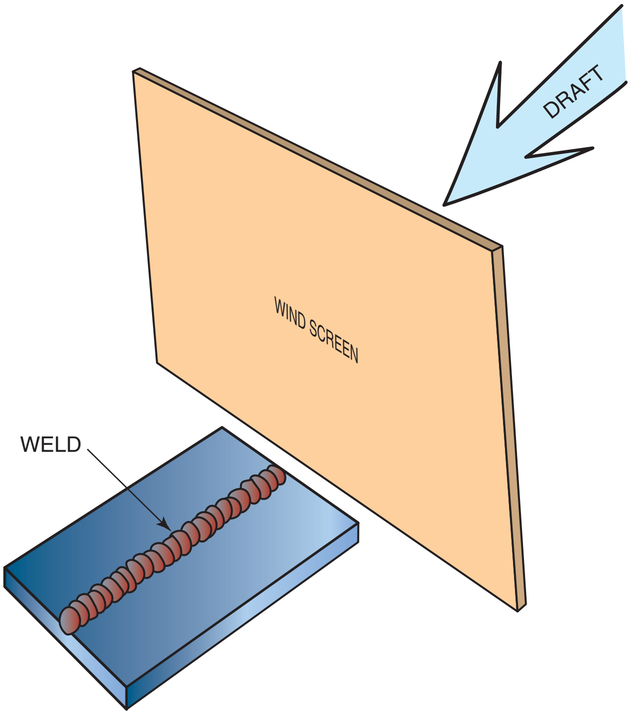

Figure 2.16 **A wind screen can help prevent the shielding gas from being blown away.**

the flow rate. Slowly turn in the adjusting screw and watch the float ball as it rises in a tube on a column of gas. The faster the gas flows, the higher the ball will float. A scale on the tube allows you to read the flow rate. Different scales are used with each type of gas being used. Since various gases have different densities (weights), the ball will float at varying levels even though the flow rates are the same, Figure 2.17. The line corresponding to the flow rate may be read as it compares to the top, center, or bottom of the ball, depending upon the manufacturer's instructions. There should be some marking or instruction on the tube or regulator to tell how it should be read, Figure 2.18.

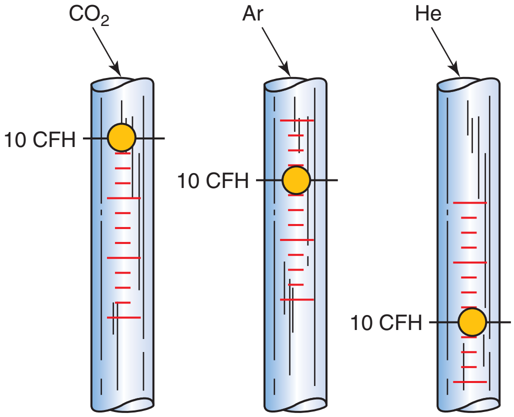

Figure 2.17 **Reading gas flow rates**
Each of these gases is flowing at the same cfh (L/min) rate. Because helium (He) is less dense, its indicator ball is the lowest. Be sure that you are reading the correct scale for the gas being used.

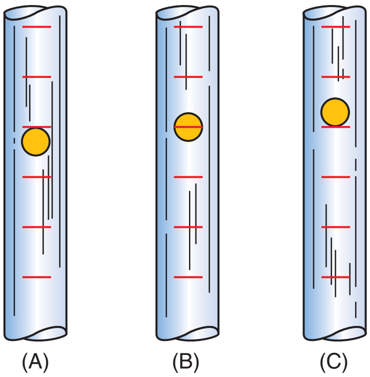

Figure 2.18 **Three methods of reading a flowmeter**
(A) Top of ball, (B) center of ball, and (C) bottom of ball.

Release the welding gun switch, and the gas flow should stop. Turn off the power and spray the hose fittings with a leak-detecting solution.

When you stop for more than a short period, close the shielding gas supply valve and release the hose pressure.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

ARC-VOLTAGE AND AMPERAGE CHARACTERISTICS

The arc-voltage and amperage characteristics of GMA welding are different from those for most other welding processes. The voltage is set on the welder, and the amperage is set by changing the wire-feed speed. At any one voltage setting the amperage required to melt the wire must change as it is fed into the weld. More amperage is required to melt the wire the faster it is fed, and less the slower it is fed.

Because changes in the wire-feed speed directly change the amperage, it is possible to set the amperage by using a chart and measuring the length of wire fed per minute, Table 2.1. The voltage and amperage required for a specific metal transfer method differ for various wire sizes, shielding gases, and metals.

The voltage and amperage setting will be specified for all welding done according to a welding procedure specification (WPS) or other codes and standards. However, most welding—like that done in small production shops, as maintenance welding, for repair work, in farm shops, and the like—is not done to a specific code or standard and therefore no specific setting exists. For that reason, it is important to learn to make the adjustments necessary to allow you to produce quality welds.

EXPERIMENT 2-2

Setting the Current

Module 1
Key Indicator 1, 2, 4

Module 5
Short-circuit Transfer
Key Indicator 4
Spray Transfer
Key Indicator 9

Using a properly assembled GMA welding machine, proper safety protection, and one piece of mild steel plate approximately 12 in. (305 mm) long × 1/4 in. (6 mm) thick, you will change the current settings and observe the effect on GMAW. On a scale of 0 to 10, set the wire-feed speed control dial at 5, or halfway between the low and high settings of

Table 2.1 Typical Amperages for Carbon Steel

	Wire Diameter			
Wire-feed Speed* (in./min)	.030 in. (0.8 mm)	.035 in. (0.9 mm)	.045 in. (1.2 mm)	.062 in. (1.6 mm)
100 (2.5)	40	65	120	190
200 (5.0)	80	120	200	330
300 (7.6)	130	170	260	425
400 (10.2)	160	210	320	490
500 (12.7)	180	245	365	–
600 (15.2)	200	265	400	–
700 (17.8)	215	280	430	–

*To check feed speed, run out wire for one minute and then measure its length.

the unit. The voltage is also set at a point halfway between the low and high settings. The shielding gas can be CO_2, argon, or a mixture. The gas flow should be adjusted to a rate of 35 cfh (16 L/min).

Hold the welding gun at a comfortable angle, lower your welding hood, and pull the trigger. As the wire feeds and contacts the plate, the weld will begin. Move the gun slowly along the plate. Note the following welding conditions as the weld progresses: voltage, amperage, weld direction, metal transfer, spatter, molten weld pool size, and penetration. Stop and record your observations in Table 2.2. Evaluate the quality of the weld as acceptable or unacceptable.

Reduce the voltage somewhat and make another weld, keeping all other weld variables (travel speed, stickout, direction, amperage) the same. Observe the weld and upon stopping record the results. Repeat this procedure until the voltage has been lowered to the minimum value indicated on the machine. Near the lower end the wire may stick, jump, or simply no longer weld.

Return the voltage indicator to the original starting position and make a short test weld. Stop and compare the results to those first observed. Then slightly increase the voltage setting and make another weld. Repeat the procedure of observing and recording the results as the voltage is increased in steps until the maximum machine capability is obtained. Near the maximum setting the spatter may become excessive if CO_2 shielding gas is used. Care must be taken to prevent the wire from fusing to the contact tube.

Return the voltage indicator to the original starting position and make a short test weld. Compare the results observed with those previously obtained.

Lower the wire-feed speed setting slightly and use the same procedure as before. First lower and then raise the voltage through a complete range and record your observations. After a complete set of test results are obtained from this amperage setting, again lower the wire-feed speed for a new series of tests. Repeat this procedure until the amperage is at the minimum setting shown on the machine. At low amperages and high voltage settings, the wire may tend to pop violently as a result of the uncontrolled arc.

Return the wire-feed speed and voltages to the original settings. Make a test weld and compare the results with the original tests. Slightly raise the wire speed and again run a set of tests as the voltage is changed in small steps. After each series, return the voltage setting to the starting point and increase the wire-feed speed. Make a new set of tests.

All of the test data can be gathered into an operational graph for the machine, wire type, size, and shielding gas. Set up a graph like that in Figure 2.19 to plot the data. The acceptable welds should be marked on

Table 2.2 **Setting the Current**

Weld Acceptability	Voltage	Amperage	Spatter	Molten Pool Size	Penetration
Good	20	75	Light	Small	Little

Electrode diameter	0.035 in. (0.9 mm)
Shielding gas CO_2	
Welding direction	Backhand

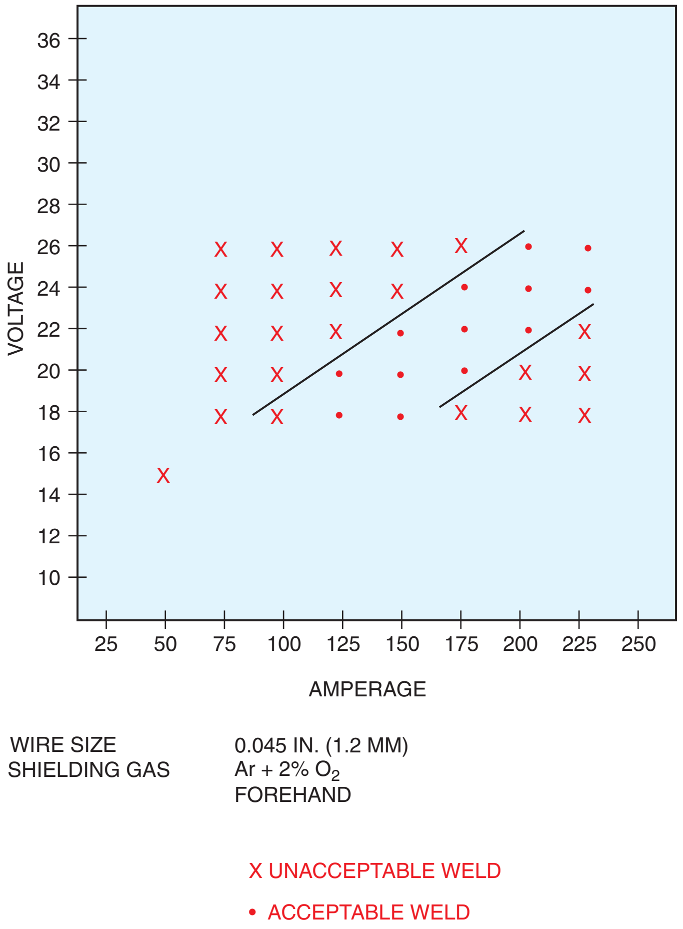

Figure 2.19 Graph for GMAW machine settings

the lines that extend from the appropriate voltages and amperages. Upon completion, the graph will give you the optimum settings for the operation of this particular GMAW setup. The optimum settings are along a line in the center of the acceptable welds.

Experienced welders will follow a much shorter version of this type of procedure anytime they are starting to work on a new machine or testing for a new job. This experiment can be repeated using different types of wire, wire sizes, shielding gases, and weld directions. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

ELECTRODE EXTENSION

Because of the constant-potential (CP) power supply, the welding current will change as the distance between the contact tube and the work changes. Although this change is slight, it is enough to affect the weld being produced. The longer the electrode extension, the greater the resistance will be to the welding current flowing through the small welding wire. This results in some of the welding current being changed to heat at the tip of the electrode, Figure 2.20. With a standard constant-current (CC) power supply for SMA welding, the heat buildup would also reduce the arc voltage, but with a CP power supply the voltage remains constant and the amperage increases. If the electrode extension is shortened, the welding current decreases.

The increase in current does not result in an increase in penetration, because the current is being used to heat the electrode tip and not being transferred to the weld metal. Penetration is reduced and buildup is increased as the electrode extension is lengthened. Penetration is increased and buildup decreased as the electrode extension is shortened. Controlling the weld penetration and buildup by changing the electrode will help maintain weld bead shape during welding. It will also help you better understand what may be happening if a weld starts out correctly but begins to change as it progresses along the joint. You may be changing the electrode extension without noticing the change. Short electrode stickout gives a hotter weld, and long stickout results in a cooler weld.

EXPERIMENT 2-3

Electrode Extension

Module 1
Key Indicator 1, 2, 4

Module 5
Short-circuit Transfer
Key Indicator 4
Spray Transfer
Key Indicator 9

Using a properly assembled GMA welding machine, proper safety protection, and a few pieces of mild steel, each about 12 in. (305 mm) long and ranging in thickness from 16 gauge to 1/2 in. (13 mm), you will observe the effect of changing electrode extension on the weld.

Start at a low current setting. Using the graph developed in Experiment 2.2, set both the voltage and amperage. The settings should be equal to those on the optimum line established for the wire type and size being used with the same shielding gas.

Holding the welding gun at a comfortable angle and height, lower your helmet and start to weld. Make a weld approximately 2 in. (51 mm) long. Then reduce the distance from the gun to the work while continuing to weld. After a few inches, again shorten the electrode extension even

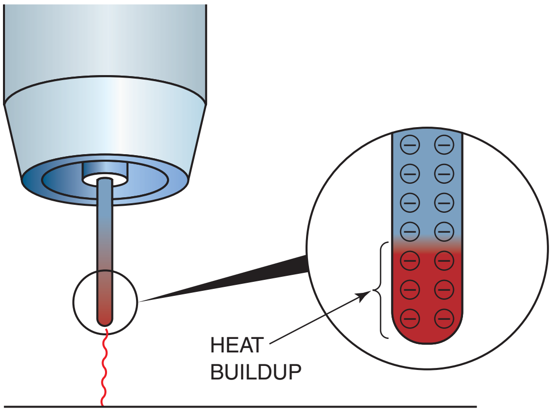

Figure 2.20 **Heat buildup due to the extremely high current for the small conductor (electrode)**

more. Keep doing this in steps until the nozzle is as close as possible to the work. Stop and return the gun to the original starting distance.

Repeat the process just described but now increase the electrode extension and make welds of a few inches at each extension. Keep increasing the electrode extension until the weld will no longer fuse or the wire becomes impossible to control.

Change the plate thickness and repeat the procedure. When the series has been completed with each plate thickness, raise the voltage and amperage to a medium setting and repeat the process. Upon completing this series of tests, adjust the voltage and amperage upward to a high setting. Make a full series of tests using the same procedures as before.

Record the results in Table 2.3 after each series of tests. The final results can be plotted on a graph, as was done in Figure 2.21, to establish the optimum electrode extension for each thickness, voltage, and amperage. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Table 2.3 **Electrode Extension**

Weld Acceptability	Voltage	Amperage	Electrode Extension	Contact Tube to Work Distance	Bead Shape
Poor	20	100	1 in. (25 mm)	1-1/4 in. (31 mm)	Narrow, high, with little penetration

Electrode diameter: 0.035 in. (0.9 mm)
Shielding gas: CO_2
Welding direction: Forehand

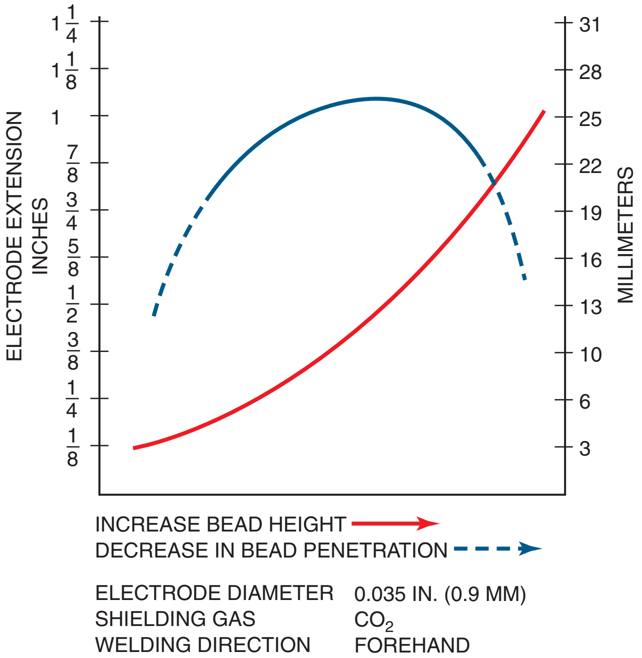

Figure 2.21 **Plot of Experiment 2-3 results**

WELDING GUN ANGLE

The term *welding gun angle* refers to the angle between the GMA welding gun and the work as it relates to the direction of travel. Backhand welding, or dragging angle, Figure 2.22, produces a weld with deep penetration and higher buildup. Forehand welding, or pushing angle, Figure 2.23, produces a weld with shallow penetration and little buildup.

Slight changes in the welding gun angle can be used to control the weld as the groove spacing changes. A narrow gap may require more penetration, but as the gap spacing increases a weld with less penetration may be required. Changing the electrode extension and welding gun angle at the same time can result in a quality weld being made under less than ideal conditions.

EXPERIMENT 2-4

Welding Gun Angle

Module 1
Key Indicator 1, 2, 4

Module 5
Short-circuit Transfer
Key Indicator 4
Spray Transfer
Key Indicator 9

Using a properly assembled GMA welding machine, proper safety protection, and some pieces of mild steel, each approximately 12 in. (305 mm) long and ranging in thickness from 16 gauge to 1/2 in. (13 mm), you will observe the effect of changing the welding gun angle on the weld bead.

Starting with a medium current setting and a plate that is 1/4 in. (6 mm) thick, hold the welding gun at a 30° angle to the plate in the direction of the weld, Figure 2.24. Lower your welding hood and depress the trigger. When the weld starts, move in a straight line and slowly pivot the gun angle as the weld progresses. Keep the travel speed, electrode

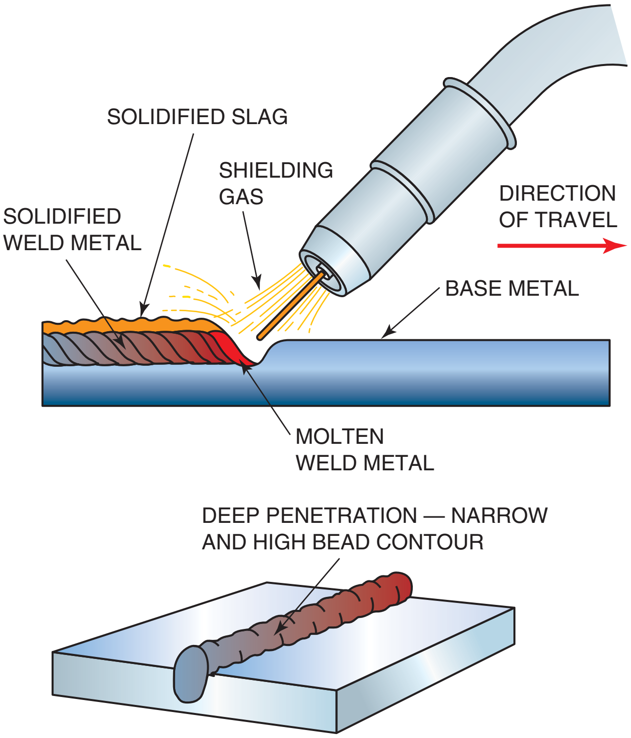
SOLIDIFIED SLAG
SHIELDING GAS
SOLIDIFIED WELD METAL
DIRECTION OF TRAVEL
BASE METAL
MOLTEN WELD METAL
DEEP PENETRATION — NARROW AND HIGH BEAD CONTOUR

Figure 2.22 Backhand welding, or dragging angle

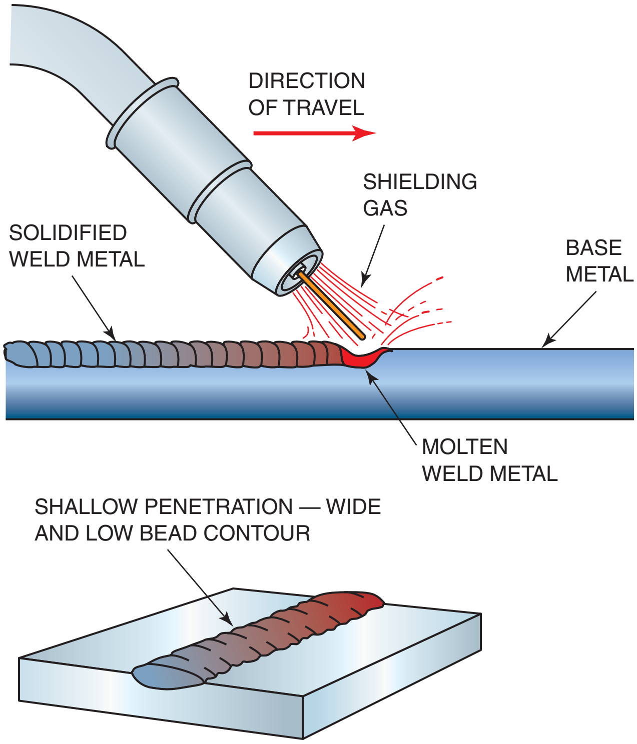
DIRECTION OF TRAVEL
SHIELDING GAS
SOLIDIFIED WELD METAL
BASE METAL
MOLTEN WELD METAL
SHALLOW PENETRATION — WIDE AND LOW BEAD CONTOUR

Figure 2.23 Forehand welding, or pushing angle

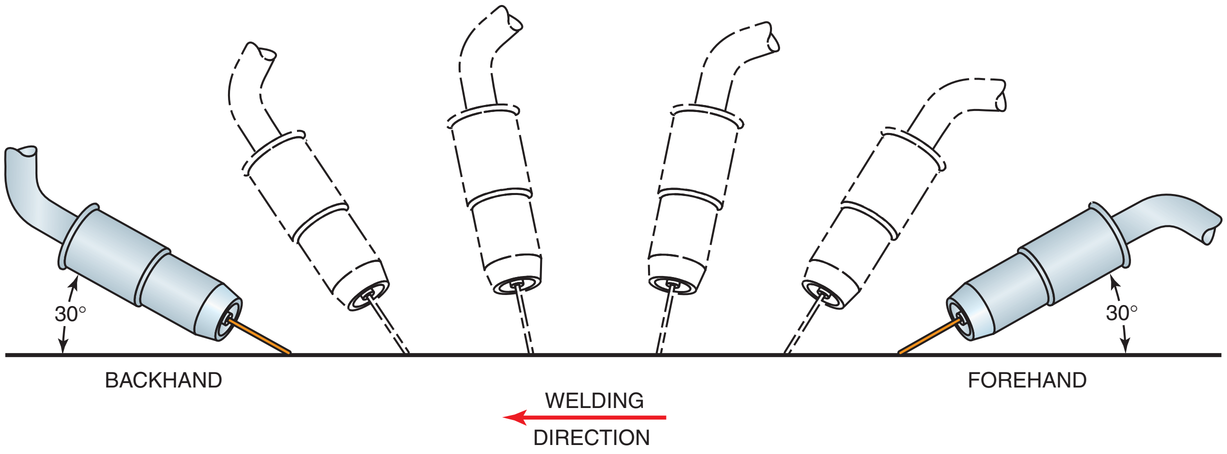

Figure 2.24 **Welding gun angle**

extension, and weave pattern (the side-to-side motion, if used) constant so that any change in the weld bead is caused by the angle change.

The pivot should be completed in the 12 in. (305 mm) of the weld. You will proceed from a 30° pushing angle to a 30° dragging angle. Repeat this procedure using different welding currents and plate thicknesses.

After the welds are complete, note the differences in width and reinforcement along the welds. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

EFFECT OF SHIELDING GAS ON WELDING

Shielding gases in the gas metal arc process are used primarily to protect the molten metal from oxidation and contamination. Other factors must be considered, however, in selecting the right gas for a particular application. Shielding gas can influence arc and metal transfer characteristics, weld penetration, width of fusion zone, surface shape patterns, welding speed, and undercut tendency. Inert gases such as argon and helium provide the necessary shielding because they do not form compounds with any other substance and are insoluble in molten metal. When used as pure gases for welding ferrous metals, argon and helium may produce an erratic arc action, promote undercutting, and result in other flaws.

It is therefore usually necessary to add controlled quantities of reactive gases to achieve good arc action and metal transfer with these materials. Adding oxygen or carbon dioxide to the inert gas tends to stabilize the arc, promote favorable metal transfer, and minimize spatter. As a result, the penetration pattern is improved and undercutting is reduced or eliminated.

Oxygen or carbon dioxide is often added to argon. The amount of reactive gas required to produce the desired effects is quite small. As little as 0.5% of oxygen will produce a noticeable change; 1% to 5% of oxygen is more common. Carbon dioxide may be added to argon in the 10% to 30% range. Mixtures of argon with less than 10% carbon dioxide may not have enough arc voltage to give the desired results.

Adding oxygen or carbon dioxide to an inert gas causes the shielding gas to become oxidizing. This in turn may cause porosity in some ferrous metals. In this case, a filler wire containing suitable deoxidizers should be used. The presence of oxygen in the shielding gas can also cause some loss of certain alloying elements, such as chromium, vanadium, aluminum, titanium, manganese, and silicon. Again, the addition of a deoxidizer to the filler wire is necessary.

Pure carbon dioxide has become widely used as a shielding gas for GMA welding of steels. It allows higher welding speed, better penetration in the short-circuiting transfer mode, and good mechanical properties, and it costs less than the inert gases. The chief drawback in the use of carbon dioxide is the less-steady-arc characteristics and considerable weld-metal-spatter losses. The spatter can be kept to a minimum by maintaining a very short, uniform arc length. Consistently sound welds can be produced using carbon dioxide shielding, provided that a filler wire having the proper deoxidizing additives is used.

EXPERIMENT 2-5

Effect of Shielding Gas Changes

Module 1
Key Indicator 1, 2, 4

Module 5
Short-circuit Transfer
Key Indicator 4
Spray Transfer
Key Indicator 9

Using a properly assembled GMA welding machine; proper safety protection; a source of CO_2, argon, and oxygen gases or a variety of premixed shielding gases; two flowmeters (or one two-gas mixing regulator); and some pieces of mild steel plate, each about 12 in. (305 mm) long and ranging in thickness from 16 gauge to 1/2 in. (13 mm), you will observe the effect of various shielding gas mixtures on the weld.

Using a mixing flowmeter regulator will allow the gases to be mixed in any desired mixture. A mixing ratio chart to be used to arrive at the approximate gas percentages appears in Figure 2.25. The exact ratios are not so important to you, as a student, as they are on code work.

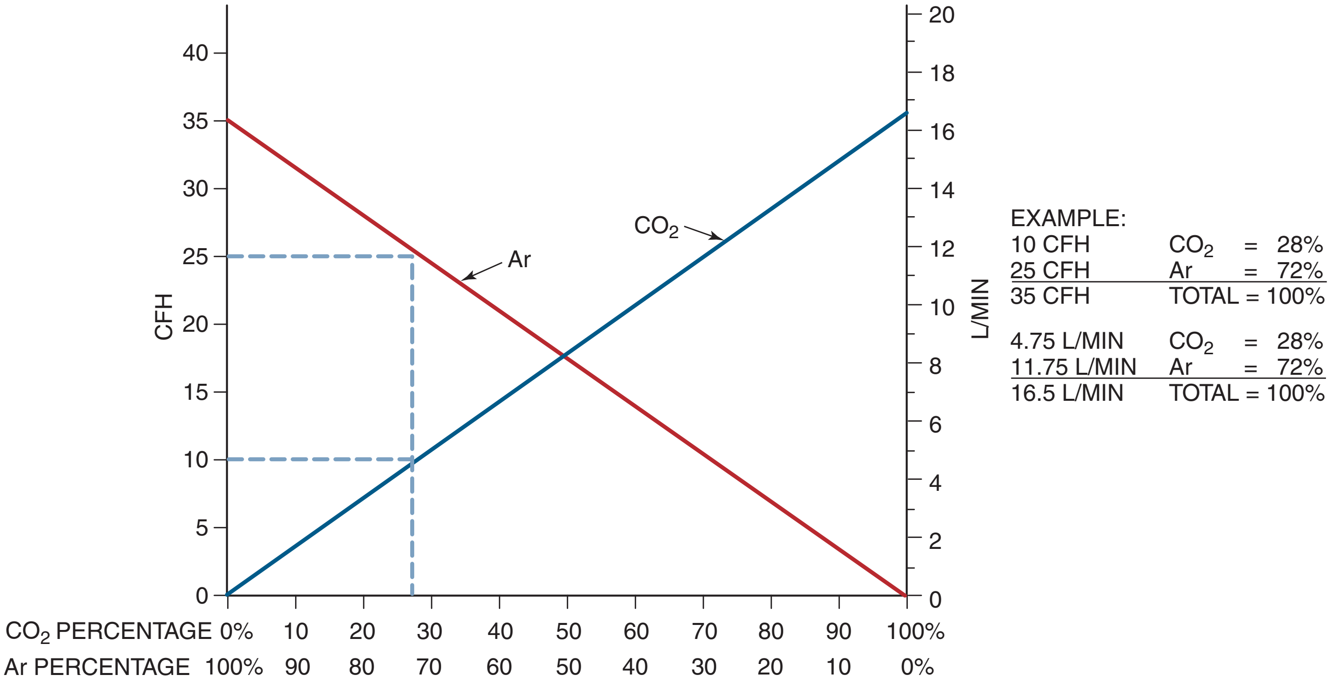

Figure 2.25 Gas mixing percentages

With a medium voltage and amperage setting and using a 100% carbon dioxide (CO_2) shielding gas, start making a weld. Either change the mixture after each weld or have another person change the shielding gas during the weld. Keep the total flow rate the same by adding argon (Ar) while reducing the CO_2 to preserve the same flow rate. During the experiment, change over the shielding gas to 100% argon (Ar).

After the weld is complete, evaluate it for spatter, penetration, undercut, buildup, width, or other noticeable changes along its length. Using Table 2.4, record the results of your evaluation.

Repeat the procedure just explained two times more with both low and high power settings. Again, record your observations.

Starting with 100% argon (Ar), add oxygen (O_2) to the shielding gas. The oxygen percentage will range from 0% to 10%, Figure 2.26. Very slight changes in the percentage will have dramatic effects on the weld. You will make three welds using low, medium, and high power settings. For each weld, you will record your observations.

Table 2.4 Shielding Gas Mixtures

Weld Acceptability	Voltage	Spatter	Penetration	Puddle Size	Bead Appearance
Good	75 Ar 25 CO_2	Very little	Deep	Large	Wide with little buildup

Electrode diameter: 0.035 in. (0.9 mm)
Welding direction: Backhand
Voltage: 25
Amperage: 150

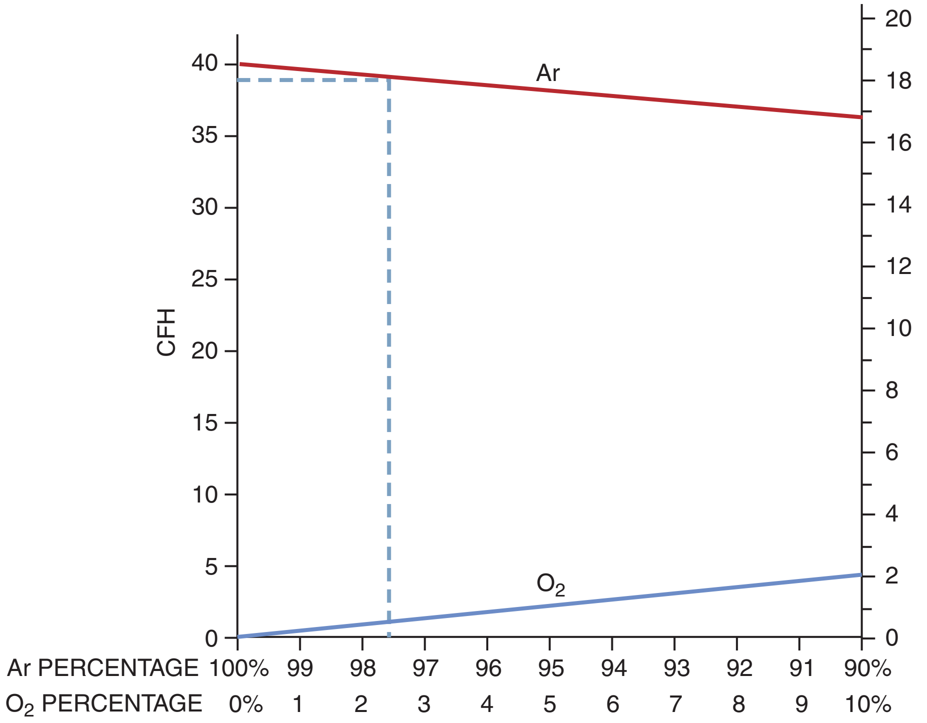

EXAMPLE:

39 CFH	Ar	= 97.5%
1 CFH	O_2	= 2.5%
40 CFH	TOTAL	= 100 %
18.4 L/MIN	Ar	= 97.5%
0.5 L/MIN	O_2	= 2.5%
18.9 L/MIN	TOTAL	= 100 %

Figure 2.26 Ar and O_2 Mixture Percentages

During some of the welding tests, you will notice a change in the method of metal transfer, weld heat, and general weld performance without a change in the current settings. The shielding gas mixture can have major effects on the rate of metal transfer and the welding speed, as well as other welding variables. Higher speeds and greater production can be obtained by using some gas mixtures. However, the savings can be completely offset by the higher gas cost. Before making a final decision about the gas to be used, all the variables must be compared. Table 2.5 lists premixed shielding gases and their uses.

Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICES

The practices in this chapter are grouped according to those requiring similar techniques and setups. To make acceptable GMA welds consistently, the major skill required is the ability to set up the equipment and weldment. Changes such as variations in material thickness, position, and type of joint require changes in both technique and setup. A correctly set up GMA welding station can, in many cases, be operated with minimum skill. Often the only difference between a welder earning a minimum wage and one earning the maximum wage is the ability to make correct machine setups.

Ideally, only a few tests would be needed for the welder to make the necessary adjustments in setup and manipulation techniques to achieve

Table 2.5 **Shielding Gases and Gas Mixtures Used for Gas Metal Arc Welding**

Shielding Gas	Chemical Behavior	Uses and Usage Notes
1. Argon	Inert	Welding virtually all metals except steel
2. Helium	Inert	Al and Cu alloys for greater heat and to minimize porosity
3. Ar and He (20% to 80% to 50% to 50%)	Inert	Al and Cu alloys for greater heat and to minimize porosity but with quieter, more readily controlled arc action
4. N_2	Reducing	On Cu, very powerful arc
5. Ar + 25% to 30% N_2	Reducing	On Cu, powerful but smoother operating, more readily controlled arc than with N_2
6. Ar + 1% to 2% O_2	Oxidizing	Stainless and alloy steels, also for some deoxidized copper alloys
7. Ar + 3% to 5% O_2	Oxidizing	Plain carbon, alloy, and stainless steels (generally requires highly deoxidized wire)
8. Ar + 3% to 5% O_2	Oxidizing	Various steels using deoxidized wire
9. Ar + 20% to 30% O_2	Oxidizing	Various steels, chiefly with short-circuiting arc
10. Ar + 5% O_2 + 15% CO_2	Oxidizing	Various steels using deoxidized wire
11. CO_2	Oxidizing	Plain-carbon and low-alloy steels, deoxidized wire essential
12. CO_2 + 3% to 10% O_2	Oxidizing	Various steels using deoxidized wire
13. CO_2 + 20% O_2	Oxidizing	Steels

a good weld. The previous welding experiments should have given the welder a graphic set of comparisons to help that welder make the correct changes. In addition to keeping the test data, you may want to keep the test plates for a more accurate comparison.

The grouping of practices in this chapter will keep the number of variables in the setup to a minimum. Often, the only change required before going on to the next weld is to adjust the power settings.

Figures that are given in some of the practices will give the welder general operating conditions, such as voltage, amperage, and shielding gas or gas mixture. These are general values, so the welder will have to make some fine adjustments. Differences in the type of machine being used and the material surface condition will affect the settings. For this reason, it is preferable to use the settings developed during the experiments.

METAL PREPARATION

All hot-rolled steel has an oxide layer, which is formed during the rolling process, called mill scale. *Mill scale* is a thin layer of dark gray or black iron oxide. Some hot-rolled steels that have had this layer removed either mechanically or chemically can be purchased. However, almost all of the hot-rolled steel used today still has this layer because it offers some protection from rusting.

Mill scale is not removed for noncode welding, because it does not prevent most welds from being suitable for service. For practice welds that will be visually inspected, mill scale can usually be left on the plate. Filler metals and fluxes usually have deoxidizers added to them so that the adverse effects of the mill scale are reduced or eliminated, Table 2.6. But with GMA welding wire it is difficult to add enough deoxidizers to remove all effects of mill scale. The porosity that mill scale causes is most often confined to the interior of the weld and is not visible on the surface, Figure 2.27. Because it is not visible on the surface, it usually goes unnoticed and the weld passes visual inspection.

If the practice results are going to be destructively tested or if the work is of a critical nature, then all welding surfaces within the weld groove and the surrounding surfaces within 1 in. (25 mm) must be cleaned to bright metal, Figure 2.28. Cleaning may be either grinding, filing, sanding, or blasting.

Table 2.6 Deoxidizing Elements in Filler Wire

Deoxidizing Element	Strength
Aluminum (Al)	Very strong
Manganese (Mn)	Weak
Silicon (Si)	Weak
Titanium (Ti)	Very strong
Zirconium (Zr)	Very strong

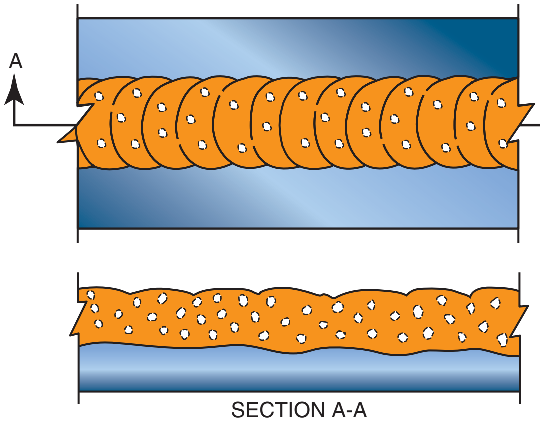

Figure 2.27 Uniformly scattered porosities

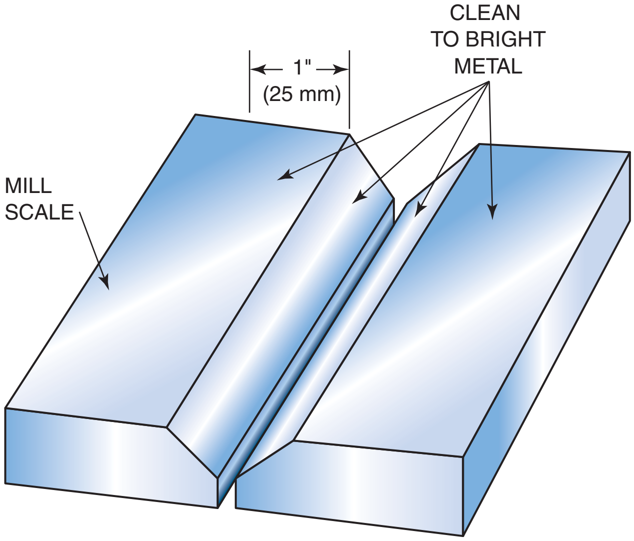

Figure 2.28 Clean all surfaces to bright metal before welding

FLAT POSITION, 1G AND 1F POSITIONS

PRACTICE 2-3

Stringer Beads Using the Short-circuiting Metal Transfer Method in the Flat Position

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Short-circuit Transfer
Key Indicator 4

Using a properly set up and adjusted GMA welding machine, Table 2.7, proper safety protection, 0.035-in. and/or 0.045-in. (0.9-mm and/or 1.2-mm) diameter wire, and two or more pieces of mild steel sheet 12 in. (305 mm) long and 16 gauge and 1/8 in. (3 mm) thick, you will make a stringer bead weld in the flat position, Figure 2.29.

Starting at one end of the plate and using either a pushing or dragging technique, make a weld bead along the entire 12-in. (305-mm) length of the metal. After the weld is complete, check its appearance. Make any needed changes in voltage, wire feed speed, or electrode extension to correct the weld (refer to Figure 2.19 and Figure 2.21). Repeat the weld and make additional adjustments. After the machine is set, start to work on improving the straightness and uniformity of the weld.

Keeping the bead straight and uniform can be hard because of the limited visibility due to the small amount of light and the size of the molten weld pool. The welder's view is further restricted by the shielding gas nozzle, Figure 2.30. Even with limited visibility, it is possible to make a

Table 2.7 Typical Welding Current Settings for Short-circuiting Metal Transfer for Mild Steel

Process	Wire Diameter	Amperage Range (Optimum)	Voltage Range (Optimum)	Shielding Gas
Short-circuiting	0.030	60 (100) 140	14 (15) 16	100% CO_2
	0.035	90 (130) 150	16 (17) 20	75% Ar + 25% CO_2 98% Ar + 2% O

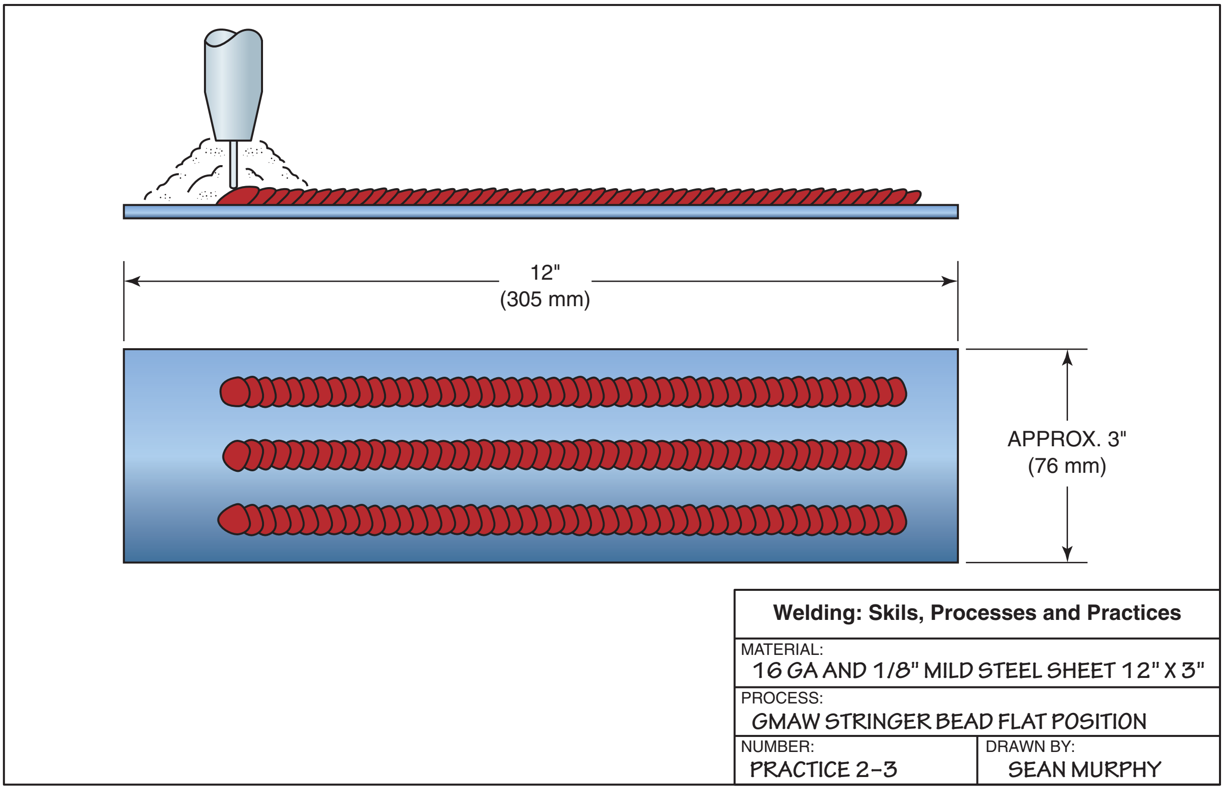
12"
(305 mm)
APPROX. 3"
(76 mm)
Welding: Skils, Processes and Practices
MATERIAL:
16 GA AND 1/8" MILD STEEL SHEET 12" X 3"
PROCESS:
GMAW STRINGER BEAD FLAT POSITION
NUMBER:
PRACTICE 2-3
DRAWN BY:
SEAN MURPHY

Figure 2.29 **Stringer beads in the flat position**

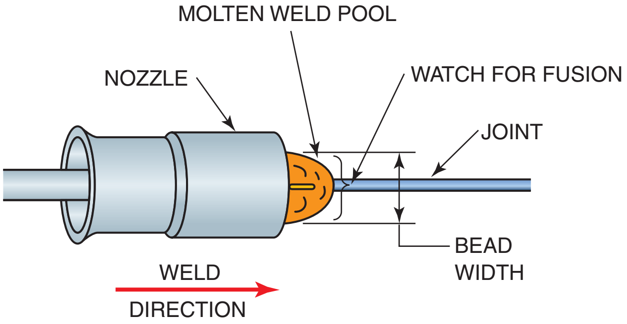
MOLTEN WELD POOL
NOZZLE
WATCH FOR FUSION
JOINT
BEAD
WIDTH
WELD
DIRECTION

Figure 2.30 **Watching the weld in forehand welding**
The shielding gas nozzle restricts the welder's view of the weld bead when pushing.

satisfactory weld by watching the edge of the molten weld pool, the sparks, and the weld bead produced. Watching the leading edge of the molten weld pool (forehand welding, pushing technique) will show you the molten weld pool fusion and width. Watching the trailing edge of the molten weld pool (backhand welding, dragging technique) will show you the amount of buildup and the relative heat input, Figure 2.31. The quantity and size of sparks produced can indicate the relative location of the filler wire in the molten weld pool. The number of sparks will increase as the wire strikes the solid metal ahead of the molten weld pool. The gun itself will begin to vibrate or bump as the wire momentarily pushes against the cooler, unmelted base metal before it melts. Changes in weld width, buildup, and proper joint tracking can be seen by watching the bead as it appears from behind the shielding gas nozzle.

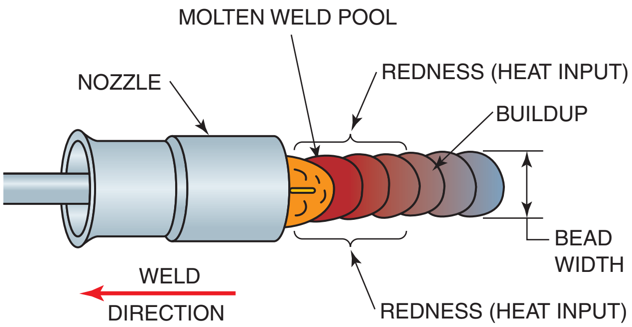
MOLTEN WELD POOL
NOZZLE
REDNESS (HEAT INPUT)
BUILDUP
BEAD WIDTH
WELD DIRECTION
REDNESS (HEAT INPUT)

Figure 2.31 **Watching the weld in backhand welding**
Watch the trailing edge of the molten weld pool.

Repeat each type of bead as needed until consistently good beads are obtained. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-4

Flat Position Butt Joint, Lap Joint, and Tee Joint

Using the same equipment, materials, and procedures listed in Practice 2-3, make welded butt joints, lap joints, and tee joints in the flat position, Figure 2.32A, 2.32B and 2.32C.

- Tack weld the sheets together and place them flat on the welding table, Figure 2.33.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Flat" position portion of the all-position requirement.

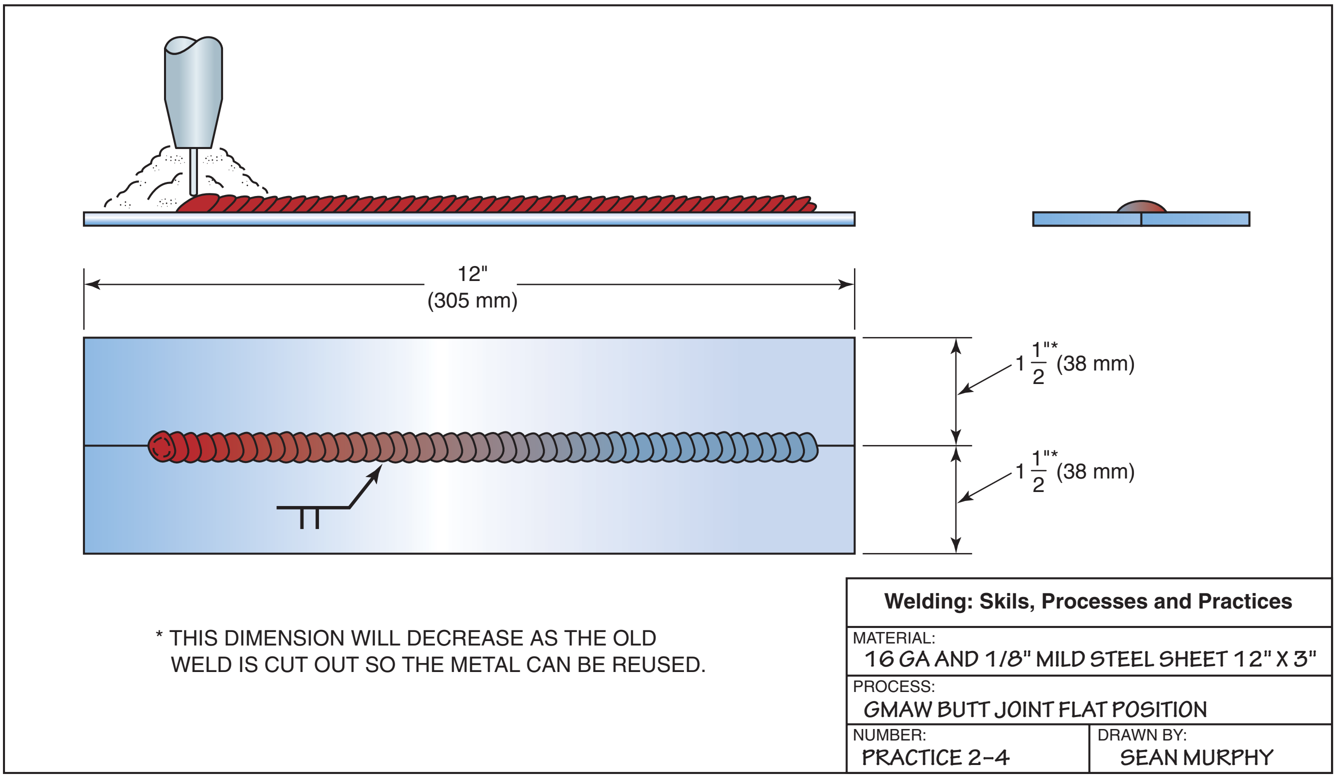
12"
(305 mm)
1 1/2"* (38 mm)
1 1/2"* (38 mm)
* THIS DIMENSION WILL DECREASE AS THE OLD WELD IS CUT OUT SO THE METAL CAN BE REUSED.
Welding: Skils, Processes and Practices
MATERIAL:
16 GA AND 1/8" MILD STEEL SHEET 12" X 3"
PROCESS:
GMAW BUTT JOINT FLAT POSITION
NUMBER:
PRACTICE 2-4
DRAWN BY:
SEAN MURPHY

Figure 2.32 **(A) Butt joint in the flat position**

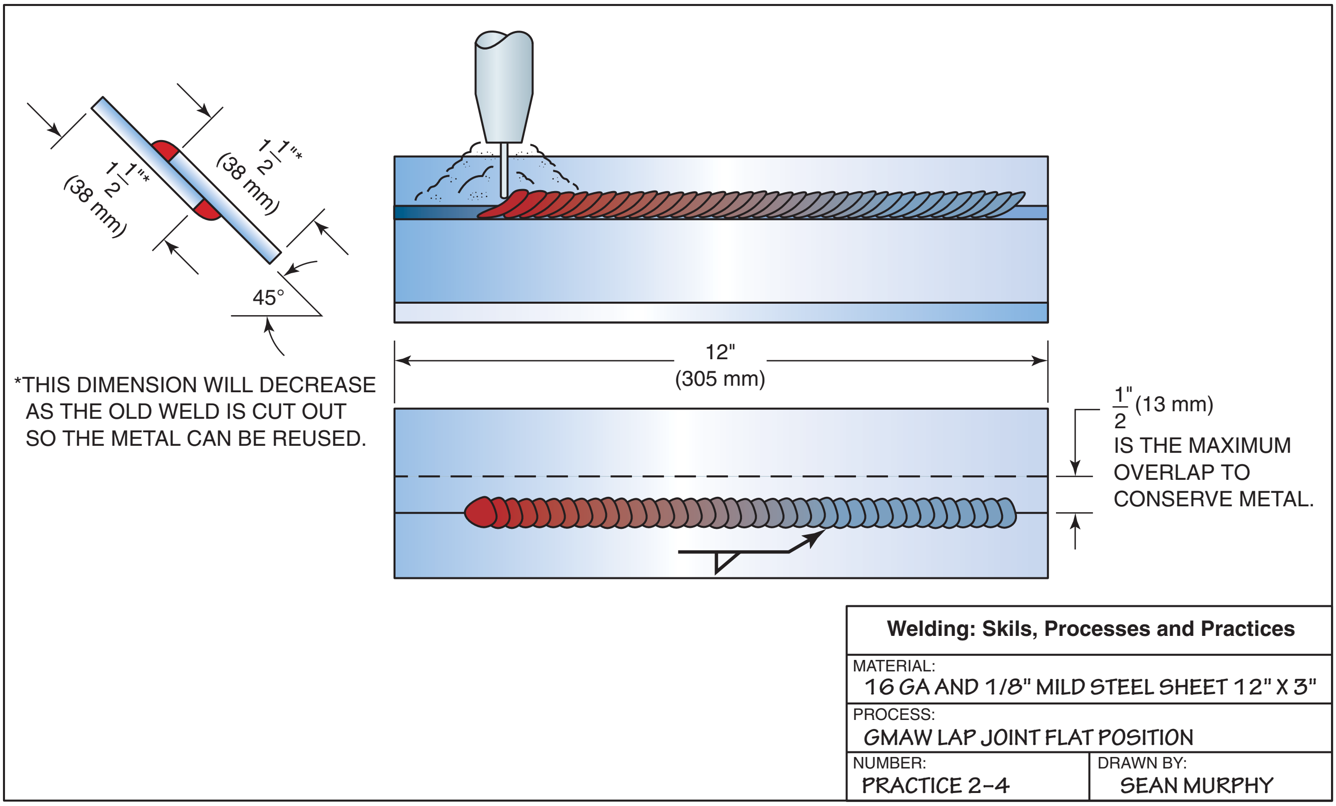
1 1/2"*
(38 mm)
1 1/2"*
(38 mm)
45°
*THIS DIMENSION WILL DECREASE AS THE OLD WELD IS CUT OUT SO THE METAL CAN BE REUSED.
12"
(305 mm)
1/2" (13 mm) IS THE MAXIMUM OVERLAP TO CONSERVE METAL.
Welding: Skils, Processes and Practices
MATERIAL:
16 GA AND 1/8" MILD STEEL SHEET 12" X 3"
PROCESS:
GMAW LAP JOINT FLAT POSITION
NUMBER:
PRACTICE 2-4
DRAWN BY:
SEAN MURPHY

Figure 2.32 (B) Lap joint in the flat position

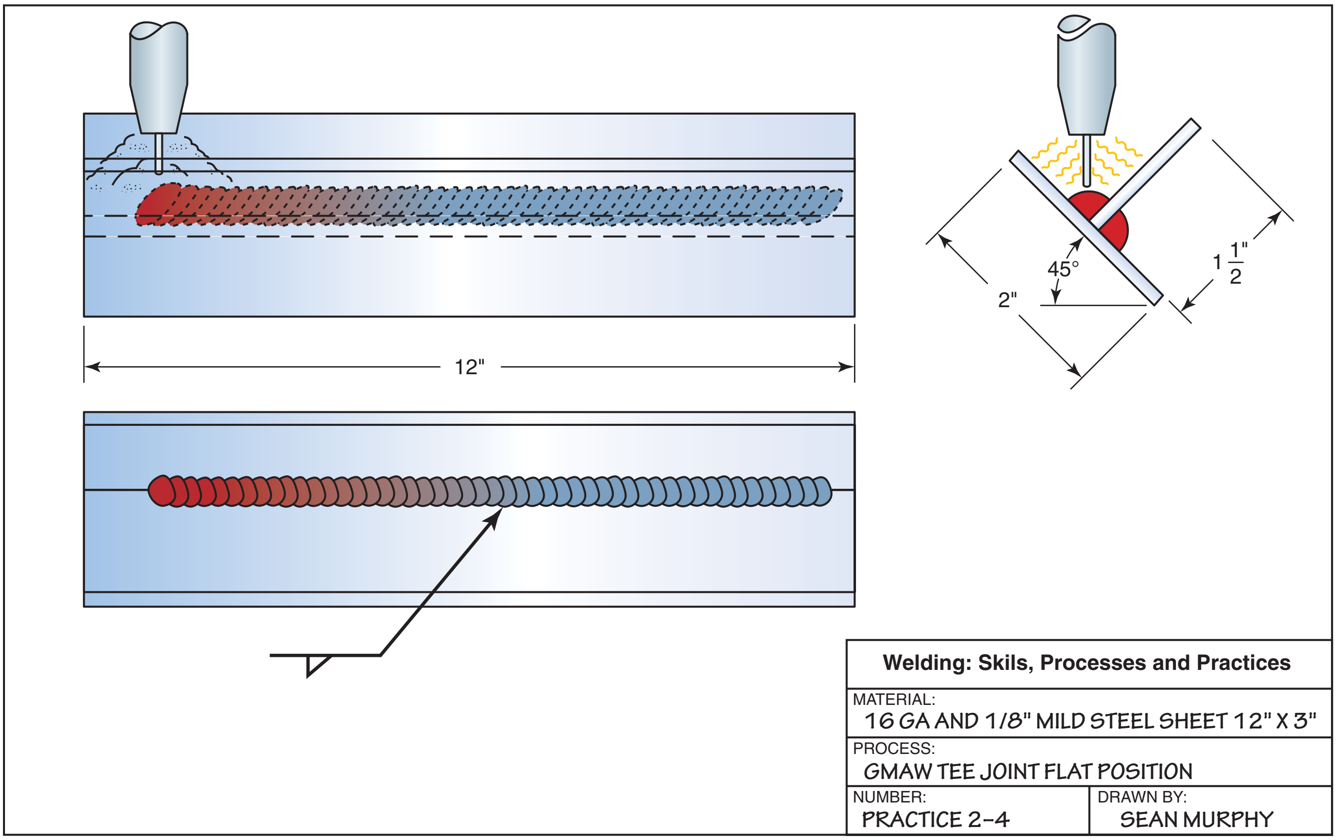
12"
45°
2"
1 1/2
Welding: Skils, Processes and Practices
MATERIAL:
16 GA AND 1/8" MILD STEEL SHEET 12" X 3"
PROCESS:
GMAW TEE JOINT FLAT POSITION
NUMBER:
PRACTICE 2-4
DRAWN BY:
SEAN MURPHY

Figure 2.32 (C) Tee joint in the flat position

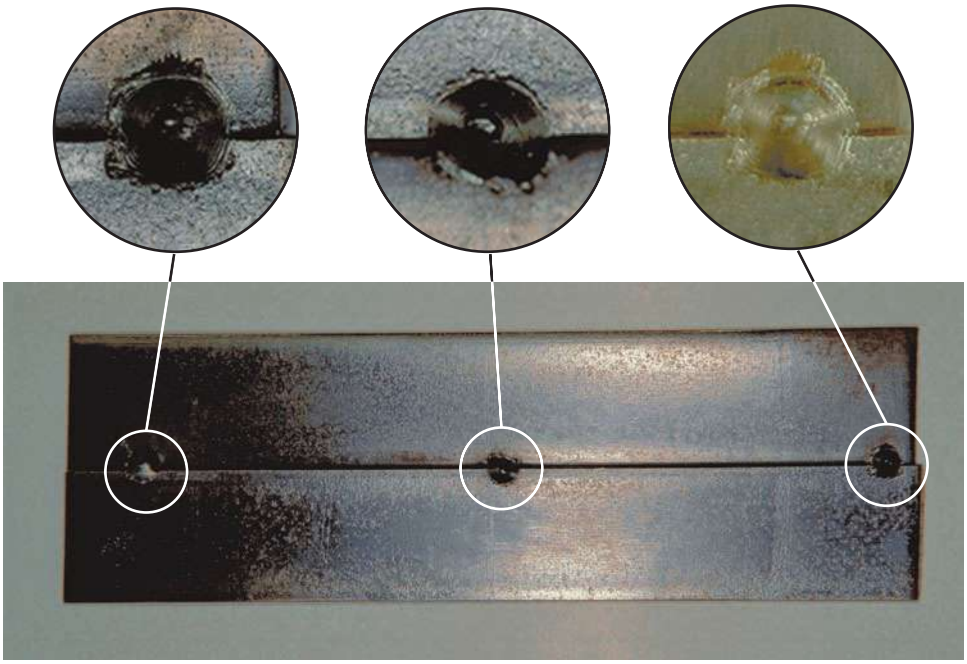

Figure 2.33 Tack welds
Use enough tack welds to keep the joint in alignment during welding. Small tack welds are easier to weld over without adversely affecting the weld.
Source: Courtesy of Larry Jeffus

- Starting at one end, run a bead along the joint. Watch the molten weld pool and bead for signs that a change in technique may be required.
- Make any needed changes as the weld progresses. By the time the weld is complete, you should be making the weld nearly perfectly.
- Using the same technique that was established in the last weld, make another weld. This time, the entire 12 in. (305 mm) of weld should be flawless.

Repeat each type of joint with both thicknesses of metal until consistently good beads are obtained. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

PRACTICE 2-5

Flat Position Butt Joint, Lap Joint, and Tee Joint, All with 100% Penetration

Using the same equipment, materials, and setup listed in Practice 2-3, make a welded joint in the flat position with 100% penetration along the entire 12-in. (305-mm) length of the welded joint. Repeat each type of joint until consistently good beads are obtained. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the “Flat” position portion of the all-position requirement.

VERTICAL UP 3G AND 3F POSITIONS

PRACTICE 2-6

Stringer Bead at a 45° Vertical Up Angle

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Vertical" position portion of the all-position requirement.

Using the same equipment, materials, and setup as listed in Practice 2-3, you will make a vertical up stringer bead on a plate at a 45° inclined angle.

Start at the bottom of the plate and hold the welding gun at a slight pushing or upward angle to the plate, Figure 2.34. Brace yourself, lower your hood, and begin to weld. Depending upon the machine settings and type of shielding gas used, you will make a weave pattern.

If the molten weld pool is large and fluid (hot), use a "C" or "J" weave pattern to allow a longer time for the molten weld pool to cool, Figure 2.35. Do not make the weave so long or fast that the wire is allowed to strike the metal ahead of the molten weld pool. If this happens, spatter increases and a spot or zone of incomplete fusion may occur, Figure 2.36.

If the molten weld pool is small and controllable, use a small "C," zigzag, or "J" weave pattern to control the width and buildup of the weld. A slower speed can also be used. Watch for complete fusion along the leading edge of the molten weld pool. Figure 2.37 shows a weld that did not fuse with the plate.

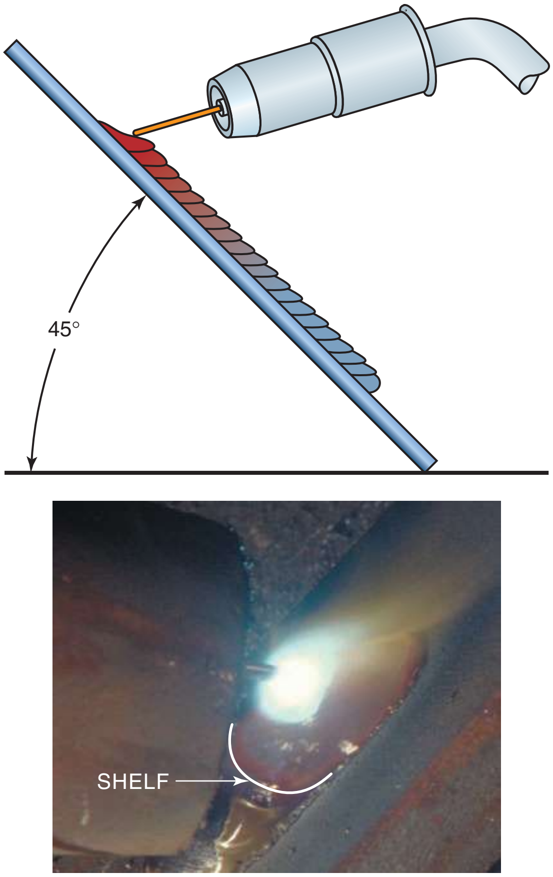

Figure 2.34 Vertical up position
Source: Photo courtesy of Larry Jeffus

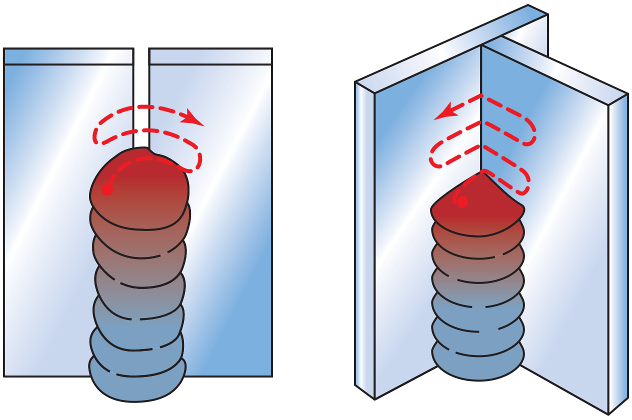

Figure 2.35 Vertical up weld patterns
Left, "C" pattern; right, "J" pattern.

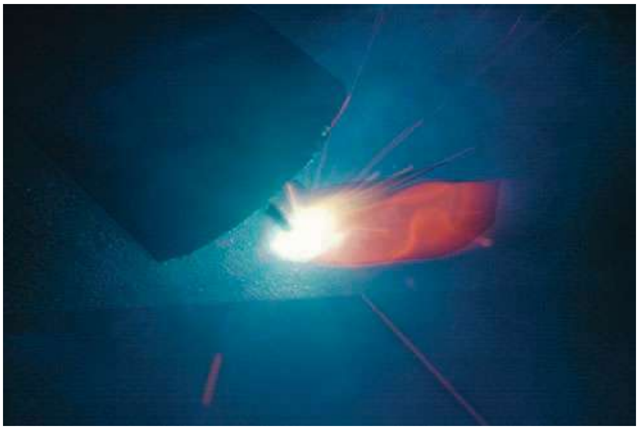

Figure 2.36 Burst of spatter caused by incorrect electrode contact with base metal
Source: Courtesy of Larry Jeffus

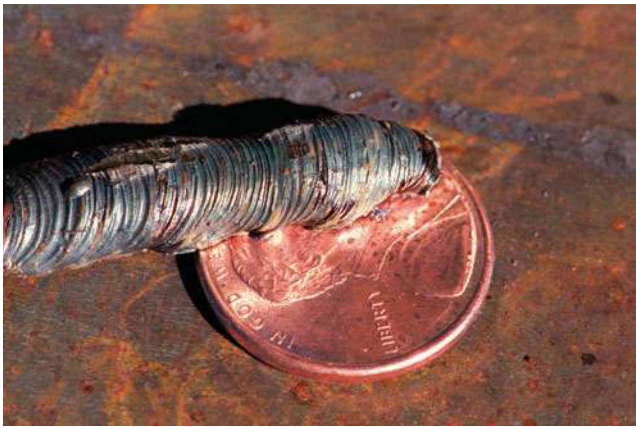

Figure 2.37 **Weld separated from the plate**
There is no fusion between the weld and plate.
Source: Courtesy of Larry Jeffus

A weld that is high and has little or no fusion is too "cold." Changing the welding technique will not correct this problem. The welder must stop welding and make the needed adjustments.

As the weld progresses up the plate, the back or trailing edge of the molten weld pool will cool, forming a shelf to support the molten metal. Watch the shelf to be sure that molten metal does not run over, forming a drip. When it appears that the metal may flow over the shelf, either increase the weave lengths, lengthen the electrode extension, or stop and start the current for brief moments to allow the weld to cool. Stopping for brief moments will not allow the shielding gas to be lost.

Continue to weld along the entire 12-in. (305-mm) length of the plate. Repeat this weld until a straight and uniform weld bead is produced. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

■ **Module 1**
Key Indicator 1, 2, 3, 4

▼ **Module 2**
Key Indicator 1, 2, 3, 4, 7

■ **Module 5**
Short-circuit Transfer
Key Indicator 4

PRACTICE 2-7

Stringer Bead in the Vertical Up Position

Repeat Practice 2-6 and increase the angle of the plate until you have mastered a straight and uniform weld bead in the vertical up position. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

▼ **Module 1**
Key Indicator 1, 2, 3, 4

▼ **Module 2**
Key Indicator 1, 2, 3, 4, 7

▼ **Module 5**
Key Indicator 5, 6
This exercise addresses the "Vertical" position portion of the all-position requirement.

PRACTICE 2-8

Butt Joint, Lap Joint, and Tee Joint in the Vertical Up Position at a 45° Angle

Using the same equipment, materials, and setup as listed in Practice 2-3, you will make vertical up welded joints on a plate at a 45° inclined angle.

Tack weld the metal pieces together and brace them in position. Check to see that you have free movement along the entire joint to prevent stopping and restarting during the weld. Avoiding stops and starts both speeds up the welding time and eliminates discontinuities.

The weave pattern should allow for adequate fusion on both edges of the joint. Watch the edges to be sure that they are being melted so that adequate fusion and penetration occur.

Repeat each type of joint as needed until consistently good beads are obtained. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-9

Butt Joint, Lap Joint, and Tee Joint in the Vertical Up Position with 100% Penetration

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Vertical" position portion of the all-position requirement.

Using the same equipment, materials, and setup as listed in Practice 2-3, you will increase the plate angle gradually as you develop skill until you are making satisfactory welds in the vertical up position.

Repeat each type of joint as needed until consistently good beads are obtained. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

VERTICAL DOWN 3G AND 3F POSITIONS

The vertical down welding technique can be useful when making some types of welds. The major advantages of this technique are the following:

- Speed—Very high rates of travel are possible.
- Shallow penetration—Thin sections or root openings can be welded with little burn-through.
- Good bead appearance—The weld has a nice width-to-height ratio and is uniform.

Vertical down welds are often used on thin sheet metals or in the root pass in grooved joints. The combination of controlled penetration and higher welding speeds makes vertical down the best choice for such welds. The ease with which welds having a good appearance can be made is deceiving. Generally, more skill is required to make sound welds with this technique than in the vertical up position. The most common problem with these welds is lack of fusion or overlap. To prevent these problems, the arc must be kept at or near the leading edge of the molten weld pool.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Short-Circuit Transfer
Key Indicator 4

PRACTICE 2-10

Stringer Bead at a 45° Vertical Down Angle

Using the same equipment, materials, and setup as listed in Practice 2-3, you will make a vertical down stringer bead on a plate at a 45° inclined angle.

Holding the welding gun at the top of the plate with a slight dragging angle, Figure 2.38, will help to increase penetration, hold back the molten weld pool, and improve visibility of the weld. Be sure that your movements along the 12-in. (305-mm) length of plate are unrestricted.

Lower your hood and start the weld. Watch both the leading edge and the sides of the molten weld pool for fusion. The leading edge should flow into the base metal, not curl over it. The sides of the molten weld pool should also show fusion into the base metal and not be flashed (ragged) along the edges.

The weld may be made with or without a weave pattern. If a weave pattern is used, it should be a "C" pattern. The "C" should follow the leading edge of the weld. Some changes on the gun angle may help to increase penetration. Experiment with the gun angle as the weld progresses.

Repeat these welds until you have established a rhythm and technique that work well for you. The welds must be straight and uniform and have complete fusion. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

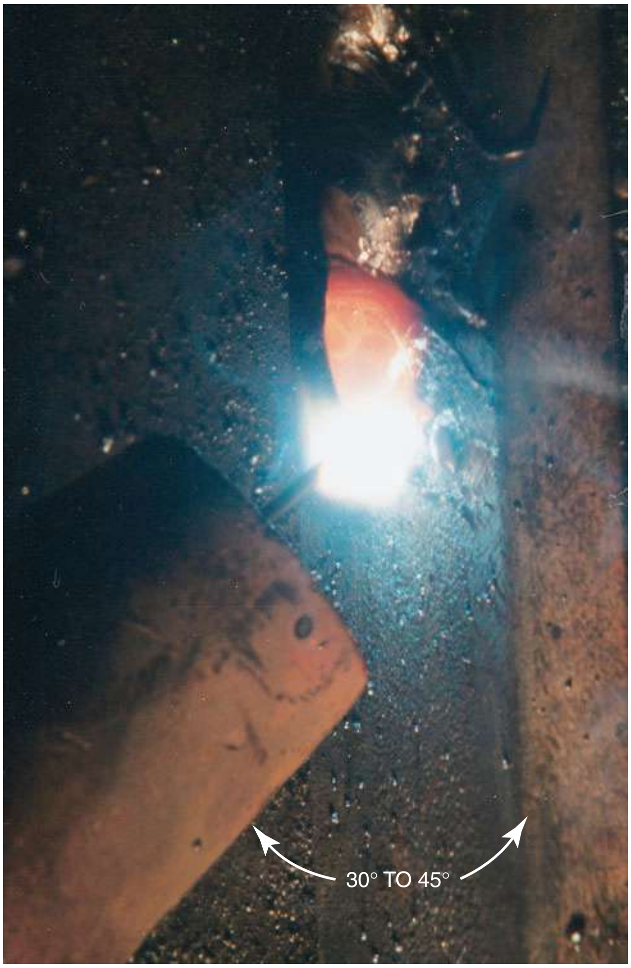

Figure 2.38 Vertical down position
Source: Courtesy of Larry Jeffus

PRACTICE 2-11

Stringer Bead in the Vertical Down Position

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Short-Circuit Transfer
Key Indicator 4

Repeat Practice 2-10 and increase the angle of the plate until you have developed the skill to repeatedly make good welds in the vertical down position. The weld bead must be straight and uniform and have complete fusion. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-12

Butt Joint, Lap Joint, and Tee Joint in the Vertical Down Position

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Vertical" position portion of the all-position requirement.

Using a properly set up and adjusted GMA welding machine, Table 2.7, proper safety protection, 0.035-in. and/or 0.045-in. (0.9-mm and/or 1.2-mm) diameter wire, and two or more pieces of mild steel sheet 12 in. (305 mm) long and 16 gauge and 1/8 in. (3 mm) thick, you will make vertical down welded joints.

Holding the welding gun at the top of the plate with a slight dragging angle, Figure 2.38, will help to increase penetration, hold back the molten weld pool, and improve visibility of the weld. Be sure that your movements along the 12-in. (305-mm) length of plate are unrestricted.

Lower your hood and start the weld. Watch both the leading edge and sides of the molten weld pool for fusion. The leading edge should flow into the base metal, not curl over it. The sides of the molten weld pool should also show fusion into the base metal and not be flashed (ragged) along the edges.

The weld may be made with or without a weave pattern. If a weave pattern is used, it should be a "C" pattern. The "C" should follow the leading edge of the weld. Some changes on the gun angle may help to increase penetration. Experiment with the gun angle as the weld progresses.

Repeat these welds until you have established a rhythm and technique that work well for you. The welds must be straight and uniform and have complete fusion. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Vertical" position portion of the all-position requirement.

PRACTICE 2-13

Butt Joint and Tee Joint in the Vertical Down Position with 100% Penetration

Using the same equipment, materials, and setup as listed in Practice 2-3, you will make welded joints with 100% weld penetration.

It may be necessary to adjust the root opening to meet the penetration requirements. The lap joint was omitted from this practice because little additional skill can be developed with it that is not already acquired with the tee joint. Repeat each type of joint until consistently good welds are obtained. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

HORIZONTAL 2G AND 2F POSITIONS

PRACTICE 2-14

Horizontal Stringer Bead at a 45° Angle

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Short-circuit Transfer
Key Indicator 4

Using the same equipment, materials, and setup as listed in Practice 2-3, you will make a horizontal stringer bead on a plate at a 45° reclined angle.

Start at one end with the gun pointed in a slightly upward direction, Figure 2.39. You may use a pushing or a dragging (a leading or a trailing) gun angle, depending upon the current setting and penetration desired. Undercutting along the top edge and overlap along the bottom edge are problems with both gun angles. Careful attention must be paid to the manipulation "weave" technique used to overcome these problems.

The most successful weave patterns are the "C" and "J" patterns. The "J" pattern is the most frequently used. It allows weld metal to be deposited along a shelf created by the previous weave, Figure 2.40. The length of the "J" can be changed to control the weld bead size. Smaller weld beads are easier to control than large ones.

Repeat these welds until you have established the rhythm and technique that work well for you. The weld must be straight and uniform and have complete fusion. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

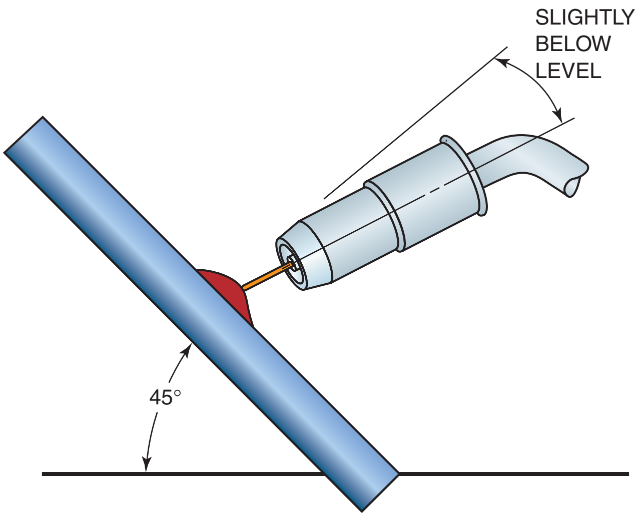

Figure 2.39 45° horizontal position

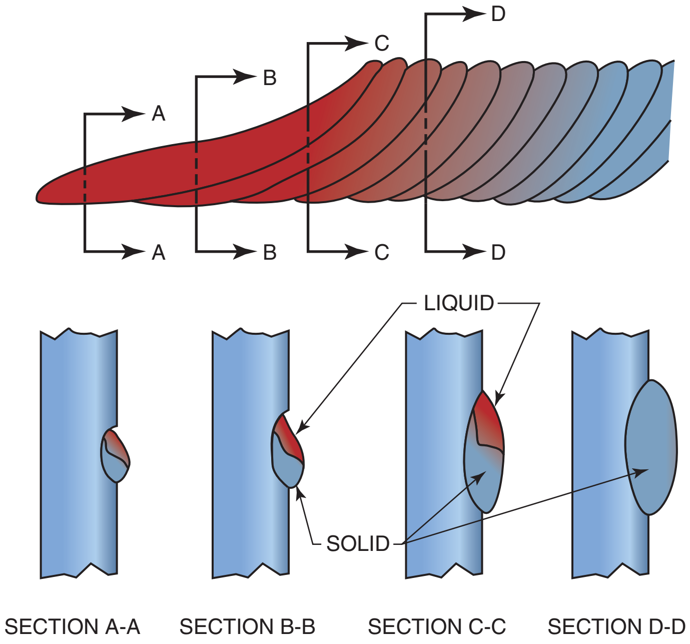

Figure 2.40 **Sections of a "J" weave pattern**
The actual size of the molten weld pool remains small along the weld.

PRACTICE 2-15

Stringer Bead in the Horizontal Position

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Repeat Practice 2-14 and increase the angle of the plate until you have developed the skill to repeatedly make good horizontal welds on a vertical surface. The weld bead must be straight and uniform and have complete fusion. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-16

Butt Joint, Lap Joint, and Tee Joint in the Horizontal Position

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Horizontal" position portion of the all-position requirement.

Using the same equipment, materials, and setup listed in Practice 2-3, you will make horizontal welded joints.

Tack weld the pieces of metal together and brace them in position using the same skills developed in Practice 2-14. Starting at one end, make a weld along the entire length of the joint. When making the butt or lap joints, it may help to recline the plates at a 45° angle until you have developed the technique required. Repeat each type of joint as needed until consistently good welds are obtained. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-17

Butt Joint and Tee Joint in the Horizontal Position with 100% Penetration

Using the same equipment, materials, and setup as listed in Practice 2-3, you will make overhead joints having 100% penetration in the horizontal position.

It may be necessary to adjust the root opening to meet the penetration requirements. Repeat each type of joint as needed until consistently good welds are obtained. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Horizontal" position portion of the all-position requirement.

OVERHEAD 4G AND 4F POSITIONS

There are several advantages to the use of short-circuiting arc metal transfer in the overhead position, including the following:

- Small molten weld pool size—The smaller size of the molten weld pool allows surface tension to hold it in place. Less molten weld pool sag results in improved bead contour with less undercut and fewer icicles, Figure 2.41.
- Direct metal transfer—The direct metal transfer method does not rely on other forces to get the filler metal into the molten weld pool. This results in efficient metal transfer and less spatter and loss of filler metal.

PRACTICE 2-18

Stringer Bead in the Overhead Position

Using the same equipment, materials, and setup as listed in Practice 2-3, you will make a welded stringer bead in the overhead position.

The molten weld pool should be kept as small as possible for easier control. A small molten weld pool can be achieved by using lower current settings, by using a longer wire stickout, by traveling faster, or by pushing the molten weld pool. The technique used is the welder's choice. Often a combination of techniques can be used with excellent results.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Short-circuit Transfer
Key Indicator 4

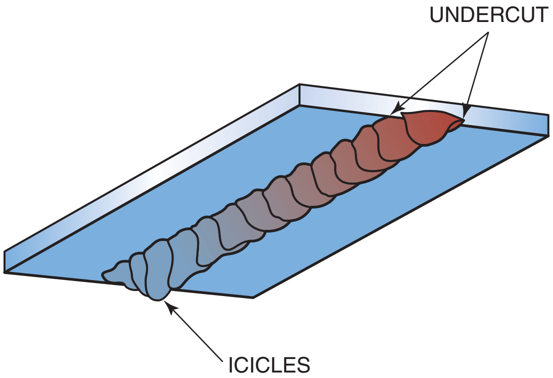

Figure 2.41 **Overhead weld**

Lower current settings require closer control of gun manipulation to ensure that the wire is fed into the molten weld pool just behind the leading edge. The low power will cause overlap and more spatter if this contact position of wire to molten weld pool is not closely maintained.

Faster travel speeds allow the welder to maintain a high production rate even if multiple passes are required to complete the weld. Weld penetration into the base metal at the start of the bead can be obtained by using a slow start or quickly reversing the weld direction. Both the slow start and reversal of weld direction put more heat into the start to increase penetration, Figure 2.42. The higher speed also reduces the amount of weld distortion by reducing the amount of time that heat is applied to a joint.

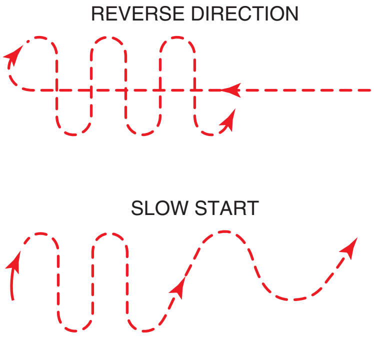

Figure 2.42 Two methods of concentrating heat at the beginning of a weld bead to aid in penetration depth

The pushing, or trailing, gun angle forces the bead to be flatter by spreading it out over a wider area as compared to the bead resulting from a dragging, or backhand, gun angle. The wider, shallow molten weld pool cools faster, resulting in less time for sagging and the formation of icicles.

When welding overhead, extra personal protection is required to reduce the danger of burns. Wear leather sleeves or a leather jacket, and a cap.

Much of the spatter created during overhead welding falls into the shielding gas nozzle. The effectiveness of the shielding gas is reduced, Figure 2.43, and the contact tube may short out to the gas nozzle, Figure 2.44. Turbulence caused by the spatter obstructing the gas may lead to weld contamination. The shorted gas nozzle may arc to the work, causing damage both to the nozzle and to the plate. To control the amount of spatter, a longer stickout and/or a sharper gun-to-plate angle is required to allow most of the spatter to fall clear of the gas nozzle. The nozzle can be dipped, sprayed, or injected automatically, Figure 2.45, with antispatter to help prevent the spatter from sticking. Applying antispatter will not stop the spatter from building up, but it does make its removal much easier. Ensure that the antispatter gel does not restrict the flow of shield gas from the nozzle.

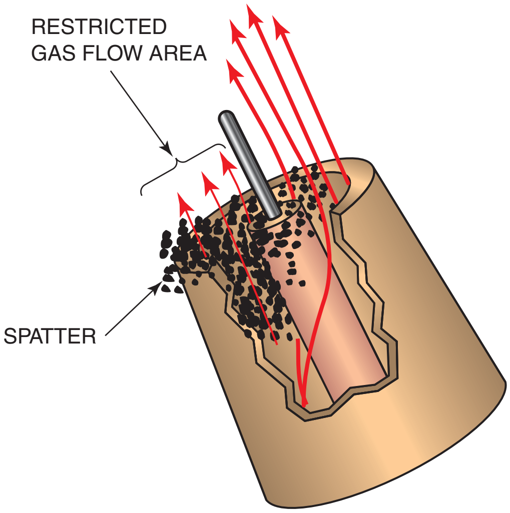

Figure 2.43 Shielding gas flow affected by excessive weld spatter in nozzle

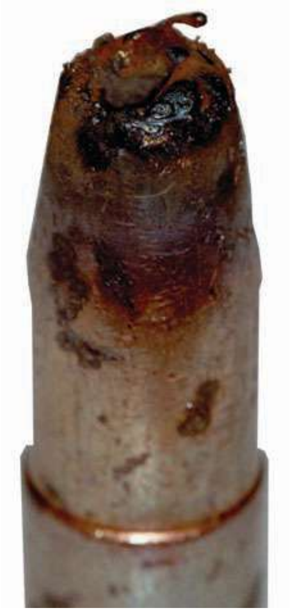

Figure 2.44 Gas nozzle damaged after shorting out against the work
Source: Courtesy of Larry Jeffus

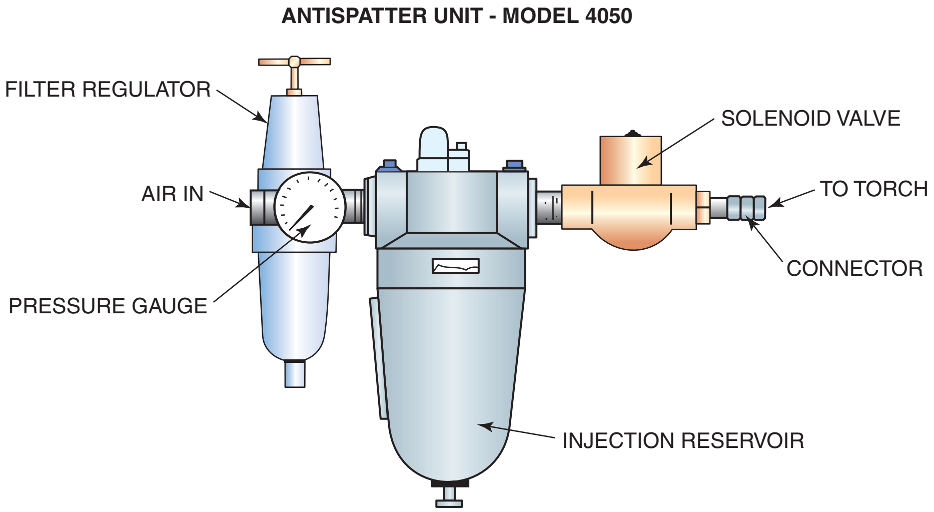

Figure 2.45 **Automatic antispatter system that can be added to a GMA welding gun**

Make several short weld beads using various techniques to establish the method that is most successful and most comfortable for you. After each weld, stop and evaluate it before making a change. When you have decided on the technique to be used, make a welded stringer bead that is 12 in. (305 mm) long.

Repeat the weld until it can be made straight, uniform, and free from any visual defects. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-19

Butt Joint, Lap Joint, and Tee Joint in the Overhead Position

Using the same equipment, materials, and setup as listed in Practice 2-3, you will make an overhead welded joint.

Tack weld the pieces of metal together and secure them in the overhead position. Be sure you have an unrestricted view and freedom of movement along the joint. Start at one end and make a weld along the joint. Use the same technique developed in Practice 2-18.

Repeat the weld until it can be made straight, uniform, and free from any visual defects. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Overhead" position portion of the all-position requirement.

PRACTICE 2-20

Butt Joint and Tee Joint in the Overhead Position with 100% Penetration

Using the same equipment, materials, and setup as listed in Practice 2-3, you will make overhead welded joints having 100% penetration.

Tack weld the metal together. It may be necessary to adjust the root opening to allow 100% weld metal penetration. During these welds, it may be necessary to use a dragging, or backhand, torch angle. When

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Key Indicator 5, 6
This exercise addresses the "Overhead" position portion of the all-position requirement.

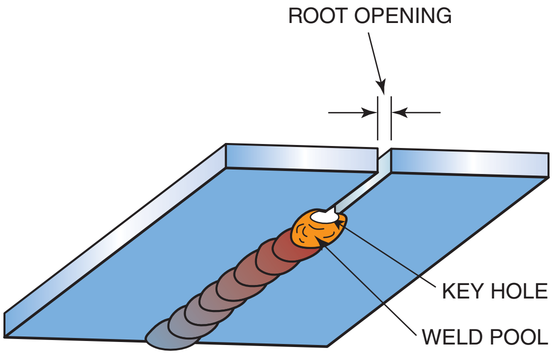

Figure 2.46 Overhead welding

used with a "C" or "J" weave pattern, this torch angle helps to achieve the desired depth of penetration. A key hole just ahead of the molten weld pool is a good sign that the metal is being penetrated, Figure 2.46.

Repeat the weld until it can be made straight, uniform, and free from any visual defects. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

AXIAL SPRAY TRANSFER

PRACTICE 2-21

Stringer Bead, 1G Position

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Spray Transfer
Key Indicator 9

Using a properly set up and adjusted GMA welding machine (see Table 2.8), proper safety protection, 0.035-in. and/or 0.045-in. (0.9-mm and/or 1.2-mm)-diameter wire, and two or more pieces of mild steel plate 12 in. (305 mm) long × 1/4 in. (6 mm) thick, you will make a welded stringer bead in the flat position.

Start at one end of the plate and use either a push or drag technique to make a weld bead along the entire 12-in. (305-mm) length of the metal using spray or pulsed-arc metal transfer. After the weld is complete, check its appearance and make any changes needed to correct the weld, Figure 2.47. Repeat the weld and make any additional adjustments required. After the machine is set, start working on improving the straightness and uniformity of the weld. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Table 2.8 Typical Welding Current Settings for Axial Spray Metal Transfer for Mild Steel

Process	Wire Diameter	Amperage Range (Optimum)	Voltage Range (Optimum)	Shielding Gas
Axial spray	0.030	115–200	15–27	98% Ar + 2% O_2
	0.035	165–300	18–32	
	0.045	200–450	20–34	

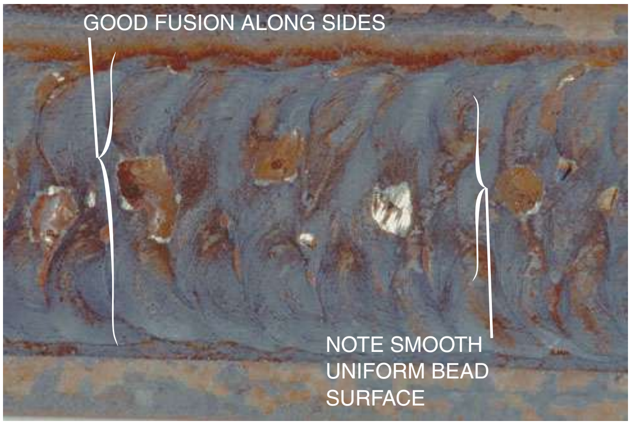

Figure 2.47 **Weld bead made with GMAW axial spray metal transfer**
Source: Courtesy of Larry Jeffus

PRACTICE 2-22

Butt Joint, Lap Joint, and Tee Joint Using the Spray Transfer Method

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Spray Transfer
Key Indicator 10, 11

Using the same equipment, materials, and setup as listed in Practice 2-21, you will make a flat and horizontal weld using spray transfer or pulsed-spray metal transfer, Figure 2.48.

Tack weld the metal together and place the assembly in the flat position on the welding table. Start at one end and make a uniform weld along the entire 12-in. (305-mm) length of the joint. Watch the sides of the fillet weld for signs of undercutting.

Repeat the weld until it can be made straight, uniform, and free from any visual defects. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

PRACTICE 2-23

Butt Joint and Tee Joint

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 5
Spray Transfer
Key Indicator 10, 11

Using the same equipment, materials, and setup as listed in Practice 2-21, you will make a flat weld using spray transfer. Each weld must pass the

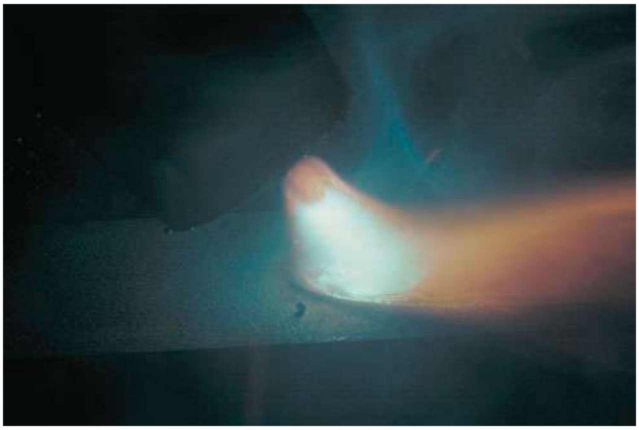

Figure 2.48 **GMAW axial spray metal transfer**
Source: Courtesy of Larry Jeffus

guided bend test. Repeat each type of weld joint as needed until the bend test can be passed. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

PRACTICE 2-24

Gas Metal Arc Welding—Short-Circuit Metal Transfer (GMAW-S) Workmanship Sample

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 3
Key Indicator 1, 2, 3

Module 5
Short-circuit Transfer
Key Indicator 3, 4, 5, 6, 7

Module 9
Key Indicator 1, 2

Welding Procedure Specification (WPS)

Welding Procedure Specification No.: Practice 2-24 Date:

Title:
Welding GMAW-S of plate to plate.

Scope:
This procedure is applicable for square groove and fillet welds within the range of 10 ga. (3.4 mm) through 14 ga. (1.9 mm).

Welding may be performed in the following positions: all.

Base Metal:
The base metal shall conform to carbon steel M-1, P-1, and S-1 Group 1 or 2.

Backing material specification: none.

Filler Metal:
The filler metal shall conform to AWS specification no. E70S-X from AWS specification A5.18. This filler metal falls into F-number F-6 and A-number A-1.

Shielding Gas:
The shielding gas, or gases, shall conform to the following compositions and purity:

CO_2 at 30 to 50 cfh or 75% Ar/25% CO_2 at 30 to 50 cfh.

Joint Design and Tolerances:

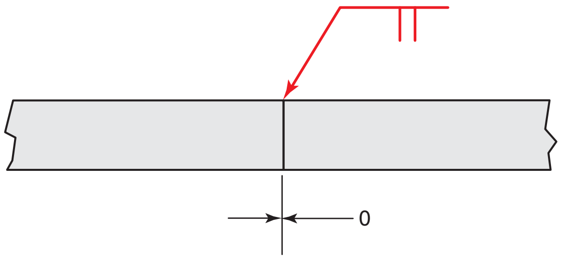

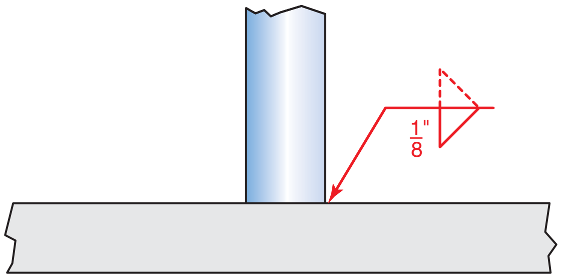

Preparation of Base Metal:

All parts may be mechanically cut or machine PAC unless specified as manual PAC.

All hydrocarbons and other contaminations, such as cutting fluids, grease, oil, and primers, must be cleaned off all parts and filler metals before welding. This cleaning can be done with any suitable solvents or detergents. The groove face and inside and outside plate surface within 1 in. (25 mm) of the joint must be mechanically cleaned of slag, rust, and mill scale. Cleaning must be done with a wire brush or grinder down to bright metal.

Electrical Characteristics:

The current shall be direct-current electrode positive (DCEP). The base metal shall be on the negative side of the line.

Electrode		Welding Power			Shielding Gas		Base Metal	
Type	Size	Amps	Wire-feed Speed, ipm (cm/min)	Volts	Type	Flow	Type	Thickness
E70S-X	0.035 in. (0.9 mm)	90 to 120	180 to 300 (457 to 762)	15 to 19	CO_2 or 75% Ar/ CO_2 25%	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)
E70S-X	0.045 in. (1.2 mm)	130 to 200	125 to 200 (318 to 508)	17 to 20	CO_2 or 75% Ar/ CO_2 25%	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)

Preheat:

The parts must be heated to a temperature higher than 50°F (10°C) before any welding is started.

Backing Gas:

N/A

Safety:

Proper protective clothing and equipment must be used. The area must be free of all hazards that may affect the welder or others in the area. The welding machine, welding leads, work clamp, electrode holder, and other equipment must be in safe working order.

Welding Technique:

Using a 1/2-in. (13-mm) or larger gas nozzle for all welding, first tack weld the plates together according to the drawing. Use the E70S-X filler metal to fuse the plates together. Clean any silicon slag, being sure to remove any trapped silicon slag along the sides of the weld.

Using the E70S-X arc welding electrodes, make a series of stringer beads, no thicker than 3/16 in. (4.7 mm). The 1/8-in. (3.1-mm) fillet welds are to be made with one pass. All welds must be placed in the orientation shown in the drawing.

Interpass Temperature:

The plate should not be heated to a temperature higher than 350°F (175°C) during the welding process. After each weld pass is completed, allow it to cool but never to a temperature below 50°F (10°C). The weldment must not be quenched in water.

Cleaning:

Any slag must be cleaned off between passes. The weld beads may be cleaned by a hand wire brush, a hand chipping, a punch and hammer, or a needle-scaler. All weld cleaning must be performed with the test plate in the welding position.

Visual Inspection:

Visually inspect the weld for uniformity and discontinuities.

1. There shall be no cracks, no incomplete fusion.
2. There shall be no incomplete joint penetration in groove welds except as permitted for partial joint penetration welds.
3. The Test Supervisor shall examine the weld for acceptable appearance, and shall be satisfied that the welder is skilled in using the process and procedure specified for the text.
4. Undercut shall not exceed the lesser of 10% of the base metal thickness or 1/32 in. (0.8 mm).
5. Where visual examination is the only criterion for acceptance, all weld passes are subject to visual examination, at the discretion of the Test Supervisor.
6. The frequency of porosity shall not exceed one in each 4 in. (100 mm) of weld length and the maximum diameter shall not exceed 3/32 in. (2.4 mm)
7. Welds shall be free from overlap.

Sketches:

See Figure 2.49.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

PRACTICE 2-25

Gas Metal Arc Welding—Short-Circuit Metal Transfer (GMAW-S) Limited Thickness Welder Performance Qualification Test Plate for 2G, 3G, and 4G Positions without Backing

This practice parallels an AWS D1.1 structural steel limited thickness welder performance qualification test.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 3
Key Indicator 1, 2, 3

Module 5
Short-circuit Transfer
Key Indicator 3, 4, 5, 6

Module 9
Key Indicator 1, 2

Welding Procedure Specification (WPS)

Welding Procedure Specification No.: Practice 2-25 Date:

Title:

Welding GMAW-S of plate to plate.

Scope:

This procedure is applicable for V-groove, bevel, or single-bevel welds within the range of 1/8 in. (3.2 mm) through 3/4 in. (19 mm).

Welding may be performed in the following positions: all.

Base Metal:

The base metal shall conform to carbon steel M-1, P-1, and S-1 Group 1 or 2.

Backing material specification: none.

Filler Metal:

The filler metal shall conform to AWS specification no. E70S-X from AWS specification A5.18. This filler metal falls into F-number F-6 and A-number A-1.

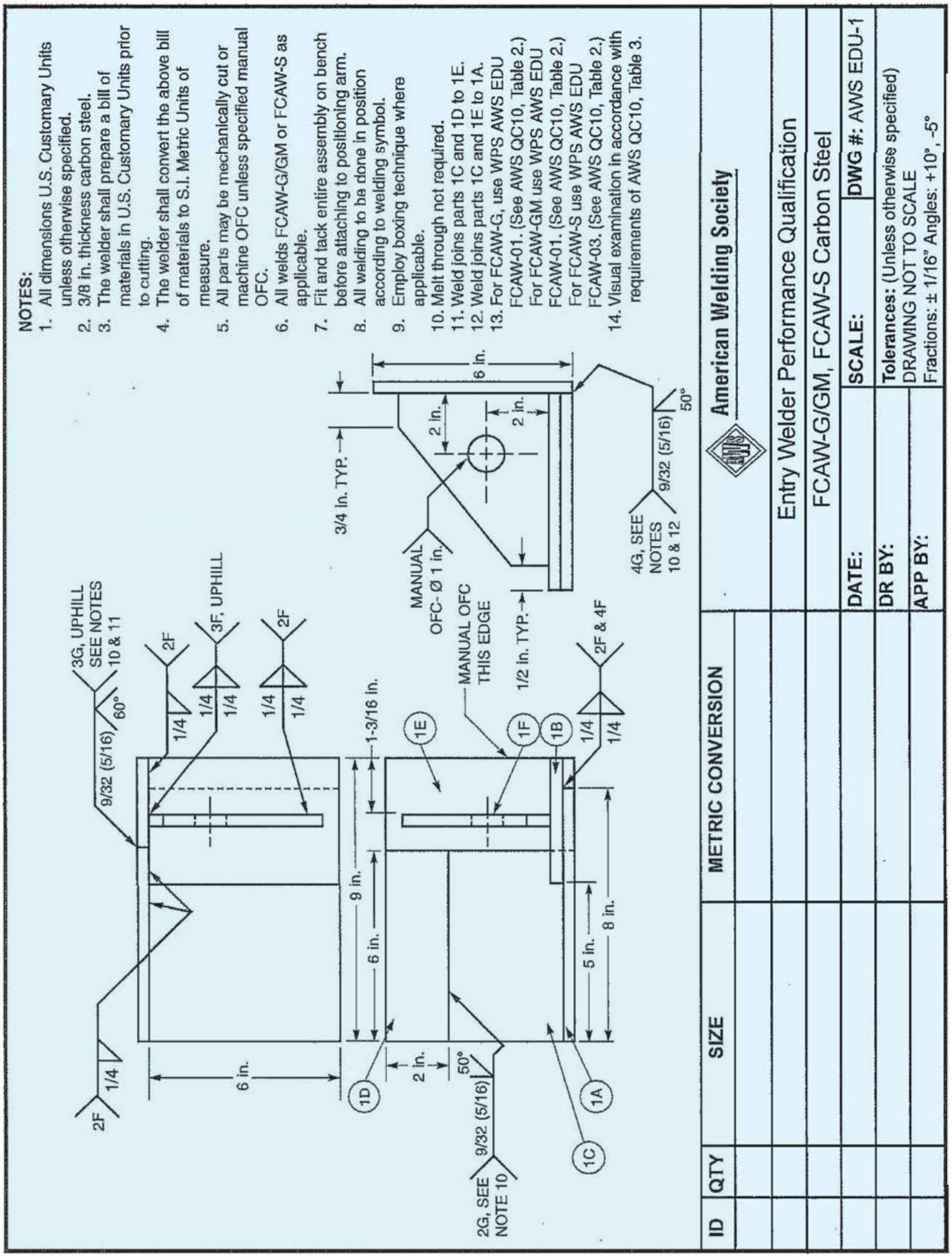
2F
1/4
6 in.
2G, SEE NOTE 10
9/32 (5/16)
50°
2 in.
1D
1C
1A
6 in.
9 in.
5 in.
8 in.
1-3/16 in.
9/32 (5/16)
60°
3G, UPHILL SEE NOTES 10 & 11
2F
1/4
3F, UPHILL
1/4
1/4
2F
1/4
1/4
1E
1F
1B
MANUAL OFC THIS EDGE
2F & 4F
1/4
1/4
1/2 in. TYP.
MANUAL OFC- Ø 1 in.
3/4 in. TYP.
2 in.
2 in.
6 in.
4G, SEE NOTES 10 & 12
9/32 (5/16)
50°

NOTES:

1. All dimensions U.S. Customary Units unless otherwise specified.
2. 3/8 in. thickness carbon steel.
3. The welder shall prepare a bill of materials in U.S. Customary Units prior to cutting.
4. The welder shall convert the above bill of materials to S.I. Metric Units of measure.
5. All parts may be mechanically cut or machine OFC unless specified manual OFC.
6. All welds FCAW-G/GM or FCAW-S as applicable.
7. Fit and tack entire assembly on bench before attaching to positioning arm.
8. All welding to be done in position according to welding symbol.
9. Employ boxing technique where applicable.
10. Melt through not required.
11. Weld joins parts 1C and 1D to 1E.
12. Weld joins parts 1C and 1E to 1A.
13. For FCAW-G, use WPS AWS EDU FCAW-01. (See AWS QC10, Table 2.) For FCAW-GM use WPS AWS EDU FCAW-01. (See AWS QC10, Table 2.) For FCAW-S use WPS AWS EDU FCAW-03. (See AWS QC10, Table 2.)
14. Visual examination in accordance with requirements of AWS QC10, Table 3.

ID	QTY	SIZE	METRIC CONVERSION

American Welding Society

Entry Welder Performance Qualification

FCAW-G/GM, FCAW-S Carbon Steel

DATE:	SCALE:	DWG #: AWS EDU-1
DR BY:	**Tolerances:** (Unless otherwise specified) DRAWING NOT TO SCALE Fractions: ± 1/16" Angles: +10°, -5°	
APP BY:		

Figure 2.49 GMAW-S workmanship qualification test

Source: Courtesy of the American Welding Society

Shielding Gas:

The shielding gas, or gases, shall conform to the following compositions and purity: CO_2 at 30 to 50 cfh or 75% Ar/25% CO_2 at 30 to 50 cfh.

Joint Design and Tolerances:

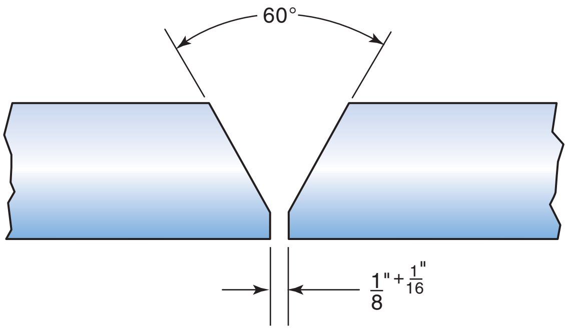

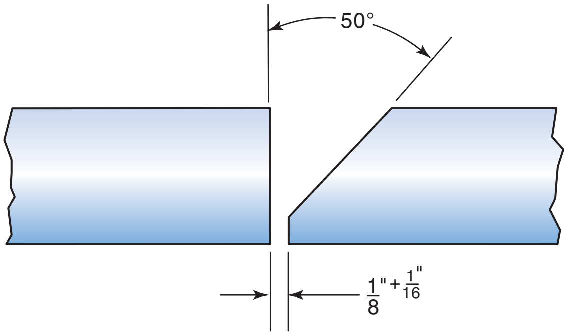

Preparation of Base Metal:

The bevels are to be flame cut on the edges of the plate before the parts are assembled. The beveled surface must be smooth and free of notches. Any roughness or notches deeper than 1/64 in. (0.4 mm) must be ground smooth.

All hydrocarbons and other contaminations, such as cutting fluids, grease, oil, and primers, must be cleaned off all parts and filler metals before welding. This cleaning can be done with any suitable solvents or detergents. The groove face and inside and outside plate surface within 1 in. (25 mm) of the joint must be mechanically cleaned of slag, rust, and mill scale. Cleaning must be done with a wire brush or grinder down to bright metal.

Electrical Characteristics:

The current shall be direct-current electrode positive (DCEP). The base metal shall be on the negative side of the line.

Electrode		Welding Power			Shielding Gas		Base Metal	
Type	**Size**	**Amps**	**Wire-feed Speed, ipm (cm/min)**	**Volts**	**Type**	**Flow**	**Type**	**Thickness**
E70S-X	0.035 in. (0.9 mm)	90 to 120	180 to 300 (457 to 762)	15 to 19	CO_2 or 75% Ar/ CO_2 25%	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)
E70S-X	0.045 in. (1.2 mm)	130 to 200	125 to 200 (318 to 508)	17 to 20	CO_2 or 75% Ar/ CO_2 25%	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)

Preheat:
The parts must be heated to a temperature higher than 50°F (10°C) before any welding is started.

Backing Gas:
N/A

Safety:
Proper protective clothing and equipment must be used. The area must be free of all hazards that may affect the welder or others in the area. The welding machine, welding leads, work clamp, electrode holder, and other equipment must be in safe working order.

Welding Technique:
Using a 1/2-in. (13-mm) or larger gas nozzle for all welding, first tack weld the plates together according to the drawing. There should be about a 1/8-in. (1.6-mm) root gap between the plates with V-grooved or beveled edges and 1/16-in. root faces. Use the E70S-X filler wire to make a root pass to fuse the plates together. Clean any silicon slag from the root pass, being sure to remove any trapped silicon slag along the sides of the weld.

Using the E70S-X filler wire, make a series of stringer or weave filler welds, no thicker than 1/4 in. (6.4 mm), in the groove until the joint is filled. Note: The horizontal (2G) weldment should be made with stringer beads only.

Interpass Temperature:
The plate should not be heated to a temperature higher than 350°F (175°C) during the welding process. After each weld pass is completed, allow it to cool but never to a temperature below 50°F (10°C). The weldment must not be quenched in water.

Cleaning:
Any slag must be cleaned off between passes. The weld beads may be cleaned by a hand wire brush, a hand chipping, a punch and hammer, or a needle-scaler. All weld cleaning must be performed with the test plate in the welding position.

Visual Inspection:*
Visually inspect the weld for uniformity and discontinuities.

1. There shall be no cracks, no incomplete fusion.
2. There shall be no incomplete joint penetration in groove welds except as permitted for partial joint penetration welds.
3. The Test Supervisor shall examine the weld for acceptable appearance, and shall be satisfied that the welder is skilled in using the process and procedure specified for the text.
4. Undercut shall not exceed the lesser of 10% of the base metal thickness or 1/32 in. (0.8 mm)
5. Where visual examination is the only criterion for acceptance, all weld passes are subject to visual examination, at the discretion of the Test Supervisor.

---

*From Table 3, AWS SENSE QC-10 2004 (Courtesy of the American Welding Society)

6. The frequency of porosity shall not exceed one in each 4 in. (100 mm) of weld length and the maximum diameter shall not exceed 3/32 in. (2.4 mm).
7. Welds shall be free from overlap.

Bend Test:

The weld is to be mechanically tested only after it has passed the visual inspection. Be sure that the test specimens are properly marked to identify the welder, the position, and the process.

Specimen Preparation:

For 3/8-in. test plates, two specimens are to be located in accordance with the requirements of the figure below left. One is to be prepared for a "transverse face bend," and the other is to be prepared for a "transverse root bend."

Transverse face bend. The weld is perpendicular to the longitudinal axis of the specimen and is bent so that the weld face becomes the tension surface of the specimen. Transverse face-bend specimens shall comply with the requirements of the figure below bottom.

Transverse root bend. The weld is perpendicular to the longitudinal axis of the specimen and is bent so that the weld root becomes the tension surface of the specimen. Transverse face-bend specimens shall comply with the requirements of the figure below bottom.

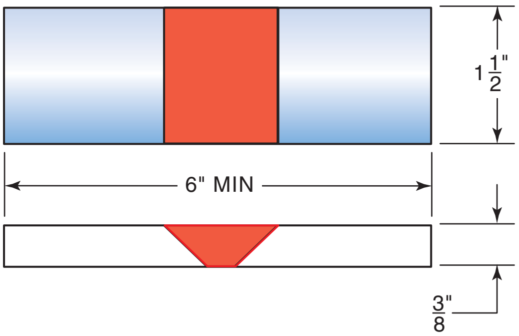

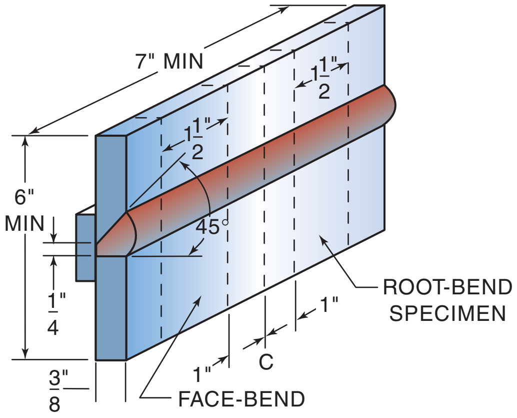

Acceptance Criteria for Face and Root Bends:*

For acceptance, the convex surface of the face- and root-bend specimens shall meet both of the following requirements:

*From Table 4, AWS-SENSE QC-10:2004 (Courtesy of the American Welding Society)

1. No single indication shall exceed 1/8 in. (3.2 mm), measured in any direction on the surface.
2. The sum of the greatest dimensions of all indications on the surface, which exceed 1/32 in. (0.8 mm) but are less than or equal to 1/8 in. (3.2 mm), shall not exceed 3/8 in. (9.6 mm).

Cracks occurring at the corner of the specimens shall not be considered unless there is definite evidence that they result from slag inclusion or other internal discontinuities.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-26

Gas Metal Arc Welding Spray Transfer (GMAW) Workmanship Sample

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 3
Key Indicator 1, 2, 3

Module 5
Spray Transfer
Key Indicator 8, 9, 10, 11, 12

Module 9
Key Indicator 1, 2

Welding Procedure Specification (WPS)

Welding Procedure Specification No.: Practice 2-26. Date:

Title:
Welding GMAW of plate to plate.

Scope:
This procedure is applicable for V-groove and fillet welds within the range of 1/8 in. (3.2 mm) through 1-1/2 in. (38 mm).

Welding may be performed in the following positions: 1G, 1F, 2F.

Base Metal:
The base metal shall conform to carbon steel M-1, P-1, and S-1, Group 1 or 2.

Backing material specification: none.

Filler Metal:
The 0.035 to 0.045 dia. filler metal shall conform to AWS specification no. E70S-X from AWS specification A5.18. This filler metal falls into F-number F-6 and A-number A-1.

Shielding Gas:
The shielding gas, or gases, shall conform to the following compositions: 98% Ar/2% O_2 or 90% Ar/10% CO_2.

Preparation of Base Metal:
The bevels are to be flame cut on the edges of the plate before the parts are assembled. The beveled surface must be smooth and free of notches. Any roughness or notches deeper than 1/64 in. (0.4 mm) must be ground smooth.

All hydrocarbons and other contaminations, such as cutting fluids, grease, oil, and primers, must be cleaned off all parts and filler metals before welding. This cleaning can be done with any suitable solvents or detergents. The groove face and inside and outside plate surface within 1 in. (25 mm) of the joint must be mechanically cleaned of slag, rust, and mill scale. Cleaning must be done with a wire brush or grinder down to bright metal.

Electrical Characteristics:
The current shall be direct-current electrode positive (DCEP). The base metal shall be on the negative side of the line.

Electrode		Welding Power			Shielding Gas		Base Metal	
Type	Size	Amps	Wire-feed Speed, ipm (cm/min)	Volts	Type	Flow	Type	Thickness
E70S-X	0.035 in. (0.9 mm)	180 to 230	400 to 550 (1016 to 1397)	25 to 27	Ar plus 2% O_2 or 90% Ar/ 10% CO_2	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)
E70S-X	0.045 in. (1.2 mm)	260 to 340	300 to 500 (762 to 1270)	25 to 30	Ar plus 2 O_2 or 90% Ar/ 10% CO_2	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)

Preheat:
The parts must be heated to a temperature higher than 50°F (10°C) before any welding is started.

Backing Gas:
N/A

Safety:
Proper protective clothing and equipment must be used. The area must be free of all hazards that may affect the welder or others in the area. The welding machine, welding leads, work clamp, electrode holder, and other equipment must be in safe working order.

Welding Technique:
Using a 3/4-in. (19-mm) or larger gas nozzle for all welding, first tack weld the plates together according to the drawing. There should be about a 1/16-in. (1.6-mm) root gap between the plates with V-grooved or beveled edges. Use the E70S-X arc welding electrodes to make a root pass to fuse the plates together. Clean any silicon slag from the root pass, being sure to remove any trapped silicon slag along the sides of the weld.

Using the E70S-X arc welding electrodes, make a series of stringer or weave filler welds, no thicker than 1/4 in. (6.4 mm), in the groove until the joint is filled. The 1/4-in. (6.4-mm) fillet welds are to be made with one pass.

Interpass Temperature:
The plate should not be heated to a temperature higher than 350°F (175°C) during the welding process. After each weld pass is completed, allow it to cool but never to a temperature below 50°F (10°C). The weldment must not be quenched in water.

Cleaning:
Any slag must be cleaned off between passes. The weld beads may be cleaned by a hand wire brush, a hand chipping, a punch and hammer, or a needle-scaler. All weld cleaning must be performed with the test plate in the welding position.

Inspection:*
Visually inspect the weld for uniformity and discontinuities.

1. There shall be no cracks, no incomplete fusion.
2. There shall be no incomplete joint penetration in groove welds except as permitted for partial joint penetration welds.
3. The Test Supervisor shall examine the weld for acceptable appearance, and shall be satisfied that the welder is skilled in using the process and procedure specified for the text.

*From Table 3, AWS SENSE QC-10:2004 (Courtesy of the American Welding Society)

4. Undercut shall not exceed the lesser of 10% of the base metal thickness or 1/32 in. (0.8 mm)
5. Where visual examination is the only criterion for acceptance, all weld passes are subject to visual examination, at the discretion of the Test Supervisor.
6. The frequency of porosity shall not exceed one in each 4 in. (100 mm) of weld length and the maximum diameter shall not exceed 3/32 in. (2.4 mm).
7. Welds shall be free from overlap.

Sketches:
See Figure 2.50.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 2-27

Gas Metal Arc Welding (GMAW) Spray Transfer Limited Thickness Welder Performance Qualification Test Plate for 1-G, Position, with Backing

Welding Procedure Specification (WPS)

Welding Procedure Specification No.: Practice (2-27). Date:

Title:
Welding GMAW of plate to plate.

Scope:
This procedure is applicable for V-groove and fillet welds within the range of 1/8 in. (3.2 mm) through 1-1/2 in. (38 mm).

Welding may be performed in the following positions: 1G, 2F.

Base Metal:
The base metal shall conform to carbon steel M-1, P-1, and S-1, Group 1 or 2.

Backing material specification: none.

Filler Metal:
The filler metal shall conform to AWS specification no. 0.035 to 0.045 dia. E70S-X from AWS specification A5.18. This filler metal falls into F-number F-6 and A-number A-1.

Shielding Gas:
The shielding gas, or gases, shall conform to the following compositions and purity: 98% Ar/2% O_2 or 90% Ar/10% CO_2

Joint Design and Tolerances:

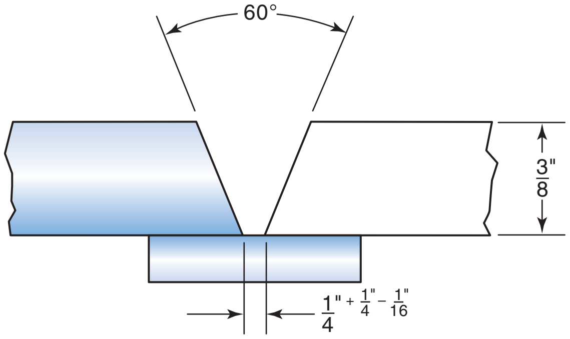

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 3
Key Indicator 1, 2, 3

Module 5
Spray Transfer
Key Indicator 8, 9, 11

Module 9
Key Indicator 1, 2

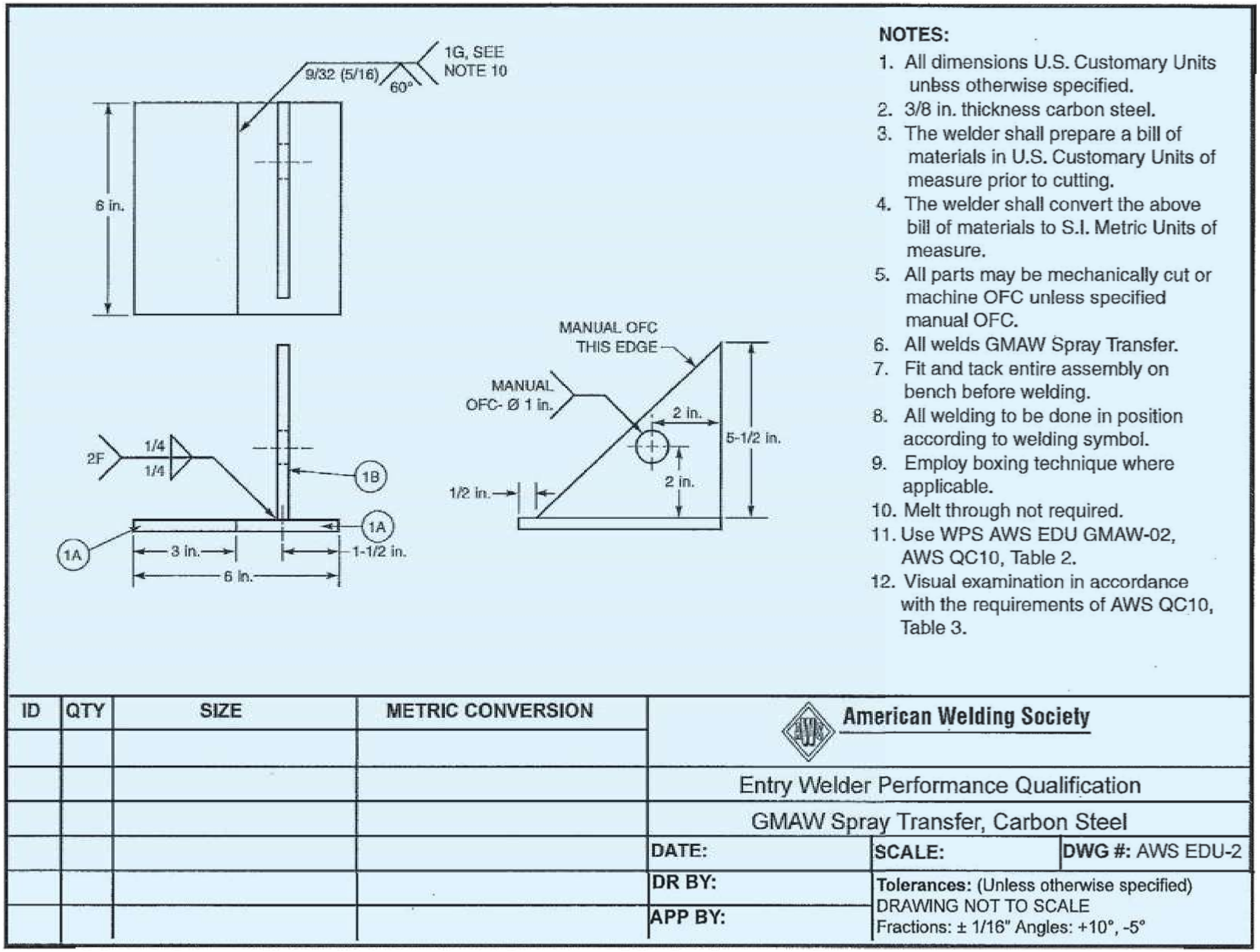
9/32 (5/16)
60°
1G, SEE NOTE 10
6 in.
2F
1/4
1/4
1B
1A
1A
3 in.
6 in.
1-1/2 in.
MANUAL OFC THIS EDGE
MANUAL OFC- Ø 1 in.
2 in.
2 in.
5-1/2 in.
1/2 in.
American Welding Society

NOTES:

1. All dimensions U.S. Customary Units unbss otherwise specified.
2. 3/8 in. thickness carbon steel.
3. The welder shall prepare a bill of materials in U.S. Customary Units of measure prior to cutting.
4. The welder shall convert the above bill of materials to S.I. Metric Units of measure.
5. All parts may be mechanically cut or machine OFC unless specified manual OFC.
6. All welds GMAW Spray Transfer.
7. Fit and tack entire assembly on bench before welding.
8. All welding to be done in position according to welding symbol.
9. Employ boxing technique where applicable.
10. Melt through not required.
11. Use WPS AWS EDU GMAW-02, AWS QC10, Table 2.
12. Visual examination in accordance with the requirements of AWS QC10, Table 3.

ID	QTY	SIZE	METRIC CONVERSION

Entry Welder Performance Qualification

GMAW Spray Transfer, Carbon Steel

DATE:

SCALE:

DWG #: AWS EDU-2

DR BY:

APP BY:

Tolerances: (Unless otherwise specified)
DRAWING NOT TO SCALE
Fractions: ± 1/16" Angles: +10°, -5°

Figure 2.50 **GMAW (spray transfer) workmanship qualification test**

Source: Courtesy of the American Welding Society

Preparation of Base Metal:

The bevels are to be flame cut on the edges of the plate before the parts are assembled. The beveled surface must be smooth and free of notches. Any roughness or notches deeper than 1/64 in. (0.4 mm) must be ground smooth.

All hydrocarbons and other contaminations, such as cutting fluids, grease, oil, and primers, must be cleaned off all parts and filler metals before welding. This cleaning can be done with any suitable solvents or detergents. The groove face and inside and outside plate surface within 1 in. (25 mm) of the joint must be mechanically cleaned of slag, rust, and mill scale. Cleaning must be done with a wire brush or grinder down to bright metal.

Electrical Characteristics:

The current shall be direct-current electrode positive (DCEP). The base metal shall be on the negative side of the line.

Electrode		Welding Power			Shielding Gas		Base Metal	
Type	Size	Amps	Wire-feed Speed ipm (cm/min)	Volts	Type	Flow	Type	Thickness
E70S-X	0.035 in. (0.9 mm)	180 to 230	400 to 550 (1016 to 1397)	25 to 28	Ar plus 2% O_2 or 90% Ar/10% CO_2	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)
E70S-X	0.045 in. (1.2 mm)	260 to 340	300 to 500 (762 to 1270)	25 to 30	Ar plus 2 O_2 or 90% Ar/10% CO_2	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)

Preheat:

The parts must be heated to a temperature higher than 50°F (10°C) before any welding is started.

Backing Gas:

N/A

Safety:

Proper protective clothing and equipment must be used. The area must be free of all hazards that may affect the welder or others in the area. The welding machine, welding leads, work clamp, electrode holder, and other equipment must be in safe working order.

Welding Technique:

Using a 3/4-in. (19-mm) or larger gas nozzle for all welding, first tack weld the plates and backing strip together according to the drawing below. There should be about a 1/4-in. (1.6-mm) root gap between the plates with V-grooved or beveled edges. Use the E70S-X arc welding electrodes to make a root pass to fuse the plates together. Clean any silicon slag from the root pass, being sure to remove any trapped silicon slag along the sides of the weld.

Using the E70S-X arc welding electrodes, make a series of stringer filler welds, no thicker than 1/4 in. (6.4 mm), in the groove until the joint is filled.

Interpass Temperature:

The plate should not be heated to a temperature higher than 350°F (175°C) during the welding process. After each weld pass is completed, allow it to cool but never to a temperature below 50°F (10°C). The weldment must not be quenched in water.

Cleaning:

Any slag must be cleaned off between passes. The weld beads may be cleaned by a hand wire brush, a hand chipping, a punch and hammer, or a needle-scaler. All weld cleaning must be performed with the test plate in the welding position.

Visual Inspection:*

Visually inspect the weld for uniformity and discontinuities.

1. There shall be no cracks, no incomplete fusion.
2. There shall be no incomplete joint penetration in groove welds except as permitted for partial joint penetration welds.
3. The Test Supervisor shall examine the weld for acceptable appearance, and shall be satisfied that the welder is skilled in using the process and procedure specified for the text.
4. Undercut shall not exceed the lesser of 10% of the base metal thickness or 1/32 in. (0.8 mm).
5. Where visual examination is the only criterion for acceptance, all weld passes are subject to visual examination, at the discretion of the Test Supervisor.
6. The frequency of porosity shall not exceed one in each 4 in. (100 mm) of weld length and the maximum diameter shall not exceed 3/32 in. (2.4 mm).
7. Welds shall be free from overlap.

Sketches:

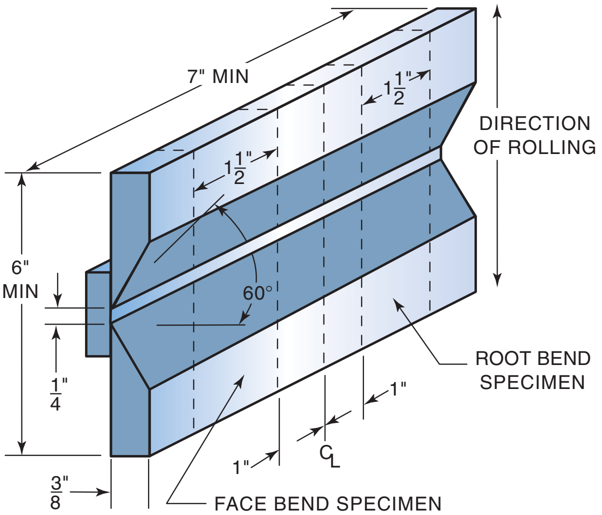

Bend-Test:

The weld is to be mechanically tested only after it has passed the visual inspection. Be sure that the test specimens are properly marked to identify the welder, the position, and the process.

Specimen Preparation

For 3/8-in. test plates, two specimens are to be located in accordance with the requirements below. One is to be prepared for a transverse face bend, and the other is to be prepared for a transverse root bend.

*From Table 3, AWS SENSE QC-10:2004 (Courtesy of the American Welding Society)

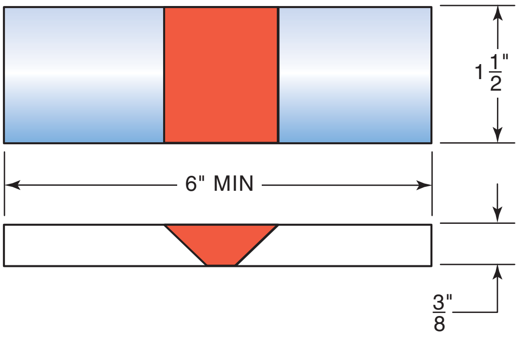

Transverse face bend. The weld is perpendicular to the longitudinal axis of the specimen and is bent so that the weld face becomes the tension surface of the specimen. Transverse face bend specimens shall comply with the requirements below.

Transverse root bend. The weld is perpendicular to the longitudinal axis of the specimen and is bent so that the weld root becomes the tension surface of the specimen. Transverse root bend specimens shall comply with the requirements below.

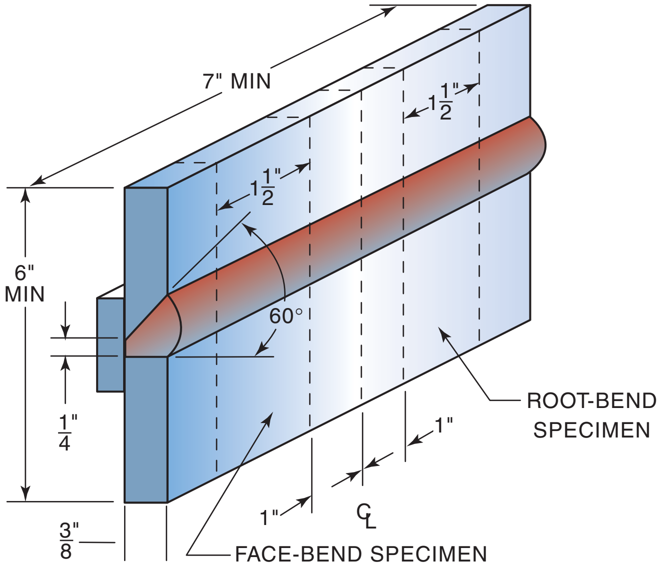

Acceptance Criteria for Face and Root Bends*

For acceptance, the convex surface of the face and root bend specimens shall meet both of the following requirements:

1. No single indication shall exceed 1/8 in. (3.2 mm), measured in any direction on the surface.
2. The sum of the greatest dimensions of all indications on the surface, which exceed 1/32 in. (0.8 mm), but are less than or equal to 1/8 in. (3.2 mm), shall not exceed 3/8 in. (9.6 mm).

Cracks occurring at the corner of the specimens shall not be considered unless there is definite evidence that they result from slag or inclusions or other internal discontinuities.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

*From Table 4, AWS SENSE QC-10:2004 (Courtesy of the American Welding Society)

SUMMARY

Slight changes in welding gun angle and electrode extension can make significant differences in the quality of the weld produced. As a new welder you might find it difficult to tell the effect of these changes if they are slight. Therefore, as you start to learn this process it is a good idea to make more radical changes so it is easier for you to see their effects on the weld. Later as you develop your skills you can use these slight changes to aid in controlling the weld's quality and appearance as it progresses along the joint. Small adjustments in your welding technique are required to compensate for slight changes that occur along a welding joint, such as joint gap and the increasing temperature of the base metal.

Variations in conditions can significantly affect welding setup for the GMA process. Before starting an actual weld in the field you should practice to test your setup. Practice on scrap metal of a similar thickness and type of metal to be welded. A practice weld before you begin welding can significantly increase the chances that your weld will meet standards or specifications. Making these sample or test welds is more important when you are welding in the field, since welds outside the shop are more difficult to control and anticipate. Think of this much as an athlete warms up before competing.

You will find it beneficial when you are initially setting up your welder to have someone assist you, so that he or she can make changes in the welding machine's settings as you are welding. This teamwork can significantly increase your setup accuracy and reduce setup time. Later on in the field, having developed a keen eye for watching the weld, you can then make these adjustments for yourself more rapidly and accurately. Working with another student in a group effort like this will also give you a better understanding of how other individuals' setup preferences affect their welds. Welding is an art, and therefore each welder may have slight differences in preference for voltage, amperage, gas flow, and other setup variables. This gives you an opportunity to learn more from others.

REVIEW

1. What items make up a basic semiautomatic welding system?
2. What must be done to the shielding gas cylinder before the valve protection cap is removed?
3. Why is the shielding gas valve "cracked" before the flowmeter regulator is attached?
4. What causes the electrode to bird-nest?
5. Why must all fittings and connections be tight?
6. What parts should be activated by depressing the gun switch?
7. What benefit does a welding wire's cast provide?
8. What can be done to determine the location of a problem that stops the wire from being successfully fed through the conduit?
9. What are the advantages of using a feed roller pressure that is as light as possible?

10. Why should the feed roller drag prevent the spool from coasting to a stop when the feed stops?
11. Why must you always wind the wire tightly into a ball or cut it into short lengths before discarding it in the proper waste container?
12. Why would the flowmeter ball float at different heights with different shielding gases if the shielding gases are flowing at the same rate?
13. Using Table 2.1, determine the amperage if 400 in. (10.2 m) of 0.45-in. (1.2-mm) steel wire is fed in one minute.
14. How is the amperage adjusted on a GMA welder?
15. What happens to the weld as the electrode extension is lengthened?
16. What is the effect on the weld of changing the welding angle from a dragging to a pushing angle?
17. What are the advantages of adding oxygen or CO_2 to argon for welds on steel?
18. What are the advantages of using CO_2 for making GMA welds on steel?
19. What is mill scale?
20. What type of porosity is most often caused by mill scale?
21. What should the welder watch if the view of the weld is obstructed by the shielding gas nozzle?
22. When you are making a vertical weld and it appears that the weld metal is going to drip over the shelf, what should you do?
23. What are the advantages of making vertical down welds?
24. How can small weld beads be maintained during overhead welds?
25. How can spatter be controlled on the nozzle when making overhead welds?
26. How should the electrode be manipulated for the deepest penetration when using the pulsed-arc metal transfer process?

This page intentionally left blank

CHAPTER

3

Flux Cored Arc Welding Equipment, Setup, and Operation

OBJECTIVES

After completing this chapter, the student should be able to

- describe the flux cored arc (FCA) welding process
- list the equipment required for an FCA welding workstation
- list five advantages of FCA welding, and explain four of its limitations
- tell how electrodes are manufactured and explain the purpose of the electrode cast and helix
- list four things flux can provide to the weld and how fluxes are classified
- explain what each of the digits in a standard FCAW electrode identification number mean
- describe the proper care and handling of FCAW electrodes
- list two common shielding gases used in FCAW, and contrast their benefits related to cost, productivity, and quality
- list three differences in an FCA weld when the gun angle is changed
- identify the two modes of metal transfer and contrast them in regard to application and quality
- list four effects that electrode extension has on FCA welding
- list three things that can cause weld porosity and how it can be prevented

KEY TERMS

air-cooled
coils
deoxidizers
dual shield
flux cored arc welding (FCAW)
lime-based flux
rutile-based flux
self-shielding
slag
smoke extraction nozzles
spools
water-cooled

AWS SENSE EG2.0

Key Indicators Addressed in this Chapter:

Module 6: Flux Cored Arc Welding (FCAW-G, FCAW-S)

Key Indicator 1: Performs safety inspections of FCAW-G/GM, FCAW-S equipment and accessories

Key Indicator 2: Makes minor external repairs to FCAW-G/GM, FCAW-S equipment and accessories

Key Indicator 3: Sets up for FCAW-G/GM, FCAW-S operations on carbon steel

INTRODUCTION

Flux cored arc welding (FCAW) is a fusion welding process in which weld heating is produced from an arc between the work and a continuously fed filler metal electrode. Atmospheric shielding is provided completely or in part by the flux sealed within the tubular electrode, Figure 3.1. Extra shielding may or may not be supplied through a nozzle in the same way as in GMAW.

Although the process was introduced in the early 1950s, it represented less than 5% of the total amount of welding done in 1965. In 2005, it passed the 50% mark and is still rising. The rapid rise in the use of FCAW has been due to a number of factors. Improvements in the fluxes, smaller electrode diameters, increased reliability of the equipment, better electrode feed systems, and improved guns have all led to the increased usage. Guns equipped with **smoke extraction nozzles** and electronic controls are the latest in a long line of improvements to this process, Figure 3.2.

PRINCIPLES OF OPERATION

FCA welding is similar in a number of ways to the operation of GMA welding, Figure 3.3. Both processes use a constant-potential (CP) or constant-voltage (CV) power supply. Constant potential and voltage are terms that have the same meaning. CP power supplies provide a controlled voltage (potential) to the welding electrode. The amperage (current) varies with the speed that the electrode is being fed into the molten weld pool. Just as in GMA welding, higher electrode feed speeds produce higher currents and slower feed speeds result in lower currents, assuming all other conditions remain constant.

The effects on the weld of electrode extension, gun angle, welding direction, travel speed, and other welder manipulations are similar to those experienced in GMA welding. As in GMA welding, having a correctly set welder does not ensure a good weld. The skill of the welder is an important factor in producing high-quality welds.

The flux inside the electrode protects the molten weld pool from the atmosphere, improves strength through chemical reactions and alloys, and improves the weld shape.

Atmospheric contamination of molten weld metal occurs as it travels across the arc gap and within the pool before it solidifies. The major atmospheric contaminations come from oxygen and nitrogen, the major elements in air. The addition of fluxing and gas-forming elements to the core electrode reduces or eliminates their effects.

Improved strength and other physical or corrosion-resistant properties of the finished weld are improved by the flux. Small additions of alloying elements, deoxidizers, and gas-forming and slag agents all can improve the desired weld properties. Carbon, chromium, and vanadium can be added to improve hardness, strength, creep resistance, and corrosion resistance. Aluminum, silicon, and titanium all help remove oxides and/or nitrides in the weld. Potassium, sodium, and zirconium are added to the flux and form a slag.

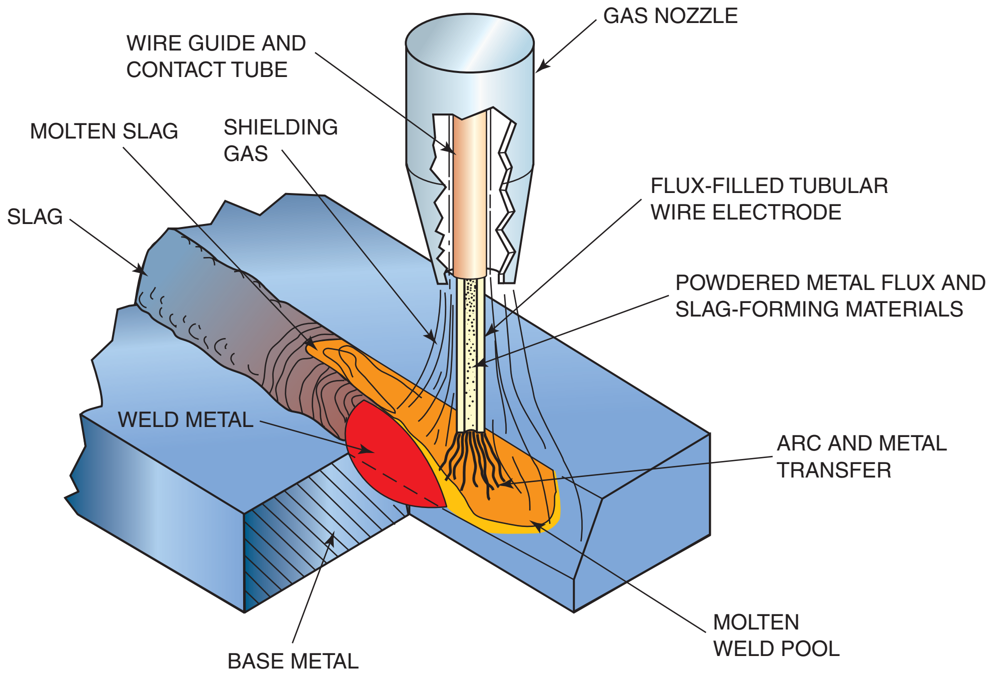
GAS NOZZLE
WIRE GUIDE AND CONTACT TUBE
MOLTEN SLAG
SHIELDING GAS
SLAG
FLUX-FILLED TUBULAR WIRE ELECTRODE
POWDERED METAL FLUX AND SLAG-FORMING MATERIALS
WELD METAL
ARC AND METAL TRANSFER
MOLTEN WELD POOL
BASE METAL

(A) GAS-SHIELDED FLUX CORED ARC WELDING (FCAW-G)

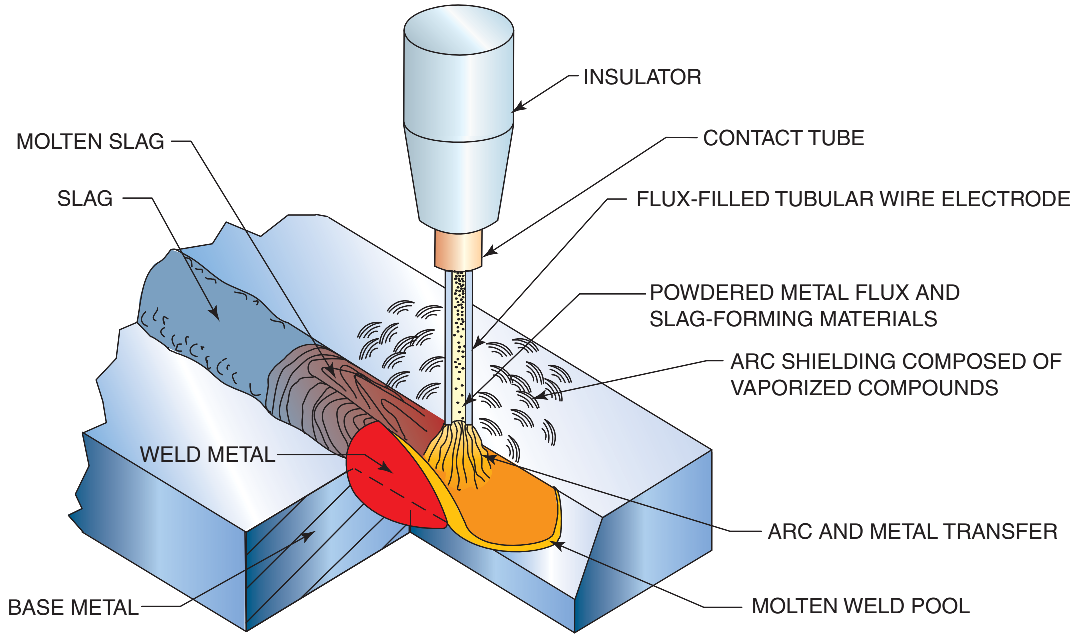
INSULATOR
MOLTEN SLAG
CONTACT TUBE
SLAG
FLUX-FILLED TUBULAR WIRE ELECTRODE
POWDERED METAL FLUX AND SLAG-FORMING MATERIALS
ARC SHIELDING COMPOSED OF VAPORIZED COMPOUNDS
WELD METAL
ARC AND METAL TRANSFER
BASE METAL
MOLTEN WELD POOL

(B) SELF-SHIELDED FLUX CORED ARC WELDING (FCAW-S)

Figure 3.1 **Two types of flux cored arc welding**
FCA welding may have extra shielding provided by a gas nozzle (A), or be self-shielding only (B).
Source: Courtesy of the American Welding Society

A discussion of weld metal additives and flux elements and their effects on the weld can be found later in this chapter.

The flux core additives that serve as deoxidizers, gas formers, and slag formers either protect the molten weld pool or help to remove impurities from the base metal. Deoxidizers may convert small amounts of surface

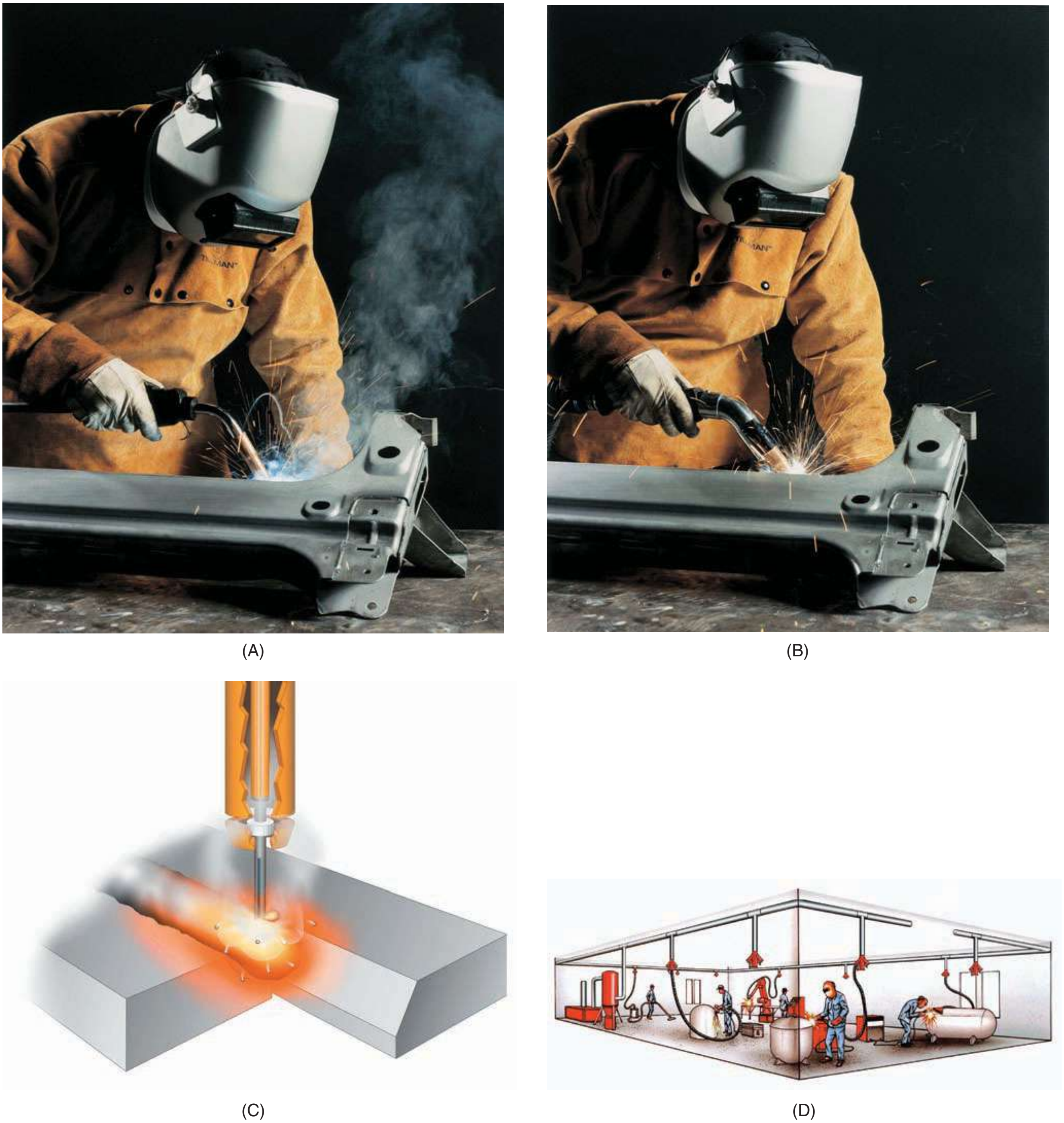

Figure 3.2 Smoke extraction
(A) FCA welding without smoke extraction and (B) with smoke extraction. (C) Typical FCAW smoke extraction gun. (D) Typical smoke exhaust system.
Source: Courtesy of Lincoln Electric Company

oxides like mill scale back into pure metal. They work much like the elements used to refine iron ore into steel.

Gas formers rapidly expand and push the surrounding air away from the molten weld pool. If oxygen in the air were to come in contact with the molten weld metal, the weld metal would quickly oxidize. Sometimes this can be seen at the end of a weld when the molten weld metal erupts in a shower of tiny sparks.

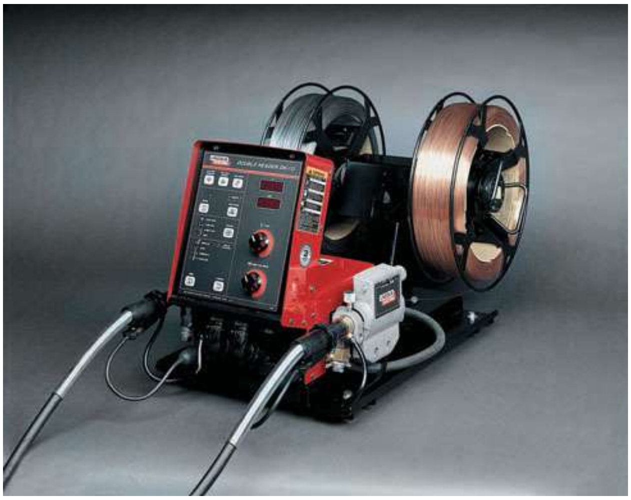

Figure 3.3 **Large-capacity wire-feed unit used with FCAW or GMAW**
Source: Courtesy of Lincoln Electric Company

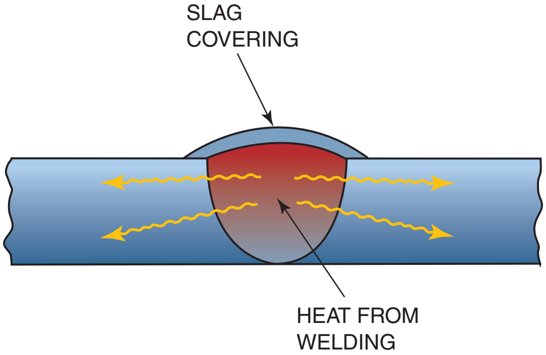

Figure 3.4 **Slag blanketing the weld**
The slag covering keeps the welding heat from escaping quickly, thus slowing the cooling rate.

The slag covering of the weld is useful for several reasons. **Slag** is a nonmetallic product resulting from the mutual dissolution of the flux and nonmetallic impurities in the base metal. Slag helps the weld by protecting the hot metal from the effects of the atmosphere, controlling the bead shape by serving as a dam or mold, and serving as a blanket to slow the weld's cooling rate, which improves its physical properties, Figure 3.4.

EQUIPMENT

Module 6
Key Indicator 1, 2, 3

Power Supply

The FCA welding power supply is the same type that is required for GMAW, called constant-potential, constant-voltage (CP, CV). The words *potential* and *voltage* have the same electrical meaning and are used interchangeably. FCAW machines can be much more powerful than GMAW machines and are available with up to 1500 amperes of welding power.

Guns

FCA welding guns are available as **water-cooled** or **air-cooled**, Figure 3.5. Although most of the FCA welding guns that you will find in schools are air-cooled, our industry often needs water-cooled guns because of the

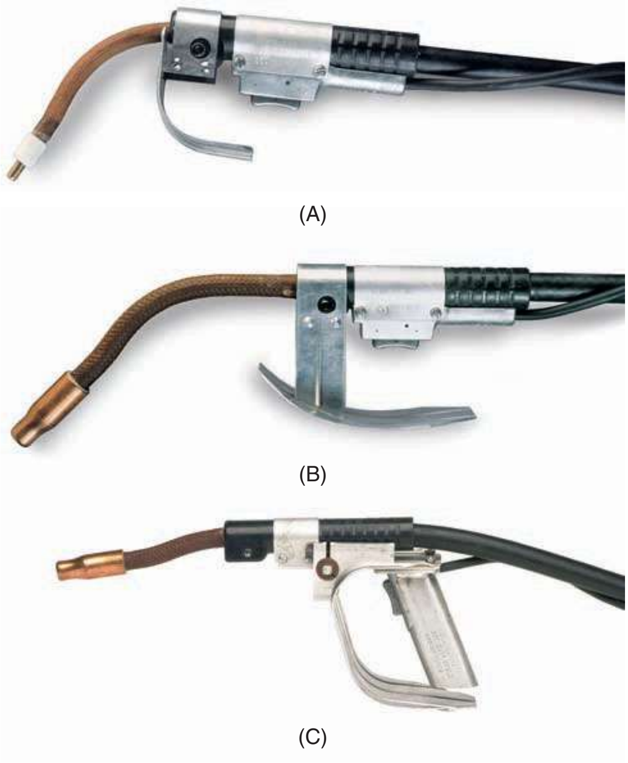

Figure 3.5 **Typical FCA welding guns**
(A) 350 ampere rating self-shielding, (B) 450 ampere rating gas-shielding, and (C) 600 ampere rating gas-shielding.
Source: Courtesy of Lincoln Electric Company

higher heat caused by longer welds made at higher currents. The water-cooled FCA welding gun is more efficient than an air-cooled gun at removing waste heat. The air-cooled gun is more portable because it has fewer hoses, and it may be made lighter so it is easier to manipulate than the water-cooled gun.

Also, the water-cooled gun requires a water reservoir or another system to give the needed cooling. There are two major ways that water can be supplied to the gun for cooling. Cooling water can be supplied directly from the building's water system, or it can be supplied from a recirculation system.

Cooling water supplied directly from the building's water system is usually dumped into a wastewater drain once it has passed through the gun. When this type of system is used, a pressure regulator must be installed to prevent pressures that are too high from damaging the hoses. Water pressures higher than 35 psi (241 kg/mm^2) may cause the water hoses to burst. Check valves must also be installed in the supply line to prevent contaminated water from being drawn back into the water supply. Some cities and states have laws that restrict the use of open systems because of the need for water conservation. Check with your city or state for any restrictions before installing an open water-cooling system.

Recirculating cooling water systems eliminate any of the problems associated with open systems. Chemicals may be added to the water in recirculating systems to prevent freezing, to aid in pump lubrication, and to prevent algae growth. Only manufacturer-approved additives should be used in a recirculation system. Read all of the manufacturer's safety and data sheets before using these chemicals.

Smoke Extraction Nozzles

Because of the large quantity of smoke that can be generated during FCA welding, systems for smoke extraction that fit on the gun have been designed, Figure 3.2B. These systems use a vacuum to pull the smoke back into a specially designed smoke extraction nozzle on the welding gun. The disadvantage of this slightly heavier gun is offset by the system's advantages. The advantages of the system are as follows:

- Cleaner air for the welder to breathe because the smoke is removed before it rises to the welder's face.
- Reduced heating and cooling cost because the smoke is concentrated, so less shop air must be removed with the smoke.

Electrode Feed

Electrode feed systems are similar to those used for GMAW; in fact many feed systems are designed with dual feeders so that solid wire and flux core may be run in sequence. The major difference is that the more robust FCAW feeders are designed to use large-diameter wire and most often have two sets of feed rollers. The two sets of rollers help reduce the drive pressure on the electrode. Excessive pressure can distort the electrode wire diameter, which can allow some flux to be dropped inside the electrode guide tube.

ADVANTAGES

FCA welding offers the welding industry a number of important advantages.

High Deposition Rate

High rates of depositing weld metal are possible. FCA welding deposition rates of more than 25 lb/hr (11 kg/hr) of weld metal are possible. This compares to about 10 lb/hr (5 kg/hr) for shielded metal arc (SMA) welding using a very large-diameter electrode of 1/4 in. (6 mm).

Minimum Electrode Waste

The FCA method makes efficient use of filler metal; from 75% to 90% of the weight of the FCAW electrode is metal, the remainder being flux. SMAW electrodes have a maximum of 75% filler metal; some SMAW electrodes have much less. Also, a stub must be left at the end of each SMA welding electrode. The stub will average 2 in. (51 mm) in length, resulting in a loss of 11% or more of the SMAW filler electrode purchased. FCA welding has no stub loss, so nearly 100% of the FCAW electrode purchased is used.

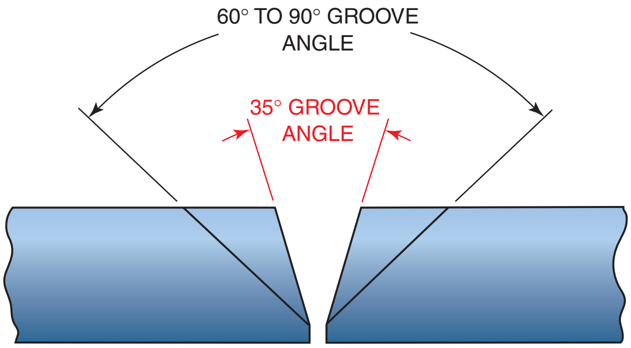

Figure 3.6 **Narrower groove angle for FCAW**
The narrower groove angle for FCAW compared to other welding processes saves on filler metal, welding time, and heat input into the part.

Narrow Groove Angle

Because of the deep penetration characteristic of FCAW, no edge-beveling preparation is required on some joints in metal up to 1/2 in. (13 mm) in thickness. When bevels are cut, the joint-included angle can be reduced to as small as 35°, Figure 3.6. The reduced groove angle results in a smaller-sized weld. This can save 50% of filler metal with about the same savings in time and weld power used.

Minimum Precleaning

The addition of **deoxidizers**, which combine with and remove harmful oxides on the base metal or its surface, and other fluxing agents permits high-quality welds to be made on plates with light surface oxides and mill scale. This eliminates most of the precleaning required before GMA welding can be performed. Often it is possible to make excellent welds on plates in the "as cut" condition; no cleanup is needed.

All-position Welding

Small-diameter electrode sizes in combination with special fluxes allow excellent welds in all positions. The slags produced assist in supporting the weld metal. This process is easy to use, and, when properly adjusted, it is much easier to use than other all-position arc welding processes.

Flexibility

Changes in power settings can permit welding to be done on thin-gauge sheet metals or thicker plates using the same electrode size. Multipass welds allow joining metals with no limit on thickness. This, too, is attainable with one size of electrode.

High Quality

Many codes permit welds to be made using FCAW. The addition of the flux gives the process the high level of reliability needed for welding on boilers, pressure vessels, and structural steel.

Excellent Control

The molten weld pool is more easily controlled with FCAW than with GMAW. The surface appearance is smooth and uniform even with less operator skill. Visibility is improved by removing the nozzle when using self-shielded electrodes.

LIMITATIONS

The main limitation of flux cored arc welding is that it is confined to ferrous metals and nickel-based alloys. Generally, all low- and medium-carbon steels and some low-alloy steels, cast irons, and a limited number of stainless steels are presently weldable using FCAW.

The equipment and electrodes used for the FCAW process are more expensive. However, the cost is quickly recoverable through higher productivity.

The removal of postweld slag requires another production step. The flux must be removed before the weldment is finished (painted) to prevent crevice corrosion.

With the increased welding output comes an increase in smoke and fume generation. The existing ventilation system in a shop might need to be increased to handle the added volume.

ELECTRODES

Methods of Manufacturing

The electrodes have flux tightly packed inside. One method used to make them is to first form a thin sheet of metal into a U-shape, Figure 3.7. A measured quantity of flux is poured into the U-shape before it is squeezed shut. The wire is then passed through a series of dies to size it and further compact the flux.

A second method of manufacturing the electrode is to start with a seamless tube. The tube is usually about 1 in. in diameter. One end of the tube is sealed, and the flux powder is poured into the open end. The tube is vibrated during the filling process to ensure that it fills completely. Once the tube is full, the open end is sealed. The tube is now sized using a series of dies, Figure 3.8.

In both these methods of manufacturing the electrode, the sheet and tube are made up of the desired alloy. Also in both cases, the flux is compacted inside the metal skin. This compacting helps make the electrode operate more smoothly and consistently.

Electrodes are available in sizes from 0.030 in. to 5/32 in. (0.8 mm to 3.9 mm) in diameter. Smaller-diameter electrodes are much more expensive per pound than the same type in a larger diameter due to the high cost of drawing and filling cored wires to small sizes. Larger-diameter electrodes produce such large welds they cannot be controlled in all positions. The most popular diameters range from 0.035 in. to 3/32 in. (0.9 mm to 2.3 mm).

The finished FCA filler metal is packaged in a number of forms for purchase by the end user, Figure 3.9. The AWS has a standard for the size of

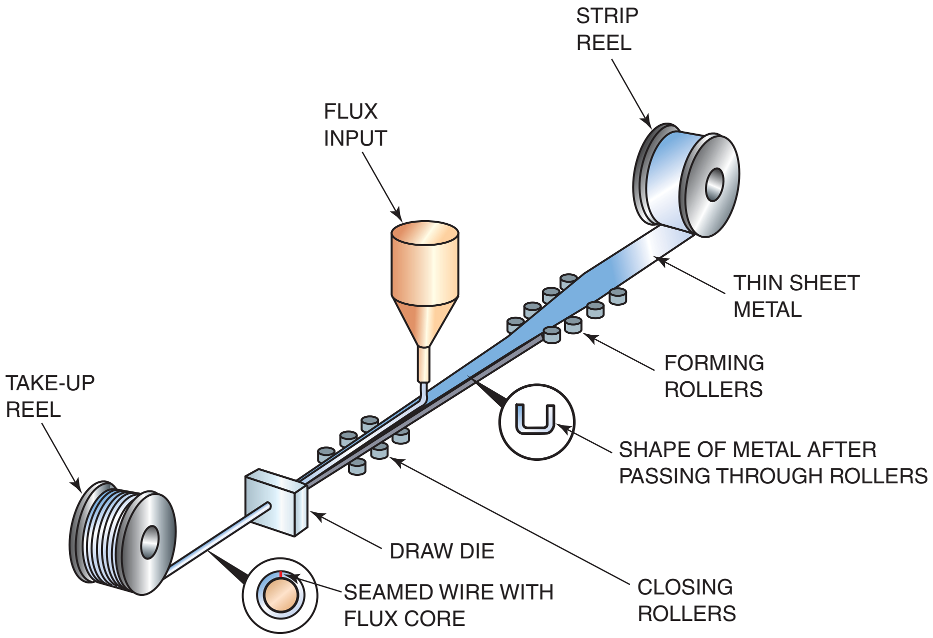
STRIP REEL
FLUX INPUT
THIN SHEET METAL
FORMING ROLLERS
TAKE-UP REEL
SHAPE OF METAL AFTER PASSING THROUGH ROLLERS
DRAW DIE
SEAMED WIRE WITH FLUX CORE
CLOSING ROLLERS

Figure 3.7 **Putting the flux in the flux cored wire**

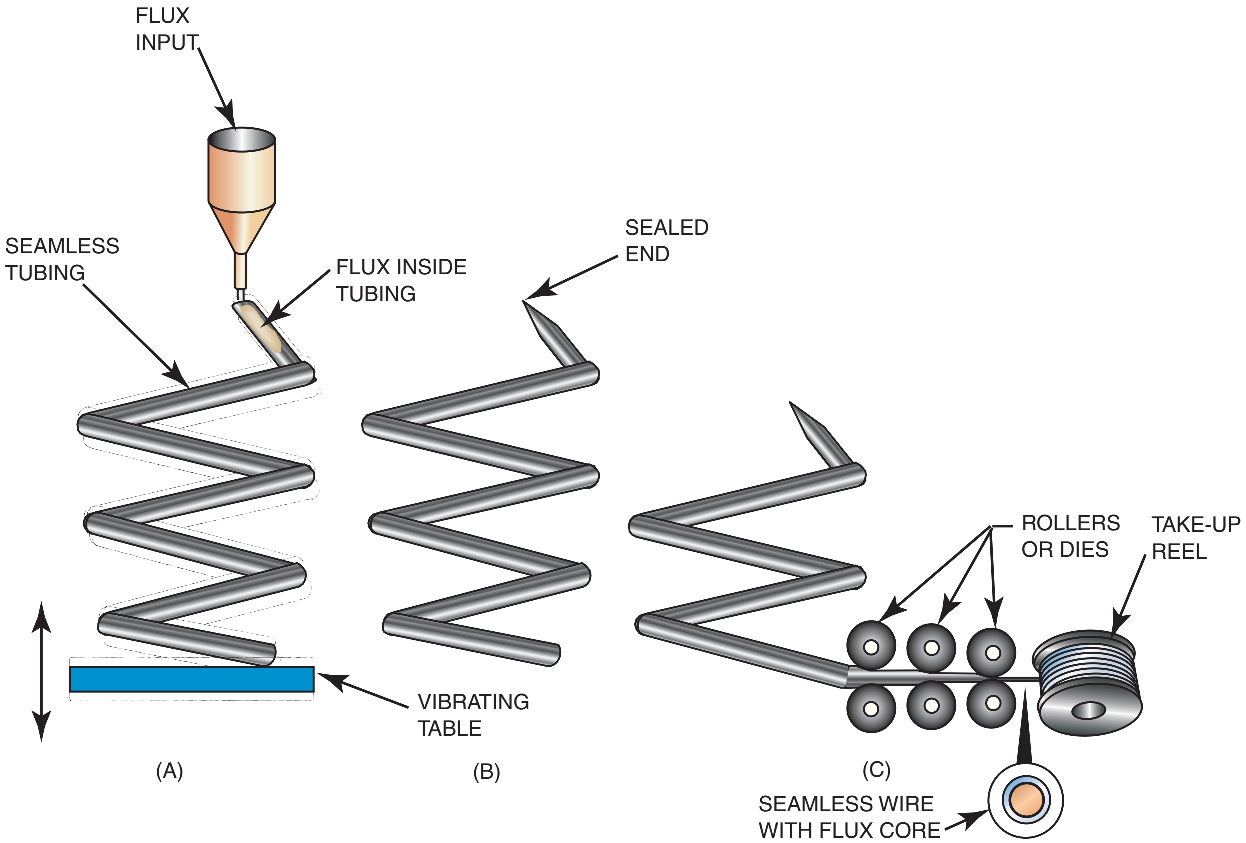
FLUX INPUT
SEAMLESS TUBING
FLUX INSIDE TUBING
SEALED END
ROLLERS OR DIES
TAKE-UP REEL
VIBRATING TABLE
(A)
(B)
(C)
SEAMLESS WIRE WITH FLUX CORE

Figure 3.8 **One method of filling seamless FCA welding filler metal with flux**
The vibration helps compact the granular flux inside the tube.

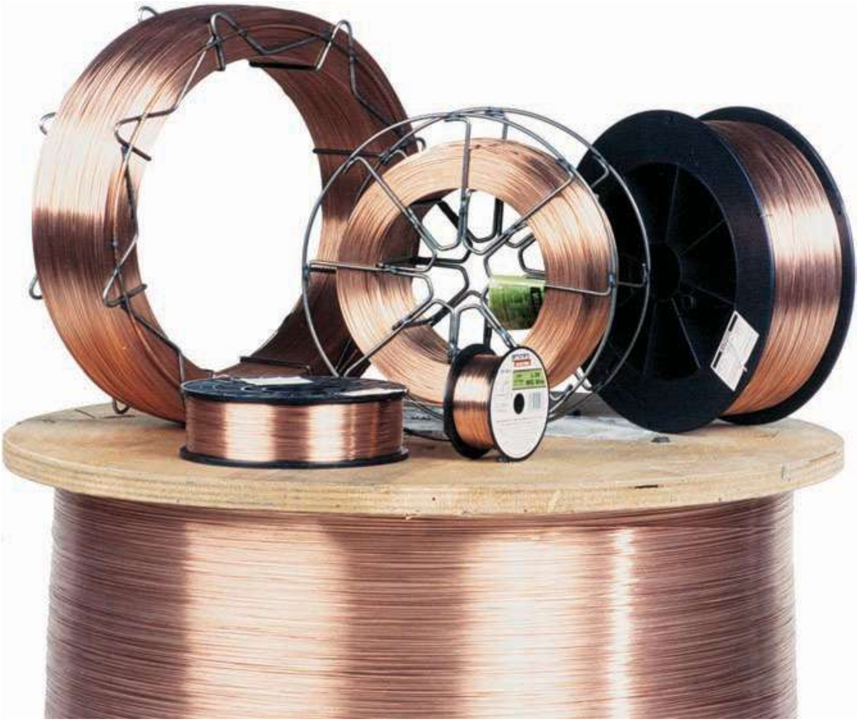

Figure 3.9 **A few packaged forms of FCA filler metal**
FCAW filler metal weights are approximate. They will vary by alloy and manufacturer.
Source: Courtesy of Lincoln Electric Company

each of the package units. Although the dimensions of the packages are standard, the weight of filler wire is not standard. More of the smaller-diameter wire can fit into the same space compared with a larger-diameter wire, so a package of 0.030-in. (0.8-mm) wire weighs more than the same-sized package of 3/32-in. (2.3-mm) wire. The standard packing units for FCAW wires are **spools**, **coils**, reels, and drums, Table 3.1.

Spools are made of plastic or fiberboard and are disposable. They are completely self-contained and are available in approximate weights from

Table 3.1 **Packaging Size Specification for Commonly Used FCA Filler Wire**

Packaging	Outside Diameter	Width	Arbor (Hole) Diameter
Spools	4 in. (102 mm)	1-3/4 in. (44.5 mm)	5/8 in. (16 mm)
	8 in. (203 mm)	2-1/4 in. (57 mm)	2-1/16 in. (52.3 mm)
	12 in. (305 mm)	4 in. (102 mm)	2-1/16 in. (52.3 mm)
	14 in. (356 mm)	4 in. (102 mm)	2-1/16 in. (52.3 mm)
Reels	22 in. (559 mm)	12-1/2 in. (318 mm)	1-5/16 in. (33.3 mm)
	30 in. (762 mm)	16 in. (406 mm)	1-5/16 in. (33.3 mm)
Coils	16-1/4 in. (413 mm)	4 in. (102 mm)	12 in. (305 mm)

	Outside Diameter	Inside Diameter	Height
Drums	23 in. (584 mm)	16 in. (406 mm)	34 in. (864 mm)

1 lb up to around 50 lb (0.5 kg to 25 kg). The smaller spools, 4 in. and 8 in. (102-mm and 203 mm), weighing from 1 lb to 7 lb, are most often used for smaller production runs or for home/hobby use; 12-in. and 14-in. (305-mm and 356-mm) spools are often used in schools and welding fabrication shops.

Coils come wrapped and/or wire tied together. They are unmounted, so they must be supported on a frame on the wire feeder in order to be used. Coils are available in weights around 60 lb (27 kg). Because FCAW wires on coils do not have the expense of a disposable core, these wires cost a little less per pound, so they are more desirable for higher-production shops.

Reels are large wooden spools, and drums are shaped like barrels. Both reels and drums are used for high-production jobs. Both can contain approximately 300 lb to 1000 lb (136 kg to 454 kg) of FCAW wire. Because of their size, they are used primarily at fixed welding stations. Such stations are often associated with some form of automation, such as turntables or robotics.

Electrode Cast and Helix

To see the cast and helix of a wire, feed out 10 ft of wire electrode and cut it off. Lay it on the floor and observe that it forms a circle. The diameter of the circle is known as the cast of the wire, Figure 3.10.

Note that the wire electrode does not lay flat. One end is slightly higher than the other. This height is the helix of the wire.

The AWS has specifications for both cast and helix for all FCA welding wires.

The cast and helix cause the wire to rub on the inside of the contact tube, Figure 3.11. The slight bend in the electrode wire ensures a positive electrical contact between the contact tube and filler wire.

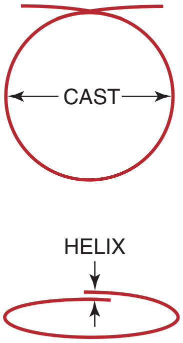

Figure 3.10 Method of measuring cast and helix of FCAW filler wire

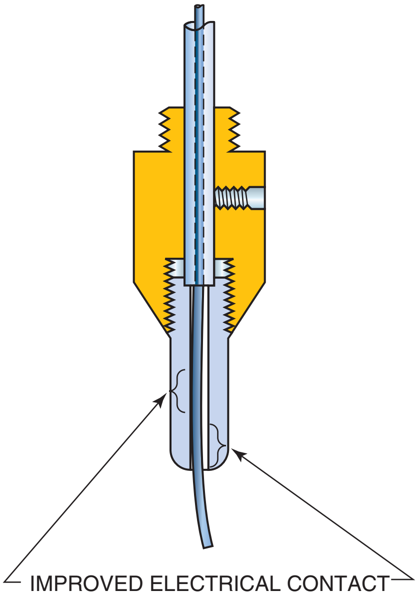

Figure 3.11 Cast forces the wire to make better electrical contact with the tube

FLUX

The fluxes used are mainly based on lime or rutile (a mineral compound consisting of titanium dioxide, usually with a little iron). The purpose of the fluxes is the same as in the shielded metal arc welding (SMAW) process. That is, they can provide all or part of the following to the weld:

- *Deoxidizers:* Oxygen that is present in the welding zone has two forms. It can exist as free oxygen from the atmosphere surrounding the weld. Oxygen can also exist as part of a compound such as an iron oxide or carbon dioxide (CO_2). In either case it can cause porosity in the weld if it is not removed or controlled. Chemicals are added that react to the presence of oxygen in either form and combine to form a harmless compound, Table 3.2. The new compound can become part of the slag that solidifies on top of the weld, or some of it may stay in the weld as very small inclusions. Both methods result in a weld with better mechanical properties because of lower porosity.
- *Slag formers:* Slag serves several vital functions for the weld. It can react with the molten weld metal chemically, and it can affect the weld bead physically. In the molten state it moves through the molten weld pool and acts as a magnet or sponge to chemically combine with impurities in the metal and remove them, Figure 3.12. Slags can be refractory, so that they become solid at a high temperature. As they solidify over the weld, they help it hold its shape and they slow its cooling rate.
- *Fluxing agents:* Molten weld metal tends to have a high surface tension, which prevents it from flowing outward toward the edges of the weld. This causes undercutting along the junction of the

Table 3.2 **Deoxidizing Elements Added to Filler Wire (to Minimize Porosity in the Molten Weld Pool)**

Deoxidizing Element	Strength
Aluminum (Al)	Very strong
Manganese (Mn)	Weak
Silicon (Si)	Weak
Titanium (Ti)	Very strong
Zirconium (Zr)	Very strong

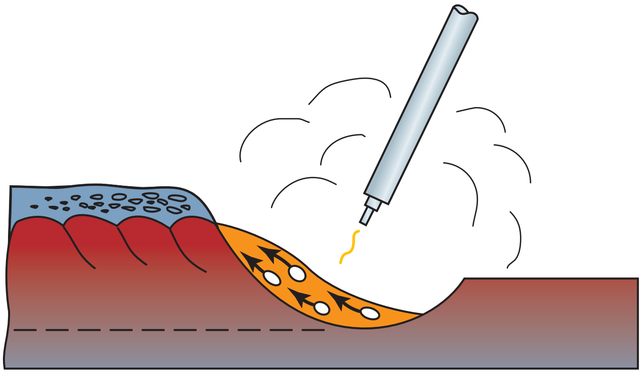

Figure 3.12 **Impurities being floated to the surface by slag**

weld and the base metal. Fluxing agents make the weld more fluid and allow it to flow outward, filling the undercut.
- *Arc stabilizers:* Chemicals in the flux affect the arc resistance. As the resistance is lowered, the arc voltage drops and penetration is reduced. When the arc resistance is increased, the arc voltage increases and weld penetration is increased. Although the resistance within the ionized arc stream may change, the arc is more stable and easier to control. It also improves the metal transfer by reducing spatter caused by an erratic arc.
- *Alloying elements:* Because of the difference in the mechanical properties of metal that is formed by rolling or forging and metal that is melted to form a weld bead, the metallurgical requirements of the two also differ. Some elements change the weld's strength, ductility, hardness, brittleness, toughness, and corrosion resistance. Other alloying elements in the form of powder metal can be added to increase deposition rates.
- *Shielding gas:* As elements in the flux are heated by the arc, some of them vaporize and form voluminous gaseous clouds hundreds of times larger than their original volume. This rapidly expanding cloud forces the air around the weld zone away from the molten weld metal, Figure 3.13. Without the protection this process affords the molten metal, it would rapidly oxidize. Such oxidization would severely affect the weld's mechanical properties, rendering it unfit for service.

All FCAW fluxes are divided into two groups based on the acid or basic chemical reactivity of the slag. The AWS classifies T-1 as acid and T-5 as basic.

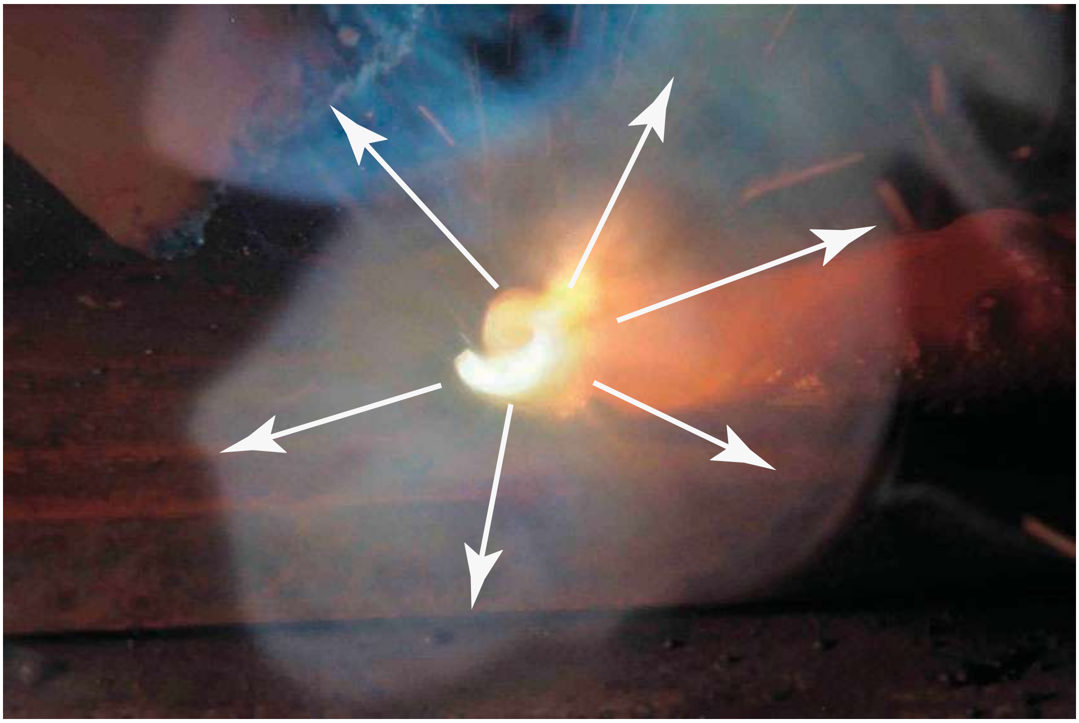

Figure 3.13 **Rapidly expanding gas cloud**
Source: Courtesy of Larry Jeffus

The **rutile-based flux** is acidic, T-1. It produces a smooth, stable arc and a refractory high-temperature slag for out-of-position welding. These electrodes produce a fine drop transfer, a relatively low fume, and an easily removed slag. The main limitation of the rutile fluxes is that their fluxing elements do not produce as high a quality deposit as do the T-5 systems.

The **lime-based flux** is basic, T-5. It is very good at removing certain impurities from the weld metal, but its low-melting-temperature slag is fluid, which makes it generally unsuitable for out-of-position welding. These electrodes produce a more globular transfer, more spatter, more fume, and a more adherent slag than do the T-1 systems. These characteristics are tolerated when it is necessary to deposit very tough weld metal and for welding materials having a low tolerance for hydrogen.

Some rutile-based electrodes allow the addition of a shielding gas. With the weld partially protected by the shielding gas, more elements can be added to the flux, which produces welds with the best of both flux systems, high-quality welds in all positions.

Some fluxes can be used on both single- and multiple-pass welds, and others are limited to single-pass welds only. Using a single-pass welding electrode for multipass welds may result in an excessive amount of manganese. The manganese is necessary to retain strength when making large, single-pass welds. However, with the lower dilution associated with multipass techniques, it can strengthen the weld metal too much and reduce its ductility. In some cases, small welds that deeply penetrate the base metal can help control this problem.

Table 3.3 lists the shielding and polarity for the flux classifications of mild steel FCAW electrodes. The letter *G* is used to indicate an unspecified classification. The *G* means that the electrode has not been classified by the American Welding Society. Often the exact composition of fluxes is kept as a manufacturer's trade secret. Therefore, only limited information about the electrode's composition will be given. The only information often supplied is current, type of shielding required, and some strength characteristics.

Table 3.3 **Welding Characteristics of Seven Flux Classifications**

Classification	Comments	Shielding Gas
T-1	Requires clean surfaces and produces little spatter. It can be used for single- and multiple-pass welds in all positions.	Carbon dioxide (CO_2) or argon/carbon dioxide mixes
T-2	Requires clean surfaces and produces little spatter. It can be used for single-pass welds in the flat (1G and 1F) and horizontal (2F) positions only.	Carbon dioxide (CO_2)
T-3	Used on thin-gauge steel for single-pass welds in the flat (1G and 1F) and horizontal (2F) positions only.	None
T-4	Low penetration and moderate tendency to crack for single- and multiple-pass welds in the flat (1G and 1F) and horizontal (2F) positions.	None
T-5	Low penetration and a thin, easily removed slag, used for single- and multiple-pass welds in the flat (1G and 1F) position only.	With or without carbon dioxide (CO_2)
T-6	Similar to T-5 without externally applied shielding gas.	None
T-G	The composition and classification of this electrode are not given in the preceding classes. It may be used for single- or multiple-pass welds.	With or without shielding

Table 3.4 Ferrite-forming Elements Used in FCA Welding Fluxes

Element	Reaction in Weld
Silicon (Si)	Ferrite former and deoxidizer
Chromium (Cr)	Ferrite and carbide former
Molybdenum (Mo)	Ferrite and carbide former
Columbium (Cb)	Strong ferrite former
Aluminum (Al)	Ferrite former and deoxidizer

As a result of the relatively rapid cooling of the weld metal, the weld may tend to become hard and brittle. This factor can be controlled by adding elements to the flux that affect the content of both the weld and the slag, Table 3.4. Ferrite is the softer, more ductile form of iron. The addition of ferrite-forming elements can control the hardness and brittleness of a weld. Refractory fluxes are sometimes called "fast-freeze" because they solidify at a higher temperature than the weld metal. By becoming solid first, this slag can cradle the molten weld pool and control its shape. This property is very important for out-of-position welds.

The impurities in the weld pool can be metallic or nonmetallic compounds. Metallic elements that are added to the metal during the manufacturing process in small quantities may be concentrated in the weld. These elements improve the grain structure, strength, hardness, resistance to corrosion, or other mechanical properties in the metal's as-rolled or formed state. But the deposited weld metal, or weld nugget, is like a small casting because the liquid weld metal freezes in a controlled shape, and some alloys adversely affect the properties of this casting (weld metal). Nonmetallic compounds are primarily slag inclusions left in the metal from the fluxes used during manufacturing. The welding fluxes form slags that are less dense than the weld metal so that they will float to the surface before the weld solidifies.

Flux Cored Steel Electrode Identification

Module 6
Key Indicator 3

The American Welding Society revised its A5.20 *Specification for Carbon Steel Electrodes for Flux Cored Arc Welding* in 1995 to reflect changes in the composition of the FCA filler metals. Table 3.5 lists the AWS specifications for flux cored filler metals.

Mild Steel

The electrode number *E70T-10* is used as an example to explain the classification system for mild steel FCAW electrodes (Figure 3.14):

- *E*—Electrode.
- *7*—Tensile strength in units of 10,000 psi for a good weld. This value is usually either *6* for 60,000 or *7* for 70,000 psi minimum

Table 3.5 Filler Metal Classification Numbers

Metal	AWS Filler Metal Classification
Mild steel	A5.20
Stainless steel	A5.22
Chromium–molybdenum	A5.29

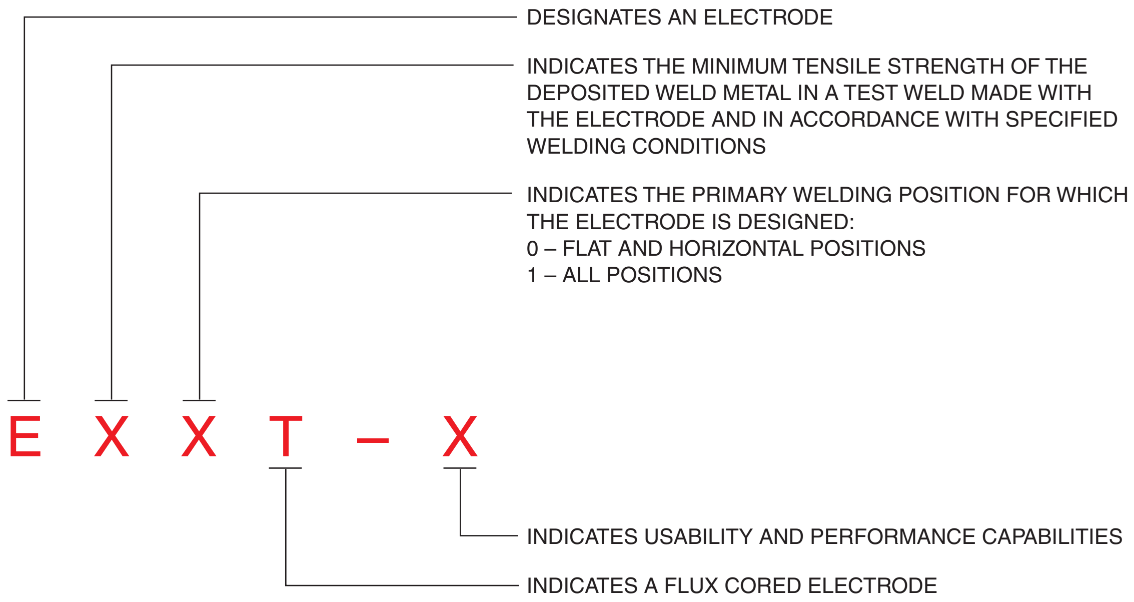
DESIGNATES AN ELECTRODE
INDICATES THE MINIMUM TENSILE STRENGTH OF THE DEPOSITED WELD METAL IN A TEST WELD MADE WITH THE ELECTRODE AND IN ACCORDANCE WITH SPECIFIED WELDING CONDITIONS
INDICATES THE PRIMARY WELDING POSITION FOR WHICH THE ELECTRODE IS DESIGNED:
0 – FLAT AND HORIZONTAL POSITIONS
1 – ALL POSITIONS
E X X T – X
INDICATES USABILITY AND PERFORMANCE CAPABILITIES
INDICATES A FLUX CORED ELECTRODE

Figure 3.14 **Identification system for mild steel FCAW electrodes**
Source: Courtesy of the American Welding Society

weld strength. An exception is for the number *12,* which is used to denote filler metals having a range from 70,000 to 90,000 psi.

- *0—0* is used for flat and horizontal fillets only, and *1* is used for all-position electrodes.
- *T*—Tubular (flux cored) electrode.
- *10*—The number in this position can range from *1* to *14* and is used to indicate the electrode's shielding gas if any, number of passes that may be applied one on top of the other, and other welding characteristics of the electrode. The letter *G* is used to indicate that the shielding gas, polarity, and impact properties are not specified. The letter *G* may or may not be followed by the letter *S. S* indicates an electrode suitable only for single-pass welding.

The electrode classification *E70T-10* can have some optional identifiers added to the end of the number, as in *E70T-10MJH8.* These additions are used to add qualifiers to the general classification so that specific codes or standards can be met. These additions have the following meanings:

- *M*—Mixed gas of 75% to 80% Ar and CO_2 for the balance. If there is no *M,* either the shielding gas is CO_2 or the electrode is self-shielded.
- *J*—Describes the Charpy V-notch impact test value of 20 ft-lb at 40°F.
- *H8*—Describes the residual hydrogen levels in the weld: *H4* equals less than 4 ml/100 g; *H8,* less than 8 ml/100 g; *H16,* less than 16 ml/100 g.

Stainless Steel Electrodes

The AWS classification for stainless steel for FCAW electrodes starts with the letter *E* as its prefix. Following the *E* prefix, the American Iron and Steel Institute's (AISI) three-digit stainless steel number is used. This number indicates the type of stainless steel in the filler metal.

To the right of the AISI number, the AWS adds a dash followed by a suffix number. The number 1 is used to indicate an all-position filler metal, and the number 3 is used to indicate an electrode to be used in the flat and horizontal positions only.

Metal Cored Steel Electrode Identification

The addition of metal powders to the flux core of FCA welding electrodes has produced a new classification of filler metals. The new filler metals evolved over time, and a new identification system was established by the AWS to identify these filler metals. Some of the earlier flux cored filler metals that already had powder metals in their core had their numbers changed to reflect the new designation. The designation was changed from the letter *T* for *tubular* to the letter *C* for *core*. For example, E70T-1 became E70C-3C. The complete explanation of the cored electrode *E70C-3C* follows:

- *E*—Electrode.
- 7—Tensile strength in units of 10,000 psi for a good weld. This value is usually either *6* for 60,000 or *7* for 70,000 psi minimum weld strength. An exception is for the number *12*, which is used to denote filler metals having a range from 70,000 to 90,000 psi.
- *0*—*0* is used for flat and horizontal fillets only, and *1* is used for all-position electrodes.
- *C*—Metal-cored (tubular) electrode.
- *3*—*3* is used for a Charpy impact of 20 ft-lb at 0°F, and *6* represents a Charpy impact of 20 ft-lb at 20°F.
- *C*—The second letter *C* indicates CO_2. The letter *M* in this position would indicate a mixed gas, 75% to 80% Ar, with the balance being CO_2. If there is no *M* or *C*, then the shielding gas is CO_2. The letter *G* is used to indicate that the shielding gas, polarity, and impact properties are not specified. The letter *G* may or may not be followed by the letter *S*. *S* indicates an electrode suitable only for single-pass welding.

NOTE

The powdered metal added to the core flux can provide additional filler metal and/or alloys. This is one way the micro-alloys can be added in very small and controlled amounts, as low as 0.0005% to 0.005%. These are very powerful alloys that dramatically improve the metal's mechanical properties.

Care of Flux Core Electrodes

Wire electrodes may be wrapped in sealed plastic bags for protection from the elements. Others may be wrapped in a special paper, and some are shipped in cans or cardboard boxes.

A small paper bag of a moisture-absorbing material, crystal desiccant, is sometimes placed in the shipping containers to protect wire electrodes from moisture. Some wire electrodes require storage in an electric rod oven to prevent contamination from excessive moisture. Read the manufacturer's recommendations located in or on the electrode shipping container for information on use and storage.

Weather conditions affect your ability to make high-quality welds. Humidity increases the chance of moisture entering the weld zone. Water (H_2O), which consists of two parts hydrogen and one part oxygen, separates in the weld pool. When only one part of hydrogen is expelled, hydrogen entrapment occurs. Hydrogen entrapment can cause weld beads to crack or become brittle. The evaporating moisture will also cause porosity.

To prevent hydrogen entrapment, porosity, and atmospheric contamination, it may be necessary to preheat the base metal to drive out moisture. Storing the wire electrode in a dry location is recommended. The electrode may develop restrictions due to the tangling of the wire or become oxidized with excessive rusting if the wire electrode package is mishandled, thrown, dropped, or stored in a damp location.

Caution

Always keep the wire electrode dry and handle it as you would any important tool or piece of equipment.

SHIELDING GAS

FCA welding wire can be manufactured so that all of the required shielding of the molten weld pool is provided by the vaporization of some of the flux within the tubular electrode. When the electrode provides all of the shielding, it is called **self-shielding** and the welding process is abbreviated FCAW-S (S for self-shielding). Other FCA welding wire must use an externally supplied shielding gas to provide the needed protection of the molten weld pool. When a shielding gas is added, the combined shielding is called **dual shield** and the process is abbreviated FCAW-G (G for gas).

Note: Sometimes the shielding gas(es) are referred to as the shielding *medium*. For example, the shielding gas, or medium, for E71T-5 is either 75% argon with 25% CO_2 or 100% CO_2.

Care must be taken to use the cored electrodes with the recommended gases, and not to use gas at all with the self-shielded electrodes. Using a self-shielding flux cored electrode with a shielding gas may produce a defective weld. The shielding gas will prevent the proper disintegration of much of the deoxidizers. This results in the transfer of these materials across the arc to the weld. In high concentrations, the deoxidizers can produce slags that become trapped in the welds, causing undesirable defects. Lower concentrations may cause brittleness only. In either case, the chance of weld failure is increased. If these electrodes are used correctly, there is no problem.

The selection of a shielding gas will affect the arc and weld properties. The weld bead width, buildup, penetration, spatter, chemical composition, and mechanical properties are all affected as a result of the shielding gas selection.

Shielding gas comes in high-pressure cylinders. These cylinders are supplied with 2000 psi of pressure. Because of this high pressure, it is important that the cylinders be handled and stored safely. For specific cylinder safety instructions see Chapter 2 in Introduction to Welding, the first book in this series.

Gases used for FCA welding include CO_2 and mixtures of argon and CO_2. Argon gas is easily ionized by the arc. Ionization results in a highly concentrated path from the electrode to the weld. This concentration results in a smaller droplet size that is associated with the axial spray mode of metal transfer, Figure 3.15. A smooth, stable arc results and there is a minimum of spatter. This transfer mode continues as CO_2 is added to the argon until the mixture contains more than 25% of CO_2.

As the percentage of CO_2 increases in the argon mixture, weld penetration increases. This increase in penetration continues until a 100% CO_2 shielding gas is reached. But as the percentage of CO_2 is increased the arc stability decreases. The less stable arc causes an increase in spatter.

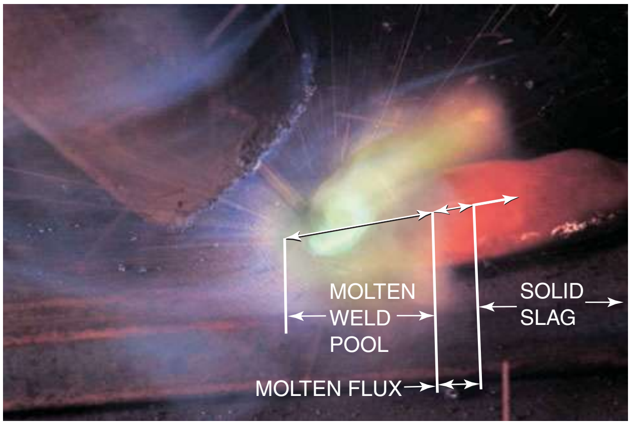

Figure 3.15 **Axial spray transfer mode**
Source: Courtesy of Larry Jeffus

A mixture of 75% argon and 25% CO_2 works best for jobs requiring a mixed gas. This mixture is sometimes called C-25.

Straight CO_2 is used for some welding. But the CO_2 gas molecule is easily broken down in the welding arc. It forms carbon monoxide (CO) and free oxygen (O). Both gases are reactive to some alloys in the electrode. As these alloys travel from the electrode to the molten weld pool, some of them form oxides. Silicon and manganese are the primary alloys that become oxidized and lost from the weld metal.

Most FCA welding electrodes are specifically designed to be used with or without shielding gas and for a specific shielding gas or percentage mixture. For example, an electrode designed specifically for use with 100% CO_2 will have higher levels of silicon and manganese to compensate for the losses to oxidization. But if 100% argon or a mixture of argon and CO_2 is used, the weld will have an excessive amount of silicon and manganese. The weld will not have the desired mechanical or metallurgical properties. Although the weld may look satisfactory, it will probably fail prematurely.

Caution

Never use an FCA welding electrode with a shielding gas it is not designated to be used with. The weld it produces may be unsafe.

WELDING TECHNIQUES

A welder can control weld beads made by FCA welding by making changes in the techniques used. The following explains how changing specific welding techniques will affect the weld produced.

Gun Angle

Gun angle, work angle, and *travel angle* are terms used to refer to the relation of the gun to the work surface, Figure 3.16. The gun angle can be used to control the weld pool. The electric arc produces an electrical force known as the arc force. The arc force can be used to counteract the gravitational pull that tends to make the liquid weld pool sag or run ahead of the arc. By manipulating the electrode travel angle for the flat and horizontal position of welding to a 20° to 45° angle from the vertical, the weld pool can be controlled. A 40° to 50° angle from the vertical plate is recommended for fillet welds.

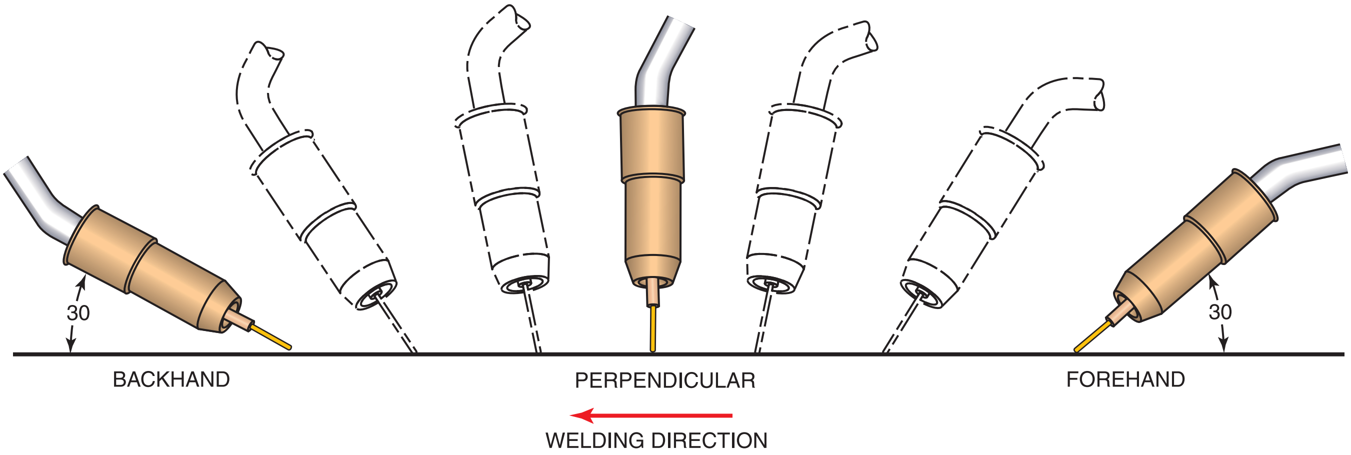

Figure 3.16 Welding gun angles

Changes in this angle will affect the weld bead shape and penetration. Shallower angles are needed when welding thinner materials to prevent burn-through. Steeper, perpendicular angles are used for thicker materials.

FCAW electrodes have a flux that is mineral based, often called low-hydrogen. These fluxes are refractory and become solid at a high temperature. If too steep a forehand, or pushing, angle is used, slag from the electrode can be pushed ahead of the weld bead and solidify quickly on the cooler plate, Figure 3.17. Because the slag remains solid at higher temperatures than the temperature of the molten weld pool, it can be trapped under the edges of the weld by the molten weld metal. To avoid this problem, most flat and horizontal welds should be performed with a backhand angle.

Vertical up welds require a forehand gun angle. The forehand angle is needed to direct the arc deep into the groove or joint for better control of the weld pool and deeper penetration, Figure 3.18. Slag entrapment associated with most forehand welding is not a problem for vertical welds.

A gun angle around 90° to the metal surface either slightly forehand or backhand works best for overhead welds, Figure 3.19. The slight angle aids with visibility of the weld, and it helps control spatter buildup in the gas nozzle.

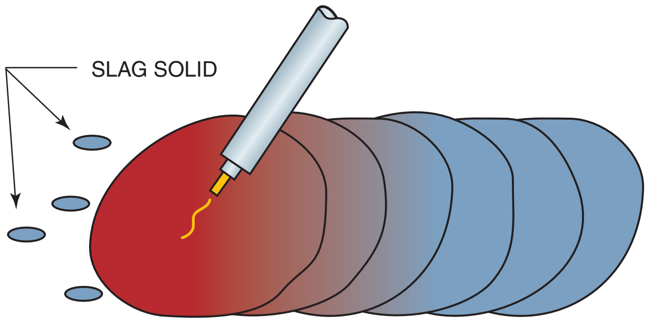

Figure 3.17 Problem of trapped slag
Large quantities of solid slag in front of a weld can cause slag to be trapped under the weld bead.

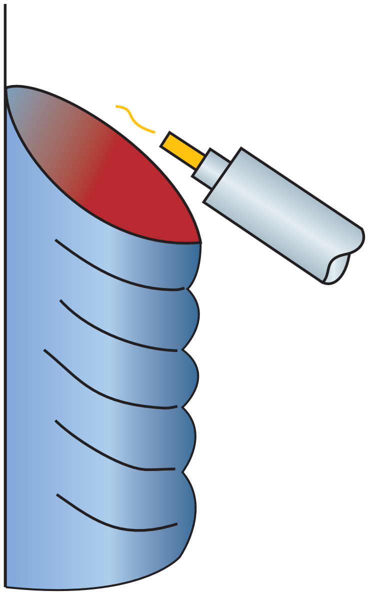

Figure 3.18 Vertical up gun angle

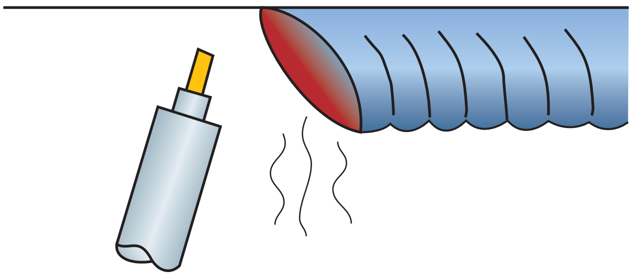

Figure 3.19 **Weld gun position to control spatter buildup on an overhead weld**

Forehand/Perpendicular/Backhand Techniques

Forehand, perpendicular, and *backhand* are the terms most often used to describe the gun angle as it relates to the work and the direction of travel. The forehand technique is sometimes referred to as *pushing* the weld bead, and backhand may be referred to as *pulling* or *dragging* the weld bead. The term *perpendicular* is used when the gun angle is at approximately 90° to the work surface, Figure 3.20.

Advantages of the Forehand Technique

The forehand welding technique has several advantages:

- Joint visibility—You can easily see the joint where the bead will be deposited, Figure 3.21.
- Electrode extension—The contact tube tip is easier to see, making it easier to maintain a constant extension length.
- Less weld penetration—It is easier to weld on thin sheet metal without melting through.
- Out-of-position welds—This technique works well on vertical up and overhead joints for better control of the weld pool.

Disadvantages of the Forehand Technique

The disadvantages of using the forehand welding technique are the following:

- Weld thickness—Thinner welds may occur because less weld reinforcement is applied to the weld joint.

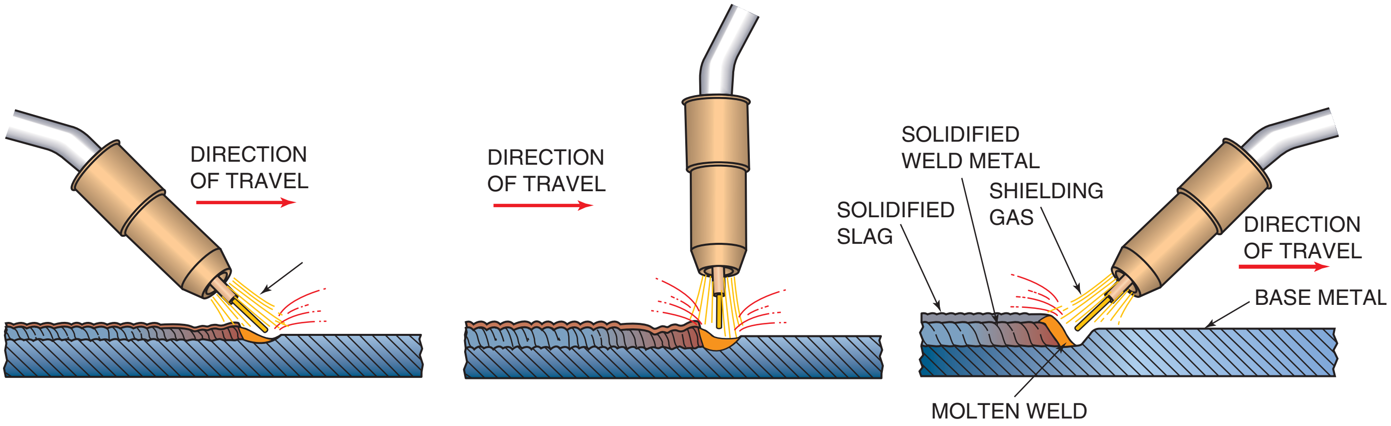

Figure 3.20 **Gun angles**
Changing the welding gun angle between forehand, perpendicular, and backhand angles will change the shape of the weld bead produced.

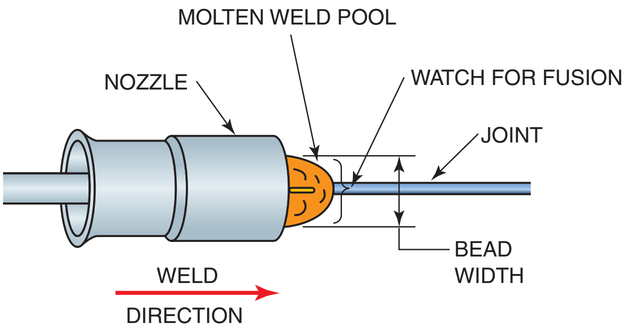

Figure 3.21 Welder's view with a forehand angle
This angle keeps the shielding gas nozzle from restricting the welder's view.

- Welding speed—Because less weld metal is being applied, the rate of travel along the joint can be faster, which may make it harder to create a uniform weld.
- Slag inclusions—Some spattered slag can be thrown in front of the weld bead and be trapped or included in the weld, resulting in a weld defect.
- Spatter—Depending on the electrode, the amount of spatter may be slightly increased with the forehand technique.

Advantages of the Perpendicular Technique

The perpendicular welding technique has the following advantages:

- Machine and robotic welding—The perpendicular gun angle is used on automated welding because there is no need to change the gun angle when the weld changes direction.
- Uniform bead shape—The weld's penetration and reinforcement are balanced between those of forehand and backhand techniques.

Disadvantages of the Perpendicular Technique

The disadvantages of using the perpendicular welding technique are the following:

- Limited visibility—Because the welding gun is directly over the weld, there is limited visibility of the weld unless you lean your head way over to the side.
- Weld spatter—Because the weld nozzle is directly under the weld in the overhead position, more weld spatter can collect in the nozzle, causing gas flow problems or even shorting the tip to the nozzle.

Advantages of the Backhand Technique

The backhand welding technique has the following advantages:

- Weld bead visibility—It is easy to see the back of the molten weld pool as you are welding, which makes it easier to control the bead shape, Figure 3.22.
- Travel speed—Because of the larger amount of weld metal being applied, the rate of travel may be slower, making it easier to create a uniform weld.
- Depth of fusion—The arc force and the greater heat from the slower travel rate both increase the depth of weld joint penetration.

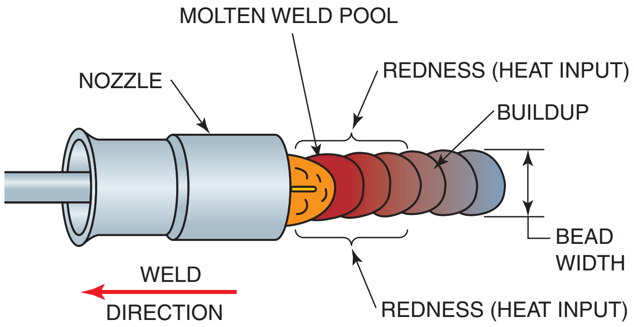

Figure 3.22 **Welder's view with a backhand angle**
Watch the trailing edge of the molten weld pool.

Disadvantages of the Backhand Technique

The disadvantages of the backhand welding technique are the following:

- Weld buildup—The weld bead may have a convex (raised or rounded) weld face when you use the backhand technique.
- Postweld finishing—Because of the weld bead shape, more work may be required if the product has to be finished by grinding smooth.
- Joint following—It is harder to follow the joint because your hand and the FCAW gun are positioned over the joint, and you may wander from the seam.
- Loss of penetration—An inexperienced welder sometimes directs the wire too far back into the weld pool causing the wire to build up in the face of the weld pool reducing joint penetration.

Travel Speed

The American Welding Society defines travel speed as the linear rate at which the arc is moved along the weld joint. Fast travel speeds deposit less filler metal. If the rate of travel increases, the filler metal cannot be deposited fast enough to adequately fill the path melted by the arc. This causes the weld bead to have a groove melted into the base metal next to the weld and left unfilled by the weld. This condition is known as undercut.

Undercut occurs along the edges or toes of the weld bead. Slower travel speeds will, at first, increase penetration and increase the filler weld metal deposited. As the filler metal increases, the weld bead will build up in the weld pool. Because of the deep penetration of flux cored wire, the angle at which you hold the gun is very important for a successful weld.

If all welding conditions are correct and remain constant, the preferred rate of travel for maximum weld penetration is a travel speed that allows you to stay within the selected welding variables and still control the fluidity of the weld pool. This is an intermediate travel speed, or progression, which is not too fast or too slow.

Another way to figure out correct travel speed is to consult the manufacturer's recommendations chart for the inches per minute (ipm) burn-off rate for the selected electrode.

Mode of Metal Transfer

The mode of metal transfer is used to describe how the molten weld metal is transferred across the arc to the base metal. The mode of metal transfer that is selected, the shape of the completed weld bead, and the depth of weld penetration depend on the welding power source, wire electrode size, type and thickness of material, type of shielding gas used, and best welding position for the task.

Spray Transfer with FCAW-G

The spray transfer mode is the most common process used with gas-shielded FCAW (FCAW-G), Figure 3.15.

As the gun trigger is depressed, the shielding gas automatically flows and the electrode bridges the distance from the contact tube to the base metal, making contact with the base metal to complete a circuit. The electrode shorts and becomes so hot that the base metal melts and forms a weld pool. The electrode melts into the weld pool and burns back toward the contact tube. A combination of high amperage and the shielding gas along with the electrode size produces a pinching effect on the molten electrode wire, causing the end of the electrode wire to spray across the arc.

The characteristic of spray-type transfer is a smooth arc, through which hundreds of small droplets per second are transferred through the arc from the electrode to the weld pool. At that moment a transfer of metal is taking place. Spray transfer can produce a high quantity of metal droplets, up to approximately 250 per second above the transition current, or critical current. This means the current required for a spray transfer to take place is dependent on the electrode size, composition of the electrode, and shielding gas. Below the transition current (critical current), globular transfer takes place.

In order to achieve a spray transfer, high current and larger-diameter electrode wire are needed. A shielding gas of carbon dioxide (CO_2), a mixture of carbon dioxide (CO_2) and argon (Ar), or an argon (Ar) and oxygen (O_2) mixture is needed. FCAW-G is a welding process that, with the correct variables, can be used

- on thin or properly prepared thick sections of material
- on a combination of thick to thin materials
- with small or large electrode diameters
- with a combination of shielding gases

Globular Transfer with FCAW-G

Globular transfer occurs when the welding current is below the transition current, Figure 3.23. The electrode forms a molten ball at its end that grows in size to approximately two to three times the original electrode diameter. These large molten balls are then transferred across the arc at the rate of several drops per second.

The arc becomes unstable because of the gravitational pull from the weight of these large drops. A spinning effect caused by a natural phenomenon takes place when argon gas is introduced to a large ball of molten metal on the electrode. The molten ball spins as it transfers across the arc to the base metal. This unstable globular transfer can produce excessive spatter.

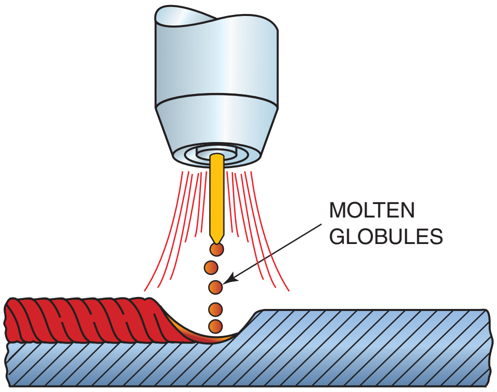

Figure 3.23 Globular transfer method

Both FCAW-S and FCAW-G use direct current electrode negative (DCEN) when welding on thin-gauge materials to keep the heat in the base metal and the small-diameter electrode at a controllable burn-off rate. The electrode can then be stabilized, and it is easier to manipulate and control the weld pool in all weld positions. Larger-diameter electrodes are welded with direct current electrode positive (DCEP) because the larger diameters can keep up with the burn-off rates.

The recommended weld position means the position in which the workpiece is placed for welding. All welding positions use either spray or globular transfer, but for now we will concentrate on the flat and horizontal welding positions.

In the flat welding position the workpiece is placed flat on the work surface. In the horizontal welding position the workpiece is positioned perpendicular to the workbench surface.

The amperage range may be from 30 to 400 amperes or more for welding materials from gauge thickness up to 1-1/2 inches. On square groove weld joints, thicker base metals can be welded with little or no edge preparation. This is one of the great advantages of FCAW. If edges are prepared and cut at an angle (beveled) to accept a complete joint weld penetration, the depth of penetration will be greatly increased. FCAW is commonly used for general repairs to mild steel in the horizontal, vertical, and overhead welding positions, sometimes referred to as out-of-position welding.

Electrode Extension

The electrode extension is measured from the end of the electrode contact tube to the point the arc begins at the end of the electrode, Figure 3.24. Compared to GMA welding, the electrode extension required for FCAW is much greater. The longer extension is required for several reasons. The electrical resistance of the wire causes the wire to heat up, which can drive out moisture from the flux. This preheating of the wire also results in a smoother arc with less spatter.

Porosity

FCA welding can produce high-quality welds in all positions, although porosity in the weld can be a persistent problem. Porosity can be caused by moisture in the flux, improper gun manipulation, or surface contamination.

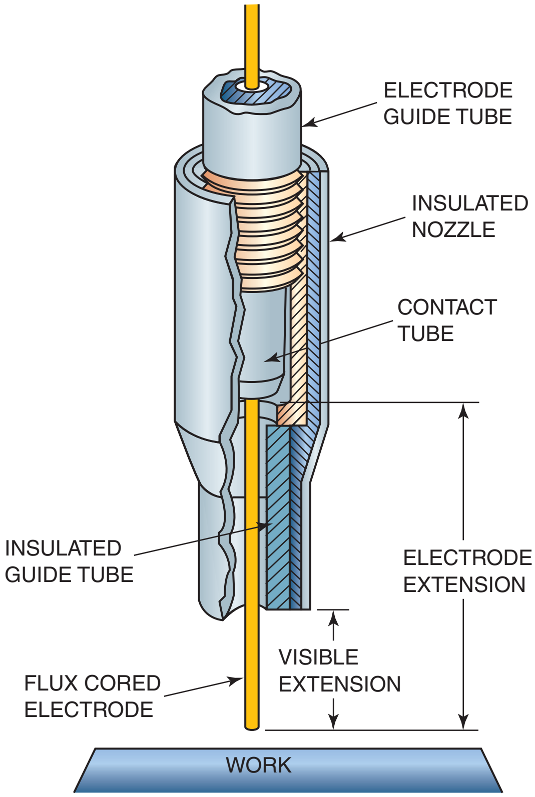

Figure 3.24 **Self-shielded electrode nozzle**
Source: Courtesy of the American Welding Society

The flux used in the FCA welding electrode is subject to picking up moisture from the surrounding atmosphere, so the electrodes must be stored in a dry area. Once the flux becomes contaminated with moisture, it is very difficult to remove. Water (H_2O) breaks down into free hydrogen and oxygen in the presence of an arc, Figure 3.25. The hydrogen can be absorbed into the molten weld metal, where it can cause postweld cracking. The oxygen is absorbed into the weld metal also, but it forms oxides in the metal.

If a shielding gas is used, the FCA welding gun gas nozzle must be close enough to the weld to provide adequate shielding gas coverage. If there is a wind or if the nozzle-to-work distance is excessive, the shielding will be inadequate and allow weld porosity. If welding is to be done outside or in an area subject to drafts, the gas flow rate must be increased or a wind shield must be placed to protect the weld, Figure 3.26.

A common misconception is that the flux within the electrode will either remove or control weld quality problems caused by surface contaminations. That is not true. The addition of flux makes FCA welding more tolerant to surface conditions than GMA welding, although it still is adversely affected by such contaminations.

New hot-rolled steel has a layer of dark gray or black iron oxide called mill scale. Although this layer is very thin, it may provide a source of enough oxygen to cause porosity in the weld. If mill scale causes porosity, it is usually uniformly scattered through the weld, Figure 3.27. Unless it is severe, uniformly scattered porosity is usually not visible in the finished weld. It is trapped under the surface as the weld cools.

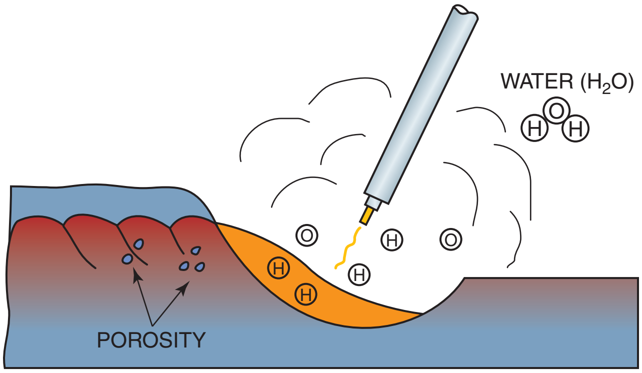
WATER (H_2O)
O
H
H
O
H
O
H
H
H
POROSITY

Figure 3.25 **Water and porosity**
Water (H_2O) breaks down in the presence of the arc and the hydrogen (H) is dissolved in the molten weld metal.

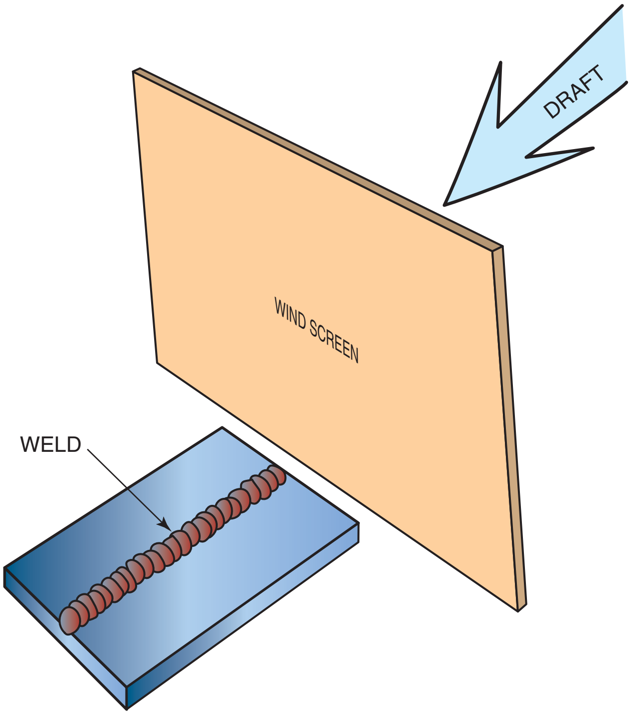
DRAFT
WIND SCREEN
WELD

Figure 3.26 **Wind and draft protection**
A wind screen can keep the welding shielding from being blown away.

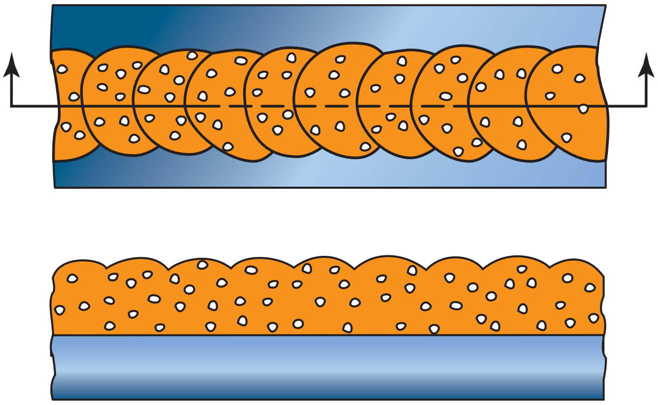

Figure 3.27 **Uniformly scattered porosity**

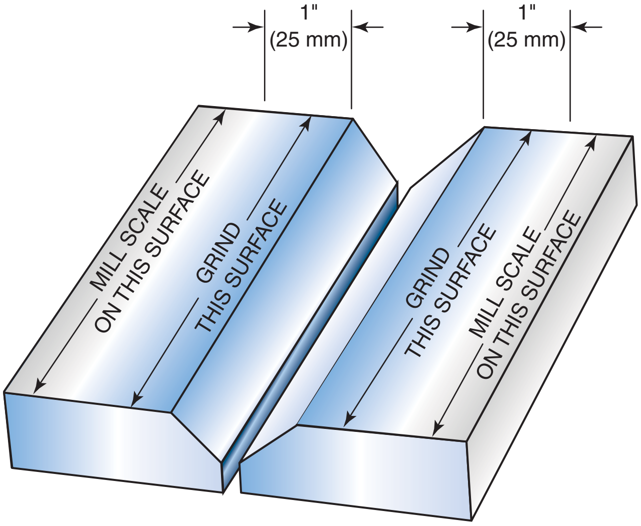

Figure 3.28 Grinding requirements
Grind mill scale off plates within 1 in. (25 mm) of the groove.

Because porosity is under the weld surface, nondestructive testing methods, including X ray, magnetic particle, and ultrasound, must be used to locate it in a weld. It can be detected by mechanical testing such as guided bend, free bend, and nick-break testing for establishing weld parameters. Often it is better to remove the mill scale before welding rather than risking the production of porosity.

All welding surfaces within the weld groove and the surrounding surfaces within 1 in. (25 mm) must be cleaned to bright metal, Figure 3.28. Cleaning may be either grinding, filing, sanding, or blasting.

Any time FCA welds are to be made on metals that are dirty, oily, rusty, or wet or that have been painted, the surface must be precleaned. Cleaning can be done chemically or mechanically.

One advantage of chemically cleaning oil and paint is that it is easier to clean larger areas. Both oil and paint smoke easily when heated, and such smoke can cause weld defects. They must be removed far enough from the weld so that weld heat does not cause them to smoke. In the case of small parts the entire part may need to be cleaned.

Caution

Chemically cleaning oil and paint off metal must be done according to the cleaner manufacturer's directions. The work must be done in an appropriate, approved area. The metal must be dry, and all residues of the cleaner must be removed before welding begins.

SUMMARY

Flux cored arc welding is used to produce more tons of welded fabrications than any other process. The ability to produce high-quality welds on a wide variety of material thicknesses and joint configurations has led to its popularity. As you learn and develop these skills, you will therefore be significantly increasing your employability and productivity in the welding industry.

A wide variety of filler metals and shielding gas combinations for flux cored arc welding are available to you in industry. These various materials aid in producing welds of high quality under various welding conditions. Although the selection of the proper filler metal and gas coverage, if used, will significantly affect the finished weld's quality in the field, there are very few differences in manipulation and setup among these filler metals. Therefore, as you practice welding in a school or training program and learn to use a specific wire and shielding gas mixture, these skills are easily transferable to the next group of materials you will encounter on the job.

REVIEW

1. List some factors that have led to the increased use of FCA welding.
2. How is FCAW similar to GMAW?
3. What does the FCA flux provide to the weld?
4. What are the major atmospheric contaminations of the molten weld metal?
5. How does slag help an FCA weld?
6. What is the electrical difference between a constant-potential and a constant-current power supply?
7. How can FCA welding guns be cooled?
8. What problems does excessive drive roller pressure cause?
9. List the advantages that FCA welding offers the welding industry.
10. Describe the two methods of manufacturing FCA electrode wire.
11. Why are large-diameter electrodes not used for all-position welding?
12. How do deoxidizers remove oxygen from the weld zone?
13. What do fluxing agents do for a weld?
14. Why are alloying elements added to the flux?
15. How does the flux form a shielding gas to protect the weld?
16. What are the main limitations of the rutile fluxes?
17. Why is it more difficult to use lime-based fluxed electrodes on out-of-position welds?
18. What benefit does adding an externally supplied shielding gas have on some rutile-based electrodes?
19. How do excessive amounts of manganese affect a weld?
20. Why are elements added that cause ferrite to form in the weld?
21. Why are some slags called refractory?
22. Why must a flux form a less dense slag?
23. Referring to Table 3.5, what is the AWS classification for FCA welding electrodes for stainless steel?
24. Describe the meaning of each part of the following FCA welding electrode identification: E81T-5.
25. What does the number 316 in E316T-1 mean?
26. What is the advantage of using an argon-CO_2 mixed shielding gas?
27. What are the primary alloying elements lost if 100% CO_2 shielding gas is used?
28. What can cause porosity in an FCA weld?
29. What happens to water in the welding arc?
30. What is the thin, dark gray or black layer on new hot-rolled steel? How can it affect the weld?
31. Why is uniformly scattered porosity hard to detect in a weld?
32. What cautions must be taken when chemically cleaning oil or paint from a piece of metal?
33. What can happen to slag that solidifies on the plate ahead of the weld?
34. How is the electrode extension measured?

CHAPTER 4

Flux Cored Arc Welding

OBJECTIVES

After completing this chapter, the student should be able to

- set up the FCA weld station
- thread the electrode wire through the system
- list three disadvantages of having to bevel a plate before welding
- make root, filler, and cover passes with the FCAW process
- make butt welds in all positions that can pass a specified standard's visual or destructive examination criteria
- make fillet welds in tee joints and lap joints in all positions that can pass a specified standard's visual or destructive examination criteria

KEY TERMS

amperage range
conduit liner
contact tube
critical weld
feed rollers
lap joint
root face
stringer bead
tee joint
voltage range
weave bead
wire-feed speed

AWS SENSE EG2.0

Key Indicators Addressed in this Chapter:

Module 1: Occupational Orientation

Key Indicator 1: Prepares time or job cards, reports, or records
Key Indicator 2: Performs housekeeping duties
Key Indicator 3: Follows verbal instructions to complete work assignments
Key Indicator 4: Follows written instructions to complete work assignments

Module 6: Flux Cored Arc Welding (FCAW-G/GM, FCAW-S)

Key Indicator 1: Performs safety inspections of FCAW equipment and accessories
Key Indicator 2: Makes minor external repairs to FCAW equipment and accessories
Gas Shielded
Key Indicator 3: Sets up for FCAW-G/GM operations on carbon steel
Key Indicator 4: Operates FCAW-G/GM equipment on carbon steel
Key Indicator 5: Makes FCAW-G/GM fillet welds, in all positions, on carbon steel

Key Indicator 6: Makes FCAW-G/GM groove welds, in all positions, on carbon steel
Key Indicator 7: Passes FCAW-G/GM welder performance qualification testing (workmanship sample) on carbon steel

Self Shielded

Key Indicator 8: Sets up for FCAW-S operations on carbon steel
Key Indicator 9: Operates FCAW-S equipment on carbon steel
Key Indicator 10: Makes FCAW-S fillet welds, in all positions, on carbon steel
Key Indicator 11: Makes FCAW-S groove welds, in all positions, on carbon steel
Key Indicator 12: Passes FCAW-S welder performance qualification test (workmanship sample) on carbon steel

Module 9: Welding Inspection and Testing Principles

Key Indicator 1: Examines cut surfaces and edges of prepared base metal parts
Key Indicator 2: Examines tacks, root passes, intermediate layers, and completed welds

INTRODUCTION

Setup of the flux cored arc welding (FCAW) work station is the key to making quality welds. It may be possible, using a poorly set up FCA welder, to make an acceptable weld in the flat position. The FCA welding process is often forgiving; thus welds can often be made even when the welder is not set correctly. However, such welds will have major defects such as excessive spatter, undercut, overlap, porosity, slag inclusions, and poor weld bead contours. Setup becomes even more important for out-of-position welds. Making vertical and overhead welds can be difficult for a student welder with a properly set up system, but it becomes impossible with a system that is out of adjustment.

Learning to set up and properly adjust the FCA welding system will allow you to produce high-quality welds at a high level of productivity.

FCAW is set up and manipulated in a manner similar to that of GMAW. The results of changes in electrode extension, voltage, amperage, and torch angle are essentially the same.

Although every manufacturer's FCA welding equipment is designed differently, all equipment is set up in a similar manner. It is always best to follow the specific manufacturer's recommendations regarding setup as provided in its equipment literature. You will find, however, that, in the field, manufacturers' literature is not always available for the equipment you are asked to use. It is therefore important to have a good general knowledge and understanding of the setup procedure for FCA welding equipment. Figure 4.1 shows all of the various components that make up an FCA welding station.

Caution

FCA welding produces a lot of ultraviolet light, heat, sparks, slag, and welding fumes. Proper personal, protective clothing and special protective clothing must be worn to prevent burns from the ultraviolet light and hot weld metal. Eye protection must be worn to prevent injury from flying sparks and slag. Forced ventilation and possibly a respirator must be used to prevent fume-related injuries. Refer to the safety precautions provided by the equipment and electrode manufacturers and to Chapter 2 in *Welding Skills, Processes and Practices for Entry-Level Welders: Book One* for additional safety help.

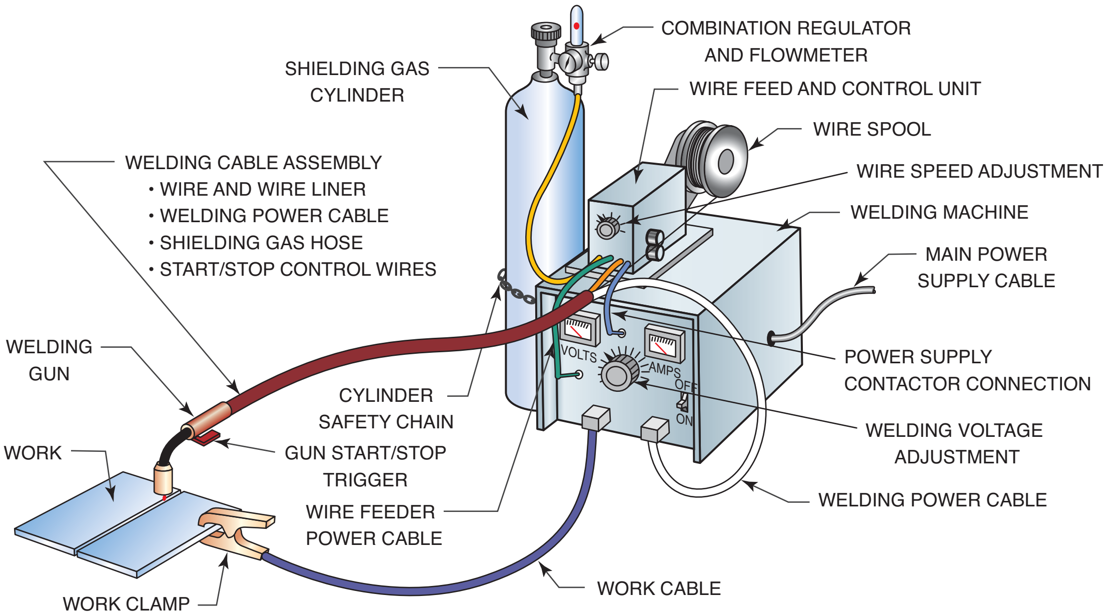
COMBINATION REGULATOR AND FLOWMETER
SHIELDING GAS CYLINDER
WIRE FEED AND CONTROL UNIT
WIRE SPOOL
WELDING CABLE ASSEMBLY
• WIRE AND WIRE LINER
• WELDING POWER CABLE
• SHIELDING GAS HOSE
• START/STOP CONTROL WIRES
WIRE SPEED ADJUSTMENT
WELDING MACHINE
MAIN POWER SUPPLY CABLE
WELDING GUN
VOLTS
AMPS
OFF
ON
POWER SUPPLY CONTACTOR CONNECTION
CYLINDER SAFETY CHAIN
WORK
GUN START/STOP TRIGGER
WELDING VOLTAGE ADJUSTMENT
WIRE FEEDER POWER CABLE
WELDING POWER CABLE
WORK CABLE
WORK CLAMP

Figure 4.1 **Basic FCA welding equipment identification**

PRACTICES

The practices in this chapter are grouped according to those requiring similar techniques and setups. Plate welds are covered first, then sheet metal. The practices start with 1/4-in. (6-mm) mild steel plates; they are used because they require the least preparation times. The thicker 3/8-in. (9.5-mm) plates provide the basics of practicing groove welding. The 3/4-in. (19-mm) and thicker plates are used to develop the skills required to pass the unlimited thickness test often given to FCA welders. Sheet metal is grouped together because it presents a unique set of learning skills.

The major skill required for making consistently acceptable FCA welds is the ability to set up the welding system. Changes such as variations in material thickness, position, and type of joint require changes both in technique and setup. A correctly set up FCA welding station can, in many cases, be operated by a less-skilled welder. Often the only difference between a welder earning a minimum wage and one earning the maximum wage is the ability to correct machine setups.

For several reasons the FCA welding practice plates will be larger than most other practice plates. Welding heat and welding speed are the major factors that necessitate this increased size. FCA welding is both high energy and fast, and the welding energy (heat) input is so great that small practice plates may glow red by the end of a single weld pass. This would seriously affect the weld quality. To prevent this from happening, wider plates are used. Because of the higher welding speeds, longer plates are usually used.

Plates less than 1/2 in. (13 mm) will be 12 in. (305 mm) long for most practices. In addition to controlling the heat buildup, the longer plates are needed to give the welder enough time to practice welding. Learning to make longer welds is a skill that must also be practiced, because the FCA welding process is used in industry to make long production welds.

Plates thicker than 1/2 in. (13 mm) can be shorter than 12 in. (305 mm). Most codes allow test plates of "unlimited thickness" to be as short as 7 in. (178 mm).

PRACTICE 4-1

FCAW Equipment Setup

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Key Indicator 2
Gas Shielded
Key Indicator 3
Self Shielded
Key Indicator 8

For this practice, you will need a semiautomatic welding power source approved for FCA welding, welding gun, electrode feed unit, electrode supply, shielding gas supply, shielding gas flowmeter regulator, electrode conduit, power and work leads, shielding gas hoses (if required), assorted hand tools, spare parts, and any other required materials. In this practice, you will demonstrate to a group of students and your instructor how to properly set up an FCA welding station. Some manufacturers include detailed setup instructions with their equipment. If such instructions are available for your equipment, follow them. Otherwise, use the following instructions.

If the shielding gas is to be used and it comes from a cylinder, the cylinder must be chained securely in place before the valve protection cap is removed, Figure 4.2. Standing to one side of the cylinder, make sure nobody is in line with the valve and quickly crack the valve to blow out any dirt in the valve before the flowmeter regulator is attached, Figure 4.3. Attach the correct hose from the regulator to the "gas-in" connection on the electrode feed unit or machine.

Install the reel of electrode (welding wire) on the holder and secure it, Figure 4.4. Check the feed roller size to ensure that it matches the wire size, Figure 4.5. Also check the **conduit liner** size for compatibility with the wire size. Connect the conduit to the feed unit. The conduit or an extension should be aligned with the groove in the roller and set as close to the roller as possible without touching, Figure 4.6. Misalignment at this point can contribute to a bird's nest, Figure 4.7. Bird-nesting of the electrode wire, so called because it looks like a bird's nest, results when the feed roller pushes the wire into a tangled ball because the wire would not go through the outfeed side conduit.

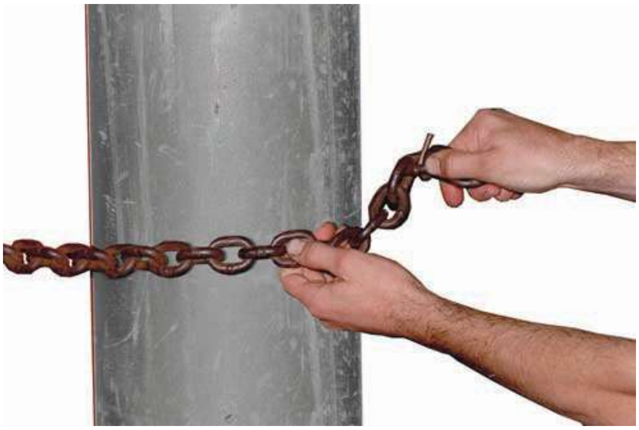

Figure 4.2 **Make sure the gas cylinder is chained securely in place before removing the safety cap**
Source: Courtesy of Larry Jeffus

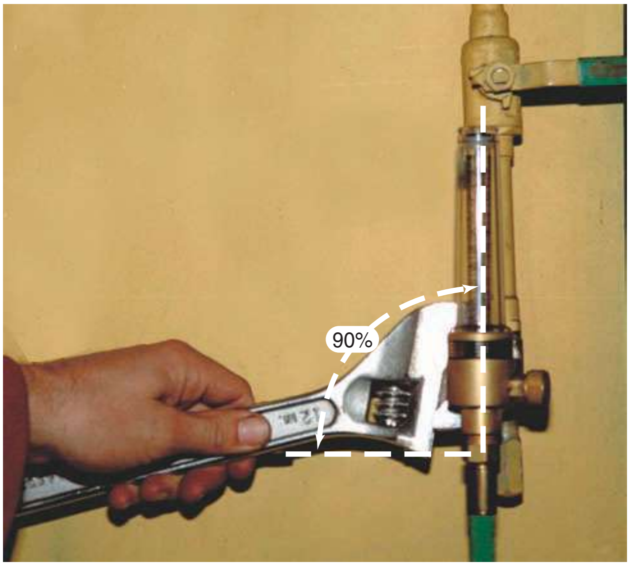

Figure 4.3 Attaching the flowmeter regulator
Be sure the tube is vertical.
Source: Courtesy of Larry Jeffus

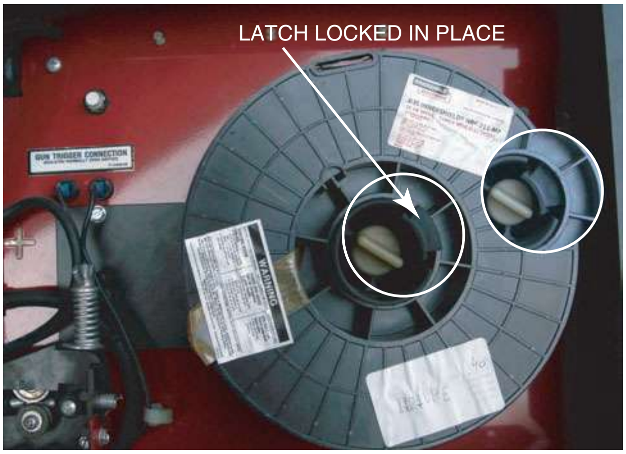
LATCH LOCKED IN PLACE

Figure 4.4 Wire reel may be secured by a center nut or locking lever
Source: Courtesy of Larry Jeffus

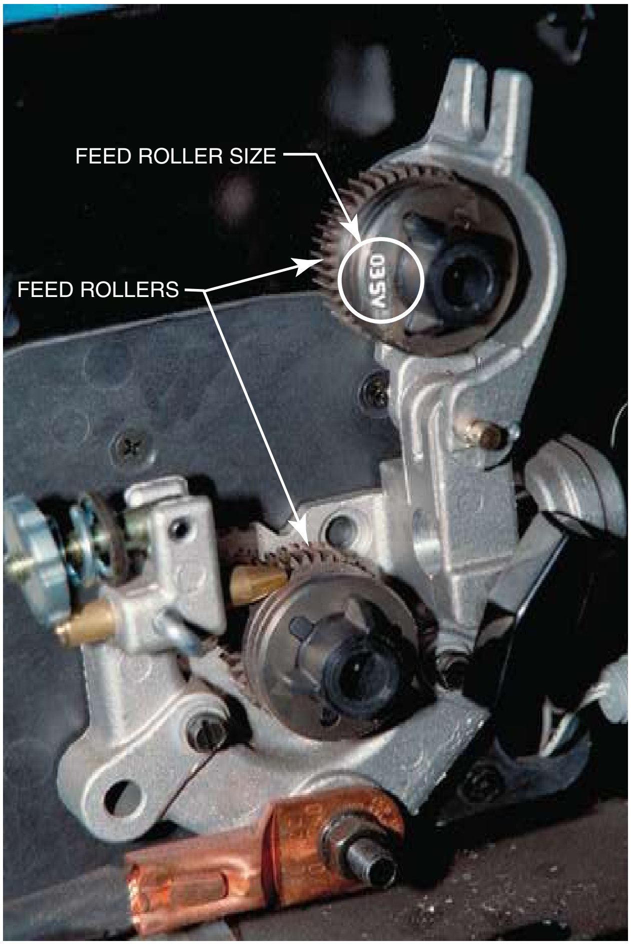
FEED ROLLER SIZE
FEED ROLLERS
035V

Figure 4.5 Checking feed roller size
Check to be certain that the feed rollers are the correct size for the wire being used.
Source: Courtesy of Larry Jeffus

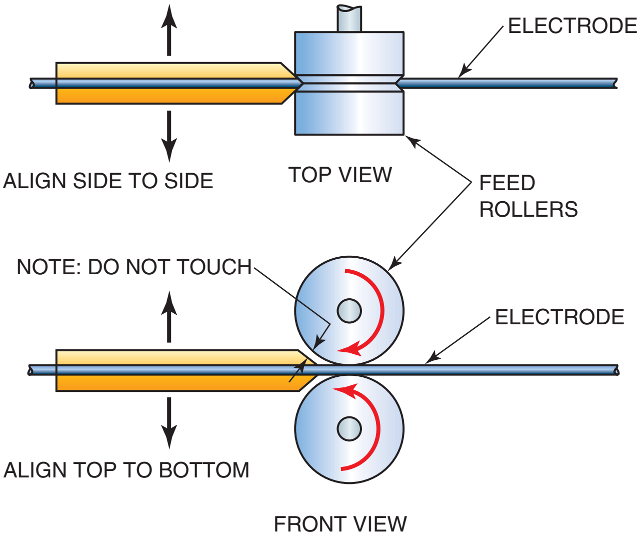

Figure 4.6 Feed roller and conduit alignment

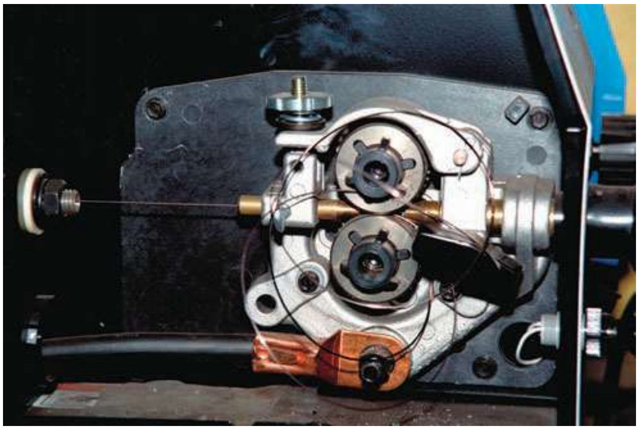

Figure 4.7 "Bird's nest" in the filler wire at the feed rollers
Source: Courtesy of Larry Jeffus

Be sure the power is off before attaching the welding cables. The electrode and work leads should be attached to the proper terminals. The electrode lead should be attached to the electrode or positive (+) terminal. If necessary, it is also attached to the power cable part of the gun lead. The work lead should be attached to the work or negative (–) terminal.

The shielding "gas-out" side of the solenoid is then also attached to the gun lead. If a separate splice is required from the gun switch circuit to the feed unit, it should be connected at this time. Check that the welding contactor circuit is connected from the feed unit to the power source.

The welding cable liner or wire conduit must be securely attached to the gas diffuser and contact tube, Figure 4.8. The **contact tube** (tip), Figure 3.1, must be the correct size to match the electrode wire size being used. If a shielding gas is to be used, a gas nozzle would be attached to complete the assembly. If a gas nozzle is not needed for a shielding gas, it may still be installed. Because it is easy for a student to touch the work with the contact tube during welding, an electrical short may occur. This short-out of the contact tube will immediately destroy the tube. Although the gas nozzle may interfere with some visibility, it may be worth the trouble for a new welder. FCA welding is more sensitive to changes in arc voltage than is SMA (stick) welding. Such variations in FCA welding voltage can dramatically and adversely affect your ability to maintain weld bead control.

A loose or poor connection will result in increased circuit resistance and a loss of welding voltage. To be sure that you have a good work

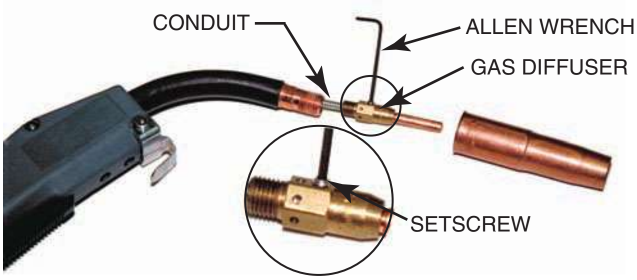

Figure 4.8 Securely attach conduit to gas diffuser and contact tube to prevent wire jams caused by misalignment
Source: Courtesy of Larry Jeffus

connection, remove any dirt, rust, oil, or other surface contamination at the point where the work clamp is connected to the weldment.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-2

Threading FCAW Wire

Using the FCAW machine that was properly assembled in Practice 4-1, you will turn the machine on and thread the electrode wire through the system.

Check that the unit is assembled correctly according to the manufacturer's specifications. Switch on the power and check the gun switch circuit by depressing the switch. The power source relays, feed relays, gas solenoid, and feed motor should all activate.

Cut off the end of the electrode wire if it is bent. When working with the wire, be sure to hold it tightly. The wire will become tangled if it is released. The wire has a natural curl known as *cast.* Straighten out about 12 in. (300 mm) of the curl to make threading easier. Separate the wire **feed rollers** and push the wire first through the guides, then between the rollers, and finally into the conduit liner, Figure 4.9. Reset the rollers so there is a slight amount of compression on the wire, Figure 4.10. Set the **wire-feed speed** control to a slow speed. Hold the welding gun so that the electrode conduit and cable are as straight as possible.

Wearing safety glasses and pointing the gun away from the welder's face, press the gun switch or the cold feed switch on the feeder. Pressing the gun switch to start the wire feeder is called triggering the gun. The cold feed switch on the feeder is a safety option built into some equipment that advances the wire without current being sent to the gun. The wire should start feeding into the liner. Watch to make certain that the wire feeds smoothly and release the gun switch as soon as the end comes through the gun.

If the wire stops feeding before it reaches the end of the gun, stop and check the system. If no obvious problem can be found, mark the wire with tape and remove it from the gun. It can then be held next to the system to determine the location of the problem.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Key Indicator 2
Gas Shielded
Key Indicator 3
Self Shielded
Key Indicator 8

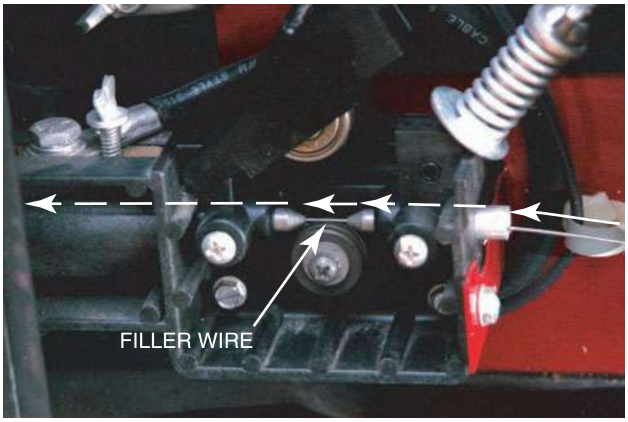

Figure 4.9 Push the wire through the guides by hand
Source: Courtesy of Larry Jeffus

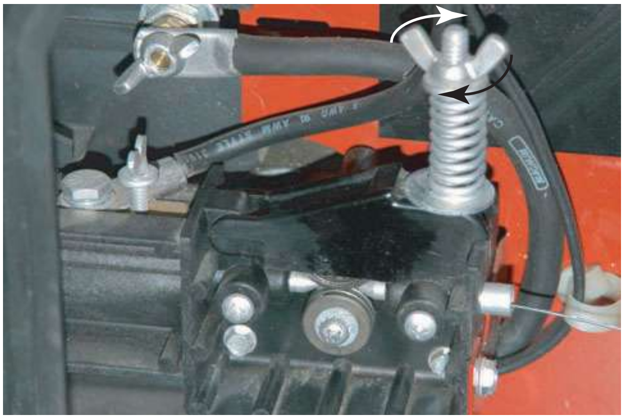

Figure 4.10 Adjust the wire-feed tensioner
Source: Courtesy of Larry Jeffus

With the wire feed running, adjust the feed roller compression so that the wire reel can be stopped easily by a slight pressure. Too light a roller pressure will cause the wire to feed erratically. Too high a pressure can crush some wires, causing some flux to be dropped inside the wire liner. If this happens, you will have a continual problem with the wire not feeding smoothly or jamming.

With the feed running, adjust the spool drag so that the reel stops when the feed stops. The reel should not coast to a stop, because it allows slack in the wire that can easily be snagged. Also, when the feed restarts, a jolt occurs when the slack in the wire is taken up. This jolt can be enough to momentarily stop the wire, possibly causing a discontinuity in the weld.

When the test runs are completed, you can either rewind or cut off the wire. Some wire-feed units have a retract button. This allows the feed driver to reverse and retract the wire automatically. To rewind the wire on units without this retraction feature, release the rollers and turn them backward by hand. If the machine will not allow the feed rollers to be released without upsetting the tension, you must cut the wire. Some wire reels have covers to prevent the collection of dust, dirt, and metal filings on the wire, Figure 4.11.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

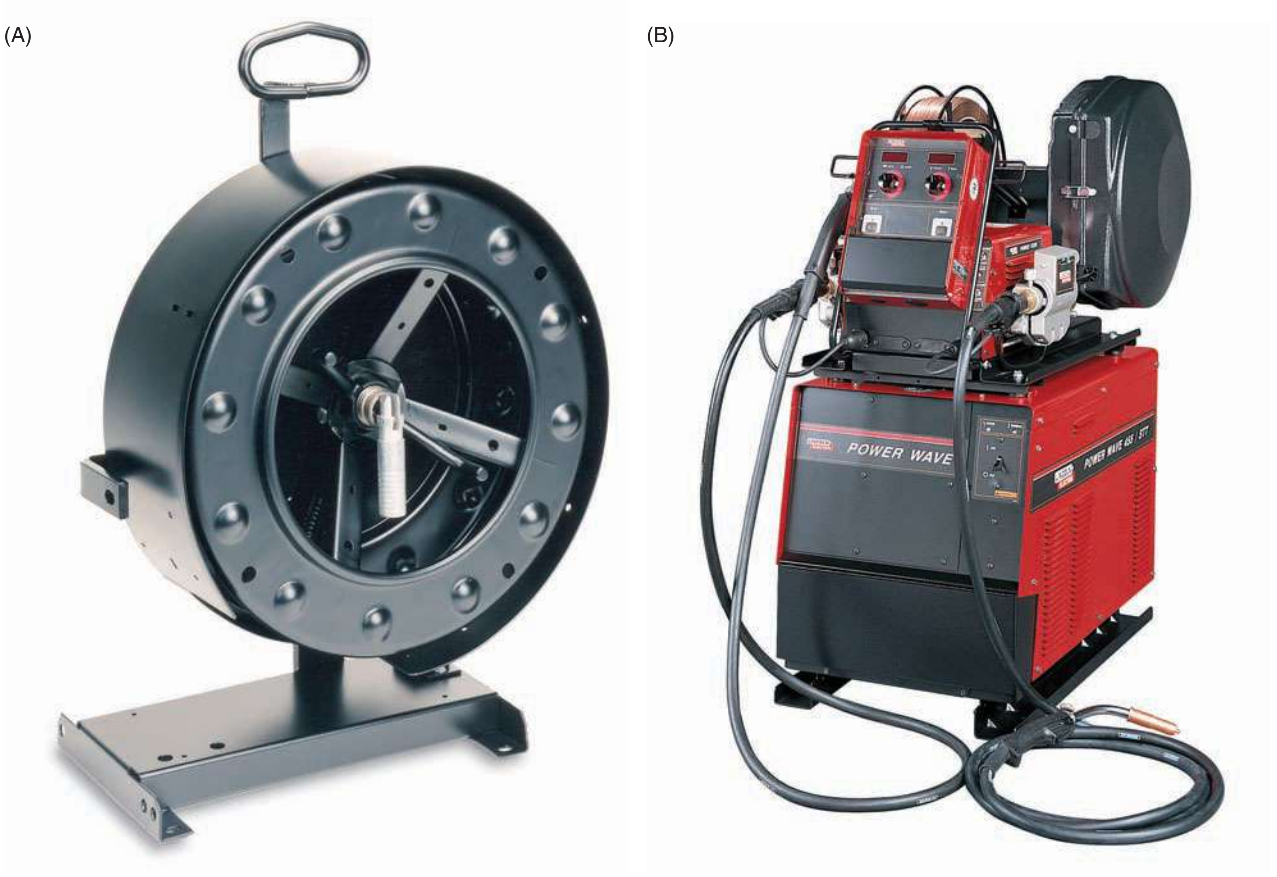

Figure 4.11 Wire protection
(A) Covered wire reel. (B) Wire cover on a dual wire-feed system.
Source: Courtesy of Lincoln Electric Company

FLAT-POSITION WELDS

PRACTICE 4-3

Stringer Beads, Flat Position

Using a properly set up and adjusted FCA welding machine, Table 4.1, proper safety protection, E70T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm), and one or more pieces of mild steel plate, 12 in. (305 mm) long and 1/4 in. (6 mm) or thicker, you will make a stringer bead weld in the flat position, Figure 4.12.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Key Indicator 1
Gas Shielded
Key Indicator 4
Self Shielded
Key Indicator 9

Table 4.1 **FCA Welding Parameters for Use if Specific Settings Are Unavailable from Electrode Manufacturer (base metal thickness 1/4 to 1/2 inch)**

Electrode		Welding Power			Shielding Gas		Base Metal	
Type	Size	Amps	Wire-feed Speed, ipm (cm/min.)	Volts	Type	Flow	Type	Thickness
E70T-1 E71T-1	0.035 in. (0.9 mm)	130 to 150	288 to 380 (732 to 975)	22 to 25	None	n/a	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)
E70T-1 E71T-1	0.045 in. (1.2 mm)	150 to 210	200 to 300 (508 to 762)	28 to 29	None	n/a	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)
E70T-5 E71T-11	0.035 in. (0.9 mm)	130 to 200	288 to 576 (732 to 1463)	20 to 28	75% argon 25% CO_2	30 cfh	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)
E70T-5 E71T-11	0.045 in. (1.2 mm)	150 to 250	200 to 400 (508 to 1016)	23 to 29	75% argon 25% CO_2	35 cfh	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)

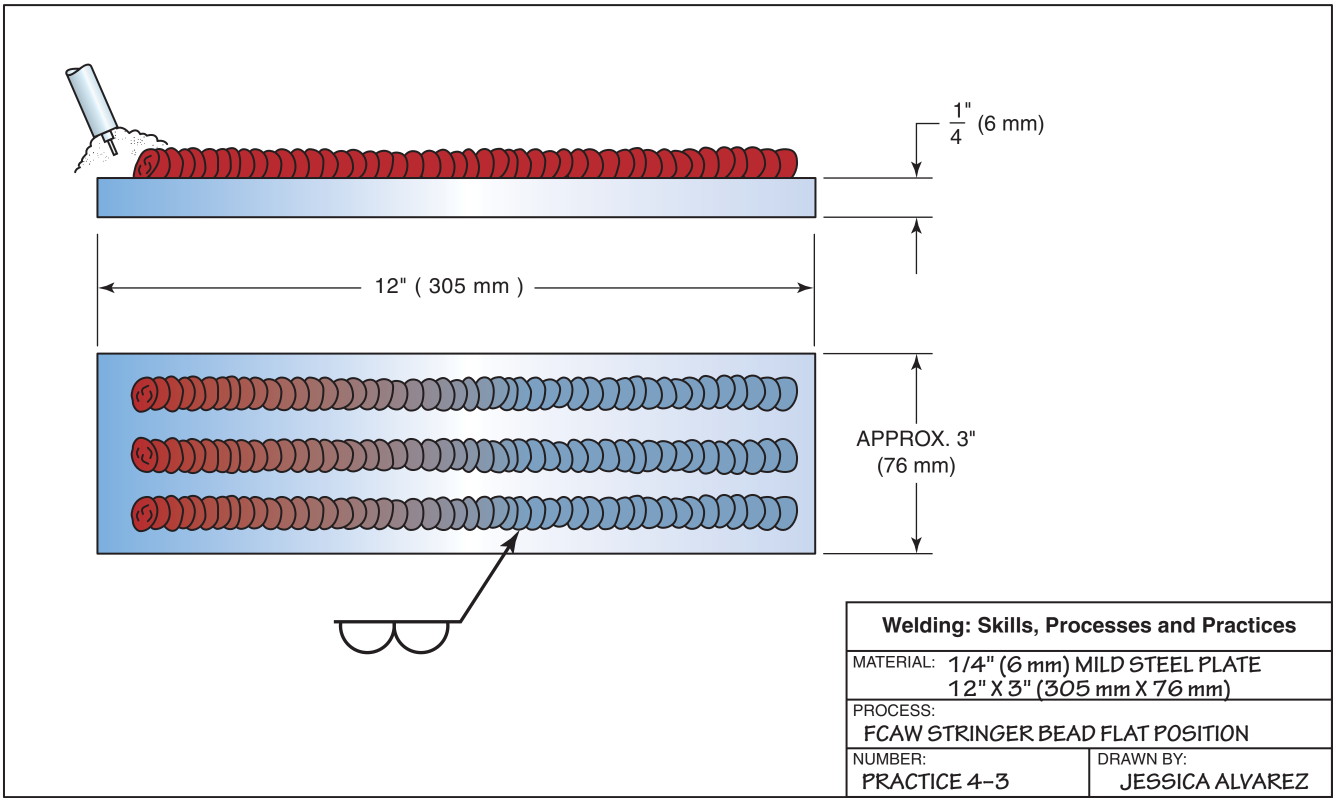

Welding: Skills, Processes and Practices
MATERIAL: 1/4" (6 mm) MILD STEEL PLATE 12" X 3" (305 mm X 76 mm)
PROCESS: FCAW STRINGER BEAD FLAT POSITION
NUMBER: PRACTICE 4–3
DRAWN BY: JESSICA ALVAREZ

Figure 4.12 **FCAW stringer bead, 1/4 in. mild steel, flat position**

Starting at one end of the plate and using a dragging technique, make a weld bead along the entire 12-in. (305-mm) length of the metal. After the weld is complete, check its appearance. Make any needed changes to correct the weld. Repeat the weld and make additional adjustments. After the machine is set, start to work on improving the straightness and uniformity of the weld. Use weave patterns of different widths and straight stringers without weaving.

Repeat with both classifications of electrodes until beads can be made straight, uniform, and free from any visual defects. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

SQUARE-GROOVE WELDS

One advantage of FCA welding is the ability to make 100%-joint-penetrating welds without beveling the edges of the plates. These full-joint-penetrating welds can be made in plates that are 1/4 in. (6 mm) or less in thickness. Welding on thicker plates risks the possibility of a lack of fusion on both sides of the **root face**, Figure 4.13.

There are several disadvantages of having to bevel a plate before welding:

- Beveling the edge of a plate adds an operation to the fabrication process.
- Both more filler metal and welding time are required to fill a beveled joint than are required to make a square jointed weld.
- Beveled joints have more heat from the thermal beveling and additional welding required to fill the groove. The lower heat input to the square joint means less distortion.

The major disadvantage of making square jointed welds is that as the plate thickness approaches 1/4 in. (6 mm) or if the weld is out of

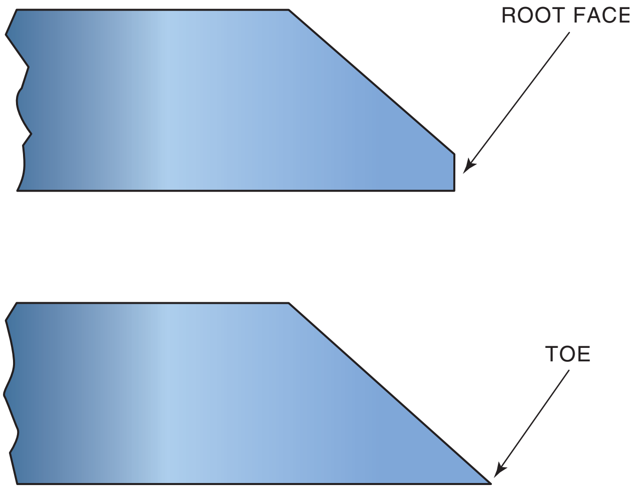

Figure 4.13 A beveled joint may or may not have a flat surface, called a root face

position, a much higher level of skill is required. The skill required to make quality square welds can be acquired by practicing on thinner metal. It is much easier to make this type of weld in metal 1/8 in. (3 mm) thick and then move up in thickness as your skills improve.

PRACTICE 4-4

Butt Joint 1G

Using a properly set up and adjusted FCA welding machine, proper safety protection, E70T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm), and one or more pieces of mild steel plate, 12 in. (305 mm) long and 1/4 in. (6 mm) or less in thickness, you will make a groove weld in the flat position, Figure 4.14.

- Tack weld the plates together and place them in position to be welded.
- Starting at one end, run a bead along the joint. Watch the molten weld pool and bead for signs that a change in technique may be required.
- Make any needed changes as the weld progresses in order to produce a uniform weld.

Repeat with both classifications of electrodes until defect-free welds can consistently be made in the 1/4-in.-thick (6-mm-thick) plate. Turn

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Flat" position portion of the all-position requirement of 6 and 11.

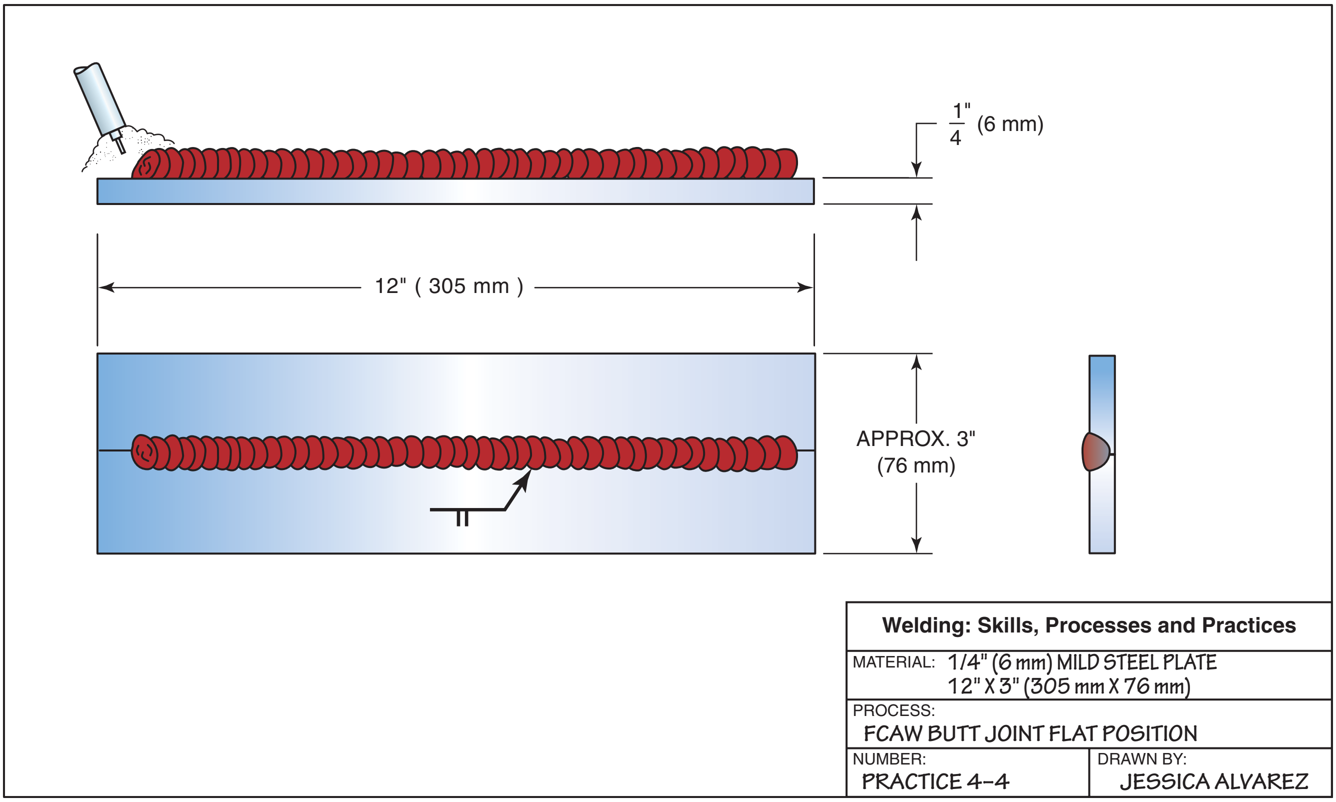

Welding: Skills, Processes and Practices
MATERIAL: 1/4" (6 mm) MILD STEEL PLATE 12" X 3" (305 mm X 76 mm)
PROCESS: FCAW BUTT JOINT FLAT POSITION
NUMBER: PRACTICE 4-4
DRAWN BY: JESSICA ALVAREZ

Figure 4.14 **FCAW butt joint, 1/4 in. mild steel, flat position**

off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

V-GROOVE AND BEVEL-GROOVE WELDS

Although for speed and economy engineers try to avoid specifying welds that require beveling the edges of plates, it is not always possible. Anytime the metal being welded is thicker than 1/4 in. (6 mm) and a 100% joint penetration weld is required, the edges of the plate must be prepared with a bevel. Fortunately, FCA welding allows a narrower groove to be made and still achieve a thorough thickness weld, due to the deeper penetration characteristics of the FCAW process, Figure 4.15.

All FCA groove welds are made using three different types of weld passes, Figure 4.16.

- *Root pass:* The first weld bead of a multiple-pass weld. The root pass fuses the two parts together and establishes the depth of weld metal penetration.
- *Filler pass:* Made after the root pass is completed and used to fill the groove with weld metal. More than one pass is often required.
- *Cover pass:* The last weld pass on a multipass weld. The cover pass may be made with one or more welds. It must be uniform in width, reinforcement, and appearance.

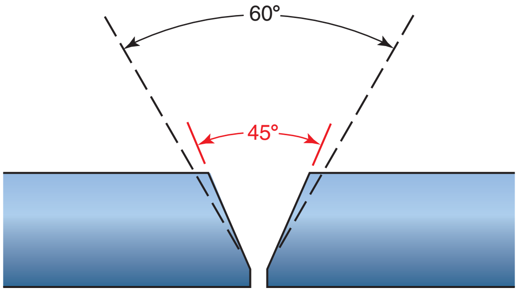

Figure 4.15 **Reduced groove angle for FCAW**
A smaller groove angle reduces both weld time and filler metal required to make the weld.

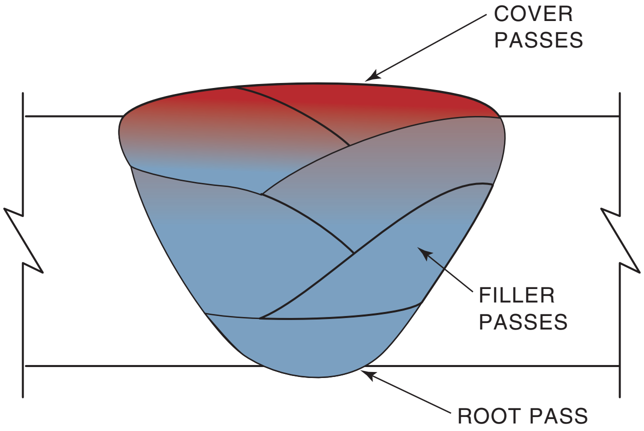

Figure 4.16 **Three types of weld passes make up a weld**

Root Pass

A good root pass is needed in order to obtain a sound weld. The root may be either open or closed using a backing strip, Figure 4.17.

The backing strips are usually made from a piece of metal 1/4 in. (6 mm) thick and 1 in. (25 mm) wide, and should be 2 in. (50 mm) longer than the base plates. The strip is attached to the plate by tack welds made on the sides of the strip, Figure 4.18.

Most production welds do not use backing strips, so they are made as open root welds. Because of the difficulty in controlling root weld face contours in FCAW, however, open-root joints are often avoided on **critical welds**. If an open-root weld is needed because of weldment design, the root pass may be put in with an SMAW electrode or the root face of the FCA weld can be retouched by grinding and/or back welding.

Care must be taken with any root pass not to have the weld face too convex, Figure 4.19. Convex weld faces tend to trap slag along the toe of

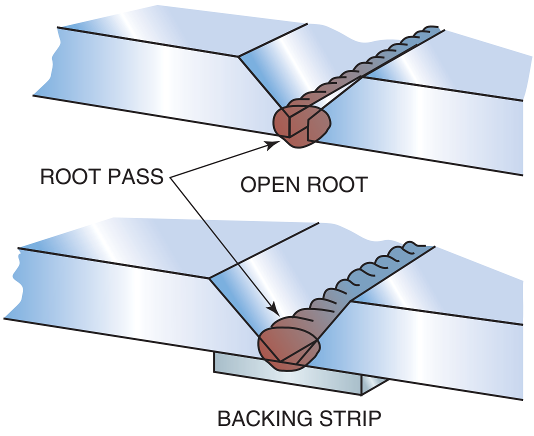

Figure 4.17 Root pass
The maximum deposit for a root pass is 1/4 in. (6 mm) thick.

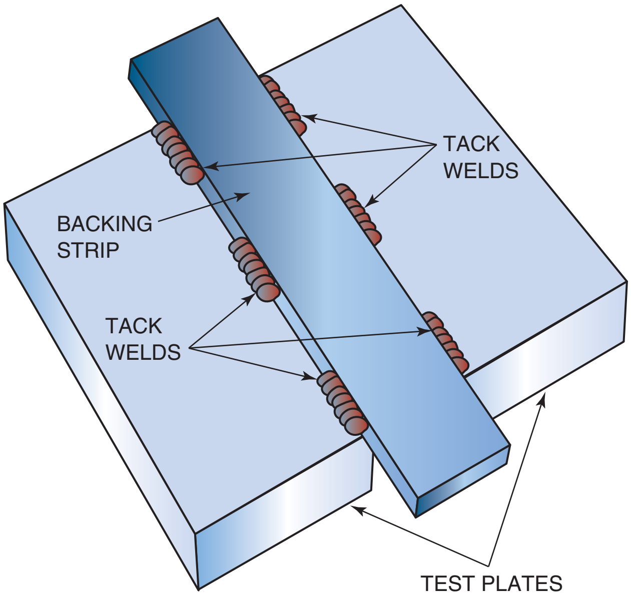

Figure 4.18 Backing strip
Securely tack weld the backing strip to the test plates.

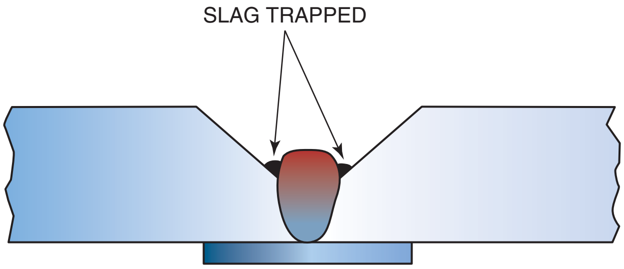

Figure 4.19 Slag trapped beside weld bead is hard to remove

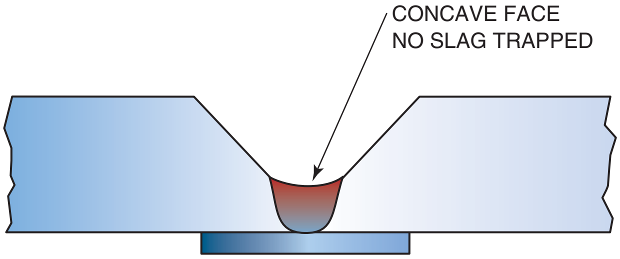

Figure 4.20 Flat or concave weld faces are easier to clean off

the weld. FCA weld slag can be extremely difficult to remove in this area, especially if there is any undercutting. To avoid this, adjust the welding power settings, speed, and weave pattern so that a flat or slightly concave weld face is produced, Figure 4.20.

Filler Pass

Filler passes are made with either **stringer beads**, which are made with a straight progression and very little gun manipulation, or **weave beads**, in which the operator oscillates the gun from side to side in order to widen the weld profile. Either bead type works well for flat or vertically positioned welds, but stringer beads work best for horizontal and overhead-positioned welds. When multiple-pass filler welds are required, each weld bead must overlap the others along the edges. Edges should overlap smoothly enough so that the finished bead is uniform, Figure 4.21. Stringer beads usually overlap about 25% to 50%, and weave beads overlap approximately 10% to 25%.

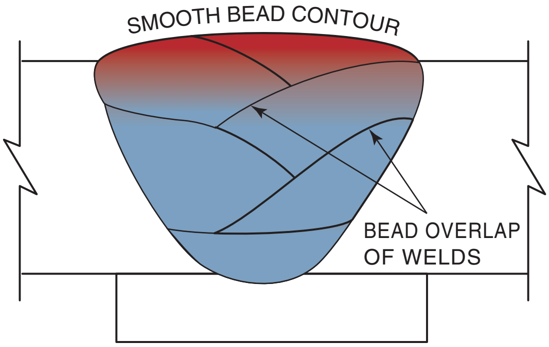

Figure 4.21 The surface of a multipass weld should be as smooth as if it were made by one weld

Each weld bead must be cleaned before the next bead is started. The filler pass ends when the groove has been filled to a level just below the plate surface.

Cover Pass

The cover pass may or may not simply be a continuation of the weld beads used to make the filler pass(es). The major difference between the filler pass and the cover pass is the weld face importance. Keeping the face and toe of the cover pass uniform in width, reinforcement, and appearance and free of defects is essential. Most welds are not tested beyond a visual inspection. For that reason the appearance might be the only factor used for accepting or rejecting welds.

The cover pass must meet a strict visual inspection standard. The visual inspection looks to see that the weld is uniform in width and reinforcement. There should be no arc strikes or hammer marks from chipping or slag removal operations on the plate other than those on the weld itself. The weld must be free of both incomplete fusion and cracks. The weld must be free of overlap, and undercut must not exceed either 10% of the base metal or 1/32 in. (0.8 mm), whichever is less. Reinforcement must have a smooth transition with the base plate and be no higher than 1/8 in. (3 mm), Figure 4.22.

PRACTICE 4-5

Butt Joint 1G

Use a properly set up and adjusted FCA welding machine, Table 4.1; proper safety protection; E70T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick; and a backing strip 14 in. (355 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the flat position, Figure 4.23.

Tack weld the backing strip to the plates. There should be a root gap of approximately 1/8 in. (3 mm) between the plates. The beveled surface can be made with or without a root face, Figure 4.24.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Gas Shielded
Key Indicator 11
This practice addresses the "Flat" position portion of the all-position requirement of 6 and 11.

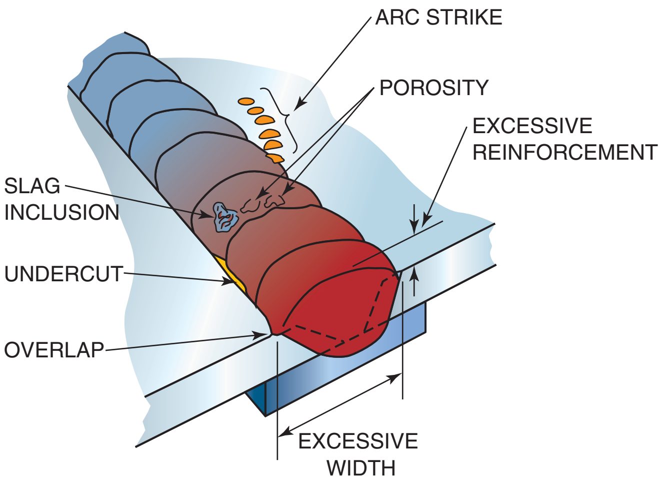

Figure 4.22 Common discontinuities found during a visual examination

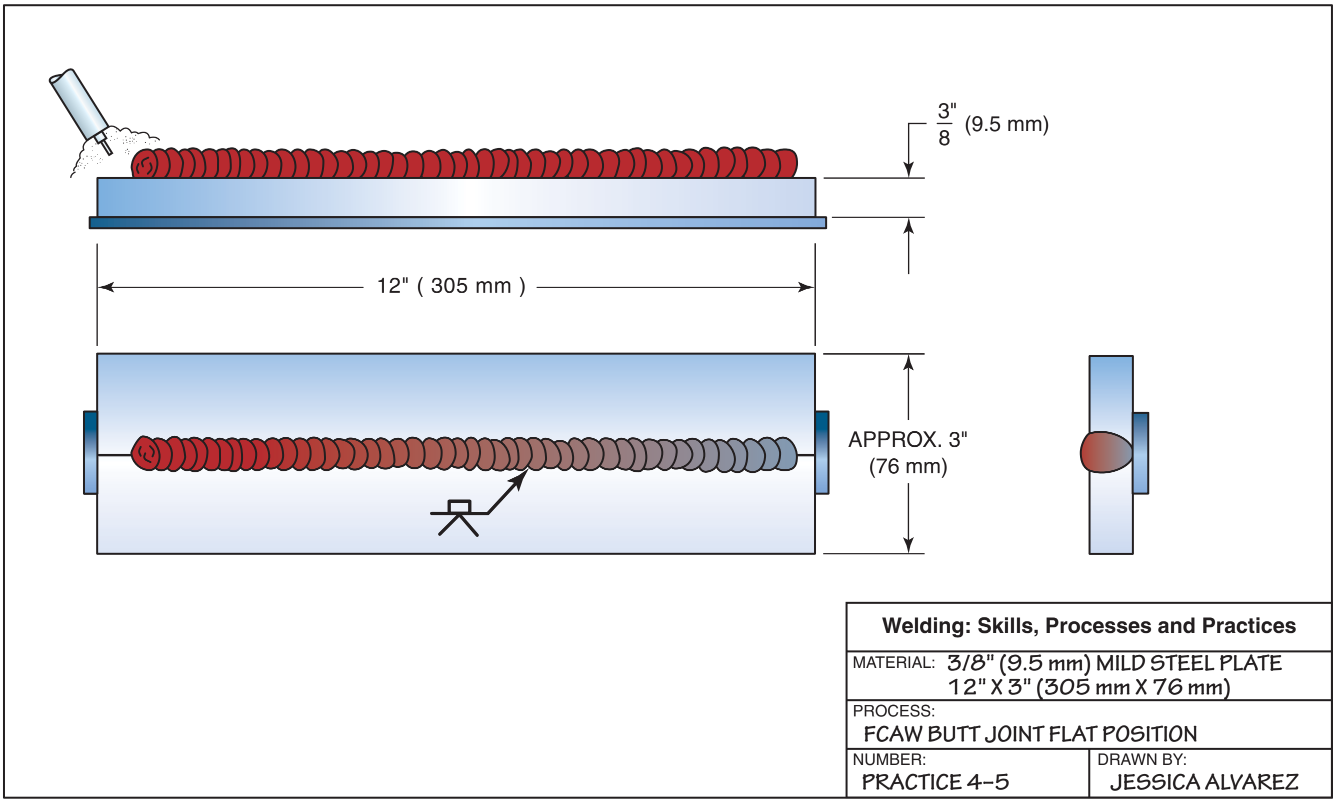
3" / 8 (9.5 mm)
12" (305 mm)
APPROX. 3" (76 mm)
Welding: Skills, Processes and Practices
MATERIAL: 3/8" (9.5 mm) MILD STEEL PLATE 12" X 3" (305 mm X 76 mm)
PROCESS: FCAW BUTT JOINT FLAT POSITION
NUMBER: PRACTICE 4-5
DRAWN BY: JESSICA ALVAREZ

Figure 4.23 FCAW butt joint, 3/8 in. mild steel, flat position

Place the test plates in position at a comfortable height and location. Be sure that you have complete and free movement along the full length of the weld joint. It is often a good idea to make a practice pass along the joint with the welding gun without power to make sure nothing will interfere with your making the weld. Be sure the welding cable is free and will not get caught on anything during the weld.

Start the weld outside the groove on the backing strip tab, Figure 4.25. This is done so that the arc is smooth and the molten weld pool size is

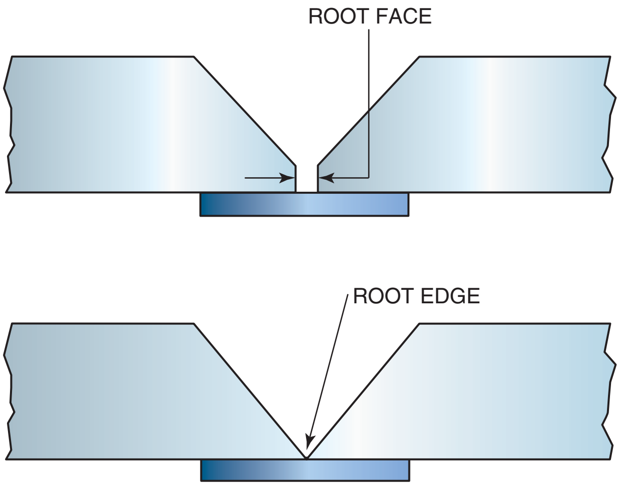
ROOT FACE
ROOT EDGE

Figure 4.24 Groove layout with and without a root face

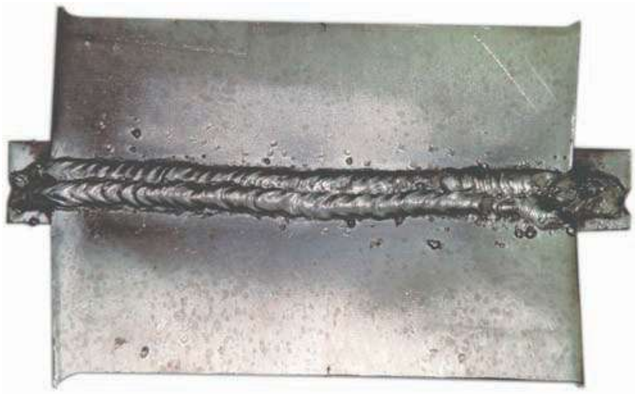

Figure 4.25 Using the ends of a backing strip
Run-off tabs, in which the weld starts and stops on the ends of the backing strip, help control possible underfill or burn-back at the starting and stopping points of a groove weld.
Source: Courtesy of Larry Jeffus

established at the beginning of the groove. Continue the weld out onto the tab at the outer end of the groove. This process ensures that the end of the groove is completely filled with weld.

Repeat with both classifications of electrodes until consistently defect-free welds can be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-6

Butt Joint 1G 100% to Be Tested

Use a properly set up and adjusted FCA welding machine; proper safety protection; E70T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick; and a backing strip 14 in. (355 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the flat position, Figure 4.26.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-5, repeat the weld until you can use each electrode type to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Flat" position portion of the all-position requirement of 6 and 11.

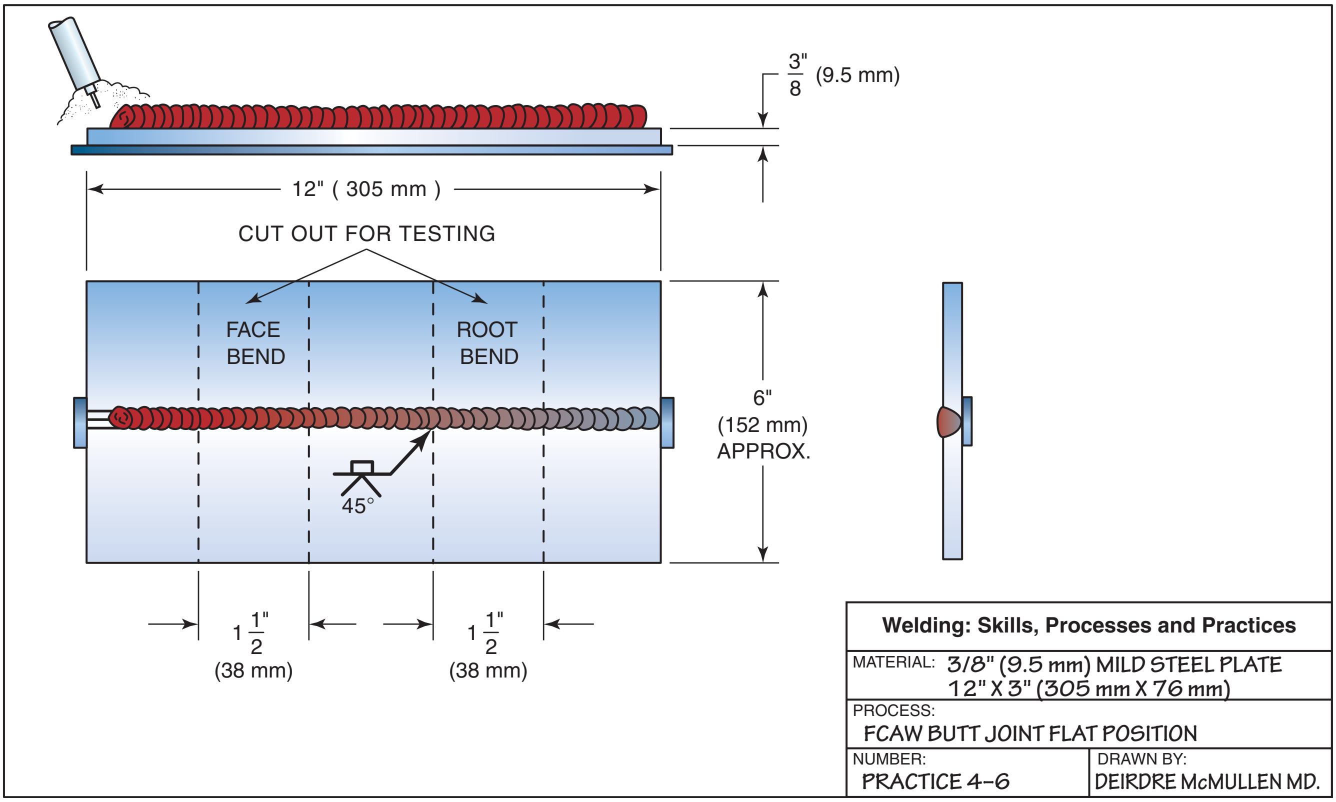

Welding: Skills, Processes and Practices	
MATERIAL: 3/8" (9.5 mm) MILD STEEL PLATE 12" X 3" (305 mm X 76 mm)	
PROCESS: FCAW BUTT JOINT FLAT POSITION	
NUMBER: PRACTICE 4-6	DRAWN BY: DEIRDRE McMULLEN MD.

Figure 4.26 **FCAW butt joint, 3/8 in. mild steel, flat position**

FILLET WELDS

A fillet weld is the type of weld made on a **lap joint** and a **tee joint**. It should be built up equal to the thickness of the plate, Figure 4.27. On thick plates the fillet must be made up of several passes as with a groove weld. The difference with a fillet weld is that a smooth transition from the plate surface to the weld is required. If this transition is abrupt, it can cause stresses that will weaken the joint.

The lap joint is made by overlapping the edges of the plates. They should be held together tightly before tack welding them together. A small tack weld may be added in the center to prevent distortion during welding, Figure 4.28. Chip and wire brush the tacks before you start to weld.

The tee joint is made by tack welding one piece of metal on another piece of metal at a right angle, Figure 4.29. After the joint is tack welded together, the slag is chipped from the tack welds. If the slag is not removed, it will cause a slag inclusion in the final weld.

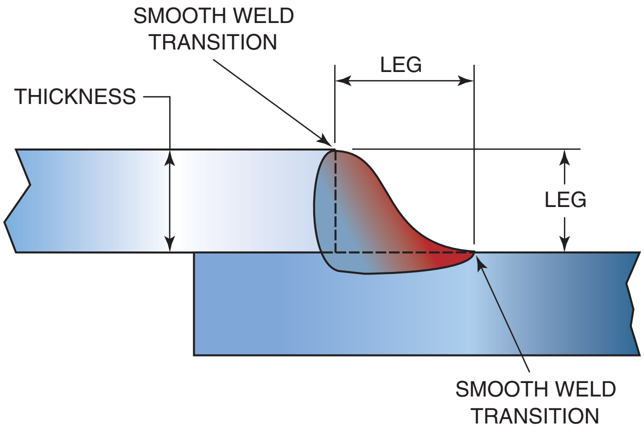

Figure 4.27 The legs of a fillet weld should generally be equal to the thickness of the base metal

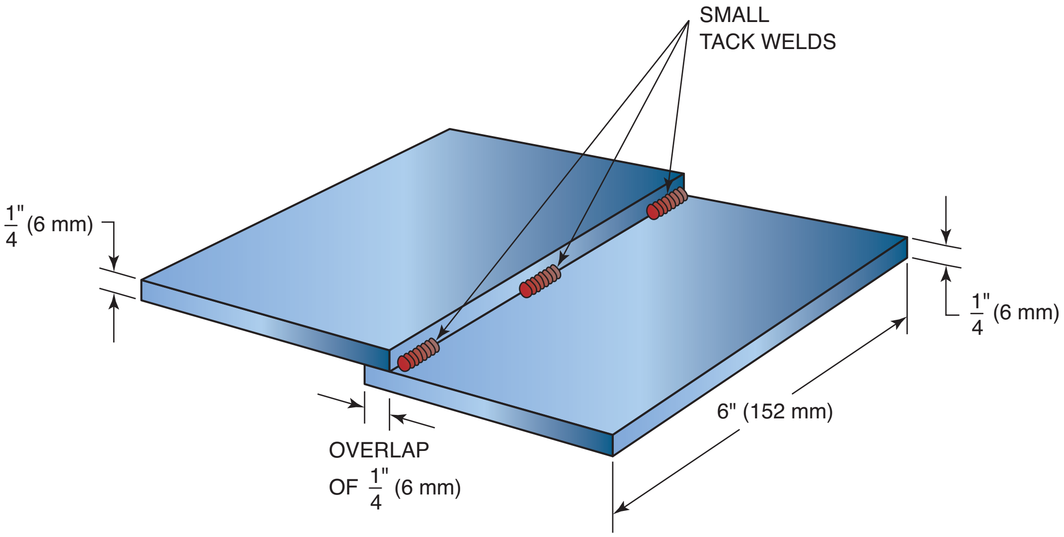

Figure 4.28 Tack welding the plates of a lap joint together

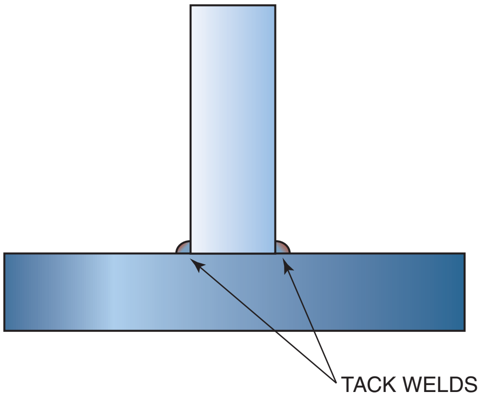

Figure 4.29 Tack welding both sides of a tee joint
Tack welding will help keep the tee square for welding.

Holding thick plates tightly together on tee joints may cause underbead cracking, or lamellar tearing, Figure 4.30. On thick plates the weld shrinkage can be great enough to pull the metal apart well below the bead or its heat-affected zone. In production welds, cracking can be controlled by not assembling the plates tightly together. The space between the two plates can be set by placing a small wire spacer between them, Figure 4.31.

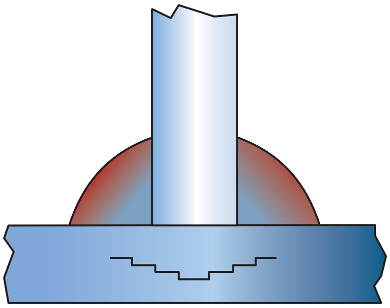

Figure 4.30 Underbead cracking, or lamellar tearing, of the base plate

A fillet welded lap or tee joint can be strong if it is welded on both sides, even without having deep penetration, Figure 4.32. Some tee joints may be prepared for welding by cutting either a bevel or a J-groove in the vertical plate. This cut is not required for strength but may be necessary because of design limitations. Unless otherwise specified, most fillet welds will be equal in size to the plates welded. A fillet weld will be as strong as the base plate if the size of the two welds equals the total thickness of the base plate. The weld bead should have a flat or slightly concave appearance to ensure the greatest strength and efficiency, Figure 4.33.

The root of fillet welds must be melted to ensure a completely fused joint. A notch along the root of the weld pool is an indication that the root is not being fused together, Figure 4.34. To achieve complete root fusion, move the arc to a point as close as possible to the leading edge of the weld pool, Figure 4.35. If the arc strikes the unmelted plate ahead of the molten weld pool, it may become erratic, which will increase weld spatter.

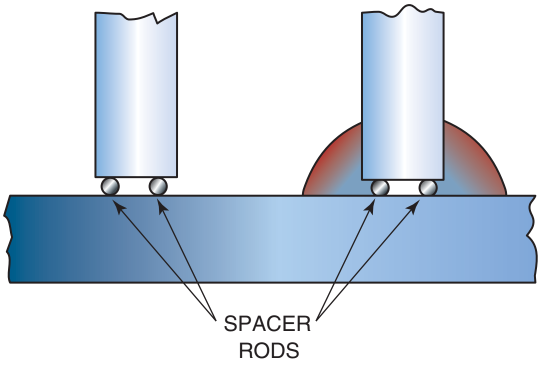

Figure 4.31 Spacers in tee joints
Base plate cracking can be controlled by placing spacers in the joint before welding.

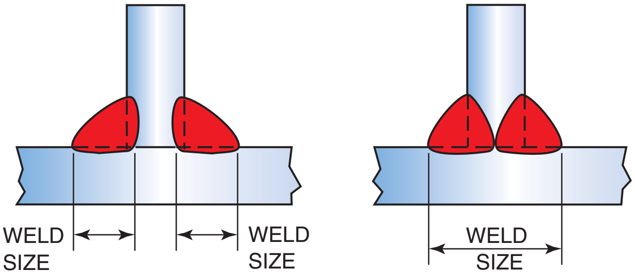

Figure 4.32 Fillet weld size
If the total weld sizes are equal, then both tee joints would have equal strength.

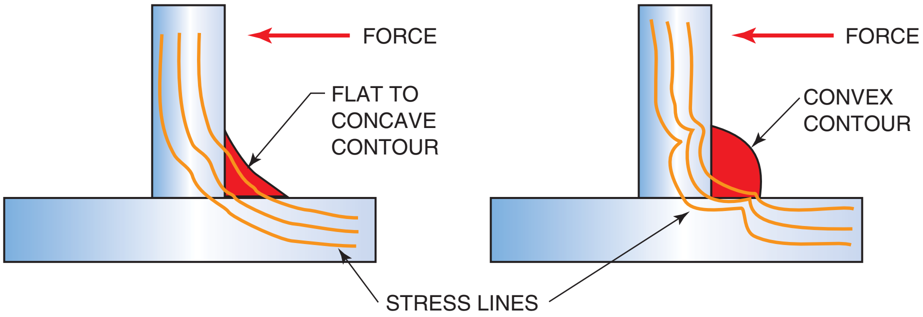

Figure 4.33 **Fillet weld shape**
The stresses are distributed more uniformly through a flat or concave fillet weld.

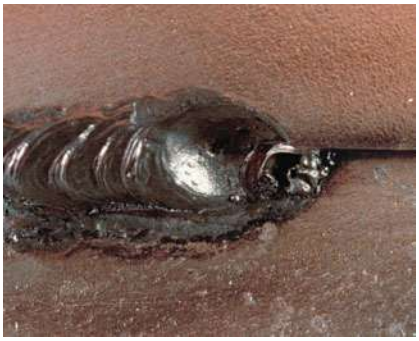

Figure 4.34 **Watch the root of the weld bead to be sure there is complete fusion**
Source: Courtesy of Larry Jeffus

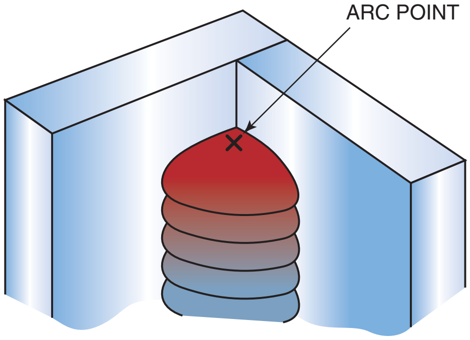

Figure 4.35 **Moving the arc as close as possible to the leading edge of the weld will provide good root fusion.**

PRACTICE 4-7

Lap Joint and Tee Joint 1F

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10
This practice addresses the "Flat" position portion of the all-position requirement for 5 and 10.

Use a properly set up and adjusted FCA welding machine; proper safety protection; E70T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); and one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick. You will make a fillet weld in the flat position.

Tack weld the pieces of metal together and brace them in position. When making the lap or tee joints in the flat position, the plates must be at a 45° angle so that the surface of the weld will be flat, Figure 4.36A and Figure 4.36B. Starting at one end, make a weld along the entire length of the joint.

Repeat each type of joint with both classifications of electrodes until consistently defect-free welds can be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-8

Tee Joint 1F

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10

Use a properly set up and adjusted FCA welding machine, Table 4.2; proper safety protection; E70T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or through 1/16 in. (0.9 mm and/or through 1.6 mm); one

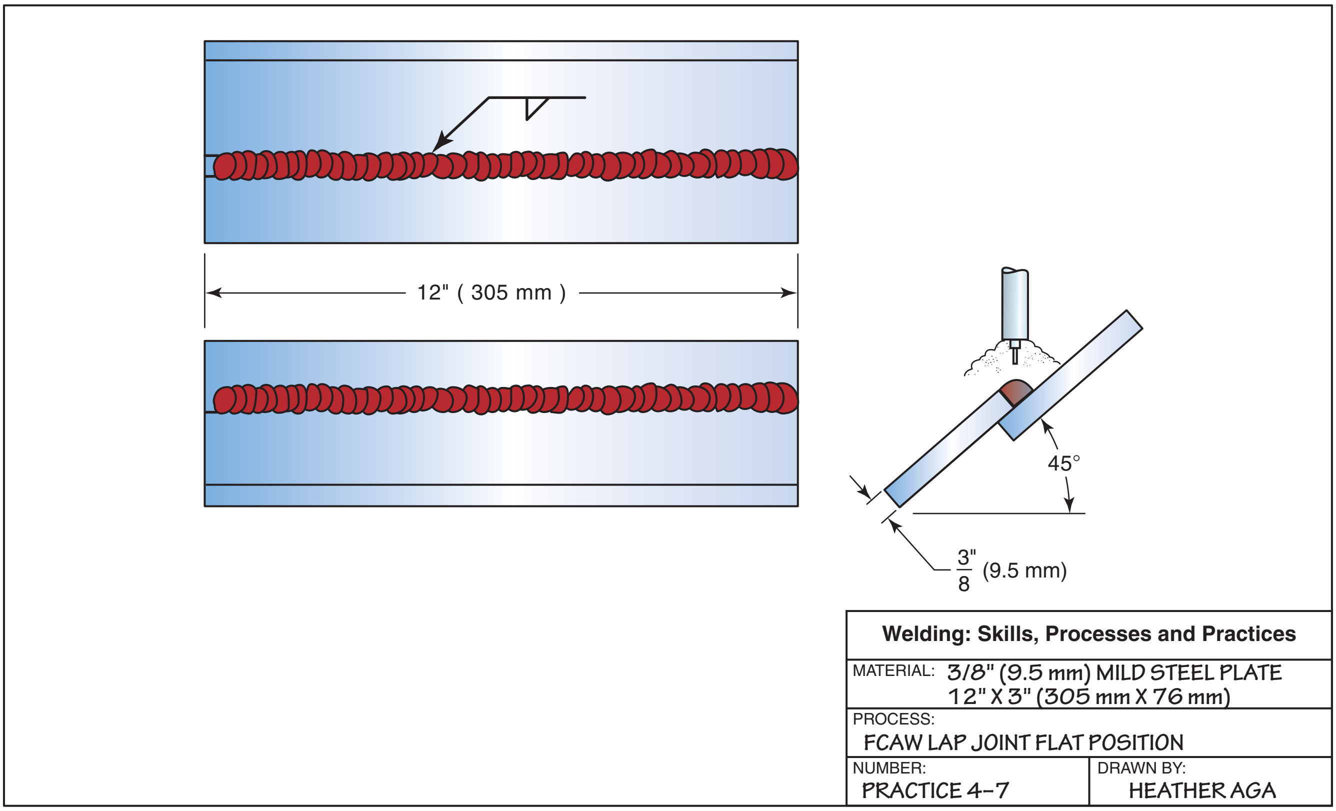
12" (305 mm)
45°
3"/8 (9.5 mm)
Welding: Skills, Processes and Practices
MATERIAL: 3/8" (9.5 mm) MILD STEEL PLATE
12" X 3" (305 mm X 76 mm)
PROCESS:
FCAW LAP JOINT FLAT POSITION
NUMBER:
PRACTICE 4-7
DRAWN BY:
HEATHER AGA

Figure 4.36 (A) FCAW lap joint, 3/8 in. mild steel, flat position

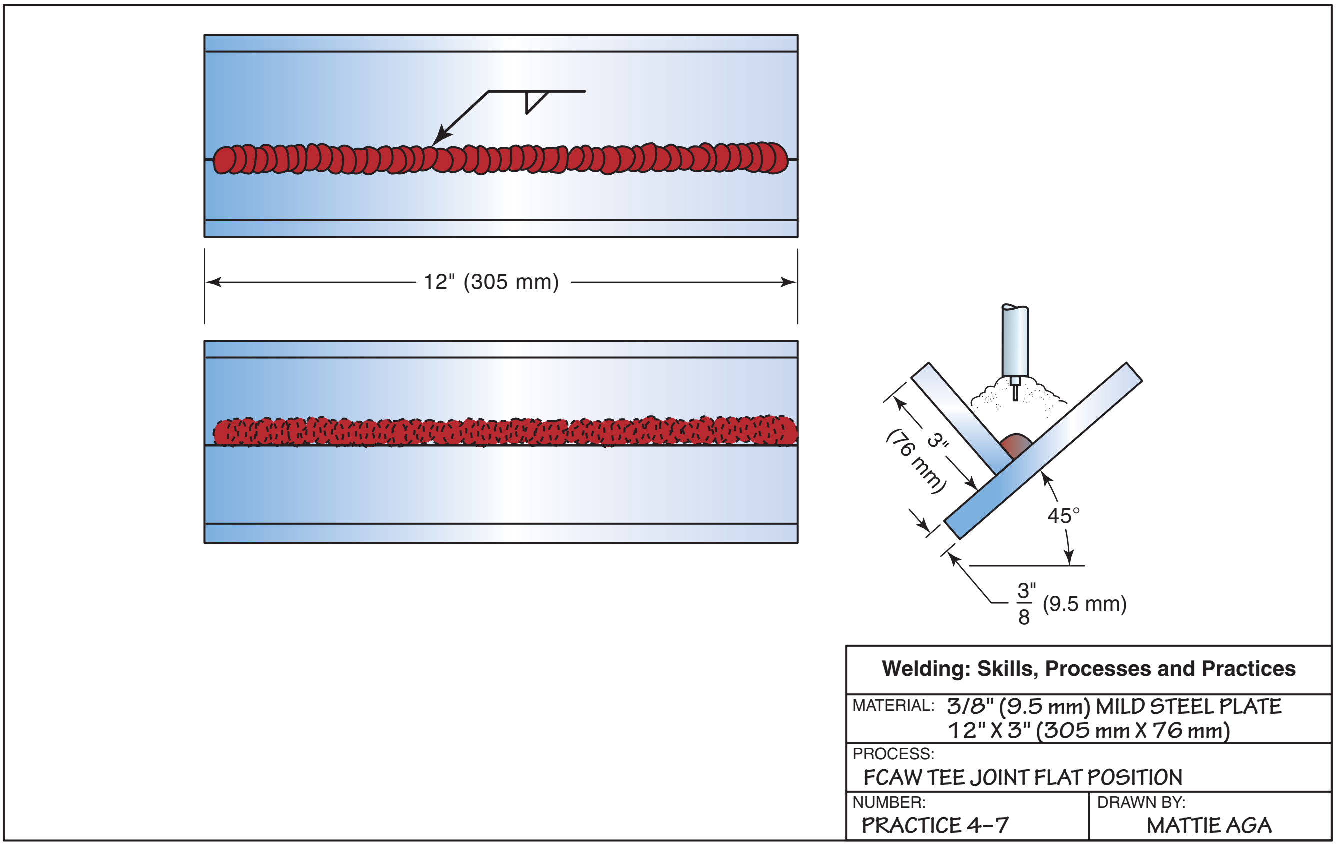
12" (305 mm)
3" (76 mm)
45°
3"/8 (9.5 mm)
Welding: Skills, Processes and Practices
MATERIAL: 3/8" (9.5 mm) MILD STEEL PLATE
12" X 3" (305 mm X 76 mm)
PROCESS:
FCAW TEE JOINT FLAT POSITION
NUMBER:
PRACTICE 4-7
DRAWN BY:
MATTIE AGA

Figure 4.36 (B) FCAW tee joint, 3/8 in. mild steel, flat position

Table 4.2 FCA Welding Parameters for Use if Specific Settings Are Unavailable from Electrode Manufacturer (base metal thickness 1/2 to 3/4 inch)

Electrode		Welding Power			Shielding Gas		Base Metal	
Type	Size	Amps	Wire-feed Speed, ipm (cm/min)	Volts	Type	Flow	Type	Thickness
E70T-1 E71T-1	0.035 in. (0.9 mm)	130 to 150	288 to 380 (732 to 975)	22 to 25	None	n/a	Low-carbon steel	1/2 in. to 3/4 in. (13 mm to 19 mm)
E70T-1 E71T-1	0.045 in. (1.2 mm)	150 to 210	200 to 300 (508 to 762)	28 to 29	None	n/a	Low-carbon steel	1/2 in. to 3/4 in. (13 mm to 19 mm)
E70T-1 E71T-1	.052 in. (1.4 mm)	150 to 300	150 to 350 (381 to 889)	25 to 33	None	n/a	Low-carbon steel	1/2 in. to 3/4 in. (13 mm to 19 mm)
E70T-1 E71T-1	1/16 in. (1.6 mm)	200 to 400	150 to 300 (381 to 762)	27 to 33	None	n/a	Low-carbon steel	1/2 in. to 3/4 in. (13 mm to 19 mm)
E70T-5 E71T-11	0.035 in. (0.9 mm)	130 to 200	288 to 576 (732 to 1463)	20 to 28	75% argon 25% CO_2	30 cfh	Low-carbon steel	1/2 in. to 3/4 in. (13 mm to 19 mm)
E70T-5 E71T-11	0.045 in. (1.2 mm)	150 to 250	200 to 400 (508 to 1016)	23 to 29	75% argon 25% CO_2	35 cfh	Low-carbon steel	1/2 in. to 3/4 in. (13 mm to 19 mm)
E70T-5 E71T-11	0.052 in. (1.4 mm)	150 to 300	150 to 350 (381 to 889)	21 to 32	75% argon 25% CO_2	35 cfh	Low-carbon steel	1/2 in. to 3/4 in. (13 mm to 19 mm)
E70T-5 E71T-11	1/16 in. (1.6 mm)	180 to 400	145 to 350 (368 to 889)	21 to 34	75% argon 25% CO_2	40 cfh	Low-carbon steel	1/2 in. to 3/4 in. (13 mm to 19 mm)

or more pieces of mild steel plate, beveled, 7 in. (178 mm) long and 3/4 in. (19 mm) thick or thicker. You will make a fillet weld in the flat position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-7, repeat each type of joint with both classifications of electrodes until consistently defect-free welds can be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

VERTICAL WELDS

PRACTICE 4-9

Butt Joint at a 45° Vertical Up Angle

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Key Indicator 1, 2
Gas Shielded
Key Indicator 3
Gas Shielded
Key Indicator 4
Self Shielded
Key Indicator 8
Self Shielded
Key Indicator 9

Using a properly set up and adjusted FCA welding machine, proper safety protection, 0.035-in. and/or 0.045-in. (0.9-mm and/or 1.2-mm)-diameter E71T-1 and/or E71T-11 electrodes, and one or more pieces of mild steel plate, 12 in. (305 mm) long and 1/4 in. (6 mm) thick or thinner, you will increase the plate angle gradually as you develop skill until you are making satisfactory welds in the vertical up position, Figure 4.37.

- Start practicing this weld with the plate at a 45° angle.
- Gradually increase the angle of the plate to vertical as skill is gained in welding this joint. A straight stringer bead or slight zigzag will work well on this joint.
- Establish a molten weld pool in the root of the joint.
- Cool, chip, and inspect the weld for uniformity and defects.

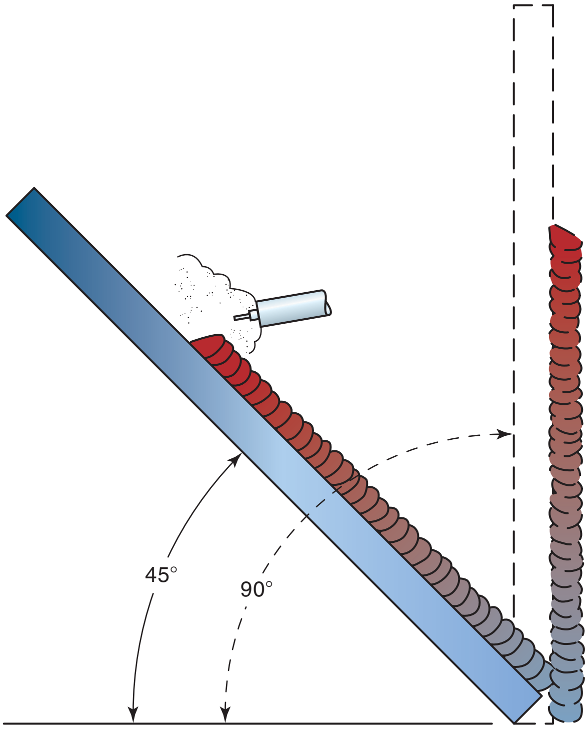

Figure 4.37 Start making welds with the plate at a 45° angle
As your skill develops, increase the angle until the plate is vertical.

It is easier to make a quality weld in the vertical up position if lower settings are used in both the **amperage range** and **voltage range**. This will make the molten weld pool smaller, less fluid, and easier to control. A problem with lower power settings is that the weld bead often can be very convex, Figure 4.38. Faster travel speed and/or slightly wider weave patterns can be used to control the bead shape.

Start at the bottom of the plate and hold the welding gun at a slight upward angle to the plate, Figure 4.39. Brace yourself, lower your hood, and begin to weld. Depending on the machine settings and type of electrode used, you will make a weave pattern.

If the molten weld pool is large and fluid (hot), use a C or J weave pattern to allow a longer time for the molten weld pool to cool, Figure 4.40. Do not make the weave so long or fast that the electrode is allowed to strike the metal ahead of the molten weld pool. If this happens, spatter increases and a spot or zone of incomplete fusion may occur.

A weld that is high and has little or no fusion is too "cold." Changing the welding technique will not correct this problem. The welder must stop welding and make the needed adjustments to the power supply or electrode feeder. Continue to weld along the entire 12-in. (305-mm) length of plate.

Repeat welds with both electrodes until defect-free welds can be consistently made vertically in the 1/4-in. (6-mm)-thick plate. Turn off the

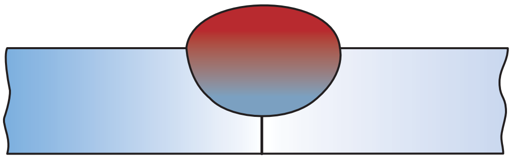

Figure 4.38 Low amperage causes too much buildup and not enough penetration

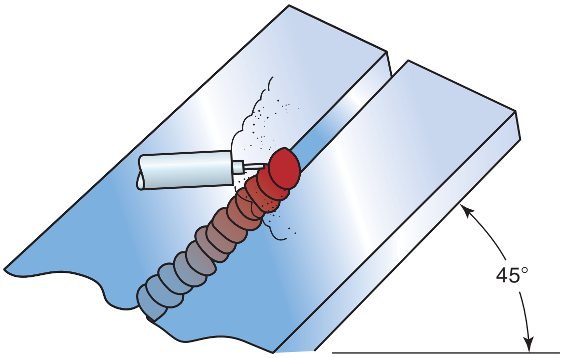

Figure 4.39 45° vertical up

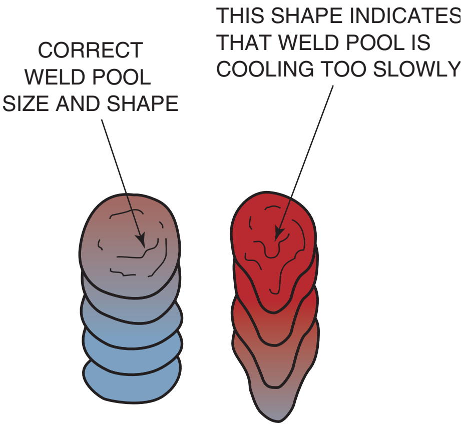
CORRECT
WELD POOL
SIZE AND SHAPE
THIS SHAPE INDICATES
THAT WELD POOL IS
COOLING TOO SLOWLY

Figure 4.40 The shape of the weld pool can indicate the temperature of the surrounding base metal

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Vertical" portion of the all-position requirement.

welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-10

Butt Joint 3G

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm), and one or more pieces of mild

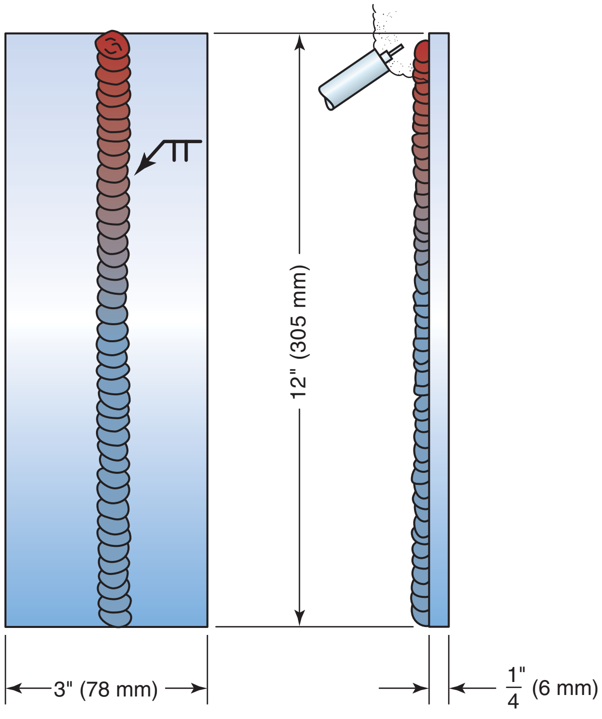
12" (305 mm)
3" (78 mm)
1"/4 (6 mm)

Welding: Skills, Processes and Practices
MATERIAL: 1/4" (6 mm) MILD STEEL PLATE
12" X 3" (305 mm X 76 mm)
PROCESS: FCAW 3G BUTT JOINT
NUMBER: PRACTICE 4-10
DRAWN BY: AMY JEFFUS

Figure 4.41 FCAW 3G butt joint, 1/4 in. mild steel

steel plate, 12 in. (305 mm) long and 1/4 in. (6 mm) thick or thinner, you will make a groove weld in the vertical position, Figure 4.41.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-9, repeat with both classifications of electrodes until defect-free welds can be consistently made in the 1/4-in. (6-mm)-thick plate. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

PRACTICE 4-11

Butt Joint 3G

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick; and a backing strip 14 in. (355 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the vertical position.

Following the same instructions for assembly and welding procedure as outlined in Practice 4-9, repeat with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

PRACTICE 4-12

Butt Joint 3G 100% to Be Tested

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2
This practice addresses the “Vertical” component of the all-position requirement of 6 and 11.

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick; and a backing strip 14 in. (355 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the vertical position, Figure 4.42.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-9, repeat the weld until you can use each electrode type to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

PRACTICE 4-13

Butt Joint 3G 100% to Be Tested

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2
This practice addresses the “Vertical” portion of the all-position requirement of 6 and 11.

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.045 in. and/or through 1/16 in. (0.9 mm and/or through 1.6 mm); one or more pieces of mild steel plate, beveled, 7 in. (178 mm) long and

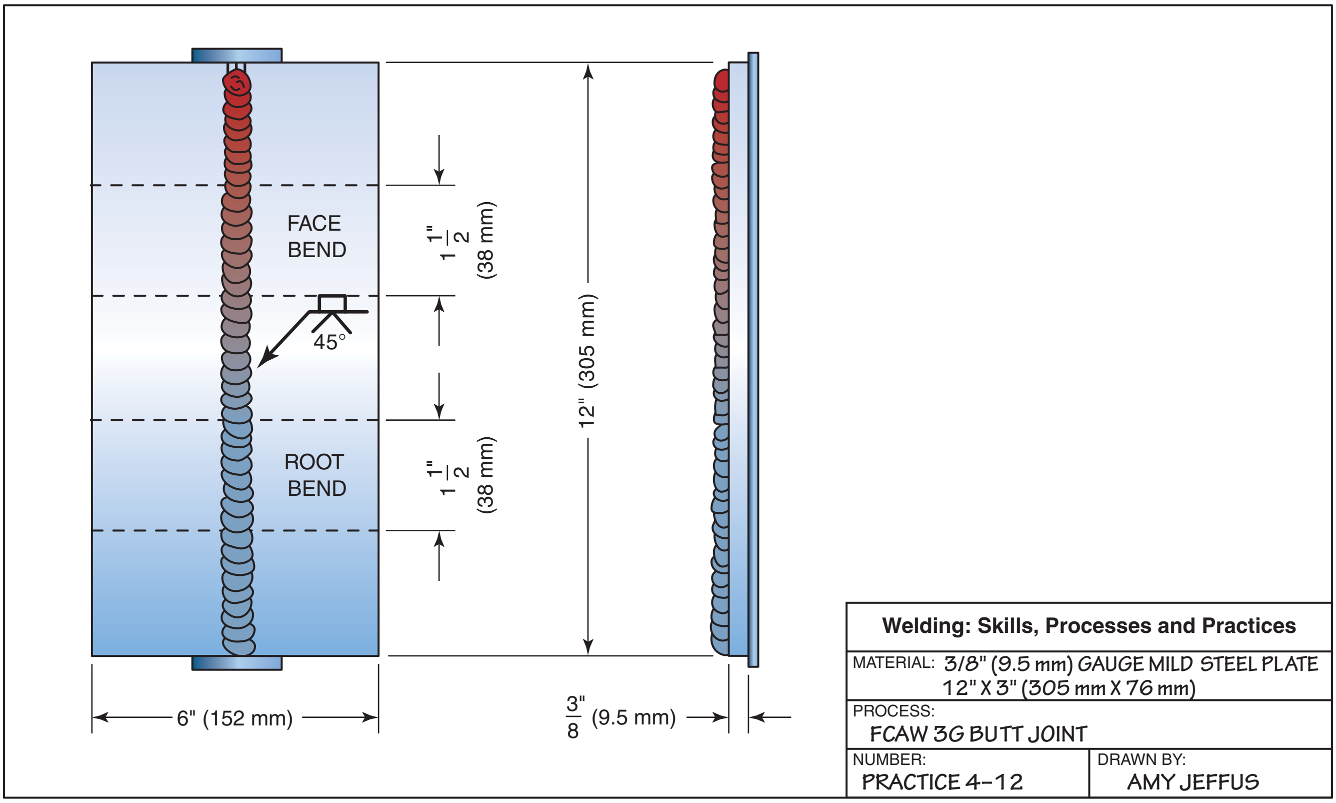
FACE
BEND
45°
ROOT
BEND
1 1/2" (38 mm)
1 1/2" (38 mm)
12" (305 mm)
6" (152 mm)
3/8" (9.5 mm)
Welding: Skills, Processes and Practices
MATERIAL: 3/8" (9.5 mm) GAUGE MILD STEEL PLATE
12" X 3" (305 mm X 76 mm)
PROCESS:
FCAW 3G BUTT JOINT
NUMBER:
PRACTICE 4-12
DRAWN BY:
AMY JEFFUS

Figure 4.42 FCAW 3G butt joint, 3/8 in. mild steel

3/4 in. (19 mm) thick or thicker; and a backing strip 9 in. (230 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the vertical position, Figure 4.43.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Key Indicator 1, 2
Gas Shielded
Key Indicator 3
Gas Shielded
Key Indicator 4
Self Shielded
Key Indicator 8
Self Shielded
Key Indicator 9
Self Shielded
Key Indicator 10
Self Shielded
Key Indicator 11
Self Shielded
Key Indicator 12

Module 9
Key Indicator 1, 2

Following the same instructions for the assembly and welding procedure outlined in Practice 4-9, repeat the weld until you can use each electrode type to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

PRACTICE 4-14

Fillet Weld Joint at a 45° Vertical Up Angle

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm), and one or more pieces of mild steel plate, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick, you will increase the plate angle gradually as you develop skill until you are making satisfactory welds in the vertical up position, Figure 4.44.

Tack weld the metal pieces together and brace them in position. Check to see that you have free movement for your gun along the entire joint to prevent stopping and restarting during the weld. Avoiding stops and starts both speeds up the welding time and eliminates discontinuities.

It is easier to make a quality weld in the vertical up position if both the amperage and voltage are set at the lower end of their ranges. This

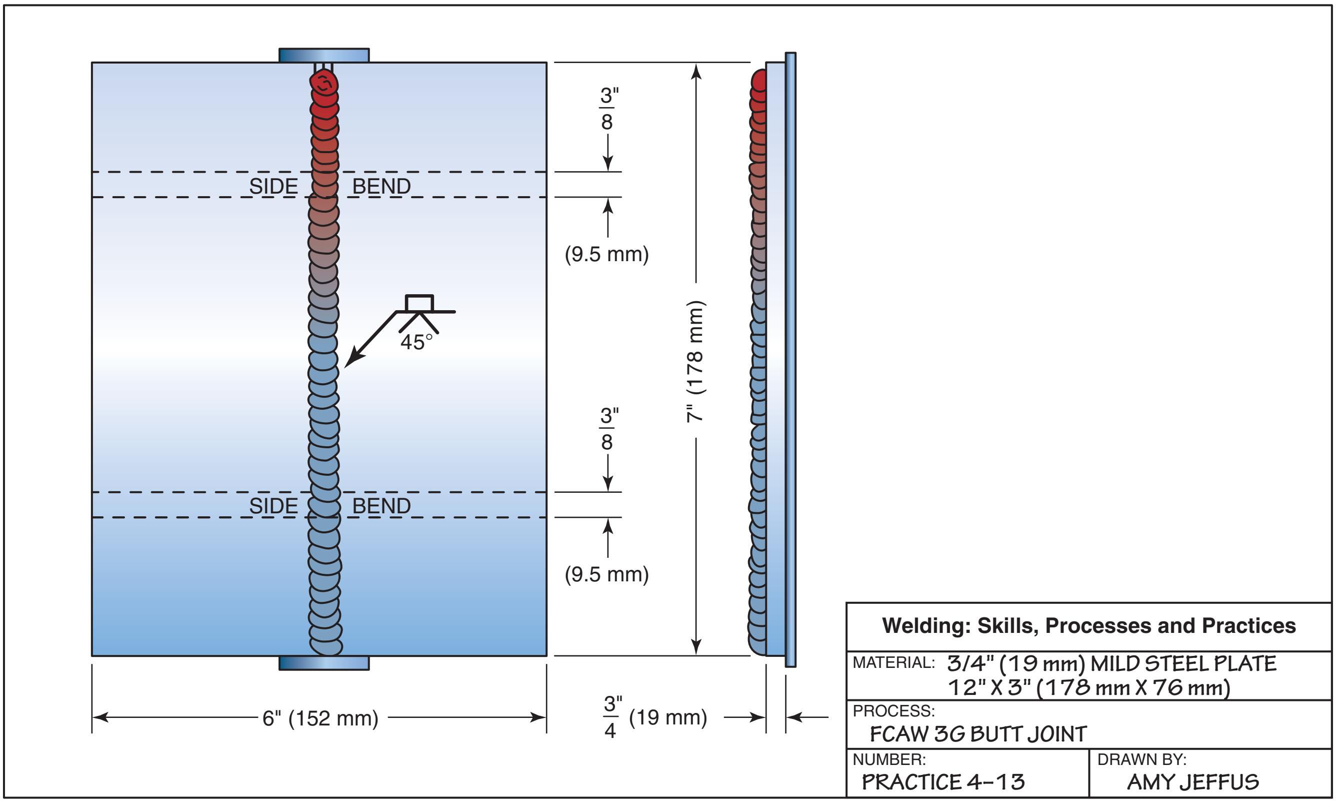
SIDE BEND
SIDE BEND
3"/8
(9.5 mm)
3"/8
(9.5 mm)
45°
7" (178 mm)
6" (152 mm)
3"/4 (19 mm)
Welding: Skills, Processes and Practices
MATERIAL: 3/4" (19 mm) MILD STEEL PLATE 12" X 3" (178 mm X 76 mm)
PROCESS: FCAW 3G BUTT JOINT
NUMBER: PRACTICE 4-13
DRAWN BY: AMY JEFFUS

Figure 4.43 FCAW 3G butt joint, 3/4 in. mild steel

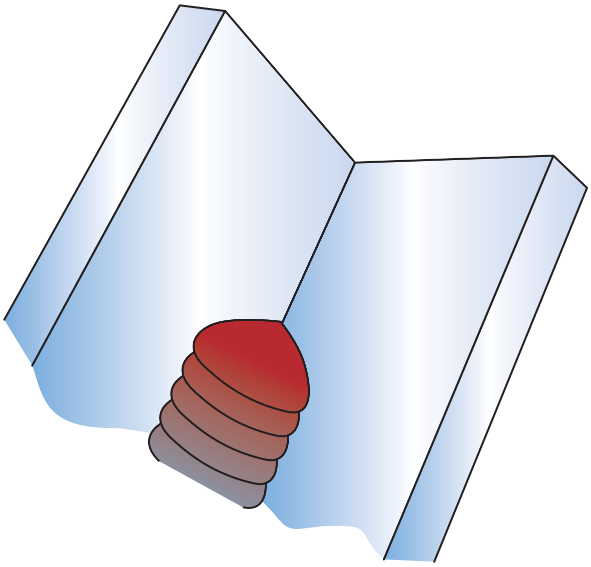

Figure 4.44 45° vertical up fillet weld

will make the molten weld pool smaller, less fluid, and easier to control. A problem with the lower power settings is that the weld bead often is very convex. A convex face on a weld bead often makes it more difficult to remove the slag along the toe of the weld.

The weave pattern should allow for adequate fusion on both edges of the joint. Watch the edges to be sure that they are being melted so that adequate fusion and penetration occur.

Repeat the weld with each electrode type until defect-free welds can consistently be made vertically. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10

Module 9
Key Indicator 1, 2
This practice addresses the "Vertical" position component of the all-position requirement.

PRACTICE 4-15

Lap Joint and Tee Joint 3F 100% to Be Tested

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); and one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick. You will make a fillet weld in the vertical position.

Following the same instructions for assembly and welding procedure as outlined in Practice 4-14, repeat each type of joint with both classifications of electrodes until you can make welds with 100% penetration that will pass the test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10
This practice addresses the "Vertical" component of the all-position requirement for 5 and 10.

PRACTICE 4-16

Tee Joint 3F

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-11 electrodes of diameter 0.045 in. and/or through 1/16 in. (0.9 mm and/or through 1.6 mm), and one or more pieces of mild steel plate, 7 in. (178 mm) long and 3/4 in. (19 mm) thick, you will make a fillet weld in the vertical position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-14, repeat each type of joint with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10
This practice addresses the "Horizontal" component of the all-position requirement for 5 and 10.

HORIZONTAL WELDS

PRACTICE 4-17

Lap Joint and Tee Joint 2F

Use a properly set up and adjusted FCA welding machine; proper safety protection; E70T-1 and/or E71T-11 (or E71T-1 and/or E70T-5) electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); and one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick. You will make a fillet weld in the horizontal position, Figure 4.45A and Figure 4.45B.

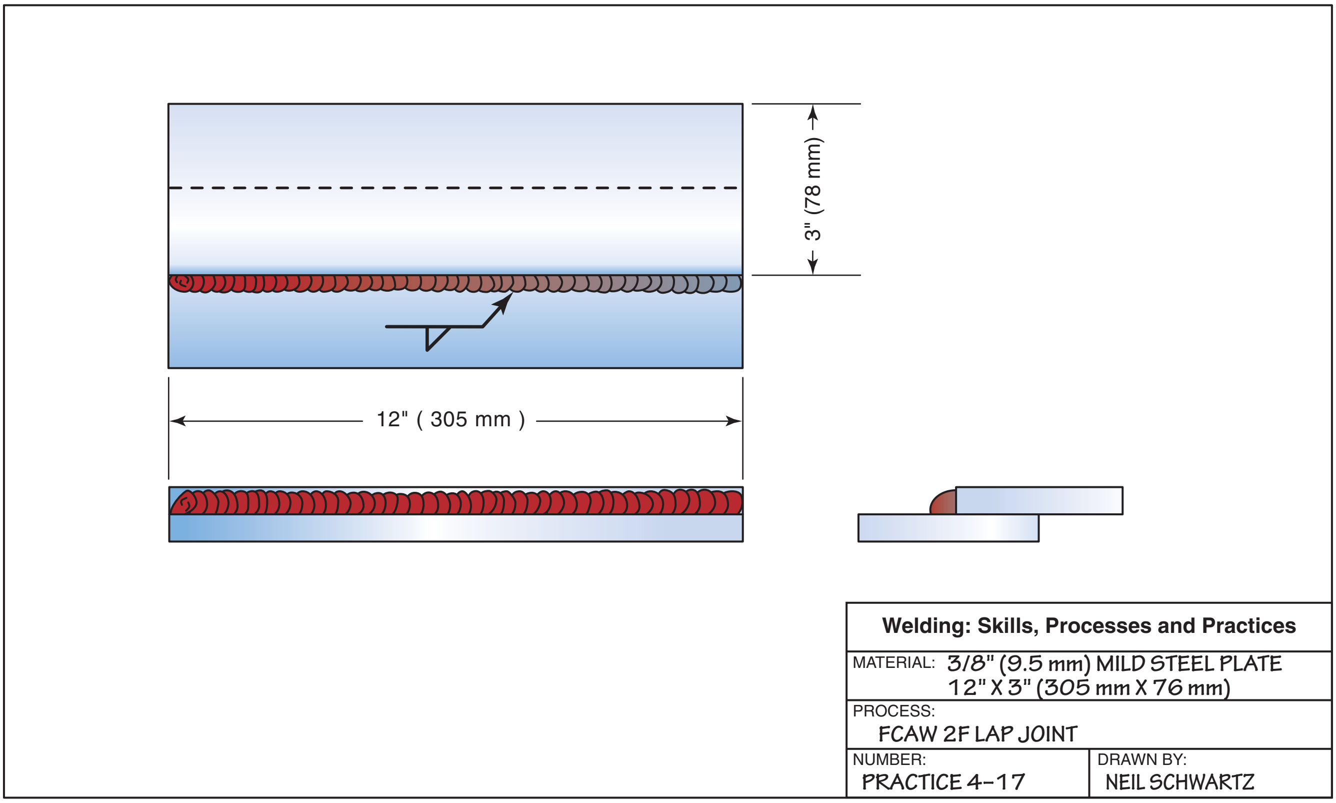
3" (78 mm)
12" (305 mm)
Welding: Skills, Processes and Practices
MATERIAL: 3/8" (9.5 mm) MILD STEEL PLATE
12" X 3" (305 mm X 76 mm)
PROCESS:
FCAW 2F LAP JOINT
NUMBER:
PRACTICE 4-17
DRAWN BY:
NEIL SCHWARTZ

Figure 4.45 **(A) FCAW 2F lap joint, 3/8 in. mild steel**

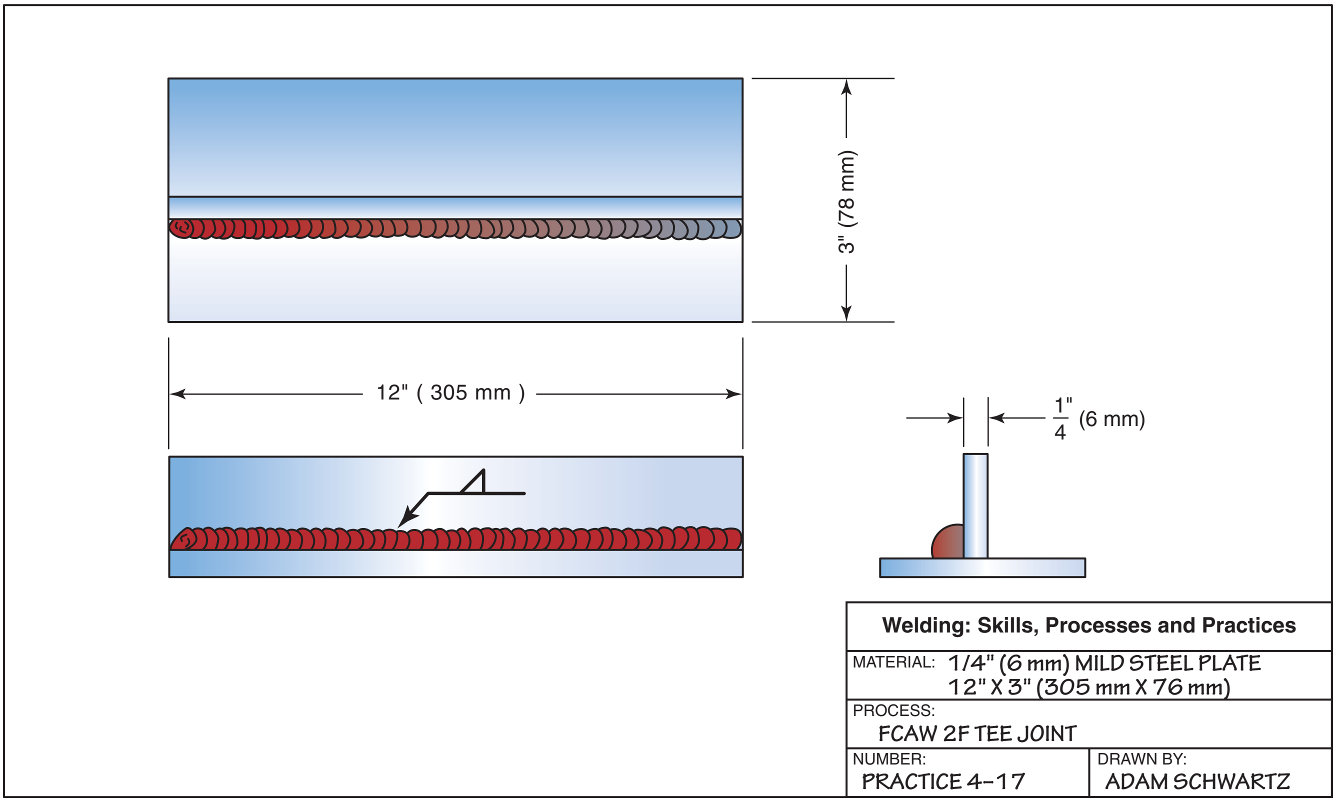
3" (78 mm)
12" (305 mm)
1"/4 (6 mm)
Welding: Skills, Processes and Practices
MATERIAL: 1/4" (6 mm) MILD STEEL PLATE
12" X 3" (305 mm X 76 mm)
PROCESS:
FCAW 2F TEE JOINT
NUMBER:
PRACTICE 4-17
DRAWN BY:
ADAM SCHWARTZ

Figure 4.45 **(B) FCAW 2F tee joint, 1/4 in. mild steel**

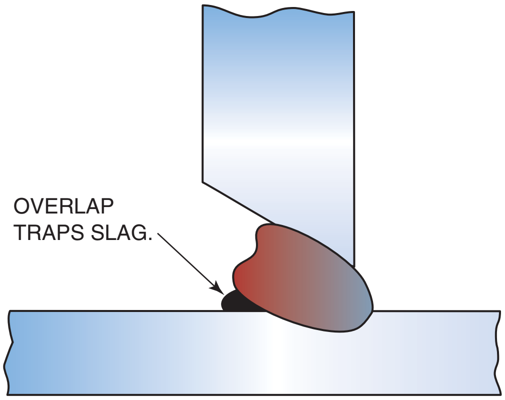

Figure 4.46 Slag can be trapped along the side of the root pass

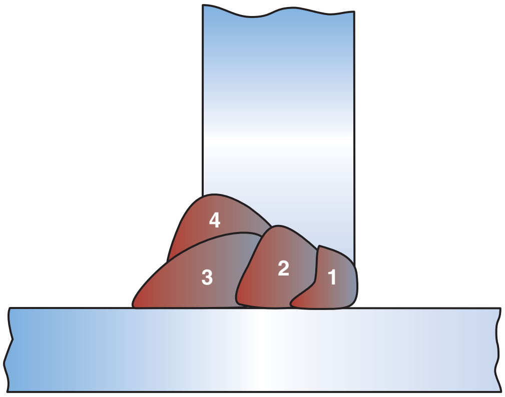

Figure 4.47 FCAW weld bead positions for a 100% penetration grooved tee joint

The root weld must be kept small so that its contour can be controlled. Too large a root pass can trap slag under overlap along the lower edge of the weld, Figure 4.46. Clean each pass thoroughly before the weld bead is started. Follow the weld bead sequence shown in Figure 4.47. Use stringer beads rather than weave beads in order to eliminate overlap. Keeping all of the weld beads small will help control their contour.

Repeat each type of joint with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10
This practice addresses the "Horizontal" component of the all-position requirement for 5 and 10.

PRACTICE 4-18

Tee Joint 2F

Use a properly set up and adjusted FCA welding machine, proper safety protection, E70T-1 and/or E71T-11 (or E71T-1 and/or E70T-5) electrodes of diameter 0.045 in. and/or through 1/16 in. (0.9 mm and/or through 1.6 mm), and one or more pieces of mild steel, beveled, 7 in. (178 mm) long and 3/4 in. (19 mm) thick or thicker. You will make a fillet weld in the horizontal position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-17, repeat each type of joint with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Horizontal" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-19

Butt Joint 2G

Use a properly set up and adjusted FCA welding machine, proper safety protection, E70T-1 and/or E71T-11 (or E71T-1 and/or E70T-5 electrodes) of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm),

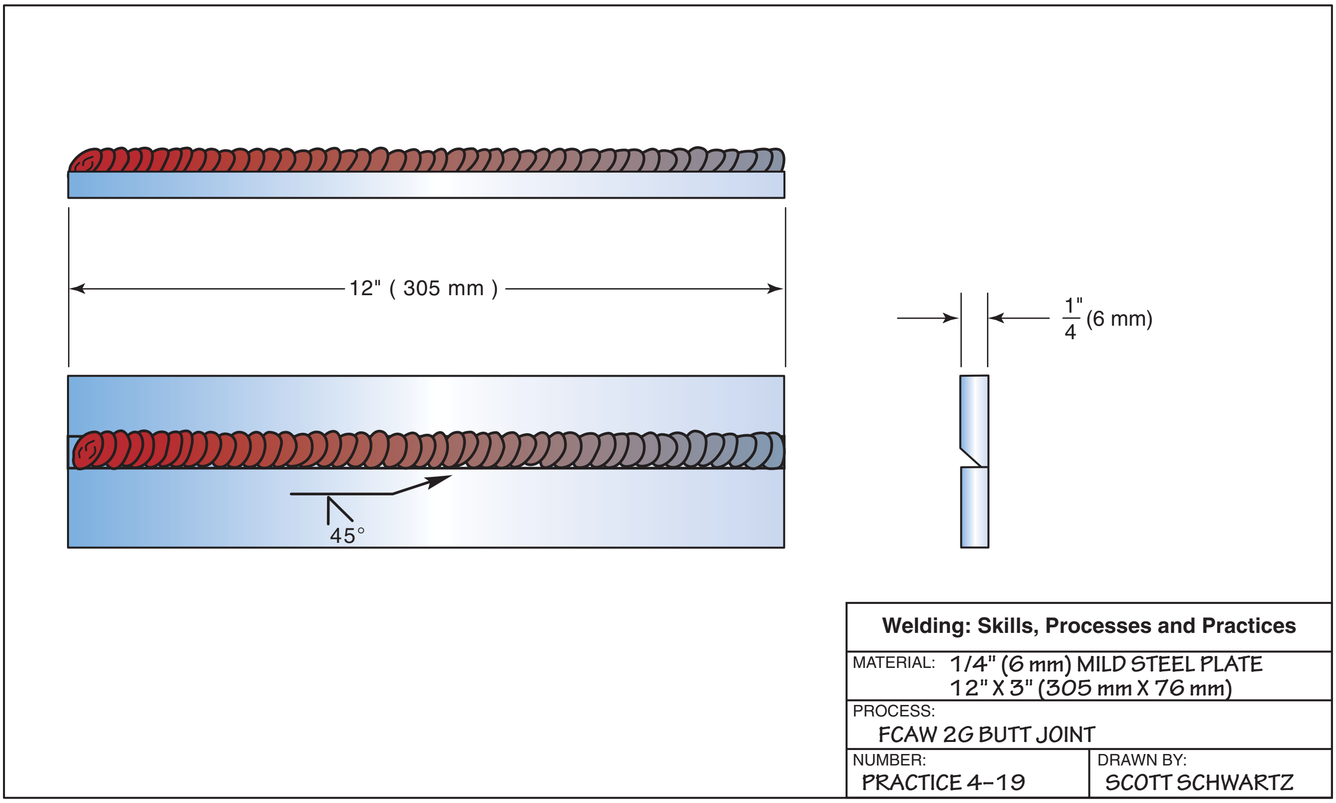
12" (305 mm)
1"/4 (6 mm)
45°
Welding: Skills, Processes and Practices
MATERIAL: 1/4" (6 mm) MILD STEEL PLATE
12" X 3" (305 mm X 76 mm)
PROCESS:
FCAW 2G BUTT JOINT
NUMBER:
PRACTICE 4–19
DRAWN BY:
SCOTT SCHWARTZ

Figure 4.48 **FCAW 2G butt joint, 1/4 in. mild steel**

and one or more pieces of mild steel plate, 12 in. (305 mm) long and 1/4 in. (6 mm) thick or thinner. You will make a groove weld in the horizontal position, Figure 4.48.

Repeat with both classifications of electrodes until defect-free welds can consistently be made in the 1/4-in. (6-mm)-thick plate. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

PRACTICE 4-20

Butt Joint 2G

Use a properly set up and adjusted FCA welding machine; proper safety protection; E70T-1 and/or E71T-11 or 5 (or E71T-1 and/or E70T-5) electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick; and a backing strip 14 in. (355 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the horizontal position.

Repeat with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the “Horizontal” position portion of the all-position requirement of 6 and 11.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2
This practice addresses the "Horizontal" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-21

Butt Joint 2G 100% to Be Tested

Use a properly set up and adjusted FCA welding machine; proper safety protection; E70T-1 and/or E71T-11 or 5 (or E71T-1 and/or E70T-5) electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick; and a backing strip 14 in. (355 mm) long, 1 in. (25 mm) wide and 1/4 in. (6 mm) thick. You will make a groove weld in the horizontal position.

Repeat the weld using each electrode classification until you can make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Self Shielded
Key Indicator 6
Gas Shielded
Key Indicator 11
This practice addresses the "Horizontal" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-22

Butt Joint 2G

Use a properly set up and adjusted FCA welding machine; proper safety protection; E70T-1 and/or E71T-11 electrodes (or E71T-1 and/or E70T-5) of diameter 0.035 in. and/or through 1/16 in. (0.9 mm and/or through 1.6 mm); one or more pieces of mild steel plate, beveled, 7 in. (178 mm) long and 3/4 in. (19 mm) thick or thicker; and a backing strip 9 in. (230 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the horizontal position, Figure 4.49.

Repeat with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2
This practice addresses the "Horizontal" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-23

Butt Joint 2G 100% to Be Tested

Use a properly set up and adjusted FCA welding machine; proper safety protection; E70T-1 and/or E71T-11 electrodes (or E71T-1 and/or E70T-5) of diameter 0.045 in. and/or through 1/16 in. (0.9 mm and/or through 1.6 mm); one or more pieces of mild steel plate, beveled, 7 in. (178 mm) long and 3/4 in. (19 mm) thick or thicker; and a backing strip 9 in. (230 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the horizontal position.

Repeat the weld until you can use each electrode type to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

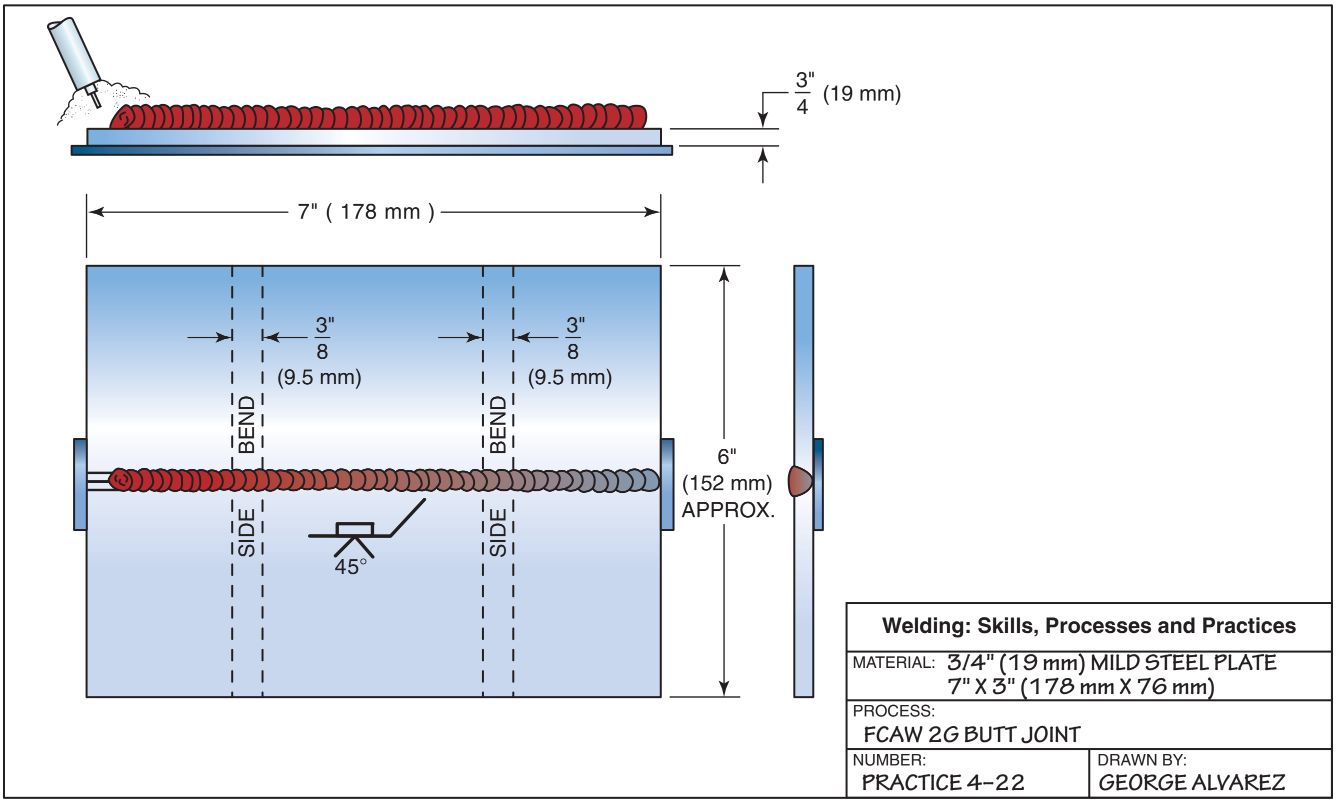
3/4" (19 mm)
7" (178 mm)
3/8 (9.5 mm)
3/8 (9.5 mm)
BEND
BEND
SIDE
SIDE
45°
6" (152 mm) APPROX.
Welding: Skills, Processes and Practices
MATERIAL: 3/4" (19 mm) MILD STEEL PLATE 7" X 3" (178 mm X 76 mm)
PROCESS: FCAW 2G BUTT JOINT
NUMBER: PRACTICE 4-22
DRAWN BY: GEORGE ALVAREZ

Figure 4.49 FCAW 2G butt joint, 3/4 in. mild steel

OVERHEAD-POSITION WELDS

PRACTICE 4-24

Butt Joint 4G

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm), and one or more pieces of mild steel plate, 12 in. (305 mm) long and 1/4 in. (6 mm) thick or thinner, you will make a groove weld in the overhead position.

The molten weld pool should be kept as small as possible for easier control. A small molten weld pool can be achieved by using lower current, faster traveling speeds, and settings.

Lower current settings require closer control of gun manipulation to ensure that the electrode is fed into the molten weld pool just behind the leading edge. The low power will cause overlap and more spatter if this electrode-to-molten weld pool contact position is not closely maintained.

Faster travel speeds allow the welder to maintain a high production rate even if multiple passes are required to complete the weld. Weld penetration into the base metal at the start of the bead can be obtained by using a slow start or quickly reversing the weld direction. Both the slow start and reversal of weld direction put more heat into the weld start to increase penetration. The higher speed also reduces the amount of weld distortion by reducing the amount of time that heat is applied to a joint.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Overhead" position portion of the all-position requirement of 6 and 11.

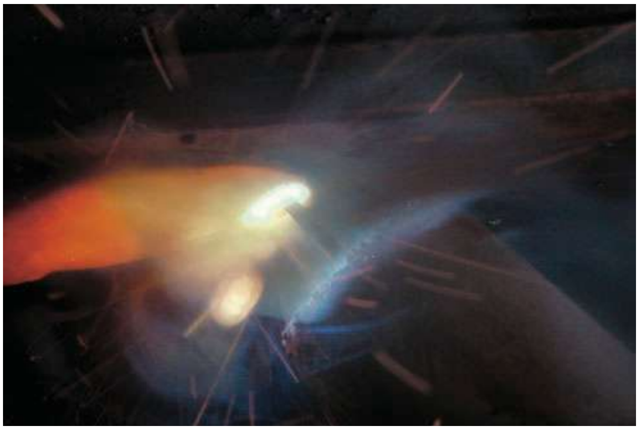

Figure 4.50 Hold the gun so that weld spatter will not fall onto the gun
Source: Courtesy of Larry Jeffus

For overhead welding, extra personal protection is required to reduce the danger of burns. Leather sleeves or leather jackets should be worn.

Much of the spatter created during overhead welding falls into or on the nozzle and contact tube. The contact tube may short out to the gas nozzle. The shorted gas nozzle may arc to the work, causing damage both to the nozzle and to the plate. To control the amount of spatter, a longer stickout and/or a sharper gun-to-plate angle is required to allow most of the spatter to fall clear of the gun or nozzle, Figure 4.50.

Make several short weld beads using various techniques to establish the method that is most successful and most comfortable for you. After each weld, stop and evaluate it before making a change. When you have decided on the technique to be used, make a welded stringer bead that is 12 in. (305 mm) long.

Repeat with both classifications of electrodes until defect-free welds can consistently be made in the 1/4-in. (6-mm)-thick plate. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2
This practice addresses the "Overhead" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-25

Butt Joint 4G 100% to Be Tested

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm), and one or more pieces of mild steel plate, 12 in. (305 mm) long and 1/4 in. (6 mm) thick, you will make a groove weld in the overhead position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-24, repeat the weld until you can use each electrode type to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-26

Butt Joint 4G

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick; and a backing strip 14 in. (355 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the overhead position, Figure 4.51.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-24, repeat with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Overhead" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-27

Butt Joint 4G 100% to Be Tested

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/8 in. (9.5 mm) thick; and a backing strip 14 in. (355 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the overhead position.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2
This practice addresses the "Overhead" position portion of the all-position requirement of 6 and 11.

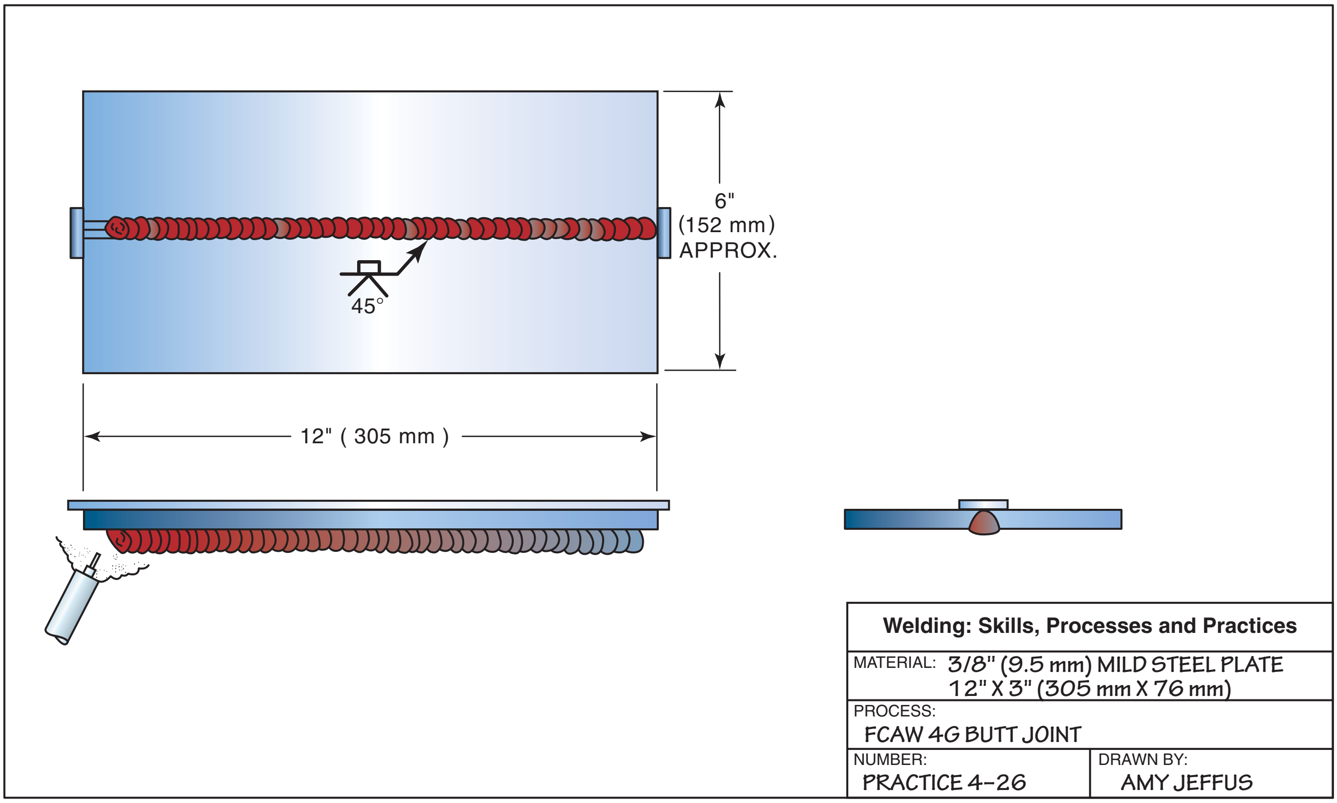
6"
(152 mm)
APPROX.
45°
12" (305 mm)
Welding: Skills, Processes and Practices
MATERIAL: 3/8" (9.5 mm) MILD STEEL PLATE
12" X 3" (305 mm X 76 mm)
PROCESS:
FCAW 4G BUTT JOINT
NUMBER:
PRACTICE 4–26
DRAWN BY:
AMY JEFFUS

Figure 4.51 **FCAW 4G butt joint, 3/8 in. mild steel**

Following the same instructions for the assembly and welding procedure outlined in Practice 4-24, repeat the weld until you can use each electrode type to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Overhead" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-28

Butt Joint 4G

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.045 in. and/or through 1/16 in. (0.9 mm and/or through 1.6 mm); one or more pieces of mild steel plate, beveled, 7 in. (178 mm) long and 3/4 in. (19 mm) thick or thicker; and a backing strip 9 in. (230 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the overhead position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-24, repeat with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2
This practice addresses the "Overhead" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-29

Butt Joint 4G 100% to Be Tested

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.045 in. and/or through 1/16 in. (0.9 mm and/or through 1.6 mm); one or more pieces of mild steel plate, beveled, 7 in. (178 mm) long and 3/4 in. (19 mm) thick or thicker; and a backing strip 9 in. (230 mm) long, 1 in. (25 mm) wide, and 1/4 in. (6 mm) thick. You will make a groove weld in the overhead position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-24, repeat the weld until you can use each electrode classification to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10
This practice addresses the "Overhead" component of the all-position requirement for 5 and 10.

PRACTICE 4-30

Lap Joint and Tee Joint 4F

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.035 in. and/or 0.045 in. (0.9 mm and/or 1.2 mm); and one or more pieces of mild steel plate, beveled, 12 in. (305 mm) long and 3/4 in. (19 mm) thick. You will make a fillet weld in the overhead position, Figure 4.52A and Figure 4.52B.

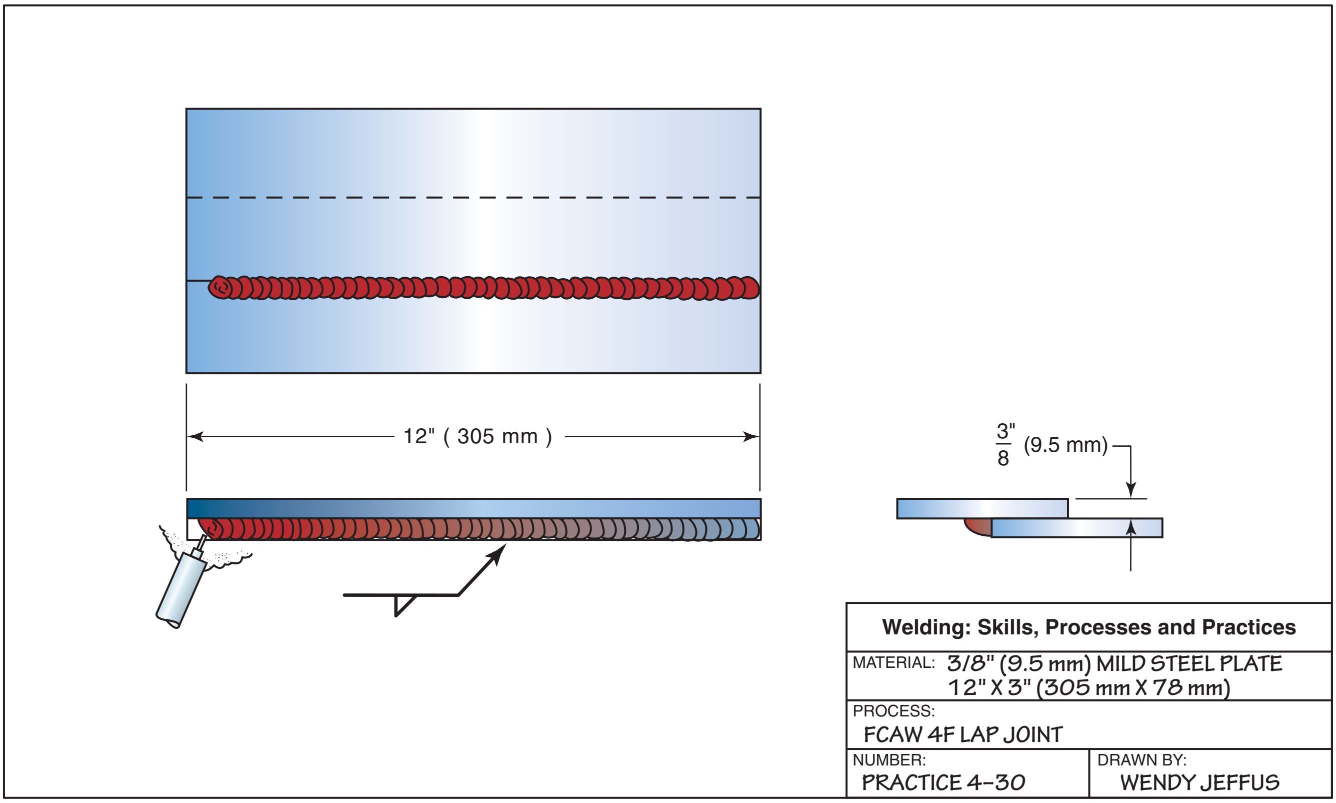

Welding: Skills, Processes and Practices

MATERIAL: 3/8" (9.5 mm) MILD STEEL PLATE
12" X 3" (305 mm X 78 mm)

PROCESS:
FCAW 4F LAP JOINT

NUMBER: PRACTICE 4-30

DRAWN BY: WENDY JEFFUS

Figure 4.52 (A) FCAW 4F lap joint, 3/8 in. mild steel

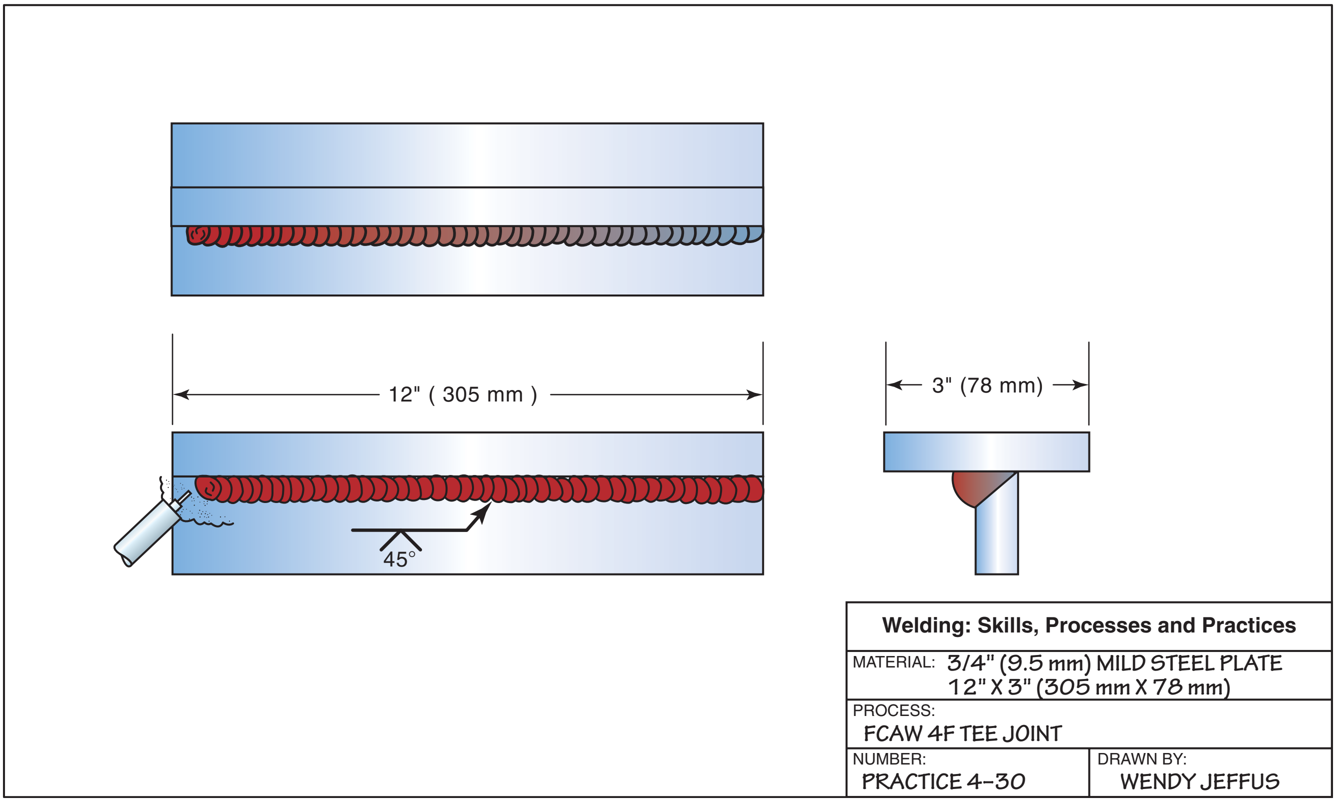

Welding: Skills, Processes and Practices

MATERIAL: 3/4" (9.5 mm) MILD STEEL PLATE
12" X 3" (305 mm X 78 mm)

PROCESS:
FCAW 4F TEE JOINT

NUMBER: PRACTICE 4-30

DRAWN BY: WENDY JEFFUS

Figure 4.52 (B) FCAW 4F tee joint, 3/4 in. mild steel

Following the same instructions for the assembly and welding procedure outlined in Practice 4-24, repeat each type of joint with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-31

Tee Joint 4F 100% to Be Tested

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10

Module 9
Key Indicator 1, 2
This practice addresses the "Overhead" component of the all-position requirement for 5 and 10.

Use a properly set up and adjusted FCA welding machine; proper safety protection; E71T-1 and/or E71T-11 electrodes of diameter 0.045 in. and/or through 1/16 in. (0.9-mm and/or through 1.6-mm); and one or more pieces of mild steel plate, beveled, 7 in. (178 mm) long and 3/4 in. (19 mm) thick or thicker. You will make a fillet weld in the overhead position.

Following the same instructions for the assembly and welding procedure as outlined in Practice 4-24, repeat the weld until you can use each electrode classification to make welds with 100% penetration that will pass the bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

THIN-GAUGE WELDING

The introduction of small electrode diameters has allowed FCA welding to be used on thin sheet metal. Usually these welds will be a fillet type, the easiest weld to make on thin stock. An effort should be made when possible to design the weld so it is not a butt-type joint. A common use for FCA welding on thin stock is to join it to a thicker member, Figure 4.53. This type of weld is used to put panels in frames.

The following practices include some butt-type joints. You will find that the vertical down welds are the easiest ones to make. If it is possible to position the weldment for a vertical down weld, production speeds for butt joints can be increased.

PRACTICE 4-32

Butt Joint 1G

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Use a properly set up and adjusted FCA welding machine, Table 4.3; proper safety protection; E71T-1 and/or E70T-5 or E70T-1 and/or E71T-11 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm); and one or more pieces of mild steel sheet, 12 in. (305 mm) long

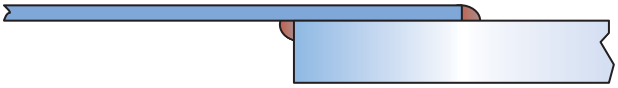

Figure 4.53 FCA welding of thin to thick metal

Table 4.3 **FCA Welding Parameters for Use if Specific Settings Are Unavailable from Electrode Manufacturer**

Electrode		Welding Power			Shielding Gas		Base Metal	
Type	**Size**	**Amps**	**Wire-feed Speed, ipm (cm/min)**	**Volts**	**Type**	**Flow**	**Type**	**Thick**
E70T-1 E71T-1	0.030 in. (0.8 mm)	40 to 145	90 to 340 (228 to 864)	20 to 27	None	n/a	Low-carbon steel	16 gauge to 18 gauge
E70T-1 E71T-1	0.035 in. (0.9 mm)	130 to 200	288 to 576 (732 to 1463)	20 to 28	None	n/a	Low-carbon steel	16 gauge to 18 gauge
E70T-5 E71T-11	0.035 in. (0.9 mm)	90 to 200	190 to 576 (483 to 1463)	16 to 29	57% argon 25% CO_2	35 cfh	Low-carbon steel	16 gauge 18 gauge

and 16-gauge to 18-gauge thick. You will make a butt weld in the flat position, Figure 4.54.

Do not leave a root opening for these welds. Even the slightest opening will result in a burn-through. If a burn-through occurs, the welder can be pulsed off and on so that the hole can be filled. This process will leave a larger than usual buildup. Excessive buildup could be ground off if necessary as part of the postweld cleanup.

Repeat with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I provided by your instructor.

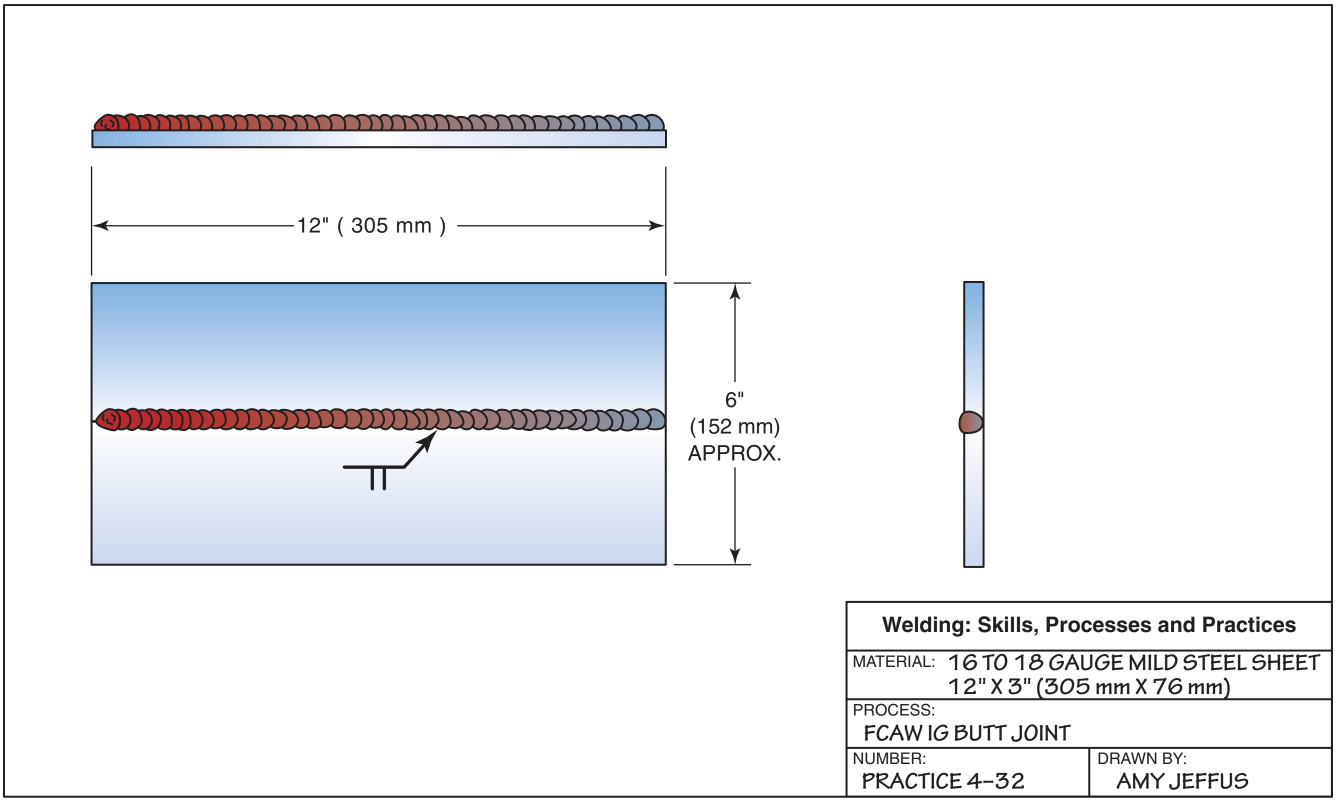

Welding: Skills, Processes and Practices

MATERIAL: 16 TO 18 GAUGE MILD STEEL SHEET
12" X 3" (305 mm X 76 mm)

PROCESS:
FCAW IG BUTT JOINT

NUMBER:
PRACTICE 4-32

DRAWN BY:
AMY JEFFUS

Figure 4.54 **FCAW 1G butt joint, 16- to 18-gauge mild steel**

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2

PRACTICE 4-33

Butt Joint 1G 100% to Be Tested

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E70T-5 or E70T-1 and/or E70T-11 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a butt weld in the flat position, Figure 4.55.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat the weld until you can use each electrode classification to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Gas Shielded
Key Indicator 10
This practice addresses the “Flat” component of the all-position requirement for 5 and 10.

PRACTICE 4-34

Lap Joint and Tee Joint 1F

Using a properly set up and adjusted FCA welding machine, proper safety protection, E70T-1 and/or E71T-11 or E71T-1 and/or E70T-5 electrodes of diameter 0.030-in. and/or 0.035-in. (0.8-mm and/or 0.9-mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a fillet weld in the flat position.

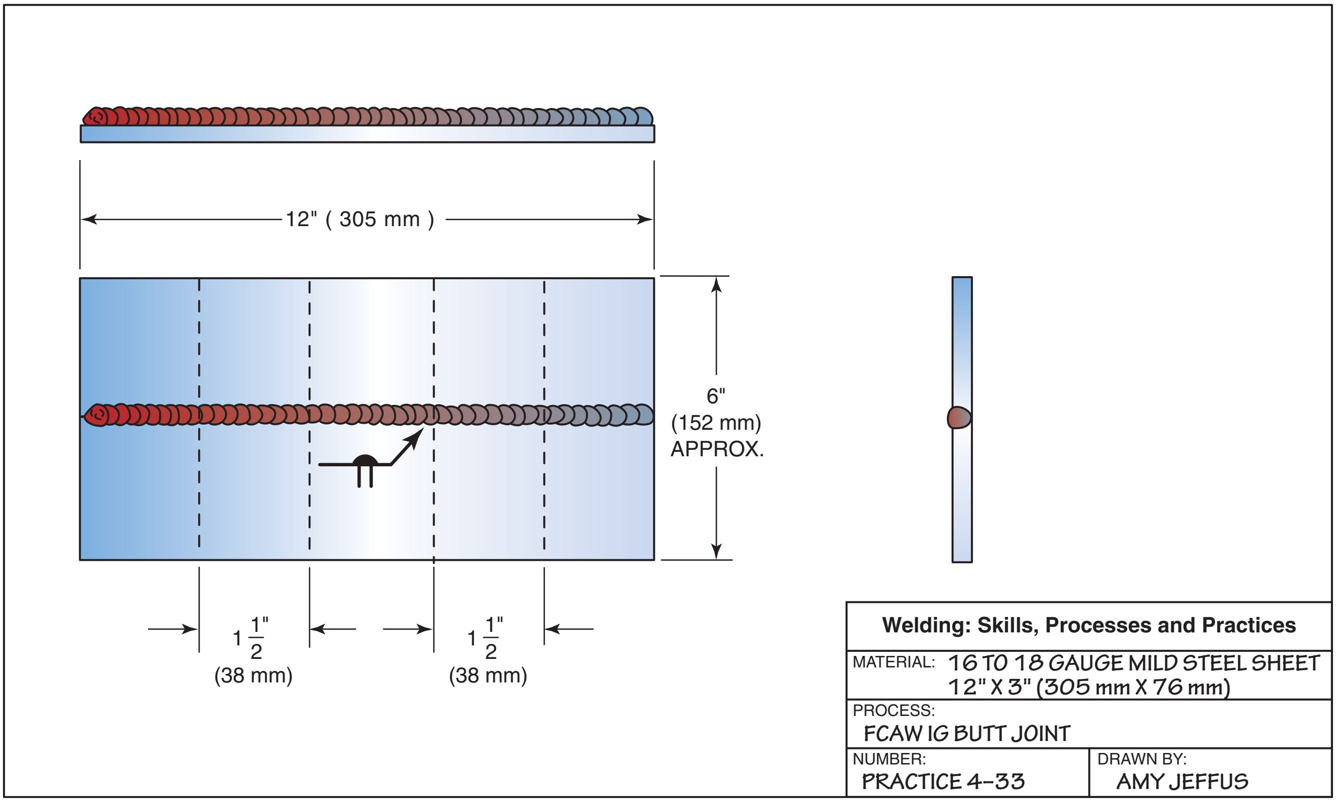
12" (305 mm)
6"
(152 mm)
APPROX.
1 1/2"
(38 mm)
1 1/2"
(38 mm)
Welding: Skills, Processes and Practices
MATERIAL: 16 TO 18 GAUGE MILD STEEL SHEET
12" X 3" (305 mm X 76 mm)
PROCESS:
FCAW IG BUTT JOINT
NUMBER:
PRACTICE 4-33
DRAWN BY:
AMY JEFFUS

Figure 4.55 FCAW 1G butt joint, 16- to 18-gauge mild steel

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat each type of joint with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10

Module 9
Key Indicator 1, 2

PRACTICE 4-35

Lap Joint and Tee Joint 1F 100% to Be Tested

Using a properly set up and adjusted FCA welding machine, proper safety protection, E70T-1 and/or E71T-11 or E71T-1 and/or E70T-5 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a fillet weld in the flat position, Figure 4.56A and Figure 4.56B.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat each type of joint with both classifications of electrodes until you can make welds with 100% penetration that will pass the bend test, Figure 4.57A and Figure 4.57B. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the “Vertical” position portion of the all-position requirement of 6 and 11.

PRACTICE 4-36

Butt Joint 3G

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-11 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild

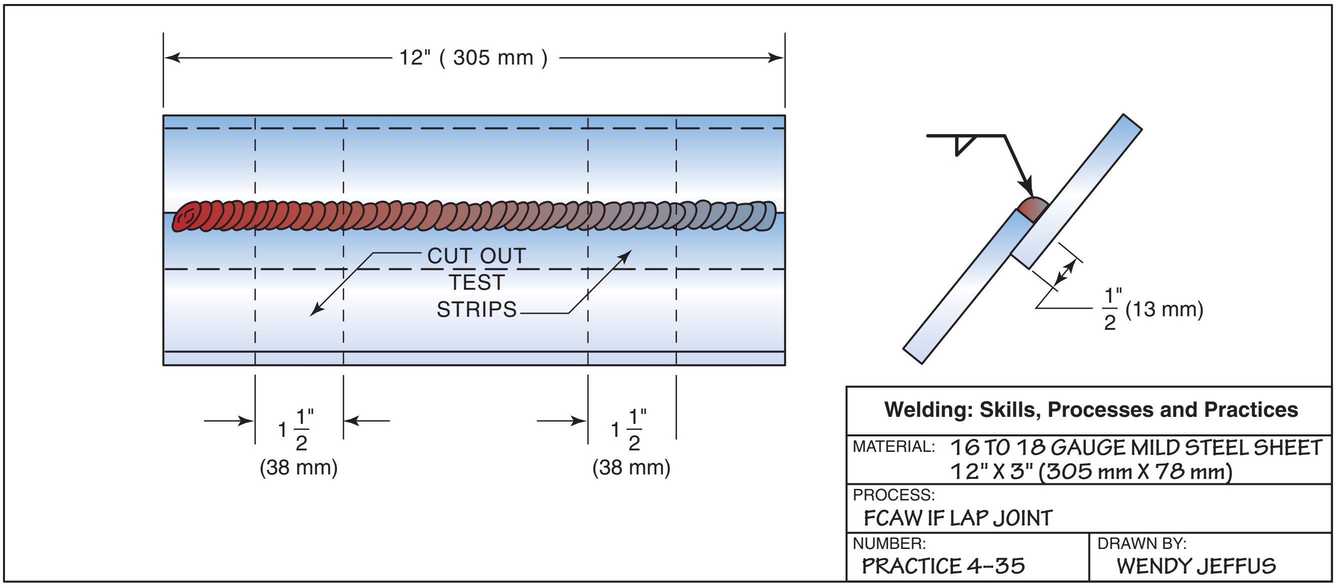
12" (305 mm)
CUT OUT TEST STRIPS
1 1/2" (38 mm)
1 1/2" (38 mm)
1/2" (13 mm)
Welding: Skills, Processes and Practices
MATERIAL: 16 TO 18 GAUGE MILD STEEL SHEET 12" X 3" (305 mm X 78 mm)
PROCESS: FCAW IF LAP JOINT
NUMBER: PRACTICE 4-35
DRAWN BY: WENDY JEFFUS

Figure 4.56 (A) FCAW 1F lap joint, 16- to 18-gauge mild steel

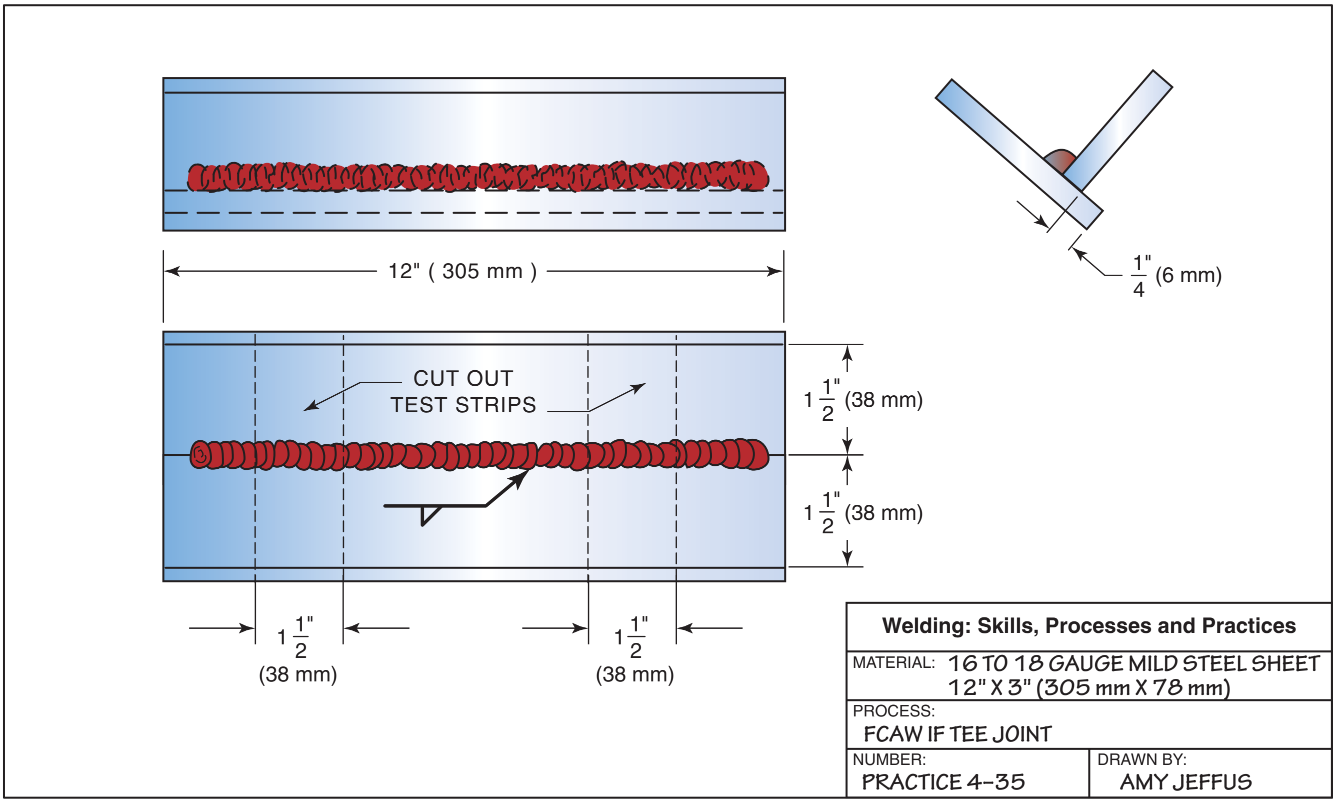
12" (305 mm)
CUT OUT
TEST STRIPS
1 1/2" (38 mm)
1 1/2" (38 mm)
1 1/2" (38 mm)
1 1/2" (38 mm)
1/4" (6 mm)
Welding: Skills, Processes and Practices
MATERIAL: 16 TO 18 GAUGE MILD STEEL SHEET
12" X 3" (305 mm X 78 mm)
PROCESS:
FCAW IF TEE JOINT
NUMBER:
PRACTICE 4-35
DRAWN BY:
AMY JEFFUS

Figure 4.56 (B) FCAW 1F tee joint, 16- to 18-gauge mild steel

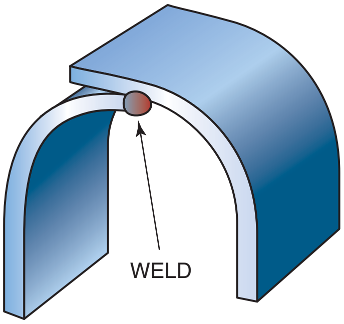

Figure 4.57 (A) 180° bend to test lap weld quality

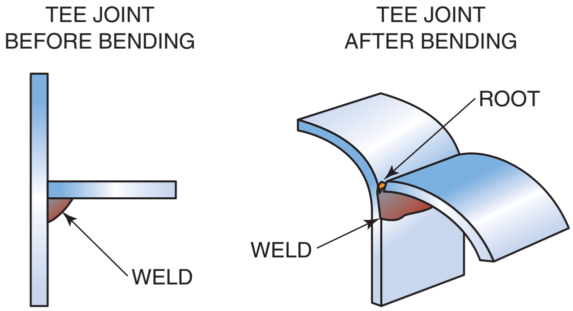
TEE JOINT
BEFORE BENDING
WELD
TEE JOINT
AFTER BENDING
ROOT
WELD

Figure 4.57 (B) Bend the test strip to be sure the weld had good root fusion

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10
This practice addresses the "Vertical" component of the all-position requirement for 5 and 10.

steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a butt weld in the vertical up or down position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-37

Lap Joint and Tee Joint 3F

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-11 electrodes of diameter 0.030 in.

and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18- gauge thick, you will make a fillet weld in the vertical up or down position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat each type of joint with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

PRACTICE 4-38

Lap Joint and Tee Joint 3F 100% to Be Tested

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-11 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a fillet weld in the vertical up or down position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat each type of joint with both classifications of electrodes until you can make welds with 100% penetration that will pass the test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10

Module 9
Key Indicator 1, 2
This practice addresses the “Vertical” component of the all-position requirement for 5 and 10.

PRACTICE 4-39

Lap Joint and Tee Joint 2F

Using a properly set up and adjusted FCA welding machine, proper safety protection, E70T-1 and/or E71T-11 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a fillet weld in the horizontal position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat each type of joint with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10
This practice addresses the “Horizontal” component of the all-position requirement for 5 and 10.

PRACTICE 4-40

Lap Joint and Tee Joint 2F 100% to Be Tested

Using a properly set up and adjusted FCA welding machine, proper safety protection, E70T-1 and/or E71T-11 or E71T-1 and/or E70T-5 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 5
Self Shielded
Key Indicator 10

Module 9
Key Indicator 1, 2
This practice addresses the “Horizontal” component of the all-position requirement for 5 and 10.

16-gauge to 18-gauge thick, you will make a fillet weld in the horizontal position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat each type of joint with both classifications of electrodes until you can make welds with 100% penetration that will pass the bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Horizontal" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-41

Butt Joint 2G

Using a properly set up and adjusted FCA welding machine, proper safety protection, E70T-1 and/or E71T-11 or E71T-1 and/or E70T-5 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a butt weld in the horizontal position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2
This practice addresses the "Horizontal" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-42

Butt Joint 2G 100% to Be Tested

Using a properly set up and adjusted FCA welding machine, proper safety protection, E70T-1 and/or E71T-11 or E71T-1 and/or E70T-5 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a butt weld in the horizontal position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat the weld until you can use each electrode classification to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11
This practice addresses the "Overhead" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-43

Butt Joint 4G

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-5 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a butt weld in the overhead position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat with both classifications of electrodes until defect-free welds can consistently be made. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

PRACTICE 4-44

Butt Joint 4G 100% to Be Tested

Using a properly set up and adjusted FCA welding machine, proper safety protection, E71T-1 and/or E71T-5 electrodes of diameter 0.030 in. and/or 0.035 in. (0.8 mm and/or 0.9 mm), and one or more pieces of mild steel sheet, 12 in. (305 mm) long and 16-gauge to 18-gauge thick, you will make a butt weld in the overhead position.

Following the same instructions for the assembly and welding procedure outlined in Practice 4-32, repeat the weld until you can use each electrode classification to make welds with 100% penetration that will pass a bend test. Turn off the welding machine and shielding gas and clean up your work area when you are finished welding.

Complete a copy of the "Student Welding Report" listed in Appendix I or provided by your instructor.

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Gas Shielded
Key Indicator 6
Self Shielded
Key Indicator 11

Module 9
Key Indicator 1, 2
This practice addresses the "Overhead" position portion of the all-position requirement of 6 and 11.

PRACTICE 4-45

AWS SENSE Entry-Level Welder Workmanship Sample for Flux Cored Arc Welding, Gas-Shielded (FCAW)

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Key Indicator 7

Welding Procedure Specification (WPS)

Welding Procedure Specification No.: Practice 4-45. Date:

Title:
Welding FCAW of plate to plate.

Scope:
This procedure is applicable for V-groove, bevel, and fillet welds within the range of 1/8 in. (3.2 mm) through 1-1/2 in. (38 mm).

Welding may be performed in the following positions: all.

Base Metal:
The base metal shall conform to carbon steel M-1, P-1, and S-1, Group 1 or 2.

Backing material specification: none.

Filler Metal:
The filler metal shall conform to AWS specification no. E71T-1 from AWS specification A5.20. This filler metal falls into F-number F-6 and A-number A-1.

Shielding Gas:
The shielding gas, or gases, shall conform to the following compositions and purity:

CO_2 at 30 to 50 cfh or 75% Ar/25% CO_2 at 30 to 50 cfh.

Joint Design and Tolerances:

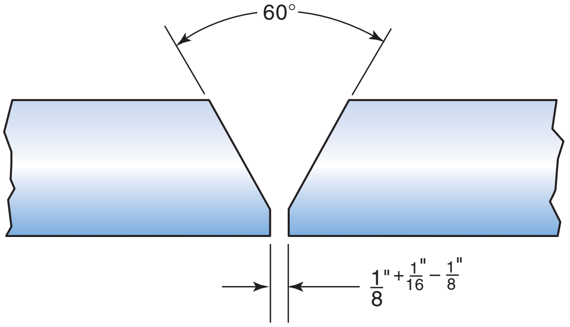

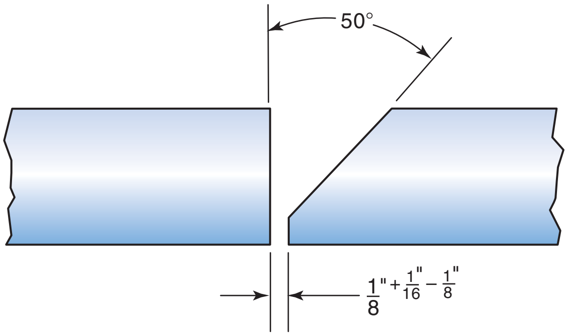

Preparation of Base Metal:

The bevels are to be flame cut on the edges of the plate before the parts are assembled. The beveled surface must be smooth and free of notches. Any roughness or notches deeper than 1/64 in. (0.4 mm) must be ground smooth.

All hydrocarbons and other contaminations, such as cutting fluids, grease, oil, and primers, must be cleaned off all parts and filler metals before welding. This cleaning can be done with any suitable solvents or detergents. The groove face and inside and outside plate surface within 1 in. (25 mm) of the joint must be mechanically cleaned of slag, rust, and mill scale. Cleaning must be done with a wire brush or grinder down to bright metal.

Electrical Characteristics:

The current shall be direct current electrode positive (DCEP). The base metal shall be on the negative side of the line.

Electrode		Welding Power			Shielding Gas		Base Metal	
Type	Size	Amps	Wire-feed Speed, ipm (cm/min)	Volts	Type	Flow	Type	Thickness
E71T-1	0.035 in. (0.9 mm)	130 to 150	288 to 380 (732 to 975)	22 to 25	CO_2 or 75% Ar/ CO_2 25%	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)
E71T-1	0.045 in. (1.2 mm)	150 to 210	200 to 300 (508 to 762)	28 to 29	CO_2 or 75% Ar/ CO_2 25%	30 to 50	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)

Preheat:

The parts must be heated to a temperature higher than 50°F (10°C) before any welding is started.

Backing Gas:

N/A

Safety:
Proper protective clothing and equipment must be used. The area must be free of all hazards that may affect the welder or others in the area. The welding machine, welding leads, work clamp, electrode holder, and other equipment must be in safe working order.

Welding Technique:
Using a 1/2-in. (13-mm) or larger gas nozzle and a distance from contact tube to work of approximately 3/4 in. (19 mm) for all welding, first tack weld the plates together according to Figure 4.58. There should be a root gap of about 1/8 in. (3.2 mm) between the plates with V-grooved or beveled edges. Use an E71T-1 arc welding electrode to make a root pass to fuse the plates together. Clean the slag from the root pass, being sure to remove any trapped slag along the sides of the weld.

Using an E71T-1 arc welding electrode, make a series of stringer or weave filler welds, no thicker than 1/4 in. (6.4 mm), in the groove until the joint is filled. The 1/4-in. (6.4-mm) fillet welds are to be made with one pass.

Interpass Temperature:
The plate should not be heated to a temperature higher than 350°F (175° C) during the welding process. After each weld pass is completed, allow it to cool but never to a temperature below 50°F (10°C). The weldment must not be quenched in water.

Cleaning:
The slag must be cleaned off between passes. The weld beads may be cleaned by a hand wire brush, a hand chipping, a punch and hammer, or a needle-scaler. All weld cleaning must be performed with the test plate in the welding position. A grinder may not be used to remove weld control problems such as undercut, overlap, or trapped slag.

Inspection:
Visually inspect the weld for uniformity and discontinuities. There shall be no cracks, no incomplete fusion, and no overlap. Undercut shall not exceed the lesser of 10% of the base metal thickness or 1/32 in. (0.8 mm). The frequency of porosity shall not exceed one in each 4 in. (100 mm) of weld length, and the maximum diameter shall not exceed 3/32 in. (2.4 mm).

Sketches:
See Figure 4.58.

Complete a copy of the “Student Welding Report” listed in Appendix I or provided by your instructor.

PRACTICE 4-46

AWS SENSE Entry-Level Welder Workmanship Sample for Flux Cored Arc Welding Self-Shielded (FCAW)

Module 1
Key Indicator 1, 2, 3, 4

Module 2
Key Indicator 1, 2, 3, 4, 7

Module 6
Key Indicator 12

Welding Procedure Specification (WPS)

Welding Procedure Specification No.: Practice 4-46 Date:

Title:
Welding FCAW of plate to plate.

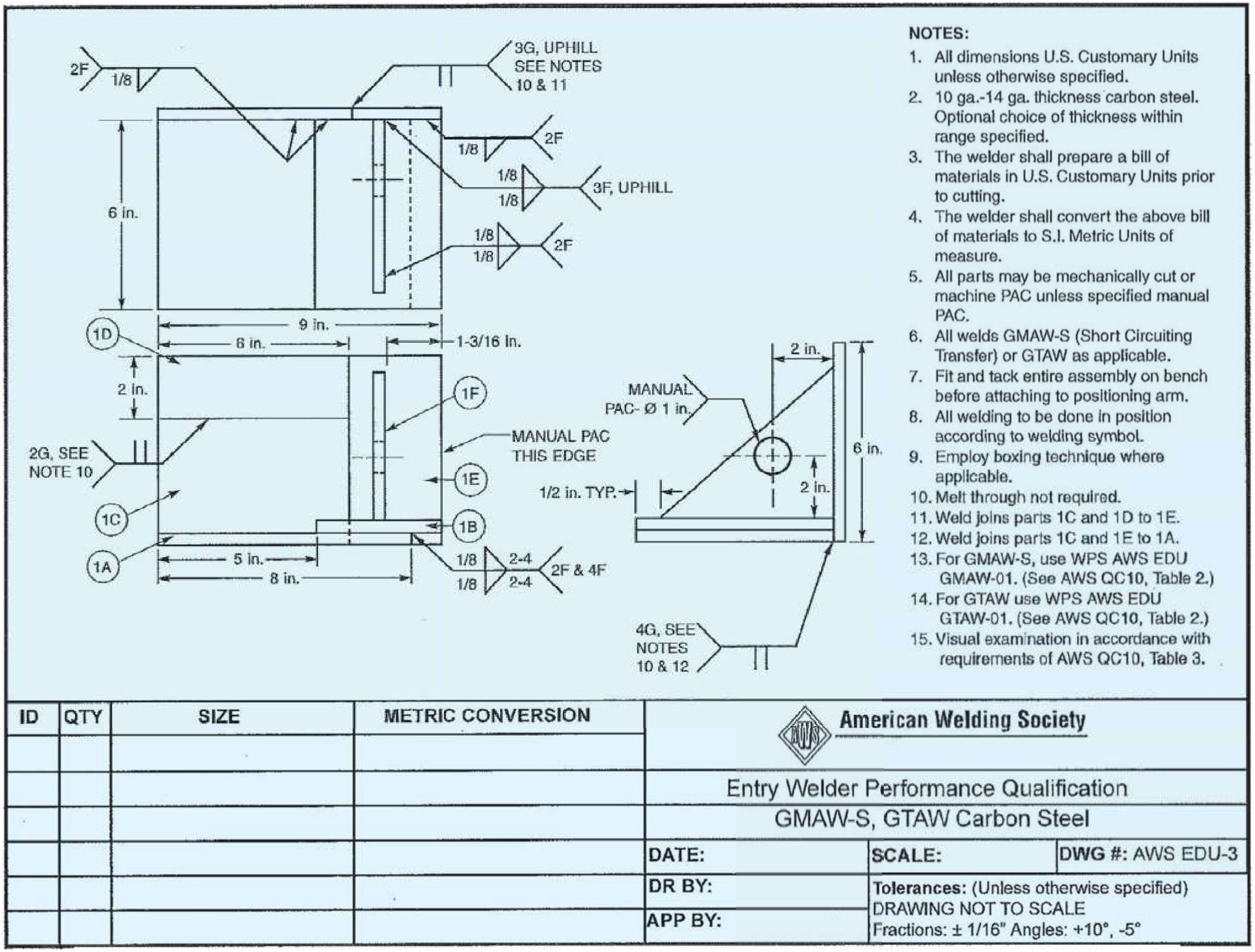
2F
1/8
3G, UPHILL
SEE NOTES
10 & 11
1/8
2F
1/8
1/8
3F, UPHILL
1/8
1/8
2F
6 in.
9 in.
6 in.
1-3/16 in.
1D
2 in.
1F
MANUAL PAC
THIS EDGE
2G, SEE
NOTE 10
1E
1C
1B
1A
5 in.
8 in.
1/8
1/8
2-4
2-4
2F & 4F
MANUAL
PAC- Ø 1 in.
2 in.
6 in.
2 in.
1/2 in. TYP.
4G, SEE
NOTES
10 & 12

NOTES:

1. All dimensions U.S. Customary Units unless otherwise specified.
2. 10 ga.-14 ga. thickness carbon steel. Optional choice of thickness within range specified.
3. The welder shall prepare a bill of materials in U.S. Customary Units prior to cutting.
4. The welder shall convert the above bill of materials to S.I. Metric Units of measure.
5. All parts may be mechanically cut or machine PAC unless specified manual PAC.
6. All welds GMAW-S (Short Circuiting Transfer) or GTAW as applicable.
7. Fit and tack entire assembly on bench before attaching to positioning arm.
8. All welding to be done in position according to welding symbol.
9. Employ boxing technique where applicable.
10. Melt through not required.
11. Weld joins parts 1C and 1D to 1E.
12. Weld joins parts 1C and 1E to 1A.
13. For GMAW-S, use WPS AWS EDU GMAW-01. (See AWS QC10, Table 2.)
14. For GTAW use WPS AWS EDU GTAW-01. (See AWS QC10, Table 2.)
15. Visual examination in accordance with requirements of AWS QC10, Table 3.

ID	QTY	SIZE	METRIC CONVERSION

American Welding Society

Entry Welder Performance Qualification

GMAW-S, GTAW Carbon Steel

DATE:	**SCALE:**	**DWG #:** AWS EDU-3
DR BY:	**Tolerances:** (Unless otherwise specified) DRAWING NOT TO SCALE Fractions: ± 1/16" Angles: +10°, -5°	
APP BY:		

Figure 4.58 **FCAW-G workmanship qualification test**

Source: Courtesy of the American Welding Society

Scope:
This procedure is applicable for V-groove, bevel, and fillet welds within the range of 1/8 in. (3.2 mm) through 1 1/2 in. (38 mm).

Welding may be performed in the following positions: all.

Base Metal:
The base metal shall conform to carbon steel M-1, P-1, and S-1, Group 1 or 2.

Backing material specification: none.

Filler Metal:
The filler metal shall conform to AWS specification no. 0.035 to 0.0415 dia. E71T-11 from AWS specification A5.20. This filler metal falls into F-number F-6 and A-number A-1.

Shielding Gas:
The shielding gas, or gases, shall conform to the following compositions and purity: none.

Joint Design and Tolerances:

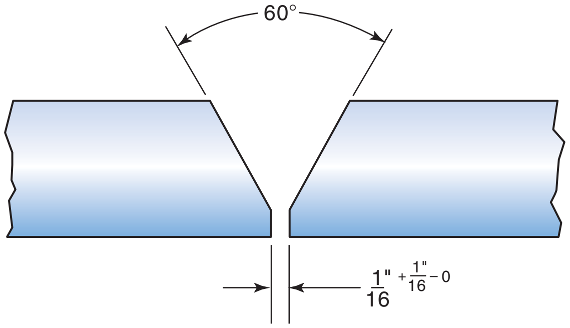

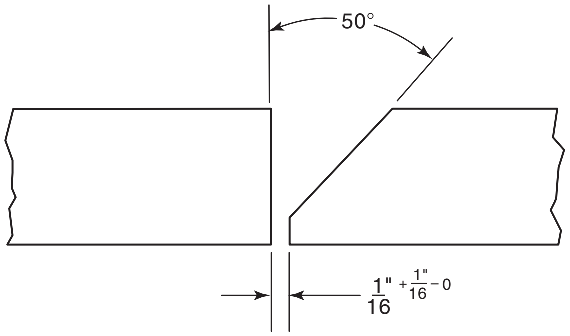

Preparation of Base Metal:
The bevels are to be flame cut on the edges of the plate before the parts are assembled. The beveled surface must be smooth and free of notches. Any roughness or notches deeper than 1/64 in. (0.4 mm) must be ground smooth.

All hydrocarbons and other contaminations, such as cutting fluids, grease, oil, and primers, must be cleaned off all parts and filler metals before welding. This cleaning can be done with any suitable solvents or detergents. The groove face and inside and outside plate surface within 1 in. (25 mm) of the joint must be mechanically cleaned of slag, rust, and mill scale. Cleaning must be done with a wire brush or grinder down to bright metal.

Electrical Characteristics:
The current shall be direct current electrode negative (DCEN). The base metal shall be on the positive side of the line.

Electrode		Welding Power			Shielding Gas		Base Metal	
Type	Size	Amps	Wire-feed Speed, ipm (cm/min)	Volts	Type	Flow	Type	Thickness
E71T-11	0.035 in. (0.9 mm)	130 to 150	150 to 225 (381 to 571)	22 to 25	None	—	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)
E71T-11	0.045 in. (1.2 mm)	150 to 210	105 to 195 (266 to 495)	15 to 18	None	—	Low-carbon steel	1/4 in. to 1/2 in. (6 mm to 13 mm)

Preheat:
The parts must be heated to a temperature higher than 50°F (10°C) before any welding is started.

Backing Gas:
N/A

Safety:
Proper protective clothing and equipment must be used. The area must be free of all hazards that may affect the welder or others in the area. The welding machine, welding leads, work clamp, electrode holder, and other equipment must be in safe working order.

Welding Technique:
Using a 1/2-in. (13-mm) or larger gas nozzle and a distance from contact tube to work of approximately 3/4 in. (19 mm) for all welding, first tack weld the plates together according to Figure 4.59. There should be a root gap of about 1/8 in. (3.2 mm) between the plates with V-grooved or beveled edges. Use an E71T-11 arc welding electrode to make a root pass to fuse the plates together. Clean the slag from the root pass, being sure to remove any trapped slag along the sides of the weld.

Using an E71T-11 arc welding electrode, make a series of stringer or weave filler welds, no thicker than 1/4 in. (6.4 mm), in the groove until the joint is filled. The 1/4-in. (6.4-mm) fillet welds are to be made with one pass.

Interpass Temperature:
The plate should not be heated to a temperature higher than 350°F (175°C) during the welding process. After each weld pass is completed, allow it to cool but never to a temperature below 50°F (10°C). The weldment must not be quenched in water.

Cleaning:
The slag must cleaned off between passes. The weld beads may be cleaned by a hand wire brush, a hand chipping, a punch and hammer, or a needle-scaler. All weld cleaning must be performed with the test plate in the welding position. A grinder may not be used to remove weld control problems such as undercut, overlap, or trapped slag.

Inspection:
Visually inspect the weld for uniformity and discontinuities. There shall be no cracks, no incomplete fusion, and no overlap. Undercut shall not exceed the lesser of 10% of the base metal thickness or 1/32 in. (0.8 mm). The frequency of porosity shall not exceed one in each 4 in. (100 mm) of weld length, and the maximum diameter shall not exceed 3/32 in. (2.4 mm).

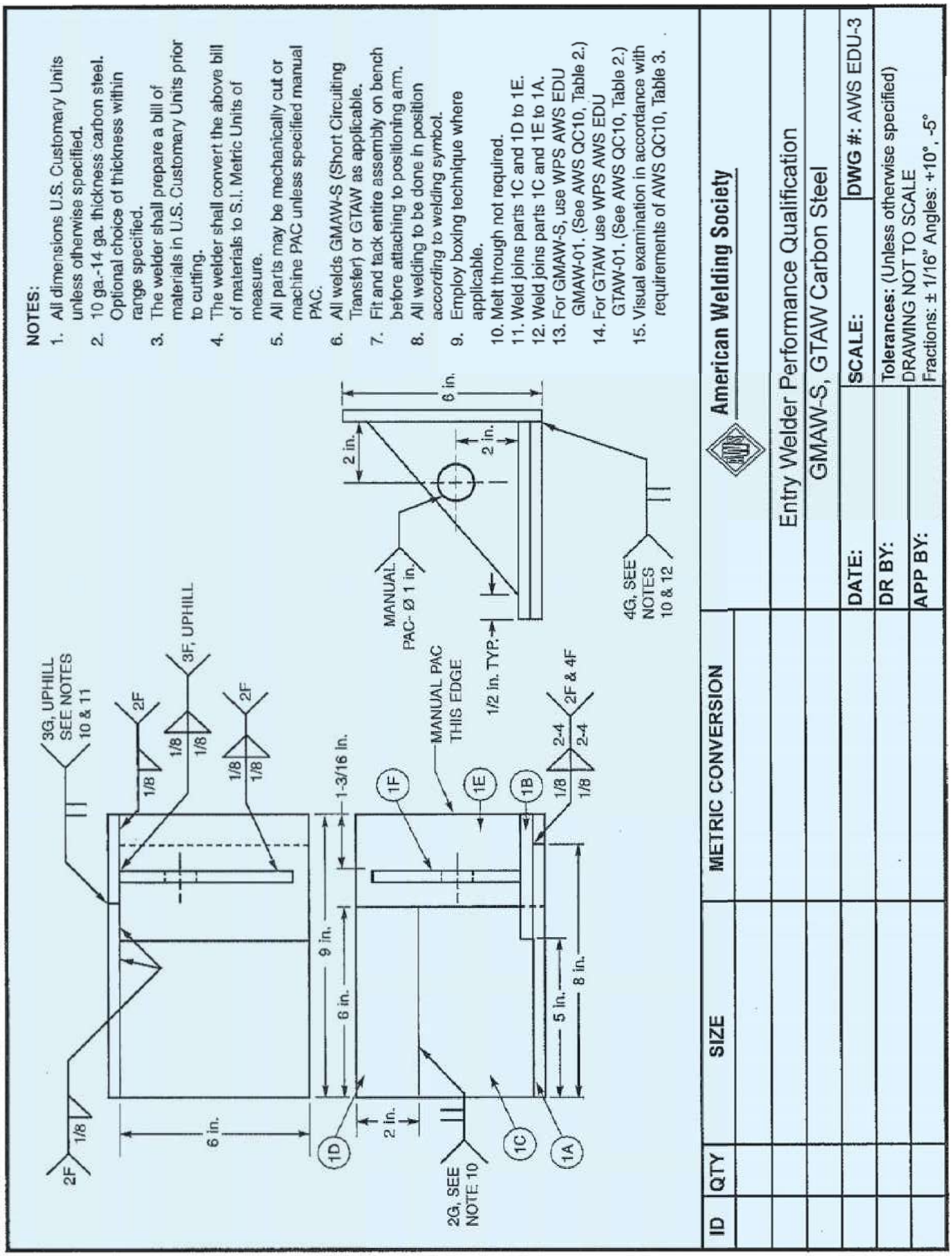
3G, UPHILL SEE NOTES 10 & 11
2F
1/8
3F, UPHILL
1/8
1/8
2F
1/8
1/8
2F
1/8
6 in.
1D
2 in.
2G, SEE NOTE 10
1C
1A
6 in.
9 in.
5 in.
8 in.
1-3/16 in.
1F
MANUAL PAC THIS EDGE
1E
1B
1/8
1/8
2-4
2-4
2F & 4F
1/2 in. TYP.
MANUAL PAC- Ø 1 in.
2 in.
2 in.
6 in.
4G, SEE NOTES 10 & 12

NOTES:

1. All dimensions U.S. Customary Units unless otherwise specified.
2. 10 ga.-14 ga. thickness carbon steel. Optional choice of thickness within range specified.
3. The welder shall prepare a bill of materials in U.S. Customary Units prior to cutting.
4. The welder shall convert the above bill of materials to S.I. Metric Units of measure.
5. All parts may be mechanically cut or machine PAC unless specified manual PAC.
6. All welds GMAW-S (Short Circuiting Transfer) or GTAW as applicable.
7. Fit and tack entire assembly on bench before attaching to positioning arm.
8. All welding to be done in position according to welding symbol.
9. Employ boxing technique where applicable.
10. Melt through not required.
11. Weld joins parts 1C and 1D to 1E.
12. Weld joins parts 1C and 1E to 1A.
13. For GMAW-S, use WPS AWS EDU GMAW-01. (See AWS QC10, Table 2.)
14. For GTAW use WPS AWS EDU GTAW-01. (See AWS QC10, Table 2.)
15. Visual examination in accordance with requirements of AWS QC10, Table 3.

ID	QTY	SIZE	METRIC CONVERSION

American Welding Society

Entry Welder Performance Qualification

GMAW-S, GTAW Carbon Steel

DATE:
SCALE:
DWG #: AWS EDU-3
DR BY:
APP BY:
Tolerances: (Unless otherwise specified)
DRAWING NOT TO SCALE
Fractions: ± 1/16" Angles: +10°, -5°

Figure 4.59 **FCAW-G workmanship qualification test**

Source: Courtesy of the American Welding Society

WELDING BASICS

An Introduction to Practical & Ornamental Welding

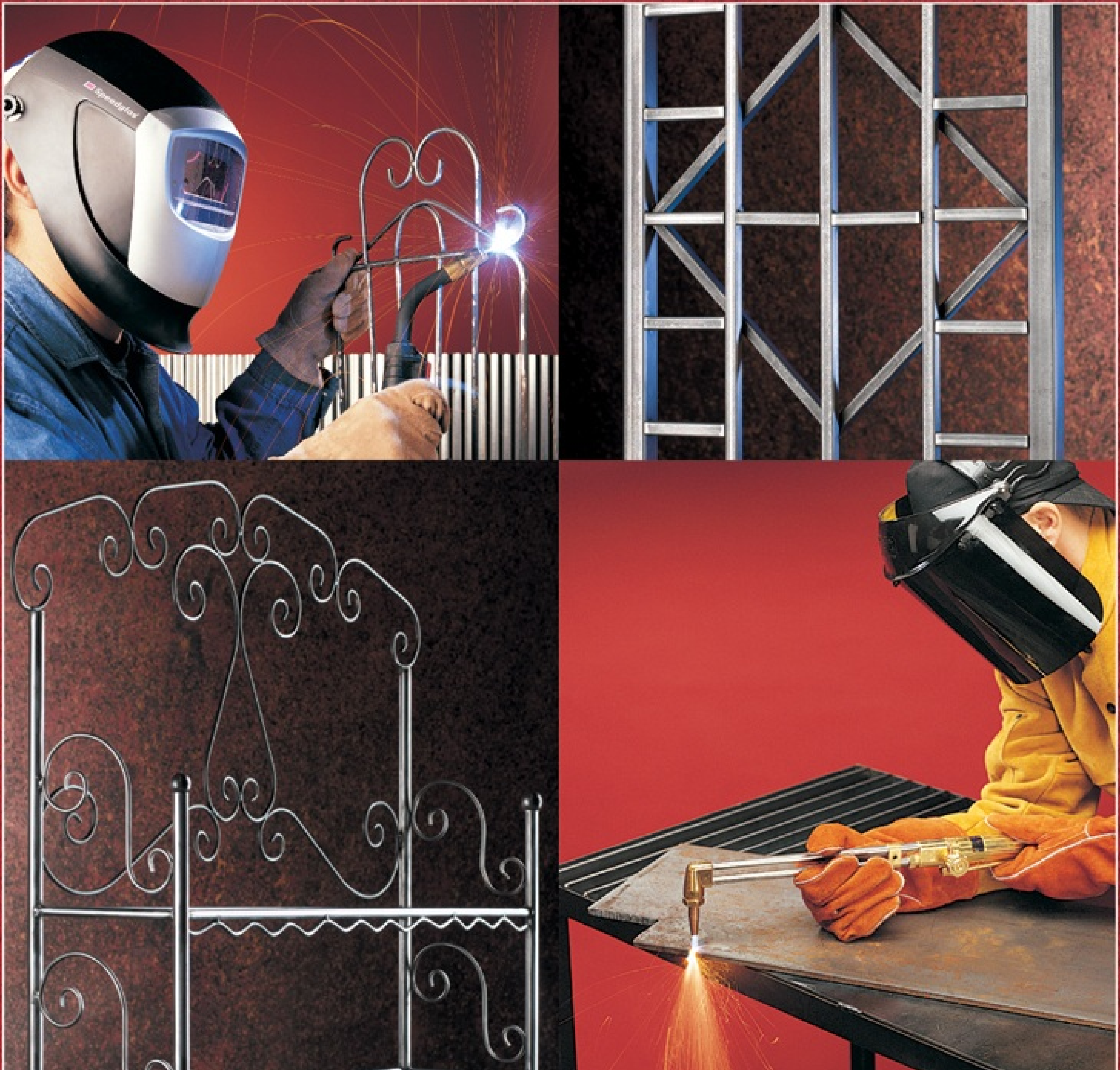

WELDING
BASICS

An Introduction to Practical
& Ornamental Welding

Creative Publishing
international

MINNEAPOLIS, MINNESOTA
www.creativepub.com

CONTENTS

Introduction
- Introduction to Welding
- Safety
- Setting Up Shop
- Shop Tools
- Metal Basics

Welding & Cutting Processes
- Oxyacetylene Welding
- Electricity for Welding
- Shielded Metal Arc Welding
- Gas Metal Arc Welding
- Gas Tungsten Arc Welding
- Plasma Cutting
- Oxyacetylene Cutting

Welding Projects
- Shop Projects
  - *Rolling Welding Curtain*
  - *Welding Table*
  - *Welding Machine Cart*
  - *Cylinder Cart*
  - *Grinder Stand*
- Lighting Projects
  - *Scroll Desk Lamp*
  - *Three Post Floor Lamp*
  - *Chandelier*
  - *Wall-mounted Candleholder*
  - *Tree-shaped Candleholder*
  - *Fireplace Candleholder*
- Furniture Projects
  - *Baker's Shelves*
  - *Coffee Table*

    - *Corner Coatrack*
    - *Kitchen Appliance Stand*
    - *Wall Hung Plate Rack*
    - *Wine Rack*
    - *Headboard*
- Outdoor Projects
    - *Stair Railing*
    - *Gate*
    - *Critter Boot Brush*
    - *Disk Fountain*
    - *Arbor*

Glossary
Resources
Index

A simple flame-cut silhouette is a charming addition to any garden.

photo/©John Gregor

INTRODUCTION

Welding is a practical skill that is also great fun. The number of welded items in our everyday lives is practically uncountable—the spot welds on the bodies of our automobile, the welded railing on our front steps, the superstructure of the buildings in which we work, and the bridges over which we drive. But welding also makes smaller, more delicate functional and decorative items possible, like patio chairs and trellises, wine racks and candleholders, baker's shelves and headboards. The photos on these pages reveal only a fraction of what welding artisans have created with heat and metal.

The strength and malleability of steel makes it perfectly suitable for forming whimsical structures like this candelabra.

©Millet/Inside/Beateworks

Cutting and welding creates beautiful and useful household items, like this fireplace screen.

Photo courtesy of Sleeper Welding/Tom Sleeper

The costs involved in setting up a well-equipped home welding shop are comparable to setting up a well-equipped woodworking shop. A great advantage of welding is that the materials are generally inexpensive, yet you can create items that sell in shops for hundreds of dollars. Because fewer people are knowledgeable about welding, there is always a rewarding “awe” factor to showing off what you have created.

This book provides basic directions and step-by-step photos that illustrate the four major welding processes and the two major cutting processes. You also will find quick reference charts that describe electrode and filler wire choices, metals and their weldability, and joint and weld types. Suggestions for setting up a welding shop, detailed safety guidelines, and ideas for finishing your metal projects are also included.

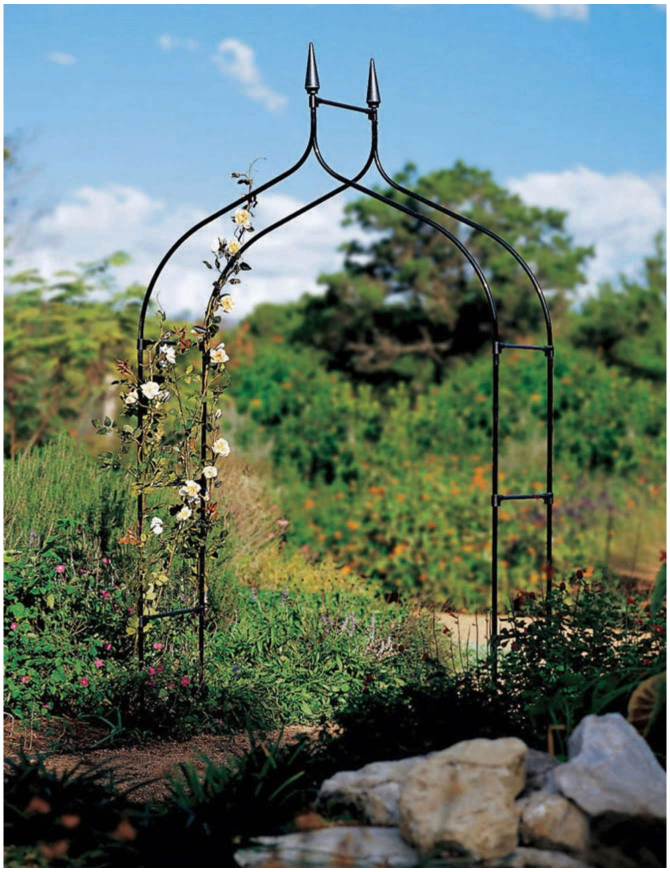

Simple and graceful arches and trellises are easily made from mild steel. A few bends, decorative finials, and a handful of welds are all it took to make this simple garden archway.

Photo courtesy of Gardeners Supply Company

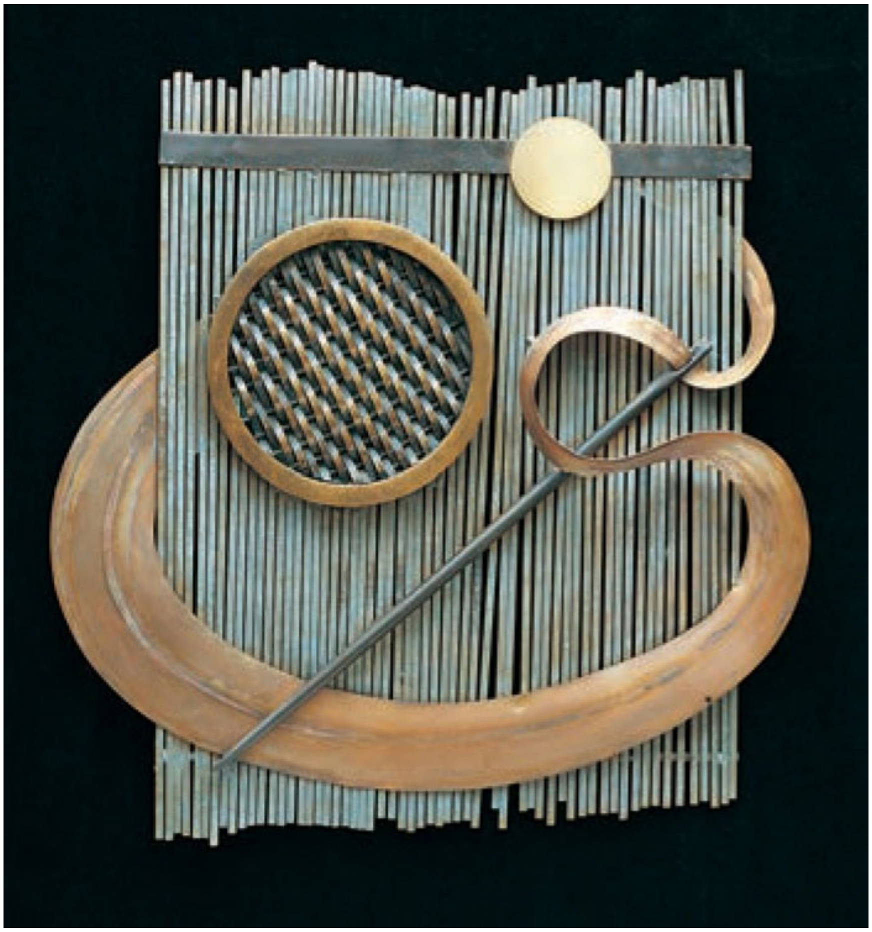

A variety of metals, shapes, and joining techniques were combined to create this beautiful artwork.

Photo courtesy of Gene Olson/Photo by Ytsma Photography

Best of all, this book includes detailed directions for 23 practical and decorative projects. You can make a welding and cutting table for your welding shop, a scrolled desktop lamp, a wrought iron railing, or a cute critter boot brush. You can practice your techniques and make accessories for your own home or as gifts. We also encourage you to use the project ideas to branch out and create your own versions of a baker's rack, coffee table, or kitchen accessory stand.

These colorful puppies are cut from junked car hoods.

Photo courtesy of Lizard Breath Ranch, Inc.

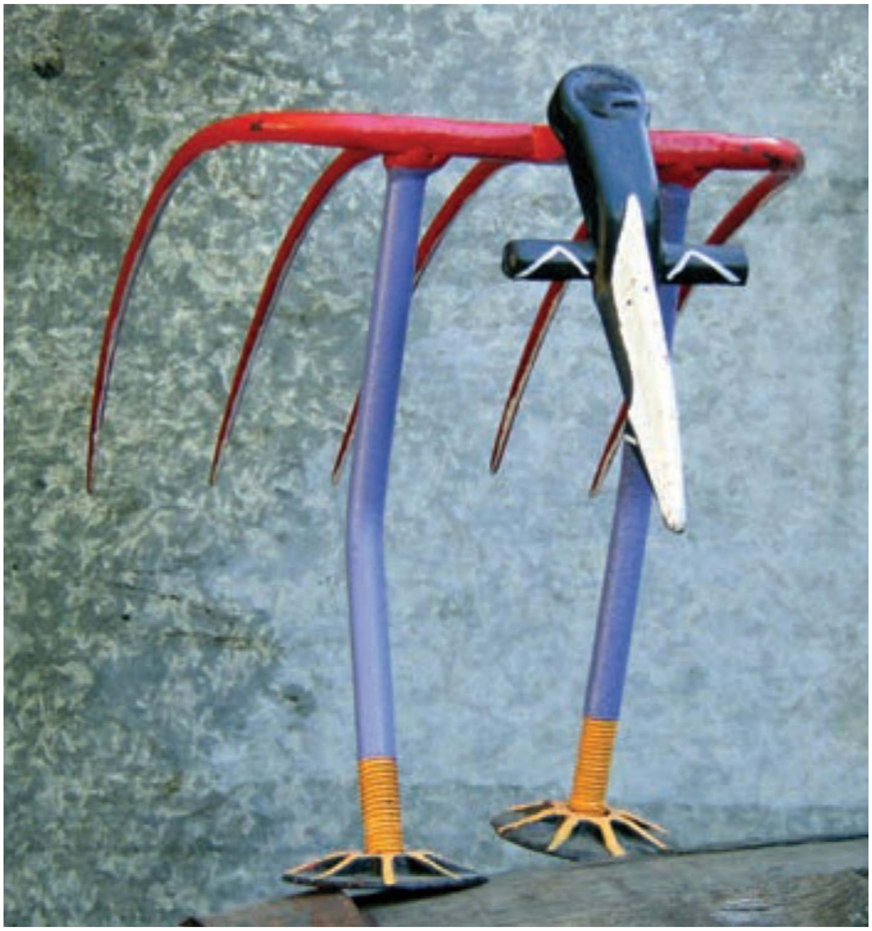

Creating fanciful creatures from garden tools is a rewarding welding project.

Photo courtesy of Heebie GBS Metalworks

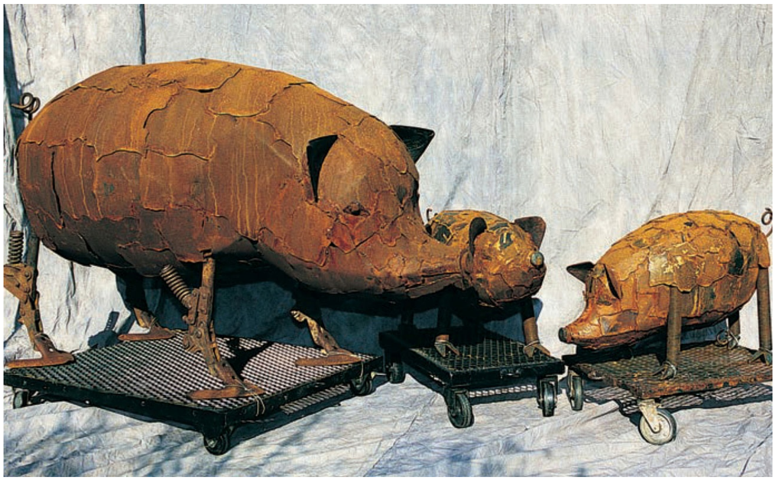

Scales of sheet metal, springy legs, and mesh trolleys—what a fun piggy collection.

Photo courtesy of Dave Johnson/Birds & Beasts

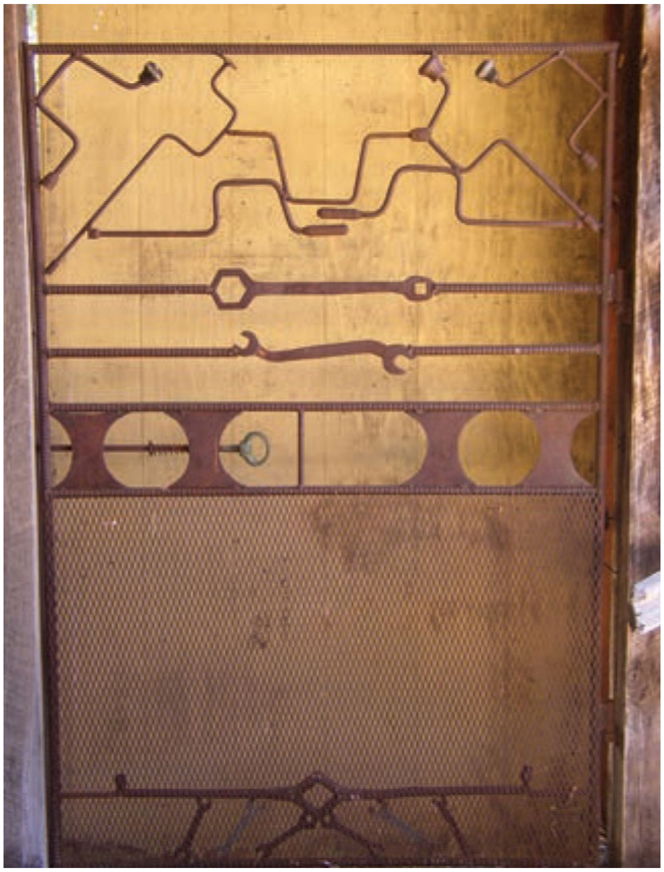

Old tools and recycled metal odds and ends make this gate geometrically pleasing.

Photo courtesy of Heebie GBS Metalworks

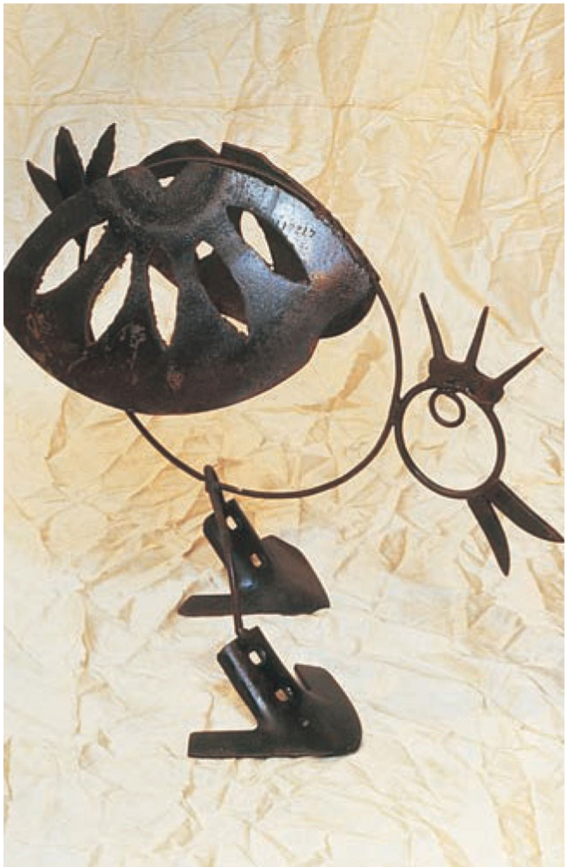

Old farm implements and spare parts are great sources of weldable material. This chicken shows what a little imagination can do with a pile of parts.

Photo courtesy of Dave Johnson/Birds & Beasts

The strength of steel allows for solid yet airy structures like this plant stand.

Photo courtesy of Gardeners Supply Company

NOTICE TO READERS:

Welding is a dangerous activity. Failure to follow safety procedures may result in serious injury or death. This book provides useful instruction, but we cannot anticipate all of your working conditions or the characteristics of your materials and tools. For safety, you should use caution, care, and good judgment when following the procedures described in this book. Consider your own skill level and the instructions and safety precautions associated with the various tools and materials shown. The publisher cannot assume responsibility for any damage to property or injury to persons as a result of misuse of the information provided.

Introduction to Welding

Welding is all about heat, about using heat to melt separate pieces of metal so they will flow together and fuse to form a single, seamless piece. Regardless of the welding or cutting process, your ability to control the heat generated by the flame or arc determines the quality of your welds and cuts.

Certain terms are used to describe the heat and action of all the welding processes. The parts being welded together are referred to as the base metal. Additional metal, called filler, is often added to the weld. The molten puddle is the area of melted base metal and filler metal that you maintain as you create your weld. To have fusion of metals, the base metal and filler metals must be the same composition. Methods for joining metal without fusion are called soldering, brazing, and braze welding. These methods can be used to join similar or dissimilar metals.

Oxyacetylene welding and cutting use flames to generate the heat to melt the metal. Shielded metal arc, gas metal arc, and gas tungsten arc welding and plasma cutting use an electric arc for heat generation. With oxyacetylene welding, you have time to watch the puddle develop as the metal turns red, then glossy and wet looking, then finally melts. With the arc processes, the puddle forms quickly and may be difficult to see because of the intensity of the arc. It is important to have ample lighting and clear vision so you can watch the puddle and move it steadily.

Oxyacetylene welding

Penetration of the weld is also a critical heat dependent factor. A strong weld penetrates all the way through the base metal. Matching filler material size and heat input to the thickness of the base metal is important. It is easy to get a good looking weld that has not penetrated the base metal at all and merely sits on the surface. At the opposite end of the spectrum from a “cold,” non-penetrating weld is burn through—where the base metal has gotten too hot and is entirely burned away, leaving a hole in the base metal and the weld.

Heat distortion is a by-product of all welding and cutting processes. It is obvious that applying a flame to metal in the oxyacetylene process makes the metal hot—and the electrical arcs are actually four to five thousand degrees hotter than the oxyacetylene flame. When metal is heated it expands; when it cools it contracts. If not taken into consideration, this expansion and contraction may cause parts to move out of alignment. This is why tack welding and clamping project pieces is critical to successful welding. It is also important to match the welding process to the base metal thickness. For example, shielded metal arc welding is generally not used on materials less than 1/8" thick. The process is too hot and too difficult to control on thin material. On the other hand, gas metal arc welding works well on very thin sheet metal, if you are able to adjust the machine to a low enough voltage.

Shielded metal arc welding

Speedglas
TILLMAN

Gas metal arc welding

Protecting the molten puddle from oxygen is also an important part of welding. Oxygen makes steel rust and causes corrosion in other metals. The exclusion of oxygen from the weld when it is in the molten state makes a stronger weld. This is accomplished in a variety of ways. A properly adjusted oxyacetylene flame burns off the ambient oxygen in a small zone around the weld puddle. Gas metal arc and gas tungsten arc welding bathe the weld in an inert gas from a pressurized cylinder. These inert gases keep oxygen away from the molten puddle. Flux cored arc welding and shielded metal arc welding use fluxes in or on the welding filler metal. When these fluxes burn, they produce shielding gases and slag, both of which protect the weld area until it has cooled.

Gas tungsten arc welding

Plasma cutting

Welding well is difficult and takes years to master, but it is possible to make many useful and decorative items with basic knowledge and a little practice. If you wish to move beyond the decorative projects outlined in this book, it is a good idea to talk with a welder and have him or her evaluate some of your welds. Remember, the safety of others is involved when you choose to make a utility trailer or spiral staircase. Take the time and make the effort to ensure that any structural project you make is safe.

This book is intended as a reference for people who have had some exposure to welding and need reminders about what steps to follow and precautions to take. It is not intended to teach welding to someone who has never handled welding equipment. Many community colleges, technical schools, and art centers offer welding classes. Such classes are an ideal way to learn the basics of welding, proper techniques for each process, and welding safety.

Oxyacetylene cutting

Safety

Welding can be a dangerous activity. Possibilities exist for getting cut, burned, electrocuted, and causing fires or explosions. Preparations for welding—grinding, burnishing, and sawing—are also dangerous. That said, it is absolutely possible, with a little care and diligence, to weld for years and never suffer more than a burnt thumb from being too eager to examine a just-completed weld.

It is very important to follow manufacturer's specific instructions and recommendations for equipment and product use and to follow general welding safety rules. Whether you are welding in a dedicated welding shop or have set up your own shop at home, you need to be aware of a number of specific safety concerns.

FIRST AID
Safety Manual

Welding safety includes protecting yourself with protective gear, following manufacturer's instructions, and being aware of the surroundings in which you are welding.

Fumes. Welding produces hazardous smoke and fumes. It is always important to keep your face out of the weld plume. Welding indoors requires either a ventilation fan, exhaust hood, or fume extractor. Wear an OSHA approved N99 particle mask or a respirator. If you are pregnant or plan to become pregnant, a respirator is a necessity.

Burns. Hot sparks and flying slag can cause burns. Protect your hands, head, and body with natural fibers—leather, wool, or cotton. Manufactured fibers such as nylon and polyester melt when ignited, which causes serious burns. Wear leather slip-on boots, and make sure your pant legs fall over the top of your footwear. Cuffed pants and pants with frayed edges are fire hazards. Also, welding sparks will find tiny holes, so don't wear holey jeans. Wear a welding cap and tie back long hair and keep it tucked under your shirt or cap.

Arc burn. Welding arcs produce ultraviolet and infrared light. Both of these can damage your eyes permanently, burn your skin, and potentially lead to skin cancer. A welding helmet with a filter lens protects your eyes and face, while long sleeve shirts and long pants protect your skin. Remember, you are also responsible for protecting the vision of curious neighbors, passersby, and pets. Use welding screens and have an extra helmet on hand for observers.

Fire. The area you're working in should be cleared of any flammable items such as lumber, rags, dropcloths, cigarette lighters, and matches. Absolutely do not weld or grind metal in a sawdust-filled shop. Sparks from welding and grinding can ignite airborne dust or fumes, or travel across the floor and come to rest against flammable materials. A fire extinguisher rated ABC is a must—mount it next to a first aid kit so you know where both are. Check your welding area one half hour after welding to make sure no sparks have found a place to smolder.

MINIMUM LENS SHADE NUMBERS FOR WELDING

APPLICATION	**SUGGESTED SHADE #**
Shielded Metal Arc Welding (SMAW)	
1/16 to 5/32" electrodes	10
3/16 to 1/4" electrodes	12

5/16 to 3/8" electrodes	14
Gas Metal Arc Welding (GMAW)	
1/16 to 5/32" non-ferrous	11
1/16 to 5/32" ferrous	12
Gas Tungsten Arc Welding (GTAW)	10 to 14
Plasma Cutting	8
Oxyacetylene Welding	5
Oxyacetylene Cutting	5
Brazing	3 to 5

Welding helmets are typically available with filter lenses in either 2" × 4-1/4" or 4-1/2" × 5-1/4" sizes. Helmets with auto-darkening lenses are also available. Full-face protective shields are available in clear, for grinding or chipping, and with a #5 filter for oxyacetylene operations.

Explosion. Even non-flammable gases such as carbon dioxide are stored in cylinders at such high pressure that they can easily become dangerous missiles. Keep the cylinder's protective cap on if regulators are not installed, and keep cylinders chained or strapped at all times. Cylinders should never be used as rollers or supports, and cylinders and protective caps should never be welded on. Shut acetylene and oxygen cylinder valves if you're away for more than 10 minutes. Cylinders must be transported right side up and chained, even if empty.

Another explosion hazard is concrete. Because of the amount of water contained in concrete, it can become a steam bomb if welded upon. Tack welding on concrete is acceptable, but using a concrete surface to back your welds is not. Use fire bricks, which have been cured to have very low water content. Never weld or cut on any closed container, tank, or cylinder. Even if it has been empty for years, it might have enough residual material to release toxic fumes or explode. Always bring tanks of any sort to a tank welding specialist for repair.

Other hazards include noise from grinding, sawing, sanding, and plasma cutting; laceration from sharp metal edges; electrocution during arc welding; and asphyxiation from inert shielding gases. Carefully read all manufacturers' instructions before starting any welding process.

A
C
D
B
N
M
J
K
L
I

Welding safety equipment includes: A. safety glasses, B. particle mask, C. low-profile respirator, D. leather slip-on boots, E. fire-retardant jacket, F. fire-retardant jacket with leather sleeves, G. welding cap, H. leather cape with apron, I. leather gloves with gauntlets, J. heavy-duty welding gloves, K. welding helmet with auto-darkening lens, L. welding helmet with flip-up lens, M. full-face #5 filter, N. full-face clear protective shield.

Setting Up Shop

If you plan to weld on a regular basis, it makes sense to set up a welding shop. The primary concern with welding is containing the hot sparks or slag and the flammable elements while exhausting dangerous fumes. It is possible to weld outside, but not all processes allow that. Gas metal and gas tungsten arc welding require that the surrounding air be still so the shielding gas is not disturbed. All the arc processes must be done in dry conditions to prevent electrical shock. Cold metals do not respond as well as metals at 70° F and may not be weldable using certain processes. A heated garage or outbuilding is best suited for a welding shop. A basement is not suitable due to the dangers of fire and compressed gases next to living spaces. Additionally, your homeowner’s insurance may not cover a welding shop if it is inside your living area as opposed to in a detached garage or shop building.

Shop space with a concrete floor and cement block walls is ideal for welding. Good ventilation is also important.

It is important to remember that tiny sparks and pieces of hot slag may scatter up to 30 or 40 feet from the source. If they come to rest on flammable materials, they may smolder, and given the right conditions, can ignite the material. Always check your welding area one half hour after you have completed welding to make certain no sparks are smoldering. A wooden table covered with metal is not a good work surface, as the transferred heat may cause the wood to smolder. Any wooden jigs or clamping devices should be doused with water to extinguish smoldering embers or stored outside after use.

Shop Tools

Power tools such as the reciprocating saw, angle grinder, portable band saw, and chop saw are useful when cutting and fitting metal parts to be welded. A drill press and metal cutting band saw are also useful tools for welding.

Specific metal working tools, including metal brakes and metal benders, are available for all sizes of metal. These can range from inexpensive sheet metal tools, such as the scroll bender shown on page 108, to expensive hydraulic tools.

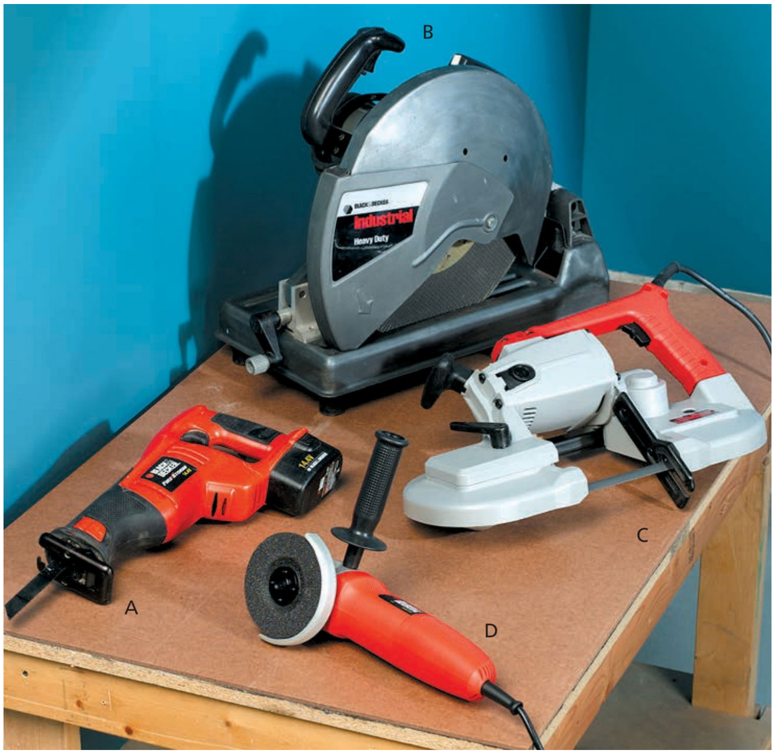

Power tools for a welding shop (above photo) include: A. reciprocating saw, B. chop saw, C. portable band saw, D. angle grinder.

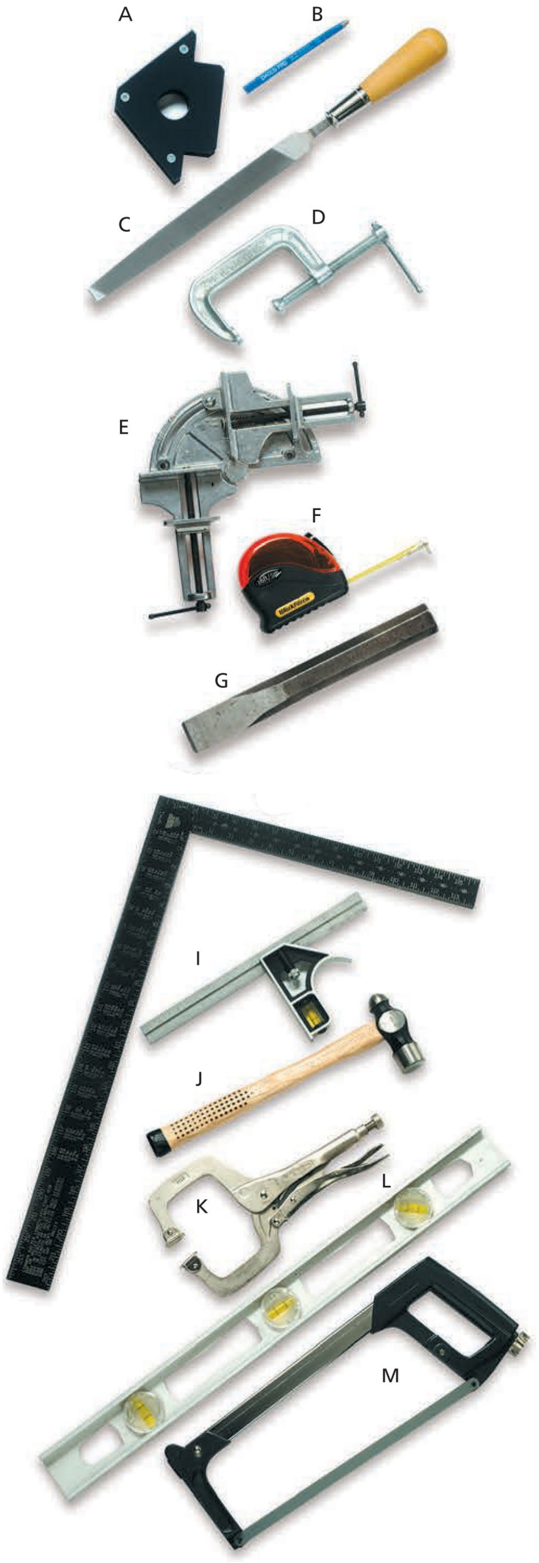

Standard tools for a welding shop include: A. magnetic clamp, B. center punch, C. metal file, D. C-clamp, E. corner clamp, F. tape measure, G. cold chisel, H. carpenter's square, I. combination square, J. ball peen hammer, K. clamping pliers, L. magnetic level, M. hacksaw.

Metal Basics

Weldability & Cutability

Different metals have different characteristics that affect their ability to be welded or cut with any of the processes described in this book. In general, only metals of the same type are welded together because welding involves melting the base metal parts and adding melted filler metal. In order to accomplish this, the parts and filler must have the same melting temperature and characteristics. Dissimilar metals can be joined by brazing and braze welding because these processes do not actually melt the base metal.

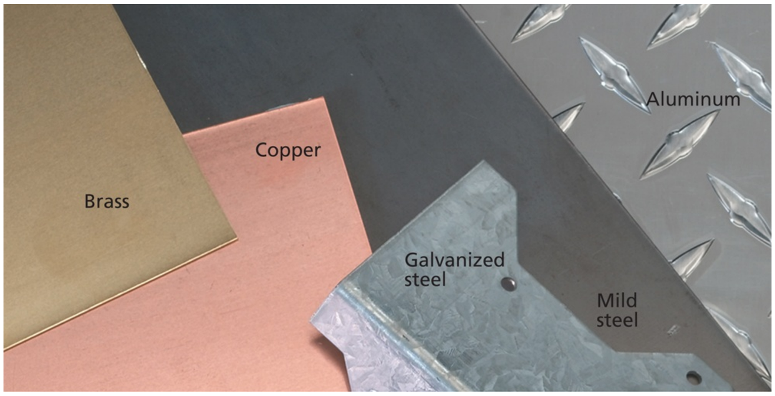
Aluminum
Copper
Brass
Galvanized steel
Mild steel

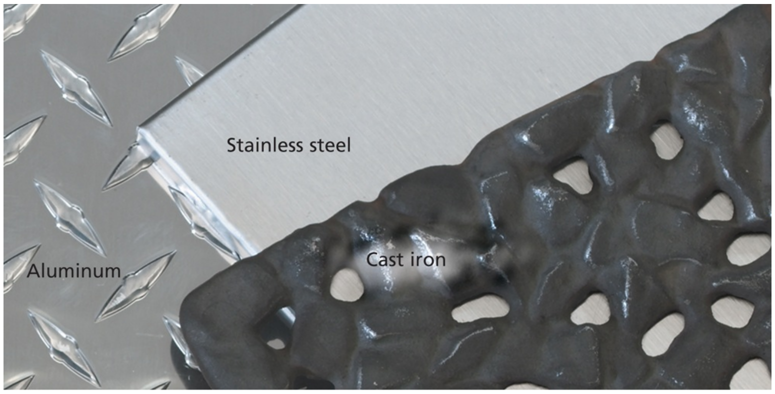
Stainless steel
Aluminum
Cast iron

Metals are divided into the categories of ferrous and non-ferrous. Ferrous metals contain iron and will generally be attracted to a magnet. Cast iron, forged steel, mild steel, and stainless steel all contain iron, along with varying amounts of carbon and other alloying elements.

Mild steel, which has a low carbon content, is the most commonly used type of steel and the easiest with which to work. Mild steel can be welded with all the processes covered in this book and can be cut with either oxyfuel or plasma cutting. It makes up most of the metal items you commonly use, make, or repair, such as automobile bodies, bicycles, railings, patio furniture, file cabinets, and shelving. Adding more carbon to steel makes it harder, more brittle, and more difficult to cut and weld. High carbon steel is called tool steel and is used to make drill bits and other metal cutting tools.

METAL	**WELDING PROCESS**	**CUTTING PROCESS**
Mild steel	All welding processes	Oxyfuel, plasma
Aluminum	Gas tungsten arc, gas metal arc	Plasma
Stainless steel	Gas tungsten arc, gas metal arc, shielded metal arc	Plasma
Chrome moly steel	Gas tungsten arc, oxyfuel	Plasma
Titanium	Gas tungsten arc	Plasma
Cast iron	Shielded metal arc, brazing	Plasma
Brass	Braze welding	Plasma

METAL MELTING POINTS

METAL	**MELTING POINT (F)**

Aluminum	1217°
Brass	1652-1724°
Bronze	1566-1832°
Chromium	3034°
Copper	1981°
Gold	1946°
Iron	2786°
Lead	621°
Mild steel	2462-2786°
Titanium	3263°
Tungsten	5432°
Zinc	786°

Other elements are added to steel to make a variety of alloys, such as chrome moly steel and stainless steel. These additions may make the steel non-magnetic. Because stainless steel does not oxidize (rust) easily, it is difficult to cut with oxyfuel. When welding steel alloys, the filler material must be matched to the alloy to get a high-quality weld.

In addition to alloying, many metals are heat-treated to improve characteristics such as hardness. Because welding and cutting would reheat these metals and destroy those characteristics, heat-treating is an important factor to consider. Many vehicle chassis parts, including motorcycles and bicycles, are made of heat-treated metals.

Aluminum is a widely used non-ferrous metal because of its light weight and corrosion resistance. Like steel, it is available in many alloys and is often heat treated. Aluminum is used for engine parts, boats, bicycles, furniture, and kitchenware. Various characteristics make aluminum difficult to weld successfully—it does not change color when it melts, it conducts heat rapidly, and it immediately develops an oxidized layer.

Metal Shapes & Sizes

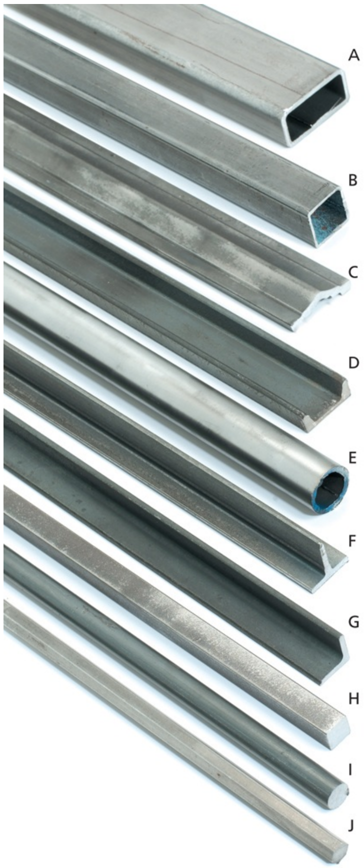
A
B
C
D
E
F
G
H
I
J

Mild steel (and most other metals) come in a variety of shapes, sizes, and thicknesses. Metal thickness may be given as a fraction of an inch, a decimal, or a gauge (see following chart).

Rectangular tube (A) and **square tube** (B) are used for structural framing, trailers, and furniture. Dimensions for rectangular and square tubing are given as width × height × wall thickness × length.

Rail cap (C) is used for making handrails. Rail cap dimensions are the overall width and the widths of the channels on the underside.

Channel (D) is often used for making handrails. Very large channel is used for truck bodies. The legs, or flanges, make it stronger than flat bars. Dimensions for channel are given as flange thickness × flange height × channel (outside) height × length.

Round tube (E) is not the same as pipe. Round tube is used for structures, and pipe is used for carrying liquids or gases. Dimensions for round tube are given as outside diameter (O.D.) × wall thickness × length. Pipe dimensions are nominal, that is, in name only. They are given as nominal inside diameter (I.D.) × length.

T-bar (F) dimensions are given as width × height × thickness of flanges × length.

Angle or **angle iron** (G) has many structural and decorative uses. Dimensions for angle iron are flange thickness × flange width × flange height × length.

Square bar (H), **round bar** (I), and **hexagonal** or **hex bar** (J) dimensions are given as width or outside diameter × length.

Flat bar or **strap** (not pictured) is available in many sizes. Dimensions are given as thickness × width × length.

Sheet metal (not pictured) is ^{3}/16" or less in thickness and is often referred to by gauge. **Plate metal** is more than 3/16" thick and is referred to by fractions or inches.

INCH EQUIVALENT FOR GAUGE THICKNESS

GAUGE	INCHES
24	0.020
22	0.026
20	0.032
18	0.043
16	0.054
14	0.069
12	0.098
11	0.113
10	0.128

Metal less than 1/8" thick is often referred to by gauge. For reference, the decimal equivalent of 1/8" is 0.125.

Purchasing Metal

Finding a metal supplier can be a challenging task. The materials that are readily available may not be the sizes and shapes needed for a project and may be expensive. Metal dealerships may not be friendly to small accounts. Because steel is so heavy, Internet or catalog shopping carries prohibitive shipping costs. With some searching, however, most necessary materials can be found.

Metal is generally priced by weight, unless you are purchasing it at a retail store. The price for small pieces of mild steel at a home center or hardware store work out to be as much as $2 to $3 per pound. The price per pound for small orders at a steel dealership may be in the range of 50 cents. Large orders or repeat orders may be priced as low as 30 cents per pound. Many metal suppliers have an odds and ends bin or rack where pieces may be as low as 10 cents per pound. Specialty metals such as stainless steel and aluminum start at $2 to $3 per pound.

Smaller sizes and shapes of mild steel and aluminum are available at home improvement and hardware stores in three-, four-, and six-foot lengths. Welding supply stores often have a selection of ten- and twelve-foot lengths of the commonly used sizes. Steel dealers have most common sizes in stock and will order other sizes for you. Most mild steel shapes and sizes come in twenty-foot lengths, and many steel dealers will make one free courtesy cut per piece. You might want to have the metal delivered, depending on the amount of material you are purchasing. You can order metal through catalogs or the Internet, but remember that shipping is expensive.

Some steel dealers are distributors for decorative metal products, but many specialty items such as wrought iron railing materials, decorative items, and weldable hardware are only available by catalog. A number of catalog supply houses sell to the public and have varied selections and reasonable prices. (See Resources, page 140.)

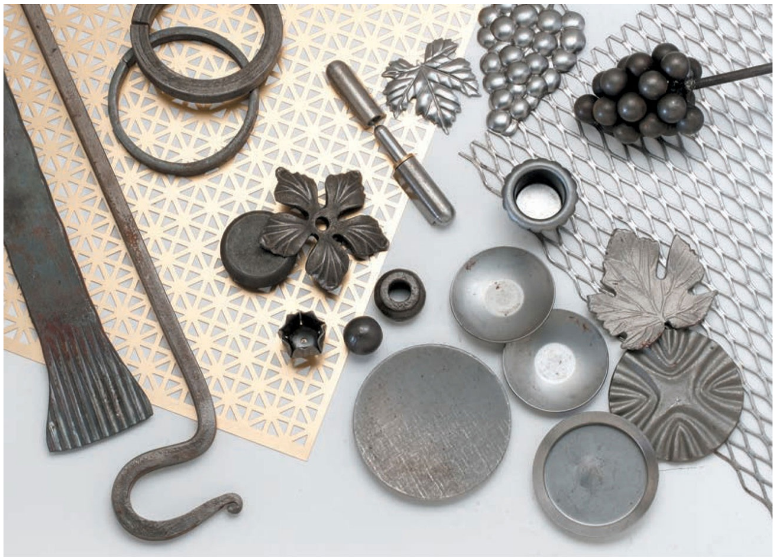

Sheet metal is available as pierced or expanded. Wall plates, hooks, rings, balls, bushings, candle cups, drip plates, and stamped or cast items are available in a wide variety of shapes, sizes, and metals.

Preparing Metal for Welding

A successful weld begins with a well-prepared piece of metal. The cleaner the pieces to be welded, the better the weld quality and appearance. When working with mild steel, it is possible to clean batches of parts ahead of time. When working with aluminum, parts need to be cleaned immediately before welding due to aluminum's nearly instant formation of a protective layer of oxidation. Mild steel is usually covered with mill scale and, often, oil or grease. Round and square tubes are usually thickly coated with oil, which helps in the manufacturing process. This oil can be removed with denatured alcohol, acetone, or a commercial degreaser.

6" THIN
BENCH G
WITH FLEXI

Cleaning the mill scale off mild steel is an important step. A bench mounted power wire brush works well on small pieces. Wire brushing the entire project prior to finishing is critical for good paint adhesion.

After the oil has been cleaned off, the mill scale can be removed by wire brushing manually or with a bench-mounted or hand-held power wire brush, with sandpaper or sanding screens, or with an angle or bench grinder. Aluminum and stainless steel need to be brushed with stainless steel brushes that are dedicated to cleaning only that metal. Small particles of mild steel will contaminate aluminum and stainless steel welds.

Finishing Metal

Though most commercially available metal items—patio furniture, indoor furniture, and garden accessories—have rough welds and spatter still present, it is nice to grind these down on your projects. An angle grinder is good for flat area welds, and a small rotary grinding tool can get into nooks and crannies. You’ll be pleased with the results of a nicely ground, good weld because it looks like a solid piece of metal. Wire brushing, sanding, or sandblasting the entire piece will further prepare it for finishing.

Painting metal projects is best accomplished by spraying, not brushing. You will need to thoroughly clean all oil, dirt, slag, and spatter from the project because even advanced rusty metal primers will not to stick to dirty, oily areas. Metal spray paints are available in various finishes and a huge array of colors.

Powder coating and brass plating are other, more expensive, finish options. Various antiquing and rusting finishes are available as well, or you can let the piece slowly rust on its own if you prefer this look.

An angle grinder works well for beveling edges on thick material. Grinding down welds gives projects a finished look.

Metal Shaping Techniques

There are a number of ways to bend and shape mild steel. Bending tools for creating right angle bends are called metal brakes. Rollers create circles, punches make holes, and shearers make cuts. You can purchase simple hand-powered metalworking tools, though they can be costly—the higher the quality or larger the capacity, the higher the price. (See page 108 for a photo of a scroll bender.) Powerful hydraulic metal shaping tools made for commercial use can cost thousands of dollars.

Simple bending jigs can be made using any round, rigid, strong form, such as pipe and salvaged flywheels, pulley wheels, or wheel rims. Using these jigs, it is fairly easy to bend round rod up to 1/4" in diameter and flat bars up to 1/8" thick into complex shapes. Rebar up to 3/8" can be bent into large curves by clamping it into a bench vise. Tubing can be bent using an electrical conduit bender. Thick material such as rail cap can be heated and bent.

You can bend materials more easily if you have a long end to apply pressure to, so you may want to cut pieces to length *after* bending. Whenever you shape metal into sharp angles, take into account the length needed for the radius of the bend. Also, cold-formed metal will spring back somewhat, so it may take trial and error to find what size jig will make the size bend you desire.

Use a sturdy hacksaw, jig saw, or reciprocating saw with bi-metal blades to saw metal. A miter box will help you make accurate cuts. A metal band saw, either portable or bench mounted, is handy if you will be making many projects that require numerous cuts. You can drill metal with a power drill or drill press, using metal cutting bits and oiling the bit often to prevent overheating. Large holes or holes in very thick material might be made more easily with an oxyfuel cutter or plasma cutter. An angle grinder or bench grinder can be used to create tapers or to correct miscut items to get better fit ups for welding.

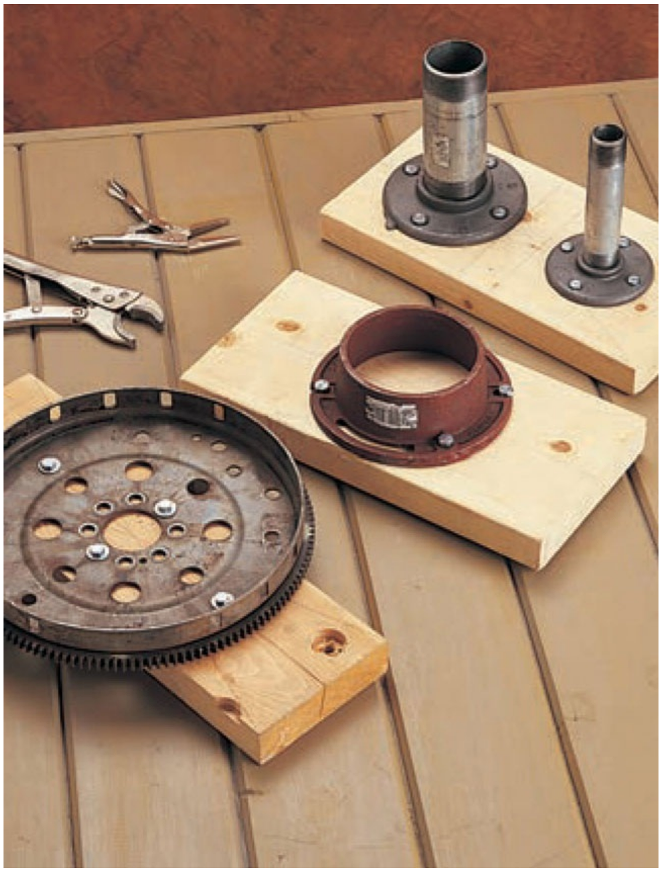

Bending jigs can be made from any rigid circular item. An engine fly-wheel, toilet flange, and various pipe sizes are shown here.

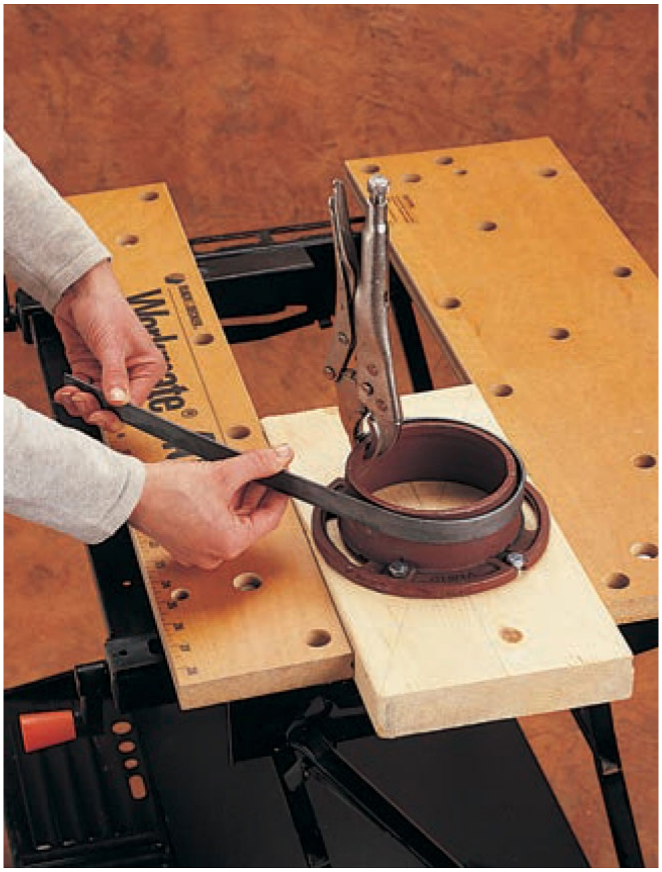

Use a vise style pliers to hold the metal to the jig. Wrap the metal around the jig in a smooth motion.

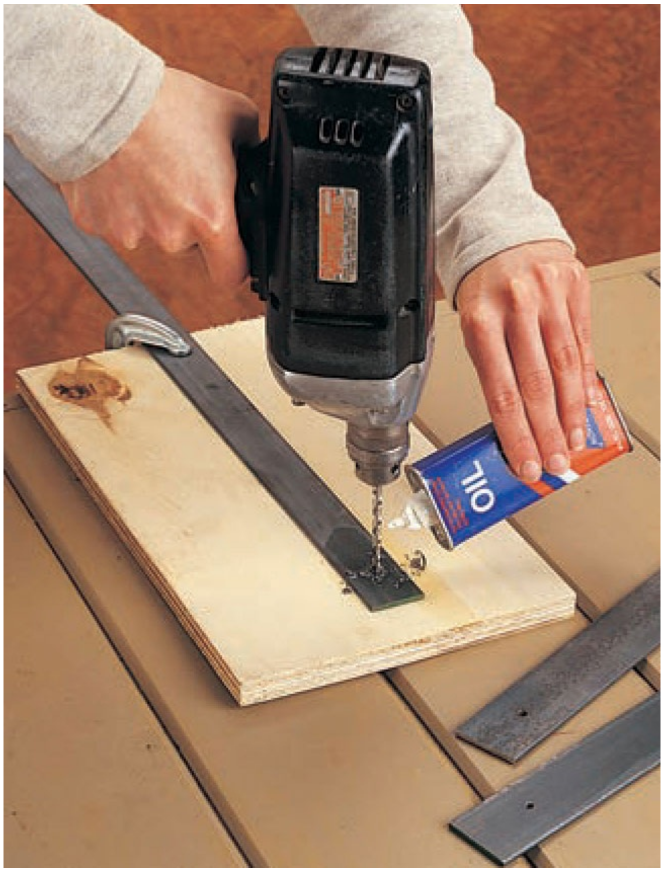

Use a slower drill speed and a metal cutting bit for metal drilling. Frequent application of oil prevents overheating.

Joint Types

The basic joints in welding are the butt joint, lap joint, corner joint, T-joint, edge joint, and saddle joint.

The **butt joint** is two pieces in the same plane butted against each other. The gap between the pieces is determined by the thickness of the pieces. Whether or not the surfaces need to be ground down (beveled) is also determined by the plate thickness. The type of weld used for a butt joint is a **groove weld.** The butt joint is a weak joint and should be avoided if at all possible.

A **lap joint** is two pieces in the same plane overlapped. The type of weld used for a lap joint is called a **fillet weld.** The strongest lap joint has welds on both sides.

A **corner joint** is two pieces coming together to make a right angle at a corner. The joint can be open, partially open, or closed. An open corner takes a fillet weld, but other corner joints may take groove welds. For hobby welding on metal 3/16" and thinner, an open corner joint is stronger and allows more penetration. A closed corner joint may be weakened if the weld is ground.

A **T-joint** is two pieces placed together to make a right angle T shape. Again, it is welded with a fillet weld. A T-joint is stronger if both sides are welded.

The **edge joint,** or **flange joint,** as it is sometimes called, is predominantly used to join thin sheet metal components. The turned up edge lessens the heat distortion to the thin sheets. With the advent of cooler welding processes, the need for the edge joint has diminished.

The **saddle joint,** or **fishmouth joint,** is used to join structural tubing. The tubes may join at any angle, and more than two tubes may be part of the joint. A fillet weld is used, most often all the way around the joint.

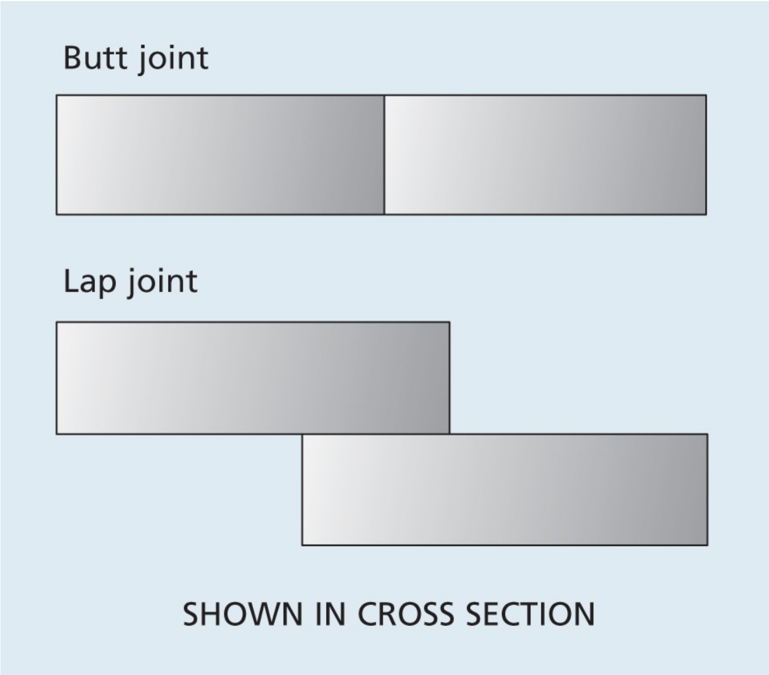
Butt joint
Lap joint
SHOWN IN CROSS SECTION

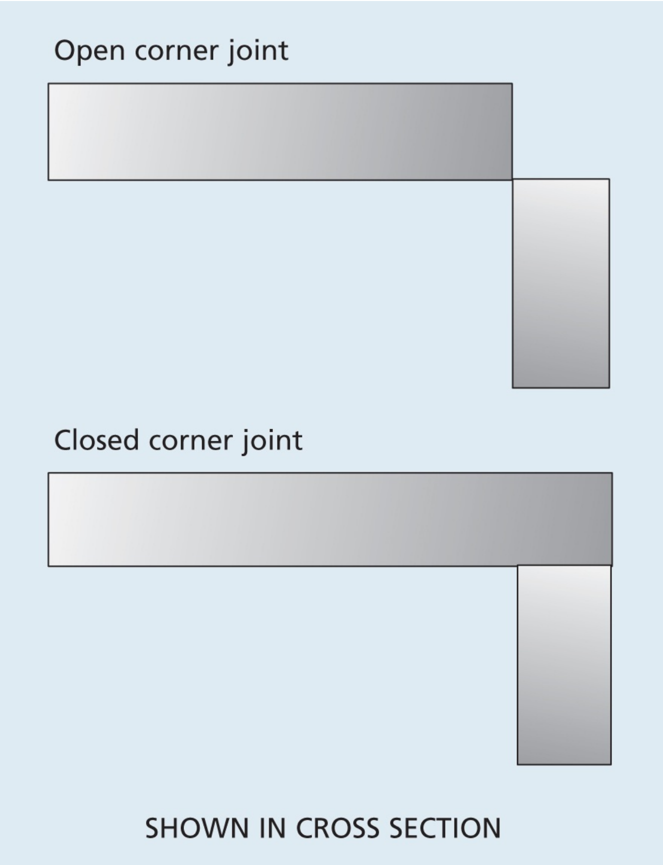
Open corner joint
Closed corner joint
SHOWN IN CROSS SECTION

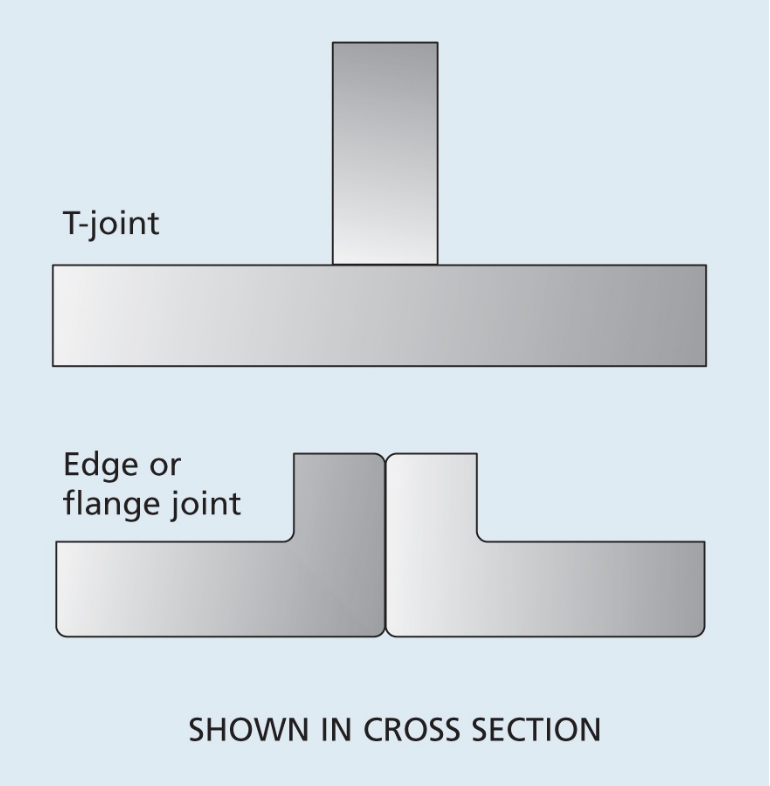
T-joint
Edge or flange joint
SHOWN IN CROSS SECTION

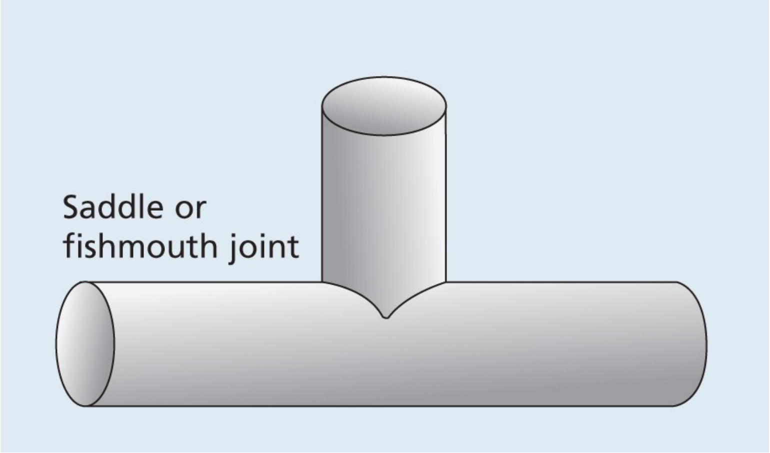
Saddle or fishmouth joint

Weld Symbols & Weld Types

Welding blueprints use specific symbols to denote weld types, locations, and other factors. The basic symbol consists of an arrow and a reference line with weld symbol. The weld symbol is placed above the reference line if the weld is located on the side opposite the arrow, and below the reference line for a weld located on the same side as the arrow. Weld symbols above and below the reference line indicate to weld both sides. A circle at the junction of the arrow line and reference line means to weld all around. Most hobby welders will not encounter these symbols.

The most common weld types used for home welding are the fillet and groove welds.

The **fillet weld** is roughly triangular in shape. It is made when welding most 90° angle joints. T-joints, open corner joints, lap joints, and saddle joints all take fillet welds.

The **groove weld** is made in a groove between pieces. The groove may be square grooved (straight sides), beveled (flat angled sides), or U shaped. The groove weld is used for butt joints, edge joints, and closed corner joints.

WELDING SYMBOLS

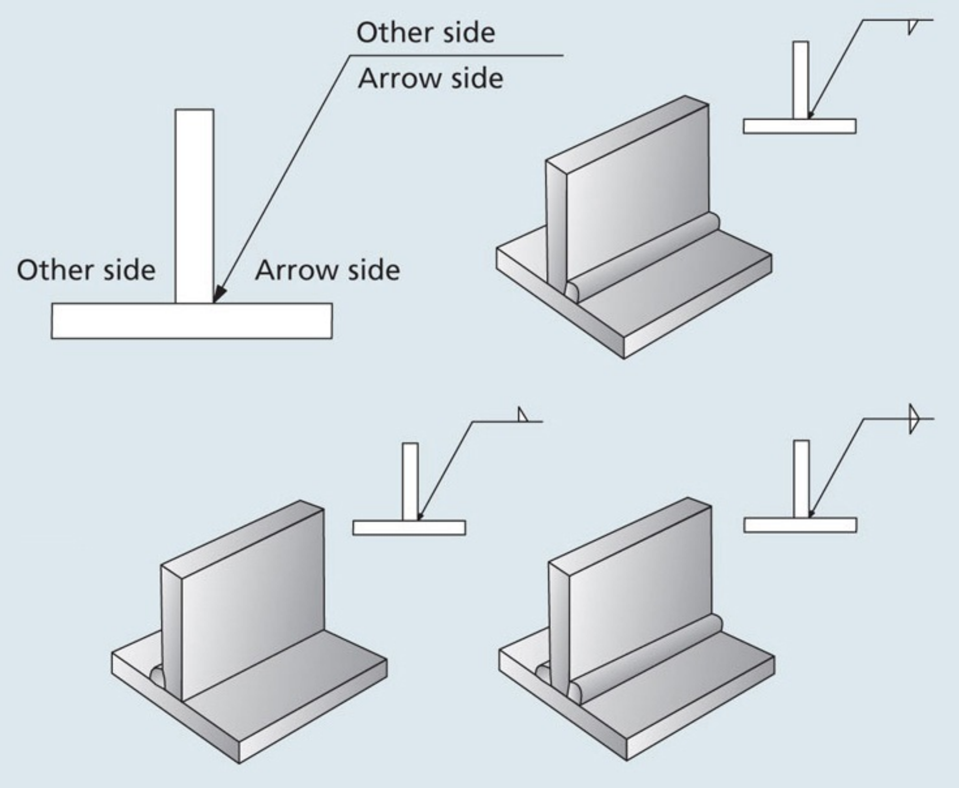
Other side
Arrow side
Other side
Arrow side

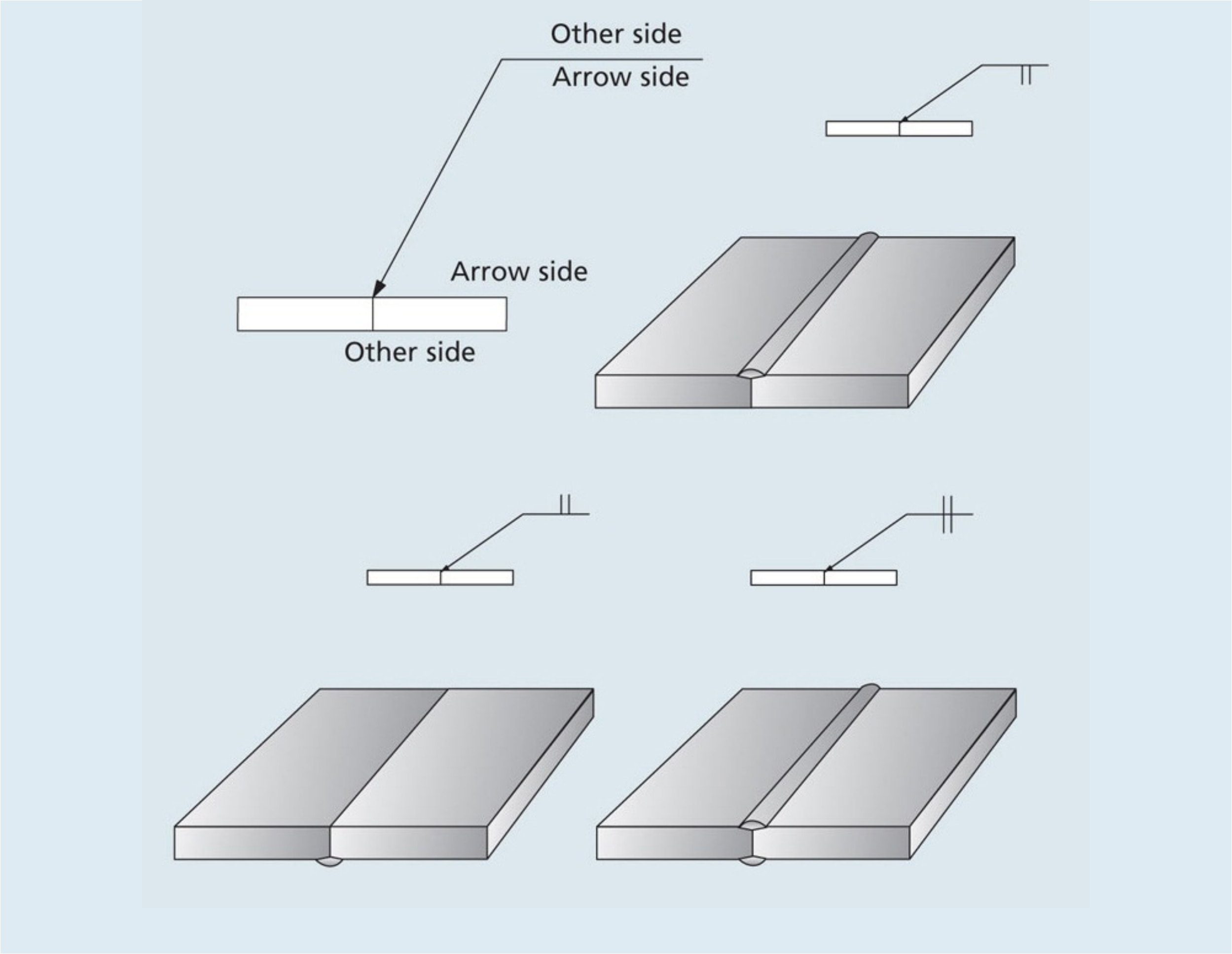
Other side
Arrow side
Arrow side
Other side

Repairing Metal

Once you begin welding, you will encounter numerous opportunities to repair items. When your friends and neighbors discover you can do repairs, even more challenges will come your way.

Performing welded repairs can be very tricky. It is important to assess your welding skills, the difficulty of the repair, and the intended use of the repaired item. Any structural or vehicle repairs, such as stairways, ladders, trailers, or truck chassis, need to meet the same safety standards as they did in their original condition.

The first step in considering a repair is determining why the item broke. If a weld was poorly executed, the repair might simply be to prepare the area and perform a good weld. If a piece has broken due to metal fatigue, simply “patching” over the crack will only cause more cracks outside the patched area. Cast iron and cast aluminum may have broken due to imperfections or inclusions in the casting, or they may have gone through rapid heating and cooling that caused cracking. Very often things break because they have been misused—which means your repair will likely be broken unless the behavior of the user changes.

All parts need to be carefully prepared before attempting a repair. Paint needs to be ground or sanded off and grease and oil need to be cleaned away. This cast iron part is being beveled to allow greater weld penetration.

The next step in a repair is determining what the base metal is. A magnet will be attracted to any metal with a certain concentration of iron in it, but stainless steel (which is sometimes non-magnetic), mild steel, and cast iron require very different welding techniques. Aluminum is non-magnetic and is discernable from stainless steel by its light weight—but which alloy is it? In the case of aluminum, some alloys are not weldable. Unfortunately, many manufactured metal items are made of alloys and may have been heat treated. Without access to the manufacturer’s specifications, it might be impossible to determine what the base metal composition is. It is important to understand the effects of welding on these materials before attempting a welded repair.

Once you have determined the feasibility of a repair on a particular piece, the piece must be prepared carefully. All rust, paint, and finishes must be removed from the area of the weld, and, if you are arc welding, from an area for the work clamp. Grease and oil must be cleaned off. If the break is on a weld joint, the old weld bead needs to be ground down.

Mild steel is the easiest material to repair. Simply prepare the material as for any other weld, and weld using any of the welding processes.

Cast iron and cast aluminum need to be preheated before welding to prevent cracking due to temperature fluctuations. If the piece is small enough, you can put it in the oven and heat it to 400 to 500° F. Otherwise, use an oxyacetylene or propane rosebud tip. Temperature crayons, which melt at specific temperatures, are available for marking metal to be preheated. Cast iron can be brazed if the fit between the broken parts is good, or it can be braze welded or welded with shielded metal arc. Specific electrodes for cast iron are available. Cast aluminum can be brazed, braze welded, or welded with gas tungsten or gas metal arc welding. Allow cast parts to cool slowly after welding.

Some aluminum is not weldable. If the piece to be repaired has been welded, you can perform a welded repair. If you can determine the composition of aluminum parts, match the electrode or filler rod to that. Some filler rods and electrodes are multipurpose and can be used on more than one alloy. Gas tungsten arc welding is the best choice for aluminum, but it can also be welded with gas metal arc, or it can be brazed.

Stainless steel comes in numerous alloys, and it is important to match the filler material to the base metal. Do not clean stainless steel with a steel wire brush, as the mild steel wires may contaminate the stainless steel. Stainless steel is best repaired with gas tungsten arc welding, although it can be brazed as well.

If you are unsure of any aspect of a potential repair, consult with your welding supply dealer or an expert welder.

Remove the paint and finish from the weld area. If you are arc welding, also remove the paint from an area close to the weld for attaching the work clamp.

Cast iron can be braze welded as shown here, or shielded metal arc welded with cast iron electrodes. Either way, the metal needs to be preheated to 450° F before being welded.

WELDING & CUTTING PROCESSES

The following chapters give basic background information on the four major welding processes and two cutting processes that home hobbyists are most likely to use. Each chapter includes the equipment needed for the process and safety precautions. How-to photographs with step-by-step directions show specific techniques for each process. Tip sections offer suggestions for improving skills and keeping equipment in good condition. A separate chapter on electricity covers some basic points that are important to all of the arc processes.

Oxyacetylene Welding 26
Electricity for Welding 38
Shielded Metal Arc Welding 40
Gas Metal Arc Welding 44
Gas Tungsten Arc Welding 54
Plasma Cutting 64
Oxyacetylene Cutting 68

OXYACETYLENE WELDING

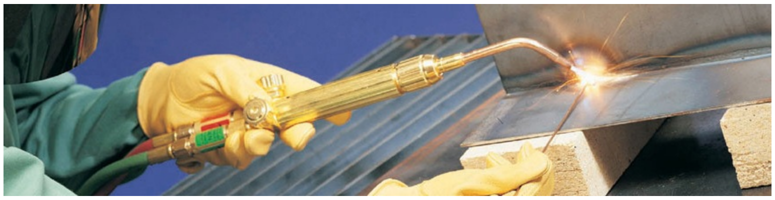

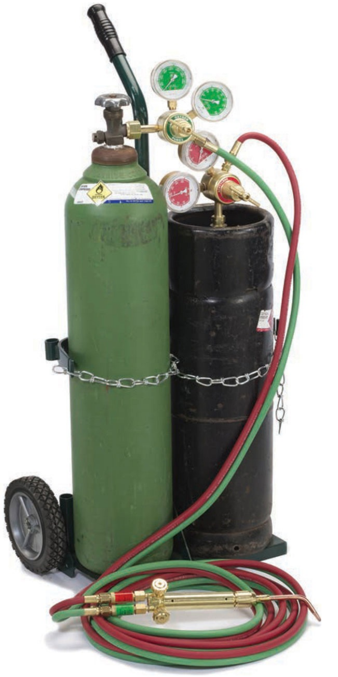

An oxyacetylene welding rig consists of two cylinders—oxygen and acetylene, regulators, hoses, and a torch with tip.

Oxyfuel welding is the process that uses the heat from a gas flame to melt base materials and cause them to join together. The gas flame is created by the combustion of oxygen and a fuel gas. Fuel gases are: acetylene, propane, butane, hydrogen, natural gas, and MPS (methyl acetylene-propa-diene, formerly known as MAPP gas). Oxygen and acetylene burn in a neutral flame at a temperature between 5600° and 6300° F, the hottest of any gas flame and capable of melting most metals. Other oxygen-fuel gas combinations are hot enough to use for soldering and brazing, but not hot enough for welding. Oxyfuel welding with acetylene is called oxyacetylene welding, but it is often simply referred to as gas welding.

The oxyacetylene process is versatile, as it can be used for both welding and cutting materials (see page 68) as well as heating, soldering, and brazing. It can be much less expensive than arc welding and is very portable because it needs no electrical power source. Oxyacetylene can be used to weld any thickness of metal, but sections over 1/4" are difficult to weld. Unfortunately, the techniques of oxyacetylene welding can be very hard to master, and there are serious safety concerns with the extreme flammability of acetylene and the high pressure gas cylinders.

Oxyacetylene welding equipment consists of an oxygen cylinder and an acetylene cylinder, each with regulators and gauges, oxygen hose, acetylene hose, torch, and tips.

Oxyacetylene welding may use fusion welding, where the base metals are melted together without filler material, or, more commonly, it may use a filler metal.

SAFETY

- Welding helmet or face shield with #5 filter
- Leather, wool, or cotton long pants, long sleeve shirt, and hat
- Leather gloves with gauntlet
- Leather boots or shoes
- Ventilation

THE EQUIPMENT

Cylinders. Oxygen comes in high pressure cylinders (sometimes called bottles or tanks) at 2200 pounds per square inch (psi) or more when full. The oxygen cylinder has a valve, and it always comes with a safety cap that screws on to the top to protect the valve. Never use the safety cap to lift an oxygen tank. The oxygen cylinder has a safety valve that will rupture to relieve tank pressure if the tank becomes overheated. Oxygen tanks should always be stored and transported in the upright position.

Acetylene. Acetylene is an extremely unstable gas. It cannot be pressurized above 15 psi in a free state. To safely pressurize the acetylene to 225 psi, the acetylene cylinder is filled with a porous material and acetone. When pressurized, the acetylene is absorbed by the acetone, which stabilizes it. When the pressure is released (the valve opened) the acetylene bubbles out of solution. Acetylene cylinders have fusible plugs that will melt in case of a fire and allow the gas to slowly escape rather than explode. Acetylene cylinders should always be stored in an upright position. Using an acetylene tank at an angle will allow acetone through the regulator and hoses, which will destroy them.

Cylinders should always be kept upright and secured with a chain or strap to prevent them from falling over. Never use a cylinder as a roller for moving heavy objects or for any purpose other than their intended use.

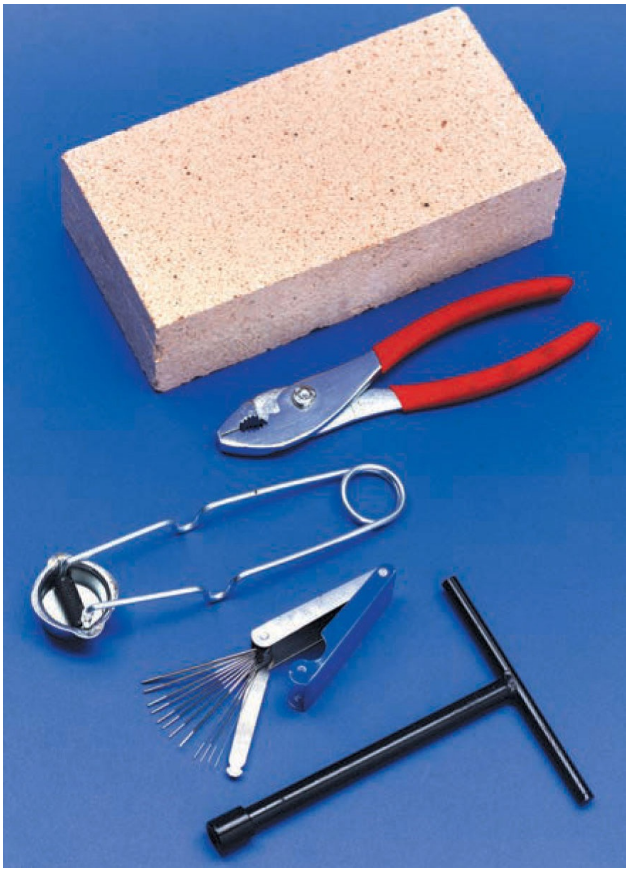

Tools for oxyacetylene welding include fire bricks, pliers, striker, tip cleaner, and acetylene tank wrench.

Cylinder Valves. Each type of cylinder has a valve that controls the flow of gas from the cylinder. A handwheel or valve wrench is used to open the valve. The oxygen cylinder valve should always be opened fully, as it is a backseating valve. A backseating valve is leakproof when closed and fully opened, but not otherwise. The acetylene valve should never be opened more than 3/4 to 1 1/2 turns. Less than this may lead to insufficient amounts of fuel and backflash; more than this makes it difficult to turn the acetylene off quickly in case of an emergency. If the acetylene valve is opened all the way, or opened suddenly, the acetone may escape from the cylinder, just like an overflowing soda can. Acetone can destroy the hoses and regulator fittings.

NOTE:

Never use oil or any other type of lubricant on regulators or other welding equipment. The combination of oil and oxygen in a spark-filled environment can lead to a fire in the regulator, which could cause a cylinder explosion.

Pressure Regulators and Gauges. Pressure regulators reduce the pressure of the gas leaving the cylinder. A single-stage regulator reduces the cylinder pressure to working pressure in one step. A two-stage regulator reduces the cylinder pressure to working pressure in two steps. A two- stage regulator is better because it gives more precise control over gas flow. but it is substantially more expensive. The regulator adjusting screw turns clockwise to increase pressure and counterclockwise to reduce pressure. The regulator is completely off when the adjusting screw is loose or “backed out.” Always back out adjusting screws as part of your post-welding routine.

Hoses. The flexible rubber hoses that move the gas from the regulator to the torch are designed to be leakproof and withstand high pressure. The oxygen hose is green, and the acetylene or fuel hose is red. Hoses are available either as a single hose or a dual hose where the oxygen and fuel hoses are paired together. The dual hoses are more common and more convenient. Always protect hoses from damage by moving them off the floor when not in use, keeping them behind the welding area so they are not burned with sparks, and keeping them out of traffic lanes so they are not stepped on or run over by vehicles. Hoses should not be allowed to come in contact with oily surfaces and should be protected from sunlight and chemical fumes to keep them in good condition. Always drain hoses when you have finished your welding session.

OXYGEN
ACETYLENE

Oxygen and acetylene regulators.

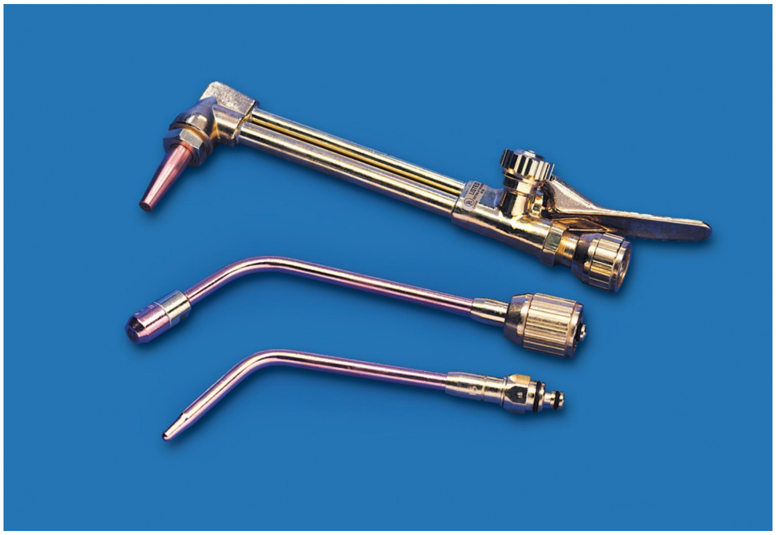

Available tips for an oxyacetylene torch body include cutting head, rosebud heating tip, and welding tip.

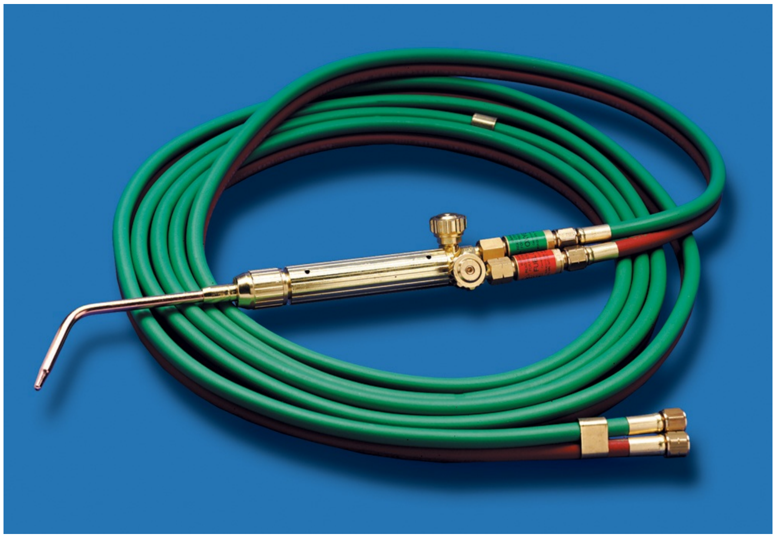

Oxygen and acetylene hoses with flashback arrestors, torch body, and welding tip.

Fittings. Fittings connect the hoses to the regulator at one end and the torch at the other. For safety, the nut for the fuel hose (red hose) is left-hand threaded and has a groove machined around the nut. The oxygen (green) hose nut is right-hand threaded and has no groove. Never interchange the oxygen and fuel hoses or fittings. Never force a fitting—brass is a very soft metal, and it is easy to damage the fitting threads. Always hand thread and hand tighten fittings before using a wrench, and be careful not to overtighten them. Never use pliers to tighten fittings, as they will damage the brass nuts.

Check Valves and Flashback Arrestors. Check valves and flashback arrestors are two safety features to prevent reverse gas flow or flashbacks. The check valve allows gas to flow from the hose to the torch. If gas pressure within the torch exceeds the hose pressure, a spring closes the valve to prevent back flow. In the event of a flashback, the check valve needs to be replaced. A flashback arrestor is installed between the torch and hose, and offers more protection than the check valve. The flashback arrestor prevents burning oxygen and fuel from flashing back into the hoses and regulator, which could cause an explosion. The flashback arrestor consists of a check valve, pressure-sensitive valve, stainless steel filter, and heat sensitive check valve. In the event of a flashback, the flashback arrestor does not need to be replaced.

Torch. The torch mixes and controls the flow of the fuel gas and oxygen. The torch consists of valves, torch body, mixing chamber, and tips. The torch is sometimes referred to as a blowpipe, but this is not technically correct.

Tips. Tips attach to the torch body and come in many sizes to create different size flames. A cutting tip, also called a cutting head or cutting torch, is attached to the torch body to use for flame cutting. A cutting tip has a number of holes for the preheating flames around a center hole for the pure oxygen. Cutting tips also come in a variety of sizes and with different numbers of preheating holes. Heating, or rosebud, tips are used to preheat metals to improve their weldability. Welding tips come in a wide range of sizes to match the thickness of the metal being welded. Matching the tip size to the welding material and gas pressure is critical for creating quality welds.

Setting Up an Oxyacetylene Outfit

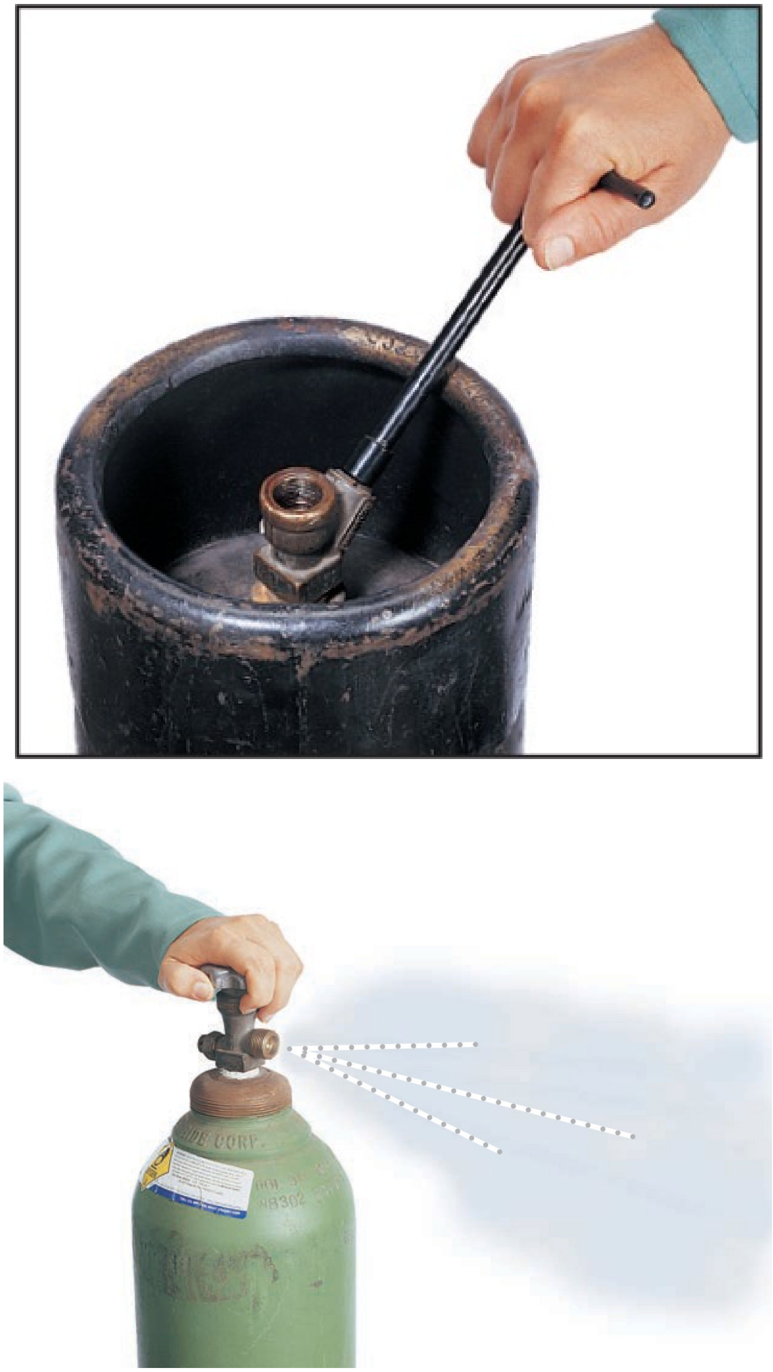

1 Secure the cylinders in an upright position, chained to a cart or strapped to a wall or post. Remove the protective cylinder caps. Wipe off the cylinder valve seats, regulator connections, and hose connections with a clean cloth. Crack open each cylinder valve briefly to expel any trapped dirt particles. “Flat top” acetylene cylinders (inset) may have antifreeze in the recessed valve seat. Use a clean rag to remove the liquid and dry the valve seat. This style acetylene cylinder requires a cylinder wrench to open the valve.

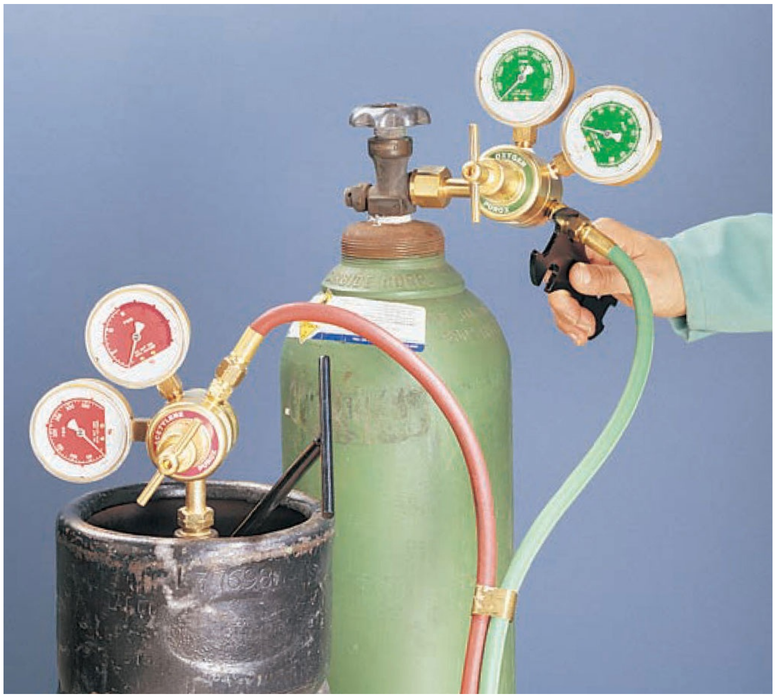

2 Attach the regulators to the cylinders. (The acetylene connectors have left-hand threads.) Always hand tighten, then use a fixed wrench, not a pliers or an adjustable wrench, to tighten. Do not overtighten—a firm seating is all that is necessary. Attach the hoses to the regulators. The acetylene hose is red and left-hand threaded. The oxygen hose is green. NEVER use grease, oil, or pipe dope to lubricate fittings. Grease and oil can ignite spontaneously when they come in contact with oxygen—even without a spark or flame present.

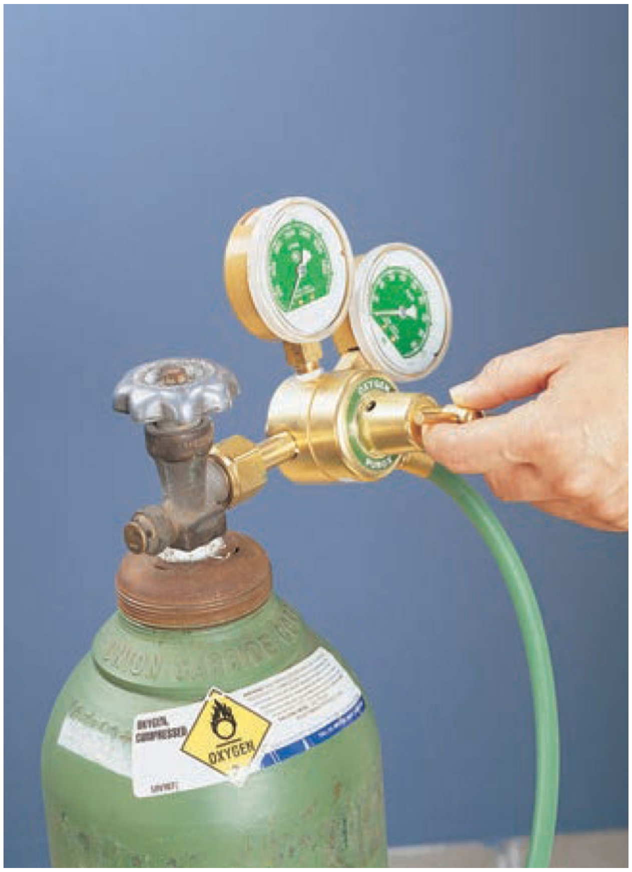

3 **Turn the regulator adjustment screws on the oxygen and acetylene regulators counterclockwise until they are loose. (Some regulators may have a knob.)**

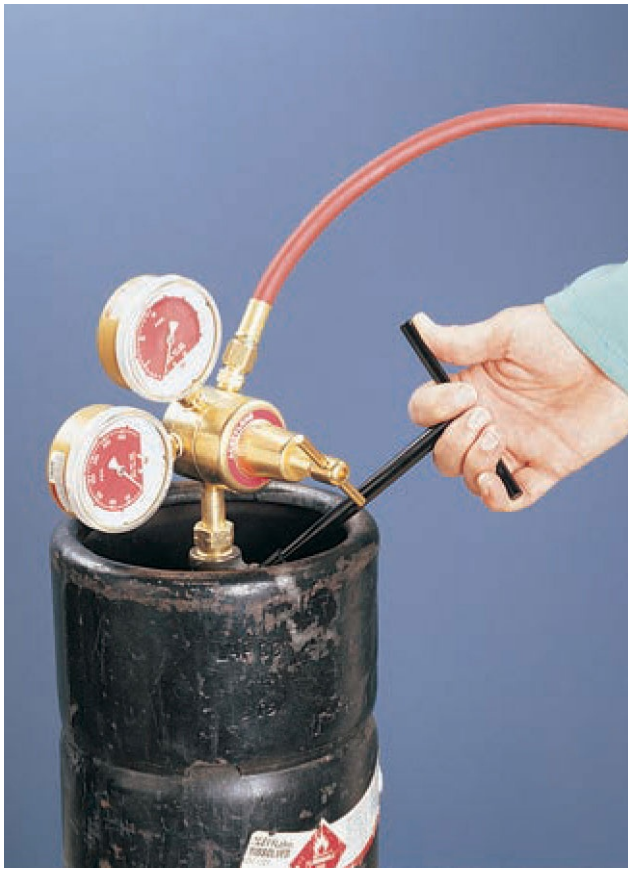

4 **Open the oxygen valve slowly all the way, while standing to the side in case the regulator gauge glass shatters. Turn the regulator adjustment screw until oxygen begins to flow through the hose, then loosen the regulator adjustment screw to stop the oxygen flow. Slowly turn the acetylene cylinder valve 3/4 to 1 1/2 turns.**

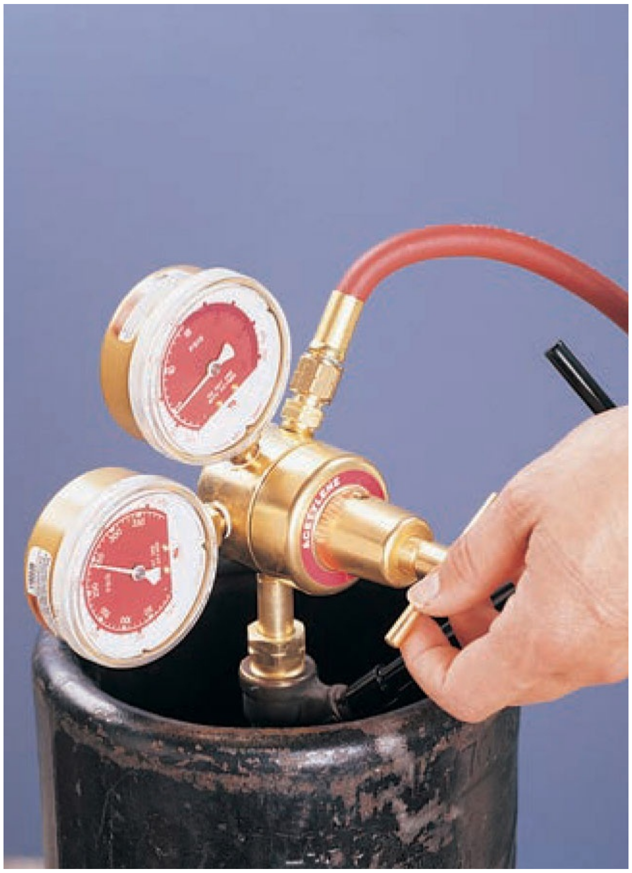

5 Adjust the acetylene regulator valve until the gas begins to flow, then loosen the regulator adjustment screw to stop the flow. Attach the torch to the hoses. Be sure to pressurize the system and check for leaks before lighting.

Pressurizing (turning on) an Oxyacetylene System

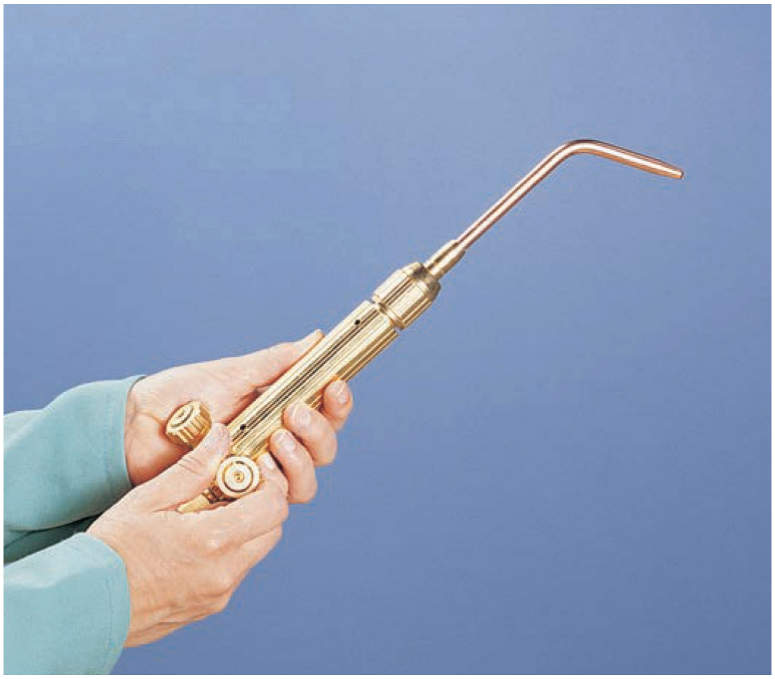

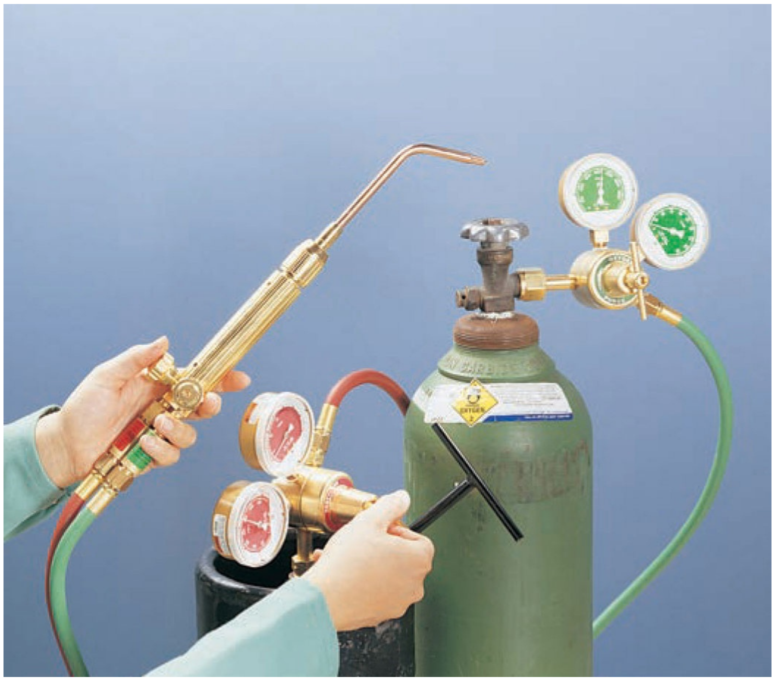

Make sure both torch valves are off. Turn both regulator adjustment screws counterclockwise until loose. Slowly turn the oxygen cylinder valve on. Once open, turn the valve all the way open to ensure proper seating without leaks. Turn the oxygen regulator adjustment screw clockwise until the gauge reads the desired pressure. (Refer to the manufacturer's specific instructions for operating pressures.) Open the oxygen valve on the torch to check the flowing oxygen pressure. Adjust if necessary and close the torch valve. Slowly open the acetylene cylinder valve 3/3 to 11/1 turns. (Leave the wrench on the valve if it is a wrench-style valve.) Turn the acetylene regulator pressure adjustment screw until the desired pressure reads on the regulator gauge. Open the acetylene valve on the torch briefly to make sure the flowing pressure matches the desired working pressure. If not, adjust the regulator until the proper pressure is reached.

Checking for Leaks

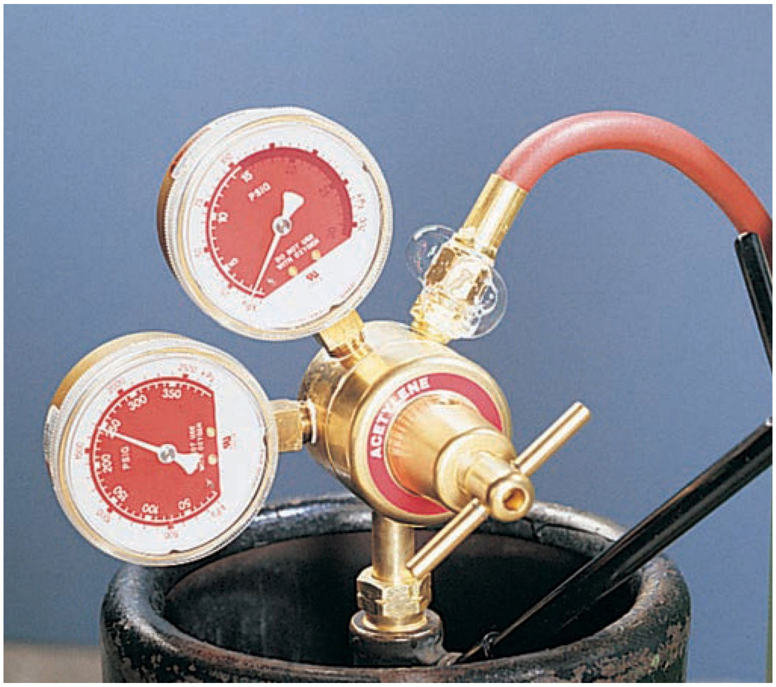

Apply leak-detecting solution to all connections with a small brush. (You can use soap and water so long as the soap is not petroleum based.) If any connections cause bubbles in the solution, tighten the connections and check again.

BACKFIRE & FLASHBACK:

Backfire and flashback are two hazardous situations that can be caused by improper gas pressures. Backfire is the pre-ignition of the acetylene and oxygen inside the tip that causes a popping sound. This may damage the tip or spray molten metal from the weld area. Flashback is the flame burning backward into the torch or hoses, causing a popping or squealing noise. Flashback can cause an explosion in the hoses. Avoid both hazards by matching the tip size to the material being welded and by using the proper pressure settings. Using lower pressures than recommended can cause backfire and possibly flashback.

Pre-lighting Checklist:

- Make sure torch valves are closed.
- Turn both regulator adjustment screws counterclockwise until loose.
- Slowly turn on the oxygen cylinder valve all the way. Turn the regulator adjustment to the proper pressure.
- Slowly turn the acetylene cylinder valve 1 1/2 turns. Turn the regulator adjustment valve to the proper pressure.

Some welders have been taught to open both the acetylene and oxygen valves before lighting the torch to avoid the smoky acetylene flame. This practice is no longer recommended.

NEVER light a torch with a match or butane lighter.

Lighting the Torch

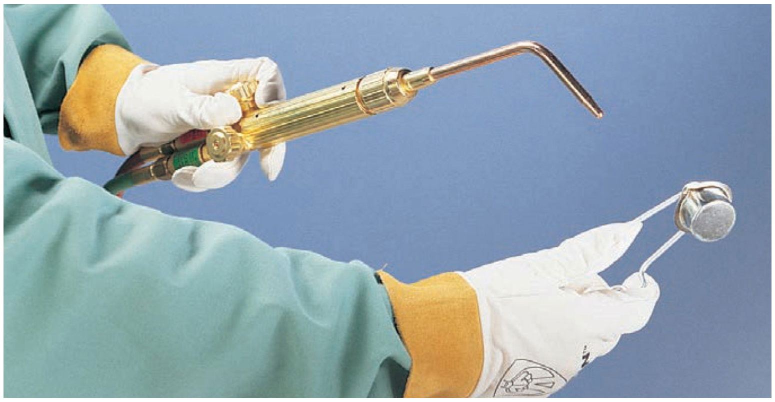

1 Hold the torch in one hand with the thumb and forefinger on the acetylene torch valve. Hold the striker in front of the torch about 3" to 6" away at a slight angle. Turn on the acetylene torch valve 1/1 to 1/1 turn.

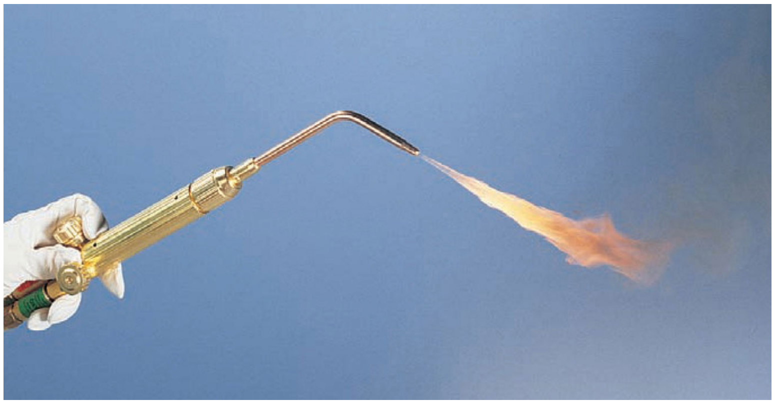

2 Immediately use the spark lighter to light the flame. The flame will be yellow and smoky.

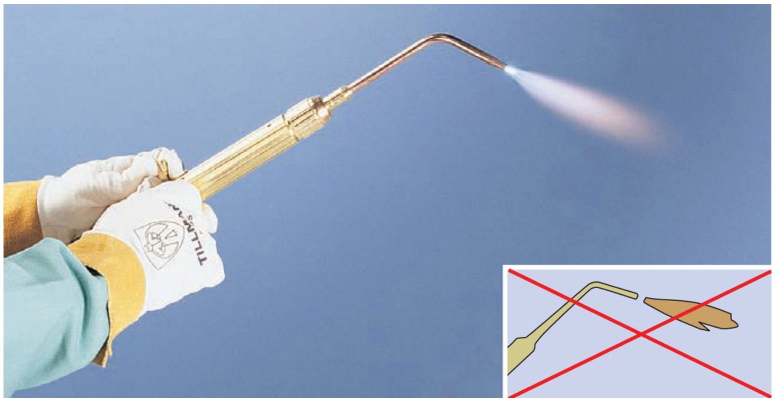

3 Put the striker down and adjust the acetylene torch valve with your right hand so the flame is burning without producing soot. The flame should not be separated from the torch (inset). Open the oxygen torch valve slowly. Adjust the oxygen to get a bright white inner flame and a bluish outer flame. Turn down the

acetylene to eliminate the excess acetylene feather if present. When you have finished welding, turn off the oxygen first, then the acetylene.

Flame & Flame States

The flame of an oxyacetylene torch has two parts—the inner or primary flame and the outer or secondary flame. The flame has different temperatures at different locations. The outside edges are cooler because they are burning with the ambient air, which is only 21% oxygen. The torch tip is cooler because complete combustion hasn't been reached. The hottest area is the tip of the primary or inner flame cone. Here the gases are completely combusted and are insulated by the secondary flame. There are three flame states for the oxyacetylene flame: carburizing, neutral, and oxidizing.

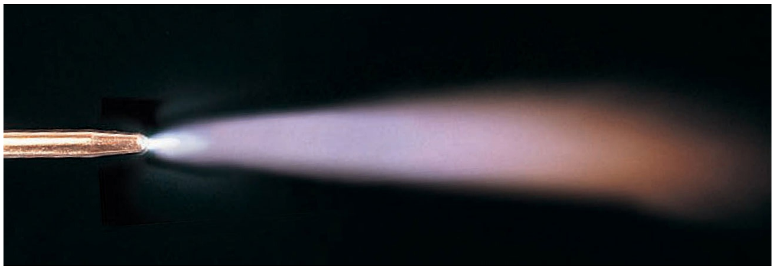

The carburizing or reducing flame has an excess of fuel. This is a useful flame as it will break down metal oxides to get at the oxygen, thus cleaning the weld area to a small degree. This process adds carbon to welds, which makes them harder. The carburizing flame has a bright white primary flame, an acetylene "feather" around the primary flame, and a bluish white secondary flame with orange edging.

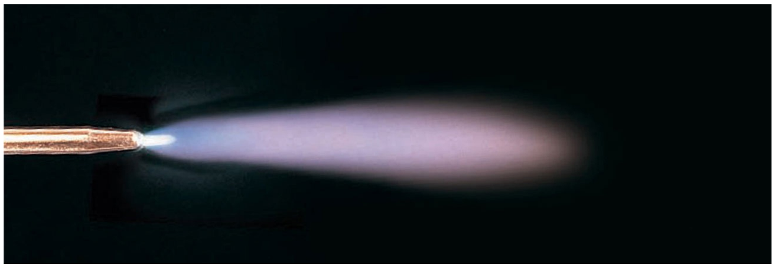

The neutral flame is the exact point where the feather and the inner cone come together. In this flame, there is exactly enough oxygen present to provide total combustion of the fuel gas. Most welding and cutting operations use a neutral flame. The neutral flame has a bright white primary flame and a colorless to bluish secondary flame.

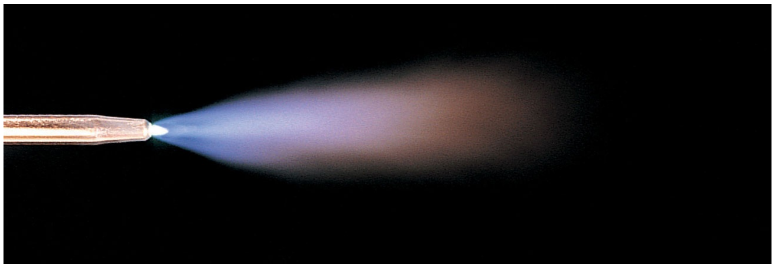

The oxidizing flame has an excess of oxygen. The white cone of this flame is small and pointed and somewhat paler than the neutral flame. A hissing sound often accompanies this flame. This flame is not particularly useful as it hastens oxidizing, which is not desirable in welding. It can, however, be used for removing carbon from molten metal, thus softening the metal.

Pre-welding Checklist:

- Check hoses for damage before pressurizing the system.
- Prepare metal for welding by wire brushing or sanding off mill scale and rust. Use acetone or denatured alcohol to remove oil or other chemical residues.
- Use fire bricks to avoid unnecessary heat loss and prevent welding to the welding table.
- See manufacturer's recommendations for appropriate tip sizes and gas pressures.
- Set up materials and clamp if necessary.

NOTE: The directions for oxyacetylene welding are for right-handed welding. Reverse the directions for left-handed welding, or if you find it easier to manipulate the filler rod with your right hand.

Welding with Oxyacetylene

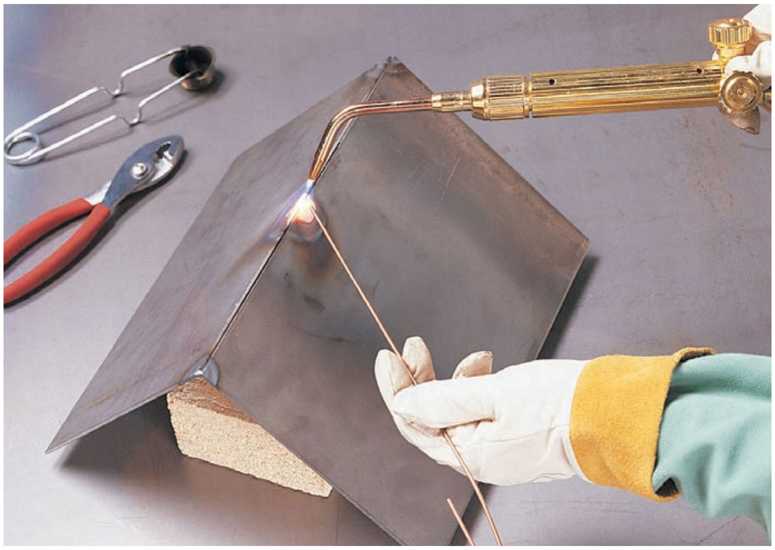

1 Select an appropriate filler rod and lay it on the table next to the bricks. Light the torch and adjust to a neutral flame. Pull down your face shield. Place small fusion tack welds at each end of the joint and in the middle if it is a long joint. (A fusion weld uses no filler rod.) Turn off the torch, oxygen first then acetylene, and check that your tacked piece is still in the desired position. If not, use a hammer to move it into position or break the tack weld and reposition.

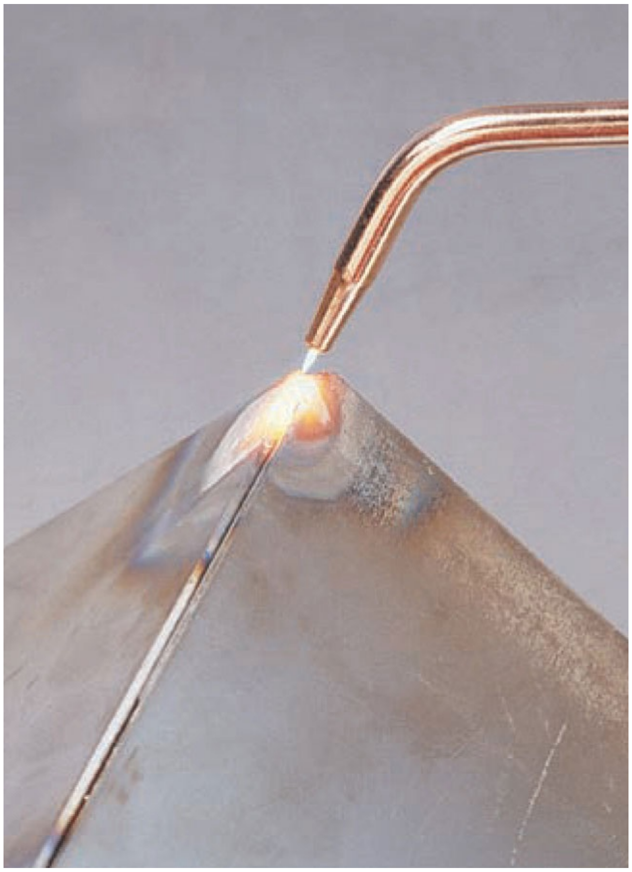

2 With the torch at a 45° angle to the right and oscillating the torch in a 1/4" to 1/2" circle over both metal pieces, create a weld puddle at the right end of your workpiece.

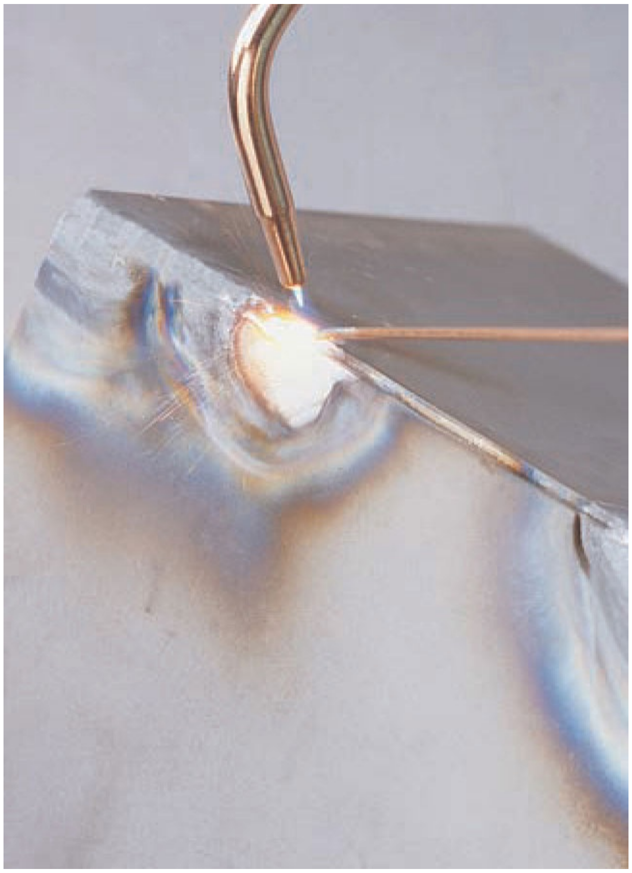

3 When a molten puddle has formed, move the filler rod close to the puddle and flame, but not in it. Begin moving slowly to the left while oscillating and maintaining the molten puddle. Dip the filler rod into the middle of the molten puddle and remove it, but keep it within the heat zone.

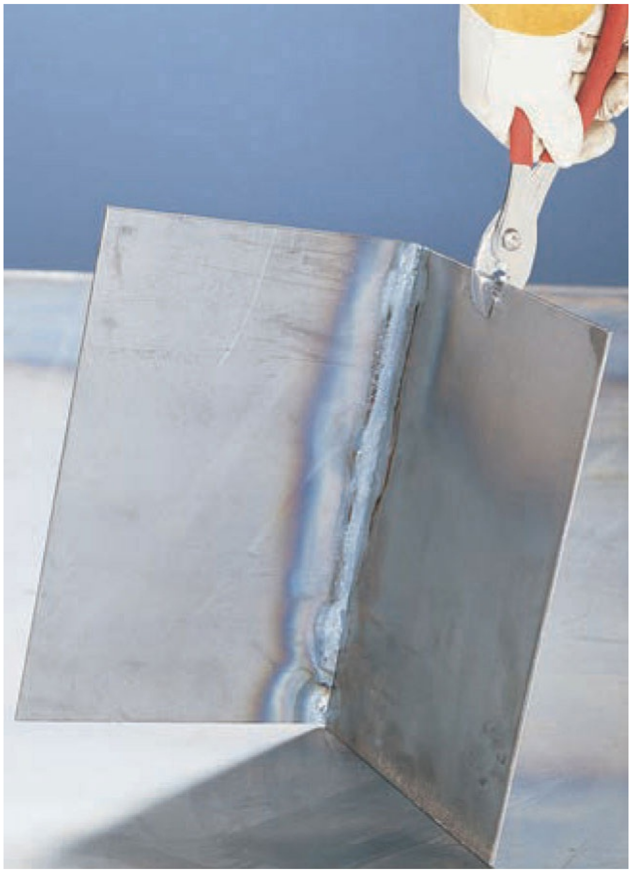

4 **Continue dipping, oscillating, and moving to the left. As you reach the end of the weld, the cumulative heat build-up may make it necessary to adjust to a shallower angle to deflect heat away from the puddle and prevent burn through. When finished, turn off the oxygen torch valve first, then the acetylene torch valve. The weld should penetrate to the back without burning through.**

Depressurizing an Oxyacetylene Outfit:

When you have finished welding, or if you are going to stop welding for more than 10 or 15 minutes, depressurize your setup. The hoses and regulators are designed to leak small amounts of fuel and oxygen if they are pressurized and not being used, so it is important for safety and economy to depressurize them. You will soon be so comfortable with the pressurizing and depressurizing steps that it won't seem like an inconvenience at all. It is important to do the fuel gas and oxygen in separate steps, to prevent having mixed, unburned oxyacetylene in the torch.

- Close the fuel gas cylinder valve.
- Open the fuel gas torch valve until both gauges read zero.
- Loosen (counterclockwise) the fuel gas regulator adjustment screw.

- Close the fuel gas torch valve.
- Close the oxygen cylinder valve.
- Open the oxygen torch valve until both gauges read zero.
- Loosen (counterclockwise) the oxygen regulator adjustment screw.
- Close the oxygen torch valve.

Oxyfuel Tips

GAS OPTIONS:

The most useful fuel gas for oxyfuel welding is acetylene. Its neutral flame generates the highest heat and heat concentration with the lowest chemical interactions with the molten metal. Other fuel gases, such as MPS or MAPP and propylene can be used for brazing and cutting applications. Natural gas and propane can be used for brazing and heating operations. Make sure that the equipment—hoses, regulators, torches, and tips—are designed for use with the gas you choose.

GAS	DENSITY (Air = 1)	TEMPERATURE OF NEUTRAL FLAME WITH OXYGEN
Acetylene	0.906	5589° F
Methylacetylene-propadiene (MPS)	1.48	5300° F
Natural gas	0.62	4600° F
Propane	1.52	4579° F
Propylene	1.48	5250° F

Maintenance. The primary maintenance for oxyacetylene welding equipment is cleaning the torch tip. A dirty tip will spark, pop, and often direct the flame sideways. Tip cleaners are inexpensive sets of cleaning rods used to clean the various sized tip orifices (see photo, page 27). The set usually comes with a small file to smooth the tip. To clean, simply insert the correctly sized tip cleaning rod into the orifice, and pull back and forth. Do not rotate the tip cleaner as this may enlarge the orifice. Also, do not use drill bits to clean tips. The sharp edges of the bit will cut grooves into the tip. Cleaning rods have smoothed edges and are sized especially for tips. Do not clean tips while the torch is lit.

FIRE BRICKS:

Fire bricks are specially made bricks that have little water content so they will not explode when heated to high temperatures. Fire bricks do not pull the heat from your weld or braze, and you cannot weld your material to them.

Brazing

Brazing is very similar to soldering since flux is applied to tightly fitted metal parts that are then heated to the point where filler material will melt and be drawn into the joint. Silver soldering and hard soldering are terms incorrectly used to refer to brazing. Brazing is different from soldering because it takes place at temperatures over 840° F and below the melting point of the base metals. The metals are not fused, but held together by the filler metal adhering to the base metals through capillary action.

PETERSON FLUXES
NO. 1 BLUE
Stay-Silv FLUX
SAFETY-SILV 56
WELCO
15FC
3/32
1 lb.
WELCO
15
3/32
1 LB

Brazing supplies include flux and silver solder (left). Braze welding requires flux coated rods, or separate flux and rods.

There are a number of industrial brazing processes, such as dip brazing, furnace brazing, and induction brazing. The home welder is likely only to do torch brazing. Torch brazing can be done with an oxyfuel torch using acetylene as the fuel gas, or any of the other fuel gases (see page 35). NOTE: Different fuel gases require different regulators, hoses, torches, and tips.

For brazing to work, the gap between the parts must be between 0.002 and 0.010 inches. If the gap is too tight, the flux and filler will not flow evenly through the joint. If the gap is too big, the strength of the joint is lessened. Gaps between parts can be measured with a feeler tool, available at automotive stores. Items to be brazed must be absolutely clean and free from rust, corrosion, grease, oil, and cleaning compound residues.

Brazing is often used commercially to join dissimilar metals such as tungsten carbide saw teeth to a steel saw blade. Another common use for brazing is in lugged bicycle frames. Because nearly every metal can be joined using brazing, it is highly suitable for art applications.

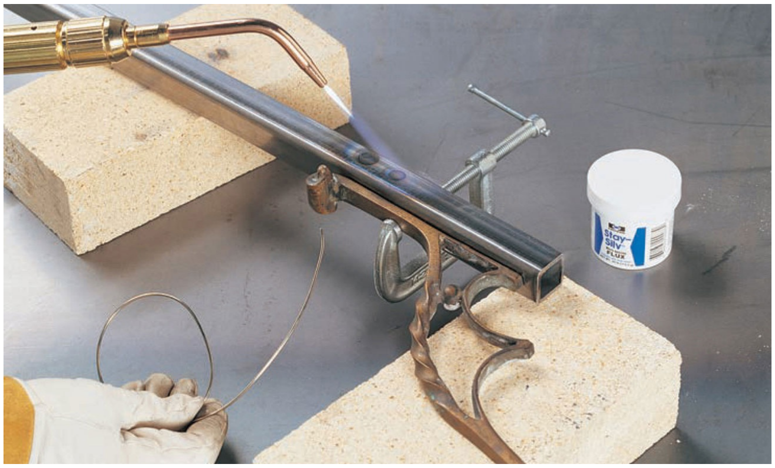

Thoroughly clean and flux both sides of the joint area. Using a small torch tip, heat the entire joint area until the flux turns clear and starts to run. Add enough filler metal to fill the joint. (Silver alloy is shown.) After the metal has cooled, the flux residue can be removed with hot water.

Braze Welding

Braze welding is similar to standard oxyacetylene gas welding except the parent or base metals are not melted, so there is no molten puddle. Instead of a steel alloy filler rod, a flux coated brass filler rod is used. Braze welding is often incorrectly referred to as brazing. Braze welding does not use capillary action to pull filler material into the joint—the filler metal is deposited as fillet or groove welds.

The brazing rod is melted by the heat of the metal and the flame, but it should not be held in the flame itself. The parts for braze welding should fit tightly, but the gap is not as important as with brazing.

Braze welding is used for joining dissimilar metals and for metals of different thicknesses. This technique is often used to repair cracked or broken cast iron.

Braze welding has less distortion than oxyacetylene welding because less heat is applied to the parts. A disadvantage is that it is not as strong as welding where the base metal is melted, but a well-made braze weld is still sufficient for most non-structural applications. Because the base metals do not need to be melted, braze welding can be done with any of the fuel gases listed on page 35.

If you are creating a piece that will be welded and braze welded, you must be careful to complete the non-brazed welds first. The heat involved with all other welding processes will boil off the brass alloy of a braze weld, ruining the weld and creating toxic fumes.

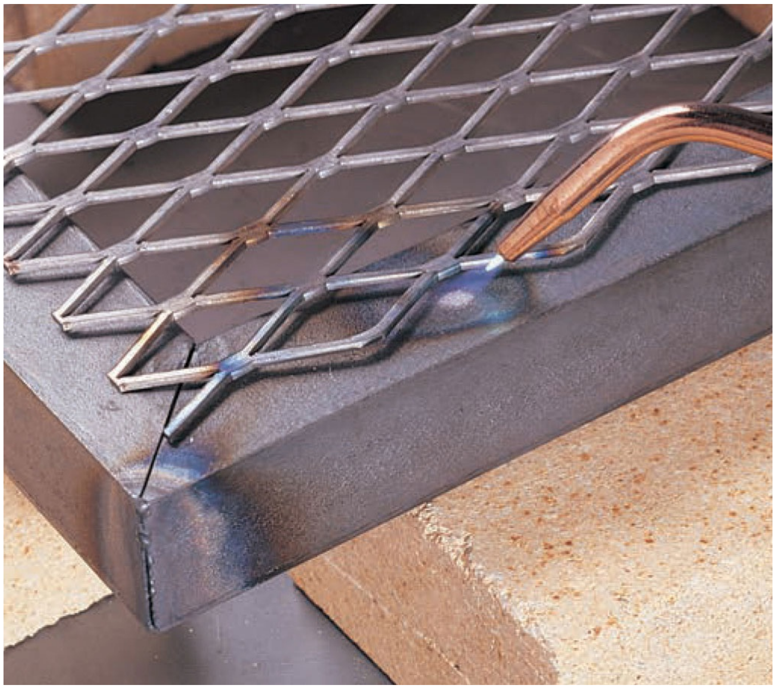

1 Braze welding is useful for joining thin metals, like this expanded metal, to thicker metals. Heat both parts, directing more heat toward the thicker part. It may take a long time for the thicker metal to heat. Using fire bricks will prevent a metal table top from absorbing any heat.

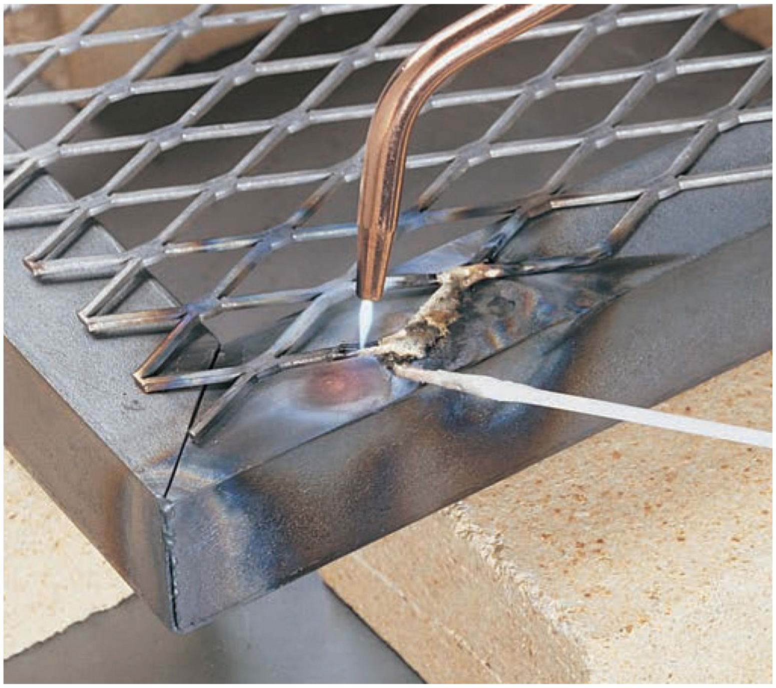

2 When both parts glow a dull cherry red, touch the flux coated rod to the joint. The flux and the filler metal will melt. If the metal is molten or the fluxed rod comes in contact with the flame, the flux will burn and the filler metal will boil. This results in a poor joint in addition to giving off toxic fumes.

ELECTRICITY FOR WELDING

Arc welding and cutting processes—shielded metal arc, gas metal arc, gas tungsten arc welding, and plasma cutting—all use electricity to generate the necessary heat. Understanding electricity is not necessary to use these processes, but knowing basic terms and concepts will help you understand why certain welding events occur.

Electric current is the flow of electrons through a conductor from a high concentration of electrons (negative charge) to a low concentration of electrons (positive charge). As electricity flows through a conductor, it generates heat based on how much resistance the conductor offers. An arc is simply a sustained electrical discharge across an air gap. Because air is highly resistant to electron flow, an enormous amount of heat is generated by the movement of electrons across this gap. The heat generated by an arc is in the area of 11,000° F, but about half of this heat is dissipated.

ELECTRICAL UNITS. The units used to refer to electricity are voltage, amperage, and wattage.

Voltage or volts (v) is the measure of electric potential or the electric pressure. The voltage controls the size of the air gap that the arc can cross. The higher the voltage, the larger the gap the arc can cross.

Amperage or amps (a) is the volume of electrons flowing through a conductor. The amperage controls the size of the arc.

Wattage or watts (w) is the measurement of the amount of electrical energy or power contained in the arc. Wattage affects the width and depth of a weld.

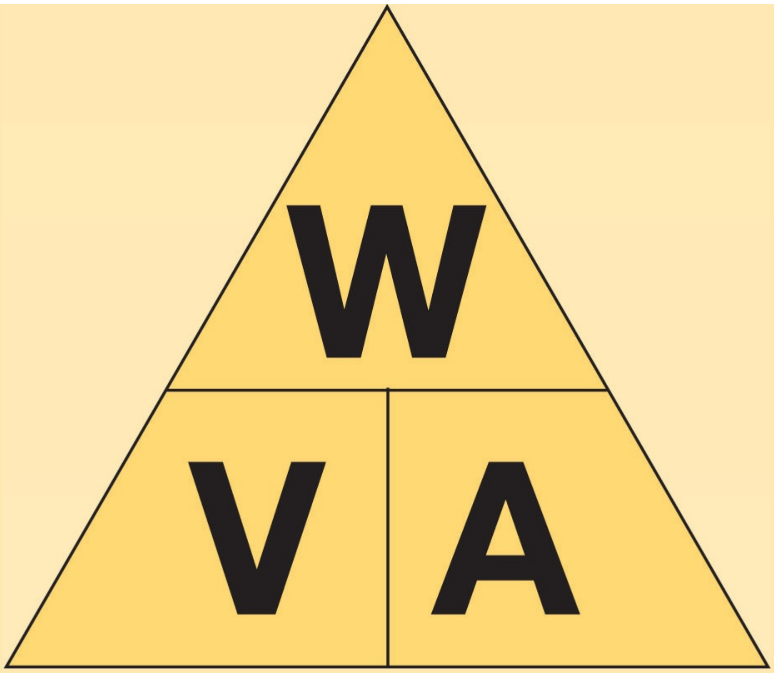

Volts, amps, and watts are related to each other in such a way that volts × amps = watts.

WELDING POWER. Power for welding may be supplied in one of the following ways:

Constant Voltage (CV). The arc voltage is maintained even when the current (amps) changes. This means if the arc length is changed, which changes the current flow, the voltage stays relatively the same.

Constant Current (CC). As the amperage fluctuates, the voltage increases or decreases to keep the total power (wattage) output the same.

OPEN CIRCUIT AND OPERATING VOLTAGE. The open circuit voltage is the voltage that exists at the electrode tip when the machine is on but no arc has been struck. The higher the open circuit voltage, the easier it is to strike an arc. Open circuit voltage is usually between 50 and 80 volts. The higher the voltage, the greater the risk of electrical shock.

The operating voltage is the voltage when the arc is struck and the circuit is completed. This is usually between 17 and 32 volts depending on the arc length, the type of electrode, the type of current, and the polarity.

POWER SOURCES FOR WELDING. There are three sources for producing the low-voltage high-amperage combination that arc welding needs.

A mechanical generator produces power through the use of a gasoline or diesel engine.

A step-down transformer takes available high voltage alternating current and changes it to a low-voltage high-amperage current.

An inverter uses solid state electronics to change the current without the weight of a transformer.

CURRENTS USED FOR WELDING. The type of current used and the polarity is important to welding.

Alternating Current (AC) is standard available household current. The electron flow changes direction two times per cycle. Welding with AC means that the electrode and the workpiece alternate between positive and negative, so the welding heat is distributed evenly to both. This balances penetration and build-up.

Direct Current (DC) electrons flow in one direction only. A rectifier is used to convert alternating current to direct current. With direct current, there are two polarity options:

> **Direct Current Electrode Negative (DCEN)—**formerly called straight polarity—means the electrode is negative, the workpiece is positive; electrons are flowing from the electrode to the workpiece and the heat is concentrated in the electrode.
>
> **Direct Current Electrode Positive (DCEP)—**formerly called reverse polarity—means the electrode is positive, the workpiece is negative; the electrons are flowing from the workpiece to the electrode. DCEP produces the best welding arc characteristics.

DUTY CYCLE. The duty cycle is the amount of continuous running time at a given power output for a welding machine, given as a percentage for a ten minute period. Because welding machines produce heat internally, they need a certain amount of cooling time. A 60% duty cycle means that a machine can run continously at that setting for six minutes, then it will need to cool for four minutes. Duty cycle is critical in construction and industrial applications, and it can be important for home shop use. A 10% duty cycle at the maximum setting means one minute of welding and nine minutes of cooling. This could be a problem for large projects on thick material. The higher the duty cycle, the more expensive the machine.

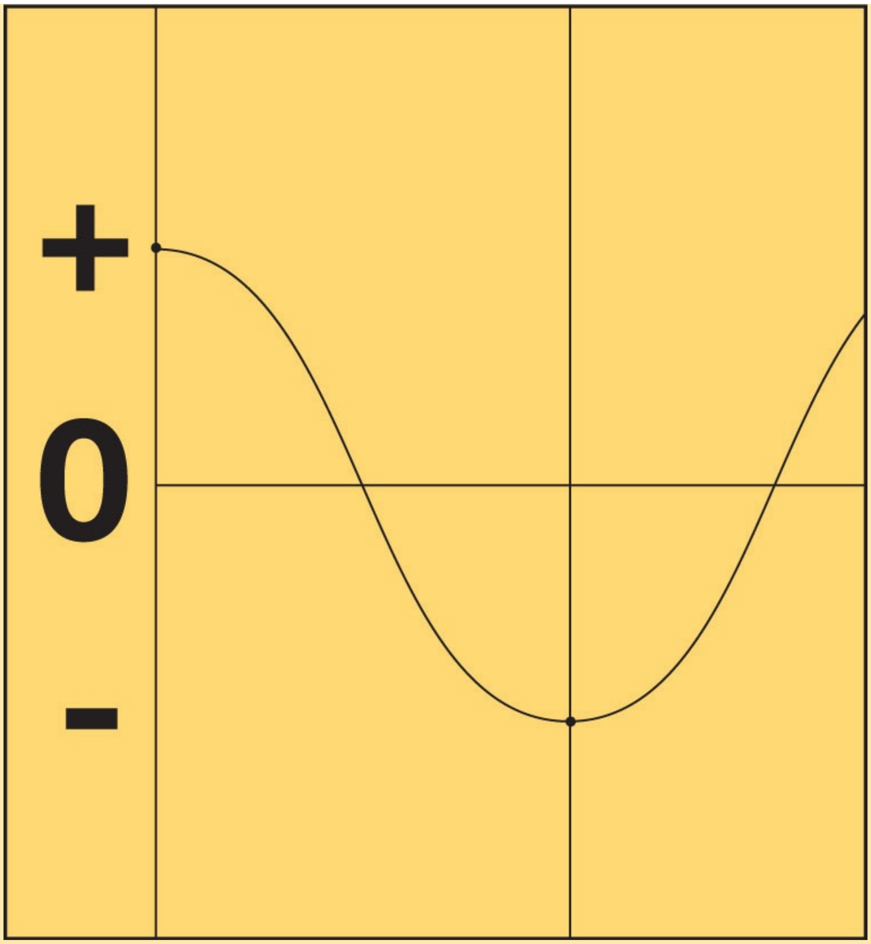

Alternating current alternates between positive and negative polarities, passing through a zero, or no current point, between the two.

NOTE:

Work leads and work clamps are sometimes referred to as ground leads and ground clamps. This is not technically correct, as the electricity is not being grounded. Instead, the work lead and clamp are allowing for completion of the circuit back to the machine.

SHIELDED METAL ARC WELDING

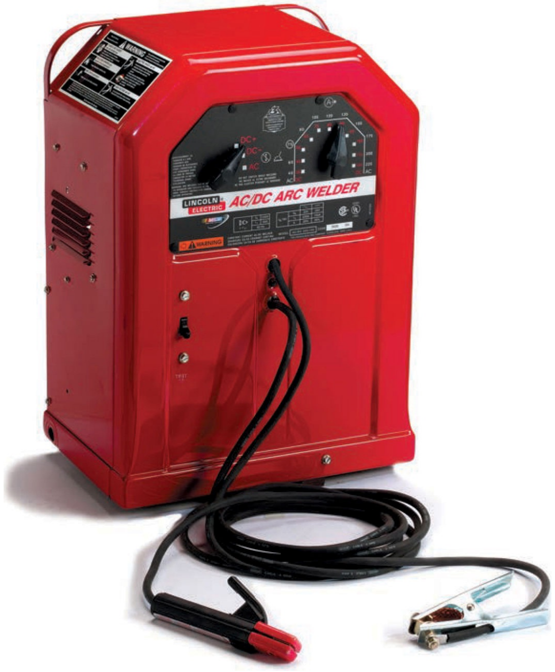
WARNING
DC+
DC-
AC
LINCOLN ELECTRIC
AC/DC ARC WELDER
WARNING

A shielded metal arc welder consists of a power source, electrode holder, and work clamp.

Shielded metal arc welding (SMAW) is also referred to as arc or stick welding. The process involves the heating of the base metal to fusion or welding temperature by an electric arc that is created between a covered metal electrode and the base metal. The coating or covering on the electrode provides both flux and shielding gas for the weld. The electrodes come in 9" to 15" straight lengths in a range of wire thicknesses from 1/1" to 3/8", hence the name “stick.” SMAW is used extensively for fabrication, construction, and repair work because the machinery is inexpensive and fairly simple, and the electrodes are inexpensive. There are drawbacks to SMAW: It doesn’t work well on thin materials (less than 1/8" is difficult), electrodes need to be changed frequently as they are used up, and the protective slag coating must be chipped off each weld.

Safety. Shielded metal arc welding uses electricity, so there is always the possibility of receiving an electric shock or being electrocuted. When an electrode (stick) is placed into the electrode holder, it is “live.” If the electrode touches anything that the work clamp is in contact with, the circuit will be completed and an arc will be struck. To prevent this from happening, always remove the electrode from the holder when you are not actively welding. Do not use your bare hands to insert or remove electrodes—always touch electrodes with dry gloves. Remember that water and electricity never mix well, so do not weld while standing on a wet or damp floor or ground, and do not weld outdoors in the rain. The electrode will be hot after use, so take care where you dispose of electrode stubs. A metal bucket is a good addition to your SMAW workshop.

Safety

- Welding helmet with #10 to #14 filter
- Leather, wool, or cotton long pants; leather jacket
- Heavy-duty welding gloves
- Safety glasses
- Hat
- Leather boots or shoes
- Ventilation

SMAW welding produces ultraviolet and infrared rays, harmful fumes, and hot spatter. Protect your eyes with a #10 to #14 filter in a full-face welding helmet or hood. Heavy-duty leather welding gloves and a welding jacket with leather sleeves are necessary to protect you from the molten spatter. Proper ventilation from an exhaust hood or fan is important since many of the chemicals used in the electrode coatings are hazardous if inhaled. Also, it is a good idea to screen off your welding area so others are protected from the intense light of the arc.

The Equipment. SMAW machines are available as either alternating current (AC), direct current (DC), or with the capability of switching between the two. The machine itself is simple in that it merely converts high-voltage low-amperage line current into low-voltage high-amperage welding current. Output is controlled with one knob. You often will hear SMAW machines referred to as “buzz boxes.”

AC welding machines meet most home and small shop needs, are inexpensive, and are readily available. Because the alternating current cycles through a zero current between the positive and negative polarities, it can be difficult to strike and maintain an arc. DC machines are easier to use and have many home and hobby applications, but they are more expensive. Because DC current can have its flow reversed (see page 39), a DC machine has more versatility in terms of the electrodes that can be used. This allows for a wider range of welding positions, metal thicknesses, and metal types that can be welded. This versatility makes a DC shielded metal arc welder well worth the extra expense.

The equipment itself consists of the welding machine, which usually has one adjustment knob, an electrode lead with electrode holder (sometimes called a stinger), and a work lead with work clamp. You will often see work leads and work clamps referred to as ground leads and ground clamps. The work lead and clamp are not grounding the electricity, they are completing the circuit back to the machine.

MILWAUKEE
BRUSH
20503

Useful tools for shielded metal arc welding are a pliers, chipping hammer, and wire brush.

SMAW Electrodes. SMAW electrodes are solid, round, metal wires coated with flux and other components. In addition to producing shielding gas and flux, the covering may also contain additional metals for filler or alloying elements for the weld.

The American Welding Society (AWS) publishes standards for the electrodes. Electrodes come in diameters ranging from 1/16" to 3/8" in increments of 1/32". The electrode diameter measures the wire itself, not the diameter of the wire with the covering. The electrode designation is inked onto the covering near the bare end of the electrode. The number classification for SMAW electrodes begins with the letter E because they are electrodes. The first two or three numbers on the left denote the tensile strength of the properly completed weld in thousands of pounds per square inch. The second number from the right indicates the welding position appropriate for that electrode: 1 = all, 2 = flat grooves and flat or horizontal fillet weld, 3 = flat only, 4 = all. The last two numbers designate the current and polarity uses and other special notes. There also may be electrode suffixes that denote alloys that have been added to the electrode.

HOBART
STICK ELECTRODES
HOBART
STICK ELECTRODES

SMAW electrodes are sold in 5- or 10-pound boxes. Specialty electrodes may be bought a pound at a time. It is important to keep electrodes dry, either in the original box or a storage container.

Each electrode manufacturer may have a number of electrodes of a specific AWS designation that are slightly different and have been tailored to specific uses. The diameter of the electrode will determine which amperage to use. Use the manufacturer’s guidelines to determine amperage. If not available, start with 90 to 120 amps minimum range for a 1/8" electrode and add 40 amps for each 1/1" increase in diameter. This will give you a rough minimum starting amperage.

6013
7018
6011

Electrodes for SMAW are stamped with a numeric code.

The most commonly used electrodes are 6011, 6013, 7014, and 7018. The 6011 and 6013 electrodes will work with AC power, the 7000 electrodes will not. For a beginner, 6013 with DC power is generally the easiest electrode to use in terms of striking an arc and maintaining a consistent arc. It is always a good idea to talk with your welding supplies dealer about what type of welder you are using, the type of welds you are making, and the materials you are working on to get the best electrode for your purposes.

Slag. Shielded metal arc welding produces a weld that is covered with a coating of ceramic-like slag. The flux and other components in the covering clean the material to be welded and also float out impurities in the weld. These impurities and flux solidify on top of the weld which protects the cooling weld from the effects of oxygen in the atmosphere. The slag must be scraped or chipped away before the weld bead is covered with another weld layer or before the weld is painted or finished. Safety glasses should be worn during this procedure.

Pre-welding Checklist:

- Electrode and work leads tightly attached to machine terminals.
- Cable insulation free of cuts, damage, and wear.
- Electrode lead tightly connected to electrode holder.
- Work lead tightly connected to clamp.
- Work clamp firmly clamped to worktable or workpiece.
- Floor is dry.
- Ventilation is adequate to draw fumes away from welder’s face.

Shielded Metal Arc Welding

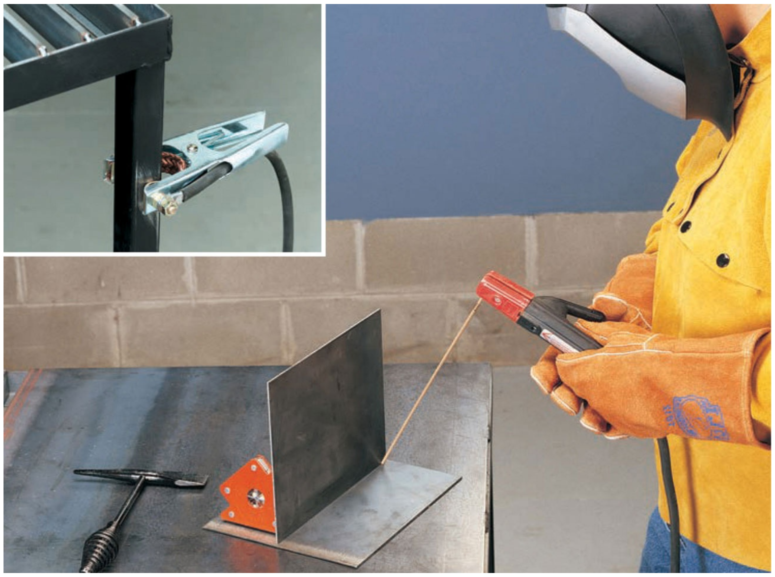

1 Set up your material to be welded. Make sure the electrode holder is not touching the workpiece or worktable. Attach the work cable clamp to the table (inset) or workpiece. Turn on machine. Adjust the range switch for the desired amperage. Wearing leather gloves, place an electrode in the electrode holder. Position the electrode over the area to be tacked, flip down your helmet, and tap or scratch the electrode on the area to be tacked to strike an arc. After making your tack welds, remove the electrode from the electrode holder. Check to see that the tacked pieces are aligned properly. If not, use a hammer to move them into alignment or break the tack welds and retack. Chip slag from tack welds so it does not contaminate the final weld.

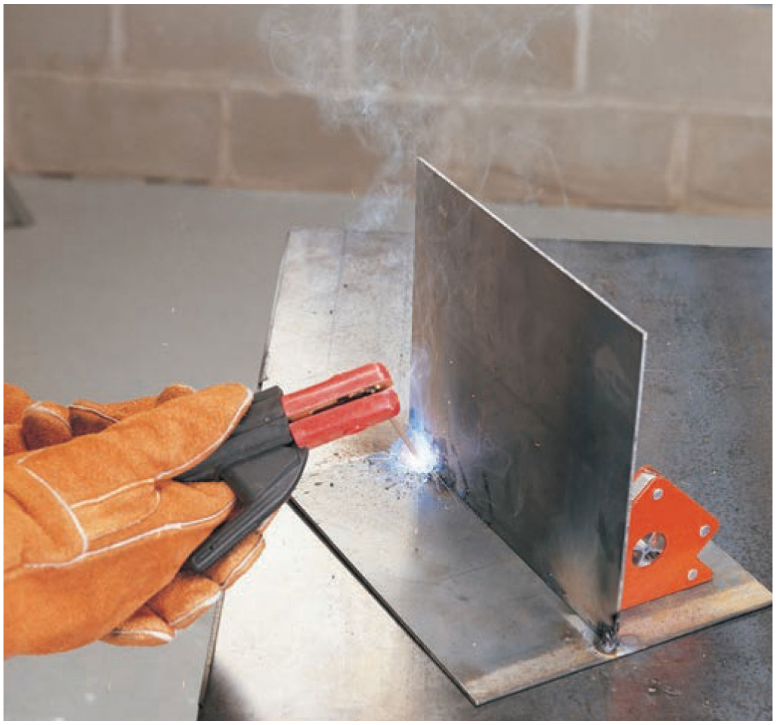

2 Replace the electrode in the electrode holder, and position the electrode over the left side of the area to be welded. Hold the electrode at a 10° to 20° angle to the right. Flip down your hood and scratch or tap to strike an arc. The distance between the metal and electrode should not exceed the thickness of the electrode bare wire diameter. Move slowly to the right until the weld is completed.

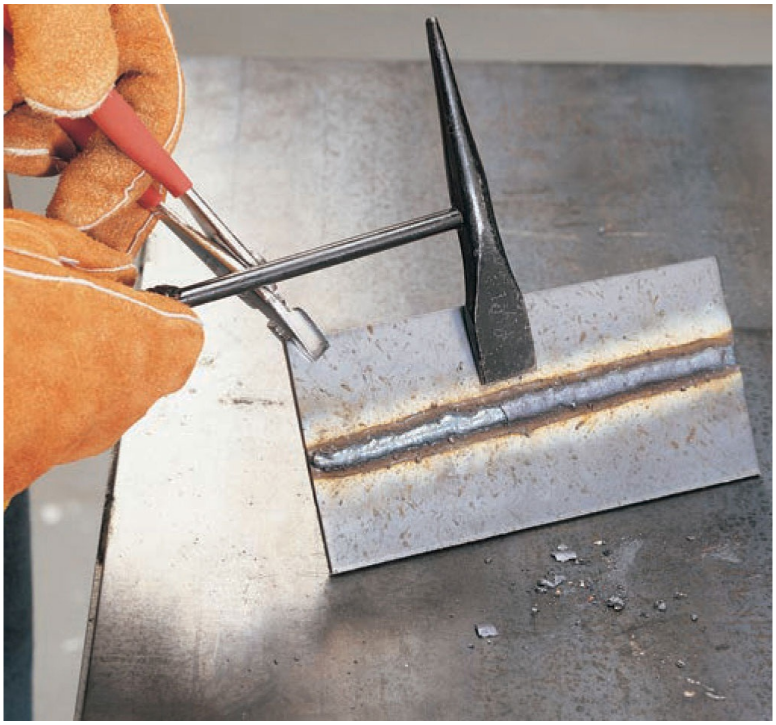

3 To remove slag, hold the workpiece with pliers at an angle and scrape or knock the slag with the flat

blade of the chipping hammer. Wear safety glasses when chipping slag.

GAS METAL ARC WELDING

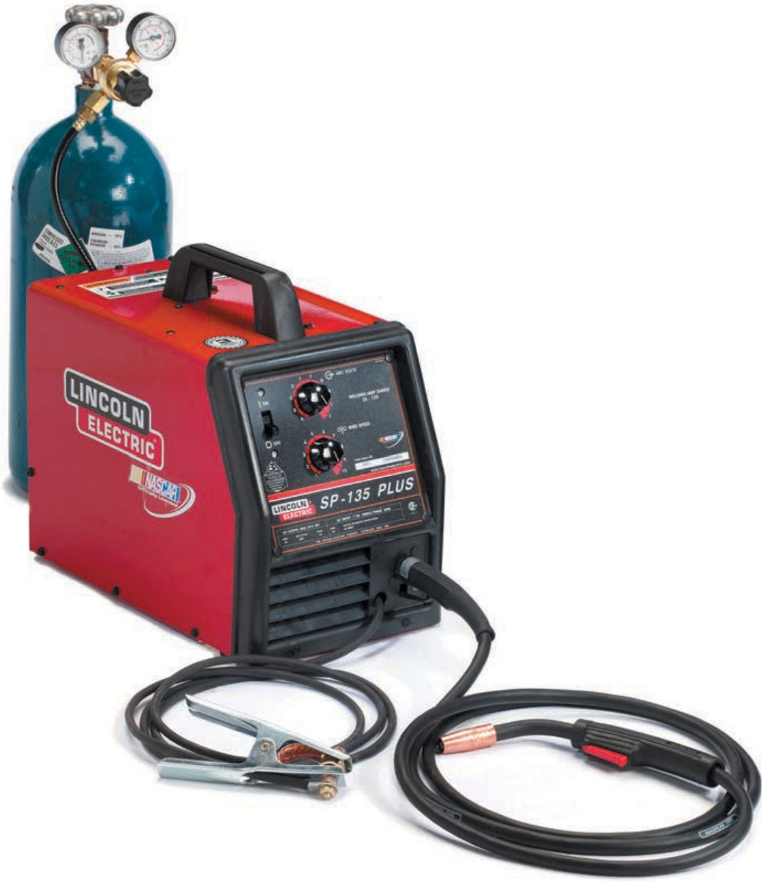
LINCOLN ELECTRIC
NASCAR
SP-135 PLUS

A gas metal arc welding setup consists of a power source, wire feed, work cable with clamp, supply cable with gun, and gas cylinder with regulator.

Gas metal arc welding (GMAW), also referred to as MIG (metal inert gas) or wire feed, is a process where a consumable electrode (wire) is automatically fed through a welding gun along with a continuous flow of a shielding gas. The actual application of filler metal is achieved with most small machines through the short circuit transfer method. Other transfer methods used by more powerful machines are short circuit globular and spray pulse.

The GMAW process has a number of advantages. The gun or torch can be held at a uniform distance from the weld, unlike shielded metal arc welding where the distance from the electrode holder to the weld becomes shorter as the electrode is consumed. The trigger power control on the welding gun also makes it possible to be completely positioned and ready to weld without accidentally striking an arc. When ready to weld, the welder can flip down the helmet and pull the trigger without getting off target. Because the electrode is the filler, the welder does not need to coordinate a filler rod in one hand and the torch in the other. The GMAW gun can be held steady with both hands to create a uniform bead. Because of the shielding gas, GMAW welds are smooth and clean, with no slag to chip or grind. The GMAW process can operate at very low amperages and is a relatively cool welding process, so 22 and 24 gauge sheet metals can be welded without distortion. The process also requires narrower beveling for thicker plate welds, so less time is spent on grinding.

Safety

- Welding helmet with #11 to #12 filter
- Leather, wool, or cotton long pants, long sleeve shirt, and hat
- Leather gloves with gauntlet
- Leather boots or shoes
- Ventilation

Some disadvantages of GMAW are that the shielding gas nozzle does not allow for welding in tight spaces without some modifications. Because the shielding gas can be disrupted or blown away by drafts easily, the process is not suited for working outdoors.

THE EQUIPMENT

Power supply. In the small GMAW units, the power supply and wire feed are integrated into the same cabinet. In larger, multipurpose machines, the wire feed is a separate unit. The power supply converts the standard alternating current (AC) into direct current (DC). Alternating current is not appropriate for gas metal arc welding because its constantly switching flow does not yield a steady, consistent arc. On the machine, you will be able to adjust the voltage, polarity, and wire feed rate; the machine adjusts the amperage. When using the power source to weld with a shielding gas, the machine should generally be set up for electrode positive, that is, the electrode is positive and the workpiece is negative. If used without shielding gas and with flux cored wire, the polarity usually needs to be switched to electrode negative. This is achieved by switching the contact wires within the machine. Larger, multipurpose machines may have a switch to change the polarity. Always check manufacturer's recommendations for polarity settings.

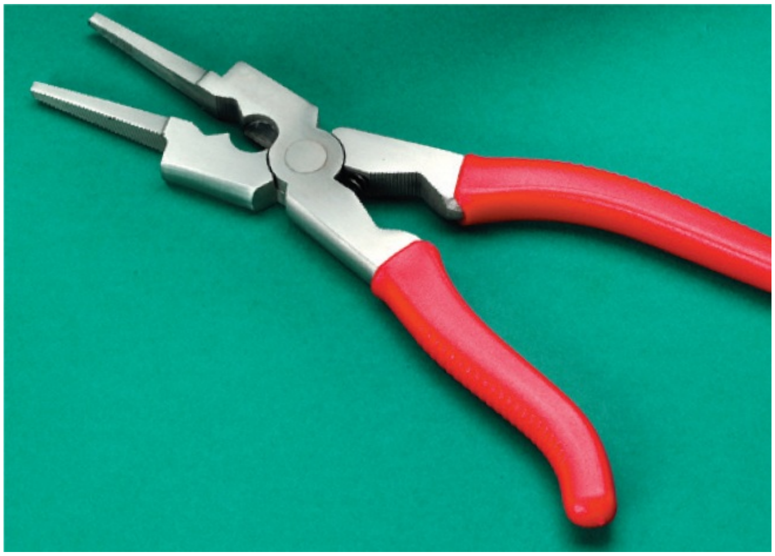

A MIG pliers is a handy tool for gas metal arc welding.

Wire feed. The wire feed consists of a spool of wire, a tension controller, and rollers and a roller motor. The wire speed is set by the speed adjustment knob, depending on the thickness of the material being welded. The rollers have V or U grooves sized to match particular wires. Most machines come with two-sided rollers that can be switched over to handle two different size wires.

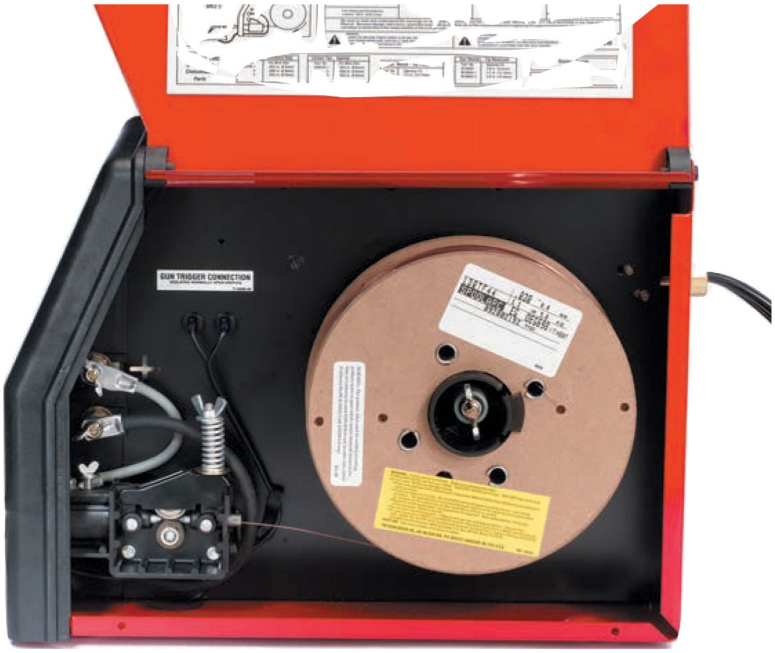

The wire feeding mechanism consists of a spindle to hold the wire spool, drive wheels, rollers, and a tension adjustment. Most welders now have a reference chart with recommended settings.

Gun. The GMAW gun, sometimes referred to as a torch, is attached to the welding machine via the supply cable. The cable carries the power, the power control, the wire in a special liner, and the shielding gas. The gun has a trigger that turns the power on and off, which also starts the wire feed and the shielding gas flow. The gun generally has a goose neck shape, although straight neck and flexible neck guns are available. A contact tip is screwed or cam locked into the gun. This tip has an orifice the same size as the wire being used. It is important that this be sized correctly—too large and good electrical contact between the source and the electrode will not be made, too small and the wire will not fit. The shielding gas is directed toward the work area with a nozzle. On most home versions, the relationship between the contact tip and the gas nozzle stickout is not adjustable; larger welders have a wider variety of nozzle sizes and shapes. It may be worthwhile to purchase an aftermarket adjustable system. When using the GMAW process, it is important that the supply cable be kept as straight as possible to prevent kinking the wire or impeding the gas flow.

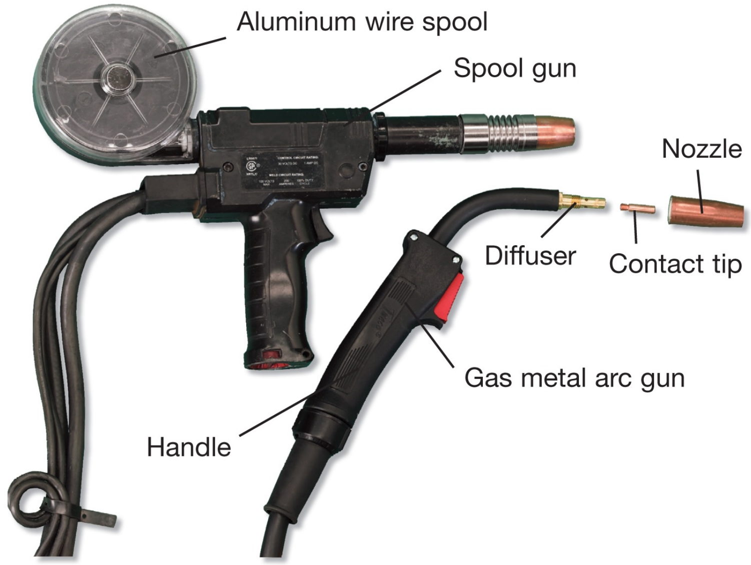

A spool gun is used for feeding aluminum wire because it often breaks or misfeeds through a standard cable. A standard gas metal arc gun consists of a handle, gas diffuser, contact tip, and nozzle.

The final piece of equipment is the work cable with clamp. This is clamped to the workpiece or to a metal work surface that the piece sits on to complete the circuit.

Electrodes. When choosing an electrode (wire), consider the composition properties, cleanliness of the base metal, and the shielding gas. If you will be welding out of position, this will also be a factor. There are dozens of wire choices for GMAW.

Most home GMAW welders use 1-pound or 10-pound wire spools. Larger spools are generally less expensive.

Electrodes are labeled with alpha-numeric codes that describe their type, tensile strength of the weld, whether it is solid or tubular (tubular is flux cored), and chemical composition. For mild steel, ER70S-3 is a good general purpose wire. ER70S-4 and ER70S-6 are good wires for dirty or rusty metals. Wire also is available for welding aluminum and stainless steel. Available wire sizes are 0.024, 0.030, 0.035, and 0.045 inches.

Using a wire feed machine without shielding gas requires the use of flux cored wire (often called innershield), and the process is then called flux cored arc welding (FCAW). Because the wire makes the electrical contact as it travels through the contact tip, the flux must be inside the wire, which makes flux cored wire more expensive than regular GMAW wire. Commercial welders may use a “dual shield” flux cored wire that does use a shielding gas. Welds made with this process are as clean as regular GMAW welds.

Flux cored arc welding can be done outdoors because there is no shielding gas to be disturbed by wind and weather. Unfortunately, using flux cored wire results in a less attractive weld due to the presence of slag and more spatter. The flux cored wire usually has greater penetration and is often used to stretch the capabilities of a small welding machine. Because 0.035" is the smallest diameter for flux cored wire, it is not appropriate for welding thin sheet metal. The lowest setting for this size wire is still too hot for this application.

Shielding Gas. The shielding gas for GMAW is supplied by a cylinder with a flow regulator and connecting hoses. The gas line is connected to the welding machine, which directs it through the cable.

Carbon dioxide is suitable for general GMAW welding. It gives good penetration and is inexpensive compared to other gases. Argon and carbon dioxide in a 75% argon and 25% carbon dioxide mix is the standard GMAW shielding gas. It is slightly more expensive than pure carbon dioxide, but it yields welds with less spatter. Pure argon is used for shielding aluminum welds. If you choose to use pure carbon dioxide, select an electrode that performs well with this shielding gas.

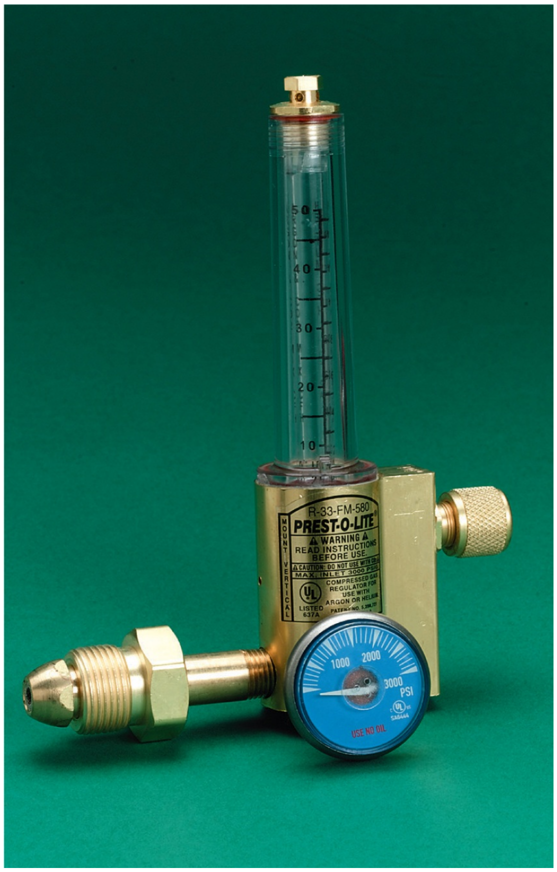
50
40
30
20
10
R-33-FM-580
PREST-O-LITE
WARNING
READ INSTRUCTIONS BEFORE USE
MOUNT VERTICAL
1000
2000
3000 PSI
USE NO OIL

A flow meter registers gas flow with a small floating ball.

Gas Cylinder Safety. Gas cylinders store compressed gases at pressures up to and greater than 2000 psi. This is sufficient pressure to send a tank through a concrete wall if the cylinder valve should be damaged. It is important to be extremely careful with gas cylinders even when they hold the non-flammable gases used in GMAW welding. Gas cylinders should always be upright and chained or strapped to a wall, post, or cart. Never drag a cylinder; never lay one down or roll it flat. To move a cylinder, chain it to a handcart first. If you must move a cylinder without a handcart, tip it slightly and roll it. Never weld anything to the cylinder or safety cap—and be careful to never accidentally strike an arc against a cylinder. If you do, you will have to purchase that cylinder.

Setting Up the Wire Feed

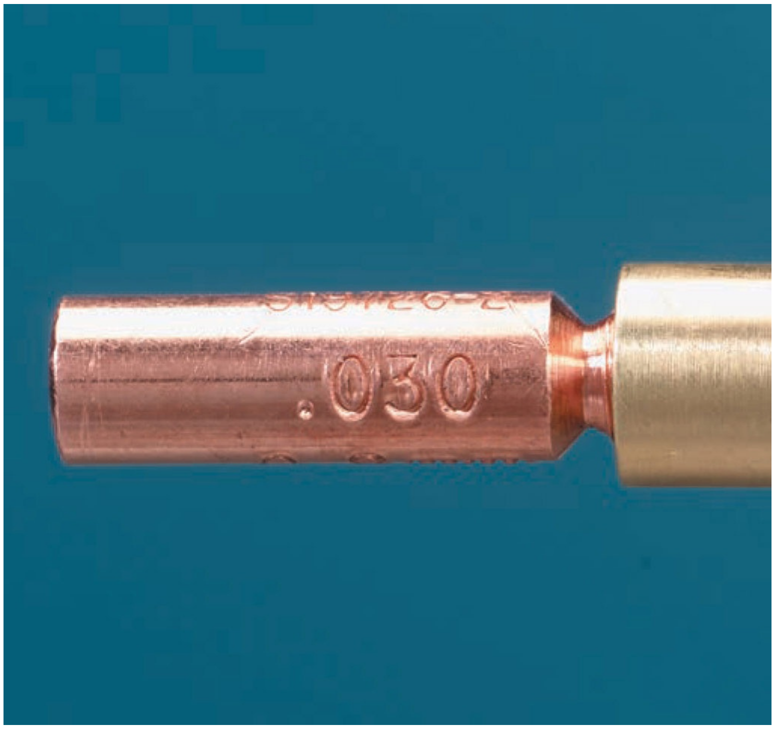

1 Contact tubes are stamped with their size. Make sure the contact tube and the drive roll grooves are the correct size for the electrode you are using. Place the wire spool on the spindle, and secure it with the pin lock, lock ring, or wing nut. Make sure it is feeding in the proper direction.

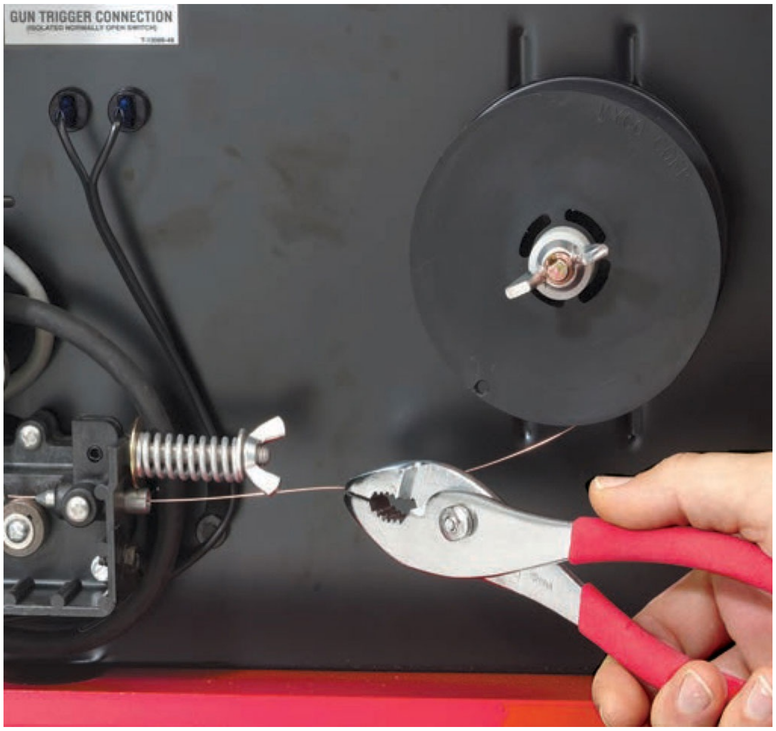

2 After releasing and cutting off the crimped wire end, hold the wire firmly with a pliers. The wire is tensioned and the entire spool will rapidly unroll if you do not hold it firmly. Swing the tension arm or pressure roll out of the way. Push the wire through the inlet wire guide, through the groove on the drive roll, and out through the outlet wire guide.

NOTE:

Unless your machine has an inch switch for advancing wire or a purge switch for activating gas flow, you will use the gun trigger to accomplish both tasks. Using the gun trigger means the wire is “live” and will arc to anything the work clamp is touching.

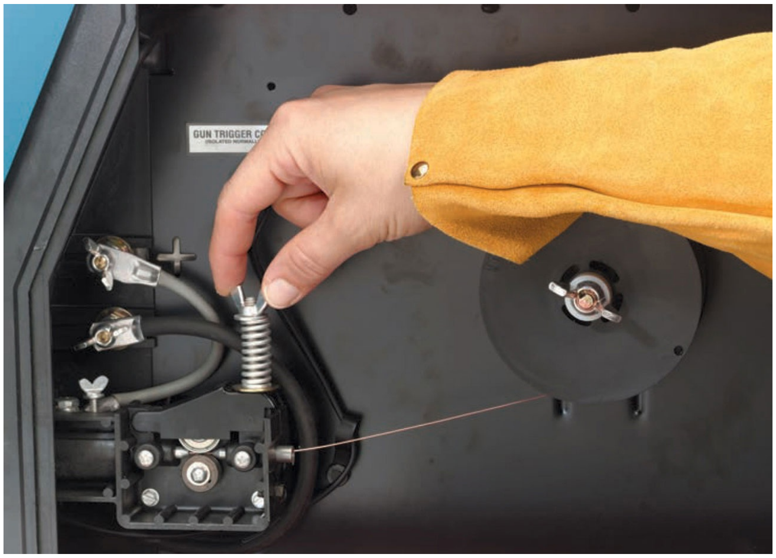

3 Swing the tension arm or pressure roll back into position. Adjust tension according to manufacturer's directions. On a new machine, the drive roll pressure will already be set. Make sure the wire is aligned perfectly straight—not up and down or side to side—on the drive roll. Turn the machine on, turn the wire speed to its highest setting, pull the cable straight, and depress the trigger. Some machines have an inch button that feeds the wire without supplying power to the contact tip or wasting shielding gas while loading wire.

Setting Up the Shielding Gas Flow Meter

1 **Open the cylinder valve briefly to clear out any dirt. Wipe the cylinder threads with a clean, dry cloth.**

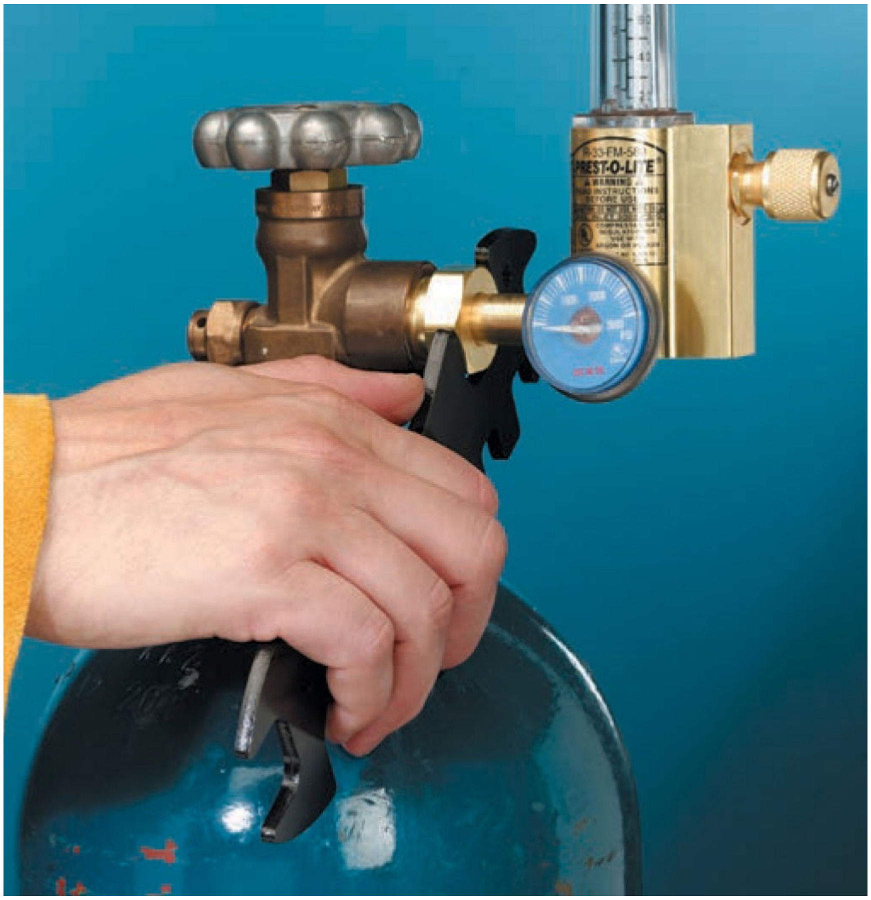

2 **Attach the flow meter to the cylinder. Use a fixed wrench to prevent damaging the brass fittings.**

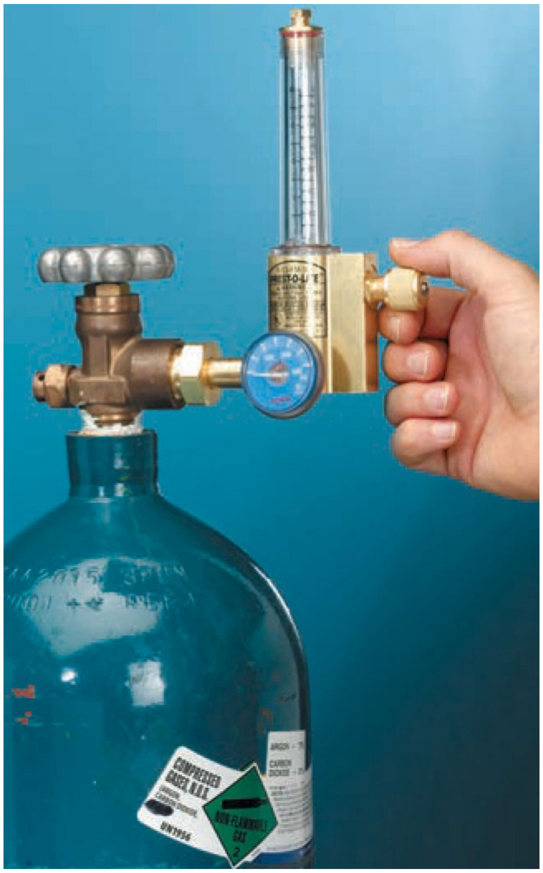

3 **Turn the knob on the flow meter clockwise to tighten it so no gas can flow through it. If the flow meter is open when you open the cylinder valve, the high pressure can damage the flow meter.**

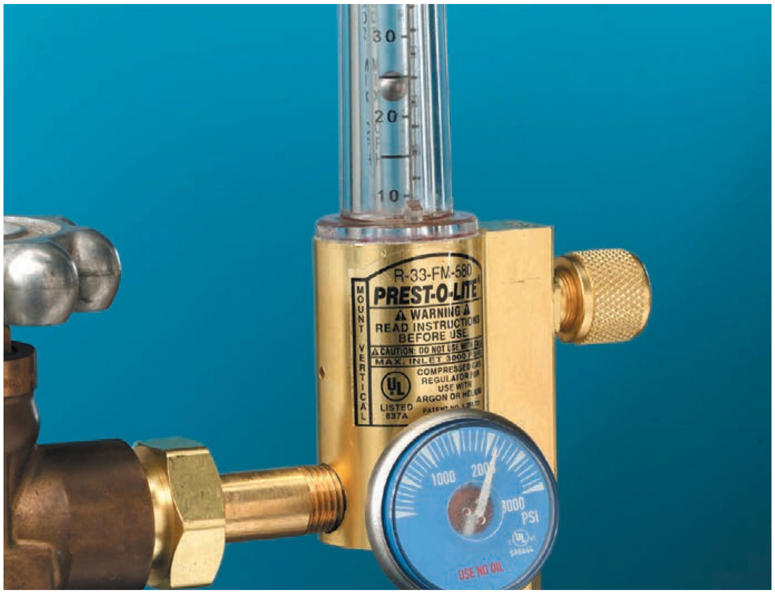
30
20
10
R-33-FM-580
PREST-O-LITE
MOUNT VERTICAL
WARNING
READ INSTRUCTIONS BEFORE USE
1000
2000
3000
PSI
USE NO OIL

4 **Attach the hose to the flow meter and to the welding machine. Slowly open the cylinder valve, then, once it is open, open the cylinder valve completely. To set the flow rate, turn on the machine. Depress the**

trigger to activate the gas flow, and turn the flow meter knob counterclockwise. While gas is flowing, continue turning the knob until the meter registers the proper flow rate. Make sure that the flow meter or flow gauge is rated for use with the shielding gas you are using. For example, an argon/carbon dioxide flow meter cannot handle pure carbon dioxide.

Pre-welding Checklist:

- Gun, shielding gas, and machine cables, hoses, and connectors are tight.
- Polarity is set correctly for type of wire being used.
- Cables are in good condition.
- Supply cable is straight and unkinked.
- Gas regulator and flow meter are correctly installed.
- Contact tube is in good condition.
- Spatter has been cleaned from nozzle and contact tube (see page 40).

Gas Metal Arc Welding

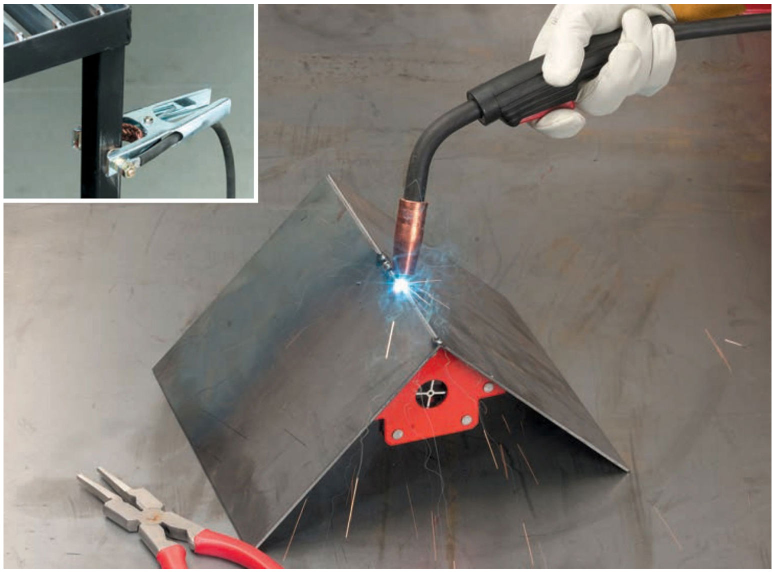

1 Prepare and fit up the material to be welded. Attach the work cable clamp to your welding table (inset) or the workpiece. Turn on the machine and shielding gas. Adjust the wire speed and voltage according to the manufacturer's recommendations. Cut the electrode to 3/3" stickout. Put the tip of the electrode at the point of the first tack weld, flip down your helmet, and pull the trigger. Place tack welds evenly around the weld area. Trim the electrode for proper stickout at the end of each use. Check to see that the workpiece is still aligned properly. If not, adjust with a hammer or break the tack welds and retack.

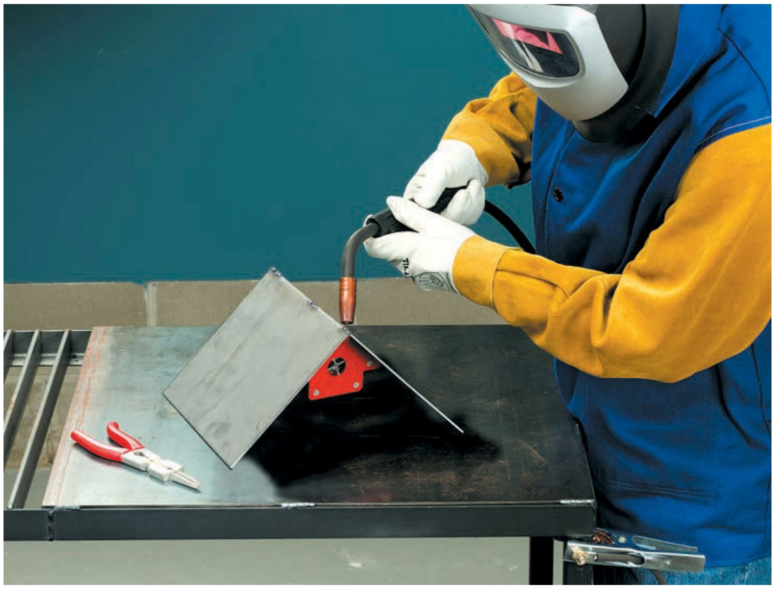

2 Start at the left end of the weld. Hold the gun with both hands and position it at a 20° angle with the tip pointing to the left. Put the tip of the electrode at the beginning of the weld, flip down your helmet, and pull the trigger.

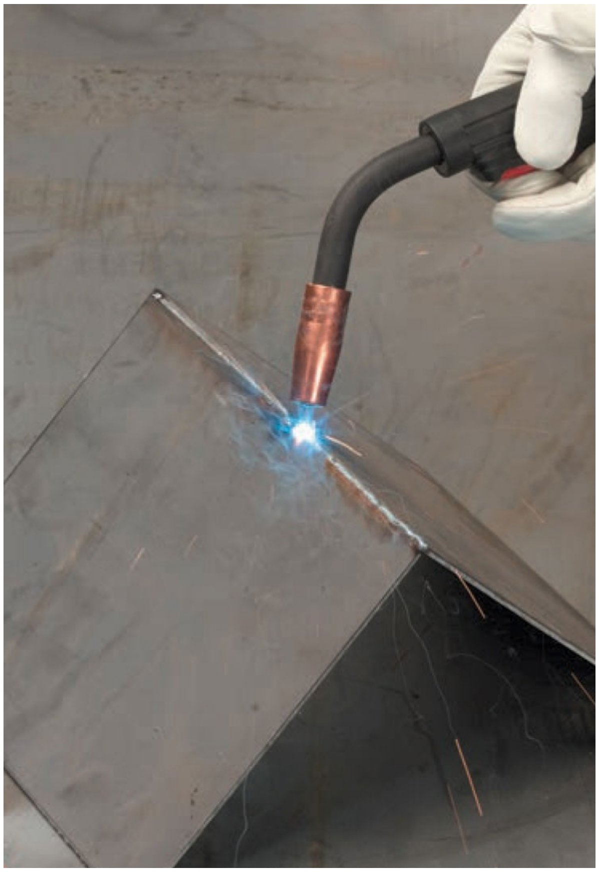

3 **Begin welding, moving steadily to the right. GMAW has a distinctive sizzling sound. Popping and snapping indicate dirty material or an improper voltage or wire speed adjustment.**

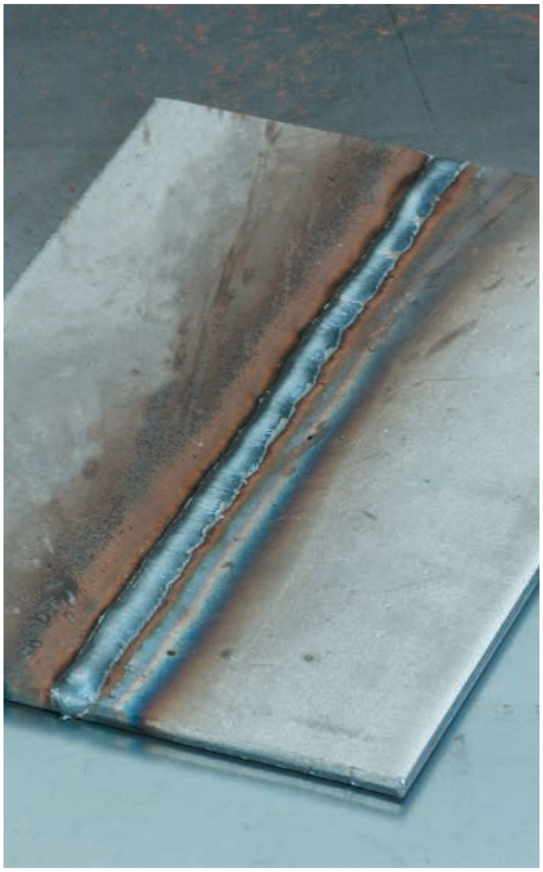

4 **A finished GMAW is smooth with even ripples or weave pattern, no slag, and little spatter.**

FLUX CORED ARC WELDING:

Flux cored arc welding uses the same wire feed and power supply as gas metal arc welding but usually without the gas shielding. This is convenient for welding outdoors in windy conditions, but the weld is not as clean. To use flux cored wires, most welding machines need to be converted. Usually this involves changing the output polarity, and installing a gasless nozzle. It also might require installing proper drive rolls and a different cable liner. Otherwise, follow the same steps to weld with flux cored wire as with solid wire.

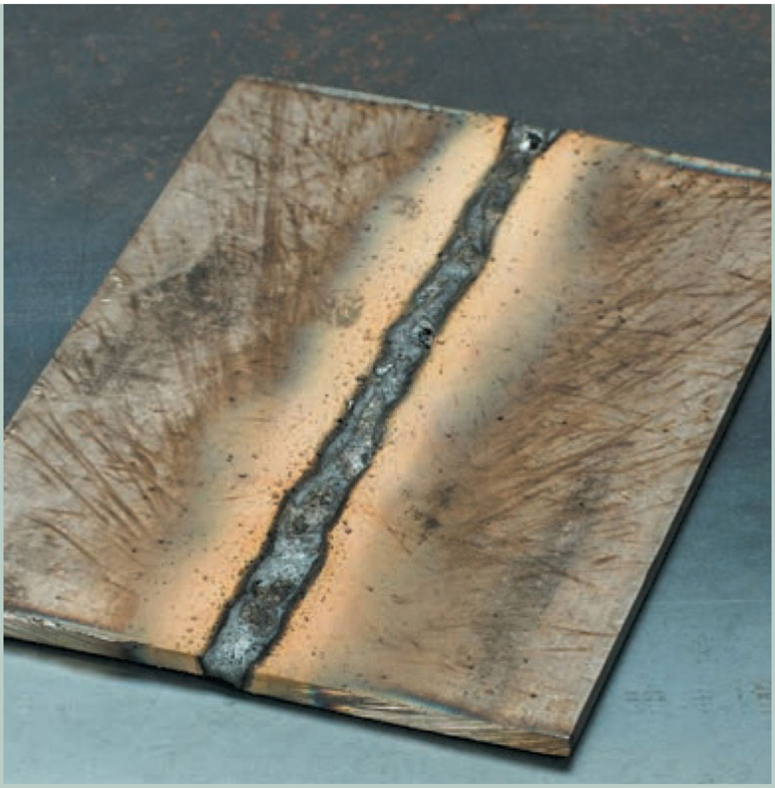

Flux cored wire welds are covered with a layer of slag.

Post-welding Sequence:

When you have finished welding with a GMAW rig, it is important to shut it down properly so you are set for your next welding session. If you do not turn off your shielding gas, you will lose much of it to bleed off even if you weld the next day.

- Close the gas cylinder valve.
- Press the gun trigger or purge switch until the flow meter or flow gauge reads zero.
- Turn off the flow meter.
- Turn off the machine.
- Coil the supply and work cables and store them off the floor.

Maintenance & Troubleshooting

A small amount of maintenance on your GMAW equipment will result in better, more consistent welds and longer lasting equipment. Spatter—tiny pieces of the electrode or base metal that have sizzled off—builds up on the nozzle and the contact tip. Remove this spatter frequently as you weld so it does not interfere with gas flow or electrical conductivity. Turn off the machine, remove the nozzle, and use the closed point of the MIG pliers (page 45) to ream out the nozzle and file the end of the contact tip. An anti-spatter gel is available in which to dip the hot nozzle and contact tip. Anti-spatter spray can be used to coat the contact tip and the nozzle. This spray is also useful for coating your welding table to prevent spatter from sticking to your work surface. Eventually you will need to replace the nozzle, but regular maintenance and proper welding techniques will extend the life of a nozzle considerably.

Contact tubes become worn because they are a soft copper alloy and the electrode is steel. The electrode will wear through the contact tip, possibly making the electrical flow irregular. Contact tips should be visually inspected for wear on a regular basis. The orifice on a worn contact tip will appear oval instead of round. Tips generally are good for about eight hours of continuous welding use. NOTE: Always turn off the welder before changing the contact tip.

It is important to keep the welder fan motor and rectifier free of dirt and dust to prevent it from overheating. If your welder is stored in a dusty shop, or you use it infrequently, keep it covered. Use low pressure air to blow dust and dirt from these assemblies. Use a vacuum to remove dirt from the wire feed mechanism.

The supply cable needs to be cleaned occasionally as well. Check the manufacturer’s recommendations for frequency, but they generally recommend cleaning after you’ve used 50 pounds of flux cored wire or 300 pounds of solid wire. With the power off, remove the cable from the machine. Remove the gas nozzle and contact tip from the gun, lay out the cable straight, and use low pressure air to blow into the gun end. (Using high pressure air may create a dirt plug that will clog the cable permanently.)

Always store the supply and work cables and the gas hose off the floor and away from chemicals, hot sparks, and sunlight.

ANCHOR BRAND
NOZZLE-DIP
16 FLUID OZ.
.473 LITERS
FOR MIG GUN NOZZLES
AND CONTACT TIPS

Use dippable anti-spatter gel to prevent spatter build up on the nozzle and contact tip. Simply dip the hot nozzle into the gel.

REMOVING A BIRD'S NEST:

A "bird's nest" is a tangle of wire in the wire drive mechanism. This happens when the wire drive rolls continue feeding wire but the cable is blocked, the outfeed tube is misaligned, the wire is stubbing out on the base metal, or the wire has been welded to the contact tip. The drive rolls are set to slip in these cases, but if the tension has been adjusted too tightly, they may continue to push wire.

To fix a bird's nest, turn off the machine and cut the wire as it comes off the spool. Remember the wire is under tension, so you must hold the spool end, then tie it off through one of the spool holes. The tangled wire also may be storing tension, so it may fly out of the drive roll area. Release the tension arm or roller arm. Pull the wire out of the drive mechanism and the supply cable. Wear safety glasses when removing a bird's nest.

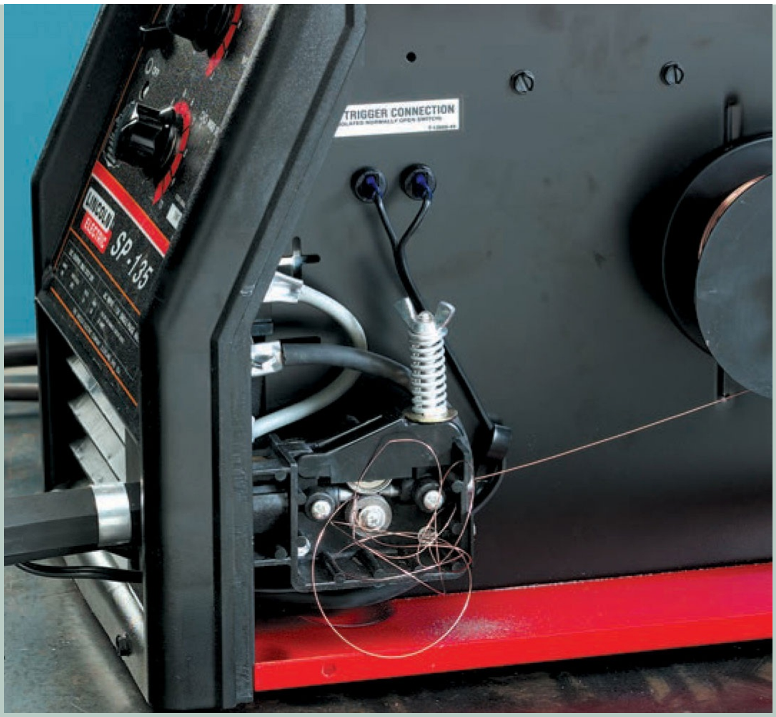
TRIGGER CONNECTION
SP-135

A "bird's nest" is a tangle of wire in the drive roll mechanism.

Stub Out or Stubbing:

“Stub out” is when the electrode welds to the base metal without melting or breaking off. This is caused by the voltage being too low, the wire feed being set too fast, or holding the gun too close to the work when starting. Correct the settings, grind off the stub outs, and restart the weld.

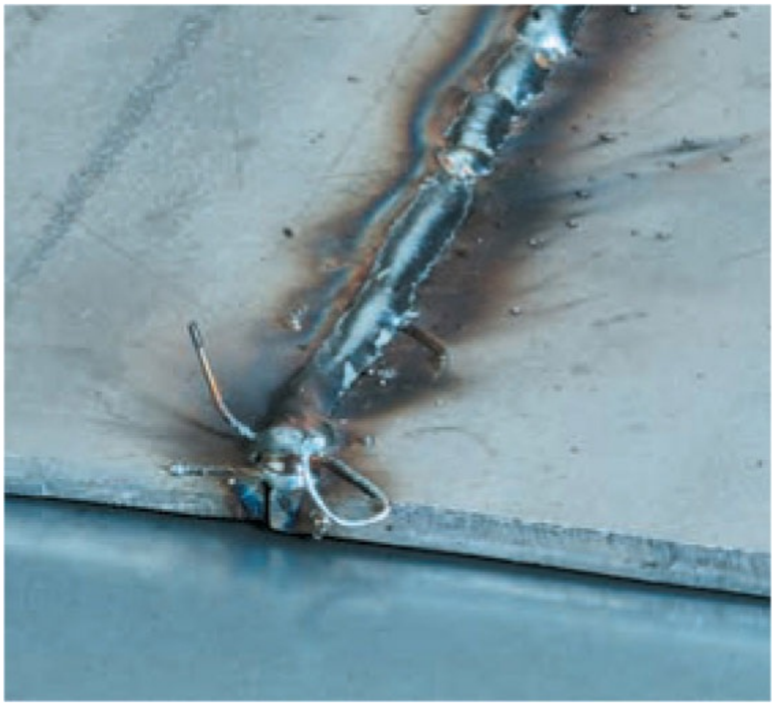

Stub out results in short sections of wire poking out of the weld.

GMAW Troubleshooting:

PROBLEM	SOLUTION
Wire feeds, but no arc can be struck	• Check to see that work clamp is connected
Weld bead is full of holes like a sponge (porosity) rate	• Make sure shielding gas is on • Adjust shielding gas to a higher flow • Eliminate drafts • Verify correct polarity
	• Make sure voltage setting is

Weld bead is tall and looks like a rope	appropriate for the material’s thickness • Reduce your travel speed
Excessive spatter or dirty welds	• Reduce the gun angle • Hold the nozzle closer to the work • Decrease voltage
Arc starts and stops	• Check the wire feed for steady feeding • Check cable connections • Clean the weld materials

GAS TUNGSTEN ARC WELDING

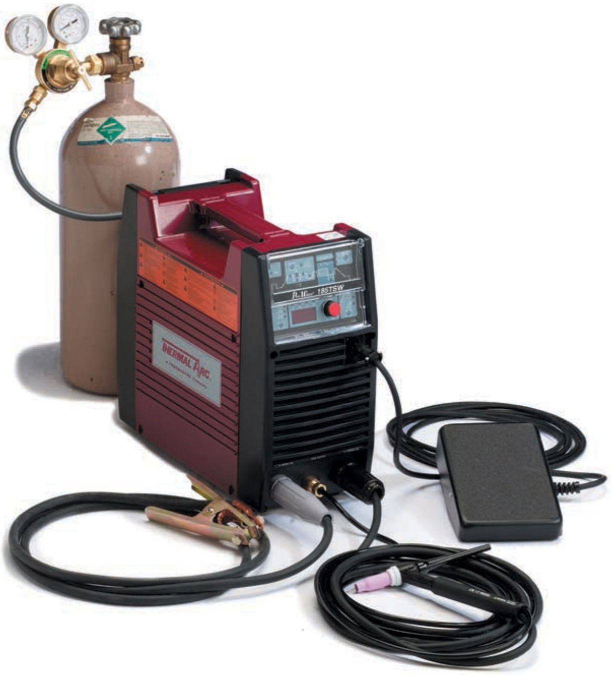

A gas tungsten arc welding setup consists of a power source, work cable and clamp, supply cable and torch, foot control, and gas cylinder with regulator.

Gas tungsten arc welding (GTAW), commonly called TIG (tungsten inert gas), and sometimes referred to as Heliarc (the L-TECH trade name), is a process that generates an arc between a non-consumable tungsten electrode and the workpiece. The electrode and the weld are protected by a shielding gas, and filler metal may or may not be used.

Gas tungsten arc welding differs from the other arc welding processes because the electrode is non-consumable and not used as filler material. GTAW is more like oxyfuel welding in terms of the skills needed to manipulate the torch with one hand and the filler rod with the other. GTAW requires another layer of coordination because most machines also use a foot activated amperage control.

Like gas metal arc welding, gas tungsten arc welding is a clean process because the shielding gas eliminates the need for flux and the resultant slag.

Safety. Because gas tungsten arc welding is such a clean process, welders often are tempted to weld without gloves or in short sleeve shirts. This is not recommended. The arc most likely produces more ultraviolet rays than other processes, and because there are no fumes or smoke, those rays are entirely unfiltered. It is important to cover all exposed skin to prevent UV burns. Filter requirements are a minimum #10 shade, and if you have an auto-darkening hood, be sure that it is rated for the GTAW process. Some entry level auto-darkening hoods are not. The arc may produce ozone, which is hazardous to breathe over long periods. Make sure your gloves, clothes, and work area are dry. Always turn off the power when changing electrodes.

Safety

- Welding helmet with #10 to #14 filter
- Long pants, long sleeved shirt, and hat
- Leather gloves
- Leather boots or shoes
- Ventilation

The Equipment. The basic equipment needed for GTAW is a constant current welding machine, cable with torch, work cable and clamp, electrode, and inert gas cylinder with regulator and flow meter. Optional equipment includes a remote amperage control and a water cooled torch with water cooler and hoses.

Although a mid-range shielded metal arc machine can be used to deliver the current for gas tungsten arc welding, a dedicated, good quality GTAW machine delivers the current as AC or DC, provides an optional high frequency output for no touch arc starting, has a remote control option for foot pedal control, and has a solenoid for shielding gas control. The combination of AC and high frequency makes it possible to weld aluminum with good results. The newest electronic GTAW machines have advanced current control capabilities and are becoming more and more affordable.

Torch & Cables. The GTAW torch holds the electrode and delivers the shielding gas. It can be air or water cooled. The torch parts are the cup or nozzle, collet body, collet, end cap, and torch body. The collet and collet body hold the electrode firmly and establish electrical contact with the electrode. The cup directs the shielding gas. A torch may be air cooled for use below 200 amps, but over 200 amps the torch usually is water cooled. The supply cable supplies electricity, shielding gas, and coolant to the torch. This may be an integrated cable or individual cables. A work cable with clamp is also necessary.

Some GTAW torches have flexible necks, which allows them to be bent into the most convenient angle for supporting the torch. Unfortunately, the non-flexible torches will bend if forced, and then they will be ruined. Some torches also have a gas on/ off valve. This is necessary for machines that are simply power sources with no internal gas flow controls.

The collet body, collet, end cap, and cups all come in a variety of sizes. The collet body and collet are sized to match electrode diameters. Each part is stamped with its size. The end caps come in a variety of lengths—some accommodate an entire 7" electrode, others are shorter for more clearance while welding.

MILWAUKEE
BRUSH
20534
Stainless
STANLEY
84-101

Pliers and a stainless steel wire brush are handy for gas tungsten arc welding.

Cups come in a variety of sizes denoted by 1/16" increments. A #4 nozzle is 4/16", a #8 is 8/16". Cups should ideally be 4 to 6 times the size of the electrode. A cup that is too large will not direct the gas in a proper stream, a cup that is too small may not provide adequate steady flow. A gas lens or screen in the torch balances the gas flow. Good shielding designs allow you to use a lower gas flow. Higher flow creates turbulence and jetting, which disturbs the shielding—more does not equal better when it comes to shielding gas flow. Cups also are available in different lengths. A longer nozzle increases the electrode extension from the collet, which may increase the electrode temperature. Use either lower amperage or larger diameter electrodes if you are using a longer nozzle. Shorter nozzles may be necessary for tight working conditions. Nozzles are also available in clear (fused quartz) types, which are more expensive but give greater visibility, especially in cramped quarters.

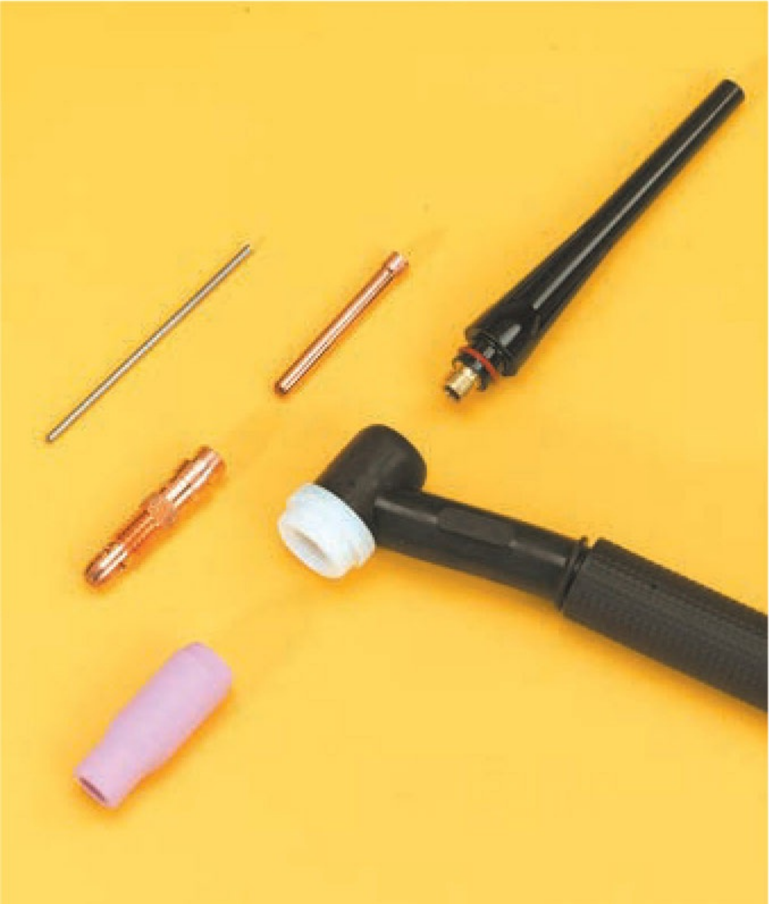

A gas tungsten arc welding torch includes the electrode, collet body, collet, cup, backcap, and torch body with heat shield.

Shielding Gases. Two inert shielding gases are used for gas tungsten arc welding: argon and helium. They may be used alone or in combination with one another. Argon yields better starts and better arc maintenance. Helium gives a hotter arc and allows faster travel speeds. Helium, however, is very expensive. Argon and helium are both contained in standard cylinders, and a pressure regulator flow meter combination will need to be attached. Because of the density difference between the two, either a separate flow meter needs to be used for each, or use a flow meter that has calibrations for each type of gas. Argon flow rates for up to 1/8" steel is 8-10 cubic feet per hour (CFH), for stainless steel 11-13 CFH, and 12-20 CFH for aluminum. Drafty locations may require higher flow rates. Check the manufacturer’s recommendations.

Electrodes. The electrode for gas tungsten arc welding is always a tungsten or tungsten alloy rod with either a pointed or rounded end. The electrodes come in a variety of diameters.

STANLEY
84-101

Tungsten is very brittle, so electrodes can be scored with a file, then snapped over a sharp table edge.

Preparing the Electrode. Prior to use, the cut end of the electrode must be sharpened to a point or melted to a ball. The tip may be ground to a point or chemically sharpened. The electrodes come in 7" lengths. To increase the number of points available, score the electrode with a file or cut-off wheel, and snap it in half. Tungsten is very hard but brittle, so it is easy to grasp each end of the electrode with pliers and snap it in half over a sharp table edge.

ELECTRODE CHARACTERISTICS:

ALLOY	AC/DC	TRAITS & USES	COLOR
Pure tungsten	AC	• Welding of aluminum • Forms ball at tip	Green
1% Thorium oxide (thoriated)	Primarily DC, but also AC	• Easier arc starts • Carries more current • Low level radioactive	Yellow
2% Thorium oxide	DC	• Long life • Easier arc starts	Red
Zirconium oxide	AC	• Similar to pure tungsten but carries more current	Brown
2% Cerium oxide (ceriated)	DC	• Not as good as thoriated but not radioactive	Orange
Lanthanum oxide	DC	• Similar to ceriated	Black

Because all tungsten electrodes look and feel the same regardless of their composition, it is important to keep them clearly separated by type. The color codes will wear off, or, if you point each end of your electrode, be ground off. It is helpful to have clearly labeled containers for each type of electrode.

Two critical factors in grinding the electrodes are the grinding wheel and the grinding direction. You must use a hard, fine grinding wheel dedicated exclusively to tungsten. Metal particles left on the wheel from grinding aluminum or steel would contaminate the tungsten, which causes erratic arc behavior and poor weld quality. An extremely hard material, tungsten will become hot as it is ground. Sharpen the electrode tip so that grinding marks run lengthwise down the tip, not in a circular or crosswise pattern. Lengthwise grinding focuses the electron flow toward the tip; circular grinding causes the arc to be unfocused and possibly jump sidewise from the electrode rather than off the tip point. Chemical means also can be used to sharpen tungsten by dipping a hot tungsten rod into a chemical agent. The length of the taper on the tungsten tip should be two to three times the diameter of the tungsten.

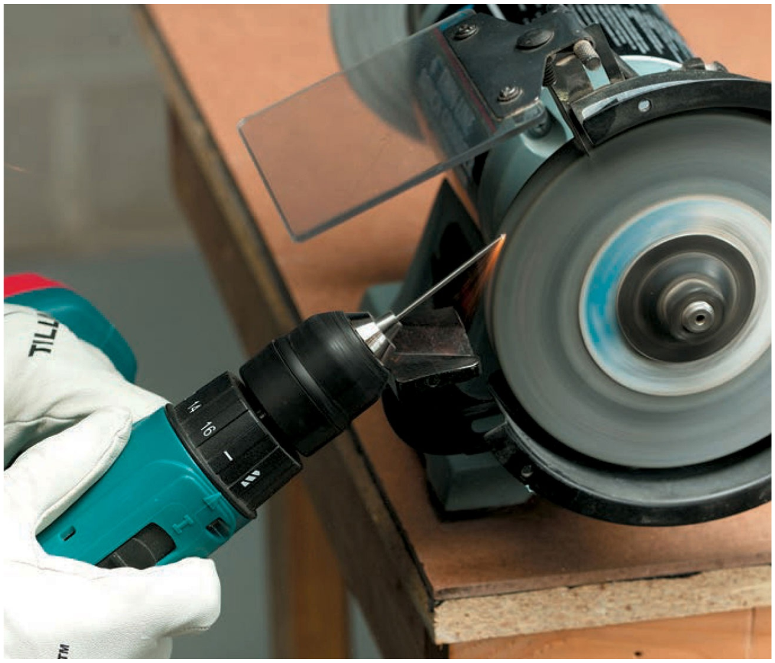

To safely hold tungstens for sharpening with a grinder, insert them into the chuck of a drill. Running the drill while sharpening the tungsten ensures a uniform point. Move the tungsten across the wheel to prevent grooves in the wheel face.

5 oz.
Anchor Brand
CHEMICAL SHARPENER

Chemical sharpeners are available for sharpening tungsten.

Welding with AC current requires that the tip of the electrode be balled, which is accomplished by running DC electrode positive or AC current and striking an arc on copper or brass. The ball should not extend beyond the diameter of the electrode, although many sources recommend a ball one and one half times the diameter of the tungsten.

If you touch the weld puddle or filler rod with the electrode during welding, you must change to a fresh electrode. Any time the electrode tip becomes pitted or blackened, is no longer pointed, or the ball becomes too large, you must change to a fresh electrode. This will happen often, so it is best to have a number of electrodes sharpened and ready to go. If you forget to turn on the shielding gas or the flow is interrupted, the electrode will give off a thick yellow tungsten oxide smoke. This smoke is hazardous, so resume welding only after the smoke has cleared. You will need to change electrodes if this happens.

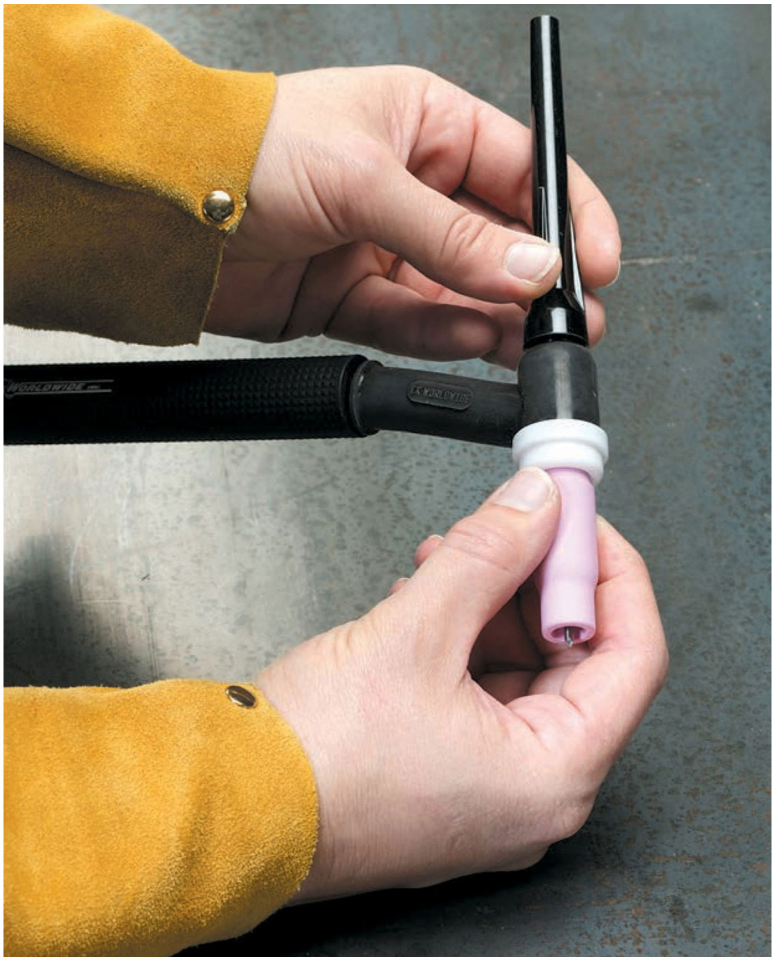

Tightening the back cap locks the tungsten electrode into place. The tip of the electrode should extend beyond the end of the cup by a distance of three times the electrode's diameter.

Anatomy of the GTAW Machine

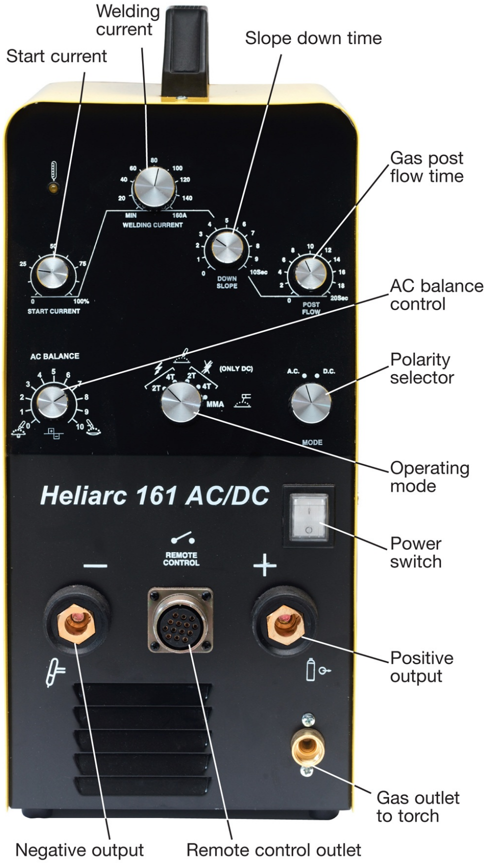
Welding current
Start current
Slope down time
Gas post flow time
AC balance control
Polarity selector
Operating mode
Heliarc 161 AC/DC
Power switch
Positive output
Gas outlet to torch
Negative output
Remote control outlet

Setting Up the GTAW Machine:

Dedicated GTAW machines, like this one and the one pictured on page 54, have many advanced features. The basic setup, however, is the same.

The positive and negative output receptacles are for the torch and work clamp connections.

The gas outlet connects to the torch gas hose.

The remote control socket is for the foot or finger control. Using a remote control allows the welder to control the amount of current while welding.

The operating mode selects high frequency or scratch starts.

Polarity selects either AC or DC.

AC balance control allows for adjusting AC power to be more positive or more negative to create either more cleaning action or deeper penetration. Advanced machines can adjust the AC power to be as much as 90% positive or negative, rather than the 50% for standard AC power.

Start current controls the current for arc starting.

Welding current sets the range for welding. The foot or finger control then operates within this range.

Slope down time gradually reduces power to the arc with-out extinguishing the arc. This allows for the weld crater to fill before the arc is extinguished.

Gas post flow controls how long the shielding gas will flow after the arc is extinguished.

Pre-welding Checklist:

- If using a water cooled torch, check for leaking.
- Check all cables for wear and damage.
- Clean and fit up parts to be welded.

Welding with GTAW

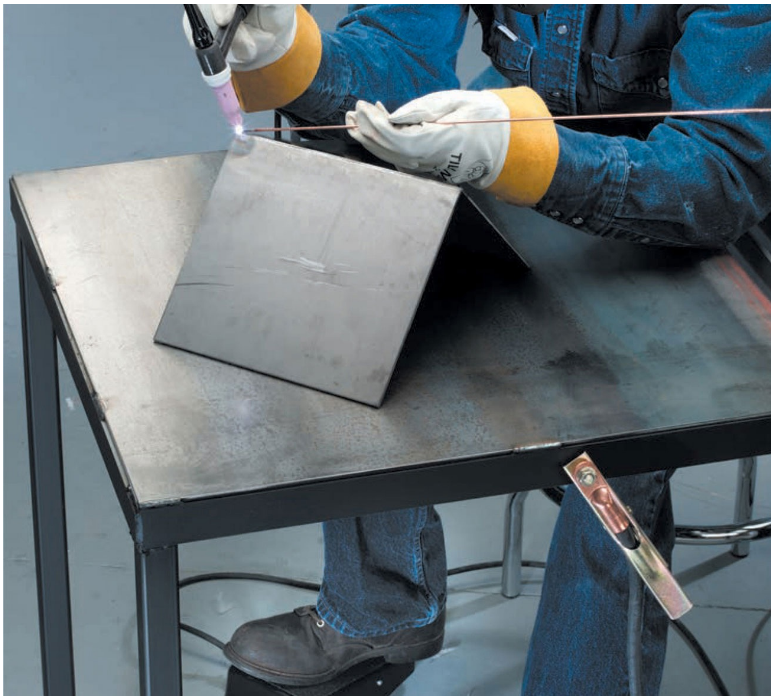

1 Set the controls based on manufacturer's recommendations for the material to be welded. Turn on the machine, and turn on the water pump, if available. Attach the work clamp to the welding table or workpiece. Flip down your helmet, activate the foot or finger control if using one, and strike an arc by scratching the tip of the tungsten against the base metal. If your welder has a high frequency option, you do not need to scratch start the arc. Place a tack weld at each end of the joint to be welded. You may be able to tack the joint by simply fusing the two pieces with the heat of the torch, or you may have to use filler rod.

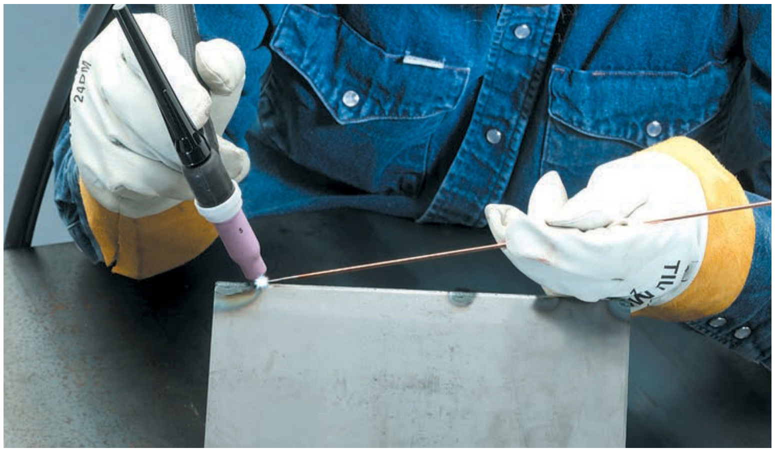

2 Position yourself to weld from right to left (if you are right handed) with the torch at a 15° angle to the right of center. Hold the filler rod in your left hand. Position yourself so you can comfortably hold the torch and filler rod for the duration of the weld.

STRIKING AN ARC:

Unless you have a high frequency option, you will need to physically strike an arc—called a scratch start. Rest the cup on the work-piece at a sharp angle. Move the tip until it briefly contacts the work, then angle it back again to start the arc. After the arc is started, lift the cup off the workpiece and establish the proper torch angle.

High frequency allows the arc to jump the gap without needing to create physical contact between the electrode and the workpiece.

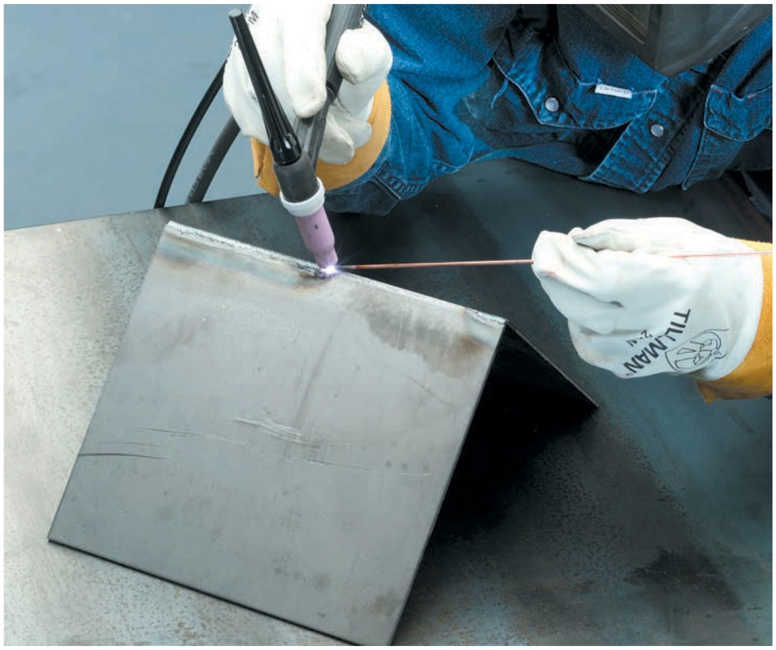

3 When a molten puddle has formed, dip the tip of the filler rod into the middle of the molten puddle. Keep the filler rod at a low angle to prevent disturbing the shielding gas. Keep the tip of the filler rod near—but not in—the puddle. Move the electrode to the left and continue the melting and dipping process.

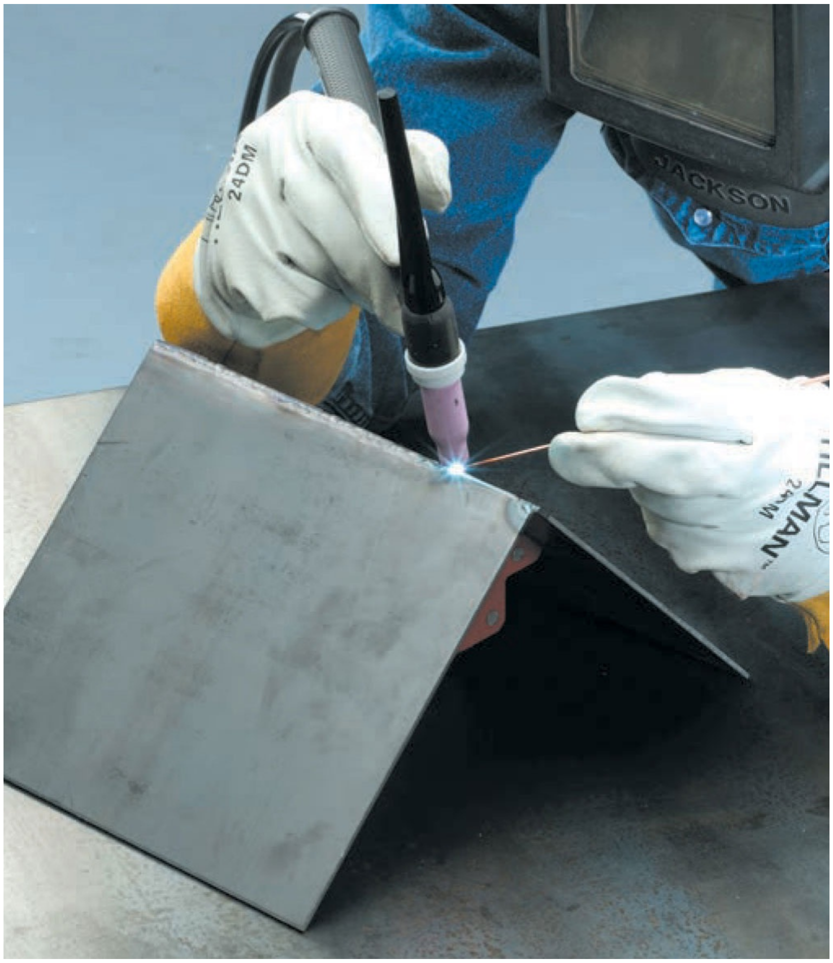

4 As you approach the end of your weld, you may need to adjust your travel speed because the buildup of heat in the material makes the molten puddle form more quickly at the end of the weld than at the beginning. You also may need to adjust the torch angle to be more shallow (not shown here) so that less heat is directed into the base metal.

Post-welding Sequence:

- Turn off cylinder valve.
- Purge gas from gas line.
- Turn off flow meter or gauge.
- Turn off machine.
- Coil hoses and cables off the floor.

Troubleshooting

Filler metal for gas tungsten arc welding comes in rod form and ranges from 1/16" to 3/16" in size. Rods are available in a variety of alloys, including aluminum, chromium and chromium nickel, copper, nickel and nickel alloys, magnesium, titanium, and zirconium. Specific alloy compositions are available for creating specific weld types on specific base metals. These filler metals are similar to those used in oxyfuel welding, with the exception of the carbon steel rods, which are not copper coated as they are for oxyfuel.

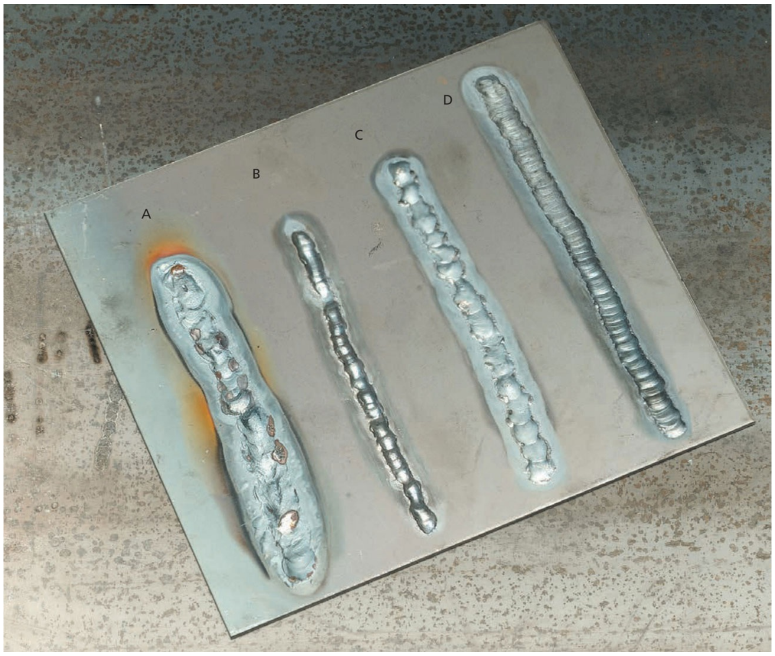

Becoming proficient in gas tungsten arc welding takes practice, and identifying problem welds is an important step. Weld A is too hot. Increase the travel speed or decrease the amperage. Weld B is too cold and is simply sitting on top of the base metal rather than penetrating it. Decrease the travel speed or increase the amperage. Weld C was done too quickly. Travel speed needs to be controlled and consistent. Weld D is a good quality weld with even ripples, good penetration, and a moderate crown.

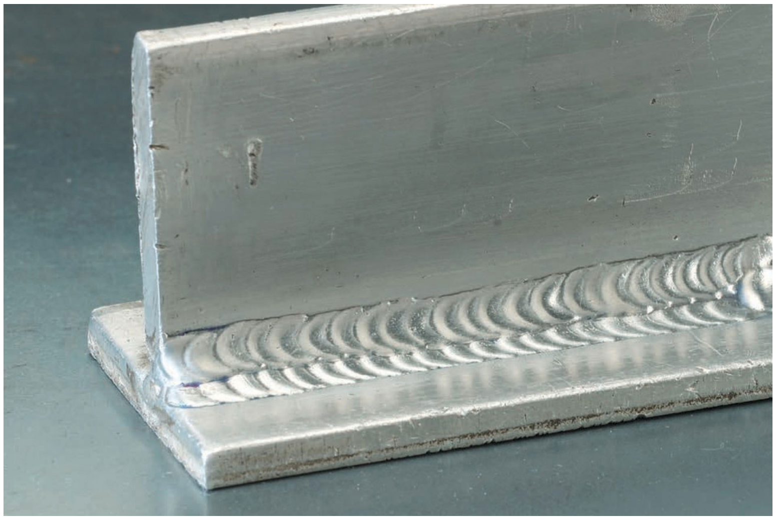

A well-done gas tungsten arc weld on aluminum has even ripples and good penetration. This sample weld shows two passes to create a fillet weld on 1/4" stock.

GTAW TROUBLESHOOTING:

PROBLEM	**SOLUTIONS**
Weld looks porous or sooty	• Make sure shielding gas is on and is correct type • Make sure shielding gas cylinder is not empty • Eliminate drafts • Make sure base metal is totally dry • Clean base metal thoroughly • Increase gas flow rate
Base metal distorts	• Tack weld parts before welding • Clamp parts down to rigid surface • Scatter welds to diminish heat buildup
Unstable arc	• Adjust electrode to work angle • Clean base metal thoroughly • Clean electrode • Connect work clamp to workpiece • Bring arc closer to work
	• Make sure polarity and current settings are correct

Electrode is rapidly consumed	• Increase electrode size • Increase gas flow • Decrease current • Increase gas post flow time • Use proper shielding gas

PLASMA CUTTING

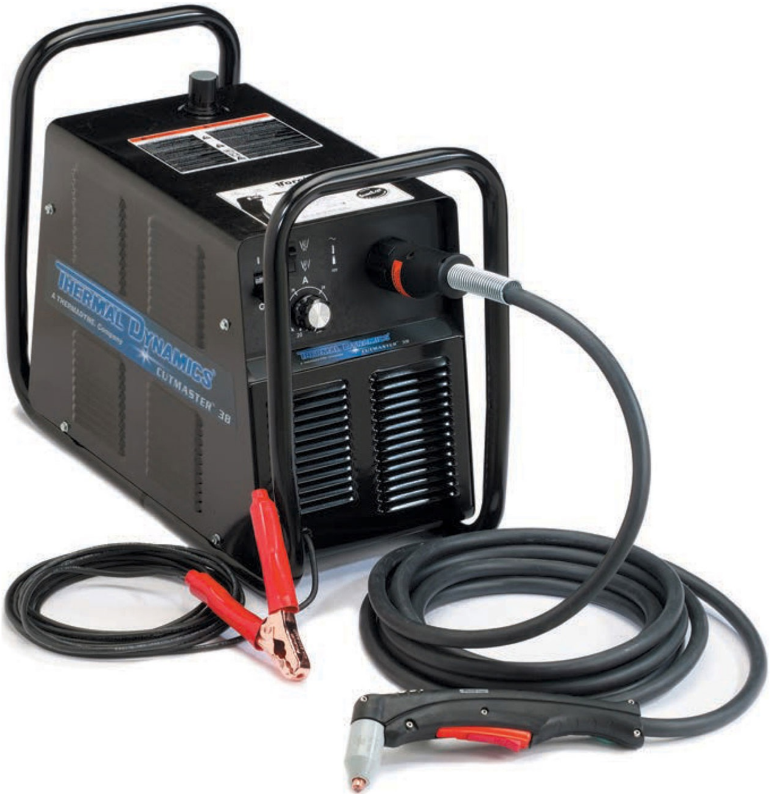
THERMAL DYNAMICS
A THERMADYNE Company
CUTMASTER 38

A plasma cutter consists of a power source with a compressed air or inert gas connection, work cable with clamp, supply cable, and torch.

Plasma cutting and welding are arc processes. Plasma is a gas that can conduct electricity because it is in a highly ionized form. The gas transfers the arc, but because of the constriction of the gas, the gas resistance is very high, which creates the extreme heat for plasma processes. Plasma cutting uses a very constricted arc to melt the base metal while compressed air or other gas blows the melted metal out of the kerf. This process can be used on any conductive metal—aluminum, brass, cast iron, copper, steel, stainless steel, and titanium—which sets it apart from oxy-acetylene cutting, which is only useable on oxidizable metals.

Though the plasma arc temperature is 40,000° F, it is so constricted and the cutting speed is so fast that the thermal distortion to the metal being cut is low. Plasma cuts are clean and are often weldable with no additional clean up if the base metal was initially clean. Properly done, the kerf is fairly straight sided, and no slag is present.

Safety. Plasma cutting produces sparks, smoke, and fumes. When cutting materials that contain chromium and nickel (such as stainless steel) or materials that have been galvanized or zinc coated, many toxins will be present in the fumes. When cutting through painted, bonded, or dirty metals, these surface additions often burn off, producing hazardous smoke. Plasma cutting often produces ozone and various nitrogen oxides that are colorless and can be odorless; both are hazardous. Therefore, good ventilation is important even when plasma cutting.

SAFETY

- Welding helmet with #8 filter
- Leather, wool, or cotton long pants, long sleeve shirt, and hat
- Heavy duty welding gloves
- Leather boots or shoes
- Ventilation

Plasma cutting may not seem as bright as other arc processes, but it is. It produces ultraviolet rays, so it is important to protect your skin and eyes. Because of the high level of open circuit voltage, plasma cutting has added electrical shock hazard. Make sure all surfaces around the cutting area are dry and all your clothing and gloves are dry. Turn off the power to the machine before changing any torch parts.

Never cut containers, tanks, or cylinders that may have held flammable materials. Even a small amount of flammable residue can cause an enclosed vessel to explode when cut.

Equipment. Plasma cutting machines are expensive, but as technological advances are made and demand increases, more companies are introducing smaller, less expensive home and small shop equipment. If you regularly need to cut stainless steel, aluminum, or the high strength steel now used in automobile bodies, this machine can be worth the cost. Small, 115-volt plasma cutters are available that will cut up to 3/8" material. Larger, 220-volt machines can cut thicker materials and have higher duty cycles. Aluminum and stainless steel require higher amperages, so check whether the machine easily can cut the materials with which you typically work.

Most plasma cutters require a compressed air source. Any shop air compressor that can deliver 65 psi is sufficient. It is critical that a filter be installed to keep the air dry and oil free. Follow the manufacturer’s directions for installing the filter. Some plasma cutters have an internal air compressor. These units often require more maintenance than models that utilize a separate compressor.

The plasma torch consists of a shield cup, cutting tip, starter cartridge, electrode, and torch handle. Most plasma torches have a manual switch on the torch body. The cutting tip and electrode are consumable parts. Check them frequently for wear and replace them when necessary.

CAMPBELL HAUSFELD

A dehumidifying filter is very important for the compressed air source for plasma cutting.

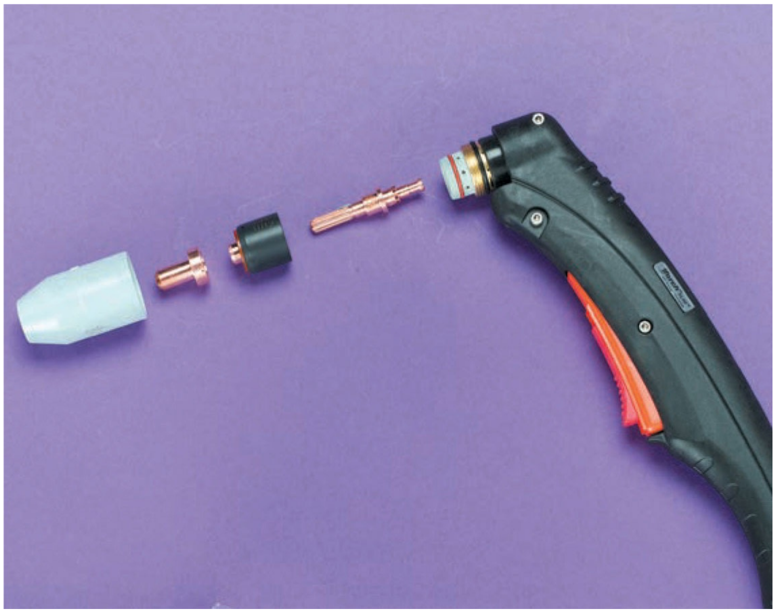

The plasma cutting torch consists of (left to right) a shield cup, cutting tip, starter cartridge, electrode, and torch handle.

Pre-cutting Checklist:

- Compressed air fittings are tightly attached.
- Cables are in good condition.
- Ventilation is sufficient to remove fumes from cutting area.
- Work area is dry.
- Flammable materials are removed from area.
- Sparks and slag from cutting will be contained.

How to Plasma Cut

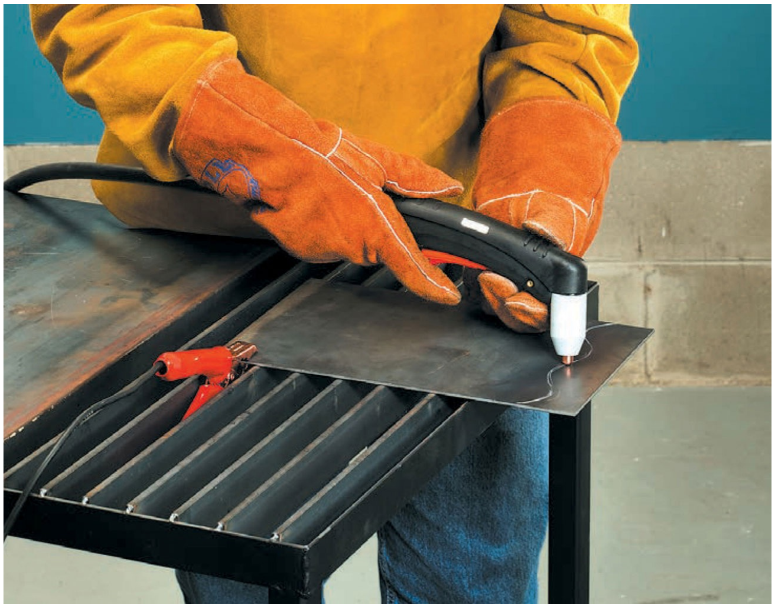

1 Check the manufacturer's recommended settings for the material to be cut. Arrange the material on a cutting table and determine the most comfortable cutting position. Attach the work clamp to the workpiece or welding table. Dry run through the cutting motion to practice speed and cutting position.

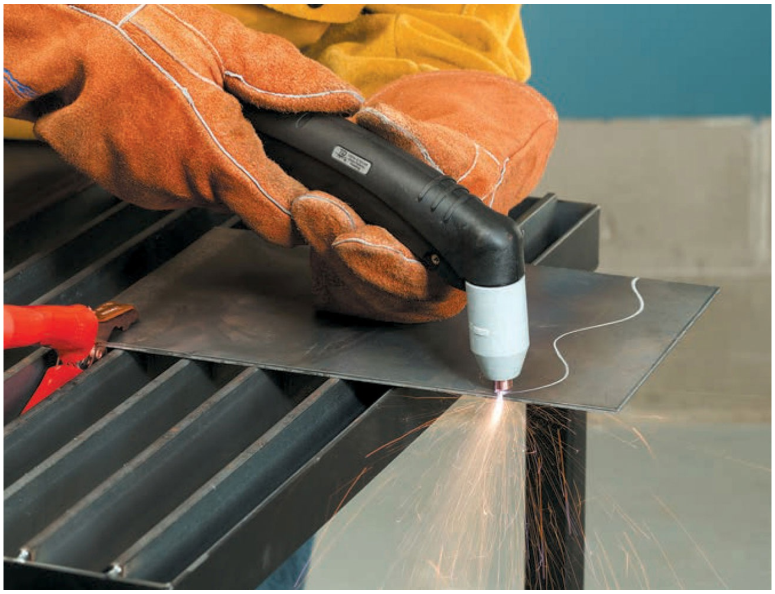

2 Check manufacturer's directions for arc starting procedure. Most current machines are designed to drag the nozzle directly on the material's surface. Activate the arc and hold the torch perpendicular to the surface of the material. Move steadily and smoothly along the entire length of the cut. Do not stop the arc until you have passed through the end of the cut.

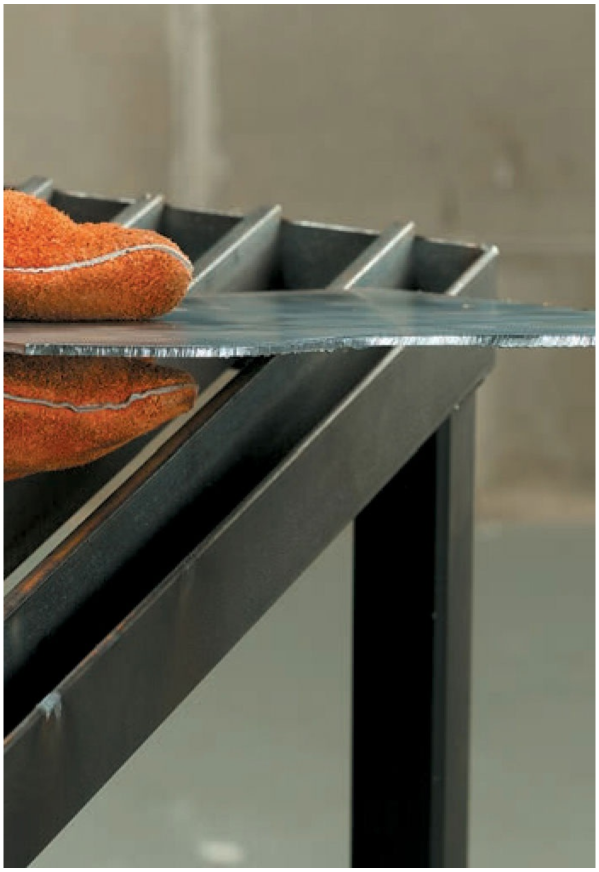

3 A good plasma cut has squared edges, small vertical ripples, and little or no slag.

Troubleshooting and Techniques

A high-quality circle cutting attachment for a plasma cutter is expensive, but worth the investment if you will be making repeated circle cuts. This unit comes with a magnetic pivot holder for steel and a suction cup pivot holder for all other metals.

PLASMA TROUBLESHOOTING:

PROBLEM	**SOLUTIONS**
Arc does not start	• Connect work clamp to workpiece • Turn on power source • Tighten cable connections • Make sure electrode is in working condition
Cut does not penetrate through material	• Cut more slowly • Make sure torch is held perpendicular to material • Make sure machine is capable of cutting that thickness • Increase amperage setting • Verify correct air flow and pressure
Excessive slag or dross	• Increase cutting speed • Decrease torch standoff distance • Replace worn torch parts • Adjust amperage setting
Torch parts consumed quickly	• Install or replace air filter to prevent oil or water from reaching torch • Cut metal within the capability of the machine • Make sure gas pressure is correct

Tip:

Making a good-quality plasma cut is dependent on a consistent travel speed. Consult the owner’s manual for cutting speeds for different material thicknesses and amperage settings. With the machine off, practice moving across the planned cut at the correct speed.

OXYACETYLENE CUTTING

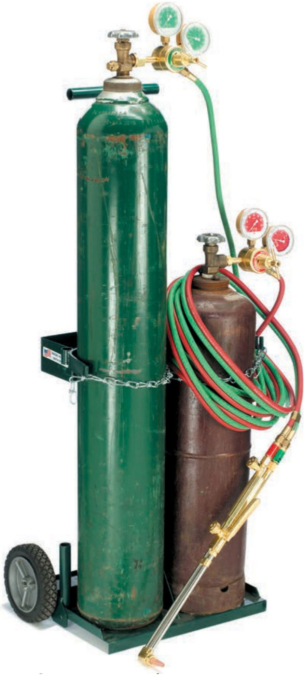

Oxyacetylene cutting equipment consists of two cylinders—oxygen and acetylene—regulators, hoses, and a cutting torch with tip.

Oxyacetylene cutting uses acetylene and oxygen to preheat metal to a temperature of 1600° F, then uses pure oxygen to burn away the heated metal. Because the cutting is achieved by oxidation of the metal, oxyacetylene cutting only works on metals that are readily oxidized at this temperature. Metals that can be cut with the oxyacetylene process are mild steel and low alloy steels. Oxyacetylene cutting can be used to cut thicknesses from less than 1/8" up to 12". Oxyacetylene cutting is inexpensive compared to machine cutting. It is very portable because no external power source is necessary, and it is faster than machine cutting. It is, however, limited to steel. The molten slag presents a fire and burn hazard, and the heat input to the base metal may cause distortion or changes to the properties of the steel.

Safety. The risk of starting a fire or being burned is high with oxyacetylene cutting. It is important to prevent molten slag and sparks from coming into contact with skin and flammable materials. Make sure you are wearing heavy-duty leather gloves with gauntlets, a leather jacket or leather-sleeved jacket, and leather boots. The work area should be cleared of flammable materials. If an item will be cut in place, make sure surrounding areas are protected from heat, sparks, and dropping slag. Never cut sealed tanks, cylinders, or items that have contained flammable materials. Never cut near gasoline tanks or fuel lines.

The Equipment. The basic equipment for oxyacetylene cutting is the same as for oxyacetylene gas welding, with the addition of a cutting torch or cutting attachment. An oxyacetylene cutting setup consists of an oxygen cylinder, an acetylene cylinder, regulators and hoses, and a torch with cutting attachment or a dedicated cutting torch. Cutting attachments and dedicated cutting torches have levers to activate the oxygen flow.

SAFETY

- Full-face #5 filter
- Leather, wool, or cotton long pants, leather jacket, and hat
- Heavy-duty welding gloves
- Leather boots or shoes
- Ventilation

Cutting Torch & Tips. Most oxyacetylene welding sets come with a cutting accessory that attaches to the torch body in the same way that the welding tips do. A dedicated cutting torch is a one-piece unit that has its own mixing chamber. Its greater overall length allows for more distance between the heat zone and the operator, and it can handle higher oxygen flow rates. A dedicated cutting torch is usually fairly expensive and not necessary unless very thick metals will be cut, which requires greater oxygen flow.

Both the cutting torch and the cutting attachment take a variety of cutting tip sizes. The cutting tip has a number of preheat holes surrounding the pure oxygen orifice. It is important to match the cutting tip to the thickness of the metal being cut. A welder often takes a one-size-fits-all approach to cutting tips, which always results in poor-quality cuts. Thinner metals require fewer preheat holes and a smaller oxygen orifice; thicker metals require more preheat holes and a larger oxygen orifice. Cutting tips are made of copper and can be damaged easily. They need to be cleaned regularly with an appropriately sized tip cleaner.

Fuel Gas Options. Because the metal being cut does not need to be brought to the melting point, various gases other than acetylene can be used. Propane, natural gas, propylene, and methylacety-lene-propadiene can be used as the preheating fuel source. Each of these gases requires specific regulators and may require different torches and torch tips. Check manufacturer’s recommended usage before substituting an alternate fuel gas.

A striker, tank wrench, pliers, and tip cleaner are needed for oxyacetylene cutting.

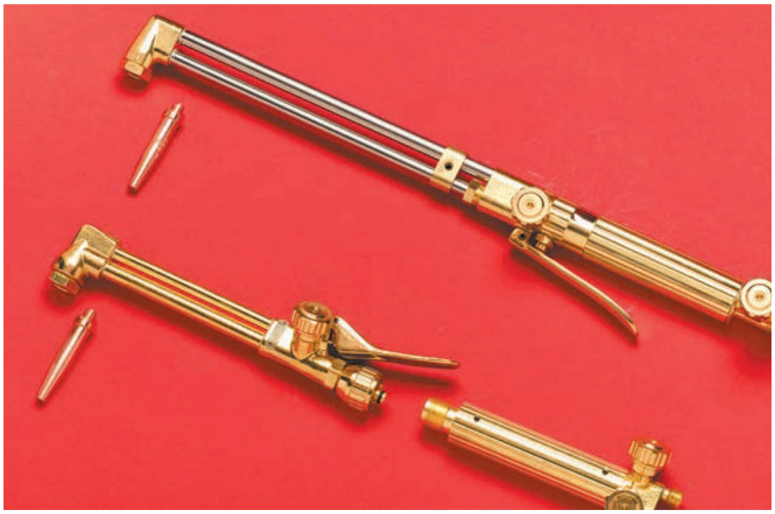

A dedicated cutting torch (top) is a one-piece unit with one oxygen and one fuel-gas valve. A cutting tip is attached to a standard torch body (bottom) and includes an additional oxygen valve. A range of tip sizes is available for both torch styles.

Pre-cutting Checklist:

- Surrounding area is free of flammable materials.
- There is room for cut metal to fall away from arms, legs, hoses, and cylinders.
- Hoses are free from damage before pressurizing the system.
- Ventilation is adequate to remove smoke and fumes from area.

How to Cut with Oxyacetylene

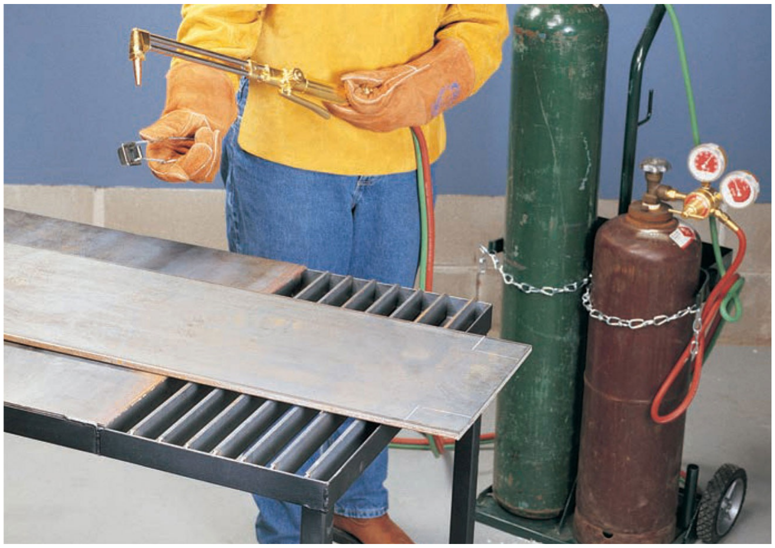

1 See pages 30 to 32 to learn how to set up, pressurize, and light an oxy-acetylene torch. Set up the material to be cut on a cutting table. Use soapstone to mark the cutting line, and practice your cutting position and bracing. Hold the torch in your left hand with your thumb and forefinger on the acetylene valve. Hold the striker in your right hand, 3" to 6" from the torch tip. Turn the acetylene valve $^1/_4$ to $^1/_2$ turn and strike sparks.

2 Adjust the acetylene so the flame is not smoking and is not separated from the tip (see page 32). Add oxygen to create a neutral flame. With a dedicated cutting torch, depress the oxygen lever to make sure neutral flame is maintained; if necessary, adjust with the oxygen lever depressed. With a cutting tip on a torch base, depress the oxygen lever and use the cutting tip adjustment valve to adjust the oxygen.

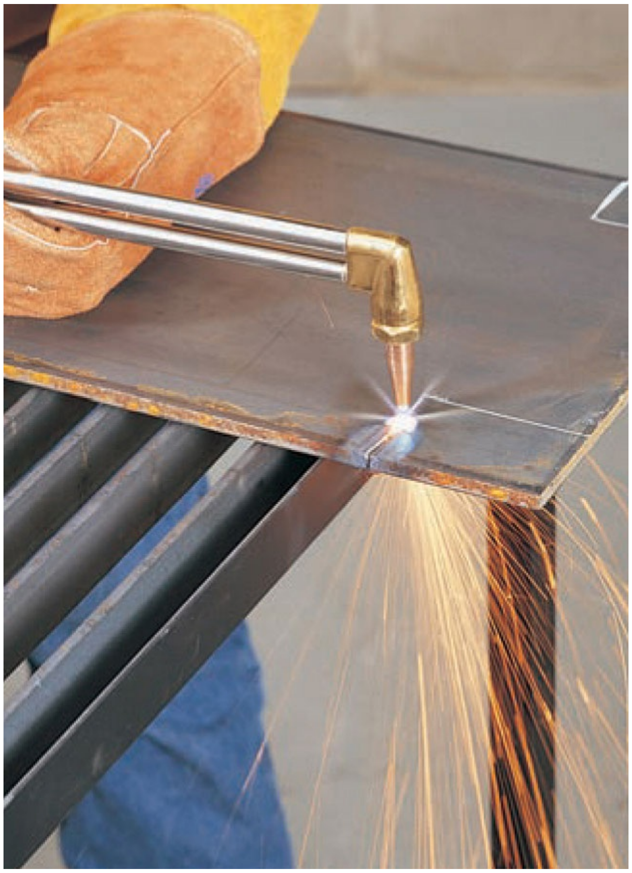

3 **Release the oxygen lever, flip down your face shield, and direct the flames at the edge of the cut. Once the metal begins to glow red and appears shiny but not yet molten, depress the oxygen lever. If it fails to pierce a hole immediately, release and continue to heat until depressing the lever pierces a hole. Move slowly and steadily along the cutting line, holding the torch at a 90° angle.**

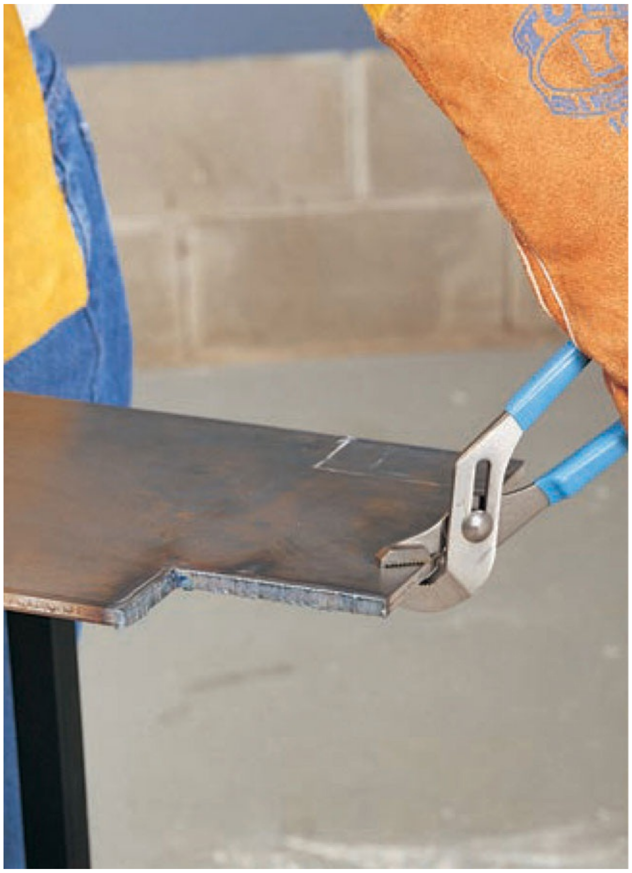

4 The finished cut should have very small vertical ridges (drag-lines), very little slag on the bottom, and the top edges should not be rounded over.

Oxyacetylene Cutting Tips

PIERCING HOLES AND MAKING CUTOUTS:

Piercing a hole: Preheat the material to a dull red. Pull back the torch from the surface and angle slightly. Squeeze the oxygen lever. As soon as the material has been pierced, return the torch tip to the perpendicular position and move back to just above the surface. Complete the hole.

Making a cutout: Begin in the center of the cutout and move to the cutout line after piercing a hole. A spiral approach to the edge works best. To give maximum protection to the surrounding material when piercing a hole, drill a 3/16" hole to start the piercing action.

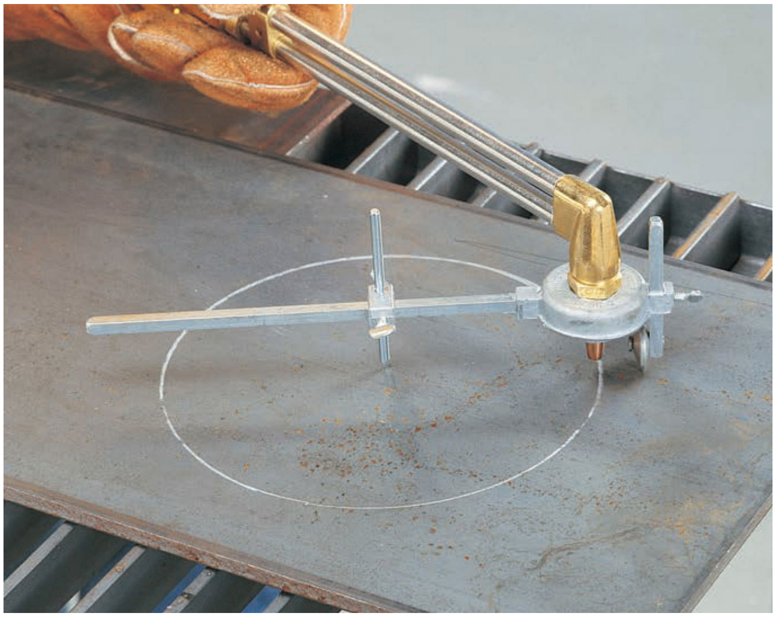

Circle cutting guides are available for making accurate circle cuts. If you need to make several circle cutouts, this can be a handy tool.

CUTTING DEFECTS:

A good oxyacetylene cut has square edges, small vertical draglines, and little slag. A dirty cutting tip is the cause of many poor-quality cuts. Clean the orifices with a properly-sized tip cleaner and use very fine sandpaper on a flat surface to polish the flat surface of the tip. Here are a few possibilities for causes of cutting defects:

Kerf is wider at the bottom than at the top (bell-mouthed).
Oxygen pressure is too high.

Top edge melted over.
Cutting speed is too slow or preheat flames are too long.

Irregular cut edge.
Speed is too fast or too slow, oxygen pressure is too high, or cutting tip is too small or too large.

WARNING:

- Never cut into a sealed container.

- Never cut into a tank or container that contains or may have contained flammable materials.
- Never cut near a gasoline tank or fuel lines.

Oxyacetylene Cutting Safety

Oxyacetylene cutting produces molten slag, sparks, and hot metal scraps. Move items that are to be cut away from burnable material. If the item cannot be moved, protect surrounding areas with sheet metal and fire retardant welding blankets (not tarps). Make sure sparks and slag will not fall into cracks, holes, or ventilation grates in the floor. Wet down any wooden material around the cutting area. Have water, sand, or a fire extinguisher on hand, and monitor the area for one half hour after cutting is completed. Plan where cut metal will fall so it does not hit your arms, legs, gas hoses, or cylinders.

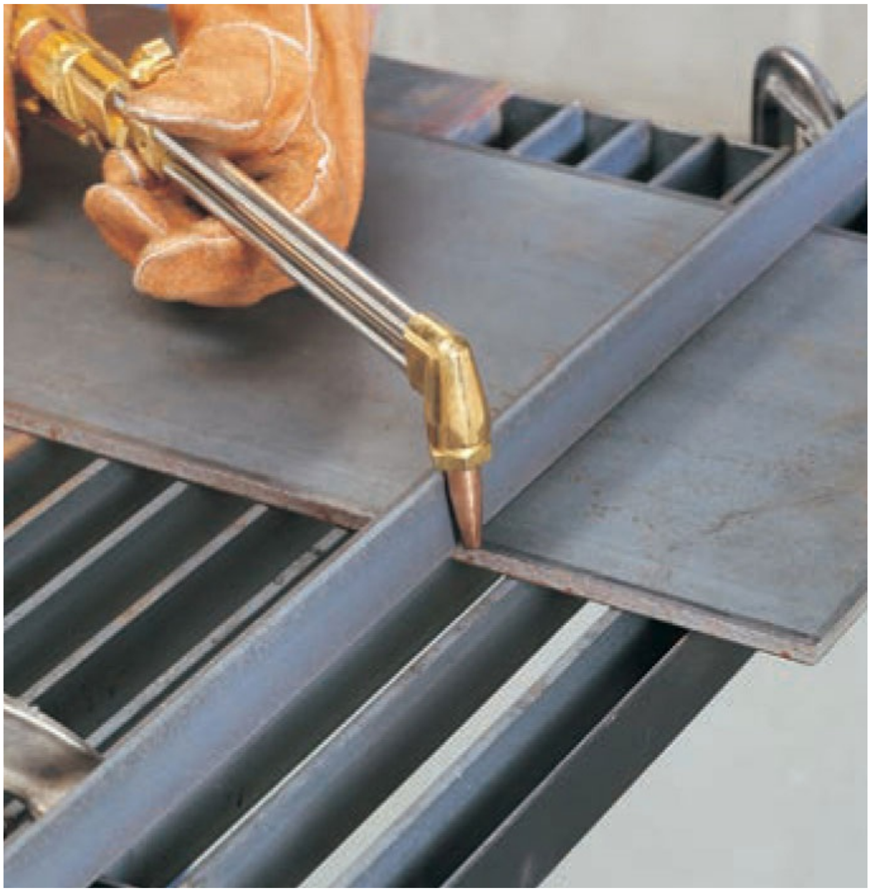

Use a piece of angle iron as a brace to support the torch or to hold pipe in place for cutting.

WELDING PROJECTS

Here's a chance to put all these welding and cutting techniques to use. The following chapters have directions for creating 23 different projects. Each contains a detailed cutting list, technical drawing, and step-by-step directions.

Among the shop projects you will find a sturdy welding table and a handy welding machine cart. The lighting projects feature a variety of candleholders and lamps. Included in the furniture projects are a delightful wine rack and a handy solution to creating a coffee table out of a slab of stone or wood. The outdoor projects will spark your creativity with an add-your-own-found-objects gate.

Shop Projects 74
Lighting Projects 90
Furniture Projects 104
Outdoor Projects 122

SHOP PROJECTS

Rolling Welding Curtain 76
Welding Table 78
Welding Machine Cart 82
Cylinder Cart 84
Grinder Stand 88

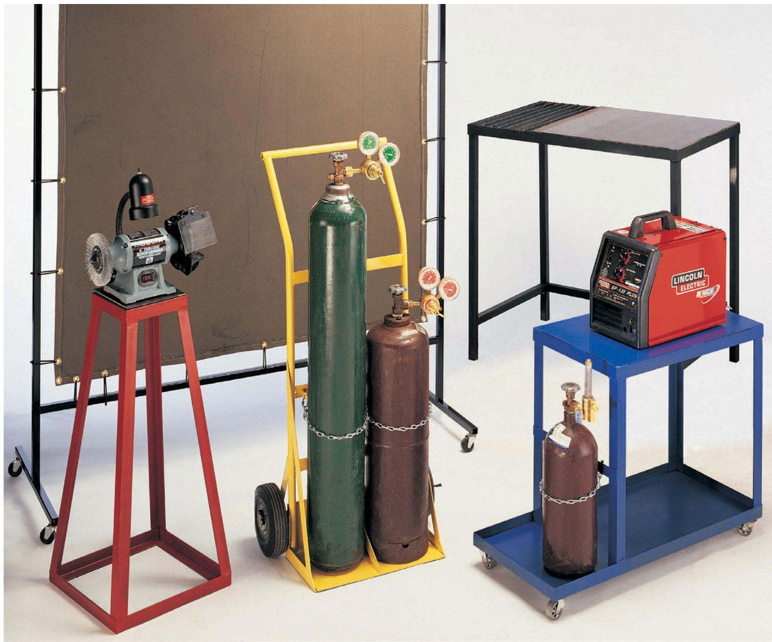
LINCOLN ELECTRIC

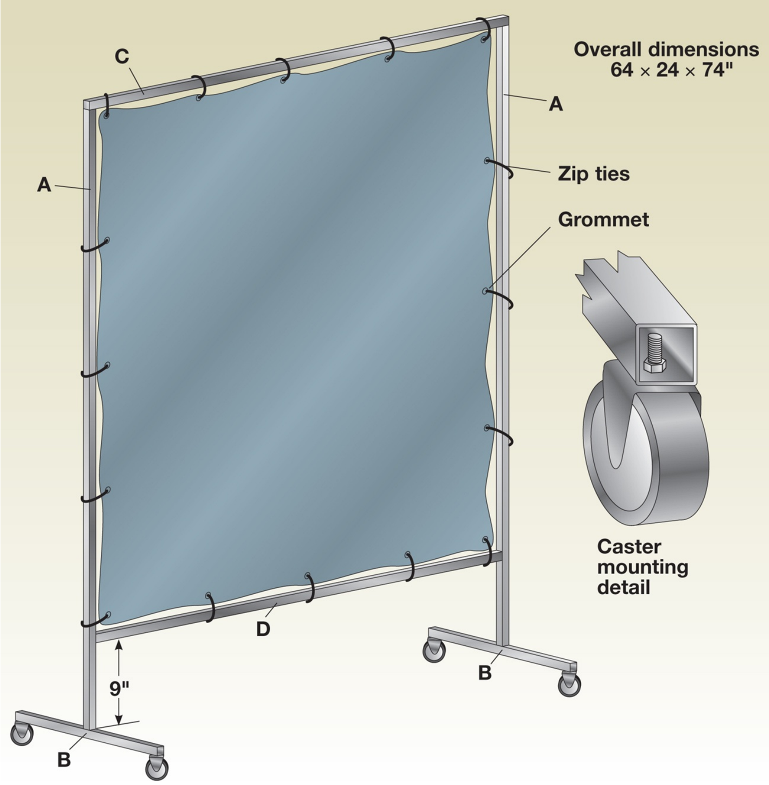
Overall dimensions
64 × 24 × 74"
C
A
A
Zip ties
Grommet
Caster
mounting
detail
D
9"
B
B

Rolling Welding Curtain

For arc welding and plasma cutting, it is very important to screen your work area so other shop workers, passersby, family members, and pets are not exposed to the damaging rays. This rolling curtain is quick and easy to make, and it offers the necessary screening while you work. You can purchase ready made welding tarps in a variety of colors, shapes, and sizes (see Resources, page 140). Or you can make your own welding tarp with 12-ounce cotton duck, fire retardant, and grommets.

PART	NAME	DIMENSION	QUANTITY
A	Sides	1 × 1" square tube × 72"	2
B	Wheel supports	1 × 1" square tube × 24"	2
C	Top crossbar	1 × 1" square tube × 64"	1
D	Bottom crossbar	1 × 1" square tube × 62"	1

Sixteen gauge or “thin wall” tube is sufficient for this project.

MATERIALS

- 1 × 1" square tube (27 feet)
- 4 threaded swivel casters, at least one locking
- 5 × 5 foot welding tarp
- Zip ties

How to Build a Rolling Welding Curtain

Attach the Wheel Supports to the Sides

1. Clean all parts with denatured alcohol, acetone, or degreaser. Prepare weld joint areas by wire brushing until shiny.

2. Cut the sides and wheel supports (A & B) to size.

3. Clamp one side piece to the work surface. Center a wheel support at the end of the side piece to form a T.

4. Check for square and tack weld along the butt joint between the two parts. Turn the assembly over and reclamp it to the work surface. Check that the wheel support is still square to the side piece. Make a final weld along the butt joint.

5. Repeat steps 3 and 4 to assemble the second side piece and wheel support.

Weld the Wheel Support T-joints

1. Clamp a side and wheel support assembly to the work surface so the wheel supports hang over the edge.

2. Weld the T-joint between the wheel support and the side piece. Turn the assembly over and weld the second T-joint.

3. Repeat steps 1 to 2 for the second side and wheel support.

Install the Crossbars

1. Cut the top and bottom crossbars (C & D) to size.

2. Place the top crossbar over the side pieces (see photo). Check for square, clamp in place, and tack weld.

3. Make a mark 10" up from the bottom of each wheel support. Align the lower edge of the bottom crossbar with the marks, check for square, and tack in place.

4. Flip the assembly over and check for square by measuring across both diagonals. If the measurements are equal, the structure is square. Clamp the assembly in place and complete the welds. Flip the assembly over and finish the welds on the other side.

Apply Finishing Touches

1. Paint the framework, if desired.

2. Drill holes for the threaded post swivels, 1/2" from the end of the wheel supports. Install the casters.

3. Attach the tarp to the framework with zip ties.

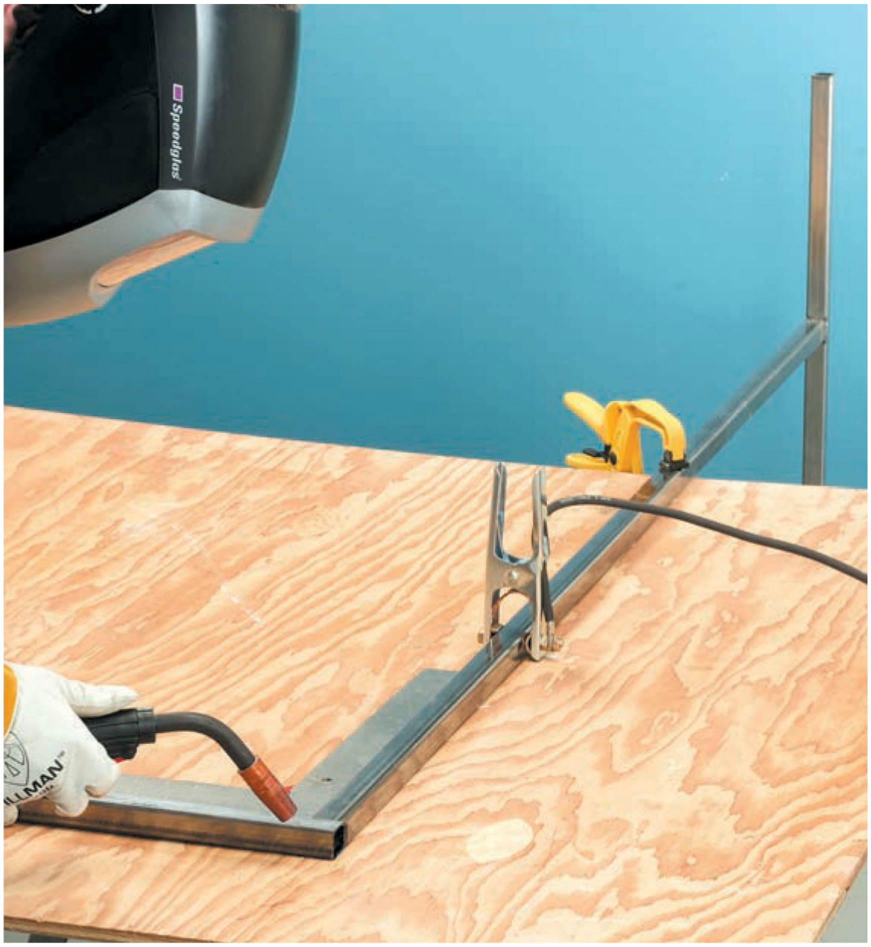

Clamp the sides to the table with the wheel supports hanging over the edge. Finish weld the wheel supports to the sides, then weld the top and bottom crossbars to the sides. Make sure the plywood work surface does not have smoldering embers when you are finished.

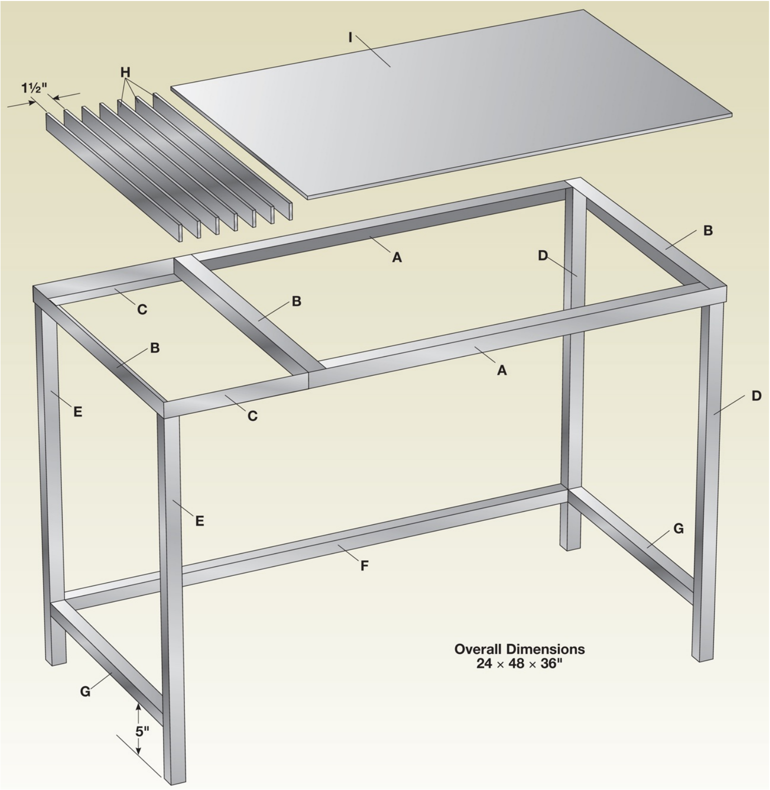
I
H
1½"
A
D
B
C
B
B
A
D
E
C
E
G
F
G
5"
Overall Dimensions
24 × 48 × 36"

Welding Table

Use it for arc welding, cutting, or oxyfuel welding—this welding table is a versatile addition to your shop. It is sized to allow you to work while sitting on a stool or while standing. You may want to check the scrap bin at the local steel yard to see what size plate steel can be gotten for a reasonable price. Some yards sell plate in 4 foot by 2 foot pre-cut sections that you can cut to size. Or you can have the piece custom cut. You can use material thinner than 3/16"—or thicker. Thicker material is better as the table top will have less distortion from the welding heat and the heat generated when you grind spatter off from the table top. You may be able to purchase a small piece of cutting grate; we've constructed our own. Use the sheet metal for the table top on top of sawhorses or a workbench as a work surface to build the rest of the table. You can enclose the area under the cutting table with sheet metal and a door to contain cutting slag and sparks.

PART	NAME	DIMENSIONS	QUANTITY
A	Table top supports (front & back)	1/8 × 1 1/2 × 1 1/2" angle iron × 24"	2
B	Table top supports (sides)	1/8 × 1 1/2 × 1 1/2" angle iron × 24"	3
C	Cutting table supports (front & back)	1/8 × 1 1/2 × 1 1/2" angle iron × 12"	2
D	Right side legs	1/8 × 1-1/4 × 1-1/4" square tube × 36"	2
E	Left side legs	1/8 × 1-1/4 × 1-1/4" square tube × 34 1/2"	2
F	Stretcher (rear)	1/8 × 1-1/4 × 1-1/4" square tube × 36"*	1
G	Stretcher (sides)	1/8 × 1-1/4 × 1-1/4" square tube × 21 1/2"*	2
H	Cutting table top	1/4 × 1 1/2" flat bar × 23-7/8"*	7
I	Table top	3/16" sheet metal 24 × 24"	1

*Approximate dimensions, cut to fit

Materials

- 1/8 × 1 1/2 × 1 1/2" angle iron (12 feet)
- 1/8 × 1-1/4 × 1-1/4" square tube (19 feet)
- 1/4 × 1 1/2" flat bar (14 feet)
- 3/16" sheet metal (2 × 2 feet)
- 4 leg levelers

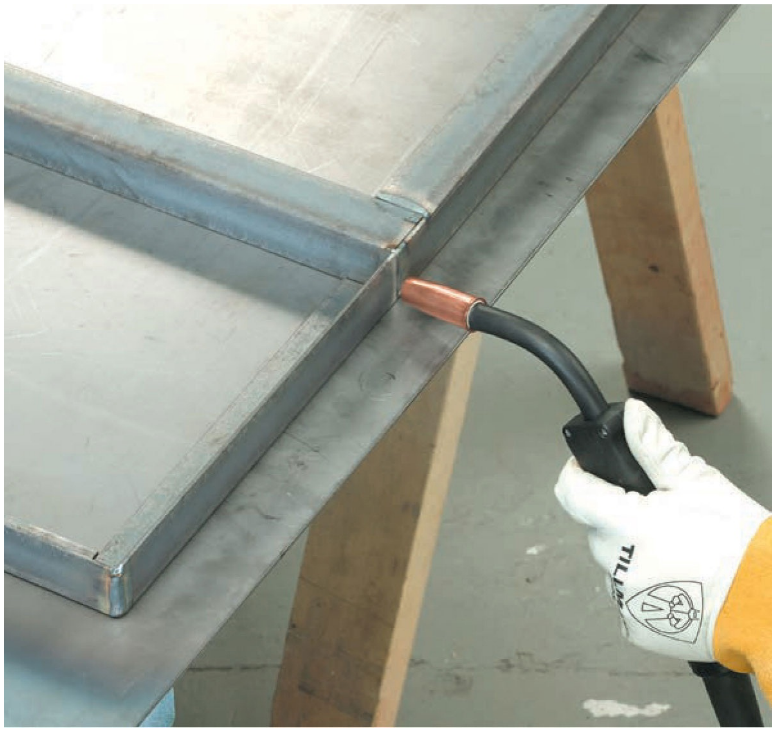

Weld the cutting table section to the table top assembly. The angle iron flange for the table top will face up to support the table top. The angle iron flange for the cutting table will face down to form a well to support the cutting table strips.

How to Build a Welding Table

Weld the Table Top Supports

Cutting notches at the end of the side supports is easier than cutting 45° miters and provides more welding surface area.

1. Cut the table top supports (A & B) and cutting table supports (C) to size. Cut 1 1/2" notches at each end of the three side supports (B) to create a 90° angle joint.

2. Place the table top front support (A) and a table top side support (B) together to form a right angle. Check for square and tack weld.

3. Repeat step 2 using the table top back support and another side support.

4. Assemble these two right angles to make a square. Check all corners for square, and check the assembly for square by measuring both diagonals—they should be equal. If not, adjust the supports so the assembly is square.

5. Complete each outside corner weld, re-checking for square as you go. Flip the assembly over and complete the remaining welds of the joints.

Complete the Table Top Supports

The cutting table supports are positioned with the flange at the bottom to hold the 1/4" strips. The table top supports are positioned with the flange at the top to support the table top (see photo).

1. Place the remaining side support (B) and the front and back cutting table supports (C) at right angles to form three sides of a rectangle. Check the corners for square and tack weld the pieces together.

2. Assemble the pieces so the table top supports have the flange at the top, the cutting table supports have the flange at the bottom, and the cutting table supports are abutting the table top supports. Tack weld where the table top sides butt together (see photo). Turn the assembly over and weld the remaining joints.

Prepare the Leg Assembly

1. Cut the legs (D & E) to size.

2. Measure the completed table-top. It may be slightly more or less than 24" deep and 36" wide. Adjust the stretcher lengths (F & G) to those measurements, minus the 1-1/4" for each leg thickness. Cut the stretchers to size.

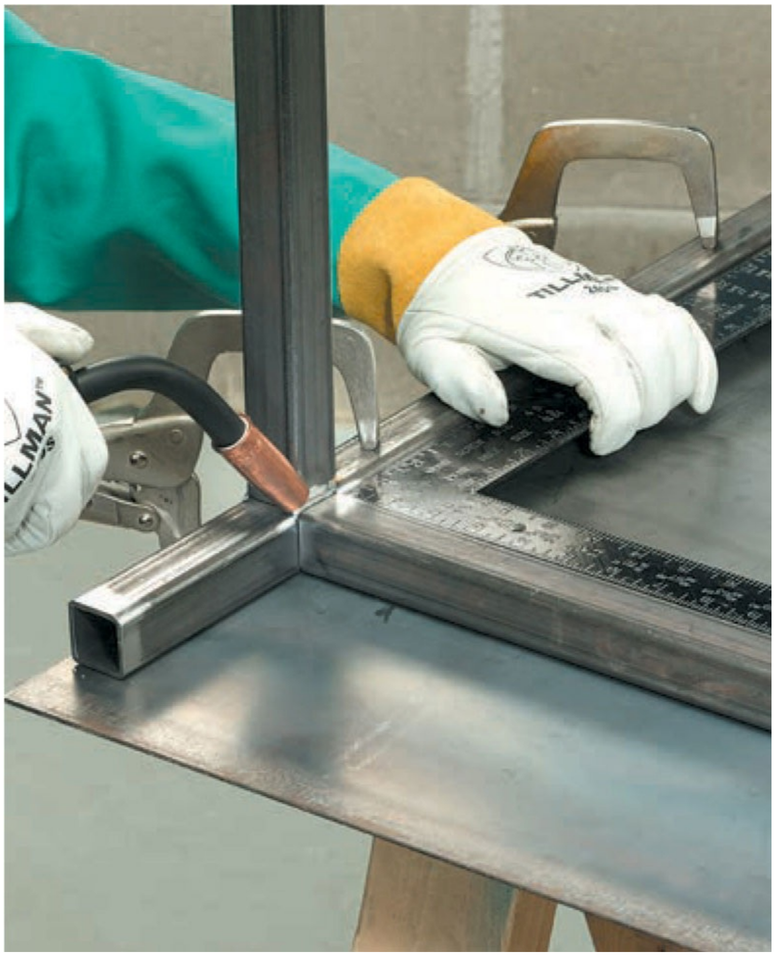

Place the back stretcher between the right and left side leg assemblies. Make sure the stretcher is square to the legs, then weld in place.

3. Mark both sets of side legs 5" up from the bottom. Starting with the right side legs (D), place a side stretcher (G) between the legs with the bottom of the stretcher aligned with the 5" mark.

4. Align the stretcher at a 90° angle to the side legs and clamp the assembly to your work surface. Tack weld the top inside angle at each end of the stretcher.

5. Repeat this process for the left side legs.

Finish the Leg Assembly

1. Place the leg assemblies on their sides with the rear side down, and clamp to your work surface.

2. Position the rear stretcher (F) between the right and left leg assemblies, aligning the bottom of the stretcher with the 5" mark.

3. Align the stretcher at a 90° angle to the assemblies and clamp in place. Tack weld the inside angles (see photo, below left).

Assemble the Table

The right legs fit inside the corner made by the angle iron, while the left legs are set back 1/8" from the edges of the angle iron.

1. Turn the table top and cutting table assembly upside down, then set the leg assembly into it.

2. Clamp the pieces in place. Check for square on both sides, front, and back.

3. Tack weld all corners. Check for square again, then weld all pieces into place.

Install the Cutting Table Top

Tack welding the cutting table grating allows you to remove and replace the 1/4" strips as they become worn from the cutting torch.

1. Cut the cutting table top grating (H) to size.

2. Place a strip of grating into the well formed by the cutting table top supports, 1-1/2" from the table top edge.

3. Tack weld the top edge of the grating to the angle iron.

4. Place another grating strip 1 1/2" from the first strip and tack weld in place. Continue building the cutting table top in this manner until complete (see photo, above).

5. Grind down the welds on the top of the table top support assembly.

Finish the Table

1. Place the table top (I) onto the assembly and tack weld twice on each side.

2. Weld the table top to the supports using 1" or 2" weld beads at both sides of each corner and twice along each side.

3. Grind down the welds, if desired. Wire brush, sand, or sandblast the entire table.

4. Paint the table, but do not paint the table top or cutting grate.

5. Install leg levelers as needed.

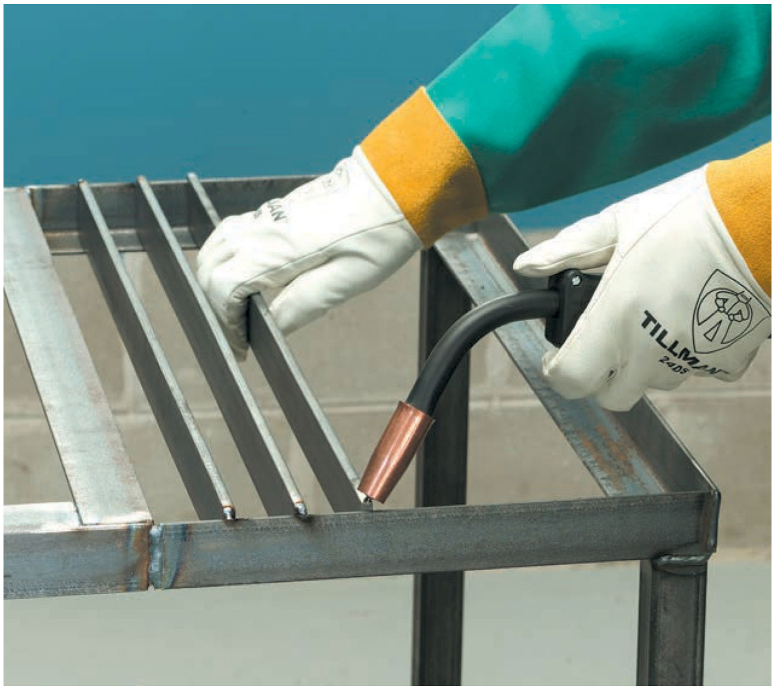

Tack weld grating strips to the angle iron every 1-1/2" to form the cutting table top.

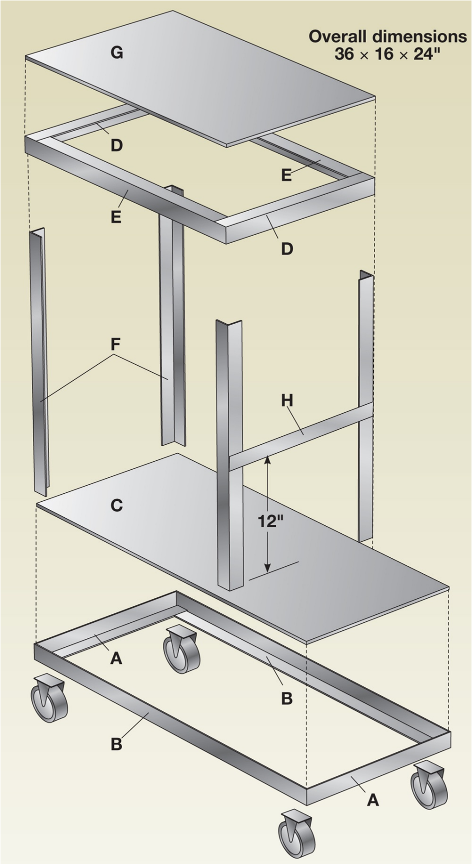
Overall dimensions
36 × 16 × 24"
G
D
E
E
D
F
H
C
12"
A
B
B
A

PART	NAME	DIMENSIONS	QUANTITY
A	Bottom front & back 1/8 × 1 1/2 × 1 1/2" angle iron × 16"	2	
B	Bottom sides	1/8 × 1 1/2 × 1 1/2" angle iron × 36"	2
C	Tray	1/8" sheet × 16 × 36"*	1
D	Top front & back	1/8 × 1 1/2 × 1 1/2" angle iron × 16"	2
E	Top sides	1/8 × 1 1/2 × 1 1/2" angle iron ×	2

		26"	
F	Legs	1/8 × 1 1/2 × 1 1/2" angle iron × 24"	4
G	Top	1/8" sheet × 16 × 26"*	1
H	Brace	1/8 × 1-1/4" flat bar × 16"	1
*Approximate dimensions, cut to fit			

Welding Machine Cart

You’ll want a cart for your GMAW or GTAW machine and gas cylinder—why not build it as your first project? It’s good practice, and the materials will cost much less than most commercial carts. If you keep your eyes open, you can probably pick up the angle iron for free the next time a neighbor throws away an old bedframe. Remember, however, to grind off the paint at the joints, otherwise you’ll have nasty fumes and poor quality welds.

Include a rubber pad and insulated chain for the gas cylinder (not shown) so an arc can’t accidentally be struck against the cylinder.

MATERIALS

- 1/8 × 1 1/2 × 1 1/2" angle iron (24 feet)
- 1/8" sheet metal (3 × 3 feet)
- 1/8 × 1-1/4" flat bar (16")
- Swivel casters (4)
- Eyebolts (2)
- Chain (2 feet)
- Threaded chain links (2)
- Rubber pad (10 × 16")
- Rubber tubing (2 feet)

How to Build a Welding Machine Cart

Assemble the Base & Tray

1. Cut the bottom front and back (A) to size. Cut the bottom side pieces (B) to size. Instead of mitering the corners, cut a 1 1/2" notch at each end (see diagram).

2. Arrange the front, back, and side pieces to form a rectangle. Check the corners for square, clamp in place, and tack weld the pieces together.

3. Measure diagonally across the corners to make sure they are square. If the measurements are equal, the base is square. If they are not, adjust the pieces until they are.

4. Clamp the base in place. Weld each outside corner. Turn the base over and weld each joint. Do not weld inside the base, as the weld beads will interfere with the tray and leg placement.

5. Measure the inside dimensions of the base and cut the tray (C) to size.

6. Insert the tray into the angles of the base. Tack weld the tray at points along the perimeter.

7. Turn the assembly over and place three 1" welds along each long side, and two 1" welds along each short side to join the tray to the base.

Assemble the Top

1. Cut the top front and back (D) and top side pieces (E) to size. Cut a 1 1/2" notch at each end of the side pieces.

2. Arrange the front, back, and side pieces to form a rectangle. Check the corners for square, clamp in place, and tack weld the pieces together.

3. Measure diagonally across the corners to make sure they are square. If the measurements are equal, the top assembly is square. If they are not, adjust the pieces until they are.

4. Clamp the pieces in place. Weld each outside corner. Turn the assembly over and weld each joint. Do not weld inside the top assembly, as the weld beads will interfere with the top and leg placement.

5. Measure the top assembly to allow a 1/8" setback on all sides, and cut the top (G) to size.

6. Grind down the welds and place the top over the assembly. Tack weld the top along the perimeter, then finish weld with two 1" welds per side.

Attach the Legs

1. Cut the legs (F) and brace (H) to size.

2. Turn the top upside down. Place a leg in one of the corners. Use a scrap piece of angle iron to assist in clamping (see photo). Make sure the leg is square to the top and tack weld into place.

3. Repeat step 2 to attach the other three legs to the corners of the top.

4. Place the legs and top assembly into the bottom assembly. The back legs should fit into the corners of the bottom assembly, and the front legs into the angles of the side pieces of the base.

5. Make sure the legs are square to the base, clamp, and tack weld.

6. Measure across the diagonals between the base and the top to make sure the unit is square. Adjust, if necessary, and finish welding the legs to the top and base.

7. Measure 12" from the bottom of the front legs Center the brace over the marks, check it for square, and clamp in place. Weld the brace to the front legs.

Add Finishing Touches

1. Sand and clean the cart, and paint.

2. Drill holes 2" in from the ends of the brace and attach the eyebolts. Use the threaded chain links to attach the chain.

3. Turn the cart over. Grind down the welds at the corners and attach the casters. If you are welding the casters, make sure you grind off the zinc coating.

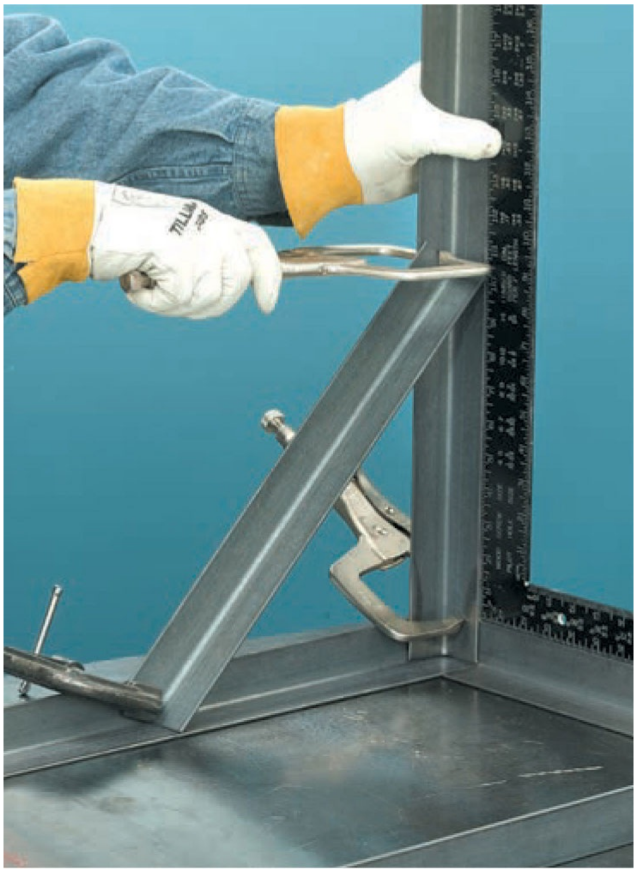

Use a scrap of angle iron to hold the legs in place for tacking.

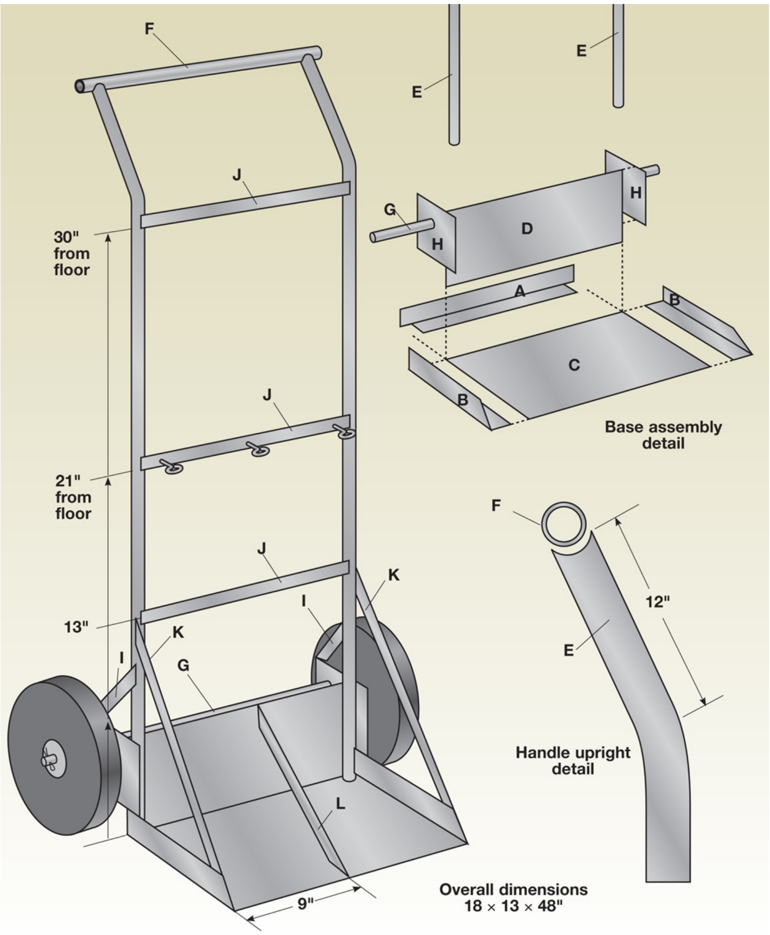
F
J
30" from floor
J
21" from floor
J
13"
K
I
K
I
G
L
9"
E
E
G
H
D
H
A
B
C
B
Base assembly detail
F
12"
E
Handle upright detail
Overall dimensions 18 × 13 × 48"

Overall dimensions
18 × 13 × 48″

Cylinder Cart

You can purchase a cylinder cart for your oxyfuel rig—your dealer might even give you a discount—but it’s fun and challenging to make your own. This cart is constructed with 1/8” stock, which is oxyfuel weldable, but we used GMAW. The base platform and back support can be flame or plasma cut. If you have access to a heavy-duty metal brake, you can make the base platform and back support from one 16-1/4 × 17-3/4” piece and bend it to 90° (allow 1/4” for the radius of the bend).

PART	NAME	DIMENSIONS	QUANTITY
A	Base back	1/8 × 1 1/2 × 1 1/2" angle iron × 18"	1
B	Base sides	1/8 × 1 1/2 × 1 1/2" angle iron × 10"	2
C	Base platform	1/8" sheet × 10 × 17-3/4"*	1
D	Back support	1/8" sheet × 6 × 17-3/4"*	1
E	Handle uprights	1/8 × 1" round tube × 60"*	2
F	Handle	1/8 × 1" round tube × 18"	1
G	Axle	5/8" round bar × 22"*	1
H	Axle brackets	1/8" sheet × 4 × 4 1/2"	2
I	Bracket supports	1/8 × 1 1/2" flat bar × 8"*	2
J	Crosspieces	1/8 × 1 1/2" flat bar × 17"*	3
K	Side supports	1/8 × 1-1/4" flat bar × 24"*	2

L	Base support	1/8 × 1 1/2" flat bar × 12"*	1

*Approximate dimensions, cut to fit

MATERIALS

- 1/8 × 1 1/2 × 1 1/2” angle iron (3 1/2 feet)
- 1/8” sheet metal (18 × 24” sheet)
- 1/8 × 1” round tube (12 feet)
- 5/8” round bar (2 feet)
- 1/8 × 1 1/2” flat bar (7 feet)
- 1/8 × 1-1/4” flat bar (4 feet)
- 1” eyebolts with nuts (3)
- Chain
- Snap closures or threaded chain links (3)
- 8” wheels with hubs (2)
- Washers and cotter pins (2)

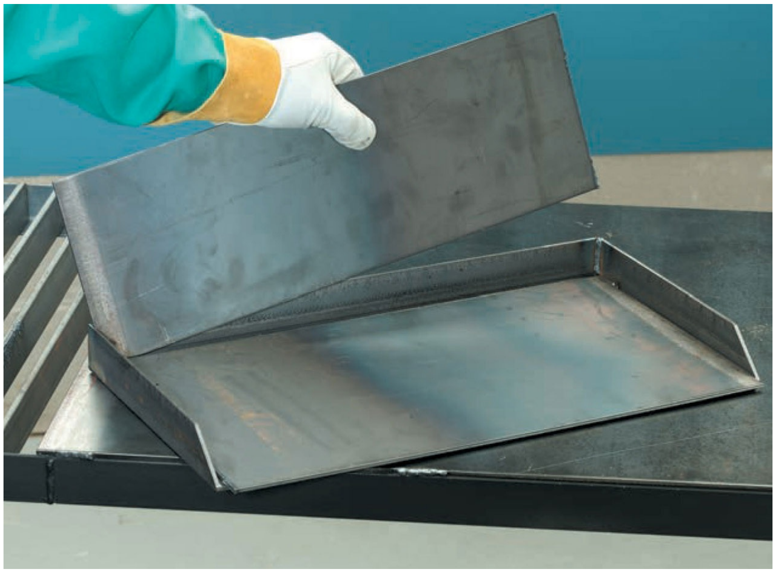

After welding the base back and sides, cut the base platform and back support to fit, then weld them in place.

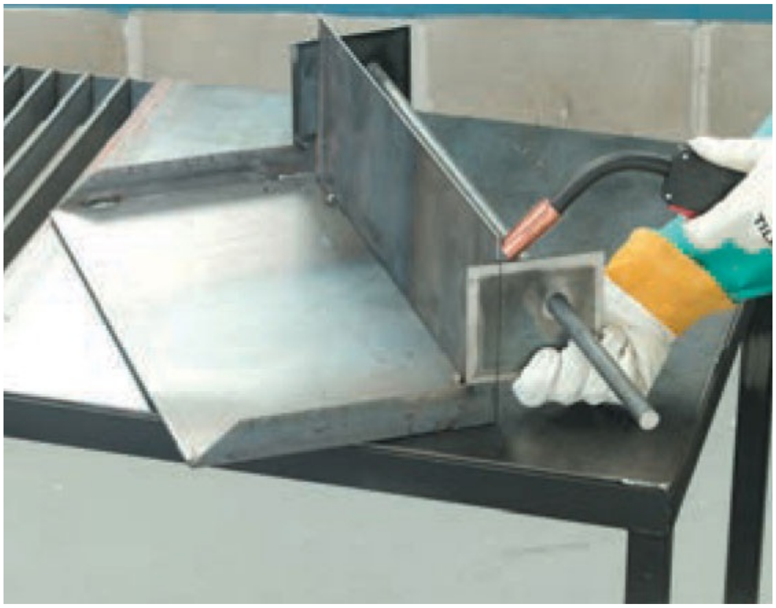

Tack the axle brackets to the base assembly. Insert the axle to check for proper alignment, then finish weld the bracket joints.

How to Build a Cylinder Cart

Prepare the Cart Base

1. Cut the base back (A) and base sides (B) to size. Rather than mitering the corners, cut a 1 1/2” notch in each end of the base back (see diagram).

2. Mark and cut a triangular section off the front end of each base side to soften the edge.

3. Lay out the base back and sides to form a three sided rectangle. Square the corners and clamp the assembly to your work surface.

4. Tack the joints at both ends of the base back. Turn the assembly over and weld the corner and butt joints.

Install the Base Platform & Back Support

1. Measure the inside width of the base assembly and cut the base platform (C) and back support (D) to size.

2. Place the base platform into the base assembly and tack weld it into place along the base sides.(Do not weld along the base back, as that would prevent the back support from fitting flat against the base platform.)

3. Turn the base over and place two 1” welds along each side where the base platform meets the back and sides.

4. Set the base flat on the work surface, right side up, and place the back support upright against the inside of the base back (see photo).

5. Clamp the back support to the base back. Place three 1” welds on the back side of the support where it meets the angle iron.

6. Place two 1” welds on the inside T-joint, making sure not to weld within two inches of the corner.

Attach the Axle Brackets

1. Cut the axle brackets (H) to size.

2. Place the base assembly on its side and align an axle bracket along the top edge of the base side, resting it on top of the back support and extending 3” off the back side.

3. Measure 4” up from the bottom of the base assembly and 1 1/2” from the back edge to mark the axle location on the axle bracket. Drill, flame cut, or plasma cut a 5/8” hole at the mark.

4. Tack weld the axle bracket in place (see photo, above left). Insert the axle to check for proper alignment, then weld both sides of the T-joint between the bracket and the back support.

5. Weld the outside butt joint between the bracket and the base side.

6. Repeat steps 2 to 5 to weld an axle bracket to the other side of the base assembly.

Prepare the Handle Uprights

1. Cut the handle uprights (E) to length.

2. Make a 20° to 30° bend in one of the handle uprights, 13” from one end, using a heavy-duty conduit bender. Bend the other handle upright to match.

3. Cut both pieces so the curve is 12” from the top end and the overall upright height of each is 48”. Grind the tops of both handle uprights so they fit around the handle tube.

Assemble & Install Handle Uprights

1. Place a handle upright against the inside corner of the base assembly. Turn the handle so the curve points straight back. Tack weld the handle in place. Repeat this process for the other handle upright.

2. Cut the handle (F) to size.

3. Place the cart on its back. Set the handle against the cutouts in the handle uprights, keeping an equal amount of overhang on each end, and tack weld in place.

4. Finish weld the handle uprights to the base assembly. Place welds between the handle uprights and top back support, and between the handle uprights and axle brackets.

Attach the Crosspieces

1. Measure and cut the crosspieces (J) so they are slightly recessed against the handle uprights.

2. Drill three 3/16” holes—one in the center and one on each end of one crosspiece. Place this crosspiece 21” up from the base.

3. Set the remaining crosspieces at 13” and 30” from the base on the forward side of the handle uprights, and weld in place (see photo).

Attach the Supports

1. Turn the assembly on its side. Place a bracket support (I) in position across the handle upright and the axle bracket. Mark the angles on the support and cut it to size.

2. Weld the bracket support in place against the handle upright and axle bracket.

3. Set the side support (K) in position against the base side and the handle upright. Mark the angles and cut to fit. Weld the support in place.

4. Turn the assembly on its opposite side and repeat steps 1 to 3.

5. Mark the angles and cut the base support (L) to size to fit between the base platform and the base back.

6. Center the base support from side to side on the base assembly. Weld both sides of each T-joint.

Complete the Cart

1. Center the axle (G) between the axle brackets. Place the wheels on the axle to make sure the wheels barely make contact with the floor (see photo, right).

2. Mark the axle for its final length and cut to size. Drill holes in each end of the axle to fit the cotter pins.

3. Center the axle between the axle brackets and weld in place.

4. Complete any unfinished welds.

5. Wire brush or sandblast the cart. Clean the cart and paint as desired.

6. Attach the wheels to the axle. Insert cotter pins in the axle holes.

7. Affix eyebolts with nuts in the 3 holes in the middle crosspiece. Attach the chain to the eyebolts with snap closures or threaded chain links.

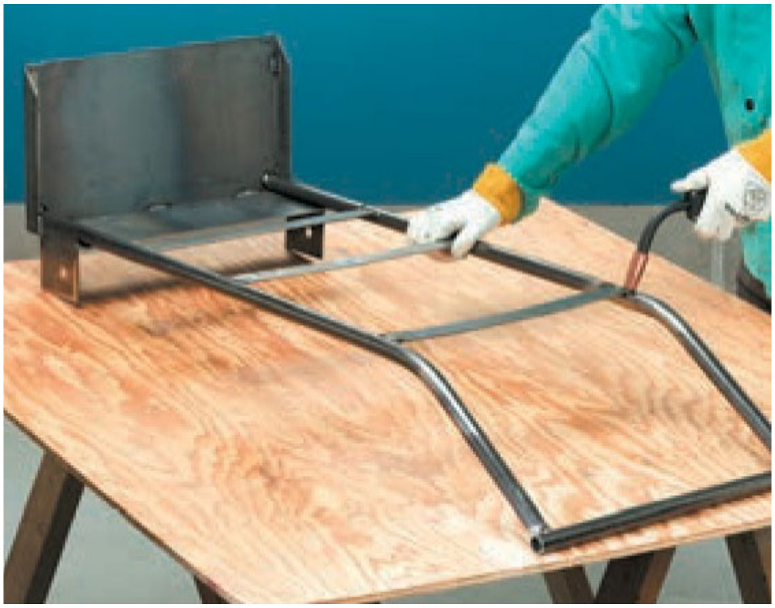

Lay the crosspieces across the handle uprights at the proper heights. Weld the pieces to the uprights.

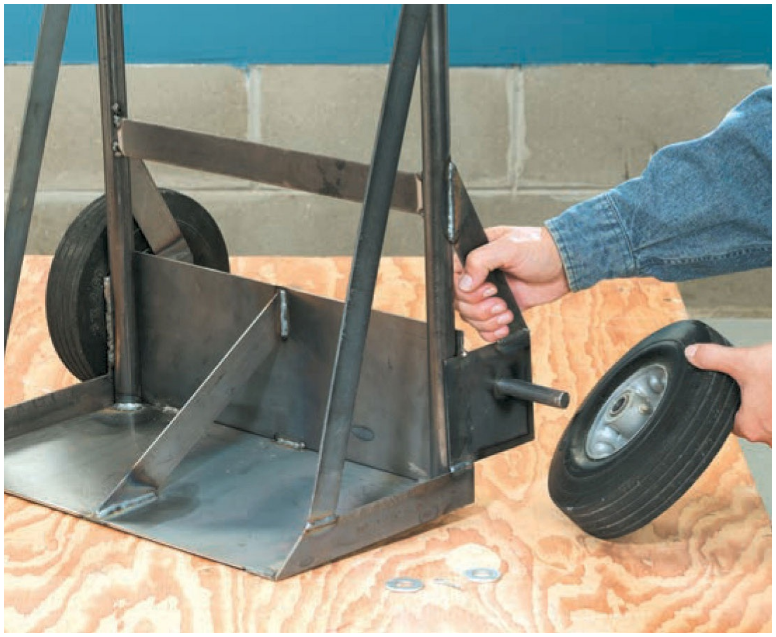

Trial fit the wheels before welding the axle into place. The wheels should just barely make contact with the floor.

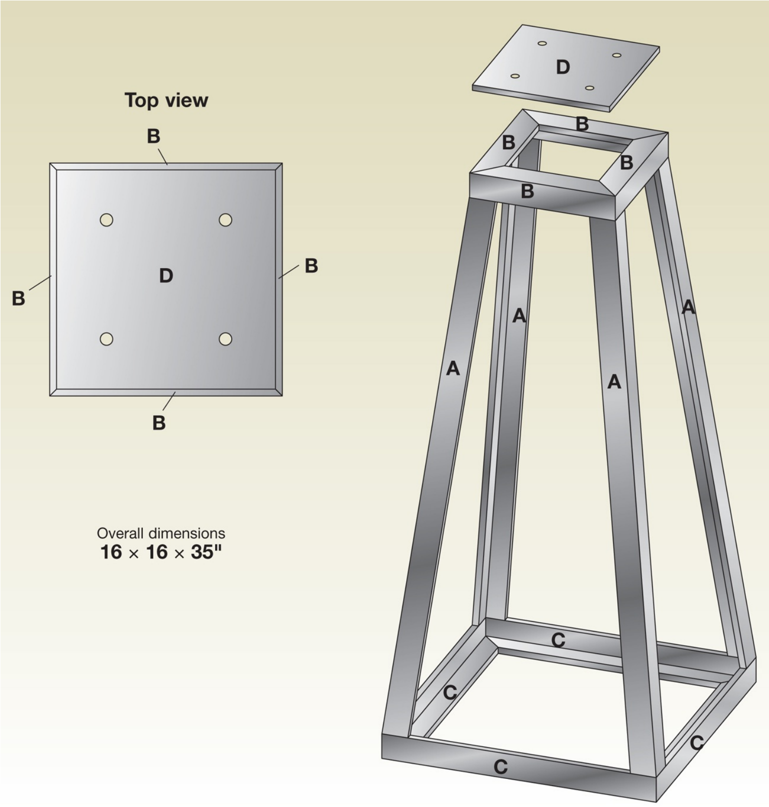
Top view
B
B
D
B
B
Overall dimensions
16 × 16 × 35"
D
B
B
B
B
A
A
A
A
C
C
C
C

Grinder Stand

This is a good project to make with angle iron from the discount bin at the local steel supply center. You can often buy short pieces for 10 cents a pound rather than the 30 to 50 cents you would normally pay. Take along your calipers or micrometer to measure thicknesses, as well as your measuring tape, cutting list, and tough gloves so you don't get cut digging through the scrap pile.

PART	NAME	DIMENSIONS	QUANTITY
A	Legs	3/16 × 1 1/2 × 1 1/2" angle iron × 32"	4
B	Top	3/16 × 1 1/2 × 1 1/2" angle iron × 8 1/2"	4
C	Bottom	3/16 × 1 1/2 × 1 1/2" angle iron × 16"	4
D	Platform	3/16 to 1/2" plate 8 × 8"	1

MATERIALS

- 3/16 × 1 1/2 × 1 1/2" angle iron (19 feet)
- 3/16" plate (8 × 8")

How to Build a Grinder Stand

Build the Top & Bottom Assemblies

1. Cut the top and bottom pieces (B & C) to length. Miter the corners at 45° or cut 1 1/2" notches into the pieces.

2. Assemble pairs of bottom pieces into right angles. Clamp the pieces to your work surface and tack weld at each corner.

3. Assemble the two right angles into a square. Check for square by measuring across each diagonal—if the measurements are equal, the assembly is square. Tack weld the corners.

4. Repeat steps 2 to 3 to assemble the top.

Mark & Cut the Legs

1. Cut the legs (A) to size.

2. Place the bottom assembly on your work surface and loosely clamp the legs into each corner.

3. Set the top assembly over the legs and clamp it in place. Make sure the top is level.

4. Mark each leg at the top and bottom where they cross the platforms (see photo).

5. Unclamp the legs and cut to size.

Assemble the Stand

1. Place one leg on top of a corner of the bottom assembly and tack weld it in place. Tack weld the other three legs in the corners the same way.

2. Align a corner of the top assembly with the top of one leg and tack it in place. Tack weld the top assembly to the other three legs the same way.

3. Adjust the stand so the top is level. You may need to remove tack welds and grind down some angles.

4. When you are satisfied with the alignment, complete the welds. To minimize distortion, alternate between the sides, top, and bottom as you weld.

Attach the Platform Plate

1. Cut the platform plate (D) to size. Drill four holes in the face to match your grinder mounting holes.

2. Set the plate over the top of the stand, aligning the edges with the outside of the top pieces. Weld the plate to the stand.

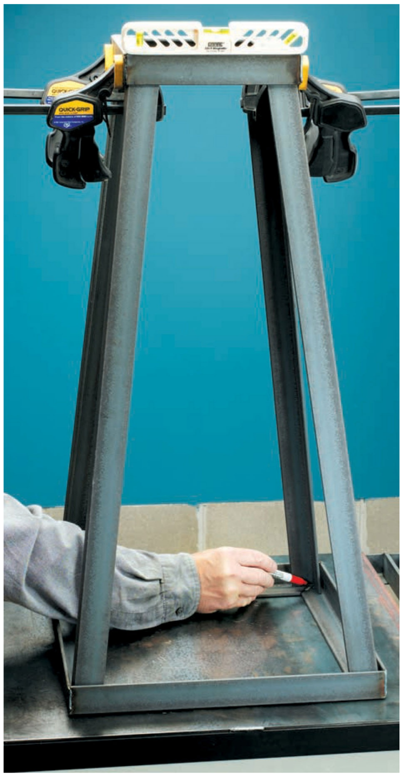

Assemble the grinder stand using clamps. Mark the top and bottom of each leg and cut to fit.

LIGHTING PROJECTS

Scroll Desk Lamp 92
Three Post Floor Lamp 94
Chandelier 96
Wall-mounted Candleholder 98
Tree-shaped Candleholder 100
Fireplace Candleholder 102

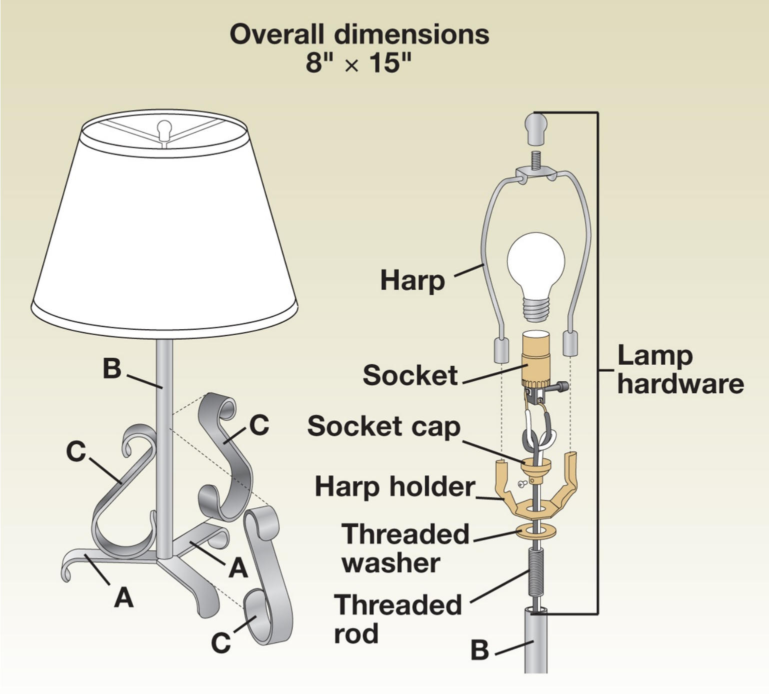
Overall dimensions
8" × 15"
B
C
C
A
A
C
Harp
Socket
Lamp hardware
Socket cap
Harp holder
Threaded washer
Threaded rod
B

Scroll Desk Lamp

In this project, you will make simple scrolls—or you can purchase wrought iron scrolls from a wrought iron supplier or blacksmith. (See Metal Shaping Techniques, page 19, or Resources, page 140.) Either way, this makes a delightful little desk lamp.

PART	NAME	DIMENSIONS	QUANTITY
A	Bases	1/8 × 3/4" flat bar × 5"	3
B	Shaft	1/2 " round tube × 9"	1
C	Scrolls	1/8 × 1/2" flat bar × 12"	3

MATERIALS

- 1/8 × 3/4" flat bar (15")
- 1/2" round tube (9")
- 1/8 × 1/2" flat bar (36")
- 1/2", 1" and 2" pipe
- Lamp hardware and shade
- 1/8" hollow threaded rod (1")

How to Make a Scroll Desk Lamp

Assemble the Base

1. Cut the base pieces (A) and shaft (B) to size.

2. Round one end of each base piece using a bench or angle grinder. Clamp the rounded ends to a 1/2" pipe. Curl the ends to form an arc.

3. Turn the base pieces upside down and arrange them so they are equally spaced at the 12, 4, and 8 o’clock positions (120° between adjacent pieces at their centerlines).

4. Use a grinder to shape the flat end on each base piece to fit against the other two legs. Once the pieces are fitted, center the 1/2" tube on the arranged pieces and mark around the tube.

5. Grind the legs so the tube fits in the shaped opening, but don't insert the shaft at this time.

6. Tack weld the undersides of the legs at each contact point.

Attach the Tube

1. Turn the base assembly right side up and insert the tube. Check for square and tack weld at three points.

2. Check for square again, adjust the pieces if necessary, and complete the weld around the base of the tube. (You can weld the tube from the bottom side only, but a solid weld all around the top looks nice.)

Prepare & Attach Scrolls

1. Cut the scrolls (C) to size.

2. Round both ends of the scroll pieces with a grinder.

3. Using a 1" pipe, make a 1" circle at one end of each scroll. Using a 2" pipe, make a 2" circle at the other end of the scrolls, facing the opposite direction. (Because of the springiness of the steel, the final circles will be slightly larger.)

4. Place one scroll against the base and the shaft, arranging it in a pleasing manner. Tack weld it in place where it touches the base and shaft.

5. Place and tack weld the second and third scrolls, taking care to align them carefully. When satisfied with the placement and alignment, tack weld on the opposite sides of the scrolls (see photo).

Apply Finishing Touches

1. Clean up the welds, and apply the finish of your choice.

2. Tap threads in the top of the shaft (or weld a small nut to the top of the shaft), and insert the hollow threaded rod. Thread the wire through the shaft and rod. Thread the washer, and place the harp holder over the threaded rod. Screw on the socket cap, and tighten the set screw. Wire the fixture (see diagram). Press the socket assembly into the socket cap, place the harp in the harp holder, and install the lightbulb and shade.

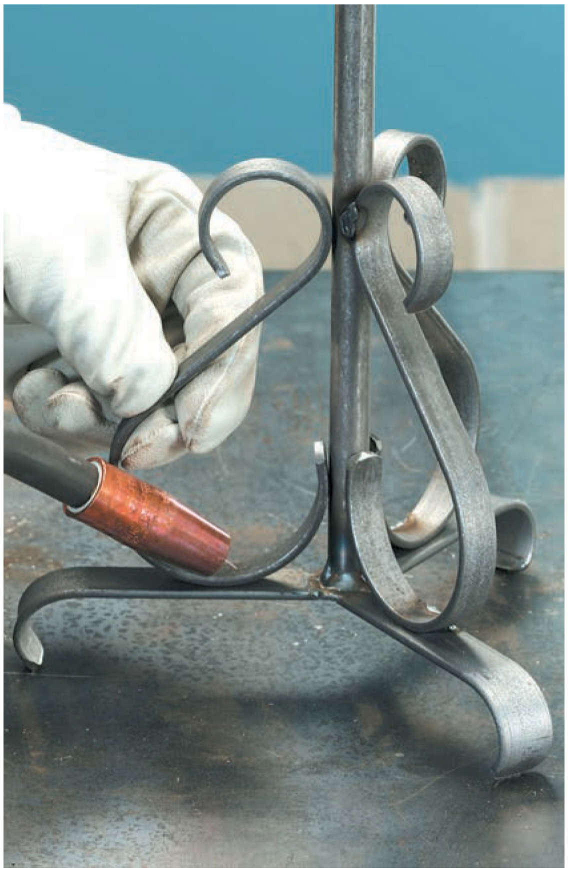

Arrange the scrolls against the base and shaft, then weld the scrolls in place on both sides.

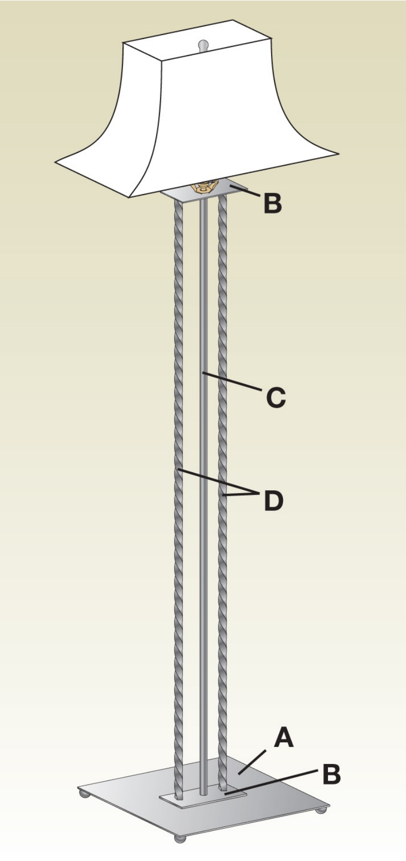
B
C
D
A
B

Three Post Floor Lamp

A welded lamp has a nice heft to it—with this 10" square base, it will be difficult to tip, no doubt! We used fully twisted rod because it creates a simple contrast of square twists and smooth rounds and is available in longer lengths. Partially twisted rod is also available. You also could use any of a variety of decorative hammered square bars or tubes from the sources listed in the Resource section on page 140. While you’re making a floor lamp, you could make a desk lamp to match. Just adjust the dimensions to fit.

PART	NAME	DIMENSIONS	QUANTITY
A	Base	3/16" sheet 10 × 10"	1
B	End caps	3/16" flat bar 2 × 4"	2
C	Center post	1/2" round tube 48-3/8"	1
D	Side posts	1/2" twisted rod 48"	2

MATERIALS

- 1/2" round tube (48-3/8")
- 1/2" twisted square rod (96")
- 3/16" sheet (10 × 10")
- 3/16 × 2" flat bar (8")
- 1" round balls (4)
- Lamp hardware and shade

How to Make a Three Post Floor Lamp

Prepare the Work Pieces

1. Cut the end caps (B) to size. Cut the center post (C) and side posts (D) to size.

2. Mark the end caps at 3/4", 2", and 3 1/4" from one end. Make a mark lengthwise through the center of each end cap.

3. At the intersection of the 2" and the centerline marks in each end cap, drill a 1/2" diameter hole.

Attach the Center Post to the End Caps

1. Insert the center post into the hole in one of the end caps. Check the post for square, then tack weld the post at the bottom.

2. Repeat step 1 to tack weld the other end cap onto the opposite end of the center post.

3. Stand the center post and end caps upright and carefully weld around the post.

Attach the Side Posts

1. Place the lamp assembly flat on your work surface. Center a side post over the 3/4" mark on one of the end caps. Check the post for square and tack weld in place.

2. Place the side post in position on the second end cap. Check for square and tack weld.

3. Repeat steps 1 and 2 to attach the second side post to the end caps at the 3-1/4" marks.

4. Square all the parts and complete each weld all around each post and end cap (see photo).

Attach the Post Assembly to the Base

1. Cut the base (A) to size.

2. Make a line 3" in from each side of the base and 4" in from the top and bottom. Draw lines at 5" lengthwise and widthwise to mark the center point.

3. Drill a 1/2" diameter hole at the center of the base for the cord.

4. Weld a 1" ball 1/2" from the edges at each corner of the base for the feet.

5. Turn the base right side up. Clamp the post assembly to the base, centering the base on the lines marked in step 2. The holes for the cord should line up. If not, enlarge the hole in the base.

6. Tack weld the post assembly to the base on two sides. Check for square. Complete the fillet weld around the base. Weld toward the tack points to minimize distortion.

Apply Finishing Touches

1. Grind down any spatter or uneven welds. Apply the finish of your choice.

2. Tap threads in the top of the shaft (or weld a small nut to the top of the shaft), and insert the hollow threaded rod. Thread the wire through the shaft and rod. Place the washer and harp holder over the threaded rod. Screw on the socket cap, and tighten the set screw. Wire the fixture (see diagram, page 92). Press the socket assembly into the socket cap, place the harp in the harp holder, and install the lightbulb and shade.

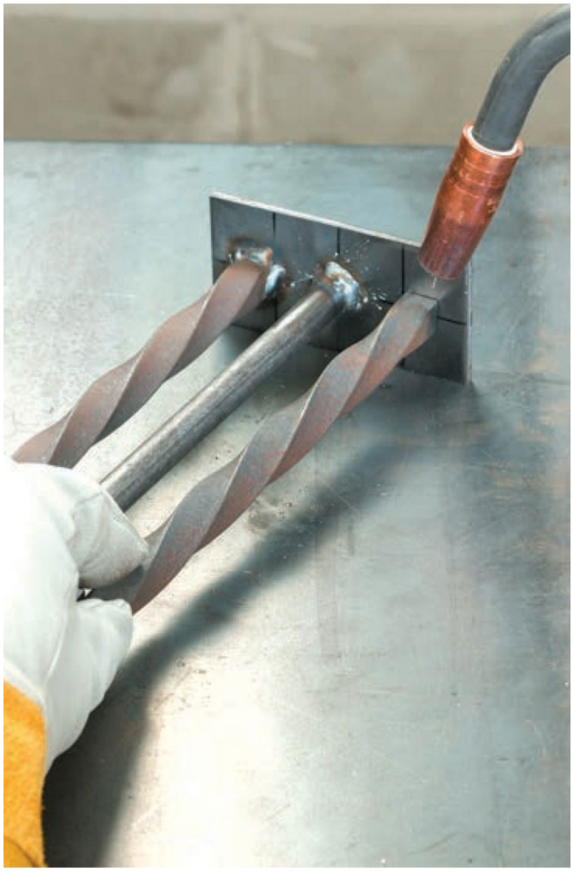

Make sure the posts are square to the end caps, then weld them in place.

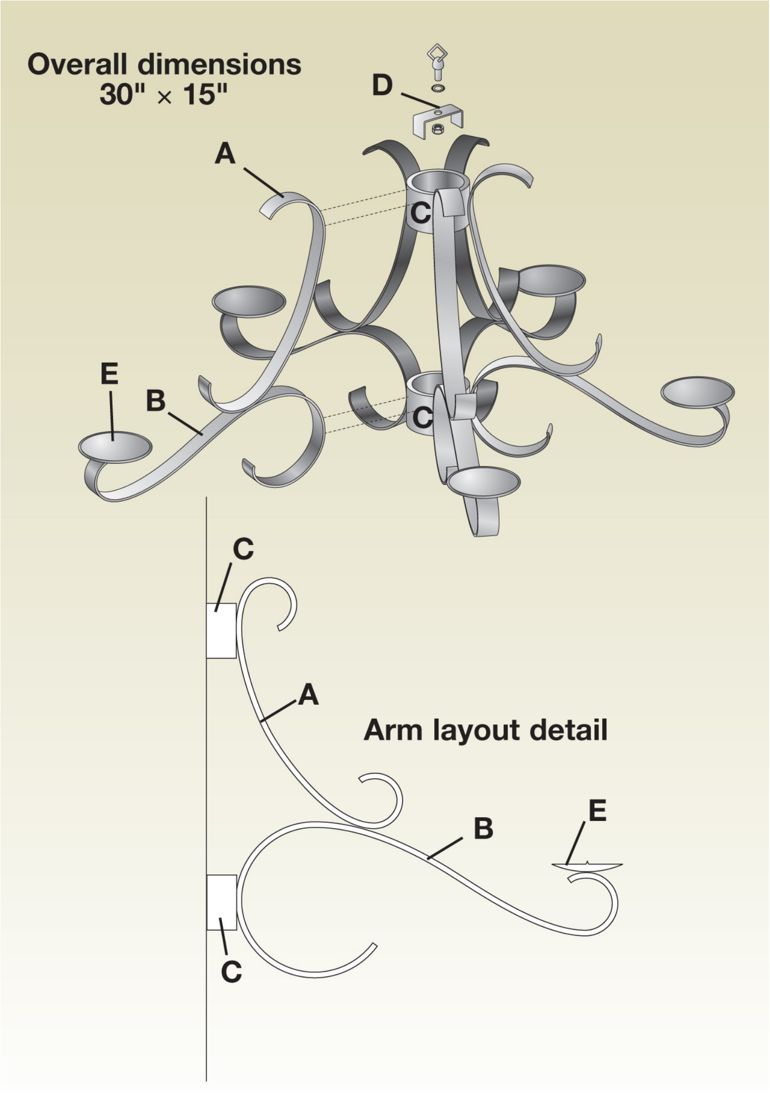
Overall dimensions
30" × 15"
D
A
C
C
E
B
C
A
Arm layout detail
E
B
C

Chandelier

This delightful candle chandelier is a wonderful accent for cabins, porches, or even outside dining. Making the scrolls is easy, but if you want something a bit fancier, you can purchase wrought iron scrolls. Place the scrolls carefully around the couplers to ensure the fixture will hang straight. If you do not want to purchase a long section of round tube for two 2" pieces and can't find any in the scrap bin at the steel yard, use threaded black pipe couplers.

PART	NAME	DIMENSIONS	QUANTITY
A	C scrolls	1/8 × 3/4" flat bar × 16"	5
B	S scrolls	1/8 × 3/4" flat bar × 24"	5
C	Couplers	2-1/4" round tube × 2"	2
D	Insert	1/8 × 3/4" flat bar × 4"	1
E	Bobeches	2-1/2"	5

MATERIALS

- 1/8 × 3/4" flat bar (17 feet)
- 10", 6", 2", and 1" bending forms
- 2-1/4" round tube (4")
- 2-1/2" bobeches (5)
- 2" threaded nipple
- Lock washers (2)
- Threaded brass washers (3)
- Finial
- Decorative chain
- Candles (5)

How to Make a Chandelier

Prepare the C Scroll Blanks

1. Cut the C scroll (A) and S scroll (B) blanks to size.

2. Clamp a C scroll blank to the 10" bending form. Shape the entire length of the scroll blank. Lightly tap the scroll with a hammer, if necessary, to get it to conform.

3. Remove the scroll from the form. Reclamp one end of the scroll to the 1" form and shape that end into a tight curve. Repeat this step to shape the opposite end.

4. Repeat steps 2 and 3 to shape the remaining C scrolls.

Prepare the S Scroll Blanks

1. Mark each S scroll blank 9-1/2" from one end. Clamp the marked end to the 6" bending form.

2. Wrap the bar around the form, shaping it just past the mark.

3. Clamp that same end to the 2" form. Wrap the bar almost all the way around the form.

4. Unwrap the large curve slightly, refining the scroll into a pleasing shape.

5. To curve the other end in the opposite direction, turn the piece over and clamp it to the 2" form. Wrap the bar almost all the way around.

Assemble the Scrolls

1. Set a straightedge on a large piece of paper. Lay out one C scroll and one S scroll along the straightedge, and adjust them until they form a pleasing shape. Mark the contact point on the bars.

2. Remove the scrolls from the paper and weld them together at the contact point.

3. Trace the outline of this assembly on the paper (see photo, top right). Use this pattern to arrange and assemble the remaining arms.

Assemble the Chandelier

1. Cut the couplers (C) and insert (D) to length.

2. Drill a hole in the center of the insert large enough for the threaded nipple. Bend tabs on each end of the insert so it will fit inside a coupler.

3. Place the insert into a coupler, with the tabs pointing down. Weld the insert to the coupler along the tabs. This forms the top coupler.

4. Draw a line around the outside of each coupler on paper.

5. Mark five equidistant points around the coupler outline. Place the coupler back on the paper and transfer the lines onto the coupler, using a combination square (see photo).

6. Weld the arms to the couplers. Weld the bobeches to the arms.

Apply Finishing Touches

1. Place a lock washer and threaded brass washer on the threaded nipple. Set this in the hole in the insert and secure it with another lock washer and threaded brass washer.

2. Add a threaded brass washer and a finial.

3. Wire brush and clean the chandelier. Apply your choice of finish.

4. Clip the chain into the finial, and it's ready to hang.

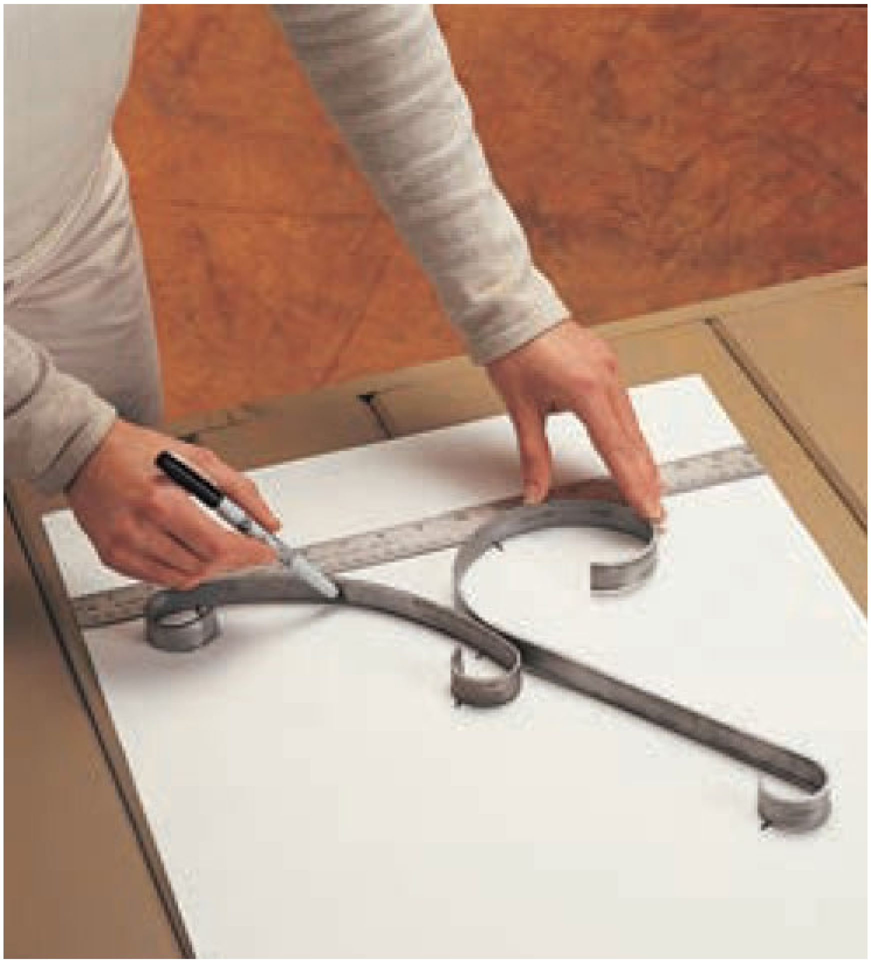

After welding the first set of scrolls, trace the outline on a sheet of paper. Use this pattern to arrange the remaining scrolls.

Use a combination square to mark the arm placement on each coupler.

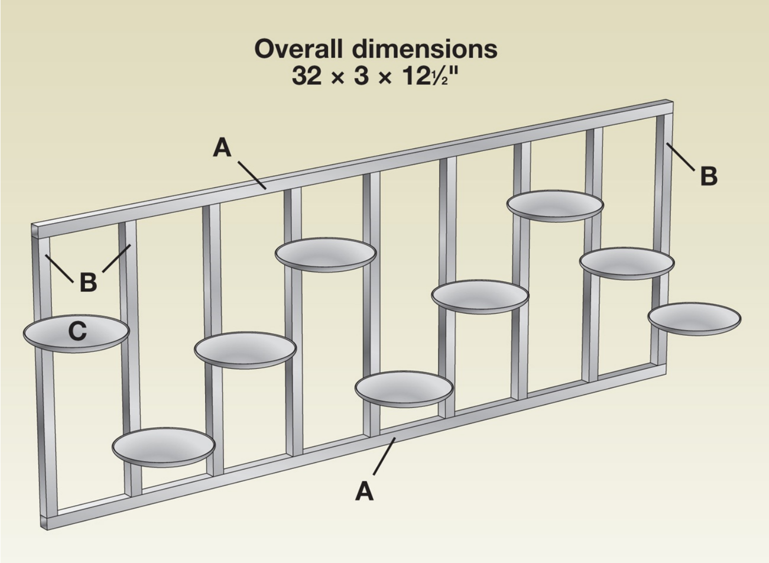
Overall dimensions
32 × 3 × 12½"
A
B
B
C
A

Wall-mounted Candleholder

This wall-mounted candleholder is perfect for a patio wall or over a buffet table. The length and height can be adjusted easily to suit your particular wall space. Or it easily can be turned to run vertically rather than horizontally. Candle drip trays, or bobeches, do not have holes, while bobeches for electrical fixtures do. See Resources, page 140, for bobeche sources. Clean all weld areas prior to welding. When hanging the candleholder, use hangers rated to hold 25 pounds. (The holder will weigh about 13 pounds, plus candles. Remember never to leave burning candles unattended.)

PART	NAME	DIMENSIONS	QUANTITY
A	Horizontal bars	1/2 " square bar × 32"	2
B	Vertical bars	1/2 " square bar × 11-1/2"	9
C	Bobeches	2-1/4"	9

MATERIALS

- 1/2" square bar (14 feet)
- 2-1/4" bobeche (9)
- Candles (9)

How to Build a Wall mounted Candleholder

Assemble the Horizontal & Vertical Bars

1. Cut the horizontal (A) and vertical (B) bars to length.

2. Butt a vertical bar against the inside left edge of the top horizontal bar. Use a framing square to ensure the pieces are set at a 90° angle. Tack weld the inside of the angle.

3. Butt another vertical bar against the inside right end of the bottom horizontal bar. Make sure the angle is 90° and tack weld on the inside.

4. Place the two angled pieces together to make a rectangle, with the vertical bars on the inside of the horizontal bars. Use the framing square to adjust the angles to 90° and clamp in place.

5. Measure diagonally across the corners. If the measurements are the same, the piece is square. If not, adjust the bars so the measurements are equal.

6. Tack the two corners, and check again for square. Weld all four outside corners.

Attach the Vertical Bars

1. Place the remaining vertical bars at 4" intervals between the end pieces. Make sure the spacing is even, then mark where each vertical piece should go.

2. Check each piece for square and weld in place.

Attach the Bobeches

1. Place the bobeches (C) on the vertical bars in a pleasingly staggered pattern.

2. Clean the edge of each bobeche and the areas on the bars where they will be placed to ensure a good weld.

3. Use a magnetic clamp to hold the bobeche at a right angle to the bar and weld from the top side (see photo). When welding, direct most of the heat toward the bar since it is much thicker than the bobeche.

4. Remove the clamp. Make sure the bobeche still meets the vertical bar at a right angle and is level from side to side. Weld the bobeche from the bottom side.

5. Repeat steps 3 to 4 to attach the remaining bobeches.

Apply Finishing Touches

1. Grind down the welds for a smooth look.

2. Finish the piece as desired. A rusted metal finish is appropriately rustic for outdoor use.

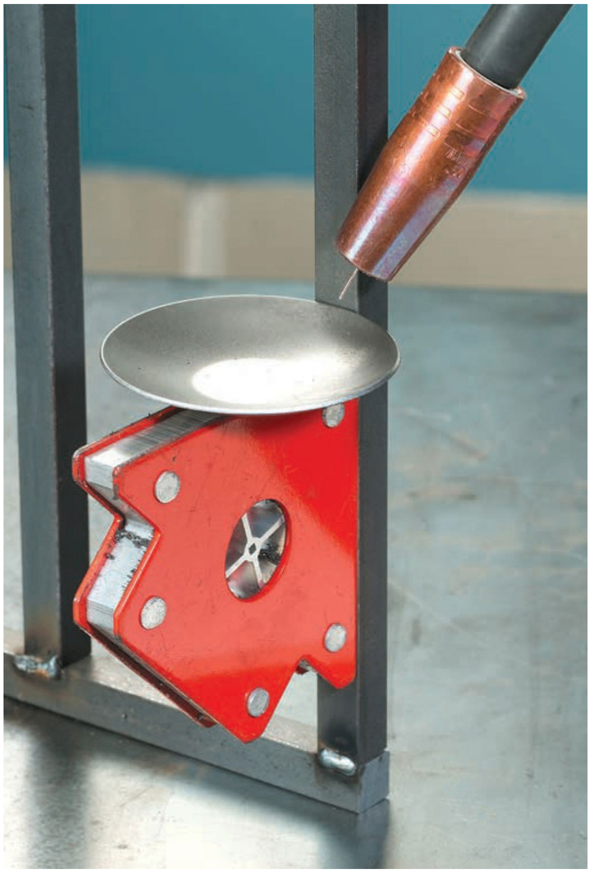

Place the bobeche at a right angle to the vertical bar using a magnetic clamp. Weld the bobeche in place, directing most of the heat toward the bar.

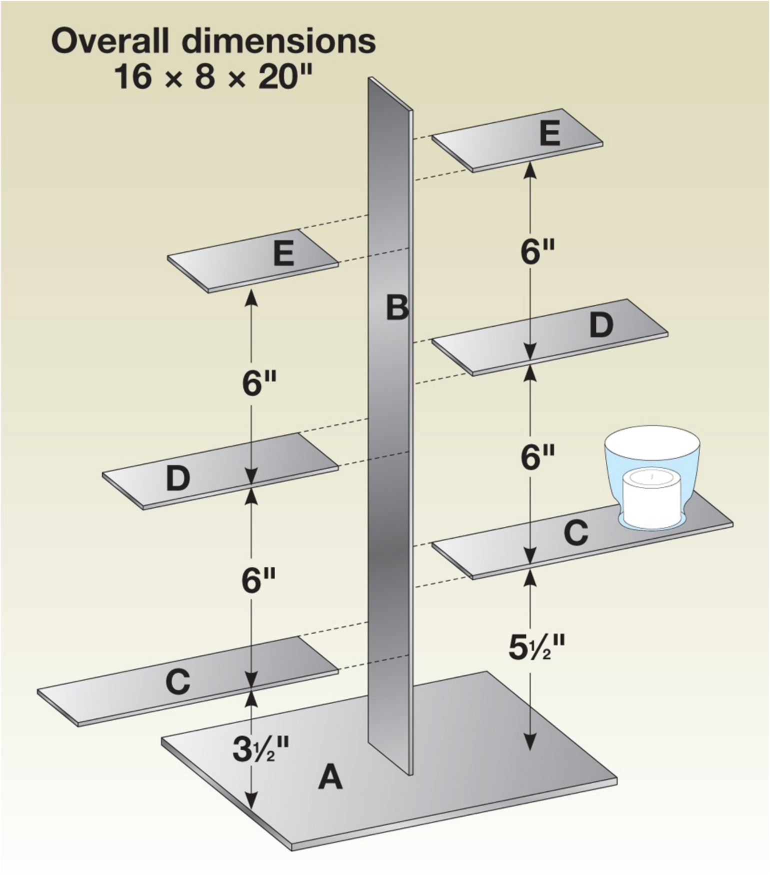
Overall dimensions
16 × 8 × 20"
E
E
B
D
D
C
C
A
6"
6"
6"
6"
3½"
5½"

Tree-shaped Candleholder

This tree-shaped candleholder displays an assortment of votive candles. The tree shape is also handy as a photo display area, if you aren’t interested in a candle-holder. An option is to use square tubing and make a four-sided tree—perhaps even a larger version as a dramatic substitute for a Christmas tree. Most steel supply shops have pre-cut squares of steel sheet and plate in a variety of sizes. A smaller, thicker base would be just as stable. You may opt to make the base round and give the tree trunk a pointed top, if you will be cutting your own pieces. Some steel distributors and welding shops do custom shearing. (Remember not to leave burning candles unattended.)

PART	NAME	DIMENSIONS	QUANTITY
A	Base	3/16" sheet × 8 × 10"	1
B	Trunk	3/16 × 2" flat bar × 20"	1
C	Bottom branches	3/16 × 2" flat bar × 8"	2
D	Middle branches	3/16 × 2" flat bar × 6"	2
E	Top branches	3/16 × 2" flat bar × 4"	2

MATERIALS

- 3/16" sheet (8 × 10")
- 3/16" × 2" flat bar (5 feet)
- Votive holders and candles
- Silicone adhesive

How to Make a Tree-shaped Candleholder

Assemble the Trunk

1. Cut the base (A), trunk (B), and branches (C, D, & E) to size.

2. Place the trunk upright in the center of the base, parallel to the 8" sides of the base. Clamp the trunk at a 90° angle to the base using a magnetic clamp.

3. Tack weld the trunk at each end on both sides.

4. Remove the clamp. Make sure the trunk is still square, then complete the welds.

Attach the Branches

1. Place the bottom branch 3-1/2" up from the base and magnetically clamp it in place at a 90° angle to the trunk.

2. Tack weld the branch to the trunk at each end on the top side (see photo).

3. Remove the clamp and complete the weld.

4. Place the second bottom branch on the opposite side of the trunk, 5-1/2" from the base. Weld in place.

5. Repeat the process by setting the middle branches 9-1/2" and 11-1/2" from the base, and the top branches 15-1/2" and 17-1/2" from the base.

Apply Finishing Touches

1. Clean the metal and the welds with a wire brush.

2. Finish the candleholder as desired.

3. Use silicone adhesive to attach a votive holder to the end of each branch if desired.

A right angle magnetic clamp makes welding the branches an easy task.

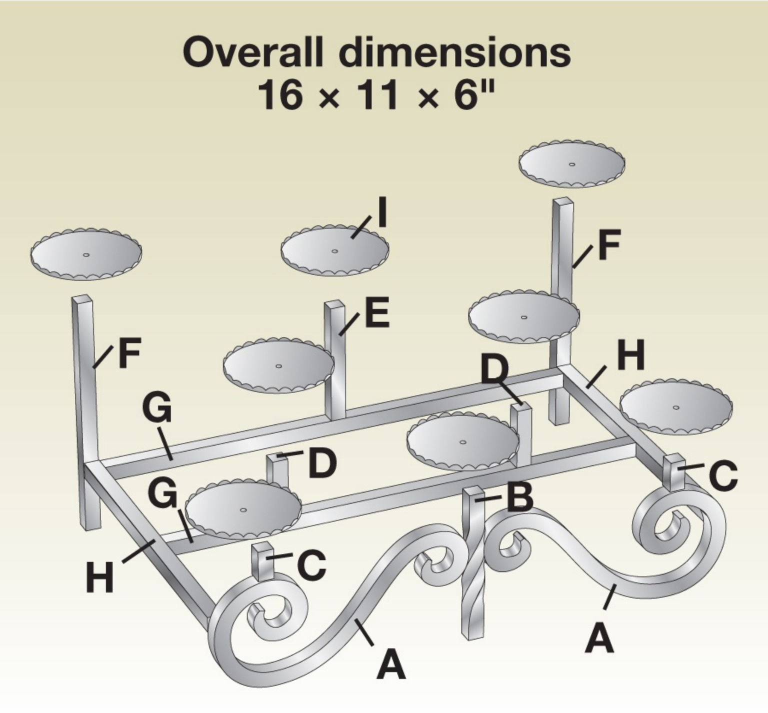
Overall dimensions
16 × 11 × 6"
I
F
E
F
D
H
G
D
C
G
B
H
C
A
A

Fireplace Candleholder

In the summertime, place this candleholder in your fireplace to enjoy flickering flames of firelight without the heat of the fire. This project uses forged scrolls, but you can bend your own, if you wish.

PART	NAME	DIMENSIONS	QUANTITY
A	Scrolls	4 × 8"	2
B	Center post	1/2 " twisted square rod × 5"	1
C	Front extenders	1/2 " square rod × 1"	2
D	Middle extenders	1/2 " square rod × 2"	2
E	Rear extender	1/2 " square rod × 3"	1
F	Rear posts	1/2 " square rod × 6"	2
G	Crosspieces	1/2 " square rod × 16"*	2
H	Side pieces	1/2 " square rod × 10"	2
I	Bobeches	3-3/8"	8

*Approximate dimensions, cut to fit

MATERIALS

- 4 × 8" scrolls (2)
- 1/2" twisted square rod (5")
- 1/2" square rod (7 feet)
- 3-3/8" bobeches (8)
- Candles (8)

How to Build a Fireplace Candleholder

Assemble the Scrolls

1. Cut the center post (B) to size.

2. Place the scrolls (A) on their sides with the small circles butted against the center post. (If the scroll ends have been hammered out to form a flared end, you will need to use shims to make the pieces lie flat.)

3. Clamp the workpieces to your work surface, and weld the scrolls to the center post.

Assemble the Holder Base

1. Measure the length of the two scrolls and center post, and subtract 1". Cut the crosspieces (G) to this length.

2. Cut the rear posts (F) and side pieces (H) to size.

3. Place the crosspieces between the side pieces; one at one end and one in the middle. Weld the crosspieces in place.

4. Mark 2" up from the bottom of the rear posts. (Depending on the size of scroll you are using, you may want to make this higher or lower.) Clamp the rear posts to the work surface, and align the cross and side piece assembly with the 2" marks. Weld into place.

5. Weld this assembly to the back of the scrolls, making sure the side pieces are level.

Attach the Bobeches

1. Cut the front (C), middle (D), and rear (E) extenders to size.

2. Weld the front extenders to the tops of the scrolls. Center the rear extender over the rear crosspiece and weld in place. Set the middle extenders 4" from each end of the middle cross piece and weld in place (see photo).

3. Center a bobeche (I) over each extender, rear post, and center post. Make sure the bobeches are level, and weld in place. You may need to grind down the tops of the posts to get a level seating for the bobeches.

4. Wire brush, clean, and apply the finish of your choice.

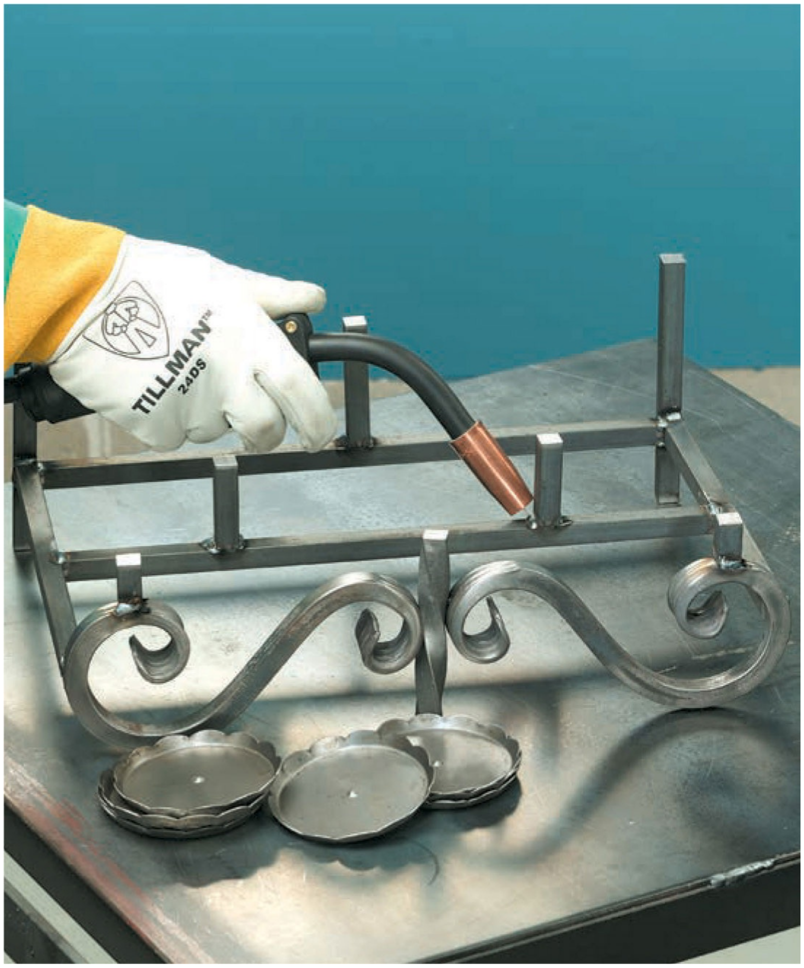

Weld the extenders into place.

FURNITURE PROJECTS

Baker’s Shelves 106
Coffee Table 110
Corner Coatrack 112
Kitchen Appliance Stand 114
Wall Hung Plate Rack 116
Wine Rack 118
Headboard 120

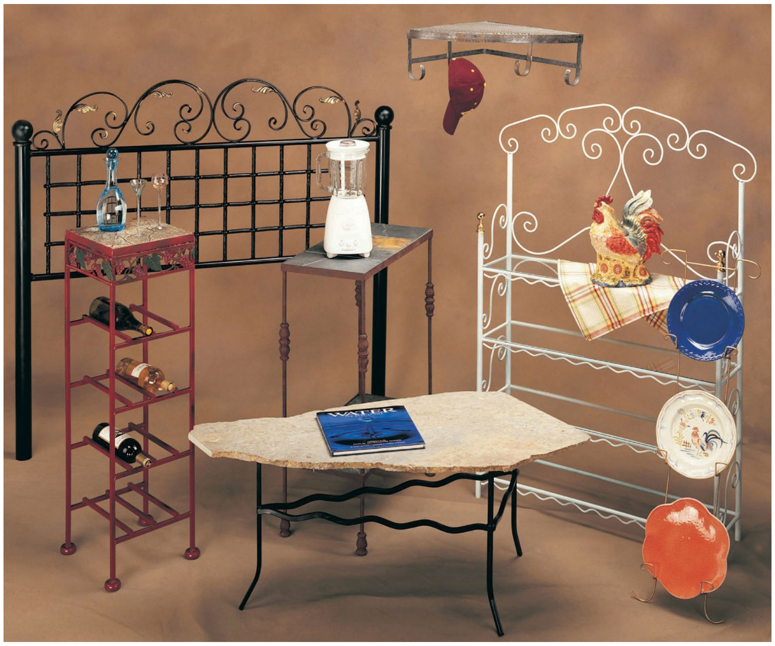

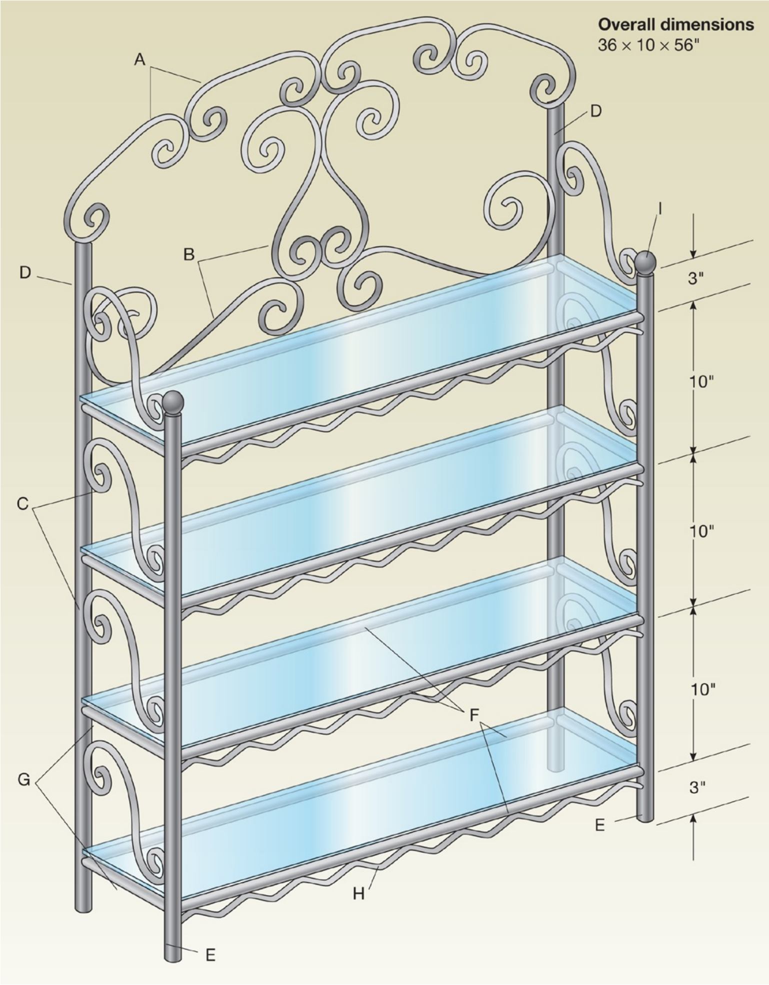
A
B
C
D
D
E
E
F
G
H
I
3"
10"
10"
10"
3"

Overall dimensions
36 × 10 × 56"

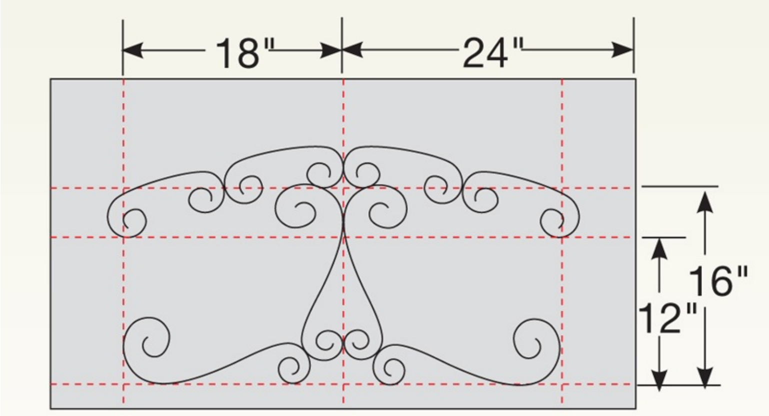

Decorative scroll layout detail

Baker’s Shelves

Baker’s shelves are a great welding project—the possibilities are nearly endless. Many gift and furnishing stores sell various baker’s shelves, but if you look closely, many have sloppy welds and thin wire scrolls. On our project, the shelves feature 1/4" round bar scrolls and a nice zigzag trim. You can make the scrolls with the homemade jigs featured on page 19, which works nicely. We also tried out a scroll making attachment for a hand-powered metal bender. This scroller was meant to use flat bar stock, but by grinding the round bar ends slightly, we could bend round stock. The ground ends actually added a nice touch to the scrolls. The metal bender could be used for the zigzag as well.

PART	NAME	DIMENSIONS	QUANTITY
A	C scrolls	1/4" round bar × 26-1/2"	4
B	Top S scrolls	1/4" round bar × 35"	4
C	Side S scrolls	1/4" round bar × 24-1/2"	8
D	Back legs	3/4" round tube × 44"	2
E	Front legs	3/4" round tube × 36"	2
F	Shelf supports, front & back	1/2 " round tube × 34-1/2"	8
G	Shelf supports, sides	1/2 " round tube × 8-1/2"	8
H	Zig zag trim	1/4" round bar × 48"	4
I	Finials	1" brass or wood ball	2

MATERIALS

- 1/4" round bar (54 feet)
- 16 gauge × 3/4" round tube (13-1/2 feet)
- 16 gauge × 1/2" round tube (29 feet)
- 4 × 6 foot plywood
- 2 × 4 foot plywood
- 1 × 2 lumber × 36" (7)
- Drywall screws
- 2 × 2 lumber × 36"

- Pre-drilled 3/4" on center flat bar (3 feet)
- 5/16 × 2-1/2" machine bolts (24)
- 3/4" rubber pipe tips or glides (4)
- Brass or wood finials (2)
- Tempered glass shelves cut to fit (4)

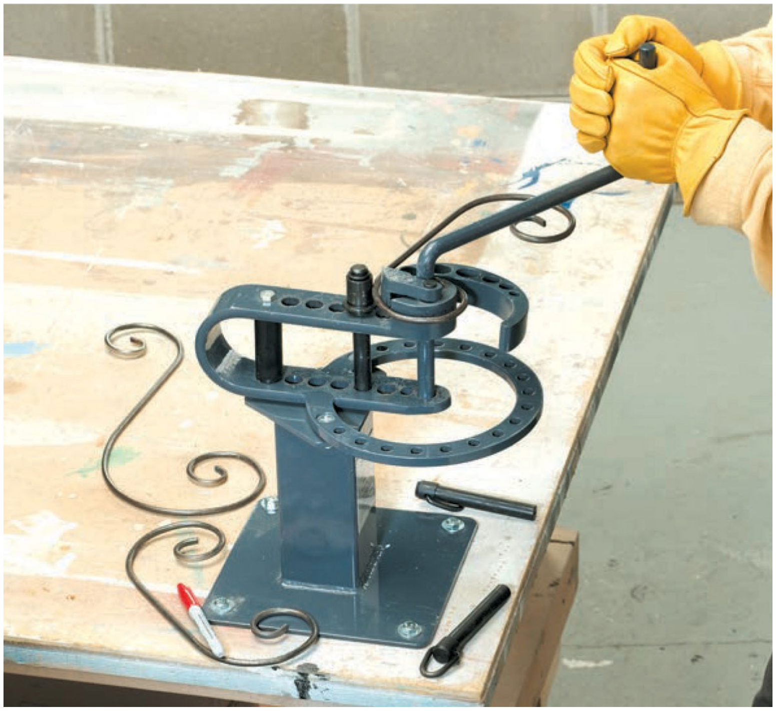

Use a scroll bender to bend the decorative scrolls.

How to Build Baker’s Shelves

Prepare the Scrolls

1. Cut the C scrolls (A) and top S scrolls (B) to size.

2. Mark a stop point 10-3/4" from each end of a C scroll. Insert one end into the scroll maker (flatten the end by grinding, if necessary), and bend it to the stop point. Do the same for the other end (see photo).

3. Repeat step 2 to shape the remaining three C scrolls.

4. Mark a stop point 20" from one end of a top S scroll. Mark a stop point 8" from the other end.

5. Insert the 20" end of the S scroll into the scroll maker and bend to the stop point for the large circle. Insert the opposite end into the scroll maker and bend to the 8" mark for the small circle.

6. Repeat steps 4 to 5 to shape the other S scrolls.

Assemble the Scrolls

1. On a 2 × 4 foot sheet of plywood, lay out the scroll pattern shown below.

2. Draw a line 6" in from one side of the plywood. Draw parallel lines at 18" and 36" from this line to mark the outside edges of the back legs and the center line. Draw a perpendicular line across the bottom of these lines for the base line.

3. On the center line, mark a point 16" up from the baseline to mark the base of the uppermost C scroll pair. On the side lines, mark points 12" up from the baseline to mark the base of the side C scrolls. The back legs will meet the C scrolls at the underside of the scroll. The S scroll will butt up against the back leg. (Because the scrolls you turn will be slightly smaller or larger, you may need to adjust the layout dimensions as you go.)

4. Lay out the scrolls on the pattern. When you are satisfied with the layout of the scrolls, weld them together at the contact points.

Create the Assembly Jig

1. Use drywall screws to fasten a 3-foot piece of 1 × 2 along the long edge of a 4 × 6 piece of plywood. Attach another 3-foot piece of 1 × 2 parallel to the first piece, 36" apart. These mark the outside edges of the legs.

2. Attach a third piece of 1 × 2 across the bottom of these two, making sure the corners are square. This will align the bottom of the legs.

3. Attach four 24" pieces of 1 × 2 perpendicular to the sides at 3", 13", 23" and 33" to mark the shelf heights. Center these pieces between the side pieces.

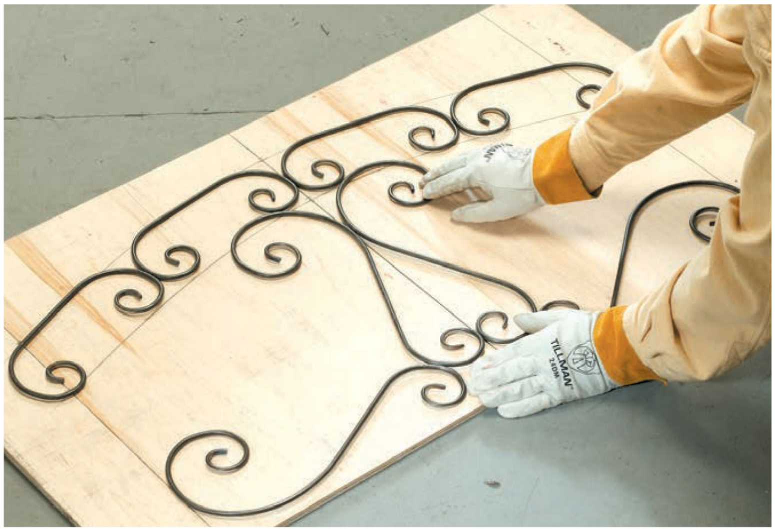

Lay out the scroll design dimensions on a piece of plywood. Place the scrolls into the layout, and weld at each contact point.

Assemble the Legs & Shelf Supports

1. Cut the back legs (D), front legs (E), and shelf supports (F & G) to size. Grind the ends of the shelf supports slightly to fit around the curve of the legs.

2. Clamp the back legs to the side 1 × 2s, making sure the leg ends are firmly against the bottom crosspiece.

3. Clamp the back shelf supports to the four shelf crosspieces of the jig.

4. Weld the back shelf supports to the back legs.

5. Repeat steps 2 to 4 to attach the front legs to the front shelf supports.

Attach the Shelf Sides

1. Stand the back legs and shelf support assembly upright. Use a magnetic clamp to hold a shelf side support even with a back shelf support. Tack weld into place.

2. Stand the front legs and shelf support assembly upright. Align with the back assembly, and tack weld the shelf side support even with its corresponding front shelf support.

3. Continue to line up and attach the shelf side supports.

4. Place the shelf assembly on its back, and align it with the scroll assembly. Weld the two assemblies together at all contact points.

Install the ZigZag Trim

1. Cut the zigzag blanks (H) to length.

2. Bend the zigzag trim to shape. Create a bending jig by attaching a 3-foot length of pre-drilled flat bar to a 3-foot length of 2 × 2. Drill 1" deep holes through every other hole. Use 2-1/2" bolts as the bending posts. Insert two bolts into the jig, and clamp the zigzag blank between them. Bend the metal around the second bolt and insert a third bolt, bend the metal around that, and continue (see photo). As you get toward the end, if you need more leverage, slip a piece of 1/2" tube over the rod.

3. Cut the zigzag trim to fit between the front legs.

4. Weld the zigzag trim to the legs and to points on the underside of each shelf.

Attach the Side S Scroll

1. Cut the side S scrolls (C) to size.

2. Mark a stop point 11-7/8" from one end for the large circle of the scroll, and 6-3/8" from the opposite end for the small circle. Do this for each S scroll.

3. Bend the scrolls at the marks.

4. Place the scrolls between the front and back legs, setting the scrolls 1/2" above the shelf supports to allow room for the glass shelf.

5. Weld the scrolls to the front and back legs.

Add Finishing Touches

1. Grind down all welds and spatter. Clean and wire brush or sandblast the baker's shelves. Apply the finish of your choice.

2. Install glides or rubber tips on the legs. Place finials on the front leg tops.

3. Measure the shelf dimensions, including the cutout dimensions to fit around the legs, and order tempered glass shelves cut to fit.

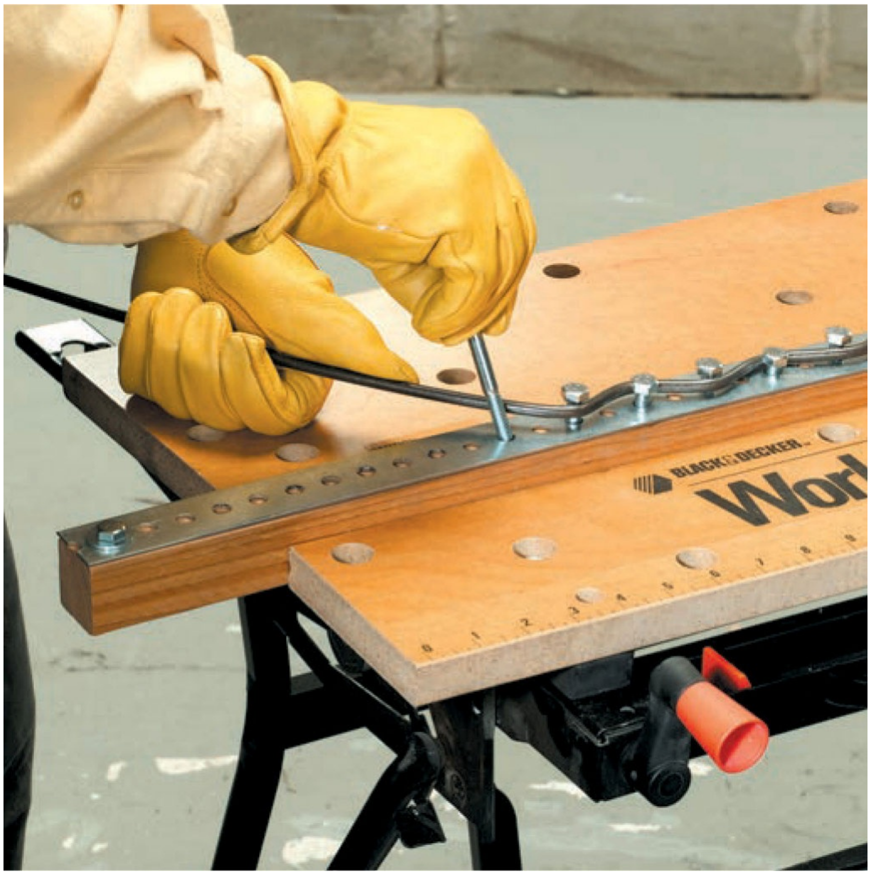

Use pre-drilled 3/4" on center flat bar attached to a 2 × 2 to create a bending jig for the zig zag trim. Clamp the rod between the first two bolts and bend. Continue adding bolts and bending until complete.

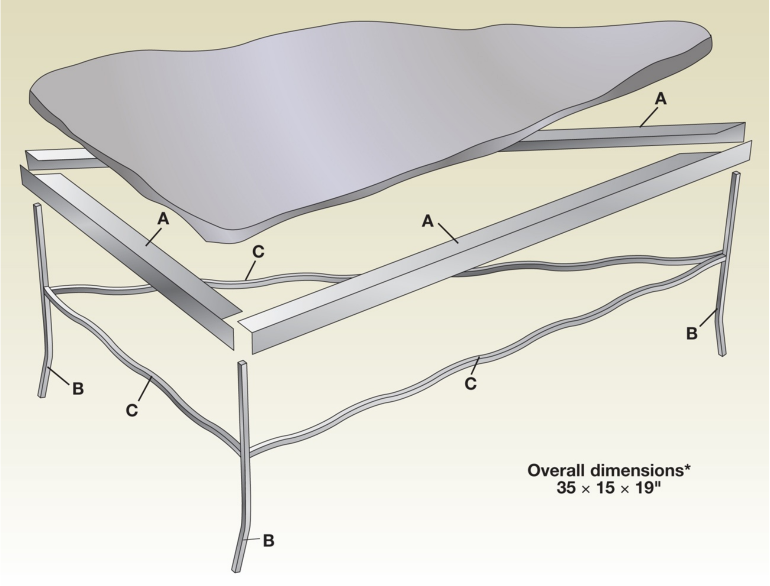
A
A
A
C
A
B
B
C
C
B

Overall dimensions*
35 × 15 × 19"

Coffee Table

A nice slab of stone or a cross section of a tree makes an appealing coffee table, but how do you support it simply and easily? A table base welded from 1" angle iron and 1/2" square bar can support the weight of stone or wood without being cumbersome or bulky. This design is for a triangular base to suit the piece of marble being used. Creating a layout with tape allows you to experiment with different sizes and locations for your triangle. If you are using fence pickets, as we have here, make sure the length of the longest side of the triangle does not exceed the length of the picket. Pickets are generally 35 to 39" long.

PART	**NAME**	**DIMENSIONS**	**QUANTITY**
A	Top supports	1/8 × 1 × 1" angle iron*	3
B	Legs	1/2 " square bar × 20"	3
C	Wavy crosspieces	1/2 " wavy bar fence pickets*	3

*Side pieces and top supports must be cut to fit the particular tabletop.

MATERIALS

- 1/8" × 1 × 1" angle iron (amount will vary)
- 1/2" square bar (5 feet)
- 1/2" wavy bar fence pickets (3)
- Construction adhesive, silicone glue, or screws
- Masking tape
- Table top

How to Build a Coffee Table

Create the Triangle

1. Turn the table top upside down, and use masking tape to lay out a triangle. The triangle should be at least 3" in from the edges of the table top material, but not so far in that the table will tip easily.

2. Measure each side of the triangle. Cut the top support pieces (A) to match these three measurements. Miter both ends of the longest piece.

3. Place the longest support piece on top of the second longest support. Set both pieces on the tape layout on the table top. Mark the angle and notch on the second support where it intersects the first support (see photo) to allow these pieces to butt together. Cut out the notch and angle on the support.

4. Place the two supports back together on the triangle layout. Set the third support on the layout under the first two support pieces. Mark the notches and angles at the intersections. Cut the third support piece. (You may need to make an additional angle cut on the second long piece, depending on the triangle.)

5. On a work surface, arrange the support pieces to form the triangle. Weld the outside corners. Turn the assembly over and weld the top butt joints.

6. Use an angle grinder to smooth the top joints so the table top rests on a flat surface.

Install the Legs

1. Cut the legs (B) to size.

2. Mark each leg 4" from one end. Place a leg in a bench vise, lining up the mark with the edge of the vise jaws, and bend the leg end 15°. Repeat this process for the other two legs.

3. Turn the top assembly upside down on your work surface. Place a leg in a corner with the bend of the leg pointing outward in the same line as the point of the triangle.

4. Check to make sure the leg is perpendicular to the top assembly, then tack weld in place. Tack weld the other two legs in the remaining corners the same way.

Attach the Side Pieces

1. Measure the distance between the legs. Cut one wavy crosspiece (C) to fit each of the three sides. Bevel the ends to fit the angled legs.

2. Clamp a crosspiece against two legs, 7” from the top assembly. Check for square and tack weld. Install the remaining crosspieces the same way.

3. Turn the assembly right side up and check for square and level. Make adjustments if necessary. Complete all the welds.

Complete the Table

1. Grind down the welds, if desired. (You may also want to grind down the feet of the table so they make flat contact with the floor.) Clean or wire brush the assembly. Apply your choice of finish.

2. Drill holes in the angle iron and attach a wooden table top with screws, or use silicone or construction adhesive to attach a stone top.

Place the mitered long support piece on top of the second longest support piece. Mark the angle and notch needed to fit the pieces together.

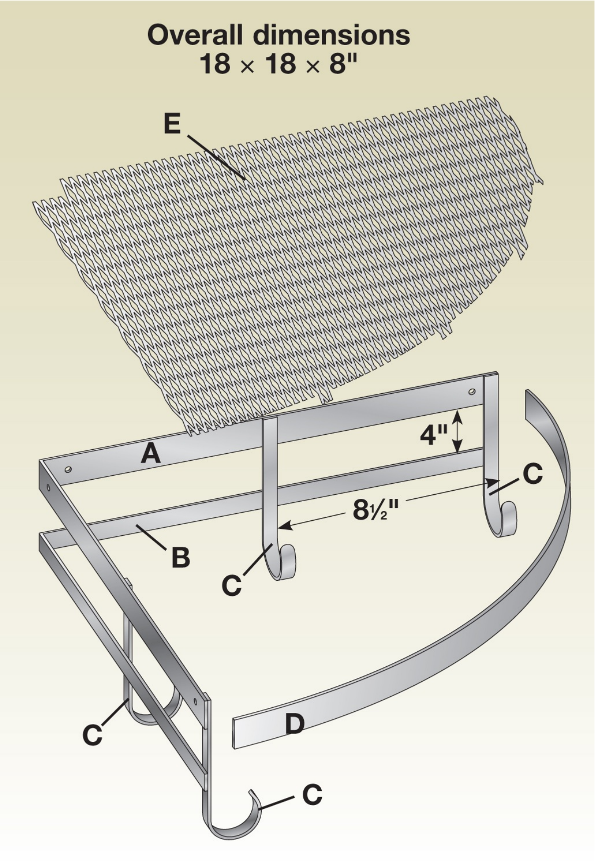
Overall dimensions
18 × 18 × 8"
E
A
4"
C
8½"
B
C
C
D
C

PART	NAME	DIMENSIONS	QUANTITY
A	Rack top	1/8 × 1-1/4" flat bar × 36"	1
B	Rack bottom	1/8 × 3/4" flat bar × 36"	1
C	Hooks	1/8 × 3/4" flat bar × 12"	4
D	Rack front	1/8 × 1-1/4" flat bar × 38"*	1
E	Shelf	Expanded sheet metal	1

*approximate dimension, cut to fit

Corner Coatrack

In a small hallway with limited or no closet space, a corner coat-rack is one of the handiest things you can add. The expanded sheet metal used for the shelf is available in a variety of thicknesses and hole sizes. You may choose to use patterned sheet metal, which is available in many designs and gauges—see the Resources section on page 140. If the shelf will be used for items heavier than hats and mittens, you will need to use thicker gauge sheet metal or add some crosspieces for support. The mounting holes are spaced to account for 16" on-center studs. If you know where you will be hanging your coatrack, locate the studs and space the mounting holes accordingly.

MATERIALS

- 1/8 × 1-1/4" flat bar (7 feet)
- 1/8 × 3/4" flat bar (7 feet)
- Expanded sheet metal (2 × 2 feet)
- 2" pipe
- Hanging hardware

How to Make a Corner Coatrack

Prepare the Rack Top & Bottom

The bend in the rack top and rack bottom takes 1/4" of material. By bending the pieces at the 17-3/4" mark, both legs will be equal length.

1. Cut the rack top (A) and rack bottom (B) to size.

2. Drill 1/4" holes in the top piece at 1", 17", 19", and 35". (These are the rack mounting holes.)

3. Mark the top piece 17-3/4" from one end. Place the piece in a bench vise with the mark at the edge of the vise and the longer end extending upward. Bend the end to 90°.

4. Repeat step 3 with the rack bottom piece.

Assemble the Rack

1. Cut the hooks (C) to size.

2. Grind one end of each hook into a smooth, rounded semi-circle. Make a 2"-radius bend in the rounded ends by clamping them to a section of 2" pipe and bending the hooks.

3. Lay the rack top and rack bottom on one side on your work surface, keeping them 4" apart. Make sure the ends are aligned and the bars are parallel.

4. Place a coat hook over the rack top and bottom, flush with the ends and with the top of the rack. Make sure the corners are square, then clamp the workpieces in place. Tack weld the coat hook to each bar.

5. Place the second hook 8-1/2" from the outside edge of the first hook. Check for square, clamp in place, and tack weld to each bar.

6. Rotate the rack so the other side is flat on the work surface. Repeat steps 4 to 5 to attach the hooks to that side.

7. Make sure that all hooks are still square to the rack. Weld the hooks to each bar, starting on the side opposite the tack.

Attach the Rack Front

1. Bend the rack front (D) into an 18"-radius arc. (The easiest way to do this is to place the bar over a table edge and make a slight bend every 2" along the length of the bar. Keep making bends until the proper radius is formed.)

2. Place one end of the arc against an end of the rack top. Make a mark where the other end of the arc meets the other end of the rack top. Cut the arc to size.

3. Place the arc against the rack top, leaving an open corner for the weld. Tack weld in place.

4. Position the other end of the arc against the opposite end of the rack top. Tack weld in place. Make sure the arc is aligned properly and not twisted. Complete the corner welds.

Attach the Shelf & Finish

1. Place the rack upside down on top of the decorative sheet metal (E), with the 90° corner of the assembly 1" from the corner of the sheet metal. Using a permanent marker, trace the inside of the arc onto the sheet metal. Draw 1/2" tabs at regular intervals in front of the arc (see photo). Trace along the sides of the rack so you will know where to bend the sides.

2. Cut the sheet metal with a plasma cutter, tin snips, or a jig saw with a bi-metal blade. Cut a notch at the 90° corner so the sides can be bent. Bend the front tabs and sides to 90°.

3. With the rack upside down, place the shelf inside the rack so the top is flush with the top of the rack. Clamp it in place and weld the sheet metal to the rack.

4. Clean and finish the rack as desired.

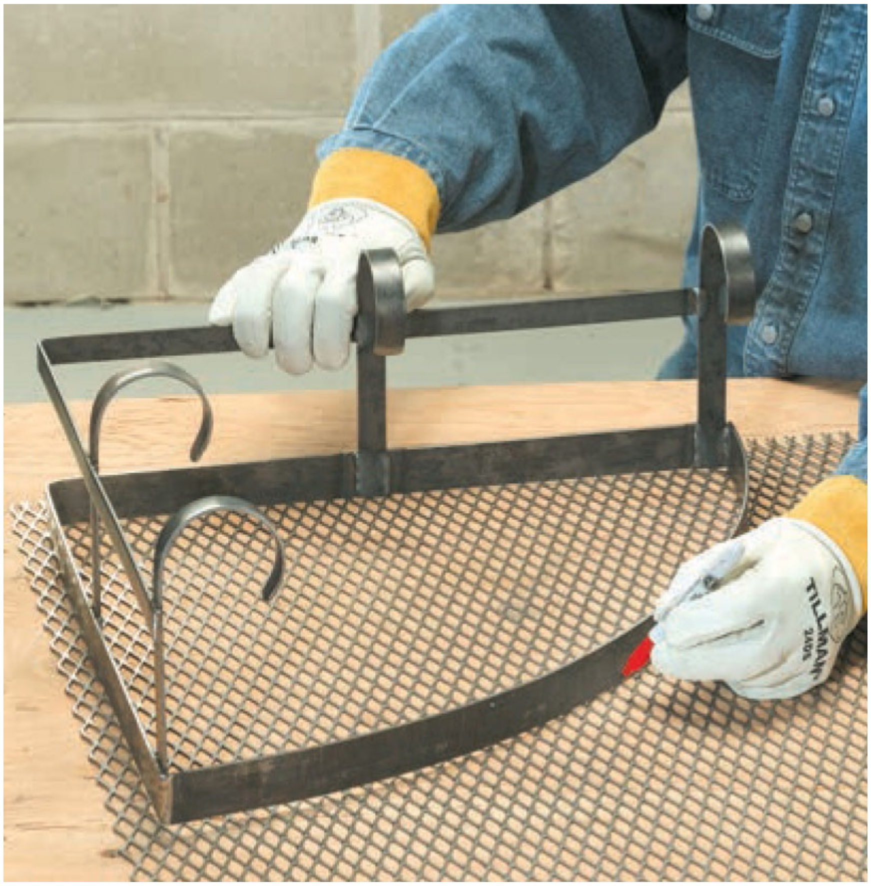

Mark the 1/2" tabs at regular intervals on the outside.

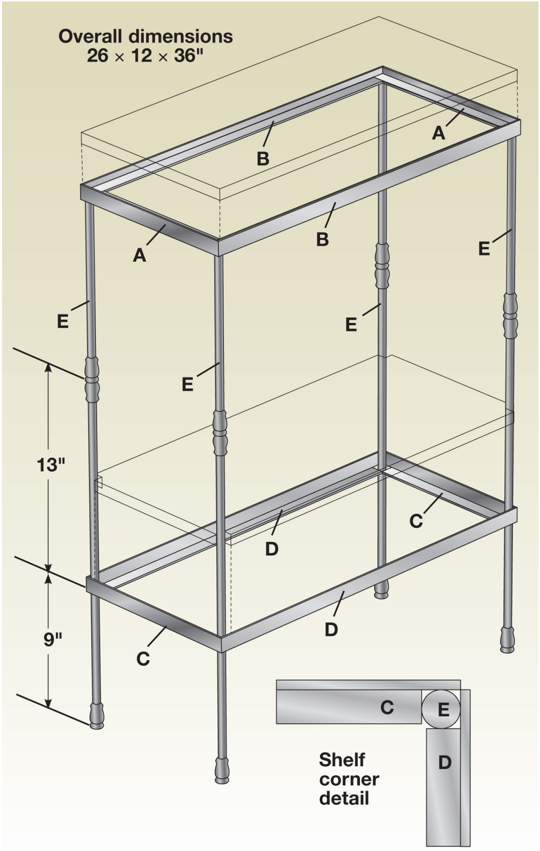
Overall dimensions
26 × 12 × 36"
A
B
A
B
E
E
E
E
13"
9"
C
D
C
D
C
E
D
Shelf corner detail

PART	NAME	DIMENSIONS	QUANTITY
A	Top sides	1/8 × 1 × 1" T-bar × 12"	2
B	Top front and back	1/8 × 1 × 1" T-bar × 26"	2
C	Shelf sides	1/8 × 1 × 1" angle iron × 12"	2
D	Shelf front and back	1/8 × 1 × 1" angle iron × 26"	2
E	Legs	1/2 " round bar × 36"	4

Kitchen Appliance Stand

This freestanding table is sized to be placed next to a stove, but it could go anywhere in the house as an incidental table or plant stand.

The stand is lightweight and easy to move, yet sturdy enough to support a good deal of weight. The decorative bushings and feet provide a touch of elegance and give the stand decorative appeal.

MATERIALS

- 1/8 × 1 × 1" T-bar (6-1/2 feet)
- 1/8 × 1 × 1" angle iron (6-1/2 feet)
- 1/2" round bar (12 feet)
- Decorative bushings (4)
- Decorative feet (4)
- Ceramic tile or wood top
- Ceramic tile or wood shelf
- Felt bumpers (4)

How to Build a Kitchen Appliance Stand

Build the Top

1. Cut the top sides (A) and the top front and back (B) to size. Rather than mitering the corners, cut a 1" notch at each end of the front and back pieces.

2. Position the top front and a side piece at a right angle. (Because the T-bars do not sit flat on the table top, it can be difficult to assemble the top pieces. Use shims or create a wooden clamping jig to hold the pieces securely while you align them.)

3. Check the top front and side piece for square and tack weld the corner.

4. Place the top back and other side piece together at a right angle. Check the corner for square and tack weld.

5. Assemble the two pieces to make a rectangle. Check for square by measuring diagonally between the corners. If the measurements are equal, the assembly is square. If it is not square, adjust until both measurements are equal.

6. Tack weld the top pieces together. Recheck the assembly for square, then finish the corner welds.

Build the Shelf

1. Cut the shelf sides (C) and the shelf front and back (D) to size.

2. Cut a 1/2" notch at both ends of all four shelf pieces to create a hole for the legs to fit through (see Shelf Corner Detail in diagram).

3. Position the shelf front and a shelf side piece at a right angle. Check the pieces for square and tack weld the corner.

4. Repeat step 3 to weld the shelf back to the other shelf side piece.

5. Assemble the two pieces to make a rectangle. Check the assembly for square and tack weld together.

6. Recheck the assembly for square, then finish the corner welds.

Attach the Legs

1. Cut the legs (E) to size.

2. Insert the legs through the cutouts in the shelf. If using decorative slide-on bushings, position them on the legs and tack weld in place.

3. Place the top over the legs. Make sure the legs are flush against the T-bar and tack weld in place.

4. Measure 9" from the bottom of the legs and make a mark. Align the shelf with the marks and tack weld in place.

5. Stand the unit upright. Check the top and shelf for level. If they are not level, break the tack welds and adjust until level. Complete the welds between the legs and the top, and the legs and the shelf.

Attach the Feet & Finish

1. Weld, braze, or braze weld the feet in place (see photo).

2. Wire brush or sandblast the appliance stand and apply the finish of your choice.

3. Attach felt bumpers to the feet.

4. Make shelves of wood, stone, or ceramic tile, and place them on the shelf and the top.

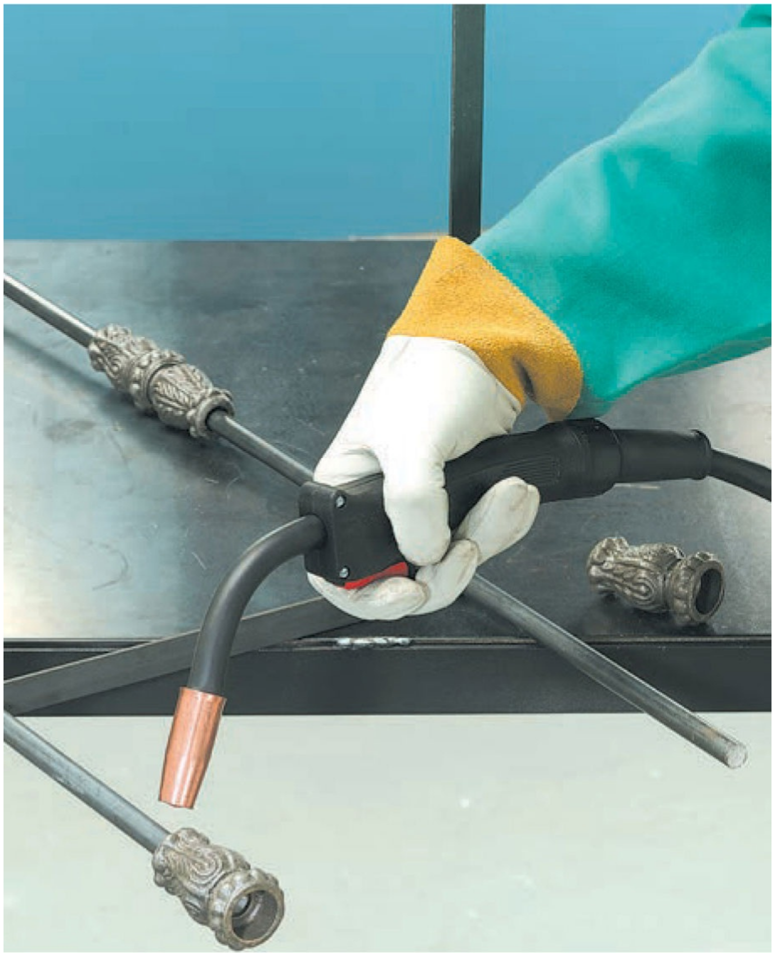

Weld around the joint between the decorative feet and the leg.

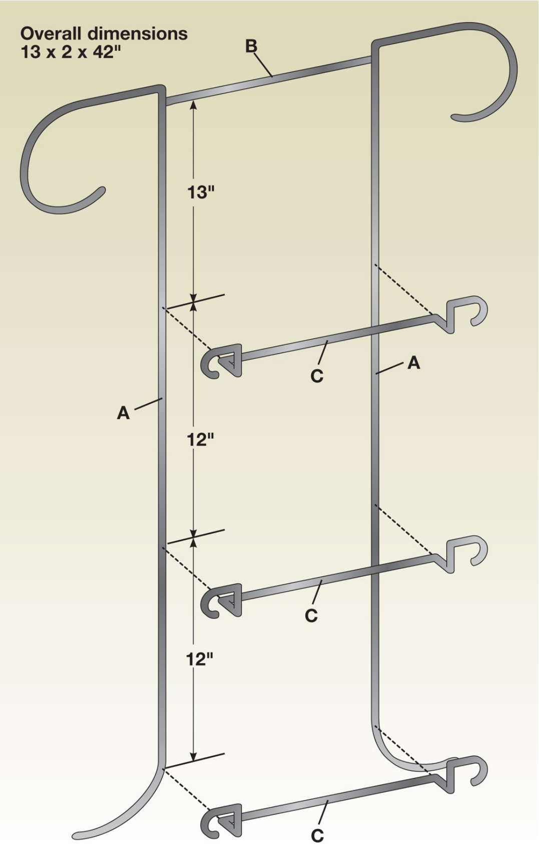
Overall dimensions
13 x 2 x 42"
B
13"
A
C
A
12"
C
12"
C

PART	NAME	DIMENSIONS	QUANTITY
A	Side supports	3/16" round rod × 48"	2
B	Top piece	3/16" round rod × 6-1/4"	1
C	Racks	3/16" round rod × 16-1/2"	3

Wall Hung Plate Rack

This rack offers a nice way to display plates and makes a great gift for a plate collector. You can use any variety of bends and scrolls to add detail to the rack. The only difficult part is the three dimensional aspect of the racks—it can be hard to follow the bend direction. You may want to adjust the spacing or the number of plate racks to fit the plates to be displayed. These directions assume a 10" diameter plate.

MATERIALS

- 3/16" round rod (13 feet)
- 1/2" and 3" pipe for bending jig

How to Make a Plate Rack

Prepare the Side Supports

1. Cut the side supports (A) to size.

2. Using the bending methods on page 19, make a 3" radius half circle at one end of each side support.

3. At the other end of the supports, make a 90° bend 6" from the end. Curl these ends into a 3" radius half circle. Make sure your work piece does not twist.

Prepare the Racks

1. Cut the racks (C) to size.

2. Make marks at 2", 4", and 5" from each end of the racks for bending points.

3. Bend a 1/2" radius curve at the first mark on each end of the racks.

4. Make a 90° bend at each of the other marks using a bench vise and ball peen hammer (see photo). Refer to the diagram on the opposite page for the orientation of the bends.

Assemble the Structure

1. Cut the top piece (B) to size.

2. Make a mark 1" down from the top of the two side supports.

3. Weld the top piece to the side supports at the 1" marks.

4. Place the first plate rack 3" from the bottom of the side supports. Square the rack to the sides and weld in place.

5. Set the remaining racks at 12" and 24" from the bottom rack. Make sure the racks are square to the sides, then weld in place.

6. Turn the rack over and weld the back side of each joint.

Apply Finishing Touches

1. Clean the structure with a wire brush.

2. Apply your choice of finish to the rack.

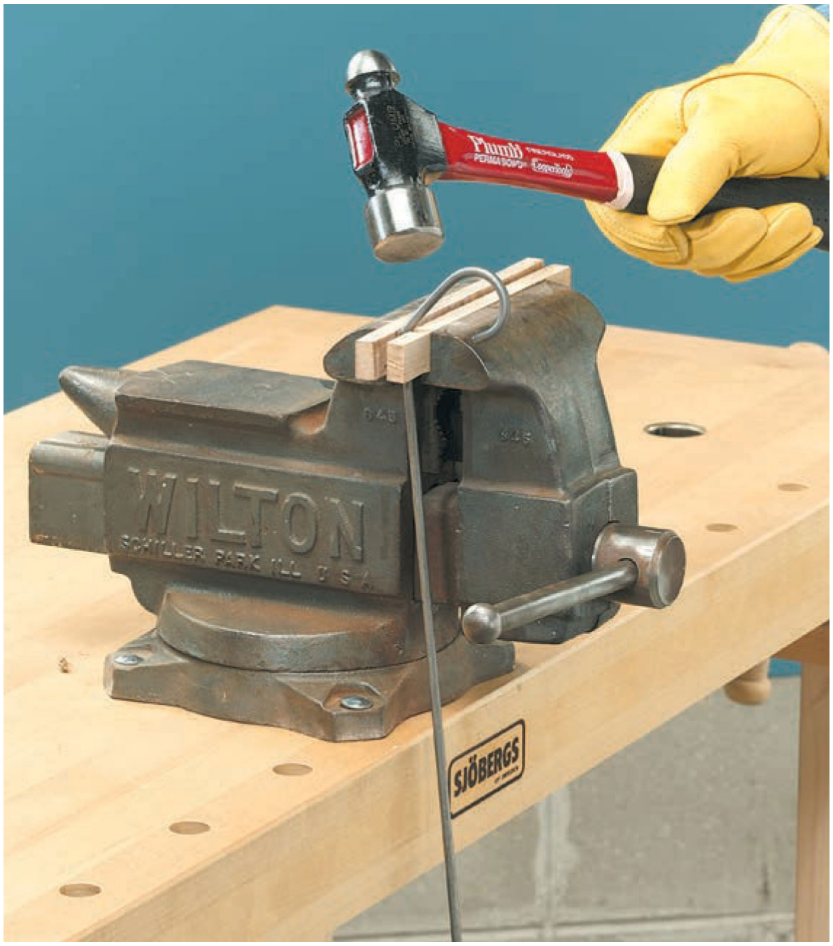
WILTON
SCHILLER PARK ILL U S A
SJÖBERGS
Plumb
PERMA BOND
Cooper

Make the right angle bends using a bench vise and a ball peen hammer.

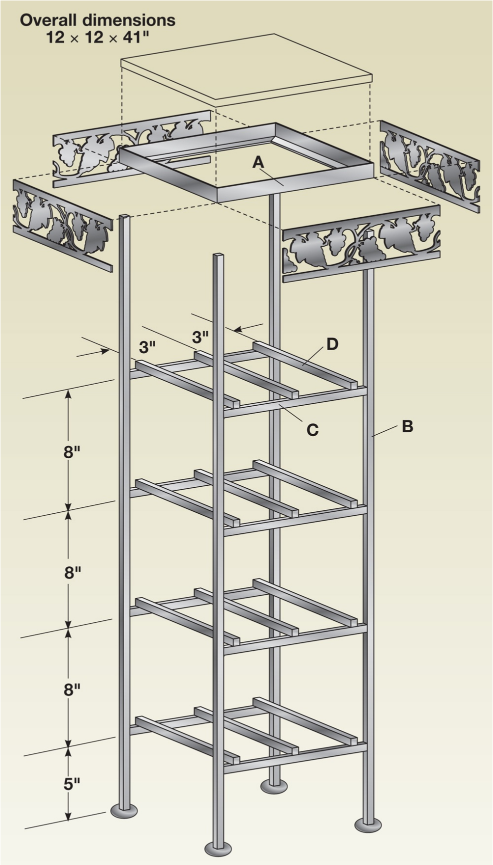
Overall dimensions
12 × 12 × 41"
A
3"
3"
D
C
B
8"
8"
8"
5"

PART	NAME	DIMENSIONS	QUANTITY
A	Top	1/8 × 1 x 1" angle iron × 12"*	4
B	Legs	1/2 " square bar × 40"	4
C	Rack supports	1/2 " square bar × 11" *	8
D	Bottle holders	1/2 " square bar × 12" *	12

*Cut to fit selected tile top

Wine Rack

This wine rack is designed so you can install a wall or floor tile as the top surface. Measure the tile and cut the top pieces to fit. Butcher block or marble would also make a nice looking top. The decorative frieze stamping and squashed ball are available from Architectural Iron Designs—see Resources on page 140.

MATERIALS

- 1/8 × 1 × 1" angle iron (4-1/2 feet)
- 1/2" square bar (33 feet)
- Decorative frieze (4-1/2 feet)
- 2" Squashed balls (4)
- 12" floor or wall tile (1)

How to Make a Wine Rack

Assemble the Top

1. Cut the top pieces (A) to size to fit your top surface, mitering the corners at 45°.

2. Clamp two top pieces together with a corner clamp. Use a carpenter’s square to check the pieces for square, and tack weld.

3. Repeat step 2 to assemble the other two top pieces.

4. Clamp the two L-shaped top pieces together, and check for square by measuring across both diagonals. If the measurements are equal, the assembly is square. If it is not square, adjust until both measurements are equal.

5. Tack weld the two corners. Unclamp the pieces and recheck for square. Finish welding the corner joints.

Assemble the Stand

1. Cut the legs (B), rack supports (C), and bottle holders (D) to size.

2. Mark the legs at 5", 13", 21", and 29" from one end. Mark the rack supports at 2-1/2", 5-1/2" and 8-1/2" from one end.

3. Position four rack supports between two of the legs at the marks. Check the pieces for square and tack weld in place.

4. Repeat step 3 to tack weld the remaining four rack supports to the other two legs.

5. Stand the leg assemblies upright and clamp a bottle holder centered on the 5-1/2" mark on the bottom rack support. Clamp the other end of the bottle holder to the other leg assembly. Check for square and tack weld the bottle holder in place.

6. Repeat step 5 to attach a bottle holder at the 5-1/2" marks on the other three rack supports (see photo).

7. Make sure the legs assemblies and bottle holders are square, then add the remaining bottle holders at the 2-1/2" and 8-1/2" marks. Finish all welds.

Attach the Top, Legs, & Frieze

1. Place the top onto the legs and weld in place.

2. Center a squashed ball at the bottom end of each leg and weld in place.

3. Bend or cut the decorative frieze to fit around the top of the legs and top. Braze, braze weld, or tack weld the frieze into place at the top of the legs and slightly below the top.

4. Grind down the welds as needed. Wire brush or sand-blast the wine rack, then apply the finish of your choice.

5. Place the tile into the top.

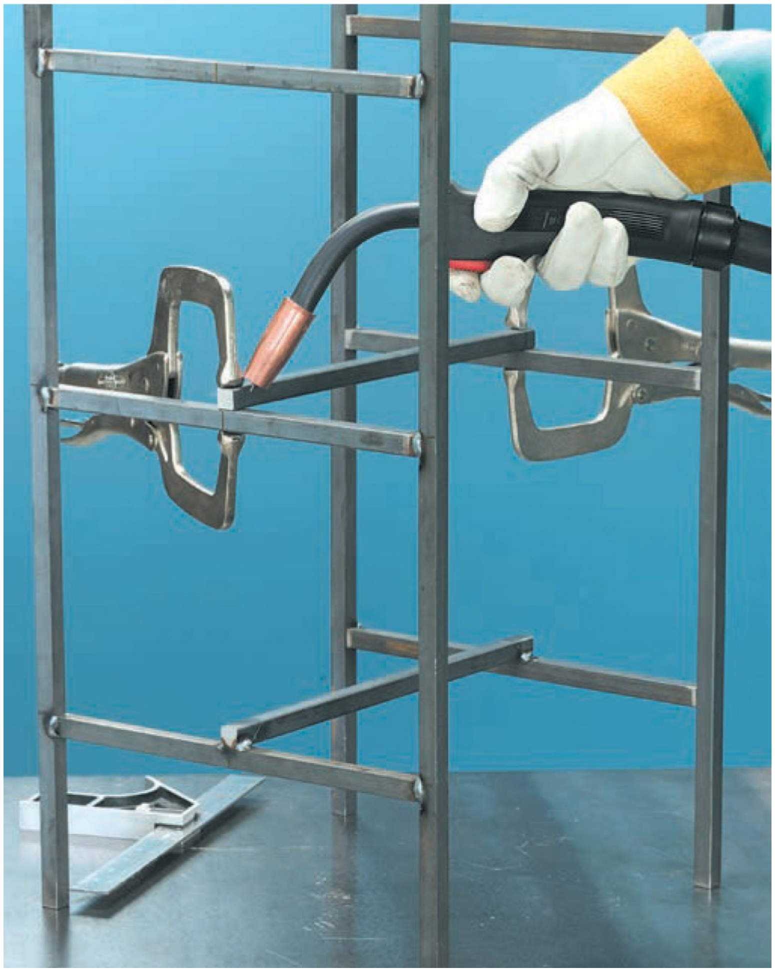

Clamp the center bottle holders to the leg assemblies and tack weld in place.

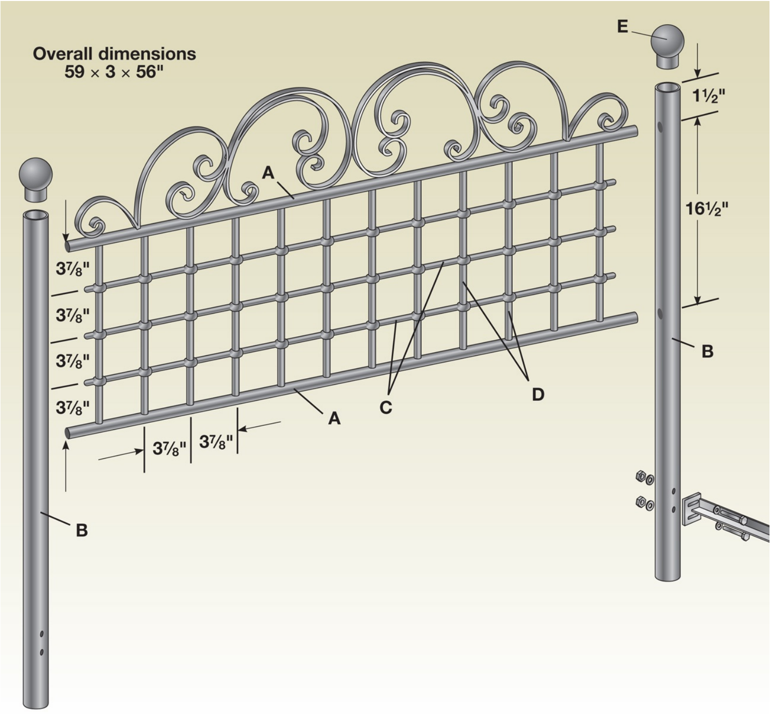
Overall dimensions
59 × 3 × 56"
E
1½"
16½"
A
B
3⅞"
3⅞"
3⅞"
3⅞"
A
C
D
3⅞"
3⅞"
B

Headboard

A custom headboard for your bed is a great welding project. Here’s an example using decorative fence and gate materials. We used top scrolls numbered 3-1/3 and 3-1/2 from Architectural Iron Designs, and hot pierced bar number SF 422US and hammered bar SF 5995 from Triple-S Steel Supply. (See Resources on page 140.)

PART	NAME	DIMENSIONS	QUANTITY
A	Top & bottom bars	1/8 × 1" round tube × 57"	2
B	End posts	1/8 × 2-1/4" round tube × 48"	2
C	Horizontal inserts	5/8 × 3-7/8" on center decorative pierced bar × 51-1/2"	3
D	Vertical inserts	5/8" hammered bar style rod × 18"	12
E	Finials/caps		2

MATERIALS

- 1/8 × 1" round tube (9-1/2 feet)
- 1/8 × 2-1/4" round tube (8 feet)
- 5/8 × 3-7/8" pierced bar (13 feet)
- 5/8" hammered rod (18 feet)
- Decorative scrolls (2)
- Ball post finals/caps (2)
- 3/8 × 2-3/4" machine bolts with washer and nuts (4)

How to Build a Headboard

Assemble the Grid

1. Cut the top and bottom bars (A) and end posts (B) to size.

2. Cut the horizontal inserts (C) to size, cutting at the outside edge of the pierced circle at each end. Cut the vertical inserts (D) to size.

3. Insert the vertical inserts through the holes in the horizontal inserts. Adjust the inserts so they are square.

4. Mark the top and bottom bars 3-1/2" from each end. Align the first and last vertical inserts with the marks. Tack weld the end inserts to the bars.

5. Check the assembly for square by measuring the diagonals. If the diagonal measurements are equal, the assembly is square. Weld each vertical insert to the top and bottom bars.

6. Adjust the horizontal inserts so they are spaced evenly. Weld them to the vertical inserts at each end and at one or two points in the middle.

Insert the Grid into the End Posts

1. Place an end post in a piece of channel or angle iron and mark a straight line down the side of the post. Mark a point 1-1/2" down from the top of the post and centered on this line.

2. Cut a 1" diameter hole centered at the mark, using a hole saw with a bi-metal blade or a plasma or flame cutter.

3. Align the top bar with the hole in the end post. Mark the placement of the hole for the bottom bar, also centered along the same line on the end post. Cut the hole in the post.

4. Repeat steps 1 to 3 to cut holes in the second end post.

5. Insert both ends of the grid assembly into the holes in the end posts (see photo). The distance between the posts should be 55". Make sure the cross bars are square to the end posts, then weld in place.

Attach Decorative Scrolls

1. Place the decorative scrolls on top of the top bar, centering them between the end posts. Weld the scrolls in place.

2. Set the post caps on top of the end posts and weld in place.

3. Wire brush or sandblast the headboard clean. Finish with your choice of finish.

4. Mark the location of the bed-frame bolt holes onto the end posts. Drill 3/8" holes through the end posts. Attach the bedframe to the posts with 3/8 × 2-3/4" machine bolts.

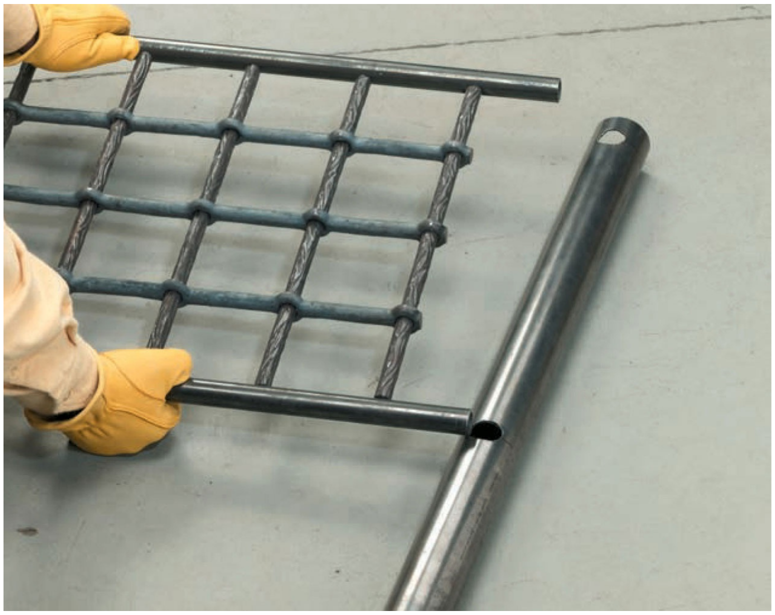

Cut holes in the end posts. Insert the grid assembly into the holes and weld in place.

OUTDOOR PROJECTS

Stair Railing 124
Gate 128
Critter Boot Brush 130
Disk Fountain 132
Arbor 134

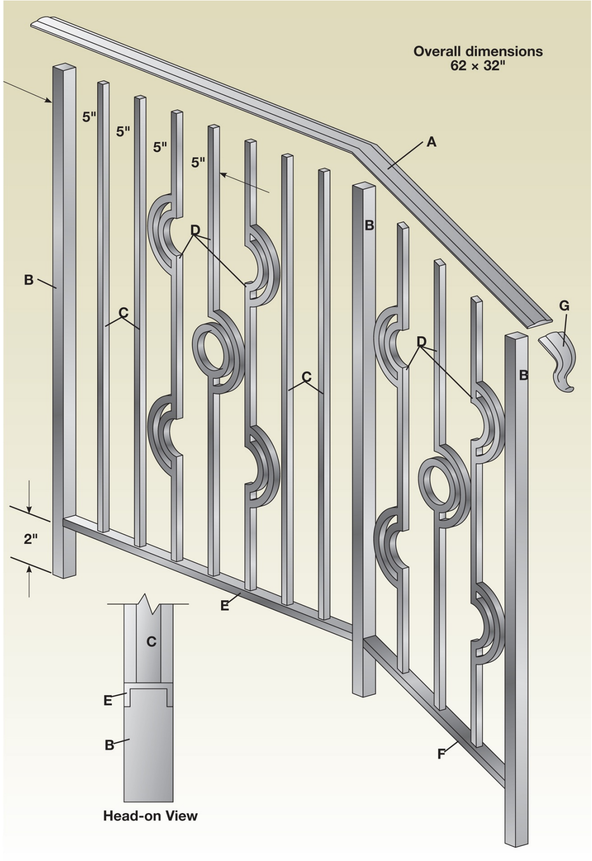
Overall dimensions
62 × 32"
A
5"
5"
5"
5"
B
D
B
C
C
G
D
B
2"
E
C
E
B
F
Head-on View

Stair Railing

With the number of companies selling decorative metal pickets and newel posts, it is easy to create a railing distinctly your own. The railing we are making is for a two step concrete stairway.

Some important information about creating railings:

The rail top should not be interrupted by knobs or other decorations—it must allow continuous hand to rail contact and should be between 1-1/2" and 2" in diameter to ensure easy gripping.

Some railings—like one that is mounted against a wall or in an area where the stair surface is not raised above the surrounding surface, like the stairs in the slope of your lawn—can be a single rail at the appropriate height. A railing that separates the stair, landing, balcony, or deck from a vertical drop has to conform to certain safety standards. The pickets or balusters must not have any gaps larger than 5-1/2" to prevent children from poking their heads through and getting stuck. The lower rail should not be more than 2 inches from the floor surface. If the vertical drop from the floor line is more than 6 feet, the railing must be a minimum of 34" tall. If the vertical drop is less than 6 feet, the railing must be a minimum of 32" in height. Railings must have turn-outs or roundovers at their ends to prevent blunt projections. The safest handrails continue 12" beyond the top and bottom stairs. Railings must be able to withstand a 250 pound force in any direction without giving way.

You can purchase rail cap, also called cap rail or handrail, at most steel supply centers. Handrail terminations come in a variety of styles. Those that scroll in an S shape are called “lamb tongues.” A volute is a spiral termination, and a lateral is a flat curl to the side. These may be ordered through specialty railing distributors. (See Resources page 140.)

MATERIALS

- 1-3/4" rail cap (as needed)
- 1" square tube (as needed)
- 1/2" square rod (as needed)
- 1/2" decorative pickets (as needed)
- 1/8 × 1/2 × 1" channel (as needed)
- Lamb tongue rail termination
- Mounting hardware

PART	NAME	DIMENSIONS	QUANTITY
A	Handrail	1-3/4" rail cap × 64"*	1
B	Newel posts	1/8 × 1 × 1" square tube × 36"	3*
C	Plain pickets	1/2 " square rod × 30"	4*
D	Decorative pickets	1/2 " decorative pickets × 30"	6*
E	Flat bottom rail	1/8 × 1/2 × 1" channel × 40"*	1
F	Angled bottom rail	1/8 × 1/2 × 1" channel × 24"*	1
G	Rail termination		1

*Dimensions and quantities must be adjusted to fit the particular stairs.

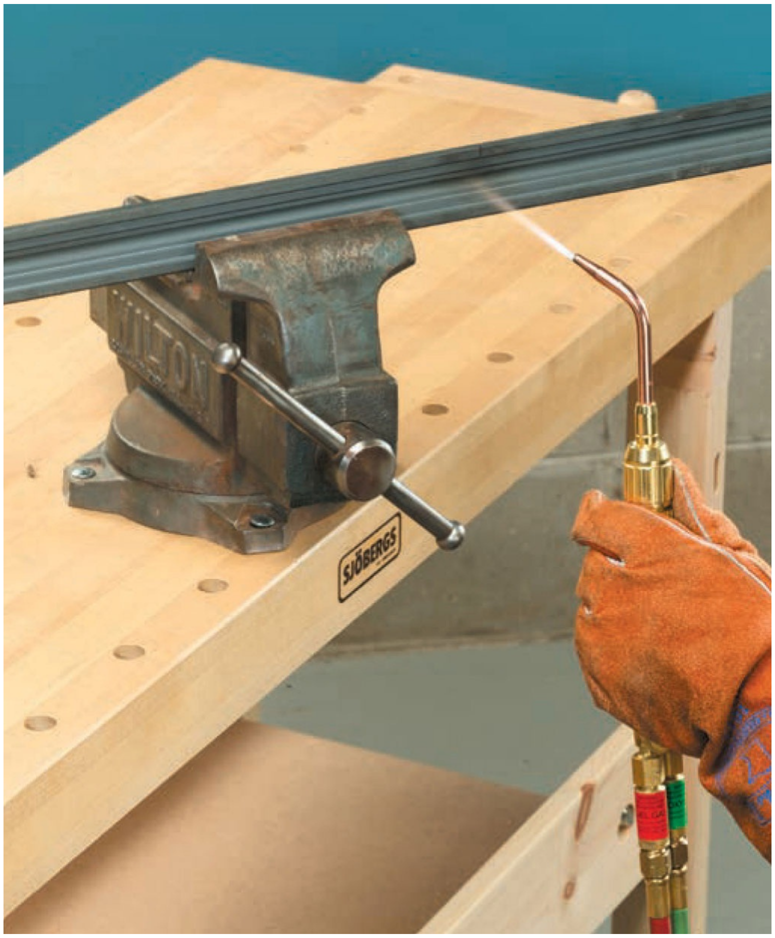

Heat the bending point of the handrail to red hot with an acetylene torch or cutting tip to make the metal easy to bend.

How to Build Stair Railing

Cut the Handrail to Length

1. Measure from the edge of the house to the edge of the landing. Measure from the edge of the landing to where the newel post will be located. Add these two measurements to get the length of the handrail.

2. Cut the handrail (A) to size.

Shape the Handrail

It is a good idea to create an angle guide for bending the handrail by screwing two pieces of wood together at the length and angle to match the stairway.

1. Set the handrail on the landing with one end butting against the house. Mark the handrail at the edge of the landing.

2. Clamp the railing in a bench vise and heat the bending point red hot with an acetylene torch (see photo). A cutting torch preheat works best, just make sure you don’t hit the oxygen and accidentally cut the metal.

3. When the metal is red hot, bend it to create the angle. It is helpful if you pull on the longer end of the rail to bend so you have more leverage.

4. When finished bending the handrail, place it on the stairs to make sure the bend is correct.

Lay Out the Newel Posts & Pickets

The newel posts will be anchored in the concrete or attached using square footings that will bolt into the concrete, so they have to be far enough from the edge not to destroy the edge of the concrete.

1. Determine the number of plain pickets (C), decorative pickets (D), and newel posts (B) you need.

2. Place the bent handrail on the floor or a large work surface. Place newel posts at each end of the handrail and near the bend.

3. Lay out the pickets in a pleasing pattern, making sure they are no more than 5-1/2" apart (see photo).

4. Mark and cut the two bottom rails (E & F) to fit between the newel posts, once you have the layout determined. Place the channel flat side up, with the legs down. Mark the picket locations on the railings.

OPTION: When you look at railings, you will see that some railings have the pickets and newel posts welded directly to the rail cap, as you see here. Other railings have a piece of channel welded into the underside of the rail cap and the pickets are welded to the flat side of the channel. Punched channel can be purchased with 1/2 × 1/2" square holes pierced through it. This is welded under the rail cap, and the pickets are inserted through the punched holes and welded in place. You also can use the punched channel for the bottom rail. Using punched channel means you cannot adjust spacing to account for the unique shapes and sizes of decorative pickets.

Cut the Posts & Pickets to Length

Purchased decorative pickets range from 36" to 39". When cutting decorative pickets, cut equal amounts from each end unless you wish the pattern to be off center.

1. Determine the height of your handrail and the depth that the newel posts will be footed in the concrete (if they are to be footed, otherwise measure to the top of the concrete).

2. Cut the newel posts and pickets to length. Cut the appropriate angle for the stair pickets.

Assemble the Railing

Concrete can explode when heated, so it is best to do your welds on a sheet of plywood that can be doused with water when you have finished.

1. Tack weld the newel posts to the rail cap. Tack weld the bottom rails to the newel posts.

2. Place the rail assembly on the stairs to make sure the dimensions are correct. If they are not, break or grind off the tack welds and make adjustments.

3. Return the assembly to the work surface, and tack weld the pickets in place, maintaining the proper spacing. Use a combination square to check each piece for square before welding.

4. Make the final welds. Weld the lamb tongue termination to the end of the rail cap.

5. Grind down any rough or unsightly welds. Wire brush or sandblast the rail assembly.

6. Install the railing by cementing the newel posts into the stairs or using bolt-down flange shoes.

7. Prime and paint the railing with a high quality outdoor metal paint.

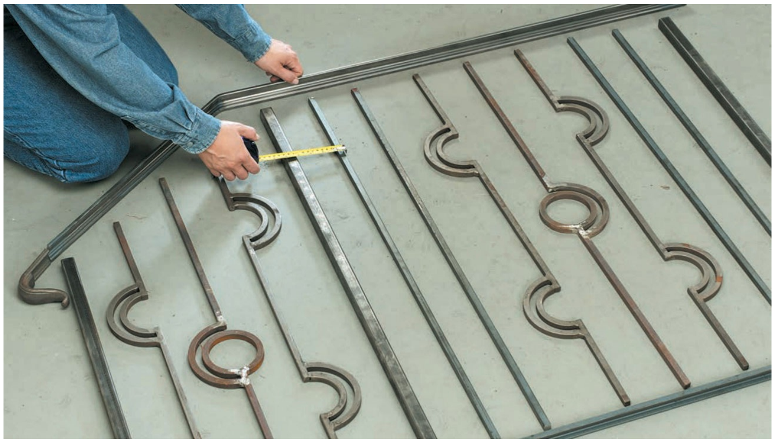

Lay out the pickets and newel posts in a pleasing arrangement, keeping them no more than 5 1/2" apart.

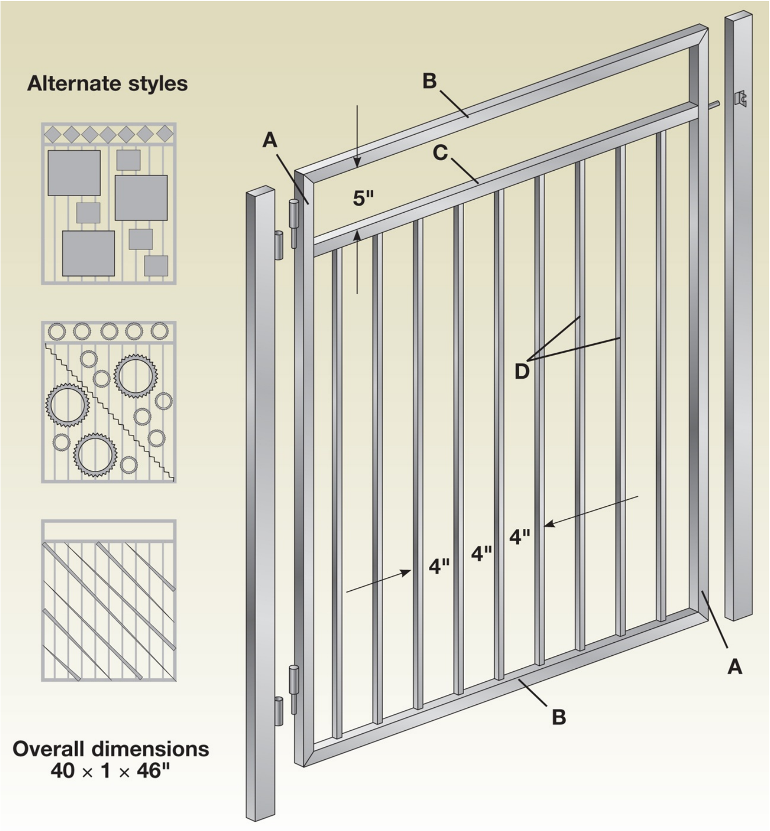
Alternate styles
B
A
C
5"
D
4"
4"
4"
A
B
Overall dimensions
40 × 1 × 46"

Gate

This gate framework can be used to hold a collection of found metal objects. Use brazing or braze welding to join non-matching metals or thicknesses of metals. Rusty objects will need to be cleaned at the point of contact. If the objects you find are large, you may want to space the uprights at 8", rather than 4". The 4" spacing meets code requirements for gates and fences—this spacing prevents children from getting their heads stuck between uprights.

PART	NAME	DIMENSIONS	QUANTITY
A	Sides	16 gauge 1 × 1" square tube × 46"	2
B	Crosspieces	16 gauge 1 × 1" square tube × 40"	2
C	Interior crosspiece	16 gauge 1 × 1" square tube × 38"	1
D	Vertical inserts	16 gauge 1/2 × 1/2" square tube × 40"	9

MATERIALS

- 16 gauge 1 × 1" square tube (17-1/2 feet)
- 16 gauge 1/2 × 1/2" square tube (30 feet)
- 4" weldable barrel hinges (2)
- Gate latch hardware
- Found objects
- Wood and brass spacers

How to Make a Gate

Assemble the Frame

1. Cut the sides and crosspieces (A & B) to size, mitering the ends at 45°.

2. Place the left side piece and bottom crosspiece together at a 90° angle to form one corner of the rectangle. Check the pieces for square, and clamp in place. Tack weld the pieces together.

3. Place the right side piece and top crosspiece together to form another corner of the rectangle. Check for square, clamp in place, and tack weld together.

4. Join the two pieces to form the rectangle. Check the corners for square and clamp in place. Tack weld the corners together.

5. Measure the diagonals of the rectangle to check for square. If the measurements of both diagonals are equal, the assembly is square. If it is not square, pull or push it into alignment. When aligned, clamp it in place, and finish the corner welds.

6. Turn the assembly over, and complete the welds.

Attach the Interior Crosspiece & Inserts

1. Cut the interior crosspiece (C) and vertical inserts (D) to size.

2. Place the interior crosspiece against the inside edges of the side pieces, 5" down from the top crosspiece. Check the pieces for square, and clamp in place. Weld the interior crosspiece to the sides.

3. Place the vertical inserts at 4" intervals between the interior and bottom crosspieces. Make sure the spacing is even—you might need to adjust to slightly less or more than 4" if the miter cuts are slightly off. Keep the outside edge of the inserts flush with the outside edge of the crosspieces.

4. When the spacing is adjusted properly and the inserts are square to the crosspieces, tack weld each insert at both ends.

5. Check the assembly and inserts for square one more time. Turn the assembly over, and weld each upright in place.

Attach the Found Objects

Objects other than mild steel will need to be brazed or braze welded. Connect non-metallic objects by wrapping or folding a thin strip of mild steel sheet metal or a short piece of 1/8" steel rod around an edge. Weld the ends of the rod or strap to the framework.

1. Arrange your found objects artfully across the interior space.

2. Carefully clean rust or paint from the areas where the found objects contact the uprights.

3. Weld the found objects in place.

Attach the hinges

Our gate is made to hang attached to a metal gate post. We chose to use barrel style hinges.

1. Place the gate between the gate posts. Use wood spacers and braces to position the gate between the gate posts, and clamp or brace solidly in place.

2. Line up the hinges on the post and the gate (see photo). Use a level to check for plumb. (For the gate to swing properly, the hinges need to be installed perfectly plumb.) Tack weld the hinges to the gate and post.

3. Remove the bracing, and check that gate swings freely. When it does, complete the hinge welds.

4. Install the gate latch hardware. If the hardware is painted or zinc coated, grind off the coating before welding, or install with screws.

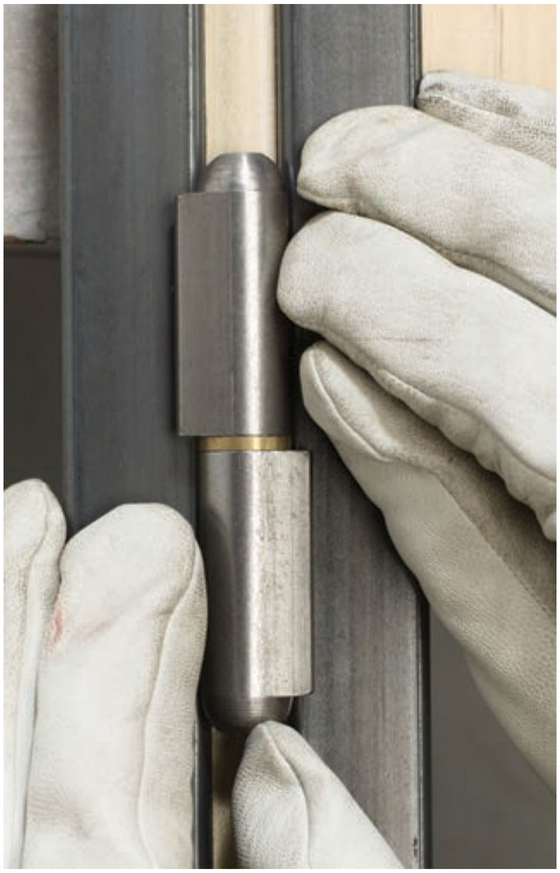

Brace the gate in position between the gate posts. Clamp a wood spacer between the gate post and the gate side. Position the hinge and check for plumb.

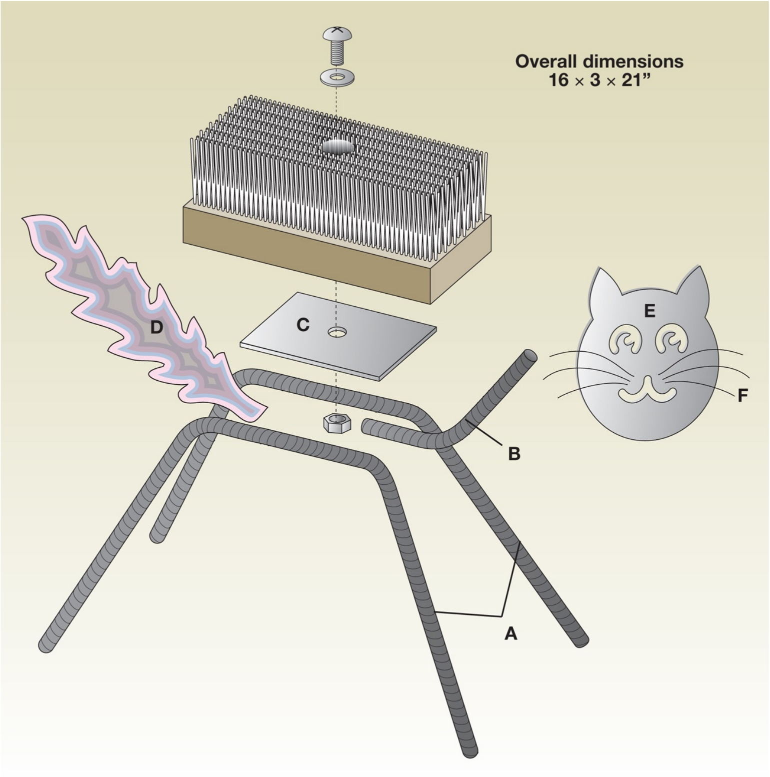
Overall dimensions
16 × 3 × 21"
C
D
E
F
B
A

Critter Boot Brush

A critter boot brush is a fun and functional addition to your back door area. This critter is made to be stuck in the ground (then the mud flies on the grass, not around the door!) but it can be mounted in cement or even on a hefty chunk of wood. This plan is for a cat because the brush mimics the classic upright fur look of an angry cat, but any animal could have this scratchy back.

PART	NAME	DIMENSIONS	QUANTITY
A	Legs	3/8 or 1/2" rebar × 30"	2
B	Neck	3/8 or 1/2" rebar × 12"	1
C	Body	1/8 or 3/16 × 3" flat bar × 6"	1
D	Tail	16 gauge sheet metal × 4 × 14"	1
E	Face	3/8 or 1/4" plate × 5 × 5"	1
F	Whiskers	16 gauge wire × 4"	6

How to Make a Critter Boot Brush

MATERIALS

- 1/8 or 1/2" rebar (6 feet)
- 1/8 or 3/16 × 3" flat bar (6")
- 16 gauge sheet metal (4 × 14")
- 1/8 or 1/4" plate (5 × 5")
- Wire or bristle brush with handle hole
- 2" machine bolt, washer, and nut
- 16 gauge wire (24")

Prepare the Legs & Neck

1. Cut the legs (A) to length.

2. Mark each leg 12" from each end. Using a bench vise, bend the legs at the marks to a 75° to 85° angle.

3. Cut the neck (B) to length.

4. Make a mark 3" from one end of the neck. Bend the neck at the mark to 85°.

Assemble the Body

1. Cut the body (C) to size.

2. Drill, flame cut, or plasma cut a hole through the center of the body for the placement of the brush.

3. Place the legs lengthwise along the edges of the body, and weld in place.

4. Position the hole in the brush over the hole in the body to get the spacing for the neck extension (if the neck is set back too far, the brush will not lay flat).

5. Weld the neck in place.

Attach the Tail

1. Use flame or plasma cutting to cut the tail (D), leaving a minimum of 4" of straight flat metal for attachment.

2. Remove any slag from the tail, and weld in place between the legs at the opposite end from the neck (see photo).

Prepare & Attach the Face

If using a flame cutter for the face, you will get the best results if you allow the metal to cool between cutting each individual feature. If you don’t let it cool, the metal begins to melt and round over, making it difficult to get a sharp outline of the features.

1. Use flame or plasma cutting to cut out the face shape (E) and features.

2. Remove any slag from the back side of the face. Cut the whiskers (F) to size and attach to the face.

3. Position the face and weld it to the end of the neck.

Attach Brush & Set in Place

1. Bolt the brush onto the body and adjust the tail in an upright position.

2. Push the legs into the ground or mount the critter in a concrete base.

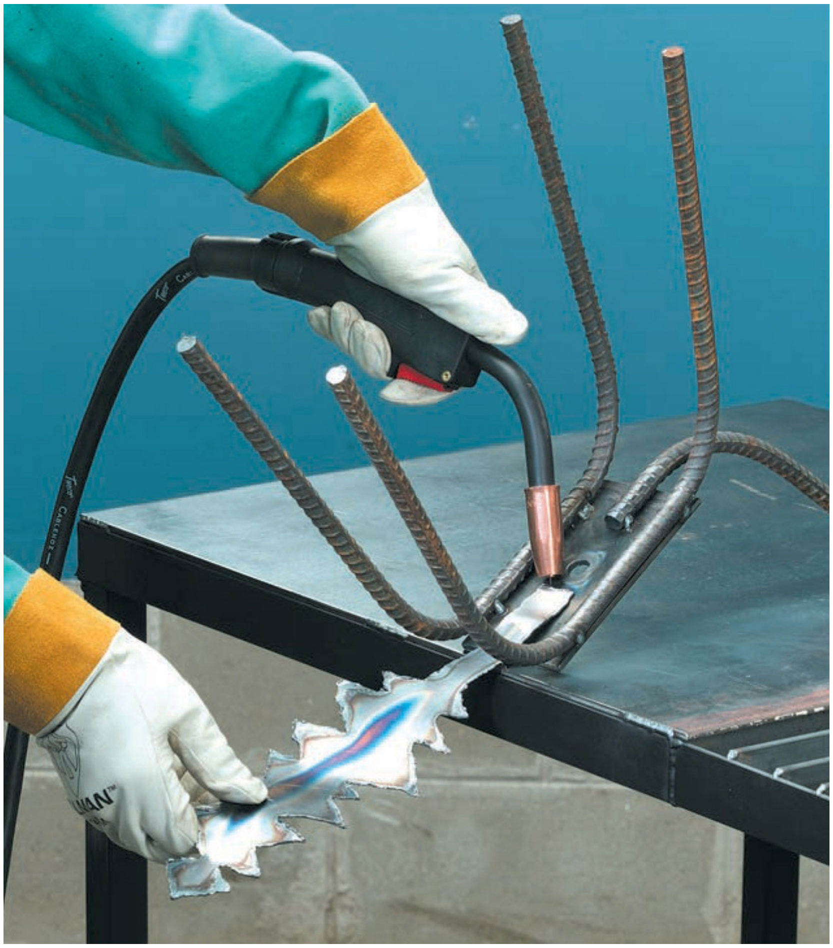

Cut and bend the legs and neck to size, then weld the legs, neck, and tail to the body.

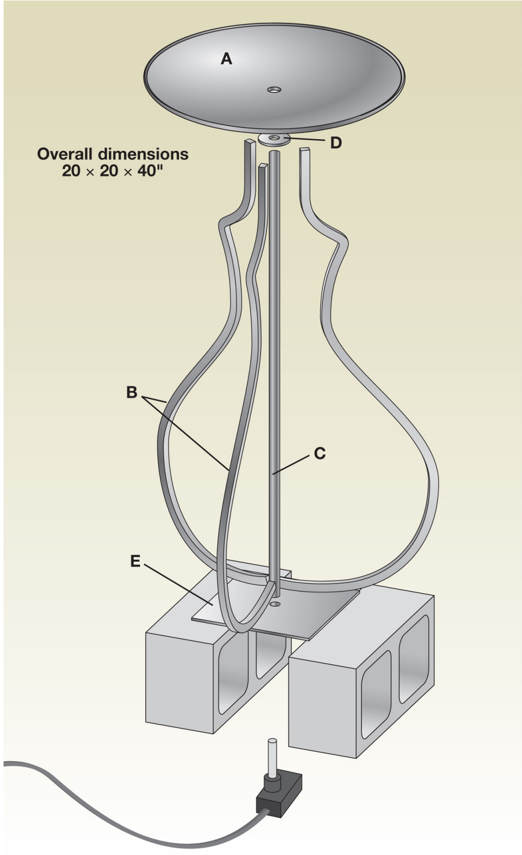
A
D
Overall dimensions
20 × 20 × 40"
B
C
E

PART	NAME	DIMENSIONS	QUANTITY
A	Basin	9 gauge × 20" plow disk	1
B	Legs	Decorative pickets × 35"	3
C	Supply pipe	1/8 × 1" round tube × 35-1/2"	1
D	Flange	3/16" sheet metal × 3" circle	1
E	Base	1/4" plate × 12 × 12"	1

Disk Fountain

This fountain uses a plow disk, which can be found at farm implement stores, and three decorative fence pickets. Decorative scrolls also could be used. You'll get to see how watertight your welds are, because you'll need to weld the supply pipe and a flange to the disk. You could use a slightly smaller disk for the base—we simply used a square plate.

MATERIALS

- 9 gauge × 20" plow disk (1)
- Decorative fence pickets (3)
- 1/8 × 1" round tube (3 feet)
- 3/16" sheet metal (3" circle)
- 1/4" plate (12 × 12")
- Fountain and pump

How to Make a Disk Fountain

Assemble the Flange

You can purchase round metal blanks in 3" sizes, or you can cut your own. We purchased one at a steel surplus store.

1. Cut the supply pipe (C), flange (D), and base (E) to size.

2. Cut a 1" diameter hole in the center of the flange and the center of the base.

3. Place the hole in the flange over the end of the supply pipe and weld all around.

4. Grind down the weld until it is flush.

Attach the Basin

1. Grind off the paint from the basin (A) in a 4" diameter circle around the center hole.

2. Turn the basin over and place the flange and pipe assembly on top of it, centering it over the hole in the basin.

3. Tack weld the flange to the disk in at least three places. Make sure the supply pipe is perpendicular to the basin, then complete the weld all around the flange.

Attach the Legs

1. Tack weld the legs to the basin assembly (see photo), keeping an equal distance between each leg.

2. Insert the supply pipe into the hole in the base and tack in place.

3. Turn the assembly upright, and check for level. Adjust if necessary. Complete the welds for the legs and the base.

Apply Finishing Touches

1. Grind down all rough welds. Thoroughly wire brush or sandblast the fountain.

2. Prime and paint the fountain with a high quality outdoor metal paint.

3. Place the fountain on a brick or cement block base in a pool or over an appropriately sized reservoir.

4. Install the fountain hardware and tubing following manufacturer's directions.

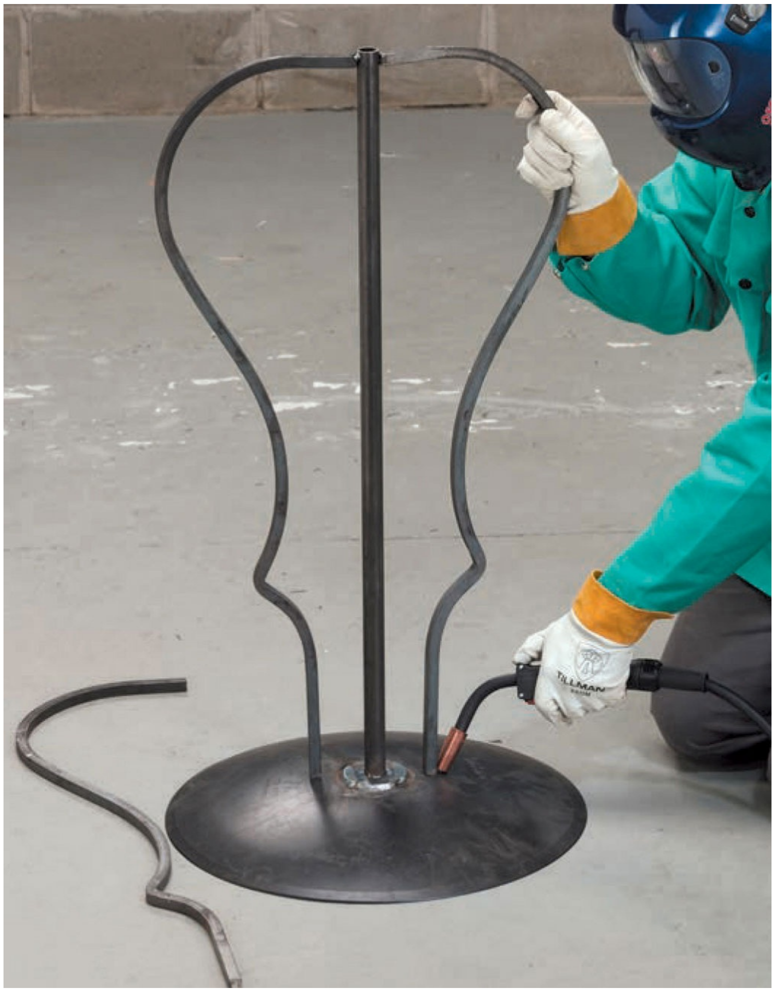

Tack weld the pickets to the basin and to the supply tube.

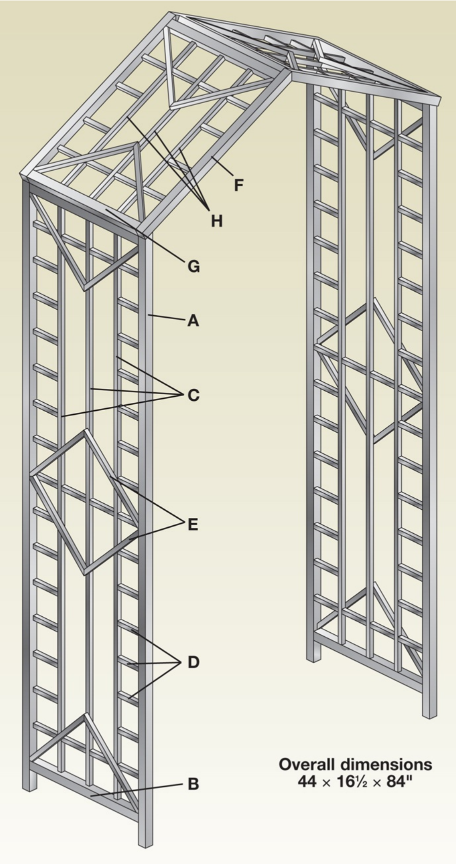
F
H
G
A
C
E
D
B
Overall dimensions
44 × 16½ × 84"

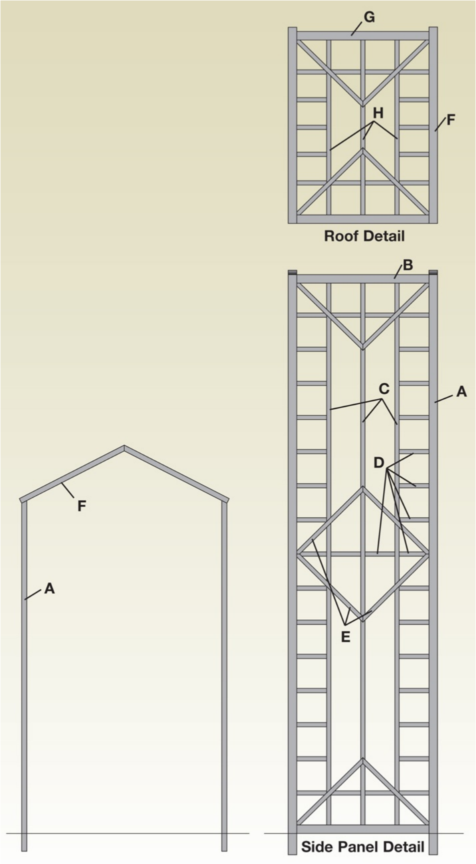

Roof Detail

Side Panel Detail

G
H
F
B
C
A
D
F
A
E

Arbor

This distinctive arbor will add a touch of class to any garden. Modeled after the “Prairie Style,” its clean lines are straightforward to cut and weld, and they provide plenty of climbing support for a variety of vines. The 1/2" inserts are placed flush with the back of the 1" sides and ends, eliminating the need for difficult centering and giving an increased sense of depth. The arbor can be made from the 6-foot steel lengths available at home improvement centers, but because of the amount of steel used, a trip to a steel supplier might be worthwhile, especially if you have a way to transport 20-foot lengths. The 1/2" square tube bows quite easily, so if you do buy 20 foot lengths, make sure they are well supported in transit. Otherwise, they will bow and not be suitable for the vertical inserts.

PART	NAME	DIMENSIONS	QUANTITY
A	Panel sides	16 gauge 1 × 1" square tube × 72"	4
B	Panel ends	16 gauge 1 × 1" square tube × 15-1/2"	4
C	Panel vertical inserts	16 gauge 1/2 × 1/2" square tube × 63-1/2"	6
D	Horizontal inserts	16 gauge 1/2 × 1/2" square tube × 3-1/2"	100
E	Diagonal inserts	16 gauge 1/2 × 1/2" square tube × 12"*	24
F	Roof ends	16 gauge 1 × 1" square tube × 15-1/2"	4
G	Roof sides	16 gauge 1 × 1" square tube × 25-1/2"	4
H	Roof vertical inserts	16 gauge 1/2 × 1/2" square tube × 23-1/2"	6

*Approximate dimension, cut to fit

MATERIALS

- 16 gauge 1 × 1" square tube (44 feet)
- 16 gauge 1/2 × 1/2" square tube (100 feet)
- Wood scraps

Check the panel for square by measuring the diagonals. If the measurements are equal, the assembly is square.

How to Build an Arbor

Assemble the Panels

Set up the project on a sheet of 3/4" plywood placed on sawhorses. Make sure the plywood is not bowed or it will cause misalignment of the project pieces. Working on a raised surface is easier than working on the floor, and you can clamp the metal to the plywood. Remember to clamp your work cable to the workpiece. If you want the arbor wider or narrower than 44", you must adjust the miter angles for the roof and side panels.

1. Cut the panel sides (A) to size, mitering one end at 30°. The mitered end is the top. Cut the panel ends (B), vertical inserts (C), and horizontal inserts (D) to size.

2. Place a panel end between the top of two panel sides, keeping the outside edge of the panel end flush with the short ends of the mitered panel sides. Clamp the pieces in place.

3. Position three vertical inserts between the two panel sides. (This is just to hold the bottom panel end in place. The exact location of the inserts is not important at this time.)

4. Place the bottom panel end between the panel sides, keeping it snug against the vertical inserts. Clamp the bottom panel end in place.

5. Remove the vertical inserts. Use a carpenter’s square to check the panel sides and ends for square. Tack weld each corner. Check for square again by measuring the two diagonals (see photo, page 135). If the measurements are equal, the panel is square. If not, adjust the workpieces until the diagonals are the same.

6. Turn the panel over and complete the welds at the four corners. Grind the tack welds flat so the panel lays flat on the work surface.

Attach the Vertical & Horizontal Inserts

1. Replace the vertical inserts between the panel sides. Place several horizontal inserts between the vertical inserts and the panel sides to ensure the correct spacing of the vertical inserts.

2. Weld the vertical inserts to the panel top and bottom.

3. Place the horizontal inserts into the vertical framework every 3-1/2". (Use horizontal inserts as spacers.)

4. Starting at the top of the panel, use a combination square to align the first row of horizontal inserts. Once they are aligned, weld them in place.

5. Continue aligning and welding the inserts one row at a time. If the vertical inserts are bowed side to side, use a clamp to hold them against the inserts. If they bow upward, weight them down with a sandbag or other weight.

Insert the Diagonal Inserts

1. Cut four diagonal inserts (E) to size, mitering one end at 45°.

2. Place two diagonal inserts against the bottom edge of the top panel. The mitered ends of the inserts should butt together over the middle vertical insert. Mark the other end of the diagonal inserts where they cross the top panel so they will fit in the corner.

3. Make the angled cuts on the diagonal inserts.

4. Grind down the welds that will be underneath the diagonal inserts so they will lie flat. Weld the diagonal inserts in place (see photo).

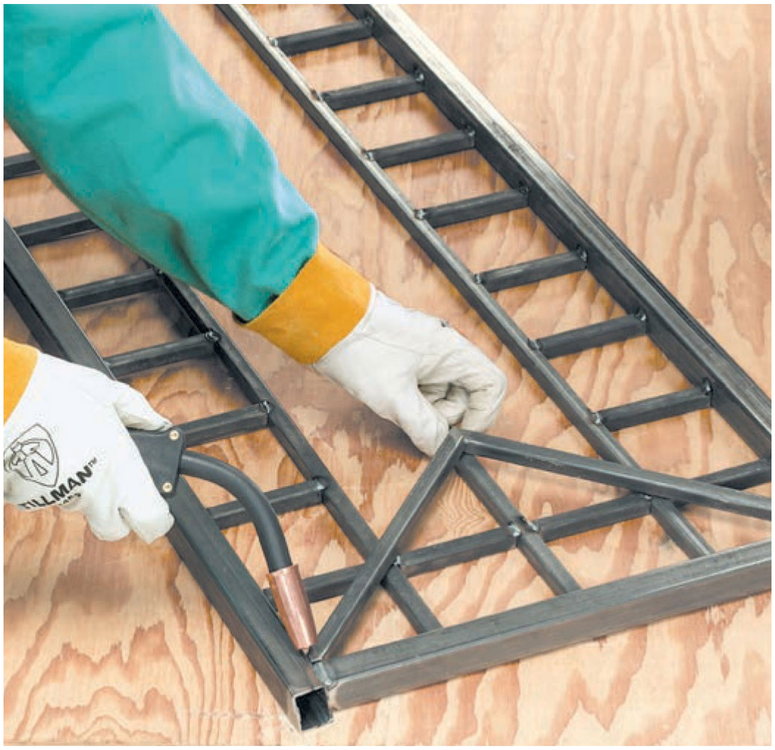

Weld the diagonal inserts to the panel framework and to the center vertical insert.

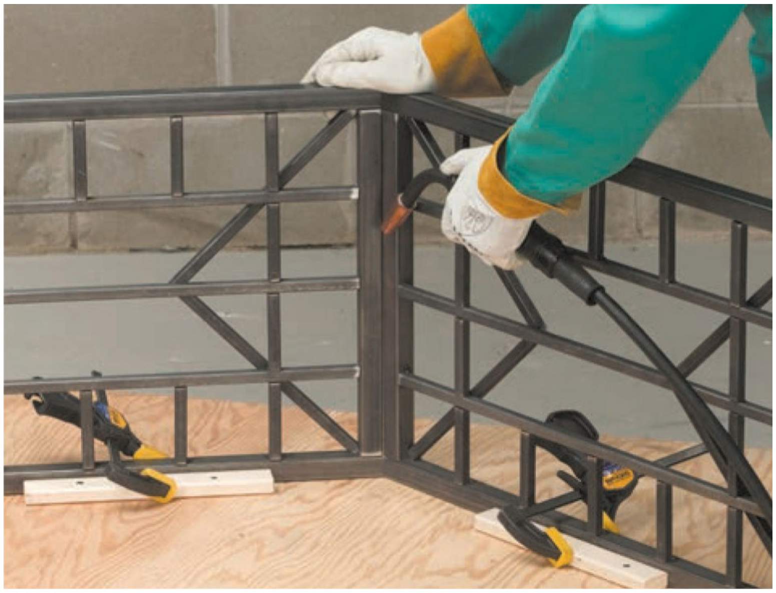

Make clamping blocks out of wood scraps. Clamp the roof panels to the blocks. Place two or three tack welds along the joint between the panel ends.

5. Repeat steps 2 to 4 to weld the diagonal inserts to the bottom of the panel.

6. Cut four more diagonal inserts, mitering both ends at 45°.

7. Center these four inserts over the eighth row of horizontal inserts to form a diamond shape. Weld the inserts in place.

8. Repeat this entire process to construct a second panel.

Build the Roof Panels

1. Construct two roof panels following the same procedure used for the side panels, except there is no central diamond insert.

2. Stand the roof panels on edge and join the mitered ends to form the 120° angle for the roof peak.

3. Clamp the roof panels in place. (You may need to attach temporary wooden clamping points to the work surface by fastening 1 × 2 or 2 × 4 scraps to match the layout. Clamp the panels to the scraps.) Place two or three tack welds along the joint between the panel ends (see photo).

Fasten the Roof to the Sides

1. Place a side panel on edge, and set it against one side of the roof panel. Line them up so the mitered edge of the side panel is making contact with the bottom edge of the roof panel. Clamp the panels in place.

2. Tack weld the panels at two points along the joint (see photo).

3. Repeat steps 1 and 2 for the other side panel.

4. Measure the distance between the tops and the bottoms of the side panels to ensure the panels are an equal distance apart at both ends.

5. Check the panels for square by measuring the diagonals. Make any necessary adjustments, and make sure the roof assembly is still fitted against the side panels.

6. Weld all the mitered roof peak joints and the mitered joints between the side panels and the roof panels.

Finishing Touches

To prevent putting stress on the welds, and to maintain the arbor’s shape, clamp wooden crosspieces between the panel sides before moving it.

1. The easiest finish for the arbor is to allow it to gently rust over the years. You may want to place plastic end caps over the exposed tube ends on the roof, or you can weld on small caps.

2. The arbor can sit on the ground or can be mounted in concrete footings. If you live in a windy location, you may want to drive four lengths of rebar into the ground and slip the arbor legs over them.

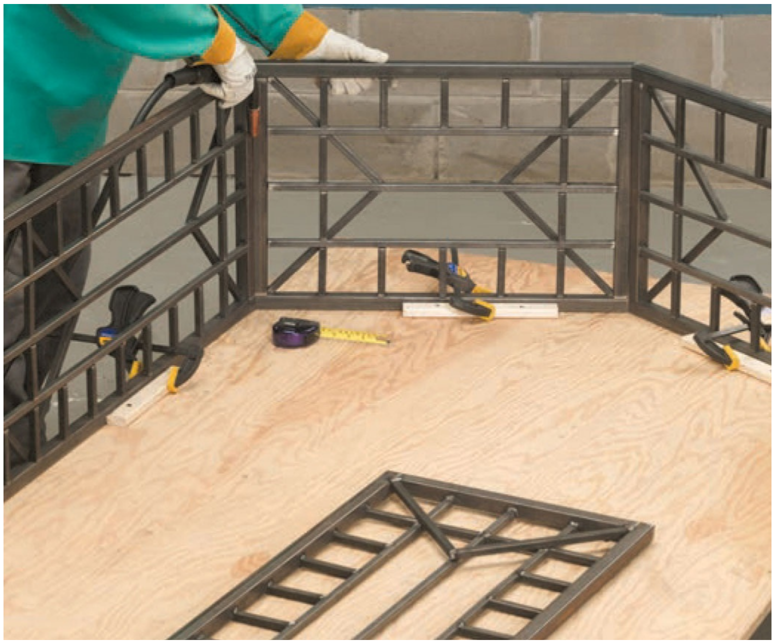

Place tack welds along the joint between the roof and side panels.

GLOSSARY

Active gas	A gas, such as oxygen, that will react with other substances.
Alloy	A substance composed of two or more metals or a metal and other non-metal material.
Arc welding	Welding processes that use an electric arc to produce heat to melt and fuse the base metals. Often used to refer specifically to shielded metal arc welding.
Backfire and flashback	Combustion taking place inside the oxyacetylene torch creating a loud pop or explosion.
Base metal	The metal that is being welded, brazed, braze welded, or cut.
Bead or weld bead	The seam between workpieces that have been joined with welding.
Braze	A process used to join metals using a non-ferrous material that melts above 850° F but below the melting point of the base metal. Brazing uses capillary action to join closely fitted parts.
Braze welding	A process used to join metals with a filler material that melts above 840° F and below the melting point of the base metal where the filler metal is not distributed by capillary action.
Butt joint	A joint between two workpieces in the same plane.
Carburizing flame	An oxyfuel flame with an excess of fuel.
Consumable electrode	An electrode that also serves as the filler material.
Corner joint	A joint between workpieces meeting at right angles and forming an L shape.
Cutting tip	Converts an oxyfuel welding torch into an oxyfuel cutting torch.

Direct current electrode negative	Direct current welding where the electrode is negative and the workpiece is positive.
Direct current electrode positive	Direct current welding where the electrode is positive and the workpiece is negative.
Duty cycle	The amount of continuous time in a 10-minute period that an arc welder can run before it needs to cool down. Expressed as a percentage at a given amperage output.
Edge joint	A joint between parallel workpieces.
Electrode	The conductive element that makes the connection with the workpiece to create an electrical arc.
FCAW	Flux cored arc welding.
Filler, filler metal	Metal added to a welded joint.
Fillet weld	A triangular shaped weld between two members that meet at 90° angles.
Flame cutting	See oxyfuel cutting.
Flux	A chemical compound that produces cleaning action and reduces the formation of oxides when heated.
Flux core, flux cored wire	An electrode for flux cored arc welding that contains flux within a wire tube.
Flux cored arc welding (FCAW)	A semi-automatic arc welding process using an electrode wire that contains flux.
Gas metal arc welding (GMAW)	A semi-automatic arc welding process using a wire electrode which also is the filler material. An inert gas is distributed over the weld area to shield the molten metal from oxygen.

	Commonly referred to as MIG (metal inert gas) or wire feed.
Gas tungsten arc welding (GTAW)	An arc welding process using a tungsten electrode, hand-held filler material, and an inert shielding gas. Also referred to as TIG (tungsten inert gas) and Heliarc.
GMAW	Gas metal arc welding.
Groove weld	A weld made in grooves between workpieces.
GTAW	Gas tungsten arc welding.
Heliarc	Gas tungsten arc welding.
Inert gas	A non-reactive or non-combining gas such as argon or helium.
Kerf	The width of a cut.
Lap joint	A joint between overlapping workpieces.
MIG	Metal inert gas—see gas metal arc welding.
Oxyacetylene cutting	Oxyfuel cutting using acetylene as the fuel gas.
Oxyacetylene welding	Oxyfuel welding using acetylene as the fuel gas.
Oxyfuel cutting	Cutting process using the combustion of a pressurized fuel gas and oxygen to heat steel to 1600° F at which time a pure oxygen stream is delivered to burn through the metal. Also called flame cutting.
Oxyfuel welding	Welding process that uses the heat produced by the combustion of a pressurized fuel gas and pressurized oxygen. A hand-held filler material is used. Also called gas welding.
Plasma cutting	An arc cutting process using a constricted arc. Compressed air or inert gas blows the molten metal from the kerf
Plate	Flat metal thicker than 3/16".
Saddle joint	A joint between round tubes where one tube has been cut to fit around the other.
Sheet metal	Flat metal 3/16" or less in thickness.
Shielded	

metal arc welding (SMAW)	An arc welding process using a flux coated consumable electrode. Also referred to as arc welding or stick welding.
Shielding gas	A gas that prevents contaminants from entering the molten weld pool.
Slag	Oxidized impurities formed as a coating over the weld bead; waste material found along the bottom edge of an oxyfuel cut.
SMAW	Shielded metal arc welding.
Spatter	Small droplets or balls of metal stuck to the base metal around the weld. Produced by shielded metal arc, gas metal arc, and flux cored arc welding.
Stick welding	Shielded metal arc welding.
T-joint	A joint between workpieces meeting at right angles and forming a T shape.
TIG	Tungsten inert gas—see gas tungsten arc welding.
Wire feed	See gas metal arc welding.

RESOURCES

Thanks to Tony Litwinchuk of Toll Gas & Welding Supply for arranging for the generous loan of the welding machinery pictured in this book. Thank you to Loren Johnson, also of Toll, for his technical advice on a variety of issues. Thank you to Jerry Foley and Dwight Affeldt, the welding instructors at Minneapolis Community and Technical College Continuing Eduction and Training Programs. Their knowledge of and enthusiasm for welding is fantastic. Thank you to Dwight also for reading the book and suggesting changes to make it more accurate. Finally, thank you to Paige Gamet for giving me the gift of my first welding class.—*Karen Ruth*

Architectural Iron Designs, Inc.
950 South 2nd Street
Plainfield, NJ 07063
800-784-7444
www.archirondesign.com

e-Tarps.com
1205-202 Winter Springs Court
Louisville, KY 40243
www.e-tarps.com

Garland's Inc.
2501 26th Avenue South
Minneapolis, MN 55406
800-809-7888
www.garlandsinc.com

Toll Gas & Welding Supply
3005 Niagara Lane
Plymouth, MN 55447
763-551-5300
www.tollgas.com

Triple-S Steel Supply Company
6000 Jensen Drive
Houston, TX 77026
800-392-3655
www.sss-steel.com

Wagner Railing Systems and Components
J.G. Braun Company
10600 West Brown Deer Road
Milwaukee, WI 53224
800-786-2111
www.rbwagner.com
www.jgbraun.com

CONTRIBUTORS

Beateworks, Inc.
Los Angeles, CA
©Millet/Inside/Beateworks: p. 5 (top)

Birds & Beasts
rustybeast@myway.com
www.birdsandbeasts.com

ColdSnap Photography
Two Harbors, MN
©John Gregor/ColdSnap Photography: p. 4

Gene Olson
gene@mettleworks.com
Elk River, MN
Photo p. 6 (top left) by Ytsma Photography
St. Paul, MN

Heebie GBS Metalworks
www.heebiegbs.com

Lizard Breath Ranch, Inc.
Critters from the "Hood"
(free standing animals cut from truck roofs and hoods)
www.lizardbreathranch.com
505-534-4881

Sleeper Welding
www.sleeperwelding.com
603-524-1597

INDEX

Acetylene, 27, 35
  See also: Oxyacetylene cutting, Oxyacetylene welding
Air compressors, 64, 65, 66
Alloys, 15, 22, 57, 62
Aluminum, 15
  preparing, 18
  repairing, 22, 23
Amperage, 38–39, 42, 43, 65
Angle iron, 16
Arbor, 134–137
Arc welding,
  electricity, 38–39
  preparing for, 23
  See also: Flux cored arc welding, Gas metal arc welding, Gas tungsten arc welding, Shielded metal arc welding
Arcs, 38, 64, 66
  burn, 10
  striking, 60
Artwork, 4, 5, 6, 7, 36
Auto-darkening lenses, 11, 54
  See also: Helmets

Backfire, 31
Backseating valve, see: Cylinders
Baker's shelves, 106–109
Base metals,
  identifying, 22
  matching electrode size, 8, 23
  matching filler material, 8, 23
  matching welding process, 9
  See also: Metals
Bird's nests, 53
Blowpipes, see: Torches Boot brush, 130–131
Brakes, 19
Braze welding, 14, 37
Brazing, 11, 14, 35, 36
Burns, 10, 68
Burn through, 8, 34
Butt joint, 20

Candelabra, 5
Candleholders,
  chandelier, 96–97
  fireplace, 102–103
  tree-shaped, 100–101
  wall-mounted, 98–99
Carburizing flames, 33
Carts, see: Cylinders, Welding machines
Cast aluminum, see: Aluminum
Cast iron, repairing, 23
Chandelier, 96–97
Channel, 16
Chemical sharpeners, 57, 58
Circle cutting, 67, 71
Clothing, protective, 10, 11, 27, 41, 45, 54, 55, 65, 68, 69
Coatrack, 112–113
Coffee table, 110–111
Collet and collet body, see: Torches
Concrete, risk of explosion, 11
Corner joint, 20
Corrosion, 9, 15
Costs, 5, 17, 26, 65, 67, 68, 69
Cutouts, making, 71
Cutting processes, 14, 38–39
  See also: Oxyacetylene cutting,
  Plasma cutting
Cylinders,
  cart, 85–87
  gas metal arc welding, 44, 47, 49
  gas tungsten arc welding, 54, 55, 61
  oxyacetylene cutting, 68
  oxyacetylene welding, 26, 30
  risk of explosion, 11, 28, 65
  safety, 27, 47, 68, 71
  valves, 27, 29, 31, 32, 49
  wrenches, 30, 31

Decorative metal products, 17, 94–95, 102–103, 110–111, 114–115, 118–119, 120–121, 124–127
Duty cycles, 39, 65

Edge joint, 20
Electrical shocks, 11, 12, 40, 65
Electricity, 38–39, 41, 55, 59, 64
Electrodes, 8, 23, 40, 42, 44, 46, 50, 52, 54, 55, 58, 65

characteristics, 57
holders, 40, 41, 43
leads, 41, 43
stub out or stubbing, 53
tungsten, 56, 57
Equipment, see: Air compressors,
Cylinders, Electrodes, Fillers, Fire bricks, Fittings, Flashback arrestors, Flow meters, Foot control, Gases, Gauges, Guns, Hoses, Power supply, Regulators, Supply cables, Tips, Tools, Torches, Welding machines, Wire feed, Work cables/leads and clamps
Explosions, risk of, 11, 28, 65

FCAW, see: Flux cored arc welding
Ferrous metals, 14
Fillers, 8, 15, 23, 34, 42
braze welding, 37
brazing, 36
gas metal arc welding, 44
gas tungsten arc welding, 54, 58, 60–61, 62
oxyacetylene welding, 26
Fillet weld, 20, 21, 37, 63
Filters, 10, 11, 27, 41, 45, 55, 65, 69
See also: Shades, Shields Finishes, 18
Fire bricks, 11, 27, 34, 35, 37
Fire extinguishers, 10, 71
Fires, risk of, 10, 12, 68
Fishmouth joint, 20
Fittings, oxyacetylene, 28
Flame cutting, see: Oxyacetylene cutting
Flames, 32, 33, 70
Flange joint, 20
Flashback, 28, 31
arrestors, 29
Flat bar, 16
Flow meters, 47, 49, 50, 51, 55, 56, 61
Flux, 9, 36, 37, 42, 45
Flux cored arc welding, 9, 47, 51
Foot control, 54, 55, 60
Fountain, 132–133
Fuel gases, see: Gases
Fumes, 10, 12, 41, 64
Furniture projects, 104–121
Fusion welding, 26, 34
Gas leaks, 30, 31
Gas metal arc welding, 8, 9, 53

  electricity, 38–39
  equipment, 44, 45–47, 52
  maintenance, 52–53
  safety, 47
  shades/lenses, 11, 45
  troubleshooting, 52–53
  See also: Flux cored arc welding
Gas tungsten arc welding, 9, 54–63
  electricity, 38–39, 59, 60
  equipment, 54, 55–56, 59
  preparing the electrode, 56–58
  safety, 54, 55
  shades/lenses, 11, 54
  troubleshooting, 62–63
Gas welding, see: Oxyacetylene welding
Gases,
  fuel, 26, 35, 69
  shielding, 42, 44, 45, 46, 47, 49, 50, 51, 54, 55, 56, 61
Gates, 7, 129
Gauges, 30, 31
Gauge thickness, inch equivalents, 17
Glasses, safety, 11
GMAW, see: Gas metal arc welding Grinder stand, 88–89
Groove weld, 20, 21, 37
Ground leads and clamps, see: Work cables/leads and clamps
GTAW, see: Gas tungsten arc welding
Guns, 44, 45, 48, 50, 51, 52
  spool, 46
Hard soldering, 36
Headboard, 120–121
Heat distortion, 8
Heat-treated metals, 15, 22
Heliarc, see: Gas tungsten arc welding, 54
Helmets, 11, 27, 41, 45, 55, 65
  auto-darkening lenses, 11, 54
  See also: Filters, Shades, Shields Hexagonal bar, 16
High-carbon steel, 15
Holes, piercing, 71
Hooks, making, 113
Hoses,
  gas metal arc welding, 47, 49, 50, 52
  gas tungsten arc welding, 61
  oxyacetylene cutting, 68, 70

  oxyacetylene welding, 26, 28, 29, 30, 34, 35

Infrared light, safety, 10, 41
Innershield, see: Flux cored arc welding
Insurance, 12

Jigs, 19, 108, 109, 115, 117
Joints, 20

Kerf, 64, 71
Kitchen appliance stand, 114–115

Lamps, 92–93, 94–95
Lap joint, 20
Lighting projects, 90–103

Melting points, 15
Metal inert gas (MIG), see: Gas metal arc welding
Metals, 14–23
  braze welding, 37
  brazing, 36
  cutability, 14, 19
  finishing, 18
  heat-treated, 15
  measuring, 16–17
  melting points, 15
  painting, 18
  preheating, 23, 69, 71
  preparing, 18, 22, 34
  purchasing, 17, 79, 88, 92, 125, 132–133, 140
  repairing, 22–23
  shapes, 16
  shaping, 19, 93, 97, 108, 109, 117
  sizes, 16
  types, 14–15
  weldability, 14
  See also: Decorative metal products, Fillers
Metalworking tools, 19, 107, 108, 109, 117
MIG, see: Gas metal arc welding MIG pliers, 45, 52
Mild steel, 14, 16
  preparing, 18
  repairing, 22
Mill scale, 18, 34
Molten puddle, 8, 9, 58, 60

Neutral flames, 33, 35, 70
Non-ferrous metals, 14
Nozzles, see: Tips

Outdoor projects, 122–137
Oxidizing flames, 33
Oxyacetylene cutting, 9, 68–71
  equipment, 68–69
  pressuring the system, 30, 31, 70
  safety, 7, 68, 69, 71
  setting up, 30, 70
  shades/lenses, 11, 69
  tools, 69, 71
Oxyacetylene welding, 8, 26–37
  depressurizing the system, 35
  equipment, 26–29
  leaks, 30, 31
  pressuring the system, 30, 31, 34
  safety, 27, 28, 29, 31, 32
  setting up, 30
  shades/lenses, 11, 27
  tools, 27
Oxyfuel brazing, see: Brazing
Oxyfuel cutting, see: Oxyacetylene cutting
Oxyfuel welding, 26
  See also: Oxyacetylene welding
Oxyfuels, 26, 35
Oxygen, 9

Particle masks, 10, 11
Penetration of the weld, 8, 22
Pipe, 16
Plant stand, 7, 114
Plasma cutting, 9, 64–67
  electricity, 38–39
  safety, 11, 64–65
  shades/lenses, 11, 65
  troubleshooting, 67
Plate metal, 16
Plate rack, 116–117
Power supply, 39, 40, 44, 45, 51, 54, 64
Preheating metals, 23, 69, 71
Pressure regulators, see: Regulators
Puddle, see: Molten puddle

Purchasing metal, 17, 79, 88, 92, 125, 132–133, 140
Racks,
  corner, 112–113
  plate, 116–117
  wine, 118–119
Rail cap, 16, 125
Railing, see: Stair railing
Rectangular tube, 16
Reducing flames, 33
Regulators,
  gas metal arc welding, 44, 47, 50
  gas tungsten arc welding, 54, 55
  oxyacetylene cutting, 68, 69
  oxyacetylene welding, 26, 28, 30, 31, 32, 35
Repairing metal, 22–23
Resources, 140
Respirators, 10, 11
Round bar, 16
Round tube, 16
Rust, 9, 15, 34

Saddle joint, 20
Safety, 7, 9, 10–11, 12, 22, 27, 28, 29, 31, 32, 40, 41, 47, 54, 55, 64–65, 68, 69, 71, 125, 128
Scratch starts, 60
  See also: Arcs
Scrolls, making, 92–93, 96–97, 102, 107–109, 116–117
Shades, 11
  See also: Filters, Shields
Sheet metal, 16, 17
Shelves,
  baker's shelves, 107–109
  corner coatrack, 112–113
  kitchen appliance stand, 114–115
Shielded metal arc welding, 8, 9, 43
  electricity, 38–39
  equipment, 40, 41–42
  safety, 40–41
  shades/lenses, 11, 41
Shielding gases, 9, 12, 42, 44, 45, 46, 47, 49, 50, 51, 54, 55, 56, 61
Shields, for helmets, 11, 27
  See also: Filters, Shades
Shocks, see: Electrical shocks
Shops, welding,
  projects, 74–89

- safety, 12, 40
- See also: Costs

Slag, 9, 40, 42, 43, 47, 51, 70, 71, 79
- burns, 10, 68
- risk of fire, 12, 68

SMAW, see: Shielded metal arc welding
Smoke, 10, 58, 64
Soldering, see: Brazing
Spark lighter, 27, 32, 69, 70
Sparks, 64, 68, 71, 79
- burns 10
- risk of fire, 12

Spatter, 41, 47, 50, 52
Square bar, 16
Square tube, 16
Stainless steel, repairing, 23
Stair railing, 125–127
Stands,
- grinder stand, 88–89
- kitchen appliance, 114–115
- See also: Tables

Starter cartridge, see: Plasma cutting
Stick welding, see: Shielded metal arc welding
Strap, see: Flat bar
Stub out or stubbing, 53
Suppliers, 17, 140
Supply cables, 44, 45, 50, 51, 52, 55, 64
Tables, see: Coffee table, Welding table
- See also: Stands

Tanks, see: Cylinders
T-bar, 16
Temperature crayons, 23
TIG, see: Gas tungsten arc welding
Tip cleaner, 27, 69, 71
Tips,
- cutting head, 28, 29, 65, 70
- gas metal arc welding, 46, 50–51, 52
- matching to material or process, 29, 31, 34
- oxyacetylene cutting, 68, 69, 70, 71
- oxyacetylene welding, 26, 28, 29
- rosebud heating, 28, 29

T-joint, 20

Tools, 13, 97, 99, 101, 119
  gas metal arc welding, 45
  gas tungsten arc welding, 55, 57
  metalworking, 19, 107, 108, 109, 117
  oxyacetylene cutting, 69, 71
  oxyacetylene welding, 27
  plasma cutting, 65
  shielded metal arc welding, 41
Torches,
  flames, 32, 33, 70
  gas tungsten arc welding, 54, 55
  lighting, 32
  oxyacetylene cutting, 68, 69, 70
  oxyacetylene welding, 26, 29, 31, 32
  plasma cutting, 64, 65
  valves, 32
  See also: Guns, Tips
Tungsten inert gas (TIG), see: Gas tungsten arc welding

Ultraviolet light, safety, 10, 41, 54, 65

Valves, see: Cylinders, Torches
Ventilation, 10, 12, 27, 41, 43, 45, 55, 64, 65, 66, 69, 70
Voltage, 8, 39, 50–51, 65

Wattage, 38–39
Welding curtain, rolling, 76–77
Welding machines, 40, 41, 43, 44, 45, 46, 47, 49, 52, 53, 55, 59
  cart, 82–83
Welding processes, 14
  See also: Flux cored arc welding, Gas metal arc welding, Gas tungsten arc welding, Oxyacetylene welding, Shielded metal arc welding
Welding table, 79–81
Welding tarp, see: Welding curtain
Welds,
  fillet weld, 20, 21
  groove weld, 20, 21
  symbols, 21
  troubleshooting, 62–63
Wine rack, 118–119
Wire brushing, 18, 23, 34, 41
Wire feed, 44, 45, 46, 51, 53

  bird’s nests, 53
  setting up, 48
Wire speed, 8, 45, 48, 50–51
Wire spools, 46
Wires,
  tangles, see: Bird’s nests
  See also: Electrodes
Work cables/leads and clamps, 39, 40, 41, 43, 44, 46, 50, 51, 52, 54, 55, 61, 64, 66

Zigzag trim, 106–109

© Copyright 2004
Creative Publishing international, Inc.
400 First Avenue North, Suite 300
Minneapolis, MN 55401
1-800-328-3895
www.creativepub.com
All rights reserved

Printed by R. R. Donnelley
10 9 8 7 6 5 4 3 2 1

President/CEO: *Michael Eleftheriou*
Vice President/Publisher: *Linda Ball*
Vice President/Retail Sales & Marketing: *Kevin Haas*

Executive Editor: *Bryan Trandem*
Creative Director: *Tim Himsel*
Managing Editor: *Michelle Skudlarek*
Editorial Director: *Jerri Farris*

Author: *Karen Ruth*
Art Director: *Kari Johnston*
Copy Editor: *Brett Martin*
Project Manager: *Tracy Stanley*
Technical Editor: *Dwight Affeldt*
Proofreader: *Linnéa Christensen*
Illustrator: *Earl Slack*
Photo Researcher: *Julie Caruso*
Studio Services Manager: *Jeanette Moss McCurdy*
Photo Team Leader: *Tate Carlson*
Photographer: *Joel Schnell*
Scene Shop Carpenter: *Randy Austin*
Scene Shop Assistant: *Jacob Austin*
Director of Production Services & Photography: *Kim Gerber*

Production Manager: *Helga Thielen*
Welding Basics: An Introduction to Practical & Ornamental Welding
Created by: The Editors of Creative Publishing international, Inc.

Other titles from Creative Publishing international include:
The New Everyday Home Repairs; Basic Wiring & Electrical Repairs; Building Decks; Home Masonry Projects & Repairs; Workshop Tips & Techniques; Bathroom Remodeling; Flooring Projects & Techniques; Decorative Accessories; Kitchen Accessories; Maximizing Minimal Space; Outdoor Wood Furnishings; Easy Wood Furniture Projects; Customizing Your Home; Carpentry: Remodeling; Carpentry: Tools • Shelves • Walls • Doors; Exterior Home Repairs & Improvements; Home Plumbing Projects & Repairs; Advanced Home Wiring; Advanced Deck Building; Built-In Projects for the Home; Landscape Design & Construction; Refinishing & Finishing Wood; Building Porches & Patios; Advanced Home Plumbing; Remodeling Kitchens; Finishing Basements & Attics; Stonework & Masonry Projects; Sheds, Gazebos & Outbuildings; Building & Finishing Walls & Ceilings; Customizing Your Home; The Complete Guide to Home Plumbing; The Complete Guide to Home Wiring; The Complete Guide to Building Decks; The Complete Guide to Painting & Decorating; The Complete Guide to Creative Landscapes; The Complete Guide to Home Masonry; The Complete Guide to Home Carpentry; The Complete Guide to Home Storage; The Complete Guide to Windows & Doors; The Complete Guide to Bathrooms; The Complete Guide to Easy Woodworking Projects; The Complete Guide to Flooring; The Complete Guide to Ceramic & Stone Tile; The Complete Photo Guide to Home Repair; The Complete Photo Guide to Home Improvement; The Complete Photo Guide to Outdoor Home Improvement; Accessible Home; Open House; Lighting Design & Installation.

Digital Edition: 978-1-61060-298-3
Softcover Edition: 978-1-61060-298-3

Library of Congress
Cataloging-in-Publication Data
Welding basics : an introduction to practical & ornamental welding.

p. cm.
ISBN 1-58923-139-2 (sc.)
1. Welding. I. Creative Publishing International.
TS227.W355 2003
671.5’2--dc22
2003055399

WELDING COMPLETE

TECHNIQUES, PROJECT PLANS & INSTRUCTIONS

WELDING COMPLETE

TECHNIQUES, PROJECT PLANS & INSTRUCTIONS

Creative Publishing
international
www.creativepub.com

CONTENTS

INTRODUCTION

Basics

- Safety
- Setting Up Shop
- Shop Tools
- Metal Basics
- Metal Shapes & Sizes
- Purchasing Metal
- Repairing Metal
- Cleaning Metal

Techniques

- Mechanical Cutting
- Plasma Cutting
- Oxyacetelen Cutting
- Shaping
- Finishing
- Jointing
- Oxyacetylene Welding
- Shielded Metal Arc Welding (SMAW)
- Metal Arc Welding (GMAW/MIG)
- Gas Tungsten Arc Welding (GTAW/TIG)

WELDING PROJECTS

Shop

- Rolling Welding Curtain
- Welding Table
- Welding Machine Cart
- Cylinder Cart
- Grinder Stand
- Portable Welding Table
- Saw Stand

Home Décor

- Coffee Table
- Wine Rack
- Swivel Mirror
- Coat Hanger
- Room Divider
- Iron Cross
- Nesting Tables
- Swing-away Coat Hooks
- Nightstand
- Fireplace Candelabrum
- Magazine Rack
- Angled Metal Shelves
- Wall-mounted Shelf
- Étagère
- Corner Étagère

Outdoor

- Truck Rack
- Trailer Frame
- Stair Railing
- Gate
- Firepit
- Vineyard Trellis
- Arch & Gate
- Gazebo
- Arbor

APPENDIX: ADDITIONAL PLANS

GLOSSARY

RESOURCES & PHOTO CREDITS

INDEX

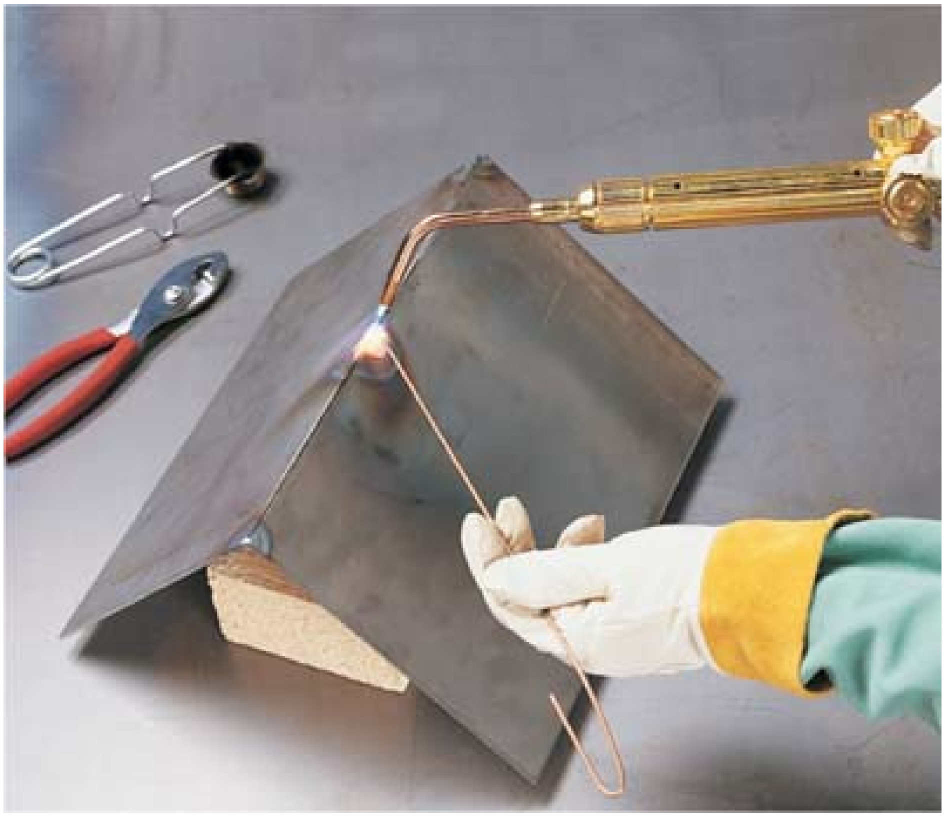

TILLMAN
TILLMAN

Introduction

Welding is a practical skill that is also great fun. The number of welded items in our everyday lives is practically uncountable—the spot welds on our cars and tractors, the welded railing on our front steps, the superstructure of the buildings in which we work, and the bridges over which we drive. Welding also makes everyday yard and interior items possible, like patio trellises, wine racks, tables, and shelves.

This book thoroughly discusses welding basics and techniques such as safety, setting up shop, tools, metals, and welding types. Once you have covered the basics, you can move on to the specific projects in this book—complete with how-to instructions, tools and materials lists, and illustrated plans—to develop your welding skills. At the same time, you'll benefit from a wide range of new furnishings in and around your home.

The home décor section includes instructions for practical and useful items commonly used in homes, such as mirrors, coat hooks, tables, and shelves, as well as decorative items like the iron cross and fireplace candelabrum.

The outdoor chapter includes heavy-duty projects suitable for hardworking DIYers, including a yard trailer and truck rack. There are also common exterior home improvement projects such as the stair railing, arbor, gate, and firepit.

Notice To Readers

Welding is a dangerous activity. Failure to follow safety procedures may result in serious injury or death. This book provides useful instruction, but we cannot anticipate all of your working conditions or the characteristics of your materials and tools. For safety, you should use caution, care, and good judgment when following the procedures described in this book. Consider your own skill level and the instructions and safety precautions associated with the various tools and materials shown. The publisher cannot assume responsibility for any damage to property or injury to persons as a result of misuse of the information provided.

Welding is all about heat, about using heat to melt separate pieces of metal so they will flow together and fuse to form a single seamless piece. Regardless of the welding or cutting process, your ability to control the heat generated by the flame or arc determines the quality of your welds and cuts.

Certain terms are used to describe the heat and action of all the welding processes. The parts being welded together are referred to as the base metal. Additional metal, called filler, is often added to the weld. The molten puddle is the area of melted base metal and filler metal that you maintain as you create your weld. To have fusion of metals, the base metal and filler metals must have the same composition. Methods for joining metal without fusion are called soldering, brazing, and braze welding. These methods can be used to join similar or dissimilar metals.

Oxyacetylene welding and cutting use flames to generate the heat to melt the metal. Shielded metal arc, gas metal arc, gas tungsten arc welding, and plasma cutting use an electric arc for heat generation. With oxyacetylene welding, you have time to watch the puddle develop as the metal turns red, then glossy and wet looking, then finally melts. With the arc processes, the puddle forms quickly and may be difficult to see because of the intensity of the arc. It is important to have ample lighting and clear vision so you can watch the puddle and move it steadily.

Penetration of the weld is also a critical heat-dependent factor. A strong weld penetrates all the way through the base metal. Matching filler material size and heat input to the thickness of the base metal is important. It is easy to get a good-looking weld that has not penetrated the base metal at all and merely sits on the surface. At the opposite end of the spectrum from a “cold,” nonpenetrating weld is burn through—where the base metal has gotten too hot and is entirely burned away, leaving a hole in the base metal and the weld.

Heat distortion is a byproduct of all welding and cutting processes. It is obvious that applying a flame to metal in the oxyacetylene process makes the metal hot—and the electrical arcs are actually four to five thousand degrees hotter than the oxyacetylene flame. When metal is heated, it expands; when it cools, it contracts. If not taken into consideration, this expansion and contraction may cause parts to move out of alignment. This is why tack welding and clamping project pieces is critical to successful welding. It is also important to match the welding process to the base metal thickness. For example, shielded metal arc welding is generally not used on materials less than 1/8″ thick. The process is too hot and too difficult to control on thin material. On the other hand, gas metal arc welding works well on very thin sheet metal, if you are able to adjust the machine to a low enough voltage.

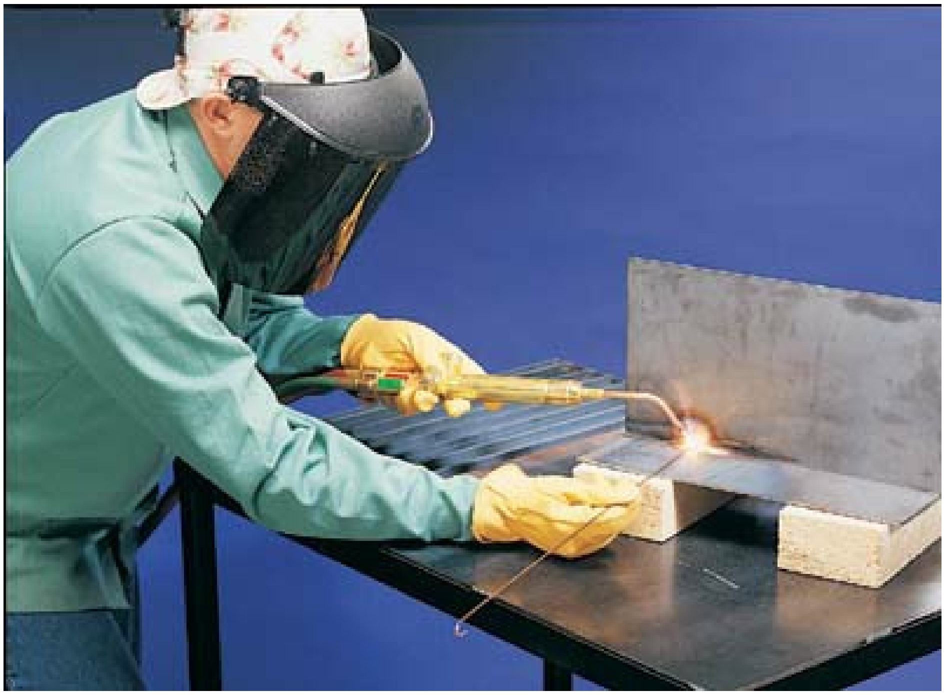

Oxyacetylene welding

Shielded metal arc welding

Gas metal arc welding

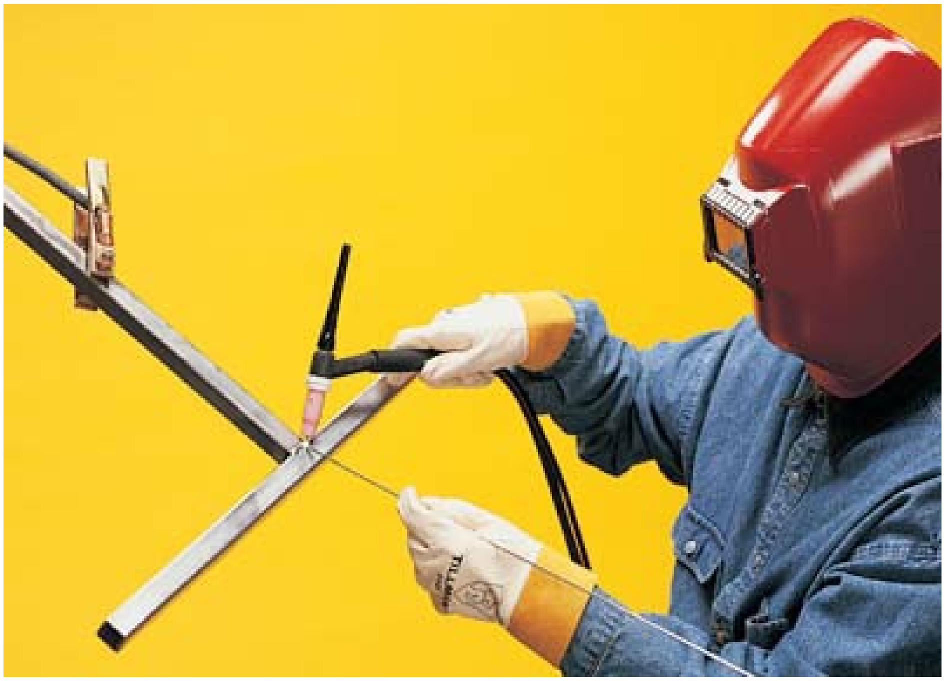

Gas tungsten arc welding

Plasma cutting

Oxyacetylene cutting

Protecting the molten puddle from oxygen is also an important part of welding. Oxygen makes steel rust and causes corrosion in other metals. The exclusion of oxygen from the weld when it is in the molten state makes a stronger weld. This is accomplished in a variety of ways. A properly adjusted oxyacetylene flame burns off the ambient oxygen in a small zone around the weld puddle. Gas metal arc and gas tungsten arc welding bathe the weld in an inert gas from a pressurized cylinder. Inert gas keeps oxygen away from the molten puddle. Flux cored arc welding and shielded metal arc welding use fluxes in or on the welding filler metal. When these fluxes burn, they produce shielding gases and slag, both of which protect the weld area until it has cooled.

Welding well is difficult and takes years to master, but it is possible to make many useful and decorative items with basic knowledge and a little practice. If you wish to move beyond the projects outlined in this book, it is a good idea to talk with a welder and have him or her evaluate some of your welds. Remember the safety of others is involved when you choose to make a utility trailer or spiral staircase. Take the time and make the effort to ensure that any project you make is safe.

This book is intended as a reference for people who have had some exposure to welding and need reminders about what steps to follow and precautions to take. It is not intended to teach welding to someone who has never handled welding equipment. If you wish to further your welding skills, many community colleges, technical schools, and art centers offer welding classes. Such classes are an ideal way to learn the basics of welding, proper techniques for each process, and welding safety.

Basics

The costs involved in setting up a well-equipped home welding shop are comparable to setting up a well-equipped woodworking shop. A great advantage of welding is that the materials are generally inexpensive, yet you can create items that sell in shops for hundreds of dollars and make repairs that would carry a considerable cost. Because fewer people are knowledgeable about welding, there is always a rewarding “awe” factor to showing off what you have created.

IN THIS CHAPTER:

Safety

Setting Up Shop

Shop Tools

Metal Basics

Metal Shapes & Sizes

Purchasing Metal

Repairing Metal

Cleaning Metal

Safety

When safety measures are overlooked or ignored welders can encounter such dangers as electric shock, overexposure to fumes and gases, arc radiation, and fire and explosion; which may result in serious or even fatal injuries. Preparations for welding—grinding, burnishing, and sawing—are also potentially dangerous. Familiarize yourself with the American National Standard Safety in Welding and Cutting practices (see Resources, page 220).

Always read, understand, and follow manufacturers' material safety data sheets (MSDS), which include specific instructions and recommendations for equipment and product use. Also follow general welding safety rules.

Fumes. Welding produces potentially hazardous fumes and gases. Always keep your face out of the weld plume and avoid breathing concentrations of those fumes. Welding indoors requires special precautions when considering ventilation and may require a ventilation fan, exhaust hood, or fume extractor. If you cannot provide adequate ventilation, an OSHA-approved respirator or particle mask may be needed. If you may be pregnant or plan to become pregnant, consult your physician.

Burns. Hot sparks and flying slag can cause burns. Protect your hands, head, and body with natural fibers—leather, wool, or cotton. Synthetic fibers such as nylon and polyester melt when ignited, which causes serious burns. Wear leather slip-on boots and pants that fall over the top of the boots. Cuffed pants, frayed edges, and pants with holes are fire hazards. Wear a welding cap and keep hair tucked away.

Arc burn. Welding arcs produce ultraviolet and infrared light. Both of these can damage your eyes permanently, burn your skin, and potentially lead to skin cancer. A welding helmet with a filter lens protects your eyes and face, while long-sleeve shirts and long pants protect your skin. Remember to use welding screens and have an extra helmet on hand for observers.

Fire. Remove flammable items such as lumber, rags, dropcloths, cigarette lighters, and matches from the work area. Do not grind or weld in a sawdust-filled shop—the sparks can ignite airborne dust and fumes or ignite flammable materials. Mount an ABC-rated fire extinguisher and first aid kit in the work area. Periodically check your welding area up to one half hour after welding to make sure no sparks have found a place to smolder.

Explosion. Non-flammable gases such as carbon dioxide are stored in high-pressure cylinders that can be dangerous. If regulators are not installed, keep the cylinder's protective cap on and keep cylinders chained or strapped. Never use cylinders as rollers or supports, and never weld protective caps to anything. Shut acetylene and oxygen cylinder valves if you’re away for more than 10 minutes. Cylinders must be transported right side up and chained, even if empty. Similarly, never weld or cut on any closed container, tank, or cylinder. Bring tanks to a welding specialist for evaluation and repair. Another explosion hazard is concrete. Tack welding is acceptable, but using a concrete surface for welds is dangerous due to the amount of water contained in concrete. Use fire bricks instead.

FIRST AID
Safety Manual

Welding safety includes protecting yourself with protective gear, following manufacturer's instructions, and being aware of the surroundings in which you are welding.

Electric Shock. Be sure to insulate yourself from the work piece and ground using dry insulation. Always wear dry, hole-free gloves. Do not touch electrically “hot” parts or electrode with bare skin or wet clothing. Only use electrode holders and cable insulation in good condition.

Other hazards include noise from grinding, sawing, sanding, and plasma cutting; laceration from sharp metal edges; electrocution during arc welding; and asphyxiation from inert shielding gases. Carefully read all manufacturers’ instructions before starting any welding process.

Minimum Lens Shade Numbers for Welding

APPLICATION	SUGGESTED SHADE
Shielded Metal Arc Welding (SMAW)	
1⁄16 to 5⁄32" electrodes	10
3⁄16 to ¼" electrodes	12
5⁄16 to ⅜" electrodes	14
Gas Metal Arc Welding (GMAW)	
1⁄16 to 5⁄32" non-ferrous	11
1⁄16 to 5⁄32" ferrous	12
Gas Tungsten Arc Welding (GTAW)	10 to 14
Plasma Cutting	8
Oxyacetylene Welding	5
Oxyacetylene Cutting	5
Brazing	3 to 5

Welding helmets are typically available with filter lenses in either 2" × 4¼" or 4½" × 5¼" sizes. Helmets with auto-darkening lenses are also available. Clear, full-face protective shields are available for grinding or chipping, and with a #5 filter for oxyacetylene operations.

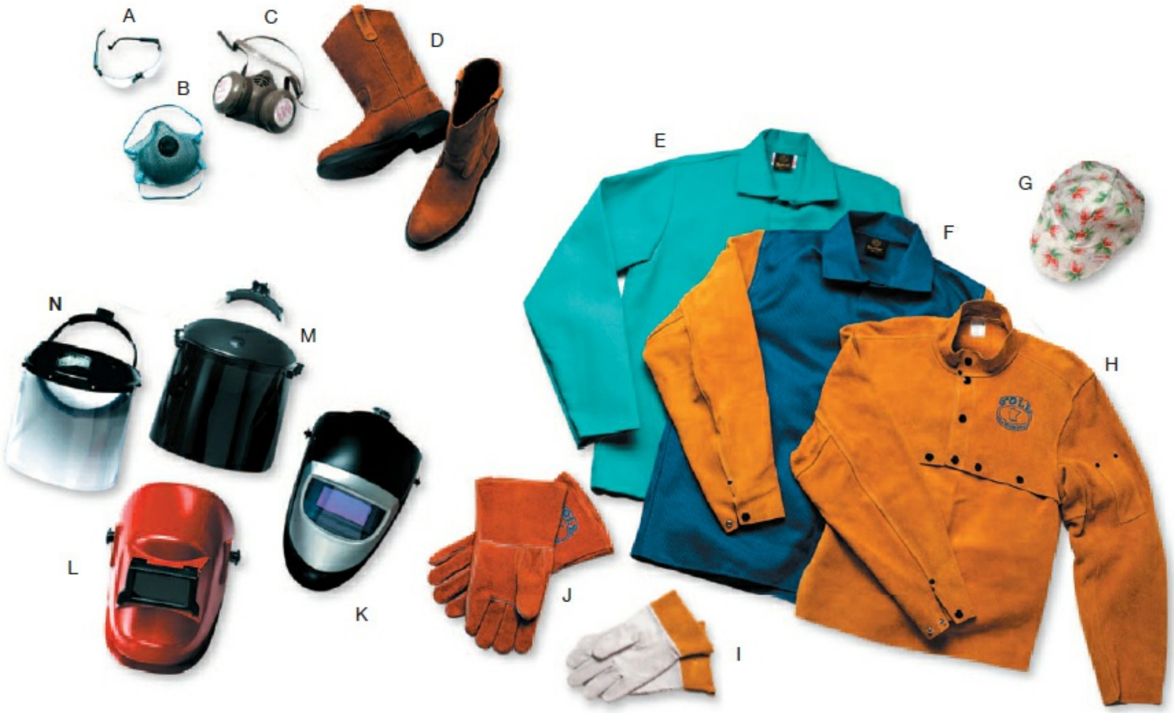

Welding safety equipment includes: (A) safety glasses, (B) particle mask, (C) low-profile respirator, (D) leather slip-on boots,(E) fire-retardant jacket, (F) fire-retardant jacket with leather sleeves, (G) welding cap, (H) leather cape with apron, (I) leather gloves with gauntlets, (J) heavy-duty welding gloves, (K) welding helmet with auto-darkening lens, (L) welding helmet with flip-up lens, (M) full-face #5 filter, (N) full-face clear protective shield.

Setting Up Shop

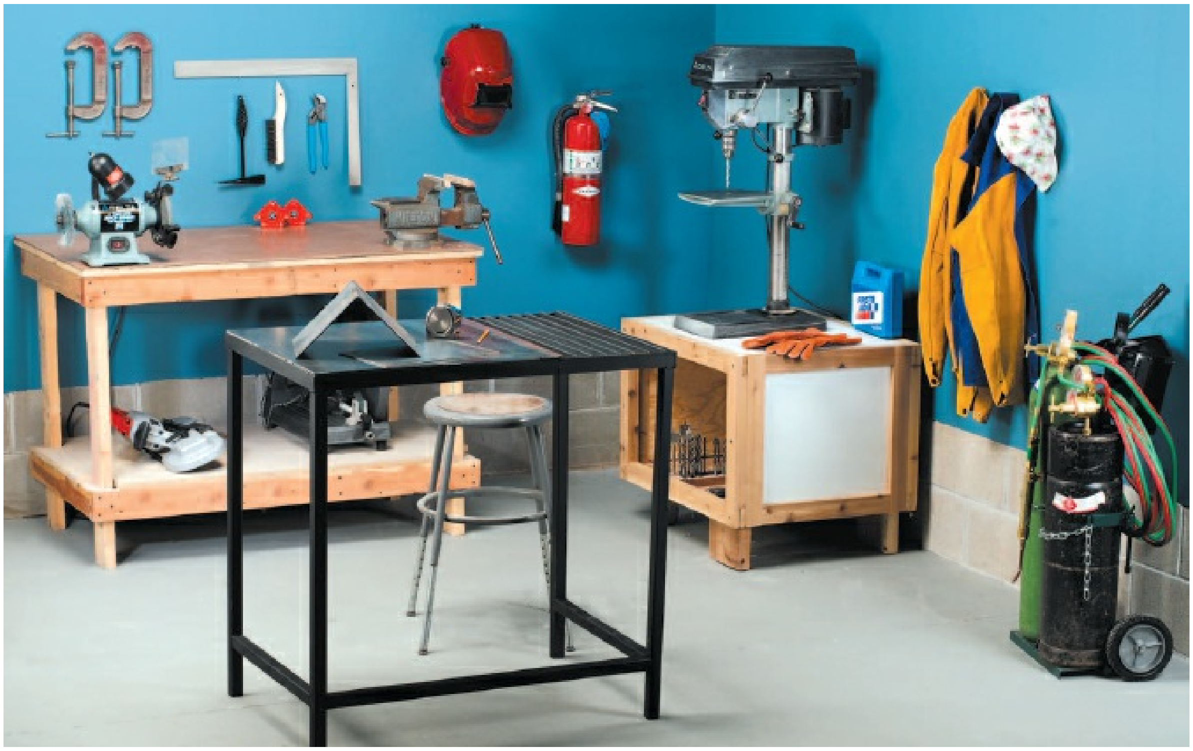

Shop space with a concrete floor and cement block walls is ideal for welding. Good ventilation is also important.

If you plan to weld on a regular basis, it makes sense to set up a welding shop. The primary concern with welding is containing the hot sparks or slag and the flammable elements while exhausting dangerous fumes. It is possible to weld outside, but not all processes allow that. Gas metal and gas tungsten arc welding require that the surrounding air be still so the shielding gas is not disturbed. All the arc processes must be done in dry conditions to prevent electrical shock. Cold metals do not respond as well as metals at 70°F and may not be weldable using certain processes. A heated garage or outbuilding is best suited for a welding shop. A basement is not suitable due to the dangers of fire and compressed gases next to living spaces. Additionally, your homeowner’s insurance may not cover a welding shop if it is inside your living area instead of in a detached garage or shop building.

It is important to remember that tiny sparks and pieces of hot slag may scatter up to 30 or 40 feet from the source. If they come to rest on flammable materials, they may smolder, and given the right conditions they can ignite the material. Always check your welding area one half hour after you have completed welding to make certain no sparks are smoldering. A wooden table covered with metal is not a good work surface, as the transferred heat may cause the wood to smolder. Any wooden jigs or clamping devices should be doused with water to extinguish smoldering embers or stored outside after use.

Shop Tools

Power tools such as the reciprocating saw, angle grinder, portable band saw, and chop saw are useful when cutting and fitting metal parts to be welded. A drill press and metal cutting band saw are also useful tools for welding.

Specific metal-working tools, including metal brakes and metal benders, are available for all sizes of metal. These can range from inexpensive sheet metal tools, such as the scroll bender shown on page 37, to expensive hydraulic tools.

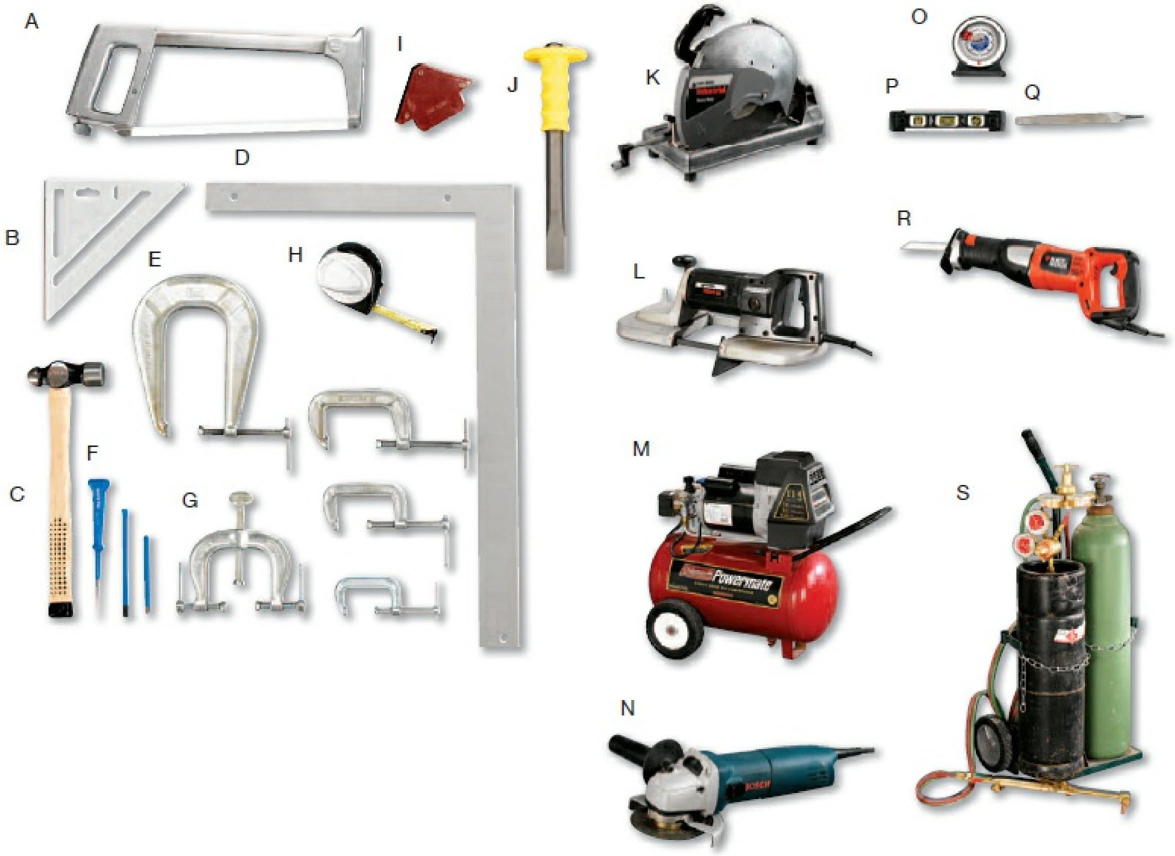

Standard tools for a welding shop include: (A) hacksaw, (B) combination square, (C) ball peen hammer, (D) framing square, (E) various sizes of C-clamps, (F) center punches, (G) alternative C-clamp, (H) measuring tape, (I) magnetic clamp, (J) cold chisel, (K) circular saw, (L) portable band saw, (M) air compressor, (N) angle grinder, (O) angle pointer/calculator, (P) level, (Q) metal file, (R) reciprocating saw, (S) oxyacetylene and oxygen (including regulators, hoses, torch with tip).

Standard tools for a welding shop include: (T) plasma cutter, (U) TIG stick welder, (V) multiple-purpose cut-off saw, (W) MIG welder.

Metal Basics

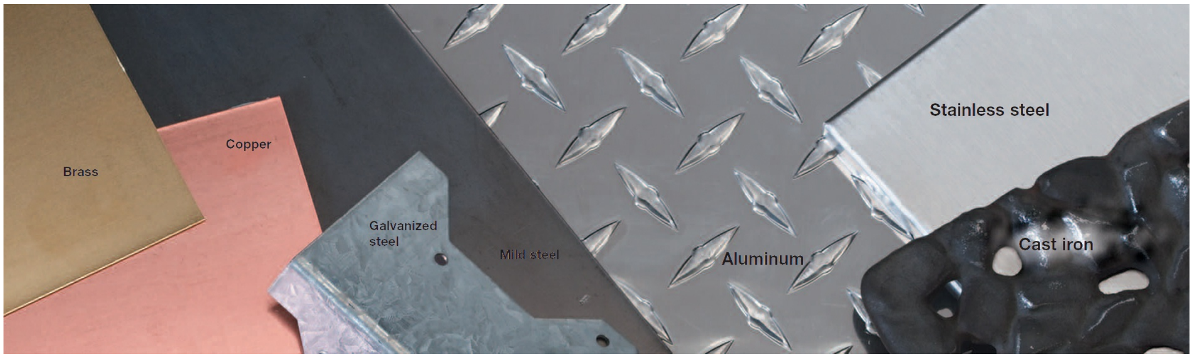

Different metals have different characteristics that affect their ability to be welded or cut with any of the processes described in this book. In general, only metals of the same type are welded together because welding involves melting the base metal parts and adding melted filler metal. In order to accomplish this, the parts and filler must have the same melting temperature and characteristics. Dissimilar metals can be joined by brazing and braze welding because these processes do not actually melt the base metal.

Metals are divided into the categories of ferrous and non-ferrous. Ferrous metals contain iron and will generally be attracted to a magnet. Cast iron, forged steel, mild steel, and stainless steel all contain iron, along with varying amounts of carbon and other alloying elements.

Mild steel, which has a low carbon content, is the most commonly used type of steel and the easiest with which to work. Mild steel can be welded with all the processes covered in this book and can be cut with either oxyfuel or plasma cutting. It makes up most of the metal items you commonly use, make, or repair, such as automobile bodies, bicycles, railings, patio furniture, file cabinets, and shelving. Adding more carbon to steel makes it harder, more brittle, and more difficult to cut and weld. High- carbon steel is called tool steel and is used to make drill bits and other metal cutting tools.

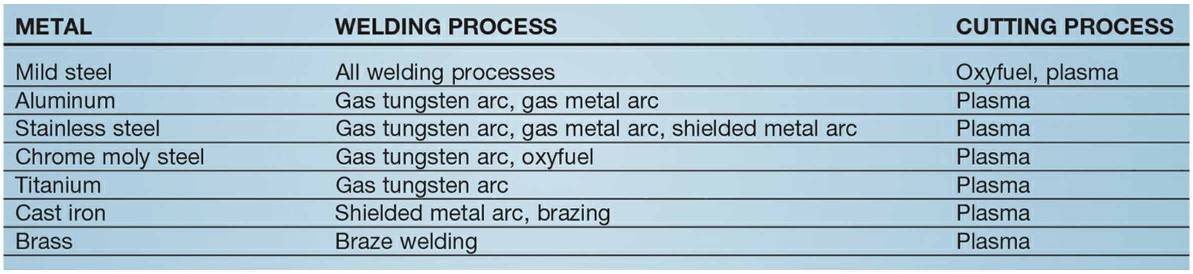

METAL	WELDING PROCESS	CUTTING PROCESS
Mild steel	All welding processes	Oxyfuel, plasma
Aluminum	Gas tungsten arc, gas metal arc	Plasma
Stainless steel	Gas tungsten arc, gas metal arc, shielded metal arc	Plasma
Chrome moly steel	Gas tungsten arc, oxyfuel	Plasma
Titanium	Gas tungsten arc	Plasma
Cast iron	Shielded metal arc, brazing	Plasma
Brass	Braze welding	Plasma

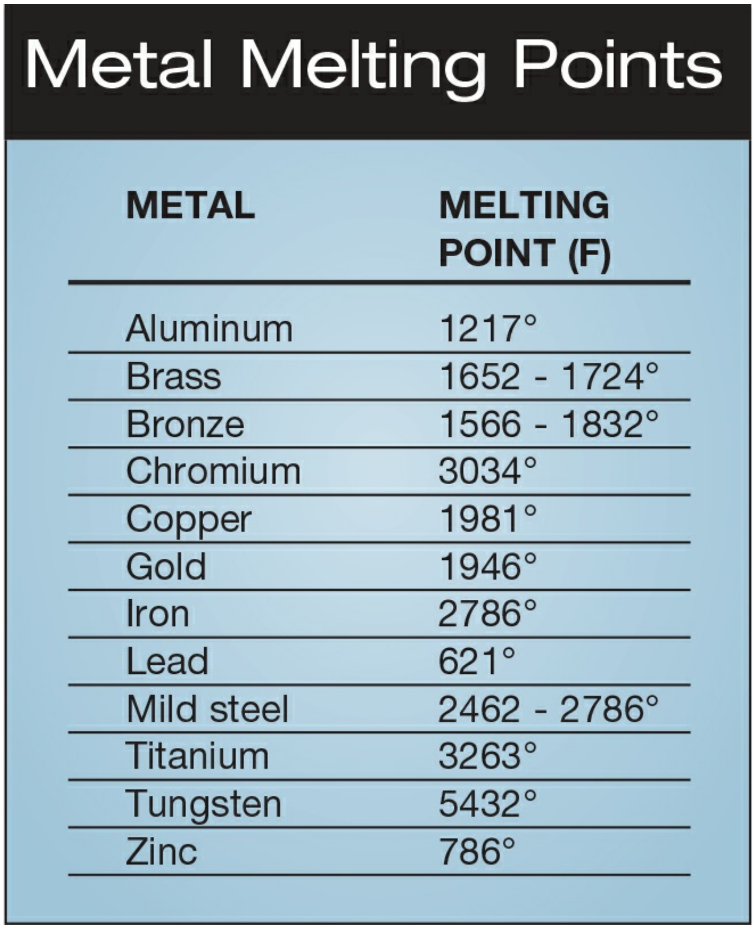

Metal Melting Points

METAL	MELTING POINT (F)
Aluminum	1217°
Brass	1652 - 1724°
Bronze	1566 - 1832°
Chromium	3034°
Copper	1981°
Gold	1946°
Iron	2786°
Lead	621°
Mild steel	2462 - 2786°
Titanium	3263°
Tungsten	5432°
Zinc	786°

Other elements are added to steel to make a variety of alloys, such as chrome moly steel and stainless steel. These additions may make the steel non-magnetic. Because stainless steel does not oxidize (rust) easily, it is difficult to cut with oxyfuel. When welding steel alloys, the filler material must be matched to the alloy to get a high-quality weld.

In addition to alloying, many metals are heat treated to improve characteristics such as hardness. Because welding and cutting would reheat these metals and destroy those characteristics, heat treating is an important factor to consider. Many vehicle chassis parts, including motorcycles and bicycles, are made of heat-treated metals.

Aluminum is a widely used non-ferrous metal because of its light weight and corrosion resistance. Like steel, it is available in many alloys and is often heat treated. Aluminum is used for engine parts, boats, bicycles, furniture, and kitchenware. Various characteristics make aluminum difficult to weld successfully—it does not change color when it melts, it conducts heat rapidly, and it immediately develops an oxidized layer.

Metal Shapes & Sizes

Mild steel (and most other metals) come in a variety of shapes, sizes, and thicknesses. Metal thickness may be given as a fraction of an inch, a decimal, or a gauge (see chart, page 18).

Rectangular tube (A) and **square tube** (B) are used for structural framing, trailers, and furniture. Dimensions for rectangular and square tubing are given as width × height × wall thickness × length.

Rail cap(C) is used for making handrails. Rail cap dimensions are the overall width and the widths of the channels on the underside.

Channel (D) is often used for making handrails. Very large channels are used for truck bodies. The legs, or flanges, make it stronger than flat bars. Dimensions for channel are given as flange thickness × flange height × channel (outside) height × length.

Round tube (E) is not the same as pipe. Round tube is used for structures, and pipe is used for carrying liquids or gases. Dimensions for a round tube are given as outside diameter (O.D.) × wall thickness × length. Pipe dimensions are nominal, that is, in name only. They are given as nominal inside diameter (I.D.) × length.

T-bar (F) dimensions are given as width × height × thickness of flanges × length.

Angle or **angle iron** (G) has many structural and decorative uses. Dimensions for angle iron are flange thickness × flange width × flange height × length.

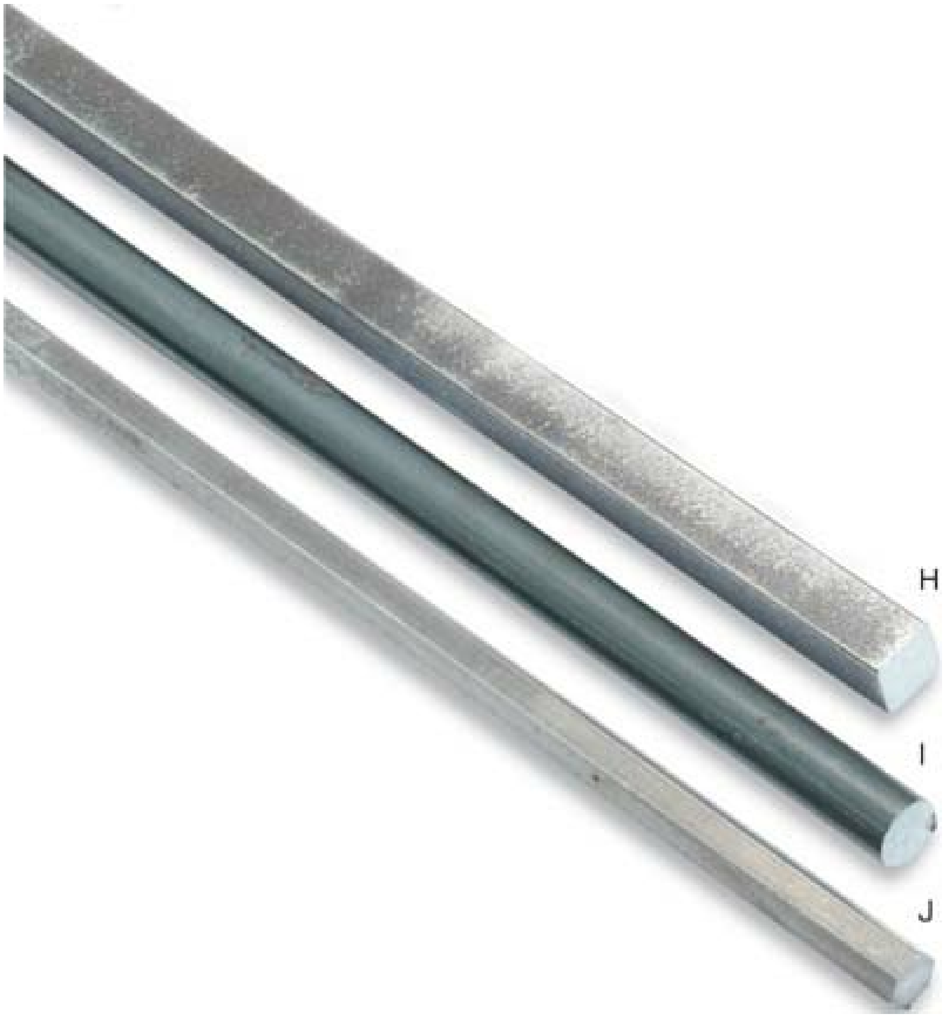

Square bar (H), **round bar** (I), and **hexagonal** or **hex bar** (J) dimensions are given as width or outside diameter × length.

Flat bar or **strap** (not pictured) is available in many sizes. Dimensions are given as thickness × width × length.

Sheet metal (not pictured) is $^3/_{16}$" or less in thickness and is often referred to by gauge. **Plate metal** (not pictured) is more than $^3/_{16}$" thick and is referred to by fractions or inches.

Diamond plate is used as a heavy-duty siding for toolboxes and carts because of its durability and strength. It is also used for decorative applications. It is sold as length × width × thickness.

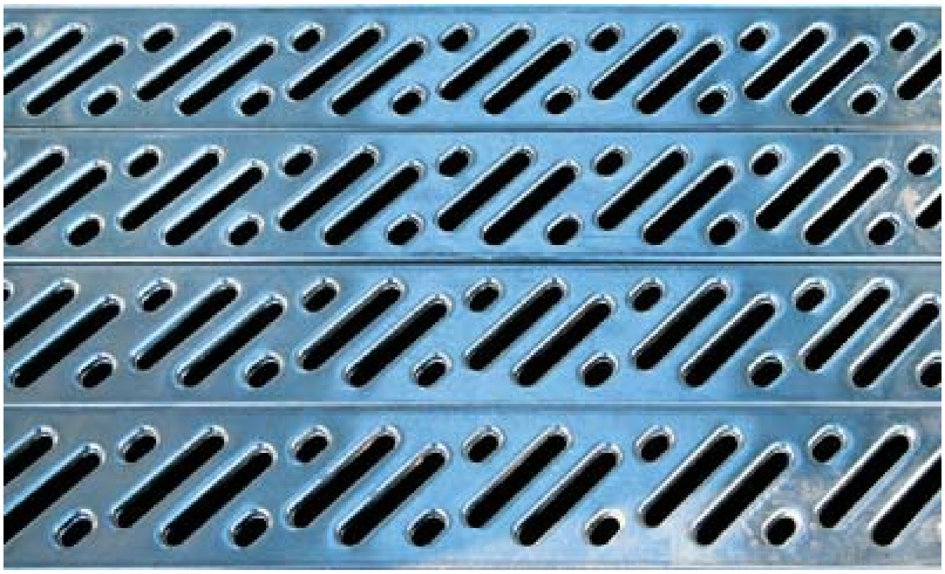

Perforated sheet metal is used for industrial shelving, fences, stair treads, grates for floor drains, and a host of decorative sidings and top treatments. There are numerous shapes and styles available. They are sometimes referred to as perforated metal screens. Perforated sheets are sold as length × width × thickness. Lighter-weight sheets are sold by the yard in a roll.

Metal Identification Chart

Test	Appearance	Fracture	Magnetic	Torch	Chip	Spark	Volume of stream
Manganese Steel	Dull cast finish	Rough grained	Non-magnetic	Turns bright red, melts quickly	Hard to chip	Bright white bursts, heavy pattern	Moderately large
Stainless Steel	Bright, smooth surface lines	Bright appearance	Depends on exact composition	Turns bright red, melts quickly	Smooth chip, smooth bright color	Very few short full red sparks with few forks	Moderate
Low Carbon Steel (mild steel <.30% carbon)	Gray, fine	Gray, bright crystalline	Highly magnetic	Gives off sparks when melted, pool solidifies rapidly	Chips easily, smooth and long chip	Long white sparks, some forks near end of stream	Moderately large
Medium Carbon Steel (.30 to .45% carbon)	Gray finish	Light gray	Highly magnetic	Melts quickly, gives off some sparks	Chips easily, smooth and long chip	Long white sparks with secondary bursts along stream	Moderately large
High Carbon Steel (>.45% carbon)	Dark gray, smooth finish	Light gray-white, finer grained than low carbon steel	Highly magnetic	Melts quickly, molten metal is brighter than low carbon steel	Difficult to chip, brittle	Large volume of brilliant white sparks	Moderate
Wrought Iron	Gray, fine surface lines	Fibrous structure, split in same direction of fibers	Highly magnetic	Melts quickly, slight tendency to spark	Chips easily, continuous chip	Straw-colored sparks near wheel, few white forks near stream end	Large
Cast Iron	Rough, very dull gray	Brittle gray	Highly magnetic	Turns dull red, first puddle is very fluid, no sparks	Very small and brittle chips	Dull red sparks formed close to wheel	Small
High Sulfur Steel	Dark Gray	Gray, very fine grain	Highly magnetic	Melts quickly, turns bright red before melting	Chips easily, smooth and long chips	Bright carrier lines with cigar-shaped swells	Large

Purchasing Metal

Finding a metal supplier can be a challenging task. The materials that are readily available at home centers and hardware stores may not be the sizes and shapes needed for a project and tend to be expensive. Because steel is so heavy, Internet or catalog shopping entails significant shipping costs. With some searching, however, most materials can be found.

Metal is generally priced by weight, unless you are purchasing it at a retail store. The price for small pieces of mild steel at a home center or hardware store work out to be as much as $3 to $5 per pound. The price per pound for small orders at a steel dealership may be in the range of $1. Large orders or repeat orders may be priced as low as 50 cents per pound. Many metal suppliers have an odds and ends bin or rack where pieces may be as low as 10 cents per pound. Specialty metals such as stainless steel and aluminum start at $2 to $3 per pound.

Smaller sizes and shapes of mild steel and aluminum are available at home improvement and hardware stores in three-, four-, and six-foot lengths. Welding supply stores often have a selection of ten- and twelve-foot lengths of the commonly used sizes. Steel dealers have most common sizes in stock and will order other sizes for you. Most mild steel shapes and sizes come in twenty-foot lengths, and many steel dealers will make one free courtesy cut per piece. You might want to have the metal delivered, depending on the amount of material you are purchasing.

Some steel dealers are distributors for decorative metal products, but many specialty items such as wrought iron railing materials, decorative accessories, and weldable hardware are only available by catalog. A number of catalog supply houses sell to the public and have varied selections and reasonable prices. (See Resources, page 220.)

Inch Equivalent for Gauge Thickness

GAUGE	INCHES
24	0.020
22	0.026
20	0.032
18	0.043
16	0.054
14	0.069
12	0.098
11	0.113
10	0.128

Metal less than ⅛" thick is often referred to by gauge. For reference, the decimal equivalent of ⅛" is 0.125.

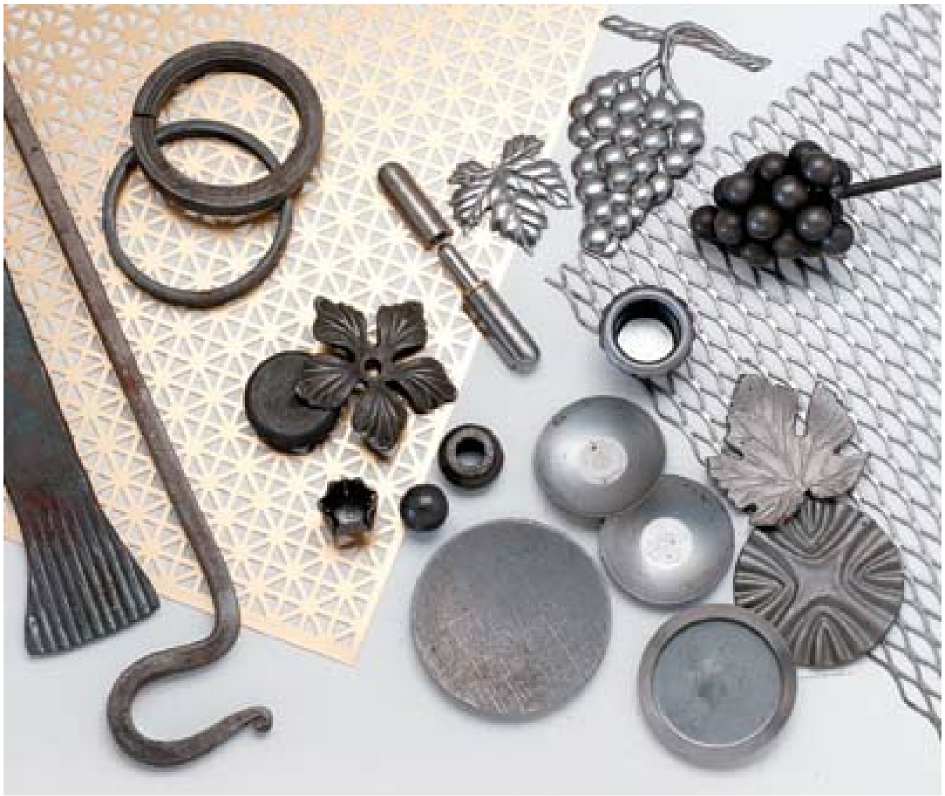

Wall plates, hooks, rings, balls, bushings, candle cups, drip plates, and stamped or cast items are available in a wide variety of shapes, sizes, and sheets.

Repairing Metal

Once you begin welding, you will encounter numerous opportunities to repair items. When your friends and neighbors discover you can do repairs, even more challenges will come your way.

Performing welded repairs can be very tricky. It is important to assess your welding skills, the difficulty of the repair, and the intended use of the repaired item. Any structural or vehicle repairs, such as stairways, ladders, trailers, or truck chassis, need to meet the same safety standards as they did in their original condition.

The first step in considering a repair is determining why the item broke. If a weld was poorly executed, the repair might simply be to prepare the area and perform a good weld. If a piece has broken due to metal fatigue, simply patching over the crack will only cause more cracks outside the patched area. Cast iron and cast aluminum may have broken due to imperfections or inclusions in the casting, or they may have gone through rapid heating and cooling that caused cracking.

The next step in a repair is determining the base metal. A magnet will be attracted to any metal with a certain concentration of iron in it, but stainless steel (which is sometimes non-magnetic), mild steel, and cast iron require very different welding techniques. Aluminum is non-magnetic and is discernable from stainless steel by its light weight—but which alloy is it? In the case of aluminum, some alloys are not weldable. Unfortunately, many manufactured metal items are made of alloys and may have been heat treated. Without access to the manufacturer’s specifications, it might be impossible to determine what the base metal composition is. It is important to understand the effects of welding on these materials before attempting a welded repair.

Once you have determined the feasibility of a repair on a particular piece, the piece must be prepared carefully. All rust, paint, and finishes must be removed from the area of the weld, and, if you are arc welding, from an area for the work clamp. Grease and oil must be cleaned off. If the break is on a weld joint, the old weld bead needs to be ground down.

Mild steel is the easiest material to repair. Simply prepare the material as for any other weld, and weld using any of the welding processes.

All parts need to be carefully prepared before attempting a repair. Paint needs to be ground or sanded off and grease and oil need to be cleaned away. This cast iron part is being beveled to allow greater weld penetration

Cast iron and cast aluminum need to be preheated before welding to prevent cracking due to temperature fluctuations. If the piece is small enough, you can put it in the oven and heat it to 400 to 500°F. Otherwise, use an oxyacetylene or propane rosebud tip. Temperature crayons, which melt at specific temperatures, are available for marking metal to be preheated. Cast iron can be brazed if the fit between the broken parts is good, or it can be braze welded or welded with shielded metal arc. Specific electrodes for cast iron are available. Cast aluminum can be brazed, braze welded, or welded with gas tungsten or gas metal arc welding.

Some aluminum is not weldable. If the piece to be repaired has been welded, you can perform a welded repair. If you can determine the composition of aluminum parts, match the electrode or filler rod to that. Some filler rods and electrodes are multipurpose and can be used on more than one alloy. Gas tungsten arc welding is the best choice for aluminum, but it can also be welded with gas metal arc, or it can be brazed.

Stainless steel comes in numerous alloys, and it is important to match the filler material to the base metal. Do not clean stainless steel with a steel wire brush, as the mild steel wires may contaminate the stainless steel. Stainless steel is best repaired with gas tungsten arc welding, although it can be brazed as well.

Remove the paint and finish from the weld area. If you are arc welding, also remove the paint from an area close to the weld for attaching the work clamp.

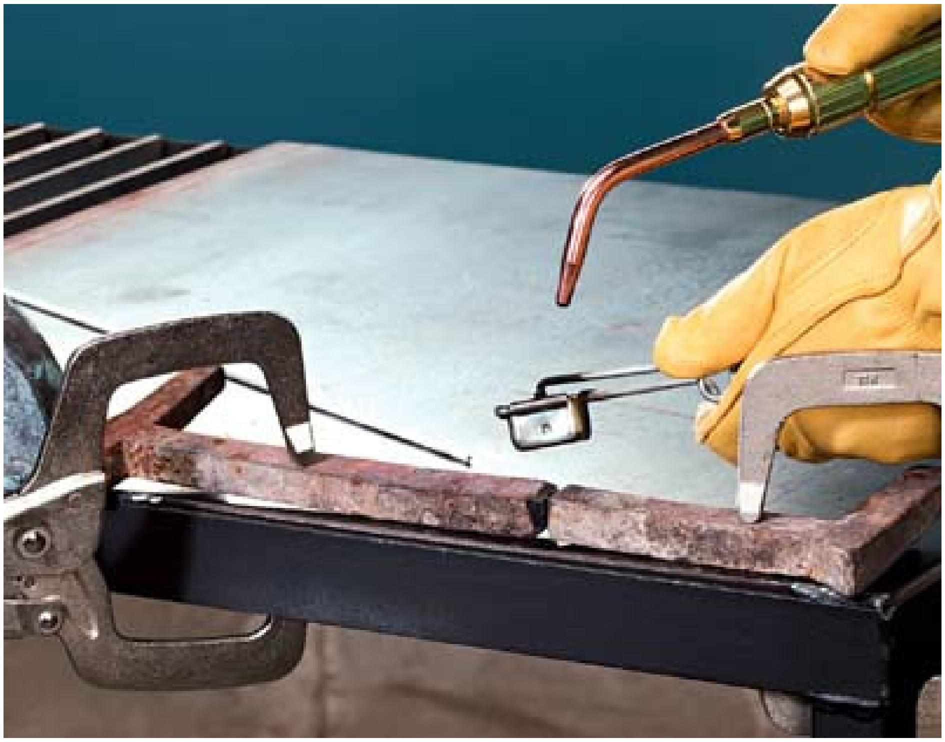

Cast iron can be braze welded as shown here, or shielded metal arc welded with cast iron electrodes. Either way, the metal needs to be preheated to 450° F before being welded.

Cleaning Metal

A successful weld begins with a well-prepared piece of metal. The cleaner the pieces to be welded, the better the weld quality and appearance. Steel is not available in a pre-cleaned state, so you will need to clean all project parts by removing oil, mill scale, dirt, and rust. Although this can be a tedious step, a well-cleaned project lasts many years and you soon forget the time spent conscientiously cleaning every piece.

For any project that will be painted, it is important to thoroughly clean the entire part. For objects that will be allowed to rust gracefully outdoors, thoroughly clean the joint areas and allow the weather to clean the remainder.

The first step in cleaning is to wipe down the part with denatured alcohol, acetone, or a commercial degreaser to remove oil and dirt. The alcohol works best as it has little odor, unlike the degreaser, and it won’t dissolve plastics, like the acetone will. Both acetone and degreasers tend to leave a residue that may diminish the quality of the final finish.

Once the grease, oil, and dirt have been removed from the surface, the mill scale (often found on mild steel) must be removed. Mill scale is a dark gray flaky oxide layer that forms on the steel as it cools. Cold-rolled steel often has a thinner layer of scale than hot-rolled steel, but even the thin layer needs to be removed before welding and finishing. Manufacturers use a pickling bath to remove scale; except for very small projects, it is unlikely that a home do-it-yourselfer would want to work with the hot acid bath needed for pickling. The remaining options are wire brushing, grinding, sanding, or sandblasting.

A bench-mounted power wire brush is perfect for cleaning the ends of parts. Unfortunately, it is not suitable for cleaning the remaining surfaces. For this you will need an angle grinder outfitted with a wire brush cup or flap sander. A drill or rotary tool with a wire brush wheel, wire brush cup, or flap sander also works well. When working with power wire brushes it is crucial you wear appropriate protective gear. A full face mask and long sleeves protect your eyes and arms from wire fragments that are thrown from the brush. Heavy leather gloves will protect your fingertips from getting skinned. These tools operate at very high speeds and can quickly do serious damage to eyes and skin.

Once the metal has been cleaned and de-scaled, it is important that you complete the project in a timely manner before the metal parts.

S-L-X
Denatured
Alcohol

Apply denatured alcohol with a clean rag to clean dirt and oil from project parts. Wear rubber gloves to protect your hands.

DEL
6" THIN-
BENCH GR
WITH FLEXIB

Cleaning the mill scale off mild steel is an important step. A bench-mounted power wire brush works well on small pieces. Wire brushing the entire project prior to finishing is critical for good paint adhesion.

Techniques

In this chapter, all of the basic welding skills are covered—from cleaning to cutting and jointing to jigging. These are the techniques that you will use again and again regardless of the type of welding you pursue. This chapter also provides basic directions and step-by-step photos that illustrate major welding and cutting processes. You will find quick reference charts that describe electrode and filler wire choices, metal types and weldability, and joint and weld types.

IN THIS CHAPTER:

Mechanical Cutting
Plasma Cutting
Oxyacetylene Cutting
Shaping
Finishing
Oxyacetylene Welding
Shielded Metal Arc Welding (SMAW)
Gas Metal Arc Welding (GMAW/MIG)
Gas Tungsten Arc Welding (GTAW/TIG)

Mechanical Cutting

For some projects, cutting the metal is the most difficult step. A variety of tools exist for cutting metal, but the thicker and larger the stock, the fewer your choices. If you want to build muscle, a hacksaw is great for tubes, rods, and bars, but a large project will soon have you wishing for some more powerful cutters. Cutting thin-gauge sheet metal can be done with sheet metal snips, but thicker gauges and plates require shears or cutting torches. You can spend a few hundred dollars or a few thousand dollars on power metal cutting tools. Here are some of your options.

Bimetal saw blades are the best for cutting metal, whether you are using them in your hacksaw or bandsaw. Make the blades in your hacksaw and portable bandsaw last longer not by forcing the cuts, but by allowing the saw to cut slowly and steadily. Excess pressure simply wears the blades out before their time.

A horizontal metal-cutting bandsaw is a bench-mounted bandsaw with clamps to hold work pieces, and an automatic shut-off feature that turns off the saw when the cut is completed. The most common cutting capacity is 4 × 6", which can cut through a rectangular piece of that dimension or a 4½" round. A portable metal-cutting bandsaw is slightly less expensive than the bench version and obviously more portable. The most common size cuts 4" stock. It is more difficult to create accurate cuts with the portable bandsaw, but the portability makes the machine more versatile in a do-it-yourself shop. A distinct advantage of the bandsaws is the very small kerf (less than $^1/_{16}$").

Metal-cutting chop saws, cut-off machines, and angle grinders use abrasive wheels for cutting. Most of these produce a great deal of heat and metal dust, and cut pieces typically have burred edges. Metal-cutting circular saws with carbide blades produce cleaner cuts with less heat. They are also faster than chop saws. Both of these types of saws have larger kerfs (at least $^3/_{32}$") than the bandsaw. A portable steel saw looks and operates like a standard circular saw, but is capable of cutting up to ¼" sheet metal, as well as cutting tubes and rods. Reciprocating saws can also be used for cutting rods and tubes.

Manual or hydraulic metal shears and punches in small sizes suitable for home shops are relatively inexpensive. They are limited to thin-gauge materials, usually less than 16 gauge. And the heavier the material capacity is, the more expensive the machine.

Manual snips are sufficient for cutting curves in sheet metal thinner than 18 gauge. Power nibblers make quick work of curves but are commonly available for 18-gauge or thicker.

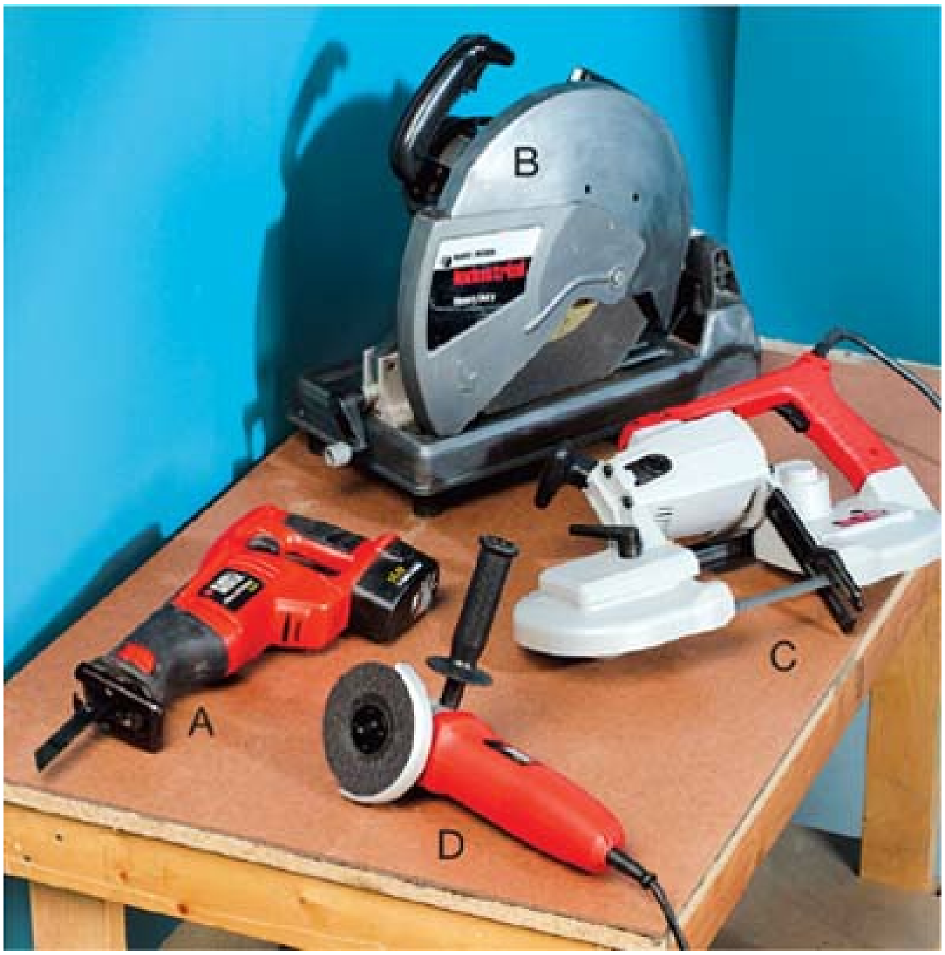

Power tools for a welding shop include: (A) reciprocating saw, (B) cut-off saw, (C) portable band saw, (D) angle grinder.

Plasma Cutting

Plasma cutting is an arc process. Plasma is a gas that can conduct electricity because it is in a highly ionized form. The gas transfers the arc, but because of the constriction of the gas, the gas resistance is very high, which creates the extreme heat for plasma processes. Plasma cutting uses a very constricted arc to melt the base metal while compressed air or other gas blows the melted metal out of the kerf. This process can be used on any conductive metal—aluminum, brass, cast iron, copper, steel, stainless steel, and titanium—which sets it apart from oxyacetylene cutting, which is only useable on oxidizable metals.

Though the plasma arc temperature is 40,000°F, it is so constricted and the cutting speed is so fast that the thermal distortion to the metal being cut is low. Plasma cuts are clean and are often weldable with no additional cleanup if the base metal was initially clean. Properly done, the kerf is fairly straight sided, and no slag is present.

Safety: Plasma cutting produces sparks, smoke, and fumes. When cutting materials that contain chromium and nickel (such as stainless steel) or materials that have been galvanized or zinc coated, many toxins will be present in the fumes. When cutting through painted, bonded, or dirty metals, these surface additions often burn off, producing hazardous smoke. Plasma cutting often produces ozone and various nitrogen oxides that are colorless and can be odorless; both are hazardous. Therefore, good ventilation is important even when plasma cutting.

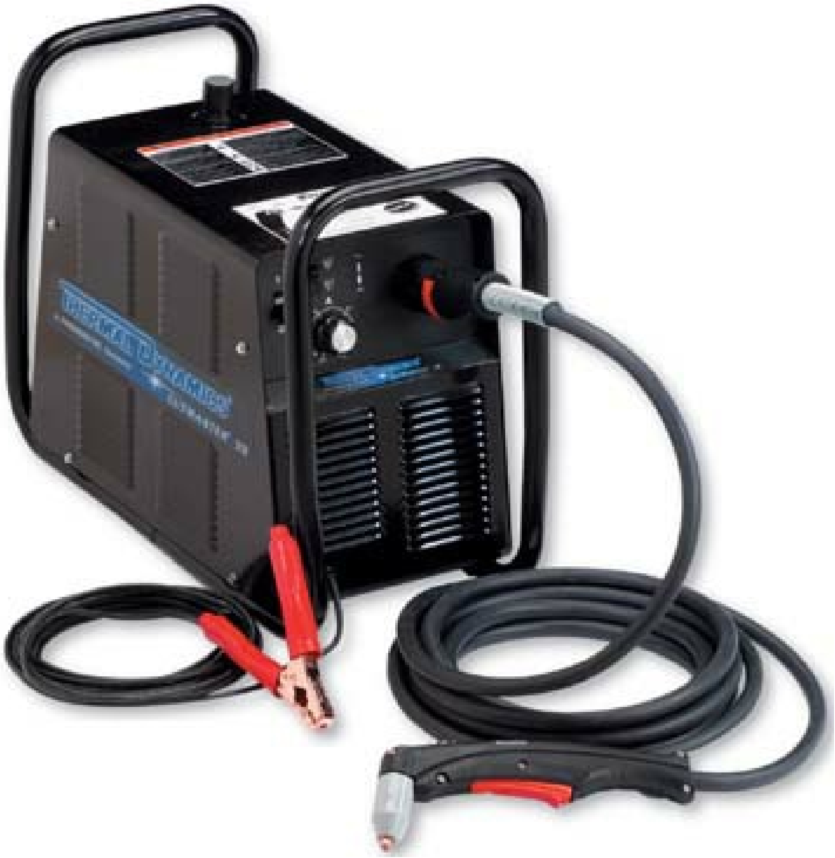

A plasma cutter consists of a power source with a compressed air or inert gas connection, work cable with clamp, supply cable, and torch.

Safety

- Welding helmet with #8 filter
- Leather, wool, or cotton long pants, long-sleeve shirt, and hat
- Heavy-duty welding gloves
- Leather boots or shoes
- Ventilation

Plasma cutting may not seem as bright as other arc processes, but it is. It produces ultraviolet rays, so it is important to protect your skin and eyes. Because of the high level of open circuit voltage, plasma cutting has added electrical shock hazard. Make sure all surfaces around the cutting area are dry, as well as your clothing and gloves. Turn off the power to the machine before changing any torch parts.

Never cut containers, tanks, or cylinders that may have held flammable materials. Even a small amount of flammable residue can cause an enclosed vessel to explode when cut.

Equipment: Plasma cutting machines are expensive, but as technological advances are made and demand increases, more companies are introducing smaller, less expensive home and small shop equipment. If you regularly need to cut stainless steel, aluminum, or the high-strength steel now used in automobile bodies, this machine can be worth the cost. Small, 115-volt plasma cutters are available that will cut up to 3/8" material. Larger 220-volt machines can cut thicker materials and have higher duty cycles. Aluminum and stainless steel require higher amperages, so check whether the machine easily can cut the materials with which you typically work.

Most plasma cutters require a compressed air source. Any shop air compressor that can deliver 65 psi is sufficient. It is critical that a filter be installed to keep the air dry and oil free. Follow the manufacturer’s directions for installing the filter. Some plasma cutters have an internal air compressor. These units often require more maintenance than models that utilize a separate compressor.

Most plasma torches have a manual switch on the torch body. The cutting tip and electrode are consumable parts. Check them frequently for wear and replace them when necessary.

CAMPBELL HAUSFELD

A dehumidifying filter is very important for the compressed air source for plasma cutting.

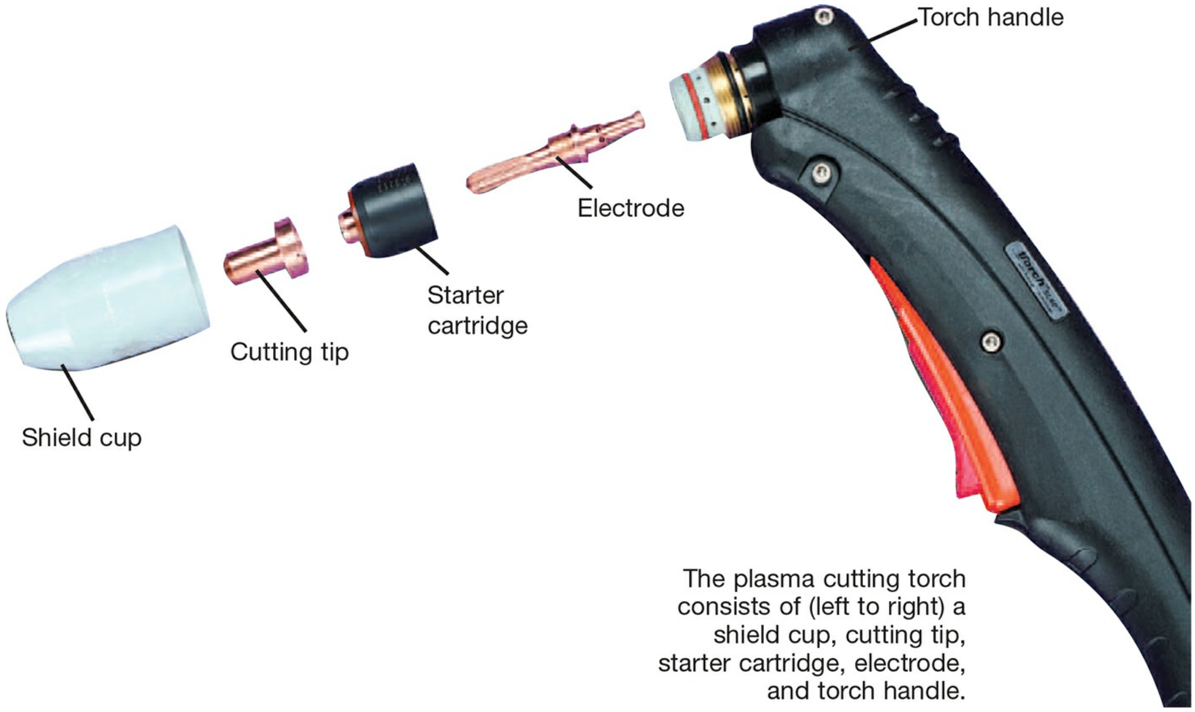
Torch handle
Electrode
Starter cartridge
Cutting tip
Shield cup
The plasma cutting torch consists of (left to right) a shield cup, cutting tip, starter cartridge, electrode, and torch handle.

The plasma cutting torch consists of (left to right) a shield cup, cutting tip, starter cartridge, electrode, and torch handle.

Precutting Checklist

- Compressed air fittings are tightly attached.
- Cables are in good condition.
- Ventilation is sufficient to remove fumes from cutting area.
- Work area is dry.
- Flammable materials are removed from area.
- Sparks and slag from cutting will be contained.

How to Plasma Cut

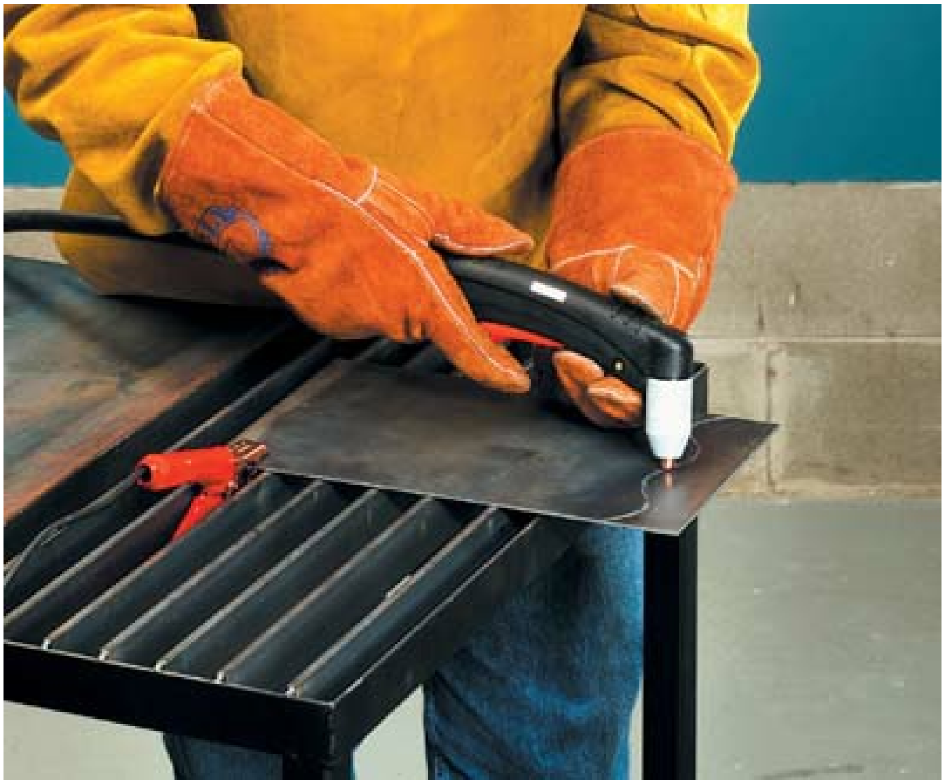

1 Check the manufacturer's recommended settings for the material to be cut. Arrange the material on a cutting table and determine the most comfortable cutting position. Attach the work clamp to the workpiece or welding table. Dry run through the cutting motion to practice speed and cutting position.

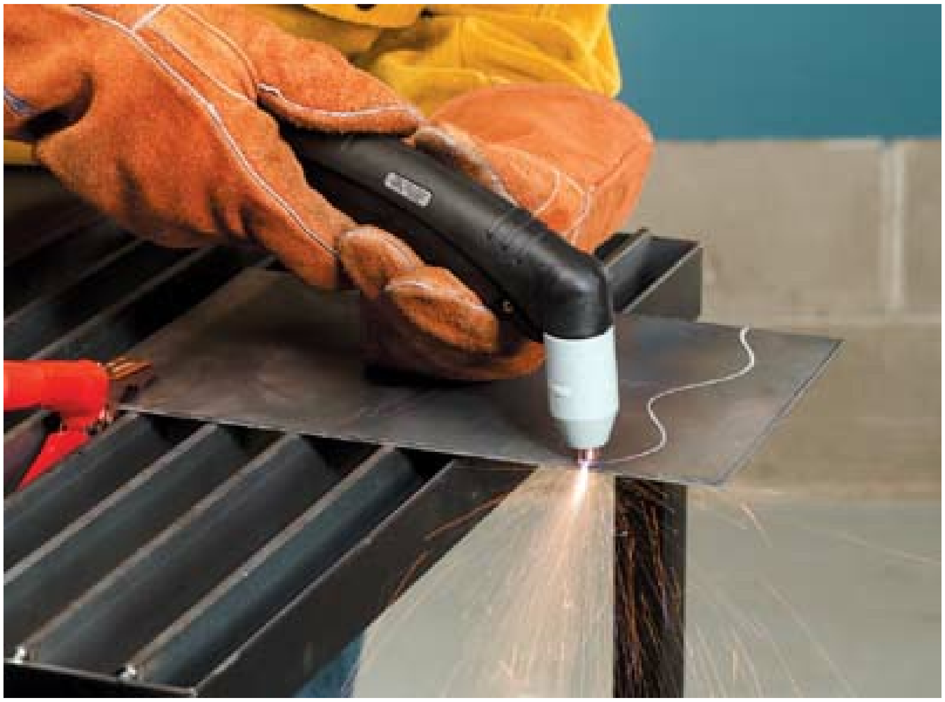

2 Check manufacturer’s directions for arc starting procedure. Most current machines are designed to drag the nozzle directly on the material’s surface. Activate the arc and hold the torch perpendicular to the surface of the material. Move steadily and smoothly along the entire length of the cut. Do not stop the arc until you have passed through the end of the cut.

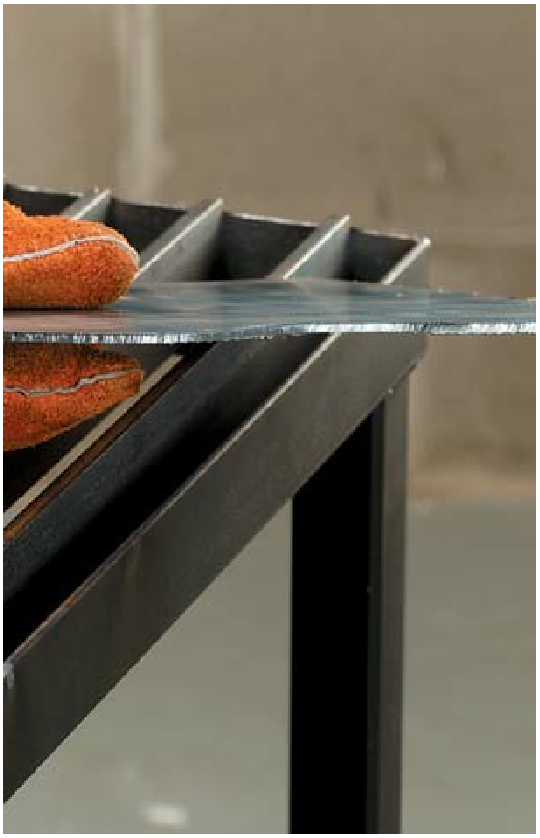

3 A good plasma cut has squared edges, small vertical ripples, and little or no slag.

Troubleshooting & Techniques

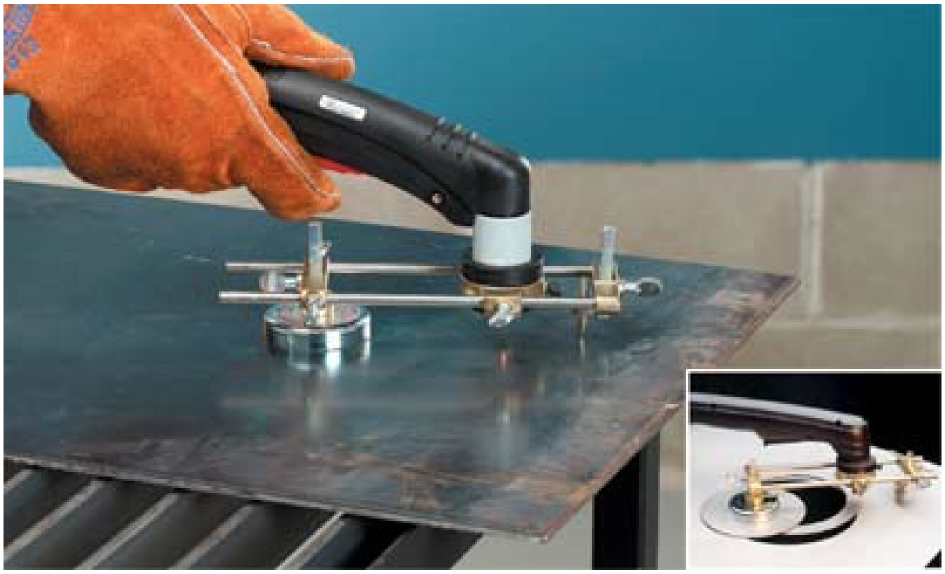

A high-quality circle cutting attachment for a plasma cutter is expensive, but worth the investment if you will be making repeated circle cuts. Inset photo: This unit comes with a magnetic pivot holder for steel and a suction cup pivot holder for all other metals.

Plasma Troubleshooting

PROBLEM	SOLUTIONS
Arc does not start	• Connect work clamp to workpiece • Turn on power source • Tighten cable connections • Make sure electrode is in working condition
Cut does not penetrate through material	• Cut more slowly • Make sure torch is held perpendicular to material • Make sure machine is capable of cutting that thickness • Increase amperage setting • Verify correct air flow and pressure
Excessive slag or dross	• Increase cutting speed • Decrease torch standoff distance • Replace worn torch parts • Adjust amperage setting
Torch parts consumed quickly	• Install or replace air filter to prevent oil or water from reaching torch • Cut metal within the capability of the machine • Make sure gas pressure is correct

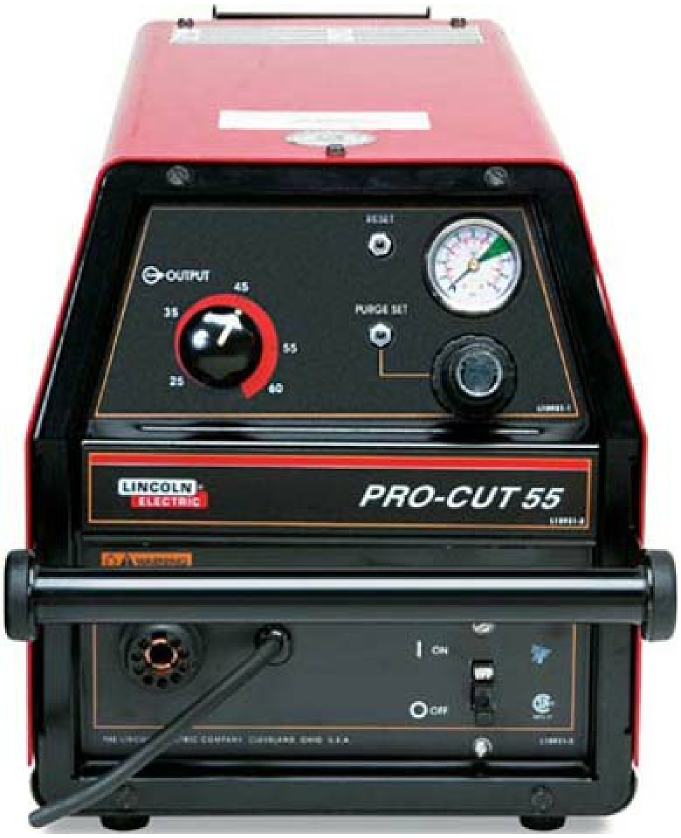
RESET
OUTPUT
35
45
55
25
60
PURGE SET
LINCOLN ELECTRIC
PRO-CUT 55
ON
OFF

The adjustable pressure regulator with gauge is located on the front of the Lincoln Pro-Cut 55 plasma cutting machine. This compact machine generates 60 amps of cutting power capable of cutting conductive materials up to ¾" thick. It has a built-in air filter.

LINCOLN
ELECTRIC

This plastic box contains the consumables for a Lincoln plasma cutter. The consumables are made of copper, which resists sticking to almost all other metals. The parts should be replaced when they become too worn out to work properly.

Tip

Making a good-quality plasma cut is dependent on a consistent travel speed. Consult the owner’s manual for cutting speeds for different material thicknesses and amperage settings. With the machine off, practice moving across the planned cut at the correct speed.

More Plasma Cutting Techniques

This operator can make precise cuts with a plasma cutter without adding the heat input of an oxyacetylene torch. This will prevent heat distortion in surrounding regions of the work area.

Plasma cutters can also pierce small pieces of thick metal, but for a cleaner cut first drill a small 1/8" - hole to start the cut.

Bending & Cutting Diamond Plate

Although diamond plate can be plasma cut, there are times when you'll want to bend the sides instead. Cut a bird's mouth with a plasma cutter (see page 97) and then use a simple break like the one shown here to bend the sheet (see page 101). Bend a little at a time until you get what you want. If you score the backside beforehand it makes the bend easier with less stress, but don't score it deep or it will crack.

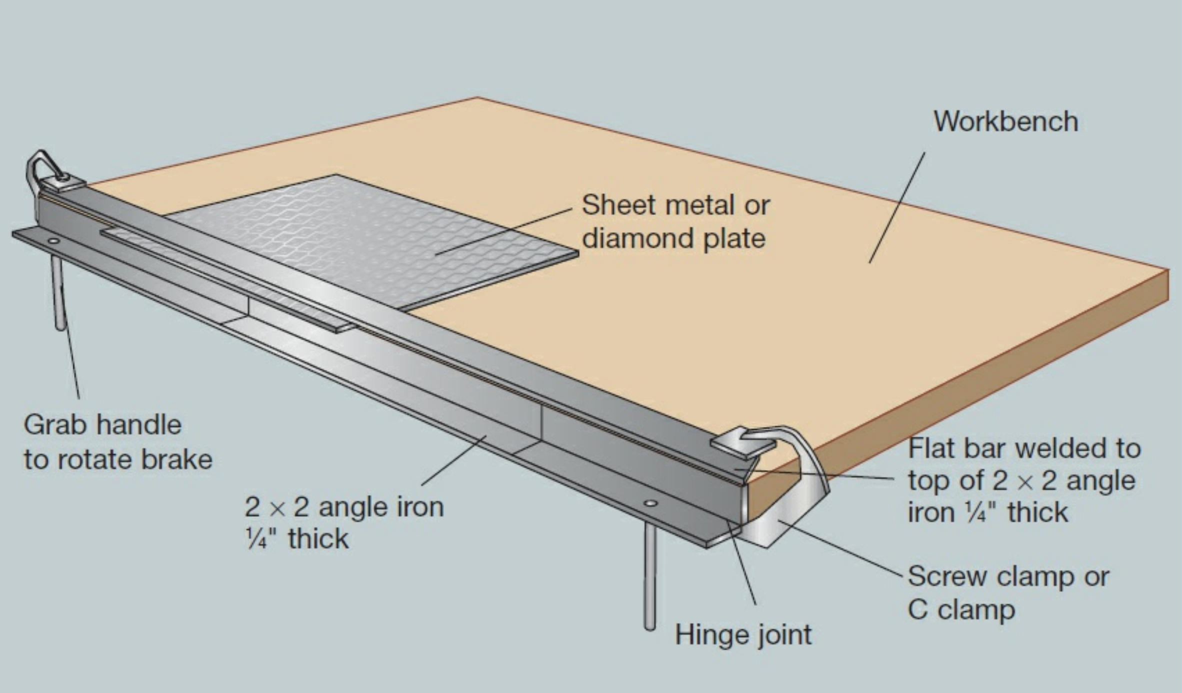
Workbench
Sheet metal or diamond plate
Grab handle to rotate brake
Flat bar welded to top of 2 × 2 angle iron ¼" thick
2 × 2 angle iron ¼" thick
Screw clamp or C clamp
Hinge joint

CAT
PONTIAC
POLARIS ATVs
CATERPILLAR

Cutting the ends off of pipes is a cinch with a plug-in plasma cutter. It even cuts through a significant amount of rust, as shown here.

Making detail cuts—even on a round tube, as shown here— is easy with a plasma cutter. Here, the welder is cutting a new, clean tube. Note the sparks are much smaller than they would be with an oxyacetylene torch.

Here, custom motorcycle builder Scott Webster (of Leroy Thompson choppers) uses the Pro-Cut 55 to cut a piece of steel over a barrel that catches the debris.

The Lincoln Pro-Cut 25 plasma cutting machine is good for fish-mouth fitting tubing, including tailpipe tubing. After the cut is made, a die grinder removes slag from the cut line.

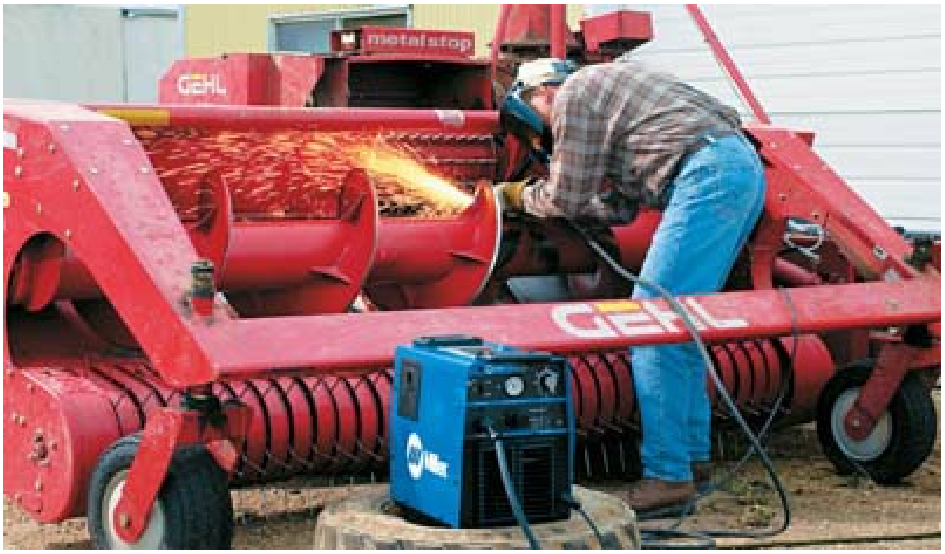
metalstop
GEHL
GEHL
Miller

A Miller plasma cutter is being operated by a gasoline generator and an air compressor to make a repair on this hay baler. When the broken part is removed, a replacement will be welded on by a wire-feed machine.

Oxyacetylene Cutting

Safety

- Full-face #5 filter
- Leather, wool, or cotton long pants, leather jacket, and hat
- Heavy-duty welding gloves
- Leather boots or shoes
- Ventilation

Oxyacetylene cutting uses acetylene and oxygen to preheat metal to a temperature of 1600°F, then uses pure oxygen to burn away the heated metal. Because the cutting is achieved by oxidation of the metal, oxyacetylene cutting only works on metals that are readily oxidized at this temperature. Metals that can be cut with the oxyacetylene process are mild steel and low alloy steels. Oxyacetylene cutting can be used to cut thicknesses from less than 1/8" up to 12". Oxyacetylene cutting is inexpensive compared to machine cutting. It is very portable because no external power source is necessary, and it is faster than machine cutting. It is, however, limited to steel. The molten slag presents a fire and burn hazard, and the heat input to the base metal may cause distortion or changes to the properties of the steel.

Safety: The risk of starting a fire or being burned is high with oxyacetylene cutting. It is important to prevent molten slag and sparks from coming into contact with skin and flammable materials. Make sure you are wearing heavy-duty leather gloves with gauntlets, a leather jacket or leather-sleeve jacket, and leather boots. The work area should be cleared of flammable materials. If an item will be cut in place, make sure surrounding areas are protected from heat, sparks, and dropping slag. Never cut sealed tanks, cylinders, or items that have contained flammable materials. Never cut near gasoline tanks or fuel lines.

Equipment: The basic equipment for oxyacetylene cutting is the same as for oxyacetylene gas welding, with the addition of a cutting torch or cutting attachment. An oxyacetylene cutting setup consists of an oxygen cylinder, an acetylene cylinder, regulators and hoses, and a torch with cutting attachment or a dedicated cutting torch. Cutting attachments and dedicated cutting torches have levers to activate the oxygen flow.

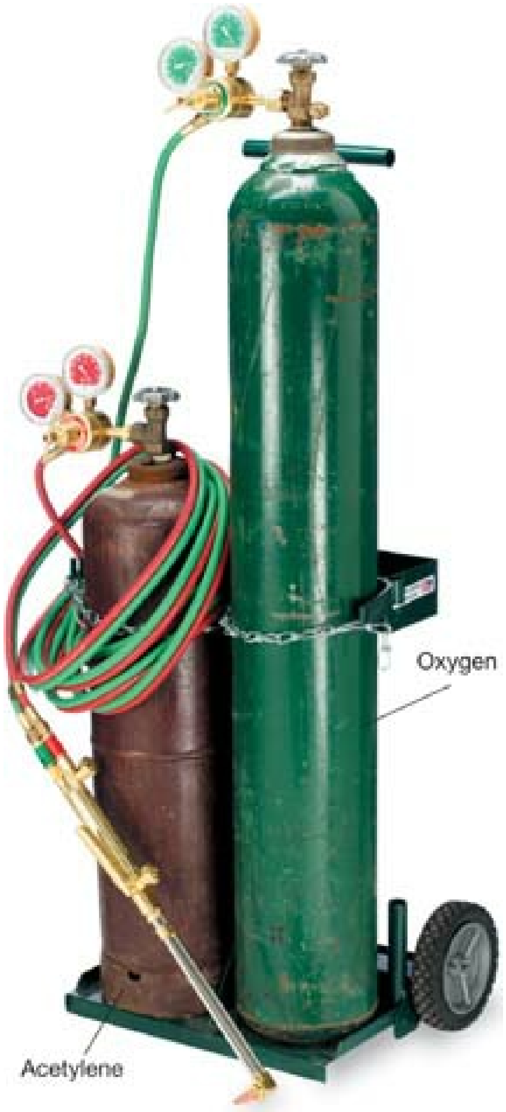
Oxygen
Acetylene

Oxyacetylene cutting equipment consists of two cylinders—oxygen and acetylene—regulators, hoses, and a cutting torch with tip.

Cutting torch and tips: Most oxyacetylene welding sets come with a cutting accessory that attaches to the torch body in the same way that the welding tips do. A dedicated cutting torch is a one-piece unit that has its own mixing chamber. Its greater overall length allows for more distance between the heat zone and the operator, and it can handle higher oxygen flow rates. A dedicated cutting torch is usually fairly expensive and not necessary unless very thick metals will be cut, which requires greater oxygen flow.

Both the cutting torch and the cutting attachment take a variety of cutting tip sizes. The cutting tip has a number of preheat holes surrounding the pure oxygen orifice. It is important to match the cutting tip to the thickness of the metal being cut. A welder often takes a one-size-fits-all approach to cutting tips, which always results in poor-quality cuts. Thinner metals require fewer preheat holes and a smaller oxygen orifice; thicker metals require more preheat holes and a larger oxygen orifice. Cutting tips are made of copper and can be damaged easily. They need to be cleaned regularly with an appropriately sized tip cleaner.

Fuel gas options: Because the metal being cut does not need to be brought to the melting point, various gases other than acetylene can be used. Propane, natural gas, propylene, and methylacetylene-propadiene can be used as the preheating fuel source. Each of these gases requires specific regulators and may require different torches and torch tips. Check manufacturer’s recommended usage before substituting an alternate fuel gas.

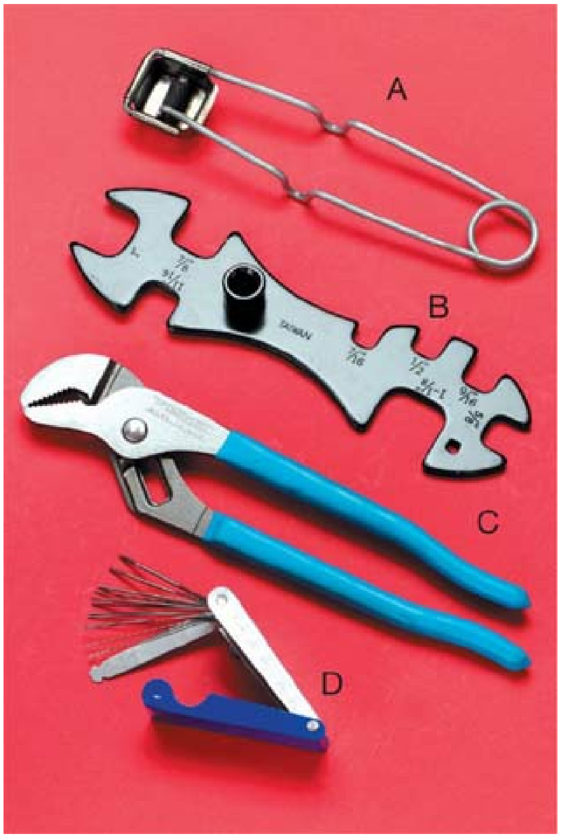

A striker (A), tank wrench (B), pliers (C), and tip cleaner (D) are needed for oxyacetylene cutting.

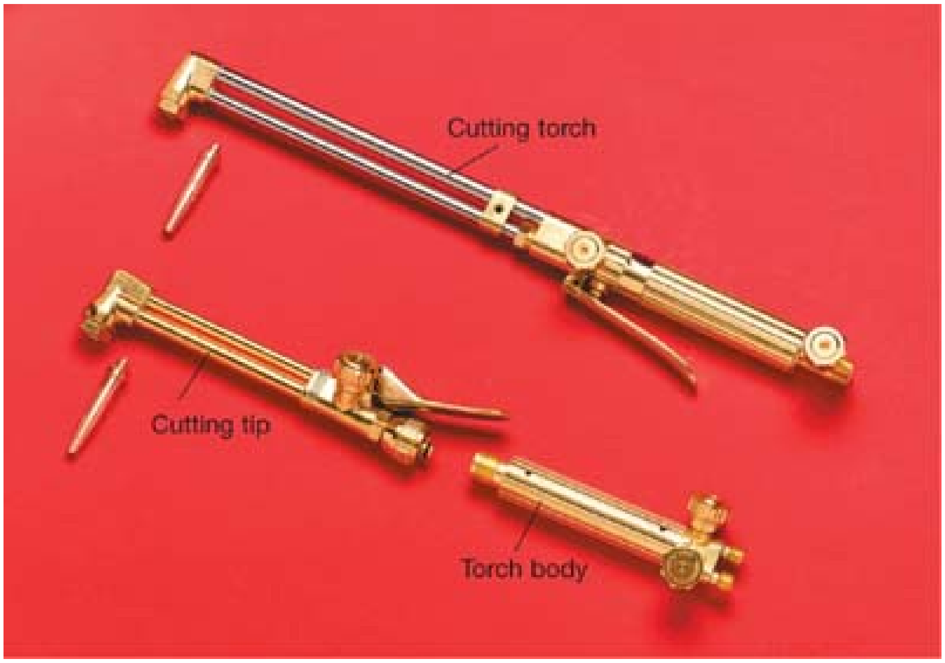
Cutting torch
Cutting tip
Torch body

A dedicated cutting torch (top) is a one-piece unit with one oxygen and one fuel-gas valve. A cutting tip is attached to a standard torch body (bottom) and includes an additional oxygen valve. A range of tip sizes is available for both torch styles.

Precutting Checklist

- Surrounding area is free of flammable materials.
- There is room for cut metal to fall away from arms, legs, hoses, and cylinders.
- Hoses are free from damage before pressurizing the system.
- Ventilation is adequate to remove smoke and fumes from area.

How to Cut with Oxyacetylene

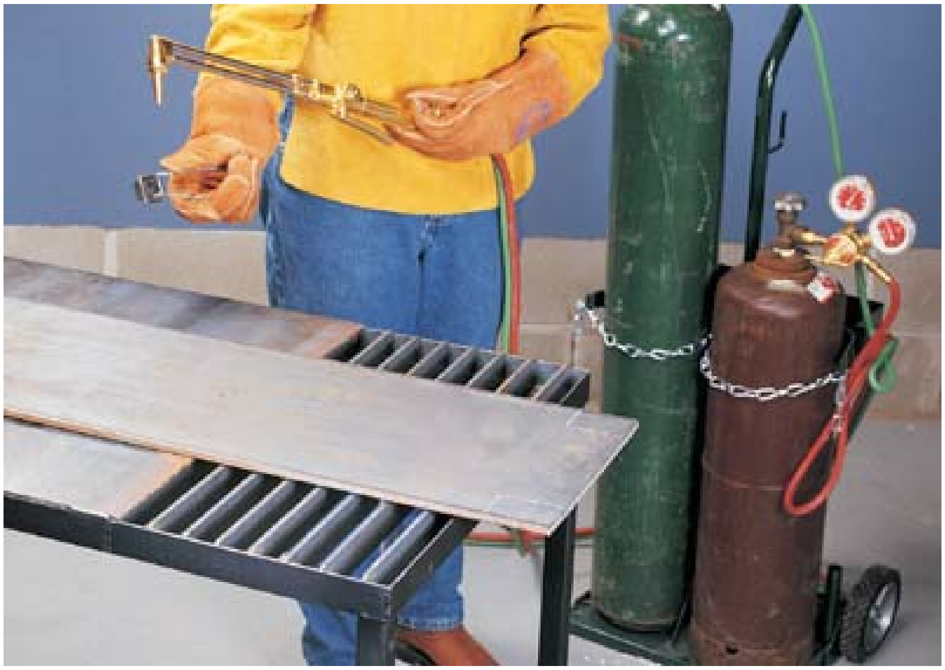

1 See pages 44 to 47 to learn how to set up, pressurize, and light an oxyacetylene torch. Set up the material to be cut on a cutting table. Use soapstone to mark the cutting line, and practice your cutting position and bracing. Hold the torch in your left hand with your thumb and forefinger on the acetylene valve. Hold the striker in your right hand, 3 to 6" from the torch tip. Turn the acetylene valve ¼ to ½ turn and strike sparks.

2 Adjust the acetylene so the flame is not smoking and is not separated from the tip (see page 47). Add oxygen to create a neutral flame. With a dedicated cutting torch, depress the oxygen lever to make sure neutral flame is maintained; if necessary, adjust with the oxygen lever depressed. With a cutting tip on a torch base, depress the oxygen lever and use the cutting tip adjustment valve to adjust the oxygen.

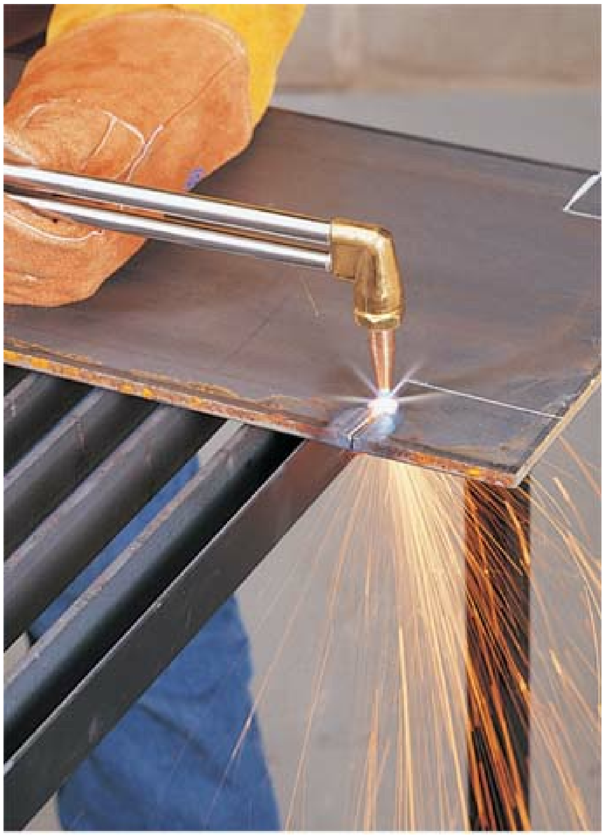

3 Release the oxygen lever, flip down your face shield, and direct the flames at the edge of the cut. Once the metal begins to glow red and appears shiny but not yet molten, depress the oxygen lever. If it fails to pierce a hole immediately, release and continue to heat until depressing the lever pierces a hole. Move slowly and steadily along the cutting line, holding the torch at a 90° angle.

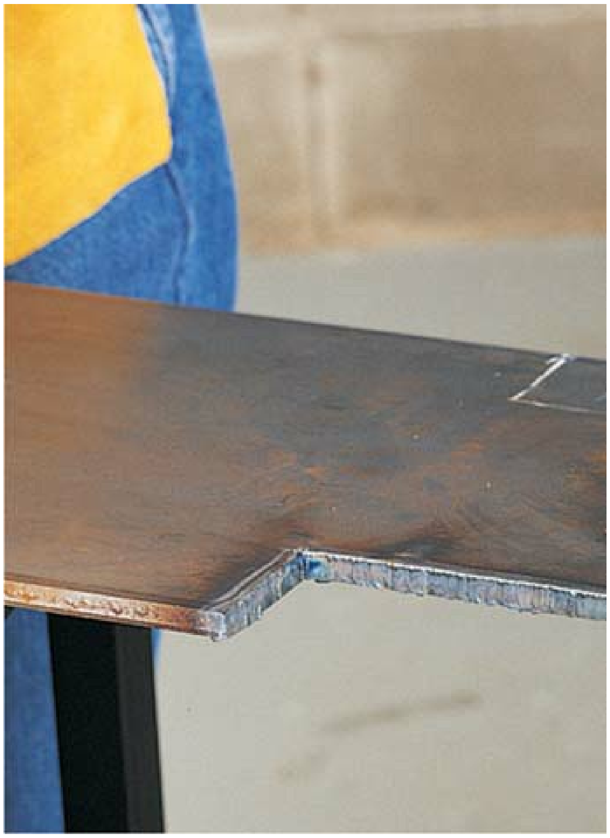

4 The finished cut should have very small vertical ridges (draglines), very little slag on the bottom, and the top edges should not be rounded over.

Oxyacetylene Cutting Tips

Oxyacetylene cutting safety.

Oxyacetylene cutting produces molten slag, sparks, and hot metal scraps. Move items that are to be cut away from burnable material. If the item cannot be moved, protect surrounding areas with sheet metal and fire retardant welding blankets (not tarps). Make sure sparks and slag will not fall into cracks, holes, or ventilation grates in the floor. Wet down any wooden material around the cutting area. Have water, sand, or a fire extinguisher on hand, and monitor the area for one half hour after cutting is completed. Plan where cut metal will fall so it does not hit your arms, legs, gas hoses, or cylinders.

Piercing Holes and Making Cutouts

Piercing a hole: Preheat the material to a dull red. Pull back the torch from the surface and angle slightly. Squeeze the oxygen lever. As soon as the material has been pierced, return the torch tip to the perpendicular position and move back to just above the surface. Complete the hole.

Making a cutout: Begin in the center of the cutout and move to the cutout line after piercing a hole. A spiral approach to the edge works best. To give maximum protection to the surrounding material when piercing a hole, drill a $^{3}/_{16}$" hole to start the piercing action.

Warning

- Never cut into a sealed container.
- Never cut into a tank or container that contains or may have contained flammable materials.
- Never cut near a gasoline tank or fuel lines.

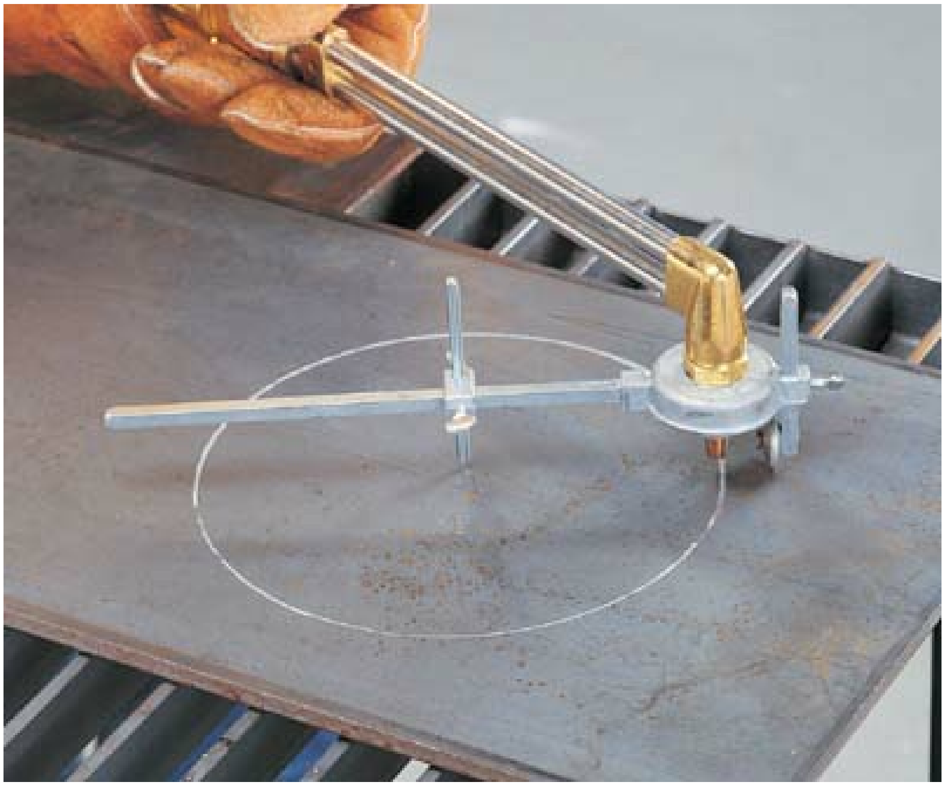

Circle cutting guides are available for making accurate circle cuts. If you need to make several circle cutouts, this can be a handy tool.

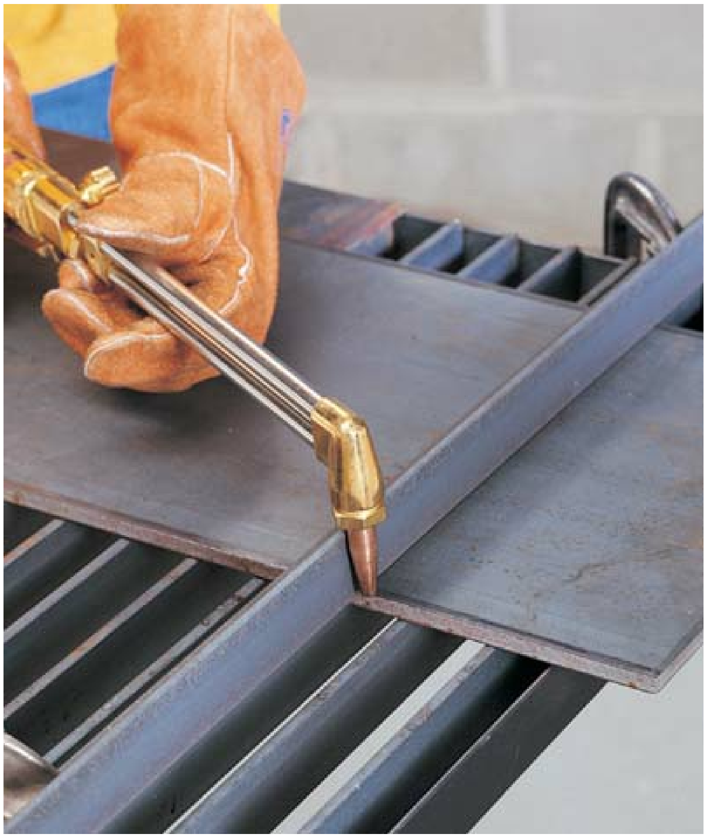

Use a piece of angle iron as a brace to support the torch or to hold the pipe in place for cutting.

Cutting Defects

A good oxyacetylene cut has square edges, small vertical draglines, and little slag. A dirty cutting tip is the cause of many poor-quality cuts. Clean the orifices with a properly-sized tip cleaner and use very fine sandpaper on a flat surface to polish the flat surface of the tip. Here are a few possibilities for causes of cutting defects:

Kerf is wider at the bottom than at the top (bell-mouthed).
Oxygen pressure is too high.

Top edge melted over.
Cutting speed is too slow or preheat flames are too long.

Irregular cut edge.
Speed is too fast or too slow, oxygen pressure is too high, or cutting tip is too small or too large.

Shaping

There are a number of ways to bend and shape mild steel. Brakes for creating angles in sheet metal, and ring rollers and scroll benders to create circles and scrolls are readily available for the home crafter. Unfortunately, the higher the capacity of these bending tools, the higher the price. You can make simple bending jigs from any round, rigid, strong form, such as pipe and salvaged flywheels, pulley wheels, or wheel rims. Using these jigs, it is fairly easy to bend a round rod up to ¼" and a flat bar up to 1/8" into complex shapes. A bench vise is handy for making acute bends. Tubing, rail caps, and channels can be bent with a heavy-duty electrical conduit bender. With practice and patience, you can create well-formed circles by making incremental bends with the conduit bender.

Thicker materials are bent more easily if you have a long end for applying pressure; it is better to cut pieces to length after bending. Sliding a pipe over a short end will increase your leverage. When shaping metal into sharp bends, take into account the length needed for the radius of the bend. Cold-formed metal springs back somewhat, so it may take trial and error to find which size jig makes the bend size you desire.

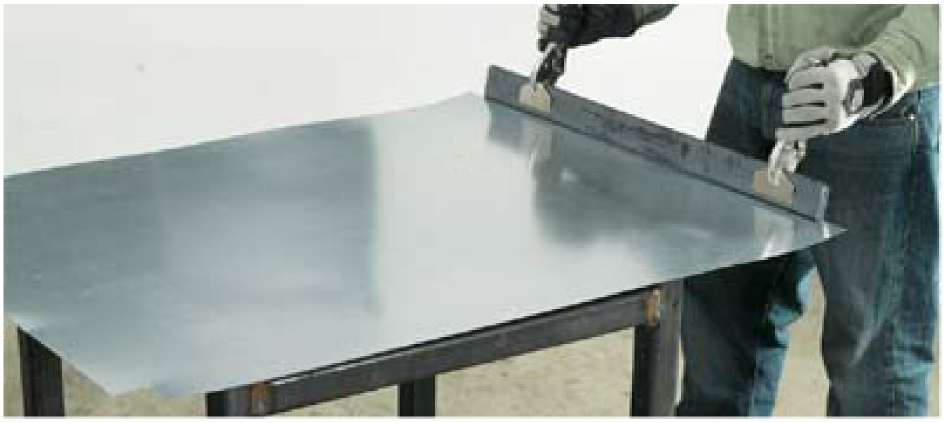

Use a flat bar and locking pliers to bend sheet metal.

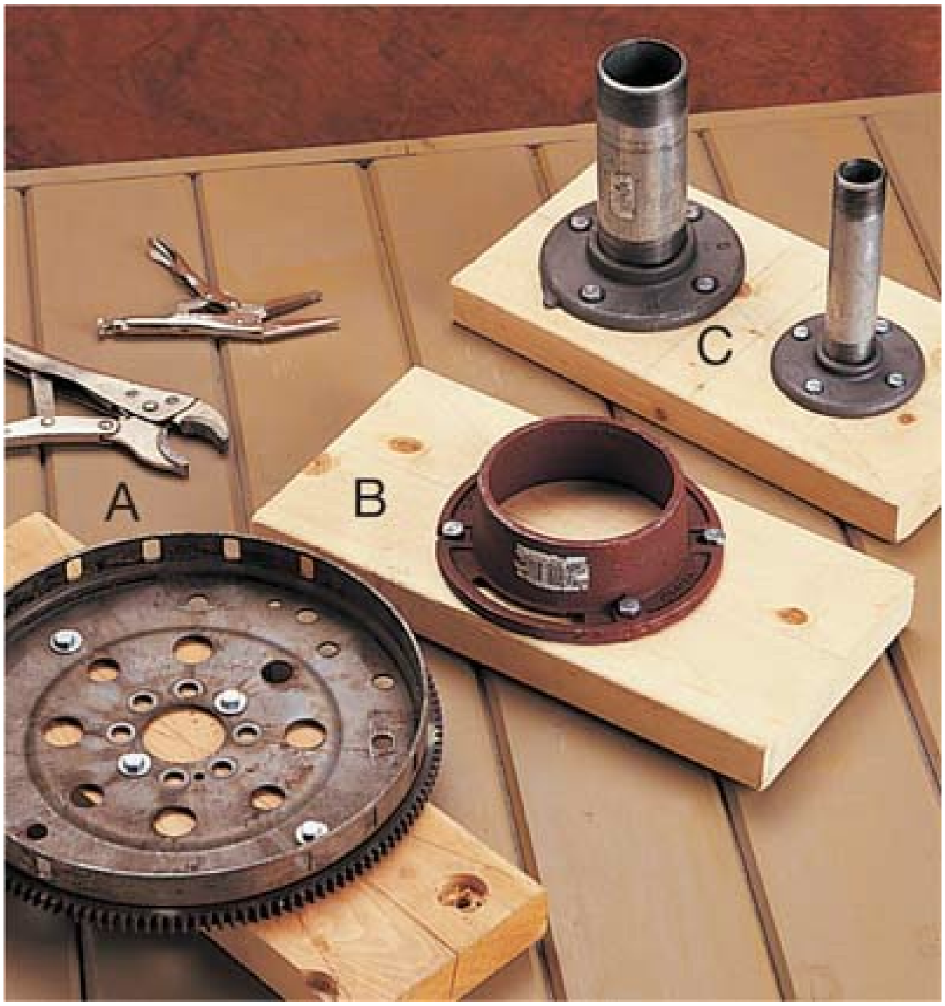

Bending jigs can be made from many rigid, circular items. An engine flywheel (A), toilet flange (B), and various pipe sizes (C) are shown here.

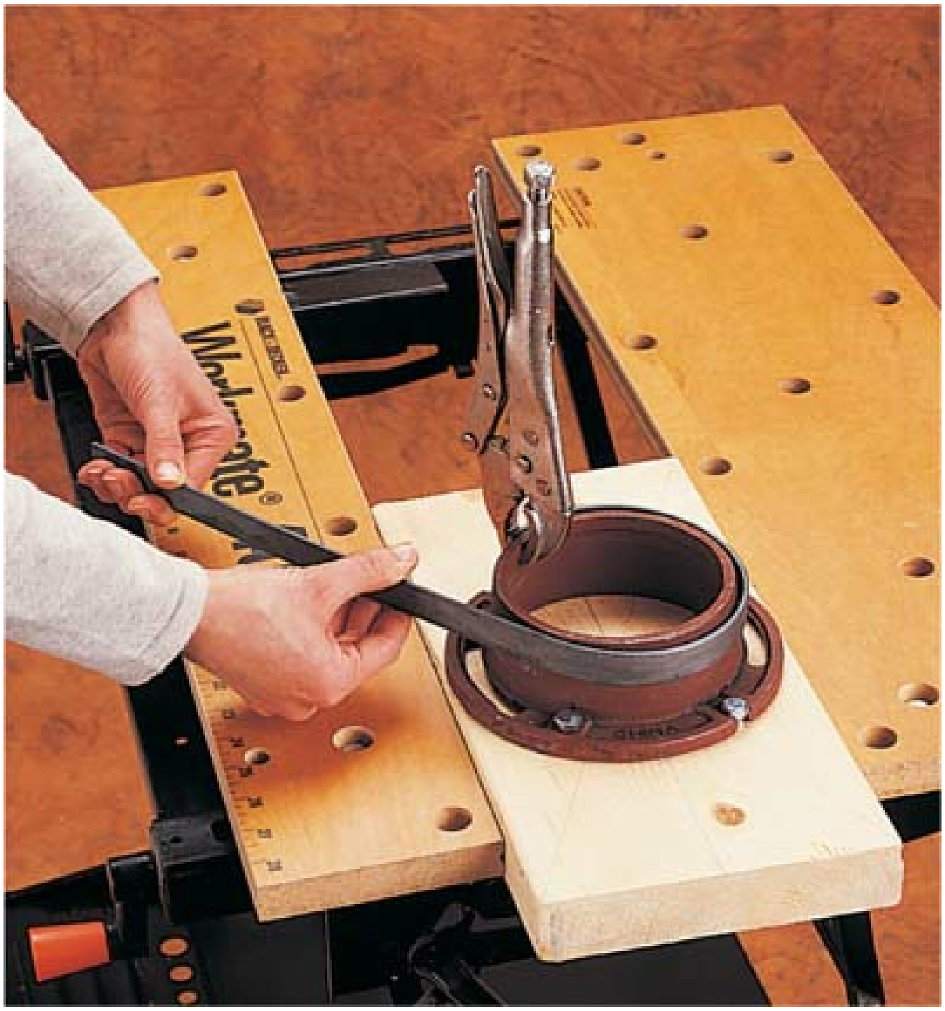
Use locking pliers to hold the metal to the jig. Wrap the metal around the jig in a smooth motion.

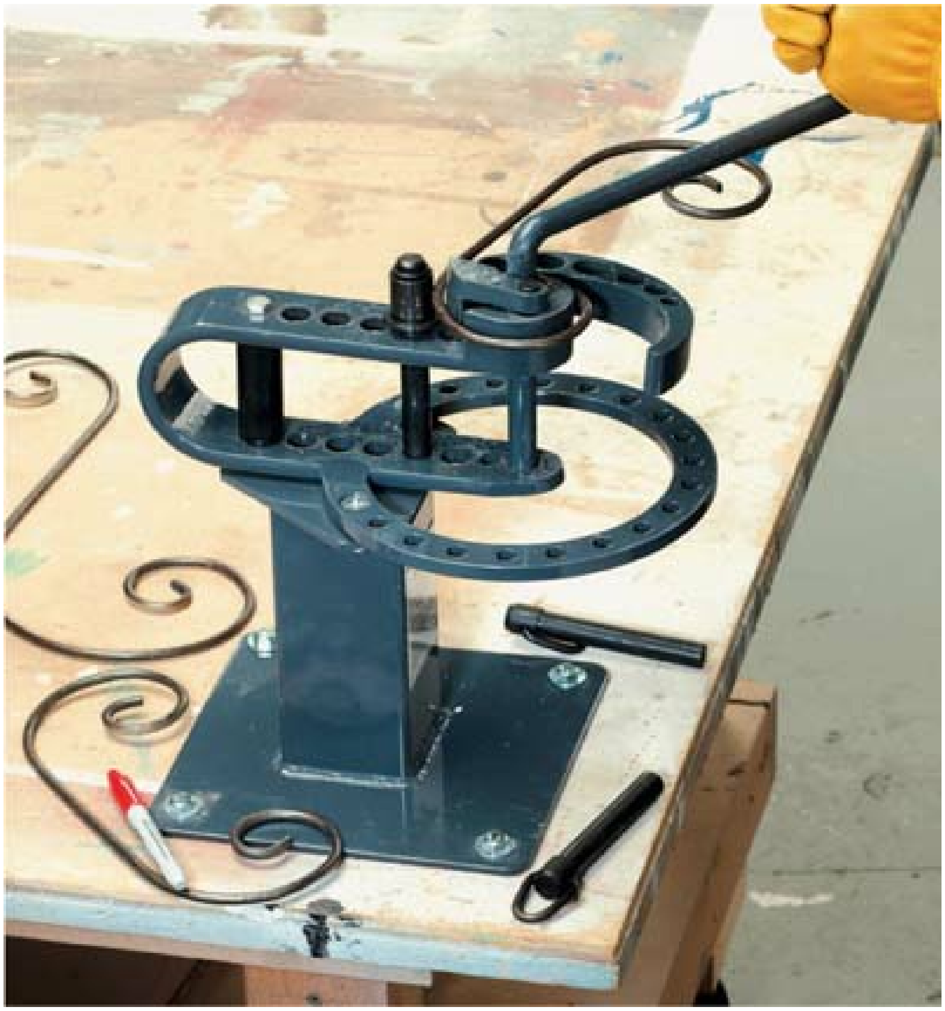

Use a scroll bender to bend decorative scrolls.

Finishing

A nice finish adds to the beauty of your welded project. Unless you intend to let your project fully rust, even bare metal needs some type of coating. Applying a paint coating is best done by spraying, not brushing.

If you look closely at most welded furniture or garden accessories you'll see that they typically have rough welds and spatter. With your own projects, especially indoor projects, taking the time to grind down welds and grind off spatter is time well spent. A nicely ground, good weld looks like a continuous solid piece of metal. This is especially important for glossy finishes, as a smooth, shiny paint job will highlight even the tiniest imperfections.

A primer coat is recommended underneath all but bare metal coatings. If you're going for bare metal, use a clear coat or a metal polish minimally to prevent rust. If already deeply rusted areas cannot be cleaned to bare metal, use a rusty metal primer or, even better, a rust converter. Carefully read product labels, as some converters require oil-based top coats and other specific finishes.

A wide variety of spray-on finishes is available. With a single application you can create beautiful, detailed faux finishes, such as a hammered texture or granite (Rust-Oleum® Stone Creations, for example). Crackle and antiquing finishes are also available as single spray-on finish paints and brush-on paints. Of course, you can create your own antique look as well (see photo, this page).

If you have more time and a higher budget, consider a verdigris finish. This finish requires more steps and time due to paint layers, but it creates a unique look that is very impressive yet natural looking for metal. The great part of simulating a patina is that you can control the hues and coloration, and then seal it all so that it remains perfect for years and years. Powder coating and brass plating are other, more expensive, finish options.

For exterior applications, look to railing and gate suppliers; they carry a variety of primers, paints, dyes, and patinas to give your project a one-of-a-kind look and withstand the constant exposure to natural elements. If the product will be near open flames, make sure to buy paint specifically for such purposes (like the square fire pit on page 186).

Never to be underestimated, bare metal creates intriguing looks just by using sanding, grinding, and heating techniques. One example of this is shown on the tapered-leg table on this page. The effect was achieved with an angle grinder that was used in a circular motion until the desired texture was achieved. It is time-consuming work that requires a little muscle action, but the finished look is unmatched.

There are a number of ways to create an antique or aged painted finish. One is to paint the surface with three or four layers of different paint colors, allowing each layer to dry thoroughly between coats. When the final coat is dry, sand off corners and high spots to reveal the paint layers. Another method is to apply one color layer and allow it to dry completely. Then apply a second layer and third layer without allowing drying time. Use a damp sponge to wipe off layers at corners and high spots (see photo page 38, top). To create a crackly, “alligatored” look, purchase crackle base and crackle top coat, either as a spray paint or brush on paint. Apply the two coats as directed to create the crackle effect (see photo page 38, middle).

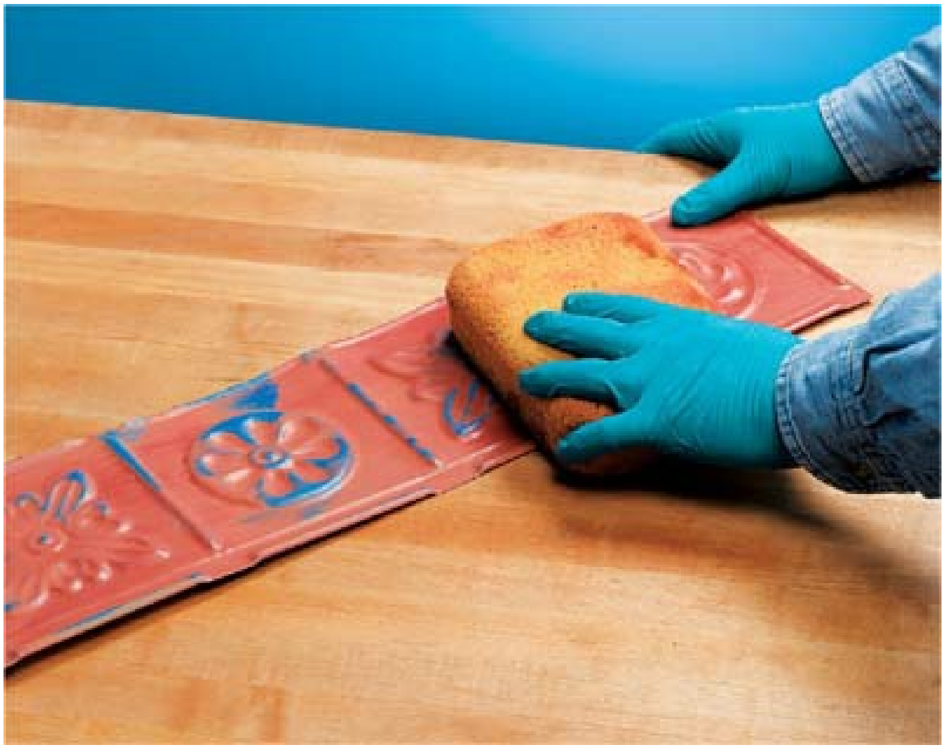

One method for creating an antique finish is to apply two wet top coats of different colors to a dry base coat of a third color. Before the top coats dry, wipe the top coats off in selected areas with a wet sponge.

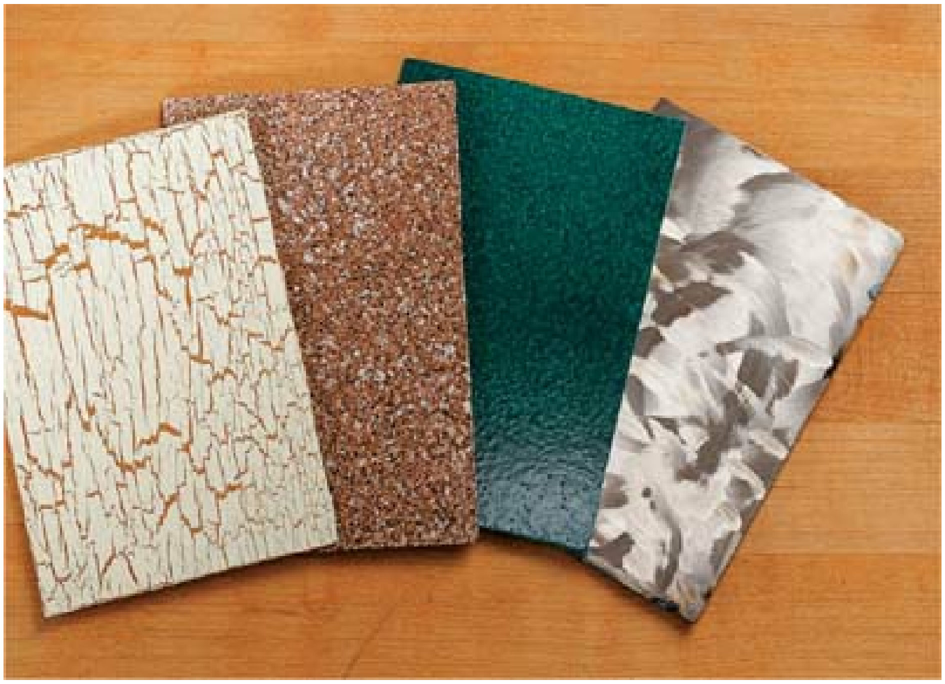

There are many possibilities for finishing metal. Pictured here, from left to right: brush-on crackle finish, spray- painted stone, spray-painted hammered metal, and a swirl pattern created with an angle grinder.

An angle grinder works well for beveling edges on thick material. Grinding down welds gives projects a finished look.

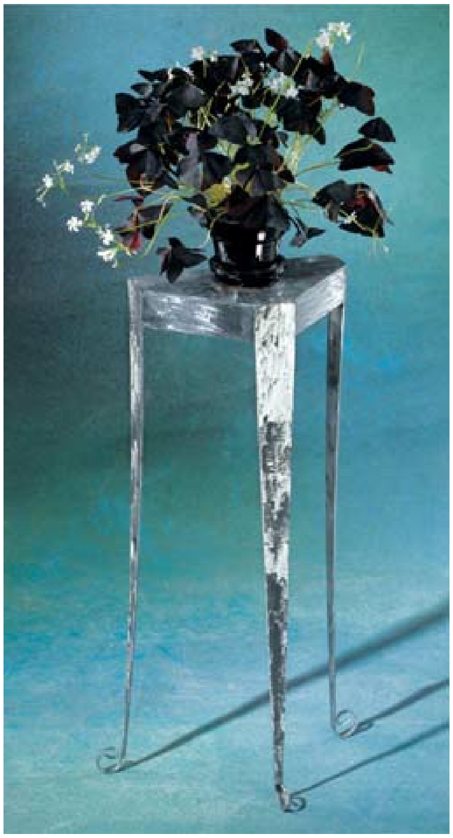

Use an angle grinder in a circular motion to achieve a shiny textured effect. On thin material, as shown here, use small circles with light pressure.

Clean and then apply primer to prevent rust. Standard white is suitable if you intend to paint the finished project.

After the primer is completely dry, add the final top coat paint. This top coat should also be formulated for metal and/or exterior applications.

Oxyacetylene Welding

Oxyfuel welding is the process that uses the heat from a gas flame to melt base materials and cause them to join together. The gas flame is created by the combustion of oxygen and a fuel gas. Fuel gases are: acetylene, propane, butane, hydrogen, natural gas, and methyl acetylene-propadiene, (MPS, formerly known as MAPP gas). Oxygen and acetylene burn in a neutral flame at a temperature between 5600° and 6300°F, the hottest of any gas flame and capable of melting most metals. Other oxygen-fuel gas combinations are hot enough to use for soldering and brazing, but not hot enough for welding. Oxyfuel welding with acetylene is called oxyacetylene welding, but it is often simply referred to as gas welding.

The oxyacetylene process is versatile, as it can be used for both welding and cutting materials (see page 33) as well as heating, soldering, and brazing. It can be much less expensive than arc welding and is very portable because it needs no electrical power source. Oxyacetylene can be used to weld any thickness of metal, but sections over ¼" are difficult to weld. Unfortunately, the techniques of oxyacetylene welding can be very hard to master, and there are serious safety concerns with the extreme flammability of acetylene and the high-pressure gas cylinders.

Oxyacetylene welding equipment consists of an oxygen cylinder and an acetylene cylinder, each with regulators and gauges, oxygen hose, acetylene hose, torch, and tips.

Oxyacetylene welding may use fusion welding, where the base metals are melted together without filler material, or, more commonly, it may use a filler metal.

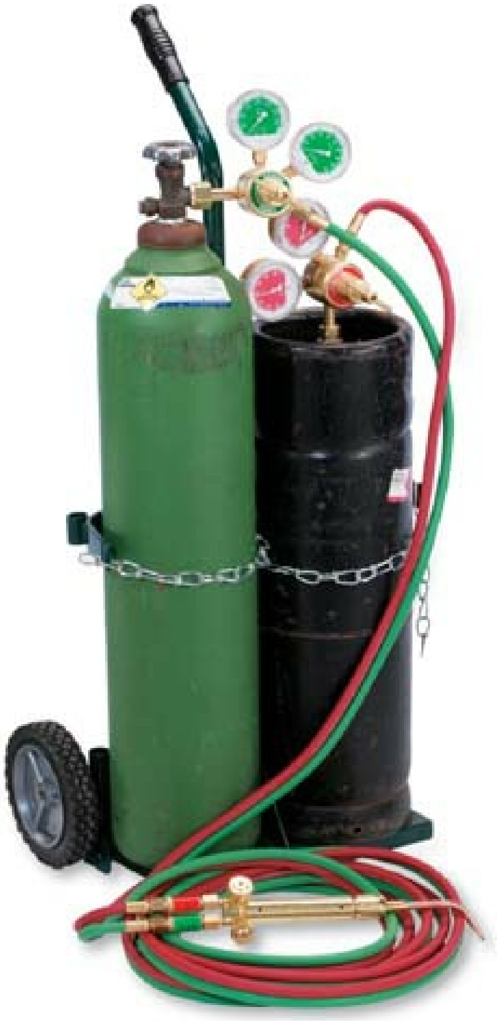

An oxyacetylene welding rig consists of two cylinders—oxygen and acetylene, regulators, hoses, and a torch with tip.

Safety

- Welding helmet or face shield with #5 filter
- Leather, wool, or cotton long pants, long sleeve shirt, and hat
- Leather gloves with gauntlet
- Leather boots or shoes
- Ventilation

The Equipment

Cylinders: Oxygen comes in high-pressure cylinders (sometimes called bottles or tanks) at 2200 pounds per square inch (psi) or more when full. The oxygen cylinder has a valve, and it always comes with a safety cap that screws on to the top to protect the valve. Never use the safety cap to lift an oxygen tank. The oxygen cylinder has a safety valve that will rupture to relieve tank pressure if the tank becomes overheated. Oxygen tanks should always be stored and transported in an upright position.

Acetylene: Acetylene is an extremely unstable gas. It cannot be pressurized above 15 psi in a free state. To safely pressurize the acetylene to 225 psi, the acetylene cylinder is filled with a porous material and acetone. When pressurized, the acetylene is absorbed by the acetone, which stabilizes it. When the pressure is released (the valve opened) the acetylene bubbles out of solution. Acetylene cylinders have fusible plugs that will melt in case of a fire and allow the gas to slowly escape rather than explode. Acetylene cylinders should always be stored in an upright position. Using an acetylene tank at an angle will allow acetone through the regulator and hoses, which will destroy them.

Cylinders should always be kept upright and secured with a chain or strap to prevent them from falling over. Never use a cylinder as a roller for moving heavy objects or for any purpose other than its intended use.

Cylinder valves: Each type of cylinder has a valve that controls the flow of gas from the cylinder. A handwheel or valve wrench is used to open the valve. The oxygen cylinder valve should always be opened fully, as it is a backseating valve. A backseating valve is leakproof when closed and fully opened, but not otherwise. The acetylene valve should never be opened more than ¾ to 1½ turns. Less than this may lead to insufficient amounts of fuel and backflash; more than this makes it difficult to turn the acetylene off quickly in case of an emergency. If the acetylene valve is opened all the way, or opened suddenly, the acetone may escape from the cylinder, just like an overflowing soda can. Acetone can destroy the hoses and regulator fittings.

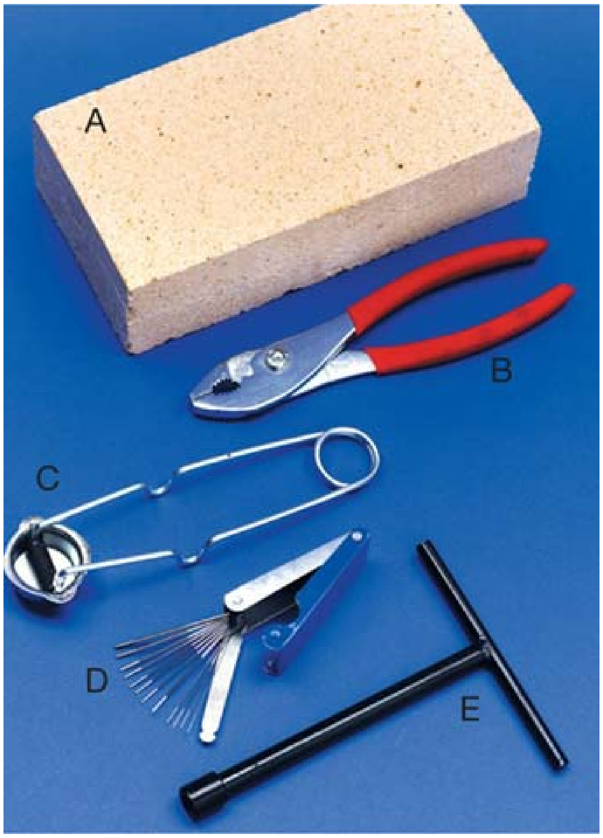

Tools for oxyacetylene welding include fire bricks (A), pliers (B), striker (C), tip cleaner (D), and acetylene tank wrench (E).

Note

Never use oil or any other type of lubricant on regulators or other welding equipment. The combination of oil and oxygen in a spark-filled environment can lead to a fire in the regulator, which could cause a cylinder explosion.

Pressure regulators and gauges: Pressure regulators reduce the pressure of the gas leaving the cylinder. A single-stage regulator reduces the cylinder pressure to working pressure in one step. A two-stage regulator reduces the cylinder pressure to working pressure in two steps. A two- stage regulator is better because it gives more precise control over gas flow. but it is substantially more expensive. The regulator adjusting screw turns clockwise to increase pressure and counterclockwise to reduce pressure. The regulator is completely off when the adjusting screw is loose or “backed out.” Always back out adjusting screws as part of your post-welding routine.

Hoses: The flexible rubber hoses that move the gas from the regulator to the torch are designed to be leakproof and withstand high pressure. The oxygen hose is green, and the acetylene or fuel hose is red. Hoses are available either as a single hose or a dual hose where the oxygen and fuel hoses are paired together. The dual hoses are more common and more convenient. Always protect hoses from damage by moving them off the floor when not in use, keeping them behind the welding area so they are not burned with sparks, and keeping them out of traffic lanes so they are not stepped on or run over by vehicles. Hoses should not be allowed to come in contact with oily surfaces and should be protected from sunlight and chemical fumes to keep them in good condition. Always drain hoses when you have finished your welding session.

Fittings: Fittings connect the hoses to the regulator at one end and the torch at the other. For safety, the nut for the fuel hose (red hose) is left-hand threaded and has a groove machined around thte nut. The oxygen (green) hose nut is right-hand threaded and has no groove. Never interchange the oxygen and fuel hoses or fittings. Never force a fitting—brass is a very soft metal, and it is easy to damage the fitting threads. Always hand thread and hand tighten fittings before using a wrench, and be careful not to overtighten them. Never use pliers to tighten fittings, as they will damage the brass nuts.

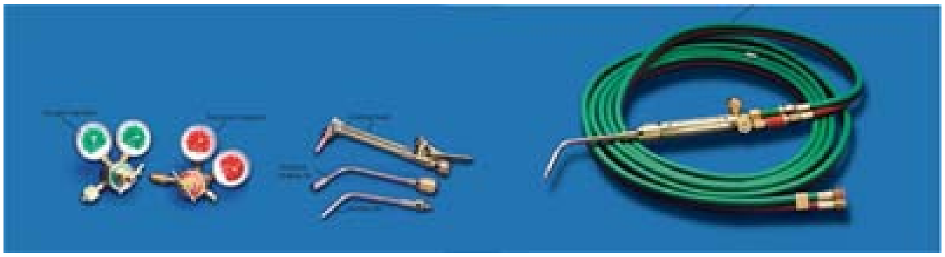

Oxygen and acetylene hoses with flashback arrestors, torch body, and welding tip.

Oxygen and acetylene regulators

Available tips for an oxyacetylene torch body include cutting head, rosebud heating tip, and welding tip.

Check valves and flashback arrestors: Check valves and flashback arrestors are two safety features to prevent reverse gas flow or flashbacks. The check valve allows gas to flow from the hose to the torch. If gas pressure within the torch exceeds the hose pressure, a spring closes the valve to prevent backflow. In the event of a flashback, the check valve needs to be replaced. A flashback arrestor is installed between the torch and the hose, and offers more protection than the check valve. The flashback arrestor prevents burning oxygen and fuel from flashing back into the hoses and regulator, which could cause an explosion. The flashback arrestor consists of a check valve, pressure-sensitive valve, stainless steel filter, and heat sensitive check valve. In the event of a flashback, the flashback arrestor does not need to be replaced.

Torch: The torch mixes and controls the flow of the fuel gas and oxygen. The torch consists of valves, torch body, mixing chamber, and tips. The torch is sometimes referred to as a blowpipe, but this is not technically correct.

Tips: Tips attach to the torch body and come in many sizes to create different-size flames. A cutting tip, also called a cutting head or cutting torch, is attached to the torch body to use for flame cutting. A cutting tip has a number of holes for the preheating flames around a center hole for the pure oxygen. Cutting tips also come in a variety of sizes and with different numbers of preheating holes. Heating, or rosebud, tips are used to preheat metals to improve their weldability. Welding tips come in a wide range of sizes to match the thickness of the metal being welded. Matching the tip size to the welding material and gas pressure is critical for creating quality welds.

How to Set Up an Oxyacetylene Outfit

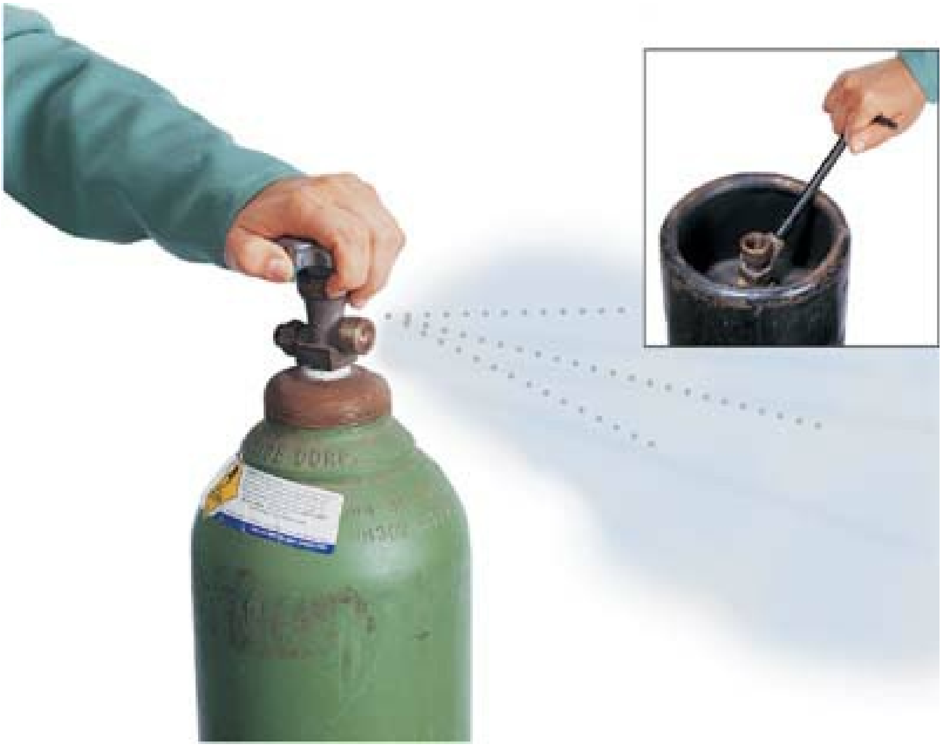

1 Secure the cylinders in an upright position, chained to a cart or strapped to a wall or post. Remove the protective cylinder caps. Wipe off the cylinder valve seats, regulator connections, and hose connections with a clean cloth. Crack open each cylinder valve briefly to expel any trapped dirt particles. “Flat top” acetylene cylinders (inset) may have antifreeze in the recessed valve seat. Use a clean rag to remove the liquid and dry the valve seat. This style acetylene cylinder requires a cylinder wrench to open the valve.

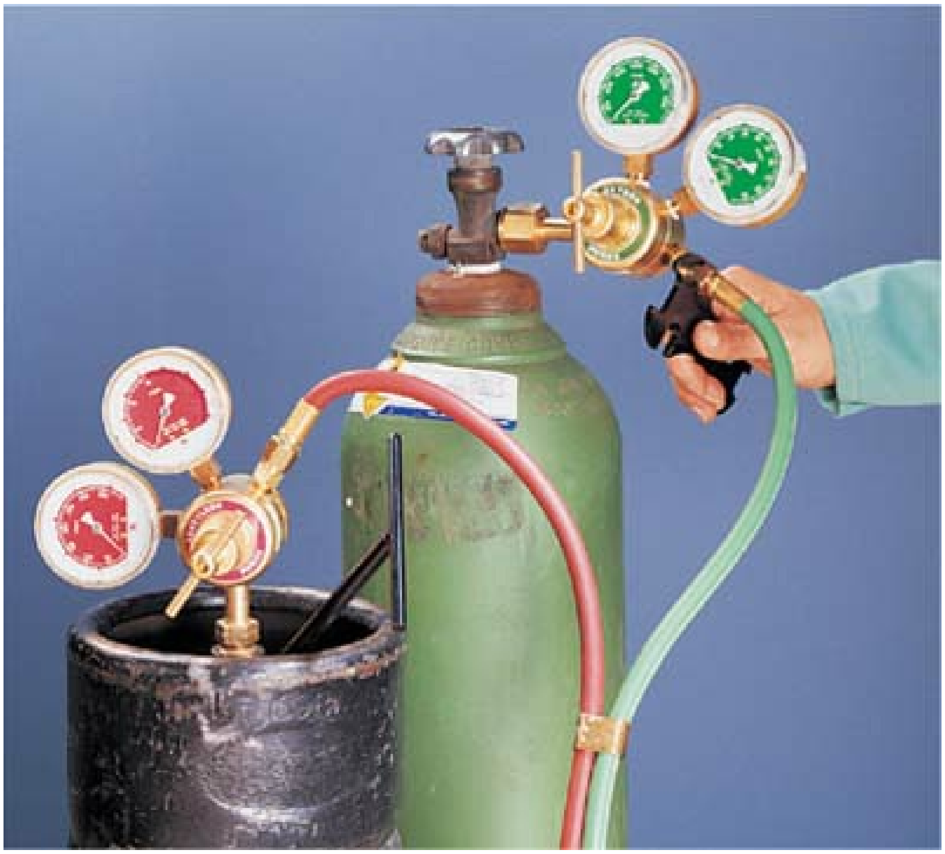

2 Attach the regulators to the cylinders. (The acetylene connectors have left-hand threads.) Always hand tighten, then use a fixed wrench, not a pliers or an adjustable wrench, to tighten. Do not overtighten—a firm seating is all that is necessary. Attach the hoses to the regulators. The acetylene hose is red and left-hand threaded. The oxygen hose is green. NEVER use grease, oil, or pipe dope to lubricate fittings. Grease and oil can ignite spontaneously when they come in contact with oxygen—even without a spark or flame present.

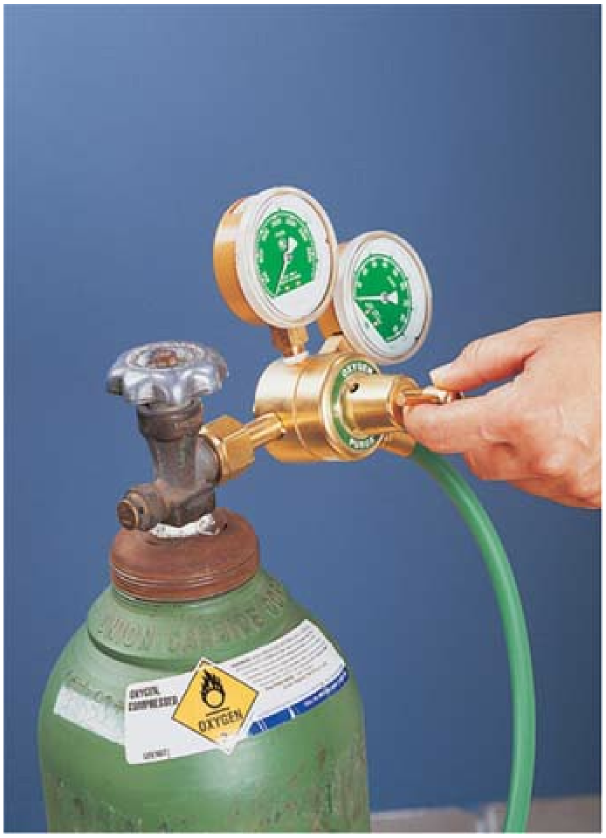

3 Turn the regulator adjustment screws on the oxygen and acetylene regulators counterclockwise until they are loose. (Some regulators have a knob.)

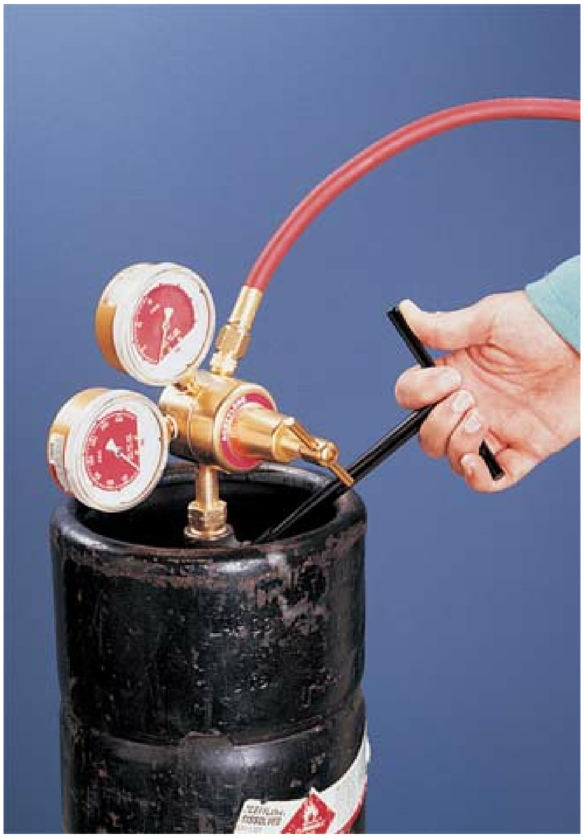

4 Open the oxygen valve slowly all the way while standing to the side in case the regulator gauge glass shatters. Turn the regulator adjustment screw until oxygen begins to flow through the hose, then loosen the regulator adjustment screw to stop the oxygen flow. Slowly turn the acetylene cylinder valve ¾ to 1½ turns.

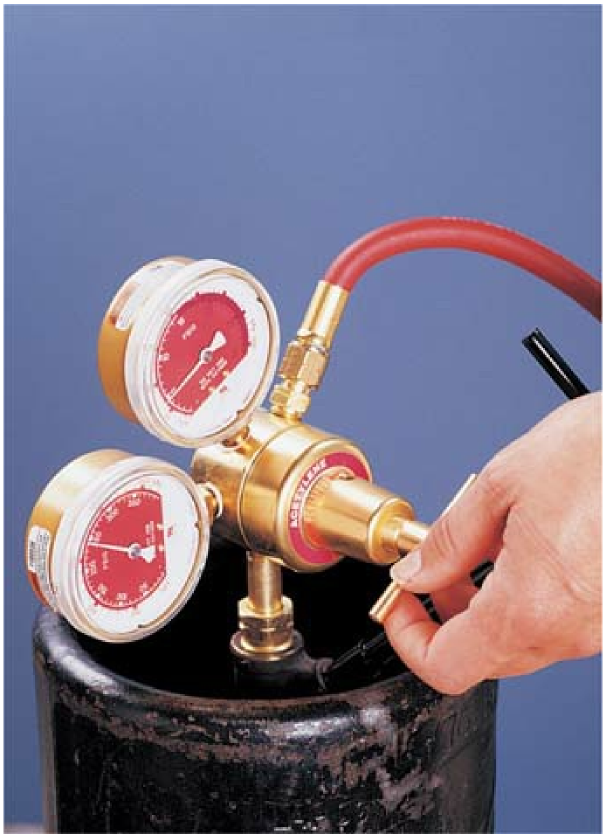

5 Adjust the acetylene regulator valve until the gas begins to flow, then loosen the regulator adjustment screw to stop the flow. Attach the torch to the hoses. Be sure to pressurize the system and check for leaks before lighting.

How to Pressurize (Turn On) an Oxyacetylene Outfit

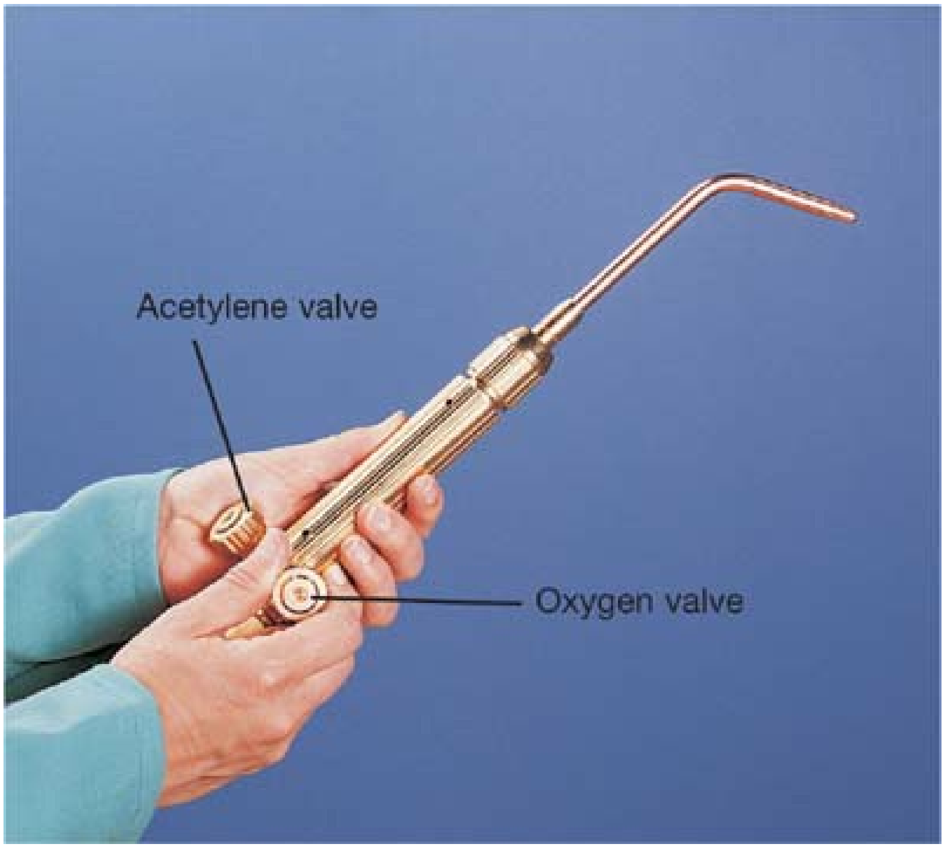

Make sure both torch valves are off. Turn both regulator adjustment screws counterclockwise until loose. Slowly turn the oxygen cylinder valve on. Once open, turn the valve all the way open to ensure proper seating without leaks. Turn the oxygen regulator adjustment screw clockwise until the gauge reads the desired pressure. (Refer to the manufacturer's specific instructions for operating pressures.) Open the oxygen valve on the torch to check the flowing oxygen pressure.

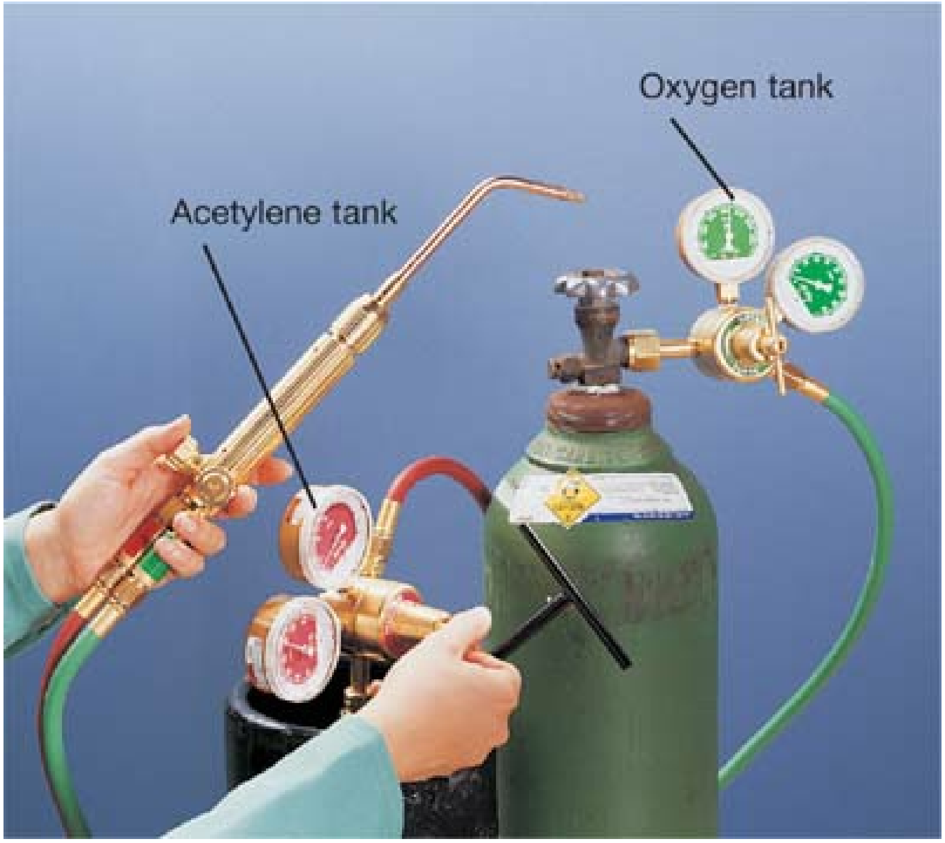

Adjust if necessary and close the torch valve. Slowly open the acetylene cylinder valve ¾ to 1½ turns. (Leave the wrench on the valve if it is a wrench-style valve.) Turn the acetylene regulator pressure adjustment screw until the desired pressure reads on the regulator gauge. Open the acetylene valve on the torch briefly to make sure the flowing pressure matches the desired working pressure. If not, adjust the regulator until the proper pressure is reached.

Checking for Leaks

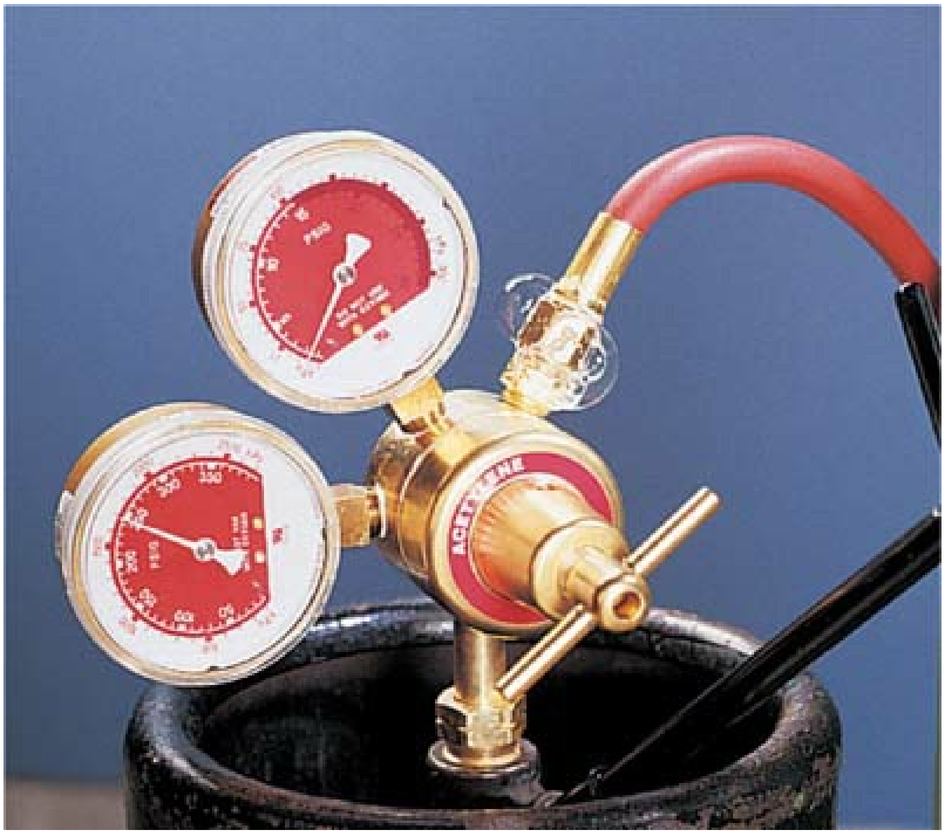

Apply leak-detecting solution to all connections with a small brush. (You can use soap and water so long as the soap is not petroleum based.) If any connections cause bubbles in the solution, tighten the connections and check again.

Backfire & Flashback

Backfire and flashback are two hazardous situations that can be caused by improper gas pressures. Backfire is the pre-ignition of the acetylene and oxygen inside the tip that causes a popping sound. This may damage the tip or spray molten metal from the weld area. Flashback is the flame burning backward into the torch or hoses, causing a popping or squealing noise. Flashback can cause an explosion in the hoses. Avoid both hazards by matching the tip size to the material being welded and by using the proper pressure settings. Using lower pressures than recommended can cause backfire and possibly flashback.

Prelighting Checklist

- Make sure torch valves are closed.
- Turn both regulator adjustment screws counterclockwise until loose.
- Slowly turn on the oxygen cylinder valve all the way. Turn the regulator adjustment to the proper pressure.
- Slowly turn the acetylene cylinder valve ¾ to 1½ turns. Turn the regulator adjustment valve to the proper pressure.

Some welders have been taught to open both the acetylene and oxygen valves before lighting the torch to avoid the smoky acetylene flame. This practice is no longer recommended.

NEVER light a torch with a match or butane lighter.

Lighting the Torch

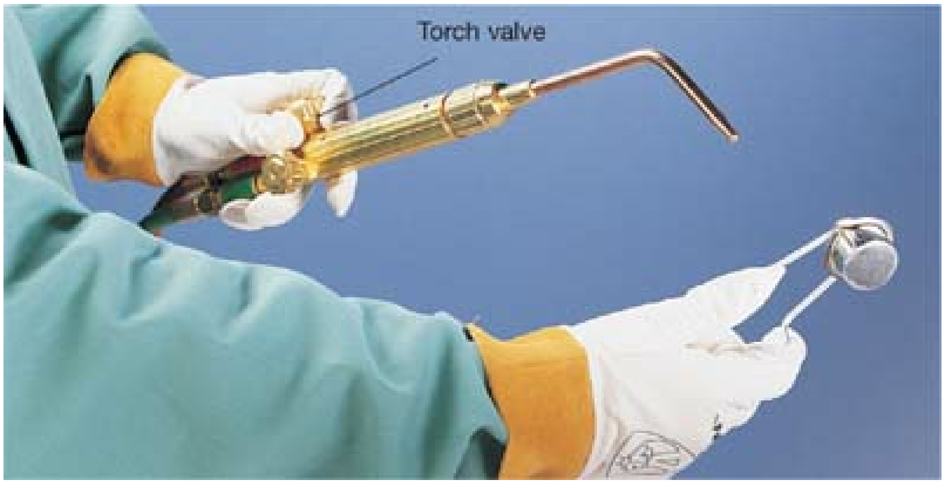

1 Hold the torch in one hand with the thumb and forefinger on the acetylene torch valve. Hold the striker in front of the torch about 3 to 6" away at a slight angle. Turn on the acetylene torch valve ¼ to ½ turn. Immediately use the spark lighter to light the flame.

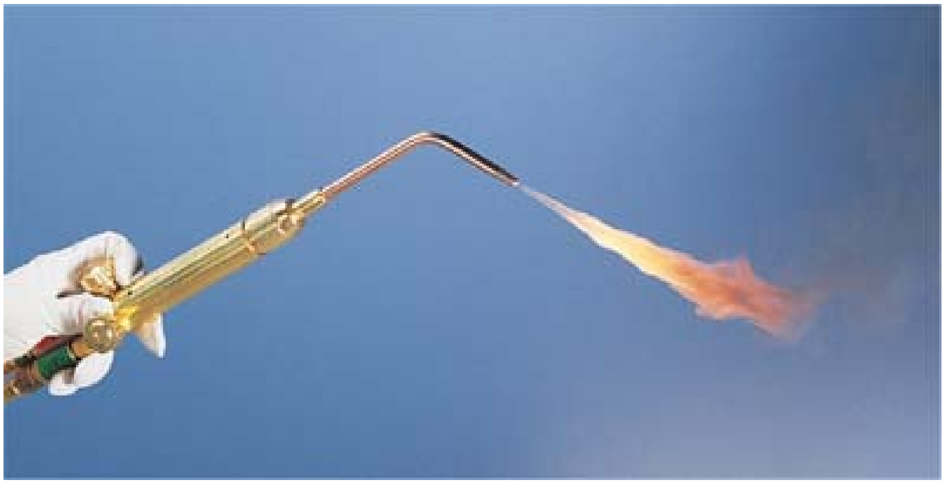

2 The flame will be yellow and smoky.

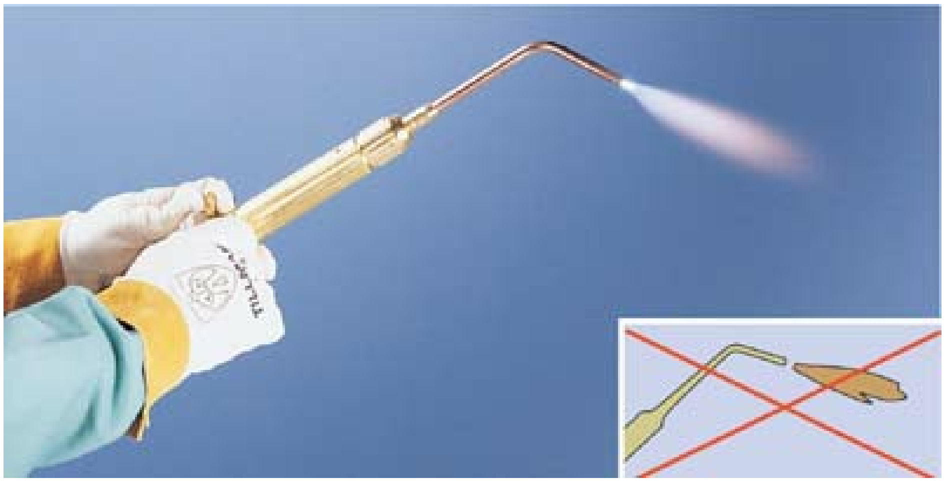

3 Put the striker down and adjust the acetylene torch valve with your right hand so the flame is burning without producing soot. The flame should not be separated from the torch (inset). Open the oxygen torch valve slowly. Adjust the oxygen to get a bright white inner flame and a bluish outer flame. Turn down the acetylene to eliminate the excess acetylene feather if present. When you have finished welding, turn off the oxygen first, then the acetylene.

Flame & Flame States

The flame of an oxyacetylene torch has two parts—the inner or primary flame and the outer or secondary flame. The flame has different temperatures at different locations. The outside edges are cooler because they are burning with the ambient air, which is only 21% oxygen. The torch tip is cooler because complete combustion hasn’t been reached. The hottest area is the tip of the primary or inner flame cone. Here the gases are completely combusted and are insulated by the secondary flame. There are three flame states for the oxyacetylene flame: carburizing, neutral, and oxidizing.

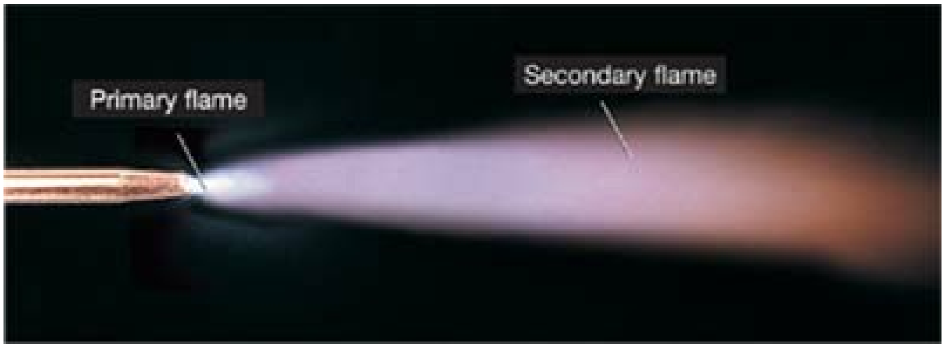

The carburizing or reducing flame has an excess of fuel. This is a useful flame as it will break down metal oxides to get at the oxygen, thus cleaning the weld area to a small degree. This process adds carbon to welds, which makes them harder. The carburizing flame has a bright white primary flame, an acetylene “feather” around the primary flame, and a bluish white secondary flame with orange edging.

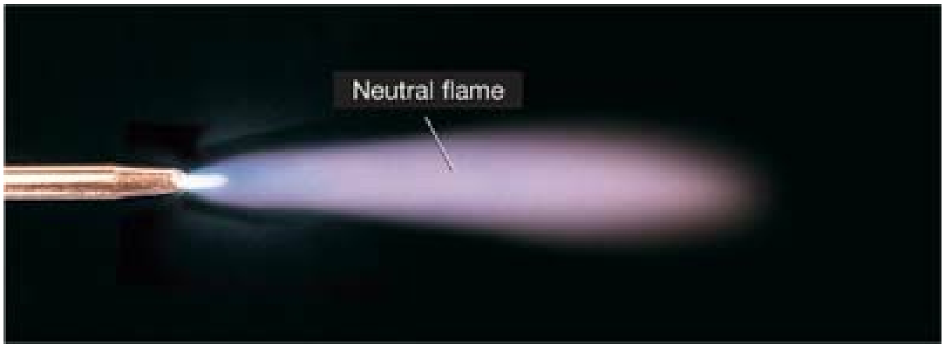

The neutral flame is the exact point where the feather and the inner cone come together. In this flame, there is exactly enough oxygen present to provide total combustion of the fuel gas. Most welding and cutting operations use a neutral flame. The neutral flame has a bright white primary flame and a colorless to bluish secondary flame.

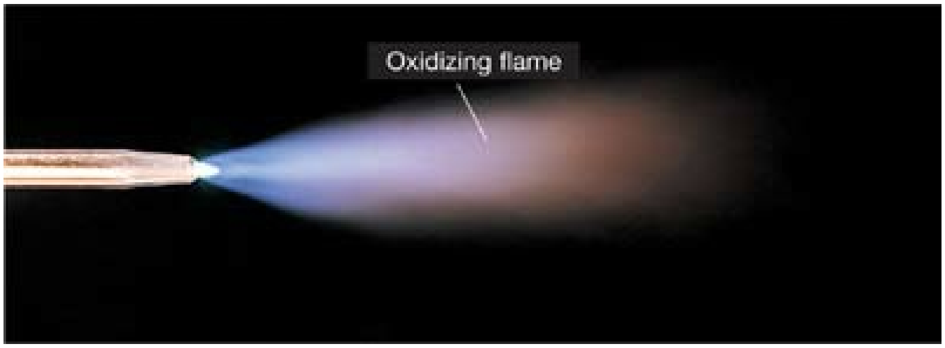

The oxidizing flame has an excess of oxygen. The white cone of this flame is small and pointed and somewhat paler than the neutral flame. A hissing sound often accompanies this flame. This flame is not particularly useful as it hastens oxidizing, which is not desirable in welding. It can, however, be used for removing carbon from molten metal, thus softening the metal.

Prewelding Checklist

- Check hoses for damage before pressurizing the system.
- Prepare metal for welding by wire brushing or sanding off mill scale and rust. Use acetone or denatured alcohol to remove oil or other chemical residues.
- Use fire bricks to avoid unnecessary heat loss and prevent welding to the welding table.
- See manufacturer's recommendations for appropriate tip sizes and gas pressures.
- Set up materials and clamp if necessary.

NOTE: The directions for oxy-acetylene welding are for right-handed welding. Reverse the directions for left-handed welding, or if you find it easier to manipulate the filler rod with your right hand.

How to Weld with Oxyacetylene

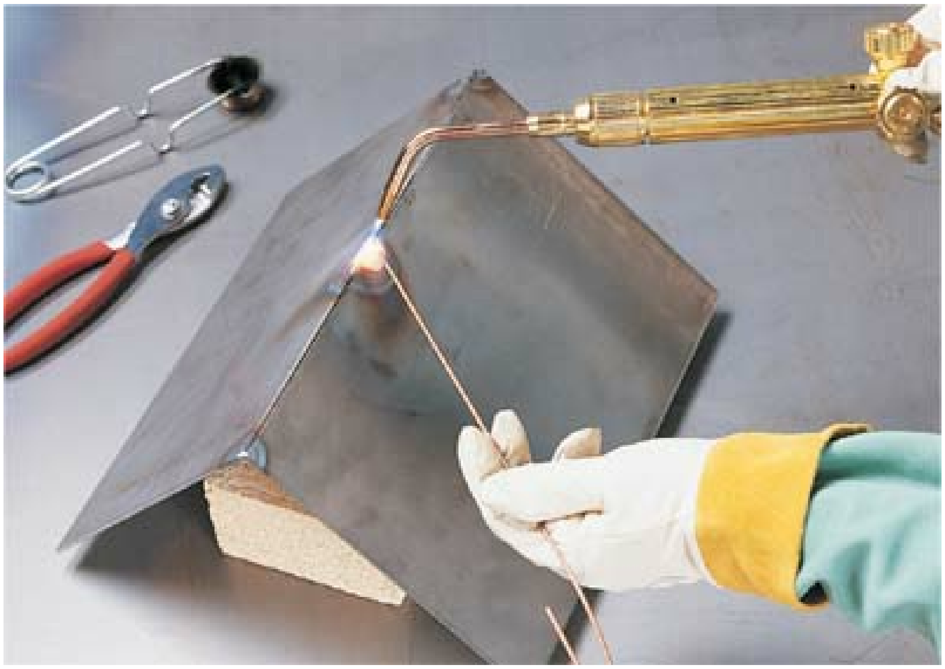

1 Select an appropriate filler rod and lay it on the table next to the bricks. Light the torch and adjust to a neutral flame. Pull down your face shield. Place small fusion tack welds at each end of the joint and in the middle if it is a long joint. (A fusion weld uses no filler rod.) Turn off the torch, oxygen first then acetylene, and check that your tacked piece is still in the desired position. If not, use a hammer to move it into position or break the tack weld and reposition.

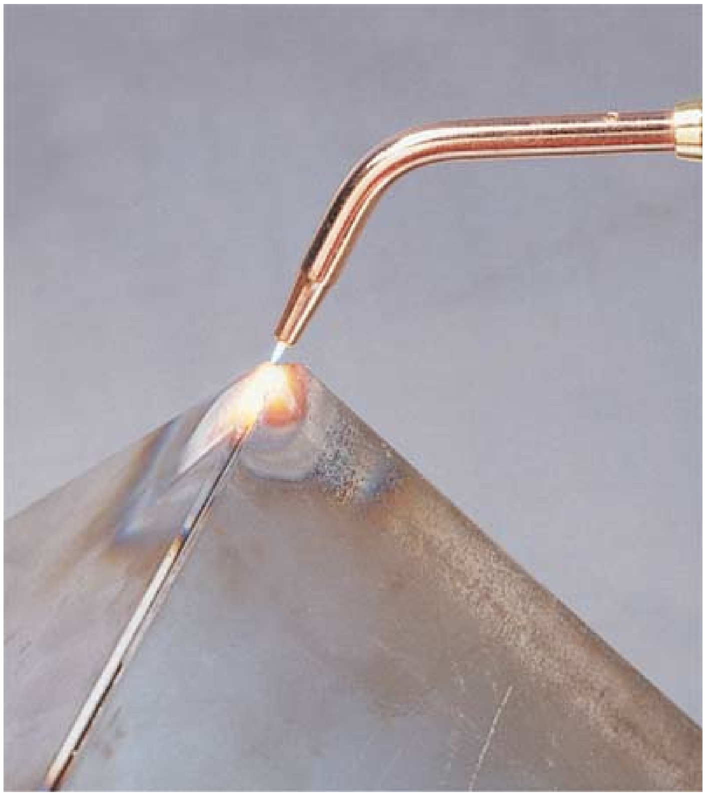

2 With the torch at a 45° angle to the right and oscillating the torch in a ¼ to ½" circle over both metal pieces, create a weld puddle at the right end of your workpiece.

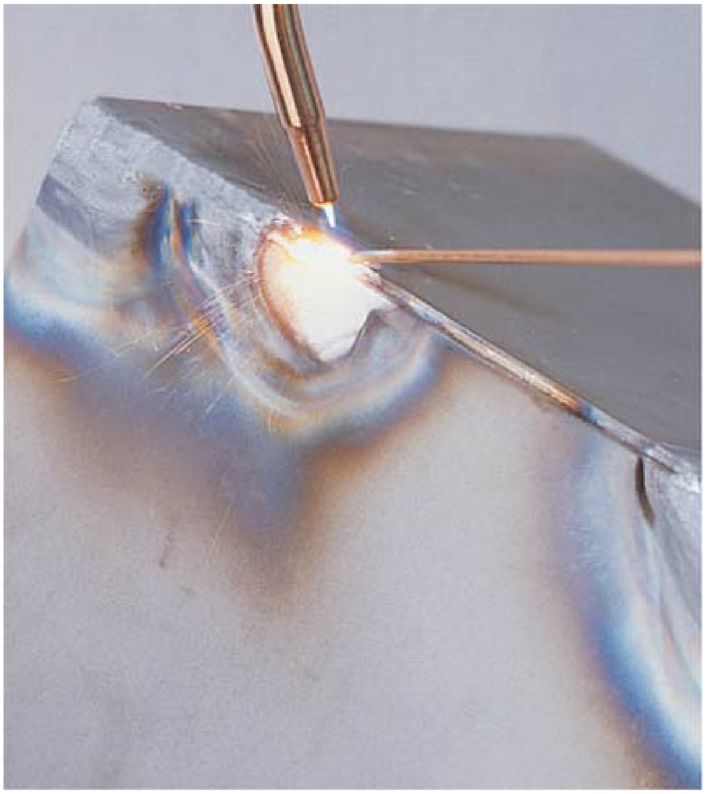

3 When a molten puddle has formed, move the filler rod close to the puddle and flame, but not in it. Begin moving slowly to the left while oscillating and maintaining the molten puddle. Dip the filler rod into the middle of the molten puddle and remove it, but keep it within the heat zone.

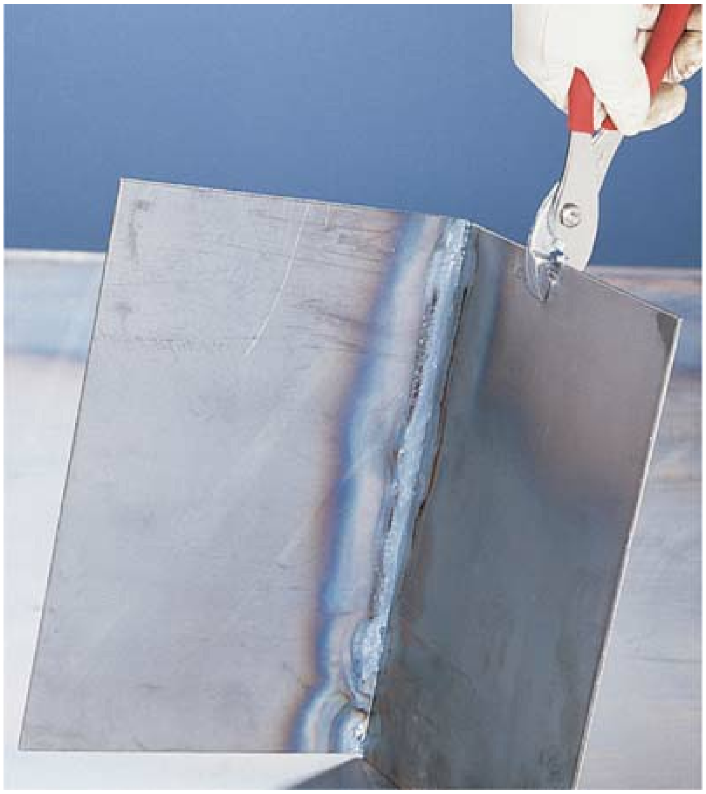

4 Continue dipping, oscillating, and moving to the left. As you reach the end of the weld, the cumulative heat buildup may make it necessary to adjust to a shallower angle to deflect heat away from the puddle and prevent burn through. When finished, turn off the oxygen torch valve first, then the acetylene torch valve. The weld should penetrate to the back without burning through.

Depressurizing an Oxyacetylene Outfit

When you have finished welding, or if you are going to stop welding for more than 10 or 15 minutes, depressurize your setup. The hoses and regulators are designed to leak small amounts of fuel and oxygen if they are pressurized and not being used, so it is important for safety and economy to depressurize them. You will soon be so comfortable with the pressurizing and depressurizing steps that it won't seem like an inconvenience at all. It is important to do the fuel gas and oxygen in separate steps, to prevent having mixed, unburned oxyacetylene in the torch.

- Close the fuel gas cylinder valve.
- Open the fuel gas torch valve until both gauges read zero.
- Loosen (counterclockwise) the fuel gas regulator adjustment screw.
- Close the fuel gas torch valve.
- Close the oxygen cylinder valve.
- Open the oxygen torch valve until both gauges read zero.
- Loosen (counterclockwise) the oxygen regulator adjustment screw.
- Close the oxygen torch valve.

Oxyfuel Tips

Gas Options:

The most useful fuel gas for oxyfuel welding is acetylene. Its neutral flame generates the highest heat and heat concentration with the lowest chemical interactions with the molten metal. Other fuel gases, such as MPS or MAPP and propylene can be used for brazing and cutting applications. Natural gas and propane can be used for brazing and heating operations. Make sure that the equipment—hoses, regulators, torches, and tips—are designed for use with the gas you choose.

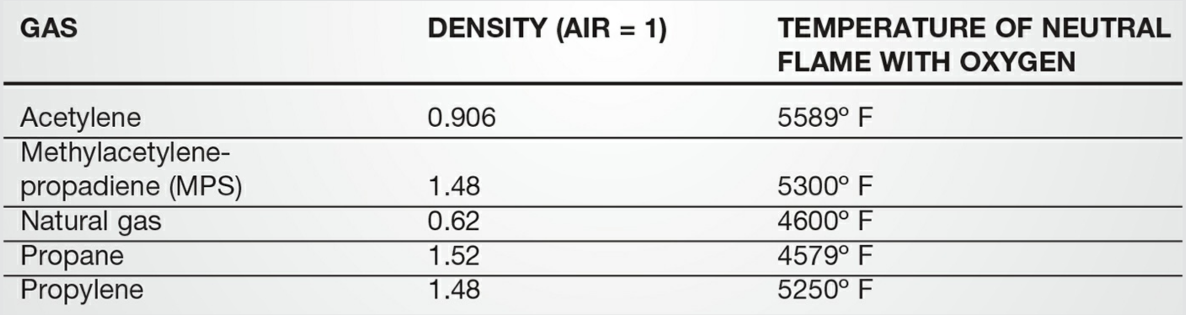

GAS	DENSITY (AIR = 1)	TEMPERATURE OF NEUTRAL FLAME WITH OXYGEN
Acetylene	0.906	5589° F
Methylacetylene-propadiene (MPS)	1.48	5300° F
Natural gas	0.62	4600° F
Propane	1.52	4579° F
Propylene	1.48	5250° F

Fire Bricks:

Fire bricks are specially made bricks that have little water content so they will not explode when heated to high temperatures. Fire bricks do not pull the heat from your weld or braze, and you cannot weld your material to them.

Maintenance. The primary maintenance for oxyacetylene welding equipment is cleaning the torch tip. A dirty tip will spark, pop, and often direct the flame sideways. Tip cleaners are inexpensive sets of cleaning rods used to clean the various sized tip orifices (see photo, page 43). The set usually comes with a small file to smooth the tip. To clean, simply insert the correctly sized tip cleaning rod into the orifice, and pull back and forth. Do not rotate the tip cleaner as this may enlarge the orifice. Also, do not use drill bits to clean tips. The sharp edges of the bit will cut grooves into the tip. Cleaning rods have smoothed edges and are sized especially for tips. Do not clean tips while the torch is lit.

Brazing

Brazing is very similar to soldering since flux is applied to tightly fitted metal parts that are then heated to the point where filler material will melt and be drawn into the joint. Silver soldering and hard soldering are terms incorrectly used to refer to brazing. Brazing is different from soldering because it takes place at temperatures over 840°F and below the melting point of the base metals. The metals are not fused, but held together by the filler metal adhering to the base metals through capillary action.

There are a number of industrial brazing processes, such as dip brazing, furnace brazing, and induction brazing. The home welder is likely only to do torch brazing. Torch brazing can be done with an oxyfuel torch using acetylene as the fuel gas, or any of the other fuel gases (see page 49). NOTE: Different fuel gases require different regulators, hoses, torches, and tips.

For brazing to work, the gap between the parts must be between 0.002 and 0.010". If the gap is too tight, the flux and filler will not flow evenly through the joint. If the gap is too big, the strength of the joint is lessened. Gaps between parts can be measured with a feeler tool, available at automotive stores. Items to be brazed must be absolutely clean and free from rust, corrosion, grease, oil, and cleaning compound residues.

Brazing is often used commercially to join dissimilar metals such as tungsten carbide saw teeth to a steel saw blade. Another common use for brazing is in lugged bicycle frames. Because nearly every metal can be joined using brazing, it is highly suitable for art applications.

PETERSON
NO. 1 BLUE
Stay-Silv
15
3/32
1 LB

Brazing supplies include flux and silver solder (left). Braze welding requires flux-coated rods or separate flux and rods (sold individually or in bulk).

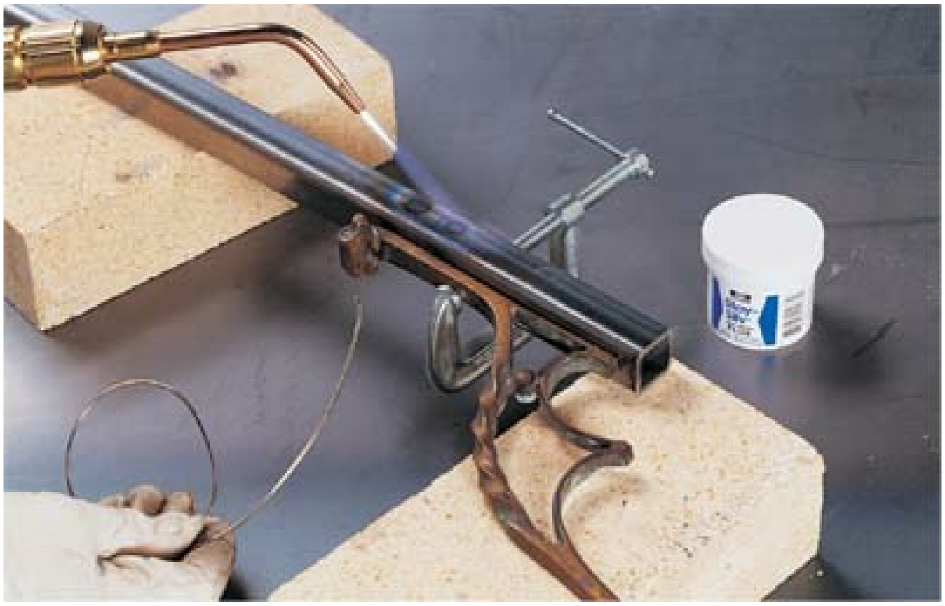

Thoroughly clean and flux both sides of the joint area. Using a small torch tip, heat the entire joint area until the flux turns clear and starts to run. Add enough filler metal to fill the joint. (Silver alloy is shown.) After the metal has cooled, the flux residue can be removed with hot water.

Braze Welding

Braze welding is similar to standard oxyacetylene gas welding except the parent or base metals are not melted, so there is no molten puddle. Instead of a steel alloy filler rod, a flux coated brass filler rod is used. Braze welding is often incorrectly referred to as brazing. Braze welding does not use capillary action to pull filler material into the joint—the filler metal is deposited as fillet or groove welds.

The brazing rod is melted by the heat of the metal and the flame, but it should not be held in the flame itself. The parts for braze welding should fit tightly, but the gap is not as important as with brazing.

Braze welding is used for joining dissimilar metals and for metals of different thicknesses. This technique is often used to repair cracked or broken cast iron.

Braze welding has less distortion than oxyacetylene welding because less heat is applied to the parts. A disadvantage is that it is not as strong as welding where the base metal is melted, but a well-made braze weld is still sufficient for most non-structural applications. Because the base metals do not need to be melted, braze welding can be done with any of the fuel gases listed on page 49.

If you are creating a piece that will be welded and braze welded, you must be careful to complete the non-brazed welds first. The heat involved with all other welding processes will boil off the brass alloy of a braze weld, ruining the weld and creating toxic fumes.

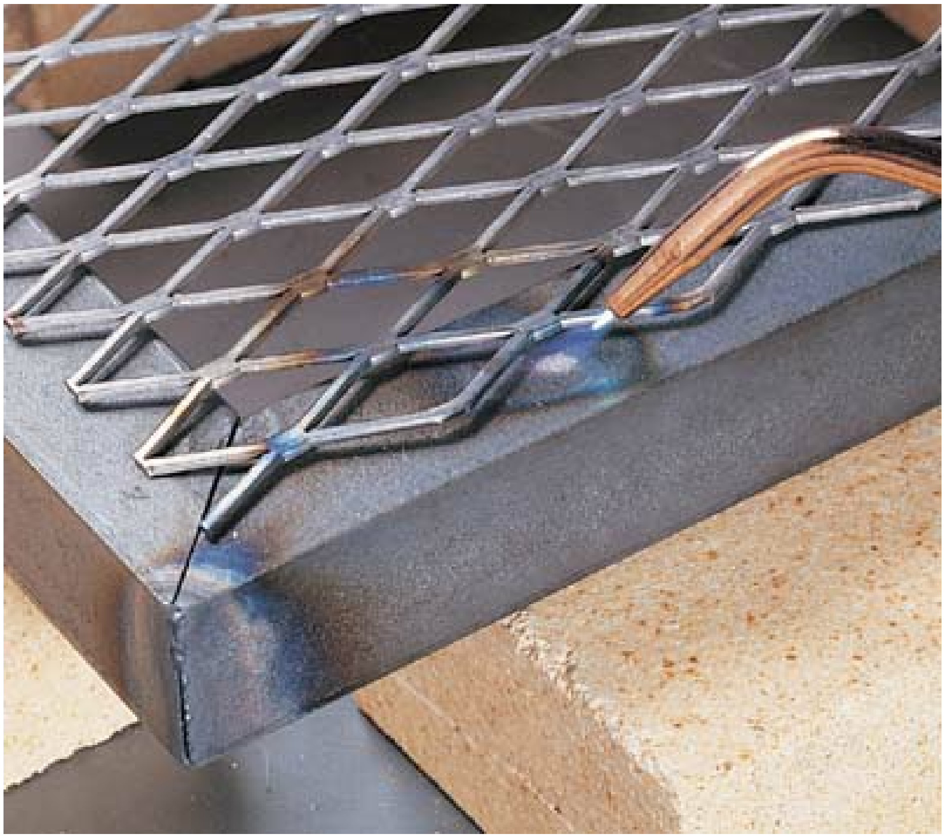

1 Braze welding is useful for joining thin metals, like this expanded metal, to thicker metals. Heat both parts, directing more heat toward the thicker part. It may take a long time for the thicker metal to heat. Using fire bricks will prevent a metal tabletop from absorbing any heat.

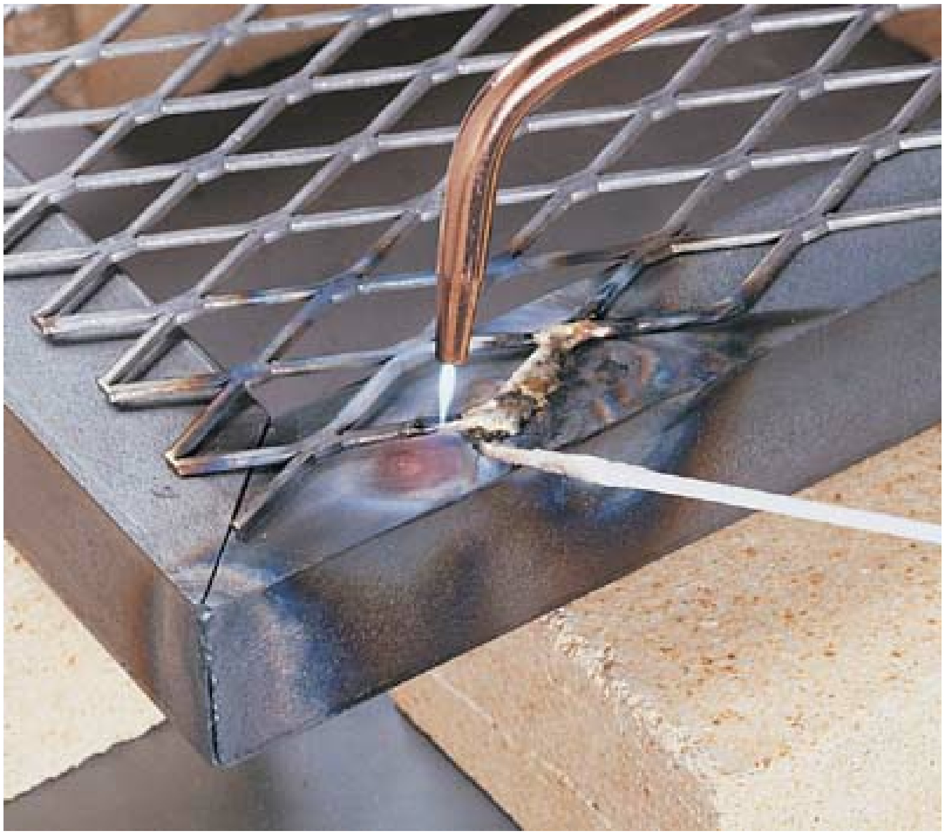

2 When both parts glow a dull cherry red, touch the flux coated rod to the joint. The flux and the filler metal will melt. If the metal is molten or the fluxed rod comes in contact with the flame, the flux will burn and the filler metal will boil. This results in a poor joint in addition to giving off toxic fumes.

Electricity For Welding

Arc welding and cutting processes—shielded metal arc, gas metal arc, gas tungsten arc welding, and plasma cutting—all use electricity to generate the necessary heat. Understanding electricity is not necessary to use these processes, but knowing basic terms and concepts will help you understand why certain welding events occur.

Electric current is the flow of electrons through a conductor from a high concentration of electrons (negative charge) to a low concentration of electrons (positive charge). As electricity flows through a conductor, it generates heat based on how much resistance the conductor offers. An arc is simply a sustained electrical discharge across an air gap. Because air is highly resistant to electron flow, an enormous amount of heat is generated by the movement of electrons across this gap. The heat generated by an arc is in the area of 11,000° F, but about half of this heat is dissipated.

Electrical units. The units used to refer to electricity are voltage, amperage, and wattage.

> **Voltage or volts (V)** is the measure of electric potential or the electric pressure. The voltage controls the size of the air gap that the arc can cross. The higher the voltage, the larger the gap the arc can cross.
>
> **Amperage or amps (A)** is the volume of electrons flowing through a conductor. The amperage controls the size of the arc.
>
> **Wattage or watts (W)** is the measurement of the amount of electrical energy or power contained in the arc. Wattage affects the width and depth of a weld.

Welding power. Power for welding may be supplied in one of the following ways:

> **Constant voltage (CV)**. The arc voltage is maintained even when the current (amps) changes. This means if the arc length is changed, which changes the current flow, the voltage stays relatively the same.
>
> **Constant current (CC)**. As the amperage fluctuates, the voltage increases or decreases to keep the total power (wattage) output the same.

Open circuit and operating voltage. The open circuit voltage is the voltage that exists at the electrode tip when the machine is on but no arc has been struck. The higher the open circuit voltage, the easier it is to strike an arc. Open circuit voltage is usually between 50 and 80 volts. The higher the voltage, the greater the risk of electrical shock.

The operating voltage is the voltage when the arc is struck and the circuit is completed. This is usually between 17 and 32 volts depending on the arc length, the type of electrode, the type of current, and the polarity.

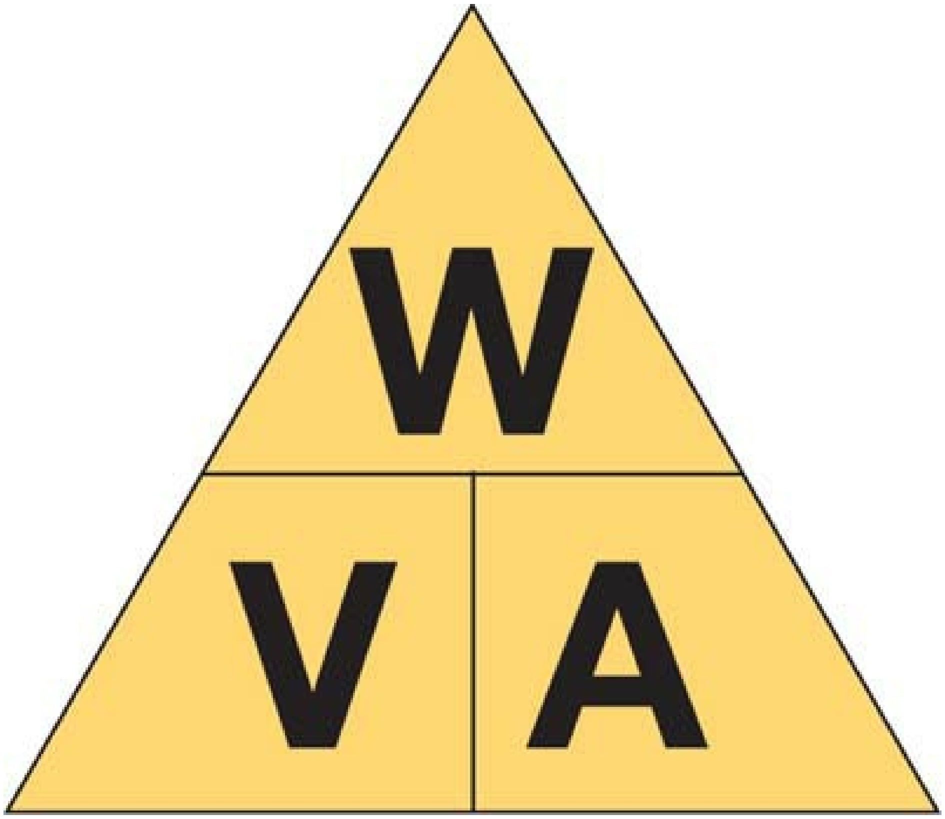

Volts, amps, and watts are related to each other in such a way that volts × amps = watts.

Power sources for welding. There are three sources for producing the low-voltage high-amperage combination that arc welding needs.

A mechanical generator produces power through the use of a gasoline or diesel engine.

A step-down transformer takes available high-voltage alternating current and changes it to a low-voltage high-amperage current.

An inverter uses solid-state electronics to change the current without the weight of a transformer.

Currents used for welding. The type of current used and the polarity is important to welding.

Alternating current (AC) is standard available household current. The electron flow changes direction two times per cycle. Welding with AC means that the electrode and the workpiece alternate between positive and negative, so the welding heat is distributed evenly to both. This balances penetration and buildup.

Direct current (DC) electrons flow in one direction only. A rectifier is used to convert alternating current to direct current. With direct current, there are two polarity options:

Direct current electrode negative (DCEN)— formerly called straight polarity—means the electrode is negative, the workpiece is positive; electrons are flowing from the electrode to the workpiece and the heat is concentrated in the electrode.

Direct current electrode positive (DCEP)— formerly called reverse polarity—means the electrode is positive, the workpiece is negative; the electrons are flowing from the workpiece to the electrode. DCEP produces the best welding arc characteristics.

Duty cycle. The duty cycle is the amount of continuous running time at a given power output for a welding machine, given as a percentage for a ten minute period. Because welding machines produce heat internally, they need a certain amount of cooling time. A 60% duty cycle means that a machine can run continously at that setting for six minutes, then it will need to cool for four minutes. Duty cycle is critical in construction and industrial applications, and it can be important for home shop use. A 10% duty cycle at the maximum setting means one minute of welding and nine minutes of cooling. This could be a problem for large projects on thick material. The higher the duty cycle, the more expensive the machine.

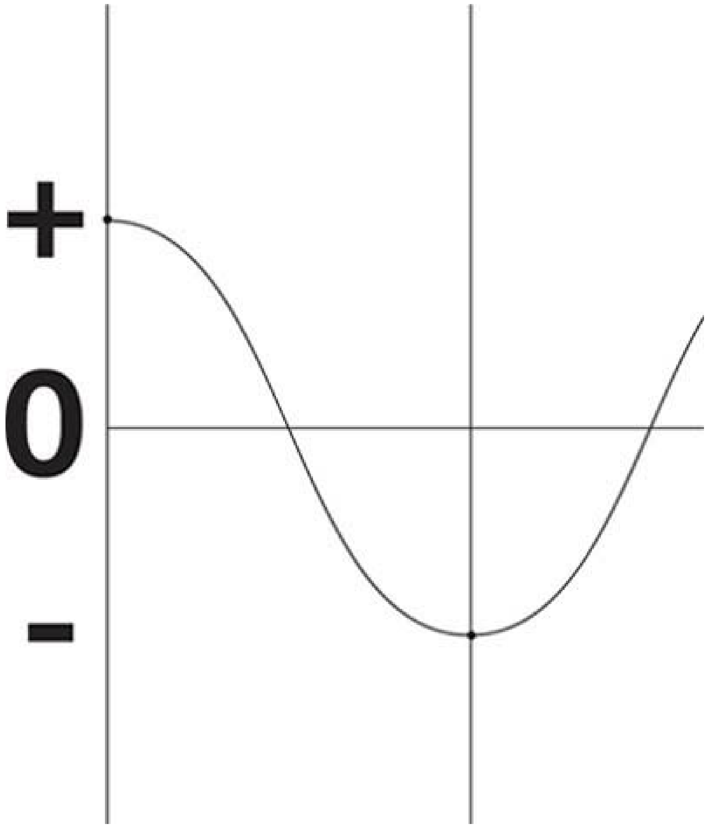

Alternating current alternates between positive and negative polarities, passing through a zero, or no current point, between the two.

Note

Work leads and work clamps are sometimes referred to as ground leads and ground clamps. This is not technically correct, as the electricity is not being grounded. Instead, the work lead and clamp are allowing for completion of the circuit back to the machine.

Shielded Metal Arc Welding (SMAW)

Shielded metal arc welding (SMAW) is also referred to as arc or stick welding. The process involves the heating of the base metal to fusion or welding temperature by an electric arc that is created between a covered metal electrode and the base metal. The coating or covering on the electrode provides both flux and shielding gas for the weld. The electrodes come in 9 to 15" straight lengths in a range of wire thicknesses from $^{1}/_{16}$ to 3/8", hence the name “stick.” SMAW is used extensively for fabrication, construction, and repair work because the machinery is inexpensive and fairly simple, and the electrodes are inexpensive. There are drawbacks to SMAW: it doesn’t work well on thin materials (less than 1/8" is difficult), electrodes need to be changed frequently as they are used up, and the protective slag coating must be chipped off each weld.

Safety: Shielded metal arc welding uses electricity, so there is always the possibility of receiving an electric shock or being electrocuted. When an electrode (stick) is placed into the electrode holder, it is “live." If the electrode touches anything that the work clamp is in contact with, the circuit will be completed and an arc will be struck. To prevent this from happening, always remove the electrode from the holder when you are not actively welding. Do not use your bare hands to insert or remove electrodes—always touch electrodes with dry gloves. Remember that water and electricity never mix well, so do not weld while standing on a wet or damp floor or ground, and do not weld outdoors in the rain. The electrode will be hot after use, so take care where you dispose of electrode stubs. A metal bucket is a good addition to your SMAW workshop.

SMAW welding produces ultraviolet and infrared rays, harmful fumes, and hot spatter. Protect your eyes with a #10 to #14 filter in a full-face welding helmet or hood. Heavy-duty leather welding gloves and a welding jacket with leather sleeves are necessary to protect you from the molten spatter. Proper ventilation from an exhaust hood or fan is important since many of the chemicals used in the electrode coatings are hazardous if inhaled. Also, it is a good idea to screen off your welding area so others are protected from the intense light of the arc.

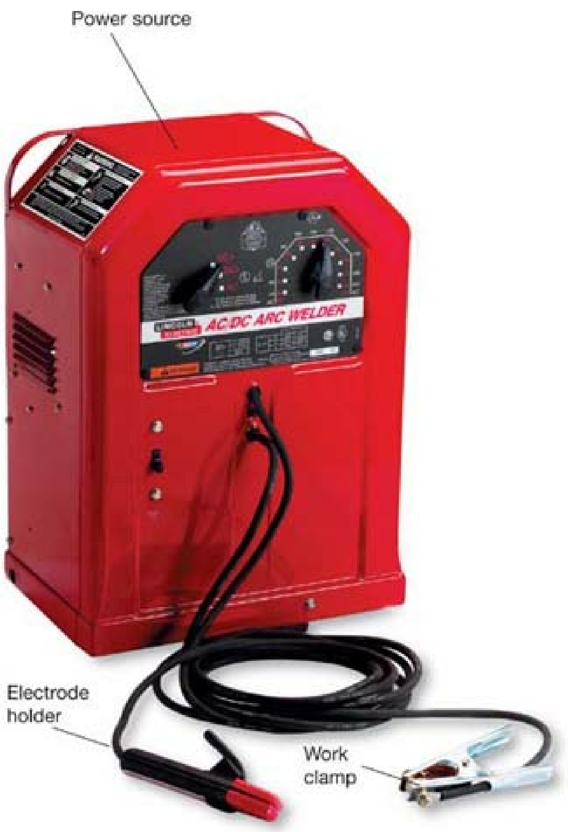
Power source
AC/DC ARC WELDER
Electrode holder
Work clamp

A shielded metal arc welder consists of a power source, electrode holder, and work clamp.

Equipment: SMAW machines are available as either alternating current (AC) or direct current (DC), or with the capability of switching between the two. The machine itself is simple in that it merely converts high-voltage low-amperage line current into low-voltage high-amperage welding current. Output is controlled with one knob. You often will hear SMAW machines referred to as “buzz boxes.”

AC welding machines meet most home and small shop needs, are inexpensive, and are readily available. Because the alternating current cycles through a zero current between the positive and negative polarities, it can be difficult to strike and maintain an arc. DC machines are easier to use and have many home and hobby applications, but they are more expensive. Because DC current can have its flow reversed (see page 53), a DC machine has more versatility in terms of the electrodes that can be used. This allows for a wider range of welding positions, metal thicknesses, and metal types that can be welded. This versatility makes a DC shielded metal arc welder well worth the extra expense.

The equipment itself consists of the welding machine, which usually has one adjustment knob, an electrode lead with electrode holder (sometimes called a stinger), and a work lead with work clamp. You will often see work leads and work clamps referred to as ground leads and ground clamps. The work lead and clamp are not grounding the electricity, they are completing the circuit back to the machine.

Safety

- Welding helmet with #10 to #14 filter
- Leather, wool, or cotton long pants; leather jacket
- Heavy-duty welding gloves
- Safety glasses
- Hat
- Leather boots or shoes
- Ventilation

Useful tools for shielded metal arc welding are a pliers, chipping hammer, and wire brush.

SMAW electrodes: SMAW electrodes are solid, round, metal wires coated with flux and other components. In addition to producing shielding gas and flux, the covering may also contain additional metals for filler or alloying elements for the weld.

The American Welding Society (AWS) publishes standards for the electrodes. Electrodes come in diameters ranging from $^{1}/_{16}$ to 3/8" in increments of $^{1}/_{32}$". The electrode diameter measures the wire itself, not the diameter of the wire with the covering. The electrode designation is inked onto the covering near the bare end of the electrode. The number classification for SMAW electrodes begins with the letter E because they are electrodes. The first two or three numbers on the left denote the tensile strength of the properly completed weld in thousands of pounds per square inch. The second number from the right indicates the welding position appropriate for that electrode: 1 = all, 2 = flat grooves and flat or horizontal fillet weld, 3 = flat only, 4 = all. The last two numbers designate the current and polarity uses and other special notes. There also may be electrode suffixes that denote alloys that have been added to the electrode.

Each electrode manufacturer may have a number of electrodes of a specific AWS designation that are slightly different and have been tailored to specific uses. The diameter of the electrode will determine which amperage to use. Use the manufacturer's guidelines to determine amperage. If not available, start with 90 to 120 amps minimum range for a 1/8" electrode and add 40 amps for each $^{1}/_{32}$" increase in diameter. This will give you a rough minimum starting amperage.

The most commonly used electrodes are 6011, 6013, 7014, and 7018. The 6011 and 6013 electrodes will work with AC power, the 7000 electrodes will not. For a beginner, 6013 with DC power is generally the easiest electrode to use in terms of striking an arc and maintaining a consistent arc. It is always a good idea to talk with your welding supplies dealer about what type of welder you are using, the type of welds you are making, and the materials you are working on to get the best electrode for your purposes.

Slag: Shielded metal arc welding produces a weld that is covered with a coating of ceramic-like slag. The flux and other components in the covering clean the material to be welded and also float out impurities in the weld. These impurities and flux solidify on top of the weld, which protects the cooling weld from the effects of oxygen in the atmosphere. The slag must be scraped or chipped away before the weld bead is covered with another weld layer or before the weld is painted or finished. Safety glasses should be worn during this procedure.

HOBART
STICK ELECTRODES
HOBART
STICK ELECTRODES

SMAW electrodes are sold in 5- or 10-pound boxes. Specialty electrodes may be bought a pound at a time. It is important to keep electrodes dry either in the original box or a storage container.

6013
7018
6011

Electrodes for SMAW are stamped with a numeric code so you can tell which to use with arc power and which are for DC.

Prewelding Checklist

- Electrode and work leads are tightly attached to machine terminals.
- Cable insulation is free of cuts, damage, and wear.
- Electrode lead is tightly connected to electrode holder.
- Work lead is tightly connected to clamp.
- Work clamp is firmly clamped to worktable or workpiece.
- Floor is dry.
- Ventilation is adequate to draw fumes away from welder’s face.

How to Weld with Shielded Metal Arc

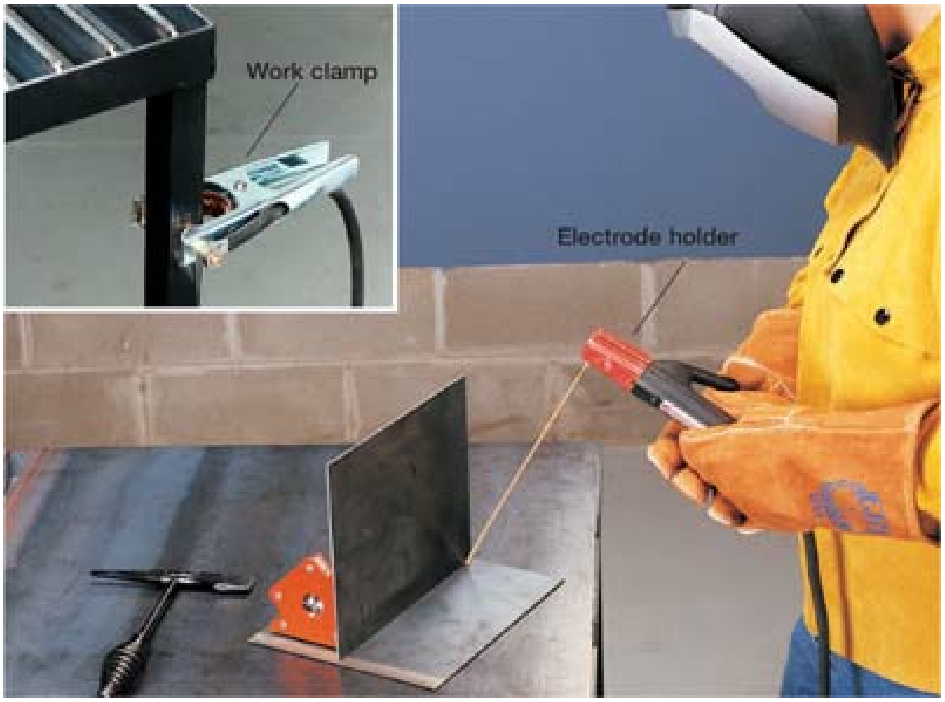

1 Set up your material to be welded. Make sure the electrode holder is not touching the workpiece or worktable. Attach the work clamp to the table (inset) or workpiece. Turn on the machine. Adjust the range switch for the desired amperage. Wearing leather gloves, place an electrode in the electrode holder. Position the electrode over the area to be tacked, flip down your helmet, and tap or scratch the electrode on the area to be tacked to strike an arc. After making your tack welds, remove the electrode from the electrode holder. Check to see that the tacked pieces are aligned properly. If not, use a hammer to move them into alignment or break the tack welds and retack. Chip slag from tack welds so it does not contaminate the final weld.

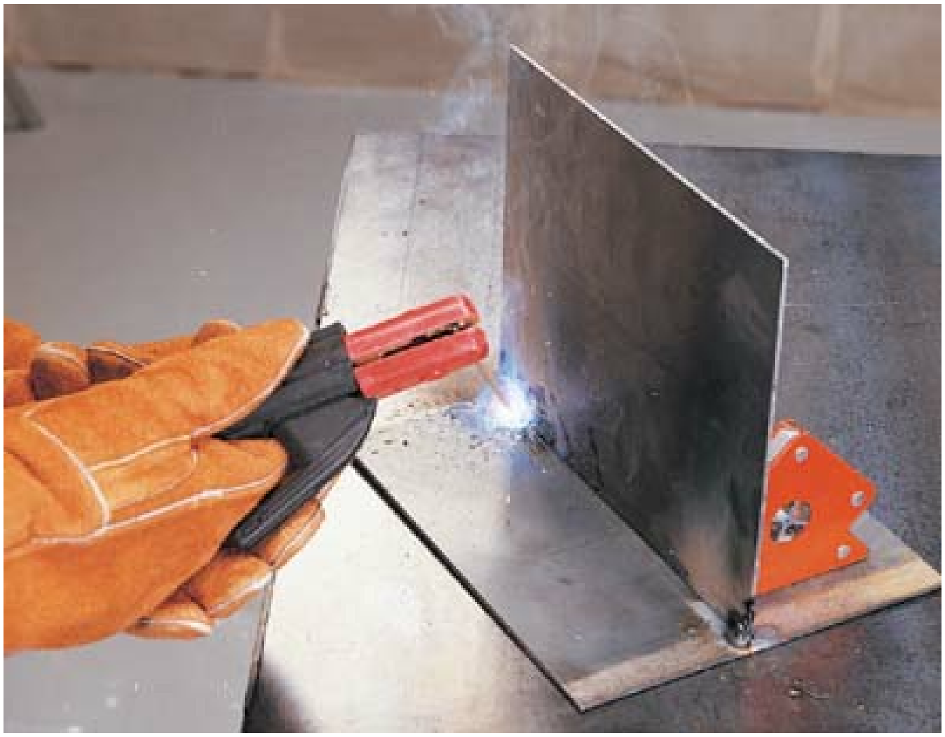

2 Replace the electrode in the electrode holder, and position the electrode over the left side of the area to be welded. Hold the electrode at a 10 to 20° angle to the right. Flip down your hood and scratch or tap to strike an arc. The distance between the metal and electrode should not exceed the thickness of the electrode bare wire diameter. Move slowly to the right until the weld is completed.

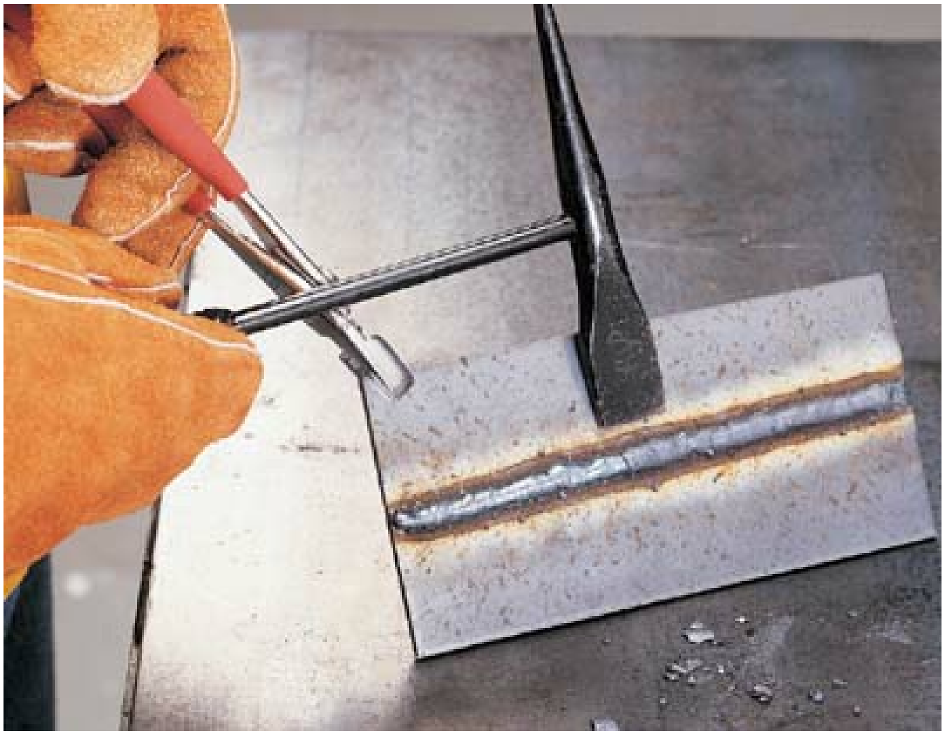

3 To remove slag, hold the workpiece with pliers at an angle and scrape or knock the slag with the flat blade of the chipping hammer. Wear safety glasses when chipping slag.

Gas Metal Arc Welding (GMAW/MIG)

Gas metal arc welding (GMAW), also referred to as MIG (metal inert gas) or wire feed, is a process where a consumable electrode (wire) is automatically fed through a welding gun along with a continuous flow of a shielding gas. The actual application of filler metal is achieved with most small machines through the short circuit transfer method. Other transfer methods used by more powerful machines are short circuit globular and spray pulse.

The GMAW process has a number of advantages. The gun or torch can be held at a uniform distance from the weld, unlike shielded metal arc welding where the distance from the electrode holder to the weld becomes shorter as the electrode is consumed. The trigger power control on the welding gun also makes it possible to be completely positioned and ready to weld without accidentally striking an arc. When ready to weld, the welder can flip down the helmet and pull the trigger without getting off target. Because the electrode is the filler, the welder does not need to coordinate a filler rod in one hand and the torch in the other. The GMAW gun can be held steady with both hands to create a uniform bead. Because of the shielding gas, GMAW welds are smooth and clean, with no slag to chip or grind. The GMAW process can operate at very low amperages and is a relatively cool welding process, so 22- and 24-gauge sheet metals can be welded without distortion. The process also requires narrower beveling for thicker plate welds, so less time is spent on grinding.

Some disadvantages of GMAW are that the shielding gas nozzle does not allow for welding in tight spaces without some modifications. Because the shielding gas can be disrupted or blown away by drafts easily, the process is not suited for working outdoors.

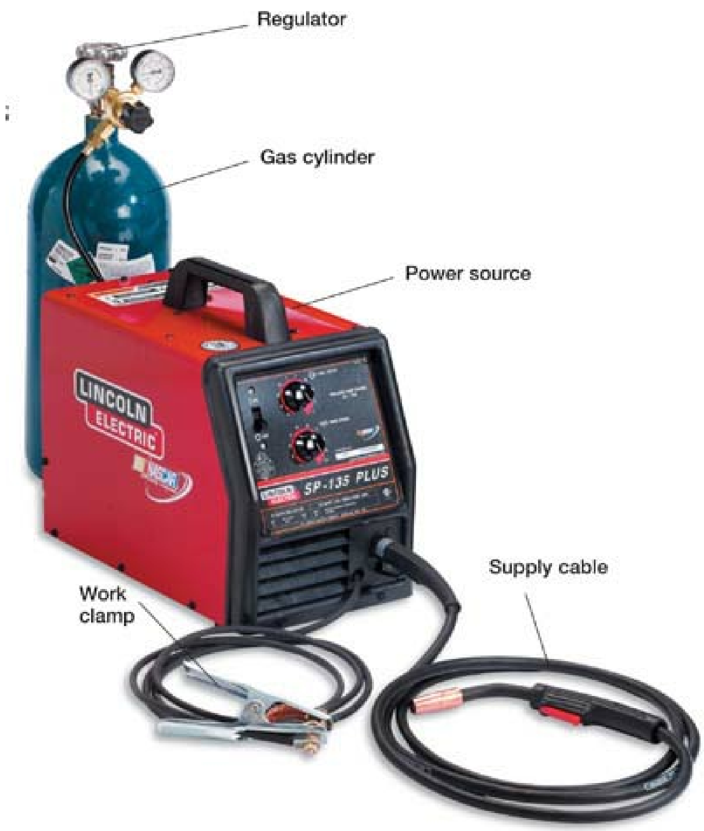
Regulator
Gas cylinder
Power source
LINCOLN ELECTRIC
SP-135 PLUS
Supply cable
Work clamp

A gas metal arc welding setup consists of a power source, wire feed (inside power source), work cable with clamp, supply cable with gun, and gas cylinder with regulator.

Safety

- Welding helmet with #11 to #12 filter
- Leather, wool, or cotton long pants, long sleeve shirt, and hat
- Leather gloves with gauntlet
- Leather boots or shoes
- Ventilation

The Equipment

Power supply: In the small GMAW units, the power supply and wire feed are integrated into the same cabinet. In larger multipurpose machines, the wire feed is a separate unit. The power supply converts the standard alternating current (AC) into direct current (DC). Alternating current is not appropriate for gas metal arc welding because its constantly switching flow does not yield a steady, consistent arc. On the machine, you will be able to adjust the voltage, polarity, and wire feed rate; the machine adjusts the amperage. When using the power source to weld with a shielding gas, the machine should generally be set up for electrode positive, that is, the electrode is positive and the workpiece is negative. If used without shielding gas and with flux cored wire, the polarity usually needs to be switched to electrode negative. This is achieved by switching the contact wires within the machine. Larger multipurpose machines may have a switch to change the polarity. Always check manufacturer's recommendations for polarity settings.

Wire feed: The wire feed consists of a spool of wire, a tension controller, and rollers and a roller motor. The wire speed is set by the speed adjustment knob, depending on the thickness of the material being welded. The rollers have V or U grooves sized to match particular wires. Most machines come with two-sided rollers that can be switched over to handle two different size wires.

Gun: The GMAW gun, sometimes referred to as a torch, is attached to the welding machine via the supply cable. The cable carries the power, the power control, the wire in a special liner, and the shielding gas. The gun has a trigger that turns the power on and off, which also starts the wire feed and the shielding gas flow. The gun generally has a goose neck shape, although straight neck and flexible neck guns are available. A contact tip is screwed or cam locked into the gun. This tip has an orifice the same size as the wire being used. It is important that this be sized correctly—too large and good electrical contact between the source and the electrode will not be made, too small and the wire will not fit. The shielding gas is directed toward the work area with a nozzle. On most home versions, the relationship between the contact tip and the gas nozzle stickout is not adjustable; larger welders have a wider variety of nozzle sizes and shapes. It may be worthwhile to purchase an aftermarket adjustable system. When using the GMAW process, it is important that the supply cable be kept as straight as possible to prevent kinking the wire or impeding the gas flow.

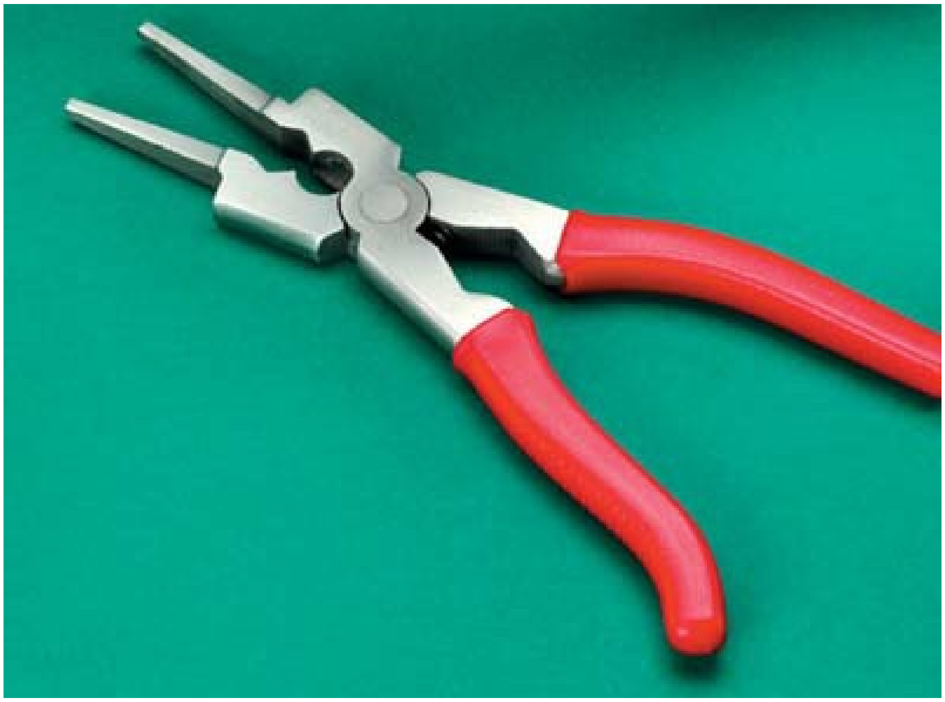

MIG pliers is a handy tool for gas metal arc welding. It is designed for hammering, wire cutting, insulation brushing removal and installation, and drawing out wire. It is also ideal for removing spatter from inside, outside, and the nozzle end.

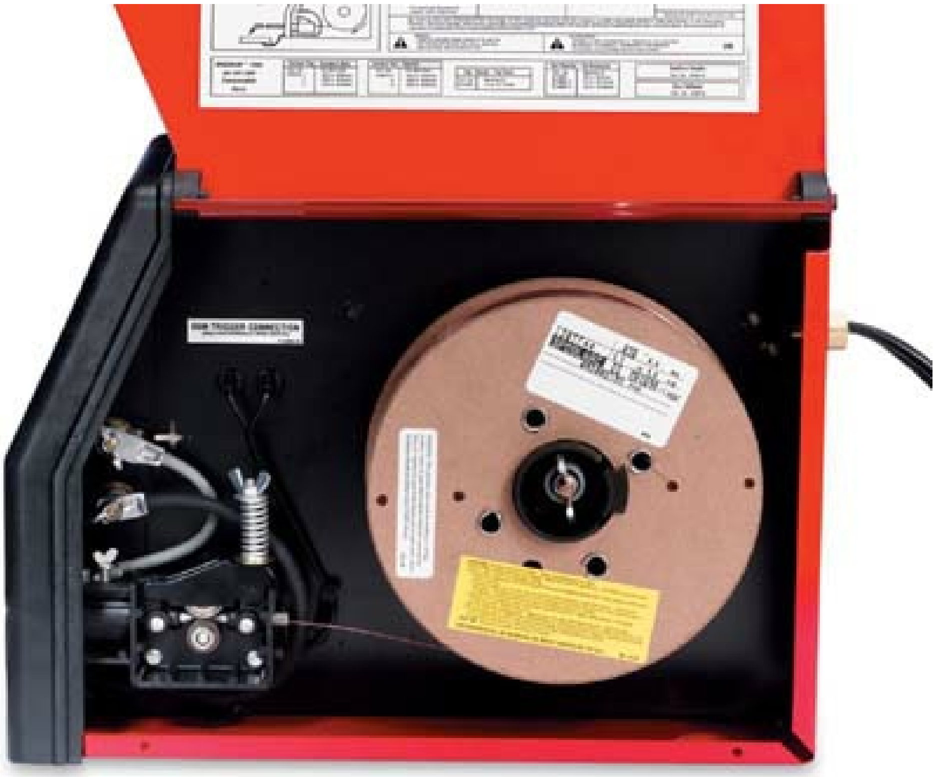

The wire feeding mechanism consists of a spindle to hold the wire spool, drive wheels, rollers, and a tension adjustment. Most welders now have a reference chart with recommended settings.

Most home GMAW welders use 1-pound or 10-pound wire spools. Larger spools are generally more economical.

Work cable with clamps: This is clamped to the workpiece or to a metal work surface that the piece sits on to complete the circuit.

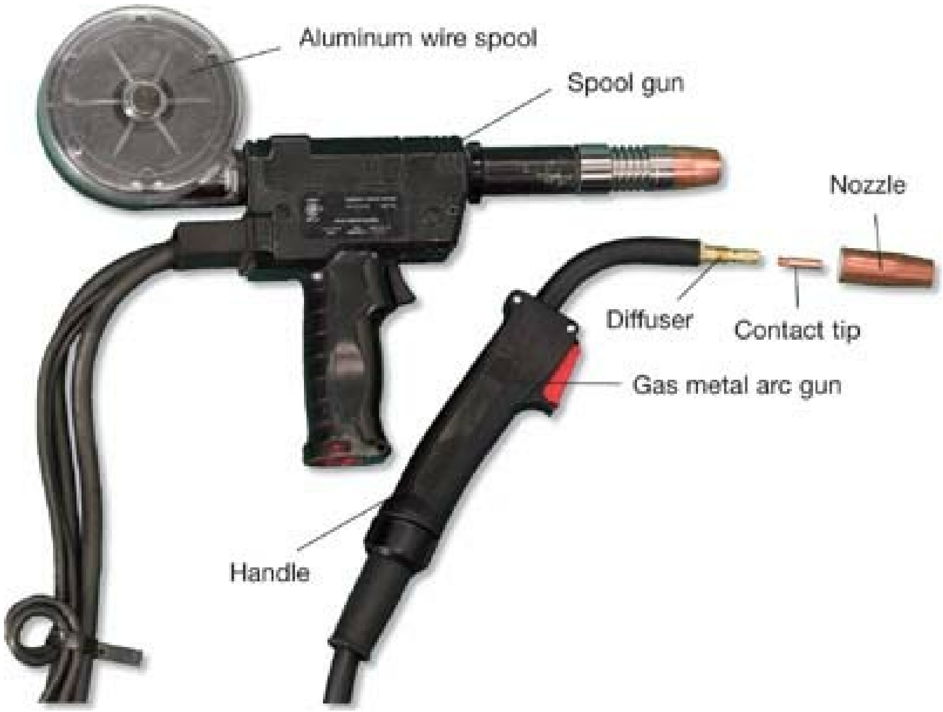
Aluminum wire spool
Spool gun
Nozzle
Diffuser
Contact tip
Gas metal arc gun
Handle

A spool gun is used for feeding aluminum wire because it often breaks or misfeeds through a standard cable. A standard gas metal arc gun consists of a handle, gas diffuser, contact tip, and nozzle.

Electrodes: When choosing an electrode (wire), consider the composition properties, cleanliness of the base metal, and the shielding gas. If you will be welding out of position, this will also be a factor. There are dozens of wire choices for GMAW.

Electrodes are labeled with alphanumeric codes that describe their type, tensile strength of the weld, whether it is solid or tubular (tubular is flux cored), and chemical composition. For mild steel, ER70S-3 is a good general-purpose wire. ER70S-4 and ER70S-6 are good wires for dirty or rusty metals. Wire also is available for welding aluminum and stainless steel. Available wire sizes are 0.024, 0.030, 0.035, and 0.045".

Using a wire feed machine without shielding gas requires the use of flux cored wire (often called innershield), and the process is then called flux cored arc welding (FCAW). Because the wire makes the electrical contact as it travels through the contact tip, the flux must be inside the wire, which makes flux cored wire more expensive than regular GMAW wire. Commercial welders may use a “dual shield” flux cored wire that does use a shielding gas. Welds made with this process are as clean as regular GMAW welds.

Flux cored arc welding can be done outdoors because there is no shielding gas to be disturbed by wind and weather. Unfortunately, using flux cored wire results in a less attractive weld due to the presence of slag and more spatter. The flux cored wire usually has greater penetration and is often used to stretch the capabilities of a small welding machine. Because 0.035" is the smallest diameter for flux cored wire, it is not appropriate for welding thin sheet metal. The lowest setting for this size wire is still too hot for this application.

Shielding gas: The shielding gas for GMAW is supplied by a cylinder with a flow regulator and connecting hoses. The gas line is connected to the welding machine, which directs it through the cable.

Carbon dioxide is suitable for general GMAW welding. It gives good penetration and is inexpensive compared to other gases. Argon and carbon dioxide in a 75% argon and 25% carbon dioxide mix is the standard GMAW shielding gas. It is slightly more expensive than pure carbon dioxide, but it yields welds with less spatter. Pure argon is used for shielding aluminum welds. If you choose to use pure carbon dioxide, select an electrode that performs well with this shielding gas.

Gas cylinder safety: Gas cylinders store compressed gases at pressures up to and greater than 2000 psi. This is sufficient pressure to send a tank through a concrete wall if the cylinder valve should be damaged. It is important to be extremely careful with gas cylinders even when they hold the non-flammable gases used in GMAW welding. Gas cylinders should always be upright and chained or strapped to a wall, post, or cart. Never drag a cylinder; never lay one down or roll it flat. To move a cylinder, chain it to a handcart first. If you must move a cylinder without a handcart, tip it slightly and roll it. Never weld anything to the cylinder or safety cap—and be careful never to accidentally strike an arc against a cylinder. If you do, you will have to purchase that cylinder.

A flow meter registers gas flow with a small floating ball.

How to Set Up the Wire Feed

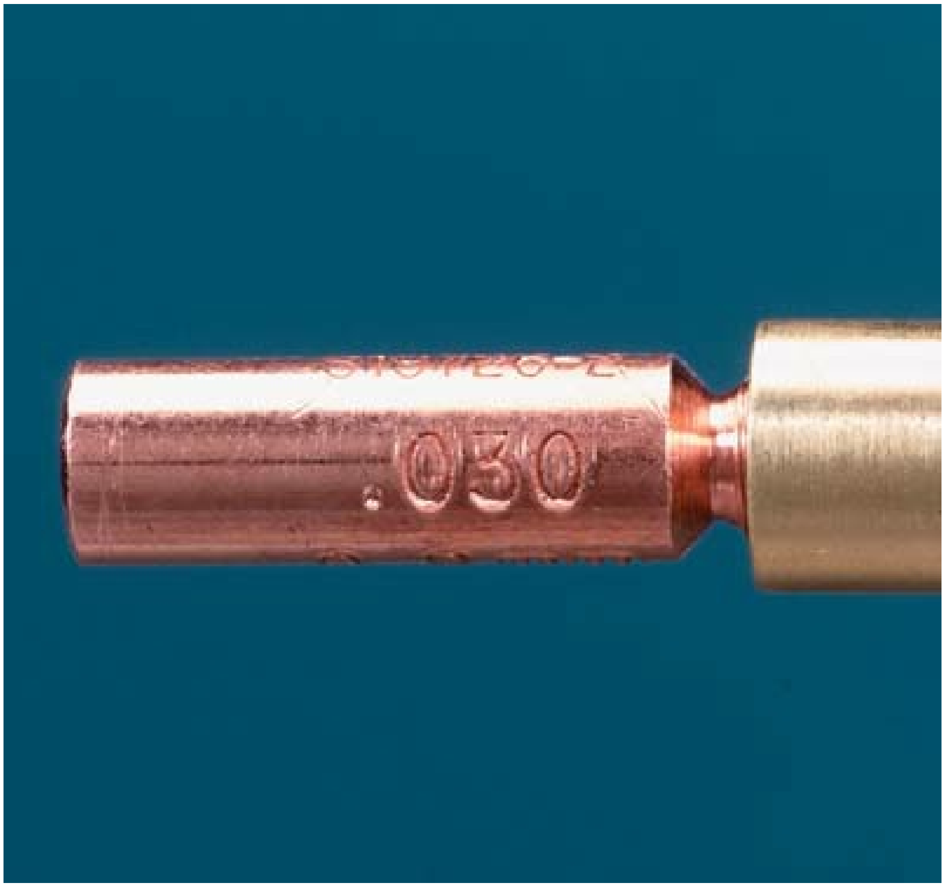

1 Contact tubes are stamped with their size. Make sure the contact tube and the drive roll grooves are the correct size for the electrode you are using. Place the wire spool on the spindle, and secure it with the pin lock, lock ring, or wing nut. Make sure it is feeding in the proper direction.

Note

Unless your machine has an inch switch for advancing wire or a purge switch for activating gas flow, you will use the gun trigger to accomplish both tasks. Using the gun trigger means the wire is “live” and will arc to anything the work clamp is touching.

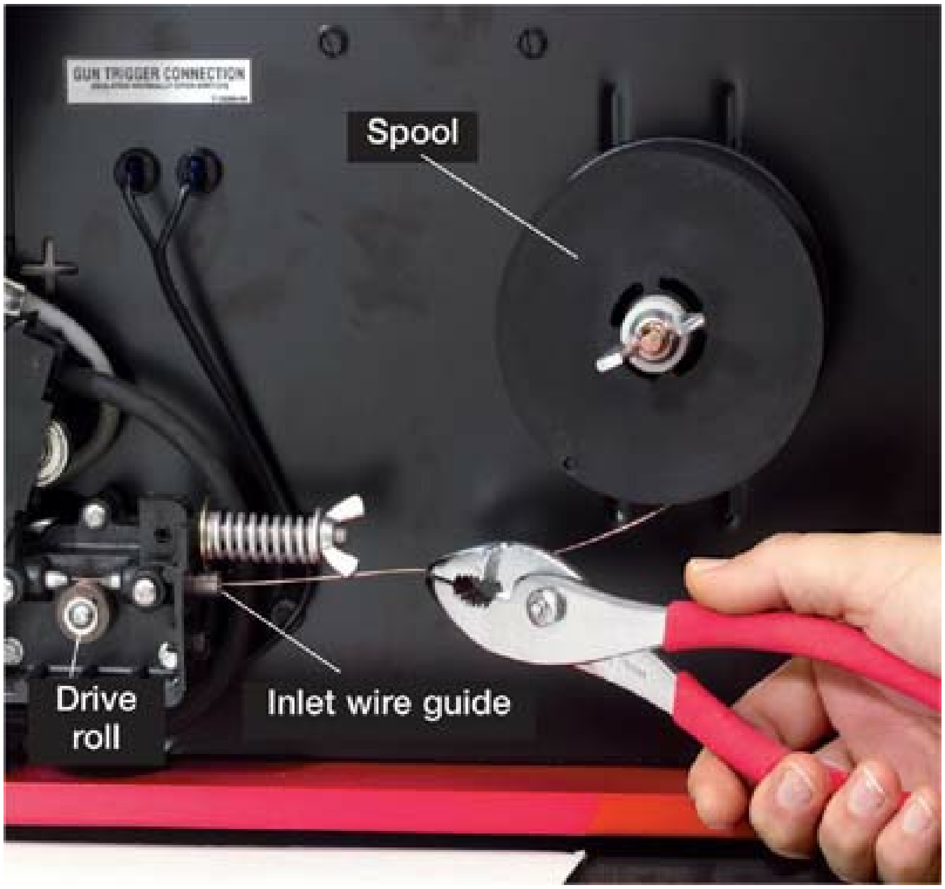
GUN TRIGGER CONNECTION
Spool
Drive roll
Inlet wire guide

2 After releasing and cutting off the crimped wire end, hold the wire firmly with a pliers. The wire is tensioned and the entire spool will rapidly unroll if you do not hold it firmly. Swing the tension arm or pressure roll out of the way. Push the wire through the inlet wire guide, through the groove on the drive roll, and out through the outlet wire guide.

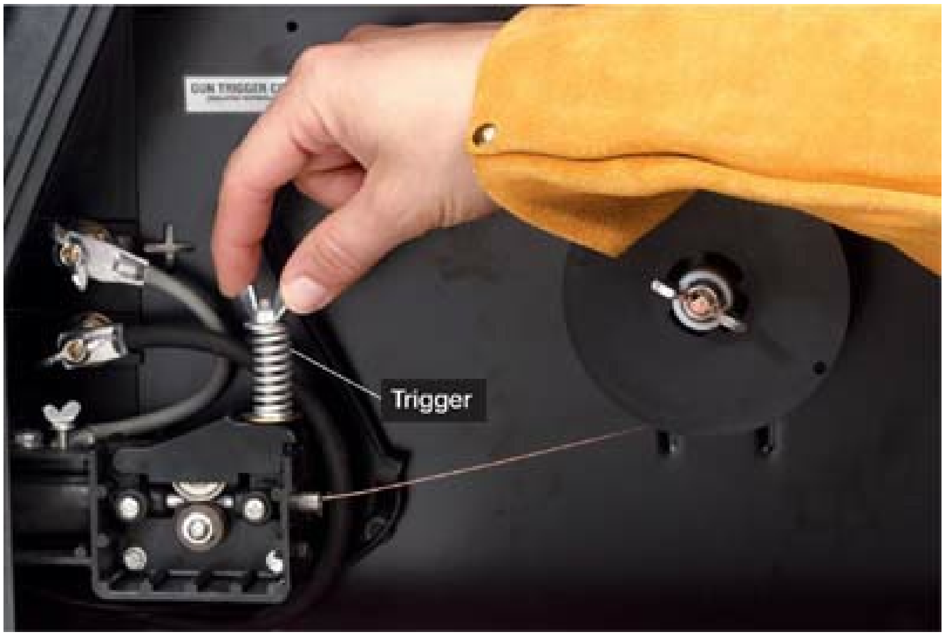

3 Swing the tension arm or pressure roll back into position. Adjust tension according to manufacturer's directions. On a new machine, the drive roll pressure will already be set. Make sure the wire is aligned perfectly straight—not up and down or side to side—on the drive roll. Turn the machine on, turn the wire speed to its highest setting, pull the cable straight, and depress the trigger (shown). Some machines have an inch button that feeds the wire without supplying power to the contact tip or wasting shielding gas while loading wire.

How to Set Up the Shielding Gas Flow Meter

1 Open the cylinder valve briefly to clear out any dirt. Wipe the cylinder threads with a clean, dry cloth.

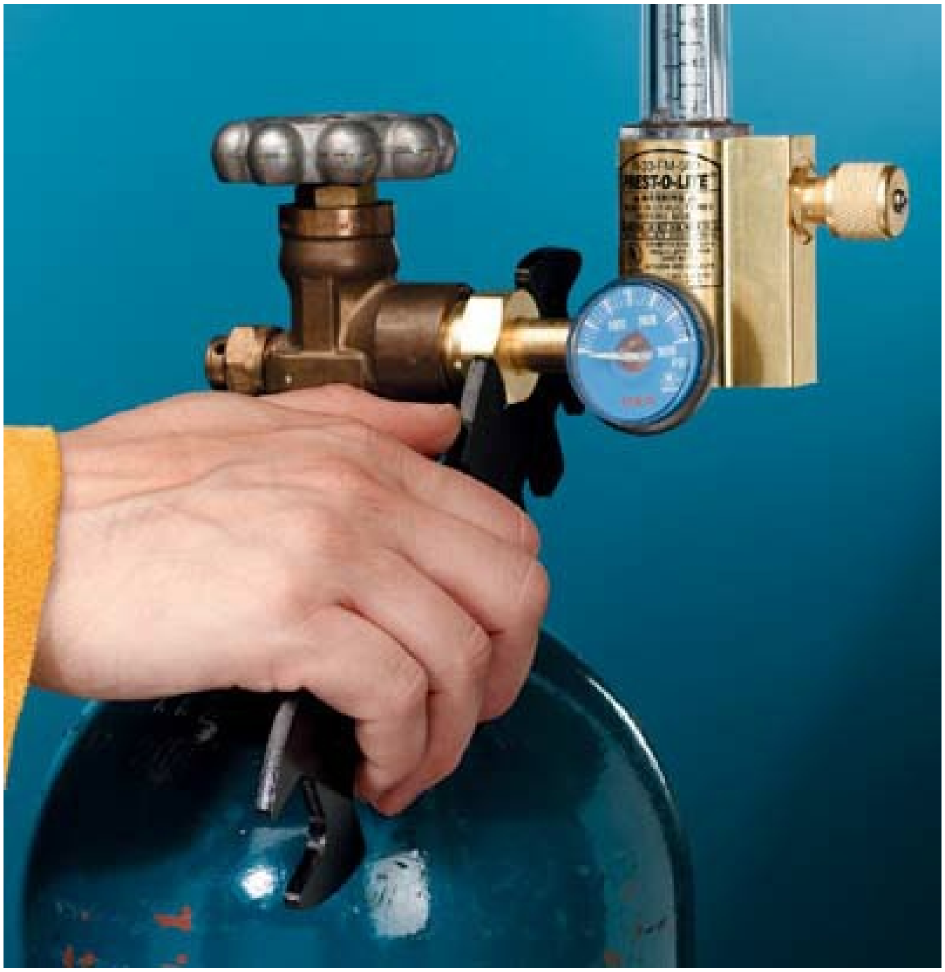

2 Attach the flow meter to the cylinder. Use a fixed wrench to prevent damaging the brass fittings.

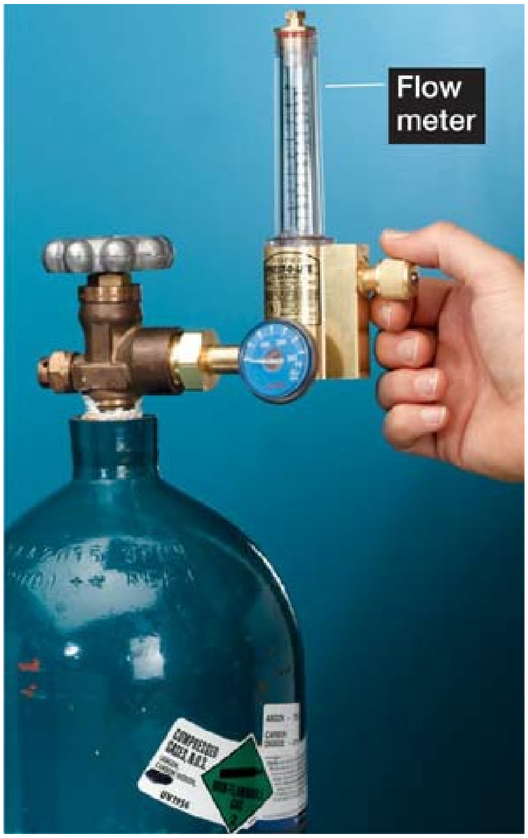

3 Turn the knob on the flow meter clockwise to tighten it so no gas can flow through it. If the flow meter is open when you open the cylinder valve, the high pressure can damage the flow meter.

4 Attach the hose to the flow meter and to the welding machine. Slowly open the cylinder valve, then, once it is open, open the cylinder valve completely. To set the flow rate, turn on the machine. Depress the trigger to activate the gas flow, and turn the flow meter knob counterclockwise. While gas is flowing, continue turning the knob until the meter registers the proper flow rate. Make sure that the flow meter or flow gauge is rated for use with the shielding gas you are using. For example, an argon/carbon dioxide flow meter cannot handle pure carbon dioxide.

Prewelding Checklist

- Gun, shielding gas, and machine cables, hoses, and connectors are tight.
- Polarity is set correctly for type of wire being used.
- Cables are in good condition.
- Supply cable is straight and unkinked.
- Gas regulator and flow meter are correctly installed.
- Contact tube is in good condition.
- Spatter has been cleaned from nozzle and contact tube (see page 66).

How to Weld with Gas Metal Arc

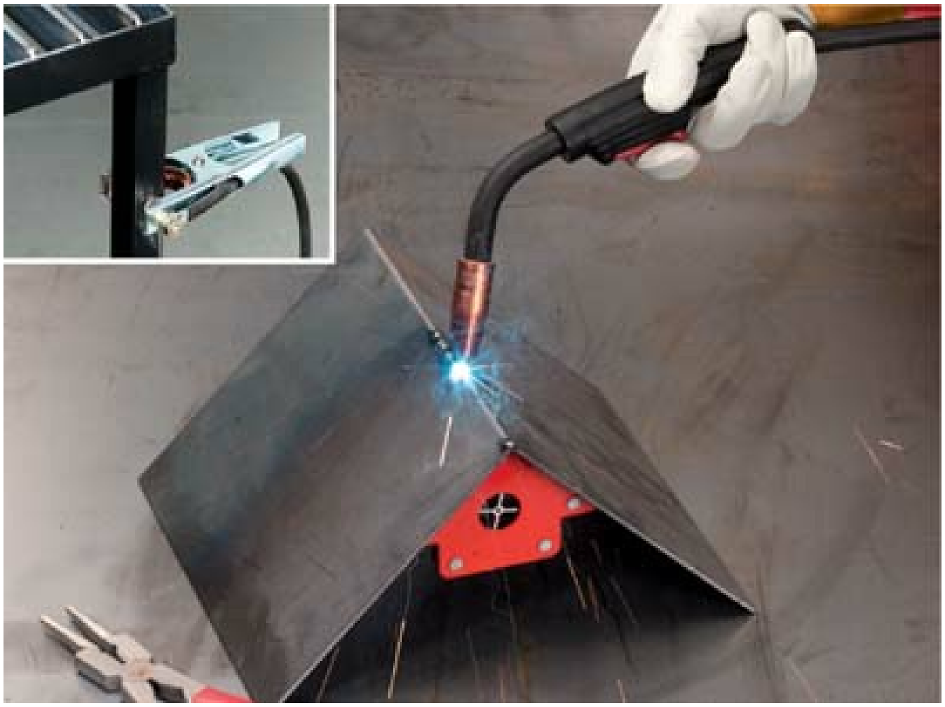

1 Prepare and fit up the material to be welded. Attach the work cable clamp to your welding table (inset) or the workpiece. Turn on the machine and shielding gas. Adjust the wire speed and voltage according to the manufacturer's recommendations. Cut the electrode to 3/8" stickout. Put the tip of the electrode at the point of the first tack weld, flip down your helmet, and pull the trigger. Place tack welds evenly around the weld area. Trim the electrode for proper stickout at the end of each use. Check to see that the workpiece is still aligned properly. If not, adjust with a hammer or break the tack welds and retack.

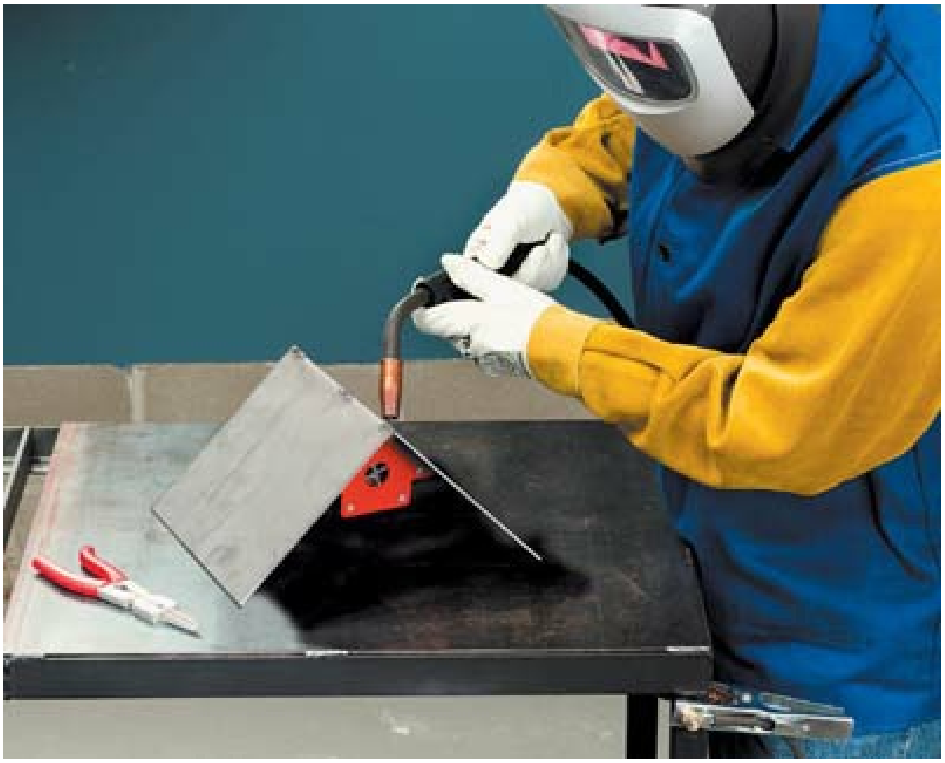

2 Start at the left end of the weld. Hold the gun with both hands and position it at a 20° angle with the tip pointing to the left. Put the tip of the electrode at the beginning of the weld, flip down your helmet, and pull the trigger.

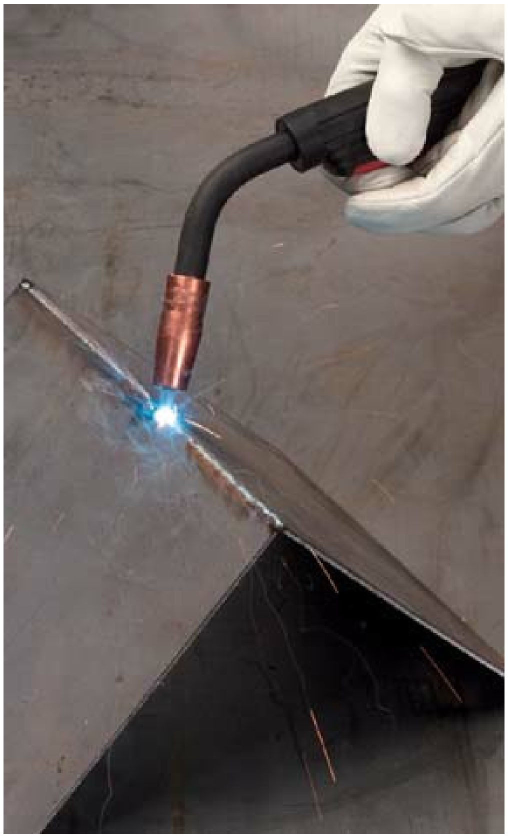

3 Begin welding, moving steadily to the right. GMAW has a distinctive sizzling sound. Popping and snapping indicate dirty material or an improper voltage or wire speed adjustment.

Postwelding Sequence

When you have finished welding with a GMAW rig, it is important to shut it down properly so you are set for your next welding session. If you do not turn off your shielding gas, you will lose much of it to bleed off even if you weld the next day.

- Close the gas cylinder valve.
- Press the gun trigger or purge switch until the flow meter or flow gauge reads zero.
- Turn off the flow meter.
- Turn off the machine.
- Coil the supply and work cables and store them off the floor.

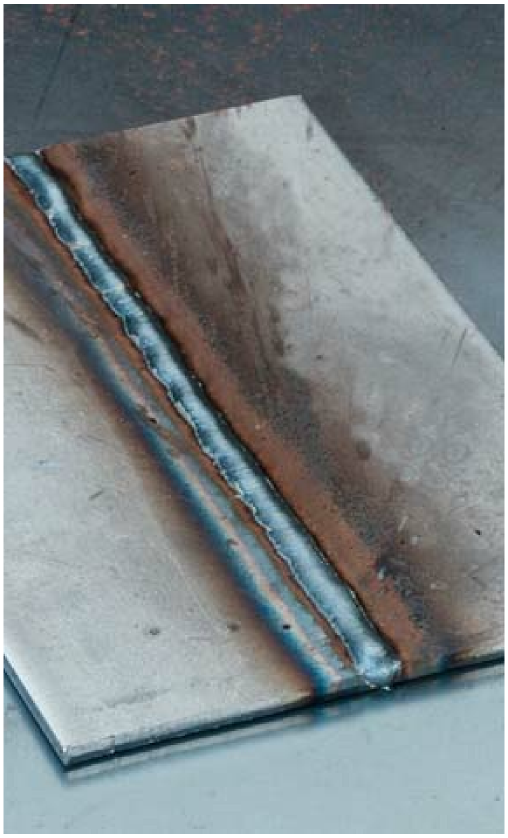

4 A finished GMAW is smooth with even ripples or weave pattern, no slag, and little spatter.

Flux Cored Arc Welding

Flux cored arc welding uses the same wire feed and power supply as gas metal arc welding but usually without the gas shielding. This is convenient for welding outdoors in windy conditions, but the weld is not as clean. To use flux cored wires, most welding machines need to be converted. Usually this involves changing the output polarity and installing a gasless nozzle. It also might require installing proper drive rolls and a different cable liner. Otherwise, follow the same steps to weld with flux cored wire as with solid wire.

Flux cored wire welds are covered with a layer of slag.

Maintenance & Troubleshooting

A small amount of maintenance on your GMAW equipment will result in better, more consistent welds and longer-lasting equipment. Spatter—tiny pieces of the electrode or base metal that have sizzled off—builds up on the nozzle and the contact tip. Remove this spatter frequently as you weld so it does not interfere with gas flow or electrical conductivity. Turn off the machine, remove the nozzle, and use the closed point of the MIG pliers (page 59) to ream out the nozzle and file the end of the contact tip. An anti-spatter gel is available in which to dip the hot nozzle and contact tip. Anti-spatter spray can be used to coat the contact tip and the nozzle. This spray is also useful for coating your welding table to prevent spatter from sticking to your work surface. Eventually you will need to replace the nozzle, but regular maintenance and proper welding techniques will extend the life of a nozzle considerably.

Contact tubes become worn because they are a soft copper alloy and the electrode is steel. The electrode will wear through the contact tip, possibly making the electrical flow irregular. Contact tips should be visually inspected for wear on a regular basis. The orifice on a worn contact tip will appear oval instead of round. Tips generally are good for about eight hours of continuous welding use. NOTE: Always turn off the welder before changing the contact tip.

It is important to keep the welder fan motor and rectifier free of dirt and dust to prevent it from overheating. If your welder is stored in a dusty shop, or you use it infrequently, keep it covered. Use low pressure air to blow dust and dirt from these assemblies. Use a vacuum to remove dirt from the wire feed mechanism.

The supply cable needs to be cleaned occasionally as well. Check the manufacturer’s recommendations for frequency, but it's generally recommended to clean after you’ve used 50 pounds of flux cored wire or 300 pounds of solid wire. With the power off, remove the cable from the machine. Remove the gas nozzle and contact tip from the gun, lay out the cable straight, and use low-pressure air to blow into the gun end. (Using high-pressure air may create a dirt plug that will clog the cable permanently.)

Always store the supply and work cables and the gas hose off the floor and away from chemicals, hot sparks, and sunlight.

Stub Out, or Stubbing

“Stub out” is when the electrode welds to the base metal without melting or breaking off. This is caused by the voltage being too low, the wire feed being set too fast, or holding the gun too close to the work when starting. Correct the settings, grind off the stub outs, and restart the weld.

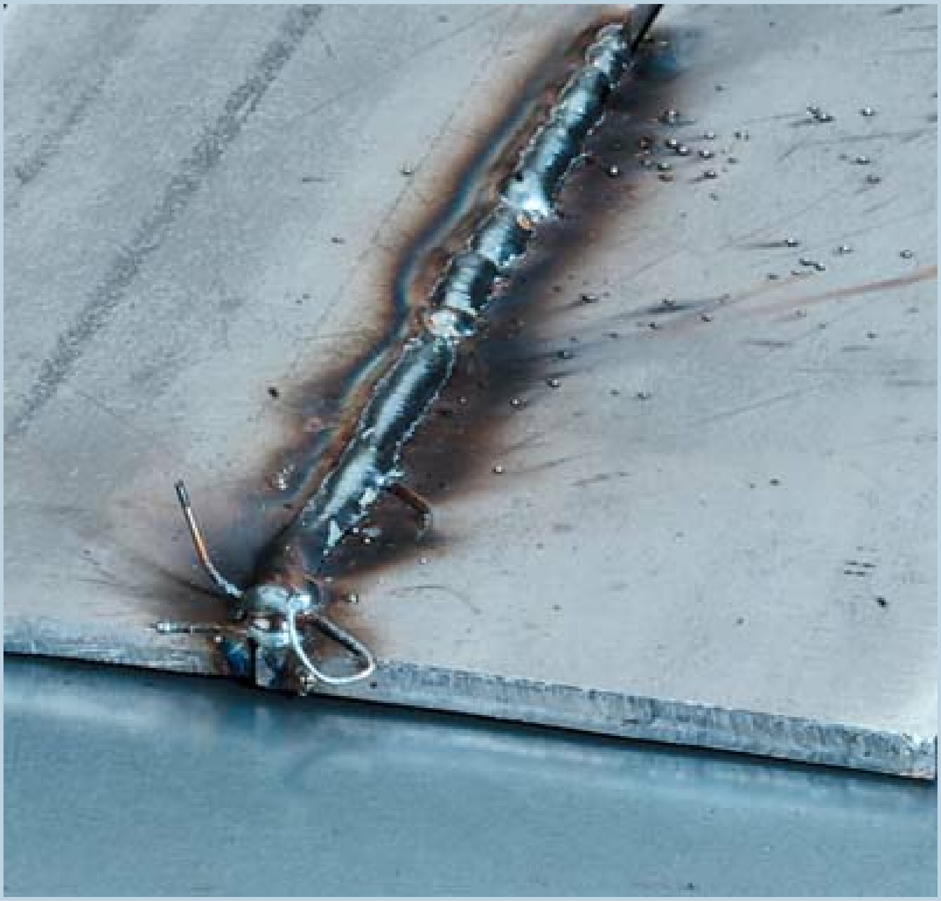

Stub out results in short sections of wire poking out of the weld.

Removing a Bird’s Nest

A bird’s nest is a tangle of wire in the wire drive mechanism. This happens when the wire drive rolls continue feeding wire but the cable is blocked, the outfeed tube is misaligned, the wire is stubbing out on the base metal, or the wire has been welded to the contact tip. The drive rolls are set to slip in these cases, but if the tension has been adjusted too tightly, they may continue to push wire.

To fix a bird’s nest, turn off the machine and cut the wire as it comes off the spool. Remember the wire is under tension, so you must hold the spool end, then tie it off through one of the spool holes. The tangled wire also may be storing tension, so it may fly out of the drive roll area. Release the tension arm or roller arm. Pull the wire out of the drive mechanism and the supply cable. Wear safety glasses when removing a bird’s nest.

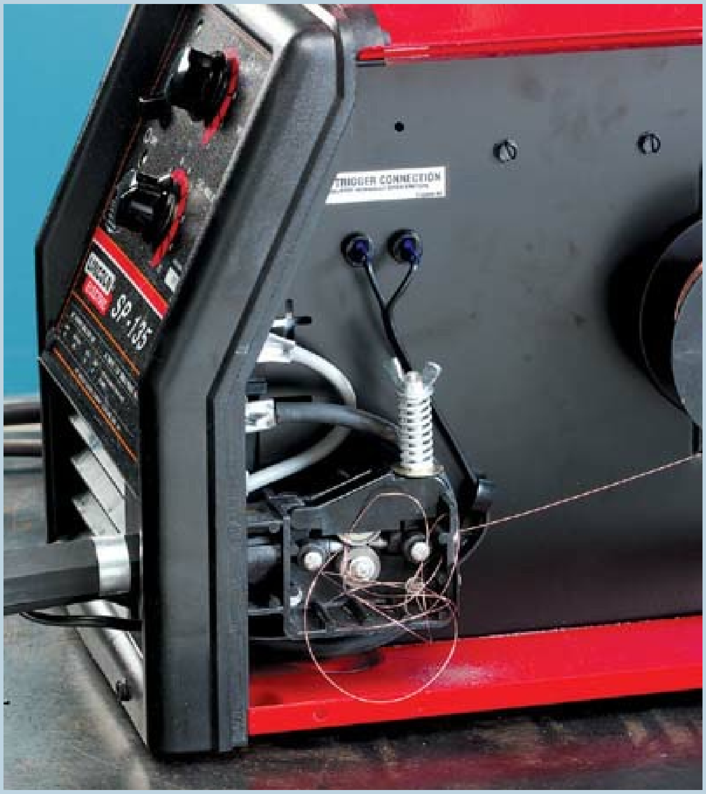

ANCHOR BRAND
NOZZLE-DIP
16 FLUID OZ
.473 LITERS
FOR MIG GUN NOZZLES
AND CONTACT TIPS

Use dippable anti-spatter gel to prevent spatter buildup on the nozzle and contact tip. Simply dip the hot nozzle into the gel.

GMAW Troubleshooting

PROBLEM	SOLUTION
Wire feeds, but no arc can be struck	• Check to see that the work clamp is connected
Weld bead is full of holes like a sponge (porosity)	• Make sure the shielding gas is on • Adjust shielding gas to a higher flow rate • Eliminate drafts • Verify correct polarity
Weld bead is tall and looks like a rope	• Make sure the voltage setting is appropriate for the material's thickness • Reduce your travel speed
Excessive spatter or dirty welds	• Reduce the gun angle • Hold the nozzle closer to the work • Decrease the voltage
Arc starts and stops	• Check the wire feed for steady feeding • Check the cable connections • Clean the weld materials

Gas Tungsten Arc Welding (GTAW/TIG)

Gas tungsten arc welding (GTAW), commonly called TIG (tungsten inert gas), and sometimes referred to as Heliarc (the L-TECH trade name), is a process that generates an arc between a non-consumable tungsten electrode and the workpiece. The electrode and the weld are protected by a shielding gas, and filler metal may or may not be used.

Gas tungsten arc welding differs from the other arc welding processes because the electrode is non-consumable and not used as filler material. GTAW is more like oxyfuel welding in terms of the skills needed to manipulate the torch with one hand and the filler rod with the other. GTAW requires another layer of coordination because most machines also use a foot activated amperage control.

Like gas metal arc welding, gas tungsten arc welding is a clean process be cause the shielding gas eliminates the need for flux and the resultant slag.

Safety: Because gas tungsten arc welding is such a clean process, welders often are tempted to weld without gloves or in short-sleeve shirts. This is not recommended. The arc most likely produces more ultraviolet rays than other processes, and because there are no fumes or smoke, those rays are entirely unfiltered. It is important to cover all exposed skin to prevent UV burns. Filter requirements are a minimum #10 shade, and if you have an auto-darkening hood, be sure that it is rated for the GTAW process. Some entry level auto-darkening hoods are not. The arc may produce ozone, which is hazardous to breathe over long periods. Make sure your gloves, clothes, and work area are dry. Always turn off the power when changing electrodes.

Equipment: The basic equipment needed for GTAW is a constant current welding machine, cable with torch, work cable and clamp, electrode, and inert gas cylinder with regulator and flow meter. Optional equipment includes a remote amperage control and a water-cooled torch with water cooler and hoses.

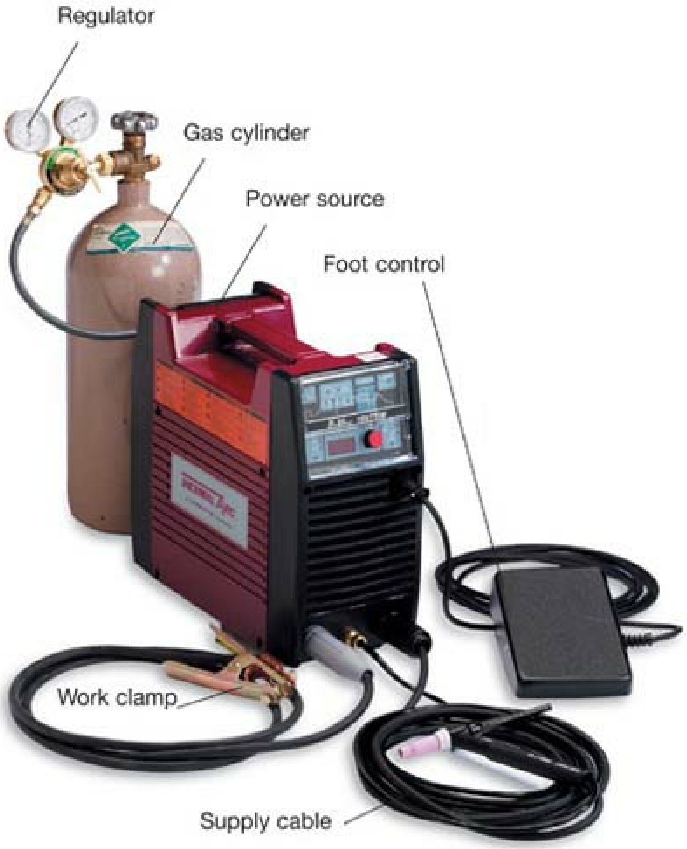
Regulator
Gas cylinder
Power source
Foot control
Work clamp
Supply cable

A gas tungsten arc welding setup consists of a power source, work cable and clamp, supply cable and torch, foot control, and gas cylinder with regulator.

Although a mid-range shielded metal arc machine can be used to deliver the current for gas tungsten arc welding, a dedicated good-quality GTAW machine delivers the current as AC or DC, provides an optional high-frequency output for no-touch arc starting, has a remote control option for foot pedal control, and has a solenoid for shielding gas control. The combination of AC and high frequency makes it possible to weld aluminum with good results. The newest electronic GTAW machines have advanced current control capabilities and are becoming more and more affordable.

Torch and cables: The GTAW torch holds the electrode and delivers the shielding gas. It can be air or water cooled. The torch parts are the cup or nozzle, collet body, collet, end cap, and torch body. The collet and collet body hold the electrode firmly and establish electrical contact with the electrode. The cup directs the shielding gas. A torch may be air cooled for use below 200 amps, but over 200 amps the torch usually is water cooled. The supply cable supplies electricity, shielding gas, and coolant to the torch. This may be an integrated cable or individual cables. A work cable with clamp is also necessary.

Some GTAW torches have flexible necks, which allows them to be bent into the most convenient angle for supporting the torch. Unfortunately, the non-flexible torches will bend if forced, and then they will be ruined. Some torches also have a gas on/off valve. This is necessary for machines that are simply power sources with no internal gas flow controls.

Safety

- Welding helmet with #10 to #14 filter
- Long pants, long- sleeve shirt, and hat
- Leather gloves
- Leather boots or shoes
- Ventilation

Pliers and a stainless steel wire brush are handy for gas tungsten arc welding.

The collet body, collet, end cap, and cups all come in a variety of sizes. The collet body and collet are sized to match electrode diameters. Each part is stamped with its size. The end caps come in a variety of lengths—some accommodate an entire 7" electrode, others are shorter for more clearance while welding.

Cups come in a variety of sizes denoted by $^{1}/_{16}$" increments. A #4 nozzle is $^{4}/_{16}$", a #8 is $^{8}/_{16}$". Cups should ideally be 4 to 6 times the size of the electrode. A cup that is too large will not direct the gas in a proper stream, a cup that is too small may not provide adequate steady flow. A gas lens or screen in the torch balances the gas flow. Good shielding designs allow you to use a lower gas flow. Higher flow creates turbulence and jetting, which disturbs the shielding—more does not equal better when it comes to shielding gas flow. Cups also are available in different lengths. A longer nozzle increases the electrode extension from the collet, which may increase the electrode temperature. Use either lower amperage or larger diameter electrodes if you are using a longer nozzle. Shorter nozzles may be necessary for tight working conditions. Nozzles are also available in clear (fused quartz) types, which are more expensive but give greater visibility, especially in cramped quarters.

Shielding gases. Two inert shielding gases are used for gas tungsten arc welding: argon and helium. They may be used alone or in combination with one another. Argon yields better starts and better arc maintenance. Helium gives a hotter arc and allows faster travel speeds. Helium, however, is very expensive. Argon and helium are both contained in standard cylinders, and a pressure regulator flow meter combination will need to be attached. Because of the density difference between the two, either a separate flow meter needs to be used for each, or use a flow meter that has calibrations for each type of gas. Argon flow rates for up to 1/8" steel is 8 to 10 cubic feet per hour (CFH), for stainless steel 11 to 13 CFH, and 12 to 20 CFH for aluminum. Drafty locations may require higher flow rates. Check the manufacturer’s recommendations.

Electrodes. The electrode for gas tungsten arc welding is always a tungsten or tungsten alloy rod with either a pointed or a rounded end. The electrodes come in a variety of diameters.

Preparing the electrode. Prior to use, the cut end of the electrode must be sharpened to a point or melted to a ball. The tip may be ground to a point or chemically sharpened. The electrodes come in 7" lengths. To increase the number of points available, score the electrode with a file or cut-off wheel, and snap it in half. Tungsten is very hard but brittle, so it is easy to grasp each end of the electrode with pliers and snap it in half over a sharp table edge.

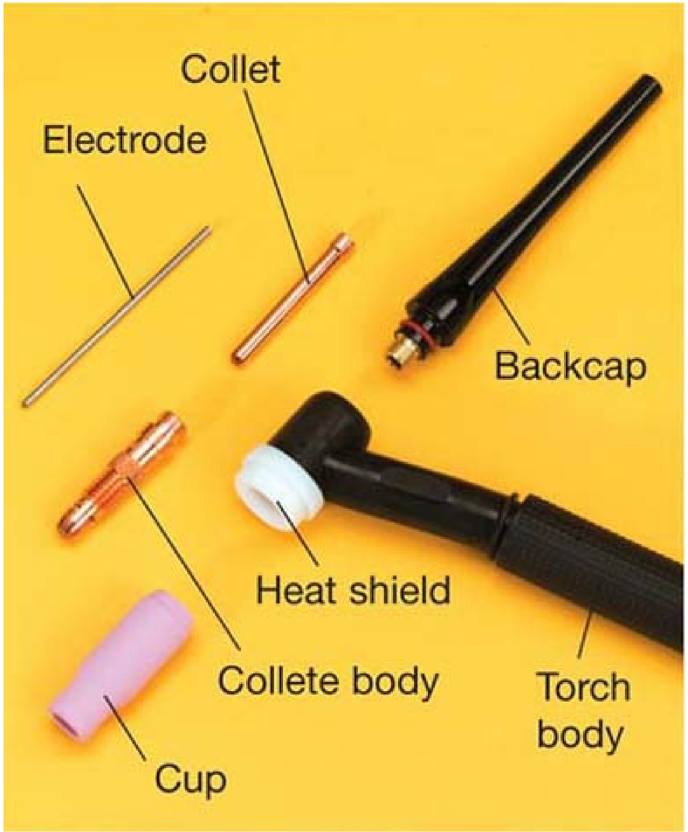
Collet
Electrode
Backcap
Heat shield
Collete body
Torch body
Cup

A gas tungsten arc welding torch includes the electrode, collet body, collet, cup, backcap, and torch body with heat shield.

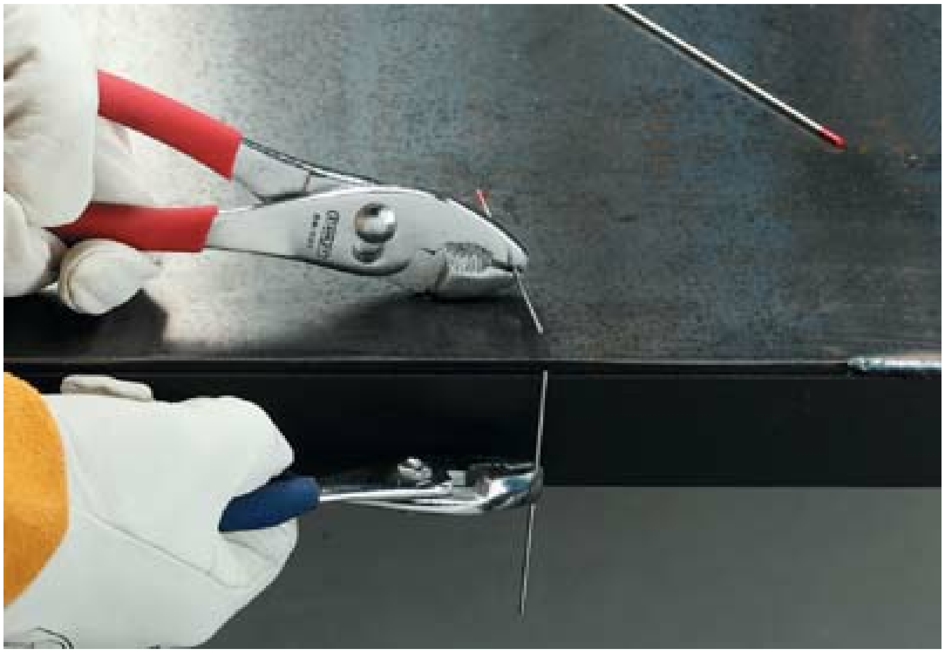

Tungsten is very brittle, so electrodes can be scored with a file, then snapped over a sharp table edge.

Because all tungsten electrodes look and feel the same regardless of their composition, it is important to keep them clearly separated by type. The color codes will wear off, or, if you point each end of your electrode, be ground off. It is helpful to have clearly labeled containers for each type of electrode.

Two critical factors in grinding the electrodes are the grinding wheel and the grinding direction. You must use a hard, fine grinding wheel dedicated exclusively to tungsten. Metal particles left on the wheel from grinding aluminum or steel would contaminate the tungsten, which causes erratic arc behavior and poor weld quality. An extremely hard material, tungsten will become hot as it is ground. Sharpen the electrode tip so that grinding marks run lengthwise down the tip, not in a circular or crosswise pattern. Lengthwise grinding focuses the electron flow toward the tip; circular grinding causes the arc to be unfocused and possibly jump sidewise from the electrode rather than off the tip point. Chemical means also can be used to sharpen tungsten by dipping a hot tungsten rod into a chemical agent. The length of the taper on the tungsten tip should be two to three times the diameter of the tungsten.

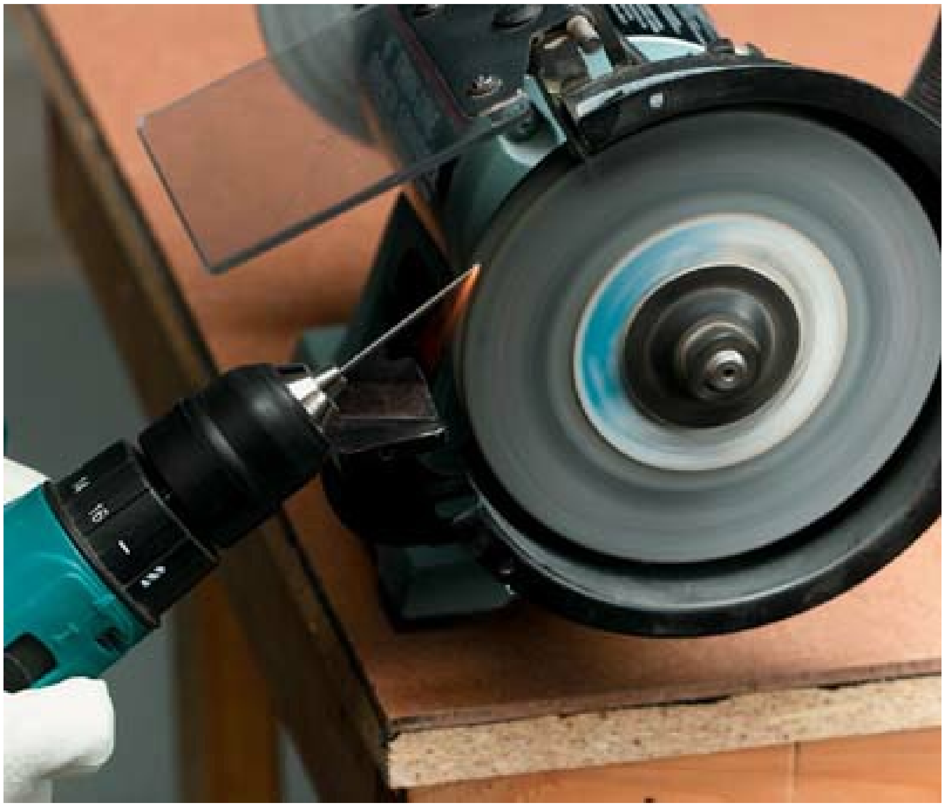

To safely hold tungsten electrodes for sharpening with a grinder, insert them into the chuck of a drill. Running the drill while sharpening the tungsten ensures a uniform point. Move the tungsten across the wheel to prevent grooves in the wheel edge.

Electrode Characteristics

ALLOY	AC/DC	TRAITS & USES	COLOR
Pure tungsten	AC	• Welding of aluminum • Forms ball at tip	Green
1% Thorium oxide (thoriated)	Primarily DC, but also AC	• Easier arc starts • Carries more current • Low-level radioactive	Yellow
2% Thorium oxide	DC	• Long life • Easier arc starts	Red
Zirconium oxide	AC	• Similar to pure tungsten but carries more current	Brown
2% Cerium oxide (ceriated)	DC	• Not as good as thoriated but not radioactive	Orange
Lanthanum oxide	DC	• Similar to ceriated	Black

Welding with AC current requires that the tip of the electrode be balled, which is accomplished by running DC electrode positive or AC current and striking an arc on copper or brass. The ball should not extend beyond the diameter of the electrode, although many sources recommend a ball one and one half times the diameter of the tungsten.

If you touch the weld puddle or filler rod with the electrode during welding, you must change to a fresh electrode. Any time the electrode tip becomes pitted or blackened, is no longer pointed, or the ball becomes too large, you must change to a fresh electrode. This will happen often, so it is best to have a number of electrodes sharpened and ready to go. If you forget to turn on the shielding gas or the flow is interrupted, the electrode will give off a thick yellow tungsten oxide smoke. This smoke is hazardous, so resume welding only after the smoke has cleared. You will need to change electrodes if this happens.

Anchor Brand
CHEMICAL SHARPENER

Chemical sharpeners are available for sharpening tungsten.

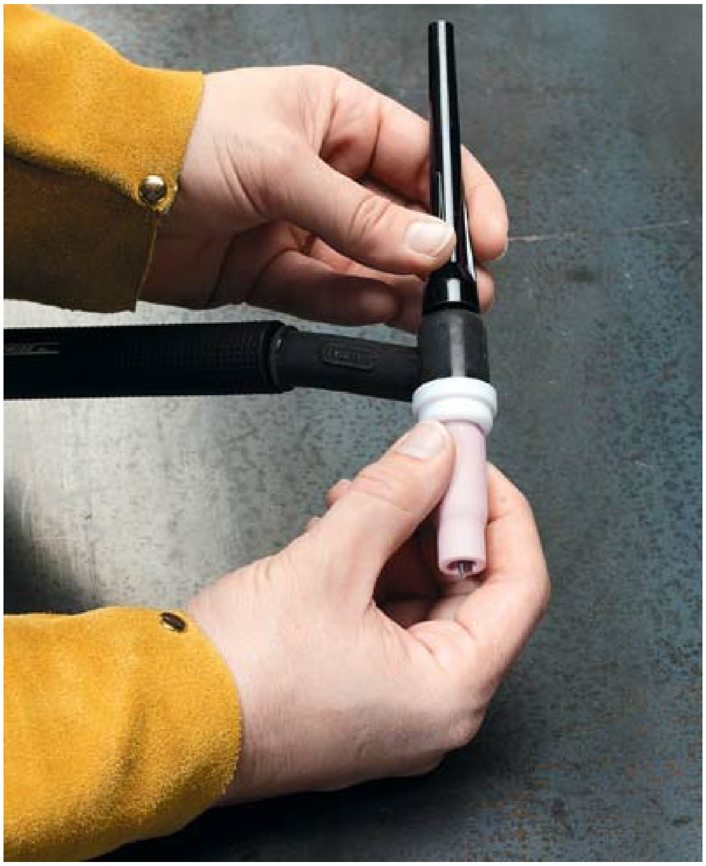

Tightening the back cap locks the tungsten electrode into place. The tip of the electrode should extend beyond the end of the cup by a distance of three times the electrode’s diameter.

Anatomy of the GTAW Machine

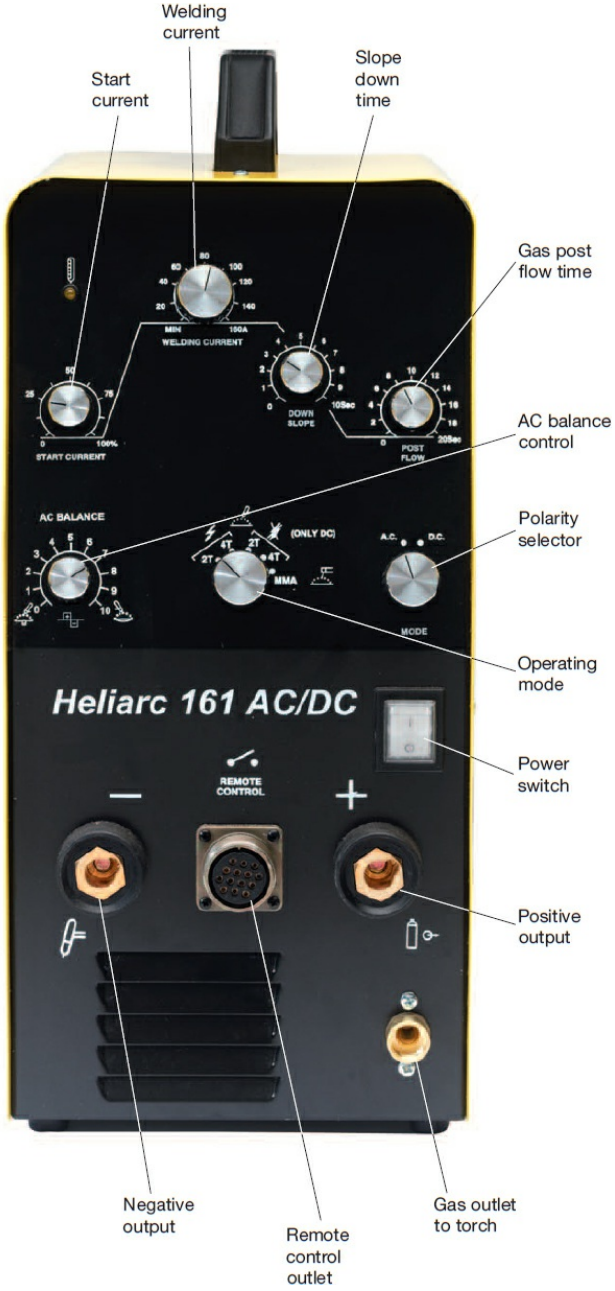
Welding current
Start current
Slope down time
Gas post flow time
AC balance control
Polarity selector
Operating mode
Power switch
Positive output
Negative output
Remote control outlet
Gas outlet to torch
Heliarc 161 AC/DC
WELDING CURRENT
START CURRENT
DOWN SLOPE
POST FLOW
AC BALANCE
(ONLY DC)
MMA
MODE
A.C.
D.C.
REMOTE CONTROL

Setting Up the GTAW Machine

Dedicated GTAW machines, like this one and the one pictured on page 68, have many advanced features. The basic setup, however, is the same.

The positive and negative output receptacles are for the torch and work clamp connections.

The gas outlet connects to the torch gas hose.

The remote control socket is for the foot or finger control. Using a remote control allows the welder to control the amount of current while welding.

The operating mode selects high frequency or scratch starts.

Polarity selects either AC or DC.

AC balance control allows for adjusting AC power to be more positive or more negative to create either more cleaning action or deeper penetration. Advanced machines can adjust the AC power to be as much as 90% positive or negative, rather than the 50% for standard AC power.

Start current controls the current for arc starting.

Welding current sets the range for welding. The foot or finger control then operates within this range.

Slope down time gradually reduces power to the arc without extinguishing the arc. This allows for the weld crater to fill before the arc is extinguished.

Gas post flow controls how long the shielding gas will flow after the arc is extinguished.

Prewelding Checklist

- If using a water-cooled torch, check for leaking.
- Check all cables for wear and damage.
- Clean and fit up parts to be welded.

How to Weld with GTAW

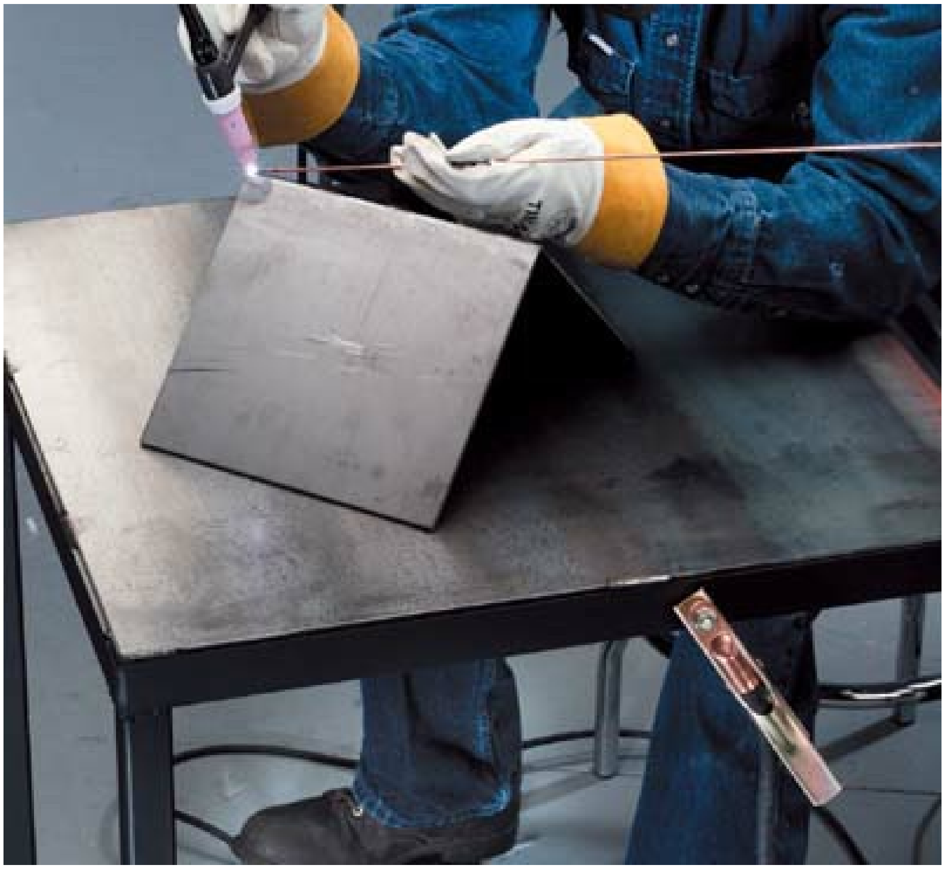

1 Set the controls based on manufacturer’s recommendations for the material to be welded. Turn on the machine, and turn on the water pump, if available. Attach the work clamp to the welding table or workpiece. Flip down your helmet, activate the foot or finger control if using one, and strike an arc by scratching the tip of the tungsten against the base metal. If your welder has a high frequency option, you do not need to scratch start the arc. Place a tack weld at each end of the joint to be welded. You may be able to tack the joint by simply fusing the two pieces with the heat of the torch, or you may have to use filler rod.

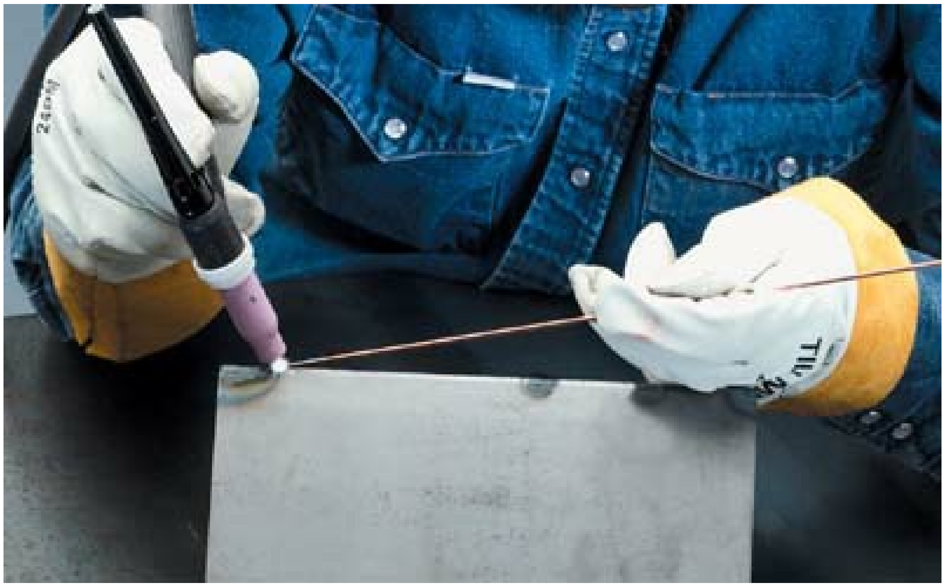

2 Position yourself to weld from right to left (if you are right handed) with the torch at a 15° angle to the right of center. Hold the filler rod in your left hand. Position yourself so you can comfortably hold the torch and filler rod for the duration of the weld.

Striking an Arc

Unless you have a high frequency option, you will need to physically strike an arc—called a scratch start. Rest the cup on the workpiece at a sharp angle. Move the tip until it briefly contacts the work, then angle it back again to start the arc. After the arc is started, lift the cup off the workpiece and establish the proper torch angle.

High frequency allows the arc to jump the gap without needing to create physical contact between the electrode and the workpiece.

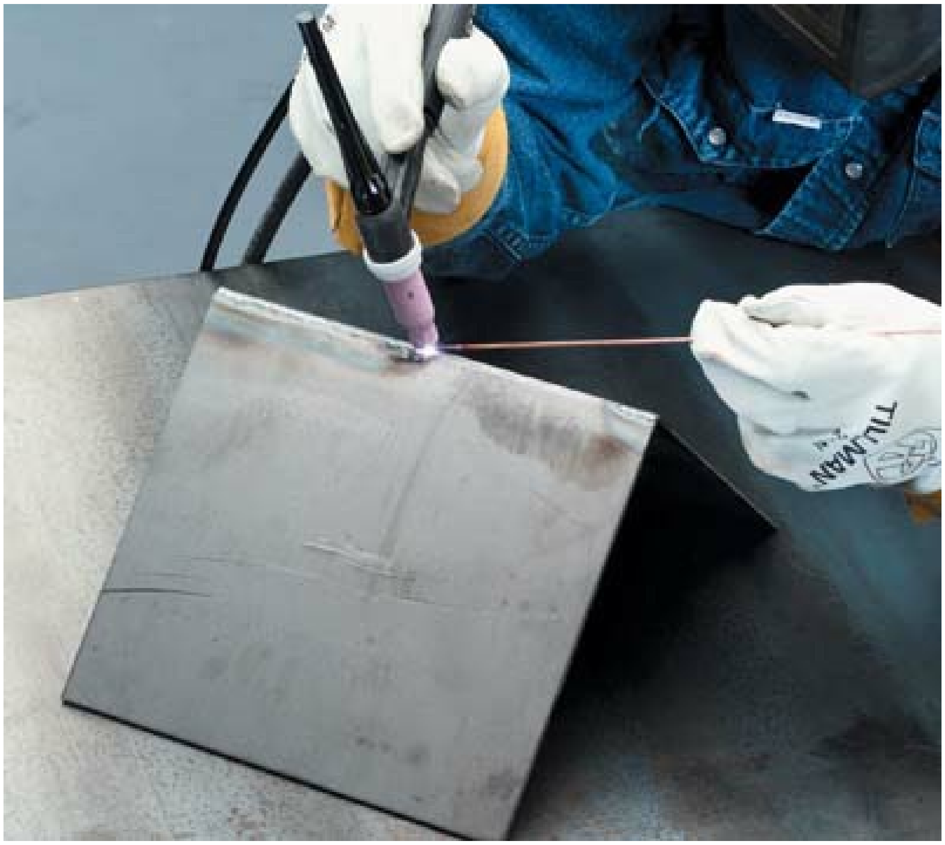

3 When a molten puddle has formed, dip the tip of the filler rod into the middle of the molten puddle. Keep the filler rod at a low angle to prevent disturbing the shielding gas. Keep the tip of the filler rod near—but not in—the puddle. Move the electrode to the left and continue the melting and dipping process.

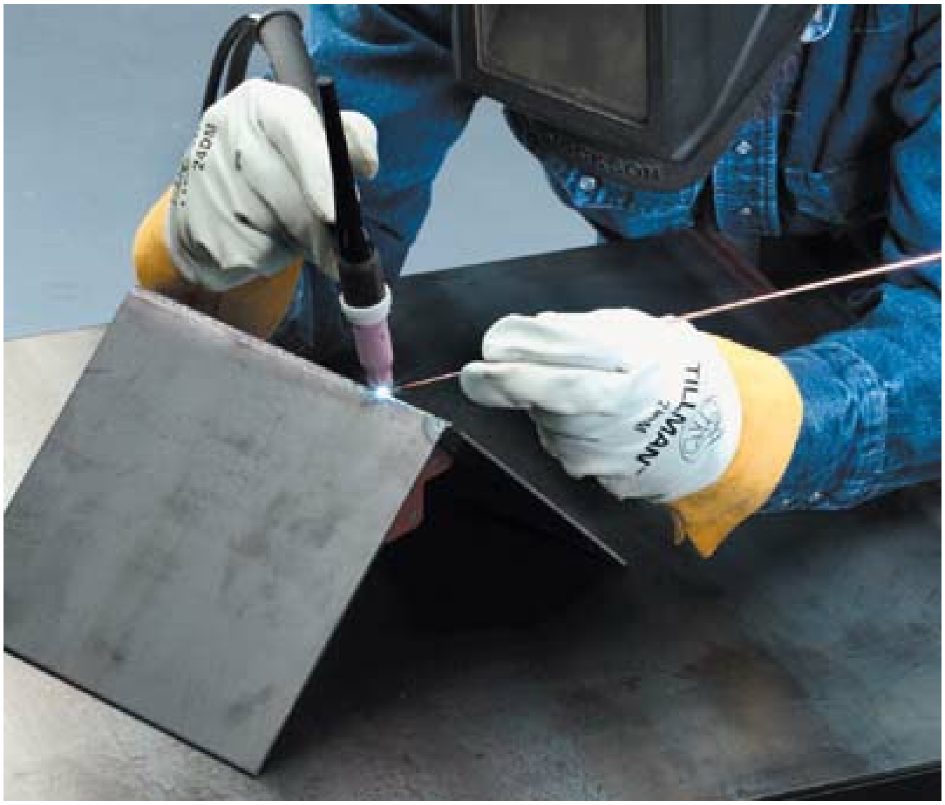

4 As you approach the end of your weld, you may need to adjust your travel speed because the buildup of heat in the material makes the molten puddle form more quickly at the end of the weld than at the beginning. You also may need to adjust the torch angle to be more shallow (not shown here) so that less heat is directed into the base metal.

Postwelding Sequence

- Turn off the cylinder valve.
- Purge gas from the gas line.
- Turn off the flow meter or gauge.
- Turn off the machine.
- Coil hoses and cables off the floor.

Troubleshooting

Filler metal for gas tungsten arc welding comes in rod form and ranges from $^1/_{16}$" to $^3/_{16}$" in size. Rods are available in a variety of alloys, including aluminum, chromium and chromium nickel, copper, nickel and nickel alloys, magnesium, titanium, and zirconium. Specific alloy compositions are available for creating specific weld types on specific base metals. These filler metals are similar to those used in oxyfuel welding, with the exception of the carbon steel rods, which are not copper coated as they are for oxyfuel.

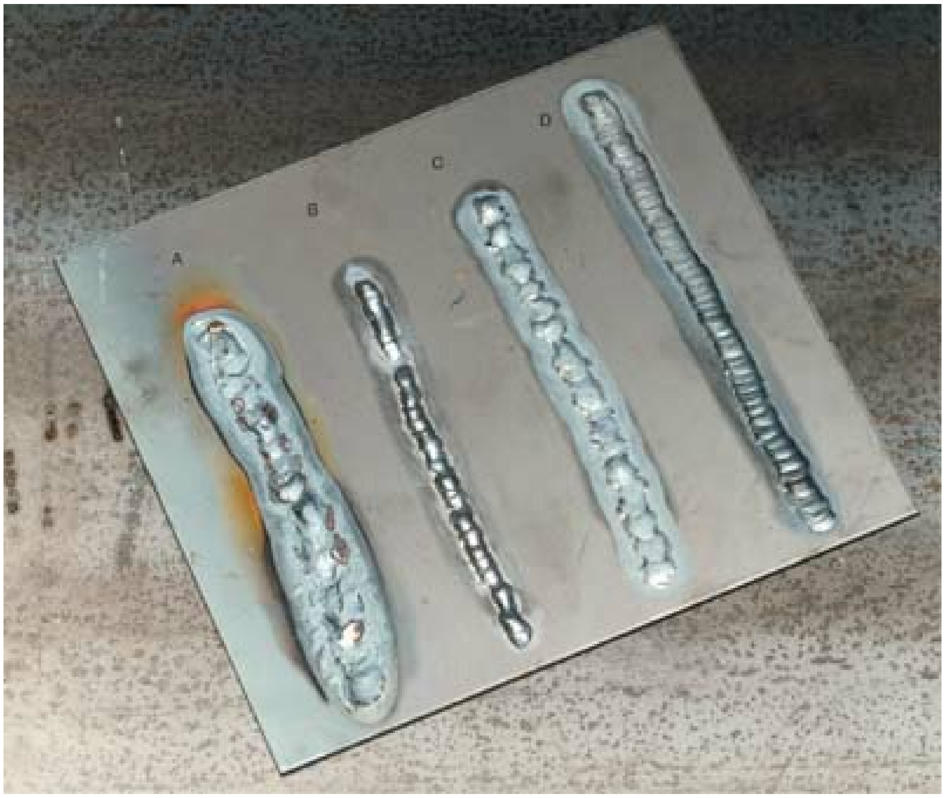

Becoming proficient in gas tungsten arc welding takes practice, and identifying problem welds is an important step. Weld A is too hot. Increase the travel speed or decrease the amperage. Weld B is too cold and is simply sitting on top of the base metal rather than penetrating it. Decrease the travel speed or increase the amperage. Weld C was done too quickly. Travel speed needs to be controlled and consistent. Weld D is a good-quality weld with even ripples, good penetration, and a moderate crown.

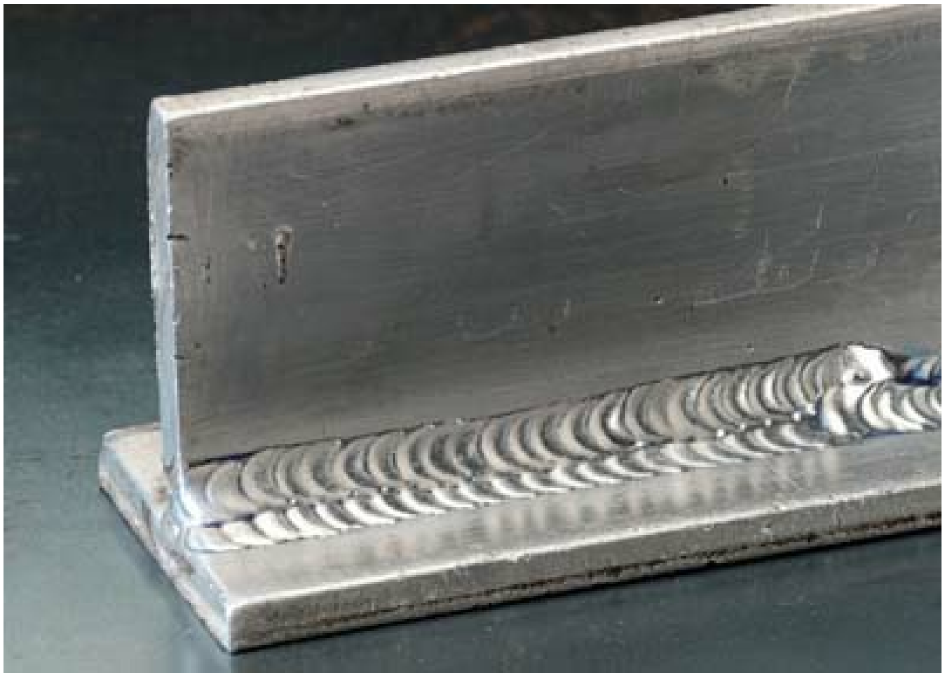

A well-done gas tungsten arc weld on aluminum has even ripples and good penetration. This sample weld shows two passes to create a fillet weld on ¼" stock.

GTAW Troubleshooting

PROBLEM	**SOLUTIONS**
Weld looks porous or sooty	• Make sure shielding gas is on and is correct type • Make sure shielding gas cylinder is not empty • Eliminate drafts • Make sure base metal is totally dry • Clean base metal thoroughly • Increase gas flow rate
Base metal distorts	• Tack weld parts before welding • Clamp parts down to rigid surface • Scatter welds to diminish heat buildup
Unstable arc	• Adjust electrode to work angle • Clean base metal thoroughly • Clean electrode • Connect work clamp to workpiece • Bring arc closer to work
Electrode is rapidly consumed	• Make sure polarity and current settings are correct • Increase electrode size • Increase gas flow • Decrease current • Increase gas postflow time • Use proper shielding gas

WELDING PROJECTS

Here’s a chance to put all these welding and cutting techniques to use. The following chapters have directions for creating 23 different projects. Each contains a detailed cutting list, technical drawing, and step-by-step directions. Among the shop projects you will find a sturdy welding table and a handy welding machine cart. The lighting projects feature a variety of candleholders and lamps. Included in the furniture projects are a delightful wine rack and a handy solution to creating a coffee table out of a slab of stone or wood. The outdoor projects will spark your creativity with an add-your-own-found-objects gate.

IN THIS SECTION:

Shop
Home Décor
Outdoor

Shop

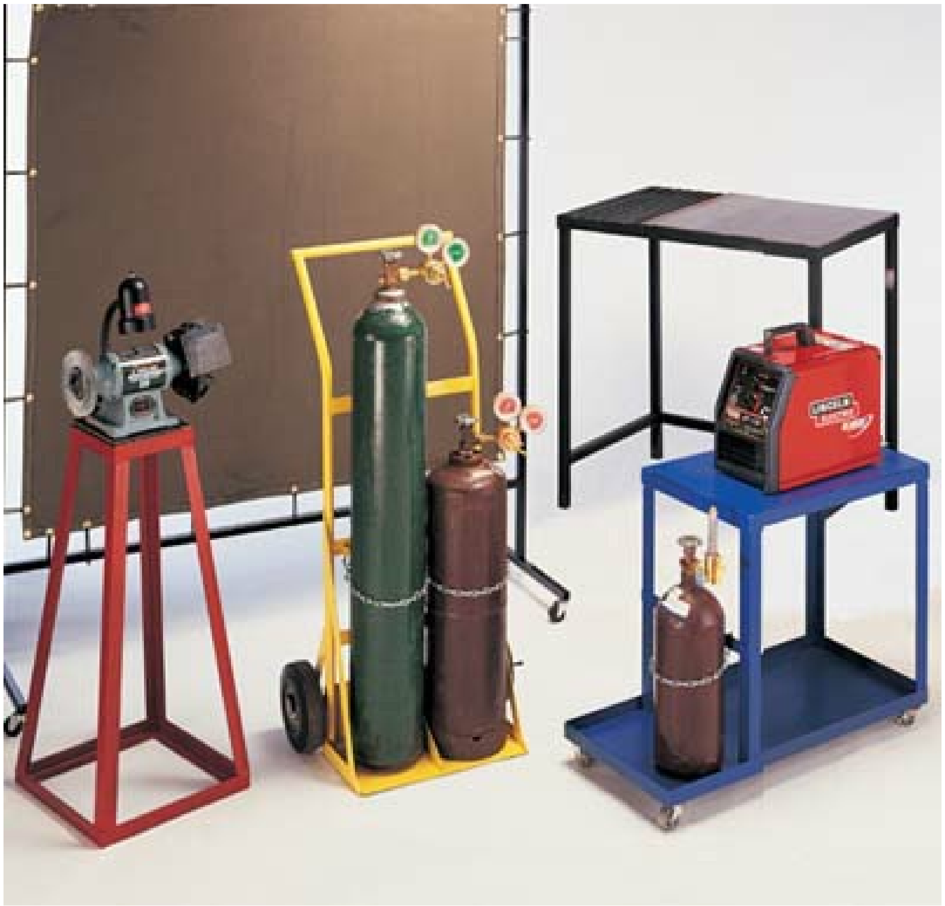

IN THIS CHAPTER:

- Rolling Welding Curtain
- Welding Table
- Welding Machine Cart
- Cylinder Cart
- Grinder Stand
- Portable Welding Table
- Saw Stand

Rolling Welding Curtain

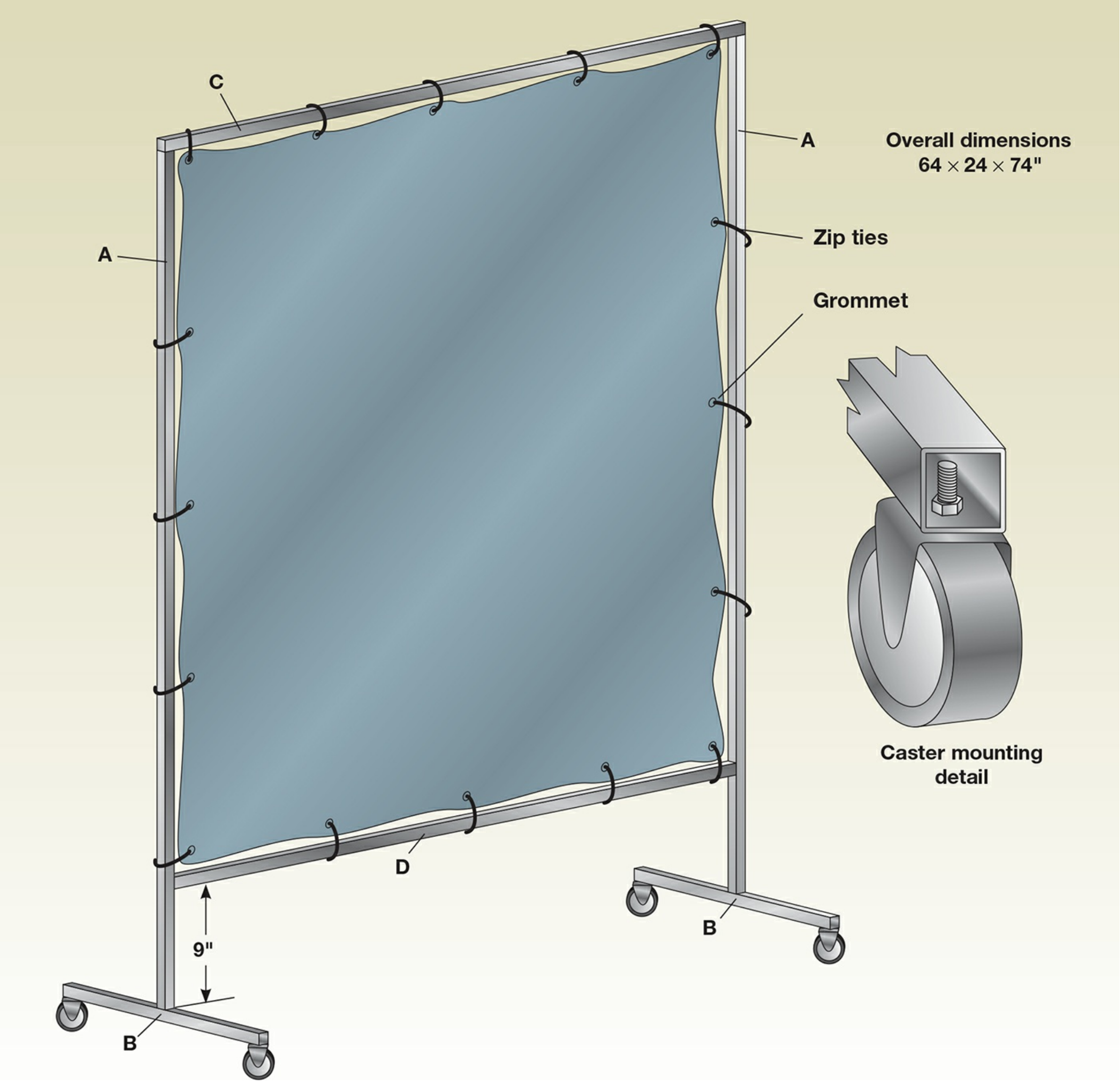
C
A
Overall dimensions
64 × 24 × 74"
A
Zip ties
Grommet
Caster mounting
detail
D
B
9"
B

For arc welding and plasma cutting, it is very important to screen your work area so other shop workers, passersby, family members, and pets are not exposed to the damaging rays. This rolling curtain is quick and easy to make, and it offers the necessary screening while you work. You can purchase ready-made welding tarps in a variety of colors, shapes, and sizes (see Resources, page 220). Or you can make your own welding tarp with 12-ounce cotton duck, fire retardant, and grommets.

Materials

- 1 × 1" square tube (27 feet)
- 4 threaded swivel casters, at least one locking
- 5 × 5 foot welding tarp
- Zip ties

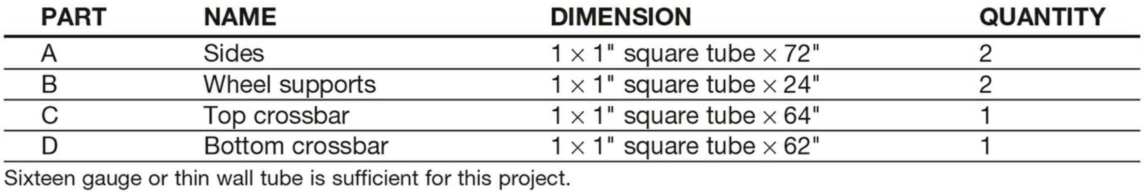

PART	NAME	DIMENSION	QUANTITY
A	Sides	1 × 1" square tube × 72"	2
B	Wheel supports	1 × 1" square tube × 24"	2
C	Top crossbar	1 × 1" square tube × 64"	1
D	Bottom crossbar	1 × 1" square tube × 62"	1

Sixteen gauge or thin wall tube is sufficient for this project.

How to Build a Rolling Welding Curtain

Attach the Wheel Supports to the Sides

1. Clean all parts with denatured alcohol, acetone, or degreaser. Prepare weld joint areas by wire brushing until shiny.

2. Cut the sides and wheel supports (A & B) to length.

3. Clamp one side piece to the work surface. Center a wheel support at the end of the side piece to form a T.

4. Check for square and tack weld along the butt joint between the two parts. Turn the assembly over and reclamp it to the work surface. Check that the wheel support is still square to the side piece. Make a final weld along the butt joint.

5. Repeat steps 3 and 4 to assemble the second side piece and wheel support.

Weld the Wheel Support T-joints

1. Clamp a side and wheel support assembly to the work surface so the wheel supports hang over the edge.

2. Weld the T-joint between the wheel support and the side piece. Turn the assembly over and weld the second T-joint.

3. Repeat steps 1 and 2 for the second side and wheel support.

Install the Crossbars

1. Cut the top and bottom crossbars (C & D) to length.

2. Place the top crossbar over the side pieces (see photo). Check for square, clamp in place, and tack weld.

3. Make a mark 10" up from the bottom of each wheel support. Align the lower edge of the bottom crossbar with the marks, check for square, and tack in place.

4. Flip the assembly over and check for square by measuring across both diagonals. If the measurements are equal, the structure is square. Clamp the assembly in place and complete the welds. Flip the assembly over and finish the welds on the other side.

Apply Finishing Touches

1. Paint the framework, if desired.

2. Drill holes for the threaded post swivels ½" from the end of the wheel supports. Install the casters.

3. Attach the welding tarp to the framework with zip ties.

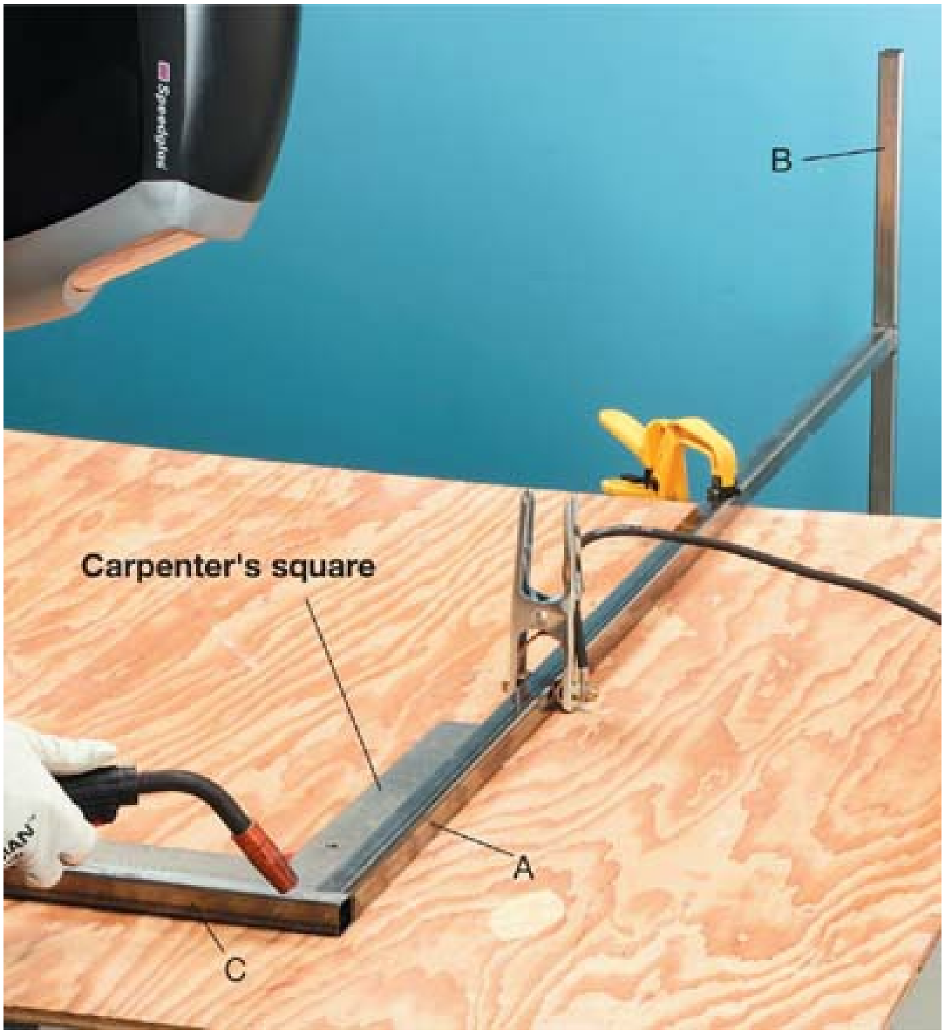

Clamp the sides to the table with the wheel supports hanging over the edge. Finish weld the wheel supports to the sides, then weld the top and bottom crossbars to the sides. Make sure the plywood work surface does not have smoldering embers when you are finished.

Welding Table

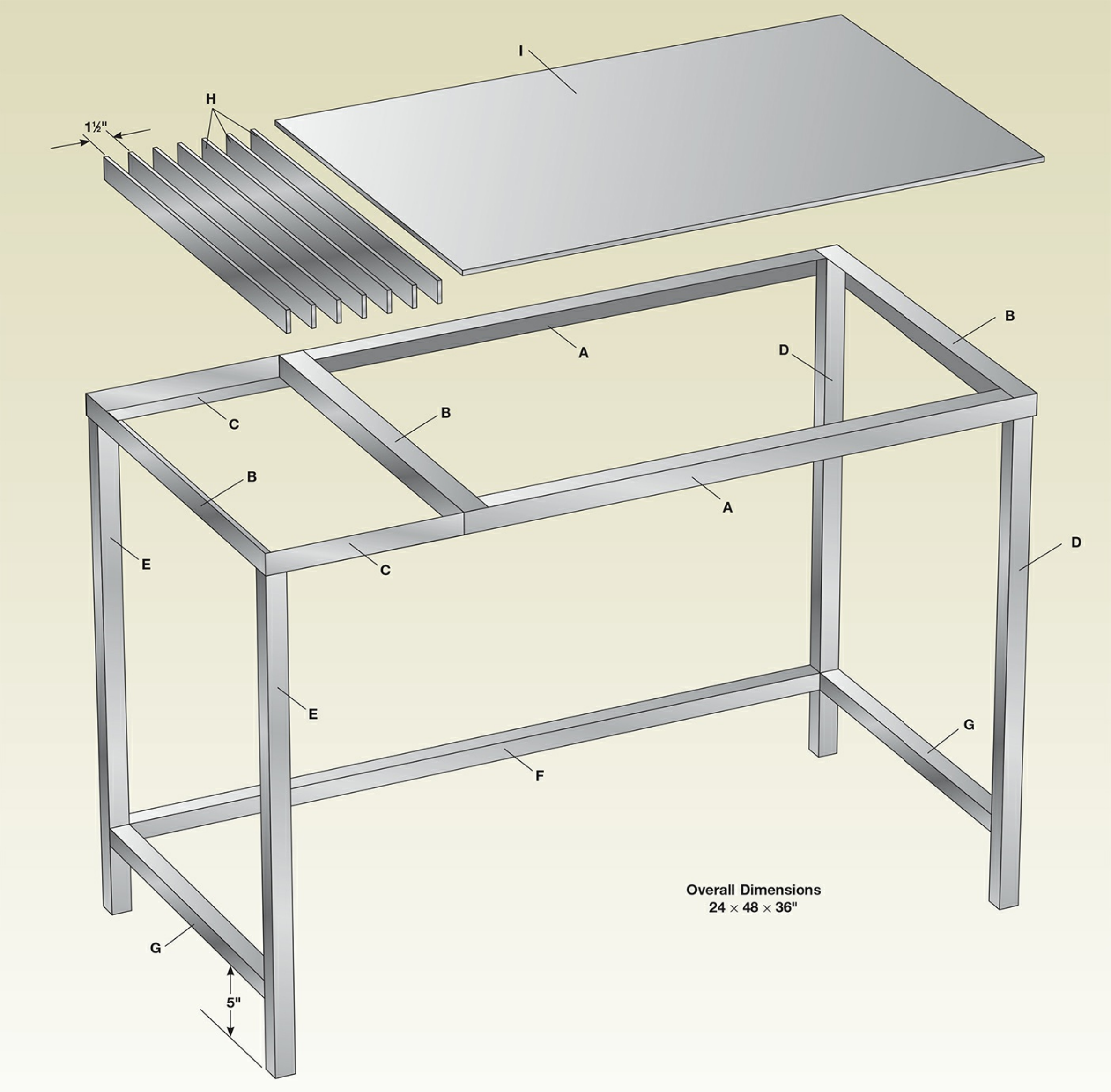
I
H
1½"
B
A
D
C
B
B
A
D
E
C
E
G
F
Overall Dimensions
24 × 48 × 36"
G
5"

A sturdy welding table is the heart of any metal shop. Used for arc welding, cutting, or oxyfuel welding, the welding table is a versatile addition to your shop. The model shown here is sized to allow you to work while sitting on a stool or while standing. You may want to check the scrap bin at the local steel yard to see if you can find a bargain price for the tabletop. Some yards sell plate in 4 foot by 2 foot pre-cut sections that you can cut to size. Or you can have the piece custom cut. You can use material that's thinner than $^{3}/_{16}$", but thicker material is better as the table top will have less distortion from the welding heat and the heat generated when you grind spatter off of it. Set the sheet metal for the table top on top of sawhorses or a workbench for use as a work surface to build the rest of the table.

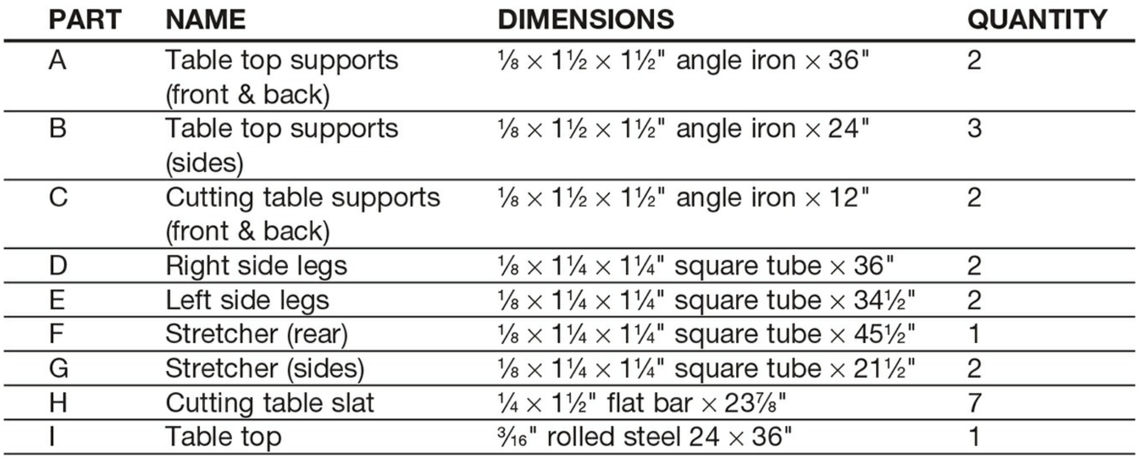

PART	NAME	DIMENSIONS	QUANTITY
A	Table top supports (front & back)	⅛ × 1½ × 1½" angle iron × 36"	2
B	Table top supports (sides)	⅛ × 1½ × 1½" angle iron × 24"	3
C	Cutting table supports (front & back)	⅛ × 1½ × 1½" angle iron × 12"	2
D	Right side legs	⅛ × 1¼ × 1¼" square tube × 36"	2
E	Left side legs	⅛ × 1¼ × 1¼" square tube × 34½"	2
F	Stretcher (rear)	⅛ × 1¼ × 1¼" square tube × 45½"	1
G	Stretcher (sides)	⅛ × 1¼ × 1¼" square tube × 21½"	2
H	Cutting table slat	¼ × 1½" flat bar × 23⅞"	7
I	Table top	³⁄₁₆" rolled steel 24 × 36"	1

Materials

- 1/8 × 1½ × 1½" angle iron (12 feet)
- 1/8 × 1¼ × 1¼" square tube (20 feet)
- ¼ × 1½" flat bar (14 feet)
- 3/16" sheet metal (2 × 2 feet)
- 4 leg levelers

How to Build a Welding Table

Weld the Table Top Supports

Cutting notches for the side supports in the front and back table top supports is easier than cutting 45° miters and provides more welding surface area.

1. Cut the table top supports (A & B) and cutting table supports (C) to length. Cut 1½" notches where each end of the three side supports (B) fits against the table top supports to create a 90° angle joint.

2. Place the table top front support (A) and a table top side support (B) together to form a right angle. Check for square and tack weld.

3. Repeat step 2 using the table top back support and another side support.

4. Assemble these two right angles to make a square. Check all corners for square, and check the assembly for square by measuring both diagonals—they should be equal. If not, adjust the supports so the assembly is square.

5. Complete each outside corner weld, re-checking for square as you go. Flip the assembly over and complete the remaining welds of the joints.

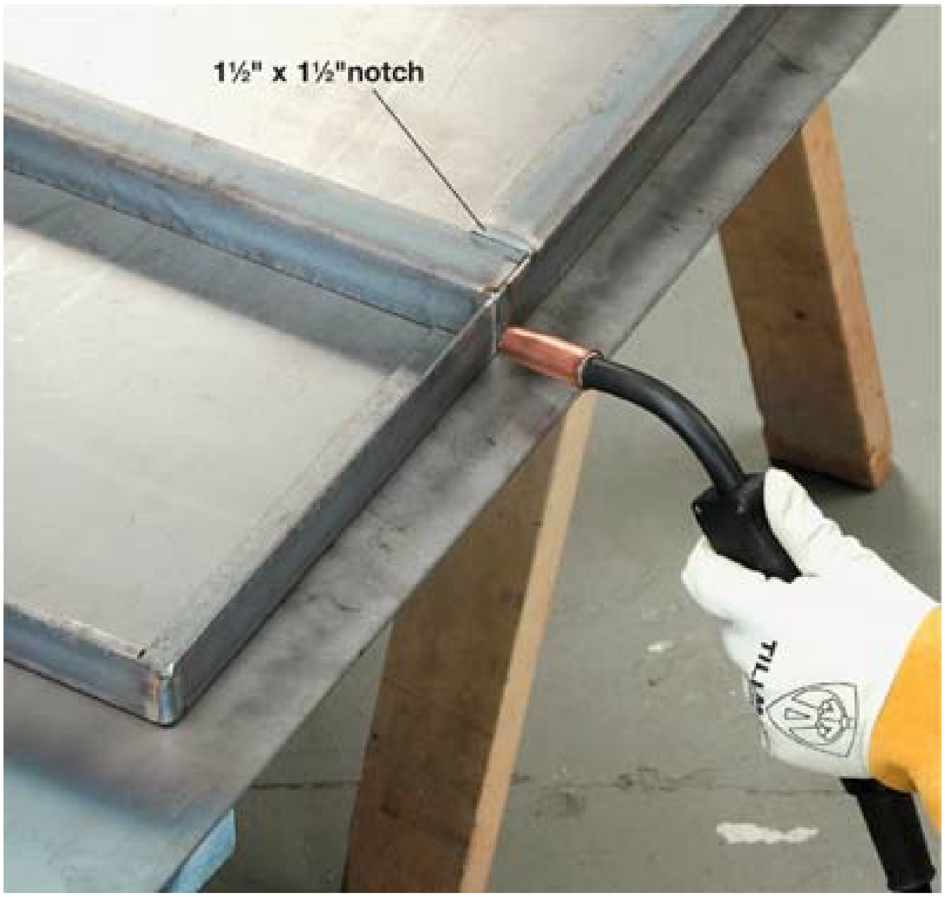

Weld the cutting table section to the table top assembly. The angle iron flange for the table top will face up to support the table top. The angle iron flange for the cutting table will face down to form a well to support the cutting table strips.

Complete the Table Top Supports

The cutting table supports are positioned with the flange at the bottom to hold the ¼" thick strips. The table top supports are positioned with the flange at the top to support the table top (see photo, above left).

1. Place the remaining side support (B) and the front and back cutting table supports (C) at right angles to form three sides of a rectangle. Check the corners for square and tack weld the pieces together.

2. Assemble the pieces so the table top supports have the flange at the top, the cutting table supports have the flange at the bottom, and the cutting table supports are abutting the table top supports. Tack weld where the table top sides butt together. Turn the assembly over and weld the remaining joints.

Prepare the Leg Assembly

1. Cut the legs (D & E) to length.

2. Measure the completed tabletop. It may be slightly more or less than 24" deep and 48" wide. Adjust the stretcher lengths (F & G) to those measurements, minus the 1¼" for each leg thickness. Cut the stretchers to length.

3. Mark both sets of side legs 5" up from the bottom. Starting with the right side legs (D), place a side stretcher (G) between the legs with the bottom of the stretcher aligned with the 5" mark.

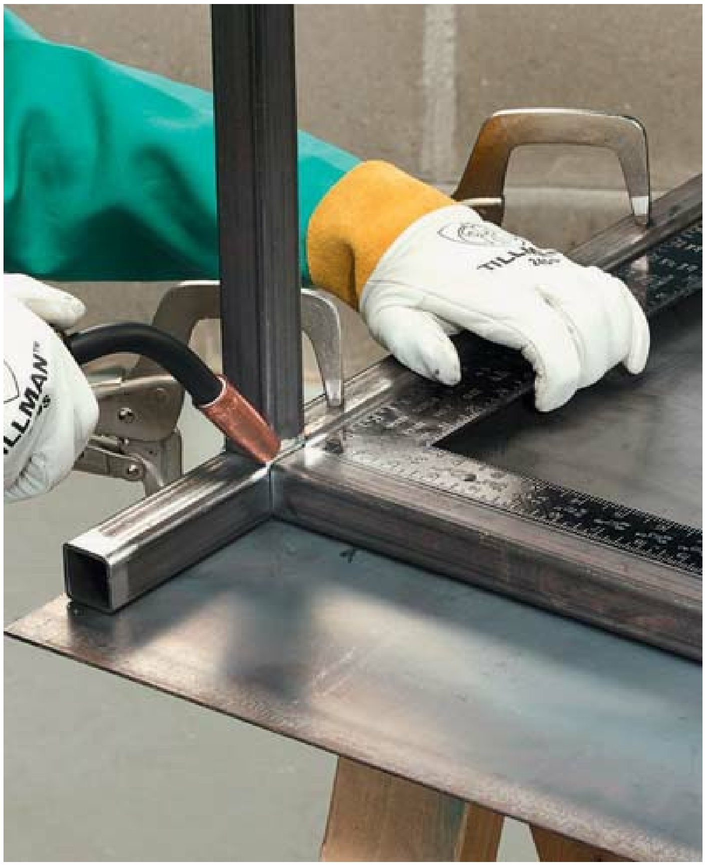

Place the back stretcher between the right and left side leg assemblies. Make sure the stretcher is square to the legs, then weld in place.

4. Align the stretcher at a 90° angle to the side legs and clamp the assembly to your work surface. Tack weld the top inside angle at each end of the stretcher.

5. Repeat this process for the left side legs.

Finish the Leg Assembly

1. Place the leg assemblies on their sides with the rear side down, and clamp to your work surface.

2. Position the rear stretcher (F) between the right and left leg assemblies, aligning the bottom of the stretcher with the 5" mark.

3. Align the stretcher at a 90° angle to the assemblies and clamp in place. Tack weld the inside angles (see photo, bottom previous page).

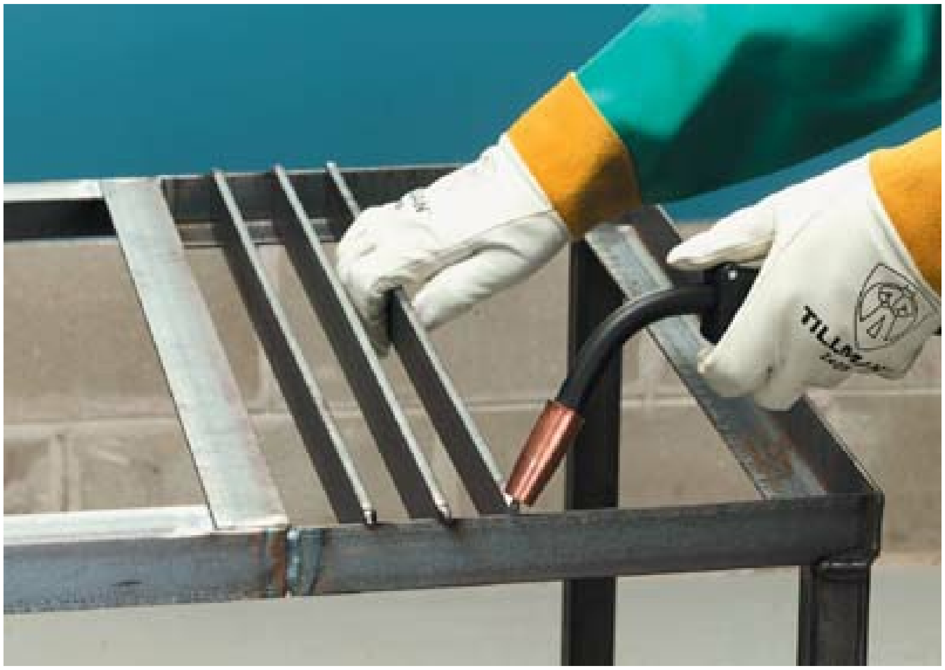

Tack weld cutting table slats to the angle iron every 1½" to form the cutting table top.

Assemble the Table

The right legs fit inside the corner made by the angle iron, while the left legs are set back 1/8" from the edges of the angle iron.

1. Turn the table top and cutting table assembly upside down, then set the leg assembly into it.

2. Clamp the pieces in place. Check for square on both sides, front, and back.

3. Tack weld all corners. Check for square again, then weld all pieces into place.

Install the Cutting Table Top

Tack welding the cutting table grating allows you to remove and replace the ¼" strips as they become worn from the cutting torch.

1. Cut the cutting table slats (H) to length.

2. Place a slat onto the ledge formed by the cutting table top supports 1½" from the table top edge.

3. Tack weld the top edge of the slat to the angle iron.

4. Place another slat 1½" from the first slat and tack weld in place. Continue building the cutting table top in this manner until complete (see photo, above).

5. Grind down the welds on the top of the table top support assembly.

Finish the Table

1. Place the table top (I) onto the assembly and tack weld twice on each side.

2. Weld the table top to the supports using 1 or 2" weld beads at both sides of each corner and twice along each side.

3. Grind down the welds, if desired. Wire brush, sand, or sandblast the entire table.

4. Paint the table, but do not paint the table top or cutting grate.

5. Install leg levelers as needed.

Welding Machine Cart

You’ll want a cart for your GMAW or GTAW machine and gas cylinder—why not build it as your first project? It’s good practice, and the materials will cost much less than most commercial carts. If you keep your eyes open, you can probably pick up the angle iron for free the next time a neighbor throws away an old bedframe. Remember, however, to grind off the paint at the joints, otherwise you’ll have nasty fumes and poor- quality welds.

Include a rubber pad and insulated chain for the gas cylinder (not shown) so an arc can’t accidentally be struck against the cylinder.

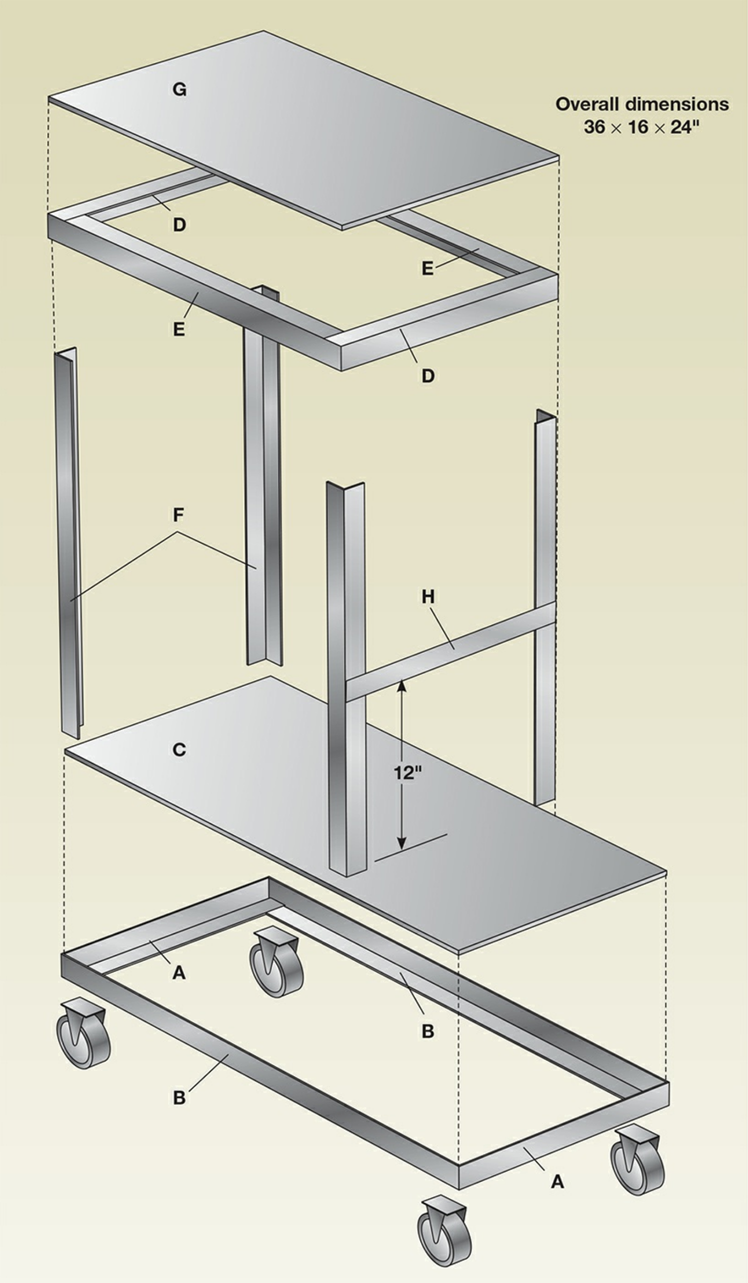
G
Overall dimensions
36 × 16 × 24"
D
E
E
D
F
H
C
12"
A
B
B
A

Materials

- 1/8 × 1½ × 1½" angle iron (24 feet)
- 1/8" sheet metal (16 × 64")
- 1/8 × 1¼" flat bar (16")
- Swivel casters (4)

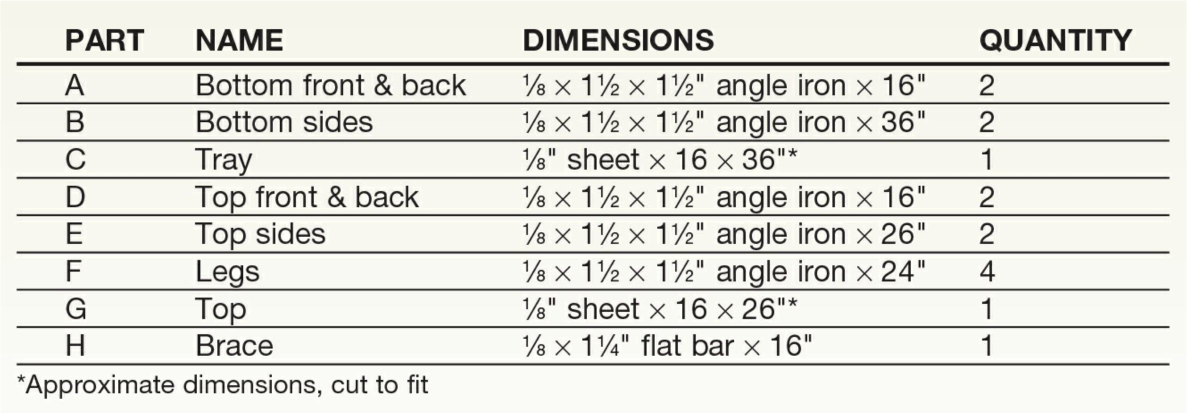

PART	NAME	DIMENSIONS	QUANTITY
A	Bottom front & back	⅛ × 1½ × 1½" angle iron × 16"	2
B	Bottom sides	⅛ × 1½ × 1½" angle iron × 36"	2
C	Tray	⅛" sheet × 16 × 36"*	1
D	Top front & back	⅛ × 1½ × 1½" angle iron × 16"	2
E	Top sides	⅛ × 1½ × 1½" angle iron × 26"	2
F	Legs	⅛ × 1½ × 1½" angle iron × 24"	4
G	Top	⅛" sheet × 16 × 26"*	1
H	Brace	⅛ × 1¼" flat bar × 16"	1

*Approximate dimensions, cut to fit

How to Build a Welding Machine Cart

Assemble the Base & Tray

1. Cut the bottom front and back (A) to length. Cut the bottom side pieces (B) to length. Instead of mitering the corners, cut a 1½" notch at each end.

2. Arrange the front, back, and side pieces to form a rectangle. Check the corners for square, clamp in place, and tack weld the pieces together.

3. Measure diagonally across the corners to make sure they are square. If the measurements are equal, the base is square. If they are not, adjust the pieces until they are.

4. Clamp the base in place. Weld each outside corner. Turn the base over and weld each joint. Do not weld inside the base, as the weld beads will interfere with the tray and leg placement.

5. Measure the inside dimensions of the base and cut the tray (C) to size.

6. Insert the tray into the angles of the base. Tack weld the tray at points along the perimeter.

7. Turn the assembly over and place three 1" welds along each long side, and two 1" welds along each short side to join the tray to the base.

Assemble the Top

1. Cut the top front and back (D) and top side pieces (E) to length. Cut a 1½" notch at each end of the side pieces.

2. Arrange the front, back, and side pieces to form a rectangle. Check the corners for square, clamp in place, and tack weld the pieces together.

3. Measure diagonally across the corners to make sure they are square. If the measurements are equal, the top assembly is square. If they are not, adjust the pieces until they are.

4. Clamp the pieces in place. Weld each outside corner. Turn the assembly over and weld each joint. Do not weld inside the top assembly, as the weld beads will interfere with the top and leg placement.

5. Measure the top assembly to allow a 1/8" setback on all sides, and cut the top (G) to size.

6. Grind down the welds and place the top over the assembly. Tack weld the top along the perimeter, then finish weld with two 1" welds per side.

Attach the Legs

1. Cut the legs (F) and brace (H) to size.

2. Turn the top upside down. Place a leg in one of the corners. Use a scrap piece of angle iron to assist in clamping (see photo). Make sure the leg is square to the top and tack weld into place.

3. Repeat step 2 to attach the other three legs to the corners of the top.

4. Place the legs and top assembly into the bottom assembly. The back legs should fit into the corners of the bottom assembly, and the front legs into the angles of the side pieces of the base.

5. Make sure the legs are square to the base, clamp, and tack weld.

6. Measure across the diagonals between the base and the top to make sure the unit is square. Adjust, if necessary, and finish welding the legs to the top and base.

7. Measure 12" from the bottom of the front legs. Center the brace over the marks, check it for square, and clamp in place. Weld the brace to the front legs.

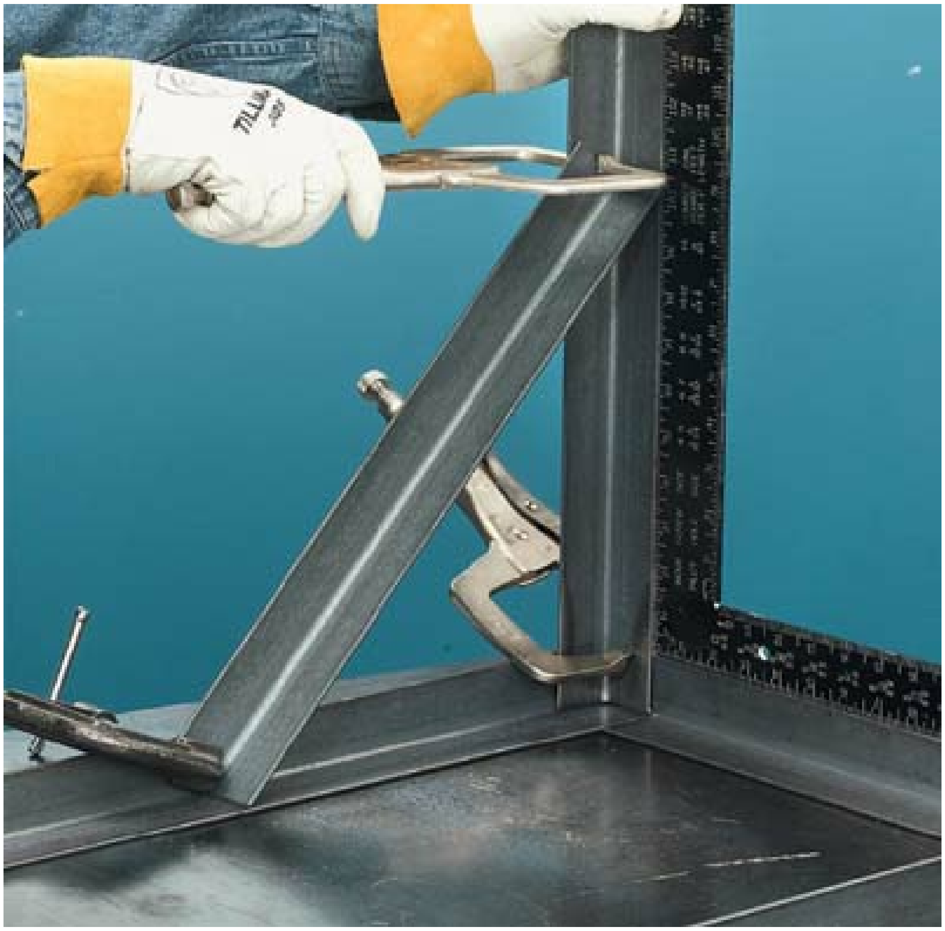

Use a scrap of angle iron to hold the legs in place for tacking.

Add Finishing Touches

1. Sand and clean the cart, and paint.

2. Drill holes 2" in from the ends of the brace and attach the eyebolts. Use the threaded chain links to attach the chain.

3. Turn the cart over. Grind down the welds at the corners and attach the casters. If you are welding the casters, make sure you grind off the zinc coating.

Cylinder Cart

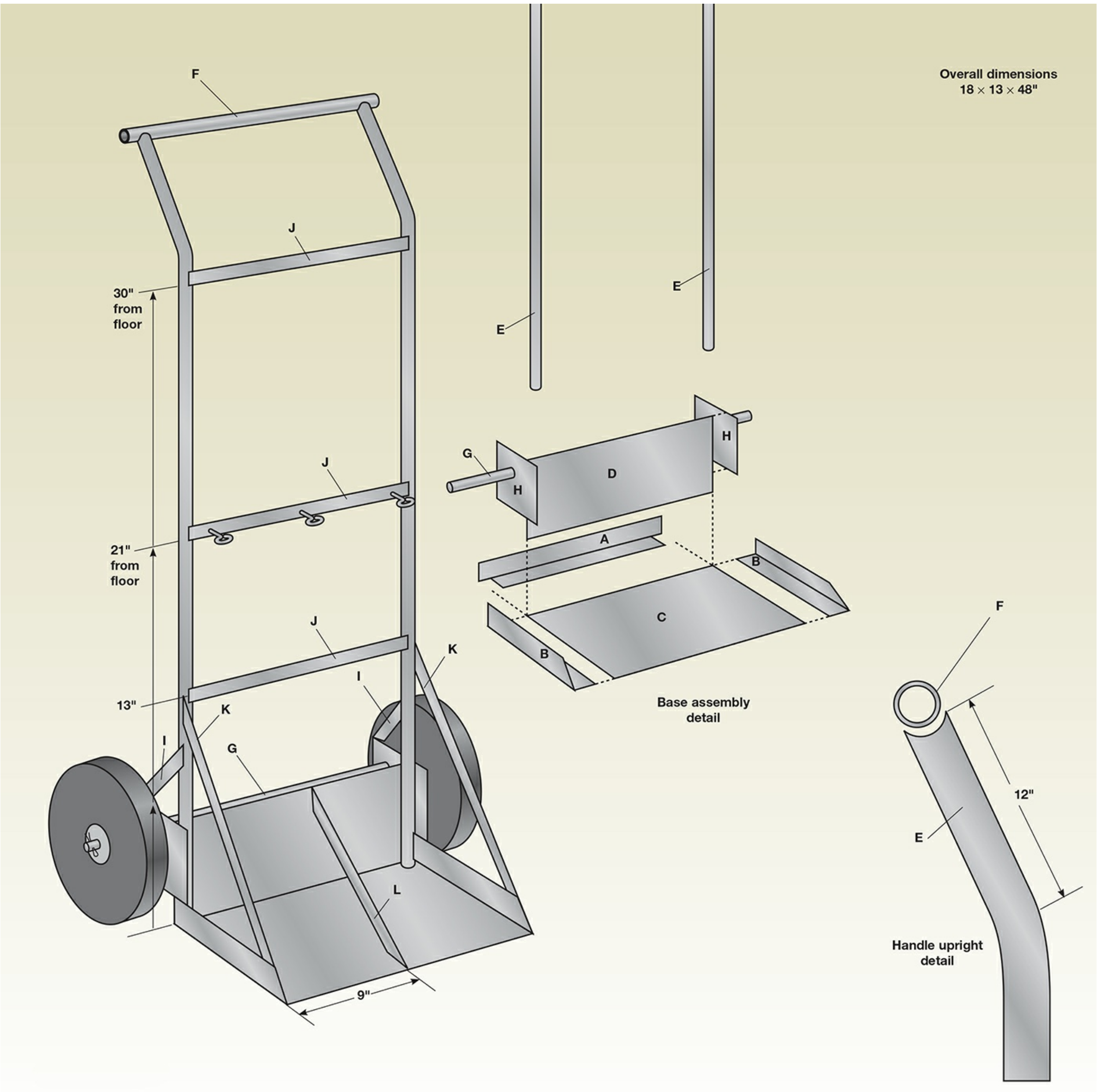
F
J
30" from floor
J
21" from floor
J
13"
K
I
K
I
G
L
9"
E
E
Overall dimensions 18 × 13 × 48"
G
H
H
D
A
B
C
B
Base assembly detail
F
12"
E
Handle upright detail

You can purchase a cylinder cart for your oxyfuel rig—your dealer might even give you a discount—but it’s fun and challenging to make your own. This cart is constructed with 1/8" stock, which is oxyfuel weldable, but we used GMAW. The base platform and back support can be flame or plasma cut. If you have access to a heavy-duty metal brake, you can make the base platform and back support from one 16¼ × 17¾" piece and bend it to 90° (allow ¼" for the radius of the bend).

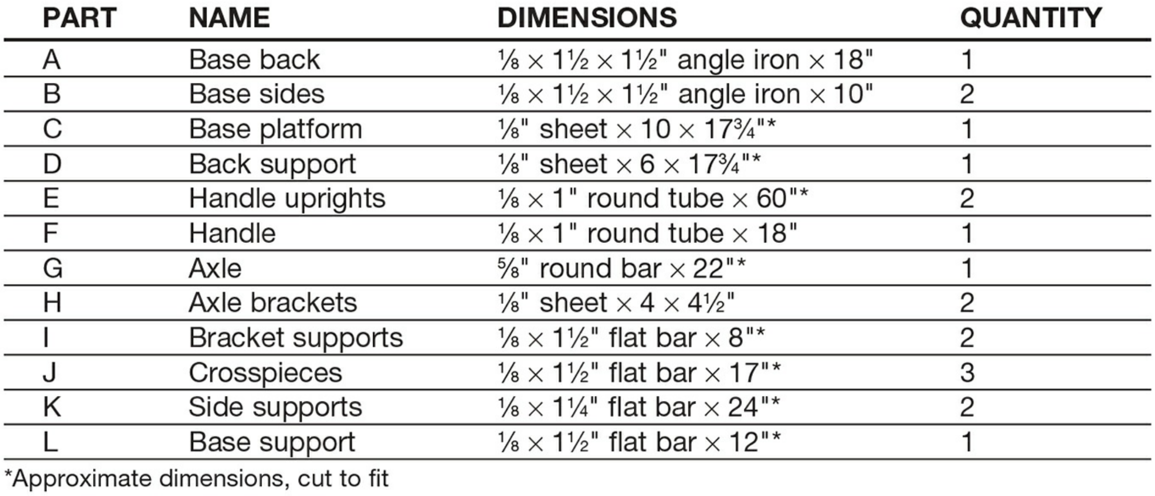

PART	NAME	DIMENSIONS	QUANTITY
A	Base back	⅛ × 1½ × 1½" angle iron × 18"	1
B	Base sides	⅛ × 1½ × 1½" angle iron × 10"	2
C	Base platform	⅛" sheet × 10 × 17¾"*	1
D	Back support	⅛" sheet × 6 × 17¾"*	1
E	Handle uprights	⅛ × 1" round tube × 60"*	2
F	Handle	⅛ × 1" round tube × 18"	1
G	Axle	⅝" round bar × 22"*	1
H	Axle brackets	⅛" sheet × 4 × 4½"	2
I	Bracket supports	⅛ × 1½" flat bar × 8"*	2
J	Crosspieces	⅛ × 1½" flat bar × 17"*	3
K	Side supports	⅛ × 1¼" flat bar × 24"*	2
L	Base support	⅛ × 1½" flat bar × 12"*	1

*Approximate dimensions, cut to fit

Materials

- 1/8 × 1½ × 1½" angle iron (3½ feet)
- 1/8" sheet metal (18 × 24" sheet)
- 1/8 × 1" round tube (12 feet)
- 5/8" round bar (2 feet)
- 1/8 × 1½" flat bar (7 feet)
- 1/8 × 1¼" flat bar (4 feet)
- 1" eyebolts with nuts (3)
- Chain
- Snap closures or threaded chain links (3)
- 8" wheels with hubs (2)
- Washers and cotter pins (2)

How to Build a Cylinder Cart

Prepare the Cart Base

1. Cut the base back (A) and base sides (B) to size. Rather than mitering the corners, cut a 1½" notch in each end of the base back (see diagram).

2. Mark and cut a triangular section off the front end of each base side to soften the edge.

3. Lay out the base back and sides to form a three-sided rectangle. Square the corners and clamp the assembly to your work surface.

4. Tack the joints at both ends of the base back. Turn the assembly over and weld the corner and butt joints.

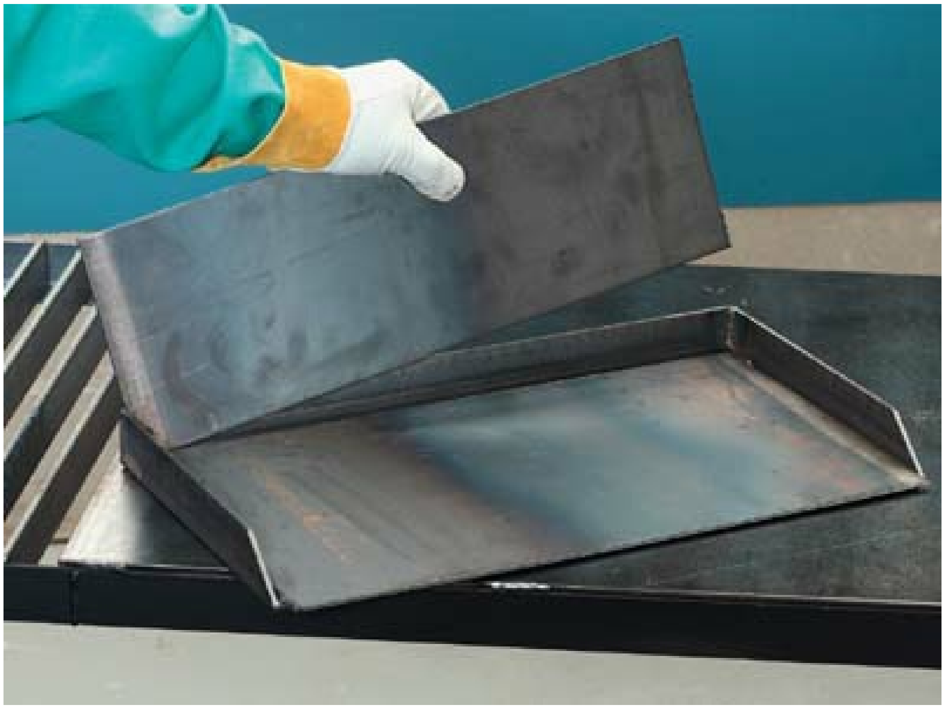

After welding the base back and sides, cut the base platform and back support to fit, then weld them in place.

Install the Base Platform & Back Support

1. Measure the inside width of the base assembly and cut the base platform (C) and back support (D) to size.

2. Place the base platform into the base assembly and tack weld it into place along the base sides.(Do not weld along the base back, as that would prevent the back support from fitting flat against the base platform.)

3. Turn the base over and place two 1" welds along each side where the base platform meets the back and sides.

4. Set the base flat on the work surface, right-side up, and place the back support upright against the inside of the base back (see photo, above).

5. Clamp the back support to the base back. Place three 1" welds on the back side of the support where it meets the angle iron.

6. Place two 1" welds on the inside T-joint, making sure not to weld within two inches of the corner.

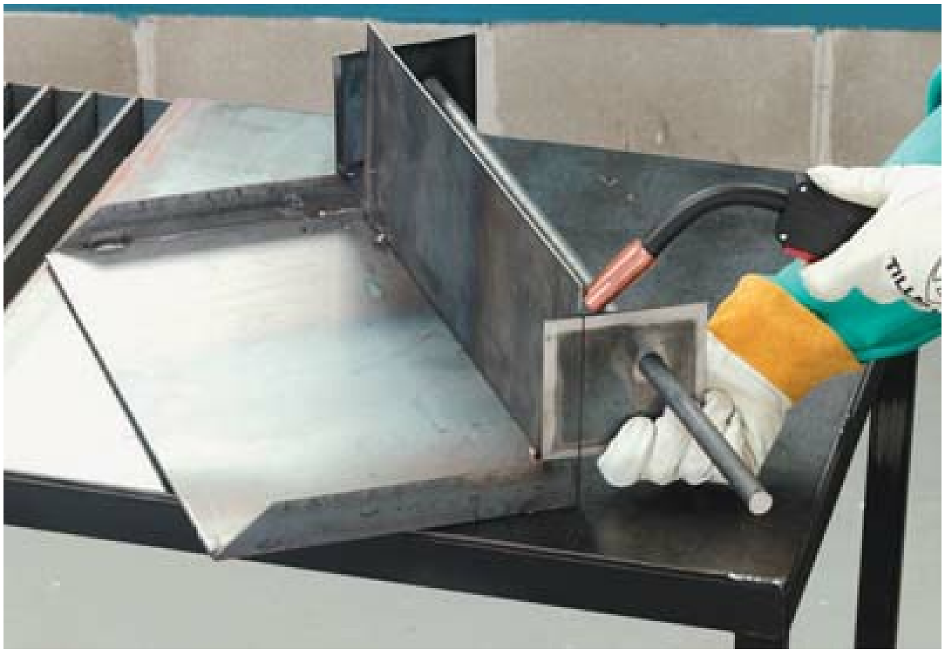

Tack the axle brackets to the base assembly. Insert the axle to check for proper alignment, then finish weld the bracket joints.

Attach the Axle Brackets

1. Cut the axle brackets (H) to size.

2. Place the base assembly on its side and align an axle bracket along the top edge of the base side, resting it on top of the back support and extending 3" off the back side.

3. Measure 4" up from the bottom of the base assembly and 1½" from the back edge to mark the axle location on the axle bracket. Drill, flame cut, or plasma cut a 5/8" hole at the mark.

4. Tack weld the axle bracket in place (see photo, above left). Insert the axle to check for proper alignment, then weld both sides of the T-joint between the bracket and the back support.

5. Weld the outside butt joint between the bracket and the base side.

6. Repeat steps 2 to 5 to weld an axle bracket to the other side of the base assembly.

Prepare the Handle Uprights

1. Cut the handle uprights (E) to length.

2. Make a 20 to 30° bend in one of the handle uprights 13" from one end using a heavy-duty conduit bender. Bend the other handle upright to match.

3. Cut both pieces so the curve is 12" from the top end and the overall upright height of each is 48". Grind the tops of both handle uprights so they fit around the handle tube.

Assemble & Install Handle Uprights

1. Place a handle upright against the inside corner of the base assembly. Turn the handle so the curve points straight back. Tack weld the handle in place. Repeat this process for the other handle upright.

2. Cut the handle (F) to size.

3. Place the cart on its back. Set the handle against the cutouts in the handle uprights keeping an equal amount of overhang on each end and tack weld in place.

4. Finish weld the handle uprights to the base assembly. Place welds between the handle uprights and top back support, and between the handle uprights and axle brackets.

Attach the Crosspieces

1. Measure and cut the crosspieces (J) so they are slightly recessed against the handle uprights.

2. Drill three 3/16" holes—one in the center and one on each end of one crosspiece. Place this crosspiece 21" up from the base.

3. Set the remaining crosspieces at 13" and 30" from the base on the forward side of the handle uprights, and weld in place (see photo, right).

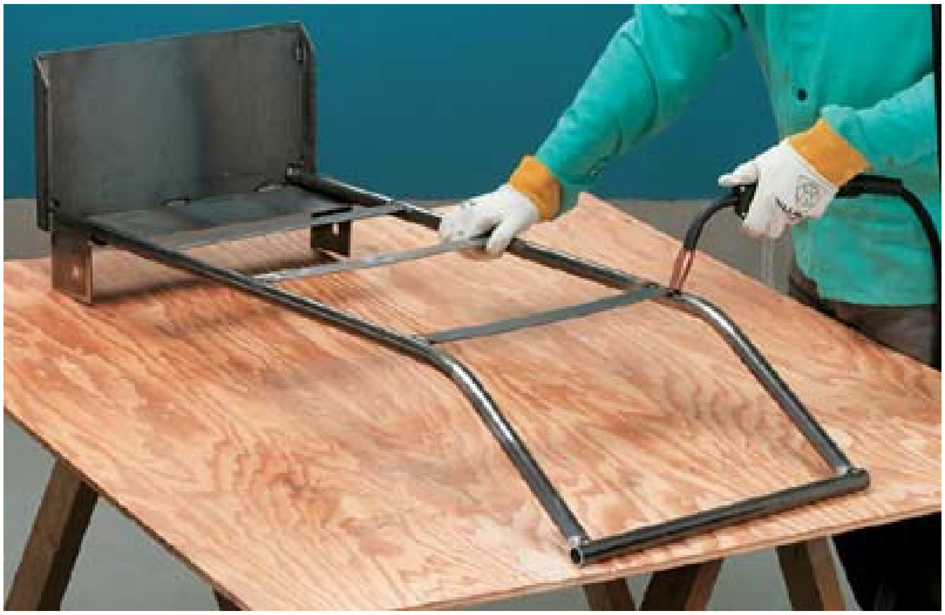

Lay the crosspieces across the handle uprights at the proper heights. Weld the pieces to the uprights.

Attach the Supports

1. Turn the assembly on its side. Place a bracket support (I) in position across the handle upright and the axle bracket. Mark the angles on the support and cut it to size.

2. Weld the bracket support in place against the handle upright and axle bracket.

3. Set the side support (K) in position against the base side and the handle upright. Mark the angles and cut to fit. Weld the support in place.

4. Turn the assembly on its opposite side and repeat steps 1 to 3.

5. Mark the angles and cut the base support (L) to size to fit between the base platform and the base back.

6. Center the base support from side to side on the base assembly. Weld both sides of each T-joint.

Complete the Cart

1. Center the axle (G) between the axle brackets. Place the wheels on the axle to make sure the wheels barely make contact with the floor (see photo, right).

2. Mark the axle for its final length and cut to size. Drill holes in each end of the axle to fit the cotter pins.

3. Center the axle between the axle brackets and weld in place.

4. Complete any unfinished welds.

5. Wire brush or sandblast the cart. Clean the cart and paint as desired.

6. Attach the wheels to the axle. Insert cotter pins in the axle holes.

7. Affix eyebolts with nuts in the 3 holes in the middle crosspiece. Attach the chain to the eyebolts with snap closures or threaded chain links.

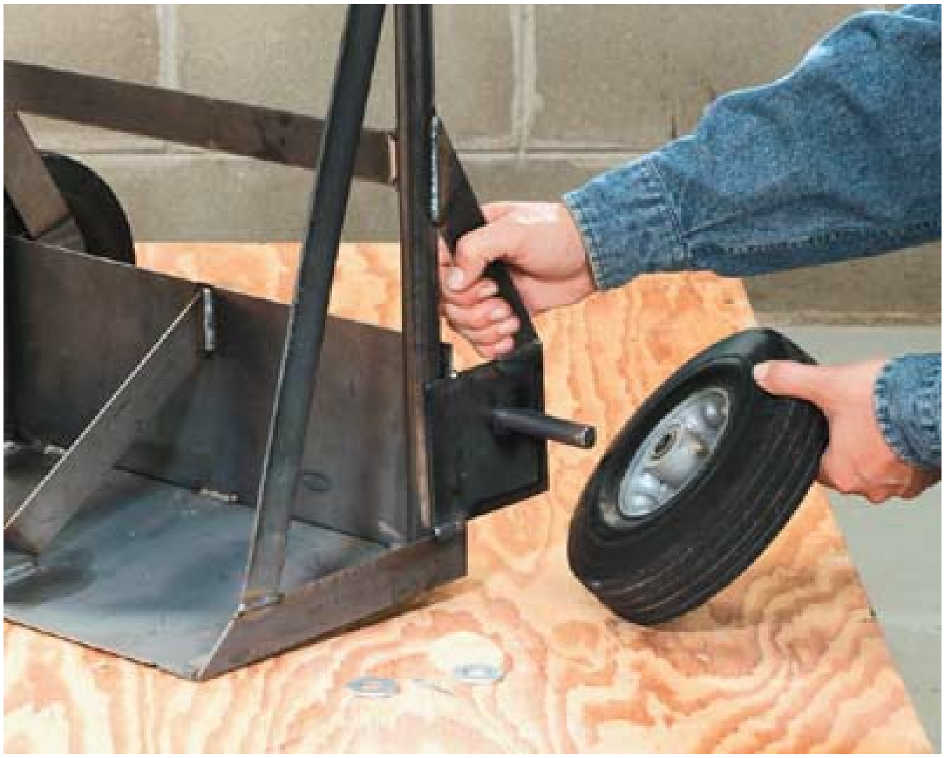

Trial fit the wheels before welding the axle into place. The wheels should just barely make contact with the floor.

Grinder Stand

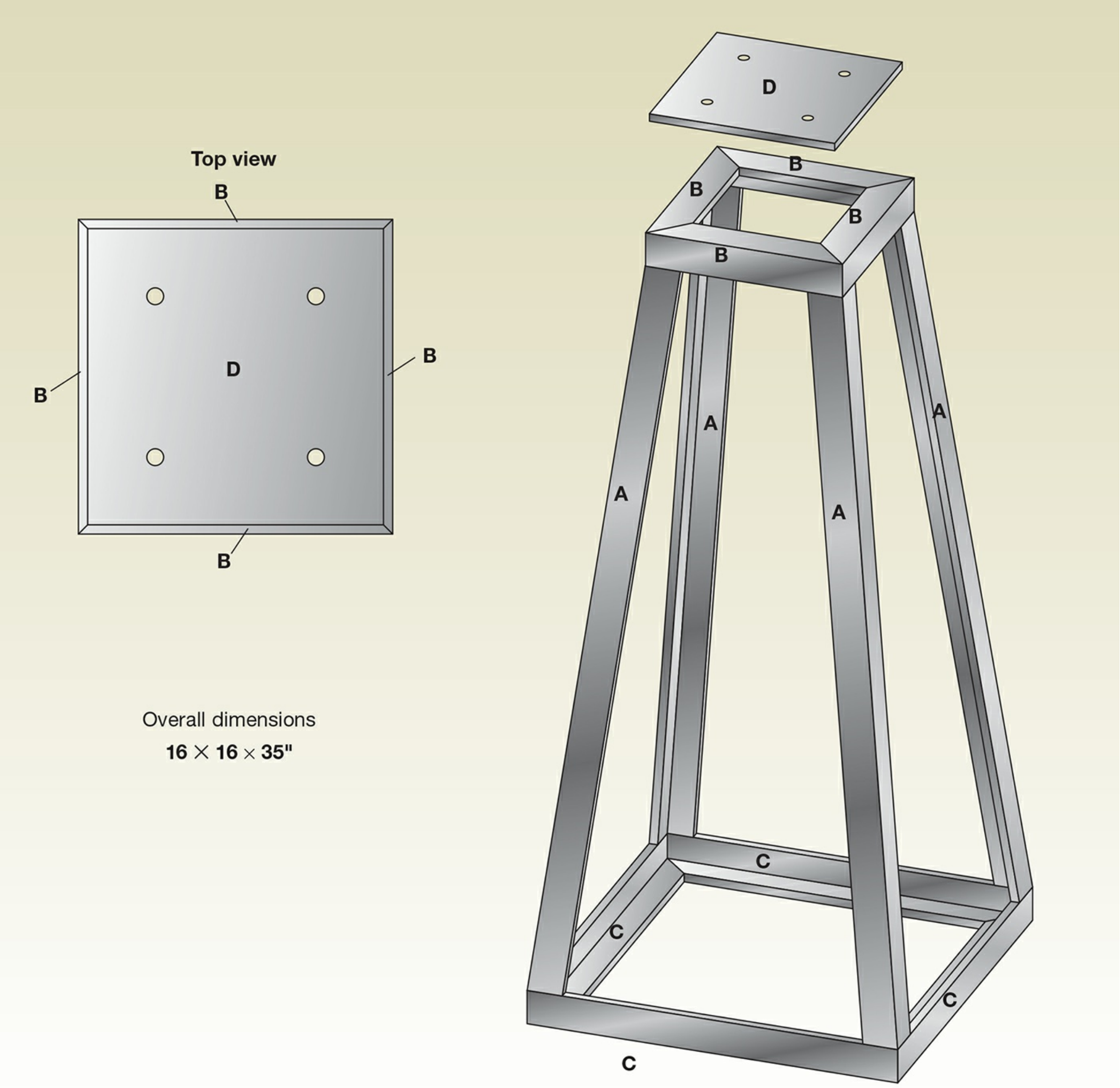
Top view
B
B
D
B
B
Overall dimensions
16 × 16 × 35"
D
B
B
B
B
A
A
A
A
C
C
C
C

This is a good project to make with angle iron from the discount bin at the local steel supply center. You can often buy short pieces for 10 cents a pound rather than the 30 to 50 cents you would normally pay. Take along your calipers or micrometer to measure thicknesses, as well as your measuring tape, cutting list, and tough gloves so you don’t get cut digging through the scrap pile.

Materials

- 3/16 × 1½ × 1½" angle iron (19 feet)
- 3/16" plate (8 × 8")

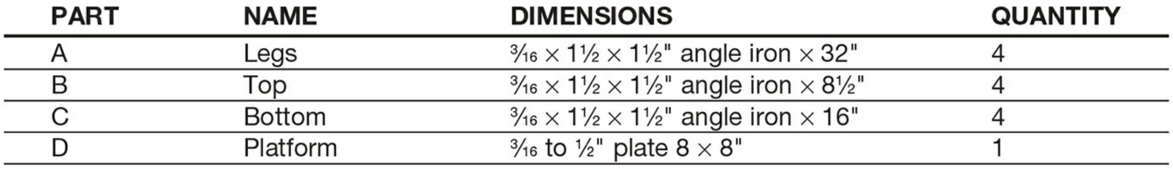

PART	NAME	DIMENSIONS	QUANTITY
A	Legs	3/16 × 1½ × 1½" angle iron × 32"	4
B	Top	3/16 × 1½ × 1½" angle iron × 8½"	4
C	Bottom	3/16 × 1½ × 1½" angle iron × 16"	4
D	Platform	3/16 to ½" plate 8 × 8"	1

How to Build a Grinder Stand

Build the Top & Bottom Assemblies

1. Cut the top and bottom pieces (B & C) to length. Miter the corners at 45° or cut 1½" notches into the pieces.

2. Assemble pairs of bottom pieces into right angles. Clamp the pieces to your work surface and tack weld at each corner.

3. Assemble the two right angles into a square. Check for square by measuring across each diagonal—if the measurements are equal, the assembly is square. Tack weld the corners.

4. Repeat steps 2 to 3 to assemble the top.

Mark & Cut the Legs

1. Cut the legs (A) to size.

2. Place the bottom assembly on your work surface and loosely clamp the legs into each corner.

3. Set the top assembly over the legs and clamp it in place. Make sure the top is level.

4. Mark each leg at the top and bottom where they cross the platforms (see photo).

5. Unclamp the legs and cut to size.

Assemble the Stand

1. Place one leg on top of a corner of the bottom assembly and tack weld it in place. Tack weld the other three legs in the corners the same way.

2. Align a corner of the top assembly with the top of one leg and tack it in place. Tack weld the top assembly to the other three legs the same way.

3. Adjust the stand so the top is level. You may need to remove tack welds and grind down some angles.

4. When you are satisfied with the alignment, complete the welds. To minimize distortion, alternate between the sides, top, and bottom as you weld.

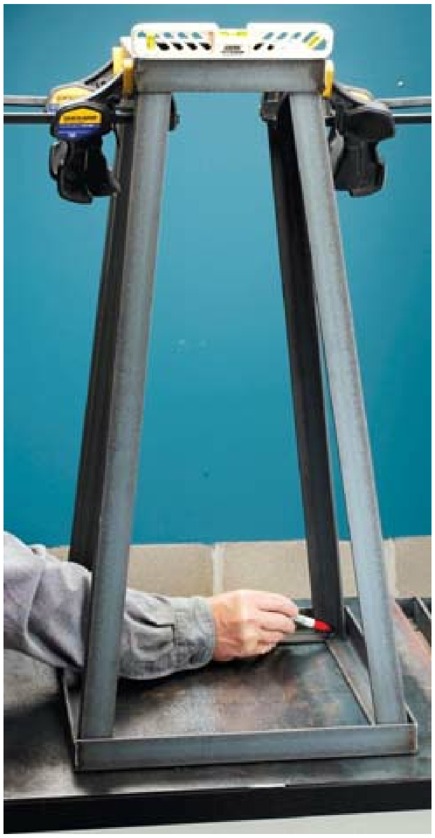

Assemble the grinder stand using clamps. Mark the top and bottom of each leg and cut to fit.

Attach the Platform Plate

1. Cut the platform plate (D) to size. Drill four holes in the face to match your grinder mounting holes.

2. Set the plate over the top of the stand, aligning the edges with the outside of the top pieces. Weld the plate to the stand.

Portable Welding Table

Building a heavy-duty welding table helps you develop your welding skills and it also provides a useful piece of equipment on which to build future projects. Because the table seen here is mounted on casters it can be rolled away into a corner of the garage when not in use. The metal-slat tabletop can be covered with a removable sheet-metal welding surface that can be cleaned and reused as necessary. You can also build a plywood work surface for doing woodworking and metalworking that doesn’t involve welding.

The welding table dimensions can be altered to meet the requirements of the welder. You could even build two tables of the same height to create an extended work surface. In short, this table is a very versatile tool and the more you use it the more you’ll appreciate it.

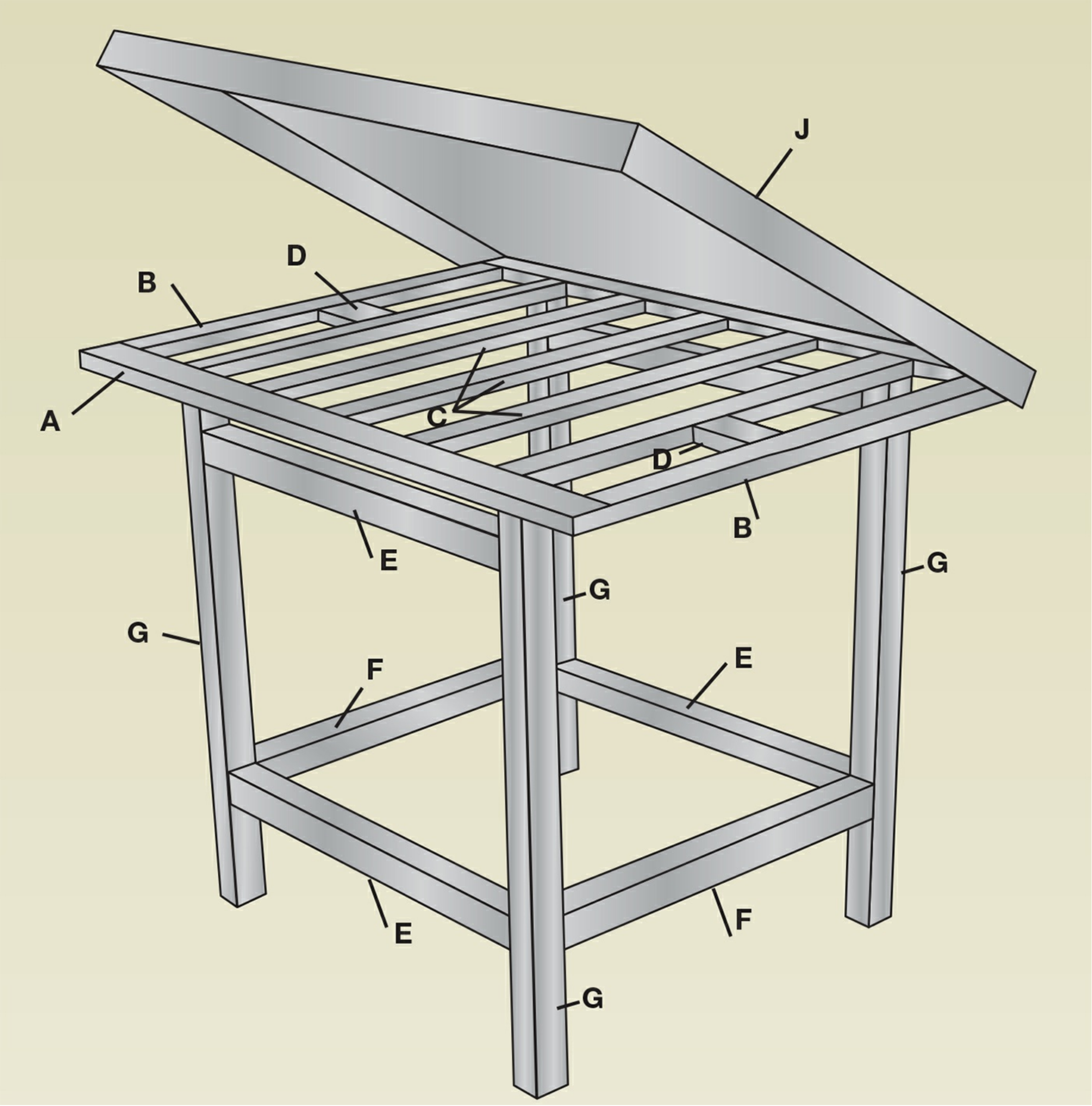
J
D
B
A
C
D
B
E
G
G
G
F
E
E
F
G

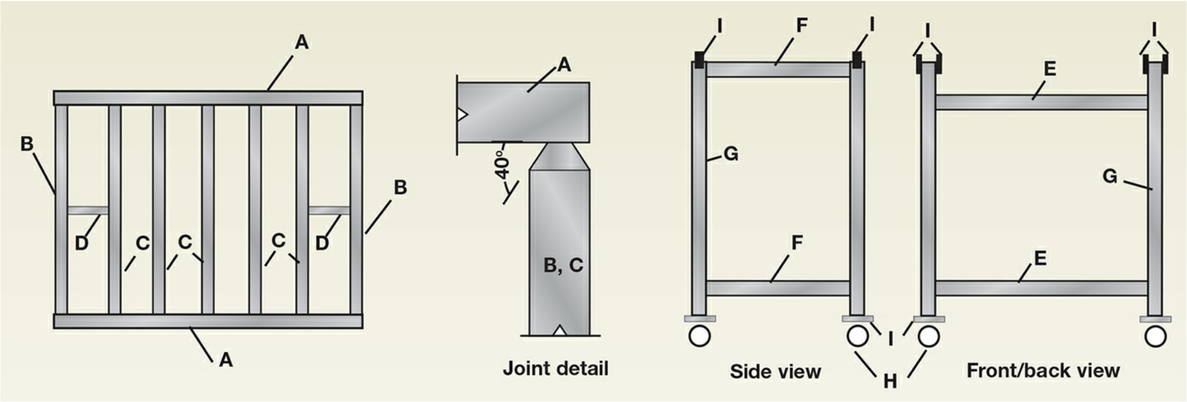
A
B
B
D
C
C
C
D
A
A
40°
B, C
I
F
I
G
F
I
H
I
I
E
G
E

Joint detail

Side view

Front/back view

Materials

- ¾ x 2" cold-rolled steel bar (CRS) stock (23 feet)
- 3/16" x 2 x 2" square tube (22 feet)
- 4 Locking swivel casters (rated for at least 500 lbs. each)
- ¼" flat stock sheet metal (3 x 4 feet, allows extra for tabs)

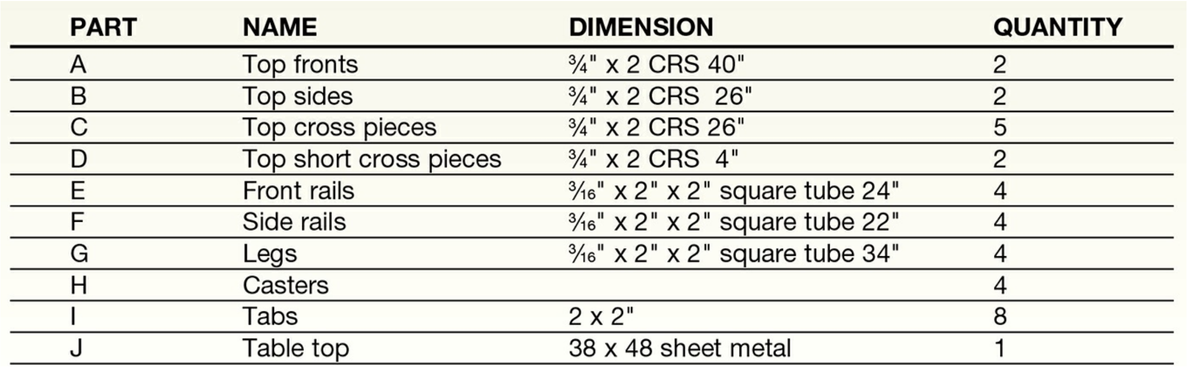

PART	NAME	DIMENSION	QUANTITY
A	Top fronts	¾" x 2 CRS 40"	2
B	Top sides	¾" x 2 CRS 26"	2
C	Top cross pieces	¾" x 2 CRS 26"	5
D	Top short cross pieces	¾" x 2 CRS 4"	2
E	Front rails	3/16" x 2" x 2" square tube 24"	4
F	Side rails	3/16" x 2" x 2" square tube 22"	4
G	Legs	3/16" x 2" x 2" square tube 34"	4
H	Casters		4
I	Tabs	2 x 2"	8
J	Table top	38 x 48 sheet metal	1

How to Build a Welding Table

Before welding, thoroughly clean all parts with denatured alcohol.

Build the Base Sides

1. Measure and mark 3/16" × 2 × 2 square tube for the four front rails (E) at 24" each, the four side rails (F) at 22" each, and the four legs (G) at 34" each.

2. Cut the pieces using a portable band saw or hardened-carbide-blade cut-off saw.

3. Align the side rails (F) flush in between two legs (see page 97): the uppermost side rail (F) is flush with the tops of the legs; the lower sides are 6" up from the bottoms of the legs. Use corner magnets to hold the lower sides square. To maintain your proper dimensions, use a 24" piece of scrap metal or 2 × 4 as a spacer between the two lower ends. Alternatively, tack weld a piece of metal in between the two pieces, and remove it later.

4. Tack the side rails in place. Repeat step 3 for the other side frame.

5. Align two front rails (E) between the side frame legs for the front: the uppermost front is 2" down from the top edge of the legs; the lower front is 6" up from the bottom edges of the legs. To maintain your proper dimensions, use a 22" piece of scrap metal or 2 × 4 as a spacer between the two lower ends. Alternatively, tack weld a piece of metal in between the two pieces, and remove it later. Tack a 6" scrap metal piece below the lowest part E so that the entire front/back frame can stand in place on its own as you tack the pieces to the legs.

6. Tack the front rails (E) to the legs in between both of the two side frames.

Measure and mark 2 x 2 x $^{3}/_{16}$" square tube for the four front rails (E) at 24" each, the four side rails (F) at 22" each, and the four legs (G) at 30" each.

Add the Tabs & Casters

1. Measure and mark ¼" flat stock for the tabs (I) at 2 × 2.

2. Check to make sure each tab is square, and then cut out eight tabs.

3. Align the tabs (I) no more than $1^{1}/_{16}$" above the top of each leg (see page 97).

4. Hold the tabs in place with clamps, and then tack the tabs to the verticals. The top cross pieces (C) closest to the top ends (B) fit in between the tabs.

5. Center casters over the ends of the legs.

6. Tack the casters, and then check for square. Adjust as necessary and then weld. Use SMAW; $^{3}/_{32}$ or 1/8" diameter E 6013 electrode to weld on caster plates. Continuous welds are not necessary.

7. Weld all the way around each joint for the entire base frame tacked together thus far. Note: don't accidentally weld your temporary spacer bars that are used only to maintain dimension.

Build the Top Frame

1. Measure and mark two ¾" cold-rolled steel bars to 40" for the top fronts (A). Cut them to length using a hardened-carbide-blade cut-off saw.

2. Also measure and mark ¾" cold-rolled steel bars to 26" each for two top sides (B) and five top cross pieces (C).

3. Measure and mark ¾" cold-rolled steel bars to 4" each for the two top short cross pieces (D).

4. Cut each of the pieces (A, B, C, D) for the tabletop frame following your cut lines.

5. Make the 40° end joints in each cross piece (see JOINT DETAIL) by grinding or flame-cutting the ends. The depth of each double bevel should be between ¼" and 1/8".

6. Align the top sides (A) and top ends (B), and then check for square. Hold them in place with corner magnets.

7. Tack together all four sides. Use enough heat to get good fusion so when you move the frame the tacks won't break (metal is heavy).

8. Lay the top frame right-side up on a level table and start to align the cross pieces. Start in the center and evenly space each piece in between the two top sides (B). Make sure the cross pieces closest to the top sides are aligned so they fit in between the tabs that are welded onto the legs.

9. Align each cross piece and hold it in place with corner magnets or clamps. Once everything is aligned properly, tack the cross pieces to the top fronts. Note: 4.5" TYP indicates the "typical" spacing between the five middle crossbars is 4½".

Cut the square tube using a metal-cutting bandsaw (shown) or cut-off saw with a hardened carbide blade.

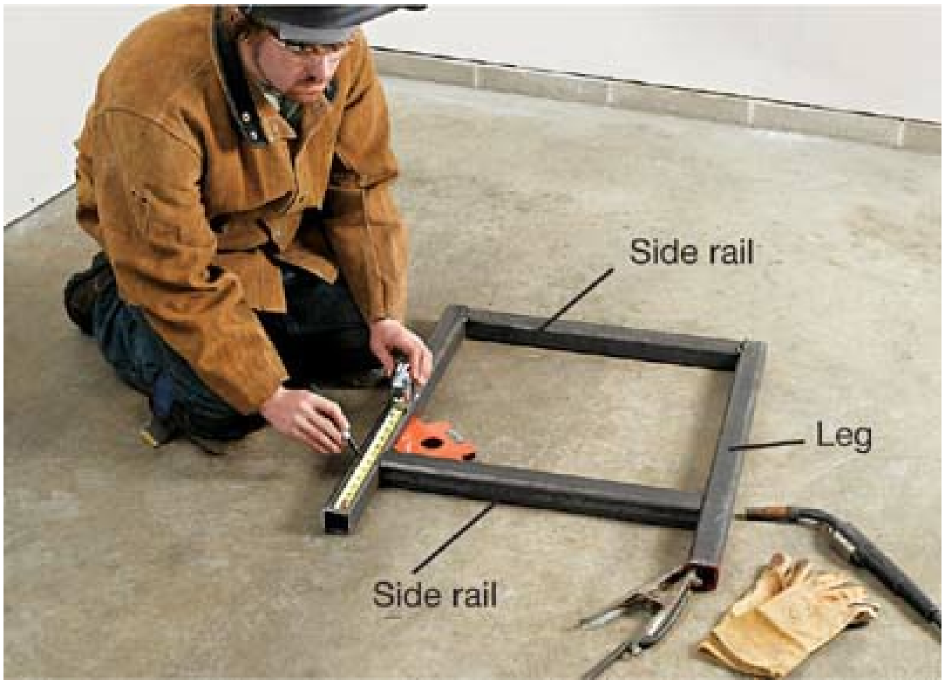
Side rail
Leg
Side rail

Align the side rails in between legs. The side rail part is 6" up from the leg bottom. You will make two of these side frames and then add front rails in between the sides for the front and back of the table base.

10. Align the top short cross pieces in between the top sides (B) and cross pieces (C). Check for square, and then tack in place.

11. Check for final square of the entire table top frame, and make any final adjustments. Align the table top frame over the base to make sure the cross pieces fit in between the tabs on legs.

12. Complete all welds using sequence welding. Alternate from side to side, starting at the center cross pieces and moving to the outer cross pieces.

Make the Table Tops

1. Measure and mark a 38 × 48" table top out of ¼" sheet metal.

2. Mark a line 4" in from the edge of each side. Also make the detailed marks for each corner bird’s mouth. This will allow one edge of each corner to bend inside of the other edge of the corner for neat and strong edges.

3. Cut the table top sheet metal with portable shears (see photo).

4. Bend up the triangular bird’s mouth tabs.

5. Use 1/8" steel as a bender. Lay the bender 4" in from the edges and then grasp the sheet metal and bender with clamps. Bend the sheet up at 90° (see photo, top opposite page). The triangular bird’s mouth tabs fold in to the inside for clean exterior corner edges.

6. Fit the table top over the top frame (see photo, bottom right, opposite page).

Finishing

1. Use a 4½" angle grinder with a 60-grit flapper disk to grind all welds until smooth.

2. Clean the metal, and then prime and paint the base, if desired. *Note: Do not paint the top*.

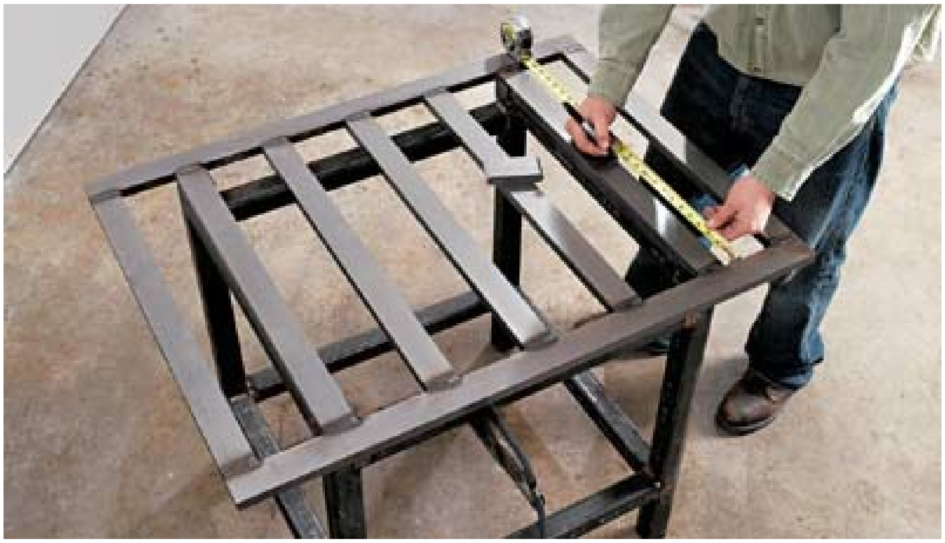

Measure and mark for the top short cross pieces. The table top frame should fit on top of the base with the cross pieces fitting in between the tabs on legs. The short cross pieces then fit in between the cross pieces and top sides (B) for extra strength.

Cut the sheet metal top to 38 x 48 using a portable metal shears.

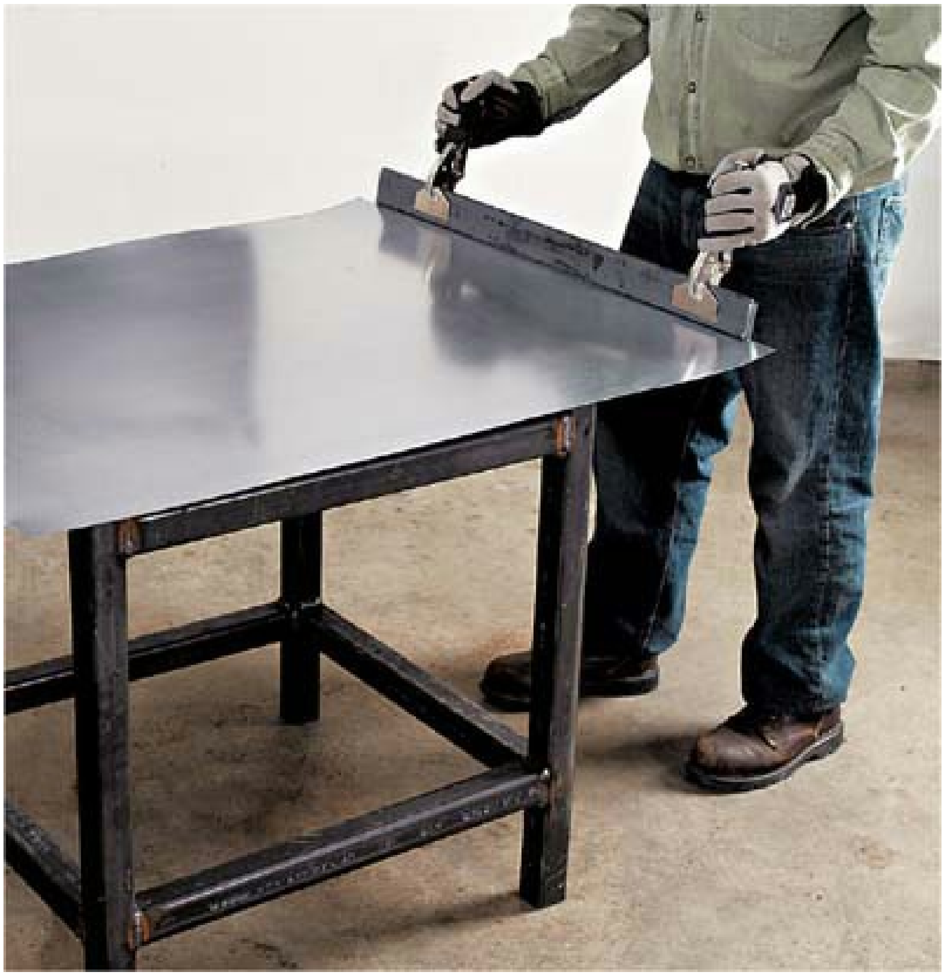

Use a strip of 1/8" -thick steel as a bender. Clamp the bender to the table top sheet metal and bend up to 90°. Shown here: first make a bird’s mouth cut on corners; bend those sides first so that the tabs are on the inside of the bend for a clean exterior edge.

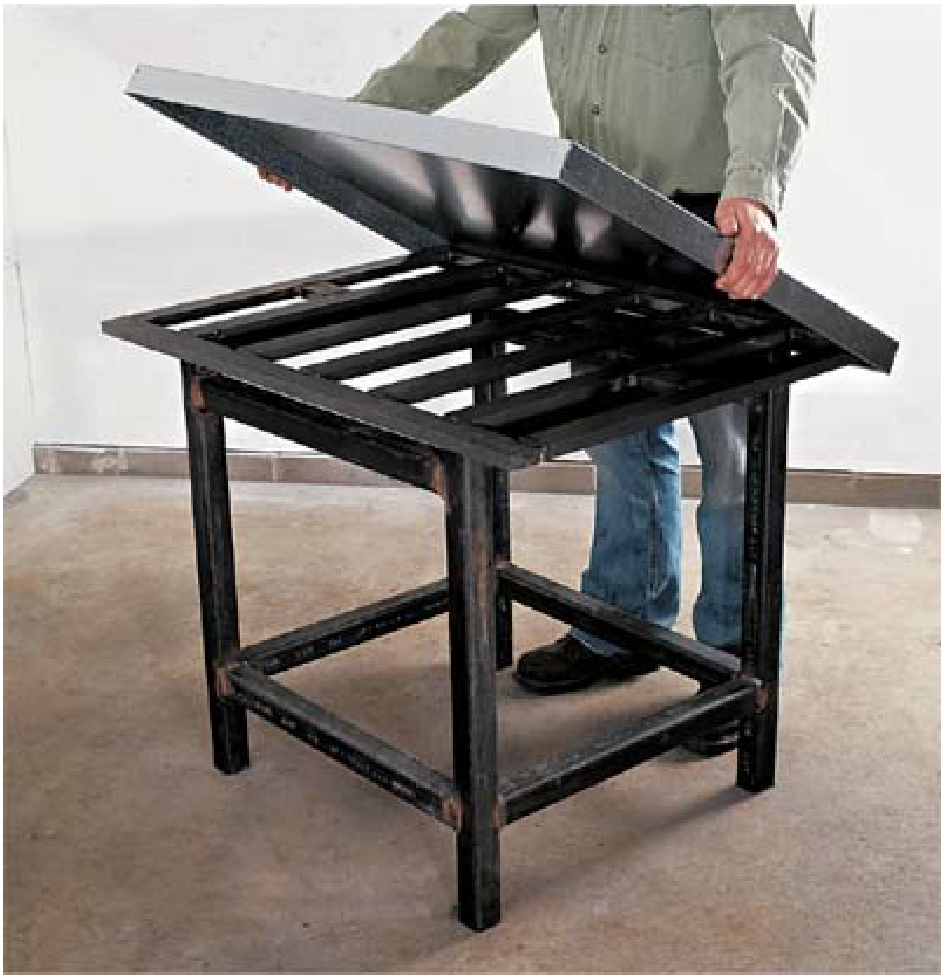

Lower the finished table top onto the table frame.

Saw Stand

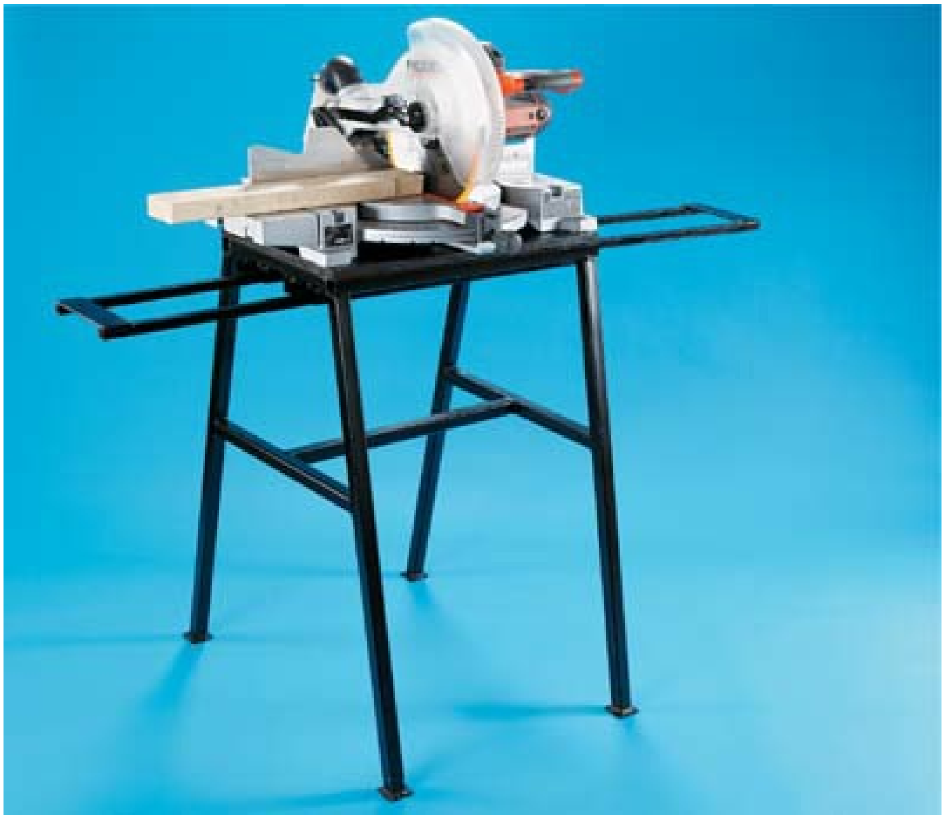

A saw table like this is commonly used in workshops, home shops, and on job sites. Designed for use with a power miter saw or metal cut-off saw, it features outriggers that are adjustable, accommodating a variety of stock lengths. The outriggers slide out on both sides, opening to their maximum length of 24". If both sides need to be extended (for extra long material), the outriggers can both be placed in the front and extend out of the same outrigger tubes. And this will be necessary because the back outriggers sit behind the saw—good for storage when the outriggers need to slide all the way in, flush with the stand sides (for moving and storage). You can add wood blocks on the ends of the outriggers to keep the object being cut level with the saw table.

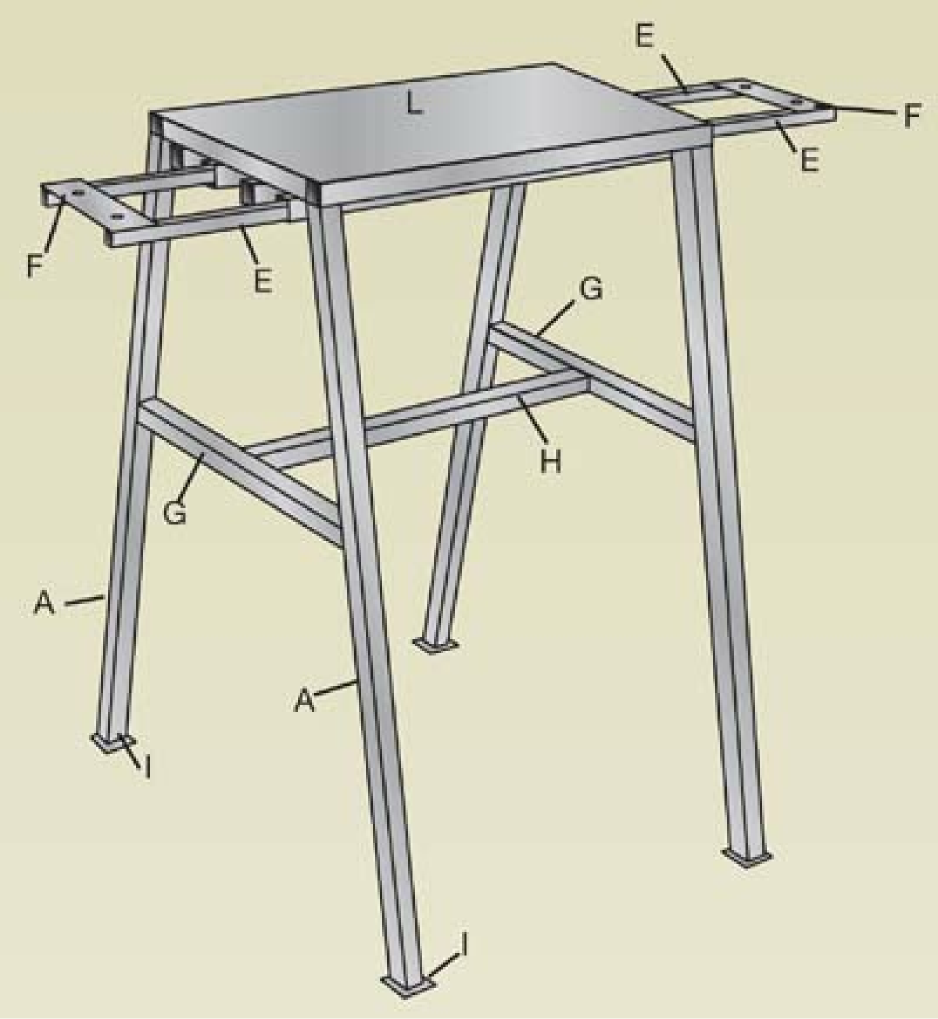
E
L
F
E
F
E
G
H
G
A
A
I
I

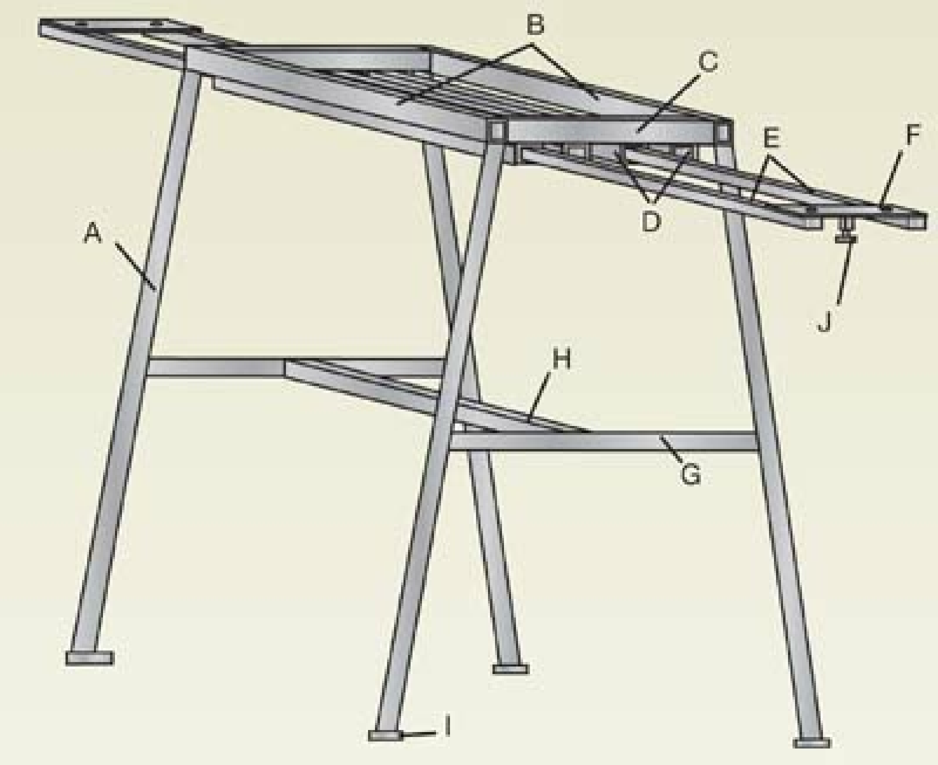
B
C
E
F
D
A
J
H
G
I

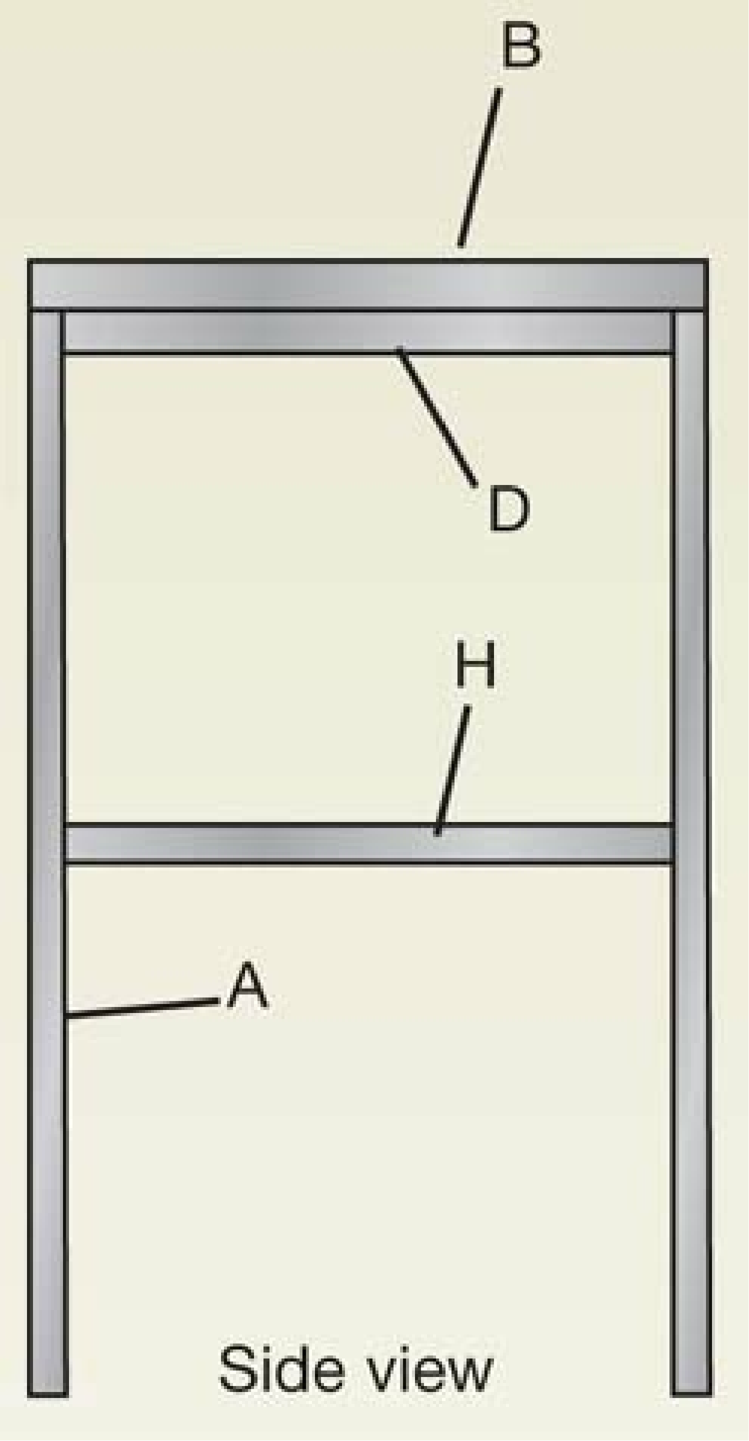
B
D
H
A
Side view

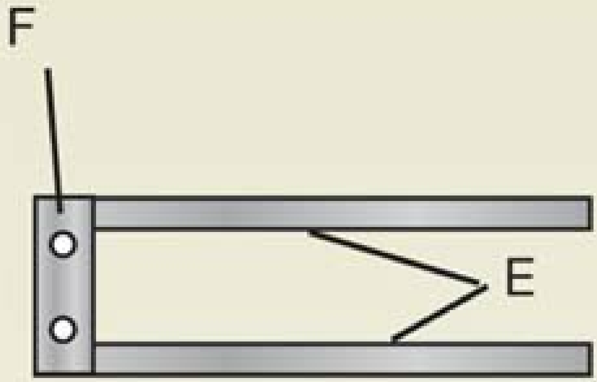

Outrigger detail

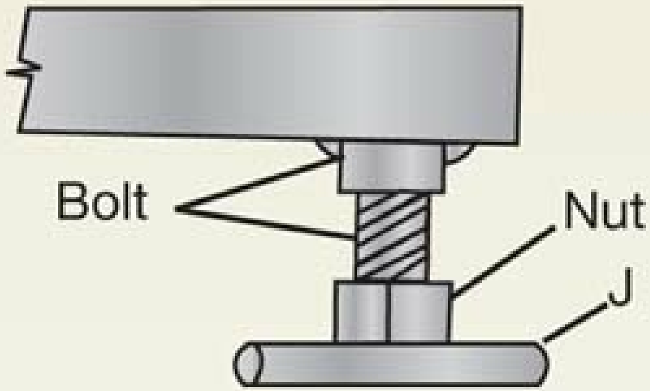

Nut & bolt detail

Materials

- 1¼ x 1¼" x 12 gauge steel (35 feet)
- 8 feet of 1" x 1 x 12-gauge steel tubing
- $^{3}/_{16}$" x 2" flat bar (2 feet)
- ¼" round stock (1 foot)
- Hex bolts and nuts $^{5}/_{16}$- 18 UNC (4)
- ¾ to 1" thick Plywood (16 x 24")

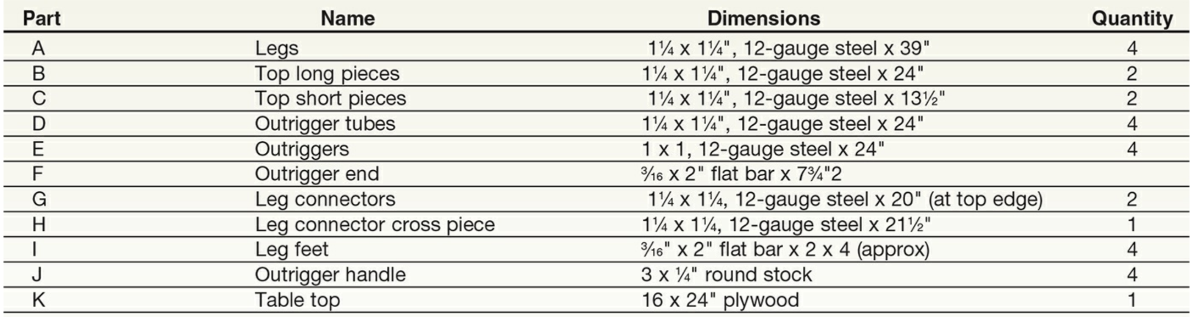

Part	Name	Dimensions	Quantity
A	Legs	1¼ x 1¼", 12-gauge steel x 39"	4
B	Top long pieces	1¼ x 1¼", 12-gauge steel x 24"	2
C	Top short pieces	1¼ x 1¼", 12-gauge steel x 13½"	2
D	Outrigger tubes	1¼ x 1¼", 12-gauge steel x 24"	4
E	Outriggers	1 x 1, 12-gauge steel x 24"	4
F	Outrigger end	3/16 x 2" flat bar x 7¾"2	
G	Leg connectors	1¼ x 1¼, 12-gauge steel x 20" (at top edge)	2
H	Leg connector cross piece	1¼ x 1¼, 12-gauge steel x 21½"	1
I	Leg feet	3/16" x 2" flat bar x 2 x 4 (approx)	4
J	Outrigger handle	3 x ¼" round stock	4
K	Table top	16 x 24" plywood	1

Building the Top Frame

1. Cut the top long pieces (B) to 24".

2. Cut the top short pieces (C) to 13½".

3. Align the top long pieces (B) and top short pieces (C) at 90°. Use a corner magnet to hold the pieces together squarely.

4. Check for square again. Use a magnetic clamp or simply hold the parts in place as you tack them together. Tack all of the long and short tops together to create the top frame.

5. Check for square again, and then weld the four pieces together.

Note: Recommended methods for all welds: SMAW $^{3}/_{32}$*" E 6013 electrode; GMAW short circuit transfer; GTAW; ER 70S-6 filler metal, or MIG (as shown here).*

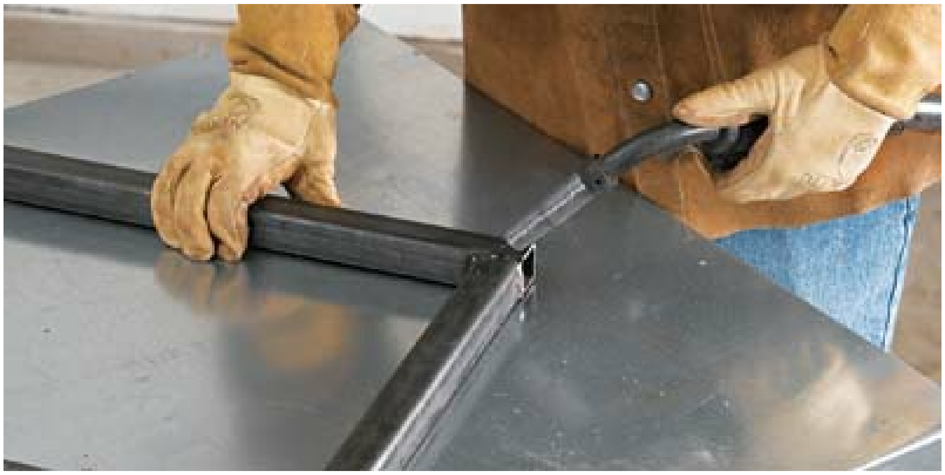

Tack together the top long (B) and short (C) parts. Shown here: MIG welder.

Building the Sides

1. Cut the legs out at 39" with parallel 10° ends (see page 103).

2. Align the 10° cut end of a leg (A) 90° to a top long piece(B).

3. Check the pieces for square, and then hold the legs in place 90° against the top long piece. Tack the leg in place.

4. Repeat steps 2 and 3 for the other leg on the same side.

5. Repeat steps 2 through 4 for the other side.

6. Cut the leg connectors (G) to length at 20" with a 10° angle on each end (from top edge of the angle to top edge of opposite angle it should measure 20").

7. Hold a leg connector (G) in between two legs. Check for proper alignment and then tack in place.

8. Check for square and then weld the connectors in place.

9. Repeat the above steps for the leg connector (G) on the other side.

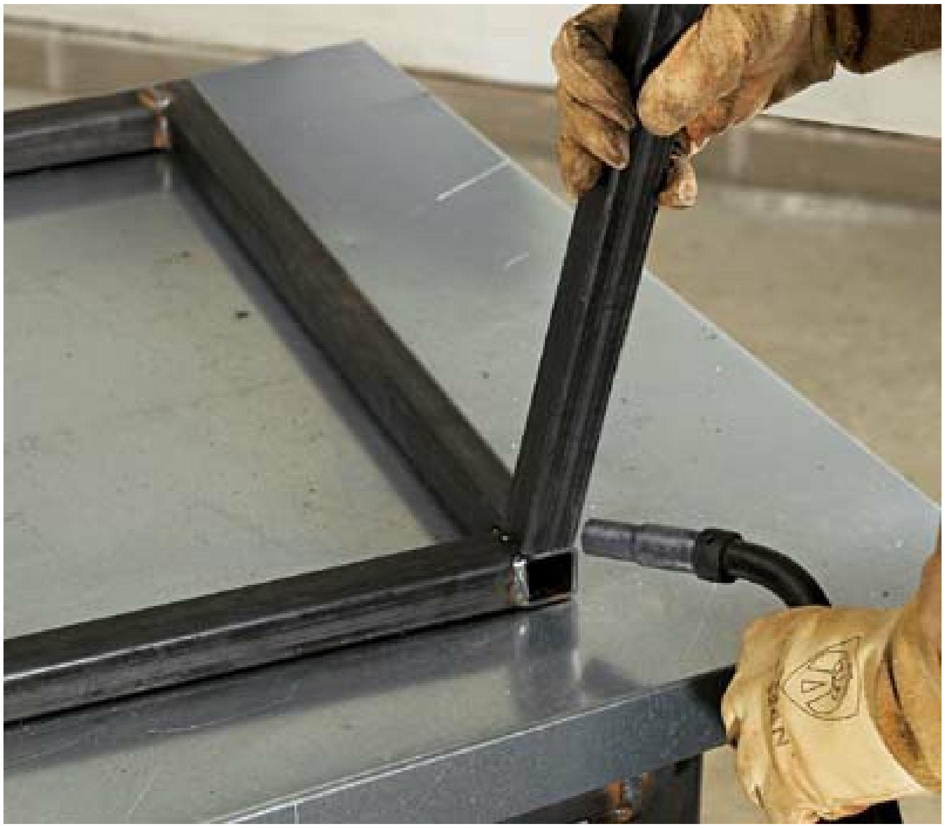

Tack the ends of the legs (A) cut to 10° onto the top frame.

10. Center the leg connector cross piece (H) in between each leg connector (G). Align at 90°, and then use a magnetic clamp to hold in place. Check for square, and then tack the cross piece to the two leg connectors.

11. Check for square and then weld in place.

12. Cut all four feet (I), each at 2".

13. Flip the table upside down on the floor, and then align the feet under the table legs. Hold them in place with pliers and then tack in place.

14. Flip the entire table right-side up to check for level. Adjust as necessary, breaking the feet off and cutting legs for level, if necessary. Shims could also be added.

15. Weld the feet to the legs.

Tack the horizontal leg connectors (G) and then center cross piece (H) that spans between the two leg connectors.

Building the Outriggers

Note: 12-gauge square tubing will always slide, one tube inside the other, when they are ¼ difference in size. More precisely, in order for the outriggers to slide into the outrigger tubes, 12-gauge tubing must be used. 12-gauge tubing has a wall thickness of .110 to .115", so tubes will slide into each other in ¼" increments. For a tighter fit,11-gauge seamless tubing can also be used, but seamless tubing is more expensive.

1. Cut outrigger tubes (D) from 1¼ × 1¼ ", 12-gauge steel to length at 24".

2. Cut the outriggers (E) from 1 × 1, 12-gauge steel to length at 24". Drill a single 3/8" hole on one end of each outrigger. This hole will eventually allow the screw/bolt on the outrigger tube to screw up into the extended outrigger, securely holding it in place when fully extended.

3. Cut the outrigger ends (F) to size at 7¾". Drill two 3/8" holes in them, 2" in from each end.

4. Align an outrigger end (F) 90° to outriggers 1 and 3 (E) at the opposite end of the 3/8" hole you drilled in step 2.

5. Align the other outrigger end 90° to outriggers 2 and 4 (also at the opposite end of the 3/8" hole you drilled in step 2).

6. Use a carpenter’s square to ensure proper alignment, and then clamp in place. Tack an outrigger end to outriggers 1 and 3, then repeat for outriggers 2 and 4. Check for square, and then weld the outrigger ends in place.

7. Drill a 3/8" hole on one end of each outrigger tube (D) underside.

8. Hold the outrigger handle (K) in place over bolt end (J), and then weld in place.

Flip the entire stand over and weld the leg feet (I) to the bottom side of the legs (A). If desired, casters could be added to the bottom for increased mobility. Be sure to pick casters that can lock.

Fastening Outriggers to Frame

1. Center a nut (J) over each of the outrigger tube (D) holes from step 7 (see NUT & BOLT DETAIL). Use pliers to hold the nut in place, and then weld. Make sure to alternate: nuts should be on the same end of 1 and 3 outrigger end (F); and 2 and 4 outrigger end (F). The nut is welded onto the tubing after drilling a hole in it. The bolt that screws in is used as a locking mechanism.

2. Clamp outrigger tubes (D) to the bottom side of the top short pieces (C) with 2" spacing. Check alignment with a ruler.

3. Insert the outriggers (E) into the outrigger tubes (D) so that 1 and 3 extend to one side and 2 and 4 extend to the opposite side. Check for ease of sliding.

4. Once the slides are positioned so that they can be pulled out easily, remove them and tack the outrigger tubes onto the top short pieces (C).

5. Double check to see that outriggers slide in and out with ease. If they do, completely weld the outrigger tubes in place. Weld only two sides of the connections, where tubes connect together. No need to weld all around it.

Adding the Tops

1. Cut a 16 x 24" table top (L) out of ¾- to 1"-thick plywood.

2. Predrill $^{5}/_{16}$" holes for the plywood table top (L) into the frame (B, C). Screw or bolt the table top in place.

Option: Cut 2 x 2" wood blocks for the outrigger tops (M). These sit directly above the holes you previously drilled in the outrigger ends (F). Predrill holes in the wood blocks (see below), and then screw the blocks onto the outrigger ends (F). These blocks keep your cutting object level as it spans out on the outriggers.

Finishing

1. Cap all ends of exposed tubing.

2. Prime and paint the saw table.

3. Prime and paint the table frame.

4. Completely remove outriggers and prime them. Once dry, they should be painted or powder-coated for extra durability. For optimal rust protection, first roughen the surfaces by sanding and apply a rust-resistant metal primer before painting.

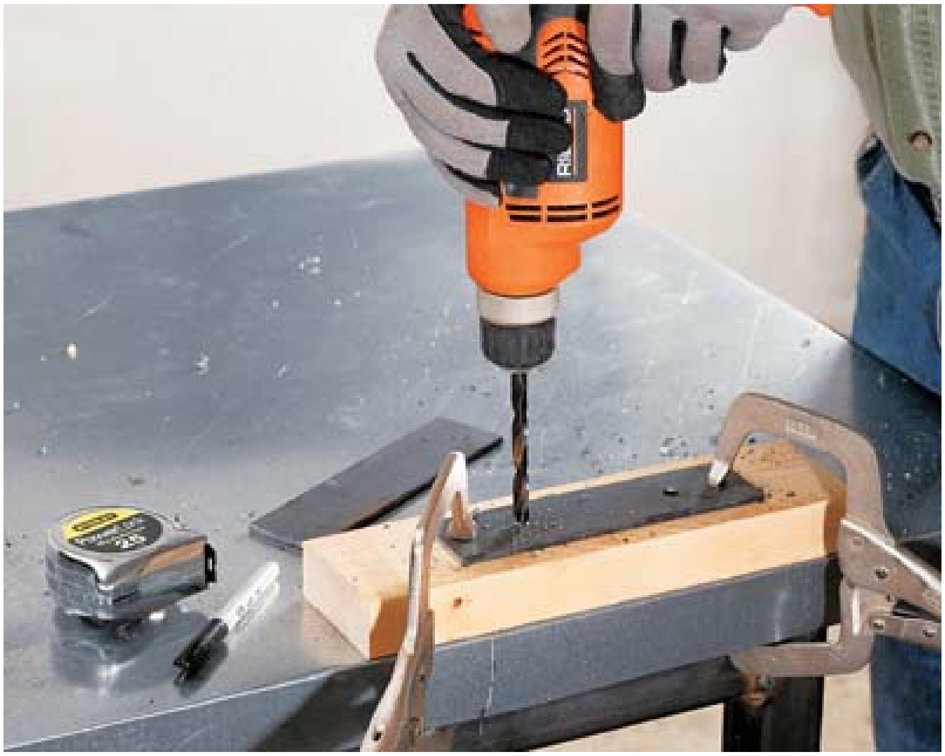

Drill 3/8" holes in the outrigger ends (F). Use a scrap piece of wood under the flat bar, and clamp the bar and wood to a table.

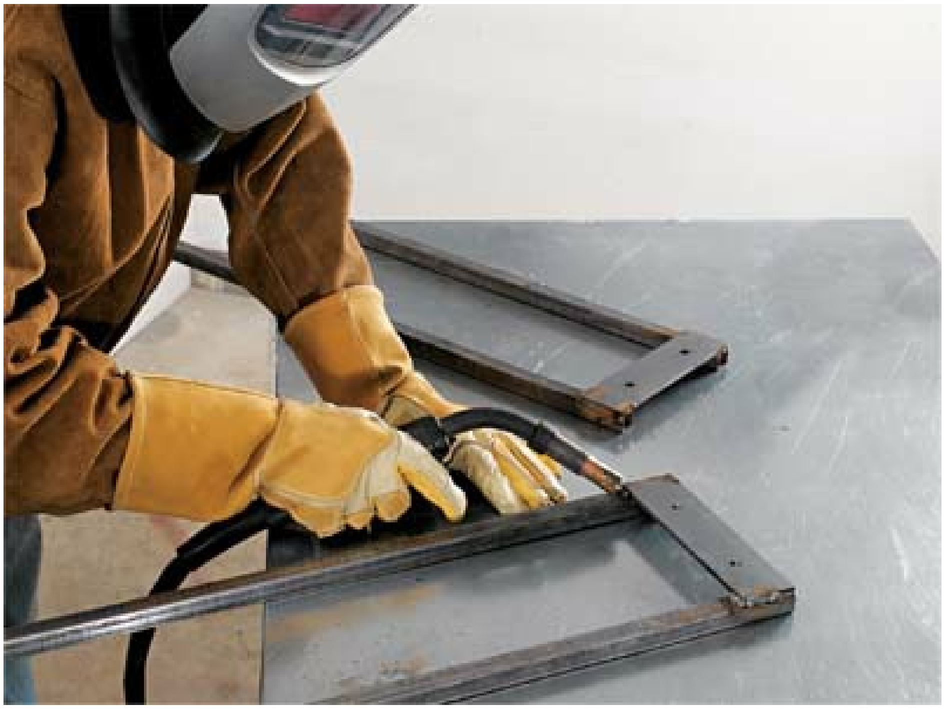

Tack weld the outrigger ends (F) to the outriggers (E). There are two complete pullout outriggers for both sides of the stand.

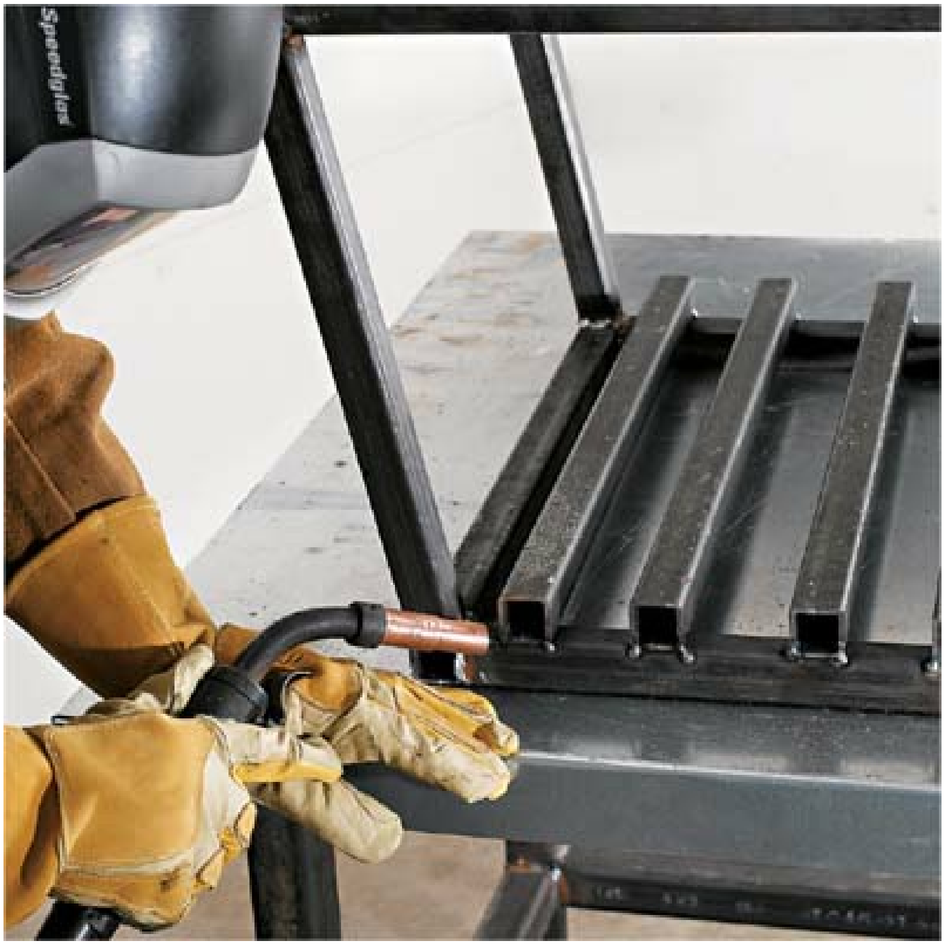

Tack weld the outrigger tubes (D) to the top frame underside.

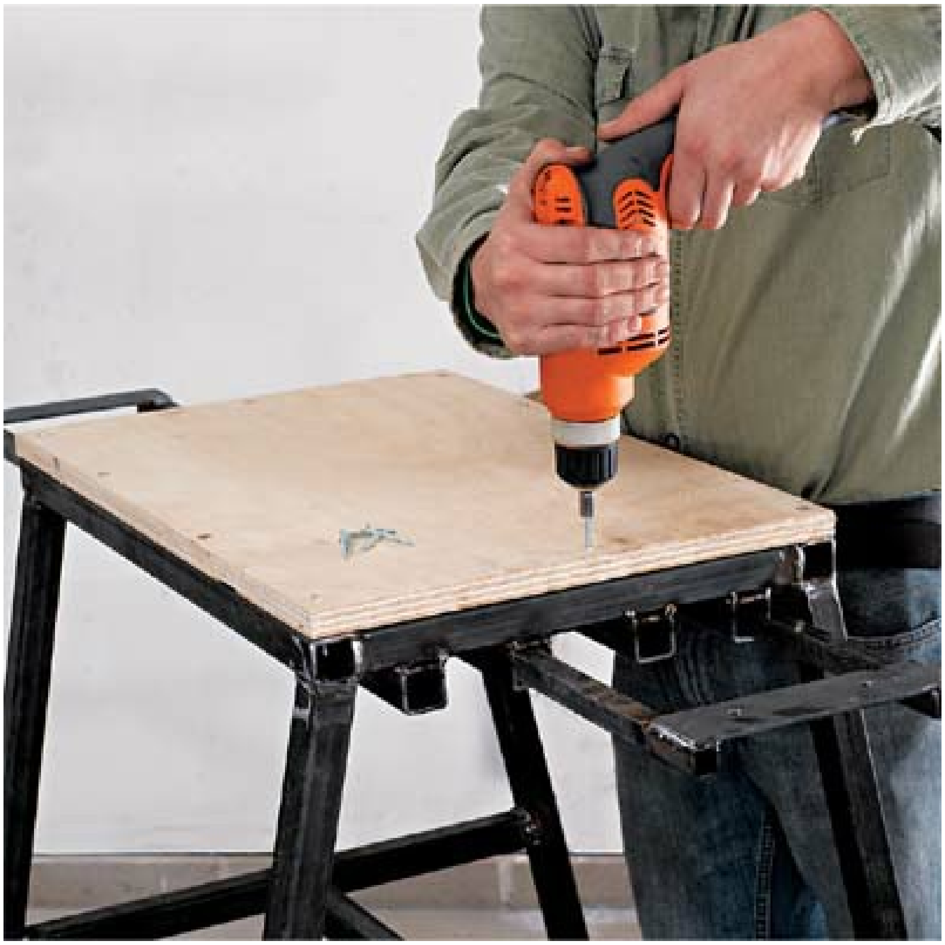

Fasten the plywood top to the frame.

Home Décor

IN THIS CHAPTER:

Coffee Table
Wine Rack
Swivel Mirror
Coat Hanger
Room Divider
Iron Cross
Nesting Tables
Swing-away Coat Hooks
Nightstand
Fireplace Candelabrum

Magazine Rack
Angled Metal Shelves
Wall-mounted Shelf
Étagère
Corner Étagère

Coffee Table

A nice slab of stone or a cross section of a tree makes an appealing coffee table, but how do you support it simply and easily? A table base welded from 1" angle iron and ½" square bar can support the weight of stone or wood without being cumbersome or bulky. This design is for a triangular base to suit the piece of marble being used. Creating a layout with tape allows you to experiment with different sizes and locations for your triangle. If you are using fence pickets, as we have here, make sure the length of the longest side of the triangle does not exceed the length of the picket. Pickets are generally 35 to 39" long.

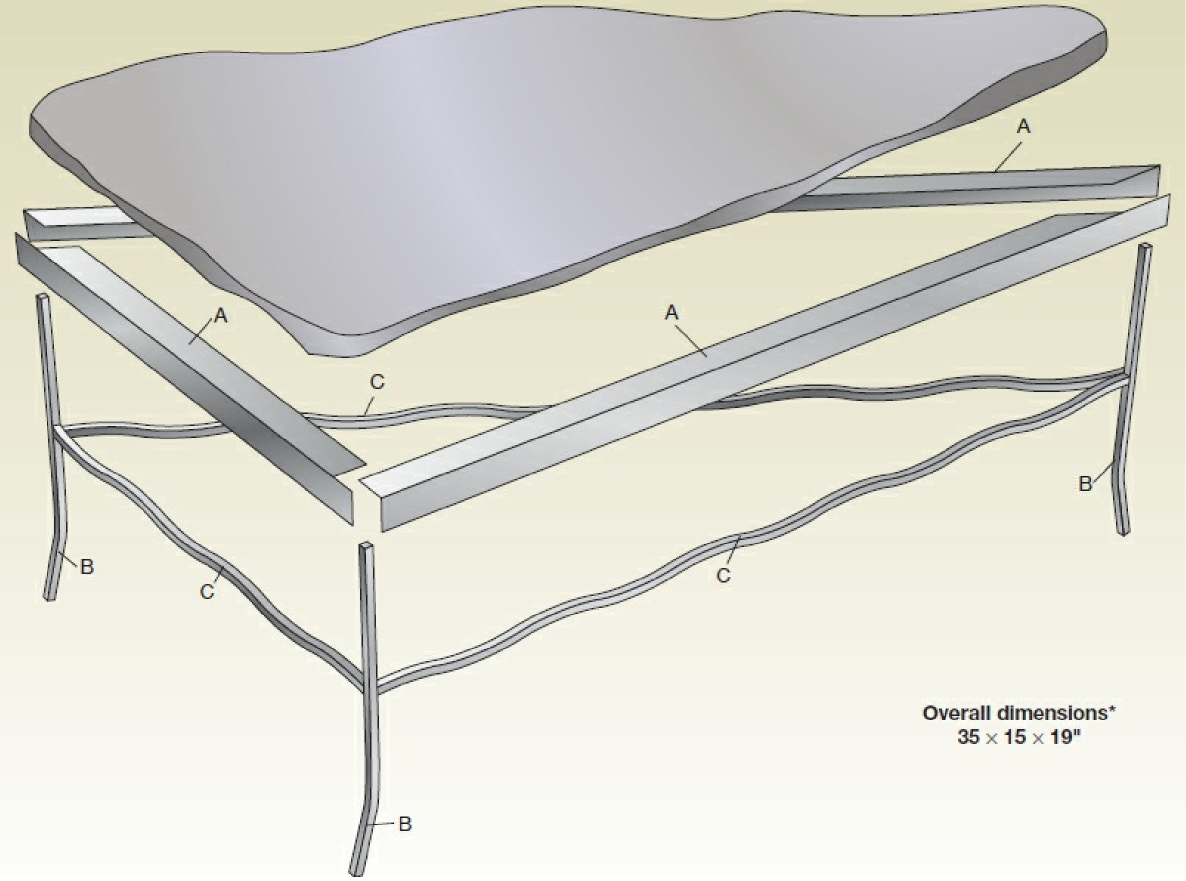

Overall dimensions*
35 × 15 × 19"

Materials

- Masking tape
- Table top
- 1/8" × 1 × 1" angle iron

(amount will vary)

- ½" square bar (5 feet)
- ½" wavy bar fence pickets (3)
- Construction adhesive,

silicone glue, or screws

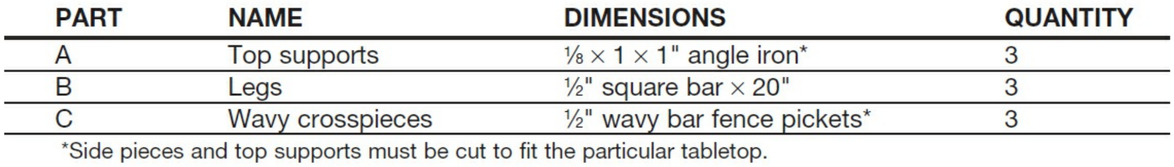

PART	NAME	DIMENSIONS	QUANTITY
A	Top supports	⅛ × 1 × 1" angle iron*	3
B	Legs	½" square bar × 20"	3
C	Wavy crosspieces	½" wavy bar fence pickets*	3

*Side pieces and top supports must be cut to fit the particular tabletop.

How to Build a Coffee Table

Create the Triangle

1. Turn the table top upside down and use masking tape to lay out a triangle. The triangle should be at least 3" in from the edges of the table top material, but not so far in that the table will tip easily.

2. Measure each side of the triangle. Cut the top support pieces (A) to match these three measurements. Miter both ends of the longest piece.

3. Place the longest support piece on top of the second longest support. Set both pieces on the tape layout on the table top. Mark the angle and notch on the second support where it intersects the first support (see photo) to allow these pieces to butt together. Cut out the notch and angle on the support.

4. Place the two supports back together on the triangle layout. Set the third support on the layout under the first two support pieces. Mark the notches and angles at the intersections. Cut the third support piece. (You may need to make an additional angle cut on the second long piece, depending on the triangle.)

5. On a work surface, arrange the support pieces to form the triangle. Weld the outside corners. Turn the assembly over and weld the top butt joints.

6. Use an angle grinder to smooth the top joints so the table top rests on a flat surface.

Install the Legs

1. Cut the legs (B) to size.

2. Mark each leg 4" from one end. Place a leg in a bench vise, lining up the mark with the edge of the vise jaws and bend the leg end 15°. Repeat this process for the other two legs.

3. Turn the top assembly upside down on your work surface. Place a leg in a corner with the bend of the leg pointing outward in the same line as the point of the triangle.

4. Check to make sure the leg is perpendicular to the top assembly, then tack weld in place. Tack weld the other two legs in the remaining corners the same way.

Attach the Side Pieces

1. Measure the distance between the legs. Cut one wavy crosspiece (C) to fit each of the three sides. Bevel the ends to fit the angled legs.

2. Clamp a crosspiece against two legs 7" from the top assembly. Check for square and tack weld. Install the remaining crosspieces the same way.

3. Turn the assembly right-side up and check for square and level. Make adjustments if necessary. Complete all the welds.

Complete the Table

1. Grind down the welds, if desired. (You may also want to grind down the feet of the table so they make flat contact with the floor.) Clean or wire brush the assembly. Apply your choice of finish.

2. Drill holes in the angle iron and attach a wooden table top with screws, or use silicone or construction adhesive to attach a stone top.

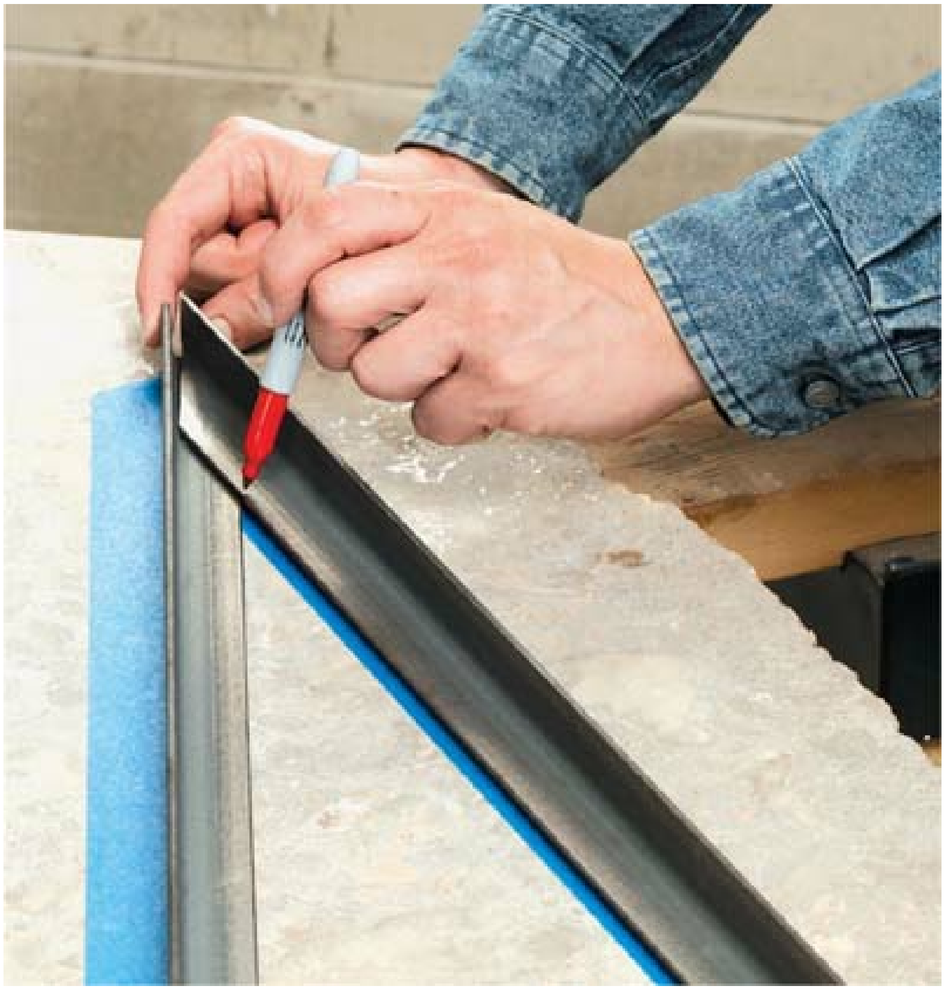

Place the mitered long support piece on top of the second longest support piece. Mark the angle and notch needed to fit the pieces together.

Wine Rack

This wine rack is designed so you can install a wall or floor tile as the top surface. Measure the tile and cut the top pieces to fit. Butcher block or marble would also make a nice-looking top. The decorative frieze stamping and squashed ball are available from Architectural Iron Designs—see Resources on page 220.

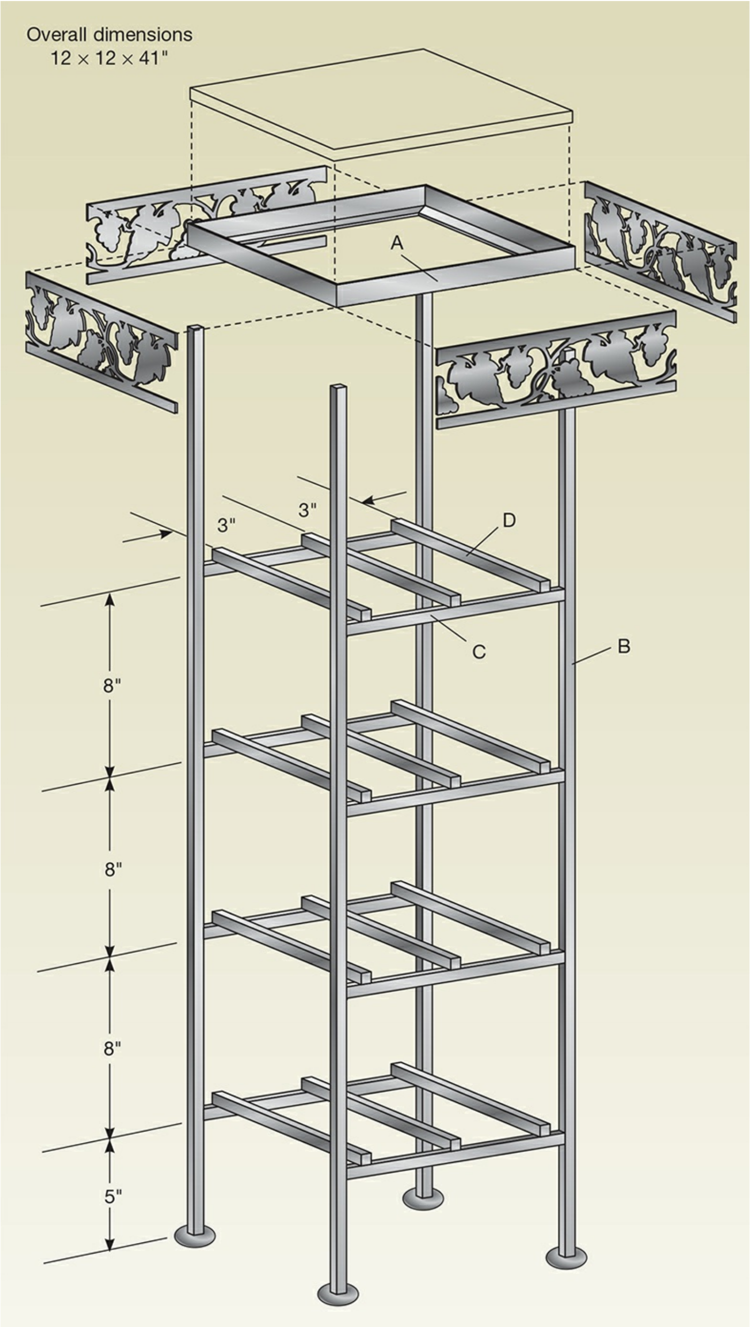
Overall dimensions
12 × 12 × 41"
A
3"
3"
D
C
B
8"
8"
8"
5"

Materials

- 1⁄8 × 1 × 1" angle iron (4½ feet)
- ½" square bar (33 feet)
- Decorative frieze (4½ feet)
- 2" squashed balls (4)
- 12" floor or wall tile (1)

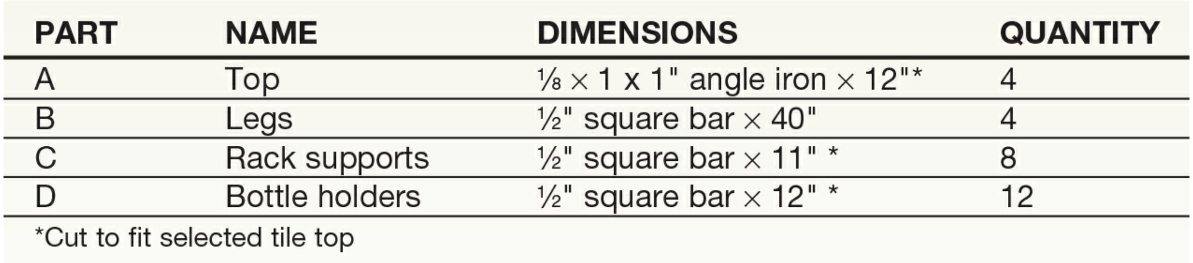

PART	NAME	DIMENSIONS	QUANTITY
A	Top	⅛ × 1 x 1" angle iron × 12"*	4
B	Legs	½" square bar × 40"	4
C	Rack supports	½" square bar × 11" *	8
D	Bottle holders	½" square bar × 12" *	12

*Cut to fit selected tile top

How to Make a Wine Rack

Assemble the Top

1. Cut the top pieces (A) to size to fit your top surface, mitering the corners at 45°.

2. Clamp two top pieces together with a corner clamp. Use a carpenter’s square to check the pieces for square, and tack weld.

3. Repeat step 2 to assemble the other two top pieces.

4. Clamp the two L-shaped top pieces together, and check for square by measuring across both diagonals. If the measurements are equal, the assembly is square. If it is not square, adjust until both measurements are equal.

5. Tack weld the two corners. Unclamp the pieces and recheck for square. Finish welding the corner joints.

Assemble the Stand

1. Cut the legs (B), rack supports (C), and bottle holders (D) to size.

2. Mark the legs at 5", 13", 21", and 29" from one end. Mark the rack supports at 2½", 5½" and 8½" from one end.

3. Position four rack supports between two of the legs at the marks. Check the pieces for square and tack weld in place.

4. Repeat step 3 to tack weld the remaining four rack supports to the other two legs.

5. Stand the leg assemblies upright and clamp a bottle holder centered on the 5½" mark on the bottom rack support. Clamp the other end of the bottle holder to the other leg assembly. Check for square and tack weld the bottle holder in place.

6. Repeat step 5 to attach a bottle holder at the 5½" marks on the other three rack supports.

7. Make sure the legs assemblies and bottle holders are square, then add the remaining bottle holders at the 2½" and 8½" marks. Finish all welds.

Attach the Top, Legs & Frieze

1. Place the top onto the legs and weld in place.

2. Center a squashed ball at the bottom end of each leg and weld in place.

3. Bend or cut the decorative frieze to fit around the top of the legs and top. Braze, braze weld, or tack weld the frieze into place at the top of the legs and slightly below the top.

4. Grind down the welds as needed. Wire brush or sandblast the wine rack, then apply the finish of your choice.

5. Place the tile into the top.

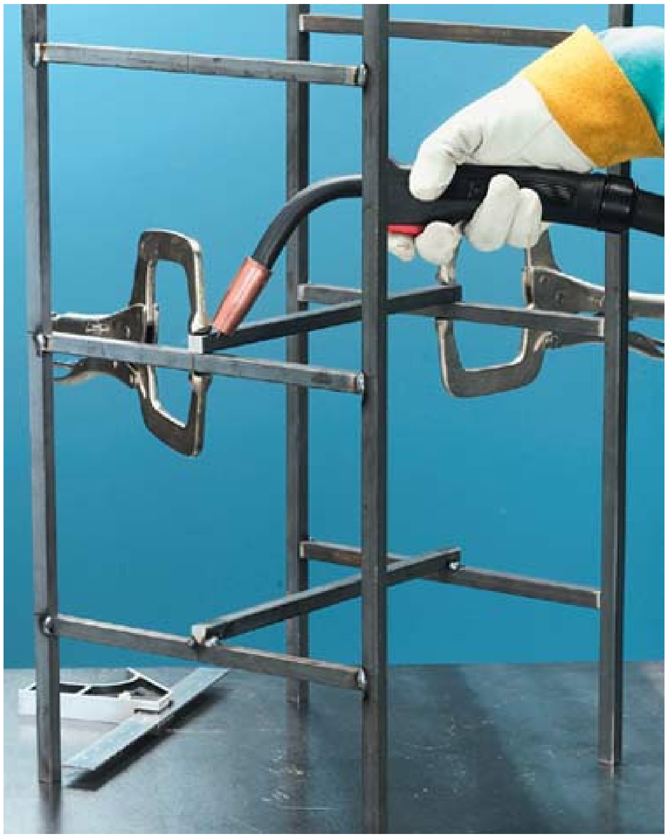

Clamp the center bottle holders to the leg assemblies and tack weld in place.

Swivel Mirror

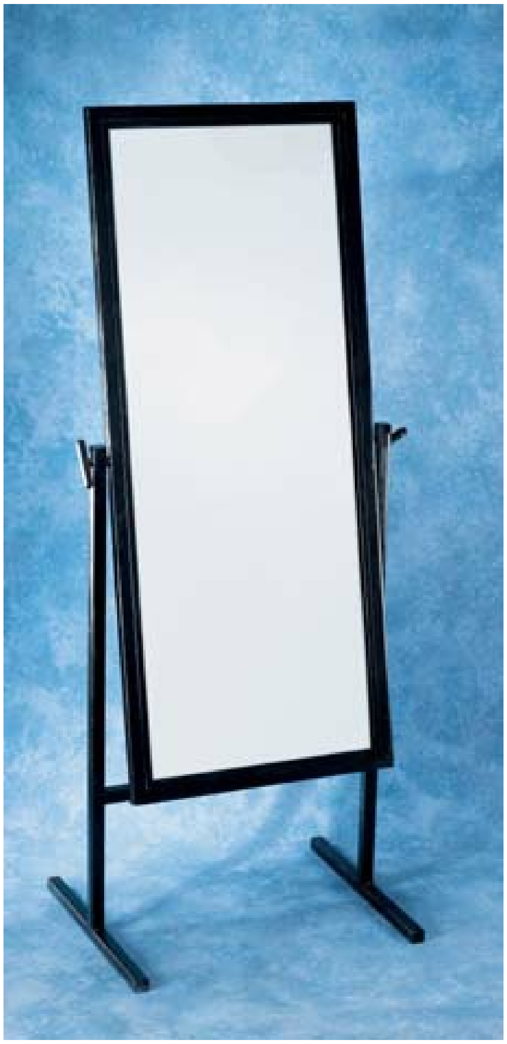

Reflect your great taste with this sturdy modern mirror. Its classic design creates a mirror built to last and will make a stunning first impression when displayed in your front entryway for years to come. The clean lines and black paint give this mirror a modern feel, but this can easily be altered by the finish treatment you choose. The adjustable mirror allows even the most petite guests to swivel it to the perfect height.

As described here, the inside frame is welded to the outside frame from the back using numerous small welds. The joints between the inside frame pieces are welded from the front, and those welds are ground down. The final look is a frame within a frame, rather than a completely smooth single frame. If you prefer a solid frame, you may continuously weld the inside frame to the outside frame from the front and then grind the welds smooth.

Materials

- 16-gauge 1 × 1" square tube (22 feet)
- 1/8 × ½" flat bar (11 feet)
- ½ × ½" square bar (7")
- 1½" × 5/16" hex bolts (2)
- ¼"–20 hex nuts (6)
- Levelers (4, Rockler #32498)
- ½ × 1" flat bar (6")
- Mirror (45 7/8" × 17 7/8")
- ¼" fiberboard (46" × 18")

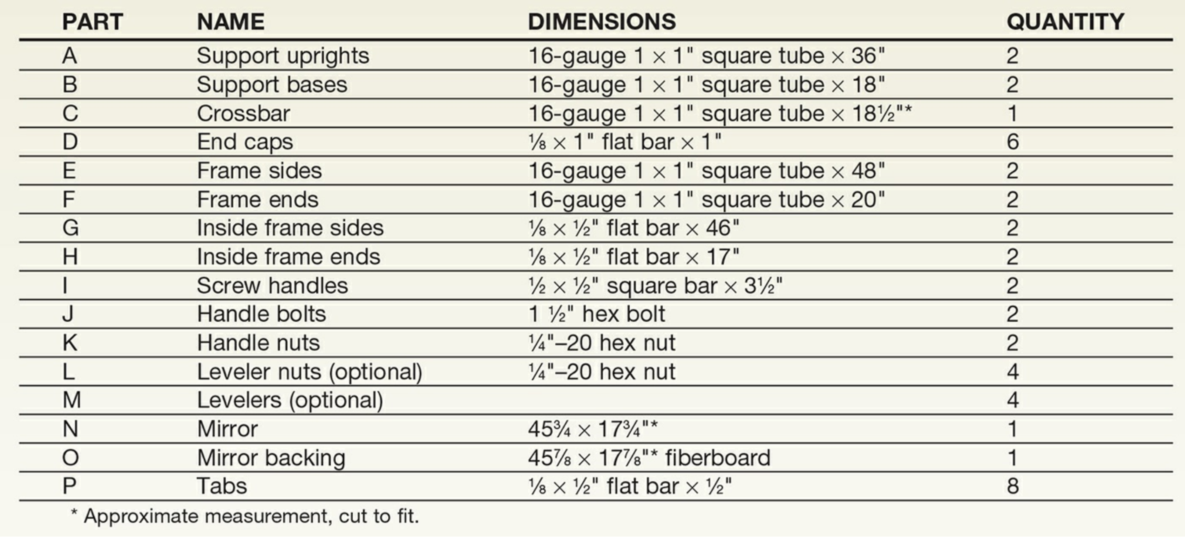

PART	NAME	DIMENSIONS	QUANTITY
A	Support uprights	16-gauge 1 × 1" square tube × 36"	2
B	Support bases	16-gauge 1 × 1" square tube × 18"	2
C	Crossbar	16-gauge 1 × 1" square tube × 18½"*	1
D	End caps	⅛ × 1" flat bar × 1"	6
E	Frame sides	16-gauge 1 × 1" square tube × 48"	2
F	Frame ends	16-gauge 1 × 1" square tube × 20"	2
G	Inside frame sides	⅛ × ½" flat bar × 46"	2
H	Inside frame ends	⅛ × ½" flat bar × 17"	2
I	Screw handles	½ × ½" square bar × 3½"	2
J	Handle bolts	1 ½" hex bolt	2
K	Handle nuts	¼"–20 hex nut	2
L	Leveler nuts (optional)	¼"–20 hex nut	4
M	Levelers (optional)		4
N	Mirror	45¾ × 17¾"*	1
O	Mirror backing	45⅞ × 17⅞"* fiberboard	1
P	Tabs	⅛ × ½" flat bar × ½"	8

* Approximate measurement, cut to fit.

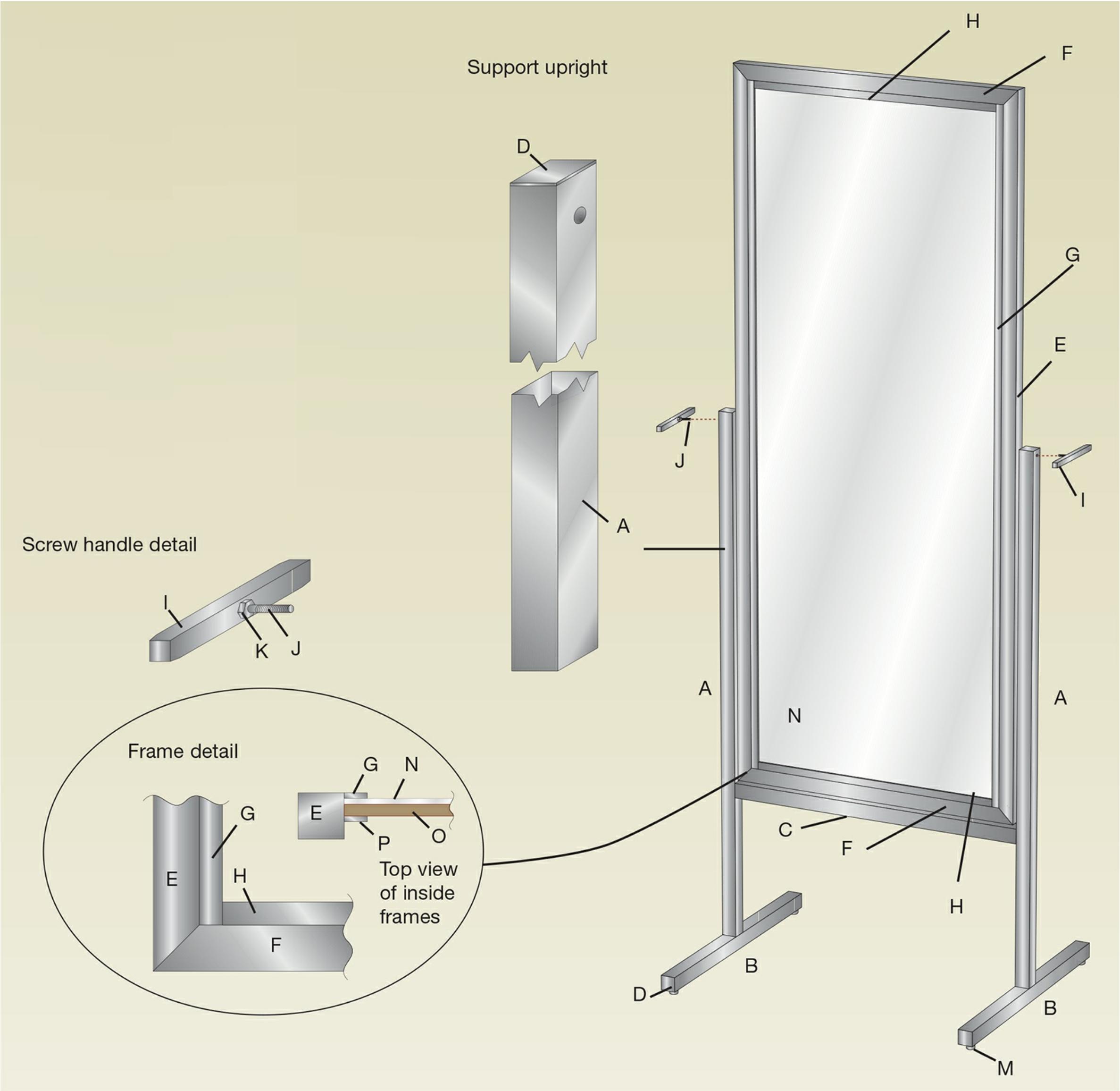
H
F
Support upright
D
G
E
J
I
A
Screw handle detail
I
K
J
A
N
A
Frame detail
G
N
E
G
P
O
E
H
F
Top view
of inside
frames
C
F
H
B
D
B
M

How to Build a Swivel Mirror

Before welding, thoroughly clean all parts with denatured alcohol.

Assemble the Supports

1. Cut the supports upright (A) and support base pieces (B) to length.

2. If using levelers, drill two 11$^{7}/_{64}$" holes in one side of each loose piece, 2" from the ends. Drill through one wall only. Weld ¼" hex nuts over the holes.

3. Center an upright on a base at 90° and check for square. Weld into place and repeat with the second upright and base.

4. At 2" from the top outside of each upright, drill a 1$^{1}/_{32}$" hole. Drill the hole through both walls of the tube.

5. Cut end caps (D) to size. Weld caps to the support upright and base ends. The two supports are in the shape of a T, thus creating the need for six end caps.

6. Grind all welds smooth.

Assemble the Frame

1. Cut the frame sides (E) and frame ends (F) to size. Miter the ends at 45°, or create a joint as shown in the Frame Detail (on page 115).

2. Drill a 1$^{1}/_{32}$" hole at the outside midpoint of each side through one wall for the mirror pivot bolts.

3. Align a frame side and a frame end to make a 90° angle. Use a carpenter’s square to check for square and clamp in place. Tack weld the pieces together to form an L. Repeat with the other frame side and frame end.

4. Align the two Ls to form a rectangle. Use a carpenter’s square to check for square. Clamp into place. Measure the diagonals of the rectangle to check for square. The diagonal measures should be equal; if not, adjust until they are. Tack weld together and recheck for square.

5. Complete all the frame welds. Grind down each weld until flush.

6. Grind the zinc coating off the handle nuts (K). Center the nuts over the holes and weld into place. Protect threads with anti-spatter gel.

Assemble the Inside Frame

1. Before cutting the inside frame sides (G) to length, measure the inside length of the assembled frame. Cut the inside frame sides to fit.

2. Place the assembled frame on a flat surface. Place the two inside frame sides in place against the two frame sides and tack weld into place.

3. Measure between the inside frame sides and cut the inside frame ends (H) to length.

4. Place the inside frame ends in place against the frame ends and tack weld into place.

5. Turn the frame unit over. The inside frame should be flush with the frame. If not, remove pieces and retack, as necessary.

6. Weld the inside frame to the frame from the back with small welds every 3 to 4". Weld the joints between the inside frame members from the front.

7. Grind down all the welds on the front so they are flush.

Tack weld the end caps in place, then weld and grind the welds smooth.

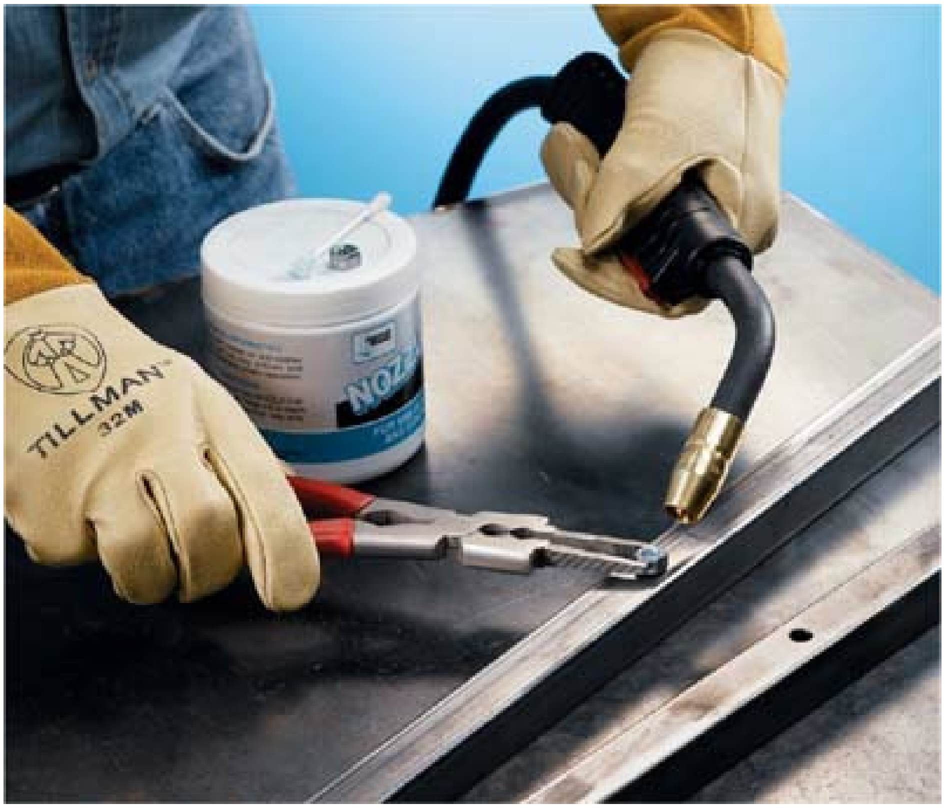
TILLMAN™
32M
NOZZ

Weld the nuts over the holes in the frame sides before assembling the frame. Use anti-spatter gel to protect the nut threads.

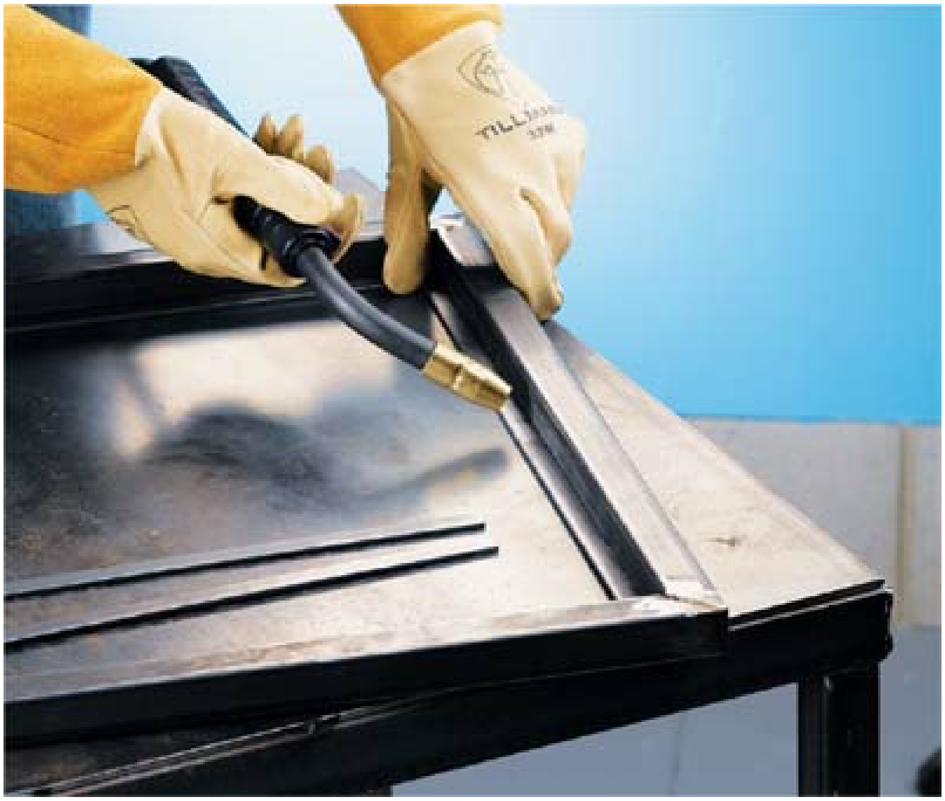

From the back, weld the inside frame to the frame. A small weld every 3 to 4" is sufficient.

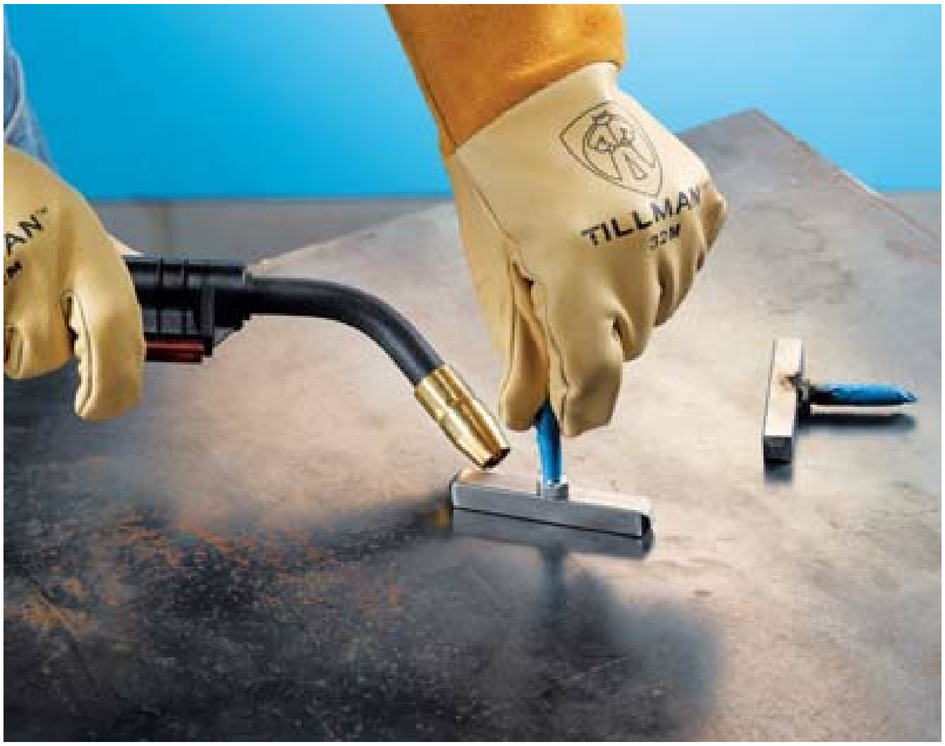
TILLMAN
32M

Weld the bolts to the beveled screw handles. Use masking tape or anti-spatter gel to protect the threads.

Make the Handles & Apply Finish

1. Cut the screw handles (I) to length. Bevel the ends.

2. Grind the zinc coating from the head of the handle bolt (J). Coat the bolt threads with anti-spatter gel or cover with masking tape. Weld the bolts to the center of the bolt handles.

3. Assemble the frame and supports. With the bolts tightened enough to prevent the frame from moving, measure the distance between the support uprights making sure the uprights are parallel during measurement. Cut the support crossbar (C) to this length.

4. Measure the distance between the frame and the uprights (it should be about ¼"). Mark the uprights below the frame at that same distance. The top of the crossbar will align with these marks so the spacing between the frame and uprights and crossbar is uniform.

5. Disassemble the frame and supports. Align the top of the crossbar with the marks on the uprights. (This is easiest with support bases hanging over the edge of the work surface.) Tack weld the crossbar into place. Use a carpenter’s square to check that all angles between the crossbar and support uprights are 90°. Complete the welds and grind them smooth.

6. Grind all welds smooth and grind off any spatter. Wipe the parts down with denatured alcohol to remove grinding dust. Prime and paint. For a natural metal look, use a clear coat. Protect bolt and nut threads with tape during painting to prevent fouling. Don’t forget to paint the handles.

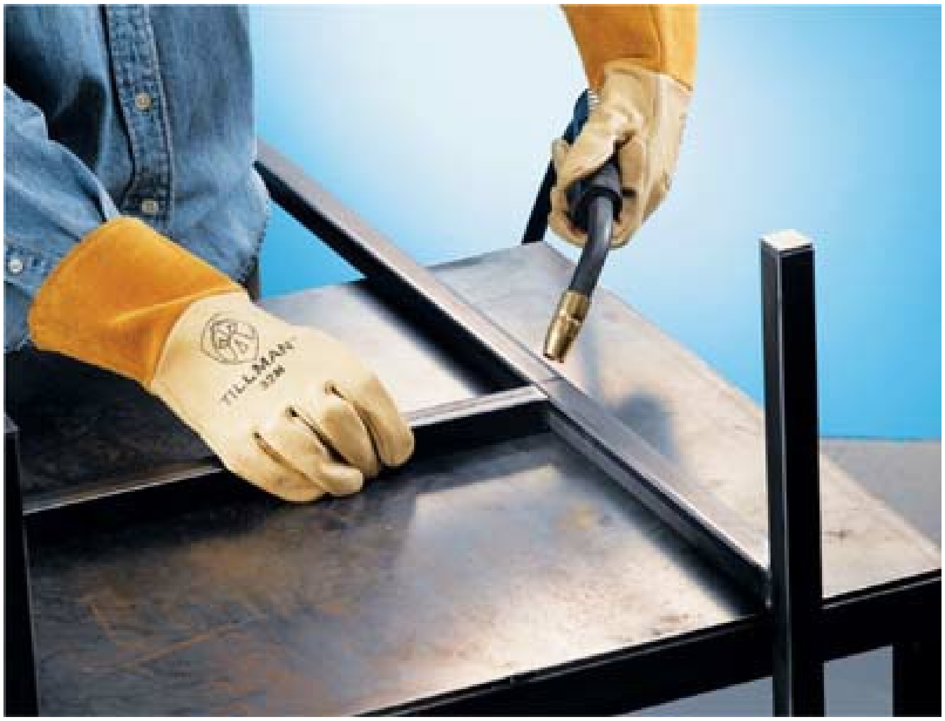

TILLMAN
1350

Assemble the frame and uprights and mark the location for the crossbar. Remove the frame and weld the crossbar in place.

Install the Mirror

1. Measure the inside of the completed frame to determine the exact mirror size to order. The mirror (N) should be ¼" shorter and ¼" narrower than the frame.

2. Cut the mirror backing (O) to size.

3. Apply small dots of silicone adhesive every 12" around the inside frame. Place the mirror into the frame.

4. Apply small dots of silicone adhesive to the back of the mirror. Place the backing against the mirror. Place the mirror assembly face down on a non-scratching surface.

5. Cut the tabs (P) to size and grind one end to a semi-circle.

6. Space the tabs at 6", 23", and 40" along the frame sides. Sand a small amount of paint off the frame at these points. Center a tab along the frame top and bottom.

7. Firmly push the tabs against the mirror backing and frame sides. Weld into place.

8. Paint the tabs and the smooth side of the mirror backing to match the frame.

9. Install the levelers, if desired, and assemble the mirror. Position the mirror within the uprights and thread the handle bolts through the support uprights and into the frame sides.

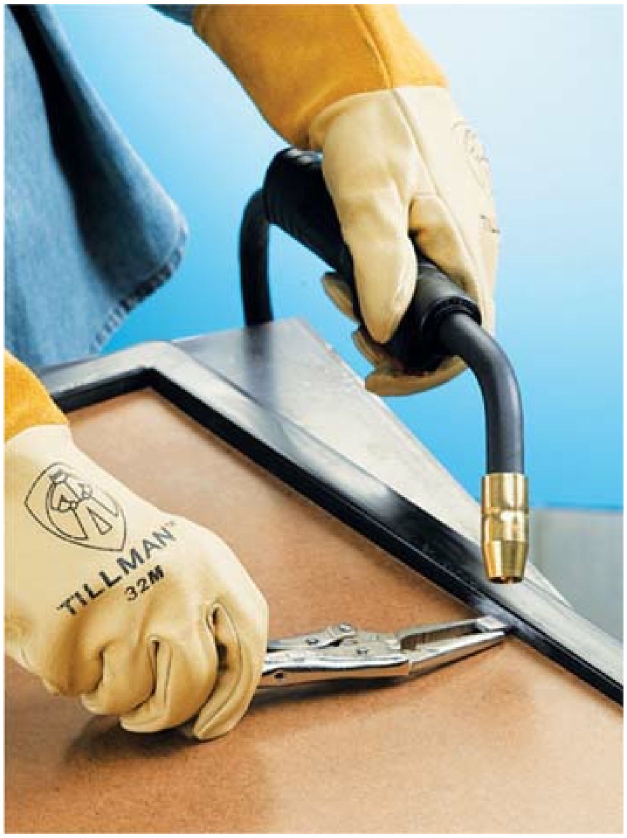
TILLMAN
32M

Install the mirror and back in the frame. Sand the finish from the frame and weld the tabs per Step 6.

Coat Hanger

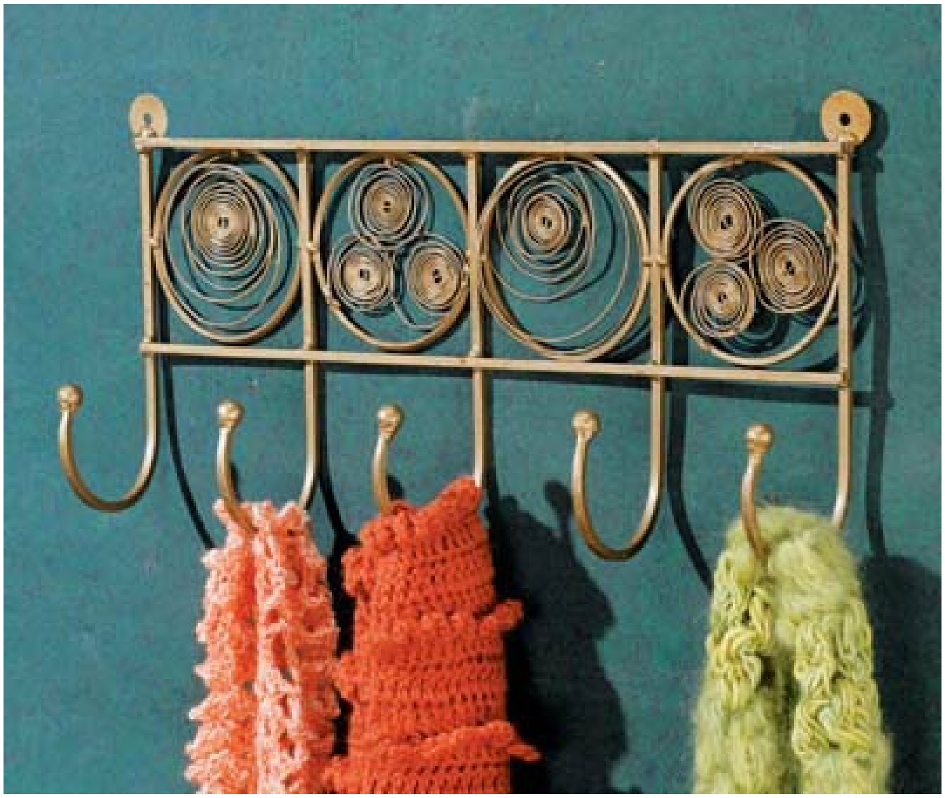

This chic coat hanger has a modern artistic flair due to the unique circular patterns. But the basic design has unlimited potential. You can customize this rack for any part of your life—the front entryway, the garage, the gardening shed, and the cabin may all have different styles. Simply replace the spirals with a punched tin pattern, decorative frieze, or dragonfly, moose, or bear stampings. The rack is sized to mount on studs spaced 16 inches apart. If you like the spirals but want more contrast, use copper or brass 14-gauge wire and leave the entire piece unpainted—but be sure to braze the spirals into place, instead of welding them.

Materials

- 3/16" square bar (8 feet)
- 1/8" square rod (8 feet)
- ½" balls (5, decorative iron #JG8851)
- ¾" washers (2)
- 18-gauge annealed stove- pipe wire (25 feet)

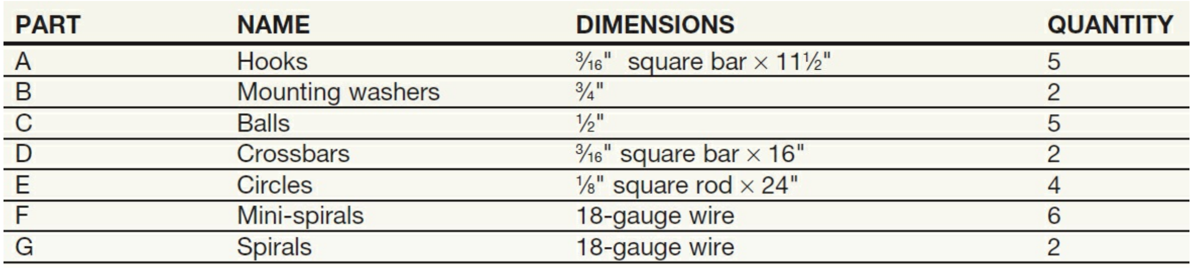

PART	NAME	DIMENSIONS	QUANTITY
A	Hooks	3/16" square bar × 11½"	5
B	Mounting washers	¾"	2
C	Balls	½"	5
D	Crossbars	3/16" square bar × 16"	2
E	Circles	1/8" square rod × 24"	4
F	Mini-spirals	18-gauge wire	6
G	Spirals	18-gauge wire	2

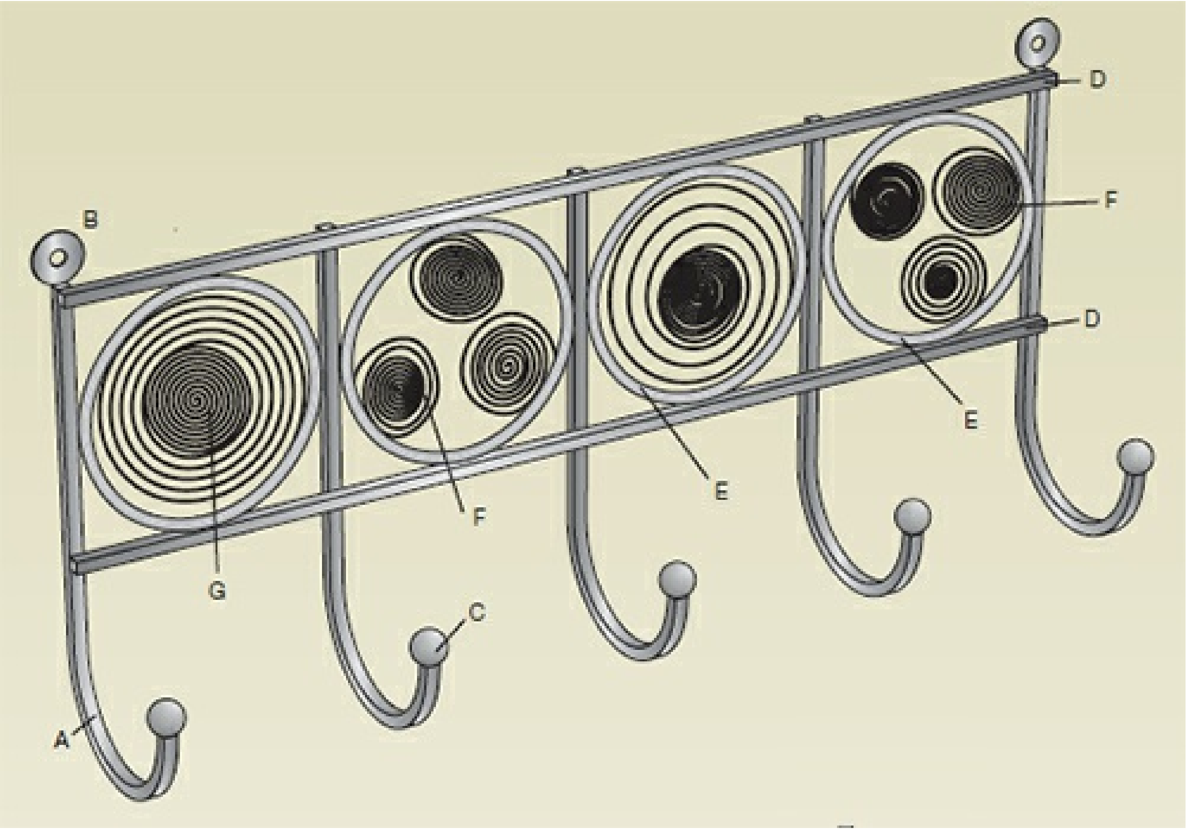
D
B
F
D
E
E
G
F
C
A

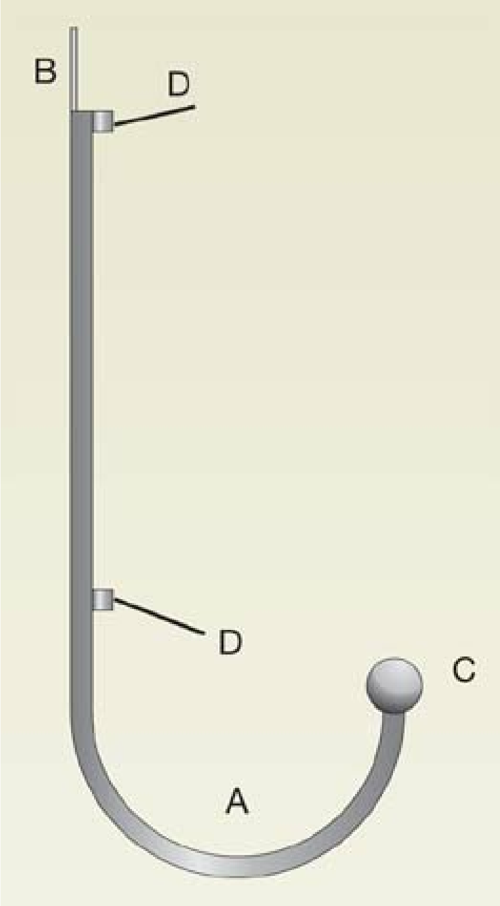
B
D
D
C
A

How to Build a Coat Hanger

Before welding, thoroughly clean all parts with denatured alcohol.

Make the Hooks & Frame

1. Cut the hooks (A) to length.

2. Clamp one end of each hook to a 2"-diameter pipe with a vise-style pliers. Bend the bar a half turn around the pipe to create each hook.

3. Weld each of the two mounting washers (B) to the top of two separate hooks. Weld a ball (C) to the end of all hooks.

4. Cut the crossbars (D) to length. Mark the crossbars at 4", 8", and 12".

5. Align the two hooks that have the mounting washers flush with the crossbar ends. Tack weld in place. Place the second crossbar 4" from the top crossbar and tack weld in place. Check that the angle between the crossbars and hooks is 90° and finish the welds.

6. Center a hook under the two crossbars at the 12" mark lining the top of the hook flush with the top crossbar. Check that the angle between the hook and crossbars is 90° and weld into place.

7. Weld the remaining two hooks to the crossbars at 4" and 8".

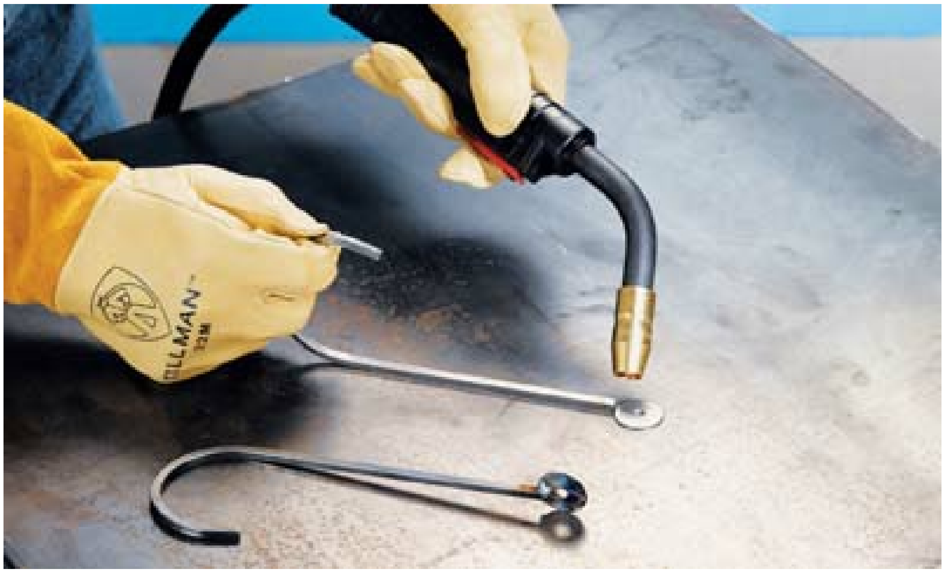
TILLMAN
22M

Weld the mounting washers to two of the hooks.

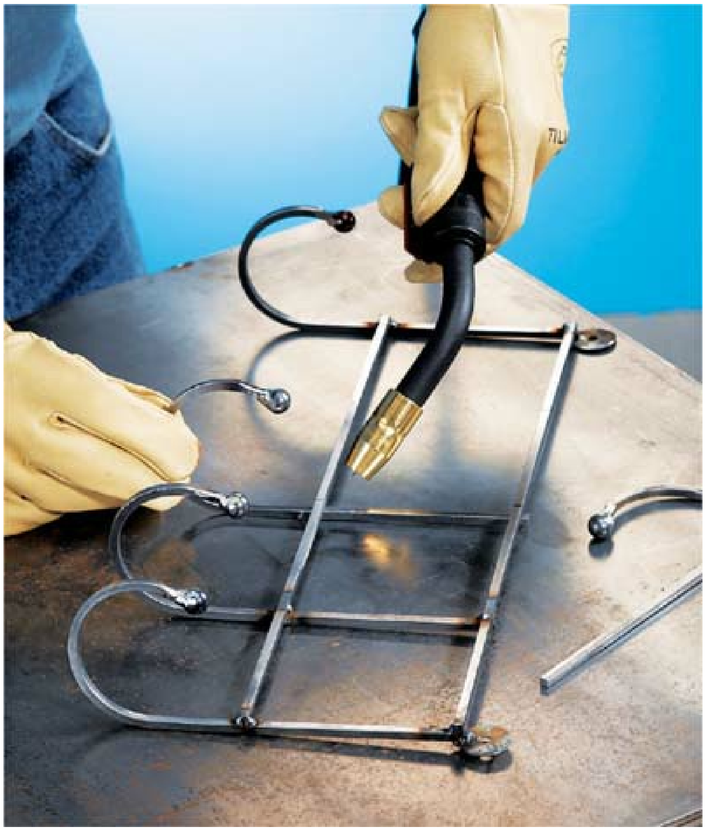

Weld the crossbars to the end hooks. Then weld the remaining hooks centered at 4", 8", and 12".

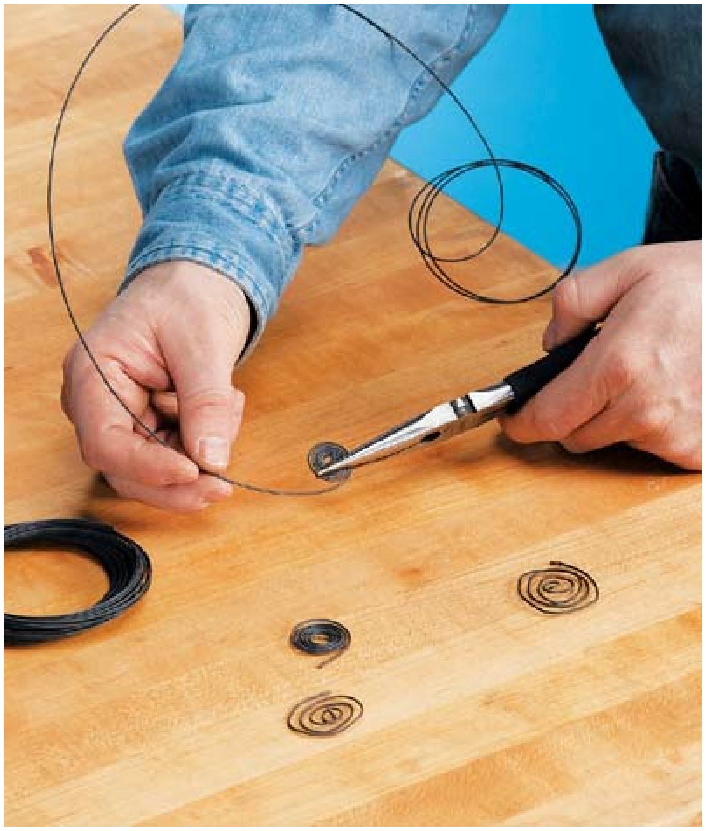

Begin a spiral by wrapping the wire around the tip of a needlenose pliers. Then grip the circle in the pliers and bend the wire until the desired size spiral is achieved.

Make the Circles & Spirals

1. Cut the circles (E) to length.

2. Clamp one end of a circle blank to a 4" pipe using a vise-style pliers. Bend the rod around the pipe one and a half turns to form a circle. Repeat to form the remaining circles.

3. Create the spirals (F) and mini-spirals (G) using a needlenose pliers. Wrap the wire around the end of the pliers to create a circle. Remove the circle from the pliers and grip it in the pliers. Wrap the wire around the circle until the desired size spiral has been formed.

Assemble the Decorative Features

1. Place a circle in the space between the first two hooks starting either on the far left or far right end. Squeeze or expand the circle to fit into the space and cut off any extra length. Weld the circle ends together and weld the circle to the two hooks.

2. Repeat step 2 with the remaining circles.

3. Arrange three mini-spirals inside the second and fourth circles. Weld them to each other and to the circle at contact points.

4. Place the large spirals between the remaining hooks and weld in place.

5. Wire brush or sandblast the coat hanger. Apply the desired finish.

Room Divider

Folding screens are versatile interior design pieces. Used as an accent behind furniture, as a room divider, or as a way to hide an unsightly outlet or wall blemish, this piece holds its own as a piece of art but is also practical furniture. In large spaces it provides a focus while also creating a seamless flow throughout the room. But don’t limit this folding screen to indoor spaces. The all-metal version presented here could easily be incorporated into a garden as a trellis or fence panel.

This project includes two options. You can make a gathered fabric panel or a decorative metal panel. Or you could create both together. As constructed, this screen is moderately sized. Measure the area where the screen will reside and change the proportions accordingly.

Any number of decorative options can be used for the screen panels. Stained glass and pierced or expanded sheet metal are options not covered here, but they would be fabulous additions. If you spend some time with the catalogs of the decorative metal suppliers (see Resources, page 220) you will discover many leaf and flower stampings that could be used to create your own personalized pattern. Grapevines, oak leaves, and sunflowers are available in many different sizes; animal figures, sporting figures, and numbers and letters are available as well.

Note: If you don’t care to cut out the leaves as listed in this project, use a small dapped leaf by Architectural Iron Designs. Try to find a leaf in a similar pattern.

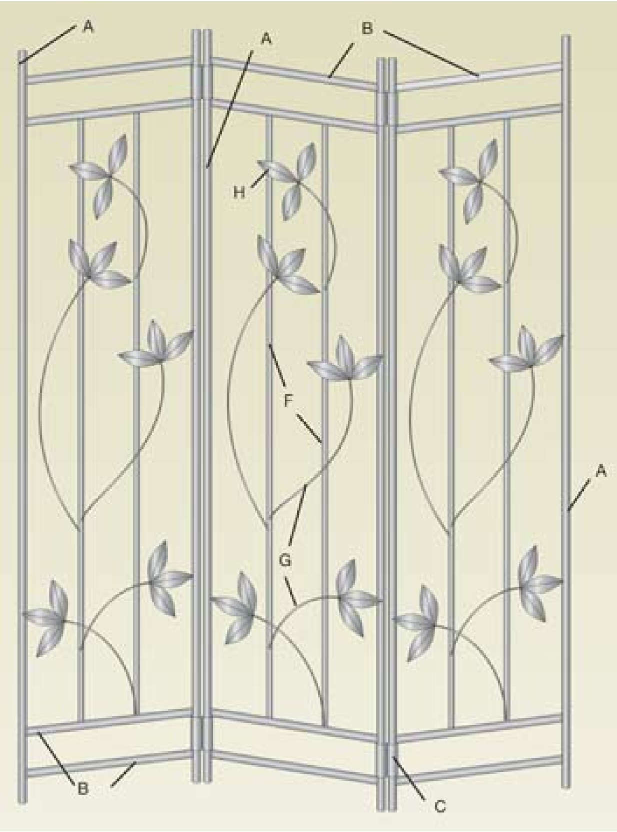
A
A
B
H
F
G
A
B
C

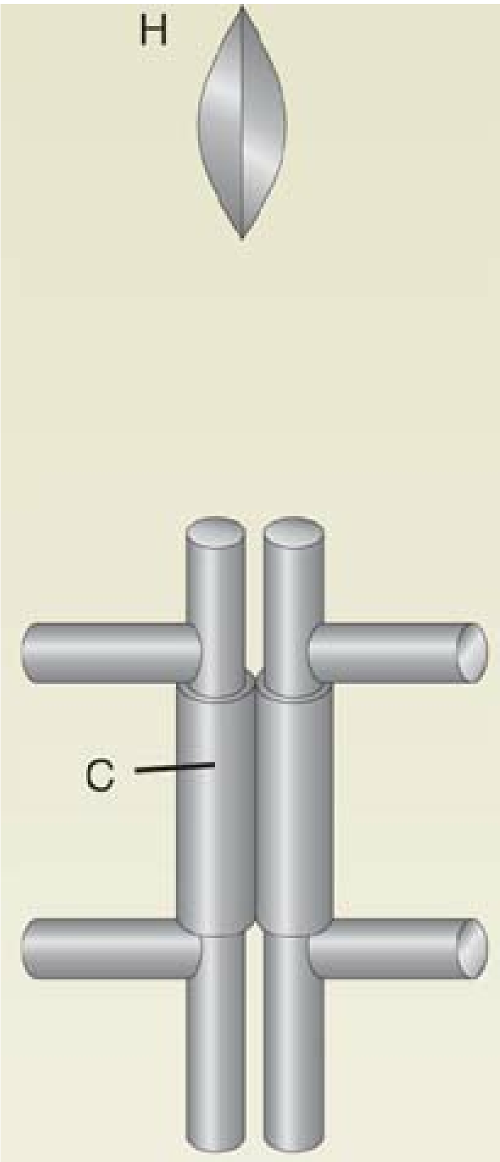
H
C

Materials

- 16-gauge ¾" round tube (49 feet)
- 16-gauge 1" round tube (2 feet)
- ¼" round rod (6 feet, optional)
- Fabric (5 yards)
- 16-gauge ½" round tube (28 feet)
- 1/8" round rod (14 feet)
- 22-gauge sheet metal (12 × 24")

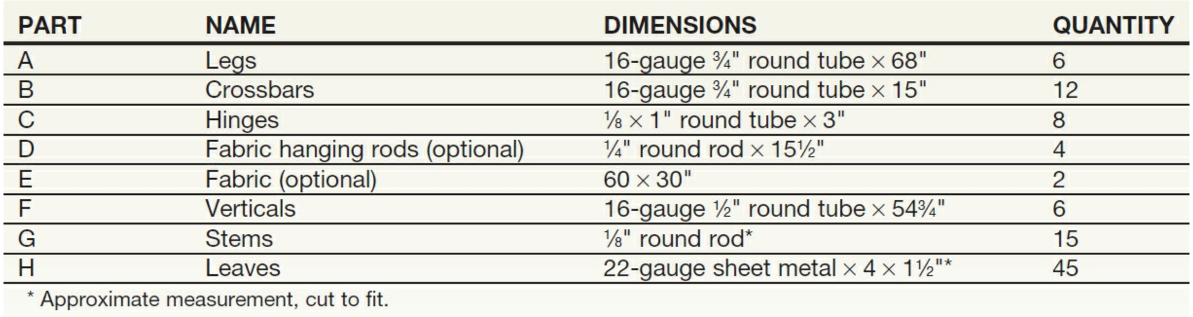

PART	NAME	DIMENSIONS	QUANTITY
A	Legs	16-gauge ¾" round tube × 68"	6
B	Crossbars	16-gauge ¾" round tube × 15"	12
C	Hinges	⅛ × 1" round tube × 3"	8
D	Fabric hanging rods (optional)	¼" round rod × 15½"	4
E	Fabric (optional)	60 × 30"	2
F	Verticals	16-gauge ½" round tube × 54¾"	6
G	Stems	⅛" round rod*	15
H	Leaves	22-gauge sheet metal × 4 × 1½"*	45

* Approximate measurement, cut to fit.

How to Build a Room Divider

Before welding, thoroughly clean all parts with denatured alcohol.

Make the Frames

1. Cut the legs and crossbars (A and B) to length.

2. Mark each leg at 6" in from each end.

3. Using a bench or angle grinder, grind the end of each crossbar to shape it to fit around the leg. Make sure the curves are oriented in the same direction at both ends.

4. Place two legs on a large work surface. Align the bottom of one crossbar and the top of another with the 6" marks. Tack weld the crossbars in place. Check the frame for square and weld in place. Grind the welds smooth.

5. Repeat step 4 for the remaining two panels.

Add the Hinges & Complete the Frames

1. Cut the hinges (C) to length.

2. Slide the hinges over all four leg ends of one panel and the top and bottom leg ends of the one side of the other two panels.

3. Align the second set of crossbars above and below the hinges and tack weld in place. Check for square and weld. Make certain you are not welding the hinges to the legs as you weld the crossbars.

Option 1: Create a Fabric Panel

The fabric panel is gathered. If you prefer a smooth look, decrease the fabric width to 15½".

Make the Hanging Rods

1. Cut the hanging rods (D) to length.

2. Mark points on the inside of each panel leg, ½" down from the inside top and bottom crossbars.

3. At the marked points, drill a $1^{7}/_{64}$" hole through the leg wall.

4. Insert one end of a hanging rod into a hole until it contacts the opposite wall. It should then be possible to insert the other end of the hanging rod into the hole on the opposite leg. If it is too long, grind it down until it fits.

5. Lay the three panels flat and parallel to each other, hinges touching. Weld each pair of hinges together.

6. Grind down any rough welds and grind off spatter. Wipe down with denatured alcohol to remove grinding dust. Finish as desired.

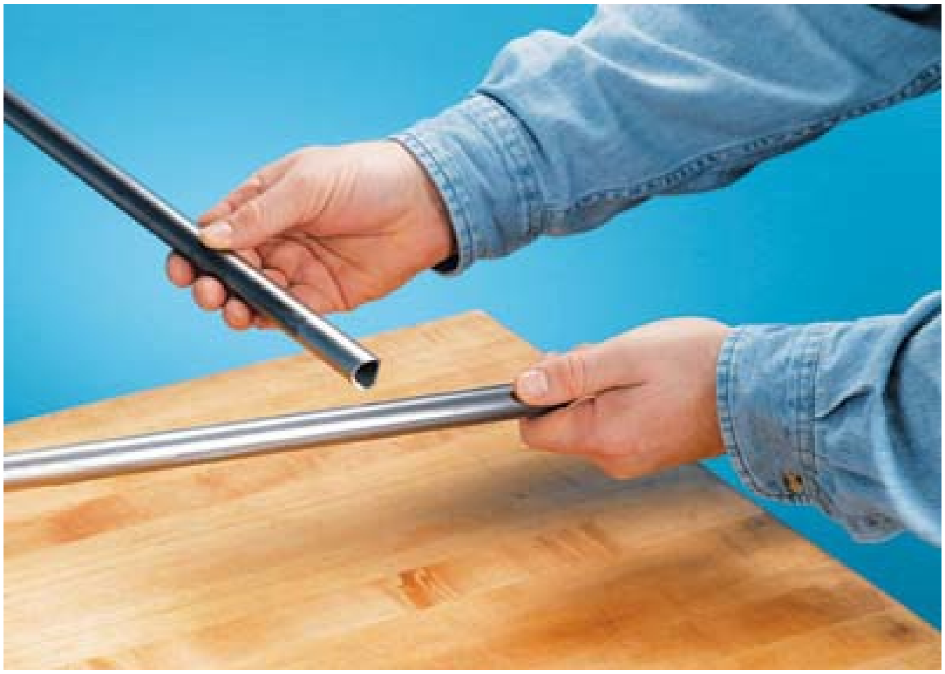

Grind a semicircle in the crossbar ends to create a saddle to fit around the legs.

Make the Fabric Panel

1. Make a ¼" flat-felled hem on the long sides of the fabric.

2. Drape the fabric over the top hanging rod so that equal amounts are extending past the top and bottom hanging rods.

3. Pin the fabric together around the rods.

4. Remove the rods and sew the rod pocket, turning the raw edge under.

5. Slide the fabric over the rods and reinstall the rods.

Option 2: Create a Decorated Panel

Make the Supports & Stems

1. Cut the verticals (F) to length.

2. Mark the interior crossbars at 5" and 10".

3. Center the supports over the marks and tack weld in place. Check for square and complete the welds. Grind the welds smooth.

4. Lay out the stems (G) in the desired pattern across the supports.

Make the Leaves

1. Cut the leaves (H) to desired size using a tin snips or plasma cutter.

2. Place a leaf lengthwise in a bench vise and slightly bend it to form a crease down the center.

3. Arrange the leaves in clusters of three. Weld the leaves together at the base.

4. Place the leaf clusters on the stem ends and weld in place.

5. Lay out the three panels parallel to each other, hinges touching. Weld each pair of hinges together.

6. Grind down any rough welds and grind off any spatter. Wipe down with denatured alcohol to remove grinding dust. Finish as desired.

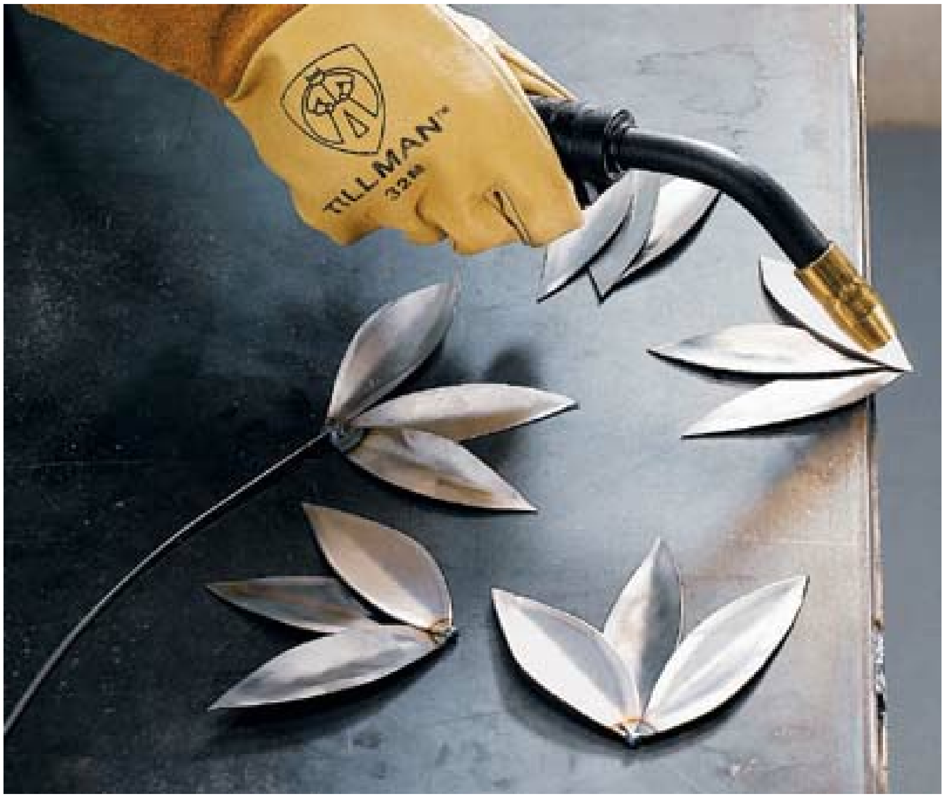
TILLMAN™
32M

Arrange the leaves in sets of three. Weld the points together, then weld the stem to the leaf sets.

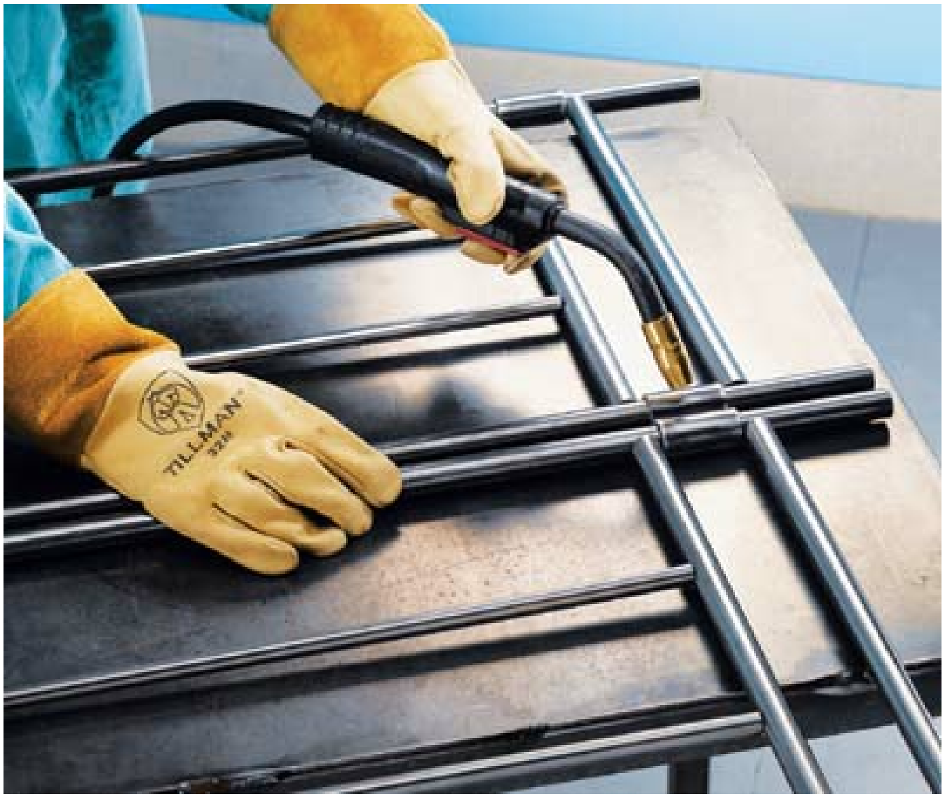
TILLMAN
3214

Align the hinge barrels and weld them together.

Iron Cross

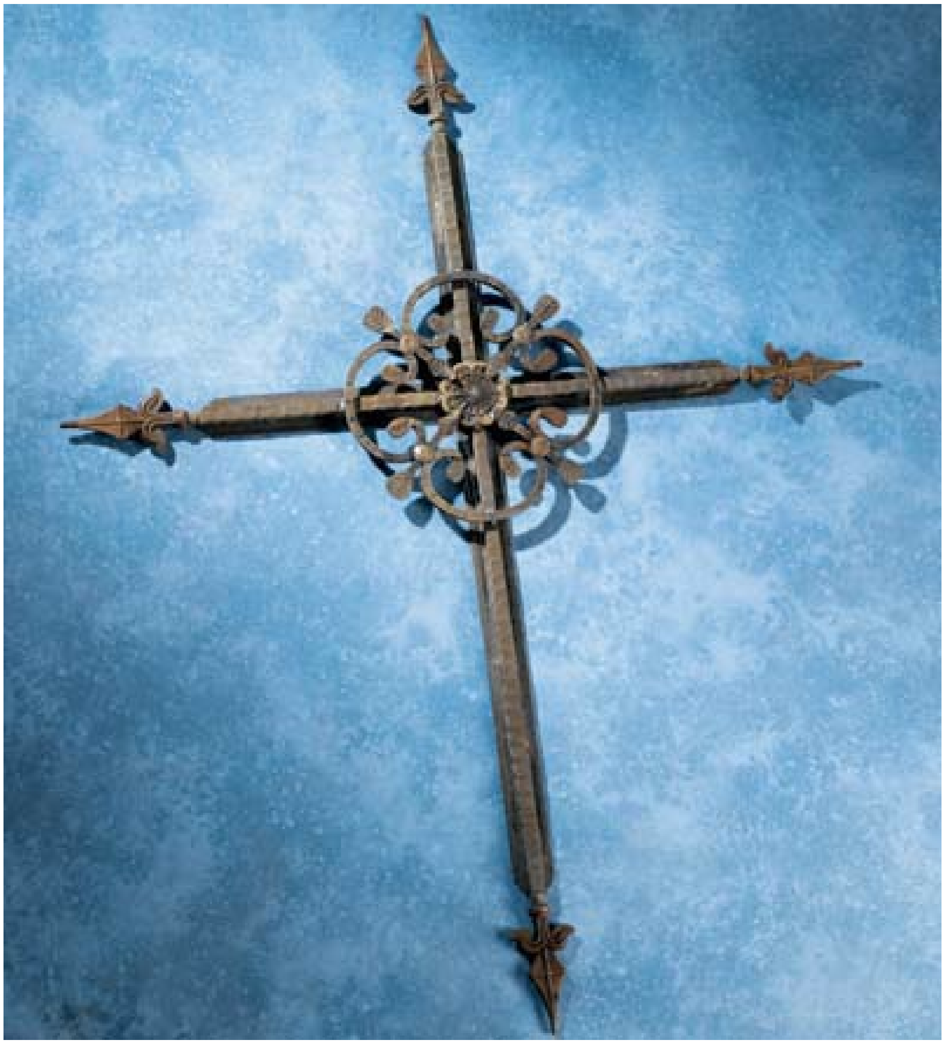

Rusted iron cross wall hangings evoke a sense of history. They create a distinctive air of mystery when hung over a fireplace or against exposed brickwork. Design centers and antique stores want hundreds of dollars for something that you can make for much less, so run wild with your imagination on this project and create a cross that personally resonates with you. The more ornate, the better. Search the catalogs from the suppliers listed in the Resource section on page 220 for just the right touches—you’ll find dozens of decorative pieces to make this project uniquely yours.

Rather than purchasing a 10- or 20-foot length of hammered or decorative bar—that’s a lot of crosses—buy two 39" hammered railing pickets. To get a natural rust finish, thoroughly clean the completed cross with denatured alcohol, wet it down with salt water, and allow it to sit outside for a few months. You’ll end up with a naturally beautiful rust.

Materials

- 1/8 × 1½" flat bar (4½ feet)
- ½" hammered bar (4½ feet, Decorative Iron PM-013)
- 2" rosette (1, Decorative Iron LRD-65)
- Scroll set (1, Architectural Iron Designs #55/9)
- Finials (4, Architectural Iron #81/3)

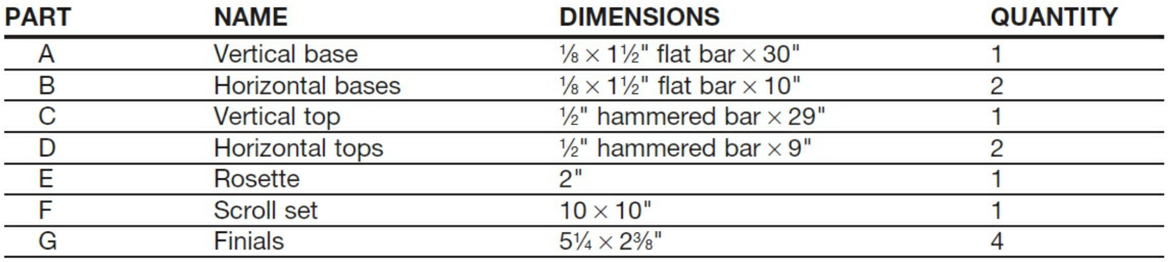

PART	NAME	DIMENSIONS	QUANTITY
A	Vertical base	⅛ × 1½" flat bar × 30"	1
B	Horizontal bases	⅛ × 1½" flat bar × 10"	2
C	Vertical top	½" hammered bar × 29"	1
D	Horizontal tops	½" hammered bar × 9"	2
E	Rosette	2"	1
F	Scroll set	10 × 10"	1
G	Finials	5¼ × 2⅜"	4

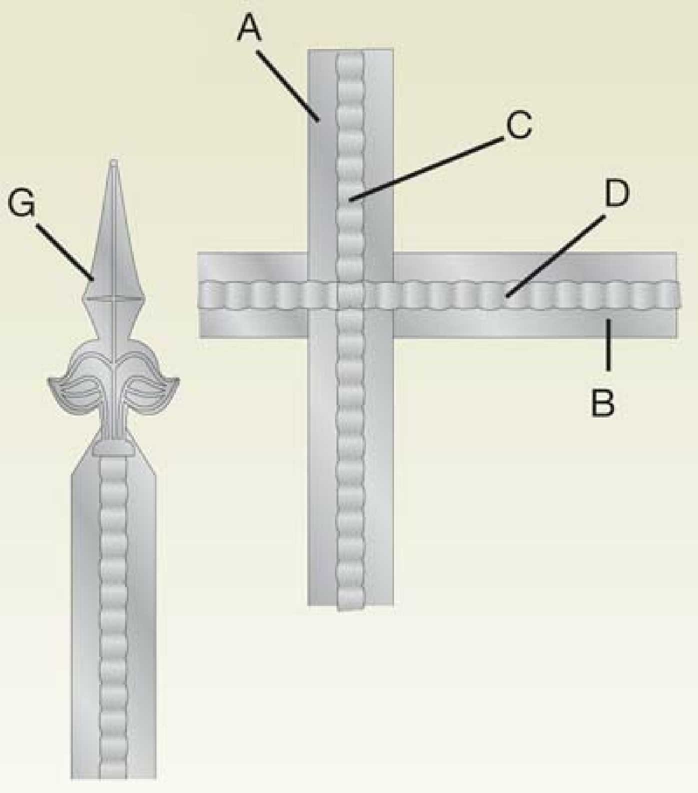
A
C
G
D
B

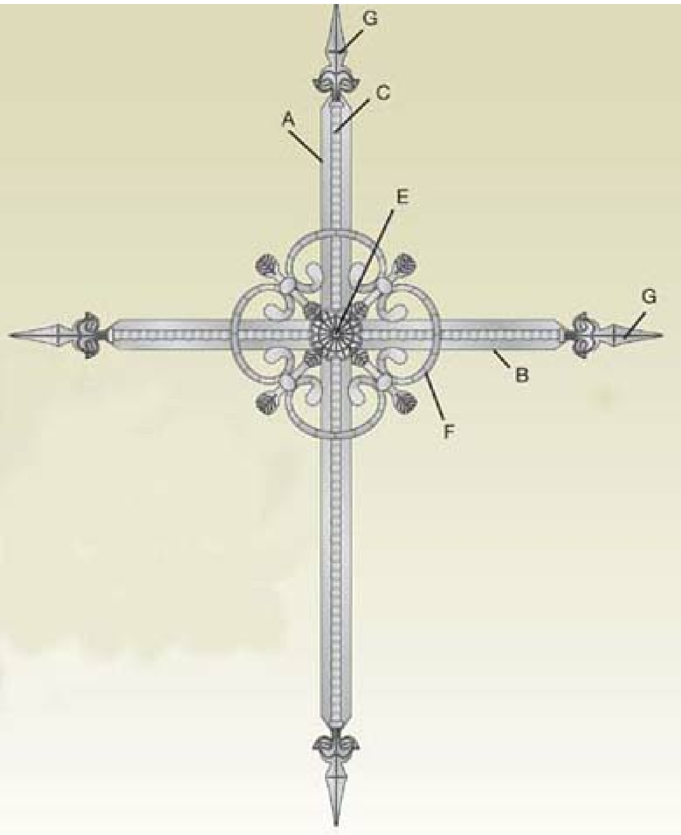
G
C
A
E
G
B
F

How to Build an Iron Cross

Before welding, thoroughly clean all parts with denatured alcohol.

Make the Base

1. Cut the base bars (A and B) to length.

2. Mark the center point of each end of the vertical base and one end of each horizontal base. Mark the bars 1" in from the ends. Draw lines from the 1" mark to the center point to create a point. Cut along the lines.

3. Mark the vertical base 11" from one end. Center the two horizontal bases at this mark to form the cross. Check for square and weld in place. Grind the welds smooth.

Make the Top

1. Cut the vertical and horizontal tops (C and D) to length.

2. Center the top vertical bar on the base. The points should extend ½ to ¾" beyond the bar ends. Weld the top bar to the base with two to three welds along each side.

3. Center the top horizontals on the base. They should touch the top vertical bar at the center and end about an inch before the points. Weld in place.

Add the Decorative Elements

1. Center the rosette (E) in the middle of the cross and weld in place.

2. Center the scroll set (F) over the rosette. Bend the arrow tips upward until the scroll set makes contact with all four arms. Weld in place.

3. Center the finials (G) on the bar ends and weld in place.

Mark the midpoint of both vertical base ends and one end of each horizontal base. Mark 1" from the ends. Connect the marks at a diagonal to create a point.

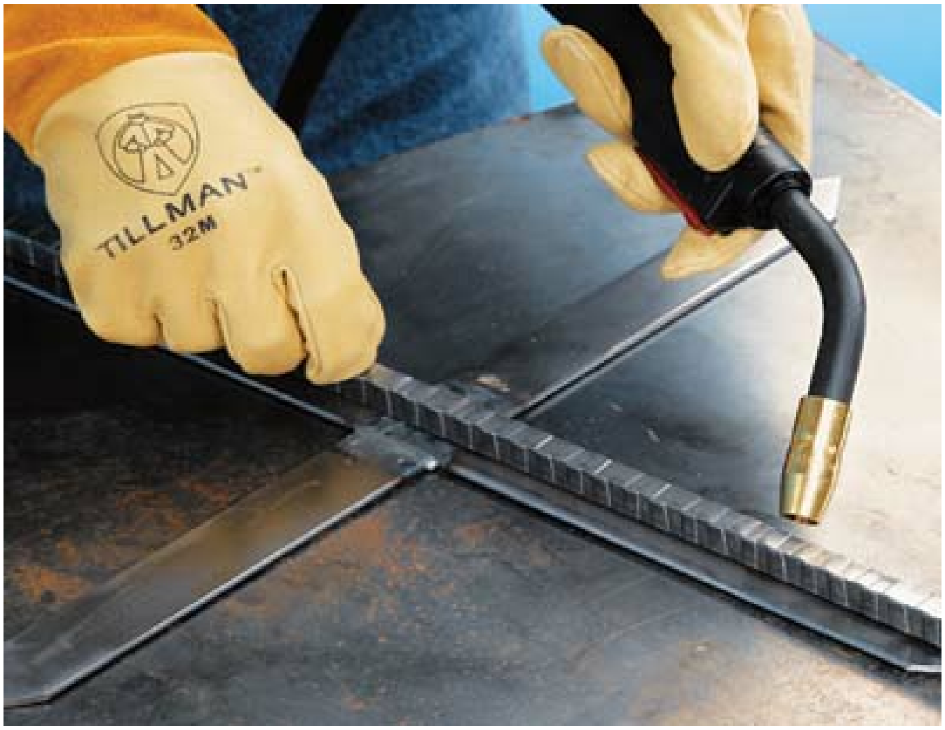
TILLMAN™
32M

Center the vertical top on the vertical base and weld in place with two to three welds per side.

Finish the Cross

1. Drill $^{3}/_{16}$" mounting holes through the vertical base on either side of the top bar and about 1" from the top and bottom.

2. Wire brush or sandblast the cross. Finish as desired.

3. Because of the weight of the cross, make sure it is mounted on a wall stud. Hang the cross with 1½" screws.

Drill mounting holes through the vertical base. Use machine oil and a slow drill speed to preserve the drill bit.

Nesting Tables

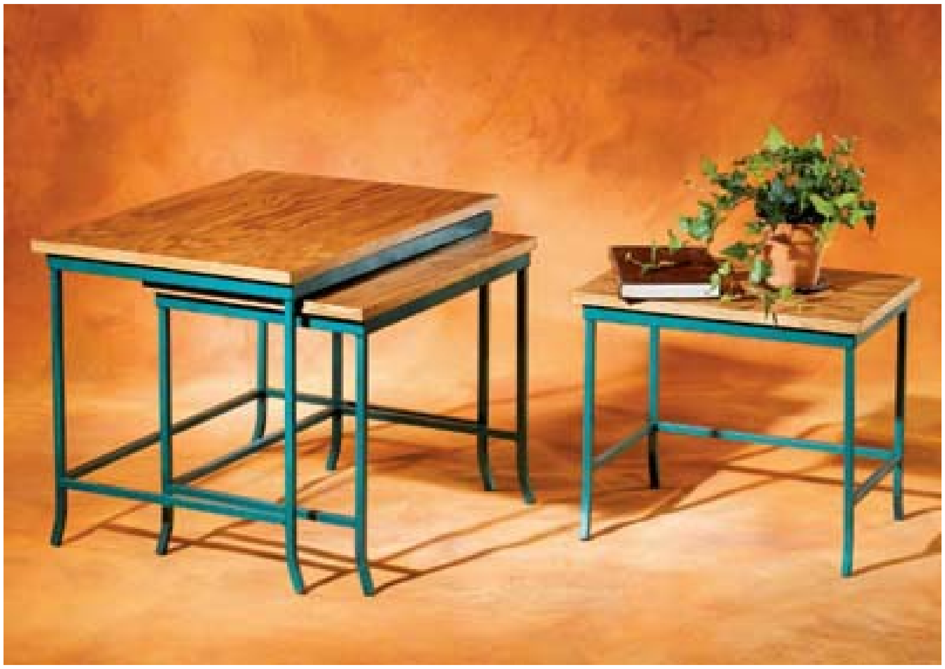

Do you like to entertain but find that you’re constantly searching for extra tables as the guests arrive? Nesting tables are the perfect way to always be ready to entertain on the fly. Best of all, they can be tucked away at the end of the evening. The basic pattern here is for three tabletops. These tables have wooden tops with a ½" overhang. You may adjust the measurements for tile or metal tops. If you want the tables to nest even closer, use ½" angle iron with a 16-gauge or 1/8" metal top. And don’t forget to dress these tables up! Consider using hammered bar or adding decorative elements to the legs to allow them to show off a bit at parties.

Materials

- ½ × ½" square bar (28 feet)
- 1 × 1 × 1/8" angle iron (17 feet)
- ½" handipanel or veneered plywood (2 ft. × 4 ft. sheet)
- ¾" #8 panhead screws (12)

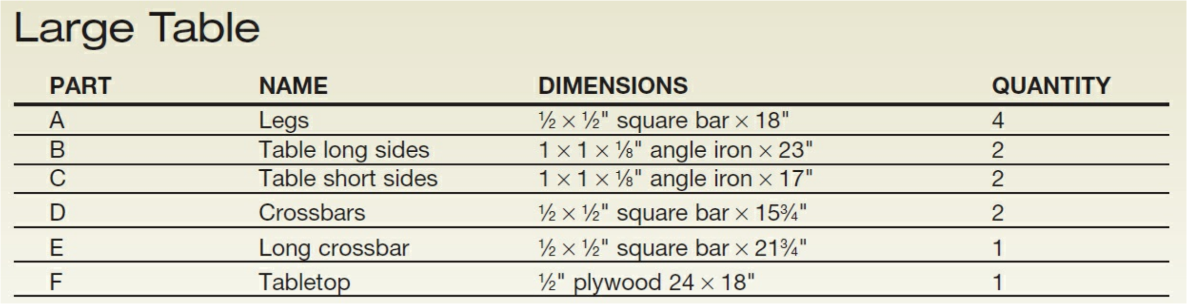

Large Table

PART	NAME	DIMENSIONS	QUANTITY
A	Legs	½ × ½" square bar × 18"	4
B	Table long sides	1 × 1 × ⅛" angle iron × 23"	2
C	Table short sides	1 × 1 × ⅛" angle iron × 17"	2
D	Crossbars	½ × ½" square bar × 15¾"	2
E	Long crossbar	½ × ½" square bar × 21¾"	1
F	Tabletop	½" plywood 24 × 18"	1

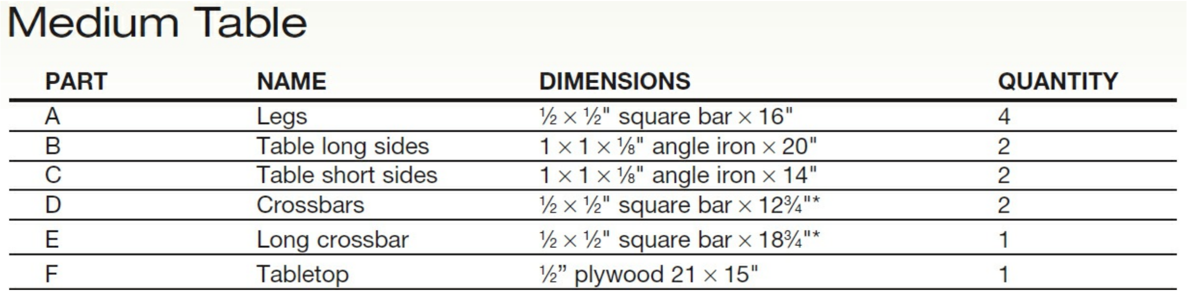

Medium Table

PART	NAME	DIMENSIONS	QUANTITY
A	Legs	½ × ½" square bar × 16"	4
B	Table long sides	1 × 1 × ⅛" angle iron × 20"	2
C	Table short sides	1 × 1 × ⅛" angle iron × 14"	2
D	Crossbars	½ × ½" square bar × 12¾"*	2
E	Long crossbar	½ × ½" square bar × 18¾"*	1
F	Tabletop	½” plywood 21 × 15"	1

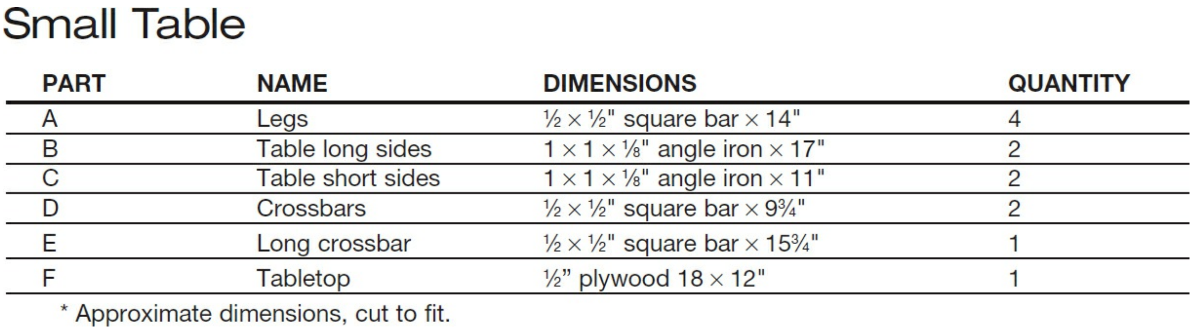

Small Table

PART	NAME	DIMENSIONS	QUANTITY
A	Legs	½ × ½" square bar × 14"	4
B	Table long sides	1 × 1 × ⅛" angle iron × 17"	2
C	Table short sides	1 × 1 × ⅛" angle iron × 11"	2
D	Crossbars	½ × ½" square bar × 9¾"	2
E	Long crossbar	½ × ½" square bar × 15¾"	1
F	Tabletop	½” plywood 18 × 12"	1

* Approximate dimensions, cut to fit.

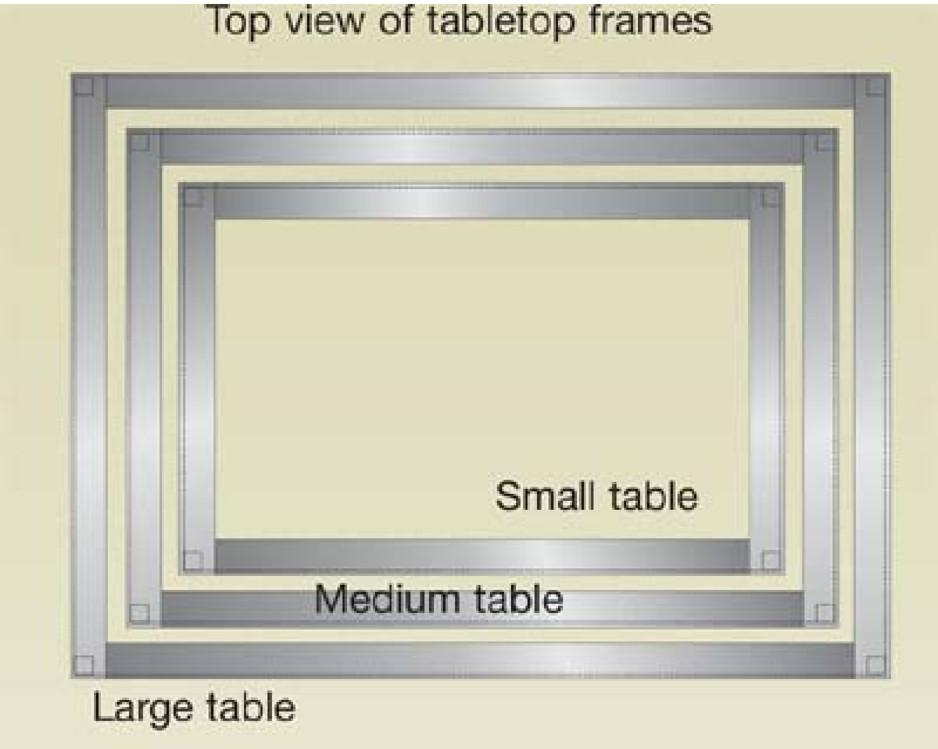
Top view of tabletop frames
Small table
Medium table
Large table

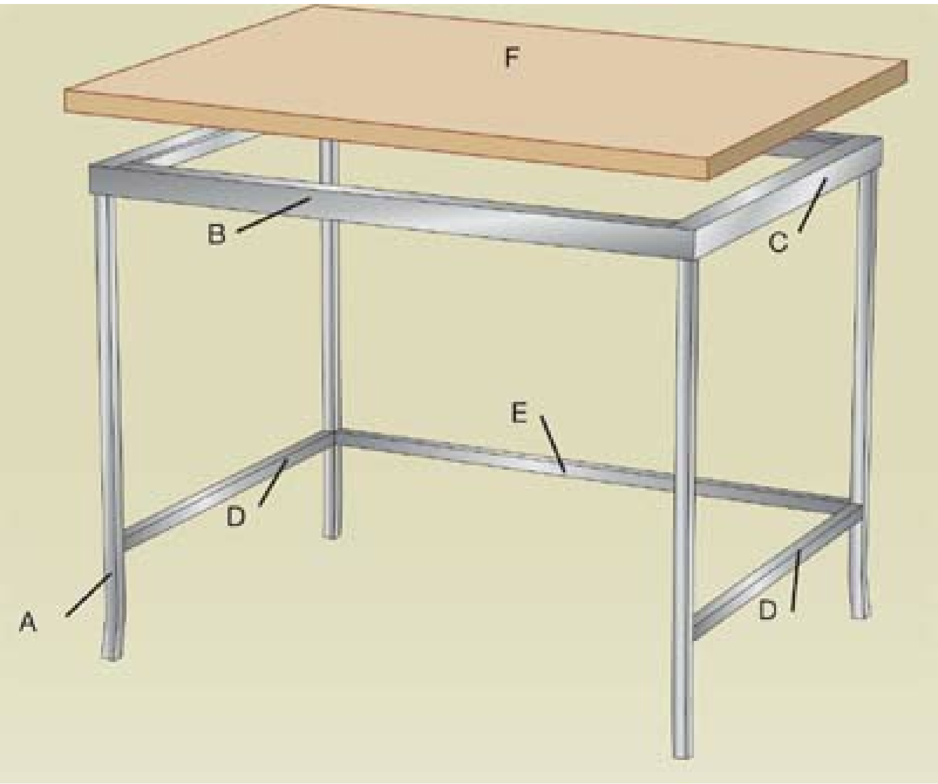
F
B
C
E
D
A
D

How to Build Nesting Tables

Before welding, thoroughly clean all parts with denatured alcohol.

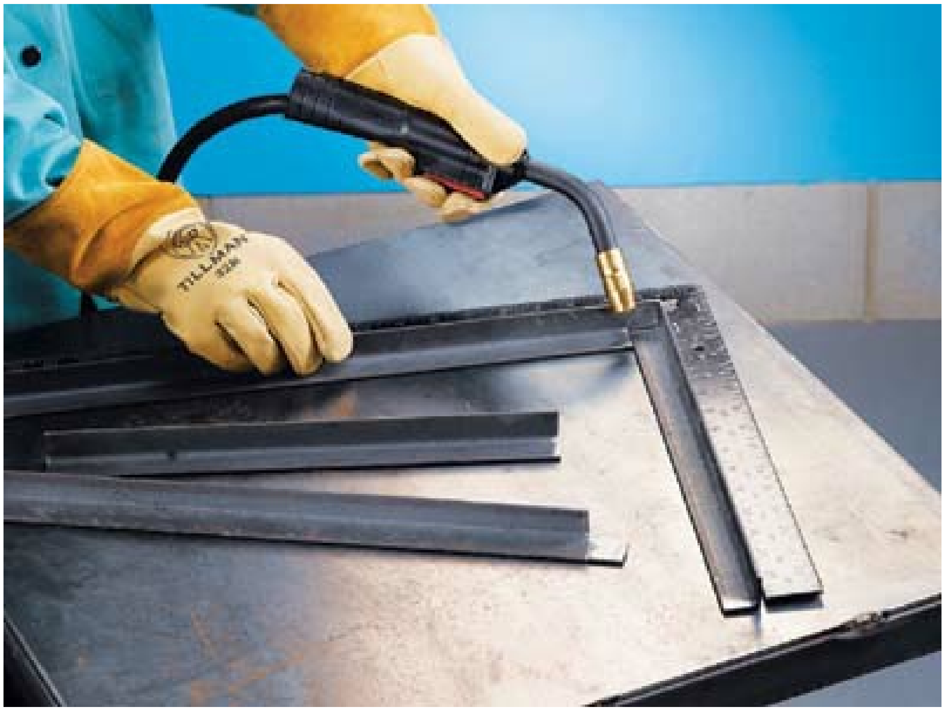
TILLMAN
1326

Using a carpenter’s square, lay out the table sides. Tack weld together and check for square before final welding.

Make the Legs

1. Cut the legs of the large table (A) to length.

2. Clamp the lower 1½" of a leg into a sturdy bench vise. Slide a 3-ft. length of ½" pipe over the leg. Bend it about 5°.

3. Using the first leg as a template, bend all the remaining legs to match.

4. Repeat steps 1 to 3 with the medium and small tables.

Make the Tabletop Frames

1. Cut the table long sides (B) and table short sides (C) to length. Miter the ends at 45°, or cut notches in the long sides.

2. Lay out one long side and one short side of the large table using a carpenter's square to check for square. Clamp into place and tack weld the corner.

3. Repeat Step 2 with the other long and short sides.

4. Use a carpenter's square to align the two Ls to form a rectangle.

5. Weld the outside joints together. Grind down the welds.

6. Repeat steps 1 to 5 with the medium and small tables.

Attach the Legs

If you will be making a wooden or metal tabletop, turn the table frame so the flanges of the angle iron are pointing downward. Weld the legs inside the corner. If you will be making a tile tabletop, turn the frame so the flanges are pointing up to create a cradle for the base and tile. Weld the legs to the bottom of the frame.

1. Place a leg on or in a corner, with the bend facing the long side. Use a carpenter's square to check that it is square from front to back and side to side. Weld in place.

2. Repeat step 1 with the remaining legs.

3. Measure the distance between the legs on the short sides, close to the frame. Cut the crossbars (D) to fit. Weld the crossbars into place at 4" from the bottom of the legs.

4. Measure the distance between the legs on a long side, close to the frame. Cut the back bar (E) to fit. Weld into place, even with the crossbars.

5. Repeat steps 1 to 4 with the medium and small tables.

6. Grind down all the welds. Sand, wire brush, or sandblast the tables. Finish as desired. (If using a metal top, install it before finishing.)

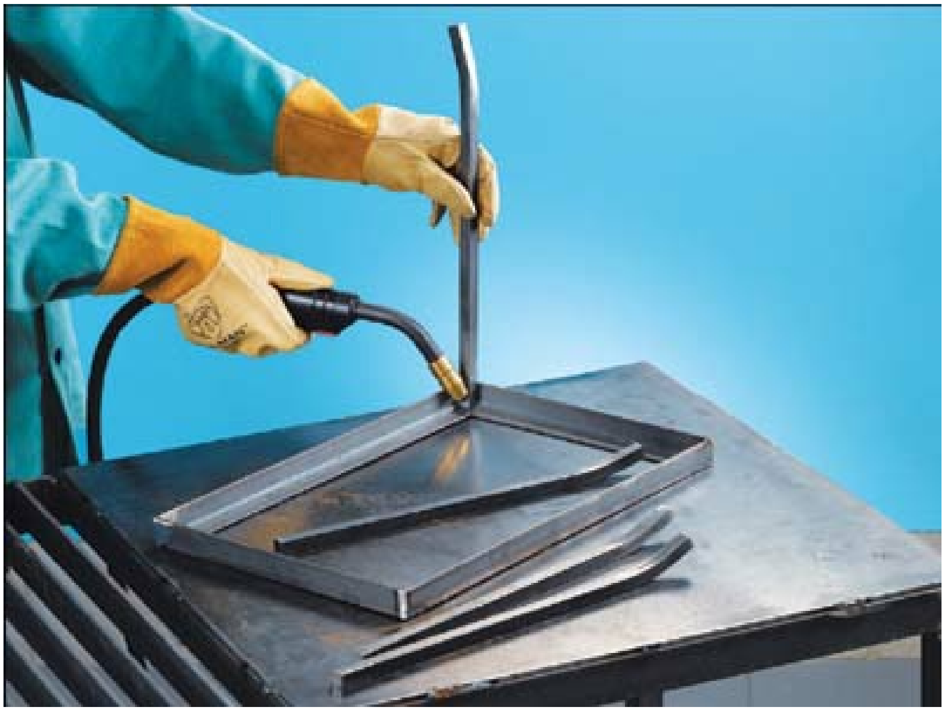

For a wood or metal tabletop, weld the legs inside the corner formed by the flanges. Align the curved legs so they point toward the long sides.

Option 1: Create a Wood Top

The wood tabletops have a ½" overhang.

1. Cut the wood tabletop (F) to size.

2. If using a handipanel, sand the edges smooth, slightly rounding the top edge and corners, or use a ½" roundover bit and rout the edges.

3. If the top is veneered plywood, apply edge banding with an iron. Trim the banding to fit and lightly sand the edges and corners.

4. Apply paint, stain, or finish.

5. Turn the table upside down and drill a $^{3}/_{16}$" hole in each side. Drill a corresponding ⅛" pilot hole in the underside of the tabletop. Attach the top to the table with ¾" panhead screws.

6. Repeat steps 1 to 5 with the medium and small tables.

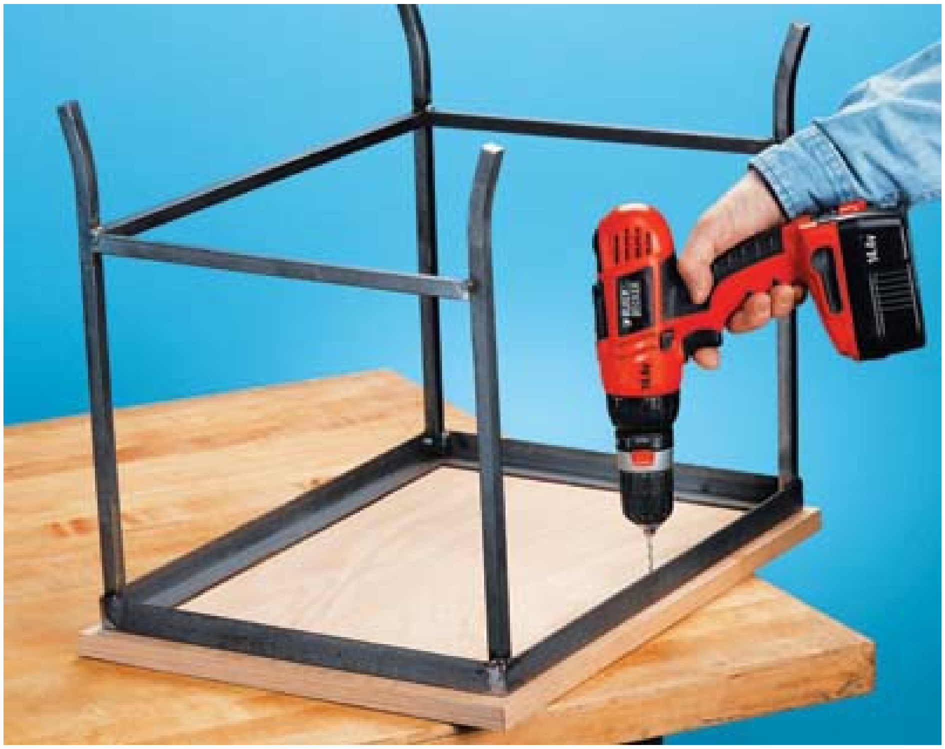

After drilling holes in each of the table’s sides, align the wood top and drill pilot holes for the mounting screws.

Option 2: Create a Tile Top

1. Measure the inside dimension of the tabletop. Cut ¾" exterior-grade plywood to fit.

2. Mark the plywood piece with your desired tile layout.

3. Apply a small amount of silicone adhesive to the lip of the table. Place the plywood onto the table.

4. Apply a layer of adhesive to the back of each tile using a notched trowel. Place the tiles on the plywood according to your pattern. After placing the tiles, use a carpet-covered 2 × 4 and a rubber mallet to set the tiles.

5. After the adhesive sets, apply grout to the tile spaces according to manufacturer's instructions. Do not apply grout to the gap between the metal frame and the tiles. Use colored silicone sealant between the metal and the tiles.

6. Repeat steps 1 to 5 with the medium and small tables.

Option 3: Create a Metal Top

You can make a metal top of matching steel or create a contrasting look with expanded sheet metal, patterned stainless steel, or aluminum. A glass panel covering patterned metal creates a smooth surface while allowing the pattern to shine through.

1. Cut the metal to fit the dimensions of the table.

2. Tack weld the top to the table from the underside, or weld around the entire perimeter. Grind the contact points to create a smooth transition.

3. Repeat steps 1 and 2 with the medium and small tables.

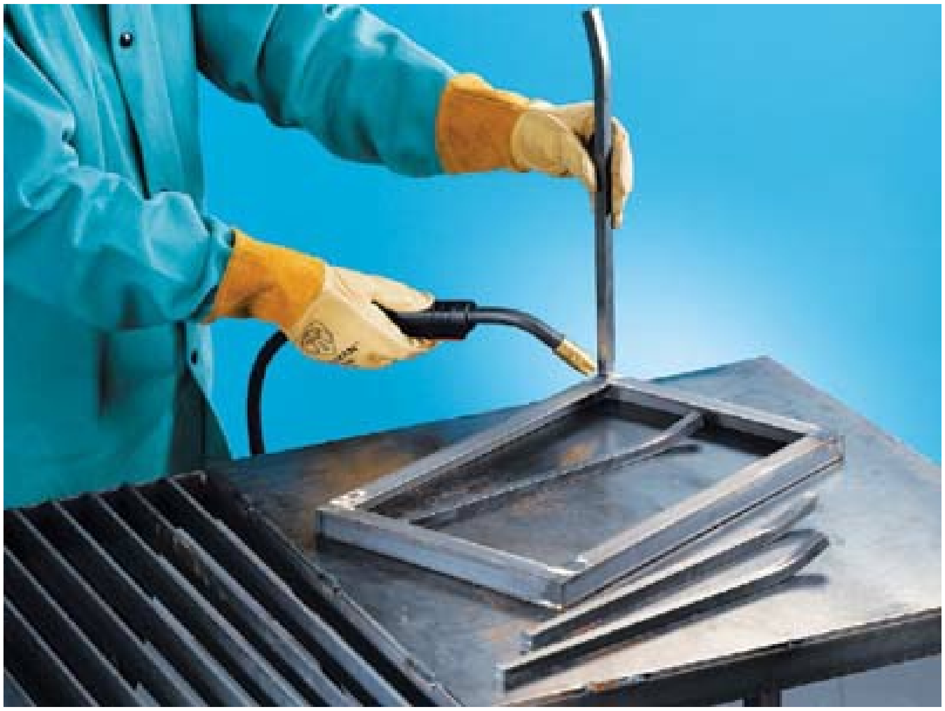

For a tiled tabletop, weld the legs to the flat underside so that the flanges form a cradle for the tile and backing material.

Swing-away Coat Hooks

These good-looking, easy-to-make coat hooks are considerably sturdier than those you might purchase at discount stores. The pivoting design lets them fold flat against the wall when not in use, so you can install them wherever you want.

Customize the hooks with a special shaped back plate or use cutout letters to personalize hooks for a child’s room. Make the hooks shorter, if desired, or use more than three hooks. Mount the hooks on a stud to prevent the weight of coats from pulling them out of the wall.

Materials

- 16-gauge ½" round tube (9")
- ¼" round rod (10")
- 3/16" round rod (6½ feet)
- 1/8 × 2" flat bar (12")
- 1" O.D. ball ¼" 20 threaded (2, Decorative Iron #C50035)

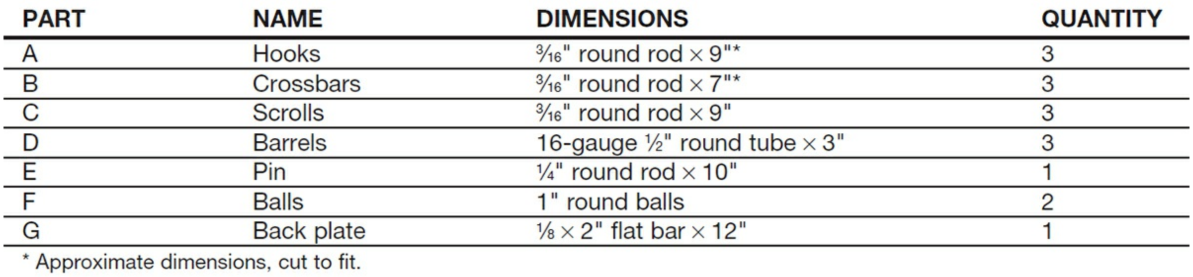

PART	NAME	DIMENSIONS	QUANTITY
A	Hooks	3⁄16" round rod × 9"*	3
B	Crossbars	3⁄16" round rod × 7"*	3
C	Scrolls	3⁄16" round rod × 9"	3
D	Barrels	16-gauge ½" round tube × 3"	3
E	Pin	¼" round rod × 10"	1
F	Balls	1" round balls	2
G	Back plate	⅛ × 2" flat bar × 12"	1

* Approximate dimensions, cut to fit.

How to Build Swing-away Coat Hooks

Before welding, thoroughly clean all parts with denatured alcohol.

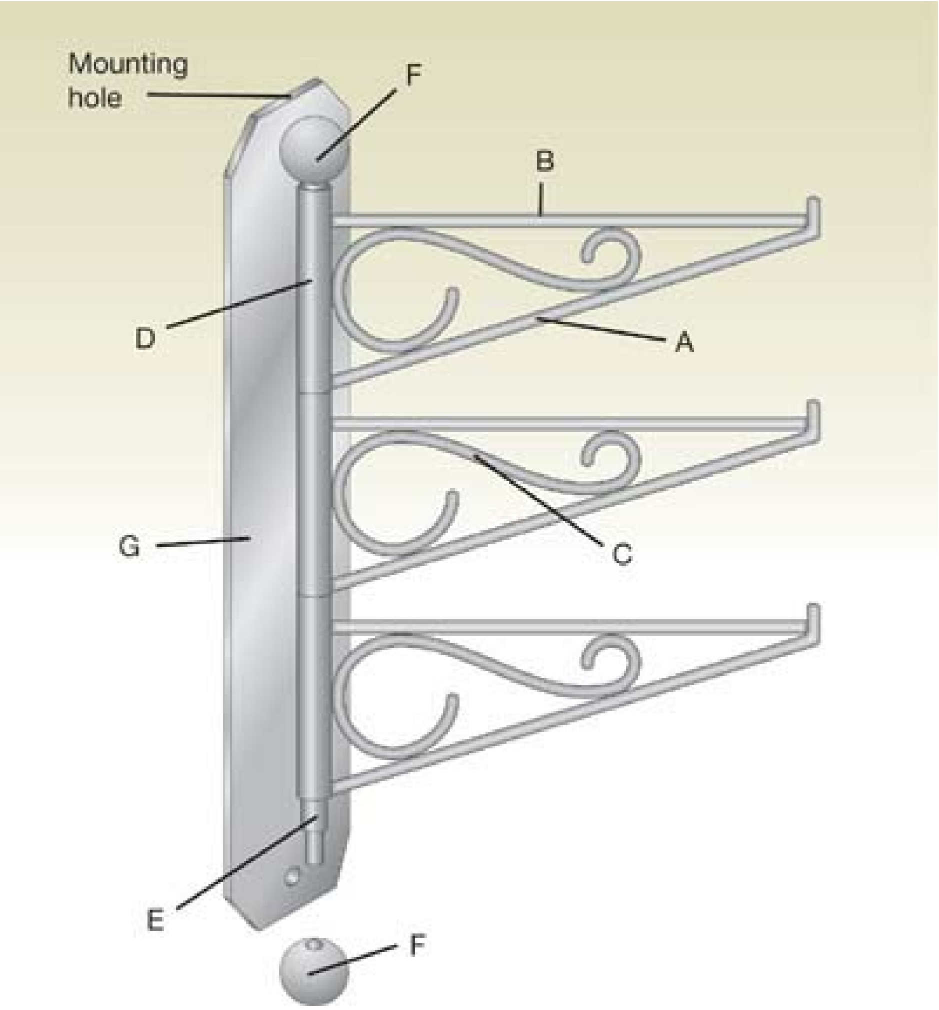
Mounting hole
F
B
D
A
C
G
E
F

Make the Scrolls

1. Cut the scroll blanks (C) to length. Round the rod ends with a grinder.

2. Clamp one end of the scroll blank to a ½" pipe. Bend the scroll around the pipe a three-quarter turn.

3. Clamp the other end of the scroll to a 1" pipe. Bend the scroll around the pipe a three-quarter turn in the opposite direction.

4. Repeat with the other two scrolls, matching the bends in the first scroll.

Make the Hooks

1. Cut the hooks and crossbars (A and B) to length. Round one hook end with a grinder.

2. Place 1" of the hook into a bench vise and bend to about 110°. Repeat with the other two hooks.

3. Align a crossbar at 90° to the hook and weld in place.

4. Trim the hook and crossbar ends to match, if necessary.

5. Place a scroll between the legs of a hook. The hook legs should be no more than 2¾" apart.

6. Arrange the other hooks and scrolls to match the first. Weld together.

Make the Hinge

1. Cut the barrels (D) to length. Draw a guideline down the center of each barrel. Make a mark at ½" down from the top of each barrel along the guideline.

2. Align the hook with the barrel so the upper leg of the hook is at the intersection of the guideline and the ½" mark.

3. Weld the hook legs to the barrel.

4. Repeat steps 2 and 3 with the other two hooks.

5. Cut the pin (E) to length. Grind down ½" at each end of the pin to fit the holes in the balls. Slide the barrels over the pin. Place the balls (F) over the pin at each end.

Assemble the Coat Hooks

1. Cut the back plate (G) to size. Round the corners with a grinder.

2. Drill $^{5}/_{32}$" mounting holes centered 1" from each end.

3. Measure the distance between the centers of the balls on the hook assembly. Drill ¼" holes this distance apart on the centerline of the back plate.

4. Place the hook assembly on the back so the balls rest in the holes. From the back side, weld the balls to the back plate.

5. Sand, wire brush, or sandblast the coat hooks. Finish as desired.

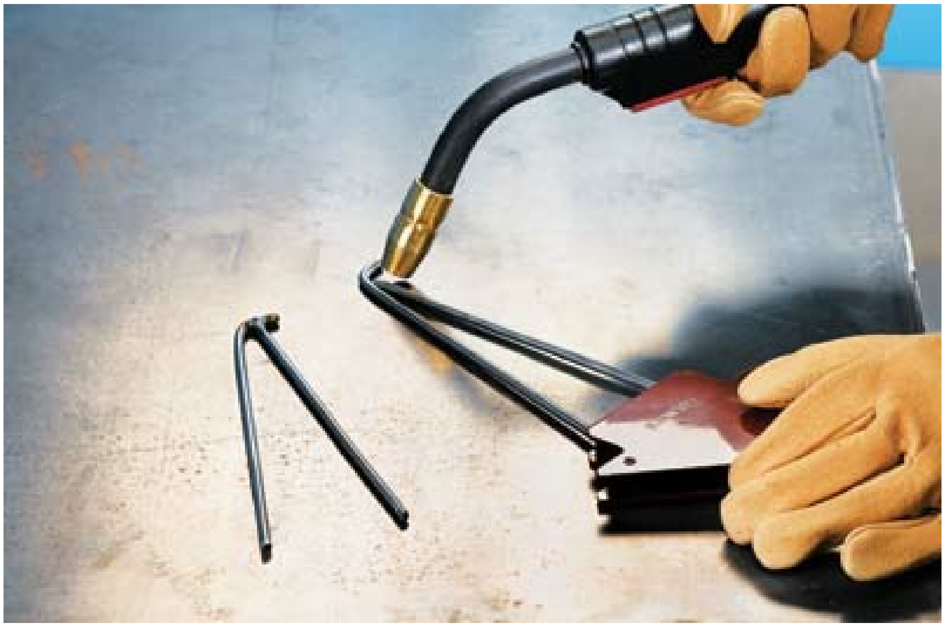

Align the crossbar with the hook and weld the ends together. A magnetic clamp helps hold the small parts in place.

Nightstand

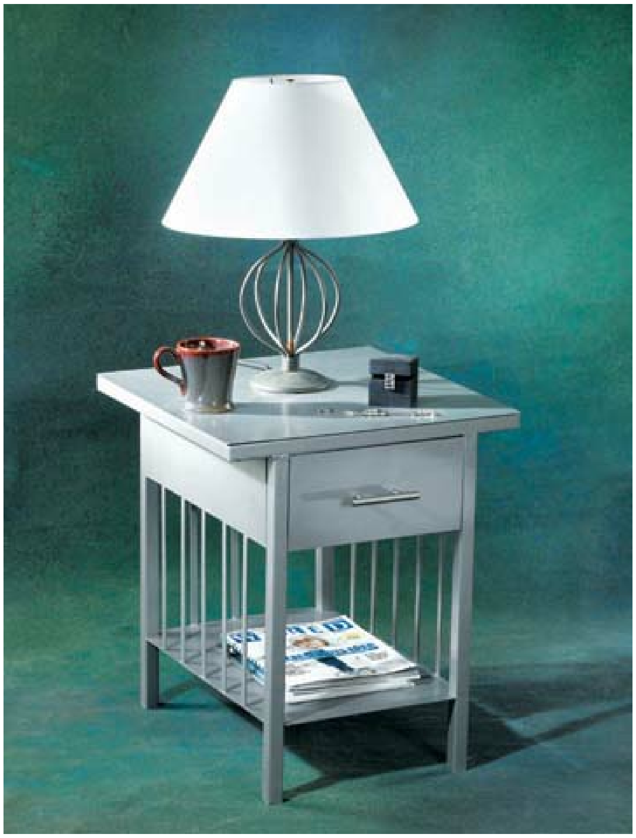

Sleek and sophisticated, this table may be designed like a wood table, but its unmatched style comes from metal. Many people find metal easier to work with than wood, and metal can offer the same great looks without all that sawdust.

When working with thinner materials, make sure to turn your welder to the appropriate setting to prevent burn through and distortion.

Materials

- 16-gauge 1 × 1" square tube (14½ feet)
- 16-gauge sheet metal (3 ft. × 5½ ft. sheet)
- 16-gauge angle iron (3 feet)
- 1/8 × ½" flat bar (14 feet)
- 22-gauge sheet metal (2 ft. × 1 ft. sheet)
- Drawer handle

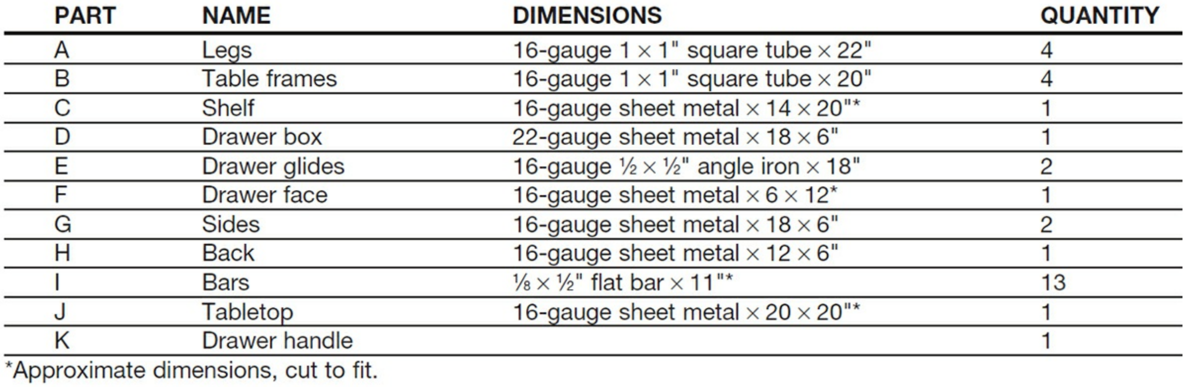

PART	NAME	DIMENSIONS	QUANTITY
A	Legs	16-gauge 1 × 1" square tube × 22"	4
B	Table frames	16-gauge 1 × 1" square tube × 20"	4
C	Shelf	16-gauge sheet metal × 14 × 20"*	1
D	Drawer box	22-gauge sheet metal × 18 × 6"	1
E	Drawer glides	16-gauge ½ × ½" angle iron × 18"	2
F	Drawer face	16-gauge sheet metal × 6 × 12*	1
G	Sides	16-gauge sheet metal × 18 × 6"	2
H	Back	16-gauge sheet metal × 12 × 6"	1
I	Bars	⅛ × ½" flat bar × 11"*	13
J	Tabletop	16-gauge sheet metal × 20 × 20"*	1
K	Drawer handle		1

*Approximate dimensions, cut to fit.

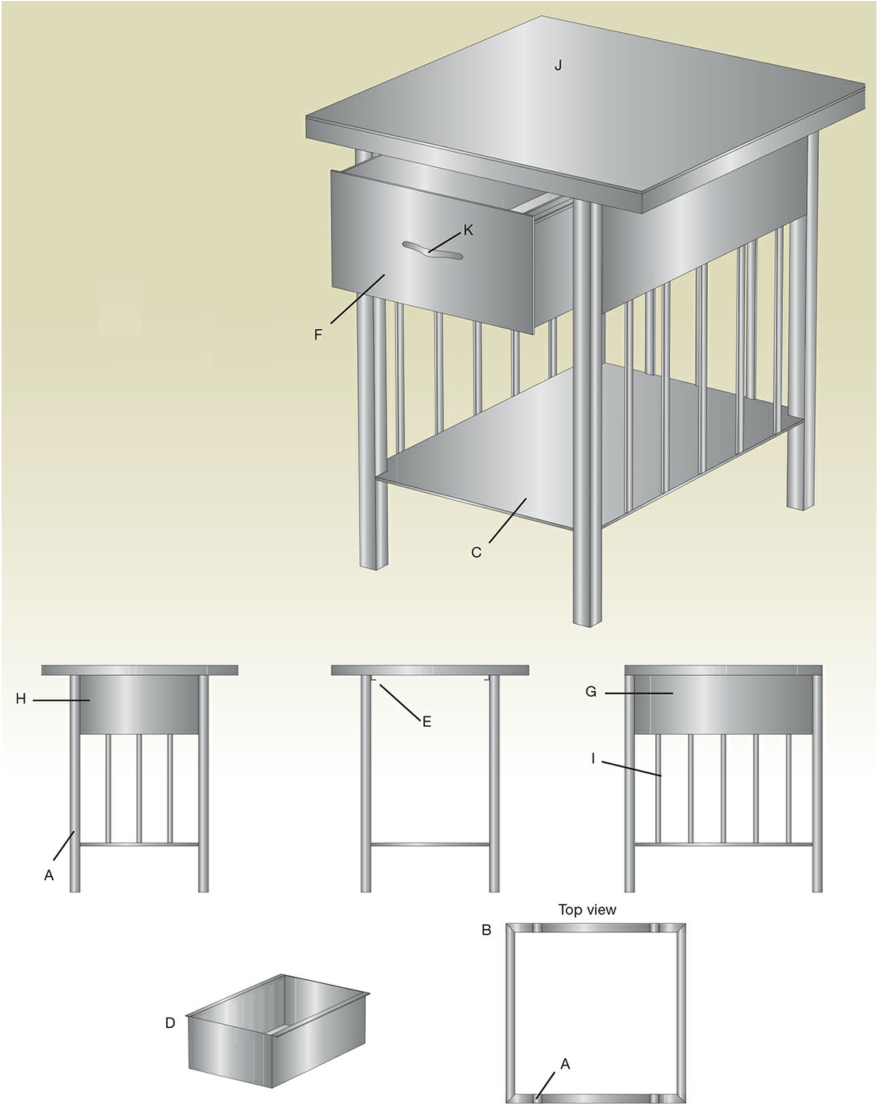
J
K
F
C
H
A
E
G
I
Top view
B
D
A

How to Build a Nightstand

Before welding, thoroughly clean all parts with denatured alcohol.

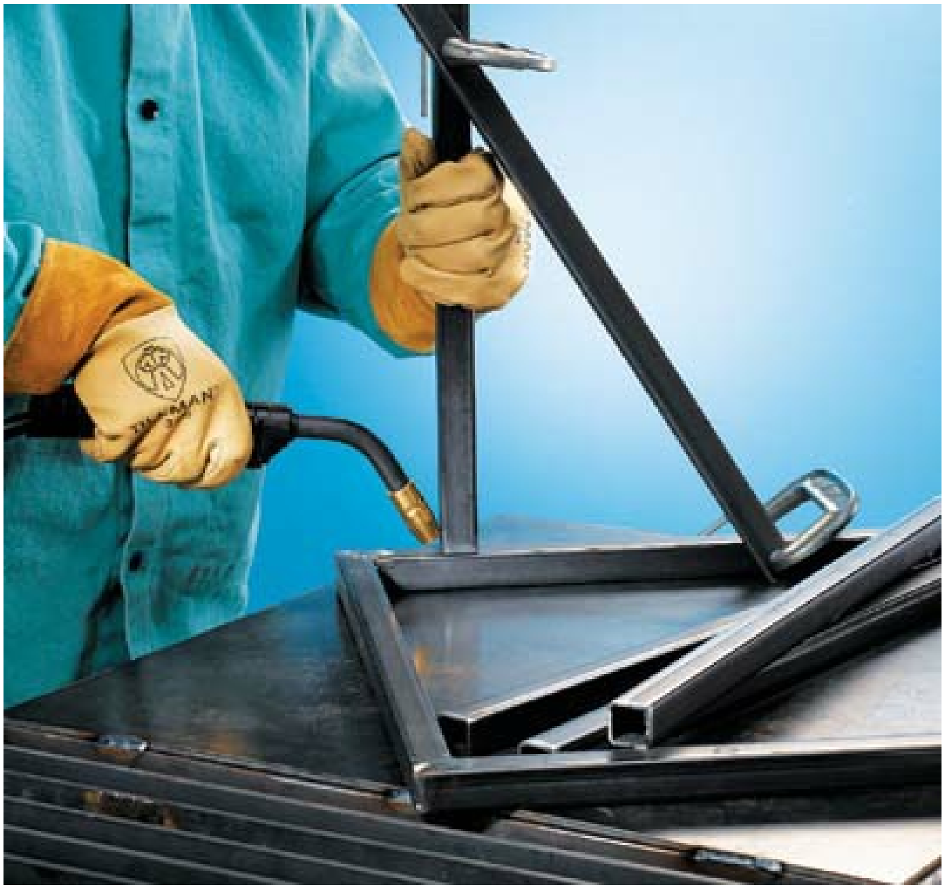

Clamp a leg in place with a piece of angle iron and two C-clamps. Check for square and weld the leg to the table frame.

Make the Table Frame

1. Cut the legs (A) and table frame (B) to size. Miter the frame ends at 45°.

2. On a flat work surface, position two sides of the frame in an L shape. Use a carpenter's square to check for square and then clamp in place. Tack weld the corner. Repeat with the other two sides.

3. Arrange the two Ls to make a square. Clamp in place. Measure across the diagonals to check for square. If the diagonal measurements are not equal, the assembly is not square. Adjust and recheck. Tack weld the corners.

4. After rechecking for square, finish welding all the joints.

5. Grind down the welds to be flush with the surrounding metal.

6. Mark points 3" in from the ends on one frame side. Mark points 3" in from the ends on the opposite frame side.

7. Align a leg with the mark and tack weld into place. Use clamps and an angle iron to securely hold the leg. Repeat with the other three legs.

8. Make sure the legs are perpendicular to the frame and weld in place.

Make the Shelf

1. Before cutting the shelf (C) to size, measure between the outside faces of the front and side legs to make sure this dimension is correct. Cut the shelf to size.

2. Make 1 × 1" cutouts in each corner of the shelf using a plasma cutter, oxyacetylene torch, or bandsaw. Note that the square tube has slightly rounded edges, so round the cutouts to match.

3. Mark each leg at 5" from the bottom.

4. Align the shelf with the marks and tack weld in place. Make sure the shelf is level before finish welding.

Make the Drawer

1. Cut the drawer (D) to size.

2. Use a hand seamer to bend the drawer sides at 3/8" to 90°.

3. Bend the sides and ends of the drawer to 90° to make a box.

4. Weld the seams.

Finish the Drawer

1. Measure the distance between the two front legs. Subtract 1/8" and cut the drawer face (F) to size.

2. With the drawer in the frame, place the drawer face over the drawer front. Align the drawer face with a 1/16" gap between the frame and legs. Placing magnetic right-angle clamps in the drawer makes this task easier.

3. Tack weld the drawer face to the drawer from the inside.

4. Remove the drawer and complete the welds.

5. Center your chosen drawer pull on the drawer face and mark the screw locations. Drill the holes for the pull.

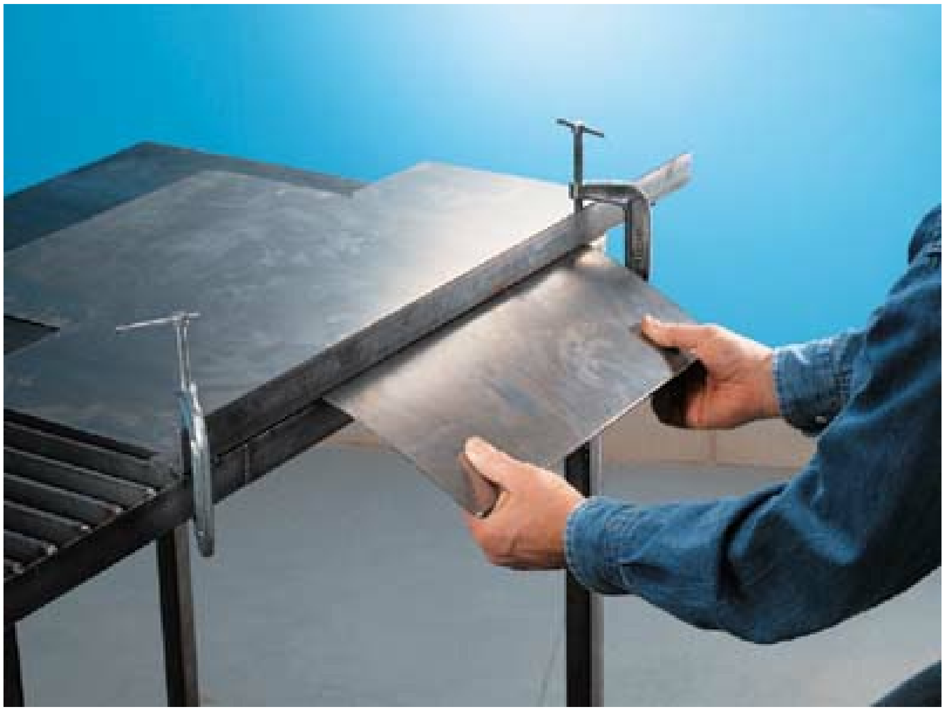

Cut the drawer to size. Clamp the drawer to a sharp-edged surface using C clamps and a piece of angle iron. Bend the sides, back and front of the drawer to 90°.

Add the Sides & Back

1. Before cutting the sides (G) and back (H), measure between the side legs and the back legs to determine the exact lengths. Cut the back and sides to size.

2. Place the sides and back and weld in place.

3. Measure from the top of the shelf to the bottom of the sides and back to determine the correct length for the bars (I). Cut the bars to length.

4. Place three bars evenly spaced (approximately 3" on center) across the back. Make sure the bars are perfectly vertical before welding. Place five bars evenly spaced (approximately 3" on center) across each side. Make sure the bars are perfectly vertical before welding.

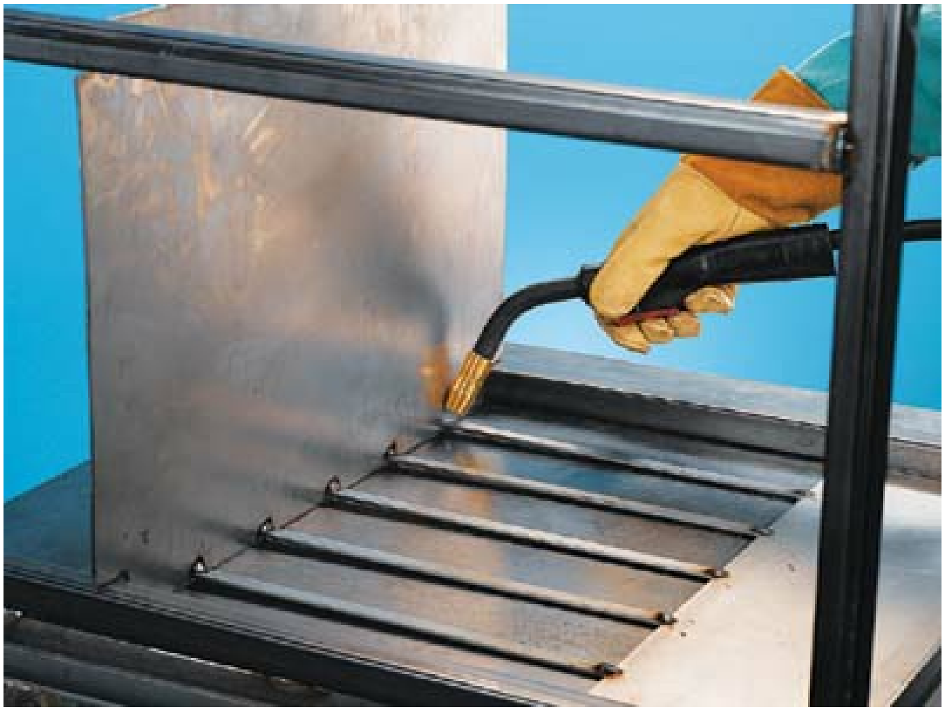

Weld the decorative bars to the table sides and shelf every 3". Weld from the inside.

Finish the Table

1. Measure the top outside dimension of the table frame. Cut the tabletop (J) to size using these dimensions.

2. Attach the tabletop from the underside using two small welds per side. If you prefer a tabletop that appears one with the frame, weld around the entire perimeter of the tabletop and grind the weld smooth.

3. Wire brush, sand, or sandblast the table and drawer. Finish as desired.

4. Attach the drawer pull.

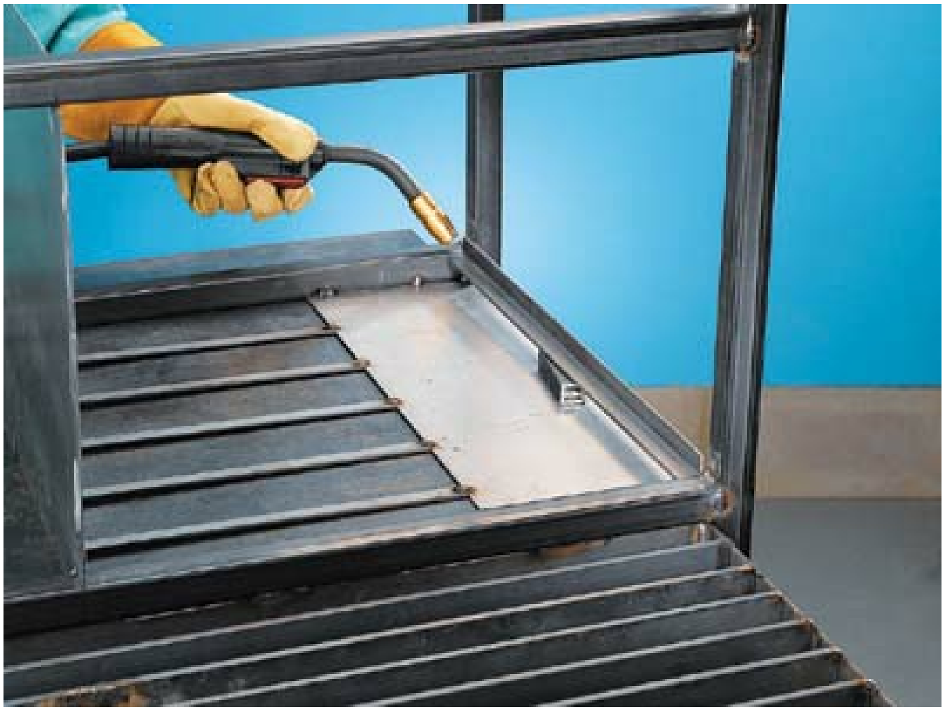

Weld the drawer glides to the legs.

Make the Drawer Glides

1. Cut the glides (E) to size.

2. Mark ¼" down from the top inside of each leg. Align the glides with the lines and tack weld in place.

3. Slide the drawer into the frame and check for fit. Adjust if necessary, and complete the welds.

Fireplace Candelabrum

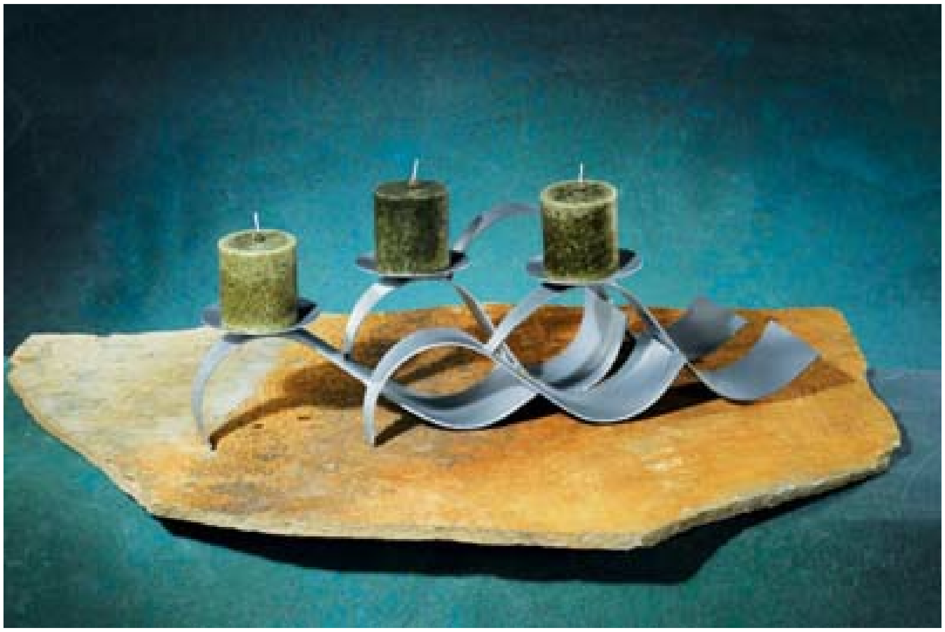

Warm up your hearth, without all the chopping and stacking of wood, with this wavy candelabrum. The simple waves imitate tongues of flames, while the large candleholders support illuminating pillar candles. Romantic, simple, and great for warm summer evenings when a hot fire might be too much.

Materials

- 1/8" sheet metal (13 × 19")
- 4½" bobeches (3, Architectural Iron Designs #79/12)

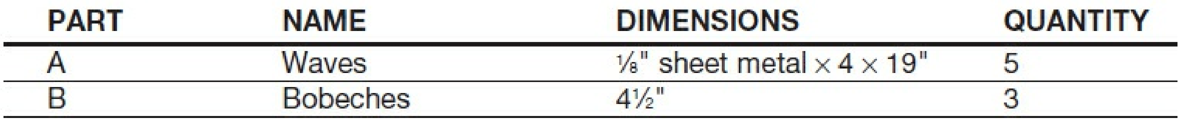

PART	NAME	DIMENSIONS	QUANTITY
A	Waves	⅛" sheet metal × 4 × 19"	5
B	Bobeches	4½"	3

How to Build a Fireplace Candelabrum

Before welding, thoroughly clean all parts with denatured alcohol.

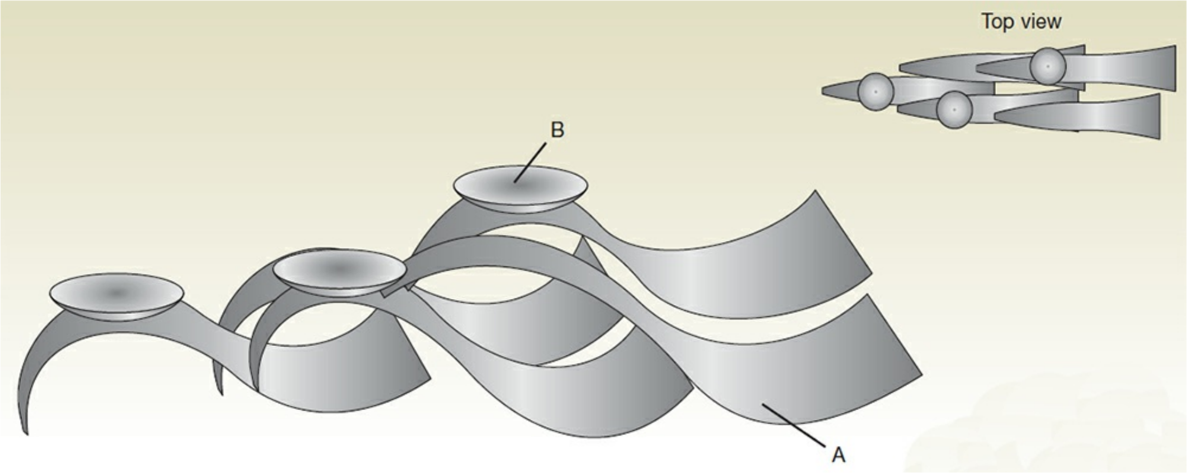

Cutting Diagram

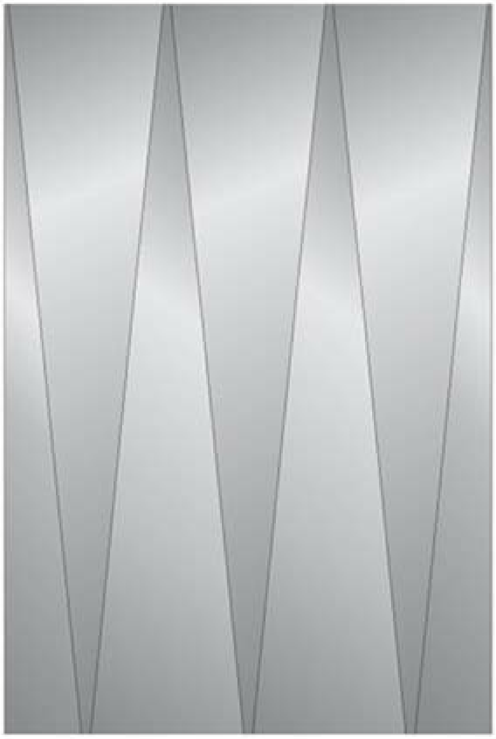

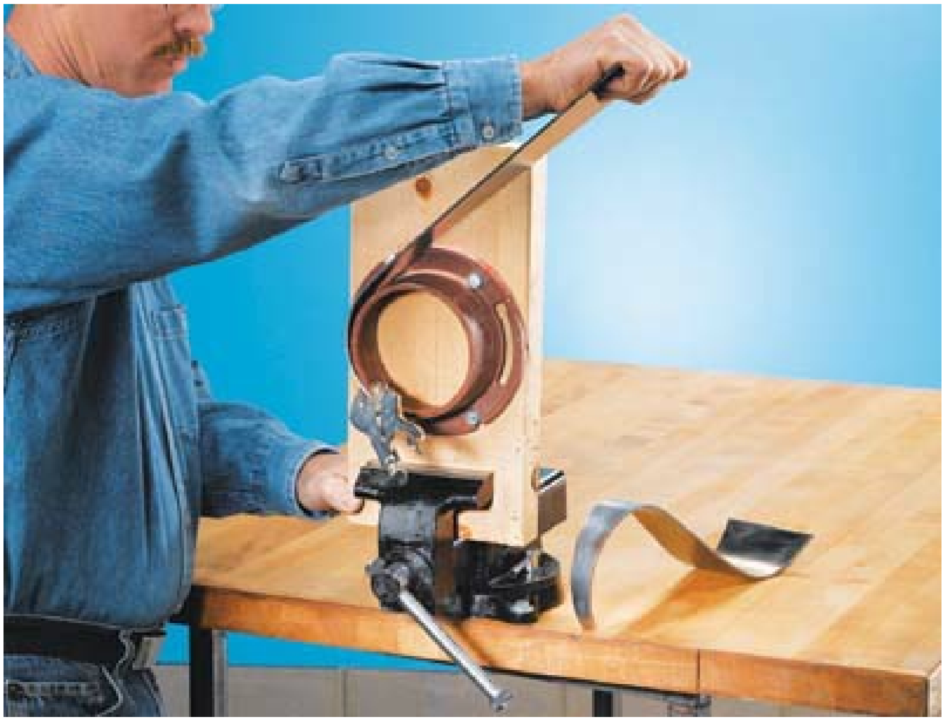

Clamp the narrow wave end to a 4" bending form and bend it slightly past halfway.

Cut the Waves

1. Mark the sheet metal at 4", 4½", 8½", and 9" along one 13" side, and 2", 2½", 6½", 7", 11", and 11½" along the opposite 13" side.

2. Use a straightedge and a permanent marker or soapstone to connect the marks (see diagram).

3. Cut the waves (A) to size using an oxyacetylene torch or plasma cutter.

4. Grind the cut edges smooth, if necessary, and sand or wire brush the metal.

Bend the Waves

1. Clamp the narrow end of a wave to a 4" bending form. Bend the wave around just slightly past halfway (see photo).

2. Clamp the wide end of the wave to a 4" bending form. Wrap the wave around the tube a half turn in the opposite direction from the first wave.

3. Repeat steps 1 and 2 with the other four waves.

Arrange & Finish

1. Arrange the waves as desired. Make sure the arrangement fits in the fireplace.

2. Weld the waves together at the contact points.

3. Place the bobeches on the waves, making sure they are level. Drill a ¼" hole in the center of each bobeche.

4. Weld the bobeches to the waves by welding through the hole. If you choose to use brass bobeches, braze or braze weld the bobeches to the waves.

5. Finish the candelabrum as desired.

Magazine Rack

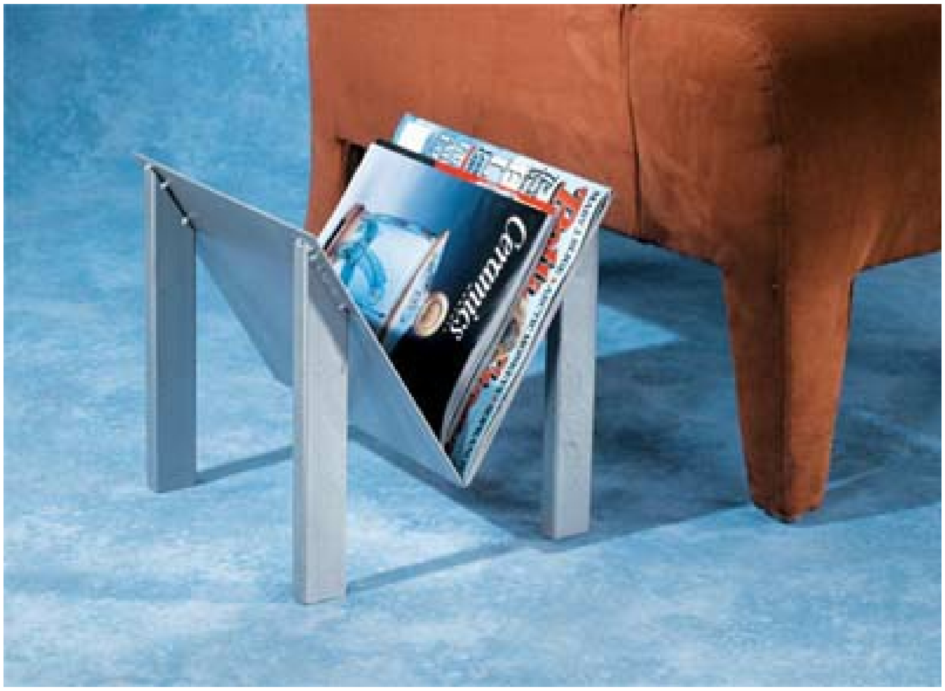

Form follows function with this nifty M design. It’s a magazine rack that’s a snap to put together and perfect for keeping magazines organized and off the floor. You can use virtually any combination of sheet metal and tubing to make this rack. In this project, 16-gauge sheet metal and ¾ x 1¾" rectangular tubing is used. You could also try it with round tubing, hammered tubing, pierced sheet metal, or expanded sheet metal. Make it to fit your favorite magazine, and it will be truly customized.

Materials

- 16-gauge sheet metal (20 × 12" sheet)
- 16-gauge ¾ × 1½" rectangular tube (40")

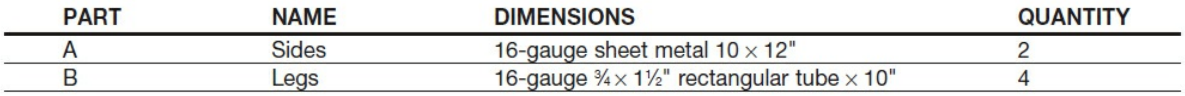

PART	NAME	DIMENSIONS	QUANTITY
A	Sides	16-gauge sheet metal 10 × 12"	2
B	Legs	16-gauge ¾ × 1½" rectangular tube × 10"	4

How to Build a Magazine Rack

Before welding, thoroughly clean all parts with denatured alcohol.

Assemble the Sides

1. Cut the sides (A) to size.

2. Remove slag and grind down any cutting imperfections.

3. Set up the sides in an upsidedown V. The V angle should be between 50° and 60°. Clamp the sides in place and tack weld the sides together (see photo).

4. Weld the sides together with four 1" welds.

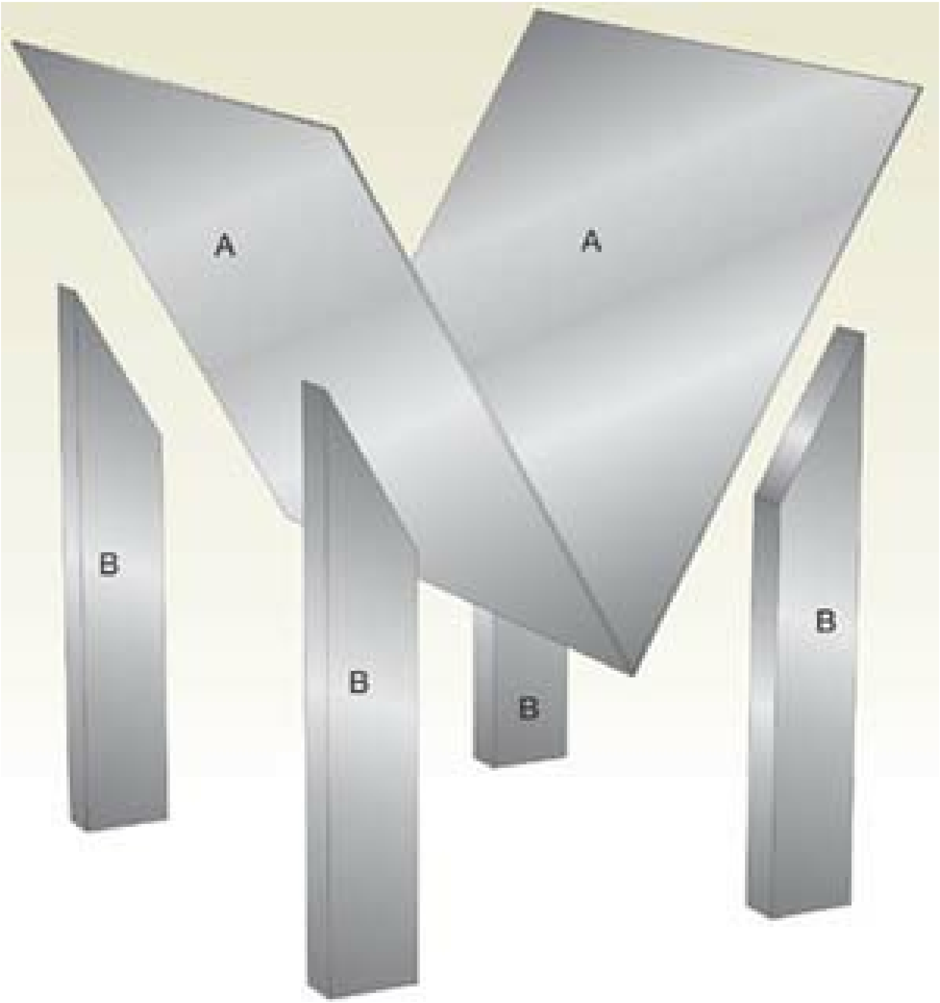
A
A
B
B
B
B

Cut the Legs

1. Cut the legs (B) to length.

2. To get the exact angle measurement for the angle at the top of the legs, trace the outside of the V on a piece of paper. Align a carpenter's square along the tops of the V, with one side of the V contacting the corner of the square. Mark the line for the leg (as though you are making an M). Repeat for the other side.

3. Flip the carpenter's square over and align it with one of the newly drawn lines. Align a leg in the corner of the square. Using a straightedge, mark the leg where it crosses the V. Repeat on the same side. Mark the remaining two legs using the other side of the pattern.

4. Cut the leg angles.

Assemble the Rack

1. Place the V upside down. Place a leg about ½" in from the front or back end of the V. Use a try square to check that the leg is perpendicular and tack weld in place.

2. Repeat step 1 with the other three legs.

3. Turn the rack upright to check for leg alignment. Make adjustments, if necessary, and complete the welds.

4. Grind down welds and spatter. Wire brush the rack and wipe down with denatured alcohol before finishing.

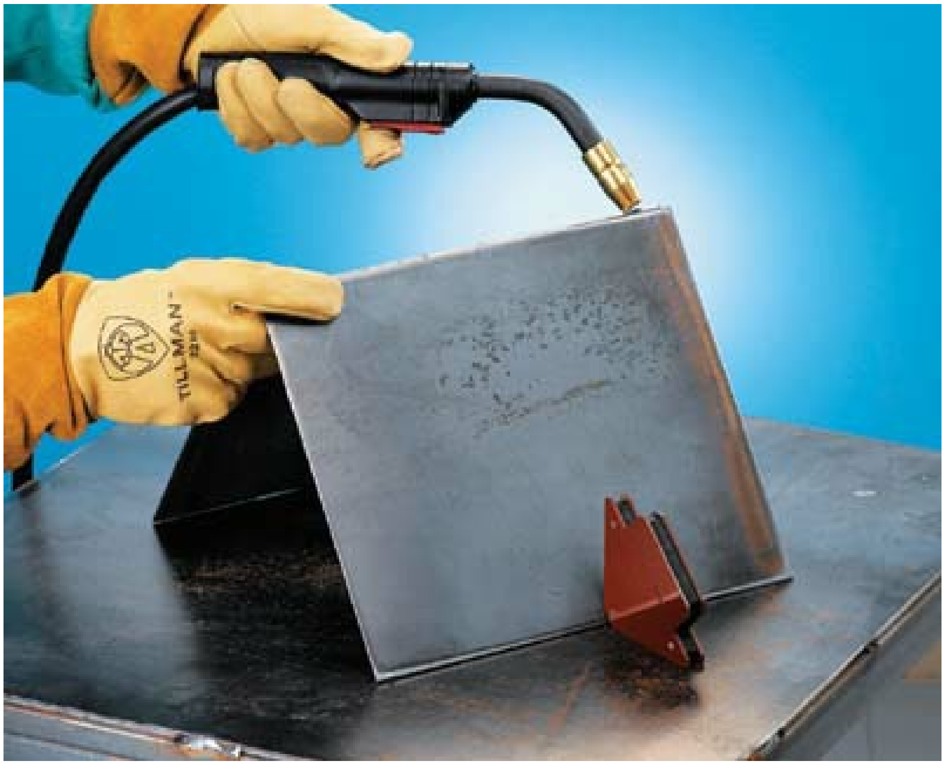

Form a V with rack sides and clamp with magnetic clamps. Weld the seam.

Angled Metal Shelves

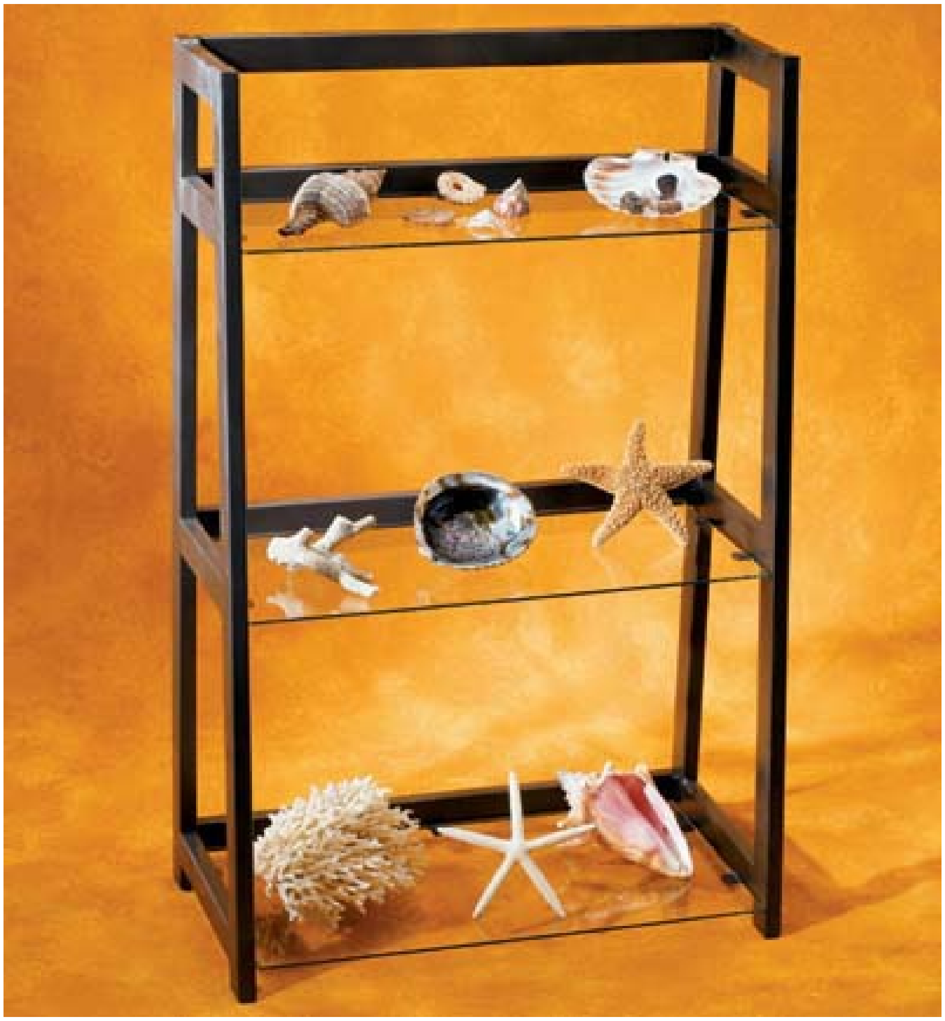

This shelf design is very versatile. The pattern can be altered infinitely—change the width, height, depth, and number of shelves to create a unit to fit in any space and display any item. Dress it up or down by using different materials. Consider using shelves with smooth tubing, a metallic finish, and glass (as we have done here), or use hammered tube, an antique finish, and weathered wood shelves for a rustic look.

Materials

- ¾ × 1½" rectangular tube (24½ feet)
- ⅛ × ¾" flat bar (2½ feet)
- Tempered glass or wood (6 feet)

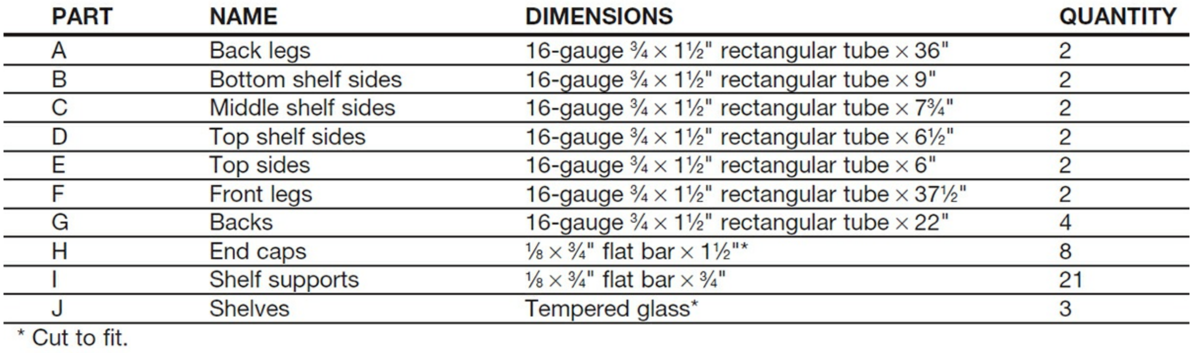

PART	NAME	DIMENSIONS	QUANTITY
A	Back legs	16-gauge ¾ × 1½" rectangular tube × 36"	2
B	Bottom shelf sides	16-gauge ¾ × 1½" rectangular tube × 9"	2
C	Middle shelf sides	16-gauge ¾ × 1½" rectangular tube × 7¾"	2
D	Top shelf sides	16-gauge ¾ × 1½" rectangular tube × 6½"	2
E	Top sides	16-gauge ¾ × 1½" rectangular tube × 6"	2
F	Front legs	16-gauge ¾ × 1½" rectangular tube × 37½"	2
G	Backs	16-gauge ¾ × 1½" rectangular tube × 22"	4
H	End caps	⅛ × ¾" flat bar × 1½"*	8
I	Shelf supports	⅛ × ¾" flat bar × ¾"	21
J	Shelves	Tempered glass*	3

* Cut to fit.

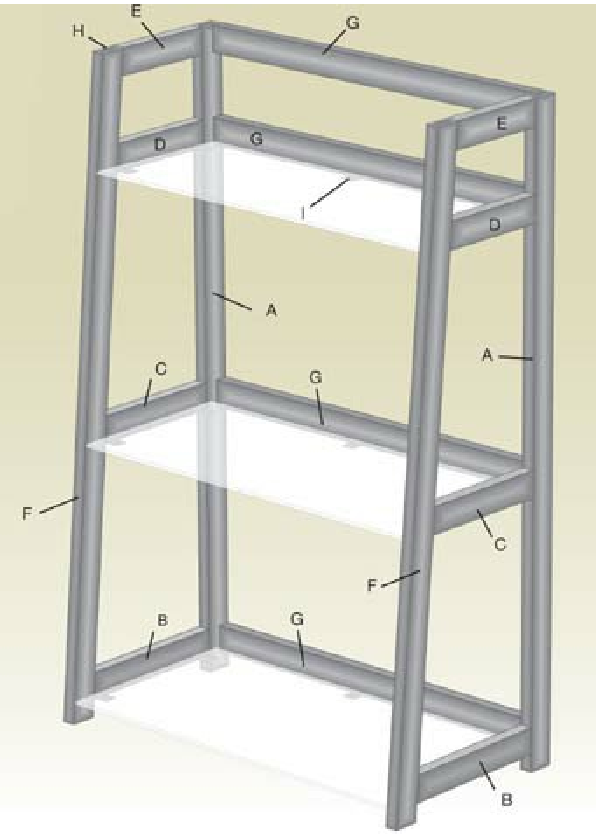
E
H
G
D
G
E
I
D
A
A
C
G
F
C
F
B
G
B

How to Build Angled Metal Shelves

Before welding, thoroughly clean all parts with denatured alcohol.

Make the Sides

1. Cut the back legs (A) to length.

2. Cut the shelf and top sides (B, C, D, E) to length with one end mitered at 83°.

3. Mark the back legs at 1", 7", and 15".

4. Align the flat end of the bottom, middle, and top shelf side pieces with the marks. Align the top side piece with the top of the back leg. Before welding, hold a straightedge against the angled side to check for alignment. Adjust if necessary. Tack weld in place. Repeat with the other back leg and sides.

5. Align a front leg (F) with the side and back leg assembly. Using a carpenter's square held against the back leg, mark the top and bottom angled cuts for the front leg. Repeat with the other front leg.

6. Cut the first front leg to length along the marks.

7. Tack weld the front leg to the side. Repeat with the second front leg.

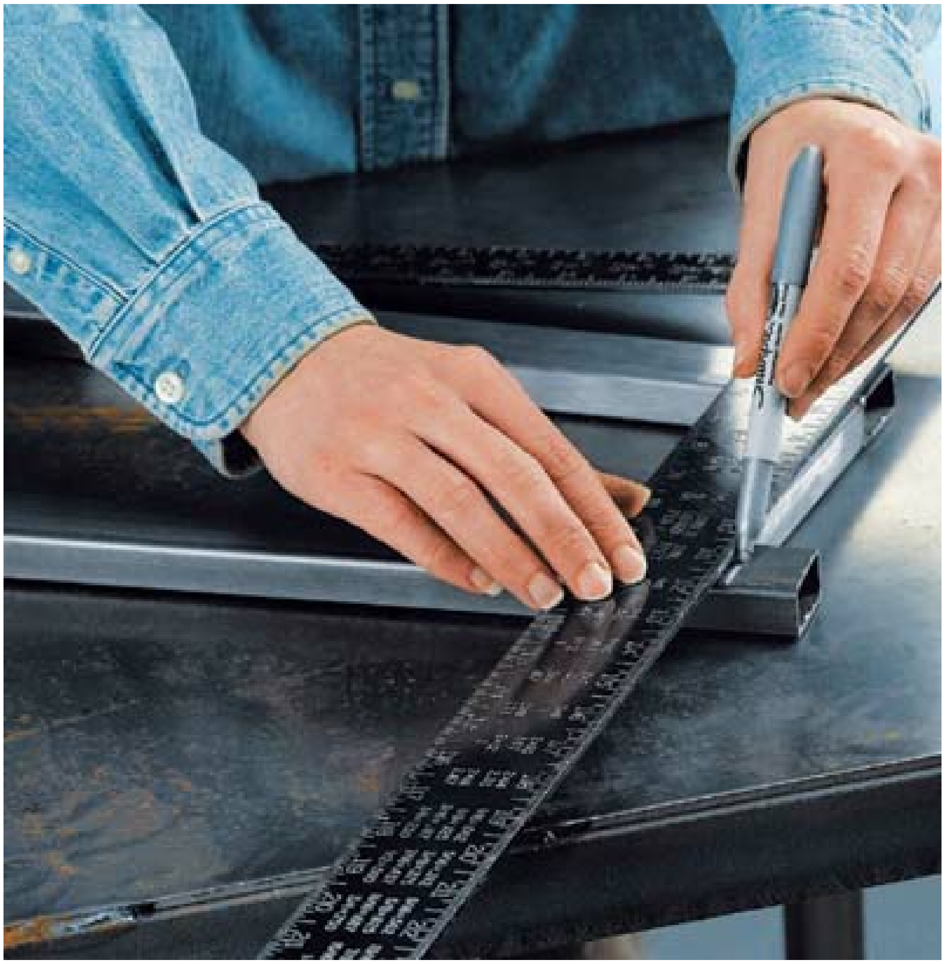

To get the exact angle for cutting the front leg, align the tube with the side pieces and use a carpenter's square to mark a line even with the top of the top side.

Assemble the Shelves

1. Cut the backs (G) to size.

2. Stand a side unit on its back and line up a back piece with the top of the unit. Use a carpenter’s square to check that the back is square to the side. Tack weld in place.

3. Align the second side unit with the back piece and tack weld in place.

4. Repeat with the other three back pieces, aligning them with the top, middle, and bottom side pieces. Check for square and complete all welds.

5. Cut the end caps (H) to fit the tops and bottoms of the front and back legs. (The front pieces will need to be longer due to the angle.) Weld the caps to the leg ends. Grind down all the welds.

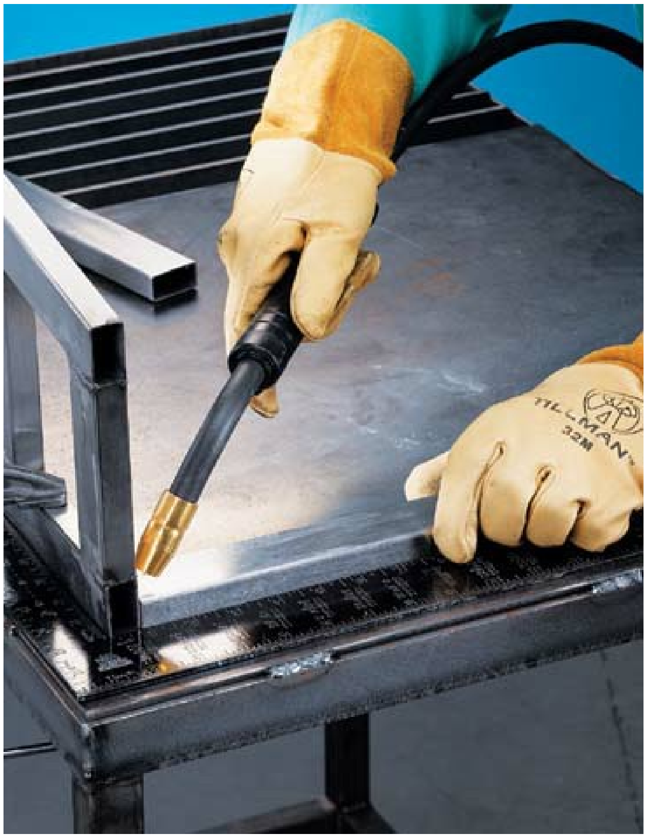
TILLMAN
32M

After assembling the sides, weld the backs in place. Check for square both vertically and horizontally.

Install Shelf Supports & Finish

1. Cut the shelf supports (I) to length.

2. Round the end of each with a bench grinder.

3. Weld two supports to the underside of each top, middle, and bottom shelf side piece. Weld three supports to each back piece. Weld two supports 1" in from the sides and center the third on each back piece.

4. Sand, sandblast, or wire brush the unit. Finish as desired.

5. Measure the dimensions for each shelf and order tempered glass to fit, or cut wooden shelves.

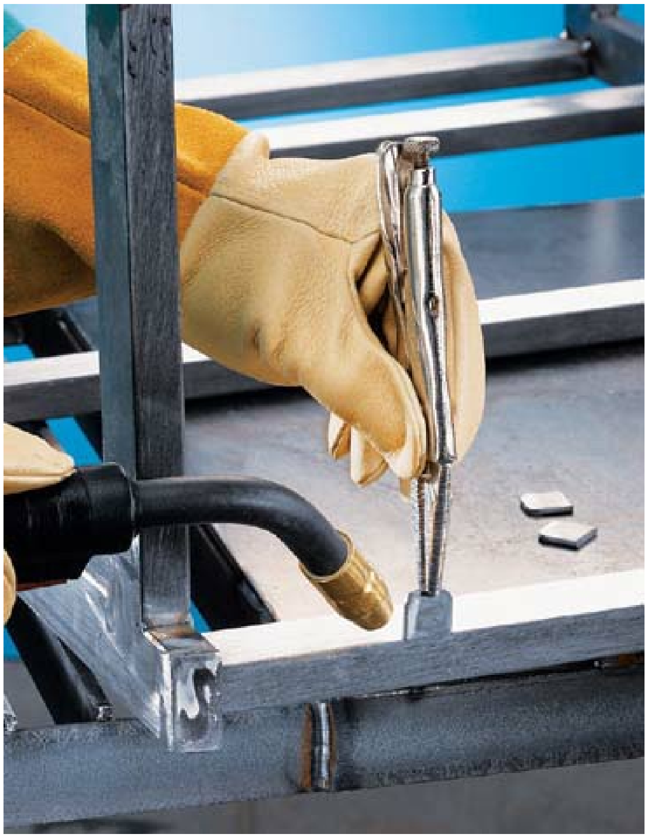

Weld the shelf supports to the bottom of the sides and backs.

Wall-mounted Shelf

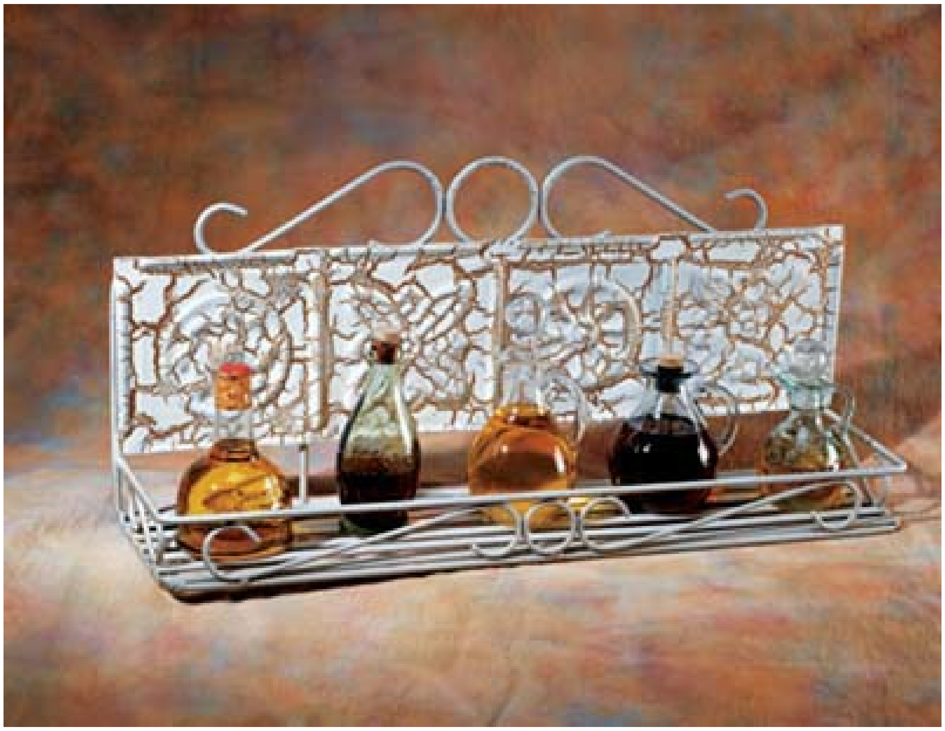

This wall-mounted shelf is perfect for displaying your decorative jars that hold cooking oils or spices. A single 24" square tile used for ceilings is cut to size at 24 × 6", which gives the appearance of four separate 6 × 6" tiles. A variety of tin tile sources are available on the Internet: see Resources on page 220 for addresses. You may also find ceiling tiles at salvage stores and antique stores—and they already have the antique character. The tiles don’t need to be welding-quality steel; installing them with two-part epoxy is sufficient.

Materials

- ¼" round rod (18½ feet)
- 1/8 × ½" flat bar (12 feet)
- 24" metal ceiling tile (1, M-BOSS #0613)
- 3/16" round rod (5 feet)

PART	NAME	DIMENSIONS	QUANTITY
A	Frame	¼" round rod × 64"	1
B	Tile supports	¼" round rod × 8"	3
C	Shelf frames	¼" round rod × 36"	2
D	Shelf supports	¼" round rod × 6"	1
E	Shelf cross pieces	⅛ × ½" flat bar × 24"	6
F	Top scrolls	¼" round rod × 15"	2
G	Top circle	¼" round rod × 12"	1
H	Side shelf scrolls	3⁄16" round rod × 9"	2
I	Front shelf scroll	3⁄16" round rod × 15"	2
J	Shelf circle	3⁄16" round rod × 10"	1
K	Tin tile	24" × 6"	1

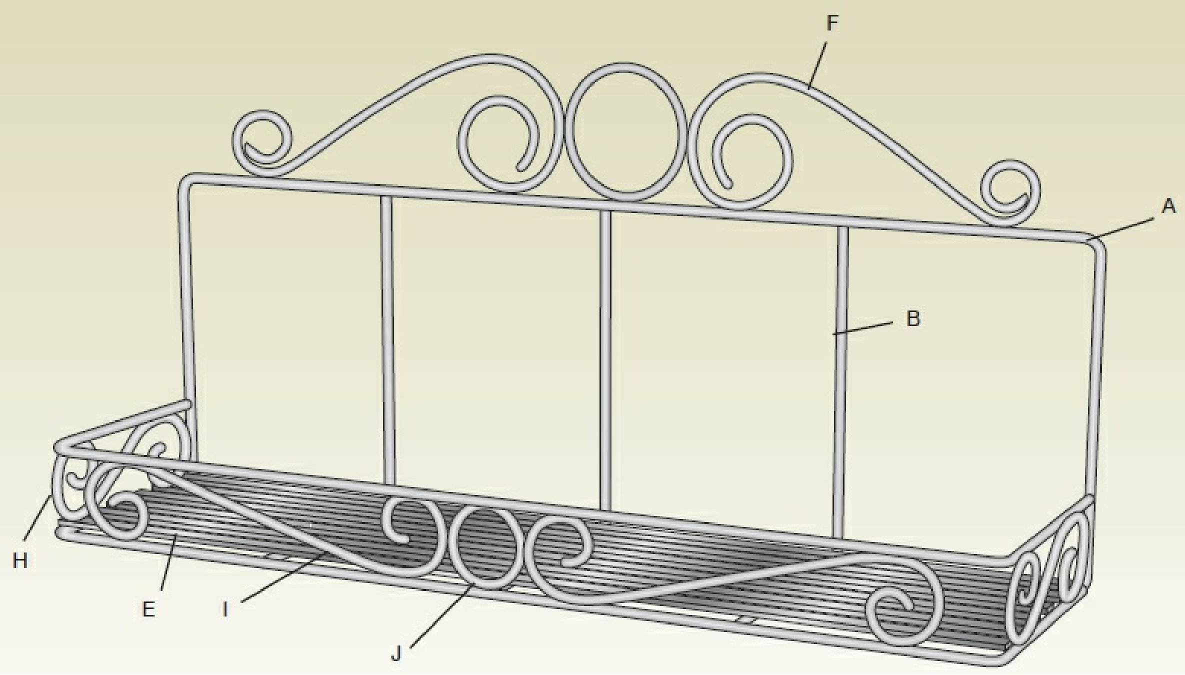
F
A
B
H
E
I
J

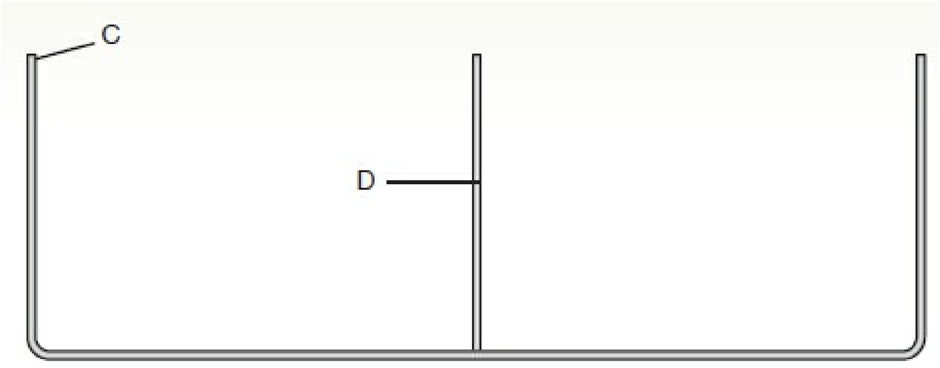

Bottom shelf

How to Build a Wall-mounted Shelf

Before welding, thoroughly clean all parts with denatured alcohol.

Create the Frame

1. Cut the frame (A) to length.

2. Mark the frame at 12", 19¾", 43½", and 52¼". Create a rectangle by making 90° bends at the marks. Weld the ends together.

3. Cut the tile supports (B) to length.

4. Weld the tile supports into the frame at 6" intervals.

Create the Shelf

1. Cut the shelf frames (C) to length.

2. Mark each of the shelf frames at 6" in from each end. Make 90° bends at the marks to create the shelf frames.

3. Weld one shelf frame to the base of the frame at a 90° angle.

4. Cut the shelf supports (D) and cross pieces (E) to length.

5. Weld the first shelf support to the installed shelf frame at the midpoint. Weld the remaining two supports centered between an outside edge of the shelf frame and the center shelf support.

6. Space the shelf cross pieces at ½" intervals across the supports. Weld in place.

7. Weld the second shelf frame to the frame 2" above the shelf at a 90° angle.

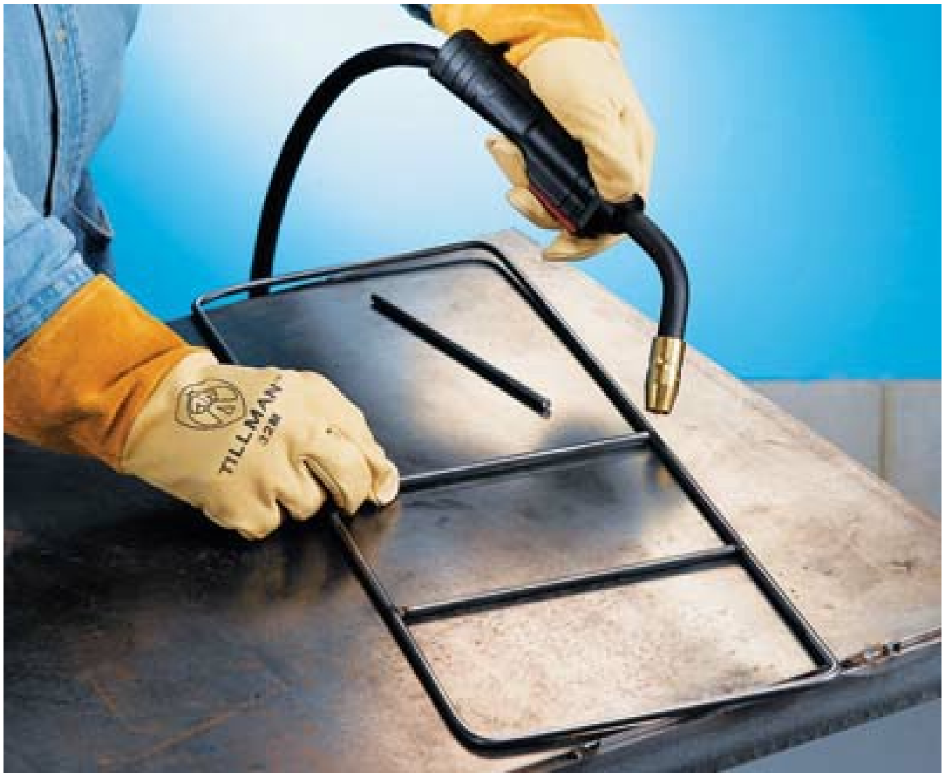
TILLMAN
328

Weld the tile supports to the frame at 6" intervals.

Make the Scrolls

1. Cut the top (F), side (H), and front (I) scrolls to length.

2. Clamp the top scroll rod to a 2" pipe. Bend the rod one full turn around the pipe. Clamp the other end around a 1" pipe and bend the rod one full turn around the pipe in the opposite direction from the first bend. Repeat with the second top scroll, making sure the scrolls match.

3. Clamp the end of a front shelf scroll to a 1" pipe. Bend the rod one full turn around the pipe. Clamp the other end to the pipe and bend the rod one full turn around the pipe in the opposite direction. Repeat with the second front scroll.

4. Clamp the end of a side scroll to a 1" pipe. Bend the rod one turn around the pipe. Clamp the other end to a 1" pipe and bend the rod one turn around the pipe in the opposite direction. Repeat with the second side scroll.

5. Bend the top circle (G) around a 2" pipe, overlapping the ends. Cut the overlap off and weld the circle ends together.

6. Bend the shelf circle (J) around a 1" pipe, overlapping the ends. Cut the overlap off and weld the circle ends together.

7. Place the shelf unit on its back and align the top scrolls and circle along the top. Weld at the contact points.

8. Align the front scrolls and circle in between the shelf and the shelf support. Weld at the contact points. Repeat with each of the side scrolls.

TILLMAN
32M

Weld the top shelf frame to the frame 2" above the shelf.

Finish the Shelf

1. Cut the tile down to a 6 × 24" strip, or 6" squares, depending on the tile and pattern you have chosen.

2. Use two-part epoxy to glue the tiles to the frame.

3. Sand, wire brush, or sandblast the shelf.

4. Finish as desired.

TILLMAN
1350

Weld the decorative scrolls and circles to the frames.

Étagère

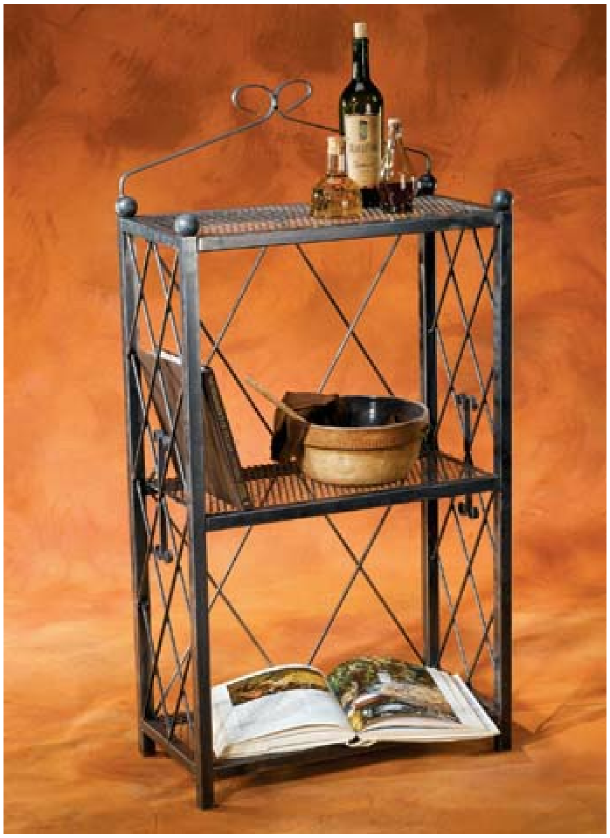

Étagère is a French term for a metal structure used to display items in the house, or plants in the garden. Étagères can be shockingly expensive, but this project is not only affordable, but it is much sturdier than many available in stores. The corner Étagère that follows (on page 160) is created to match the style of this project.

This decorated shelf may be used inside or outside the home. Paint it or allow it to gently rust. The decorative elements are not structural, so you may change them to suit your desired level of fuss. The flat expanded sheet metal was chosen because it repeats the diamond pattern, but anything from metal to wood to tile can be used as shelving material. Check out the decorative iron suppliers listed in the Resources section on page 220 for dozens of friezes and metal stampings that may be used instead of the diagonals and scrolls.

Materials

- 16-gauge 1 × 1" square tube (23 feet)

- 16-gauge ½" flat expanded sheet metal (2 ft. × 3 ft.)
- ¼" round rod (31 feet)
- 1/8 × ½" flat bar (5 feet)
- 1½" balls (4, Triple S Steel #SF116F4)
- 1 × 1" × 1/8" angle iron (6 feet)

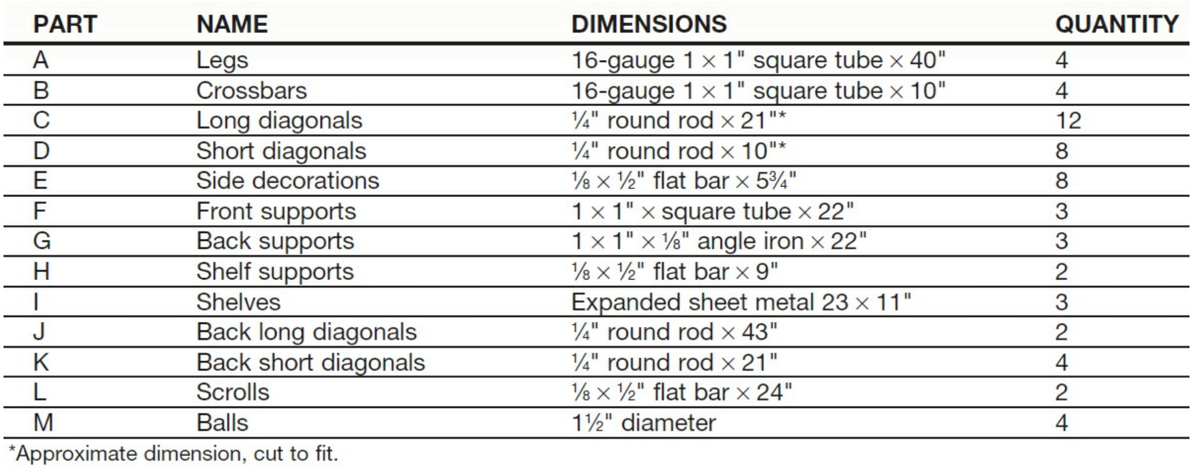

PART	NAME	DIMENSIONS	QUANTITY
A	Legs	16-gauge 1 × 1" square tube × 40"	4
B	Crossbars	16-gauge 1 × 1" square tube × 10"	4
C	Long diagonals	¼" round rod × 21"*	12
D	Short diagonals	¼" round rod × 10"*	8
E	Side decorations	⅛ × ½" flat bar × 5¾"	8
F	Front supports	1 × 1" × square tube × 22"	3
G	Back supports	1 × 1" × ⅛" angle iron × 22"	3
H	Shelf supports	⅛ × ½" flat bar × 9"	2
I	Shelves	Expanded sheet metal 23 × 11"	3
J	Back long diagonals	¼" round rod × 43"	2
K	Back short diagonals	¼" round rod × 21"	4
L	Scrolls	⅛ × ½" flat bar × 24"	2
M	Balls	1½" diameter	4

*Approximate dimension, cut to fit.

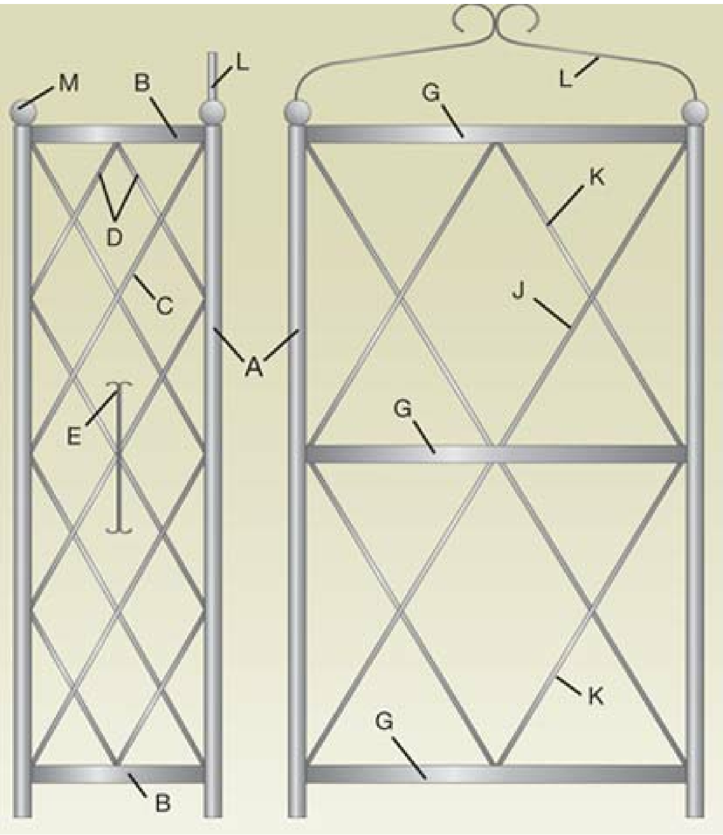
M
B
L
G
L
K
D
C
J
A
E
G
K
G
B

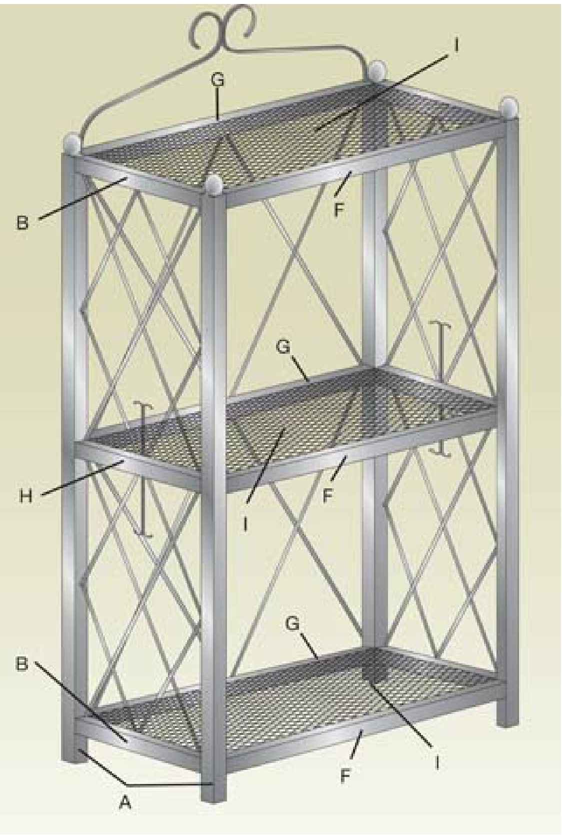
I
G
B
F
G
H
F
I
G
B
I
F
A

How to Build an Étagère

Before welding, thoroughly clean all parts with denatured alcohol.

Make the Sides

1. Cut the legs (A) and crossbars (B) to length.

2. Mark each leg 2" from one end. This is the bottom of the leg.

3. On a flat work surface, lay out two legs and two crossbars. Align the bottom of one crossbar with the marks. Align the top of the other crossbar with the top of the legs. Clamp in place.

4. Check all corners for square. Tack weld all corners. Recheck for square by measuring across the diagonals. If the two measurements are equal, the unit is square. If the measurements are not equal, make adjustments until they are. Complete the welds.

5. Repeat steps 3 and 4 to make the second side panel.

Cut the Decorative Side Diagonals

1. Find the midpoint of the top and bottom crossbars and mark.

2. Find the midpoint of the front and back legs between the crossbars and mark.

3. Mark the points halfway between the midpoints and the crossbars. These are the quarterpoints.

4. Determine the lengths for the long diagonals (C) by measuring from a corner to the opposite midpoint and from an upper quarterpoint to a lower, opposite quarterpoint. These measurements should be equal if the panel is square. Cut the long diagonals slightly longer than this measurement.

5. Determine the lengths for the short diagonals (D) by measuring from a quarterpoint to the adjoining crossbar midpoint. Cut the short diagonals slightly longer than this measurement.

6. Hold the diagonals in place against the side panels and mark the necessary cutting angle. Cut at the marks.

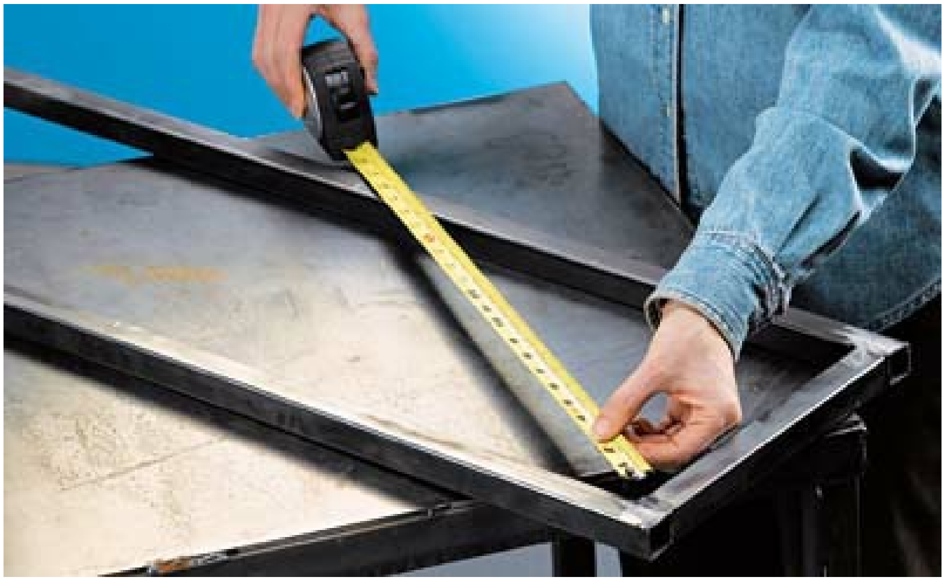

Measure between a corner and the opposite midpoint to find the length for the long diagonals.

Assemble the Decorative Side Diagonals

1. With the side panel on a flat surface, lay the three long and two short diagonals that slant down from upper left to lower right in place. Weld to the frame.

2. Lay the three long and two short diagonals that slant from upper right to lower left in place on top of the other set of diagonals. Weld in place.

3. Cut the side decorations (E) to length. Round one end of each to a smooth semicircle using a bench grinder.

4. Clamp the rounded end to a ½" pipe and bend the bar a half turn around the tube. Repeat with all the pieces.

5. Place two pieces back to back and weld together. Repeat with the other pieces.

6. Grind the flat ends of the decoration to a V.

7. Place one welded piece vertically at the middle of the diagonal pattern and weld in place. Place the second welded piece upside down at the middle and weld in place.

8. Cut and assemble the diagonals for the second side panel.

Weld the first layer of diagonals in place.

Assemble the Étagère

1. Cut the front supports (F) and back supports (G) to length.

2. The decorations on the side panels should be flush on one side and recessed on the other. The flush side is the outside. Place the two sides parallel to each other lying on their backs. Align the front supports with the top and bottom crossbars and centered over the midpoint. Check for square and tack weld in place.

3. Orient the top back support so the flat faces are to the top and front and aligned with the top crossbar. Tack weld in place.

4. Orient the midpoint and bottom back supports with the flat faces to the top and back. Align with the front of the back leg. (This will allow the back panel decorations to be recessed).

5. Cut the shelf supports (H) to length.

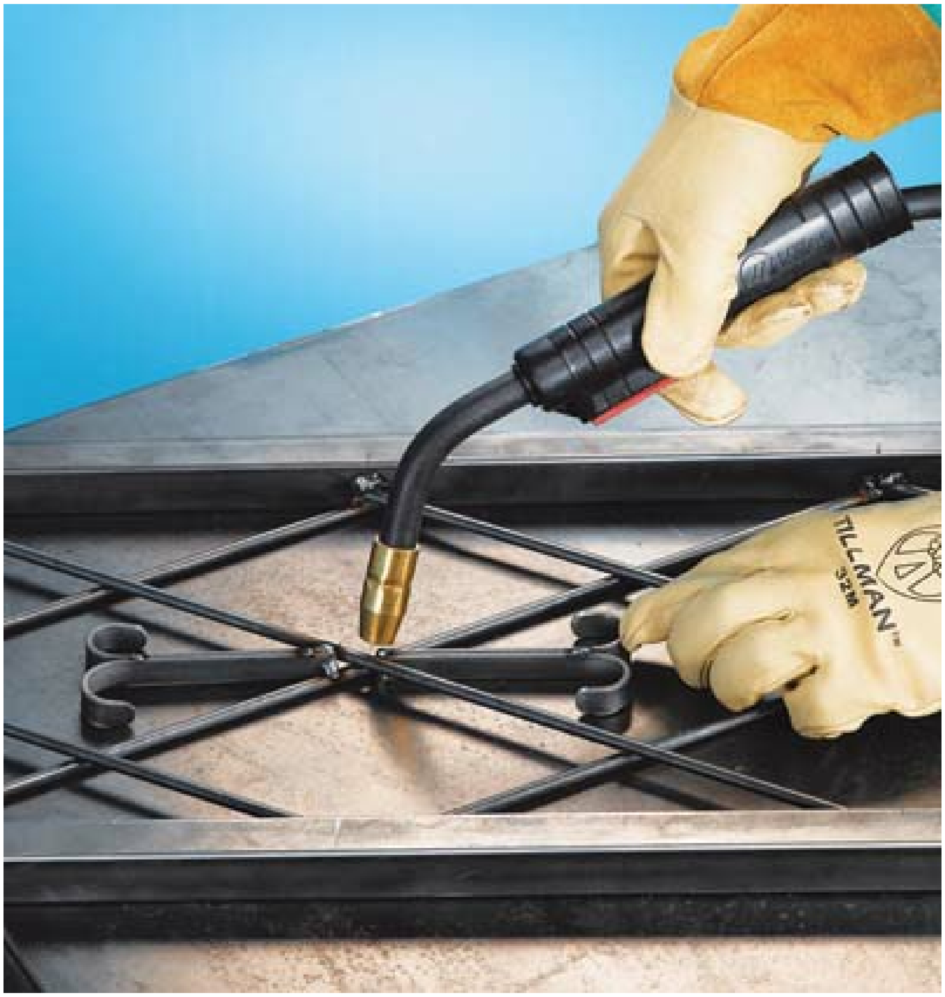
TILLMAN™
32M

Weld the decorations to the middle X of the diagonals.

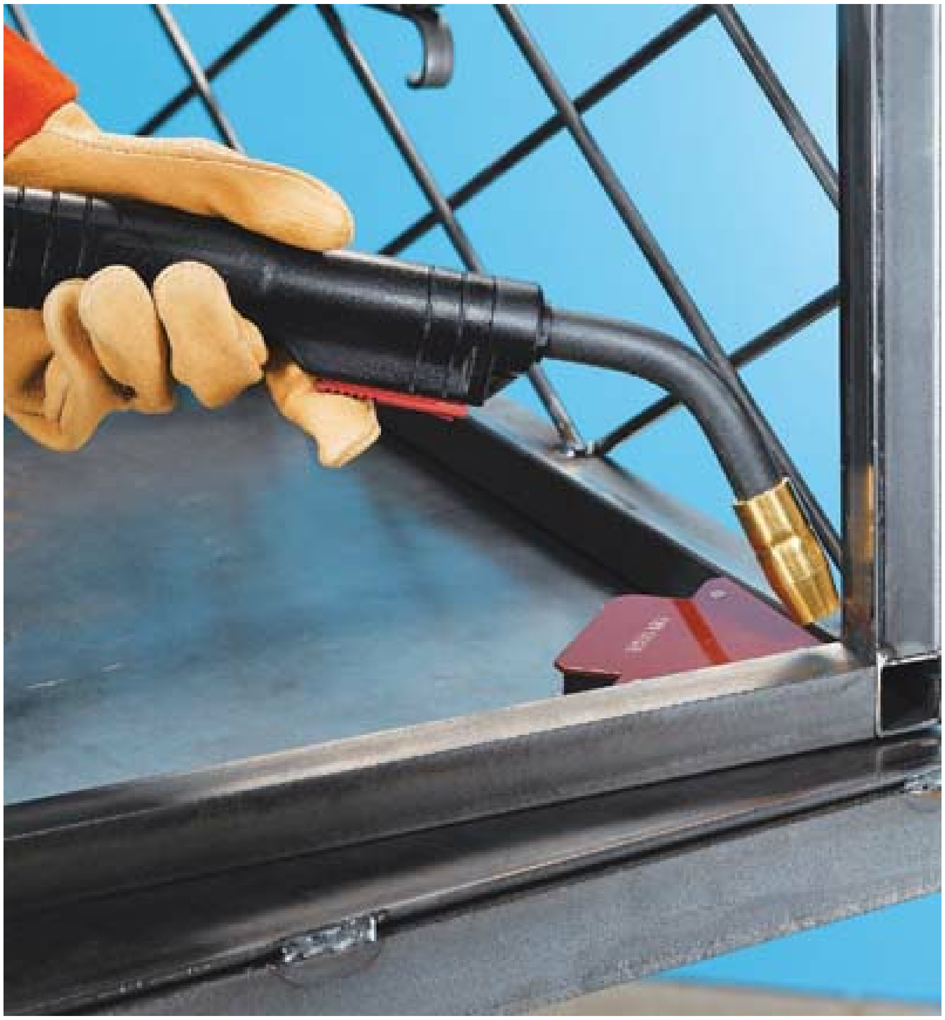

Tack weld the front supports in place and check for square.

6. Weld the shelf supports flush with the top of the middle shelf frames at the ends. Bend the top shelf to fit inside the four supports. Weld in place.

7. Cut the shelves (I) to size.

8. Bend the front of the middle shelf to fit behind the front support. The shelf lies on top of the shelf suport and back support. Weld in place.

9. Bend the front and sides of the bottom shelf to fit between the supports. The back lies on top of the back support. Weld in place.

Make the Back Panel Decorations

1. Measure the diagonal across the back of the shelves from corner to corner. Cut the back long diagonals (J) to this length.

2. Measure from the midpoint of the top back shelf support to the midpoint of the leg panel. Cut the short diagonals (K) to this length.

3. Place the left to right descending diagonals in place and weld. Place the right to left diagonals and weld.

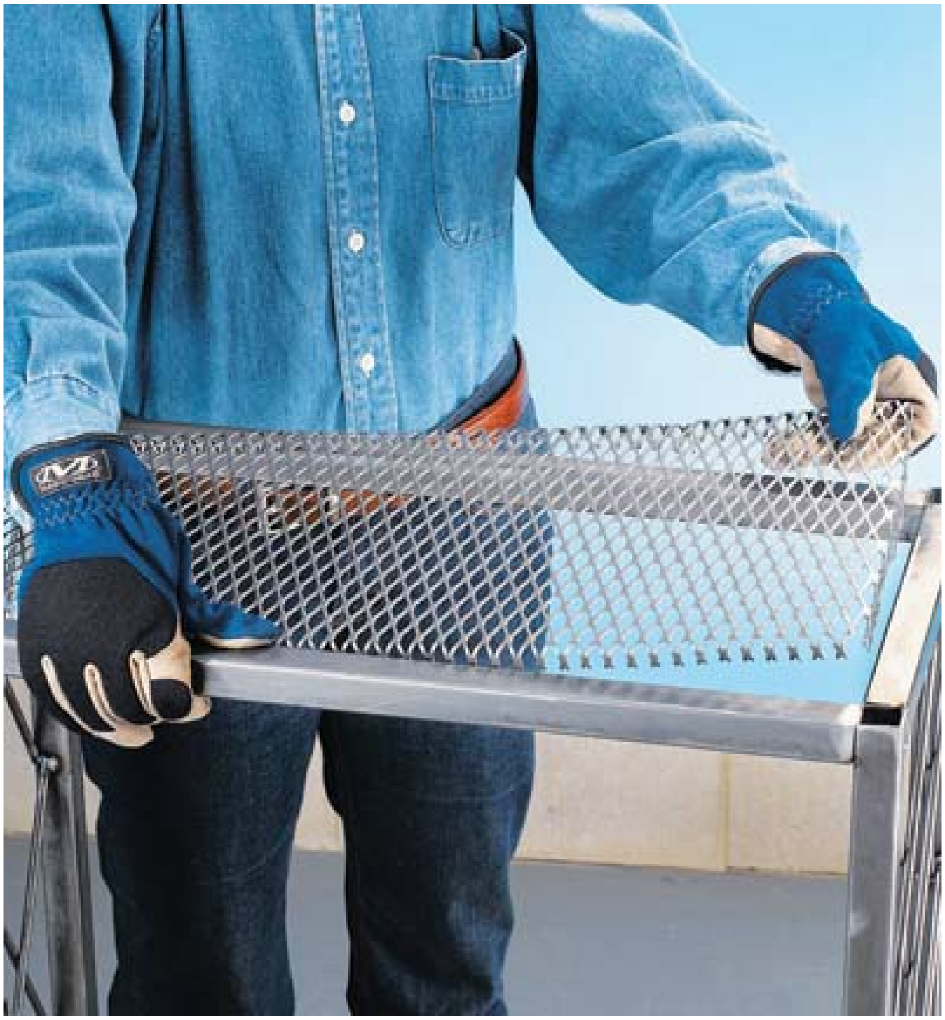

Bend the top shelf to fit inside the sides and supports.

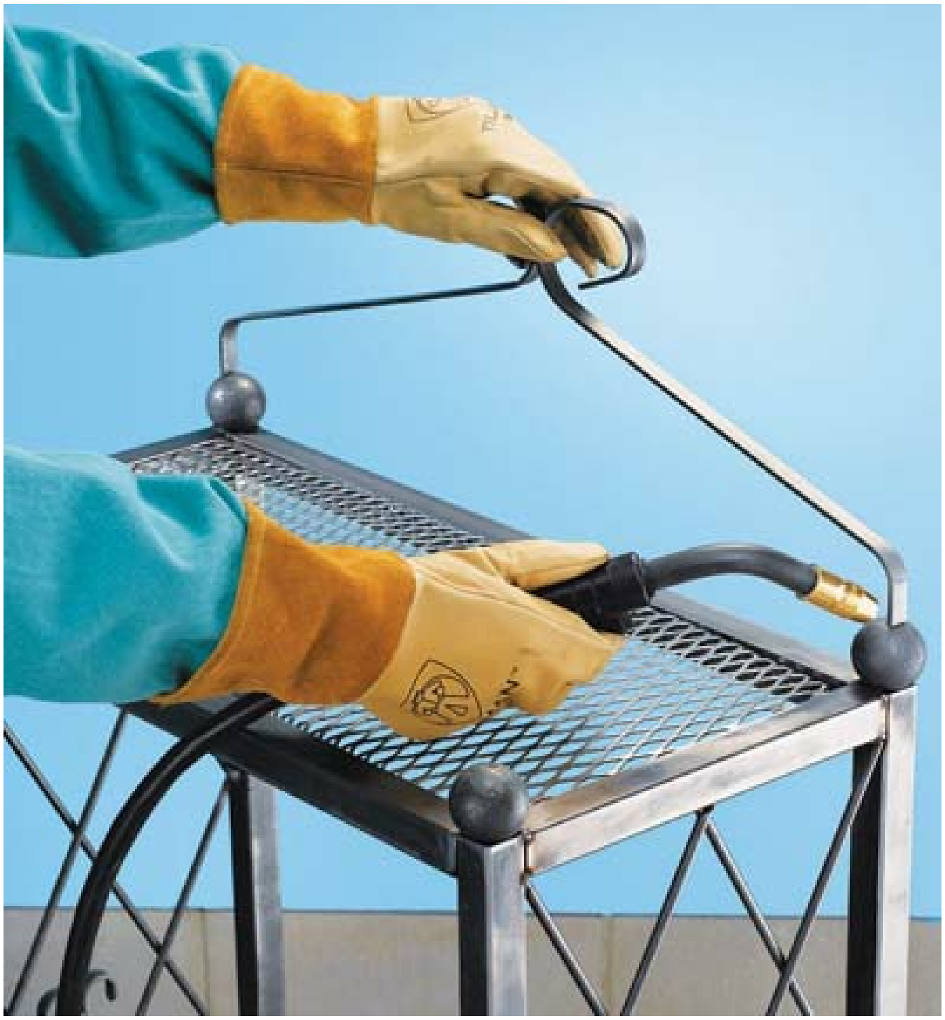

Weld the spheres to the tops of the legs and weld the scroll to the tops of the rear spheres.

Make the Decorative Scrolls

1. Cut the scrolls (L) to length.

2. Round one end of each scroll to a smooth semicircle using a bench grinder.

3. Clamp the flat end to a 4" pipe and bend slightly past a one-quarter turn to create a rounded 90° angle. Repeat with the second scroll.

4. Cut a 1/8" notch in a 2" pipe.

5. Insert the rounded end of the scroll through both notches. Bend the scroll a three-quarter turn around the pipe. Repeat with the second scroll.

6. Weld the balls (M) into the tops of the legs.

7. Align the two scrolls so they meet in the middle and are symmetrical. Weld at the contact points.

8. Weld the scrolls to the top of the rear balls.

9. Sand, wire brush, or sandblast the shelf unit. Finish as desired.

Tip

Welding thick material to thin material.

Welding solid spheres to 16-gauge material won’t be successful without an extra step. The spheres will not weld easily because they absorb a great deal of heat before melting. Heating them first with a propane torch assures a good weld.

Corner Étagère

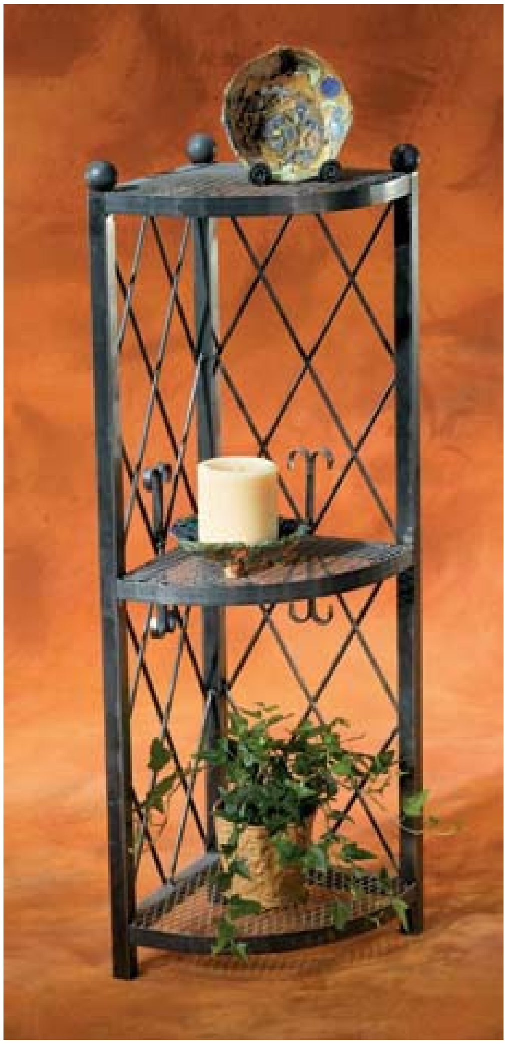

Why have an empty corner when it can display a beautiful shelf? Designed to match the étagère, this shelf unit is perfect for displaying plants or other decorative curios. You don’t need to hide it in a corner—it can stand on its own in your backyard, or you can build two to create a semi-circle shelf to stand out along a wall.

The arcing shelf fronts look fancy but are actually fairly simple to execute. Face the front with a decorative frieze or metal stampings for a different look. The diagonals do not provide structural support, so you may personalize the sides in any way you choose.

Materials

- 16-gauge 1 × 1" square tube (10 feet)
- ¼" round rod (15½ feet)
- 1/8 × 1" flat bar (4½ feet)
- 16-gauge ½" flat expanded sheet metal (2 ft. × 2 ft.)
- 1½" balls (3, Triple S Steel #SF116F4)
- 1 × 1 × 1/8" angle iron (20")

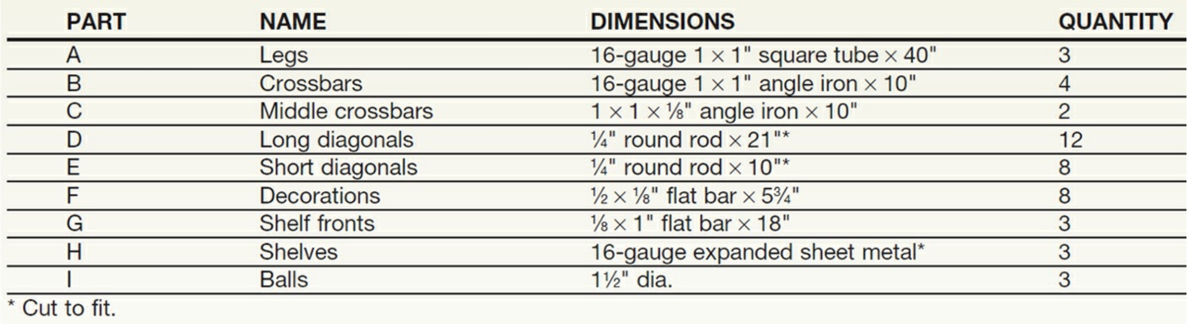

PART	NAME	DIMENSIONS	QUANTITY
A	Legs	16-gauge 1 × 1" square tube × 40"	3
B	Crossbars	16-gauge 1 × 1" angle iron × 10"	4
C	Middle crossbars	1 × 1 × ⅛" angle iron × 10"	2
D	Long diagonals	¼" round rod × 21"*	12
E	Short diagonals	¼" round rod × 10"*	8
F	Decorations	½ × ⅛" flat bar × 5¾"	8
G	Shelf fronts	⅛ × 1" flat bar × 18"	3
H	Shelves	16-gauge expanded sheet metal*	3
I	Balls	1½" dia.	3

* Cut to fit.

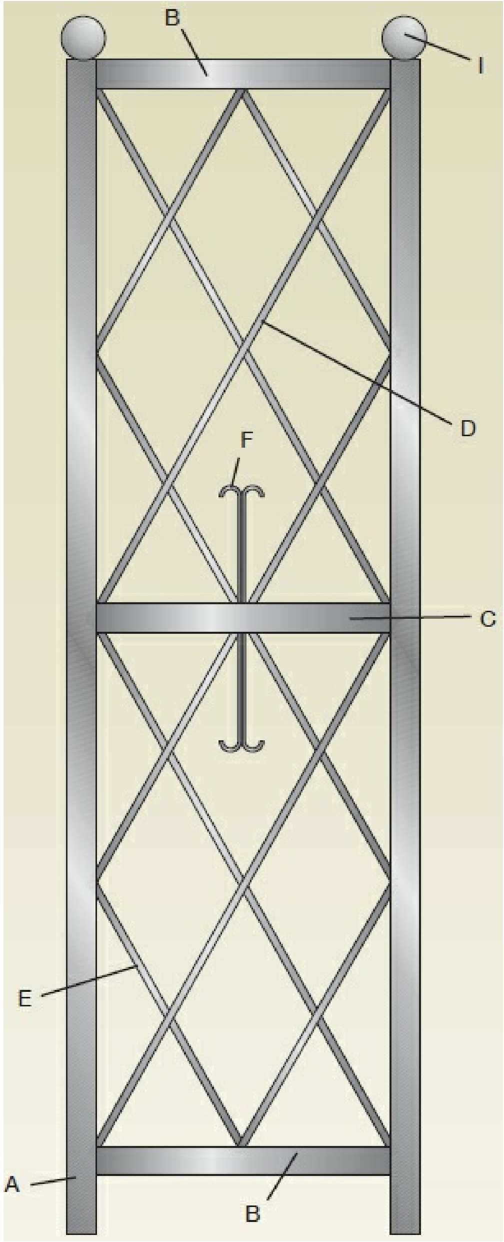
B
I
D
F
C
E
A
B

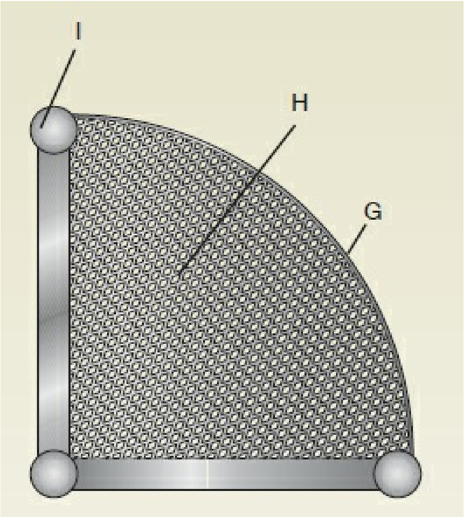
I
H
G

How to Build a Corner Étagère

Before welding, thoroughly clean all parts with denatured alcohol.

Make the Frame

1. Cut the legs (A) and crossbars (B) to length.

2. Measure 2" up from the ends of the legs and make a mark.

3. Align one crossbar in between and flush with the tops of two legs. Check for square and tack weld in place. Align the bottom of a second crossbar with the end marks, check for square and tack weld in place.

4. Measure across the diagonals of this unit to check for square. The measurements should be equal; if not, adjust until they are. Weld in place.

5. Repeat steps 3 to 4, attaching the third leg and the crosspieces to the leg unit at right angles to form a triangular frame.

6. Cut the middle crossbars (C) to length, mitering one end at 45°. Center the crossbars over the midpoints and even with the insides of the legs. Weld in place.

Cut the Decorative Diagonals

1. Find the midpoint of the top and bottom crossbars and mark.

2. Find the midpoint of the front and back legs between the crossbars and mark.

3. Mark the points halfway between the midpoints and the crossbars. These are the quarterpoints.

4. Determine the lengths for the long diagonals (D) by measuring from a corner to the opposite midpoint and from an upper quarterpoint to a lower, opposite quarterpoint. Cut the long diagonals to length. Place the diagonals and mark the angled cut necessary for an exact fit.

5. Determine the lengths for the short diagonals (E) by measuring from a quarterpoint to the adjoining crossbar midpoint. Cut the short diagonals to length. Cut the long diagonals to length. Place the diagonals and mark the angled cut necessary for an exact fit.

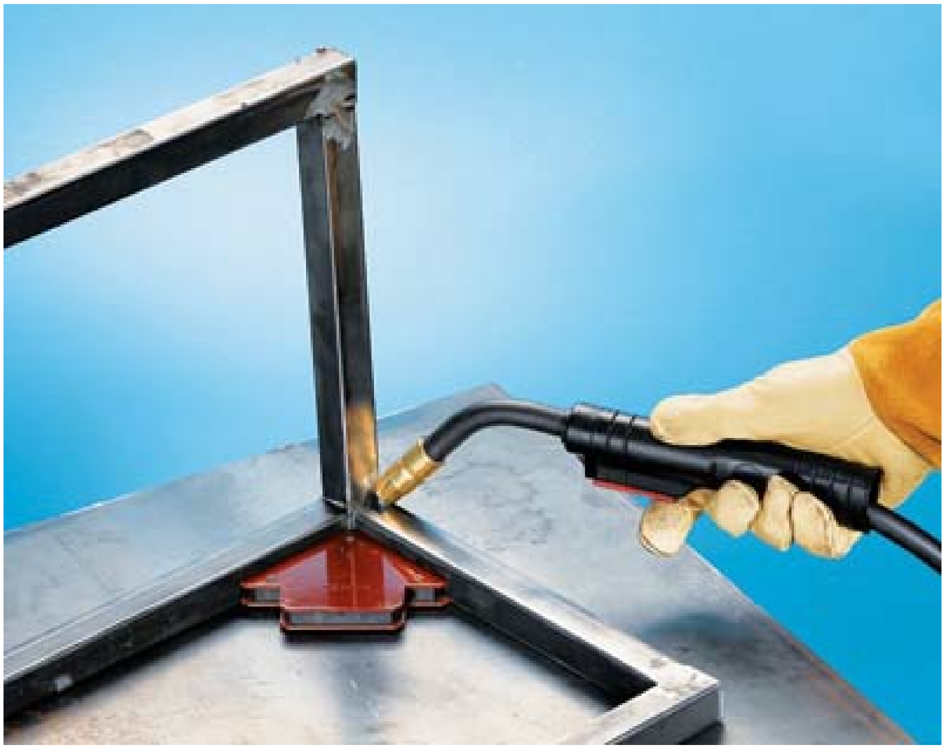

Attach the third leg and crosspieces to create a 90° angle.

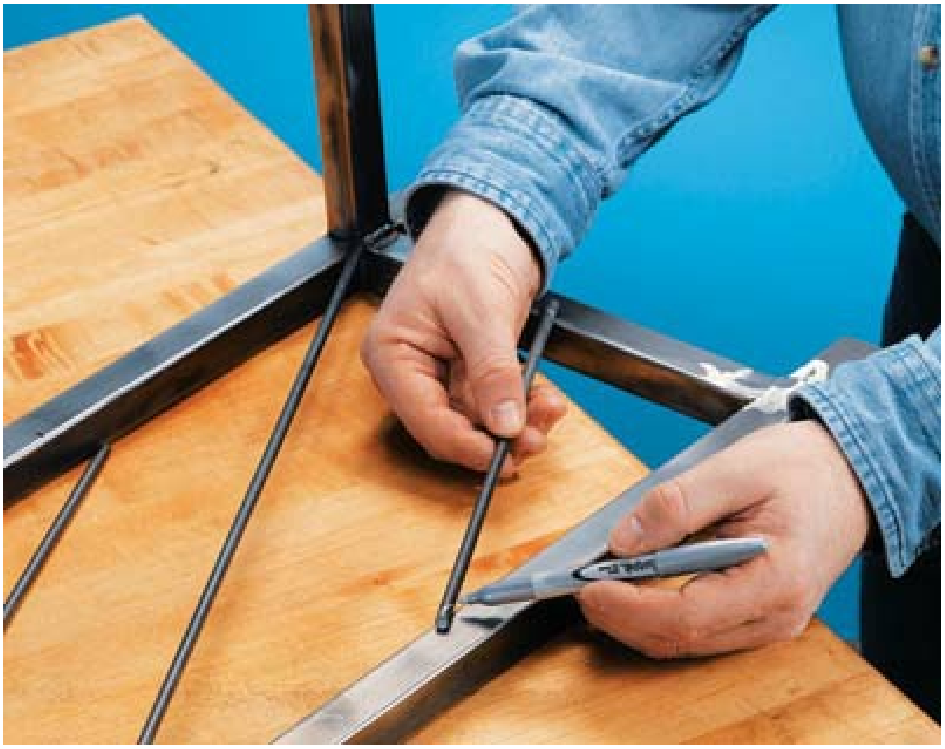

After cutting the diagonals to approximate length, mark the angles where the diagonal crosses the framework.

Assemble the Decorative Diagonals

1. With the side panel on a flat surface, lay the three long and two short diagonals that slant down from upper left to lower right in place. Weld to the frame.

2. Lay the three long and two short diagonals that slant from upper right to lower left in place on top of the other set of diagonals. Weld in place.

3. Cut the side decorations (F) to length. Round one end of each to a smooth semicircle, using a bench grinder.

4. Clamp the rounded end to a ½" round tube and bend the bar a half turn around the tube. Repeat with all the pieces.

5. Place two pieces back to back and weld together. Repeat with the other pieces.

6. Place one welded piece vertically upright at the middle X of the diagonal pattern and weld in place. Place the second welded piece upside down at the middle X and weld in place.

7. Cut and assemble the diagonals for the second side panel.

Make the Shelves

1. Cut the shelf fronts (G) to length.

2. Bend a shelf front into a 12" radius arc by making small bends every ½". Check that it fits in between the two front legs. Adjust, if necessary. Bend the other two shelf fronts.

3. Weld the shelf fronts to the forward inner corners of the front legs.

4. Cut the shelves (H) to size. Allow ¾" bending allowance on all sides.

5. Use pliers and a hand seamer to bend the shelf to fit inside the top and bottom crossbars and shelf fronts. The middle shelf rests on top of the crossbars. Cut notches in the sheet metal, if necessary, to help it fit.

6. Weld the shelves in place.

Finish

1. Place balls (I) in the tops of each leg and weld in place.

2. Sand, wire brush, or sandblast the shelf unit. Finish as desired.

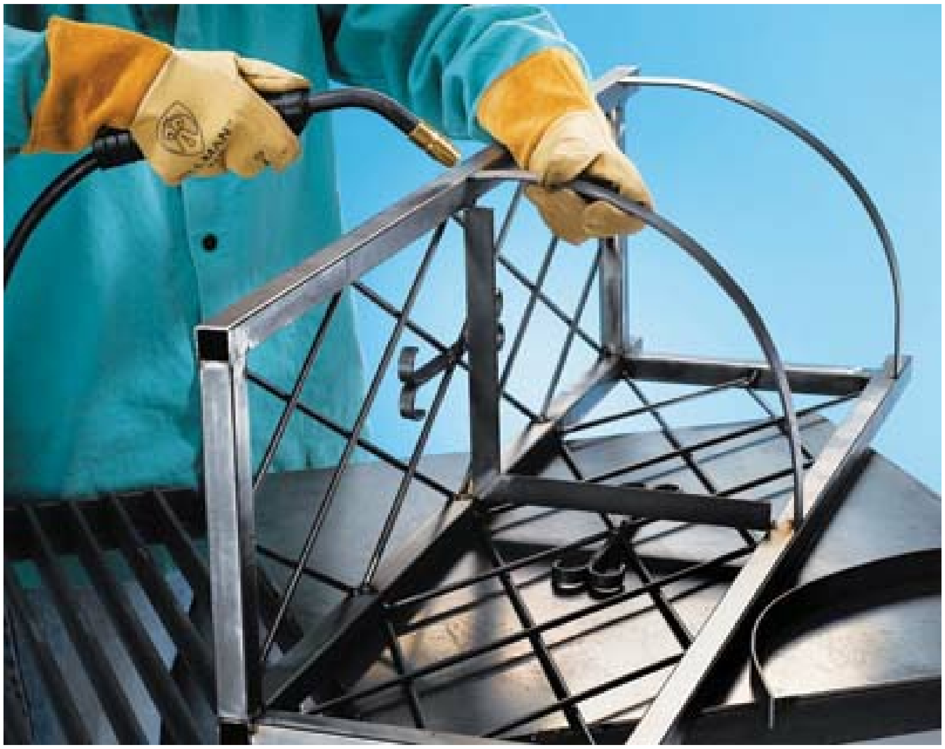

Weld the arched fronts to the sides, even with the crossbars.

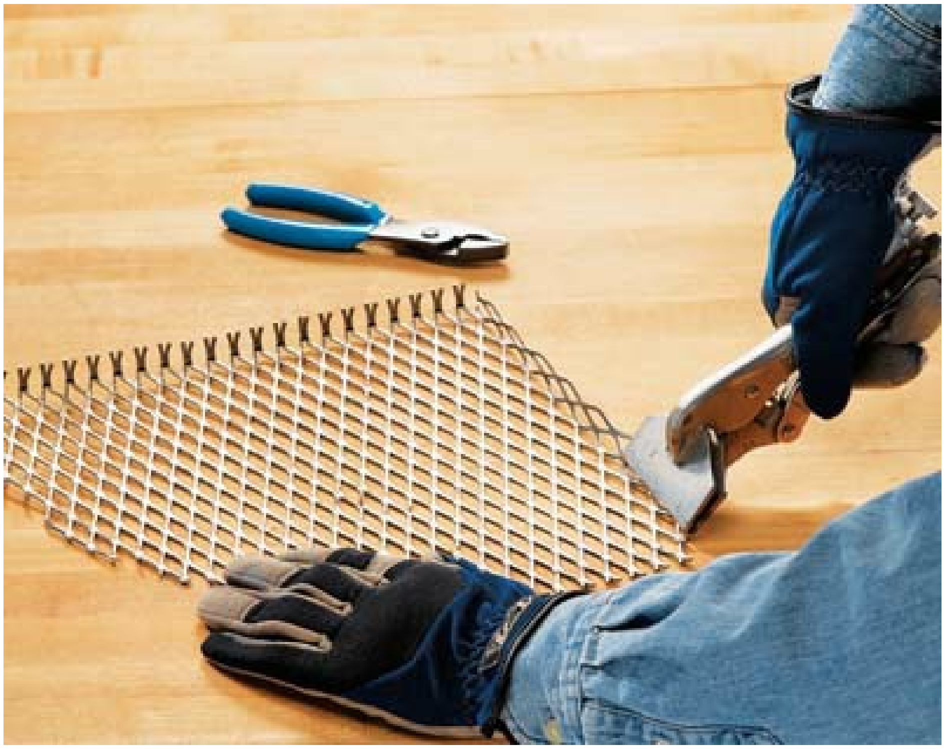

Use a hand seamer and pliers to bend the expanded sheet metal to fit inside the top and bottom crossbars.

Outdoor

IN THIS CHAPTER:

Truck Rack
Trailer
Stair Railing
Gate
Firepit
Vineyard Trellis
Arch & Gate
Gazebo
Arbor

Truck Rack

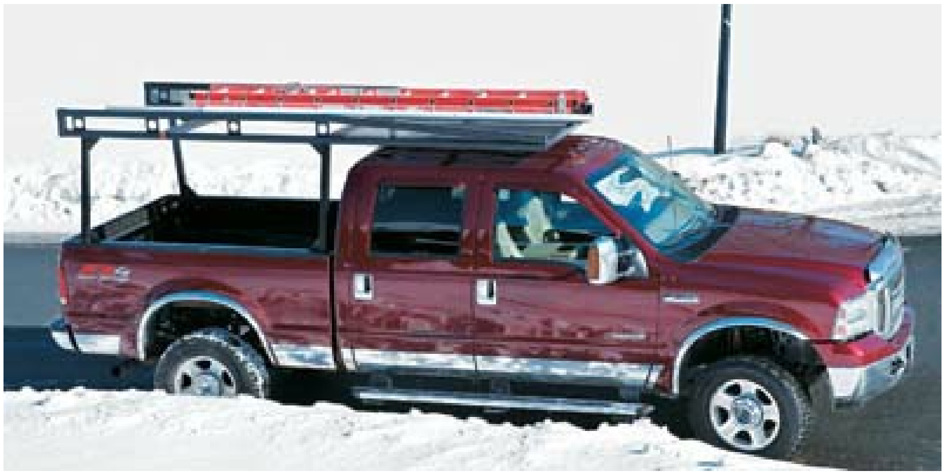

A carrying rack that mounts on the walls of your truck bed is a very handy accessory for serious metalworkers and DIYers with a frequent need to transport building materials. The heavy-duty metal rack shown here is made from 12-gauge steel tubing and can be rescaled to fit onto just about any pickup truck, from compact to full-size. The full-length bedrails distribute the load, protect the truck’s bedrail, and provide rack attachment points along the entire length of the truck bed. They also provide a flat base for mounting a toolbox or cover. If your truck has stake holes and you prefer to align the uprights to fit into those spaces, you must alter the rack design a bit—this is a valid design consideration for ease of installation and less drilling into your truck.

Perhaps the most unique feature of this rack is the round spacer plates that fit between the upper rack bars. The circular cutouts are very handy for tying down a load.

To protect the rear window of the cab, consider attaching a panel of expanded metal mesh to span between the front uprights. You can weld the mesh directly to the uprights or build a metal sub-frame to fit between the uprights and weld the mesh to the subframe. Also consider mesh for over the cab to protect the truck, as we have done.

Even when scaled for a compact pickup, this rack will weigh at least 200 pounds. The final weight, of course, depends on the size of the rack and the gauge of the metal you use. In any event, you will need at least one helper to lift the rack into position. A person for each corner is best. Finally, be sure to finish the metal so that it does not rust.

Materials

- 12-gauge (1/8") 2 × 2" steel tubing, approx. (*dim. a* + *dim. c*) × 4) + (*dim. b* × 4)
- 1/8 × 3 × 3" angle iron (*dim. a.* × 2)
- ¼ × 4 × 4" flat stock steel (5½ feet)
- ¼ × 2 × 2" flat stock steel

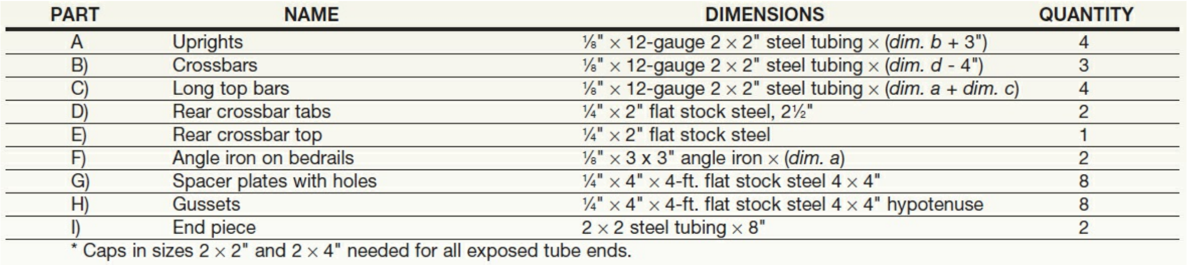

PART	NAME	DIMENSIONS	QUANTITY
A	Uprights	⅛" × 12-gauge 2 × 2" steel tubing × (*dim. b* + 3")	4
B)	Crossbars	⅛" × 12-gauge 2 × 2" steel tubing × (*dim. d* - 4")	3
C)	Long top bars	⅛" × 12-gauge 2 × 2" steel tubing × (*dim. a* + *dim. c*)	4
D)	Rear crossbar tabs	¼" × 2" flat stock steel, 2½"	2
E)	Rear crossbar top	¼" × 2" flat stock steel	1
F)	Angle iron on bedrails	⅛" × 3 x 3" angle iron × (*dim. a*)	2
G)	Spacer plates with holes	¼" × 4" × 4-ft. flat stock steel 4 × 4"	8
H)	Gussets	¼" × 4" × 4-ft. flat stock steel 4 × 4" hypotenuse	8
I)	End piece	2 × 2 steel tubing × 8"	2

* Caps in sizes 2 × 2" and 2 × 4" needed for all exposed tube ends.

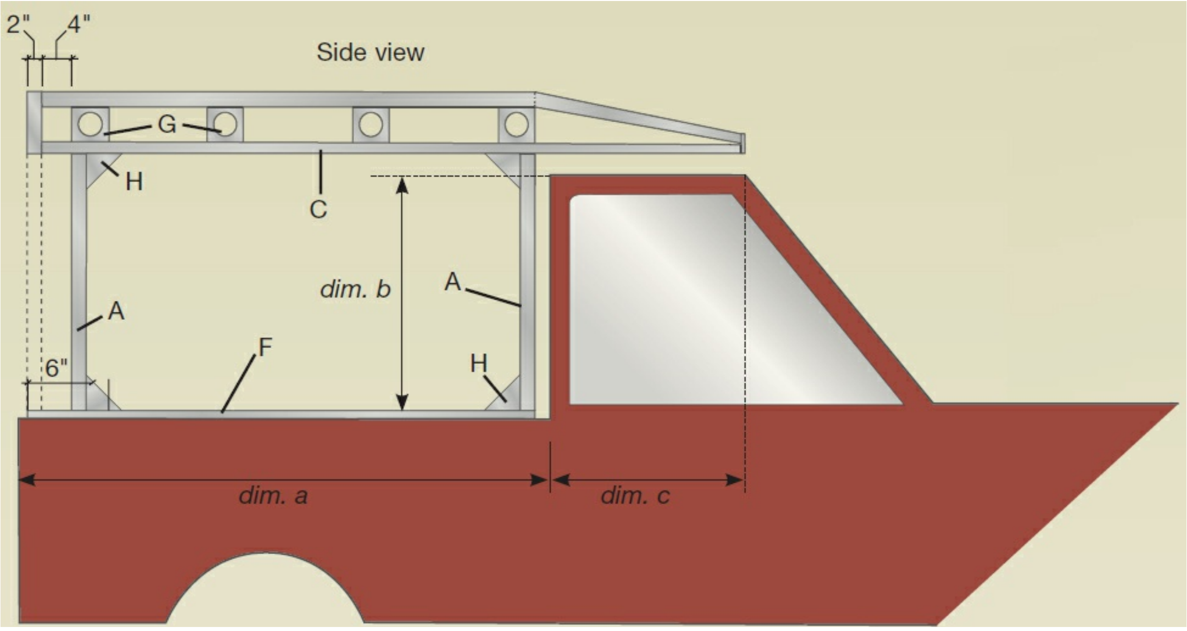
2"
4"
Side view
G
H
C
dim. b
A
A
F
H
6"
dim. a
dim. c

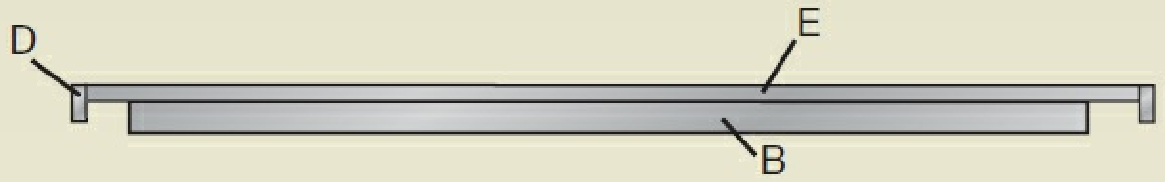

Rear Crossbar Detail

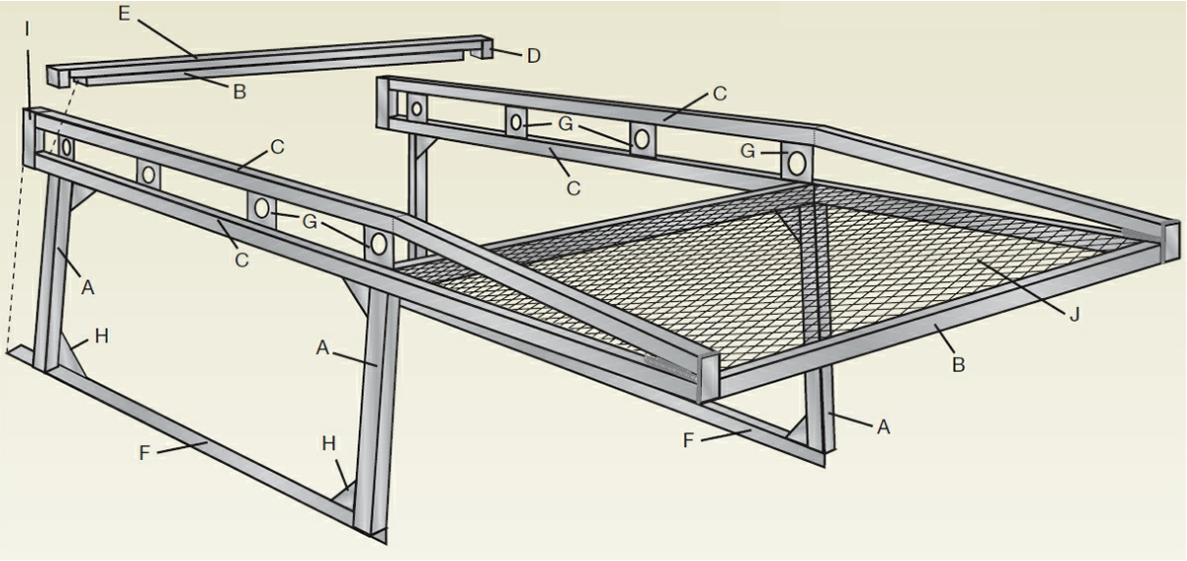
I
E
D
B
C
C
G
G
C
C
G
C
A
J
H
A
B
F
A
F
H

How to Build a Truck Rack

Safety
NOTICE: For safe use of this rack project only use approved, safety-rated products to attach the rack securely to your vehicle. These may include clamp-based, bolt-based, or stake-pocket systems. Inspect clamps and fasteners regularly to make sure they are tightened properly.

Measure the Truck

1. Measure corner to corner on your truck bed for square.

2. Measure and mark your truck bedrails for angle iron (F) placement. The full length of the truck bed is *dimension a*. This dimension is required to design your rack so that it custom fits your truck.

3. Measure the length of your truck cab roof from front to back (*dimension c*)—refer to drawing on page 167. Add this measurement to *dimension a* for the total length of the long top bars (C).

4. Measure from the truck bedrail to the top of the cab. Add 1" for additional clearance above the cab. The height from the truck bedrail to the top of cab is *dimension b*.

Note: The uprights on this rack lean toward the center at a 10° angle so that the long top bars (C) are aligned roughly with the inside walls of the truck bed.

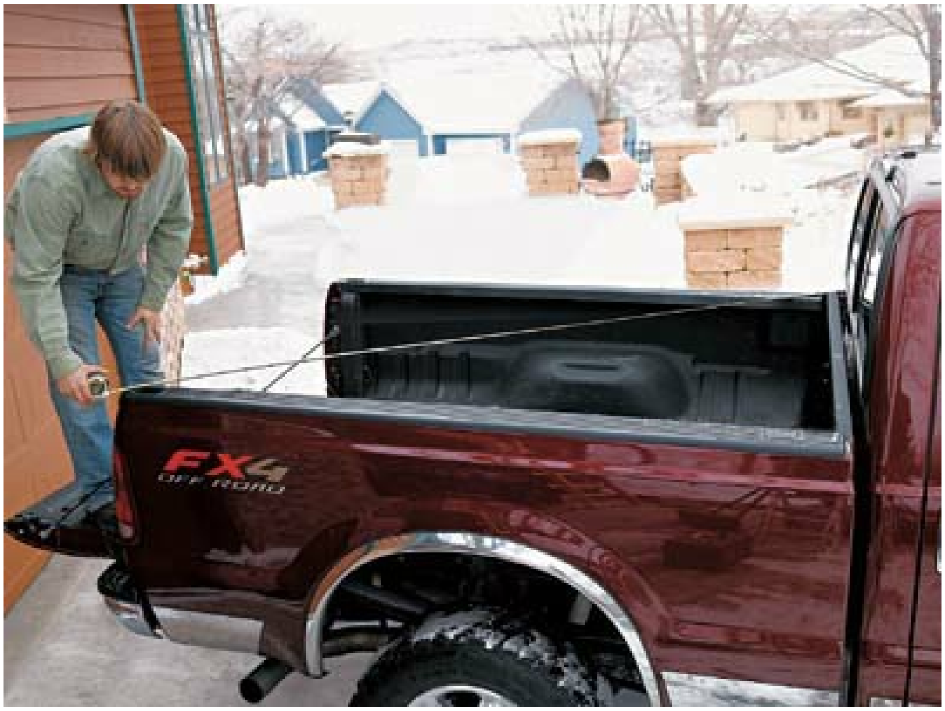
FX4
OFF ROAD

Measure from corner to corner of the truck bed for square.

Prepare the Angle Irons

1. Mark two 1/8 × 3 × 3" angle irons (F) to length (*dimension a*). Use a speed square to ensure a straight cutting line across the parts.

2. Cut the two angle irons at marks using a cutoff saw.

Option: If you choose not to bolt the angle iron to the truck and instead use clamping all along the bedrail angle iron or some other fastening technique, skip to Cut & Fasten Uprights to Angle Irons.

3. Align the angle irons on the truck bedrails. Align the solid edge butted up flush against the inside edge of the truck bed so it caps over the bedrails and clamp in place.

4. Make a mark 6" in from the tailgate edge of the angle irons. The 2 × 2 uprights are aligned 6" in from the tailgate edge. The cab-side upright sits flush to the ends of the angle irons (F) on the cab-side edge, so make a mark 2" in from the cab-side edge of the angle iron as well.

5. Measure in from the marks made in step 4 for four to five evenly spaced bolts along the angle iron (F). It is standard to use four bolts to fasten a truck rack to a smaller truck; larger trucks will require more bolts. Check with your truck manufacturer and the fastener manufacturer’s recommendations.

6. Remove the angle irons and bring them back to your shop.

7. Slowly drill 5/16" holes through the angle iron (F) for grade 8¾" hex bolts at your bolt marks using a drill press or handheld drill. Use cutting oil and drill slowly. Repeat for the other angle iron (F).

FX4
OFF ROAD

Measure in 6" from the tailgate ends of the angle iron and make a mark for the tailgate uprights. Make a mark 2" in from the cab-side ends of the angle irons for cab-side uprights. Then measure and mark the bolt (or other fastener) locations evenly spaced along the angle iron while it is still clamped in place on the truck bedrail.

Cut & Fasten Uprights to Angle Irons

1. On a flat surface, measure and mark 12-gauge (1/8") × 2 × 2" steel tubing lengths at *dimension b* for the four uprights (A). This will leave a 4" clearance between the rack and your truck cab roof.

2. Mark for a 10° angle cut on each upright end. This angle cut ensures the rack fits roughly parallel with the cab roof.

3. Cut the uprights to length using a cutoff saw. Refine the 10° angle with an angle grinder.

4. Align the cab-side uprights flush with the angle iron (F) cab-side ends. Tack along two sides so that you can easily break the tacks and realign, if necessary. *Note: The location of the cab-side uprights determines where you need to bend the long top bar. The edges of A and G and the bend on C should all be aligned.*

5. Align the tailgate-end uprights 6" in from the angle iron ends. Tack along two sides so that you can easily break the tacks and realign, if necessary.

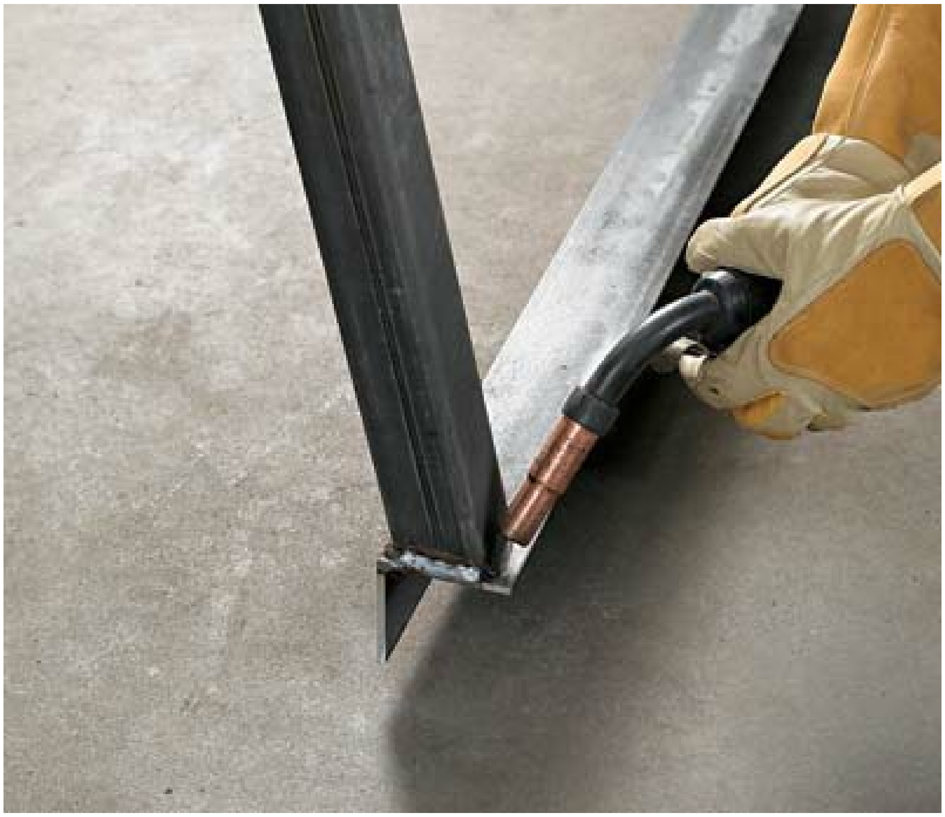

Tack the cab-end uprights (A) onto the angle iron bedrails (F). Then tack the tailgate-side uprights onto the angle iron bedrails 6" in from the edge. Tack only along two sides because you may need to break your tack welds to realign parts several times.

Build the Basic Frame

1. Measure and mark the two lower long bars (C) to the length of the truck bed (minus 2") plus the length of the cab (*dim. a* – 2") + *dim. c*.

Note: When planning out the four long top bar lengths, remember to account for end caps. There is a 2 × 2" tubing cap at the tailgate end and ¼" flat stock steel cap at the front end. Some racks hang slightly over the cab front while others (like our rack) sit behind the cab front. The length of your rack depends on your truck size and how much length you need for typical loads. Consult with your truck manufacturer on recommended clearances.

2. Measure and mark the two top long bars (C) for length (*dim. a* + *dim. c*).

Note: The bend itself creates a variable in the final length of the top bars. When the long top bar bends to meet the lower long bar, the two parts do not meet flush at the ends. The square end is created with a ¼" flat stock steel end cap(s); also, the small triangular space in between the two parts is filled with slag.

3. Cut all four long top bars to length using a cutoff saw.

4. Align the two lower long bars flush on top of the uprights. The long bars will extend beyond the tailgate upright by 4".

Note: The top and lower long bars will be joined with a 2 × 2" steel tubing (8" long) cap at the tailgate end. To have the top bars align with the tailgate end of your truck, the top bars must be aligned 2" in from the tailgate end.

5. Tack the lower long bars to the uprights. Make minimal tacks so that you can break the welds and realign parts, if necessary.

Note: Consider adding a temporary brace made of scrap steel tubing that spans horizontally from the tailgate ends of the two long top bars to help maintain the proper angle. We realigned the top frame later and then introduced the brace, but it may save you some time to add the brace now to avoid making time-consuming breaks and re-tacking later.

6. Measure between the front two edges of the long top bars. Also measure between the long top bars at the cab-side (in between two uprights). Use these measurements to mark the lengths of the two permanent crossbars (B).

7. Cut the two crossbars to length with a cutoff saw.

8. Align the crossbars in between the two lower long bars and then tack them to the long bars. Use four 1" welds evenly spaced 1 to 2"; do not weld all the way down the joint.

9. Measure between the two long bars at the tailgate side of the truck for the removable crossbar. Measure, mark, and then cut this crossbar to length using a cutoff saw.

10. Measure and mark the two rear crossbar tabs (D) to 2 × 2½". Also measure and mark the rear crossbar top (E) to length (length of removable crossbar + 2"). Cut the tabs and top to size.

11. Center the crossbar top (E) over the removable crossbar (B) and clamp together with a corner clamp. Tack weld the two parts together.

12. Align the crossbar tabs (D) flush with the top edges of the crossbar top (E) ends, and then weld them together (see Rear Crossbar Detail).

13. Measure and mark eight 4 × 4" square spacer plates (G) from ¼ × 4 × 4" flat stock steel.

14. Cut the spacer plates using a portable band saw.

15. Mark the center circle on each spacer plate using a piece of 2½" (o.d.) pipe as a guide.

16. Cut the center circles out of the spacer plates using a plasma cutter.

17. Center the spacer closest to the tailgate 5" in from the long bar end and make a mark (C). This leaves a 4" space in between the end piece (I) and spacer for the rear removable crossbar.

18. Align a spacer with its cab-side edge flush with the cab-side edge of the upright. Mark this placement on the long bar (C).

19. Evenly space the other two spacers along the long bar and mark the placements onto the lower long bars (C).

20. Repeat the alignment of the spacer plates (G) along the other long bar (C).

21. Align each spacer plate on the lower long bars (C) in alignment with your marks.

22. Hold the spacer plates in place with pliers and tack weld each spacer to the long bars.

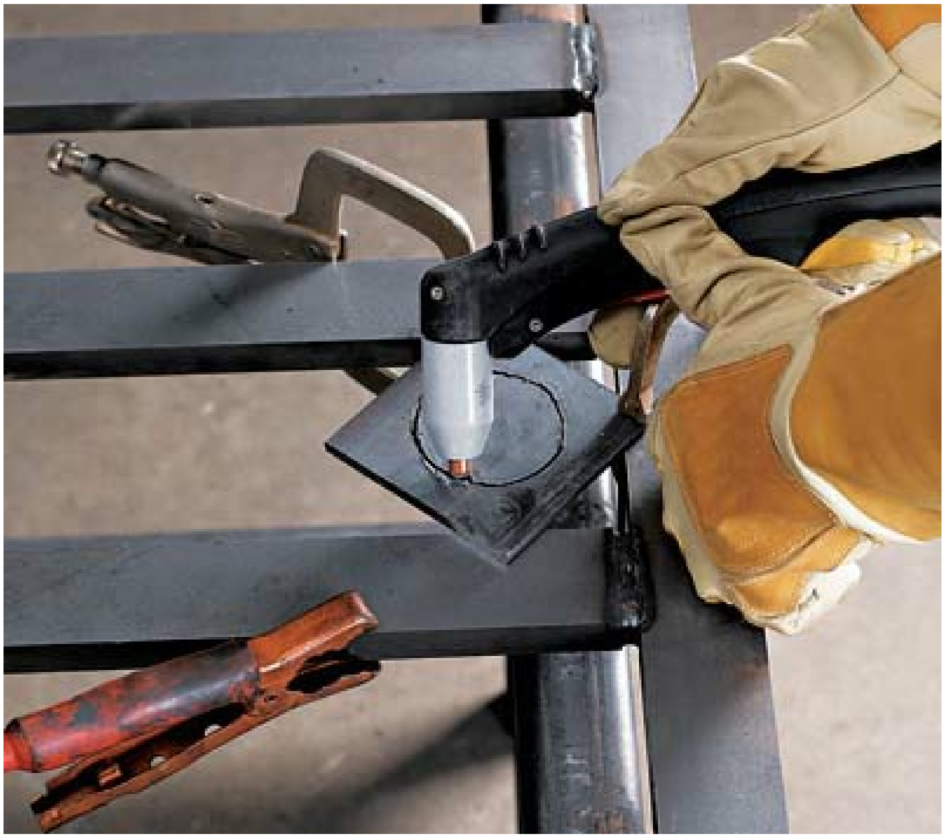

Cut the circle in center of spacer plates using a plasma cutter. Shown here: clamped 4 × 4" flat stock (with circle cutout premarked) and ground secured to table.

Add Top Long Bars

1. Align the two uppermost long top bars (C) directly above the lower long top bars. Center them over the spacer plates.

2. Tack weld the uppermost long top bars to the spacer plates every several inches.

3. Mark the bend location flush with the edge of the cab-side spacer plate and upright.

4. Check the entire rack for proper alignment. Break tack welds and realign parts as necessary. Make adjustment marks and re-tack parts together in proper alignment. *Note: We had to break the welds at the lower top bar and uprights in order to adjust the uprights to proper alignment with 10° angles at each upright*.

5. When the long bars are removed from the uprights (this is very likely going to be necessary; see Note above), take advantage of this situation to make the long top bar bends on a sturdy worktable.

6. Score along the bend mark using an angle grinder or reciprocating saw. Score only as deep as necessary to make the bend.

7. Bend the top long bars to meet the bottom long bars.

8. Mark the cutting line at the front (cab-side) of the uppermost long top bars for the final length. Cut the excess length off the top bars with a portable bandsaw.

9. Tack weld the top long bar to the lower long bar at the front only. *Note: Leave the bend open (not welded) for now.*

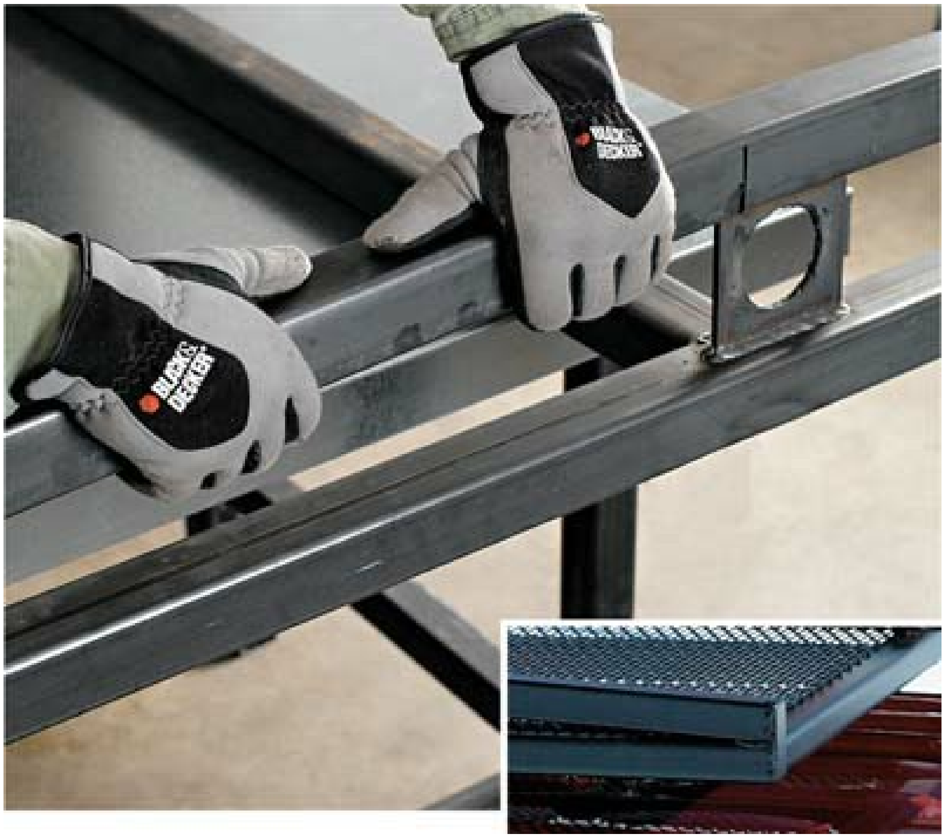

Bend the notched top long bar to reach the lower bar. The front finished triangular ends fill in with slag. The ends are capped with 2× flat stock steel. Inset: The top long bar was capped with a 2 × 2" flat stock steel to bring ends flush; then a 2 × 4" flat stock steel piece was used for the final end cap that spans from top to bottom of the two long bars.

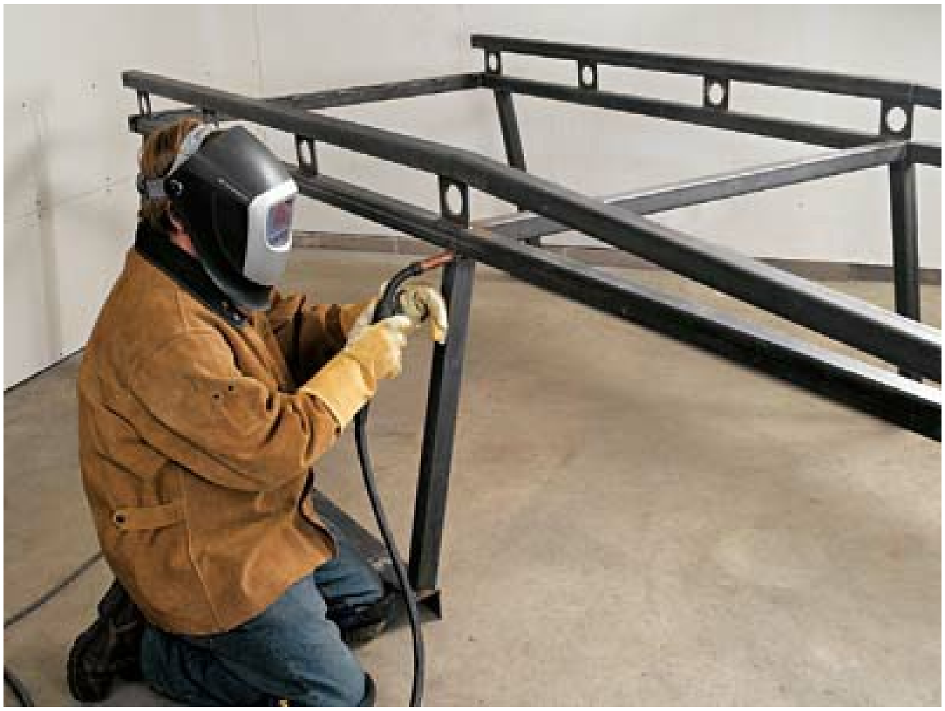

Re-attach the top frame to the uprights, welding at the inside corner of each crosspiece and long bar.

Re-Fasten Top & Bottom Frames Together

1. With the long bars tacked, realign the top/lower long top bars (top frame) atop the uprights and check for proper alignment. *Note: We also tacked on a temporary brace made of scrap steel tubing over the tailgate ends of the lower long bars (C) to help maintain alignment*.

2. Re-tack the long top bars in place on the uprights. Check for proper alignment and make final adjustments, as necessary.

3. Finish the welds at the bend (previously tacked to allow for adjustments), allowing slag to fill the void.

4. Finish the welds at the fronts of the long bars, allowing slag to fill in between the triangular corner.

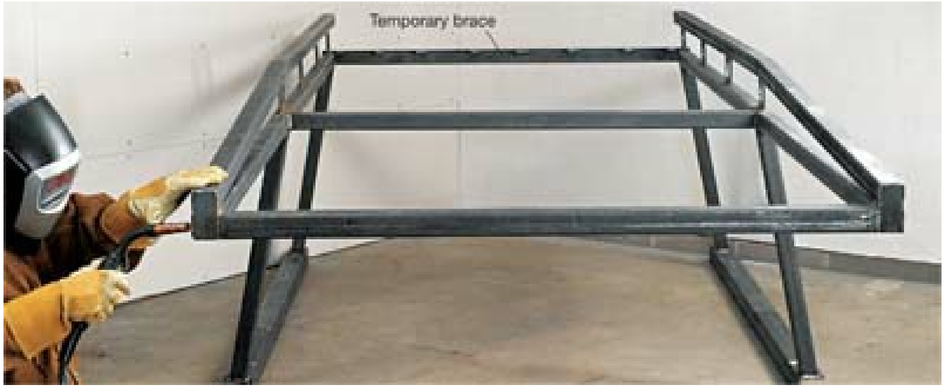
Temporary brace

Weld the final 2 × 4" caps onto the ends of the long top bars for flush front edges.

Add End Caps & Gussets

1. Cap the exposed ends of long bar tubing for a flush front edge. *Note: We had to use a 2 × 2 " end cap on the top long bar and then another 2 × 4 " end cap to cover both the top and lower long bars.*

2. Cap all remaining exposed tube ends.

3. Cut 12-gauge (1/8") 2 × 2" steel tubing to 8" (I). Align each piece against the lower and uppermost top long bars at the tailgate end of the rack. Weld in place.

4. Measure and mark the eight triangular gussets (H) to fit in four inside corners of the rack (see illustration).

5. Cut the gussets with a band saw.

6. Align the gusset plates in each inside corner of the rack, and then weld them in place.

Finish

1. Fully weld all final joints. Make at least 2" welds at the inside corner of each crosspiece and long bar. Do not weld any braces used. And do not weld the rear removable crossbar.

2. Grind down all of the welds and sand all surfaces in preparation for primer.

3. Break the tacks on any temporary braces and remove them. Grind down those final edges.

4. Prime the entire truck rack to protect it from the elements. Allow the rack to dry.

5. Paint or powder-coat the entire truck rack. Use exterior-rated enamel spray paint. Powder coating is best, but it is also the most expensive and requires special equipment.

Install the Rack

1. To install the rack onto your truck, you’ll need a helper or two. Lift the rack up onto the truck and align the angle irons onto the truck bedrail.

2. Fasten the rack angle irons to the truck with clamps, bolts, and/or your chosen fastening method.

Grind all welds and caps smooth with an angle grinder.

Trailer Frame

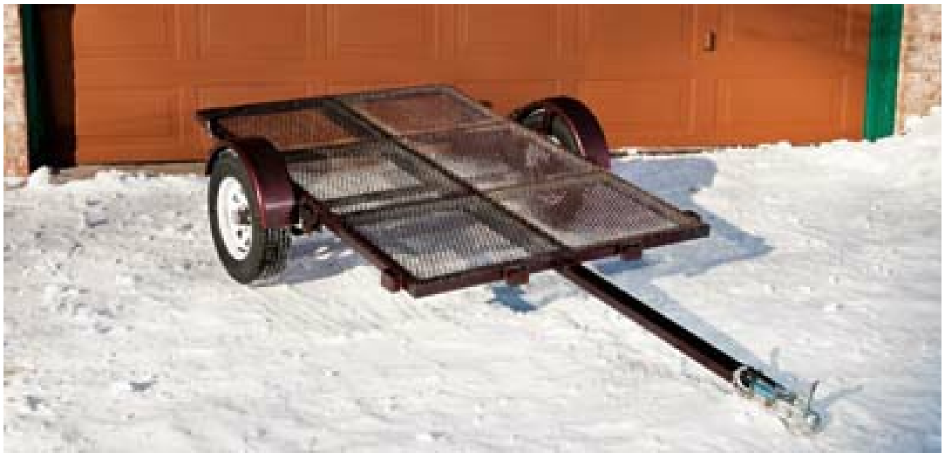

Most trailers have a very specific purpose. It could be to transport lawn equipment or snow removal machines for a small business, it might be to haul all-terrain vehicles or dirt bikes, or perhaps to transport building materials for DIY projects or cart away trash to the dump on the weekend. Each purpose carries special requirements that are reflected in the trailer construction. The welding project seen here is a highly adaptable trailer frame that can be finished and accessorized in just about any way you can imagine to serve your own needs.

A trailer is basically a metal frame with axles and wheels that you attach with a hitch to your vehicle. When choosing an axle kit consider weight capacity, type of suspension (leaf spring, torsion bar, etc.), width of frame, and height of the hitch. Axle kits typically include the axle, brackets to fasten the axle to the trailer frame, suspension (ours included leaf springs), fastening system to attach the leaf springs to the axle (ours included U-bolts, plates, bolts, and nuts), and a hub set. Our kit also included the trailer fenders, tires, and wheels—some kits include necessary parts for brakes and lights as well. The axle kit shown in this project was purchased from Northern Tool Company. The cost for a similar axle kit ranges from $170 to $300.

You’ll also need to choose and purchase a floor material. For vehicle transport—such as snowmobiles or ATVs—expanded metal lath is an excellent choice. If you’ll be hauling construction debris or yard waste, exterior-rated plywood is a good choice; hardwood boards (such as white oak) are a good option for high durability. Walls can also be formed with similar materials.

Further customization of your trailer can include tie-downs and anchors, removable or hinged ramps, and custom tongue or coupler hardware.

Materials

- 14-gauge 2 × 2" steel tubing
- $^{3}/_{16}$ × 2" HRS flat bar (7 feet)
- ¼" × 1¾" HRS flat bar (1 foot)
- Axle kit (incl. leaf springs, U-bolts, U-brackets, axle)
- Hubs and wheels
- Side stakes (2 × 4 treated lumber or metal tubes)
- Trailer sides, front/back, and bed (treated plywood or expanded metal)
- Trailer coupler

PART	NAME	DIMENSIONS	QUANTITY
A	Frame sides	14-gauge 2 × 2" steel tubing × (*dim. a*)	2
B	Frame front/back	14-gauge 2 × 2" steel tubing × (*dim. b*)	2
C	Trailer tongue	14-gauge 2 × 2" steel tubing × (variable dim.)	1
D	Cross bars	14-gauge 2 × 2" steel tubing × (variable dim.)	2
E	Stake pocket sides	3⁄16 × 2" HRS × 1¾"	24
F	Stake pocket fronts	3⁄16 × 2" HRS × 3½"	12
G	Axle kit	Incl. axle, 2 leaf springs, 2 hubs, 4 U-bolts (2 plates, 8 nuts), 4 suspension brackets	1
H	Wheels, tires		2 ea.
I	Trailer side stakes	2 × 4 treated lumbar or metal tubes × 15" (not shown)	12
J	Trailer sides*	1" thick × (*dim. a*) × 12" (not shown)	2
K	Trailer front/back*	1" thick × (*dim. b*) × 12"	2
L	Coupler plates	1¼ × 1¾" HRS flat bar	2
M	Trailer bed*	(*dim. a* × *dim. b*) (not shown)	1

* Treated 2 × 4" lumber, plywood, planks, or expanded metal.

Note: caps in sizes 2 × 2" and 2 × 4" will also be needed for all exposed tube ends.

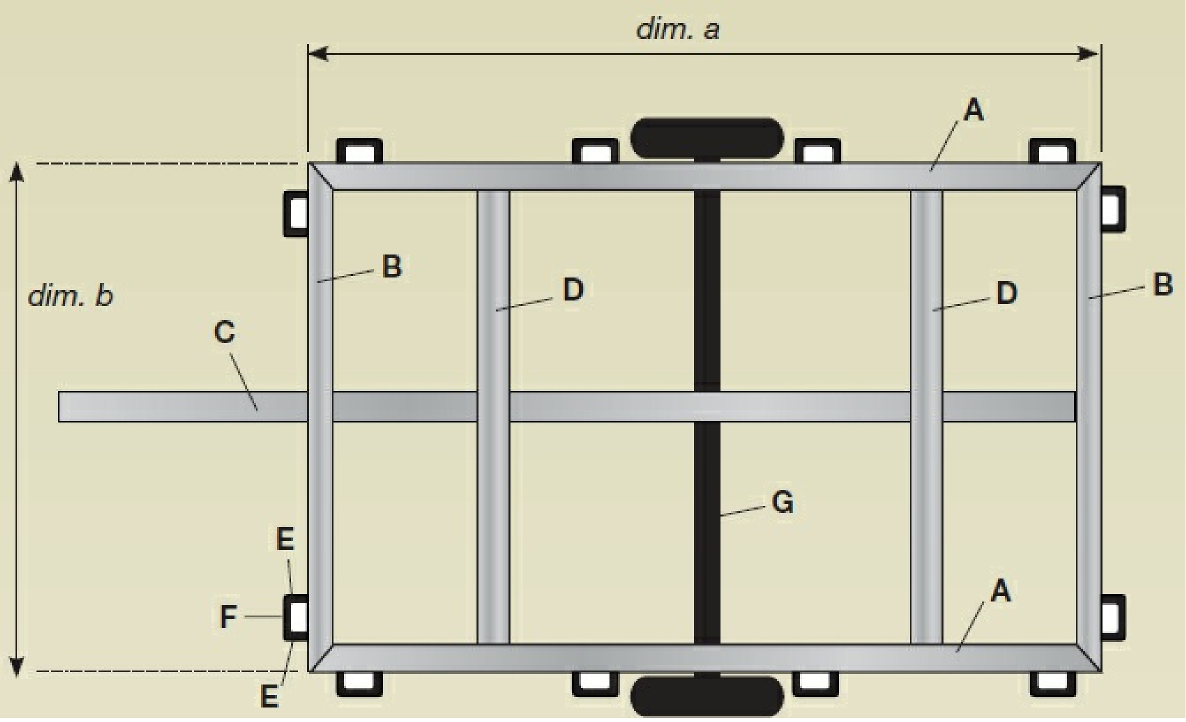
dim. a
A
dim. b
B
D
D
B
C
G
E
A
F
E

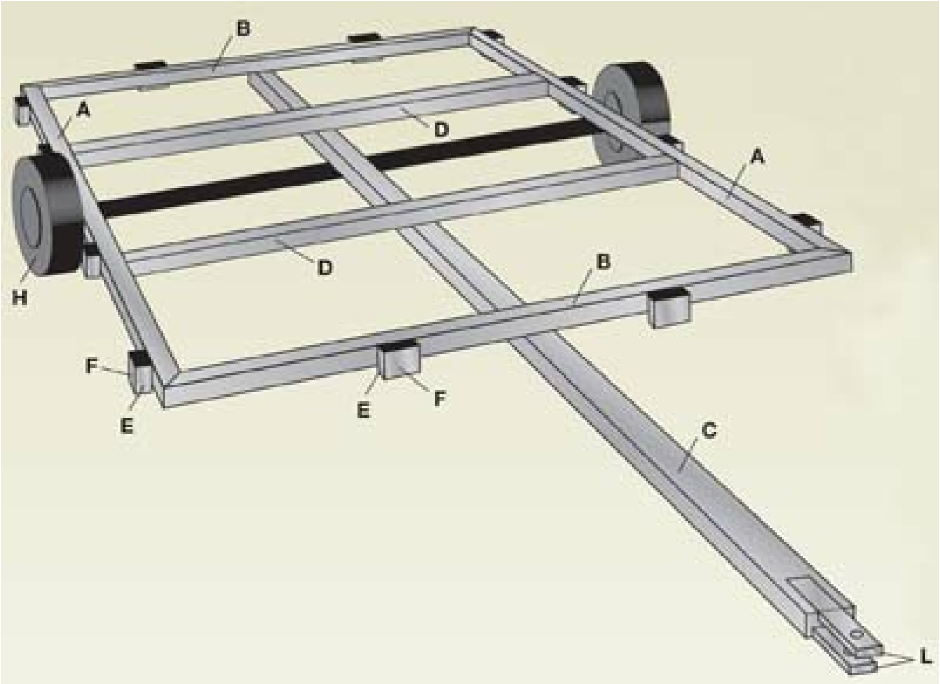
B
A
D
A
H
D
B
F
E
E
F
C
L

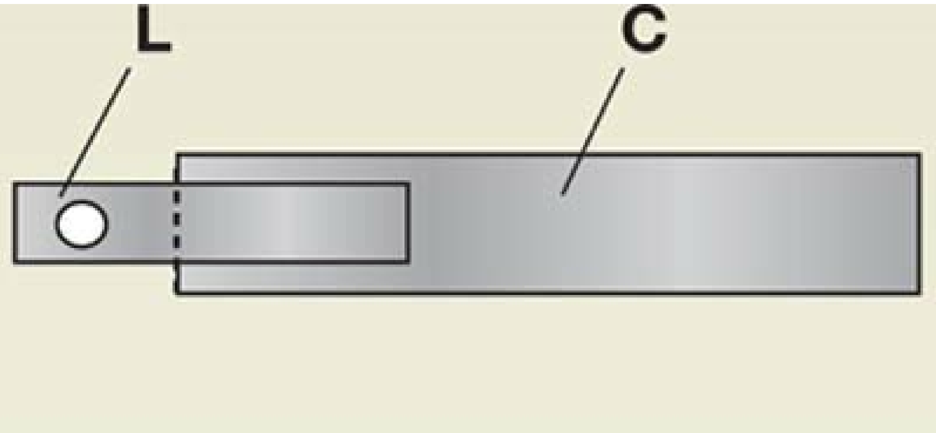

Trailer hitch detail

How to Build a Trailer Frame

Build the Frame

1. Measure, mark, and then cut the frame sides (A) and frame front/back (B) parts to length (*dimension a* and *dimension b* respectively). *Note: We cut the ends at a 45 ° angle for mitered corners.*

2. Square one frame side tube (A) with one frame front tube (B), and tack them together to make an L.

3. Repeat step 2 for the other frame side tube and back tube.

4. Tack the two Ls together and check for square using a combination square or framing square.

5. Weld the inside corners; then weld the outside corners; lastly, weld along the top and bottom.

6. Cut the two cross bars (D) to length.

7. Align the cross bars in between the frame sides. *Note: The exact location is determined by the axle size and placement. The cross bars add overall support to the frame. Most axle kits come with instructions and placement recommendations based on the size of your trailer and the loads you intend to carry*.

8. Check each cross bar for square, and then tack them onto the frame.

Tack weld together the trailer sides and front/back to make the basic frame.

Safety

NOTICE: If you plan on using your trailer on streets or roads it must be licensed and, in most cases, it must have working brake lights, reflectors, and safety chains in case of hitch failure. Some areas may also require that your trailer pass a safety inspection.

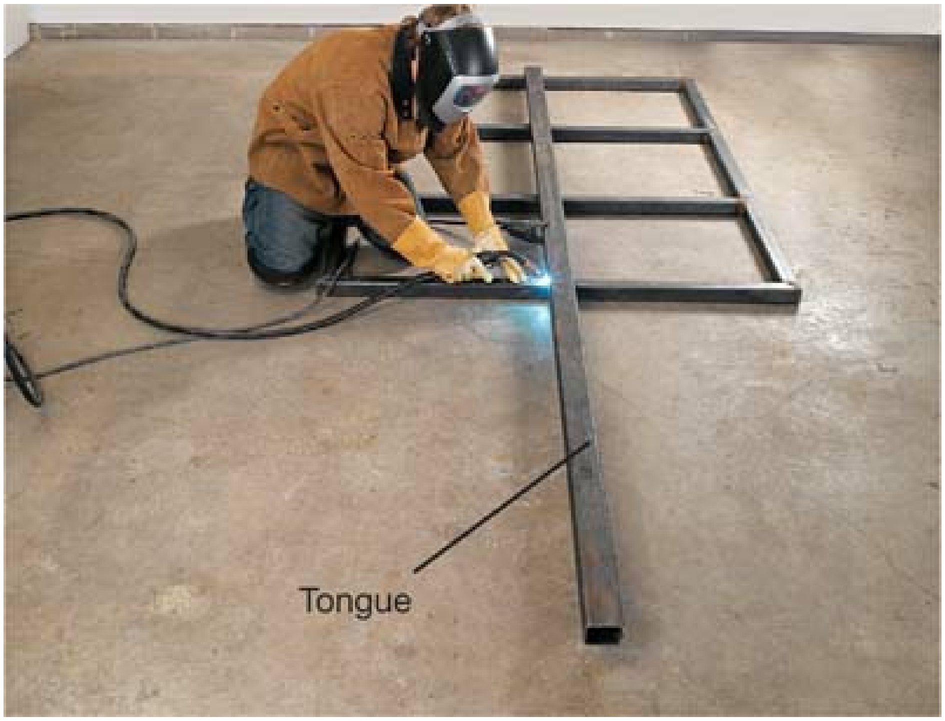

Weld the trailer tongue to the underside of the frame.

Connect the Trailer Tongue to Cross Bars

1. Measure, mark, and then cut the trailer tongue (C) to length.

2. Align the trailer tongue down the center of the frame, on top of the cross bars. Clamp the trailer tongue to the cross bars. Measure to ensure the tongue is centered and square. Adjust as necessary. *Note: We extended the trailer tongue to the far back of the trailer frame for optimal strength.*

3. Tack the trailer tongue to the two cross bars, and then weld it onto the frame.

Install the Axle Kit & Wheels

1. Align a bracket onto each underside of the trailer frame sides. Follow the axle kit instructions, and make sure the brackets are square and aligned at the exact same location on both sides of the frame.

2. Tack the brackets onto the frame.

3. Align the leaf spring into the bracket tacked onto the trailer following the manufacturer's instructions. Locate the placement for the other bracket under the leaf spring, and then align it on the trailer frame. Repeat for the other trailer side.

4. Check the brackets on both sides of the trailer for squareness to each other.

5. Set the leaf spring aside, and then weld all brackets in place.

6. Bolt the leaf spring to the brackets following the manufacturer's instructions.

7. Align the axle in place under the leaf spring following the manufacturer's instructions. Fasten the axle to each leaf spring. Our kit came with two heavy-duty U-bolts that wrap around the axle and then bolt onto each leaf spring.

8. Install the hubs and wheels onto the axle following the axle kit instructions. Inflate the tires.

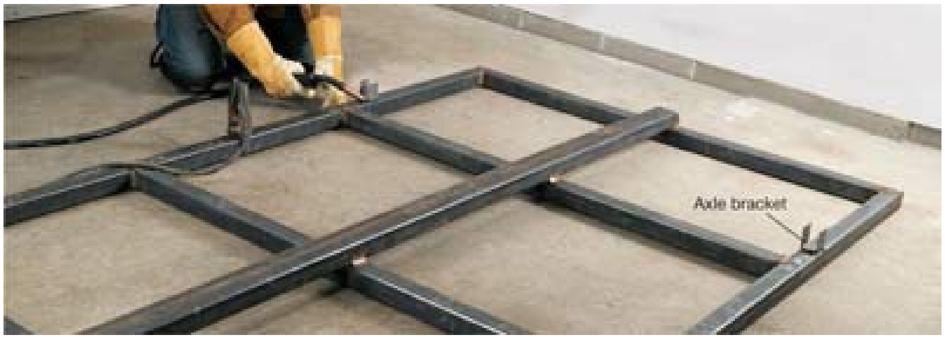

Tack the first axle bracket onto the underside of the trailer frame. Repeat for the other trailer side. The brackets need to be perfectly square to each other. Hold the leaf spring in place over the bracket to align the second bracket on each side. Weld the brackets to the frame, and then bolt the leaf spring to the brackets.

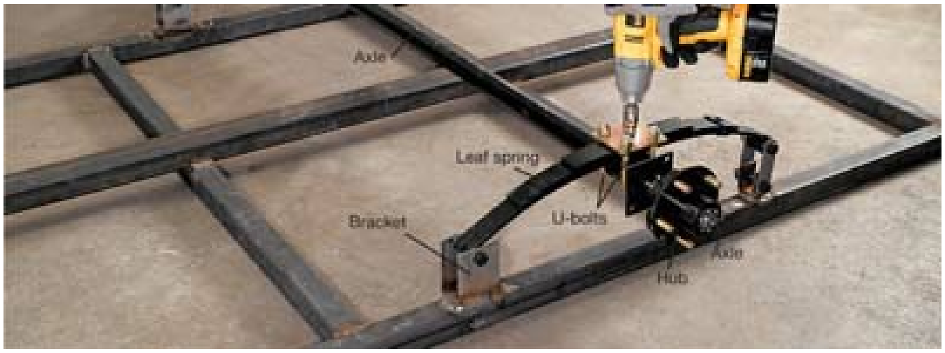
Axle
Leaf spring
Bracket
U-bolts
Axle
Hub

Bolt the axle to the leaf spring with large U-bolts.

Make the Trailer Coupler Plates

1. Cut the coupler plates (L) to size, and then drill holes in the plates. Match the hole size and location to the coupler used. *Note: Our coupler slid over the plates and fastened to the tongue on the top and bottom so we aligned the plates with holes at top and bottom of tongue, but some couplers fasten to tongues at the sides and the hole locations will vary. The coupler and hitch type is determined by your vehicle type, final trailer construction, and the weight of your trailer and anticipated loads (see Resources).*

2. Visually align, clamp, and then tack the coupler plates to the trailer tongue. Extend the coupler plates at least 3" past the trailer tongue end.

3. Mark the angle bend location(s) on the tongue. Determine the tongue angle(s) by measuring from the hitch of the towing vehicle to the trailer tongue (remember to inflate tires). *Note: If the angle is significantly steep, making multiple bends at incremental angles will provide a stronger tongue.*

4. Score the angle mark using a 4½ angle grinder with a fiber cut-off wheel.

5. Bend the tongue to the proper angles.

6. Tack the tongue back together at the scores. Double-check the fit next to the towing vehicle. The trailer should be level.

7. Weld the coupler plates to the tongue.

8. Make all final welds at bends along tongue.

9. Slide the coupler over the coupler plates. Bolt the coupler to the plates, and then weld along all points on which the coupler and tongue connect.

Cut the coupler plates 8” long using a chop saw.

Make the Stake Side Pockets

1. Measure and cut the two sides for each stake pocket (E) to $^{3}/_{16}$ × 2" HRS × 1¾".

2. Measure and cut one front for each stake pocket (F) to $^{3}/_{16}$ × 2" HRS × 3½".

3. Align a pocket side (E) and front (F) onto a 90° magnetic square to form open corner joints. Tack the pieces together to form an L, and then make the final weld.

4. Use a magnetic square to weld on the second pocket side to form the complete U for all twelve pockets—be sure to use an open corner joint again. Make all final welds.

5. Repeat for all twelve pockets.

6. Align pockets around the trailer so they are evenly spaced along the long frame sides, starting at 2" in from the corners.

7. Tack the pockets to the trailer frame. Check for square.

Finish

1. Roughen the surfaces by sanding.

2. Apply a rusty metal primer for optimal rust protection.

3. Paint or powder coat the trailer with paint rated for exterior applications for extra durability and weatherproofing.

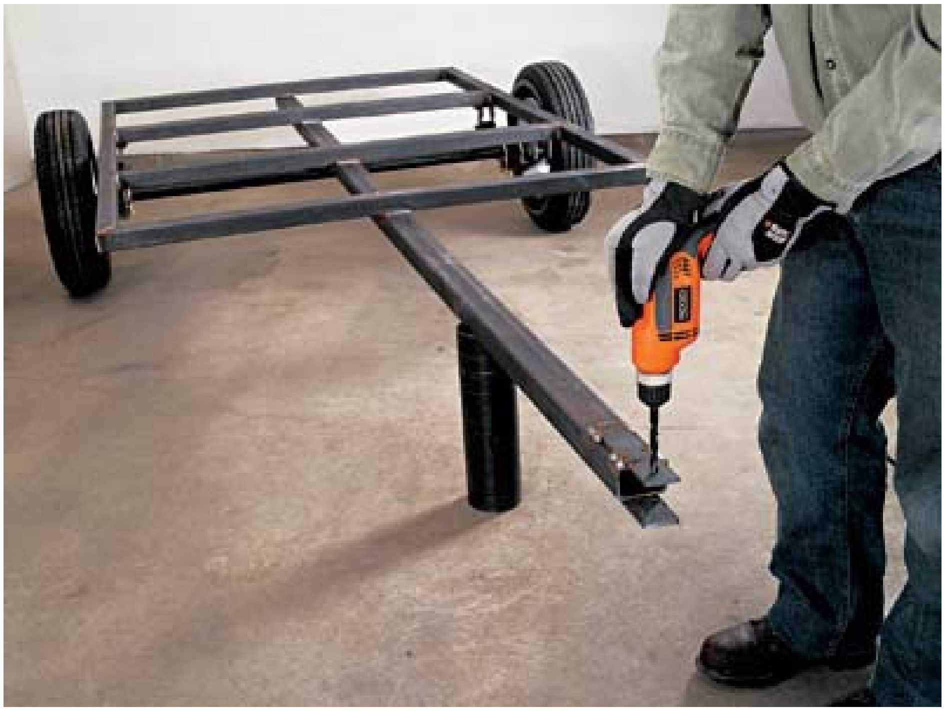

Drill holes into the coupler plates according to the coupler specifications.

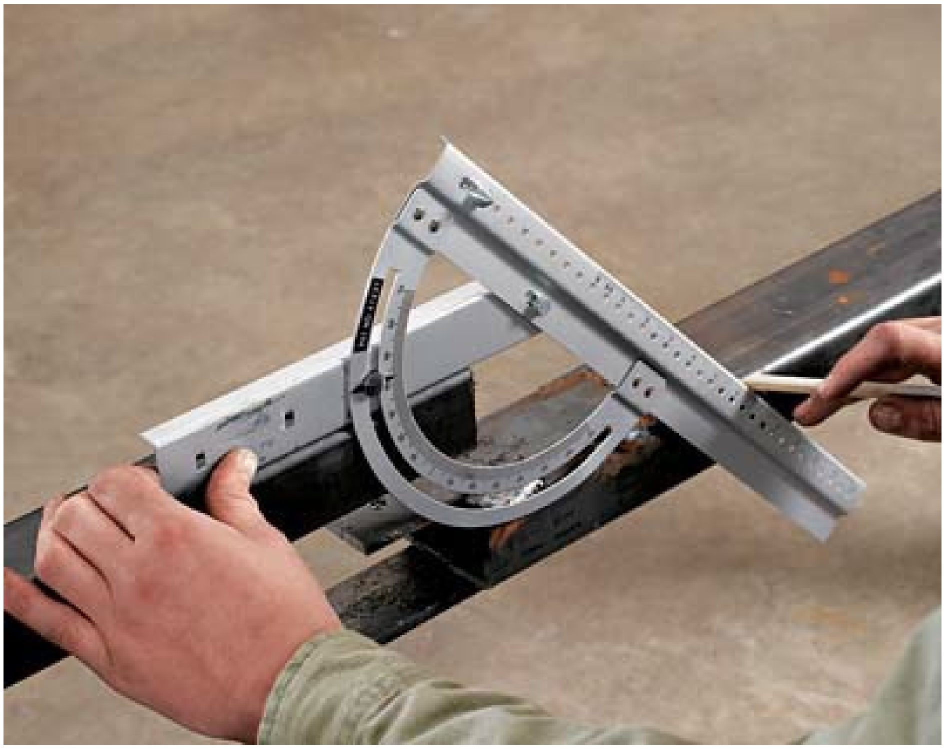

To determine the angle for the trailer tongue, align an angle finder on your truck or ATV hitch and on the trailer tongue. Mark this angle at the appropriate location on the tongue.

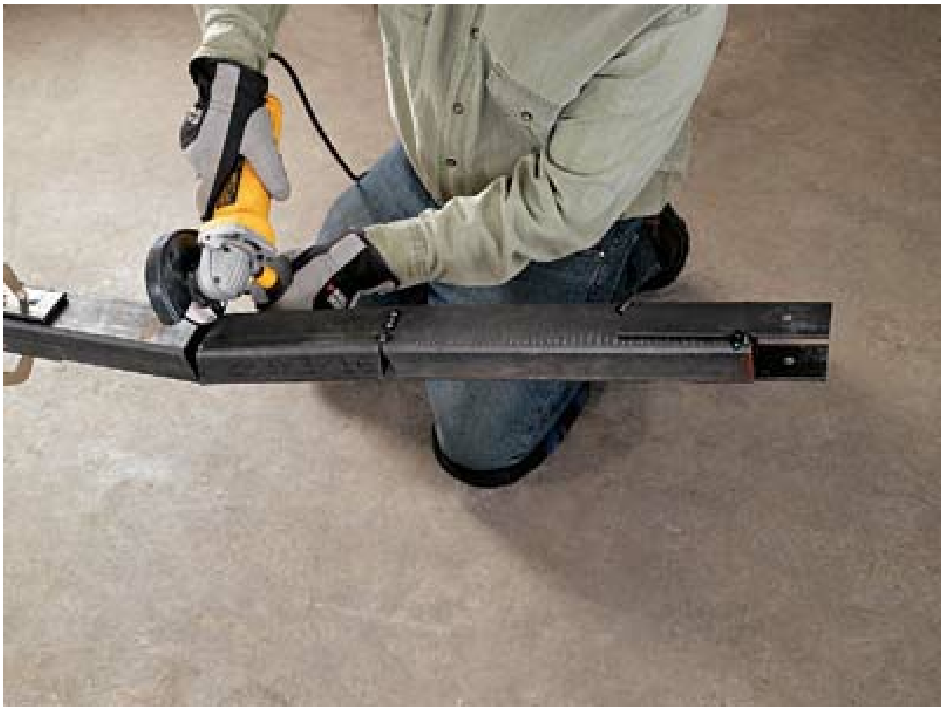

Score notches along the tongue to achieve the desired angle using a handheld grinder. Tack the bends and then hook up the trailer to your towing vehicle—the trailer should be level. If necessary, break tacks and re-bend tongue. Make final welds along each bend.

Powder Coating

Powder coating is the most effective painting technique to protect exterior metal projects, as it produces a very thick coating, results in little to no paint drips or runs, and produces an abrasion-resistant and tough finish. Home powder-coating systems range in price from $100 to $700. See Resources.

Assemble the side pockets. These pockets are available pre-made if you find this to be too time-consuming. They are relatively inexpensive.

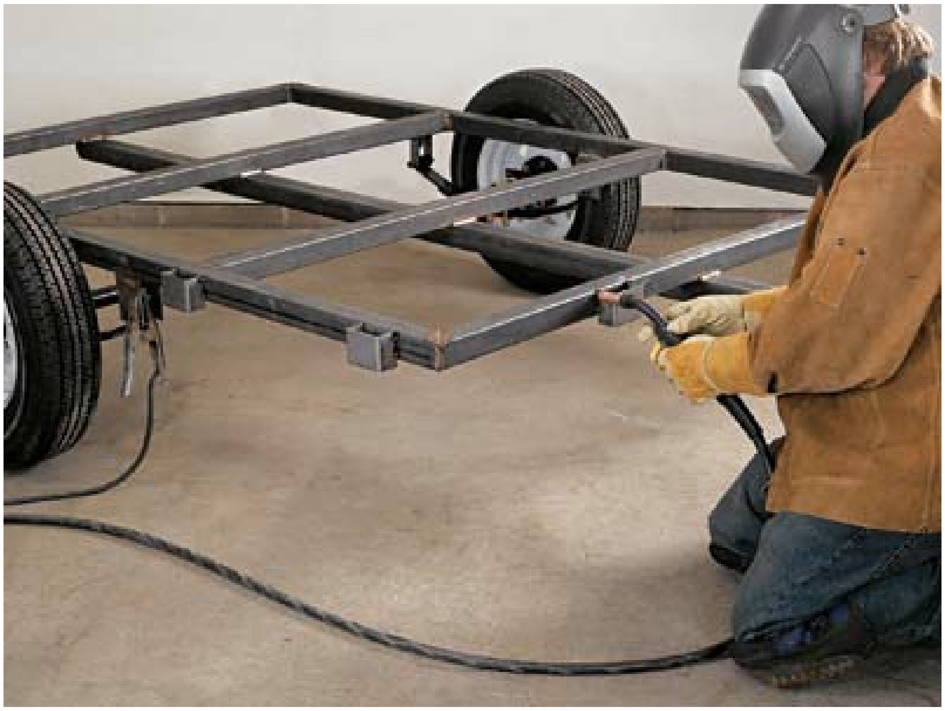

Tack the pockets around the perimeter of the frame.

Floor & Side Options

If you plan to use your trailer for yardwork and hauling debris, you’ll want to add a solid floor and sides to your trailer. A simple method, which we show here, is to attach plywood siding to 2 × 4 exterior-rated stakes set into the pockets along the trailer frame. The stakes are exterior-treated 2 × 4 lumber and the siding is ¾" × 4 × 8" plywood. The plywood floor is drilled directly to the trailer frame.

Make Removable Sides & Trailer Bed

1. Measure and cut twelve 2 × 4s (I) into 15" lengths.

2. Measure and cut the plywood bed (M) to *dimension a* × *dimension b*.

3. Measure and cut the plywood sides (J) to *dimension a* × 12".

4. Measure and cut the plywood front/back pieces to *dimension b* × 12" (h).

5. Insert the twelve 2 × 4 stakes into the trailer pockets.

6. Fasten the plywood sides and front/back to the 2 × 4s. If a 2 × 4 does not quite fit, use a chisel and hammer to adjust it.

7. Fasten the trailer bed to the frame with bolts or self-tapping metal screws spaced every 6 to 12".

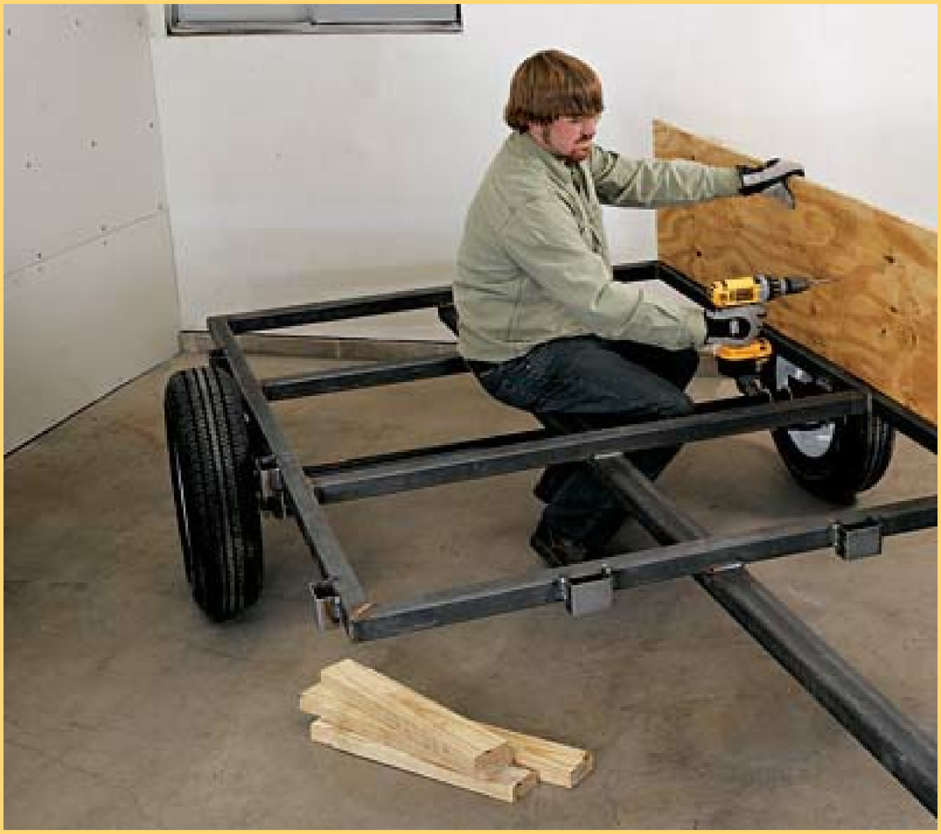

Fasten ¾"-thick plywood trailer sides to the exterior-rated 2 × 4 stakes.

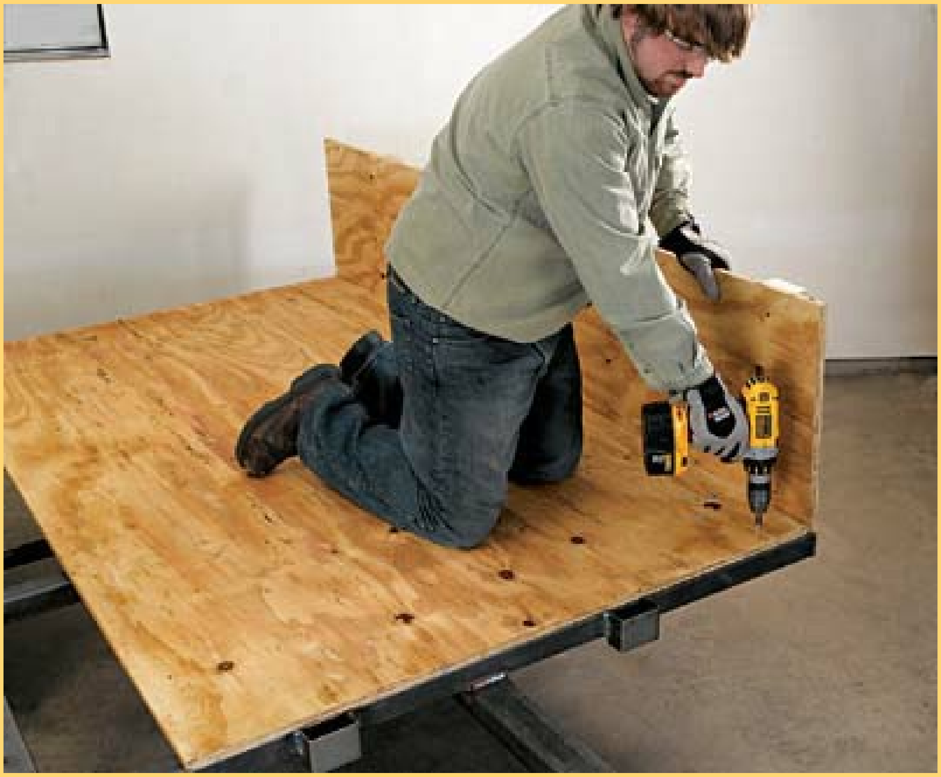

Drill the exterior-rated plywood sheet for flooring directly to the trailer frame.

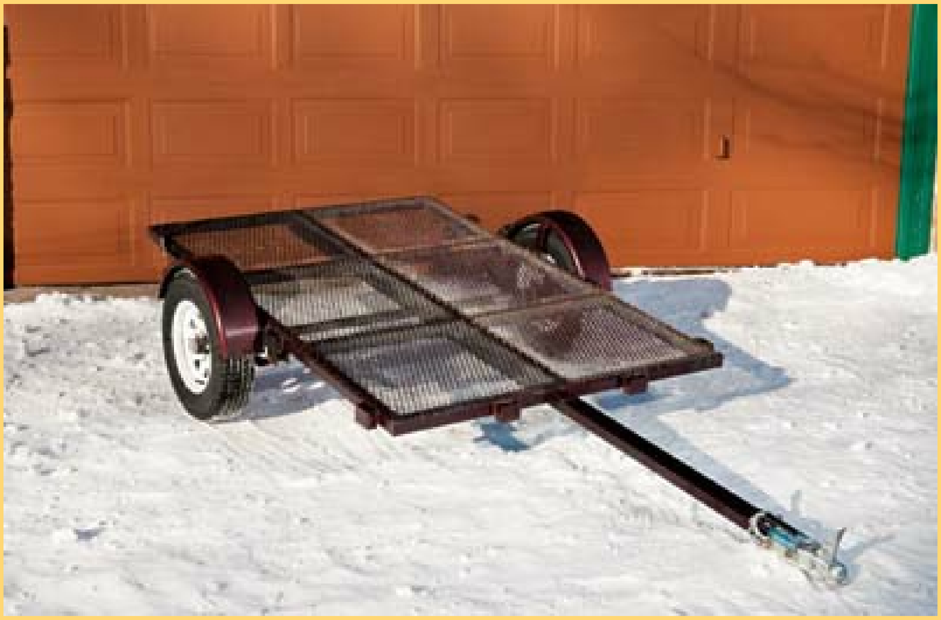

Alternative: Weld expanded metal lath all along the top of the frame perimeter. This is more suitable than wood for hauling ATVs, motorcycles, and snowmobiles.

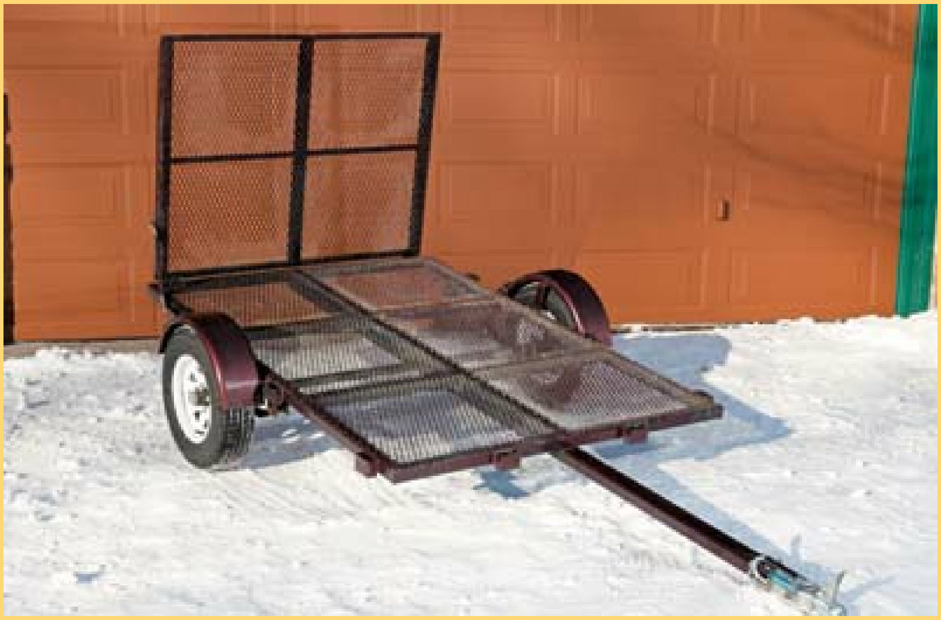

If the primary use of your trailer is to transport vehicles, another feature to consider is a drop-down back. This requires building out a frame and then welding expanded metal lath onto the frame—just as you did with the trailer floor—and then using a purchased or custom-made bolt-on hinge system with latches.

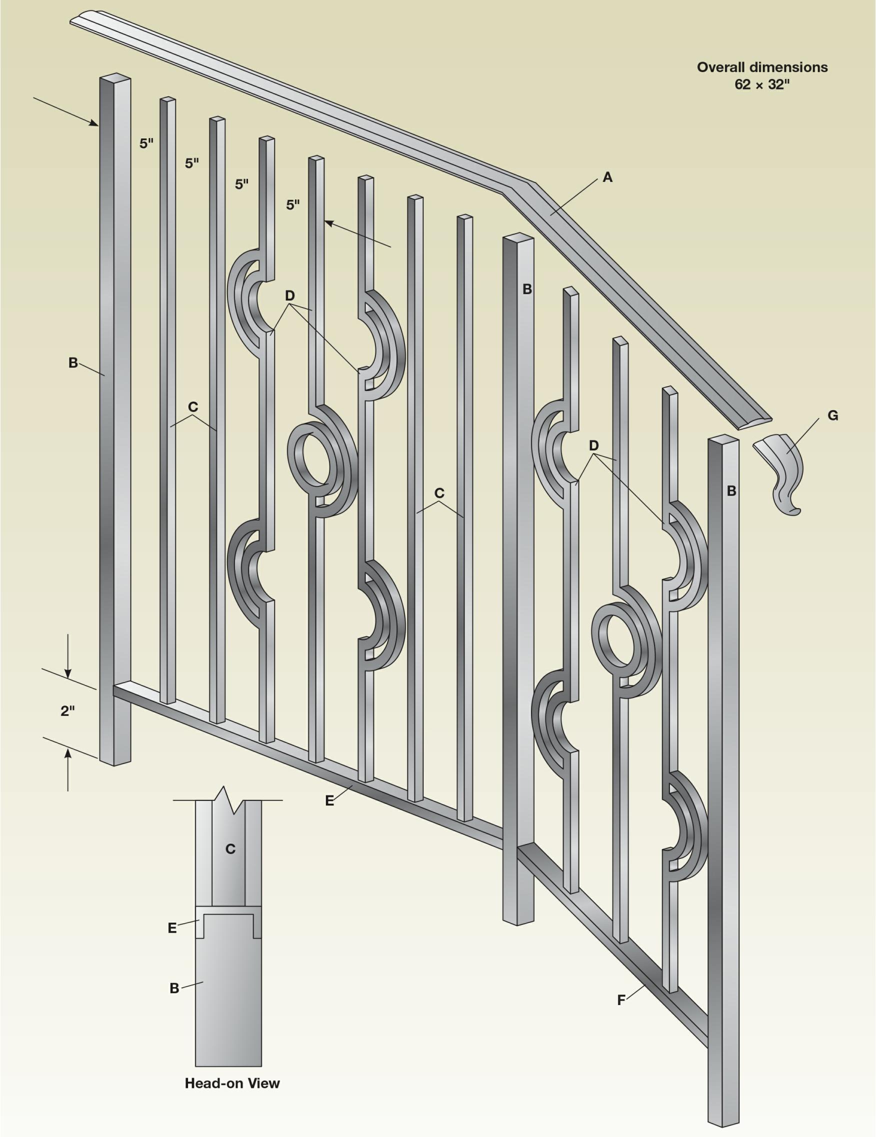
Overall dimensions
62 × 32"
5"
5"
5"
5"
A
B
D
B
C
D
G
C
B
2"
E
C
E
B
F
Head-on View

Stair Railing

With the number of companies selling decorative metal pickets and newel posts, it is easy to create a railing distinctly your own. The railing we are making is for a two step concrete stairway.

Some important information about creating railings:

The rail top should not be interrupted by knobs or other decorations—it must allow continuous hand to rail contact and should be between 1½ and 2" in diameter to ensure easy gripping.

Some railings—those mounted against a wall or in an area where the stair surface is not raised above the surrounding surface, like the stairs in the slope of your lawn—can be a single rail at the appropriate height. A railing that separates the stair, landing, balcony, or deck from a vertical drop has to conform to certain safety standards. Typically, the pickets or balusters must not have any gaps larger than 5½" for child safety. The lower rail should not be more than 2" from the floor surface. If the vertical drop from the floor line is more than 6 feet, the railing must be a minimum of 34" tall. If the vertical drop is less than 6 feet, the railing must be a minimum of 32" in height. Railings must have turn-outs or roundovers at their ends to prevent blunt projections. The safest handrails continue 12" beyond the top and bottom stairs. Railings must be able to withstand a 250-pound force in any direction without giving way. Check your local building codes for specific requirements.

You can purchase rail cap, also called cap rail or handrail, at most steel supply centers. Handrail terminations come in a variety of styles. Those that scroll in an S shape are called “lamb tongues.” A “volute” is a spiral termination, and a “lateral” is a flat curl to the side. These may be ordered through specialty railing distributors. See Resources page 220.

Materials

- 1¾" rail cap (as needed)
- 1" square tube (as needed)
- ½" square rod (as needed)
- ½" decorative pickets (as needed)
- 1/8 × ½ × 1" channel (as needed)

- Lamb tongue rail termination
- Mounting hardware

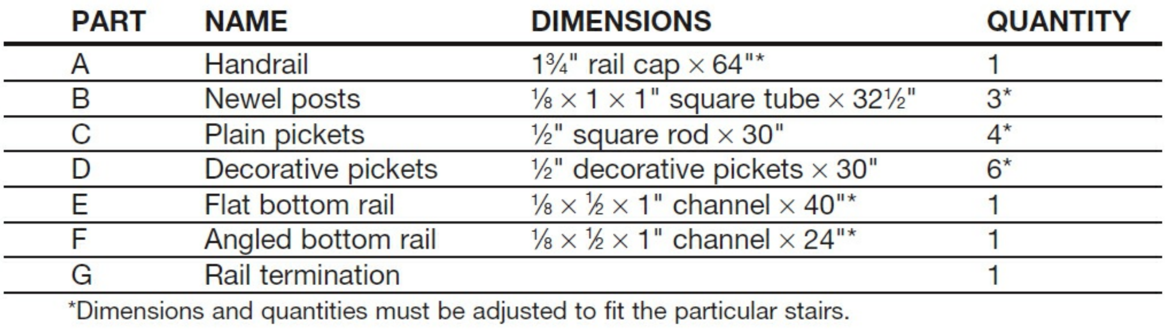

PART	NAME	DIMENSIONS	QUANTITY
A	Handrail	1¾" rail cap × 64"*	1
B	Newel posts	⅛ × 1 × 1" square tube × 32½"	3*
C	Plain pickets	½" square rod × 30"	4*
D	Decorative pickets	½" decorative pickets × 30"	6*
E	Flat bottom rail	⅛ × ½ × 1" channel × 40"*	1
F	Angled bottom rail	⅛ × ½ × 1" channel × 24"*	1
G	Rail termination		1

*Dimensions and quantities must be adjusted to fit the particular stairs.

How to Build Stair Railing

Cut the Handrail to Length

1. Measure from the edge of the house to the edge of the landing. Measure from the edge of the landing to where the newel post will be located. Add these two measurements to get the length of the handrail.

2. Cut the handrail (A) to size.

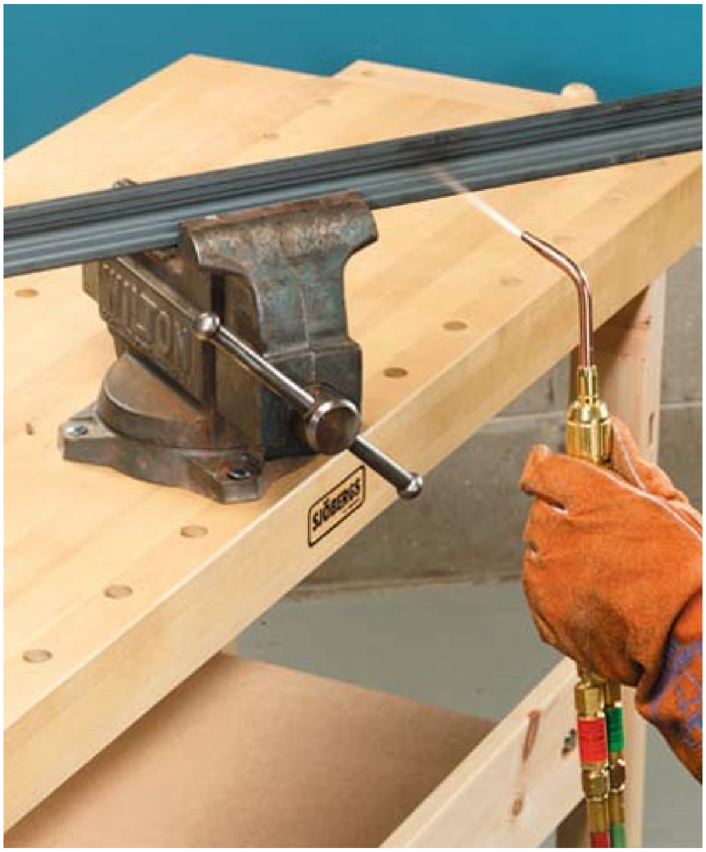
WILTON
SJÖBERGS

Heat the bending point of the handrail to red hot with an acetylene torch or cutting tip to make the metal easy to bend.

Shape the Handrail

It is a good idea to create an angle guide for bending the handrail by screwing two pieces of wood together at the length and angle to match the stairway.

1. Set the handrail on the landing with one end butting against the house. Mark the handrail at the edge of the landing.

2. Clamp the railing in a bench vise and heat the bending point red hot with an acetylene torch. A cutting torch preheat works best, just make sure you don’t hit the oxygen and accidentally cut the metal.

3. When the metal is red hot, bend it to create the angle. It is helpful if you pull on the longer end of the rail to bend so you have more leverage.

4. When finished bending the handrail, place it on the stairs to make sure the bend is correct.

Lay Out the Newel Posts & Pickets

The newel posts will be anchored in the concrete or attached using square footings that will bolt into the concrete, so they have to be far enough from the edge not to destroy the edge of the concrete.

1. Determine the number of plain pickets (C), decorative pickets (D), and newel posts (B) you need.

2. Place the bent handrail on the floor or a large work surface. Place newel posts at each end of the handrail and near the bend.

3. Lay out the pickets in a pleasing pattern, making sure they are no more than 5½" apart.

4. Mark and cut the two bottom rails (E & F) to fit between the newel posts, once you have the layout determined. Place the channel flat side up, with the legs down. Mark the picket locations on the railings.

Option: *When you look at railings, you will see that some railings have the pickets and newel posts welded directly to the rail cap, as you see here. Other railings have a piece of channel welded into the underside of the rail cap and the pickets are welded to the flat side of the channel. Punched channel can be purchased with ½ × ½" square holes pierced through it. This is welded under the rail cap, and the pickets are inserted through the punched holes and welded in place. You also can use the punched channel for the bottom rail. Using punched channel means you cannot adjust spacing to account for the unique shapes and sizes of decorative pickets.*

Cut the Posts & Pickets to Length

Purchased decorative pickets range from 36 to 39". When cutting decorative pickets, cut equal amounts from each end unless you wish the pattern to be off center.

1. Determine the height of your handrail and the depth that the newel posts will be footed in the concrete (if they are to be footed, otherwise measure to the top of the concrete).

2. Cut the newel posts and pickets to length. Cut the appropriate angle for the stair pickets.

Assemble the Railing

Concrete can explode when heated, so it is best to do your welds on a sheet of plywood that can be doused with water when you have finished.

1. Tack weld the newel posts to the rail cap. Tack weld the bottom rails to the newel posts.

2. Place the rail assembly on the stairs to make sure the dimensions are correct. If they are not, break or grind off the tack welds and make adjustments.

3. Return the assembly to the work surface, and tack weld the pickets in place, maintaining the proper spacing. Use a combination square to check each piece for square before welding.

4. Make the final welds. Weld the lamb’s tongue termination to the end of the rail cap.

5. Grind down any rough or unsightly welds. Wire brush or sandblast the rail assembly.

6. Install the railing by cementing the newel posts into the stairs or using bolt-down flange shoes.

7. Prime and paint the railing with a high quality outdoor metal paint.

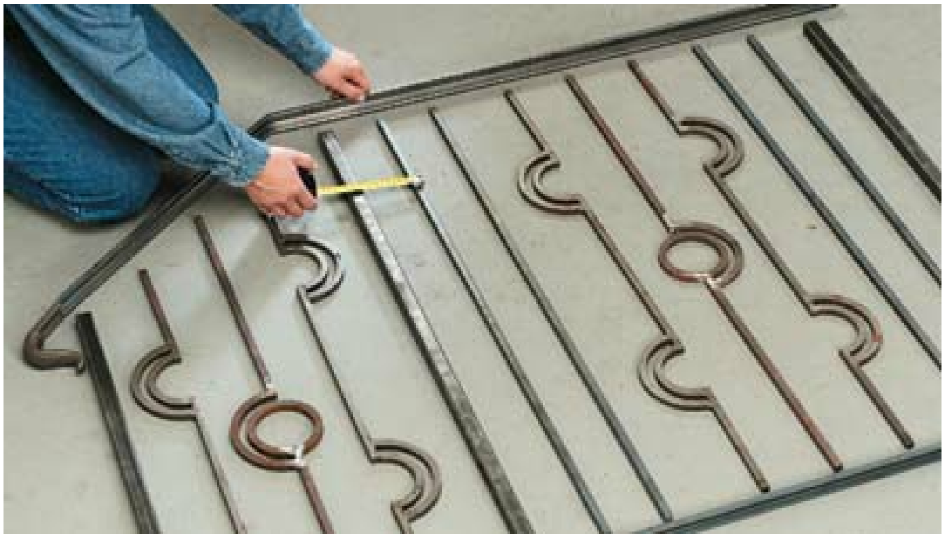

Lay out the pickets and newel posts in a pleasing arrangement, keeping them no more than 5½" apart.

Gate

This gate framework can be used to hold a collection of found metal objects. Use brazing or braze welding to join non-matching metals or thicknesses of metals. Rusty objects will need to be cleaned at the point of contact. If the objects you find are large, you may want to space the uprights at 8", rather than 4". The 4" spacing meets code requirements for gates and fences—this spacing prevents children from getting their heads stuck between uprights.

Materials

- 16 gauge 1 × 1" square tube (17½ feet)
- 16 gauge ½ × ½" square tube (30 feet)
- 4" weldable barrel hinges (2)
- Gate latch hardware
- Found objects
- Wood and brass spacers

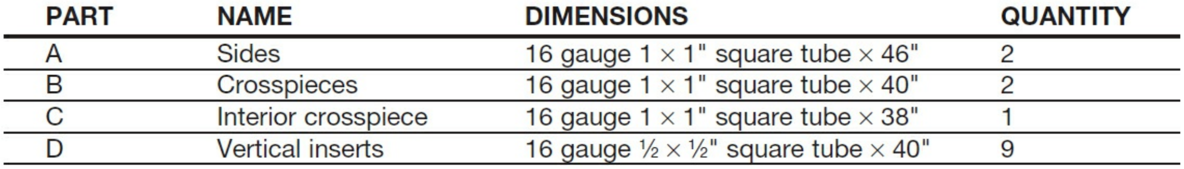

PART	NAME	DIMENSIONS	QUANTITY
A	Sides	16 gauge 1 × 1" square tube × 46"	2
B	Crosspieces	16 gauge 1 × 1" square tube × 40"	2
C	Interior crosspiece	16 gauge 1 × 1" square tube × 38"	1
D	Vertical inserts	16 gauge ½ × ½" square tube × 40"	9

Alternate styles

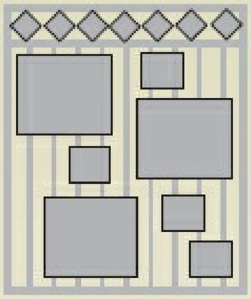

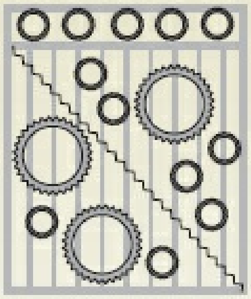

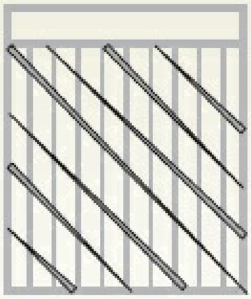

Overall dimensions
40 × 1 × 46"

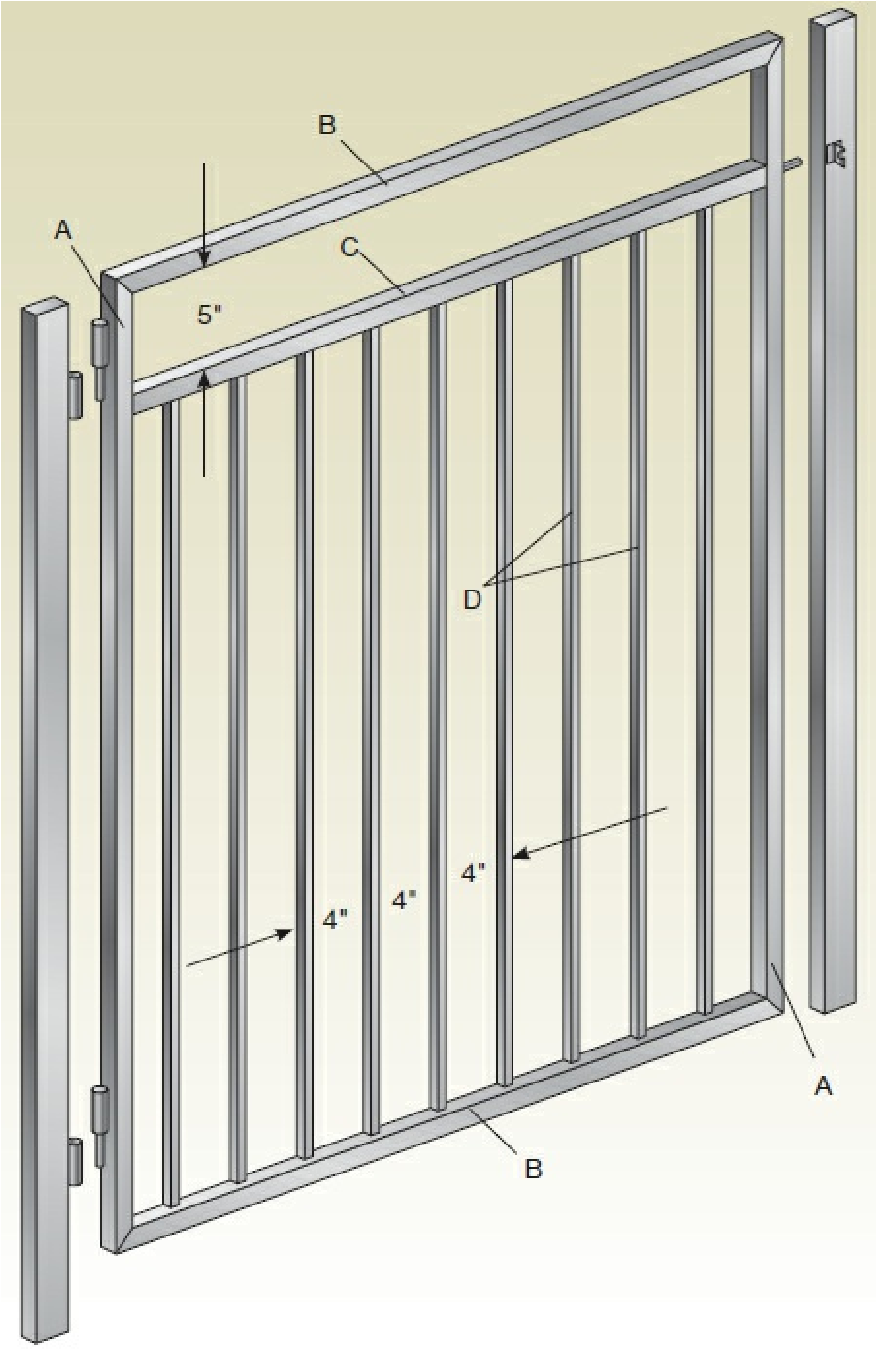
B
A
C
5"
D
4"
4"
4"
A
B

How to Make a Gate

Assemble the Frame

1. Cut the sides and crosspieces (A & B) to size, mitering the ends at 45°.

2. Place the left side piece and bottom crosspiece together at a 90° angle to form one corner of the rectangle. Check the pieces for square, and clamp in place. Tack weld the pieces together.

3. Place the right side piece and top crosspiece together to form another corner of the rectangle. Check for square, clamp in place, and tack weld together.

4. Join the two pieces to form the rectangle. Check the corners for square and clamp in place. Tack weld the corners together.

5. Measure the diagonals of the rectangle to check for square. If the measurements of both diagonals are equal, the assembly is square. If it is not square, pull or push it into alignment. When aligned, clamp it in place, and finish the corner welds.

6. Turn the assembly over, and complete the welds.

Attach the Interior Crosspiece & Inserts

1. Cut the interior crosspiece (C) and vertical inserts (D) to size.

2. Place the interior crosspiece against the inside edges of the side pieces 5" down from the top crosspiece. Check the pieces for square, and clamp in place. Weld the interior crosspiece to the sides.

3. Place the vertical inserts at 4" intervals between the interior and bottom crosspieces. Make sure the spacing is even—you might need to adjust to slightly less or more than 4" if the miter cuts are slightly off. Keep the outside edge of the inserts flush with the outside edge of the crosspieces.

4. When the spacing is adjusted properly and the inserts are square to the crosspieces, tack weld each insert at both ends.

5. Check the assembly and inserts for square one more time. Turn the assembly over, and weld each upright in place.

Attach the Found Objects

Objects other than mild steel will need to be brazed or braze welded. Connect non-metallic objects by wrapping or folding a thin strip of mild steel sheet metal or a short piece of 1/8" steel rod around an edge. Weld the ends of the rod or strap to the framework.

1. Arrange your found objects artfully across the interior space.

2. Carefully clean rust or paint from the areas where the found objects contact the uprights.

3. Weld the found objects in place.

Attach the Hinges

Our gate is made to hang attached to a metal gate post. We chose to use barrel style hinges.

1. Place the gate between the gate posts. Use wood spacers and braces to position the gate between the gate posts, and clamp or brace solidly in place.

2. Line up the hinges on the post and the gate (see photo). Use a level to check for plumb. (For the gate to swing properly, the hinges need to be installed perfectly plumb.) Tack weld the hinges to the gate and post.

3. Remove the bracing, and check that gate swings freely. When it does, complete the hinge welds.

4. Install the gate latch hardware. If the hardware is painted or zinc coated, grind off the coating before welding, or install with screws.

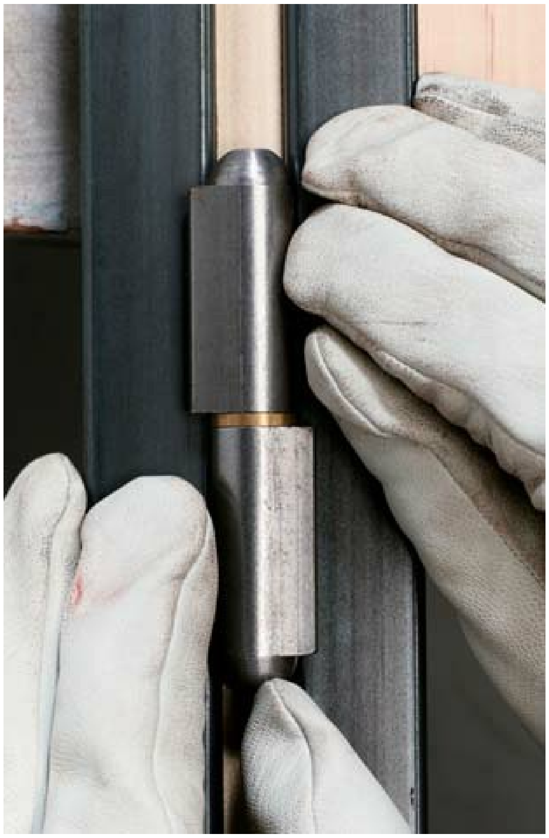

Brace the gate in position between the gate posts. Clamp a wood spacer between the gate post and the gate side. Position the hinge and check for plumb.

Firepit

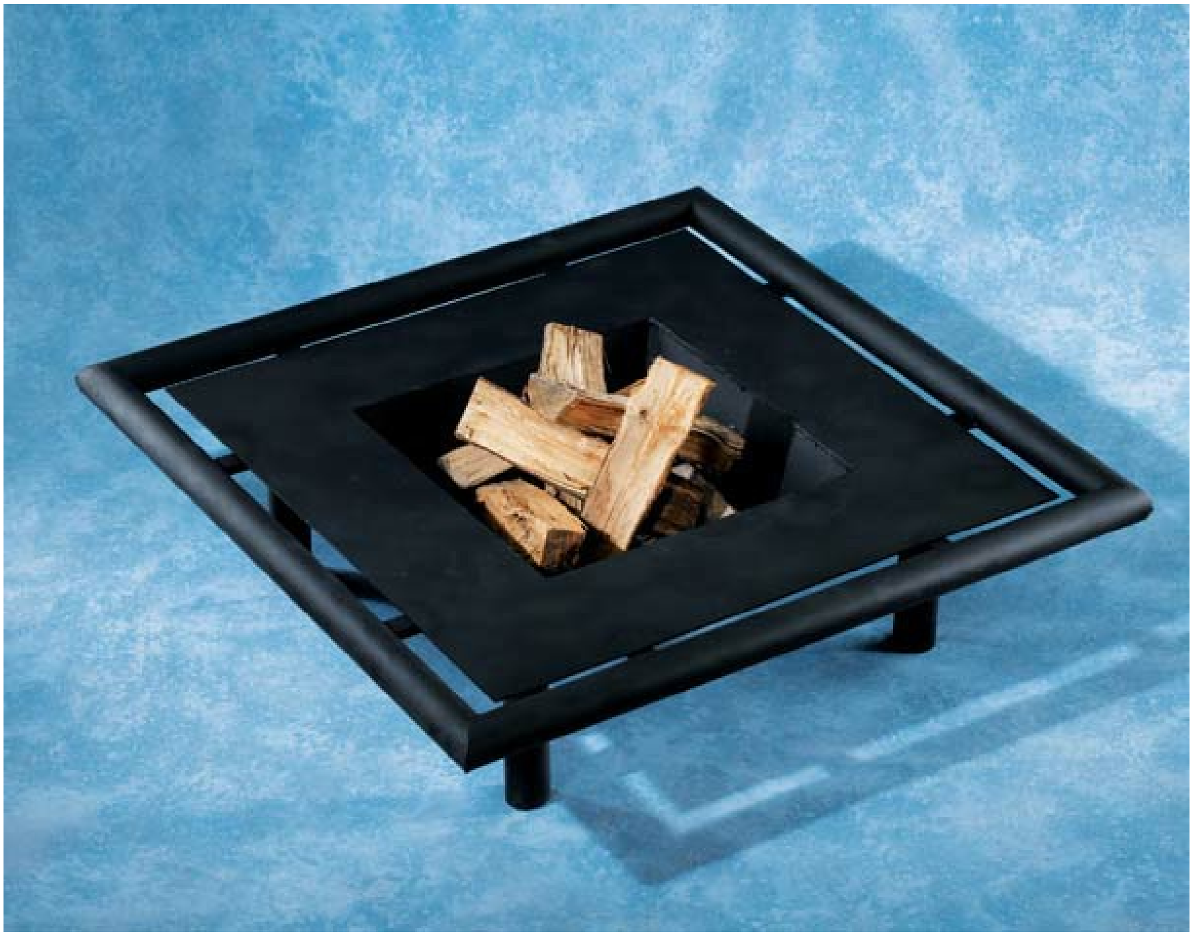

All kids and most adults love having a fire in the backyard. Here’s a firepit that takes into account safety precautions, and it is easy to build. The deep box holds large chunks of wood, and the wide table lip minimizes scattering sparks. The footrests are great for kicking back by the fire, and they also help keep small hands away from the heat. If you want a spark screen, Smith & Hawken sells a clamshell screen on its website (see Resources). This firepit has been sized to use such a screen.

Materials

- 1/8" sheet metal (30 × 30", 6 × 60½")
- 1/8 × ½ × 1" channel × 4 feet
- 16-gauge 2" round tube (15½ feet)
- Spark screen (Smith & Hawken #749622, optional)

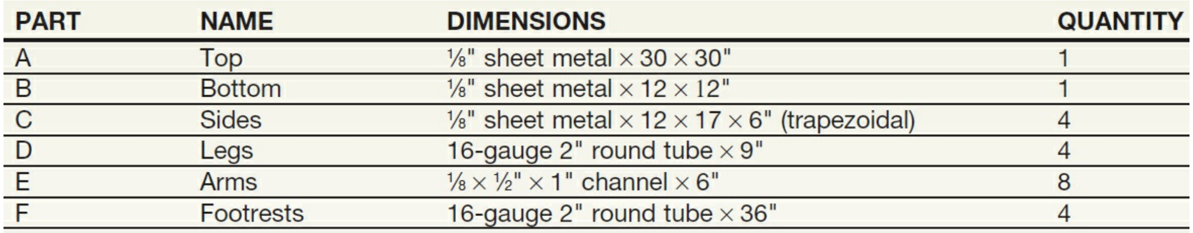

PART	NAME	DIMENSIONS	QUANTITY
A	Top	⅛" sheet metal × 30 × 30"	1
B	Bottom	⅛" sheet metal × 12 × 12"	1
C	Sides	⅛" sheet metal × 12 × 17 × 6" (trapezoidal)	4
D	Legs	16-gauge 2" round tube × 9"	4
E	Arms	⅛ × ½" × 1" channel × 6"	8
F	Footrests	16-gauge 2" round tube × 36"	4

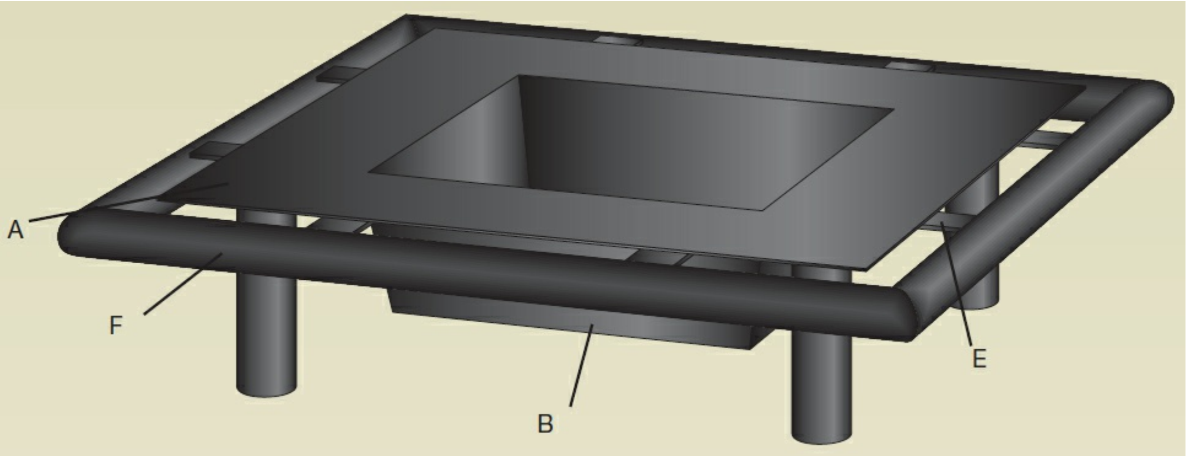
A
F
B
E

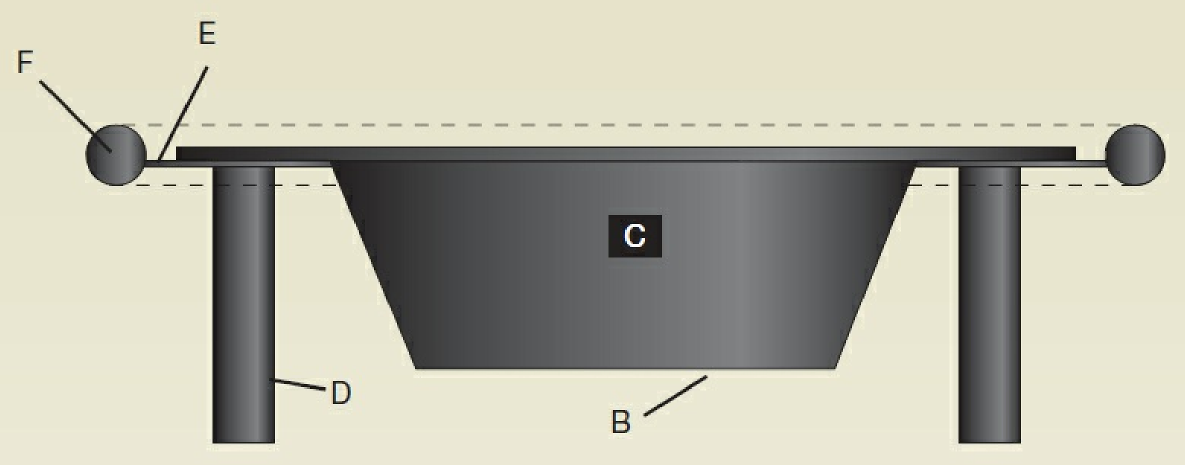
F
E
C
D
B

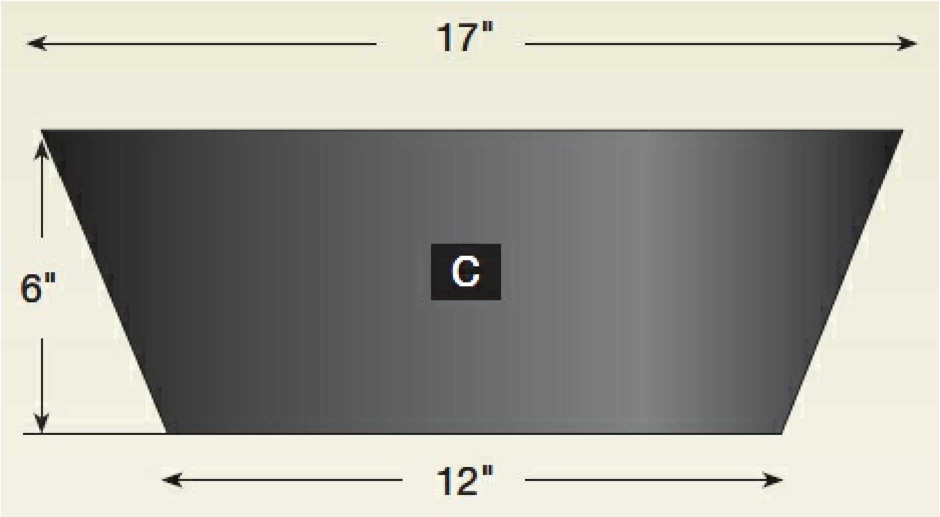
17"
6"
C
12"

How to Build a Firepit

Before welding, clean all parts thoroughly with denatured alcohol.

Make the Top

1. Cut the top (A) to size.

2. Mark a straight line 6½" in from each side. This will create a center square.

3. Using a plasma cutter or acetylene torch, cut the inside square out of the top.

Make the Box

1. Cut the bottom (B) to size by cutting down the scrap from the top. Mark in 5" from two sides of the scrap to form the 12" square. Cut along the lines.

2. On a 6 × 60½" piece of sheet metal, mark one long side at 17", 29", 46", and 58". Mark the other side at 2½", 14½", 31½", and 43½". Use a straightedge and marker or soapstone to draw lines from the corner to the 2½" mark, from the 14½" mark to the 17" mark and so on to create four trapezoids with the parallel sides being 17" and 12". Cut the sides (C) to size.

3. Tack weld a side to the bottom. Tack weld an adjacent side to the bottom. Bend the sides so their seam aligns, if necessary, and tack-weld. Continue with the other sides.

4. Turn the box over and weld all the seams.

5. Place the box over the cutout in the table and tack weld in place.

6. Complete all the welds.

On the scrap piece cut from the top, mark 5" in from two edges to create a 12" square.

Make the Legs

1. Cut the legs (D) to length.

2. Place the legs 2" from each corner of the box. Weld in place.

Make the Footrest

1. Cut the footrests (F) to length, mitering each end at 45°. Make sure the miters are perfectly aligned. Use a clamping jig to prevent rolling.

2. Lay out the footrests to form a square. Tack weld the corners.

3. Turn the table and fire box upside down on top of 1" spacers. Place the footrest around the top. Align the footrest evenly with the table edges.

4. Cut the arms (E) to length.

5. Place the arms so they touch the footrest and are aligned with the box edges.

6. Tack weld the arms to the footrest. Tack weld the arms to the table underside. Make sure the footrest is still aligned with the top.

7. Complete all the welds for the arms and footrests.

Finish the Project

1. Grind down all welds.

2. Wire brush and clean fire pit.

3. Finish with Rust-Oleum High Heat spray enamel.

Tack weld the sides to the base from the inside. Then turn the pit over and weld all the seams.

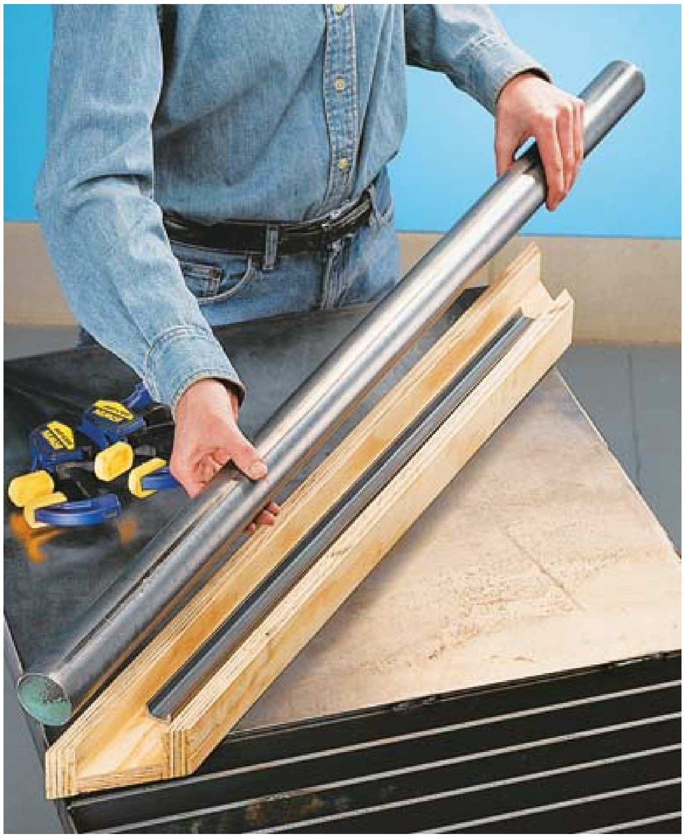

Creating a jig using a piece of channel iron and plywood makes cutting the 45° angles easier.

Vineyard Trellis

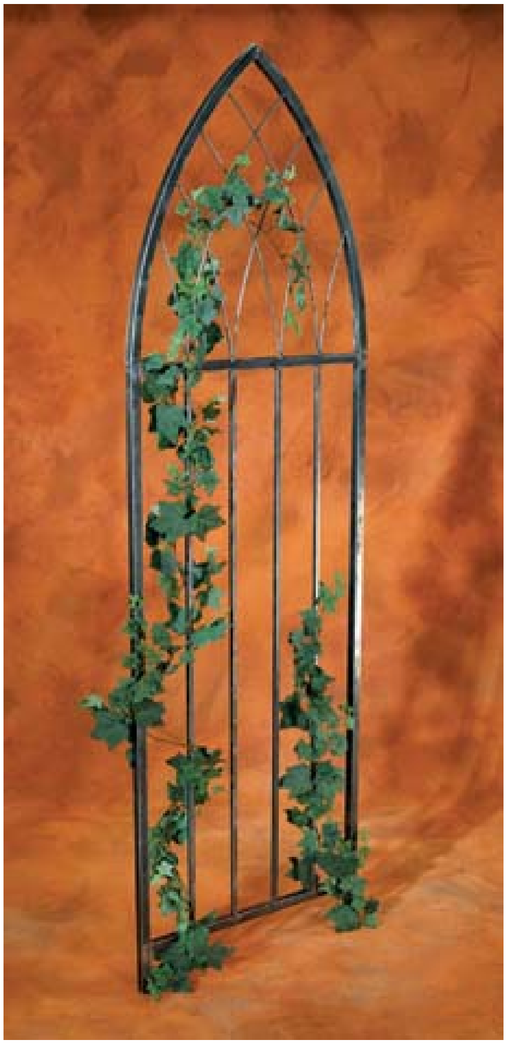

This arbor is a beautiful stand- alone piece, but it is also a brilliant addition to the Arch & Gate. Either way it is an impressive addition to your yard. Mount it against a wall or fence, or use it as a panel within a fence. Better yet, make an entire fence of trellises! To do so, screw mount the panels to 4 × 4 wood posts, or weld the panels to 3 × 3 steel posts.

You may want to cap off your trellis with a finial or decorate the panels with permanent trailing vines for the off season. See Resources for finial, frieze, and stamped metal sources. Make sure to order parts that can be welded. Cast iron cannot be easily welded to the steel trellis parts.

Materials

- 16-gauge 1 × 1" square tube (23 feet)
- 16-gauge ½ × ½" square tube (18 feet)
- ¾" round rod (6 feet)
- 1⁄8 × ½ × ½" angle iron (4")
- ¼" round rod (18 feet)

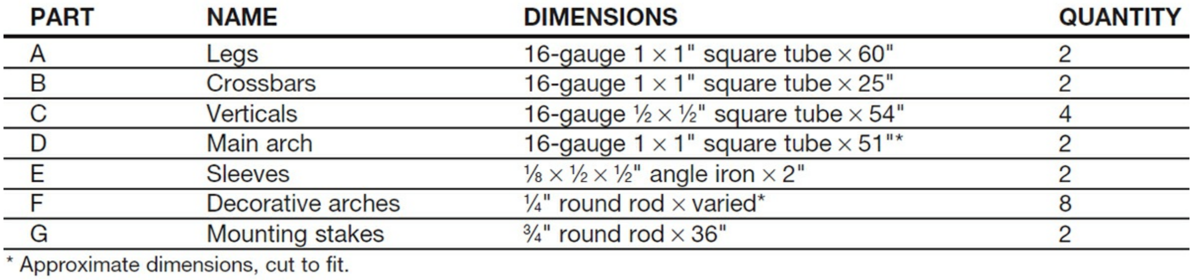

PART	NAME	DIMENSIONS	QUANTITY
A	Legs	16-gauge 1 × 1" square tube × 60"	2
B	Crossbars	16-gauge 1 × 1" square tube × 25"	2
C	Verticals	16-gauge ½ × ½" square tube × 54"	4
D	Main arch	16-gauge 1 × 1" square tube × 51"*	2
E	Sleeves	⅛ × ½ × ½" angle iron × 2"	2
F	Decorative arches	¼" round rod × varied*	8
G	Mounting stakes	¾" round rod × 36"	2

* Approximate dimensions, cut to fit.

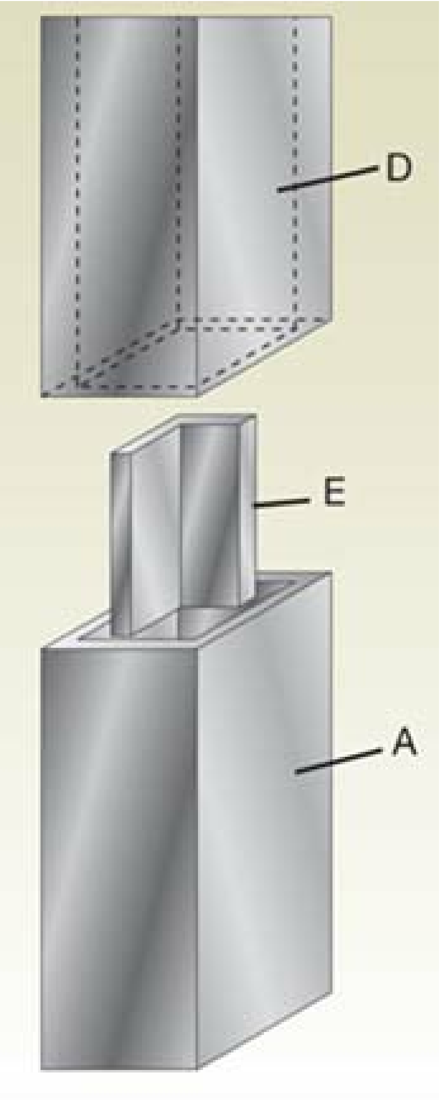

Sleeve detail

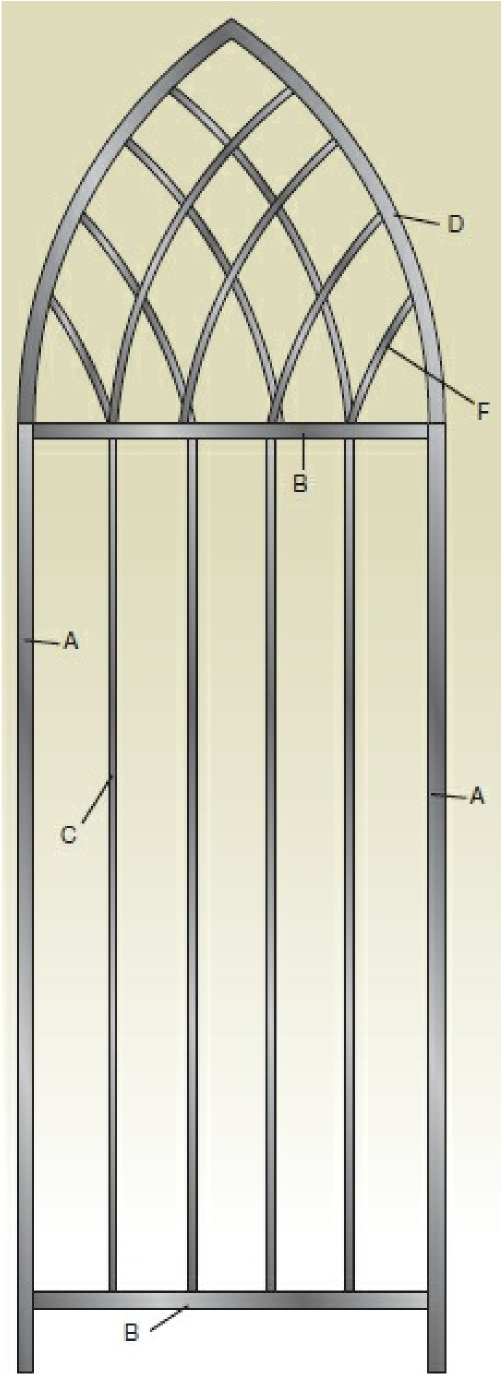
D
F
B
A
C
A
B

How to Build a Vineyard Trellis

Before welding, clean all parts thoroughly with denatured alcohol.

Make the panel

1. Cut the legs (A) and crossbars (B) to length.

2. Lay out the crossbars between the legs. Align the top crossbar with the top of the legs. Place the lower crossbar 4" from the bottom of the legs.

3. Check for square, clamp in place, and tack weld together.

4. Measure across the diagonals to check for square. If the diagonals are not equal, adjust until they are. Complete the welds.

5. Cut the verticals (C) to length.

6. Mark the crossbars at 5", 10", 15", and 20". Center the verticals at the marks, check for square, and weld in place.

Bend the Main Arches

1. Cut a 28" radius semicircle from ½" plywood. Cut the semicircle in half.

2. Stack the quarter circles and screw them onto a 3 × 3 sheet of plywood. Screw 2 × 4 stops along the bottom.

3. Clamp this bending jig to a stable work surface. Slide an arch piece (D) between the stops so it extends 2" below the bottom of the semicircle. Bend the tube around the curve. Bend the second arch in the same manner.

Finish the Main Arches

1. Remove the quarter circles from the bending jig. From the right angles of the quarter circles, measure in 17½" along one flat side. At this point, draw a line perpendicular to the side. Cut along the line. Repeat with the second quarter circle.

2. Mount these two arcs at the top of a 4 × 8 sheet of plywood, abutting them together on the new cuts to create a template for the arch.

3. Place one arch into the left side of the jig. Mark a vertical line on the arch to match the joint of the two wood arcs. Cut the arch along the line.

4. Place the other arch into the right side of the jig. Mark the centerline and cut.

5. Place both arches back into the jig and weld together at the peak.

6. Cut the sleeves (E) to length. Slide a sleeve 1" into each end of the arch. Slide the top of the panel over the sleeves. Weld the arch to the legs.

Make the Decorative Arches

1. Attach a stop block ½" from the base of the arch jig. Bend a decorative arch piece (F) around the jig to make a curve. Continue bending curves as needed.

2. At the 5" mark on the top crossbar, center one end of the arch piece. Align the arch parallel to the main arch. Mark the arch where it crosses the main arch. Cut on this line. At both ends, tack weld in place.

3. Repeat step 2 at the 10", 15", and 20" marks.

4. Repeat the process of placing, cutting, and welding arches for arches curving in the opposite direction. Finish all welds.

5. To prevent rattling, tack weld the arches to each other where they cross.

Finish the Arbor

1. Wire brush or sandblast the arbor. Finish as desired.

2. Pound the mounting stakes (G) 2 ft. into the ground, 26" apart. Slide the legs over the mounting stakes.

3. Using wire, secure the top or sides of the arbor to a wall or fence.

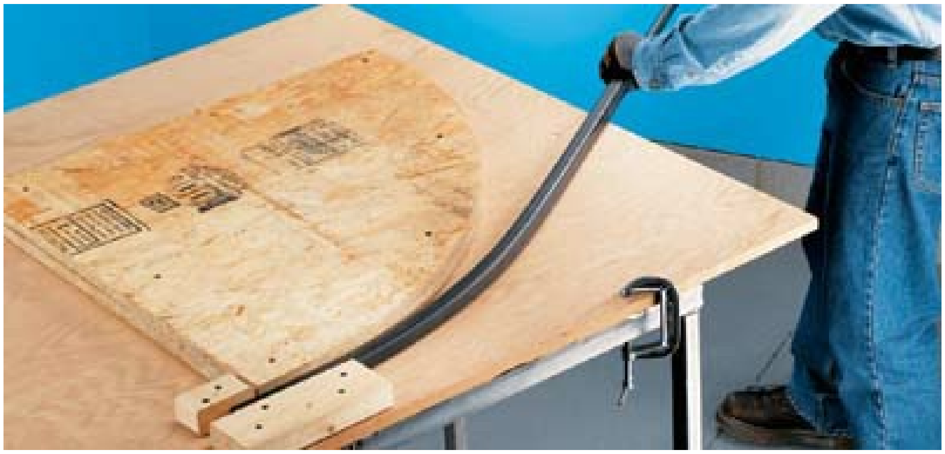

Bend the arch around the jig. Insert a ½" black pipe into the tube end to create more leverage.

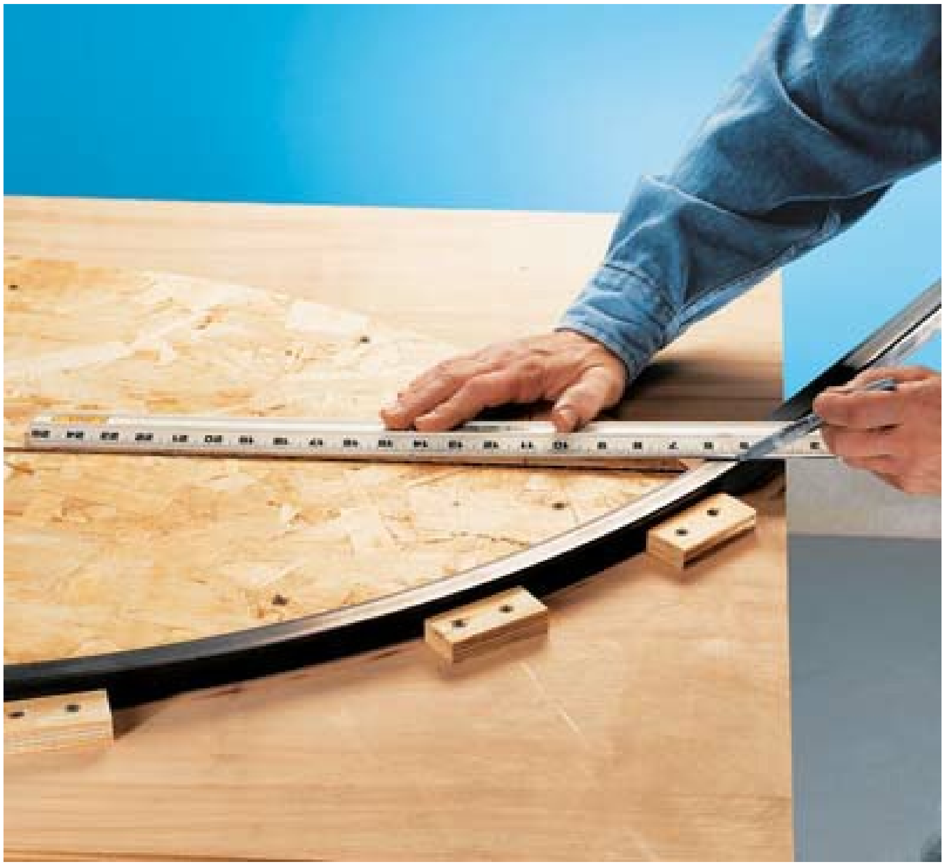

Place an arch in the second jig and mark where the arc crosses the vertical line between the two half circles.

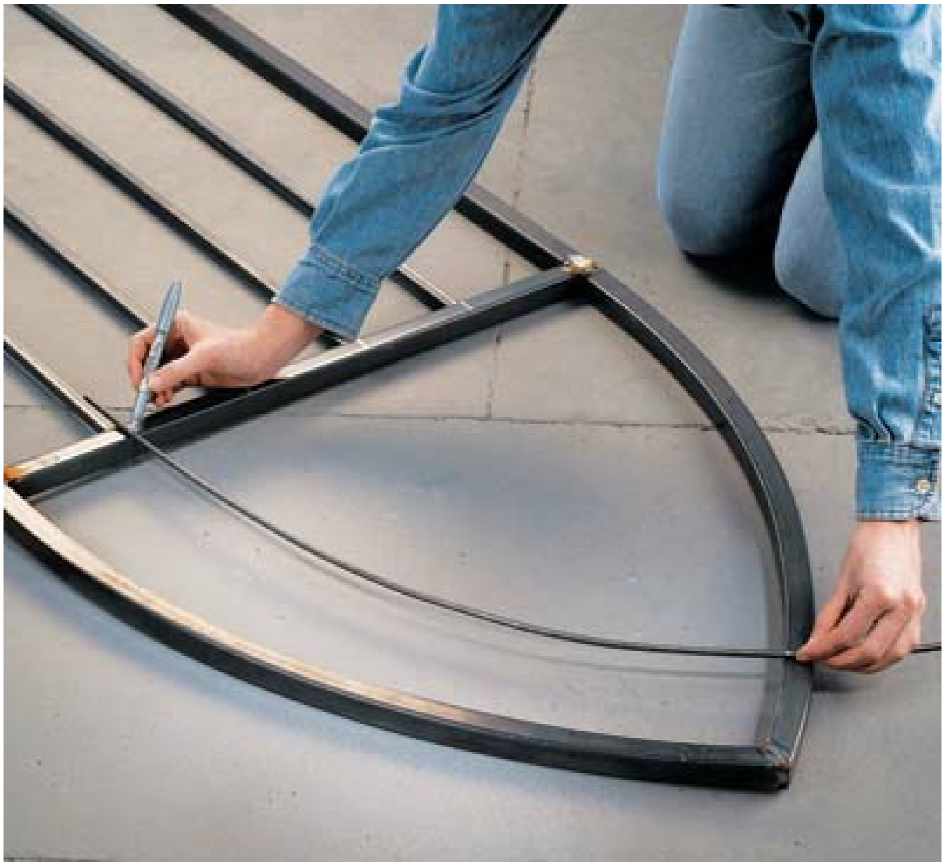

After bending the decorative arches, mark where they cross the crossbar and main arches.

Arch & Gate

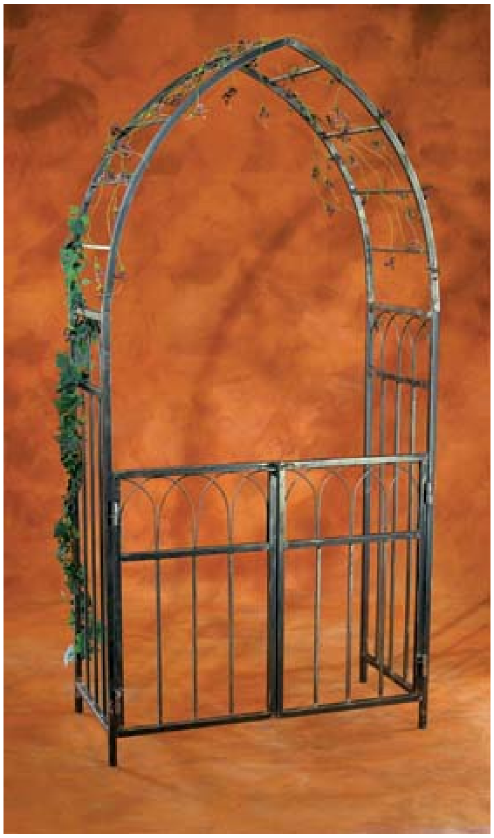

What a grand entryway this arch and gate provides! Unlike many manufactured archways, this one stands a full 8 feet tall, allowing easy passage underneath and plenty of room for trailing foliage. It is proportioned to fit over a 42" walkway. The arching pattern is beautiful and fairly easy to create. When paired with the matching vineyard trellis, this project will really add class to your yard or garden.

Materials

- 16-gauge ½ × ½" square tube (48½ feet)
- 16-gauge 1 × 1" square tube (70 feet)
- 1⁄8 × ½ × ½" angle iron (8")
- ¼" round rod (48 feet)
- ¾" round tube (8 feet)
- 2" weldable barrel hinges (2 pair, Decorative Iron #14.1120)
- Gate latch (1, Decorative Iron #14.2012)
- ½" plywood (2, 4 × 8) sheets

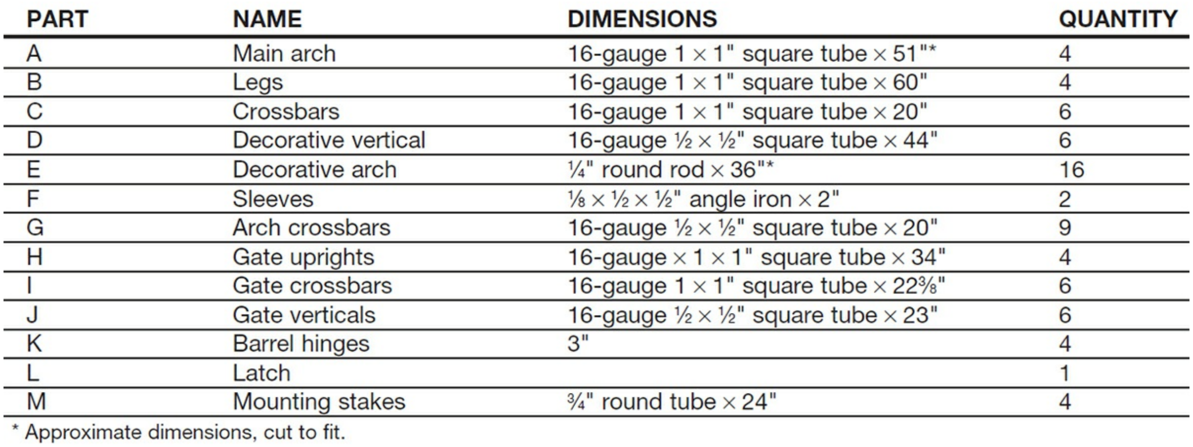

PART	NAME	DIMENSIONS	QUANTITY
A	Main arch	16-gauge 1 × 1" square tube × 51"*	4
B	Legs	16-gauge 1 × 1" square tube × 60"	4
C	Crossbars	16-gauge 1 × 1" square tube × 20"	6
D	Decorative vertical	16-gauge ½ × ½" square tube × 44"	6
E	Decorative arch	¼" round rod × 36"*	16
F	Sleeves	⅛ × ½ × ½" angle iron × 2"	2
G	Arch crossbars	16-gauge ½ × ½" square tube × 20"	9
H	Gate uprights	16-gauge × 1 × 1" square tube × 34"	4
I	Gate crossbars	16-gauge 1 × 1" square tube × 22⅜"	6
J	Gate verticals	16-gauge ½ × ½" square tube × 23"	6
K	Barrel hinges	3"	4
L	Latch		1
M	Mounting stakes	¾" round tube × 24"	4

* Approximate dimensions, cut to fit.

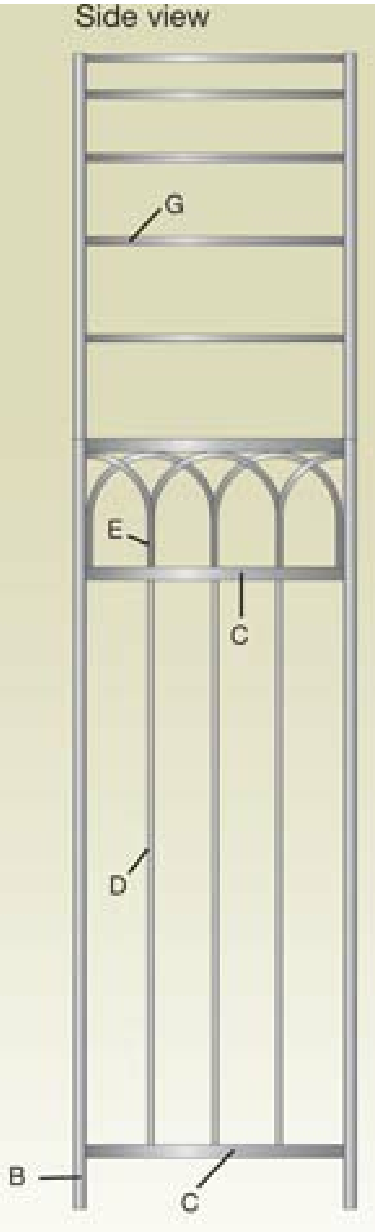
Side view
G
E
C
D
B
C

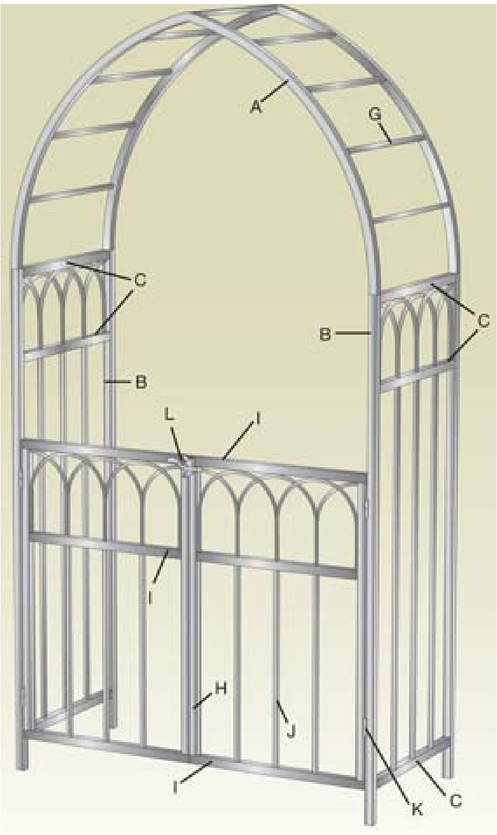
A
G
C
C
B
B
L
I
I
H
J
I
K
C

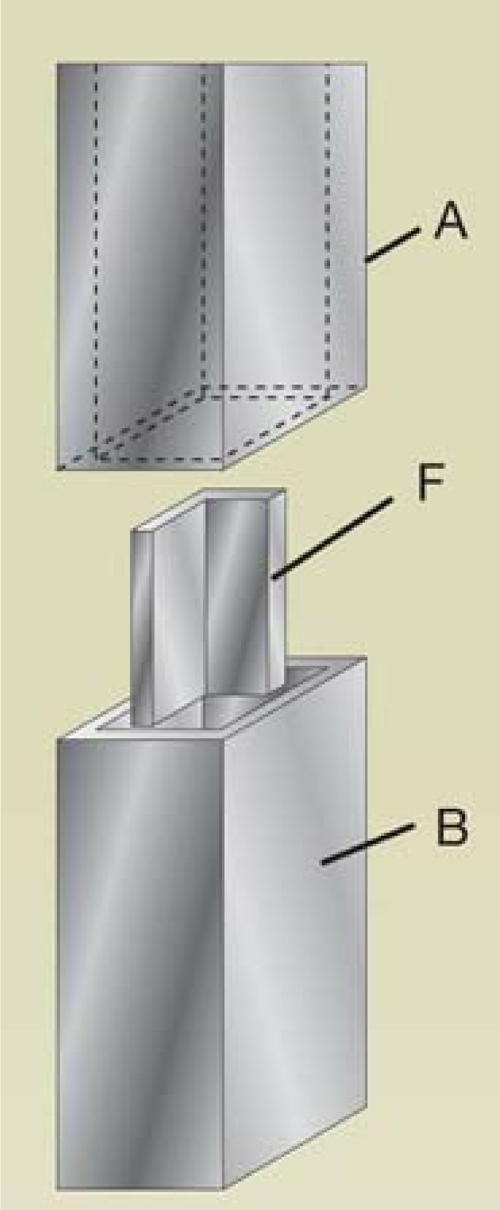
Leg arch detail

How to Build an Arch Gate

Before welding, clean all parts thoroughly with denatured alcohol.

Make the Main Arches

1. Cut a 28" radius semicircle from ½" plywood. Cut the semicircle in half.

2. Stack the quarter circles and screw them onto a 3 × 3 sheet of plywood. Screw 2 × 4 stops along the bottom.

3. Clamp this bending jig to a stable work surface. Slide an arch piece (A) between the stops, so it extends 4" below the bottom of the semicircle. Bend the tube around the curve. Bend the remaining arches.

Make the Side Panels

1. Cut the legs (B) and crossbars (C) to length.

2. Lay out two legs and three crossbars on a flat surface. Align the top crossbar flush with the top of the legs and the base of the bottom crossbar 4" from the bottom of the legs. Align the top of the third crossbar 10" down from the top of the legs. Check for square, clamp in place, and tack weld. Repeat to form the second side panel.

3. Before cutting the decorative verticals (D), measure between the bottom and middle crossbars for an exact length. Cut the decorative verticals to length. Mark all three crossbars at 5", 10", and 15".

4. Center the verticals at the marks. Check for square and weld in place. Finish all welds.

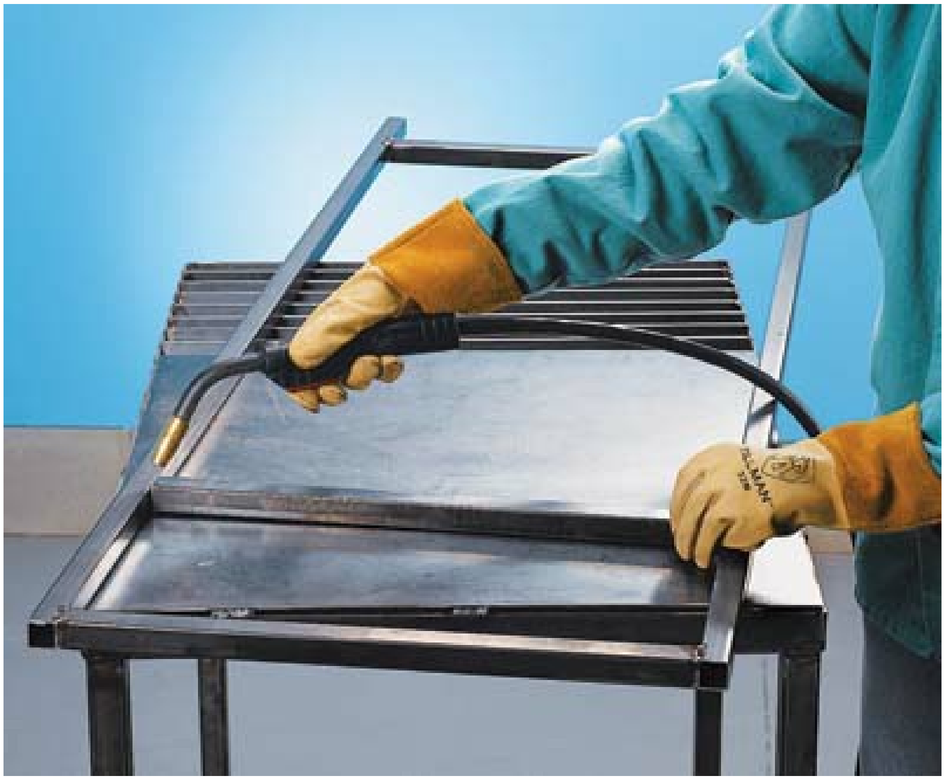

Weld the crossbars to the legs.

Create the Decorative Arches

1. Make a bending jig by attaching a 9" diameter half circle and 10" long squares to a 15 × 19" rectangle. Slide a decorative arch piece (E) between the stops and bend it around the curve.

2. Place two hoops between the top and middle crossbars aligned with the legs and the 10" mark. Mark where the legs cross the middle crossbar, cut and weld in place.

3. Place a hoop with one leg at the 5" mark and the other at the 15" mark. Mark, cut, and weld (see photo, middle right).

4. Cut a hoop in half and place the halves starting from the 5" and 15" weld in place.

5. Repeat steps 2 to 4 for the other side panel.

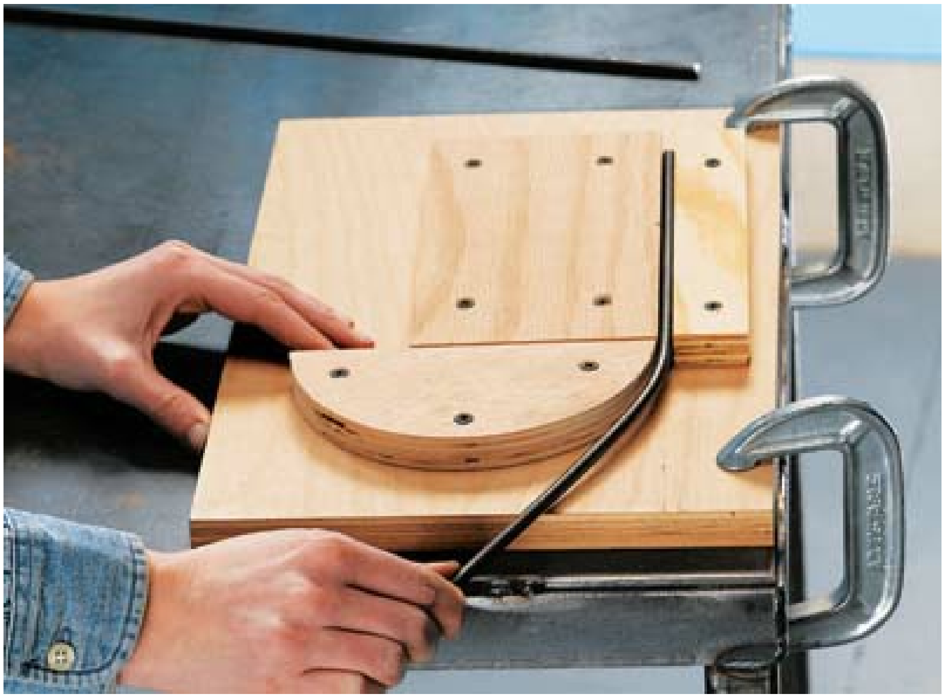

Make a bending jig to create the hoops for the decorative arches.

Finish the Arches

1. Remove the quarter circles from the bending jig. From the right angles of the quarter circles, measure in 7½" along one flat side. At this point, draw a line perpendicular to the side. Cut along the line. Repeat with the second quarter circle.

2. Mount these two arches at the top of a 4 × 8 sheet of plywood, abutting them together on the new cuts to create a template for the arches. Place stop blocks 1" from the arch.

3. Place one arch into the left side of the jig. Mark a vertical line on the arch to match the joint of the two wood arches. Cut the arch along the line.

4. Place the other arch into the right side of the jig. Mark the centerline and cut.

5. Place both arches back into the jig and weld together at the peak. Repeat steps 3 and 4 for the second arch. Remove the arch jig from the plywood sheet.

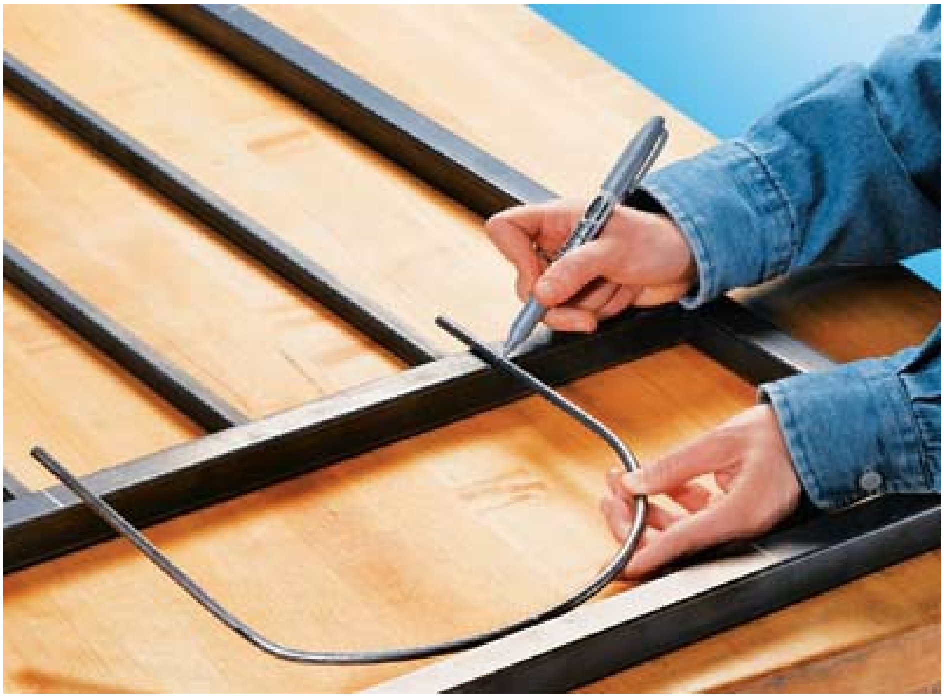

Align the top of the hoop with the top crossbar and mark where the legs cross the middle crossbar.

Assemble the Archway

1. Cut the sleeves (F) to length. Cut the arch crossbars (G) to length.

2. Mark two parallel lines 45" apart on the sides of the plywood. Attach 2 × 4 clamping blocks along these lines.

3. Clamp the two side panels, lying on their backs, to the jig.

4. Slide sleeves into the legs 1", and weld in place.

5. Slide an arch over the sleeves and weld in place. Repeat with the second arch.

6. Weld a crossbar (G) between the two arches at the peak. Weld crossbars spaced 9" apart on the sides of the arches.

After bending the main arches, remove the quarter circles from the jig. Mark in 7½" from one flat edge and cut the form down.

Make the Gate Frame

1. Cut the gate uprights (H) and crossbars (I) to length.

2. Lay out the two gate frames with the crossbars on the top and bottom of the uprights. Place the third crossbar 10" from the top of the uprights. Check for square and tack weld together. Check the diagonals and adjust if necessary. Finish the welds.

3. Cut the gate verticals (J) to length. Mark all three crossbars at $5^3/_{16}$", $10^3/_{16}$", and $15^3/_{16}$". Center the verticals at the marks. Check for square and weld in place (see center photo, opposite page).

4. Make and assemble the gate decorative arches using the same method as described under Create the Decorative Arches on page 197.

Install the Gate

1. Stand the archway upright. Using scrap metal or wood, clamp the archway legs so that the space between the legs is 45" and the panels are parallel.

2. Align the gates flush with a set of legs (the legs and gate are flush front to back) and clamp in place. Allow a slight gap between the gates and legs (the gap is between the legs and gate side to side).

3. Tack weld the hinges (K) in place (see bottom photo, opposite page). Remove the clamps holding the gates and check that the gates swing freely. Adjust if necessary and complete the welds. Install the gate latch (L).

Finish the Archway

1. If you plan to finish the archway, sand or wire brush or sandblast it. Apply the finish of your choice.

2. Install the archway by clamping the side panels 45" apart.

3. On the ground where the arch will be installed, mark a 21 × 46" rectangle. At the corners of the rectangle, pound the mounting stakes 1 ft. into the ground. Lift the archway onto the mounting stakes and remove the clamps.

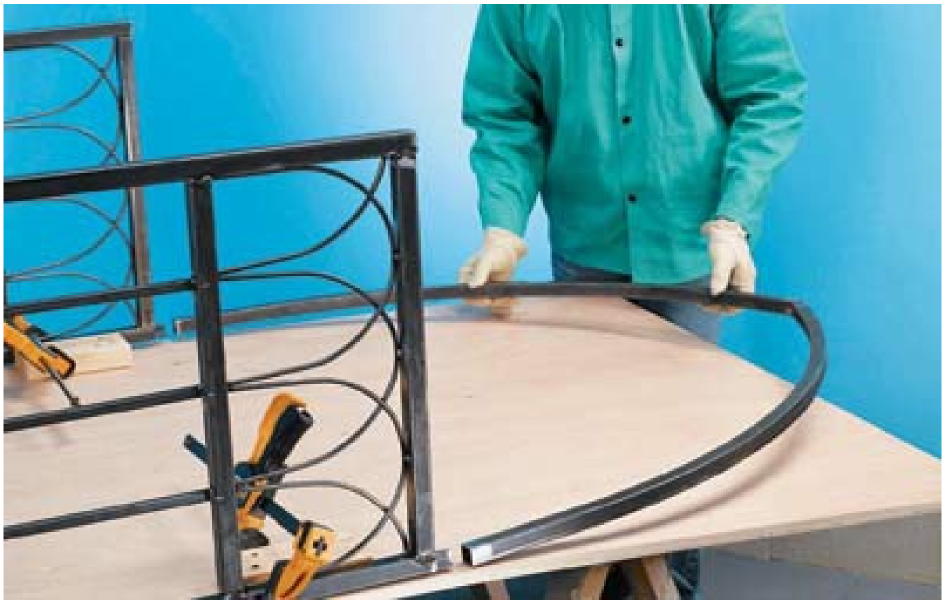

Clamp the sides to clamping blocks on a 4 × 8 sheet of plywood. Slide the main arches over the angle iron sleeves in the legs.

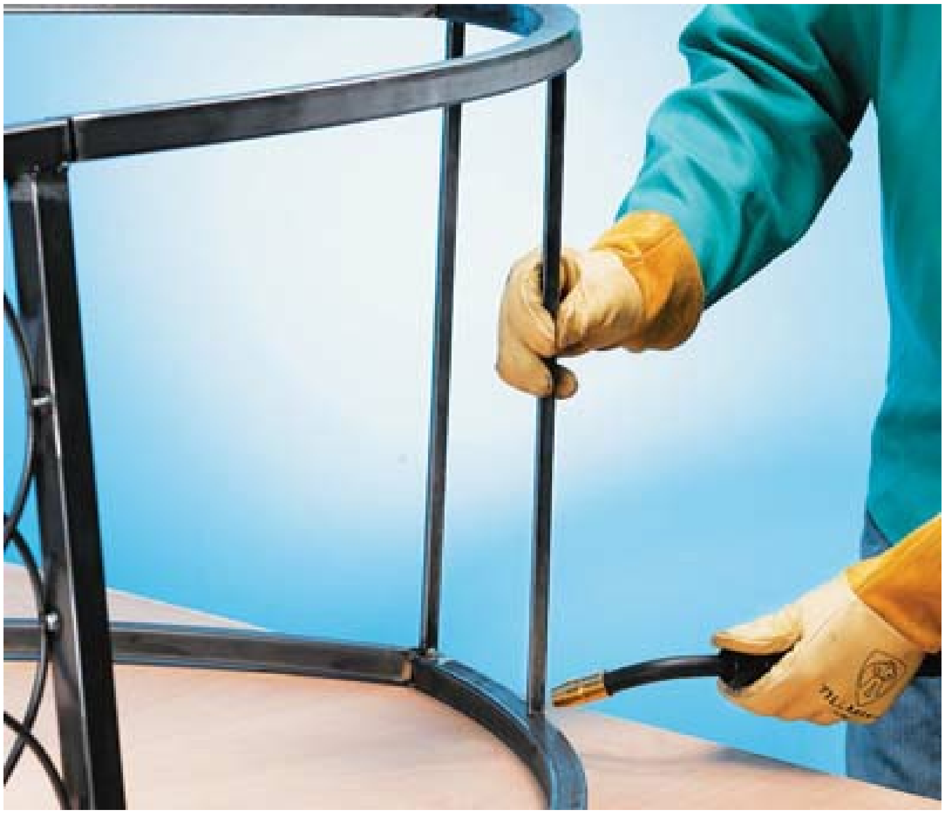

Weld the arch crossbars to the arches.

Center the gate verticals at the marks and weld in place.

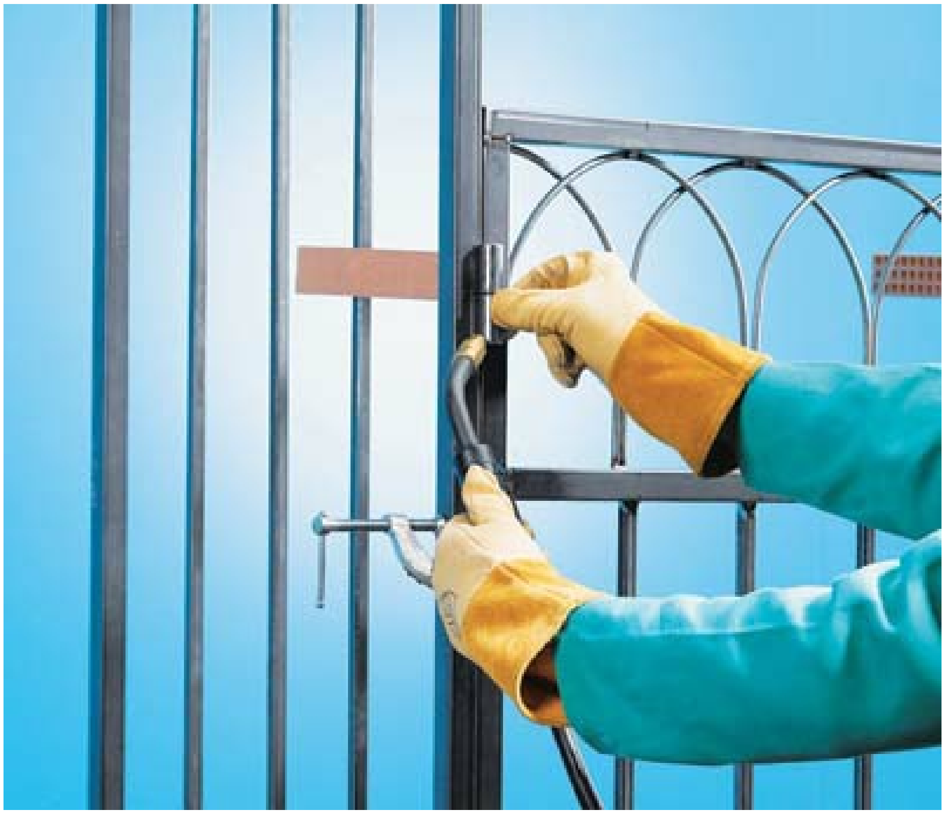

Use shims and clamps to hold the gates in place. Weld the base, pin side of the hinge to the frame, and the top to the gate.

Gazebo

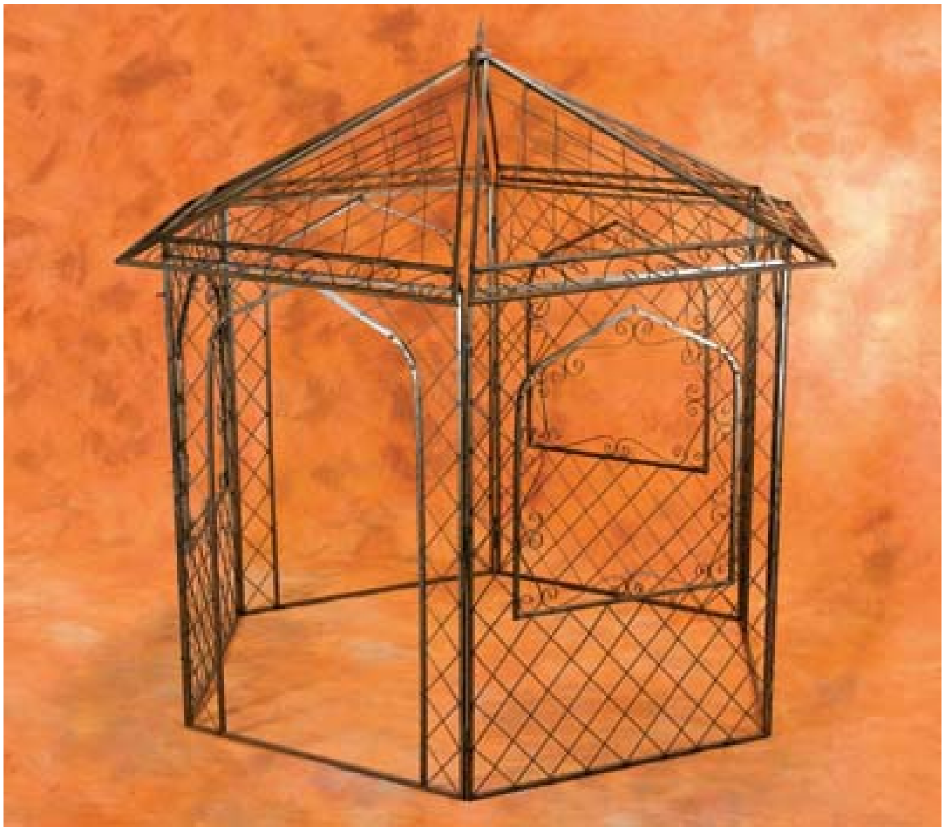

This moderate-size gazebo is also a special addition to any yard. Big enough for a table and chairs, it is small enough to fit in all but the tiniest of yards. Covered with vines in the summer, it provides delightful shade. In the off-season it supplies a point of visual interest with its scrolling patterns.

Want a different scroll pattern? Simply peruse the online catalogs of the companies listed in the Resources on for dozens of alternative patterns.

The directions here describe how to build door, window, and roof panels. The material quantities are for a gazebo with three door panels and three window panels. You can choose to construct your gazebo with any combination, just remember to adjust your quantities list accordingly.

Materials

- 4" finial (1, Decorative Iron #126)
- 1 × 1" square tube (344 feet)
- 3/16" round rod (1,000 feet)
- ½" round tube (6 feet)
- 3/8" round rod (10½ feet)
- ¼" round rod (3 feet)
- 4½" drip trays (2)
- 2" round tube (2")
- ½ × 6" carriage bolt, nut, and washer (1)
- ¾" round rod or rebar (9 feet)

PART	NAME	DIMENSIONS	QUANTITY
A	Legs	16-gauge 1 × 1" square tube × 80"	12
B	Door frames	16-gauge 1 × 1" square tube × 96"	6
C	Top crossbars	16-gauge 1 × 1" square tube × 54"	6
D	Bottom crossbars	16-gauge 1 × 1" square tube × 52"	6
E	Window frames	16-gauge 1 × 1" square tube × 40"	6
F	Window crossbars	16-gauge 1 × 1" square tube × 30"*	6
G	Roof sides	16-gauge 1 × 1" square tube × 66½"	12
H	Roof bases	16-gauge 1 × 1" square tube × 61"	6
I	Roof crossbars	16-gauge 1 × 1" square tube × 61"*	6
J	Hinge barrels	16-gauge ½" round tube × 3"	24
K	Hinge pins	⅜" round rod × 7½"	12
L	Finial pegs	¼" round rod × 6"	6
M	Finial sleeve	16-gauge 2" round tube × 2"	1
N	Finial	4"	1
O	Drip trays	4½"	2
P	Roof pegs	⅜" round rod × 3"	12
Q	Peg sleeves	16-gauge ½" round tube × 3"	12
R	Window scrolls	3⁄16" round rod × 24"	30
S	Roof scrolls	3⁄16" round rod × 36"	24
T	Diagonals	3⁄16" round rod × varied*	624

* Cut to fit.

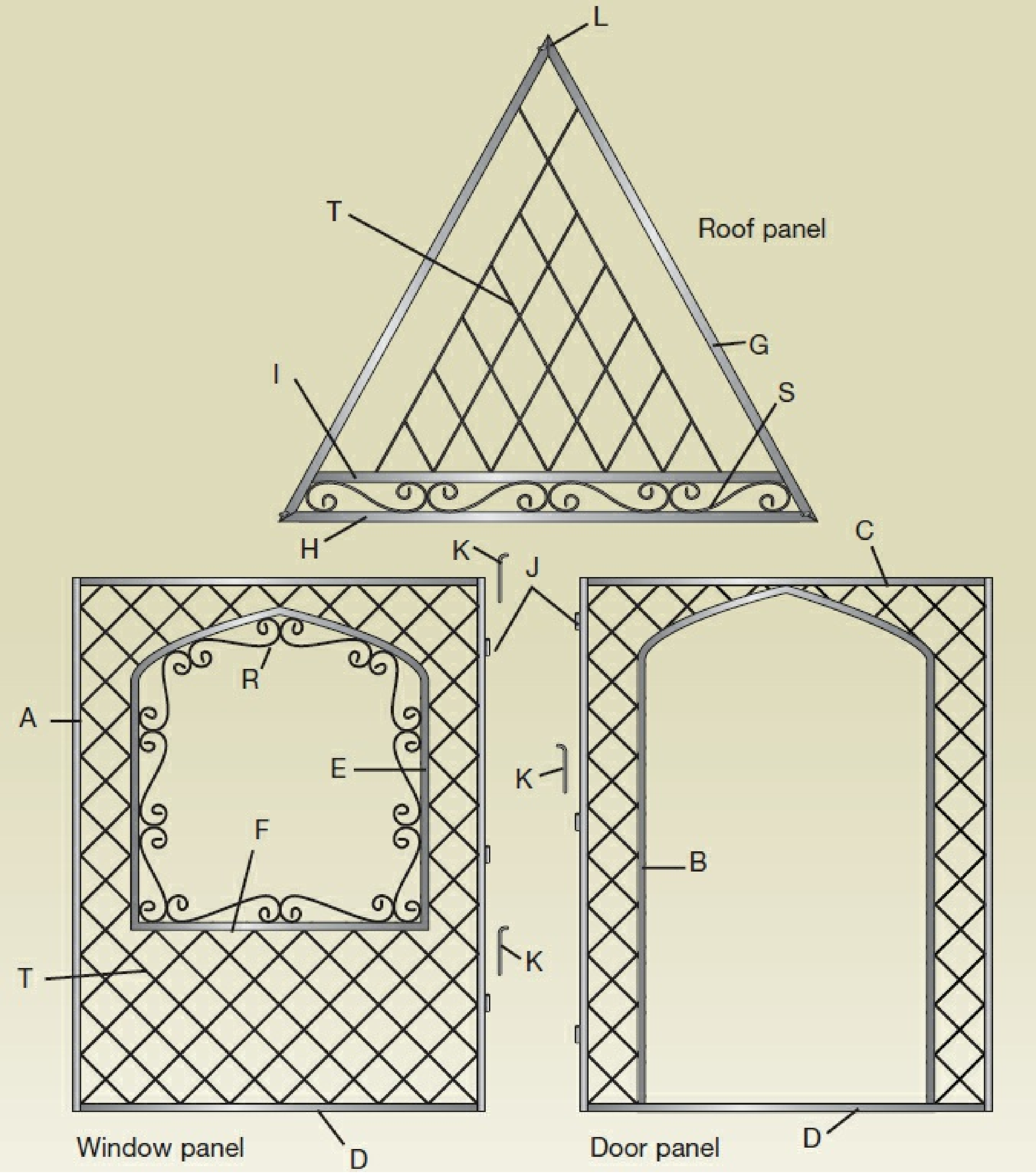
L
T
Roof panel
G
I
S
H
K
J
C
R
A
E
K
F
B
T
K
Window panel
D
Door panel
D

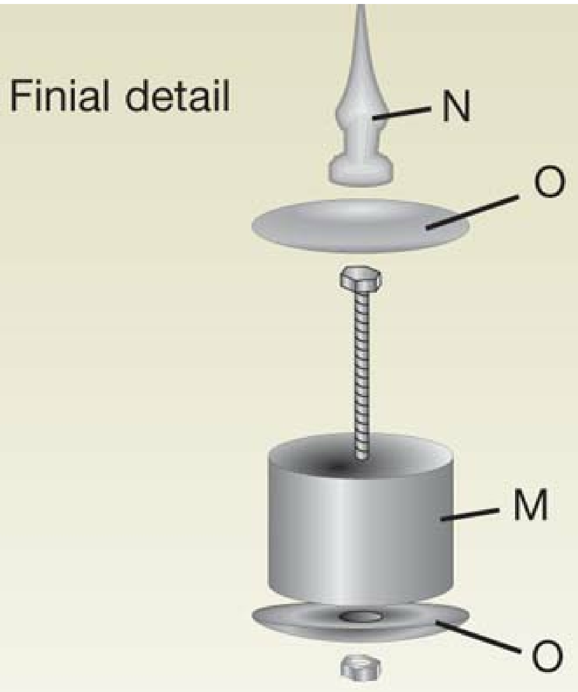
Finial detail
N
O
M
O

How to Build a Gazebo

Before welding, thoroughly clean all parts with denatured alcohol.

Make the Door Panels

1. Cut two legs (A), two door frames (B), a top crossbar (C), and bottom crossbar (D) to length. Miter one end of each leg at 45°, and both ends of the top crossbar at 45°.

2. Lay out the legs, top crossbar, and bottom crossbar to form a rectangle. The bottom crossbar fits between the legs.

3. Check the rectangle for square and tack weld all corners.

4. Mark the door frames at 24". Using a conduit bender, bend the frames at the mark to 50°.

5. Place the door frames inside the assembled panel frame. Align the door frames 8" in from the panel sides. Mark vertical lines on the door frames where they overlap at the peak. Cut on the marks and tack weld the door frame arch together.

6. Align the door frame arch inside the panel frame with the point of the arch touching the top crossbar. Mark the legs where they cross the bottom crossbar. Cut the legs to length. Tack weld the door frame to the top and bottom crossbars.

7. Repeat steps 1 through 6 to make the desired number of door panels. Complete the welds.

Make the Window Panels

1. Cut two window frames (E) and a window crossbar (F) to length.

2. Mark the window frames at 24". Using a conduit bender, bend the frames at the mark to 50°.

3. Lay out the window frames with the window crossbar between two legs (A). Mark the vertical line at the peak where the frames cross each other. Cut the frames on the line and tack weld together.

4. Trim the legs to length, if necessary. Tack weld the window crossbar in place. Put the window frame aside for later installation.

5. Repeat steps 1 through 4 to make the desired number of window panels.

Make the Roof Panels

1. Cut two roof sides (G) to length. Miter one end of each side at 29°. Align the mitered edges and tack weld.

2. Cut a roof base (H) to size. Evenly align it underneath the sides. Mark where the sides cross the base. Cut on this line. Tack weld the base to the sides.

3. Cut a roof crossbar (I) to size. Place the crossbar 5" up from the roof base and mark the angles where the roof sides cross the crossbar.

4. Repeat steps 1 to 3 to make the remaining five roof panels. Complete the welds.

Assemble the Gazebo

Assembling the gazebo before all the decorations are added makes it lighter and easier to maneuver. In addition, if a panel is slightly skewed, it is easier to make corrections without having to rearrange decorative elements. Because the hinges will be perfectly aligned only for one arrangement, make sure you assemble the panels in the order desired.

1. Cut the hinge barrels (J) and pins (K) to length. With a bench vise, bend a right angle ½" in from one end of each pin.

2. Slide a pin into two barrels. Place two panels on the floor or upright against a wall. Place the barrels against the joint between the panels at 20" from the top. Tack weld one barrel to one panel and the other barrel to the second panel.

3. Tack weld another hinge to the panels at about 20" from the bottom.

4. Repeat steps 2 and 3 with the two other sets of two panels.

5. With the panels standing, arrange the three sets of two panels to form a hexagon. The angle between each panel pair should be 120°.

6. Install hinges at the remaining three panel joints.

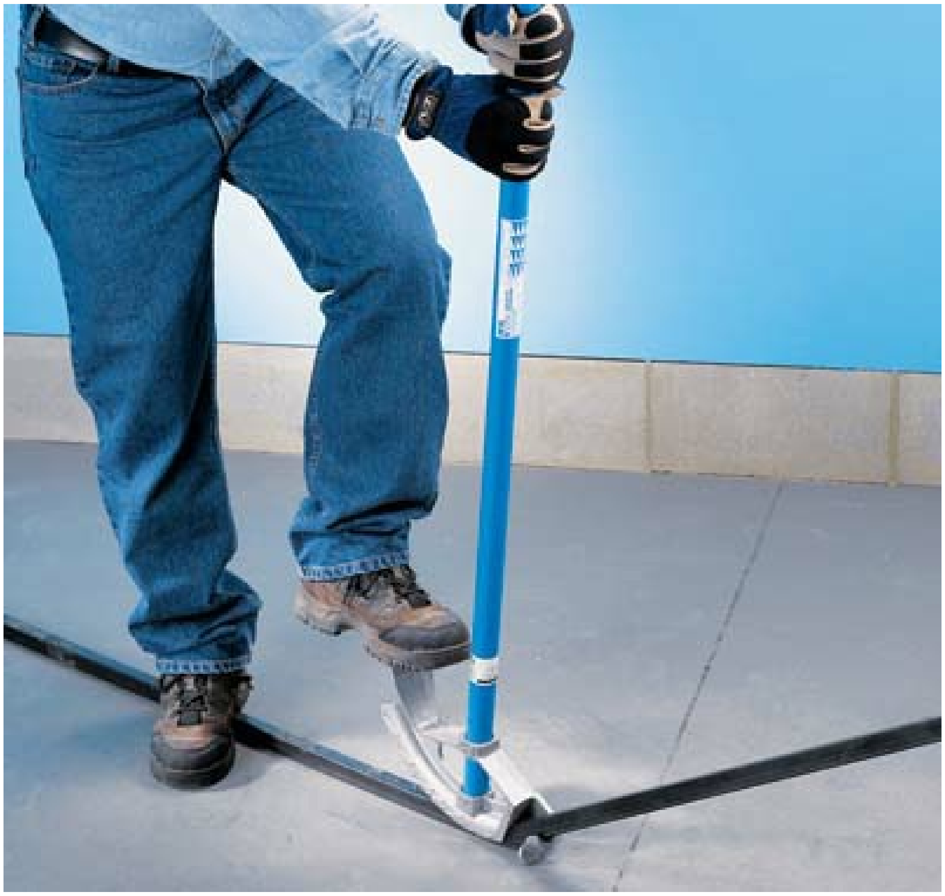

Create the window and door arches by bending the tubing with a large conduit bender.

Assemble the Finial

1. Cut the finial pegs (L) to length. Bend the finial pegs in half to 110°. Support the roof sections on a cinder block or stool and tack weld a finial peg to the point of each roof section.

2. Cut the finial sleeve (M) to length.

3. Weld the finial (N) to the center of one drip tray (O) on the convex side.

4. Grind down the head of a ½ × 6" carriage bolt until it matches the concave curve of the drip tray. Welt the carriage bolt to the middle of the concave side of the drip tray from step 3.

5. Center the finial sleeve over the bolt and weld in place. Drill a $^{9}/_{16}$" hole through the center of the second drip tray.

6. With the peaks of the roof panels resting on a cinder block or low stool, slide the finial assembly over the finial pegs. Bend the finial pegs if necessary to fit inside the finial sleeve.

Assemble the Roof

Assembling the roof is easiest with an assistant or two.

1. Cut the roof pegs (P) and peg sleeves (Q) to length.

2. Weld two sleeves to the tops of each side panel ¾" in from the sides of the legs.

3. Measure the distance between a pair of sleeves. Use this measurement to mark the peg locations on a roof panel. Grind the pegs to 60° on one end and weld the ground end to the roof panel.

4. With the side panels assembled, place a roof section on top of a wall panel, sliding the pegs into the sleeves. Have an assistant hold up the center of the roof, or create awood support.

5. Place the remaining roof sections. Slide the finial over the finial pegs, slide the drip cup over the bolt, and install the washer and nut.

6. Adjust roof and finial pegs as needed.

7. Number or otherwise mark the side and roof panels for reassembly.

8. Remove the finial and take the roof panels down. Disassemble the panels by pounding out the hinge pins.

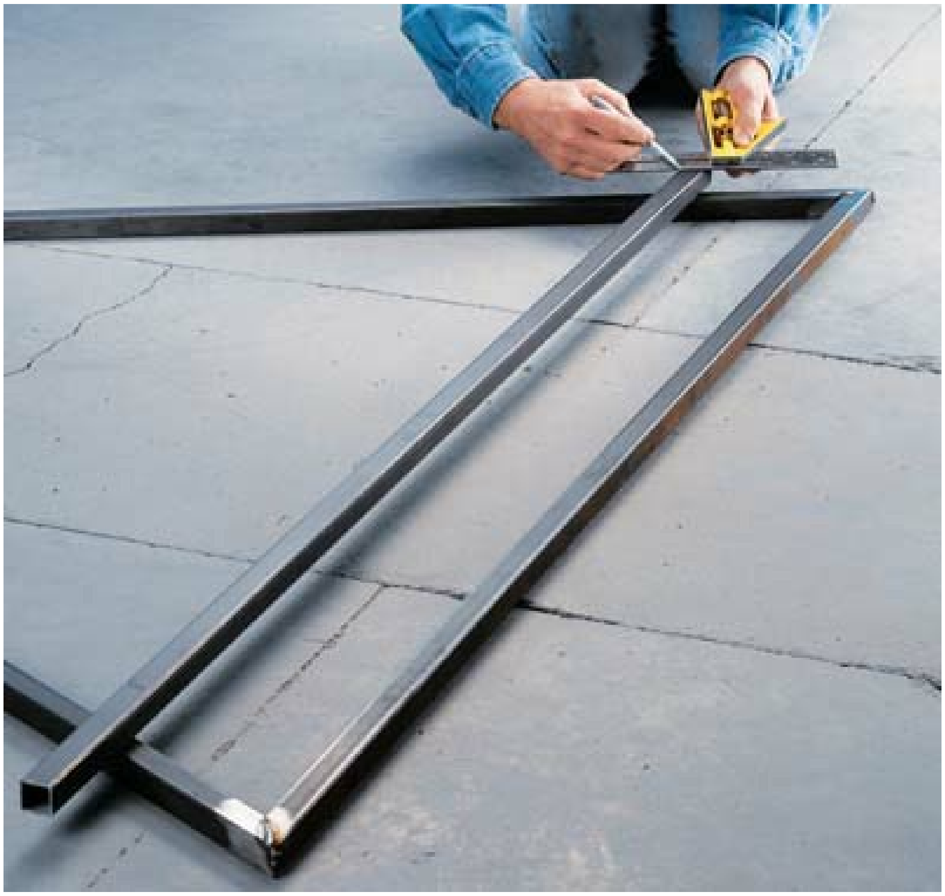

Mark the roof crossbar for cutting.

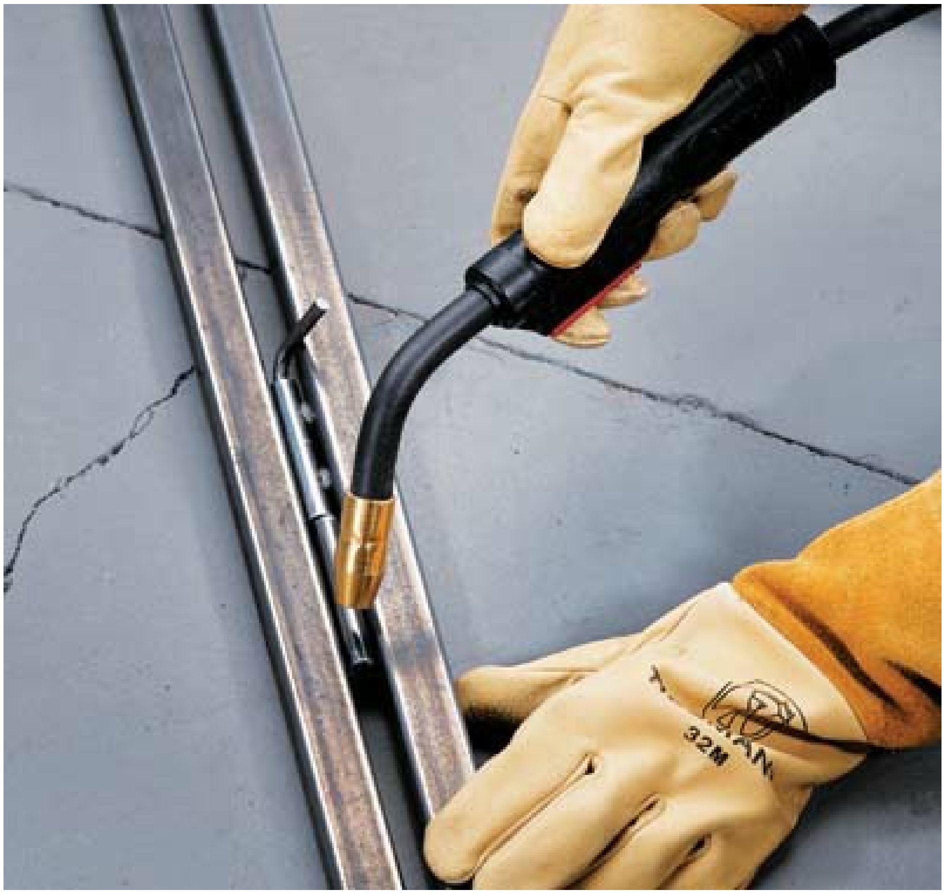

Insert a hinge pin into two hinge barrels. Place the hinge between two side panels and weld one barrel to one panel and the other barrel to the adjoining panel.

Make the Window Scrolls

1. Cut the scroll blanks (R) to length.

2. Cut a $^{3}/_{16}$" slot in a 2" pipe. Place the rod in both slots, then bend the rod around the outside of the pipe one and a half times.

3. Place the other end of the rod in the slot and bend one and a half turns around the outside of the pipe in the opposite direction.

4. Slightly open the bends to create a pleasing scroll.

5. Repeat to form the remaining window scrolls.

6. Weld the scrolls into the window frames as pictured.

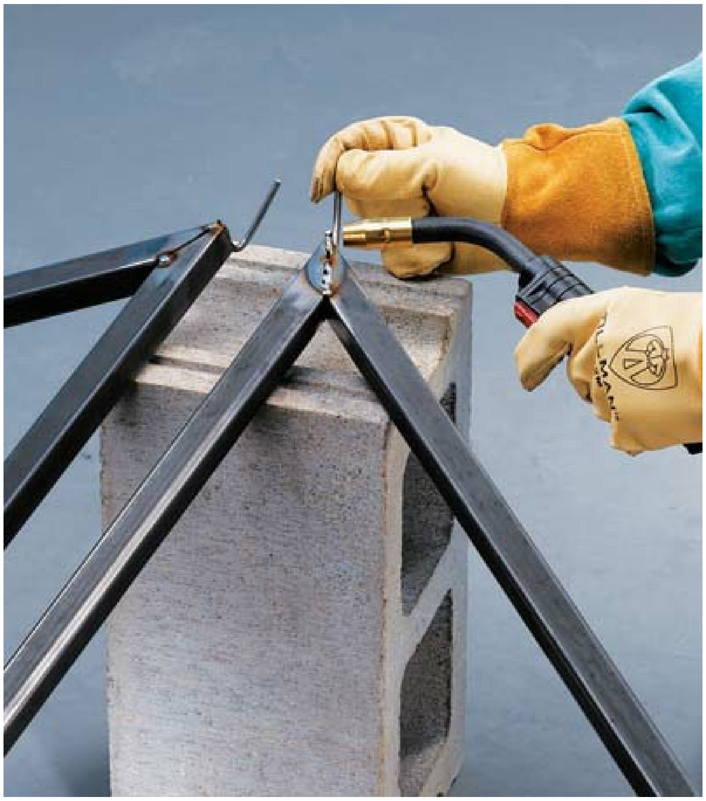

Support roof sections on a cinder block. Weld the finial pegs to the points of the panels.

Make the Roof Scrolls

1. Cut the scroll blanks (S) to length.

2. Cut a $^3/_{16}$" in a 3" pipe. Place the rod through both slots, then bend the rod around the outside of the pipe one and a half times.

3. Place the other end of the rod in the slot and bend one and a half turns around the outside of the pipe in the opposite direction.

4. Slightly open or close the bends to make the scroll fit between the roof base and roof crossbar.

5. Repeat for the other roof scrolls.

6. Weld the scrolls between the roof base and roof crossbar as pictured.

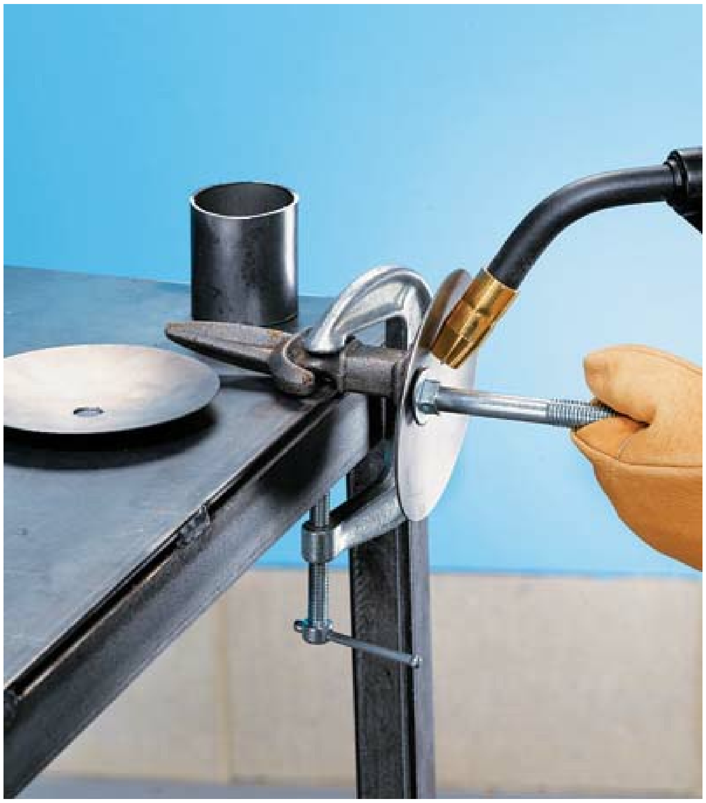

Create the finial assembly by welding the finial to the drip tray, then welding the bolt to the drip tray.

Add Decoration to the Roof

You can make the roof a simple diamond pattern.

1. To make the diagonals (T), mark the roof sides every 6" down from the peak and 6" down from the roof crossbar. Lay out two rods from the peak marks to the opposite first two marks on the roof crossbar to match the roof panel diagram (page 201). Cut the rods to fit and weld in place.

2. Mark the diagonals from step 1 every 6".

3. Make the internal diagonals by cutting pieces to fit between the marks on the right side and base. Weld in place.

4. Finish the internal diagonals by cutting pieces to fit between the marks on the left side and base. Weld in place.

Add Diagonal Filler to Side Panels

Diagonal rods can fill the space below and above the windows, or you can add scrolls and hearts.

1. On the door panels, mark the legs, door frames, and top crossbars every 8".

2. Cut diagonals to fit between the marks.

3. Weld the descending left to right diagonals in place, then weld the ascending left to right diagonals in place.

4. On the window panels, mark the legs, top and bottom crossbars, window frames, and window crossbars every 8". Align the window frame between the legs, 8" in from each side with the 8" marks aligned. The bottom of the window should be 32" from the bottom crossbar.

5. Cut diagonals to fit between the marks on the window frame and legs.

6. Cut diagonals to fit between the marks on the crossbars and the window frames and window crossbars.

7. Weld the descending left to right diagonals in place, then weld the ascending left to right diagonals in place.

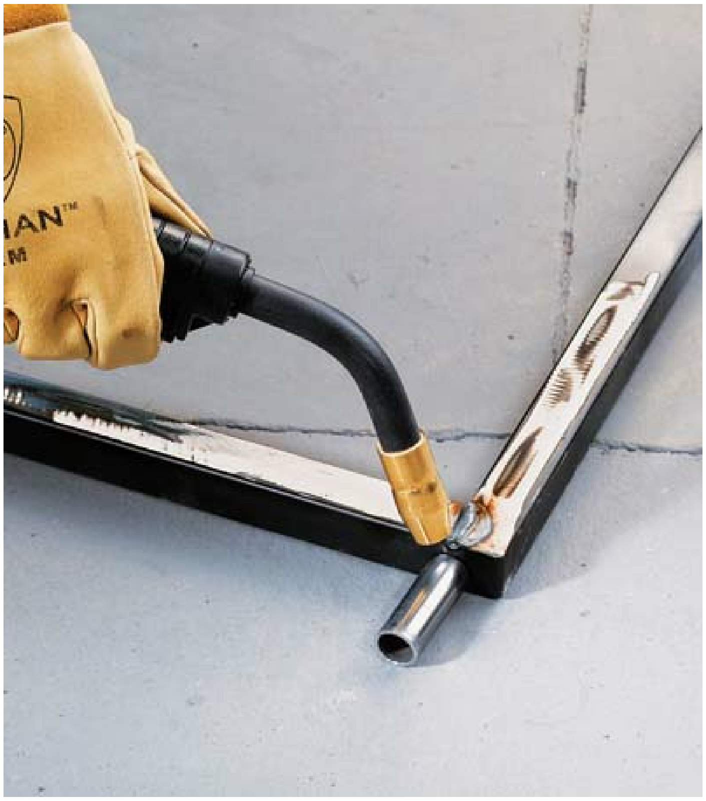

Weld the roof sleeves to the tops of the side panels.

Finish the Gazebo

1. Wire brush or sand blast the gazebo. Finish as desired. *Note: Take care to track the panel numbers so you can reassemble the gazebo.*

2. Reassemble the gazebo.

3. For at least three corners of the gazebo, pound 3-ft. lengths of ¾" round bar or 18" rebar into the ground and slide the leg ends over the bars.

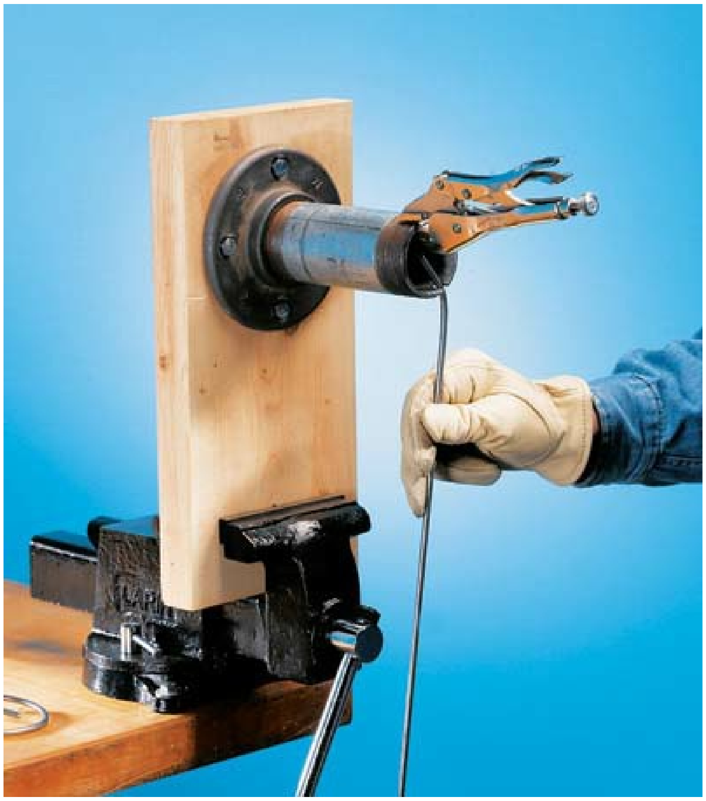

Cut a $^{3}/_{16}$" notch in a 2" pipe to create the bending jig for the scrolls. Slide the rod into the notch, then bend the rod around the pipe.

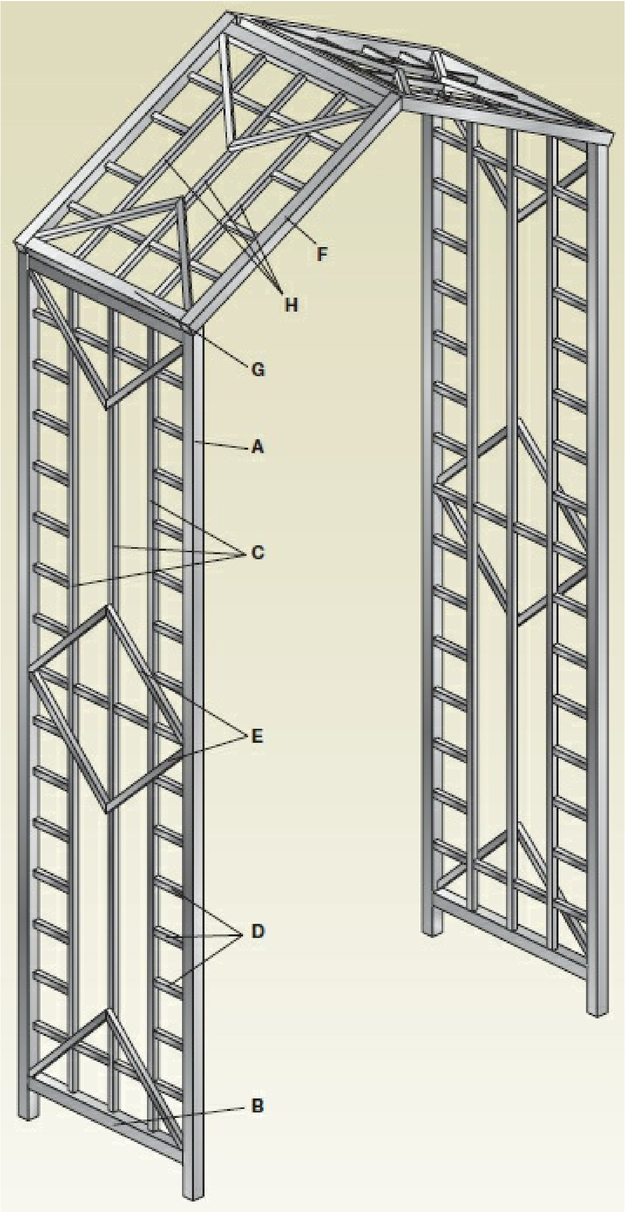
F
H
G
A
C
E
D
B

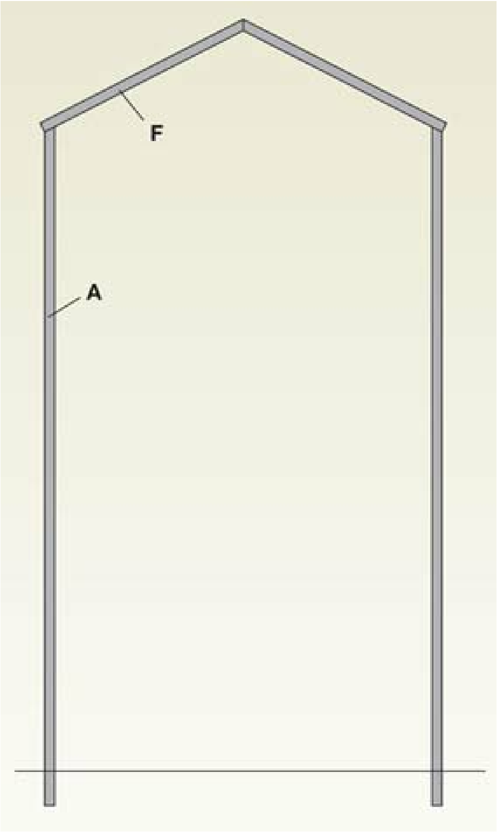

Overall dimensions

44 × 16½ × 84"

Arbor

This distinctive arbor will add a touch of class to any garden. Modeled after the prairie style, its clean lines are straightforward to cut and weld, and they provide plenty of climbing support for a variety of vines. The ½" inserts are placed flush with the back of the 1" sides and ends, eliminating the need for difficult centering and giving an increased sense of depth. The arbor can be made from the 6-foot steel lengths available at home improvement centers, but because of the amount of steel used, a trip to a steel supplier might be worthwhile, especially if you have a way to transport 20-foot lengths. The ½" square tube bows quite easily, so if you do buy 20 foot lengths, make sure they are well supported in transit. Otherwise, they will bow and not be suitable for the vertical inserts.

Materials

- 16 gauge 1 × 1" square tube (44 feet)
- 16 gauge ½ × ½" square tube (100 feet)
- Wood scraps

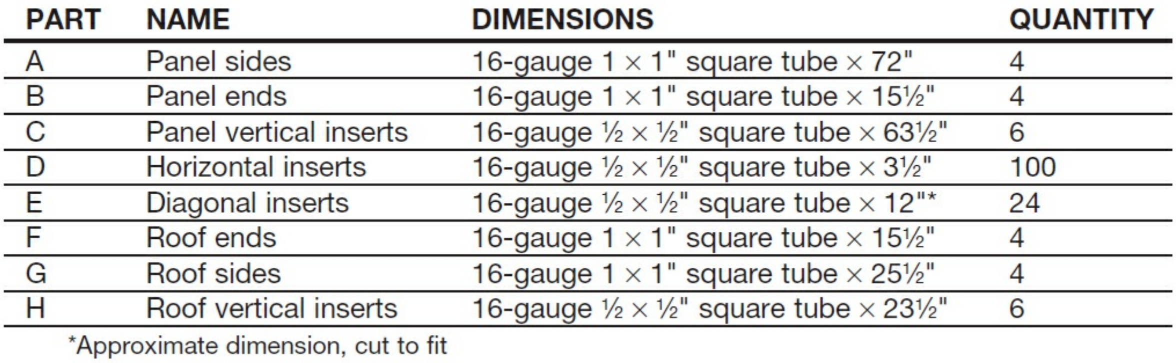

PART	NAME	DIMENSIONS	QUANTITY
A	Panel sides	16-gauge 1 × 1" square tube × 72"	4
B	Panel ends	16-gauge 1 × 1" square tube × 15½"	4
C	Panel vertical inserts	16-gauge ½ × ½" square tube × 63½"	6
D	Horizontal inserts	16-gauge ½ × ½" square tube × 3½"	100
E	Diagonal inserts	16-gauge ½ × ½" square tube × 12"*	24
F	Roof ends	16-gauge 1 × 1" square tube × 15½"	4
G	Roof sides	16-gauge 1 × 1" square tube × 25½"	4
H	Roof vertical inserts	16-gauge ½ × ½" square tube × 23½"	6

*Approximate dimension, cut to fit

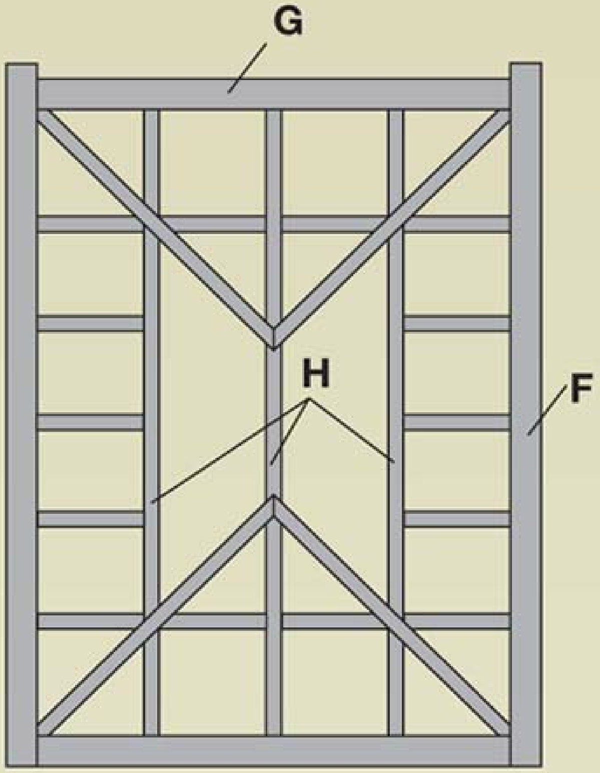

Roof Detail

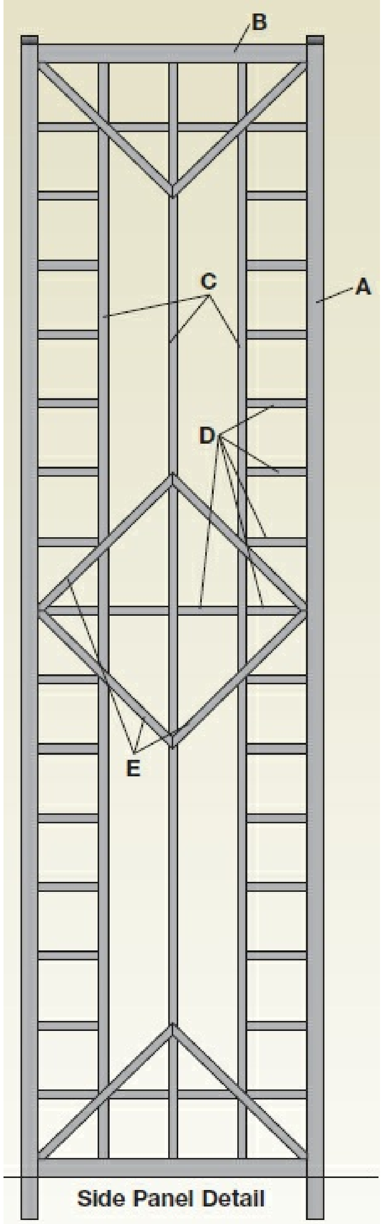

Side Panel Detail

Check the panel for square by measuring the diagonals. If the measurements are equal, the assembly is square.

How to Build an Arbor

Assemble the Panels

Set up the project on a sheet of ¾" plywood placed on sawhorses. Make sure the plywood is not bowed or it will cause misalignment of the project pieces. Working on a raised surface is easier than working on the floor, and you can clamp the metal to the plywood. Remember to clamp your work cable to the workpiece. If you want the arbor wider or narrower than 44", you must adjust the miter angles for the roof and side panels.

1. Cut the panel sides (A) to size, mitering one end at 30°. The mitered end is the top. Cut the panel ends (B), vertical inserts (C), and horizontal inserts (D) to size.

2. Place a panel end between the top of two panel sides, keeping the outside edge of the panel end flush with the short ends of the mitered panel sides. Clamp the pieces in place.

3. Position three vertical inserts between the two panel sides. (This is just to hold the bottom panel end in place. The exact location of the inserts is not important at this time.)

4. Place the bottom panel end between the panel sides, keeping it snug against the vertical inserts. Clamp the bottom panel end in place.

5. Remove the vertical inserts. Use a carpenter’s square to check the panel sides and ends for square. Tack weld each corner. Check for square again by measuring the two diagonals. If the measurements are equal, the panel is square. If not, adjust the workpieces until the diagonals are the same.

6. Turn the panel over and complete the welds at the four corners. Grind the tack welds flat so the panel lays flat on the work surface.

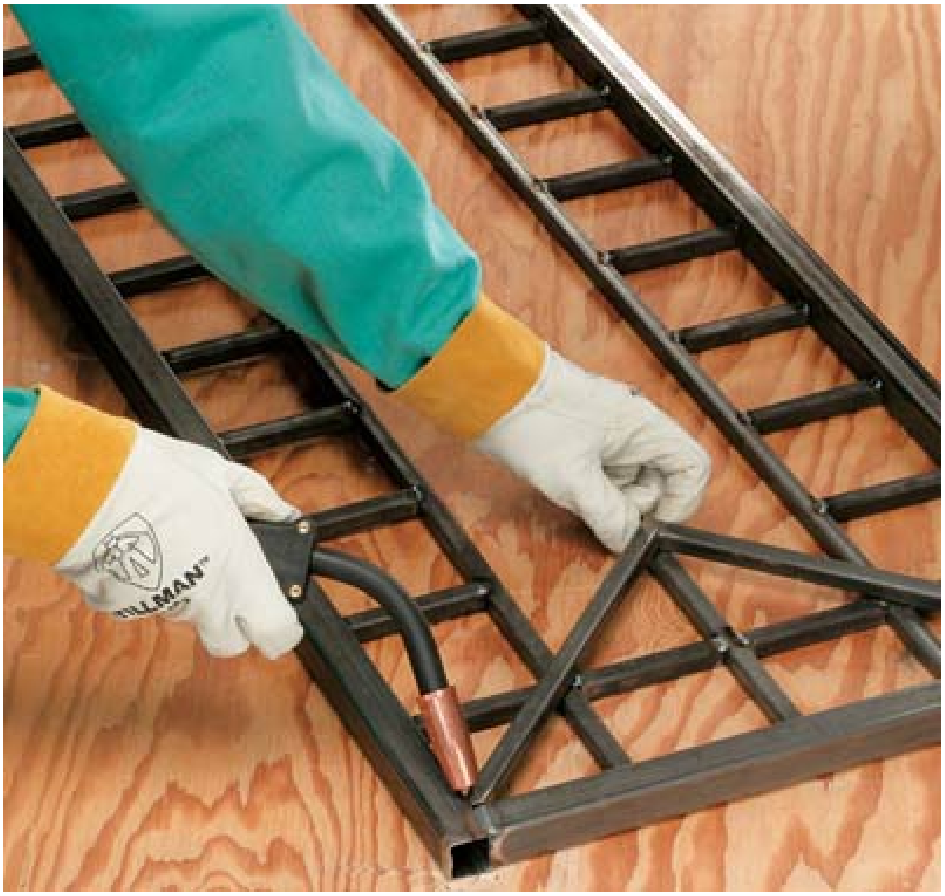

Weld the diagonal inserts to the panel framework and to the center vertical insert.

Attach the Vertical & Horizontal Inserts

1. Replace the vertical inserts between the panel sides. Place several horizontal inserts between the vertical inserts and the panel sides to ensure the correct spacing of the vertical inserts.

2. Weld the vertical inserts to the panel top and bottom.

3. Place the horizontal inserts into the vertical framework every 3½". (Use horizontal inserts as spacers.)

4. Starting at the top of the panel, use a combination square to align the first row of horizontal inserts. Once they are aligned, weld them in place.

5. Continue aligning and welding the inserts one row at a time. If the vertical inserts are bowed side to side, use a clamp to hold them against the inserts. If they bow upward, weigh them down with a sandbag or other weight.

Insert the Diagonal Inserts

1. Cut four diagonal inserts (E) to size, mitering one end at 45°.

2. Place two diagonal inserts against the bottom edge of the top panel. The mitered ends of the inserts should butt together over the middle vertical insert. Mark the other end of the diagonal inserts where they cross the top panel so they will fit in the corner.

3. Make the angled cuts on the diagonal inserts.

4. Grind down the welds that will be underneath the diagonal inserts so they will lie flat. Weld the diagonal inserts in place.

5. Repeat steps 2 to 4 to weld the diagonal inserts to the bottom of the panel.

6. Cut four more diagonal inserts, mitering both ends at 45°.

7. Center these four inserts over the eighth row of horizontal inserts to form a diamond shape. Weld the inserts in place.

8. Repeat this entire process to construct a second panel.

Make clamping blocks out of wood scraps. Clamp the roof panels to the blocks. Place two or three tack welds along the joint between the panel ends.

Build the Roof Panels

1. Construct two roof panels following the same procedure used for the side panels, except with no central diamond insert.

2. Stand the roof panels on edge and join the mitered ends to form the 120° angle for the roof peak.

3. Clamp the roof panels in place. (You may need to attach temporary wooden clamping points to the work surface by fastening 1 × 2 or 2 × 4 scraps to match the layout. Clamp the panels to the scraps.) Place two or three tack welds along the joint between the panel ends.

Fasten the Roof to the Sides

1. Place a side panel on edge, and set it against one side of the roof panel. Line them up so the mitered edge of the side panel is making contact with the bottom edge of the roof panel. Clamp the panels in place.

2. Tack weld the panels at two points along the joint.

3. Repeat steps 1 and 2 for the other side panel.

4. Measure the distance between the tops and the bottoms of the side panels to ensure the panels are an equal distance apart at both ends.

5. Check the panels for square by measuring the diagonals. Make any necessary adjustments, and make sure the roof assembly is still fitted against the side panels.

6. Weld all the mitered roof peak joints and the mitered joints between the side panels and the roof panels.

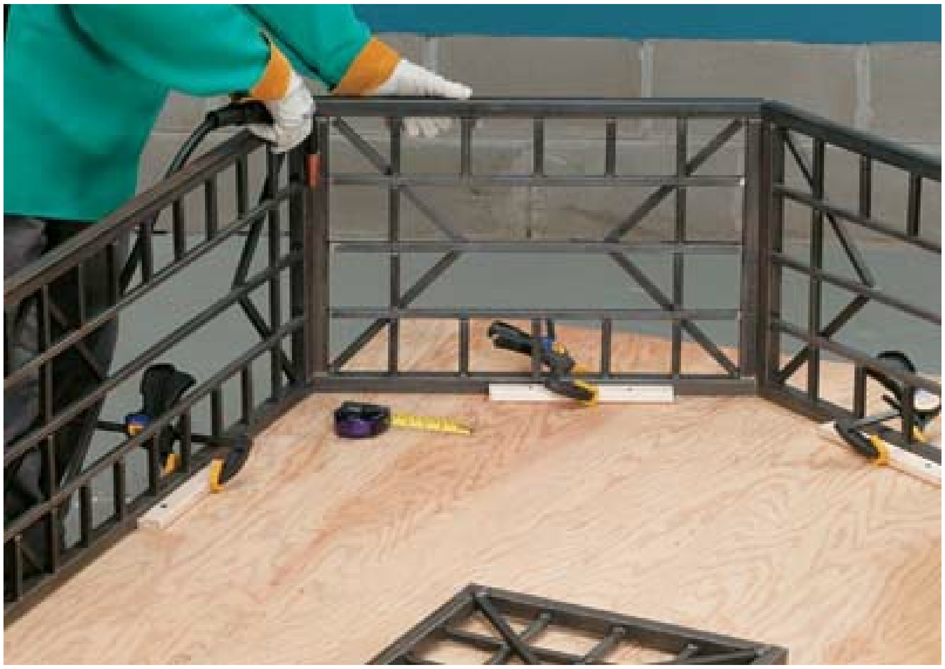

Place tack welds along the joint between the roof and side panels.

Finishing Touches

To prevent putting stress on the welds, and to maintain the arbor’s shape, clamp wooden crosspieces between the panel sides before moving it.

1. The easiest finish for the arbor is to allow it to gently rust over the years. You may want to place plastic end caps over the exposed tube ends on the roof, or you can weld on small caps.

2. The arbor can sit on the ground or can be mounted in concrete footings. If you live in a windy location, you may want to drive four lengths of rebar into the ground and slip the arbor legs over them.

Appendix: Additional Plans

Scroll Desk Lamp

Materials

- 1/8 × ¾" flat bar (15")
- ½" round tube (9")
- 1/8 × ½" flat bar (36")
- ½", 1", and 2" pipe
- Lamp hardware and shade
- 3/8" hollow threaded rod (1")

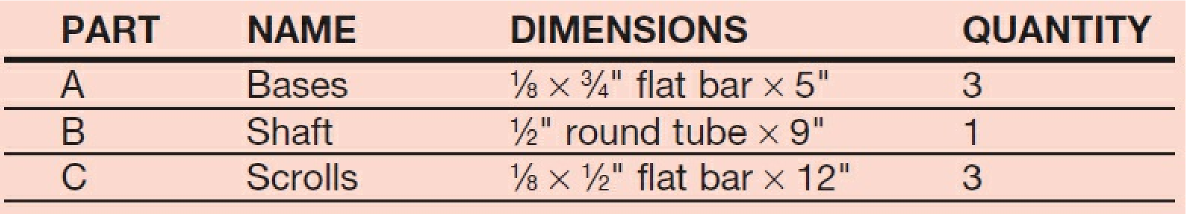

PART	NAME	DIMENSIONS	QUANTITY
A	Bases	⅛ × ¾" flat bar × 5"	3
B	Shaft	½" round tube × 9"	1
C	Scrolls	⅛ × ½" flat bar × 12"	3

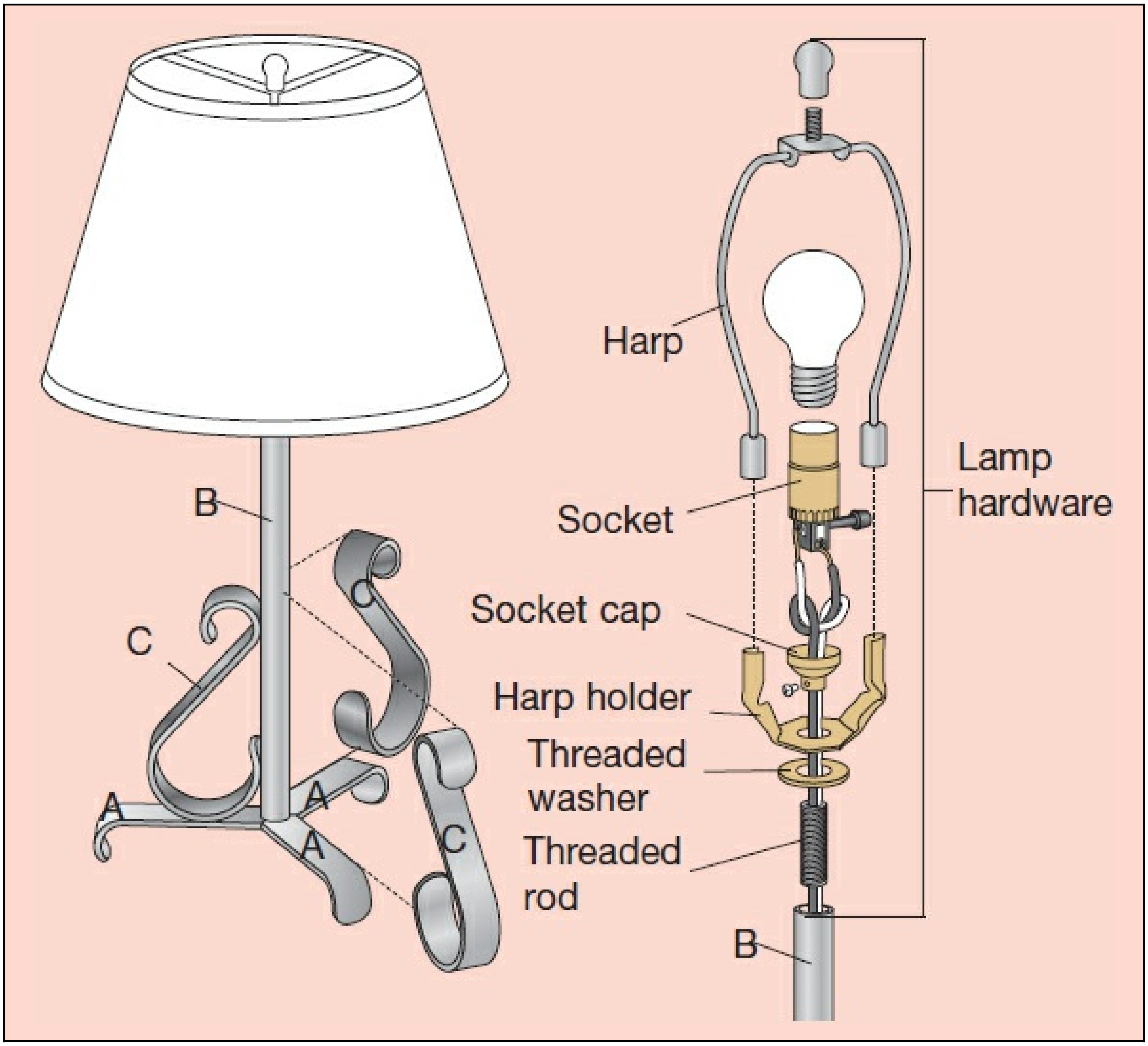
B
C
C
C
A
A
A
Harp
Socket
Socket cap
Harp holder
Threaded washer
Threaded rod
B
Lamp hardware

Tree-shaped Candleholder

Materials

- $^{3}/_{16}$" sheet (8 × 10")
- $^{3}/_{16}$" × 2" flat bar (5 feet)
- Votive holders and candles
- Silicone adhesive

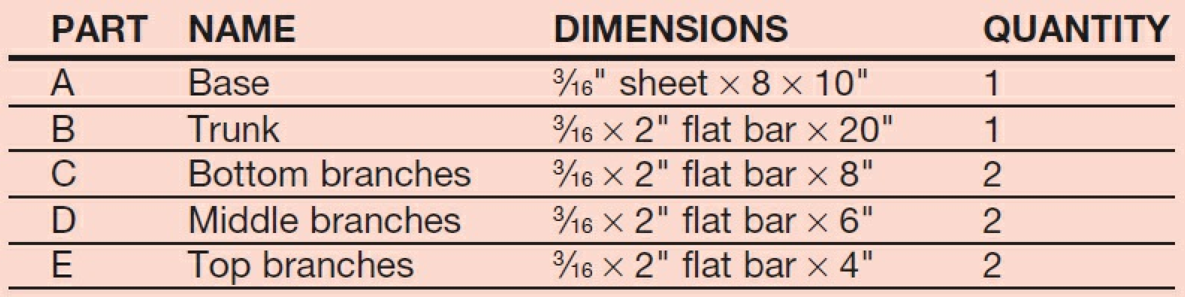

PART	NAME	DIMENSIONS	QUANTITY
A	Base	$^{3}/_{16}$" sheet × 8 × 10"	1
B	Trunk	$^{3}/_{16}$ × 2" flat bar × 20"	1
C	Bottom branches	$^{3}/_{16}$ × 2" flat bar × 8"	2
D	Middle branches	$^{3}/_{16}$ × 2" flat bar × 6"	2
E	Top branches	$^{3}/_{16}$ × 2" flat bar × 4"	2

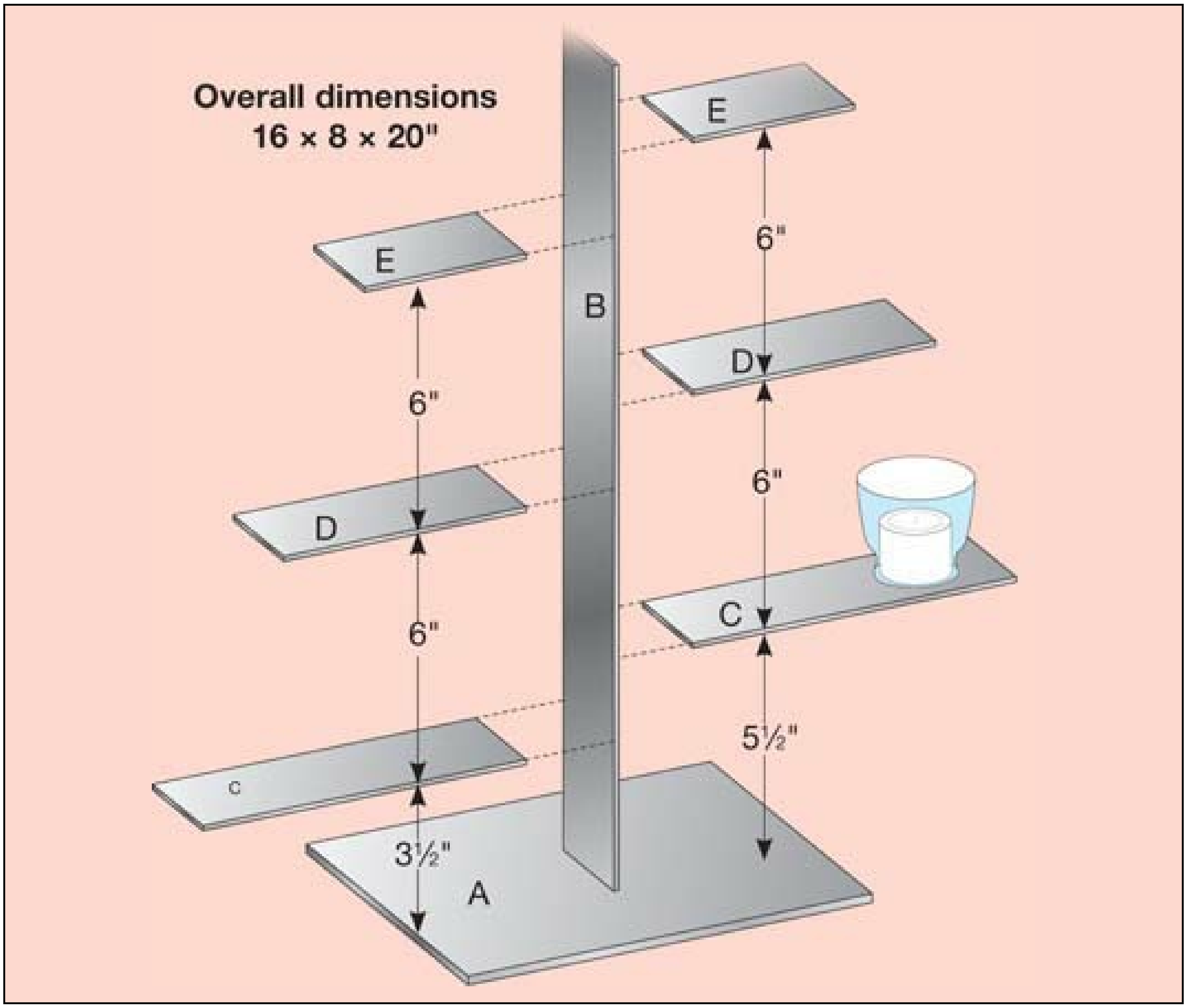
Overall dimensions
16 × 8 × 20"
E
E
B
6"
D
6"
6"
D
C
6"
C
5½"
3½"
A

Fireplace Candleholder

Materials

- 4 × 8" scrolls (2)
- ½" twisted square rod (5")
- ½" square rod (7 feet)
- 3⅜" bobeches (8)
- Candles (8)

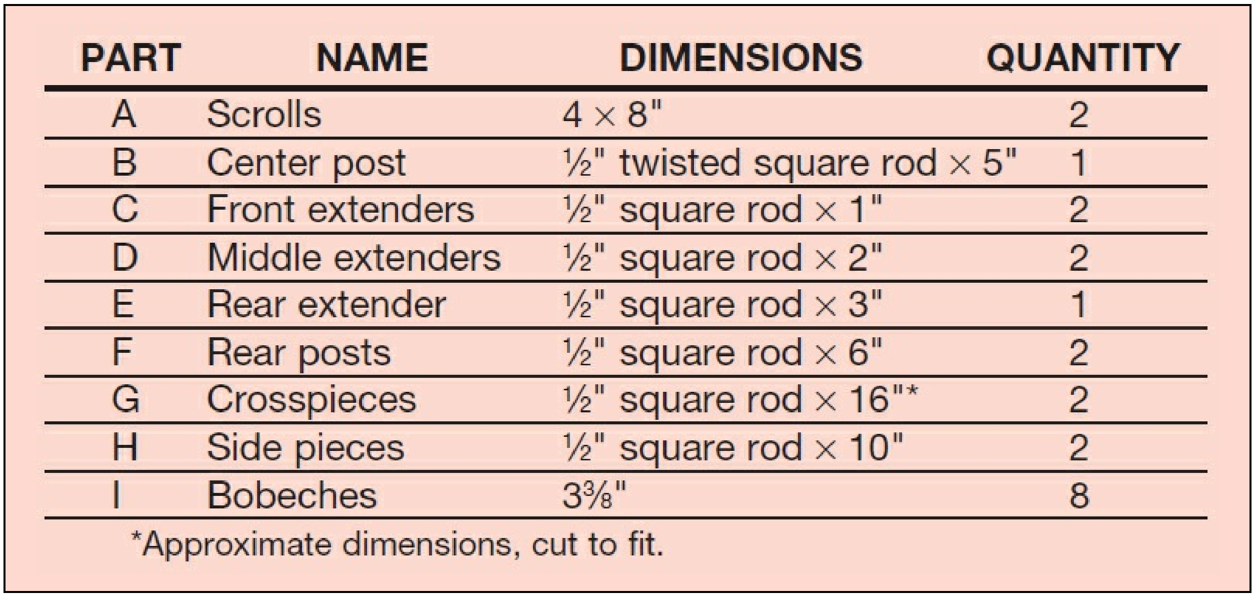

PART	NAME	DIMENSIONS	QUANTITY
A	Scrolls	4 × 8"	2
B	Center post	½" twisted square rod × 5"	1
C	Front extenders	½" square rod × 1"	2
D	Middle extenders	½" square rod × 2"	2
E	Rear extender	½" square rod × 3"	1
F	Rear posts	½" square rod × 6"	2
G	Crosspieces	½" square rod × 16"*	2
H	Side pieces	½" square rod × 10"	2
I	Bobeches	3⅜"	8

*Approximate dimensions, cut to fit.

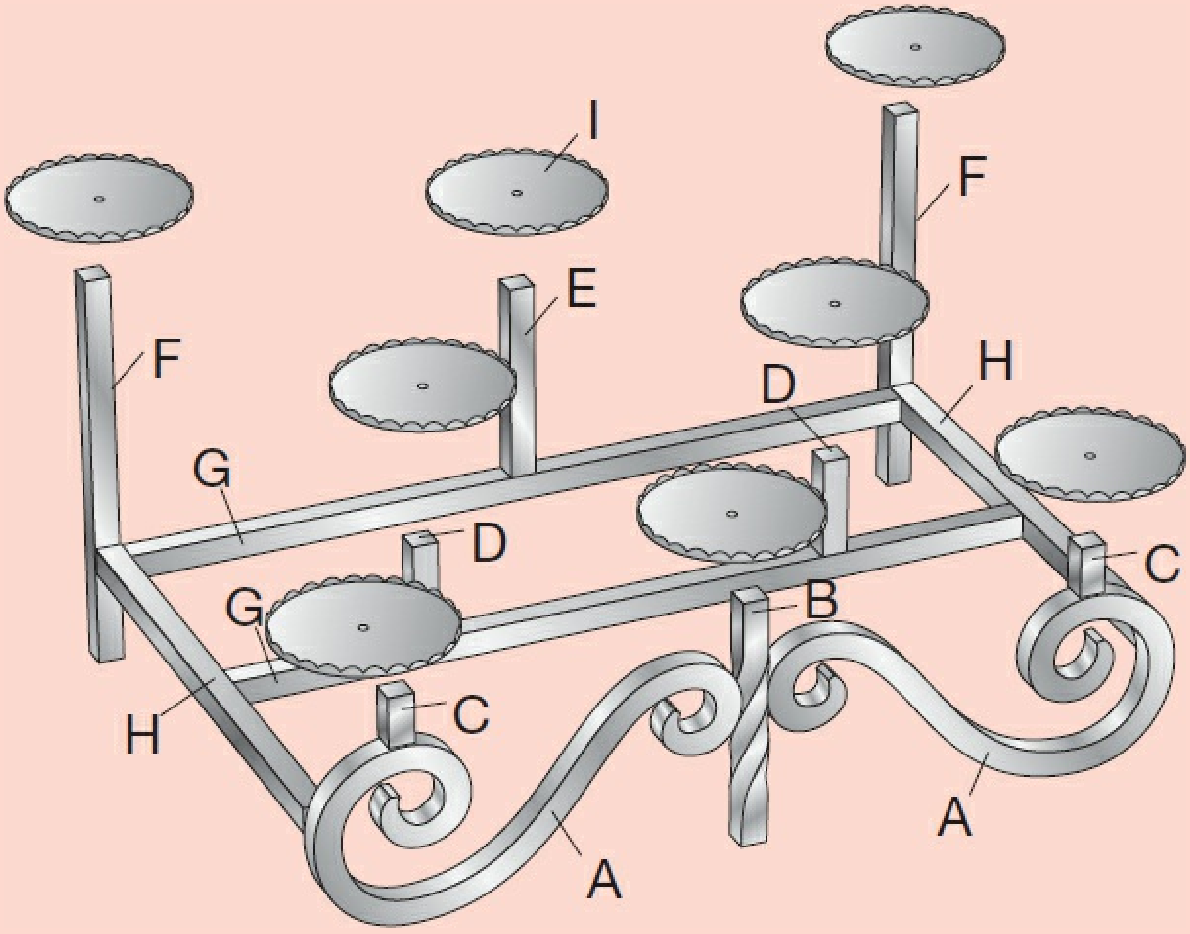
I
F
E
F
H
D
G
D
C
G
B
C
H
A
A

Overall dimensions

16 × 11 × 6"

Wall-mounted Candleholder

Materials

- ½" square bar (14 feet)
- 2¼" bobeche (9)
- Candles (9)

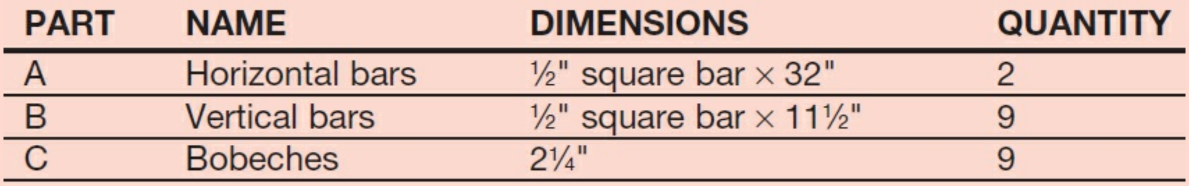

PART	NAME	DIMENSIONS	QUANTITY
A	Horizontal bars	½" square bar × 32"	2
B	Vertical bars	½" square bar × 11½"	9
C	Bobeches	2¼"	9

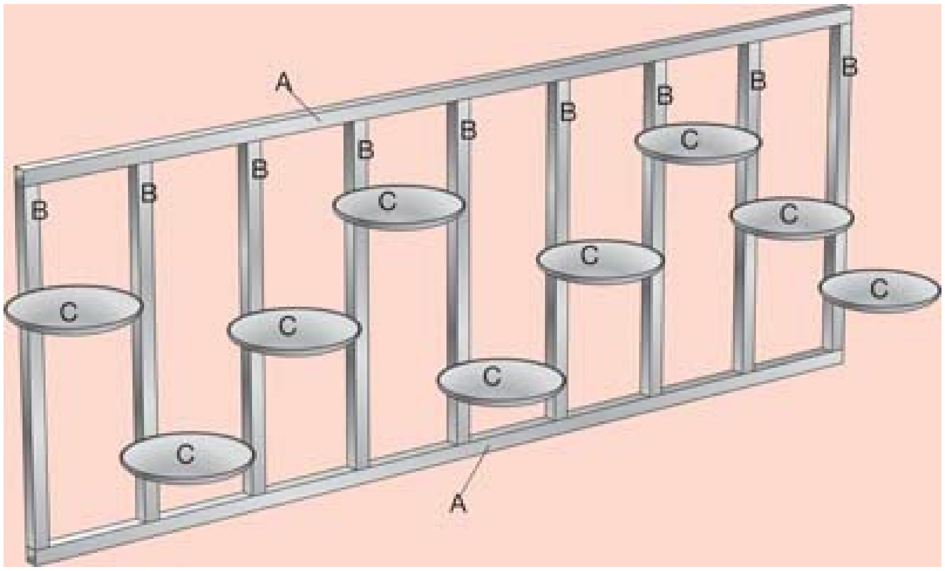
A
B
B
B
B
B
B
B
B
B
B
C
C
C
C
C
C
C
C
C
A

Overall dimensions

32 × 3 × 12½"

Chandelier

Materials

- 1/8 × ¾" flat bar (17 feet)
- 10", 6", 2", and 1" bending forms
- 2¼" round tube (4")
- 2½" bobeches (5)
- 2" threaded nipple
- Lock washers (2)
- Threaded brass washers (3)
- Finial
- Decorative chain
- Candles (5)

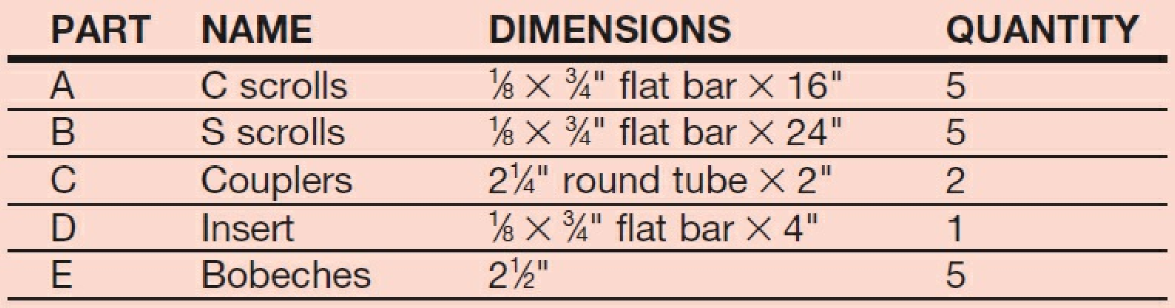

PART	NAME	DIMENSIONS	QUANTITY
A	C scrolls	⅛ × ¾" flat bar × 16"	5
B	S scrolls	⅛ × ¾" flat bar × 24"	5
C	Couplers	2¼" round tube × 2"	2
D	Insert	⅛ × ¾" flat bar × 4"	1
E	Bobeches	2½"	5

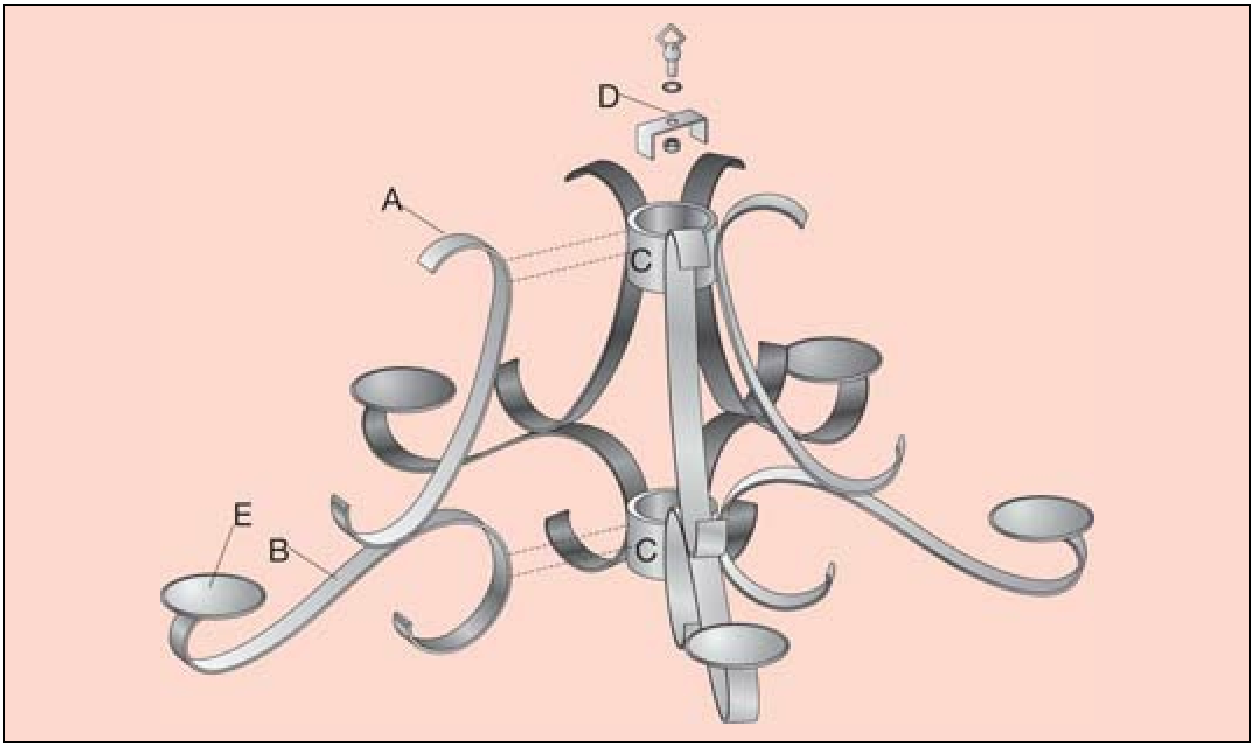
D
A
C
C
E
B

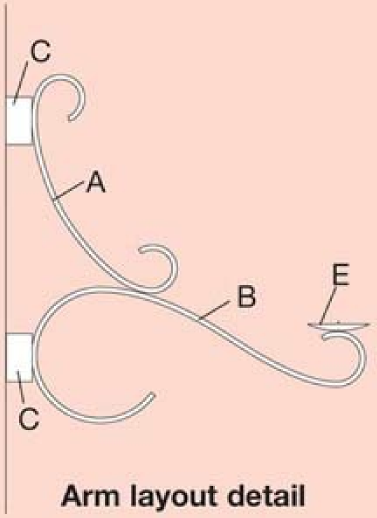

Arm layout detail

Corner Coat Rack

Materials

- 1/8 × 1¼" flat bar (7 feet)
- 1/8 × ¾" flat bar (7 feet)
- Expanded sheet metal (2 × 2 feet)
- 2" pipe
- Hanging hardware

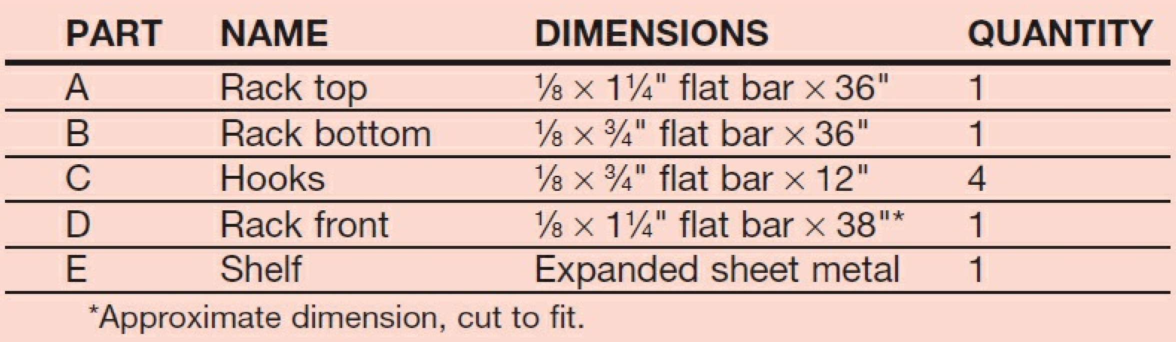

PART	NAME	DIMENSIONS	QUANTITY
A	Rack top	⅛ × 1¼" flat bar × 36"	1
B	Rack bottom	⅛ × ¾" flat bar × 36"	1
C	Hooks	⅛ × ¾" flat bar × 12"	4
D	Rack front	⅛ × 1¼" flat bar × 38"*	1
E	Shelf	Expanded sheet metal	1

*Approximate dimension, cut to fit.

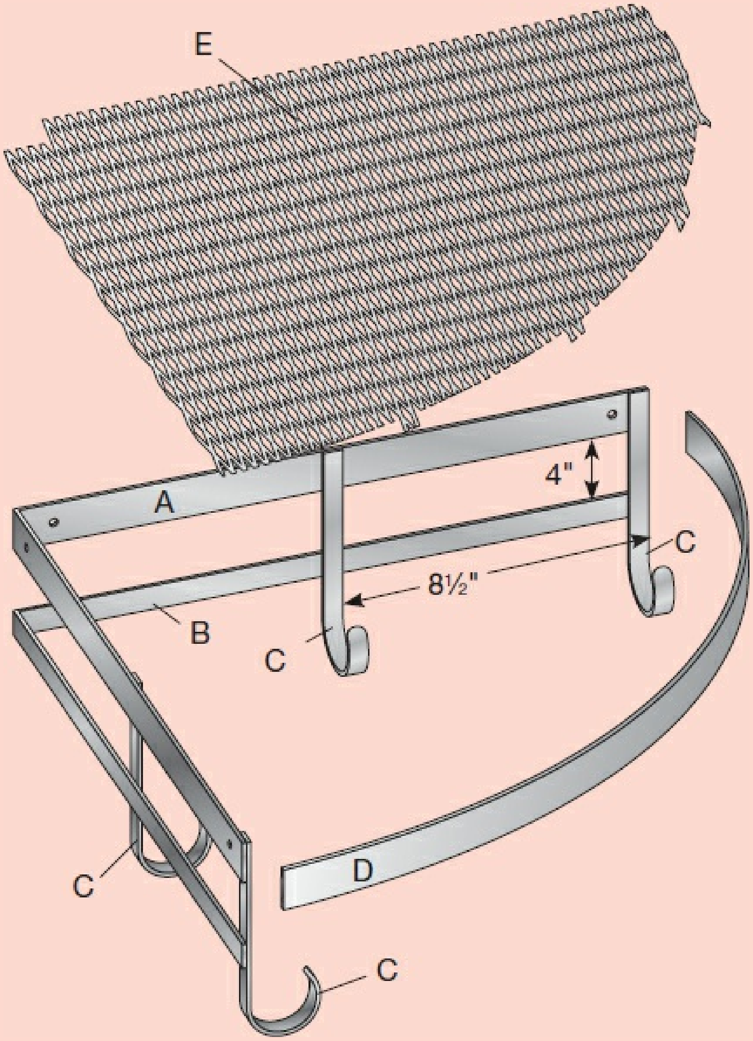
E
A
4"
B
C
8½"
C
C
D
C
C

Overall dimensions

18 × 18 × 8"

Baker’s Shelves

Materials

- ¼" round bar (54 feet)
- 16 gauge × ¾" round tube (13½ feet)
- 16 gauge × ½" round tube (29 feet)
- 4 × 6 foot plywood
- 2 × 4 foot plywood
- 1 × 2 lumber × 36" (7)
- Drywall screws
- 2 × 2 lumber × 36"
- Pre-drilled ¾" on center flat bar (3 feet)
- $^{5}/_{16}$ × 2½" machine bolts (24)
- ¾" rubber pipe tips or glides (4)
- Brass or wood finials (2)
- Tempered glass shelves cut to fit (4)

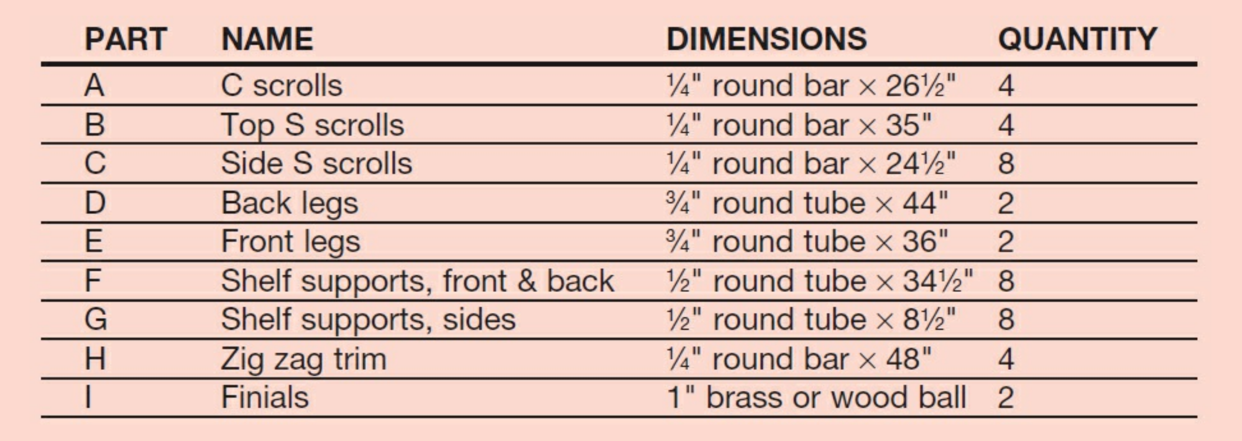

PART	NAME	DIMENSIONS	QUANTITY
A	C scrolls	¼" round bar × 26½"	4
B	Top S scrolls	¼" round bar × 35"	4
C	Side S scrolls	¼" round bar × 24½"	8
D	Back legs	¾" round tube × 44"	2
E	Front legs	¾" round tube × 36"	2
F	Shelf supports, front & back	½" round tube × 34½"	8
G	Shelf supports, sides	½" round tube × 8½"	8
H	Zig zag trim	¼" round bar × 48"	4
I	Finials	1" brass or wood ball	2

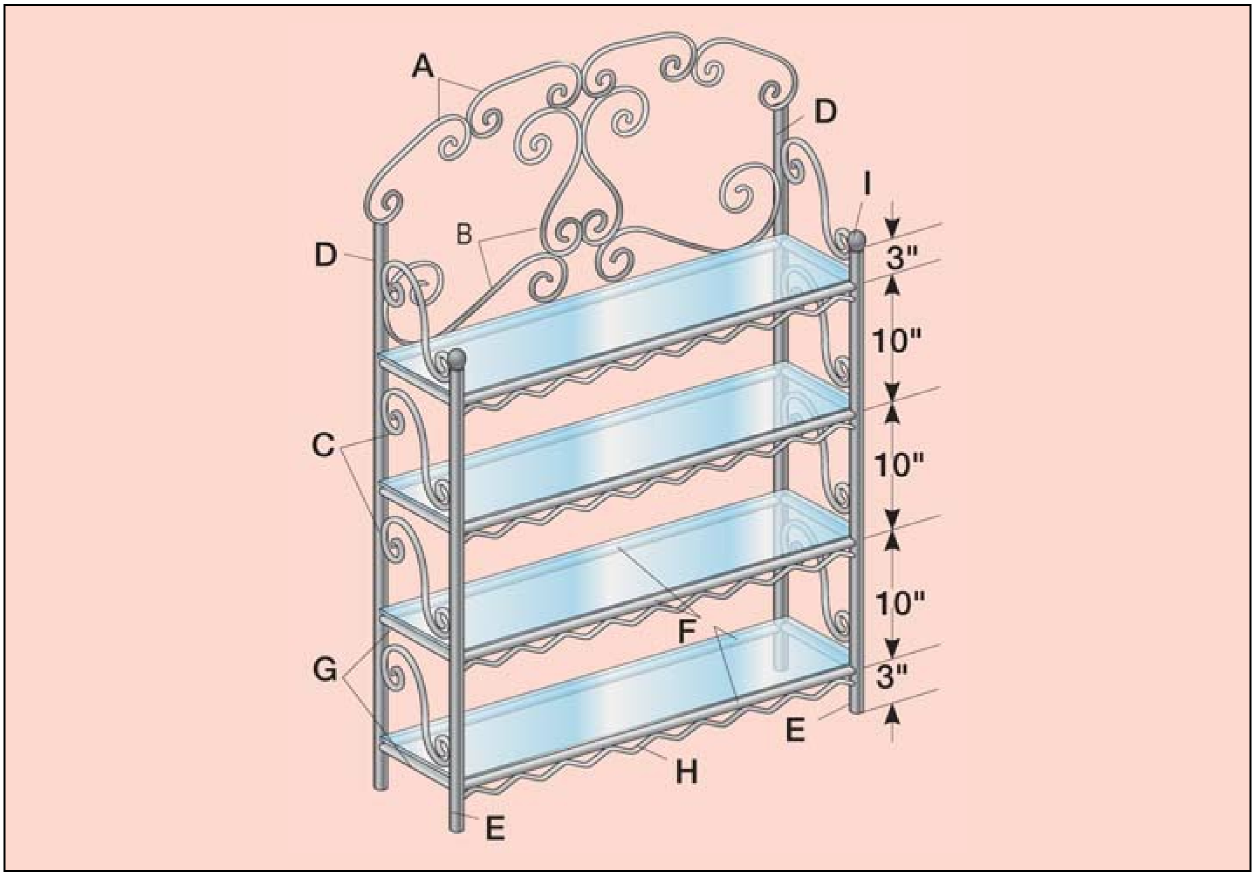
A
D
I
B
D
3"
10"
C
10"
10"
F
G
3"
E
H
E

Tapered-Leg Table

Materials

- 1/8" sheet metal (6 × 30" rectangle, 12" square)
- 1/8 × 2" flat bar (2½ feet)
- ½" black pipe (6")

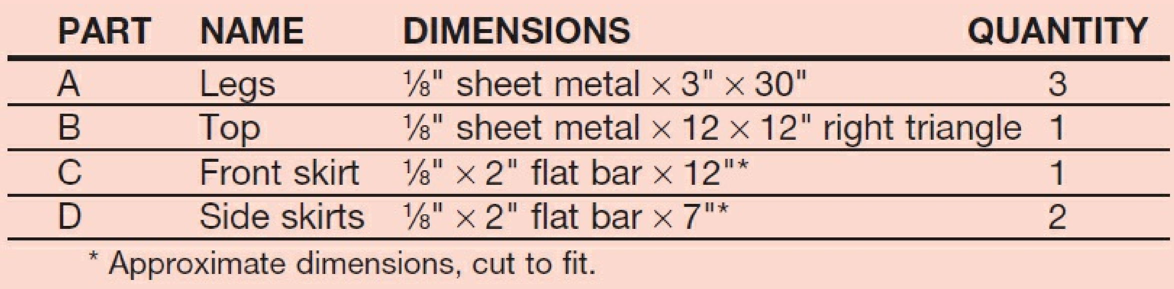

PART	NAME	DIMENSIONS	QUANTITY
A	Legs	⅛" sheet metal × 3" × 30"	3
B	Top	⅛" sheet metal × 12 × 12" right triangle	1
C	Front skirt	⅛" × 2" flat bar × 12"*	1
D	Side skirts	⅛" × 2" flat bar × 7"*	2

* Approximate dimensions, cut to fit.

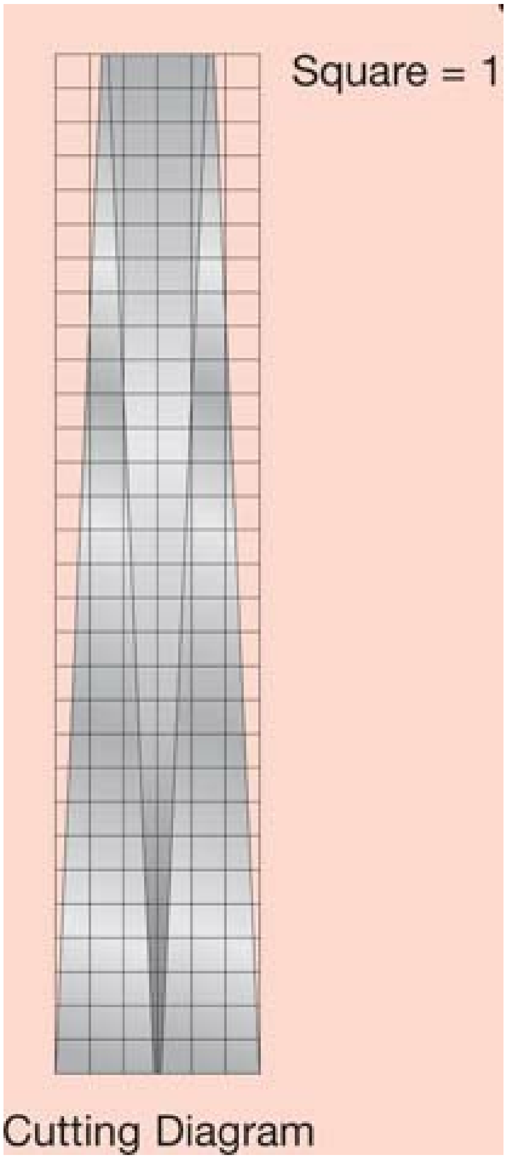
Square = 1
Cutting Diagram

Cutting Diagram

Square = 1"

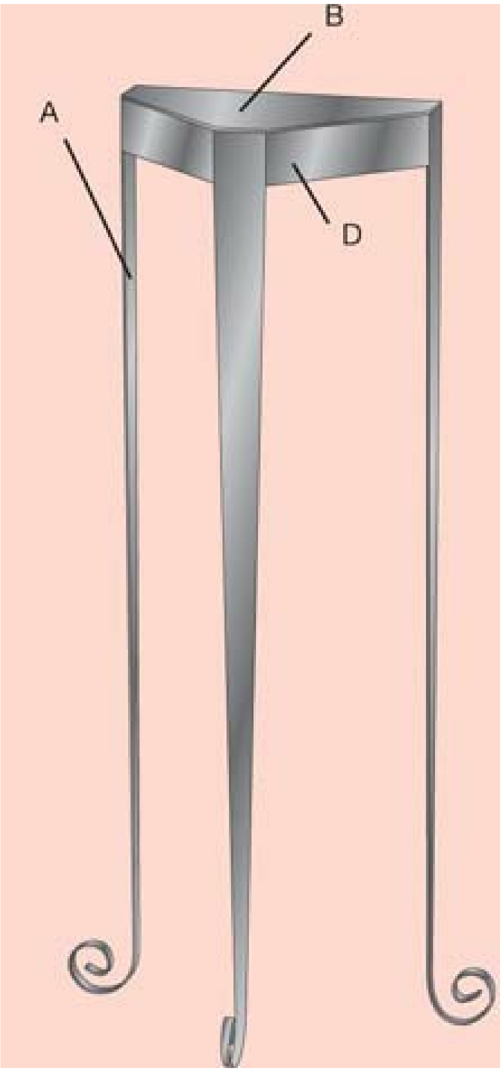
B
A
D

Coat Rack

Materials

- 16-gauge ½ × ½" square tube (58 feet)
- 16-gauge 1 × 1" square tube (50 feet)
- 4" O.D. square tube rings (20, Triple S Steel #SR 4)
- 1/8 × 1" round tube (39")
- 1/8 × ½" flat bar (39 feet)
- 1½" spheres (8, Triple S Steel SF116F4)

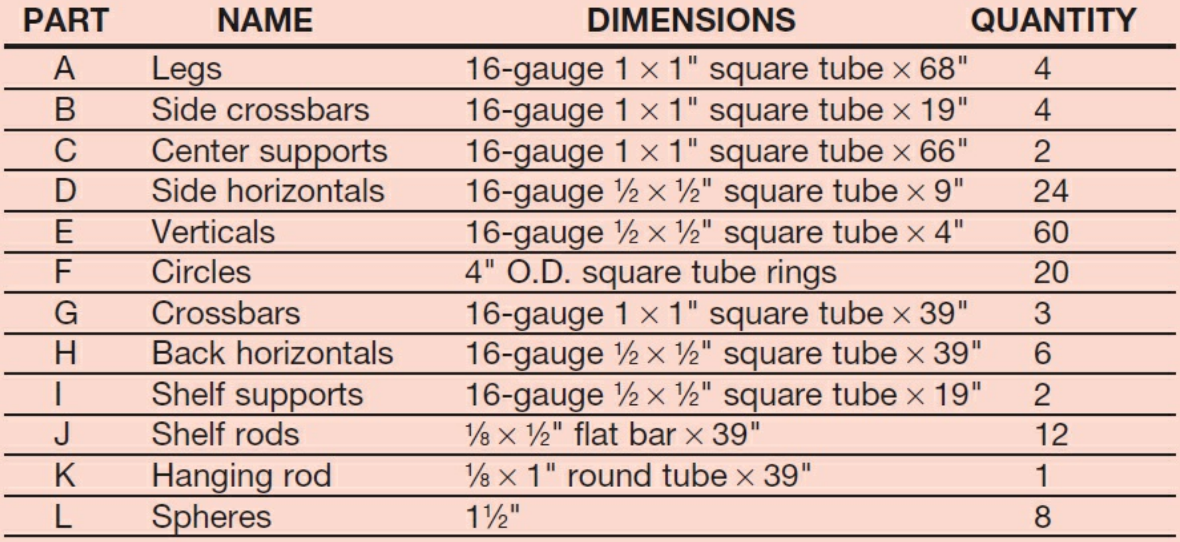

PART	NAME	DIMENSIONS	QUANTITY
A	Legs	16-gauge 1 × 1" square tube × 68"	4
B	Side crossbars	16-gauge 1 × 1" square tube × 19"	4
C	Center supports	16-gauge 1 × 1" square tube × 66"	2
D	Side horizontals	16-gauge ½ × ½" square tube × 9"	24
E	Verticals	16-gauge ½ × ½" square tube × 4"	60
F	Circles	4" O.D. square tube rings	20
G	Crossbars	16-gauge 1 × 1" square tube × 39"	3
H	Back horizontals	16-gauge ½ × ½" square tube × 39"	6
I	Shelf supports	16-gauge ½ × ½" square tube × 19"	2
J	Shelf rods	⅛ × ½" flat bar × 39"	12
K	Hanging rod	⅛ × 1" round tube × 39"	1
L	Spheres	1½"	8

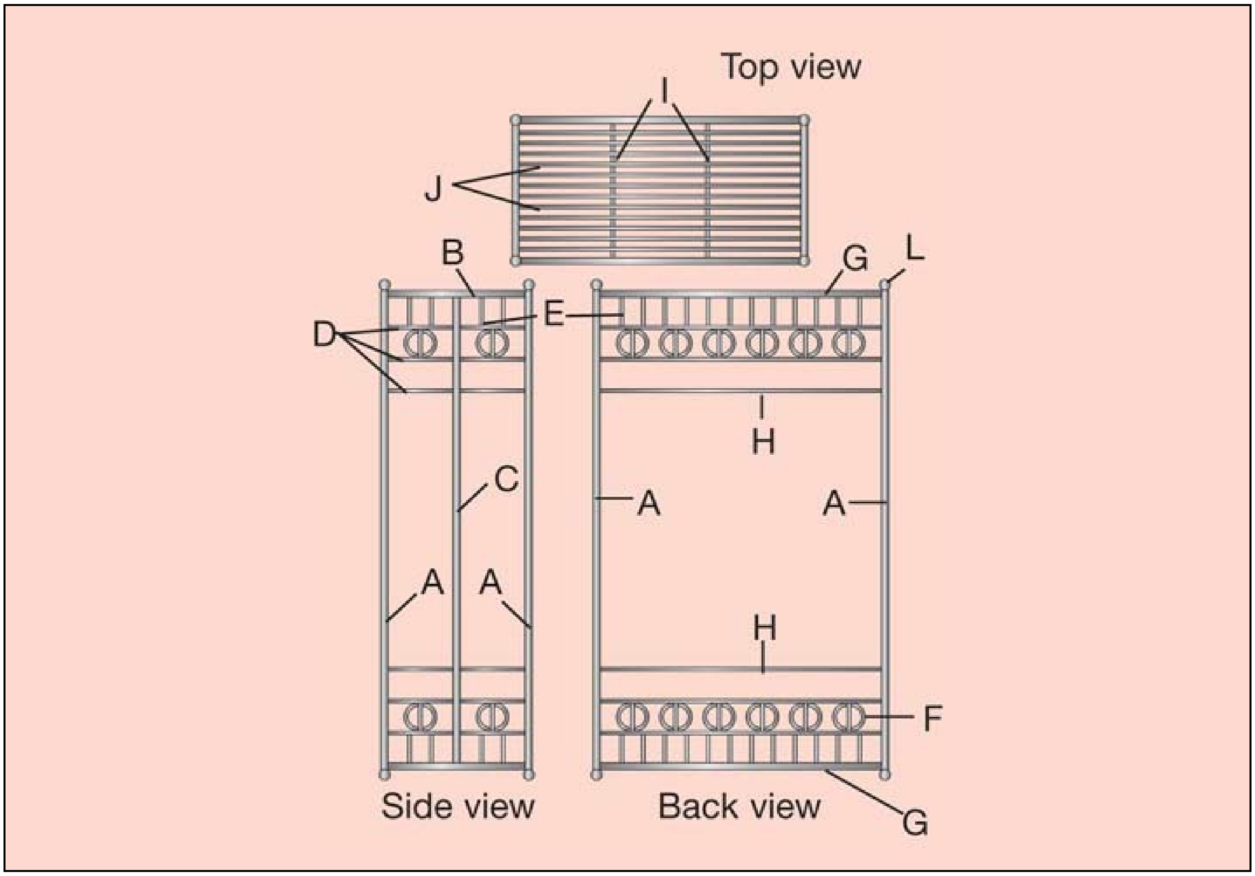
Top view
I
J
B
G
L
E
D
C
A
A
H
H
A
A
F
Side view
Back view
G

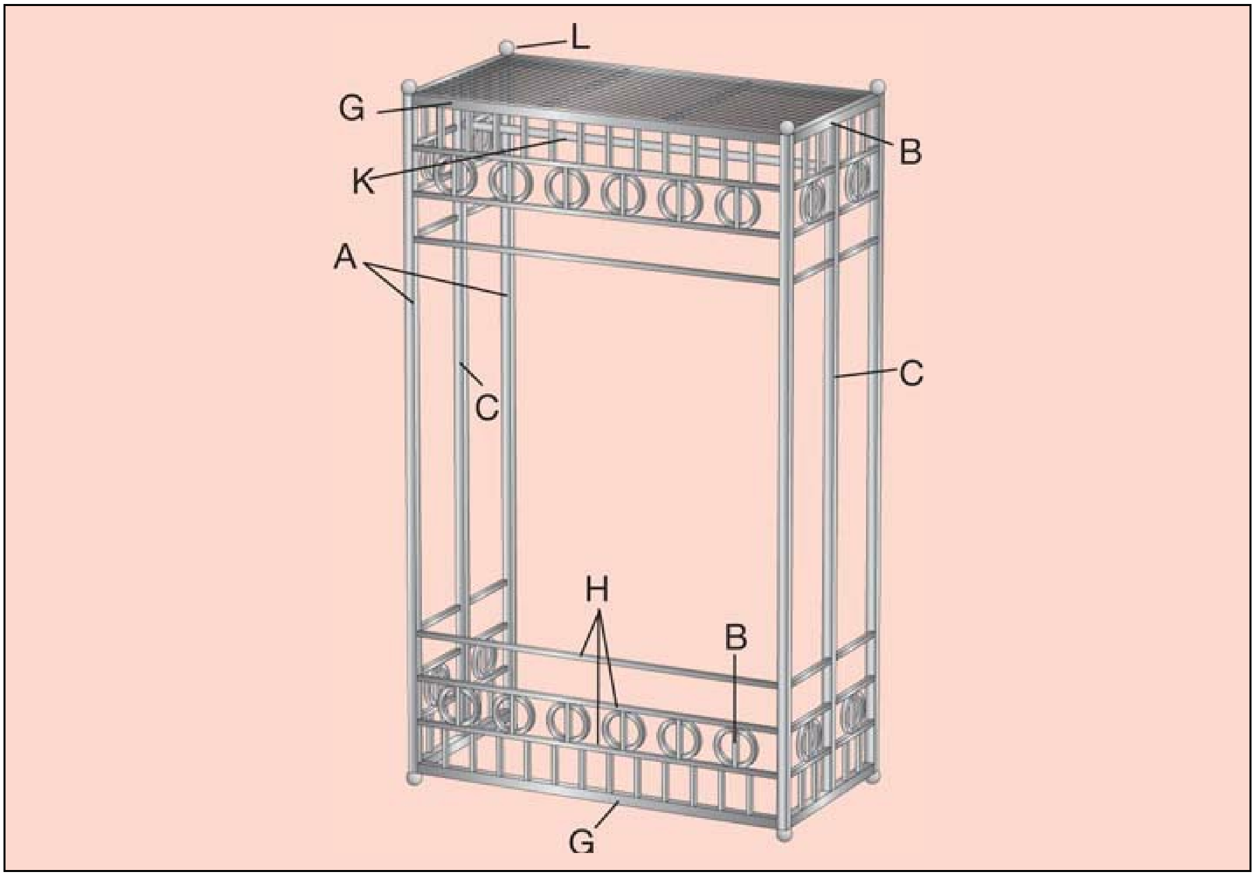
L
G
B
K
A
C
C
H
B
G

Votive Lantern with Chain and Holder

Materials

- 22-gauge sheet metal (3, 6 × 18" sheets)
- Glass (7 × 11")
- 1/8" round rod (4 feet)
- ¼" round rod (26")
- 3/16" sheet metal (10 × 10")
- Hardwood stick
- 1¼" wire nails (7)

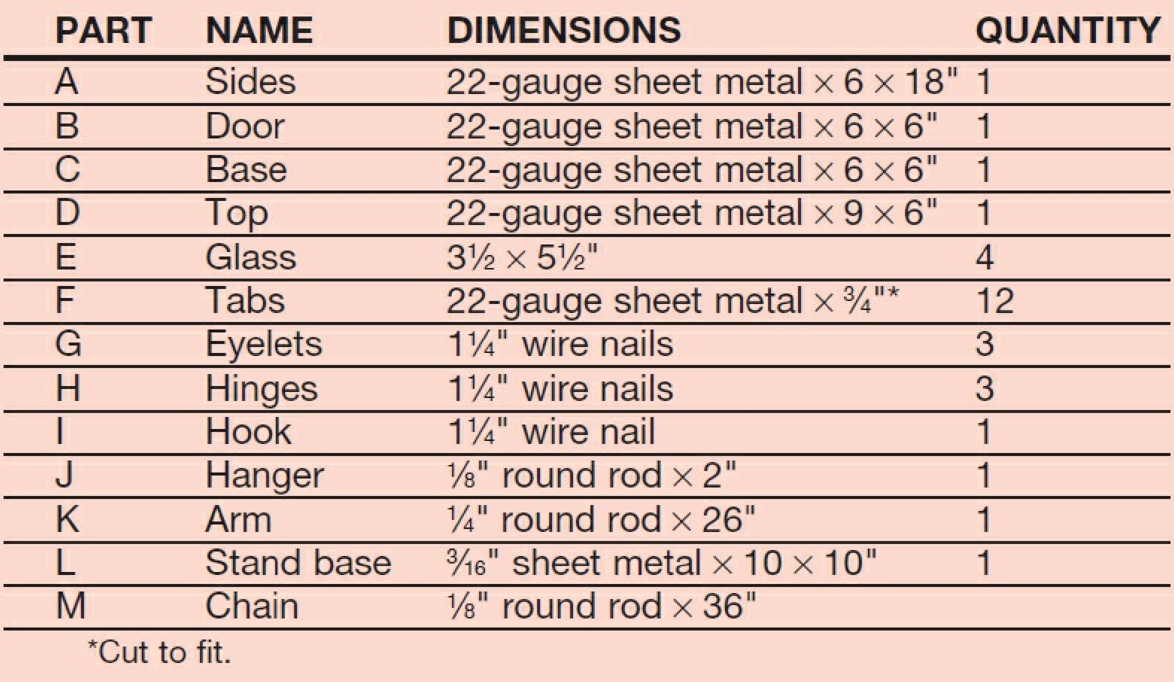

PART	NAME	DIMENSIONS	QUANTITY
A	Sides	22-gauge sheet metal × 6 × 18"	1
B	Door	22-gauge sheet metal × 6 × 6"	1
C	Base	22-gauge sheet metal × 6 × 6"	1
D	Top	22-gauge sheet metal × 9 × 6"	1
E	Glass	3½ × 5½"	4
F	Tabs	22-gauge sheet metal × ¾"*	12
G	Eyelets	1¼" wire nails	3
H	Hinges	1¼" wire nails	3
I	Hook	1¼" wire nail	1
J	Hanger	⅛" round rod × 2"	1
K	Arm	¼" round rod × 26"	1
L	Stand base	3⁄16" sheet metal × 10 × 10"	1
M	Chain	⅛" round rod × 36"	

*Cut to fit.

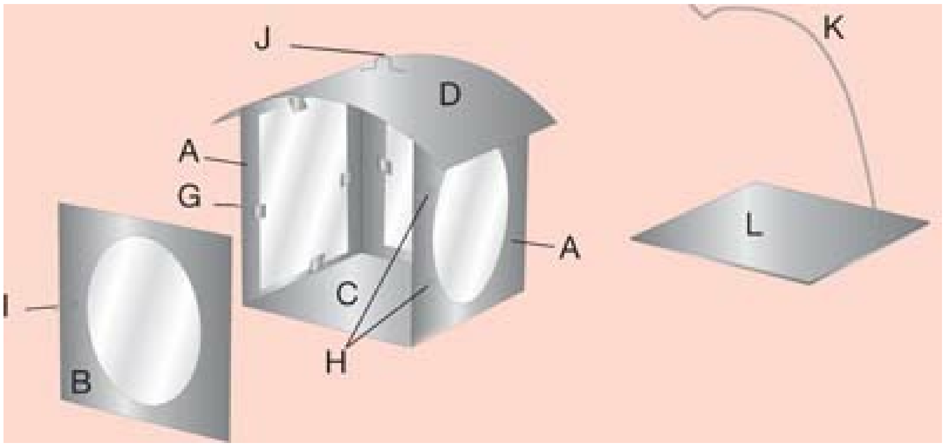
J
K
D
A
G
L
A
I
C
H
B

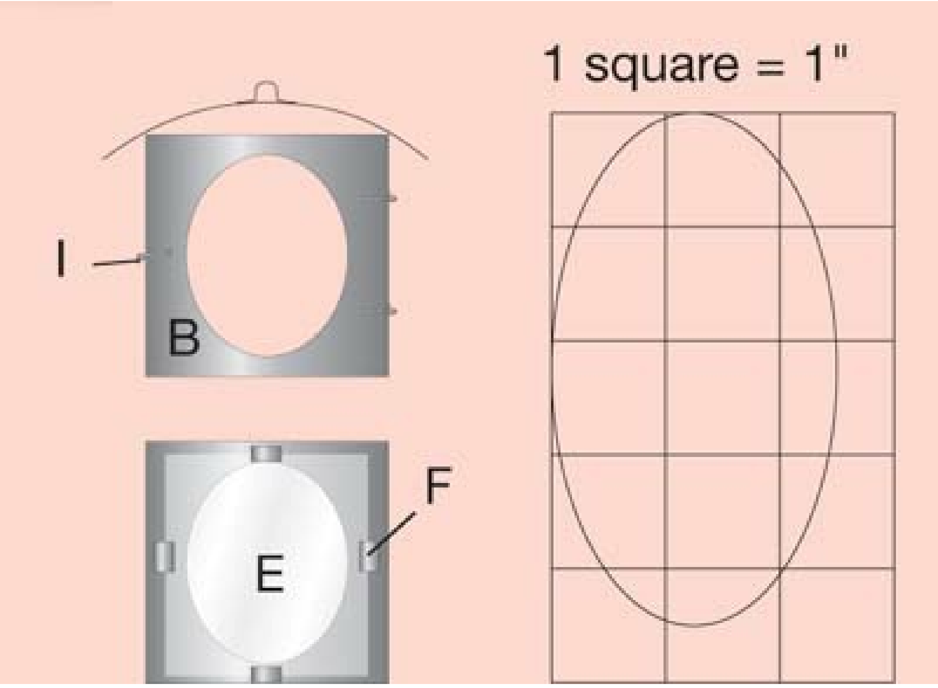
1 square = 1"
I
B
F
E

Hanging Chandelier

Materials

- 3/16" round rod (24 feet)
- 2⅜" bobeches (6, Architectural Iron Designs #79/6)
- Candle cups (6, Architectural Iron Designs #78/4)
- 8½" acanthuses (6, Architectural Iron Designs #11942 8½)
- 1" teardrop crystals (24, Chandelierparts)
- 3" hangdrop crystal prisms (5, Chandelierparts)
- 3" teardrop prisms (12, Chandelierparts)

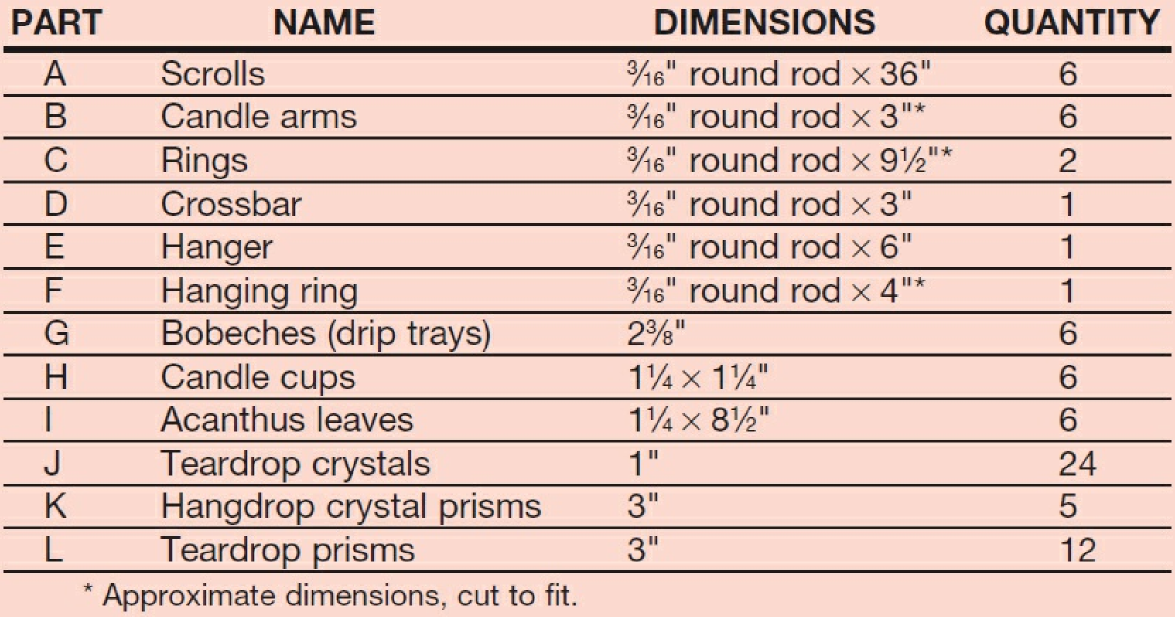

PART	NAME	DIMENSIONS	QUANTITY
A	Scrolls	3/16" round rod × 36"	6
B	Candle arms	3/16" round rod × 3"*	6
C	Rings	3/16" round rod × 9½"*	2
D	Crossbar	3/16" round rod × 3"	1
E	Hanger	3/16" round rod × 6"	1
F	Hanging ring	3/16" round rod × 4"*	1
G	Bobeches (drip trays)	2⅜"	6
H	Candle cups	1¼ × 1¼"	6
I	Acanthus leaves	1¼ × 8½"	6
J	Teardrop crystals	1"	24
K	Hangdrop crystal prisms	3"	5
L	Teardrop prisms	3"	12

* Approximate dimensions, cut to fit.

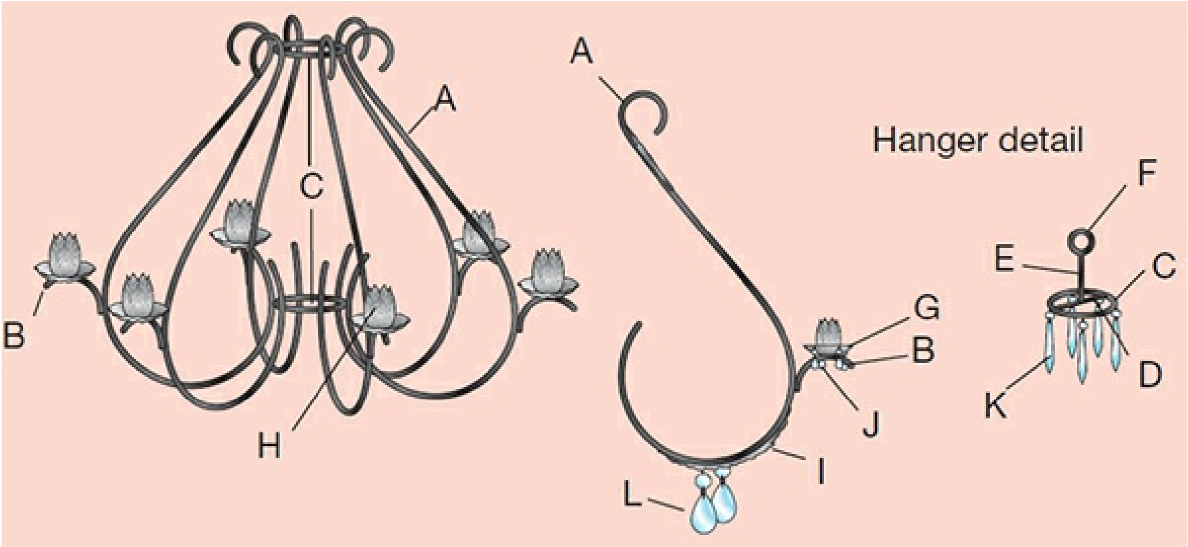
A
C
B
H
A
G
B
J
I
L
Hanger detail
F
E
C
K
D

Window Box

Materials

- 1 × ½ × 1/8" channel (18 feet)
- 1" square tube (20")
- ½" square tube (16 feet)
- ½" square bar (8 feet)
- 1/8 × ½" flat bar (14½ feet)
- 1/8 × 1" flat bar (8")
- Finials (2, Architectural Iron #8¾)

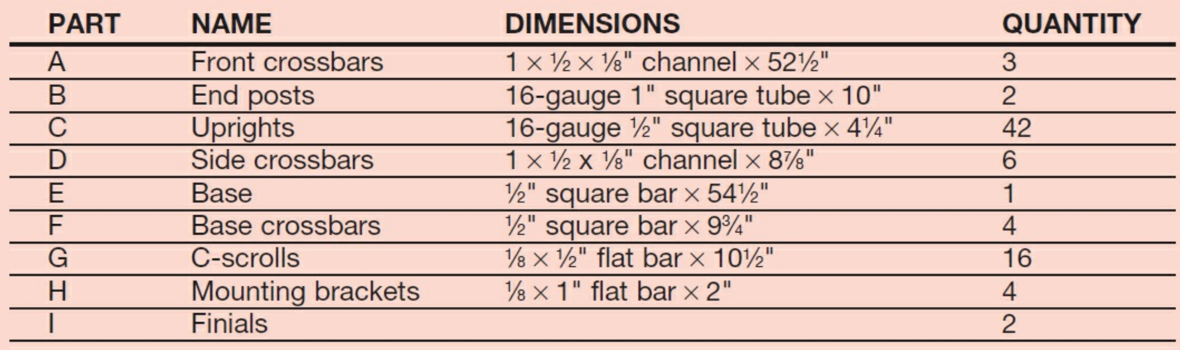

PART	NAME	DIMENSIONS	QUANTITY
A	Front crossbars	1 × ½ × ⅛" channel × 52½"	3
B	End posts	16-gauge 1" square tube × 10"	2
C	Uprights	16-gauge ½" square tube × 4¼"	42
D	Side crossbars	1 × ½ x ⅛" channel × 8⅞"	6
E	Base	½" square bar × 54½"	1
F	Base crossbars	½" square bar × 9¾"	4
G	C-scrolls	⅛ × ½" flat bar × 10½"	16
H	Mounting brackets	⅛ × 1" flat bar × 2"	4
I	Finials		2

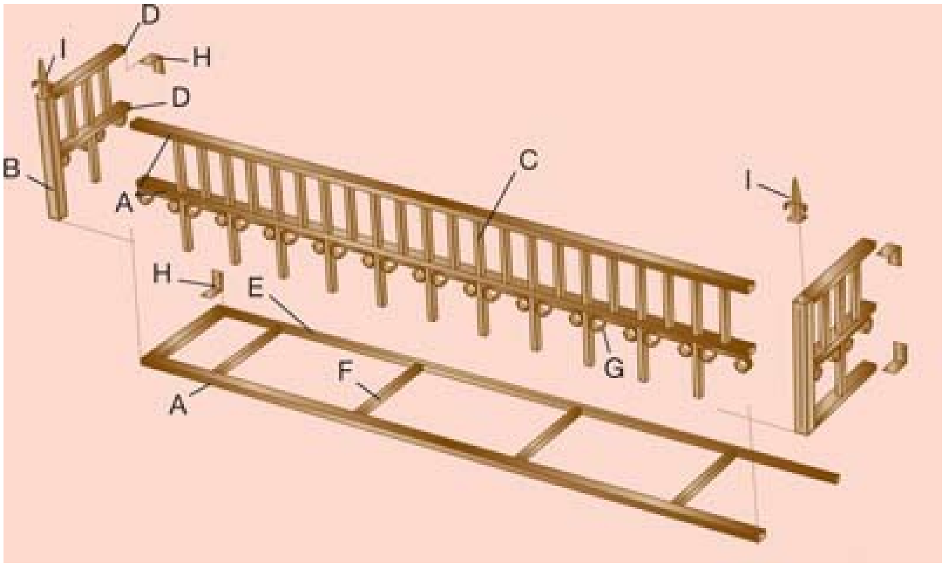
D
I
H
D
B
A
C
I
H
E
F
G
A

Laundry-Tub Planter

Materials

- 3/8" round bar (30 feet)
- 1/8 × 1" flat bar (9 feet)
- 10.5 gal. laundry tub (1, Main St. Supply)
- ½" plywood (4 × 4 sheet)
- 2 × 4 scraps
- Wood screws
- 3/16" round bar (10 feet)

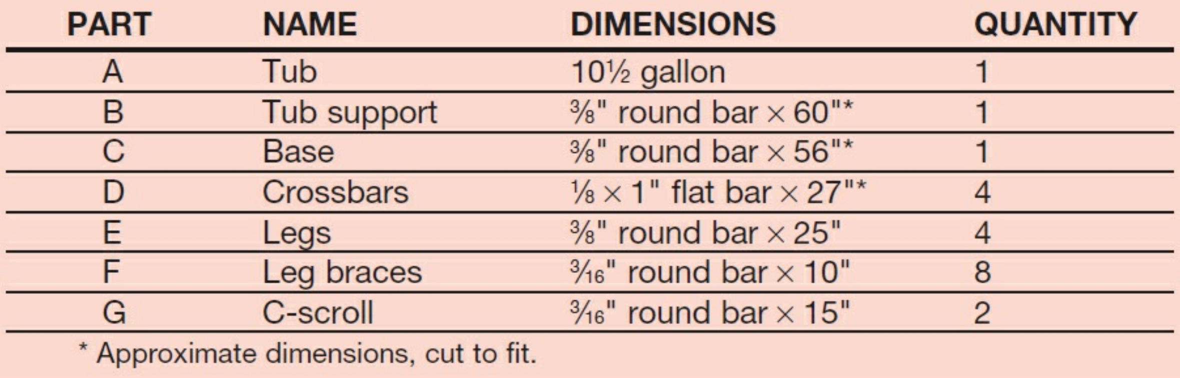

PART	NAME	DIMENSIONS	QUANTITY
A	Tub	10½ gallon	1
B	Tub support	⅜" round bar × 60"*	1
C	Base	⅜" round bar × 56"*	1
D	Crossbars	⅛ × 1" flat bar × 27"*	4
E	Legs	⅜" round bar × 25"	4
F	Leg braces	3⁄16" round bar × 10"	8
G	C-scroll	3⁄16" round bar × 15"	2

* Approximate dimensions, cut to fit.

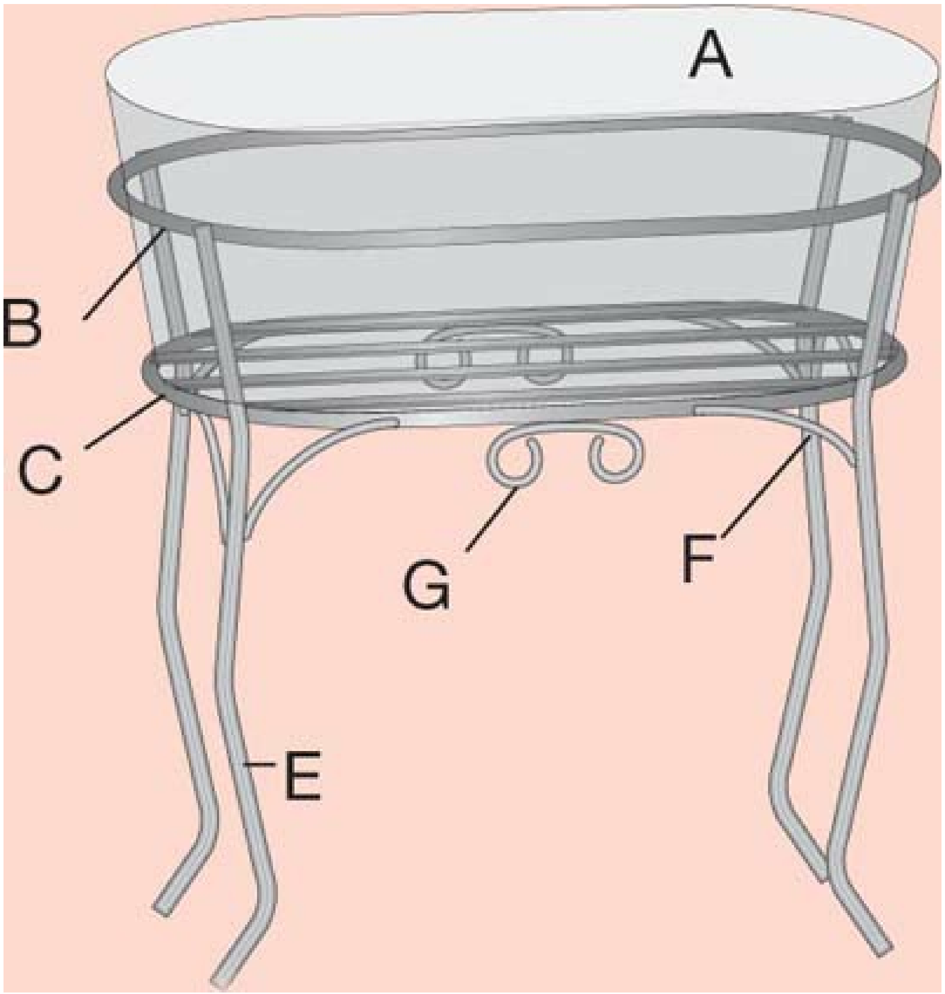
A
B
C
G
F
E

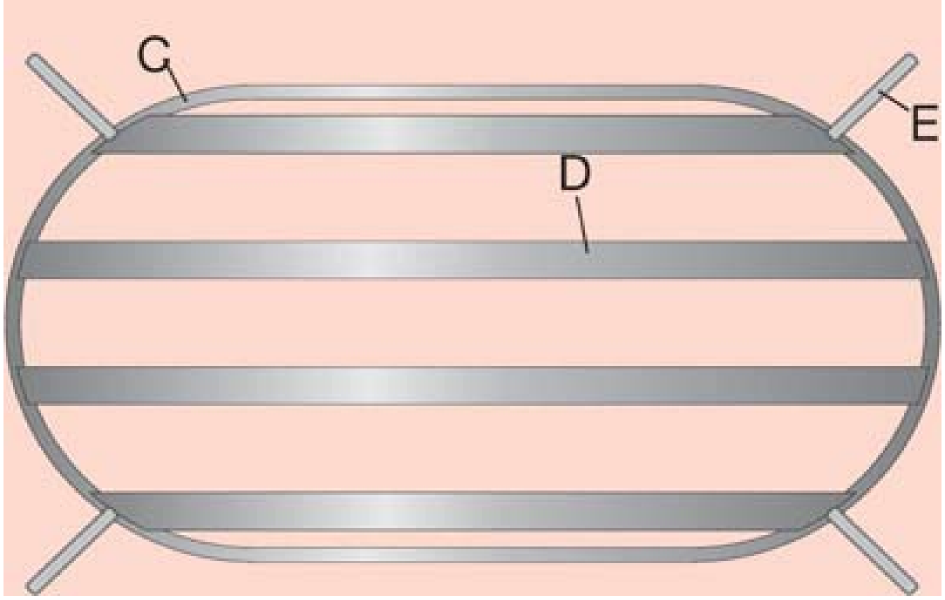
C
E
D

Three Post Floor Lamp

Materials

- ½" round tube (48 3/8")
- ½" twisted square rod (96")
- 3/16" sheet (10 × 10")
- 3/16 × 2" flat bar (8")
- 1" round balls (4)
- Lamp hardware and shade

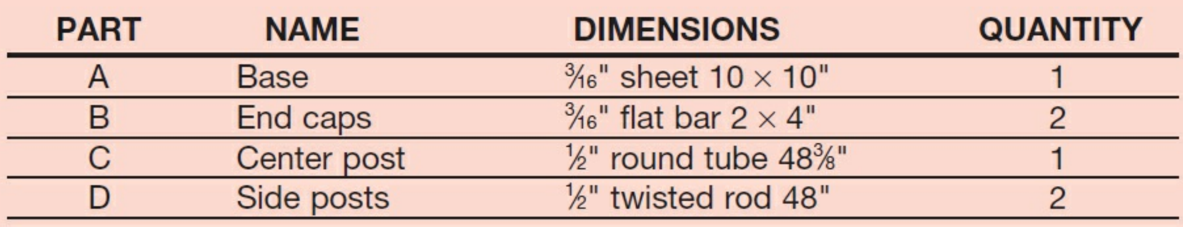

PART	NAME	DIMENSIONS	QUANTITY
A	Base	3/16" sheet 10 × 10"	1
B	End caps	3/16" flat bar 2 × 4"	2
C	Center post	½" round tube 48⅜"	1
D	Side posts	½" twisted rod 48"	2

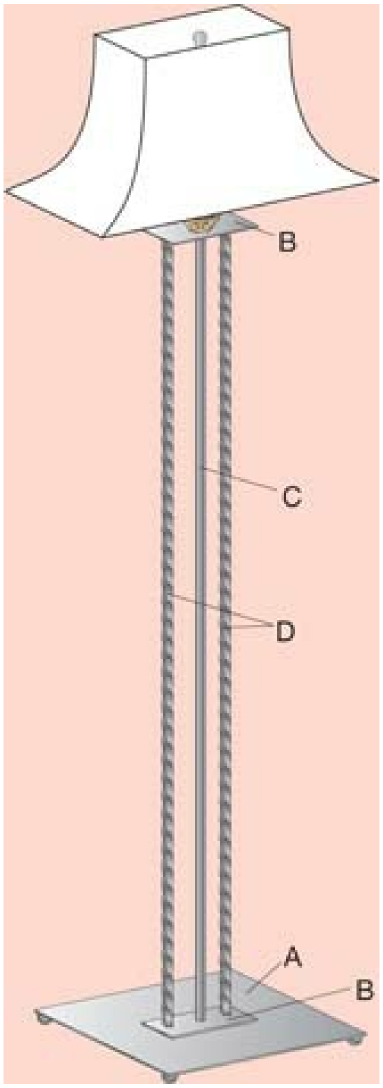
B
C
D
A
B

Glossary

Active gas	A gas, such as oxygen, that will react with other substances.
Alloy	A substance composed of two or more metals or a metal and other non-metal material.
Arc welding	Welding processes that use an electric arc to produce heat to melt and fuse the base metals. Often used to refer specifically to shielded metal arc welding.
Backfire and flashback	Combustion taking place inside the oxyacetylene torch creating a loud pop or explosion.
Base metal	The metal that is being welded, brazed, braze welded, or cut.
Bead or weld bead	The seam between workpieces that have been joined with welding.
Braze	A process used to join metals using a non-ferrous material that melts above 850° F but below the melting point of the base metal. Brazing uses capillary action to join closely fitted parts.
Braze welding	A process used to join metals with a filler material that melts above 840° F and below the melting point of the base metal where the filler metal is not distributed by capillary action.
Butt joint	A joint between two workpieces in the same plane.
Carburizing flame	An oxyfuel flame with an excess of fuel.
Consumable electrode	An electrode that also serves as the filler material.

Corner joint	A joint between workpieces meeting at right angles and forming an L shape.
Cutting tip	Converts an oxyfuel welding torch into an oxyfuel cutting torch.
Direct current electrode negative	Direct current welding where the electrode is negative and the workpiece is positive.
Direct current electrode positive	Direct current welding where the electrode is positive and the workpiece is negative.
Duty cycle	The amount of continuous time in a 10-minute period that an arc welder can run before it needs to cool down. Expressed as a percentage at a given amperage output.
Edge joint	A joint between parallel workpieces.
Electrode	The conductive element that makes the connection with the workpiece to create an electrical arc.
FCAW	Flux cored arc welding.
Filler, filler metal	Metal added to a welded joint.
Fillet weld	A triangular shaped weld between two members that meet at 90° angles.
Flame cutting	See oxyfuel cutting.
Flux	A chemical compound that produces cleaning action and reduces the formation of oxides when heated.

Flux core, flux cored wire	An electrode for flux cored arc welding that contains flux within a wire tube.
Flux cored arc welding (FCAW)	A semi-automatic arc welding process using an electrode wire that contains flux.
Gas metal arc welding (GMAW)	A semi-automatic arc welding process using a wire electrode, which also is the filler material. An inert gas is distributed over the weld area to shield the molten metal from oxygen. Commonly referred to as MIG (metal inert gas) or wire feed.
Gas tungsten arc welding (GTAW)	An arc welding process using a tungsten electrode, hand-held filler material, and an inert shielding gas. Also referred to as TIG (tungsten inert gas) and Heliarc.
GMAW	Gas metal arc welding.
Groove weld	A weld made in grooves between workpieces.
GTAW	Gas tungsten arc welding.
Heliarc	Gas tungsten arc welding.
Inert gas	A non-reactive or non-combining gas such as argon or helium.
Kerf	The width of a cut.
Lap joint	A joint between overlapping workpieces.
MIG	Metal inert gas—see gas metal arc welding.
Oxyacetylene cutting	Oxyfuel cutting using acetylene as the fuel gas.
Oxyacetylene welding	Oxyfuel welding using acetylene as the fuel gas.

Oxyfuel cutting	Cutting process using the combustion of a pressurized fuel gas and oxygen to heat steel to 1600° F at which time a pure oxygen stream is delivered to burn through the metal. Also called flame cutting.
Oxyfuel welding	Welding process that uses the heat produced by the combustion of a pressurized fuel gas and pressurized oxygen. A hand-held filler material is used. Also called gas welding.
Plasma cutting	An arc cutting process using a constricted arc. Compressed air or inert gas blows the molten metal from the kerf.
Plate	Flat metal thicker than $^{3}/_{16}$".
Saddle joint	A joint between round tubes where one tube has been cut to fit around the other.
Sheet metal	Flat metal $^{3}/_{16}$" or less in thickness.
Shielded metal arc welding (SMAW)	An arc welding process using a flux coated consumable electrode. Also referred to as arc welding or stick welding.
Shielding gas	A gas that prevents contaminants from entering the molten weld pool.
Slag	Oxidized impurities formed as a coating over the weld bead; waste material found along the bottom edge of an oxyfuel cut.
SMAW	Shielded metal arc welding.
Spatter	Small droplets or balls of metal stuck to the base metal around the weld. Produced by shielded metal arc, gas metal arc, and flux cored arc welding.
Stick welding	Shielded metal arc welding.

T-joint	A joint between workpieces meeting at right angles and forming a T shape.
TIG	Tungsten inert gas—see gas tungsten arc welding.
Wire feed	See gas metal arc welding.

Resources

Architectural Iron Designs
950 South 2nd Street
Plainfield, NJ 07063
www.archirondesign.com

Century Trailer Inc.
800 366 3395
www.centurytrailer.com

Chandelierparts.com
218 763 7000

Decorative Iron
10600 Telephone Road
Houston, TX 77075
888 380 9278
www.decorativeiron.com

Eastwood Company
800 343 9353
www.eastwoodco.com

e-Tarps.com
1205-202 Winter Springs Court
Louisville, KY 40243
www.e-tarps.com

Garland's Inc.
2501 26th Avenue South
Minneapolis, MN 55406
800 809 7888
www.garlandsinc.com

Hooker Glass Mirror & Beveling Group
1127 East El Dorado Street
Decatur, IL 62521
800 274 0114
www.miror.com

The Lincoln Electric Company
22801 St. Clair Ave.
Cleveland, OH 44117
216 481 8100
www.lincolnelectric.com

M-Boss, Inc.
4400 willow Parkway
Cleveland, OH 44125
866 886 2677
www.mbossinc.com

Main Street Supply
PO Box 729
Monroe, NC 2811
800 624 8373
www.mainstsupply.com

Miller Electric Manufacturing Co.
1635 W. Spencer St.
P.O. Box 1079
Appleton, WI 54912-1079
920 734 9821
www.millerwelds.com

National Ornamental & Miscellaneous Metals Association (NOMMA)
1535 Pennsylvnia Ave.
McDonogh, GA 30253
888 516 8586
www.nomma.org

Northern Tool + Equipment
800 221 0516
www.northerntool.com

Norton Metals
1350 Lawson Road
Fort Worth, TX 76131
817 232 0404
www.nortonmetals.com

Red Trailers
Website helps you locate the perfect hitch, drawbar, and wiring to fit your vehicle.

215 538 1155
www.redtrailers.com

SIC Metals & Fabrication, LLC
Diamond plate, aluminum materials, stainless steel, and fabrication; also welding services and CAD design.
10883 Rt 322
Shippenville, PA 16254
814 227 5546 (fax 814 227 2536)
email sales@sicmetalfab.com
www.sicmetalfab.com

Silent Source
58 Nonotuck Street
Northampton, MA 01062
413 584 7944
www.silentsource.com

Smith & Hawken
PO Box 8690
Pueblo, CO 81008-99998
800 940 1170
www.smithandhawken.com

Toll Gas & Welding Supply
3005 Niagara Lane
Plymouth, MN 55447
763 551 5300
www.tollgas.com

Triple-S Steel Supply Co.
6000 Jensen Drive
Houston, TX 77026
713 697 7105
www.sss-steel.com

Wagner Railing Systems & Components/JG Braun Co.
10600 West Brown Deer Road
Milwaukee, WI 53224
800 786 2111
www.rbwagner.com
www.jgbraun.com

Photo Credits

Devilbis Industrial Finishing
195 Internationale Blvd.
Glendale Heights, IL 60139
630 237 5000
© p. 179

iStock Photo
p. 17 (top)

Lincoln / www.lincolnelectric.com
© p. 29, 30 (top), 31, 32 (top)

Miller / www.millerwelds.com
© p. 4, 8, 32 (lower)

SIC Metals & Fabrication, LLC / www.sicmetalfab.com
p. 30 (lower) Brake design courtesy of SIC

Safety

Read, understand, and follow safety precautions and rules when welding. In addition to general safety rules, also reference the material safety data sheet (MSDS) as provided by manufacturers as well as the American National Standard (ANSI Z49.1) as available at these websites for free download:

American Welding Society

www.aws.org

The Lincoln Electric Company

www.lincolnelectric.com/community/safety

Welding safety publications available as PDF download (see E205 document).

Index

A

Acetone for cleaning, 21
Acetylene basics, 41
Alcohol for cleaning, 21
Alternating current (AC), 59
Aluminum
  basics, 15
  cutting process for, 14
  repairing cast, 20
  welding processes for, 14
Amperage/amps, defined, 52
Angle/angle iron, uses & dimensions, 16
Angled metal shelves, building, 146–149
Antique finishes, 38, 39
Arbor, building, 206–210
Arc burns, preventing, 10
Arch and gate, building, 194–199
Arcs, striking, 74
Arc welding
  building
  rolling welding curtain for, 82–83
  welding table for, 84–87
  electricity terms & concepts for, 52–53
  See also specific types

B

Backfire, 45
Baker's shelves, plans for, 212
Base metal, defined, 6
Bird's nest, removing, 67
Blowpipes. See Torches
Brass welding & cutting processes, 14

Brazing/braze welding, 50–51, 184–185
Burns, preventing, 10

C

Candelabrum, building, 142–143
Candleholders
  building for fireplace, 142–143
  plans for
  fireplace, 212
  tree-shaped, 211
  wall-mounted, 212
Cast aluminum, repairing, 20
Cast iron
  characteristics of, 17
  cutting process for, 14
  repairing, 20
  steel and, 190
  welding processes for, 14
Chandeliers, plans for, 213, 216
Channels, uses & dimensions, 16
Check valve basics, 43
Chrome moly steel welding & cutting processes, 14
Cleaning metals, 21
Clothing, safe, 10, 11, 21
Coat hanger rack, building, 120–123
Coat hooks, building swing-away, 136–137
Coat racks
  building
  with swing-away hooks, 136–137
  wall-mounted, 120–123
  plans for
  corner, 213
  free-standing, 215
Coffee table, building, 110–111
Constant current (CC), described, 52
Constant voltage (CV), described, 52

Corner coat rack, plans for, 213
Cross, building, 128–131
Currents used, described, 53
Cutouts, making, 36
Cutting techniques
  mechanical, 24
  plasma, 27, 29, 30–31
Cylinder basics, 41
Cylinder cart, building, 90–93
Cylinders, gas, 61
Cylinder valve basics, 41

D
Decorative products, 18
Degreasers for cleaning, 21
Desk lamp, plans for, 211
Diamond plate metal, uses for, 17
Direct current (DC), described, 53
Direct current electrode negative (DCEN), described, 53
Direct current electrode positive (DCEP), described, 53
Duty cycle, described, 53

E
Electrical units, described, 52
Electricity, terms & concepts, 52–53
Electric shocks, preventing, 11
Electrodes
  direct current electrode negative (DCEN), described, 53
  direct current electrode positive (DCEP), described, 53
  for gas metal arc welding, 61
  for gas tungsten arc welding, 70–72
Equipment
  building
  cylinder cart, 90–93
  grind stand, 94–95
  portable welding table, 96–101

rolling welding curtain, 82–83
saw stand, 102–107
welding machine cart, 88–89
welding table, 84–87
for cleaning metal, 21
for gas metal arc welding, 59–61
for gas tungsten arc welding, 68–73
for mechanical cutting, 24
for oxyacetylene welding, 40–43
for plasma cutting, 26, 29, 32
for shielded metal arc welding (SMAW), 55–56
shop tools, 13
Étagères, building
corner, 160–163
straight, 154–159
Explosions, preventing, 10–11

F
Fence panel, building, 124–127, 190–193
Ferrous metals, 14
Filler metals, defined, 6
Finishing, 38–39
Fire, preventing, 10
Fire bricks, 48, 49
Firepit, building, 186–189
Fireplace candelabrum, building, 142–143
Fireplace candleholder, plans for, 212
Fittings basics, 42
Flashback, 45
Flashback arrestor basics, 43
Flat bars, dimensions, 16
Floor lamp, plans for three post, 217
Flux cord arc welding, described, 7, 65
Free-standing coat rack, plans for, 215
Fuels
for oxyacetylene cutting, 34

for oxyacetylene welding, 49
Fumes, avoiding, 10

G
Garden trellis, building, 124–127
Gas cylinder safety, 61
Gas metal arc welding (GMAW/MIG)
  building welding machine cart for, 88–89
  described, 6, 7, 58
  directions, 64–65
  setting up shielding gas flow meter, 63
  setting up wire feed, 62
  electricity terms & concepts for, 52–53
  equipment for, 59–61
  lens shade numbers for, 11
  maintenance, 66–67
  requirements, 12
  safety, 59
  troubleshooting, 66–67
Gas options for oxyfuel welding, 49
Gas tungsten arc welding (GTAW/TIG)
  building welding machine cart for, 88–89
  described, 6, 7, 68
  directions, 74–75
  electricity terms & concepts for, 52–53
  equipment for, 68–73
  lens shade numbers for, 11
  requirements, 12
  safety, 69
  troubleshooting, 76–77
Gate, building, 184–185, 194–199
Gauge thickness, inch equivalents, 18
Gazebo, building, 200–205
Glossary, 218–219
Grind stand, building, 94–95

Gun basics, 59–60

H

Heat distortion, described, 6–7
Hex/hexagonal bars, dimensions, 16
Holes, piercing, 36
Hose basics, 42

I

Inverters, described, 53
Iron cross, building, 128–131

L

Lamps, plans for
    scroll desk, 211
    three post floor, 217
Laundry-tub planter, plans for, 217
Leaks, checking for, 45
Lens shade numbers, 11

M

Magazine rack, building, 143–145
Mechanical cutting techniques, 24
Mechanical generators, described, 53
Melting points, 15
Metals
    basics, 14–15
    cleaning, 21
    decorative, 18
    ferrous, 14
    filler, 6
    identification chart, 17
    melting points, 15
    methods to join, 6
    purchasing, 18
    repairing, 19–20
    shapes & sizes, 16–17

Mild steel
  basics, 14
  cutting processes for, 14
  repairing, 19
  shaping, 37
  welding processes for, 14
Mirror, building swivel, 114–119
Molten puddle
  defined, 6
  protecting from oxygen, 7

N
Nesting tables, building, 132–135
Nightstand, building, 138–141

O
Open circuit voltage, described, 52
Operating voltage, described, 52
Oxyacetylene cutting
  equipment for, 33–34
  fuel options, 34
  precutting checklist, 35
  safety, 32, 36
  techniques, 35–36
  troubleshooting, 36
Oxyacetylene welding
  depressurizing outfit, 49
  described, 6, 7, 40
  directions, 48
  equipment for, 40–43
  flame & flame states, 47
  fuel, 49
  requirements, 12
  safety, 41, 42
  setting up outfit, 44
  torch lighting, 46

  - troubleshooting, 45
  - turning on outfit, 45

Oxyfuel welding
  - building
  - cylinder cart for, 90–93
  - welding table for, 84–87
  - gas options for, 49

Oxygen, protecting molten puddle from, 7

Oxygen cylinder basics, 41

P

Penetration, described, 6

Perforated sheet metal basics, 17

Planter, plans for laundry-tub, 217

Plasma cutting
  - bending & cutting diamond plate, 30
  - building rolling welding curtain for, 82–83
  - electricity terms & concepts for, 52–53
  - equipment for, 26, 29, 32
  - overview of, 25
  - precutting checklist, 27
  - safety and, 25–26
  - techniques, 27, 29, 30–31
  - troubleshooting, 28

Plate metal, thicknesses, 16

Portable welding table, building, 96–101

Power sources, described, 53

Pressure regulator & gauge basics, 42

Primers and finishing, 38, 39

R

Rail caps, uses & dimensions, 16

Railing, building, 180–183

Rectangular tubes, uses & dimensions, 16

Repairing metal directions, 19–20

Reverse polarity, 53

Rolling welding curtain, building, 82–83
Room divider, building, 124–127
Round bars, dimensions, 16
Round tubes, uses & dimensions, 16

S
Safety, 10–11
  with gas cylinders, 61
  truck rack materials and, 168
  when cleaning metal, 21
  when gas metal arc welding, 59
  when gas tungsten arc welding, 68, 69
  when oxyacetylene cutting, 26, 33, 36
  when oxyacetylene welding, 41, 42
  when plasma cutting, 25–26
  when shielded metal arc welding, 54–55
Saw stand, building, 102–107
Scratch starts, 74
Scroll desk lamp, plans for, 211
Shapes of metals, 16
Shaping, 37
Sheet metal, thicknesses, 16
Shelves
  building
  angled metal, 146–149
  étagères
  corner, 160–163
  straight, 154–159
  wall-mounted, 150–153
  plans for baker's, 212
Shielded metal arc welding (SMAW)
  described, 6, 7, 54
  directions, 57
  electricity terms & concepts for, 52–53
  equipment for, 55–56

- lens shade numbers for, 11
- requirements, 12

Shielding gases
- for gas metal arc welding, 61
- for gas tungsten arc welding, 70

Shop requirements, 12

Shop tools needed, 13

Spray-on finishes, 38

Square bars, dimensions, 16

Square tubes, uses & dimensions, 16

Stainless steel
- characteristics of, 17
- cutting process for, 14
- repairing, 20
- welding processes for, 14

Stair railing, building, 180–183

Steel
- basics, 14–15
- cast iron and, 190
- characteristics of, 17
- repairing, 20

Step-down transformer, 53

Stick welding. *See* Shielded metal arc welding (SMAW)

Straight polarity, 53

Straps, dimensions, 16

Stub out/stubbing, 66

Swing-away coat hooks, building, 136–137

Swivel mirror, building, 114–119

T

Tables
- building
- for arc welding, 84–87
- coffee, 110–111
- nesting, 132–135
- nightstand, 138–141

  portable welding, 96–101
  plans for tapered-leg, 212
T-bars, dimensions, 16
Thick materials, welding to thin, 159
Three post floor lamp, plans for, 217
Tips for torches, 43
Titanium welding & cutting processes, 14
Tools. *See* Equipment
Torches
  for gas tungsten arc welding, 69–70
  for oxyacetylene welding
  described, 43
  flame & flame states, 47
  lighting, 46
  maintaining, 49
  tips for, 43
Trailer frame, building, 172–179
Tree-shaped candleholder, plans for, 211
Trellises, building, 124–127, 190–193
Truck rack, building, 166–171

V
Verdigris finishes, 38–39
Vineyard trellis, building, 190–193
Voltage/volts, defined, 52
Votive lantern with chain & holder, plans for, 215

W
Wall-mounted candleholder, plans for, 212
Wall-mounted coat rack, building, 120–123
Wall-mounted shelves, building, 150–153
Wattage/watts, defined, 52
Welding
  electricity terms & concepts for, 52–53
  overview of, 6–7
Welding curtain, building, 82–83

Welding machine cart, building, 88–89
Welding table, building, 84–87
Window box, plans for, 216
Wine rack, building, 112–113
Wire feed basics, 59
Work clamps, 53, 60
Work leads, 53
Wrought iron, characteristics of, 17

Copyright 2009
Creative Publishing international, Inc.
400 First Avenue North, Suite 300
Minneapolis, MN 55401
1-800-328-0590
www.creativepub.com
All rights reserved

Library of Congress Cataloging-in-Publication Data

Welding complete : techniques, project plans & instructions.
p. cm.
Includes index.
Summary: “Explores the practical and decorative aspects of welding, including information on tools, materials, safety and dozens of metalwork projects” – –Provided by publisher.
ISBN-13: 978-1-58923-455-0 (soft cover)
ISBN-10: 1-58923-455-3 (soft cover)
Digital edition: 978-1-58923-455-0
1. Welding. 2. Welding– –Equipment and supplies. I. Creative Publishing International. II. Title.

TT211.W45 2009
739'.14– –dc22

2009010736

President/CEO: Ken Fund
VP for Sales & Marketing: Kevin Hamric

Home Improvement Group
Publisher: Bryan Trandem
Managing Editor: Tracy Stanley
Senior Editor: Mark Johanson
Editor: Jennifer Gehlhar
Senior Design Managers: Jon Simpson, Brad Springer
Design Manager: James Kegley

Technical Review: Kris Kersten
Page Layout Artist: Rick Penn-Kraus

Lead Photographer: Joel Schnell
Photo Coordinator: Cesar Fernandez Rodriguez
Photo Set Builder: Bryan McLain
Shop Help: Scott LaSalle, Max LaSalle

Production Managers: Laura Hokkanen, Linda Halls

Additional Contributors
Todd Bridigum, Carl Peters, Jenny Ogborn, Bruce Reed at SIC.

Cover photograph: Fotosearch Stock Photography; inset images CPi

THE SCIENCE AND PRACTICE OF WELDING

Volume 2

The practice of welding

A. C. DAVIES

Tenth Edition

The Science and Practice of Welding, now in its tenth edition and published in two volumes, is an introduction to the theory and practice of welding processes and their applications.

Volume 1, *Welding Science and Technology*, explains the basic principles of physics, chemistry and metallurgy as applied to welding. The section on electrical principles includes a simple description of the silicon diode and resistor, the production and use of square wave, and one-knob stepless control of welding current. There is a comprehensive section on non-destructive testing (NDT) and destructive testing of welds and crack tip opening displacement testing. The text has been brought completely up to date and now includes a new chapter devoted to the inverter power unit. Duplex stainless steel has been included in the list of materials described.

Volume 2, *The Practice of Welding*, is a comprehensive survey of the welding methods in use today, and gives up-to-date information on all types of welding methods and tools. Processes described include: manual metal arc welding (MMA or SMAW); gas shielded arc welding (MIG, MAG or GMAW); tungsten inert gas welding (TIG or GTAW) and plasma arc welding (PA) and cutting. Resistance, flash butt and oxy-acetylene welding are also included. Cutting processes are given a separate chapter. This new edition has been brought right up to date with a new chapter on the welding of plastics, and new sections on the welding of duplex stainless steel and air plasma cutting. The text is illustrated by up-to-date photographs of plant and equipment. As in previous editions, the appendices bring together a wealth of essential information, including British and American welding symbols, tables of conversion, information on proprietary welding gases and mixtures, testing practices, safety features and tables of brazing alloys and fluxes.

Both volumes contain numerous questions of the type set at craftsman and technician grade of the City and Guilds of London Institute examinations.

The science and practice of welding
Volume 2
The practice of welding

The science and practice of welding

VOLUME 2

The practice of welding

A. C. DAVIES

B.Sc (London Hons. and Liverpool), C.Eng., M.I.E.E., Fellow of the Welding Institute

TENTH EDITION

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011–4211, USA
477 Williamstown Road, Port Melbourne VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

© Cambridge University Press 1963, 1972, 1977, 1984, 1989, 1993

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 1941
Second edition 1943
Third edition 1945
Reprinted 1947, 1950
Fourth edition 1955
Reprinted 1959
Fifth edition 1963
Reprinted 1966, 1969, 1971
Sixth edition 1972
Reprinted 1975
Seventh edition 1977
Reprinted with revisions 1981
Eighth edition 1984
Reprinted with revisions 1986
Ninth edition 1989
Reprinted 1990
Tenth edition 1993
Reprinted with additional Appendixes 12 and 13 1994
Reprinted with additional Appendix 14 1996
Reprinted 1998, 2000, 2002

Typeset in M Times New Roman

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data
Davies, A. C. (Arthur Cyril)
The science and practice of welding.
Includes indexes.
Contents: v. 1. Welding science and technology –
v. 2. The practice of welding.
1. Welding. I. Title.
TS227.D22 1993 671.5′2 88-6041

ISBN 0 521 43404 1 hardback
ISBN 0 521 43566 8 paperback

Transferred to digital printing 2003

Contents

Preface xi

1 **Manual metal arc welding** 1

The electric arc 1

Electrode classification (British) 14

Electrode classification (American) 22

Welders' accessories 23

The practice of manual metal arc welding 31

Welding of pipelines 73

Welding of carbon and carbon–manganese and low-alloy steels 78

Low-alloy steel classifications 83

2 **Gas shielded metal arc welding** 92

Metal inert gas (MIG), metal active gas (MAG) including CO_2 and mixed gas processes 92

Techniques 119

CO_2 welding of mild steel 129

Automatic welding 131

MIG pulsed arc welding 137

Synergic pulse MIG welding 140

3 **Tungsten electrode, inert gas shielded welding processes (TIG), and the plasma arc process** 147

Technology and equipment 147

Welding techniques 173

Automatic welding 185

Plasma-arc welding 189

4 **Resistance welding and flash butt welding** 197

Spot welding 197

Electrodes 204

Seam welding 208

Projection welding 211

Resistance butt welding 212
Flash butt welding 213
5 **Additional processes of welding** 216
Submerged arc welding 216
Electroslag welding 224
Mechanized MIG welding with robots (robotics) 226
Pressure welding 231
Friction welding 234
Electron beam welding 237
Laser beam welding 239
Stud welding 242
Capacitor discharge stud welding 245
Explosive welding 247
Gravity welding 250
Thermit welding 251
Underwater welding 253
6 **Oxy-acetylene welding** 255
Principles and equipment 255
Methods of welding 274
Cast iron welding 286
Braze welding and bronze welding 288
Copper welding 298
Aluminium welding 301
Welding of nickel and nickel alloys 306
Hard surfacing and stelliting 310
Brazing 317
General precautions 323
7 **Cutting processes** 326
Gas cutting of iron and steel 326
Oxygen or thermic lance 333
Flame gouging by the oxy-acetylene process 336
Oxy-arc cutting process 338
Arc-air cutting and gouging process 339
Plasma cutting 340
Powder injection cutting 349
Underwater cutting 353
8 **The welding of plastics** 356
Plastics 356
Machine (tractor) welding 369
High frequency spark testing 370
Hot plate (or hot iron) welding 371
Hot plate welding by machine 372
Electric fusion welding 373
Ultrasonic welding 374

Vibration welding (linear) 375
Joining plastics and rubbers by means of adhesives 377

Appendixes

1 Welding symbols: British, American 381
2 Simplified notes on the operation of a thyristor 390
3 Proprietary gases and mixtures 393
4 Tests for wear-resistant surfaces 403
5 Conversion factors 404
6 Low hydrogen electrode, downhill pipe welding 405
7 The manufacture of extruded MMA electrodes 409
8 Notes on fire extinguishing 412
9 Tables of brazing alloys and fluxes 417
10 Latest plant and equipment 429
11 Refraction and reflection 442
12 Notes on the welding of Duplex and Super Duplex stainless steels 445
13 Welding fumes: health and safety 449
14 Lasers and laser welding 458
City and Guilds of London Institute examination questions 467
Manual metal arc 467
Gas shielded metal arc (MIG and TIG) 480
Other welding processes 497
Cutting processes 500
Oxy-acetylene 502
Multiple choice 507
Welding engineering craft studies 512
Fabrication and welding engineering (technical grade) 536
Index 538

Preface

The Science and Practice of Welding was divided into two volumes for the eighth edition: Volume 1, *Welding Science and Technology*; Volume 2, *The Practice of Welding*. Volume 1 covers all the basic material on the physics and chemistry of welding, including metallurgy, equilibrium diagrams, testing of welds, drawing and welding symbols and an appendix with tables and conversions. Volume 2 gives a comprehensive survey of the welding methods in use today and the tools involved.

This tenth edition has been brought completely up to date throughout. Volume 1 has a new chapter on the inverter, which has become popular as a power unit because of its reduced weight and size compared with a conventional unit. There is also an up-to-date section on the classification of stainless steels. Volume 2 has a new chapter on welding plastics and new sections on welding duplex stainless steel and air plasma cutting. There are two new appendices (one illustrating the latest plant and equipment, and one on refraction). The appendix on proprietary welding gases has been completely revised.

My thanks are due to the following firms for their help and cooperation by supplying much technical information and photographs as indicated:

AGA Gas Ltd: Industrial gases

Air Products Ltd, Crewe: special gases and mixtures, welding of aluminium, stainless and heat resistant steels.

Air Products Ltd, Ruabon: welding of aluminium and its alloys, stainless and 9% nickel steels and plasma cutting.

Alcan Wire Ltd: aluminium welding techniques and applications.

Alpha Electronics for information on measuring instruments.

Aluminium Federation: aluminium and its alloys. Welding techniques and applications.

American Welding Society: welding symbols and classifications.

Babcock Wire Equipment Ltd: cold pressure welding with photographs.

Bernard Division, Armco Ltd: MIG welding guns.

Bielomatic London Ltd: ultrasonic, hot plate and linear vibration machines and photographs.

BOC Ltd: Industrial gases and classification of stainless steels.

Brush Electrical Machines Ltd: details of thyristor controls.

Bullfinch Gas Equipment Ltd: brazing techniques and torch photographs.

Butters' Welding Equipment (Eland Group): power units.

Deloro-Stellite Ltd and Cabot Corporation: wear technology and laying down of wear resistant surfaces.

Distillers AG Ltd: Industrial gases, heaters and gauges for CO_2 cylinders.

ESAB Ltd: TIG and MIG and submerged arc processes and accessories including backing strips, robot welding systems and MIG applications, power units for MMA, MIG and TIG welding, positioners, manipulators and add on units, photographs of equipment.

Filarc Welding Industries BV (successors to Philips Export BV): low hydrogen electrode, downhill pipeline welding.

Goodburn Plastics Ltd: information and photographs of torches, nozzles and HF spark tester.

Hermann Ultrasonics Ltd: resonant unit for plastic welding.

Hobart Brothers Co Ltd: illustration of synergic welding unit and MIG unit wire drive.

INCO Alloys International Ltd, Hereford; details of the nickel–copper and chromium–nickel alloys.

Interlas Ltd: Miller synchrowave TIG unit and Pulsar MIG unit techniques and photographs of welding units including those of Miller (US) and Hitachi.

Johnson Matthey and Co.: brazing alloys and fluxes.

Loctite UK Ltd: Technology and photographs of plastics joining adhesives.

Megger Instruments Ltd: for information on instruments and testing.

Murex Welding Products Ltd (ESAB Group): MIG and TIG welding units; synergic welding plant and photographs; 230 bar gauges (Saffire) and photographs; sub arc fluxes; technology of air plasma cutting and photographs of plant.

Nederman Ltd: sketches of fume extraction plant.

Oerlikon Bührle Ltd: electrode coatings and their manufacture.

Palco Ltd: details of seam tracking and automatic magnetic arc welding with illustrations.

Power Con Incorporated (US): photographs of power units.

RGA Stainless Ltd (Smethwick): technology and classification of stainless steels.

Stewart and Lloyds Plastics Ltd: technology and photographs of electro-fusion and hot plate welding of plastics.

Welding Institute: photographs of laser welding.

Welding Torches Co Ltd (WTC) Wigan: technology and photographs of air plasma welding and cutting.

Welwyn Tool Co Ltd: technology and photographs of Leister plastic welding equipment.

Wescol Ltd: photographs of gauges for gases and flashback arrestor.

Wharton Williams Taylor 2W: underwater welding and cutting techniques.

Union Carbide Co. Ltd: plasma cutting and photographs.

I would like to express my thanks to Mr J. Crouch, Mr H. J. Davies (Murex Welding Products ESAB (UK) Group Ltd), Mr P. V. C. Watkins (ESAB (UK)), Mr J. B. Stokes and Mr W. F. J. Thomas for help in compilation of information, to Mr P. V. C. Watkins for help with the section on the synergic MIG welding process and to Mr R. P. Hampson and Mr R. M. Wilson for help with Appendix 2, Mr H. J. Davies for general help and information on air plasma cutting, Mr A. Ellis for general help in air plasma cutting and the welding of plastics, Mr J. B. Stokes for help with general compilation and Mr C. Dawes for information on laser welding.

I would also like to express my thanks to the City and Guilds of London Institute for permission to reproduce, with some amendments, examination questions set in recent years.

Extracts from British Standards are reproduced by permission of The British Standards Institution. Copies of the latest standards can be obtained from The British Standards Institution (Sales Department), Linford Wood, Milton Keynes MK14 6LE.

Extracts from American Welding Society Standards are reproduced by permission of The American Welding Society Inc., 550 North LeJeune Road, Miami, Florida 33126, USA, from whom copies of all Standards can be obtained.

Oswestry 1992 *A. C. Davies*

Note added at 1994 reprinting
I have taken this opportunity to add two new appendixes: Appendix 12 – Notes on the welding of Duplex and Super Duplex stainless steels, and Appendix 13 – Welding fumes: health and safety.

Note added at 1996 reprinting
I have now added a new appendix on Lasers and laser welding.

Oswestry 1996 *A. C. Davies*

1

Manual metal arc welding[*]

The electric arc

An electric arc is formed when an electric current passes between two electrodes separated by a short distance from each other. In arc welding (we will first consider direct-current welding) one electrode is the welding rod or wire, while the other is the metal to be welded (we will call this the plate). The electrode and plate are connected to the supply, one to the + ve pole and one to the − ve pole, and we will discuss later the difference which occurs when the electrode is connected to −ve or +ve pole. The arc is started by momentarily touching the electrode on to the plate and then withdrawing it to about 3 to 4 mm from the plate. When the electrode touches the plate, a current flows, and as it is withdrawn from the plate the current continues to flow in the form of a 'spark' across the very small gap first formed. This causes the air gap to become ionized or made conducting, and as a result the current is able to flow across the gap, even when it is quite wide, in the form of an arc. The electrode must always be touched on to the plate before the arc can be started, since the smallest air gap will not conduct a current (at the voltages used in welding) unless the air gap is first ionized or made conducting.

The arc is generated by electrons (small negatively charged particles) flowing from the −ve to the +ve pole and the electrical energy is changed in the arc into heat and light. Approximately two-thirds of the heat is developed near the +ve pole, which burns into the form of a crater, the temperature near the crater being about 6000–7000°C, while the remaining third is developed near to the −ve pole. As a result an electrode connected to the +pole will burn away 50% faster than if connected to the −ve pole. For this reason it is usual to connect medium-coated electrodes and bare rods to the −ve pole, so that they will not burn away too quickly. Heavily coated rods are connected to the +ve pole because, due to the extra heat required to melt the heavy coating, they burn more slowly than the other types of rods when carrying the same current. The thicker the electrode used, the more heat is required to melt it, and thus the more current is required. The welding current may vary from 20 to 600 A in manual metal arc welding.

---

[*] Abbreviated MMA. American designation: shielded metal arc welding (SMAW).

When alternating current is used, heat is developed equally at plate and rod, since the electrode and plate are changing polarity at the frequency of the supply.

If a bare wire is used as the electrode it is found that the arc is difficult to control, the arc stream wandering hither and thither over the molten pool. The globules are being exposed to the atmosphere in their travel from the rod to the pool and absorption of oxygen and nitrogen takes place even when a short arc is held. The result is that the weld tends to be porous and brittle.

The arc can be rendered easy to control and the absorption of atmospheric gases reduced to a minimum by 'shielding' the arc. This is done by covering the electrode with one of the various types of covering previously discussed, and as a result gases such as hydrogen and carbon dioxide are released from the covering as it melts and form an envelope around the arc and molten pool, excluding the atmosphere with its harmful effects on the weld metal. Under the heat of the arc chemical compounds in the electrode covering also react to form a slag which is liquid and lighter than the molten metal. It rises to the surface, cools and solidifies, forming a protective covering over the hot metal while cooling and protecting it from atmospheric effects, and also slows down the cooling rate of the weld. Some slags are self-removing while others have to be lightly chipped (Fig. 1.1).

The electrode covering usually melts at a higher temperature than the wire core so that it extends a little beyond the core, concentrating and directing the arc stream, making the arc stable and easier to control. The difference in controllability when using lightly covered electrodes and various medium- and heavily covered electrodes will be quickly noticed by the operator at a very early stage in practical manual metal arc welding.

With bare wire electrodes much metal is lost by volatilization, that is turning into a vapour. The use of covered electrodes reduces this loss.

An arc cannot be maintained with a voltage lower than about 14 V and is not very satisfactory above 45 V. With d.c. sources the voltage can be varied by a switch or regulator, but with a.c. supply by transformer the open circuit voltage (OCV) choice is less, being 80 or 100 V on larger units, down to 50 V on small units.

The greater the volts drop across the arc the greater the energy liberated in heat for a given current.

Arc energy is usually expressed in kilojoules per millimetre length of the weld (kJ/mm) and

$$\text{Arc energy (kJ/mm)} = \frac{\text{arc voltage} \times \text{welding current}}{\text{welding speed (mm/s)} \times 1000}.$$

The volts drop can be varied by altering the type of gas shield liberated by the electrode covering, hydrogen giving a higher volts drop than carbon dioxide for example. As the length of the arc increases so does the voltage drop, but since there is an increased resistance in this long arc the current is decreased. Long arcs are difficult to control and maintain and they lower the efficiency of the gas shield because of the greater length. As a result, absorption of oxygen and nitrogen from the atmosphere can take place, resulting in poor mechanical properties of the weld. It is essential that the welder should keep as short an arc as possible to ensure sound welds.

Transference of metal across the arc gap

When an arc is struck between the electrode and plate, the heat generated forms a molten pool in the plate and the electrode begins to melt away, the metal being transferred from the electrode to the plate. The transference takes place whether the electrode is positive or negative and also when it has a changing polarity, as when used on a.c. Similarly it is

Fig. 1.1. The shielded arc. Manual arc weld on steel base plate with a covered electrode.

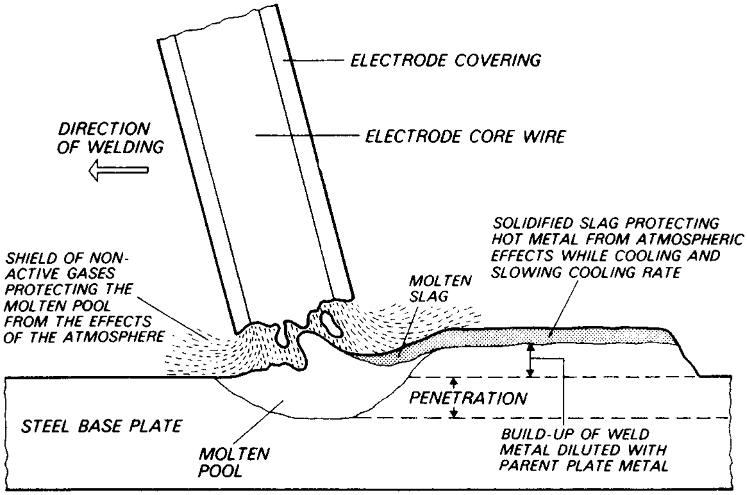
ELECTRODE COVERING
DIRECTION OF WELDING
ELECTRODE CORE WIRE
SOLIDIFIED SLAG PROTECTING HOT METAL FROM ATMOSPHERIC EFFECTS WHILE COOLING AND SLOWING COOLING RATE
SHIELD OF NON-ACTIVE GASES PROTECTING THE MOLTEN POOL FROM THE EFFECTS OF THE ATMOSPHERE
MOLTEN SLAG
PENETRATION
STEEL BASE PLATE
BUILD-UP OF WELD METAL DILUTED WITH PARENT PLATE METAL
MOLTEN POOL

transferred upwards against the action of gravity, as when making an overhead weld. Surface tension plays an important part in overhead welding and a very short arc must be held to weld in the overhead position successfully.

The forces which cause the transfer appear to be due to: (1) its own weight, (2) the electro-magnetic (Lorentz) forces, (3) gas entrainment, (4) magneto-dynamic forces producing movement and (5) surface tension. The globule is finally necked off by the magnetic pinch effect.

If the arc is observed very closely, or better still if photographs are taken of it with a slow-motion cine-camera, it can be seen that the metal is transferred from the electrode to the plate in the form of drops or globules, and these globules vary in size according to the current and type of electrode covering. Larger globules are transferred at longer intervals than smaller globules and the globules form, elongate with a neck connecting them to the electrode, the neck gets reduced in size until it breaks, and the drop is projected into the molten pool, which is agitated by the arc stream, and this helps to ensure a sound bond between weld and parent metal. Drops of water falling from a tap give an excellent idea of the method of transference (see Fig. 1.2). Other methods of transfer known as dip (short circuiting arc) and spray (free flight transfer) are discussed in the section on MIG welding process.

Arc blow

We have seen that whenever a current flows in a conductor a magnetic field is formed around the conductor. Since the arc stream is also a flow of current, it would be expected that a magnetic field would exist around it, and that this is so can be shown by bringing a magnet near the arc. It is seen that the arc is blown to one side by the magnet, due to the interaction of its field with that of the magnet (just as two wires carrying a current will attract each other if the current flows in the same direction in each, or repel if the currents are in opposite directions), and the arc may even be extinguished if the field due to the magnet is strong enough. When welding, particularly with d.c., it is sometimes found that the arc tends to wander and becomes rather uncontrollable, as though it was being blown to and fro. This is known as arc blow and is experienced most when using

Fig. 1.2. Detachment of molten globule in the metal arc process.

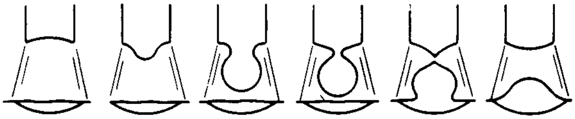

currents above 200 or below 40 A, though it may be quite troublesome, especially when welding in corners, in between this range. It is due to the interaction of the magnetic field of the arc stream with the magnetic fields set up by the currents in the metal of the work or supply cables. The best methods of correction are:

1. Weld away from the earth connexion.
2. Change the position of the earth wire on the work.
3. Wrap the welding cable a few turns around the work, if possible, on such work as girders, etc.
4. Change the position of the work on the table if working on a bench.

In most cases the blow can be corrected by experimenting on the above lines, but occasionally it can be very troublesome and difficult to eliminate. Alternating-current welding has the advantage that since the magnetic field due to the arc stream is constantly alternating in direction at the frequency of the supply, there is much less trouble with arc blow, and consequently this is very advantageous when heavy currents are being used. Arc blow can be troublesome in the TIG and MIG processes, particularly when welding with d.c.

Spatter

At the conclusion of a weld small particles or globules of metal may sometimes be observed scattered around the vicinity of the weld along its length. This is known as 'spatter' and may occur through:

1. Arc blow making the arc uncontrollable.
2. The use of too long an arc or too high an arc voltage.
3. The use of an excessive current.

The latter is the most frequent cause.

Spatter may also be caused by bubbles of gas becoming entrapped in the molten globules of metal, expanding with great violence and projecting the small drops of metal outside the arc stream, or by the magnetic pinch effect, by the magnetic fields set up, and thus the globules of metal getting projected outside the arc stream.

Spatter can be reduced by controlling the arc correctly, by varying current and voltage, and by preventing arc blow in the manner previously explained. Spatter release sprays ensure easy removal.

Eccentricity of the core wire in an MMA welding electrode

If the core wire of a flux-coated electrode is displaced excessively from the centre of the flux coating because of errors in manufacture, the arc may not function satisfactorily. The arc tends to be directed towards one side as if influenced by 'arc blow' and accurate placing of the deposited metal is prevented (Fig. 1.3*a*). A workshop test to establish whether the core wire is displaced outside the manufacturer's tolerance is to clean off the flux covering on one side at varying points down the length of the electrode and measure the distance L (Fig. 1.3*b*). The difference between the maximum and minimum reading is an approximate indication of the eccentricity.

Electrode efficiency. The efficiency of an electrode is the mass of metal actually deposited compared with the mass of that portion of the electrode consumed. It can be expressed as a percentage thus:

$$\text{efficiency } \% = \frac{\text{mass of metal deposited}}{\text{mass of metal of the electrode consumed}} \times 100.$$

With ordinary electrodes the efficiency varies from 75% to 95% but with electrodes containing metallic components in the covering the efficiency can approach 200% (e.g. electrodes containing iron powder).

In the electrode classification (British), efficiencies of 110% and above are indicated by a three-digit figure in the additional section of the electrode coding, giving the efficiency rounded to the nearest 10, with values ending in 5 being rounded up (BS 639 (1986)).

The efficiency of a particular type of electrode can be obtained by taking a clean steel plate (up to 0.25% carbon) and weighing it, say m_1 g. Electrodes of a mass of 20% m_1 g are completely stripped of their covering and weighed, say m_3 g (weighing tolerances are ± 1 g). A similar number of electrodes to those stripped are then deposited on the plate with an interpass temperature not greater than 100 °C, until each stub length is 50 mm. Slag and spatter are cleaned from each run when deposited and the plate completely cleaned after the final deposit. It is then weighed, say

Fig. 1.3. (*a*)

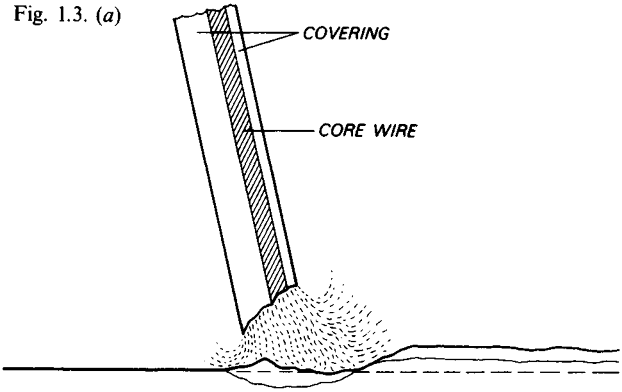

m_2 g, so that the increase in mass of the plate (i.e. the mass of the deposited metal) is $(m_2 - m_1)$ g.

The stubs have any covering remaining on them removed and are weighed, say m_4 g; thus the mass of metal to be deposited is $(m_3 - m_4)$ g.

The nominal (N) electrode efficiency R_N % is given by:

$$R_N = \frac{(m_2 - m_1)}{(m_3 - m_4)} \times 100$$

$$\text{or} \quad \frac{\text{mass of deposited metal}}{\text{mass of core wire before depositing} - \text{mass of stub wire}}$$

Electrodes should be tested according to the maker's instructions regarding d.c. or a.c. and polarity. Full details for obtaining the nominal electrode efficiency are given in BS 639 (1986), appendix C.

High efficiency electrodes

The deposition rate of a given electrode is dependent upon the welding current used, and for maximum deposition rate, maximum current should be used. This maximum current depends upon the diameter of the core wire, and for any given diameter of wire there is a maximum current beyond which increasing current will eventually get the wire red hot and cause overheating and hence deterioration of the covering.

To enable higher currents to be used an electrode of larger diameter core wire must be used, but if metallic components such as iron powder are added to the covering of the electrode, this covering becomes conducting, and a higher welding current can now be used on an electrode of given core wire diameter. The deposition rate is now increased and in addition the iron powder content is added to the weld metal, giving greater

Fig. 1.3. (*b*) Eccentric core wire.

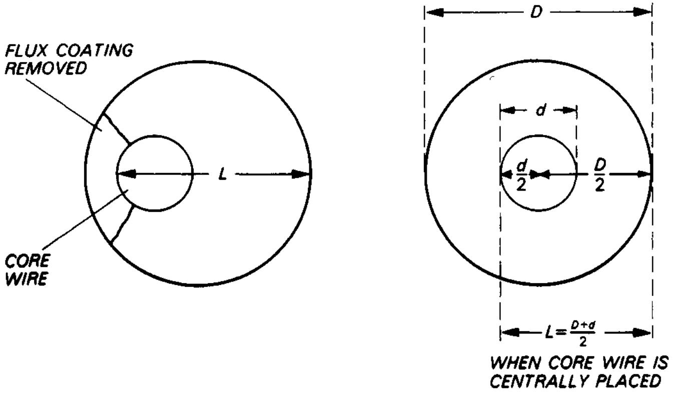

efficiency, that is enabling more than the core wire weight of metal to be deposited because of the extra iron powder. Efficiencies of up to 200 % are possible, this meaning that twice the core wire weight of weld metal is being deposited. These electrodes can have coverings of rutile or basic type or a mixture of these. The iron powder ionizes easily, giving a smoother arc with little spatter, and the cup which forms as the core wire burns somewhat more quickly than the covering gives the arc directional properties and reduces loss due to metal volatilization. See also Electrode efficiency (metal recovery and deposition coefficient).

Hydrogen-controlled electrodes (basic covered)*

If oxygen is present with molten iron or steel a chemical reaction occurs and the iron combines chemically with the oxygen to form a chemical compound, iron oxide. Similarly with nitrogen, iron nitride being formed if the temperature is high enough as in metal arc welding. When hydrogen is present however there is no chemical reaction and the hydrogen simply goes into solution in the steel, its presence being described as x millilitres of hydrogen in y grams of weld metal.

This hydrogen can diffuse out of the iron lattice when in the solid state resulting in a lowering of the mechanical properties of the weld and increasing the tendency to cracking. By the use of basic hydrogen-controlled electrodes, and by keeping the electrodes very dry, the absorption of hydrogen by the weld metal is reduced to a minimum and welds can be produced that have great resistance to cracking even under conditions of very severe restraint.

The coverings of these electrodes are of calcium or other basic carbonates and fluorspar bonded with sodium or potassium silicate. When the basic carbonate is heated carbon dioxide is given off and provides the shield of protective gas thus:

calcium carbonate (limestone) heated → calcium oxide (quicklime) + carbon dioxide.

There is no hydrogen in the chemicals of the covering, so that if they are kept absolutely dry, the deposited weld metal will have a low hydrogen content. Electrodes which will give deposited metal having a maximum of 15 millilitres of hydrogen per 100 grams of deposited metal (15 ml/100 g) are indicated by the letter H in BS 639 (1986) classification. The absence of diffusible hydrogen enables free cutting steels to be welded with absence of porosity and cracking and the electrodes are particularly suitable for welding in all conditions of very severe restraint. They can be used on a.c. or d.c. supply according to the makers' instructions and are available also in iron powder form and for welding in all positions. Low and medium alloy steels which normally would require considerable pre-heat if welded with rutile-coated electrodes can be welded with very much less pre-heating, the welds resisting cracking under severe restraint conditions and also being very suitable for welding in sub-zero temperature conditions. By correct storage and drying of these electrodes the hydrogen content can be reduced to 5 ml/100 g of weld metal for special applications. Details of these drying methods are given in the section on storage and drying of electrodes (q.v.).

---

* Typical AWS classification of these electrodes may be E 7015 or E 7018, for example.

Experiment to illustrate the diffusible hydrogen content in weld metal. Make a run of weld metal about 80 mm long with the metal arc on a small square of steel plate using an ordinary steel welding rod with a cellulose, rutile or iron oxide coating. Deslag, cool out quickly and dry off with a cloth and place the steel plate in a beaker or glass jar of paraffin. It will be noted that minute bubbles of gas stream out of the weld metal and continue to do so even after some considerable time. If this gas is collected as shown in Fig. 1.4 it is found to be hydrogen which has come from the flux covering and the moisture it contains. A steel weld may contain hydrogen dissolved in the weld metal and also in the molecular form in any small voids which may be present. Hydrogen in steel produces embrittlement and a reduction in fatigue strength. If a run of one of these hydrogen-controlled electrodes is made on a test plate and the previous experiment repeated it will be noted that no hydrogen diffuses out of the weld.

Fig. 1.4. Collecting diffusible hydrogen from a mild steel weld.

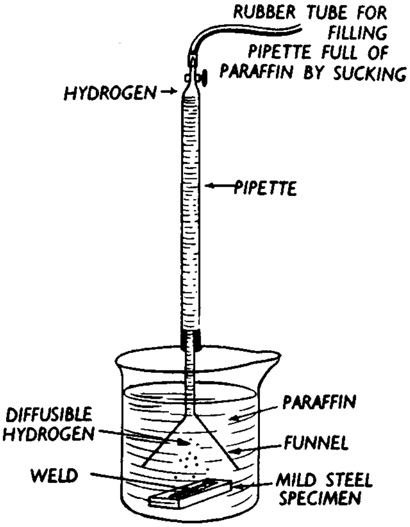

BS 6693, Pt 1 (3-day collection) and Pt 2 (collection continued until there is no further increase of hydrogen) give methods of determining the quantity of diffusible hydrogen present in a specimen, together with drawings of apparatus for collecting this hydrogen.

Deep penetration electrodes

A deep penetration electrode is defined in BS 499, Part 1 (*Welding terms and symbols*) as 'A covered electrode in which the covering aids the production of a penetrating arc to give a deeper than normal fusion in the root of a joint'. For butt joints with a gap not exceeding 0.25 mm the penetration should be not less than half the plate thickness, the plate being twice the electrode core thickness. For fillet welds the gap at the joint should not exceed 0.25 mm and penetration beyond the root should be 4 mm minimum when using a 4 mm diameter electrode.

Welding position

Weld slope is the angle between line of the root of the weld and the horizontal (Fig. 1.5).

Weld rotation. Draw a line from the foot of the weld at right angles to the line welding to bisect the weld profile. The angle that this line makes with the vertical is the angle of weld rotation.

The table indicates the five welding positions used for electrode classification. Any intermediate position not specified may be referred to as 'inclined'. (See also Figs. 1.6, 1.7, 1.8.)

Position	Slope	Rotation	Symbol	Fig.
Flat	0–5°	0–10°	F	1.5
Horizontal–vertical	0–5°	30–90°	H	1.6
Vertical–up	80–90°	0–180°	V	1.7a
Vertical–down	80–90°	0–180°	D	1.7b
Overhead	0–15°	115–180°	O	1.8

Storage of electrodes

The flux coverings on modern electrodes are somewhat porous and absorb moisture to a certain extent. The moisture content (or humidity) of the atmosphere is continually varying and hence the moisture content of the covering will be varying. Moisture could be excluded by providing a non-porous covering, but any moisture entrapped would be liable to cause rupture of the coating when the moisture was turned to steam by the heating effect of the passage of the current through the electrode. Cellulosic electrodes absorb quite an appreciable amount of moisture, and it does not affect their properties since they function quite well with a moisture content. They should not be over-dried or the organic compounds of which

Fig. 1.5. Flat position.

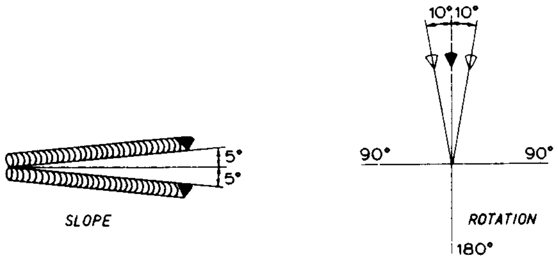
5°
5°
SLOPE
10° 10°
90°
90°
ROTATION
180°

Fig. 1.6. Horizontal-vertical position.

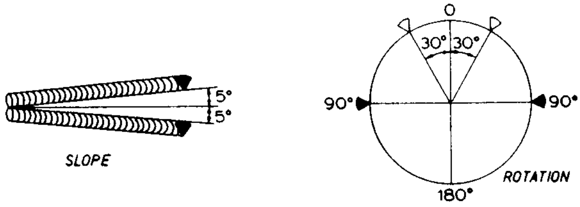
5°
5°
SLOPE
0
30° 30°
90°
90°
ROTATION
180°

Fig. 1.7. (*a*) Vertical-up position. (*b*) Vertical-down position.

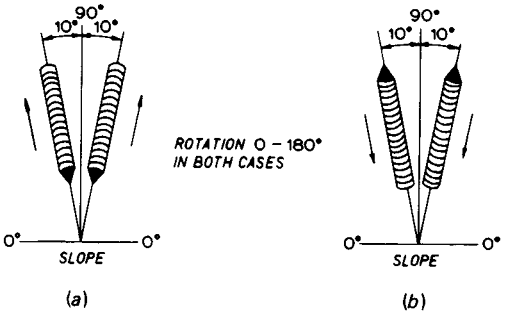
90°
10° 10°
0°
0°
SLOPE
(a)
ROTATION 0 – 180°
IN BOTH CASES
90°
10° 10°
0°
0°
SLOPE
(b)

Fig. 1.8. Overhead position.

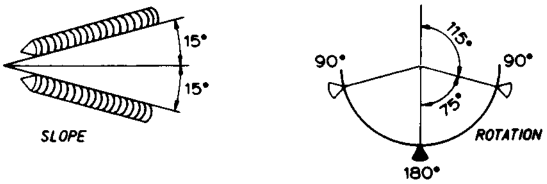
15°
15°
SLOPE
115°
90°
90°
75°
ROTATION
180°

they are composed tend to char, affecting the voltage and arc properties. The extruded electrodes with rutile, iron oxide and silicate coatings do not pick up so much moisture from the atmosphere and function quite well with a small absorbed content. If they get damp they can be satisfactorily dried out, but it should be noted that if they get excessively wet, rusting of the core wire may occur and the coating may break away. In this case the electrodes should be discarded.

Storage temperatures should be about 12° C above that of external air temperature with 0–60% humidity. Cellulose covered electrodes are not so critical: but they should be protected against condensation and stored in a humidity of 0–90%.

Drying of electrodes. The best drying conditions are when the electrodes are removed from their package and well spaced out in a drying oven which has a good circulation of air. Longer drying times are required if the electrodes are not spaced out. The following table gives an indication of temperatures and times required, but see also the special conditions for drying basic electrodes (Fig. 1.9).

Drying of electrodes: approximate times and temperatures with electrodes spaced apart. Times will vary with air circulation, electrode spacing and oven loading

Electrode type	Diameter mm	Temperature °C	Time in mins air circulation good	Time in mins air circulation poor
Rutile mild steel	1.6–2.5	110	10–30	20–30
	3.2–5.0	110	20–45	30–60
	6.0–10.0	110	30–60	45–120
Cellulose	2.5–6.0	110	10–15	15–20

Hydrogen-controlled (basic) electrodes

The coatings of these electrodes contain no hydrogen-forming compounds, but if moisture is absorbed by the coating it becomes a source of hydrogen and cannot be tolerated. They must therefore be stored in a dry, heated and well-ventilated store on racks above floor level and unused electrodes should be returned to the store rather than left in the colder and moister conditions of the workshop where they could absorb moisture. A temperature of about 12°C above that of the external air temperature is suitable. Before use they should be removed from their package and spread out in the drying oven, the drying time and temperature depending upon the permissible volume of hydrogen in the weld deposit. Suggested figures are given in the following table.

Hydrogen content in millilitres of hydrogen per 100 grams of weld metal	Temperature °C	Time minutes	Use
10–15 ml H_2/100 g	150	60	To give resistance to HAZ cracking in thick sections of mild steel, high restraint.
5–10 ml H_2/100 g	200	60	High quality welds in pressure vessel and structural applications.
below 5 ml H_2/100 g	450	60	Thick sections to avoid lamellar tearing and critical applications.

In order to obtain high radiographic standards of deposited weld metal the drying periods given above may be extended. The following periods are given as an indication of prolonged drying times such that the electrode coating will not suffer a decrease in coating strength.

Drying temperature	Maximum time
150°C	72 hours
250°C	12 hours
450°C	2 hours

The makers' instructions for drying should be strictly adhered to.

Fig. 1.9

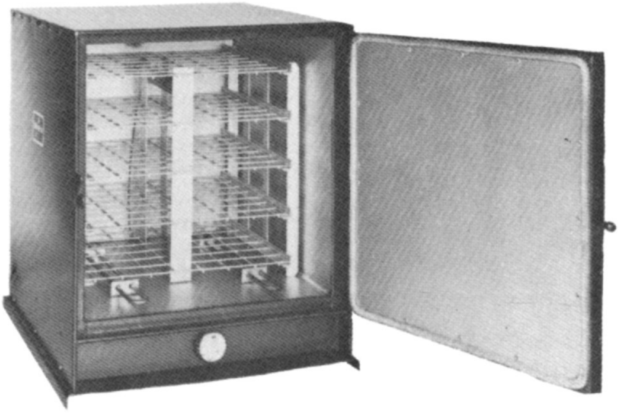

Many electrodes if stored in damp situations get a white fur on their coverings. This is sodium carbonate produced by the action of the carbon dioxide (carbonic acid) of the atmosphere on the sodium silicate of the binder in the flux covering. The fur appears to have little detrimental effect on the weld but shows that the electrodes are being stored in too damp a situation.

Electrode classification (British)

Abridged classification for covered carbon and carbon–manganese steel electrodes for manual metal arc welding. BS 639 (1986)

Note: This classification is for deposited weld metal having a tensile strength not greater than 650 N/mm^2. Students should study the whole text of BS 639 (1986), which gives full details of the classification together with the tests involved. Weld metals with tensile strength greater than 650 N/mm^2 are dealt with in BS 2493 and BS 2926.

The classification is denoted by a code consisting of two parts: (*a*) a general code, followed by (*b*) an additional code in parentheses, for example E 43 2 2 RR (2 1).

(*a*) *General code* (strength, toughness and covering (STC) code). There are five elements in the general code (in the order given)

(1) the letter E indicating a covered electrode for manual metal arc welding
(2) two digits indicating the strength (tensile, yield and elongation properties of the weld metal) (see Table 1)
(3) one digit indicating the temperature for a minimum average impact value of 28 J (see Table 2)
(4) one digit indicating the temperature for a minimum average impact value of 47 J (see Table 3)
(5) either one or two letters indicating the type of covering, namely: B – basic; BB – basic, high efficiency; C – cellulosic; R – rutile; RR – rutile heavy coated; S – other types.

(*b*) *Additional code.* This has four elements, of which (1) and (4) are included only if appropriate.

(1) where appropriate, three digits indicating the nominal electrode efficiency (see p. 6) included only if this is equal to, or greater than 110, the figures being rounded off to the nearest multiple of 10, those of 5 and upwards being rounded up.

(2) a digit indicating the recommended welding positions for the electrode:
1 – all positions,
2 – all positions except vertical/down,
3 – flat and for fillet welds, horizontal/vertical,
4 – flat,
5 – flat, vertical/down and, for fillet welds, horizontal/vertical,
9 – any other position or combination of positions not classified above.

(3) a digit indicating the power requirements (Table 4)

(4) a letter H, where appropriate, indicating a hydrogen controlled electrode (see below).

Hydrogen-controlled electrodes. The coding is followed by the letter H if the electrode is hydrogen controlled. For this coding the electrodes deposit not more than 15 ml of diffusible hydrogen per 100 g of deposited weld metal when determined in accordance with the method given in BS 6693, Part 2. The manufacturer shall provide information on the recommended drying conditions to obtain hydrogen levels in the following ranges:

not exceeding 15 ml }
not exceeding 10 ml } per 100 g of deposited weld metal.
not exceeding 5 ml }

Examples of the use of BS 639 classification

Note: the tables referred to in Examples 1–3 are on pp. 17 and 18.

Example 1

Complete classification of electrode considered E 43 2 2 RR (2 1)

	STC code
E	Arc welding electrode
43	Strength 430–550 N/mm^2 (Table 1)
2	Temperature for minimum average impact strength of 28 J, 0 °C (Table 2)
2	Temperature for minimum average impact strength of 47 J, 0 °C (Table 3)
RR	Covering, rutile heavy coated (see list of coverings)
	Additional code
2	Welding position: all positions except vertical/down (see list of welding positions)
1	Welding current and voltage: a.c., 50 V; d.c. electrode +/− (Table 4)

Example 2

Complete classification of electrode considered E 51 5 4 B (120 1 0 H)

STC code

E	Arc welding electrode
51	Strength 510–650 N/mm^2 (Table 1)
5	Temperature for minimum average impact strength of 28 J, −40 °C (Table 2)
4	Temperature for minimum average impact strength of 47 J, −30 °C (Table 3)
R	Covering, basic (see list of coverings)

Additional code

120	Efficiency
1	Welding position: all positions (see list of welding positions)
0	Welding current and voltage: polarity d.c. as recommended by makers Not suitable for use on a.c. (Table 4)
H	Hydrogen controlled

Example 3

Complete classification of electrode considered E 51 5 3 BB (160 3 6 H)

STC code

E	Arc welding electrode
51	Strength 510–650 N/mm^2 (Table 1)
5	Temperature for minimum average impact strength of 28 J, −40 °C (Table 2)
3	Temperature for minimum average impact strength of 47 J, −20 °C (Table 3)
BB	Covering, basic, high efficiency (see list of coverings)

Additional code

160	Efficiency
3	Welding position: flat and for fillets, horizontal/vertical (see list of welding positions)
6	Welding current and voltage: a.c., 70 V; d.c. electrode + (Table 4)
H	Hydrogen controlled

Table 1. *Tensile properties*

Electrode designation	Tensile strength N/mm^2	Minimum yield stress N/mm^2	Minimum elongation %		
			When digit of Table 2 is 0 or 1	When digit of Table 2 is 2	When digit of Table 2 is 3, 4, or 5
E 43	430–550	330	20	22	24
E 51	510–650	360	18	18	20

Table 2. *First digit for impact value*

Digit (in position in the classification)	Temperature for minimum average impact value of 28 J using 4 mm diameter electrodes only, °C
E -- 0 - -	not specified
E -- 1 - -	+20
E -- 2 - -	0
E -- 3 - -	−20
E -- 4 - -	−30
E -- 5 - -	−40

Table 3. *Second digit for impact value*

Digit (in position in the classification)	Temperature for minimum average impact value of 47 J using 4 mm diameter electrodes and the largest diameter electrodes submitted for classification, °C
E – – – 0 –	not specified
E – – – 1 –	+20
E – – – 2 –	0
E – – – 3 –	−20
E – – – 4 –	−30
E – – – 5 –	−40
E – – – 6 –	−50
E – – – 7 –	−60
E – – – 8 –	−70

Table 4. *Welding current and voltage conditions*

Digit	Direct current recommended electrode polarity	Alternating current minimum open circuit voltage, V
0	polarity as recommended by the manufacturer	not suitable for use on a.c.
1	+ or −	50
2	−	50
3	+	50
4	+ or −	70
5	−	70
6	+	70
7	+ or −	80
8	−	80
9	+	80

Table 5. *Types of electrode flux coverings*

Class	Composition of covering	Characteristics
C cellulosic	Organic materials containing cellulose $(C_6H_{10}O_5)_n$	Voluminous gas shield; good penetration; fast welding speeds; easily removable slag. Suitable for steel welding a.c. or d.c. electrode +ve. Used for first (stringer) run in downhill pipe-line welding. High level of hydrogen.
B basic	Calcium carbonate ($CaCO_3$), calcium fluoride (CaF_2) and other basic carbonates	The calcium carbonate decomposes in the arc heat to release carbon dioxide which gives the gas shield. The calcium oxide (CaO) combines with the calcium fluoride to form a basic slag having a low melting point. The low hydrogen content of the weld metal results in a weld that is resistant to solidification cracking and to a high sulphur content in the steel. Because there are no organic or hydrated materials in the covering the electrodes can be baked at high temperature giving a low level of hydrogen in the weld metal and reducing the danger of cold cracking, particularly in highly restrained joints and thick sections. Because of the relatively small gas shield, a short arc should be used and the electrodes are suitable for a.c. or d.c. electrode +ve. They should be stored under warm dry conditions and preferably baked before use.
BB basic high efficiency	Similar to basic electrode covering but have additional metallic materials (e.g. iron powder) in the covering which raise the efficiency to 130% and more	These electrodes are suitable for welding in the flat and horizontal/vertical position with a greatly increased rate of metal deposition. Their high efficiency covering makes them unsuitable for welding in the vertical and overhead positions. They can be used either a.c. or d.c., generally with electrode +ve. Efficiency is indicated by a three-figure digit beginning the additional coding.

Class	Composition of covering	Characteristics
R rutile	Titanium dioxide (rutile) and other hydrated minerals and/or organic cellulose materials	Easy to use, with a smooth weld finish and medium penetration. High level of hydrogen in the weld metal limits their use in thick sections or restrained joints. Suitable for a.c. or d.c.; the fast freezing of weld metal and fluid slag makes them suitable for vertical and overhead welding.
RR rutile heavy coating	Similar covering to the previous rutile electrode but containing, in addition, metallic substances (e.g. iron powder), which raise the efficiency to 130% or more.	Similar characteristics to rutile electrodes but generally unsuitable for vertical and overhead welding because of increased slag. Increased rate of metal deposition. Efficiency is indicated by a three-figure digit beginning the additional coding.
S other types		This class includes electrodes that do not fall into any of the previous classes. They range from little used types, such as those with acid covering (containing oxides and carbonates of lime and manganese with deoxidizers such as ferromanganese), to the oxide types (containing iron oxide with or without manganese oxide and silicates). This class also includes any newly developed flux systems. Manufacturer's advice should be followed when using them.

Table 6. *AWS electrode classification*

AWS classification	Type of covering	Capable of producing satisfactory welds in position shown[a]	Type of current[b]
E60 series electrodes			
E6010	High cellulose sodium	F, V, OH, H	DCEP
E6011	High cellulose potassium	F, V, OH, H	AC or DCEP
E6012	High titania sodium	F, V, OH, H	AC or DCEN
E6013	High titania potassium	F, V, OH, H	AC or DC, either polarity
E6020	High iron oxide	H-fillets	AC or DCEN
E6022[c]	High iron oxide	F	AC or DC, either polarity
E6027	High iron oxide, iron powder	H-fillets, F	AC or DCEN
E70 series electrodes			
E7014	Iron powder, titania	F, V, OH, H	AC or DC, either polarity
E7015	Low hydrogen sodium	F, V, OH, H	DCEP
E7016	Low hydrogen potassium	F, V, OH, H	AC or DCEP
E7018	Low hydrogen potassium, iron powder	F, V, OH, H	AC or DCEP
E7024	Iron powder, titania	H-fillets, F	AC or DC, either polarity
E7027	High iron oxide, iron powder	H-fillets, F	AC or DCEN
E7028	Low hydrogen potassium, iron powder	H-fillets, F	AC or DCEP
E7048	Low hydrogen potassium, iron powder	F, OH, H, V-down	AC or DCEP

Note the American use of AC and DC for alternating current and direct current.

[a] F = flat, H = horizontal, H-fillets = horizontal fillets, V-down = vertical down.

$\left.\begin{array}{c}\text{V = vertical}\\ \text{OH = overhead}\end{array}\right\}$ for electrodes 3/16 in and under, except 5/32 in and under for classifications E7014, E7015, E7016, E7018.

[b] DCEP direct current electrode positive (DC reverse polarity). DCEN direct current electrode negative (DC straight polarity).

[c] E6022 classification electrodes are for single-pass welds.

Electrode classification (American)

American Welding Society (AWS) electrode classification: abridged specification for covered carbon steel arc welding electrodes ANSI*/AWS A5 1 81 (Reprint 1984)

In the AWS classification a four-digit number is used, preceded by the letter E (indicating a covered arc welding electrode).

The first two digits indicate the minimum tensile strength of the weld metal deposited in thousands of pounds force per square inch (1000 psi) (Table 6). For example, E70XX could represent an electrode giving weld metal having a minimum tensile strength of 72 000 psi or 72 ksi (500 MPa or 500 N/mm^2). Details of chemical composition, mechanical usability and soundness (radiographic) test, all-weld-metal tension test and impact (Charpy) test are given in the specification.

The last two digits indicate the characteristics of the covering (see Table 7).

Examples of AWS classification

Example 1

E6013: Electrode with weld metal tensile strength 60 ksi or 412 N/mm^2 high rutile covering bonded with potassium silicate AC or DC either polarity.

Table 7. *Characteristics of coverings*

Exx10	High cellulose, bonded with sodium silicate. Deeply penetrating, forceful spray-type arc. Thin friable slag.
Exx11	Similar to Exx10 but bonded with potassium silicate to allow use on AC or DC; covering bonded with sodium silicate.
Exx12	High rutile. Quiet arc, medium penetration.
Exx13	Similar to Exx12 but bonded with potassium silicate and other easily ionized materials. Fluid slag, easily removed.
Exx14	Similar to Exx12 and Exx13 types with the addition of iron powder.
Exx15	Basic low hydrogen type bonded with sodium silicate. For high tensile steels.
Exx16	Similar to Exx15 but bonded with potassium silicate.
Exx18	Covering similar to Exx15 and Exx16 but with addition of iron powder.
Exx20	High iron oxide coating bonded with sodium silicate.
Exx24	Similar to Exx12 and Exx13. Heavily coated plus iron powder.
Exx27	Flux ingredients similar to Exx20 with the addition of iron powder.
Exx28	Similar to Exx18 but with heavier covering. Low hydrogen, potassium silicate as binder. Flat and horizontal fillets.
Exx48	Similar to Exx28, low hydrogen but suitable for most positions.

---

* American National Standards Institute.

Example 2

E7018: Electrode with weld metal of tensile strength 70 ksi, with basic covering, iron powder, low hydrogen. Bonded with potassium silicate F, V, OH, H, AC or DCEP.

Core wire for all electrodes in this specification is usually of composition 0.10% carbon, 0.45% manganese, 0.03% sulphur, 0.02% phosphorus and 0.01% silicon.

Welders' accessories (Fig. 1.10*a*)

Electrode holder

This is an arrangement which enables the welder to hold the electrode when welding. It has an insulated body and head which reduces the danger of electric shock when working in damp situations and also reduces the amount of heat conducted to the hand whilst welding. The electrode is clamped between copper jaws which are usually spring loaded, and a simple movement of a side lever enables the electrode to be changed easily and quickly. In another type the electrode end fits into a socket in the holder head and is held there by a twist on the handle. The welding flexible cable is attached to the holder by clamping pieces or it may be sweated into a terminal lug. The holder should be of light yet robust construction and well insulated, and electrode changing must be a simple operation. Fig. 1.10*b* shows typical holders.

Head shields, lenses and general protective gear. The rays from the metallic arc are rich in infra-red and ultra-violet radiation, and it is essential that the eyes and face of the welder should be protected from these rays and from the intense brightness of the arc. The welding shield can either fit on to the head (Fig. 1.11*a*), leaving both hands free, or may be carried in one hand. It should extend so as to cover the sides of the face, especially when welding is done in the vicinity of other welders, so as to prevent stray flashes reaching the eyes. The shield must be light and, thus, preferably made of fibre.

Helmets and hand shields are now available, weighing little more than ordinary shields, that give maximum comfort when welding in restricted conditions. They are similar in appearance to the standard shield but have double-wall construction, the inner face being perforated with small holes, through which pure air is supplied to the interior of the helmet. Head and face are kept cool and the operator's eyes and lungs are protected from dust and fumes as well as radiation and spatter as with ordinary shields. The air is supplied from any standard compressor or air line, the supply tube being fitted with a pressure-reducing valve to give a pressure of 1.7 bar (Fig. 1.11*b*).

The arc emits infra-red and ultra-violet radiation in addition to light in the visible spectrum. The filter or lens is designed to protect the eyes from the ultra-violet and infra-red radiation which would injure the eyes and also to reduce the amount of visible light so that there is no discomfort. The filters are graded by numbers, followed by the letters EW (electrical welding) denoting the process, according to BS 679, and increase in shade depth with increasing number of increasing currents. The filters recommended are: up to 100 A, 8 or 9/EW; 100–300 A, 10 or 11/EW; over 300 A,

Fig. 1.10. (*a*)

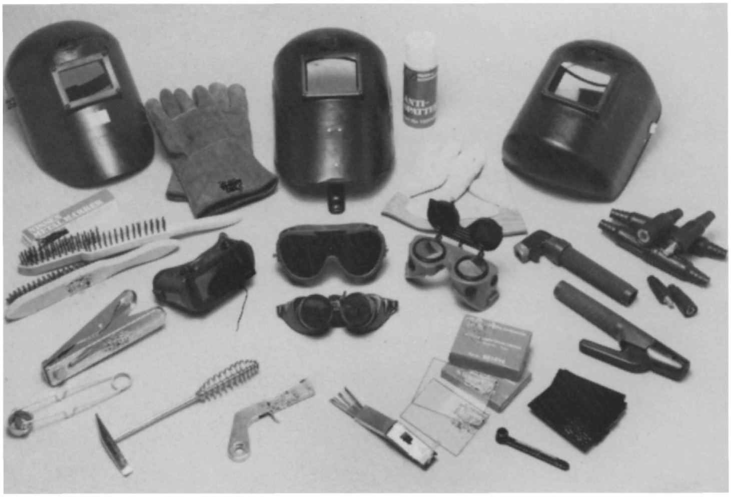

Fig. 1.10. (*b*)

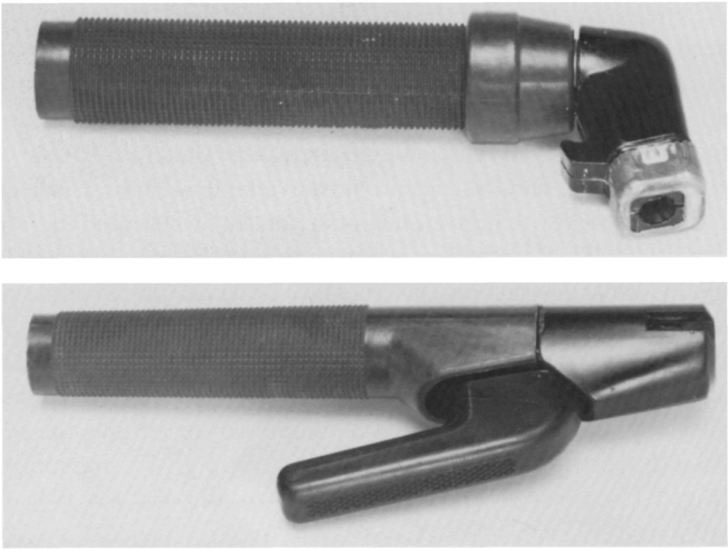

Fig. 1.11. (*a*)

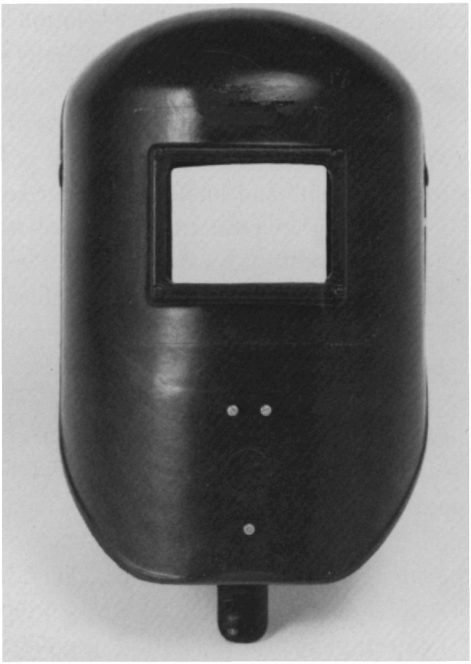

Fig. 1.11. (*b*)

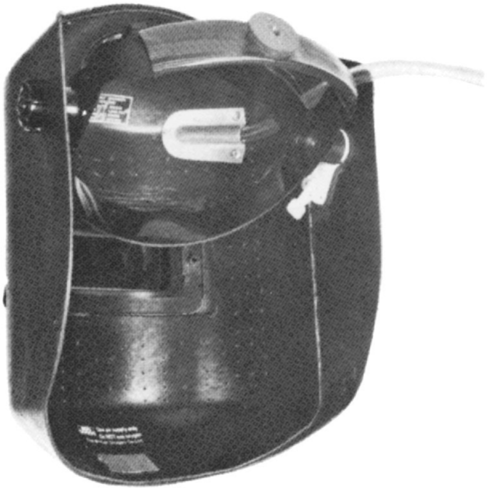

12, 13, 14/EW.* The choice of the correct filter is the safeguard against eye damage. Occasional accidental exposure to direct or reflected rays may result in the condition known as arc flash. This causes no permanent damage but is painful, with a feeling of sand in the eyes accompanied by watering. Bathing the eyes with eye lotion and wearing dark glasses reduces the discomfort and the condition usually passes with no adverse effects in from 12 to 24 hours. If it persists a doctor should be consulted.

---

* Many headshields are now fitted with light-reactant lenses, leaving both hands free. The normal view that the welder has before welding is quite clear. Upon the arc being struck, the lens darkens to the shade previously selected, the time lag being of the order of 0.001–0.0015 sec.

A lens of this type is expensive and is protected from metallic spatter on *both* sides by plain glass, which should be renewed from time to time, as it becomes opaque and uncleanable, due to spatter and fumes.

The welding area must be adequately screened so that other personnel are not affected by rays from the arc.

The eyes must also be protected against foreign bodies such as particles of dust and metal, which are all potential dangers in the welding shop.

Spectacles with clear or tinted lenses are available and also fit over normal spectacles if required. For use particularly on grinding operations there are clear goggles and there are also face shields with plain or tinted visors which fit on the head like a welding head mask but have the shield hinged so that it is easily raised when required. For conditions where there is much noise there are hearing protectors which fit snugly over the ears in the same way as radio headphones. There are also ear plugs which are made in a soft foam and which are easily inserted, and are very effective.

Protection against inhalation of polluted atmosphere is given by single cartridge dust respirators and various types of moulded plastic masks with adjustable headbands in which the filter is replaceable. A protective cap is also available with a hard shell, foam lining and elastic grip giving good fitting to the head.

Leather or skin aprons are excellent protection for the clothes against sparks and molten metal. Trouser clips are worn to prevent molten metal lodging in the turn-ups, and great care should be taken that no metal can drop inside the shoe, as very bad burns can result before the metal can be removed. Leather spats are worn to prevent this. Gauntlet gloves are worn for the same reason, especially in vertical and overhead welding. In welding in confined spaces, the welder should be fully protected, so that his clothes cannot take fire due to molten metal falling on him, otherwise he may be badly burnt before he can be extracted from the confined space.

The welding bench in the welding shop should have a flat top of steel plate, about 1.5 m × 0.75 m being a handy size. On one end a vice should be fitted, while a small insulated hook on which to hang the electrode holder when not in use is very handy.

Jigs and fixtures

These are a great aid to the rapid setting up of parts and holding them in position for welding. In the case of repetition work they are essential equipment for economical working. Any device used in this way comes under this heading, and jigs and fixtures of all types can be built easily, quickly and economically by arc welding. They are of convenience to the welder, reduce the cost of the operation, and standardize and increase the accuracy of fabrication.

Jigs may be regarded as specialized devices which enable the parts being welded to be easily and rapidly set up, held and positioned. They should be rigid and strong since they have to stand contractional stresses without deforming unless this is required; simple to operate, yet they must be accurate. Their design must be such that it is not possible to put the work in them the wrong way, and any parts of them which have to stand continual wear should be faced with wear-resistant material. In some cases, as in inert gas welding (q.v.), the jig is used as a means of directing the inert gas on to the underside of the weld (backpurge) and jigs may also incorporate a backing strip.

Fixtures are of a more general character and not so specialized as jigs. They may include rollers, clamps, wedges, etc., used for convenience in manipulation of the work. Universal fixtures are now available, and these greatly reduce the amount of time of handling of the parts to be welded and can be adapted to suit most types of work.

Manipulators, positioners, columns and booms

Positioners are appliances which enable work to be moved easily, quickly and accurately into any desired position for welding – generally in the downhand position since this speeds up production by making welding easier, and is safer than crane handling. Universal positioners are mounted on trunions or rockers and can vary in size from quite small bench types to very large ones with a capacity of several tons. Manually operated types are generally operated through a worm gearing with safety locks to prevent undesired movement after positioning. The larger types are motor-driven, and on the types fitted with a table, for example, the work can be swung through any angle, rotated, and moved up and down so that if required it can be positioned under a welding head.

As automatic welding has become more and more important so has the design of positioners and rollers improved. Welding-columns and booms may be fixed or wheel mounted and have the automatic welding head mounted on a horizontal boom which can slide up and down a vertical column. The column can be swivel mounted to rotate through 360° and can be locked in any position. In the positioning ram-type boom there is horizontal and vertical movement of the boom carrying the welding head and they are used for positioning the head over work which moves beneath them at welding speed. In the manipulating ram-type boom, the boom is

Fig. 1.12. (*a*) Head and tailstock manipulator.

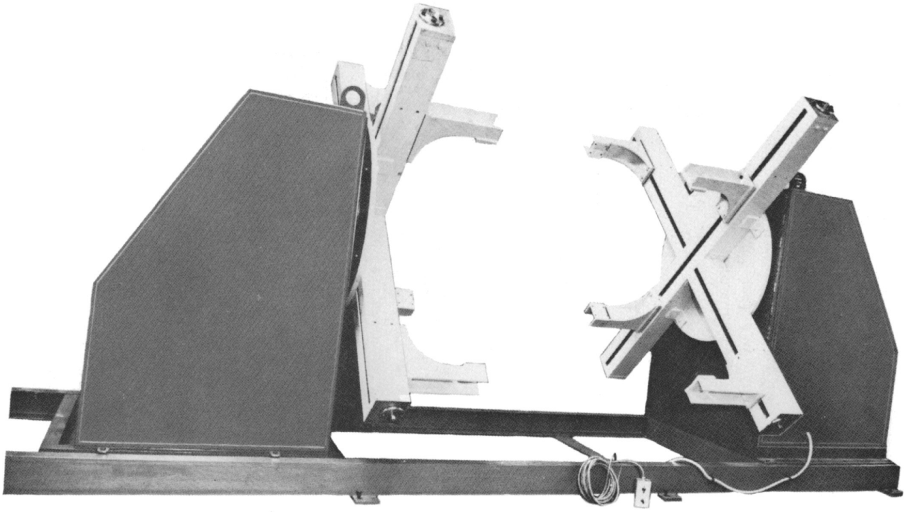

Fig. 1.12. (*b*) Power rotation, power tilt through 135°.

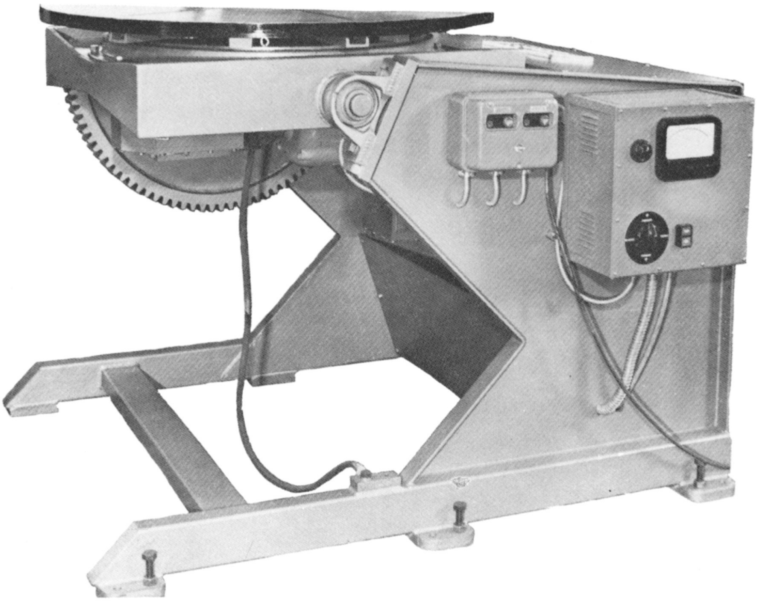

Fig. 1.12. Roller bed, travelling type, motorized.

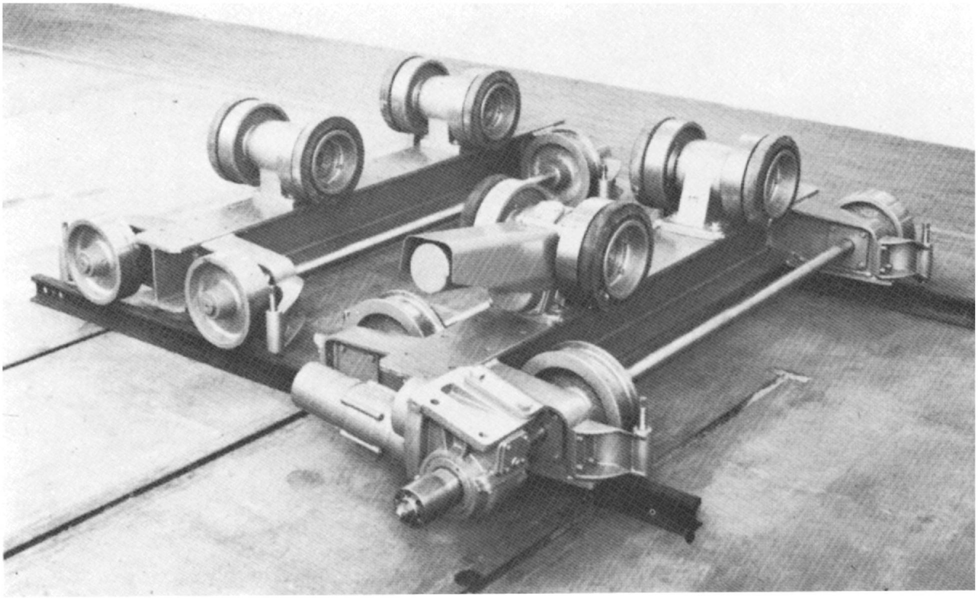

Fig. 1.12. (*d*) Power rotation, manual tilt.

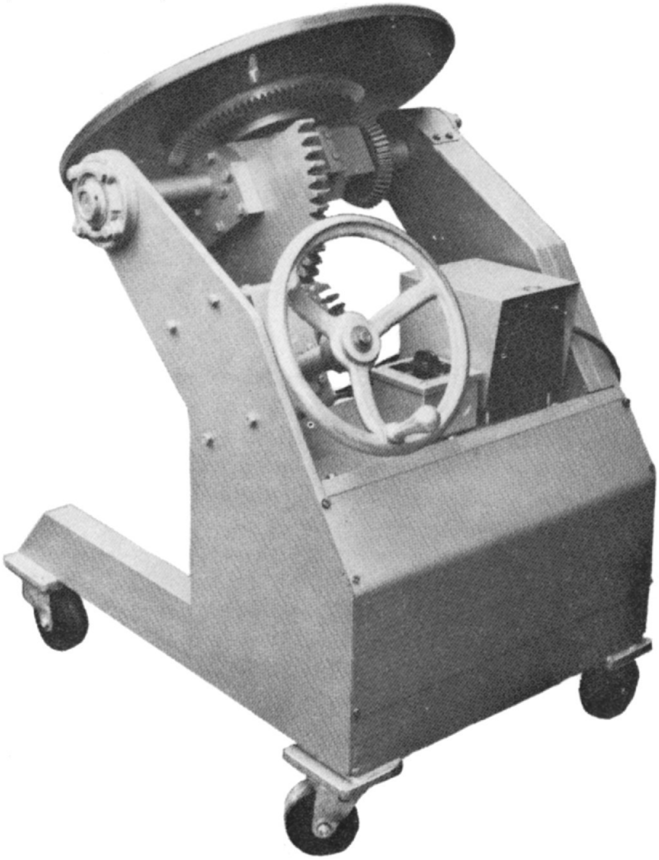

provided with a variable speed drive of range of about 150–1500 mm per min enabling the boom to move the welding head over the stationary work. Both types of boom in the larger sizes can be equipped with a platform to carry the operator, who can control all movements of head and boom from this position in Fig. 1.12. Various types are shown.

*The practice of manual metal arc welding**

Electrode lengths. The actual length must be within ± 2 mm of the nominal value

Diameter, mm	Nominal length, mm
1.6	200
	250
2.0	200
	250
	300
	350
2.5	250
	300
	350
3.2, 4.0 or 5.0	350
	450
6.0 to 8.0	350
	450
	500
	600
	700
	900

Electrode diameter		Nearest fractional equivalent, in.
mm	inch	
1.6	0.06	—
2.0	0.08	$\frac{5}{64}$
2.5	0.10	$\frac{3}{32}$
3.0	0.12	—
3.2	0.13	$\frac{1}{8}$
4.0	0.16	$\frac{5}{32}$
5.0	0.19	$\frac{3}{16}$
6.0	0.23	—
6.3	0.25	$\frac{1}{4}$
8.0	0.32	$\frac{5}{16}$

* American designation: shielded metal arc welding (SMAW).

Safety precautions

Protection of the skin and eyes. Welding gloves should be worn and no part of the body should be exposed to the rays from the arc otherwise burning will result. Filter glasses (p. 23) should be chosen according to BS recommendations. Where there are alternatives the lower shade numbers should be used in bright light or out of doors and the higher shade numbers for use in dark surroundings.

Do not weld in positions where other personnel may receive direct or reflected radiation. Use screens. If possible do not weld in buildings with light coloured walls (white) as this increases the reflected light and introduces greater eye strain.

Do not chip or deslag metal unless glasses are worn.

Do not weld while standing on a damp floor. Use boards.

Switch off apparatus when not in use.

Make sure that welding return leads make good contact, thus improving welding conditions and reduced fire risk.

Avoid having inflammable materials in the welding shop.

Degreasing, using chemical compounds such as trichlorethylene, perchlorethylene, carbon tetrachloride or methyl chloride, should be carried out away from welding operations, and the chemical allowed to evaporate completely from the surface of the component before beginning welding. The first-named compound gives off a poisonous gas, phosgene, when heated or subjected to ultra-violet radiation, and should be used with care. Special precautions should be taken before entering or commencing any welding operations on tanks which have contained inflammable or poisonous liquids, because of the risk of explosion or suffocation. The student is advised to study the following publications of the Department of Employment (SHW–Safety, health and welfare):

SHW 386. *Explosions in drums and tanks following welding and cutting*. HMSO.

SHW Booklet 32. *Repair of drums and tanks. Explosions and fire risks.* HMSO.

SHW Booklet 38. *Electric arc welding*. HMSO.

Memo No 814. *Explosions – Stills and tanks.* HMSO.

Technical data notes No. 18. *Repair and demolition of large storage tanks.* No. 2. *Threshold limit values.* HMSO.

Health and safety in welding and allied processes. The Welding Institute.

Filters for use during welding. BS 679. British Standards Institution.

Protection for eyes, face and neck. BS 1542. British Standards Institution.

Metal arc welding

This section on practical welding techniques applies equally well to all the following steels:

BS 970 *Wrought steels for mechanical and engineering purposes*
Part 1: *General inspection and testing procedures and specific requirements for carbon, carbon manganese, alloy and stainless steel.*
BS 1501 Part 1: *Steel for fired and unfired pressure vessels*
Part 2: *Aluminium steel for fired and unfired pressure vessels.*

In each and every case the electrode recommended by the manufacturer for the particular steel being welded should be used and the procedure advised for the electrode (such as pre- or post-heat) strictly adhered to.

Sources of supply. The source of welding supply can be a.c. or d.c. The relatively cheap a.c. sets with an output maximum of about 150 A are extremely useful for repair and maintenance in small workshops, but are used mainly for steel. Larger a.c. sets are available for fabrication shops including shipyards.

The d.c. output units are used for both steel and non-ferrous metals and can be obtained in current ranges of 300 A, 400 A, as required. The latest sets have thyristor controlled rectifiers with a one knob stepless control of current, and a remote control can also be supplied.

In order to assist the operator, tables are given indicating the approximate current values with various types and sizes of electrodes. These tables are approximate only and the actual value of the current employed will depend to a great extent upon the work. In general the higher the current in the range given for one electrode size, the deeper the penetration and the faster the rate of deposit. Too much current leads to undercutting and spatter. Too small a current will result in insufficient penetration and too small a deposit of metal. As a general rule, slightly increase the arc voltage as the electrode size increases. The angle of the electrode can be varied between 60° and 90° to the line of the weld. As the angle between electrode and plate is reduced, the gas shield becomes less effective, the possibility of adverse effects of the atmosphere on the weld increases, and penetration is reduced.

Full details of currents suitable for the particular electrodes being used are usually found on the electrode packet and should be adhered to; in the following pages it is assumed that, if a.c. is being used, no notice should be taken of the polarity rules, and covered electrodes only should be used.

The techniques of welding mild steel and low-alloy steel are similar.

Diameter of rod (mm)	Current (amperes)		
	Min.	Max.	Average
1.6	25	45	40
2.0	34	65	50
2.5	50	90	90
3.2	60	130	115
4.0	100	180	140
5.0	150	250	200
6.0	200	310	280
6.3	215	350	300
8.0	250	500	350

Note. American designation: straight polarity, electrode −ve, DCEN; reverse polarity, electrode +ve, DCEP. These apply when welding with d.c.

Striking and maintaining the arc. With an electrode of 4 mm diameter and a current of 130–150 A and a mild steel plate 6–8 mm thick, the first exercise is to strike and maintain the arc. The operator should be comfortably placed with the cable to the electrode exerting no pull on the holder (e.g. cable over the operator's shoulder). The electrode tip is scratched on to the plate at the same time as the hand shield with filter glass is drawn across the face. Never expose eyes, face or any part of bare skin to arc flashes, as the effects can be most painful. The electrode is lifted to about 3 mm from the plate, drawing the arc, and the molten pool can be clearly seen (Fig. 1.13). The lighter, brighter colour is the slag while the more viscous, darker colour is the molten metal.

Straight runs on the plate can now be made once the arc can be struck and maintained. As the electrode melts it deposits the metal on the plate and this deposit should be in a straight line and continuous with no gaps or entrapped slag. Too high a current for a given size of electrode will result in

Fig. 1.13. Striking and maintaining the arc.

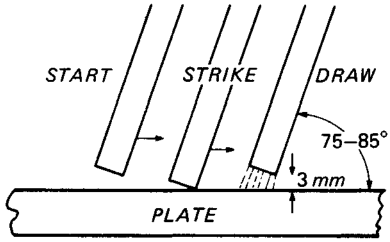

the electrode becoming overheated and eventually red hot. A 4 mm electrode with about 160 A will deposit about 200–235 mm (about 9 in.) of weld metal, if the speed is correct. Slower travel will result in a wider run and it will not be so long, while faster travel results in a longer and narrower bead with the danger of gaps and slag inclusions. The fault of most beginners is that the run is made too quickly. The depth of the crater indicates the amount of penetration (Fig. 1.14*a*).

If, when striking the arc or welding, the electrode touches the plate and 'freezes' it should be freed with a sharp twist of the holder. If this fails, switch off the electric supply at the set and remember that the electrode will still be very hot.

When welding with a d.c. machine, considerable difference can be seen according to whether the electrode is made positive or negative polarity. Since greater heat is generated at the positive pole, making the electrode negative reduces the burn-off rate and deepens the crater and thus the penetration. With the electrode positive the penetration is reduced but the burn-off rate is increased, most heat being generated near the electrode. For this reason (together with the type of electrode coating) most special steels (stainless, heat resistant, etc.), vertical and overhead runs in steel, cast iron and aluminium and bronze are nearly always welded with electrode positive. The instructions on the electrode package issued by the makers should be strictly adhered to.

It will be noticed that when welding on a cold plate the deposited bead rises well above the level of the metal when it is cold. As welding progresses, however, and the parent metal gets heated up, the penetration becomes deeper and the head has a lower contour. For this reason the current setting can be higher when welding is commenced and then should be reduced slightly as the metal heats up, resulting in a bead having an even contour throughout its length.

This exercise should be continued until a straight, even, uniform bead about 250 mm long can be run with good penetration.

If a bead is to be continued after stopping, as for example to change an

Fig. 1.14. (*a*)

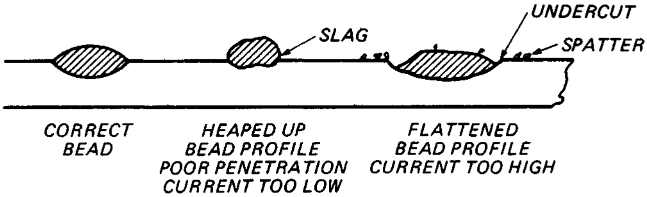
UNDERCUT
SLAG
SPATTER
CORRECT BEAD
HEAPED UP BEAD PROFILE POOR PENETRATION CURRENT TOO LOW
FLATTENED BEAD PROFILE CURRENT TOO HIGH

electrode, the end of the bead and the crater must be deslagged and brushed clean and the arc struck at the forward end of the crater. The electrode is then quickly moved to the rear end and the bead continued, so that no interruption can be detected (Fig. 1.14*b*). This should be practised until no discontinuity in the finished bead can be observed. Since in welding long runs, the welder may have to change rods many times, the importance of this exercise will be appreciated, otherwise a weakness or irregularity would exist whenever the welding operation was stopped.

Beads can now be laid welding away from the operator and also welding from left to right or right to left. A figure-of-eight about 120 mm long provides good practice in this and in changing direction of the bead when welding. The bead should be laid continuously around the figure.

A transformer unit usually has two voltage settings, one 80 open circuit volts and the other 100 OCV, the latter being used for thinner sheet when the current setting is low. Other sets may have lower OCV's, the figure required generally being given on the electrode packet. With an a.c. volts setting of 80 V the current setting for a given electrode can be varied from a low value, giving poor penetration and a shallow crater, with the metal heaping up on the plate producing overlap, to a high value, giving an excessively deep crater, too much penetration with a fierce hissing and not well controllable arc, much spatter, and a flat bead with undercut and an overheated electrode. Intermediate between these extreme values is a correct value of current, say 115–125 A for a 3.2 mm electrode. There is good penetration, the metal flows well, the arc is controllable with no spatter or undercut, and the arc sound is a steady crackle.

These results are tabulated below. If a d.c. set is being used with a voltage control this also can be varied for the operator to find out the effect of this control.

Fig. 1.14 (*b*) Arc struck in front of crater and moved back to continue the run preserves weld profile and helps eliminate any start porosity.

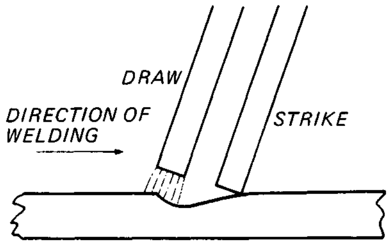

Welding circuit	Effect
Too low current	Poor penetration; shallow crater; metal heaps up on plate with overlap; arc has unsteady spluttering sound.
Too high current	Deep crater; too deep penetration; flat bead; fierce arc with loud crackle; electrode becomes red hot; much spatter.
Too high voltage	Noisy hissing arc; fierce and wandering arc; bead tends to be porous and flat; spatter.
Correct voltage and current and welding speed	Steady crackle; medium crater giving good penetration; easily controlled stable arc; smooth even bead.

Weaving (see pp. 134–6 for automatic weaving)

This may be attempted before or after the preceding exercise according to the inclination of the operator. Weaving is a side to side motion of the electrode, as it progresses down the weld, which helps to give better fusion on the sides of the weld, and also enables the metal to be built up or reinforced along any desired line, according to the type of weave used. It increases the dilution of weld metal with parent metal and should be reduced to a minimum when welding alloy steel.

There are many different methods of weaving, and the method adopted depends on the welder and the work being done. The simplest type is shown in Fig. 1.15*a*, and is a simple regular side to side motion, the circular portion helping to pile the metal in the bead in ripples. Fig. 1.15*b* is a circular motion favoured by many welders and has the same effect as Fig. 1.15*a*. Care should be taken with this method when using heavy coated rods that the slag is not entrapped in the weld. Fig. 1.15*c* is a figure-of-eight method and gives increased penetration on the lines of fusion, but care must again be taken that slag is not entrapped in the overlap of the weave on the edges. It is useful when reinforcing and building up deposits of wear-

Fig. 1.15

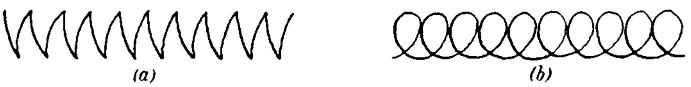

NORMAL WEAVES FOR HORIZONTAL BEAD

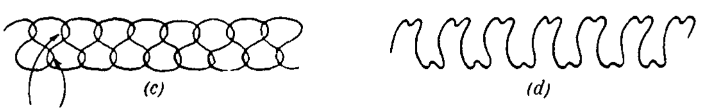

THIS PRODUCES DEEP PENETRATION AND REINFORCEMENT ON THESE LINES

USEFUL FOR HORIZONTAL BEAD ON VERTICAL PLATE. HESITATING ACTION CONTROLS METAL ON EACH SIDE OF BEAD resisting steels. Fig. 1.15*d* is a weave that is useful when running horizontal beads on a vertical plane, since by the hesitating movement at the side of the bead, the metal may be heaped up as required. The longer the period of hesitation at any point in a weave, the more metal will be deposited at this point.

Weaving should be practised until the bead laid down has an even surface with evenly spaced ripples. The width of the weave can be varied, resulting in a narrow or wide bead as required.

Padding and building up shafts

This exercise provides a good test of continued accuracy in laying a weld. A plate of 8–9.5 mm mild steel about 150 mm square is chosen and a series of parallel beads are laid side by side across the surface of the plate and so as to almost half overlap each other. If the beads are laid side by side with no overlap, slag becomes entrapped in the line where the beads meet, being difficult to remove and causing blowholes (Fig. 1.16). Each bead is

Fig. 1.16

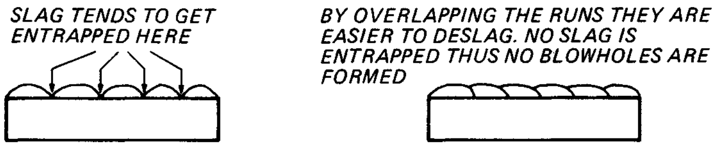
SLAG TENDS TO GET ENTRAPPED HERE
BY OVERLAPPING THE RUNS THEY ARE EASIER TO DESLAG. NO SLAG IS ENTRAPPED THUS NO BLOWHOLES ARE FORMED

Fig. 1.17

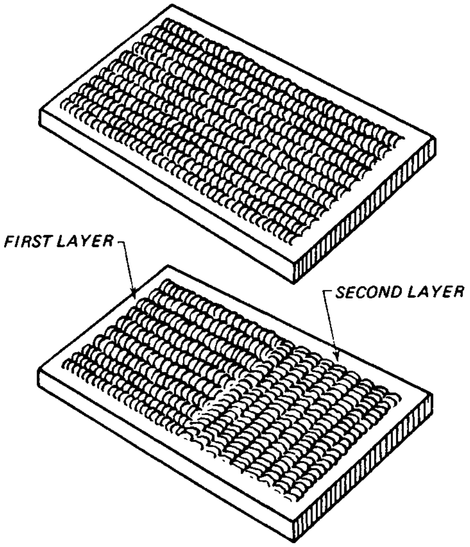
FIRST LAYER
SECOND LAYER

deslagged before the next is laid. The result is a build-up layer of weld metal.

After thoroughly cleaning and brushing all slag and impurities from this layer, another layer is deposited on top of this with the heads at right angles to those of the first layer (Fig. 1.17), or they may be laid in the same direction as those of the first layer. This can be continued for several layers, and the finished pad can then be sawn through and the section etched, when defects such as entrapped slag and blowholes can at once be seen.

Odd lengths of steel pipe, about 6 mm or more thick, may be used for the next exercise, which again consists in building up layers as before. The beads should be welded on opposite diameters to prevent distortion (Fig. 1.18*a*). After building up two or more layers, the pipe can be turned down on the lathe and the deposit examined for closeness of texture and absence of slag and blowholes. Let each bead overlap the one next to it as this greatly reduces the liability of pin-holes in the finished deposit.

The same method is adopted in building up worn shafts and a weld may be run around the ends as shown to finish the deposit. Another method sometimes used consists of mounting the shaft on V blocks and welding spirally (Fig. 1.18*a* and *b*).

Tack welding

Tack welds are essential in welding and fabrication to ensure that there is correct gap and line-up of the components to be welded. The tack welds should be made with a higher current than normal to ensure good fusion and penetration and should be strong enough and of sufficient length and frequency of spacing to ensure rigidity against the distortional

Fig. 1.18. (*a*)

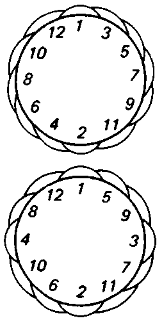

ALTERNATIVE METHODS OF LAYING SEQUENCE OF BEADS. EACH BEAD IS DESLAGGED BEFORE THE OVERLAPPING LAYER IS RUN AND TAKE CARE THERE IS NO ENTRAPPED SLAG IN POCKETS.

Fig. 1.18. (*b*)

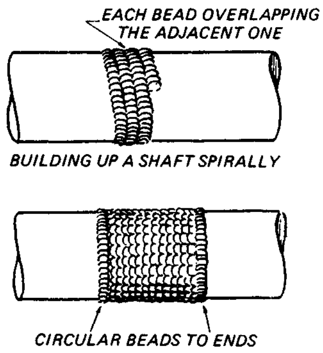
EACH BEAD OVERLAPPING THE ADJACENT ONE
BUILDING UP A SHAFT SPIRALLY
CIRCULAR BEADS TO ENDS

NORMAL METHOD OF BUILDING UP

stresses set up in the welding process. The tack welds should be deslagged and examined for cracks, which may occur because of the rapid cooling of the tack weld if there is no pre-heat. If any cracks are found they should be gouged or ground out and the tack weld remade. Cracks left in may cause cracking in the finished weld.

Thicker plate may need stronger tack welds. These should be made in two runs, the second tack weld being made after the first has been deslagged and examined. Remember when making the first run that the tack weld must be fused into the run so that there is no evidence of the tack weld. This may mean welding at a slightly increased rate at these points. Tack welds are often made by the TIG process in fabrication of plate and pipe work for lining up and then welded by the MIG process.

Plates set up for fillet welding should have the tack welds made at about twice the frequency of pitch used for butt welds.

The operator should now be able to proceed to the making of welded joints, and it will be well to consider first of all the method of preparation of plates of various thicknesses for butt welding. Although a U preparation may prove to be more expensive than a V preparation, it requires less weld metal and distortion will be lessened. For thicker sections the split-weave or stringer bead technique is often used (Fig. 1.21) since the stresses due to contraction are less with this method than with a wide weave.

When a weld is made on a plate inclined at not more than 5° to the horizontal, this is termed the flat position, and wherever possible welding should be done in this position, since it is the easiest from the welder's point of view.

Fig. 1.19 shows some simple welding joints.

Butt welding in the flat position

Joints are prepared with an included angle of 60° with root face and root gap as required for the plate thickness. (An angle of 60° is chosen because it combines accessibility to the bottom of the joint with the least amount of weld metal (Fig. 1.21).)

The plates can be tack welded in position to prevent movement due to expansion and contraction during welding and the tack welds must be strong enough to prevent movement of the plates. Using say a 4 mm electrode with about 125–135 A for a 3 mm gap and 160–170 A for a 1.5 mm gap the welds are kept to about 3–3.5 mm thickness. The length of the run depends upon the size of the electrode and the width of the weld, which gets greater as the joint is filled up. To avoid having the top deposits in thick plate too wide, the stringer or split-weave method is used. If the plates are to be welded from one side only, the penetration *must* be right through to the bottom of the V and the underbead must be continuous along the length of the joint (Fig. 1.20). If a sealing run is to be made the underside of the joint should be back chipped (by electrically or pneumatically operated chipping hammer) so that any defects such as inclusions, porosity, etc., are removed and a sealing run is then made, using a rod about 5 mm diameter.

In many cases the joint to be welded is accessible from one side only, so that the penetration bead cannot be examined for defects and it is not possible to employ a sealing run. In this case a backing bar can be used. The plates are prepared to a sharp edge and the backing bar, usually about 25 mm wide and 5 mm thick, is tack welded to one plate and the plates set apart to the required gap. The first run is made with a higher current than previously used as there is no possibility of burn through. This ensures that the roots of the parent plate and the backing bar are well fused together, and the other runs are then made, as previously, with lower current.

Double-V butt joint preparation is often used with thicker plate because

Fig. 1.19. Simple welding joints.

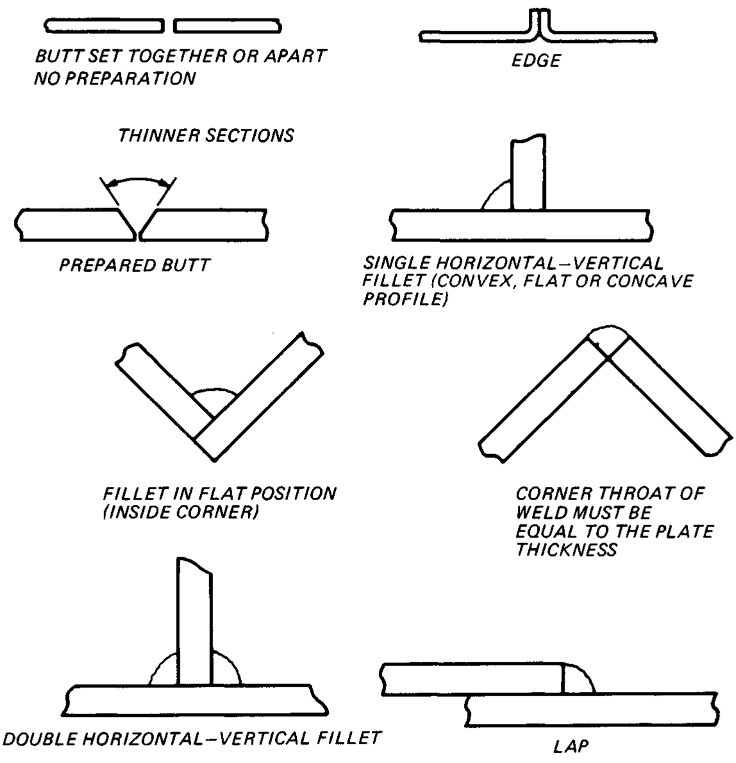
BUTT SET TOGETHER OR APART
NO PREPARATION
EDGE
THINNER SECTIONS
PREPARED BUTT
SINGLE HORIZONTAL–VERTICAL FILLET (CONVEX, FLAT OR CONCAVE PROFILE)
FILLET IN FLAT POSITION (INSIDE CORNER)
CORNER THROAT OF WELD MUST BE EQUAL TO THE PLATE THICKNESS
DOUBLE HORIZONTAL–VERTICAL FILLET
LAP

of the saving in weld metal, which in the case of a 19 mm (¾″) plate is about half of that required for single-V preparation. The plates are prepared to a sharp edge (Fig. 1.21) and the first run made with little or no weave and absolutely continuous. The plates are turned over, the joint root deslagged, after which the next run is made on to the root of the first run, ensuring that there is no entrapped slag or porosity in the centre of the joint. The remaining runs can be made on alternate sides of the joint where possible, reducing the distortional stresses.

Butt welds on pipes provide good exercise in arc manipulation. The pipes are prepared in the same way as for plates by V'ing. They are then lined up in a clamp, or are lined up and tack welded in position and then placed across two V blocks.

Right angle outside corner joints

Tack weld two 12.5 mm (½ in.) steel plates together with a 2 mm gap between them and an angle of 90° between the plates (Fig. 1.22), making a right angled corner joint. With an electrode of 4 mm diameter and

Fig. 1.20

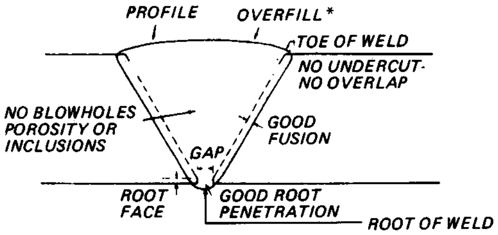
PROFILE
OVERFILL*
TOE OF WELD
NO UNDERCUT-
NO OVERLAP
NO BLOWHOLES
POROSITY OR
INCLUSIONS
GOOD
FUSION
GAP
ROOT
FACE
GOOD ROOT
PENETRATION
ROOT OF WELD

FEATURES OF A GOOD WELD

(a)

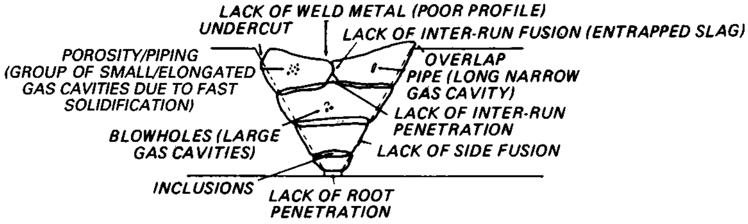
LACK OF WELD METAL (POOR PROFILE)
UNDERCUT
LACK OF INTER-RUN FUSION (ENTRAPPED SLAG)
POROSITY/PIPING
(GROUP OF SMALL/ELONGATED
GAS CAVITIES DUE TO FAST
SOLIDIFICATION)
OVERLAP
PIPE (LONG NARROW
GAS CAVITY)
LACK OF INTER-RUN
PENETRATION
BLOWHOLES (LARGE
GAS CAVITIES)
LACK OF SIDE FUSION
INCLUSIONS
LACK OF ROOT
PENETRATION

IMPERFECTIONS IN A WELD

(b)

---

* The BS term is 'excess weld metal' but the term 'overfill' is now used because in certain circumstances, such as under fatigue loading, excess weld metal (or overfill) is detrimenta to the weld.

about 160 A current make a run about 180 mm long keeping the electrode at an angle of about 60° and ensuring that penetration is adequate but not excessive. The speed of welding should be increased where the tack welds occur to ensure the uniform appearance of the upper side of the weld. Deslag upper and underside of the joint and observe that correct penetration has been achieved. Then make the second (and subsequent) runs with an electrode angle of about 70–75° using a slight weave and avoiding any undercut. Deslagging is followed by the final run to obtain the correct throat thickness and profile as shown. A sealing run can be made if required on the underside of the joint using a somewhat higher current.

Fig. 1.21. Flat butt welds, preparation.

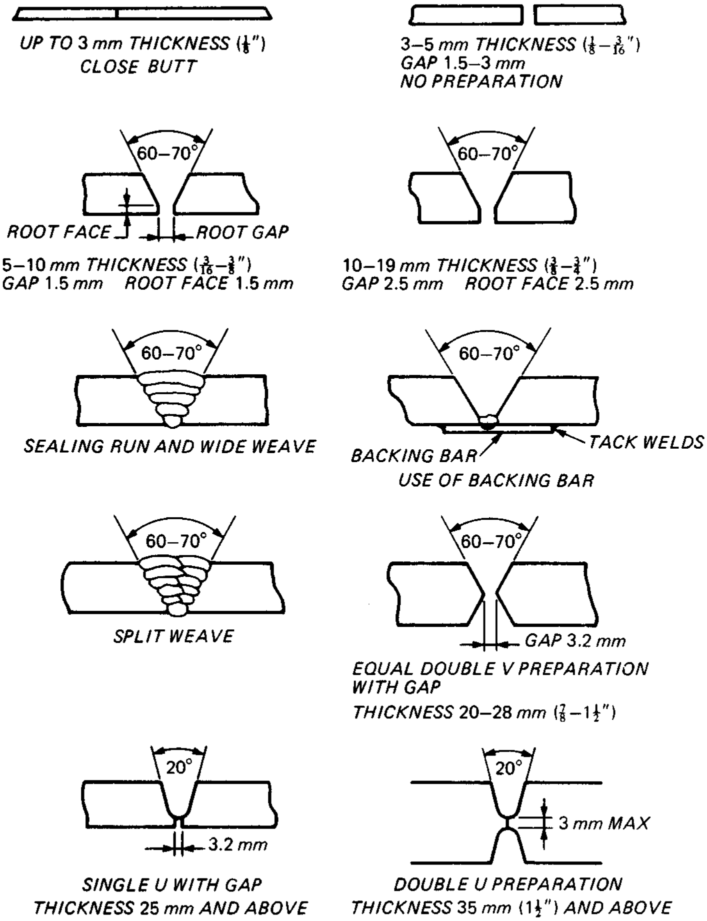
UP TO 3 mm THICKNESS (1/8")
CLOSE BUTT
3–5 mm THICKNESS (1/8–3/16")
GAP 1.5–3 mm
NO PREPARATION
60–70°
ROOT FACE
ROOT GAP
5–10 mm THICKNESS (3/16–3/8")
GAP 1.5 mm ROOT FACE 1.5 mm
60–70°
10–19 mm THICKNESS (3/8–3/4")
GAP 2.5 mm ROOT FACE 2.5 mm
60–70°
SEALING RUN AND WIDE WEAVE
60–70°
BACKING BAR
TACK WELDS
USE OF BACKING BAR
60–70°
SPLIT WEAVE
60–70°
GAP 3.2 mm
EQUAL DOUBLE V PREPARATION
WITH GAP
THICKNESS 20–28 mm (7/8–1½")
20°
3.2 mm
SINGLE U WITH GAP
THICKNESS 25 mm AND ABOVE
20°
3 mm MAX
DOUBLE U PREPARATION
THICKNESS 35 mm (1½") AND ABOVE

Fig. 1.22. Right angle corner welds.

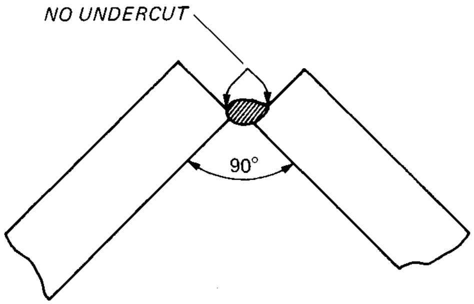
NO UNDERCUT
90°

PLATES TACK WELDED IN POSITION. AFTER DESLAGGING, FIRST RUN GIVES UNIFORM PENETRATION ON UNDERSIDE OF JOINT, TACK WELD MUST BE INCORPORATED INTO THE MAIN WELD, WITHOUT TRACE

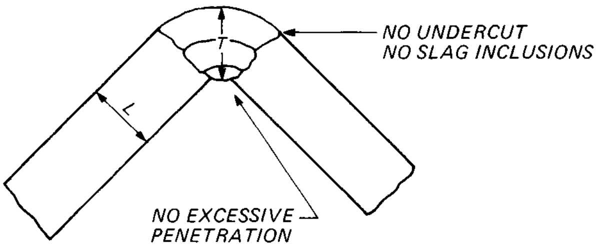
T
L
NO UNDERCUT
NO SLAG INCLUSIONS
NO EXCESSIVE
PENETRATION

THROAT THICKNESS EQUALS PLATE THICKNESS

Fig. 1.23. Fillet weld flat position.

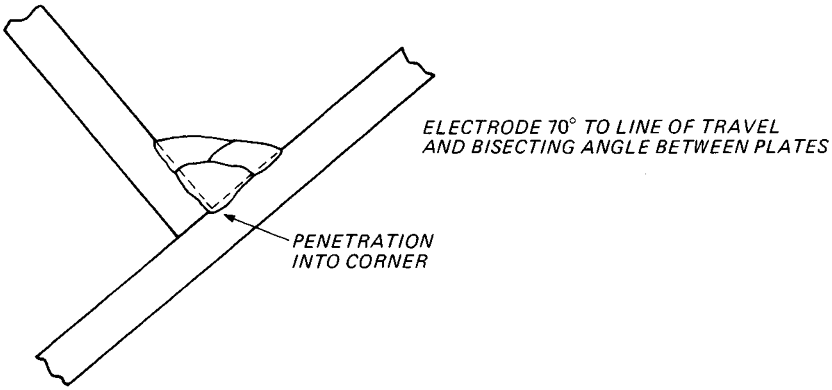
ELECTRODE 70° TO LINE OF TRAVEL
AND BISECTING ANGLE BETWEEN PLATES
PENETRATION
INTO CORNER

Fillet welds in the flat position

Two plates about 9 mm thick are tack welded to make a tee joint and set at an angle of 45° to the horizontal as shown in Fig. 1.23. The weld is made with the electrode at an angle of about 70° to the line of travel and bisecting the angle between the plates. The weld metal must penetrate and fuse into the corner of the joint and there must be no undercutting with equal leg length. A 5 mm electrode at about 200 A will give a 6 mm fillet (throat) about 280 mm long. Make sure that the tack welds, which have been previously deslagged and examined for cracks, etc., are well incorporated into the weld by slightly increasing the welding speed at these points.

Fillet welds in the horizontal–vertical position

With the one plate in the vertical position there will be more tendency for the weld metal to fall downwards to the horizontal plate. To counteract this the electrode is directed at about 40° to the horizontal plate and more vertically to the line of travel – say about 60–80° increasing with the rod diameter. Although the slag, which is more fluid, tends to fall to the horizontal plate, make sure that the weld has equal leg length and avoid the most common error of undercutting in the vertical plate (Fig. 1.24).

Electrode diam. (mm)	Fillet leg length (mm)	Length of run (mm)
3.25 ($\frac{1}{8}$ in.)	4	300 (12 in.)
4.0	5	300
6.0 ($\frac{1}{4}$ in.)	6.0	300

Fig. 1.24. Horizontal–vertical fillet weld.

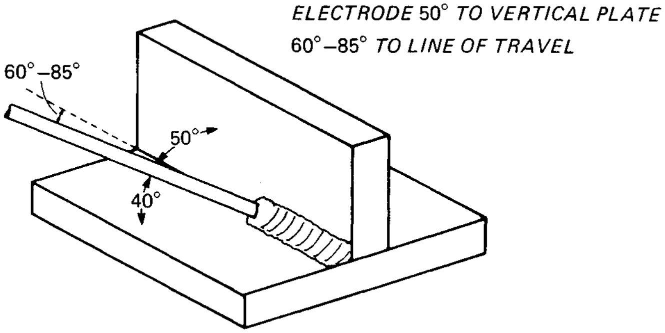

Too long a length of run gives lack of root fusion and too small leg length. Too short a length of weld causes excessive undercut and uncontrolled slag trouble. Root penetration (Figs. 1.26*d* and 1.27) is essential, and avoid the other faults shown.

A lap joint is a particular case of a fillet weld. Take care not to cause breaking away of the edges as shown (Fig. 1.25). Fig. 1.26*a*, *b* and *c* shows various types of fillet welds while Fig. 1.26*e* shows profiles and definitions of leg and throat.

Fig. 1.25. Lap joint.

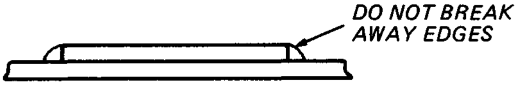
DO NOT BREAK AWAY EDGES

Fig. 1.26. Horizontal–vertical fillet welds. (*a*) Double fillet, no preparation. (*b*) Prepared fillet, welded one side only. (*c*) Multi-run fillet, equal leg length. (*d*) Faults in horizontal–vertical fillet welds. (*e*) Weld profiles.

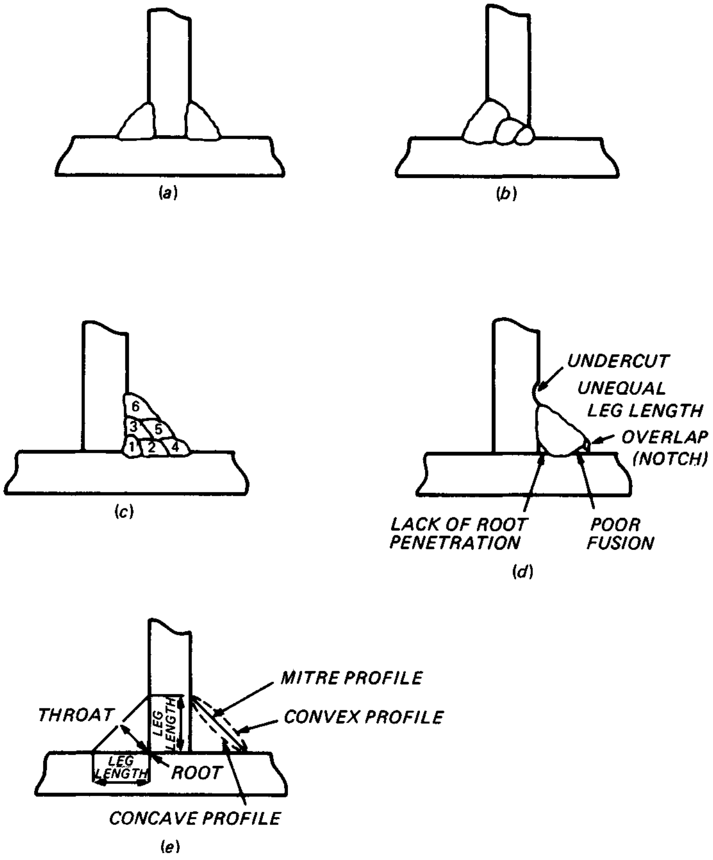
(a)
(b)
6
3
5
1
2
4
(c)
UNDERCUT
UNEQUAL LEG LENGTH
OVERLAP (NOTCH)
LACK OF ROOT PENETRATION
POOR FUSION
(d)
MITRE PROFILE
THROAT
LEG LENGTH
CONVEX PROFILE
LEG LENGTH
ROOT
CONCAVE PROFILE
(e)

Special electrodes for fillet welding help greatly in producing welds having uniform surface, and good penetration with no undercut. Control of the slag often presents difficulties. Keeping the rod inclined at about 70° as before stated helps to prevent the slag running ahead of the molten pool. Too fast a rate of travel will result in the slag appearing on the surface in uneven thicknesses, while too slow a rate of travel will cause it to pile up and flow off the bead. Observation of the slag layer will enable the welder to tell whether his speed is correct or not.

The plain fillet or tee joint (Fig. 1.19) is suitable for all normal purposes and has considerable strength. The prepared fillet (Fig. 1.26*b*) is suitable for heavier loads, and is welded from one side only.

In thick sections a thinner electrode is used for the first bead, followed by final runs with thicker rods, each run being well deslagged before the next is laid.

Vertical welding

All welds inclined at a greater angle than 45° to the horizontal can be classed as vertical welds.

Vertical welding may be performed either upwards or downwards, and in both methods a short arc should be held to enable surface tension to pull the drop across into the molten pool. Electrodes 4 mm diameter or smaller

Fig. 1.27. (*a*) A horizontal–vertical fillet weld showing good penetration. (*b*) Showing poor penetration at corner of fillet.

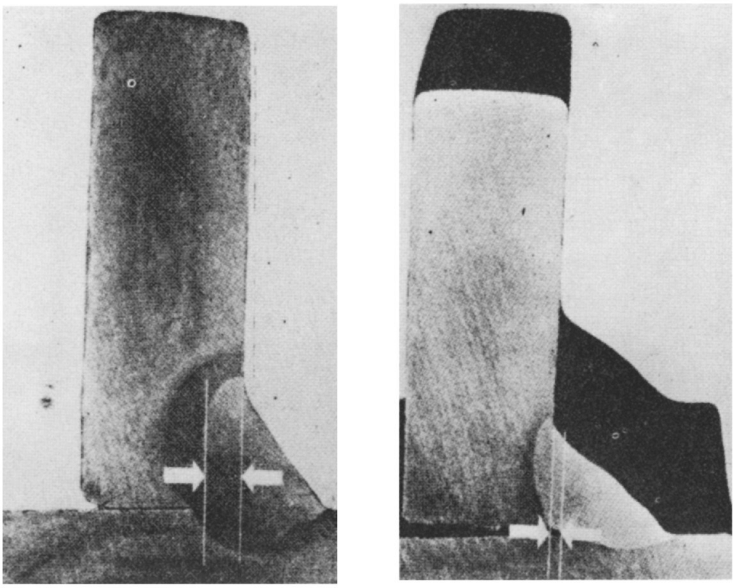

are generally used, and special rods having light coatings to reduce difficulties with slag are available. Vertical beads should first be run on mild steel plate, the electrode being held at 75–90° to the plate (Fig. 1.28).

Downward welding produces a concave bead, and is generally used for lighter runs, since a heavy deposit cannot be laid. If it is, the metal will not freeze immediately it is deposited on the plate, and will drop and run down the plate. This method is, therefore, usually only used as a finishing or washing run over an upward weld because of its neat appearance, or for thin sections.

Fig. 1.28

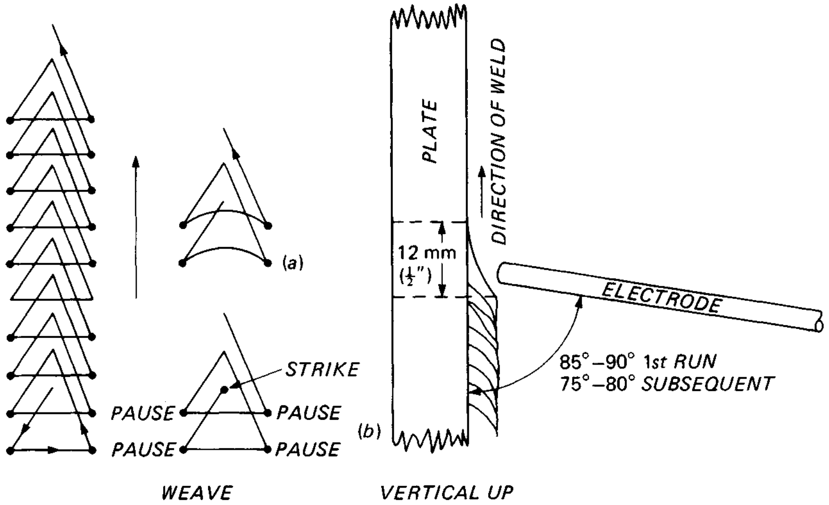
(a)
STRIKE
PAUSE
PAUSE
PAUSE
PAUSE
(b)
WEAVE
PLATE
DIRECTION OF WELD
12 mm
(½")
ELECTRODE
85°–90° 1st RUN
75°–80° SUBSEQUENT
VERTICAL UP

Fig. 1.29

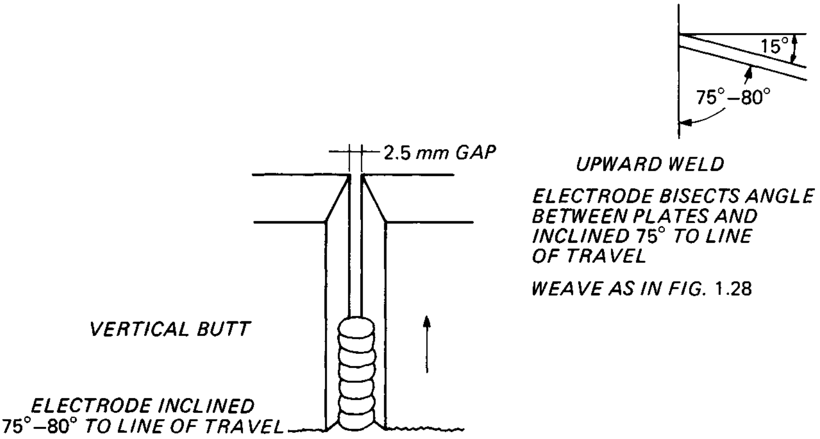
15°
75°–80°
2.5 mm GAP
UPWARD WELD
ELECTRODE BISECTS ANGLE BETWEEN PLATES AND INCLINED 75° TO LINE OF TRAVEL
WEAVE AS IN FIG. 1.28
VERTICAL BUTT
ELECTRODE INCLINED 75°–80° TO LINE OF TRAVEL

Upward welding (Fig. 1.29) produces a convex bead, and is used on sections of about 6 mm thickness. The metal just deposited is used as a step on which to continue the deposit, and the slag flows away from the pool and does not hinder penetration as it does in the downward method.

See Fig. 1.28 for angle of electrode and method of weaving. Hesitate at each side and travel from left to right by moving upwards as shown. This keeps the upper side of the pool shallow and enables the slag to run away easily with no danger of entrapping. Fig. 1.30 shows a vertical fillet weld.

Horizontal butt welding

This type of weld is very useful in fabrication, repair and reinforcement. A weaving motion is used for the first runs having prepared the plates as shown in Fig. 1.31 with a medium current and a short arc. Subsequent runs can be made with very little weave and again a short arc – avoid undercutting the upper plate. (Note that there is little if any preparation of the lower plate.)

In reinforcement of vertical surfaces, the horizontal layer may be first laid along the bottom of the surface and succeeding beads built up above this, using the lower bead in each case as a step. The next layer is then deposited with the beads at right angles to those in the first layer as in padding, this second layer consisting of normal vertical beads. An alternative method is to lay all the beads vertically, though this is scarcely so satisfactory.

Fig. 1.30. Vertical fillet weld.

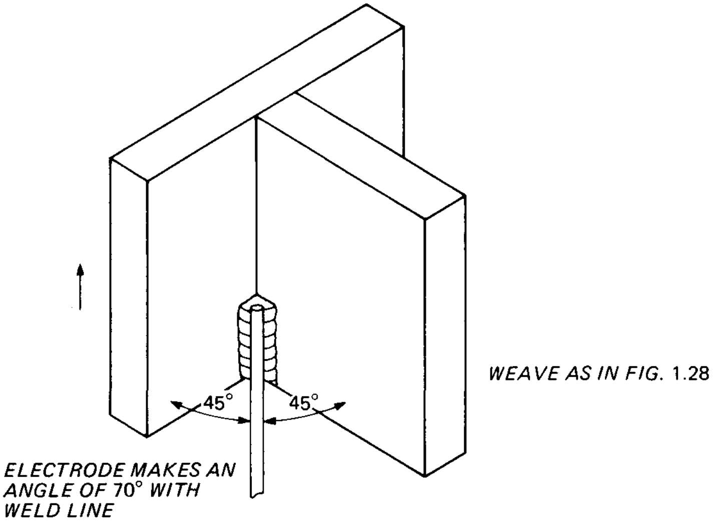

Overhead welding

Overhead welding takes a great deal of practice before the operator can deposit an even weld with no blemishes. The weld should be made with electrodes of 3.2 or 4 mm diameter specially designed for all positional welding. The slag from these rods is fluid and does not give much trouble when welding. Because the electrodes are relatively small in diameter, overhead welding is a relatively slow process and should only be used when it is not possible to perform the welding in any other position such as downhand. An example is the overhead position when welding on the underside of a pipe joint in pipeline welding.

The operator must wear suitable protective clothes, tight wrist protection, gauntlet gloves, protective apron over shoulders and neck, no turn-ups on trouser bottoms and protective leather apron. The welding mask must completely cover face, head and neck. The welding position must be comfortable with the weight of the welding cable taken over the shoulder and the position should be such that molten slag (or even metal) does not fall on the operator.

Overhead butt weld (Fig. 1.32*a*). Using about 110 A for a 3.2 mm diameter rod make the first or root run with no weave and the rod at about 80–85° to the weld line. The arc must be as short as possible to enable surface tension and the magnetic forces to pull the metal up into the pool against the pull of gravity. After cleaning the first run, the subsequent runs are made with a little weave using a 4 mm diameter electrode and a current of about 160 A.

Overhead fillet welds are made with the same electrode and current settings as the overhead butt joint. Keep the arc very short and use no weave. If,

Fig. 1.31. Horizontal butt welds.

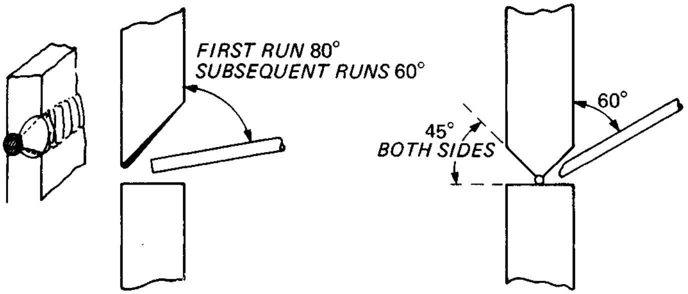

HORIZONTAL BUTT WELD FROM ONE SIDE ONLY

HORIZONTAL BUTT WELD FROM BOTH SIDES: 2 WELDERS WORKING SIMULTANEOUSLY REDUCE DISTORTION PROBLEMS AND THUS INTERNAL STRESS because of the short arc, the electrode persists in sticking, raise the current a little, and avoid undercutting of the upper plate (Fig. 1.32*b*).

Aluminium and aluminium alloys

Aluminium can be arc welded using electrodes coated with fluxes consisting of mixtures of fluorides and chlorides. The flux dissolves the layer of aluminium oxide (alumina) on the surface of the metal, and also prevents oxidation during welding. The melting point is about 660°C but aluminium has a high thermal conductivity and as a result, since there is little if any change of colour before melting point is reached, as with steel, the rapid melting which occurs presents the operator with some difficulty when first beginning welding aluminium. Except for the thinnest sheets, pre-heating to the range 100–300°C should be carried out and great pains taken that the area of welding is free from all oxides, paints, and other contamination and the joint should be cleaned by wire brushing or machining immediately before beginning to weld.

Fig. 1.32. (*a*) Overhead butt weld with backing bar.

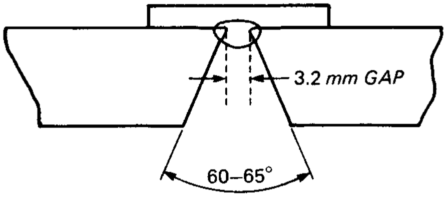

FIRST RUN FUSES INTO BACKING

ELECTRODE ANGLE 80–85°
WITH LINE OF WELD

Fig. 1.32. (*b*) Overhead fillet weld.

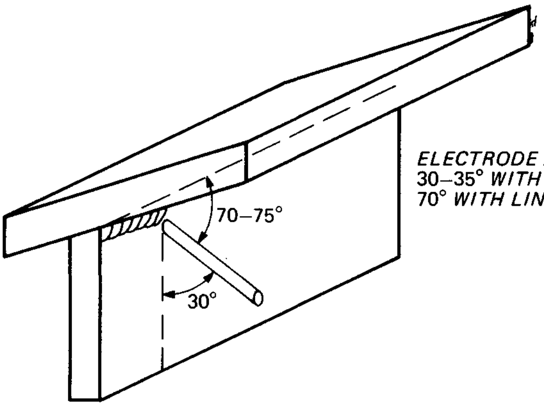

ELECTRODE ANGLE
30–35° *WITH VERTICAL PLATE*
70° *WITH LINE OF WELD*

Preparation. Fig. 1.33 shows the preparation for various sheet thicknesses. Tack welds, jigs, or other clamps should be used to position the work and tack welding should be done at very frequent intervals, then deslagged and smoothed before welding commences. The thermal expansion is about twice that of mild steel so allowance must be made accordingly. Molten aluminium absorbs hydrogen and this results in porosity so that preheating the work and making sure that the electrodes are kept dry and heating to 130–160°C prevents this.

The table shows the more generally welded alloys. The electrodes used are of the 10–12% silicon type, so that if heat treatable alloys are welded the heat treatment is lost and the alloy needs heat treatment again. The Al–Si electrode does not have the same mechanical properties as the parent plate in every case, and it should be remembered that the higher Al–Mg alloys are not satisfactorily arc welded. Aluminium and the work hardening Al–Mg alloys are now fabricated by the TIG and MIG processes (q.v.).

Electrodes	*Material examples*
Aluminium–10% Si	Pure aluminium, 5251 (N4), 6063 (H9), 6061 (H20), 6082 (H30), Castings LM18, LM6, LM8, LM9, LM20.

Technique. The rod is connected to the positive pole of a d.c. supply, and the arc struck by scratching action, as explained for mild steel. It will be found that, as a layer of flux generally forms over the end of the rod, it has to be struck very hard to start the arc. The rod is held at right angles to the work and a *short* arc must be held (Fig. 1.34), keeping the end pushed down into the molten pool. This short arc, together with the shielding action of the coating of the rod, reduces oxidation to a minimum. A long arc will result in a weak, brittle weld. No weaving need be performed, and the rate of

Fig. 1.33. Preparation of aluminium.

CLOSE SQUARE BUTT
WITH BACKING STRIP
NO PREHEAT

CLOSE SQUARE BUTT.
WELD BOTH SIDES 4.8–9.5 *mm*
PRE-HEAT 100–300 °C

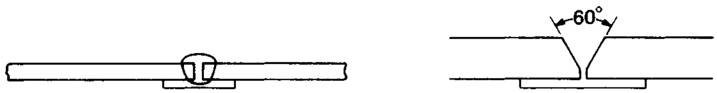

SQUARE BUTT WITH BACKING STRIP
AND WITH GAP. 4.8–9.5 *mm*
PRE-HEAT 100–200° *C*

60° *V PREPARATION*
ROOT GAP 1.6–3.2 *mm*
ROOT FACE 4–5 *mm*
PRE-HEAT 100–300 °*C*

welding must be uniform. As the metal warms up, the speed of welding must be increased.

Castings are welded in the same way after preparation, but owing to their larger mass, care must be taken to get good fusion right down into the parent metal, since if the arc is held for too short a time on a given portion of the weld the deposited aluminium is merely 'stuck' on the surface as a bead with no fusion. This is a very common fault. Castings should be preheated to 200°C to reduce the cracking tendency and to make welding easier.

Lap joints and fillet joints should be avoided since they tend to trap corrosive flux, where it cannot be removed by cleaning. Fillet welds are performed with no weave and with the rod bisecting the angle between the plates.

After treatment. The flux is very corrosive and the weld must be thoroughly washed and brushed in hot water after it has cooled out. Immersion in a 5% solution of nitric acid in water is an even better method of removing the flux, this being followed by brushing and washing in hot water.

Cast iron

The following types of electrode are in general use for the welding of cast iron.

(1) Mild steel, low carbon, basic coated (low hydrogen), electrode d.c. + ve, a.c. 40 OCV.
(2) Nickel cored, electrode d.c. + ve, a.c. 40 OCV.
(3) MONEL® nickel–copper alloy* cored electrode, d.c. + ve, a.c. 40 OCV.†
(4) Nickel iron cored electrode, d.c. – ve, a.c. 40 OCV.

(1) When steel-based weld metal is deposited on cold cast iron, quick cooling results, due to the large mass of cold metal near the weld. This quick

Fig. 1.34. Aluminium welding.

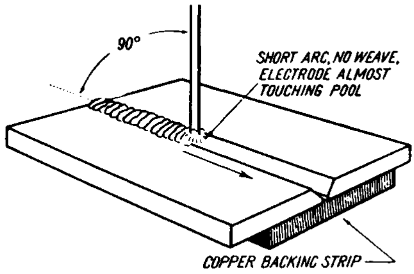

---

* MONEL, INCOLOY, INCONEL, etc. are trademarks of the Inco family of companies.

† See table, p. 71.

cooling results in much of the carbon in an area adjacent to the weld being retained in the combined form (cementite) and thus a hardened zone exists near the weld. In addition, the steel weld metal absorbs carbon and the quick cooling causes this to harden also. As a result welds made with this type of rod have hard zones and cannot always be machined.

If, however, there can be pre-heat to about 500–600 °C followed by slow cooling the deposit may produce machinable welds. In many cases, however, machining is not necessary, thus this drawback is no disadvantage. The weld is strong and electrodes of about 3.2 mm diameter are generally used with a low current which introduces a minimum of heat into the work.

(2) The nickel cored electrode may be used on cast iron without any pre-heating and it gives a deposit that is easily machinable with easy deslagging. It can be used in all positions and is used for buttering layers.

(3) The MONEL® alloy cored electrode is easy to deposit on cold cast iron and is again easily machinable. Nickel and MONEL® alloy electrodes have reduced carbon pick-up and thus a reduced hardening effect, but pre-heating should be done with castings of complicated shape, followed by slow cooling though, with care, even a complicated casting may be welded satisfactorily without pre-heat if the welding is performed slowly. The lower the heat input the less the hardening effect in the HAZ.

(4) Electrodes which deposit nickel–iron alloy are generally used where high strength is required, as for example with the SG irons.

In the repair of cast iron using high nickel electrodes the weld metal is strong and ductile.. The electrode coating has a high carbon content and gives up to $1\frac{1}{2}\%$ carbon, as graphite, in the weld. The carbon has a low solubility in the nickel and excess carbon appears in the weld, increasing weld volume and reducing shrinkage. Pre-heating may be done whenever the casting is of complicated shape and liable to fracture easily, though, with care, even a complicated casting may be welded satisfactorily without pre-heating if the welding is done slowly.

Nickel alloy electrodes are also used for the welding of SG cast irons, but the heat input will affect the pearlitic and ferritic structures in the heat-affected zone, precipitating eutectic carbide and martensite in a narrow zone at the weld interface even with slow cooling. For increased strength, annealing or normalizing should be carried out after welding. The lower **the heat input the less the hardening effect in the HAZ. See table, p. 71, for suitable electrodes.**

Preparation. Cracks in thin castings should be V'd or, better still, U'd, as for example with a bull-nosed chisel. Thicker castings should be prepared with a single V below 9.5 mm thick and a double U above this. Studding (q.v.) can be recommended for thicker sections and the surrounding metal should be thoroughly cleaned. The polarity of the electrode depends on the electrode being used and the maker's instructions should be followed, though with d.c. it is generally + ve. With a.c. an OCV of 80 V is required.

Since the heat in the work must be kept to a minimum, a small-gauge electrode, with the lowest current setting that will give sufficient penetration, should be used. A 3.2 mm rod with 70 to 90 A is very suitable for many classes of work. Thick rods with correspondingly heavier currents may be used, but are only advisable in cases where there is no danger of cracking. Full considerations of the effect of expansion and contraction must be given to each particular job.

Technique. The rod is held as for mild steel, and a slight weave can be used as required. Short beads of about 50 to 60 mm should be run. If longer beads are deposited, cracking will occur unless the casting is of the simplest shape. In the case of a long weld the welding can be done by the skip method, since this will reduce the period of waiting for the section welded to cool. It may be found that with steel base rods, welding fairly thin sections, fine cracks often appear down the centre of the weld on cooling. This can often be prevented, and the weld greatly improved, by peening the weld immediately after depositing a run with quick light blows with a ball-paned hammer. If cracks do appear a further light 'stitching' run will seal them. Remember that the cooler the casting is kept, the less will be the risk of cracking, and the better the result. Therefore take time and let each bead cool before laying another. The weld should be cool enough for the hand to be held on it before proceeding with the next bead. In welding a deep V, lay a deposit on the sides of the V first and follow up by filling in the centre of the V. This reduces risk of cracking. If the weld has been prepared by studding (q.v.), take care that the studs are fused well into the parent metal.

In depositing non-ferrous rods, the welding is performed in the same way, holding a short arc and welding *slowly*. Too fast a welding speed results in porosity. In many cases a nickel–copper rod may be deposited first on the cast iron and then a steel base rod used to complete the weld. The nickel–copper rod deposit prevents the absorption of carbon into the weld metal and makes the resulting weld softer. Where a soft deposit is required on the surface of the weld for machining purposes, the weld may be made in the ordinary way with a steel base rod and the final top runs with a non-ferrous rod. The steel base rod often gives a weld which has hard spots in it that can only be ground down, hence this weld can never be completely guaranteed machinable.

Studding

We have seen that whenever either steel base or non-ferrous rods are used for cast iron welding, there is a brittle zone near the line of fusion, and since contraction stresses are set up, a weakness exists along this line. This weakened area can be greatly strengthened in thick section castings by studding. Welds made by this method have proved to be exceptionally strong and durable. Studding consists of preparing the casting for welding with the usual single or double V, and then drilling and tapping holes along the V and screwing steel studs into the holes to a depth slightly deeper than the diameter of the studs. Studs of 4.8 mm diameter and larger are generally used, depending on the thickness of the casting, and they must project about 6 mm above the surface. The number of studs can be such that their area is about $\frac{1}{5}$ to $\frac{1}{4}$ of the area of the weld, though a lesser number can be used in many cases (Fig. 1.35). Welding is performed around the area near each stud, using steel or steel base electrodes, so as to ensure good fusion between the stud and the parent casting. These areas are then welded together with intermittent beads, as before explained, always doing a little at a time and keeping the casting as cool as possible. This method should always be adopted for the repair, by arc welding, of large castings subjected

Fig. 1.35. Methods of studding.

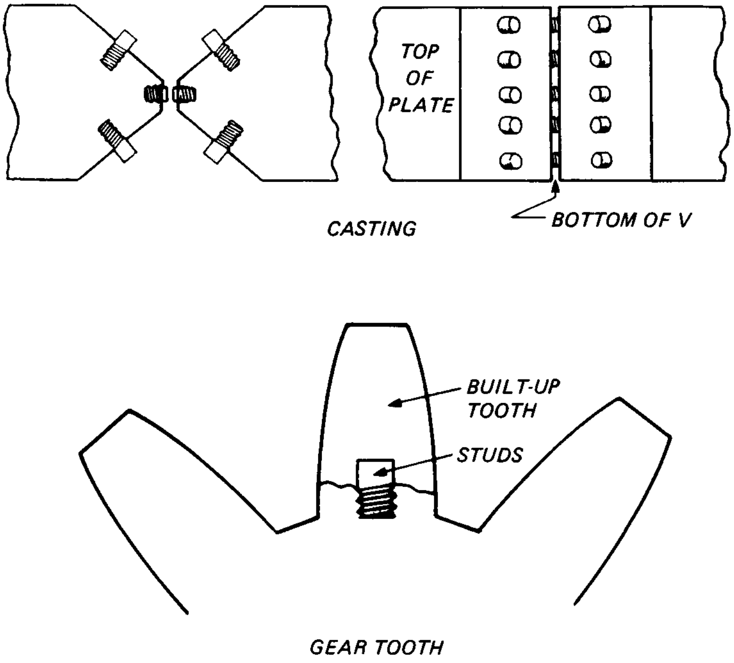

to severe stress. An alternative method is to weld steel bars across the projecting studs as additional reinforcement.

Copper, bronze and brass welding

If the weld must have the same characteristics as the parent metal, as for example for electrical conductivity, great difficulty is encountered with the welding of copper by the MMA process because the weld is very porous and unsatisfactory and the TIG and MIG processes should be chosen. Similarly for brass but with the use of an alloy electrode of bronze (about 80% copper, 20% tin or similar) welding can be satisfactorily carried out on most of the copper alloys with certain exceptions. Welding is performed with d.c. and the electrode positive.

Copper, copper–tin (bronze) and copper–zinc (brass) have a much greater coefficient of thermal conductivity than mild steel and the thermal expansion is also greater so that more heat is required in the welding process, so that it is almost imperative to pre-heat except with the smallest sections, and since tack welds lose their strength at elevated temperatures, line up of the parts to be welded should be with clamps, jigs or fixtures, and due allowance made for expansion and contraction during and after welding. In general, great care must be taken to avoid porosity due to the presence of gas which is expelled on solidification. Stress relief (post weld) can be carried out on the work hardening alloys to about 250–300 °C.

Copper. Tough pitch copper should not be welded as the weld is always porous. Phosphorus deoxidized copper is satisfactorily welded but the weld, made with a rod of copper–tin (bronze) will not have the same properties as the parent plate. Pre-heating is up to 250 °C.

Bronze. The tin bronzes up to about 9% Sn are weldable, but those with higher tin content are not welded due to solidification cracking and any alloy with more than 0.5% Pb is usually not welded.

Brasses. These are alloyed mostly with zinc and those with single-phase composition, as for example cartridge brass (70% Cu, 30% Zn) and those with up to about 37% Zn, are weldable, but yellow or Muntz metal (60% Cu, 40% Zn) and manganese bronze (58% Cu, 38% Zn, rem, Mn, Fe, Ni or Sn) really need pre-heat to about 450 °C.

Preparation. Plates below 3 mm thickness need no preparation, but above this they are prepared as for mild steel with a 60° V and the surfaces thoroughly cleaned. The work must be supported during welding. Sections above about 4 mm must be pre-heated to 250°C because of the high thermal conductivity of the metal.

Technique. The rod should be connected to the +ve pole of a d.c. supply except where otherwise stated. The current value is generally the same as or slightly less than that for the same gauge mild-steel electrode, with an arc voltage drop of 20–25 V. The actual values will depend on the particular job. The electrode should be held steeply to the line of weld, and a short arc held keeping the electrode well down into the molten pool. The weld must be made *slowly*, since a quick rate of welding will produce a porous deposit even when bronze rods are used. As welding proceeds and the part heats up, it is usually advisable to reduce the current, or there is a liability to melt through the weld, especially in thinner sections. Light hammering, while still hot, greatly improves the structure and strength of the weld. The extent to which this should be done depends on the thickness of the metal.

Hard surfacing

Surface resistance to abrasion, impact, corrosion and heat. The advantages of hard surfacing are that the surfaces can be deposited on relatively much cheaper base metal to give the wear-resistant or other qualities exactly where required, with a great saving in cost. In addition built-up parts save time and replacement costs. The chief causes of wear in machine parts are abrasion, impact, corrosion and heat. In order to resist impact the surface must be sufficiently hard to resist deformation yet not hard enough to allow cracks to develop. On the other hand, to resist abrasion a surface must be very hard and if subject to severe impact conditions cracking may occur, so that in general the higher the abrasion resistance the less the impact resistance, and evidently it is not possible to obtain a surface which has the highest values of both impact and abrasion resistance. In choosing an electrode for building up surfaces consideration must be given as to whether high abrasion or high impact resistance is required. High-impact electrodes will give moderate abrasion resistance and vice versa, so that the final choice must be made as to the degree of (1) hardness, (2) toughness, (3) corrosion resistance, (4) temperature of working, (5) type of base metal and whether pre- and post-heating is possible. The main types of wear- and abrasion-resistant surfacing electrodes are:

(1) Fused granules of tungsten carbide in an austenitic matrix. The deposit has highest resistance to wear but is brittle and has medium impact strength. The electrodes are tubular and therefore moisture-resistant with high deposition rates.

(2) Chromium carbide. The basic-coated electrodes deposit a dense network of chromium carbide in an austenitic matrix and have high resistance to wear with good impact resistance.

(3) Cobalt–chromium–tungsten non-ferrous alloys. These have a high carbon content, are corrosion-resistant, have a low coefficient of friction and retain their hardness at red heat.

(4) Nickel-base alloys containing chromium, molybdenum, iron and tungsten. These have good abrasion resistance and metal-to-metal impact resistance. They work-harden and have resistance to high-temperature wear and corrosive conditions.

(5) Air-hardening martensitic steels. These have a high hardness value due to their martensitic structure and there is a variety of electrodes available in this group. Dilution plays an important part in the hardness of the deposit, as with all surfacing applications. A single run on mild steel may be only a little harder than the parent plate, but if deposited on carbon or alloy steel, carbon and alloy pick-up greatly increases the hardness of the deposit.

(6) Austenitic steels: (1) 12/14% manganese deposits develop their hardness by cold work-hardening so that the deposit has the strength of its austenitic core with the hard surface. With approximately 3% Ni there is reduced tendency to cracking and brittleness due to heat compared with plain 12/14% Mn weld metal. (2) Chromium–nickel, chromium–molybdenum–nickel and chromium–manganese deposits work-harden and give resistance to heavy impact.

Thus, in general, low-alloy deposits give medium abrasion with high impact resistance, medium-alloy deposits give high abrasion and medium impact properties.

Austenitic, including 13% manganese, deposits give high impact resistance and work-harden as a result of this impact and work to give abrasion-resistant qualities. Of the carbides, chromium deposits give the best abrasion and impact properties while tungsten has the hardest surface and thus the highest resistance to abrasion with medium impact properties.

The table overleaf gives a selection of electrodes available.

Preparation and technique

The surface should be ground all over and loose or frittered metal removed. Because of the danger of cracking, large areas should be divided up and the welding done in skipped sections so that the heat is distributed as evenly as possible over the whole area. Sharp corners should be rounded and thick deposits should be avoided as they tend to splinter or spall. The

Type (R – rutile) (B – basic) (T – tubular)		Hardening	HV	Use	Abrasion (H – high) (M – medium)	Impact
pearlitic P	Cr–Mn	air	250	Used also for butter layers.	M	M
martensitic R	Cr–Mn	air	350	For and with buffer layers.	M	H
martensitic R	Cr–Mn–Mo	air	650	After buffer layers.	H	M
martensitic B	Cr–Mo–V	air	700	After buffer layers for heavy reinforcement.	H	M
martensitic B	Cr–Mo with borides	air	800	Surfacing generally restricted to two layers.	H	M
austenitic B	Cr–Mo–Ni	work	250 500	For joining and depositing on 13% Mn steel and joining this steel to carbon steel.	M	M
austenitic B	Cr–Ni–W–B	work (slight)	450 500	Hard at elevated temperatures, corrosion-resistant.	M	M
austenitic B	13% Mn	work	170 500		H	M
austenitic T	Cr–Mn	work	300 480	For reinforcing 13% Mn steel castings and as buffer layer for harder surfaces.	M	H
non-ferrous B	Co–Cr–W		630	Red hard and corrosion-resistant, various grades.	H	M
austenitic T matrix	Chromium–carbide–Mn		560 matrix 1400 carbides	Heat resistant to about 1100°C.	H	H
non-ferrous T matrix	tungsten–carbide		600 matrix 1800 carbides	Used at all temperatures.	H	M

first runs on any surface are subject to considerable dilution, so it is advisable to lay down a buffer layer using a nickel-based or austenitic stainless steel electrode, especially when welding on high carbon or high alloy steels. The buffer layer should be chosen so as to be of intermediate hardness between the parent metal and the deposit, or two layers should be used with increasing hardness if the parent metal is very soft compared with the deposit.

Mild pre-heating to 150–200 °C is advantageous if the base metal has sufficient carbon or alloying elements to make it hardenable but is not usual on steel below 0.3% C or stainless steel. If the base metal is very hard and brittle, slow pre-heating to 400–600 °C with slow cooling after welding may be necessary to prevent the formation of brittle areas in the HAZ. Below a hardness value of 350 HV (hardness Vickers), surfaces are machinable generally with carbide tools but above this grinding is usually necessary. In depositing a surface on manganese steel there should be no pre-heating or stress relieving and the electrode should be connected to the +ve pole. Use only sufficient current to ensure fusion, keep a short arc, hold the electrode as in welding mild steel and introduce as little heat as possible into the casting by staggering welding so that the temperature does not rise much above 200 °C. Austenitic stainless steel electrodes should be used for joining broken sections as these electrodes work-harden in the same way as the parent metal and 13% manganese, or 13% manganese, 3.5% nickel electrodes used for building up.

Dilution is about 25–35% in the first layer, and to obtain a dilution of about 5% at least three layers must be laid down with a thickness of 6–10 mm.

Note. The manual metal arc method of surfacing is rather slow compared with semi-automatic and automatic processes such as MIG and submerged arc so that it is best suited for smaller areas and complex shapes.

Corrosion-resistant surfacing

Nickel and its alloys can be laid as surfaces on low carbon and other other steel to give corrosion- and heat-resistant surfaces. Electrode diameter should be as large as possible, with minimum current compatible with good fusion to reduce dilution effects. The first bead is laid down at slow speed and subsequent runs overlapped with minimum weaving, reducing dilution to 15–20% compared with 25–35% for the first bead. Subsequent layers should be put down with interpass temperature below 180 °C to minimize dilution and avoid micro-fissures in the deposit. Dilution is reduced to about 5% after three layers with 6–10 mm thickness. Suitable electrodes are in the Nickel, MONEL®, INCONEL®, INCOLOY® and INCO® alloy ranges, for which the student should refer to manufacturers' instructions.

Tipping tool steel

Cutting tools for lathes, milling machines and high-speed cutting tools of all types can be made by depositing a layer of high-speed steel on to a shank of lower carbon steel. Special electrodes are made for this purpose, and give very good results.

The surface to be tipped is ground so as to receive the deposit. The electrode is connected to the + ve pole when d.c. is used, and held vertically to the line of weld, a narrow deposit being laid as a general rule. More than one layer is generally advisable, since the first layer tends to become alloyed with the parent metal. For this reason the current setting should be as low as possible, giving small penetration. Use a very narrow weave so as to prevent porosity, which is very usual unless great care is taken. Each bead should be allowed to cool out and then be deslagged before depositing the next. The deposited metal when allowed to cool out slowly usually has a Brinell hardness of 500–700, depending upon the rod used. The hardness can be increased by heat treatment in the same way as for high-speed tool steel, and the deposit retains its hardness at fairly high temperatures.

These types of rods are very suitable for depositing cutting edges on drills, chisels, shearing blades, dies and tappets, etc. Fig. 1.36 shows the method of depositing the surface on a lathe tool.

Stellite (Cobalt base and nickel base alloys for hard surfacing)

Stellite is an alloy of cobalt, chromium, tungsten and carbon, and when deposited on steel, steel alloy or cast iron it gives a surface having excellent wear-resisting qualities and one that will stand up well to corrosive action. It preserves its hardness of surface even when red-hot (650 °C), and is thus suitable for use in places where heat is likely to be generated.

It may be deposited very satisfactorily with the arc, though the deposit is not as smooth as one deposited with the oxy-acetylene flame. The arc method of application, however, is specially recommended where it is essential not to introduce undue heat into the part, due to danger of warping or cracking. In many cases, especially where the part has large mass, Stelliting by the arc saves time and money compared with the flame. Bare rods of stellite may be used when welding with d.c. and are connected to the + ve pole. If a.c. is employed, covered rods must be used, and better

Fig. 1.36

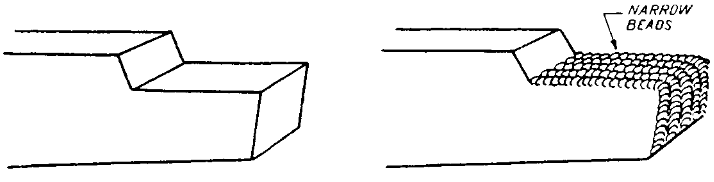

TOOL PRIOR TO DEPOSIT

TOOL TIPPED AND READY FOR GRINDING

results are also obtained with covered rods when using d.c., since the deposit is closer grained and the arc more stable.

Covered rods can be obtained or bare rods may be fluxed with a covering of equal parts of calcium carbonate (chalk), silica flour and either borax or sodium carbonate (baking soda), mixed with shellac as a binder. Coated electrodes give a gas shield which protects the molten metal from oxidation in the welding process and provides a protective slag over the metal when cooling. The deposit resists heat, abrasion, impact, corrosion, galling, oxidation, thermal shock and cavitation erosion to varying degrees depending upon the alloy. Types of Stellite alloys available are:

No. 1. 2.5% C, resistant to abrasion and solid particle erosion, slightly reduces toughness. Used for screw components and pump sleeves. 46 C (Rockwell).

No. 6. Most generally useful cobalt alloy, 1% C, 28% Cr, 4% W, rem. Co. Excellent resistance to mechanical and chemical action over a wide temperature range. Widely used as valve seat material as it has high temperature hardness and high resistance to cavity erosion.

No. 12. Very similar to alloy No. 6 but with higher hardness. Used for cutting edge in carpet, plastics, paper and chemical industries. 40 C.

No. 21. Low carbon, Co–Cr alloy with molybdenum. Has high temperature strength, resistant to cavitation erosion, corrosion and galling. Used for hot die material, fluid valve seat facing.

No. 238. Co–Fe–Cr–Mo alloy, low carbon. Has excellent resistance to mechanical and thermal shock with good hot hardness.

No. 2006. Low cobalt alternative to No. 6 to which it is similar in galling and cavitation erosion properties, but has better abrasive wear and metal-to-metal sliding conditions. Used for hot working dies and shears.

No. 2012. Low cobalt alternative to No. 12. Has improved resistance to abrasion but is not as tough. Can be used for cutting edges.

It should be noted that in some cases the nickel-based alloys, such as C 0.15%, Cr 17%, Mo 17%, Fe. 6%, Rem. Ni, available in covered electrode form may be superior to the cobalt-base alloys.

See also the section on hard surfacing in Chapter 6.

Preparation. The surface to be Stellited must be thoroughly cleaned of all rust and scale and all sharp corners removed. A portable grinder is extremely useful for this purpose. In some cases, where the shape is complicated, pre-heating to prevent cracking is definitely an advantage. A slightly longer arc than usual is held, with the rod nearly perpendicular to the surface, as this helps to spread the Stellite more evenly. Care must be taken not to get the penetration too deep, otherwise the Stellite will become alloyed with the base metal and a poor deposit will result. Since Stellite has no ductility, cooling must be at an even rate throughout to avoid danger of cracking. The surface is finally ground to shape. Lathe centres, valve seats, rock drills and tool tips, cams, bucket lips, dies, punches, shear knives, valve tappet surfaces, thrust washers, stillson teeth, etc., are a few examples of the many applications of hard surfacing by this method.

Stainless steels

Note on BS 2926. This standard includes the chromium–nickel austenitic steels and chromium steels and uses a code by which weld metal content and coating can be identified.

The first figure is the % chromium content, the second figure the % nickel content and the third figure the % molybdenum content. The letter L indicates the low carbon version, Nb indicates stabilization with niobium, W indicates that there is tungsten present. R indicates a rutile coating, usually either d.c. or a.c., and B a basic coating, usually d.c. electrode +ve only. A suffix MP indicates a mild steel core.

For example: 19.12.3.Nb.R is a nobium-stabilized 19% Cr, 12% Ni, 3% Mo, rutile-coated electrode.

Stainless steel welding

Martensitic (chromium) stainless steels. These steels contain 12–16% chromium and harden when welded. The carbon content is usually up to 0.3% but may be more. They are used for cutlery and sharp-edged tools and for circumstances in which anti-corrosion properties are important. They harden when welded so that they should be pre-heated to 200–300 °C and then allowed to cool slowly. After welding they should then be post-heated to about 700 °C to remove brittleness. By using an austenitic electrode of the 25.12.Cr.Ni type with 3% tungsten (see table, pp. 66–7) the weld is more ductile and freer from cracking. If, however, the weld must match the parent plate as nearly as possible, electrodes of the chromium type must be chosen but are not very satisfactory. Use the austenitic rod where possible.

Ferritic stainless steel. Ferritic steels contain 16–30% Cr, do not harden when welded but suffer from grain growth when heated in the 950–1100 °C range, so that they are brittle at ordinary temperatures but may be tougher at red heat. Pre-heating to 200 °C should be carried out and the weld completed, followed by post-heating to 750 °C. For mildly corrosive conditions, electrodes of matching or higher chromium content should be used, e.g. 25.12.3.W, which give a tough deposit. If joint metal properties need not match the parent plate, electrodes of 26.20 can be used. It should be noted that the weld deposit of these electrodes contains nickel which is attacked by sulphurous atmosphere, so they are only suitable for mildly corrosive conditions.

Austenitic stainless steels. These largely non-magnetic steels form the largest and most important group from the fabrication point of view. They do not harden when welded because of their austenitic structure, the largest group being the 18% Cr 8% Ni type with other groups such as 25% Cr 20% Ni and 18% Cr 12% Ni. Other elements such as molybdenum are added to make them more acid-resistant, and they are available with rutile coatings suitable for d.c. electrode +ve or a.c. minimum OCV 55–80 V depending upon the electrode; or with basic coatings usually for d.c. only, electrode +ve, these being especially suitable for vertical and overhead positions. The table gives a selection of electrodes available.

Technique. These steels have a greater coefficient of expansion than mild steel so that the tendency to distort is greater. Tack welds should be placed at about twice the frequency as for mild steel but clamps and jigs are very suitable. The technique is similar to mild steel but the arc should be struck on a striking plate and avoid any stray arcing. Welding should be done with a short arc to avoid loss of alloying metals in the transference over the arc gap. Keep the electrode at about 80° to the line of travel using as low a current as possible to ensure penetration, since the lower conductivity of stainless steel reduces the width of the HAZ and excess current will cause burn through. After welding, clean with a brush having stainless steel wires – grinding is also satisfactory. Weaving should be about twice rod width and watch for any crack down the centre line of the weld indicating hot cracking due to restraint (see table on p. 66).

***Duplex stainless steels**.** These are austenitic/ferritic steels having 35–75 % ferrite, 22–25½ % Cr, 5–6½ % Ni and 3 % Mo. They are used in the petrochemical industry for both plant and pipe-lines. They can be welded using electrodes specially designed for the purpose (such as Nirex DPC®), containing 22 % Cr, 9 % Ni and 3 % Mo. The coating is generally of acid-rutile type, requiring no pre- or post-heat. Since these steels are not used in conditions above 300 °C, the interpass temperature must not exceed 200 °C. Nitric acid grade resistant steel, NAG (no. 329) is extremely resistant to nitric acid. Precipitation-hardening steels (e.g. FV520) are fully austenitic or martensitic. The latter type should be age-hardened after welding, but some austenitic types are unweldable due to hot cracking. Follow the manufacturer's instructions.

Most stainless steels can be welded by MMA (MMAW), MIG, GMAW, TIG (GTAW) as well as by laser and electron beam processes, the wire rod being matched to the parent plate.

---

* See also Appendix 12.

Examples of electrodes for stainless and heat resistant steels

Electrode a.c. OCV	d.c.	British Standard 2926 Cr	Ni	Mo		A.W.S. equivalent	Coating	UTS (N/mm^2)	Applications
80	+	19	9	–	Nb	E347–16	R	650	For the welding of plain or stabilized stainless steels.
60	+	19	12	3	L	E316L–16	R	610	For welding Mo bearing stainless steels of 316 or 316L type.
60	+	19	9	–	L	E308–16	R	610	For 18/8 Cr–Ni steels with low carbon deposit (0.03).
100	+	25	20	–		E310–15	B	640	For heat- and corrosion-resistant stainless steels of similar composition. Also low alloy and mild steel to stainless steel with low restraint – non-magnetic.
85	+	19	9	–		E308–16L	R	585	For unstabilized stainless steel of this composition. Also for welding and surfacing 14% Mn steels.
65	+	19	9	–	Nb	E347–16	R	650	Austenitic electrode for Nb stabilized stainless steels of the 19/9 class. Can be used all positional.
70	+	19	12	3	Nb	E316–16	R	640	Austenitic electrode for unstabilized stainless steels and similar types.
55	+	19	12	3	Nb	E318–16	R		For all plain or Ti or Nb stabilized stainless steels.
—	+	19	12	3	Nb	E318–15	B	640	Austenitic electrode for Ti or Nb stabilized stainless steels. Suitable for vertical and overhead welding, e.g. fixed pipe work, d.c. only.
60	+	19	13	4	L	E317	R	630	For AISI 317 steels. Carbon below 0.08%.
85	+	17	8	2		E16–8–2	B	630	Ferrite controlled with 2% Mn. Suitable for all similar types especially for long service at elevated temperatures, e.g. in steam pipes and gas turbines.
—	+	23	12	–		E309–15	B	587	Deposits of 22/12 Cr–Ni for welding clad steels of Cr–Ni of similar composition and for dissimilar metals.
55	+	23	12	2		E309–Mo–16	R	569	For clad steels with a deposit of AISI 309Mo type, also for dissimilar metals and stainless steel to mild steel.
70	+	29	10	–		E312–16	R	800	For dissimilar steels, nickel steels, hardenable steels or steel of the same composition. Ferrite about 35%.

Stainless steel butt weld preparation

	Plate thickness (mm)	Preparation		Suggested electrode diameter (mm)
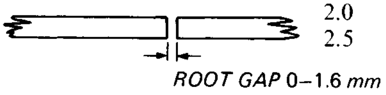 ROOT GAP 0–1.6 mm	1.2–1.6	square edge		1.6
	2.0	square edge		2.0
	2.5	square edge		2.5
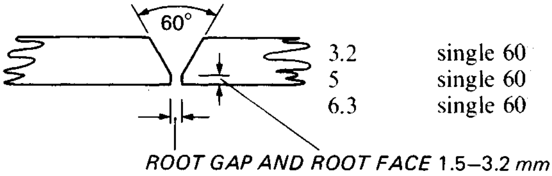 60° ROOT GAP AND ROOT FACE 1.5–3.2 mm	3.2	single 60°		2.5
	5	single 60°		3.2
	6.3	single 60°		3.2 and 4
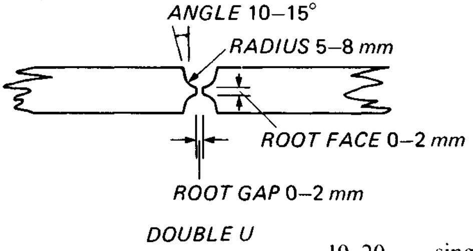 ANGLE 10–15° RADIUS 5–8 mm ROOT FACE 0–2 mm ROOT GAP 0–2 mm DOUBLE U	10–20	single 60°	or double 70°	4 and 5
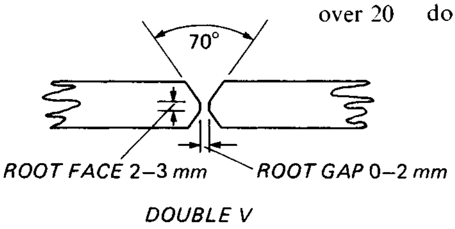 70° ROOT FACE 2–3 mm ROOT GAP 0–2 mm DOUBLE V	over 20	double 70°	or double U	4 and 5

Shaeffler constitutional diagram for stainless steel weld metal (Fig. 1.37)

The various alloying elements in a stainless steel are expressed in terms of (1) nickel, which tends to form austenite, and (2) chromium, which tends to form ferrite. Thus

Nickel equivalent $= \%\,\mathrm{Ni} + (30 \times \%\,\mathrm{C}) + (0.5 \times \%\,\mathrm{Mn})$

and

Chromium equivalent $= \%\,\mathrm{Cr} + (\%\,\mathrm{Mo}) + (1.5 \times \%\,\mathrm{Si}) + (0.5 \times \%\,\mathrm{Nb}$ or $\mathrm{Cb})$.

From the lines on the diagram we see that the low-alloy steels are hardenable because they contain the martensitic phase as welded. With increasing alloying elements, austenite and ferrite phases become more stable and the steel is no longer quench hardenable. Thus steels with high nickel, manganese and carbon become austenitic (shown in diagram) while those with high chromium and molybdenum are more ferritic (also shown). In the area $A + F$ the weld will contain both austenite and ferrite (duplex phase) and it can be seen how this leads to the designation of stainless steels as austenitic, martensitic and ferritic.

The shaded areas on the Shaeffler diagram indicate regions in which defects may under certain circumstances appear in stainless steel welds.

Example

Stainless steel BS 2926. 19.9.L.R. AWS equivalent E.308 L. Carbon 0.03 %, manganese 0.7 %, silicon 0.8 %, chromium 19.0 %, nickel 10.0 %.

$$\begin{aligned}\text{Nickel equivalent} &= \%\ \text{Ni} + (30 \times \%\ \text{C}) + (0.5 \times \%\ \text{Mn}).\\ &= 10 + (30 \times 0.03) + (0.5 \times 0.7)\\ &= 10 + 0.9 + 0.35\\ &= 11.25.\\ \text{Chromium equivalent} &= \%\ \text{Cr} + \%\ \text{Mo} + (1.5 \times \%\ \text{Si}) + (0.5 \times \%\ \text{Nb})\\ &= 19 + 0 + (1.5 \times 0.8) + (0.5 \times 0)\\ &= 19 + 0 + 1.2 + 0\\ &= 20.2\end{aligned}$$

Fig. 1.37. Constitution diagram for stainless steel welds (after Schaeffler) with approximate regions of defects and phase balance, depending on composition.

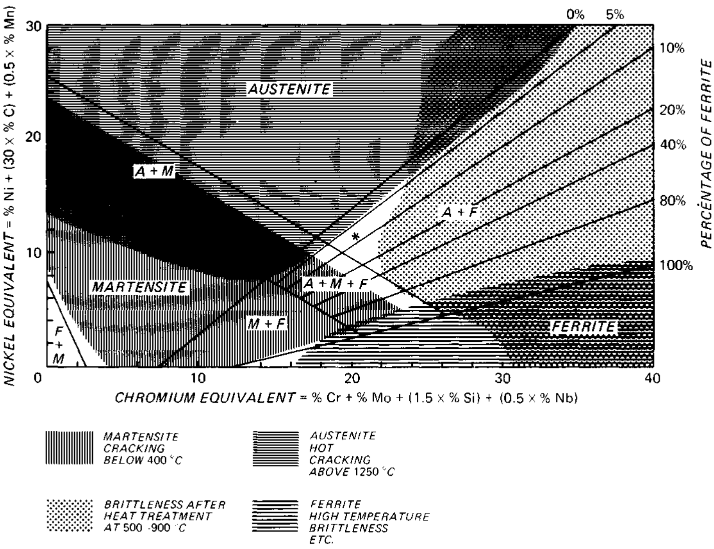

By plotting the two values on the diagram we get the point marked * in the clear part of the diagram, giving the main phases present and giving some indication of how it will behave during welding.

The welding of stainless steel to low alloy steel* and mild steel

The technique for welding is similar to that for stainless steel but dilution is the greatest problem with this work. If the electrode used is a mild or low alloy steel, the weld will have a low nickel and chromium enrichment and will be subject to cracking problems. Because of this the joint between the steels should be welded with an austenitic stainless steel electrode of much higher chromium–nickel content, say 25–12 or 23–12–2 or 29–10, which will result in weld metal which will have up to 18% Cr and 8% Ni after allowing for dilution. This now has a high resistance to hot cracking but the 25–20 type electrode must not be subject to high weld restraint. See Vol. 1, p. 93, for composition of core wire and coverings for steel electrodes.

Nickel and nickel alloys

The welding of nickel and its alloys is widely practised using similar techniques to those used for ferrous metals. The electrodes should be dried before use by heating to 120 °C or, if they are really damp, by heating for 1–2 hours at about 260 °C. Direct current from generator or rectifier with electrode positive gives the best results. A short arc should be held with the electrode making an angle of 20–30° to the vertical and when the arc is broken it should first be reduced in length as much as possible and held almost stationary, or the arc can be moved backwards over the weld already laid and gradually lengthened to break it. This reduces the crater effect and reduces the tendency to oxidation. The arc should be restruck by striking at the end of a run and moving quickly back over the crater, afterwards moving forward with a slight weave over the crater area, thus eliminating starting porosity.

Fig. 3.28 shows the most satisfactory joint preparation and it should be remembered that the molten metal of the nickel alloys is not as fluid as that of steel so that a wider V preparation is required with a smaller root face to obtain satisfactory penetration.

Each run of a multi-run weld should be deslagged by chipping and wire-brushed. Grinding should not be undertaken as it may lead to particles of slag being driven into the weld surface with consequent loss of corrosion resistance. Stray arcing should be avoided and minimum weaving performed because it results in poorer quality of weld deposit due to increased dilution.

---

* British and American classifications for low alloy steel electrodes are given on pp. 83–9.

Although it is preferable to make welds in the flat position, vertical welds can be made either upwards or downwards holding the electrode at right angles to the line of weld, using reduced current compared with similar flat conditions.

For fillet welds the electrode angle should roughly bisect the angle between the plates and be at right angles to the line of weld. If the plates are of unequal thickness the arc should be held more on to the thicker plate to obtain better fusion, and tilted fillets give equal leg length more easily.

The table gives a list of the alloys suitable for welding by this process and the type of electrode recommended.

Steels containing 9% nickel used for low temperature (– 196°C) are welded with basic-coated electrodes using d.c., such as 80/20 nickel/chromium or INCO-WELD B® alloy electrodes. See table.

Material	Recommended covered MMA electrodes
Nickel and nickel alloys 200, 201	Nickel alloy 141
MONEL® alloy 400	MONEL® alloy 190
MONEL® alloy K500	MONEL® alloy 190
Copper–nickel alloys	MONEL® alloy 187
INCONEL® alloy 600	INCONEL® alloy 182, INCO-WELD® alloys A, B
INCONEL® alloy 601	INCONEL® alloys 182, 112, INCO-WELD® alloy A
INCONEL® alloy 625	INCONEL® alloy 112
INCONEL® alloy 617	INCONEL® alloy 117
INCONEL® alloy 690	INCO-WELD® alloy A, INCONEL alloy 112
INCO® alloy C276	INCO® alloy C276
INCO® alloy HX	INCO® alloy HX, INCONEL® alloy 117
INCO® alloy G3	INCONEL® alloy 112
INCOLOY® alloys 800, 800H, 800HT	INCO-WELD® alloys A, B, INCONEL® alloys 112, 117
INCOLOY® alloy 825	INCOLOY® alloy 135
INCOLOY® alloy DS	INCO-WELD® alloys A, B
NIMONIC® alloy 75	INCONEL® alloys 182, 112, INCO-WELD® alloys A, B
BRIGHT® alloy series	INCO-WELD® alloys A, B
NILO® alloy series	Nickel alloy 141, INCO-WELD® alloys A, B
Cast irons	NI-ROD® 44, 55X, 99X
Mo bearing steels	INCONEL® alloy 112

INCO-WELD® alloy A electrode is suitable for welding (cryogenic) 9% nickel steels.

Clad steels

Stainless clad steel is a mild or low alloy steel backing faced with stainless steel such as 18% Cr 8% Ni or 18% Cr 10% Ni with or without Mo, Ti and Nb, or a martensitic 13% Cr steel, the thickness of the cladding being 10–20% of the total plate thickness.

Preparation and technique. The backing should be welded first and the mild steel root run should not come into contact with the cladding, so that preparation should be either with V preparation close butted with a deep root face, or the cladding should be cut away from the joint at the root (see Fig. 1.38). The clad side is then back grooved and the stainless side welded with an electrode of similar composition. Generally an austenitic stainless steel electrode of the 25% Cr 20% Ni type should be used for the root run on the clad side because of dilution effects, and at least two layers or more if possible should be laid on the clad side to prevent dilution effects affecting the corrosion-resistant properties. First runs should be made with low current values to reduce dilution effects. For martensitic 13% Cr cladding, pre-heating to 240°C is advisable followed by post-heating and using an austenitic stainless steel electrode such as 22% Cr 12% Ni 3% Mo with about 15% ferrite, which gives weld metal of approximately 18% Cr 8% Ni with about 6% ferrite. Welding beads not adjacent to the backing plate can be made with 18% Cr 12% Ni 3% Mo electrodes. If the heat input is kept as low as possible welding may be carried out without heat treatment, the HAZ being tempered by the heat from successive runs.

Nickel clad steel

Mild and low alloy steel can be clad with Nickel, MONEL® or INCONEL® alloys for corrosion resistance at lower cost compared to solid nickel base material and with an increase in thermal conductivity and greater strength, the thickness of cladding usually being not greater than 6 mm. When welding clad steels it is essential to ensure the continuity of

Fig. 1.38. Alternative methods of welding clad steel.

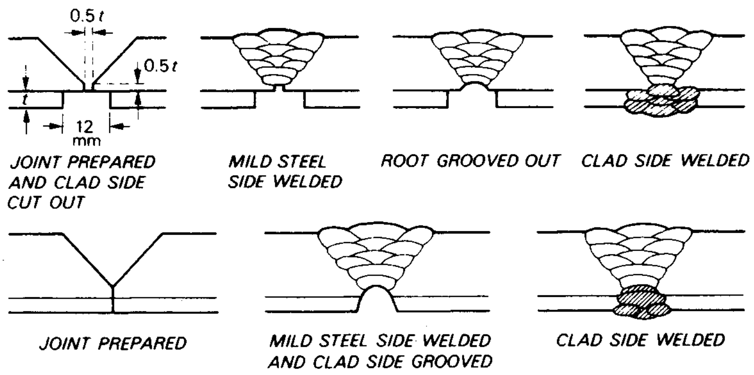

the cladding, and because of this butt joints are favoured. Dilution of the weld metal with iron occurs when welding the clad side, but the electrode alloy can accept this. First runs should be made with low current values to reduce the dilution.

Preparation of joints is similar to that for stainless clad steel. Recommended electrodes are given in the table on p. 71.

Welding of pipelines

The following brief account will indicate to the welder the chief methods used in the welding of pipelines for gas, oil, water, etc. The lengths of pipes to be welded are placed on rollers, so that they can easily be rotated. The lengths are then lined up and held by clamps and tack welded in four places around the circumference, as many lengths as can be handled conveniently, depending on the nature of the country, being tacked together to make a section. The tack welder is followed by the main squad of welders, and the pipes are rolled on the rollers by assistants using chain wrenches, so that the welding of the joint is entirely done in the flat position; hence the name *roll welding*.

After careful inspection, each welded section is lifted off the rollers by tractor-driven derricks and rested on timber baulks, either over or near the trench in which it is to be laid. The sections are then *bell hole* welded together. The operator welds right down the pipe, the top portion being done downhand, the sides vertical, and the underside as an overhead weld. Electrodes of 4 and 5 mm diameter are used in this type of weld.

Stove pipe welding (see also Appendix 6)

This type of welding has a different technique from conventional positional welding methods and has enabled steel pipelines to be laid across long distances at high rates. The vertically downward technique is used, welding from 12 o'clock to 6 o'clock in multiple runs.

Cellulose or cellulose–iron powder coated electrodes are used. These give a high burn-off rate, forceful arc and a light, fast-freezing slag and are very suitable for the vertical downward technique. The coating provides a gas shield which is less affected by wind than other electrodes but, generally, welding should not be carried out where the quality of the completed weld would be impaired by prevailing weather conditions, which may include rain, blowing sand and high winds. Weather protection equipment can be used wherever practicable.

Preparation is usually with 60° between weld faces, increased sometimes to 70° with a 1.5 mm root face and 1.5 mm root gap; internal alignment clamps are used. The stringer bead is forced with no weave into the root, then the hot pass with increased current fuses the sides and fills up any burn-through which may have occurred. Filler runs, stripper runs and capping run complete the welding (Fig. 1.39*a*). A diesel driven welding generator with tractor unit or mounted with the pipe laying plant is used with the electrode –ve (reverse polarity).

The welding of pipelines is usually performed by a team of welders; the larger the diameter of pipe, the greater the number of welders. In most cases the welder performs the same type of weld on each successive joint. The first team deposits the stringer bead and then moves on to the next joint. The second team carries out the hot pass, and they are followed by the third with the hot fill and the three teams follow along the pipeline from one joint to another. On a 42″ pipe there could be twelve welders spread over three adjacent joints all welding simultaneously. The first group is followed by further groups of fillers and cappers. It is important that the first three passes are deposited without allowing the previous one to cool.

Note. The wire brushes used are of strong stainless steel wire and power driven. As the wire is so strong there must be adequate protection of eyes and face by using a transparent thick face mask.

For clarity Fig. 1.39*a* and *b* shows the various passes and the clock face, while Fig. 1.40 shows pipe preparation including the use of backing rings and Fig. 1.41 shows preparation for a cut and shut bend and a gusset bend.

The CO_2 semi-automatic process is also used for pipeline welding. The supply unit is usually an engine-driven generator set and the technique used is similar to the stove pipe method, but it is important to obtain good line-up to avoid defects in the penetration and correct manipulation to avoid cold shuts. Fully automatic orbital pipe welders are also available using the TIG process and are used on pipes from 25 to 65 mm outside diameter.

For full details of the various methods of preparation and welding standards for pipes, the student should refer to BS 2633, class 1: *Steel pipework*, and BS 2971, class 2: *Steel pipelines*; BS 2910: *Radiographic examination of pipe joints*, and BS 938: *Metal arc welding of structural steel tubes*. See also Appendix 6, Low hydrogen electrode, downhill pipe welding.

Fig. 1.39. (*a*)

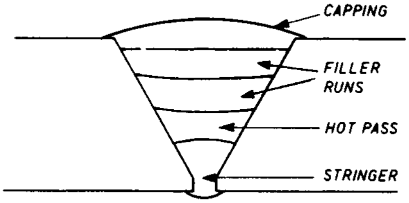

PREPARATION 60° INCLUDED ANGLE
ROOT GAP 1.6 mm ROOT FACE 1.6 mm.

For river crossings the pipe thickness is increased by 50–100% and the lengths are welded to the length required for the crossing. Each joint is further reinforced by a sleeve, and large clamps bolted to the line serve as anchor points. The line is then laid in a trench in the river bed.

For water lines from 1 to 1½ m diameter, bell and spigot joints are often used. On larger diameter pipes the usual joint is the reinforced butt. The

Fig. 1.39. (*b*) Clock face for reference.

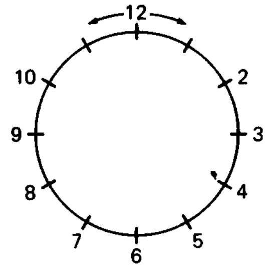
12
10
2
9
3
8
4
7
6
5

Fig. 1.39. (*c*) Electrode angles (hot fill).

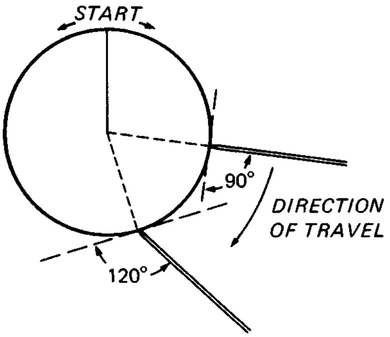
START
90°
DIRECTION OF TRAVEL
120°

Fig. 1.40

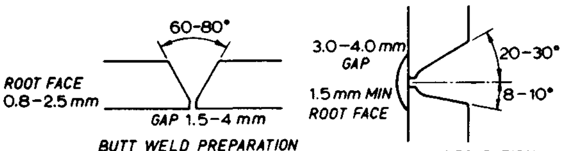
60-80°
ROOT FACE 0.8-2.5 mm
GAP 1.5-4 mm
3.0-4.0 mm GAP
1.5 mm MIN ROOT FACE
20-30°
8-10°

BUTT WELD PREPARATION UP TO 16 mm THICK

BUTT WELD PREPARATION VERTICAL PIPE

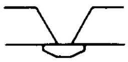

BACKING RING, WELDED IN WITH ROOT RUN

Stove pipe welding – example for pipes of 300–400 mm diameter

Pass	Electrode diam. (mm)	Current (A)	Technique
Stringer bead	3.25	100–130	Preheat to 50–100 °C. Strike electrode at 12 o'clock and keep at approximately 90° to line of travel. To control fine penetration, vary electrode angle to 6 o'clock for more to 12 o'clock for less. Use straight touch technique. If necessary someone can control current from generator according to welder's instructions. Arc should be visible on inside of pipe. Do not stop for small 'windows' or burn throughs, which will be eliminated by the hot pass. Grind bead when finished taking slightly more from sides until weld tracks are just visible.
Hot pass	4	180–220	This is deposited immediately after the stringer bead and melts out any inclusions, fills in 'windows' and *must* be deposited on a hot stringer bead which it anneals. It is a light drag weld with a backward and forward 'flicking' movement of the electrode held generally at 60°–70° to the line of travel but this can vary to 90° at 5 o'clock to 6 o'clock and 7 o'clock to 6 o'clock. Rotary wire brushing with a heavy duty power brush completes the pass.
Hot fill	4	160–190	This is deposited immediately after the hot pass, while the pipe is still warm. It cleans the weld profile and adds to the strength of the joint. The electrode is held at 60° to 70° to the line of travel at 12 o'clock, increasing to 120° at 6 o'clock (see Fig. 1.39*c*).
Fill and stripper passes	4	150–190	If the joint has cooled, preheat to 50 to 100 °C. All runs are single passes with no weaving and each run is well wire brushed. The electrode is held at 60°–70° to line of travel at 12 o'clock, changing gradually to 90°–100° at 6 o'clock. If the areas between 2 o'clock and 5 o'clock and between 10 o'clock and 7 o'clock are low, stripper passes are made until the weld is of a uniform level.
Capping run	4	130–190	The electrode is held at 60°–80° at 12 o'clock, to 120° at 6 o'clock. A fast side to side weave is permitted, covering not more than 2½ times the diameter of the rod. Lengthening the arc slightly between 5 o'clock and 7 o'clock avoids undercut. Manipulation between 4 o'clock and 8 o'clock can be by a slight 'flicking' technique. After completion the joint is power wire brushed and wrapped in a dry asbestos or similar cloth with waterproof backing and the joint is allowed to cool uniformly.

pipe is V'd on the inside and welded from the inside. A steel reinforcing band is then slipped over the joint and fillet welded in position.

The types of joint are illustrated in Fig. 1.42.

Fig. 1.41. Steel pipe preparation.

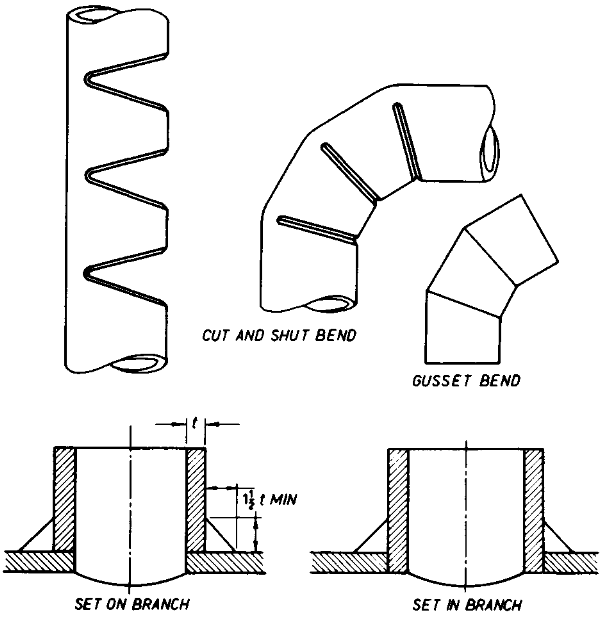
CUT AND SHUT BEND
GUSSET BEND
t
1½ t MIN
SET ON BRANCH
SET IN BRANCH

Fig. 1.42

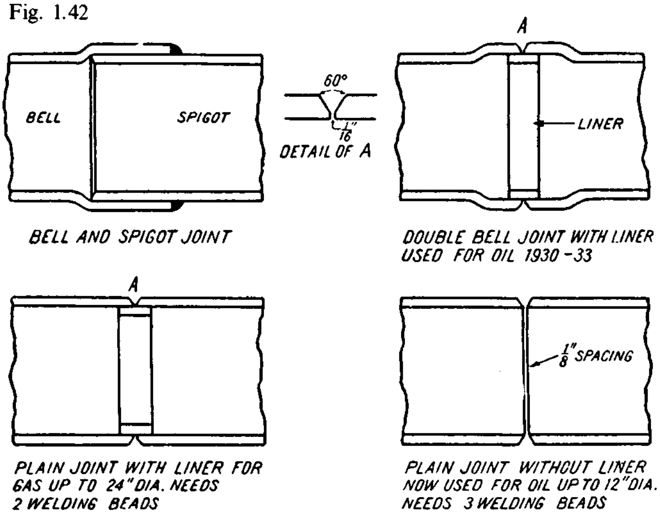
BELL
SPIGOT
BELL AND SPIGOT JOINT
60°
1/16″
DETAIL OF A
A
LINER
DOUBLE BELL JOINT WITH LINER USED FOR OIL 1930-33
A
PLAIN JOINT WITH LINER FOR GAS UP TO 24″DIA. NEEDS 2 WELDING BEADS
1/8″SPACING
PLAIN JOINT WITHOUT LINER NOW USED FOR OIL UP TO 12″DIA. NEEDS 3 WELDING BEADS

Welding of carbon and carbon–manganese and low-alloy steels

Steels with a carbon content up to about 0.25% carbon (mild steels) are easy to weld and fabricate because, due to their low carbon content, they do not harden by heat treatment so that the weld and HAZ do not have hardened zones even though there is quick cooling. As the carbon content of a steel increases it becomes more difficult to weld, because as the weld cools quickly owing to the quenching action of the adjacent cold mass of metal hardened zones are formed in the HAZ, resulting in brittleness and possible cracking if the joint is under restraint. Pre-heat can be used to overcome this tendency to cracking.

For low-alloy steels the core wire of electrodes may be based on a rimming steel with the alloying elements added to the covering. For stainless steel electrodes the core wire is generally of stainless steel, say 19% Cr/9% Ni, with additions to the covering of alloying elements to make the higher ranges of alloy steels, e.g. 19% Cr, 13% Ni, 1% Mo etc. (See also Vol. 1, p. 93.)

One or more alloying elements such as Ni, Cr, Mo, Mn, Si, V and Cu are added to steel to increase tensile, impact and shear strength, resistance to corrosion and heat and in some cases to give these improved properties either at high temperature or low (cryogenic) temperatures. The alloying elements do not impair the weldability as would an increase in carbon content, but the steels are susceptible to cracking, and it is this problem which has to be considered in more detail.

Cracking may appear as: (1) delayed cold cracking caused by the presence of H_2; (2) high-temperature liquation cracking (hot cracking); (3) solidification cracking; (4) lamellar tearing; hot cracking and cold cracking have been already considered in Vol. 1. Hot cracks usually appear down the centre of a weld, which is the last part to solidify, while cold cracks occur in the HAZ (Fig. 1.43). These latter may not occur until the weld is subject to stress in service, and since they are often below the metal surface they cannot be seen, so that the first indication of their presence is failure of the joint.

The factors which lead to cold cracking are:

(1) The composition of steel being welded.
(2) The presence of hydrogen.
(3) The rate of cooling of the welded joint.
(4) The degree of restraint (stress) on the joint.

(1) Composition of the steel and carbon equivalent (CE). The tendency to crack increases as the carbon and alloying element content increases, and since there is a great variety in the types of steel to be welded it is convenient to convert the varying amounts of alloying elements present in a given steel into terms of a simple equivalent carbon steel, thus giving an indication of the tendency to crack. This 'carbon equivalent' can be calculated from a formula such as the following:

$$\mathrm{CE} = \mathrm{C}\% + \frac{\mathrm{Mn}\%}{6} + \frac{\mathrm{Cr}\% + \mathrm{Mo}\% + \mathrm{V}\%}{5} + \frac{\mathrm{Ni}\% + \mathrm{Cu}\%}{15}$$

(which is a variation of the original formula of Dearden and O'Neill) so that to take a simple example, if steel has the following composition: C 0.13%, Mn 0.3%, Ni 0.5%, Cr 1.0%, application of the formula gives a CE of 0.41%.

(2) ***The presence of hydrogen*** in the welded zone greatly increases the tendency to crack, the amount of hydrogen present depending upon the type of electrode used and the moisture content of its coating. A rutile coating may have a high moisture content giving up to 30 ml of hydrogen in 100 g of weld metal. The hydrogen diffuses into the HAZ, and on cooling quickly a hard martensitic zone exists with a liability of cracks occurring. Even a small amount of hydrogen present can result in cracking in severely restrained joints. Basic (hydrogen controlled) electrodes correctly dried before use (p. 8) result in a very low hydrogen content. Austenitic stainless steel electrodes deposit weld metal in which the hydrogen is retained and does not diffuse into the HAZ. The weld has a relatively low yield strength and, when stressed, yields and reduces the restraint on the joint, so they are used, for example, to weld steels such as armour plate which may crack when welded with basic-coated mild steel electrodes. Gas-shielded processes using CO_2 or argon–CO_2 mixtures give welds of a very low hydrogen content.

Fig. 1.43

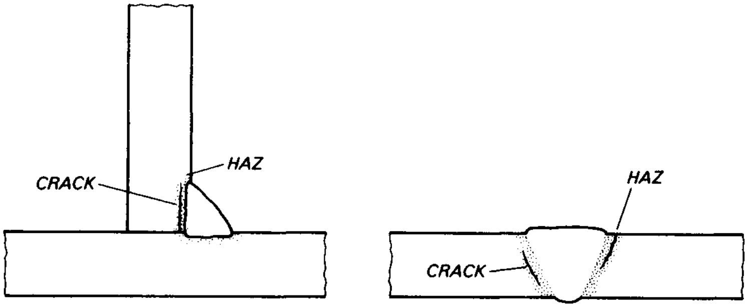

HAZ CRACKING

(3) Rate of cooling of the welded zone. The rate of cooling depends upon (1) the heat energy put into the joint and (2) the combined thickness of the metal forming the joint. Arc energy is measured in kilojoules per mm length of weld and can be found from the formula

$$\text{Arc energy (kJ/mm)} = \frac{\text{arc voltage} \times \text{welding current}}{\text{welding speed (mm/s)} \times 1000}.$$

The greater the heat input into the joint the slower the rate of cooling so that the use of a large-diameter electrode with high current reduces the quenching effect and thus the cracking tendency. Similarly, smaller-diameter electrodes with lower currents reduce the heat input and give a quicker cooling rate, increasing the tendency to crack due to the formation of hardened zones. Subsequent runs made immediately afterwards are not quenched as is the first run, but if the first or subsequent runs are allowed to cool, conditions then return to those of the first run. For this reason interpass temperature is often stipulated so as to ensure that the weld is not allowed to cool too much before the next run or pass is made. The use of large electrodes with high currents, however, does not necessarily give good impact properties at low temperatures. For cryogenic work it is essential to obtain the greatest possible refining of each layer of weld metal by using smaller-diameter electrodes with stringer or split-weave technique. As the 'combined thickness', that is the total thickness of the sections at the joint, increases, so the cooling rate increases, since there is increased section through which the heat can be conducted away from the joint. The cooling rate of a fillet joint is greater than that for a butt weld of the same section plate since the combined thickness is greater (Fig. 1.44).

(4) Restraint. When a joint is being welded the heat causes expansion, which is followed by rapid cooling. If the joint is part of a very

Fig. 1.44

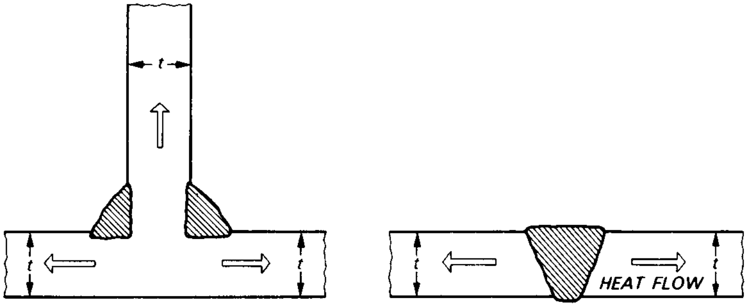

FILLET WELDS COMBINED THICKNESS = 3*t*   BUTT WELD COMBINED THICKNESS = 2*t* rigid structure the welded zone has to accommodate the stresses due to these effects and if the weld is not ductile enough, cracking may occur. The degree of restraint is a variable factor and is important when estimating the tendency to crack.

Controlled Thermal Severity (CTS) tests in which degrees of restraint are placed upon the joint to be welded and on which pre- and post-heat can be applied are used to establish the liability to crack. (See Reeve test, Volume 1.)

Hydrogen cracking can be avoided by (1) using basic hydrogen-controlled electrodes, correctly dried and (2) pre-heating.

The temperature of pre-heat depends upon (1) the CE of the steel, (2) the process used and in the MMA process, the type of electrode (rutile or basic), (3) the type of weld, whether butt or fillet and the run out length (*x* mm electrode giving *y* mm weld), (4) the combined thickness of the joint and (5) arc energy.

Reference tables are given in BS 5135 from which the pre-heat temperature can be ascertained from the above variables. Pre-heat temperatures may vary from 0 to 150 °C for carbon and carbon–manganese steels and be up to 300 °C for higher-carbon low-alloy steels containing chromium and molybdenum. The pre-heating temperature is specified as the temperature of the plate immediately before welding begins and measured for a distance of at least 75 mm on each side of the joint, preferably on the opposite face from that which was heated. The combined thickness is the sum of the plate thickness up to a distance of 75 mm from the joint. If the thickness increases greatly near the 75 mm zone higher combined thickness values should be used and it should be noted that if the whole unit being welded (or up to twice the distance given above) can be pre-heated, pre-heat temperatures can be reduced by about 50 °C. Austenitic electrodes can generally be used without pre-heat.

Steels used for cryogenic (low temperature) applications can be carbon–manganese types which have good impact properties down to −30 °C and should be welded with electrodes containing nickel; 3% nickel steels are used for temperatures down to −100 °C and are welded with matching electrodes whilst 9% nickel steels, used down to −196 °C (liquid nitrogen), are welded with nickel–chromium–iron electrodes since the 3 % nickel electrodes are subject to solidification cracking.

Creep-resistant steels usually contain chromium and molybdenum and occasionally vanadium and are welded with basic-coated low-alloy electrodes with similar chromium and molybdenum content. Two types of cracking are encountered: (1) transverse cracks in the weld metal, (2) HAZ cracking in the parent plate. Pre-heating and interpass temperatures of 200–300 °C with post-heat stress relief to about 700 °C is usually advisable.

Lamellar tearing

In large, highly stressed structures cracks may occur in the material of the parent plate or the HAZ of a joint, the cracks usually running parallel to the plate surface (Fig. 1.45). This is known as lamellar tearing, and it is the result of very severe restraint on the joint and poor ductility, due to the presence of non-metallic inclusions running parallel to the plate surface which are difficult to detect by the usual non-destructive tests. Certain types of joint such as T, cruciform and corner are more susceptible than others. Should lamellar tearing occur the joint design should be modified and tests made on the parent plate to indicate its sensitivity to tearing, while buttering of the surface may also help.

For structural steels in general, basic-coated electrodes are used and current electrode lists of the electrode makers should be consulted for the most up-to-date information.

When welding these steels the following points should be observed:

(*a*) When tack welds are used to position the work, as is the general practice, they should be well fused into the weld because, due to the rapid cooling and consequent hardening of the area around the tack weld, cracks may develop.

(*b*) Since the chilling effect is most marked on the first run, careful watch should be kept for any cracks which may develop. Any sealing run applied to the back of the joint should preferably be made either while the joint is still hot, or with a large electrode. The least possible number of runs should be used to fill up the V to minimize distortion.

(*c*) When austenitic rods are used to obtain a weld free from cracks, all the runs should be made with this type of rod and ordinary steel rods not used for subsequent runs. Plates thicker than 15 mm are preferably pre-

Fig. 1.45. Lamellar tearing.

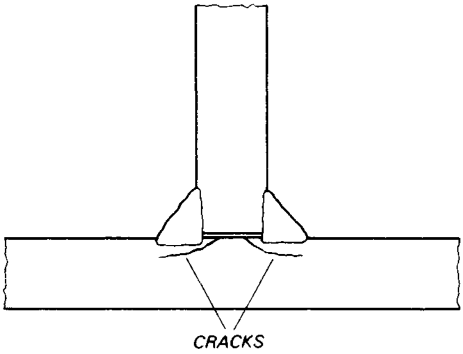

heated to 100–200 °C to avoid cracking. Cold work and interrupted welds should generally be pre-heated.

(*d*) The electrode or holder should not be struck momentarily or 'flashed' by design or accident on to the plate prior to welding, since the rapid cooling of the small crater produced leads to areas of intense hardness that may result in fatigue cracks developing.

(*e*) Hard spots in the parent plate may be softened by post-heat applied locally but this may reduce the endurance value of the joint.

Welding of steel castings

The use of steel castings in welded assemblies generally falls under two methods.

1. Welding steel castings together to form a complex casting that would be difficult to cast as a whole.
2. Castings and wrought material welded together to form a 'composite' fabrication.

The same rules regarding weldability apply as when welding wrought material. Castings may be in low-carbon-content mild steel, low-alloy steels, or high-alloy steels (such as austenitic manganese and chrome nickel stainless types) and for the alloy steels the same precautions must be taken against cracking as for the wrought form (q.v.). In general the stringer bead technique is used for thick butt welds since it reduces the cracking tendency and horizontal vertical fillets of greater than 8 mm leg length may be multi-run to avoid the tendency towards undercutting of the vertical plate.

Low-alloy steel classifications (British)

Abridged classification for low-alloy steel electrodes for manual metal arc welding. BS 2493 (1985)

Classification. The alloying elements are indicated by their chemical symbols: Mn manganese, Ni nickel, Cr chromium, Mo molybdenum, W tungsten (wolfram), V vanadium. Each symbol is preceded by a number indicating the approximate percentage of the first element given as the alloying addition, e.g. 2CrMoB shows that the alloy will contain 2.0–2.5 % chromium and that the other alloying element is molybdenum (see Table 8). The covering is indicated by B for basic, R for rutile and C for cellulosic. Almost all coverings in this standard are basic (hydrogen controlled) because of the resistance to cracking.

The letter H indicates a hydrogen-controlled electrode and is included for electrodes that deposit not more than 15 ml of diffusible hydrogen per 100 g of weld metal. The manufacturer indicates on the package the

Table 8. *Composition of deposited weld metal*

		Major alloying elements (%)												
Composition and covering code	Nearest AWS equivalent	Cr min	Cr max	Mo min	Mo max	Mn min	Mn max	Ni min	Ni max	Si max	W min	W max	C max	V max
Molybdenum steel electrodes														
MoB	E70XX-A1			0.4	0.7	0.75	1.2			0.6			0.1	
MoC	E70XX-A1			0.4	0.7	0.35	1.2			0.8			0.1	
MoR														
Chromium–molybdenum steel electrodes														
(for creep applications a maximum copper content of 0.15% is recommended)														
1CrMoLB	E80XX-B2L	1.0	1.8	0.4	0.7	0.5	1.2			0.5			0.05	
1CrMoB	E80XX-B2	1.0	1.5	0.4	0.7	0.5	1.2			0.5			0.1	
1CrMoR		1.0	1.5	0.4	0.7	0.35	1.2			0.3			0.1	
2CrMoLB	E90XX-B3L	2.0	2.5	0.9	1.2	0.5	1.2			0.5			0.05	
2CrMoB	E90XX-B3	2.0	2.5	0.9	1.2	0.5	1.2			0.5			0.1	
2CrMoR		2.0	2.5	0.9	1.2	0.35	1.2			0.3			0.1	
5CrMoB	E502	4.0	6.0	0.4	0.7	0.5	1.0			0.5			0.1	
7CrMoB	E7Cr	6.0	8.0	0.4	0.7	0.5	1.0			0.8			0.1	
9CrMoB	E505	8.0	10.0	0.9	1.2	0.5	1.0			0.8			0.1	
12CrMoB		11.0	12.5	0.8	1.2	0.3	1.0	0.3	0.8	0.8			0.23	0.3
12CrMoVB		11.0	13.0	0.8	1.2	0.3	1.0			0.8			0.23	0.5
12CrMoWVB		11.0	13.0	0.8	1.2	0.5	1.5	0.3	0.8	0.8	0.7	1.0	0.28	0.5
Nickel steel electrodes														
1NiLB	E70XX-C3L					0.3	1.1	0.8	1.2	0.6			0.07	
2NiLB	E70XX-C1L					0.3	1.1	2.0	2.75	0.6			0.05	
3NiLB	E70XX-C2L					0.5	1.1	2.8	3.75	0.6			0.05	
1NiB	E80XX-C3					0.5	1.8	0.8	1.1	0.8			0.1	
2NiB	E80XX-C1					0.5	1.2	2.0	2.75	0.8			0.1	
3NiB	E80XX-C2					0.5	1.2	2.8	3.75	0.8			0.1	
1NiC	E70XX-C3					0.5	1.2	0.8	1.2	0.8			0.15	
2NiC	E70XX-C2					0.5	1.2	2.0	2.75	0.8			0.15	
Manganese–molybdenum electrodes														
MnMoB	E90XX-D1			0.25	0.45	1.2	1.8			0.8			0.1	
2MnMoB	E100XX-D2			0.25	0.45	1.6	2.0			0.8			0.1	
High strength low-alloy steel electrodes														
MnNiB	E90XX-M					0.6	1.2	1.0	1.8	0.8			0.1	
NiMoB	E100XX-M			0.2	0.5	0.8	1.6	1.2	1.9	0.8			0.1	
1NiMoC	E100XX-G			0.2	0.5	0.5	1.2	1.2	1.9	0.8			0.18	
2NiMoB	E110XX-M			0.2	0.5	1.3	1.8	1.5	2.5	0.8			0.1	
2NiCrMoB	E120XX-M	0.7	1.5	0.2	0.0	1.3	2.2	1.5	2.5	0.8			0.1	

Table 9. *American Welding Society (AWS) abridged classification for low-alloy steel covered arc welding electrodes. A5.5–81 (reprinted 1985)*

The higher tensile strength electrodes, such as E8010, which do not have low hydrogen coverings are used for pipe welding and are usually matched to the pipe.

AWS classification	Type of covering	Capable of producing satisfactory welds in position shown	Type of current
E70 series: minimum tensile strength of deposited metal 70000 psi (480 MPa)			
E7010-X[a]	High cellulose sodium	F, V, OH, H	DCEP
E7011-X	High cellulose potassium	F, V, OH, H	AC or DCEP
E7015-X	Low hydrogen sodium	F, V, OH, H	DCEP
E7016-X	Low hydrogen potassium	F, V, OH, H	AC or DCEP
E7018-X	Iron powder low hydrogen	F, V, OH, H	AC or DCEP
E7020-X	High iron oxide	H-fillets	AC or DCEN
		F	AC or DC, either polarity
E7027-X	Iron powder iron oxide	H-fillets	AC or DCEN
		F	AC or DC, either polarity
E80 series: minimum tensile strength of deposited metal 80000 psi (550 MPa)			
E8010-X	High cellulose sodium	F, V, OH, H	DCEP
E8011-X	High cellulose potassium	F, V, OH, H	AC or DCEP
E8013-X	High titania potassium	F, V, OH, H	AC or DC, either polarity
E8015-X	Low hydrogen sodium	F, V, OH, H	DCEP
E8016-X	Low hydrogen potassium	F, V, OH, H	AC or DCEP
E8018-X	Iron powder low hydrogen	F, V, OH, H	AC or DCEP
E90 series: minimum tensile strength of deposited weld metal 90000 psi (620 MPa)			
E9010-X	High cellulose sodium	F, V, OH, H	DCEP
E9011-X	High cellulose potassium	F, V, OH, H	AC or DCEP
E9013-X	High titania potassium	F, V, OH, H	AC or DC, either polarity
E9015-X	High hydrogen sodium	F, V, OH, H	DCEP
E9016-X	Low hydrogen potassium	F, V, OH, H	AC or DCEP
E9018-X	Iron powder low hydrogen	F, V, OH, H	AC or DCEP
E100 series: minimum tensile strength of deposited metal 100000 psi (690 MPa)			
E10010-X	High cellulose sodium	F, V, OH, H	DCEP
E10011-X	High cellulose potassium	F, V, OH, H	AC or DCEP
E10013-X	High titania potassium	F, V, OH, H	AC or DC, either polarity
E10015-X	Low hydrogen sodium	F, V, OH, H	DCEP
E10016-X	Low hydrogen potassium	F, V, OH, H	AC or DCEP
E10018-X	Iron powder, low hydrogen	F, V, OH, H	AC or DCEP
E110 series: minimum tensile strength of deposited weld metal 110000 psi (760 MPa)			
E11015-X	Low hydrogen sodium	F, V, OH, H	DCEP
E11016-X	Low hydrogen potassium	F, V, OH, H	AC or DCEP
E11018-X	Iron powder, low hydrogen	F, V, OH, H	AC or DCEP
E120 series: minimum tensile strength of deposited weld metal 120000 psi (830 MPa)			
E12015-X	Low hydrogen sodium	F, V, OH, H	DCEP
E12016-X	Low hydrogen potassium	F, V, OH, H	AC or DCEP
E12018-X	Iron powder, low hydrogen	F, V, OH, H	AC or DCEP

[a] X stands for the various suffixes A1, B1, C1, etc., which denote the types of chemical composition of the electrodes (see Table 10).

Table 10. *Chemical composition of electrodes*

Suffix	Type of steel	Example	Suffix	Type of steel	Example
A1	Carbon–molybdenum	E7011-A1	C2L	Nickel steel	E7016-C2L
B1	Chromium–molybdenum	E8016-B1	C3[c]	Nickel steel	E8018-C3[c]
B2	Chromium–molybdenum	E8018-B2	NM[d]	Nickel-molybdenum	E8018-NM[d]
B2L	Chromium–molybdenum	E8015-B2L	D1	Manganese–molybdenum	E9018-D1
B3	Chromium–molybdenum	E9016-B3	D3	Manganese–molybdenum	E8016-D3
B3L	Chromium–molybdenum	E9018-B3L	D2	Manganese–molybdenum	E10015-D2
B4L	Chromium–molybdenum	E8015-B4L	G	All other low alloy steel electrodes	E7020-G
B5	Chromium–molybdenum	E8016-B5			
C1	Nickel steel	E8016-C1	M	Military specification	E11018-M
C1L	Nickel steel	E7018-C1L	W	Containing small % copper	E7018-W
C2	Nickel steel	E8018-C2			

L indicates a low carbon steel, generally less than 0.15%.

[c] These classifications conform to those covered by military specifications for similar compositions.

[d] Copper shall be 0.10% max and aluminium 0.05% max for E8018-NM electrodes.

necessary drying conditions for the electrodes that will give the necessary hydrogen levels per 100 g of weld metal (see pp. 8 and 12–13). Electrode diameters are 2, 2.5, 3.2, 4, 5, 6 and 6.3 mm.

Details of the full chemical composition, mechanical and radiographic tests, all weld metal tension test and impact (Charpy) test are given in BS 2493 (1985).

Example. 2NiB. A 2% nickel steel electrode (Ni 2.0–2.75%) with basic covering (AWS E80XX-C1)

Most low-alloy steel electrodes have a core wire of rimming steel (steel that has not been fully deoxidized before casting) with the alloying elements added to the covering. Under certain conditions, such as in pipe welding, the thickness of the covering makes positional welding difficult so the cellulosic covered electrodes have a core wire of the correct alloy steel e.g. chromium–molybdenum type as given in Table 8 (1CrMo, 2CrMo and 5CrMo).

Table giving some of the chief types of metal arc welding electrodes available for welding alloy steels

Electrode	Description and uses
Mild steel heavy duty	Rutile or basic coated for medium and heavy duty fabrications. The latter suitable for carbon and alloy steels and mild steel under restraint and for thick sections and root runs in thick plate.
High-tensile alloy steels	Austenitic rutile or basic coated for high-tensile steels including armour plate and joints between low-alloy and stainless steel and in conditions where pre-heat is not possible to avoid cracking.
Structural steels	Basic coated for high-strength structural steels 300–425 N/mm^2 tensile strength and for copper bearing weathering quality steels.
Notch ductile steels for low-temperature service	Basic coated for steels containing 2–5% Ni, 3% Ni and carbon–manganese steels. Also austenitic high nickel electrodes for 9% Ni steel for service to – 196°C and for dissimilar metal welding and for high Ni–Cr alloys for use at elevated temperatures.
Creep-resisting steels	Ferritic, basic coating for (1) 1.25% Cr, 0.5% Mo, (2) 2.5% Cr, 1.0% Mo, (3) 4–6% Cr, 0.5% Mo, steels with pre- and post-heat. Austenitic ferrite controlled for creep-resistant steels and for thick stainless steel sections requiring prolonged heat treatment after welding.
Heat- and corrosion-resisting steels	(1) Rutile or basic coating 19% Cr 9% Ni for extra low carbon stainless steels. (2) Basic coating Nb stabilized for plain or Ti or Nb stabilized 18/8 stainless steels and for a wide range of corrosive- and heat-resisting applications. Variations of these electrodes are for positional welding, smooth finish, and for high deposition rates. (3) Rutile or basic coating Mo bearing for 18/10 Mo steels. (4) Rutile or basic coating, low-carbon austenitic electtodes for low-carbon Mo bearing stainless steels and for welding mild to stainless steel. (5) Basic coating austenitic for heat-resisting 25% Cr, 12% Ni steels, for welding mild and low-alloy steels to stainless steel and for joints in stainless-clad mild steel. (6) Rutile coating austenitic for 23% Cr, 11% Ni heat-resistant steels containing tungsten. (7) Basic coating 25% Cr, 20% Ni (non-magnetic) for welding austenitic 25/20 steels and for mild and low-alloy steels to stainless steel under mild restraint.

Electrode	Description and uses
	(8) Basic coating 60% Ni, 15% Cr for high-nickel alloys of similar composition (incoloy DS, cronite, etc.) and for welding these alloys and stainless steel to mild steel in low restraint conditions.
	(9) A range of electrodes for high-nickel, monel, inconel and incoloy welding. These electrodes are also suitable for a wide range of dissimilar metal welding in these alloys.
Hard-surfacing and abrasion-resisting alloys	(1) High impact moderate abrasion-resistance rutile coating for rebuilding carbon steel rails, shafts, axles and machine parts subject to abrasive wear. 250 HV.
	(2) Medium impact medium abrasion-resistance rutile or basic coating for rebuilding tractor links and rollers, roller shafts, blades, punching die sets, reasonably machinable. The basic coated electrode gives maximum resistance to underbead cracking and is suitable for resurfacing low-alloy and hardenable steels. 360 HV.
	(3) High abrasion medium impact. (*a*) Rutile or basic coating for hard-surfacing bull-dozer blades, excavator teeth, bucket lips, etc. The basic coated electrode has greater resistance to underbead cracking and eliminates the need for a buffer layer on high-carbon steels. 650 HV. Unmachinable. (*b*) Tubular electrodes depositing chromium carbide. For worn carbon steel such as dredger buckets, excavator shovels etc. Matrix 560 HV. Carbide 1400 HV.
	(4) Severe abrasion moderate impact. (*a*) Tubular electrode depositing fused tungsten carbide giving highest resistance to abrasion with moderate impact resistance. Matrix 600 HV. Carbide 1800 HV. (*b*) High-alloy weld deposit for severe abrasion, suitable for use on sand and gravel excavators. Resistant to oxidation. Matrix 700 HV. Carbides 1400 HV.
12–14% Manganese steels	(1) Basic coating, 14% manganese steel (work-hardening) for 12–14% Mn steel parts – steel excavators and mining equipment. 240 HV.
	(2) Austenitic stainless steel weld deposit suitable for joining 12–14% Mn steels and for reinforcing and for butter layer for 12–14% Mn electrodes. 250 HV, work-hardened to 500 HV.
	(3) Tubular-type electrodes for hardfacing 12–14 Mn steel parts with high resistance to abrasion and heavy impact. Matrix 640 HV. Carbides 1400 HV.

2

Gas shielded metal arc welding[†]

Metal inert gas (MIG), metal active gas (MAG) including CO_2 and mixed gas processes[*]

The MIG semi-automatic and automatic processes are increasing in use and are displacing some of the more traditional oxy-acetylene and MMA uses.

For repair work on thin sheet as in the motor trade, semi-automatic MIG using argon–CO_2 mixtures has displaced the traditional oxy-acetylene methods because of the reduced heat input and narrower HAZ, thus reducing distortion. For larger fabrication work, mechanical handling equipment with automatic MIG welding heads has revolutionized the fabrication industry, while the advent or robots, which are program controlled and use a fully automated MIG welding head with self-contained wire feed, make less demands on the skilled welder.

Argon could not be used alone as a shielding gas for mild, low-alloy and stainless steel because of arc instability but now sophisticated gas mixtures of argon, helium, CO_2 and oxygen have greatly increased the use of the process. Much research is proceeding regarding the welding of stainless and 9% nickel steels by this method, using magnetic arc oscillation and various gas combinations to obtain positional welds of great reliability and freedom from defects.

The process has very many applications and should be studied by the student as one of the major processes of the future.

It is convenient to consider, under this heading, those applications which involve shielding the arc with argon, helium and carbon dioxide (CO_2) and mixtures of argon with oxygen and/or CO_2 and helium, since the power source and equipment are essentially similar except for the gas supply. These processes fall within the heading MIG/MAG.

---

[*] American designation: gas metal-arc welding (GMAW).

[†] See also BS 3571 MIG welding of aluminium and aluminium alloys.

With the tungsten inert gas shielded arc welding process, inclusions of tungsten become troublesome with currents above 300 A. The MIG process does not suffer from these disadvantages and larger welding currents giving greater deposition rates can be achieved. The process is suitable for welding aluminium, magnesium alloys, plain and low-alloy steels, stainless and heat-resistant steels, copper and bronze, the variation being filler wire and type of gas shielding the arc.

The consumable electrode of bare wire is carried on a spool and is fed to a manually operated or fully automatic gun through an outer flexible cable by motor-driven rollers of an adjustable speed, and rate of burn-off of the electrode wire must be balanced by the rate of wire feed. Wire feed rate determines the current used.

In addition, a shielding gas or gas mixture is fed to the gun together with welding current supply, cooling water flow and return (if the gun is water cooled) and a control cable from gun switch to control contractors. A d.c. power supply is required with the wire electrode connected to the positive pole (Fig. 2.1).

Fig. 2.1. Components of gas shielded metal arc welding process.

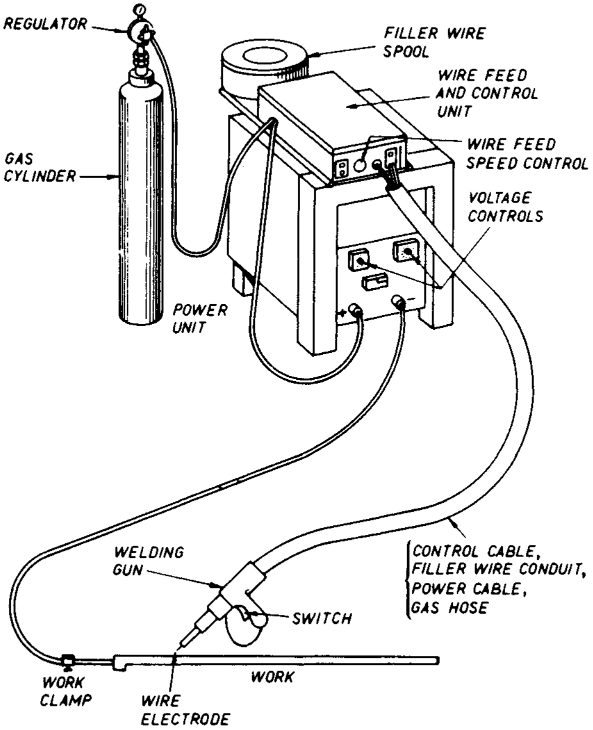
REGULATOR
FILLER WIRE SPOOL
WIRE FEED AND CONTROL UNIT
WIRE FEED SPEED CONTROL
GAS CYLINDER
VOLTAGE CONTROLS
POWER UNIT
+
-
WELDING GUN
SWITCH
CONTROL CABLE, FILLER WIRE CONDUIT, POWER CABLE, GAS HOSE
WORK CLAMP
WIRE ELECTRODE
WORK

Spray transfer

In manual metal arc welding, metal is transferred in globules or droplets from electrode to work. If the current is increased to the continuously fed, gas-shielded wire, the rate at which the droplets are projected across the arc increases and they become smaller in volume, the transfer occurring in the form of a fine spray.

The type of gas being used as a shield greatly affects the values of current at which spray transfer occurs. Much greater current densities are required with CO_2 than with argon to obtain the same droplet rate. The arc is not extinguished during the operation period so that arc energy output is high, rate of deposition of metal is high, penetration is deep and there is considerable dilution. If currents become excessively high, oxide may be entrapped in the weld metal, producing oxide enfoldment or puckering (in Al). For spray transfer therefore there is a high voltage drop across the arc (30–45 V) and a high current density in the wire electrode, making the process suitable for thicker sections, mostly in the flat position.

The high currents used produce strong magnetic fields and a very directional arc. With argon shielding the forces on the droplets are well balanced during transfer so that they move smoothly from wire to work with little spatter. With CO_2 shielding the forces on the droplet are less balanced so that the arc is less smooth and spatter tendency is greater (Fig. 2.2). The power source required for this type of transfer is of the

Fig. 2.2. Types of arc transfer. (*a*) Spray transfer: arc volts 27–45 V. Shielding gases: argon, argon–1 or 2% oxygen, argon–20% CO_2, argon–2% oxygen–5% CO_2. High current and deposition rate, used for flat welding of thicker sections. (*b*) Short-circuit or dip transfer: arc volts 15–22 V. Shielding gases as for spray transfer. Lower heat output and lower deposition rate than spray transfer. Minimizes distortion, low dilution. Used for thinner sections and positional welding of thicker sections.

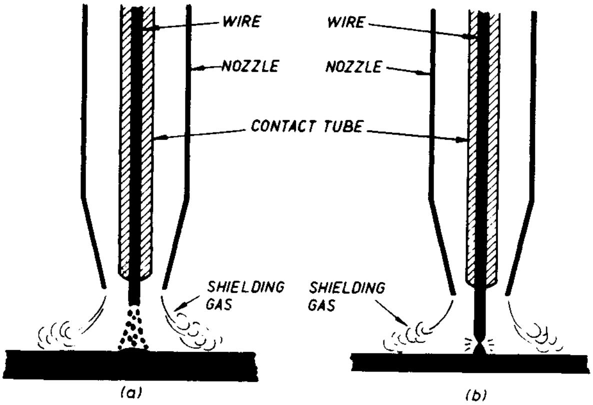

constant voltage type described later. Spray transfer is also termed free flight transfer.

Short circuit or dip transfer and controlled dip transfer

With lower arc volts and currents transfer takes place in globular form but with intermittent short-circuiting of the arc. The wire feed rate must just exceed the burn-off rate so that the intermittent short-circuiting will occur. When the wire touches the pool and short-circuits the arc there is a momentary rise of current, which must be sufficient to make the wire tip molten, a neck is then formed in it due to magnetic pinch effect and it melts off in the form of a droplet being sucked into the molten pool aided by surface tension. The arc is then re-established, gradually reducing in length as the wire feed rate gains on the burn-off until short-circuiting again occurs (Fig. 2.2). The power source must supply sufficient current on short-circuit to ensure melt-off or otherwise the wire will stick into the pool, and it must also be able to provide sufficient voltage immediately after short-circuit to establish the arc. The short-circuit frequency depends upon arc voltage and current, type of shielding gas, diameter of wire, and the power source characteristic. The heat output of this type of arc is much less than that of the spray transfer type and makes the process very suitable for the welding of thinner sections and for positional welding, in addition to multi-run thicker sections, and it gives much greater welding speed than manual arc on light gauge steel, for example. Dip transfer has the lowest weld metal dilution value of all the arc processes and this method is also pulsed.

Semi-short-circuiting arc

In between the spray transfer and dip transfer ranges is an intermediate range in which the frequency of droplet transfer is approaching that of spray yet at the same time short-circuiting is taking place, but is of very short duration. This semi-short-circuiting arc has certain applications, as for example the automatic welding of medium-thickness steel plate with CO_2 as the shielding gas.

Power supply d.c. and arc control*

There are two methods of automatic arc control:

(1) Constant voltage or potential, known as the self-adjusting arc (CV).

(2) Drooping characteristic or controlled arc (constant current, CC).

The former is more usually employed both on MIG and CO_2 welding plant though the latter may be used very occasionally with larger diameter wires and higher currents and with the flux cored welding process.

* See Fig. 2.26(*f*), p. 144 and Appendix 10.

Constant voltage d.c. supply

Power can be supplied from a welding generator with level characteristic or from a natural or forced draught cooled three-phase or one-phase transformer and rectifier arranged to give output voltages of approximately 14–50 V and ranges of current according to the output of the unit.

The voltage–current characteristic curve, which should be flat or level in a true constant voltage supply, is usually designed to have a slight droop as shown in Fig. 2.3*a*. Evidently this unit maintains an almost constant arc voltage irrespective of the current flowing. The wire feed motor has an adjustable speed control with which the wire feed speed must be pre-set for a given welding operation. Once pre-set the motor feeds the wire to the arc at constant speed. Supplies for the auxiliaries are generally at 110 V a.c. for the wire feed motor and 25 v a.c. for the torch switch circuit, but some units have a 120 V d.c. supply. For the arc to function correctly the rate of wire feed must be exactly balanced by the burn-off rate to keep the arc length constant. Suppose the normal arc length is that with voltage drop V_M

Fig. 2.3. Volt–ampere curves of constant voltage and drooping characteristic sources

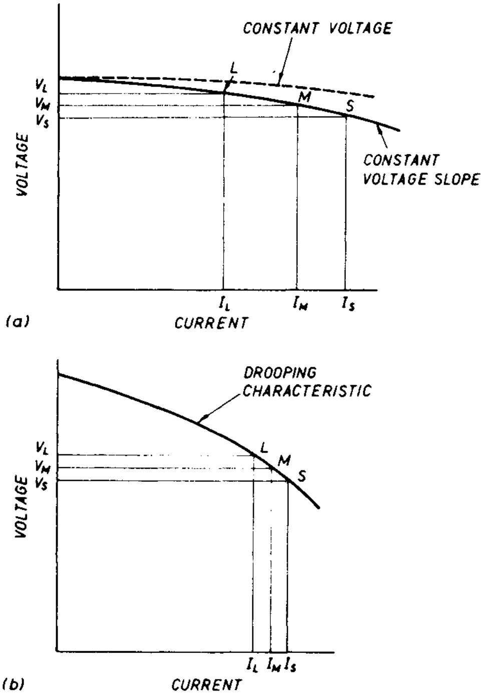

indicated in Fig. 2.3*a* at M and the current for this length is I_M amperes. If the arc shortens (manually or due to slight variation in motor speed) to S (the volts drop is now V_S) the current now increases to I_S, increasing the burn-off rate, and the arc is lengthened to M. Similarly if the arc lengthens to L, current decreases to I_L and burn-off rate decreases, and the arc shortens to M.

Evidently the gradient of slope of the output curve affects the welding characteristics and slope-controlled units are produced in which the gradient or steepness of the slope can be varied as required and the correct slope selected for given welding conditions.

Drooping characteristic d.c. supply

With this system the d.c. supply is obtained from a welding generator with a drooping characteristic or more usually from a transformer–rectifier unit. If a.c. is required it is supplied at the correct voltages from a transformer.

The characteristic curve of this type of supply (Fig. 2.3*b*) shows that the voltage falls considerably as the current increases, hence the name. If normal arc length M has volts drop V_M and if the arc length increases to L, the volts drop increases substantially to V_L. If the arc is shortened the volts drop falls to V_S while the current does not vary greatly, hence the name constant current which is often given to this type of supply. The variations in voltage due to changing arc length are fed through control gear to the wire feed motor, the speed of which is thus varied so as to keep a constant arc length, the motor speeding up as the arc lengthens and slowing down as the arc shortens. With this system, therefore, the welding current must be selected for given welding conditions and the control circuits are more complicated than those for the constant voltage method.

Power source–dip transfer (or short-circuit transfer)

In order to keep stable welding conditions with a low voltage arc (17–20 V) which is being rapidly short-circuited, the power source must have the right characteristics. If the short-circuit current is low the electrode will freeze to the plate when welding with low currents and voltages. If the short-circuit current is too high a hole may be formed in the plate or excessive spatter may occur due to scattering of the arc pool when the arc is re-established. The power supply must fulfil the following conditions:

(1) During short-circuit the current must increase enough to melt the wire tip but not so much that it causes spatter when the arc is re-established.

(2) The inductance in the circuit must store enough energy during short-circuit to help to start the arc again and assist in maintaining it during the decay of voltage and current. If an inductive reactor or choke connected in the arc circuit when the arc is short-circuited the current does not rise to a maximum immediately, so the effect of the choke is to limit the rate of rise of current, and the amount by which it limits it depends upon the inductance of the choke. This limitation is used to prevent spatter in CO_2 welding. When the current reaches its maximum value there is maximum energy stored in the magnetic field of the choke. When the droplet necks off in dip transfer and the arc is re-struck, the current is reduced and hence the magnetic field of the choke is reduced in strength, the reduction in energy being fed into the circuit helping to re-establish the arc. If the circuit is to have variable inductance so that the choke can be adjusted to given conditions the coil is usually tapped to a selector switch and by varying the number of turns in circuit the inductive effect is varied. The inductance can also be varied by using a variable air gap in the magnetic circuit of the choke.

To summarize: the voltage of a constant voltage power source remains substantially constant as the current increases. In the case of a welding power unit the voltage drop may be one or two volts per hundred amperes of welding current, and in these circumstances the short-circuit current will be high. This presents no problem with spray transfer where the current adjusts to arc length and thus prevents short-circuiting, but in the case of short-circuiting (dip) transfer, excessive short-circuit currents would cause much spatter.

The steeper the slope of the power unit volt–ampere characteristic curve the less the short-circuit current and the less the 'pinch effect' (which is the resultant inward magnetic force acting on the molten metal in transfer) so that spatter is reduced. Too much reduction of the short-circuit current, however, may lead to difficulty in arc initiation and stubbing. Power units are available having slope control so that the slope can be varied to suit welding conditions, and the control can be by tapped reactor or by infinitely variable reactor, the power factor of the latter being better than that of the former.

Three variables can thus be provided on the power unit – slope control, voltage and inductance. Machines with all three controls give the most accurate control of welding conditions but are more expensive and require more setting than those with only two variables, slope–voltage or voltage–inductance. In general, units with voltage–inductance control offer better characteristics for short-circuit transfer than those with slope–voltage control. For spray transfer conditions all types perform well, with the proviso that for aluminium welding the unit should have sufficient slope.

Transformer–rectifier power unit

Fig. 2.4*a* shows a transformer–rectifier power unit with output voltages by tapped transformer giving various wire speeds and feeds.

Fig. 2.4*b* shows an engine driven transformer–rectifier unit.

Fig. 2.5 shows a general method of connection for 3 phase input.

Wire feed and control cabinet

In the most popular system the wire is pushed through the guide tube to the gun and is particularly suitable for hard wires such as steel and stainless and heat-resistant steel. In the second method the motor is in the handle of the gun (Fig. 2.7*f*) and the wire is pulled from the reel through the guide tube to the gun. This allows a greater distance to be worked from power unit to welding position and is particularly useful for aluminium as the roll pressure is not so great but the reel capacity is smaller. In some cases a combination of these two methods, the push–pull type, is used applicable to both hard and soft wires but higher in initial cost.

In the push type, the most popular, the wires of diameter 0.8, 1.0, 1.2, 1.6, 2.4 mm hard and 1.2, 1.6, 2.4 soft, and 3.2 mm flux cored are supplied on reels of approximately 15, 25, 30 kg, etc., and are supplied for steel, low-alloy steel, creep-resistant and weathering steels, stainless heat-resistant steel, hard facing, aluminium, bronze and copper. The wire passes between motor-driven feed and pressure rollers which may have serrations or grooves to provide grip and which drive the wire at speeds between 2.5 and 15 m per min. The pressure on this drive can be varied and care must be taken that there is enough pressure to prevent slipping, in which case the arc lengthens and may burn back to the contact tube, and on the other hand that the pressure is not so great as to cause distortion of the wire or the flaking off of small metal particles with consequent increased wear on the guide tubes and possibility of jamming. Some units have a steel channel between feed and pressure rollers through which the wire passes and which helps prevent kinking. Other machines have a small removable magnet fixed after the wire drive to pick up such particles when ferrous wires are used. The flexible outer cables through which the wire is fed to the gun may have nylon liners for smoother feeding of fine wire sizes. The cabinet houses the wire drive motor and assembly, gas and cooling water valves and main

Fig. 2.4. (*a*) A 380–415 V 3-phase 50 Hz transformer–rectifier power source with integral wire feed system, 235 A. Other capacities available for single and 3-phase, e.g. 135 A, 165 A, 245 A, 285 A. CO_2- and Ar-rich gases. Continuous, spot and stitch modes. Capability: mild steel up to 10 mm thick, stainless steel, and aluminium. Spot 0.5–1.0 mm thick, output current 40–230 A at 15–25 arc volts. At 60% duty cycle 155 A. Weight 75 kg. (See Appendix 10, p. 429, for a photograph of a similar unit.)

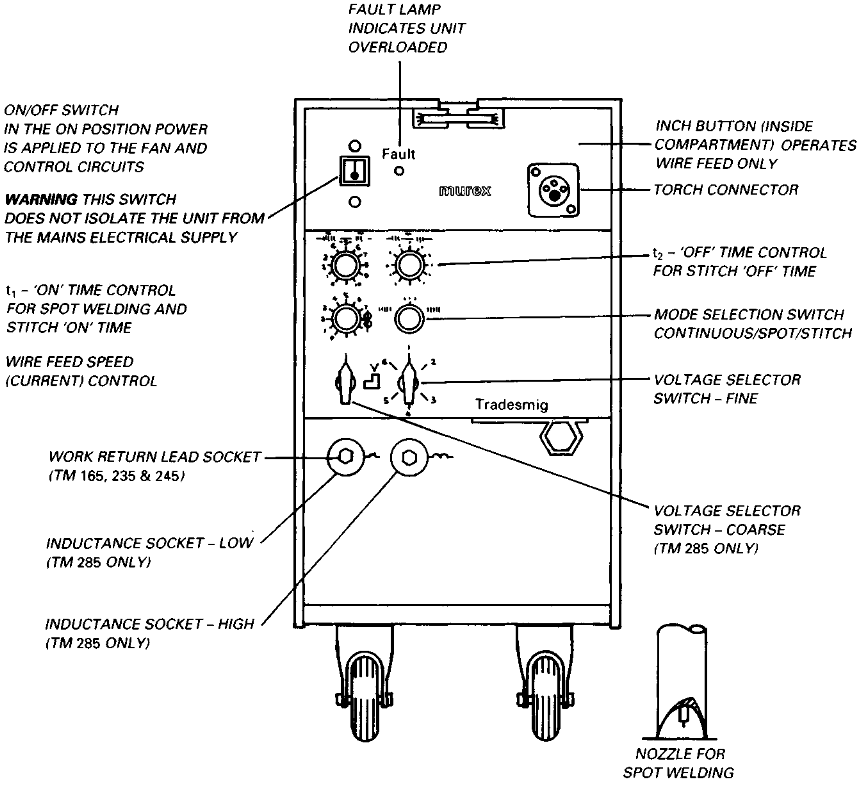
FAULT LAMP INDICATES UNIT OVERLOADED
ON/OFF SWITCH IN THE ON POSITION POWER IS APPLIED TO THE FAN AND CONTROL CIRCUITS
WARNING THIS SWITCH DOES NOT ISOLATE THE UNIT FROM THE MAINS ELECTRICAL SUPPLY
t_1 - 'ON' TIME CONTROL FOR SPOT WELDING AND STITCH 'ON' TIME
WIRE FEED SPEED (CURRENT) CONTROL
WORK RETURN LEAD SOCKET (TM 165, 235 & 245)
INDUCTANCE SOCKET - LOW (TM 285 ONLY)
INDUCTANCE SOCKET - HIGH (TM 285 ONLY)
Fault
murex
Tradesmig
INCH BUTTON (INSIDE COMPARTMENT) OPERATES WIRE FEED ONLY
TORCH CONNECTOR
t_2 - 'OFF' TIME CONTROL FOR STITCH 'OFF' TIME
MODE SELECTION SWITCH CONTINUOUS/SPOT/STITCH
VOLTAGE SELECTOR SWITCH - FINE
VOLTAGE SELECTOR SWITCH - COARSE (TM 285 ONLY)
NOZZLE FOR SPOT WELDING

For spot welding the nozzle is changed for one that has a cut-away end, so that the contact tip is distanced from the work. Upon pressing the torch switch the wire moves down to the work to make the spot weld. The size of spot weld depends upon the thickness of plate being spot welded and is determined by the time setting on switch 1.

Fig. 2.4. (*a*) *cont.*

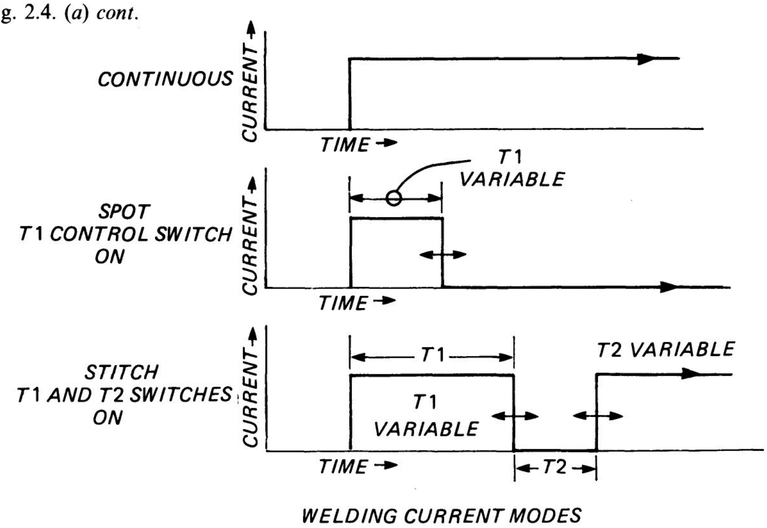
CONTINUOUS
CURRENT
TIME
SPOT
T1 CONTROL SWITCH
ON
CURRENT
T1
VARIABLE
TIME
STITCH
T1 AND T2 SWITCHES
ON
CURRENT
T1
T1
VARIABLE
T2 VARIABLE
TIME
T2
WELDING CURRENT MODES

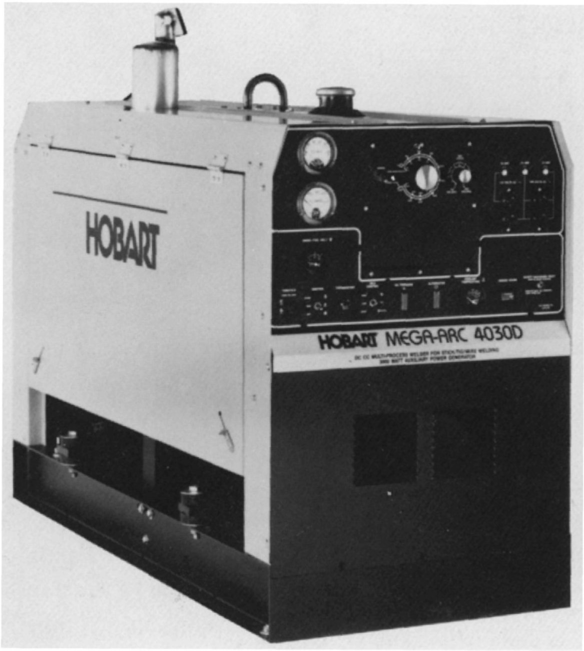
HOBART
HOBART MEGA-ARC 4030D

Fig. 2.4. (*b*) Mega-arc diesel engine driven, 1800 rpm. Water-cooled Perkins engine. Models for 400 A, 450 A and 500 A at 60 % duty cycle. High and low current ranges, OCV 75. Auxiliary power 115/230 receptacle. For d.c. MMA, TIG, flux core with or without gas, solid wire spray transfer, and some dip transfer applications using 75 % Ar, 25 % CO_2. Arc force control. Optional remote control.

current contactor (controlled from the switch on the gun or automatic head) and the wire speed control, which is also the current control, is pre-set manually or automatically (Figs. 2.6*a* and *b*). There are also gas purging and inching switches, and in some units regenerative braking on the wire drive motor prevents over-run at the end of a weld.

Fig. 2.5. 3-phase rectification (capacitors not shown).

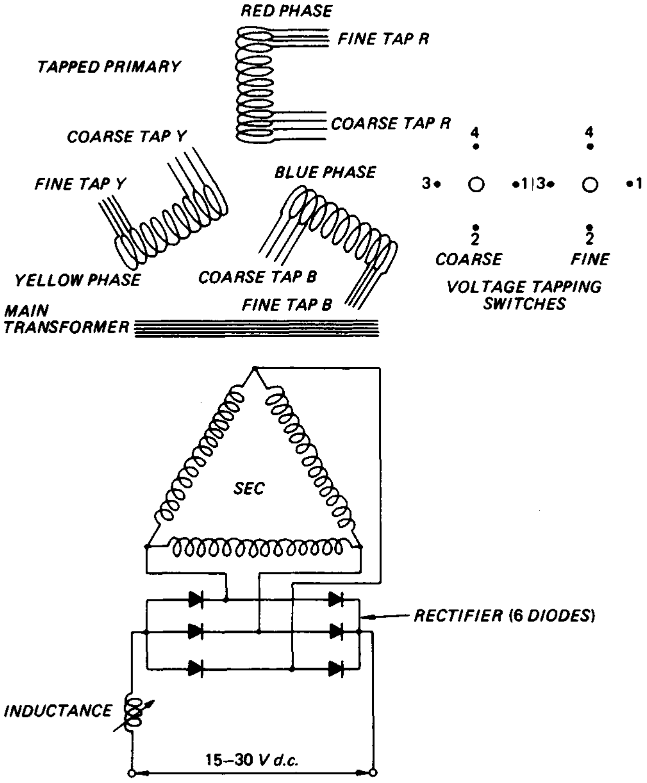
RED PHASE
FINE TAP R
TAPPED PRIMARY
COARSE TAP R
COARSE TAP Y
FINE TAP Y
BLUE PHASE
YELLOW PHASE
COARSE TAP B
FINE TAP B
MAIN TRANSFORMER
4
3
1
2
COARSE
FINE
VOLTAGE TAPPING SWITCHES
SEC
RECTIFIER (6 DIODES)
INDUCTANCE
15–30 V d.c.

Sets are now also available with programmable power sources. Using known quantities such as amperes, seconds, metres per minute feed, the welding program is divided into a chosen number of sections and the welding parameters as indicated previously are used to program the computer which controls the welding source. The program can be stored in the computer memory of up to say 50 numbered welding programs or it can be stored on a separate magnetic data card for external storage or use on another unit. By pressing the correct numbers on the keyboard of the unit any programs can be selected and the chosen program begins, controlling welding current, shielding and backing gas, gas pre-flow, wire feed speed, arc length, pulsed welding current and slope control, etc. All safety controls are fitted and changes in the welding program can be made without affecting other data.

Fig. 2.6. (*a*) Transmig 400 inverter for MIG/MAG, MMA (stick), d.c. TIG, with pulsed MIG. Input 380–415 V, 50–60 Hz, 3-phase with wire feed. Remote control optional. 60–70 OCV MIG/MAG, and 65–75 OCV with MMA and TIG. At 60% duty cycle 315 A with 50 arc volts. At 40% duty cycle 400 A at 36 V. Stable arc, low spatter, HF start. Good positional capability. Remote control, pulse unit and water adapter. Weight 48 kg, the inverter giving weight reduction. (See Appendix 10, p. 430, for photograph, and Chapter 5 of Volume 1 for complete description of the inverter.)

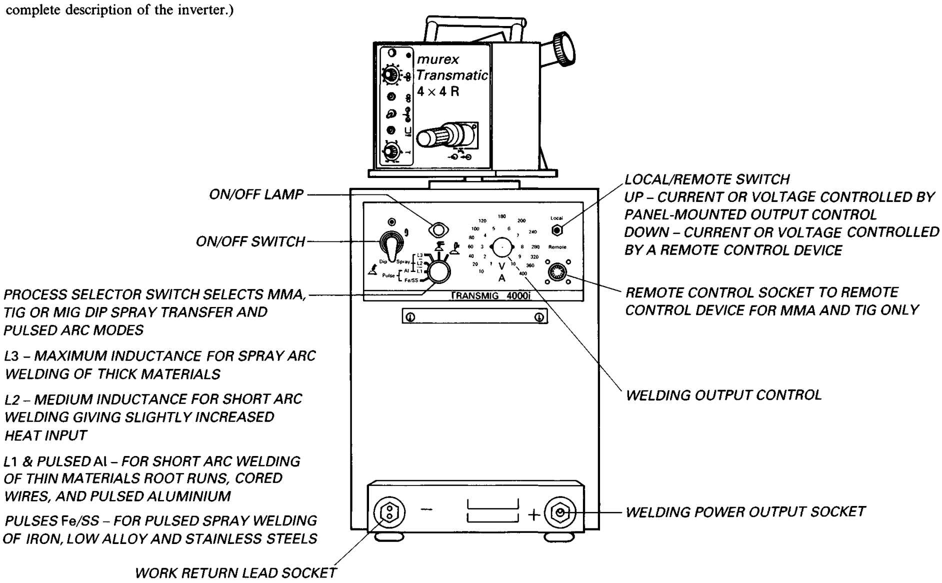

ON/OFF LAMP

ON/OFF SWITCH

PROCESS SELECTOR SWITCH SELECTS MMA, TIG OR MIG DIP SPRAY TRANSFER AND PULSED ARC MODES

L3 – MAXIMUM INDUCTANCE FOR SPRAY ARC WELDING OF THICK MATERIALS

L2 – MEDIUM INDUCTANCE FOR SHORT ARC WELDING GIVING SLIGHTLY INCREASED HEAT INPUT

L1 & PULSED Al – FOR SHORT ARC WELDING OF THIN MATERIALS ROOT RUNS, CORED WIRES, AND PULSED ALUMINIUM

PULSES Fe/SS – FOR PULSED SPRAY WELDING OF IRON, LOW ALLOY AND STAINLESS STEELS

WORK RETURN LEAD SOCKET

LOCAL/REMOTE SWITCH
UP – CURRENT OR VOLTAGE CONTROLLED BY PANEL-MOUNTED OUTPUT CONTROL
DOWN – CURRENT OR VOLTAGE CONTROLLED BY A REMOTE CONTROL DEVICE

REMOTE CONTROL SOCKET TO REMOTE CONTROL DEVICE FOR MMA AND TIG ONLY

WELDING OUTPUT CONTROL

WELDING POWER OUTPUT SOCKET

Fig. 2.6. (*b*) Tectron or Maxtron 400, CC/CV inverter power source. Input 380–415 V. Output 400 A at 36 V (60% duty cycle) OCV 80. Solid state power and control. Suitable for MMA (GMAW), MIG (SMAW) with pulsed spray for MIG/MAG, TIG (GTAW) and pulsed arc with 0.25–25 pulses/sec; scratch start and flux cored arc welding (FCAW). Can be used with all types of wire feeders. Remote control when required. Line voltage compensation 10%. Two-position process switch for selected welding method. Voltage range in CV mode 10–36. Amperage in CC mode 3–400. Weight 51 kg. This unit is termed Tectron in Europe. See Appendix 10 for illustration of unit, c/w wire feed etc.

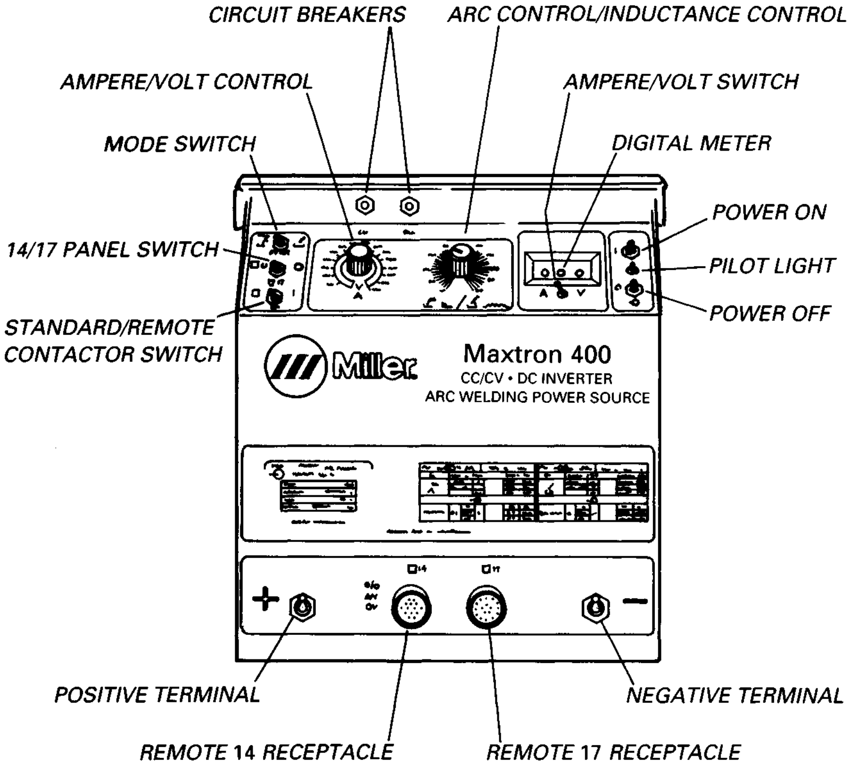
CIRCUIT BREAKERS
ARC CONTROL/INDUCTANCE CONTROL
AMPERE/VOLT CONTROL
AMPERE/VOLT SWITCH
MODE SWITCH
DIGITAL METER
POWER ON
14/17 PANEL SWITCH
PILOT LIGHT
POWER OFF
STANDARD/REMOTE CONTACTOR SWITCH
Miller
Maxtron 400
CC/CV • DC INVERTER
ARC WELDING POWER SOURCE
A
V
+
−
POSITIVE TERMINAL
NEGATIVE TERMINAL
REMOTE 14 RECEPTACLE
REMOTE 17 RECEPTACLE

The inverter

The inverter power source, is used in the latest modern technology for MIG, MAG, TIG etc. The a.c. from the mains is rectified to d.c., passed into the inverter, converted to high frequency square wave d.c., transformed to HF a.c., passed into a transformer still at HF and brought down to a voltage suitable for welding. It then passes through a reactor which smooths the current and a final rectifier gives d.c. for the arc. A final electronic switching arrangement enables the welder to select either d.c. or a.c. at high frequency and the resulting arc is easy to strike, smooth and stable. This method has many advantages compared with the conventional power unit. The subject is dealt with in Volume 1.

Torch (Fig. 2.7*a–g*). To the welding torch of either gooseneck or pistol type or an automatic head are connected the following supplies:

(1) Flexible cable through which the wire electrode is fed.
(2) Tubes carrying shielding gas and cooling water flow and return (if water cooled).
(3) Cable carrying the main current and control wire cable.

A centrally placed and replaceable contact tube or tip screws into the torch head and is chosen to be a sliding fit in the diameter of wire being used. Contact from power unit to welding wire is made at the contact tube, which must be removed and cleaned at intervals and replaced as required. A metal shield or nozzle surrounds the wire emerging from the tube through which the shielding gas flows and surrounds arc and molten pool. Air-cooled torches are used up to 400 A and water-cooled up to 600 A. In the latter type the cooling water return flows around the copper cable carrying the welding current and thus this cable can be of smaller cross-sectional area and thus lighter and more flexible. A water-cooled fuse in the circuit ensures that the water-cooling flow must be in operation before welding commences and thus protects the circuit.

The wire feed can be contained in the head (Fig. 2.7*f*) only when the feed is by 'pull' or there can be a 'pull' torch unit, a 'push' unit, and a control box. These units handle wires of 0.8, 1.0, 1.2 or 1.6 mm diameter in soft aluminium or 0.8, 1.0 and 1.2 mm in hard steel wire, the unit being rated at 300 A. Pre-weld and post-weld gas flows are operated by the gas trigger and operate automatically. They operate from the standard MIG or CO_2 rectifier units and are suitable for mixed gas or CO_2 shielding, the very compact bulk greatly adding to their usefulness since they have a very large working radius.

Fig. 2.7. (*a*) Straight-necked, air-cooled torch, variable in length by 38 mm. Current: at 75% duty cycle, 300 A with argon-rich gases, 500 A with CO_2; at 50% duty cycle, 350 and 550 A respectively. Wires: hard and soft, 1.2 and 1.6 mm, flux cored 1.6, 2.0 and 2.4 mm. (*b*) air-cooled torch, 45° neck angle. Current: at 75% duty cycle 300 A with argon-rich gases, 350 A with CO_2; at 50% duty cycle 350 A and 400 A respectively. Wires: hard 0.8, 1.0, 1.2 and 1.6 mm; soft 1.2 and 1.6 mm, flux cored, 1.6 mm. (*c*) Air-cooled torch, 60° neck angle. Current: at 75% duty cycle 300 A with argon-rich gases, 350 A with CO_2; at 50% duty cycle 350 and 400 A respectively. Wires: hard 0.8, 1.0, 1.2 and 1.6 mm.

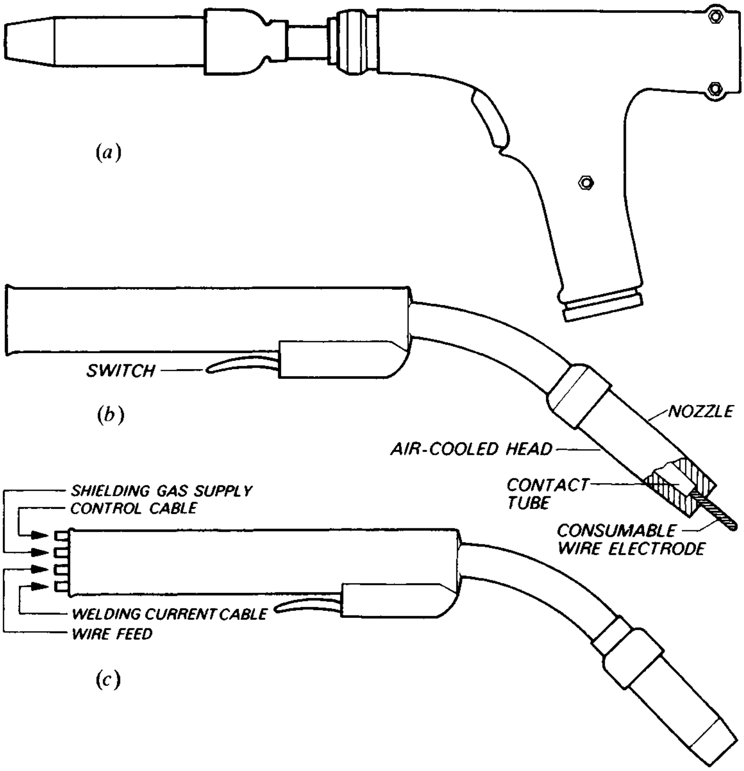
(*a*)
SWITCH
(*b*)
NOZZLE
AIR-COOLED HEAD
CONTACT TUBE
CONSUMABLE WIRE ELECTRODE
SHIELDING GAS SUPPLY
CONTROL CABLE
WELDING CURRENT CABLE
WIRE FEED
(*c*)

Fig. 2.7. (*d*) MIG welding guns: 600 A, 400 A, 300 A.

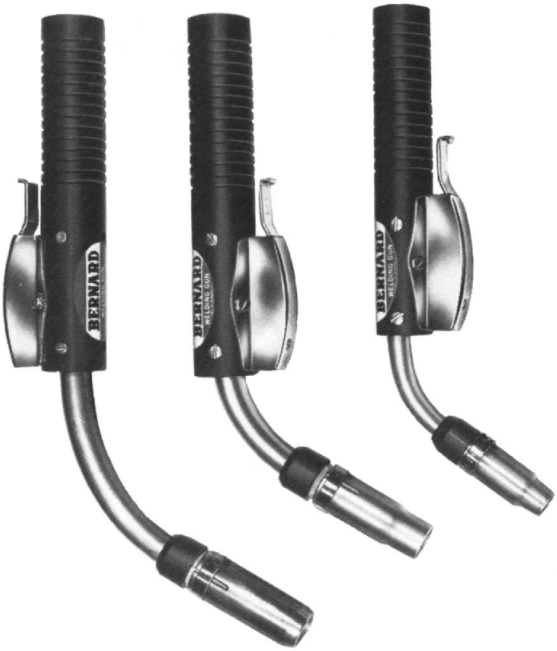
BERNARD
BERNARD
WELDING GUN
BERNARD
WELDING GUN

Fig. 2.7. (*e*) 400 A gun showing 'quick disconnect'.

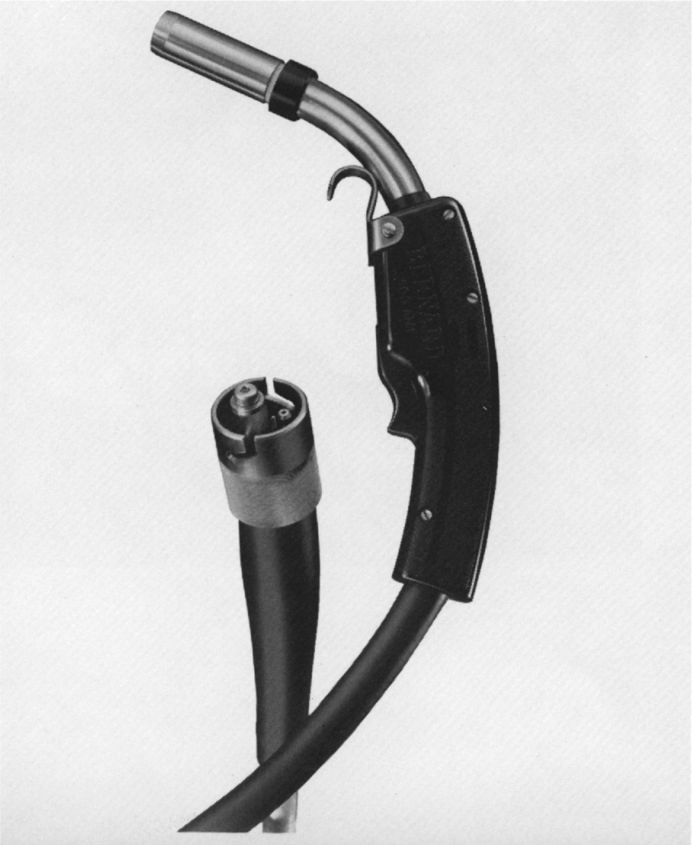

Fig. 2.7. (*f*) Welding torch with air motor for wire feed drive built into the handle. Speed of motor is set by means of a knob which controls an air valve. Guns can be fitted with swan necks rotatable through 360°. Available in sizes up to 400 A, these latter have a heat shield fitted to protect the hand. A fume extraction unit is also fitted.

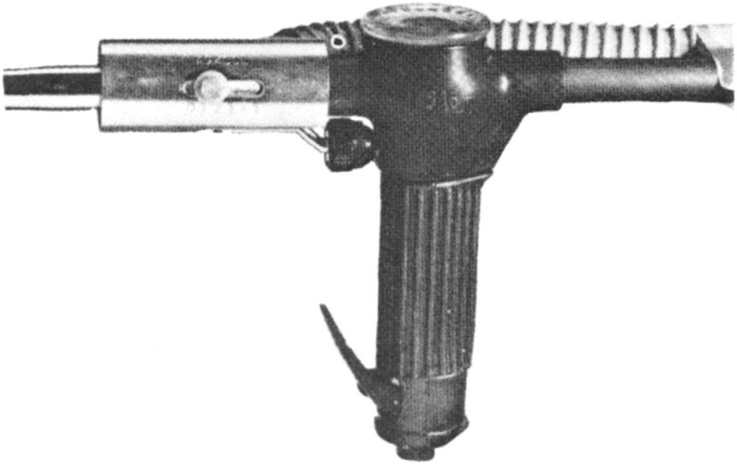

Fig. 2.7. (*g*) Torch with fume extraction unit.

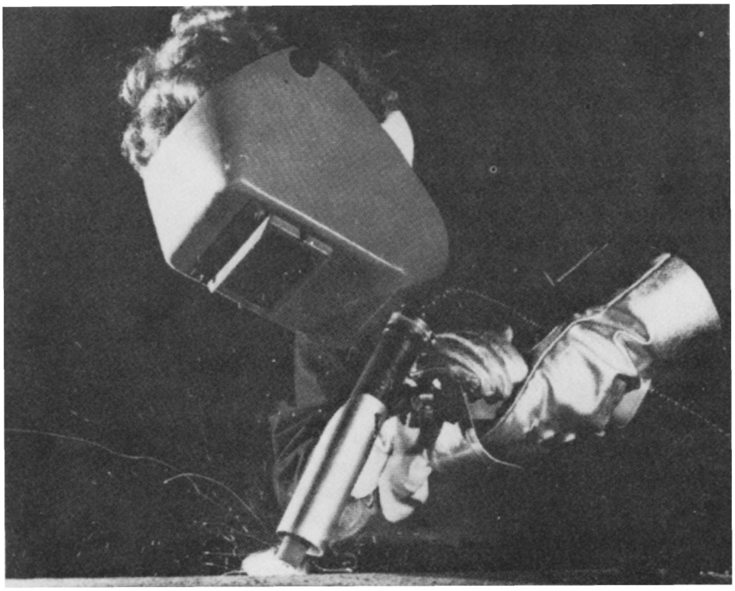

Fig. 2.7. (*h*) Welding torch or gun.

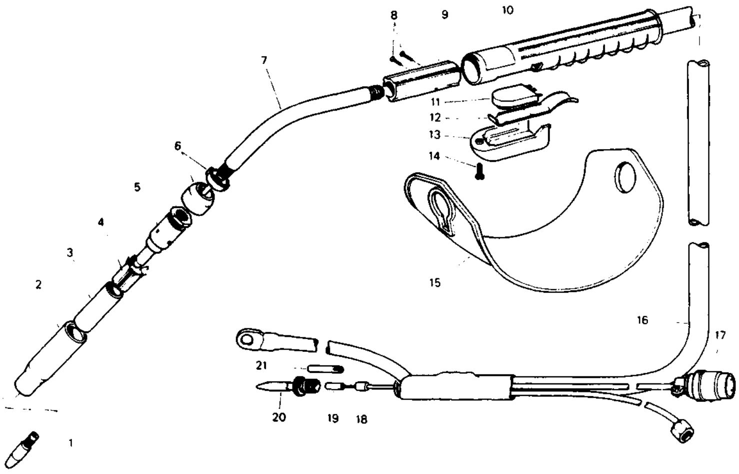

1. Contact tip for 1.0, 1.2 and 1.6 mm hard and soft wire; 1.6, 2.0 and 2.4 mm tubular wire.
2. Nozzle.
3. Nozzle insulator.
4. Nozzle spring clip.
5. Torch head.
6. Head insulator and ○ clip.
7. Neck.
8. Self-tapping screw.
9. Spider.
10. Handle mounting.
11. Microswitch.
12. Switch lever.
13. Switch housing.
14. Screw.
15. Heat shield.
16. Integrated cable.
17. Plug: 7 pin.
18. Basic liner.
19. Liner; 1.2, 1.6 mm soft wire, 1.2, 1.6 mm hard wire, 1.6, 2.0 mm tubular wire.
20. Outlet guide; 1.2, 1.6 mm soft wire, 1.2, 1.6 mm hard wire, 1.6, 2.0, 2.4 mm tubular wire.
21. Collet (for soft wire outlet guides).

Gases

Since CO_2 and oxygen are not inert gases, the title metallic inert gas is not true when either of these gases is mixed with argon or CO_2 is used on its own. The title *metallic active gas* (MAG) is sometimes used in these cases.

Argon, Ar. Commercial grade purity argon (99.996%) is obtained by fractional distillation of liquid air from the atmosphere, in which it is present to about 1% (0.932%) by volume. It is supplied in blue-painted cylinders containing 1.7, 2.0, 8.48 and 9.66 m^3 of gas at 175 or 200 bar maximum pressure or from bulk supply. It is used as a shielding gas because it is chemically inert and forms no compounds.

Carbon dioxide, CO_2. This is produced as a by-product of industrial processes such as the manufacture of ammonia, from the burning of fuels in an oxygen-rich atmosphere or from the fermentation processes in alcohol production, and is supplied in black-painted steel cylinders containing up to 35 kg of liquid CO_2.* To avoid increase of water vapour above the limit of 0.015% in the gas as the cylinder is emptied, a dip tube or syphon is fitted so that the liquid CO_2 is drawn from the cylinder, producing little fall in temperature. Fig. 2.8 shows a CO_2 vaporizer or heater with pressure reducing regulator and ball type flowmeter. A cartridge type electric heating element at 110 V is in direct contact with the liquid CO_2 to vaporize it. The 150 W version gives a flow rate of 21 litres/min while the 200 W version gives 28 litres/min. A neon warning light connected via a thermostat indicates when the element is heating and is extinguished when the heater has warmed up sufficiently. The syphon cylinder has a white longitudinal stripe down the black cylinder while the non-syphon cylinder, used when the volume of CO_2 to be taken from it is less than 15 litres/min, is all black. Manifold cylinders can be fed into a single vaporizer, and if the supply is in a bulk storage tank, this is fed into an evaporator, and then fed to the welding points at correct pressure as with bulk argon and oxygen supplies.

Helium, He. Atomic weight 4, liquifying point −268 °C, is thus much lighter than argon, atomic weight 40. It is present in the atmosphere in extremely small quantities but is found in association with natural gas in

---

* The cylinder pressure depends upon the temperature being approximately 33 bar at 0 °C and 50 bar at 15 °C.

Texas, Oklahoma, Kansas, Alberta, etc., which are the main sources of supply. There is little associated with North Sea gas. Being lighter it requires a greater flow rate than argon and has a 'hotter' arc which may give rise to certain health hazards. Cylinder identifying colour is brown.

Oxygen, O_2. Atomic weight 16, boiling point − 183°C. Obtained by distillation from liquid air. Used in connection with the inert gas in small percentages to assist in wetting and stabilizing the arc. Wetting is discussed on pp. 318–19 under the heading Surface tension.

Application of gases

(1) Argon. Although argon is a very suitable shielding gas for the non-ferrous metals and alloys, if it is used for the welding of steel there exists an unstable negative pole in the work-piece (the wire being positive) which produces an uneven weld profile. Mixtures of argon and oxygen are selected to give optimum welding conditions for various metals.

Fig. 2.8

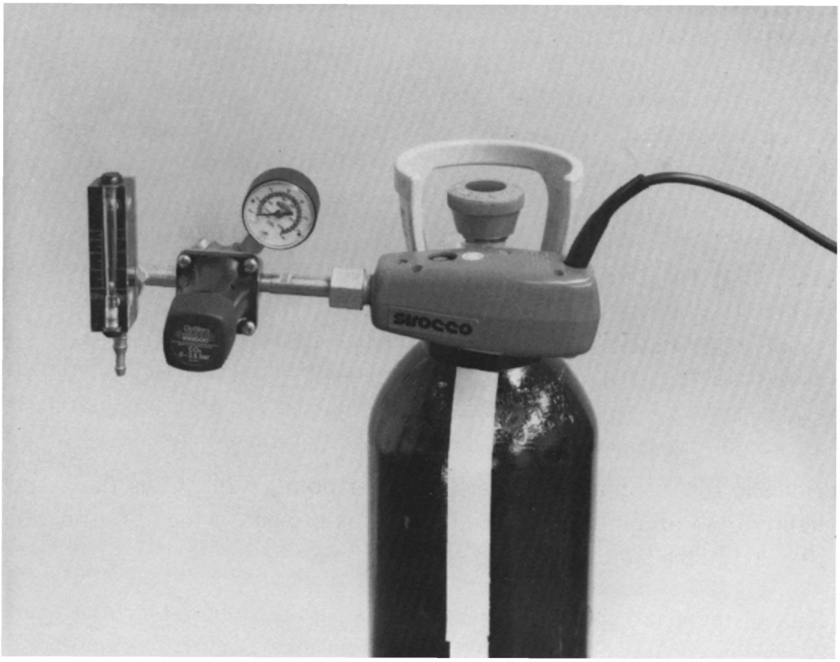

(2) Argon + 1% or 2% oxygen. The addition of oxygen as a small percentage to argon gives higher arc temperatures and the oxygen acts as a wetting agent to the molten pool, making it more fluid and stabilizing the arc. It reduces surface tension and produces good fusion and penetration. Argon-rich mixtures such as argon and up to 25% CO_2 with the addition of oxygen make the dip transfer process applicable to positional steel work and to thin sheet. Argon + 1–2% O_2 is used for stainless steel and helium, argon, CO_2 can be used on thin sections similarly and has good weld profile. The argon + 1% mixture is used for stainless steels with spray and pulse methods. Dip transfer is better with argon +2% oxygen as long as the increased oxidation can be tolerated. The addition of 5% hydrogen is the maximum for titanium-stabilized stainless steel and larger amounts than this increase porosity.

(3) Helium is nearly always found in mixed gases. Because of the greater arc temperature, mixing it with argon, oxygen or CO_2 controls the pool temperature, increases wetting and stabilizes the arc. The higher the helium content the higher the arc voltage and the greater the heat output. It is used in gas mixtures for aluminium, nickel, cupro-nickels, etc., and is particularly applicable to stainless steels with the helium–argon–oxygen or CO_2 mixtures. Helium–argon mixtures are also used for the welding of 9% nickel steels.

(4) Carbon dioxide. Pure CO_2 is the cheapest of the shielding gases and can be used as a shield for welding steel up to 0.4% C and low-alloy steel. Because there is some dissociation of the CO_2 in the arc resulting in carbon monoxide and oxygen being formed, the filler wire is triple deoxidized to prevent porosity, and this adds somewhat to its cost and results in some small areas of slag being present in the finished weld. The droplet rate is less than that with pure argon, the arc voltage drop is higher, and the threshold value for spray transfer much higher than with argon. The forces on the droplets being transferred across the arc are less balanced than with argon–oxygen so that the arc is not as smooth and there is some spatter, the arc conditions being more critical than with argon–oxygen.

Using spray transfer there is a high rate of metal deposition and low-hydrogen properties of the weld metal. The dip transfer process is especially suited for positional work. The process has not displaced the submerged arc and electro-slag methods for welding thick steel sections, but complements them and competes in some fields with the manual metal arc process using iron powder electrodes. It offers the most competitive method for repetitive welding operations, and the use of flux cored wire greatly increases the scope of the CO_2 process. Thickness up to 75 mm can be welded in steel using fully automatic heads.

With stainless steel, because of the loss of stabilizers (titanium and niobium) in the CO_2 shielded arc, there is some carbon pick-up resulting in some precipitation of chromium carbide along the grain boundaries and increased carbon content of the weld, reducing the corrosion resistance. Multi-pass runs result in further reduction in corrosion resistance, but with stabilized filler wire and dip transfer on thinner sections satisfactory single-pass welds can be made very economically. Non-toxic, non-flammable paste is available to reduce spatter problems.

(5) Argon + 5% CO_2, Argon + 20% CO_2. The addition of CO_2 to argon for the welding of steel improves the 'wetting' action, reduces surface tension and makes the molten pool more fluid. Both mixtures give excellent results with dip and spray transfer, but the 20% mixture gives poor results with pulse while the 5% mixture gives much better results. The mixtures are more expensive than pure CO_2 but give a smoother, less critical arc with reduced spatter and a flatter weld profile, especially in fillets (Fig. 2.9*b*). The current required for spray transfer is less than for similar conditions with CO_2 and the 5% mixture is suitable for single-run welding of stainless steel with the exception of those with extra low carbon content (L), below 0.06%. Both mixtures contain a small amount of oxygen.

If the CO_2 content is increased above 25% the mixture behaves more and more like pure CO_2. If argon and CO_2 are on bulk supply, mixers enable the percentage to be varied as required. Figs 2.9*a* and 2.9*b* show penetration beads and weld profiles with differing gas mixtures.

Fig. 2.9. (*a*) Stainless steel weld bead profiles: dip transfer.

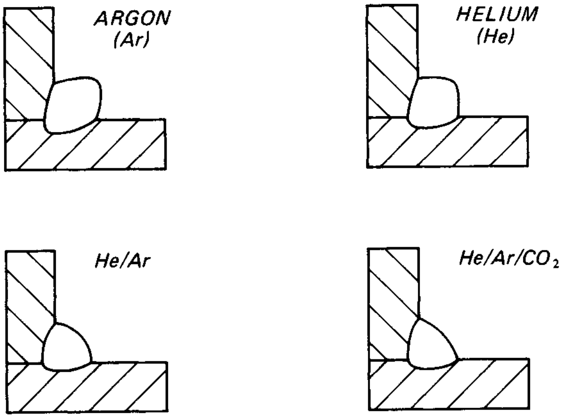

Filler wires*

Note. BS 2901 gives specifications for *Filler rods and wires for inert gas welding.* Part 1, *Ferritic steels;* Part 2, *Austenitic stainless steels;* Part 3, *Copper and copper alloys;* Part 4, *Aluminium and aluminium alloys and magnesium alloys;* Part 5, *Nickel and nickel alloys.*

Filler wires are supplied on convenient reels of 300 mm or more diameter and of varying capacities with wire diameters of 0.6, 0.8, 1.0, 1.2, 1.6, and 2.4 mm. The bare steel wire is usually copper coated to improve conductivity, reduce friction at high feed speeds and minimize corrosion while in stock. Manganese and silicon are used as deoxidizers in many cases but triple deoxidized wire using aluminium, titanium and zirconium gives high-quality welds and is especially suitable for use with CO_2. The following are examples of available steel wires and can be used with argon–5% CO_2 argon–20% CO_2 and CO_2, the designation being to BS 2901 Part 1 (1970).

A 18. General purpose mild steel used for mild and certain low alloy steels. Analysis: 0.12% C; 0.9–1.6% Mn; 0.7–1.2% Si. 0.04% max S and P.

Fig. 2.9. (*b*) Penetration beads: spray transfer.

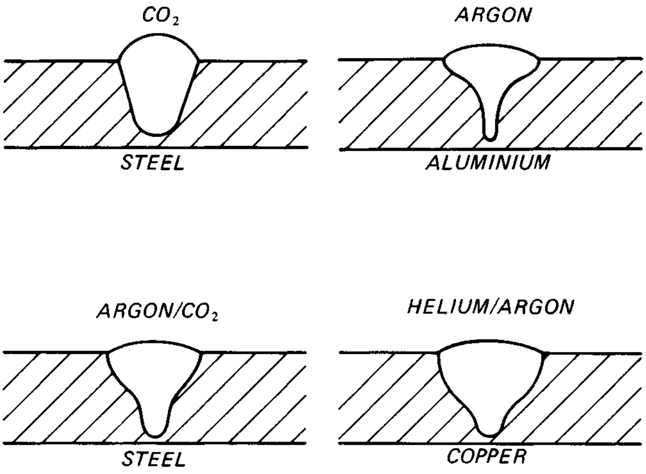

---

* American classification of mild steel welding wires AWS A5 18–69. *Example* E 70S–3 E = electrode, S = bare solid electrode, 3 is a particular classification based on the as manufactured chemical composition.

Metal arc gas shielded process. Recommended gases and gas mixtures for various metals and alloys

Metal type	Gas shield	Remarks
Carbon and low-alloy steels	CO_2	For dip transfer, and spray transfer Spatter problems. Use deoxidized wire
	Ar–15/20% CO_2	For dip or spray transfer. Minimum spatter
	Ar–5% CO_2	For dip and spray transfer
	Ar–5% O_2	Spray transfer. High impact properties
	Ar–5% CO_2–O_2 2%	For pulsed arc and thin sections
Stainless steels	Ar–1/2% O_2	Spray transfer
	75 He 23.5% Ar 1.5% CO_2	High quality dip transfer. For thin sections and positional work. Good profile
	He 75%–Ar 24%–O_2 1%	
Aluminium and its alloys	Argon	Stable arc with little spatter
	Helium	Hotter arc, less pre-heat, more spatter
	He 75% Ar 25%	Stable arc, high heat input. Good penetration. Recommended for thicknesses above 16 mm
Magnesium and its alloys	Argon	Stable arc
	He 75% Ar 25%	Hotter arc. Less porosity
Copper and its alloys	Argon	For sections up to 9.5 mm thickness
	Helium	
	He 75% Ar 25%	For medium and heavy sections. High heat input
Nickel and its alloys	Argon	Sections up to 9.5 mm thickness. Pulsed arc
	Ar 70% He 30%	High heat input, less cracking in thicker sections of 9% Ni
	Ar 25% He 75%	
Cupro-nickel	Argon	Stable arc
	Ar 70% He 30%	Stable arc with less cracking risk
Titanium, zirconium and alloys	High purity argon	Very reactive metals. High purity shielding gases are essential

Note. O_2 increases the wetting action.
See appendix 3 for proprietary gases.

The following colour codes are used for cylinders (BS 349, BS 381 C)

Gas or mixture	Colour code
Acetylene, dissolved	Maroon; also with name ACETYLENE.
Argon	Peacock blue.
Argon–CO_2	Peacock blue. Light brunswick green band round middle of cylinder.
Argon–hydrogen	Peacock blue. Signal red band round middle of cylinder.
Argon–nitrogen	Peacock blue. French grey band round middle of cylinder.
Argon–oxygen	Peacock blue. Black band round middle of cylinder. % of oxygen indicated.
CO_2, commercial vapour withdrawal	Black.
CO_2, commercial liquid withdrawal	Black. Narrow white band down length of cylinder (syphon tube).
Helium	Brown.
Hydrogen	Signal red.
Hydrogen–high purity	Signal red. White circle on shoulder.
Hydrogen–nitrogen	Signal red. French grey band round middle of cylinder.
Nitrogen	French grey. Black cylinder shoulder.
Nitrogen (oxygen free)	French grey. Black cylinder shoulder with white circle.
Nitrogen–CO_2	French grey. Light brunswick green band round middle of cylinder and black cylinder shoulder.
Nitrogen–oxygen	French grey. Black band round middle of cylinder and black cylinder shoulder.
Oxygen, commercial	Black.
Propane	Signal red with name PROPANE.
Air	French grey.
Air–CO_2	French grey. Light brunswick green band round middle of cylinder.

Note. Colour bands on shoulder of cylinder denote hazard properties. Cylinders having a red band on the shoulder contain a flammable gas; cylinders having a golden yellow band on the shoulder contain a toxic (poisonous) gas. Cylinders having a red band on the upper shoulder and a yellow band on the lower shoulder contain a flammable and toxic gas. Carbon monoxide has a red cylinder with a yellow shoulder.

A 15. Triple deoxidized steel wire recommended for pipe welds and root runs in heavy vessel construction. Analysis: 0.12% max. C; 0.9–1.6% Mn; 0.3–0.9% Si; 0.04–0.4% Al; 0.15% max. Ti; 0.15% max. Zr; 0.04% max. S and P.

A31. Molybdenum bearing mild steel wire. Used for most mild-steel applications requiring extra strength; high-tensile and quench and tempered steel; and suitable for offshore pipeline application and root runs in thick joints. Analysis: 0.14% max. C; 1.6–2.1% Mn; 0.5–0.9% Si; 0.4–0.6% Mo; 0.03% max. S and P.

Flux cored and metal powder cored filler wire*

Flux cored and metal powder cored wires, used with a gas shield of CO_2 or argon–5% CO_2 to argon–20% CO_2 give rapid deposition of metal and high-quality welds in steel. Metal powder wire used with argon–20% CO_2 gives a smooth arc with little spatter and is used with wire negative polarity. It can be used for positional work and because of the small amount of slag, which is produced intermittently, there is no need to deslag between runs.

The wire is manufactured from a continuous narrow flat steel strip formed into a U shape which is then filled with flux and formed into a tube. It is then pulled through reducing dies which reduce the diameter and compress the flux uniformly and tightly into a centre core (Fig. 2.10*a*). It is supplied on reels as for MIG welding in diameters 1.6, 2.0, 2.4, 3.2 mm and

Fig. 2.10. (*a*) Flux cored process with gas shield.

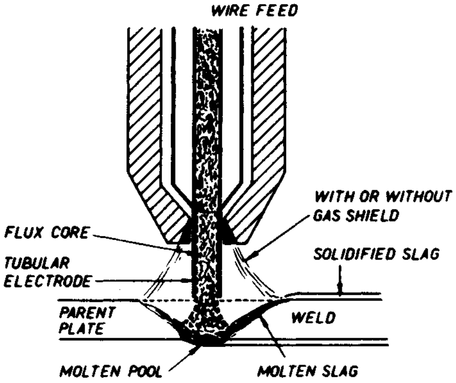

---

* American classification of mild steel tubular welding wires; flux cored arc welding (FCAW), with and without external gas shield. *Example* E 70T–1: E = electrode; 70 = tensile strength of weld metal in 1000lb/sq in. (ksi); T = tubular continuous electrode with powdered flux in core; 1 = gas type and current. (1, 2 – CO_2 deep; 3, 4, 6, 8 – none, deep: 5 – CO_2 or none, deep; 7 – none, deep; 9 – miscellaneous).

fed to the feed rollers of a MIG type unit. The welding gun can be straight or swan-necked and should be water cooled for currents above 400 A, and the power unit is the same as for MIG welding.

There is deep penetration with smooth weld finish and minimum spatter, and deposition rate is of the order of 10 kg/h using 2.4 mm diameter wire. The deep penetration characteristics enable a narrower V preparation to be found for butt joints resulting in a saving of filler metal, and fillet size can be reduced by 15–20%.

Some of the filler wires available are:

(1) Rutile type, steel, giving a smooth arc and good weld appearance with easy slag removal. Its uses are for butt and fillet welds in mild and medium-tensile steel in the flat and horizontal–vertical position.
(2) Basic hydrogen-controlled type, steel. This is an all-positional type and gives welds having good low-temperature impact values.
(3) Basic hydrogen-controlled type, steel, with $2\frac{1}{2}\%$ Ni for applications at − 50°C.
(4) Basic hydrogen-controlled type for welding 1% Cr, 0.5% Mo steels.
(5) Basic hydrogen-controlled type for welding 2.25% Cr, 1% Mo steels.

Self-shielded flux cored wire

Self-shielded flux cored wires are used without an additional gas shield and can be usefully employed in outdoor or other on site draughty situations where a cylinder-supplied gas shield would be difficult to establish.

The core of these wires contains powdered metal together with gas-forming compounds and deoxidizers and cleaners. The gas shield formed protects the molten metal through the arc and slag-forming compounds form a slag over the metal during cooling, protecting it during solidification. To help prevent absorption of nitrogen from the atmosphere by the weld pool, additions of elements are made to the flux and electrode wire to effectively reduce the soluble nitrogen.

This process can be used semi- or fully automatically and is particularly useful for on-site work (Fig. 2.10*b*).

Metal cored filler wire

The wire has a core containing metallic powders with minimal slag-forming constituents and there is good recovery rate (95%) with no interpass deslagging. It is used with CO_2 or argon–CO_2 gas shield to give welds with low hydrogen level (less than 5 ml/100 g weld metal). The equipment is similar to that used for MIG welding, and deposition rates are higher than with stick electrodes especially on root runs.

Safety precautions

The precautions to be taken are similar to those when metal arc welding, given on pp. 29–30. The BS 679 recommended welding filters are up to 200 A, 10 or 11 EW (electric welding); over 200 A, 12, 13, or 14 EW. When welding in dark surroundings choose the higher shade number, and in bright light the lower shade number. Because there is greater emission of infra-red energy in this process a heat absorbing filter should be used. The student should consult the following publications for further information: BS 679, *Filters for use during welding*, British Standards Institution; *Electric arc welding*, new series no. 38, *Safety, health and welfare*. Published for the Department of Employment by HMSO.

Techniques

There are three methods of initiating the arc. (1) The gun switch operates the gas and water solenoids and when released the wire drive is switched on together with the welding current. (2) The gun switch operates the gas and water solenoids and striking the wire end on the plate operates the wire drive and welding current (known as 'scratch start'). (3) The gun switch operates gas and water solenoids and wire feed with welding current, known as 'punch start'.

The table on p. 115 indicates the various gases and mixtures at present in use. As a general rule dip transfer is used for thinner sections up to 6.4 mm

Fig. 2.10. (*b*) Flux cored process (self-shielded).

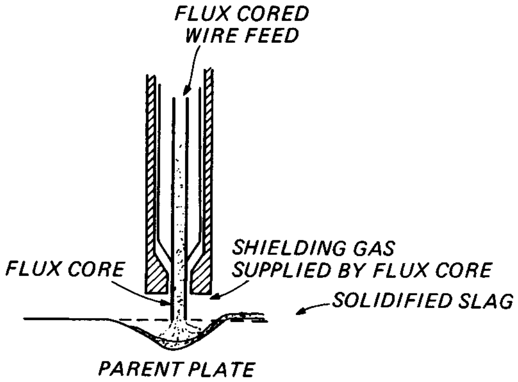

and for positional welding, whilst spray transfer is used for thicker sections. The gun is held at an angle of 80° or slightly less to the line of the weld to obtain a good view of the weld pool, and welding proceeds with the nozzle held 6–12 mm from the work (see Fig. 2.11*a*). Except under special conditions welding takes place from right to left. If welding takes place for special reasons from left to right the torch has to be held at almost 90° to the line of travel and care must be taken that the gas shield is covering the work.

The further the nozzle is held from the work the less the efficiency of the gas shield, leading to porosity. If the nozzle is held too close to the work spatter may build up, necessitating frequent cleaning of the nozzle, while arcing between nozzle and work can be caused by a bent wire guide tube allowing the wire to touch the nozzle, or by spatter build-up short-circuiting wire and nozzle. If the wire burns back to the guide tube this may be caused by a late start of the wire feed, fouling of the wire in the feed conduit or the feed rolls being too tight. Intermittent wire feed is generally due to insufficient feed roll pressure or looseness due to wear in the rolls. Excessively sharp bends in the flexible guide tubes can also lead to this trouble.

Root runs are performed with no weave and filler runs with as little weave as possible consistent with good fusion since excessive weaving tends to promote porosity. The amount of wire projecting beyond the contact tube is important because the greater the projection, the greater the I^2R effect and the greater the voltage drop, which may reduce the welding current and affect penetration. The least projection commensurate with

Fig. 2.11. (*a*) Angle of torch (flat position). (*b*) Stubbing. (*c*) Dip transfer on thin sheet.

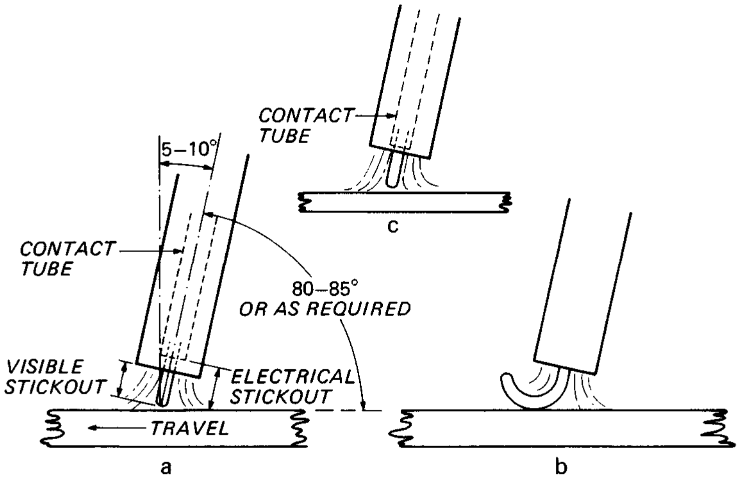

accessibility to the joint being welded should be aimed at. Backing strips which are welded permanently on to the reverse side of the plate by the root run are often used to ensure sound root fusion. Backing bars of copper or ceramics with grooves of the required penetration bead profile can be used and are removed after welding. It is not necessary to back-chip the roof run of the light alloys but with stainless steel this is often done and a sealing run put down. This is a more expensive way compared with efficient back purging. The importance of fit-up in securing continuity and evenness of the penetration bead cannot be over-emphasized.

Flat welds may be slightly tilted to allow the molten metal to flow against the deposited metal and thus give a better profile. If the first run has a very convex profile poor manipulation of the gun may cause cold laps in the subsequent run.

Run-on and run-off plates can be used as in TIG welding to obviate cold start and crater finishing leading to cracks. Slope in and out controls obviate much of this.

Stubbing

If the wire speed is too high the rate of feed of the wire is greater than the burn-off rate and the wire stubs. If a reduction in the wire speed feed rate does not cure the stubbing, check the contact tube for poor current pick-up and replace if necessary (Fig. 2.11*b*).

Burn back control

This operates a variable delay so that the wire burns back to the correct 'stick-out' beyond the nozzle before being switched off. It is then ready for the next welding sequence. Excessive burn back would cause the wire to burn to the contact tip. Varying the stick-out length (Fig. 2.11*a*) varies the temperature of the molten pool. Increasing the length gives a somewhat 'cooler' pool and is useful in cases of poor fit-up, for example.

The nearer that the contact tube end is to the outer end of the torch nozzle, the greater the likelihood of the contact tube being contaminated with spatter. With dip transfer, in which the spatter is least, the distance is about 2–3 mm, but for spray transfer it should be in the region of 5–12 mm. The contact tube may even be used slightly projecting from the gas nozzle when being used on thin sheet (e.g. body panels) using dip transfer. This gives a very clear view of the arc but is not recommended as the contact tube becomes rapidly contaminated even when treated with anti-spatter paint or paste and must be cleaned frequently (Fig. 2.11*c*.) Whatever the stick-out, the flow of shielding gas should be such that molten pool and immediate surrounding areas are covered.

Positional welding

This is best performed by the dip transfer method since the lower arc energy enables the molten metal to solidify more quickly after deposition. Vertical welds in thin sections are usually made downwards with no weave. Thicker sections are welded upwards or with the root run downwards and subsequent runs upwards, weaving as required. Overhead welding, which is performed only when absolutely necessary, is performed with no weave.

Fillet welds are performed with the gun held backwards to the line of welding, as near as possible to the vertical consistent with a good view of the molten pool, bisecting the angle between the plates and with a contact tip-to-work distance of 16–20 mm (Fig. 2.12). On unequal sections the arc is held more towards the thicker section. The root run is performed with no weave and subsequent runs with enough weave to ensure equal fusion on the legs. Tilted fillets give better weld profile and equal leg length more easily.

Some of the filler wires available are:

Note. Except where indicated they can be used with CO_2, argon–5% CO_2 and argon–20% CO_2. CO_2 gives good results with dip transfer whereas the argon mixtures give better weld profile. L indicates 0.08% C or lower.

(1) 1.5% Mn, L. Smooth arc, little slag, so no deslagging, Argon–20% CO_2, cored wire −ve.

(2) 1.8% Mn, 0.5% Mo, 0.14% C. High strength mild steels and for highet tensile steels, solid wire +ve.

(3) 1.4% Mn, 0.5% Mo, L. Similar wire to (1) but molybdenum gives extra strength. Argon–20% CO_2, cored wire −ve.

(4) 1.4% Mn, 0.45% Cu, L. A similar wire to (1) but with copper added for weathering steels such as Corten. Use argon–5% or argon–20% CO_2, cored wire −ve.

Fig. 2.12. Angle of torch: horizontal–vertical fillet.

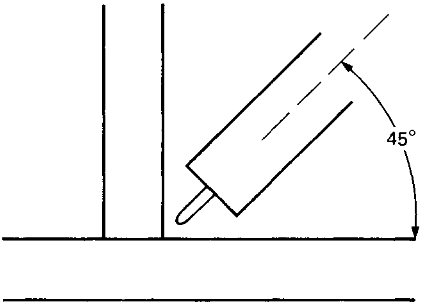

(5) 1.3% Cr, 1.1% Mn, 0.2% Mo, 1.0% Ni, L. High strength welds, tensile strengths of 750 N/mm^2. Rutile basic flux cored wire +ve.
(6) 2.0% Mn, 1.0% Mo, L. Basic flux core with iron powder, creep-resistant steels, cored wire −ve.
(7) 2.2% Cr, 1.0% Mo, 0.12% C. Basic flux core for creep-resistant steels to temperatures of 580°C, solid wire − ve.
(8) 2.2% Ni, L. For low temperatures to −50 °C. Good resistance to cracking cored wire −ve, use argon–CO_2 mixtures.

Stainless steels

Preparation angle is usually similar for semi-automatic and automatic welding for butt welds in stainless steel, being 70–80° with a 0–1.5 mm gap (Fig. 3.22). Torch angle should be 80–90° with an arc length long enough to prevent spatter but not long enough to introduce instability. All oxides should be removed by stainless steel wire-brushing. Back chipping is usually performed by grinding. With spray transfer using high currents there is considerable dilution effect and welds can only be made flat, but excessive weaving should be avoided as this increases dilution. Direct current with the torch (electrode) positive is used with a shield of argon + 1% oxygen.

When welding stainless steel of any thickness it is imperative to obtain good penetration and this is only possible if there is a good back purge. Without this the penetration has to be back chipped and a sealing run made which adds considerably to the cost of the joint.

Back purging can be obtained simply by tack welding a strip on the underside of the weld and removing after welding (Fig. 2.13). This uses argon from the torch and reduces oxide formation on the penetration bead. Back purging of tubes can often be done using soluble paper dams or piped bladders which are blown up in position inside the pipe and serve to contain the inert gas on the underside of the weld, thus effecting considerable saving.

Fig. 2.13. Butt welding stainless steel MIG, 99% A–1% O_2, d.c. wire positive. Automatic or semi-automatic.

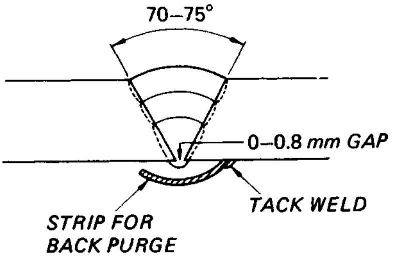

In the welding of dissimilar metals to stainless steel the shortcircuiting arc (dip transfer) gives much less dilution with lower heat input and is generally preferred. The arc should be kept on the edge of the parent metal next to the molten pool and not in the pool itself, to reduce the tendency to the formation of cold laps. Argon 1–2% oxygen is used as the shielding gas and filler wires should be chosen to match the analysis of the parent metal while allowing for some loss in the transfer across the arc. Examples are given in the table. Figs. 2.13 and 3.27 give suitable methods of preparation.

Stainless steel filler wires available include the following

BS 2901 Pt 2	AISI	AWS	
308S92	304, 304L	A59 ER308L	Low carbon (0.03% max.) 19.5–22% Cr, 9–11% Ni, suitable for welding 18% Cr, 8% Ni steels including low carbon types.
	321, 347		18% Cr, 8 Ni, Nb stabilized at temperatures up to 400 °C
316S92	316, 316L 318	A59 ER 316L	Low carbon 0.03% max., 18–20% Cr, 11–14% Ni, 2–3% Mo, for welding 18% Cr, 8% Ni, Mo and 18% Cr, 8% Ni, Mo, Nb types
316S96	316, 318		Carbon 0.08% max., 18–20% Cr, 11–14% Ni for steels of similar composition to give high impact strength at low temperatures
347S96	347, 321, 304		Carbon 0.08% max., 19–21.5% Cr, 9–11% Ni niobium stabilized. For steels of 20% Cr, 10% Ni type, giving high corrosion resistance
309S94	347		Carbon 0.12% max., 25% Cr, 12% Ni for welding mild and low-alloy steel to stainless steel. for buttering, when welding type 347 to mild steel and for temperatures up to 300 °C

Note. Argon 1 or 2% oxygen is used as the shielding gas.
These wires are also used for submerged arc welding of stainless steels.
AISI: American Iron and Steel Institute.

Nickel alloys

The welding of nickel and nickel alloys can be done using spray transfer and short circuit (dip transfer) arc methods.

Spray transfer with its higher heat input gives high welding speeds and deposition rate and is used for thicker sections, usually downhand because of the large molten pool. Argon is used as the shielding gas but 10–20%

helium can be added for welding Nickel 200 and INCONEL® alloy 600, giving a wider and flatter bead with reduced penetration. The gun should be held as near to 90° to the work as possible, to preserve the efficiency of the gas shield, and the arc should be kept just long enough to prevent spatter yet not long enough to affect arc control.

Short circuiting arc conditions are used for thinner sections and positional welding, the lower heat input giving a more controllable molten pool and minimum dilution. The shielding gas is pure argon but the addition of helium gives a hotter arc and more wetting action so that the danger of cold laps is reduced. To further reduce this danger the gun should be held as near as possible to 90° to the work and moved so as to keep the arc on the plate and not on the pool. High-crowned profile welds increase the danger of cold laps.

The table in chapter 3 (p. 181) on TIG welding of nickel alloys indicates materials and filler metal suitable for welding by the MIG process and Fig. 3.28 the recommended methods of preparation.

Aluminium alloys

When fabricating aluminium alloy sections and vessels, the plates and sections are cut and profiled using shears, cold saw, band saw or arc plasma and are bent and rolled as required and the edges cut to the necessary angle for welding where required. (*Note.* As the welding currents employed in the welding of aluminium are high, welding lenses of a deeper than normal shade should be used to ensure eye protection.)

The sections are degreased using a degreaser such as methyl chloride and are tack welded on the reverse side using either TIG or MIG so as to give as far as possible a penetration gap of close tolerance. Where this is excessive it can be filled using the TIG process, and pre-heating may be performed for drying. The areas to be welded are stainless steel wire brushed to remove all oxide and the root run is made. Each successive run is similarly wire brushed and any stop–start irregularities should be removed. It is important that the tack welds should be incorporated into the overbead completely so that there is no variation in it. Back-chipping of the root run, where required for a sealing run, should be performed with chisels, routers or saws and not ground, as this can introduce impurities into the weld.

Back-chipping and a sealing run add substantially to the cost of a welded seam. In many cases it can be avoided by correctly backpurging the seam and welding from one side only. Fig. 2.14 and table indicate typical butt weld preparation in the flat position using 1.6 mm filler wire. Fig. 2.15*a* shows the section of a MIG weld in aluminium and Fig. 2.15*b* a TIG weld on the preparation bead of Fig. 2.15*a*.

Plate thickness (mm)	Number of runs	Approximate current (A)		Root face (mm)
		Root	Subsequent	
6 and 8	2	220	250–280	1.6
9.5 and 11	3	230	260–280	1.6
12.7 and 16	4	230–240	270–290	3.2
19	5	240–250	280–310	3.2
22 and 25	6	240–260	280–330	4.8

Refer also to BS 3571 *General recommendations for manual inert gas metal arc welding of aluminium and aluminium alloys.*

Suitable filler wires are given on p. 175.

Fig. 2.14. Flat butt weld preparation, aluminium plate MIG, semi-automatic process argon shield. Work negative.

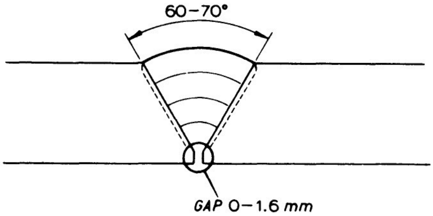

Fig. 2.15. (*a*) MIG weld in aluminium 5083 (NS8) plate 8 mm thick with 5556 A (NG61) wire. Stainless steel backing bar.
1st run: 250–270 A, 26–27 V, speed 9–10 mm/second.
2nd run: 260–280 A, 28–29 V, speed 8–10 mm/second.
Preparation: 90° V, 0–1 mm root gap, 3–4 mm root face.

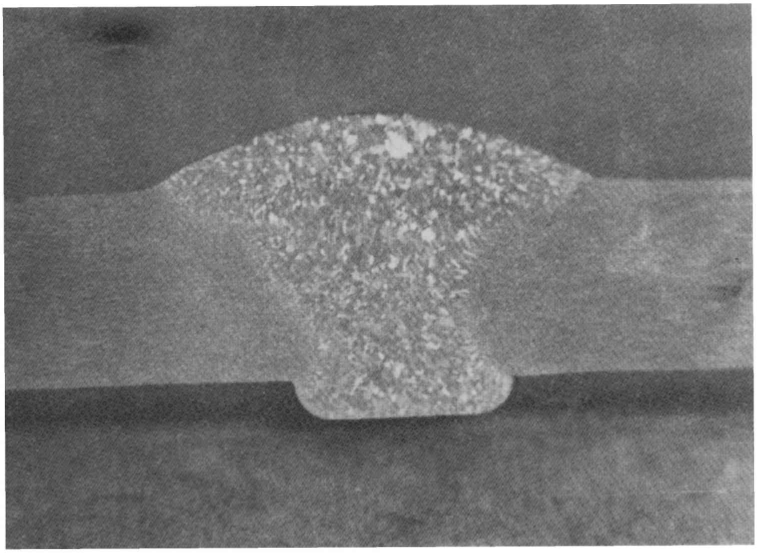

Copper and copper alloys

Many of the problems associated with the welding of copper and copper alloys are discussed in the chapter on the TIG process (Chapter 3) and apply equally to the MIG process. Because of the high thermal conductivity of copper and to reduce the amount of pre-heating, it is usual in all but the thinnest sections to use high currents with spray transfer conditions, which are obtainable with argon and argon–helium mixtures. The addition of nitrogen to argon destroys the spray transfer conditions but arc conditions are improved with up to 50% helium added to argon and this results in an increased heat output. For thin sections, fine feed wires can be used giving spray transfer conditions with lower current densities, thus preventing burn-through. Fig. 2.16 shows edge preparation for TIG and MIG butt welds in copper.

SG cast irons

Pearlitic and ferritic cast irons are very satisfactorily welded using dip transfer conditions (e.g. 150 A, 22 V with 0.8 mm diameter wire) to give low heat input using filler wire of nickel or the NI-ROD® alloy range. Carbide precipitation in the HAZ is confined to thin envelopes around some of the spheroids, unlike the continuous film associated with MMA welding. SG iron can be welded to other metals and MONEL®, INCONEL® or INCO® alloys can be used to give a corrosion-resistant surface on SG iron castings or as a buffer layer for other weld deposits. See

Fig. 2.15. (*b*) TIG weld on penetration bead of the above weld. No back chipping, no filler wire (can be added if required). 350 A, 32 V, speed 3–4 mm/second. Section etched with cupric chloride $CuCl_2$ followed by a 50% nitric acid wash.

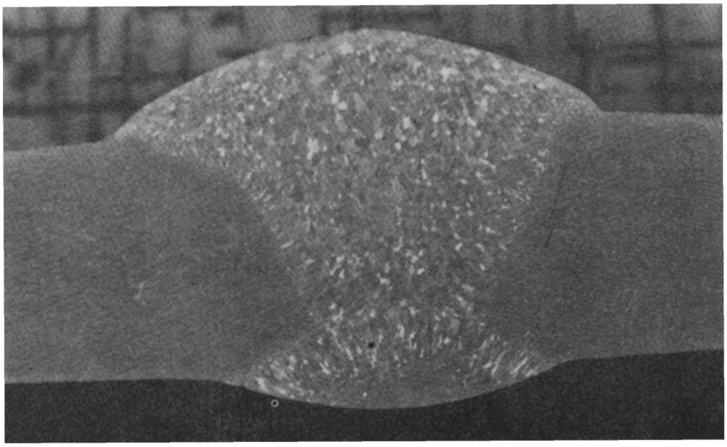

table, p. 71. Cleaning and degreasing should be performed before welding, and pre-heat is only required for heavy pearlitic section or joints under heavy restraint when pre-heat of 200 °C is suitable. Minimum-heat input compatible with adequate fusion should always be used (Fig. 2.17).

MIG Process. Recommended filler wires for copper welding

Type (BS 2870–2875)	Grade	Filler wire Argon or helium shield	 Nitrogen shield
C106	Phosphorus deoxidized, non-arsenical	C7, C8, C21	Not recommended
C107	Phosphorus deoxidized, arsenical		
C101	Electrolytic tough pitch, high conductivity	C7, C8, C21	Not recommended
C102	Fire refined tough pitch, high conductivity		
C103	Oxygen-free, high conductivity	C7, C21	Not recommended

Fig. 2.16. Edge preparation for TIG and MIG butt welds in copper.

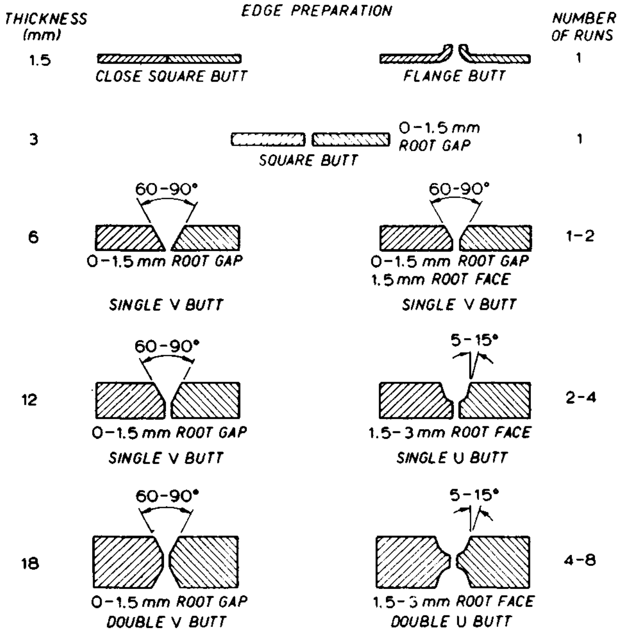
THICKNESS (mm)
EDGE PREPARATION
NUMBER OF RUNS
1.5
CLOSE SQUARE BUTT
FLANGE BUTT
1
3
0-1.5 mm ROOT GAP
SQUARE BUTT
1
60-90°
6
0-1.5 mm ROOT GAP
SINGLE V BUTT
60-90°
0-1.5 mm ROOT GAP
1.5 mm ROOT FACE
SINGLE V BUTT
1-2
60-90°
12
0-1.5 mm ROOT GAP
SINGLE V BUTT
5-15°
1.5-3 mm ROOT FACE
SINGLE U BUTT
2-4
60-90°
18
0-1.5 mm ROOT GAP
DOUBLE V BUTT
5-15°
1.5-3 mm ROOT FACE
DOUBLE U BUTT
4-8

CO_2 welding of mild steel

There are four controls to enable optimum welding conditions to be achieved: (1) wire feed speed which also controls the welding current, (2) voltage, (3) choke or series inductance and (4) gas flow.

For a given wire diameter the wire feed rate must be above a certain minimum value to obtain a droplet transfer rate of above about 20 per second, below which transfer is unsatisfactory. With increasing wire feed rate the droplet transfer rate and hence the burn-off increases and the upper limit is usually determined by the capacity of the wire feed unit. The voltage setting also affects the droplet frequency rate and determines the type of transfer, about 15–20 V for short-circuit or dip, and 27–45 V for spray, with an intermediate lesser used zone of about 22–27 V for the semi-short-circuiting arc.

The choke, which limits the rate of current rise and decay, is also important because too low a value can give a noisy arc with much spatter and poor weld profile, while too high a value can give unstable arc conditions with more difficult start and even occasional arc extinguishing. Between these limits there is a value which with correct arc voltage and wire feed rate gives a smooth arc with minimum spatter and good weld profile. Penetration is also affected by the choke value, see Vol. 1, p. 223.

As stated before, the short-circuiting arc is generally used for welding thinner sections, positional welding, tacking and on thicknesses up to 6.5 mm. In positional welding the root run may be made downwards with no weave and subsequent runs upward. The lower heat output of this type of arc reduces distortion on fabrications in thinner sections and minimizes

Fig. 2.17. Preparation for welding of SG cast irons.

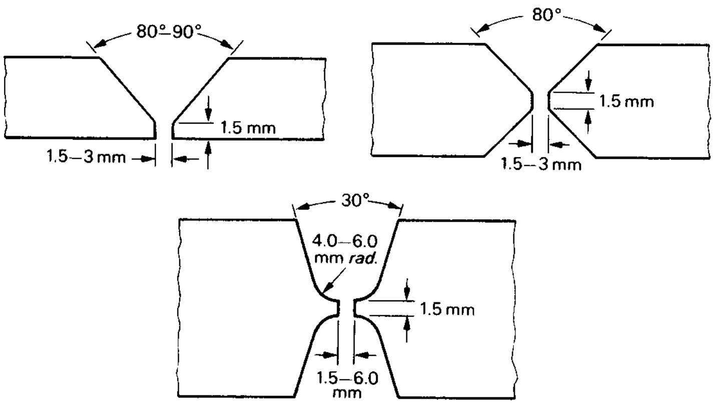

over-penetration. The spray-type arc is used for flat welding of thicker sections and gives high deposition rates.

Gas flow rate can greatly affect the quality of the weld. Too low a flow rate gives inadequate gas shielding and leads to the inclusion of oxides and nitrides, while too high a rate can introduce a turbulent flow of the CO_2 which occurs at a lower rate than with argon. This affects the efficiency of the shield and leads to a porosity in the weld. The aim should be to achieve an even non-turbulent flow and for this reason spatter should not be allowed to accumulate on the nozzle, which should be directed as nearly as possible at 90° to the weld, again to avoid turbulence.

The torch angle is, in practice, about 70–80° to the line of travel consistent with good visibility and the nozzle is held about 10–18 mm from the work. If the torch is held too close, excess spatter build-up necessitates frequent cleaning, and in deep U or V preparation the angle can be increased to obtain better access. Weaving is generally kept as low as convenient to preserve the efficiency of the gas shield and reduce the tendency to porosity. Wide weld beads can be made up of narrower 'stringer' runs, and tilted fillets compared with HV fillets give equal leg length more easily, with better profile.

Mild steel sheet, butt welds, CO_2 shielding, flat, 0.8 mm diam. wire (approximate values)

Thickness (mm)	Gap (mm)	Wire feed (m/minute)	Arc (volts)	Current (A)
1	0	2.8–3.8	16–17	65–80
1.2	0	3.2–4.0	18–19	70–85
1.6	0.5	4.0–4.8	19–20	85–95
2.0	0.8	5.8–7.0	19–20	110–125
2.5	0.8	7.0–8.4	20–21	125–140
3.0	1.5	7.0–8.4	20–21	125–140

Economic considerations

Although filler wire for the CO_2 process, together with the cost of the shielding gas, is more expensive than conventional electrodes, other factors greatly affect the economic viability of the process. The deposition rate governs the welding speed which in turn governs the labour charge on a given fabrication.

The deposition rate of the filler metal is a direct function of the welding current. With metal arc welding the upper limit is governed by the overheating of the electrode. The current I amperes flows through the electrode, the wire of which has an electrical resistance R ohms, so that the heating effect is $\propto I^2R$. The resistance of any metallic conductor increases with the rise in temperature, so that, as the electrode becomes hotter, the resistance and hence the I^2R loss increases so that with excessive currents, when half of the electrode has been consumed, the remaining half has become red hot and the coating ruined. Iron powder electrodes have a greater current-carrying capacity due to the conductivity of the coating, the electrical resistance being reduced. With the CO_2 process the distance from contact tube to wire tip is of the order of 20 mm so that the electrical resistance is greatly reduced even though the wire diameter is smaller. The current can thus be increased greatly, resulting in higher deposition rates, greater welding speeds and reduced labour charges. In addition the duty cycle is increased since there is no constant electrode change and need for deslagging, but there may be greater spatter loss. Economically therefore the CO_2 process shows an advantage in very many applications, though the final choice of process is governed by the application and working conditions.

By using the argon–CO_2 gas mixtures, certain advantages are obtained over CO_2 which may offset the greater price of the mixtures. These are faster welding speed and reduced spatter, with an improved weld profile.

Fig. 2.18. Various preparations for thicker plate (varying with applications). CO_2 welding.

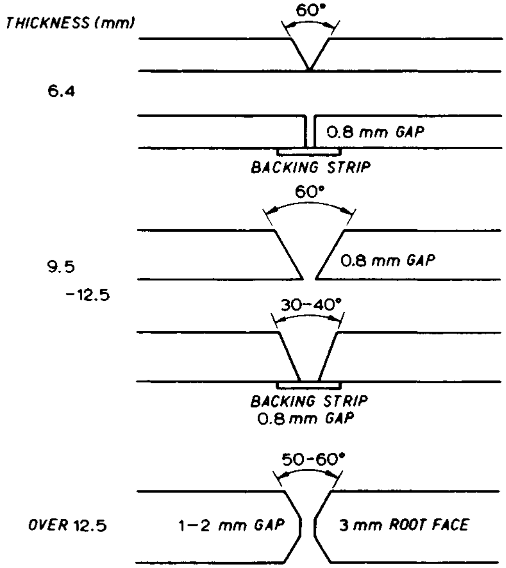

When welding thin steel sections with a dip transfer arc, argon–CO_2 mixtures are often preferred because there is a lower voltage drop across the arc which thus gives a cooler molten pool and prevents burn-through, especially if the fit-up is poor.

Details of plate preparation are given as a guide only, since there are too many variables to give generalized recommendations (see Figs. 2.18 and 2.19).

Automatic welding

Automatic welding, by MIG, pulse and CO_2 processes, now plays an important part in welding fabrication practice. It enables welds of consistently high quality and accuracy to radiographic standard to be performed at high welding speeds because of the close degree of control over the rate of travel and nozzle-to-work distance. It is less tolerant than semi-automatic welding to variations of root gap and fit-up but reduces the number of start–stop breaks in long sequences. The choice between semi-automatic and automatic process becomes a question of economics, involving the length of runs, number involved, volume of deposited metal if the sections are thick, method of mechanization and set-up time. The torches are now usually air cooled even for currents up to 450 A and are carried on welding heads fitted with controls similar to those used for semi-automatic welding, and may be remote controlled (Fig, 2.20).

The head may be: (1) fixed, with the work arranged to move or be rotated beneath it, (2) mounted on a boom and column which can either be of the

Fig. 2.19. (*a*) Unprepared fillets. (*b*) multi-run prepared fillets in thick plate, (*c*) deep preparation fillets, (*d*) multi-run unprepared fillet. CO_2 welding.

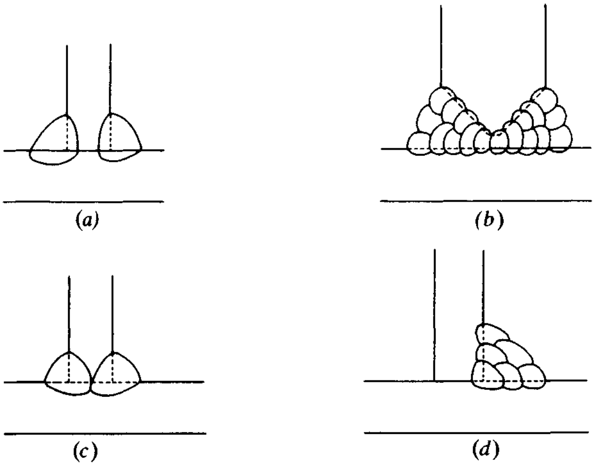

positioning type in which the work moves, or the boom can traverse over the work (Figs. 2.21*a* and *b*), (3) gantry mounted so as to traverse over the stationary work, (4) tractor mounted, running on guide rails to move over the fixed work, (5) mounted on a special machine or fixture designed for a specific production. A head may carry two torches arranged to weld simultaneously, thus greatly reducing the welding time.

For the CO_2 process in steel a typical example would be with wire of 1.2 mm diameter using 150–170 A on thinner sections and multiple passes up to 30 on thicknesses up to 75 mm with current in the 400–500 A range and 2.4 mm wire. With automatic surge (pulse) arc welding on stainless

Fig. 2.20. MIG head showing wire feed.

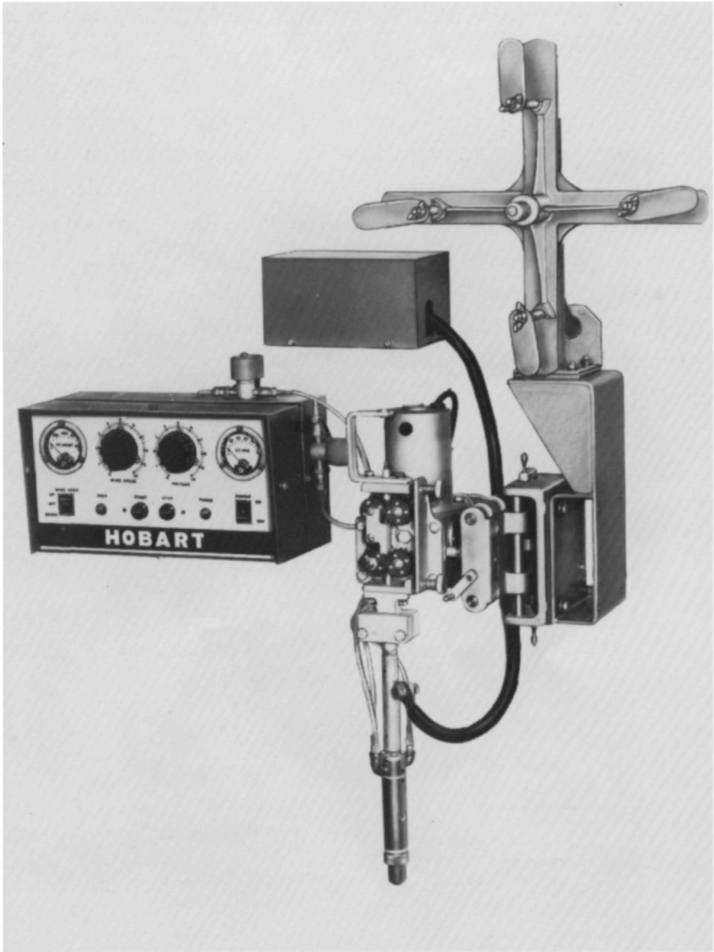

steel, accurate control of the underbead is achieved, obviating the necessity for back-chipping and sealing run. In the case of aluminium welding on plate above 10 mm thickness the accuracy of the underbead produced with the MIG automatic process results in more economical welding than by the double-operator vertical TIG method. In general the full automation of these processes results in greater productivity with high-quality welds.

Magnetic arc control systems

Magnetic probes with electronic solid state control and with water cooling if required can be fitted onto or near to the welding head and are applicable to spray, MIG, TIG, plasma and submerged arc systems (Fig. 2.22*a* and *b*).

Fig. 2.21. (*a*) A ram-type boom with welding head.

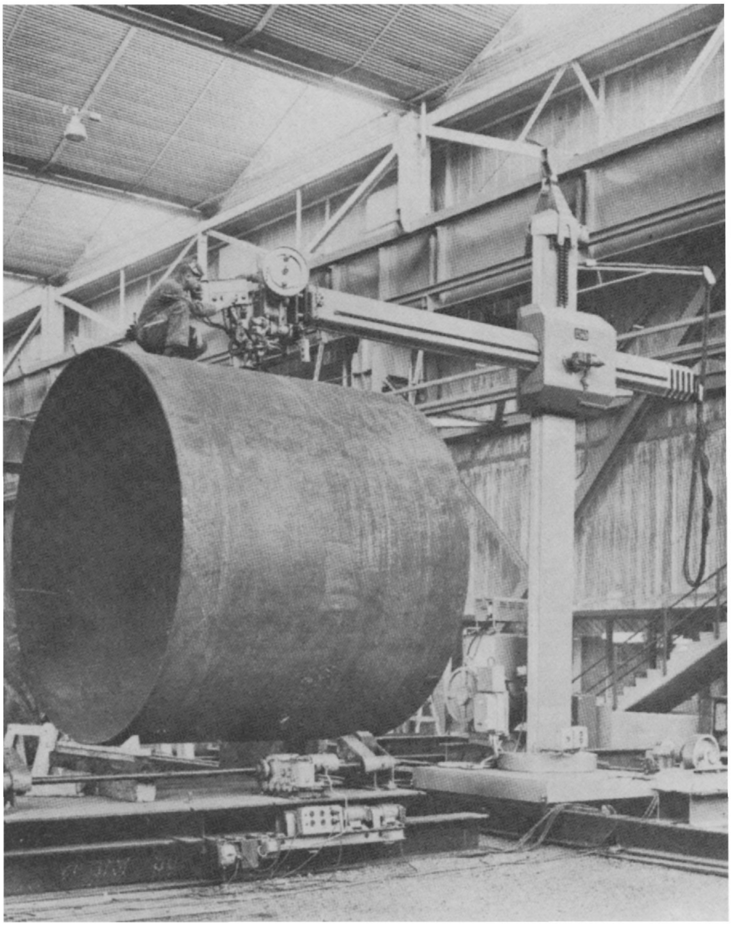

Fig. 2.21. (*b*) Extra heavy duty column and boom and roller bed.

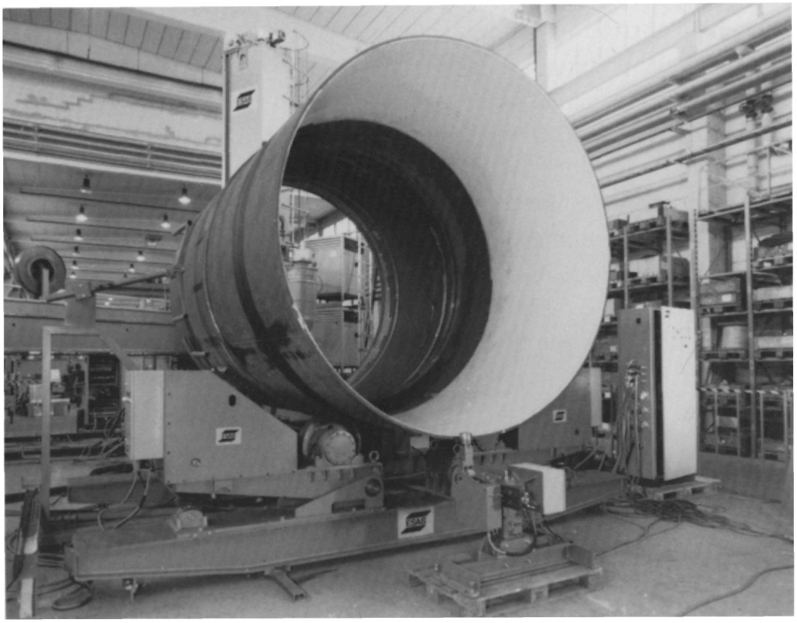

Fig. 2.22. (*a*) Types of probe. (*b*) Magnetic arc control.

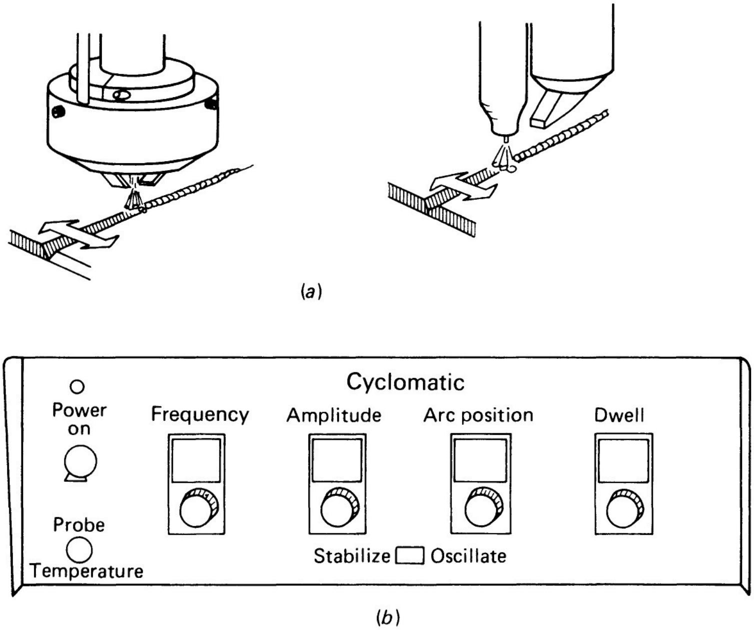
(a)
Cyclomatic
Power on
Frequency
Amplitude
Arc position
Dwell
Probe
Temperature
Stabilize
Oscillate
(b)

The arc can be stabilized or oscillated by the applied magnetic field and magnetic and non-magnetic materials are equally well welded when using this system. Controls on the unit enable the sweep frequency to be varied from 0 to 100 oscillations per second while sweep amplitude and arc position are controllable and proportional to the arc length in an approximately 1 : 1 ratio. There is controllable variable dwell on each side of the oscillation in order to obtain good fusion, arc blow is counteracted and the heat imput can be balanced when welding thin to thick sections, and undercut is minimized. Porosity is decreased and the system is applicable, for example, to the welding of thicker stainless steel sections resulting in reduced danger from cracking, reduced porosity and greater weld reliability. This method of arc stabilization and control is now becoming increasingly popular as MIG and submerged arc welding become more automated.

Seam tracker

This unit bolts on to the head of the welding torch and enables the torch to follow accurately the line of the joint to be welded.

The torch head is placed over the seam or joint and the tracker keeps the arc over the seam with an accuracy of less than 0.25 mm.

One type has an electro-mechanical sensor which responds to the movement of the probe tip which moves along the seam. The signals are sent by the probe to the solid state control which operates two linear slides whose axes are at right angles to each other, one horizontal and the other vertical. These are driven by small d.c. motors and keep the welding head in place exactly over the seam to be welded, which keeps the torch-to-work distance constant (Fig. 2.23).

Fig. 2.23. Programmable seam tracker.

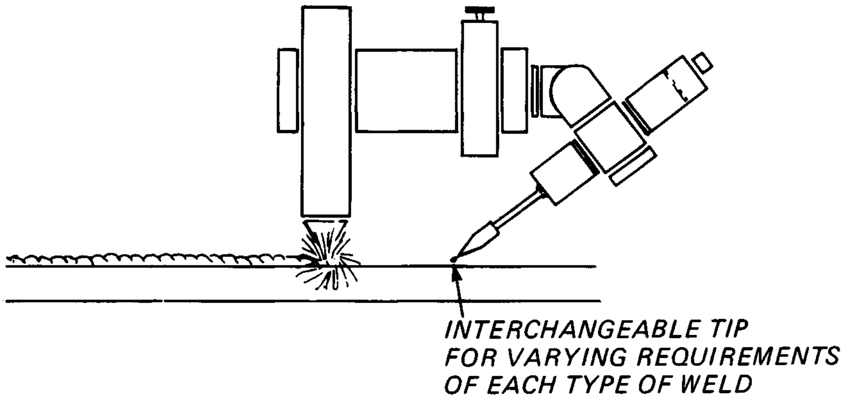

The optical seam tracker using flexible fibre optics for operation and with microcircuit control operating the cross slides as in the previous unit keeps the torch head accurately over the joint, which should be a tight machined edge butt weld, although this system can tolerate a gap of 2 mm and still keep the torch central to the joint.

MIG spot welding

Spot welding with this process needs access to the joint on one side only and consists of a MIG weld, held in one spot only, for a controlled period of time. Modified nozzles are fitted to the gun, the contact tip being set 8–12 mm inside the nozzle, and the timing unit can either be built into the power unit or fitted externally, in which case an on–off switch does away with the necessity of disconnecting the unit when it is required for continuous welding. The timer controls the arcing time, and welds can be made in ferrous material with full or partial penetration as required. A typical application of a smaller unit is that of welding thin sheet as used on car bodies. The spot welding equipment timer is built into a MIG unit and is selected by a switch. Wire of 0.6 mm diameter is used with argon + CO_2 5% + O_2 2% as the shielding gas. Currents of 40–100 A are available at 14–17 V for continuous welding with a maximum duty cycle of 60%. The spot welding control gives a maximum current of 160 A at 27 V, enabling spot welds to be made in material down to 0.5 mm thickness, the timing control varying from 0.5–2.0 seconds.

On larger units arcing time can be controlled from 0.3 to 4.5 seconds with selected heavier currents enabling full penetration to be achieved on ferrous plate from 0.7 to 2.0 mm thick and up to 2.5 mm sheet can be welded on to plate of any thickness.

Cycle arc welding is similar to spot welding except that the cycle keeps on repeating itself automatically as long as the gun switch is pressed. The duration of the pause between welds is constant at about 0.35 seconds while the weld time can be varied from 0.1 to 1.5 seconds. This process is used for welding light-section components which are prone to burn-through. In the pause between the welding period, the molten pool cools and just solidifies, thus giving more accurate control over the molten metal.

MIG pulsed arc welding

Pulsed arc welding is a modified form of spray transfer in which there is a controlled and periodic melting off of the droplets followed by projection across the arc. A pulse of voltage is applied for a brief duration at regular frequency and thus results in a lower heat output than with pure spray transfer, yet greater than that with dip transfer. Because of this, thinner sections can be welded than with spray transfer and there is no danger of poor fusion in a root run as sometimes occurs with dip transfer and positional welding is performed much more easily. There is regular and even penetration, no spatter and the welds are of high quality and appearance.

To obtain these conditions of transfer it is necessary to have two currents fed to the arc: (*a*) a background current which keeps the gap ionized and maintains the arc and (*b*) the pulsed current which is applied at 50 or 100 Hz and which melts off the wire tip into a droplet which is then projected across the arc gap. These two currents, which have critical values if satisfactory welding conditions are to be obtained, are supplied from two sources, a background source and a pulse source contained in one unit, and their voltages are selected separately. The background current, of much lower value than the pulse, is half a cycle out of phase with it (see Figs. 2.24*a* and *b*). A switch enables the pulse source to be used for dip and spray transfer methods as required. The power supply is a silicon rectifier with constant voltage output and a maximum current value of about 350 A at open-circuit voltages of 11–45 V similar to those already described.

To operate the unit, the 'pulse height' (the value of pulse current) and the background current are selected on separate switches, the wire is adjusted to protrude about 10 mm beyond the nozzle and welding is commenced

Fig. 2.24. (*a*) Wave shape of pulse supply.

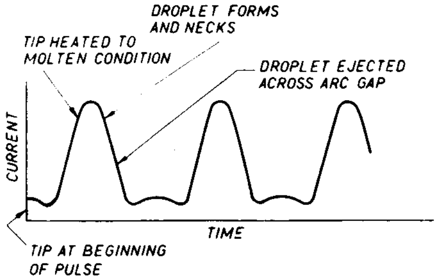

Fig. 2.24. (*b*) Modified pulse supply with background pulse.

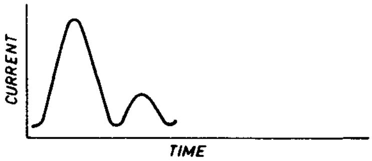

moving from right to left down the line of the weld with the gun making an angle of 60–70° with the line of weld. When welding fillets the gun is usually held at right angles to the line of weld with the wire pointing directly into the joint, and thus an excellent view is obtained of the degree of fusion into the root; welding is again performed from right to left. The process gives good root fusion with even penetration and good fill-in and is especially efficient when used fully automatically. As in all other welding processes, welding is best performed flat but positional welding with pulsed arc is very satisfactory and relatively easy to perform. Thicknesses between 2 mm and 6.5 mm which fall intermediate in the ranges for dip and spray transfer are easily welded. One of the drawbacks to pulse arc is the necessity to ensure accurate fit-up. If there are any sizeable gaps a keyhole effect is produced and it is impossible to obtain a regular underbead. The gap should be of the order of from 0 to about 1.0 mm max. The shielding gas is argon with 1% or 2% oxygen or argon with 5% CO_2 and 2% oxygen for welding mild and low-alloy steels, stainless steel and heat-resistant steel, and pure argon for aluminium and its alloys, 9% nickel steel and nickel alloys. Pulse arc is especially useful when used automatically for stainless steel welding, since the accurately formed under- or penetration bead obviates the expensive operation of back-chipping and a sealing run and there is little carbon pick-up and thus little increase of carbon content in the weld. Aluminium requires no back purge and for the back purge on stainless steel a thin strip of plate can be tack welded on the underside of the joint (see Fig. 2.13) and removed after welding. The argon from the torch supplies the back purge and prevents oxidation of the underside of the weld.

There is good alloy recovery when welding alloy steel and because of the accurate heat control, welding in aluminium is consistently good without porosity and a regular underbead so that it can be used in place of double argon TIG with a saving of time and cost.

Fabrications and vessels can be fully tack weld fabricated with TIG on the underside of the seam and pulse arc welded, greatly reducing the

Fig. 2.25. Automatic pulsed arc welding of stainless steel, flat, 99% argon–1% oxygen.

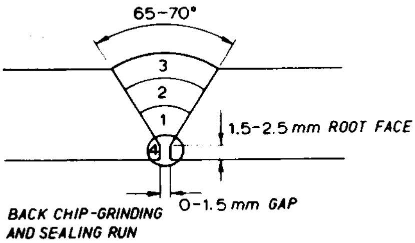

tendency to distortion. The torch should be held at 75–80° to the line of the weld and good results are also obtained fully automatically by welding from left to right with the torch held vertically to the seam. This method allows greater tolerance in fit-up and preparation with better penetration control. Because this process is sensitive to the accuracy of fit-up and preparation, care should be taken to work to close tolerances and excessive gaps should be made up from the reverse side with, for example, the TIG process, and then carefully cleaned.

Fig. 2.25 and the table give a typical preparation for automatic flat butt welds in stainless steel using 99% A, 1% O_2 as shielding gas and 1.6 mm diameter filler wire. For the alloys of nickel the recommended shielding gas is pure argon. A similar technique is used as for the MIG welding of stainless steel, with a slight pause each side of the weave to avoid undercut.

Plate thickness (mm)	Runs	Current (A)	Volts (approx.)
		Roots and subsequent	
4.8, 6.4, 8	2	180–200	27–28
9.5	3	180–210	27–28

There appears to be no advantage in using the pulse method over the normal shielded metal arc with unpulsed wave when using CO_2 as the shielding gas. (See Appendix 10).

Synergic pulse MIG welding

In the first generation of pulsed MIG units the output waveform consisted essentially of a half or full wave rectified sine wave (pp. 138–9) with the pulse frequency fixed by that of the main supply to either 50 or 100 Hz (UK), see Figs. 2.24*a*, *b* and 2.26*a*.

In order to establish the welding condition the operator was required to set four variables, namely peak current or voltage, pulse frequency, background current and wire feed speed. Hence, although pulse MIG was found to have significant benefits for many applications, especially aluminium and stainless steel, the complex nature of the process and high degree of skill required to establish the correct welding parameters, heavily limited its acceptance by the welding industry. Common complaints regarding the process included variation in transfer with 'stick out', narrow operating range for a given consumable, undercut and spatter. It is clear with these original units that the quantity of metal detached per pulse was directly related to wire feed speed (Fig. 2.26*b*).

Research work during the late 60s and early 70s had demonstrated that for each wire diameter/material combination there is a specific droplet size for optimum metal transfer. In fact, the diameter of this droplet is very nearly equal to the wire diameter. Further, for each wire, specific pulse parameters, i.e. peak current and pulse width, can be defined to detach such a droplet.

With the advent of electronically controlled and transistorized power sources, it became possible to generate wave shapes that were rectangular instead of sinusoidal (Fig. 2.26*c*). Equally, since the output was no longer synchronized to mains frequency, the pulse frequency could now be varied from well below 50 Hz to in excess of 300 Hz and functions like

Fig. 2.26(*a*)

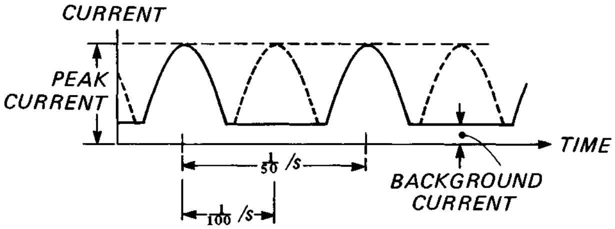
CURRENT
PEAK CURRENT
TIME
$\frac{1}{50}$ /s
$\frac{1}{100}$ /s
BACKGROUND CURRENT

Fig. 2.26(*b*)

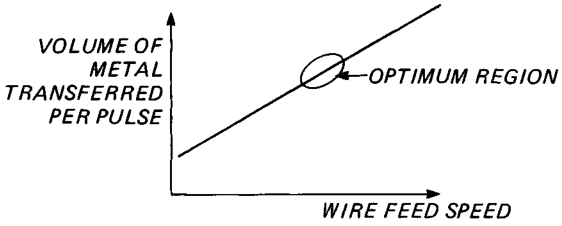

Fig. 2.26(*c*)

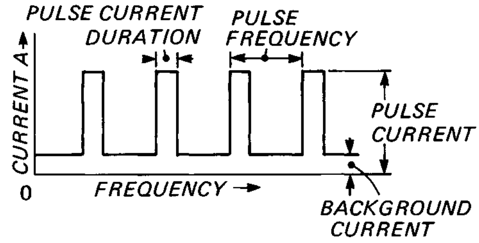
PULSE CURRENT DURATION
PULSE FREQUENCY
CURRENT A
PULSE CURRENT
0
FREQUENCY
BACKGROUND CURRENT

output volt–ampère relationships and pulse waveform could be accurately predicted and controlled. See also p. 160.

In the synergic (Gr. *syn*, together; *ergon*, work) pulse MIG process, the correct pulse parameters are selected, usually automatically, depending upon the wire diameter, material and shielding gas combination chosen. The pulse frequency and wire feed speed are electronically linked together so that as the wire speed is increased so the pulse frequency is increased and vice versa. A stream of droplets of constant size are transferred, one per pulse, with the transfer frequency linked directly to the wire feed speed and hence the average current demand (Fig. 2.26*d*). As a result, the operator is relieved of the onus of establishing the various parameters each time that the work is changed. When the torch switch is pressed, the average current demand set by the one-knob control is automatically translated into wire speed and pulse frequency information, the actual value of pulse frequency depending on the wire diameter selected. The correct value of inductance is also 'electronic' so that no inductance 'tappings' are required.

The pulsed current gives a projected spray type open-arc transfer over a wide range of currents and, in particular, at levels far below the normal dip-to-spray transition. This has the benefit of offering an all-positional

Fig. 2.26(*d*)

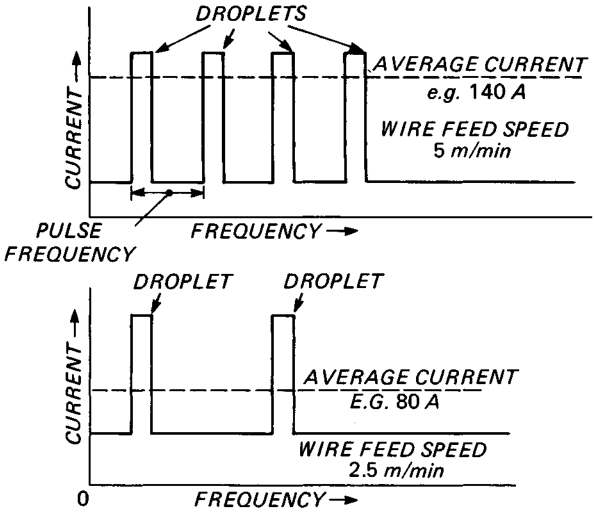
DROPLETS
AVERAGE CURRENT
e.g. 140 A
WIRE FEED SPEED
5 m/min
CURRENT →
PULSE FREQUENCY
FREQUENCY →
DROPLET
DROPLET
AVERAGE CURRENT
E.G. 80 A
WIRE FEED SPEED
2.5 m/min
CURRENT →
0
FREQUENCY →

PULSE FREQUENCY AND WIRE FEED RELATIONSHIP

INCREASE OF WIRE FEED (AVERAGE CURRENT) GIVES INCREASE OF PULSE FREQUENCY.

welding method with virtually no spatter, and the regularity of actual transfer of the droplet gives a smooth stable weld pool.

Most synergic machines offer the ability to run a wide range of wires including mild and stainless steels in synergic mode. Some models can operate in all three process modes (dip, pulse and spray transfer) with various shielding gases, using the one-knob control system. Most have an

Fig. 2.26. (*e*) Synergic 500 for MMA, TIG and MIG processes with dip/spray or pulse modes. Input 3-phase 380–415 V. Adjustment of background current for pulse MIG. KVA input 34. Auxiliary power 1 A at 220 V and 4 A at 42 V. Voltage trim/arc force controls permits arc length/arc voltage to be adjusted. Remote control if required. Inch button. Pulsed arc is very suitable for plate 2–6 mm thick, and gives good results on stainless steel with good alloy, recovers on alloy steels. Used with pure argon it is very suitable for aluminium welding. (See Appendix 10, p. 432, for photograph of unit.)

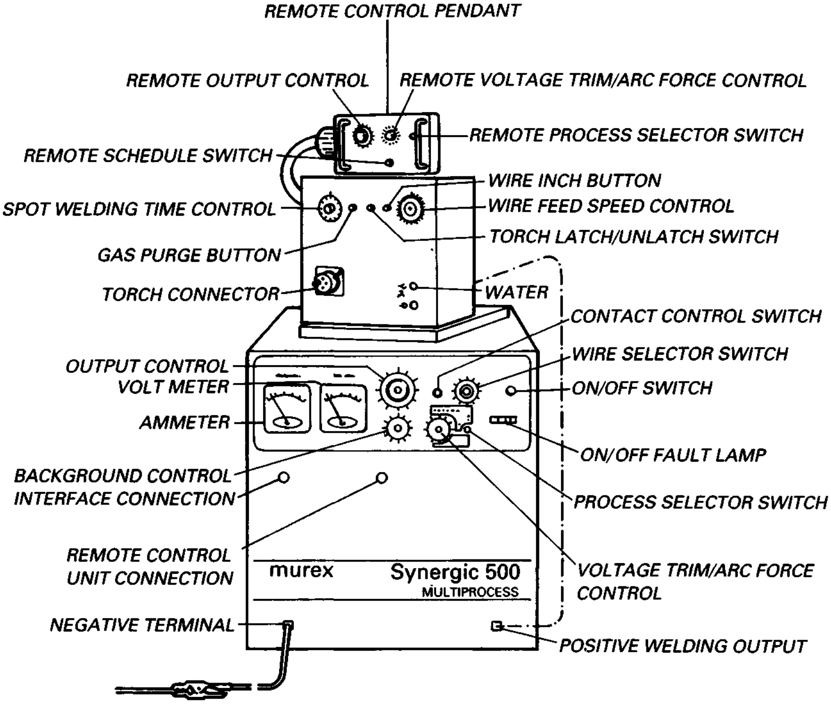
REMOTE CONTROL PENDANT
REMOTE OUTPUT CONTROL
REMOTE VOLTAGE TRIM/ARC FORCE CONTROL
REMOTE PROCESS SELECTOR SWITCH
REMOTE SCHEDULE SWITCH
WIRE INCH BUTTON
SPOT WELDING TIME CONTROL
WIRE FEED SPEED CONTROL
TORCH LATCH/UNLATCH SWITCH
GAS PURGE BUTTON
TORCH CONNECTOR
WATER
CONTACT CONTROL SWITCH
OUTPUT CONTROL
WIRE SELECTOR SWITCH
VOLT METER
ON/OFF SWITCH
AMMETER
ON/OFF FAULT LAMP
BACKGROUND CONTROL
INTERFACE CONNECTION
PROCESS SELECTOR SWITCH
REMOTE CONTROL UNIT CONNECTION
murex
Synergic 500
MULTIPROCESS
VOLTAGE TRIM/ARC FORCE CONTROL
NEGATIVE TERMINAL
POSITIVE WELDING OUTPUT

arc voltage trim function enabling the operator to adjust the arc length to suit specific needs.

Two such modern synergic units are shown in Figs. 2.26*e* and 2.26*f*.

Plasma MIG process. A process is being developed in which MIG filler wire is fed through a contact tube situated centrally in the torch head. A tungsten electrode is set at an angle to this contact tube and projects nearly to the mouth of the shielding nozzle through which the plasma gas issues. An outer nozzle provides a gas shield. The current through the wire assists

Fig. 2.26. (*f*) Ultra-arc 350 synergic transistorized pulsed MIG. Spray transfer one-knob synergic control of five parameters: pulse current, pulse time, background current, pulse frequency and wire feed speed. Input 200, 230 or 460 V 3-phase. Single current control 50–350 A. Constant energy CC and CV which reduces shrinkages and distortion. The constant energy is irrespective of the stick out and makes welding of difficult ferrous metals easier, smooth arc. Micro processor control has nine preset programs e.g. mild steel of two sizes using 95% Ar, 5% CO_2; stainless steel of three sizes using 98% Ar, 2% O_2; silicon bronze of two sizes using 100% Ar; and aluminium of two sizes using 100% Ar. (See Appendix 10, p. 433, for photograph of plant.)

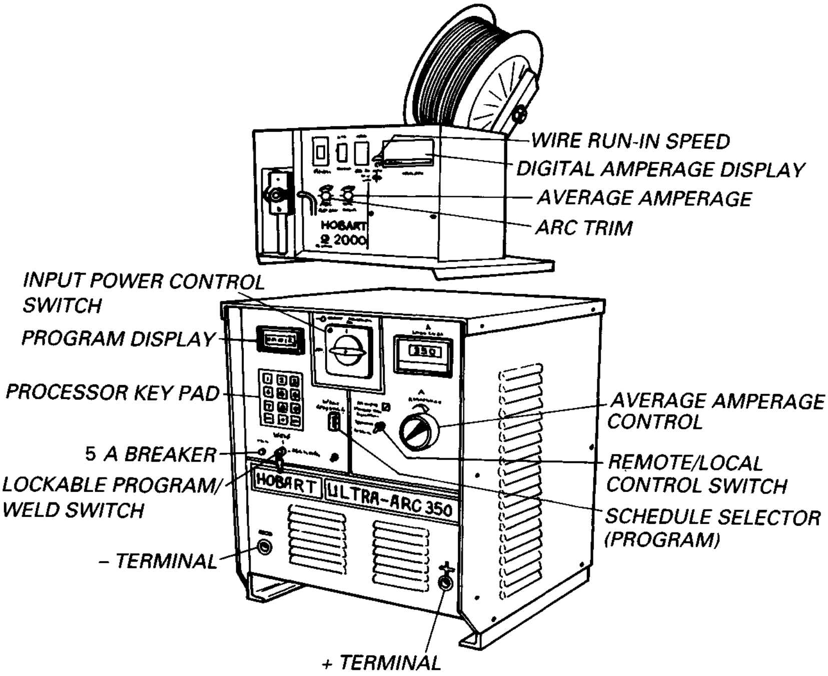
WIRE RUN-IN SPEED
DIGITAL AMPERAGE DISPLAY
AVERAGE AMPERAGE
ARC TRIM
HOBART
2000
INPUT POWER CONTROL SWITCH
PROGRAM DISPLAY
PROCESSOR KEY PAD
5 A BREAKER
LOCKABLE PROGRAM/ WELD SWITCH
- TERMINAL
350
HOBART
ULTRA-ARC 350
AVERAGE AMPERAGE CONTROL
REMOTE/LOCAL CONTROL SWITCH
SCHEDULE SELECTOR (PROGRAM)
+ TERMINAL

the melting and gives good starting characteristics and a stable arc, so that thin plate is weldable at high welding speeds.

Repair of cast iron by automated MIG process using tubular electrodes. With the advent of flux cored wire (2.10*a*) carbon can be added to the core and high speed welding performed using high nickel iron wire. As an example, nickel FC 55 is a tubular wire, the core being filled with carbon, slagging constituents and deoxidizers. It can be used with or without a shielding gas such as CO_2 and no post-weld heat treatment is required. Standard gas–metal arc equipment is used, the joints have high strength and the HAZ is free from the carbide complex which can cause embrittlement. It makes possible the automatic welding of cast irons to themselves or to stainless steel or high nickel alloys and can be used for overlaying and metal arc spot welding.

Fig. 2.27. Gas mixing system for mixing shielding gases.

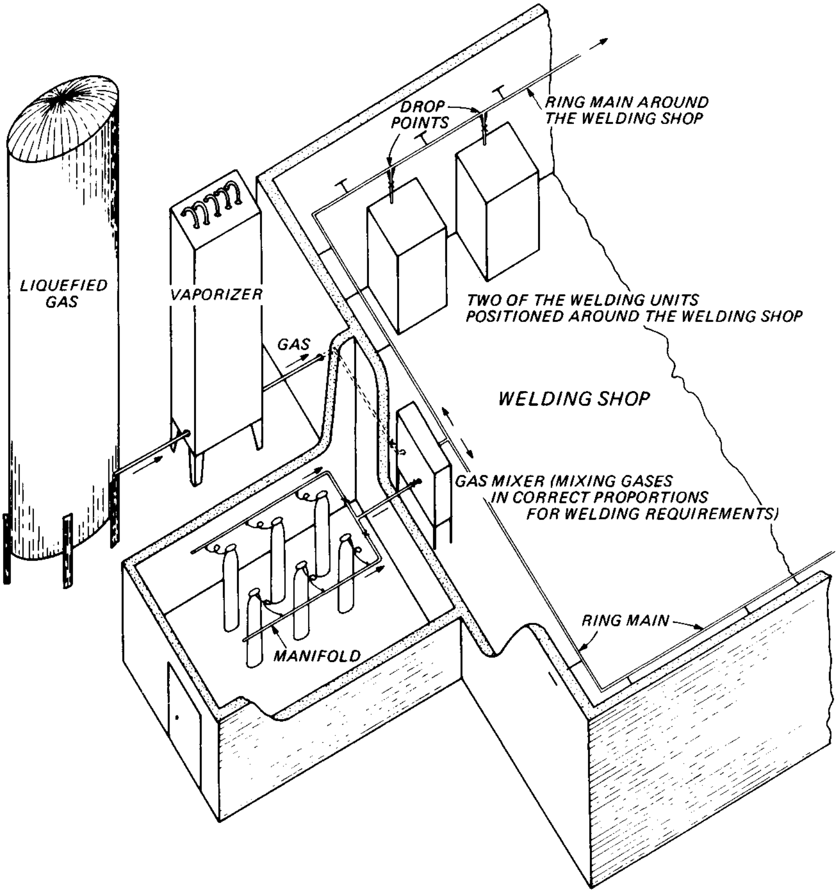
DROP POINTS
RING MAIN AROUND THE WELDING SHOP
LIQUEFIED GAS
VAPORIZER
TWO OF THE WELDING UNITS POSITIONED AROUND THE WELDING SHOP
GAS
WELDING SHOP
GAS MIXER (MIXING GASES IN CORRECT PROPORTIONS FOR WELDING REQUIREMENTS)
RING MAIN
MANIFOLD

Liquid (gas) storage and gas mixing systems

The gas mixer (Fig. 2.27) mixes the gases in proportions suitable for welding. The proportion of each gas can be adjusted and the panel maintains this proportion over a wide range of flow requirements. A typical example is the mixing of argon and carbon dioxide as in the Coogar or Argoshield range (see appendix 3).

3

*Tungsten electrode, inert gas shielded welding processes (TIG), and the plasma arc process**

Technology and equipment

The welding of aluminium and magnesium alloys by the oxy-acetylene and manual metal arc processes is limited by the necessity to use a corrosive flux. The gas shielded, tungsten arc process (Fig. 3.1) enables these metals and a wide range of ferrous alloys to be welded without the use of the flux. The choice of either a.c. or d.c. depends upon the metal to be welded. for metals having refractory surface oxides such as aluminium and its alloys, magnesium alloys and aluminium bronze, a.c. is used whilst d.c. is used for carbon and alloy steels, heat-resistant and stainless steels, copper and its alloys, nickel and its alloys, titanium, zirconium and silver.

Fig. 3.1. Connexions for inert gas welding using air-cooled torch.

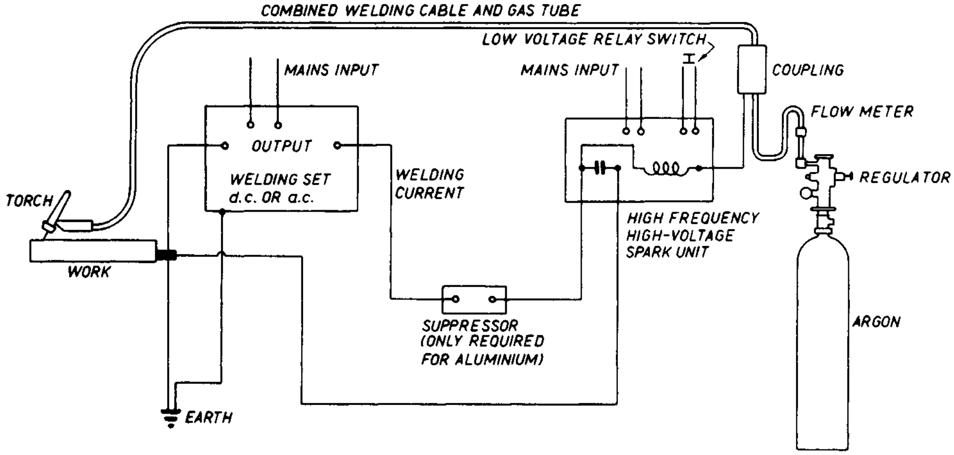
COMBINED WELDING CABLE AND GAS TUBE
LOW VOLTAGE RELAY SWITCH
MAINS INPUT
MAINS INPUT
COUPLING
FLOW METER
OUTPUT
WELDING SET
d.c. OR a.c.
WELDING CURRENT
REGULATOR
TORCH
WORK
HIGH FREQUENCY
HIGH-VOLTAGE
SPARK UNIT
SUPPRESSOR
(ONLY REQUIRED
FOR ALUMINIUM)
ARGON
EARTH

---

*American designation: gas tungsten arc welding (GTAW).
See also BS 3019 TIG welding, Parts 1 and 2.

(See also BS 3019, *General recommendations for manual inert gas tungsten arc welding*. Part 1, *Wrought aluminium, aluminium alloys and magnesium alloys*; Part 2, *Austenitic stainless and heat resisting steels.*)

The arc burns between a tungsten electrode and the workpiece within a shield of the inert gas argon, which excludes the atmosphere and prevents contamination of electrode and molten metal. The hot tungsten arc ionizes argon atoms within the shield to form a gas plasma consisting of almost equal numbers of free electrons and positive ions. Unlike the electrode in the manual metal arc process, the tungsten is not transferred to the work and evaporates very slowly, being classed as 'non-consumable'. Small amounts of other elements are added to the tungsten to improve electron emission. It is, however, a relatively slow method of welding.

Gases

Argon in its commercial purity state (99.996%) is used for metals named above, but for titanium extreme purity is required. Argon with 5% hydrogen gives increased welding speed and/or penetration in the welding of stainless steel and nickel alloys; nitrogen can be used for copper welding on deoxidized coppers only. Helium may be used for aluminium and its alloys and copper, but it is more expensive than argon and, due to its lower density, a greater volume is required than with argon to ensure adequate shielding, and small variations in arc length cause greater changes in weld conditions. A mixture of 30% helium and 70% argon is now used, and gives fast welding speeds. The mechanized d.c. welding of aluminium with helium gives deep penetration and high speeds.

The characteristics of the arc are changed considerably with change of direction of flow of current, that is with arc polarity.

Electrode positive

The electron stream is from work to electrode while the heavier positive ions travel from electrode to work-piece (Fig. 3.2*a*). If the work is of aluminium or magnesium alloys there is always a thin layer of refractory oxide of melting point about 2000°C present over the surface and which has to be dispersed in other processes by means of a corrosive flux to ensure weldability. The positive ions in the TIG arc bombard this oxide and, together with the electron emission from the plate, break up and disperse the oxide film. It is this characteristic which has made the process so successful for the welding of the light alloys. The electrons streaming to the tungsten electrode generate great heat, so its diameter must be relatively large and it forms a bulbous end. It is this overheating with consequent vaporization of the tungsten and the possibility of tungsten being transferred to the molten pool (pick-up) and contaminating it that is the drawback to the use of the process with electrode positive. Very much less heat is generated at the molten pool and this is therefore wide and shallow.

Electrode negative

The electron stream is now from electrode to work with the zone of greatest heat concentrated in the workpiece so that penetration is deep and the pool is narrower. The ion flow is from work to electrode so that there is no dispersal of oxide film and this polarity cannot be used for welding the light alloys. The electrode is now near the zone of lesser heat and needs be of reduced diameter compared with that with positive polarity. For a given diameter the electrode, when negative, will carry from four to eight times

Fig. 3.2. Electron streams between electrode and work: (*a*) d.c., tungsten electrode +ve of large diameter tends to overheat; (*b*) d.c., tungsten electrode −ve of small diameter; (*c*) a.c., electrode diameter between that of electrode +ve and −ve electrode.

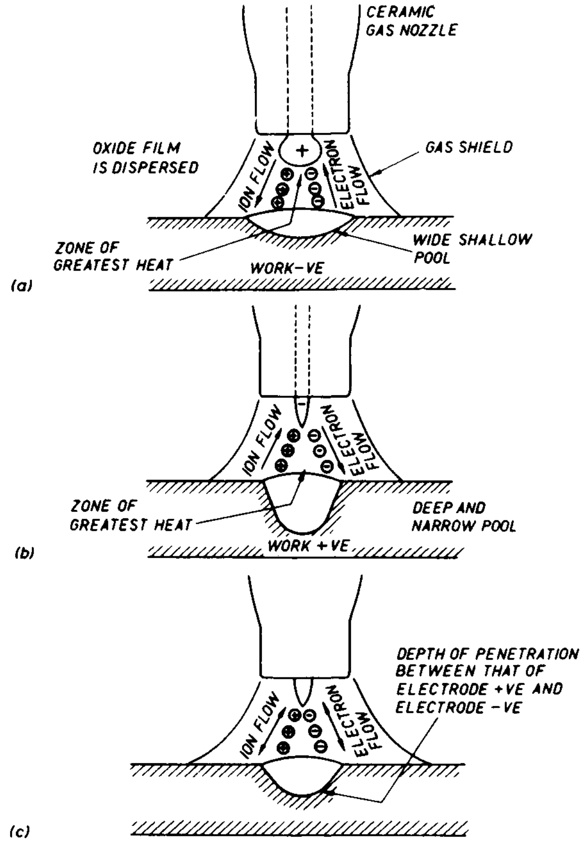
CERAMIC GAS NOZZLE
OXIDE FILM IS DISPERSED
GAS SHIELD
+
ION FLOW
ELECTRON FLOW
ZONE OF GREATEST HEAT
WORK-VE
WIDE SHALLOW POOL
(a)
−
ION FLOW
ELECTRON FLOW
ZONE OF GREATEST HEAT
DEEP AND NARROW POOL
WORK +VE
(b)
ION FLOW
ELECTRON FLOW
DEPTH OF PENETRATION BETWEEN THAT OF ELECTRODE +VE AND ELECTRODE −VE
(c)

the current than when it is positive and twice as much as when a.c. is used. (Fig. 3.2*b*.)

Alternating current

When a.c. is used on a 50 Hz supply, voltage and current are reversing direction 100 times a second so that there is a state of affairs between that of electrode positive and electrode negative, the heat being fairly evenly distributed between electrode and work (Fig. 3.2*c*). Depth of penetration is between that of electrode positive and electrode negative and the electrode diameter is between the previous diameters. When the electrode is positive it is termed the positive half-cycle and when negative the negative half-cycle. Oxide removal takes place on the positive half-cycle.

See note on square wave equipment and wave balance control (pp. 160–3).

Inherent rectification in the a.c. arc

In the a.c. arc the current in the positive half-cycle is less than that in the negative half-cycle (Fig. 3.3*b*). This is known as inherent rectification and is a characteristic of arcs between dissimilar metals such as tungsten and aluminium. It is due to the layer of oxide acting as a barrier layer to the current flowing in one direction and to the greater emission of electrons from the tungsten electrode when it is of negative polarity. The result of this imbalance is that an excess pulsating current flows in one direction only and the unbalanced wave can be considered as a balanced a.c. wave, plus an excess pulsating current flowing in one direction only on the negative half-cycle. This latter is known as the d.c. component and can be measured with

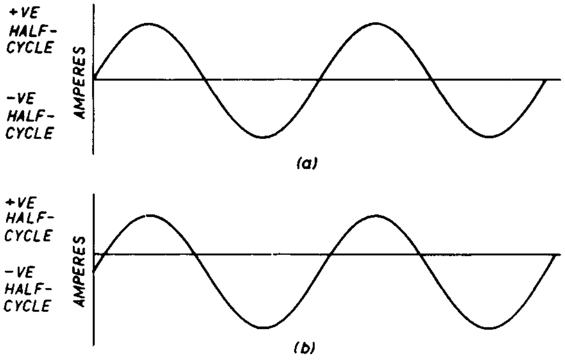

(a)

(b)

Fig. 3.3. Alternating current: (*a*) balanced wave; (*b*) unbalanced wave, inherent rectification (current in +ve half-cycle less than that in −ve half cycle).

a d.c. ammeter. (The suppression of this d.c. component is discussed later.) The reduction of current in the positive half-cycle due to the inherent rectification results in a reduction of oxide removal.

Partial rectification

A greater voltage is required to strike the arc than to maintain it and re-ignition on the negative half-cycle requires a lower voltage than for the positive half-cycle, partly due to the greater electron emission from the tungsten when it is negative polarity, but actual re-ignition depends upon many factors including the surface condition of the weld pool and electrode, the temperature of the pool and the type of shielding gas. There may be a delay in an arc re-ignition on the positive half-cycle until sufficient voltage is available and this will result in a short period of zero current (Fig. 3.4*a*) until the arc ignites. This delay reduces the current in the positive half-cycle and this state is known as partial rectification. If the available voltage is not sufficient, ignition of the arc may not occur at all on the positive half-cycle, the arc is extinguished on the one half-pulse and continues burning on the uni-directional pulses of the negative half-cycle, and we have complete rectification with gradual extinguishing of the arc (Fig. 3.4*b*).

Re-ignition voltages

To ensure re-ignition of the arc on the positive half-cycle, the available voltage should be of the order of 150 V, which is greater than that of the supply transformer. To ensure re-ignition, auxiliary devices are used which obviate the need for high open-circuit transformer voltages.

Fig. 3.4. (*a*) Partial rectification; (*b*) half-wave rectification (+ve half-cycle misssing).

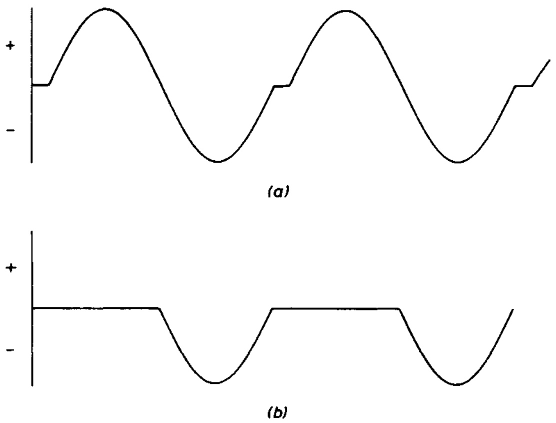

Ignition and re-ignition equipment*

High-frequency, high-voltage, spark gap oscillator. This device enables the arc to be ignited without touching down the electrode on the work and thus it prevents electrode contamination. It also helps arc re-ignition at the beginning of the positive half-cycle.

The oscillator consists of an iron-cored transformer with a high voltage secondary winding, a capacitor, a spark gap and an air core transformer or inductive circuit, one coil of which is in the high voltage circuit and the other in the welding circuit (Fig. 3.5). The capacitor is charged every half cycle to 3000–5000 V and discharges across the spark gap. The discharge is oscillatory, that is, it is not a single spark but a series of sparks oscillating across the spark gap during discharge. This discharge, occurring on every half-cycle, sets up oscillatory currents in the circuit and these are induced and superimposed on the welding current through the inductance of the coils *L* (Fig. 3.6).

The spark discharge is phased to occur at the beginning of each half-cycle

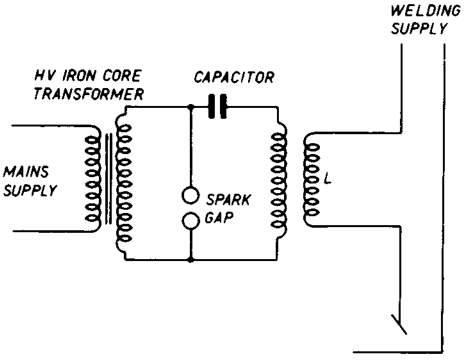

Fig. 3.5. (*left*) High-frequency spark oscillator.

Fig. 3.6. (*below*) Arc voltage with superimposed HF spark main for re-ignition and stabilization; the HF is injected on both +ve and −ve half-cycles, but is only required at the beginning of the former.

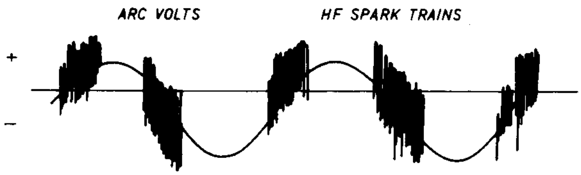

---

* This equipment produces radio-frequency interference (RFI). EEC regulations will, in future, prohibit the manufacture and use of this sort of equipment for starting and maintaining the TIG arc because of interference with computers, electronic systems, etc. Consequently, non-HF starting and stabilization of the TIG arc, such as lift arc (LIFTIG), pp. 153–4, is necessary. Square wave a.c. does not allow the arc to be extinguished at each cycle changeover and should be used for welding aluminium where necessary with a.c. (p. 160). Use of the inverter (Vol. 1, ch. 5) gives a high frequency a.c. with smooth, stable arc and no additional HF stablization.

(although for re-ignition purposes it is required only for the positive half-cycle) and is of about 5 milliseconds or less duration compared with the half-cycle duration of 10 milliseconds (Fig. 3.7). To initiate the arc, the electrode is brought to about 6 mm from the work with the HF unit and welding current switched on. Groups of sparks pass across the gap, ionizing it, and the welding current flows in the form of an arc without contamination of the electrode by touching down. The HF unit can give rise to considerable radio and TV interference and adequate suppression and screening must be provided to eliminate this as far as possible. The use of HF stabilization with the a.c. arc enables this method to be used for aluminium welding, although inherent rectification is still present, but partial rectification can be reduced to a minimum by correct phasing of the spark train.

Scratch start. Some units do not employ an HF unit for starting the arc without touch down. The tungsten is scratched momentarily onto the work and the arc is ionized. Naturally there will be some very small tungsten contamination but this will not greatly interfere with the mechanical properties of the joint in non-critical conditions.

Lift start. In this method no HF unit is used, the tungsten electrode is placed down upon the work where the welding is to begin. The operator presses a switch which connects it to the electronic control but, as yet, no current flows.

The torch is then lifted a little from the work (or tilted on the gas nozzle), the arc is struck as the electrode makes a gap from the work and it then slopes up to the values set for the work in hand. Upon releasing the switch the current slopes down to zero and thus prevents the formation of a finishing crater. With this method there is no HF interference with any other HF apparatus.

Surge injection unit. This device supplies a single pulse surge of about 300 V phased to occur at the point when the negative half-cycle changes to the

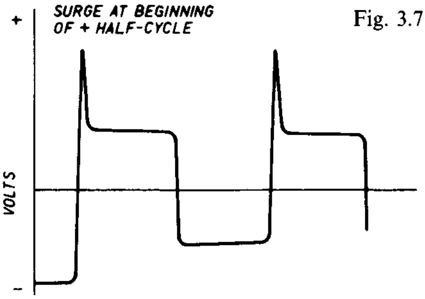

Fig. 3.7. Surge voltage superimposed on arc voltage.

positive half-cycle (Fig. 3.7), so that the pulse occurs at 20 millisecond intervals and lasts for only a few microseconds. This unit enables transformers of 80 V open-circuit voltage to be used and it does not produce the high frequency radiation of the spark oscillator and thus does not interfere to any extent with radio and TV apparatus.

The unit consists of a rectifier which supplies d.c. to a circuit containing resistance and capacitance, a surge valve which supplies the short pulses and a trigger valve which releases the pulses into the welding circuit. The trigger valve, which is sensitive to change in arc voltage, releases the pulse at the end of the negative half-cycle, just when the positive half-cycle is beginning. As the pulse is usually unable to initiate the arc from cold, a spark oscillator is also included in the unit and is cut out of circuit automatically when the arc is established so that the surge injection unit is an alternative method to the HF oscillator for re-igniting the arc on the positive half-cycle (Fig. 3.8).

Suppression of the d.c. component in the a.c. arc

In spite of the use of the HF unit, the imbalance between the positive and negative half-cycles remains and there is a d.c. component flowing. This direct current flows through the transformer winding and saturates the iron core magnetically, giving rise to high primary currents with such heating effect that the rating of the transformer is lowered, that is, it cannot supply its rated output without overheating. The insertion of banks of electrolytic capacitors in series with the welding circuit has two effects:

(1) They offer low impedence to the a.c. which flows practically uninterrupted, but they offer very high impedance to the d.c. which is therefore suppressed or blocked.

(2) During the negative half-cycle the capacitors receive a greater charge (because of the imbalance) and during the following

Fig. 3.8. Current and arc voltage wave-forms showing re-ignition.

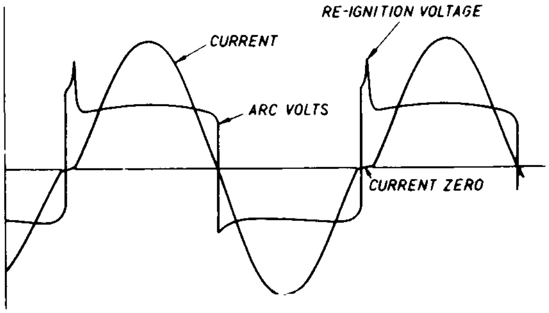

positive half-cycle this excess adds to the positive half-cycle voltage so that it is increased, and if the open-circuit voltage is greater than about 100 V the arc is re-ignited on each half-cycle without the aid of high-frequency currents, which can then be used for starting only. The effect of this increase in voltage on the positive half-cycle is to improve the balance of the wave (Fig. 3.8), so that the heating effect between electrode and plate is more equal and disposal of the oxide film is increased.

The value of the capacitor must be chosen so that there is no danger of electrical resonance in the circuit which contains resistance, inductance and capacitance, which would result in dangerous excessive currents and voltages irrespective of the transformer output (Fig. 3.9*a*). If external capacitors are used they may affect the current output of the transformer by altering the power factor, and this must be taken into account since the current calibrations on the set will be increased.

The cost of large banks of capacitors however is considerable and is accompanied by a high voltage across them, so that this method has been largely superseded in the units which use solid state technology and thyristor control (see later).

Power sources a.c. and d.c.

Equipment can be chosen to give a.c. or d.c. or both a.c. and d.c. from one unit and may even be designed for specific industries: (1) d.c. output for the fabrication of a variety of steels and special steels such as stainless, heat resistant and 9% Ni, etc., (2) a.c. output for fabrication aluminium and its alloys, (3) a.c. and d.c., which includes the fabrication of both the above ferrous and non-ferrous metals and it is this type that covers most types of fabrication.

Power unit a.c. For the light alloys of aluminium and magnesium a transformer similar to that used for MMA welding can be used. Cooling can be by forced draught or oil, usually the former, and primary tappings are provided on the input side for single- or three-phase 50 Hz supply, with an output voltage of about 80 V. Current control can be by tapped choke or saturable reactor and auxiliary units such as HF oscillator, d.c. component suppressor, and surge injector can be built in or fitted externally. Scratch start obviates the use of the HF oscillator but has certain disadvantages.

Power unit d.c. These units consist of a step-down transformer with input from single- or three-phase 50 Hz mains which feeds into a thyristor or silicon controlled rectifier (SCR) which gives a stepless control of current and obviates the use of other methods of current control. On larger units there may be a two position switch for high or low current ranges, and these are usually constant voltage units.

Power units, d.c. and a.c. These units, which supply either a.c. or d.c., are connected to single- or three-phase 50 Hz mains, fed into a step-down transformer and then into a thyristor which may also act as a contactor so that there are no mechanical parts to open and close when welding is begun and ended.

The thyristors, which can be switched under load, are fitted to a heat sink and the whole is forced draught cooled. Current can be supplied for MMA and TIG and output can be for electrode +ve or −ve or a.c. by means of the controlling printed circuit.

The auxiliary supply is at a lower voltage of about 110 V and there is automatic regulation for variation of the mains voltage. In addition there can be burn back control, soft start, pulse and spot welding facilities and the units are suitable for using hard or soft solid core wires and flux cored wire using either short circuit or spray transfer.*

Duty cycle

Over a period of 10 minutes, a 60% duty cycle means that the unit can be used at that particular current (about 275 A in Fig. 3.9*b*) for 6 minutes and then given 4 minutes to cool down. Increasing the welding current above that recommended for the particular duty cycle increases the quantity of heat evolved and the excessive temperature rise can lead to failure. Decreasing the welding current increases the period within the 10 minutes for which the unit can be operated. Thermal cut-outs may be fitted which trip and prevent serious damage while thyristors are bolted to a heat sink and the whole unit force draught cooled.

Exceeding the duty cycle value of current can cause overheating and may damage the unit.

Fig. 3.9. (*a*) Series capacitor used to block d.c. component.

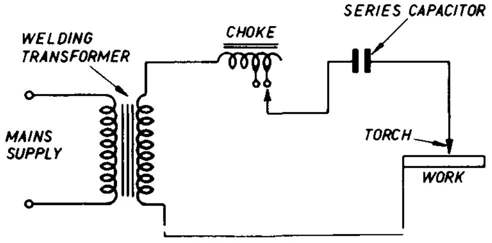

---

* A typical formula for obtaining the voltage drop V, across the arc when a current A is flowing is: $V = 10+(0.04A)$. Thus a current of 70 A gives a drop of ~12.8 V.

Slope-in (slope-up) and soft start. The use of normal welding current at the beginning of the weld often causes burn-through and increases the risk of tungsten contamination. To prevent this a slope-in control is provided which gives a controlled rise of the current to its normal value over a preselected period of time, which can be from 0 to 10 seconds. The soft start control is available on some equipment and performs the same function but it has a fixed time of operation and can be switched in or out as required.

Slope-out (slope-down) (crater fill). The crater which would normally form at a weld termination can be filled to the correct amount with the use of this control. The crater-filling device reduces the current from that used for the welding operation to the minimum that the equipment can supply in a series of steps. The slope-out control performs the same function, but the current is reduced over a period of about 20 seconds depending upon the setting of the control (not steps). The tendency to cracking is reduced by the use of this control (Fig. 3.10).

Fig. 3.9. (*b*) Illustrating how the welding current decreases as the % duty cycle of a power unit increases.

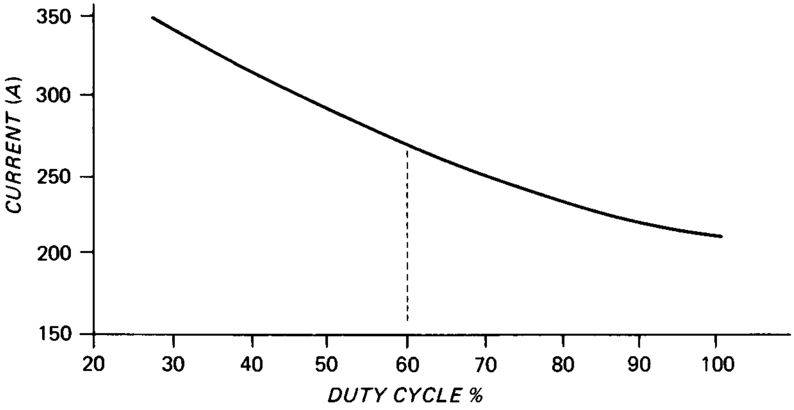

Fig. 3.10

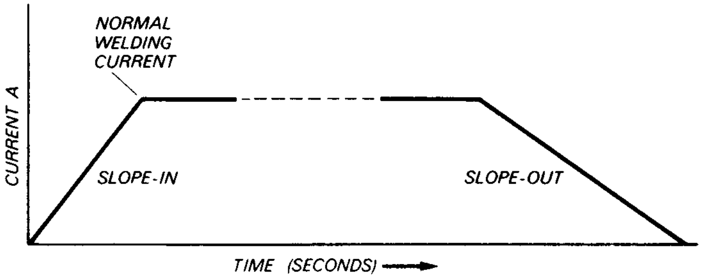

Thyristor or silicon controlled rectifier (SCR) control of welding current

The value of welding current required is set by the single dial or knob current control on the front panel of the unit. This value is compared by the control circuit (p.c.b.) with the value of output current received from the sensor in the outgoing welding line and the firing angle of the SCR's is altered to bring the output current to the value set by the operator on the control dial. Hence stepless control is achieved (Fig. 3.11*a*, *b*). A simple

Fig. 3.11. (*a*) For a.c. or d.c. TIG, a.c./d.c. low frequency pulsed TIG, a.c. or d.c. middle frequency pulsed d.c. TIG spot welding and d.c. MMA. 380–415 V 3-phase 50–60 Hz. Duty cycle 40% (10 min). Currents: d.c. TIG 5–300 A, a.c. TIG 10–300 A, d.c. MMA 10–250 A. Crater filling. Up and down slope 0–1.5 A. Pre and after flow gas. Cleaning effect control adjustment ratio 10–50%. Remote control. Weight 40 kg. HF or lift start.

WARNING
POWER
HITACHI
Inverter Pair 300GP
INVERTER PAIR 300 REMOTE CONTROL BOX
WELDING CURRENT(A)
(PULSE CURRENT(A))
BASE CURRENT(A)
AD-GP300

Fig. 3.11. (*b*) 250 a.c./d.c. TIG MMA square wave with OC/CV feeder. Available voltages 230/380/415/500 a.c. 50 Hz OCV 80. Welding mode switch 5–310 A, a.c. or d.c.e.n., d.c.e.p. Pre and post flow gas 250 A at 40% duty cycle. Balance control penetration/cleaning. Arc force control (soft or digging arc). (See Appendix 10, p. 436, for photograph of unit.)

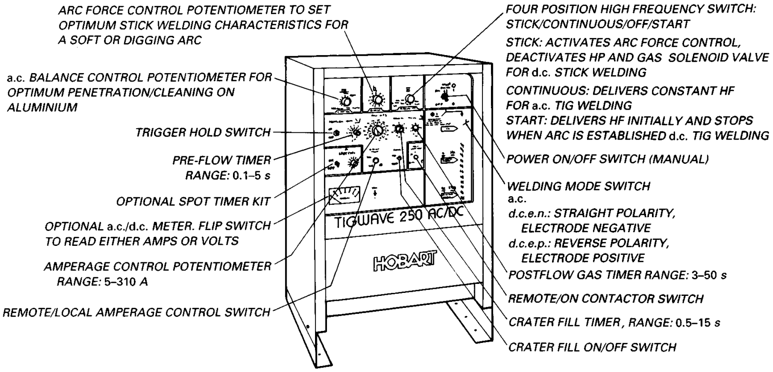
ARC FORCE CONTROL POTENTIOMETER TO SET OPTIMUM STICK WELDING CHARACTERISTICS FOR A SOFT OR DIGGING ARC
a.c. BALANCE CONTROL POTENTIOMETER FOR OPTIMUM PENETRATION/CLEANING ON ALUMINIUM
TRIGGER HOLD SWITCH
PRE-FLOW TIMER RANGE: 0.1–5 s
OPTIONAL SPOT TIMER KIT
OPTIONAL a.c./d.c. METER. FLIP SWITCH TO READ EITHER AMPS OR VOLTS
AMPERAGE CONTROL POTENTIOMETER RANGE: 5–310 A
REMOTE/LOCAL AMPERAGE CONTROL SWITCH
FOUR POSITION HIGH FREQUENCY SWITCH: STICK/CONTINUOUS/OFF/START
STICK: ACTIVATES ARC FORCE CONTROL, DEACTIVATES HP AND GAS SOLENOID VALVE FOR d.c. STICK WELDING
CONTINUOUS: DELIVERS CONSTANT HF FOR a.c. TIG WELDING
START: DELIVERS HF INITIALLY AND STOPS WHEN ARC IS ESTABLISHED d.c. TIG WELDING
POWER ON/OFF SWITCH (MANUAL)
WELDING MODE SWITCH
a.c.
d.c.e.n.: STRAIGHT POLARITY, ELECTRODE NEGATIVE
d.c.e.p.: REVERSE POLARITY, ELECTRODE POSITIVE
POSTFLOW GAS TIMER RANGE: 3–50 s
REMOTE/ON CONTACTOR SWITCH
CRATER FILL TIMER, RANGE: 0.5–15 s
CRATER FILL ON/OFF SWITCH
TIGWAVE 250 AC/DC
HOBART

explanation of the operation of an SCR, with diagrams, is given in appendix 2.

Square wave output units (a.c./d.c.)

Power units are now available in which the voltage and current waves are not sinusoidal, but have been modified by solid state technology and using printed circuit boards (p.c.b.'s) to give a very rapid rise of the a.c. wave from zero value to maximum value to give a 'square wave' output (Fig. 3.12).

When TIG welding aluminium using a sinusoidal wave form current the arc tends to become unstable and the electrode is easily overloaded. This gives tungsten inclusions in the weld (spitting) and leads to faults in the weld bead and more rapid consumption of the tungsten electrode. Square wave current overcomes these drawbacks and the arc is greatly stabilised and risk of inclusions greatly reduced. In addition MMA welding characteristics are greatly improved and the arc is smooth with reduced spatter.

The units are designed for precision a.c. TIG welding of aluminium, etc., and for d.c. TIG and manual metal arc welding. They have a transformer and a silicon bridge rectifier, SCR or thyristor with a square wave output. A memory core stores energy proportional to the previous half-cycle and then injects it into the circuit just as the wave passes through zero at the beginning of the next half-cycle. The rapid rise of the wave from zero to maximum, of about 80 microseconds from peak to peak (Figs. 3.13 and 14) means that the high frequency and high voltage at the beginning of the cycle needed to initiate the arc and often to keep it ionized without touchdown, need only be applied when the arc is first initiated. After this first initiation the HF is switched off automatically and the rapid rise of voltage from peak

Fig. 3.12. a.c. square wave output.

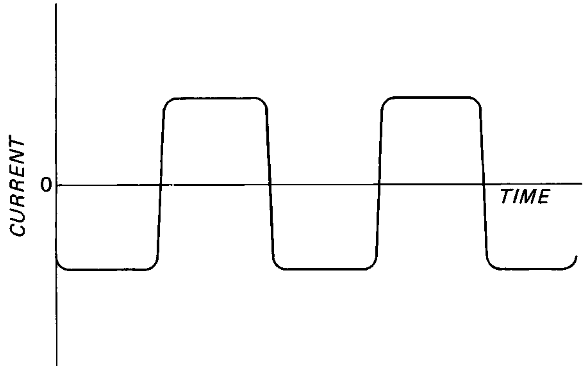

to peak keeps the arc ionized without application of the HF. A switch enables this HF to be used continuously if required although its continuous use may interfere with radio frequency apparatus in the vicinity.

Variation in mains voltage is compensated automatically, there is provision for MMA, pulse and spot welding and often there may be two ranges of current, either of which can be controlled by the one knob control (Fig. 3.13).

Soft start (slope-up) and crater fill (slope-down) are provided with pre- and post-gas and water supply controls. Pulse duration, height and background and wave balance controls are also fitted.

Fig. 3.13. Syncrowave 300 and 500 A square wave a.c., d.c., power unit for TIG and MMA welding. Input 220–576 V as required, mains voltage compensated. 500 A model gives 500 A at 100% duty cycle. Solid state contactors, slope-in and slope-out. Pulsed mode for improved penetration and control of weld pool. Rectification by SCR's with one knob control of welding current. 300 A model has dual current range, 5–75 A and 15–375 A, 500 A model has single range 25–625 A. Square wave output has wave balance control. Post timer for gas and water. Slope-up and slope-down (crater fill).

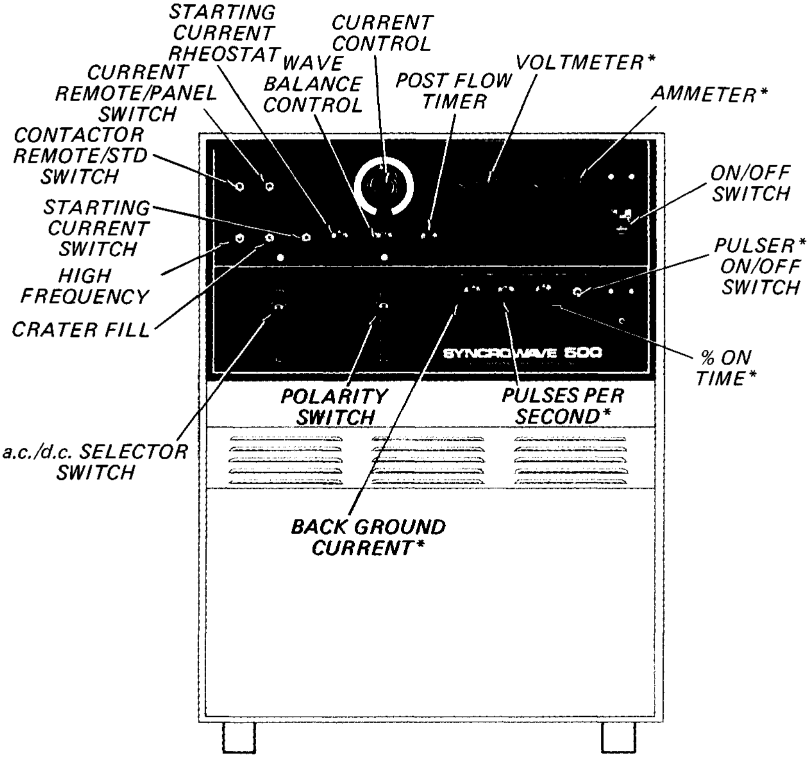
STARTING CURRENT RHEOSTAT
CURRENT CONTROL
CURRENT REMOTE/PANEL SWITCH
WAVE BALANCE CONTROL
POST FLOW TIMER
VOLTMETER*
AMMETER*
CONTACTOR REMOTE/STD SWITCH
ON/OFF SWITCH
STARTING CURRENT SWITCH
PULSER* ON/OFF SWITCH
HIGH FREQUENCY
CRATER FILL
SYNCROWAVE 500
% ON TIME*
POLARITY SWITCH
PULSES PER SECOND*
a.c./d.c. SELECTOR SWITCH
BACK GROUND CURRENT*

Wave balance control a.c.

On square wave equipment the wave balance control enables either tungsten electrode or work to be biased as required. Fig. 3.15 indicates how this is used for controlling penetration and cleaning action. (1) is a balanced wave. Moving the control to one side gives a wave balance (2) in which the tungsten electrode is positive polarity for a larger period,

Fig. 3.14. Square wave a.c. and d.c. 260 A (also a 360 A model) for a.c. and d.c. TIG and MMA, 50 Hz. Input 3-phase 380–415 V (or single phase 220 V, 200 A model). Output 10–375 A, 30% duty cycle, 375 A, 35 V; 60% cycle, 300 A, 32 V 100%; cycle 235 A, 29 V. Two current ranges, minimum current 10 A. Pre and post flow gas, slope up and slope down, spot timer. 80 OCV. Start lift arc and HF. TIG switch latching torch. Remote and pulse control optional extras. (See Appendix 10, p. 417, for photograph of unit.)

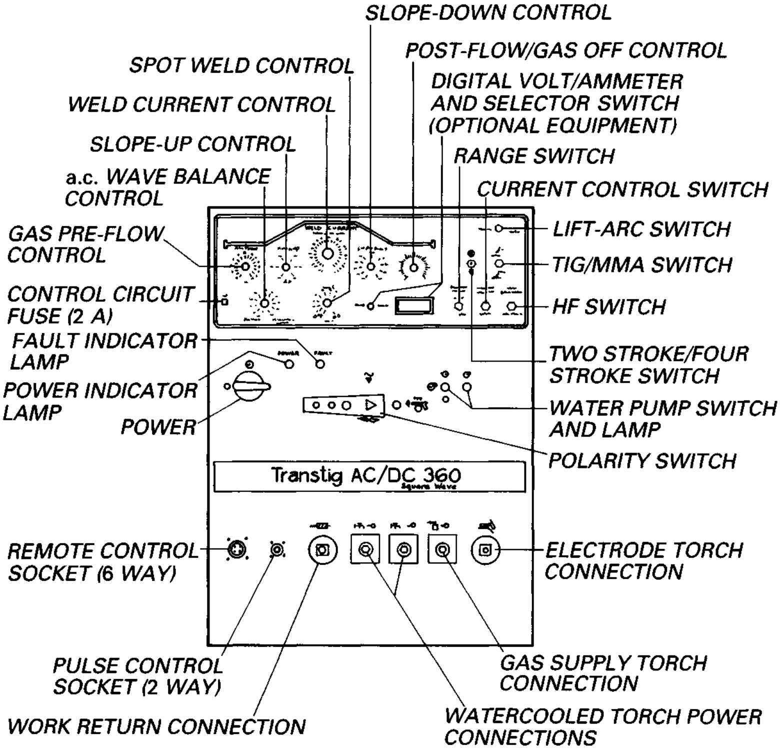
SLOPE-DOWN CONTROL
SPOT WELD CONTROL
POST-FLOW/GAS OFF CONTROL
DIGITAL VOLT/AMMETER AND SELECTOR SWITCH (OPTIONAL EQUIPMENT)
WELD CURRENT CONTROL
SLOPE-UP CONTROL
RANGE SWITCH
a.c. WAVE BALANCE CONTROL
CURRENT CONTROL SWITCH
GAS PRE-FLOW CONTROL
LIFT-ARC SWITCH
TIG/MMA SWITCH
CONTROL CIRCUIT FUSE (2 A)
HF SWITCH
FAULT INDICATOR LAMP
TWO STROKE/FOUR STROKE SWITCH
POWER INDICATOR LAMP
WATER PUMP SWITCH AND LAMP
POWER
POLARITY SWITCH
Transtig AC/DC 360 Square Wave
REMOTE CONTROL SOCKET (6 WAY)
ELECTRODE TORCH CONNECTION
PULSE CONTROL SOCKET (2 WAY)
GAS SUPPLY TORCH CONNECTION
WORK RETURN CONNECTION
WATERCOOLED TORCH POWER CONNECTIONS

giving maximum cleaning but with minimum penetration. With the control in the other direction the greatest heat is in the work so that we have maximum penetration but with minimum cleaning (3). With d.c. TIG welding as with MMA the control is set for a balanced wave.

Electrical contactors

The contactors control the various circuits for welding, inert gas, water flow and ancillary equipment. The contactor control voltage is of the order of 25–45 V and if the TIG head is machine mounted, a 110 V supply for the wire feed and tractor is provided. The post-weld argon flow contactor allows the arc to be extinguished without removing the torch with its argon shield from the hot weld area, thus safeguarding the hot electrode end and weld area from contamination whilst cooling. An argon–water purge is also provided and contactors may be foot-switch operated for convenience. Where the ancillary equipment is built into one unit an off-manual metal arc–TIG switch enables the unit to be used for either process.

With the advanced SCR method of control the SCR's act as a contactor so that there are no mechanical contactors in the welding current circuit. Cooling is by heat sink and forced draught and this method is very reliable since there are no moving parts.

Gas regulator, flowmeter and economizer

The gas regulator reduces the pressure in the argon cylinder from 175 or 200 bar down to 0–3.5 bar for supply to the torch (Fig. 3.16*a* and *b*). The flowmeter, which has a manually operated needle valve, controls the argon flow from 0–600 litres/hour to 0–2100 litres/hour according to type (Fig. 3.16*c*).

The economizer may be fitted in a convenient position near the welder and when the torch is hung from the projecting lever on the unit, argon gas and (if fitted) water supplies are cut off. A micro-switch operated by the lever can also be used to control the HF unit.

Fig. 3.15. Wave balance control.

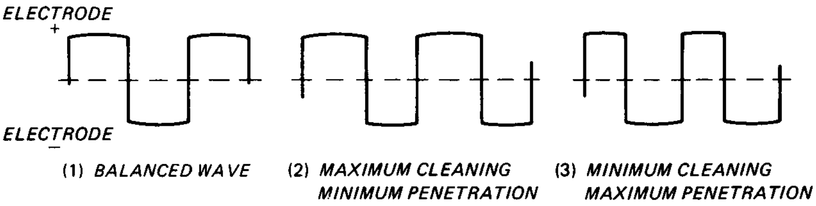
ELECTRODE +
ELECTRODE −
(1) BALANCED WAVE
(2) MAXIMUM CLEANING MINIMUM PENETRATION
(3) MINIMUM CLEANING MAXIMUM PENETRATION

Fig. 3.16. (*a*) Single-stage preset argon regulator and flowmeter.

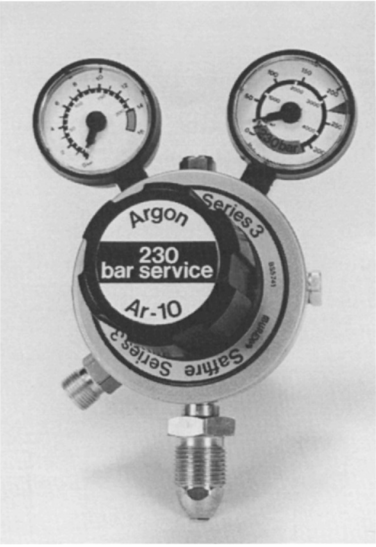
Argon
Series 3
230
bar service
Ar-10
Saffire Series 3

Fig. 3.16. (*b*) Two-stage argon regulator.

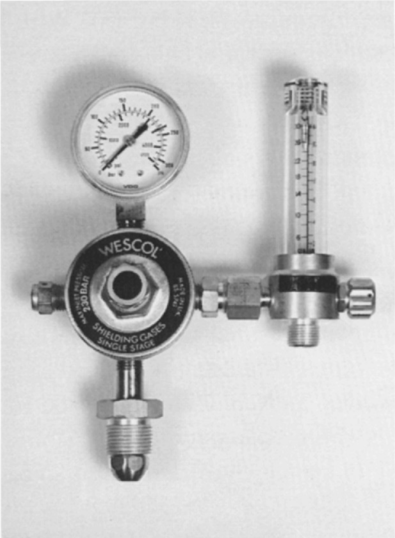
WESCOL
230 BAR
SHIELDING GASES
SINGLE STAGE

Additional equipment

Add-on equipment can be used with existing transformers and rectifiers. The equipment shown in Fig. 3.17 is fitted with a pulse generator and can be connected to a thyristor controlled rectifier power unit or to a.c. or d.c. arc welding sources. When connected to a d.c. source the pulse generator is used for arc ignition only, being automatically cut off after ignition, but it cuts in again if required to establish the arc. When used with an a.c. source the pulse generator ensures arc ignition without touchdown and ensures the re-ignition of the arc when welding. Soft start, crater fill, current control with one knob and remote control facilities are also fitted.

Torch

There is a variety of torches available varying from lightweight air cooled to heavy duty water cooled types (Fig. 3.18*a* and *b*). The main factors to be considered in choosing a torch are:

(1) Current-carrying capacity for the work in hand.
(2) Weight, balance and accessibility of the torch head to the work in hand.

The torch body holds a top-loading compression-type collet assembly which accommodates electrodes of various diameters. They are securely gripped yet the collet is easily slackened for removal or reposition of the electrode. As the thickness of plate to be welded increases, size of torch and

Fig. 3.16. (*c*) Argon flowmeter.

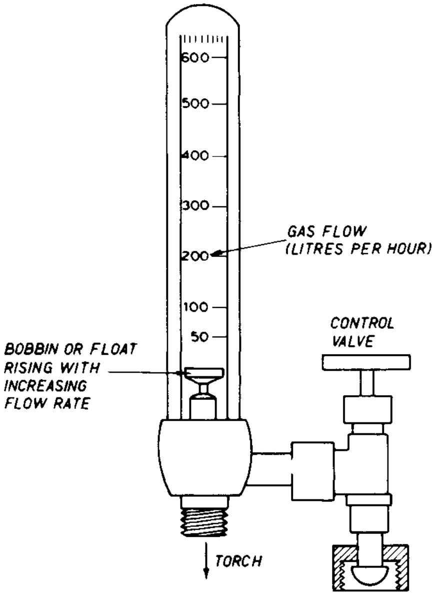
600
500
400
300
200
100
50
GAS FLOW (LITRES PER HOUR)
CONTROL VALVE
BOBBIN OR FLOAT RISING WITH INCREASING FLOW RATE
TORCH

electrode diameter must increase to deal with the larger welding currents required.

Small lightweight air cooled torches rated at 75 A d.c. and 55 A a.c. are ideal for small fittings and welds in awkward places and may be of pencil or swivel head type. Collet sizes on these are generally 0.8 mm, 1.2 mm and 1.6 mm diameter. Larger air cooled torches of 75 A d.c. or a.c. continuous rating or 100 A intermittent usually have a collet of 1.6 mm diameter. Air or water cooled torches rated at 300 A intermittent may be used with electrodes from 1.6 to 6.35 mm diameter and can be fitted with water cooled shields while heavy duty water cooled torches with a water cooled nozzle of 500 A a.c. or d.c. continuous rating and 600 A intermittent employ larger electrodes. A gas lens can be fitted to the torch to give better gas coverage and to obtain greater accessibility or visibility.

Normally, because of turbulence in the flow of gas from the nozzle, the electrode is adjusted to project up to a maximum of 4–9 mm beyond the nozzle (Fig. 3.20). By the use of a lens which contains wire gauzes of coarse and fine mesh, turbulence is prevented and a smooth even gas stream is obtained, enveloping the electrode which, if the gas flow is suitably increased, can be used on a flat surface projecting up to 19 mm from the nozzle orifice, greatly improving accessibility. The lens is screwed on to the torch body in place of the standard nozzle and as the projection of electrode from nozzle is increased the torch must be held more vertically to the work to obtain good gas coverage.

The ceramic nozzles (of alumina, silicon carbide or silicon nitride),

Fig. 3.17. Add-on unit for TIG welding.

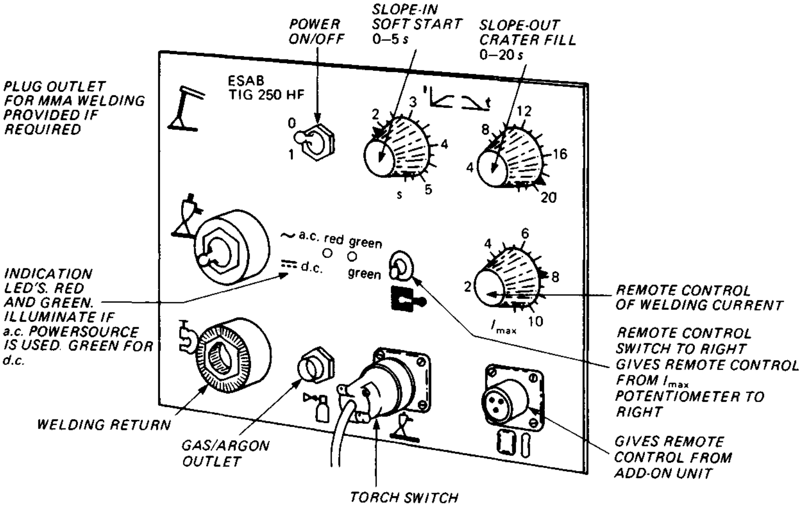
POWER ON/OFF
SLOPE-IN SOFT START 0–5 s
SLOPE-OUT CRATER FILL 0–20 s
PLUG OUTLET FOR MMA WELDING PROVIDED IF REQUIRED
ESAB TIG 250 HF
0
1
2
3
4
5
s
8
12
4
16
20
a.c. red green
d.c. green
4
6
2
8
10
I_{max}
INDICATION LED'S. RED AND GREEN. ILLUMINATE IF a.c. POWERSOURCE IS USED. GREEN FOR d.c.
REMOTE CONTROL OF WELDING CURRENT
REMOTE CONTROL SWITCH TO RIGHT GIVES REMOTE CONTROL FROM I_{max} POTENTIOMETER TO RIGHT
GIVES REMOTE CONTROL FROM ADD-ON UNIT
WELDING RETURN
GAS/ARGON OUTLET
TORCH SWITCH

Fig. 3.18. (*a*) 160 A air-cooled torch (manual). Torches are air cooled for lower currents and water cooled for higher currents.

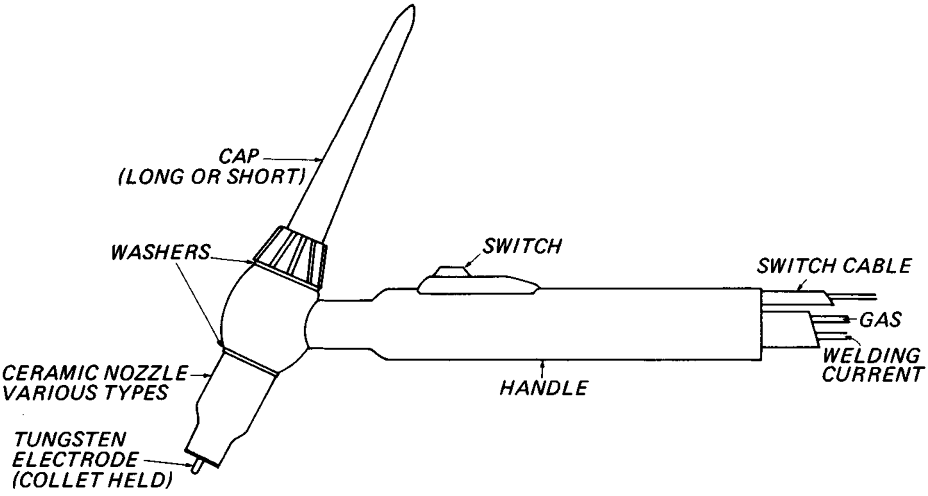
CAP (LONG OR SHORT)
WASHERS
SWITCH
SWITCH CABLE
GAS
WELDING CURRENT
CERAMIC NOZZLE VARIOUS TYPES
HANDLE
TUNGSTEN ELECTRODE (COLLET HELD)

Fig. 3.18. (*b*) A water-cooled torch.

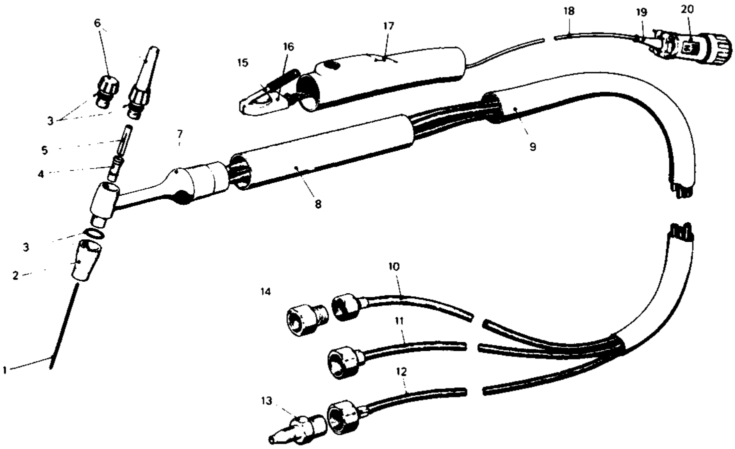
6
3
5
4
3
2
1
7
15
16
17
18
19
20
9
8
10
14
11
12
13

Key:

1. Thoriated or zirconiated tungsten electrode (0.8, 1.2, 1.6, 2.4, 3.2 mm diameter).
2. Ceramic nozzle.
3. O ring.
4. Collet holder.
5. Collet (sizes as above to take various diameters of electrodes).
6. Electrode cap (long and short).
7. Body assembly.
8. Sheath.
10. Argon hose assembly.
11. Water hose assembly.
12. Power cable assembly.
13. Adapter power/water; required only in certain cases.
14. Adapter argon; required only in certain cases.
15. Switch actuator.
16. Switch.
17. Switch-retaining sheath.
18. Cable, 2 core.
19. Insulating sleeve.
20. Plug.

which direct the flow of gas, screw on to the torch head and are easily removable for cleaning and replacement. Nozzle orifices range from 9.5 to 15.9 mm in diameter and they are available in a variety of patterns for various applications. Ceramic nozzles are generally used up to 200 A a.c. or d.c. but above this water cooled nozzles or shields are recommended because they avoid constant replacement.

Electrodes (see also BS 3019, Pts 1 and 2)

The electrode may be of pure tungsten but is more generally of tungsten alloyed with thorium oxide (thoria ThO_2) or zirconium oxide (zirconia ZrO_2).

The thoriated electrode is generally used with d.c.; 1% thoriated is often used but 2% thoriated gives good arc striking characteristics at low d.c. values. Although thoriated electrodes may be used for a.c. welding, it is generally better to use zirconiated.

Tungsten has a melting point of 3380 °C and a boiling point of 5950 °C so that there is only little vaporization in the arc and it retains its hardness when red hot. Though the electrodes are costly they are classed as virtually 'non-consumable'. They are supplied with a ground grey finish to ensure good collet contact, electrode diameters being 0.5, 1.2, 2.4, 3.2, 4.0, 4.8, 5.6 and 6.4 mm for zirconiated electrodes with additional sizes for 1% thoriated electrodes (e.g. 0.8, 8.0, 9.5 mm).

Though pure tungsten electrodes may be used, thoriated tungsten electrodes give easier starting on d.c. with a more stable arc and little possibility of tungsten contamination in the weld and they have a greater current carrying capacity for a given diameter than pure tungsten. However, when they are used on a.c. difficulty is encountered in maintaining a hemispherical end on the electrode. Thus zirconiated electrodes are preferred for a.c. welding because of good arc starting characteristics and the reduced risk of tungsten contamination. Zirconiated tungsten electrodes are used for high quality welds in aluminium and magnesium.

Selection of electrode size is usually made by choosing one near the maximum range for electrode and work. Too small an electrode will result in overheating and thus contamination of the work with tungsten, while too large an electrode results in arc control difficulty. Aim for a shining hemispherical end on the electrode, the lengths being usually 75 or 150 mm.

In the table of electrode current ratings it should be noted that the figures given are for class 1 welding using balanced a.c. waveform. Higher currents than these may be used by experienced welders.

Electrode grinding

Usually electrodes need grinding to a point only when thin materials are to be welded. They should be ground on a fine grit, hard abrasive wheel used only for this purpose to avoid contamination, ground with the electrode in the plane of the wheel and rotated while grinding (Fig. 3.19).

Electrode current ratings (BS 3019, Pts 2 and 3)

d.c.; thoriated electrodes								
Electrode diameter (mm)	0.5	1.2	1.6	2.4	3.2	4.0	4.8	
Max current (A)	20	60	70	120	200	300	370	
Balanced wave a.c.[a]								
Electrode diameter (mm)	0.5	1.2	1.6	2.4	3.2	4.0	4.8	6.4
Maximum current (A)	15	50	60	80	120	160	250	320

[a] Square wave has slightly higher values. Thoriated may be used but zirconiated is preferable.

Note. These are BS values. Experienced welders may use higher currents. Ratings on a.c. are roughly 40% lower than using d.c. electrode −ve (DCSP) and 60% duty cycle.

Pulsed current TIG welding (see also Chapter 2, pulsed MIG welding)

In MMA welding the operator chooses the current for a given joint and then, when making the weld, uses his skill to balance the heat input by the arc with the heat output due to the melting of the metal and conduction, convection and radiation by the work. He keeps this balance by electrode angle, amount of weave and rate of travel and achieves correct penetration.

In manual TIG welding this balance is achieved by weaving and the addition of filler metal. With pulsed current TIG welding the melting of the metal is controlled by pulsing the current at a higher value and at regular intervals.

There are two current values, the background current which is steady throughout and the pulsed current which is intermittent but at regular intervals. This pulse increases the energy in the arc and melts off an amount of filler rod to allow correct penetration, and then when the pulse ceases there is left only the background current to keep the arc ionized and the weld metal freezes.

The weld thus progresses as a series of overlapping spot welds and by judicious selection of background current and pulse height (amount of current) and duration the operator has complete control of welding parameters. It can be seen that this method is very advantageous for thin and medium sections, positional welding and in mechanized TIG welding. Pulse facilities are now incorporated on most power units.

Firstly the background current is set at about 5–12 A, then the pulse is set

Fig. 3.19. Tungsten electrode preparation.

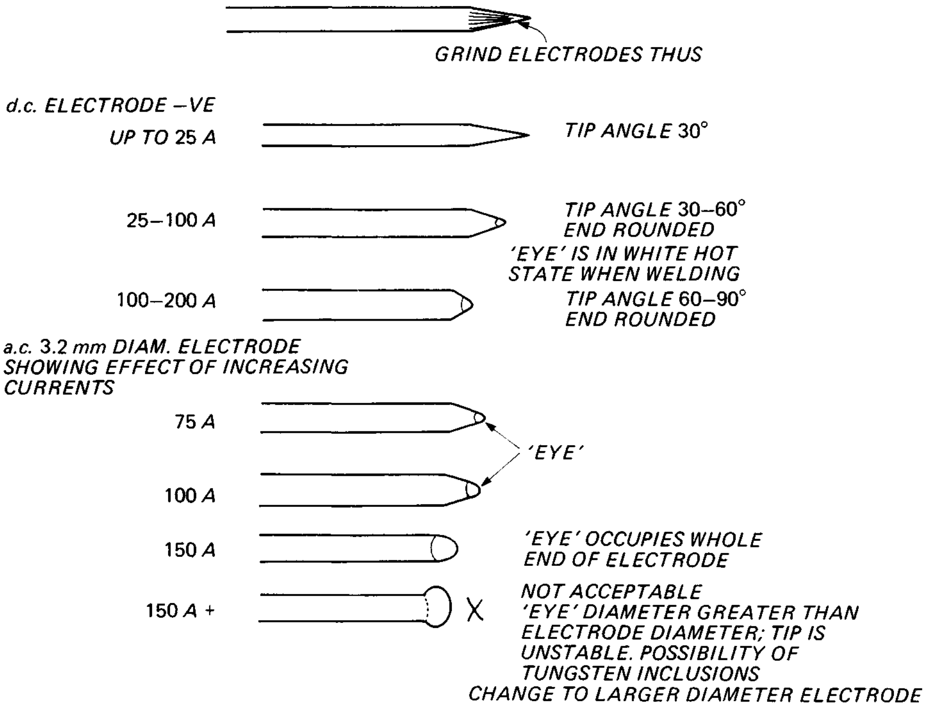
GRIND ELECTRODES THUS
d.c. ELECTRODE –VE
UP TO 25 A
TIP ANGLE 30°
25–100 A
TIP ANGLE 30–60°
END ROUNDED
'EYE' IS IN WHITE HOT STATE WHEN WELDING
100–200 A
TIP ANGLE 60–90°
END ROUNDED
a.c. 3.2 mm DIAM. ELECTRODE SHOWING EFFECT OF INCREASING CURRENTS
75 A
'EYE'
100 A
150 A
'EYE' OCCUPIES WHOLE END OF ELECTRODE
150 A +
X
NOT ACCEPTABLE
'EYE' DIAMETER GREATER THAN ELECTRODE DIAMETER; TIP IS UNSTABLE. POSSIBILITY OF TUNGSTEN INCLUSIONS
CHANGE TO LARGER DIAMETER ELECTRODE

Fig. 3.20. Electrode protrusion. No gas lens. In awkward places protrusion can be increased. Gas shield *must* cover area being welded.

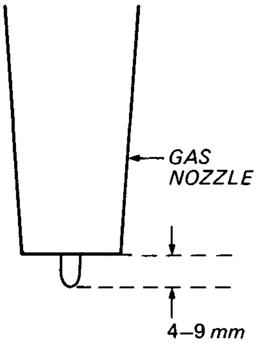
GAS NOZZLE
4–9 mm

at a value which depends upon the metal to be welded, say 140 A, and the duration is varied until the correct amount of penetration is achieved (remember that the heat input depends upon (pulse current × duration). The pulse heats up the molten pool, more filler wire is melted off, penetration is achieved and the pulse ceases; the metal then solidifies, cooling slightly ready for the next pulse.

Fig. 3.21 shows a basic pulse for d.c. and a.c. and since there is such control over the molten pool, fit-up tolerances are not so severe as in welding with non-pulsing current. Pulsed TIG is used on mild and low-alloy steels, stainless steels, heat-resistant steels, titanium, inconel, monel, nickel and aluminium and its alloys and for the orbital welding of thin pipes.

The pulse frequency can be increased from about 1 to 10 pulses per second by the use of the pulse control, while the width control varies pulse width and the background current control varies this as a constant percentage of the weld current as shown in Fig. 3.22. Using these controls

Fig. 3.21. Pulsed current. Pulses are variable in frequency (pulses per second) and width (peak current).

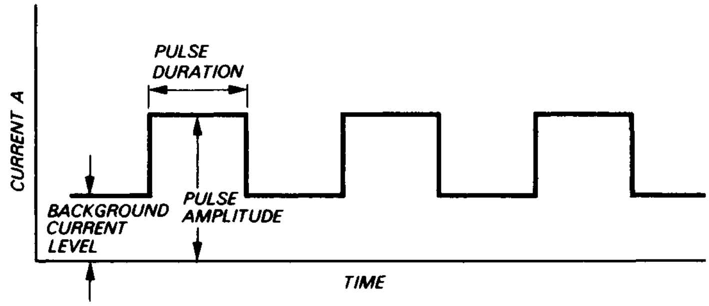

PULSED CURRENT d.c.

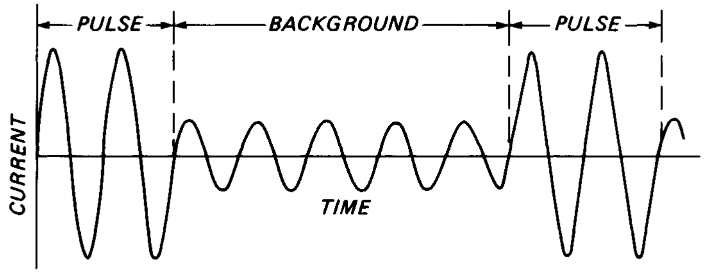

PULSED CURRENT a.c.

enables the operator to control the heat input into the weld and thus the root penetration.

Safety precautions

The precautions to be taken are similar to those when manual metal arc welding. In addition however there is the high voltage unit used for arc initiation without touchdown. This unit is similar in output to that of the car ignition coil and care should be taken when it is switched on. Do not let the electrode touch the body when switched on or a severe shock will be felt. BS 639 recommended EW filters are graded according to welding current thus: up to 50 A, no. 8; 15–75 A, no. 9; 75–100 A, no. 10; 100–200 A, no 11; 200–250 A, no, 12; 250–300 A, no. 13 or 14.

It will be noted that these filters are darker than those used for similar current ranges in MMA welding. The TIG (and MIG) arcs are richer in infra-red and ultra-violet radiation, the former requiring some provision for absorbing the extra associated heat, and the latter requiring the use of darker lenses.

It is recommended that the student reads the booklet *Electric arc welding, safety, health and welfare*, new series no. 38; published for the Department

Fig. 3.22

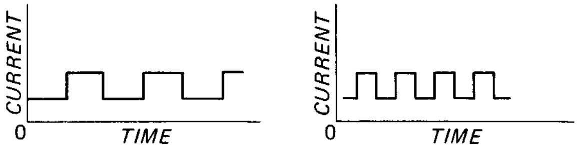

PULSE FREQUENCY INCREASE

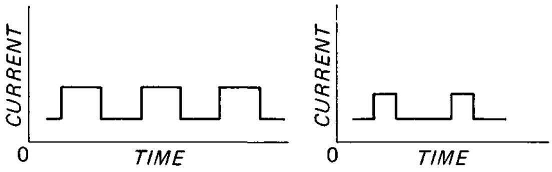

PULSE WIDTH DECREASE

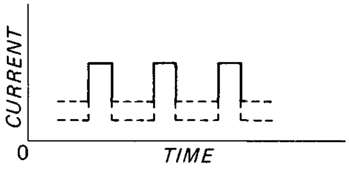

BACKGROUND CURRENT

of Employment by HMSO; and also BS 679, *Filters for use during welding*; British Standards Institution.

Welding techniques

Carbon steel

TIG welding can be used with excellent results on carbon steel sheet and pipework, and for this and other steel welds the angle of torch and rod are similar to that for aluminium, namely torch 80–90° to line of travel and filler wire at 20–30° to the plate (Fig. 3.23). The filler wire is chosen to match the analysis of the plate or pipe.

The welding of aluminium and its alloys

The table (p. 175) gives details of the filler wire types for welding the various alloys. Angles of torch and filler wire are shown in Fig. 3.23.

Casting alloys. Those with high silicon content, LM2, LM6, LM8 and LM9, can be welded to the Al–Mg–Si alloys with Al–Si 4043 wire. Similarly the Al–Mg casting alloys LM5 and LM10 can be welded to the wrought alloys using 5356 wire. Serious instability of LM10 may occur due to local heating and the alloy must be solution treated after welding.

Fig. 3.23. Angles of torch and rod showing typical weld: aluminium alloy plate 11 mm thick 70–80° prep., 1.5 mm root face, no gap, a.c. volts 102, arc volts drop, 17–18 volts: electrode 4.8 mm diameter, current 255 A, arc length approx. 4 mm.

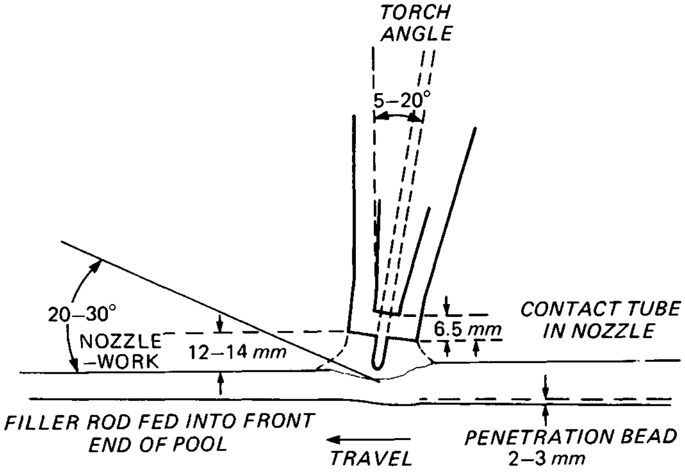
TORCH ANGLE
5–20°
20–30°
NOZZLE–WORK
12–14 mm
6.5 mm
CONTACT TUBE IN NOZZLE
FILLER ROD FED INTO FRONT END OF POOL
TRAVEL
PENETRATION BEAD 2–3 mm

Technique. The supply is a.c. and the shielding gas pure argon. Remember that, with TIG and MIG processes using shielding gases, the welding position must be protected from stray draughts and winds which will destroy the argon shield. This is particularly important when site welding and (waterproof) covers or a tent to protect welder and gas shields from wind and rain must be erected if they are to be fully protected. Pre-heating is required for drying only to produce welds of the highest quality. All surfaces and welding wire should be degreased and the area near the joint and the welding wire should be stainless steel wire brushed or scraped to remove oxide and each run brushed before the next is laid. After switching on the gas, water, welding current and HF unit, the arc is struck by bringing the tungsten electrode near the work (without touching down). The HF sparks jump the gap and the welding current flows.

Arc length should be about 3 mm. Practise starting by laying the holder on its side and bringing it to the vertical position using the gas nozzle as a fulcrum, or use a striking plate and get the tungsten hot before starting the weld. (This does not apply to scratch start.) The arc is held in one position on the plate until a molten pool is obtained and welding is commenced, proceeding from right to left, the rod being fed into the forward edge of the molten pool and always kept within the gas shield. It must not be allowed to touch the electrode or contamination occurs. A black appearance on the weld metal indicates insufficient argon supply. The flow rate should be checked and the line inspected for leaks. A brown film on the weld metal indicates presence of oxygen in the argon while a chalky white appearance of the weld metal accompanied by difficulty in controlling the weld indicates excessive current and overheating. The weld continues with the edge of the portion sinking through, clearly visible, and the amount of sinking, which determines the size of the penetration bead, is controlled by the welding rate.

Run-on and run-off plates are often used to prevent cold starts and craters at the end of a run, which may lead to cracking. Modern sets have slope-up and slope-down controls which greatly assist the welding operation but it is as well to practise using these plates.

Preparations for single V butt joints are shown in Fig. 3.24*a* while 3.24*b* shows a backing strip in position.

Tack welding and the use of jigs for line-up is similar to that used for steel. The tacks should be reduced in size by chiseling or grinding before welding the seam and care should be taken, when breaking any jigs away that were temporarily welded in, that damage does not occur to the structure or casting.

Aluminium and its alloys. Filler rod guide – TIG and MIG

Parent metal	6061 6063 6082	5083	5454	5154A 5251	3103	1050A
1050A	4043 5356(3)	5356 4043	4043 5183 5356	4043 5183 5356	4043 5356 1050A(1)	4043 1050A(1) (2)
3103	4043 5356(3)	5356 5183	5154A 5183 5356	5154A 5183 5356	4043 1050A 5356	
5154A 5251	5356 5183 4043	5356 5183 5556A	5356 5183 5154A	5356 5183		
5454	5356 5183 4043	5356 5183 5556A	5154A 5554(2) 5356			
5083	5356 5183 5556A	5183 5356 55556A				
6061 6063 6083	4043 5356(3) 5183					

(1) corrosive conditions
(2) elevated temperatures
(3) colour matching (after anodizing)

Old designations

1050A	1B
5154A	NG5
5356	NG6
5183	NG8
4043	NG21
5554	NG 52
5556	NG 61

Examples of preparation: Flat butt joints, no backing

Plate thickness (mm)	No. of runs	Current (A)	Filler rod diam. (mm)
2.0	1	100–110	2.4
2.4	1	120	3.2
3.2	1	130–160	3.2
4.8	2	230	4.8
6.4	2	240	4.8

Plate thickness (mm)		
2	Butt no gap	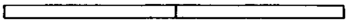
2.4	Butt 0.8 mm gap	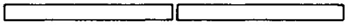
3.2	Butt 0.8 mm gap	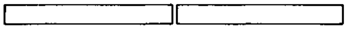

TIG tack welding is used for line-up of plate and tube before finally MIG welding the seam. Temporary backing bars are grooved to the shape of the underbead and made of mild or stainless steel. These help to form and shape the penetration bead when root runs cannot be made and control of penetration is difficult. When ungrooved bars are used they should be removed after welding, the root run back-chipped and a sealing run made. Any backing strips to be welded in place should be of aluminium alloy, tack welded in place and fused in with the root run.

The vertical double-operator method (Fig. 3.25), using filler wire on one side only, gives sound non-porous welds with accurate penetration bead controlled by one operator 'pulling through' on the reverse side. There is excellent argon shielding from both sides but the method is expensive in man-hours of work and needs more skill. It however reduces residual stresses as there is less total heat input and welding speed is increased.

Magnesium and magnesium alloys

Equipment is similar to that for the aluminium alloys and the technique similar, welding from right to left with a short arc and with the same angles of torch and filler rod, with pure argon as the shielding gas and an a.c. supply. Little movement of the filler rod is required but with material over 3 mm thick some weaving may be used. For fillet welding the torch is held so as to bisect the angle between the plates and at sufficient angle from the vertical to obtain a clear view of the molten pool, with

Fig. 3.24. (*a*) Manual TIG process: preparation of aluminium and aluminium alloy plate. Single V butt joints. Flat. Plate thickness 4.8 mm–9.5 mm, root run may be back chipped and a sealing run made, though this adds to the cost of the weld.

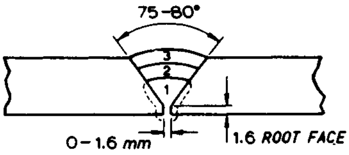

Fig. 3.24. (*b*) Use of backing strip and preparation.

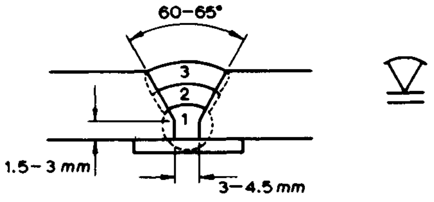

enough weave to obtain equal leg fusion. Tilted fillets are used to obtain equal length.

The material is supplied greased or chromated. Degreasing removes the grease and it is usual to remove the chromate by wire brushing from the side to be welded for about 12 mm on each side of the weld and to leave the chromate on the underside where it helps to support the penetration bead. No back purge of argon is required. The surface and edges should be wire brushed and the filler rod cleaned before use. Each run should be brushed before the next is laid.

Backing plates can be used to profile the underbead and can be of mild steel, or of aluminium or copper 6.4–9.5 mm wide, with grooves 1.6–3.2 mm deep for material 1.6–6.4 mm thick. (Fig. 3.26.)

Jigs, together with correct welding sequence, can be used to prevent distortion, but if it occurs the parts can be raised to stress relief temperatures of approximately 250°–300°C. Ensuring sufficient flow of argon will prevent any oxidized areas, porosity, and entrapped oxides and nitrides.

Welding rods are of similar composition to that of the plate and the table indicates the relative weldability of similar and dissimilar alloys. It is not recommended that the Mg–Zr alloys should be welded to alloys containing Al or Mn.

Fig. 3.25. (*a*) Angles of torches and filler wire. (*b*) Double operator vertical welding aluminium alloys.

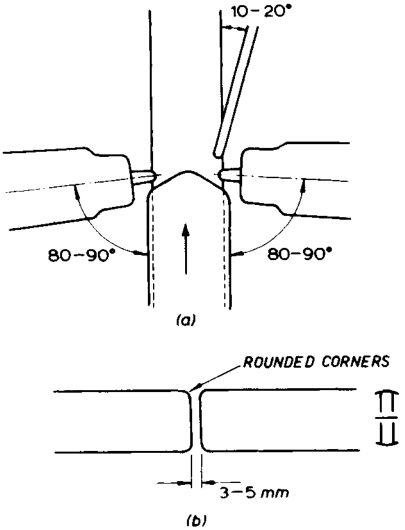

Typical magnesium alloy welding rod, composition and suitability

Elektron alloy	Zn%	Zr%	RE%	Others%	MEL welding rod type
RZE 1	2.0–3.0	0.5–1.0	2.5–4.0		W 6
RZ5	3.5–5.0	0.4–1.0	1.2–1.75		W 7
ZT1	1.7–2.5	0.5–1.0	0.1	2.5–4.0 Th	W 1
TZ6	5.2–6.0	0.5–1.0	0.2	1.5–2.2 Th	W 12
MSR B	0.2	0.4–1.0	2.2–2.7	2.0–3.0 Ag	W 13
AM503	0.03	0.01	—	1.3–1.7 Mn, 0.02 Ca	W 2
AZ31	0.7–1.3	—	—	2.5–3.5 Al, 0.04 Ca	W 15
A8	0.4–1.0	—	—	0.1 Sn, Cu, Si, Fe, Ni, to 0.4 max, 7.5–8.1 Al	W 14

Note. RE – Rare earths.

Fig. 3.26. Suitable edge preparations for magnesium alloy butt welds (dimensions in mm).

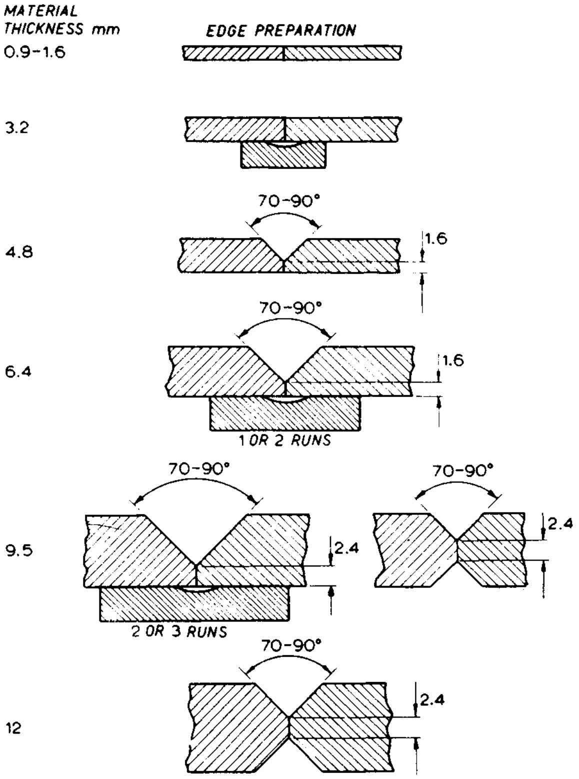
MATERIAL THICKNESS mm
EDGE PREPARATION
0.9–1.6
3.2
70–90°
4.8
1.6
70–90°
6.4
1.6
1 OR 2 RUNS
70–90°
70–90°
9.5
2.4
2.4
2 OR 3 RUNS
70–90°
12
2.4

For fillet welds the torch is held at about 90° to the line of travel and roughly bisecting the angle between the plates, so that the nozzle is clear of either plate, and it is often necessary to have more projection of the wire beyond the nozzle to give good visibility, but it should be kept to a minimum. Tilted fillets give a better weld profile and equal leg length. The table gives recommended filler wires for the various alloys.

Stainless and heat-resistant steel

The production welding of stainless steel by this process is apt to be rather slow compared with MMA, MIG or pulsed arc. (See the section on the MMA welding of stainless steel.)

Areas adjacent to the weld should be thoroughly cleaned with stainless steel wire brushes and a d.c. supply with the torch negative used. The shielding gas is pure argon or argon–hydrogen mixtures (up to 5%) which give a more fluid weld pool, faster rate of deposition (due to the higher temperature of the arc), better 'wetting' and reduction of slag skin by the hydrogen.

The addition of oxygen, CO_2 or nitrogen is not recommended. Joint preparation is given in Fig. 3.27.

Fig. 3.27. Preparation of stainless steel and corrosion- and heat-resistant steels. (TIG and MIG.)

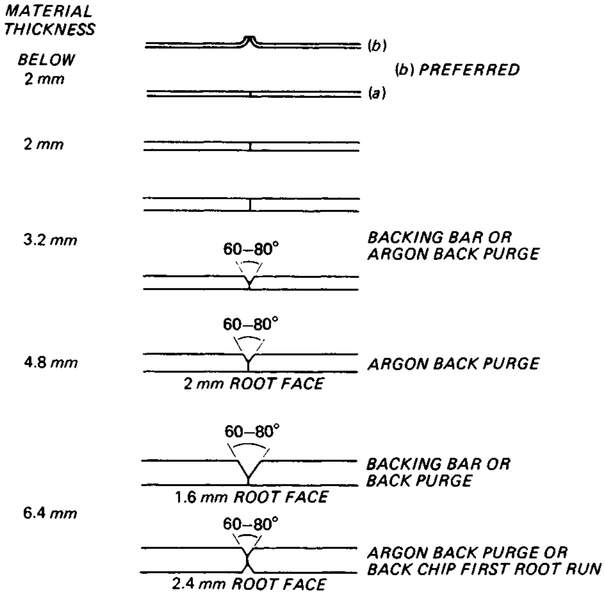
MATERIAL THICKNESS
BELOW 2 mm
(b)
(b) PREFERRED
(a)
2 mm
3.2 mm
60–80°
BACKING BAR OR ARGON BACK PURGE
60–80°
4.8 mm
ARGON BACK PURGE
2 mm ROOT FACE
60–80°
BACKING BAR OR BACK PURGE
1.6 mm ROOT FACE
6.4 mm
60–80°
ARGON BACK PURGE OR BACK CHIP FIRST ROOT RUN
2.4 mm ROOT FACE

Technique. The torch is held almost vertically to the line of travel and the filler rod fed into the leading edge of the molten pool, the hot end never being removed from the argon shield. Electrode extension beyond the nozzle should be as short as possible: 3.0–5.0 mm for butt and 6.0–11.0 for fillets. To prevent excessive dilution, which occurs with a wide weave, multi-run fillet welds can be made with a series of 'stringer beads'. The arc length should be about 2 mm when no filler wire is used and 3–4 mm with filler. Gas flow should be generous and a gas lens can be used to advantage to give a non-turbulent flow. Any draughts should be avoided.

Nickel alloys

Although this process is very suitable for welding nickel and its alloys it is generally considered rather slow, so that it is mostly done on thinner sections of sheet and tube.

Shielding gas should be commercially pure argon and addition of hydrogen up to 10% helps to reduce porosity and increases welding speeds especially for MONEL® alloy and Nickel alloy 200. Helium has the same advantages, greatly increasing welding speed. Great care should be taken to avoid disturbance of the protective inert gas shield by draughts, and the largest nozzle diameter possible should be used, with minimum distance between nozzle and work. The use of a gas lens increases the efficiency of the shield and gives increased gas flow without turbulence. Argon flow is of the order of 17–35 litres per hour for manual operation and higher for automatic welding. Helium flow should be about 1½–3 times this flow rate.

The torch is held as near 90° to the work as possible – the more acute the angle the greater the danger of an aspirating effect causing contamination of the gas shield. Electrodes of pure, thoriated or zirconiated tungsten can be used, ground to a point to give good arc control and should project 3–4.5 mm beyond the nozzle for butt welds and 6–12 mm for fillets.

Stray arcing should be avoided as it contaminates the parent plate, and there should be no weaving or puddling of the pool, the hot filler rod end being kept always within the gas shield, and fed into the pool by touching it on the outer rim ahead of the weld. The arc is kept as short as possible – up to 0.5 mm maximum when no filler is used.

To avoid crater shrinkage at the end of a weld a crater-filling unit can be used or failing this, extension tabs for run-off are left on the work and removed on completion. Fig. 3.28 gives suitable joint preparations.

For vertical and overhead joints the technique is similar to that for steel, but downhand welding gives the best quality welds.

Filler wires for TIG, PLASMA and MIG dip, spray and pulsed submerged arc filler metals and fluxes

Material	Filler wire TIG/plasma MIG dip, pulsed	Filler wire MIG spray	Filler wire/flux submerged arc
Nickel alloys 200, 201	Nickel alloy 61	Nickel alloy 61	Nickel alloy 61, INCOFLUX® 6
MONEL® alloy 400	MONEL® alloy 60	MONEL® alloy 60	MONEL® alloy 60, INCOFLUX® 5
MONEL® alloy K500	MONEL® alloy 60	Not recommended	Not recommended
Cupro–nickel alloys	MONEL® alloy 67	MONEL® alloy 67	MONEL® alloy 67, INCOFLUX® 8
INCONEL® alloy 600	INCONEL® alloy 82	INCONEL® alloy 82	INCONEL® alloy 82, INCOFLUX® 4
INCONEL® alloy 601	INCONEL® alloys 601, 82, 625	INCONEL® alloys 82, 625	INCONEL® alloys 82, 625, INCOFLUX® 4, 7
INCONEL® alloy 625	INCONEL® alloy 625	INCONEL® alloy 625	INCONEL® alloy 625, INCOFLUX® 7
INCONEL® alloy 617	INCONEL® alloy 617	INCONEL® alloy 617	Not recommended
INCONEL® alloy 690	INCONEL® alloys 82, 625	INCONEL® alloys 82, 625	Not recommended
INCONEL® alloy 718	INCONEL® alloys 625, 718	Not recommended	Not recommended
INCONEL® alloy X-750	INCONEL® alloy 718	Not recommended	Not recommended
INCO® alloy C-276	INCO® alloy C-276	INCO® alloy 276	Not recommended
INCO® alloy HX	INCO® alloy HX, INCONEL® alloy 617	INCO® alloy HX, INCONEL® alloy 617	Not recommended
INCO® alloy G-3	INCONEL® alloy 625	INCONEL® alloy 625	Not recommended
INCOLOY® alloys 800, 800H, 800HT	INCONEL® alloys 82, 625, 617	INCONEL® alloys 82, 625, 617	INCONEL® alloys 82, 625, INCOFLUX® 4, 7
INCOLOY® alloy 825	INCOLOY® alloy 65	INCOLOY® alloy 625	INCONEL® alloy 625, INCOFLUX® 7
INCOLOY® alloy DS	NC 80/20	Not recommended	Not recommended
NIMONIC® alloy 75	NC 80/20	NC 80/20	INCONEL® alloy 82, INCOFLUX® 4
BRIGHTRAY® alloy series	NC 80/20	Not recommended	Not recommended
NILO® alloy series	Nickel alloy 61, INCONEL® alloy 82	Not recommended	Not recommended
Cast irons	NI-ROD® alloys 44, 99, NILO® alloy 55	NI-ROD® alloys 44, FC55, NILO® alloy 55	NI-ROD® alloy FC55, INCOFLUX® 5
Steels containing molybdenum	INCONEL® alloy 625	INCONEL® alloy 625	INCONEL® alloy 625, INCOFLUX® 7

Copper and copper alloys

Direct current, electrode negative is used with argon as the shielding gas for copper and most of its alloys with currents up to 400 A. and angle of torch and rod roughly as for aluminium welding. With aluminium bronze and copper–chromium alloys however, dispersal of the oxide film is difficult when using d.c., and for these alloys a.c. is generally used, but d.c. can be used with helium as the shielding gas. The weld areas should be bronze wire brushed and degreased and each run brushed to remove oxide film. Jigs or tack welds can be used for accurate positioning and to prevent distortion.

Mild steel or stainless steel backing bars, coated with graphite or anti-spatter compound to prevent fusion, can be used to control the penetration bead and also to prevent heat dissipation, or backing may be welded into the joint.

Because of the high coefficient of thermal expansion of copper the root gap has a tendency to close up as welding proceeds, and due to its high thermal conductivity pre-heating from 400–700 °C according to thickness is essential on all but the thinnest sections to obtain a good molten pool. The thermal conductivity of most of the copper alloys is much lower than copper so that pre-heating to about 150 °C is usually sufficient. In the range 400–700 °C a reduction in ductility occurs in most of the alloys so that a cooling period should be allowed between runs. The main grades of copper are (1) tough pitch (oxygen containing) high conductivity, (2) oxygen free, high conductivity, (3) phosphorus deoxidized, and in most cases it is the last type that is used for pressure vessels, heat exchangers and food processing equipment, etc. If tough pitch copper, in the cast form, is to be welded, a deoxidizing filler rod should be used to give a deoxidized weld and in all cases the weld should be performed as quickly as possible to prevent overheating.

Argon, helium and nitrogen can be used as shielding gases. Argon has a lower heat output than helium and nitrogen because it has a lower arc voltage. Argon can be mixed with helium and the mixture increases the heat output proportionally with the helium percentage, so that using the mixture for copper welding a lower pre-heat input is required and

Fig. 3.28

	Nickel alloys. Examples of joint design for fusion welding. Type of joint	Material thickness (mm) T	Width of groove (top) (mm) W	Root space (mm) S
Square butt	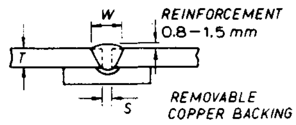 W REINFORCEMENT 0.8–1.5 mm T S REMOVABLE COPPER BACKING	1.0	3.2	0
		1.2	4.0	0
		1.6	4.8	0
		2.4	4.8–6.4	0–0.8
		3.2	6.4	0–0.8
V groove	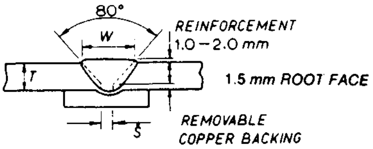 80° W REINFORCEMENT 1.0–2.0 mm T 1.5 mm ROOT FACE S REMOVABLE COPPER BACKING	4.8	8.9	3.2
		6.4	13.0	4.8
		8.0	15.5	4.8
		9.5	18.0	4.8
		12.7	23.0	4.8
		16.0	29.5	4.8
V groove	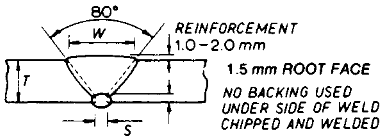 80° W REINFORCEMENT 1.0–2.0 mm T 1.5 mm ROOT FACE NO BACKING USED UNDER SIDE OF WELD CHIPPED AND WELDED S	6.4	10.4	2.4
		8.0	13.0	2.4
		9.5	16.5	3.2
		12.7	21.6	3.2
		16.0	27.0	3.2

penetration is increased. Nitrogen gives the greatest heat output but although the welds are sound they are of rough appearance.

Recommended filler alloys are:

C7 0.05–0.35% Mn, 0.2–0.35% Si, up to 1.0% Sn, remainder Cu.
C8 0.1–0.3% Al, 0.1–0.3% Ti, remainder Cu.
C21 0.02–0.1% B, 99.8% Cu.

Note. Filler alloys for copper and its alloys are to BS 2901, Pt 3, 1970.

Copper–silicon (silicon bronzes). d.c. supply, electrode negative and argon (or helium) shield. These alloys contain about 3% silicon and 1% manganese. They tend to be hot short in the temperature range 800–950 °C so that cooling should be rapid through this range. Too rapid cooling, however, may promote weld cracking.

Filler alloy C9, 96% Cu, 3% Si, 1% Mn is recommended.

Copper–aluminium (aluminium bronzes). a.c. with argon shield or d.c. with helium shield. The single phase alloys contain about 7% Al and the two phase alloys up to 11 % Al with additions of Fe, Mn and Ni. Welding may affect the ductility of both types. Weld root cracking in the 6–8% Al, 2–2.3% Fe alloy used for heat exchangers can be avoided by using a non-matching filler rod. Due to a reduction in ductility at welding temperatures difficulty may be experienced in welding these alloys. The alloy containing 9% Al and 12% Mn with additions of Fe and Ni has good weldability but requires heat treatment after welding to restore corrosion resistance and other mechanical properties.

Recommended filler alloys are:

C12 6.0–7.0% Al, remainder Cu.
C13 9.0–11.0% Al, remainder Cu.
C20 8.0–10.0% Al, 1.5–3.5% Fe, 4.0–7.0% Ni, remainder Cu.

Copper–nickel (cupro–nickels). d.c. with argon shield and electrode negative. These alloys, used for example for pipe work and heat exchangers, contain from 5 to 30% Ni, often with addition of Fe or Mn to increase corrosion resistance. They are very successfully welded by both TIG and MIG processes but since they are prone to contamination by oxygen and hydrogen, an adequate supply of shielding gas must be used including a back purge in certain cases, if possible.

Recommended filler alloys are:

C17 19–21% Ni, 0.2–0.5% Ti, 0.2–1.0% Mn, remainder Cu.
C19 29–31% Ni, 0.2–0.5% Ti, 0.2–1.0% Mn, remainder Cu.

Copper–tin (phosphor bronzes and gunmetal), d.c. with argon shield. The wrought phosphor bronzes contain up to 8% Sn with phosphorus additions up to 0.4%. Gunmetal is a tin bronze containing zinc and often some lead. Welding of these alloys is most often associated with repair work using a phosphor bronze filler wire but to remove the danger of porosity, non-matching filler wires containing deoxidizers should be used.

Filler alloys recommended are:

C10 4.5–6.0% Sn, 0.4% P, remainder Cu.
C11 6.0–7.5% Sn, 0.4% P, remainder Cu.

Copper–zinc (brass and nickel silvers), a.c. with argon shield, d.c. with helium shield. The copper–zinc alloys most frequently welded are Admiralty brass (70% Cu, 29% Zn, 1% Sn), Naval brass (62% Cu, 36.75% Zn, 1.25% Sn) and aluminium brass (76% Cu, 22% Zn, 2% Al). Porosity, due to the formation of zinc oxide, occurs when matching filler wire is used so that it is often preferable to use a silicon–bronze wire which reduces evolution of fumes. This may, however, induce cracking in the HAZ. Post-weld stress relief in the 250–300 °C range reduces the risk of stress corrosion cracking. The range of nickel silver alloys is often brazed.

Filler alloys recommended are:

C14 70–73% Cu, 1.0–1.5% Sn, 0.02–0.06% As.
C15 76–78% Cu, 1.8–2.3% Sn, 0.02–0.06% As.

Work-hardening and precipitation-hardening copper-rich alloys

The work-hardening alloys are not often welded because mechanical properties are lost in the welding operation. Heat-treatable alloys such as copper–chromium and copper–beryllium are welded with matching filler rods using an a.c. supply to disperse the surface oxides. They are usually welded in the solution-treated or in the over-aged condition and then finally heat-treated.

Hardfacing

The tungsten arc is a good method for deposition of stellite and other hard surfacing materials on a base metal. The heat is more concentrated than that of the oxy-acetylene flame and although there may be some base metal pick-up this can be kept to a minimum by using correct technique and there is not a great deal of dilution. It produces a clean, sound deposit and is very useful for reactive base metals.

The electrode is connected to the –ve pole (straight polarity) with the largest diameter filler rod possible to reduce tungsten contamination. Deposition is similar to that when the oxy-acetylene flame is used and deposition moves from right to left. Filler rods of 3.2, 4.0, 4.8, 6.4 and 8 mm are available and alloys include stellite and nickel-based alloys, and deposition proceeds in the normal way for hard surfacing – do not puddle the pool as this increases dilution.

Easily hardfaced

Low and medium carbon steel to 0.4% C. Above 0.4% C, oxy-acetylene only.

Low-alloy steels.

Nickel and nickel–copper alloys.

Chrome–nickel stainless steels, niobium stabilized. (Not free machining nor titanium stabilized).

11–14% manganese steels.

More difficult to hardface

Cast iron.

Stainless steel, straight and titanium stabilized.

Tool and die steels, water-hardening, oil-hardening and air-hardening types and hot work grade.

Straight chromium stainless steels.

Automatic welding

For automatic welding the TIG torch is usually water cooled and may be carried on a tractor moving along a track or mounted on a boom so as to move over the work or for the work to move under the head. The head has a control panel for spark starter, water and gas and current contactor, and the torch has lateral and vertical movement, the arc length being kept constant by a motor-driven movement controlled by circuits operating from the arc length. Filler wire is supplied from the reel to the weld pool by

TIG process. Recommended filler wires for copper welding

		Filler wire	
Type (BS 2870–2875)	Grade	Argon or helium shield	Nitrogen shield
C106	Phosphorus deoxidized non-arsenical	C7, C21	C8
C107	Phosphorus deoxidized arsenical		
C101	Electrolytic tough pitch high conductivity	C7, C21	Not recommended
C102	Fire-refined tough pitch high conductivity		
C103	Oxygen-free high conductivity	C7, C21	Not recommended

rollers driven by an adjustable speed motor. Heavier currents can be used than with manual operation resulting in greater deposition rates, and the accurate control of speed of travel and arc length results in welds of high quality. Plate up to 9 mm thick can be welded in one run, the welding being of course downhand. Fig. 3.29 illustrates a typical tractor-driven head.

Orbital pipe welding unit

This can operate from a standard TIG power source and enables circumferential joints on pipes and tubes to be made automatically. It is of caliper type enabling rapid adjustments to be made, and rotational drive is by a small motor mounted in the handle of the unit, rotational speed being controlled by transistor regulator. A wire feed unit feeds wire to the arc, the feed speed of the wire being controlled by a thyristor regulator. Three sizes are available for pipes with outside diameters of 18–40, 36–80 and 71–160 mm (Fig. 3.30).

Mechanized TIG hard facing

In this process the argon shielded TIG arc oscillates over a width from 6 to 25 mm and is operated by variable controls for speed and width. Rod and torch also have controls for vertical height and horizontal positioning and for feeding the hard surfacing rod.

Small quantities of hydrogen are added to the argon by means of a gas blender to give a more consistent deposit. The stellite alloy, in rod form, can be laid on large circular components of 100–600 mm diameter and this is ideal for valves and valve seatings, rings and sleeves, etc., the work being mounted on a manipulator.

A smaller system enables components from 25–220 mm diameter to be hard faced. The unit can be adapted for flat straight working if required and the usual power unit for TIG welding is required. A complete range of hard facing alloys is available.

Fig. 3.29. Automatic TIG welding head mounted on electrically driven tractor; 500 A water cooled torch.

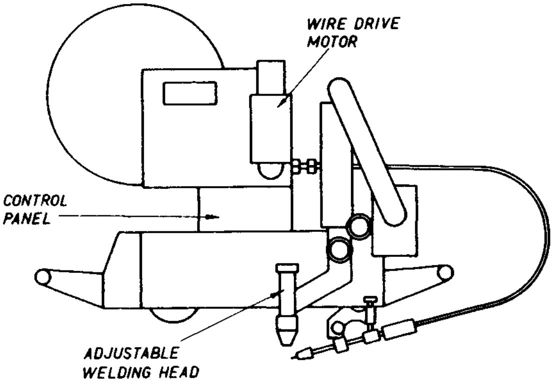

TIG robot welding

TIG welding has now been applied to the robot system with two manipulators. In this method of welding the workpieces are held in jigs on the plates of the manipulators, one being welded while the other is being loaded. The six axes of the robot enable it to handle the filler wire and place it in the best position without limitation of the wrist movement of the robot.

Since the HF necessary for the arc initiation may cause unwanted signals to the microprocesser all conductors have been screened. The control computer selects and records the particular welding program and together with the robot achieves synchronized movements, and welding operations are commenced when the operator has loaded one of the two manipulators of the system. (See Chapter 5, Welding with robots, for illustration of this system.)

Various parameter sequences for various operations can be pre-recorded and recalled as required.

Spot welding

Mild steel, stainless and alloy steel and the rarer metals such as titanium can be spot welded using a TIG spot welding gun. The pistol grip water cooled gun carries a collet assembly holding a thoriated tungsten electrode positioned so that the tip of the electrode is approximately 5 mm

Fig. 3.30. Orbital TIG pipe welding.

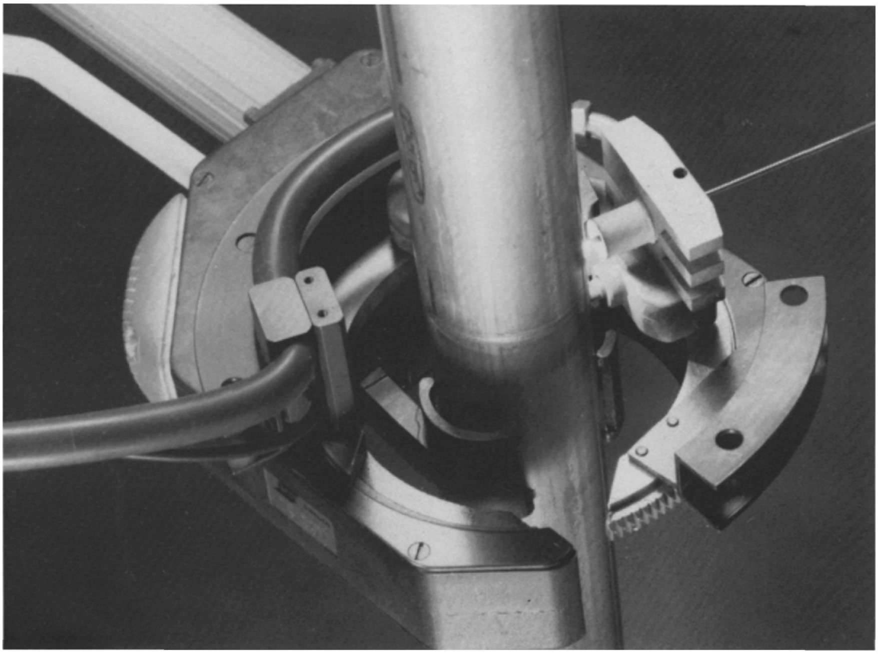

within the nozzle orifice, thus determining the arc length. Interchangeable nozzles enable various types of joints to be welded flat, or vertically. A variable timing control on the power unit allows the current to be timed to flow to ensure fusion of the joint and a switch on the gun controls gas, water, HF unit and current contactor. To make a weld, the nozzle of the gun is pressed on to the work in the required position and when the gun switch is pressed the following sequence occurs: (1) gas and water flow to the gun, (2) the starter ignites the arc and the current flows for a period controlled by the timer, (3) current is automatically switched off and post-weld argon flows to cool electrode tip and weld.

Note. Refer also to BS 3019, *General recommendations for inert gas, tungsten arc welding*, Part 1: *Wrought aluminium, aluminium alloys and magnesium alloys*, Part 2: *Austenitic stainless and heat resisting steels.*

Titanium

Titanium is a silvery coloured metal with atomic weight 47.9, specific gravity 4.5 (cf. Al 2.7 and Fe 7.8) melting point 1800°C and UTS 310 N/mm^2 in the pure state. It is used as deoxidizer for steel and sometimes in stainless steel to prevent weld decay and is being used in increasing amounts in the aircraft and atomic energy industries because of its high strength-to-weight ratio.

Titanium absorbs nitrogen and oxygen rapidly at temperatures above 1000°C and most commercial grades contain small amounts of these gases; the difficulty in reducing the amounts present makes titanium expensive. Because of this absorption it can be seen that special precautions have to be taken when welding this metal.

Where seams are reasonably straight a TIG torch with tungsten electrode gives reasonable shrouding effect on the upper side of the weld, and jigs may be employed to concentrate the argon shroud even more over the weld. In addition the underbead must be protected and this may be done by welding over a grooved plate, argon being fed into this groove so that a uniform distribution is obtained on the underside. It can be seen therefore that the success of welding in these conditions depends upon the successful shrouding of the seam to be welded.

For more complicated shapes a vacuum chamber may be employed. The chamber into which the part to be welded is placed is fitted with hand holes, and over these are bolted long-sleeve rubber gloves into which the operator places his hands, and operates the TIG torch and filler rod inside the chamber. An inspection window on the sloping top fitted with welding glass enables a clear view to be obtained. All cable entries are sealed tightly and the air is extracted from the chamber as completely as possible by pumps, and the chamber filled with argon, this being carried out again if necessary to get the level of oxygen and nitrogen down to the lowest permissible level. The welding may be performed with a continuous flow of argon through the chamber or in a static argon atmosphere. Since the whole success of the operation depends upon keeping oxygen and nitrogen levels down to an absolute minimum great care must be taken to avoid leaks as welding progresses and to watch for discoloration of the weld indicating absorption of the gases.

Tantalum

Tantalum is a metallic element of atomic weight 180.88, HV 45, melting point 2910 °C, specific gravity 16.6, it is a good conductor of heat and is used in sheet form for heat exchangers, condensers, small tubes, etc. Like titanium it readily absorbs nitrogen and oxygen at high temperatures and also readily combines with other metals so that it is necessary to weld it with a TIG torch in a vacuum-purged argon chamber as for titanium.

Beryllium

Beryllium is a light, steely coloured metallic element, atomic weight 9, melting point 1280–1300 °C, specific gravity 1.8, HV 55–60. It is used as an alloy with copper (beryllium bronze) to give high strength with elasticity. As with tantalum and titanium it can be welded in a vacuum-purged argon chamber using a TIG torch. It can also be pressure welded and resistance welded.

Electron beam welding is now being applied to the welding of the rare metals such as the foregoing, in the vacuum chamber. A beam of electrons is concentrated on the spot where the weld is required. The beam can be focused so as to give a small or large spot, and the power and thus the heating effect controlled. Successful welds in a vacuum-purged argon chamber are made in beryllium and tantalum.

*Plasma-arc welding (plasma welding)**

Plasma welding is a process which complements, and in some cases is a substitute for, the TIG process, offering, for certain applications, greater welding speed, better weld quality and less sensitivity to process variations. The constricted arc allows lower current operation than TIG for similar joints and gives a very stable controllable arc at currents down to 0.1 A, below the range of the TIG arc, for welding thin metal foil sections. Manual plasma welding is operated over a range 0.1–100 A from foil to 3–4 mm thickness in stainless steel, nickel and nickel alloys, copper, titanium and other rare earth metals (not aluminium or magnesium).

---

* American designation: plasma-arc welding (PAW).

Ions and plasma

An ion is an atom (or group of atoms bound into a molecule) which has gained or lost an electron or electrons. In its normal state an atom exhibits no external charge but when transference of electrons takes place an atom will exhibit a positive or negative charge depending upon whether it has lost or gained electrons. The charged atom is called an ion and when a group of atoms is involved in this transference the gas becomes ionized.

All elements can be ionized by heat to varying degrees (thermal ionization) and each varies in the amount of heat required to produce a given degree of ionization, for example argon is more easily ionized than helium. In a mass of ionized gas there will be electrons, positive ions and neutral atoms of gas, the ratio of these depending upon the degree of ionization.

A plasma is the gas region in which there is practically no resultant charge, that is, where positive ions and electrons are equal in number; the region is an electrical conductor and is affected by electric and magnetic fields. The TIG torch produces a plasma effect due to the shield of argon and the tungsten arc but a plasma jet can be produced by placing a tungsten electrode centrally within a water-cooled constricted copper nozzle. The tungsten is connected to the negative pole (cathode) of a d.c. supply and the nozzle to the positive pole (anode). Gas is fed into the nozzle and when an arc is struck between tungsten electrode and nozzle, the gas is ionized in its passage through the arc and, due to the restricted shape of the nozzle orifice, ionization is greatly increased and the gas issues from the nozzle orifice, as a high-temperature, high-velocity plasma jet, cylindrical in shape and of very narrow diameter realizing temperatures up to 10000°C. This type is known as the non-transferred plasma (Fig. 3.31*a*). With the transferred arc process used for welding, cutting and surfacing, the restricting orifice is in an inner water-cooled nozzle within which the tungsten electrode is centrally placed. Both work and nozzle are connected to the anode and the tungsten electrode to the cathode of a d.c. supply (in American terms, d.c.s.p., direct current straight polarity). Relatively low plasma gas flow (of argon, argon–helium or argon–hydrogen) is necessary to prevent turbulence and disturbance of the weld pool, so a further supply of argon is fed to the outer shielding nozzle to protect the weld (Fig. 3.31*b*). In the lead from work to power unit there is a contactor switch as shown.

A high-frequency unit fed from a separate source from the mains supply initiates the pilot arc and the torch nozzle is positioned exactly over the work. Upon closing the contactor switch in the work-to-power unit connexion, the arc is transferred from electrode to work via the plasma. Temperatures up to 17000 °C can be obtained with this arc. To shape the arc two auxiliary gas passages on each side of the main orifice may be included in the nozzle design. The flow of cooler gas through these squeezes the circular pattern of the jet into oval form, giving a narrower heat-affected zone and increased welding speed. If a copper electrode is used instead of tungsten as in the welding of zirconium, it is made the anode. The low-current arc plasma with currents in the range 0.1–15 A has a longer operating length than the TIG arc, with much greater tolerance to change

Fig. 3.31. (*a*) Non-transferred plasma arc. (*b*) Transferred plasma arc.

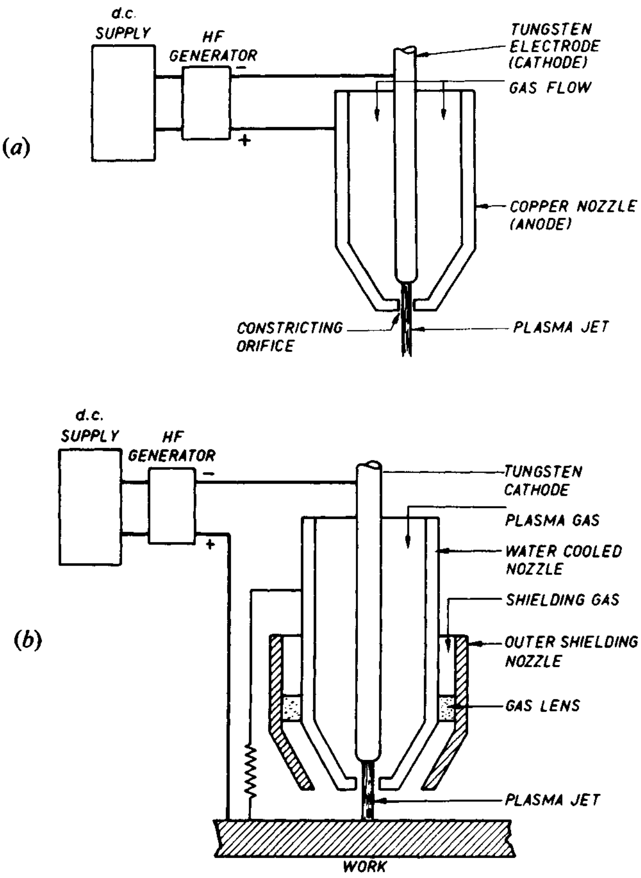
(a)
d.c. SUPPLY
HF GENERATOR
−
+
TUNGSTEN ELECTRODE (CATHODE)
GAS FLOW
COPPER NOZZLE (ANODE)
CONSTRICTING ORIFICE
PLASMA JET
(b)
d.c. SUPPLY
HF GENERATOR
−
+
TUNGSTEN CATHODE
PLASMA GAS
WATER COOLED NOZZLE
SHIELDING GAS
OUTER SHIELDING NOZZLE
GAS LENS
PLASMA JET
WORK

in arc length without significant variation in the heat energy input into the weld. This is because it is straight, of narrow diameter, directional and cylindrical, giving a smaller weld pool, deeper penetration and less heat spread whereas the TIG arc is conical so that small changes in arc length have much more effect on the heat output. Fig. 3.32 compares the two arcs. Since the tungsten electrode is well inside the nozzle (about 3 mm) in plasma welding, tungsten contamination by touchdown or by filler rod is avoided, making welding easier.

Equipment. In the range 5–200 A a d.c. rectifier power unit with drooping characteristic and an OCV of 70 V can be used for argon and argon–hydrogen mixtures. If more than 5% hydrogen is used, 100 V or more is required for pilot arc ignition. This arc may be left in circuit with the main arc to give added stability at low current values. Existing TIG power sources such as that in Fig. 3.11 may be used satisfactorily, reducing capital cost, the extra equipment required being in the form of a console placed on or near the power unit. Input is 380–440 V, 50 Hz, a.c. single-phase with approximately 3.5 A full load current. It houses relays and solenoid valves controlling safety interlocks to prevent arc initiation unless gas and water pressures are correct; flowmeters for plasma and shielding gases, gas purge and post-weld gas delay and cooling water controls. For low-current welding in the range 0.1–15 A, often referred to in Europe as micro-plasma and in America as needle plasma welding, the power unit is about 3 KVA, 200–250 V, 50 Hz single-phase input, fan cooled with OCV of 100 V nominal and 150 V peak d.c., for the main arc, and output current ranges 0.1–2.0 A and 1.0–15 A. The pilot arc takes 6A at 25 V start and 2.5 A at 24 V running, the main arc being cylindrical and only 0.8 mm wide. Tungsten electrodes are 1.6, 2.4 and 3.2 mm diameter, depending upon application.

Gases. Very pure argon is used for plasma and shield (or orifice) when welding reactive metals such as titanium and zirconium which have a strong affinity for hydrogen. For stainless steel and high-strength nickel alloys argon, or argon–hydrogen mixtures are used. Argon–5% hydrogen mixtures are applied for this purpose (cylinders coloured blue with a wide red band around the middle) and argon–8% hydrogen and even up to 15% hydrogen are also used. With these mixtures the arc voltage is increased giving higher welding speeds (up to 40% higher) and the thin oxide film present even on stainless and alloy steels is removed by hydrogen reduction giving a clean bright weld, and the wetting action is improved. For copper, nickel and their alloys argon is used in keyhole welding (Fig. 3.33) for thinner sections, for both orifice and shielding gas. Argon and helium is used in the melt-in welding of thinner sections and helium for orifice and shielding gas for sections over 3 mm thick.

De-ionized cooling water should be used, and gas flow rates at 2.0 bar pressure are about 0.25–3.3 litres per minute plasma and 3.8–7 litres per minute shielding in the 5–100 A range. Cleaning and preparation are

Fig. 3.32. (*a*) Plasma arc. (*b*) TIG arc.

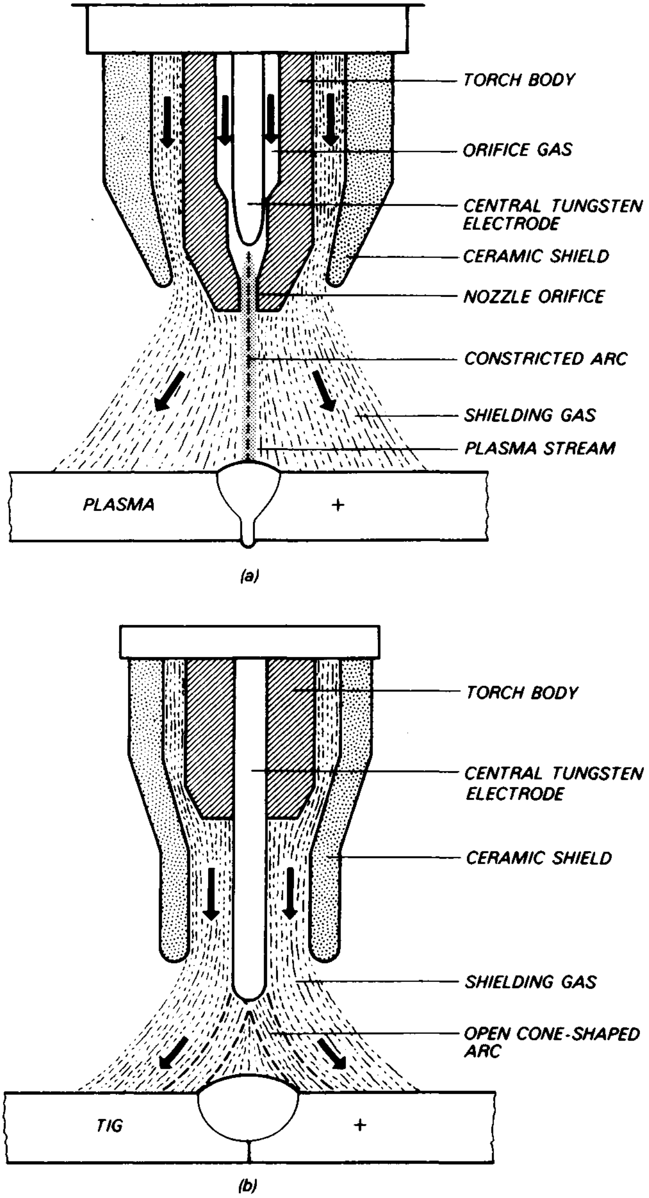
TORCH BODY
ORIFICE GAS
CENTRAL TUNGSTEN ELECTRODE
CERAMIC SHIELD
NOZZLE ORIFICE
CONSTRICTED ARC
SHIELDING GAS
PLASMA STREAM
PLASMA
+
(a)
TORCH BODY
CENTRAL TUNGSTEN ELECTRODE
CERAMIC SHIELD
SHIELDING GAS
OPEN CONE-SHAPED ARC
TIG
+
(b)

similar to those for TIG, but when welding very thin sections, oil films or even fingerprints can vary the degree of melting, so that degreasing should be thorough and the parts not handled in the vicinity of the weld after cleaning.

Technique. Using currents of 25–100 A, square butt joints in stainless steel can be made in thicknesses of 0.8–3.2 mm with or without filler rod, the angle of torch and rod being similar to that for TIG welding. The variables are current, gas flow and welding rate. Too high currents may break down the stabilizing effect of the gas, the arc wanders and rapid nozzle wear may occur. In 'double arcing' the arc extends from electrode to nozzle and then to the work. Increasing plasma gas flow improves weld appearance and increases penetration, and the reverse applies with decreased flow. A turbulent pool gives poor weld appearance.

In 'keyhole' welding of thicknesses of 2.5–6.5 mm a hole is formed in the square-edge butt joint at the front edge of the molten pool, with the arc passing through the section (Fig. 3.33). As the weld proceeds, surface tension causes the molten metal to flow up behind the hole to form the welding bead, indicating complete penetration, and gas backing for the underbead is required. When butt welding very thin sections, the edges of the joint must be in continuous contact so that each edge melts and fuses into the weld bead. Separation of the edges gives separate melting and no weld. Holding clamps spaced close together near the weld joint and a backing bar should be used to give good alignment, and gas backing is recommended to ensure a fluid pool and good wetting action. Flanging is recommended for all butt joints below 0.25 mm thickness and allows for greater tolerance in alignment. In the higher current ranges joints up to 6 mm thick can be welded in one run. Lap and fillet welds made with filler rod are similar in appearance and employ a similar technique to those made with the manual TIG process. Edge welds are the best type of joint for foil thickness, an example being a plug welded into a thin-walled tube.

In tube welding any inert gas can be used within the tube for underbead protection.

Faults in welding very thin sections are: excessive gaps which the metal cannot bridge; poor clamping allowing the joint to warp; oil films varying the degree of melting; oxidation or base metal oxides preventing good

Fig. 3.33. Surface tension causes molten metal to flow and close keyhole.

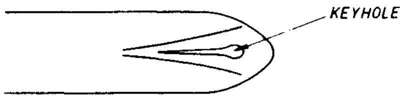

'wetting'; and nicking at the ends of a butt joint. This latter can be avoided by using run-off plates or by using filler rod for the start and finish of the weld.

Plasma spray hard facing

In this method the arc plasma melts the hard facing powder particles and the high-velocity gas stream carries the molten particles on to the surface (Fig. 3.34). Although there may be some voids caused, 90–95% densities can be obtained and the method is particularly suitable for applying refractory coatings because this method has a better bond with the parent plate than the spray and fuse method. Because the heat is localized, surfaces can be applied to finished parts with no distortion and the finish is very smooth. The cobalt, nickel- and tungsten-based powders are all supplied for this process and suitable applications are gas turbine parts, such as sealing rings of gas turbines, mixer and feed parts, sleeves, sheets and wear pads and slides, etc.

Plasma transferred arc hardfacing

This powder deposition process is fully automated and uses a solid state SCR, the power output being 300 A, 32 V, at 100% duty cycle. The pilot arc is initiated by HF and the unit has pulse facilities. The powder is supplied from a powder feeder operating on the rotating drum principle. Powder feed rate is controlled by varying the feed opening over the knurled rotating wheel upon which a regulated ribbon of powder is applied, and the powder feed can be regulated to provide an upslope and downslope feed rate. This powder is carried from the container to the work in a stream of argon and is melted in the plasma, the argon providing a shield around the plasma heat zone (Fig. 3.35). Torches are of varying size and cover most

Fig. 3.34. Plasma surfacing.

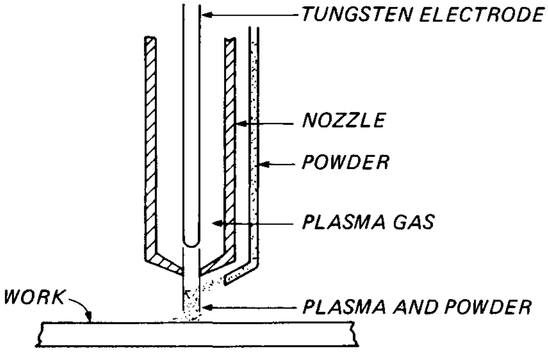

hard surfacing requirements. As can be seen in the illustration the plasma and powder are supplied through separate passages. Cobalt-based (stellite), nickel-based, iron-based with chrome or molybdenum alloys are available and applications include valve seats, oil drill tool joints, gate valve inserts, diesel engine crossheads, etc. Stellite alloys Nos. 1, 6, 12, 21, 156, 157, 158, F of the stellite range are suitable.

Fig. 3.35. Plasma transferred arc.

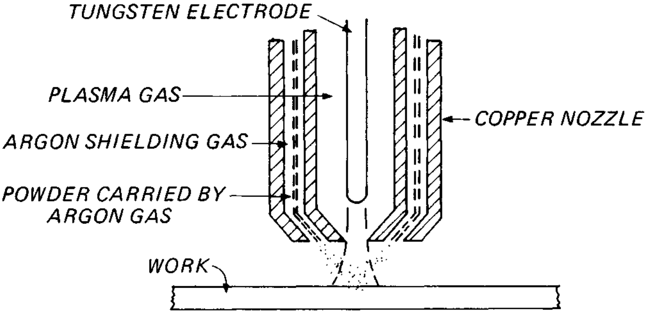
TUNGSTEN ELECTRODE
PLASMA GAS
ARGON SHIELDING GAS
POWDER CARRIED BY ARGON GAS
WORK
COPPER NOZZLE

4

Resistance welding and flash butt welding

Spot welding

In this method of welding use is made of the heating effect which occurs when a current flows through a resistance. Suppose two steel plates A and B (Fig. 4.1) are to be welded together at X. Two copper electrodes are pressed against the plates squeezing them together. The electrical resistance is greatest at the interface where the plates are in contact, and if a large current at low voltage is passed between the electrodes through the plates, heat is evolved at the interface, the heat evolved being equal to I^2Rt joules, where I is the current in amperes, R the resistance in ohms and t the time in seconds. A transformer supplies a.c. at low voltage and high current to the electrodes (Figs. 4.1 and 4.2).

Fig. 4.1

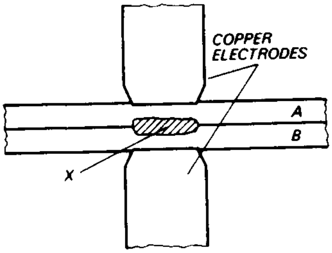
COPPER ELECTRODES
A
B
X

Fig. 4.2

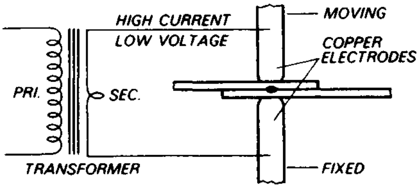
HIGH CURRENT
LOW VOLTAGE
MOVING
COPPER ELECTRODES
PRI.
SEC.
TRANSFORMER
FIXED

For any given joint between two sheets a suitable time is selected and the current varied until a sound weld is obtained. If welds are made near each other some of the current is shunted through the adjacent weld (Fig. 4.3) so that a single weld is not representative of what may occur when several welds are made. Once current and time are set, other welds will be of consistent quality. If the apparatus is controlled by a pedal as in the simplest form of welder, then pressure is applied mechanically and the time for which the current flows is switch controlled, as for example in certain types of welding guns used for car body repairs. This method has the great disadvantage that the time cannot be accurately controlled, so that if it is too short there is insufficient heat and there is no fusion between the plates, whilst if the time is too long there is too much heat generated and the section of the plates between the electrodes melts, the molten metal is spattered out due to the pressure of the electrodes and the result is a hole in the plates. In modern spot-welding machines the pressure can be applied pneumatically or hydraulically or a combination of both and can be accurately controlled. The current is selected by a tapping switch on the primary winding of the transformer and the time is controlled electronically, the making and breaking of the circuit being performed by thyristors.

When the current flows across the interface between the plates, the heating effect causes melting and fusion occurs at *A* (Fig. 4.4). Around this there is a narrow heat-affected zone (HAZ) since there is a quenching effect

Fig. 4.3

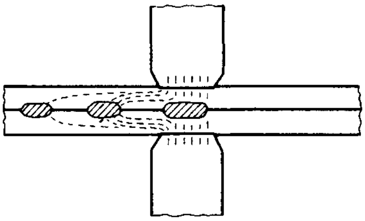

Fig. 4.4

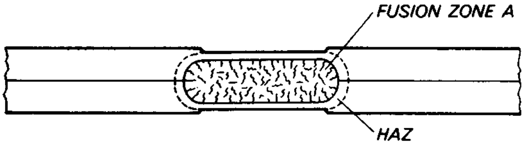

due to the electrodes, which are often water cooled. There are equi-axed crystals in the centre of the nugget and small columnar crystals grow inwards towards the centre of greatest heat.

Types of spot welders

In the pedestal type there is a fixed vertical pedestal frame and integral transformer and control cabinet. The bottom arm is fixed to the frame and is stationary during welding and takes the weight of the workpiece. The top arm may be hinged so as to move down in the arc of a circle (Fig. 4.5*a*) or it may be moved down in a straight line (Fig. 4.5*b*). Pivoting arms are adjustable so as to have a large gap between the electrodes, the arms are easily adjusted in the hubs and various length arms are easily fitted giving easy access to difficult joints. The vertical travel machine has arms of great rigidity so that high pressures can be applied, and the electrode tips remain in line irrespective of the length of stroke, so that the machine is easily adapted for projection welding. Additionally the spot can be accurately positioned on the work, and since the moving parts have low inertia, high welding speeds can be achieved without hammering (see Fig. 4.6*a* and *b*).

Welding guns

Portable welding guns are extensively used in mass production. The equipment consists of the welding station often with hydro-pneumatic booster to apply the pressures, a water manifold, one or two welding guns

Fig. 4.5. (*a*) Pivoting arms, (*b*) vertical travel.

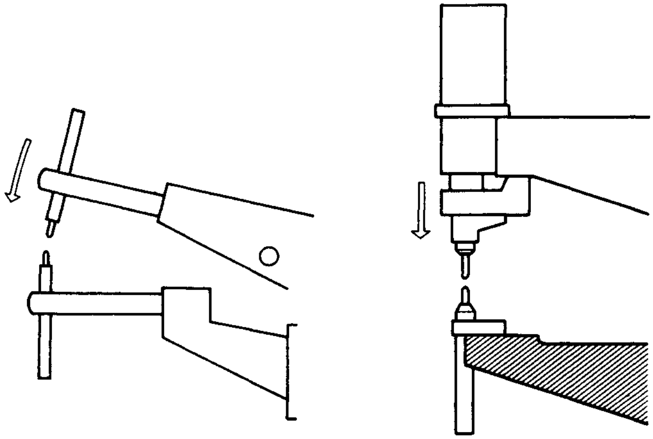

(a) PIVOTTING ARMS *(b) VERTICAL TRAVEL*

with balancers, cable between transformer and guns and a control station. In modern machines the composite station with built-in cabinet comprises sequence controls with integrated circuits, thyristor contactors and disconnect switch. Articulated guns (Figs. 4.7*a* and *b*) have both arms articulated as in a pair of scissors, giving a wide aperture between electrodes, and are used for welding joints difficult of access. Small

Fig. 4.6. (*a*) Air-operated spot welding machine with pivoting arms.

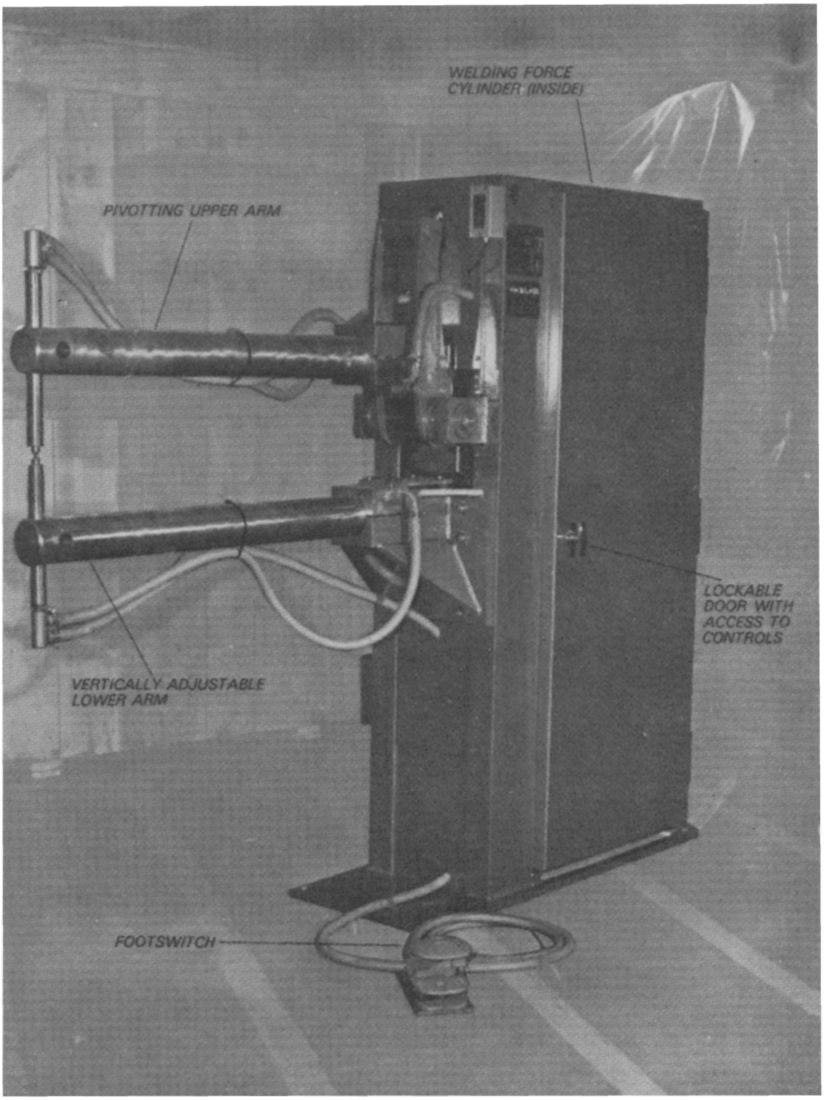
WELDING FORCE CYLINDER (INSIDE)
PIVOTTING UPPER ARM
LOCKABLE DOOR WITH ACCESS TO CONTROLS
VERTICALLY ADJUSTABLE LOWER ARM
FOOTSWITCH

articulated guns are used for example in car body manufacture for welding small flanges, and in corners and recesses. C guns (Fig. 4.7*c*) have the piston-type ram of the pressure cylinder connected to the moving electrode, which thus moves in a straight line. There is great rigidity and a high working speed because of the low inertia of the moving parts. The electrodes are always parallel and the precision motion is independent of

Fig. 4.6. (*b*) Air-operated spot welding machine with vertical action.

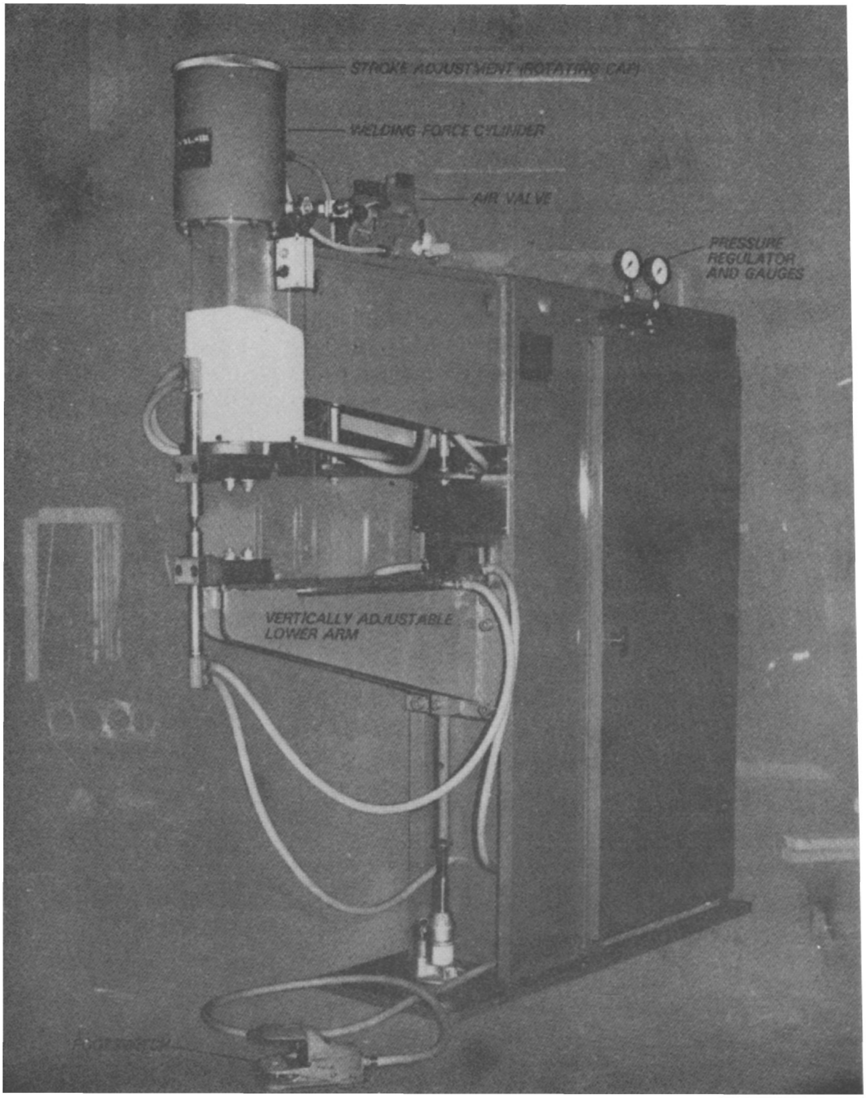
STROKE ADJUSTMENT (ROTATING CAP)
WELDING FORCE CYLINDER
AIR VALVE
PRESSURE REGULATOR AND GAUGES
VERTICALLY ADJUSTABLE LOWER ARM

Fig. 4.7. (*a*) Articulated type welding head; air operated.

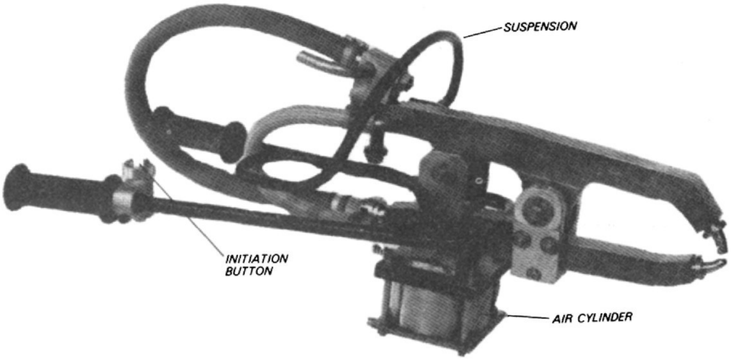
SUSPENSION
INITIATION BUTTON
AIR CYLINDER

Fig. 4.7. (*b*) Articulated type welding head; hydraulically operated.

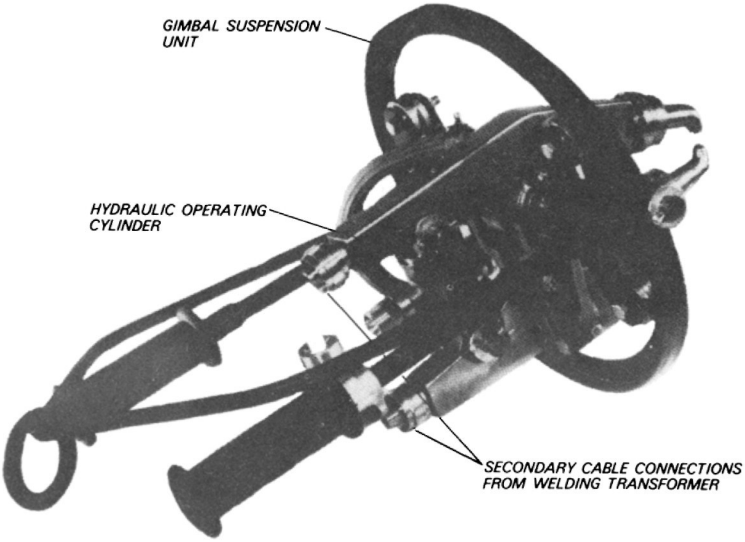
GIMBAL SUSPENSION UNIT
HYDRAULIC OPERATING CYLINDER
SECONDARY CABLE CONNECTIONS FROM WELDING TRANSFORMER

Fig. 4.7. (*c*) Air-operated C type welding head.

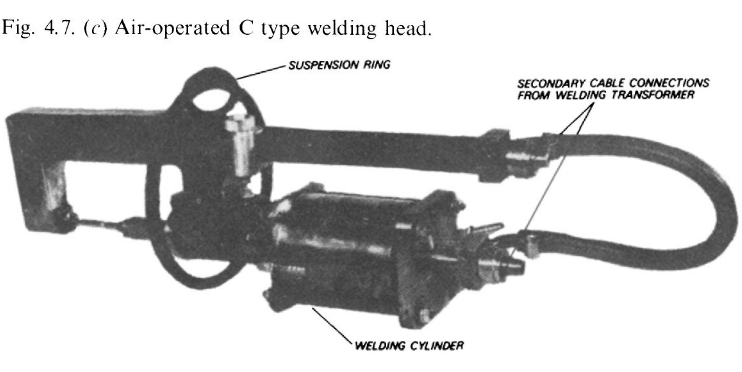
SUSPENSION RING
SECONDARY CABLE CONNECTIONS FROM WELDING TRANSFORMER
WELDING CYLINDER

the arm length with easily determined point of welding contact. They are used for quality welds where the height does not preclude them. Integral transformer guns are used to reduce handling costs of bulky fabrications, and the manual air-cooled gun and twin spot gun are used in repair work on car bodies and agricultural equipment and for general maintenance, the latter being used where there is access for only one arm as in closed box sections and car side panels.

The welding cycle

The cycle of operations of most modern machines is completely automatic. Once the hand or foot switch is pressed, the cycle proceeds to completion, the latest form having digital sequence controls with printed circuits, semi-conductors, digital counting and air valve operation by relays. The simplest cycle has one function, namely weld time, the electrode force or pressure being pre-set, and it can be best illustrated by two graphs, one the electrode pressure plotted against time and the other the current plotted against the same time axis (Fig. 4.8). The time axis is divided into 100 parts corresponding to the 50 Hz frequency of the power supply. Fig. 4.8 shows a simple four-event control. First the squeeze is applied and held constant (*a–b* on the graph). The current is switched on at *X*, held for seven complete cycles (7/50 s) and switched off at *Y*. The pressure is held until *c* on the graph, the weld cooling in this period. Finally the squeeze is released and the repeat cycle of operations begins again at a_1. The switches comprise repeat–non-repeat, thyristors on–off and heat control in–out.

For more complex welding cycles, necessary when welding certain metals to give them correct thermal treatment, the graphs are more complex with up to ten functions with variable pressure cycle, and these machines are often operated from three phases. Machines for the aeronautical and space industries are specially designed and incorporate variable pressure heads

Fig. 4.8. Standard four-event sequence for spot welding.

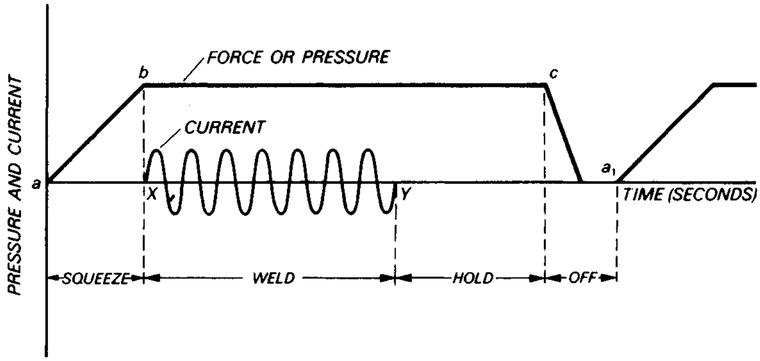

and up to ten functions. Most aluminium alloys lose the properties given to them by work hardening or heat treatment when they are heated. To enable the correct thermal treatment of heating and cooling to be given to them during the welding cycle three-phase machines are often employed.

The transformer

The transformer steps the voltage down from that of the mains to the few volts necessary to send the heavy current through the secondary welding circuit. When the current is flowing the voltage drop across the secondary may be as low as 3–4 volts and in larger machines the current can be up to 35000 A. Because of these large currents there is considerable force acting between the conductors due to the magnetic field, so transformers must be robustly constructed or movement may cause breakdown of the insulation, and in addition they may be water cooled because of the heating effect of the current. The secondary winding consists of one or two turns of copper strip or plates over which the primary coils fit, the whole being mounted on the laminated iron core. Because any infiltration of moisture may lead to breakdown, modern transformers have the primary winding and secondary plates assembled as a unit which is then vacuum encapsulated in a block of epoxy resin. The primary winding has tappings which are taken to a rotary switch which selects the current to be used.

The current flowing in the secondary circuit of the transformer provides the heating effect and depends upon (1) the open circuit voltage and (2) the impedance of the circuit, which depends upon the gap, throat depth and magnetic mass introduced during welding and the resistance of the metal to be welded. The duty cycle is important, as with all welding machines, as it affects the temperature rise of the transformer. For example, if a spot welder is making 48 spot welds per minute, each of 0.25 seconds duration, the duty cycle is $(48 \times 0.25 \times 100) \div 60 = 20\%$. Evidently knowing the duty cycle and welding time, the number of welds that can be made per minute can be calculated.

Electrodes

The electrode arms and tips (Fig. 4.9*a*) which must carry the heavy currents involved and apply the necessary pressure must have the following properties: (1) high electrical conductivity so as to keep the I^2R loss (the heating due to the resistance) to a minimum, (2) high thermal conductivity to dissipate any heat generated, (3) high resistance to deformation under large squeeze pressures, (4) must keep their physical properties at elevated temperatures, (5) must not pick up metal from the surface of the workpiece,

Electrode	Properties	Uses
(1) Copper, 1% silver	High conductivity, medium hardness.	Light alloys, coated sheets, scaly steel.
(2) Copper, 0.6% Cr or 0.5% Cr and Be	Best electrical conductivity with greatest hardness.	Clean or lightly oxidized steel, brass and cupro-nickel. Used for the arms of the machine and electrical conductors.
(3) Copper, 2.5% Co and 0.5% Be	Poor conductivity but very hard.	For welding hard metals with high resistivity, e.g. stainless steel, heat-resistant and special steels.
(4) Copper, beryllium	Poor conductivity, great hardness.	For clamping jaws of flash butt welders.
(5) Molybdenum		Drawn bar or forged buttons used as pressed fit inserts in supports of (2) above. For welding thin sheet or wires or electro-brazing silver-based metals.
(6) Sintered copper-tungsten	Fair conductivity, very hard.	Keeps its mechanical properties when hot.

Note. The addition of various elements increases the hardness but reduces the conductivity.

(6) must be of reasonable cost. The following types of electrodes are chiefly used, and are given in the table. Electrolytic copper (99.95% Cu) has high electrical and thermal conductivity and, when work hardened, resists deformation but at elevated temperatures that part of the electrode tip in contact with the work becomes annealed due to the heating and the tip softens and deforms. Because of this it is usual to water-cool the electrodes to prevent excessive temperature rise.*

Water cooling

Adequate electrode cooling is the most essential factor to ensure optimum tip life; the object is to prevent the electrode material from reaching its softening temperature, at which point it will lose its hardness

Fig. 4.9. (*a*) Electrodes and holders.

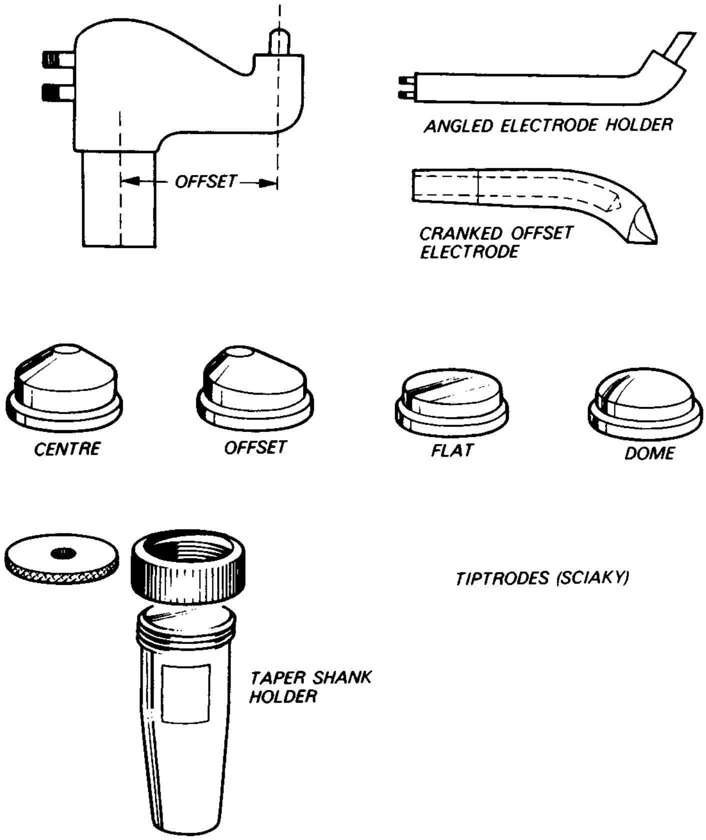
OFFSET
ANGLED ELECTRODE HOLDER
CRANKED OFFSET ELECTRODE
CENTRE
OFFSET
FLAT
DOME
TIPTRODES (SCIAKY)
TAPER SHANK HOLDER

---

* See also BS 807 and BS 4215. Spot welding electrodes and electrode holders.

and rapidly deteriorate. The normal cooling method is by internal water circulation where the water is fed via a central tube arranged to direct the water against the end of the electrode cooling hole (Fig. 4.9*b*). In some cases a short telescopic extension tube is used inside the main tube and must be adjusted to suit the length of the electrode used.

Electrodes are available with a taper fit to suit the electrode arms and also as tips to fit on to a tapered shank body, which makes them easily interchangeable. The face that makes contact with the metal to be welded can be a truncated cone with central or offset face, flat or domed (Fig. 4.9) and tip diameters can be calculated from formulae which depend upon the plate thickness, but in all cases these are approximate only and do not replace a test on the actual part. Electrodes are subject to great wear and tear in service due to the constant heating and cooling and varying pressure

Fig. 4.9. (*b*)

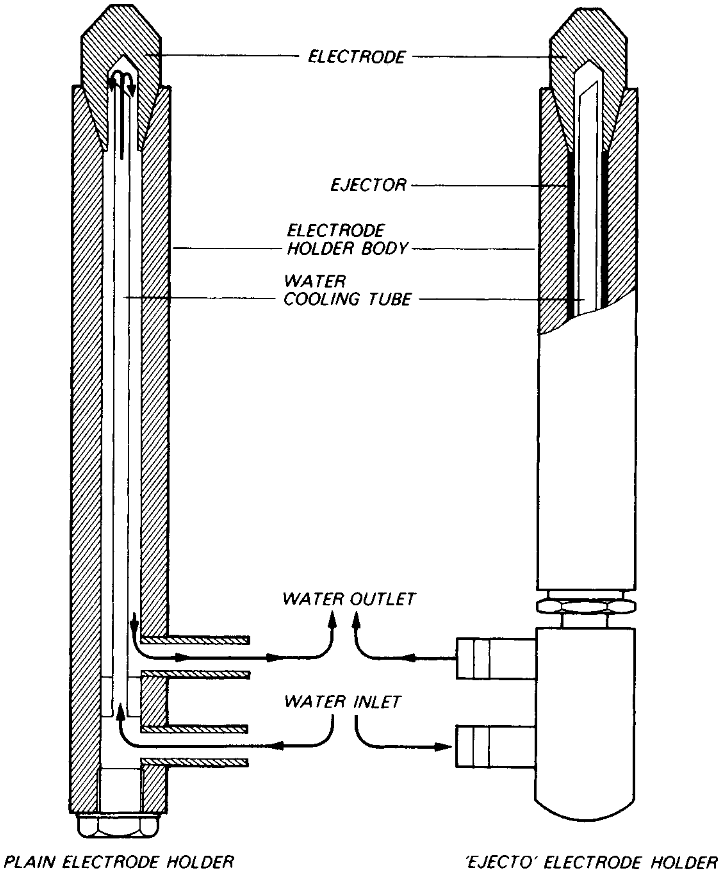
ELECTRODE
EJECTOR
ELECTRODE HOLDER BODY
WATER COOLING TUBE
WATER OUTLET
WATER INLET
PLAIN ELECTRODE HOLDER
'EJECTO' ELECTRODE HOLDER

cycles. The chief causes of wear are: electrical; wrong electrode material, poor surface being welded, contacts not in line; mechanical; electrode hammering, high squeeze, weld and forge pressures, abrasion in loading and unloading and tearing due to the parting of the electrodes.

Seam welding

In a seam welding machine the electrodes of a spot welder are replaced by copper alloy rollers or wheels which press on the work to be welded (Fig. 4.10*a* and *b*). Either one or both are driven and thus the work passes between them. Current is taken to the wheels through the rotary bearings by silver contacts with radial pressure and the drive may be by knurled wheel or the more usual shaft drive which enables various types of wheel to be easily fitted. If the current is passed continuously a continuous seam weld results but, as there is a shunt effect causing the current to flow through that part of the weld already completed, overheating may occur resulting in burning of the sheets. To avoid this the current can be pulsed, allowing sufficient displacement of the already welded portion to take place and thus obviating most of the shunt effect. For materials less than 0.8 mm thick or at high welding speeds (6 m/min) no pulsing is required, the 50 Hz frequency of the supply providing a natural pulse. Above 2 × 0.8 mm thickness pulsation is advisable, and essential above 2 × 1.5 mm thickness, while for pressure-tight seams the welds can be arranged to overlap and if the seams are given a small overlap with wide-faced wheels and high pressure a mash weld can be obtained.

By the use of more complex electro-mechanical bearing assemblies, longitudinal and circumferential welds can be made (Fig. 4.10*c*).

Seam welding guns are extremely useful for fabricating all types of tanks, exhaust systems, barrels, drip-mouldings on car body shells, etc. They have electrode drive which automatically propels the gun along the seam so that

Fig. 4.10. (*a*)

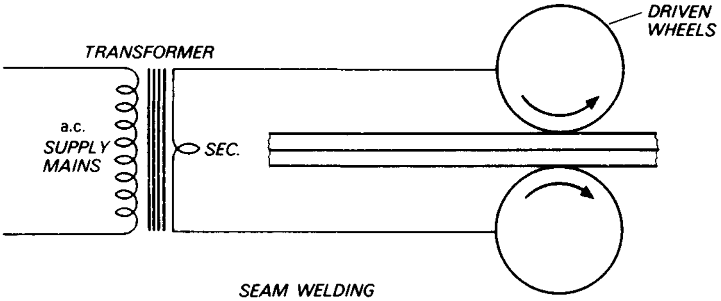

Fig. 4.10. (*b*) 'Thin wheel' seam welding machine with silver bearings.

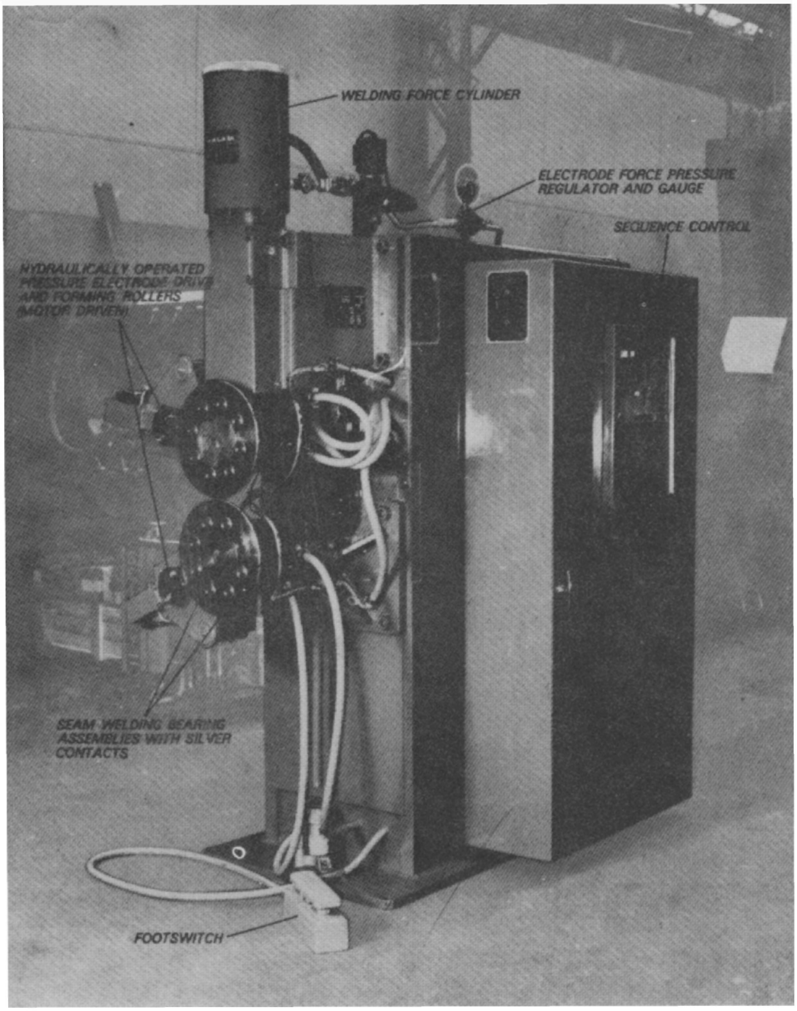
WELDING FORCE CYLINDER
ELECTRODE FORCE PRESSURE REGULATOR AND GAUGE
SEQUENCE CONTROL
HYDRAULICALLY OPERATED PRESSURE ELECTRODE DRIVE AND FORMING ROLLERS (MOTOR DRIVEN)
SEAM WELDING BEARING ASSEMBLIES WITH SILVER CONTACTS
FOOTSWITCH

Fig. 4.10. (*c*)

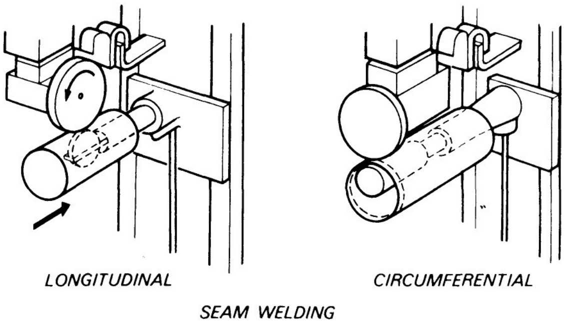
LONGITUDINAL
CIRCUMFERENTIAL
SEAM WELDING

it only requires guidance, and they are operated in the same way as spot welding guns.

Direct current spot welding

We have seen that when iron or steel to be welded is placed in a spot or seam welder the impedance of the secondary circuit is increased and the secondary current varies according to the mass introduced. By using d.c. this loss is obviated, the power consumed is reduced and the electrode life increased because skin effect is eliminated. This enables coated steels, stainless and other special steels in addition to aluminium alloys to be welded to high standards. The machine is similar in appearance to the a.c. machine but has silicon diodes as the rectifying elements.

Three-phase machines

Three-phase machines have been developed to give impulse of current in the secondary circuit at low frequency, with modulated wave form to give correct thermal treatment to the material being welded. These machines have greatly increased the field of application of spot welding into light alloys, stainless, and heat-resistant steels, etc. A typical sequence of operations is: squeeze, which multiplies the number of high spots between the contact faces; welding pressure, which diminishes just before the welding current flows; immediately after the passage of the current, forging or recompression pressure is applied which is above the elastic limit of the material and completes the weld. Setting of the welding current can be done by welding test pieces with increasing current until the diameter of the nugget, found by 'peeling' the joint, is $(2t + 3)$, where t is the thickness of the thinner sheet. This current is noted. It is then increased until spatter of the nugget occurs and the welding current is taken as midway between these two values, with final adjustments made on test pieces.

Nickel and nickel alloys have higher electrical resistivity and lower thermal conductivity than steel and are usually welded in thin sections as lap joints, although the crevice between the sheets may act as a stress raiser and affect the corrosion and fatigue resistance. High pressures are required for the high-nickel alloys so as to forge the solidifying nugget, and the machine should have low inertia of the moving parts so that the electrode has rapid follow-up during welding. Current may be set as in the previous section, and in some cases it is advantageous to use an initial squeeze followed by an increasing (up-sloping) current. When the nugget is just beginning to form with diffusion of the interface, the squeeze force is reduced to about one-third of its initial value and held until the current is switched off; this reduces danger of expulsion of molten metal and gives better penetration of the weld into the two sheets.

Projection welding

Protrusions are pressed on one of the sheets or strips to be welded and determine the exact location of the weld (Fig. 4.11). Upon passage of the current the projection collapses under the electrode pressure and the sheets are welded together. The machines are basically presses, the tipped electrodes of the spot welder being replaced by flat platens with T slots for the attachment of special tools, and special platens are available which allow the machine to be used as a spot welder by fitting arms and electrodes (Fig. 4.12), and automatic indexing tables can be used to give increased output. Projection welding is carried out for a variety of components such as steel radiator coupling elements, brake shoes, tin-plate tank handles and spouts, etc. The press type of machine is also used for resistance brazing in which the joining of the parts is achieved with the use of an alloy with a lower melting point than the parent metal being welded so that there is no melting involved.

Cross wire welding is a form of projection welding, the point of contact of the two wires being the point of location of the current flow. Low-carbon mild steel, brass, 18/8 stainless steel, copper-coated mild steel and galvanized steel wire can be welded but usually the bulk of the work is done with clean mild steel wire, bright galvanized, or copper coated as used for milk bottle containers, cages, cooker and refrigerator grids, etc., and generally several joints are welded simultaneously.

Modern machines are essentially spot welders in ratings of 25, 70, or 150 kVA fitted with platens upon which suitable fixtures are mounted and having a fully controlled pneumatically operated vertical head, and the electrical capacity to weld as many joints as possible simultaneously. Large programmed machines are manufactured for producing reinforcement mesh automatically.

Fig. 4.11. Projection welding.

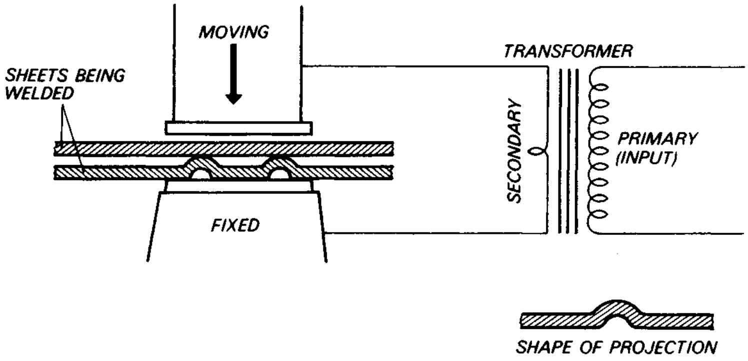

Resistance butt welding

This method is similar to spot welding except that the parts to be butt welded now take the place of the electrodes. The two ends are prepared so that they butt together with good contact. They are then placed in the jaws of the machine, which presses them close together end to end (Fig. 4.13). When a given pressure has been reached, the heavy current is switched on, and the current flowing through the contact resistance

Fig. 4.12. Projection welder.

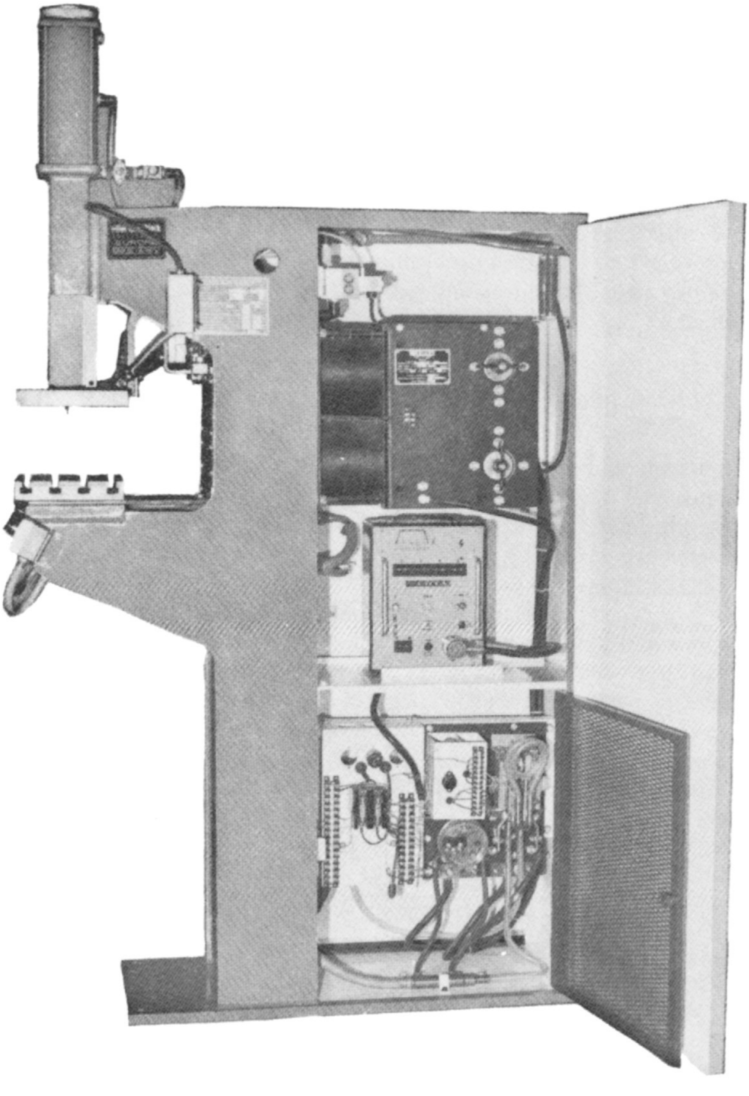

between the ends brings them to welding heat. Extra pressure is now applied and the ends are pushed into each other, the white hot metal welding together and an enlargement of section taking place. The section may be machined to size after the operation if necessary.

*Flash butt welding**

Although this is not a resistance welding process it is convenient to consider it here. Flash butt welding machines must be very robustly made and have great rigidity because considerable pressures are exerted and exact alignment of the components is of prime importance. The clamping dies of copper alloy, which carry the current to the components and hold them during butting up under high pressure, should grip over as large an area as possible to reduce distortion tendencies. The clamping pressure, which is about twice the butting pressure, is usually done pneumatically or hydraulically, and current is of the order of 7–10 A/mm^2 of joint area. The following are the stages in the welding cycle.

Pre-heat. The components are butted together and a current passing across the joint heats the ends to red heat.

Flash. The parts are separated and an arc is established between them until metal begins to melt, one of the components being moved to keep the arc length constant.

Upset. The parts are butted together under high pressures with the current still flowing. Impurities are forced out of the joint in the butting process and an impact ridge or flash is formed (Fig. 4.14). Post-heat treatment can be given by a variation of current and pressure after welding. For welding light alloys, pre-heating is generally dispensed with and the flashing is of short duration.

As an example of currents involved, a butt weld in 6 mm thick 18/18/3

Fig. 4.13. Resistance butt welding.

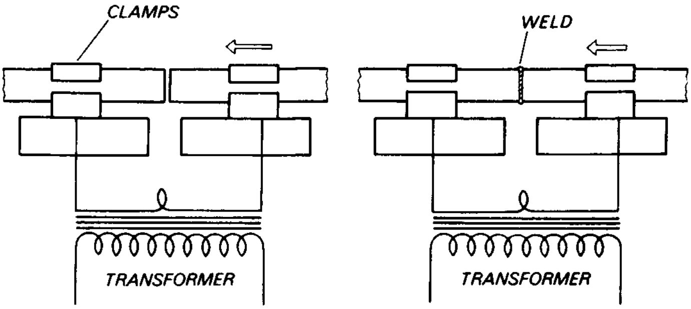

---

* BS 499 gives this as a resistance process.

stainless steel with a cross-section of 600 mm^2 involves currents of 20 000 A with a 9 second flashing time.

Flash butt welding is used very extensively by railway systems of the world to weld rails into continuous lengths. British Rail for example weld rails approximately 18.3 m long into continuous lengths from 91 to 366 m and conductor rails are welded in the same way. Fig. 4.15 shows a modern rail welding machine. Hydraulic rams, equally spaced on each side of the rail section, apply the forging load of 200–400 kN to the rails of section approximately 7200 mm^2. The rail is clamped by two vertically acting cylinders and horizontally acting cylinders align each rail to a common datum and an anti-twist device removes axial twist. When welding long lengths of rail it is more convenient to move the machine to the exact position for welding rather than the rail and for this the machine can be rail mounted. Machines of this type can make 150–200 welds per 8-hour shift. In this case the sequence of operations is pre-flash, pre-heat, flashing and forging.

The upset is removed by a purpose-designed machine with hydraulic power-shearing action usually mounted in line with the welding machine.

The flash butt welded lengths are welded *in situ* on the track by the thermit process.

Fig. 4.14. Flash butt welding.

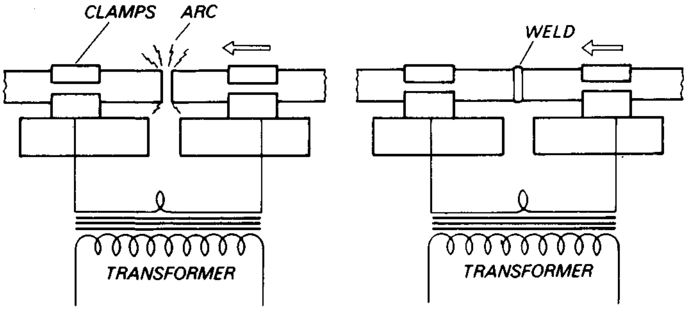

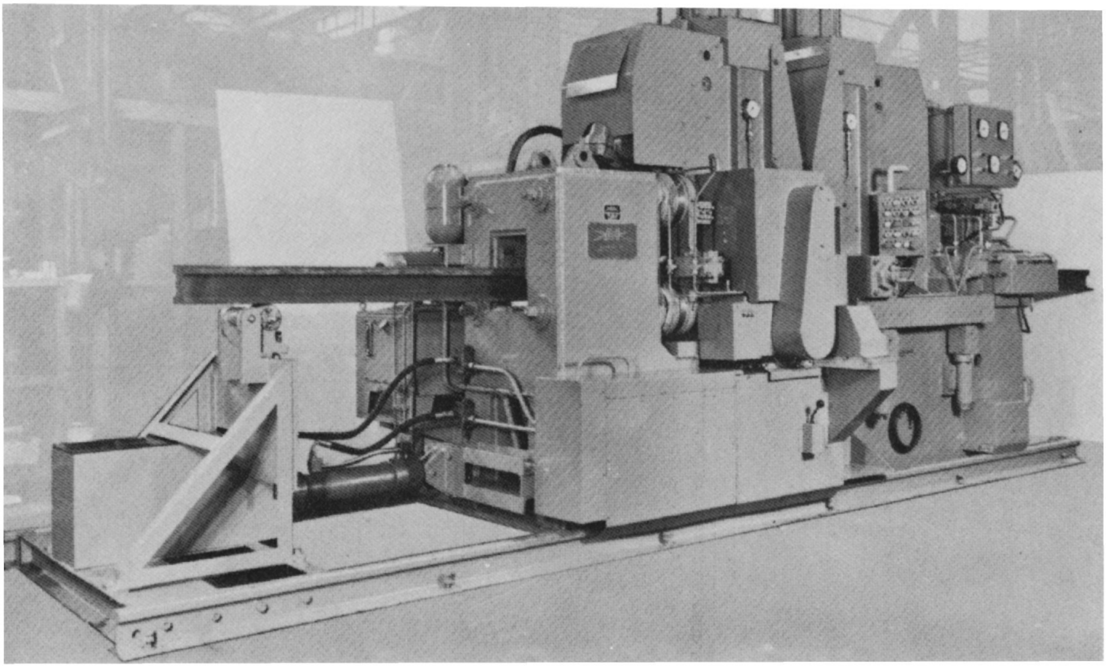

Fig. 4.15. Machine for welding rail sections. 1000 kVA. 3-phase, 440 V. Welding sequence: (1) Both rails securely clamped in welder. (2) Rails aligned vertically and horizontally under full clamping pressure. (3) Weld initiated by push button control. (4) Moving head moves forward on the burn-off or pre-flashing stroke used to square up the ends of the rails if necessary. On completion of this, pre-heating begins until the requisite number of pre-heats have lapsed. At this point the rails should be in a suitably plastic state to allow for straight flashing and finally forging.

5

Additional processes of welding

There has been a great increase in the number of automatic processes designed to speed up welding production. Automatic welding gives high rates of metal deposition because high currents from 400 to 2000 A can be used, compared with the limit of about 600 A with manual arc welding. Automatic arc control gives uniformly good weld quality and finish and the high heat input reduces distortion and the number of runs for a given plate thickness is reduced. Twin welding heads still further reduce welding time, and when used, for example, one on each side of a plate being fillet welded, distortion is reduced. The welding head may be:

1. Fixed, with the work arranged to move beneath it.
2. Mounted on a boom and column which can either be of the positioning type in which the work moves or the boom can traverse at welding speed over the fixed work.
3. Gantry mounted so that it can traverse over the stationary work.
4. Self propelled on a motor-driven carriage.

The processes which have been described previously in this book, namely TIG, MIG and CO_2 (gas shielded metal arc) with their modifications, are extensively used fully automatically. Heads are now available which, by changing simple components, enable one item of equipment to be used for MIG (inert gas), CO_2 and tubular wire, and submerged arc processes.

Submerged arc welding (sub-arc)*

In this automatic process the arc is struck between bare or flux cored wire and the parent plate, the arc, electrode end and the molten pool are submerged or enveloped in an agglomerated or fused powder which turns into a slag in its lower layers under the heat of the arc and protects the weld from contamination. The wire electrode is fed continuously to the arc by a feed unit of motor-driven rollers which is voltage-controlled in the same way as the wire feed in other automatic processes and ensures an arc of constant length. The flux is fed from a hopper fixed to the welding head, and a tube from the hopper spreads the flux in a continuous mound in front of the arc along the line of weld and of sufficient depth to completely submerge the arc, so that there is no spatter, the weld is shielded from the atmosphere and there are no radiation effects (UV and IR) in the vicinity. (Fig. 5.1*a* and *b*.) A latest development to increase deposition rate is to lay down a powder in front of flux feed and welding head; this is generally used with a tandem a.c./d.c. head.

---

* American designation SAW (similar to British).

Welding heads

Fully automated welding heads for this process can also be used with modification for gas shielded metal arc welding including CO_2, solid and flux cored, thus greatly increasing the usefulness of the equipment. The head can be stationary and the work moved below it, as for example in the welding of circumferential and longitudinal seams, or the head may be used with positioners or booms or incorporated into custom-built mass production welding units for fabricating such components as brake shoes, axle housings, refrigerator compressor housings, brake vacuum cylinders, etc., and hard surfacing can also be carried out. The unit can also be tractor mounted (cf. Fig. 5.2*c*) and is self-propelled, with a range of speeds of 100 mm to 2.25 m per minute, and arranged to run on guide bars or rails. Oscillating heads can be used for root runs on butt joints to maintain a

Fig. 5.1. (*a*)

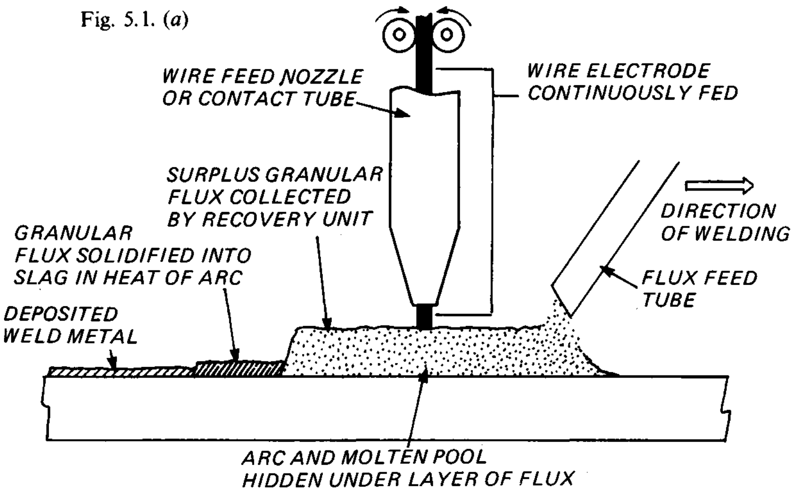
WIRE FEED NOZZLE OR CONTACT TUBE
WIRE ELECTRODE CONTINUOUSLY FED
SURPLUS GRANULAR FLUX COLLECTED BY RECOVERY UNIT
DIRECTION OF WELDING
GRANULAR FLUX SOLIDIFIED INTO SLAG IN HEAT OF ARC
FLUX FEED TUBE
DEPOSITED WELD METAL
ARC AND MOLTEN POOL HIDDEN UNDER LAYER OF FLUX

constant welding bead on the underside of the joint, and two and even three heads can be mounted together or the heads can be arranged side by side to give a wide deposit as in hard surfacing. Fillet welding can be performed by inclining two heads, one on each side of the joint, with flux feeds and recovery, the heads being mounted on a carriage which travels along a gantry over the work (Fig. 5.2). Two heads mounted in tandem and travelling either along a guide rail or directly on the workpiece are used for butt joints on thick plate, and both can operate on d.c. or the leading head can operate on d.c. and the trailing head on a.c.

Three electrode heads can be gantry mounted on a carriage, the leading electrode being d.c. operated with the trailing electrodes a.c. This method gives high deposition rates with deep penetration. Special guide units ensure in all cases that the electrode is correctly positioned relative to the joint.

The main components in the control box are: welding voltage and arc

Fig. 5.1. (*b*)

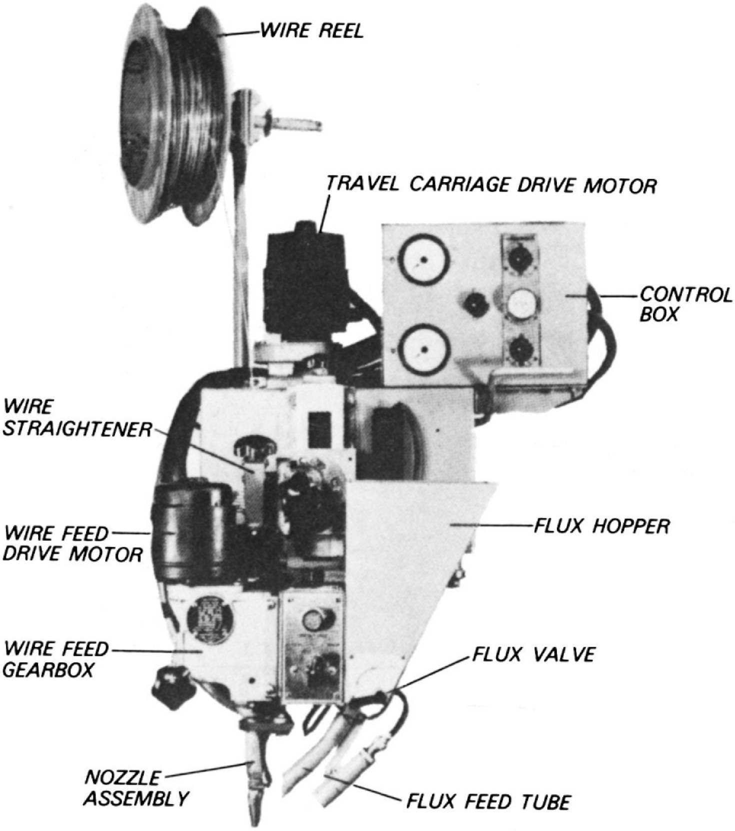
WIRE REEL
TRAVEL CARRIAGE DRIVE MOTOR
CONTROL BOX
WIRE STRAIGHTENER
FLUX HOPPER
WIRE FEED DRIVE MOTOR
WIRE FEED GEARBOX
FLUX VALVE
NOZZLE ASSEMBLY
FLUX FEED TUBE

current controls, and wire feed controlled by a thyristor regulator which maintains set values of arc voltage. The head is accurately positioned by slide adjusters for horizontal and vertical movement and has angular adjustment also. The wire feed motor has an integral gearbox and wire-straightening rolls give smooth wire feed. The gear ratio for the metal arc process is much higher than for submerged arc and each wire diameter usually requires its own feed rolls, which are easily interchanged. For fine wires less than 3 mm diameter a fine-wire-straightening unit can be fitted.

Current is passed to the electrode wire through a contact tube and jaws

Fig. 5.2. Various mountings for automatic welding equipment.
(*a*) An automatic welding machine in which the head is mounted on a carriage which travels along a beam.
(*b*) An automatic welding head as in (*a*) designed for stationary mounting on a manipulator column or boom in order to be an integral part of a mechanized welding system.
(*c*) A tractor-mounted automatic welding head as in (*a*). The machine has a single welding head and is designed for welding butt joints and for making fillet welds in the flat or horizontal–vertical position.

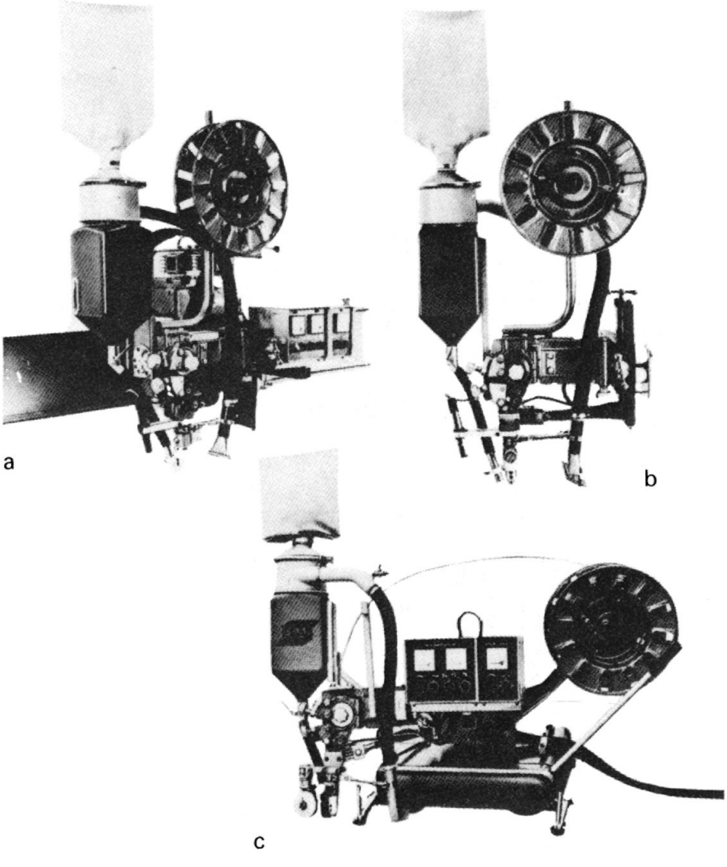

which fit the wire diameter being used, and the contact tube is used for the shielding gas when gas shielded welding is being performed and is water cooled. The coil arm holder has a brake hub with adjustable braking effect and carries 300 mm i.d. coils, and a flux hopper is connected to a flux funnel attached to the contact tube by a flexible hose.

A guide lamp which is attached to the contact tube provides a spot of light which indicates the position of the wire, thus enabling accurate positioning of the head along the joint, and a flux recovery unit collects unfused flux and returns it to the hopper.

A typical sequence of operations for a boom-mounted carriage carrying multiple welding head is: power on, carriage positioned, welding heads 1, 2, etc., down, electrode feed on, wire tips set, flux valve open, welding speed set, welding current set, welding voltage set, flux recovery on; press switch to commence welding.

Power unit

The power unit can be a motor- or engine-driven d.c. generator or transformer-rectifier with outputs in the 30–55 V range and with currents from 200 to 1600 A with the wire generally positive. In the case of multiple head units in which the leading electrode is d.c. and the trailing electrode is a.c. a transformer is also required. In general any power source designed for automatic welding is usually suitable when feeding a single head.

Wires are available in diameters of 1.6, 2.0, 2.4, 3.2, 4.0, 5.0, 6.0 mm on plastic reels or steel formers. Wrappings should be kept on the wire until ready for use and the reel should not be exposed to damp or dirty conditions.

A variety of wires are available including the following (they are usually copper coated). For mild steel types with varying manganese content, e.g. 0.5%, 1.0%, 1.5%, 2.0% manganese, a typical analysis being 0.1% C, 1.0% Mn, 0.25% Si, with S and P below 0.03%. For low alloy steels, 1,25% Cr, 0.5% Mo; 2.25% Cr, 1.0% Mo; 0.2% Mn, 0.5% Mo; and 1.5% Mn 0.5% Mo are examples, whilst for the stainless steel range there are: 20% Cr, 10% Ni for unstabilized steels; 20% Cr, 10% Ni, 0.03% C, for low-carbon 18/8 steels; 19% Cr, 12% Ni, 3% Mo for similar steels; 20% Cr, 10% Ni, nobium stabilized; 24% Cr, 13% Ni for steels of similar composition and for welding mild and low-alloy steels to stainless steel.

Many factors affect the quality of the deposited weld metal: electrode wire, slag basicity, welding variables (process), cleanliness, cooling rate, etc. For hardfacing, the alloy additions necessary to give the hard surface usually come from the welding wire and a neutral flux, and tubular wire with internal flux core is also used in conjunction with the external flux.

Hardness values as welded, using three layers on a mild steel base, are between 230 and 650 HV depending upon the wire and flux chosen.

Flux (see also Vol. 1, pp. 61–3)

Fluxes are suitable for use with d.c. or a.c. They are graded according to their form, whether (1) fused or (2) agglomerated. Fused fluxes have solid glassy particles, low tendency to form dust, good recycling properties, good slag-flux compatibility, low combined water and little sensitivity to humid conditions. Agglomerated fluxes have irregular-sized grains with low bulk density, low weight consumption at high energy inputs with active deoxidizers and added alloying elements where required.

Fluxes are further classified as to whether they are acidic or basic, the basicity being the ratio of basic oxides to acidic oxides which they contain. In general the higher the basicity the greater the absorption of moisture and the more difficult it is to remove.*

The general types of flux include manganese silicate, calcium manganese aluminium sulphate, rutile, zirconia and bauxite, and the choice of flux affects the mechanical properties of the weld metal. Manufacturers supply full details of the chemical composition and mechanical properties of the deposited metal when using wires of varying compositions with various selected fluxes (i.e. UTS, % elongation Charpy impact value, and CTOD† figures at various temperatures).

Fluxes that have absorbed moisture should be dried in accordance with the makers' instructions, as the presence of moisture will affect the mechanical properties of the deposited metal. The flux, whether fused or agglomerated, should be chosen to give the weld as near as possible the same characteristics as the parent plate. Wire composition and chosen flux must be compatible. Tensile and yield strength, together with impact values of the welded joint must all be considered.

The strength of carbon–manganese steels depends upon the carbon, silicon and manganese content. Of these, the manganese content can most easily be varied by additions to the flux to obtain the various levels of strength. In general, the manganese content of the weld metal should be equal to or exceed that of the parent plate. Too much manganese will result in brittleness. The impact value is governed by the basicity of the flux; an increase of basicity results in a decrease of arc stability, a reduction in weld appearance and more difficult slag removal. In general the flux of highest basicity consistent with stable arc, good weld appearance and easy slag removal should be chosen.

---

* Basicity (BI) $= \dfrac{\mathrm{CaO+CaF_2+MgO+K_2O+Na_2O+\frac{1}{2}(MnO+FeO)}}{\mathrm{SiO_2+\frac{1}{2}(Al_2O_3+TiO_2+ZrO_2)}}$. The basicity may vary from 0.7 to 3.1 (see also Vol. 1, pp. 61–4.

† Crack tip opening displacement Vol. 1, pp. 284–9.

Electrode wires, wound on reels, are available, together with compatible fluxes for carbon–manganese steels, low-alloy steels and stainless steels. For duplex stainless steels use a flux such as OK 16.93 or 18.86, and for super duplex use flux OK 16.88. See table, p. 181, for filler wire and fluxes.

Joint preparation

Joint edges should be carefully prepared and free from scale, paint, rust and oil, etc., and butt seams should fit tightly together. If the fit-up has gaps greater than 0.8 mm these should be sealed with a fast manual weld.

When welding curved circumferential seams there is a tendency for the molten metal and slag, which is very fluid, to run off the seam. This can be avoided or reduced by having the welding point 15–65 mm before top dead centre in the opposite direction to the rotation of the work and in some cases the speed of welding and current can be reduced. Preparation of joints is dependent upon the service to which the joint is to be put and the following preparations are given as examples only (Fig. 5.3).

Backing

As the cost of back-chipping and making a sealing run has escalated it becomes more and more necessary to be able to weld plates and cylinders of large size with a run on one side only. This may be the case, for example, if the fit-up of the sections is poor and a weld in the root of the section may not be able to bridge a wide root gap successfully. In these cases a backing can be used so that the weld is performed from one side only and with which a good profile of underbead is obtained even when fit-up and alignment are not good. The following are examples of differing types of backing strips available.

Ceramic tile backing strip. This is shown in section in Fig. 5.4*a* and is suitable for slag-forming processes such as submerged arc, or flux cored and MMA can be used for vertical and horizontal–vertical butts. A recess in the tile allows the slag to form below the underbead and is stripped off and discarded with the aluminium foil which holds it in position as the weld progresses.

Fibreglass tape backing strip is a closely woven flexible material of about four to six layers and fibreglass tape which gives good support to the underbead or root run of the weld and is usually used in conjunction with a copper or aluminium backing bar. It is non-hygroscopic and has low fume level. Sizes are from 30 mm wide heavy single layer, 35 mm wide four layer and 65 mm wide six layer.

Fibreglass tape, sintered sand backing plate. This is typical for submerged arc single or twin wire as in large structures, e.g. deck plates in shipyards, etc. The backing is of sintered sand (silica) about 600 mm long, 50 mm wide and 10 mm thick reinforced with steel wires. It has a fibreglass tape fitted to the upper surface to support the root of the weld and has adhesive outer edges to allow for attachment to the joint, which should be dry and which should have a 40–50° included angle preparation and a root gap of about 4 mm. The backing is slightly flexible to allow for errors of alignment and has 45° bevelled ends. Overlapping tape prevents burn-through at the

Fig. 5.3. Types of butt welds.

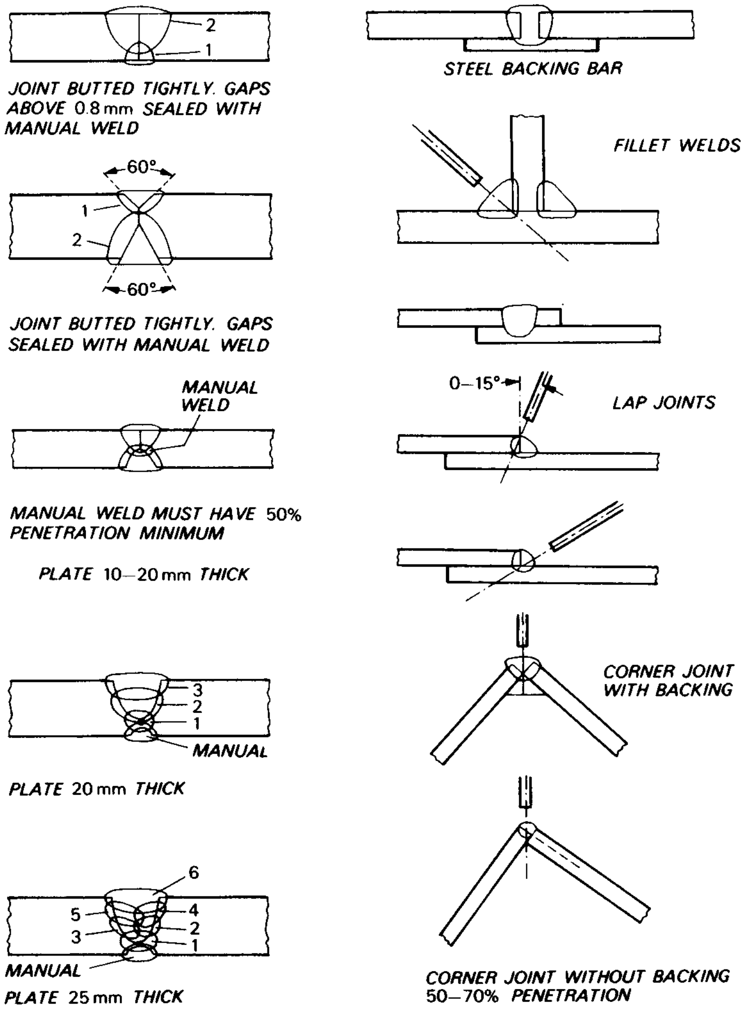
2
1
JOINT BUTTED TIGHTLY. GAPS ABOVE 0.8 mm SEALED WITH MANUAL WELD
60°
1
2
60°
JOINT BUTTED TIGHTLY. GAPS SEALED WITH MANUAL WELD
MANUAL WELD
MANUAL WELD MUST HAVE 50% PENETRATION MINIMUM
PLATE 10–20 mm THICK
3
2
1
MANUAL
PLATE 20 mm THICK
6
5
4
3
2
1
MANUAL
PLATE 25 mm THICK
STEEL BACKING BAR
FILLET WELDS
0–15°
LAP JOINTS
CORNER JOINT WITH BACKING
CORNER JOINT WITHOUT BACKING 50–70% PENETRATION

backing junction. An aluminium section can be used as undersupport if required (Fig. 5.4*b*).

Electroslag welding (Fig. 5.5)

Developed in Russia, this process is used for butt welding steel sections usually above 60 mm in thickness although plates down to 10 mm thick can be welded. The sections to be welded are fixed in the vertical position and part of the joint line, where welding is to commence, is enclosed with water-cooled copper plates or dams which serve to position the molten weld metal and slag. The dams are pressed tightly against each side of the joint to prevent leakage. There may be from one to three electrode wires depending upon the thickness of the section and they are fed continuously from spools. The self-adjusting arc is struck on to a run-off plate beneath a coating of powder flux which is converted into a liquid in about half a minute. The current is then transferred, not as an arc but through the liquid slag, which gives the same order of voltage drop as would the arc. During welding some slag is lost in forming a skin between the molten metal and the copper dams, and a flow of powder, carefully metered to avoid disturbing the welding conditions, is fed in to make good wastage. The vertical traverse may be obtained by mounting the welding

Fig. 5.4. (*a*) Ceramic tile backing strip.

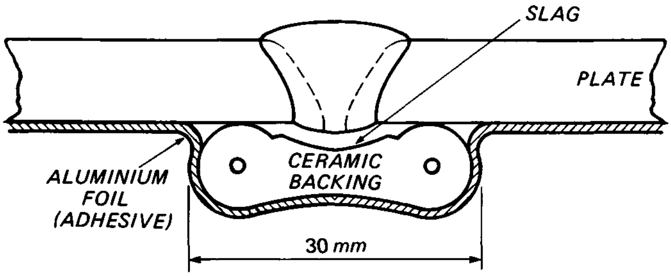
SLAG
PLATE
ALUMINIUM FOIL (ADHESIVE)
CERAMIC BACKING
30 mm

Fig. 5.4. (*b*)

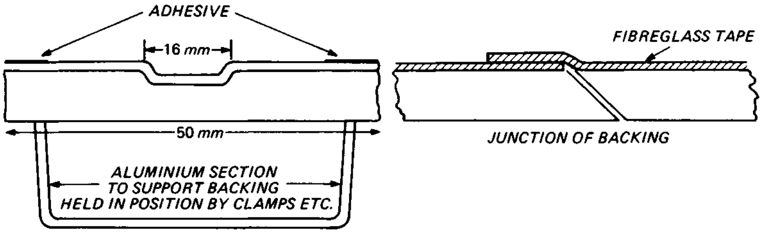
ADHESIVE
16 mm
50 mm
ALUMINIUM SECTION TO SUPPORT BACKING HELD IN POSITION BY CLAMPS ETC.
FIBREGLASS TAPE
JUNCTION OF BACKING

head on a carriage which is motor-driven and travels up a rack on a vertical column in alignment with the joint to be welded. The rate of travel is controlled so that the electrode nozzle and copper dams are in the correct position with regard to the molten pool, and since the electrode is at right angles to the pool, variations in fit-up are not troublesome. For thick sections the electrodes are given a weaving motion across the metal.

The welds produced are free from slag inclusions, porosity and cracks, and the process is rapid, preparation is reduced, and there is no deslagging. Composite wires containing deoxidizers and alloying elements can be used when required. A variation of the process uses a CO_2 shield instead of the flux powder, the CO_2 being introduced through pipes in the copper dams just above the molten metal level.

Electroslag welding with consumable guides or nozzles

Consumable guide welding is a development of the electroslag process for welding straight joints in thick plate in the vertical or near vertical position, in a range of 15–40 mm thick plate and with joints up to 2 m long. The set-up gap between plates is 25–30 mm, but when welding

Fig. 5.5. Principle of electroslag or vertical submerged melt welding.

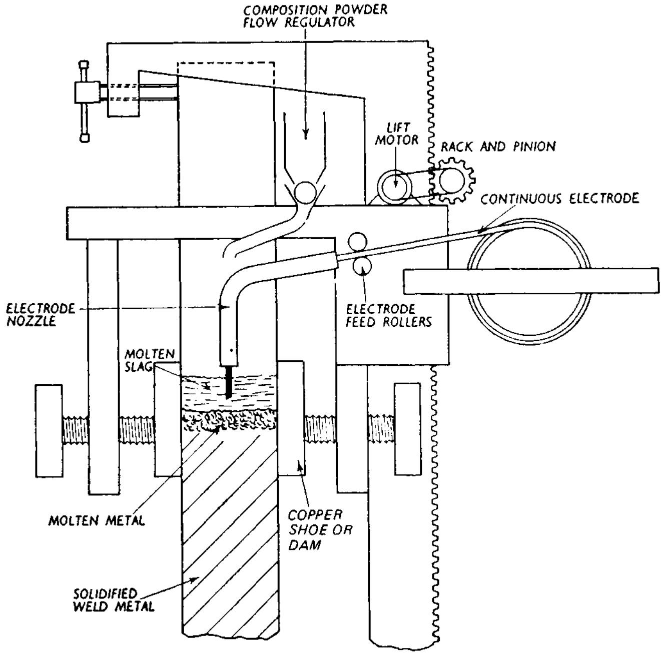
COMPOSITION POWDER FLOW REGULATOR
LIFT MOTOR
RACK AND PINION
CONTINUOUS ELECTRODE
ELECTRODE NOZZLE
ELECTRODE FEED ROLLERS
MOLTEN SLAG
MOLTEN METAL
COPPER SHOE OR DAM
SOLIDIFIED WELD METAL

thicknesses less than 20 mm the joints can be reduced to 18–24 mm, the gap ensuring that the guide tube does not touch the plate edges. Water-cooled copper shoes act as dams and position the molten metal and also give it the required weld profile. As with electroslag welding the current passing through the molten slag generates enough heat to melt the electrode end, guide and edges of the joint, ensuring a good fusion weld. If a plain uncoated guide tube is used, flux is added to cover the electrode and guide end before welding commences. To start the process the arc is struck on the work. It continues burning under the slag with no visible arc or spatter. The slag should be viewed through dark glasses as in gas cutting because of its brightness when molten.

An a.c. or d.c. power source in the range 300–750 A is suitable, such as is used for automatic and MMA processes. Striking voltage is of the order of 70–80 V, with arc voltages of 30–50 V, higher with a.c. than d.c.

The advantages claimed for the process are: relatively simpler, cheaper, and more adaptable than other similar types, faster welding than MMA of thick plate, cheaper joint preparation, even heat input into the joint thus reducing distortion problems, no spatter losses, freedom from weld metal defects and low consumption of flux. Fig. 5.6 illustrates the layout of the machine.

Mechanized MIG welding with robots (robotics)

Fully automatic welding using the gas shielded metal arc or submerged arc processes in conjunction with columns and booms, positioners, rotators and other equipment such as jigs is extensively used for the making of welds either straight, circumferential or circular. The use of robots in the car industry for spot welding the body shell, spray painting and general assembly is well known and it has resulted in increased speed of production and less overall cost with reduced monotony and fatigue for the operator.

Now consider the robot adapted to the fabrication of large numbers of similar components with welds in all positions performed by the gas shielded metal arc process.

The word robot comes from the Czech word *robota* which, in translation, means any class of work that involves monotony, repetition or drudgery. When operating the robot the welder becomes the supervisor for whom the robot works, performing this monotonous work at speed, with accuracy and not getting tired or needing breaks.

To serve as an example we can consider one particular modern robot illustrated in Fig. 5.7.

This has six degrees of freedom, (one being optional) based on the human hand, wrist and forearm and the following gives the approximate movement values:

> rotational 340°, radial arm 550 mm max., vertical arm 850 mm max., wrist bending 90°, wrist turning 180°.

The accuracy of these movements is 0.1–0.2 mm at 500 mm distance and it should be noticed that the robot can be equipped with heads for selection, grinding, polishing and spot welding, if required.

Evidently there must be large production runs of components with repetitive welding to be performed for the cost of the station, consisting of positioners, control cabinets, robot with gas shielded metal arc welding,

Fig. 5.6. (*a*) Consumable guide layout showing water-cooled dams.

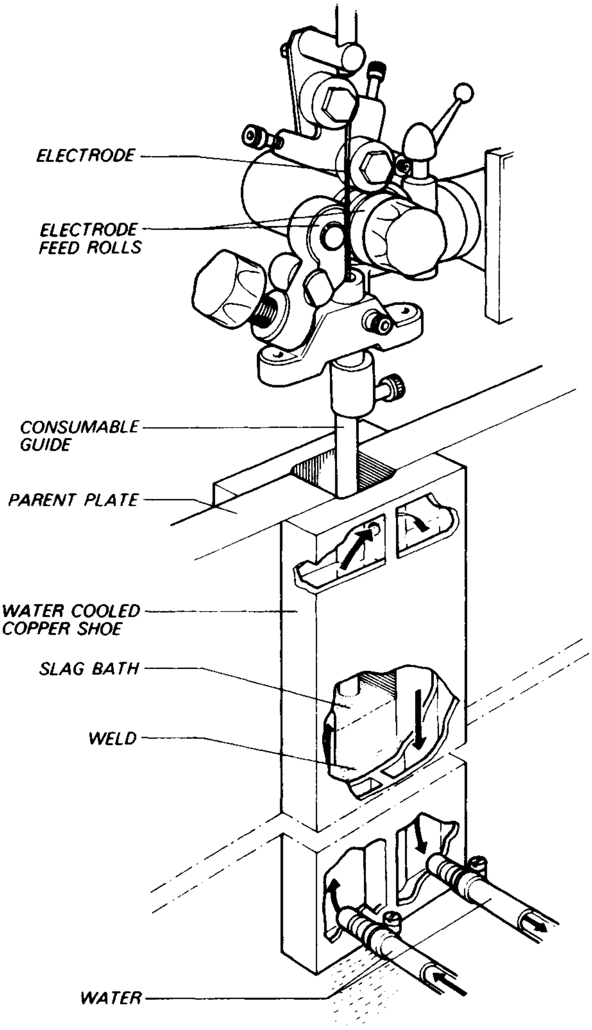
ELECTRODE
ELECTRODE FEED ROLLS
CONSUMABLE GUIDE
PARENT PLATE
WATER COOLED COPPER SHOE
SLAG BATH
WELD
WATER

Fig. 5.6. (*b*) Completed weld.

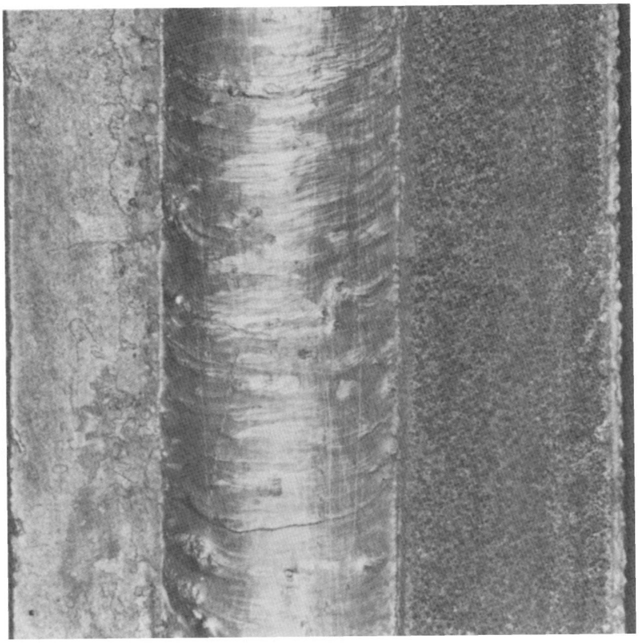

Fig. 5.6. (*c*) The consumable guide welding process using, in this case, twin wire/tube system. The dams have been removed to show the position of the guide tubes.

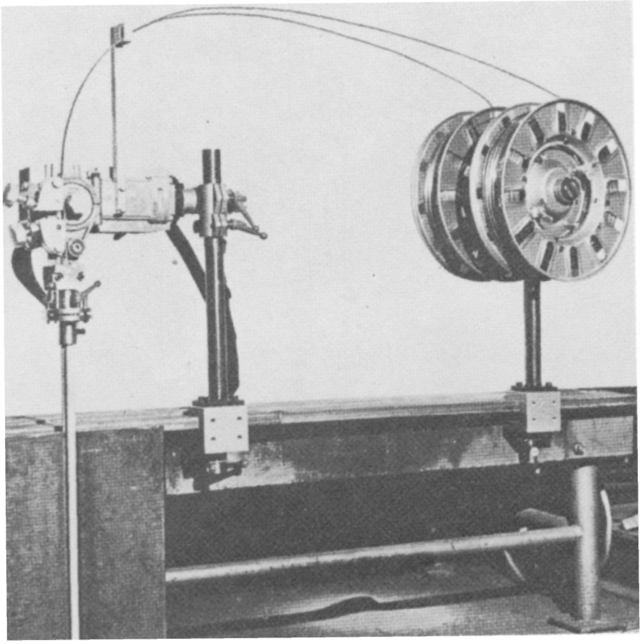

Fig. 5.7

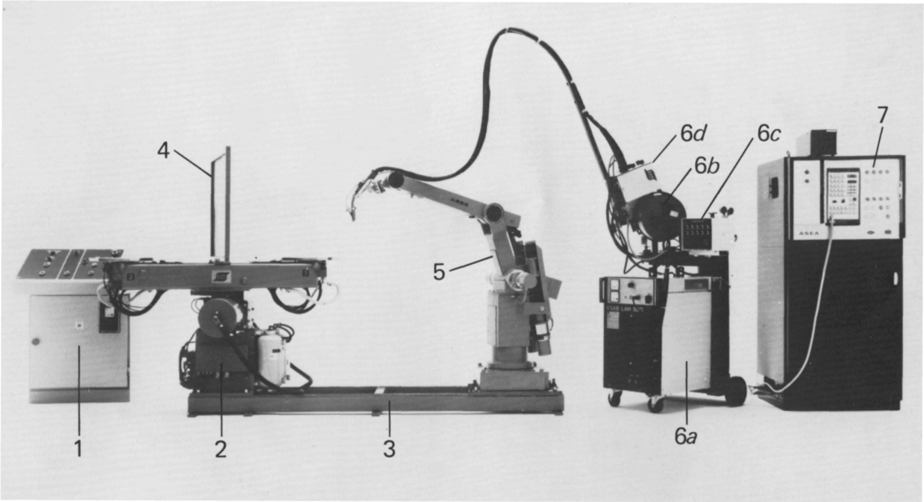

Key:

1. Control panel, by which operator determines when robot should start welding the next work piece.
2. Handling unit. Different types are available. On this one fixtures for the component are mounted on tilting turntables.
3. Handling unit on same foundation as robot giving stability and accuracy.
4. Screen.
5. Welding robot with rapid, accurate movements.

6*a*. Constant voltage power source for welding head. Wire feed unit can be positioned separately.
6*b*. Wire reel.
6*c*. Programming unit.
6*d*. Wire feed.
7. Control cabinet, controlling movements and where removable programming unit i located. Program is stored on tape cassette.

head and power and control unit, to be justified. Once the decision has been taken for the outlay to be justified the advantages that accrue are great.

The production station, which varies according to the size of the work pieces, can consist, for example, of two positioners, on which the work is jig held. These positioners (Fig. 5.8) and the data required in the welding operations are controlled by a microcomputer and are servo-steered.

The robot head has a gas shielded metal arc welding gun adapted to fit the head, the power source is of the thyristor-controlled constant voltage type and the wire is fed from a unit which controls speed of feed and compensates for variations of mains voltage and friction of the feed rollers.

To program the robot, the programming unit is taken from the control cabinet and the robot run through the complete welding sequence for the part to be fabricated. The welding gun is moved from point to point and each section is fed with the speed required and the welding parameters (current and voltage, etc.), and the accelerated movement from one welding point to the next is also programmed and mistakes are easily corrected. The computer memory has say 500 position capacity with additional instructions with a tape recorder increasing the memory. The storage of the program is on a digital tape cassette so that the switch from one program to another can be made without starting from scratch. As is usual pre- and post-flow of the shielding gases and crater filling are all part of the program. (*Note.* The TIG method of welding may also be performed by robots. See Chapter 3.)

Fig. 5.8. Computer control, servo-powered positioner. Dual axes, fast and accurate handling, Self-braking worm gears.

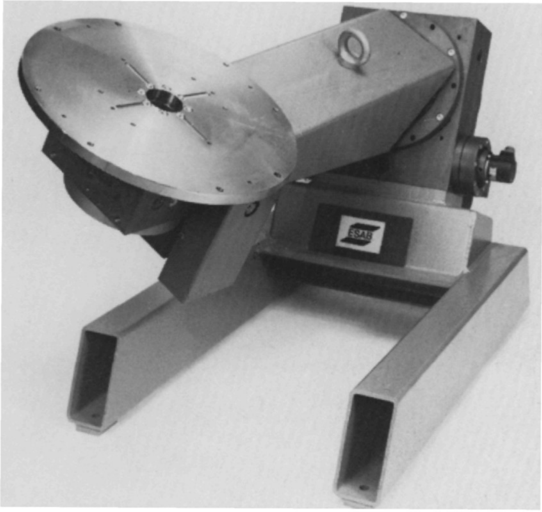

Pressure welding

This is the joining together of metals in the plastic condition (not fusion) by the application of heat and pressure as typified by the blacksmith's weld. In general the process is confined to butt welding. The parts (tubes are a typical example) are placed on a jig which can apply pressure to force the parts together.

The faces to be welded are heated by oxy-acetylene flames, and when the temperature is high enough for easy plastic flow to take place, heating ceases and the tubes are pushed together causing an upset at the welded face. The welding temperature is about 1200 °C for steel. It is considered that atoms diffuse across the interfaces and recrystallization takes place, the grains growing from one side to the other of the welded faces since they are in close contact due to the applied pressure. Any oxide is completely broken up at a temperature well below that of fusion welding but due to the heating time concerned, grain growth is often considerable. Steel, some alloy steels, copper, brass, and silver can be welded by this process.

Cold pressure welding

Cold pressure welding is a method of joining sections of metal together by the application of pressure alone using no heat or flux. Pressure

Fig. 5.9

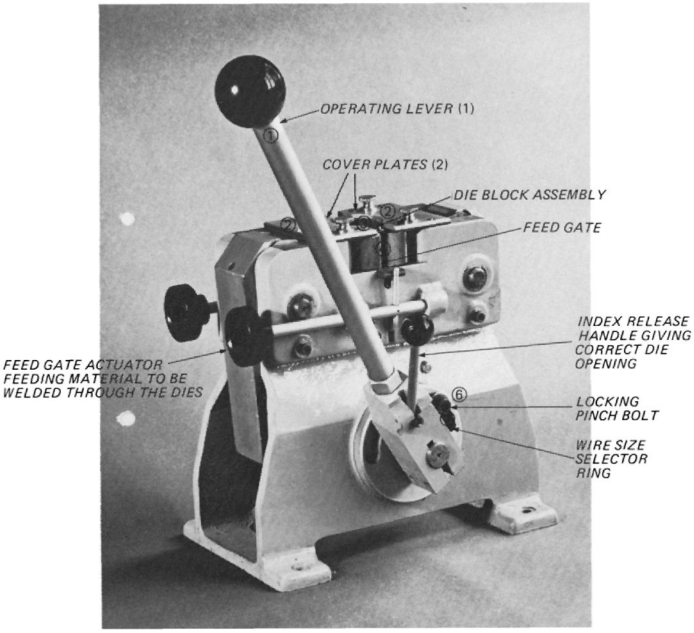
OPERATING LEVER (1)
①
COVER PLATES (2)
②
DIE BLOCK ASSEMBLY
FEED GATE
INDEX RELEASE HANDLE GIVING CORRECT DIE OPENING
FEED GATE ACTUATOR FEEDING MATERIAL TO BE WELDED THROUGH THE DIES
⑥
LOCKING PINCH BOLT
WIRE SIZE SELECTOR RING

Fig. 5.10. (*a*) Cold pressure weld, aluminium to copper (wire 9.5 mm diameter). × 25.

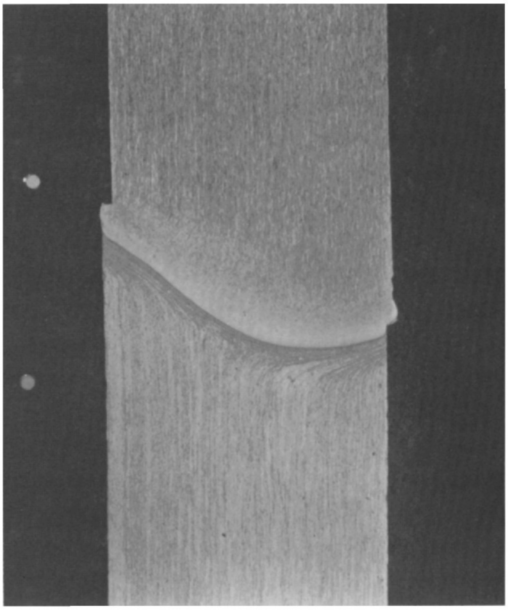

Fig. 5.10. (*b*) Cold pressure weld, copper to copper (wire 8 mm diameter). × 25.

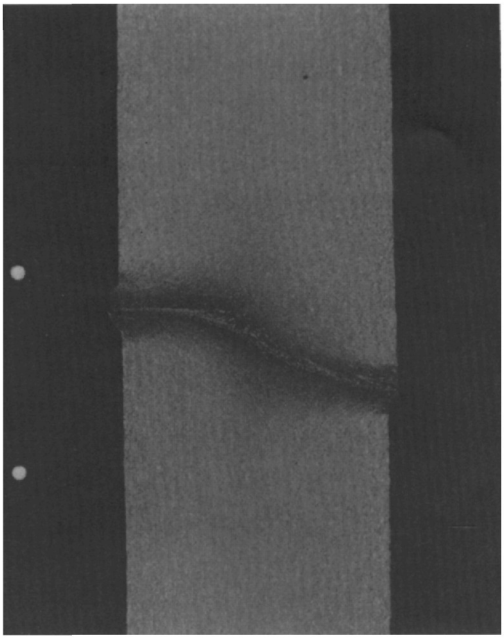

is applied to the points to be welded at temperatures below the recrystallization temperature of the metals involved. This applied pressure brings the atoms on the interface to be welded into such close contact that they diffuse across the interface and a cold pressure weld is made.

It is a method for relatively ductile metals such as aluminium, copper cupro-nickel, gold, silver, platinum, lead, tin and lead–tin alloys, etc., and it is particularly suited to welds in circular wire section. In the type described of the multiple upset type, the surfaces to be welded are placed in contact and held in position by gripping dies and are fed together in small increments by a lever (Fig. 5.9). Each lever movement giving interface pressure displaces the original surfaces by plastic flow and after about four to six upsets, the last movement completes the weld and the flash is easily removed. In the machine illustrated, copper wire 1.1–3.5 mm diameter max. or aluminium wire to 4.75 max. can be butt welded. Fig. 5.10*a* and *b* illustrates the micrographs of welds made in copper to copper and copper to aluminium wire.

Ultrasonic welding

Ultrasonic vibrations of several megacycles per second (the limit of audibility is 20 000–30 000 Hz) are applied to the region of the faces to be welded. These vibrations help to break up the grease and oxide film and heat the interface region. Deformation then occurs with the result that welding is possible with very greatly reduced pressure compared with an ordinary pressure weld. Very thin section and dissimilar metals can be welded and because of the reduced pressure there is reduced deformation (Fig. 5.11).

Fig. 5.11. Ultrasonic welding.

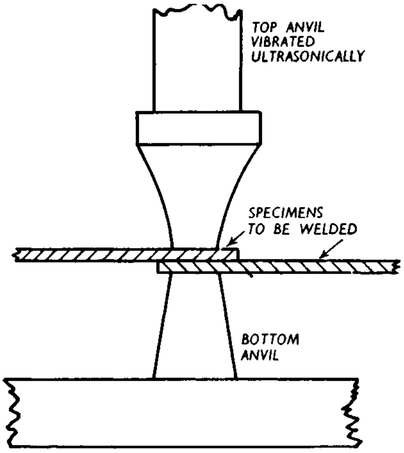

Friction welding

The principle of operation of this process is the changing of mechanical energy into heat energy. One component is gripped and rotated about its axis while the other component to be welded to it is gripped and does not rotate but can be moved axially to make contact with the rotating component. When contact is made between rotating and non-rotating parts heat is developed at the contact faces due to friction and the applied pressure ensures that the temperature rises to that required for welding. Rotation is then stopped and forging pressure applied, causing more extrusion at the joint area, forcing out surface oxides and impurities in the form of a flash (Fig. 5.12). The heat is concentrated and localized at the interface, grain structure is refined by hot work and there is little diffusion across the interface so that welding of dissimilar metals is possible.

In general at least one component must be circular in shape, the ideal situation being equal diameter tubes, and equal heating must take place over the whole contact area. If there is an angular relationship between the final parts the process is not yet suitable.

The parameters involved are: (1) the power required, (2) the peripheral speed of the rotating component, (3) the pressure applied and (4) the time of duration of the operation. By adjusting (1), (2) and (3), the time can be reduced to the lowest possible value consistent with a good weld.

Power required. When the interfaces are first brought into contact, maximum power is required, breaking up the surface film. The power required then falls and remains nearly constant while the joint is raised to welding temperature. The power required for a given machine can be chosen so that the peak power falls within the overload capacity of the driving motor. It is the contact areas which determine the capacity of a machine. The rotational speed can be as low as 1 metre per second peripheral and the pressure depends upon the materials being welded, for example for mild steel it can be of the order of 50 N/mm^2 for the first part of the cycle followed by 140 N/mm^2 for the forging operation. Non-ferrous metals require a somewhat greater difference between the two operations. The faster the rotation of the component and the greater the pressure, the shorter the weld cycle, but some materials suffer from hot cracking if the cycle is too short and the time is increased with lower pressure to increase the width of the HAZ. At the present time most steels can be welded including stainless, but excluding free cutting. Non-ferrous metals are also weldable and aluminium (99.7% Al) can be welded to steel.

Fig. 5.12. (*a*) Friction welding.

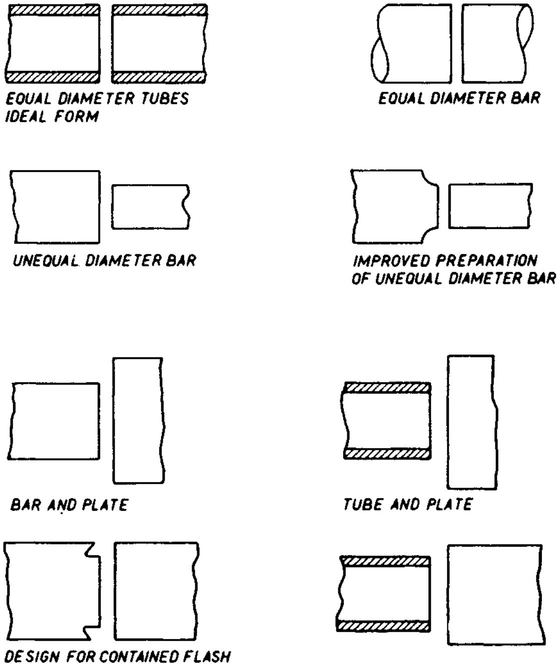
EQUAL DIAMETER TUBES
IDEAL FORM
EQUAL DIAMETER BAR
UNEQUAL DIAMETER BAR
IMPROVED PREPARATION
OF UNEQUAL DIAMETER BAR
BAR AND PLATE
TUBE AND PLATE
DESIGN FOR CONTAINED FLASH

Fig. 5.12. (*b*) Suitable forms of friction welding.

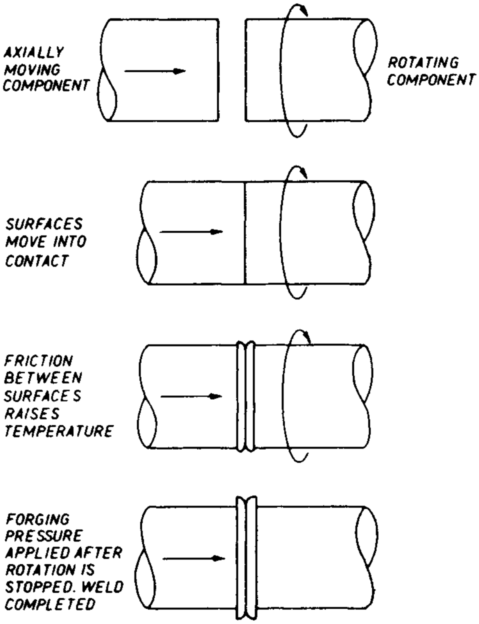
AXIALLY
MOVING
COMPONENT
ROTATING
COMPONENT
SURFACES
MOVE INTO
CONTACT
FRICTION
BETWEEN
SURFACES
RAISES
TEMPERATURE
FORGING
PRESSURE
APPLIED AFTER
ROTATION IS
STOPPED. WELD
COMPLETED

Fig. 5.13. Friction welder.

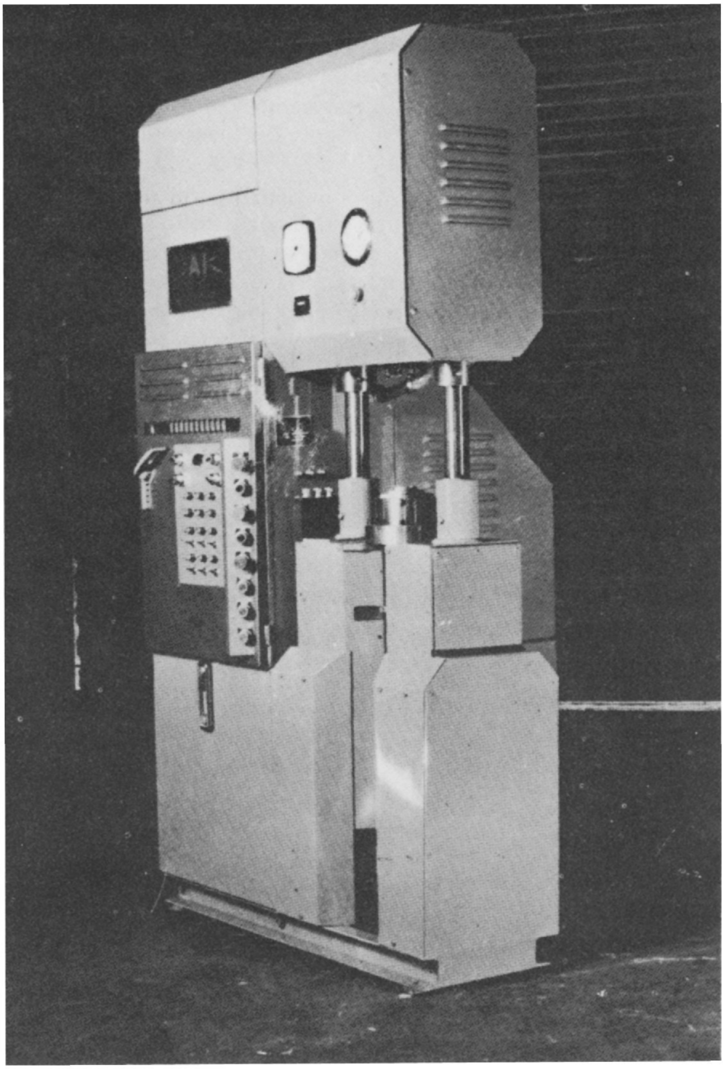

There are various control systems: (1) Time control, in which after a given set time period after contact of the faces, rotation is stopped and forge pressure is applied. There is no control of length with this method. (2) Burn-off to length: parts contact and heating and forging take place within a given pre-determined length through which the axially moving component moves. (3) Burn-off control: a pre-determined shortening of the component is measured off by the control system when minimum pressures are reached. Weld quality and amount of extrusion are thus controlled, but not the length. (4) Forging to length: the axially moving work holder moves up to a stop during the forging operation irrespective of the state of the weld and generally in this case extrusion tends to be excessive.

The extrusion of flash can be removed by a subsequent operation or, for example for tubes of equal diameter, a shearing unit can be built into the machine operating immediately after forging and while the component is hot, thus requiring much less power. Fig. 5.12(*a*) illustrates a joint designed to contain the flash.

In the process known as inertia welding the rotating component is held in a fixture attached to a flywheel which is accelerated to a given speed and then uncoupled from its drive. The parts are brought together under high thrust and the advantage claimed is that there is no possibility of the driving unit stalling before the flywheel energy is dissipated.

Friction welding machines resemble machine tools in appearance, as illustrated in Fig. 5.13.

Electron beam welding

If a filament of tungsten or tantalum is heated to high temperature in a vacuum either directly by means of an electric current or indirectly by means of an adjacent heater, a great number of electrons are given off from the filament, which slowly evaporates. This emission has been mentioned previously in the study of the tungsten arc welding process. The greater the filament current the higher the temperature and the greater the electron emission, and if a metal disc with a central hole is placed near the filament and charged to a high positive potential relative to the filament, so that the filament is the cathode and the disc the anode, the emitted electrons are attracted to the disc and because of their kinetic energy pass through the hole as a divergent beam. This can then be focused electrostatically, or magnetically, by means of coils situated adjacent to the beam and through which a current is passed. The beam is now convergent and can be spot focused. The basic arrangement, an electron 'gun', is similar to that used for television tubes and electron microscopes (Fig. 5.14).

If the beam is focused on to a metal surface the beam can have sufficient energy to raise the temperature to melting point, the heating effect depending upon the kinetic energy of the electrons. The kinetic energy of an electron is $\frac{1}{2}mV^2$, where m is the mass of an electron (9.1×10^{-28} g) and V its velocity. The electron mass is small, but increasing the emission from the filament by raising the filament current increases the number of electrons and hence the mass effect. Because the kinetic energy varies directly as the square of the velocity, accelerating the electrons up to velocities comparable with the velocity of light by using anode voltages (up to 200 kV) greatly increases the beam energy. The smaller the spot into which the beam is focused the greater the energy density but final spot size is often decided by working conditions, by aberration in the focusing system, etc., so that spot size may be of the order of 0.25–2.5 mm.

When the beam strikes a metal surface X-rays are generated, so that adequate precautions must be taken for screening personnel from the rays by using lead or other metal screens or making the metal walls of the gun chamber sufficiently thick. If the beam emerges into the atmosphere the energy is reduced by collision of the electrons with atmospheric molecules and focus is impaired. Because of this it has been the practice to perform many welding operations in a vacuum, either in the gun chamber in which case each time the component is loaded the chamber must be evacuated to high vacuum conditions, thus increasing the time and cost of the operation, or in a separate steel component chamber fixed to the gun chamber. This can be made of a size to suit the component being welded and is evacuated

Fig. 5.14. Electron beam welding.

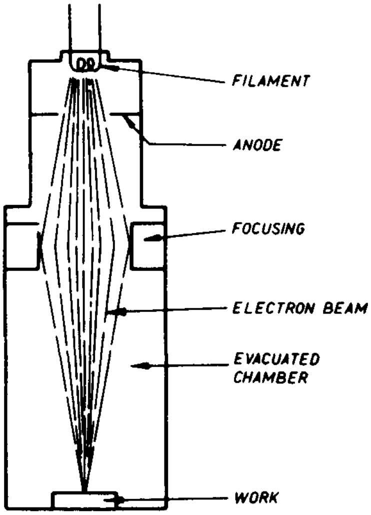

to a relatively low vacuum after each loading. In either case welds suffer no contamination because of vacuum conditions. Viewing of the spot for set-up, focusing and welding is done by various optical arrangements.

Welding in non-vacuum conditions requires much greater power than for the preceding method because of the effects of the atmosphere on the beam and the greater distance from gun to work, and a shielding gas may be required around the weld area. Research work is proceeding in this field involving guns of higher power consumption. Difficulties may also be encountered in focusing the beam if there is a variation in the gun-to-work distance, as on a weld on a component of irregular shape.

Welds made with this process on thicker sections are narrow with deep penetration with minimum thermal disturbance and at present welds are performed in titanium, niobium, tungsten, tantalum, beryllium, nickel alloys (e.g. nimonic), inconel, aluminium alloys and magnesium, mostly in the aero and space research industries. The advantages of the process are that being performed in a vacuum there is no atmospheric contamination and the electrons do not affect the weld properties, accurate control over welding conditions is possible by control of electron emission and beam focus, and there is low thermal disturbance in areas adjacent to the weld (Fig. 5.15). Because of the vacuum conditions it is possible to weld the more reactive metals successfully. On the other hand the equipment is very costly, production of vacuum conditions is necessary in many cases and there must be protection against radiation hazards.

Laser beam welding*

Radio waves, visible light, ultra-violet and infra-red radiations are electro-magnetic radiations which have two component fields, one electric and the other magnetic, vibrating at right angles to each other and to the direction of propagation of the beam. If either one of these components is suppressed the resulting radiations are said to be polarized and the direction in which the resultant electric or magnetic forces act is the plane of polarization. Light from a source such as a tungsten filament electric light consists of several frequencies involving various shades of colour. These waves are not in phase and are of various amplitudes and planes of polarization, and the light is said to be non-coherent. Light of a single wave-length or frequency is termed monochromatic. The wave-length of light is measured in metres or micro-metres, termed microns (μm), visible light being in the range 0.4–0.7 μm. The frequency is related to the wave-length by the expression $V = n\lambda$, where n is the frequency in Hz, λ the wave-length in metres and V the velocity of electro-magnetic radiation,

---

* The velocity of light is 3×10^{10} cm/s (~ 186000 mps).

3×10^8 m/s, so that the frequency range of visible light is in the range $430–750 \times 10^{12}$ Hz.

Atoms of matter can absorb and give out energy and the energy of any atomic system is thereby raised or lowered about a mean or 'ground' level. When an atom or molecule is at an energy level higher than that of the ground state (or level) it is said to be excited, and in this condition two similar atoms can combine to form a molecule called a 'dimer'. The combination is for an extremely short time, and only when in the excited state. Energy can only be absorbed by atoms in definite small amounts (quanta) termed photons, and the relationship between the energy level and the frequency of the photon is $E = h\nu$, where E is the energy level, h is Planck's constant and ν the photon frequency, so that the energy level depends upon the frequency of the photon. An atom can return to a lower energy level by emitting a photon and this takes place in an exceedingly short space of time from when the photon was absorbed, so that if a photon of the correct frequency strikes an atom at a higher energy level, the photon which is released is the same in phase and direction as the incident photon.

The principle of the laser (Light Amplification by Stimulated Emission of Radiation) is the use of this stimulated energy to produce a beam of

Fig. 5.15. Electron beam welding machine with indexing table, tooled for welding distributor shafts to plates at a production rate of 450 per hour. The gun is fitted with optical viewing system. Power 7.5 kW at 60 kV, 125 mA. Vacuum sealing is achieved by seals fitted in the tooling support plate and at the bottom of the work chamber. As the six individual tooling stations reach the welding station they are elevated to the weld position and then rotated by an electronically controlled d.c. motor.

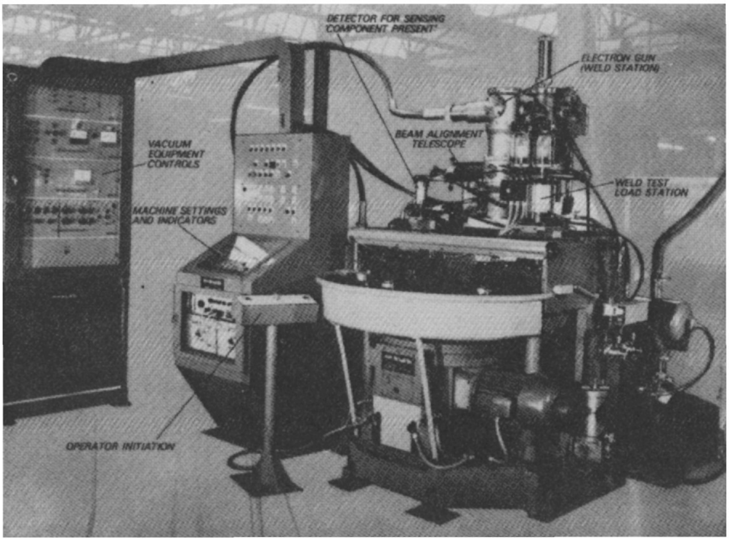
DETECTOR FOR SENSING COMPONENT PRESENT
ELECTRON GUN (WELD STATION)
BEAM ALIGNMENT TELESCOPE
VACUUM EQUIPMENT CONTROLS
WELD TEST LOAD STATION
MACHINE SETTINGS AND INDICATORS
OPERATOR INITIATION

coherent light, that is one which is monochromatic, and the radiation has the same plane of polarization and is in phase. Lasers operate with wavelengths in the visible and infra-red region of the spectrum. When the beam is focused into a small spot and there is sufficient energy, welding, cutting and piercing operations can be performed on metals.

The ruby laser has a cylindrical rod of ruby crystal (Al_2O_3) in which there is a trace of chromium as an impurity. An electronic flash gun, usually containing neon, is used to provide the radiation for stimulation of the atoms. This type of gun can emit intense flashes of light of one or two milliseconds duration and the gun is placed so that the radiations impinge on the crystal. The chromium atoms are stimulated to higher energy levels, returning to lower levels with the emission of photons. The stimulation continues until an 'inversion' point is reached when there are more chromium atoms at the higher levels than at the lower levels, and photons impinging on atoms at the higher energy level cause them to emit photons. The effect builds up until large numbers of photons are travelling along the axis of the crystal, being reflected by the ends of the crystal back along the axis, until they reach an intensity when a coherent pulse of light, the laser beam (of wave-length about 0.06 μm), emerges from the semi-transparent rod end. The emergent pulses may have high energy for a short time period, in which case vaporization may occur when the beam falls on a metal surface, or the beam may have lower energy for a longer time period, in which case melting may occur, while a beam of intermediate power and duration may produce intermediate conditions of melting and vaporization, so that control of the time and energy of the beam and focusing of the spot exercise control over the working conditions.

Developments of the ruby laser include the use of calcium tungstate and glass as the 'host' material with chromium, neodymium, etc., as impurities, a particular example being yttrium–aluminium–garnet with neodymium (YAG), used for operations on small components.

The CO_2 laser uses carbon dioxide for its main gas, with a little helium or nitrogen. The tube may be some metres long, the average length of the beam (10.6 μm) being longer than that of the solid state laser and other continuous wave or pulsed (the power increasing with the length of the tube). To increase the power without increasing the length of the unit, folded-beam lasers have many shorter tubes set parallel to each other; the beam passes down each tube and is reflected from specially ground polished copper mirrors set at the end of each tube. As with other gas lasers, electrical stimulation is by means of an HF discharge from a tube (similar to a fluorescent tube).

A solid state laser (ruby, Nd–YAG), with operational wave-length of 1.06 μm in the infra-red wave band, uses the heating effect of the beam.

Gas lasers use argon fluoride, krypton fluoride and xenon fluoride (and chlorides), giving a pulsed beam of 25–50 W at the ultra-violet end of the spectrum with little or no heat. On starting the tube, electrical stimulation from a radio frequency (r.f.) generator is connected to each end of the tube. The helium atoms, or those of other seeding gases, are stimulated and their energy level raised above ground level. These atoms transfer their energy to the other gas in the tube, until inversion occurs and a coherent beam is emitted. This is known as collision excitation and gas lasers are used for visual effects of various colours.

The excimer laser is of the helium–neon type and stimulation of the gas is to an energy level such that two atoms combine to form a 'dimer'. Energy is obtained from an HF electrical discharge through the tube. When the laser beam strikes a surface it dissociates the bond between the molecules and removes the surface by chemical action. It is used at present for micro-machining metallic and non-metallic surfaces, in rubbers, polymers, papers, glass etc., for surface hardening, marking and cutting thin foils, and in eye operations to alter the contour of the cornea to correct short sight (a laser beam from the ultraviolet part of the spectrum is used).

2 kW CO_2 lasers can be used to weld up to 3 mm thick material and are an alternative to the electron beam for thin gauge material. The width of the weld may be increased at speeds below 12 mm/s due to interaction between the beam and an ionized plasma which occurs near the work. At speeds of 20 mm/s and over, laser and electron beam welds are practically indistinguishable from one another.

Stud welding

This is a rapid, reliable and economical method of fixing studs and fasteners of a variety of shapes and diameters to parent plate. The studs may be of circular or rectangular cross-section, plain or threaded internally or externally and vary from heavy support pins to clips used in component assembly.

There are two main methods of stud welding: (1) arc (drawn arc), (2) capacitor discharge, and the process selected for a given operation depends upon the size, shape and material of the stud and the composition and thickness of the parent plate, the arc method generally being used for heavier studs and plate, and capacitor discharge for lighter gauge work.

Arc (drawn arc) process

This is used in both engineering and construction work and the equipment consists of a d.c. power source, controller and a hand-operated or bench-mounted tool or gun.

A typical unit consists of a forced-draught-cooled power source with a 380/440 V, 3-phase, 50 Hz transformer, the secondary of which is connected to a full-wave, bridge-connected silicon rectifier (Vol. 1, p. 202), with tappings giving ranges of 120–180 A, 170–320 A, and 265–1500 A approximately (equipment is also available up to 1800 A), at about 65 V. The two lower ranges can also be used for MMA welding and two or more units can be connected in parallel if required, the loading being shared automatically.

The power controller contains a main current contactor and timer relays and a pilot arc device, energized from the d.c. source. The solid-state timer has a multi-position switch or switches for the selection of operating time and current for the diameter of stud in use. Work is usually connected to the positive pole with the stud negative except for aluminium, when the polarity is reversed.

The hand gun (Fig. 5.16) has the operating solenoid and return spring within the gun body, which also carries the operating switch and adjustable legs to accommodate varying lengths of studs, an interchangeable chuck for varying stud diameters and a foot adaptor to maintain concentricity between the stud and ferrule or arc shield, this latter being held by ferrule grips. Studs are fluxed on the contact end, which is slightly pointed (Fig. 5.17), and are supplied with ferrules.

To operate the equipment, the welding current and time for the diameter of stud in use are selected, the stud is loaded into the appropriate chuck, the legs adjusted for length and the stud positioned on the plate. A centre punch can be used for locating the stud point and the plate should be free of contamination.

Sequence of operations. The gun switch is pressed, a low current flows between pointed stud end and workpiece and immediately the stud is raised, drawing an arc and ionizing the gap. The main current contactor

Fig. 5.16

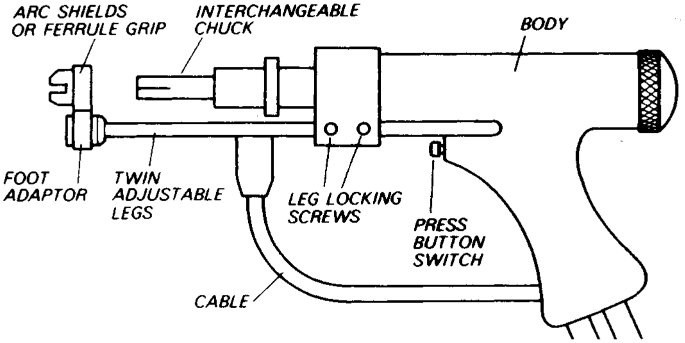

now closes and full welding current flows in an arc, creating a molten state in plate and stud end. The solenoid is de-energized and the stud is pushed under controlled spring pressure into the molten pool in the plate. Finally the main current contactor opens, the current is switched off, and the operation cycle is complete, having taken only a few hundredths of a second (Fig. 5.17).

This method, which is usually employed, whereby the welding current is kept flowing until the return cycle is completed, is termed 'hot plunge'. If the current is cut off just before the stud enters the molten pool it is termed 'cold plunge'. The metal displaced during the movement of the stud into the plate is moulded into fillet form by the ceramic ferrule or arc shield held against the workpiece by the foot. This also protects the operator, retains heat and helps to reduce oxidation.

Studs from 3.3 mm to 20 mm and above in diameter can be used on parent plate thicker than 1.6 mm, and the types include split, U shaped, J bent anchors, etc. in circular and rectangular cross-section for the engineering and construction industries, and can be in mild steel (low carbon), austenitic stainless steel, aluminium and aluminium alloys (3–4% Mg). The rate of welding varies with the type of work, jigging, location, etc., but can be of the order of 8 per minute for the larger diameters and 20 per minute for smaller diameters.

Fig. 5.17. Arc stud welding, cycle of operations.

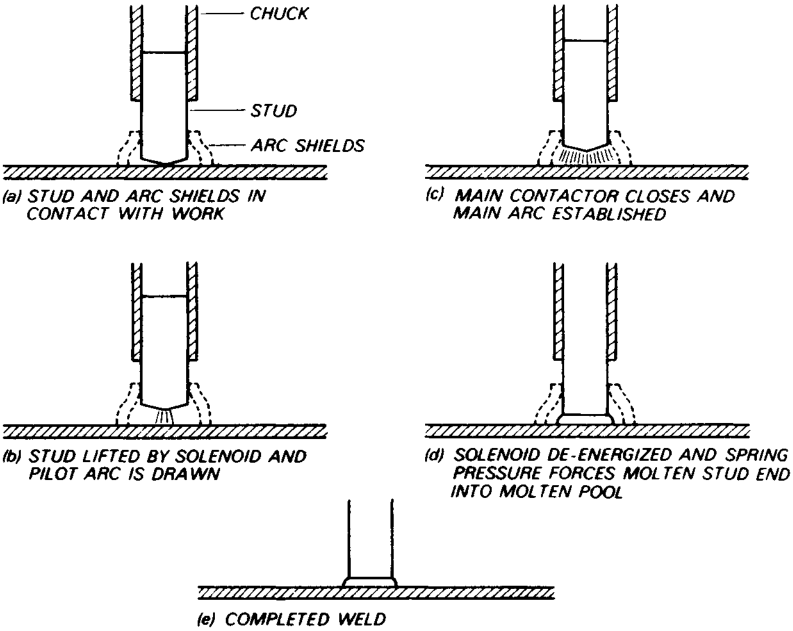

(a) STUD AND ARC SHIELDS IN CONTACT WITH WORK

(c) MAIN CONTACTOR CLOSES AND MAIN ARC ESTABLISHED

(b) STUD LIFTED BY SOLENOID AND PILOT ARC IS DRAWN

(d) SOLENOID DE-ENERGIZED AND SPRING PRESSURE FORCES MOLTEN STUD END INTO MOLTEN POOL

(e) COMPLETED WELD

Capacitor discharge stud welding

In this process a small projection on the end of the stud makes contact with the workpiece and the energy from a bank of charged capacitors (see Volume 1) is discharged across the contact, melting the stud projection, ionizing the zone and producing a molten end on the stud and a shallow molten pool in the parent plate. At this time the stud is pushed into the workpiece under controlled spring pressure, completing the weld (Fig. 5.18).

If C is the capacitance of the capacitor, Q the charge and V the potential difference across the capacitor, then $C = Q/V$ or *capacitance = charge/potential.* The energy of a charged capacitor is $\frac{1}{2}QV$, and since $Q = CV$, the energy is $\frac{1}{2}CV^2$, so that the energy available when a capacitor is discharged is dependent upon: (1) the capacitance – the greater the capacitance the greater the energy; (2) the square of the potential difference (voltage) across the capacitance. The greater this voltage, the greater the energy. Thus the energy required for a given welding operation is obtained by selection of the voltage across the capacitor, i.e. that to which it is charged, and the total capacitance in the circuit.

In the contact method of capacitor discharge welding the small projection on the stud end is placed in contact with the workpiece as explained above. In the hold-off method, useful for thin gauge plate to avoid reverse marking, a holding coil in the hand gun is energized when the welding power is selected in the 'hold-off' position on the capacitor switch. When the stud is pushed into the gun-chuck both stud and chuck move into the hold-off position giving a pre-set clearance between the projection on the stud end and the workpiece. When the gun switch is pressed the hold-off

Fig. 5.18. Capacitor discharge stud welding, sequence of operations.

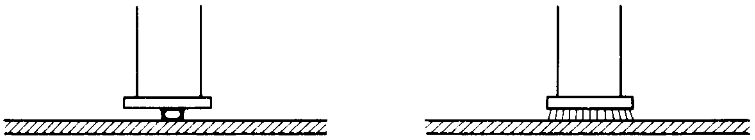

(a) SPIGOT ON STUD CONTACTS WORK

(c) ARC PRODUCES MOLTEN POOL AND MOLTEN END ON STUD

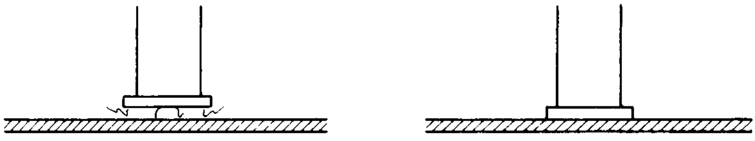

(b) TRIGGER PRESSED, CURRENT FLOWS, SPIGOT DISINTEGRATES ESTABLISHING AN ARC BETWEEN STUD AND WORK

(d) SPRING PRESSURE FORCES STUD INTO POOL. WELD COMPLETED IN A FEW MILLISECONDS coil is de-energized and spring pressure pushes the stud into contact with the workpiece, the discharge takes place and the weld is made.

This process minimizes the depth of penetration into the parent metal surface and is used for welding smaller diameter ferrous and non-ferrous studs and fasteners to light gauge material down to 0.45 mm thickness in low-carbon and austenitic stainless steel, and 0.7 mm in aluminium and its alloys and brass, the studs being from 2.8 to 6.5 mm diameter. Studs are usually supplied with a standard flange on the end to be welded (Fig. 5.18) but this can be reduced to stud diameter if required and centre punch marks should not be used for location. The studs can be welded on to the reverse side of finished or coated sheets with little or no marking on the finished side. The studs are not fluxed and no arc shield or ferrule is required.

Typical equipment with a weld time cycle of 3–7 milliseconds consists of a control unit and a hand- or bench-mounted tool or gun.

The control unit houses the banks of capacitors of 100 000–200 000 μF capacitance depending upon the size of the unit, the capacitance required for a given operation being selected by a switch on the front panel. The control circuits comprise a charging stage embodying a mains transformer and a bridge-connected full-wave silicon rectifier and the solid-state circuits for charging the capacitors to the voltage predetermined by the voltage sensing module. Interlocking prevents the energy being discharged by operating the gun switch until the capacitors have reached the power pre-set by voltage and capacitance controls.

The printed circuit voltage sensing module controls the voltage to which the capacitors are charged and switches them out of circuit when they are charged to the selected voltage, and is controlled by a voltage dial on the panel. A panel switch is also provided to discharge the capacitors if required.

Solid-state switching controls the discharge of the capacitors between stud and work, so there are no moving contactors. The gun, similar in appearance to the arc stud welding gun, contains the adjustable spring pressure unit which enables variation to be made in the speed of return of the stud into the workpiece, and a chuck for holding the stud. Legs are provided for positioning or there can be a nosecap gas shroud for use when welding aluminium or its alloys using an argon gas shield. The weld cannot be performed until legs or shroud are firmly in contact with the workpiece.

Welding rates attainable are 12 per minute at 6.35 mm diameter to 28 per minute at 3.2 mm diameter in mild steel. Similar rates apply to stainless steel and brass but are lower for aluminium because of the necessity of argon purging. Partial scorch marks may indicate cold laps (insufficient energy) with the possibility of the stud seating too high on the work. The weld should show even scorch marks all round the stud, indicating a sound weld. Excessive spatter indicates the use of too much energy.

Automatic single- and multiple-head machines, pneumatically operated and with gravity feed for the fasteners, are currently available.

Explosive welding

This process is very successfully applied to the welding of tubes to tube plates in heat exchangers, feedwater heaters, boiler tubes to clad tube plates, etc.; and also for welding plugs into leaking tubes to seal the leaks.

The welds made are sound and allow higher operating pressures and temperatures than with fusion welding, and the tubes may be of steel, stainless steel or copper; aluminium brass and bronze tubes in naval brass tube plates are also successfully welded.

The tube and tube plate may be parallel to each other (parallel geometry) with a small distance between them and the tube plate can be counterbored as in Fig. 5.19*a*. The explosive (e.g. trimonite) must have a low detonation velocity, below the velocity of sound in the material, and there is no limit to the joint area so that the method can be used for cladding surfaces. For the tube plate welds several charges can be fired simultaneously, the explosive being in cartridge form.

In the oblique geometry method now considered (YIMpact patent) the two surfaces are inclined at an angle to each other, Fig. 5.19*b*, the tube plate being machined or swaged as shown. As distance between tube and plate continuously increases because of the obliquity, there is a limit to the surface area which can be welded because the distance between tube and plate becomes too great.

The detonation value of the explosive can be above the velocity of sound

Fig. 5.19

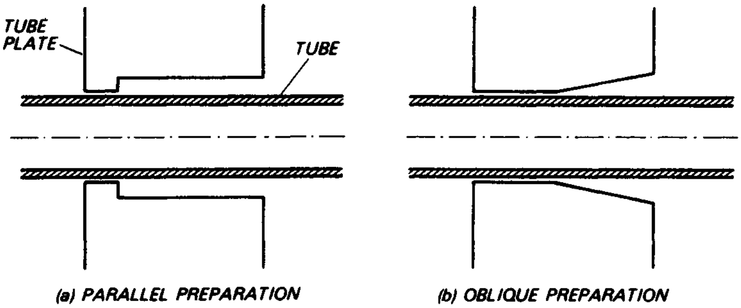

(a) PARALLEL PREPARATION (b) OBLIQUE PREPARATION in the material and the explosive (e.g. PETN) can be of pre-fabricated shape and is relatively cheap. The charge is fired electrically from a fuse head on the inner end of the charge and initiates the explosion, the detonation front then passing progressively through the charge.

The size of the charge depends upon the following variables: surface finish, angle of inclination of tube and plate, yield strength and melting point of the materials, the tube thickness and diameter; and the upper limit of the explosive is dependent upon the size of ligament of the tube plate between tubes, this usually being kept to a minimum in the interest of efficient heat exchange.

The tube plate is tapered towards the outer surface otherwise there would be a bulge in the tube on the inner side after welding and the tube would be difficult to remove. The charge must be fired from the inner end so that the weld will progress from contact point of tube and plate and thus the detonator wires must pass backwards along the charge to the outer end of the tube, the charge being withn a polythene insert (Fig. 5.20*a*).

Upon initiation of the explosion, tube and tube plate collide at the inner end of the taper and, due to the release of energy, proceed along it, and ideally the jet of molten metal formed at the collision point is ejected at the tube mouth. In effect the welded surfaces assume a sinusoidal wave form, some of the molten metal, which is rather porous and brittle, being entrapped in the troughs and crests of the wave. This entrapment can be reduced to a minimum by having a good surface finish (e.g. of the order of 0.003 mm) and the angle between tube and tube plate from 10° to 15°. At the lower angle the waves are pronounced and of shorter wavelength while at the higher angle the wavelength is longer and the waves more undulating so that 15° is the usual angle (Fig. 5.20*b* and *c*).

Surface oxide and impurities between the surfaces increase the charge required compared with a smooth surface and the positioning of the charge is important. If it is too far in, the energy at the mouth of the tube is not sufficient to produce a weld in this region, while if the charge is not far enough in the tube, welding is not commenced until some distance along the taper. To position the charge correctly and quickly a polythene insert has been developed to contain the charge and is positioned by a brass plug.

At present tubes of any thickness and of diameters 16–57 mm are weldable, with ·plate thickness greater than 32 mm and 9.5 mm plate ligaments.

The plug for explosive plugging of leaking tubes is of tubular form, the end of which is swaged to give the necessary taper, the polythene insert protruding beyond the open end of the plug to allow for extraction in case of misfire, thus increasing the safety factor (Fig. 5.21).

Fig. 5.20

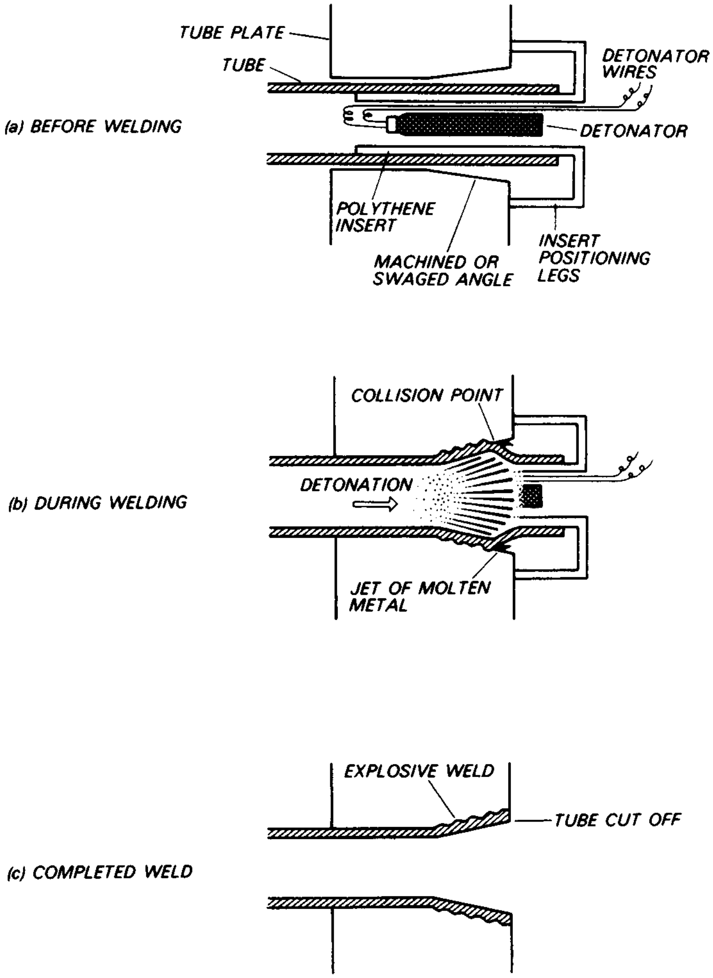
TUBE PLATE
TUBE
(a) BEFORE WELDING
DETONATOR WIRES
DETONATOR
POLYTHENE INSERT
MACHINED OR SWAGED ANGLE
INSERT POSITIONING LEGS
COLLISION POINT
(b) DURING WELDING
DETONATION
JET OF MOLTEN METAL
EXPLOSIVE WELD
TUBE CUT OFF
(c) COMPLETED WELD

Fig. 5.21. A swaged explosive plug.

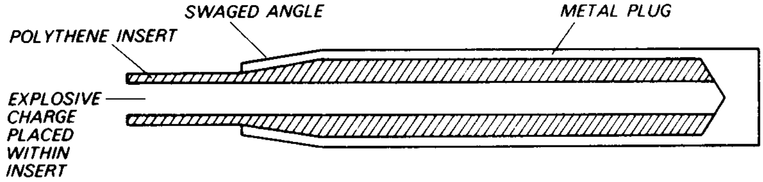
SWAGED ANGLE
METAL PLUG
POLYTHENE INSERT
EXPLOSIVE CHARGE PLACED WITHIN INSERT

Since all the configuration is confined to the plug, the tube plate hole is cleaned by grinding, the plug with explosive charge is inserted and the detonator wires connected. The plugs are available in diameters with 1.5 mm increments.

This method has proved very satisfactory in reducing the time and cost in repair work of this nature.

Personnel can be trained by the manufacturers or trained personnel are available under contract.

Gravity welding

This is a method for economically welding long fillets in the flat position using gravity to feed the electrode in and to traverse it along the plate. The equipment is usually used in pairs, welding two fillets at a time, one on each side of a plate, giving symmetrical welds and reducing stress and distortion. The electrode holder is mounted on a ball-bearing carriage and slides smoothly down a guide bar, the angle of which to the weld can be adjusted to give faster or slower traverse and thus vary the length of deposit of the electrode and the leg length of the weld.

The special copper alloy electrode socket is changed for varying electrode diameters and screws into the electrode holder, the electrode being pushed into a slightly larger hole in the socket and held there by the weight of the carriage. Turning the electrode holder varies the angle of electrode inclination. The base upon which the guide and support bar is mounted has two small ball bearings fitted so that it is easy to move along the base plate when resetting. If the horizontal plate is wider than about 280 mm a counterweight can be used with the base, otherwise the base can be attached to the vertical plate by means of two magnets or a jig can be used in place of the base to which the segment is fixed. A flexible cable connects the electrode holder to a disconnector switch carried on the support arm. This enables the current to be switched on and off so that electrodes can be changed without danger of shock. A simple mechanism at the bottom of the guide bar switches the arc off when the carriage reaches the bottom of its travel (Fig. 5.22).

At the present time electrodes up to 700 mm long are available in diameters of 3.5, 4.0, 4.5, 5.0 and 5.5 mm using currents of 220–315 A with rutile, rutile-basic and acid coatings suitable for various grades of steel.

Gravity welding is usually used for fillets with leg lengths of 5–8 mm, the lengths being varied by altering the length of deposit per electrode. An a.c. power source is used for each unit with an OCV of 60 and arc volts about 40 V with currents up to 300 A. Sources are available for supplying up to 6 units (three pairs), manageable by one welding operator and so arranged that when the current setting for one unit is chosen, the remaining units are supplied at this value. In general gravity welding is particularly suitable for welding, for example, long parallel stiffeners on large unit panels, enabling one operator to carry out three or four times the deposit length as when welding manually.

Thermit welding

Thermit (or alumino-thermic) is the name given to a mixture of finely divided iron oxide and powdered aluminium. If this mixture is placed in a fireclay crucible and ignited by means of a special powder, the action, once started, continues throughout the mass of the mixture, giving out great heat. The aluminium is a strong reducing agent, and combines with the oxygen from the iron oxide, the iron oxide being reduced to iron.

The intense heat that results, because of the chemical action, not only melts the iron, but raises it to a temperature of about 3000 °C. The aluminium oxide floats to the top of the molten metal as a slag. The crucible is then tapped and the superheated metal runs around the parts to be welded, which are contained in a mould. The high temperature of the iron results in excellent fusion taking place with the parts to be welded. Additions may be made to the mixture in the form of good steel scrap, or a small percentage of manganese or other alloying elements, thereby producing a good quality thermit steel. The thermit mixture may consist of about 5 parts of aluminium to 8 parts of iron oxide, and the weight of thermit used will depend on the size of the parts to be welded. The ignition

Fig. 5.22. Gravity welding.

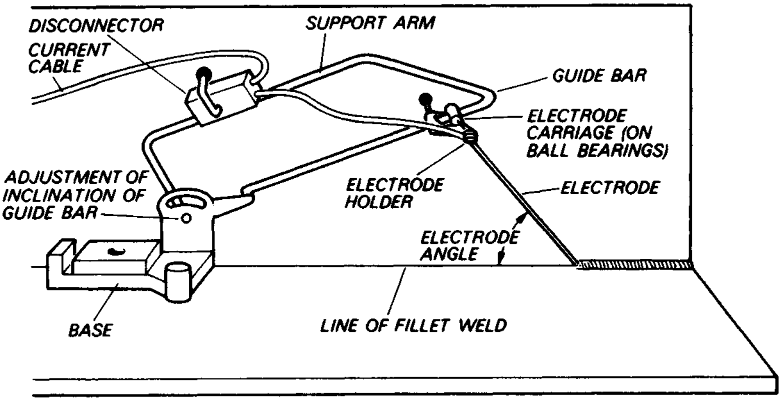
DISCONNECTOR
SUPPORT ARM
CURRENT CABLE
GUIDE BAR
ELECTRODE CARRIAGE (ON BALL BEARINGS)
ADJUSTMENT OF INCLINATION OF GUIDE BAR
ELECTRODE HOLDER
ELECTRODE
ELECTRODE ANGLE
BASE
LINE OF FILLET WELD

powder usually consists of powdered magnesium or a mixture of aluminium and barium peroxide.

Preparation

The ends which are to be welded are thoroughly cleaned of scale and rust and prepared so that there is a gap between them for the molten metal to penetrate well into the joint. Wax is then moulded into this gap, and also moulded into a collar round the fracture. This is important, as it gives the necessary reinforcement to the weld section. The moulding box is now placed around the joint and a mould of fireclay and sand made, a riser, pouring gate and pre-heating gate being included. The ends to be welded are now heated through the pre-heating gate by means of a flame and the wax is first melted from between the ends of the joint. The heating is continued until the ends to be welded are at red heat. This prevents the thermit steel being chilled, as it would be if it came into contact with cold metal. The pre-heating gate is now sealed off with sand and the thermit process started by igniting the powder. The thermit reaction takes up to about 1 minute, depending upon the size of the charge and the additions that have been made in the form of steel scrap and alloying elements etc. When the action is completed the steel is poured from the crucible through the pouring gate, and it flows around the red-hot ends to be welded, excellent fusion resulting. The riser allows extra metal to be drawn by the welded section when contraction occurs on cooling, that is it acts as a reservoir. The weld should be left in the mould as long as possible (up to 12 hours), since this anneals the steel and improves the weld.

Thermit steel is pure and contains few inclusions. It has a tensile strength of about 460 N/mm^2. The process is especially useful in welding together parts of large section, such as locomotive frames, ships' stern posts and rudders, etc. It is also being used in place of flash butt welding for the welding together of rail sections into long lengths.

At the present time British Rail practice is to weld 18 m long rails into lengths of 91 to 366 m at various depots by flash butt welding. These lengths are then welded into continuous very long lengths *in situ* by means of the thermit process and conductor rails are welded in the same way. Normal running rails have a cross-sectional area of 7184 mm^2 and two techniques are employed, one requiring a 7 minute pre-heat before pouring the molten thermit steel into the moulds and the other requiring only $1\frac{1}{2}$ minutes pre-heat, the latter being the technique usually employed. Excess metal can be removed by pneumatic hammer and hot chisel or by portable trimming machine.

Underwater welding

The three main methods of underwater welding at present are (1) wet, (2) localized dry chamber, (3) dry habitat (surroundings).

Wet welding

Wet welding is performed by the diver-welder normally using surface diving or saturation diving techniques using the MMA process with electrodes specially coated with insulating varnishes to keep them dry.

Current is supplied by a generator of some 400 A capacity at 60–100 OCV, directly to the welding torch head. Interchangeable collets of 4 mm and 4.8 mm ($\frac{5}{32}''$ and $\frac{3}{16}''$), hold electrodes of these diameters and are tightened by the twist-grip control, which is also used to eject the stub when electrodes are changed. Watertight glands and washers prevent seepage of water into the body of the torch, which is tough rubber covered thus reducing the danger of electric shock.

Arc stability is less than that in air due to the large volumes of gas and steam which are evolved, making vision difficult and as a result touch welding is generally used. The presence of water in the immediate vicinity of the weld except at the molten pool under the arc stream results in the very rapid quenching of the weld metal, which produces a hard, narrow heat affected zone and can give rise to severe hydrogen cracking. At present this method is used for non-critical welds, critical welding being carried out under dry conditions.

Localized dry chamber (hyperbaric)

These chambers, in which welding is performed at pressure above one atmosphere, are generally made of steel and are constructed with antichambers around the section to be welded. Water is then displaced from the chamber by gas pressure. The gas in the chamber used by the diver-welder is typically a mixture of helium and oxygen (heliox). This mixture eliminates the harmful effect of nitrogen, the helium being the carrier for the oxygen – a partial pressure of oxygen above 2 bar can be poisonous. The oxygen content is adjusted until the diver-welder does not use a breathing mask, but at the moment of welding he puts on a breathing–welding mask which has a gas supply separate from that of the chamber and umbilicals connected to the mask to enable the breathing gases to be exhausted outside the chamber.

TIG, MMA and flux cored methods of welding are used down to a depth of 300 m of water but TIG, using high purity argon as the shielding gas, is the slowest method as it is a low deposition rate process. It can be used for root runs and hot passes, but MMA or the flux cored method is used for filling and capping. Stringer bead techniques may be used to impart toughness, and temper beads to minimize HAZ hardness. Generally the high-strength, basic-coated steel electrodes which may contain iron powder are used. They are pre-heated in a drying oven immediately before use and pre-heat of about 100°C can be applied locally to the joint.

After welding, the joints are subject to X-ray or ultra-sonic testing according to the welding codes used.

Coffer dam

A coffer dam is a watertight case of steel piling erected around a given point to keep out the water. By pumping the dam dry, welding can be performed at atmospheric pressure up to about 18 m deep and welding can proceed irrespective of the state of the tides if in tidal water.

In this case the welding is performed exactly as if it were on shore and the welders are in audio and visual contact with the welding process engineer. Much of the damage to offshore platforms is in the splash zone so there is quite a saving in cost using this method as opposed to hyperbaric welding. Fig. 5.23 shows a typical underwater torch, which can also be used with small modifications for underwater cutting.

Fig. 5.23. Underwater welding (and cutting) torch.

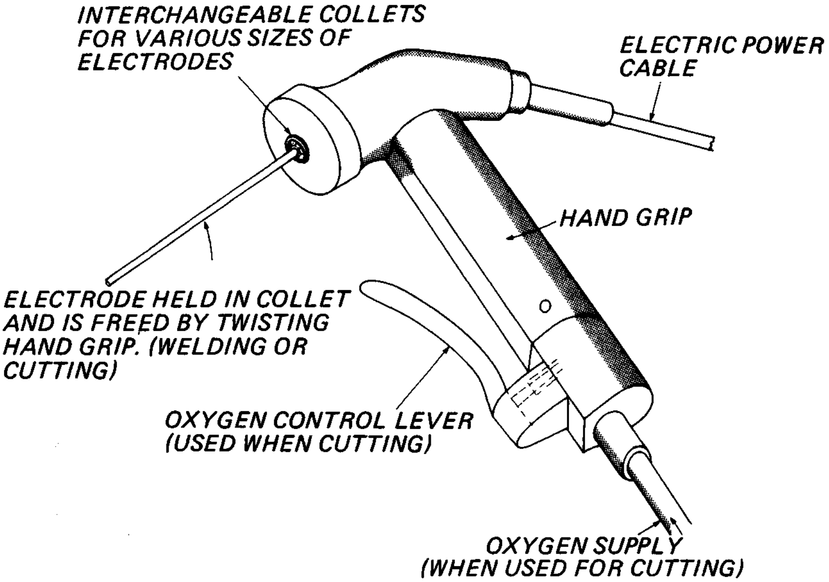
INTERCHANGEABLE COLLETS FOR VARIOUS SIZES OF ELECTRODES
ELECTRIC POWER CABLE
HAND GRIP
ELECTRODE HELD IN COLLET AND IS FREED BY TWISTING HAND GRIP. (WELDING OR CUTTING)
OXYGEN CONTROL LEVER (USED WHEN CUTTING)
OXYGEN SUPPLY (WHEN USED FOR CUTTING)

6

Oxy-acetylene welding

Principles and equipment

(See also pp. 323–5 for general safety precautions)

For oxy-acetylene welding, the oxygen is supplied from steel cylinders and the acetylene either from cylinders or from an acetylene generator which can be of the medium-pressure or low-pressure type.

With cylinder gas, the pressure is reduced to 0.06 N/mm^2 or under, according to the work, by means of a pressure-reducing valve and the acetylene is passed to the blowpipe where it is mixed with oxygen in approximately equal proportions, and finally passed into the nozzle or tip to be burnt.

The medium-pressure acetylene generator delivers gas to any desired pressure up to a maximum of 0.06 N/mm^2 (0.6 bar) in the same way as cylinder gas.

The low-pressure generator produces gas at a pressure of only a few millimetres water column, necessitating the use of 'injector' blowpipes where the high-pressure oxygen injects or sucks the acetylene into the blowpipe. In some cases, as for example if the supply pipes are small for the volume of gas to be carried or if it is desirable to use the equal pressure type of blowpipe, a booster is fitted to the low-pressure generator and the gas pressure increased to 0.06 N/mm^2.

Without Home Office approval, the maximum pressure at which acetylene may be used in England is 0.06 N/mm^2. With approval, the pressure may be increased to 0.15 N/mm^2, but this is rare and applies only in special cases.

Oxygen

The oxygen for both high- and low-pressure systems is supplied in solid drawn steel cylinders at pressures up to 200 bar at 15 °C.* The cylinders are rated according to the amount of gas they contain, varying from 0.68 m^3 at a pressure of 137 bar to 9.66 m^3 at 200 bar.

---

* *Note.* 1 bar = 0.1 N/mm^2 = approximately 15 lbf/in^2.

The volume of oxygen contained in the cylinder is approximately proportional to the pressure; hence for every 10 litres of oxygen used, the pressure drops about 0.02 N/mm^2. This enables us to tell how much oxygen remains in a cylinder. The oxygen cylinder is provided with a valve threaded right hand and is painted black. On to this valve, which contains a screw-type tap, the pressure regulator and pressure gauge are screwed. The regulator adjusts the pressure to that required at the blowpipe. Since grease and oil can catch fire spontaneously when in contact with pure oxygen under pressure, they must never be used on any account upon any part of the apparatus. Leakages of oxygen can be detected by the application of a soap solution, when the leak is indicated by the soap bubbles. Never test for leakages with a naked flame.

Liquid oxygen

Liquid oxygen, nitrogen, argon and LPG are available in bulk supply from tankers to vacuum-insulated evaporators (VIE) in which the liquid is stored at temperatures of −160 to −200°C and are very convenient for larger industrial users.

There is no interruption in the supply of gas nor drop in pressure during filling.

The inner vessel, of austenitic stainless steel welded construction, has dished ends and is fitted with safety valve and bursting disc and is available in various sizes with nominal capacities from 844 to 33 900 m^3. Nominal capacity is the gaseous equivalent of the amount of liquid that the vessel will hold at atmospheric pressure. The outer vessel is of carbon steel and fitted with pressure release valve. The inner vessel is vacuum and pearlite insulated from the outer vessel, thus reducing the thermal conductivity to a minimum. The inner vessel *A* (Fig. 6.1) contains the liquid with gas above, and gas is withdrawn from the vessel through the gaseous withdrawal line *B* and rises to ambient temperature in the superheater-vaporizer *C*, from which it passes to the supply pipeline. If the pressure in the supply falls below the required level the pressure control valve *D* opens and liquid flows under gravity to the pressure-raising vaporizer *E*, where heat is absorbed from the atmosphere, and the liquid vaporizes and passes through the gas withdrawal point *H* raising the pressure to the required pre-set level which can be up to 1.7 N/mm^2 (250 lbf/in^2), and the valve *D* then shuts.

In larger units, to allow for heavy gas demand, when the pressure falls on the remote side of the restrictive plate *F*, liquid flows from the vessel via the withdrawal line *G* and passes to the superheater-vaporizer where it changes to the gaseous form and is heated to ambient temperature, finally passing to the pipeline.

During periods when the VIE is not in use the valve *D* remains shut. Heat from the outside atmosphere gradually flows through the insulation between the vessels so that more liquid is vaporized and the pressure of the gas rises. This rate of heat leakage is slow, however, and it usually takes about seven days for the pressure to rise sufficiently to lift the safety valve, so that under normal working conditions there is almost zero loss.

Acetylene

In the high-pressure system the acetylene is stored in steel cylinders similar to the oxygen cylinders. Acetylene gas, however, is unstable when compressed to high pressures, and because of this it is contained in the bottles dissolved in a chemical called acetone; hence the name 'dissolved acetylene'.

The acetone is contained in a porous spongy mass of a substance such as charcoal, asbestos, kapok or other such material. Acetone can absorb 25 times its own volume of acetylene at normal temperature and pressure and for every increase of one atmosphere of pressure (0.1 N/mm^2 or 1 bar) it can absorb an equal amount.

Fig. 6.1. Vacuum insulated evaporator.

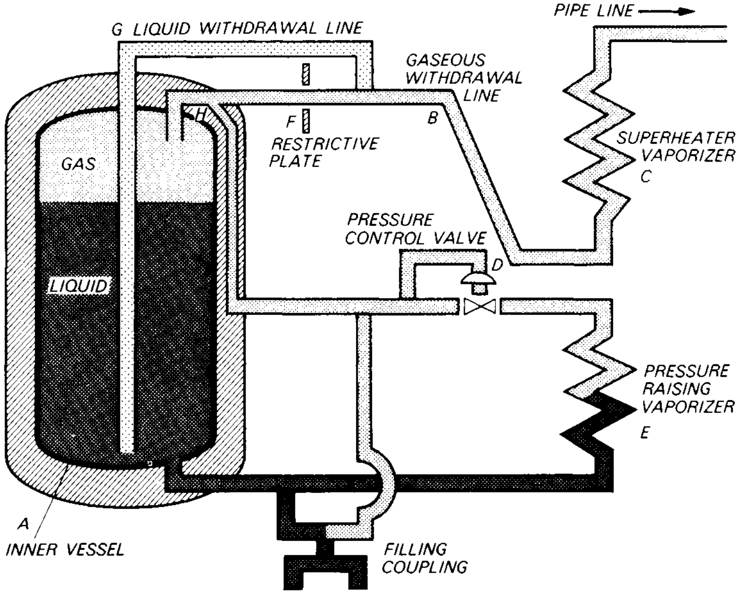
G LIQUID WITHDRAWAL LINE
PIPE LINE
GASEOUS WITHDRAWAL LINE
H
F
RESTRICTIVE PLATE
B
GAS
SUPERHEATER VAPORIZER C
PRESSURE CONTROL VALVE
D
LIQUID
PRESSURE RAISING VAPORIZER E
A
INNER VESSEL
FILLING COUPLING

The inlet pressure of the acetylene is usually about 20 bar maximum, typical capacities varying from 0.57 to 8.69 m^3, depending on the cylinder type. The gas leaves the cylinder through a valve after passing through a filter pad. The valve has a screw tap fitted and is screwed left-hand. The cylinder is painted maroon and a regulator (also screwed left-hand) is necessary to reduce the pressure to 1.3 bar maximum as required by the operator.

The amount of dissolved acetylene in a cylinder cannot be determined with any accuracy from the pressure gauge reading since it is in the dissolved condition. The most accurate way to determine the quantity of gas in a cylinder is to weigh it, and subtract this weight from the weight of the full cylinder, which is usually stamped on the label attached to the cylinder. The volume of gas remaining in the cylinder is calculated by remembering that 1 litre of acetylene weighs 1.1 g.

As long as the volume of acetylene drawn from the cylinder is not greater than $\frac{1}{5}$ of its capacity per hour, there is no appreciable amount of acetone contained in gas; hence this rate of supply should not be exceeded. The advantages of dissolved acetylene are that no licence is required for storage of the cylinders, there is no fluctuation of pressure in use, and the gas is always dry, clean and chemically pure, resulting in a reliable welding flame, and it avoids the need to charge and maintain a generator and to dispose of the sludge. There is no discernible difference in efficiency between generated and dissolved acetylene when both are used under normal conditions. Acetylene is highly inflammable and no naked lights should be held near a leaking cylinder, valve or tube. Leaks can be detected by smell or by soap bubbles. If any part of the acetylene apparatus catches fire, immediately shut the acetylene valve on the cylinder. The cylinder should be stored and used in an upright position.

The following is a summary of the main safety precautions to be taken when storing and using cylinders of compressed gas.

Storage

(1) Store in a well ventilated, fire-proof room with flame-proof electrical fittings. Do not smoke, wear greasy clothing or have exposed flames in the storage room.

(2) Protect cylinders from snow and ice and from the direct rays of the sun, if stored outside.

(3) Store away from sources of heat and greasy and oily sources. (Heat increases the pressure of the gas and may weaken the cylinder wall. Oil and grease may ignite spontaneously in the presence of pure oxygen.)

(4) Store acetylene cylinders in an upright position and do not store oxygen and combustible gases such as acetylene and propane together.
(5) Keep full and empty cylinders apart from each other.
(6) Avoid dropping and bumping cylinders violently together.

Use

(7) Keep cylinders away from electrical apparatus or live wires where there may be danger of arcing taking place.
(8) Protect them from the sparks and flames of welding and cutting operations.
(9) Always use pressure-reducing regulators to obtain a supply of gas from cylinders.
(10) Make sure that cylinder outlet valves are clear of oil, water and foreign matter, otherwise leakage may occur when the pressure-reducing regulators are fitted.
(11) Do not use lifting magnets on the cylinders. Rope slings may be used on single cylinders taking due precautions against the cylinder slipping from the sling. Otherwise use a cradle with chain suspension.
(12) Deload the diaphragm of the regulator by unscrewing before fitting to a full cylinder, and open the cylinder valve slowly to avoid sudden application of high pressure on to the regulator.
(13) Do not overtighten the valve when shutting off the gas supply; just tighten enough to prevent any leakage.
(14) Always shut off the gas supply when not in use for even a short time, and always shut off when moving cylinders.
(15) Never test for leaks with a naked flame; use soapy water.
(16) Make sure that oxygen cylinders with round bases are fastened when standing vertically, to prevent damage by falling.
(17) Thaw out frozen spindle valves with hot water NOT with a flame.
(18) Use no copper or copper alloy fitting with more than 70% copper because of the explosive compounds which can be formed when in contact with acetylene.
(19) Do not use oil or grease or other lubricant on valves or other apparatus, and do not use any jointing compound.
(20) Blow out the cylinder outlet by quickly opening and closing the valve before fitting the regulator.
(21) Should an acetylene cylinder become heated due to any cause, immediately take it outdoors, immerse in water or spray it with water, open the valve and keep as cool as possible until the cylinder is empty. Then contact the suppliers.

(22) Do not force a regulator on to a damaged outlet thread. Report damage to cylinders to the suppliers.

Note. Also refer to Form 1704, 'Safety measures required in the use of acetylene gas and in oxy-acetylene processes in factories' (HMSO) and also 'Safety in the use of compressed gas cylinders', a booklet published by the British Oxygen Gases Ltd.

The cylinder outlet union is screwed left hand for combustible gases and right hand for non-combustible gases, the thread being $\frac{5}{8}$ in (16 mm) BSP except for CO_2 which is 21.8 mm, 14 TPI male outlet.

Acetylene (dissolved) is contained in a maroon coloured cylinder with the name ACETYLENE. Oxygen (commercial) is contained in a black coloured cylinder. (See Appendix 3 for full description of colour codes used on cylinders (BS 349, BS 381 C).) Consult BS 679, Gas welding filters (GWF) for the correct shade of filter for eye protection and pp. 323–5.

Medium-pressure acetylene generators

The medium-pressure generator uses small granulated carbide. 50 kg of good calcium carbide will produce about 14 000 litres of acetylene, an average practical value being 250 litres of acetylene per kilogram of carbide. The most popular sizes are 20/50 mm, 15/25 mm and 10/15 mm. The generator is self-contained (i.e. there is no separate gas holder) and consists of a water tank surmounted by a carbide hopper, in the top of which is a diaphragm. The carbide feed valve is controlled by the diaphragm, which is actuated by the pressure of the gas generated in the tank. When the pressure falls, carbide flows into the tank; as the pressure builds up the flow ceases. The gas pressure at which the generator will work is adjusted by means of a spring fitted to the opposite side of the diaphragm, ensuring close control of pressure with generation strictly in accordance with the demand. The carbide hopper is either made of glass or is fitted with windows so that the quantity of carbide remaining in the hopper can be ascertained at a glance.

Owing to the relatively large volume of water into which the small grains of carbide fall, there is no possibility of overheating and the carbide is completely slaked. The sludge, which collects at the bottom of the tank, and is emptied each time the generator is charged, consists of a thin milky fluid.

The impurities in crude acetylene consist chiefly of ammonia, hydrides of phosphorus, sulphur and nitrogen, and there are also present water vapour and particles of lime.

These impurities must be removed before the gas is suitable for welding use; the gas is filtered and washed and chemically purified by passing it through salts of ferric iron.

The normal method of testing acetylene to ascertain whether it is being efficiently purified is to hold a silver nitrate test paper (a piece of filter paper soaked in a solution of silver nitrate) in the stream of gas for about 10 seconds. If the acetylene is being properly purified, there will be no trace of stain on the silver nitrate paper.

The reducing valve or pressure regulator

In order to reduce the pressure of either oxygen or dissolved acetylene from the high pressure of the storage cylinder to that required at the blowpipe, a regulator or reducing valve is necessary. Good regulators are essential to ensure the even flow of gas to the blowpipe. A reference to Fig. 6.2 will make the principle of operation of the regulator clear. The gas enters the regulator at the base via a fine sintered metal filter and the cylinder pressure is indicated on the first gauge. The gas then enters the body of the regulator *R* through the aperture *A*, which is controlled by the

Fig. 6.2. Pressure regulator (single stage).

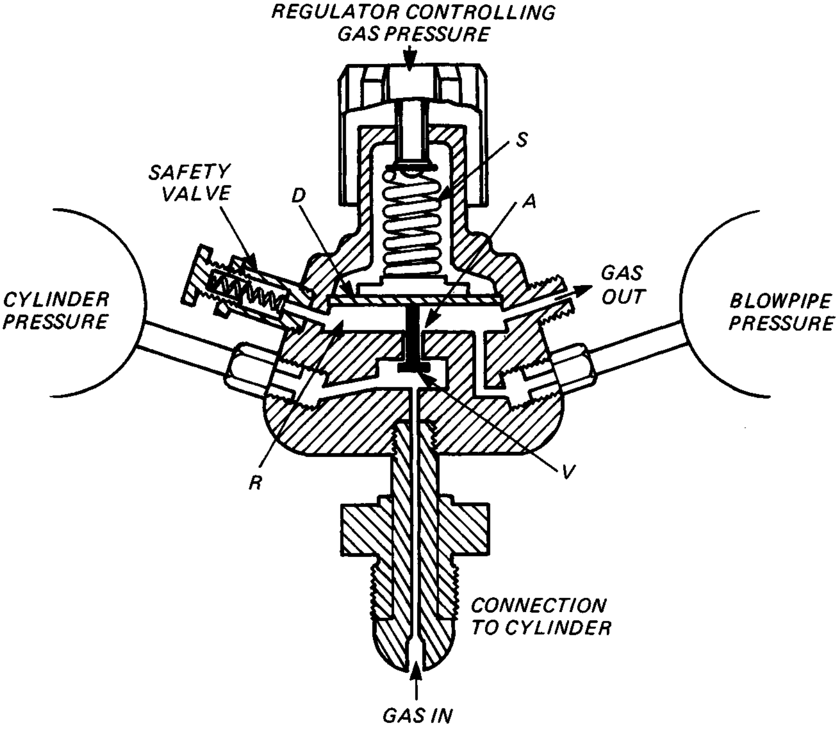
REGULATOR CONTROLLING GAS PRESSURE
S
SAFETY VALVE
D
A
CYLINDER PRESSURE
GAS OUT
BLOWPIPE PRESSURE
R
V
CONNECTION TO CYLINDER
GAS IN

NOTE: RH CONNECTIONS FOR OXYGEN, ARGON, NITROGEN, HELIUM,
LH CONNECTIONS FOR ACETYLENE, HYDROGEN.

valve *V*. The pressure inside the regulator rises until it is sufficient to overcome the pressure of the spring *S*, which loads the diaphragm *D*. The diaphragm is therefore pushed back and the valve *V*, to which it is attached, closes the aperture *A* and prevents any more gas from entering the regulator.

The outlet side is also fitted with a pressure gauge (although in some cases this may be dispensed with) which indicates the working pressure on the blowpipe. Upon gas being drawn off from the outlet side the pressure inside the regulator body falls, the diaphragm is pushed back by the spring, and the valve opens, letting more gas in from the cylinder. The pressure in the body *R* therefore depends on the pressure of the springs and this can be adjusted by means of a regulator as shown.

Regulator bodies are made from brass forgings and single stage regulators are fitted with one safety valve set to relieve pressures of 16–20 bar, and should it be rendered inoperative, as for example by misuse, it ruptures at pressures of 70–80 bar and vents to the atmosphere through a vent in the bonnet. Single-stage regulators are suitable for general welding with maximum outlet pressure of 2.1 bar and for scrap cutting and heavy-duty cutting, thermic lancing and boring with outlet pressures 8.3–14 bar.

Figure 6.3*a*, *b* and *c* shows a two-stage regulator. This reduces the pressure in two stages and gives a much more stable output pressure than the single-stage regulator.

It really consists of two single stages in series within one body forging. The first stage, which is pre-set, reduces the pressure from that of the cylinder to 23 bar, and gas at that pressure passes into the second stage, from which it emerges at a pressure set by the pressure-adjusting control screw attached to the diaphragm. High-pressure regulators except those for acetylene and carbon dioxide are designed for inlet pressures of 230 bar and tested to four times working pressure. A high-flow safety relief valve is fitted and ensures that any abnormal pressures are quickly vented to atmosphere, the valve reseating itself when pressures return to normal. If a safety valve is blowing, the main valve is not seating and the regulator should immediately be taken out of service and sent for overhaul. Needle-type control valves can be fitted to regulator outlets.

The correct blowpipe pressure is obtained by adjusting the pressure of the spring with the control knob and noting the pressure in bar on the blowpipe pressure gauge. On changing a tip the regulator is set with its finger on the tip number on the scale, and final accurate adjustment of the flame is made with the blowpipe regulating valves. This is a simple and convenient method. The regulators are supplied with a table indicating the suitable pressures for various nozzles, which are stamped with their consumption of gas in litres per minute or suitable numbers. With practice the welder soon recognizes the correct pressures for various tips without reference to the table.

Regulators can be obtained for a wide variety of gases (given here with maximum outlet pressures): oxygen (welding and medium cutting and heating 3.5 bar, heavy cutting and heating 10 bar); acetylene (1.3 bar); argon (3.5 bar); nitrogen (3.5 bar); propane (heavy cutting and heating 3.5 bar); hydrogen (3.5 bar, for pressurizing 10 bar); carbon dioxide (10 bar).

To enable two, three or more cylinders of gas to be connected together, as may be required when heavy cutting work is to be done and the oxygen consumption is very great, special adaptors are available, and these feed the bottles into one gauge. In this way a much steadier supply of oxygen is obtained.

Two operators may also be fed from one cylinder of oxygen or acetylene by using a branched gauge with two regulators. The type of branched gauge which has only one regulator feeding two outlet pipes is not recommended, since any alteration of the blowpipe pressure by one operator will affect the flame of the other operator.

Owing to the rapid expansion of the oxygen in cases where large

Fig. 6.3. (*a*) Passage of gas through a two-stage regulator.

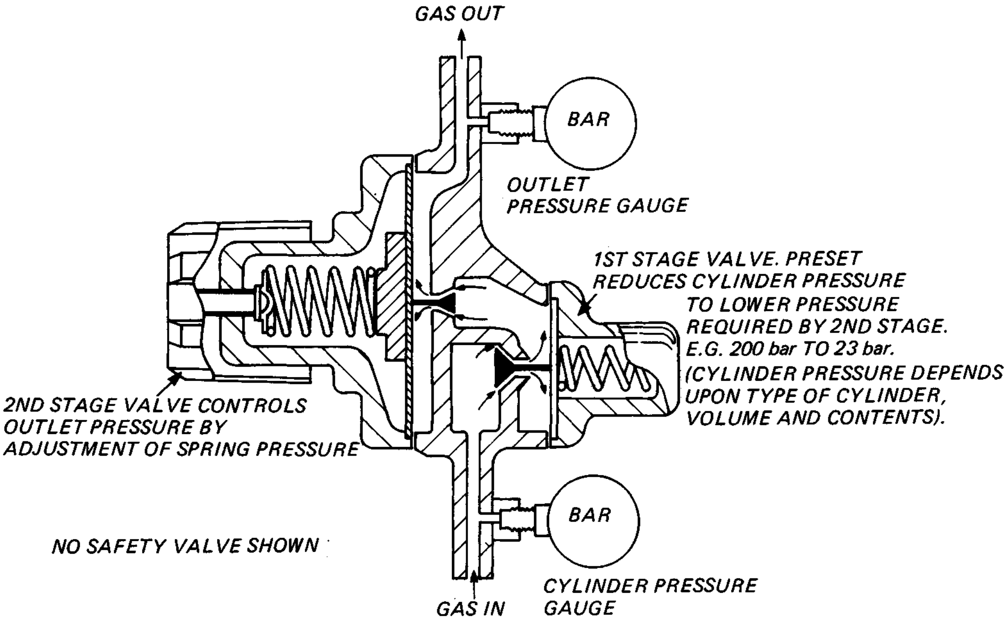
GAS OUT
BAR
OUTLET PRESSURE GAUGE
1ST STAGE VALVE. PRESET REDUCES CYLINDER PRESSURE TO LOWER PRESSURE REQUIRED BY 2ND STAGE. E.G. 200 bar TO 23 bar. (CYLINDER PRESSURE DEPENDS UPON TYPE OF CYLINDER, VOLUME AND CONTENTS).
2ND STAGE VALVE CONTROLS OUTLET PRESSURE BY ADJUSTMENT OF SPRING PRESSURE
BAR
NO SAFETY VALVE SHOWN
CYLINDER PRESSURE GAUGE
GAS IN

Fig. 6.3. (*b*) Multi-stage regulator.

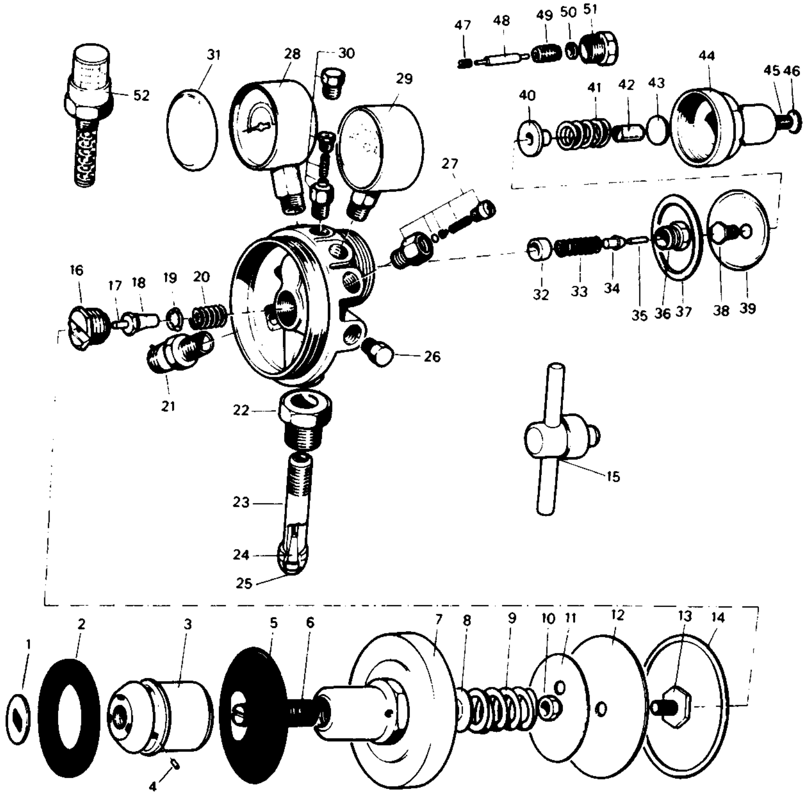
47
48
49
50
51
31
28
30
52
29
44
45
46
40
41
42
43
27
16
17
18
19
20
32
33
34
35
36
37
38
39
26
21
22
15
23
24
25
1
2
3
4
5
6
7
8
9
10
11
12
13
14

Key:

1. Disc monogram.
2. Ring cover.
3. Knob.
4. Set screw.
5. Name plate
6. Screw P. A.
7. Bonnet.
8. Spring centre.
9. Spring.
10. Nut.
11. Packing plate.
12. Diaphragm.
13. Diaphragm carrier.
14. Washer.
15. Screw P. A.
16. Nozzle.
17. Valve pin.
18. Valve.
19. Washer.
20. Spring.
21. Outlet adaptor LH.
    Outlet adaptor RH.
22. Inlet nut LH.
    Inlet nut RH.
23. Inlet nipple.
24. Filter.
25. Retaining ring.
26. Plug.
27. Safety valve LP.
28. Gauge.
29. Gauge.
30. Relief valve HP.
31. Gauge glass.
32. Sleeve.
33. Spring.
34. Valve.
35. Plunger.
36. Nozzle and seat.
37. Sealing ring.
38. Diaphragm carrier.
39. Diaphragm.
40. Disc.
41. Spring.
42. Damper plug.
43. Pivot.
44. Bonnet.
45. Screw.
46. Disc anti-tamper.
47. Spring.
48. Valve.
49. Seat retainer.
50. Seat.
51. Valve holder.
52. Indicator assembly.

Fig. 6.3. (*c*) Two-stage oxygen regulator. Output 0–10 bar.

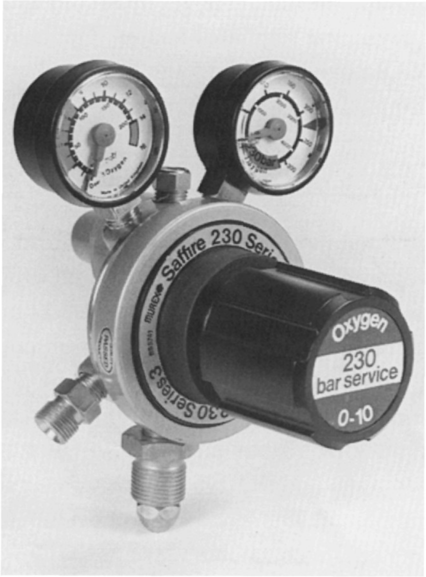
Saffire 230 Seri
Oxygen
230
bar service
0-10
230 Series 3

Fig. 6.3. (*d*) Two-stage regulator. 230 bar inlet fitted with resettable flashback arrestor.

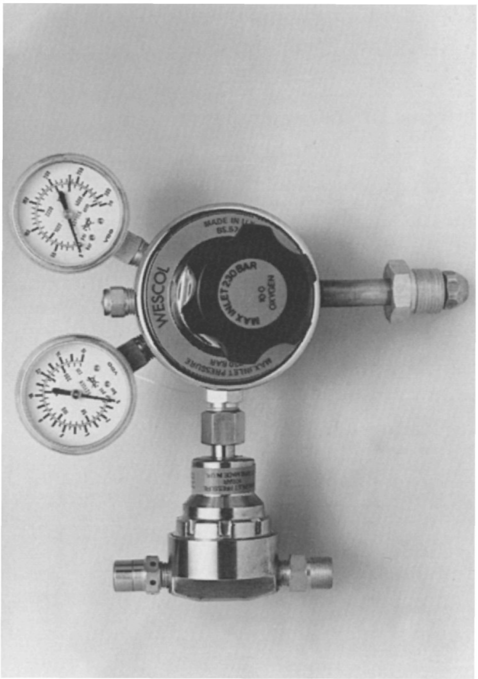
WESCOL
MAX INLET 230 BAR
10-0
OXYGEN

quantities are being used, the regulator may become blocked with particles of ice, causing stoppage. This happens most frequently in cold weather, and can be prevented by use of an electric regulator heater. The heater screws into the cylinder and the regulator screws into the heater. The heater is plugged into a source of electric supply, the connexion being by flexible cable.

Hoses

Hoses are usually of a seamless flexible tube reinforced with plies of closely woven fabric impregnated with rubber and covered overall with a tough, flexible, abrasion-resistant sheath giving a light-weight hose. They are coloured blue for oxygen, red for fuel gases, black for non-combustible gases and orange for LPG, Mapp and natural gas. Available lengths are from 5 to 20 m, with bore diameters 4.5 mm for maximum working pressure of 7 bar, 8 mm for a maximum of 12 bar and 10 mm for a maximum working pressure of 15 bar. Nipple- and nut-type connexions and couplers are available for 4.5 mm ($\frac{3}{8}$ in.), 8 and 10 mm hoses with 6.4 mm ($\frac{1}{4}$ in. BSP) and 10 mm ($\frac{3}{8}$ in. BSP) nuts. A hose check valve is used to prevent feeding back of gases from higher or lower pressures and reduces the danger of a flashback due to a blocked nozzle, leaking valve, etc. It is connected in the hose at the blowpipe end or to the economizer or regulator, and consists of a self-aligning spring-loaded valve which seals off the line in the event of backflow. BS 924 J and 796 J apply to hoses.

The welding blowpipe or torch

There are two types of blowpipes: (1) high pressure, (2) low pressure, and each type consists of a variety of designs depending on the duty for which the pipe is required. Special designs are available for rightward and leftward methods of welding (the angle of the head is different in these designs), thin gauge or thick plate, etc., in addition to blowpipes designed for general purposes.

The high-pressure blowpipe is simply a mixing device to supply approximately equal volumes of oxygen and acetylene to the nozzle, and is fitted with regulating valves to vary the pressure of the gases as required (Fig. 6.4*a*, *b*). A selection of shanks is supplied with each blowpipe, having orifices of varying sizes, each stamped with a number or with the consumption in litres per hour (l/h). Various sizes of pipes are available, from a small light variety, suitable for thin gauge sheet, to a heavy duty pipe. A high-pressure pipe cannot be used on a low-pressure system.

The low-pressure blowpipe has an injector nozzle inside its body through which the high-pressure oxygen streams (Fig. 6.5). This oxygen draws the

Fig. 6.4. (*a*) Principle of the high-pressure blowpipe.

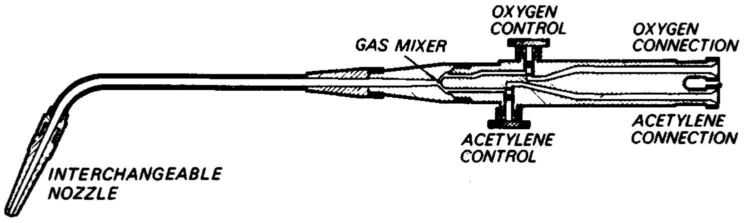
OXYGEN CONTROL
GAS MIXER
OXYGEN CONNECTION
ACETYLENE CONNECTION
ACETYLENE CONTROL
INTERCHANGEABLE NOZZLE

Fig. 6.4. (*b*) Blowpipe.

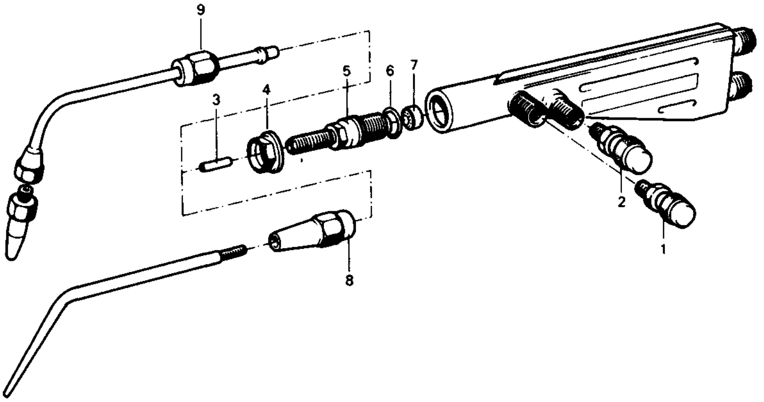

Key:

1. Spindle assembly RH.	6. 'O' ring.
2. Spindle assembly LH.	7. Mixer spool.
3. Insert.	8. Adaptor nut.
4. Nut–locking.	9. Neck assembly.
5. Mixer.	

Fig. 6.5. Principle of the low-pressure blowpipe.

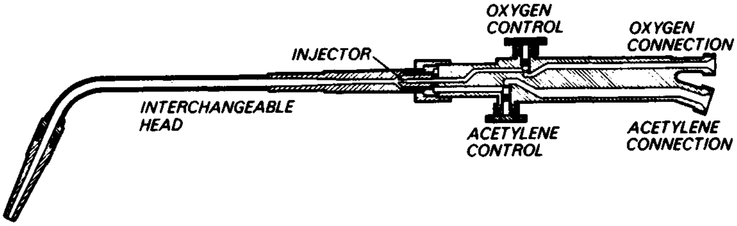
OXYGEN CONTROL
OXYGEN CONNECTION
INJECTOR
INTERCHANGEABLE HEAD
ACETYLENE CONTROL
ACETYLENE CONNECTION

low-pressure acetylene into the mixing chamber and gives it the necessary velocity to preserve a steady flame, and the injector also helps to prevent backfiring. The velocity of a 1/1 mixture of oxygen/acetylene may be 200 m per minute, while the maximum gas velocity occurs for a 30% acetylene mixture and may be up to 460 m per min. (These figures are approximate only.)

It is usual for the whole head to be interchangeable in this type of pipe, the head containing both nozzle and injector. This is necessary, since there is a corresponding injector size for each nozzle. Regulating valves, as on the high-pressure pipes, enable the gas to be adjusted as required. The low-pressure pipe is more expensive than the high-pressure pipe; and it can be used on a high-pressure system if required, but it is now used on a very small scale.

A very useful type of combined welding blowpipe and metal-cutting torch is shown in Fig. 6.6. The shank is arranged so that a full range of nozzles, or a cutting head, can be fitted. The design is cheaper than for a corresponding separate set for welding and cutting, and the cutter is sufficient for most work.

The oxy-acetylene flame

The chemical actions which occur in the flame have already been discussed in Volume 1, and we will now consider the control and regulation of the flame to a condition suitable for welding.

Fig. 6.6. Combined welding and cutting pipes. Will weld sections from 1.6 mm to 32 mm thick, and cut steel up to 150 mm thick with acetylene and 75 mm with propane.

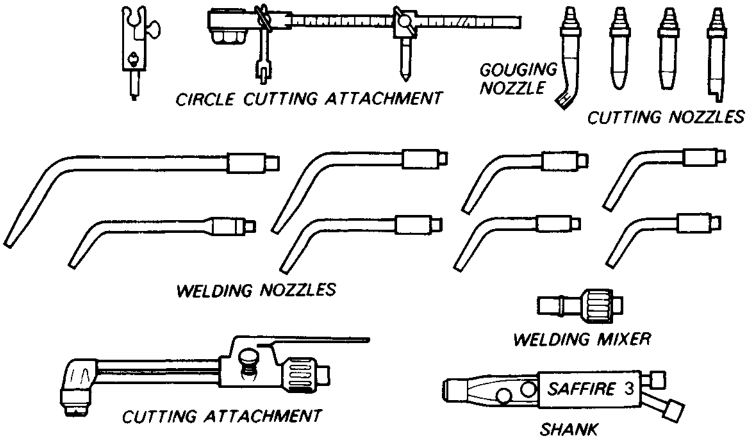

Adjustment of the flame.[*] To adjust the flame to the neutral condition the acetylene is first turned on and lit. The flame is yellow and smoky. The acetylene pressure is then increased by means of the valve on the pipe until the smokiness has just disappeared and the flame is quite bright. The condition gives approximately the correct amount of acetylene for the particular jet in use. The oxygen is then turned on as quickly as possible, and as its pressure is increased the flame ceases to be luminous. It will now be noticed that around the inner blue luminous cone, which has appeared on the tip of the jet, there is a feathery white plume which denotes excess acetylene (Fig. 6.7*a*). As more oxygen is supplied this plume decreases in size until there is a clear-cut blue cone with no white fringe (Fig. 6.7*b*). This is the neutral flame used for most welding operations. If the oxygen supply is further increased, the inner blue cone becomes smaller and thinner and the outer envelope becomes streaky; the flame is now oxidizing (Fig. 6.7*c*). Since the oxidizing flame is more difficult to distinguish than the carbonizing or carburizing (excess acetylene) flame, it is always better to start with excess acetylene and increase the oxygen supply until the neutral condition is reached, than to try to obtain the neutral flame from the oxidizing condition.

Some welders prefer to regulate the oxygen pressure at the regulator itself. The acetylene is lighted as before, and with the oxygen valve on the blowpipe turned full on, the pressure is adjusted correctly at the regulator until the flame is neutral. In this way the welder is certain that the regulator is supplying the correct pressure to the blowpipe for the particular nozzle being used.

Fig. 6.7

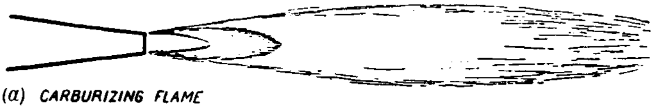

(*a*) CARBURIZING FLAME

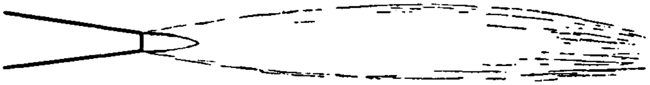

(*b*) NEUTRAL FLAME

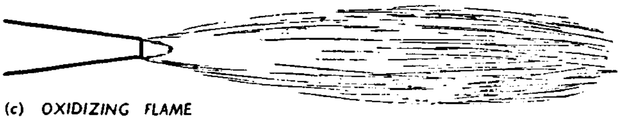

(c) OXIDIZING FLAME

---

[*] The pressure on oxygen and acetylene gauges is approximately that given in the table on p. 276.

Selection of correct nozzle. As the thickness of the work to be welded increases, the flame will have to supply more heat, and this is made possible by increasing the size of the nozzle. The nozzle selected may cover one or two thicknesses of plate; for example, a nozzle suitable for welding 6.4 mm plate will weld both 4.8 mm and 8 mm plate by suitable regulation of the pressure valves. This is because the blowpipe continues to mix the gases in the correct proportion over a range of pressures. If, however, one is tempted to weld a thickness of plate with a nozzle which is too large, by cutting down the supply of gas at the valves instead of changing the nozzle for one smaller, it will be noticed that explosions occur at the nozzle when welding, these making the operation impossible. These explosions indicate too low a pressure for the nozzle being used.

If, on the other hand, one attempts to weld too great a thickness of metal with a certain nozzle, it will be noticed that as one attempts to increase the pressure of the gases beyond a certain point to obtain a sufficiently powerful flame, the flame leaves the end of the nozzle. This indicates too high a pressure and results in a *hard* noisy flame. It is always better to work with a soft flame which is obtained by using the correct nozzle and pressure. Thus, although there is considerable elasticity as to the thickness weldable with a given nozzle, care should be taken not to overtax it.

Use and care of blowpipe

Oil or grease should upon no account be used on any part of the blowpipe, but a non-oily graphite may be used and is useful for preventing wear and any small leaks.

A backfire is the appearance of the flame in the neck or body of the blowpipe and which rapidly extinguishes itself.

A flashback is the appearance of the flame beyond the blowpipe body into the hose and even the gauge, with subsequent explosion. It can be prevented by fitting a flashback arrestor.

Backfiring may occur at the pipe through several causes:

(1) Insufficient pressure for the nozzle being used. This can be cured by increasing the pressure on the gauge.

(2) Metal particles adhering to the nozzle. The nozzle can be freed of particles by rubbing it on a leather or wooden surface. (The gases should be first shut off and then relit.)

(3) The welder touching the plate or weld metal with the nozzle. In this case the gases should be shut off and then relit.

(4) Overheating of the blowpipe. A can of water should be kept near so as to cool the nozzle from time to time, especially when using a large flame. Oxygen should be allowed to pass slowly through the

nozzle, when immersed in the water, to prevent the water entering the inside of the blowpipe.

(5) Should the flame backfire into the mixing chamber with a squealing sound, and a thin plume of black smoke be emitted from the nozzle, serious damage will be done to the blowpipe unless the valves are immediately turned off. This fault may be caused by particles having lodged inside the pipe, or even under the regulating valves. The pipe should be thoroughly inspected for defects before being relit.

In the event of a backfire, therefore, immediately shut off the acetylene cylinder valve, and then the oxygen, before investigating the cause.

Blowpipe nozzles should be cleaned by using a soft copper or brass pin. They should be taken off the shank and cleaned from the inside, as this prevents enlarging the hole. A clean nozzle is essential, since a dirty one gives an uneven-shaped flame with which good welds are impossible to make. Special sets of reamers can be obtained for this work.

Flashback arrestor

Note. Flashback arrestors should be fitted to all welding equipment. A flashback occurs when the flame moves from the blow torch into the supply system against the flow of the gases. It is potentially dangerous and its effects can range from sooty deposits in the blowpipe and hoses to a fire in the gauge or cylinder often accompanied by squealing or popping noises. It is generally due to incorrect operating practice such as overheated blowpipe and wrong pressures.

The automatic flashback arrestor is made for acetylene, propane, hydrogen, and oxygen and is generally connected to the regulator outlet (Fig. 6.3*c*) and prevents flame movement into gauges and cylinder, causing regulator damage and even cylinder fires.

A dense sintered stainless steel plate filter of up to 100 micron mesh (1) in Fig. 6.8*a* arrests the flame and is designed to quench the flame from even the most violent explosion of the gas mixture. The large surface area of the filter does not offer much resistance to the forward flow of the gas. The pressure-sensitive cut-off valve (2) automatically cuts off the supply of gas to the blow torch preventing sustained flashback and it will remain closed until it is set manually, a spring plus the gas pressure holding it closed. There should never be any attempt to reset the valve until the cause of the flashback has been thoroughly investigated and put right and it cannot be reset until the pressure is taken off the system, so close the cylinder valve to cut off the supply at source. To reset, on the model shown, the reset pin (attached to the gauge by a chain to prevent loss) is inserted into the gas inlet orifice, centred against the valve and pushed hard home resetting the valve mechanism. On some models there is a visible lever which is actuated when the arrestor cut-off valve operates. This lever actuates the mechanism but should never be reset until the cause of the flashback is determined and put to rights.

After several actuations, carbon deposits may interfere with the correct functioning of the arrestor so it should be exchanged. Arrestors are generally used up to oxygen pressures of 10 bar, propane of 5 bar and acetylene of 1.5 bar.

Operation. The spring-loaded valve (or piston) is held by a low melting point solder against the tension of a spring. If a flashback occurs the temperature rising to 120–150 °C melts the solder and the non-return valve is forced onto its seat and prevents further reverse flow of gas. There is little restriction of gas flow under normal working conditions but in the case of the piston being activated the whole unit must be replaced (Figs. 6.8*b*, *c*).

Fig. 6.8. (*a*) Section through resettable flashback arrestor.

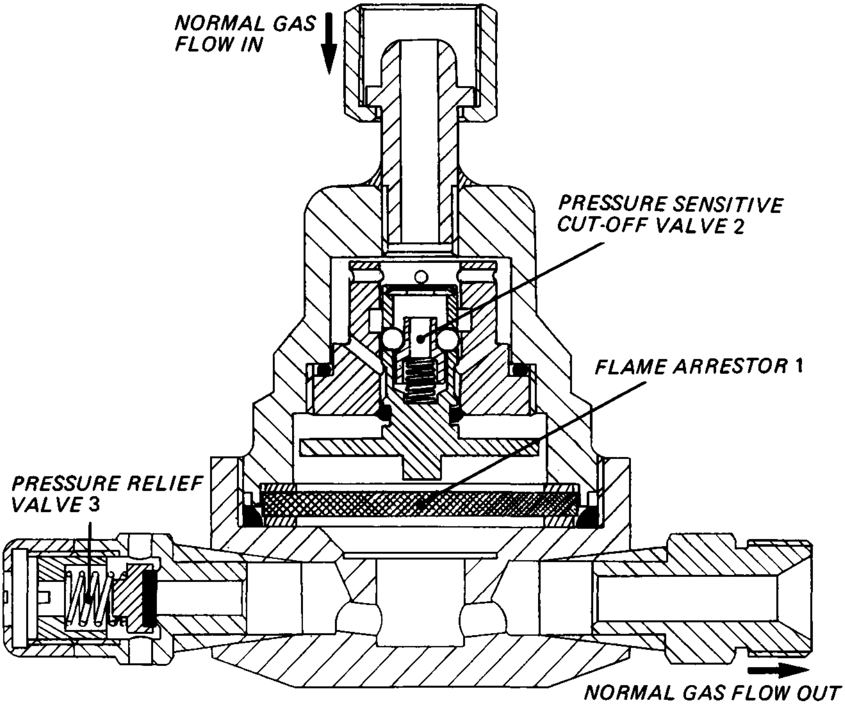

Technique of welding

Before attempting any actual welding operations, the beginner should acquire a sense of fusion and a knowledge of blowpipe control. This can be obtained by running lines of fusion on thin-gauge steel plate.

The flame is regulated to the neutral condition and strips of 1.6 or 2 mm steel plate are placed on firebricks on the welding bench. Holding the blowpipe at approximately 60° to the plate, with the inner blue cone near the metal surface, and beginning a little from the right-hand edge of the sheet, the metal is brought to the melting point and a puddle formed with a rotational movement of the blowpipe. Before the sheet has time to melt through into a hole, the pipe is moved steadily forward, still keeping the steady rotating motion, and the line of fusion is made in a straight line. This exercise should be continued on various thicknesses of thin-gauge plate until an even line is obtained, and the underside shows a regular continuous bead, indicating good penetration, the student thus acquiring a sense of fusion and of blowpipe control.

In the following pages various methods of welding techniques are considered, and it would be well to state at this point what constitutes a good weld, and what features are present in a bad weld.

Fig. 6.8. (*b*) Heavy duty high flow flashback arrestor, thermally operated. (*c*) Section through flashback arrestor.

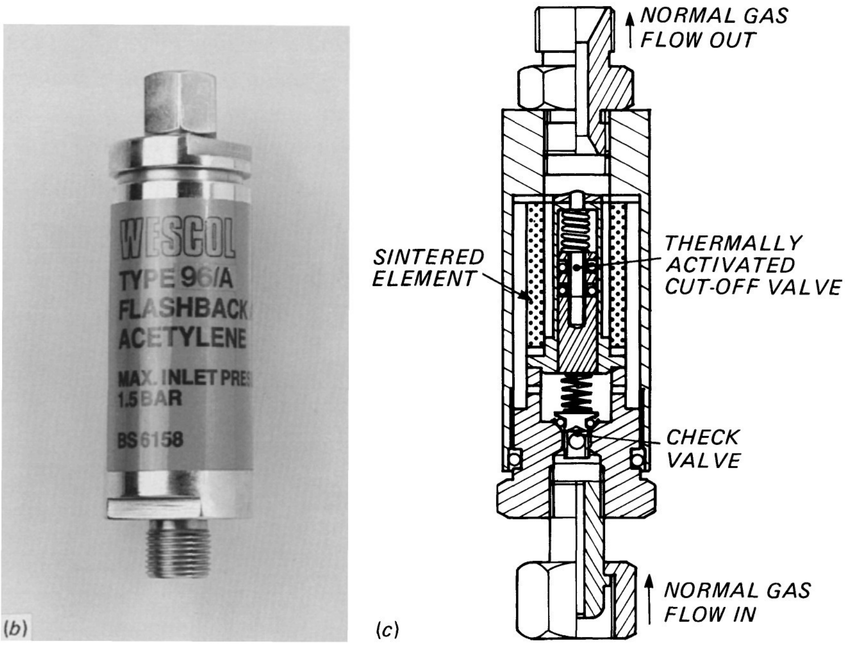
WESCOL
TYPE 96/A
FLASHBACK
ACETYLENE
MAX. INLET PRES
1.5 BAR
BS6158
NORMAL GAS FLOW OUT
SINTERED ELEMENT
THERMALLY ACTIVATED CUT-OFF VALVE
CHECK VALVE
NORMAL GAS FLOW IN

(*b*) (*c*)

Fig. 6.9*a* indicates the main features of a good fusion weld, with the following features:

(*a*) Good fusion over the whole side surface of the V.
(*b*) Penetration of the weld metal to the underside of the parent plate.
(*c*) Slight reinforcement of the weld above the parent plate.
(*d*) No entrapped slag, oxide, or blowholes.

Fig. 6.9*b* indicates the following faults in a weld:

(*a*) Bad flame manipulation and too large a flame has caused molten metal to flow on to unfused plate, giving no fusion (i.e. adhesion).
(*b*) Wrong position of work, incorrect temperature of molten metal, and bad flame manipulation has caused slag and oxide to be entrapped and channels may be formed on each line of fusion, causing undercutting.
(*c*) The blowpipe flame may have been too small, or the speed of welding too rapid, and this with lack of skill in manipulation has caused bad penetration.

(N.B. Reinforcement on the face of a weld *will not* make up for lack of penetration.)

Methods of welding

The following British Standards apply to this section: BS 1845, *Filler alloys for brazing*; BS 1723, *Brazing*; *Part 1 Specification for brazing, Part 2 Guide to brazing*; BS 1724, *Bronze welding by gas*; BS 1453, *Filler materials for gas welding (ferritic steels, cast iron, austenitic stainless steels, copper and copper alloys, aluminium and aluminium alloys and magnesium alloys)*; BS 1821, *Class 1 oxy-acetylene welding of ferritic steel*

Fig. 6.9

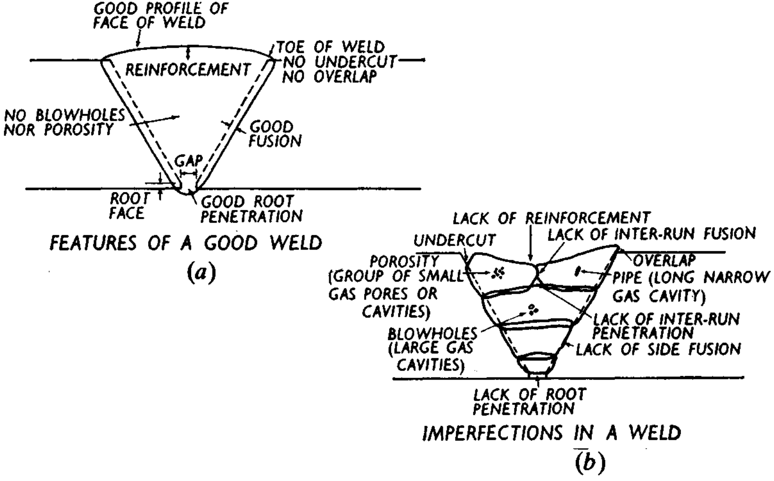
GOOD PROFILE OF FACE OF WELD
TOE OF WELD
REINFORCEMENT
NO UNDERCUT
NO OVERLAP
NO BLOWHOLES NOR POROSITY
GOOD FUSION
GAP
ROOT FACE
GOOD ROOT PENETRATION
LACK OF REINFORCEMENT
UNDERCUT
LACK OF INTER-RUN FUSION
POROSITY (GROUP OF SMALL GAS PORES OR CAVITIES)
OVERLAP
PIPE (LONG NARROW GAS CAVITY)
LACK OF INTER-RUN PENETRATION
BLOWHOLES (LARGE GAS CAVITIES)
LACK OF SIDE FUSION
LACK OF ROOT PENETRATION

FEATURES OF A GOOD WELD
(*a*)

IMPERFECTIONS IN A WELD
(*b*)

pipework for the carrying of fluids; BS 2640, *Class 2 oxy-acetylene welding of carbon steel pipework for carrying fluids.*

Leftward or forward welding

This method is used nowadays for welding steel plate under 6.5 mm thick and for welding non-ferrous metals. The welding rod precedes the blowpipe along the seam, and the weld travels from right to left when the pipe is held in the right hand. The inner cone of the flame, which is adjusted to the neutral condition, is held near the metal, the blowpipe making an angle of 60 to 70° with the plate, while the filler rod is held at an angle of 30 to 40°. This gives an angle of approximately 90° between the rod and the blowpipe. The flame is given a rotational, circular, or side-to-side motion, to obtain even fusion on each side of the weld. The flame is first played on the joint until a molten pool is obtained and the weld then proceeds, the rod being fed into the molten pool and not melted off by the flame itself. If the flame is used to melt the rod itself into the pool, it becomes easy to melt off too much and thus reduce the temperature of the molten pool in the parent metal to such an extent that good fusion cannot be obtained. Fig. 6.10 will make this clear.

The first exercise in welding with the filler rod is done with the technique just described and consists of running lines of weld on 1.6 or 2 mm plate, using the filler rod. Butt welds of thin plate up to 2.4 mm can be made by flanging the edges and melting the edges down. When a uniform weld is obtained, with good penetration, the exercises can be repeated on plate up to 3.2 mm thick, and butt welds on this thickness attempted. Above 3.2 mm thick, the plates are bevelled, chamfered, or V'd to an angle of 80 to 90° (Fig. 6.11). The large area of this V means that a large quantity of weld metal is required to fill it. If, however, the V is reduced to less than 80°, it is found

Fig. 6.10. Leftward welding.

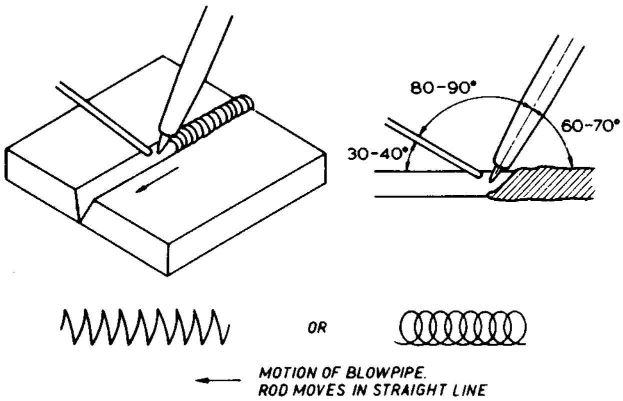

that as the V becomes narrower the blowpipe flame tends to push the molten metal from the pool, forward along onto the unmelted sides of the V, resulting in poor fusion or adhesion. This gives an unsound weld, and the narrower the V the greater this effect.

As the plate to be welded increases in thickness, a larger nozzle is required on the blowpipe, and the control of the molten pool becomes more difficult; the volume of metal required to fill the V becomes increasingly greater, and the size of nozzle which can be used does not increase in proportion to the thickness of the plate, and thus welding speed decreases. Also with thicker plates the side-to-side motion of the blowpipe over a wide V makes it difficult to obtain even fusion on the sides and penetration to the bottom, while the large volume of molten metal present causes considerable expansion. As a result it is necessary to weld thicker plate with two or more layers if this method is used. From these considerations it can be seen that above 6.4 mm plate the leftward method suffers from several drawbacks. It is essential, however, that the beginner should become efficient in this method before proceeding to the other methods, since for general work, including the non-ferrous metals (see later), it is the most used.

Edge preparation (Letters refer to Fig. 6.11)	Thickness of plate (mm)	Nozzle size (mm)	Oxygen and acetylene pressure (bar)	Oxygen and acetylene gas consumption (l/h)
(*a*)	0.9–1.6	0.9–1	0.14	28
		1.2–2	0.14	57
(*b*)	2.4–3.0	2–3	0.14	86
		2.6–5	0.14	140
(*c*)	3.0–4.0	3.2–7	0.14	200
		4.0–10	0.21	280

Fig. 6.11. Leftward welding: edge preparations.

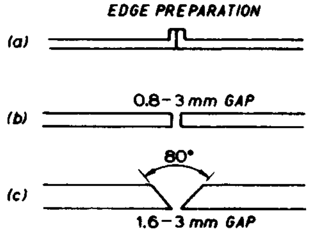

The preparation of various thicknesses of plate for butt joints by the leftward method is given in the table accompanying Fig. 6.11.

Rightward welding

This method was introduced some years ago to compete with electric arc welding in the welding of plate over 4.8 mm thick, since the leftward method has the disadvantage just mentioned on welding thick plate. This method has definite advantages over the leftward method on thick plate, but the student should be quite aware of its limitations and use it only where it has a definite advantage.

In this method the weld progresses along the seam from left to right, the rod following the blowpipe. The rod is given a rotational or circular motion, while the blowpipe moves in practically a straight line, as illustrated in Fig. 6.12. The angle between blowpipe and rod is greater than that used in the leftward method.

When using this method good fusion can be obtained without a V up to 8 mm plate. Above 8 mm the plates are prepared with a 60° V, and since the blowpipe has no side motion the heat is all concentrated in the narrow V, giving good fusion. The blowpipe is pointing backwards towards the part that has been welded and thus there is no likelihood of the molten metal being pushed over any of the unheated surface, giving poor fusion.

A larger blowpipe nozzle is required for a given size plate than in leftward welding, because the molten pool is controlled by the pipe and rod but the pipe has no side to side motion. This larger flame gives greater welding speed, and less filler rod is used in the narrower V. The metal is under good control and plates up to 16 mm thick can be welded in one pass. Because the blowpipe does not move except in a straight line, the molten metal is agitated very little and excess oxidation is prevented. The flame playing on the metal just deposited helps to anneal it, while the smaller volume of molten metal in the V reduces the amount of expansion. In

Fig. 6.12. Rightward welding.

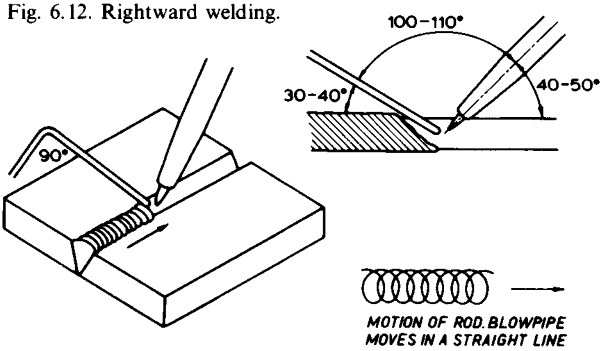

addition, a better view is obtained of the molten pool, resulting in better penetration.

It is essential, however, in order to ensure good welds by this method, that blowpipe and rod should be held at the correct angle, the correct size nozzle and filler rod used, and the edges prepared properly (Fig. 6.13). The rod diameter is about half the thickness of the plate being welded up to 8 mm plate, and half the thickness + 0.8 mm when welding V'd plate. The blowpipe nozzle is increased in size from one using about 300 litres per hour, with the leftward method, to one using about 350 litres per hour, when welding 3.2 mm plate. If too large diameter filler rods are used, they melt too slowly causing poor penetration, and poor fusion. Small rods melt too quickly and reinforcement of the weld is difficult. Rightward welding has no advantage on plates below 6.4 mm thick and is rarely used below this thickness, the leftward method being preferred.

The advantages of the rightward method on thicker plate are:

(1) Less cost per metre run due to less filler rod being used and increased speed.
(2) Less expansion and contraction.
(3) Annealing action of the flame on the weld metal.
(4) Better view of the molten pool, giving better control of the weld.

See later for 'all-position rightward welding'.

Edge preparation (Letters refer to Fig. 6.13)	Thickness of plate (mm)	Nozzle size (mm)	Oxygen and acetylene pressure (bar)	Oxygen and acetylene gas consumption (1/h)
(*a*)	4.8–8.2	5.5–13	0.28	370
		6.5–18	0.28	520
		8.2–25	0.42	710
(*b*)	8.2–15	10–35	0.63	1000
		13–45	0.35 (heavy duty mixer)	1300

Fig. 6.13. Rightward welding: edge preparations.

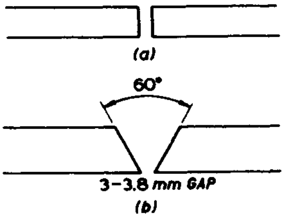

Vertical welding

The preparation of the plate for welding greatly affects the cost of the weld, since it takes time to prepare the edges, and the preparation given affects the amount of filler rod and gas used. Square edges need no preparation and require a minimum of filler rod. In leftward welding square edges are limited to 3.2 mm thickness and less. In vertical welding, up to 4.8 mm plate can be welded with no V'ing with the single-operator method while up to 16 mm plate can be welded with no V'ing with the two-operator method, the welders working simultaneously on the weld from each side of the plate. The single-operator method is the most economical up to 4.8 mm plate.

The single-operator method (up to 4.8 mm plate) requires more skill in the control of the molten metal than in downhand welding. Welding is performed either from the bottom upwards, and the rod precedes the flame as in the leftward method, or from the top downwards, in which case the metal is held in place by the blowpipe flame. This may be regarded as the rightward method of vertical welding, since the flame precedes the rod down the seam. In the upward method the aim of the welder is to use the weld metal which has just solidified as a 'step' on which to place the molten pool. A hole is first blown right through the seam, and this hole is maintained as the weld proceeds up the seam, thus ensuring correct penetration and giving an even back bead.

In the vertical welding of thin plate where the edges are close together, as for example in a cracked automobile chassis, little filler rod is needed and the molten pool can be worked upwards using the metal from the sides of the weld. Little blowpipe movement is necessary when the edges are close together, the rod being fed into the molten pool as required. When the edges are farther apart, the blowpipe can be given the usual semicircular movement to ensure even fusion of the sides.

From Fig. 6.14 it will be noted that as the thickness of the plate increases, the angle of the blowpipe becomes much steeper.

When welding downwards much practice is required (together with the correct size flame and rod), in order to prevent the molten metal from falling downwards. This method is excellent practice to obtain perfect control of the molten pool.

Double-operator vertical welding

The flames of each welder are adjusted to the neutral condition, both flames being of equal size. To ensure even supply of gas to each pipe the blowpipes can be supplied from the same gas supply. It is possible to use much smaller jets with this method, the combined consumption of which is less than that of a single blowpipe on the same thickness plate. Blowpipes and rods are held at the same angles by each operator, and it is well that a third person should check this when practice runs are being done. To avoid fatigue a sitting position is desirable, while, as for all types of vertical welding, the pipes and tubes should be as light as possible. Angles of blowpipes and rods are shown in Fig. 6.15.

This method has the advantage that plate up to 16 mm thick can be welded without preparation, reducing the gas consumption and filler rod used, and cutting out the time required for preparation. When two operators are welding 12.5 mm plate, the gas used by both is less than 50%

Fig. 6.14. Single-operator vertical welding.

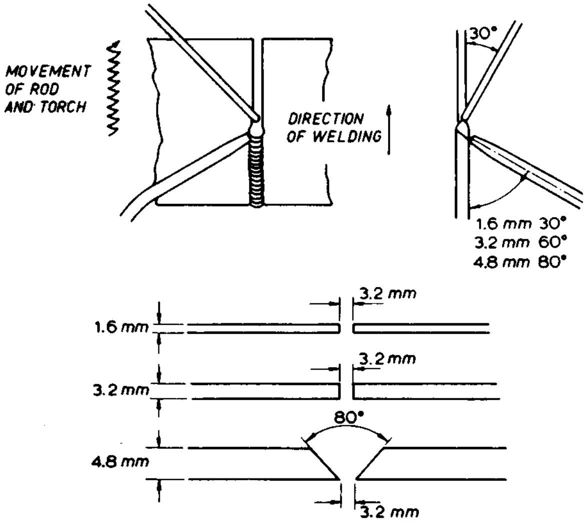
MOVEMENT OF ROD AND TORCH
DIRECTION OF WELDING
30°
1.6 mm 30°
3.2 mm 60°
4.8 mm 80°
3.2 mm
1.6 mm
3.2mm
3.2 mm
80°
4.8 mm
3.2 mm

Fig. 6.15. Double-operator vertical welding.

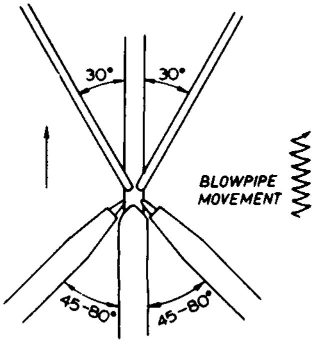
30°
30°
BLOWPIPE MOVEMENT
45-80°
45-80°

of the total consumption of the blowpipe when welding the same thickness by the downhand rightward method. Owing to the increased speed of welding and the reduced volume of molten metal, there is a reduction in the heating effect, which reduces the effects of expansion and contraction.

Overhead welding

Overhead welding is usually performed by holding the blowpipe at a very steep angle to the plate being welded. The molten pool is entirely controlled by the flame and by surface tension, and holding the flame almost at right angles to the plate enables the pool to be kept in position.

Difficulty is most frequently found in obtaining the correct amount of penetration. This is due to the fact that as sufficient heat to obtain the required penetration is applied, the molten pool becomes more fluid and tends to become uncontrollable. With correct size of flame and rod, however, and practice, this difficulty can be entirely overcome and sound welds made. Care should also be taken that there is no undercutting along the edges of the weld.

A comfortable position and light blowpipe and tubes are essential if the weld is to be made to any fair length, as fatigue of the operator rapidly occurs in this position and precludes the making of a good weld.

The positions of blowpipe and rod for the leftward techniques are shown in Fig. 6.16*a*, while Fig. 6.16*b* shows their relative positions when the rightward method is used. The rightward method is generally favoured, but

Fig. 6.16. (*a*) Leftward overhead welding. The flame is used to position the molten metal. (*b*) Rightward overhead welding. Blowpipe has little motion. Rod moves criss-cross from side to side.

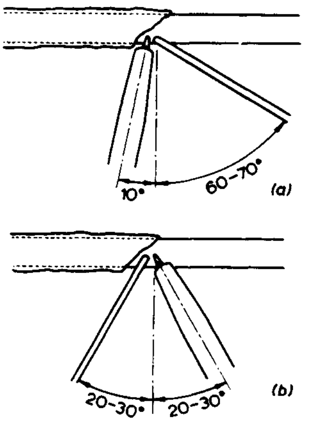

it must again be stressed that considerable practice is required for a welder to become skilled in overhead welding.

Lindewelding

This method was devised by the Linde Co. of the United States for the welding of pipelines (gas and oil), and for this type of work it is very suitable. Its operation is based on the following facts:

1. When steel is heated in the presence of carbon, the carbon will reduce any iron oxide present, by combining with the oxygen and leaving pure iron. The heated surface of the steel then readily absorbs any carbon present.
2. The absorption of carbon by the steel lowers the melting point of the steel (e.g. pure iron melts at 1500 °C, while cast iron with $3\frac{1}{2}\%$ carbon melts at 1130 °C).

In Lindewelding the carbon for the above action is supplied by using a carburizing flame. This deoxidizes any iron oxide present and then the carbon is absorbed by the surface layers, lowering their melting point. By using a special rod, a good sound weld can be made in this way at increased speed. The method is almost exclusively used on pipe work and is performed in the downhand position only.

Block welding

Block welding is a method especially applicable to steel pipes and thick walled tubes in which the weld is carried out to the full depth of the joint in steps. This can more easily be understood by reference to the figure. The first run is laid giving good penetration for as great a length as is convenient, say *AB* (Fig. 6.17).

The second run is now started a little way short of *A* and finished short of *B*, as *CD*. The third run is laid in a similar manner from *E* to *F*, if required, and so on for the full depth of the weld required. We thus have a series of ledges or platforms at the beginning and end of the weld. The welding is now continued with the first run from *B* with full penetration. The second run starts at *D* and has the ledge *BD* deposited before it gets to where the

Fig. 6.17

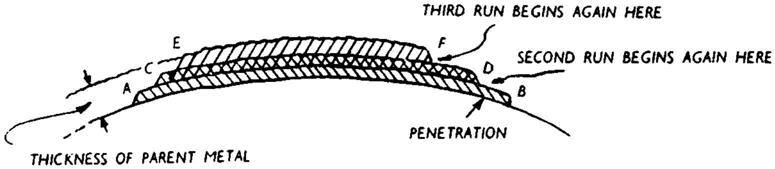
THIRD RUN BEGINS AGAIN HERE
E
F
C
D
SECOND RUN BEGINS AGAIN HERE
A
B
PENETRATION
THICKNESS OF PARENT METAL

first run started. Similarly with the third run, which starts at *F*. Upon completing the weld in the case of a pipe, the first run finishes at *A*, the second one at *C* and the third one at *E*, giving a good anchorage on to the previous run.*

Horizontal-vertical fillet welding

In making fillet welds (Fig. 6.18), care must be taken that, in addition to the precautions taken regarding fusion and penetration, the vertical plate is not undercut as in Fig. 6.19*b*, and the weld is not of a weakened section. A lap joint may be regarded as a fillet. No difficulty will be experienced with undercutting, since there is no vertical leg, but care should be taken not to melt the edge of the lapped plate.

The blowpipe and rod must be held at the correct angles. Holding the

Fig. 6.18. Types of fillet joints.

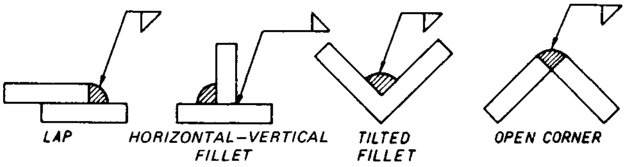
LAP
HORIZONTAL-VERTICAL FILLET
TILTED FILLET
OPEN CORNER

Fig. 6.19. (*a*)

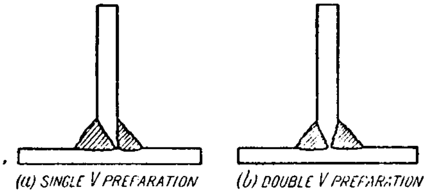
(a) SINGLE V PREPARATION

(b) DOUBLE V PREPARATION

Fig. 6.19. (*b*)

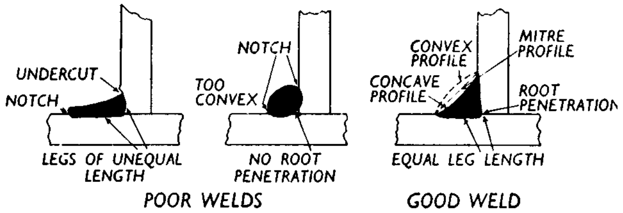
UNDERCUT
NOTCH
LEGS OF UNEQUAL LENGTH
NOTCH
TOO CONVEX
NO ROOT PENETRATION
POOR WELDS
CONVEX PROFILE
CONCAVE PROFILE
MITRE PROFILE
ROOT PENETRATION
EQUAL LEG LENGTH
GOOD WELD

---

* Refer also to BS 1821 and 2640, *Class I and class II oxy-acetylene welding of steel pipelines and assemblies for carrying fluids.*

flame too high produces undercutting, and the nozzle of the cone should be held rather more towards the lower plate, since there is a greater mass of metal to be heated in this than in the vertical plate (Fig. 6.19*a* and *b*).

Figure 6.20 *a* and *b* show the angles of the blowpipe and rod, the latter being held at a steeper angle than the blowpipe. Fillet welding requires a larger size nozzle than when butt welding the same section plate, owing to the greater amount of metal adjacent to the weld. Because of this, multi-jet blowpipes can be used to great advantage for fillet welding. The single (Fig. 6.19*a*) preparation is used for joints which are subjected to severe loading, while the double V preparation (Fig. 6.19*b*) is used for thick section plate when the welding can be done from both sides. The type of preparation therefore depends entirely on the service conditions, the unprepared joint being quite suitable for most normal work.

All-position rightward welding

This method can be used for vertical, overhead and horizontal–vertical positions, the blowpipe preceding the rod as for downhand welding. For vertical welding the blowpipe is held 10° below the horizontal line (welding upwards) while the rod is held alongside the pipe at 45–60° to the vertical plate. Overhead welding is done similarly. The advantages are similar to those of the downhand position but considerable practice is required to become proficient. See Fig. 6.21.

For the preparation and welding of steel pipes the student should refer to BS 1821, *Class I steel pipelines* and BS 2640, *Class II steel pipelines.*

Fig. 6.20. (*a*) Fillet weld.

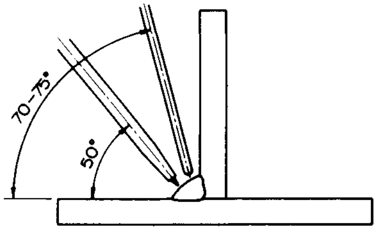

Fig. 6.20. (*b*) Fillet weld.

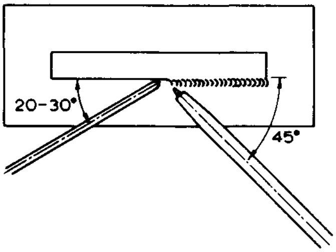

Fig. 6.21. All-position rightward welding.

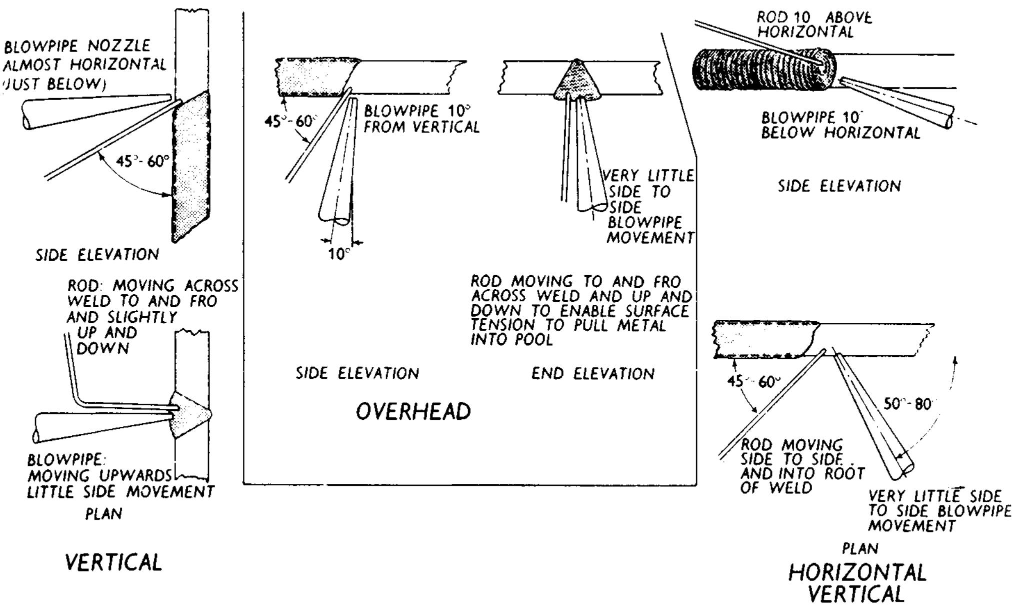
BLOWPIPE NOZZLE ALMOST HORIZONTAL (JUST BELOW)
45°-60°
SIDE ELEVATION
ROD: MOVING ACROSS WELD TO AND FRO AND SLIGHTLY UP AND DOWN
BLOWPIPE: MOVING UPWARDS LITTLE SIDE MOVEMENT
PLAN
VERTICAL
45°-60°
BLOWPIPE 10° FROM VERTICAL
10°
VERY LITTLE SIDE TO SIDE BLOWPIPE MOVEMENT
ROD MOVING TO AND FRO ACROSS WELD AND UP AND DOWN TO ENABLE SURFACE TENSION TO PULL METAL INTO POOL
SIDE ELEVATION
END ELEVATION
OVERHEAD
ROD 10° ABOVE HORIZONTAL
BLOWPIPE 10° BELOW HORIZONTAL
SIDE ELEVATION
45°-60°
50°-80°
ROD MOVING SIDE TO SIDE AND INTO ROOT OF WELD
VERY LITTLE SIDE TO SIDE BLOWPIPE MOVEMENT
PLAN
HORIZONTAL VERTICAL

Steel filler rods for oxy-acetylene welding

Rod type	Composition %					Application
	C	Si	Mn	Ni	Cr	
Mild steel (copper coated)	0.1	—	0.6	0.25	—	General utility low-carbon-steel rod for low and mild carbon steel and wrought iron. UTS 386 N/mm^2 hardness 120 Brinell, melting point 1490 °C.
Medium carbon steel (copper coated)	0.25–0.3	0.3–0.5	1.3–1.6	0.25	0.25	For medium carbon steels, high strength with toughness, UTS 552 N/mm^2, hardness 150 Brinell, melting point 1400 °C.
Pipe-welding steel	0.1–0.2 Plus Al, Ti, Zr as deoxidizers up to 0.15% max.	0.1–0.35	1.0–1.6	—	—	Low carbon steel for high-strength welds in steel pipes. UTS 492 N/mm^2, hardness 145 Brinell, melting point 1450°.

Note. Phosphorus and sulphur 0.04% max. for all types.

Cast iron welding

Cast iron, because of its brittleness, presents a different problem in welding from steel. We may consider three types – grey, white and malleable.

The grey cast iron is softer and tougher than the white, which is hard and brittle. The good mechanical properties of grey cast iron are due to the presence of particles of free carbon or graphite, which separate out during slow cooling. When the cooling is rapid, it is impossible for the cementite (iron carbide) to decompose into ferrite and graphite; hence the structure consists of masses of cementite embedded in pearlite, this giving the white variety of cast iron with its hardness and brittleness.

The other constituents of cast iron are silicon, sulphur, manganese and phosphorus. Silicon is very important, because it helps to increase the formation of graphite, and this helps to soften the cast iron. Manganese makes the casting harder and stronger. It has a great affinity for sulphur and, by combining with it, prevents the formation of iron sulphide, which makes the metal hard and brittle. Phosphorus reduces the melting point and increases the fluidity. If present in a greater proportion than 1% it tends to increase the brittleness. Sulphur tends to prevent the formation of graphite and should generally not exceed 0.1%. It may be added to enable the outer layers to have a hard surface (chill casting), while the body of the casting is still kept in the grey state.

The aim of the welder should be always, therefore, to form grey cast iron in the weld (Fig. 6.22).

Fig. 6.22. Oxy-acetylene fusion weld in cast iron. × 100.

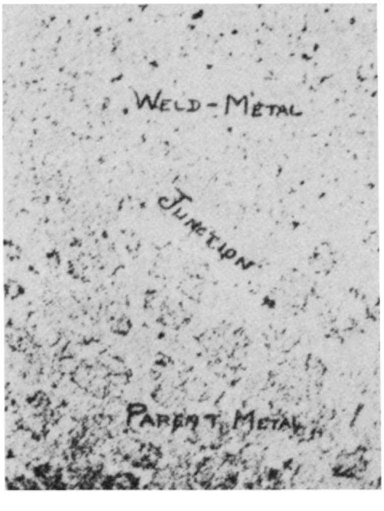

Preparation. The edges are V'd out to 90° on one side only up to 10 mm thickness and from both sides for greater thickness. Pre-heating is essential, because not only does it prevent cracking due to expansion and contraction, but by enabling the weld to cool down slowly, it causes grey cast iron to be formed instead of the hard white unmachinable deposit which would result if the weld cooled off rapidly. Pre-heating may be done by blowpipe, forge or furnace, according to the size of the casting.

Blowpipe, flame, flux and rod. The neutral flame is used, care being taken that there is not the slightest trace of excess oxygen which would cause a weak weld through oxidation, and that the metal is not overheated. The inner cone should be about 3–4 mm away from the molten metal. If it touches the molten metal, hard spots will result. The flux (of alkaline borates) should be of good quality to dissolve the oxide and prevent oxidation, and it will cause a coating of slag to form on the surface of the weld, preventing atmospheric oxidation. Ferro-silicon and super-silicon rods, containing a high percentage of silicon, are the most suitable rods to use.

Technique. The welding operation should be performed on the dull red-hot casting with the rod and blowpipe at the angles shown in Fig. 6.23 and done in the leftward manner. The rod is dipped into the flux at intervals, using only enough flux to remove the oxide. It will be noticed that, as the flux is added, the metal flows more easily and looks brighter, and this gives a good indication as to when flux is required. The excessive use of flux causes blowholes and a weak weld. The rod should be pressed down well into the weld, removing a good percentage of the slag and oxide. Very little motion of the blowpipe is required and the rod should not be stirred round continuously, as this means the formation of more slag with more danger of entrapping it in the weld. Cast-iron welds made by beginners often have brittle parts along the edges, due to their not getting under the oxide and floating it to the top.

Fig. 6.23. Cast iron welding.

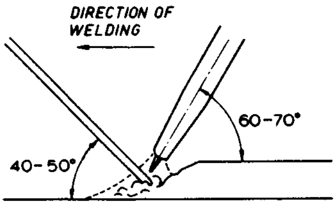

After treatment. The slag and oxide on the surface of the finished weld can be removed by scraping and brushing with a wire brush, but the weld should not be hammered. The casting is then allowed to cool off very slowly, either in the furnace or fire, or if it has been pre-heated with the flame, it can be put in a heap of lime, ashes or coke, to cool. Rapid cooling will result in a hard weld with possibly cracking or distortion.

Malleable cast iron welding

This is unsuitable for welding with cast-iron rods because of its structure. If attempted it invariably results in hard, brittle welds having no strength. The best method of welding is with the use of bronze rods, and is described below.

Braze welding and bronze welding

Braze welding is the general term applied to the process in which the metal filler wire or strip has a lower melting point than that of the parent metal, using a technique similar to that used in fusion welding, except that the parent metal is not melted nor is capillary action involved as in brazing. When the filler metal is made of a copper-rich alloy the process is referred to as bronze welding.

By the use of bronze filler wires such as those given in the table on p. 290, a sound weld can be made in steel, wrought iron, cast iron and copper and between dissimilar metals such as steel and copper, at a lower temperature than that usually employed in fusion welding. The bronze filler metal, which has a lower melting point than the parent metal, is melted at the joint so as to flow and 'wet' the surfaces of the joint to form a sound

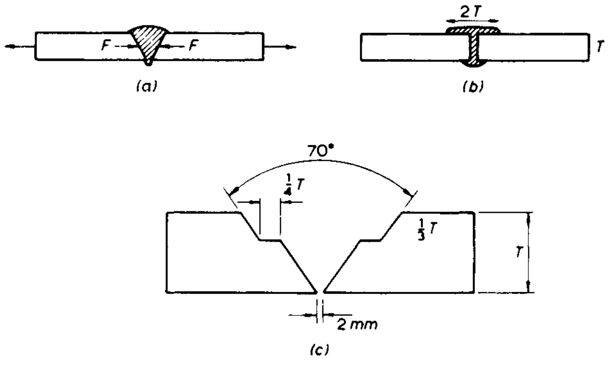

Fig. 6.24. When specimen (*a*) is subjected to a tensile pull as shown the faces *F* of the joint are partly in shear. (*c*) 'Shear V' preparation.

bond without fusion of the parent metal. There is less thermal disturbance than with fusion welding due to the lower temperatures involved and the process is simple and relatively cheap to perform. Unlike capillary brazing (see pp. 317–19) the strength of the joint is not solely dependent upon the areas of the surfaces involved in the joint but rather upon the tensile strength of the filler metal. A bronze weld has relatively great strength in shear, and joints are often designed to make use of this (Fig. 6.24*a*, *b* and *c*).

The final strength of the welded joint depends upon the bond between filler metal and parent metal so that thorough cleanliness of the joints and immediate surroundings is essential to ensure that the molten filler metal should flow over the areas and 'wet' them completely without excessive penetration of the parent metal, and there should be freedom from porosity. Care, therefore, should be taken to avoid excessive overheating. Figs. 6.25 and 6.26 show various joint designs for plate and tube.

If bronze filler wires are used on alloys such as brasses and bronzes the melting points of wire and parent metal are so nearly equal that the result is a fusion weld.

General method of preparation for bronze welding. All impurities such as scale, oxide, grease, etc., should be removed, as these would prevent the bronze wetting the parent metal. The metal should be well cleaned on both upper and lower faces for at least 6 mm on each side of the joint, so that the bronze can overlap the sides of the joint, running through and under on the lower face.

Bronze welding is unlike brazing in that the heat must be kept as local as possible by using a small flame and welding quickly. The bronze must flow in front of the flame for a short distance only, wetting the surface, and by having sufficient control over the molten bronze, welding may be done in

Fig. 6.25. Typical joint designs for bronze welding (sheet and plate).

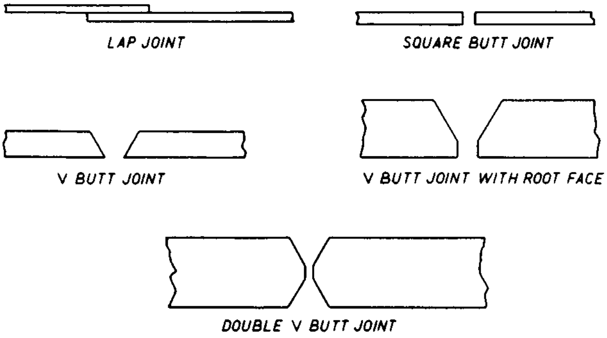

the overhead position. Too much heat prevents satisfactory wetting. We will now consider the bronze welding of special metals using a flux of a mixture of alkaline fluoride and borax, but bronze filler rods are also supplied flux-coated.

Cast iron

Bevel the edges to a 90° V, round off the sharp edges of the V, and clean the casting well. Pre-heating may be dispensed with unless the casting is of complicated shape, and the welding may often be done without dismantling the work. If pre-heating is necessary it should be heated to

British Standards designations, compositions and recommended usage of filler alloys for bronze welding

Alloy designation		Composition % (by weight) Deleterious impurities, e.g. Al and Pb, are each restricted to 0.3% max.				
BS 1453	BS 1845 (Group CZ)			Recommended for usage on	Approx. melting point (°C)	Applications
C2	CZ6	Cu 57.00 Si 0.20 Zn balance Sn optional	to 63.00 to 0.50 to 0.50 max.	Copper Mild steel	875–895	A silicon-bronze used for copper sheet and tube mild steel and line production applications.
C4	CZ7	Cu 57.00 Si 0.15 Mn 0.05 Fe 0.10 Zn balance Sn optional	to 63.00 to 0.30 to 0.25 to 0.50 to 0.50 max.	Copper Cast iron Wrought iron	870–900	Similar to C2. (CZ6).
C5	CZ8	Cu 45.00 Si 0.15 Ni 8.00 Zn balance Sn optional Mn optional Fe optional	to 53.00 to 0.50 to 11.00 to 0.50 max. to 0.50 max. to 0.50 max.	Mild steel Cast iron Wrought iron	970–980	A nickel-bronze for bronze welding steel and malleable iron, building up worn surfaces and welding Cu–Zn–Ni alloys of similar composition.
C6	—	Cu 41.00 Si 0.20 Ni 14.00 Zn balance Sn optional Mn optional Fe optional	to 45.00 to 0.50 to 16.00 to 1.00 max. to 0.20 max. to 0.30 max.	Cast iron Wrought iron	— —	Similar to C5. (CZ8).

450 °C, and on completion cooling should be as slow as possible as in the fusion welding of cast iron (Fig. 6.27).

Blowpipe, flame and rod. The blowpipe nozzle may be about two sizes smaller than for the same thickness steel plate, and the flame is adjusted so as to give a slight excess of oxygen. If a second deposit is to be run over the first, the flame is adjusted to a more oxidizing condition still for the subsequent runs, the inner cone being usually only about $\frac{3}{4}$ of its neutral length. The best flame condition can easily be found by trial. Suitable filler

Fig. 6.26. Typical joint designs for bronze welding (tube).

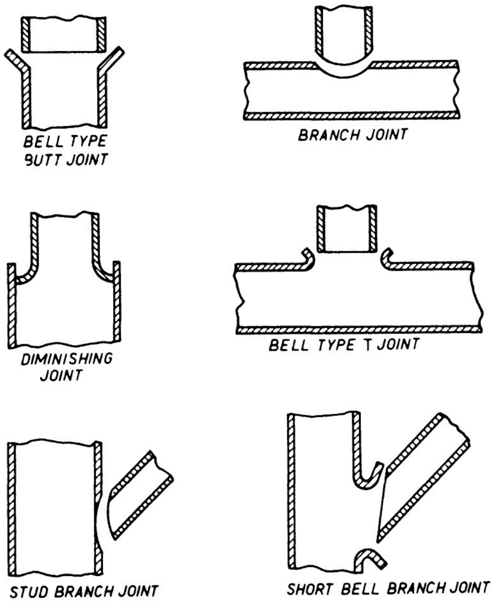
BELL TYPE
BUTT JOINT
BRANCH JOINT
DIMINISHING
JOINT
BELL TYPE T JOINT
STUD BRANCH JOINT
SHORT BELL BRANCH JOINT

Fig. 6.27. Bronze weld in cast iron.

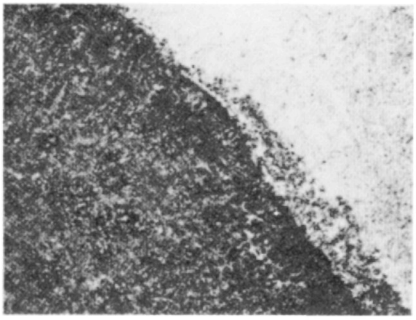

alloys are given in the table, those containing nickel giving greater strength, the bronze flux being of the borax type.

Technique. The leftward method is used with the rod and blowpipe held as in Fig. 6.28, the inner cone being held well away from the molten metal. The rod is wiped on the edges of the cast iron and the bronze wets the surface. It is sometimes advisable to tilt the work so that the welding is done uphill, as shown in Fig. 6.28. This gives better control. Do not get the work too hot.

Vertical bronze welding of cast iron can be done by the two-operator method, and often results in saving of time, gas and rods and reduces the risk of cracking and distortion.

The edges are prepared with a double 90° V and thoroughly cleaned for 12 mm on each side of the edges. The blowpipes are held at the angle shown in Fig. 6.29*a*. Blowpipe and rod are given a side to side motion, as indicated in Fig. 6.29*b* as the weld proceeds upwards, so as to tin the surfaces.

Malleable cast iron

The bronze welding of malleable castings may be stated to be the only way to ensure any degree of success in welding them. Both types (blackheart and whiteheart) can be welded satisfactorily in this way, since the heat of the process does not materially alter the structure. The method is the same as for cast iron, using nickel bronze rods (C5) and a borax-type flux.

Steel

In cases where excessive distortion must be avoided, or where thin sections are to be joined to thick ones, the bronze welding of steel is often

Fig. 6.28. Bronze welding cast iron.

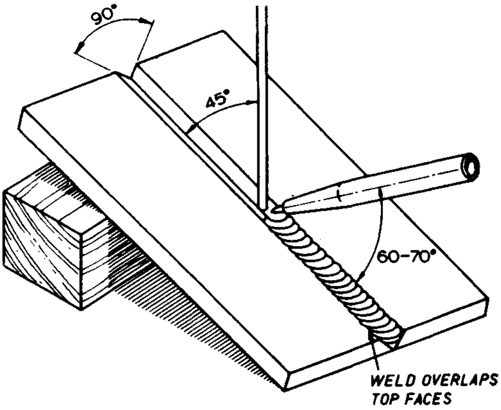

used, the technique being similar to that for cast iron, except, of course, that no pre-heating is necessary.

Galvanized iron

This can be easily bronze welded, and will result in a strong corrosion-resisting joint, with no damage to the zinc coating. If fusion welding is used, the heat of the process would of course burn the zinc (or galvanizing) off the joint and the joint would then not resist corrosion.

Preparation, flame and rod. For galvanized sheet welding, the edges of the joint are tack welded or held in a jig and smeared with a silver–copper flux. Thicker plates and galvanized pipes are bevelled 60–80° and tacked to position them. The smallest possible nozzle should be used (for the sheet thickness) and the flame adjusted to be slightly oxidizing. Suitable filler alloys are given in the table (p. 290) as for steel.

Technique. No side to side motion of the blowpipes is given, the flame being directed on to the rod, so as to avoid overheating the parent sheet. The rod is stroked on the edges of the joint so as to wet them. Excessive flux *must* be washed off with hot water.

Fig. 6.29. (*a*) Two-operator vertical bronze welding of cast iron. (*b*) Showing motion of pipe and rod.

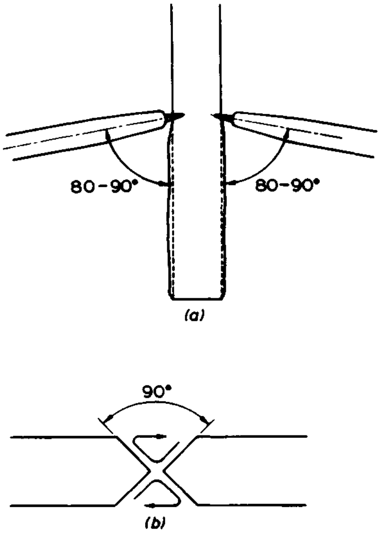

Copper

Tough pitch copper can be readily bronze welded owing to the much lower temperature of the process compared with fusion welding (Fig. 6.30).

Preparation, flame and rod. Preparation is similar to that for cast iron. Copper tubes can be bell mouthed (Fig. 6.31). Special joints are available for multiple branches. The blowpipe nozzle should be small and will depend on the size of the work or the diameter of the pipe to be welded, and should be chosen so that the bronze flows freely but no overheating occurs. The flame should be slightly oxidizing, and if a second run is made, it should be adjusted slightly to be more oxidizing still (inner cone about $\frac{3}{4}$ of its normal neutral length). Suitable filler alloys are given in the table (p. 290) and are used with a bronze-base flux.

Fig. 6.30. Bronze weld in copper. × 250.

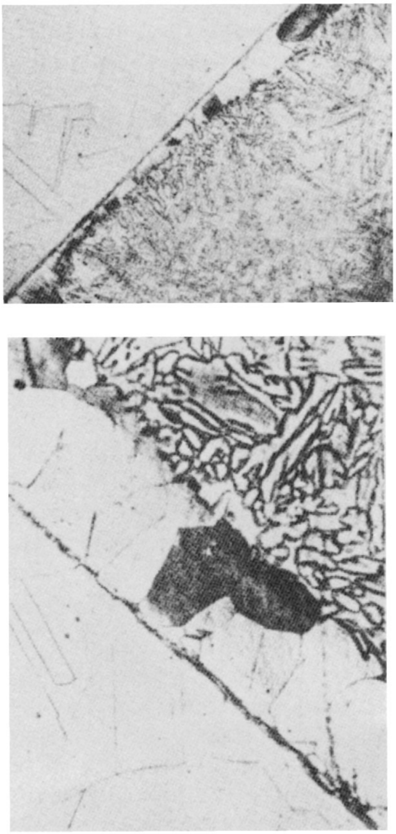

Technique. The method is similar to that for cast iron, and the final difference between the bronze-welded and brazed joint is that the former has the usual wavy appearance of the oxy-acetylene weld, while the latter has a smooth appearance, due to the larger area over which the heat was applied. The bronze joint is, of course, much stronger than the brazed one.

Brasses and zinc-containing bronzes

Since the filler rod now melts at approximately the same temperature as the parent metal, this may now be called fusion welding. When these alloys are heated to melting point, the zinc is oxidized, with copious evolution of fumes of zinc oxide, and if this continued, the weld would be full of bubble holes and weak (Fig. 6.32). This can be prevented by using an oxidizing flame, so as to form a layer of zinc oxide over the molten metal, and thus prevent further formation of zinc oxide and vaporization.

Preparation. The edges of the faces of the joint are cleaned and prepared as usual, sheets above 3.2 mm thickness being V'd to 90°. Flux can be applied by making it into a paste, or by dipping the rod into it in the usual manner, or a flux-coated rod can be used.

Flame and rod. Suitable filler alloys are given in the table (p. 290), while a brass rod is used for brass welding, the colour of the weld then being similar to that of the parent metal. Owing to the greater heat conductivity, a larger size jet is required than for the same thickness of steel plate. The flame is adjusted to be oxidizing, as for bronze welding cast iron, and the exact flame condition is best found by trial as follows. A small test piece of the brass or bronze to be welded is heated with a neutral flame and gives off copious fumes of zinc oxide when molten. The acetylene is now cut down until no more fumes are given off. If any blowholes are seen in the metal on solidifying, the acetylene should be reduced slightly further. The inner cone will now be about half its normal neutral length. Too

Fig. 6.31

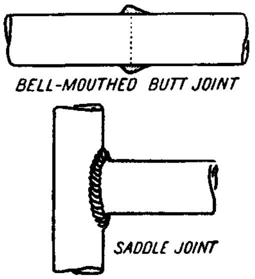

Fig. 6.32. (*a*) Unsatisfactory brass weld made with neutral flame, Unetched × 2.5. (*b*) Brass weld made with insufficient excess of oxygen. Unetched. (*c*) Correct brass weld made with adequate excess of oxygen. Unetched. × 2.5.

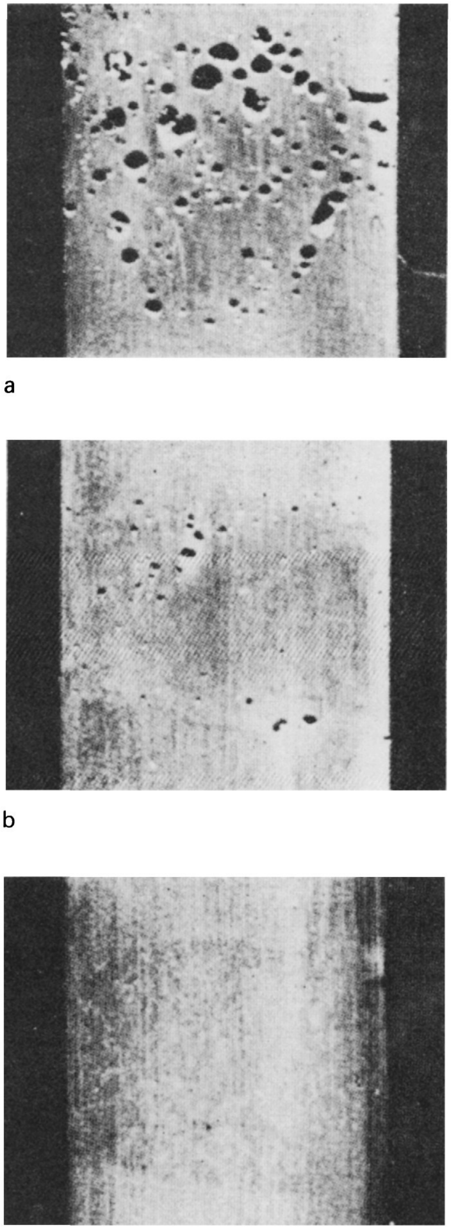

much oxygen should be avoided, as it will form a thick layer of zinc oxide over the metal and make the filler rod less fluid.

The weld is formed in the 'as cast' condition, and hammering improves its strength. Where 60/40 brass rods have been used the weld should be hammered while hot, while if 70/30 rods have been used the weld should be hammered cold and finally annealed from dull red heat.

Tin bronze

Tin bronze cannot be welded using an oxidizing flame. Special rods and fluxes are available, however, with which good welds can be made. Urgent repairs may be safely carried out using a *neutral* flame and a silicon–bronze rod with borax-type flux.

Gilding metal

For the weld to be satisfactory on completion, the weld metal must have the same colour as the parent metal. Special rods of various compositions are available, so that the colours will 'match'.

Aluminium bronze

Aluminium bronze can be welded using a filler rod of approximately 90% copper, 10% aluminium (C13, BS 2901 Pt 3), melting point 1040 °C, a rod also suitable for welding copper, manganese bronze and alloy steels where resistance to shock, fatigue and sea-water corrosion is required. The aluminium bronze flux (melting point 940 °C) can be mixed with water to form a paste if required.

Preparation. The edges of the joint should be thoroughly cleaned by filing or wire brushing to remove the oxide film which is difficult to dissolve.

Up to 4.8 mm thickness no preparation is required – just a butt joint with gap. Above 4.8 mm the usual V preparation is required and a double V above 16 mm thick. Sheets should not be clamped for welding, as this tends to cause the weld to crack, they must be allowed to contract freely on cooling, and it is advisable to weld a seam continuously and not make starts in various places.

Flame and rod. A neutral flame is usually used – any excess of acetylene tends to produce hydrogen with porosity of the weld, while excess oxygen causes oxidation. Flame size should be carefully chosen according to the thickness of the plate – too small a flame causes the weld metal to solidify too quickly while there is a danger of burning through with too large a flame.

The filler rod should be a little thicker (0.8 mm) than the sheet to be welded to avoid overheating, and it should be added quickly to give complete penetration without deep fusion.

Technique. The leftward method is used with a steep blowpipe angle (80°) to start the weld, this being reduced to 60–70° as welding progresses. The parent metal should be well pre-heated prior to starting welding and during welding a large area should be kept hot to avoid cracks. The rod should be used with a scraping motion to clean the molten pool and remove any entrapped gas.

In welding the single-constituent (or α phase) aluminium bronze, i.e. 5–7% Al, 93% Cu, the weld metal should be deposited in a single run or at most two runs to avoid intergranular cracks. Since the metal is hot short in the range 500–700 °C it should cool quickly through the range, and should not be peened. Cold peening is sometimes an advantage. The two-constituent (α and β) or duplex aluminium bronzes contain 10% aluminium. They have a wide application, are not as prone to porosity, and are easier to weld than the 7% Al type, and also their hot short range is smaller.

After treatment. Stresses can be relieved by heat treating at low temperature, and any required heat treatment can be carried out as required after welding.

Copper welding

Tough pitch copper (that containing copper oxide) is difficult to weld, and so much depends on the operator's skill that it is advisable to specify deoxidized copper for all work in which welding is to be used as the method of jointing. Welds made on tough pitch copper often crack along the edge of the weld if they are bent (Fig. 6.33*a*), showing that the weld is unsound due to the presence of oxide, often along the lines of fusion. A *good* copper weld (Fig. 6.33*b*), on the other hand, can be bent through 180° without cracking and can be hammered and twisted without breaking. This type of copper weld is strong and sound, free from corrosion effects, and is eminently satisfactory as a method of jointing.

Preparation. The surfaces are thoroughly cleaned and the edges are prepared according to the thickness, as shown in Fig. 6.34. In flanging thin sheet the height of the flange is about twice the plate thickness and the flanges are bent at right angles. Copper has a high coefficient of expansion, and it is necessary therefore to set the plates diverging at the rate of 3–4 mm per 100 mm run, because they come together so much on being welded. Since copper is weak at high temperatures, the weld should be well supported if possible and an asbestos sheet between the weld and the backing strip of steel prevents loss of heat.

Tacking to preserve alignment is not advised owing to the weakness of the copper tacks when hot. When welding long seams, tapered spacing clamps or jigs should be used to ensure correct spacing of the joint, care

Fig. 6.33. (*a*) Poor copper weld. Crack developed when bent. (*b*) Oxy-acetylene weld in deoxidized copper. × 100.

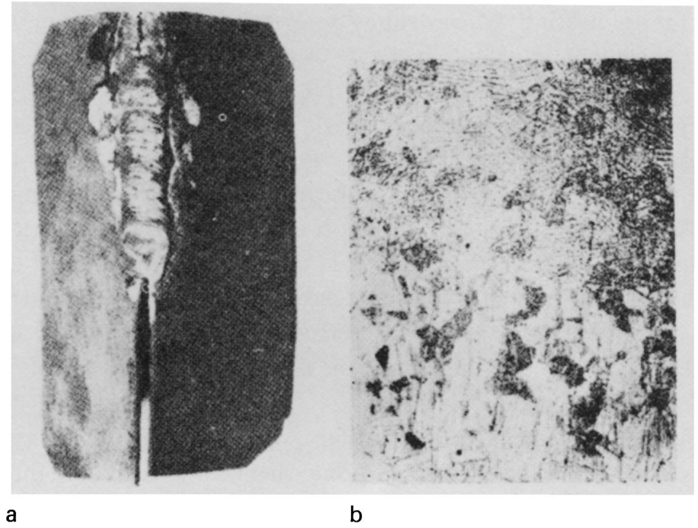

Fig. 6.34. Preparation of copper plates for welding.

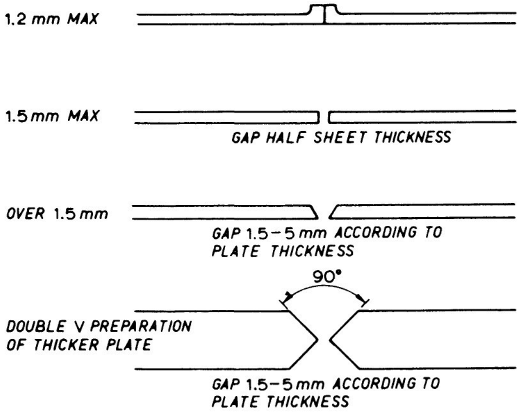
1.2 mm MAX
1.5 mm MAX
GAP HALF SHEET THICKNESS
OVER 1.5 mm
GAP 1.5–5 mm ACCORDING TO PLATE THICKNESS
90°
DOUBLE V PREPARATION OF THICKER PLATE
GAP 1.5–5 mm ACCORDING TO PLATE THICKNESS

being taken that these do not put sufficient pressure on the edges to indent them when hot. A very satisfactory method of procedure is to place a clamp *C* at the centre of the seam and commence welding at a point say about one-third along the seam.

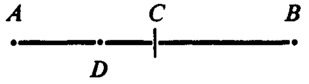

Welding is performed from *D* to *B* and then from *D* to *A*.

Because of the high conductivity of copper it is essential to pre-heat the surface, so as to avoid the heat being taken from the weld too rapidly. If the surface is large or the metal thick, two blowpipes must be used, one being used for pre-heating. When welding pipes they may be flanged or plain butt welded, while T joints can be made as saddles.

Blowpipe, flame and rod. A larger nozzle than for the same thickness of steel should be used and the flame adjusted to be neutral or very slightly carbonizing. Too much oxygen will cause the formation of copper oxide and the weld will be brittle. Too much acetylene will cause steam to form, giving a porous weld, therefore close the acetylene valve until the white feathery plume has almost disappeared. The welding rod should be of the deoxidized type, and many alloy rods, containing deoxidizers and other elements such as silver to increase the fluidity, are now available and give excellent results.

The weld may be made without flux, or a flux of the borax type used. Proprietary fluxes containing additional chemicals greatly help the welding operation and make it easier.

Technique. The blowpipe is held at a fairly steep angle, as shown in Fig. 6.35, to conserve the heat as much as possible. Great care must be taken to keep the tip of the inner cone 6–9 mm away from the molten metal, since the weld is then in an envelope of reducing gases, which prevent oxidation. The weld proceeds in the leftward manner, with a slight sideways motion of the blowpipe. Avoid agitating the molten metal, and do not remove the rod from the flame but keep it in the molten pool. Copper may

Fig. 6.35. Copper welding.

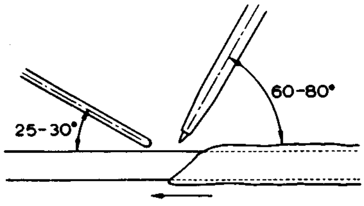

also be welded by the rightward method, which may be used when the filler rod is not particularly fluid. The technique is similar to that for rightward welding of mild steel, with the flame adjusted as for leftward welding of copper.

Welding can also be performed in the vertical position by either single- or double-operator method, the latter giving increased welding speed.

After treatment. Light peening, performed while the weld is still hot, increases the strength of the weld. The effect of cold hammering is to consolidate the metal, but whether or not it should be done depends on the type of weld and in general is not advised. Annealing, if required on small articles, can be carried out by heating to 600–650 °C.

Aluminium welding

The welding of aluminium, either pure or alloyed, presents no difficulty (Fig. 6.36) provided the operator understands the problems which must be overcome and the technique employed.

The oxide of aluminium (alumina Al_2O_3), which is always present as a surface film and which is formed when aluminium is heated, has a very high melting point, much higher than that of aluminium, and if it is not removed it would become distributed throughout the weld, resulting in weakness and brittleness. A good flux, melting point 570 °C, is necessary to dissolve this oxide and to prevent its formation.

Fig. 6.36. Oxy-acetylene weld in aluminium. × 45.

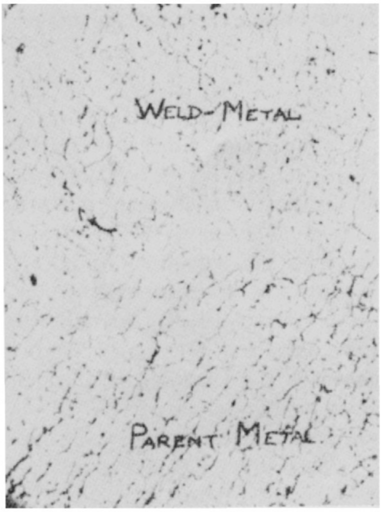

Aluminium and its alloys

Preparation. The work should be cleaned of grease and brushed with a wire brush. Sheets below 1 mm thickness can be turned up at right angles (as for mild steel) and the weld made without a filler rod. Over 3.2 mm thick the edges should have a 90° V and over 6 mm thick a double 90° V. Tubes may be bevelled if thick or simply butted with a gap between them. It is advisable always to support the work with backing strips of asbestos or other material, to prevent collapse when welding. Aluminium, when near its melting point, is extremely weak, and much trouble can be avoided by seeing that no collapsing can occur during the welding operation.

Blowpipe, flame, flux and rod. The flame is adjusted to have a very slight excess of acetylene and then adjusted to neutral, and the rod of pure aluminium or 5% silicon–aluminium alloy should be a little thicker than the section to be welded. A good aluminium flux must be used and should be applied to the rod as a varnish coat, by heating the end of the rod, dipping it in the flux and letting the tuft, which adheres to the rod, run over the surface for about 150 mm of its length. This ensures an even supply. Too much flux is detrimental to the weld.

Technique. The angles of the blowpipe and rod are shown in Fig. 6.37 (a slightly larger angle between the blowpipe and rod than for mild steel), and the welding proceeds in the leftward manner, keeping the inner cone well above the molten pool. The work may be tacked at about 150 mm intervals to preserve alignment, or else due allowance made for the joint coming in as welded, like mild steel. As the weld progresses and the metal becomes

Fig. 6.37. (*a*) Aluminium flat welding. Rightward technique may also be employed using approximately the same angles of blowpipe and rod. (*b*) Aluminium welding by the double-operator method.

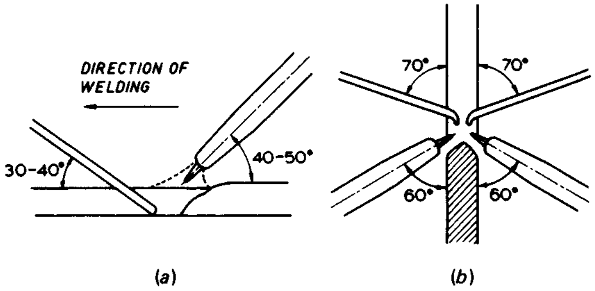

hotter, the rate of welding increases, and it is usual to reduce the angle between blowpipe and weld somewhat to prevent melting a hole in the weld. Learners are afraid of applying sufficient heat to the joint as a rule, because they find it difficult to tell exactly when the metal is molten, since it does not change colour and is not very fluid. When they do apply enough heat, owing to the above difficulty, the blowpipe is played on one spot for too long a period and a hole is the result.

If the rightward technique is used the blowpipe angle is 45° and the rod angle 30–40°. Distortion may be reduced when welding sheets, and the flame anneals the deposited metal.

The two-operator vertical method may be employed (as for cast iron) on sheets above 6 mm thickness, the angle of the blowpipes being 50–60° and the rods 70–80°. This method gives a great increase in welding speed (Fig. 6.37)

After treatment. All the corrosive flux must be removed first by washing and scrubbing in hot soapy water. This can be followed by dipping the article in a 5% nitric acid solution followed by a washing again in hot water.

Where it is not possible to get at the parts for scrubbing, such as in tanks, etc., the following method of removal is suitable. Great care, however, should be taken when using the hydrofluoric acid as it is dangerous, and rubber gloves should be worn, together with a face mask.

A solution is made up as follows in a heavy duty polythene container.

Nitric acid – 100 g to 1 litre water.
Hydrofluoric acid – 6 g to 1 litre water.

The nitric acid is added to the water first, followed by the hydrofluoric acid.

Articles immersed in this solution for about ten minutes will have all the flux removed from them, and will have a uniformly etched surface. They should then be rinsed in cold water followed by a hot rinse, the time of the latter not exceeding three minutes, otherwise staining may occur.

Hammering of the completed weld greatly improves and consolidates the structure of the weld metal, and increases its strength, since the deposited metal is originally in the 'as cast' condition and is coarse grained and weak. Annealing may also be performed if required.

Aluminium alloy castings and sheets

The process for the welding of castings is very similar to that for the welding of sheet aluminium. See Volume 1 for explanation of alloy coding letters.

Preparation. The work is prepared by V'ing if the section is thicker than 3 mm, and the joint is thoroughly cleaned of grease and impurities. Castings such as aluminium crank-cases are usually greasy and oily (if they have been in service), the oil saturating into any crack or break which may have occurred. If the work is not to be pre-heated, this oil *must* be removed. It may be washed first in petrol, then in a 10% caustic soda solution, and this followed by a 10% nitric acid or sulphuric acid solution. A final washing in hot water should result in a clean casting. In normal cases in which pre-heating is to be done, filing and a wire brush will produce a clean enough joint, since the pre-heating will burn off the remainder.

Aluminium alloys. Recommended filler rods

Alloys	Composition % (remainder aluminium)	Filler rod (old designation in brackets)
Casting alloys BS 1490		
LM2	0.7–2.5 Cu, 9–11.5 Si	4047A, 10–12% Si (NG2)
LM4	2–4 Cu, 4–6 Si	4047A or 4043A, 4.5–6% Si (NG2 or NG21)
LM5	3–6 Mg	5356, 4.6–5.5 Mg (NG6)
LM6	10–13 Si, 0.5 Mn	4043, 415–6.0 Si (NG21)
LM9	10–13 Si, 0.3–0.7 Mn	4043 or 4047 (NG21 or NG2)
LM18	4.5–6 Si	4043 (NG21)
LM20	10–13 Si, 0.4 Cu	4047 (NG2)
Wrought alloys BS 1470–1477		
1080A	99.8 Al	1080A (G1B)
1050A	99.5 Al	1080A (G1C)
3103 (N3)	1–1.5 Mn	3103 (NG3)
6063 (H9)	0.4–0.9 Mg, 0.3–0.7 Mn	4043 or 5056A (NG21 or NG6)
061 (H20)	0.15 C 0.4 Cu, 0.8–1.2 Mg, 0.4–0.8 Si, 0.2–0.8 Mn, 0.15–0.35 Cr	4043 or 5056A (NG21 or NG6)
6082 (H30)	0.5–1.2 Mg, 0.7–1.3 Si, 0.4–1.0 Mn	4043 or 5056A (NG21 or NG6)

If the casting is large or complicated, pre-heating should be done as for cast iron.* In any case it is advantageous to heat the work well with the blowpipe flame before commencing the weld.

---

* Large complicated castings can be pre-heated to about 400 °C, smaller castings to 300–350 °C and small castings to 250–300 °C. No visible change in the appearance of the aluminium occurs at these temperatures.

Blowpipe, flame, rod and flux. The blowpipe is adjusted as for pure aluminium and a similar flame used. The welding rod should preferably be of the same composition as the alloy being welded (see table on p. 290) but for general use a 5% silicon–aluminium rod is very satisfactory. This type of rod has strength, ductility, low shrinkage, and is reasonably fluid. A 10% silicon–aluminium rod is used for high silicon castings, while 5% copper–aluminium rods are used for the alloys containing copper, such as Y alloy, and are very useful in automobile and aircraft industries. The deposit from this type of rod is harder than from the other types.

When welding the Al–Mg alloys the oxide film consists of both aluminium and magnesium oxide making the fluxing more difficult so that as the magnesium proportion increases welding may become more difficult. Alloys containing more than 2½% Mg, e.g. 5154A (N5) and 5183 (N8) are difficult to weld and require considerable experience as do the high strength alloys 6061 (H15) and 6082 (H30). The inert gas arc processes are to be preferred for welding these alloys.

Since there is also a loss of Mg in the welding process note that the filler rod recommended has a greater Mg content than the parent plate. The flux used is similar to that for pure aluminium and its removal must be carried out in the same way.

Technique. The welding is carried out as for aluminium sheet, and the cooling of the casting after welding must be gradual.*

After-treatment. After welding, the metal is in the 'as cast' condition and is weaker than the surrounding areas of parent metal, and the structure of the deposited metal may be improved by hammering. The area near the welded zone, however, is annealed during the welding process and failure thus often tends to occur in the area alongside the weld, and not in the weld itself. In the case of heat-treatable alloys, the welded zone can be given back much of its strength by first lightly hammering the weld itself and then heat-treating the whole of the work.

For this to be quite successful it is essential that the weld should be of the same composition as the parent metal. If oxidation has occurred, however, this will result in a weld metal whose structure will differ from that of the parent metal and the weld will not respond to heat treatment. Since many of this type of alloy are 'hot short', cracking may occur as a result of the welding process.

---

* When repairing cracked castings, any impurities which appear in the molten pool should be floated to the top, using excess flux if necessary.

If sheets are anodized, the welding disturbs the area and changes its appearance. Avoidance of overheating, localizing the heat as much as possible, and hammering, will reduce this disturbance to a minimum, but heat treatment will make the weld most inconspicuous.

Welding of nickel and nickel alloys

The alloys include MONEL® (nickel–copper) and INCONEL® (nickel–chromium) alloys and modifications of these constituents to give variations in properties.

Oxy-acetylene welding is used only for welding Nickel 200 alloy, certain MONEL® alloys, INCO® and INCONEL® alloys. Welding of nickel–chromium–cobalt alloys (NIMONIC® alloy) is not usually performed. (Consult the manufacturers for up to date details.)

Preparation. Sheets of thinner than 1 mm can be bent up through an angle of about 75°, as shown in Fig. 6.38, and the edges melted together. The ridge formed by welding can then be hammered flat. Sheets thicker than 1 mm are bevelled with the usual 90° V and butted together. For corner welds on sheet less than 1 mm the corners are flanged as shown, while for thicker plate the weld is treated as an open corner joint (see Fig. 6.38). Castings should be treated as for cast iron.

In tube welding, preparation should be an 80° V with no root gap.

Blowpipe, flame, flux and rod, High-purity acetylene as supplied from DA cylinders is required, and in general the blowpipe nozzle size is the same as that for the same thickness mild steel. For nickel 200 a size larger can be used. No flux is required for nickel 200 and for Incoloy DS the use of flux is

Fig. 6.38. Preparation of plates.

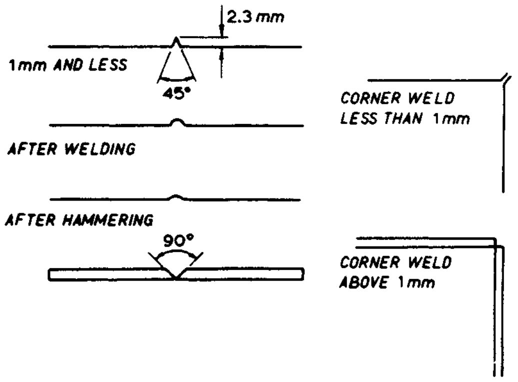

optional and special fluxes are available for the other alloys. Flux containing boron should not be used with alloys containing chromium as it tends to cause hot cracking in the weld. The joint is tacked in position without flux. The flux is made into a thin paste and painted onto both sides of the joint and the filler rod and allowed to dry before welding. Fused flux remaining should be removed by wire brushing and unfused flux with hot water. Flux remaining (after welding the nickel chromium alloys) can be removed by treatment with a solution of equal parts of nitric acid and water for 15–30 minutes followed by washing with water. Flux removal is important when high-temperature applications are involved, as the flux may react with the metal.

Technique. Leftward technique is used. Weaving and puddling of the molten pool should be avoided as agitation of the pool causes porosity due to the absorption of gases by the high-nickel alloys, and the filler rod should be kept within the protective envelope of the flame to prevent oxidation. Keep the flame tip above the pool but for MONEL® alloy K500 let it just touch the pool. Nickel 200 melts sluggishly. INCONEL® alloy 600, NIMONIC® alloy 75 and the BRIGHTRAY® alloys are more fluid while MONEL® alloy 400 and INCOLOY® DS alloy flow easily. A slightly carburizing soft flame (excess acetylene) should be used for the nickel and nickel–copper alloys, whilst for the chromium-containing alloys the flame should be a little more carburizing.

There are no pronounced ripples on the weld surface. They should be smooth without roughness, burning or signs or porosity.

Examples of filler metals

Material	
Nickel alloys 200, 201	Nickel filler metal 61
MONEL® alloys 400, R405, K500	MONEL® filler metal 60
INCONEL® alloy 600	INCONEL® filler metal 62
INCONEL® alloys 600, 601	INCONEL® filler metal 82
INCOLOY® alloys 800, 800HT	
INCO® alloy 330	
INCOLOY® DS	N/C (nickel–chromium) filler metal

When welding tube with 80° feather edge preparation and no root gap, the first run is made with no filler rod, the edges being well fused together to give an even penetration bead, followed by filler runs. If the work is rotatable, welding in the two o'clock position gives good metal control. If the joint is fixed, vertical runs should be made first downwards followed by a run upwards.

Nickel clad steel*

This steel is produced by the hot rolling of pure nickel sheet on to steel plate, the two surfaces uniting to form a permanent bond. It gives a material having the advantage of nickel, but at much less cost than the solid nickel plate. It can be successfully welded by the oxy-acetylene process.

Preparation. The bevelled 90° butt joint is the best type. For fillet welds it is usual to remove (by grinding) the nickel cladding on the one side of the joint, as shown in Fig. 6.39, so as to ensure a good bond of steel to steel.

Technique. The weld is made on the steel side first, using a mild steel rod and the same technique as for mild steel. The nickel side is then welded, using a nickel rod and a slightly reducing flame as for welding monel and nickel. The penetration should be such that the nickel penetrates and welds itself into the steel weld.

Stainless steels

Stainless steels of the martensitic and austenitic class can be welded by the oxy-acetylene process.

Preparation. Thin gauges may be flanged at 90° as for mild steel sheet and the edges fused together. Thicker sections are prepared with the usual 90° V and the surfaces cleaned of all impurities. The coefficient of expansion of the 18/8 austenitic steels is about 50% greater than that of mild steel, and consequently the sheets should be set diverging much more than for mild

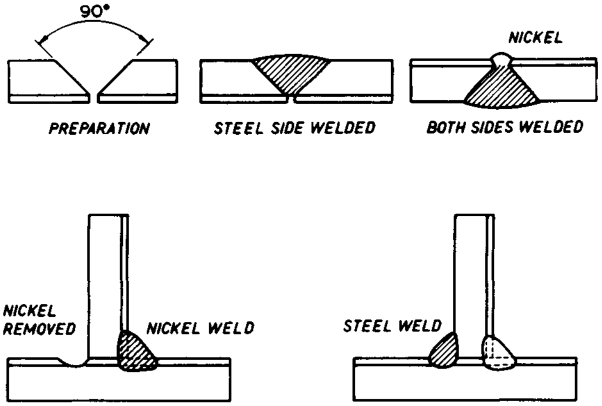

Fig. 6.39. Double fillet weld. Nickel clad steel.

---

* Also stainless clad steel

steel to allow for them coming together during welding. An alternative method is to tack the weld at both ends first, and then for sheets thinner than 1 mm tack at 25 mm intervals, while for thicker sheets the tacks can be at 50 mm intervals. Cooling clamps are advised for the austenitic steels. The thermal conductivity is less than for mild steel, and thus the heat remains more localized. Unless care is taken, therefore, the beginner tends to penetrate the sheet when welding thin gauges.

Flame. The flame should be very slightly carburizing (excess acetylene), there being the smallest trace of a white plume around the inner cone. The flame should be checked from time to time to make sure that it is in this condition, since any excess oxygen is fatal to a good weld, producing porosity, while too great an excess of acetylene causes a brittle weld.

Rod and flux. The welding rod must be of the same composition as the steel being welded, the two types normally used being the 18/8 class, and one having a higher nickel and chromium content and suitable for welding steels up to a 25/12 chrome–nickel content. Since there is considerable variety in the amounts of the various elements added to stainless steels to improve their physical properties, such as molybdenum and tungsten in addition to titanium or niobium, it is essential that the analysis of the parent metal be known so that a suitable rod can be chosen. The steel makers and electrode makers will always co-operate in this matter. The rod should be of the same gauge or slightly thicker than the sheet being welded. No flux is necessary, but if difficulty is experienced with some steels in penetration, a special flux prepared for these steels should be used, and can be applied as a paste mixed with water.

Technique. Welding is performed in the leftward manner, and the flame is played over a larger area than usual because of the low thermal conductivity. This lessens the risk of melting a hole in the sheet. The tip of the inner cone is kept very close to the surface of the molten pool and the welding performed exactly as for the leftward welding of mild steel. No puddling should be done and the least possible amount of blowpipe motion used.

After treatment. Martensitic stainless steels should be heated to 700 to 800°C and then left to air cool. The non-stabilized steels require heat treatment after welding if they are to encounter corrosive conditions, otherwise no heat treatment is necessary. The weld decay-proof stabilized steels and those with a very low carbon content require no heat treatment after welding.

The original silver-like surface can be restored to stainless steels which have scaled or oxidized due to heat by immersing in a bath made up as follows: sulphuric acid 8%, hydrochloric acid 6%, nitric acid 1%, water 85%. The temperature should be 65 °C and the steel should be left in for 10 to 15 minutes, and then immersed for 15 minutes or until clean in a bath of 10% nitric acid, 90% water, at 25 °C.

Other baths made up with special proprietary chemicals with trade names will give a brighter surface. These chemicals are obtainable from the makers of the steel.

Stainless iron

The welding of stainless iron results in a brittle region adjacent to the weld. This brittleness cannot be completely removed by heat treatment, and thus the welding of stainless iron cannot be regarded as completely satisfactory. When, however, it has to be welded the welding should be done as for stainless steel, using a rod specially produced for this type of iron.

Hard surfacing and stelliting

Surfaces of intense hardness can be applied to steel, steel alloys, cast iron or monel metal, by means of the oxy-acetylene flame. The surfaces are hard and have intense resistance to wear and corrosion. By depositing a surface on a more ductile metal we have an excellent combination. Metals easily hardfaced are low- and medium-carbon steels, steels above 0.4% C, low-alloy steels, nickel, nickel–copper alloys, chrome-nickel and nobium bearing stainless steels (not titanium bearing nor free machining). Those metals faced with difficulty include cast iron, straight chromium stainless steels, tool and die steels and water-, air- and oil-hardening steels while copper, brass and other low melting point alloys are not suitable for hard surfacing. Hard surfaces may be deposited either on new parts so that they will have increased resistance to wear at reduced cost, or on old parts which may be worn, thus renewing their usefulness. In addition, non-ferrous surfaces, such as bronze or stellite, may be deposited and are described under their respective headings.

The surfaces deposited may be hard and/or wear and corrosion resisting. It is usual in the case of hard surfaces for the metal to be machinable as deposited, but to be capable of being hardened by quenching or by work hardening, e.g. 12/14% Mn (see table of rods available on p. 323).

We may divide the methods as follows:

(1) Building up worn parts with a deposit similar to that of the parent metal. This process is very widely used, being cheap and economical, and is used for building up, for example, gear wheels, shafts, key splines, etc.

The technique employed is similar to that for mild steel, using a neutral flame and the leftward method. The deposit should be laid a little at a time and the usual precautions taken for expansion and distortion. Large parts should be pre-heated and allowed to cool out very slowly, and there are no difficulties in the application.

(2) Building up surfaces with rods containing, for example, carbon, manganese, chromium or silicon to give surfaces which have the required degree of hardness or resistance to wear and corrosion. These surfaces differ in composition from that of the parent metal, and as a result the fusion method of depositing cannot be used, since the surface deposit would become alloyed with the base metal and its hardness or wear-resisting properties would be thus greatly reduced.

(3) Hard facing with tungsten carbide. The rods consist of a steel tube containing fused granules of tungsten carbide HV 1800 in a matrix of chromium iron HV 850. Various grades are available with granules of different mesh size. Large granules embedded in the steel surface do not round off as wear takes place, but chip because they are brittle and maintain good serrated cutting edges. Finer mesh granules give a more regular cutting edge so that the mesh of the granules is determined by the working conditions.

Technique. The flame is adjusted to have excess acetylene with the white plume from 2½ to 3 times the length of the inner cone (Fig. 6.40*a*). As the heated surface absorbs carbon, its melting point is reduced and the surface sweats. The rod is melted on to this sweating surface and a sound bond is

Fig. 6.40. (*a*) Hard surfacing flame making the surface sweat.

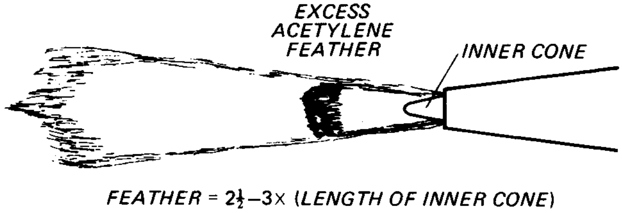

made between deposit and parent metal with the minimum amount of alloying taking place.

This type of deposit, used for its wear-resisting properties, is usually very tough and is practically unmachinable. The surface is therefore usually ground to shape, but in many applications, such as in reinforcing tramway and rail crossings, it is convenient and suitable to hammer the deposit to shape while hot, and thus the deposit requires a minimum amount of grinding. The hammering, in addition, improves the structure.

In the case of the high-carbon deposits, they can be machined or ground to shape and afterwards heat treated to the requisite degree of hardness.

When using tungsten carbide rods the cone of the flame should be played on the rod and pool to allow gases to escape, preventing porosity. Weaving is necessary to give an even distribution of carbide granules in the matrix. Second runs should be applied with the same technique as the first, with no puddling of the first run since this would give dilution of the hard surface with the parent plate.

Another method frequently used to build up a hard surface is to deposit a surface of cast iron in the normal way, using a silicon cast iron rod and flux. Immediately the required depth of deposit has been built up, the part is quenched in oil or water depending on the hardness required.

This results in a hard deposit of white cast iron which can be ground to shape, and which possesses excellent wear-resisting properties. This method is suitable only for parts of relatively simple shape, that will not distort on quenching, such as camshafts, shackle pins, pump parts, etc.

The use of carbon and copper fences in building up and resurfacing results in the deposit being built much nearer to the required shape, reducing time in welding and finishing and also saving material.

When hard surfacing cast iron, the surface will not sweat. In this case the deposit is first laid as a fusion deposit, with a neutral flame, and a second layer is then 'sweated' on to this first layer. In this way the second layer is obtained practically free from any contamination of the base metal.

It will be seen, therefore, that the actual composition of the deposit will depend entirely upon the conditions under which it is required to operate. A table of alloy steel rods and uses is given at the end of this chapter.

Stellite

Stellite alloys of the cobalt–chrome type with carbon were first developed in 1900 in the US and were the first wear-resistant cobalt-based alloys to be used. Later tungsten and/or molybdenum were added. They are hard and have a great resistance to wear and corrosion and there are now about 20 Stellite alloys available, covering a wide field of resistant surfaces. They can be divided into:

(1) Co–Cr–W–C,
(2) Co–Cr–W/Mo–Ni/Fe–C.

The latter alloys of group (2) have been developed for high-temperature, high-impact wear operations.

The carbon content of the alloys varies from 0.1% to 3.0% and over half of them have a carbon content above 1%. Those with less carbon have a combined carbon plus boron content greater than 1%. (The cobalt-based alloys are discussed in Chapter 1 on hard facing by MMA.)

Stellite alloys available for use with the oxy-acetylene process with the hardness Rockwell on the C scale and VPH are:

> No. 1 53R, 590VPH; No. 4 47R 480VPH; No. 6 42R, 420VPH; No. 12 48R, 500VPH; No. 20 55R, 620VPH; No. 190 52R, 580VPH; No. 2006 44R, 440VPH; No. 2012 49R, 520VPH.

In general, wear resistance increases as the hardness increases, owing to the dispersal of the hard carbides in the somewhat softer matrix. Stellite alloys preserve their hardness at elevated temperatures dependent upon the W and/or the Mo content. Bare tube rods of granulated tungsten carbide or of a 60/40 percentage of tungsten carbide with either Stellite No. 6 steel tube filled with the tungsten carbide or the 60/40 blend.

Tips of Stellite can be brazed on to lathe and cutting tools of all types, giving an excellent cutting edge on a less brittle shank. Stellite, however, can be deposited directly on to surfaces, and in this form it is used for all types of duty, such as surfaces on shafts which have to stand up to great wear and corrosion, lathe centres, drill tips, etc.

Stelliting steel

Preparation. Scale, dirt and impurities are thoroughly removed and the parent metal is pre-heated. Pre-heating and slow cooling are essential to avoid cracking.

Small jets of water, playing on each side of the weld, can be used to limit the flow of heat and reduce distortion, when building up deposits on hardened parts such as camshafts.

Flame and rod. A flame with an excess of acetylene is used, the white plume being about 2½–3 times the length of the inner blue cone. Too little acetylene will cause the Stellite to foam and bubble, giving rise to blowholes, while too much acetylene will cause carbon to be deposited around the molten metal. The tip should be one size larger than that for the same thickness steel plate, but the pressure should be reduced, giving a softer flame. For small parts 5 mm diameter Stellite rods can be used, while thicker rods are used for larger surfaces. Too much heat prevents a sound deposit being obtained, because some of the base metal may melt and mix with the molten stellite, thus modifying its structure.

Technique. The flame is directed on to the part to be surfaced, but the inner cone should not touch the work, both blowpipe and rod making an angle of 45–60° with the plate (Fig. 6.40*b* and *c*). When the steel begins to sweat the Stelliting rod is brought into the flame and a drop of Stellite melted on to the sweating surface of the base metal, and it will spread evenly and make an excellent bond with the base metal. The surfacing is continued in this way.

After treatment. The part should be allowed to cool out very slowly to prevent cracks developing.

Fig. 6.40. (*b*) and (*c*) Alternative methods of hard facing.

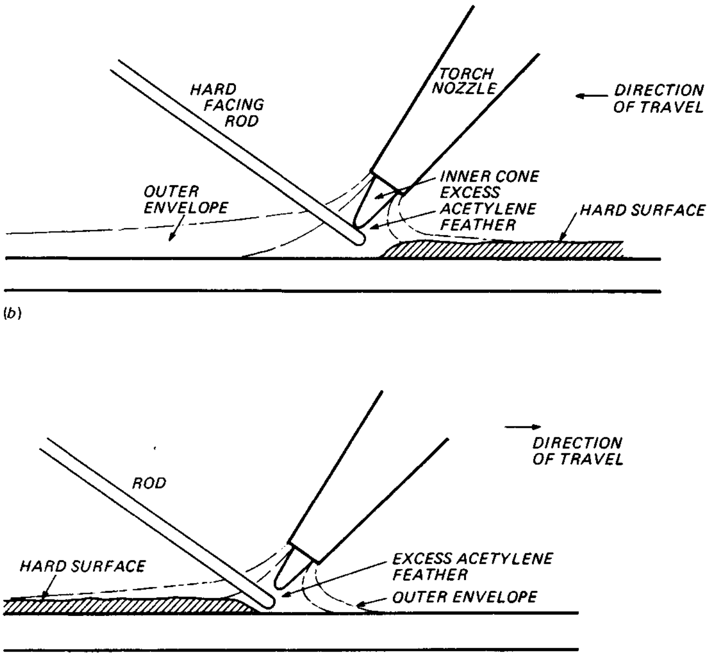
HARD FACING ROD
TORCH NOZZLE
DIRECTION OF TRAVEL
OUTER ENVELOPE
INNER CONE
EXCESS ACETYLENE FEATHER
HARD SURFACE
(b)
ROD
DIRECTION OF TRAVEL
HARD SURFACE
EXCESS ACETYLENE FEATHER
OUTER ENVELOPE

Stelliting cast iron

Preparation. Clean the casting thoroughly of oil and grease and pre-heat to a dull red heat.

Technique. Using the same type rod and flame as for Stelliting steel, it is advisable first to lay down a thin layer and then build up a second layer on this. The reason for this is that more of the surface of the cast iron is melted than in the case of steel, and as a result the first layer of Stellite will be diffused with impurities from the cast iron.

The flame is then played on the cast iron and the rod used to push away any scale. A drop of Stellite is then flowed on to the surface, and the flame kept a little ahead of the molten pool so as to heat the cast iron to the right temperature before the Stellite is run on. Cast iron flux may be used to flux the oxide and produce a better bond.

Heat treatment for depositing Stellite

(*a*) Small components of mild steel and steel up to 0.4% carbon. Pre-heat with the torch the area to be faced; hard face the area and cool away from draughts.

(*b*) Large components of mild steel and steel up to 0.4% carbon; small components of high-carbon and low-alloy steels. Pre-heat to 250–350 °C. Hard face whilst keeping at this heat with auxiliary heating flame. Cover and bring to an even heat with flame and cool slowly in dry kieselguhr, mica, slaked lime, ashes or sand.

(*c*) Large components of high-carbon or low-alloy steel; cast iron components; bulky components of mild steel with large areas of Stellite facing. Pre-heat to 400–500°C. Hard face whilst at this temperature. Bring to a dull red even heat and cool slowly as for (*b*).

(*d*) Air-hardening steel (not stainless). When a large area of deposit is required these steels should be avoided. Otherwise pre-heat to 600–650 °C and deposit hard surface whilst at this heat. Then place in a furnace at 650 °C for 30 minutes and cool large components in the furnace and small ones as (*b*).

(*e*) 18/8 austenitic stainless steel non-hardening welding type. Pre-heat to 600–650 °C. Hard surface whilst at this temperature. Bring to an even temperature and cool out as (*b*).

(*f*) 12–14% manganese austentic steel. Use arc process.

Spray fuse process for depositing Stellite

In this process a stellite powder (HV 425–750) or nickel powder (HV 375–750) is sprayed on to the part to be hardened and this layer is then fused on with an oxy-acetylene flame. In this way thin layers up to 2 mm thick can be deposited having little dilution with the parent metal. The alloys are self-fluxing, the deposit has a fine structure, and any inclusions are well distributed. Surfaces having sharp corners and sudden change of section should be avoided and the areas should be rough turned and then shot blasted to obtain a rough surface. The hardfacing powder, cobalt based, nickel based or tungsten carbide, is applied with a gas spray gun using aspirating gas pressure to mix the powder and to disperse it evenly through the flame on to the pre-heated surface to be treated. The particles are in the form of a semi-molten spray when they hit the surface. Jobs should be pre-heated and the gun held about 150 mm from the surface, and after deposition the deposit is porous like other cold-sprayed deposits. The next operation is to fuse the deposit into a sound, wear-resistant coating securely bonded to the parent metal. Because of stresses remaining in the deposit, fusing should be carried out immediately after spraying. It is done with an oxy-acetylene torch with a multi-jet nozzle, the part being first raised to about 350 °C. One area is selected to begin fusing and the temperature raised to 700–800 °C over a small area and part of this is then raised to 1100 °C when glazing of the surface begins, indicating that fusing is taking place. The torch is moved over the area until it is all fused and the whole part finally brought up to an even temperature, cooling being carried out by covering it with heat-insulating substance or heat treatment being given, if required for the parent plate.

With this method of application localized heat is reduced, so that distortion is minimized and thin deposits can be applied with little dilution. Shrinkage amounting to 25% takes place during fusing so this must be allowed for, together with grinding tolerance if required, when calculating the thickness of the sprayed coat.

In the manual torch powder method, the gun and torch are combined in one unit and deposition and fusing are carried out in one operation. The incoming gas is used for mixing and carrying the powder to the flame. The usual method of deposit is by using a carburizing flame with a feather about 2½–3 times the length of the inner cone. A dilution of 1–5% occurs and the thickness of deposit is governed by the rate of flow of powder and movement of the torch (Fig. 6.41*a* and *b*).

Typical applications are: dies, cast iron parts, cement dies, fan blades, feed screws, hammer mill hammers, knives, cams and other metal-to-metal parts, plough shares, pump parts, bearing surfaces of steel shafts, etc. See also hard facing by plasma spray, plasma transferred arc, mechanized TIG, MIG and MMA methods.

Brazing (see also Appendix 9 for rods and fluxes)

Brazing may be defined as 'a process of joining metals in which, during or after heating, molten filler metal is drawn by capillary attraction into the space between closely adjacent surfaces of the parts to be joined'. In general the melting point of the filler metal is above 500 °C but always below the melting temperature of the parent metal.

Since capillarity and hence surface tension are involved in the process it may be convenient to give a brief explanation of some of the principles involved.

The student should refer to BS 1845 which lists the chemical compositions and approximate melting ranges of filler metals grouped under the following headings: aluminium brazing alloys, silver brazing alloys, copper–phosphorus brazing alloys, copper brazing alloys, brazing brasses, nickel-base brazing alloys, palladium bearing brazing alloys, and gold bearing brazing alloys.

Fig. 6.41. (*a*) Manual torch operation.

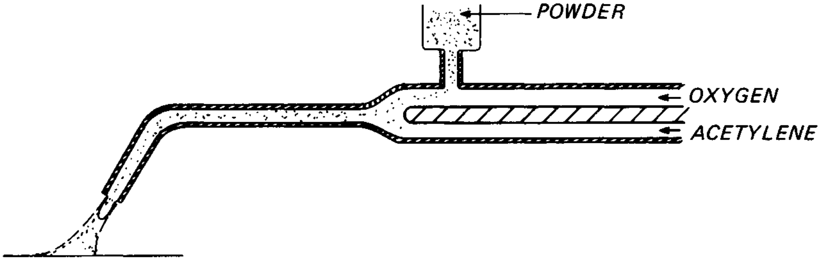
POWDER
OXYGEN
ACETYLENE

Fig. 6.41. (*b*) Manual power torch–stellite.

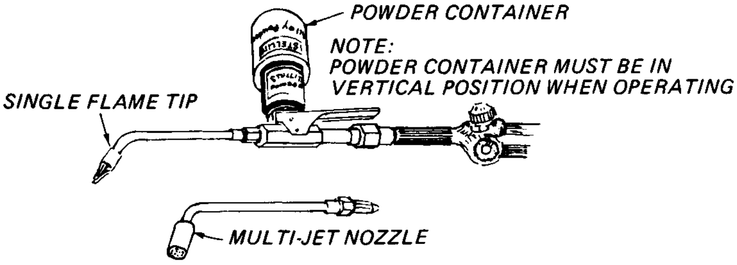
POWDER CONTAINER
NOTE:
POWDER CONTAINER MUST BE IN VERTICAL POSITION WHEN OPERATING
SINGLE FLAME TIP
MULTI-JET NOZZLE

Surface tension

If drops of mercury rest on a level plate it will be noticed that the smaller the drop the more nearly spherical it is in shape, and if any drop is deformed it always returns to its original shape. If the only force acting on any drop were that due to its own weight, the mercury would spread out over the plate to bring its centre of gravity (the point at which the whole weight of the drop may be conceived to be concentrated) to the lowest point so that to keep the shape of the drop other forces must be present. As the drop gets smaller the force due to its own weight decreases and these other forces act so as to make the drop more spherical, that is to take up a shape which has the smallest surface area for a given volume. Other examples of these forces, termed surface tension, are the floating of a dry needle on the surface of water, soap bubbles and water dripping from a tap. In the first example the small dry needle must be laid carefully horizontally on the surface. If it is pushed slightly below the surface it will sink because of its greater density. Evidently the surface of the water exhibits a force (surface tension) which will sustain the weight of the needle. If a wire framework $ABCD$, with CD, length x, able to slide along BC and AD, holds a soap film, the film tends to contract, and to prevent this a force F must be applied at right angles to CD (Fig. 6.42). The surface tension is defined as the force per unit length S on a line drawn in the film surface, and since there are two surfaces to the film $F = 2Sx$.

Angle of contact(θ). The angle of contact between a liquid and a solid may be defined as the angle between the tangent to the liquid surface at the point of contact and the solid surface. For mercury on glass the contact angle is about 140°, while for other liquids the angle is acute and may approach zero (Fig. 6.43).

Fig. 6.42

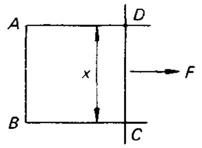

Fig. 6.43

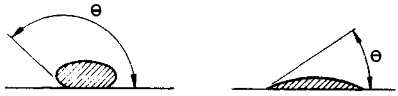

CONTACT ANGLE
(NON-WETTING AND WETTING)

Wetting. If the contact angle approaches zero the liquid spreads and wets the surface and may do so in an upward direction. If the solid and liquid are such that the forces of attraction experienced by the molecules towards the interior of the liquid are less than the forces of attraction towards the solid, the area of contact will increase and the liquid spreads.

Capillarity. If a narrow bore (capillary) tube with open ends is placed vertically in a liquid which will wet the surface of the tube, the liquid rises in the tube and the narrower the bore of the tube the greater the rise. The wall thickness of the tube does not affect the rise and a similar rise takes place if the tube is replaced by two plates mounted vertically and held close together. If the tube or the plates are held out of the vertical the effect is similar and the vertical rise the same. If the liquid does not wet the tube (e.g. mercury) a depression occurs, and the shape of the liquid surface (the meniscus) is shown. The rise is due to the spreading or wetting action already considered – the liquid rises until the vertical upward force due to surface tension acting all round the contact surface with the tube is equal to the vertical downward force due to the weight of the column of liquid (Fig. 6.44).

This wetting action and capillary attraction are involved in the brazing process. The flux, which melts at a lower temperature than the brazing alloys, wets the surface to be brazed, removes the oxide film and gives clean surfaces which promote wetting by a reduction of the contact angle between the molten filler alloy and the parent plate at the joint. The molten filler alloy flows into the narrow space or joint between the surfaces by capillary attraction and the narrower the joint the further will be the capillary flow. Similarly solder flows into the narrow space between tube and union when 'capillary fittings' are used in copper pipe work.

Brazing can be performed on many metals including copper, steel and aluminium, and in all cases cleanliness and freedom from grease is essential. The filler alloy used for aluminium brazing has already been mentioned in the section on oxy-acetylene welding. In the case of the nickel alloys, time and temperature are important. For example copper alloys mix readily with nickel alloy 200 or MONEL® alloy 400 and can pick up sufficient

Fig. 6.44

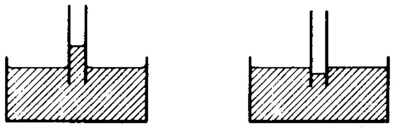
CAPILLARY ELEVATION CAPILLARY DEPRESSION

nickel to raise the melting point and hinder the flow of the filler metal. Also chromium and aluminium form refractory oxides which make brazing difficult, so that the use of a flux is necessary. A wide range of brazing alloys are available having a variety of melting points. Fig. 6.45 shows a modern brazing torch using LPG. See appendix for table of general purpose silver brazing alloys and fluxes.

Aluminium brazing

The fusion welding of fillet and lap joints in thin aluminium sections presents considerable difficulty owing to the way in which the edges melt back. In addition the corrosive fluoride flux is very liable to be entrapped between contracting metal surfaces so it is advantageous to modify the design wherever possible so as to include butt joints instead of fillet and lap.

In many cases, however, corner joints are unavoidable and in these cases flame brazing overcomes the difficulty. It can be done more quickly and cheaply than fusion welding, less skill is required and the finished joint is neat and strong.

Fig. 6.45. LP gas brazing torch for brazing, hard and soft soldering. Propane gas recommended, cylinder pressure 7–8 bar. Standard torch pressure 4 bar. Use smaller nozzles for smaller flames. Temperatures up to 950 °C, piezo-electric ignition.

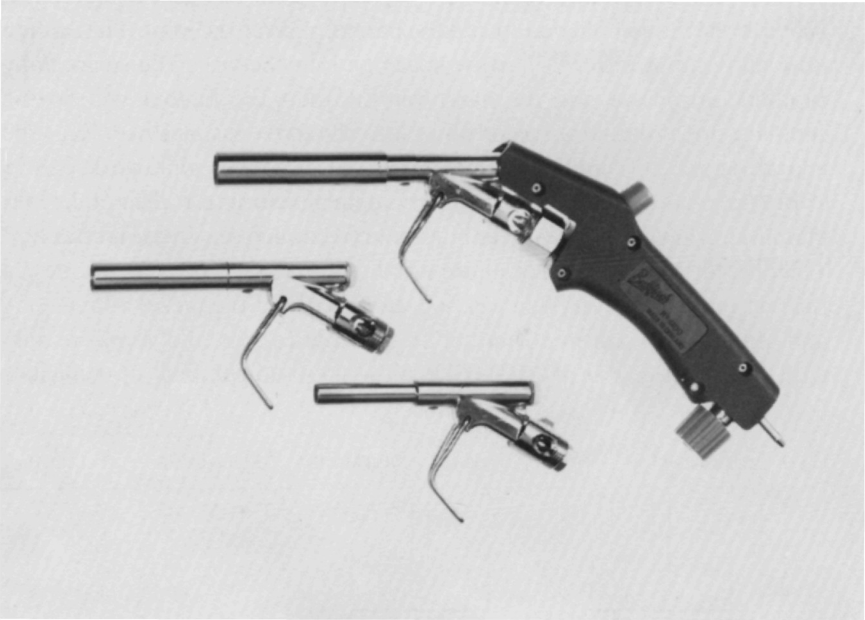

Aluminium brazing is suitable for pure aluminium and for alloys such as LM4, LM18 and for the aluminium–manganese and aluminium–manganese–magnesium alloys, as long as the magnesium content is not greater than 2%.

Preparation. It is always advisable to allow clearance at the joints since the weld metal is less fluid as it diffuses between adjacent surfaces. Clearance joints enable full penetration of both flux and filler rod to be obtained.

Surface oxide should be removed by wire brush or file and grease impurities by cleaning or degreasing. Burrs such as result from sawing or shearing and other irregularities should be removed so that the filler metal will run easily across the surface. Socket joints should have a 45° belling or chamfer at the mouth to allow a lead in for the flux and metal, and to prevent possibility of cracking on cooling, the sections of the surfaces to be jointed should be reduced to approximately the same thickness where possible.

Blowpipe, flame, flux and rod. The blowpipe and rod are held at the normal angle for leftward welding, and a nozzle giving a consumption of about 700 litres of both oxygen and acetylene each per hour is used. The flame is adjusted to the excess acetylene condition with the white acetylene plume approximately $1\frac{1}{2}$ to 2 times the length of the inner blue cone. The rod can be of 10–13% silicon–aluminium alloy melting in the range 565–595 °C (compared with 659 °C for pure aluminium). This is suitable for welding and brazing the high silicon aluminium alloys and for general aluminium brazing. Another type of rod containing 10–13% Si and 2–5% Cu is also used for pure aluminium and aluminium alloys except those with 5% silicon, or with more than 2% magnesium, but is not so suitable due to its copper content if corrosive conditions are to be encountered, but it has the advantage of being heat-treatable after brazing, giving greater mechanical strength.

BS 1845, Filler metals for brazing

	Major alloying elements only (%)				Melting range (°C)	
Type	Silicon	Copper	Iron	Aluminium	Solidus	Liquidus
AL1	10–13	2–5	0.6	Rem.	535	595
AL2	10–13	0–0.1	0.6	Rem.	565	595
AL3	7–8	0–0.1	—	Rem.	565	610
AL4	4.5–6.0	0–0.1	—	Rem.	565	630

In general, ordinary finely divided aluminium welding flux prepared commercially is quite suitable for brazing, but there is also available a brazing flux of similar composition which has a lower melting point.

Technique. The flame is held well away from the joint to be brazed and preheating is done with the outer envelope of the flame for about three minutes – this procedure ensures an even temperature, which is essential so that first the flux flows evenly into the joint followed by the filler metal which must displace *all* the flux, otherwise if islands of flux are entrapped corrosion of the joint will occur.

The rod is warmed, dipped in the flux and the tuft which adheres to the rod is touched down on the heated joint. When the correct temperature has been reached the flux will melt and flow over and between the surfaces smoothly and easily. The blowpipe is now lowered to the normal welding position, some rod is melted on to the joint and the blow-pipe moved forward along the seam running the filler rod into the joint. The blowpipe is then raised and brought back a little and lowered again, the above operation being repeated – the blowpipe thus describes an elliptical motion – each operator modifying the technique according to his individual style (Fig. 6.46).

Flux should only be added when the filler rod does not appear to be running freely. Too much flux is detrimental to the finished joint and great care should be taken that the filler rod flows freely into the joint, so as to displace all the flux.

After treatment. The corrosive flux should be removed by the treatment as given for aluminium welding.

Fig. 6.46. Aluminium brazing.

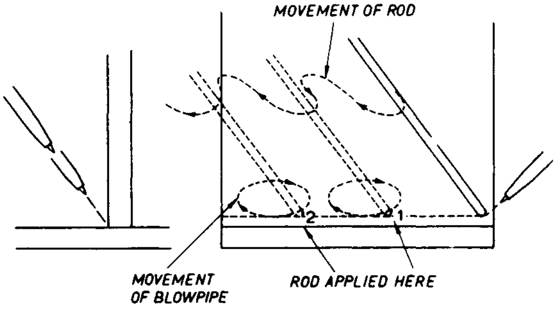

General precautions

The following general precautions should be taken in welding:

(1) Always use goggles of proved design when welding or cutting. The intense light of the flame is harmful to the eyes and in addition small particles of metal are continually flying about and may cause serious damage if they lodge in the eyes. Welding filters or glasses are graded according to BS 679 by numbers followed by the letters GW or GWF. The former are for welding operations without a flux and the latter with flux because there is an additional amount

Table giving some of the chief types of oxy-acetylene welding rods available for welding carbon and alloy steels.

Rod	Description and use
Low-carbon steel	A general purpose rod for mild steel. Easily filed and machined. Deposit can be case-hardened.
High-tensile steel	Gives a machinable deposit of greater tensile strength than the previous rod. Can be used instead of the above wherever greater strength is required.
High-carbon steel	Gives a deposit which is machinable as deposited, but which can be heat treated to give a hard abrasion-resisting surface. When used for welding broken parts, these should be of high carbon steel, and heat treatment given after welding.
High-nickel steel (3½–4%)	Produces a machinable deposit with good wear-resisting properties. Suitable for building up teeth in gear and chain wheels, splines and keyways in shafts, etc.
Wear-resisting steel (12–14% manganese)	Gives a dense tough unmachinable deposit which must be ground or forged into shape, and can be heat treated. Useful for building up worn sliding surfaces, cam profiles, teeth on excavators, tracks, etc.
Stellite	For hard surfacing and wear- and abrasion-resisting surfaces.
Chrome molybdenum steel (creep resisting)	High-tensile alloy steel deposit for pressure vessels and high-pressure steam pipes, etc. Rod should match the analysis of the parent metal.
Chrome-vanadium steel	A high-tensile alloy rod for very highly stressed parts.
Tool steel	Suitable for making cutting tools by tipping the ends of mild or low-carbon steel shanks.
Stainless steel	Decay-proof. Rod should match the analysis of the parent metal.

of glare. The grades range from 3/GW and 3/GWF to 6/GW and 6/GWF, the lightest shade having the lowest grade number. For aluminium welding and light oxy-acetylene cutting, 3/GW or 3/GWF is recommended, and for general welding of steel and heavier welding in copper, nickel and bronze, 5/GW or 5/GWF is recommended. A full list of recommendations is given in the BS.

(2) When welding galvanized articles the operator should be in a well-ventilated position and if welding is to be performed for any length of time a respirator should be used. (In cases of sickness caused by zinc fumes, as in welding galvanized articles or brass, milk should be drunk.)

(3) In heavy duty welding or cutting and in overhead welding, asbestos or leather gauntlet gloves, ankle and feet spats and protective clothing should be worn to prevent burns. When working inside closed vessels such as boilers or tanks, take every precaution to ensure a good supply of fresh air.

(4) In welding or cutting tanks which have contained inflammable liquids or gases, precautions must be taken to prevent danger of explosion. One method for tanks which have contained volatile liquids and gases is to pass steam through the tank for some hours according to its size. Any liquid remaining will be vaporized by the heat of the steam and the vapours removed by displacement.

Tanks should never be merely swilled out with water and then welded; many fatal explosions have occurred as a result of this method of preparation. Carbon dioxide in the compressed form can be used to displace the vapours and thus fill the tank, and is quite satisfactory but is not always available. Tanks which have contained heavier types of oil, such as fuel oil, tar, etc., present a more difficult problem since air and steam will not vaporize them. One method is to fill the tank with water, letting the water overflow for some time. The tank should then be closed and turned until the fracture is on top. The water level should be adjusted (by letting a little water out if necessary) until it is just below the fracture. Welding can then be done without fear as long as the level of the water does not drop much more than a fraction of an inch below the level of the fracture.

The welder should study the Department of Employment memorandum on *Safety measures for the use of oxy-acetylene equipment in factories* (Form 1704). Toxic gases and fine airborne particles can provide a hazard to a welder's health. The Threshold Limit Value (TLV) is a system by which concentrations of these are classified, and is explained in the Department of Employment Technical data notes No. 2, *Threshold limit values.*

Relevant publications. *Safe working, maintenance and repair of gas cylinder and pipeline regulators used with compressed gases for welding, cutting and related processes* (BCGA CP1).

Safe working, maintenance and repair of hand held oxygen and fuel-gas blowpipes used for welding, cutting and related processes (BCGA CP2).

Technical Information sheets; TIS 1 *Pressure gauges,* TIS2 *Hoses.*

All the above together with British Acetylene Association publications may be obtained from the British Compressed Gases Association, 93 Albert Embankment, London SE1 7TU.

In the 'Health and Safety at Work' series *Noise and the worker No. 25* and *Welding and flame-cutting using compressed gases No. 50* are obtainable from Her Majesty's Stationery Office.

7

Cutting processes

Gas cutting of iron and steel

Iron and steel can be cut by the oxy-hydrogen, oxy-propane, oxy-natural gas and oxy-acetylene cutting blowpipes with ease, speed and a cleanness of cut.

Principle of cutting operation

There are two operations in gas cutting. A heating flame is directed on the metal to be cut and raises it to bright red heat or ignition point. Then a stream of high-pressure oxygen is directed on to the hot metal. The iron is immediately oxidized to magnetic oxide of iron (Fe_3O_4) and, since the melting point of this oxide is well below that of the iron, it is melted immediately and blown away by the oxygen stream.

It will be noted that the metal is cut entirely by the exothermic chemical action and the iron or steel itself is not melted. Because of the rapid rate at which the oxide is produced, melted and blown away, the conduction of the metal is not sufficiently high to conduct the heat away too rapidly and prevent the edge of the cut from being kept at ignition point.

The heat to keep the cut going once it has started is provided partly by the heating jet, and partly by the heat of the chemical action.

The cutting torch or blowpipe (Fig. 7.1)

Cutting blowpipes may be either high or low pressure. The high pressure pipe, using cylinder acetylene or propane as the fuel gas,* can have the mixer in the head (Fig. 7.2), or in the shank, while the low pressure pipe with injector mixing can be used with natural gas at low pressure.

---

* Proprietary gases are available and contain mixtures of some of the following: methyl acetylene, propadiene, propylene, butane, butadiene, ethane, methane, diethyl ether, dimethyl ether, etc.

In the high pressure pipe, fuel gas and heating oxygen are mixed in the head (Fig. 7.2), and emerge from annular slots for propane or holes for acetylene (Fig. 7.4).

The cutting oxygen is controlled by a spring loaded lever, pressure on which releases the stream of cutting oxygen which emerges from the central

Fig. 7.1. Cutter.

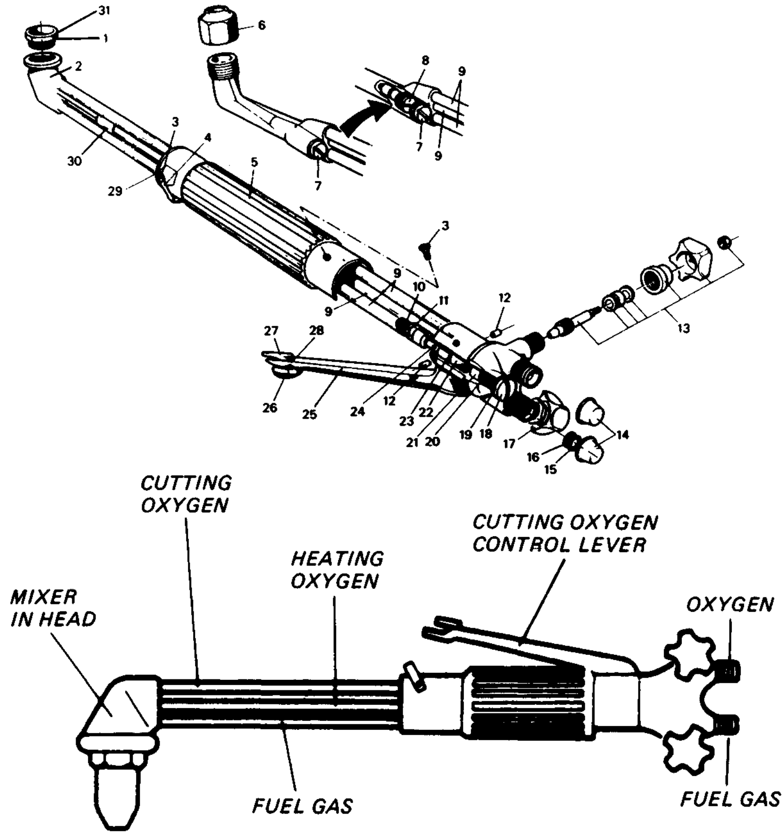
31
1
2
6
8
9
3
4
30
29
5
7
9
7
3
9
10
11
9
12
13
27
28
26
25
24
12
23
22
21
20
19
18
17
16
15
14
CUTTING OXYGEN
HEATING OXYGEN
MIXER IN HEAD
CUTTING OXYGEN CONTROL LEVER
OXYGEN
FUEL GAS
FUEL GAS

Key:

1. Nozzle nut.
2. Head 90°.
   Head 75°.
   Head 180°.
3. Screw.
4. Bracket latch.
5. Handle.
6. Nozzle nut.
7. Injector cap.
8. Injector assembly.
9. Tube.

9a. Tube.

10. Push rod.
11. Lock nut.
12. Pivot pin.
13. Control valve fuel gas.
14. Red cap.
15. Spring clip.
16. Filter.
17. Control valve oxygen.
18. Rear cap.
19. Cap washer.
20. Valve spring disc.
21. Rear valve.
22. Plunger.
23. O ring.
24. Lever pin.
25. Lever.
26. Button.
    Grub screw.
27. Lever latch.
28. Spring washer.
29. Forward tube.
30. Tube support.
31. Nut.

orifice the diameter of which increases as thickness of plate to be cut increases. An oxygen gauge giving a higher outlet pressure (up to 10 bar) than for welding is required, but the gauge for cylinder acetylene can be the same with pressures up to 1.3 bar. Propane pressure is usually up to 3.5 bar and if natural gas is used no regulator is required but a non-return valve should be fitted in the supply line to prevent flash back.

The size of blowpipe used depends upon whether it is for light duty or

Fig. 7.2

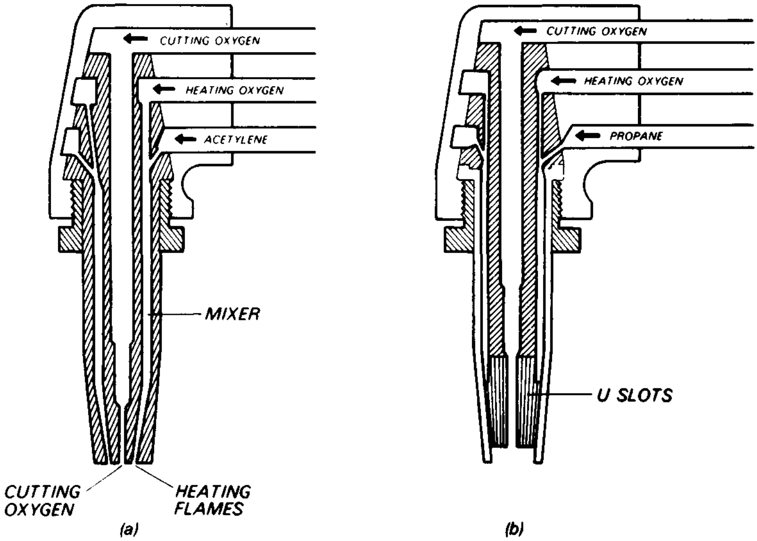
CUTTING OXYGEN
HEATING OXYGEN
ACETYLENE
MIXER
CUTTING OXYGEN
HEATING FLAMES
(a)
CUTTING OXYGEN
HEATING OXYGEN
PROPANE
U SLOTS
(b)

Fig. 7.3. Cutting head for thin steel sheet.

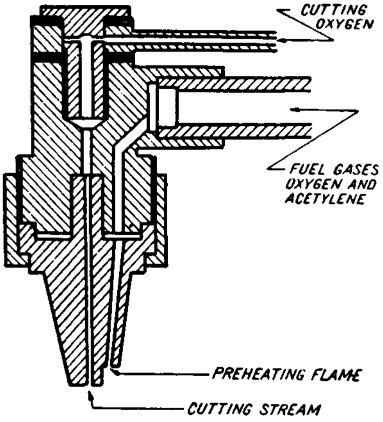
CUTTING OXYGEN
FUEL GASES OXYGEN AND ACETYLENE
PREHEATING FLAME
CUTTING STREAM

heavy continuous cutting and the volume of oxygen used is much greater than of fuel gas (measured in litres per hour, l/h).

Fig. 7.3 shows a stepped nozzle used for cutting steel sheet up to 4 mm thick and this type is also available with head mixing.

The size of torch varies with the thickness of work it is required to cut and whether it is for light duty, or heavy, continuous cutting.

Adjustment of flame

Oxy-hydrogen and Oxy-propane. The correctly adjusted pre-heating flame is a small non-luminous central cone with a pale blue envelope.

Oxy-natural gas. This is adjusted until the luminous inner cone assumes a clear, definite shape, that may be up to 8–10 mm in length for heavy cuts.

Oxy-acetylene. This flame is adjusted until there is a circular short blue luminous cone, if the nozzle is of the concentric ring type, or until there is a series of short, blue, luminous cones (similar to the neutral welding flame), if it is of the multi-hole type (Fig. 7.5*a* and *b*). The effect of too much oxygen is indicated in Fig. 7.5*c*.

Fig. 7.4

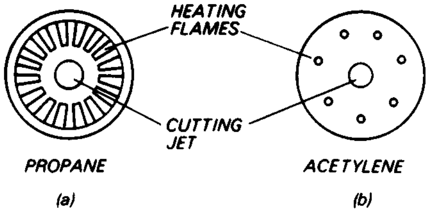

Fig. 7.5

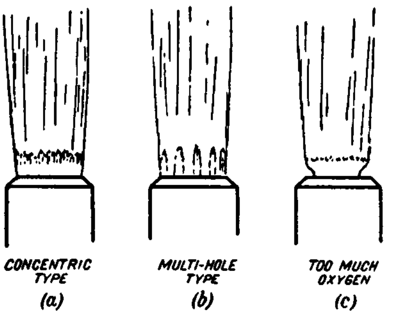

It may be observed that when the cutting valve is released the flame may show a white feather, denoting excess acetylene. This is due to the slightly decreased pressure of the oxygen to the heating jet when the cutting oxygen is released. The flame should be adjusted in this case so that it is neutral when the cutting oxygen is released.

Care should be taken to see that the cutter nozzle is the correct size for the thickness to be cut and that the oxygen pressure is correct (the nozzle sizes and oxygen pressures vary according to the type of blowpipe used).

The nozzle should be cleaned regularly, since it becomes clogged with metallic particles during use. In the case of the concentric type of burner, the outer ring should be of even width all round, otherwise it will produce an irregular-shaped inner cone, detrimental to good cutting (Fig. 7.6).

Technique of cutting

The surface of the metal to be cut should be free of grease and oil, and the heating flame held above the edge of the metal to be cut, farthest from the operator, with the nozzle at right angles to the plate. The distance of the nozzle from the plate depends on the thickness of the metal to be cut, varying from 3 to 5 mm for metal up to 50 mm thick, up to 6 mm for metal 50 to 150 mm thick. Since the oxide must be removed quickly to ensure a good, clean cut, it is always preferable to begin on the edge of the metal.

The metal is brought to white heat and then the cutting valve is released, and the cut is then proceeded with at a steady rate. If the cutter is moved along too quickly, the edge loses its heat and the cut is halted. In this case, the cutter should be returned to the edge of the cut, the heating flame applied and the cut restarted in the usual manner. Round bars are best nicked with a chisel, as this makes the starting of the cut much easier. Rivet heads can be cut off flush by the use of a special type nozzle while if galvanized plates are to be cut for any length of time, a respirator is advisable, owing to the poisonous nature of the fumes.

Fig. 7.6

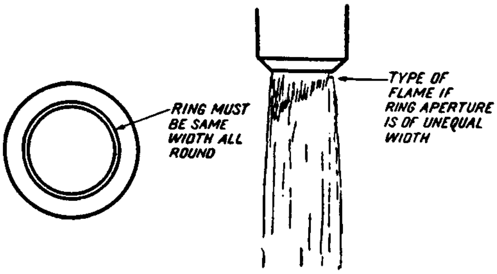
RING MUST BE SAME WIDTH ALL ROUND
TYPE OF FLAME IF RING APERTURE IS OF UNEQUAL WIDTH

To cut a girder, for example, the cut may be commenced at A and taken to B (Fig. 7.7). Then commenced at C and taken to B, that is, the flange is cut first. Then the bottom flange is cut in a similar manner. The cut is then commenced at B and taken to E along the web, this completing the operation. By cutting the flanges first the strength of the girder is altered but little until the web itself is finally cut.

Rollers and point guides can be affixed to the cutter in order to ensure a steady rate of travel and to enable the operator to execute straight lines or circles, etc., with greater ease (Fig. 7.8*a*).

The position of the flame and the shape of the cut are illustrated in Fig. 7.8*b*.

To close down, first shut off the cutting stream, then the propane or acetylene and then the oxygen valve. Close the cylinder valves and release the pressure in the tubes by momentarily opening the cutter valves.

To cut holes in plates a slightly higher oxygen pressure may be used. The spot where the hole is required is heated as usual and the cutting valve

Fig. 7.7

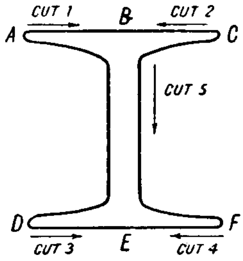
CUT 1
B
CUT 2
A
C
CUT 5
D
F
CUT 3
E
CUT 4

Fig. 7.8

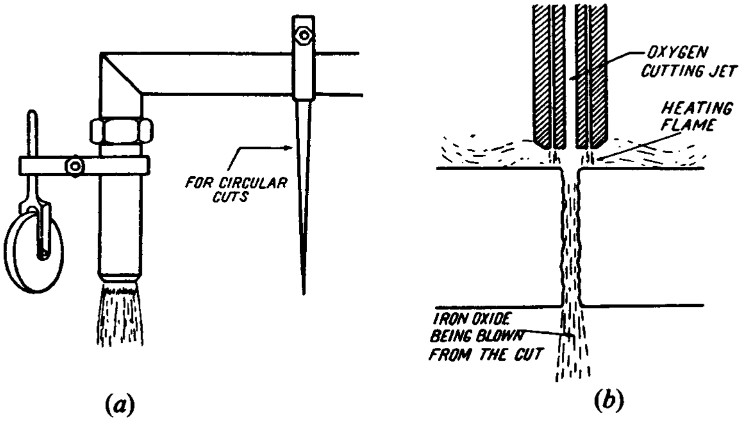
FOR CIRCULAR CUTS
OXYGEN CUTTING JET
HEATING FLAME
IRON OXIDE BEING BLOWN FROM THE CUT

(*a*) (*b*)

released gently, at the same time withdrawing the cutter from the plate. The extra oxygen pressure assists in blowing away the oxide, and withdrawing the nozzle from the plate helps to prevent oxide from being blown on to the nozzle and clogging it. The cutting valve is then closed and the lower surface now exposed is heated again, and this is then blown away, these operations continuing until the hole is blown through. The edges of the hole are easily trimmed up afterwards with the cutter.

When propane or natural gas is used instead of acetylene, the flame temperature is lower; consequently it takes longer to raise the metal to ignition point and to start the cut, and it is not suitable for hand cutting over 100 mm thick nor for cast iron cutting. Because of this lower temperature, the speed of cutting is also slower. The advantages, however, are in its ease of adjustment and control; it is cheap to instal and operate and, most important of all, since the metal is not raised to such a high temperature as with the oxy-acetylene flame, the rate of cooling of the metal is slower and hence the edges of the cut are not so hard. This is especially so in the case of low-carbon and alloy steels. For this reason, oxy-propane or natural gas is often used in works where cutting machines are operated (see later).

The oxy-hydrogen flame can also be used (the hydrogen being supplied in cylinders, as the oxygen). It is similar in operation to oxy-propane and has about the same flame temperature. It is advantageous where cutting has to be done in confined spaces when the ventilation is bad, since the products of combustion are not so harmful as in the case of oxy-acetylene and oxy-propane, but is not so convenient and quick to operate.

The effect of gas cutting on steel

It would be expected that the cut edge would present great hardness, owing to its being raised to a high temperature and then subjected to rapid cooling, due to the rapid rate at which heat is dissipated from the cut edges. Many factors, however, influence the hardness of the edge.

Steels of below 0.3% carbon can be easily cut, but the cut edges will definitely harden, although the hardness rarely extends more than 3 mm inwards and the increase is only 30 to 50 points Brinell.

Steels of 0.3% carbon and above and also alloy steels are best preheated before cutting, as this reduces the liability to crack.

Nickel, molybdenum, manganese and chrome steels (below 5%) can be cut in this way. Steels having a high tungsten, cobalt or chrome content, however, cannot be cut satisfactorily. Manganese steel, which is machined with difficulty owing to the work hardening, can be cut without any bad effects at all.

The oxy-acetylene flame produces greater hardening effect than the oxy-propane flame, as before mentioned, owing to its higher temperature. Excessive cutting speeds also cause increased hardness, since the heat is thereby confined to a narrower zone near the cut and cooling is thus more rapid. Similarly, a thick plate will harden more than a thin one, owing to its more rapid rate of cooling from the increased mass of metal being capable of absorbing the heat more quickly. The hardening effect for low carbon steels, however, can be removed either by preheating or heat treatment after the cut. The hardening effect in mild steel is very small. On thicknesses of plate over 12 mm it is advisable to grind off the top edge of the cut, as this tends to be very hard and becomes liable to crack on bending.

The structure of the edges of the cut and the nearby areas will naturally depend on the rate of cooling. Should the cutting speed be high and the cooling be very rapid in carbon steels, a hard martensitic zone may occur, while with a slower rate of cutting and reduced rate of cooling the structure will be softer. A band of pearlite is usually found, however, very near to the cut edges and because of this, the hardness zone, containing increased carbon, is naturally very narrow. When the cut edge is welded on directly, without preparation, all this concentration of carbon is removed.

Thus, we may say that, for steels of less than 0.3% carbon, if the edges of the cut are smooth and free from slag and loose oxide, the weld can be made directly on to the gas-cut edge without preparation.

Oxygen or thermic lance

This is a method of boring or cutting holes in concrete, brick, granite, etc., by means of the heat generated by chemical reactions.

The lance consists of a tube about 3 m long and 6.5–9.5 mm diameter which is packed with steel wires. One end of the tube is threaded or snap connected and is connected by means of a flexible hose to an oxygen supply. To operate the lance the free end is heated and oxygen passed down the tube. Rapid oxidation of the wires begins at the heated end with great evolution of heat. Magnesium and aluminium are often added to the packing to increase the heat output. The operator can be protected by a shield and protective clothing should be worn. The exothermic reaction melts concrete and other hard materials to a fluid slag and cast iron is satisfactorily bored. Standard gas pipe can be used for thick steel sections.

As an example, a hole of 30 mm diameter and 300 mm deep can be bored in concrete using 1.9 m of lance and 1.0–1.3 m^3 of oxygen at a pressure of 7 bar (100 lb per in^2) in 120 seconds.

Cutting machines

Profiles cut by hand methods are apt to be very irregular and, where accuracy of the cut edge is required, cutting machines are used. The heating flame is similar to that used in the hand cutter and is usually oxy-propane or oxy-acetylene (either dissolved or generated), while the thickness of the cut depends on the nozzle and gas pressures.

The mechanical devices of the machine vary greatly, depending upon the type of cut for which they are required. In many types a tracing head on the upper table moves over the drawing or template of the shape to be cut. Underneath the table or on the opposite side of the machine (depending on the type) the cutting head describes the same motion, being worked through an intermediate mechanism. The steel being cut is placed on supports below the table (Fig. 7.9*a* and *b*).

Simpler machines for easier types of cuts, such as straight line and circles, bevels, etc., are also made.

A typical machine is capable of cutting from 1.5 mm to 350 mm thick, 3 m in a straight line and up to 1.5 m diameter circle. The machine incorporates a magnetic tracing roller, which follows round a steel or iron template the exact shape of the cut required, while the cutting head cuts the replica of this shape below the table.

Fig. 7.9. (*a*) Cutting machine fitted with four cutting heads.

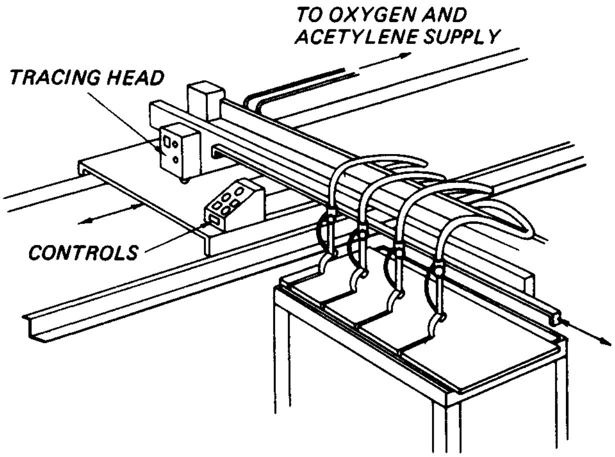

Stack cutting

Thin plates which are required in quantities can be cut by clamping them tightly in the form of a stack and, due to the accuracy of the modern machines, this gives excellent results and the edges are left smooth and even. Best results are obtained with a stack 75–100 mm thick, while G clamps can be used for the simpler types of stack cutting.

Cast iron cutting

Cast iron cutting is made difficult by the fact that the graphite and silicon present are not easily oxidized. Reasonably clean cuts can now be made, however, using a blowpipe capable of working at high pressure of oxygen and acetylene. Cast iron cannot be cut with hydrogen.

Since great heat is evolved in the cutting process, it is advisable to wear protective clothing, face mask and gloves.

The oxygen pressure varies from 7 N/mm^2 for 35 mm thick cast iron to 11 N/mm^2 for 350 to 400 mm thick, while the acetylene pressure is increased accordingly.

The flame is adjusted to have a large excess of acetylene, the length of the white plume being from 50–100 mm long (e.g. 75 mm long for 35–50 mm thick plate). The speed of cutting is low, being about 2.5 m per hour for 75–125 mm thick metal.

Fig. 7.9. (*b*)

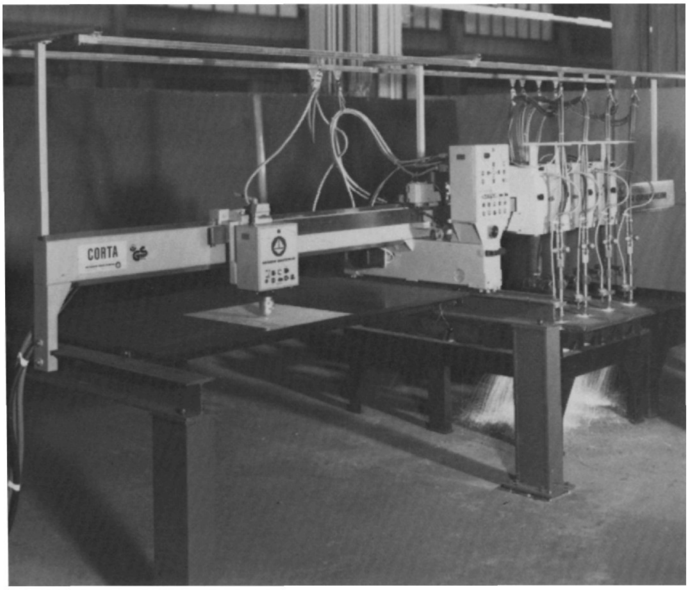

Technique. The nozzle is held at 45° to the plate with inner cone 5–6 mm from the plate, and the edge where the cut is to be started is heated to red heat over an area about 12–18 mm diameter. The oxygen is then released and this area burnt out. The blowpipe is given a zig-zag movement, and the cut must not be too narrow or the slag and metal removed will clog the cut. About 12 to 18 mm is the normal width. After the cut is commenced the blowpipe may be raised to an angle of 70–80°, which will produce a lag in the cut, as shown in Fig. 7.10.

Owing to the fact that high pressures are used in order to supply sufficient heat for oxidation, large volumes of gas are required, and this is often obtained by connecting several bottles together.

Flame gouging by the oxy-acetylene process

Flame gouging is an important extension of the principle of oxy-acetylene cutting by which grooves with very smooth contours can be cut easily in steel plates without the plate being penetrated.

Principle of operation

This is the same as that in the oxy-acetylene process, except that a special type of nozzle is used in the standard cutting blowpipe. A pre-heating flame heats the metal to red heat (ignition temperature), the cutting oxygen is switched on, oxidation occurs and the cut continues as previously explained.

Equipment

The cutting torch may have a straight 75° or 180° angle head together with a range of special gouging nozzles, as shown in Fig. 7.11. The nozzle sizes are designated by numbers and they are bent at an angle which is best for the gouging process. Regulators and other equipment are as for cutting.

Fig. 7.10

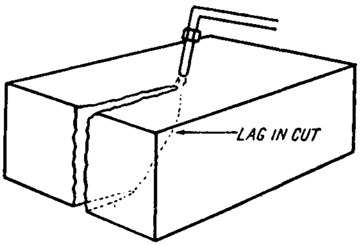

Operation

There are two main techniques:

1. progressive,
2. spot.

In the former the groove is cut continuously along the plate – it may be started at the plate edge or anywhere in the plate area. It can be used for removing the underbeads of welds prior to depositing a sealing run, or it may be used for preparing the edges of plates. Spot gouging, however, is used for removing small localized areas such as defective spots in welds.

To start the groove at the edge of a plate for continuous or progressive gouging the nozzle is held at an angle of about 20° so that the pre-heating flames are just on the plate and when this area gets red hot the cutting oxygen stream is released and at the same time the nozzle is brought at a shallower angle to the plate as shown, depending upon depth of gouge required. The nozzle is held so that the nozzle end clears the bottom of the cut and the pre-heating flames are about 1.5 mm above the plate.

The same method is adopted for a groove that does not start at the plate edge. The starting point is pre-heated with the nozzle making a fairly steep angle with the plate at 20–40°. When the pre-heated spot is red hot the cutting oxygen stream is released and the angle of the nozzle reduced to 5–10° depending upon the depth of gouge required (Fig. 7.12).

To gouge a single spot, it is pre-heated as usual where required, but when red hot and the cutting stream of oxygen is turned on, the angle of the nozzle is *increased* (instead of, as previously, decreased) so as to make the gouge deep.

The depth of groove cut depends upon nozzle size, speed of progress and angle between nozzle and plate (i.e. angle at which the cutting stream of oxygen hits the plate). The sharper this angle, the deeper the groove. If the

Fig. 7.11. Flame gouging nozzle.

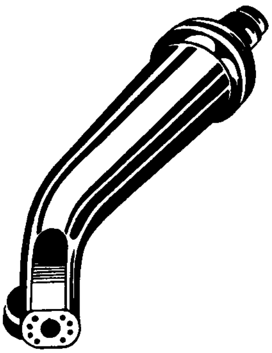

cutting oxygen pressure is too low ripples are left on the base of the groove. If the pressure is too high the cut at the surface proceeds too far in advance of the molten pool, and eventually the cut is lost and must be restarted.

Use of flame gouging

Certain specifications, such as those for fabrication of butt welded tanks, etc., stipulate that the underbead (or back bead) should be removed and a sealing run laid in place. This can easily and efficiently be done by flame gouging, as also can the removal of weld defects, tack welds, lugs, cleats, and also the removal of local areas of cracking in armour plate, and flashing left after upset welding (Fig. 7.13).

Oxy-arc cutting process

In this process the electric arc takes the place of the heating flame of the oxy-acetylene cutter. The covered electrodes of mild steel are in 4 sizes and are of tubular construction, the one selected depending upon whether it is required for cutting, piercing, gouging, etc. They are about 5 to 6 mm outside diameter with a fine hole about 1.5 mm diameter or more through which the cutting oxygen stream passes, down the centre. The gun type holder secures the electrode by means of a split collet and has a trigger controlling the oxygen supply. A d.c. or a.c. supply of 100–300 A is suitable.

The arc is struck with the oxygen off, the oxygen valve is released immediately and cutting begins, the electrode being held at an angle of 60° to the line of cut, except at the finish when it is raised to 90°, and is

Fig. 7.12. Method and uses of flame gouging.

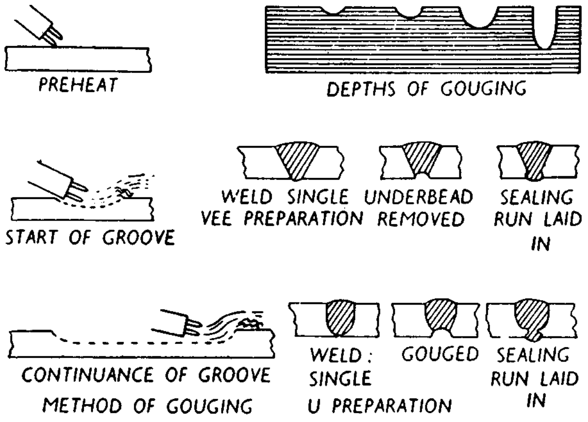
PREHEAT
DEPTHS OF GOUGING
START OF GROOVE
WELD SINGLE VEE PREPARATION
UNDERBEAD REMOVED
SEALING RUN LAID IN
CONTINUANCE OF GROOVE
METHOD OF GOUGING
WELD: SINGLE U PREPARATION
GOUGED
SEALING RUN LAID IN

consumed in the process. The oxygen pressure varies with the thickness of steel and with the size of electrode being used, about 4 bar (60 lb per in^2) for mild and low alloy steel plate of 8 to 10 mm thick and about 4.5–5.5 bar (70–80 lb per in.^{2}) for 23 to 25 mm thick. In addition to mild and low-alloy steel, and stainless steel copper, bronze, brass, monel and cupro-nickels can be roughly cut by this process.

Arc-air cutting and gouging process

In this process a carbon arc is used with a d.c. supply from a welding generator, or rectifier, together with a compressed air supply at 5–8 N/mm^2.

The equipment comprises a holder for the carbon electrode to which is supplied the first current with *electrode* +ve, and the compressed air which is controlled by a lever-operated valve on the electrode holder.

The jaws of the electrode holder are rotatable enabling the carbon electrode to be held in any position and twin jets of compressed air emerge from the head on each side of the carbon and parallel to it, the two jet streams converging at the point where the arc is burning. Air is also circulated internally in the holder to keep it cool.

Operation

The carbon electrode is placed in the holder with 75–150 mm projecting, with the twin jet holes pointing towards the arc end of the carbon.

The carbon is held at approximately 45° to the job to be cut while for shallow gouging the angle may be reduced to 20°. This angle together with the speed of travel affects the depth of cut. The carbons have been specially designed for the process and are a mixture of carbon and graphite covered over all with a thin sheath of copper. The copper coating prevents tapering

Fig. 7.13. Flame-gouged edge of plate, prepared for welding.

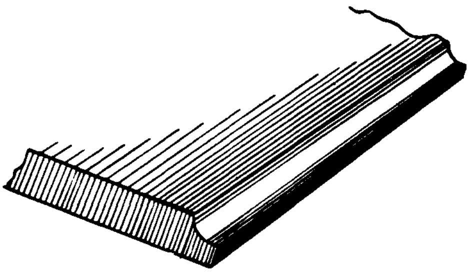

and ensures a cut of regular width in addition to enabling higher current to be used and consequently greater speeds of cutting to be obtained.

The carbons range in diameter from 4 mm (75–150 A) to 16 mm (550–700 A) and the higher the current density the more efficient the operation. Too high a current for a given size of electrode destroys the copper coating and burns the carbon at an excessive rate. Normally the copper coating burns away about 20 mm from the arc.

The process is applicable to work in all positions and is used for removing defects in castings, in addition to cutting and gouging. A reasonably clean surface of cut is produced with no adverse effect so that welding can be carried out on the cut surface without further grinding.

Cutting (and gouging) with the carbon arc. The carbon arc, owing to its high temperature, can be used for cutting steel. A high current is required, and the cut must be started in such a spot that the molten metal can flow away easily. The cut should also be wide enough so that the electrode (of carbon or graphite) can be used well down in it, especially when the metal is thick, so as to melt the lower layers. Cast iron is much more difficult to cut, since the changing of the iron into iron oxide is not easily performed owing to the presence of the graphite. Carbon arc cutting does not produce a neat cut, and because of this is only used in special circumstances. Gouging is performed with a very high current and the carbon is held at an angle of 45° to the work. This process is little used as it is very noisy.

Plasma cutting

(see also pp. 190–3, Ions and plasma, and Appendix 14 on lasers)

This process gives clean cuts at fast cutting speeds in aluminium, mild steel and stainless steel. All electrically conducting materials can be cut using this process, including the rare earth metals and tungsten and titanium and it has superseded powder cutting.

The arc is struck between the electrode and the conducting body of the torch. A gas mixture or air passes under pressure through the restricted nozzle orifice and around the arc, emerging as a high-temperature (up to 25000 °C) ionized plasma stream, and the arc is transferred from the nozzle and passes between the electrode and work (Fig. 7.14).

Great improvements have been made in plasma cutting torches by greatly constricting the nozzle and thus narrowing the plasma stream, giving a narrower and cleaner cut with less consumption of power. In certain cases 'double arcing' may occur, in which the main arc jumps from tungsten electrode to nozzle and then to work, damaging both nozzle and electrode.

Power unit

A 24 kVA fan cooled unit has input voltages of 220, 380–415, and 500 and cutting currents ranging from 50 A minimum to 240 A maximum at 100% duty cycle and OCV of 200 V d.c. with heavy duty rectifiers for main and pilot arc control. An automatic pilot arc switch cuts off when the cutting arc is transferred. This size of unit enables metal of up to 40 mm thickness to be cut and edges prepared quickly and cleanly in stainless steel, alloy steels, aluminium and copper (Fig. 7.15).

A 75 kVA unit of similar input voltage has cutting currents of 50–250 A with OCV of 300 V and cutting voltage of 170 V enabling thicknesses of 2.7 to 50 mm upwards, in the previously mentioned materials, to be cut and edges prepared at high speed. The cutting head of either unit can be fitted to a carriage to obtain high speeds of cut with greater precision than with manual cutting.

The gases used are argon, hydrogen, nitrogen, and it should be noted that since high voltages of up to 300 V on open circuit and about 170 V during the cutting process, are encountered, great care should be taken to follow all safety precautions and avoid contact with live parts of the torch when the unit is switched on. All work to be done on the torch head should be with the power supply switched off.

Argon–hydrogen and argon–nitrogen are used for cutting. Any combination of the gases can be selected at will to suit the material and thickness and the actual ratio of the gases will depend upon the operator. The tungsten electrode may be of 1.6 or 2.0 mm diameter and gives a cut of width about 2.5–3.5 mm with a torch-to-work distance of about 10 mm when cutting with about 250 A (the work is +ve).

Since hydrogen is an explosive gas and nitrogen combines with oxygen of the atmosphere to form the oxides of nitrogen (NO, N_2O, and N_2O_4) in the heat of the arc, cutting should be done in a well-ventilated shop (not enclosed in any way) and protective clothing and the correct eye lens protection always worn, with earmuffs for longer periods.

Operation. The unit is switched 'on', current is selected on the potentiometer knob and checked on the ammeter, gases are adjusted for mixture and the torch switch pressed. The pilot arc gas flows and the pilot arc is ignited. The torch is now brought down to about 10 mm from the work (take care not to touch down) and, making sure that all protective clothing including head mask is in order, the torch switch is released and the main arc is transferred to the work. If this does not occur a safety cut-off circuit ensures that the pilot arc current circuit is de-energized after a few seconds.

The arc commences with slow start current and when the thickness of work is pierced, the torch is moved along the plate and if cutting current and speed of travel are correct a steady stream of molten metal flows from the underside of the cut giving a smoothly cut surface. Average currents are in the region of 250 A.

The arc is extinguished automatically at the end of the cut or the torch can be pulled away from the work.

Torch (Fig. 7.14). These can be either hand or machine operated and are supplied with spare electrodes, cutting tips and heat shields. If water-cooled torches of either air or mixed gas are used electrolytic action occurs in the waterways of the torch, due to dissolved minerals in the water. Corrosion occurs and the torch is severely damaged. To prevent this, de-ionized water (about 2 litres per minute) can be used (see p. 346).

Fig. 7.14. Plasma cutting torch.

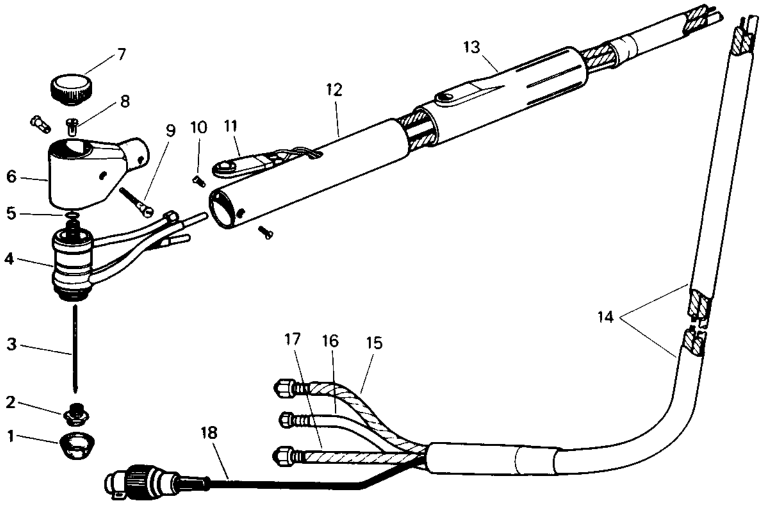

Key:

1. Heat shield.
2. Cutting tip.
3. Tungsten electrode 3.2 mm diameter.
4. Torch body.
5. O ring
6. Torch head cover.
7. Torch cap.
8. Collet.
9. Cover retaining nuts and screw.
10. Screw.
11. Torch switch.
12. Handle.
13. Switch boot.
14. Sheath.
15. Main arc cable.
16. Gas hose.
17. Pilot arc cable.
18. Control cable with plug.

Fig. 7.15. (*a*) Control panel for power unit for air (gas) plasma cutting.

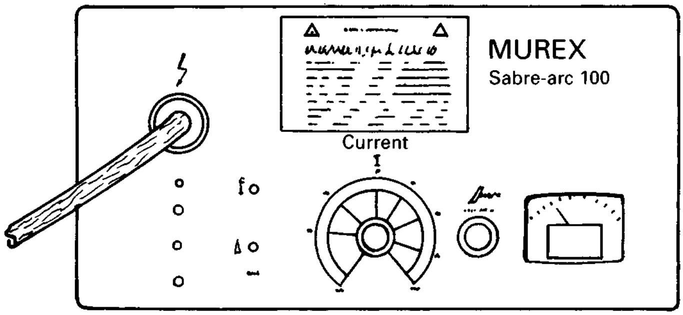

Models available:

Sabre-arc Model	Input V	OCV output	Cutting current	Thickness range mild steel	Rating	Plasma cutting gas	Cooling gas
40i Inverter weight 21 kg	380/415 1(2) phase	200 d.c.	10–40 A	0.5–15 mm	40 A at 60%	Air	Air
100	220/380/415	370 d.c.	25–100 A	1–25 mm	100 A at 100%	Air or N_2 or ArH_2	Air or N_2 or CO_2 or O
150	220/380/415	370 d.c.	25–150 A	1–37 mm	150 A at 75%	Air or N_2 or ArH_2	Air or N_2 or CO_2 or O_2

Fig. 7.15. (*b*) Sabre-arc unit (left); WTC (Wigan Torch Co.) unit 100–150 A (right).

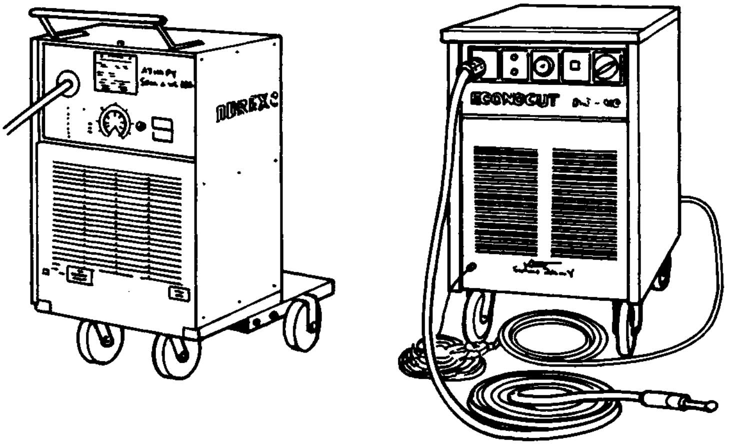

Air plasma cutting

In this process, instead of using argon, nitrogen and hydrogen as with the tungsten arc, air is used in its compressed form for both plasma gas and cooling. The high plasma temperature provides rapid cutting with little distortion. Sections up to 30–40 mm thick can be cut in steel, low alloy steel, carbon steel, stainless steel, aluminium and bronze. The process is particularly useful for cutting thinner sections. Air contains oxygen, so rapid oxidation takes place within the torch unless steps are taken to reduce it. The tungsten electrode is therefore dispensed with and, instead, electrode and tip are made as small as possible so that they can be discarded when eroded, and are the consumables in this process. The tip has a body of copper into which a small pellet of hafnium wire, with a fine hole down the middle, is pressed centrally in the nose (Fig. 7.16*d*). Hafnium has a lower melting point (2200 °C) than that of tungsten (3400 °C) but it forms a conductive coating with the oxygen which increases its life due to its higher breakdown temperature. An air tube of various sizes controls the amount of air to the tip. The tip fits into the electrode but does touch it (the electrode is the cathode). In addition, as with plasma welding, the work piece, connected to the return cable to the power unit, is also the anode (Fig. 7.16*c*).

Air is supplied from compressor or cylinders to the torch at a pressure of 5–10 bar and with a volume of about 150 litres per minute. The air must be very finely filtered, dry and free from any oily residues. Of the air fed to the torch head, some passes through the tip and the remainder cools the rear of the tip and, with the rest of the air, circulates through the torch head, cooling it and leaving through a small space between heat shield and diffuser (shroud) (Fig. 7.16). The air which passes through the electrode tip emerges at high velocity to form the plasma stream when ionized.

The power unit is similar to that for gas tungsten cutting (Fig. 7.15). It supplies two electrical circuits. One is supplied at 300 V d.c. open circuit, falling to 100–150 V drop across the plasma when using a current of 20–150 A.

The other circuit is of high frequency (HF) about 4 kV, with a d.c. component of 300 pulses per second. When the switch on the torch handle is pressed, the air gap between tip (nozzle) and electrode breaks down, establishing an arc between them (the pilot arc for ionizing the gap), the air flows through and becomes the plasma stream, for transfer to the cut in the work.

Operation. To operate the unit, the air flow tube, air electrode (jet) and

Fig. 7.16

(*a*) Sabre arc torch

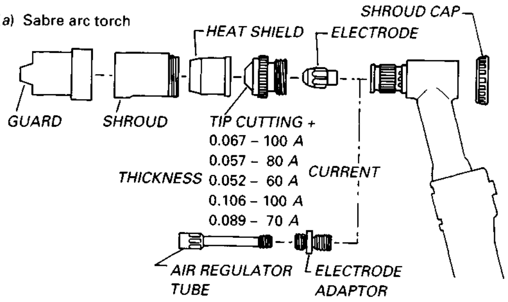
SHROUD CAP
HEAT SHIELD
ELECTRODE
GUARD
SHROUD
TIP CUTTING +
THICKNESS
0.067 – 100 *A*
0.057 – 80 *A*
0.052 – 60 *A*
0.106 – 100 *A*
0.089 – 70 *A*
CURRENT
AIR REGULATOR TUBE
ELECTRODE ADAPTOR

(*b*) WTC torch

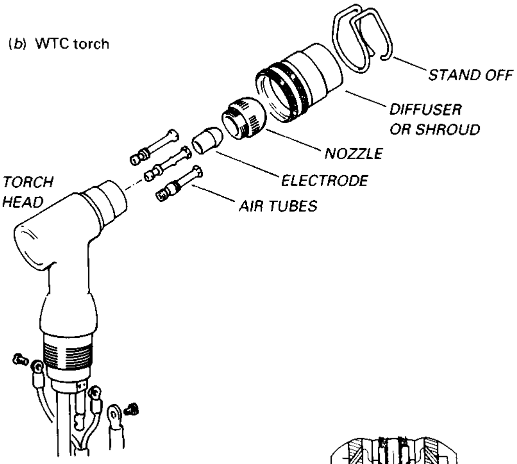
STAND OFF
DIFFUSER OR SHROUD
NOZZLE
ELECTRODE
AIR TUBES
TORCH HEAD

(*c*) Method of obtaining narrow width of cut

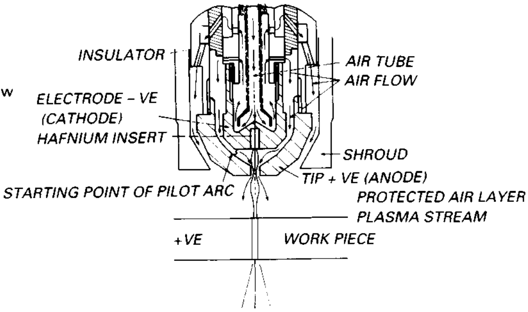
INSULATOR
AIR TUBE
AIR FLOW
ELECTRODE – VE (CATHODE)
HAFNIUM INSERT
SHROUD
TIP + VE (ANODE)
STARTING POINT OF PILOT ARC
PROTECTED AIR LAYER
PLASMA STREAM
+VE
WORK PIECE

(*d*) Maximum wear in hafnium electrode tip

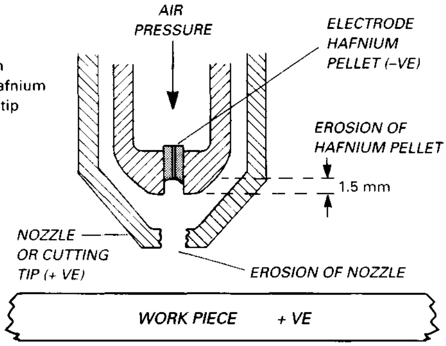
AIR PRESSURE
ELECTRODE HAFNIUM PELLET (–VE)
EROSION OF HAFNIUM PELLET
1.5 mm
NOZZLE OR CUTTING TIP (+ VE)
EROSION OF NOZZLE
WORK PIECE + VE

tip and the amperage required for the thickness to be cut must be chosen. See that the flow of air is unhindered. The torch head is now brought down and held over the plate to be cut. The switch on the gun handle is pressed, the pilot arc established and the gap ionized. The torch head is now brought down to the correct height, as determined by the stand off guide or guard at the plate edge. The plasma is transferred to the work (it is of positive polarity) and the cut commences. The torch head is moved along so that the cut is smooth and continuous and there is a slight trailing angle as it passes through the metal. At the end of the cut the trigger is released and the plasma extinguished, but the air continues to flow for about a minute in order to cool the torch head. This is controlled by a control circuit in the power unit.

The life of the electrode decreases with increasing current. At 150 A for 40 mm thick plate the life is about 30 min, while at 50 A for 15 mm thick plate the life is about 110 min. In any case the electrode should be changed when the insert has eroded to a depth of about 1.5 mm (Fig. 7.16*c*). The nozzle or tip has a longer life but should be changed if the orifice is damaged or enlarged. A loss of cutting power may be the result of melting down due to over-use and is accompanied by a green flame.

A ceramic or high-temperature plastic assembly covers the tip (nozzle) to protect it from spatter and a heat shield is often included between tip and shroud. Air passing through the narrow gap between nozzle and diffuser cools the torch head and guides the cooling air into the cut. Attachments, supplied as extras, can be fitted to the torch to give a longer arc for plasma gouging, this method being quieter than the carbon arc method.

Multiple starts cause rapid erosion of the hafnium pellet, and the use of higher currents has the same effect. The higher the current the shorter the life of the tip (nozzle). Remember that tip and electrode are consumables and should be examined constantly for wear. Most complaints about this process are due to trying to operate the tip or electrode (or both) after their efficient life is past.

Examine tip and electrode and replace or clean, if eroded.

De-ionized water

Water taken directly from a tap contains dissolved carbon dioxide (CO_2) with possibly hydrogen sulphide (H_2S) and sulphur dioxide (SO_2) together with mineral salts such as calcium bicarbonate and magnesium sulphate which have dissolved into the water in its passage through various strata in the earth. These minerals separate into ions in the water, for example, magnesium sulphate $MgSO_4$ separates into Mg^{2+} ions (positively charged) and sulphate ions SO_4^{2-} which are negatively charged. Similarly with calcium bicarbonate ($Ca(HCO_3)_2$) which gives metallic ions, Ca^{2+} and bicarbonate ions HCO_3^{2-}.

By boiling or distilling the water the minerals are not carried over in the steam (the distillate) so that the distilled water is largely free of mineral salts but may contain some dissolved CO_2. Distillation is however rather slow and costly in fuel and a more efficient method of obtaining de-ionized water is now available and uses ion exchange resins. The resins are organic compounds which are insoluble in water. Some resins behave like acids and others behave like alkalis and the two kinds can be mixed without any chemical change taking place.

When water containing dissolved salts is passed through a column containing the resins the metal ions change places with the hydrogen ions of the acidic resin so that the water contains hydrogen ions instead of metallic ions. Similarly the bicarbonate ions are replaced by hydroxyl (OH^-) ions from the alkaline resin. During this exchange insoluble metallic salts of the acid resin are formed and the insoluble alkali resin is slowly converted into insoluble salts of the acids corresponding to the acid radials previously in solution.

The water emerging from the column is thus completely de-ionized and now has a greater resistivity (resistance) than ordinary tap water so that it conducts a current less easily. A meter can be incorporated in the supply of de-ionized water to measure its resistivity (or conductivity) and this will indicate the degree of ionization of the water in the cooling circuit. When the ionization rises above a certain value the resins must be regenerated. This is done by passing hydrochloric acid over the acidic resin, so that the free hydrogen ions in the solution replace the metallic ions (Ca^{2+} and Mg^{2+}). This is followed by passing a strong solution of sodium hydroxide (NaOH) through the column, when the hydroxyl ions displace the sulphate and bicarbonate ions from the alkaline resin.

This process produces de-ionized water, purer than distilled water, easily and quickly.

Water injection plasma cutting

In this process, water is injected through four small-diameter jets tangentially into an annular swirl chamber, concentric with the nozzle, to produce a vortex which rotates in the same direction as the cutting gas (Fig. 7.17). The water velocity is such as to produce a uniform and stable film around the high-temperature plasma stream, constricting it and reducing the possibility of double arcing. Most of the water emerges from the nozzle in a conical jet which helps to cool the work surface.

The cut produced is square within about 2° on the right-hand side (viewed in the direction of cutting), whilst the other side is slightly bevelled, caused in general by the clockwise rotation of the cutting gas, which is commercial purity nitrogen. The use of nitrogen reduces cutting costs as it is cheaper than gas mixtures.

The process gives accuracy of cut at high cutting speeds with very smooth cut surfaces. There is little or no adherent dross and the life of the cutting nozzles is greatly increased; mild and carbon steels, alloy and stainless steels, titanium and aluminium are among the metals which can be efficiently cut, in thicknesses from 3 to 75 mm for stainless steel.

The noise and fumes associated with plasma cutting can be reduced by the use of a water muffler fitted to the torch to reduce the noise and by the use of a water table which replaces the normal cutting table and which removes up to 99.5% of the particles and fumes by scrubbing, using the kinetic energy of the hot gases and molten metal stream from the kerf.

This cutting equipment is currently fitted to existing cutting installations up to the largest sizes used in shipyards and including those with numerical control.

Fig. 7.17. The principle of water-injection plasma cutting.

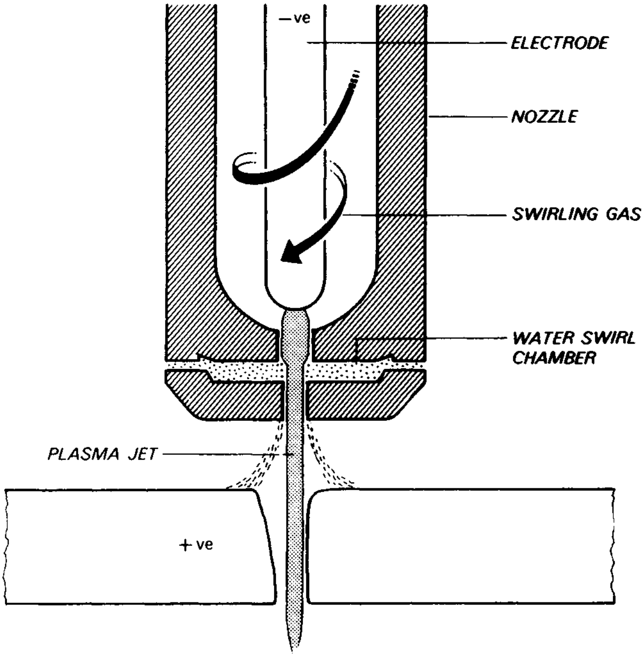

Powder injection cutting

The process of powder injection cutting causes a great deal of fumes to be given off and is one of the less liked operations. As a result it has, in many cases, been superseded by the plasma arc method of cutting of stainless steels, 9% nickel steel and non-ferrous metals. Powder cutting enables stainless steel to be cut, bevelled and profiled with much the same ease and facility as with the oxygen cutting of low-carbon steels.

In cutting low-carbon steels a pre-heating flame raises the temperature of a small area to ignition point. This is the temperature at which oxidation of the iron occurs, and iron oxide is formed when a jet of oxygen is blown on to the area. The heat of chemical combination, together with that of the pre-heating flame enables the oxidation, and hence the cut to continue, the oxide being removed by being blown away by the oxygen jet, resulting in a narrow cut.

For this sequence to occur, the melting point of the oxide formed must be lower than that of the metal being cut. This is the case in low-carbon steel. In the case of stainless steels and non-ferrous metals the oxides formed have a melting point higher than that of the parent metal.

When attempting to cut stainless steel with the ordinary oxy-acetylene equipment the chromium combines with oxygen at high temperature and forms a thin coating of oxide which has a melting point higher than that of the parent steel, and since it is difficult to remove, further oxidation does not occur and the cut cannot continue.

In the powder-cutting process a finely divided iron powder is sprayed by compressed air or nitrogen into the cutting oxygen stream on the line of the cut. The combustion of this iron powder so greatly increases the ambient temperature that the refractory oxides are melted, fluxed, and to a certain extent eroded by the action of the particles of the powder, so that a clean surface is exposed on to which the cutting oxygen impinges and thus the cut continues. The quality of the cut is very little inferior to that of a cut in a low-carbon steel.

Equipment

For hand or machine cutting the powder is delivered to the reaction zone of the cut by means of an attachment fitted to the cutting blowpipe (Fig. 7.18). The attachment consists of powder valve, powder nozzle and tubing. The nozzle is fitted over the standard cutting nozzle and the powder valve is clamped near the gas valves. The iron powder is carried down the outside of the cutting nozzle and after passing through inclined ports, is injected through the heating flame into the cutting stream of oxygen which it meets at approximately 25 mm below the end of the cutting nozzle.

The nozzle is normally one size larger than for cutting the same thickness of low-carbon steel and is held as for normal cutting except that a clearance of 25 to 35 mm is given between nozzle and plate to be cut to allow the powder to burn in the oxygen stream. The great heat produced makes preheating unnecessary on stainless steel and what may be termed a 'flying start' can be made.

The powder dispenser unit (Fig. 7.19) is of the injector type and is a pressure vessel which incorporates:

(1) A hopper for filling.
(2) An air filter.
(3) An air pressure regulator.
(4) A dryer.
(5) An injector unit.

The removable cover enables the hopper to be filled and is fitted with a pressure relief valve which lifts at 0.15 N/mm². A screen for removing overlarge particles from the powder and a tray for the drying agent are fitted.

Compressed air flows into the dispenser, picks up powder and carries it along a rubber hose to the cutting nozzle. Nitrogen may be used instead of air but never use oxygen, as it is dangerous to do so. The powder flow from the dispenser is regulated by adjusting the nozzle or by varying the air pressure.

The dispenser should be fed from an air supply of 1.5 m³ per hour at a pressure of 0.3 N/mm². The usual pressure for operating is from 0.02 to 0.03 N/mm².

Since the whole process depends upon a uniform and smooth flow of powder every care must be taken to ensure this. Any moisture in either the powder or in the compressed air can cause erratic operation and affect the quality of the cut.

Silica gel (a drying agent) is incorporated in the dispenser to dry the powder but the amount is not sufficient to dry out the compressed air and a

Fig. 7.18. Cutter equipped for powder cutting.

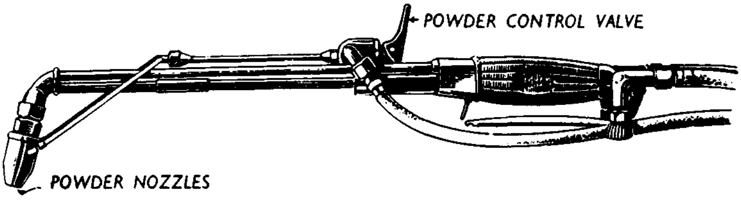

separate drying and filtering unit should be installed in the air line to ensure dry, clean air. The oxygen and acetylene is supplied through not less than 9 mm bore hose with 6 mm bore for the powder supply.

The single-tube attachment discharges a single stream of powder into the cutting oxygen and is used for straight line machine cutting of stainless steel. The multi-jet type (Fig. 7.18) has a nozzle adaptor which fits over a standard cutting nozzle and has a ring of ports encircling it. The powder is fed through these ports and passes through the heating flame into the cutting oxygen. This type is recommended for hand cutting and for profile, straight and bevel machine cutting. Special cutting nozzles are available for this process. These give a high velocity parallel cutting oxygen stream from their bore being convergent–divergent. The pre-heater holes are smaller in number but more numerous and are set closer to the cutting oxygen orifice than in the standard nozzle. This gives a soft narrow pre-heating flame, giving a narrower cut and better finish with a faster cutting speed. Since the iron powder is very abrasive, wear occurs in ports and passages through which the powder passes. By using stainless tube where possible, avoiding sharp internal bends and reinforcing certain parts, wear is reduced to a minimum.

Fig. 7.19. Powder dispenser unit.

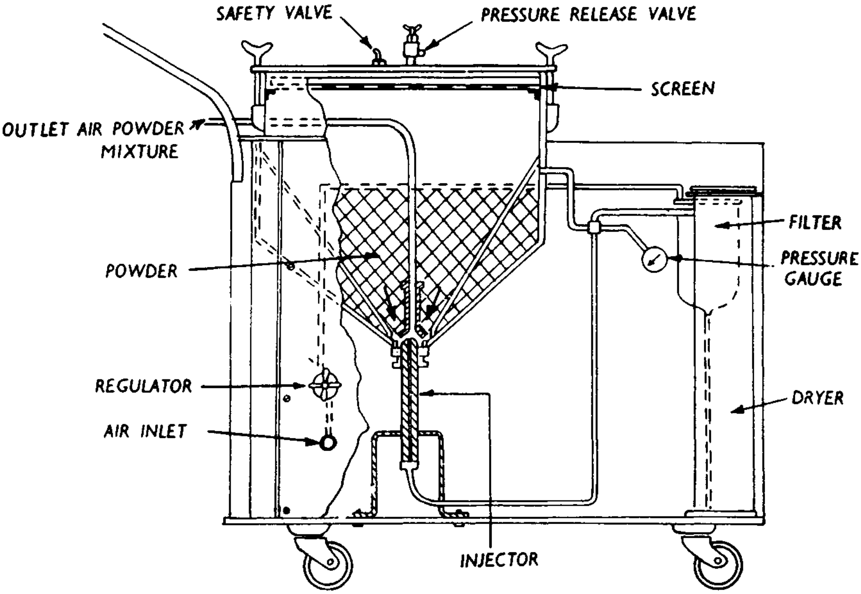

Technique

In cutting any metal by this process, correct dispenser setting, dry air and powder, clean powder passages and leak-tight joints all help towards ensuring a good quality cut. The rate of powder flow is first adjusted to the correct amount for the particular work in hand by trial cuts.

Stainless steel. Size of nozzle is one size larger than for the same thickness low-carbon steel. For thicknesses up to 75 or 100 mm the nozzle is held about 25 mm from the plate, for thicknesses up to 150 mm the distance is increased to about 35 mm, while for heavier sections it can be about 50 mm distant. No pre-heating is necessary – an immediate or flying start can be made. Scrap plate may be loosely put together and a single cut can be made, the great heat enabling this to be done in spite of gap between the plates.

Cast iron and high-alloy steels. Technique for cast iron is similar to that for stainless steel but cutting speed is up to 50% slower. With high-alloy steels for example, a 25/20 nickel–chrome steel takes 30–40% longer than for the 18/8 nickel–chrome steel. Pre-heating of high-alloy tool steels is advisable to prevent risk of cracking due to localized heat.

Round carbon steel bars are easily cut by laying them side by side. The cut is done from bar to bar, powder being switched on to start the cut on each bar giving no interruption in cut from bar to bar.

Nickel and nickel alloys. Pure nickel is extremely difficult to cut because it does not readily oxidize and it can only be cut at slow speed in thicknesses up to 25 mm section, whereas INCONEL® and NIMONIC® alloys can be cut up to 125 mm thick, speeds being up to 60% slower than for the same thickness of stainless steel.

Copper and copper alloys. In this operation copper is melted (and not cut) the particles of iron powder removing the molten metal and eroding the cut.

Much pre-heating is required due to the high conductivity of copper and copper alloys – for example – a powder cutting blowpipe capable of cutting 180 mm thick stainless steel can only deal with copper up to 25 mm thick and brass and bronze up to 100 mm thick. A lower cutting oxygen pressure than normal is used and the nozzle moved in a forwards, upwards, backwards, downwards motion in the line of the cut, thus helping to remove the molten metal and avoiding cold spots.

Aluminium and aluminium alloys. The quality of the cut is poor and of ragged profile. Alloys containing magnesium such as MG7 develop a hard surface to the cut due to formation of oxide and this may extend to a depth of 16 mm. As a result powder cutting is limited to scrap recovery. Cutting (or melting) speeds on thinner sections are much the same as for stainless steel, but with thicker sections molten metal chokes the cut, so that large nozzles with reduced cutting oxygen pressures should be used.

Powder cutting can be incorporated on cutting machines and gouging can be carried out as with the normal blowpipe with the same ease and speed, the nozzle being again held further away from the work, to allow space for the combustion of the powder.

Underwater cutting

Underwater cutting of steel and non ferrous metals is carried out on the oxy-arc principle using the standard type of electrode which has a small hole down its centre through which the pressurized oxygen flows to make the cut. By using a thermic electrode, this oxygen stream causes intense oxidation of the material of the electrode, giving a greatly increased quantity of heat for the cut. In both cases the products of the cut are removed or blown away by the oxygen stream.

A d.c. generator is used with the torch negative and the special type earth connexion, connected to clean, non-rusted metal in good electrical contact with the cut, is positive polarity: a.c. is not suitable as there is increased danger of electric shock.

The torch head for ordinary or thermic electrodes is similar. It is tough rubber covered and has connections for oxygen and electric current. The electrodes, usually of 4 mm ($\frac{5}{32}''$) and 4.8 mm ($\frac{3}{16}''$) diameter are securely clamped in the head by a collet for the size electrode being used. A twist of the torch grip clamps and releases the collet and is used for stub ejection. These collets can be changed to allow the torch head to be used for welding (Fig. 5.23).

The oxygen cylinder, fed from a manifold, has a high volume regulator which will give free flow and pressures from 8 bar (118 psi) at 12 m depth to 20 bar (285 psi) at 107 m. In general the oxygen supply to the torch should give about 6 bar (90 psi) over the bottom pressure at the electrode tip.

Ingress of water to the torch head is prevented by suitable valves and washers and replaceable flash arrester cartridge and screen are fitted.

The seal for the electrode in the torch head is made with washers and collet and tightened in place with the torch head locknut. When the electrode is fitted to the head it must bottom on to the washer and be held tight to prevent leakage of oxygen. Note that the thermic type of electrode will continue burning once it has been ignited, even when the electrical supply is switched off, due to the oxidizing action of the gas stream.

Use of equipment

An eyeshield with the correct lens fitted is attached to the outside of the diver's helmet. No oil nor grease should be used on the equipment and there should be no combustible nor explosive materials near to the point where cutting (or welding) is to be performed. Hose connexions should be checked for leaks and all electrical connexions tight, especially check the earth connexion and see that it is in good electrical connexion with the position of the cut. As electrolysis can cause rapid deterioration, all equipment should be continually inspected for signs of this. If any part of the work is above water level, connect the earth clamp to this after checking that there is a good return path from the cut to the earth clamp.

A double pole single throw switch of about 400 A capacity is connected in the main generator circuit as a safety switch. This should always be kept in the 'off' position except when the cut is actually taking place. To begin cutting, strike the arc and open the oxygen valve by pressing the lever on the torch.

When working, the diver-welder must always face the earth connexion so that the cutting is done *between* the diver and the earth connexion. When the electrode has burnt to a stub of about 75 mm in length the diver should call to the surface to shut off the current by opening the safety switch. This having been done the collet is loosened by a twist of the wrist, the oxygen lever is pressed and the stub is blown out of the holder. A new electrode is now fitted, making sure that it sits firmly against the sealing washer, and the handle twisted to lock it in place. Stubs should not be burnt close to the holder for fear of damage. When the safety switch is again placed in the 'on' position cutting can again commence.

When cutting, bear down on the torch so as to keep the cutting electrode against the work so that the electrode tip is in the molten pool and proceed with the cut as fast as possible consistent with good cutting. Spray back from the cut indicates that it is not through the work completely. Keep all cables and hoses from underfoot and where anything may fall on them. Pipes and tubes may be cut accurately to size under water by a hydraulically operated milling machine which is driven around a circumferential guide.

Safety precautions

The general safety precautions which should be observed when welding are listed at the end of the section on oxy-acetylene welding but special precautions should be taken when cutting processes using oxygen are used in the welding shop or confined spaces. Oxygen is present in the atmosphere at about 29.9% by volume and below 6–7% life cannot be sustained.

In Appendix 8 on fire precautions we state that for a fire to occur we need (1) oxygen, (2) fuel and (3) heat.

Oxygen rapidly oxidizes combustible materials and in the case of grease or oil the oxidation reaction produces so much heat that the material may be ignited, thus causing fire or explosion. When cutting with the oxy-acetylene or oxy-propane flame the oxygen (cutting) lever should not be kept open longer than necessary because not only is the compressed gas expensive but it also causes oxygen enrichment of the atmosphere in confined spaces. Should the operator be wearing gloves and clothing that are greasy or oily, the oxidation of this grease and oil may cause burning with evolvement of heat so that the operator may sustain burns and a fire or explosion may endanger life.

Never use compressed oxygen for any other purpose (e.g. in place of compressed air) other than cutting (or welding). Make sure that the ventilation of the area is good – cutting in confined spaces always presents a hazard. Oil, grease and other combustible material should never be in the vicinity where cutting is taking place.

Since oxygen has no smell or taste its presence is difficult to detect and no smoking should be allowed in areas where oxygen is being used.

8

The welding of plastics

Plastics

Plastics are replacing steel and other metals used in construction. They are now used in the home, in gas engineering, farming, sports equipment, and in the automobile industry etc. Equipment made of plastic is lighter than that made of steel, and since welding and fabrication is relatively easy and the cost of repair lower, it will be well to consider some of the processes involved.

Plastics are derived mainly from oil, coal and wood. Polystyrene and polyvinylchloride (PVC) are examples of those derived from oil and coal while the cellulose types (such as cellulose acetate and ethylcellulose) are produced from wood. Plastics consist of larger molecules built up into chains arranged randomly, from single molecules (monomers) by a process of polymerization, assisted by a catalyst. (A catalyst is a substance that accelerates a chemical reaction but undergoes no permanent change during the reaction.) If differing monomers are used, copolymers are produced from which the characteristics of the original monomers cannot be obtained.

There are two main types of plastic: (1) thermo-setting and (2) thermoplastic. A thermo-setting plastic is one that softens when first heated and is then moulded into the required form. Upon cooling it sets hard and gains its full strength. Upon reheating it does not again soften and thus cannot be reshaped. An example of a thermo-setting plastic is that made from phenol and formaldehyde, which with an acid catalyst gives a soluble and fusible resin used in varnishes, and with an alkaline catalyst gives an infusible and insoluble resin (Bakelite).

Thermoplastic substances soften each time they are heated and can be reshaped again and again, regaining their rigidity and strength each time upon cooling. Foil, thin and thicker sections, fabrications, repairs to automobile components and welding in general, can be successfully carried out on thermoplastics.

Welding may be by: (1) external heating, i.e. hot air, hot plate, resistance and RF processes; (2) external movement, i.e. friction or spin, ultra-sonic and other vibrations (100–240 Hz); (3) adhesive, i.e. cyanoacrylates (super-glue). Method (1) is of the most interest to the welder, and hot air and hot plate will be considered in detail.

Table 1 gives the thermoplastics most commonly used for fabrication and in repair work, while Table 2 indicates tests which may be carried out to identify a plastic by removing a shaving from it and burning it in a low flame, carefully observing the results. It is essential that rod and parent plastic should be compatible. To ensure this the makers of plastic welding equipment supply small bundles of various rods as given in Table 1. Tests are carried out on the given plastic with each of the rods until one is found that fuses well and easily into the parent plastic. Compatibility of rod and plate cannot be stressed enough. Note, however, that flame or burn tests

Table 1. *Types of plastic suitable for welding*

Identification	Plastic	Welding temperature (°C)
ABS and ABS/PC	Acrylonitrile-butadiene-styrene and ABS/polycarbonate	350
PVC high density	Polyvinylchloride hard	300
PVC low density	Polyvinylchloride soft	400–500
PC	Polycarbonate and polycarbonate modified	350
PE high density (HDPE)	Polyethylene high density	300
PE soft density (SDPE) (polythene)	Polyethylene soft density	270
PP	Polypropylene	300
PA	Polyamide	400
PP/EPDE	Polypropylene 3-ethylene-diene rubber	300
PUR (polymer alloy)	Polyurethene (not all types are weldable)	300–350
PBT	Polybutylene tetraphthalate	350
Acrylics (e.g. PMMA)	Resins. Usually transparent, e.g. polymethyl methacrylate	350

Note. Glass-reinforced plastics are not generally weldable and foam-filled PU should not be welded.

are little used nowadays because of the difficulty in differentiating between plastics with the same burning characteristics.

Welding thin sheets

Thin sheets are welded by applying heat and pressure to the joint. The sheets are overlapped and then passed between two rollers; the width of roller determines the width of the weld. One of the rollers is motor driven and can be heated by an electrical element. The other roller is free rolling and serves only as a back pressure for the heated roller. The variables are therefore: (1) the temperature of the heated roller, (2) the speed at which the surfaces to be welded pass between the rollers, and (3) the pressure between the rollers. The two surfaces to be welded together pass when in the spongy state – just below melting. The heating element for the main roller has a potentiometer control and the element conducts its heat to the roller. To operate, the temperature is selected for the given plastic (e.g. PVC), for which welding temperature is 300–500 °C. The motor is set at the correct speed to rotate the driven roller, and finally the pressure between the rollers is set. If the current is too low (temperature), or the pressure and speed of travel between the rollers too high, the two sheets will not weld together. If the current is too high and the speed of travel too low, the sheets will be welded together only in certain parts and holes will appear. It will be appreciated that the three variables are notoriously difficult to adjust correctly. Modern machines greatly simplify this procedure. Note that the process is similar in many ways to seam welding (p. 208). If the roller is hand-operated (usually for simple welds only) a smooth surface, under the part to be welded, is fixed so that it takes the place of the bottom roller (Fig. 8.1*a*). Again the temperature of the heating element is set and the roller unit moved down the seam with a medium pressure. As before, if the speed is too high the sheets will not

Table 2. *Tests for identification of plastics*

Plastic	Observations
ABS	Sweet smell; burns with a black sooty flame not self-extinguishing
PC	Burns with black sooty smoke; not reliably self-extinguishing
PE	Smelly and feels like a wax; burns like a wax candle and drips
PP	Similar to PE; feels like wax and smells and drips like wax
PA	Smells like burnt horn; flame not self-extinguishing; stringy
ABS/PC	Sweet smell; black sooty smoke; flame not self-extinguishing
PVC	Burns with black smoke and an acrid smell; not self-extinguishing

be welded together, while too slow a speed results in an unsatisfactory weld with holes down the length of it. The weld must be mechanically sound and continuous. Fig. 8.1(*b*) shows a taped seam welded with hot air from a gun.

Hot air (or gas) welding

Guns or torches. The guns for this method of welding operate on the same principle as a hair dryer, in which a stream of air, propelled by a fan driven by an electric motor, flows over a heated electric element. They are available in various voltages and power loadings: 200–240 V, 1500 W, or 110–120 V, 1700 W and lower voltages if required. Hot gas guns were originally supplied with compressed carbon dioxide or nitrogen but most of them use air nowadays, although the term hot air gun includes hot gas guns.

Some guns have an electric motor in the body of the gun, from which

Fig. 8.1. (*a*) Sheet welding. (*b*) Welding of taped seam with hot air and pressure.

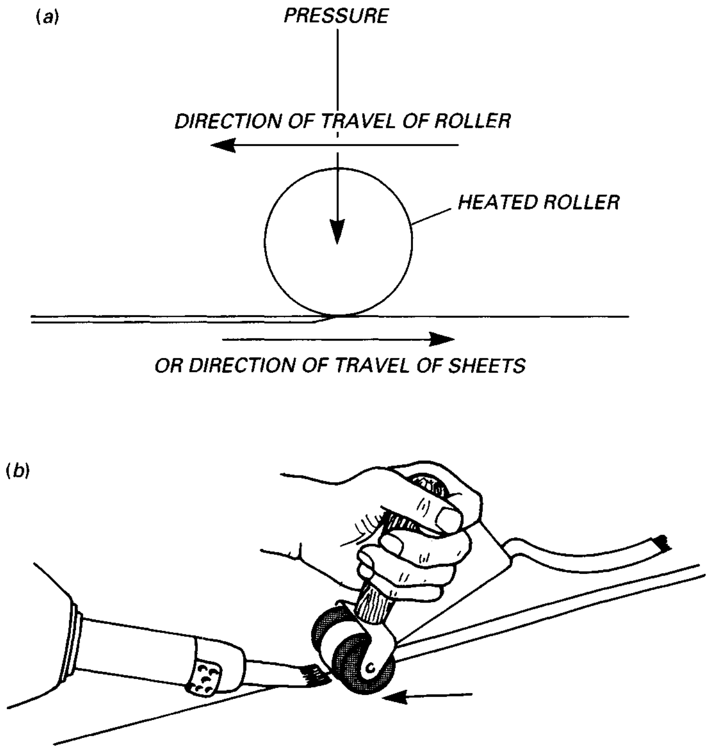
(a)
PRESSURE
DIRECTION OF TRAVEL OF ROLLER
HEATED ROLLER
OR DIRECTION OF TRAVEL OF SHEETS
(b)

the air is supplied, while others use air supplied from an external compressor. In the latter, there are two leads to the gun: the electricity supply for the heating element and the compressed air supply. This type of gun is lighter and of smaller diameter than the type with an electric motor in the body. However, the gun with the motor has the advantage that there is only one cable feeding the motor and element (Fig. 8.2). In both types the air is blown over an electric element, the temperature being accurately controlled (by potentiometer). There is a switch on the outside of the body of the gun, clearly marked so that any temperature between 300 °C and 600–700 °C can be selected for the air issuing from the gun. In some guns the heating element is easily replaceable. The volume of air passing over the element is controlled by a movable shutter over the air intake. The gun barrel gets very hot during use and should on no account be touched. Some guns have an outer shield over the barrel to prevent it being touched accidentally.

Nozzles. A variety of nozzles are available, all having either a tight push-on fit, locked on with a clamp ring and screw, or a screwed-on fit. These

Fig. 8.2. (*a*) Hot air torches for general use.

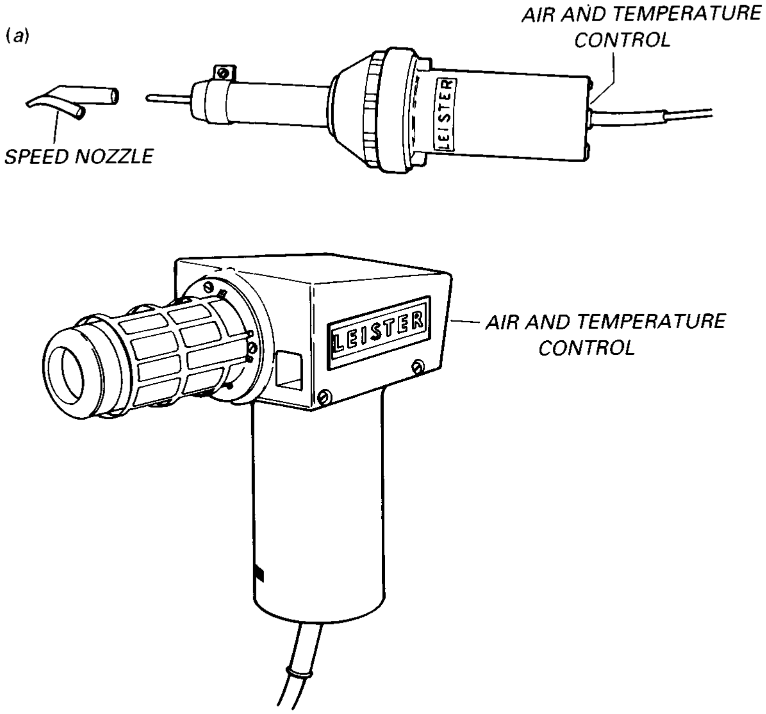

Fig. 8.2. (*b*) High speed torch or gun showing relative size (Goodburn).

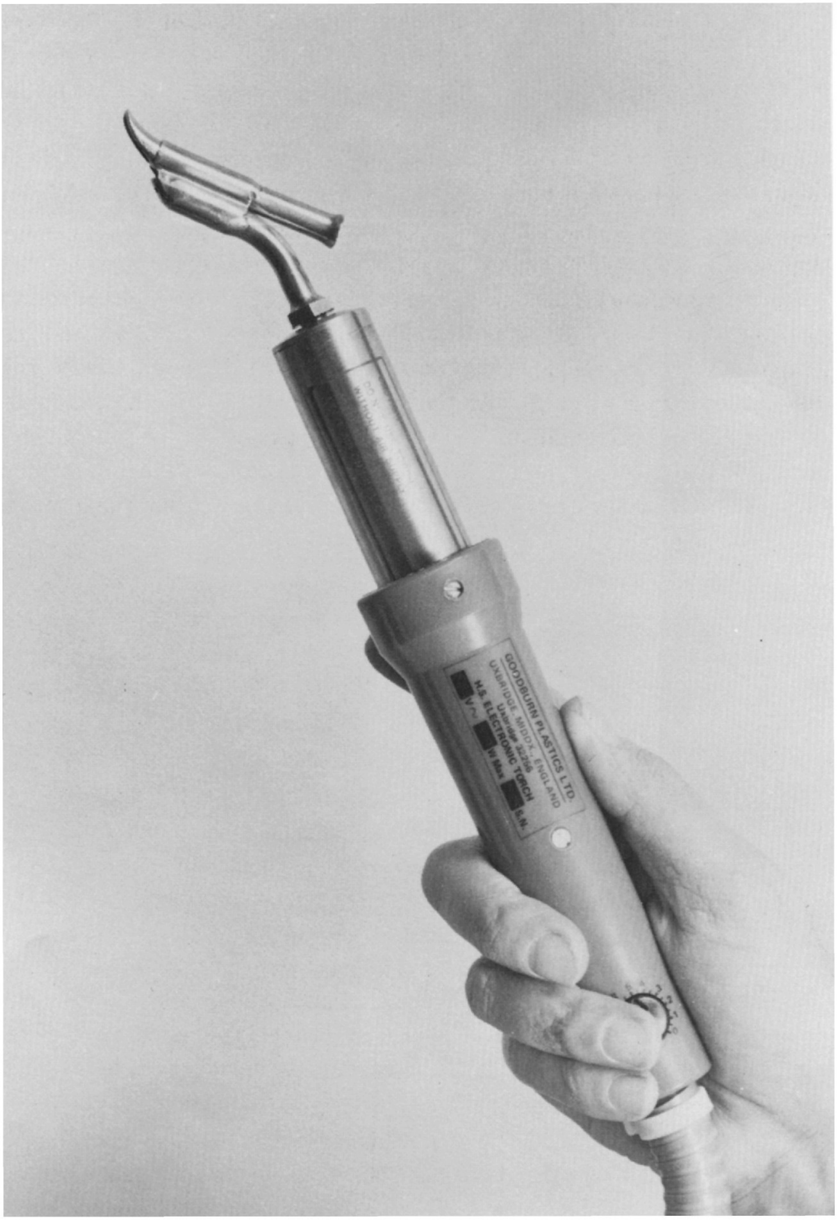
GOODBURN PLASTICS LTD.
UXBRIDGE, MIDDX., ENGLAND
Uxbridge 32266
H.S. ELECTRONIC TORCH
V~
W Max
S.N.

nozzles are obtainable with diameters 3 mm, 4 mm and 5 mm I/D. The following are the most important.

The tack welding nozzle has an open mouth for the hot air, and there is a blunt toe on its front or toe end (Fig. 8.3). The sheets to be welded are close butted together, the hot air from the gun softens the edges of the sheets for the preparation to welding condition and the blunt toe draws the edges together as the gun moves along the surfaces, with the angle of the base of the nozzle about 15–30° to the seam of the joint. The tack weld has little strength and should not be subjected to any load. It is for positioning only and to ensure fusion at the bottom of the weld.

The speed welding nozzle (Fig. 8.4) consists of two tubes – either round or triangular – joined at the bottom end. One tube is for hot air and the other for the rod (round, triangular (profiled) or tape as required). The unit is pushed or screwed onto the gun. The hot air tube has a small-diameter tube inside it down which the hot air passes. At the bottom end is a baffle plate with a 3 mm hole in it, and the bottom of the hot air tube finishes about 20 mm from the baffle so that during welding some of the

Fig. 8.3. Types of nozzle.

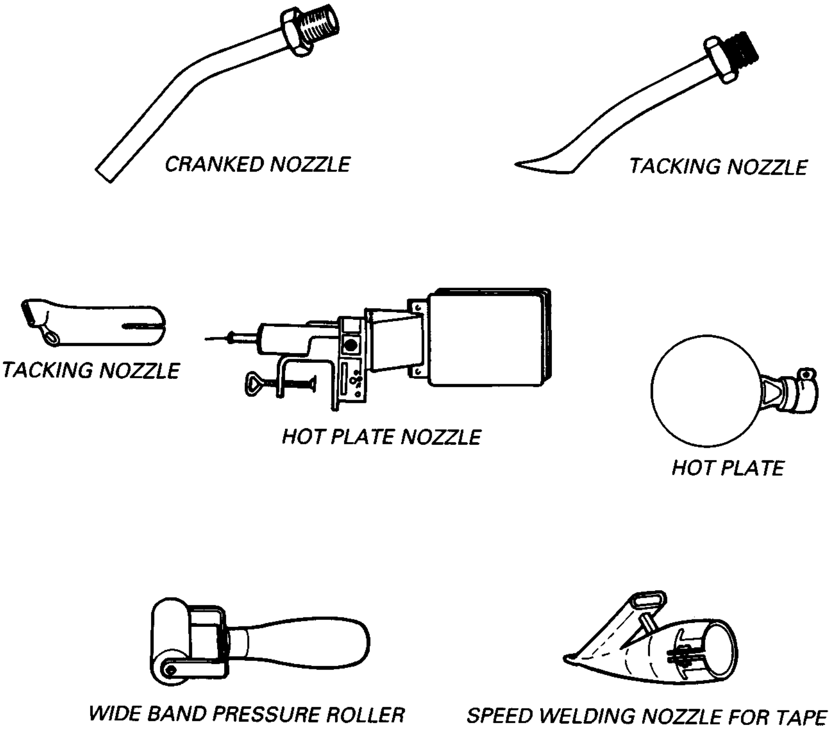
CRANKED NOZZLE
TACKING NOZZLE
TACKING NOZZLE
HOT PLATE NOZZLE
HOT PLATE
WIDE BAND PRESSURE ROLLER
SPEED WELDING NOZZLE FOR TAPE

air passes through the baffle hole and preheats the joint, while the rest of the hot air is available for the welding operation, enveloping the base of the rod and the edges of the preparation. The rod to be used is passed down the other tube in which it should be an easy sliding fit.

The rod is preheated by the tube, and a shoe on the front of the nozzle rests on the completed weld. There is also a speed welding nozzle with no baffle plate and a simple mixing chamber at the base. Speed welding is dealt with in detail later. Some units have a swing back tacking unit to save time when changing nozzles. There are special high speed nozzles for left-hand and right-hand welding in tight corners, and tank openings etc. Nozzles are available for 3, 4, 4.5, 5 and 6 mm round, 4 × 4 × 4 mm,

Fig. 8.4. Speed welding nozzles.

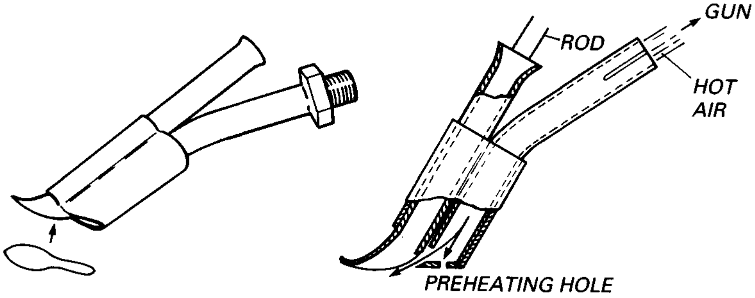

Fig. 8.5. (*a*) Sections of welding rod. (*b*) Profiled welding rods. (*c*) Multi-run.

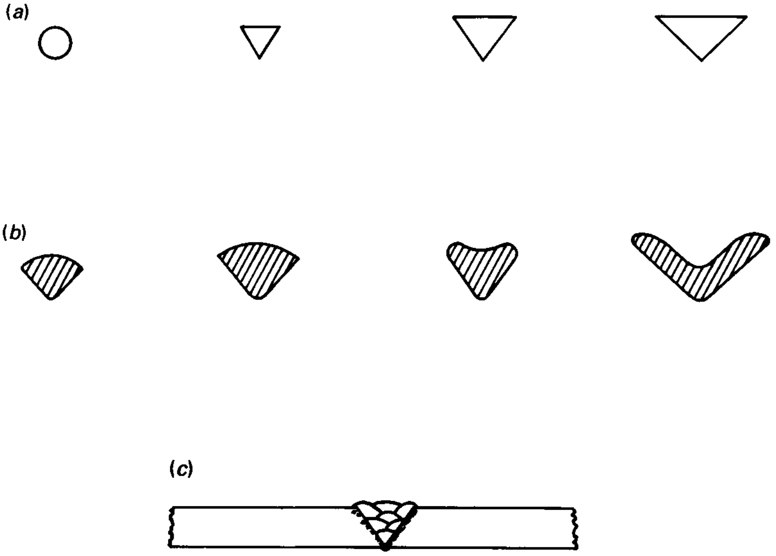

5 × 5 × 5 mm, 6 × 6 × 6 mm triangular rods, and sections for special use such as 5 × 5 × 4 mm, 6 × 4 × 4 mm profiled (see Fig. 8.5).

Preparation

Sheets up to 3 mm thickness require no preparation. They can be butted together and a tape of the same material and about 20–40 mm wide welded along the seam (Fig. 8.1*b*). This is followed immediately by using a hand-pressed roller to complete the operation. Very thin sheets of

Fig. 8.6. Types of joint and preparations.

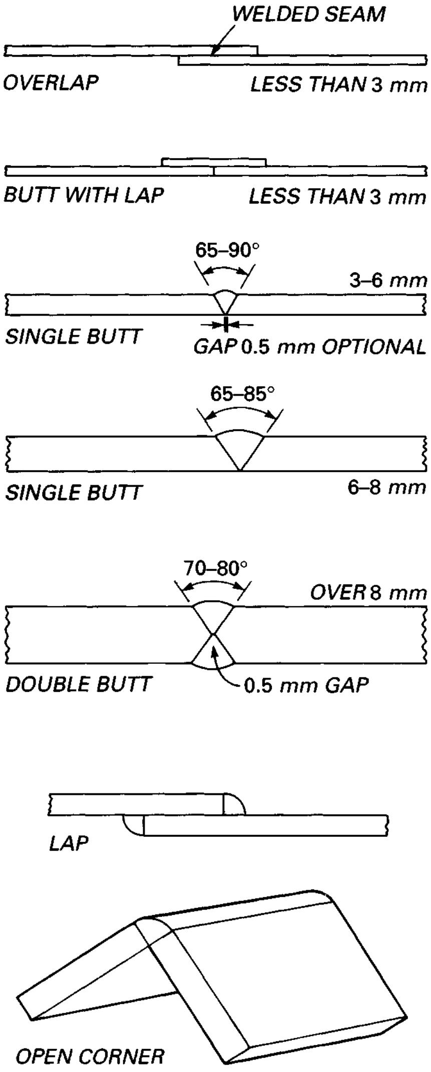
WELDED SEAM
OVERLAP
LESS THAN 3 mm
BUTT WITH LAP
LESS THAN 3 mm
65–90°
3–6 mm
SINGLE BUTT
GAP 0.5 mm OPTIONAL
65–85°
SINGLE BUTT
6–8 mm
70–80°
OVER 8 mm
DOUBLE BUTT
0.5 mm GAP
LAP
OPEN CORNER

PVC or polythene can be welded together by a heated wheel and hand pressure along the seam, or automatically in a machine which heats the roller and supplies pressure and movement.

When hand welding, sections up to 6 mm thick can be prepared with a 65–90° V, while sections of greater thickness can have either a single V of 65–85° or a double V of 70–80° (see Fig. 8.6). If the tack welding nozzle is used there is no need to allow any gap between sections.

If cracks are to be welded up in, for example, a car spoiler or bumper, they should be carefully inspected and a small hole, say 2 mm diameter, drilled at each end of the crack to prevent it travelling any further during welding. The plastic to be welded is then identified if possible, and a suitable rod selected. If identification is not possible, one rod at a time is selected from a bundle of rods of varying types, supplied by the manufacturers, until a rod is found that will fuse very well with the parent plate when heat is applied. It is essential to have rod and parent plate 'compatible' so that fusion results in a sound bond. In each case the temperature of the hot air to be used is selected on the gun.

Tack welding

The parts to be welded are placed in position and lined up. The gun, with tack welding nozzle on, is allowed about 2 minutes to heat up, and the front or toe of the nozzle is placed on the bottom of the commencement of the seam to be welded. The nozzle is held in this position until the sides of the weld begin to sweat and the gun is kept at an angle of about 30–40° to the seam and drawn along the seam bringing the two sides together and the tack weld is completed (Fig. 8.7). When cool, the tacking nozzle is exchanged for a straight or cranked standard nozzle.

Fig. 8.7. Tack weld.

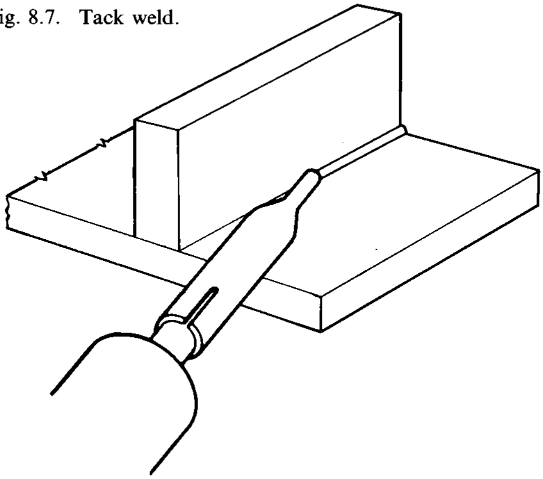

General welding

The rod is held at right angles to the weld at its commencement and hot air from the gun is played on the bottom of the rod and the parent plate until the rod begins to melt, spongily, into the prepared joint. The gun is held at an angle of about 45° to the seam, and moved up and down pendulum fashion with an amplitude of about 20–30 mm and a slight elliptical bias, as shown in Fig. 8.8. As the weld progresses down the seam with the gun preceding the rod, there should be a slight hesitation at the bottom of the stroke to ensure good fusion with the parent plastic. The rod softens and flows easily into the prepared groove with a minimum of hand pressure, say 1.5–2.5 kg. Any sudden reduction of pressure (e.g. when changing position of the fingers on the rod) may cause the rod to lift from the weld and, by trapping air, weaken the weld.

The rod angle is 90° to the weld for PVC, but for polypropylene and polyethylene the rod angle should be 45–60° in order to exert sufficient downward pressure on the rod. A reduced angle for PVC may produce cracks in the finished weld on cooling. To finish a weld, rod movement is stopped and heat applied to where the rod meets the parent plastic. The heat is then removed, keeping a downward pressure on the rod. The rod

Fig. 8.8. Down hand butt weld.

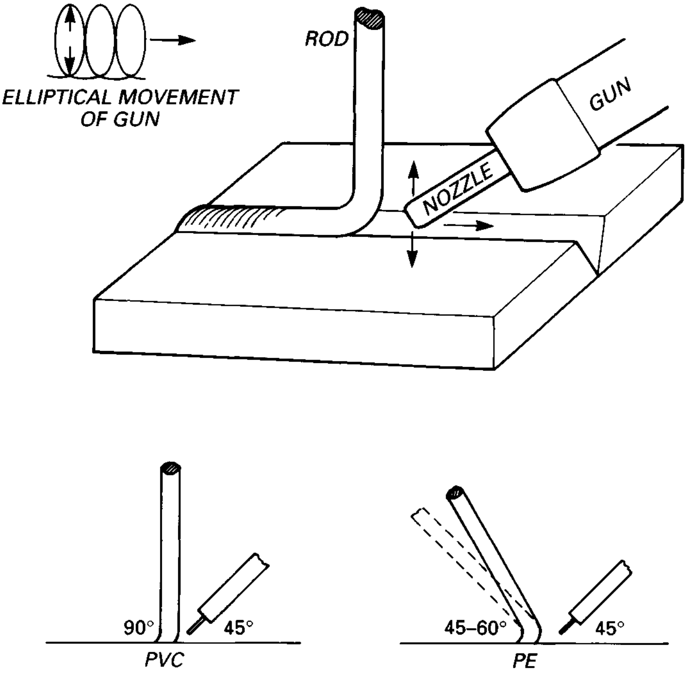

is then allowed to cool and is twisted off. Continuation of a weld can be performed by cutting the weld where it finishes at an angle of about 30° and starting the new weld at this point.

If a weld is to be made right around a pipe, the finish of the weld should overlap the beginning of the weld by about 10–15 mm and not finish alongside it. Using a wide slot nozzle, a tape welded seam can be made to fasten two sheets of plastic butted together. (See section on machine welding.)

Speed welding

The speed welding nozzle, as already described, is fitted to the gun. The welding rod or coil is prepared by sharpening one end to a pencil or chisel point. It is then pushed into the feed tube on the nozzle (and must be an easy sliding fit) until about 2.5–3.5 mm protrudes from the end. The tube on the nozzle is obtainable in various types of section, corresponding to the available rods (see Fig. 8.5). The gun is now held vertically at a point about 65–85 mm above the surface, thus preheating it. The gun, still held vertically, is brought to the commencement of the weld and the sharp end of the rod pushed in so that the sharpened end projects beyond the start of the weld at the front of the nozzle. Hand pressure is kept on the rod at the start, in order to obtain good fusion. When rod and parent plastic are completely softened and the rod begins to fuse into the plastic, the weld commences and the angle of the gun is lowered to about 45° to the line of weld. It is then drawn slowly down the weld preparation, the rod needing little or no feeding and only slight guidance. Remember that

Fig. 8.9. Speed welding.

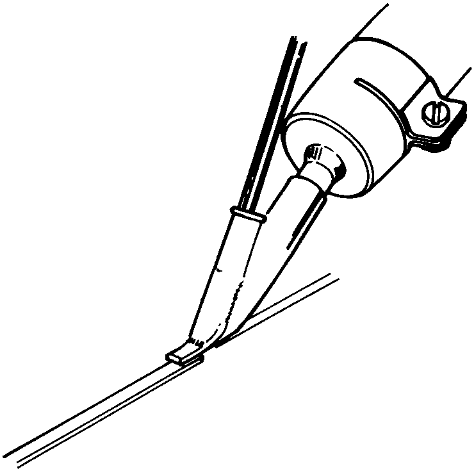

the angle between the gun and the weld determines the rate of welding, since the preheat hole varies the amount of preheat (Fig. 8.4). When the welding rate is too slow there are charred edges to the weld (PVC), when the rate is too high, there is incomplete fusion with no flow lines. The completed weld should be smooth with a faint wash visible down each side, an indication of good fusion of rod and parent plastic (Fig. 8.9). Note that if cracks are being repaired they should be welded in a continuous run, with, if possible, no stopping. To terminate a weld the gun is brought to the near vertical position and the rod is cut off by slipping the nozzle off the rod, leaving it standing attached to the weld. When cool it can be cut off close to the weld. If multi-layers are required (Fig. 8.5) each run should be allowed to cool before the next is applied. Since thermoplastics are poor conductors of heat, the stresses that arise during welding are concentrated in a smaller area than when welding metals. When welding V-prepared joints, the use of a profile rod and the process of speed welding reduces the stresses and accompanying strains. Note that the expansion coefficient of plastics is many times greater than that of steel.

Practical hints

At first, speed welding appears more difficult than that with an ordinary nozzle, because of the difficulty in obtaining penetration of the weld when starting. The first attempts result in no penetration until the weld has started to be made. It is difficult to heat both rod and plate (keeping some pressure on the rod) until the plate is hot enough to begin the operation. If the welder is too slow, the welding rod in the tube feeding the weld (which is fairly spongy due to preheating in the tube) quickly becomes charred, and the tube becomes blocked by the charred rod. It must then be cleared by pushing a steel rod of just smaller diameter than the tube through it. To overcome this the plates to be welded are preheated with the gun held vertically and then the pointed rod is immediately pushed quickly down the tube and the end guided under the shoe of the nozzle. The gun is held in this position until the rod melts into the parent plate; the angle of the gun is quickly reduced to about 45° and the weld commences immediately and must not stop, the rod being fed automatically by the welding operation and only guided by the fingers. If it is a longer weld, the coil of welding rod can be slipped over the welder's arm onto the shoulder whence it feeds itself automatically into the weld. If difficulty is met in starting the weld, use a small run on plate until the desired time for a good start is found. Keep the underside of the nozzle clean with a wire brush.

Machine (tractor) welding

Long seams in plastic sheets (e.g. PVC and polythene) can be welded together by moving a hot air gun suitably mounted and driven by an electrically driven tractor; for ease of replacement the heating element is of the plug-in type (Fig. 8.10).

The sheets may be butted together and a tape of the same material as the sheet and about 20–50 mm wide placed across the length of the joint and the tape welded to the butted seam, followed, where required, by a roller which presses the tape tightly to the seam. The machine is available as 220–240 V, 4.2 kW and 380–400 V, 5.2 kW, with other voltage ranges and loadings for various models. The machines are available for tilt welding (welding on roofs) and the speed is infinitely adjustable between set ranges, e.g. 1–3, 1–6 and 1–12 m/min as required.

The second method much used is overlap welding. The plastic sheets are overlapped up to 100 mm and sand or other material is placed beneath, so as to give a firm underside and one that will not damage the welding process. A test channel may be included, so that the joint can be tested from 2 to 5 bar for any leaks, which can be sealed by hand operation.

With the correct temperature selected and the correct speed of travel

Fig. 8.10. Leister hot air automatic welding machine for overlap welding of thermoplastic film.

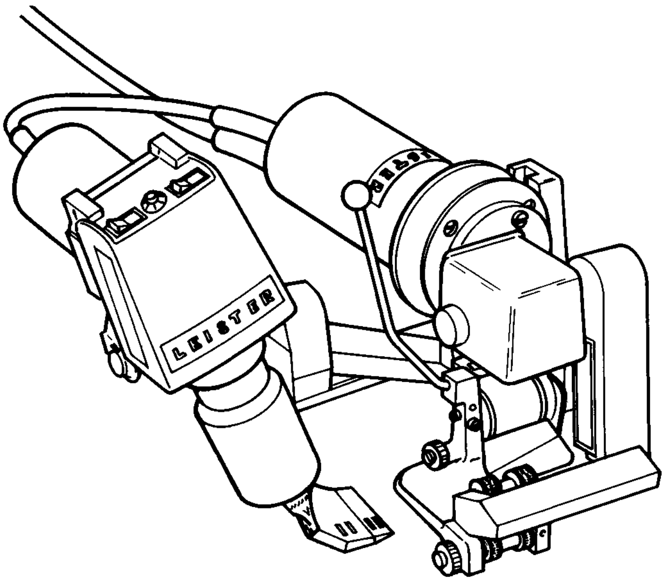

the machine produces an excellent weld between the sheets using a wide slot nozzle. For thin plastic sheets (e.g. bags) welding can be performed by a hand or machine driven roller. In both cases speed and temperature must be absolutely correct. Using a hand-held automatic welder, vertical and overhead seams about 10 mm thick can be tape welded and a test channel incorporated.

High frequency spark testing

Among the many uses of the high frequency spark tester is the testing of welded joints for porosity and, where required, the testing of the parent plastic for the same defect.

The spark tester is in the form of a light weight gun (Fig. 8.11), and is made for 100–125 V or 220–250 V supply, the loading being only 30 W so that it can be used from any 3-pin socket. The testing voltage is 55 kV maximum and the frequency 200 kHz.

Operation

There is a voltage control on the spark gun, which is important since any sheet or dielectric will break down if the applied voltage is too high. There is a probe fitted to the gun, onto which the high voltage is applied; the probe is then brought near an earthed metal plate and the trigger pressed. The sparks which emanate from the probe are adjusted in length until they are just a little longer than the thickness of the plastic and weld to be tested. The thinner the plastic, the smaller the voltage that

Fig. 8.11. High frequency spark gun (Goodburn).

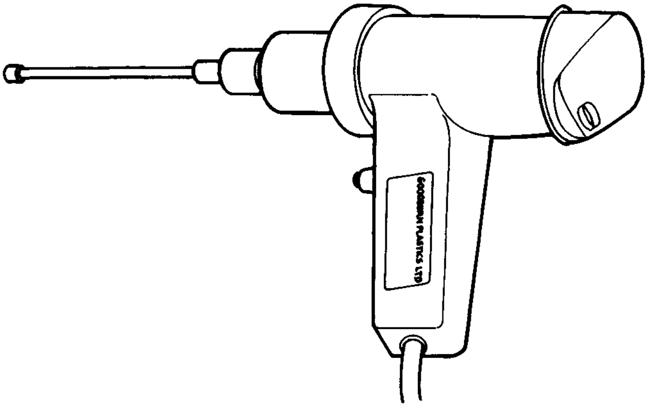

must be applied, because too high a voltage puts greater strain on the plastic (the dielectric).

The plastic weld under test is placed on the earthed metal plate and the gun switched on. The probe is then moved over the surface of the joint or other surface under test. When any defect is passed over, there is a bright spark and a hissing sound. The position of the defect is marked and the remainder tested as before. The defects are then repaired and the tester again moved over the positions of the faults to ensure that they have been correctly repaired. All plastics including nylon can be tested for plate or weld porosity using the probe from the gun, the material to be tested lying on the earthed metal plate.

Hot plate (or hot iron) welding

The parts of, say, polyethylene or PVC should have uniform profiles and diameters (e.g. pipes). The ends are prepared with a pipe-cutting tool and are lined up with one another using a jig. The ends are then lightly cleaned, kept separated and not touched. The hot plate, which may be of nickel-plated steel or aluminium, for example, is placed between the ends to be welded and the parts moved up to the plate.

The plate (Fig. 8.3) can be heated electrically or with a hot air gun until the ends of the parts are spongy and soft, the time for which they are in contact with the plate being critical, as this determines the temperature rise of the parts. With electric heating the current is switched off automatically after a given time, both current and time being variable. With the hot air gun care must be taken not to overheat the parts. When the current is switched off the plate is immediately removed and the part ends, now in the spongy condition, are brought together and pressure applied by hand or mechanically until the parts are cool. The flash is then removed and the weld is complete. Fusion has taken place throughout the part ends and an excellent bond is obtained.

Pipes of polyethylene, coloured yellow for gas, blue for water and black for sewage, are much used today and are jointed in this way. The generator used in the field is generally 110 V and equipped with an automatic timing device. Sockets, caps, reducers, tees and simple and self-tapping saddles are available for use with gas and water pipes. Pipes of length 50, 100 or 150 m are available, longer lengths being supplied on a drum. Wall thickness should be selected according to the pressure within the pipe and can vary from 20 mm to 500 mm O/D. Absolute cleanliness is essential with excellent fit, and scraping and other cutting tools are used.

Hot plate welding by machine

Hot plate welding is now used for the production of more complicated assemblies (e.g. in the automobile and domestic industries) which are not suitable for injection, blow moulding or extrusion. Several types of jigs and fixtures are used for variety of mouldings giving watertight sealing, enabling different types of material to be welded together with accurate dimensions and small tolerances.

Fig. 8.12. Horizontal hot plate plastics welding machine. Hand loaded, fully automatic welding cycle (Bielomatik). Optional extras include: vacuum pumps, mechanical clamping. Also available are special adaptations of machine sequences.

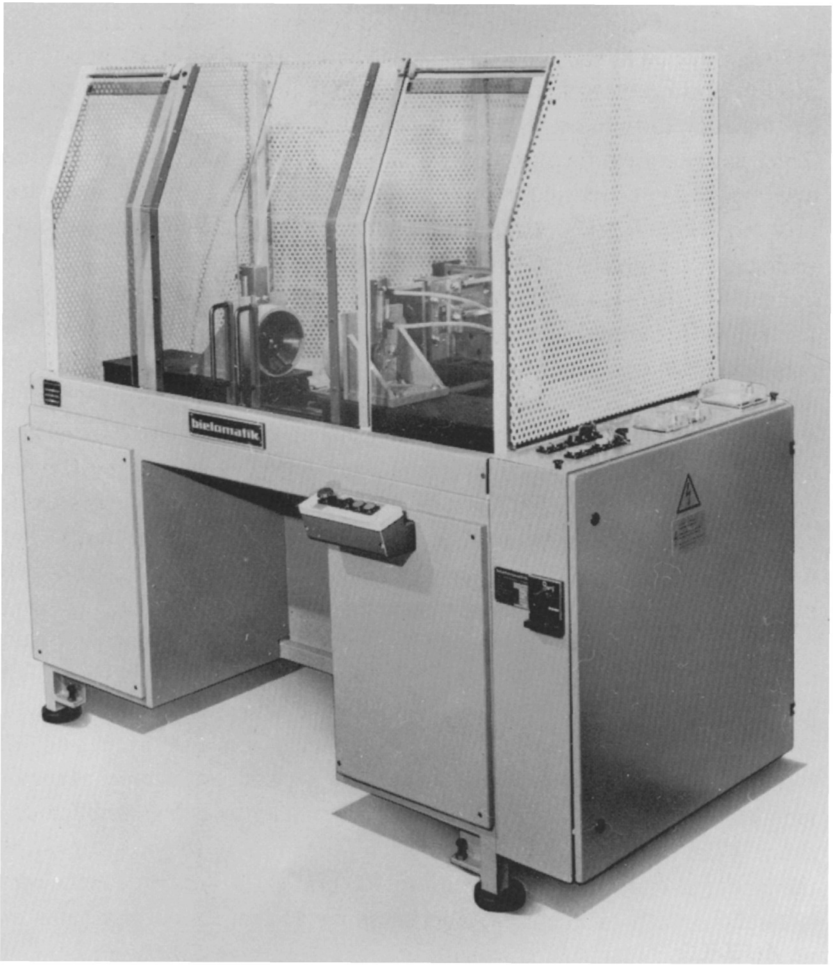

The parts to be welded are loaded onto the upper and lower plattens of the machine (Fig. 8.12). The fusion faces are then plasticized by being brought into contact with the hot plates, which are retracted automatically when the correct plasticizing point is reached. The mouldings are then brought together under pressure as with other types of hot plate welding. Upon cooling the assembly is unloaded from the machine.

Complicated assemblies can be produced more cheaply in this way, because if one part of the assembly is damaged or cracked, that part can be repaired by welding after correct preparation in the manner described in the section on hot air welding.

Electric fusion welding

This method can be performed using fittings such as couplers, reducers, branch saddles etc., which have electrical resistance coils moulded into each fitting. The ends of the heating elements in each fitting are brought to two connecting pillars. Power for the element is supplied from a generator fitted with an automatic timing control, which is essential since it identifies the fitting and gives the welding current and time for each operation. The generator is rated as 40 V and 50 A maximum. Before pipes are joined, they are prepared by machining the ends with a cutter, scraped if necessary, and scrupulously cleaned. They are then pushed into the coupling which has a central circlip to ensure that they are pushed in by an equal amount (Fig. 8.13). The terminals of the fitting are connected to the generator and the correct current and timing for the welding operation is automatically selected. This method is suitable for laying pipes in trenches. When pipes and coupling are brought to the correct temperature they are quite spongy or in a plasticized state; coupling and pipes fuse together at this stage. The electrical wires are disconnected from the pillars and the coupling is allowed to cool. The work may be done in a narrow trench when required and the coupling, or

Fig. 8.13. Types of joints and fittings required for electric fusion welding.

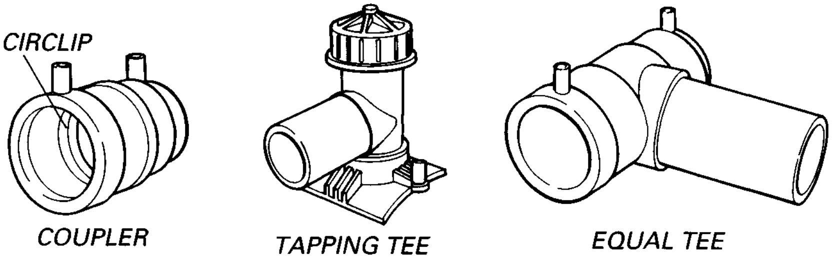

any type of fitting being used, is left in position in the trench, which is then filled in. The various joints made in this way can be subjected to tests for internal pressure (say 10 bar), tensile strength and impact for saddles.

It will be noted that with hot plate and electric fusion methods of welding, no extra filling rods are required. The electric fusion jointing of pipes is more convenient than the hot plate method and is largely used today.

Ultrasonic welding

With this method of machine welding, an alternating current of frequency 50 Hz is connected to a generator where its frequency is changed to 20–40 kHz (20000–40000 Hz) the capacity of the generator being 100–3000 W as required (Fig. 8.14). This ultrasonic current, say at 20 kHz, is fed into a transducer (which is generally a piezo-electric device) which changes the ultrasonic electric vibrations into a mechanical vibration also at 20 kHz. (The magneto-strictive types are little used nowadays.)

A piezo-electric material increases in length when a current flows through a coil surrounding the device and changes the alternating electric field into mechanical movement. Examples of piezo-electric materials are quartz, tourmaline (a silicate of boron and aluminium with additions of

Fig. 8.14. Resonant unit, showing various amplitude gains in ultrasonic welding (piezo-electric crystal transducer) (Bielomatik).

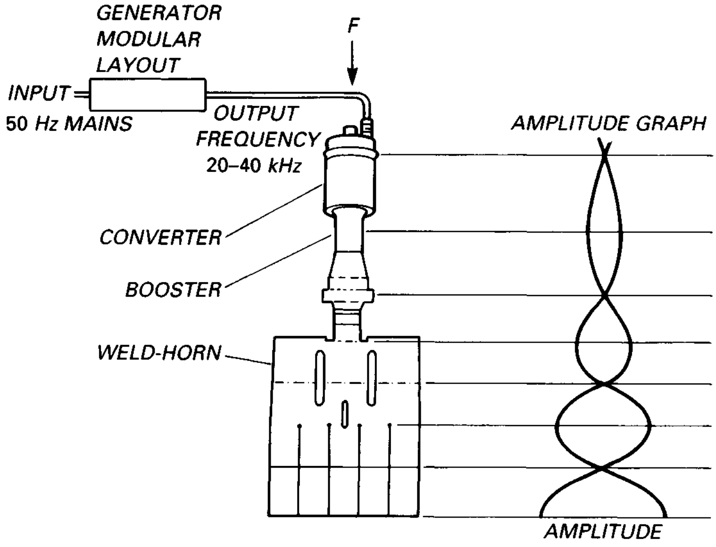

magnesium iron and other alkaline metals and fluorine) and Rochelle salt (sodium potassium tartrate etc.). Present-day transducers have thin sheets or wafers of piezo-electric material bonded to a central metal laminated core while two metal strips are bonded to the other faces of the piezo-electric material, the lengths of these strips being about one half that of sound in the metal core, so that the assembly is resonant at the given frequency and vibrates ultrasonically at the frequency of the input current. The welding head to which the transducer is fixed is inserted to get the gain in amplitude which cannot be obtained from the horn alone. It consists of two parts: the booster and the horn, or tool holder. The booster increases the amplitude of the vibration and is clamped at its nodal point (the point of least vibration). If the booster has a larger diameter where it is attached to the transducer body than where it is attached to the horn, the amplitude is increased and the ratio of the masses of metal on each side of the node determines the increase or decrease of the vibration.

The horn is specially designed to have the correct sonic properties, and transmits pressure to the welded surfaces and vibrates to make the weld. It must have good mechanical strength; the end which makes the weld can be of cylindrical shape, bar shape, or more complex shapes according to the parts to be welded. The horn can be made of steel alloy, aluminium alloy or titanium, which all have good ultrasonic properties.

The sequence of operations is: (1) machine is loaded; (2) force is applied; (3) ultrasonic power flows for a given time; (4) hold time during cooling of the weld; (5) force is removed and machine deloaded.

Vibration welding (linear)

This method, which is mechanized (Fig. 8.15) has an upper plate (or platten) on which one of the parts to be welded is secured; the lower platten carries the other part to be welded. The upper platten is vibrated hydraulically and is connected to springs. The frequency of vibration (180–300 Hz) is infinitely adjustable, as is the amplitude of vibration (0–1.2 mm each way, i.e. 2.4 mm in total). When the plattens are in the unloaded position there is ample space between them to mount the parts, which can be clamped hydraulically if required. When the parts are brought together, the upper platten vibrates, the frequency being varied as required. When the parts are plasticized they are brought together with up to 60 bar pressure. The above figures are for reference only. Particular examples of machines will naturally have varying degrees of vibration frequency, amplitude and pressure. Friction between the moving and fixed parts generates the heat required for the welding process and the pressure assures good fusion. The pressure is time dependent and is reduced to a lower value for some of the welding time and the cooling time. The time cycle varies from 1 to 4 seconds in most cases.

This method is especially useful for welding thermoplastics of low

Fig. 8.15. Bielomatic linear vibration welding machine features a short amplitude and high frequency, enabling accurate welding of components even with very small clearances.

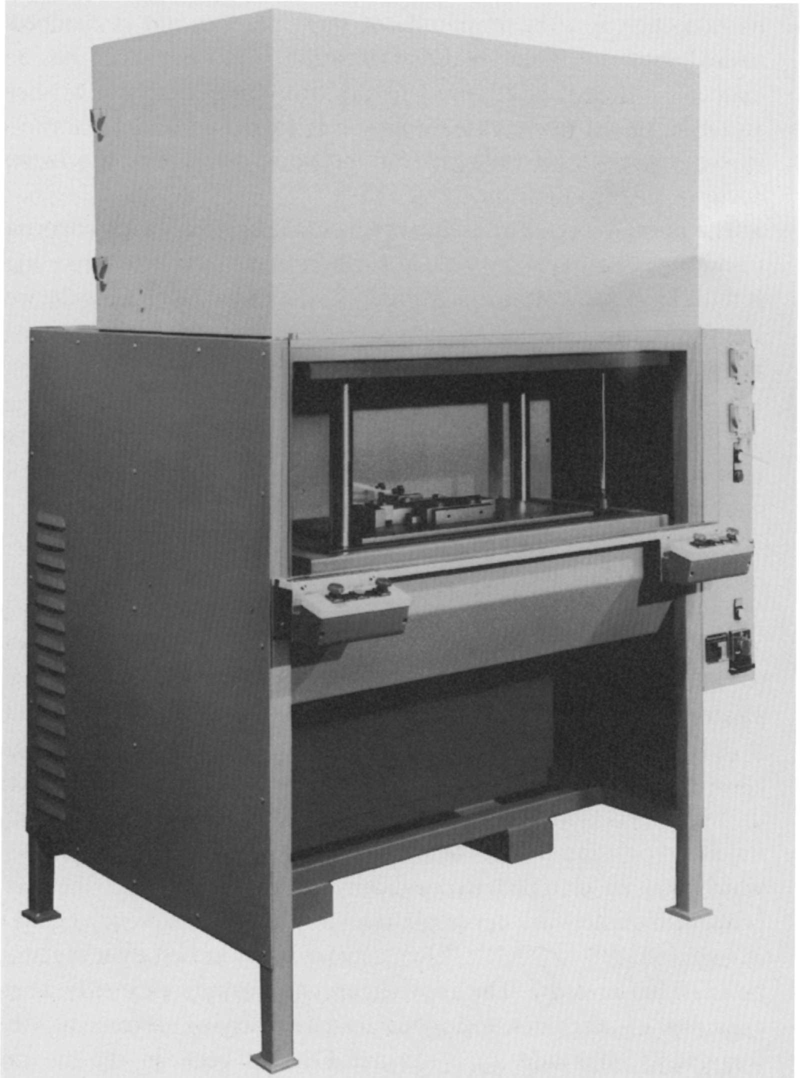

viscosity, such as various grades of nylon. It is evident that the machine is suitable for mass production of parts as there is considerable tooling required for each job. Plastics that have a low coefficient of friction, e.g. PTFE (polytetrafluorethylene), are not satisfactorily welded by this method.

Joining plastics and rubbers by means of adhesives

Plastics are being increasingly used by production engineers for stressed parts of assemblies. From the welding point of view plastics can also be joined by means of adhesives. To join surfaces together they should be completely cleaned and if possible roughened by rubbing down with fine sandpaper – metallic surfaces should be grit-blasted. The surfaces should not be touched after the final cleaning. The adhesive is then applied, enough being spread to ensure that any gaps which remain between the surfaces are filled with adhesive. The surfaces are then brought together and the adhesive soon becomes hard or 'cures': the higher the ambient temperature the quicker the cure. Also the thinner the line of bond between the surfaces, the greater will be the strength of joint and the more quickly will it cure. Slight irregularities of the surfaces increase both the grip and the locking effect.

One type of adhesive in use today is the 'instant' cyanoacrylate resin. These adhesives solidify because of the reaction between them and the thin layer of moisture always present on surfaces being bonded. The liquid adhesive is prevented from congealing when in its container, by the addition of an acidic stabilizer which keeps the molecules of the adhesive apart. On being applied to the surfaces being bonded the molecules of the stabilizer are repelled or removed and the molecules of the adhesive join together and curing commences. The 'instant' description of the resin indicates that the curing process is extremely short and can be just a second or so. Not all surfaces are suitable for successful bonding because they have insufficient attraction for the molecules of the adhesive and they must be treated with other preparations which prepare their surfaces for bonding. Surfaces of polyesters and ABS (acrylonitrile-butadiene-styrene) are examples of those with good attraction while others, such as polyethylene, have little attraction.

In the original acrylic resin adhesive, the smallest molecules are those of the methyl group of esters. The polymer chains of molecules are large in number, resulting in high tensile strength bonds. The ethyl group have larger molecules so that there are fewer on a given area than for methyl esters, with the result that there is a less rigid joint, but one suitable for bonding rubbers and plastics. In addition, the curing time is increased but with less odour.

There is a wide range of adhesives covering most applications for locking threaded parts, for acidic and porous surfaces and those having great 'peel' strength with a top temperature of 100 °C and a tensile strength of 26 N/mm^2.

The 'gel' type adhesives are used for fabrics, paper, phenolics, PVC, neoprine, and nitrile rubber, with a curing time of a few seconds only. One gel has no associated fumes or odour and can be used in confined spaces, while another is clear and thick and does not run. Its top temperature is over 100 °C and it resists heat that would destroy a normal bond made with instant adhesive.

Another type is used for the assembly of pulleys, sprockets, bearings, bushes, sleeves etc., and other adhesives are used for pressed fits for cylinder liners of diesel engines. There are many other applications of this type. There is also a range of adhesives which bond the substrates of glass, ceramics, ferrites, epoxy moulding compounds, phenolic resins, laminated paper, and glass fibre laminate epoxy materials.

By the use of single-component liquids, drying rapidly at room

Fig. 8.16. Bondmatic® with foot switch and pencil.

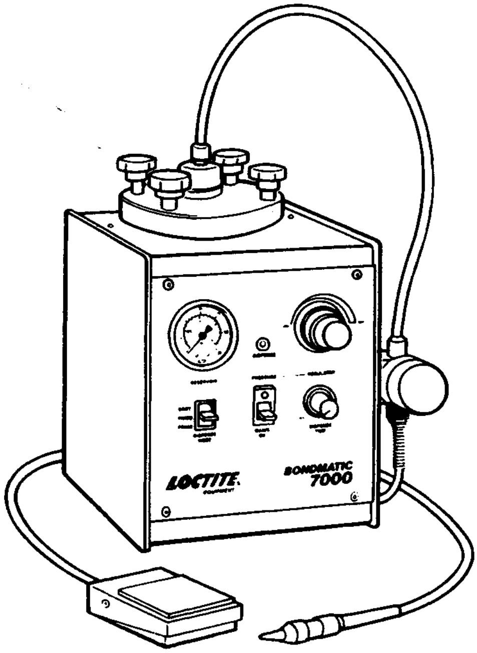

temperature, surfaces are made suitable for bonding polyolefins and other low-energy surfaces. The liquid is applied by brush, spraying or dipping and is a polyolefin primer so that it should only be used for difficult-to-bond surfaces, e.g. polyolefins, polypropylene, PTFE and thermoplastic rubber.

With such a variety of adhesives available, each developed for a specific purpose, it is essential that the operator should contact the makers of the adhesives and obtain a full list of the adhesives available and their method of usage (Figs. 8.16 and 8.17).

Fig. 8.17. Loctite® adhesive squeeze bottle.

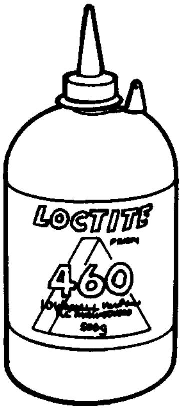

Some of the plastics considered with a few of their properties

Plastic	Properties
PVC (HDPVC and LDPVC)[a]	Low resistance to heat; rigid PVC higher. Low stress application, but up to 75 °C use HDPVC. Expansion 6–7 times that of steel. Low notch impact test, but this is improved by compounding with other substances. Expansion and contraction problems important when lining tanks or pipe-lines. Rigid PVC widely used for welding. Compounded with ABS, improves its impact strength, but chemical resistance suffers.
ABS (ABS/PVC and ABS/MAT: hard ABS reinforced with fibreglass, thermosetting)	Light, with high impact strength, good resistance to chemicals (alkalis and dilute acids). Used for radiator grilles, etc.

Plastic	Properties
Polypropylene (PPHD and PPLD)	Rigid and m.p. 170 °C. Low SG can be used up to 120 °C, has good mechanical strength and is resistant to most corrosive substances. Welded polypropylene can be used up to 120 °C as long as little stress is applied. Has replaced PVC and polythene in many fields. Used for sheeting, ropes and webbing. Compounded with other substances to improve qualities such as anti-cracking.
Polyethylene (Ethylene C_2H_4)	Soft and semi-flexible (makes squeeze bottles). Used for bags of all kinds. Similar in many ways to polypropylene. Used for pipes such as those used by British Gas which demand ease of connection and flexibility, for mains and smaller services. Used for injection moulded houseware (sink tidies, bowls etc.).
PTFE (Ethylene with four fluorine atoms substituted for hydrogen atoms)	Polytetrafluorethylene is mixed with other substances to give good heat resistance up to 200 °C and good dielectric properties. Welding temperatures 320–400 °C, low friction coefficient. Used for heat-resistant non-lubricated parts.
Polyacrylates (PMMA)	Polymethyl-methacrylate transmits about 92% light, has excellent optical properties and does not shatter. Chemical resistance only fair. Polyacrylates which are modified cannot be welded but can be cemented.
Polycarbonates	Low tendency to crystallize. High melting point, absorb little water, are tough and shock resistant. Yellowish or colourless. Can stand deformation but are soluble in some solvents such as methylene chloride. Resistance to chemicals rather low. Attacked by concentrated mineral acids, alkalis, ammonia, amines, ketones etc., and thus are used for dilute acids (organic and inorganic) mineral and vegetable oils and fats.
Polyamides	Contain amide group NHCO. Better known as nylon, certain types of which can be welded. Tough with high melting point. Resistant to oils, dilute acids and alkalis and some solvents. Used for textiles, hosiery, pipes, bearings and moulded parts where their higher melting point is an advantage. Some types absorb water others do not. Properties vary with the final composition.

[a] HD, high density; LD, low density.

Appendix 1

Welding symbols

A weld is indicated on a drawing by (1) a symbol (Fig. A1.1*a*) and (2) an arrow connected at an angle to a reference line usually drawn parallel to the bottom of the drawing (Fig. A1.1*b*). The side of the joint on which the arrow is placed is known as the 'arrow side' to differentiate it from the 'other side' (Fig. A1.1*c*). If the weld symbol is placed *below* the reference line, the weld face is on the *arrow-side* of the joint, while if the symbol is above the reference line the weld face is on the other side of the joint (Fig. A1.1*c*). Symbols on both sides of the reference line indicate welds to be made on both sides of the joint while if the symbol is across the reference line the weld is within the plane of the joint. A circle where arrow line meets reference line indicates that it should be a peripheral (all round) weld, while a blacked in flag at this point denotes an 'on site' weld (Fig. A1.1*d*). Intermittent runs of welding are indicated by figures denoting the welded portions, and figures in brackets the non-welded portions, after the symbol (Fig. A1.1*e*). A figure before the symbol for a fillet weld indicates the leg length. If the design throat thickness is to be included, the leg length is prefixed with the letter '*b*' and the throat thickness with the letter '*a*'. Unequal leg lengths have a × sign separating the dimensions (Fig. A1.1*f*). A fork at the end of the reference line with a number within it indicates the welding process to be employed (e.g. 131 is MIG, see Table A1.1) while a circle at this point containing the letters NDT indicates that non-destructive testing is required (Fig. A1.1*g*). Weld profiles, flat (or flush), convex and concave profiles are shown as supplementary symbols in Fig. A1.1*h*.

Figure A1.2*a* gives the elementary symbols and A1.2*b* typical uses of them. The student should study BS 499, Part 2, *Symbols for welding*, for a complete account of this subject.

Table A1.1 shows the numerical indication of processes complying with International Standard ISO 4063.

Fig. A1.1

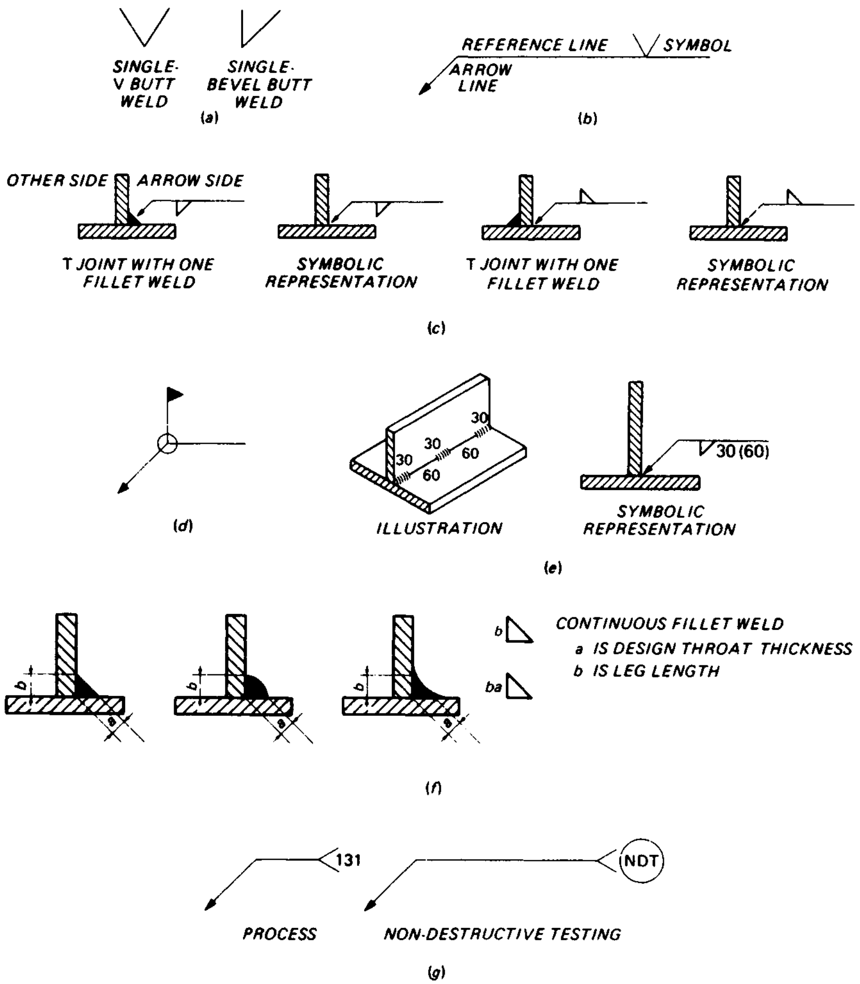
SINGLE-V BUTT WELD
SINGLE-BEVEL BUTT WELD
(a)
REFERENCE LINE
SYMBOL
ARROW LINE
(b)
OTHER SIDE
ARROW SIDE
T JOINT WITH ONE FILLET WELD
SYMBOLIC REPRESENTATION
T JOINT WITH ONE FILLET WELD
SYMBOLIC REPRESENTATION
(c)
(d)
30
30
30
60
60
ILLUSTRATION
30 (60)
SYMBOLIC REPRESENTATION
(e)
b
a
a
b
a
b
a
b
ba
CONTINUOUS FILLET WELD
a IS DESIGN THROAT THICKNESS
b IS LEG LENGTH
(f)
131
PROCESS
NDT
NON-DESTRUCTIVE TESTING
(g)

Supplementary symbols

SHAPE	SYMBOL
FLAT (USUALLY FINISHED FLUSH)	—
CONVEX	⌒
CONCAVE	‿

Examples of supplementary symbols

DESCRIPTION	ILLUSTRATION	SYMBOL
FLAT (FLUSH) SINGLE-V BUTT WELD	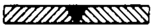	
CONVEX DOUBLE-V BUTT WELD	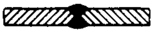	
CONCAVE FILLET WELD	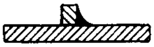	
FLAT (FLUSH) SINGLE-V BUTT WELD WITH FLAT (FLUSH) BACKING RUN	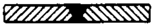	

(h)

Table A1.1. *Numerical indication of process* (*BS 499, Part 2, 1980*)*

No.	Process
1	**Arc welding**
11	Metal-arc welding without gas protection
111	Metal-arc welding with covered electrode
112	Gravity arc welding with covered electrode
113	Bare wire metal arc welding
114	Flux cored metal-arc welding
115	Coated wire metal-arc welding
118	Firecracker welding
12	Submerged arc welding
121	Submerged arc welding with wire electrode
122	Submerged arc welding with strip electrode
13	Gas shielded metal-arc welding
131	MIG welding
135	MAG welding: metal-arc welding with non-inert gas shield
136	Flux cored metal-arc welding with non-inert gas shield
14	Gas-shielded welding with non-consumable electrode
141	TIG welding
149	Atomic-hydrogen welding
15	Plasma arc welding
18	Other arc welding processes
181	Carbon arc welding
185	Rotating arc welding
2	**Resistance welding**
21	Spot welding
22	Seam welding
221	Lap seam welding
225	Seam welding with strip
23	Projection welding
24	Flash welding
25	Resistance butt welding
29	Other resistance welding processes
291	HF resistance welding
3	**Gas welding**
31	Oxy-fuel gas welding
311	Oxy-acetylene welding
312	Oxy-propane welding
313	Oxy-hydrogen welding
32	Air fuel gas welding
321	Air-acetylene welding
322	Air-propane welding
4	**Solid phase welding; Pressure welding**
41	Ultrasonic welding
42	Friction welding
43	Forge welding
44	Welding by high mechanical energy
441	Explosive welding
45	Diffusion welding
47	Gas pressure welding
48	Cold welding
7	**Other welding processes**
71	Thermit welding
72	Electroslag welding
73	Electrogas welding
74	Induction welding
75	Light radiation welding
751	Laser welding
752	Arc image welding
753	Infrared welding
76	Electron beam welding
77	Percussion welding
78	Stud welding
781	Arc stud welding
782	Resistance stud welding
9	**Brazing, soldering and braze welding**
91	Brazing
911	Infrared brazing
912	Flame brazing
913	Furnace brazing
914	Dip brazing
915	Salt bath brazing
916	Induction brazing
917	Ultrasonic brazing
918	Resistance brazing
919	Diffusion brazing
923	Friction brazing
924	Vacuum brazing
93	Other brazing processes
94	Soldering
941	Infrared soldering
942	Flame soldering
943	Furnace soldering
944	Dip soldering
945	Salt bath soldering
946	Induction soldering
947	Ultrasonic soldering
948	Resistance soldering
949	Diffusion soldering
951	Flow soldering
952	Soldering with soldering iron
953	Friction soldering
954	Vacuum soldering
96	Other soldering processes
97	Braze welding
971	Gas braze welding
972	Arc braze welding

---

* This table complies with International Standard ISO 4063.

Fig. A1.2. (*a*) Elementary welding symbols (BS 499, Part 2, 1980).

DESCRIPTION	SECTIONAL REPRESENTATION	SYMBOL
1. BUTT WELD BETWEEN FLANGED PLATES (FLANGES MELTED DOWN COMPLETELY)	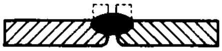	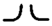
2. SQUARE BUTT WELD	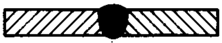	||
3. SINGLE-V BUTT WELD	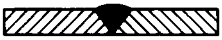	V
4. SINGLE-BEVEL BUTT WELD	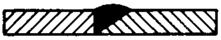	
5. SINGLE-V BUTT WELD WITH BROAD ROOT FACE	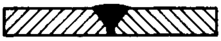	Y
6. SINGLE-BEVEL BUTT WELD WITH BROAD ROOT FACE	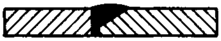	
7. SINGLE-U BUTT WELD	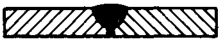	
8. SINGLE-J BUTT WELD	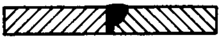	
9. BACKING OR SEALING RUN	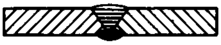	
10. FILLET WELD	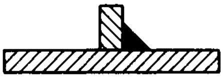	
11. PLUG WELD (CIRCULAR OR ELONGATED HOLE, COMPLETELY FILLED)	ILLUSTRATION 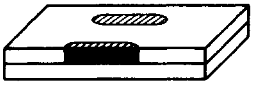	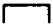
12. SPOT WELD (RESISTANCE OR ARC WELDING) OR PROJECTION WELD	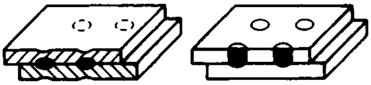 (*a*) RESISTANCE (*b*) ARC	○
13. SEAM WELD	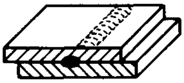	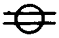

Fig. A1.2. (*b*) Examples of uses of symbols (BS 499, Part 2, 1980).

DESCRIPTION SYMBOL	GRAPHIC REPRESENTATION	SYMBOLIC REPRESENTATION
SINGLE-V BUTT WELD V	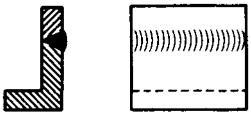	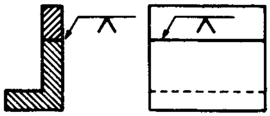
SINGLE-V BUTT WELD V AND BACKING RUN ◡	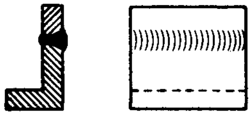	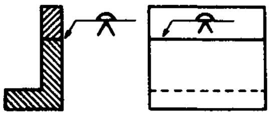
FILLET WELD ⊿	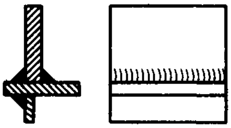	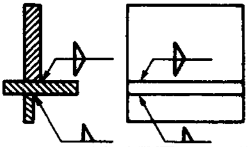
SINGLE-BEVEL BUTT WELD |/ WITH FILLET WELD SUPERIMPOSED ⊿	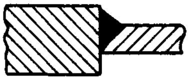	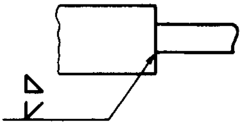
SQUARE BUTT WELD 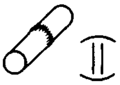	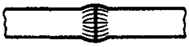	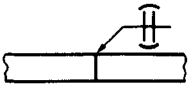
STAGGERED INTERMITTENT FILLET WELD	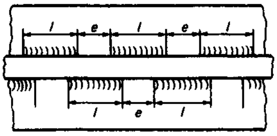	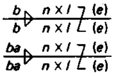 *a* IS THE DESIGN THROAT THICKNESS *b* IS THE LEG LENGTH *e* IS THE DISTANCE BETWEEN ADJACENT WELD ELEMENTS *l* IS THE LENGTH OF THE WELD (WITHOUT END CRATERS) *n* IS THE NUMBER OF WELD ELEMENTS

American welding symbols

From *Standard Welding Symbols and Rules for their Use*, A2.4.79 (reprint 1983), published by The American Welding Society

Note: A groove weld is a weld made in the groove between two members to be joined.

Fig. A1.3. Location of information on welding symbols.

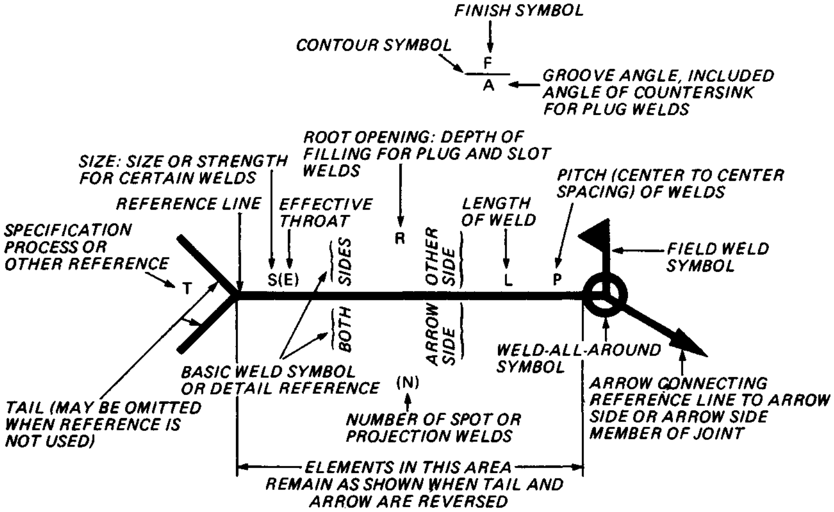
FINISH SYMBOL
CONTOUR SYMBOL
F
A
GROOVE ANGLE, INCLUDED ANGLE OF COUNTERSINK FOR PLUG WELDS
ROOT OPENING: DEPTH OF FILLING FOR PLUG AND SLOT WELDS
SIZE: SIZE OR STRENGTH FOR CERTAIN WELDS
PITCH (CENTER TO CENTER SPACING) OF WELDS
REFERENCE LINE
EFFECTIVE THROAT
LENGTH OF WELD
SPECIFICATION PROCESS OR OTHER REFERENCE
R
SIDES
OTHER SIDE
FIELD WELD SYMBOL
T
S(E)
L
P
BOTH
ARROW SIDE
WELD-ALL-AROUND SYMBOL
BASIC WELD SYMBOL OR DETAIL REFERENCE
(N)
TAIL (MAY BE OMITTED WHEN REFERENCE IS NOT USED)
NUMBER OF SPOT OR PROJECTION WELDS
ARROW CONNECTING REFERENCE LINE TO ARROW SIDE OR ARROW SIDE MEMBER OF JOINT
ELEMENTS IN THIS AREA REMAIN AS SHOWN WHEN TAIL AND ARROW ARE REVERSED

Fig. A1.4. Welding symbols.

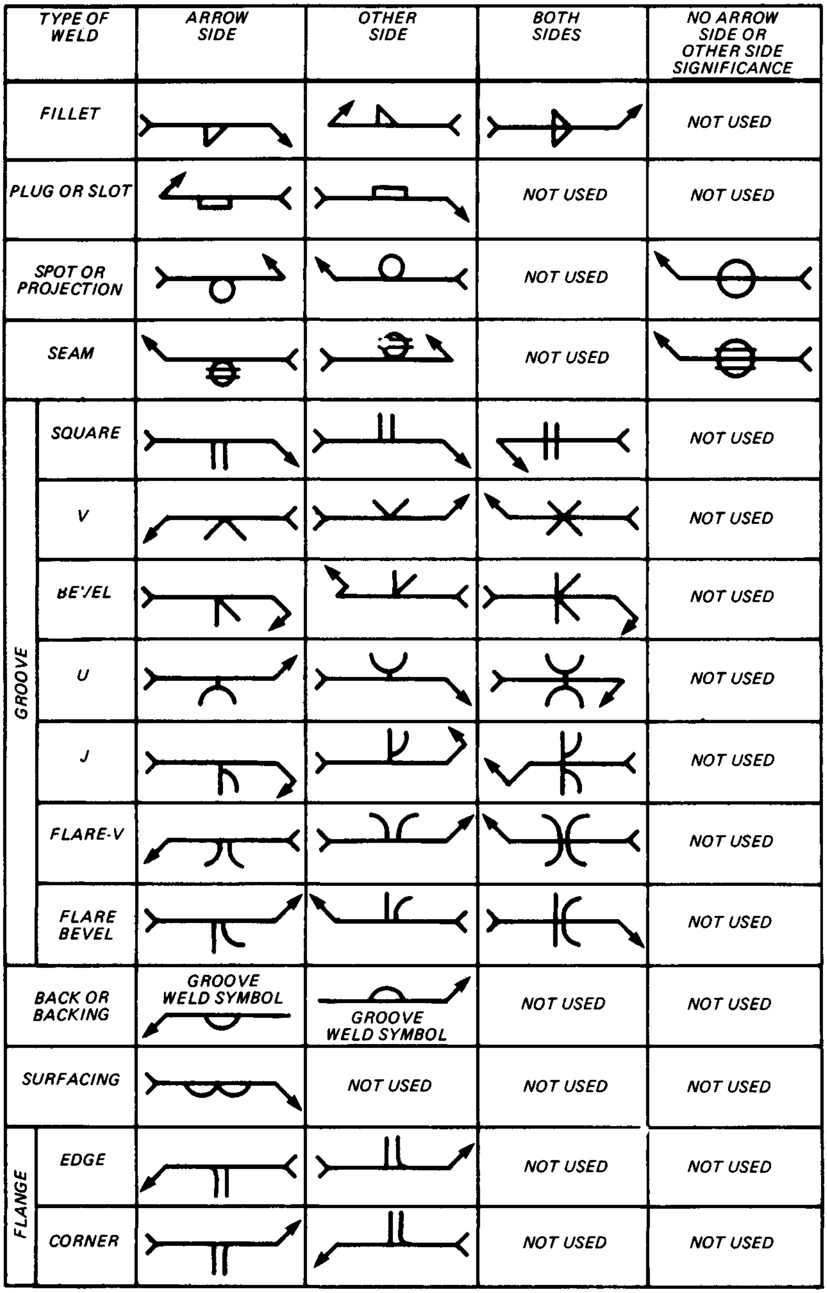
TYPE OF WELD
ARROW SIDE
OTHER SIDE
BOTH SIDES
NO ARROW SIDE OR OTHER SIDE SIGNIFICANCE
FILLET
NOT USED
PLUG OR SLOT
NOT USED
NOT USED
SPOT OR PROJECTION
NOT USED
SEAM
NOT USED
GROOVE
SQUARE
NOT USED
V
NOT USED
BEVEL
NOT USED
U
NOT USED
J
NOT USED
FLARE-V
NOT USED
FLARE BEVEL
NOT USED
BACK OR BACKING
GROOVE WELD SYMBOL
GROOVE WELD SYMBOL
NOT USED
NOT USED
SURFACING
NOT USED
NOT USED
NOT USED
FLANGE
EDGE
NOT USED
NOT USED
CORNER
NOT USED
NOT USED

Fig. A1.5. Supplementary welding symbols.

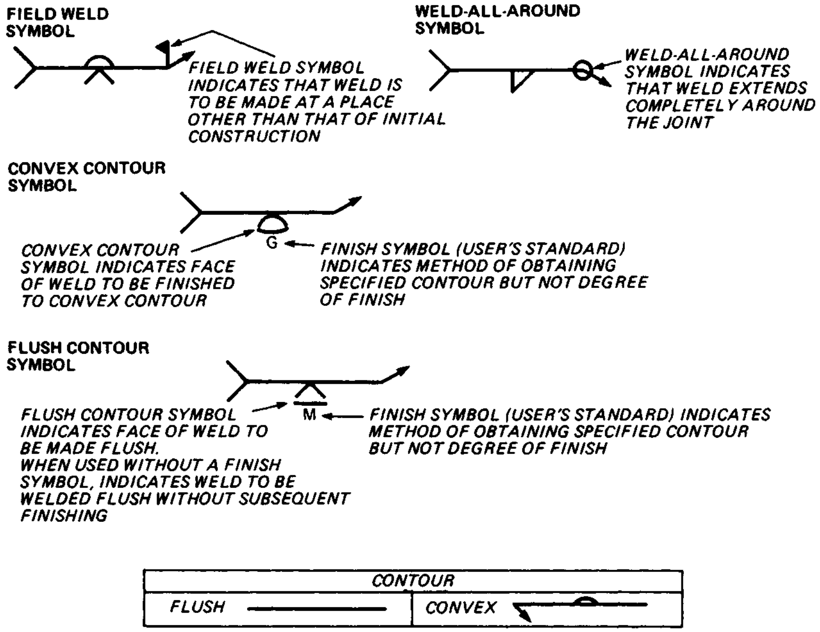
FIELD WELD SYMBOL
FIELD WELD SYMBOL INDICATES THAT WELD IS TO BE MADE AT A PLACE OTHER THAN THAT OF INITIAL CONSTRUCTION
WELD-ALL-AROUND SYMBOL
WELD-ALL-AROUND SYMBOL INDICATES THAT WELD EXTENDS COMPLETELY AROUND THE JOINT
CONVEX CONTOUR SYMBOL
CONVEX CONTOUR SYMBOL INDICATES FACE OF WELD TO BE FINISHED TO CONVEX CONTOUR
G
FINISH SYMBOL (USER'S STANDARD) INDICATES METHOD OF OBTAINING SPECIFIED CONTOUR BUT NOT DEGREE OF FINISH
FLUSH CONTOUR SYMBOL
FLUSH CONTOUR SYMBOL INDICATES FACE OF WELD TO BE MADE FLUSH. WHEN USED WITHOUT A FINISH SYMBOL, INDICATES WELD TO BE WELDED FLUSH WITHOUT SUBSEQUENT FINISHING
M
FINISH SYMBOL (USER'S STANDARD) INDICATES METHOD OF OBTAINING SPECIFIED CONTOUR BUT NOT DEGREE OF FINISH
CONTOUR
FLUSH
CONVEX

Fig. A1.6. Other typical welding symbols.

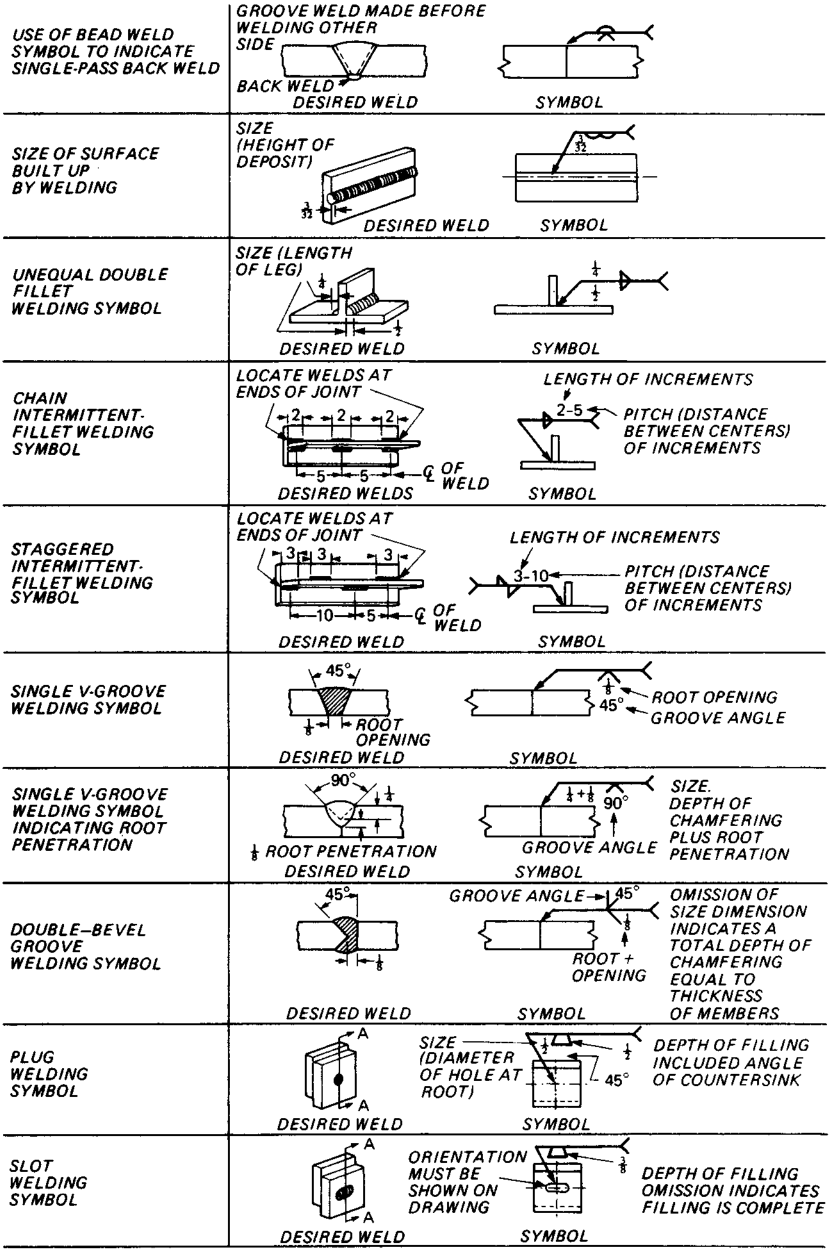
USE OF BEAD WELD SYMBOL TO INDICATE SINGLE-PASS BACK WELD
GROOVE WELD MADE BEFORE WELDING OTHER SIDE
BACK WELD
DESIRED WELD
SYMBOL
SIZE OF SURFACE BUILT UP BY WELDING
SIZE (HEIGHT OF DEPOSIT)
3/32
DESIRED WELD
SYMBOL
UNEQUAL DOUBLE FILLET WELDING SYMBOL
SIZE (LENGTH OF LEG)
1/4
1/2
DESIRED WELD
SYMBOL
CHAIN INTERMITTENT-FILLET WELDING SYMBOL
LOCATE WELDS AT ENDS OF JOINT
2
5
℄ OF WELD
DESIRED WELDS
LENGTH OF INCREMENTS
2-5
PITCH (DISTANCE BETWEEN CENTERS) OF INCREMENTS
SYMBOL
STAGGERED INTERMITTENT-FILLET WELDING SYMBOL
LOCATE WELDS AT ENDS OF JOINT
3
10
5
℄ OF WELD
DESIRED WELD
LENGTH OF INCREMENTS
3-10
PITCH (DISTANCE BETWEEN CENTERS) OF INCREMENTS
SYMBOL
SINGLE V-GROOVE WELDING SYMBOL
45°
1/8
ROOT OPENING
DESIRED WELD
ROOT OPENING
GROOVE ANGLE
SYMBOL
SINGLE V-GROOVE WELDING SYMBOL INDICATING ROOT PENETRATION
90°
1/4
1/8 ROOT PENETRATION
DESIRED WELD
1/4+1/8
SIZE. DEPTH OF CHAMFERING PLUS ROOT PENETRATION
GROOVE ANGLE
SYMBOL
DOUBLE-BEVEL GROOVE WELDING SYMBOL
45°
1/8
DESIRED WELD
GROOVE ANGLE
ROOT + OPENING
OMISSION OF SIZE DIMENSION INDICATES A TOTAL DEPTH OF CHAMFERING EQUAL TO THICKNESS OF MEMBERS
SYMBOL
PLUG WELDING SYMBOL
A
DESIRED WELD
SIZE (DIAMETER OF HOLE AT ROOT)
1/2
45°
DEPTH OF FILLING
INCLUDED ANGLE OF COUNTERSINK
SYMBOL
SLOT WELDING SYMBOL
A
DESIRED WELD
ORIENTATION MUST BE SHOWN ON DRAWING
3/8
DEPTH OF FILLING
OMISSION INDICATES FILLING IS COMPLETE
SYMBOL

MEASUREMENTS ARE IN INCHES

Appendix 2

Simplified notes on the operation of a thyristor

An ‘n’ type silicon wafer or element has impurities such as antimony, arsenic or phosphorus added to it to increase the number of negatively charged electrons available, termed ‘doping’.

A ‘p’ type has impurities such as indium, gallium or boron added to increase the number of positive carriers or ‘holes’ available.

Both types have increased conductivity compared with the pure silicon.

When ‘n’ and ‘p’ types are placed in contact this forms a diode (two elements) and this will pass a current in one direction but not in the other (subject to certain temperature considerations), converting the a.c. input into half-wave pulsating d.c. The electrons flow from the ‘n’ type where there are free electrons to the ‘p’ type where there are positive ‘holes’. By connecting four diodes in bridge connection full wave rectification is obtained (Figs. A2.1 and A2.2).

Fig. A2.1

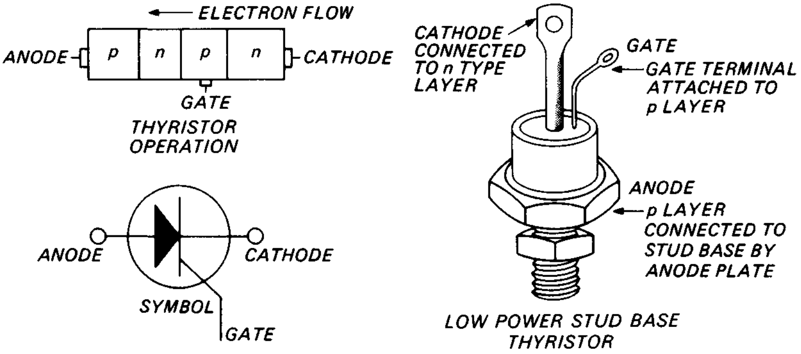
ELECTRON FLOW
ANODE
p
n
p
n
CATHODE
GATE
THYRISTOR OPERATION
ANODE
CATHODE
SYMBOL
GATE
CATHODE CONNECTED TO n TYPE LAYER
GATE
GATE TERMINAL ATTACHED TO p LAYER
ANODE
p LAYER CONNECTED TO STUD BASE BY ANODE PLATE
LOW POWER STUD BASE THYRISTOR

---

* Gallium arsenide is now regarded as a suitable material for integrated circuits (IC) and other devices, not supplanting silicon, but being used for high speed and high frequency applications where silicon is not suitable.

The thyristor (sometimes called the silicon controlled rectifier or SCR) has four elements with three junctions and a further terminal called the 'gate' which is connected to the 'p' layer adjacent to cathode (Fig. A2.1). If the gate is connected directly to the cathode, the device blocks the flow of current in either direction. However, if, while the anode is positive with respect to the cathode, the gate is also made positive (typically by two or three volts), current will flow from anode to cathode. The current required in the gate is usually between 1/100 and 1/1000 of the maximum anode current. Once conduction has been so initiated, the gate has no further effect, and can be de-energized. Conduction will continue until the external circuit causes the current to fall to zero, when the device will revert to its blocking state. Like a diode, it will not permit reverse (cathode to anode) current flow under any conditions.

If half the diodes in a bridge rectifier are replaced by thyristors (Fig. A2.2), and an adjustable time delay is incorporated in the circuit generating the gate current, the d.c. output of the bridge may be varied from zero up to the equivalent of an ordinary diode bridge, by varying the time delay.

Fig. A2.3 shows typical waveforms for half output. A diode bridge connected to the supply shown in Fig. A2.3(*a*) would produce a d.c. output voltage as in (*b*). However, by using thyristors and delaying the firing pulses by a quarter of a cycle after the thyristor anodes become positive, the output voltage is halved. Fig. A2.3(*c*) shows the gate pulse timing, and (*d*) the resultant d.c. voltage.

Fig. A2.2 also shows, in simplified form, automatic control of this firing delay. The desired current is set by a single control knob on potentiometer *C*. This is compared electronically with the measured value of current from sensor *H* and the pulse timing adjusted automatically to maintain the current at the set level.

Fig. A2.2 Simplified schematic diagram for one knob control of welding current by thyristor (SCR).

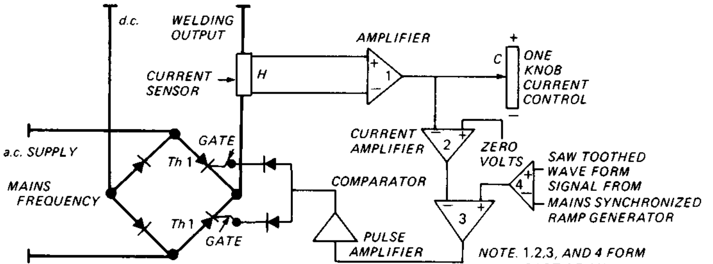
d.c.
WELDING OUTPUT
AMPLIFIER
+
C
ONE KNOB CURRENT CONTROL
−
CURRENT SENSOR
H
+
1
−
a.c. SUPPLY
MAINS FREQUENCY
GATE
Th1
Th1
GATE
CURRENT AMPLIFIER
2
ZERO VOLTS
SAW TOOTHED WAVE FORM SIGNAL FROM MAINS SYNCHRONIZED RAMP GENERATOR
4
COMPARATOR
3
PULSE AMPLIFIER

NOTE. 1,2,3, AND 4 FORM PART OF AN INTEGRATED OPERATIONAL AMPLIFIER CIRCUIT

Fig. A2.3

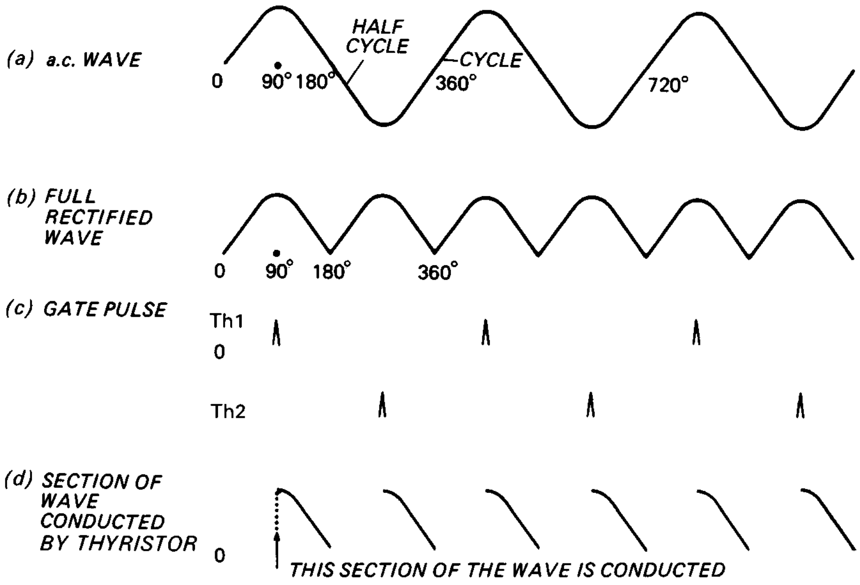
(a) a.c. WAVE
HALF CYCLE
CYCLE
0
90°
180°
360°
720°
(b) FULL RECTIFIED WAVE
0
90°
180°
360°
(c) GATE PULSE
Th1
0
Th2
(d) SECTION OF WAVE CONDUCTED BY THYRISTOR
0
THIS SECTION OF THE WAVE IS CONDUCTED

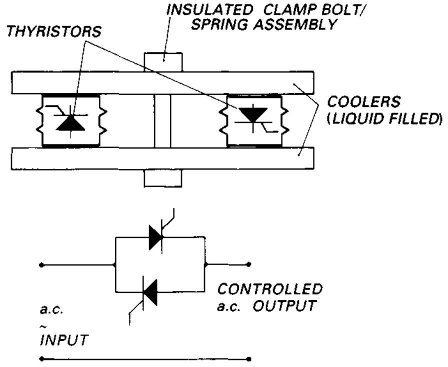
THYRISTORS
INSULATED CLAMP BOLT/ SPRING ASSEMBLY
COOLERS (LIQUID FILLED)
a.c. ~ INPUT
CONTROLLED a.c. OUTPUT

Fig. A2.4 Typical a.c. controller
(thyristors in anti-parallel).

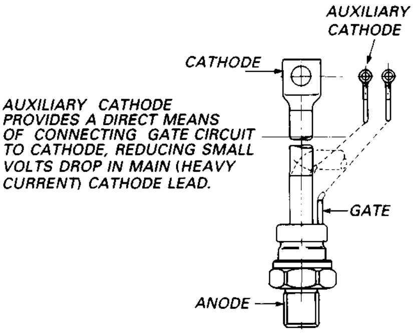
AUXILIARY CATHODE
CATHODE
AUXILIARY CATHODE PROVIDES A DIRECT MEANS OF CONNECTING GATE CIRCUIT TO CATHODE, REDUCING SMALL VOLTS DROP IN MAIN (HEAVY CURRENT) CATHODE LEAD.
GATE
ANODE

Fig. A2.4 Thyristor: 325 A (RMS);
length ~ 210 mm.

A number of other circuits are available, including three phase and a.c. controllers, the latter producing a regulated *alternating* output (Fig. A2.4).

Appendix 3

Proprietary gases and mixtures

Cylinder colour identification

Gases and gas mixtures are identified by colours as recommended in BS 349, 1973. The contents of the cylinder are indicated by the name(s) and chemical formula(e) on a label on the shoulder of the cylinder. Secondary identification is by the use of recommended base colours on the body of the cylinder and coloured bands on the cylinder to indicate the type of gas.

Gas	Cylinder colour
Acetylene	Maroon
Air (not breathing quality)	Grey
Argon	Blue
Carbon dioxide, commercial vapour withdrawal	Black
Carbon dioxide, commercial liquid withdrawal	Black with white stripes down length of cylinder
Hydrogen	Red
Nitrogen	Grey with black shoulder
Oxygen	Black
Propane	Red, wider and shorter cylinder
Argon/carbon dioxide	Blue, green band on cylinder
Argon/helium	Blue, green band on shoulder
Argon/hydrogen	Blue, red band on shoulder
Argon/oxygen	Blue, black band on shoulder
Nitrogen/hydrogen	Grey with red band or red with grey band

Fig. A3.1. Special mixtures.

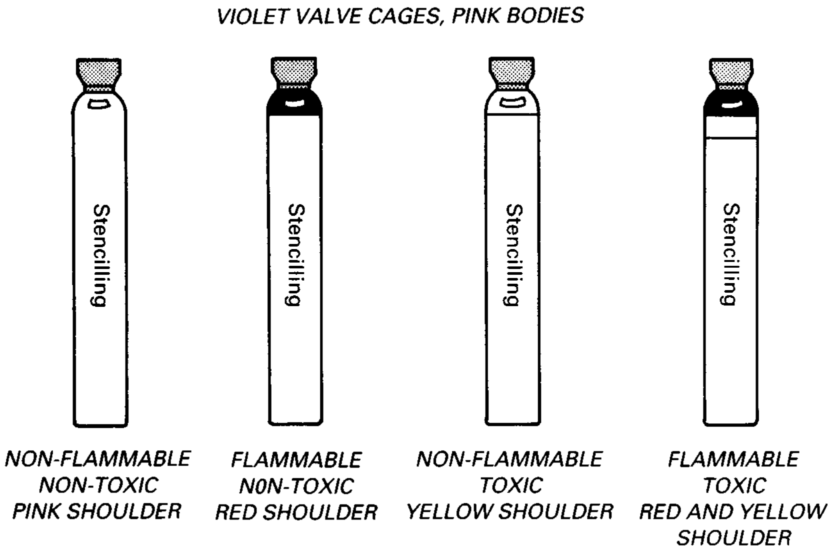
VIOLET VALVE CAGES, PINK BODIES
Stencilling
Stencilling
Stencilling
Stencilling
NON-FLAMMABLE NON-TOXIC PINK SHOULDER
FLAMMABLE NON-TOXIC RED SHOULDER
NON-FLAMMABLE TOXIC YELLOW SHOULDER
FLAMMABLE TOXIC RED AND YELLOW SHOULDER

AGA GAS Ltd. Selection of gases

Gas	Cylinder contents	Applications
Argon	Ar 99.99 %	TIG, MIG (GTAW, GMAW) and plasma welding, root purging. Nickel based alloys and high alloy steels (stainless steel). Copper and its alloys
Helium	He	Aluminium welding with a.c. and also straight polarity d.c. (d.c.e.n.)
AGA MIX® UN	Ar 80 %, CO_2 20 %	MAG (GMAW) welding mild and low alloy steel
AGA MIX® ST	Ar 98 %, CO_2 2 %	MIG, MAG (GMAW) welding stainless and low alloy steel. Not low carbon stainless steel
AGA MIX® SP	Ar 92 %, CO_2 8 %	MAG welding low alloy steel and mild steel, mainly spray and pulse
AGA MIX® SH	Ar 75 %, CO_2 25 %	MAG welding mild and low alloy steel, mainly short arc
AGA MIX® O_2	Ar 98 %, O_2 2 %	MAG welding stainless steel not less than 0.03 % C. Mainly for spray arc
AGA MIX® H5	Ar 95 %, H_2 5 %	Plasma cutting
AGA MIX® H20	Ar 80 %, H_2 20 %	Plasma cutting
AGA MIX® H35	Ar 65 %, H_2 35 %	Plasma cutting
AGA MIX® He30	Ar 70 %, He 30 %	Welding copper and its alloys
AGA MIX® He70	Ar 30 %, He 70 %	Welding copper and its alloys and titanium
AGA MIX® N_2	Ar 98 %, N 2 %	Welding duplex stainless steels
AGA MISON®	Ar with less than 0.03 % NO	In every case where argon is indicated
AGA MIX® NH	H_2 10 %, N_2 90 %	Used as root protection when welding stainless steel
Argon + (pure)		Welding titanium
Nitrogen	N_2	Gas purging. Plasma cutting

Gases marked ® are trade names of AGA GAS Ltd.

Industrial gases and mixtures supplied by Air Products plc

Name	Cylinder contents	Cylinder colours	Applications
Oxygen	O_2	Black	
Nitrogen	N_2	Grey with black shoulder	
Helium	He	Brown	
Argon	Ar	Peacock blue	
Hydrogen	H_2	Red	
Coogar® 5	Ar 93%, CO_2 5%, O_2 2%	Peacock blue with pastel purple shoulder and contents stencilled in black	MIG/MAG welding of steel 10 mm thick and below
Coogar® 20	Ar 78%, CO_2 20%, O_2 2%		MIG/MAG welding of steel 10 mm thick and above
Coogar® SG	Ar 83%, CO_2 15%, O_2 2%		Synergic pulse and standard MIG/MAG welding of steels, all thicknesses and positions
Astec® 15	He 75%, Ar 23.5%, CO_2 1.5%	Pale blue with contents stencilled in black	MIG/MAG welding of low alloy steels and austenitic stainless steels with dip transfer
Astec® S3	He 35%, Ar 63%, CO_2 2%		Synergic pulse and standard MIG/MAG for high quality welds in stainless steels
Astec® 25	He 75%, Ar 25%		MIG and TIG applications on high thermal conductivity – non-ferrous metals
Astec® 30	He 30%, Ar 70%		MIG and TIG applications on high thermal conductivity – non-ferrous metals
Astec® 50	He 50%, Ar 50%		MIG/MAG and synergic pulse weld overlay/surfacing applications

Quasar®	Ar 98%, O_2 2%	Supplied in liquid form by Minitank	MIG/MAG welding of steels and stainless steels, recommended for use with robotics or automatic applications on thinner section metals
Hytec® 1	Ar 99%, H_2 1%	Pastel purple with red shoulder and contents stencilled in black	Manual TIG applications on stainless steel
Hytec® 3	Ar 97%, H_2 3%		Manual TIG applications on stainless steel
Hytec® 5	Ar 95%, H_2 5%		Automatic and orbital TIG, plasma welding
Hytec® 35	Ar 65%, H_2 35%		High quality plasma cutting of stainless steel, aluminium and nickel alloys
Acetylene	C_2H_2	Maroon	Oxy-acetylene cutting, brazing and welding applications
Apachi-plus®	C_3H_8	Orange with red neck	Fuel gas for cutting, brazing and thermal spraying applications
Propane	C_4H_{10}	Red with green neck	Heating and cutting

Gases marked ® are the trade names of Air Products plc.

BOC Ltd (British Oxygen Co Ltd): Shieldmaster range of gases for welding and cutting

Gas		Cylinder contents	Cylinder colours	Applications
Oxygen	h[a]	Oxygen O_2 (99.7% purity)	Black	
Nitrogen	l[b]	Nitrogen N_2 (99.99% or 99.999% purity)	Grey with black shoulder; two white circles on shoulder indicate the higher purity	
Helium	l	Helium He	Brown	
Argon	h	Pure argon Ar (99.997% purity)	Blue	
		High purity argon (99.999% purity)	Blue plus a certificate of quality	
Carbon dioxide	l	Carbon dioxide CO_2 (to BS 4105 99.8% purity)	Vapour withdrawal black; liquid withdrawal black with two vertical white stripes, full length of the cylinder	
Argoshield® 5	h	Ar 93%, CO_2 5%, O_2 2%	Blue with green band round middle of cylinder and green shoulder; individual gases denoted by shoulder label and identity disc round neck	GMA/MIG welding thin mild steel sections (dip transfer) up to 3 mm thick
Argoshield® TC	h	Ar 86%, CO_2 12%, O_2 2%		GMA/MIG welding mild steel (dip and spray transfer) 3–10 mm thick; mini MIG welding mild steel
Argoshield® 20	h	Ar 78%, CO_2 20%, O_2 2%		GMA/MIG welding mild steel (spray transfer) above 10 mm thick
Pureshield® P1		Ar 98.5%, H_2 1.5%	Blue with red band round middle of cylinder; identity disc round neck; applications and identity label above red band	GTA/TIG welding austenitic stainless steel
Pureshield® P2		Ar 65%, H_2 35%		Plasma cutting aluminium and stainless steel
Helishield® H1	l	He 83%, Ar 15%, CO_2 2% and a trace of oxygen	Brown with blue band round middle of cylinder; identity disc round neck; applications and identity label above blue band	GMA/MIG welding stainless steel – dip transfer
Helishield® H2	l	He 75%, Ar 25%		GMA/MIG and GTA/TIG welding copper and thicker sections of aluminium
Helishield® H3	h	He 30%, Ar 70%		GMA/MIG and GTA/TIG welding aluminium and aluminium alloys
Helishield® H5	l	He 50%, Ar 50%		GMA/MIG and GTA/TIG welding nickel and nickel alloys
Helishield® H101	l	He 40%, Ar 60% plus some CO_2		GMA/MIG welding stainless steel, low alloy steel; spray transfer
Argomix®	h	Ar 80%, CO_2 20%	Blue with green band round middle of cylinder	Special applications of GMA/MIG welding mild steel; spray transfer
Argon/Hydrogen	l	Special mixtures as required	Blue with red band round middle of cylinder	Special GTA/TIG welding applications
Argonox®	h	99/1, 98/2, 95/5% argon/oxygen	Blue with black band round middle of cylinder	GMA/MIG welding stainless and high alloy steels
Acetylene	l	Acetylene C_2H_2	Maroon	Bevelling, high speed flame cutting and welding mild steel; brazing
Propylene	h	Propylene C_3H_6	Red body, pink top and bottom	Metal spraying
Propane	h	Propane C_3H_8 to BS 4250	Red	Flame cutting mild steel; pre-heating

[a] h – heavier than air.

[b] l – lighter than air.

Gases marked ® are trade names of BOC Ltd.

Note GMA (gas metal arc) is the American term for MTG); GTA (gas tungsten arc) is the American term for TIG.

Manifolded cylinder pallet (MCP) contains 15 N size cylinders connected to a single outlet. The N cylinder is usually 200 bar pressure as is the MCP manifold.

Distillers AG Ltd: gases for welding and cutting

Gas or gas mixture	Cylinder contents			Cylinder colours	Applications
Oxygen	O_2			Black	Welding and cutting
Nitrogen	N_2			Grey, with black shoulder and guard	Freezing, purging, certain steels, plasma cutting
Helium	He			Brown cylinder, shoulder and guard	TIG welding of non-ferrous metals and alloys. General TIG applications
Argon	Ar			Peacock blue	Non-toxic, smooth arc, easy start, little spatter. Used in MIG and TIG welding and in gas mixtures
Hydrogen	H_2			Red cylinder, shoulder and guard	MAG welding of stainless steel. Very inflammable
Carbon dioxide	CO_2			Black with white stripe down cylinder length. For commercial liquid withdrawal	Sometimes used on its own for welding. Also mixed with other gases (see below)
Propane	C_3H_8			Red cylinder with valve core blue	For cutting and heating when used with oxygen
Acetylene	C_2H_2			Maroon	Welding and cutting
Air (blended)	Air			Grey cylinder and cage	Cutting
Argon and argon-based					
	Ar	CO_2	O_2		
Krysal® 5	93%	5%	2%	Peacock blue, with pink shoulder and guard	Thin low carbon steel. Low spatter, pulse arc. All positions
Krysal® 20	78%	20%	2%	Peacock blue, with pink shoulder and guard	Thick low carbon steel. Reduced spatter. Not suitable for pulse

Krysal® SM	85%	13.5%	1.5%	Peacock blue, with pink shoulder and guard	Most sections low carbon steel. Smooth arc. Minimum spatter. Suitable pulse arc and has robotic uses
Krysal® G8	92%	8%	—	Peacock blue with pink shoulder and guard	CO_2 component controls penetration and bead shape for weld strength. Improves wetting and viscosity
Krysal® 155	80%	15%	5%	Peacock blue, with pink shoulder and guard	The oxygen content stabilizes the arc, improves wetting and gives good fusion
Excellar® 8	Argon-based			Bulk liquid form	Developed for MAG welding of carbon steels and good for automatic operations
Argon–oxygen					
	Ar	O_2			
Argon/oxygen 1%	99%	1%		Blue, with black shoulder; contents stencilled in black	Used for the MIG and MMA welding for stainless steel. Reasonable spatter
Argon/oxygen 2%	98%	2%			
Argon/oxygen 3%	97%	3%			
Helium based					
	Ar	He	CO_2		
Stellar® 30	70%	30%	—	Mid-brown with pink shoulder and guard	Used when argon–hydrogen is inappropriate. Deep penetration. As the helium percentage rises the arc becomes less stable. Higher helium mixtures are ideal for non-automated processes. For MIG welding of non-ferrous metals, aluminium, copper etc. TIG welding of most materials. Stellar 50 and Stellar 75 are lighter than air
Stellar® 50	50%	50%	—		
Stellar® 75	25%	75%	—		

Gas or gas mixture	Cylinder contents			Cylinder colours	Applications
Stellar® T1	minor	major	trace	Mid-brown with pink shoulder and guard	For MAG welding of stainless steels. T1 and T6 are lighter than air
Stellar® T6	major	minor	trace		
Stellar® T8	major	minor	trace		
Helihi®	Helium–air			Cylinder body brown, cage pink	Balloon gas
Argon–hydrogen					
	Ar	H_2			
Hylar® 1	99%	1%		Peacock blue, with signal red shoulder and pink guard	For TIG welding of stainless steels
Hylar® 2	98%	2%			
Hylar® 3	97.5	1.5%			
Hylar® 5	95%	5%			
Hylar® 35	65%	35%			
Shielding gases for plasma welding					
Argon					Used both as orifice and shielding gas with some equipment.
Hylar® 5					Used as a shielding gas for plasma welding thin material.
Hylar® 35					Used for plasma cutting in certain applications

Mixtures of gases containing more than 5% hydrogen can be explosive.

Argon is non-toxic, but both argon and carbon dioxide are asphyxiants in high concentrations. Carbon dioxide dissociates in the arc; the oxygen produced increases wetting. The occupational exposure standard (OES) has been set at 5000 vpm. Carbon dioxide in the atmosphere is approximately 350 vpm.

A toxic gas is one that in its pure form has an OES as defined by the Health and Safety Executive of less than 500 vpm.

Gases marked ® are trade names of Distillers AG Ltd.

Appendix 4

Tests for wear-resistant surfaces

The following tests for wear are used

(1) *Adhesive wear test* in which a test specimen in the form of a block is pressed or loaded against a rotating alloy wear test specimen in the form of a ring or of a comparison standard. The block and ring may be of similar or dissimilar alloy and the ring may be lubricated (optional). The volume loss for various loads is measured for a given number of revolutions.

(2) *Abrasive wear test* in which dry sand is fed between the surfaces of a wear test specimen which is pressed or loaded against a rotating rubber wheel. The volume loss of the specimen for a given number of revolutions and load is measured.

(3) *Galling test* in which a rotating cylindrical pin made from undiluted weld metal (e.g. TIG) is pressed (loaded) against a standard ground block and the stress in kg/mm^2 at which visual sign of material transfer (galling) occurs after one revolution of the pin is noted.

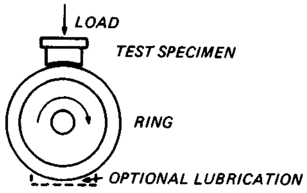
LOAD
TEST SPECIMEN
RING
OPTIONAL LUBRICATION

a ADHESIVE WEAR

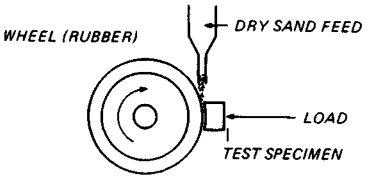
WHEEL (RUBBER)
DRY SAND FEED
LOAD
TEST SPECIMEN

b ABRASIVE WEAR

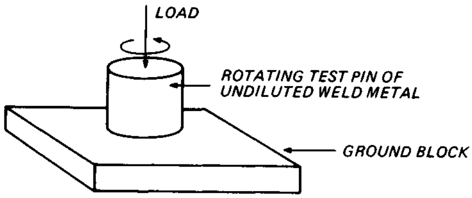
LOAD
ROTATING TEST PIN OF UNDILUTED WELD METAL
GROUND BLOCK

c GALLING (MATERIAL TRANSFER)

Appendix 5

Conversion factors

inch	mm	Preferred mm	inch	mm	Preferred mm
$\frac{1}{64}$	0.4	—	$\frac{5}{16}$	8.0	8
$\frac{1}{32}$	0.8	—	$\frac{3}{8}$	9.5	10
$\frac{1}{16}$	1.6	—	$\frac{7}{16}$	11.1	11
$\frac{3}{32}$	2.4	—	$\frac{1}{2}$	12.7	12
$\frac{1}{8}$	3.2	3	$\frac{9}{16}$	14.4	—
$\frac{5}{32}$	4.0	4	$\frac{5}{8}$	16.0	15
$\frac{3}{16}$	4.8	5	$\frac{11}{16}$	17.6	18
$\frac{1}{4}$	6.4	6	$\frac{3}{4}$	19.0	20
			$\frac{7}{8}$	22.4	22
			1	25.4	25

SWG	mm	SWG	mm
28	0.38	16	1.6
26	0.46	14	2.0
24	0.56	12	2.5
22	0.71	10	3.25
20	0.91	8	4.0
18	1.22	6	5.0
		4	5.9

1 kgf = 9.8 N (10 N approx.).
1 bar = 14.5 lbf/in^2 = 0.1 N/mm^2.
1 tonf/in^2 = 15.4 N/mm^2 = 1.54 hbar.
1000 mb = 1 bar = 14.5 lbf/in^2 = 1 kgf/cm^2.
1 hbar = 100 bar.
1 cu. ft = 0.028 m^3 = 28.3 litres.

Appendix 6

Low hydrogen electrode, downhill pipe welding

A d.c. power source is used as with cellulosic rods (p. 73) and is usually engine driven in the field, the electrode being connected to the – ve pole. Weld metal deposition is faster than with cellulosic or uphill low hydrogen methods and a short arc is used with no 'pumping' of the electrode. Since shrinkage is important, if internal clamps are used there should be more spacers. The preparation of a joint using different diameter pipes is show in Fig. A6.1 and the tack welds should be tapered at each end.

For the root run the electrode should lightly touch the joint and the weld should continue through 6 o'clock about 2 mm towards 7 o'clock to facilitate link up with the run to be deposited on the other side. There is no weave and keeping a short arc and high travel rate the slag is kept above the

Fig. A6.1. Joint preparation.

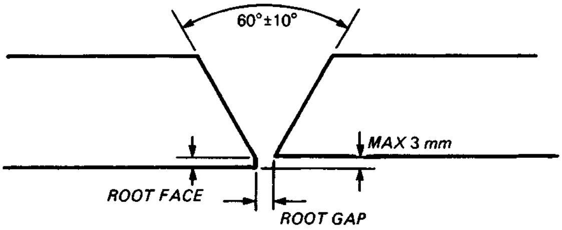

ROOT FACE

ELECTRODE DIAMETER	FACE
2.5 mm 3.25 mm	

ROOT GAP

ELECTRODE DIAMETER	GAP
2.5 mm	2.5 mm + 1.0 mm
3.25 mm	2.5 mm + 1.0 mm

Note: further reduction of the root gap is possible in thinner wall thicknesses, for experienced pipeline welders using the 3.25 mm electrode.

Fig. A6.2

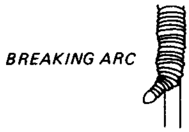
BREAKING ARC

Fig. A6.3

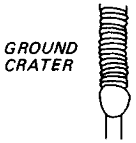
GROUND
CRATER

arc at a distance of about the coating diameter. The arc is broken by turning into the bevels and the metal is tailed off as shown in Fig. A6.2.

The finishing crater should be ground to a pear-shaped taper (Fig. A6.3) and the next electrode is started about 10 mm above the top end of the tapered crater. Slag is removed from each run with power brush and chipping hammer and grinding is used to remove any excess metal generally from 4 to 6 o'clock. The penetration bead may be slightly concave but has a smooth tie-in with the pipe metal and is acceptable. Too much heat gives increasingly convex penetration and is due to too slow travel or too high a current.

Avoid burn through between 12 and 1 o'clock when using a leading arc by keeping the arc as short as possible and directing it alternately to each side of the joint. Fig A6.4 shows angles of electrode, electrode diameters and currents.

For the filling and capping runs the arc is kept short with a weave not exceeding 3 to 4 times the core wire diameter and the slag–electrode distance

Fig. A6.4. Angles of electrode, electrode diameters and currents, root run.

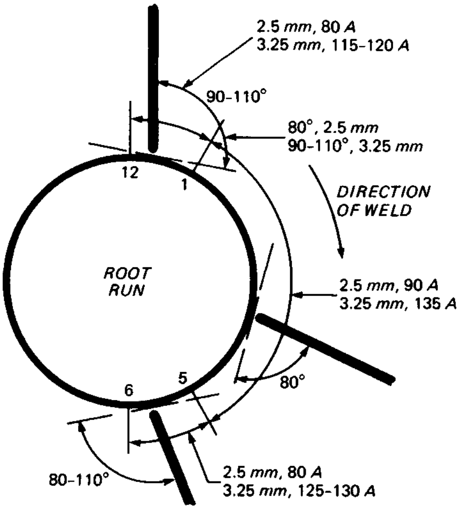
2.5 mm, 80 A
3.25 mm, 115-120 A
90-110°
80°, 2.5 mm
90-110°, 3.25 mm
DIRECTION
OF WELD
12
1
ROOT
RUN
2.5 mm, 90 A
3.25 mm, 135 A
80°
6
5
80-110°
2.5 mm, 80 A
3.25 mm, 125-130 A

should be about 10 mm (Fig. A6.6). Fig A6.5 shows angles of electrode, electrode diameter and currents, Fig. A6.6 the slag–electrode tip distance and Fig A6.7 the restrike and electrode movement.

Low hydrogen electrodes are supplied in sealed air-tight tins and should be stored as on p.13, the hydrogen content being 5 ml per 100 g of weld metal (IS03690). If the electrodes are left unused for 8 hours or more they should be rebaked at 280 °C + 20°C for 1 hour and stored in a holding oven at 120–150 °C until used. They should not be exposed to any damp conditions

Fig. A6.5. Angles of electrode, electrode diameters and currents, filling and capping runs.

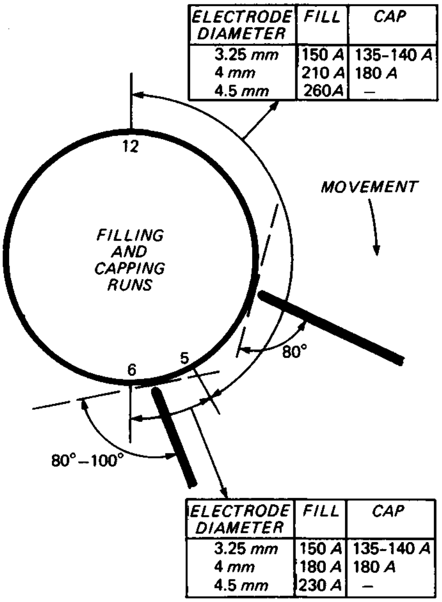
12
FILLING AND CAPPING RUNS
MOVEMENT
80°
6
5
80°–100°

ELECTRODE DIAMETER	FILL	CAP
3.25 *mm*	150 *A*	135-140 *A*
4 *mm*	210 *A*	180 *A*
4.5 *mm*	260*A*	–

ELECTRODE DIAMETER	FILL	CAP
3.25 *mm*	150 *A*	135-140 *A*
4 *mm*	180 *A*	180 *A*
4.5 *mm*	230 *A*	–

Fig. A6.6

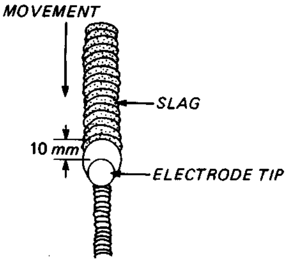
MOVEMENT
SLAG
10 mm
ELECTRODE TIP

Fig. A6.7

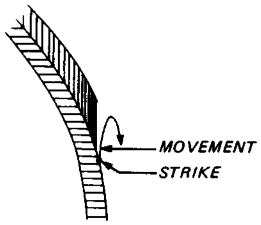
MOVEMENT
STRIKE

during use and the joint should be completely free of all moisture. The weld metal has high resistance to cracking and is of composition C 0.06–0.09%, Mn 1.00–1.40%, Si 0.03–0.07% and classification AWS A5.5 E8018-G BS 639 E5154B 120 90(H).

Appendix 7

The manufacture of extruded MMA electrodes

As previously stated, almost all electrodes are produced by the extrusion process. The present day electrode consists of a core wire, centrally placed in a coating, usually silicate bonded (Vol. 1, p. 60). Double coated electrodes are also produced.

A semi-automatic production line may consist of the following:

(1) *Wire drawing.* Non-slip hydraulic machines draw the wire through dies to the correct size, after which the wire is stored on reels.

(2) *Wire cutting.* The wire is then straightened and cut to electrode length, giving burr-free wire rods of any required length which are stacked in bins.

(3) *The coating.* The raw materials of the coatings are kept in raw material silos and from these the materials are sieved and dosed. It is the type and composition of these materials that determines the characteristic of the particular electrode. The dosing installations consist of a number of silos each fitted with a discharge system, dosing control and weighing facilities.

(4) *Silicate storing* consists of tanks that are temperature controlled and from which the silicate is pumped to the mixer.

(5) *Dry blending and wet mixing.* The batch is first thoroughly mixed in the dry state and then a measured quantity of silicate is added from a dosing unit. The wetted batch is then roughly mixed, emptied and removed to the slug press, where it is pressed into slugs ready for extrusion.

(6) *Extrusion.* The slugs are placed into the charging trough and when this is full the load of slugs is thrust into the extrusion cylinder. The breech of the hydraulic extrusion cylinder is then locked by a simple lever movement. The cut and deburred wires are stored adjacent to the wire feed from which, by opening a slide in the bin

side, the wire rods drop into the feed magazine. The extruder piston is then operated and when the flux flow pressure is reached the wire feed commences. Accelerator and drive rolls take the rods dropping from the magazine and push them axially through the press to the extrusion head where they receive their coating.

(7) *Concentricity check and brushing.* It is important that the core wire should be centrally placed in the covering (pp. 5–6). This is checked electronically and adjustments made if required. A transfer belt then takes the electrodes to a brushing station where the ends for the electrode holder and the striking tip are formed, after which the number or type of electrode is printed on the coating by a printing machine.

(8) *Drying and baking.* The electrodes are air dried to prevent cracking of the coating and are then loaded into baking ovens, after which they are ready for packing.

Fig. A7.1. Simplified diagram of electrode production process.

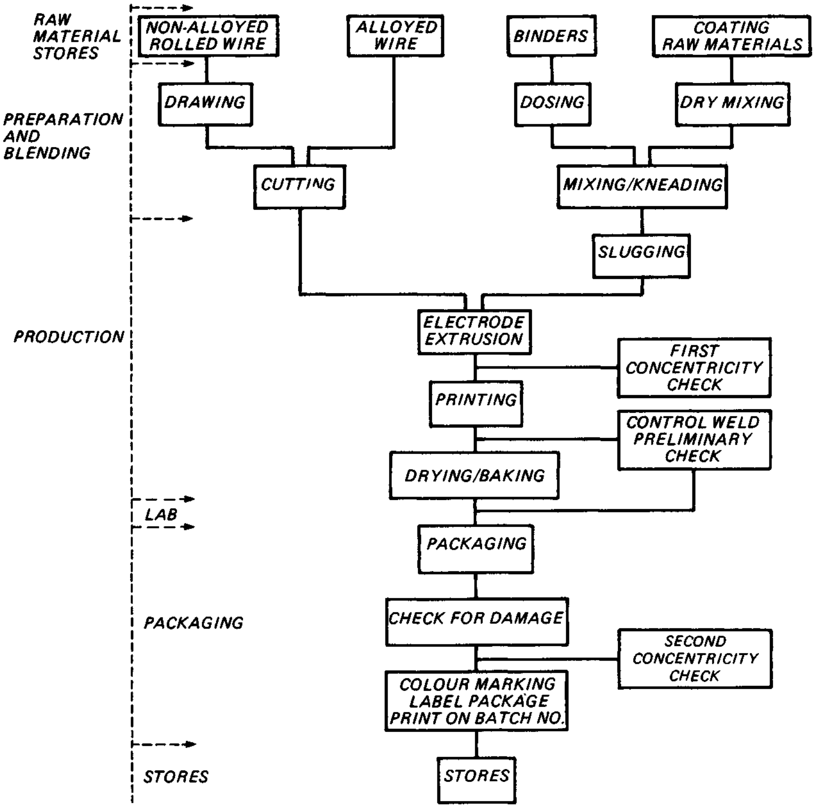
RAW MATERIAL STORES
NON-ALLOYED ROLLED WIRE
ALLOYED WIRE
BINDERS
COATING RAW MATERIALS
PREPARATION AND BLENDING
DRAWING
DOSING
DRY MIXING
CUTTING
MIXING/KNEADING
SLUGGING
PRODUCTION
ELECTRODE EXTRUSION
FIRST CONCENTRICITY CHECK
PRINTING
CONTROL WELD PRELIMINARY CHECK
DRYING/BAKING
LAB
PACKAGING
PACKAGING
CHECK FOR DAMAGE
SECOND CONCENTRICITY CHECK
COLOUR MARKING LABEL PACKAGE PRINT ON BATCH NO.
STORES
STORES

Where high capacity is required, two extruders are placed in parallel. When one is being charged the other is working, so that the same brushing and drying ovens are used and delay in production is minimised. In many cases the high speed wire drawing and cutting machines are fully automatic and the drying ovens are of the bucket type with conveyor chains up to 1000 m long and temperatures up to 450 °C.

Appendix 8

Notes on fire extinguishing (BS 5306, Part 3: Code of practice for selection, installation and maintenance of portable fire extinguishers)

For a fire to occur we need *oxygen, fuel* and *heat.*

Oxygen is present in the atmosphere and in the compressed form contained in cylinders used for oxy-acetylene or oxy-propane cutting.

Fuel may be any combustible substance such as petrol, diesel oil, paper, cleaning materials, rubber or litter etc. – anything that will burn.

Heat may be supplied by discarded cigarette ends, match sticks and, in the welding shop, by sparks from the arc or bad connections such as those at the earth or return terminal to the set. The arc is an intense source of heat and, together with the sparking, will rapidly cause ignition of any combustible solids or liquids in the immediate vicinity.

If any one of these above causes is removed, the fire will be contained. For example the oxygen supply from the atmosphere may be contained by smothering the fire with a blanket or dry sand or by turning off the oxygen supply if it is from a cylinder. The heat can be controlled by ceasing to use the arc and switching off all power supplies to prevent further arcing at poor contacts.

Classes of fire with suitable extinguishing media

(*A*) Solid materials (carbonaceous fires forming glowing embers). Use water type extinguisher (cooling method) except on electrical apparatus.

(*B*) Liquids or liquifiable solids (petrol, paraffin, wax, etc.). Use foam, dry powder, CO_2, vaporizing liquid, fire blanket, dry sand. (smothering method).

(*C*) Flammable gases, liquid petroleum gas including propane and butane, acetylene. Turn off the supply (starvation method).

(*D*) Metals (magnesium, metallic powders). Special powders containing sodium or potassium chloride are required. Such fires should be dealt with by the fire brigade.

Types of extinguishers

(1) *Carbon dioxide CO_2.* This contains liquid CO_2 under pressure, which is released as a gas when the trigger or plunger on the extinguisher is operated.
(2) *Extinguishing powder (gas cartridge).* The powder is expelled by the pressure from a gas cartridge.
(3) *Extinguishing powder (stored pressure).* The powder is expelled by pressure stored within the body of the extinguisher.
(4) *Foam (chemical).* Chemicals stored in the body of the extinguisher are allowed to mix and react, producing foam, which is expelled from the extinguisher.
(5) *Foam (gas cartridge).* Foam is expelled from the extinguisher by the pressure from a gas cartridge.
(6) *Foam (mechanical stored pressure).* Pressure stored in the body of the extinguisher expels the foam.
(7) *Halon (stored pressure).* The body of the extinguisher contains a halon which is expelled by pressure stored within the body.
(8) *Water (gas cartridge).* Water is expelled from the extinguisher by the pressure from a gas cartridge.
(9) *Water (stored pressure).* Water is expelled by pressure stored within the extinguisher body.
(10) *Water (soda-acid).* An acid–alkaline reaction within the body produces the pressure to expel the water.

Extinguishers containing dry powder, CO_2 and vaporizing liquids

These extinguishers are suitable for fires involving flammable liquids and live electrical apparatus (as are met with under welding conditions). For fires in liquids (containers or spillage) the jet or horn should be directed at the near side of the fire, which should be driven away from the operator to the far side with a side-to-side motion of the jet. If falling liquid is alight begin at the bottom and move upwards. On electrical apparatus direct the jet directly at the fire after first switching off the mains supply. If the latter is not possible, contain the fire and summon the fire brigade.

If there is no shut-off valve on the extinguisher continue discharging it over the fire area until empty. If the discharge is controlled, shut off the extinguisher discharge when the fire is extinguished and hold it ready for any further outbreak which may occur when the atmosphere has cleared.

Do not use vaporizing liquid extinguishers in confined spaces where there is poor ventilation and a danger of fumes being inhaled.

Foam extinguishers give semi-permanent protection against contained fires and will partially extinguish a fire, which will not gain its full intensity until the foam covering is destroyed. They may be used on liquids, in tanks to protect them from ignition or from giving off flammable vapours. Dry powder and foam should not be used together unless a specially compatible powder/foam are used.

BCF extinguishers

When this type is used the bromochlorodifluoromethane, which is a stable compound, appears as a heavy mist that blankets the fire and excludes the atmospheric oxygen. It is more efficient than chlorobromomethane (CB), methyl bromide (MB) or carbon tetrachloride (CTC). It does not damage normal materials and leaves no deposit. It decomposes slightly in the heat of a flame but the amount of harmful gas produced (which has a pungent smell) has no effect on adjacent personnel even when excess BCF is used. It has high insulating properties and can be used on electrical equipment and is particularly suitable for the welding or fabrication shop since it can be used on Class *A* fires in addition to those of Class *B* and *C*.

Fire prevention where welding processes are involved

General precautions

(1) All areas must be free of litter and combustible waste with all combustible material stored in metal bins, with lids, well away from the welding processes.

(2) Make certain that all electrical cables are not chafing and connections are secure and will not cause arcing. This applies particularly to the earth or return connection to the welding set.

(3) All chemicals including cleaning fluids should be kept away from the welding area, in a store room.

(4) Oxygen cylinders should be kept in a room apart from fuel gases such as acetylene and propane (LPG), which should be kept in another room. All cylinders and manifolds should be turned off except when in use. Keep cylinder keys with the appropriate cylinders in case of urgent need.

(5) No smoking signs should be prominently displayed and should be strictly adhered to.

(6) Defects in plant and equipment should be reported at once and indicated. Faulty electrical equipment is always a fire hazard.

(7) Always read the instructions for operation of the extinguisher situated in your shop.

Notes on procedures to be followed when using welding apparatus

Leakage of oxygen, as for example from a cylinder used in the cutting process, may lead to the ignition of any glowing or red-hot material due to the rapid oxidizing action of the escaping oxygen. The gas must be turned off at the cylinder valve (leave the cylinder key attached to the cylinder so that it is never misplaced). Remember that should flames play on a cylinder containing oxygen, there is a danger of explosion due to the high pressure involved – although oxygen is non-combustible it is a powerful supporter of combustion and may cause re-ignition.

Fuel gas cylinders

LPG, acetylene, etc. and accompanying apparatus and appliances

(1) Gas leak on fire. If the fire is away from regulators and cylinders first turn off the gas supply at the cylinder or appliance valve. If this is not possible and the fire is spreading to nearby combustible material, extinguish the flames and then seal the leak, taking care that re-ignition is not possible.

If the gas is on fire near a regulator or cylinder valve first extinguish the flames if this is possible within two or three minutes of the beginning of the combustion and use a cloth or gloves (e.g. welders) to turn off the valve and guard against re-ignition.

If the valve cannot be shut, evacuate the area and call the fire brigade.

If the gas is on fire and the flames are impinging on the gas cylinders, first extinguish the flames if this is possible within two to three minutes of the conflagration. If this is not possible, divert the flames from the affected cylinder away from the other cylinders and combustible material, evacuate the area and call the fire brigade.

Any cylinders exposed to the heat of fire should be removed within 2–3 minutes. If this is not possible, evacuate the area and call the fire brigade.

A burning gas leak can be extinguished by a dry powder, BCF or water type fire extinguisher in this order of preference but small fires may be smothered with a wet blanket or cloth.

A water type fire extinguisher should be used on any adjacent burning material and whilst dry powder or BCF types can be used in this case they will have only a limited effect.

Do not continue to fight a fire if there is a danger of gas cylinders exploding or if the fire continues to grow in spite of your efforts to contain it. Remember that extinguishers are only intended to deal with small fires so that the quicker the extinguisher is brought into use the better as fires spread quickly. Position yourself between the fire and the exit from the shop so that your escape will not be cut off.

If there is any possibility of the conflagration getting out of control in the slightest degree, call the Fire Brigade whose personnel are trained in fire fighting. They must be met on arrival and the officer in charge advised of the presence and position of all gas cylinders.

(2) Gas leak. No fire.

(*a*) The danger is from an explosion if the fuel gas is ignited. Keep all sources of ignition away from the gas. Do not smoke, switch on electric lights, electrical apparatus or telephones, etc. and do not drive vehicles in the vicinity.

(*b*) Shut off the cylinder valve or pipeline. If this is impossible move the cylinder(s) to the open air so as to dissipate the gas.

(*c*) Do not re-occupy the shops until they are properly ventilated and pay attention to pockets of gas that may exist in ducting, sumps and pits and see that they are completely purged.

Appendix 9

Table of brazing alloys and fluxes

General purpose silver brazing alloys: cadmium bearing

Description	Nominal composition	Melting range (°C)		BS 1845	Remarks
		Solidus	Liquidus		
Easy-flo	50% Ag:Cu:Cd:Zn	620	630	AG1	Cadmium-bearing alloys generally offer the best combination of melting range, flow characteristics and mechanical properties. Thus, where the presence of cadmium is acceptable, they would normally be recommended as the most economical alloys. Easy-flo, Easy-flo No. 2 and DIN Argo-flo are the lowest melting point alloys available. All three alloys have excellent flow characteristics. The highest joint strengths are obtained with Easy-flo. Alloys with 30–40% silver are normally considered as general-purpose alloys. They still offer good flow characteristics and melting ranges and are suitable for mechanized brazing operations. The low silver alloys have good fillet-forming characteristics and are widely used in fabricating brass fittings and furniture. Owing to their wide melting ranges they are not generally used as preforms due to the danger of liquation.
Easy-flo No. 2	42% Ag:Cu:Cd:Zn	608	617	AG2	
DIN Argo-flo	40% Ag:Cu:Cd:Zn	595	630	—	
Argo-flo	38% Ag:Cu:Cd:Zn	608	655	—	
Mattibraze 34	34% Ag:Cu:Cd:Zn	612	668	AG11	
Argo-swift	30% Ag:Cu:Cd:Zn	607	685	AG12	
Metalflo	25% Ag:Cu:Cd:Zn:Si	606	720	—	
Argo-bond	23% Ag:Cu:Cd:Zn	616	735	—	
Metalsil	20% Ag:Cu:Cd:Zn:Si	603	766	—	
Metalbond	17% Ag:Cu:Cd:Zn:Si	610	782	—	
Metaloy	13% Ag:Cu:Cd:Zn:Si	605	795	—	

General purpose silver brazing alloys: cadmium free

Description	Nominal composition	Melting range (°C)		BS 1845	Remarks
		Solidus	Liquidus		
Silver-flo 55	55% Ag:Cu:Zn:Sn	630	660	AG14	The range of cadmium-free alloys has been developed for applications where the presence of cadmium is not allowed due to the brazing environment or the service conditions of the assembly, i.e. food-handling equipment. Silver-flo 55 is the lowest melting-point alloy in the group and has been widely adopted as a substitute for Easy-flo No. 2. Silver-flos 30 to 45 are general-purpose alloys suitable for use on most common engineering materials. The alloys with less than 24% silver are widely used on copper and steel. They give a good colour match on brass, but require close temperature control due to their high melting temperature. Unlike the cadmium-bearing range, this group contains low silver content alloys such as Silver-flo 33, 24 and 16, with relatively high melting points, which make them ideal for step brazing.
Silver-flo 55	45%Ag:Cu:Zn:Sn	640	680	—	
Silver-flo 452	45%Ag:Cu:Zn:Sn	640	680	—	
Silver-flo 45	45% Ag:Cu:Zn	680	700	—	
Silver-flo 44	44%Ag:Cu:Zn	675	735	—	
Silver-flo 40	40%Ag:Cu:Zn:Sn	650	710	AG20	
Silver-flo 38	38% Ag:Cu:Zn:Sn	660	720	—	
Silver-flo 34	34%Ag:Cu:Zn:Sn	630	740	—	
Silver-flo 33	33%Ag:Cu:Zn	700	740	—	
Silver-flo 302	30%Ag:Cu:Zn:Sn	665	755	AG21	
Silver-flo 30	30%Ag:Cu:Zn	695	770	—	
Silver-flo 25	25%Ag:Cu:Zn	700	800	—	
Silver-flo 24	24% Ag:Cu:Zn	740	780	–	
Silver-flo 20	20% Ag:Cu:Zn:Si	776	815	—	
Silver-flo 18	18%Ag:Cu:Zn:Si	784	816	—	
Silver-flo 16	16%Ag:Cu:Zn	790	830	—	
Silver-flo 12	12%Ag:Cu:Zn	820	840	—	
Silver-flo 4	4% Ag:Cu:Zn	870	890	—	
Silver-flo 1	1% Ag:Cu:Zn:Si	880	890	—	

General purpose silver brazing alloys: phosphorus bearing

		Melting range (°C)			
Description	Nominal composition	Solidus	Liquidus	BS 1845	Remarks
Sil-fos	15% Ag:Cu:P	644	700**	CP1	These alloys are recommended for fluxless brazing of copper. They may also be used on brass and bronze with the application of Easy-flo Flux or Tenacity 4A Flux. They should not be used on ferrous- or nickel-base materials or nickel-bearing copper alloys. The silver bearing alloys are more ductile than the Copper-flos and are recommended where joints will be subjected to significant stress levels.
Sil-fos 5	5% Ag:Cu:P	650	710**	CP4	
Silbralloy	2% Ag:Cu:P	644	740**	CP2	
Copper-flo	7% P:Cu	714	800	CP3	
Copper-flo No. 2	6% P:Sb:Cu	690	800	CP5	
Copper-flo No. 3	6% P:Cu	714	850	CP6	

** A small percentage of higher melting point material remains at this temperature.

Silver brazing alloys for tungsten carbide

		Melting range (°C)			
Description	Nominal composition	Solidus	Liquidus	BS 1845	Remarks
Easy-flo No. 3	50% Ag:Cu:Zn:Cd:Ni	634	656	AG9	General-purpose cadmium-containing alloy with excellent mechanical properties.
Argo-braze 38	38% Ag:Cu:Zn:Cd:Ni	615	655	—	An economical alternative to Easy-Flo No. 3 which has proved adequate for many applications.
Argo-braze 49H	49% Ag:Cu:Zn:Ni:Mn	680	705	—	A cadmium-free alternative to Easy-flo No. 3. The addition of manganese enhances wetting on difficult carbides.
Argo-braze 49LM	49% Ag:Cu:Zn:Ni:Mn	670	710	—	A free-flowing alloy giving good wetting. Suitable for small carbides.
Argo-braze 40	40% Ag:Cu:Zn:Ni	670	780	—	Economical alloys with good wetting characteristics. Their application is restricted by their relatively high melting points.
Argo-braze 25	25% Ag:Cu:Zn:Ni:Mn	710	810	—	
Easy-flo Tri-foil 'C'	Easy-flo bonded to both sides of a copper shim	620	630	—	For brazing large pieces of cemented carbide. The copper ensures that a thick stress-absorbing joint is achieved.
Easy-flo Tri-foil 'CN'	Easy-flo bonded to both sides of a copper-nickel shim	620	630	—	Recommended for joints subjected to high compressive stresses during service.
AB49 Tri-foil	Argo-braze 49LM bonded to both sides of a copper shim	670	710	—	A cadmium-free alternative to Easy-flo Tri-foil 'C'.

Brazing fluxes

Fluxes are normally selected on the basis of two main criteria: 1. The melting range of the brazing alloy – both solidus and liquidus should fall well within the quoted working range of the flux. 2. The parent metals to be joined – some alloys, such as aluminium bronze, require special fluxes. Copper, brass and mild steel are effectively cleaned by all fluxes.

Description	Working range °C	DIN 8511	Availability	Remarks	Residue removal
Easy-flo Flux	550–800	F-SH1	Powder/Paste	A general-purpose flux with good fluxing activity and long life at temperature. The powder has good hot-rodding characteristics.	
Easy-flo Flux Dipping Grade	550–775	F-SH1	Paste	A highly active and fluid flux which exhibits a minimum of bubbling, making it ideal for induction brazing. This paste is formulated to give a thin, stable consistency suitable for dipping.	Residues are generally soluble in hot water. Where difficulty is encountered, immersion in 10% caustic soda is suggested.
Easy-flo Flux Stainless Steel Grade	550–775	F-SH1	Powder	More active than Easy-flo Flux. Recommended for use on stainless steel with the lower melting-point brazing alloys.	
Easy-flo Flux Aluminium Bronze Grade	550–800	F-SH1	Paste	Similar for Easy-flo Flux Paste but modified for use on alloys containing up to 10% aluminium.	When finished components are heavily oxidized, cleaning and flux removal may be accomplished by the use of 10% sulphuric acid
Tenacity Flux No. 6	550–800	F-SH1	Powder/Paste	An opaque flux due to the addition of elemental boron. Recommended for tungsten carbide, refractory metals and stainless steel.*	
Tenacity Flux No. 14	550–750	F-SH1	Powder	An extremely fluid flux principally developed to prevent red staining on brass.	
Tenacity Flux No. 4A	600–850	F-SH1/2	Powder	A general-purpose flux with good resistance to overheating, used with higher melting temperature.	
Tenacity Flux No. 5	600–900	F-SH1/2	Powder	Recommended for stainless steel and heavy assemblies, where flux exhaustion is likely to occur due to prolonged heating	Residues are virtually insoluble in water.
Tenacity Flux No. 125	750–1200	F-SH2/3	Powder/Paste	A general-purpose high temperature flux for use with the bronze alloys.	Immersion in 10% caustic soda or mechanical removal is recommended.
Tenacity Flux No. 12	800–1300	F-SH3	Powder	A fluoride-free, high temperature flux for use with 'B' Bronze and 'G' Bronze.	
Silver-flo	550–775	F-SH1	Paste	A general purpose flux paste which combines exceptional activity with low fluxing temperature and melt viscosity.	
Mattiflux 100	550–800	F-SH1	Paste	Combines a relatively low melt viscosity with a good resistance to overheating and is particularly well suited to brazing mild steel assemblies.	

* This flux is unsuitable for use on stainless steel where crevice corrosion is likely to be a hazard in service.

Specialised silver brazing alloys

Description	Nominal composition	Melting range (°C)		BS 1845	Remarks
		Solidus	Liquidus		
Silver-flo 60	60% Ag:Cu:Zn	695	730	AG13	Principally used for brazing assemblies which will be exposed to sea water. Easy-flo No. 3 and Silver-flo 55 are also suitable for marine service.
Silver-flo 43	43% Ag:Cu:Zn	690	775	AG5	
Argo-braze 56	56% Ag:Cu:In:Ni	600	711	—	Specifically developed to prevent crevice corrosion, which may result in rapid failure of joints in stainless steel when exposed to water.
Argo-braze 35	35% Ag:Cu:Sn:Ni	685	887	—	An economical alloy suitable for fluxless brazing. Principally used on stainless steel and nickel silver.
Silver-copper eutectic	72% Ag:Cu	778	778	AG7 and AG7V	Ideal for fluxless brazing of copper, nickel and metallized ceramics. Available in two grades: Grade 1 (AG7V), a high purity material for brazing assemblies which will operate in vacuum. Grade 3 (AG7) for general engineering.
RTSN	60% Ag:Cu:Sn	602	718	—	Suitable for fluxless furnace brazing. Relatively fast heating rates necessary to avoid liquation.
15% Manganese-silver	85% Ag:Mn	951	960	AG19	A copper-free alloy for brazing assemblies which will be in contact with ammonia.

Flux coated rods

Alloy	Nominal composition	Melting range (°C)	BS 1845	Remarks
Cadmium bearing				
Easy-flo No. 2	42% Ag:Cu:Zn:Cd	608–617	AG2	A very fluid alloy, ideal for small components.
DIN Argo-flo	40% Ag:Cu:Zn:Cd	595–630	—	A general-purpose alloy with good flow characteristics.
Mattibraze 34	34% Ag:Cu:Zn:Cd	612–668	AG11	Fillet-forming alloys with limited flow characteristics.
Argo-swift	30% Ag:Cu:Zn:Cd	607–685	AG12	
Cadmium free				
Silver-flo 55	55% Ag:Cu:Zn:Sn	630–660	AG14	A free-flowing alloy generally suitable for all applications where a neat appearance is required.
Silver-flo 40	40% Ag:Cu:Zn:Sn	650–710	AG20	
Silver-flo 33	33% Ag:Cu:Zn	700–740	—	A range of cadmium-free alloys with fillet-forming properties.
Silver-flo 30	30% Ag:Cu:Zn	695–770	—	
Silver-flo 20	20% Ag:Cu:Zn:Si	776–815	—	An economical alloy suitable for brazing mild steel and copper.

Copper base brazing alloys

Description	Nominal composition	Melting range (°C)		BS 1845	Remarks
		Solidus	Liquidus		
'B' Bronze	97% Cu:Ni:B	1081	1101	CU7	Ideal substitute for pure copper in furnace brazing where joint gaps range from interference to 0.5 mm.
'C' Bronze	86% Cu:Mn:Ni	965	995	—	Suitable for furnace brazing. Requires a dewpoint of better than − 40°C, or a small addition of flux.
'D' Bronze	86% Cu:Mn:Co	980	1030	—	Recommended for brazing cemented carbide into rock drills and similar percussively loaded joints.
'F' Bronze	58% Cu:Zn:Mn:Co	890	930	—	
'G' Bronze	97% Cu:Ni:Si	1090	1100	—	The group offers a range of brazing characteristics to match different material and heat-treatment requirements.
Argentel	49% Cu:Zn:Ni:Si	913	930	CZ8	Low cost alloys ideal for use on tubular steel furniture and display racks. The high strength of Argentel makes it particularly suitable for butt joints while Argentel B offers improved flow characteristics and a lower working temperature.
Argentel B	57% Cu: Zn: Ni: Mn: Si	890	910	—	

Silver-bearing soft solders

Description	Nominal composition	Melting range (°C)		Remarks
		Solidus	Liquidus	
Plumbsol	2.5% Ag:Sn	221	225	Lead-free alloys ideal for use in plumbing and the fabrication of assemblies where toxicity is a problem.
P35	3.5% Ag:Sn	221	221	Good colour match on stainless steel.
P5	5% Ag:Sn	221	235	P35 and P5 are stronger than tin-lead alloys at elevated temperatures.
Ceramic No. 1	9% Ag:Sn:Pb	183	250	For soldering of silver-metallized components in the electronics industry.
Comsol	1.5% Ag:Sn:Pb	296	296	Good creep resistance at elevated temperatures. Suitable for soldering armature windings and radiators.
A25	2.5% Ag:Pb	304	304	For use in 'hard chlorinated waters where tin-bearing alloys
A5	5% Ag:Pb	304	370	suffer corrosion.
LM10A	10% Ag:Sn:Cu	214	275	A free-flowing alloy with above average tensile strength and creep resistance.
LM15	5% Ag:Cd:Zn	280	320	The strongest alloy in the soft solder range. Used where temperature limitations do not permit silver brazing.
LM5	5% Ag:Cd	338	390	Can be considered for applications where good strength at moderate temperatures is necessary.

Soft solder fluxes

Description	Flux residue removal	Working range (°C)	Remarks
Soft Solder Flux No. 1	Corrosive – can be removed with cold water	150–400	A general-purpose acid flux suitable for copper, brass, mild steel and stainless steel.
Soft Solder Flux No. 2	Non-corrosive – can be removed with cold water	150–400	Recommended where flux residues cannot be removed. Suitable for copper, brass and most plated surfaces.

Appendix 10

Latest plant and equipment

Fig. A10.1. For MIG process. See p. 100 for description of unit.

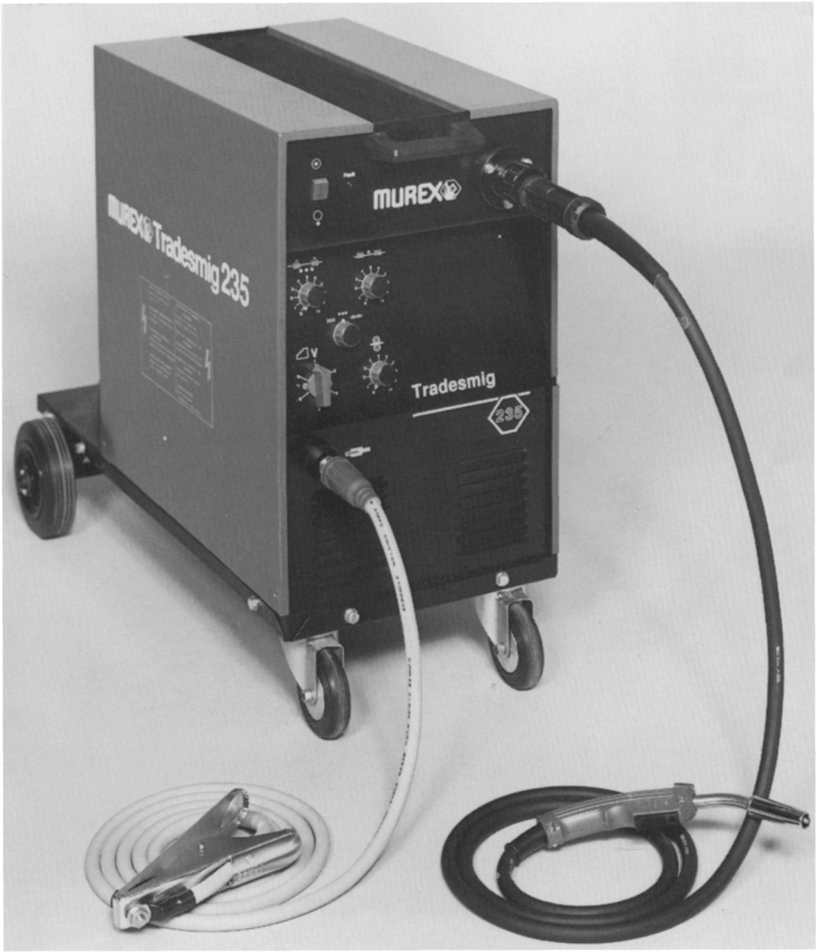
MUREX
MUREX Tradesmig 235
Tradesmig
235

Fig. A10.2. For MIG, MAG (GMAW), MMA (SMAW), d.c. TIG (GTAW) and pulsed MIG. See p. 103 for description of unit.

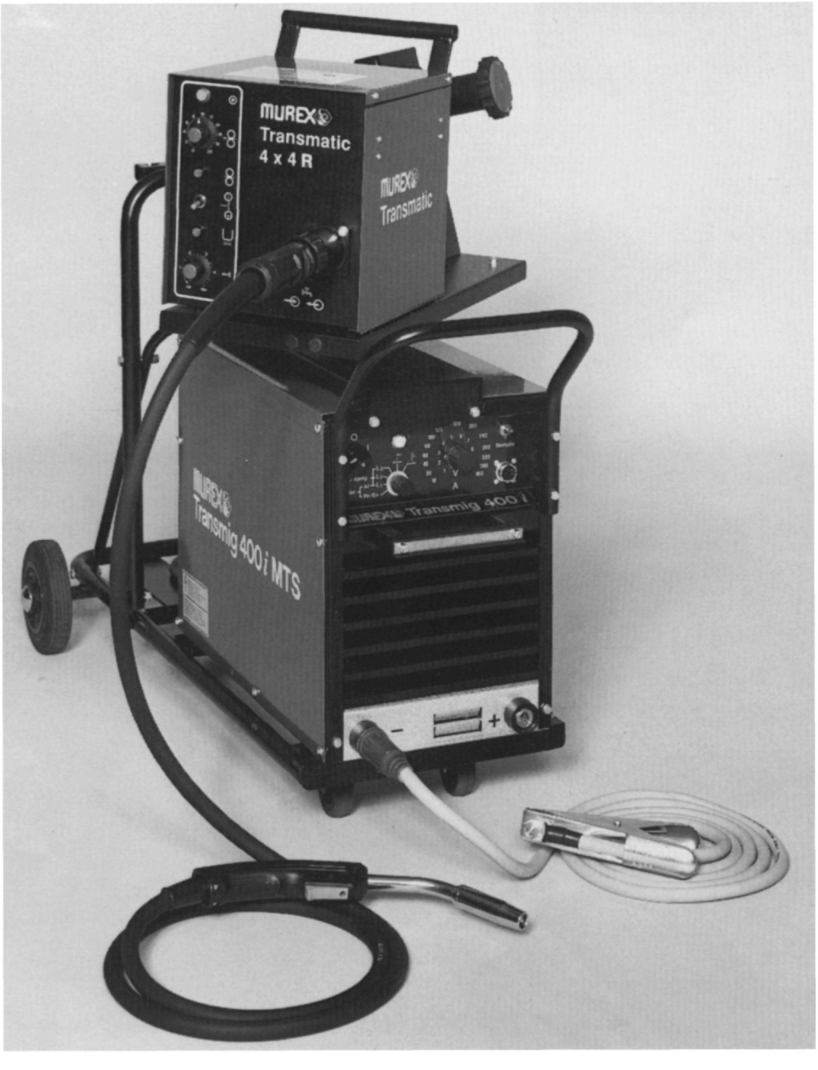
MUREX
Transmatic
4 x 4 R
MUREX
Transmatic
MUREX
Transmig 400 *i* MTS
MUREX Transmig 400 *i*
−
+

Fig. A10.3. Tectron or Maxtron unit with wire feed (termed Tectron in Europe and Maxtron in the USA). See pp. 104–5 for description of unit.

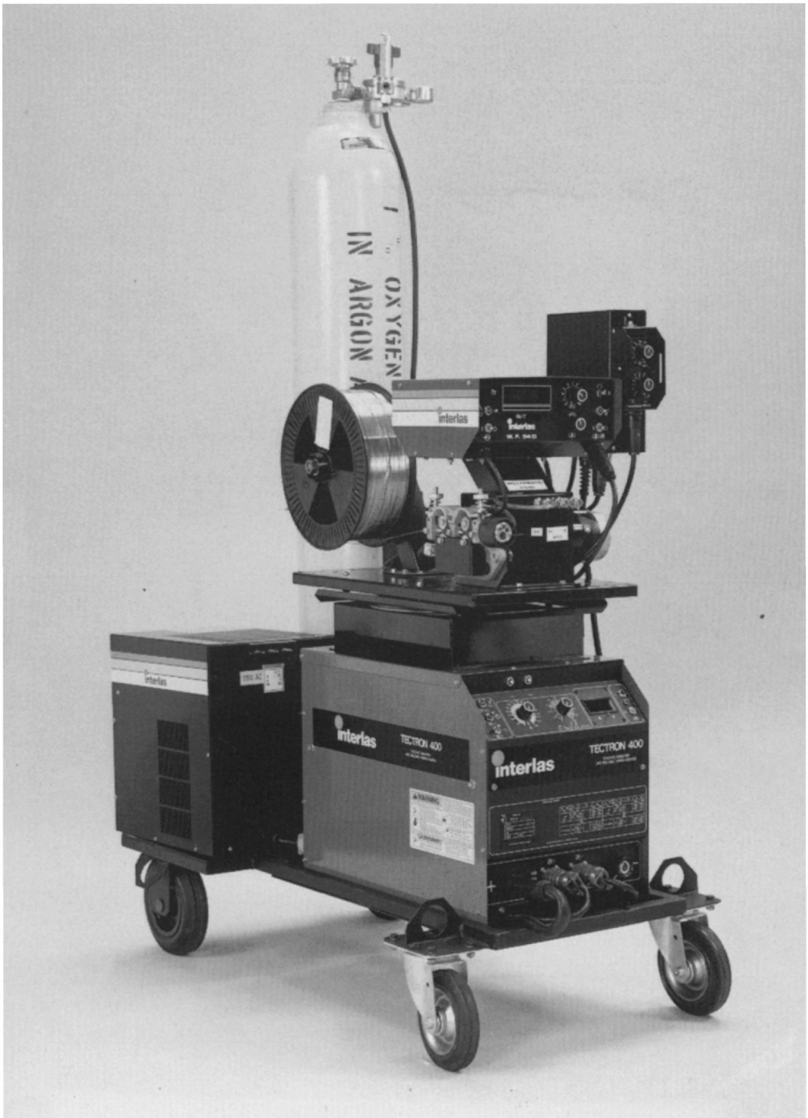
OXYGEN
IN ARGON
interlas
interlas
TECTRON 400
interlas
TECTRON 400

Fig. A10.4. For MMA (SMAW) and MIG (GMAW), dip spray or pulsed modes. See p. 143 for description of unit.

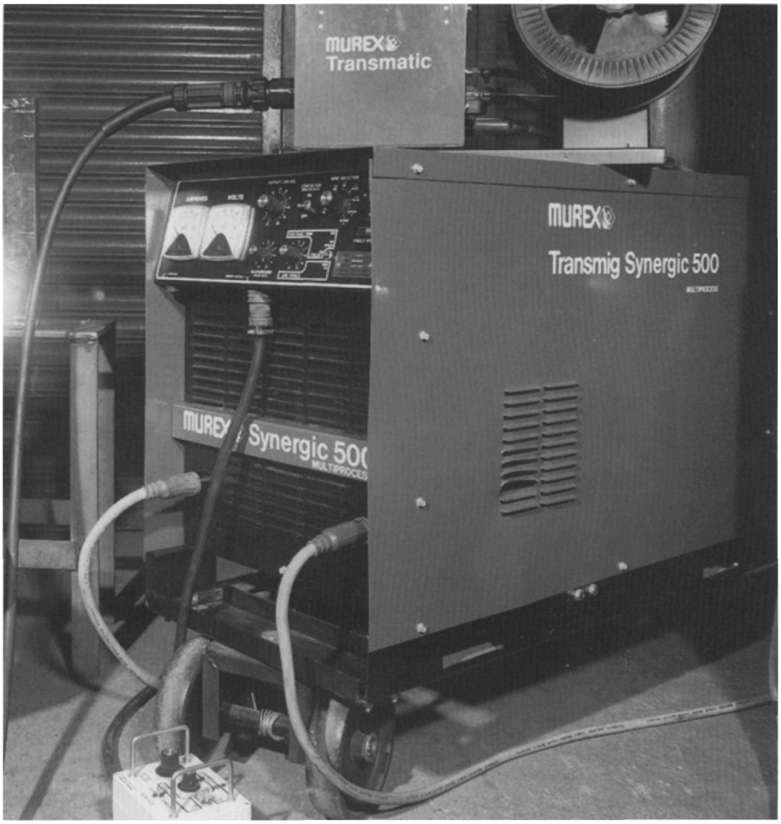
MUREX
Transmatic
MUREX
Transmig Synergic 500
MULTIPROCESS
AMPERES
VOLTS
MUREX Synergic 500
MULTIPROCESS

Fig. A10.5. For synergic pulsed MIG (GMAW). See p. 144 for description of unit.

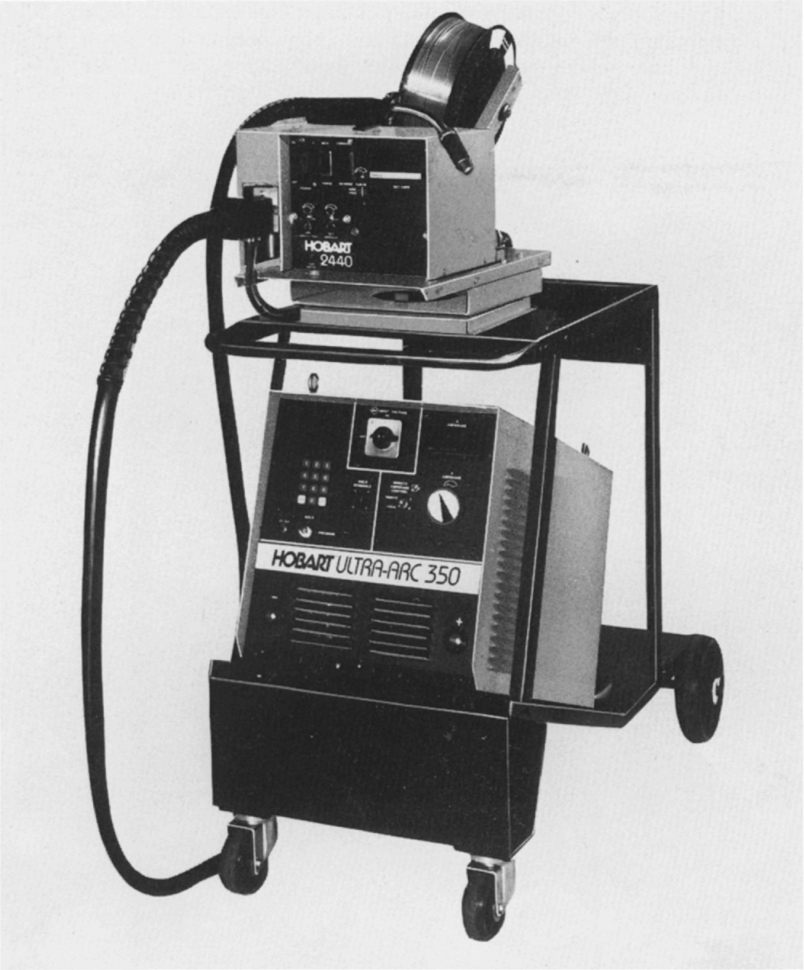
HOBART
2440
HOBART ULTRA-ARC 350

Fig. A10.6. 150 A single phase illustrated (200 A, 300 A and 500 A also available). For the 300 A unit, phase 50–60 Hz 380–415 V inverter unit for d.c. TIG and MMA. Output MMA 5–250 A TIG 5–300 A; arc volts 30, OCV 62. Up and down slope 0–1.5 sec. Arc spot timer. Low and middle pulse frequency with adjustment range 15–85%. Stepless control of current. Smooth shift from welding to crater-fill current. Remote control optional. HF or lift start.

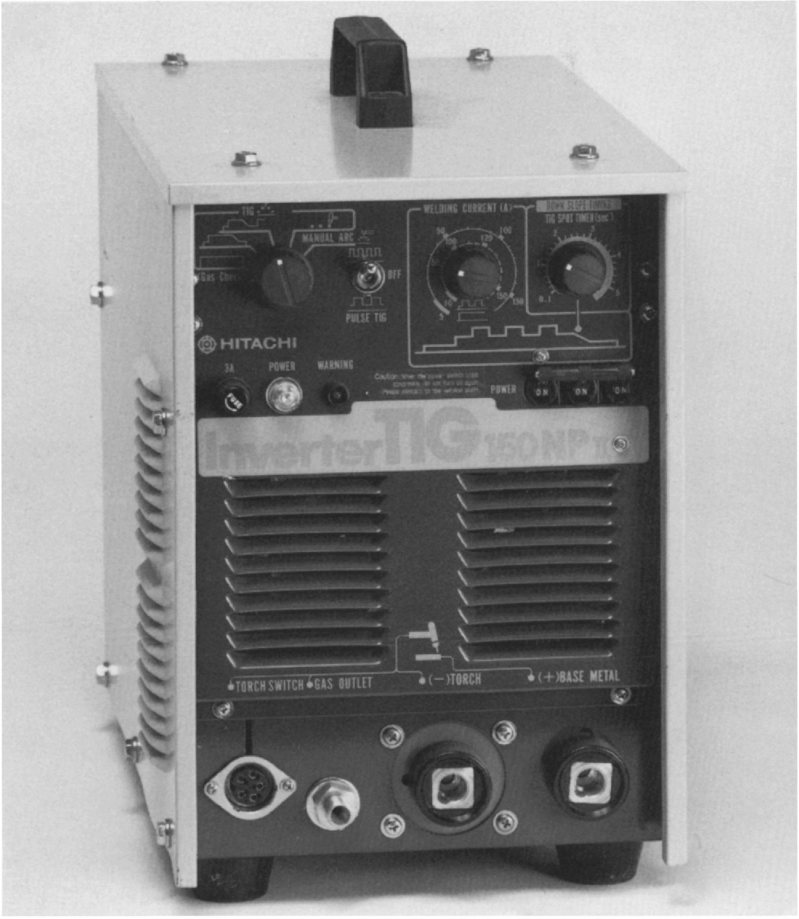
TIG
MANUAL ARC
OFF
PULSE TIG
WELDING CURRENT (A)
TIG SPOT TIMER (sec)
HITACHI
3A
POWER
WARNING
POWER
ON
ON
ON
InverterTIG 150NP
TORCH SWITCH
GAS OUTLET
(−)TORCH
(+)BASE METAL

Fig. A10.7 (*opposite*). (*a*) NBC (Butters) MIG power sources to BS 5970 and ISO 9000 in three outputs, 320 A, 350 A and 450 A. Input 3-phase 415 V, 50 Hz. Ratings 10.8 kVA, 13.5 kVA and 17.1 kVA in each case. OCV of each 13–43. Outputs at 60% duty cycle 300 A, 375 A, 450 A; at 100% duty cycle 250 A, 300 A, 350 A, for each unit. Boom supplied if required. (*b*) Ultra TIG 250 square wave. Electronically chopped d.c. with square wave output; solid state control; gas pre and post flow; slope up and down current; craterfill; current settings 5–25 A and 5–250 A; frequency selection 25–450 Hz; cleaning (balanced) 99–1% or 1–99%; latch selector; pulse module optional; digital meters.

(*a*)

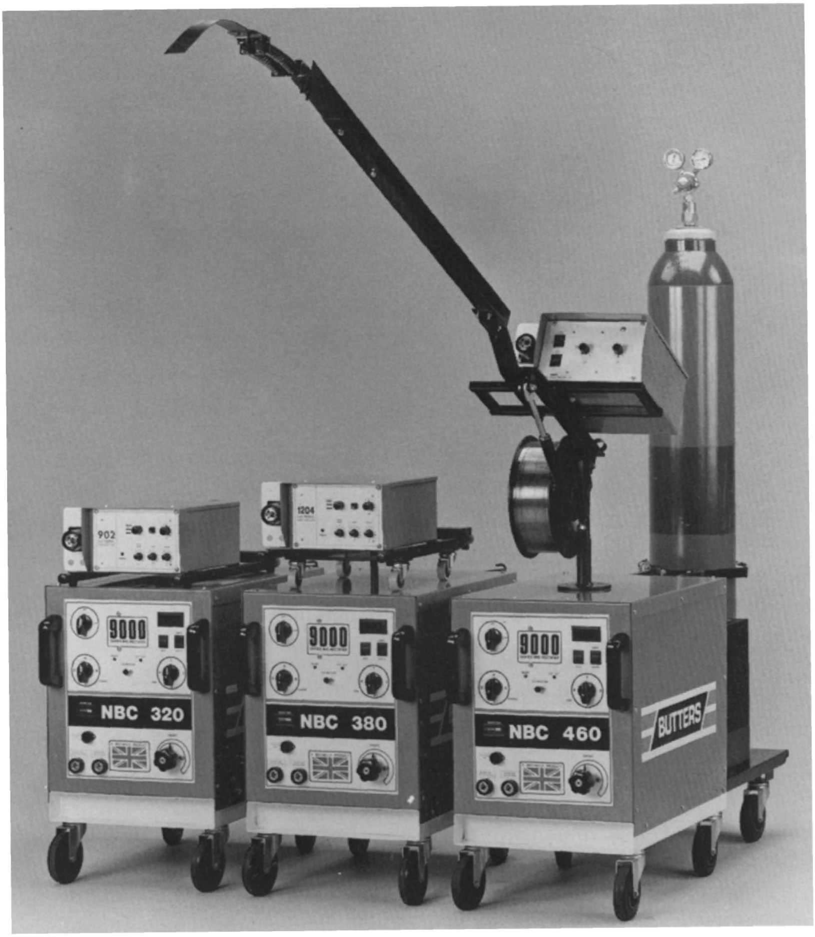
902
1204
9000
9000
9000
NBC 320
NBC 380
NBC 460
BUTTERS

(*b*)

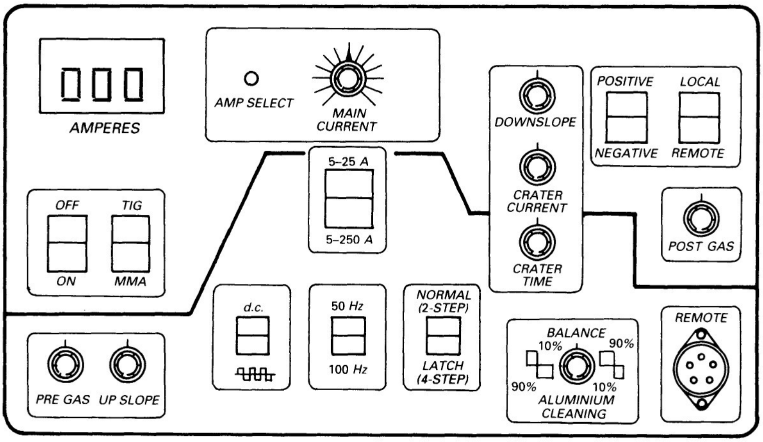
AMPERES
AMP SELECT
MAIN CURRENT
DOWNSLOPE
POSITIVE
NEGATIVE
LOCAL
REMOTE
5–25 A
5–250 A
CRATER CURRENT
CRATER TIME
POST GAS
OFF
ON
TIG
MMA
d.c.
50 Hz
100 Hz
NORMAL (2-STEP)
LATCH (4-STEP)
BALANCE
10%
90%
90%
10%
ALUMINIUM CLEANING
REMOTE
PRE GAS
UP SLOPE

Fig. A10.8. TIGWAVE 250 a.c./d.c. for a.c. adjustable square-wave TIG welding; d.c. TIG, a.c./d.c. MMA welding with arc force control. Flux cored wire welding with CC CV feeders. Input 220, 240, 415, 500 V, the 415 V model at 35 A, frequency 50 Hz. Output 80 OCV single output 5–310 A, 250 A at 30 V and 40 % duty cycle, 200 A at 28 V and 60 % duty cycle. Adjustable square-wave control. Crater fill control. Arc force control. Trigger hold which in 'ON' position enables use of torch button when arc is established. Depressing button second time triggers the crater fill function. Optional spot timer. Pre- and post-flow gas timer. See p. 159 for drawing of this power unit.

HOBART
INPUT POWER
ON
OFF
AMPS
TRIGGER HOLD
PREFLOW
CRATER FILL
POSTFLOW
DO NOT SWITCH WHILE WELDING
NE PAS COMMUTER DURANT LE SOUDAGE
SPOT TIMER
AMPERAGE CONTROL
CRATER FILL
OUTPUT CONTACTOR
METER
TIGWAVE 250 AC/DC
HOBART

Fig. A10.9. For a.c. and d.c. TIG (GTAW) square wave. See p. 162 for description of unit.

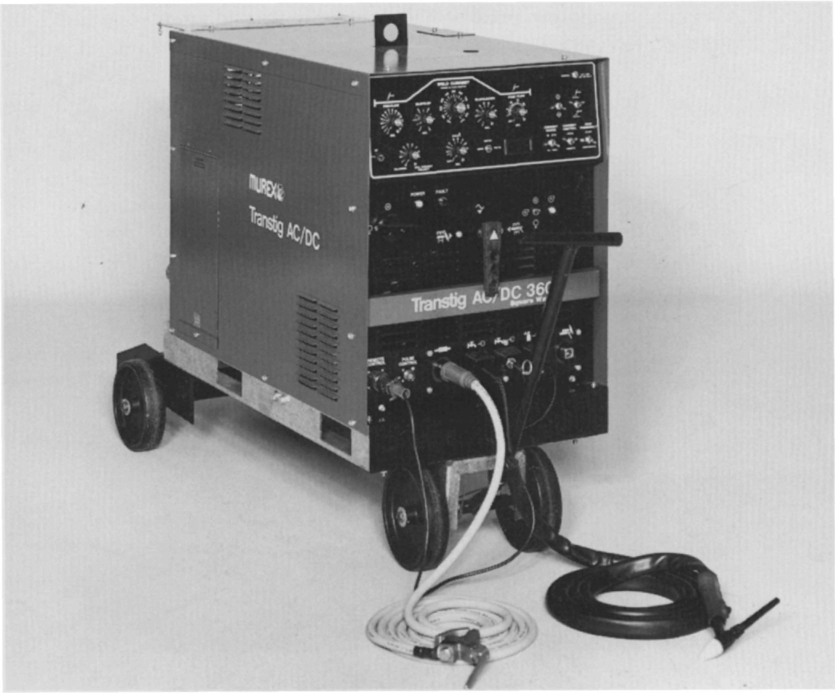
MUREX
Transtig AC/DC
Transtig AC/DC 36

Fig. A10.10. Cyber wave 300 A, a.c. or d.c. rated output 300 A, OCV 80. Input 3-phase, voltages 230/460/575. For a.c. square wave TIG, square wave a.c. and d.c. plasma, a.c. and d.c. MMA (stick). For aluminium, magnesium, steel and stainless steel. Auxiliary power 1 kVA. Arc force control. Balance control d.c.e.n. 60–80 % for maximum penetration; d.c.e.p. maximum cleaning. Single current range. Solid state program module, calibrated for use with various units, e.g. pulse and arc spot timer, up and down slope, gas prepurge, HF start.

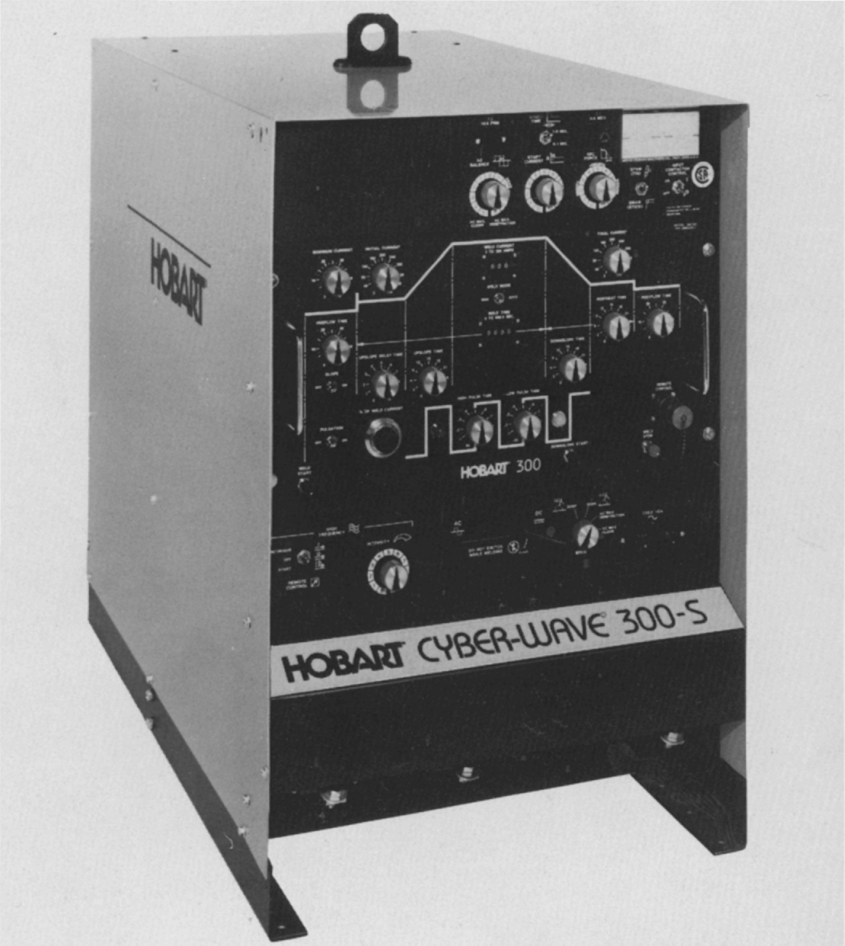

Fig. A10.11. PowCon 300 A power source. Inverter, single and 3-phase; suitable for MIG/MAG (GTAW), dip spray and flux cored (FCAW) TIG (GTAW) d.c., MMA (SMAW) and CO_2 processes. Smooth low ripple arc. For 3-phase model, input 380–575 V, with OCV 80. At 100% duty cycle 150 A at 30 V, at 60% duty cycle 300 A at 32 V. Optional extras: pulsed MIG, pulsed TIG. Hand or foot remote control. Generator can be engine-driven if required. Weight 45–48 kg, depending on model.

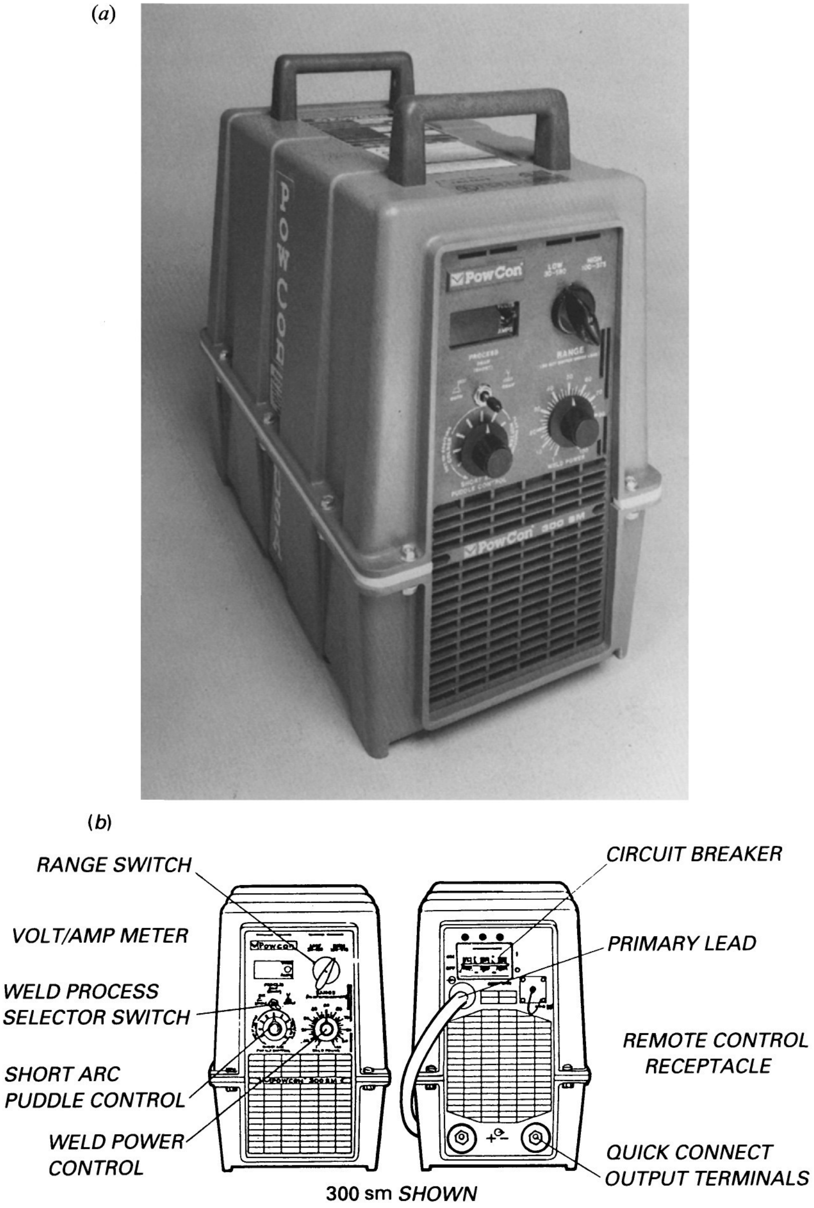
(a)
PowCon
PowCon 300 SM
(b)
RANGE SWITCH
VOLT/AMP METER
WELD PROCESS SELECTOR SWITCH
SHORT ARC PUDDLE CONTROL
WELD POWER CONTROL
CIRCUIT BREAKER
PRIMARY LEAD
REMOTE CONTROL RECEPTACLE
QUICK CONNECT OUTPUT TERMINALS
300 sm SHOWN

Fig. A10.12. Water-cooled diesel unit. Automatic idle, a.c. and d.c. constant current (CC) output for MMA (stick) and d.c. constant voltage (CV) output for gas, metal and flux-cored arc welding; a.c. TIG by addition of HF unit; 115 and 230 V single-phase auxiliary power for tools etc.

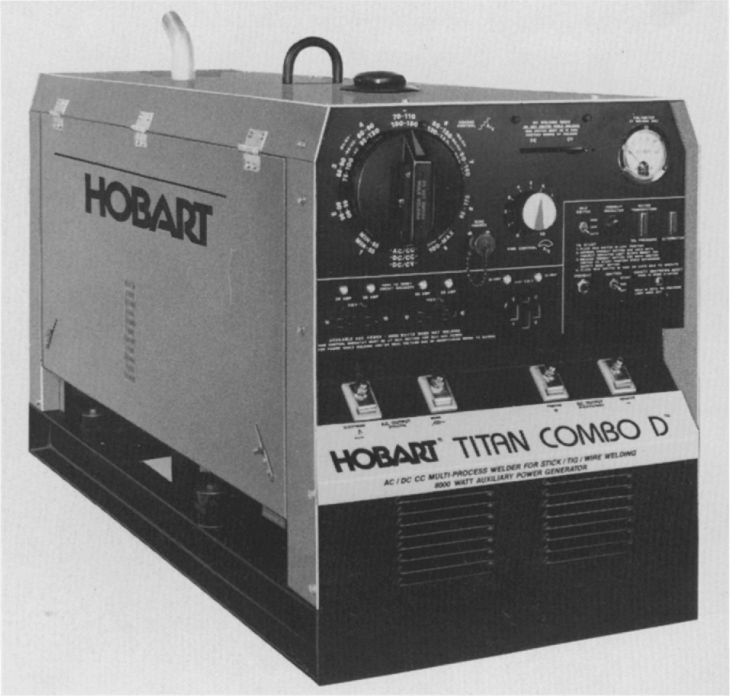
HOBART
HOBART TITAN COMBO D
AC / DC CC MULTI-PROCESS WELDER FOR STICK / TIG / WIRE WELDING
8000 WATT AUXILIARY POWER GENERATOR

Fig. A10.13. Wire feeder, CC/CV with selector switch. The CV operation controls the wire-feed speed. The CC operation controls the arc voltage. Increase in arc volts means a decrease in wire-feed speed (wfs), one driven roll and one idler roll. Speed of wire 0.127–15.25 m/min. Spool weight 13.6 kg. Wire: tubular 2.0 mm or smaller, solid 1.3 mm or smaller. Weight of unit 11.8 kg. Motor 20 V d.c. permanent magnet motor. Quick connect gun connection. Optional remote control. Suitable for shipyards, pipe-lines, manufacturing, offshore, power plants etc.

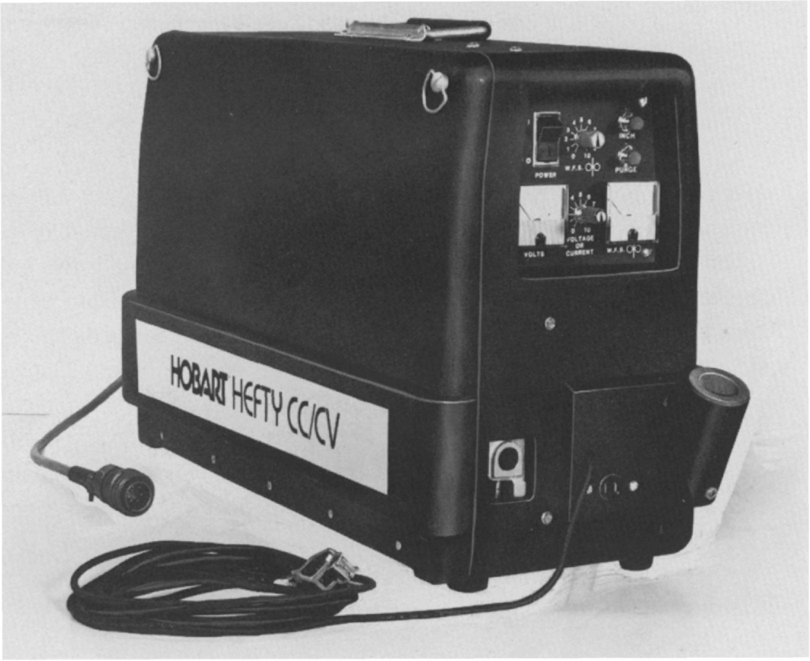
HOBART HEFTY CC/CV
POWER
W.F.S.
INCH
PURGE
VOLTS
VOLTAGE OR CURRENT
W.F.S.

Appendix 11

Refraction and reflection

If a ray of light is incident obliquely at a point O on a glass slab and the normal N is drawn at the point O, the light ray is refracted or bent in towards the normal. If i is the angle of incidence and r the angle of refraction to the normal it can be seen that if a ray passes from a rare (air) to a denser medium (glass or water) it is refracted towards the normal (Fig. A11.1*a*). We then have $\sin i/\sin r = \eta$ (refractive index) The value of η is constant for a given pair of media (Snell's law).

Upon emerging from the denser to the rarer medium the ray is refracted away from the normal by the same amount but is displaced by an amount which depends upon the distance travelled through the denser medium (Fig. A11.1*b*).

A stick partly immersed obliquely in water appears to be bent due to the refraction of the light from the immersed end of the stick.

Fig. A11.1 (*a*)

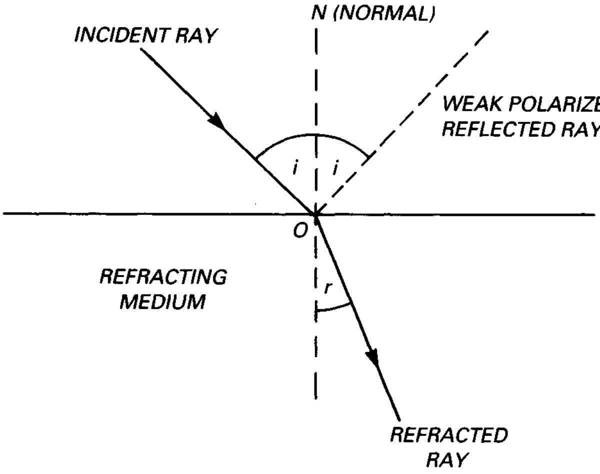

Fig. A11.1 (*b*)

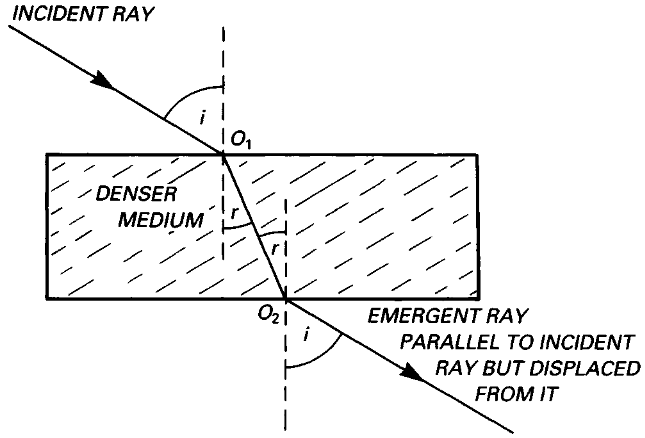

Polarization

We have seen that a ray of light is an electromagnetic radiation having two component fields, one electric and the other magnetic, vibrating or oscillating at random, at right angles to each other and at right angles to the direction of travel of the ray.

If the ray is passed through certain substances that allow the electric forces in one plane only to pass through the substance, the emergent light is said to be plane polarized with the electric forces vibrating in the plane of polarization.

Polaroid is the trade name given to a substance of paper-like consistency impregnated with chemicals, which gives an emergent polarized ray when a light ray is passed through it. It is cheaper than the Nicol prism (calcite or Iceland spar) and is supplied in sheets of various sizes. The emergent polarized ray has the electric forces vibrating in the plane of polarization of the sheet.

If a second sheet is placed at a little distance from, and parallel to, the first sheet, with its plane of polarization in the same direction, the ray of polarized light emergent from the second sheet suffers virtually no loss. Now rotate the second sheet about the axis of the ray and it will be seen that the polarized ray emergent from the second sheet becomes darker and darker until darkness occurs when the planes of polarization of the two sheets are at right angles to each other. There is now no emergent ray, the polarized ray from the first sheet being completely quenched by the second sheet. The first sheet is termed the polarizer and the second sheet the analyser.

If a solution of a salt is placed in a tube between the sheets, the analyser has now to be rotated through an angle greater than a right angle to the first sheet to obtain darkness. The solution has rotated the plane of polarization of the ray by an amount depending upon its strength, so that by comparing the increased rotation per unit length of various solutions their strengths may be compared.

The eye cannot differentiate between unpolarized and polarized light. When an unpolarized ray of light is incident at an oblique angle on a surface such as glass or water, part is reflected at the surface and the remainder is refracted into the medium (Fig. A11.2). The reflected component is part plane polarized and this is a maximum when the angle between the reflected polarized ray and the refracted ray is a right angle. This leads to Brewster's law, which states that the angle of polarization at the surface of a medium is the angle whose tangent is equal to the refractive index of the medium.

Fig. A11.2

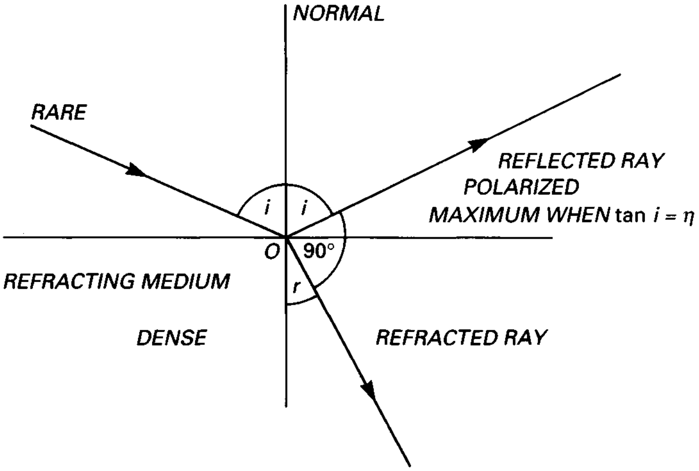

Appendix 12

*Notes on the welding of Duplex and Super Duplex stainless steels**

Duplex and Super Duplex stainless steels are particularly useful in the petrochemical industry because of their pitting and corrosion resistance. Duplex stainless steels are so called because they have two phases in their microstructure, ferrite and austenite, so they are ferritic–austenitic steels. One of the most common of the Duplex steels has 22 % Cr, 5 % Ni, 3 % Mo and 1 % N_2, remainder ferrite. The ferrite gives the high tensile strength and the austenite gives the toughness. The latest of these steels is called Super Duplex stainless steel and has a greater resistance to corrosion and a high pitting resistance. A typical example of a Super Duplex steel has 25 % Cr, 7 % Ni, 4 % Mo and 0.15 % N_2, remainder ferrite. Duplex steels are not sensitive to corrosion cracking but must not be worked above 300 °C, as above this embrittlement occurs. The ferritic–austenitic balance is important as a high ferrite content will cause brittleness and a low one will cause a lowering of resistance to stress corrosion. Contents of 30–70 FN (ferrite number) are normal. The balance between ferrite and austenite depends upon the heat treatment and the composition of the filler metal, the heat input when welding and the interpass temperature. A filler wire should always be used and the result will be a strong and sound weld and HAZ. Welding without any filler rod and excessive dilution of the weld and parent metal produce an excess of ferrite causing the weld metal to be brittle. The interpass temperature must *not* exceed 150 °C (Super Duplex steels are particularly sensitive to high heat input and the temperature of the interpass runs).

To summarize: high heat input and high interpass temperature give a low cooling rate resulting in brittle phases forming during cooling, whereas low heat input and low interpass temperature during welding give a high cooling rate, resulting in more ferrite and lower toughness resistance. The gases used for the various processes are: argon or argon–helium mixtures for TIG; argon–oxygen, argon–carbon dioxide or argon–helium–oxygen mixtures for MIG; and argon–carbon dioxide or pure carbon dioxide for flux cored arc welding (FCAW).

---

* Details of the article and table of consumables by permission of ESAB (UK) Ltd.

When welding Duplex steels it is essential to use root gas shielding and this should be argon, or argon with a small amount of hydrogen or nitrogen, and the mixture should be checked to ensure its efficiency. The addition of nitrogen to the backing or shielding gas improves the corrosion resistance while chromium, molybdenum and nitrogen determine the resistance to pitting. The pitting resistance equivalent $(\text{PRE}) = \%\text{Cr} + 3.3\,\%\ \text{Mo} + 16\,\%\ \text{N}_2$. A PRE of 40 is often used as a definition of a Super Duplex stainless steel. For the estimation of the critical pitting temperature (CPT) ferric chloride is used with a minimum CPT of 40 °C for Super Duplex, and this may be as high as 60 °C for weld metal that has to stand immersion in sea water (Super Duplex OK 68.53 is a good example). Welds made in Super Duplex steels containing small percentages of copper and tungsten often give CPTs of 45–50 °C in this test. This is a good test for sea water corrosion and pitting. The amount of ferrite in the final weld can be estimated by constructing a constitutional diagram using the chromium equivalent along the horizontal axis and the nickel equivalent along the vertical axis. The varying dilutions are indicated by lines on the diagram representing all the compositions of the weld metal that can occur. The chosen dilution is made and other lines drawn to indicate the possible ferrite content of the weld (see, for example, p. 69).

When welding Duplex steels the parts should be thoroughly cleaned and brushed with a hand-held brush with stainless steel wires. Use dry electrodes and if necessary warm them before use. Do *not* preheat. The heat input of the run will vary with the plate thickness but must not be too high or too low. Do not strike the arc anywhere outside the joint as this may lead to the initiation of corrosion and cracking. Avoid excessive weaving as this may over-heat the weld. See that the backing gas of pure argon with small amounts of nitrogen or hydrogen is functioning. After welding, clean off all slag and any other foreign matter with a hand-held stainless steel brush. Do not use rotary brushes as they make almost invisible scratches which can lead to corrosion. No post-heating is required, but the part may be solution-treated (see Vol. 1, p. 130, for description of solution treatment). Do not attempt any stress relief. In other words, the part or welded structure should *not* have any heat treatment.

The arc energy can be calculated in the following way:

$$\text{arc energy (kJ/mm)} = \frac{\text{arc voltage} \times \text{welding current}}{\text{welding speed (mm/s)} \times 1000} \quad \text{(see p. 80).}$$

Consumables for welding Duplex and Super Duplex stainless steel

Composition of base metal (%)	FN	MMA	MIG	TIG	FCAW	SAW
23 Cr, 4 Ni, 0.1 N	30–45	67 50 Rutile 67 53 Rutile	Autrod 16 86	Tigrod 16 86	Tubrod 14 37	Flux 10 93 Autrod 16 86
22 Cr, 5.5 Ni, 3 Mo, 0.15 N	30–45	67 50 Rutile 67 53 Rutile	Autrod 16 86	Tigrod 16 86	Tubrod 14 37	Flux 10 93 Autrod 16 86
25 Cr, 7 Ni, 0.14 N	30–50	68 53 Rutile 68 55 Rutile		Tigrod 16 88		Flux 10 94 Autrod 16 88
25 Cr, 6.5 Ni, 3 Mo, 0.18 N	30–50	68 53 Basic 68 55 Basic		Tigrod 16 88		Flux 10 94 Autrod 16 88
25 Cr, 6 Ni, 3 Mo, 2 Cu, 0.20 N	30–50	68 53 Basic 68 55 Basic		Tigrod 16 88		Flux 10 94 Autrod 16 88
25 Cr, 6 Ni, 3 Mo, 0.7 Cu, 0.7 W, 0.25 N	30–50	68 53 Basic 68 55 Basic		Tigrod 16 88		Flux 10 94 Autrod 16 88
25 Cr, 6 Ni, 3 Mo, 0.7 Cu, 0.7 W, 0.25 N	30–50	68 53 Basic		Tigrod 16 88		Flux 10 94 Autrod 16 88

Note: All electrodes, wires and fluxes are prefixed with the letters OK (e.g. OK 67 50); FN is the ferrite number.

As an example, suppose that for Duplex steels the arc energy is within the range 0.5–2.5 kJ/mm, and for Super Duplex stainless steels the arc energy is within the range 0.2–1.5 kJ/mm. Suppose also that a Super Duplex steel has an arc voltage drop of 20, a welding current of 80 A and a welding speed of 1.5 mm/s. Then from the above formula:

$$\text{arc energy} = \frac{20 \times 80}{1.5 \times 1000} \approx 1.1 \text{ kJ/mm}.$$

This setting would be suitable for both steels as an arc energy of 1.1 kJ/mm falls within the above limits for both steels.

Suppose, now, that for a Duplex steel the arc voltage drop is 20, with a welding current 100 A and a welding speed of 1.0 mm/s, then

$$\text{arc energy} = \frac{20 \times 100}{1.0 \times 1000} \approx 2.0\ \text{kJ/mm}.$$

This setting of current and weld speed is suitable for Duplex steels only, as the arc energy falls outside the figure given for Super Duplex stainless steels.

Overalloying

Note that in all consumables used for welding Duplex and Super Duplex stainless steels there is more nickel than normal. This 'overalloying' is to help the formation of austenite so that there is not too much ferrite in the final weld metal. The use of a filler material and only a small weave, in addition to the excess nickel, prevents excessive dilution, which would increase the ferrite content and hence the brittleness of the weld.

Appendix 13

Welding fumes: health and safety

The electric arc, used in MMA, MIG–MAG, TIG, and other welding processes, causes visible and invisible fumes to be given off by the melting flux when the arc is burning. These fumes can be harmful if inhaled through the nose and mouth of the welder. The content of the fumes may be either asphyxiating or toxic.

Asphyxiating fumes

Asphyxiating fumes are largely due to the shielding gases used around the arc. They are invisible and produce their effect by displacement of the surrounding air and thus, of its oxygen. (The air which we breathe consists mainly of nitrogen and oxygen, roughly in the ratio of four parts nitrogen to one part oxygen.) Asphyxiation begins when the oxygen content of the air falls below about 18 %.

The gases usually used in welding are argon, carbon dioxide, helium and their mixtures. Argon and carbon dioxide are heavier than air (the mixtures are known under trade names). A mixture is selected to give the best welding conditions for a particular process. A simple mixture is the rutile coating for mild steel electrodes in the MMA welding of mild steel. The rutile gives a smooth arc, which is easily managed, but it gives off some hydrogen, which is absorbed by the molten metal and leads to hydrogen cracking under certain conditions of stress. To prevent this the covering is of calcium carbonate ($CaCO_3$), mixed with a little rutile in order to give a smooth and easily managed arc. This mixture now gives off only a little hydrogen (a low hydrogen electrode), which can be used where higher stresses are encountered. The proportion of asphyxiating gas present is generally given in parts per million (p.p.m.) (this can be converted into milligrams per cubic metre of the atmosphere (mg/m^3)).

The mixtures with argon are, like the argon, generally heavier than air

Asphyxiating gases

Asphyxiant	Notes
Argon (Ar)	Non-toxic. Used as a shielding gas, alone or mixed with other gases. Heavier than air and can accumulate at the base of any closed vessel being welded and can form layers at the bottom of the welding operation in a badly ventilated welding shop. Essential to have extraction and circulation of the atmosphere around the welding point.
Helium (He)	Non-toxic. Used generally mixed with other gases (e.g. 50% helium 50% argon). Helium is not produced in this country but is imported from the United States and is therefore much more expensive than argon.
Oxygen (O_2)	Non-toxic but it promotes rapid oxidation especially in the pure state. Atmosphere roughly 4 parts nitrogen to 1 part oxygen (the proportions required by the human body to enable it to function). Used in small quantities (1% to 2%) mixed with argon for stainless steel welding.
Nitrogen (N_2)	Although nitrogen is used for displacing the atmosphere in tanks and pipelines it is not a neutral gas like argon because it forms compounds with other substances. Two of these are produced in welding, namely nitric oxide (NO) and nitrogen dioxide (NO_2), both of which are toxic (q.v.) as nitrogen itself is an asphyxiant. Nitrogen is produced in large quantities as a by-product in oxygen-making plants.

and invisible. They tend to sink to the floor, making an asphyxiating layer which becomes deeper as welding continues. Evaporation and vaporization of the metal from the molten pool rise with the other fumes but are invisible.

*Toxic fumes**

Toxic fumes are made up of very many tiny particles of the flux covering of the electrode, the central core of the flux (if used), and the vapours from the metallic evaporation of the molten metal of the arc; the fumes rise to give the familiar welding fumes. Note that the size of the particles is extremely small. The student should read publications EH40, EH54 and EH55†, together with the COSHH regulations to get a complete

---

* *Note.* Spatter, which appears as sparks (which are globules of molten metal ejected from the arc), is not classified as fumes.

† All booklets mentioned are published by HMSO and can be obtained from any branch office, or by post from HMSO Books, PO Box 276, London SW8 5DT. They are written by the Health and Safety Executive (HSE) and are available in most libraries, colleges and universities. The full titles are:

HSE Guidance Notes EH40: Occupational Exposure Limits
Guidance Notes EH54: Assessment of Exposure to Fume from Welding and Allied Processes
Guidance Notes EH55: Control, Exposure, Fumes, Welding, Brazing and Similar Processes

picture of this important subject. Booklet EH40 xx (xx indicates the year of printing) gives, in detail, an explanation of the exposure limits. Because the limits are updated each year the values have been omitted from the following table. The student is advised to fill them in using pencil and to keep them up to date. Note that the total fumes inhaled should not exceed 5 mg/m^3.

The following terms are used in EH40:

OES	Occupational exposure standard (given in mg/m^3 or p.p.m. over a given time period).
MEL	Maximum exposure limit (given in mg/m^3 or p.p.m.).
TWA	Time-weighted average. Allowance is made for the periods when the arc is not burning.
ID	Inhalable dust (total dust inhaled).
RD	Respirable dust (quantity of dust that enters the respiratory tract and lungs).
STEL	Short-term exposure limit.

The removal of asphyxiating and toxic welding fumes

If welding is performed outside, there is a constantly changing atmosphere around the welding area, so that the extraction of fumes may not be necessary, but careful assessment should be made to confirm this. If a wind is blowing, however, screens should be set up to prevent the gas shield around the arc being disturbed. If necessary the electrode wire should be flux cored to lessen the effect.

The welding shop should be as large, airy and high as possible. Sometimes there may be long narrow ducts at ground level through which clean air is blown, and similar ducts at ceiling height to enable the polluted air to be exhausted. Even with heated clean air being blown in through the bottom ducts the shop will be very cold in winter time and the heating of the air will add to the cost of welding.

A popular method of eliminating fumes, in use today, consists of a suction unit fitted with a filter, from the suction side of which a large-diameter, fixed, rigid tube is fitted. This tube passes down the shop (well above head height) and over every welding position or bench. Flexible tubes of a similar large diameter are fitted, which reach down to the welding position or bench. At the lower end of the flexible tube a collecting head is fitted (see Figs. A13.1 and A13.2). The tube is sufficiently rigid to remain in any position in which it is placed and the collecting head may be fitted with a magnet to help positioning of the collector. For long

Table of substances in fluxes which produce toxic fumes

Substance	Notes	OES mg/m^3	TWA 8 h STEL	MEL mg/m^3	TWA 8 h STEL
Aluminium and its alloys, aluminium oxides Al_2O_3, $Al(OH)_3$, Al(OOH))	Fumes are produced when welding aluminium and its alloys. Inhaling the fumes over long periods may affect the respiratory tract.				
Cadmium and its compounds, cadmium oxide fume	These are highly toxic and are generally formed when welding articles that have been cadmium plated. Clean off in the areas to be welded.				
Calcium, calcium oxide, calcium hydroxide, calcium carbonate, calcium silicate	Though non-toxic, fumes may irritate the respiratory tract when in large concentrations.				
Carbon dioxide	Present mixed with many types of shielding gases. It is mildly toxic and, being slightly heavier than air, sinks to the base of closed containers when welding. With the other gases it acts as an asphyxiant.				
Carbon monoxide	Very toxic. Small amounts are formed when CO_2 is present as a shielding gas.				
Chromium and its compounds, chromium II, chromium III (trivalent), chromium VI (hexavalent)	Fumes are produced when welding stainless steel and other alloys with a high chromium content. These are all toxic and those of chromium III and VI may be carcinogenic.				
Copper and its compounds, fumes, dust and mists	Toxic fumes are produced when welding copper. Most welding wires for welding steel by the MIG process are copper coated to prevent rusting, but produce very few copper fumes.				
Iron and its compounds, iron oxide fumes Fe_2O_3, iron salts	Only slightly toxic and found when welding steels. Welding may cause deposits upon the lungs.				

Table continued

Substance	Notes	OES mg/m^3	TWA 8 h STEL	MEL mg/m^3	TWA 8 h STEL
Lead and its compounds, lead in the atmosphere (lead in the air is taken as standard)	Very toxic. Large amounts of fumes are produced when welding heavily lead-painted articles (clean off paint where required). Fumes are given off when lead is welded. These are very toxic.				
Magnesium, magnesium oxide (MgO)	Formed when welding high magnesium content aluminium alloys, which work-harden.				
Manganese and its compounds	Fumes are toxic. Manganese is used as an alloying element in steel making, 12–14% manganese steel work-hardens and is work resistant. Fumes affect the nervous system and respiratory tract.				
Barium (soluble) and its compounds	Highly toxic and found in some flux cored wires.				
Nickel and its compounds, soluble compounds	Present as an alloy in many steels and in stainless steel as a major alloying element. Fumes are very toxic. Affects the respiratory tract and produces metal fume fever. May be carcinogenic.				
Silicon, silica (SiO_2)	Non-toxic. Contained in some fluxes in core or covering. Little effect on welding fumes as only small amounts are formed.				
Zinc and its compounds, zinc oxide (ZnO), zinc chloride (ZnCl)	Toxic. Fumes of zinc oxide are given off copiously when welding galvanized articles, the coating of which is zinc. Grind off the coating from areas to be welded.				

Fig. A13.1. Practically all fumes are collected at this distance.

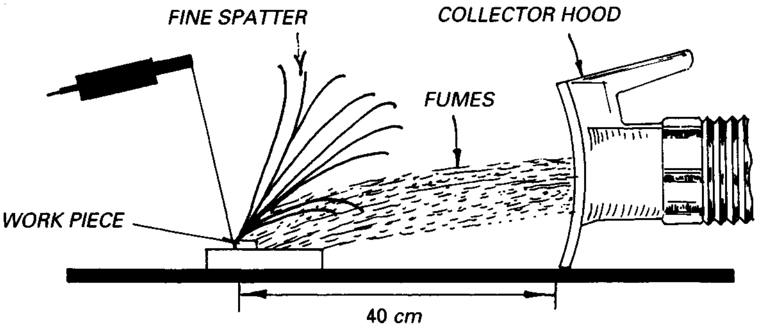
FINE SPATTER
COLLECTOR HOOD
FUMES
WORK PIECE
40 cm

Fig. A13.2. Twin fume extractors with fan, light and control box.

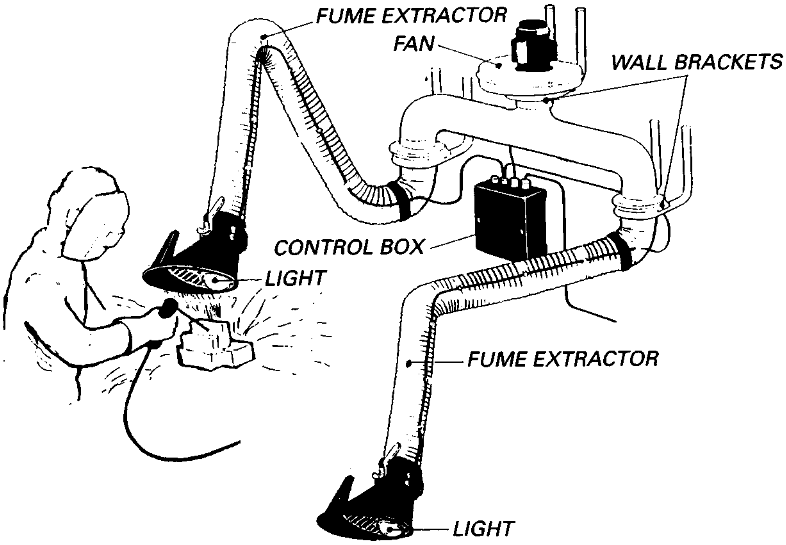
FUME EXTRACTOR
FAN
WALL BRACKETS
CONTROL BOX
LIGHT
FUME EXTRACTOR
LIGHT

extensions from the overhead tube, the flexible tube may be fitted with several jointed or articulated arms passing inside the tube. These arms are a great help in positioning the collector in cases on the shop floor when the cantilever effect is maximum. The collecting head may be fitted with a shutter, hand-operated from outside the head, so that an operator can switch off the suction without affecting other welders on the same circuit. In addition, there may be a protected electric light to assist the operator in positioning the work. For larger articles and fabrications a portable unit can be used.

The filter, starting unit and electric motor are all mounted on wheels and can easily be moved around. A large-bore flexible tube is fitted to the suction side of the unit and the collector is positioned by the welder, the collector being above weld level and as close as possible to the weld without impairing the work of the welder. The extracted air is filtered to remove the fumes, and the air is then returned to the atmosphere.

For this kind of work the suction rail method is useful. This consists of a fixed large-diameter tube fitted with rails. A collector on the rails and running on wheels is pulled to cover any desired point in the fabrication, and the tube on the collector delivers the air and fumes directly into the suction tube. The clean air exhausts to the atmosphere no matter where the trolley is placed.

In some cases, when MIG welding, the gun may incorporate fume-collecting ducts. In the body of such a gun, towards the lower end, are some holes or ducts (see Fig. A.13.3). The fumes are sucked in through the holes,

Fig. A13.3. MIG gun with built-in fume collector (Murex).

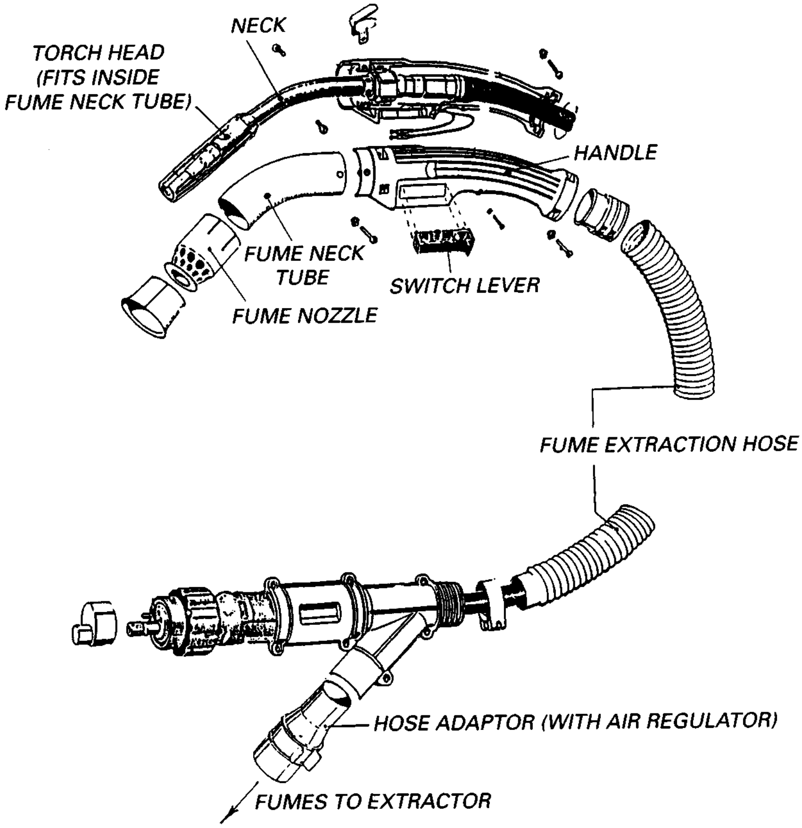
NECK
TORCH HEAD (FITS INSIDE FUME NECK TUBE)
HANDLE
FUME NECK TUBE
SWITCH LEVER
FUME NOZZLE
FUME EXTRACTION HOSE
HOSE ADAPTOR (WITH AIR REGULATOR)
FUMES TO EXTRACTOR

up the bore of the gun and are quite separate from the shielding gas and water-cooling (if any) supply. Higher up the gun is an angled unit attached to the suction tube of the fume-cleaning unit. The fumes are collected as they are made, and collection does not interfere with the shielding gas supply. The fumes pass up the gun and are filtered and exhausted as described earlier. Note that these guns are available only for MIG welding.

Electrostatic precipitation can be utilized where it is inconvenient to install a fixed collector. Upon entering the precipitator, fume particles are charged electrostatically and then pass through an insulated sleeved tube to plates of opposite polarity to that of the particles. They are deposited on the plates, which are part of the filtering elements, and the air issuing from the unit is fumeless and is returned to the atmosphere. Fig. A13.4 shows a Sectional view of a filter.

Efficient extraction of fumes is essential to the long-term health of the welder, as it may be many years until the effects of fumes inhaled over a long period are apparent. Almost all the toxic substances in the table above have a long-term effect on the respiratory tract and lungs, and in time seriously affect the health of the operator. Therefore, take all precautions possible:

> Never weld with head and helmet directly in the path of rising fumes.

Fig. A13.4. Mobile unit with filter. Arrows show passage of fumes through the replaceable paper filter. In the electronic model an additional motor cleans the filter elements and indicates when the main container needs emptying.

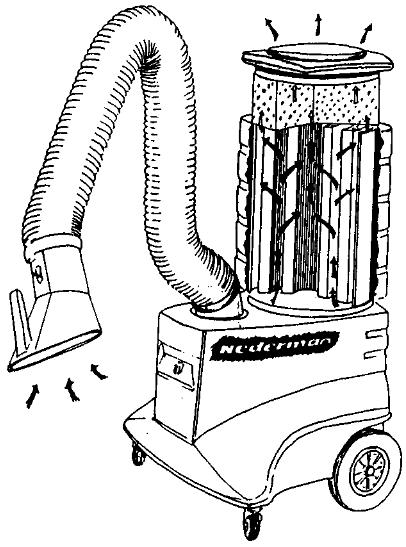

Never use a flat shield.
Use a head shield that protects the face, head-sides and neck.
Use a filter glass of the deepest shade compatible with the weld being performed.
Wear gloves to protect hands and wrists from the harmful effects of UV rays.
Never weld with the sleeves rolled up.
Protect the body skin by covering it.

However carefully employers safeguard the health of their welding operators in the welding shop or fabricating area within the COSHH regulations and the HSE recommendations, it is finally up to the welder to safeguard his or her health, by following these regulations regarding toxic fumes and asphyxiating gases. Figs. A13.1–A13.4 should make some of the apparatus for toxic fume removal quite clear.

Free information booklets on the hazards to health from welding fumes are provided by the makers of the electrodes and gas suppliers, for example:

Air Products Plc, 4th Avenue, Crewe, Cheshire
BOC Ltd, Guildford, Surrey
Eland Group Butters Ltd, Hartlebury, Worcestershire
Murex Ltd and ESAB Ltd, Waltham Cross, Hertfordshire
Nederman Ltd, Bamber Bridge, Preston, Lancashire
Oerlikon Welding Ltd, Hayes, Middlesex
Welding Manufacturers Association, Leicester Street, London WC2

Appendix 14*

Lasers and laser welding

Notes on the laser

The cost of the CO_2 lasers has mitigated against their general acceptance by industry, except for heavier work. They can perform welding, cutting, drilling, spot welding and marking operations quickly and efficiently, with very little distortion, due to the narrow HAZ, but it is still only the larger firms that use them where there is much repetition work. We will discuss the following types of laser here: solid state, such as the Nd–YAG, and gas, such as the CO_2 axial flow and the CO_2 transverse or cross flow.

The Nd–YAG laser

The Nd–YAG laser is of neodymium and yttrium–aluminium garnet (the last word indicating that its crystals are cubic). This laser is relatively light and portable, and its use is greatly increased by the fact that the laser beam can be conducted along a fibre-optic cable to a gun head, and can thus reach awkward places. It is used for marking hardened parts and no matter how thin, these are clearly marked using a low-power laser (Fig. A14.1). The drilling of thin sheet and stainless steel is performed with a very small focussed beam; the width of the beam determines the hole diameter.

The power output of a laser depends upon its length and its capacity to keep cool. The laser is contained within a unit called a resonator and the whole is water-cooled. The Nd–YAG laser passes the electric field, the beam and photons with ease and is used in the power range 50 W to 3 kW, the low input giving a very narrow HAZ. Above this power range the heat produced renders its operation greatly restricted. It is fast in the welding

---

* The student should read pp. 239–42 on laser beam welding.

Fig. A14.1. Simplified Nd–YAG laser.

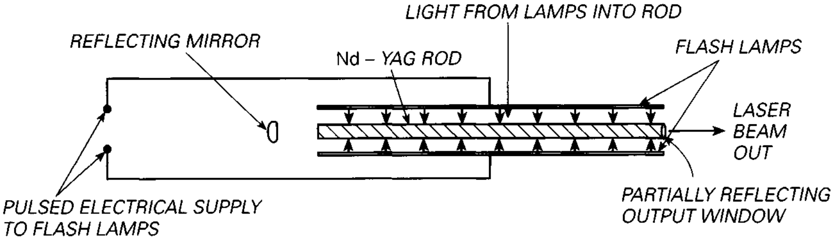
LIGHT FROM LAMPS INTO ROD
REFLECTING MIRROR
Nd - YAG ROD
FLASH LAMPS
LASER BEAM OUT
PULSED ELECTRICAL SUPPLY TO FLASH LAMPS
PARTIALLY REFLECTING OUTPUT WINDOW

Fig. A14.2. Keyhole laser welding.

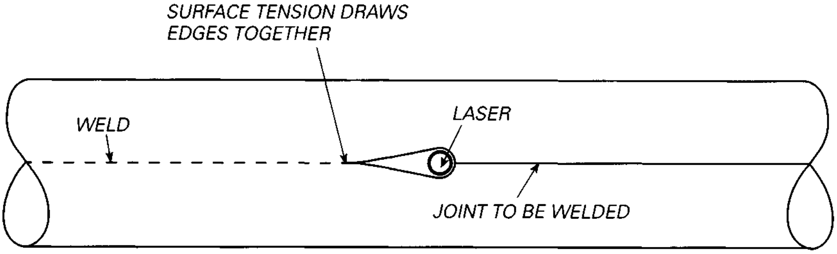
SURFACE TENSION DRAWS EDGES TOGETHER
WELD
LASER
JOINT TO BE WELDED

operation and is particularly useful in welding thin sections of steel and stainless steel of about 2 mm or less. Fit up is extremely important since any gaps between the sections being welded allow the beam to traverse in this area without welding. For this reason butt welds in shafts and those in tubes of stainless steel or steel must be perfectly butted. The spot welding of thin sheets can also be carried out without distortion. No filler rod is used and the weld is performed as a keyhole weld (described on p. 194 and illustrated in Fig. A14.2). Any gases used are described on pp. 36 and 106 of Vol. 1, and focussing the spot on p. 241.

The CO_2 laser

For welding thicker materials, the CO_2 laser is usually used, and is available in sizes 0.4–30 kW. Many lasers in the USA have been produced with very much greater outputs than this but information on them is difficult to obtain. These high-powered lasers have uses other than for welding and cutting. The solid state Nd–YAG laser is used for materials (steel) down to 0.2 mm and less. Others of this type are being developed, but the YAG rod gets very hot at these higher powers and so the CO_2 laser is used. There are two types: axial flow, in which the gas flow

Fig. A14.3. Simplified drawing of an axial flow CO_2 laser. Beam is in direction of optical axis. The whole is water-cooled.

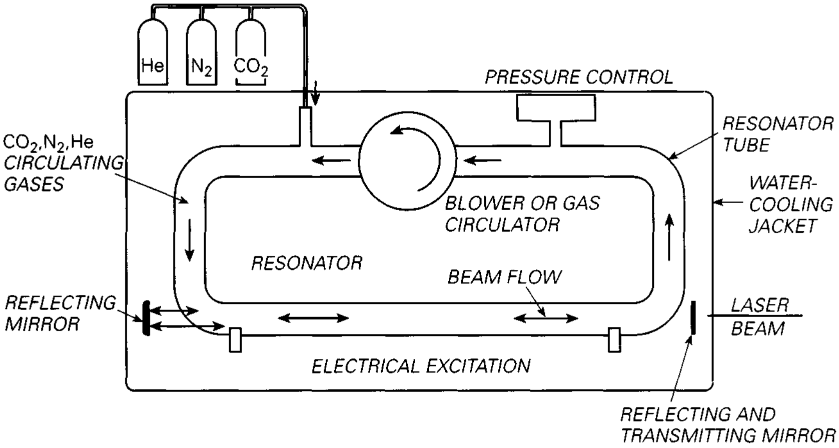
He
N2
CO2
PRESSURE CONTROL
CO2,N2,He CIRCULATING GASES
RESONATOR TUBE
WATER-COOLING JACKET
BLOWER OR GAS CIRCULATOR
RESONATOR
BEAM FLOW
REFLECTING MIRROR
LASER BEAM
ELECTRICAL EXCITATION
REFLECTING AND TRANSMITTING MIRROR

Fig. 14.4. Layout of a cross flow CO_2 laser.

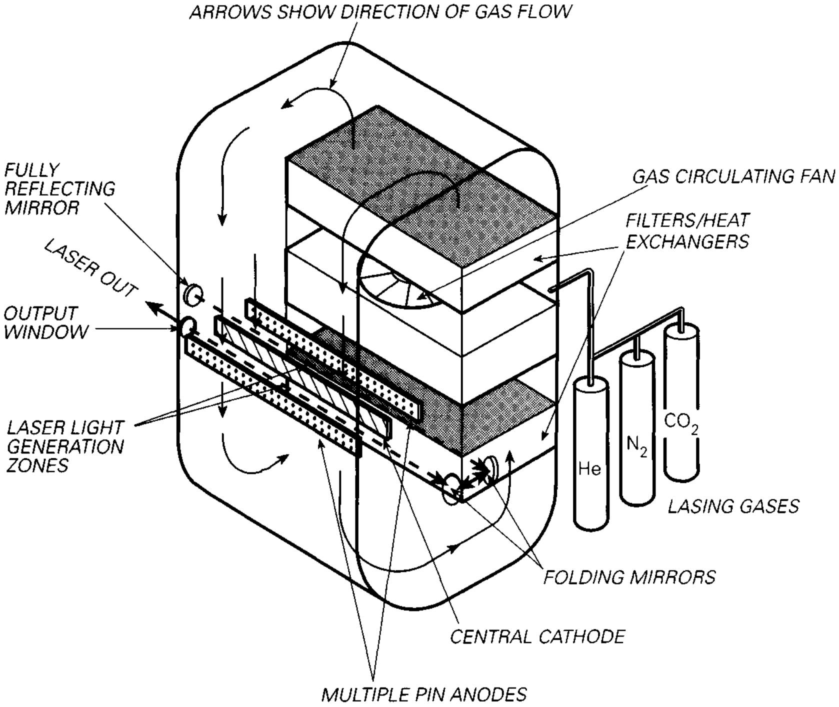
ARROWS SHOW DIRECTION OF GAS FLOW
FULLY REFLECTING MIRROR
LASER OUT
OUTPUT WINDOW
GAS CIRCULATING FAN
FILTERS/HEAT EXCHANGERS
LASER LIGHT GENERATION ZONES
He
N2
CO2
LASING GASES
FOLDING MIRRORS
CENTRAL CATHODE
MULTIPLE PIN ANODES

is in line with the optical axis, and cross or transverse flow, in which the gas flow is at right angles to the optical axis.

The axial laser consists of a resonator tube or chamber containing CO_2 and traces of He (helium) and N_2 (nitrogen) to give easy starting and good beam characteristics at a pressure of about 60 millibars. Air is pumped out of this resonator before the CO_2, He and N_2 are introduced. The gases are circulated by a fan or compressor and this keeps the pressure in the tube constant. During operation there is some degrading of these gases and this is slowly bled off and replaced by fresh CO_2, He and N_2, as required, keeping the pressure constant. Heat exchangers keep the unit cool. The electrical stimulation may be a.c. or d.c. and the power of the unit (electrical) is many times that of the laser beam, whose wavelength is approximately 1.06 μm. A simplified version of the axial flow laser is shown in Fig. A14.3, and the cross flow laser in Fig. A14.4.

The cross or transverse flow laser has the gases as before but the electrical stimulation is now at right angles to the optical axis and this type is usually of higher output than the axial flow type (Fig. A14.5).

In both cases the resonator tube usually consists of smaller tubes (lengthwise) fitted with reflective mirrors set at an angle of 45° to the axis

Fig. 14.5. Direction of gases in cross flow CO_2 laser.

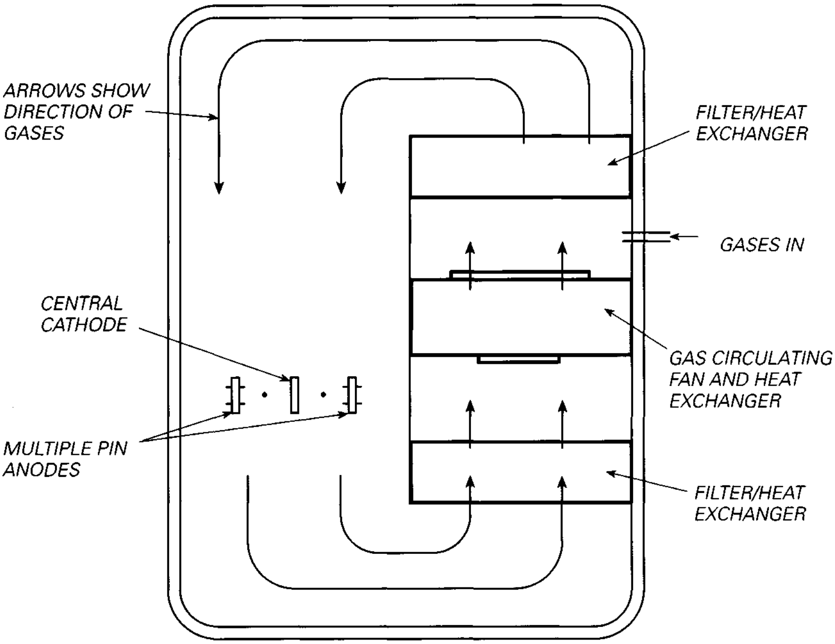
ARROWS SHOW DIRECTION OF GASES
FILTER/HEAT EXCHANGER
GASES IN
CENTRAL CATHODE
GAS CIRCULATING FAN AND HEAT EXCHANGER
MULTIPLE PIN ANODES
FILTER/HEAT EXCHANGER

Fig. 14.6. Folded-beam laser.

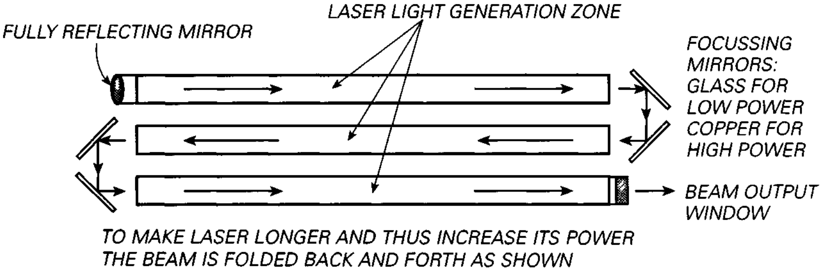
LASER LIGHT GENERATION ZONE
FULLY REFLECTING MIRROR
FOCUSSING MIRRORS: GLASS FOR LOW POWER COPPER FOR HIGH POWER
BEAM OUTPUT WINDOW
TO MAKE LASER LONGER AND THUS INCREASE ITS POWER THE BEAM IS FOLDED BACK AND FORTH AS SHOWN

of the beam (Fig. A14.6). These mirrors are generally of polished copper to stand the extra strength of the beam. The beam energy is evenly distributed over the width of the emergent beam in the case of the axial flow laser, while in the cross flow laser any parts of the beam that would miss the outgoing mirror are collected by a scraper at 45° to the emergent beam and the resultant beam is strengthened. The advantage of the folding-beam laser is that the power is increased without increasing the length of the unit.

Fig. A14.7(*a*), (*b*) shows sections through a laser weld. The parts to be welded have their ends machined so that there is an excellent fit up between them. If there is any poor fit up between the parts the beam passes this point and does not weld it. Note that the HAZ is extremely narrow, being widest where the weld commences. Because of this narrow zone there is little heat input to the parts and hence very little distortion. Sheet and plate are welded in the same manner or by spot welding in both steel and stainless steel. The method of welding is called 'keyhole' welding (Fig. A14.2, see also p. 194). No filler material is required. Since the width of the weld bead varies considerably for different size lasers it can be seen that the focussing of the beam to a small spot, say 0.15 to 2.2 mm diameter, exactly where the weld is to commence is essential. Focussing is done by an optical arrangement in the path of the emergent beam.

The laser is extremely useful in cutting both steel and stainless steel down to 0.5 mm. It is often preferable to use nitrogen as a shielding gas rather than add oxygen, since the edge of the cut is not oxidized in any way and, of course, there is little distortion.

The depth of penetration is an important feature of the laser beam. It has been found, in use, that a spot which was focussed correctly when work commenced may alter due to slight irregularities in the plates. For this reason some units are fitted with a self-adjusting mechanism (like a zoom lens) and this keeps the spot always correctly in focus and can save much time.

Fig. 14.7. (*a*) Weld in stainless steel pipe showing narrow HAZ.

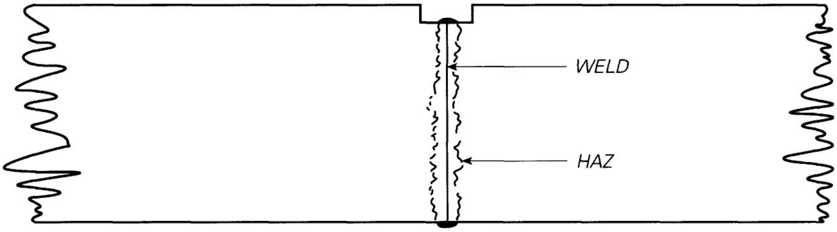

Fig. 14.7. (*b*) Section through a weld in stainless steel showing the narrow HAZ.

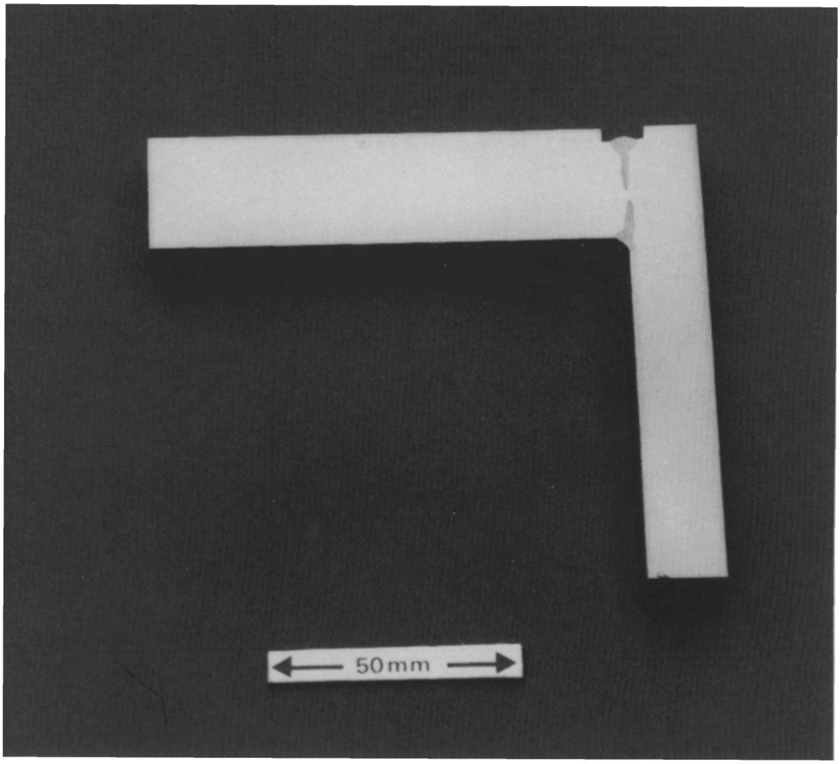

Laser welding of other metals

The laser welding of steel to steel (and stainless steel to stainless steel) can be performed quite easily since in both cases the parts to be welded have the same melting temperatures. Similarly stainless steel to ordinary steel welding is very satisfactory. When welding dissimilar metals together the melting temperatures and the ability of the metals to form a solid solution at the point of welding are very important (for example, steel to aluminium forms no solid solution). The table gives a list of dissimilar metals welded together, with notes on the weld.

Laser welding of steel, stainless steel, titanium, other alloys and cast iron

Material	Notes
Aluminium and its alloys	Not particularly satisfactory
Nickel and high nickel alloys such as MONEL®, INCONEL®, NIMONIC®, KOVAR®, HASTELOY®	Some of these weld well. Consult manufacturers
Cast iron	Any cast iron containing nickel can be welded if a nickel filler rod is used
Steel	Weldable. If high carbon content, care is required, special precautions must be taken
Low-carbon and high-strength low-alloy steels	These are very weldable if the sulphur and phosphorus content is low
Alloy steel	Weldable. Care must be taken to ensure that there is no hardening due to a fast welding rate
Stainless steels	
Austenitic	Most grades of steel give good welds except the free cutting steels with a high percentage of sulphur
Martensitic	Owing to hardening, steels with a high carbon and chromium content are satisfactorily welded. Otherwise welding is very difficult
Ferritic	Steels having low carbon and chromium (below 0.4%) give good welding properties. After welding, grain growth will affect the toughness
Titanium and its alloys	Good welds are obtainable with fine grain if the material is clean before welding. Note that high quality of shielding gas is required over the weld pool

Fig. 14.8. M/S plate. The sides of the rim are ring-pressed to shape, and laser welded outside onto the M/S ring.

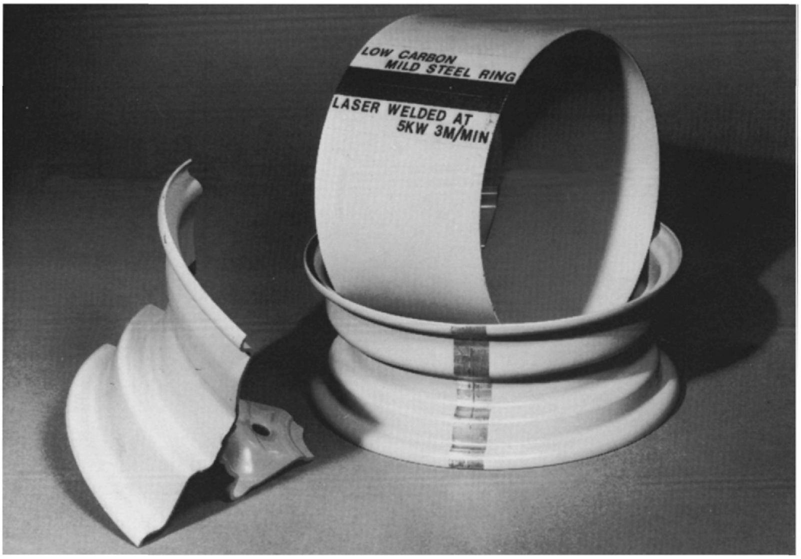
LOW CARBON
MILD STEEL RING
LASER WELDED AT
5KW 3M/MIN

Fig. 14.8. Laser-welded food mixer.

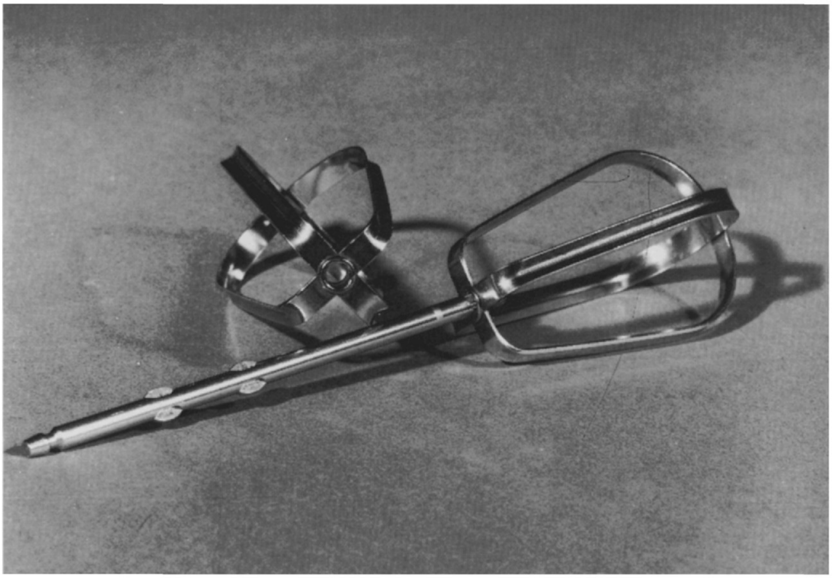

Because the beam from an Nd–YAG laser can be transmitted down a 1 mm optic fibre cable it can be used to cut holes of this size in thin steel sheet such as used in car bodies. A 25 W Nd–YAG laser can cut steel down to 1 mm thickness.

Nitrogen or argon is used for the welding of motor vehicle gears in an inert atmosphere. Glove boxes are designed so that the welds can be carried out in an inert atmosphere. This is very useful when welding any reactive metals such as titanium.

An Nd–YAG laser of 35 W capacity can be used for very small work such as engraving.

Figs. A14.8 and A14.9 show further examples of laser welding.

Advantages and disadvantages of laser welding and cutting

Advantages

(1) Very narrow HAZ, so little or no distortion (e.g. finished gears welded together).
(2) Depth of penetration limited only by power of laser.
(3) Welding speeds are increased.
(4) Useful if the point of access is poor.
(5) Useful for welding, cutting, heat treatment and marking.
(6) Very thin sections can be welded or cut with little distortion.
(7) Surface treatment is easily carried out.

Disadvantages

(1) Surface to be welded must fit up accurately.
(2) Joint alignment and focussing of beam must be accurately set.
(3) Cost of equipment and its layout is very high.
(4) Beam energy must be accurately controlled.
(5) Equipment is not easily transportable.
(6) The beam is highly dangerous and must never be allowed to fall on human skin.

Sizes of laser are roughly:

Welding – Pulsed Nd–YAG laser welding of sizes 3–4 kW, distortion-free, deep weld penetration.

Cutting – $CO_2 + He + N_2$, 0.75–25 kW and continuous wave. Cooled lasers of 2 kW and upwards output.

Marking – Lasers 25–150 W in metals, ceramics and plastics.

City and Guilds of London Institute Examination questions

Note: All dimensions are given in millimetres, unless otherwise stated.

Manual metal arc p. 467
MIG and TIG (gas shielded) p. 480
Other welding processes p. 497
Cutting processes p. 500
Oxy-acetylene p. 502
Multiple choice p. 507
Craft studies (welding engineering) p. 512
Technical grade (fabrication and welding engineering) p. 536

Manual metal arc

1 State *four* precautions which should be taken to protect the eyes and skin of the welder and other workers exposed to radiation from arc welding operations.

2 Sketch and label *three* defects that may occur in a double-vee butt welded joint which would affect fatigue strength.

3 Give *three* reasons why hydrogen-controlled electrodes are often used in preference to general purpose electrodes for welding low-alloy steels.

4 (*a*) Explain why the degree of penetration is greater using a cellulose-covered electrode compared to a rutile-covered electrode using the same heat input.
  (*b*) An electrode used for manual metal-arc welding has over 100 per cent metal recovery. State
    (1) the addition which must have been made to the covering,
    (2) *two* advantages of using this type of electrode.

5 The figure shows two views of a fractured grey cast-iron bracket which is to be repaired by the manual metal-arc process.

(*a*) Make a pictorial sketch of the bracket.

(*b*) List *eight* items of information that should be included on a welding procedure sheet for the repair of the bracket.

(*c*) State *four* additional difficulties that would occur if the bracket has been made from 3% silicon alluminium alloy.

6 Give *three* reasons why hot cracking may be a problem when manual metal-arc welding austenitic stainless steel.

7 List *four* factors that should be considered when determining the pre-heating temperature to be used for welding a steel fabrication.

Question 5

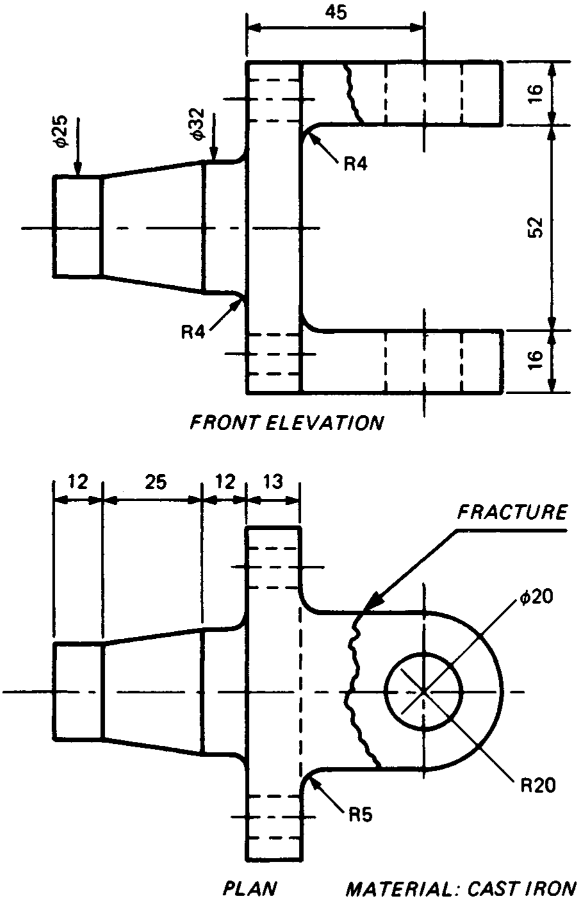
45
16
52
16
ϕ25
ϕ32
R4
R4
FRONT ELEVATION
12
25
12
13
FRACTURE
ϕ20
R20
R5
PLAN
MATERIAL: CAST IRON

8 (*a*) A 30 mm thick alloy of the composition shown has to be manual metal-arc welded using rutile-coated electrodes.

Composition: 0.2% carbon; 2% chromium; 0.5% molybdenum; 0.8% manganese; 0.4% silicon; 0.35% sulphur; 0.035% phosphorus; remainder iron.

$$\text{Carbon Equivalent} = \%\text{C} + \frac{\%\text{Mn}}{20} + \frac{\%\text{Ni}}{15} + \frac{\%\text{Cr} + \%\text{Mo} + \%\text{V}}{10}$$

(1) Using the formula given calculate the carbon equivalent of the alloy.

(2) From the information given in the table below construct a graph and from it determine the pre-heating temperature of the alloy.

Carbon equivalent %	0.3	0.35	0.4	0.45	0.5	0.55	0.6
Pre-heat temp. (°C)	100	150	200	250	300	350	400

(3) If a hydrogen-controlled electrode is used to weld the material explain why the pre-heating temperature may be lowered by 100°C.

(*b*) Two of the hazards which may arise in welding are explosions and fumes. For any *one* of these state *four* possible causes and the precautions that should be taken to avoid them.

9 Brittle fracture may occur in welded structures in low-alloy steels.

(1) List *four* possible causes.

(2) State *four* precautions that should be taken in order to reduce its occurrence.

(3) Describe one test procedure which may be used to give information on the parent plate's susceptibility to this type of failure.

10 (*a*) Explain what is meant by percentage metal recovery in manual metal-arc welding.

(*b*) Electrode manufacturers claim that over 100% metal recovery is possible when using certain of their electrodes.

(1) State the addition which must be made to the coatings of these electrodes.

(2) State *two* advantages.

11 (*a*) State *two* factors that should be considered before selecting a filler electrode for a hard surfacing application.
(*b*) State *three* precautions which may be taken by the welder to control the level of dilution when surfacing using an arc welding process.

12 Steel components are to be hardfaced by the manual metal-arc process.
(*a*) Give *two* reasons why cracking may occur.
(*b*) State *three* precautions that can be taken in order to reduce the possibility of cracking.
(*c*) Name *one* other problem that may arise when hardfacing.

13 (*a*) With the aid of sketches, show *two* examples of distortion caused by manual metal-arc welding.
(*b*) For *each* example shown, suggest a suitable method of controlling the distortion.

14 The figure shows a typical beam-to-stanchion connection used in structural steelwork. Copy the drawing as shown and insert the appropriate welding symbols to indicate the following:
(*a*) the gusset to be fillet welded on each side to the stanchion in the shop;
(*b*) the beam to be added on site, and fillet welded all round to the stanchion;
(*c*) the gusset to be fillet welded each side to the beam.

Question 14

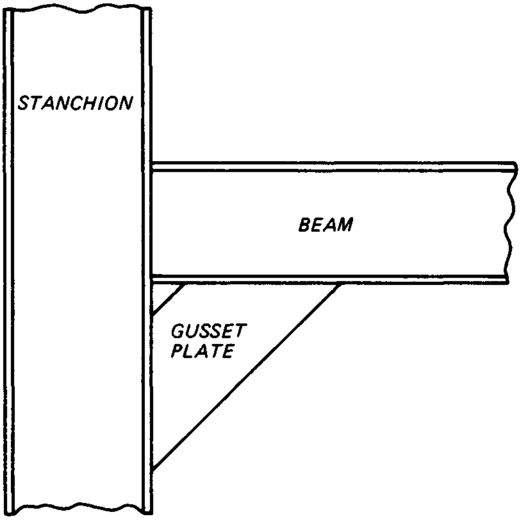

15 (*a*) Name two functions of the filter glasses used when metal-arc welding.
  (*b*) Why should the cable used for the welding lead connexion of an arc welding circuit be flexible?
  (*c*) Name two types of arc welding processes in which non-consumable electrodes are used.
  (*d*) Give one example of a semi-automatic arc welding process.

16 Name *four* different types of metal-arc welding plant which fulfil the electric power supply requirements for welding. Describe, with aid of a sketch, one of the plants named.

17 What would be the practical effect of each of the following during metal arc welding:
  (*a*) loose circuit connexions;
  (*b*) variations in the mains supply voltage?

18 Describe in detail, with the aid of sectional sketches, three defects which may be produced during the metal-arc welding in the horizontal–vertical position of close-square-tee joints in mild steel plate. In *each* case state the cause and explain how the defects should be avoided.

19 Explain briefly the effect of any two of the following upon the production of an efficient joint in a partially chamfered butt weld:
  (*a*) depth of root face,
  (*b*) the angle of the vee,
  (*c*) the gap setting.

20 Describe, with the aid of a sketch, one manual metal-arc welding technique used for making butt welds in mild steel pipelines.

21 Explain, by means of sketches, what is meant by *each* of the following:
  (*a*) The *slope and tilt* of the filler rod and blowpipe *when making* a butt weld in low-carbon steel, 2 mm thick, in the vertical position by the oxy-acetylene process.
  (*b*) The *slope and tilt* of the cutting electrode *when making* a straight cut in 8 mm low-carbon steel plate in the flat position by the oxygen-arc process.

22 (*a*) State *three* factors which may influence slag control during manual metal-arc welding.
  (*b*) Give *one* example of when each of the following are used in manual metal-arc welding:
    (1) tong test ammeter,
    (2) voltmeter.

23 (*a*) What is meant by (1) open circuit voltage, and (2) arc voltage in manual metal arc welding?

(*b*) State the likely effect on root penetration and weld deposit when manual metal-arc welding low-carbon steel 6 mm thick with too long an arc.

24 State *two* safety precautions which should be observed with *each* of the following for manual metal-arc welding:

(*a*) treating components prior to welding by the use of trichloroethylene degreasing plant,

(*b*) welding in confined spaces,

(*c*) welding in close proximity to glossy finished surfaces,

(*d*) preparing vessels which have contained liquids with flammable vapours for repair by welding.

25 (*a*) Sketch the joint set-up and state the diameter of electrodes and current values to be used when making butt welds in the flat position in *each* of the following thicknesses of low-carbon steel:

(1) 3 mm,
(2) 6 mm,
(3) 10 mm.
(4) 14 mm,

(*b*) Discuss the effect of *each* of the following factors on depth of root penetration when making butt welded joints in the flat position:

(1) current,
(2) arc length,
(3) speed of travel,
(4) angle of electrode (slope and tilt).

26 State why pre-heating is to be recommended when welded joints are to be made in *each* of the following:

(*a*) low-alloy, high-tensile steel in cold weather,

(*b*) 50 mm thick low-carbon steel,

(*c*) 6 mm thick copper plate.

27 The figure shows two views of a cast iron support bracket which has fractured in service.

(*a*) If the casting is to be replaced by a low-carbon steel welded fabrication:

(1) make a sketch showing the complete bracket assembled ready for welding. By the use of weld symbols (BS499) indicate the type and location of the weld joints;

(2) detail freehand the steel parts that you would require.

(*b*) If the casting is to be repaired by manual metal arc welding state:

(1) all preparations necessary to be made before welding;

(2) the weld procedure, including details of electrode and current values;

(3) any precautions to be carried out after welding.

28 Explain briefly why nickel is liable to crack when metal-arc welded. State briefly how this defect may be avoided.

29 With the aid of sketches, give two examples of distortion of welded work that may result from the application of metal-arc welding.

30 (*a*) Explain what is meant by the term 'arc eye'.

(*b*) What precautions should be taken to avoid 'arc eye'?

(*c*) What action should be taken in the case of severe 'arc eye'?

31 When each of the following materials is welded by the manual metal-arc process state, in each case, two difficulties which may arise:

(1) copper, (2) grey cast iron.

(*a*) State two difficulties which may be encountered when manual-metal arc welding dissimilar metals.

(*b*) Explain how these difficulties can be overcome.

32 Give *three* precautions which should be taken to produce acceptable joints when manual metal-arc welding austenitic stainless steel. Butt welds are to be made in the construction of a pipe line.

(*a*) Give *two* reasons why excessive penetration should be avoided.

Question 27

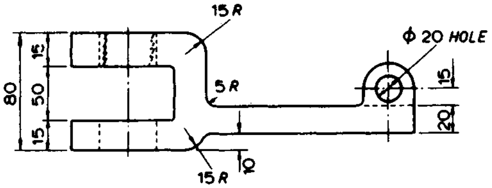

ELEVATION

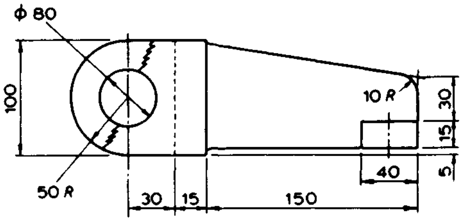

PLAN

(*b*) Describe briefly *two* methods of controlling root penetration.

33 Describe, with the aid of sketches, any *four* of the following. In *each* case state *one* typical application, *one* advantage and *one* limitation:

(*a*) stove pipe technique,
(*b*) tongue bend test,
(*c*) manipulators,
(*d*) arc on gas techniques,
(*e*) 'studding' when welding cast iron.

34 Describe, with the aid of sketches, each of the following:

(*a*) back-step welding,
(*b*) backing strips,
(*c*) backing bars.

35 The figure shows details of a pipe assembly. Answer the following:

(*a*) Recommend a suitable method in each case for producing to correct shape and size (1) the flanges, (2) the main pipe and the branch pipe.

(*b*) Describe, with the aid of sketches, a suitable edge preparation that may be used to ensure satisfactory penetration at the following joints: (1) a flange to the main pipe, (2) branch pipe to the main pipe.

(*c*) Show, with the aid of sketches, the procedure for welding both of these joints, stating the sequence of runs, the current value and diameter of electrode used in each case.

(*d*) Describe briefly how the finished welds may be tested for surface defects.

36 The figure shows details of a pipe assembly. Answer the following:

(*a*) Show, by means of a sketch, how the main pipe and the flanges may have distorted as a result of welding.

(*b*) Sketch a simple fixture that would be suitable for locating the branch to the main pipe and assist in controlling distortion.

(*c*) Describe briefly a suitable method for pressure testing the completed assembly.

(*d*) Calculate the total weight of the three flanges; 1 m^3 of low-carbon steel weighs 7750 kg.

37 Briefly explain the function of *each* of the following in manual metal-arc welding:

(*a*) a low-voltage safety device,
(*b*) a rectifier.

38 (*a*) What are the *three* basic types of wear to which hard-faced components are subjected?

(*b*) Select *one* of these types and suggest a suitable type of electrode to be used for surfacing to meet requirements.

(*c*) State *two* precautions which should be taken to minimize cracking.

39 An aluminium alloy casting is to be repaired by the manual metal-arc welding process. Give *each* of the following:

(*a*) a workshop method of indicating the pre-heating temperature,

(*b*) the type of current to be used.

(*c*) *three* factors which make this material more difficult to weld than low-carbon steel.

40 Select a suitable electrode for welding *each* of the following low-carbon steel joints:

(*a*) a severely stressed single-vee butt joint welded in the flat position,

(*b*) a lightly stressed fillet welded joint in the vertical position,

(*c*) fillet welds in the flat position where a high metal recovery rate is required.

41 State one important safety precaution that must be observed for each of the following:

(*a*) before commencing welding repair work on a two-compartment tractor fuel tank, having one compartment used for petrol and the other for fuel oil;

Question 35 and 36

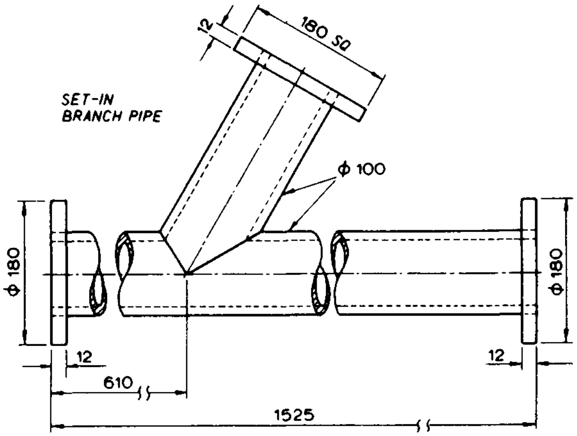

(*b*) when flame gouging in the vertical position a defective vertical joint out of a heavy fabrication in preparation for rewelding.

42 State *two* advantages which may be obtained by the use of a U preparation instead of a V preparation for welded butt joints in thicker materials.

43 Sketch the joint preparation and set-up, and state the size of the electrodes, number of runs and current values to be used for making
- (1) butt welds, without backing bar in the flat position,
- (2) tee fillet welds in the horizontal–vertical position, in each of the following thicknesses of mild steel: 6.4 mm and 12.7 mm.

44 State the plate thickness limitations of the upward–vertical and the downward–vertical metal arc welding techniques.

45 (*a*) Describe, with the aid of sketches, what effect each of the following would have on the depth of root penetration and the quality of the deposited metal:
- (1) the use of too high current value,
- (2) incorrect angle of electrode slope,
- (3) too fast a speed of travel,
- (4) incorrect arc length.

(*b*) Explain how measuring equipment could be used to check current and voltage values available in a metal arc welding circuit during welding.

46 State one safety precaution that must be carried out:
- (*a*) before commencing welding repair work on a tank which has contained acids,
- (*b*) when metal arc welding from a multi-operator set is carried out during the construction of multiple-storey steel framed structures.

47 State what is meant by:
- (*a*) deep-penetration coated electrode,
- (*b*) non-consumable electrode.

48 What is the purpose of a choke reactance used in manual metal arc welding?

49 (*a*) What is meant by arc blow in manual metal arc welding?

(*b*) Arc blow may arise when manual metal arc welding with either direct current or alternating current supply. Give *two* possible causes.

50 (*a*) State what is meant by:
- (1) rutile-coated electrode,
- (2) hydrogen-controlled coated electrode.

(*b*) Give *one* defect which may arise when manual metal arc welding low-carbon steel, using an eccentrically coated electrode.

51 Describe with the aid of sketches the *technique* required when cutting low-carbon steel plate in the flat position by the use of *each* of the following arc cutting processes:
(1) air-arc,
(2) oxygen-arc.

52 Sketch and label *two* different types of edge preparation for butt joints other than close-square butt, suitable for manual metal arc welding.

53 Give *one* possible cause of *each* of the following defects when manual metal arc welding close-square-tee fillet joints in low-carbon steel in the vertical position by the upwards technique:
(*a*) incomplete root penetration,
(*b*) undercut,
(*c*) unequal leg-length.

54 (*a*) Explain what is meant by 'percentage metal recovery' in metal arc welding.
(*b*) If over 100% is claimed, what additions could have been made to the electrode coating?

55 Manufacturers of metal arc welding electrodes take care to specify a minimum and maximum current for each size and type of mild steel electrode. Discuss in detail the important consequences of using
(*a*) insufficient current,
(*b*) excessive current.

56 (*a*) State *three* reasons for using welding fixtures in welding fabrication.
(*b*) State *three* basic factors that must be considered for the effective operation of a welding fixture.
(*c*) Describe briefly, with the aid of a sketch, the principle of operation of any welding fixture with which you are acquainted.

57 A number of butt welded joints have to be made between the ends of 100 mm diameter low-carbon steel pipe. Sketch a simple jig for holding them in position for tack welding.

58 Give *four* important functions of a slag.
(*a*) Briefly describe one test, other than visual inspection, used to reveal surface defect in welded joints.
(*b*) Give one limitation of such a test.
(*c*) Name four constituents used in electrode coatings.

(*d*) State the most important function of any one of these constituents.

59 Give *four* differences in the behaviour of the metal that a welder will find between the metal arc welding of (*a*) aluminium and (*b*) low-carbon steel.

60 (*a*) Describe in detail the metal arc welding process commonly used for manually welding mild steel.

(*b*) Give *two* examples showing how the process in (*a*), using standard electrodes, may be partially automated so as to reduce costs.

61 State *three* factors that may influence metal transfer phenomena when using a metal arc welding process.

62 Give *three* reasons why stray-arcing is undesirable.

63 Give *three* reasons why damp flux-coated electrodes should not be used for welding mild steel.

64 With the aid of a sketch, show the principle of the buttering technique.

65 (*a*) Construct a table showing the ranges of welding currents typically used with 2.5 mm, 3.2 mm, 5.0 mm and 6.0 mm general-purpose, mild steel electrodes.

(*b*) Plot a graph from the figures in the table in part (*a*).

(*c*) From the graph in part (*b*) determine the probable current range to be used with a 4.0 mm electrode.

66 State *three* safety precautions to be observed when using engine-driven metal arc welding equipment.

67 Give any *two* advantages of using multi-operator a.c. metal arc welding equipment.

68 The figure shows two views of a bracket which has to be fabricated by manual metal arc welding.

(*a*) Make a pictorial sketch of the bracket and indicate, by the use of weld symbols, according to BS499, the location and the type of weld joints required.

(*b*) List *ten* items of information that could be included on a Welding Procedure Sheet for the bracket shown.

69 The figure shows two views of a cast iron bracket which has fractured in service.

(*a*) If the casting is to be repaired by manual metal arc welding,

(1) state *all* preparations necessary to be made before welding,

(2) describe a suitable procedure, including details of the type of electrodes and the current values,

(3) state *two* precautions carried out after welding.

Question 68

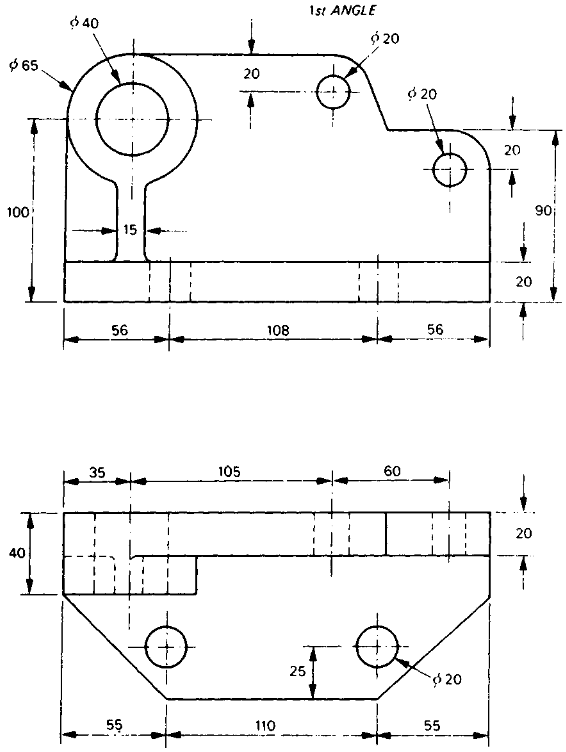
1st ANGLE
ϕ 40
ϕ 65
ϕ 20
ϕ 20
20
20
90
100
15
20
56
108
56
35
105
60
40
20
25
ϕ 20
55
110
55

ALL DIMENSIONS IN MILLIMETRES

Question 69

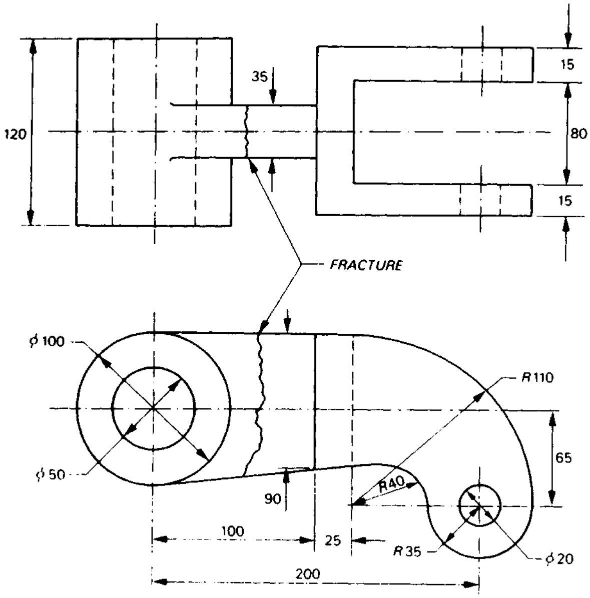
120
35
15
80
15
FRACTURE
ϕ 100
R110
65
ϕ 50
R40
90
100
25
R35
ϕ 20
200

ALL DIMENSIONS IN MILLIMETRES

(*b*) Make a sketch to show the complete bracket assembled ready for welding, if the casting is to be replaced by a low-carbon steel welded fabrication. By the use of weld symbols (BS499), indicate the type and location of the weld joints to be used.

70 Give *one* reason why, in particular situations, it might be difficult to control the arc when arc welding mild steel.

71 Give what you consider to be *three* important factors likely to affect the cost of the welded fabrication of a very large one-off component.

72 (*a*) By means of a labelled sketch show what is meant by buttering.

(*b*) Name *one* material on which buttering could helpfully be used to make an effective arc fusion welded joint.

73 Explain how the difficulties which may arise from the presence of hydrogen may be overcome during the welding of low-alloy steels.

74 What is an alloy? Give *two* reasons why an alloying addition might be necessary in the composition of a metal intended for use as weld metal for joints to be made by gas-shielded arc welding.

Gas shielded metal arc MIG *and* TIG

1 (*a*) List *four* welding problems to be overcome when gas-shielded metal-arc welding austenitic heat-resisting steels to be used for high-temperature service.

(*b*) State *three* features which should be present in a manipulator intended for extensive gas-shielded metal-arc welding of joints in a large assembly.

(*c*) Outline *three* functions of jigs designed to be used for the metal-arc gas-shielded welding of joints in components during flow production.

2 (*a*) State *five* factors which influence the cost of metal-arc gas-shielded welding in the batch production of welded fabrications in non-ferrous metals.

(*b*) A butt welded joint in a flat, low-carbon steel bar L mm wide and t mm thick is to be made by the metal-arc gas-shielded process. When in service a static tensile load of P newtons will be carried by the butt welded joint. Show by means of a simple formula how the tensile stress in the butt weld may be calculated.

(*c*) Describe, with the aid of sketches, a technique which may be used for making a fillet welded part-lapped close outside corner joint in the flat position in 6 mm thick aluminium plate by the metal-arc gas-shielded process.

3 (*a*) With the aid of a labelled block diagram to show the arrangement, describe the equipment necessary to fulfil the electrical gas-shielding and ancillary requirements when one operator is required to carry out effective tungsten-arc gas-shielded welding with
  (1) alternating current,
  (2) direct current as required for different types of material.
(*b*) State the purpose of *each* item of equipment in the system and give details of current control together with a labelled sketch to show the power source characteristic.

4 (*a*) State *four* different-types of material suitable for tungsten-arc gas-shielded welding.
(*b*) In the case of *two* of these materials state
  (1) *two* difficulties that may arise during the welding of the material,
  (2) the methods used to overcome the difficulties encountered.

5 (*a*) List *four* gases *or* gas mixtures used in tungsten-arc welding and give *one* typical application of *each.*
(*b*) Describe with the aid of sketches, a technique used for making a close-square butt joint in the vertical position by the upward method in 2.5 mm thick austenitic stainless steel by the tungsten-arc process.
(*c*) State *four* possible causes of incomplete penetration when making butt welds by the tungsten-arc process in the flat position.

6 (*a*) Outline *two* essential functions of the gas shield required for effective gas-shielded metal-arc welding.
(*b*) List *four* gases *or* gas mixtures, other than carbon dioxide, used in metal-arc gas-shielded welding and give *one* typical application of *each.*
(*c*) State *two* possible causes of *each* of the following defects when encountered in welds made by the metal-arc gas-shielded process in low alloy steel:
  (1) porosity,
  (2) cracking.
(*d*) Explain why care should be taken to avoid leakage at joints in the shielding gas supply lines when gas-shielded metal-arc welding aluminium alloys.

7 (*a*) Show, by means of a labelled diagram, the voltage waveform obtained when using *either* a high-frequency unit *or* a surge injector for gas-shielded tungsten-arc welding with alternating current supply.

(*b*) State the function of each of the regularly alternating high and low current levels used in controlled spray (pulse) mode of metal transfer in gas-shielded metal-arc welding.

8 (*a*) Outline *one* method of arc length control which may be used in metal-arc gas-shielded welding.

(*b*) Show, by means of labelled sketches, the stages in spray mode of metal transfer when metal-arc gas-shielded welding.

9 (*a*) Explain, with the aid of a sketch, how the level of dilution of a butt-welded joint may be determined.

(*b*) Show by means of a labelled sketch *one* type of edge preparation used to control pick-up effects when making a butt-welded joint in clad steel.

10 The following gas shielding mixtures are used in gas-shielded arc welding:

(*a*) argon,
(*b*) argon and hydrogen,
(*c*) argon and oxygen,
(*d*) argon and nitrogen,
(*e*) argon and helium.

State, in *each case*, a material and the welding process for which these mixtures are best suited.

11 (*a*) Name the type of current and electrode required for gas-shielded tungsten-arc welding *each* of the following materials:

(1) low-carbon steel,
(2) aluminium alloy,
(3) copper alloy.
(4) austenitic stainless steel.

(*b*) State *two* factors, other than thickness of material, which should be considered when selecting the filler wire diameter to be used for tungsten-arc welding.

12 (*a*) Outline the procedure for rectifying a fault which has been indicated in a metal-arc gas-shielded welding circuit by failure to strike an arc and a falling voltage reading at the power source.

(*b*) Show, by means of a sectional sketch, the contact tube setting and electrode extension required for dip transfer welding of low-carbon steel.

13 (*a*) Sketch a tungsten-arc welding torch fitted with a gas lens and increased electrode extension, indicating the pattern of gas flow which would be obtained from the torch.

(*b*) State *two* reasons why rectification must be controlled when alternating current is used for the gas-shielded tungsten-arc welding of aluminium.

14 (*a*) State *two* safety precautions, other than ventilation, which are necessary when using gas-shielded arc welding processes.

(*b*) State how
   (1) ozone,
   (2) carbon monoxide
may be formed during gas-shielded arc welding.

15 (*a*) State the purpose of a choke reactance when used for gas-shielded metal-arc welding.

(*b*) State the function of
   (1) a surge injector,
   (2) a high frequency unit,
when tungsten arc welding with alternating current.

16 Show by means of a labelled diagram of output curves the essential differences between the characteristics of the power source suitable for

(*a*) manual tungsten-arc gas-shielded welding,
(*b*) semi-automatic metal-arc gas-shielded welding with a self-adjusting arc.

17 (*a*) State what current values are being shown by the ammeter when welding with controlled spray (pulse) mode of metal transfer.

(*b*) Show, by means of labelled sketches, the stages in metal transfer when metal-arc gas-shielded welding by means of dip mode of metal transfer.

18 State, in *each* case, *two* precautions which must be taken in the storage of

(*a*) electrode wires for metal-arc gas-shielded welding,
(*b*) filler wires for tungsten-arc gas-shielded welding,

to ensure effective welding when required.

19 (*a*) During gas-shielded tungsten-arc welding, porosity has appeared in the weld face and metal deposition has become difficult. State *two* possible causes that may produce this condition.

(*b*) Give *one* suitable material in *each* case which may be used for a backing bar insert when tungsten-arc gas-shielded welding
 (1) ferrous metals,
 (2) non-ferrous metals.

20 (*a*) Give *two* undesirable effects which may occur when using excessive wire feed speed during gas-shielded metal-arc welding.

(*b*) State the shielding gas or gas mixture which is best suited to obtain the required modes of metal transfer for the efficient metal-arc gas-shielded welding of *each* of the following metals:
 (1) low-carbon steel and low-alloy steel by spray transfer in the flat position,
 (2) austenitic stainless steel and aluminium by controlled spray (pulse) transfer in the vertical position.

21 (*a*) State how gas flow rate is controlled and measured to enable the production of efficient welded joints in tungsten-arc gas-shielded welding.

(*b*) Name *four* factors upon which the selection of the correct current value will depend in tungsten-arc welding.

22 (*a*) Give *two* advantages which may be obtained by the use of the metal-arc gas-shielded spot welding process instead of the electric resistance spot welding process.

(*b*) State *one* typical application in each case for the following welding process:
 (1) electron beam,
 (2) friction.

23 (*a*) Outline the procedure for rectifying a fault which has been indicated in a metal-arc gas-shielded welding circuit by failure to strike an arc and a falling voltage reading at the power source.

(*b*) Show, by means of a sectional sketch, the contact tube setting and electrode extension required for dip transfer welding of low-carbon steel.

24 (*a*) Sketch a tungsten-arc welding torch fitted with a gas lens and increased electrode extension, indicating the pattern of gas flow which would be obtained from the torch.

(*b*) State *two* reasons why rectification must be controlled when alternating current is used for the gas-shielded tungsten-arc welding of aluminium.

25 Explain why the presence of moisture should be avoided when gas-shielded arc welding low-alloy steels.

26 (*a*) State in *each* case *one* probable cause of the following weld defects in joints made by tungsten-arc welding:
(1) lack of root penetration,
(2) porosity.

(*b*) State *two* reasons why abrupt changes of section should be avoided in the contours of a welded joint.

27 (*a*) Outline *two* essential functions of the gas shield required for tungsten-arc welding.

(*b*) State *two* factors which influence the selection of the flow rate of argon gas necessary to obtain efficient welded joints.

(*c*) State in *each* case *two* difficulties that may arise during the tungsten-arc welding of
(1) stainless steel,
(2) aluminium,
(3) magnesium alloy.

28 (*a*) Describe, with the aid of a sketch, the distortion effects which are likely to be produced when an unrestrained 450 mm long single-vee butt joint between 6 mm thick austenitic stainless steel plates 225 mm wide is made by the tungsten-arc process in two runs from one side.

(*b*) Give *two* methods used to control distortion.

(*c*) Outline, with the aid of sketches, *two* methods of providing gas backing for the butt welding of pipe joints.

29 (*a*) Outline the shut-down procedure to be followed at the end of work with tungsten-arc welding plant in order to ensure safety and continued efficiency of equipment.

(*b*) Show, by means of a labelled sketch, the gas paths and the function of *each* of the main components of a water-cooled tungsten-arc welding torch.

(*c*) Describe, with the aid of sketches, a tungsten-arc technique used for making a close-square butt-welded joint in the vertical position by the upward method in 1.5 mm thick austenitic stainless steel.

30 (*a*) Describe, with the aid of a labelled block arrangement diagram, the purpose of each part of the equipment essential for argon-shielded metal-arc welding aluminium alloys.

(*b*) State *one* precaution which must be taken with the temporary backing bars used in welding aluminium alloys in order to ensure the production of efficient joints.

(*c*) State the shielding gas or gas mixture which is most suited to obtain controlled spray (pulse) mode of metal transfer for welding
  (1) low alloy steel,
  (2) aluminium.

31 (*a*) List *four* welding problems to be overcome when welding steel containing 9% nickel, which is to be used for low-temperature service.

(*b*) State *three* features which should be present in a jig designed for use in the welding of joints made in components during batch production.

(*c*) Describe, with the aid of sketches, a gas-shielded metal-arc technique which may be used for making a lap-fillet welded joint in the vertical position by the downward method in 3 mm thick low-carbon steel plate.

32 (*a*) List *three* advantages and *three* disadvantages in *each* case which may be obtained by the use of
  (1) carbon dioxide,
  (2) argon,
for metal-arc gas-shielded welding.

(*b*) Outline *two* effects of the chromium content on the weldability of low-alloy high-tensile steel.

(*c*) State *two* essential factors which must be considered before making welded joints between dissimilar metals by the metal-arc gas-shielded welding process.

33 (*a*) Give *two* reasons why it is necessary to be particularly careful with the insulation of components in the welding circuit when using high-frequency current in tungsten-arc gas-shielded welding operations.

(*b*) Give *two* safety precautions which should be taken when preparing to use the metal-arc gas-shielded welding process for making welds on site in an exposed position which is 12 metres above ground level.

34 (*a*) State the purpose of a drooping voltage transformer-rectifier when used for tungsten-arc gas-shielded welding.

(*b*) Outline *two* methods used to control the current surges at the short-circuitings which take place when metal-arc gas-shielded welding with dip mode of metal transfer.

35 (*a*) Explain what is meant by *each* of the following in fusion welding:
  (1) dilution,
  (2) pick-up.

(*b*) Show by means of a labelled sketch *one* type of joint preparation used to control pick-up effects when making a butt-welded joint in clad steel.

36 (*a*) State *two* reasons why the d.c. component must be suppressed when alternating current is used for gas-shielded tungsten-arc welding.

(*b*) Outline, with the aid of a sketch, *one* method of providing back purging for the butt welding of a joint in flat plate by the tungsten-arc process.

37 (*a*) State *one* undesirable effect in *each* case which is likely to be produced when metal-arc gas-shielded welding with wire feed drive roll pressure adjustment which is

(1) insufficient,
(2) excessive.

(*b*) Give *four* routine checks which should be made on the electrode wire before fitting a new spool into a gas-shielded metal-arc welding plant.

38 Distortion may result from gas-shielded arc welding operations. State *two* possible causes and give *two* methods of control.

39 (*a*) Give *one* reason in *each* case for the use of

(1) a remote control switch,
(2) a welding contactor,

in tungsten-arc welding.

(*b*) State the likely effect of allowing cables in the welding circuit to coil when using high-frequency currents in gas-shielded tungsten-arc welding.

40 (*a*) Explain the technique to be followed to avoid contamination of the electrode when tungsten-arc welding butt joints in thicker section material in the flat position.

(*b*) Give *two* advantages which may be obtained by the use of the two-operator upward-vertical technique for tungsten-arc welding.

41 (*a*) Give *two* reasons why flux-cored electrode wire is used with carbon dioxide shielding for the metal-arc gas-shielded welding of steel.

(*b*) When using a gas-shielded arc welding process, on an exposed site, which *two* problems are most likely to be encountered as a result of weather conditions?

42 (*a*) What advantageous feature is provided by a constant potential power source when the arc is initiated to start the weld in metal-arc gas-shielded welding?

(*b*) Outline the procedure for rectification when a fault condition has been indicated in a metal-arc gas-shielded welding circuit, by falling voltage reading at the power source, when the end of the electrode wire is touched on to the work with the torch control trigger actuated.

43 Name *four* factors which may affect the cost of metal-arc gas-shielded welding a fabrication in a non-ferrous alloy.

44 With the aid of a sketch, explain briefly the principles of the tungsten-arc spot welding process.

45 (*a*) Describe, with the aid of a sketch, the distortion effects which are likely to be produced when an unrestrained 300 mm long single-vee butt joint between 5 mm thick austenitic stainless steel plates, 150 mm wide, is made in two runs from one side by the tungsten arc process.

(*b*) Outline *two* methods used to control distortion in tungsten-arc welding operations.

46 List *four* of the main welding problems encountered when tungsten-arc gas-shielded welding magnesium alloys in thicknesses ranging from 2 mm to 10 mm, indicating how *each* of these problems is overcome, and mentioning the effects of surface preparation.

47 Twenty 2 metre pipe lengths are to be fabricated from existing stock of 1 metre lengths of 150 mm internal diameter low-carbon steel pipe of 10 mm wall thickness by metal-arc gas-shielded welding. Describe, with the aid of sketches, *each* of the following for the welding of one joint:

(*a*) a suitable joint preparation and set-up,
(*b*) the assembly and tack welding procedure,
(*c*) the mode of metal transfer, shielding gas and electrode wire size to be used,
(*d*) an effective welding procedure.

48 (*a*) Compare the use of carbon dioxide with the use of argon for metal-arc gas-shielded welding by listing *two* relative advantages and *two* relative disadvantages which may be obtained by the use of *each* gas.

(*b*) State *two* advantages which may be obtained by the use of a manipulator as an aid to fabrication by metal-arc gas-shielded welding.

49 State the sequence which could be used to complete the joints numbered 1, 2, 3, and 4 in the figure of a welded fabrication made from five parts as shown. Give reasons for your answer.

50 For joint 1 in the figure give details of (*a*) the type of preparation and (*b*) the welding procedure to be used.

51 For joint 2 in the figure give details of (*a*) the type of preparation and (*b*) the welding procedure to be used.

52 Describe briefly four problems which may be encountered in the welding of the fabrication shown in the figure.

53 State why a self-adjusting arc is 'not likely to operate effectively in CO_2 shielded metal-arc welding'.

54 By means of a labelled block diagram show clearly the equipment required to give effective control of gas flow to the torch or welding head during gas-shielded arc welding. Assume that a suitable gas is being used.

55 By means of a labelled sketch indicate any *three* types of weld defect likely to occur in gas-shielded arc welded joints.

56 What is the purpose of a suppressor unit in a gas-shielded arc welding circuit?

57 State *three* different types of material suitable for tungsten arc gas-shielded welding. In *each* case state

(*a*) two difficulties that may arise during the welding of the material,

(*b*) the methods used to overcome the difficulties encountered.

58 With the aid of a simple labelled block diagram show the name, location and purpose of each part of the equipment essential for the effective tungsten-arc gas welding of aluminium.

Questions 49–52. Section through a heater shield. Material: 18% Cr, 8% Ni austenitic steel, Ti stabilized. Process: gas-shielded tungsten arc welding.

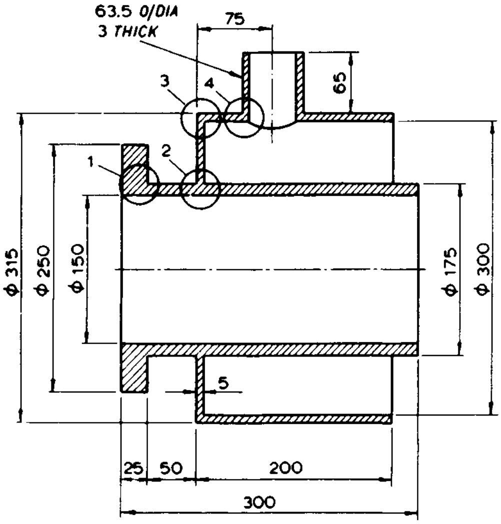

59 (*a*) Sketch a tungsten-arc welding torch fitted with gas lens, indicating the pattern of gas flow which would be obtained from the torch.

(*b*) State *two* advantages which may be obtained by the use of a gas lens in tungsten-arc welding.

60 (*a*) Describe briefly the recommended procedure to follow for clearing a 'burn back' in gas-shielded metal-arc welding.

(*b*) State *four* causes, other than joint preparation, of lack of root penetration in butt-welded joints made by gas-shielded metal-arc welding.

61 (*a*) Outline the typical main functions of a jig or fixture suitable for use in tungsten-arc gas-shielded welding.

(*b*) Give *two* situations in which the use of a jig or fixture would be considered essential, in the manufacture of a component by tungsten arc gas-shielded welding.

62 Explain what is meant by gas-backing in tungsten-arc welding and show *two* ways in which it may be applied.

63 List *two* advantages and *two* limitations of gas-shielded metal-arc welding as a process for general-purpose welding repair work.

64 With the aid of a simple labelled block diagram show the name, location and purpose of each part of the equipment essential for the effective CO_2 shielded metal-arc welding of low-carbon steels.

65 State *four* features which should be present in a manipulator intended for extensive gas-shielded metal-arc welding of joints in a large component.

66 (*a*) State the type of filter which should be used to provide eye and skin protection against radiation effects when gas-shielded arc welding with:

(1) the tungsten arc process,
(2) the metal arc process.

(*b*) Give *one* reason, in *each* case, for the type of filter used.

67 (*a*) What is the purpose of a choke reactance when used for gas-shielded metal-arc welding?

(*b*) State *two* effects which may be produced by using excessive electrode wire extension during welding.

68 (*a*) Give *two* advantages of tungsten-alloyed electrodes over plain tungsten electrodes when used for tungsten-arc welding.

(*b*) Show by means of simple labelled sketches what is meant by the electrode tip (vertex) angle when tungsten-arc welding with:

(1) alternating current,
(2) direct current.

69 Show by means of a single labelled sketch the essential differences between welding power source characteristics suitable for:
(*a*) manual tungsten-arc welding,
(*b*) welding with a self-adjusting arc.

70 (*a*) State *two* factors which influence the amount of dilution of the weld deposit in gas-shielded arc welding.
(*b*) With the aid of an outline sketch, show how the level of dilution of the welded joint may be determined.

71 (*a*) State *two* modes of metal transfer, other than spray, used for gas-shielded metal-arc welding.
(*b*) Give *one* application of *each* mode of transfer named.

72 (*a*) Give *two* important physical properties of ceramic gas nozzles.
(*b*) State *two* of the factors which govern the gas-shielding necessary to obtain efficient welded joints by tungsten-arc welding.

73 (*a*) State the type of current and tungsten alloyed electrode recommended to be used for tungsten-arc welding *each* of the following materials:
(1) low-carbon steel up to 2 mm thick,
(2) austenitic stainless steel up to 5 mm thick,
(3) aluminium over 3 mm thick,
(4) magnesium alloy up to 5 mm thick.
(*b*) List *three* gases or gas mixtures in tungsten-arc welding and give one typical use of each.

74 (*a*) State *four* factors that have to be taken into account when costing for tungsten-arc gas-shielded welding.
(*b*) Explain what is meant by:
(1) arcing time,
(2) floor to floor time.

75 (*a*) List the main welding problems when tungsten-arc gas-shielded welding copper plates in thicknesses ranging from 3 to 10 mm.
(*b*) Outline with the aid of sketches how each of these problems may be counteracted.

76 (*a*) Show by means of a labelled sketch *one* method of providing gas backing for the tungsten-arc welding of butt joints in plate in the flat position.
(*b*) Why should syphon-type cylinders be used to supply carbon dioxide for shielding gas in welding?

77 With the aid of simple sketches show how you would prepare the joints for welding the top central part of the beam shown in the

figure, where the central boss and the top flange join with each other and where the vertical web joins with the top plate.

78 State the welding sequence which should be used for welding the beam shown in the figure. Give reasons for each step.

79 Describe four of the problems which are likely to arise in welding the ends of the beam shown in the figure.

80 On a sketch of the figure give all the information needed to show the type and location of each welded joint, using the system given in BS 499 (Part 2).

81 Give *three* reasons why a suitable shade of filter glass should be used for viewing the arc during gas-shielded arc welding.

82 The following gas-shielding mixtures are used in gas-shielded arc welding:

(*a*) argon + carbon dioxide + oxygen,
(*b*) argon + oxygen,
(*c*) argon + hydrogen,
(*d*) helium + argon,
(*e*) helium + argon + carbon dioxide.

State in *each* case:

(1) the material for which the mixture is best suited,
(2) the approximate percentages of gases in the mixture.

83 State *five* desirable features in jigs intended for gas-shielded metal-arc welding of joints in components to be mass produced.

Questions 77–80. Support beam. Material: high yield stress structural steel (BS 4360 Grade 50) welded by CO_2 shielded metal arc process.

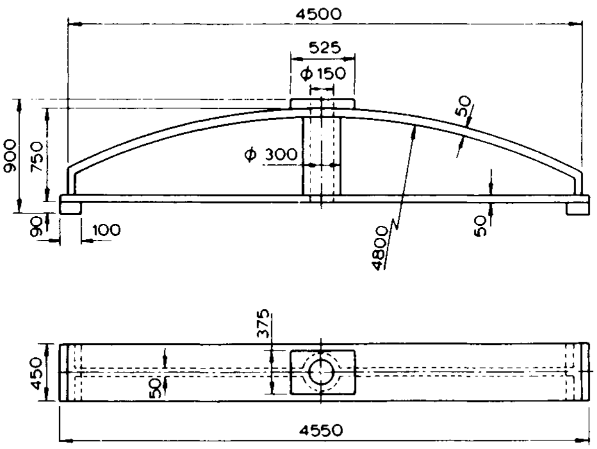

84 (*a*) With the aid of a labelled block diagram to show the arrangement, describe the purpose of *each* part of the equipment necessary for the effective argon-shielded metal-arc welding of aluminium alloys.

(*b*) State *one* precaution which must be taken in *each* case, with:

(1) the shielding gas supply lines,
(2) temporary backing bars,

in order to assist in the production of efficient joints.

85 (*a*) Give *one* safety precaution to be taken before tungsten-arc gas-shielded welding when the gas cylinder is located near the welding area.

(*b*) Give *two* hazards present when metal arc gas-shielded welding equipment is inadequately earthed.

86 (*a*) State the purpose of a combined transformer-rectifier when used for tungsten-arc welding.

(*b*) Give *two* important reasons for the use of a welding contactor in tungsten-arc welding.

87 (*a*) State the type of current and tungsten alloyed electrode recommended to be used for tungsten-arc welding *each* of the following materials:

(1) low-carbon steel up to 2 mm thick,
(2) austenitic stainless steel up to 5 mm thick,
(3) aluminium over 3 mm thick,
(4) magnesium alloy up to 5 mm thick.

(*b*) List *three* gases or gas mixtures in tungsten-arc welding and give one typical use of each.

88 (*a*) State the purpose of a surge injector unit as used in arc welding.

(*b*) State the purpose of a suppressor unit as used in arc welding.

89 A cylindrical vessel 620 mm diameter by 2.5 metres long with low-carbon steel flanged ends 25 mm thick is to have two austenitic stainless steel pipes coming out at right angles to the vessel axis and to each other midway along the vessel axis. The pipes are 100 mm inside diameter with a wall thickness of 9.5 mm and the vessel is made of 12.5 mm thick low-carbon steel, clad inside with austenitic steel to a further 1.6 mm thickness. The pipes and flanges are to be joined to the vessel by gas-shielded tungsten-arc welding.

(*a*) Give particulars of (1) the type of joint, (2) the joint preparation and (3) the welding procedure that you recommend for the welding of flanges to the vessel.

(*b*) Give particulars of (1) the type of joint, (2) the joint preparation and (3) the welding procedure that you recommend for welding the pipes into the vessel.

90 A special I-section girder 6 metres long has the following sectional dimensions: flanges 450 mm wide by 50 mm thick, web 19 mm thick, and overall depth 1 metre. Stiffening ribs 150 mm wide by 12.5 mm thick, running from flange to flange, are located opposite each other against the web at right angles to the rider axis at 1.2 m intervals. The girder is to be fabricated from high-tensile constructional steel of welding quality by gas-shielded metal-arc welding.

(*a*) Give particulars of (1) the type of joint, (2) the joint preparation, and (3) the welding procedure that you would use for joining the web to the flanges.

(*b*) Give particulars of (1) the type of joint, (2) the joint preparation, and (3) the welding procedure that you would use for attaching the stiffening ribs.

91 There are particular economic and metallurgical problems associated with the effective use of gas-shielded metal-arc welding in production applications. Outline these problems and explain how they may be overcome.

92 Austenitic heat-resisting steels containing higher nickel and chromium contents are used for fabricating structures for high-temperature service.

(*a*) State *five* main welding problems that are likely to arise in the metal-arc gas-shielded welding of this type of material.

(*b*) Explain how *each* of these problems may be effectively overcome to ensure efficient welded joints for service at high temperatures.

93 (*a*) When a d.c. arc is operating, what proportion of the heat may be generated at each side of the arc?

(*b*) Would the heating situation be the same when tungsten-arc gas-shielded welding aluminium with the electrode positive?

94 A gas pressure regulator and a flowmeter are each essential for the successful operation of a gas-shielded arc welding process.

(*a*) With the aid of an outline diagram show the usual location of each in a conventional gas-shielded tungsten-arc welding arrangement.

(*b*) Why is it necessary to have a gas pressure regulator?

95 In gas-shielded tungsten-arc welding each of the following plays an important part: (1) length of electrode projecting beyond the nozzle, (2) length of arc, (3) angle of electrode relative to the workpiece.

Give *two* important reasons why small diameter filler wire is used for gas-shielded metal-arc welding.

96 (*a*) State one difficulty with the dip-transfer mode of metal deposition as it is used in CO_2 shielded metal-arc welding.

(*b*) State briefly how this difficulty is overcome.

97 On a simple labelled block diagram show the name and purpose of each essential part of the plant and equipment needed for making tungsten-arc gas-shielded welds in a variety of metals and a range of thicknesses.

98 Give two checks you should make on the electrode wire before fitting a new spool into a gas-shielded metal-arc welding machine.

99 Why is a double-bevel or a double J butt-welded tee joint preferable to a double fillet welded close-square-tee joint for welding two highly stressed members?

100 (*a*) Explain why, in spite of the overall efficiency of argon as a shielding gas, so much effort is spent in developing the use of gases such as carbon dioxide.

(*b*) Name two difficulties likely to be encountered in using carbon dioxide as a shielding gas in arc welding.

(*c*) For each of the difficulties given in your answer to (*b*) name one method used for overcoming the problem.

101 State the main difficulty you would expect to have to overcome in joining each of the following types of material by gas-shielded tungsten-arc welding:

(*a*) malleable cast iron,

(*b*) solution-treatable aluminium alloy.

102 (*a*) Name any *two* limitations of the gas-shielded tungsten-arc welding process.

(*b*) Name any *two* limitations of the gas-shielded metal-arc welding process.

(*c*) Why is a gas shield needed in the two types of arc processes in which it is used?

103 (*a*) In tungsten-arc welding how should the welding arc be initiated without contact between electrode and parent metal?

(*b*) With the aid of a simple line diagram show all the essential equipment for the conditions that you outline in (*a*) and clearly label each part.

104 (*a*) With the aid of sketches show how a single-vee butt joint preparation for tungsten-arc welding should differ from that of gas-shielded metal-arc welding in the same thicknesses of similar materials.

(*b*) Give the main reason why they should differ.

105 The support column shown in the figure is to be made up from 25 mm thick plates of high-strength low-alloy structural steel mounted on a square mild steel base plate. The sizes are given on the diagram.

(*a*) Outline the main material problems to be overcome in welding this construction.

(*b*) Indicate where you would locate the four transverse joints required to make up the vertical member from the two plates.

106 For the support column shown in the figure and the materials quoted in question, give

(*a*) the details of the appropriate longitudinal joint preparation used to complete the hollow section,

(*b*) the number and sequence of deposition of the runs required to complete the section,

(*c*) any precautions necessary to ensure sound welds.

107 For the support column shown in the figure and the materials quoted in question, give

(*a*) the details of the appropriate joint preparations used to attach the vertical member to the base,

(*b*) the number and sequence of deposition of runs used to complete this joint,

(*c*) any precautions necessary to ensure sound welds.

Questions 105–108. Overall height 5.6 m. Material available for main stem: plate 3.5 m × 1.2 m × 25 mm, plate 2.0 m × 1.2 m × 25 mm.
Note. Gas shielded metal arc welding to be used for all except tack welds.

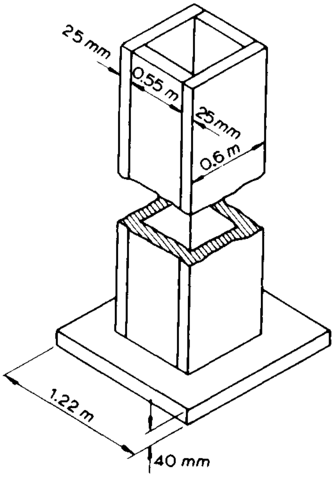

108 Discuss the main factors affecting the desirability of making up a special welding fixture for use when making the longitudinal welds in the vertical member of the support column shown in the figure.

109 (*a*) Give *two* reasons why it is necessary to be particularly careful of plant insulation when using HF current.
   (*b*) Why may it be dangerous to weld with an exposed arc near to a tank filled with non-flammable degreasing liquid?

110 (*a*) Give the principle reason why carbon dioxide gas is used for the gas-shielded metal-arc welding of mild steel.
   (*b*) Give *one* specific difficulty to be overcome when using carbon dioxide as a shielding gas for arc welding mild steel.

111 With the aid of an outline sectional sketch show the relative positions and proportions of (*a*) the electrode, (*b*) the collet, (*c*) the gas entry and (*d*) the nozzle, in a typical gas-shielded tungsten-arc welding torch head.

112 On a simple outline sketch locate and name the function of any *three* parts essential to the operation of the welding head of a fully automatic gas-shielded metal-arc welding plant.

113 Give any *three* possible differences between the respective preparation and deposition techniques for making a flat single V butt weld in 12 mm thick material by (*a*) gas-shielded tungsten-arc welding and (*b*) gas-shielded metal-arc welding.

Other welding processes

1 (*a*) Explain *four* principles involved in the production of a weld.
   (*b*) State *four* variables which influence the production of friction welds.

2 The electroslag process is used for a wide range of welded work.
   (*a*) Outline the principles of this process.
   (*b*) Sketch a sectional view of the weld area during welding showing all the essential parts.
   (*c*) Explain how the joint may be set up to retain its shape and minimize transverse shrinkage.
   (*d*) State *three* reasons why run-on and run-off plates are required.
   (*e*) Describe a suitable method of testing the weld metal to assess its toughness.
   (*f*) Describe a heat treatment process that can be used to improve the toughness of the completed welded joint.

3 (*a*) Explain how heat is produced in the resistance spot welding process.

(*b*) The figure shows a time, pressure, current diagram which illustrates the sequence of operations for the formation of a resistance spot weld.

(1) Calculate the percentage of the total time during which the current flows to form the weld.

(2) Calculate the total heat energy generated in joules using the formula $J = I^2Rt$, where

I = current in amperes,
R = 0.001 ohm resistance,
t = time in seconds,
J = heat energy.

4 (*a*) Explain *four* principles involved in the production of a friction weld.

(*b*) State *four* variables which influence the production of friction welds.

5 For each of three of the following give a description of (1) the principles of operation, (2) a typical application:

(*a*) submerged arc welding.
(*b*) electric resistance spot welding,
(*c*) electroslag welding,
(*d*) arc stud welding.

6 With the aid of sketches show how resistance projection welding differs from resistance spot welding.

Question 3

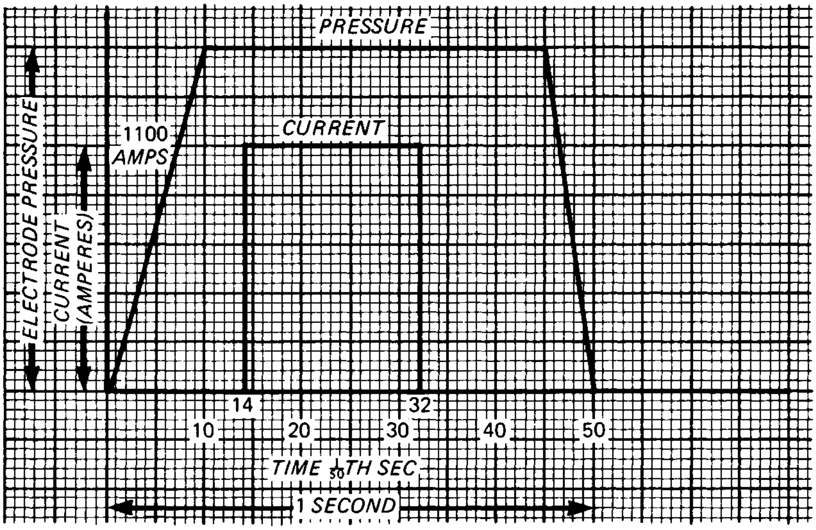

7 Describe briefly the principles of operation, and give *one* suitable industrial application, for *three* of the following welding processes:
  (*a*) electroslag,
  (*b*) electric resistance flash butt,
  (*c*) electron beam,
  (*d*) friction.

8 The figure shows a fixture arrangement for the automatic arc welding of copper floats using the gas-shielded tungsten-arc process.
  (*a*) State the influence the thermal conductivity of the metal would have on weld penetration.
  (*b*) State *three* factors other than thermal conductivity which would influence weld profile.

9 (*a*) List the following welding processes in their order of usefulness for welding steel components 50 mm thick, in the vertical position:
    (1) arc welding with coated electrodes,
    (2) electron beam welding,
    (3) electroslag welding,
    (4) laser beam welding.
  (*b*) Give a brief outline of friction welding.

10 (*a*) Explain why a flux is not required for resistance spot welding.
  (*b*) Give *three* possible causes of defective resistance spot welds.

11 The submerged arc welding process is used for a wide range of welded work.
  (*a*) Outline the principles of this process.
  (*b*) Explain why run-on and run-off plates are used.
  (*c*) State the purpose of vacuum recovery.

Question 8

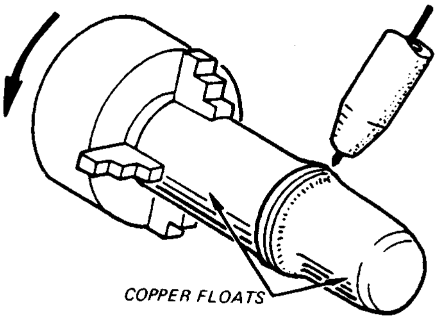

(*d*) State *four* precautions to be taken in order to avoid porosity of the welded joint.

(*e*) State *four* advantages this process has over the manual metal-arc process.

Cutting processes

1 (*a*) State *two* methods used for gas cutting bevels on plate edges in preparation for welding.

(*b*) Describe, with the aid of a sectional sketch indicating the gas paths, the operation of an oxy-fuel gas cutting blowpipe.

2 (*a*) Make labelled sketches to show the essential difference between the nozzle assemblies used with

(1) acetylene,
(2) propane,

for the oxy-fuel gas cutting of low-carbon steel.

(*b*) After manual oxy-fuel gas cutting 50 mm thick low-carbon steel, it is required to cut 6 mm thick low-carbon steel plate. State the adjustment which will need to be made for this operation to be efficiently carried out.

3 Explain the action which produces the severance of the metal when oxy-fuel gas cutting low-carbon steel and state the difference between the basic principles of gas cutting and flame gouging.

4 (*a*) Give three operations which are essential during the assembly of a high-pressure oxy-acetylene cutting plant to ensure safe and effective service.

(*b*) Show by means of labelled sketches the angles of slope and tilt of the electrode when making a straight cut in 8 mm thick low-carbon steel plate in the flat position by the air-arc process.

5 (*a*) What is meant by 'oxygen lance cutting'? Give *two* examples of its use.

(*b*) Describe with the aid of sketches the technique required when cutting low-carbon steel-plate in the flat position by the use of each of the following arc cutting processes:

(1) air-arc,
(2) oxygen arc.

6 Unsuitable variations during oxy-fuel gas cutting of steel may lead to faults along the cut face. Show, by means of labelled sketches of the cut face, the faults which would be caused by either:

(1) the speed of travel being too fast, or

(2) the nozzle being too high above the work surface.

7 (*a*) Explain how the correct size of cutting nozzle to be used in the oxy-fuel gas cutting process should be determined.

(*b*) Discuss the influence on the oxy-fuel gas cutting operation of any two alloying elements which may be present in the steel.

(*c*) Describe how cast iron can be cut by the oxy-fuel gas cutting process.

8 (*a*) Explain the action which causes the severance of the metal when using the oxy-fuel gas cutting process for the cutting of austenitic stainless steel.

(*b*) What is meant by stack cutting? List *two* advantages of the use of stack cutting.

9 (*a*) Give *two* advantages in *each* case which may be obtained when oxy-fuel gas cutting:

(1) manually,
(2) by machine.

(*b*) Describe, with the aid of a sketch, a suitable arc cutting process for the cutting of ferrous and non-ferrous metals.

10 (*a*) What is arc plasma?
By means of a simple labelled outline diagram, show one method of using arc plasma for cutting purposes.

(*b*) Explain the differences between

(1) a single port nozzle,
(2) a multi-port nozzle as used with plasma arc cutting.

(*c*) Show, by means of diagrams, the finished cut that is produced by each of these nozzles after cutting 4.5 mm thick austenitic stainless steel.

11 The following thermal cutting equipment is available for use:
oxy-acetylene manual cutting equipment,
arc plasma profile cutting machine,
oxy-acetylene cutting equipment complete with lance,
oxy-acetylene cutting equipment complete with dispenser unit.
From the cutting equipment available select, with reasons, which cutting process should be used for cutting each of the following materials to the shape or form specified. Do not use the same process for cutting both materials.

(1) 200 mm diameter discs to be cut from 20 mm thick austenitic stainless steel plate.
(2) The removal of excess material from a newly cast grey iron casting.

(3) Outline the principles of the thermal cutting process used.
(4) State *two* aspects of safety that are particularly appropriate to each process used.

12 State why the arc plasma process would be used in preference to oxygen and fuel gas cutting of austenitic stainless steel.

13 (*a*) Why is it necessary to pre-heat medium- and high-carbon steels before cutting by the oxy-fuel gas cutting process?
(*b*) State the factors that prevent the use of natural gas as an oxy-fuel cutting gas.

14 (*a*) Discuss the health hazards that can be encountered when cutting galvanized steel sheet.
(*b*) State the safety precautions that should be adopted to ensure that workshop personnel can be protected against these hazards.

15 (*a*) Describe two suitable thermal cutting processes which could be used to cut 1 metre diameter blanks from 12 mm thick austenitic steel plate.
(*b*) Compare the *two* methods selected in (*a*) with regard to:
(1) cost,
(2) speed of cutting,
(3) distortion,
(4) treatment after cutting.

16 (*a*) List the following cutting process in sequence with regard to best quality of cut face for the thermal cutting of austenitic stainless steel 38 mm thick in the flat position:
(1) oxygen-arc,
(2) air-arc,
(3) oxy-fuel-gas powder,
(4) arc-plasma.
(*b*) Explain, with the aid of a sketch, how the heat is produced for cutting by the plasma cutting process.

17 (*a*) Explain briefly the principles of *laser cutting*.
(*b*) List the advantages of laser cutting over conventional methods of cutting such as oxy-fuel gas and oxy-arc cutting.

Oxy-acetylene

1 (*a*) State *four* problems associated with the welding of copper.
(*b*) State what type of flame setting should be used and how it is attained.
(*c*) Describe in detail the welding procedure.

2 (*a*) Unsuitable variations during oxy-fuel gas cutting of steel may lead to faults along the cut face. Show, by means of labelled sketches of the cut face, the faults which would be caused by *either*
  (1) the speed of travel being too fast, *or*
  (2) the nozzle being too high above the work surface.
(*b*) State *two* methods which may be used for back gouging the root, ready for a sealing run, on the reverse side of welded butt joints in low-carbon steel.

3 Describe briefly the fundamental differences between the following: (*a*) the technique for the bronze welding of cast iron, and (*b*) the technique for the fusion of welding of cast iron.

4 (*a*) Describe, with the aid of sketches, the all-position rightward technique to be used when making a butt joint in low-carbon steel pipe 100 mm diameter by 5 mm wall thickness, the pipe axis to be in the fixed vertical position throughout.
(*b*) Compare the respective advantages and limitations of the leftward and the rightward techniques when used for the butt welding of low-carbon steel plate, 5 mm thick, in the flat position.

5 (*a*) Describe briefly, with the aid of a sectional sketch and indicating the gas paths, the mode of operation of *either*
  (1) a non-injector type welding blowpipe, *or*
  (2) a single-stage gas pressure regulator.
(*b*) What precautions are necessary during the assembly of a high-pressure oxy-acetylene cutting plant in order to ensure safe and effective operation?
(*c*) Explain why the specified discharge rate of a dissolved acetylene cylinder must not be exceeded when in use.

6 (*a*) Describe in detail, with the aid of sectional sketches, *four* defects which may occur during the oxy-acetylene welding of low-carbon steel, stating in *each* case its cause and explaining how it may be avoided.

7 Describe the technique necessary to form a tee-fillet joint in 1.6 mm commercially pure aluminium sheet by the flame brazing process.

State the type of filler wire and flame setting needed. You may use a sketch to illustrate your answer.

8 (*a*) Name three impurities found in acetylene gas immediately after generation.
(*b*) Explain how acetylene gas may be tested for purity.

9 (*a*) Summarize the problems encountered in the oxy-acetylene repair welding of each of the following cast materials: (1) zinc base die cast alloy, (2) magnesium alloy.
(*b*) Describe the preparation, welding technique and post-weld treatment necessary for each.

10 (*a*) When would a carburizing flame be used for joining metals?
(*b*) What purpose is served by the use of such a flame adjustment?

11 Describe, with the aid of a simple outline sketch, one example of a repair welding operation where *studding* could be employed with advantage.
(*a*) Name two advantages of using a high silicon content in cast iron, or in cast iron filler rods for welding purposes.
(*b*) Give the approximate percentage of silicon contained in a 'super-silicon' cast iron filler rod.

12 (*a*) What is meant by the term 'hard-facing'?
(*b*) Name two types of filler rod that may be used for hard-facing operations.

13 (*a*) Name the fuel gas generally used for underwater flame cutting operations.
(*b*) State one good reason why this gas is used.

14 (*a*) Explain, in detail, four safety measures that should be taken before carrying out welding repairs on a 150 litre (30 gallon) low-carbon steel petrol tank.
(*b*) Describe the hazards encountered when welding galvanized metals. State the precautions to be taken in the interests of health and safety.

15 A 6 mm thick magnesium alloy casting is cracked for a length of 150 mm and is to be repaired by oxy-acetylene welding.
(*a*) Outline the preparation which may be required before welding.
(*b*) Give *two* methods of indicating the correct pre-heat temperature.
(*c*) Describe a suitable method of stress relieving after welding.
(*d*) Explain how the flux residue should be removed.

16 A cobalt-based, hard-facing alloy is required to be deposited on to a low-carbon steel component.
(*a*) State how the component could be prepared.
(*b*) Name the type of flame setting required for a single layer deposit and give *two* reasons for your choice.

(*c*) State *two* precautions to be taken to avoid defects.

17 What is the included angle of preparation necessary for single-vee butt welded joints to be made by the oxy-acetylene process using (1) the leftward technique, and (2) the rightward technique?

18 (*a*) What is meant by 'all-position rightward welding'?
(*b*) Give two important advantages which may be obtained by the use of this technique for the welding of pipe joints.

19 (*a*) Explain briefly how (1) carburizing and (2) oxidizing oxy-acetylene flame settings are obtained.
(*b*) Give one important application of each flame setting.

20 What is the purpose of (*a*) portable cylinder couplers in oxy-acetylene welding, (*b*) a brushless d.c. generator for metal-arc welding?

21 (*a*) Make a labelled section-sketch to show the arrangement of the nozzle assembly used for progressive flame gouging.
(*b*) State two methods used for gas-cutting bevels on plate edges in preparation for welding.

22 Name one possible cause of each of the following, when oxy-acetylene welding butt joints in mild steel: (*a*) excessive penetration, (*b*) incomplete penetration, (*c*) adhesion.

23 (*a*) What is the difference between a backfire and a flashback?
(*b*) Name five possible causes of backfiring when using a gas welding blow-pipe. In each case explain how backfiring could have been avoided.
(*c*) State two difficulties which may be experienced in the efficient operation of a gas pressure regulator.
(*d*) State the purpose of each part of the high-pressure oxy-acetylene welding system.

24 Sketch the joint preparation and set-up, and state the nozzle size (in cubic feet or litres per hour) and filler rod size to be used, for making: (1) butt welds in the vertical position in each of the following thickness of mild steel: 3.2 mm ($\frac{1}{8}$ in.) and 5 mm ($\frac{3}{16}$ in.); (2) a tee fillet-weld in the horizontal–vertical position in 5 mm ($\frac{3}{16}$ in.) thick mild steel plate.
Describe in detail, with the aid of sketches, the fusion welding of a butt joint in cast iron 10 mm ($\frac{3}{8}$ in.) thick.

25 Explain what is meant by gas velocity and outline how this is controlled in welding practice.

26 Describe, with the aid of a sectional sketch indicating the gas paths, the operation of an oxy-fuel gas cutting blowpipe.

27 (*a*) Sketch the joint preparation and set-up and state the nozzle and filler rod sizes to be used for making:
   (1) a butt weld in the vertical position in 3 mm thick low-carbon steel,
   (2) a close-square tee fillet weld in the horizontal–vertical position in 5 mm thick low-carbon steel,
   (3) a fusion welded butt joint in the flat position in 10 mm thick cast iron.
(*b*) Describe in detail, with the aid of sketches, the procedure and the technique required for making any *one* of the joints in part (*a*) above.

28 Outline one method which may be used during the oxy-acetylene welding of vertical butt joints, in 4 mm thick low-carbon steel, to ensure freedom from weld defects when
(*a*) starting the weld at the beginning of the joint,
(*b*) restarting the weld at a stop-point along the joint.

29 State *two* safety precautions which should be taken for *each* of the following:
(*a*) storage or use of dissolved acetylene cylinders,
(*b*) using gas pressure regulators,
(*c*) oxy-acetylene cutting operations in a confined space,
(*d*) using non-injector type gas welding blowpipes.

30 (*a*) Explain the action which produces the severance of the metal when oxy-fuel gas cutting low-carbon steel and state the difference between the basic principles of gas cutting and flame gouging.
(*b*) Describe, with the aid of a sectional sketch indicating the gas paths, the operation of an oxy-fuel gas cutting blowpipe.

31 Explain what is meant by capillary attraction and describe how this affects the making of certain brazed joints.

32 The figure shows a cast iron wheel with two broken spokes. With the aid of a sketch, indicate how partial pre-heating may be used to minimize the risk of cracking.

33 (*a*) Describe with the aid of sketches the technique for making a lap fillet weld in the horizontal–vertical position by the oxy-acetylene process.
(*b*) Compare the respective advantages and limitations of the leftward and the Lindewelding techniques when butt welding mild steel pipes.

Question 33

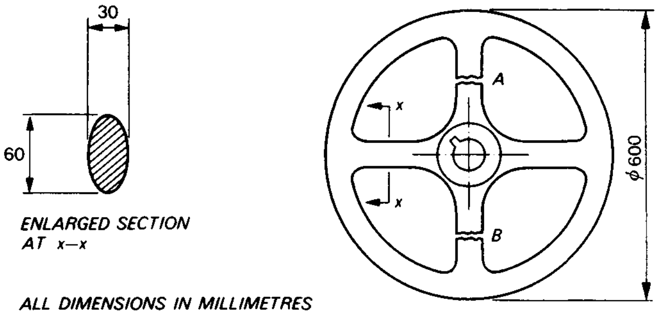

Multiple choice

Note: The following are examples of the multiple choice type of question but may not be representative of the entire scope of the examination either in content or difficulty.

1 Because of the possibility of explosions, acetylene line fittings should *not* be made from
(*a*) steel
(*b*) copper
(*c*) aluminium
(*d*) cast iron.

2 One reason why low-carbon steel may be successfully welded by oxy-acetylene without the use of a flux, is that the oxide
(*a*) is under the surface
(*b*) has a higher melting point than the parent metal
(*c*) has a lower melting point than the parent metal
(*d*) melts at the same temperature as the parent metal.

3 An undesirable property of an aluminium flux residue is that it
  (*a*) is corrosive
  (*b*) obstructs the vision of the molten pool
  (*c*) decreases fluidity
  (*d*) requires great heat to melt it.

4 When a low-alloy steel has a hard and brittle structure it may be rendered soft and malleable by
  (*a*) recrystallization
  (*b*) cold working
  (*c*) lowering its temperature
  (*d*) hot quenching.

5 What happens to the mechanical properties of steel if the carbon content is increased to 0.5%?
  (*a*) The material becomes softer.
  (*b*) Malleability is increased.
  (*c*) The tensile strength is increased.
  (*d*) Ductility is increased.

6 The main reason for pre-heating medium- and high-carbon steels before cutting by the oxy-fuel gas technique is to
  (*a*) improve the quality of cut
  (*b*) increase the cutting speed
  (*c*) refine the grain structure
  (*d*) prevent hardening and cracking.

7 Which one of the following factors restricts the use of town gas as an oxy-fuel cutting gas?
  (*a*) Its low calorific value.
  (*b*) Its tendency to cause rapid melting.
  (*c*) Its unsuitability for cutting plates less than 12 mm thick.
  (*d*) Its relatively high cost.

8 A suitable filler wire for brazing pure aluminium would consist of:
  (*a*) aluminium bronze
  (*b*) aluminium alloy containing 10/13% silicon
  (*c*) aluminium alloy containing 5% magnesium
  (*d*) pure aluminium.

9 Columnar growth takes place when a metal is
  (*a*) cold
  (*b*) losing heat
  (*c*) being heated
  (*d*) being rolled.

10 Difficulty may be encountered when welding aluminium because
  (*a*) the weld metal expands during solidification

(*b*) its coefficient of expansion is low compared to steel
(*c*) no colour change takes place to indicate its melting points
(*d*) its thermal conductivity is low compared to steel.

11 One purpose of a microscopic examination of a weld is to establish the
(*a*) strength of the weld
(*b*) number of alloying elements
(*c*) grain size
(*d*) number of runs used.

12 Which one of the following components is employed to control amperage in an a.c. arc-welding circuit?
(*a*) rheostat
(*b*) choke
(*c*) voltmeter
(*d*) resistor.

13 When carrying out welds in low-carbon steel, using the carbon dioxide welding process, one purpose of the inductance control is to reduce
(*a*) porosity
(*b*) penetration
(*c*) undercut
(*d*) spatter.

14 One purpose of a reactor (choke) when manual metal arc welding is to
(*a*) change alternating current to direct current
(*b*) allow the correct amperage to be selected
(*c*) allow the desired arc voltage to be selected
(*d*) enable the correct polarity to be chosen.

15 Which shielding gas is generally recommended when butt-welding 6 mm nickel alloy sheet by the metal-arc gas-shielded process?
(*a*) Argon
(*b*) CO_2
(*c*) Hydrogen
(*d*) Nitrogen.

16 When TIG welding using a.c. output, which one of the following is essential in the circuit to stabilize the arc?
(*a*) A surge injector
(*b*) An open circuit voltage of 100 volts
(*c*) A flow meter
(*d*) An amperage regulator.

17 In manual metal-arc welding the flux coating to give deep penetration characteristics would contain

(*a*) iron oxide
(*b*) manganese
(*c*) cellulose
(*d*) calcium carbonate.

18 Which element is used as a deoxidant in copper filler rods?

(*a*) Aluminium
(*b*) Tin
(*c*) Sulphur
(*d*) Phosphorus.

19 An oxygen cylinder regulator being used in a flame-cutting supply may freeze up if the

(*a*) gas withdrawal rate is exceeded
(*b*) cylinder content is too low
(*c*) cylinder is on its side
(*d*) needle valve on the regulator is not fully open.

20 To test a component part for a vibrational loading, a suitable mechanical test would be

(*a*) impact
(*b*) tensile
(*c*) compressive
(*d*) fatigue.

21 The principal advantage of arc-on-gas welding is that it

(*a*) allows controlled penetration of initial bead
(*b*) requires less operator skill
(*c*) entirely eliminates distortion
(*d*) improves surface finish.

22 One reason why a grey cast iron casting should be slowly cooled after welding is to keep it

(*a*) soft
(*b*) spheroidal
(*c*) hard
(*d*) brittle.

23 An iron casting has a crack in it. Before oxy-acetylene fusion welding it may be necessary to drill the ends of the crack. One reason for this is to

(*a*) balance out any shrinkage stresses
(*b*) stop the crack from spreading
(*c*) prevent the ends of the crack from being carburized
(*d*) prevent grain growth.

24 Which one of the following metals may require the studding

techniques to be used when being repaired by manual metal arc welding?

(*a*) Low-carbon steel
(*b*) Aluminium
(*c*) Nickel
(*d*) Cast iron.

25 During the deposition of a manual metal-arc electrode, a certain percentage of the core wire is lost. This is due to

(*a*) voltage drop across the arc
(*b*) short arc length
(*c*) spatter
(*d*) excessive build-up.

26 Which one of the following can be welded by d.c. using the tungsten arc gas-shielded process?

(*a*) Copper
(*b*) Commercial pure aluminium
(*c*) Silicon–aluminium
(*d*) Magnesium alloys.

27 Backing bars for manual metal-arc welding of low-carbon steel should be made from

(*a*) copper
(*b*) low-carbon steel
(*c*) tool steel
(*d*) cast iron.

28 Peening may be carried out when manual metal-arc welding cast iron in order to

(*a*) reduce the effects of contraction
(*b*) make the bond more firmly adhering
(*c*) refine the grain structure
(*d*) speed up the welding.

29 The fillet welds on the support brackets in the figure should have 5 mm leg length with 50 mm intermittent welds as shown. The symbol at *A* to communicate this information should be

(a)

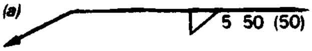

(b)

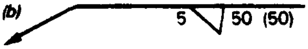

(c)

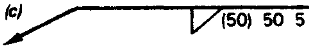

(d)

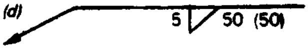

30 What is the volume of deposited metal in a fillet weld indicated by the symbol in the figure neglecting reinforcement?
(*a*) 14000 mm^3
(*b*) 14900 mm^3
(*c*) 15000 mm^3
(*d*) 15 100 mm^3.

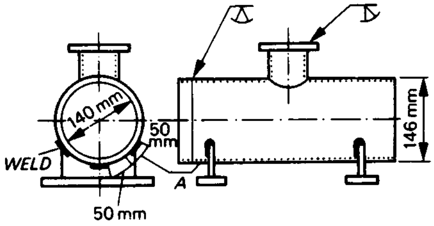

Question 29. The figure shows a component fabricated from stabilized austenitic stainless steel sheet 3 mm thick.

Question 30

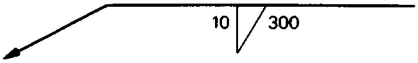

Answer key

1 – *b*
2 – *c*
3 – *a*
4 – *a*
5 – *c*
6 – *d*
7 – *a*
8 – *b*
9 – *b*
10 – *c*
11 – *c*
12 – *b*
13 – *b*
14 – *b*
15 – *a*
16 – *a*
17 – *c*
18 – *d*
19 – *a*
20 – *d*
21 – *a*
22 – *a*
23 – *b*
24 – *d*
25 – *c*
26 – *a*
27 – *b*
28 – *a*
29 – *d*
30 – *c*

Welding engineering craft studies

All dimensions are in millimetres

1 Spot welds are to be made by the tungsten arc gas-shielded welding process.
(*a*) Make a sectional sketch through one of these spot welds.

(*b*) State *one* advantage that this process has over the resistance spot welding process for making this type of weld.

(*c*) Name the additional equipment that would be necessary for making the spot welds using standard tungsten arc gas-shielded welding equipment.

2 (*a*) How is the size of a fillet weld with normal penetration determined in accordance to BS 499 in
  (1) a convex fillet weld,
  (2) a concave fillet weld?

(*b*) Make a sectional sketch of a mitre fillet weld.

3 State *five* factors that would influence the pre-heating temperature to be used for a welded steel fabrication.

4 (*a*) Give *two* reasons why hydrogen-controlled electrodes are preferred for manual metal-arc welding restrained joints in low-alloy steel.

(*b*) Explain the influence of the cellulose in the coating of an electrode on
  (1) voltage,
  (2) penetration.

5 A circular low-carbon steel plate of 750 mm in diameter and 10 mm thick has to be fitted and welded into a deck plate.

(*a*) State why hot cracking is likely to occur in this particular type of assembly.

(*b*) Outline *either* a suitable weld sequence *or* a change in the form of the plate insert that could be used to avoid the occurrence of hot cracking.

6 A worn press die has to be built up using the oxy-acetylene flame powder spraying process.

(*a*) From the table select the powder to be used for this repair.

(*b*) List *three* advantages that powder spraying has over arc welding for this type of repair.

(*c*) Name the type of flame setting to be used.

Powder no.	Resistance to:		
	Abrasion	Impact	Machinability
1	Fair	Excellent	Very good
2	Very good	Excellent	Very good
3	Excellent	Poor	Grind only

7 From the information given below, list in the correct order the welding sequence that should be carried out when friction welding two 20 mm diameter low-carbon steel bars.
- (*a*) Place parts lightly in contact.
- (*b*) Load machine.
- (*c*) Apply axial force.
- (*d*) Apply upset force.
- (*e*) Rotate chuck and close gap.
- (*f*) Release upset force.
- (*g*) Arrest chuck movement.
- (*h*) Remove specimen.

8 Steel cylindrical tanks 5 m long and rolled to 2 m internal diameter from 50 mm thick low-carbon steel plate are to be welded, using the electroslag welding process for the longitudinal seams.
- (*a*) With the aid of a sketch outline how the weld area is protected from atmospheric contamination.
- (*b*) (1) State *three* advantages that the electroslag welding process has over arc welding processes for welding process for the longitudinal seams.
  - (2) Give *two* limitations of the electroslag welding process.
- (*c*) (1) What is the purpose of run-on and run-off plates when electroslag welding?
  - (2) After electroslag welding has commenced small additions of flux must continue to be added to the weld pool. Why is this?
- (*d*) If it was considered that the cylinder would be working under conditions that may cause stress corrosion cracking:
  - (1) State what is meant by the term stress corrosion cracking,
  - (2) State *two* precautions that could be taken to reduce the occurrence of failure from this form of attack.

9 The figure shows a low-carbon steel nut to be resistance projection welded to 2 mm thick low-carbon steel sheet.
- (*a*) Explain the principles involved in making this welded connection using the resistance welding process.
- (*b*) Make a sectional sketch through *A–A* of the welded assembly shown in the figure.
- (*c*) State *two* advantages of resistance welding over manual metal-arc welding for making this welded connection.
- (*d*) List *three* defects that may be found when resistance *spot*

welding, and in each case state the probable cause of the defect.

10 As part of a welding sequence, stiffeners require to be welded to the aluminium alloy plate shown in the figure.

(*a*) Select a suitable process for welding the stiffeners.

(*b*) Sketch and label the essential components of the welding circuit for the process selected.

Question 9

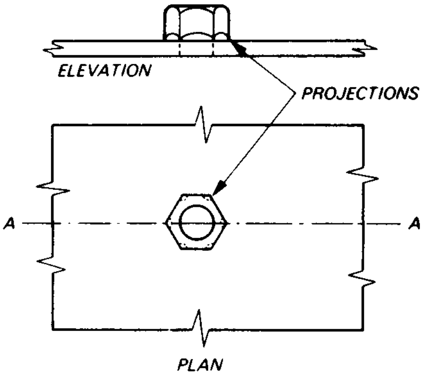
ELEVATION
PROJECTIONS
A
A
PLAN

Question 10

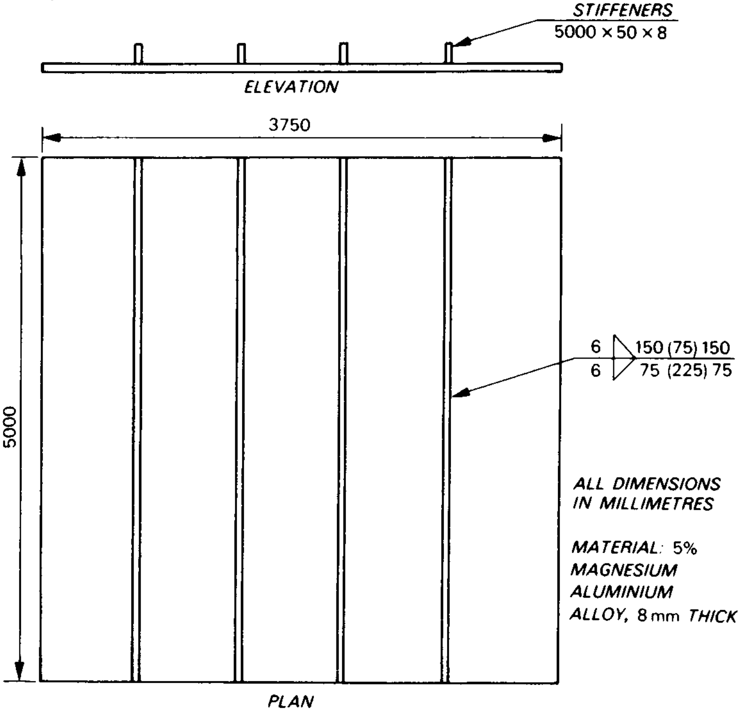
STIFFENERS
5000 × 50 × 8
ELEVATION
3750
5000
6 150 (75) 150
6 75 (225) 75
ALL DIMENSIONS IN MILLIMETRES
MATERIAL: 5% MAGNESIUM ALUMINIUM ALLOY, 8 mm THICK
PLAN

(*c*) Using a graph to illustrate your answer, give a brief outline of the output characteristics of the power source needed for the process used.

(*d*) What information is indicated by the weld symbol shown in the drawing?

11 Alloy steel of the following % composition has to be manual metal-arc welded.

Carbon (C)	Chromium (Cr)	Molybdenum (Mo)	Manganese (Mn)	Silicon (Si)	Nickel (Ni)	Vanadium (V)
0.16	1.5	0.5	0.8	0.4	0.0	0.0

Remainder: iron with acceptable limits of impurities.

(*a*) State *three* advantages of using alloying elements in steel.

(*b*) (1) Using the carbon equivalent formula given below determine the carbon equivalent of the alloy steel outlined above.

Carbon equivalent (%) =

$$C\% + \frac{Mn\%}{20} + \frac{Ni\%}{15} + \frac{Cr\% + Mo\% + V\%}{10}$$

(2) Indicate how the information produced may be used to determine the welding procedure used.

(*c*) Explain why pre-heating this alloy may reduce the occurrence of underbead cracking.

(*d*) Welds made on low-alloy steels are generally heat treated by normalizing.

(1) What is meant by normalizing?

(2) State *three* advantages produced by normalizing welded fabrications.

12 State *five* factors that will influence dilution during fusion welding.

13 (*a*) Give *three* reasons for pre-heating air-hardenable steels before or during oxy-fuel gas cutting.

(*b*) Give the main constituents in an air-hardenable steel.

14 Briefly explain how a friction weld is produced.

15 (*a*) State *three* advantages of using fully automatic welding processes in preference to manual welding processes.

(*b*) Under what circumstances may manual welding be preferred to fully automatic welding?

16 (*a*) Explain how heat is produced during cutting when using the oxygen lance.

(*b*) State *two* safety precautions that should be taken when using the oxygen lance for cutting.

17 List *five* problems which may be encountered when fusion welding aluminium.

18 Explain briefly why single-pass welds made by the arc plasma process may have the form shown in the macrograph in the figure.

19 One thousand circular containers as shown in the figure have been fabricated from low-carbon steel by using the metal-arc gas-shielded welding process with carbon dioxide as shielding gas.

   (*a*) Name the main type of stress, in *each* case, that weld *A* and *B* would be subjected to during hydraulic testing.

   (*b*) During testing some welds were found to contain porosity. State *three* probable causes of this defect.

20 (*a*) The figure shows a time and current graph making an electric resistance spot weld. The following two parts of this question refer to this diagram.

   (1) Calculate the weld time as a percentage of the welding cycle.

   (2) Explain why it is generally necessary for the forging time to be longer than the squeeze time.

   (*b*) A resistance spot welding machine was set up for welding low-carbon steel. If austenitic stainless steel of the same

Question 18

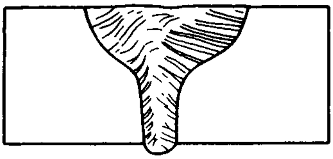

Question 19

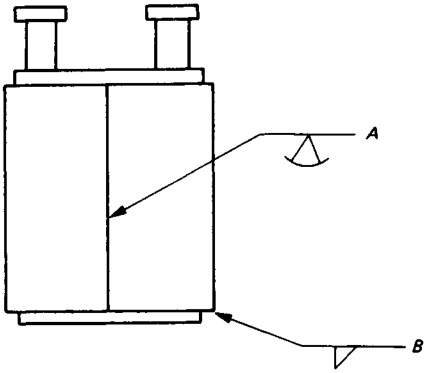

thickness is to be welded, what alterations would require to be made to

(1) the welding current,

(2) the welding time?

Explain why these alterations are necessary.

(*c*) Spot welding of sheet metal may also be carried out by fusion welding.

(1) Select a fusion welding process and describe how a spot weld is made using this process.

(2) What additional equipment would be necessary for making spot welds using standard equipment?

21 The figure shows a low-alloy steel rotary shear blade, used for trimming steel plate. The cutting edge has become worn due to severe abrasion during service and is to be repaired by using the manual metal-arc welding process.

(*a*) The general composition of three electrodes is shown below. Select the most suitable electrode for use in the building up of the cutting edge, and give a reason for your selection.

(1) Austenitic stainless steel.

(2) Medium-carbon low-alloy steel.

(3) Pure nickel.

(*b*) Outline the welding procedure that should be used to carry out this repair.

Question 20

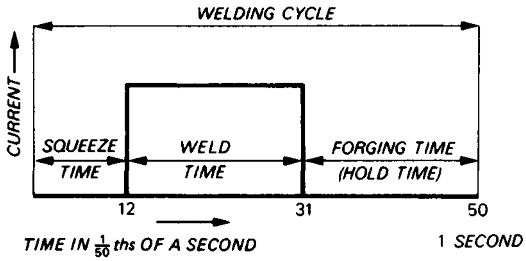

Question 21

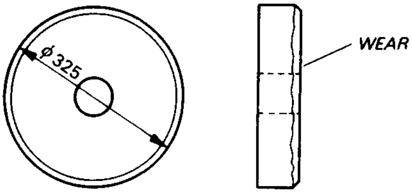

(*c*) Explain the influence that the dilution of the weld metal by the parent metal would have on the weld's mechanical properties.

(*d*) Under what circumstances would the oxy-acetylene welding process be preferred to arc-welding processes for hard surfacing?

22 The figure shows a low-carbon steel shaft to be fabricated by welding using the submerged arc welding process.

(*a*) With the aid of a sketch, explain how the weld area is protected from atmospheric contamination.

(*b*) State *three* advantages of using a backing strip for this joint.

(*c*) Sketch in detail the weld preparation that would be used for the butt weld shown at *A*.

(*d*) Why is the filler used in submerged arc welding copper coated?

(*e*) When submerged arc welding, explain why it is an advantage to use a multipower source which has one electrode using alternating current and the other direct current.

23 50,000 shear connectors are to be welded to the low-carbon steel box girder shown in the figure, using the drawn arc stud welding process.

(*a*) In what respect does the purpose-built power source used for this type of equipment differ from the power source used for metal-arc welding?

(*b*) Sketch in section, and label the main parts of, a ceramic ferrule used for stud welding.

(*c*) If the shear connectors are to be positioned at 100 mm centres, explain how this could be best achieved.

Question 22

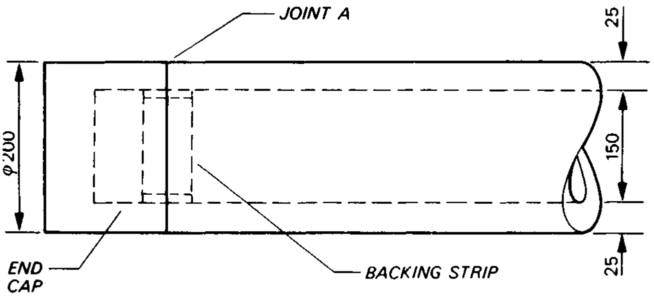

ALL DIMENSIONS IN mm

(*d*) (1) Describe a method of testing the stud welds.
(2) Welds having the section shown in *d*(2) below failed during testing due to the defects shown. State *four* possible causes of these defects.

24 Power souces used for electron beam welding may be rated as 30 kV.
(*a*) Explain the term 30 kV.
(*b*) Give *two* reasons why electron beam welding is generally carried out in a vacuum.

25 (*a*) State *four* of the variables involved in the production of friction welds.
(*b*) State the range or temperature necessary for the production of friction welds in low-carbon steel.

26 Sketch in section, and label the parts of, the head of an arc plasma torch for welding.

27 (*a*) State *three* advantages of introducing iron powder into the flame when oxy-acetylene powder cutting.
(*b*) State *two* hazards that the operator should guard against when oxy-acetylene powder cutting.

28 Low-carbon steel plate 25 mm thick has to be surfaced by the submerged arc welding process using stabilized austenitic stainless steel filler. List *five* variables that would influence the degree of dilution found in the weld.

29 Explain why it is recommended that materials containing sulphur should be removed from the weld area before welding nickel and nickel alloys.

Question 23

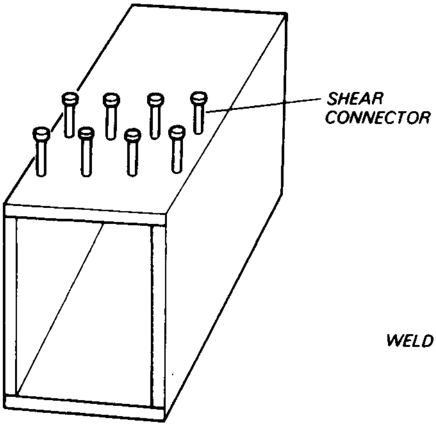

Question 23. (*d*) (2)

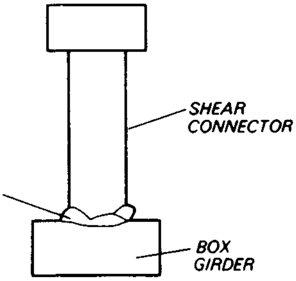

30 State *five* factors that would need to be considered before deciding the pre-heating temperature to be used on low-alloy steel plate.

31 (*a*) With the aid of a graph, explain why a constant potential power source is often preferred for automatic arc welding processes.

(*b*) List *four* variables, under the control of the welder, that can influence weld quality when using the submerged arc welding process.

(*c*) Explain why the quality of submerged arc welds made in low-alloy steels is particularly high compared with the quality of welds made by other arc welding processes.

(*d*) Inspection authorities may specify that welds made by the submerged arc welding process should be heat treated. Outline a suitable heat treatment.

32 The grey cast iron angle plate shown in the figure has to be repaired by manual metal-arc welding and then machined flush. Rutile-covered low-carbon steel and nickel alloy electrodes are available for making the weld.

(*a*) Select the most suitable electrode to carry out this repair.

(*b*) State *two* advantages and *one* disadvantage of the electrode selected.

33 The aluminium–magnesium alloy components shown in the figure are to be welded using the metal-arc gas-shielded welding process.

(*a*) Explain why an argon/carbon dioxide gas mixture would be unsuitable as the gas shield to be used for welding the components.

Question 32

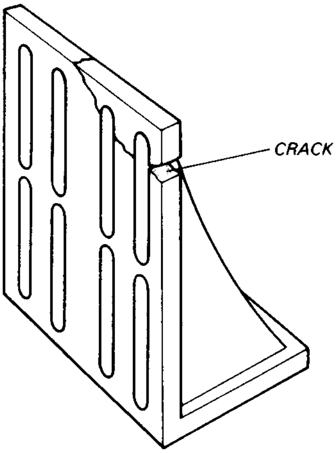

(*b*) State *three* factors that would influence the weld profile.

(*c*) Explain why the oxy-acetylene fusion welding process would be unsuitable for welding the components.

(*d*) Outline the influence that welding would have on the metal's mechanical properties and grain structure.

(*e*) If 1000 components are to be fabricated, calculate the total length of welding carried out to complete the contract. Each joint is to be made by means of a single-run weld deposit.

34 (*a*) Calculate the volume of metal deposited in the welded joint shown in the figure if the length of the joint is 10 m, and the total weld reinforcement is taken as one quarter of the total volume of the joint gap.

(*b*) Outline the influence of *each* of the following factors on the cost of welded fabrication:

(1) current density,

(2) joint set-up.

(*c*) Outline how *four* factors in the welding procedure can influence welding costs.

(*d*) Sketch an example of a non-load bearing fillet weld, indicating by an arrow the direction of the applied load on the component when in service.

Question 33

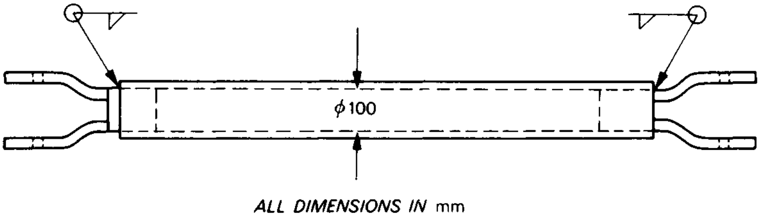

Question 34

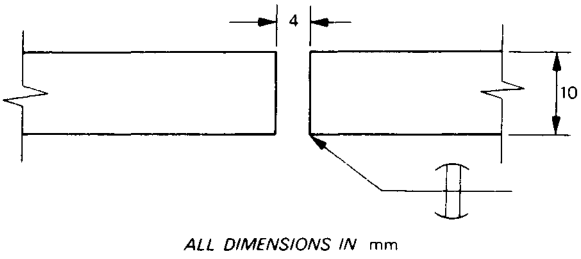

35 A fabricated tank is to be manufactured from stabilized austenitic stainless steel. The weld preparation and holes are to be thermally cut using the arc plasma process.

- (*a*) Explain why the arc plasma process would be used in preference to oxygen and fuel gas for cutting this material.
- (*b*) Sketch in detail the design of the welded butt joint shown at *A* in the figure.
- (*c*) State *two* problems which may be encountered during the arc welding of this joint.
- (*d*) Explain why the depth of penetration would be greater in welds made in austenitic stainless steel, compared with similar welds made in low-carbon steel, assuming the energy input to be the same.
- (*e*) Explain, with the use of a graph, why a drooping characteristic power source should be used when manual metal-arc welding, rather than a constant potential power source.

36 Component parts for an oil rig, having the form shown in the figure, are to be fabricated and joined by welding.

- (*a*) Calculate the carbon equivalent of the steel given the following:

  Material composition: Carbon, 0.22%; Silicon, 0.5%;

Question 35

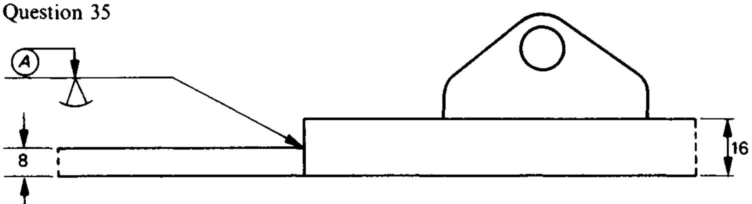

ALL DIMENSIONS IN mm

Question 36

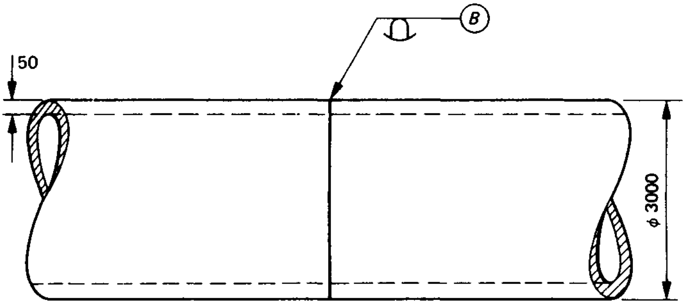

ALL DIMENSIONS IN mm

Manganese, 1.5%; Niobium, 0.1%; Vanadium, 0.1%; Sulphur, 0.05%; Phosphorus, 0.05%; remainder iron.

Carbon equivalent =

$$C\% + \frac{Mn\%}{6} + \frac{Cr\% + Mo\% + V\%}{5} + \frac{Ni\% + Cu\%}{15}$$

(*b*) Sketch in detail the butt joint preparation shown at *B*.

(*c*) Outline a welding procedure that could be used to make the joint.

(*d*) Since this component will be subjected to fatigue conditions during service, state *three* precautions which should be taken when welding has been completed.

37 (*a*) The figure shows a cruciform welded joint. Using the formula $l = 1.4\, t$ calculate the throat size (t) of a fillet weld if the leg length (l) is 21 mm.

(*b*) Show by means of a labelled sectional sketch a fillet weld leg length of 21 mm.

(*c*) A load is applied in the direction of the arrows indicated in the figure. State whether the welds are load bearing or non-load bearing.

(*d*) The material shown in the figure has a relatively high carbon equivalent and pre-heating temperatures have not been maintained during welding so that underbead cracking has resulted. Explain how the underbead cracking would take place.

(*e*) What is meant by the term 'fatigue fracture'?

(*f*) State *four* factors under the control of the welder which may adversely affect the fatigue life of a single-vee butt-welded joint.

Question 37

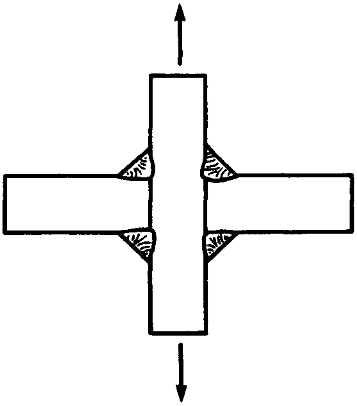

(*g*) State *three* precautions that could be taken to reduce the occurrence of a brittle fracture failure in a welded joint.

38 The figure shows an austenitic stainless steel hopper which has to be fabricated from 5 mm thick sheet.

(*a*) State *four* methods which may be used to ensure a low heat input procedure for arc welding the hopper.

(*b*) (1) State the value of the included angle *X* for the joint shown at *B*.

(2) State *two* reasons why the included angle *X* should be increased if welding has to be carried out.

(*c*) Four different welding procedures could be used for making the butt-welded joints shown at *A* and *B*. State *one* advantage and *one* disadvantage for *each* of the following welding procedures.

(1) Completely weld the joints using the tungsten arc gas-shielded process.

(2) Completely weld the joints using the gas-shielded metal-arc process.

(3) Weld the first run using the tungsten arc gas-shielded process and complete the weld using the manual metal-arc process.

Question 38

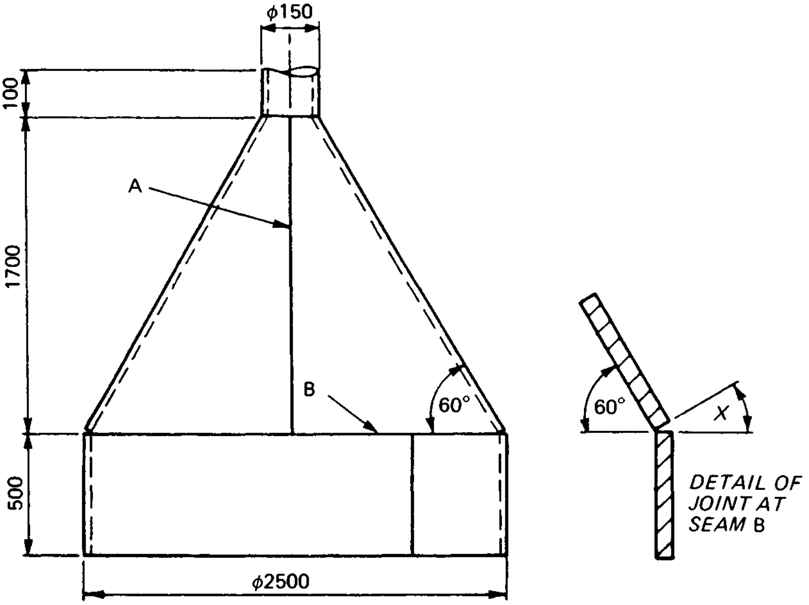

(4) Completely weld the joints using the manual metal-arc process.

(*d*) (1) Explain why distortion is likely to be a problem when welding this material.

(2) Describe a method that could be used to control the distortion when welding joint *A*.

39 The figure shows a fabrication which is to be made from 25 mm thick low-carbon steel using the gas-shielded metal-arc welding process with flux cored electrode wire.

(*a*) Give *three* reasons why flux cored electrode wire may be preferred for welding this fabrication in preference to solid wire.

(*b*) State *two* advantages of using a gas shield in addition to the flux cored electrode wire when using this process.

(*c*) A root crack was observed when welding the single bevel butt weld joint shown in (*b*). State *three* possible causes for its formation.

(*d*) State *three* factors which may cause porosity during welding, assuming the joint surfaces were clean before welding was carried out.

(*e*) Calculate the area of the side plate shown as *A* on (*a*).

40 (*a*) State *one* reason why

(1) some welding specifications permit a small amount of slag inclusion in a welded joint,

(2) it would not be realistic for a specification to state that a welded joint must be free from defects,

(3) welding specifications do not permit welded joints containing cracks to be accepted.

(*b*) Show, with the aid of a sketch, how the design size of a butt weld is measured.

41 (*a*) The figure shows butt and fillet joints which are to be joined by manual metal-arc welding. Explain why the welding specification may state that for the same heat energy input the butt-welded joint would not require to be pre-heated, although the fillet welded joint should be.

(*b*) The hardness of the gas cut edge shown in '*X*' in (*b*) was found to be unacceptably high. State *three* changes in the oxy-fuel gas cutting procedure that could have been taken to reduce the hardness value.

42 The figure shows a low-carbon steel oil storage tank which is to be fabricated from 18 mm thick low-carbon steel plate using the submerged arc welding process.

Question 39

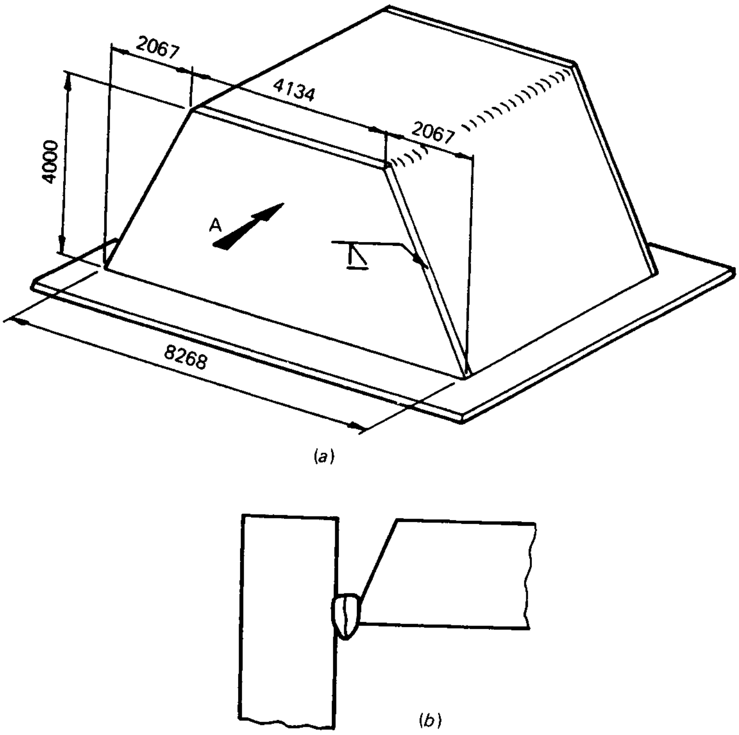
2067
4134
2067
4000
8268
A

(a)

(b)

Question 41

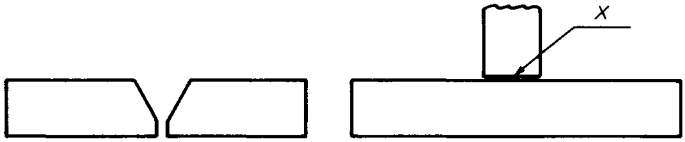

(a)

(b)

Question 42

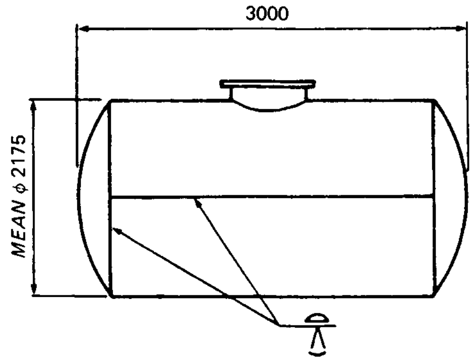
3000
MEAN ϕ 2175

(*a*) Calculate the length of plate required to be gas cut and rolled to form the cylindrical section of the tank.
(*b*) Explain why run-on and run-off plates are used when welding the longitudinal seam.
(*c*) Explain why welds made using the submerged arc welding process usually have a low level of defects.
(*d*) If the weld deposit was required to have good impact properties state whether it would be made with either a large number of runs or as few a number of runs as possible.
(*e*) State *four* variables under the control of the welder which will influence the weld profile.
(*f*) Name *two* other mechanized welding processes that could be used for welding the joints and state in *each* case *one* reason for using the process in preference to the submerged arc process.
(*g*) State in *each* case *one* advantage which may be obtained by using
  (1) alternating current for submerged arc welding,
  (2) direct current for submerged arc welding.

43 In relation to the plasma arc welding process
(*a*) reproduce the sketch shown in the figure and identify the
  (1) negative pole within the circuit,
  (2) shielding gas ports.
(*b*) State whether the parent metal forms part of the welding circuit when a non-transferred arc is used for welding.
(*c*) State a temperature within the temperature range of the arc.
(*d*) Sketch a transverse section through the weld made by means of a transferred arc on low-carbon steel 10 mm thick.

44 (*a*) List *four* variables under the control of the operator which will influence the quality of a resistance spot weld.

Question 43

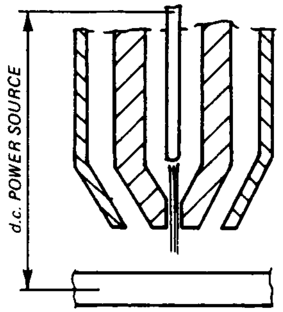

(*b*) Low-carbon steel sheet has been resistance spot welded. State whether the welding current would need to be increased if this process was to be used for welding aluminium sheet of the same thickness.

(*c*) Describe the operational sequence used for making resistance butt welds in short lengths of 50 mm diameter low-carbon steel pipes of wall thickness 6 mm.

(*d*) (1) A number of steel studs having the form shown in the figure are to be welded to low-carbon steel plate. Describe, with the aid of a sketch, the operational sequence used to make a drawn arc stud welded joint.

(2) The figure shows a ceramic ferrule used in the drawn arc stud welding. Explain the purpose of the serrations.

45 Each term used in column *A* in the table is directly related to *one* of the terms used in the column *B*. Pair *each* of the terms listed in column *A* with its appropriate term in column *B*.

Column *A*	Column *B*
d.c. negative polarity	Force
Nugget	Oxy-fuel cutting
Kerf	Welding standard
API	Resistance spot welding
newton	Hard surfacing

46 A box section column 750 mm square × 6500 mm long is to be fabricated from 40 mm thick low-carbon steel plate.

(*a*) Select and describe a suitable automatic welding process.

(*b*) Sketch the corner joint and show, if necessary, any edge preparation.

Question 44. (*d*) (1)

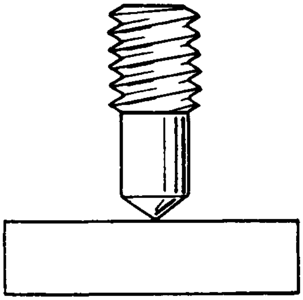

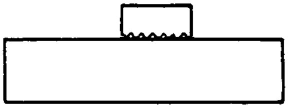

(*c*) Give *three* advantages of using automatic welding as opposed to manual metal-arc.

47 A spherical vessel, 3.6 m diameter, is fabricated from 75 mm thick low-carbon steel, the welding of the circumferential seams being carried out by the electroslag process.

(*a*) Sketch the macrostructure of the plate and weld area.

(*b*) Describe a suitable method of heat treatment to be carried out on completion of the welding.

(*c*) State *three* advantages to be obtained by the form of heat treatment chosen.

(*d*) State *two* safety precautions to be taken when pressure testing the vessel.

48 (*a*) Explain why the effect of electric shock is much less when using a flat characteristic instead of a drooping characteristic power source for welding.

(*b*) Name *two* gases which may be produced during arc welding operations which are harmful to the health of the welder.

49 (*a*) State *two* factors that should be considered before selecting a filler/electrode for a hard surfacing application.

(*b*) State *three* precautions which may be taken by the welder to control the level of dilution when surfacing using an arc welding process.

50 (*a*) The figure shows three views of a platform bracket. Using graph isometric paper, make an isometric sketch of the bracket in the direction of arrow *A*. Omit all hidden lines.

(*b*) Parts of the bracket are to be assembled and welded using the electrical resistance spot welding process.

(1) Outline the principles of electrical resistance spot welding.

(2) State why the stiffener shown at *B* could not be resistance spot welded for the design shown.

(3) Sketch a modification to the stiffener that would enable spot welding to be used.

(4) State *three* advantages of using the resistance spot welding process for welding the bracket in preference to a fusion welded process.

51 State *five* variables, other than electrode diameter, that will influence the production of an acceptable quality weld when using the mechanized flux shielded visible arc process.

52 Name *each* of the welding or thermal cutting processes illustrated in the figure. Identify your answer by stating the appropriate letter and name the process concerned.

53 (*a*) State *one* safety precaution that must be carried out.

   (1) when oxy-acetylene cutting inside a steel container,
   (2) when manual metal-arc welding galvanized steel,
   (3) when gas-shielded metal-arc welding inside a container.

(*b*) State why gas-shielded tungsten arc welding should not be carried out inside a container that has been degreased using carbon tetrachloride until all the degreasing compound and vapours have been removed.

54 The figure shows an aluminium–magnesium alloy safety rail which is to be welded by the gas-shielded metal-arc welding process using 1 mm diameter filler wire.

(*a*) State a suitable welding voltage and amperage setting for making the fillet joint shown at *A*, in the horizontal–vertical position.

Question 50

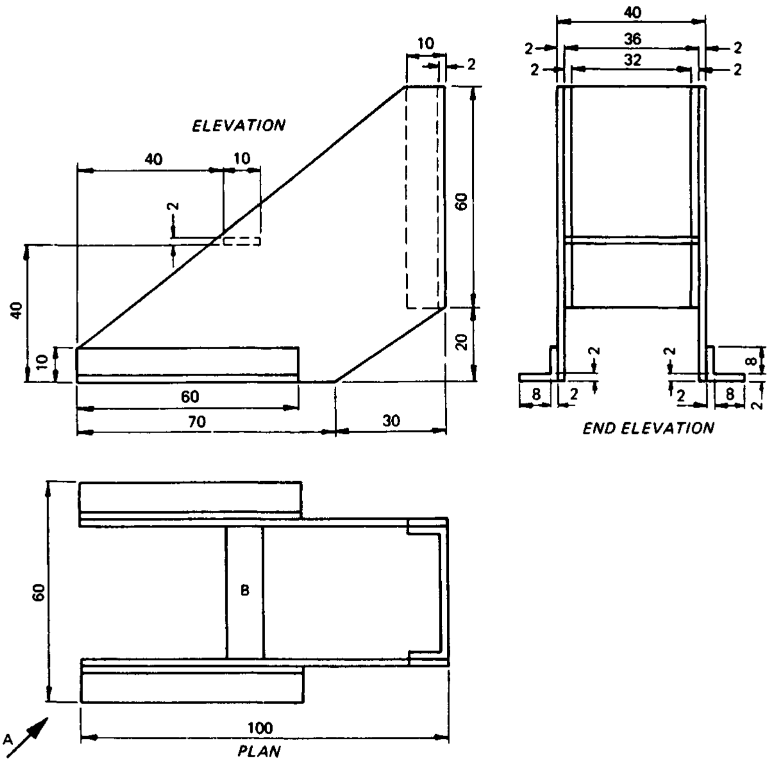

ELEVATION
40
10
2
40
10
60
70
30
10
2
60
20
40
36
32
2
2
2
2
2
2
8
2
2
8
8
2
END ELEVATION
60
B
100
A
PLAN

(*b*) Name the mode of metal transfer which would be used for welding *each* of the following:
  (1) the fillet joint shown at *A* in the horizontal position,
  (2) the fillet joint shown at *B* in the vertical position.

(*c*) When welding aluminium with this process, electrode wire may be supplied by pull feed from the torch or by push feed from a separate wire feed unit. State *two* advantages and *two* disadvantages of *each* method of wire supply.

Question 52

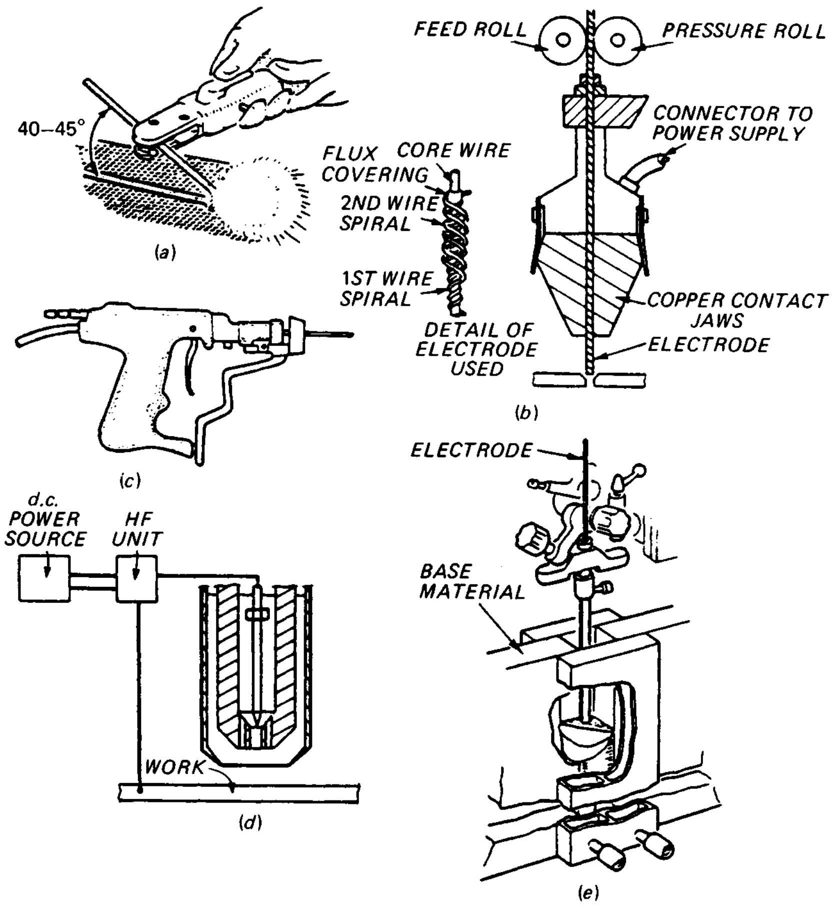
40–45°
(a)
FLUX COVERING
CORE WIRE
2ND WIRE SPIRAL
1ST WIRE SPIRAL
DETAIL OF ELECTRODE USED
FEED ROLL
PRESSURE ROLL
CONNECTOR TO POWER SUPPLY
COPPER CONTACT JAWS
ELECTRODE
(b)
(c)
d.c. POWER SOURCE
HF UNIT
WORK
(d)
ELECTRODE
BASE MATERIAL
(e)

(*d*) List *four* problems which may be encountered when fusion welding aluminium.

(*e*) State *two* reasons why an 80% argon, 20% carbon dioxide gas mixture should *not* be used as the shielding gas for welding aluminium.

(*f*) Give *two* reasons for using helium as a shielding gas for welding aluminium with the gas-shielded tungsten arc welding process.

Question 54

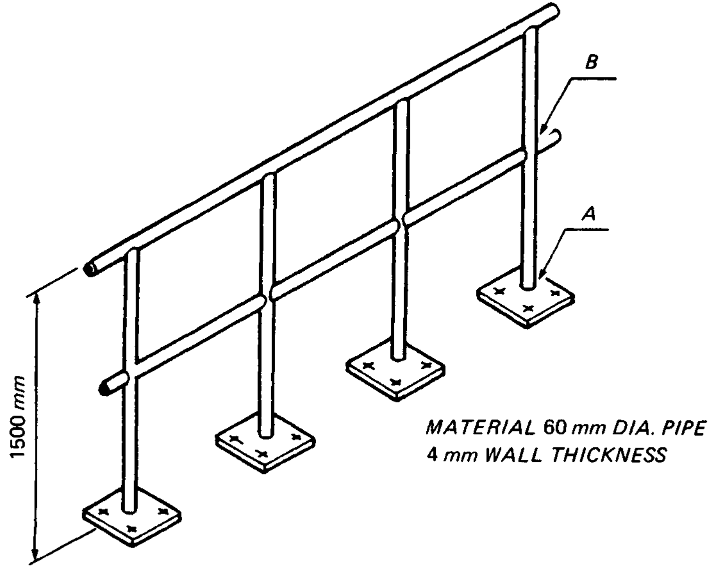
B
A
1500 mm
MATERIAL 60 mm DIA. PIPE
4 mm WALL THICKNESS

Question 55

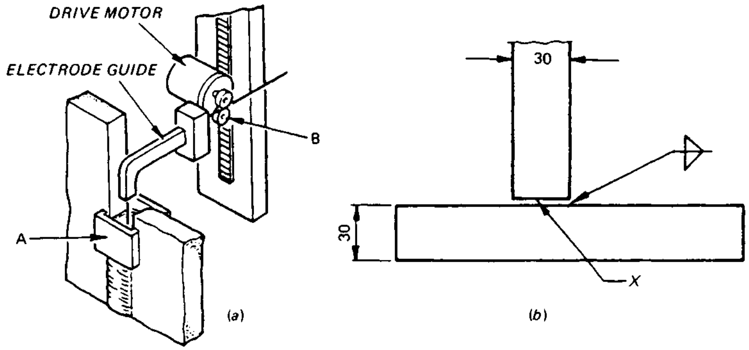
DRIVE MOTOR
ELECTRODE GUIDE
B
A
30
30
X

(*a*) (*b*)

55 The figure (*a*) shows a section of a joint being welded using the electroslag process.

(*a*) Name the parts identified in the sketch at points *A* and *B*.
(*b*) State why run-off plates are required to complete the weld.
(*c*) State whether or not an arc is maintained throughout electroslag welding.
(*d*) Explain why small additions of flux may require to be added during electroslag welding.
(*e*) Give *two* reasons why electroslag welds may require to be heat treated after welding.
(*f*) State *three* basic differences that exist between conventional electroslag welding and consumable guide welding.
(*g*) Sketch in section a cruciform joint made by electroslag welding.
(*h*) Figure (*b*) shows a joint in low-alloy steel set-up for fillet welding using the manual metal-arc welding process. Explain why the edge shown at *X* was oxy-acetylene flame cut, then machined before welding.

56 In relation to welded fabrications:

(*a*) state what is meant by the term brittle fracture,
(*b*) state *five* conditions in which an otherwise ductile steel may behave in a brittle manner after welding,
(*c*) name a test that may be used to assess the notch ductility of a steel,
(*d*) state *five* factors to be considered when selecting a manual metal-arc electrode to be used for a given application.

57 The figure shows a beam to be fabricated from low-alloy steel, using the gas-shielded metal-arc process with flux cored electrode wire.

(*a*) Using the carbon equivalent formula given, calculate the carbon equivalent of the steel beam when its composition is as shown in the table.
(*b*) Explain why steel components with a high carbon equivalent should be
  (1) pre-heated before welding is carried out,
  (2) immediately post-heated after welding.
(*c*) The figure shows that the flanges of the beam have been stiffened using gusset plates. Explain why the corners have been removed.
(*d*) Name the main stress acting on the underside of flange *B* if a load is applied to the beam in the direction of arrow *A*.

Composition	Carbon	Manganese	Silicon	Niobium	Vanadium	Remainder/ Iron
As a percentage	0.22	1.2	0.5	0.1	0.2	Sulphur and phosphorus in acceptable quantities

$$CE = C\% + \frac{Mn\%}{6} + \frac{Cr\% + Mn\% + V\%}{5} + \frac{Ni\% + Cu\%}{15}$$

(*e*) State *four* methods that may be used to reduce undercut at the toes of the welds joining the flanges to the web.

(*f*) Outline, with the aid of a sectional sketch, the effect of using too low a voltage setting for depositing a single run fillet weld by this process.

Question 57

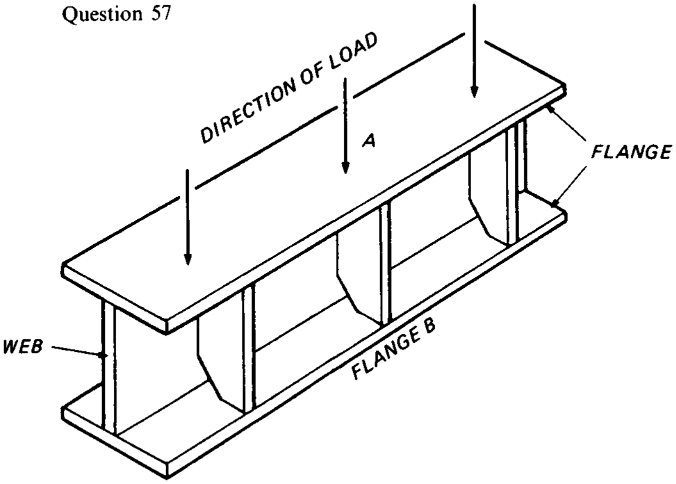

Fabrication and welding engineering (technical grade)

1 (*a*) Give the main feature which distinguishes between cylinders for (1) combustible gas, and (2) non-combustible gas.
  (*b*) Explain the function of a 30 A fuse.
  (*c*) Why is a cylinder normally rolled to a slight overlap of the butting edges in a plate binding rolls?
  (*d*) Show, on an appropriate cross-section of rolled steel member, the correct positioning of a tapered washer.
  (*e*) State *four* factors which would ensure good quality soldering.
  (*f*) What information may be obtained by an examination of an etched cross-section of a fillet weld?
  (*g*) A large triangular plate of 30 mm thickness is lying on a workshop floor. Describe how the plate could be marked off to find its centre of area, for lifting purposes.
  (*h*) Give *five* reasons for using false wired edge on thin sheet metal.

2 (*a*) Explain the essential differences between ductility and malleability.
  (*b*) What is meant by (1) hardness and (2) toughness? Give practical examples of the use of these *two* properties in different materials.

3 Compare and contrast the following cutting processes: (1) oxy-fuel gas cutting, (2) guillotine cutting, with respect to:
  (*a*) cost,
  (*b*) limitations of cut,
  (*c*) type of material to be cut.

4 Discuss the factors which must be considered when deciding whether to form a given metal by hot or cold processes.

5 Describe, with the aid of sketches, *three* methods of stiffening and strengthening metal platework without the attachment of additional members.

6 When producing a riveted watertight butt joint, what are the important features to look for:
  (*a*) when setting up the work,
  (*b*) after riveting.

7 Explain the meaning of the following terms when used in connection with the control of distortion in welding:
  (*a*) preset,
  (*b*) restraint,

- (*c*) back-step,
- (*d*) intermittent,
- (*c*) staggered.

8 It is suggested that the following requirements are essential for safe and adequate electrical circuits in a fabrication workshop:

- (1) colour coding,
- (2) notices,
- (3) maintenance,
- (4) notification of faults,
- (5) earthing,
- (6) load protection,
- (7) insulation,
- (8) protection,
- (9) emergency switches.

Write explanatory notes on any *five* of the above.

9 With the aid of sketches of cross-section show *each* of the following weld defects:

- (*a*) undercut on a fillet weld,
- (*b*) undercut on a butt weld,
- (*c*) lack of root penetration in a fillet weld,
- (*d*) lack of penetration in an open square-edge butt joint,
- (*e*) excessive root penetration in a single bevel butt joint.

10 One hundred welded fabrications each weighing 48 kgf are at a temperature of 25 °C. They are placed in a furnace and heated to 500 °C for the purpose of stress relieving. Calculate the quantity of heat to be supplied so that the fabrications reach the required temperature. Specific heat capacity of the welded fabrication is 0.5 kJ kg °C. Furnace efficiency is 20%.

Index

abrasion surfacing 403
  MMA 58–63
  oxy-acetylene 310–16
a.c. wave balance control 162–3
acetylene
  absorption by acetone; dissolved 257
  flash back arrestor 271–3
  gauge 261–6
  generator medium pressure 260
  impurities in 260
  safety precautions 259
  storage 258–9
  test for leaks 258
air plasma cutting 344 *et seq.*
alloy steel, welding of 78 *et seq.*
all position rightward welding 277–8
aluminium and its alloys
  brazing 320
  filler rods: oxy-acetylene 303–5; manual metal arc 51–5
  filler wire: TIG 175; MIG 175
  manual metal arc welding 51–3
  MIG 125–6
  oxy-acetylene welding 303–5
  TIG welding 175
aluminium bronze
  oxy-acetylene welding 297–8
  TIG welding 183
arc
  alternating current 2
  bare wire 2
  blow 4–5
  characteristics of 4–5
  energy 3, 80
  forces 4
  metal 1–5
  MIG 94–5
  non-transferred 190–1
  plasma 189–94
  shielded 2, 92–96
  striking 34–6
  TIG 152–3
  transference of metal in
    MMA 9–10
    MIG 94–5
    TIG 148–9
  transferred 190–1
  voltage drop 2, 94
arc-air cutting 339
argon 110, 112
  flow meter 163–5
  gas regulator 163–4
  MIG welding 92, 115
  TIG welding 148
asphyxiating fumes 459–50
austenitic stainless steel welding
  MMA 64–70
  MIG 115
  oxy-acetylene 308–9
  TIG 179–80
automatic welding
  MIG 132–4
  spot MIG 137
  seam 208
  TIG 185–7

back chipping 121, 123
back purging 123
backfire 270
backing bar 41, 51, 121, 181, 222
backing strip 222–3
bell and spigot joint 77
bell hole welding 73
beryllium, welding of 189
block welding 282
blow pipe, oxyacetylene 266 *et seq.*
  care of 270–1
  cleaning 271
  cutting 326–9
  high pressure 266–7
  low pressure 266–7
  nozzles 266–7
  overheating 270
  see also torch
boom ram 132–4
brazing alloys and fluxes 395 *et seq.*
braze welding 288 *et seq.*
brazing 288 *et seq.*, 317 *et seq.*
  aluminium 320
bronze welding, oxy-acetylene 288 *et seq.*
  aluminium bronze 297–8
  galvanized iron 293
  gilding metal 297
  malleable cast iron 292
  tin bronze 297
burnback control 121
butt welding
  CO_2 129–30
  MMA 40–2
  oxy-acetylene 274–6, 277–8

capacitor discharge stud welding 245–6
capillarity 319
carbon, carbon-manganese and low alloy steels, welding of 78 *et seq.*, 173
carbon equivalent 78–9
cast iron
  bronze welding 290–2

  MMA welding 53–5
  MIG welding 127, 145
  oxy-acetylene welding 286–8
characteristics of power units
  MIG 95–8
  TIG 155–6
chromium and nickel steels
  MMA welding 70–1, 78–9
  MIG welding 123–5
  TIG welding 179–81
City and Guilds of London Institute questions
  MMA 467–80
  MIG, TIG 480–97
  other processes 497–500
  cutting 500–2
  oxy-acetylene 502–7
  multiple choice 507–12
  craft studies 512–35
  technician grade 536–7
clad steels, MMA welding of 71–2
CO_2 welding 129–31
coffer dam 254
consumable guide welding 225–6
contact tube
  MIG 99, 105, 120–1
  submerged arc 217
copper
  alloy filler rods 184
  bronze welding of 294–5
  oxy-acetylene 298–301
  MMA 57
  MIG welding 127–8
  TIG welding 181–5
copper–aluminium (aluminium bronze), TIG welding 183
copper silicon alloys
  MIG welding 127–8
  TIG welding 183
crater filling control 157
cross wire welding 211
cupro-nickels
  MIG welding 127
  TIG welding 183
current control by thyristors SCR's 158, 363
cutter (gas)
  oxy-acetylene 326–9
  oxy-propane 329
cutting process 326 *et seq.*
  air plasma 341 *et seq.*
  alloy steel 340–6
  arc air 339
  carbon arc 340
  cast iron 335
  effect on steel 332–3
  machines 334
  oxy-acetylene 326–30
  oxy-arc 338–9
  oxy-hydrogen 329
  oxy-natural gas 329
  oxy-propane 329
  plasma 340–6
  powder injection 349–51
  stack cutting 335
  safety precautions 354–5
  thermic lance 333
  under water 353–4
cutting tools (tipping)
  MMA 62
  oxy-acetylene 313
cycle-arc welding 137
cylinders (gas)
  MIG 110–13
  oxy-acetylene 255–8
  TIG 148

deep penetration electrode 10
degreasing 125
degrees of freedom (robot) 226–7
deionized water 346–7
deoxidizers 114
diffusible hydrogen in weld metal 9–10
dip transfer 95
dissolved acetylene 257–8
double operator oxy-acetylene welding 279–81
drawn arc stud welding 242–4
drooping characteristics 97–8
droplet, *see* transfer of metal in arc
Duplex stainless steel 66, 123, 445–8
duty cycle 156

eccentricity of electrode core 5–6
electrode
  classification of: British 14–20; American 21–2
  composition of coating 19–20
  deep penetration 10
  drying of 12–4
  efficiency 6–7
  flux cored 117–18
  grinding (TIG) 169–70
  hydrogen controlled 8, 12–13
  hydrogen releasing 8
  iron powder (high efficiency) 8
  manual metal arc welding 1–91
  manufacture of coated 376–8
  metrication (conversion factors) 371
  MIG welding 92–146
  polarity: MIG 93, TIG 148–50
  storage 10–12
  tungsten 168–9
electrode holder
  MIG 105 *et seq.*
  MMA 23
  TIG 165–8
electron beam welding 237–9
electro slag welding (submerged melt) 224–5
  consumable guide 225–6
explosive welding 247–50
  preparation 248

eyes, protection of
  in MMA welding 23–4, 32
  in MIG welding 119
  in oxy-acetylene welding 323–4
  in TIG welding 172

faults in a weld 42
ferritic stainless steel, welding of MMA 65, 66–7
  MIG 114–15
  TIG 179
fillet weld, horizontal–vertical
  MMA 45–7
  MIG 122–3
  oxy-acetylene 283–5
  TIG 179
fire extinguishing 412–16
fixtures 26–7
flame
  adjustments for welding 268–70
  carburizing 269
  cutting 329–30
  gouging 326–8, 336–8
  neutral 269
  oxidizing 269
flash back arrestor 271–2
flash butt welding 213–15
flux basicity 221
flux cored electrode welding 117–18
flux cored self-shielded filler wires 118
flux, metal powder cored 117–18
free flight transfer 95
friction welding 234 *et seq.*
  forms for 235
fumes 449–57
  asphyxiating 449–50
  removal of 451–7
  toxic 450–1

galvanized articles, oxy-acetylene welding 293
gas cutting, *see* cutting processes
gas lens 166
gases 393–402
  identification of cylinders 116
  inert 110
  mixtures 366–402
  table: for MIG welding 115; for TIG welding 148
gas mixer 145–6
gas-shielded metal arc welding 92–146
globular transfer 3–4, 94–5
gouging, flame 336–8
gravity welding 250–1
gusset bend 77

hard and corrosion resistant surfacing
  MMA 58–61
  plasma 195–6
  oxy-acetylene 310–15
  spray fuse 316
  TIG 184–5
head shield 23–4
helium
  in MIG welding 110–11, 115
  in TIG welding 148
high efficiency electrodes 7–8
high frequency spark gap oscillator (TIG) 152–3
horizontal–vertical welding
  MMA 45–7
  MIG 113, 122–3
  oxy-acetylene 283–4
  TIG 179
hoses 266
hot pass or fill 74, 76
hydrogen-controlled electrodes 8, 12–13
  effect on a weld 8–9
  solubility in weld metal 8–9
hyperbaric welding 253–4

ID, definition of 451
ignition of TIG arc 151–4
ions 190–1

jigs 26–7
joints
  butt 40–1
  horizontal–vertical fillet 40–1, 45–7
  lap 40–1
  open corner 40–1

keyhole welding 194

lamellar tearing 82–3
laser beam welding 239–42, 458–66
leftward welding (forward welding) 275
  flat butt 275–6
  HV fillet 275–6, 283–4
  vertical 279–81, 284–5
  overhead 281–2, 284–5
lens, gas for fitting to TIG gas nozzle 166
lenses for eye protection, *see* eyes, protection of
lift start (TIG) 153
Lindewelding 282
liquefied gas
  oxy-acetylene 256
  MIG 146
low alloy steel classifications 83 *et seq.*
low alloy steel, welding of 78–83

magnesium alloys
  table of 178
  TIG welding 176–9
magnetic arc control 134–7
malleable iron, oxy-acetylene welding 288
manganese, welding of steels containing 59, 78
manipulators 27–31
manual metal arc welding
  cast iron 53–5
  copper, bronze and brass 57–8

flat butt 40–2
hard surfacing 58–9
horizontal butt 49
horizontal–vertical fillet 45–6
open corner 42–4
overhead 50
padding 38–9
pipe lines 73–7
safety precautions 32
stainless steel 64–70
steel castings 83
Stellite 62–4
studding 56
tipping tool steel 62
vertical 47–8
martensitic stainless steel, MMA welding of 64–5, 69
MEL, definition of 451
metal inert gas welding 92–146
methyl chloride 125
metrication, electrode 34, 371
mild steel welding
  CO_2 129–31
  MMA 40–50
  MIG 92–123
  oxy-acetylene 274–85
MONEL® alloy, welding of, *see* nickel alloys
monochromatic light 240

neutral flame 269
nickel and its alloys
  MMA welding 70–1
  oxy-acetylene welding 306–7
  TIG welding 180–1
nickel–chromium steel, *see* stainless steel
nickel clad steel
  MMA welding 72
  oxy-acetylene welding 308
nugget 199
nozzle, selection of 270

OES, definition of 451
orbital pipe welding 186
oscillator (HF) for TIG 152–3
overalloying 448
overfill 42
overhead welding
  MMA 50
  MIG 122
  oxy-acetylene 281, 285
oxidizing flame 269
oxy-acetylene flame 269
  adjustment 269
oxy-acetylene welding
  all position rightward 284–5
  aluminium and alloys 301–5
  block 282
  blow pipe 266–8
  brazing 317–22
  brazing welding and bronze welding 288–97
  cast iron 286–8
  copper 298–301
  double operator vertical 279–80
  general precautions 323–5
  hard surfacing 310–15
  horizontal–vertical fillet 283–4
  INCONEL® alloy 306–7
  leftward 275–6
  malleable cast iron 288
  MONEL® alloy 306–7
  nickel and nickel alloys 306–7
  nickel clad steel 308
  overhead 281
  rightward 277–8
  stainless steel 308–10
  Stellite 312–16
  table of filler rods 323
  tanks 324
  vertical 279–80
oxy-arc cutting 338–9
oxygen
  as wetting agent 111–112, 318–19
  compressed 255–6
  cylinders 255–6
  liquid 256–7
  safety precautions in the use of 353
  test for leaks 256
oxygen lance 333–4

padding 38–9
partial rectification 151
pipe lines
  low hydrogen electrode, downhill welding of 405–8
  MMA welding of 73–7
  orbital welding of 186
plasma
  comparison of TIG and plasma arcs 193
  cutting 340–6
  keyhole welding 194
  MIG process 144
  non-transferred arc 191
  surfacing 195–6
  transferred arc 191
  welding 189–94
plastics, manual welding
  HF spark tester 370
  hot air or gas 359
  hot plate 371
  general 366
  guns, *see* torches
  nozzle types 362–3
  practical hints 368
  speed welding 367–8
  rod profiles 361
  tack welding 365
  thin sheets 358

  - torches 359–60
  - types of plastic 356–7, 379–80
- plastics, mechanical welding
  - adhesives 377–8
  - electric fusion 373–4
  - hot plate 372–3
  - tractor 369, 372
  - ultrasonic 374–5
  - vibration (linear) 375–6
- polarity of electrode
  - effect on welding
  - MMA 1, 35
  - MIG 93
  - TIG 147–50
- positioners 27–31
  - powered 229–30
- powder injection cutting 347–51
- powder surfacing 195, 316
- pressure regulator
  - details of 261–6
  - single-stage 261
  - multi-stage 263
  - TIG 163–4
- pressure welding 231
  - cold 231–3
- programmable power source for MIG welding 102, 144, fig. 2. 26(*f*)
- projection welding 211
- propane 329
- protection of eyes and skin when welding 23–4, 32, 172, 323–5
- pulsed current welding
  - MIG 137–40
  - TIG 169–71

- RD, definition of 451
- recovery rate, *see* electrode efficiency
- reducing valve, *see* pressure regulator
- regulator, oxy-acetylene pressure 261–6
- reignition voltage 151–3
- resistance butt welding 212
- restraint in a joint (carbon and alloy steels) 80–1
- rightward welding, all positional 284–5
- robot welding (robotics) 226–30
- roll welding 73

- safety precautions
  - HF in TIG welding 172
  - in use of compressed gases 258–60, 325
  - in use of oxygen 354–5
  - MMA 32
  - MIG 119
  - oxy-acetylene 323–5
  - tanks 324
  - TIG 172
- scratch start 153
- seam tracker 136
- seam welding 208
- SG cast iron, MIG welding of 127, 145
- Shaeffler's diagram 69
- shafts rebuilding, MMA 38–9
- shield, hand and head 23
- shielded arc, principle of 1–3
- short circuit current transfer (MIG), 95
- slope in 157
- slope out (crater fill) 157
- soft start 157
- spatter 5
- spot welding 197–203
  - electrodes 204–6
  - MIG 137
  - TIG 187
  - sequence of welding 203
- spray fuse hard surfacing 316
- spray transfer (free flight) 94
- square wave output unit 141–4, 160–1
- stainless steel, table of steels and electrodes 66–7, 124
  - duplex 65
- stainless steel welding
  - MIG 123–4
  - MMA 64–8
  - oxy-acetylene 308–10
  - TIG 179
- STEL, definition of 000
- stick out (MIG) 120
- stubbing 121
- studding 56
- stud welding
  - arc 242–4
  - capacitor discharge 244–6
- submerged arc welding 216–23
  - automatic 217
  - power unit 220
  - fluxes 221
- Super Duplex stainless steels 445–8
- suppression of dc component (TIG) 154–5
- surface tension 318
- surge unit 153–4
- symbols, welding, British and American 381–9
- synergic pulsed MIG welding 140–4

- tables
  - electric arc welding rods 90–1
  - oxy-acetylene welding rods 323
  - MIG welding wires 115, 124
  - silver brazing alloys and fluxes 417–28
  - TIG welding wires, Al 175, Mg 178, Ni 181, Cu 184
- tack welding 39–40
- tantalum welding of 189
- tanks, precaution and safety in welding 324
- thermic lance, oxygen lance 333–4

thoriated tungsten 168
thermit welding 251–2
thyristor, control of current 158
threshold limit value (TLV) 324
TIG welding
  aluminium alloys 173–6
  automatic 185
  copper and alloys 181–4
  hardfacing 184–5
  magnesium and alloys 176–8
  nickel and alloys 180–1
  orbital 186
  robot welding 187
  spot 187
  stainless steel 179–80
tin bronze, oxy-acetylene welding 297
tipping tool steel 62
titanium
  oxide 20
  welding 188
torch
  MIG 105–9
  oxy-acetylene 266–8
  TIG 165–8
toxic fumes 50–1
transfer of metal in arc welding 3–4, 94–5
triple deoxidized wire 114
tubular electrode
  automated MIG 145
  flux cored 117
  metal powder cored 117–18
  self shielded 118

tungsten arc 147–56
tungsten electrode
  current ratings 169
  grinding 169
  protrusion 170
  thoriated 168
  zirconiated 168
TWA, definition of 451
tungsten inert gas welding 147–89
types of plant 429–41

ultrasonic welding 233
underwater
  cutting 353
  welding 253–4

vertical welding
  MMA 47–8
  MIG 122
  oxy-acetylene 279–80, 284–5
  TIG 176–7
voltage control, MIG 94–99
voltage drop
  MIG 94, 96–9
  MMA 2–3

water injection plasma cutting 347–8
wave balance control 162
wear resistant surfaces, building up of
  MMA 58–64
  oxy-acetylene 310–16
  plasma spray 195–6
  spray fuse 316
  TIG 184–5
weaving 37
weld
  defects 42
  features of a good 42
  profiles 42, 381
  rotation 10–11
  slope 10–11
welding gun (spot)
  MIG 137 & fig. 2.4(*a*)
  TIG 187–8
welders' accessories 23–6
welding positions 10–11
wire feed (MIG), control 99–105

zirconiated tungsten 168

PRICE $7.95
68-1015

1000TH STORE COMMEMORATIVE EDITION • UNABRIDGED

ALLIED RADIO SHACK

MODERN DICTIONARY of ELECTRONICS

Howard W. Sams PUBLICATION

Rudolf F. Graf

OVER 16,000 DEFINITIONS!
Including the latest terms
in the fields of:
- COMMUNICATIONS
- MICROELECTRONICS
- FIBER OPTICS
- RELIABILITY
- COMPUTERS

OVER 450 ILLUSTRATIONS
ABBREVIATIONS
GREEK ALPHABET
SCHEMATIC SYMBOLS
PRONUNCIATION GUIDE
OHM'S LAW FOR AC OR DC
TRIGONOMETRIC RELATIONSHIPS
MATHEMATICAL SYMBOLS

608 PAGES
AUTHORITATIVE • COMPREHENSIVE
COMPLETELY UP-TO-DATE

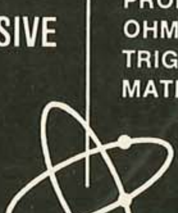

MODERN DICTIONARY *of* ELECTRONICS

Rudolf F. Graf

HOWARD W. SAMS & CO., INC.
THE BOBBS-MERRILL CO., INC.
INDIANAPOLIS • KANSAS CITY • NEW YORK

THIRD EDITION

THIRD PRINTING — 1970

Copyright © 1962, 1963 and 1968 by Howard W. Sams & Co., Inc., Indianapolis, Indiana 46206. Printed in the United States of America.

All rights reserved. Reproduction or use, without express permission, of editorial or pictorial content, in any manner, is prohibited. No patent liability is assumed with respect to the use of the information contained herein.

Library of Congress Catalog Card Number 68-13873

PREFACE

No truly authoritative dictionary can be the work of one person. Rather, it is the culmination of the unceasing efforts of many people. The field of electronics is constantly expanding and so its language and vocabulary must be augmented with newly originated words and phrases if effective communication of thoughts and ideas are to be established. The originators give the words their initial meanings, but the exact definitions change with technological advances and actual usage by others.

The content of this dictionary is thus an analysis of words and their meanings as determined by common usage, and as researched and written by its author and editors. Because the field of electronics is constantly expanding, its language and vocabulary must grow along with it. More than 4000 new terms and additions to existing terms have been added in this third edition. Many of the new terms are related to fields, phenomena, or processes that were little-known or even nonexistent when the first edition was published in 1962.

Every entry in the second edition was reviewed and many were revised for this new book in order to keep pace with current usage. Where more than one definition exists for a term, the different meanings are arranged numerically. This method, however, does not necessarily imply a preferred order of meanings.

The emphasis of this dictionary is on the broad subject of electronics. Within this field it contains over 16,000 definitions of current terms from radio, television, communications, radar, electronic instrumentation, industrial and medical electronics, microelectronics, computers, data processing, lasers, logic, semiconductors, and fiberoptics.

I gratefully acknowledge the permission granted once again by Mr. E. K. Gannett of the IEEE to use definitions and illustrations which originally were prepared by them. My thanks also go to the other organizations and trade sources who have so kindly supplied their material and granted permission to use it.

I should like to take this opportunity to dedicate this book to my children, Jeffrey and Debbie, while at the same time extending gratitude to my lovely wife Bettina, whose untiring efforts and truly invaluable assistance have made it possible for me to prepare this edition of the dictionary.

Samuel Johnson once said, "Dictionaries are like watches. The worst is better than none, and the best cannot expect to be quite true." With that thought in mind, the users of this volume are cordially invited to submit to the publisher comments and suggestions for additions or changes of definitions given here, so that future editions will better reflect the current usage of the language of electronics.

Rudolf F. Graf

HOW TO USE THIS DICTIONARY

This Modern Dictionary of Electronics follows the standards accepted by prominent lexicographers. All terms of more than one word are treated as one word. For example, "bridged-T network" appears between "bridged tap" and "bridge duplex system." Abbreviations are also treated alphabetically; the initials "ARRL" follow the term "arrester" rather than appearing at the beginning of the A's.

For ease in quickly locating a specific term, catchwords for the first and last entries which appear on each page are shown at the top.

Illustrations have been positioned with the terms they depict and are clearly captioned so they can be immediately associated with the proper definition.

Moderate cross-referencing has been used as an aid in locating terms which you might look for in more than one place. For example, when looking up "Esaki diode" you'll be referred to "tunnel diode." However, occasionally you may look for a term and not find it. In such instances, always think of the term in its most logical form; i.e., you will find "acoustic resonator" in the A's and not in the R's. In other words, when looking up the definition for a specific device, such as a "dipole antenna," refer to the modifier "dipole" rather than to "antenna."

A unique feature is the *Pronunciation Guide* beginning after the definitions. This Guide shows syllabic division and pronunciations, based on accepted industry usage, for over 1100 commonly mispronounced words. Because the language is in constant flux, pronunciations and spellings acceptable five years ago may be obsolete today. Witness the evolution of the word "ampere." Originally it was pronounced "AHMpair," the French pronunciation. In this country, accepted usage simplified the term to "AMpeer" (just as "Schmidt" became "Smith"). Today the accepted pronunciation is "AMper."

Since it follows the most authoritative standards of the industry, this dictionary will serve as an excellent guide on spelling, hyphenation, abbreviation, capitalization, etc.

It is hoped you will find Modern Dictionary of Electronics helpful, informative, and satisfactory in every way. Should you care to pass along any comments or suggestions which come to mind as a result of its use, we will be most happy to hear from you.

Howard W. Sams

A

A—1. Abbreviation for angstrom unit, used in expressing wavelength of light. Its length is 10^{-8} centimeter. 2. Chemical symbol for argon, an inert gas used in some electron tubes. 3. Symbol for area of a plane surface.

a—1. Abbreviation for atto (10^{-18}). 2. Abbreviation for ampere.

A− (A-minus or A-negative)—Sometimes called F−. Negative terminal of an A-battery or negative polarity of other sources of filament voltage. Denotes the terminal to which the negative side of the filament-voltage source should be connected.

A+ (A-plus or A-positive)—Sometimes called F+. Positive terminal of an A-battery or positive polarity of other sources of filament voltage. The terminal to which the positive side of the filament voltage source should be connected.

ab—The prefix attached to names of practical electric units to indicate the corresponding unit in the cgs (centimeter grams second) electromagnetic systems—e.g., abampere, abvolt, abcoulomb.

abac—*See* Alignment Chart.

abampere—Centimeter-gram-second electromagnetic unit of current. The current which, when flowing through a wire one centimeter long bent into an arc with a radius of one centimeter, produces a magnetic field intensity of one oersted. One abampere is equal to 10 amperes.

A-battery—Source of energy which heats the filaments of vacuum tubes in battery-operated equipment.

abc—Abbreviation for automatic bass compensation, a circuit used in some equipment to increase the amplitude of the bass notes to make them appear more natural at low volume settings.

abcoulomb—Centimeter-gram-second electromagnetic unit of electrical quantity. The quantity of electricity passing any point in an electrical circuit in one second when the current is one abampere. One abcoulomb is equal to 10 coulombs.

aberration—In lenses a defect that produces inexact focusing. Aberration may also occur in electron optical systems, causing a halo around the light spot.

abfarad—Centimeter-gram-second electromagnetic unit of capacitance. The capacitance of a capacitor when a charge of one abcoulomb produces a difference of potential of one abvolt between its plates. One abfarad is equal to 10^{9} farads.

abhenry—Centimeter-gram-second electromagnetic unit of inductance. The inductance in a circuit in which an electromotive force of one abvolt is induced by a current changing at the rate of one abampere per second. One abhenry is equal to 10^{9} henrys.

abmho—Centimeter-gram-second electromagnetic unit of conductance. A conductor or circuit has a conductance of one abmho when a difference of potential of one abvolt between its terminals will cause a current of one abampere to flow through the conductor. One abmho is equal to 10^{9} mho.

abnormal glow—In a glow tube, a current discharge of such magnitude that the cathode area is entirely surrounded by a glow. A further increase in current results in a rise in its density and a drop in voltage.

abnormal propagation—The phenomenon of unstable or changing atmospheric and/or inospheric conditions acting upon transmitted radio waves. Such waves are prevented from following their normal path through space, causing difficulties and disruptions of communications.

abnormal reflections—*See* Sporadic Reflections.

abohm—Centimeter-gram-second electromagnetic unit of resistance. The resistance of a conductor when, with an unvarying current of one abampere flowing through it, the potential difference between the ends of the conductor is one abvolt. One abohm is equal to 10^{-9} ohm.

abort—To cut short or break off (an action, operation, or procedure) with an aircraft, guided missile, or the like—especially because of equipment failure. An abort may occur at any point from start to countdown or takeoff to the destination. An abort can be caused by human technical or meteorological errors, miscalculation, or malfunctions.

AB power pack—Assembly in a single unit of the A- and B-batteries of a battery-operated circuit. Also, a unit that supplies the necessary A and B voltages from an a-c source of power.

abrasion machine—A laboratory device for determining the abrasive resistance of wire or cable. The two standard types of machines are the squirrel cage with square steel bars and the abrasive grit types.

abrasion resistance—A measure of the ability of a wire or wire covering to resist damage due to mechanical causes. Usually expressed as inches of abrasive tape travel.

abscissa—Horizontal, or X-, axis on a chart or graph.

absolute address—An address used to specify the location in storage of a word in a computer program, not its position in the program.

absolute altimeter—Electronic instrument which furnishes altitude data with regard to the surface of the earth or any other surface immediately below the instrument—as distinguished from an aneroid altimeter, the readings of which depend on air pressure.

absolute altitude—Altitude with respect to the earth's surface, as differentiated from the altitude with respect to sea level.

absolute code—A code using absolute addresses and absolute operation codes—i.e., a code that indicates the exact location where the reference operand is to be found or stored.

absolute coding—Computer coding using absolute addresses.

absolute delay—The time interval between the transmission of two synchronized radio, loran, or radar signals from the same or different stations.

absolute efficiency—Ratio of the actual output of a transducer to that of a corresponding ideal transducer under similar conditions.

absolute error—1. The amount of error expressed in the same units as the quantity containing the error. 2. Loosely, the absolute value of the error—i.e., the magnitude of the error without regard to its algebraic sign.

absolute gain of an antenna—The gain in a given direction when the reference antenna is an isotropic antenna isolated in space.

absolute humidity—Amount of water vapor present in a unit volume of atmosphere.

absolute maximum rating—Limiting values of operating and environmental conditions, applicable to any electron device of a specified type as defined by its published data, and not to be exceeded under the worst probable conditions. Those ratings beyond which the life and reliability of a device can be expected to decline.

absolute minimum resistance—The resistance between the wiper and the termination of a potentiometer, when the wiper is adjusted to minimize that resistance.

absolute Peltier coefficient—The product of the absolute temperature and the absolute Seebeck coefficient of a material.

absolute pressure—Pressure of a liquid or gas measured relative to a vacuum (zero pressure).

absolute pressure transducer — A pressure transducer that accepts simultaneously two independent pressure sources, and the output of which is proportional to the pressure difference between the sources.

absolute scale—*See* Kelvin Scale.

absolute Seebeck coefficient—The integral from absolute zero to the given temperature of the quotient of the Thomson coefficient of a material divided by its absolute temperature.

absolute system of units—Also called coherent system of units. A system of units in which a small number of units is chosen as fundamental—e.g., units of mass, length, time, and charge. Such units are termed absolute units. All other units are derived from them by taking a definite proportional factor in each of those laws which is chosen as the basic laws for expressing the relationships between the physical quantities. The proportional factor is generally taken as unity.

absolute temperature—Temperature measured from absolute zero, a theoretical level defined as −273.2°C or −459.7°F or 0°K.

absolute units—A system of units based on physical principles, in which a small number of units are chosen as fundamental and all other units derived from them—i.e. abohm, abcoulomb, abhenry, etc.

absolute value—The numerical value of a number or symbol without reference to its algebraic sign. Thus, $|3|$ is the absolute value of +3 or −3. An absolute value is signified by placing vertical lines on both sides of the number or symbol.

absolute value device—A computing element that produces an output equal to the magnitude of the input signal, but always of one polarity.

absolute zero—Lowest possible point on the scale of absolute temperature; the point at which all molecular activity ceases. Absolute zero is defined as −273.2°C, −459.7°F, or 0°K.

absorber—1. In a nuclear reactor, a substance that absorbs neutrons without reproducing them. Such a substance may be useful in control of a reactor or, if unavoidably present, may impair the neutron economy. 2. Any material or device which absorbs and dissipates radiated energy.

absorption—1. Dissipation of the energy of a radio or sound wave into other forms as a result of its interaction with matter. 2. The process by which the number of particles or photons entering a body of matter is reduced by interaction of the particle or radiation with the matter. Similarly, the reduction of the energy of a particle while traversing a body of matter. This term is sometimes erroneously used for capture. 3. Penetration of a substance into the body of another.

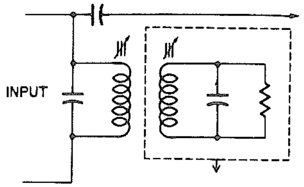

Absorption circuit.

absorption circuit—A tuned circuit that dissipates energy taken from another circuit.

absorption coefficient—1. Measure of sound-absorbing characteristics of a unit area of a given material, compared with the sound-absorbing characteristics of an open space (total absorption) having the same area. 2. Ratio or loss of intensity caused by absorption, to the total original intensity of radiation.

absorption control—Control of a nuclear reactor by use of a neutron absorber. Adjustment is made by varying the effective amount of absorber in or near the core. The most common arrangement is to incorporate the absorber in rods which can be moved in or out to produce the desired effect.

absorption current — The current flowing into a capacitor following its initial charge, due to a gradual penetration of the electric stress into the dielectric. Also, the current which flows out of a capacitor following its initial discharge.

absorption dynamometer—An instrument for measuring power, in which the energy of a revolving wheel or shaft is absorbed by the friction of a brake.

absorption fading—A slow type of fading, primarily caused by variations in the absorption rate along the radio path.

absorption frequency meter—*See* Absorption Wavemeter.

absorption loss—That part of transmission loss due to dissipation or conversion of electrical energy into other forms (e.g., heat), either within the medium or attendant upon a reflection.

absorption marker—A sharp dip on a frequency-response curve due to the absorption of energy by a circuit sharply tuned to the frequency at which the dip occurs.

absorption modulation — Also called loss modulation. A system for amplitude-modulating the output of a radio transmitter by means of a variable-impedance device (such as a microphone or vacuum-tube circuit) inserted into or coupled to the output circuit.

absorption trap — A parallel-tuned circuit coupled either magnetically or capacitively to absorb and attenuate interfering signals.

absorption wavemeter—Also called absorption frequency meter. An instrument for measuring frequency. Its operation depends on the use of a tuned electrical circuit or cavity loosely coupled to the source. Maximum energy will be absorbed at the resonant frequency, as indicated by a meter or other device. Frequency can then be determined by reference to a calibrated dial or chart.

absorptivity—A measure of the portion of incident radiation or sound energy absorbed by a material.

A-B test—Direct comparison of two sounds by playing first one and then the other. May be done with two tape recorders playing identical tapes (or the same tape), two speakers playing alternately from the same tape recorder, or two amplifiers playing alternately through one speaker, etc.

abvolt—Centimeter-gram-second electromagnetic unit of potential difference. The potential difference between two points when one erg of work is required to transfer one abcoulomb of positive electricity from a lower to a higher potential. An abvolt is equal to 10^{-8} volt.

ac—Abbreviation for alternating current.

a-c bias—The alternating current, usually of a frequency several times higher than the highest signal frequency, that is fed to a record head in addition to the signal current. The a-c bias serves to linearize the recording process.

accelerating conductor or relay—One which causes the operation of a succeeding device to begin in the starting sequence after the proper conditions have been established.

accelerating electrode—An electrode in a cathode-ray or other electronic tube to which a positive potential is applied to increase the velocity of electrons or ions toward the anode. A klystron tube does not have an anode but does have accelerating electrodes.

acceleration—The rate of change in velocity. Often expressed as a multiple of the acceleration of gravity ($g = 32.2$ ft/sec^2).

acceleration at stall—The value of servomotor angular acceleration calculated from the stall torque of the motor and the moment of inertia of the rotor. Also called torque-to-inertia ratio.

acceleration switch—A type of limit switch which senses acceleration forces above a preset value and in a particular direction. Commonly used in aerospace equipment, it triggers sensors or arming networks when a specific gravitational force has been reached.

acceleration time—In a computer, the elapsed time between the interpretation of instructions to read or write on tape and the possibility of information transfer from the tape to the internal storage, or vice versa.

acceleration voltage — Potential between a cathode and anode or other accelerating element in a vacuum tube. Its value determines the average velocity of the electrons.

accelerator—A device for imparting a very high velocity to charged particles such as electrons or protons. Fast-moving particles of this type are used in research or in studying the structure of the atom itself.

accelerator dynamic test—A test performed on an accelerometer by means of which information is gathered pertaining to the overall behavior frequency response and/or natural frequency of the device.

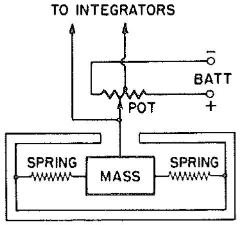

Accelerometer.

accelerometer — An instrument or device, often mounted in an aircraft, guided missile, or the like; used to sense accelerative forces and convert them into corresponding electrical quantities usually for measuring, indicating or recording purposes. It does not measure velocity or distance, only changes in velocity.

accentuation—Also called pre-emphasis. The emphasizing of any certain band of frequencies, to the exclusion of all others, in an amplifier or electronic device. Applied particularly to the higher audio frequencies in frequency-modulated (f-m) transmitters.

accentuator—Network or circuit used for pre-emphasis or increase in amplitude of a given band of frequencies, usually audio.

acceptable-environmental-range test — A test to determine the range of environmental conditions for which an equipment maintains at least the minimum required reliability.

acceptable quality level—The maximum percentage of defective components considered to be acceptable as an average for a process or the lowest quality a supplier is permitted to present continually for acceptance. (Abbreviated AQL.)

acceptance sampling plan—A plan for the inspection of a sample as a basis for acceptance or rejection of a lot or taking another sample.

acceptance test—A test to demonstrate the degree of compliance of a purchaser's equipment with his requirements and specifications.

acceptor—Also called acceptor impurity. An impurity lacking sufficient valence electrons to complete the bonding arrangement in the crystal structure. When added to a semiconductor crystal, it accepts an electron from a neighboring atom and thus creates a hole in the lattice structure of the crystal.

acceptor circuit—A circuit which offers minimum opposition to a given signal.

acceptor impurity—*See* Acceptor.

access arm—In a computer storage unit, a mechanical device which positions the reading and writing mechanism.

access time—Also called waiting time. The time interval between the instant a memory or storage device requests information and the instant this information begins to be available in useful form.

a-c circuit breaker—A device which is used to close and interrupt an a-c power circuit under normal conditions, or to interrupt this circuit under faulty or emergency conditions.

accompaniment manual — In an organ, the keyboard used for playing the accompaniment to the melody. Also called the lower manual, or great manual.

accompanying audio (sound) channel—Also known as co-channel sound frequency. The r-f carrier frequency which supplies the sound to accompany a television picture.

accordion—A type of contact used in some printed-circuit connectors. The contact spring is given a *z* shape to permit high deflection without excessive stress.

a-c coupling—Coupling of one circuit to another circuit through a capacitor or other device which passes the varying portion but not the static (d-c) characteristics of an electrical signal.

accumulator—In an electronic computer, a device which stores a number and which, on receipt of another number, adds the two and stores the sum. An accumulator may have properties such as shifting, sensing signals, clearing, complementing, and so forth.

accuracy—1. The maximum error in the measurement of a physical quantity in terms of the output of an instrument when referred to the individual instrument calibration. Usually given as a percentage of full scale. 2. The quality of freedom from mistake or error in an electronic computer—that is, of conformity to truth or to a rule.

accuracy rating of an instrument — The limit, usually expressed as a percentage of full-scale value, not exceeded by errors when the instrument is used under reference conditions.

ac/dc—Electronic equipment capable of operation from either an a-c or d-c primary power source.

a-c/d-c receiver—A radio receiver designed to operate directly from either an a-c or a d-c source.

a-c directional overcurrent relay—A device which functions on a desired value of a-c overcurrent flowing in a predetermined direction.

a-c dump—The intentional, accidental, or conditional removal of all alternating-current power from a system or component. An a-c dump usually results in the removal of all power, since direct current is usually supplied through a rectifier or converter.

a-c erasing head—In magnetic recording, a device using alternating current to produce the magnetic field necessary for removal of previously recorded information.

acetate—A basic chemical compound in the mixture used to coat recording discs.

acetate base—The transparent plastic film which forms the tough backing for acetate magnetic recording tape.

acetate disc—A mechanical recording disc, either solid or laminated, made mostly from cellulose nitrate lacquer plus a lubricant.

acetate tape—A sound-recording tape with a smooth, transparent acetate backing. One side is coated with an oxide capable of being magnetized.

a-c generator—1. A rotating electrical machine that converts mechanical power into alternating current. Also known as an alternator. 2. A device, usually an oscillator, designed for the purpose of producing alternating current.

A-channel—One of two stereo channels (usually the left) to the microphones, speakers, or other equipment associated with this channel.

achieved reliability—Reliability determined on the basis of actual performance of nominally identical items under equivalent environmental conditions. Also called operational reliability.

achromatic—1. In color television, a term meaning a shade of gray from black to white, or the absence of color (without color). 2. Black-and-white television, as distinguished from color television. 3. A term applied to lenses, signifying their more or less complete correction for chromatic aberration.

achromatic lens—A lens which has been corrected for chromatic aberration. Such a lens is capable of bringing all colors of light rays to approximately the same point of focus. This it does by combining a concave lens of flint glass with a convex lens of crown glass.

achromatic locus—Also called achromatic region. On a chromaticity diagram an area that contains all points representing acceptable reference white standards.

achromatic region—*See* Achromatic Locus.

acid—A chemical compound which dissociates and forms hydrogen ions when in aqueous solution.

acid depolarizer—An acid, such as nitric acid, sometimes introduced into a primary cell to prevent polarization.

aclinic line—Also called isoclinic line. On a magnetic map, an imaginary line which connects points of equal magnetic inclination or dip.

a-c magnetic biasing—In magnetic recording, the method used to remove random noise and/or previously recorded material from the wire or tape. This is done by introducing an alternating magnetic field at a substantially higher frequency than the highest frequency to be recorded.

a-c noise—Noise which displays a rate of change which is fast relative to the response capability of the device.

a-c noise immunity—A measure of a logic circuit's ability to maintain the prescribed logic state in the presence of such noise. It is defined in terms of the amplitude and pulse width of an input noise signal to which the element will not respond.

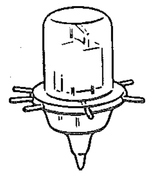

Acorn tube.

acorn tube—A button- or acorn-shaped vacuum tube with no base, for uhf applications. Electrodes are brought out through the glass envelope on the side, top, and bottom.

acoustic—Also acoustical. Pertaining to sound or the science of sound.

acoustic absorption loss—The energy lost by conversion into heat or other forms when sound passes through or is reflected by a medium.

acoustic absorptivity—The ratio of sound energy absorbed by a surface to the sound energy arriving at the surface. Equal to 1 minus the reflectivity of the surface.

acoustical attenuation constant—The real part of the acoustical propagation constant. The commonly used unit is the neper per section or per unit distance.

acoustical-electrical transducer — A device designed to transform sound energy into electrical energy and vice versa.

acoustical mode — A mode of crystal-lattice vibration that does not produce an oscillating dipole.

acoustical ohm—A measure of acoustic resistance, reactance, or impedance. One acoustical ohm is equal to a volume velocity of 1 cubic centimeter per second when produced by a sound pressure of 1 microbar.

acoustical phase constant—The imaginary part of the acoustical propagation constant. The commonly used unit is the radian per section or per unit distance.

acoustical reflectivity—*See* Sound-Reflection Coefficient.

acoustical transmittivity—*See* Sound-Transmission Coefficient.

acoustic burglar alarm—Also called acoustic intrusion detector. A burglar alarm that is responsive to sounds produced by an intruder. Concealed microphones connected to an audio amplifier trip an alarm when sounds exceed a predetermined normal level.

acoustic capacitance—In a sound medium, a measure of volume displacement per dyne per square centimeter. The unit is centimeter to the fifth power per dyne.

acoustic clarifier—A system of cones loosely attached to the baffle of a speaker and designed to vibrate and absorb energy during sudden loud sounds, thereby suppressing them.

acoustic compliance—1. The measure of volume displacement of a sound medium when subjected to sound waves. 2. That type of acoustic reactance which corresponds to capacitive reactance in an electrical circuit.

acoustic delay line—A device which retards one or more signal vibrations by causing them to pass through a solid or liquid.

acoustic depth finder—*See* Fathometer.

acoustic dispersion — The change of the speed of sound with frequency.

acoustic elasticity—1. The compressibility of the air in a speaker enclosure as the cone moves backward. 2. The compressibility of any material through which sound is passed.

acoustic feedback—Also called acoustic regeneration. The mechanical coupling of a portion of the sound waves from the output of an audio-amplifying system to a preceding part or input circuit (such as the microphone) of the system. When excessive, acoustic feedback will produce a howling sound in the speaker.

acoustic filter — A sound-absorbing device that selectively suppresses certain audio frequencies while allowing others to pass.

acoustic frequency response—The voltage-attenuation frequency measured into a resistive load, producing a bandwidth approaching sufficiently close to the maximum.

acoustic generator—A transducer such as a speaker, headphones, or a bell, which converts electrical, mechanical, or other forms of energy into sound.

acoustic homing system—A missile guidance system which responds to noise radiated by the target.

acoustic horn—Also called horn. A tube of varying cross section having different terminal areas which change the acoustic impedance to control the directivity of the sound pattern.

acoustic impedance—Total opposition of a medium to sound waves. Equal to the force per unit area on the surface of the medium, divided by the flux (volume velocity or linear velocity multiplied by area) through that surface. Expressed in ohms and equal to the mechanical impedance divided by the square of the surface area. One unit of acoustic impedance is equal to a volume velocity of one cubic centimeter per second produced by a pressure of 1 microbar. Acoustic impedance contains both acoustic resistance and acoustic reactance.

acoustic inertance—A type of acoustic reactance which corresponds to inductive reactance in an electrical circuit. (The resistance to movement or reactance offered by the sound medium because of the inertia of the effective mass of the medium.) Measured in acoustic ohms.

acoustic intensity—The limit approached by the quotient of acoustical power being transmitted at a given time through a given area divided by the area as the area approaches zero.

acoustic interferometer—An instrument for measuring the velocity or frequency of sound waves in a liquid or gas. This is done by observing the variations of sound pressure in a standing wave, established in the medium between a sound source and a reflector, as the reflector is moved or the frequency is varied.

acoustic intrusion detector—*See* Acoustic Burglar Alarm.

acoustic labyrinth—Special speaker enclosure having partitions and passages to prevent cavity resonance and to reinforce bass response.

acoustic lens—An array of obstacles that refract sound waves in the same way that an optical lens refracts light waves. The dimensions of these obstacles are small compared to the wavelengths of the sounds being focused.

acoustic line—Mechanical equivalent of an electrical transmission line. Baffles, labyrinths, or resonators are placed at the rear of a speaker to help reproduce the very low audio frequencies.

acoustic memory—A computer memory using an acoustic delay line. The line employs a train of pulses in a medium such as mercury or quartz.

acoustic mine—Also called sonic mine. An underwater mine that is detonated by sound waves, such as those from a ship's propeller or engines.

acoustic mirage—The distortion of a sound wavefront by a large temperature gradient in air or water. This creates the illusion of two sound sources.

acoustic ohm—The unit of acoustic resistance, reactance, or impedance. One acoustic ohm is present when a sound pressure of 1 dyne per square centimeter produces a volume velocity of 1 cubic centimeter per second.

acoustic phase constant—The imaginary part of the acoustic propagation constant. The commonly used unit is the radian per section or per unit distance.

acoustic pickup — In nonelectrical phonographs, the method of reproducing the material on a record by linking the needle directly to a flexible diaphragm.

acoustic radiator — In an electroacoustic transducer, the part that initiates the radiation of sound vibration. A speaker cone or headphone diaphragm is an example.

acoustic radiometer — An instrument for measuring sound intensity by determining the unidirectional steady-state pressure caused by the reflection or absorption of a sound wave at a boundary.

acoustic reactance—That part of acoustic impedance due to the effective mass of the medium—that is, to the inertia and elasticity of the medium through which the sound travels. The imaginary component of acoustic impedance and expressed in acoustic ohms.

acoustic reflectivity—The ratio of the rate of flow of sound energy reflected from the surface on the side of incidence, to the incident rate of flow.

acoustic refraction—A bending of sound waves when passing obliquely from one medium to another in which the velocity of sound is different.

acoustic regeneration — *See* Acoustic Feedback.

acoustic resistance — That component of acoustic impedance responsible for the dissipation of energy due to friction between molecules of the air or other medium through which sound travels. Measured in acoustic ohms and analogous to electrical resistance.

acoustic resonance—An increase in sound intensity as reflected waves and direct waves which are in phase combine. May also be due to the natural vibration of air columns or solid bodies at a particular sound frequency.

acoustic resonator—An enclosure which intensifies those audio frequencies at which the enclosed air is set into natural vibration.

acoustics—1. Science of production, transmission, reception, and effects of sound. 2. In a room or other location, those characteristics which control reflections of sound waves and thus the sound reception in it.

acoustic scattering — The irregular reflection, refraction, or diffraction of a sound wave in many directions.

acoustic shock—Physical pain, dizziness, and sometimes nausea brought on by hearing a loud, sudden sound.

acoustic system—Arrangement of components in devices designed to reproduce audio frequencies in a specified manner.

acoustic transmission system—An assembly of elements adapted for the transmission of sound.

acoustic treatment—Use of certain sound-absorbing materials to control the amount of reverberation in a room, hall, or other enclosure.

acoustic wave—A traveling vibration by which sound energy is transmitted in air, in water, or in the earth. The characteristics of these waves may be described in terms of change of pressure, of particle displacement, or of density.

acoustic wave filter—A device designed to separate sound waves of different frequencies. (Through electroacoustic transducers, such a filter may be associated with electric circuits.)

acoustoelectric effect—Generation of an electric current in a crystal by a traveling longitudinal sound wave.

a-c plate resistance—Also called dynamic plate resistance. Internal resistance of a vacuum tube to the flow of alternating current. Expressed in ohms, the ratio of a small change in plate voltage to the resultant change in plate current, other voltages being held constant.

a-c power supply—A power supply that provides one or more a-c output voltages—e.g., a-c generator, dynamotor, inverter, or transformer.

acquisition—1. The process of locating the orbit of a satellite or trajectory of a space probe so that tracking or telemetry data can be gathered. 2. The process of pointing an antenna or telescope so that it is properly oriented to allow gathering of tracking or telemetry data from a satellite or space probe.

acquisition and tracking radar—A radar set which locks on to a strong signal and tracks the object emitting or reflecting the signal. May be airborne or on the ground. Tracking radars use a dish-type antenna reflector to produce a searchlight-type beam.

acquisition radar—A radar set that detects an approaching target and feeds approximate position data to a fire-control or missile-guidance radar, which then takes over the function of tracking the target.

a-c receiver—A radio receiver designed to operate from an a-c source only.

a-c reclosing relay—A device which controls the automatic reclosing and locking out of an a-c circuit interrupter.

a-c relay—A relay designed to operate from an alternating-current source.

a-c resistance—Total resistance of a device in an a-c circuit. (*Also see* High-Frequency Resistance.)

a-c time overcurrent relay—A device which has either a definite or an inverse time characteristic and functions when the current in an a-c circuit exceeds a predetermined value.

actinic—In radiation, the property of producing a chemical change, such as the photographic action of light.

actinium—A radioactive element discovered in pitchblende by the French chemist Debierne in 1889. Its atomic number is 89; its atomic weight, 227.

actino electric effect—The property of some special materials whereby when an electric current is impressed on them, their resistance changes with light.

actino electricity—Electricity produced by the action of radiant energy on crystals.

actinometer—An instrument that measures the intensity of radiation by determining the amount of fluorescence produced by that radiation.

action area—In the rectifying junction of a metallic rectifier, that portion which carries the forward current.

action current—A brief and very small electric current which flows in a nerve during a nervous impulse.

action potential—The instantaneous value of the voltage between excited and resting portions of an excitable living structure.

activation—1. Making a substance artificially radioactive by placing it in an accelerator such as a cyclotron, or by bombarding it with neutrons. 2. To treat the cathode or target of an electron tube in order to create or increase its emission. 3. The process of adding electrolyte to a cell to make it ready for operation.

activation time—The time interval from the moment activation is initiated to the moment the desired operating voltage is obtained in a cell or battery.

activator—An additive that improves the action of an accelerator.

active area—The portion of the rectifying junction of a metallic rectifier that carries forward current.

active communications satellite—*See* Active Comsat.

active component—1. An electrical or electronic element capable of controlling voltages or currents to produce gain or switching action in a circuit (e.g., transistor, vacuum tube, or saturable reactor). Also called active device, or active element. 2. A device, the output of which is dependent on a source of power other than the main input signal.

active comsat—Also called active communication satellite. A communication satellite in which on-board receivers and transmitters receive signals beamed at them from a ground terminal, amplify them greatly, and retransmit them to another ground terminal. Less sensitive receivers and less powerful transmitters can be used on the ground than are needed for passive comsats.

active current—In an alternating current, a component in phase with the voltage. The working component as distinguished from the idle or wattless component.

active decoder—A device that is associated with a ground station and automatically indicates the radar beacon reply code that is received in terms of its number or letter designation.

active device—*See* Active Component.

active electric network—An electric network containing one or more sources of energy.

active element—*See* Active Component.

active filter—A device employing passive network elements and amplifiers. It is used for transmitting or rejecting signals in certain frequency ranges, or for controlling the relative output of signals as a function of frequency.

active guidance—*See* Active Homing.

active homing—Also called active guidance. A missile system using a radar system in the missile itself to provide target information and to guide itself to the target.

active infrared detection—An infrared detection system in which a beam of infrared rays is transmitted toward one or more possible targets, and the rays reflected from the target are detected.

active jamming—Intentional radiation or reradiation of electromagnetic waves to impair the use of a specific portion of the electromagnetic-wave spectrum.

active leg—Within a transducer, an electrical element which changes its electrical characteristics as a function of the applied stimulus.

active line—A horizontal line which produces the tv picture, as opposed to the lines occurring during blanking (horizontal and vertical retrace).

active maintenance downtime—The time during which work is actually being done on an item from the recognition of an occurrence of failure to the time of restoration to normal operation. This includes both preventive and corrective maintenance.

active material—1. In the plates of a storage battery, lead oxide or some other active substance which reacts chemically to produce electrical energy. 2. The fluorescent material, such as calcium tungstate, used on the screen of a cathode-ray tube.

active mixer and modulator—A device requiring a source of electrical power and using nonlinear network elements to heterodyne or combine two or more electrical signals.

active network—A network containing passive and active (gain) elements.

active pressure—In an a-c circuit, the pressure which produces a current, as distinguished from the voltage impressed upon the circuit.

active RC network—A network formed by resistors, capacitors, and active elements.

active repair time—That portion of corrective maintenance downtime during which repair work is being done on the item, including preparation, fault-location, part-replacement, adjustment and recalibration, and final test time. It may also include part procurement time under shipboard or field conditions.

active satellite—A satellite which receives, regenerates, and retransmits signals between stations. *See also* Communications Satellite.

active sonar—*See* Sonar.

active substrate—A substrate in which active elements are formed to provide discrete or integrated devices. Examples of active substrates are single crystals of semiconductor materials within which are transistors, resistors, and diodes, or combinations of these elements. Another example is ferrite substrates within which electromagnetic fields are used to perform logical, gating, or memory functions.

active systems—In radio and radar, systems which require transmitting equipment, such as a beacon or transponder, to be carried in the vehicle.

active tracking system—Usually, a system which requires the addition of a transponder or responder on board the vehicle to repeat or retransmit information to the tracking equipment—e.g., dovap, secor, azusa.

active transducer—A type of transducer in which its output waves depend on one or more sources of power, apart from the actuating waves.

active wire—The wire of an armature winding that produces useful voltage. That portion of the winding in which induction takes place.

activity—1. In a piezoelectric crystal, the magnitude of oscillation relative to the exciting voltage. 2. The intensity of a radioactive source. 3. Operations that result in the use or modification of the information in a computer file.

activity curve—A graph showing how the activity of a radioactive source varies with time

activity ratio—The ratio of the number of records in a computer file which have activity to the total number of records in the file.

a-c transducer—A transducer which, for proper operation, must be excited with alternating currents only. Also a device, the output of which appears in the form of an alternating current.

actual power—The average of values of instantaneous power taken over 1 cycle.

actuating system—1. In a device or vehicle, a system that supplies and transmits energy for the operation of a mechanism or other device. 2. A manually or automatically operated mechanical or electrical device which operates electrical contacts to effect signal transmission.

actuating time—The time at which a specified contact functions.

actuator—1. In a servo system, the device which moves the load. 2. The part of a relay that converts electrical energy into mechanical motion.

a-c voltage—*See* Alternating Voltage.

acyclic machine—A direct-current machine in which the voltage generated in the active conductors maintains the same direction with respect to those conductors at all times.

adapter—A fitting designed to change the terminal arrangement of a jack, plug, socket, or other receptacle, so that other than the original electrical connections are possible.

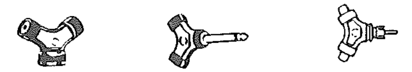

Adapter.

adaptive control system—A device the parameters of which are automatically adjusted to compensate for changes in the dynamics of the process to be controlled. An afc circuit utilizing temperature-compensating capacitors to correct for temperature changes is an example.

adaptive telemetry—Telemetry having the ability to select certain vital information or any change in a given signal.

Adcock antenna—A pair of vertical antennas separated by one-half wavelength or less and connected in phase opposition to produce a figure-8 directional pattern.

Adcock direction finder—A radio direction finder using one or more pairs of Adcock antennas for directional reception of vertically polarized radio waves.

Adcock radio range—A type of radio range utilizing four vertical antennas (Adcock antennas) placed at the corners of a square, with a fifth antenna in the center.

a-d converter—Abbreviation for analog-to-digital converter.

add-and-subtract relay—A stepping relay capable of being operated so as to rotate the movable contact arm in either direction.

addend—A quantity which, when added to another quantity (called the augend), produces a result called the sum.

adder—1. A device which forms the sum of two or more numbers, or quantities, impressed on it. 2. In a color tv receiver, a circuit which amplifies the receiver primary signal coming from the matrix. Usually there is one adder circuit for each receiver primary channel.

addition record—A new record created during the processing of a file in a computer.

additive—Sometimes referred to as the key. A number, series of numbers, or alphabetical intervals added to a code to put it in a cipher.

additive color—A system which combines two colored lights to form a third.

additive primaries—Primary colors which can be mixed to form other colors, but which cannot themselves be produced by mixing other primaries. Red, green, and blue are the primaries in television because, when added in various proportions, they produce a wide range of other colors. For this reason, red, green, and blue are called additive primaries.

additron—An electrostatically focused, beam-switching tube used as a binary adder in high-speed digital computers.

address—1. An expression, usually numerical, which designates a specific location in a storage or memory device or other source or destination of information in a computer. 2. An identification, as represented by a name, label, or number, for a register, location in storage, or any other data source or destination such as the location of a station in a communications network. 3. Loosely, any part of an instruction that specifies the location of an operand for the instruction. 4. *Also see* Instruction Code.

address computation—The process by which the address part of an instruction in a digital computer is produced or modified.

addressed memory—In a computer, memory sections containing each individual register.

address modification—In a computer, a change in the address portion of an instruction or command such that, if the routine which contains that instruction or command is repeated, the computer will go to a new address or location for data or instructions.

address part—In an electronic-computer instruction, a portion of an expression designating location. (*Also see* Instruction Code.)

add-subtract time—The time required by a digital computer to perform addition or subtraction. It does not include the time required to obtain the quantities from storage and put the result back into storage.

add time—The time required in a digital computer to perform addition. It does not include the time required to obtain the quantities from storage and put the result back into storage.

ADF—*See* Automatic Direction Finder.

adiabatic damping—A reduction in the size of an accelerator beam as the energy of the beam is increased.

adiabatic demagnetization—A technique used to obtain temperatures within thousandths of a degree of absolute zero. It consists of applying a magnetic field to a substance at a low temperature and in good thermal contact with its surroundings, insulating the substance thermally, and then removing the magnetic field.

A-display—Also called A-scan. A radar scope presentation in which time (distance or range) is one co-ordinate (usually horizontal) and the target appears displaced perpendicular to the time base.

adjacent- and alternate-channel selectivity—A measure of the ability of a receiver to differentiate between a desired signal and between signals which differ in frequency from the desired signal by the width of one channel or two channels, respectively.

adjacent audio (sound) channel—The r-f carrier frequency which contains the sound modulation associated with the next lower-frequency television channel.

adjacent channel—That frequency band immediately above or below the one being considered.

adjacent-channel attenuation—*See* Selectance.

adjacent-channel interference—Undesired signals received on one communication channel from a transmitter operating on a channel immediately above or below.

adjacent-channel selectivity—The ability of a receiver to reject signals on channels adjacent to the channel of the desired station.

adjacent video carrier—The r-f carrier that carries the picture modulation for the television channel immediately above the channel to which the receiver is tuned.

adjustable resistor—A resistor which has the resistance wire partly exposed to enable the amount of resistance in use to be adjusted occasionally by the user. Adjustment requires the loosening of a screw, the subsequent moving of the lug, and retightening of the screw.

adjustable voltage divider—A wirewound resistor with one or more movable terminals that can be slid along the length of the exposed resistance wire until the desired voltage values are obtained.

adjusted circuit—Also called bolted-fault level. In a circuit the current measured under short-circuit conditions with the leads that are normally connected to the circuit breaker bolted together.

adjusted decibels—The noise level in a transmission system at any point in the system above a reference noise, when the noise meter used has been adjusted to represent the interfering effect under specified conditions. Abbreviated dba.

admittance—The ease with which an alternating current flows in a circuit. The reciprocal of impedance and usually expressed in mhos. Symbol is Y or y.

ADP—Abbreviation for automatic data processing.

adsorption—The deposition of a thin layer of gas or vapor particles or gas onto the surface of a solid. The process is known as chemisorption if the deposited material is bound to the surface by a simple chemical bond.

advance ball—In mechanical recording, a rounded support (often sapphire) which is attached to a cutter and rides on the surface of the recording medium. Its purpose is to maintain a uniform mean depth of cut and to correct for small irregularities on the surface of the disc.

advance wire—An alloy of copper and nickel, used in the manufacture of electric heating units and some wirewound resistors.

aeolight—A glow lamp which employs a cold cathode and a mixture of inert gases and in which the intensity of illumination varies with the applied signal voltage. This lamp is used to produce a modulated light for motion-picture sound recordings.

aerial—*See* Antenna.

aerial cable—A cable installed on a pole line or similar overhead structure.

aerodiscone antenna—An aircraft antenna that is aerodynamically shaped and is physically small compared to other antennas having similar electrical characteristics. Its radiation pattern is omnidirectional and linearly polarized.

aerodrome control radio station—A radio station providing communications between an aerodrome control tower and aircraft or mobile aeronautical radio station.

aerodynamics—The science of the motion of air and other gases. Also, the forces acting on bodies when they move through such gases, or when such gases move against or around the bodies.

aeromagnetic—Pertaining to the magnetic field of the earth as surveyed from the air.

aeronautical advisory station—A station used for civil defense and advisory communications with private aircraft stations.

aeronautical broadcasting service—The broadcasting service intended for the transmission of information related to air navigation.

aeronautical broadcast station—A radio station which broadcasts meteorological information and notices to airmen.

aeronautical fixed service—A fixed service intended for the transmission of information relating to air navigation and preparation for and safety of flight.

aeronautical fixed station—A station operating in the aeronautical fixed service.

aeronautical ground station—A radio station operated for the purpose of providing air-to-ground communications in connection with the operation of aircraft.

aeronautical marker-beacon station—A land station operating in the aeronautical radionavigation service and providing a signal to designate a small area above the station.

aeronautical mobile service—A radio service between aircraft and land stations or between aircraft stations.

aeronautical radio beacon station—A radionavigation land station in the aeronautical radionavigation service, the emission of which enables an aircraft or other mobile service to determine its bearing or its position in relation to the aeronautical radio beacon station.

aeronautical radionavigation service—A radionavigation service intended for use in the operation of aircraft.

aeronautical radio service—1. Service carried on between aircraft stations and/or land stations. 2. Special radio for air navigation.

aeronautical station—A land station (or in certain instances a shipboard station) in the aeronautical mobile service that carries on communications with aircraft stations.

aeronautical telecommunication agency—An agency to which is assigned the responsibility for operating a station or stations in the aeronautical telecommunication service.

aeronautical telecommunication log—A record of the activities of an aeronautical telecommunication station.

aeronautical telecommunications—Any telegraph or telephone communications of sign signals, writing, images, and sounds of any nature, by wire, radio or other system or process of signaling, used in the aeronautical service.

aeronautical telecommunication service—Telecommunication service provided for aeronautical purposes.

aeronautical telecommunication station—A station in the aeronautical telecommunication service.

aeronautical utility land station—A land station located at an airport control tower and used for communications connected with the control of ground vehicles and aircraft on the ground.

aeronautical utility mobile station—A mobile station used at an airport for communications with aeronautical utility land stations, ground vehicles, and aircraft on the ground.

aerophare—*See* Radio Beacon.

AES—Abbreviation for Audio Engineering Society.

af—*See* Audio Frequency.

afc—*See* Automatic Frequency Control.

afterglow—Also called phosphorescence. The light that remains in a gas-discharge tube after the voltage has been removed, or on the phosphorescent screen of a cathode-ray tube after the exciting electron beam has been removed.

afterheat—The heat produced by the continuing decay of radioactive atoms in a reactor after fission has stopped. Most of the afterheat is due to the radioactive decay of fission products.

afterpulse—In a photomultiplier, a spurious pulse induced by a preceding pulse.

agc—*See* Automatic Gain Control.

aging—1. Storing a permanent magnet, capacitor, rectifier, meter, or other device, sometimes with voltage applied, until its desired characteristics become essentially constant. 2. The change of a component or a material with time under defined environmental conditions, leading to improvement or deterioration of properties.

agonic line—An imaginary line on the earth's surface, all points of which have zero magnetic declination.

AGREE—Advisory Group on Reliability of Electronics Equipment.

aided tracking—A system of tracking a target signal in bearing, elevation, or range (or any combination of these variables) in which manual correction of the tracking error automatically corrects the rate at which the tracking mechanism moves.

AIEE—Abbreviation for American Institute of Electrical Engineers. Now merged with IRE to form IEEE.

airborne intercept radar—Short-range airborne radar employed by fighter and interceptor planes to track down their targets.

airborne noise—Undesired sound in the form of fluctuations of air pressure about the atmospheric pressure as a mean.

air capacitor—A capacitor in which air is the only dielectric material between its plates.

aircarrier aircraft station—A radio station aboard an aircraft that is engaged in or essential to the transportation of passengers or cargo for hire.

air cell—A cell in which depolarization at the positive electrode is accomplished chemically by reduction of the oxygen in the air.

air column—The air space within a horn or acoustic chamber.

air condenser—*See* Air Capacitor.

air-cooled tube—An electron tube in which the generated heat is dissipated to the surrounding air directly, through metal heat-radiating fins, or with the aid of channels or chimneys that increase the air flow.

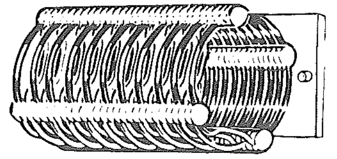

Air-core coil.

air-core coil—A number of turns of spiral wire in which no metal is used in the center.

air-core transformer—A transformer (usually rf) having two or more coils wound around a nonmetallic core. Transformers wound around a solid insulating substance or on an insulating coil form are included in this category.

aircraft bonding—Electrically connecting together all of the metal structure of the aircraft, including the engine and metal covering of the wiring.

aircraft flutter—Flickering in a tv picture as the signal is reflected from flying aircraft. The reflected signal arrives in or out of phase with the normal signal and thus strengthens or weakens the latter.

aircraft station—A radio station installed on aircraft and continuously subject to human control.

airdrome control station—A station used for communication between an airport control tower and aircraft.

air gap—1. A nonmagnetic discontinuity in a ferromagnetic circuit. For example, the space between the poles of a magnet—although filled with brass, wood or any other nonmagnetic material—is nevertheless called an air gap. This gap reduces the tendency toward saturation. 2. The air space between two magnetically or electrically related objects.

air lock—A small chamber, located at the entrance to an area, the doors of which are so interlocked that only one can be opened at a time. This acts as an air seal to maintain the condition of the air within the area.

airport runway beacon—A radio-range beacon which defines one or more approaches to an airport.

airport surveillance radar — Radar equipment or a radar system used in air-traffic control. Used in conjunction with precision approach radar to scan the airspace for a distance of approximately thirty to sixty miles around an airport. It shows, on an indicator in the airport control tower, the location of all airborne aircraft below a certain altitude, as well as obstructions to flight within its range.

air-position indicator—Airborne computing system which presents a continuous indication of aircraft position on the basis of aircraft heading, air speed, and elapsed time.

air-spaced coax—A coaxial cable in which air is basically the dielectric material. The conductor may be centered by means of a spirally wound synthetic filament, by beads, or by braided filaments. This construction is also referred to as an air dielectric.

air-to-ground communication — Transmission of radio signals from an aircraft to stations or other locations on the earth's surface, as differentiated from ground-to-air, air-to-air, or ground-to-ground.

air-to-ground radio frequency—The frequency or band of frequencies agreed upon for transmission from an aircraft to an aeronautical ground station.

air-to-surface missile—A missile designed to be dropped from an aircraft. An internal homing device or the aircraft's radio guides it to a surface target.

airwaves—Slang expression for radio waves used in radio and television broadcasting.

alacritized switch—A mercury switch treated to yield a low adhesional force between the rolling surface and mercury pool, resulting in a decreased differential angle.

alarm relay—A relay, other than an annunciator, used to operate, or to operate in connection with, a visual or audible alarm.

albedo—The reflecting ability of an object. It is the ratio of the amount of light reflected compared to the amount received.

Alexanderson alternator—An early mechanical generator used as a source of low-frequency power for transmission or induction heating. It is capable of generating frequencies as high as 200,000 hertz.

algebraic adder—A computer circuit which can form an algebraic sum.

algol—An international problem language designed for the concise, efficient expression of arithmetic and logical processes and the control (iterative, etc.) of these processes. From ALGOrithmic Language.

algorithm—A fixed step-by-step procedure for finding a solution to a problem.

align—To adjust the tuned circuits of a receiver or transmitter for maximum signal response.

aligned-grid tube—A multigrid vacuum tube in which at least two of the grids are aligned one behind the other to give such effects as beam formation and noise suppression.

alignment—The process of adjusting components of a system for proper interrelationship. The term is applied especially to (1) the adjustment of tuned circuits in a receiver to obtain the desired frequency response, and (2) the synchronization of components in a system.

alignment chart—Also called nomograph or abac. Chart or diagram consisting of two or more lines on which equations can be solved graphically. This is done by laying a straightedge on the two known values and reading the answer at the point where the straightedge intersects the scale for the value sought.

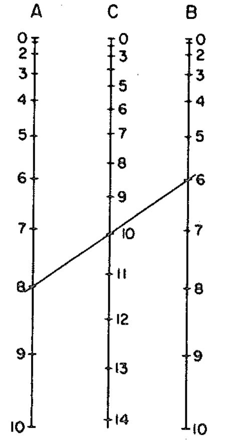

Alignment chart.

alignment pin—1. A pin in the center of the base of a tube. A projecting rib on the pin assures that the tube is correctly inserted into its socket. 2. Any pin or device that will insure the correct mating of two components designed to be connected.

alignment tool—A special screwdriver or socket wrench used for adjusting trimmer or padder capacitors or cores in tuning inductances. It is usually constructed partly or entirely of nonmagnetic material. (*Also see* Neutralizing Tool.)

alive—Electrically connected to a source of potential difference, or electrically charged to have a potential different from that of the earth.

alive circuit—One which is energized.

alkali—A compound which forms hydroxyl ions when in aqueous solution. Also called a base.

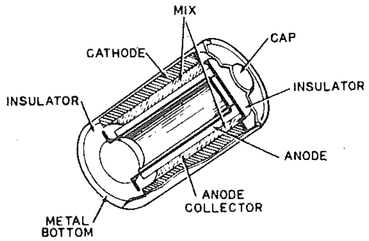

Alkaline cell.

alkaline cell—Also called alkaline-manganese. A primary cell similar to the zinc-carbon cell except that a potassium hydroxide (KOH) electrolyte is used. This cell has about 50 to 100 percent more capacity than the zinc-carbon cell. Nominal cell voltage is 1.5 volts.

all-diffused monolithic integrated circuit—Also called compatible monolithic integrated circuit. A microcircuit consisting of a silicon substrate into which all of the circuit parts (both active and passive elements) are fabricated by diffusion and related processes.

allen screw—A screw having a hexagonal hole or socket in its head. Often used as a setscrew.

allen wrench—A straight or bent hexagonal rod used to turn an Allen screw.

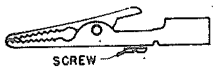

Alligator clip.

alligator clip—A spring-loaded metal clip with long, narrow meshing jaws, used for making temporary electrical connections.

allocate—In a computer, to assign storage locations to main routines and subroutines, thus fixing the absolute values of symbolic addresses.

allocated channels—A channel assigned to a specific user.

allocated frequency band—A segment of the radio-frequency spectrum established by competent authority designating the use which may be made of the frequencies contained therein.

allocated-use circuit—1. A circuit in which one or more channels have been allocated for the exclusive use of one or more services by a proprietary service; may be a unilateral or joint circuit. 2. Communication link specifically assigned to user(s) warranting such facilities.

alloy—A composition of two or more elements, of which at least one is a metal. It may be either a solid solution, a heterogeneous mixture, or a combination of both.

alloy deposition—The process of depositing an alloy on a substrate.

alloy-diffused transistor—A transistor with a diffused base and alloyed emitter.

alloyed contact—An ohmic contact formed by an alloy process.

alloy junction—Also called fused junction. A junction produced by alloying one or more impurity metals to a semiconductor. A small button of impurity metal is placed at each desired location on the semiconductor wafer, heated to its melting point, and cooled rapidly. The impurity metal alloys with the semiconductor material to form a P or N region, depending on the impurity used.

alloy-junction photocell—A photodiode in which an alloy junction is produced by alloying (mixing) an indium disc with a thin wafer of N-type germanium.

alloy-junction transistor—Also called fused-junction transistor. A semiconductor wafer of P- or N-type material with two dots containing P- or N-type impurities fused, or alloyed, into opposite sides of the wafer to provide emitter and base junctions. The base region comprises the original semiconductor wafer.

alloy process—A fabrication technique in which a small part of the semiconductor material is melted together with the desired metal and allowed to recrystallize. The alloy developed is usually intended to form a p-n junction or an ohmic contact.

alloy transistor—A transistor in which the emitter and collector junctions are both alloy junctions.

all-pass network—A network designed to introduce phase shift or delay but not appreciable attenuation at any frequency.

all-wave antenna—A receiving antenna suitable for use over a wide range of frequencies.

all-wave receiver—A receiver capable of receiving stations on all the commonly used wavelengths in short-wave bands as well as in the broadcast band.

alnico—An alloy consisting mainly of ALuminum, NIckel, and CObalt plus iron. Capable of very high flux density and magnetic retentivity. Used in permanent magnets for speakers, magnetrons, etc.

alpha—Emitter-to-collector current gain of a transistor connected as a common-base amplifier. For a junction transistor, alpha is less than unity, or 1.

alphabetic coding—A system of abbreviation used in preparing information for input into a computer. Information may then be reported in the form of letters and words as well as in numbers.

alphabetic-numeric—Having to do with the alphabetic letters, numerical digits, and special characters used in electronic data processing work.

alpha cutoff frequency—The frequency at which the current gain of a common-base transistor stage has decreased to 0.707 of its low-frequency value. Gives a rough indication of the useful frequency range of the device.

alphameric (alphanumeric)—Generic term for alphabetic letters, numerical digits, and special characters which are machine-processable.

alphanumeric—Pertaining to a character set that contains both letters and numerals and usually other characters.

alphanumeric code—A code used to express numerically the letters of the alphabet.

alpha particle—A small, electrically charged particle thrown off at a very high velocity by many radioactive materials including uranium and radium. Identical to the nucleus of a helium atom, it is made up of two neutrons and two protons. Its electrical charge is positive and is equal in magnitude to twice that of an electron.

alpha ray—A stream of fast-moving alpha particles which produce intense ionization in gases through which they pass, are easily absorbed by matter, and produce a glow on a fluorescent screen. The lowest-frequency radioactive emissions.

alternate channel—A channel located two channels above or below the reference channel.

alternate-channel interference — Interference caused in one communication channel by a transmitter operating in the channel after an adjacent channel. (*Also see* Second-Channel Interference.)

alternate frequency—The frequency assigned for use at a certain time, or for a certain purpose, to replace or supplement the frequency normally used.

alternate mode—A means of displaying on an oscilloscope the output signals of two or more channels by switching the channels, in sequence, after each sweep.

alternate routine—A secondary or backup communications path to be used if the normal (primary) routing is not possible.

alternating-charge characteristic — The function relating, under steady-state conditions, the instantaneous values of the alternating component of transferred charge to the corresponding instantaneous values of a specified periodic voltage applied to a nonlinear capacitor.

alternating current—Abbreviated ac. A flow of electricity which reaches maximum in one direction, decreases to zero, then reverses itself and reaches maximum in the opposite direction. The cycle is repeated continuously. The number of such cycles per second is the frequency. The average value of voltage during any cycle is zero.

alternating current/direct current — A term applied to electronic equipment indicating it is capable of operation from either an alternating-current or direct-current primary power source.

alternating-current erasing head—An erasing head, used in magnetic recording in which alternating current produces the magnetic field necessary for erasing. Alternating-current erasing is achieved by subjecting the medium to a number of cycles of a magnetic field of a decreasing magnitude. The medium is, therefore, essentially magnetically neutralized.

alternating-current pulse — An alternating-current wave of brief duration.

alternating-current transmission—In television, that form of transmission in which a fixed setting of the controls makes any instantaneous value of signal correspond to the same value of brightness only for a short time.

alternating quantity—A periodic quantity which has alternately positive and negative values, the average value of which is zero over a complete cycle.

alternating voltage—Also called a-c voltage. Voltage that is continually varying in value and reversing its direction at regular intervals, such as that generated by an alternator or developed across a resistance or impedance through which alternating current is flowing.

alternation—One-half of a cycle—either when an alternating current goes positive and returns to zero, or when it goes negative and returns to zero. Two alternations make one cycle.

alternator—A device for converting mechanical energy into electrical energy in the form of an alternating current.

alternator transmitter—A radio transmitter that generates power by means of a radio-frequency alternator.

altimeter—An instrument that indicates the altitude of an aircraft above a specific reference level, usually sea level or the ground below the aircraft. It may be similar to an aneroid barometer that utilizes the change of atmospheric pressure with altitude, or it may be electronic.

altimeter station—An airborne transmitter, the emissions from which are used to determine the altitude of an aircraft above the surface of the earth.

altitude—The vertical distance of an aircraft or other object above a given reference plane such as the ground or sea level.

altitude delay—The synchronization delay introduced between the line of transmission of the radar pulse and the start of the trace on the indicator. This is done to eliminate the altitude circle on the plan-position indicator-type display.

alto-troposphere—A portion of the atmosphere about 40 to 60 miles above the surface of the earth.

alumina—A ceramic used for insulators in electron tubes for substrates in thin-film circuits. It can withstand continuously high temperatures and has a low dielectric loss over a wide frequency range.

aluminized-screen picture tube—A cathode-ray picture tube which has a thin layer of aluminum deposited on the back of its fluorescent surface to improve the brilliance of the image and also prevent ion-spot formation.

aluminum-electrolytic capacitor—A capacitor with two aluminum electrodes (the anode has the oxide film) separated by layers of absorbent paper saturated with the operating electrolyte. The aluminum-oxide film or dielectric is repairable in the presence of an operating electrolyte.

am—*See* Amplitude Modulation.

amateur—Also called a ham. A person licensed to operate radio transmitters as a hobby. Any amateur radio operator.

amateur bands—Certain radio frequencies assigned exclusively to radio amateurs. In the United States of America, the Federal Communications Commission (FCC) makes these assignments.

amateur station—A radio transmitting station operated by one or more licensed amateur operators.

amateur-station call letters—A group of numbers and letters assigned exclusively to a licensed amateur operator to identify his station.

ambient—Surrounding. (*Also see* Ambient Noise; Ambient Temperature.)

ambient level—The level of interference emanating from sources other than the test sample, such as inherent noise of the measuring device and extraneous radiated fields.

ambient light—Normal room light.

ambient-light filter—A filter used in front of a television picture-tube screen to reduce the amount of ambient light reaching the screen, and to minimize reflections of light from the glass face of the tube.

ambient noise—Acoustic noise in a room or other location. Usually measured with a sound-level meter. The term "room noise" commonly designates ambient noise at a telephone station.

ambient pressure—The general surrounding atmospheric pressure.

ambient temperature—Temperature of air or liquid surrounding any electrical part or device. Usually refers to the effect of such temperature in aiding or retarding removal of heat by radiation and convection from the part or device in question.

ambient temperature range—The range of environmental temperatures in the vicinity of the unit, over which the unit may be operated safely and within specifications. For forced-air cooled units, the ambient temperature is measured at the air intake.

ambiguity—1. An undesirable tendency of a synchro or servo system to seek a false null position, in addition to the proper null position. 2. Inherent error resulting from multiple-bit changes in a polystrophic code. (Proper logic design prevents such errors.)

ambiguous count—A count on an electronic scaler that is obviously impossible.

American Institute of Electrical Engineers (AIEE) — Now merged with IRE to form IEEE.

American Morse code—A system of dot-and-dash signals originated by Samuel F. B. Morse and still used to a limited extent for wire telegraphy in North America. It differs from the International Morse code used in radiotelegraph transmission.

American Radio Relay League (ARRL)—An organization of amateur radio operators.

American Standards Association (ASA)—A national federation of 123 trade associations, technical and professional societies, and consumer groups whose purpose is to develop uniform and voluntary standards of performance, dimension, and terminology.

American wire gage (AWG)—The system of notation generally adopted in the United States for measuring the size of solid wires.

a-m/f-m receiver—A device capable of converting either amplitude- or frequency-modulated signals into audio frequencies.

a-m/f-m tuner—A device capable of converting either amplitude- or frequency- modulated signals into low-level audio frequencies.

ammeter — An instrument for measuring either direct or alternating electric current (depending on its construction). Its scale is usually graduated in amperes, milliamperes, microamperes, or kiloamperes.

ammeter shunt—A low-resistance conductor placed in parallel with the meter movement so most of the current flows through this conductor and only a small part passes through the movement itself. This extends the usable range of the meter.

amp—Abbreviation for ampere.

ampacity—Current-carrying capacity expressed in amperes.

amperage—The number of amperes flowing in an electrical conductor or circuit.

ampere—1. A unit of electrical current or rate of flow of electrons. One volt across 1 ohm of resistance causes a current flow of 1 ampere. A flow of 1 coulomb per second equals 1 ampere. An unvarying current is passed through a solution of silver nitrate of standard concentration at a fixed temperature. A current that deposits silver at the rate of .001118 gram per second is equal to 1 ampere, or 6.25×10^{18} electrons per second passing a given point in a circuit. 2. The constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular sections, and placed 1 meter apart in a vacuum will produce between these conductors a force equal to 2×10^{-7} newton per meter of length.

ampere-hour—A current of one ampere flowing for one hour. Multiplying the current in amperes by the time of flow in hours gives the total number of ampere-hours. Used mostly to indicate the amount of energy a storage battery can deliver before it needs recharging, or the energy a primary battery can deliver before it needs replacing. One ampere-hour equals 3,600 coulombs.

ampere-hour capacity—The amount of current a battery can deliver in a specified length of time under specified conditions.

ampere-hour meter—An electrical meter which measures the amount of current (amperes) per unit of time (hours) which has been consumed in a circuit.

Ampere's rule—Current in a certain direction is equivalent to the motion of positive charges in that direction. The magnetic flux generated by a current in a wire encircles the current in the counter clockwise direction when the current is approaching the observer.

ampere-turn—The magnetomotive force produced by a coil, derived by multiplying the number of turns of wire in the coil by the current (in amperes) flowing through it.

amp-hr—Abbreviation for ampere-hour or ampere-hours.

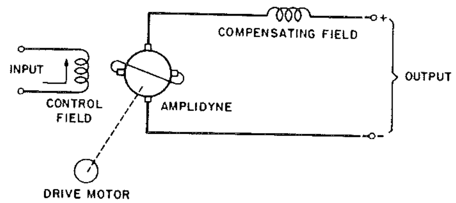

Amplidyne.

amplidyne—A special direct-current generator used extensively in servo systems as a power amplifier. The response of its output voltage to changes in field excitation is very rapid, and its amplification factor is high.

amplification—Increase in size of a medium in its transmission from one point to another. May be expressed as a ratio or, by extension of the term, in decibels.

amplification factor (μ)—1. In a vacuum tube, the ratio of a small change in plate voltage to a small change in grid voltage required to produce the same change in plate current (all other electrode voltages and currents being held constant). 2. In any device, the ratio of output magnitude to input magnitude.

amplified agc—An automatic gain-control circuit in which the control voltage is amplified before being applied to the tube or transistor, the gain of which is to be controlled in accordance with the strength of the incoming signal.

amplified back bias—Degenerative voltage developed across a fast time-constant circuit within a stage of an amplifier and fed back into a preceding stage.

amplifier—A device which draws power from a source other than the input signal and which produces as an output an enlarged reproduction of the essential features of its input. The amplifying element may be an electron tube, transistor, magnetic circuit, or any of various devices.

amplifier noise—All spurious or unwanted signals, random or otherwise, that can be observed in a completely isolated amplifier in the absence of a genuine input signal.

amplifier nonlinearity—The inability of an amplifier to produce an output at all times proportionate to its input.

amplify—To increase in magnitude or strength, usually said of a current or voltage.

amplistat—A self-saturating type of magnetic amplifier.

Amplitron—A broad-band crossed-field amplifier with a re-entrant electron stream. The electron stream interacts with the backward wave of a nonre-entrant r-f structure. (Raytheon.)

amplitude—The magnitude of variation in a changing quantity from its zero value. The word must be modified with an adjective such as peak, rms, maximum, etc., which designates the specific amplitude in question.

amplitude-controlled rectifier—A rectifier circuit in which a thyratron is the rectifying element.

amplitude distortion—In an amplifier or other device, that which occurs when the output amplitude is not a linear function of the input amplitude. (Note: Amplitude distortion is measured with the system operating under steady-state conditions and with a sinusoidal input signal. When other frequencies are present, the term "amplitude" applies to the fundamental only.)

amplitude fading—Fading in which the amplitudes of all frequency components of a modulated carrier wave are uniformly attenuated.

amplitude-frequency response—The variation of gain, loss, amplification, or attenuation of a device or system as a function of frequency. Usually measured in the region where the transfer characteristic is essentially linear.

amplitude gate—*See* Slicer.

amplitude-level selection—The choice of the voltage level at which an oscilloscope sweep is triggered.

amplitude limiter—A circuit or stage which automatically reduces the amplification to prevent signal peaks from exceeding a predetermined level.

amplitude-modulated transmitter—A transmitter in which the amplitude of its radio-frequency wave is varied at a low frequency rate—usually in the audio or video range. This low frequency is the intelligence (information) to be conveyed.

amplitude-modulated wave—A constant-frequency waveform in which the amplitude varies in step with the frequency of an impressed signal.

amplitude modulation—1. Modulation in which the amplitude of a wave is the characteristic subject to variation. 2. Those systems of modulation in which each component frequency (f) of the transmitted intelligence produces a pair of sideband frequencies at carrier frequency plus (f) and carrier frequency minus (f). 3. In special cases, the carrier may be suppressed; either the lower or upper sets of sideband frequencies may be suppressed; the lower set of sideband frequencies may be produced by one or more channels of information. The carrier may be transmitted without intelligence-carrying sideband frequencies. The resulting emission bandwidth is proportional to the highest frequency component of the intelligence transmitted.

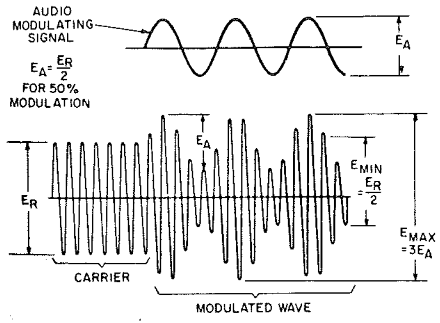

Amplitude modulation.

amplitude-modulation noise level—Undesired amplitude variations of a constant radio-frequency signal, especially in the absence of any intended modulation.

amplitude noise—The effect on radar accuracy of the fluctuations in amplitude of the signal returned by the target. These fluctuations are caused by any change in aspect if the target is not a point source.

amplitude of noise—When impulse-type noise is of random occurrence and so closely spaced that the individual wave shapes are not separated by the receiving equipment, then the noise has the waveshape and characteristics of random noise. Random-noise amplitude is proportional to the square root of the bandwidth. If the impulses are separated, the noise no longer has the wave shape of random noise and its amplitude is directly proportional to the bandwidth of the transmission system.

amplitude permeability—The relative permeability at a stated value of field strength and under stated conditions, the field strength varying periodically with time and no direct magnetic-field component being present.

amplitude range—The ratio, usually expressed in decibels, between the upper and lower limits of program amplitudes which contain all significant energy contributions.

amplitude resonance—The condition that exists when any change in the period or frequency of the periodic agency (but not its amplitude) decreases the amplitude of the oscillation or vibration of the system.

amplitude selection—The process of selecting that portion of a waveform which lies above or below a given value, or between two given values.

amplitude separator—A television-receiver circuit that separates the control impulses from the video signal.

amplitude-suppression ratio—In frequency modulation, the ratio of the magnitude of the undesired output to the magnitude of the desired output of an f-m receiver when the applied signal is simultaneously amplitude- and frequency-modulated. Generally measured with an applied signal that is amplitude-modulated 30% at a 400-hertz rate and is frequency-modulated 30% of the maximum system deviation at a 1,000-hertz rate.

amplitude versus frequency distortion—Distortion caused by the nonuniform attenuation or gain of the system, with respect to frequency under specified terminal conditions.

a-m tuner—A device capable of converting amplitude-modulated signals into low-level audio frequencies.

amu—Abbreviation for atomic mass unit.

anacoustic zone—Zone of silence in space where distances between air molecules are so great that sound waves are not propagated.

analog—In electronic computers a physical system in which the performance of measurements yields information concerning a class of mathematical problems.

analog channel—A computer channel in which the transmitted information can have any value between the defined limits of the channel.

analog communications—A system of telecommunications employing a nominally continuous electrical signal that varies in frequency, amplitude, etc., in some direct correlation to nonelectrical information (sound, light, etc.) impressed on a transducer.

analog computer—A computer operating on the principle of creating a physical (often electrical) analogy of the mathematical problem to be solved. Variables such as temperature or flow are represented by the magnitude of a physical phenomenon such as voltage or current. The computer manipulates these variables in accordance with the mathematical formulas "analogued" on it.

analog data—A physical representation of information such that the representation bears an exact relationship to the original information. The electrical signals on a telephone channel are an analog data representation of the original voice. 2. Data represented in a continuous form, as contrasted with digital data represented in a discrete (discontinued) form. Analog data are usually represented by physical variables, such as voltage, resistance, rotation, etc.

analog network—A circuit or circuits that represent physical variables in such a manner as to permit the expression and solution of mathematical relationships between the variables, or to permit the solution directly by electric or electronic means.

analog output—As distinguished from digital output. Here the amplitude is continuously proportionate to the stimulus, the proportionality being limited by the resolution of the device.

analog recording—A method of recording in which some characteristic of the record current, such as amplitude or frequency, is continuously varied in a manner analogous to the time variations of the original signal.

analog representation — A representation that does not have discrete values, but is continuously variable.

analog switch—A device that either transmits an analog signal without distortion, or completely blocks it.

analog-to-digital converter—A device which produces a digital output from an input in the form of physical motion or electrical voltages.

Analytical Engine—An early form of general-purpose digital computer invented in 1833 by Charges Babbage.

analyzer—1. An instrument or other device designed to examine the functions of components, circuits, or systems, and their relations to each other, as contrasted with an instrument designed to measure some specific parameter of such a system or circuit. 2. Of computers, a routine the purpose of which is to analyze a program written for the same or a different computer. This analysis may consist of summarizing instruction references to storage and tracing sequences of jumps.

anastigmat—A lens system designed so as to be free from the aberration called astigmatism.

anchor—An object, such as a metal rod, set into the ground to hold the end of a guy wire.

ancillary equipment — Equipment not directly employed in the operation of a system, but necessary for logistic support, preparation of flight or assessment of target damage —e.g. test equipment, vehicle transport.

AND circuit—Synonym for AND gate.

Anderson bridge—A bridge normally used for the comparison of self-inductance with capacitance. It is a 6-branch network in which an outer loop of four arms is formed by four nonreactive resistors and the unknown inductor. An inner loop of three arms is formed by a capacitor and a fifth resistor in series with each other and in parallel with the arm opposite the unknown inductor. The detector is connected between the junction of the capacitor and the fifth resistor and at that end of the unknown inductor separated from a terminal of the capacitor by only one resistor. The source is connected to the other end of the unknown inductor and to the junction of the capacitor with two resistors of the outer loop. The balance is independent of frequency.

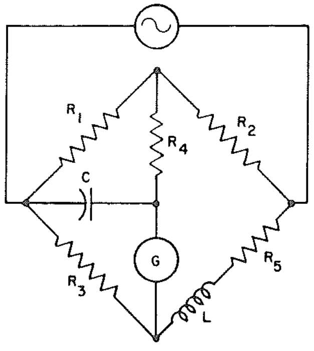

Anderson bridge.

AND gate—In an electronic computer, a gate circuit with more than one control (input) terminal. No output signal will be produced unless a pulse is applied to all inputs simultaneously.

anechoic enclosure—A term originally applied to audio-frequency low-reflection enclosures. Now applied generally to low-reflection enclosures.

anechoic room—An enclosure in which reflected sound energy is negligible. Used for measurement of microphone and speaker characteristics.

anemometer—An instrument used for measuring the force or speed of wind.

angels — Short-duration radar reflections in the lower atmosphere. Most often caused by birds, insects, organic particles, tropospheric layers, or water vapor.

angle—1. A fundamental mathematical concept formed when two straight lines meet at a point. The lines are the sides of the angle, and the point of intersection is the vertex. 2. A measure of the distance along a wave or part of a cycle, measured in degrees. 3. The distance through which a rotating vector has progressed.

angle modulation—Modulation in which the angle of a sine-wave carrier is the characteristic varied from its normal value. Phase and frequency modulation are particular forms of angle modulation.

angle noise—Tracking error introduced into radar by variations in the apparent angle of arrival of the echo from a target due to finite target size. (This effect is caused by variations in the phase front of the radiation from a multiple-point target as the target changes its aspect with respect to the observer.)

angle of advance—*See* Braid Angle.

angle of arrival—Angle made between the line of propagation of a radio wave and the earth's surface at the receiving antenna.

angle of azimuth—The angle measured clockwise in a horizontal plane, usually from the north. The north used may be true north, Y-north, or magnetic north.

angle of beam—The angle which encloses most of the transmitted energy from a directional-antenna system.

angle of convergence—Angle formed by the lines of sight of both eyes when focusing on an object.

angle of deflection—The angle formed between the new position of the electron beam in a cathode-ray tube and the normal position before deflection.

angle of departure—The angle of the line of propagation of a radio wave with respect to a horizontal plane at the transmitting antenna.

angle of divergence—In cathode-ray tubes, a measure of its spread as the electron beam travels from the cathode to the screen. The angle formed by an imaginary center line and the border line of the electron beam. In good tubes, this angle is less than 2°.

angle of elevation—The angle between the horizontal plane and the line ascending to the object.

angle of incidence—The angle between a wave or beam striking a surface and a line perpendicular to that surface.

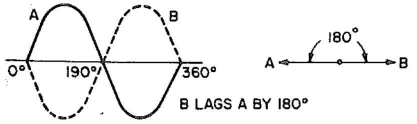

Angle of lag.

angle of lag—The angular phase difference between one sinusoidal function and a second having the same frequency. Expressed in degrees, the amount the second function must be retarded to coincide with the first.

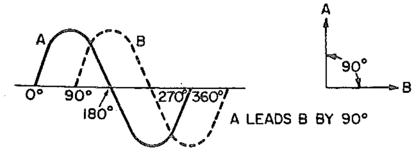

Angle of lead.

angle of lead—1. The time or angle by which one alternating electrical quantity leads another of the same cyclic period. 2. The angle through which the commutator brushes of a generator or motor must be moved from the normal position to prevent sparking.

angle of radiation—The angle between the surface of the earth and the center of the beam of energy radiated upward into the sky from a transmitting antenna.

angle of reflection—The angle between a wave or beam leaving a surface and a line perpendicular to that surface.

angle of refraction—The angle between a wave or beam as it passes through a medium and a line perpendicular to the surface of that medium.

angle tracking noise—Any deviation of the tracking axis from the center of reflectivity of a target. The resultant of servonoise, receiver noise, angle noise, and amplitude noise.

angstrom unit—A unit of measurement of a wavelength of light and other radiation. Equal to one ten-thousandth of a micron or one hundred-millionth of a centimeter (10^{-8} cm). The visible spectrum extends from about 4,000 to 8,000 angstrom units. Blue light has a wavelength in the region of 4,700 angstroms; yellow, 5,800; and red, 6,500.

angular acceleration — The time rate of change in angular velocity.

angular accelerometer—A device capable of measuring the magnitude of, and/or variations in, angular acceleration.

angular aperture—The largest angular extent of wave surface which an objective can transmit.

angular deviation loss—The ratio of the response of a microphone or speaker on its principal axis to the response at a specified angle from the principal axis (expressed in decibels).

angular distance—The angle subtended by two bodies at the point of observation. It is equal to the distance in wavelengths multiplied by 2π radians or by 360°.

angular frequency—Frequency expressed in radians per second. It is equal to the number of hertz (cycles per second) multiplied by 2π.

angular length—Length expressed in radians or equivalent angular measure equal to 2π radians, or 360°, multiplied by the length in wavelengths.

angular phase difference — Phase difference between two sinusoidal functions expressed as an angle.

angular rate—The rate of change of bearing.

angular velocity—The rate at which an angle changes. Expressed in radians per second, the angular velocity of a periodic quantity is the frequency multiplied by 2π. If the periodic quantity results from uniform rotation of a vector, the angular velocity is the number of radians per second passed over by the rotating vector. Generally designated by the Greek letter omega (ω).

anhysteresis—The process whereby a material is magnetized by applying a unidirectional field upon which is superimposed an alternating field of gradually decreasing amplitude.

anion—A negatively charged ion which, during electrolysis, is attracted toward the anode. A corresponding positive ion is called the cation.

anisotropic—A material that has better magnetic characteristics along one axis than any other.

anisotropic body—A body in which the value of any given property depends on the direction of measurement, as opposed to a body that is isotropic.

anisotropic magnet—A magnetic material having a better magnetic characteristic along the preferred axis than along any other.

anisotropy—Directional dependence of magnetic properties, leading to the existence of easy or preferred directions of magnetization. Anisotropy of a particle may be related to its shape, to its crystalline structure, or to the existence of strains within it.

annealed wire—A wire which has been softened by heating (soft-drawn wire).

annealing—The process of heating any solid material such as glass or metal, followed by slow cooling. This generally lowers the tensile strength and thereby improves the ductility.

annular—Ringed; ring-shaped.

annular transistor—A mesa transistor in which the semiconductor regions are arranged in concentric circles about the emitter.

annulling network—An arrangement of impedance elements connected in parallel with filters to annul or cancel capacitive or inductive impedance at the extremes of the passband of a filter.

annunciation relay—1. An electromagnetically operated signaling apparatus which indicates whether a current is flowing or has flowed in one or more circuits. 2. A nonautomatic reset device which gives a number of separate visual indications upon the functioning of protective devices, and which may also be arranged to perform a lockout function.

annunciator—A visual device consisting of a number of pilot lights or drops. Each light or drop indicates the condition which exists or has existed in an associated circuit and is labeled accordingly.

anode—1. The positive electrode such as the plate of a vacuum tube; the element to which the principal stream of electrons flows. 2. In a cathode-ray tube, the electrodes connected to a source of positive potential. These anodes are used to concentrate and accelerate the electron beam for focusing. 3. The less noble and/or higher-potential electrode of an electrolytic cell at which corrosion occurs. This may be an area on the surface of a metal or alloy, the more active metal in a cell composed of two dissimilar metals, or the positive electrode of an impressed-current system.

anode-balancing coil—A set of mutually coupled windings used to maintain approximately equal currents in anodes operating in parallel from the same transformer terminal.

anode breakdown voltage—The potential required to cause conduction across the main gap of a gas tube when the starter gap is not conducting and all other tube elements are held at cathode potential.

anode-bypass capacitor—Also called plate-bypass capacitor. A capacitor connected between the anode and ground in an electron tube circuit. Its purpose is to bypass high-frequency currents and keep them out of the load.

anode characteristic—A graph that shows how the anode current of an electron tube is affected by changes in the anode voltage.

anode circuit breaker—A device used in the anode circuits of a power rectifier for the primary purpose of interrupting the rectifier circuit if an arc-back should occur.

anode current—The electron flow in the element designated as the anode. Usually signifies plate current.

anode dark space—In a gas tube, a narrow, dark zone next to the surface of the anode.

anode dissipation—The power dissipated as heat in the anode of an electron tube because of the bombardment by electrons and ions.

anode efficiency—*See* Plate Efficiency.

anode-load impedance—*See* Plate-Load Impedance.

anode modulation—*See* Plate Modulation.

anode neutralization—Also called plate neutralization. A method of neutralization in which a portion of the anode-cathode a-c voltage is shifted 180° and applied to the grid-cathode circuit through a neutralizing capacitor.

anode power input—*See* Plate Power Input.

anode power supply—The means for supplying power to the plate of an electron tube at a more positive voltage than that of the cathode. (*Also see* Plate Power Supply.)

anode pulse modulation—*See* Plate Pulse Modulation.

anode rays—Positive ions coming from the anode of an electron tube; these ions are generally due to impurities in the metal of the anode.

anode saturation—*See* Plate Saturation.

anode sheath—A layer of electrons surrounding the anode in mercury-pool arc tubes.

anode strap—A metallic connector between selected anode segments of a multicavity magnetron, used principally for mode separation.

anode supply—Also called plate supply. The direct-voltage sources used in an electron-tube circuit to place the anode at a high positive potential with respect to the cathode.

anode terminal—The semiconductor-diode terminal that is positive with respect to the other terminal when the diode is biased in the forward direction.

anode voltage—The potential difference existing between the anode and cathode.

anode voltage drop (of a glow-discharge, cold-cathode tube)—Difference in potential between cathode and anode during conduction, caused by the electron flow through the tube resistance (IR drop).

anodizing—An electrochemical oxidation process used to improve the corrosion resistance or to enhance the appearance of a metal surface. Aluminum and magnesium parts are frequently anodized.

anomalous propagation—1. Propagation that is unusual or abnormal. 2. The conduction of uhf signals through atmospheric ducts or layers in a manner similar to that of a waveguide. These atmospheric ducts carry the signals with less than normal attenuation over distances far beyond the optical path taken by uhf signals. Also called superrefraction. 3. In sonar, pronounced and rapid variations in the strength of the echo due to large, rapid focal fluctuations in propagation conditions.

A – N radio range—A navigational aid which provides four equisignal zones for aircraft guidance. Deviation from the assigned course is indicated aurally by the Morse code letters A (• —) or N (— •). On-course position is indicated by an audible merging of the A and N code signals into a continuous tone.

A – N signal—A radio-range, quadrant-designation signal which indicates to the pilot whether he is on course or to the right or left.

answerback—The response of a terminal to remote-control signals. *Also see* Handshaking.

antenna—Also called aerial. That portion, usually wires or rods, of a radio transmitter or receiver station used for radiating waves into or receiving them from space.

antenna array—A system of two or more antennas coupled together to obtain desirable directional effects.

antenna bandwidth—The range of frequencies over which the impedance characteristics of the antenna are sufficiently uniform that the quality of the radiated signal is not significantly impaired.

antenna beam width—The angle, in degrees, between two opposite half-power points of an antenna beam.

antenna coil—In a radio receiver or transmitter, the inductance through which antenna current flows.

antenna coincidence—That instance when two rotating, highly directional antennas are pointed toward each other.

antenna-conducted interference—Any signal that is generated within a transmitter or receiver, and appears as an undesired signal at the antenna terminals of the device—e.g., harmonics of a transmitter signal, or the local-oscillator signal of the receiver.

antenna cores—Ferrite cores of various cross-sections for use in radio antennas.

antenna coupler—1. A radio-frequency transformer used to connect an antenna to a transmission line or to connect a transmission line to a radio receiver. 2. A radio-frequency transformer, link circuit, or tuned line used to transfer radio-frequency energy from the final plate-tank circuit of a transmitter to the transmission line feeding the antenna.

antenna cross talk—A measure of undesired power transfer through space from one antenna to another. Usually expressed in decibels, the ratio of power received by one antenna to the power transmitted by the other.

antenna current—The radio-frequency current that flows in an antenna.

antenna detector—A device consisting of an antenna and electronic equipment to warn aircraft crew members of their being observed by radar sets. (Usually located in the nose or tail of the aircraft and illuminates a light on one or more panels when radar signals are detected.)

antenna-directivity diagram—A curve representing, in polar or Cartesian coordinates, a quantity proportional to the gain of an antenna in the various directions in a particular plane or cone.

antenna disconnect switch—A safety switch or interlock plug used to remove driving power from the antenna to prevent rotation while work is being performed.

antenna duplexer—A circuit that permits two transmitters to transmit simultaneously from the same antenna without interaction between them.

antenna effect—1. Cause of error in a loop antenna due to the capacitance to ground. 2. In a navigational system, any undesirable output signal that results when a directional antenna acts as a nondirectional antenna.

antenna effective area—In any specified direction, the square of the wavelength multiplied by the power gain (or directive gain) in that direction, and divided by 4π. (When power gain is used, the effective area is that for power reception; when directive gain is used, the effective area is that for directivity.)

antenna factor—The value in db that must be added to a two-terminal voltmeter reading to obtain the actual induced antenna open-circuit voltage or the electric-field strength.

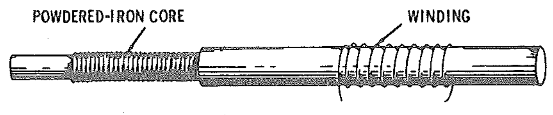

Antenna core.

antenna field—1. The region defined by a group of antennas. 2. A group of antennas placed in a geometric configuration which is specific for a particular trajectory measuring system. 3. The effective free-space energy distribution produced by an antenna or group of antennas.

antennafier—An integrated low-profile antenna and amplifier for use with compact, portable communications systems.

antenna gain—The effectiveness of a directional antenna in a particular direction, compared against a standard (usually an isotropic antenna). The ratio of standard antenna power to the directional antenna power that will produce the same field strength in the desired direction.

antenna ground system—That portion of an antenna closely associated with the earth and including an extensive conducting surface which may be the earth itself.

antenna height—The average antenna height above the terrain from two to ten miles from the antenna. In general, the antenna height will be different in each direction from the antenna. The average of these various heights is considered the antenna height above average terrain.

antenna-induced microvolts—The voltage that exists across the open-circuited antenna terminals, as calculated from a measurement.

antenna lens—An arrangement of metal vanes or dielectric material used to focus a microwave beam in a manner similar to an optical lens.

antenna lobe—*See* Lobe.

antenna matching—Selection of components to make the impedance of an antenna equal to the characteristic impedance of its transmission line.

antennamitter—An integrated low-profile antenna and oscillator for use with compact, portable communications systems.

antenna pair—Two antennas located on a base line of accurately surveyed length. The signals received by these antennas are used to determine quantities related to a target position.

antenna pattern—Also called antenna polar diagram. A plot of angle *versus* free-space field intensity at a fixed distance in the horizontal plane passing through the center of the antenna.

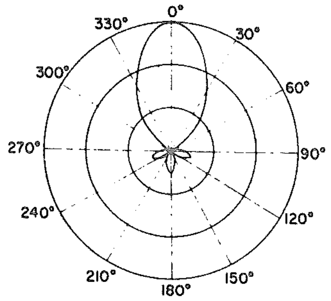

Antenna pattern.

antenna-pattern measuring equipment—Devices used to measure the relative field strength or intensity existing at any point or points in the space immediately surrounding an antenna.

antenna pedestal—A structure which supports an antenna assembly (motors, gears, synchros, rotating joints, etc.).

antenna polar diagram—*See* Antenna Pattern.

antenna power—The square of the antenna current of a transmitter, multiplied by the antenna resistance at the point where the current is measured.

antenna power gain—The power gain of an antenna in a given direction is four times the ratio of the radiation intensity in that direction to the total power delivered to the antenna. (The term is also applied to receiving antennas.)

antenna reflector—In a directional-antenna array, an element which modifies the field pattern in order to reduce the field intensity behind the array and increase it in front. In a receiving antenna, the reflector reduces interference from stations behind the antenna.

antenna relay—A relay used in radio stations to automatically switch the antenna to the receiver or transmitter and thus protect the receiver circuits from the r-f power of the transmitter.

antenna resistance—The total resistance of a transmitting antenna system at the operating frequency. The power supplied to the entire antenna circuit, divided by the square of the effective antenna current referred to the feed point. Antenna resistance is made up of such components as radiation resistance, ground resistance, radio-frequency resistance of conductors in the antenna circuit, and equivalent resistance due to corona, eddy currents, insulator leakage, and dielectric power loss.

antenna resonant frequency—The frequency (or frequencies) at which an antenna appears to be a pure resistance.

antenna stabilization—A system for holding a radar beam steady despite the roll and pitch of a ship or airplane.

antenna switch—Switch used for connecting an antenna to or disconnecting it from a circuit.

antenna system—An assembly consisting of the antenna and the necessary electrical and mechanical devices for insulating, supporting, and/or rotating it.

antenna terminals—On an antenna, the points to which the lead-in (transmission line) is attached.

antenna tilt error—The angular difference between the antenna tilt angle shown on the mechanical indicator, and the electrical center of the radar beam.

antenna wire—A wire, usually of high tensile strength such as copperweld, bronze, etc., with or without insulation, used as an antenna for radio and electronic equipment.

antiaircraft missile—A guided missile launched from the surface against an airborne target.

anticapacitance switch—A switch with widely separated legs, designed to keep capacitance at a minimum in the circuits being switched.

anticathode—Also called target. The target of an X-ray tube on which the stream of electrons from the cathode is focused and from which the X rays are radiated.

anticlutter circuit—In a radar receiver, an auxiliary circuit which reduces undesired reflection, to permit the detection of targets which otherwise would be obscured by such reflections.

anticlutter gain control—A device which automatically and gradually increases the gain of a radar receiver from low to maximum within a specified period after each transmitter pulse. In this way, short-range echoes producing clutter are amplified less than long-range echoes.

anticoincidence—A nonsimultaneous occurrence of two or more events (usually, ionizing events).

anticoincidence circuit—A counter circuit that produces an output pulse when either of two input circuits receives a pulse, but not when the two inputs receive pulses simultaneously.

anticollision radar—A radar system used in an aircraft or ship to warn of possible collision.

antiferroelectricity—The property of a class of crystals which also undergo phase transitions from a higher to a lower symmetry. They differ from the ferroelectrics in having no electric dipole moment.

antiferroelectric materials—Those materials in which spontaneous electric polarization occurs in lines of ions; adjacent lines are polarized in an antiparallel arrangement.

antiferromagnetic materials—Those materials in which spontaneous magnetic polarization occurs in equivalent sublattices; the polarization in one sublattice is aligned antiparallel to the other.

antiferromagnetic resonance—The absorption of energy from an oscillating electromagnetic field by a system of processing spins located on two sublattices, with the spins on one sublattice going in one direction and the spins on the other sublattice in the opposite direction.

antihunt—A stabilizing signal or equalizing circuit used in a closed-loop feedback system of a servomechanism to prevent the system from hunting, or oscillating. Special types of antihunt circuits are the anticipator, derivative, velocity feedback, and damper.

antihunt circuit—A circuit used to prevent excessive correction in a control system.

antihunt device—A device used in positioning systems to prevent hunting or oscillation of the load around an ordered position. The device may be mechanical or electrical. It usually involves some form of feedback.

antijamming—1. Art of minimizing the effect of enemy electronic countermeasures to permit echoes from targets detected by radar to be visible on the indicator. 2. Controls or circuit features incorporated to minimize jamming.

antijamming radar data processing—Use of data from one or more radar sources to determine target range in the presence of jamming.

antilogarithm—The number from which a given logarithm is derived. For example, the logarithm of 4261 is 3.6295. Therefore the antilogarithm of 3.6295 is 4261.

antimagnetic—Made of alloys that will not remain in a magnetized state.

antimicrophonic—Specifically designed to prevent microphonics. Possessing the characteristic of not introducing undesirable noise or howling into a system.

antimissile missile—A missile which is launched to intercept and destroy another missile in flight.

antinodes—Also called loops. The points of maximum displacement in a series of standing waves. Two similar and equal wave trains traveling at the same velocity in opposite directions along a straight line result in alternate antinodes and nodes along the line. Antinodes are separated from their adjacent nodes by half the wavelength of the wave motion.

antinoise carrier-operated device—A device commonly used to mute the audio output of a receiver during standby or no carrier periods. Usually the automatic volume control voltage is used to control a squelch tube which, in turn, controls the bias applied to the first audio tube so that it is permitted to operate only when a carrier is present at the receiver input. Thus, the receiver output is heard when a signal is received, and is muted when no signal is present.

antinoise microphone—A microphone which discriminates against acoustic noise. A lip or throat microphone is an example.

antiproton—An elementary atomic particle which has the same mass as a proton but is negatively charged.

antiresonant circuit—A parallel-resonant circuit offering maximum impedance to the series passage of the resonant frequency.

antiresonant frequency—1. The frequency at which the impedance of a system is very high. 2. Of a crystal unit, the frequency for a particular mode of vibration at which, neglecting dissipation, the effective impedance of the crystal unit is infinite.

antiresonant frequency of a crystal unit—For a particular mode of vibration, the frequency at which (neglecting dissipation) the effective impedance of the crystal unit is infinite.

antisidetone—In a telephone circuit, special circuits and equipment which are so arranged that only a negligible amount of the power generated in the transmitter reaches the associated receiver.

antisidetone circuit—A telephone circuit which prevents sound, introduced in the local transmitter, from being reproduced in the local receiver. (Reduces sidetones.)

antisidetone induction coil—An induction coil designed for use in an antisidetone telephone set.

antisidetone telephone set—A telephone set with an antisidetone circuit.

antistickoff voltage—A small voltage, usually applied to the rotor winding of the coarse synchro control transformer in a two-speed system. The antistickoff voltage acts to eliminate the possibility of ambiguous behavior in the system.

antitransmit-receive switch—Abbreviated ATR switch. An automatic device employed in a radar system to prevent received energy from being absorbed in the transmitter.

aperiodic—Having no fixed resonant frequency or repetitive characteristics or no tendency to vibrate. A circuit that will not resonate within its tuning range is often called aperiodic.

aperiodic antenna—An antenna designed to have a constant impedance over a wide frequency range (for example, a terminated rhombic antenna).

aperiodic damping—Also called overdamping. The condition of a system when the amount of damping is so large that, when the system is subjected to a single disturbance, either constant or instantaneous, the system comes to a position of rest without passing through that position. While an aperiodically damped system is not strictly an oscillating system, it has such properties that it should become an oscillating system if the damping is sufficiently reduced.

aperiodic function—A function having no repetitive characteristics.

aperture—1. In a unidirectional antenna, that portion of the plane surface which is perpendicular to the direction of maximum radiation and through which the major part of the radiation passes. 2. In an opaque disc, the hole or window placed on either side of a lens to control the amount of light passing through. 3. Also called aperture time. The amount of uncertainty about the exact time when the encoder input was at the value represented by a given output code. In general, the aperture is equal to the conversion time; it may be reduced by the use of sample-and-hold circuits.

aperture compensation—Reduction of aperture distortion by boosting the high-frequency response of a television-camera video amplifier.

aperture distortion—In a television signal, the distortion due to the finite dimension of the camera-tube scanning beam. The beam covers several mosaic globules simultaneously, resulting in a loss of picture detail.

aperture illumination—The field distribution in amplitude and phase through the aperture.

aperture mask—Also called shadow mask. A thin sheet of perforated material placed directly behind the viewing screen in a three-gun color picture tube to prevent the excitation of any one color phosphor by either of the two electron beams not associated with that color.

aperture plate—A ferrite memory plate containing a large number of uniformly spaced holes arranged in parallel rows and interconnected by plated conductors to provide a magnetic memory plane.

A power supply—A power supply used as a source of heating current for the cathode or filament of a vacuum tube.

apparent bearing—The direction from which the signal arrives with respect to some reference direction.

apparent power—The product of voltage and current of a single-phase circuit in which the two reach their peaks at different times.

apparent source—*See* Effective Acoustic Center.

Applegate diagram—A graphical representation of electron bunching in a velocity-modulated tube, showing their positions along the drift space. This bunching is plotted on the vertical coordinate, against time along the horizontal axis.

Appleton layer—In the ionosphere, a region of highly ionized air capable of reflecting or refracting radio waves back to earth. It is made up of the F_1 and F_2 layers.

appliance—Any electrical equipment used in the home and capable of being operated by a nontechnical person. Included are units that perform some task that could be accomplished by other, more difficult means, but usually not those used for entertainment (radios, tv's, hi-fi sets, etc.).

application factor—A modifier of the failure rate. It is based on deviations from rated operating stress (usually temperature and one electrical parameter).

applicator (applicator electrodes)—Appropriately shaped conducting surfaces between which an alternating electric field is established for the purpose of producing dielectric heating.

applied voltage—The potential between a terminal and a reference point in any circuit or device.

applique circuit—A special circuit provided to modify existing equipment in order to allow for some special usage.

approach-control radar—Any radar set or system used in a ground-controlled approach system—e.g., an airport-surveillance radar, precision approach radar, etc.

approach path—In radio aircraft navigation, that portion of the flight path in the immediate vicinity of a landing area where such flight path terminates at the touchdown point.

AQL—Abbreviation for acceptable quality level.

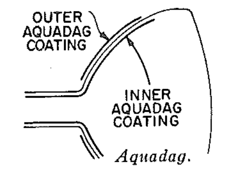

Aquadag.

aquadag—A conductive graphite coating on the inner side walls of some cathode-ray tubes. It serves as an electrostatic shield or as a postdeflection and an accelerating anode. Also applied to outer walls and grounded; here it serves, with the inner coating, as a capacitor to filter the applied high voltage.

arbitrary function fitter—A circuit having an output voltage or current that is a presettable, adjustable, usually nonlinear function of the input voltage(s) or current(s) fed to it.

arc—1. A luminous discharge of electricity through a gas. Characterized by a change in space potential in the immediate vicinity of the cathode; this change is approximately equal to the ionization potential of the gas. 2. A prolonged electrical discharge, or series of prolonged discharges, between two electrodes. (Both produce a bright-colored flame, as contrasted to a dim corona-glow discharge).

arcback—Also called backfire. Failure of the rectifying action in a tube, resulting from the flow of a principal electron stream in the reverse direction due to the formation of a cathode spot on the anode. This action limits the peak inverse voltage which may be applied to a particular rectifier tube.

arc converter—A form of oscillator utilizing an electric arc to generate an alternating or pulsating current.

arc discharge—A discharge between electrodes in gas or vapor. Characterized by a relatively low voltage drop and a high current density.

arc drop—The voltage drop between the anode and cathode of a gas rectifier tube during conduction.

arc-drop loss—In a gas tube, the product of the instantaneous values of arc-drop voltage and current averaged over a complete cycle of operation.

arc-drop voltage—The voltage drop between the anode and cathode of a gas rectifier tube during conduction.

arc failure—1. A flashover in the air near an insulation surface. 2. An electrical failure in the surface heated by a flashover arc. 3. An electrical failure in the surface damaged by the flashover arc.

arc furnace—An electric furnace heated by arcs between two or more electrodes.

arcing—The production of an arc—for example, at the brushes of a motor or at the contact of a switch.

arcing contacts—Special contacts on which the arc is drawn after the main contacts of a switch or circuit breaker have opened.

arcing time—1. The interval between the parting, in a switch or circuit breaker, of the arcing contacts and the extension of the arc. 2. The time elapsing, in a fuse, from the severance of the fuse link to the final interruption of the circuit under the specified condition.

arc lamp—Application of an electric arc to produce a brilliant light. Used extensively in spotlights and motion-picture projectors.

arc oscillator—A negative-resistance oscillator comprised of a sustained d-c arc and a resonant circuit.

arcover voltage—Under specified conditions, the minimum voltage required to create an arc between electrodes separated by a gas or liquid insulation.

arc percussive welding—A type of welding in which the materials to be welded are separated by a gap, across which an arc is struck; the arc melts the surfaces of the materials, and the materials are simultaneously brought together. *See also* Pulse Arc Welding.

arc resistance—The length of time that a material can resist the formation of a conductive path by an arc adjacent to the surface of the material. Also called tracking resistance.

arc suppressor—*See* Spark Suppressor.

arc-through—A loss of control in multielectrode gas tubes. As a result, the principal electron stream flows during a scheduled nonconducting period.

area redistribution—A method of measuring the duration of irregularly shaped pulses. A rectangle is drawn having the same peak amplitude and the same area as the original pulse under consideration. Because the same time units are used in measuring the original and the new pulse, the width of the rectangle is considered the duration of the pulse.

argon—An inert gas used in discharge tubes and some electric lamps. It gives off a purple glow when ionized.

argon glow lamp—A glow lamp containing argon gas which produces a pale blue violet light.

argument—The independent variable of a function.

arithmetic check—A check of a computation making use of the arithmetical properties of the computation.

arithmetic element—Synonym for arithmetic unit.

arithmetic mean—Usually, the same as average. It is obtained by first adding quantities together and then dividing by the number of quantities involved. It also means a figure midway between two extremes and is found by adding the minimum and maximum together and dividing by two.

arithmetic operation—In an electronic computer, the operations in which numerical quantities form the elements of the calculation, including the fundamental operations of arithmetic (addition, subtraction, multiplication, comparison, and division).

arithmetic organ—*See* Arithmetic Unit.

arithmetic shift—In a digital computer, the multiplication or division of a quantity by a power of the base used in the notation.

arithmetic unit—Also called arithmetic element or arithmetic organ. In an automatic digital computer, that portion in which arithmetical and logical operations are performed on elements of information.

armature—The moving element in an electromechanical device such as the rotating part of a generator or motor, the movable part of a relay, or the spring-mounted, iron portion of a bell or buzzer.

armature contacts—1. Contacts mounted directly on the armature. 2. Sometimes used for movable contacts.

armature core—An assembly of laminations forming the magnetic circuit of an armature.

armature gap—The space between the armature and pole face.

armature hesitation—A delay or momentary reversal of the motion of the armature.

armature-hesitation contact chatter—Chatter caused by delay or momentary reversal in direction of the armature motion of a relay during either the operate or the release stroke.

armature-impact contact chatter—Chatter caused by impact of the armature of a relay on the pole piece in operation, or on the backstop in release.

armature overtravel—That portion of the available stroke occurring after the contacts of a relay have touched.

armature reaction—In an armature, the reaction of the magnetic field produced by the current on the magnetic lines of force produced by the field coil of an electric motor or generator.

armature rebound — Return motion of the armature of a relay after it strikes the backstop.

armature-rebound contact chatter—Chatter caused by the partial return of the armature of a relay to its operated position as a result of rebound from the backstop in release.

armature relay—A relay operated by an electromagnet which, when energized, causes an armature to be attracted to a fixed pole or poles.

armature slot—In the core of an armature, a slot or groove into which the coils or windings are placed.

armature stud — In a relay, an insulating member that transmits the motion of the armature to an adjacent contact member.

armature travel—The distance traveled during operation by a specified point on the armature of a relay.

armature voltage control—A means of controlling the speed of a motor by changing the voltage applied to its armature windings.

armature wire—Stranded annealed copper wire, straight lay, soft loose white cotton braid. It is used for low-voltage, high-current rotor winding motors and generators. Straight lay permits forming in armature slots, and compressibility.

armchair copy—Amateur term for clear, static-free signals.

armed sweep—*See* Single Sweep.

arming the oscilloscope sweep—Closing a switch which enables the oscilloscope to trigger on the next pulse.

armor—A metallic cover placed over the insulation of wire or cable to protect it from abrasion or crushing.

armored cable—Two or more insulated wires collectively provided with a metallic covering, primarily to protect the insulated wires from damage.

armor clamp—A fitting for gripping the armor of a cable at the point where the armor terminates or where the cable enters a junction box.

Armstrong frequency-modulation system—A phase-shift modulation system originally proposed by E. H. Armstrong.

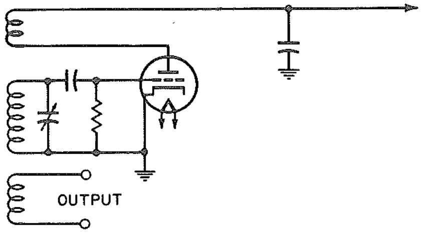

Armstrong oscillator.

Armstrong oscillator—An inductive feedback oscillator that consists of a tuned grid circuit and an untuned tickler coil in the plate circuit. Feedback is controlled by varying the coupling between the tickler and the grid circuit.

array—In an antenna, a group of elements arranged to provide the desired directional characteristics. These elements may be antennas, reflectors, directors, etc.

array antenna—An antenna comprising a number of radiating elements, generally similar, arranged and excited to obtain directional effects.

array device—A group of many similar, basic, complex, or integrated devices without separate enclosures. Each has at least one of its electrodes connected to a common conductor, or all are connected in series.

arrester—1. Also called a lightning arrester. A protective device used to provide a bypass path directly to ground for lightning discharges that strike an antenna or other conductor. 2. A power-line device capable of reducing the voltage of a surge applied to its terminals, interrupting current if present, and restoring itself to original operating conditions.

ARRL — Abbreviation for American Radio Relay League.

arrowhead—A linearly polarized, frequency-independent, log-periodic antenna.

ARSR—Abbreviation for Air Route Surveillance Radar.

articulation— (Sometimes called intelligibility.) In a communications system, the percentage of speech units understood by a listener. The word "articulation" is customarily used when the contextual relationships among the units of speech material are thought to play an unimportant role; the word "intelligibility," when the context is thought to play an important role in determining the listener's perception.

articulation equivalent—The articulation of speech reproduced over a complete telephone connection, expressed numerically in terms of the trunk loss of a working reference system which is adjusted to give equal articulation.

artificial antenna—Also called dummy antenna. A device which simulates a real antenna in its essential impedance characteristics and has the necessary power-handling capabilities, but which does not radiate or receive radio waves. Used mainly for testing and adjusting transmitters.

artificial ear—A microphone-equipped device for measuring the sound pressures developed by an earphone. To the earphone it presents an acoustic impedance equivalent to the impedance presented by the human ear.

artificial echo—1. Received reflections of a transmitted pulse from an artificial target, such as an echo box, corner reflector, or other metallic reflecting surface. 2. A delayed signal from a pulsed radio-frequency signal generator.

artificial horizon—A gyroscopically operated instrument that shows, within limited degrees, the pitching and banking of an aircraft with respect to the horizon. Lines or marks on the face of the instrument represent the aircraft and the horizon. The relative positions of the two are then easily discernible.

artificial ionization—Introduction of an artificial reflecting or scattering layer into the atmosphere to permit beyond-the-horizon communications.

artificial line—A lumped-constant network designed to simulate some or all the characteristics of a transmission line over a desired frequency range.

artificial line duct—A balancing network simulating the impedance of the real line and distant terminal apparatus. It is employed in a duplex circuit to make the receiving device unresponsive to outgoing signal currents.

artificial load—Also called dummy load. A dissipative but essentially nonradiating device having the impedance characteristics of an antenna, transmission line, or other practical utilization circuit.

artificial radioactivity — Radioactivity induced in stable elements under controlled conditions by bombarding them with neutrons or high-energy, charged particles. Artificially radioactive elements emit beta and/or gamma rays.

artificial voice—A small speaker mounted in a specially shaped baffle that is proportioned to simulate the acoustical constants of the human head. It is used for calibrating and testing close-talking microphones.

artos stripper—A machine which, when properly adjusted, will automatically measure to a predetermined length, cut, strip, count, and tie wire in bundles.

artwork—An accurately scaled configuration used to produce a master pattern.

ASA—Abbreviation for American Standards Association.

ASA code—A code recommended by the American Standards Association for industry-wide use in the transmission of information.

asbestos—A nonflammable material generally used for heat insulation, such as in a line-cord resistor.

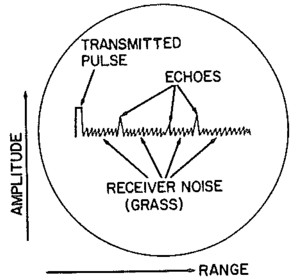

A-scan.

A-scan—Also called A-display. On a cathode-ray indicator, a presentation in which time (range or distance) is one co-ordinate (horizontal) and signals appear as perpendicular deflections to the time scale (vertical).

A-scope—An oscilloscope that uses an A-scan to present the range of a target as the distance along a horizontal line from the transmitted pulse pip to the target, or echo pip. Signals appear as vertical excursions of the horizontal line, or trace.

aspect ratio—Ratio of frame width to frame height. In the United States the television standard is 4/3.

asperities—Local microscopic points on an electrode surface at which there is considerable field enhancement. They lead to a dependence of electric strength on electrode area (area effect).

aspheric—Not spherical; an optical element having one or more surfaces that are not spherical.

ASR—1. Abbreviation for airport surveillance radar. 2. Abbreviation for automatic send-receive set. A combination teletypewriter, transmitter, and receiver with transmission capability from either a keyboard or paper tape. Most often used in a half-duplex circuit.

ASRA — Abbreviation for automatic stereophonic recording amplifier. An instrument developed by Columbia Broadcasting System for stereo recording. Compression of the vertical component of the stereo recording signal is automatically decreased or increased as required by the recording conditions.

assemble—In digital computer programming, to put together subprograms into a complete program.

assembly—1. A complete operating unit, such as a radio receiver, made up of subassemblies such as an amplifier and various components. 2. Process in which instructions written in symbolic form by the programmer are changed to machine language by the computer.

assembly program — A computer program flexible enough to incorporate subroutines into the main program.

assignable cause—A definitely identified factor contributing to a quality variation.

assigned frequency—The center of the frequency band assigned to a station.

assigned frequency band—The frequency band, the center of which coincides with the frequency assigned to the station, and the width of which equals the necessary bandwidth plus twice the absolute value of the frequency tolerance.

astable circuit—A circuit which continuously alternates between its two unstable states at a frequency determined by the circuit constants. It can be readily synchronized by applying a repetitive input signal of slightly higher frequency. A blocking oscillator is an example of an astable circuit.

astable multivibrator (free-running)—A circuit having two momentarily stable states, between which it continuously alternates, remaining in each for a period controlled by the circuit parameters and switching rapidly from one to the other.

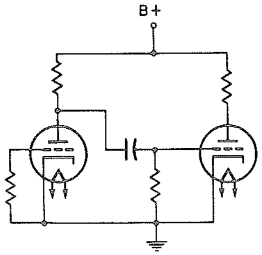

Astable multivibrator.

astatic—1. Having no particular orientation or directional characteristics; such as a vertical antenna. 2. Being in neutral equilibrium; having no tendency toward any change of position.

astatic galvanometer—A sensitive galvanometer used for detecting small currents. Consists of two small magnetized needles of equal size and strength arranged in parallel and with their north and south poles adjacent, suspended inside the galvanometer coil. Since the resultant magnetic moment is zero, the earth's magnetic field does not affect the system.

astigmatism—A type of spherical aberration in which the rays from a single point of an object do not converge on the image, thereby causing a blurred image. Astigmatism in an electron-beam tube is a focus defect in which electrons in different axial planes come to focus at different points.

astrionics—Electronics as involved with astronautics.

astrocompass—An instrument for determining direction relative to the stars. It is unaffected by the errors to which magnetic or gyrocompasses are subject.

astrodome—A rigid hemispherical structure used to cover large tracking instruments to protect them from the elements. It is usually constructed so that the dome rotates with the instrument.

astronautics—The science and art of operating space vehicles.

astrotracker—A device for tracking stars.

A-supply—The A-battery, transformer filament winding, or other voltage source that supplies power for heating the filaments of vacuum tubes.

asymmetrical cell—A cell, such as a photoelectric cell, in which the impedance to the flow of current is greater in one direction than in the other direction.

asymmetrical distortion—Distortion affecting a two-condition or binary modulation or restitution, in which all the significant intervals corresponding to one of the two significant conditions have longer or shorter durations than the corresponding theoretical durations of the excitation. If this particular requirement is not met, distortion is present.

asymmetric sideband transmission — *See* Vestigial Sideband Transmission.

asymmetry control—In pH meters, an adjustment sometimes provided to compensate for differences in the electrodes.

asymptote—A line which comes nearer and nearer a given curve but never touches it.

asymptotic breakdown voltage—A voltage that will break down insulation if applied over a long period of time.

asynchronous—Not having a constant period.

asynchronous computer—An automatic digital computer in which an operation is started by a signal denoting that the previous operation has been completed.

asynchronous device—A device in which the speed of operation is not related to any frequency in the system to which it is connected.

asynchronous machine — Any machine in which its speed of operation is not proportionate to the frequency of the system to which the machine is connected.

AT-cut crystal—A quartz-crystal slab cut at a 35° angle with respect to the optical, or Z-axis, of the crystal. It has practically a zero temperature coefficient and is used at frequencies of about 0.5 to 10 megahertz.

atmosphere—1. The body of air surrounding the earth. 2. A unit of pressure defined as the pressure of 760 mm of mercury at 0° C. Approximately 14.7 pounds per square inch.

atmospheric absorption—The energy lost in the transmission of radio waves due to dissipation in the atmosphere.

atmospheric absorption noise—The dominant noise factor, at frequencies above 1000 MHz, caused by the absorption of energy from radio waves by oxygen and water vapor in the atmosphere.

atmospheric duct—Within the troposphere, a condition in which the variation of refractive index is such that the propagation of an abnormally large proportion of any radiation of sufficiently high frequency is confined within the limits of a stratum. This effect is most noticeable above 3,000 MHz.

atmospheric electricity — Static electricity between clouds, or between clouds and the earth.

atmospheric noise—Also called atmospherics. The noise heard during radio reception because of atmospheric interference.

atmospheric pressure—The barometric pressure of air at a particular location on the earth's surface. The nominal, or standard, value of atmospheric pressure is 760 mm of mercury (14.7 pounds per square inch) at sea level. Atmospheric pressure decreases at higher altitudes.

atmospheric radio wave—A radio wave that is propagated by reflections in the atmosphere. May include the ionospheric wave, the tropospheric wave, or both.

atmospheric radio window—That portion of the frequency spectrum that will allow radio-frequency waves to pass through the earth's atmosphere (approximately 10 to 10,000 MHz).

atmospheric refraction — The bending of the path of electromagnetic radiation from a distant point as the radiation passes obliquely through varying air densities.

atmospherics—Also referred to as static, atmospheric noise, and strays. In a radio tuner or receiver, noise due to natural weather phenomena and electrical charges existing in the atmosphere.

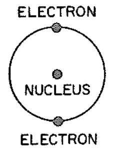

Atom.

atom—The smallest portion of an element which exhibits all properties of the element. It is pictured as composed of a positively charged nucleus containing almost all the mass of the atom, surrounded by one or more electrons. In the neutral atom, the number of electrons is such that their total charge (negative) exactly equals the positive charge in the nucleus.

atomic battery—*See* Nuclear Battery.

atomic charge—The electronic charge of an ion, equal to the number of ionization multiplied by the charge on one electron.

atomic energy—*See* Nuclear Energy.

atomic fission—*See* Fission.

atomic frequency — The natural vibration frequency of an atom.

atomic fuel—A fissionable material—i.e., one in which the atomic nucleus may be split to release energy.

atomic fusion—*See* Fusion.

atomic mass unit—Abbreviated amu. Used to express the relative masses of isotopes. It is so proportioned that the mass of a neutral atom of the naturally most abundant isotope of oxygen (O16) is 16.00000 atomic mass units.

atomic migration—The progressive transfer of a valance electron from one atom to another within the same molecule.

atomic number—The number of protons (positively charged particles) in the nucleus of an atom. All elements have different atomic numbers, which determine their positions in the periodic table. For example, the atomic number of hydrogen is 1; that of oxygen is 8; iron, 26; lead, 82; uranium, 92.

atomic power—*See* Nuclear Energy.

atomic ratio—The ratio of quantities of different substances to the number of atoms of each.

atomic reactor—*See* Nuclear Reactor.

atomic theory—A generally accepted theory concerning the structure and composition of substances and compounds. It states that everything is composed of various combinations of ultimate particles called atoms.

atomic weight—The approximate weight of the number of protons and neutrons in the nucleus of an atom. The atomic weight of oxygen, for example, is approximately 16 (actually it is 16.0044)—it contains 8 neutrons and 8 protons. Aluminum is 27 and contains 14 neutrons and 13 protons. If expressed in grams, these weights are called gram atomic weights.

ATR tube — Abbreviation for antitransmit-receive tube. A gas-filled, radio-frequency switching tube used to isolate the transmitter while a pulse is being received.

attachment cord—*See* Patch Cord.

attack—The length of time it takes for a tone in an organ to reach full intensity after a key is depressed. On most organs this effect is adjustable by either a switch or potentiometer.

attack time—The interval required for an

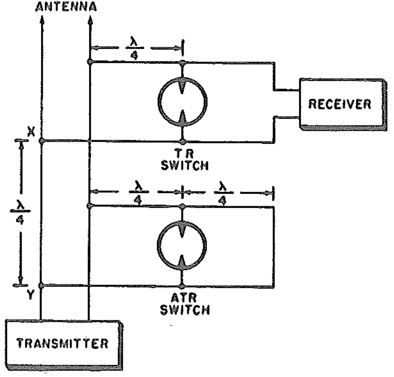

ATR (switch) tube.

input signal, after suddenly increasing in amplitude, to attain a specified percentage (usually 63%) of the ultimate change in amplification or attenuation due to this increase.

attenuating — Decreasing electrical current, voltage, or power in a communicating channel. Refers to audio, radio, or carrier frequencies.

attenuation—The decrease in amplitude of a signal during its transmission from one point to another. It may be expressed as a ratio or, by extension of the term, in decibels.

attenuation constant—1. The real component of the propagation constant. 2. For a traveling plane wave at a given frequency, the rate at which the amplitude of a field component (or the voltage or current) decreases exponentially in the direction of propagation, in nepers or decibels per unit length.

attenuation distortion—1. In a circuit or system, its departure from uniform amplification or attenuation over the frequency range required for transmission. 2. Distortion that causes a decrease in the amplitude of a field component (voltage or current) in the direction of propagation.

attenuation equalizer—A corrective network designed to make the absolute value of the transfer impedance of two chosen pairs of terminals substantially constant for all frequencies within a desired range.

attenuation-frequency distortion — A form of wave distortion in which the relative magnitudes of the different frequency components of the wave are changed.

attenuation network—1. A network providing relatively slight phase shift and substantially constant attenuation over a range of frequencies. 2. The arrangement of circuit elements, usually impedance elements, inserted in circuitry to introduce a known loss or to reduce the impedance level without reflections.

attenuation ratio — The magnitude of the propagation ratio which indicates the relative decrease in energy.

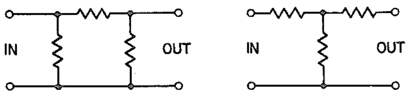

Attenuators.

attenuator—1. An adjustable resistive network for reducing the amplitude of an electrical signal without introducing appreciable phase or frequency distortion. 2. A distributed network that absorbs part of a signal and transmits the remainder with a minimum of distortion or delay.

attenuator tube—A gas-filled, radio-frequency switching tube in which a gas discharge, initiated and regulated independently from the radio-frequency power, is used to control this r-f power by reflection or absorption.

atto—Prefix meaning 10^{-18}. Abbreviated a.

audibility—1. The ability to be heard, usually construed as being heard by the human ear. 2. The ratio of the strength of a specific sound to the strength of a sound that can barely be heard. Usually expressed in decibels.

audible—Capable of being heard, in most contexts by the average human ear.

audible tones—Sounds composed of frequencies which the average human can detect.

audio—Pertaining to frequencies corresponding to a normally audible sound wave. These frequencies range roughly from 15 to 20,000 hertz.

audio amplifier — *See* Audio-Frequency Amplifier.

audio band—The range of audio frequencies passed by an amplifier, receiver, transmitter, etc. (*Also see* Audio Frequency.)

audio-channel wire—A small-diameter wire, shielded and jacketed, used primarily in radio and television for wiring consoles, panels, etc.

audio component—That portion of any wave or signal which contains frequencies in the audible range (between 15 and 20,000 hertz).

audio frequency—Abbreviated af. Any frequency corresponding to a normally audible sound wave. Audio frequencies range roughly from 15 to 20,000 hertz.

audio-frequency amplification—An increase in voltage, current, or power of a signal at the audio frequency.

audio-frequency amplifier—Also called audio amplifier. A device which contains one or more electron tubes or transistors (or both) and which is designed to amplify signals within a frequency range of about 20 to 20,000 hertz.

audio-frequency choke—An inductance used to impede the flow of audio-frequency currents.

audio-frequency noise — In the audio-frequency range, any electrical disturbance introduced from a source extraneous to the signal.

audio-frequency oscillator—An oscillator circuit using an electron tube, transistor, or other nonrotating device capable of producing audio signals.

audio-frequency peak limiter — A circuit generally used in the audio system of a radio transmitter to prevent overmodulation. It keeps the signal amplitude from exceeding a predetermined value.

audio-frequency shift modulator—A system of facsimile transmission over radio, in which the frequency shift required is applied through an 800-hertz shift of an audio signal, rather than shifting the transmitter frequency. The radio signal is modulated by the shifting audio signal, usually at 1500 to 2300 hertz.

audio-frequency transformer — Also called audio transformer. An iron-core transformer for use with audio-frequency currents to transfer signals from one circuit to another. Used for impedance matching or to permit maximum transfer of power.

audiogram—Also called threshold audiogram. A graph showing hearing loss, percent of hearing loss, or percent of hearing as a function of frequency.

audio-level meter—An instrument that measures audio-frequency power with reference to a predetermined level. Usually calibrated in decibels.

audiometer — An instrument for measuring hearing acuity. Measurements may be made with speech signals (usually recorded) or with tone signals.

audion—A three-electrode vacuum tube introduced by Dr. Lee de Forest.

audio oscillator—*See* Audio-Frequency Oscillator.

audio patch bay—Specific patch panels provided for termination of all audio circuits and equipment used in the channel and technical-control facility. This equipment can also be found in transmitting and receiving stations.

audio peak limiter — *See* Audio-Frequency Peak Limiter.

audiophile—A person who is interested in good musical reproduction for his own personal listening and who uses the latest audio equipment and techniques.

audio signal—An electrical signal the frequency of which is within the audio range.

audio spectrum—The continuous range of audio frequencies extending from the lowest to the highest (from about 15 to 20,000 hertz.

audio taper — Semilogarithmic change of resistance. Used on tone controls in audio amplifiers to compensate for the lower sensitivity of the human ear when listening to low-volume sounds.

audio transformer — *See* Audio-Frequency Transformer.

audiovisual—Involving both sight and sound. (As audiovisual education uses films, slides, phonograph records, and the like, to supplement instruction.)

audiovisual system—A system of communications which simultaneously transmits pictorial and audio signals.

augend—In an arithmetic addition, the number increased by having another number (called the addend) added to it.

augmented operation code—In a computer, an operation code that is further defined by information contained in another portion of an instruction.

aural—Pertaining to the ear or to the sense of hearing. (*Also see* Audio.)

aural radio range—A radio range the courses of which are normally followed by interpretation of an aural signal.

aural signal—The signal corresponding to the sound portion of a television program. In general, the audible component of a signal.

aural transmitter—The equipment used to transmit the aural (sound) signals from a television broadcast station.

aurora—Sheets, streamers, or streaks of pale light often seen in the skies of the northern and southern hemispheres. The aurora borealis and aurora australis.

auroral absorption — Absorption of radio waves due to auroral activity. (*Also see* Aurora.)

authorized carrier frequency — A specific carrier frequency authorized for use, from which the actual carrier frequency is permitted to deviate, solely because of frequency instability, by an amount not to exceed the frequency tolerance.

autoalarm—Also called automatic alarm receiver. A device which is tuned to the international distress frequency of 500 kHz and which automatically actuates an alarm if any signal is received.

auto call—An alerting device which sounds a preset code of signals in a building, to page those persons whose code is being sounded.

autodyne circuit—A vacuum-tube circuit which serves simultaneously as an oscillator and as a heterodyne detector.

autodyne reception—A type of radio reception employed in regenerative receivers for the reception of c-w code signals. In this system the incoming signal beats with the signal from an oscillating detector to produce an audible beat frequency.

auto-man—A locking switch which controls the method of operation—i.e., automatic or manual.

automata—A plural form of automaton.

automatic—Self-regulating or self-acting; capable of producing a desired response to certain predetermined conditions.

automatic-alarm receiver — Complete receiving, selecting, and warning device capable of being actuated automatically by intercepted radio-frequency signals forming the international automatic-alarm signal.

automatic back bias—A radar-receiver technique which consists of one or more automatic gain-control loops to prevent large signals from overloading a receiver, whether by jamming or by actual echoes.

automatic bass compensation—A circuit used in a receiver or audio amplifier to make the bass notes sound more natural at low volume settings. The circuit, which usually consists of resistors and capacitors connected to taps on the volume control, automatically compensates for the poor response of the human ear to weak sounds.

automatic bias—*See* Self-Bias.

automatic brightness control—A circuit used in television receivers to keep the average brightness of the reproduced image essentially constant. Its action is similar to that of an automatic volume-control circuit.

automatic check—An operation performed by equipment built into an electronic computer to automatically verify proper operation.

automatic chrominance control—A color-television circuit which automatically controls the gain of the chrominance bandpass amplifier by varying the bias.

automatic circuit breaker—A device that automatically opens a circuit, usually by electromagnetic means, when the current exceeds a safe value. Unlike a fuse, which must be replaced once it blows, the circuit breaker can be reset manually when the current is again within safe limits.

automatic coding—A technique by which a digital computer is programmed to perform a significant portion of the coding of a problem.

automatic color purifier—*See* Automatic Degausser.

automatic computer—A computer which automatically handles long sequences of reasonable operations without further intervention from a human being.

automatic contrast control—A television circuit which automatically changes the gain of the video i-f and r-f stages to maintain proper contrast in the television picture.

automatic crossover—A type of current-limiting circuit on a power supply provided with an adjustment for setting the short-circuit current to an adjustable maximum value.

automatic current limiting—An overload-protection mechanism designed to limit the maximum output current of a power supply to a preset value. Usually it automatically restores the output when the overload is removed.

automatic cutout—A device operated by electromagnetism or centrifugal force, to automatically disconnect some parts of an equipment after a predetermined operating limit has been reached.

automatic data processing—The processing of digital information by automatic computers and other machines. Abbreviated ADP. Also called integrated data processing.

automatic degausser—Also called automatic color purifier and degausser. An arrangement of degaussing coils mounted around a color-television picture tube. These coils are energized only for a short while after the set is turned on. They serve to demagnetize any parts of the picture tube that may have been affected by the earth's magnetic field or the magnetic field of any nearby home appliance.

automatic direction finder—Abbreviated ADF. Also called an automatic radio compass. An electronic device, usually for marine or aviation application, which provides a radio bearing to any transmitter whose frequency is known but whose direction and location are not.

automatic error correction—A technique, usually requiring the use of special codes and/or automatic retransmission, which detects and corrects errors occurring in transmission. The degree of correction depends upon coding and equipment configuration.

automatic exchange—A telephone exchange in which connections are made between subscribers by means of devices set in operation by the originating subscriber's instrument and without the intervention of an operator.

automatic focusing—A method of electrostatically focusing a television picture tube; the focusing anode is internally connected through a resistor to the cathode and thus requires no external focusing voltage.

automatic frequency control—Abbreviated afc. A circuit that automatically maintains the frequency of an oscillator within specified limits.

automatic gain control—1. Abbreviated agc. A type of circuit used to maintain the output volume of a receiver constant, regardless of variations in the signal strength applied to the receiver. 2. A self-acting compensation device which maintains the output of a transmission system constant within narrow limits in the face of wide variations in the attenuation of the system. 3. A radar circuit which prevents saturation of the radar receiver by long blocks of received signals, or by a carrier modulated at low frequency.

automatic gain stabilization—A circuit, used in certain identification friend-or-foe equipment and radar beacon systems, which serves to maintain optimum sensitivity in a superregenerative stage by keeping the noise-pulse load constant. The system prevents random noises from triggering the automatic transmitter associated with the receiver.

automatic grid bias—Grid-bias voltage provided by the difference in potential across a resistance (or resistances) in the grid or cathode circuit due to grid or cathode current or both.

automatic level compensation—A system which automatically compensates for variations in the circuit. *See also* Automatic Volume Control.

automatic light control—The process by

which the illumination incident upon the face of a television pickup device is automatically adjusted as a function of scene brightness.

automatic modulation control—A transmitter circuit that reduces the gain for excessively strong audio input signals without affecting the strength of normal signals, thus permitting higher average modulation without overmodulation; this is equivalent to an increase in the carrier-frequency power output.

automatic noise limiter—A vacuum-tube circuit that automatically clips off all noise peaks above the highest peak of the desired signal being received. This circuit prevents strong atmospheric or man-made interference from being troublesome.

automatic pedestal control—The process by which the pedestal height of a television signal is automatically adjusted as a function of the input or other specified parameter.

automatic phase control—A circuit used in color television receivers to synchronize the burst signal with the 3.58-MHz color oscillator.

automatic pilot—*See* Autopilot.

automatic programming—Any technique designed to simplify the writing and execution of programs in a computer. Examples are assembly programs which translate from the programmer's symbolic language to the machine language, those which assign absolute addresses to instruction and data words, and those which integrate subroutines into the main routine.

automatic radio compass—A radio direction finder which automatically rotates the loop antenna to the correct position. A bearing can then be secured from the indicator dials without mechanical adjustments or calculation. (*Also see* Automatic Direction Finder.)

automatic record changer—An electrically operated mechanism which automatically feeds, plays, and rejects a number of records in a preset sequence. It consists of a motor, turntable, pickup arm, and changer. Modern changers are designed to play automatically 16⅔-, 33⅓-, 45-, and 78-rpm records.

automatic relay—A means of selective switching which causes automatic equipment to record and retransmit communications.

automatic reset—1. A stepping relay that returns to its home position either when it reaches a predetermined contact position or when a pulsing circuit fails to energize its driving coil within a given time. 2. An overload relay that restores the circuit as soon as the cause of the overload is corrected.

automatic reset relay—1. A stepping relay that returns to its home position either when it reaches a predetermined contact position or when a pulsing circuit fails to energize the driving coil within a given time. It may either pulse forward or be spring-reset to the home position. 2. An overload relay that restores the circuit as soon as an overcurrent situation is corrected.

automatic scanning—A variable-speed sweep of the entire frequency range of an RFI meter. It may also include the scanning of a portion of this frequency range over a predetermined sector.

automatic scanning receiver—A receiver which can automatically and continuously sweep across a preselected frequency, either to stop when a signal is found or to plot signal occupancy within the frequency spectrum being swept.

automatic selective control (or transfer) relay—A device which operates to select automatically between certain sources or conditions in an equipment, or performs a transfer operation automatically.

automatic sensitivity control—A circuit used for automatically maintaining the receiver sensitivity at a predetermined level. It is similar to an automatic gain control, but it affects the receiver constantly rather than during the brief interval selected by the range gate. 2. The self-acting mechanism which varies the system sensitivity as a function of the specified control parameters. This may include automatic target control, automatic light control, or any combination thereof.

automatic sequencing—The ability of a computer to perform successive operations without additional instructions from a human being.

automatic short-circuit protection—An automatic current-limiting system which enables a power supply to continue operating at a limited current, and without damage, into any output overload, including a short circuit. The output voltage is restored to normal when the overload is removed, as distinguished from a fuse or circuit-breaker system which opens with overload and must be reclosed manually to restore power.

automatic shutoff—In a tape recorder, a switching arrangement which automatically shuts the recorder off when the tape breaks or runs out. Also, a switching arrangement which stops the record changer after the last record.

automatic starter—A device which, after being given the initial impulse by means of a push button or similar device, starts a system or motor automatically in the proper sequence.

automatic target control—The self-acting mechanism which controls the vidicon target potential as a function of the scene brightness.

automatic telegraph transmission—A form of telegraphy in which signals are transmitted mechanically from a perforated tape.

automatic telegraphy—A form of telegraphy in which signals are transmitted and/or received automatically.

automatic threshold variation—A constant false-alarm rate scheme which is an open-loop type of automatic gain control in which the decision threshold is varied continuously in proportion to the incoming intermediate frequency and video noise level.

automatic time switch—A combination of a switch with an electric or spring-wound clock arranged to turn an apparatus on and off at predetermined times.

automatic tracking—In radar, the process whereby a mechanism, actuated by the echo, automatically keeps the radar beam locked on the target, and may also determine the range simultaneously.

automatic transfer equipment—Equipment which automatically transfers a load so that a source of power may be selected from one of several incoming lines.

automatic tuning—An electrical, mechanical, or electrical/mechanical system that automatically tunes a circuit to a predetermined frequency when a button or other control is operated.

automatic video-noise leveling—A constant false-alarm rate scheme in which the video-noise level at the output of the receiver is sampled at the end of each range sweep and the receiver gain is readjusted accordingly to maintain a constant video-noise level at the output.

automatic voltage regulator—A device or circuit which maintains a constant voltage, regardless of any variation in input voltage or load.

automatic volume compression—*See* Volume Compression.

automatic volume control (avc)—1. A self-acting compensation device which maintains the output of a transmission system constant within narrow limits in the face of wide variations in attenuation in that system. 2. A self-acting device which maintains the output of a radio receiver or amplifier substantially constant within relatively narrow limits while the input voltage varies over a wide range. 3. *See also* Automatic Level Compensation.

automatic volume expansion—Also called volume expansion. An audio-frequency circuit that automatically increases the volume range by making loud portions louder and weak ones weaker. This is done to make radio reception sound more like the actual program, because the volume range of programs is generally compressed at the point of broadcast.

automatic zero and full-scale calibration correction—A system of zero and sensitivity stabilization in which electronic servos are used that compare demodulated "zero" and "full-scale" signals with reference voltages.

automation—1. The method or act of making a manufacturing or processing system partially or fully automatic. 2. The entire field of investigation, design, development, application, and methods of rendering or making processes or machines self-acting or self-moving; rendering automatic.

automaton—1. A device that automatically follows predetermined operations or responds to encoded instructions. 2. Any communication-linked set of elements. 3. A machine which exhibits living properties.

automonitor—1. To instruct an automatic digital computer to produce a record of its information-handling operations. 2. A program or routine for this purpose.

autopilot—Also called automatic pilot, gyropilot, or robot pilot. A device containing amplifiers, gyroscopes, and servomotors which automatically control and guide the flight of an aircraft or guided missile. The autopilot detects any deviation from the planned flight and automatically applies the necessary corrections to keep the aircraft or missile on course.

auto radio—A radio receiver designed to be installed in an automobile and powered by the storage battery of the automobile.

autoradiography—Self-portraits of radioactive sources made by placing the radioactive material next to photographic film. The radiations fog the film and thus leave an image of the source.

autoregulation induction heater—An induction heater in which a desired control is effected by the change in characteristics of a magnetic charge as it is heated at or near its Curie point.

autostarter—*See* Autotransformer Starter.

Autosyn—A trade name of the Bendix Corp. for a remote-indicating instrument or system based on the synchronous-motor principle in which the angular position of the rotor of a motor at the measuring source is duplicated by the rotor of the indicator motor.

auto tracking—Also called automatic tracking operation. A master/slave connection of two or more power supplies, each of which has one of its output terminals in common with one of the output terminals of all of the other supplies, such connections being characterized by one-knob control and proportional output voltage from all supplies. Useful where simultaneous turn-up, turn-down, or proportional control of all power supplies in a system is required.

autotransformer—1. A transformer with a single winding (electrically) in which the whole winding acts as the primary winding, and only part of the winding acts as the secondary (step-down); or part of the winding acts as the primary and the whole winding acts as the secondary (step-up). 2. A voltage, current, or impedance transforming device in which parts of one winding are common to both the primary and secondary circuits.

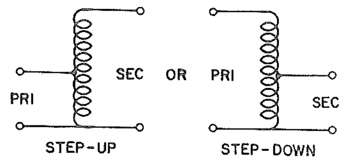

Autotransformer.

autotransformer starter—Also called a compensator or autostarter. A motor starter having an autotransformer to furnish a reduced voltage for starting. Includes the necessary switching mechanism.

auxiliary actuator—A mechanism which may be attached to a switch to modify its characteristics.

auxiliary circuit—Any circuit other than the main circuit.

auxiliary contacts—In a switching device, contacts, in addition to the main circuit contacts, which function with the movement of the latter.

auxiliary memory—*See* Auxiliary Storage.

auxiliary relay—1. A relay which responds to the opening or closing of its operating circuit, to assist another relay or device in the performance of a function. 2. A relay actuated by another relay and used to control secondary circuit functions such as signals, lights, or other devices. Also called slave relay.

auxiliary-station line filter—A line filter for use at repeated points to separate frequencies of different carrier systems, using the same line pair. For example, such a filter might be used at a high-frequency carrier-system repeater to bypass the low-frequency carrier system and voice frequencies around the repeater.

auxiliary storage—Also called auxiliary memory. Storage capacity in a digital computer, in addition to the main memory.

auxiliary switch—A switch actuated by some device such as a circuit breaker, for signaling, interlocking, or other purpose.

auxiliary transmitter—A transmitter held in readiness in case the main transmitter of a broadcasting station fails.

availability—1. The ratio, expressed as a percent, of the time during a given period that an equipment is correctly operating to the total time in that period. Also called operating ratio. 2. The probability that a system is operating satisfactorily at any point in time when used under stated conditions, where the total time considered includes operating time, active repair time, administrative time, and logistic time.

available conversion gain—Ratio of available output-frequency power from the output terminals of a transducer to the available input-frequency power from the driving generator, with terminating conditions specified for all frequencies which may affect the result. Applies to outputs of such magnitude that the conversion transducer is operating in a substantially linear condition.

available gain—The ratio of the available power at the output terminals of the network to the available power at the input terminals of the network.

available line—In a facsimile system, that portion of a scanning line which can be used for picture signals. Expressed as a percentage of the length of the scanning line.

available machine time—Time after the application of power during which a computer is operating correctly.

available power—1. The mean square of the open-circuit terminal voltage of a linear source, divided by four times the resistive component of the source impedance. 2. Of a network, the power that would be delivered to a conjugately matched load. It is the maximum power that a network can deliver. Available power, though defined in terms of an output load impedance, is independent of that impedance.

available power gain—Sometimes called completely matched power gain. Ratio of the available power from the output terminals of a linear transducer, under specified input-termination conditions, to the available power from the driving generator. The available power gain of an electrical transducer is maximum when the input-termination admittance is the conjugate of the driving-point admittance at the input terminals of the transducer.

available signal-to-noise ratio—Ratio of the available signal power at a point in a circuit, to the available random-noise power.

avalanche—Rapid generation of a current flow with reverse-bias conditions as electrons sweep across a junction with enough energy to ionize other bonds and create electron-hole pairs, making the action regenerative.

avalanche breakdown—In a semiconductor diode, a nondestructive breakdown caused by the cumulative multiplication of carriers through field-induced impact ionization.

avalanche conduction—A form of conduction in a semiconductor in which charged-particle collisions create additional hole-electron pairs.

avalanche diode—Also called breakdown diode. A silicon diode that has a high ratio of reverse-to-forward resistance until avalanche breakdown occurs. After breakdown the voltage drop across the diode is essentially constant and is independent of the current. Used for voltage regulating and voltage limiting. Originally called zener diode, before it was found that the Zener effect had no significant role in the operation of diodes of this type.

avalanche impedance—*See* Breakdown Impedance.

avalanche noise—A phenomenon in a semiconductor junction in which carriers in a high-voltage gradient develop sufficient energy to dislodge additional carriers through physical impact.

avc—*See* Automatic Volume Control.

average—*See* Arithmetic Mean.

average absolute pulse amplitude—The average of the absolute value of instantaneous amplitude taken over the pulse duration. Absolute value means the arithmetic value regardless of algebraic sign.

average brightness—The average illumination in a television picture.

average calculating operation—A typical computer calculating operation longer than an addition and shorter than a multiplication, often taken as the mean of nine additions and one multiplication.

average electrode current—The value obtained by integrating the instantaneous electrode current over an averaging time and dividing by the average time.

average life—*See* Mean Life.

average noise factor—*See* Average Noise Figure.

average noise figure—Also called average noise factor. In a transducer, the ratio of total output noise power to the portion attributable to thermal noise in the input termination, the total noise being summed over frequencies from zero to infinity and the noise temperature of the input termination being standard (290° K).

average power output of an amplitude-modulated transmitter—The radio-frequency power delivered to the transmitter output terminals, averaged over a modulation cycle.

average pulse amplitude—The average of the instantaneous amplitudes taken over the pulse duration.

average rate of transmission—Effective speed of transmission.

average value—The value obtained by dividing the sum of a number of quantities by the number of quantities. The average value of a sine wave is 0.637 times the peak value.

average voltage—The sum of the instantaneous voltages in a half-cycle waveshape, divided by the number of instantaneous voltages. In a sine wave, the average voltage is equal to 0.637 times the peak voltage.

aviation channels—A band of frequencies, below and above the standard broadcast band, assigned exclusively for aircraft and aviation applications.

aviation services—The aeronautical mobile and radionavigation services.

avionics—An acronym designating the field of AVIation electrONICS.

Avogadro's number—The number (N) of molecules in a gram-molecule or the number of atoms in a gram-atom of a substance. The number N = 6.02 × 10^{23} molecules.

AWG—American Wire Gage. A means of specifying wire diameter. The higher the number, the smaller the diameter.

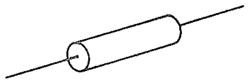

Axial leads.

axial leads—Leads coming out the ends and along the axes of a resistor, capacitor, or other axial part, rather than out the side.

axial ratio—Ratio of the major axis to the minor axis of the polarization ellipse of a wave guide. This term is preferred over ellipticity, because mathematically ellipticity is 1 minus the reciprocal of the axial ratio.

axis—The straight line, either real or imaginary, passing through a body around which the body revolves or around which parts of a body are symmetrically arranged.

Ayrton-Perry winding—Two conductors connected in parallel so that the current flows in opposite directions in each conductor and thus neutralizes the inductance between the two.

Ayrton shunt—Also called universal shunt. A high-resistance parallel connection used to increase the range of a galvanometer without changing the damping.

azel display—A modified type of plan-position indicator presentation showing two separate radar displays on one cathode-ray screen. One display presents bearing information, and the other shows elevation.

azimuth—The angular measurement in a horizontal plane and in a clockwise direction.

azimuth alignment—Alignment of the recording and reproducing gaps so that their center lines lie parallel with one another. Misalignment of the gaps causes a loss in output at short wavelengths.

azimuth gain reduction—A technique which allows control of the radar receiver system throughout any two azimuth sectors.

azimuth rate—The rate of change of true bearing.

azimuth resolution—The angle or distance by which two targets must be separated in azimuth to be distinguished by a radar set, when the targets are at the same range.

azimuth stabilization—The presentation of indications on a radar display so that north, or any specific reference line of direction, is always at the top of the screen.

azimuth-stabilized, plan-position indicator—A ppi scope on which the reference bearing (usually true or magnetic north) remains fixed with respect to the indicator, regardless of the vehicle orientation.

azimuth versus amplitude—An electronic counter-countermeasures receiver with plan-position indicator-type display attached to the main antenna, used to display strobes due to jamming aircraft. It is useful in making passive fixes when two or more radar sites can operate together.

azusa—A short base-line continuous-wave phase-comparison electronic tracking system operating on the C-band, wherein a single station provides two direction cosines and slant range.

B

B—1. Symbol for the base of a transistor. 2. Symbol for magnetic flux. 3. Abbreviation for photometric brightness. 4. B or b. Abbreviation for susceptance.

B− (B-minus or B-negative)—Negative terminal of a B battery or the negative polarity of other sources of anode voltage. Denotes the terminal to which the negative side of the anode-voltage source should be connected.

B+ (B-plus or B-positive)—Positive terminal of a B battery or the positive polarity of other sources of anode voltage. The terminal to which the positive side of the anode voltage source should be connected.

babble—1. The aggregate cross talk from a large number of disturbing channels. 2. In a carrier, or other multiple-channel system, the unwanted disturbing sounds which result from the aggregate cross talk or mutual interference from other channels.

babble signal—A type of electronic deception signal used to confuse enemy receivers. Generally it has characteristics of enemy transmission signals. It can be composed by superimposing incoming signals on previously recorded intercepted signals. This composite signal can then be radiated as a jamming signal.

BABS (Blind Approach Beacon System)—A pulse-type ground-based navigation beacon used for runway approach. The BABS ground beacon is installed beyond the far end of the runway on the extended center line. When interrogated by an aircraft, it retransmits two diverging beams, one of short and the other of long duration pulses. The beams are transmitted alternately, but because of the fast switching, the aircraft receives what appears to be a continuous transmission of both beams. The cathode-ray tube in the aircraft displays both long and short pulses superimposed on each other. When the aircraft is properly aligned with the runway, the pulses will be of equal amplitude.

back bias—1. A degenerative or regenerative voltage which is fed back to circuits before its originating point. Usually applied to a control anode of a tube. 2. A voltage applied to a grid of a tube (or tubes) to restore a condition which has been upset by some external cause. 3. *Also see* Reverse Bias.

backbone—A high-voltage, high-capacity transmission line or group of lines having a limited number of large-capacity connections between loads and points of generation.

back contact—Relay, key, jack, or other contact designed to close a circuit and permit current to flow when, in the case of a relay, the armature has released or fallen back, or in other cases, when the equipment is inoperative.

back current—Also called reverse current. The current which flows when reverse bias is applied to a semiconductor junction.

back diode—A tunnel diode that is usually chosen for its reverse-conduction characteristics.

back echo—An echo due to the back lobe of an antenna.

back emission—Also called reverse emission. Emission from an electrode occurring only when the electrode has the opposite polarity from that required for normal conduction. A form of primary emission common to rectifiers during the inverse portion of their cycles.

backfire—*See* Arcback.

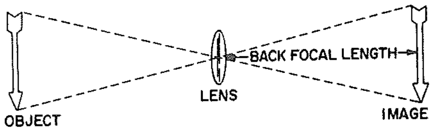

Back focal length.

back focal length—Distance from the center of a lens to its principal focus on the side of the lens away from the object.

background—1. The picture white of the copy being scanned when the picture is black and white only. Also undesired printing in the recorded copy of the picture being transmitted, resulting in shading of the background area. 2. Noise heard during a radio program; this noise is caused by atmospheric interference or operation of the receiver at such high gain that inherent tube and circuit noises become noticeable.

background control—In color television, a potentiometer used as a means of controlling the d-c level of a color signal at one input of a tricolor picture tube. The setting of this control determines the average (or background) illumination produced by the associated color phosphor.

background count—Count caused by radiation from sources other than the one being measured.

background noise—1. The total system noise, independent from the presence or absence of a signal. The signal is not included as part of the noise. 2. In a receiver, the noise in the absence of signal modulation in the carrier. 3. Any unwanted sound that intrudes upon program material, such as sounds produced as a result of surface imperfections of a disc record.

background radiation—Radiation due to the presence of radioactive material in the vicinity of the measuring instrument.

background response—In radiation detectors, response caused by ionizing radiation from sources other than that to be measured.

backlash—1. In a potentiometer, the maximum difference that occurs in shaft position when the shaft is moved to the same actual output-ratio point from opposite directions. Resolution and contact-width effects must be excluded from this measure. 2. The shortest distance between nondriving tooth surfaces of adjacent teeth in mating gears or the amount of clearance between the teeth or mating gears. Usually measured at the common pitch circle.

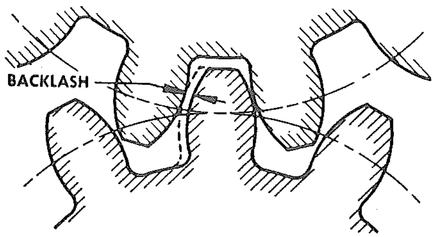

Backlash.

backloaded horn—A speaker enclosure arranged so the sound from the front of the cone feeds directly into the room, while the sound from the rear feeds into the room via a folded horn.

back loading—A form of horn loading particularly applicable to low-frequency speakers; the rear radiating surface of the speaker feeds the horn and the front part of the speaker is directly exposed to the room.

back lobe—In the radiation pattern of a directional antenna, that part which extends backward from the main lobe.

back pitch—The winding pitch of the back end of the armature—that is, the end opposite the commutator.

backplate—In a camera tube, the electrode to which the stored-charge image is capacitively coupled.

back porch—In a composite picture signal, that portion which lies between the trailing edge of a horizontal-sync pulse and the trailing edge of the corresponding blanking pulse. A color burst, if present, is not considered part of the back porch.

back-porch effect—The continuation of collector current in a transistor for a short time after the input signal has dropped to zero. The effect is due to storage of minority carriers in the base region. It also occurs in junction diodes.

back-porch tilt—The slope of the back porch from its normal horizontal position. Positive and negative refer respectively to upward and downward tilt to the right.

back scatter — Return, due to scatter, of a portion of electromagnetic radiation toward its source.

backscatter—Propagation of extraneous signals by F- or E- region reflection in addition to the desired ionospheric scatter mode. The undesired signal enters the antenna through the back lobes.

back scattering—1. Radiation of unwanted energy to the rear of an antenna. 2. The reflected radiation of energy from a target toward the illuminating radar.

back-shunt keying—A method of keying a transmitter, in which the radio-frequency energy is fed to the antenna when the telegraph key is closed and to an artificial load when the key is open.

backstop—That part of the relay which limits the movement of the armature away from the pole face or core. In some relays, a normally closed contact may serve as a backstop.

backswing—The amplitude of the first maximum excursion in the negative direction after the trailing edge of a pulse expressed as a percentage of the 100% amplitude.

back-to-back circuit — Two tubes or semiconductor devices connected in parallel but in opposite directions so that they can be used to control current without introducing rectification. Also called inverse-parallel connection.

backup—1. An item kept available to replace an item which fails to perform satisfactorily. 2. An item under development intended to perform the same general function performed by another item also under development.

backup facility—A communications-electronics facility which is established for the purpose of replacing or supplementing another facility or facilities, under real or simulated emergency conditions.

backup item—An additional item to perform the general functions of another item. It may be secondary to an identified primary item or a parallel development to improve the probability of success in performing the general function.

backwall—The plate in a pot core which connects the center post to the sleeve.

backwall photovoltaic cell—A cell in which light must pass through the front electrode and a semiconductor layer before reaching the barrier layer.

backward-acting regulator—A transmission regulator in which the adjustment made by the regulator affects the quantity which caused the adjustment.

backward diode — A highly doped, alloyed germanium junction that operates on the principle of quantum-mechanical tunneling. The diode is "backward" because its easy-current direction is in the negative-voltage rather than the positive-voltage region of the i-v curve. The backward diode has a negative-resistance region, but the resultant "valley" of its i-v curve is much less pronounced than in tunnel diodes.

backward wave—In a traveling-wave tube, a wave having a group velocity opposite the direction of electron-stream motion.

backward-wave oscillator—An oscillator employing a special vacuum tube in which oscillatory currents are produced by using an oscillatory electromagnetic field to bunch the electrons as they flow from cathode to anode.

backward-wave tube—A traveling-wave tube in which the electrons travel in a direction opposite to that in which the wave is propagated.

back wave—*See* Spacing Wave.

baffle—1. In acoustics, a shielding structure or partition used to increase the effective length of the external transmission path between two points (for example, between the front and back of an electroacoustic transducer). A baffle is often used in a speaker to increase the acoustic loading of the diaphragm. 2. In a gas tube, an auxiliary member placed in the arc path and having no separate external connection. 3. A device for deflecting oil or gas in a circuit breaker.

baffle plate—A metal plate inserted into a waveguide to reduce the cross-sectional area for wave-conversion purposes.

bail—A loop of wire used to prevent permanent separation of two or more parts assembled together—e.g., the bail holding dustcaps on round connectors.

Bakelite—A trademark of the Bakelite Corp. for its line of plastic and resins. Formerly, the term applied only to its phenolic compound used as an insulating material in the construction of radio parts.

balance—1. The effect of blending the volume of various sounds coming over different microphones in order to present them in correct proportion. 2. The maintenance of equal average volume from both speaker systems of a stereo installation.

balance control—On a stereo amplifier, a differential gain control used to vary the volume of one speaker system relative to the other without affecting the over-all volume level. As the volume of one speaker increases and the other decreases, the sound appears to shift from left to center to right, or vice versa.

balanced—1. Electrically alike and symmetrical with respect to ground. 2. Arranged to provide balance between certain sets of terminals.

balanced amplifier—An amplifier circuit with two identical signal branches, connected to operate in phase opposition and with their input and output connections each balanced to ground; for example, a push-pull amplifier.

balanced armature—An armature which is approximately in equilibrium with respect to both static and dynamic forces.

balanced-armature unit—The driving unit used in magnetic speakers, consisting of an iron armature pivoted between the poles of a permanent magnet and surrounded by coils carrying the audio-frequency current. Variations in the audio-frequency current cause corresponding changes in the armature magnetism and corresponding movements of the armature with respect to the poles of the permanent magnet.

balanced circuit—1. A circuit with two sides electrically alike and symmetrical to a common reference point, usually ground. 2. A circuit terminated by a network that has infinite impedance losses.

balanced converter—*See* Balun.

balanced currents — Also called push-pull currents. In the two conductors of a balanced line, currents which are equal in value and opposite in direction at every point along the line.

balanced detector—A demodulator for frequency-modulation systems. In one form, the output consists of the rectified difference of the two voltages produced across two resonant circuits, one circuit being tuned slightly above the carrier frequency and the other slightly below.

balanced line—A line or circuit utilizing two identical conductors. Each conductor is operated so that the voltages on them at any transverse plane are equal in magnitude and opposite in polarity with respect to ground. Thus, the currents on the line are equal in magnitude and opposite in direction. A balanced line is preferred where minimum noise and cross talk are desired.

balanced-line system—A system consisting of a generator, balanced line, and load adjusted so that the voltages of the two conductors at all transverse planes are equal in magnitude and opposite in polarity with respect to ground.

balanced low-pass filter — A low-pass filter designed to be used with a balanced line.

balanced method — A method of measurement in which the reading is taken at zero. It may be a visual or audible reading, and in the latter case the null is the no-sound setting.

balanced modulator—An amplitude modulator in which the control grids of two tubes are connected for parallel operation, and the screen grids and plates for push-pull operation. After modulation, the output contains the two sidebands without the carrier.

balanced network — A hybrid network in which the impedances of the opposite branches are equal.

balanced oscillator—Any oscillator in which (1) the impedance centers of the tank circuits are at ground potential and (2) the voltages between either end and the centers are equal in magnitude and opposite in phase.

balanced termination—For a system or network having two output terminals, a load presenting the same impedance to ground for each output terminal.

balanced transmission line—A transmission line having equal conductor resistances per unit length and equal impedances from each conductor to earth and to other electrical circuits.

balanced voltages — Also called push-pull voltages. On the two conductors of a balanced line, voltages (relative to ground) which are equal in magnitude and opposite in polarity at every point along the line.

balanced-wire circuit—A circuit with two sides electrically alike and symmetrical to ground and other conductors. Commonly refers to a circuit the two sides of which differ only by chance.

balancer—In a direction finder, that portion used for improving the sharpness of the direction indication. It balances out the capacitance effect between the loop and ground.

balance-to-unbalance transformer — A device for matching a pair of lines, balanced with respect to earth, to a pair of lines not balanced with respect to earth.

balancing network — An electrical network designed for use in a circuit in such a way that two branches of the circuit are made substantially conjugate (i.e., such that an electromotive force inserted into one branch produces no current in the other).

balancing unit—1. An antenna-matching device used to permit efficient coupling of a transmitter or receiver having an unbalanced output circuit to an antenna having a balanced transmission line. 2. A device for converting balanced to unbalanced transmission lines, and vice versa, by placing suitable discontinuities at the junction between the lines instead of using lumped components.

ball—In face bonding, a method of providing chips with contact.

ballast lamp—A lamp which maintains a nearly constant current by increasing its resistance as the current increases.

ballast resistor—A special type of resistor used to compensate for fluctuations in alternating-current power-line voltage. It is usually connected in series with the power supply to a receiver or amplifier. The resistance of a ballast resistor increases rapidly with increases in current through it, thereby tending to maintain an essentially constant current despite variations in the line voltage.

ballast tube—A current-controlling resistance device designed to maintain a substantially constant current over a specified range of variations in the applied voltage to a series circuit.

ball bonding—A type of thermocompression bonding which uses a flame to cut a wire as it leaves a capillary feed tube. The molten end of the wire solidifies into a ball, which is pressed against the bonding pad on an integrated circuit.

ballistic galvanometer—An instrument that indicates the effect of a sudden rush of electrical energy, such as the discharge current of a capacitor.

ballistic-missile early-warning system—An electronic system for providing detection and early warning of attack by enemy intercontinental ballistic missiles.

ballistic trajectory—In the trajectory of a missile, the curve traced after the propulsive force is cut off and the body of the missile is acted upon only by gravity, aerodynamic drag, and wind.

ballistocardiogram—A waveform of the impulse imparted to or incurred by the body as a result of the displacement of blood upon each heartbeat. The period of the cycle of this waveform is the time interval between heart beats. Typically taken in one of two ways—by measuring the seismic disturbance imparted to the table upon which the patient is lying supine, or by measuring the deflection of a heavy metal bar placed across the patient's ankles while he is lying supine.

balop—Contraction of balopticon, an apparatus for the projection of opaque images in conjunction with a television camera.

balopticon—*See* Balop.

balun — Also called balanced converter or "bazooka." An acronym from BALanced to UNbalanced. A device used for matching an unbalanced coaxial transmission line to a balanced two-wire system.

banana jack—A jack that accepts a banana plug. Generally designed for panel mounting.

banana plug—A plug with a banana-shaped spring-metal tip and with elongated springs to provide a low-resistance compression fit.

band—1. Any range of frequencies which lies between two defined limits. 2. A group of radio channels assigned by the FCC to a particular type of radio service:

Very low freq.	(vlf)	10-30 kHz.
Low freq.	(lf)	30-300 kHz.
Medium freq.	(mf)	300-3,000 kHz.
High freq.	(hf)	3-30 MHz.
Very high freq.	(vhf)	30-300 MHz.
Ultrahigh freq.	(uhf)	300-3,000 MHz.
Superhigh freq.	(shf)	3,000-30,000 MHz.
Extremely high freq.	(ehf)	30-300 GHz.

3. A group of tracks or channels on a magnetic drum in an electronic computer. (*Also see* Track.)

band center—The geometric mean between the limits of a band of frequencies.

banded cable — Two or more cables banded together by stainless-steel strapping.

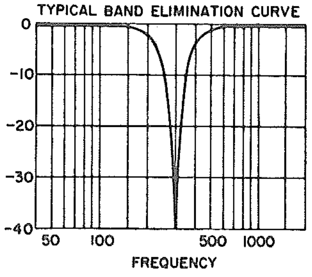

Band-elimination filter characteristic.

band-elimination filter—Also called band-stop filter. A wave filter with a single attenuation band, neither of the cutoff frequencies being zero or infinite. The filter passes frequencies on either side of this band.

bandpass—The number of hertz expressing the difference between the limiting frequencies at which the desired fraction (usually half-power) of the maximum output is obtained.

bandpass-amplifier circuit—A stage designed to uniformly amplify signals of certain frequencies only.

bandpass filter—A wave filter with a single transmission band, neither of the cutoff frequencies being zero or infinite. The filter attenuates frequencies on either side of this band.

bandpass flatness—The variations in gain in the bandpass of a filter or tuned circuit.

bandpass response—Also called flat-top response. The response characteristic in which a definite band of frequencies is transmitted uniformly.

band-rejection filter—A filter having a single attenuation band, with neither of the cutoff frequencies being zero or infinite.

band selector—Also called bandswitch. A switch used to select any one of the bands in which a receiver, signal generator, or transmitter is designed to operate.

B and S gage—Brown and Sharpe wire gage, where the conductor sizes rise in geometrical progression. Adopted as the American Wire Gage standard.

bandspreading—1. The spreading of tuning indications over a wide range to facilitate tuning in a crowded band of frequencies. 2. The method of double-sideband transmission in which the frequency band of the modulating wave is shifted upward so that the sidebands produced by modulation are separated from the carrier by a frequency at least equal to the bandwidth of the original modulating wave. In this way, second-order distortion products may be filtered out of the demodulator output.

bandspread tuning control—A separate tuning control provided on some short-wave receivers to spread the stations in a single band of frequencies over an entire tuning dial.

band-stop filter — *See* Band-Elimination Filter.

bandswitch—A switch used to select any one of the frequency bands in which an electronic apparatus may operate.

bandwidth—1. The range within the limits of a band. The width of a bandpass filter is generally taken as the limits between which its attenuation is not more than 3.0 decibels greater than its average attenuation throughout its passband. Also used in connection with receiver selectivity, transmitted frequency spectrum occupancy, etc. 2. In a given facsimile system, the difference in hertz between the highest and lowest frequency components required for the adequate transmission of the facsimile signals. 3. The least frequency interval of a wave, outside of which the power spectrum of a time-varying quantity is everywhere less than some specified fraction of its value at a reference frequency. 4. The range of frequencies of a device, within which its performance, with respect to some characteristic, conforms to a specified standard.

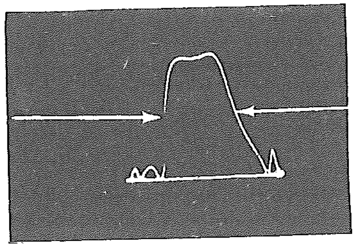

Bandwidth.

bang-bang controller — A discontinuity-type nonlinear system that contains time delay, dead space, and hysteresis.

bank—An aggregation of similar devices (e.g. transformers, lamps, etc.) connected and used together. In automatic switching, a bank is an assemblage of fixed contacts over which one or more wipers or brushes move to establish electric connections.

bank-and-wiper switch—A switch in which the electromagnetic ratchets or other mechanisms are used, first, to move the wipers to a desired group of terminals, and second, to move the wipers over the terminals of this group to the desired bank contacts.

bank winding—Also called banked winding. A compact multilayer form of coil winding used for reducing distributed capacitance. Single turns are wound successively in two or more layers, the entire winding proceeding from one end of the coil to the other without being returned.

bantam tube — A compact tube having a standard octal base but a considerably smaller glass envelope than the standard glass octal tube has.

bar—1. *See* Microbar. 2. A subdivision of a crystal slab. 3. A vertical or horizontal line on a television screen, used for testing.

bare conductor — A conductor not covered with any insulating material.

bar generator—A generator of pulses or repeating waves which are equally separated in time. These pulses are synchronized by the synchronizing pulses of a television system so that they produce a stationary bar pattern on a television screen.

bar-graph monitoring oscilloscope—An oscilloscope for observation of commutated signals appearing as a series of bars with lengths proportional to channel modulation. The same oscilloscope is commonly used for setup and troubleshooting observations.

barium—An element the oxide of which is used in the cathode coating of vacuum tubes.

barium titanate—A ceramic that has electric properties and is capable of withstanding much higher temperatures than Rochelle salt crystals. Used in crystal pickups and sonar transducers.

Barkhausen effect—A succession of abrupt changes which occur when the magnetizing force acting on a piece of iron or other magnetic material is varied.

Barkhausen interference — Interference caused by Barkhausen oscillations.

Barkhausen-Kurz oscillator — Circuit for generating ultrahigh frequencies. Its operation depends on the variation in the electrical field around the positive grid and less positive plate of a triode; the variation is caused by oscillatory electrons in the interelectrode spaces.

Barkhausen oscillation—A form of parasitic oscillation in the horizontal-output tube of a television receiver; it results in one or more narrow, dark, ragged vertical lines near the left side of the picture or raster.

Barkhausen oscillator—*See* Barkhausen-Kurz Oscillator.

Barkhausen tube—*See* Positive-Grid Oscillator Tube.

bar magnet—A bar of metal that has been so strongly magnetized that it holds its magnetism and thereby serves as a permanent magnet.

barn—A unit of measure of nuclear cross sections. Equal to 10^{-24} square centimeters.

Barnett effect—The magnetization resulting from the rotation of a magnetic specimen. The rotation of a ferromagnet produces the same effect as placing the ferromagnet in a magnetic field directed along the axis of rotation. On the macroscopic model, the domains of a ferromagnet can be considered a group of electron systems, each acting as an independent gyroscope or gyrostat.

barometer—An instrument for measuring atmospheric pressure. There are two types of barometers commonly used in meteorology—the mercury barometer and the aneroid barometer.

barometric pressure—The weight of the atmosphere per unit of surface. The standard barometer reading at sea level and 59° F is 29.92 inches of mercury absolute.

bar pattern—A pattern of repeating lines or bars on a television screen. When such a pattern is produced by pulses which are equally separated in time, the spacing between the bars on the television screen can be used to measure the linearity of the horizontal or vertical scanning systems.

bar quad—*See* B-quad.

barrage jamming—Simultaneous jamming of a number of adjacent channels or frequencies.

bar relay—A relay in which a bar actuates several contacts simultaneously.

barrel distortion—In camera or image tubes, the distortion which results in a monotonic decrease in radial magnification in the reproduced image away from the axis of symmetry of the electron optical system. In tv receivers, barrel distortion makes all four sides of the raster curve out like a barrel.

barretter—A voltage-regulator tube consisting of an iron-wire filament in a hydrogen-filled envelope. The filament is connected in series with the circuit to be regulated and maintains a constant current over a given voltage variation.

barretter mount — A wave-guide mount in which a barretter can be inserted to measure electromagnetic power.

barricade shield—A type of movable shield for protection from radiation.

barrier—A partition for the insulation or isolation of electric circuits or electric arcs.

barrier capacitance — *See* Depletion-Layer Capacitance.

barrier-film rectifier — A rectifier in which a film having unilateral (single-direction) conductivity is in contact with metal or other normally conducting plates.

barrier grid—A grid close to, or in contact with, a storage surface of a charge storage tube. It establishes an equilibrium voltage for secondary-emission charging and serves to minimize redistribution.

barrier layer—*See* Depletion Layer.

barrier-layer cell — A type of photovoltaic cell in which light acting on the surface of the contact between layers of copper and cuprous oxide causes an electromotive force to be produced.

barrier-layer rectification—*See* Depletion-Layer Rectification.

barrier region—*See* Depletion Region.

barrier shield—A wall or enclosure shielding the operator from an area where radioactive material is being used or processed by remote-control equipment.

barrier voltage — The voltage necessary to cause electrical conduction in a junction of two dissimilar materials, such as a P-N junction diode.

base—1. The region between the emitter and collector of a transistor which receives minority carriers injected from the emitter. It is the element which corresponds to the control grid of an electron tube. 2. In a vacuum tube, the insulated portion through which the electrodes are connected to the pins. 3. On a printed-circuit board, the portion that supports the printed pattern. 4. *Also see* Positional Notation and Radix. 5. *See* Alkali.

baseband—The frequency band occupied by the aggregate of the transmitted signals used to modulate a carrier. The term is commonly applied where the ratio of the upper to the lower limit of the frequency band is larger than unity.

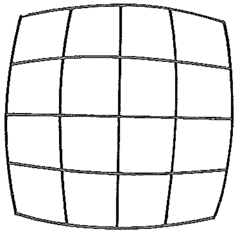

Barrel distortion.

base electrode—An ohmic or majority-carrier contact to the base region of a transistor.

base film — The plastic substrate that supports the coating of magnetic recording tape. The base film of most instrumentation and computer tapes is made of polyester. For less critical uses, cellulose acetate and polyvinyl chloride are employed.

basegroup—Designating a number of carrier channels in a particular frequency range that forms a basic unit (channel bank) for further modulation to a final frequency band.

base insulator — Heavy-duty insulator used to support the weight of an antenna mast and to insulate the mast from the ground of some other surface.

base line — 1. In radar displays, the visual line representing the track of the radar scanning beam. 2. In graphical presentations, the horizontal scale, often representing time, bias, or some other variable.

base-line break — In radar, a technique which uses the characteristic break in the base line on an A-scope display due to a pulse signal of significant strength in noise jamming.

base load—In a d-c converter, the current which must be taken from the base to maintain a saturated state.

base-loaded antenna—A vertical antenna the electrical height of which is increased by adding inductance in series at the base.

base material—An insulating material (usually a copper-clad laminate) used to support a conductive pattern.

base number—The radix of a number system (10 is the base number, or radix, for the decimal system; 2 for the binary system).

base pin—*See* Pin.

base point—*See* Radix Point.

base region—In a transistor, the interelectrode region into which minority carriers are injected.

base resistance—Resistance in series with the base lead in the common-T equivalent circuit of a transistor.

base ring—Ohmic contact to the base region of power transistors; so called because it is ring-shaped.

base spreading resistance—In a transistor, the resistance of the base region caused by the resistance of the bulk material of the base region.

base station—A land station, in the land mobile service, carrying on a service with land mobile stations. (A base station may secondarily communicate with other base stations incident to communications with land mobile stations.) Sometimes defined as a station in a land mobile system which remains in a fixed location and communicates with mobile stations.

base voltage—The voltage between the base terminals of a transistor and the reference point.

basic frequency—In any wave, the frequency which is considered the most important. In a driven system, it would in general be the driving frequency, while in most periodic waves it would correspond to the fundamental frequency.

basic linkage—In a computer, a linkage that is used repeatedly in one routine, program, or system and that follows the same set of rules each time it is used.

basic-*Q*—*See* Nonloaded *Q*.

basic rectifier—A metallic rectifier in which each rectifying element consists of a single metallic rectifier cell.

basic switch—A precision snap-acting switch of either the open type, skeleton type, or the type encased within an envelope. A basic switch is normally used within a protective enclosure or as a component.

basket winding—A coil winding in which adjacent turns are separated except at the points of crossing.

bass—Sounds in the low audio-frequency range. On the standard piano keyboard, all notes below middle *C* (256 hertz).

bass boost—A deliberate adjustment of the amplitude-frequency response of a system or component to accentuate the lower audio frequencies.

bass-boosting circuit—A circuit that attenuates the higher audio frequencies in order that low or bass frequencies will be emphasized by comparison.

bass compensation—Emphasizing the low-frequency response of an audio amplifier at low volume levels to compensate for the lowered sensitivity of the human ear to weak low frequencies.

bass control—A manual tone control that has the effect of changing the level of bass frequencies reproduced by an audio amplifier.

bass reflex—A method of extending the low-frequency range of a speaker and its response. Accomplished by an opening in the cabinet enclosing the speaker.

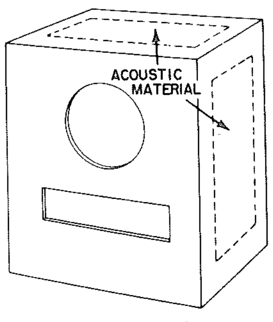

Bass-reflex enclosure.

bass-reflex enclosure—*See* Vented Baffle.

bass response—1. The extent to which a speaker or audio-frequency amplifier handles low audio frequencies. 2. The ability of any device to pick up or reproduce low audio frequencies.

batch process—A method of fabricating monolithic resistors, capacitors, and diodes with the same process at the same time.

batch processing—In a computer, a method of processing in which a number of similar input items are grouped for processing during the same machine run.

bat handle—Standard form of a toggle-switch lever having a shape similar to that of a baseball bat.

bathtub capacitor—A type of capacitor enclosed in a metal housing having broadly rounded corners like those on a bathtub.

bathyconductorgraph—A device used from a moving ship to measure the electrical conductivity of sea water at various depths.

bathythermograph—A device that automatically plots a graph showing temperature as a function of depth when lowered into the sea.

Batten system—A method developed by W. E. Batten for coordinating single words in a computer to identify a document. Sometimes called peek-a-boo system.

battery—A d-c voltage source consisting of two or more cells which converts chemical, nuclear, solar, or thermal energy into electrical energy.

battery acid—A solution that serves as the electrolyte in a storage battery. In the common lead-acid storage battery, the electrolyte is diluted sulphuric acid.

battery capacity—The amount of energy obtainable from a storage battery, usually expressed in ampere-hours.

battery charger—Device used to convert alternating current into a pulsating direct current which can be used for charging a storage battery.

battery clip—A metal clip with a terminal to which a connecting wire can be attached, and with spring jaws that can be quickly snapped onto a battery terminal or other point to which a temporary connection is desired.

battery receiver—A radio receiver that obtains its operating power from one or more batteries.

battle short—A switch for short-circuiting safety interlocks and lighting a red warning light.

bat wing—An element on an f-m or tv transmitting or receiving antenna; so called because of its shape.

baud—A unit of signaling speed derived from the duration of the shortest code element. Speed in bauds is the number of code elements per second.

bay—1. A portion of an antenna array. 2. A vertical compartment in which a radio transmitter or other equipment is housed.

bayonet base—A base having two projecting pins on opposite sides of a smooth cylindrical base; the pins engage corresponding slots in a bayonet socket and hold the base firmly in the socket.

bayonet coupling—A quick-coupling device. Connection is accomplished by rotating two parts under pressure. Pins on the side of the male connector engage slots on the side of the female connector.

bayonet socket—A socket for bayonet-base tubes or lamps; it has slots on opposite sides and one or more contact buttons at the bottom.

bazooka—*See* Balun.

B battery—The battery that furnishes the required d-c voltages to the plate and screen-grid electrodes of the vacuum tubes in a battery-operated circuit.

BCD—Abbreviation for Binary-Coded Decimal.

B channel—One of two stereo channels, usually the right, together with the microphone, speakers, or other equipment associated with this channel.

BCI—Abbreviation for BroadCast Interference, a term denoting interference by transmitters with reception of broadcast signals on standard broadcast receivers.

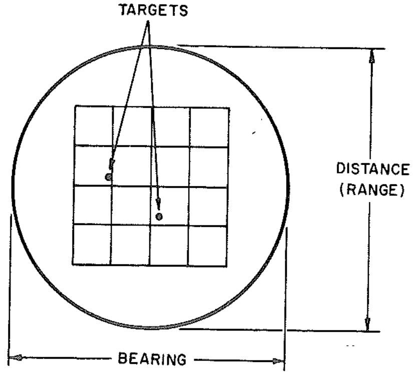

B-display.

B-display—On a radarscope, a type of presentation in which the target appears as a bright spot. Its bearing is indicated by the horizontal coordinate and its range by the vertical coordinate.

beacon—A device which emits a signal for use as a guidance or warning aid. Radar beacons aid the radar set in locating and identifying special targets which may otherwise be difficult or impossible to sense.

beacon delay—The amount of inherent delay within the beacon—i.e., the time between the arrival of a signal and the response of the beacon. In a pulse beacon, delay ordinarily is measured between the leading db points of the triggering pulse and the reply pulse.

beacon receiver—A radio receiver for converting into perceptible signals the waves emanating from a radio beacon.

beacon skipping—A term used to describe a condition where beacon return pulses are missing at the interrogating radar. Beacon skipping can be caused by interference, over-interrogation of the beacon, antenna nulls, or pattern minima.

beacon stealing—The loss of beacon tracking by a (desired) radar due to (interfering) interrogation signals from another radar.

beacon time sharing—A technique by which two or more radars may interrogate and track a long-recovery type of beacon without exceeding the duty cycle of the beacon. This technique is accomplished by the proper sequencing of the various radar interrogations. It is necessary to insure that the total of all interrogations does not exceed the beacon duty cycle and that enough time is allowed for the modulator section of the beacon to recover before it receives the next interrogation.

beaded coax—A coaxial cable in which the dielectric consists of beads made of various materials.

beaded support—Ceramic and plastic beads used to support the inner conductor in coaxial transmission lines.

beaded transmission line—A line using beads to support the inner conductor in coaxial transmission lines.

bead thermistor—A thermistor consisting of a small bead of semiconducting material such as germanium placed between two wire leads. Used for microwave power measurement, temperature measurement, and as a protective device. The resistance decreases as the temperature increases.

beam—1. A flow of electromagnetic radiation concentrated in a parallel, converging, or diverging pattern. 2. The unidirectional or approximately unidirectional flow of radiated energy or particles.

beam alignment—The adjustment of the electron beam in a camera tube (on tubes employing low-velocity scanning) to cause the beam to be perpendicular to the target at the target surface.

beam angle—The angle between the directions, on either side of the axis, at which the intensity of the radio-frequency field drops to one-half the value it has on the axis.

beam antenna—An antenna that concentrates its radiation into a narrow beam in a definite direction.

beam bender—*See* Ion-Trap.

beam bending—Deflection of the scanning beam of a camera tube by the electrostatic field of the charges stored on the target.

beam blanking—Interruption of the electron beam in a cathode-ray tube by the application of a pulse to the control grid or cathode.

beam candlepower—The candlepower of a bare source which, if located at the same distance as the beam, would produce the same illumination as the beam.

beam convergence—The converging of the three electron beams of a three-gun color picture tube at a shadow mask opening.

beam crossover—The point of overlap of a beam from an antenna that is mutated or rotated about the center line of the antenna radiation direction. The crossover point is normally at the 0.7 power point. The received energy, when commutated into four quadrants, provides the necessary information for the servoamplifier error signal used to align the antenna to a target.

beam-coupling coefficient—In a microwave tube, the ratio of the amplitude, expressed in volts, of the velocity modulation produced by a gap to the radio-frequency gap voltage.

beam current—The current carried by the electron stream that forms the beam in a cathode-ray tube.

beam cutoff—In a television picture tube or cathode-ray tube, the condition in which the control-grid potential is so negative with respect to the cathode that electrons cannot flow and thereby form the beam.

beam-deflection tube—An electron-beam tube in which current to an output electrode is controlled by the transverse movement of an electron beam.

beam-forming electrode—Electron-beam focusing elements in power tetrodes and cathode-ray tubes.

beam hole—An opening through a reactor shield and, generally, through the reactor reflector which permits a beam of radioactive particles or radiation to be used for experiments outside the reactor.

beam-index color tube—A color picture tube in which the signal generated by an electron beam after deflection is fed back to a control device or element in such a way that an image in color is provided.

beam-lead bonding—A free-bonding technique in which thick, gold extensions of the thin-film terminals of the semiconductor devices and circuits are electroformed so they extend beyond the edges of the chips.

beam modulation—*See* Z-Axis Modulation.

beam parametric amplifier—A parametric amplifier in which a modulated electron beam provides a variable reactance.

beam-positioning magnet—A magnet used with a tricolor picture tube to influence the direction of one of the electron beams so that it will have the proper spatial relationship with the other two beams.

beam-power tube—An electron-beam tube in which directed electron beams are used to contribute substantially to its power-handling capability, and in which the control and screen grids essentially are aligned.—*See* illustration on page 52.

beam relaxor—A type of sawtooth scanning-oscillator circuit which generates but does not amplify the current wave required for magnetic deflection in a single beam-power pentode.

beam-rider control system—A system whereby the control station sends a beam to

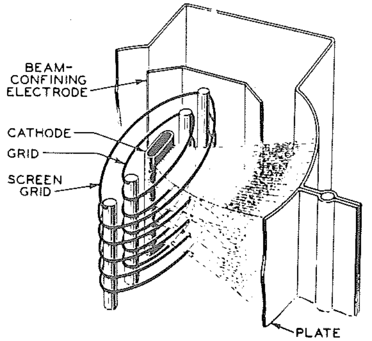

Beam-power tube.

the target, and the missile follows this beam until it collides with the target.

beam-rider guidance—A form of missile guidance wherein a missile, through a self-contained mechanism, automatically guides itself along a radar beam.

beam splitting—A process for increasing the accuracy in locating targets by radar. By noting the azimuths at which one radar scan first discloses a target and at which radar data from it ceases, beam splitting calculates the mean azimuth for the target.

beam spreader—An optical element the purpose of which is to impart a small angular divergence to a colliminated incident beam.

beam switching—A method of obtaining more accurately the bearing and/or elevation of an object by comparing the signals received when the beam is in a direction differing slightly in bearing and/or elevation. When these signals are equal, the object lies midway between the beam axes.

beam-switching tube—A multiposition, high-vacuum, constant-current distributor. The beam-switching tube consists of many identical "arrays" around a central cathode. Each array comprises (1) a spade which automatically forms and locks the electron beam, (2) a target-output electrode which gives the beam current its constant characteristics, and (3) a high-impedance switching grid which switches the beam from target to target. A small cylindrical magnet, permanently attached to the glass envelope, provides a magnetic field. This field, in conjunction with an applied electric field, comprises the crossed fields necessary for operation of this tube. It is used in electronic switching and in distributing such as counting, timing, sampling, frequency dividing, coding, matrixing, telemetering, and controlling.

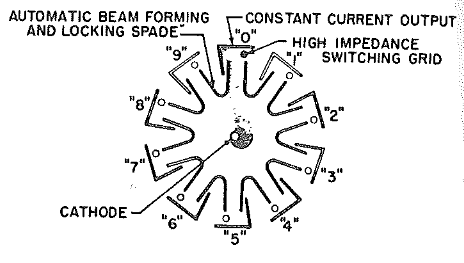
AUTOMATIC BEAM FORMING AND LOCKING SPADE
CONSTANT CURRENT OUTPUT
HIGH IMPEDANCE SWITCHING GRID
"0"
"1"
"2"
"3"
"4"
"5"
"6"
"7"
"8"
"9"
CATHODE

Beam-switching tube.

beam width—1. The angular width of a radio, radar, or other beam measured between two reference lines. 2. The width of a radar beam measured between lines of half-power intensity.

bearing—1. The horizontal direction of an object or point, usually measured clockwise from a reference line or direction through 360°. 2. Support for a rotating shaft. 3. The horizontal angle at a given point, measured from a specific reference datum, to a second point relative to another as measured from a specific reference datum.

bearing cursor—A mechanical bearing line of a plan-position indicator type of display for reading the target bearing.

bearing loss—The loss of power through friction in the bearings of an electric motor (brushes removed and no current in the windings).

bearing resolution—The minimum angular separation in a horizontal plane between two targets at the same range that will allow an operator to obtain data on either individual target.

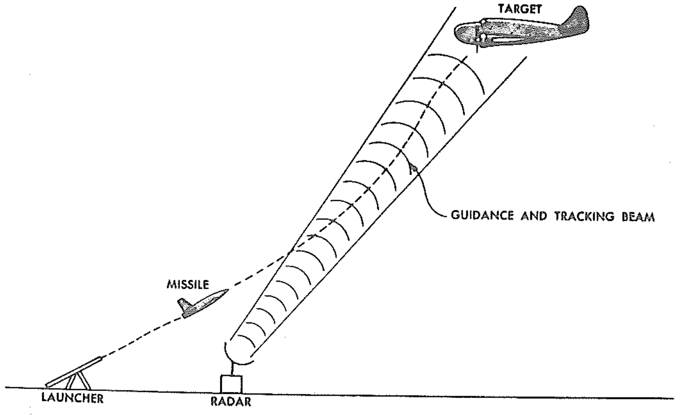

Beam-rider guidance.

beat—Periodic variations that result from the superimposition of waves having different frequencies. The term is applied both to the linear addition of two waves, resulting in a periodic variation of amplitude, and to the nonlinear addition of two waves, resulting in new frequencies, of which the most important usually are the sum and difference of the original frequencies.

beat frequency—One of the two additional frequencies produced when two different frequencies are combined. One beat frequency is the sum of the two original frequencies; the other is the difference between them.

beat-frequency oscillator — Abbreviated bfo. An oscillator that produces a signal which mixes with another signal to provide frequencies equal to the sum and difference of the combined frequencies.

beating—The combining of two or more frequencies to produce sum and difference frequencies called beats.

beating-in—Interconnecting two transmitter oscillators and adjusting one until no beat frequency is heard in a connected receiver. The oscillators are then at the same frequency.

beating oscillator—*See* Local Oscillator.

beat note—The wave of difference frequency created when two waves of different frequencies are combined. The beat note is equal to the difference in frequency between the two.

beat reception—*See* Heterodyne Reception.

beats—Beat notes which are generally at a sufficiently low audio frequency that they can be counted.

beat tone—Musical tone due to beats, produced by the heterodyning of two high-frequency wave trains.

beaver tail—A fan-shaped radar beam, wide in the horizontal plane and narrow in the vertical plane. The beaver tail is swept up and down for height finding.

bedspring—A broadside antenna array with a flat reflector.

before start—The interval before the starting circuit to a timer has been operated. The timer is fully reset and all contacts are in the precycle position.

bel—The fundamental unit in a logarithmic scale for expressing the ratio of two amounts of power. The number of bels is equal to the $\log_{10} P_1/P_2$, where P_1 is the power level being considered and P_2 is an arbitrary reference level. The decibel, equal to 1/10 bel, is a more commonly used unit.

B eliminator—A power pack that changes the a-c power-line voltage to the d-c source required by the vacuum tubes. In this way, batteries can be eliminated.

bell—An electrical device consisting of a hammer vibrated by an electromagnet. The hammer strikes the sides of the bell and emits a ringing noise. The electromagnet attracts an armature or piece of soft iron forming part of the hammer lever. A contact breaker then opens the circuit and cuts off the attraction. A spring draws the hammer back to its original position, closing the circuit and repeating the action.

Bellini-Tosi direction finder—An early radio direction-finder system consisting of two loop antennas at right angles to each other and connected to a goniometer.

bellows—A pressure-sensing element consisting of a ridged metal cylinder closed at one end. A pressure difference between its outside and inside will cause the cylinder to expand or contract along its axis.

bellows contact—A contact in which a multileaf spring is folded. This type provides a more uniform spring rate over the full tolerance range of the mating unit.

bell transformer—A small iron-core transformer; its primary coil is connected to an a-c primary line, and its secondary coil delivers 10 to 20 volts for operation of a doorbell, buzzer, or chimes.

bell wire—Cotton-covered copper wire, usually No. 18, used for doorbell and thermostat connections in homes and for similar low-voltage work.

bend—A change in the direction of the longitudinal axis of a waveguide.

bend waveguide—A section of waveguide in which the direction of the longitudinal axis is changed.

Benito—A c-w navigational system in which the distance to an aircraft is determined on the ground by measuring the phase difference of an audio signal transmitted from the ground and retransmitted by the aircraft. Bearing information is obtained by ground direction finding of the aircraft signals.

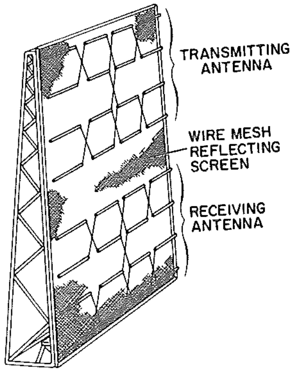

Bedspring antenna.

beta—Symbolized by the Greek letter *beta* (β). Also called current-transfer ratio. 1. The current transfer ratio (gain) of a transistor in the common-emitter circuit arrangement. This is the ratio of a-c collector current to a-c base current. 2. A symbol used to denote B quartz.

beta circuit—In a feedback amplifier, the circuit which transmits a portion of the amplifier output back to the input.

beta cutoff frequency — The frequency at which the beta of a transistor is 3 decibels below the low-frequency value.

beta particle—A small electrically charged particle thrown off by many radioactive materials. It is identical to the electron and possesses the smallest negative electrical charge found in nature. Beta particles emerge from radioactive material at high speeds, sometimes close to the speed of light.

beta ray—A stream of beta particles.

betatron—A large doughnut-shaped accelerator which produces artificial beta radiation. Electrons (beta particles) are whirled through a changing magnetic field. They gain speed with each trip and emerge with high energies (on the order of 100 million electron volts in some instances).

bev—A Billion Electron Volts. An electron possessing this much energy travels at a speed close to that of light—186,000 miles a second.

bevatron—A very large circular accelerator in which protons are whirled between the poles of a huge magnet to produce energies in excess of 1 billion electron volts.

Beverage antenna—*See* Wave Antenna.

beyond-the-horizon propagation—*See* Scatter Propagation.

beyond-the-horizon transmission—*See* Scatter Propagation.

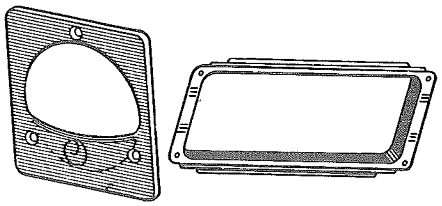

Bezel.

bezel—1. A holder designed to receive and position the edges of a lens, meter, window, or dial glass. 2. The flange or cover used for holding an external graticule or crt cover in front of the crt in an oscilloscope. May also be used for mounting a trace-recording camera or other accessory item.

bfo—Abbreviation for beat-frequency oscillator.

B-H curve—Curve plotted on a graph to show successive states during magnetization of a ferromagnetic material. A normal magnetization curve is a portion of a symmetrical hysteresis loop. A virgin magnetization curve shows what happens the first time the material is magnetized.

B-H meter—A device for measuring the intrinsic hysteresis loop of a sample of magnetic material.

bias—1. The electrical, mechanical, or magnetic force applied to a relay, vacuum tube, or other device for the purpose of establishing an electrical or mechanical reference level for the operation of the device. 2. Direct-current potential applied to the control grid of a vacuum tube. 3. Bias derived from a direct current, used on signaling or telegraph relays or electromagnets to secure the desired time spacing of transitions from marking to spacing. 4. A method of restraining a relay armature, by means of spring tension, to secure a desired time spacing of transitions from marking to spacing. 5. The average direct-current voltage between the control grid and cathode of a vacuum tube. 6. The effect on teletypewriter signals produced by the electrical characteristics of the line and the equipment. 7. Energy applied to a relay to hold it in a given position.

bias cell—A dry cell used in the grid circuit of a vacuum tube to provide the necessary C-bias voltage.

bias current—The current through the base-emitter junction of a transistor. It is adjusted to set the operating point of the transistor.

bias distortion—1. Distortion resulting from operation on a nonlinear portion of the characteristic curve of a vacuum tube or other device, due to improper biasing. 2. In teletype circuits, the uniform shifting of mark pulses from their proper position in relationship to the start pulses.

biased induction—Symbolized by B_b. The biased induction at a point in a magnetic material which is subjected simultaneously to a periodically varying magnetizing force and a biasing magnetizing force and is the algebraic mean of the maximum and minimum values of the magnetic induction at the point.

bias-induced noise—The difference between bulk-erased and zero-modulation noise.

biasing magnetizing force—Symbolized by H_b. A biasing magnetizing force at a point in a magnetic material which is subjected simultaneously to a periodically varying magnetizing force and a constant magnetizing force and is the algebraic mean of the maximum and minimum values of the combined magnetizing forces.

bias meter—Used in teletypewriter work for measuring signal bias directly in percent. A positive reading on the meter indicates a marking-signal bias; a negative reading indicates a spacing-signal bias.

bias oscillator—An oscillator used in magnetic recorders to generate an a-c signal in the range of 40 to 80 kHz for the purpose of magnetic biasing to obtain a linear recording characteristic. Usually the bias oscillator also serves as the erase oscillator.

bias resistor—A resistance connected into a self-biasing vacuum-tube circuit to produce the voltage drop necessary to provide a desired biasing voltage.

bias-set frequency—In direct magnetic tape recording, a specified recording frequency employed during the adjustment of bias level for optimum record performance (not the frequency of the bias).

bias telegraph distortion — Distortion in which all mark pulses are lengthened (positive bias) or shortened (negative bias). It can be measured with a steady stream of "unbiased reversals" (square waves having equal-length mark and space pulses). The average lengthening or shortening does not give true bias distortion unless other types of distortion are negligible.

bias windings—Control windings of a saturable reactor, by means of which the operating condition is translated by an arbitrary amount.

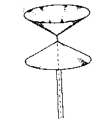

Biconical antenna.

biconical antenna—An antenna which is formed by two conical conductors, having a common axis and vertex, and excited at the vertex. When the vertex angle of one of the cones is 180°, the antenna is called a discone.

bidirectional—Responsive in opposite directions. An ordinary loop antenna is bidirectional because it has maximum response from the opposite directions in the plane of the loop.

bidirectional antenna—An antenna having two directions of maximum response.

bidirectional current—A current which is both positive and negative.

bidirectional microphone—A microphone in which the response predominates for sound incidences of 0° and 180°.

bidirectional pulses—Pulses, some of which rise in one direction and the remainder in the other direction.

bidirectional pulse train—A pulse train, some pulses of which rise in one direction and the remainder in the other direction.

bidirectional transducer — *See* Bilateral Transducer.

bidirectional transistor—A transistor which is specified with parameter limits in both the normal and inverted configuration and has substantially the same electrical characteristics when the terminals normally designed as emitter and collector are interchanged.

bifilar resistor — A resistor wound with a wire doubled back on itself to reduce the inductance.

bifilar transformer—A transformer in which the turns of the primary and secondary windings are wound together side-by-side and in the same direction. This type of winding results in near unity coupling, so that there is a very efficient transfer of energy from primary to secondary.

bifilar winding—A method of winding noninductive resistors in which the wire is folded back on itself and then wound double, with the winding starting from the point at which the wire is folded.

bifurcated contact — A movable contact which is forked (divided) to provide two contact-mating surfaces in parallel for a more reliable contact.

bilateral amplifier—An amplifier capable of receiving as well as transmitting signals; it is used primarily in transceivers.

bilateral antenna—An antenna, such as a loop, having maximum response in exactly opposite directions (180° apart).

bilateral bearing—A bearing which indicates two possible directions of wave arrival. One of these is the true bearing, and the other is a bearing displaced 180° from the true bearing.

bilateral circuit — A circuit wherein equipment at opposite ends is managed, operated, and maintained by different services.

bilateral element—A two-terminal element, the voltage-current characteristic of which has odd symmetry around the origin.

bilateral network—A network in which a given current flow in either direction results in the same voltage drop.

bilateral transducer — Also called bidirectional transducer. A transducer capable of transmission simultaneously in both directions between at least two terminations.

billboard antenna—Broadside array consisting of stacked dipoles with flat reflectors.

bimag—*See* Tape-wound Core.

bimetallic strip—A strip formed of two dissimilar metals welded together. Because the metals have different temperature coefficients of expansion, the strip bends or curls when the temperature changes.

bimetallic thermometer—A device containing a bimetallic strip which expands or contracts as the temperature changes. A calibrated scale indicates the amount of change in temperature.

bimorph cell—Two crystal elements (usually Rochelle salt) in rigid combination, arranged to act as a mechanical transformer in headphones, microphones, pickups, and speakers.

binary—1. A numbering system using a base number, or radix, of 2. In the binary system, there are only two digits—1 and 0. 2. Pertaining to a characteristic or property involving a selection, choice, or condition in which there are two possibilities. 3. Another range for bistable multivibrator.

binary cell—In an electronic computer, an elementary unit of storage which can be placed in either of two stable states.

binary chain — A series of binary circuits, each of which can exist in either one of two states, arranged so each circuit can affect or modify the condition of the next circuit.

binary channel—A transmission facility limited to the use of two symbols.

binary code—A code in which each code element is one of two distinct kinds or values—for example, the presence and absence of a pulse.

binary-coded—Expressed by a series of binary digits (0's and 1's).

binary-coded character—A decimal digit, alphabetic letter, punctuation mark, etc., represented by a fixed number of consecutive binary digits.

binary-coded decimal—Abbreviated BCD. A system of number representation in which each decimal digit of a number is expressed by binary numbers.

binary-coded digit—One element of a notation system for representing a decimal digit by a fixed number of binary positions.

binary-coded octal system—An octal numbering system in which each octal digit is represented by a three-place binary number.

binary counter—*See* Binary Scaler.

binary digit—*See* Bit.

binary magnetic core—A ring-shaped magnetic material which can be made to take either of two stable states of magnetic polarization.

binary notation—*See* Binary Number System.

binary number system—A number system using two symbols (usually denoted 0 and 1) and having 2 as its base, just as the decimal system uses 10 symbols (0, 1, . . . , 9) and has a base of 10. Also called binary notation.

binary numeral—The binary representation of a number—e.g., "101" is the binary numeral and "V" is the Roman numeral of the number of fingers on one hand.

binary point—The point which marks the place between integral powers of two and fractional powers of two in a binary number.

binary pulse-code modulation—A form of pulse-code modulation in which the code for each element of information consists of one of two distinct kinds, e.g., pulses and spaces.

binary scaler—Also called binary counter. 1. A counter which produces one output pulse for every two input pulses. 2. A counting circuit, each stage of which has two distinguishable states.

binary search—Also called dichotomizing search. A search in which a set of items is divided into two parts; one part is rejected, and the process is repeated on the accepted part until those items with the desired property are found.

binary signaling—A communications mode in which information is passed by the presence and absence, or plus and minus variations, of one parameter of the signaling medium.

binary system—A system of mathematical computation based on powers of 2.

binary-to-decimal conversion—The process of converting a number written in binary notation to the equivalent number written in the ordinary decimal notation.

binaural—Two-eared; applied to a kind of recording that is played into separate headphones to create a sense of depth.

binaural disc—A stereo record with two separate signals, one on each wall of its recorded grooves. Stereophonic sound is obtained by feeding each signal into its own speaker.

binaural effect—The effect which makes it possible for a person to distinguish the difference in arrival time or intensity of sound at his ears and thereby determine the direction from which a sound is arriving.

binaural recorder—A tape recorder which employs two separate recording channels, or systems, each with its own microphone, amplifier, recording and playback heads, and earphones. Recordings using both channels are made simultaneously on a single magnetic tape having two parallel tracks. During playback, the original sound is reproduced with a depth and realism unequalled by any other recording method. For a true binaural effect, headphones are necessary.

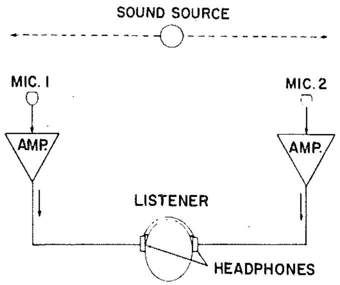

Binaural sound reproducing system.

binaural sound—Sound recorded or transmitted by pairs of equipment so as to give the listener the effect of having heard the original sound.

binder—A substance, like cement, used to hold particles together and thus provide mechanical strength in carbon resistors and phonograph records.

binding energy—The minimum energy required to dissociate a nucleus into its component neutrons and protons. Neutron or proton binding energies are those energies required to remove a neutron or a proton, respectively, from a nucleus. Electron binding energy is that energy required to remove an electron from an atom or a molecule.

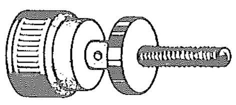

Binding post.

binding post—A bolt-and-nut terminal for making temporary electrical connections.

binistor—A four-terminal controlled rectifier semiconductor which provides bistable, negative-resistance characteristics.

binomial array—A directional antenna array used for reducing minor lobes and providing maximum response in opposite directions.

biochemical fuel cell—An electrochemical generator of electrical power in which bi-organic matter is used as the fuel source. In the usual electrochemical reaction, air serves as the oxidant at the cathode and microorganisms are used to catalyze the oxidation of the bi-organic matter at the anode.

bioelectrogenesis—The practical application of electricity drawn directly from the bodies of animals, including humans, to power electronic devices and appliances.

bioelectronics—The application of electronic theories and techniques to the problems of biology.

biological shield—A mass of absorbing material placed around a reactor or radioactive source to reduce the radiation to a level that is safe for human beings.

biometrics—The science of statistics as applied to biological observations.

bionics—1. The science of applying biological techniques to the design of electronic devices and systems. 2. The science of systems which function in a manner resembling living systems.

biotelemetry—The technique of measuring or monitoring vital processes and transmitting data without wires to a point (or points) remote from the subject.

biotelescanner—A device that can analyze and radio data on life forms during space exploration.

bipolar—Having two poles.

bipolar electrode—An electrode without metallic connection with the current supply, one face of which acts as an anode surface and the opposite face as a cathode surface when an electric current is passed through a cell.

bipolar magnetic driving unit—A headphone or speaker unit having two magnetic poles acting directly on a flexible iron diaphragm.

bipolar transistor—A transistor that uses both negative and positive charge carriers.

biquinary notation—A numerical system of notation based on a double scale of five.

biscuit—*See* Preform.

bistable—1. A circuit element with two stable operating states—e.g., a flip-flop in which one transistor is saturated while the other is turned off. It changes state for each input pulse or trigger. 2. A device capable of assuming either one of two stable states.

bistable multivibrator (flip-flop)—A circuit having two stable states; it will stay in either one indefinitely until appropriately triggered after which it immediately switches to the other state.

bistable relay—A relay that requires two pulses to complete one cycle composed of two conditions of operation. Also called locked, interlocked, and latching relay.

biswitch—A two-terminal integrated device which basically performs the function of two PNPN switches interconnected so as to provide bilateral switching.

bit—Abbreviation for binary digit. A unit of information equal to one binary decision, or the designation of one of two possible and equally likely values or states (such as 1 or 0) of anything used to store or convey information. It may also mean "yes" or "no."

bit density—The number of bits of information contained in a given area, such as the number of bits written along an inch of magnetic tape.

bit rate—The speed of transmission of bits.

black—A signal produced at any point in a facsimile system by the scanning of a selected area of subject copy having maximum density.

black and white—*See* Monochrome.

black and white transmission—*See* Monochrome Transmission.

black area—An area with encrypted signal only present.

black-body—1. An idealized emitter for which total radiated energy and the spectral distribution of the energy are accurately known functions of temperature. These relationships are given by Planck's law. 2. A solid that radiates or absorbs energy with no internal reflection of the energy at any wavelength. Physically it may be a hollow sphere coated on the inside with lampblack and with an opening through which energy may enter or leave.

black box—1. A term used loosely to refer to any subcomponent that is equipped with connects and disconnects so that it can be readily inserted into or removed from a specified place in a larger system (e.g., the complete missile or some major subdivision) without benefit of knowledge of its detailed

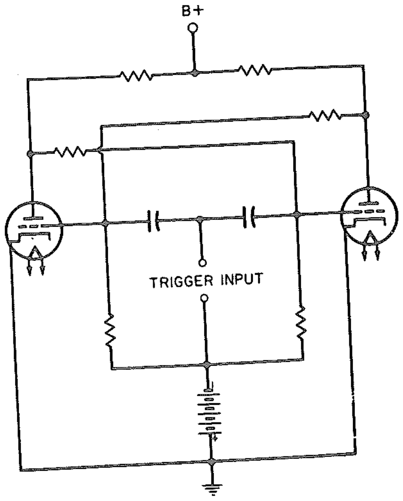

Bistable multivibrator.

internal structure. 2. A term pertaining to either the functional transformation that acts upon a specified input to give a particular output or to the apparatus for accomplishing this transformation (without regard to the detailed circuitry used). 3. A useful mathematical approach to an electronic circuit which concerns itself only with the input and output, and ignores the interior elements, discrete or integrated.

black compression—Also called black saturation. The reduction in the gain of a television picture signal at those levels corresponding to dark areas in the picture with respect to the gain at that level corresponding to the midrange light value in the picture. The over-all effect of black compression is to reduce contrast in the low lights of the picture.

blacker-than-black—The amplitude region of the composite video signal below the reference black level in the direction of the synchronizing pulses.

blacker-than-black level—A voltage value used in an electronic television system for control impulses. It is greater than the value representing the black portions of the image.

black level—That level of the picture signal corresponding to the maximum limit of black peaks.

black light—Invisible light radiation. May be either ultraviolet or infrared radiation, both of which are invisible.

black-light emitter—A source of electromagnetic radiation in the ultraviolet or infrared region, just outside the visible spectrum.

black peak—A peak excursion of the picture signal in the black direction.

black saturation—*See* Black Compression.

black signal—Also called picture black. A signal produced at any point in a facsimile system by the scanning of a maximum-density area of the subject copy.

blackout—1. Interruption of radiocommunication due to excess absorption caused by solar flares. During severe blackouts, all frequencies above approximately 1500 kHz are absorbed excessively in the daylight zone. 2. Passive defense that consists of interrupting all forms of communication or identification.

black transmission—1. In an amplitude-modulation facsimile system a form of transmission in which the maximum transmitted power corresponds to the maximum density of the copy. 2. In a frequency-modulation system, a form of transmission, in which the lowest transmitted frequency corresponds to the maximum density of the copy.

blank—1. The result of the final operation on a crystal. 2. To cut off the electron beam of a cathode-ray tube. 3. A code character to denote the presence of no information rather than the absence of information. In the Baudot code, it is composed of all spacing pulses. In paper tape, it is represented by a feedhole without intelligence holes.

blanked picture signal—The signal resulting from adding blanking to a picture signal. Adding the sync signal to the blanked picture signal forms the composite picture signal.

blanketing—The overriding of a signal by a more powerful one or by interference, so that a receiver is unable to receive the desired signal.

blank groove—*See* Unmodulated Groove.

blanking—The process of making a channel or device noneffective for the desired interval. In television, blanking is the substitution for the picture signal, during prescribed intervals, of a signal the instantaneous amplitude of which is such as to make the return trace invisible. *See also* Gating.

blanking level—Also called pedestal level. In a composite picture signal, the level that separates the range of the composite picture signal containing picture information from the range containing synchronizing information.

blanking pulse—A square wave (positive or negative) used to switch off electronically a part of a television or radar set for a predetermined length of time.

blanking signal—A wave made up of recurrent pulses related in time to the scanning process and used to effect blanking. In television, this signal is composed of pulses at line and field frequencies, which usually originate in a central sync generator and are combined with the picture signal at the pickup equipment in order to form the blanked picture signal. The addition of a sync signal completes the composite picture signal.

blanking time—The length of time the electron beam of a cathode-ray tube is cut off.

blanking zone—*See* Blanking Pulse.

blank tape—*See* Raw Tape.

blasting—Overloading of an amplifier or speaker, resulting in severe distortion of loud sounds.

bleeder current—The current drawn continuously from a power supply by a resistor. Used to improve the voltage regulation of the power supply.

bleeder resistor—1. A resistor used to draw a fixed current. Also used, as a safety measure, to discharge a filter capacitor after the circuit is de-energized. 2. A resistor placed in the power supply of a radio receiver or other electronic device to stabilize the voltage supply.

bleeding whites—An overloading condition in which white areas in a television picture appear to flow into the black areas.

blemish—On the storage surface of a charge-storage tube, an imperfection which produces a spurious output.

blind approach—An aircraft landing approach when visibility is poor, usually made with the aid of instruments and radiocommunication.

blind zone—An area from which echoes cannot be received; generally, an area shielded from the transmitter by some natural obstruction and therefore from which there can be no return.

blip—Sometimes referred to as pip. On a cathode-ray display, a spot of light or a base-line irregularity representing the radar reflection from an object.

blip-scan ratio—The ratio between a single recognizable blip on a radar scope and the number of scans necessary to produce it. The blip-scan ratio of any given radar set varies with the range, antenna tilt, level of operator and set performance, target aspect, wind, etc.

blivet—*See* Land.

Bloch wall—The transition layer separating adjacent ferromagnetic domains.

block—A set of things, such as words, characters, or digits, handled as a unit by a computer.

block diagram—A diagram in which the essential units of any system are drawn in the form of blocks, and their relationship to each other is indicated by appropriately connected lines. The path of the signal or energy may be indicated by lines and/or arrows.

blocked impedance—The input impedance of a transducer when its output is connected to a load of infinite impedance.

blockette—In digital computer programming, a subgroup, or subdivision, of a group of consecutive machine words transferred as a unit.

block-grid keying—A method of keying a continuous-wave transmitter by operating the amplifier stage as an electronic switch. During the spacing interval when the key is open, the bias on the control grid becomes highly negative and prevents the flow of plate current so that the tube has no output. During the marking interval when the key is closed, this bias is removed and full plate current flows.

blocking—Application of an extremely high bias voltage to a transistor, vacuum tube, or metallic rectifier to prevent current from flowing in the forward direction.

blocking capacitor—A capacitor which introduces a comparatively high series impedance for limiting the flow of low-frequency alternating or direct current without materially affecting the flow of high-frequency alternating current.

blocking layer—*See* Depletion Layer.

blocking oscillator—Also called squegging oscillator. An electron-tube oscillator that operates intermittently as its grid bias increases during oscillation to a point where the oscillations stop, and then decreases until oscillation resumes.

blocking-oscillator driver—A circuit which develops a square pulse used to drive the modulator tubes, and which usually contains a line-controlled blocking oscillator that shapes the pulse into the square wave.

blocking relay—A device which initiates a pilot signal to block tripping on external faults in a transmission line or in other apparatus under predetermined conditions, or which co-operates with other devices to block tripping or to block reclosing on an out-of-step condition or on power swings.

block protector—A rectangular piece of carbon, *Bakelite* with a metal insert, or porcelain with a carbon insert which, in combination with each other, make one element of a protector. They form a gap which will break down and provide a path to ground for excessive voltages.

block sort—A computer sorting technique in which the file is first divided according to the most significant character of the key, and the separate portions are then sorted one at a time. It is used particularly for large files.

blooming—An increase in the size of the scanning spot on a cathode-ray tube, caused by defocusing when the brightness control is set too high. The result is expansion and consequent distortion of the image. May also be caused by insufficient high voltage.

blow—The opening of a circuit because of excessive current, particularly when the current is heavy and a melting or breakdown point is reached.

blower—An electric fan used to supply moving air for cooling purposes.

blown jacket—The common term given to an outer covering of insulation of a cable that was applied by the controlled inflation of the cured jacket tube and the pulling of the cable through it.

blowout coil—An electromagnetic device used to establish a magnetic field in the space where an electrical circuit is broken and thus displace and extinguish the arc.

blowout magnet—A strong permanent magnet or electromagnet used for reducing or deflecting the arc between electrodes or contacts.

blue-beam magnet—A small permanent magnet used to adjust the static convergence of the electron beam for blue phosphor dots in a three-gun color picture tube.

blue glow—The glow normally seen in vacuum tubes containing mercury vapor; it is due to ionization of the molecules of mercury vapor.

blue gun—In a three-gun, color picture tube, the electron gun whose beam strikes the phosphor dots emitting the blue primary color.

blue restorer—The d-c restorer in the blue channel of a three-gun color-television picture-tube circuit.

blue video voltage—The signal voltage that controls the grid of the blue gun in a three-gun picture tube. This signal is a reproduction of the blue output signal of the color-tv camera.

bobbin—A small insulated spool which serves as a support for a coil or wirewound resistor.

bobbin core—*See* Tape-wound Core.

body capacitance — Capacitance introduced into an electrical circuit by the proximity of the human body.

body electrodes—Electrodes placed on or in the body to couple electrical impulses from the body to an external measuring or recording device.

boffle — A speaker enclosure, developed by H. A. Hartley, containing a group of stretched, resilient sound-absorbing screens.

bogey—The average, or published, value for a tube characteristic. A bogey tube would be one having all characteristics of a bogey value.

bogey electron device—An electron device the characteristics of which have the published nominal values for the type.

boiling point—The temperature at which a liquid vaporizes when heated. The exact point depends on the absolute pressure at the liquid-vapor surface.

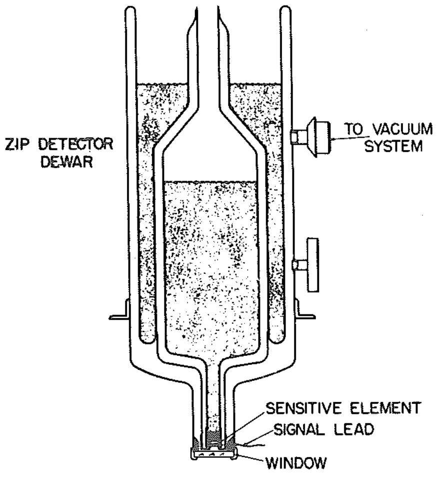

Superconducting bolometer.

bolometer—A radiation detector that converts incident radiation into heat which, in turn, causes a temperature change in the material used in the detector. This change is then measured to give an indication of the amount of incident radiant energy.

bolted fault level—*See* Adjusted Circuit.

bombardment — 1. The directing of high-speed electrons at an electrode, causing secondary emission of electrons, fluorescence, disintegration, or the production of X rays. 2. The process of directing high-speed particles at atoms to cause ionization or transmutation.

bond—1. Electrical interconnection made with a low-resistance material between a chassis, metal shield cans, or cable shielding braid, in order to eliminate undesirable interaction and interference resulting from high-impedance paths between them. 2. *See* Valence Bond.

bonded assembly—An assembly, the supporting frame and metallic noncircuit elements of which are connected so as to be electrically shorted together.

bonded-barrier transistor — A transistor made by alloying the base with the alloying material on the end of a wire.

bonded NR diode—An N junction semiconductor device in which the negative resistance arises from a combination of avalanche breakdown and conductivity modulation due to the current through the junction.

bonded pickup—*See* Bonded Transducer.

bonded strain gage—A pressure transducer that uses a pressure-sensing system consisting of strain-gage elements firmly bonded to a pressure-responsive member. Thermal stability and insensitivity to shock and vibration are improved by means of this bonded construction.

bonded transducer—Also called bonded pickup. A transducer which employs the bonded strain-gage principle of transduction.

bonding—1. Soldering or welding together various elements, shields, or housings of a device to prevent potential differences and possible interference. 2. A method used to produce good electrical contact between metallic parts of any device. Used extensively in automobile and aircraft to prevent static buildup. Also refers to the connectors and straps used to bond equipment. 3. The means employed to obtain an electromagnetically homogenous mass having an equipotential surface.

bonding conductor—A conductor which serves to connect exposed metal surfaces together.

bond strength—A measure of the amount of stress required to separate a layer of material from the base to which it is bonded. Peel strength, measured in pounds per inch of width, is obtained by peeling the layer; pull strength, measured in pounds per square inch is obtained by a perpendicular pull applied to a surface of the layer.

bone conduction — The process by which sound is conducted to the inner ear through the cranial bones.

book capacitor—A two-plate trimmer capacitor that has its plates hinged together like the pages of a book. The capacitance is varied by changing the angle between the plates.

Boolean algebra—A system of mathematical logic dealing with classes, propositions, on-off circuit elements, etc. Associated by operators as AND, OR, NOT, EXCEPT, IF . . . THEN, etc., thereby permitting computations and demonstration as in any mathematical system. Named after George Boole, famous English mathematician and logician, who introduced it in 1847.

Boolean calculus—Boolean algebra modified to include time.

Boolean function—A mathematical function in Boolean algebra.

boom—A mechanical support for a microphone, used in a television studio to suspend the microphone within range of the actors' voices, but out of camera range.

boost capacitor — A capacitor used in the damper circuit of a television receiver to supply a boosted B voltage.

boost charge—The partial charge of a storage battery, usually at a high current rate for a short period.

boosted B voltage — In television receivers the voltage resulting from the combination of the B-plus voltage from the power supply and the average value of voltage pulses coming through the damper tube from the horizontal deflection-coil circuit. The pulses are partially or wholly smoothed by filtering. This boosted voltage may be several hundred volts higher than the B-plus voltage.

booster — 1. A carrier-frequency amplifier, usually a self-contained unit, connected between the antenna or transmission line and a television or radio receiver. 2. An intermediate radio or tv station which retransmits signals from one fixed station to another. 3. A small, self-contained transformer designed to be connected to a cathode-ray tube socket to increase the filament voltage and thereby extend the life of the tube.

booster amplifier—A circuit used to increase the output current or the voltage capabilities of an operational amplifier circuit without loss of accuracy (ideally) or inversion of polarity. Usually applied inside the loop for accuracy.

booster voltage—The additional voltage supplied by the damper tube of a television receiver to the horizontal-output, horizontal-oscillator, and vertical-output tubes, resulting in a greater sawtooth sweep output.

boot—1. A form placed around the wire termination of a multiple-contact connector for the purpose of containing the liquid potting compound until it hardens. 2. A protective housing, usually made from a resilient material, used to protect connector or other terminals from moisture.

bootstrap circuit—A single-stage amplifier in which the output load is connected between the negative end of the plate supply and the cathode, the signal voltage being applied between the grid and cathode. The name "bootstrap" arises from the fact that the change in grid voltage also changes the potential of the input source (with respect to ground) by an amount equal to the output signal.

boresight error—The angular deviation of the electrical boresight of an antenna from its reference boresight.

boss—*See* Land.

bottom metalization — The metalization which may be provided over the back portion of an uncased IC chip facilitating its face-up attachment.

bounce—An unnatural, sudden variation in the brightness of a television picture.

boundary—An interface between P and N material at which donor and acceptor concentrations are equal.

boundary defect—In a crystal, the boundary area between two adjacent perfect crystal regions that are tilted slightly with respect to each other.

bound charge—On a conductor, the charge which, owing to the inductive action of a neighboring charge, will not escape to ground; residual charge.

bound circuit — A circuit designed to limit the excursion of a signal. The limit value it establishes may be nominal when used for protection, or highly precise when used operationally.

bound electron—An electron bound to the nucleus of an atom by electrostatic attraction.

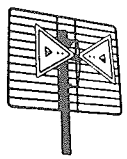

Bow-tie antenna.

bow-tie antenna—An antenna generally used for uhf reception. It consists of two triangles in the same plane, with a reflector behind them. The transmission line is connected to the points which form a gap.

boxcar circuit—A circuit used in radar to sample voltage waveforms and store the latest value sampled.

boxcar lengthener—A circuit that lengthens a series of pulses without changing their heights.

boxcars—Long pulses that are separated by very short spaces.

B power supply—A power supply which provides the plate and screen voltages applied to a vacuum tube.

B-quad—A quad arrangement similar to the S-quad except for a short between the junctions of the two sets of series elements. Also called bridge quad or bar quad.

Bragg's law—An expression of the conditions under which a system of parallel atomic layers in a crystal will reflect an X-ray beam with maximum intensity.

braid—1. A weave of organic or inorganic fiber used as a covering for a conductor or group of conductors. 2. A woven metal tube used as shielding around a conductor or group of conductors. When flattened, it is used as a grounding strap.

braided wire—A flexible wire made up of small strands woven together.

brain waves—The patterns of lines produced on the moving chart of an electroencephalograph as the result of electrical potentials produced by the brain, picked up by electrodes, and amplified in the machine.

braking magnet—*See* Retarding Magnet.

branch—1. In an electronic network, a section between two adjacent branch points. 2. A portion of a network consisting of one or more two-terminal elements in series.

branch circuit—In a wiring system, that portion extending beyond the final overcurrent device protecting the circuit.

branch current—The current in the branch of a network.

branch impedance—In a passive branch, the impedance obtained by assuming a driving force across and a corresponding response in the branch, no other branch being electrically connected to the one under consideration.

branching—In a computer, a method of selecting, on the basis of the computer results, the next operation to execute while the program is in progress.

branch order—An instruction used to link subroutines into the main program of a computer.

branch point—1. In an electric network, the junction of more than two conductors. *(Also see* Node.) 2. In a computer, a point in the routine where one of two or more choices is selected under control of a routine.

branch voltage—The voltage across a branch of a network.

brass pounder—Amateur term for a Morse code operator, especially one who spends long hours handling "traffic."

breadboard construction—An arrangement in which electronic components are fastened temporarily to a board for experimental work.

breadboard model — An assembly in rough form to prove the feasibility of a circuit, device, system, or principle.

break—1. An open circuit. 2. An interruption in radio transmission to permit transmission from the other end.

break-before-make contacts — Contacts which interrupt one circuit before establishing another.

break contact — In a switching device, the contact which opens a circuit upon operation of the device (normally closed contact).

breakdown—1. An electric discharge through an insulator, insulation on wire, or other circuit separator, or between electrodes in a vacuum or gas-filled tube. 2. The phenomenon occurring in a reverse-biased semiconductor diode. The start of the phenomenon is observed as a transition from a region of high dynamic resistance to one of substantially lower dynamic resistance. This is done to boost the reverse current.

breakdown diode—*See* Avalanche Diode.

breakdown impedance—Also called avalanche impedance. The small-signal impedance at a specified direct current in the breakdown region of a semiconductor diode.

breakdown region—The entire region of the volt-ampere characteristic beyond the initiation of breakdown due to reverse current in a semiconductor-diode characteristic curve.

breakdown strength—*See* Dielectric Strength.

breakdown torque—The maximum torque a motor will develop, without an abrupt drop in speed, as the rated voltage is applied at the rated frequency.

breakdown voltage — 1. That voltage at which an insulator or dielectric ruptures, or at which ionization and conduction take place in a gas or a vapor. 2. The voltage measured at a specified current in the breakdown region of a semiconductor diode. Also called zener voltage. 3. The voltage required to jump an air gap.

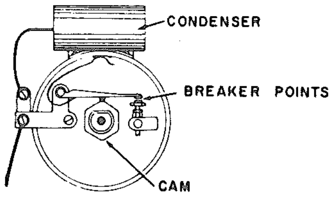

Breaker points.

breaker points — The low-voltage contacts that interrupt the current in the primary circuit of the ignition system of a gasoline engine.

break frequency—In a plot of log gain (attenuation) versus log frequency, the frequency at which the asymptotes of two adjacent linear slope segments meet.

break-in keying—In the operation of a radiotelegraph communication system, a method by which the receiver is capable of receiving signals during transmission spacing intervals.

break-in operation—A method of radiocommunication involving break-in keying which allows the receiving operator to interrupt the transmission.

break-in relay — A relay used for break-in operation.

break-make contact — Also called transfer contact. A contact form in which one contact opens its connection to another contact and then closes its connection to a third contact.

breakout—The exit point of a conductor or number of conductors along various points of a main cable of which the conductors are a part. This point is usually harnessed or sealed with some synthetic-rubber compound.

breakover voltage — The value of positive anode voltage at which an scr with the gate circuit open switches into the conductive state.

breakpoint—A place in a routine specified by an instruction, instruction digit, or other condition, where the routine may be interrupted by external intervention or by a monitor routine.

breakpoint instruction — In the programming of a digital computer an instruction which, together with a manual control, causes the computer to stop.

breakpoint switch — A manually operated switch which controls conditional operation at breakpoints; it is used primarily in debugging.

breakup—*See* Color Breakup.

breathing—Amplitude variations similar to "bounce," but at a slow, regular rate.

breezeway—In the NTSC color system, that portion of the back porch between the trailing edge of the sync pulse and the start of the color burst.

B-register—A computer register that stores a word which will change an instruction before the computer carries out that instruction.

bremsstrahlung — Electromagnetic radiation emitted by a fast-moving charged particle (usually an electron) when it is slowed down (or accelerated) and deflected by the electric field surrounding a positively charged atomic nucleus. X rays produced in ordinary X-ray machines are bremsstrahlung. (In German, the term means "braking radiation.")

brevity code—A code which has as its sole purpose the shortening of messages rather than the concealment of their content.

Brewster angle—The angle of incidence at which the reflection of parallel-polarized electromagnetic radiation at the interface between two dielectric media equals zero.

bridge—1. In a measuring system, an instrument in which part or all of a bridge circuit is used to measure one or more electrical quantities. 2. In a fully electronic stringed instrument, the part that converts the mechanical vibrations produced by the strings into electrical signals.

bridge circuit—A network arranged so that, when an electromotive force is present in one branch, the response of a suitable detecting device in another branch may be zeroed by suitable adjustment of the electrical constants of still other branches.

bridged tap—The portion of a cable pair connected to a circuit which is not a part of the useful path.

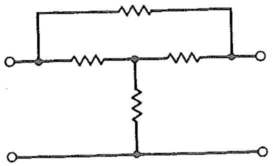

Bridged-T network.

bridged-T network—A T-network in which the two series impedances of the T are bridged by a fourth impedance.

bridge duplex system — A duplex system based on the Wheatstone-bridge principle in which a substantial neutrality of the receiving apparatus to the transmitted currents is obtained by an impedance balance. Received currents pass through the receiving relay, which is bridged between the points that are equipotential for the transmitted currents.

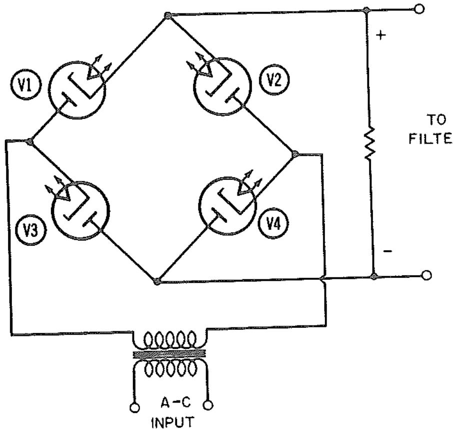

Bridge rectifier.

bridge hybrid—*See* Hybrid Junction

bridge quad—*See* B-quad.

bridge rectifier—A full-wave rectifier with four elements connected in the form of a bridge circuit so that d-c voltage is obtained from one pair of opposite junctions when an alternating voltage is applied to the other pair of junctions.

bridge transformer—*See* Bridging Transformer.

bridging—The shunting of one electrical circuit by another.

bridging amplifier—An amplifier with an input impedance sufficiently high that its input may be bridged across a circuit without substantially affecting the signal level of the circuit.

bridging connection—A parallel connection by means of which some of the signal energy in a circuit may be withdrawn with imperceptible effect on the normal operation of the circuit.

bridging contacts—A set of contacts in which the moving contact touches two stationary contacts simultaneously during transfer.

bridging gain—Ratio between the power a transducer delivers to a specified load impedance under specified operating conditions, and the power dissipated in the reference impedance across which the transducer input is bridged. Usually expressed in decibels.

bridging loss—Ratio between the power dissipated in the reference impedance across which the input of a transducer is bridged, and the power the transducer delivers to a specified load impedance under specified operating conditions. Usually expressed in decibels.

bridging transformer—Also called bridge transformer and hybrid coil. A transformer designed to couple two circuits having at least nominal ohmic isolation and operating at different impedance levels, without introducing significant frequency or phase distortion, or significant phase shift.

brightness—The attribute of visual perception in accordance with which an area appears to emit more or less light. Used with cathode-ray tubes.

brightness control—In a television receiver, the control which varies the average brightness of the reproduced image.

brightness signal—*See* Luminance Signal.

brilliance—1. The degree of brightness and clarity in a reproduced cathode-ray tube. 2. The degree to which the higher audio frequencies sound like the original when reproduced by a receiver or public-address amplifier.

brilliance control—A potentiometer used in a three-way speaker system to adjust the output level of the tweeter for proper relative volume between the treble and the lower audio frequencies produced by the complete speaker system.

broad band—As applied to data transmission, the term denotes transmission facilities capable of handling frequencies greater than those required for high-grade voice communications (higher than 3 to 4 kilohertz).

broad-band amplifier—An amplifier which has an essentially flat response over a wide frequency range.

broad-band antenna—An antenna which is capable of receiving a wide range of frequencies.

broad-band electrical noise—Also called random noise. A signal that contains a wide range of frequencies and has a randomly varying instantaneous amplitude.

broad-band interference—Interference occupying a frequency range that is much greater than the bandwidth of the equipment being used to measure it.

broad-band random vibration—Single-frequency component vibrations that are random in both phase and amplitude at minute increments of frequency throughout a specified bandwidth. Typically it occurs when intense, high-power-level noise impinges on structures.

broad-band tube (TR and pre-TR tubes)—A gas-filled, fixed-tuned tube incorporating a bandpass filter suitable for radio-frequency switching.

broadcast—Radio or television transmission intended for public reception, for which receiving stations make no receipt.

broadcast band—The band of frequencies extending from 535 to 1605 kHz.

broadcasting—The transmitting of speech, music, or visual programs for commercial or public-service motives to a relatively large audience (as opposed to two-way radio, for example, which is utilitarian and is directed toward a limited audience).

broadcasting service—A radiocommunication service in which the transmissions are intended for direct reception by the general public. This service may include sound transmissions, television transmissions, or other types of transmissions.

broadcasting station—A station in the broadcasting service.

broadside—A direction perpendicular to an axis or plane.

broadside array—An antenna array whose direction of maximum radiation is perpendicular to the line or plane of the array (depending on whether the elements lie on a line or a plane). A uniform broadside array is a linear array whose elements contribute fields of equal amplitude and phase.

broad tuning—A tuned circuit or circuits which respond to frequencies within one band or channel, as well as to a considerable range of frequencies on each side.

brush—A piece of conductive material, usually carbon or graphite, which rides on the commutator of a motor and forms the electrical connection between it and the power source.

brush discharge—An intermittent discharge of electricity which starts from a conductor when its potential exceeds a certain value but is too low for the formation of an actual spark. It is generally accompanied by a whistling or crackling noise.

brush-discharge resistance—*See* Corona Resistance.

brush rocker—A movable rocker, or "yoke," on which the brush holders of a dynamo or motor are fixed so that the position of the brushes on the commutator can be adjusted.

brush station—In a computer, a position where the holes of a punched card are sensed, particularly when this is done by a row of brushes sweeping electrical contacts.

brute-force filter—A type of power-pack filter which depends on large values of capacitance and inductance to smooth out pulsations, rather than on the resonant effects of a tuned filter.

brute supply—One that is completely unregulated. It employs no circuitry to maintain the output voltage constant with a changing input line or load variations. The output voltage varies in the same percentage and in the same direction as the input line voltage. If the load current changes, the output voltage changes inversely—an increase in load current drops the output voltage.

B & S—Abbreviation for Brown and Sharpe gage. A wire-diameter standard that is the same as AWG.

B scope—A type of radarscope that presents the range of an object by a vertical deflection of the signal on the screen, and the bearing by a horizontal deflection.

"B" service—FAA service pertaining to transmission and reception by teletype or radio of messages containing requests for and approval to conduct an aircraft flight, flight plans, in-flight progress reports, and aircraft arrival reports.

B supply or B+ supply—A source for supplying a positive voltage to the anodes and other positive electrodes of electron tubes.

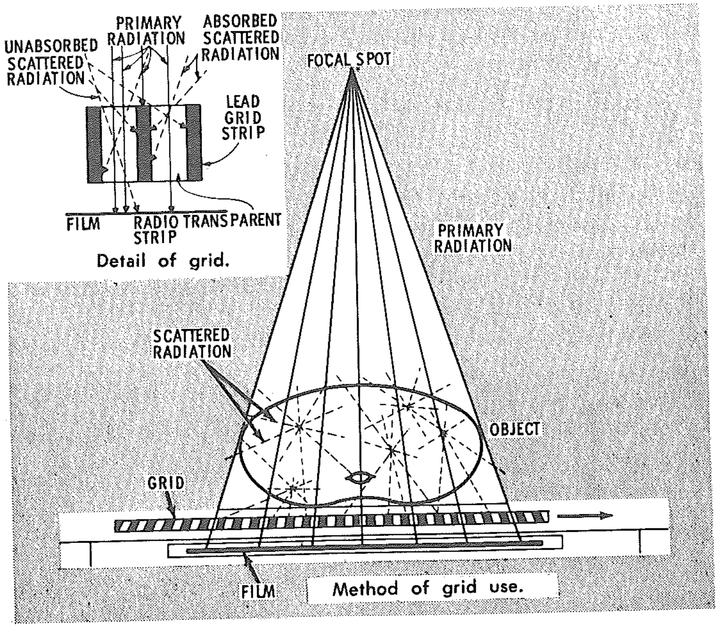

Bucky diaphragm.

buck—To oppose, as one voltage bucking another, or the magnetic fields of two coils bucking each other.

bucket—A general term for a specific reference in storage in a computer, such as a section of storage, the location of a word, a storage cell, etc.

bucking coil—A coil connected and positioned in such a way that its magnetic field opposes the magnetic field of another coil. The hum-bucking coil of an electrodynamic speaker is an example.

bucking voltage—A voltage which is opposite in polarity to another voltage in the circuit and hence bucks, or opposes, the latter voltage.

buckling—The warping of the plates of a battery due to an excessively high rate of charge or discharge.

Bucky diaphragm—A grid composed of narrow strips of lead arranged with X-ray–transparent spaces between adjacent strips and placed between the specimen and the X-ray film. Used to reduce scattered radiation.

buffer—1. A circuit or component which isolates one electrical circuit from another. Usually refers to electron-tube amplifiers used for this purpose. 2. A vacuum-tube stage used chiefly to prevent undesirable interaction between two other stages. In a transmitter, it generally follows the master-oscillator stage. 3. An isolating circuit used in an electronic computer to avoid reaction of a driven circuit on the corresponding driver circuit. 4. A storage device used to compensate for a difference in the rate of flow of information or the time or occurrence of events when transmitting information from one device to another. 5. A computer circuit having an output and a multiplicity of inputs and designed so that the output is energized whenever one or more inputs are energized. Thus, a buffer performs the circuit function equivalent to the logical OR.

buffer amplifier—An amplifier designed to isolate a preceding circuit from the effects of a following circuit.

buffer capacitor—A capacitor connected across the secondary of a vibrator transformer, or between the anode and cathode of a cold-cathode rectifier tube, to suppress voltage surges that might otherwise damage other parts in the circuit.

buffer element—A low-impedance inverting driver circuit. Because of its very low source impedance the element can supply substantially more output current than the basic circuit. As a consequence, the buffer element is valuable in driving heavily loaded circuits or minimizing rise-time deterioration due to capacitive loading.

buffer storage—In a computer, a synchronized element of storage between two other forms of storage. Computation may continue while data are being transferred between the buffer and the other storage.

bug—1. A semiautomatic telegraph sending key consisting of a lever which is moved to one side to produce a series of correctly spaced dots, and to the other side to produce a dash. 2. A circuit fault due to improper design or construction. 3. An electronic listening device, generally small and concealed, used for commercial or military espionage.

build—The increase of diameter due to insulation.

building-out—Addition to an electric structure of an element or elements electrically similar to an element or elements of the structure, to bring a certain property or characteristic to a desired value.

building-out circuit—A short section of transmission line, or a network which is shunted across a transmission line, for the purpose of impedance matching.

building-out network—A network designed to be connected to a basic network so that the combination will simulate the sending-end impedance, neglecting dissipation, of a line having a termination other than that for which the basic network was designed.

building-out section—A short section of transmission line, either open or short-circuited at the far end, shunted across another transmission line for use on an impedance-matching transformer.

bulb—The glass envelope which encloses an incandescent lamp or an electronic tube.

bulb-temperature pickup—A temperature transducer in which the sensing element is enclosed in a metal tube or sheath to protect it against corrosive liquids or other contaminants.

bulge—The difference between the actual characteristic and a linear characteristic of the attenuation-frequency characteristic of a transmission line.

bulk degausser—*See* Bulk Eraser.

bulk-erased noise—The noise arising when a bulk-erased tape is reproduced with the erase and record heads completely de-energized. Ideally, this noise is governed by the number of magnetic particles that pass by the head in unit time.

bulk eraser—Also called bulk degausser or degausser. Equipment for erasing a roll of tape. The roll is usually rotated while a 60-hertz a-c erasing field is decreased, either by withdrawing the roll from an electromagnet or by reducing the a-c supply to an electromagnet.

bulk noise—*See* Excess Noise.

bulk resistor—A resistor made by providing ohmic contacts between two points of a homogeneous, uniformly doped crystal of silicon material.

bump contracts—A small mound of material formed on the chip substrate, to register with terminal pads, as then the chip is employed in "flip-chip circuits."

bunched pair—A group of pairs tied together or otherwise associated for identification.

buncher—1. The input resonant cavity in a conventional klystron oscillator. 2. In a velocity-modulated tube, the electrode which concentrates the electrons in the constant-current electron beam into bunches.

buncher gap—*See* Input Gap.

buncher resonator—The input cavity resonator in a velocity-modulated tube. It serves to modify the velocity of the electrons in the beam.

bunching—1. Grouping pairs together for identification and testing. 2. In a velocity-modulated electron stream the action that produces an alternating convection-current component as a direct result of the differences of electron transit time produced by the velocity modulation.

bunching parameter—One-half of the product of the bunching angle in the absence of velocity modulation and the depth of velocity modulation.

bunching time—The time in the armature motion of a relay during which all three contacts of a bridging-contact combination are electrically connected.

bunching voltage—The radio-frequency voltage between the grids of the buncher resonator in a velocity-modulated tube such as a klystron. Generally, the term implies the peak value of this oscillating voltage.

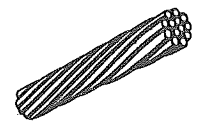

Bunch stranding.

bunch stranding—A method in which a number of wires are twisted together in a common direction and with a uniform pitch to form a finished, stranded wire.

bundled cable—Individual insulated wires laced together to form a bundle to facilitate handling.

buried cable—A cable installed underground and not removable except by disturbing the soil.

buried layer—A low-resistivity material introduced into a silicon substrate.

burned-in image—An image which remains in a fixed position in the output signal of a camera tube after the camera has been turned to a different scene.

burn-in—Operation of a device to stabilize its failure rate.

burn-in period—*See* Early Failure Period.

burst—1. A sudden increase in the strength of a signal. 2. The cosmic-ray effect upon matter, causing a sudden intense ionization that gives rise to great numbers of ion pairs at once. (*Also see* Color Burst.)

burst pedestal—A rectangular pulselike television signal which may be part of the color burst. The amplitude of the color-burst pedestal is measured from the alternating-current axis of the sine-wave portion of the horizontal pedestal.

burst pressure—The maximum pressure to which a device can be subjected without rupturing.

bus—1. In a computer, one or more conductors used as a path over which information is transmitted from any of several sources to any of several destinations. 2. The term used to specify an uninsulated conductor (a bar or wire); may be solid, hollow, square, or round. 3. Sometimes used to specify a bus bar. 4. The communications path between two switching points.

bus bar—A heavy copper strap or bar used to carry heavy currents or to make a common connection between several circuits.

bus reactor—A current-limiting reactor connected between two different buses or two sections of the same bus for the purpose of limiting and localizing the disturbance due to a fault on either bus.

bussback—The connection, by a common carrier, of the output portion of a circuit back to the input portion of a circuit. (*Also see* Loopback Test.)

busy test—In telephony, a test to find out whether certain facilities which may be desired, such as a subscriber line or a trunk, are available for use.

Butler oscillator—A two-tube (or transistor) crystal-controlled oscillator in which the crystal forms the positive feedback path when excited in its series-resonant mode.

butterfly resonator—Also called butterfly capacitor. A tuning device that combines both inductance and capacitance in such a manner that it exhibits resonant properties at very high and ultrahigh frequencies (characterized by a high tuning ratio and Q). So called because the shape of the rotor resembles the opened wings of a butterfly.

Butterworth filter—A filter network that exhibits the flattest possible response in the passband. The response is monotonic, rolling off smoothly into the stop band, where it approaches a constant slope of 6 db/octave.

butt joint—1. A splice or other connection formed by placing the ends of two conductors together and joining them by welding, brazing, or soldering. 2. A connection between two waveguides which maintains electrical continuity by providing physical contact between the ends.

button—1. The metal container in which the carbon granules of a carbon microphone are held. 2. Also called dot. A piece of metal used for alloying onto the base wafer in making alloy transistors.

button silver-mica capacitor—A stack of silvered mica sheets encased in a silver-plated brass housing. The high-potential terminal is connected through the center of the stack. The other capacitor terminal is formed by the metal shell, which connects at all points around the outer edge of the electrodes. This design permits the current to fan out in a 360° pattern from the center terminal, providing the shortest possible electrical path between the center terminal and chassis. The internal series inductance is thus kept small.

button stem—In a tube, the glass base onto which the mount structure is assembled. The pins may be sealed into the glass; if so, no base is needed. In some large tubes, the stiff wires are passed directly into the base pins to give added strength. (*Also see* Pressed Stem.)

button up—To close or completely seal any operating device.

buzzer—A signaling device in which an armature vibrates to produce a raucous, nonresonant sound.

BV_{EBO}—The reverse-breakdown voltage of the emitter-to-base junction of a transistor with the collector open-circuited.

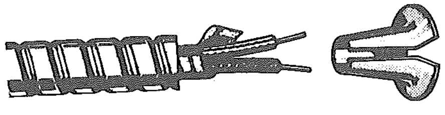

BX cable.

BX cable—Insulated wires in flexible metal tubing, used for wiring buildings and electrical and electronic equipment.

bypass—A shunt (parallel) path around one or more elements of a circuit.

bypass capacitor—A capacitor used for providing a comparatively low-impedance a-c path around a circuit element.

bypass filter—A filter providing a low-attenuation path around some other circuit or equipment.

B–Y signal—One of the three color-difference signals in color television. The B–Y signal forms a blue primary signal for the picture tube when combined—either inside or outside the picture tube—with a luminance or Y signal.

byte—A single group of bits processed together (in parallel). It can consist of a variable number of bits.

C

C—Symbol for capacitor, capacitance, carbon, coulomb, centigrade or Celsius, transistor collector, candle, velocity of light.

C– (C minus)—The negative terminal of a C battery, or the negative polarity of other sources of grid-bias voltage. Used to denote the terminal to which the negative side of the grid-bias voltage source should be connected.

C+ (C plus)—Positive terminal of a C battery, or the positive polarity of other sources of grid-bias voltage. The terminal to which the positive side of the grid-bias voltage source should be connected.

cabinet—A protective housing for electrical or electronic equipment.

cable—1. An assembly of one or more conductors, usually within a protective sheath, and so arranged that the conductors can be used separately or in groups. 2. A stranded conductor (single-conductor cable) or a combination of conductors insulated from one another (multiple-conductor cable).

cable assembly—A cable with plugs or connectors on each end for a specific purpose. It may be formed in various configurations.

cable clamp—A device used to give mechanical support to the wire bundle or cable at the rear of a plug or receptacle.

cable core—That portion of an insulated cable lying under the protective covering or jacket.

cable coupler—A device used to join lengths of similar or dissimilar cable having the same electrical characteristics.

cable fill—The ratio of the number in use to the total number of pairs in a cable.

cable filler—Material used in multiple-conductor cables to occupy the spaces between the insulated conductors.

cable Morse code—A three-element code used mainly in submarine-cable telegraphy. Dots and dashes are represented by positive and negative current impulses of equal length, and a space by the absence of current.

cable run—The path occupied by a cable on cable racks or other support from one termination to another.

cable sheath—A protective covering of rubber, neoprene, resin, or lead over a wire or cable core.

cable splice—A connection between two or more separate lengths of cable. The conductors in one length are individually connected to conductors in the other length, and the protecting sheaths are so connected that protection is extended over the joint.

cable terminal—A means of electrically connecting a predetermined number of cable conductors in such a way that they can be individually selected and extended by conductors outside the cable.

cable vault—A vault in which the outside plant cables are spliced to the tipping cables.

cabling—The method by which a group of insulated conductors is mechanically assembled (or twisted together).

cactus needle—A phonograph needle made from the thorn of a cactus plant.

cadmium—A metallic element widely used for plating steel hardware or chassis to improve its appearance and solderability and to prevent corrosion. It is also used in the manufacture of photocells.

cadmium selenide photoconductive cell—A photoconductive cell that uses cadmium selenide as the semiconductor material. It has a fast response time and high sensitivity to longer light wavelengths, such as those emitted by incandescent lamps and some infrared light sources.

cadmium sulfide photoconductive cell—A photoconductive cell in which a small wafer of cadmium sulfide is used to provide an extremely high dark-light resistance ratio. Some of the cells can be used directly as a light-controlled switch operated directly from the 120-volt a-c power line.

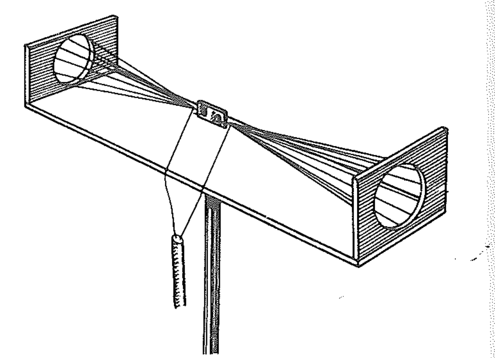

Cage antenna.

cage antenna—An antenna comprising a number of wires connected in parallel and arranged in the form of a cage. This is done to reduce the copper losses and increase the effective capacity.

calculating—Computing a result by multiplication, division, addition, or subtraction, or by a combination of these operations. A data-processing function.

calculating punch—A punched-card machine that reads data from a group of cards and punches new data in the same or other cards.

calculator—1. A device capable of performing arithmetic. 2. A calculator as in 1 which requires frequent manual intervention. 3. Generally and historically, a device for carrying out logical and arithmetical digital operations of any kind.

calendar age—Age of an item or object measured in terms of time elapsed since it was manufactured.

calendar life—That period of time expressed in days, months, or years during which an item may remain installed and in operation, but at the end of which the item should be removed and returned for repair, overhaul, or other maintenance.

calibrate—To ascertain, by measurement or by comparison with a standard, any variations in the readings of another instrument, or to correct the readings.

calibrated interference measurements—Measurements in which the meter is calibrated against a known sine wave or impulse-generator output and interference readings are taken directly from the output indicating meter on the receiver.

calibrated triggered sweep—In a cathode-ray oscilloscope, a sweep that occurs only when initiated by a pulse and moves horizontally at a known rate.

calibration—1. The process of comparing an instrument or device with a standard to determine its accuracy or to devise a corrected scale. 2. Taking measurements of various parts of electronic equipment to determine the performance level of the equipment and whether it conforms to technical order specifications.

calibration accuracy—Finite degree to which a device can be calibrated (influenced by sensitivity, resolution, and reproducibility of the device itself and the calibrating equipment). Expressed as a per cent of full scale.

calibration curve—A smooth curve connecting a series of calibration points.

calibration marker—On the screen of a radar indicator, the markings that divide the range scale into accurate intervals for range determination or checking against mechanical indicating dials, scales, or counters.

call—1. A transmission made for the purpose of identifying the transmitting station and the station for which the transmission is intended. 2. To transfer control to a specified closed subroutine. 3. In communication the action performed by the calling party, or the operations necessary in making a call, or the effective use made of a connection between two stations.

call circuit—A communication circuit between switching points used by traffic forces for transmitting switching instructions.

call in—To transfer control of a digital computer temporarily from a main routine to a subroutine, which is inserted in the sequence of calculating operations to fulfill a subsidiary purpose.

call letters—A series of government-assigned letters, or letters and numbers, which identify a transmitting station.

call sign—Any combination of characters or pronounceable words which identifies a communication facility, a command, an authority, an activity, or a unit; used primarily for establishing and maintaining communications.

call word—A call number which is exactly the size of one machine word.

calomel electrode—An electrode consisting of mercury in contact with a solution of potassium chloride saturated with mercurouse chloride (calomel). (*Also see* Glass Electrode.)

calorimeter—An apparatus for measuring quantities of heat. Used to measure microwave power in terms of heat generated.

calorimeter system—A precision r-f wattmeter as well as an efficient dummy load able to absorb energy at any frequency band. It can absorb and measure any level of microwave energy, and functions by circulating a known amount of liquid through a suitably designed low vswr load. The load is located in a wave-guide or coaxial section that mates to the termination of the r-f energy source being measured or absorbed.

camera—*See* Television Camera.

camera cable—A cable or group of wires that carry the picture signal from the television camera to the control room.

camera chain—A television camera, associated control units, power supplies, monitor, and connecting cables necessary to deliver a picture for broadcasting.

camera signal—The video-output signal of a television camera.

camera tube—An electron-beam tube in which an electron current or charge-density image is formed from an optical image and scanned in a predetermined sequence to provide an electric signal.

can—A metal shield placed around a tube, coil, or transformer to prevent electromagnetic or electrostatic interaction.

canal ray—Also called positive ray. Streams of positive ions that flow from the anode to the cathode in an evacuated tube.

candle—The unit of luminous intensity. One candle is defined as the luminous intensity of 1/60th square centimeter of a black-body radiator operating at the solidification temperature of platinum.

candlepower—Luminous intensity expressed in terms of standard candles.

candoluminescence—A phenomenon that produces white light without need for very high temperatures.

cannibalization—A method of maintenance or modification in which the required parts are removed from one system or assembly for installation on a similar system or assembly.

capacitance—1. Abbreviated C. Also called capacity. In a capacitor or a system of conductors and dielectrics, that property which permits the storage of electrically separated charges when potential differences exist between the conductors. The capacitance of a capacitor is defined as the ratio between the electric charge that has been transferred from one electrode to the other and the resultant difference in potential between the electrodes. The value of this ratio is dependent on the magnitude of the transferred charge.

$$C\ (\text{farad}) = \frac{Q\ (\text{coulomb})}{V\ (\text{volt})}$$

2. In the case of fluid devices, it accumulates fluid by virtue of its volume and the pressure head available.

capacitance between two conductors—The ratio between the charge transferred from one conductor to the other and the resultant difference in the potentials of the two conductors when insulated from each other and from all other conductors.

capacitance divider—A circuit made up of capacitors and used for measuring the value of a high-voltage pulse by making available only a small, known fraction of the total pulse voltage for measurement.

capacitance meter—An instrument for measuring capacitance. If the scale is graduated in microfarads, the instrument is usually designated a microfaradmeter.

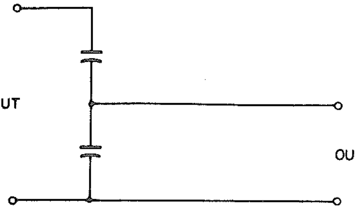

Capacitance divider.

capacitance-operated intrusion detector—A boundary alarm system in which the approach of an intruder to an antenna wire encircling the protected area (a few feet above ground) changes the antenna-ground capacitance and thereby sets off the alarm.

capacitance ratio—The ratio of maximum to minimum capacitance, as determined from a capacitance characteristic, over a specified voltage range.

capacitance relay—An electronic circuit incorporating a relay which responds to a small change in capacitance, such as that created by bringing the hand or body near a pickup wire or plate.

capacitive coupling—Also called electrostatic coupling. The association of two or more circuits with one another by means of mutual capacitance between them. Between stages of an amplifier, that type of interconnection which employs a capacitor in the circuit, between the plate of one tube and the grid of the succeeding tube.

capacitive divider—Two or more capacitors placed in series across a source, making available a portion of the source voltage across each capacitor. The voltage across each capacitor will be inversely proportional to its capacitance.

capacitive feedback—The process of returning part of the energy in the plate or output circuit of a vacuum tube to the grid, or input, circuit by means of a capacitance common to both circuits.

capacitive load—A predominantly capacitive load—i.e., one in which the current leads the voltage.

capacitive reactance—Symbolized by X_c. The impedance a capacitor offers to ac or pulsating dc. Measured in ohms and equal to $\frac{1}{2\pi fC}$, where f is in hertz and C is in farads.

capacitive speaker — *See* Electrostatic Speaker.

capacitive storage welding — A particular type of resistance welding whereby the energy is stored in banks of capacitors, which are then discharged through the primary of the welding transformer. The secondary current generates enough heat to produce the weld.

capacitive transduction—Conversion of the measurand into a change in capacitance.

capacitive tuning—Tuning by means of a variable capacitor.

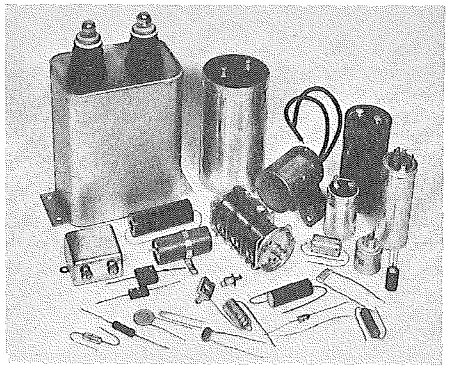

Capacitors.

capacitivity—*See* Dielectric Constant.

capacitor—A device consisting essentially of two conducting surfaces separated by an insulating material or dielectric such as air, paper, mica, glass, plastic film, or oil. A capacitor stores electrical energy, blocks the flow of direct current, and permits the flow of alternating current to a degree dependent upon the capacitance and the frequency.

capacitor antenna—Also called condenser antenna. An antenna which consists of two conductors or systems of conductors and the essential characteristic of which is its capacitance.

capacitor bank—A number of capacitors connected together in series, parallel, or in series-parallel.

capacitor color code—Color dots or bands placed on capacitors to indicate one or more of the following: capacitance, capacitance tolerance, voltage rating, temperature coefficient, and the outside foil (on paper capacitors).

capacitor-input filter—A power-supply filter in which a capacitor is connected directly across, or in parallel with, the rectifier output.

capacitor microphone—*See* Electrostatic Microphone.

capacitor motor—A single-phase induction motor with the main winding arranged for direct connection to the power source, together with an auxiliary winding connected in series with a capacitor.

capacitor pickup — A phonograph pickup which depends for its operation on the variation of its electrical capacitance.

capacitor series resistance—An equivalent resistance in series with a pure capacitance which gives the same resultant losses as the actual capacitor. This equivalent circuit does not represent the variation in capacitor losses with frequency.

capacitor speaker—*See* Electrostatic Speaker.

capacitor-start motor—An a-c split-phase induction motor in which a capacitor is connected in series with an auxiliary winding to provide a means of starting. The auxiliary circuit opens when the motor reaches a predetermined speed.

capacitor voltage — The voltage across the terminals of a capacitor.

capacity—1. The current-output capability of a cell or battery over a period of time. Usually expressed in ampere-hours (amp-hr). 2. Capacitance. 3. The limits, both upper and lower, of the items or numbers which may be processed in a computer register, in the accumulator. When quantities exceed the capacity, a computer interrupt develops and requires special handling. 4. The total quantity of data that a part of a computer can hold or handle, usually expressed as words per unit of time. 5. The capability of a specific system to store data, accept transactions, process data, and generate reports.

capstan—The driven spindle or shaft in a tape recorder—sometimes the motor shaft itself—which rotates against the tape, pulling it through the machine at a constant speed during recording and playback modes of operation. The rotational speed and diameter of the capstan thus determine the tape speed.

capstan idler—*See* Pressure Roller.

captive screw—Screw-type fastener that is retained when unscrewed and cannot easily be separated from the part it secured.

capture effect — 1. The selection of the stronger of two frequency-modulated signals of the same frequency, with the complete rejection of the weaker signal. If both signals are of equal strength, both may be accepted and no intelligible signal will result. 2. An effect occurring in a transducer (usually a demodulator) whereby the input wave having the largest magnitude controls the output.

capture ratio—The ability of a tuner to reject unwanted f-m stations and interference on the same frequency as a desired one, measured in db. The lower the figure, the better the performance of the tuner.

carbon—One of the elements, consisting of a nonmetallic conductive material occurring as graphite, lampblack, diamond, etc. Its resistance is fifty to several hundred times that of copper and decreases as the temperature increases.

carbon brush — A current-carrying brush made of carbon, carbon and graphite, or carbon and copper.

carbon-contact pickup—A phonograph pickup which depends for its operation on the variation in resistance of carbon contacts.

carbon-film resistor—A resistor formed by vacuum-depositing a thin carbon film on a ceramic form.

carbonize—To coat with carbon.

carbonized filament — A thoriated-tungsten filament treated with carbon. A layer of tungsten carbide formed on the surface slows down the evaporation of the active emitting thorium and thus permits higher operating temperatures and much greater electron emission.

carbonized plate—An electron-tube anode that has been blackened with carbon to increase its heat dissipation.

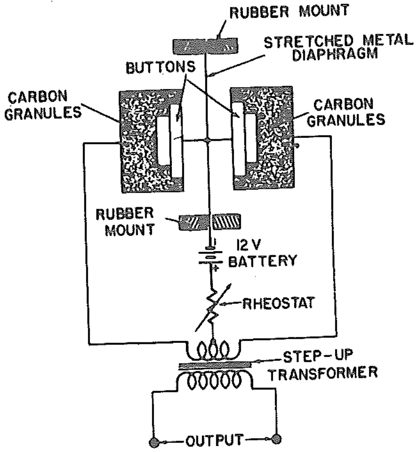

Carbon microphone.

carbon microphone—A microphone which depends for its operation on the variation in resistance of carbon contacts.

carbon-pile regulator—An arrangement of carbon discs the series resistance of which decreases as more pressure or compression is applied.

carbon resistor—Also called composition resistor. A resistor consisting of carbon particles which are mixed with a binder molded into a cylindrical shape, and then baked. Terminal leads are attached to opposite ends. The resistance of a carbon resistor decreases as the temperature increases.

carbon transfer recording—A type of facsimile recording in which carbon particles are deposited on the record sheet in response to the received signal.

carborundum—A compound of carbon and silicon used in crystals to rectify or detect radio waves.

carcinotron—A voltage-tuned, backward-wave oscillator tube used to generate frequencies ranging from uhf up to 100GHz or more.

card—*See* Punched Card and Printed Circuit Board.

card bed — A mechanical device for holding punch cards to be transported past the punching and reading stations.

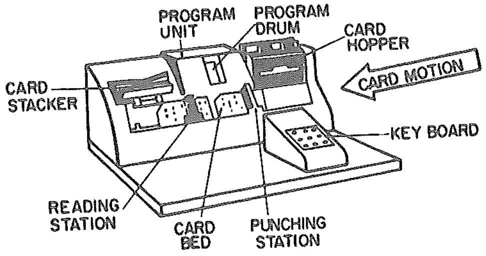

Card bed.

card code—An arbitrary code in which holes punched in a card are assigned numeric or alphabetic values.

card column—One of 20 to 90 single-digit columns in a tabulating card. When punched, a column contains only one digit, one letter, or one special code.

Cardew voltmeter—The earliest type of hot-wire instrument. It consisted of a small-diameter platinum-silver wire sufficiently long to give a resistance high enough to be connected directly across the circuit being measured. The wire was looped over pulleys and it expanded as current flowed, causing the pointer to rotate.

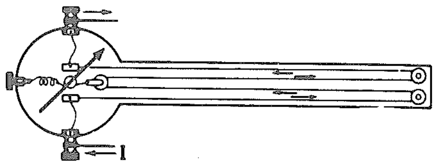

Cardew voltmeter.

card face—The printed side of a punched card if only one side is printed.

card feed—A mechanism that moves punch cards, one at a time, into a machine.

card field—On a punch card, the fixed columns in which the same type of information is routinely entered.

card hopper—*See* Card Stacker.

cardiac monitor—An instrument that usually has an oscilloscope display of the heart wave, and combines the features of several cardiac instruments such as an electrocardiograph, cardiotachometer, etc. May also allow upper and lower limits to be set and trigger audible and/or visual alarms when these limits are exceeded.

card image—1. A representation in storage of the holes punched in a card, in such a manner that the holes are represented by one binary digit and the unpunched spaces are represented by the other binary digit. 2. In machine language, a duplication of the data contained in a punch card.

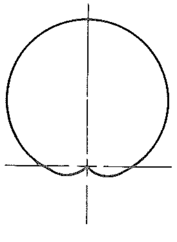

Cardioid diagram.

cardiograph—An instrument (or recording of instruments) for measuring the form or force of heart motion.

cardioid diagram—Heart-shaped polar diagram showing the response or radiation characteristic of certain directional antennas, or the response characteristic of certain types of microphones.

cardioid microphone—A microphone that has a heart-shaped response pattern that gives nearly uniform response for a range of about 180° in one direction and a minimum response in the opposite direction.

cardiostimulator—A device used to stimulate the heart and/or regulate its beat. *Also see* Pacemaker and Defibrillator.

card jam—A pile-up of cards in a machine.

card machine—A machine used to transfer information from or to punched cards.

card programmed—1. The capability of being programmed by punched cards. 2. The capability of performing sequences of calculating operations according to instructions contained in a stack of punched cards.

card punch—A device to record information in cards by punching holes in the cards to represent letters, digits, and special characters.

card reader—A device designed to read punched cards and convert each hole into an electrical impulse for use in a computer system.

card row—On a punched card, one of the horizontal lines of punching positions.

card sensing—The process of sensing or reading the information in punched cards and converting this information, usually into electrical pulses.

card stacker—A mechanism that stacks cards in a pocket or bin after they have passed through a computer. Also called card hopper.

carillon—A bell tower designed to play from a keyboard. In an organ, this may be achieved by tube synthesis of bell-like tones struck with felt hammers, or completely electronically.

Carnot theorem—A thermodynamic principle which states that a cycle continuously operating between a low temperature and a high temperature can be no more efficient than a reversible cycle operating between the same temperatures.

carriage tape—*See* Control Tape.

carrier—1. A wave of constant amplitude, frequency, and phase which can be modulated by changing amplitude, frequency, or phase. 2. An entity capable of carrying an electric charge through a solid (e.g., holes and conduction electrons in semiconductors).

carrier-amplitude regulation—The change in amplitude of the carrier wave in an amplitude-modulated transmitter when symmetrical modulation is applied.

carrier beat—In facsimile transmission, an undesirable heterodyne of signals, each synchronous with a different stable oscillator, causing a pattern in received copy. Where one or more of the oscillators is fork-controlled, this is called fork beat.

carrier chrominance signal—In color television, sidebands of a modulated chrominance subcarrier, plus any unsuppressed subcarrier, added to the monochrome signal to convey color information.

carrier color signal—In color television, sidebands of a modulated chrominance subcarrier, plus the chrominance subcarrier, if not suppressed, which are added to the monochrome signal to convey color information.

carrier concentration—The number of carriers in a cubic centimeter of semiconductor material.

carrier control—A control by the presence or absence of an r-f carrier.

carrier current—The current associated with a carrier wave.

carrier-current communication—The superimposing of a high-frequency alternating current on ordinary telephone, telegraph, and power-line frequencies for telephone communication and control.

carrier-current control—Remote control in which the receiver and transmitter are coupled together through power lines.

carrier frequency—1. The frequency (hertz) of the wave modulated by the intelligence wave; usually a radio frequency (rf). 2. The reciprocal of the period of a periodic carrier.

carrier-frequency interconnection—In the formation of carrier networks, the transfer of groups of channels between terminals of wire-line cable or radio carrier systems at carrier frequencies.

carrier-frequency peak-pulse power—The power, averaged over that carrier-frequency cycle, which occurs at the maximum pulse of power (usually half the maximum instantaneous power).

carrier-frequency pulse—A carrier that is amplitude-modulated by a pulse. The amplitude of the modulated carrier is zero before and after the pulse.

carrier-frequency range—The continuous range of frequencies within which a transmitter may normally operate. A transmitter may have more than one carrier-frequency range.

carrier-frequency stereo disc—A stereo disc with two laterally-cut channels. One channel is cut in the usual manner. The second channel is employed to frequency-modulate a supersonic carrier frequency. The playback cartridge delivers the signal for one channel plus the carrier frequency containing the other channel. The latter must then be demodulated to obtain the second channel.

carrier-isolating choke coil—An inductor inserted in series with a line on which carrier energy is applied to impede the flow of carrier energy beyond that point.

carrier leak—The carrier-frequency signal remaining after suppression in a suppressed carrier system.

carrier level—The strength, expressed in decibels, of an unmodulated carrier signal at a particular point in a system.

carrier lifetime—The time required for excess carriers doped into a semiconductor to recombine with other carriers of the opposite sign.

carrier line—A transmission line used for multiple-channel carrier communications.

carrier loading—The insertion of additional lump inductance in a cable section of a transmission line utilized for carrier transmission up to about 35 kHz. Loading serves to minimize impedance mismatch between cable and open wire, and to reduce the cable attenuation.

carrier mobility—The average drift velocity of carriers per unit electric field in a homogeneous semiconductor. The mobility of electrons is usually different from that of holes.

carrier noise level—Also called residual modulation. The noise level produced by undesired variations of a radio-frequency signal in the absence of any intended modulation.

carrier power output rating—The power available at the output terminals of a transmitter when the output terminals are connected to the normal-load circuit or to a circuit equivalent thereto.

carrier repeater—An assembly, including an amplifier (or amplifiers), filters, equalizers, level controls, etc., used to raise the carrier signal level to a value suitable for traversing a succeeding line section while maintaining an adequate signal-to-noise ratio.

carrier shift—1. The transmission of radio teletypewriter messages by shifting the carrier frequency in one direction for a marking signal and in the opposite direction for a spacing signal. 2. The condition resulting from imperfect modulation, whereby the positive and negative excursions of the envelope pattern are unequal, thus effecting a change in the power associated with the carrier.

carrier signaling—In a telephone system, the method by which ringing, busy signals, or dial-signaling relays are operated by the transmission of a carrier-frequency tone.

carrier suppression—The method of operation in which the carrier wave is not transmitted.

carrier swing—The total deviation of a frequency- or phase-modulated wave from the lowest to the highest instantaneous frequency.

carrier system—A means of obtaining a number of channels over a single path by modulating each channel upon a different carrier frequency and demodulating at the receiving point to restore the signals to their original form.

carrier tap choke coil—A carrier-isolating choke coil inserted in series with a line tap.

carrier tap transmission choke coil—An inductor inserted in series with a line tap to control the amount of carrier energy flowing into the tap.

carrier telegraphy—That form of telegraphy in which the transmitted signal is formed by modulating the alternating current, under control of the transmitting apparatus, before supplying it to the line.

carrier telephony—Ordinarily applied only to wire telephony. That form of telephony in which carrier transmission is used, the modulating wave being at an audio frequency.

carrier terminal—Apparatus at one end of a carrier transmission system, whereby the processes of modulation, demodulation, filtering, amplification, and associated functions are effected.

carrier-to-noise ratio—Ratio of the magnitude of the carrier to the magnitude of the noise after selection and before any nonlinear process such as amplitude limiting and detection. This ratio is expressed in many different ways—for example, in terms of peak values in the case of impulse noise, and in terms of root-mean-square values in the case of random noise.

carrier-transfer filter—A group of filters arranged to form a carrier-frequency crossover or bridge between two transmission circuits.

carrier transmission—That form of electrical transmission in which a single-frequency wave is modulated by another wave containing the information.

carrier type d-c amplifier—An amplifier system which converts a d-c input to modulated a-c amplifiers, and synchronously detects to provide amplified d-c output.

carrier wave—The single-frequency wave which is transmitted and which is modulated by another wave containing the information.

carry—1. A signal or expression produced in an electronic computer by an arithmetic operation on a one-digit place of two or more numbers expressed in positional notation and transferred to the next higher place for processing there. 2. A signal or expression—as defined in (1)—which arises when the sum of two digits in the same digit place equals or exceeds the base of the number system in use. If a carry into a digit place will result in a carry out of the same digit place and the normal adding circuit is bypassed when this new carry is generated, the result is called a high-speed or "stand-on-nines" carry. If the normal adding circuit is used, the result is called a cascade carry. If a carry resulting from the addition of carries is not allowed to propagate, the process is called a partial carry; if it is, a complete carry. A carry generated in the most significant digit place and sent directly to the least significant place is called an end-around carry. 3. In direct subtraction, a signal or expression—as defined in (1) above—which arises when the difference between the digits is less than zero. Such a carry is frequently called a borrow. 4. The action of forwarding a carry. 5. The command directing a carry to be forwarded.

carry time—The time required for a computer to transfer a carry digit to the next higher column and add it there.

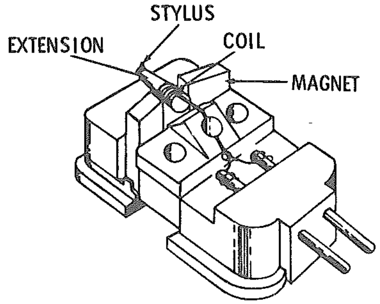

Cartridge.

cartridge—A removable phonograph pickup.

cartridge fuse—A tubular fuse the end caps of which are enclosed in a glass or composition insulating tube to confine the arc or vapor when the fuse blows.

cascade—*See* Tandem.

cascade amplifier—A multiple-stage amplifier in which the output of each stage is connected to the input of the next stage.

cascade connection—Two or more similar component devices arranged in tandem, with the output of one connected to the input of the next.

cascade control—Also called piggy-back control. An automatic control system in which the control units, linked in chain fashion, feed into one another in succession. Each unit thus regulates the operation of the next in line.

cascaded carry—In a computer, a system of executing the carry process in which carry information cannot be passed on to place (N + 1) unless the Nth place has received carry information or produced a carry.

cascaded feedback canceler—Also called velocity-shaped canceler. A sophisticated moving-target indicator canceler which provides clutter and chaff rejection.

cascaded thermoelectric device—A thermoelectric device having two or more stages that are arranged thermally in series.

cascode amplifier—An amplifier using a neutralized grounded-cathode input stage followed by a grounded-grid output stage. The circuit has high gain, high input impedance, and low noise.

case pressure—The total differential pressure in the internal cavity of a transducer and the ambient pressure. The term is commonly used to summarize the limiting combined differential and/or line-pressure capabilities of differential transducers.

case temperature—The temperature on the surface of the case at a designated point.

Cassegrain antenna—An antenna the feed of which is positioned near the vertex of the reflector, with a small subreflector placed near the focal point. The feed illuminates the subreflector, and the subreflector redirects the waves toward the main reflector, which then forms the radiated beam (called the secondary beam or pattern). The shapes of the subreflector and main reflector are so chosen that the secondary rays will emerge parallel to the main-reflector axis.

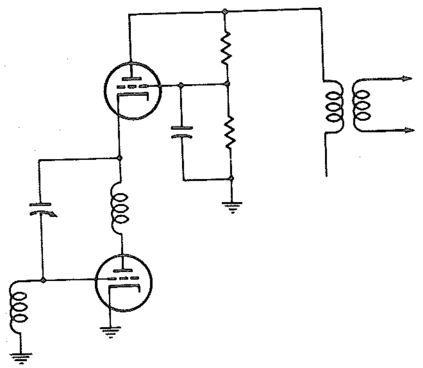

Cascode amplifier.

casting-out-nines check—A partial verification of an arithmetical operation on two or more numbers. It involves casting out nines from the numbers and from the results.

catalyst—Also called hardener and promoter. A substance which markedly speeds up the cure of a casting or molding compound when added in a minor quantity as compared to the amounts of primary reactants.

catatastrophic failure—A sudden failure without warning, as opposed to degradation failure. Or, a failure the occurrence of which can prevent the satisfactory performance of an entire assembly or system.

catcher—In a velocity-modulated vacuum tube, an electrode on which the spaced electron groups induce a signal. The output of the tube is taken from this element.

catena—A chain or connected series.

catenate—*See* Concatenate.

catenation—*See* Concatenation.

cathamplifier—A push-pull amplifier in which the push-pull transformer is in the cathode circuit.

cathode—1. In an electron tube the electrode through which a primary source of electrons enters the interelectrode space. 2. General name for any negative electrode. 3. The lower-potential electrode of a corrosion cell where the action of the corrosion current may reduce or eliminate corrosion, or the negatively charged metallic parts of an impressed-current system. 4. When a semiconductor diode is biased in the forward direction, that terminal of the diode which is negative with respect to the other terminal.

cathode activity—Measure of the efficiency of an emitter. The mathematical relationship between two values of emission current measured under two conditions of cathode temperature.

cathode bias—A method of biasing a vacuum tube by placing the biasing resistor in the common cathode-return circuit, thereby making the cathode more positive—rather than the grid more negative—with respect to ground.

cathode-coupled amplifier—A cascade amplifier in which the coupling between two stages is accomplished by a common-cathode resistor.

cathode coupling—The use of an input or output element in the cathode circuit for coupling energy to another stage.

cathode current—*See* Electrode Current.

cathode-current density—The current per square centimeter of cathode area, expressed as amperes or milliamperes per centimeter squared.

cathode dark space—Also called Crookes' dark space. The relatively nonluminous region between the cathode and negative glow in a glow-discharge—cold-cathode tube.

cathode emission—The process whereby electrons are emitted from the cathode structure.

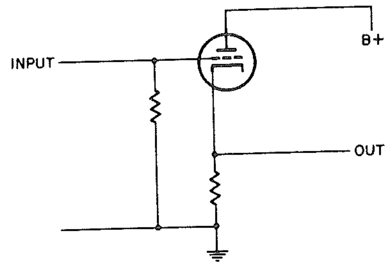

Cathode follower.

cathode follower—Also called grounded-plate amplifier. A vacuum-tube circuit in which the input signal is applied to the control grid, and the output is taken from the cathode. Electrically, such a circuit possesses a high input impedance and a low output impedance characteristic and a gain of less than unity. The equivalent circuit using a transistor is called an emitter follower.

cathode glow—The luminous glow which covers the surface of the cathode in a glow-discharge—cold-cathode tube, between the cathode and the cathode dark space.

cathode guide—The element of a glow tube used in switching the neon glow from one indicated number to the next.

cathode heating time—The time required for the cathode to attain a specified condition—for example, a specified value of emission or a specified rate of change in emission.

cathode interface—A resistive and capacitive layer formed between the nickel sleeve and oxide coating of an indirectly-heated cathode. Raising the cathode temperature will largely nullify the layer.

cathode keying—A method of keying a radiotelegraph transmitter by opening the plate return lead to the cathode or filament center tap.

cathode luminous sensitivity (of a multiplier phototube)—The photocathode current divided by the incident luminous flux.

cathode modulation—A form of amplitude modulation in which the modulating voltage is applied to the cathode circuit.

cathode pulse modulation—Modulation produced in an amplifier or oscillator by applying externally generated pulses to the cathode circuit.

cathode radiant sensitivity—The current leaving the photocathode divided by the incident radiant power of a given wavelength.

cathode ray—A stream of electrons emitted, under the influence of an electric field, from the cathode of an evacuated tube or from the ionized region nearby.

cathode-ray charge-storage tube — A charge-storage tube in which the desired information is written by means of a cathode-ray beam.

cathode-ray instrument—*See* Electron-Beam Instrument.

cathode-ray oscillograph — An apparatus capable of producing, from a cathode-ray tube, a permanent record of the value of an electrical quantity as a function of time.

cathode-ray oscilloscope—A test instrument which, when properly adjusted, makes possible the visual inspection of alternating-current signals. It consists of an amplifier, time-base generating circuits, and a cathode-ray tube for transformation of electrical energy into light energy.

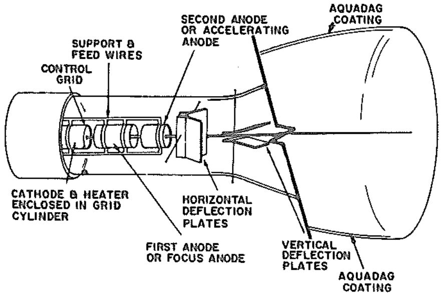
SECOND ANODE OR ACCELERATING ANODE
AQUADAG COATING
SUPPORT & FEED WIRES
CONTROL GRID
CATHODE & HEATER ENCLOSED IN GRID CYLINDER
HORIZONTAL DEFLECTION PLATES
FIRST ANODE OR FOCUS ANODE
VERTICAL DEFLECTION PLATES
AQUADAG COATING

Cathode-ray tube.

cathode-ray tube—Abbreviated crt. A tube in which its electron beam can be focused to a small cross section on a luminescent screen and can be varied in position and intensity to produce a visible pattern.

cathode-ray tuning indicator — Commonly called magic eye. A small-diameter cathode-ray tube that visually indicates whether an apparatus such as a radio receiver is tuned precisely to a station.

cathode resistor—A resistance connected in the cathode circuit of a tube so that the voltage drop across it will supply the proper cathode-biasing voltage.

cathode spot—On the cathode of an arc, the area from which electrons are emitted at a current density of thousands of amperes per square centimeter and where the temperature of the electrode is too low to account for such currents by thermionic emission.

cathode sputtering—*See* Sputtering.

cathodic protection—The control of the electrolytic corrosion of an underground or underwater metallic structure by the application of an electric current through a sacrificial anode in such a way that the structure is made to act as a cathode of an electrolytic cell.

cathodolumninescence — Luminescence produced by the bombardment with high-velocity electrons of a metal in a vacuum. Small amounts of the metal are vaporized in an excited state by the bombardment and emit radiation characteristic of the metal.

cation—A positive ion that moves toward the cathode in a discharge tube, electrolytic cell, or similar equipment. The corresponding negative ion is called an "anion."

catv — Abbreviation for community antenna television.

catwhisker—A small, pointed wire used to make contact with a sensitive area on the surface of a crystal or semiconductor.

cavitation—In ultrasonic cleaning, the local formation of cavities in a liquid as the total pressure in that region is reduced.

cavitation noise — The noise produced in a liquid by the collapse of the bubbles created by cavitation.

cavity — A metallic enclosure inside which resonant fields may be excited at a microwave frequency.

cavity filter—A selective tuned device having the proper coupling means for insertion into a transmission line to produce attenuation of unwanted off-frequency signals.

cavity impedance—The impedance that appears across the gap of the cavity of a microwave tube.

cavity magnetron—A magnetron having a number of resonant cavities forming the anode; used as a microwave-transmitting oscillator.

cavity radiation — The radiation (heat) emerging from a small hole leading to a constant-temperature enclosure. Such radiation is identical with black-body radiation at the same temperature, no matter what the nature of the inner surface of the enclosure.

cavity resonator—A space which is normally bounded by an electrically conducting surface and in which oscillating electromagnetic energy is stored; the resonant frequency is determined by the geometry of the enclosure.

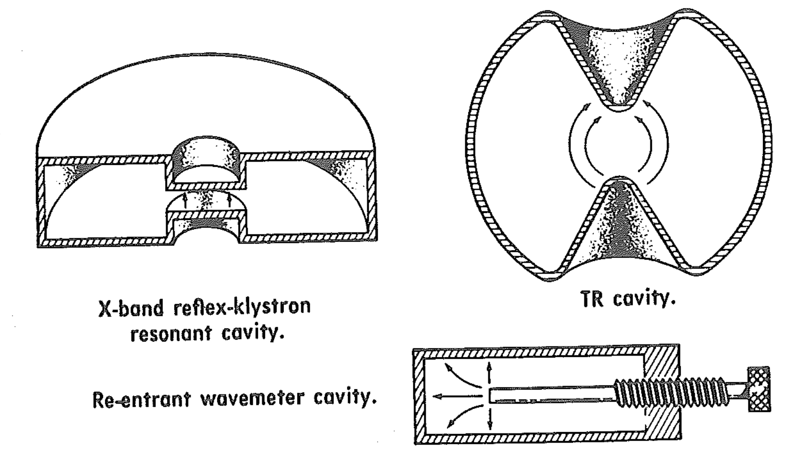

X-band reflex-klystron resonant cavity.

TR cavity.

Re-entrant wavemeter cavity.

Cavities.

cavity-resonator frequency meter—A cavity resonator used for determining the frequency of an electromagnetic wave.

cavity-tuned, absorption-type frequency meter—A device used for measuring frequency. Its operation depends on the use of an enclosure with a conductive inner wall; the resonant frequency of the wall is determined by its internal dimensions.

cavity-tuned, heterodyne-type frequency meter—A device for measuring frequency. Its operation depends on the use of an enclosure, the resonant frequency of which is determined by its internal dimensions.

cavity-tuned, transmission-type frequency meter—A device for measuring frequency. Its operation depends on the use of an enclosure with a conductive inner wall; the resonant frequency of the wall is determined by its internal dimensions.

C-band—A radio-frequency band of 3.9 to 6.2 GHz, with wavelengths of 7.69 to 4.84 cm. It includes the top two sidebands of S-band, and the bottom three sidebands of the X-band.

C-battery—Also called grid battery. The energy source which supplies the voltage for biasing the grid of a vacuum tube.

C-bias—*See* Grid Bias.

CCIF — Abbreviation for International Telephone Consultative Committee.

CCIR—Abbreviation for International Radio Consultative Committee.

CCIT — Abbreviation for International Telegraph Consultative Committee.

CCS—Abbreviation for continuous commercial service. Refers to the power rating of transformers, tubes, resistors, etc. Used for rating components in broadcasting stations and some industrial applications.

ccw—Abbreviation for counterclockwise.

C-display—A type of radar display in which the signal is a bright spot, with the bearing as the horizontal and the elevation angle as the vertical co-ordinate.

celestial guidance—A system of guidance in which star sightings that are automatically taken during the flight of a missile provide position information used by the guidance equipment.

cell—1. A single unit that produces a direct voltage by converting chemical energy into electrical energy. 2. A single unit that produces a direct voltage by converting radiant energy into electrical energy—for example, a solar or photovoltaic cell. 3. A single unit that produces a varying voltage drop because its resistance varies with illumination. 4. Elementary unit of storage. 5. In corrosion processes, a source of electric potential that is responsible for corrosion. It consists of an anode and a cathode immersed in an electrolyte and electrically bonded together. The anode and cathode may be separate metals or dissimilar areas on the same metal. The different metals will develop a difference in potential that is accompanied by corrosion of the anode. When this cell involves an electrolyte as it does in corrosion processes, it is referred to as an electrolytic cell.

cell counter—An electronic instrument used to count white or red blood cells or other very small particles.

cell-type enclosure — A prefabricated basic shielded enclosure of double-walled copper-mesh construction. The original screen-room design.

cell-type tube (TR, ATR, and PRE-TR tubes)—A gas-filled, radio-frequency switching tube which operates in an external resonant circuit. A tuning mechanism may be incorporated into the external resonant circuit or the tube.

cellulose-nitrate discs — *See* Lacquer Disc.

Celsius temperature scale—Also called centigrade temperature scale. A temperature scale based on the freezing point of water defined as 0°C. and the boiling point defined as 100°C. both under conditions of normal atmospheric pressure (760 mm of mercury).

center frequency—1. Also called resting frequency. The average frequency of the emitted wave when modulated by a symmetrical signal. 2. The frequency at the center of a spectrum display (for linear frequency scanning). It is usually tunable.

centering control—One of two controls used to shift the position of the entire image on the screen of a cathode-ray tube. The horizontal-centering control moves the image to the right or left, and the vertical-centering control moves it up or down.

centering diode—A clamping circuit used in some types of plan-position indicators.

center of gravity—A point inside or outside a body and around which all parts of the body balance each other.

center of mass—On a line between two bodies, the point around which the two bodies would revolve freely as a system.

center poise—Scale of viscosity for insulating varnishes.

center ring—The part that supports the stator in an induction-motor housing. The motor end shields are attached to the ends of the center ring.

center tap—A connection at the electrical center of a winding, or midway between the electrical ends of a resistor or other portion of a circuit.

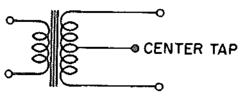

Center tap.

center-tapped winding—An internal tap at the center of a winding (three external leads).

center wire—A fine loop of wire used in proportional counters as an anode. A high voltage is applied to it to set the conditions for radiation measurement.

centigrade temperature scale—The older name for a Celsius temperature scale in the English-speaking countries. Officially abandoned by international agreement in 1948, but still in common usage.

centimeter waves—Microwave frequencies between 3 and 30 GHz, corresponding to wavelengths of 1 to 10 centimeters.

central processing unit—Also called central processor. Part of a computer system which contains the main storage, arithmetic unit, and special register groups. Performs arithmetic operations, controls instruction processing, and provides timing signals and other housekeeping operations.

Central Radio Propagation Laboratory—An organization, reporting to the Department of Commerce through the National Bureau of Standards, which is responsible for the collection, correlation, and analysis of data on which radio-propagation predictions are based, and issuance of radio-propagation predictions, and research on radio propagation and measurement methods and standards.

centrifugal force—The force which acts on a rotating body and which tends to throw the body farther from the axis of its rotation.

centripetal force—The force which compels a rotating body to move inward toward the center of rotation.

Ceracircuits—A trademark of the Sprague Electric Company for hybrid thick-thin film integrated circuits that consist of discrete passive and semiconductor active elements attached to precision resistor substrates to form functional electronic modules.

ceramic—A clay-like material, consisting primarily of magnesium and aluminum oxides, which after molding and firing is used as an insulating material. It withstands high temperatures and is less fragile than glass. When glazed, it is called porcelain.

ceramic-based microcircuit—A microminiature circuit printed on a ceramic substrate. Usually consists of combinations of resistive, capacitive, or conductive elements fired on a waferlike piece of ceramic.

ceramic capacitor—A capacitor the dielectric of which is a ceramic material such as steatite or barium titanate, the composition of which can be varied to give a wide range of temperature coefficients. The electrodes are usually silver coatings, fired on opposite sides of the ceramic disc or slab, or fired on the inside and outside of a ceramic tube. After connecting leads are soldered to the electrodes, the unit is usually given a protective insulating coating.

ceramic microphone—A microphone with a ceramic cartridge.

ceramic permanent magnet—A permanent, nonmetallic magnet made from pressed and sintered mixtures of metallic-oxide powders, usually oxides of barium and iron.

ceramic pickup—A phonograph pickup with a ceramic cartridge.

ceramic transducer—*See* Piezoelectric Transducer.

Cerenkov radiation—Light emitted when charged particles pass through a transparent material at a velocity greater than that of light in that material. It can be seen, for example, as a blue glow in the water around the fuel elements of pool reactors. P.A. Cerenkov was the Russian scientist who first explained the origin of this light.

Cerenkov rebatron radiator—A device in which a tightly bunched velocity-modulated electron beam is passed through a hole in a dielectric. The reaction between the higher velocity of the electrons passing through the hole and the slower velocity of the electromagnetic energy passing through the dielectric results in radiation at some frequency higher than the frequency of modulation of the electron beam.

cermet—A metal-dielectric mixture used in making thin-film resistive elements. The first half of the term is derived from *cer*amic and the second half from *met*al.

cesium—A chemical element having a low work function. Used as a getter in vacuum tubes and in cesium-oxygen-silver photocell cathodes.

cesium-vapor lamp—A low-voltage arc lamp for producing infrared radiation.

C_{gk}—Symbol for grid-cathode capacitance in a vacuum tube.

C_{gp}—Symbol for grid-plate capacitance in a vacuum tube.

cgs electromagnetic system of units—A coherent system of units for expressing the magnitude of electrical and magnetic quantities. The most common fundamental units of these quantities are the centimeter, gram, and second. Their unit of current (abampere) is of such a magnitude that if maintained constant in two straight parallel conductors having an infinite length and negligible circular sections and placed one centimeter apart in a vacuum, a force equal to 2 dynes per centimeter of length will be produced.

cgs electrostatic system of units—A coherent system of units for expressing the magnitude of electrical and magnetic quantities. The most common fundamental units of these quantities are the centimeter, gram, and second. Their unit of electrical charge (statcoulomb) is of such a magnitude that two equal unit point charges one centimeter apart in a vacuum will repel each other with a force of one dyne.

chad—The piece of material removed when a hole or notch is formed in a storage medium such as punched tape or punched cards.

chadless tape—A type of punched paper tape in which each chad is left fastened by about a quarter of the circumference of the hole. Chadless punched paper tape must be sensed by mechanical fingers, because chad interferes with reliable electrical or photoelectrical reading.

chafe—Undesirable rubbing with friction.

chaff—A general name applied to radar-confusion reflectors which consist of thin, narrow, metallic strips of various length and frequency responses used to reflect radar echoes.

chaining—In a computer, a system of storing records such that each record belongs to a list or group of records and has a linking field for tracing the chain.

chain printer—In a computer, a high-speed printer having type slugs carried on the links of a revolving chain.

chain radar system—A radar system comprising a number of radars or radar stations located at various sites along a missile-flight path. These radar stations are linked together by data and communication lines for target acquisition, target positioning, and/or data-recording purposes. The target-acquisition link makes it possible for any radar to position any other radar on target.

chain reaction—A reaction that stimulates its own repetition. In a fission chain reaction a fissionable nucleus absorbs a neutron and fissions, releasing additional neutrons. These in turn can be absorbed by other fissionable nuclei, releasing still more neutrons. A fission chain reaction is self-sustaining when the number of neutrons released in a given time equals or exceeds the number of neutrons lost by absorption in nonfissioning material or by escape from the system.

challenger—*See* Interrogator.

chance failure—*See* Random Failure.

channel—1. A portion of the spectrum assigned for the operation of a specific carrier and the minimum number of sidebands necessary to convey intelligence. 2. A single path for transmitting electric signals. (Note: The word "path" includes separation by frequency division or time division. Channel may signify either a one-way or two-way path, providing communication in either one direction only or in two directions.) *Also see* Alternate Channel. 3. In electric computers, that portion of a storage medium which is accessible to a given reading station. 4. The path along which information, particularly a series of digits or characters, may flow. 5. In computer circulating storage, one recirculating path containing a fixed number of words stored serially by word. 6. The main current path between the source and the drain in a field-effect transistor or other semiconductor.

channel balance—Equal response on both left and right channels.

channel capacity—The maximum possible information rate through a channel, subject to the constraints of that channel. May be either per second or per symbol.

channel frequency—The band of frequencies which must be handled by a carrier system to transmit a specific quantity of information.

channeling—The utilization of a modulation-frequency band for the simultaneous transmission from two or more communication channels in which the channel separation is accomplished by the use of carriers or subcarriers, each in a different discrete frequency band forming a subdivision of the main band. This covers a special case of multiplex transmission.

channel interval—The time allocated to a channel, including on and off time.

channelizing—The process of subdividing wideband transmission facilities for the purpose of putting many different circuits requiring comparatively narrow bandwidths on a single wideband facility.

channel pulse—Intelligence on a channel, by virtue of the modulation characteristic.

channel pulse synchronization—Synchronization of a local-channel rate oscillator by comparison and phase lock with separate channel synchronizing pulses.

channel reliability—The percentage of time that the channel meets the arbitrary standards established by the user.

channel reversal—Shifting the outputs of a stereo system so the channel formerly heard from the left speaker now comes from the right and vice versa.

channel-reversing switch—A switch that reverses the connections of two speakers in a stereo system with respect to the channels, so that the channel heard previously from the right speaker is heard from the left and vice versa.

channel sampling rate — The number of times per second that individual channels are sampled. This is different from commutation rate, since it is possible for more than one channel to be applied to a given commutator input (with subcommutation).

channel selector—A switch or dial used for selecting a desired channel.

channel separation—In stereo the electrical or acoustical difference between the left and right channels. Inadequate separation can lessen the stereo effect; excessive separation can exaggerate it beyond natural proportions.

channel strip—An amplifier, or other device having a sufficiently wide bandpass to amplify one television channel.

channel subcarrier—The channel required to convey telemetric data involving a subcarrier band.

channel synchronizing pulse separator—A device for separating channel synchronizing pulses from commutated signals.

channel-to-channel connection — A device for rapid data transfer between two computers. A channel adapter is available that permits the connection between any two channels on any two systems. Data are transferred at the rate of the slower channel.

channel-utilization index—In a computer, the ratio between the information rate (per second) through a channel and the channel capacity (per second).

channel wave—Any elastic wave propagated in a sound channel because of a low-velocity layer in the solid earth, the sea, or the atmosphere.

character—1. In electronic computers, one of a set of elementary symbols which may be collectively arranged in order to express information. These symbols may include the decimal digits 0 through 9, the letters A through Z, punctuation and typewriter symbols, and any other single symbol which a computer may read, store, or write. 2. One of a set of symbols used to present information on a display tube.

character crowding—The effect of reducing the time interval between subsequent characters read from tape. It is caused by a combination of mechanical skew, gap scatter, jitter, amplitude variation, etc. Also called packing.

character emitter—In a computer, an electromechanical device that puts out coded pulses.

characteristic—An inherent and measurable property of a device. Such a property may be electrical, mechanical, thermal, hydraulic, electromagnetic, or nuclear; and it can be expressed as a value for stated or recognized conditions. A characteristic may also be a set of related values (usually in graphical form).

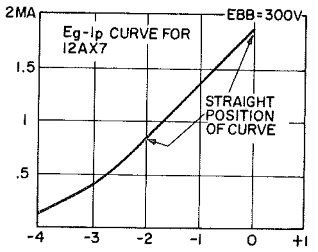

Characteristic curve.

characteristic curve—A graph plotted to show the relationship between changing values. An example would be a curve showing plate-current changes as the grid voltage varies.

characteristic distortion — Displacement of signal transitions due to the persistence of transients caused by preceding transitions.

characteristic frequency — The frequency which can be easily identified and measured in a given emission.

characteristic impedance—Also called surge impedance. 1. The driving-point impedance of a line if it were of infinite length. 2. In a delay line, the value of terminating resistance which provides minimum reflection to the network input and output. 3. The ratio of voltage to current at every point along a transmission line on which there are no standing waves. 4. The square root of the product of the open- and short-circuit impedance of the line.

characteristic impedance of free space—The relationship between the electric and magnetic intensities of space due to the expansion of the impedance concept to electromagnetic fields.

characteristic spread—The range between the minimum and maximum values for a given characteristic that is considered normal in any large group of tubes or other devices.

characteristic telegraph distortion—Distortion which does not affect all signal pulses alike. Rather, the effect on each transition depends on the signal previously sent, because remnants of previous transitions or transients persist for one or more pulse lengths.

characteristic wave impedance—The ratio of the transverse electric vector to the transverse magnetic vector at the point it is crossed by an electromagnetic wave.

character reader — A computer input device which can directly recognize printed or written characters; they need not first be converted into punched holes in cards or paper or polarized magnetic spots.

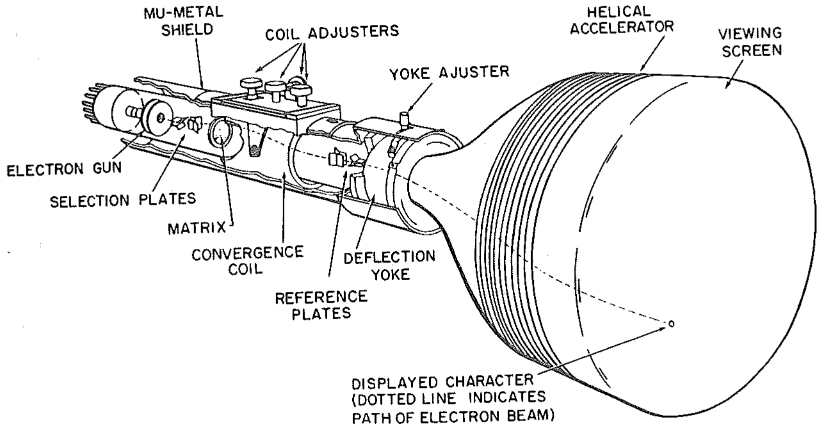

Charactron.

character recognition — In a computer, the act of reading, identifying, and encoding a printed character.

character sensing — To detect the presence of characters optically, magnetically, electrostatically, etc.

Charactron—Trade name of General Dynamics/Electronics for a specially constructed cathode-ray tube used to display alphanumeric characters and other special symbols directly on its screen.

charge—1. The electrical energy stored in a capacitor or battery or held on an insulated object. 2. The quantity of electrical energy in (1) above.

charge carrier—A mobile hole or conduction electron in a semiconductor.

charge density—The charge per unit area on a surface, or charge per unit volume in space.

charged particle — An ion, an elementary particle that carries a positive or negative electric charge.

charger—A device used to convert alternating current into a pulsating direct current which can be used for charging a storage battery.

charge retention—The ability of a battery to hold its energy once it has been charged.

charge-storage tube—A storage tube which retains information on its surface in the form of electric charges.

charge transfer—The process in which an ion takes an electron from a neutral atom of the same type with a resultant transfer of electronic charge.

charging—The process of converting electrical energy to stored chemical energy.

charging current—The current that flows into a capacitor when a voltage is first applied.

charging rate—The rate of current flow used in charging a battery.

chassis—1. A sheet-metal box, frame, or simple plate on which electronic components and their associated circuitry can be mounted. 2. The entire equipment (less cabinet) when so assembled. (*Also see* Printed-Circuit Board.)

chassis ground—A connection to the metal structure that supports the electrical components which make up the unit or system.

chatter — 1. Prolonged undesirable opening and closing of electrical contacts. 2. The vibration of a cutting stylus in a direction other than the direction in which it is driven.

chatter time—The interval of time from initial actuation of a contact to the end of chatter.

check—The partial or complete verification of the correctness of equipment operations, the existence of certain prescribed conditions, and/or the correctness of results.

check digit—A digit carried along with a machine word and used to report information about the other digits in the word so that if a single error occurs, the check fails and an error alarm signal is initiated.

checkerboard—*See* Worst-Case Noise Pattern.

check point—In a computer routine, a point at which it is possible to store sufficient information to permit restarting the computation from that point.

check-point routine—A computer routine in which information for a check point is stored.

check problem—A problem which, when incorrectly solved, indicates an error in the programming or operation of a computer.

check register—A special register provided in some computers to temporarily store transferred information for comparison with a second transfer of the same information in order to verify that the information transferred each time agrees precisely.

check routine — A program the purpose of which is to determine whether a computer or a program is operating correctly.

checksum—In a computer, a summation of digits or bits summed according to an arbitrary set of rules and primarily used for checking purposes.

cheese antenna—An antenna with a cylindrical parabolic reflector enclosed by two plates perpendicular to the cylinder and so spaced that more than one mode can be propagated in the desired direction of polarization. It is fed on the focal line.

chelate—A molecule in which a central inorganic ion is covalently bonded to one or more organic molecules, with at least two bonds to each molecule; used as a laser dopant.

chemical deposition—The process of depositing a substance on a surface by means of the chemical reduction of a solution.

chemically deposited printed circuit—A printed circuit formed on a base by the reaction of chemicals alone. Dielectric, magnetic, and conductive circuits can be applied.

chemically reduced printed circuit—A printed circuit formed by chemically reducing a metallic compound.

chemisorption—*See* Adsorption.

Child's law—Also known as the three-halves power equation. It states that the current in a thermionic diode varies directly with the three-halves power of the anode voltage and inversely with the square of the distance between electrodes.

chip—1. Also called thread. In mechanical recording, the material removed from the recording medium by the recording stylus as it cuts the groove. 2. In punched cards, a piece of cardboard removed in the punching process. 3. A single substrate on which all the active and passive elements of an electronic circuit have been fabricated utilizing the semiconductor technologies of diffusion, passivation, masking, photoresist, and epitaxial growth. A chip is not ready for use until it is packaged and provided with terminals for connection to the outside world.

chirp—1. An all-encompassing term for the various techniques of pulse expansion-pulse compression applied to pulse radar. A technique to expand narrow pulses to wide pulses for transmission, and to compress wide received pulses to the original narrow pulse width and waveshape. This improves the signal-to-noise ratio without degradation to the range resolution and range discrimination. 2. A colloquial expression for a coded pulse. In coding the pulse, the carrier frequency is increased in a linear manner for the duration of the pulse, and when the pulse is translated to an audio frequency, it sounds like a chirp. 3. A change in the pitch of code signals, generally due to poor regulation of the transmitter power supply.

choke—An inductance used to impede the flow of pulsating direct current or alternating current by means of its self-inductance.

choke coil—Also called impedance coil. An inductor (reactor) used to limit or suppress the flow of alternating current without appreciable effect on the flow of direct current.

choke flange—A waveguide flange with a grooved surface; the groove is so dimensioned that the flange forms part of a choke joint.

choke-input filter—A power-supply filter in which a choke is the first element in series with the input.

choke joint—1. A connector between two sections of transmission line in which the gap between sections to be connected is built out to form a series-branching transmission line carrying a standing wave, in which actual contact falls at or near a current minimum. 2. A joint for connecting two sections of waveguide together. Permits efficient energy transfer without the necessity of an electrical contact at the inside surface of the guide.

chopped mode—A time-sharing method of displaying output signals of two or more channels with a single crt gun, in sequence, at a rate not referenced to the sweep.

chopper—1. A device for interrupting a current or a light beam at regular intervals. Choppers are frequently used to facilitate amplification. 2. An electromechanical switch for the production of modified square waves. The waves are of the same frequency as a driving sine wave and bear a definite relationship to it. 3. An electromechanical or electronic device used to interrupt a d-c or low-frequency a-c signal at regular intervals to permit amplification of the signal by an a-c amplifier. It may also be used as a demodulator to convert an a-c signal to dc. 4. A rotating shutter for interrupting an otherwise continuous stream of particles. Choppers can release short bursts of neutrons with known energies. Used to measure nuclear cross sections.

chopper amplifier—A circuit that amplifies a low-level signal after it has gone through a chopper.

chopper stabilization—1. The addition of a chopper amplifier to the regulator input circuitry of a regulated power supply in order to reduce output drift. 2. A method of improving the d-c drift of an amplifier by utilization of chopper circuits.

chopper-stabilized amplifier—An amplifier configuration utilizing a carrier-type d-c amplifier to reduce the effect of input offset and drift of a direct-coupled amplifier.

chopping—Removal by electronic means of one or both extremities of a wave at a predetermined level.

chord—A harmonious combination of tones sounded together through the use of one or more fingers on either or both hands. On chord organs a full chord is selected by depressing a single chord button.

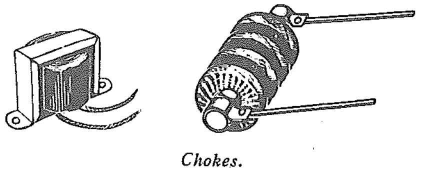

Chokes.

chord organ—An organ with provision for playing a variety of chords, each produced by means of a single button or key.

chorus—A natural electromagnetic phenomenon in the vlf range. Probably originates in the exosphere. Also called "dawn chorus" because it sounds like birds at dawn. It generally consists of a multitude of rising tones, each tone rising from 1-2 kHz to 3-4 kHz and usually lasting 0.1 to 0.5 second.

Christmas-tree pattern—1. *See* Optical Pattern. 2. A pattern resembling a Christmas tree, sometimes produced on the screen of a television receiver when the horizontal oscillator falls out of sync.

chroma—That quality which characterizes a color without reference to its brightness; that quality which embraces hue and saturation. White, black, and gray have no chroma.

chroma-clear raster—Also called white raster. Looks like a clear raster, but each of the three guns in the crt is operating under the influence of a color level determined by a white video signal. In this case, all tv set chroma circuits are working as though the tv set were receiving a color transmission of a completely white scene.

chroma control—A variable resistor which controls saturation by varying the level of chrominance signal fed to the demodulators of a color television receiver.

chromatic aberration—An effect which causes refracted white light to produce an image with colored fringes due to the various colors being bent at different angles.

chromaticity—1. The combination of the hue and saturation attributes of color. 2. A term quantitatively descriptive of a color, and dependent upon both hue and saturation, but without reference to brilliance.

chromaticity coordinate—The ratio of any one of the tristimulus values of a sample color to the sum of the three tristimulus values.

chromaticity diagram—A plane diagram formed when any one of the three chromaticity co-ordinates is plotted against another.

chromaticity flicker—The flicker which results from fluctuation of the chromaticity only.

chrominance—Colorimetric difference between any color and a reference color of equal luminance, the reference color having a specified chromaticity. In standard color-television transmission, the specified chromaticity is that of the zero subcarrier.

chrominance amplifier—The amplifier that separates the chrominance signal from the total video signal.

chrominance cancellation—A cancellation of the brightness variations produced by the chrominance signal on the screen of a monochrome picture tube.

chrominance-carrier reference—A continuous signal having the same frequency as the chrominance subcarrier, and a fixed phase with the color burst. The phase reference of carrier-chrominance signals for modulation or demodulation.

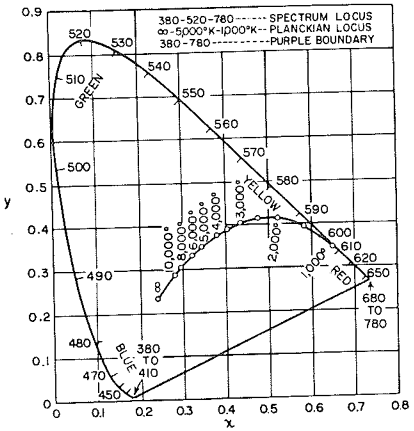

Chromaticity diagram.

chrominance channel—In color television, a combination of circuits designed to pass only those signals having to do with the reproduction of color.

chrominance component—Either of the I and Q signals which add to produce the complete chrominance signal in NTSC systems.

chrominance demodulator—A demodulator used in color-television reception for deriving video-frequency chrominance components from the chrominance signal and a sine wave of the chrominance subcarrier frequency.

chrominance gain control—In red, green, and blue matrix channels, variable resistors which individually adjust the primary-signal levels. Used in color television.

chrominance modulator—A modulator used in color-television transmission for generating the chrominance signal from the video-frequency chrominance components and the chrominance subcarrier.

chrominance primary—One of two transmission primaries, the amounts of which deter-

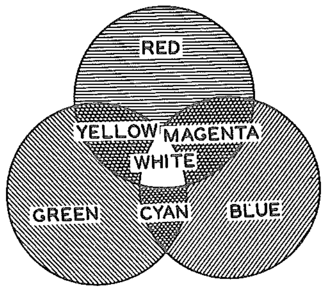

Chrominance primaries.

mine the chrominance of a color. Chrominance primaries have zero luminance and are not physical.

chrominance signal—That portion of the composite color signal used to represent electrically the hues and saturation levels of the colors in a televised scene.

chrominance subcarrier — Also called color carrier. An r-f signal which has a specific frequency of 3.579545 MHz and which is used as a carrier for the I and Q signals.

chrominance-subcarrier oscillator — In a color-tv receiver, a crystal-controlled oscillator which generates the subcarrier signal for use in the chrominance demodulators.

chrominance video signals—Output voltages from the red, green, and blue sections of a color-television camera or receiver matrix.

chronistor—A subminiature elapsed-time indicator that uses electroplating principles to totalize the operating time of equipment up to several thousand hours.

chronograph—An instrument for producing a graphical record of time as shown by a clock or other device.

chronometer — A portable timekeeper with compensated balance, capable of showing time with high precision and accuracy.

chronoscope—An instrument for measuring very small intervals of time.

CIE — Initials of the Commission Internationale de l'Eclairage or International Commission on Illumination.

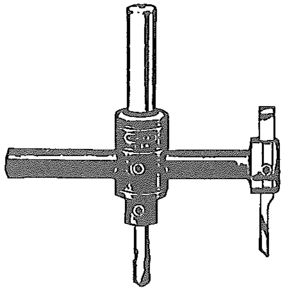

Circle cutter.

circle cutter—A tool consisting of a center drill with an adjustable extension-arm cutter and used to cut holes in panels and chassis.

circle of confusion—The circular image of a point source due to the inherent aberrations in an optical system.

circlotron amplifier—A one-port, nonlinear cross-field high power microwave amplifier which uses a magnetron as a negative-resistance element, much as a maser uses an active material.

circuit—1. An electronic path between two or more points capable of providing a number of channels. 2. A number of conductors connected together for the purpose of carrying an electrical current. 3. The interconnection of a number of devices in one or more closed paths to perform a desired electrical or electronic function. Examples of simple circuits are high- or low-pass filters, multivibrators, oscillators, and amplifiers.

circuit analyzer — Also called multimeter. Several instruments or instrument circuits combined in a single enclosure and used in measuring two or more electrical quantities in a circuit.

circuit breaker—An automatic device which, under abnormal conditions, will open a current-carrying circuit without damaging itself (unlike a fuse, which must be replaced when it blows).

circuit-breaker cascade system—A system wherein the protective devices are arranged in order of ratings such that those in series will coordinate and provide the required protection.

circuit capacity—The number of communication channels which can be handled by a given circuit at the same time.

circuit dropout—A momentary interruption of a transmission because of the complete failure of a circuit.

circuit efficiency (of the output circuit of electron tubes)—Ratio of the power, at the desired frequency, delivered to a load at the output-circuit terminals of an oscillator or amplifier, to the power, at the desired frequency, delivered to the output circuit by the electron stream.

circuit element—Any basic constituent of a circuit except the interconnections.

circuit hole — On a printed-circuit board, a hole that lies partially or completely within the conductive area.

circuit noise—The noise brought to the receiver electrically from a telephone system, but not the noise picked up acoustically by the telephone transmitters.

circuit noise level—At any point in a transmission system, the ratio of the circuit noise at that point to some arbitrary amount of circuit noise chosen as a reference. This ratio is usually expressed in decibels above reference noise, abbreviated dbrn, signifying the reading of a circuit noise meter; or in adjusted decibels, abbreviated dba, signifying the circuit noise meter reading adjusted to represent interfering effect under specified conditions.

circuit-noise meter—Also called noise-measuring set. An instrument for measuring the circuit-noise level. Through the use of a suitable frequency-weighting network and other characteristics, the instrument gives equal readings for noises of approximately equal interference. The readings are expressed in decibels above the reference noise.

circuit parameters—The values of the physical quantities associated with circuit elements—for example, the resistance (parameter) of a resistor (element), the amplification factor and plate resistance (parameters) of a tube (element), the inductance per unit length (parameter) of a transmission line (element).

circuit protection—Automatic protection of a consequence-limiting nature used to minimize the danger of fire or smoke, as well as the disturbance to the rest of the system which may result from electrical faults or prolonged electrical overloads.

circuit re-entrancy—*See* Re-entrancy.

circuit reliability—The percentage of time the circuit meets arbitrary standards set by the user.

circuitry—A complex of circuits describing interconnection within or between systems.

circular antenna—A horizontally polarized antenna derived essentially from a half-wave antenna but having its elements bent into a circle.

circular electric wave—A wave with circular electric lines of force.

circular magnetic wave—A wave with circular magnetic lines of force.

circular mil—The universal term used to define cross-sectional areas. Equal to the area of a circle one mil (.001 inch) in diameter.

circular mil area—The square of the diameter of a round conductor measured in thousandths of an inch. The circular mil area of a braid is the sum of the circular mil area of each of the wires that make up the braid.

circular polarization—Polarization such that the vector representing the wave has a constant magnitude and rotates continuously about a point.

circular polarized wave—Applied usually to transverse waves. An electromagnetic wave for which the electric and/or magnetic field vector at a point describes a circle.

circular scan—A radar scan in which a given point on the direct beam describes a circle in space, with the antenna at the center.

circular scanning—Scanning in which the direction of maximum radiation generates a plane or a right circular cone with a vertex angle close to 180°.

circular trace—A cro time base produced by applying sine waves of the same frequency and amplitude, but 90° out of phase, to the horizontal- and vertical-deflection plates of a cathode-ray tube. This results in a circular trace, and signals then give inward or outward radial deflections from the circle.

circular waveguide—A waveguide having a circular cross-sectional area.

circulating memory — *See* Circulating Register.

circulating register—Also called circulating memory. A register (or memory) consisting of a means for delaying the information and a means for regenerating and reinserting it into the delaying means. This is accomplished as the information moves around a loop and returns to its starting place after a fixed delay.

circulating storage—A device using a delay line to store information in a train or pattern of pulses. The pulses at the output end are sensed, amplified, reshaped, and re-inserted into the input end of the delay line.

circulator — A microwave coupling device having a number of terminals so arranged that energy entering one terminal is transmitted to the next adjacent terminal in a particular direction.

Citizens band channels—

Channel No.	*Frequency (MHz)*	*Channel No.*	*Frequency (MHz)*
1	26.965	13	27.115
2	26.975	14	27.125
3	26.985	15	27.135
4	27.005	16	27.155
5	27.015	17	27.165
6	27.025	18	27.175
7	27.035	19	27.185
8	27.055	20	27.205
9	27.065	21	27.215
10	27.075	22	27.225
11	27.085	23*	27.255
12	27.105		

*This channel is shared with class-C radio control.

Citizens Radio Service—A type of radio service intended for private or personal use for communication, radio control, and similar purposes. A license can be obtained from the FCC; there is no need to take a technical examination. The frequencies assigned are in the 460-470-MHz and 26.965-27.255-MHz bands.

clamp—*See* Clamping Circuit.

clamper—*See* Clamping Circuit.

clamping — The process that establishes a fixed level for the picture signal at the beginning of each scanning line.

clamping circuit — Also called clamper or clamp. A circuit which adds a fixed bias to a wave at each occurrence of some predetermined feature of the wave. This is done to hold the voltage or current of the feature at (clamp it to) a specified fixed or variable level. (*Also see* D-C Restorer.)

clamping diode—A diode used to fix a voltage level at some point in a circuit.

clapotis—A standing wave in which there is no horizontal motion of the crest. This phenomenon results from the interference caused by the reflection of a wave train from a barrier which the wave train approaches with its crest parallel to the barrier.

clapper—A hinged or pivoted armature.

Clapp oscillator — A Colpitts-type oscillator using a series-resonant tank circuit for improved stability.

Clark cell—An early standard cell that used an anode of mercury, a cathode of zinc amalgam, and an electrolyte containing zinc sulphate and mercurous sulphate. Its voltage is 1.433 at 15°C.

class-A amplifier—An amplifier in which the grid bias and alternating grid voltage are such that plate current flows at all times. To denote that no grid current flows during any part of the input cycle, the suffix *1* is sometimes added to the letter or letters of the class identification. The suffix *2* denotes that grid current flows during part of the input cycle.

class-AB amplifier—An amplifier in which the grid bias and alternating grid voltage are such that plate current flows for more than half but less than the entire electrical cycle. To denote that no grid current flows during any part of the input cycle, the suffix *1* is sometimes added to the letter or letters of the class identification. The suffix *2* denotes that grid current flows during part of the cycle.

class-AO emission—The incidental radiation of an unmodulated carrier wave from a station.

class-A1 emission—A carrier wave (unmodulated by an audio frequency) keyed normally for telegraphy to transmit intelligence in the International Morse Code at a speed not exceeding 40 words per minute (the average word is composed of five letters).

class-A2 emission—A carrier wave which is amplitude-modulated at audio frequencies not exceeding 1,250 hertz. The modulated carrier wave is keyed normally for telegraphy to transmit intelligence in the International Morse Code at a speed not exceeding 40 words per minute, the average word being composed of five letters.

class-A3 emission—A carrier wave which is amplitude-modulated at audio frequencies corresponding to those necessary for intelligible speech transmitted at the speed of conversation.

class-A insulating material—A material or combination of materials such as cotton, silk, and paper suitably impregnated, coated, or immersed in a dielectric liquid such as oil. Other materials or combinations of materials may be included if shown to be capable of satisfactory operation at 105°C.

class-A modulator—A class-A amplifier used for supplying the signal power needed to modulate the carrier.

class-A operation—Operation of a vacuum tube with grid bias such that plate current flows throughout 360° of the input cycle.

class-A signal area—A strong tv signal area for vhf, 2,500 microvolts per meter or greater; and for uhf, 5,000 microvolts per meter or greater.

class-A transistor amplifier—An amplifier in which the input electrode and alternating input signal are biased so that output current flows at all times.

class-B amplifier—An amplifier in which the grid bias is approximately equal to the cutoff value so that, when no exciting grid voltage is applied, the plate current will be approximately zero and will flow for approximately half of each cycle when an alternating grid voltage is applied. To denote that no grid current flows during any part of the input cycle, the suffix *1* is sometimes added to the letter or letters of the class identification. The suffix *2* denotes that grid current flows during part of the cycle.

class-B insulating material — A material or combination of materials such as mica, glass fiber, asbestos, etc., suitaby bonded. Other materials or combinations, not necessarily inorganic may be included if shown to be capable of satisfactory operation at 130°C.

class-B modulator—A class-B amplifier used specifically for supplying the signal power needed to modulate a carrier.

class-B operation—Operation of a vacuum tube with the grid bias set at or very near cutoff, so that plate current flows for approximately the positive half of each cycle of the input signal.

class-B transistor amplifier—An amplifier in which the input electrode is biased so that when no alternating input signal is applied, the output current is approximately zero, and when an alternating input signal is applied, the output current flows for approximately half a cycle.

class-C amplifier—An amplifier in which the grid bias is appreciably beyond the cutoff point, so that plate current is zero when no alternating grid voltage is applied, and plate current flows for appreciably less than half of each cycle when an alternating grid voltage is applied. To denote that no grid current flows during any part of the input cycle, the suffix *1* is sometimes added to the letter or letters of the class identification. The suffix *2* denotes that grid current flows during part of the cycle.

class-C insulating material—Insulation consisting entirely of mica, porcelain, glass, quartz, or similar inorganic materials. Other materials or combinations of materials may be included if shown to be capable of satisfactory operation at temperatures over 220°C.

class-C operation—Operation of a vacuum tube with grid bias considerably greater than cutoff. The plate current is zero with no input signal to the grid and flows for appreciably less than one-half of each cycle of the input signal.

class-F insulating material—A material or combination of materials such as mica, glass fiber, asbestos, etc., suitably bonded. Other materials or combinations of materials, not necessarily inorganic, may be included if shown to be capable of satisfactory operation at 155°C.

class-H insulating material—A material or combination of materials such as silicone elastomer, mica, glass fiber, asbestos, etc., suitably bonded. Other materials or combinations of materials may be included if shown to be capable of satisfactory operation at 180°C.

class-J oscilloscope—*See* J-Scope.

class-O insulating material—An unimpregnated material or combination of materials such as cotton, silk, or paper. Other materials or combinations of materials may be included if shown to be capable of satisfactory operation at 90°C.

clavier — Any keyboard, either hand or foot operated.

clean room—An area in which high standards of control of humidity, temperature, dust and all forms of contamination are maintained.

clear—Also called reset. To restore a storage or memory device to a prescribed state, usually to zero.

clearance — The shortest distance through space between two live parts, between live parts and supports or other objects, or between any live part and grounded part.

clearing ends — The operation of removing the sheath from the end of a cable, eliminating all moisture, and checking for crosses, shorts, and grounds in preparation for testing.

clear raster—A raster free of snow such as would be obtained in the absence of a video signal on either the cathodes or the grids of the three guns in the color crt (mostly a function of bias conditions).

click filter—A capacitor and resistor connected across the contacts of a switch or relay to prevent a surge being introduced into an adjacent circuit. (*Also see* Key-Click Filter.)

clipped-noise modulation—A clipping action performed to increase the bandwidth of a jamming signal. Results in more energy in the sidebands, correspondingly less energy in the carrier, and an increase in the ratio of average power to peak power.

clipper—A device the output of which is zero or a fixed value for instantaneous input amplitudes up to a certain value, but is a function of the input for amplitudes exceeding the critical value.

clipper amplifier—An amplifier designed to limit the instantaneous value of its output to a predetermined maximum.

clipper-limiter—Also called slicer. A device the output of which is a function of the instantaneous input amplitude for a range of values lying between two predetermined limits, but is approximately constant at another level for input values above the range.

clipping—The loss of initial or final parts of words or syllables due to less-than-ideal operation of voice-operated devices.

clock—The primary source of synchronizing signals in electronic computers.

clocked flip-flops—Directly coupled units operating on the "master-slave" principle. Information enters the "master" while the trigger input voltage is high, and transfers to the "slave" when the trigger input voltage goes low.

clock frequency—In digital computers, the master frequency of periodic pulses that are used to schedule the operation of the computer.

clock rate—The rate at which a word or characters of a word (bits) are transferred from one internal computer element to another. Clock rate is expressed in cycles (in a parallel-operation machine, in words; in a serial-operation machine, in bits) per second.

clockwise-polarized wave — *See* Right-Handed Polarized Wave.

close coupling—Also called tight coupling. Any degree of coupling greater than critical coupling.

closed array—An array which cannot be extended at either end.

closed circuit—1. A complete electric circuit through which current may flow when a voltage is applied. 2. A program source that is not broadcast for general consumption, but is fed to remote monitoring units by wire.

closed-circuit communication systems — Certain communication systems which are entirely self-contained and do not exchange intelligence with other facilities and systems.

closed-circuit signaling—Signaling in which current flows in the idle conditions and a signal is initiated by increasing or decreasing the current.

closed-circuit television—A television system in which the television signals are not broadcast, but are transmitted over a closed circuit and received only by interconnected receivers.

closed entry—A design that places a limit on the size of a mating part.

closed loop—1. A circuit in which the output is continuously fed back to the input for constant comparison. 2. In a computer, a group of indefinitely repeated instructions.

closed-loop bandwidth — The frequency at which the closed-loop gain drops 3 db from its midband or d-c value.

closed-loop gain — The response of a feedback circuit to a voltage inserted in series with the amplifier input.

closed-loop input impedance—The impedance looking into the input port of an amplifier with feedback.

closed-loop output impedance—The impedance looking into the output port of an operational amplifier with feedback.

closed-loop voltage gain—The voltage gain of an amplifier with feedback.

closed magnetic circuit—A circuit in which the magnetic flux is conducted continuously around a closed path through ferromagnetic materials.

closed routine — In a computer, a routine that is entered by basic linkage from the main routine rather than being inserted as a block of instructions within a main routine.

close-talking microphone—Also called noise-cancelling microphone. A microphone designed to be held close to the mouth of the speaker.

closing rating—In a relay, conditions under which the contact must close, with a prescribed duty cycle and contact life.

cloverleaf antenna—An nondirectional vhf transmitting antenna that consists of a number of horizontal four-element radiators arranged much like a four-leaf clover, stacked a half-wave apart vertically. These horizontal units are energized to give maximum radiation in the horizontal plane.

clutter—Confusing, unwanted echoes which interfere with the observation of desired signals on a radar display.

clutter gating—The technique which provides switching between moving-target indicators and normal videos. This results in normal video being displayed in regions with no clutter, and moving-target indicator video being switched in only for the clutter areas.

CML—Abbreviation for current mode logic.

C-network—A network composed of three impedance branches in series. The free ends are connected to one pair of terminals, and the junction points to another pair.

coarse-chrominance primary—Also called the Q signal. A zero-luminance transmission primary associated with the minimum bandwidth of chrominance transmission and chosen for its relatively small importance in contributing to the subjective sharpness of the color picture.

coast—On a radar, a memory feature which, when activated, causes the range and/or angle systems to continue to move in the same direction and at the same speed as an original target was moving. Used to prevent lock-up to a stronger target if approached by the target being tracked.

coast station—A land-based radio station in the maritime mobile service. It carries on communications with shipboard stations.

coated cathode—In a vacuum tube, a cathode that has been coated with compounds so as to increase its electron emission (e.g., an oxide-coated cathode).

coated filament—A vacuum-tube filament that has been coated with metal oxides to increase its electron emission.

coating—The magnetic layer, consisting of oxide particles held in a binder that is applied to the base film, used for magnetic recordings.

coating thickness—The thickness of the magnetic coating applied to the base film. Modern tape coatings range in thickness from 170 to 650 microinches, with a preponderance of coatings being approximately 400 microinches thick. In general, thin coatings give good resolution, at the expense of reduced output at long wavelengths; thick coatings give a high output at long wavelengths, at the expense of degraded resolution.

coated tape—*See* Magnetic-Powder—Coated Tape.

coaxial antenna—An antenna comprised of a quarter-wavelength extension to the inner conductor of a coaxial line, and a radiating sleeve which in effect is formed by folding back the outer conductor of the coaxial line for approximately one-quarter wavelength.

coaxial cable—*See* Coaxial Line.

coaxial cavity—A cylindical reasonating cavity that has a central conductor in contact with its movable pistons or other reflecting devices. The conductor serves to pick up a desired wave in the microwave region.

coaxial diode—A diode that has the same outer diameter and terminations as a coaxial cable into which the diode is designed to be inserted.

coaxial-fed linear array—A beacon antenna having a uniform azimuth pattern.

coaxial filters—A passive, linear, essentially nondissipative network that transmits certain frequencies and rejects others.

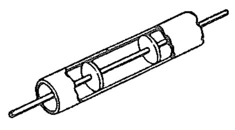

Coaxial line.

coaxial line—Also called coaxial cable, coaxial transmission line, and concentric line. A transmission line in which one conductor completely surrounds the other, the two being coaxial and separated by a continuous solid dielectric or by dielectric spacers. Such a line has no external field and is not susceptible to external fields from other sources.

coaxial-line connector—A connection between two coaxial lines, or between a coaxial line and the equipment.

coaxial-line frequency meter—A shorted section of coaxial line which acts as a resonant circuit and is calibrated in terms of frequency or wavelength.

coaxial relay—A type of relay used for switching high-frequency circuits.

coaxial speaker—A single speaker comprising a high and a low-frequency unit plus an electrical crossover network.

coaxial stub—A short length of coaxial cable joined as a branch to another coaxial cable. Frequently a coaxial stub is short-circuited at the outer end, and its length is so chosen that a high or low impedance is presented to the main coaxial cable at a certain frequency range.

coaxial transistor—A diffused-base, alloy-emitter, epitaxial mesa germanium semiconductor device with a bandwidth product up to 3 gigahertz capable of being operated at medium power.

coaxial transmission line—*See* Coaxial Line.

COBOL—Acronym for COmmon Business Oriented Language. Used to express problems of data manipulation and processing in English narrative form.

co-channel interference—Interference between two signals of the same type from transmitters operating on the same channel.

Cockcroft-Walton accelerator—A device for accelerating charged particles by the action of a high direct-current voltage on a stream of gas ions, in a straight insulated tube. The voltage is generated by a voltage-multiplier system consisting essentially of a number of condenser pairs connected through switching devices (vacuum tubes). The particles (which are nuclei of an ionized gas, such as protons from hydrogen) gain energies of up to several million electron-volts from the single acceleration so produced. Named for the British physicists J. D. Cockcroft and E.T.S. Walton, who developed this machine in the 1930's.

codan—Acronym for Carrier-Operated Device, Anti-Noise. An electronic circuit that keeps a receiver inactive except when a signal is received.

codan lamp—A visual indication that a usable transmitted signal has been received by a particular radio receiver.

code—1. A communications system in which arbitrary groups of symbols represent units of plain text of varying length. Codes may be used for brevity or security. 2. System of signaling by utilizing dot-dash-space, mark-space, or some other method where each letter or figure is represented by prearranged combinations. 3. System of characters and rules for representing information. 4. Loosely, the set of characters resulting from the use of a code.

code character—One of the elements which make up a code and which represent a specified symbol or value to be encoded. Dot-dot-dot-dash is the Morse code character for the letter V.

coded decimal digit—A decimal digit expressed in terms of four or more ones and zeros.

coded passive reflector antenna—An object intended to reflect Hertzian waves and having variable reflecting properties according to a predetermined code for the purpose of producing an indication on a radar receiver.

coded program—A description of a procedure for solving a problem with a digital computer. It may vary in detail from a mere outline of the procedure to an explicit list of instructions coded in the machine's language.

code element—One of the finite set of parts of which the characters in a code may be composed.

code-practice oscillator—An audio oscillator with a key and either headphones or a speaker, used to practice sending and receiving Morse code.

coder—1. A device which sets up a series of signals in code form. 2. A beacon circuit which forms the trigger-pulse output of a discriminator into a series of pulses and then feeds them to a modulator circuit. 3. A person who prepares instruction sequences from detailed flowcharts and other algorithmic procedures prepared by others, as contrasted with a programmer who prepares the procedures and flowcharts.

coding—1. Converting program flowcharts into the language used by the computer. 2. The assignment of identification codes to transactions, such as a customer code number. 3. A method of representing characters within a computer.

codiphase radar—A radar system including a phased-array radar antenna and signal-processing and beam-forming techniques.

codistor—A multijunction semiconductor device which provides noise-rejection and voltage-regulation functions.

coefficient—The ratio of change under specified conditions of temperature, length, volume, etc.

coefficient of coupling—*See* Coupling Coefficient.

coefficient of performance of a thermoelectric cooling couple—The quotient of the net rate of heat removal from the cold junction divided by the electrical power input to the thermoelectric couple, assuming perfect thermal insulation of the thermoelectric arms.

coefficient of performance of a thermoelectric cooling device—The quotient of the rate at which heat is removed from the cooled body divided by the electrical power input to the device.

coefficient of performance of a thermoelectric heating couple—The quotient of the rate at which heat is added to the hot junction divided by the electrical power input to the thermoelectric couple, assuming perfect thermal insulation of the thermoelectric arms.

coefficient of performance of a thermoelectric heating device—The quotient of the rate at which heat is added to the heated body divided by the electrical power input to the device.

coercive force—Symbolized by H. The magnetizing force that must be applied to a magnetic material in a direction opposite the residual induction in order to reduce the induction to zero.

coercivity—The property of a magnetic material measured by the coercive force corresponding to the saturation induction for the material.

cogging—Nonuniform angular velocity. The armature coil of a motor tends to speed up when it enters the magnetic field produced by the field coils, and to slow down when leaving it. This becomes apparent at low speeds, and the fewer the coils, the more noticeable it is.

coherent carrier—A carrier, derived from a c-w signal, the frequency and phase of which have a fixed relationship to the frequency and phase of the reference signal.

coherent detection—A method of deriving additional information from the phase of the carrier.

coherent electroluminescence device—*See* Diode Laser.

coherent interrupted waves—Interrupted continuous waves occurring in wave trains in which the phase of the waves is maintained through successive wave trains.

coherent light—A single frequency of light. Light having characteristics similar to a radiated radio wave that has a single frequency.

coherent light detection and ranging—*See* Colidar.

coherent oscillator—An oscillator within some radar sets which furnishes phase references for target returns during intervals between transmitter pulse, and its output compared with the returns so that the echo becomes coherent video. Coherent video is applied to a cancellation circuit which eliminates nonmoving targets, and only moving targets are supplied to the indicator.

coherent-pulse operation — The method of pulse operation in which a fixed phase relationship is maintained from one pulse to the next.

coherent radiation—A form of radiation in which definite phase relationships exist between radiation at different positions in a cross section of the radiant beam.

coherent reference—The reference signal, usually of stable frequency, to which other signals are phase-locked to establish coherency throughout a system.

coherent system of units—Also called absolute system of units. A system of units in which the magnitude and dimensions of each unit are related to those of the other units by definite simple relationships in which the proportionality factors are usually chosen to be unity.

coherent transponder—A transponder, the output signal of which is coherent with the input signal. Fixed relations between frequency and phase of input and output signals are maintained.

coherer—An early form of detector used in wireless telegraphy.

coil—A number of turns of wire wound around an iron core or onto a form made of insulating material, or one which is self-supporting. A coil offers considerable opposition to the passage of alternating current, but very little opposition to direct current.

coil dissipation—The amount of electrical power consumed by a winding. For the most practical purposes, this equals the I^2R loss.

coil form—An insulating support of ceramic, plastic, or cardboard onto which coils are wound.

coil loading—As commonly understood, the insertion of coils into a line at uniformly spaced intervals. However, the coils are sometimes inserted in parallel.

coil neutralization — *See* Inductive Neutralization.

coil rating—*See* Input-Power Rating.

coil resistance—The total terminal-to-terminal resistance of a coil at a specified temperature.

coil serving—A covering, such as thread or tape, that serves to protect a coil winding from mechanical damage.

coil temperature rise — The increase in winding temperature above the ambient temperature when energized under specified conditions for a given period of time, usually that required to reach a stable temperature.

coil tube — A tubular coil form. (*Also see* Spool.)

coincidence amplifier—An amplifier which produces no output unless two input pulses are applied simultaneously to the circuit.

coincidence circuit—A circuit that produces a specified output pulse when and only when a specified number (two or more) or combination of input terminals receives pulses within a specified time interval.

coincidence gate—A circuit with the ability to produce an output which is dependent upon a specified type or the coincident nature of the output.

coincident-current selection—The selection of a magnetic core, for reading or writing, by the simultaneous application of two or more currents.

cold cathode—A cathode the operation of which does not depend on its temperature being above the ambient temperature.

cold-cathode tube—An electron tube containing a cathode which will emit electrons without being heated.

cold emission—*See* Field Emission.

cold flow—Change of dimension or distortion, caused by sustained application of a force.

cold-pressure welding—A method of making an electrical connection in which the members to be joined are compressed to the plastic range of the metals.

cold rolling—Rolling a magnetic core alloy into the form of a rod so that the metallic grains are oriented in the long direction of the rod.

cold weld—A joint between two metals (without an intermediate material) produced by the application of pressure only.

colidar — Abbreviation for COherent LIght Detection And Ranging. An optical radar system that uses the direct output from a ruby laser source without further pulse modulation.

collate—To combine two or more similarly ordered sets of items to produce another ordered set. Both the number and the size of the individual items may differ from those of the original sets or their sums.

collator—A device to collate sets of punched cards or other documents into a sequence.

collating sequence—1. In digital computers, the sequence in which the characters acceptable to a computer are ordered. The British term is marshalling sequence. 2. An ordering assigned to a set of items, such that any two sets in that assigned order can be collated.

collector—1. An electrode that collects electrons or ions which have completed their functions within an electron tube. 2. The collector of a transistor is an electrode through which a primary flow of carriers leaves the interelectrode region.

collector capacitance — Depletion-layer capacitance associated with the collector junction of a transistor.

collector-coupled logic—*See* Current-Sourcing Logic.

collector current—The direct current flowing in the collector of a transistor.

collector cutoff—The operating condition of a transistor when the collector current is reduced to the leakage current of the collector-base junction.

collector cutoff current—The minimum current that will flow in the collector circuit of a transistor with zero current in the emitter circuit.

collector efficiency—The ratio, usually expressed in percentage, of useful power output to final-stage power-supply power input of a transistor.

collector family—Set of transistor characteristic curves in which the collector current and voltage are variables.

collector junction—Of a semiconductor device, a junction normally biased in the high-resistance direction, so that the current through the junction can be controlled by the introduction of minority carriers.

collector resistance — Resistance in series with the collector lead in the common-T equivalent circuit of a transistor.

collector rings (slip rings)—Metal rings suitably mounted on an electric machine serving, through stationary brushes bearing thereon, to conduct current into or out of the rotating member.

collector transition capacitance — The capacitance across the collector-to-base transition region of a transistor.

collector voltage—The d-c collector supply voltage applied between the base and collector of a transistor.

collimation—The process of adjusting an instrument so that its reference axis is aligned in a desired direction within a predetermined tolerance.

collimation equipment—Equipment designed specifically for aligning optical equipment.

collimator—An optical device that creates a beam made up of parallel rays of light, used in testing and adjusting certain optical instruments.

collinear array—An antenna array in which half-wave elements are arranged end to end on the same vertical or horizontal line.

color balance—The adjustment of electron-gun emissions to compensate for the difference in the light emitting efficiencies of the three phosphors on the screen of the color picture tube.

color-bar pattern—A test pattern (usually a series of vertical bars, each having a specific hue), used for checking the performance of a color-television receiver.

color breakup—Any fleeting or partial separation of a color picture into its display primary components because of a rapid change in the condition of viewing. For example, fast movement of the head, abrupt interruption of the line of sight, and blinking of the eyes are illustrations of rapid changes in the conditions of viewing.

color burst—Also called reference burst. Approximately nine cycles of the chrominance subcarrier added to the back porch of the horizontal blanking pedestal of the composite color signal and used in the color receiver as a phase reference for the 3.579545-MHz oscillator.

color carrier—*See* Chrominance Subcarrier.

colorcast—A color television broadcast.

color code—A system of colors for specifying the electrical value of a component part or for identifying terminals and leads. Also used to distinguish between cable conductors.

color coder—Also called color encoder. In a color-tv transmitter, that circuit or section which combines the camera signals and the chrominance subcarrier to form the transmitted color picture signal.

color contamination—An error in color rendition due to incomplete separation of the paths carrying different color components of the picture. Such errors can arise in the optical, electronic, or mechanical portions of a color-television system as well as in the electrical portions.

color-coordinate transformation — Computation of the tristimulus values of colors in terms of one set of primaries from the tristimulus values of the same colors in another set of primaries. Such computation may be performed electrically in a color-television system.

color decoder—A section or circuit of a color-television receiver used for deriving the signals for the color-display device from the color-picture signal and the color burst.

color-difference signal—The signal produced when the amplitude of a color signal is reduced by an amount equal to the amplitude of the luminance signal. Color-difference signals are usually designated R-Y, B-Y, and G-Y. In a sense, I and Q signals are also color-difference signals because they are formed when specific proportions of R-Y and B-Y color-difference signals have been combined.

color edging—Spurious color at the boundaries of differently colored areas in a picture.

color encoder—*See* Color Coder.

color fidelity—The degree to which a color television system is capable of faithfully reproducing the colors in an original scene.

color flicker—The flicker which results from fluctuation of both the chromaticity and the luminance.

color fringing—Spurious chromaticity at the boundaries of objects in a color-tv picture. It can be caused by the change in relative position of the televised object from field to field, or by misregistration. Color fringing may cause small objects to appear separated into different colors.

colorimeter—An optical instrument designed to compare the color of a sample with that of a standard sample or a synthesized stimulus. (In a three-color colorimeter, the synthesized stimulus consists of three colors of constant chromaticity but variable luminance.)

colorimetric—The measurement of color characteristics, particularly wavelength and primary-color content.

colorimetry—The technique of measuring color and interpreting the results.

color killer—A stage designed to prevent signals in a color receiver from passing through the chrominance channel during monochrome telecasts.

color match—The condition in which the two halves of a structureless photometric field look exactly the same.

color mixture—Color produced by the combination of lights of different colors. The combination may be accomplished by successive presentation of the components, provided the rate of alternation is sufficiently high; or the combination may be accomplished by simultaneous presentation, either in the same or in adjacent areas, provided they are small enough and close enough together to eliminate pattern effects.

color oscillator—In a color-television receiver, the oscillator operating at the burst frequency of 3.579545 MHz. Its frequency and phase are synchronized by the master oscillator at the transmitter.

color phase—The difference in phase between a chrominance primary signal (I or Q) and the chrominance carrier reference.

color-phase alternation—Periodic changing of the color phase of one or more components of the color-television subcarrier between two sets of assigned values.

color-phase diagram—A vector diagram which denotes the phase difference between the color-burst signal and the chrominance signal for each of the three primary and complementary colors. This diagram also designates vectorially the peak amplitude of the chrominance signal for each of these colors, and the polarities and peak amplitudes of the in-phase and quadrature portions required to form these chrominance signals.

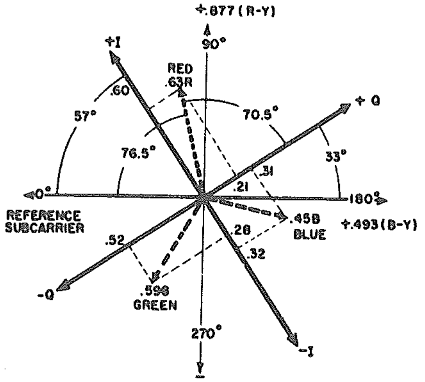

Color-phase diagram.

color-picture signal—1. A signal which represents electrically the three color attributes (brightness, hue, and saturation) of a scene. 2. A combination of the luminance and chrominance signals, excluding all blanking and synchronizing signals.

color-picture tube—An electron tube that provides an image in color by scanning a raster and by varying the intensity at which it excites the phosphors on the screen to produce light of the chosen primary colors.

color primaries—In the color receiver, the saturated colors of definite hue and variable luminance produced by the receiver. These color primaries, when mixed in proper proportions, form other colors.

color purity—Freedom of a color from white light or any colored light not used to produce the desired color.

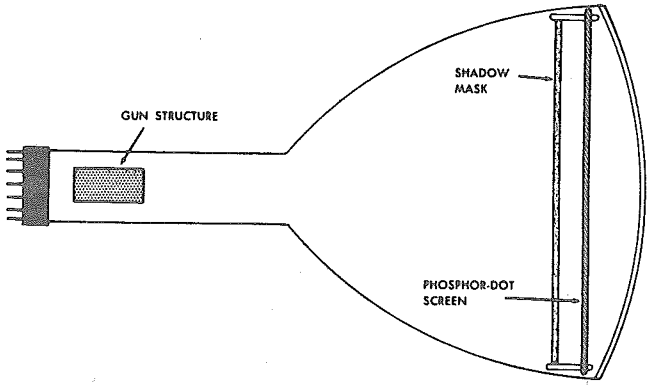

Color picture tube.

color-purity magnet—A magnet placed in the neck region of a color-picture tube to alter the electron-beam path and thereby improve color purity.

color registration—The accurate superimposing of the red, green, and blue images that are used to form a complete color picture in a color television receiver.

color sampling rate—In a color television system, the number of times per second that each primary color is sampled.

color saturation—The degree to which white light is absent in a particular color. A fully saturated color contains no white light. If 50% of the light intensity is due to the presence of white light, the color is said to have a saturation of 50%.

color sensitivity—The spectral sensitivity of a light-sensitive device such as a phototube or camera tube.

color signal—Any signal at any point in a color-television system, used for wholly or partially controlling the chromaticity values of a color-television picture. This is a general term encompassing many specific connotations, such as are conveyed by the words, "color-picture signal," "chrominance signal," "carrier-color signal," "monochrome signal" (in color television), etc.

color subcarrier—A monochrome signal to which modulation sidebands have been added to convey color information.

color-sync signal—The series of color bursts (pulses of subcarrier-reference signal) applied to the back porch of the horizontal-sync pedestal in the composite video signal.

color-television receiver—A standard monochrome receiver to which special circuits have been added. Phosphors capable of glowing in the three primary colors are used on the special screen. By using these primary colors and mixing them to produce complementary colors, and by varying their intensity, it is possible to reproduce an image in somewhat the original colors.

color-television signal—The complete signal used to transmit a color picture. Included are horizontal-, vertical-, and color-sync components.

color temperature—The temperature to which a perfectly black body must be heated to match the color of the source being measured. Color-temperature measurements begin at absolute zero and are expressed in degrees Kelvin.

color transmission—The transmission of color-television signals which can be reproduced with different values of hue, saturation, and luminance.

color triad—One cell of a three-color, phosphor-dot screen of a phosphor-dot, color-picture tube. Each triad contains one dot of each of the three color-producing phosphors.

color triangle—A triangle drawn on a chromaticity diagram to represent the entire range of chromaticities obtainable when the 3 prescribed primaries are added. These are represented by the corners of the triangle.

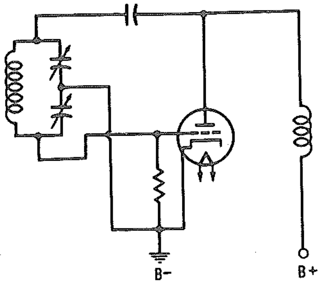

Colpitts oscillator.

Colpitts oscillator—An electron-tube oscillator comprising a parallel-tuned tank circuit connected between grid and plate. The tank capacitance contains two voltage-dividing capacitors in series, with their common connection at cathode potential. When these two capacitances are the plate-to-cathode and grid-to-cathode capacitances of the tube, the circuit is known as an Ultra-audion oscillator.

column—Also called place. In positional notation, a position corresponding to a given power of the radix. A digit located in any particular column is a coefficient of a corresponding power of the radix.

coma—1. An aberration of spherical lenses, occuring in the case of oblique incidence, when the bundle of rays forming the image is unsymmetrical. The image of a point is comet shaped, hence the name. 2. A cathode-ray tube image defect that makes the spot on the screen appear comet shaped when away from the center of the screen.

comb filter—A type of filter network that is, in effect, a multiple-bandpass design that passes only frequencies within a number of narrow bands, or provides outputs corresponding to each of its passbands. Comb filters are so named because their response characteristics have the appearance of a comb.

combination cable—A cable in which the conductors are grouped in combinations such as pairs and quads.

combination microphone—A microphone consisting of two or more similar or dissimilar microphones combined into one.

combination tones—Frequencies produced in a nonlinear device, such as in an audio amplifier having appreciable harmonic distortion.

combiner — A circuit for mixing video, trigger, and scan data from the synchronizer for the modulation of a link.

combiner circuit—One that combines the luminance and chroma channels with the sync signals in color-television cameras.

combustible—*See* Flammable.

command—1. In a computer, one or more sets of signals which occur as the result of an instruction. (*Also see* Instruction.) 2. An independent signal from which the dependent signals in a feedback-control system are controlled according to the prescribed system relationships. 3. A signal which initiates or triggers an action in the device that receives the signal.

command control—A system whereby functions are performed as the result of a transmitted signal.

command destruct signal—A radio signal for destroying a missile in flight.

command guidance system—A missile guidance system in which both the missile and the target are tracked by radar. The missile is guided by signals transmitted to it while it is in flight. (*Also see* Command Link.)

command link—The portion of a command guidance system used to transmit steering commands to the missile. (*Also see* Command Guidance System.)

command module—*See* Module.

command net— A communication network which connects an echelon of command with some or all of its subordinate echelons for the purpose of command and control.

command reference—In a servo or control system, the voltage or current to which the feedback signal is compared. As an independent variable, the command reference exercises complete control over the system output.

comment—An expression which explains or identifies a particular step in a routine, but which has no effect on the operation of the computer in performing the instructions for the routine.

commercial test (measuring) equipment—Devices used as "working instruments" selected from suppliers' catalogs as suitable for the user's measuring needs and procured as standard ("off-the-shelf") items. This includes measuring devices installed, intact and in consoles. Devices procured from suppliers but modified by user-imposed specifications, in such a manner as to affect console performance, do not fall into this category.

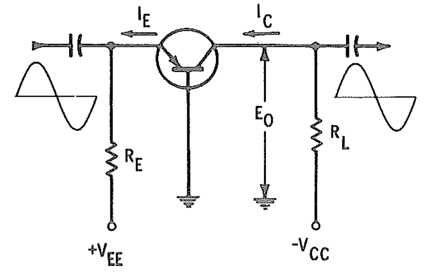

Common-base amplifier.

common-base amplifier — Also called a grounded-base amplifier. A transistor amplifier in which the base element is common to both the input and the output circuit. It is comparable to the grounded-grid configuration of a triode electron-tube amplifier.

common battery—A system of current supply in which all direct-current energy for a unit of a telephone system is supplied by one source in a central office or exchange.

common-battery office — A central office which supplies transmitter and signal current for its associated stations, and for signaling by the central office equipment from a power source located in the central office.

common business oriented language—Specific language by which business data-processing procedures may be precisely described in a standard form. Intended not only as a means for directly presenting any business program to any suitable computer for which a compiler exists, but also as a means of communicating such procedures among individuals.

common-carrier fixed station—A fixed station that is open to public correspondence.

common-collector amplifier—Also known as an emitter-follower and a grounded-collector amplifier. A transistor amplifier in which the collector element is common to both the input and the output circuit. This configuration is comparable to an electron-tube cathode follower.

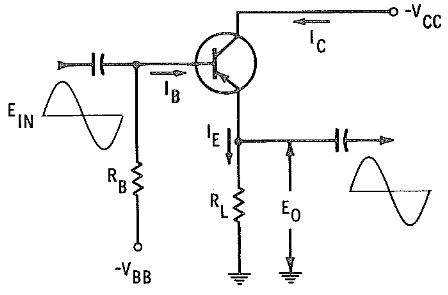

Common-collector amplifier.

common communications carrier— A company recognized by an appropriate regulatory agency as having a vested interest in furnishing communications services.

common-emitter amplifier — Also called grounded-emitter amplifier. A transistor amplifier in which the emitter element is common to both the input and the output circuit. This configuration is comparable to a conventional electron-tube amplifier.

common language—A form of representing information which a machine can read and which is common to a group of computers and data-processing machines.

common mode—Signals that are identical with respect to both amplitude and time. Also used to identify the respective parts of two signals that are identical with respect to amplitude and time.

common-mode characteristics—The characteristics pertaining to performance of an operational amplifier where the inverting and noninverting inputs have a common signal.

common-mode gain—The ratio of the common-mode output voltage of a differential amplifier divided by the common-mode input voltage.

common-mode impedance input—The internal impedance between either one of the input terminals of a differential operational amplifier and signal ground.

common-mode input—That signal applied in phase (i.e., common mode) equally to both inputs of a differential amplifier.

common-mode input capacitance — The equivalent capacitance of both inverting and noninverting inputs of an operational amplifier with respect to ground.

common-mode input impedance — The open-loop input impedance of both inverting and noninverting inputs of an operational amplifier with respect to ground.

common-mode input voltage — The maximum voltage that can be applied simultaneously between the two inputs of a differential amplifier and ground without causing damage.

common-mode rejection — Also called in-phase rejection. A measure of how well a differential amplifier ignores a signal which appears simultaneously and in phase at both input terminals (called a common-mode signal). Usually and preferably stated as a voltage ratio, but more often stated in the db equivalent of said ratio at a specified frequency—e.g., "120 db at 60 hertz with 1K source impedance."

common-mode rejection ratio—The ratio of the common-mode input voltage to the output voltage expressed in db. The extent to which a differential amplifier does not provide an output voltage when the same signal is applied to both inputs.

common-mode resistance — The resistance between the input- and output-signal lines or circuit ground. In an isolated amplifier, this is its insulation resistance. (Common-mode resistance has no connection with common-mode rejection.)

common-mode signal — The instantaneous algebraic average of two signals applied to a balanced circuit (i.e., two ungrounded inputs of a balanced amplifier), all signals referred to a common reference.

common-mode voltage—1. The amount of voltage common to both input lines of a balanced amplifier. Usually specified as the maximum voltage which can be applied without breaking down the insulation between the input circuit and ground. (Common-mode voltage has no connection with common-mode rejection.) 2. The voltage on both inputs of a differential input amplifier.

common-mode voltage gain—The ratio of the signal voltage applied at the inputs (in parallel) to the signal voltage between the outputs.

common-mode voltage range—The range of voltage that may be applied to both inputs of an operational amplifier without saturating the input stage. This may limit the output capabilities in the voltage-follower connection.

common-user channels — Communication channels which are available to all authorized agencies for transmission of command, administrative, and logistic traffic.

communication—The transmission of information from one point, person, or equipment to another.

communication band—The band of frequencies due to the modulation (including keying) necessary for a given type of transmission.

communication channel—Part of a radio or wire circuit, or a combination of wire and radio which connects two or more terminals.

communication link—The physical means of connecting one location to another for the purpose of transmitting and receiving information.

communications common carriers — Companies that furnish communications services to the public, regulated by the FCC or appropriate state agencies.

communications receiver — A receiver designed for reception of voice or code signals from stations operating in the communications service.

communication switch — A device used to execute repetitive sequential switching.

communication zone indicator—A device that indicates whether or not long-distance high-frequency broadcasts are successfully reaching their destination.

community antenna television—A television system that receives and retransmits commercial television broadcasts. Microwave transmitters and coaxial cables are used to bring the television signals to subscribers in a community. Abbreviated catv.

commutation—1. A mechanical process of converting the alternating current in the armature of direct-current generators into the direct-current generator output. 2. Sampling of various quantities in a repetitive manner, for transmission over a single channel in telemetering. 3. Repetitive sampling of two or more information channels for transmission over a single link.

commutation capacitors — Cross-connected capacitors in a thyratron inverter. They provide a path such that the start of conduction in one thyratron causes an extinguishing pulse to be applied to the alternate thyratron. Also used in inverter circuits employing semiconductor devices.

commutator—The part of the armature to which the coils of a motor are connected. It consists of wedge-shaped copper segments arranged around a steel hub and insulated from it and from one another. The motor brushes ride on the outer edges of the commutator bars and thereby connect the armature coils to the power source.

compactron—An electron tube based on a building-block concept which involves the standardizing of basic tube sections, diodes, triodes, pentodes; clipping them together as required; and sealing them in a single envelope.

compander—A combination comprised of a compressor at one point in a communication path to reduce the volume range of signals, followed by an expander at another point to improve the ratio of the signal to the interference entering the path between the compressor and expander.

companding—A process in which compression is followed by expansion. Companding is often used for noise reduction, in which case the compression is applied before the noise exposure and the expansion afterward.

compandor—*See* Compander.

companion keyboard — A remote keyboard connected by a multiwire cable to an ordinary keyboard and able to operate it.

comparator—A circuit which compares two signals and supplies an indication of agreement or disagreement.

compare—A computer operation in which two quantities are matched for the purpose of discovering their relative magnitudes or algebraic values.

comparison bridge—A type of voltage-comparison circuit resembling a four-arm electrical bridge. The elements are so arranged that if a balance exists in the circuit, a zero error signal is derived.

compatible IC—A hybrid IC in which the active circuit element is within the silicon planar integrated structure. A passive network, which may be separately optimized, is deposited onto its insulating surface to complete the IC device.

compatible monolithic integrated circuit—A device in which passive components are deposited by thin-film techniques on top of a basic silicon-substrate circuit containing the active components and some passive parts. *Also see* All-Diffused Monolithic Integrated Circuit.

compatibility—1. That property of a color-television system which permits typical, unaltered monochrome receivers to receive substantially normal monochrome from the transmitted signal. 2. In stereo, a system usable with a monophonic program source, or a stereo program that can be reproduced monophonically on a monophonic system.

compensated amplifier — A broad-band amplifier the frequency range of which is extended by the proper choice of circuit constants.

compensated-impurity resistor—A diffused-layer resistor into which are introduced additional N- and P-type impurities.

compensated-loop direction finder—A direction finder employing a loop antenna and a second antenna system to compensate for polarization error.

compensated semiconductor—A semiconductor in which one type of impurity or imperfection (donor) partially cancels the electrical effects of the other (acceptor).

compensated volume control—*See* Loudness Control.

compensating filter—A filter used to alter the spectral emission of an emulsion to a specified response to different wavelengths.

compensation — The controlling elements which compensate for or offset the undesirable characteristics of the process to be controlled in a system.

compensation signal—A signal recorded on the tape, along with the computer data and on the same track as the data; this signal is used during the playback of data to electrically correct for the effects of tape-speed errors.

compensation theorem—An impedance in a network may be replaced by a generator of zero internal impedance, the generated voltage of which at any instant is equal to the instantaneous potential difference produced across the replaced impedance by the current flowing through it.

compensator—1. In a direction finder, the portion which automatically applies to the direction indication all or part of the correction for the deviation. 2. An electronic circuit for altering the frequency response of an amplifier system to achieve a specified result. This refers to record equalization or loudness correction.

compile — To bring digital-computer programming subroutines together into a main routine or program.

compiler—An automatic coding system in a computer which generates and assembles a program from instructions written by a programmer.

compiling routine — A routine by means of which a computer can itself construct the program used to solve a problem.

complement—1. In an electronic computer, a number the representation of which is derived from the finite positional notation of another by one of the following rules:

True complement—Subtract each digit from 1 less than the base; then add 1 to the least significant digit and execute all required carries.

Base minus 1's complement—Subtract each digit from 1 less than the base (e.g., "9's complement" in the base 10, "1's complement" in the base 2, etc.)

2. To form the complement of a number. (In many machines, a negative number is represented as a complement of the corresponding positive number.)

complementary colors—Two colors are complementary if, when added together in proper proportion such as by projection, they produce white light.

complementary rectifier—Half-wave rectifying circuit elements which are not self-saturating rectifiers in the output of a magnetic amplifier.

complementary-symmetry circuit — An arrangement of PNP and NPN transistors that provides push-pull operation from one input signal.

complementary tracking—A system of interconnection of two or more devices in which one (the master) operates to control the others (the slaves).

complementary transistor amplifier — An amplifier that utilizes the complementary symmetry of NPN and PNP transistors.

complementary transistor logic—A digital logic circuit configuration making use of a complementary transistor emitter coupled AND-OR gate. Basically a two-level diode gate using simultaneous NPN and PNP action.

complementary transistors — Two transistors of opposite conductivity (PNP and NPN) that are operated in the same functional unit.

complementary wave—A wave brought into existence at the ends of a coaxial cable or two-conductor transmission lines, or any discontinuity along the line.

complementary wavelength — The wavelength of light of a single frequency. When combined with a sample color in suitable proportions, the wavelength matches the reference-standard light.

complementor—A circuit or device that produces a Boolean complement. A NOT circuit.

complete carry—A system of executing the carry process in a computer. All carries, and any other carries to which they give rise, are allowed to propagate to completion in this system.

complex components—Indivisible and nonrepairable components having more than one function.

complex function—An integrated device in which three or more circuits are integral to a single silicon chip. In addition, the circuits are interconnected on the chip itself to form some electronic function at a higher level of organization than a single circuit. The interconnection pattern for the function is predetermined by the fixed mask; no wiring discretion is available for yield purposes. The input and outputs of all the circuits are not normally exposed to the package terminals. An example of a complex function would be a full adder or a multibit serial shift register.

complex parallel permeability — The complex relative permeability measured under stated conditions on a core with the aid of a coil. The parameter characterizing the induction is the impedance of the coil when placed on the core, expressed as a parallel connection of reactance and resistance. The parameter characterizing the field strength is the reactance the coil would have if placed on a core of the same dimensions but with unity relative permeability, the distribution of the magnetic field being identical in both cases. The coil should have negligible copper losses.

complex permeability—Under stated conditions, the complex quotient of the moduli of the parallel vectors representing induction and field strength in a material. One of the moduli varies sinusoidally with time, and the component chosen from the other modulus varies sinusoidally at the same frequency.

complex series permeability—The complex relative permeability measured under stated conditions on a core with the aid of a coil. The parameter characterizing the induction in the core is the impedance of the coil when placed on the core, expressed as a series connection of reactance and resistance. The parameter characterizing the field strength is the reactance this coil would have if placed on a core of the same dimensions but with unity relative permeability. The coil should have negligible copper losses, etc.

complex steady-state vibration—A periodic vibration of more than one sinusoid. It includes repeating square waves, sawtooth waves, etc., because these waveforms can be expressed in terms of a Fourier series of sinusoidal terms.

complex tone—A sound wave produced by the combination of simple sinusoidal components of different frequencies. A sound sensation characterized by more than one pitch.

complex wave—A periodic wave made up of a combination of several frequencies or several sine waves superimposed on one another.

complex-wave generator — A device which generates a nonsinusoidal signal having a desired repetitive characteristic and waveform.

compliance—1. The reciprocal of stiffness—i.e., the ability to yield or flex. 2. The ease with which a phonograph stylus responds to an outside force. 3. The mechanical and acoustical equivalent of capacitance.

compliance extension—A form of master/slave interconnection of two or more current-regulated supplies to increase their output-voltage range through series connection.

compliance range — The range of voltage needed to sustain a given constant current throughout a range of load resistances.

compliance voltage—The output voltage of a constant-current supply.

compliance voltage range—The output voltage range of a d-c power supply operating in a constant-current mode.

component — 1. An essential functional part of a subsystem or equipment. It may be any self-contained element with a specific function, or it may consist of a combination of parts, assemblies, accessories, and attachments. 2. In vector analysis, one of the parts of a wave, voltage, or current considered separately.

component density—The volume of a circuit assembly divided by the total number of discrete circuit components utilized, usually expressed in components per cubic foot.

component layout—The physical arrangement of the components in a chassis or printed circuit.

component operating hours — A unit of measurement for the period of successful operation of one or more components (of a specified type) which have endured a given set of environmental conditions.

component stress—Those factors of usage or test, such as voltage, power, temperature, frequency, etc., which tend to affect the failure rate of component parts.

composite cable—A cable in which conductors of different gages or types are combined under one sheath.

composite circuit—A circuit which can be used simultaneously for telephony and direct-current telegraphy or signaling, the two being separated by frequency discrimination.

composite color signal—The color-picture signal plus all blanking and synchronizing signals. Includes luminance and chrominance signals, vertical- and horizontal-sync pulses, vertical- and horizontal-blanking pulses, and the color-burst signal.

composite color sync—The signal comprising all the sync signals necessary for proper operation of a color receiver. Includes the deflection sync signals to which the color sync signal is added in the proper time relationship.

composite conductor—Two or more strands of different metals, such as aluminum and steel or copper and steel, assembled and operated in parallel.

composite controlling voltage—The voltage of the anode of an equivalent diode, combining the effects of all individual electrode voltages in establishing the space-charge limited current.

composited circuit—A circuit which can be used simultaneously for telephony and for direct-current telegraphy or signaling, separation between the two being accomplished by frequency discrimination.

composite filter—A filter with two or more sections.

composite guidance system—A guidance system using a combination of more than one individual guidance system.

composite picture signal—The television signal produced by combining a blanked picture signal with the sync signal.

composite video signal—The complete video signal. For monochrome, it consists of the picture signal and the blanking and synchronizing signals. For color, color-synchronizing signals and color-picture information are added.

composite wave filter—A combination filter consisting of two or more low-pass, high-pass, bandpass, or band-elimination filters.

composition resistor—*See* Carbon Resistor.

compound-connected transistor—Two transistors which are combined to increase the current amplification factor at high emitter currents. This combination is generally employed in power-amplifier circuits.

compound-filled transformer — A transformer which is contained within a case and in which the structural insulating material is supplemented by submergence in a solid or semisolid insulating material introduced into the case in a fluid state.

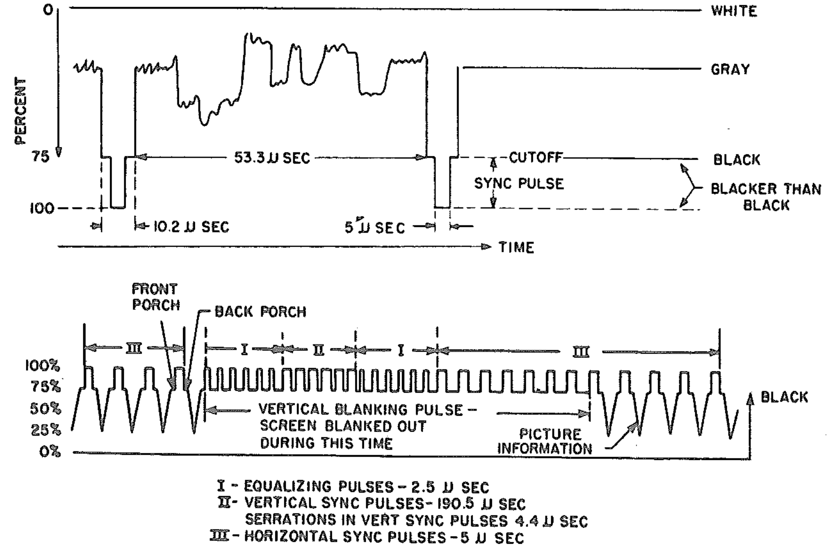
0
PERCENT
75
100
WHITE
GRAY
BLACK
BLACKER THAN BLACK
CUTOFF
SYNC PULSE
53.3 μ SEC
10.2 μ SEC
5 μ SEC
TIME
FRONT PORCH
BACK PORCH
III
I
II
I
III
100%
75%
50%
25%
0%
BLACK
VERTICAL BLANKING PULSE - SCREEN BLANKED OUT DURING THIS TIME
PICTURE INFORMATION
I - EQUALIZING PULSES - 2.5 μ SEC
II - VERTICAL SYNC PULSES - 190.5 μ SEC
SERRATIONS IN VERT SYNC PULSES 4.4 μ SEC
III - HORIZONTAL SYNC PULSES - 5 μ SEC

Composite picture signal.

compound horn—An electromagnetic horn of rectangular cross section. The four sides of the horn diverge in such a way that they coincide with or approach four planes, with the provision that the two opposite planes do not intersect the remaining planes.

compound modulation—*See* Multiple Modulation.

compound-wound motor—A d-c motor having two separate field windings. One, usually the predominant field, is connected in parallel with the armature circuit and the other is connected in series.

compressed-air loudspeaker—A loudspeaker that has an electrically actuated valve to modulate a stream of compressed air.

compression—A process in which the effective amplification of a signal is varied as a function of the signal magnitude, the effective gain being greater for small than for large signals. In television, the reduction in gain at one level of a picture signal with respect to the gain at another level of the same signal.

compressional wave—In an elastic medium, a wave which causes a change in volume of an element of the medium without rotation of the element.

compression driver unit—A speaker driver unit that does not radiate directly from the vibrating surface. Instead, it requires acoustic loading from a horn which connects through a small throat to an air space adjacent to the diaphragm.

compression ratio—The ratio between the magnitude of the gain (or amplification) at a reference signal level and its magnitude at a higher stated signal level.

compressor—A transducer which, for a given amplitude range of input voltages, produces a smaller range of output voltages. In one important type of compressor, the envelope of speech signals is used to reduce their volume range.

compromise network—In a telephone system, a network used in conjunction with a hybrid coil to balance a subscriber's loop. The network which is adjusted for an average loop length, an average subscriber's set, or both, and gives compromise (not precision) isolation between the two directional paths of the hybrid coil.

Compton effect — The elastic scattering of photons by electrons. Because the total energy and momentum are conserved in the collisions, the wavelength of the scattered radiation undergoes a change that depends on the scattering angle.

computer — Any device, usually electronic, capable of accepting information, comparing, adding, subtracting, multiplying, dividing, and integrating this information, and then supplying the results of these processes in acceptable form.

computer code — Also called machine language. The code by which data are represented within a computer system. An example is a binary-coded decimal.

computing—To perform basic and more involved mathematical processes of comparing, adding, subtracting, multiplying, dividing, integrating, etc.

computing machine — An automatic device which carries out well defined mathematical operations.

concatenate—To link together or unite in a series.

concatenation — The joining or linking of sets or series.

concave—Curved inward.

concentrated-arc lamp—A type of low-voltage arc lamp having nonvaporizing electrodes sealed in an atmosphere of inert gas and producing a small, brilliant, incandescent cathode spot.

concentration gradient—A difference in carrier concentration (holes or free electrons) from point to point in a semiconductor.

concentric cable—*See* Coaxial Line.

concentric groove—*See* Locked Groove.

concentric-lay conductor—A conductor composed of a central core surrounded by one or more layers of helically laid wires. In the most common type of concentric-lay conductor, all wires are of the same size, and the central core is a single wire.

concentric line—*See* Coaxial Line.

concentric stranding—A method of stranding wire in which the final wire is built up in layers such that the inner diameter of a succeeding layer always equals the outer diameter of the underlying layer.

concentric-wound coil—A coil with two or more insulated windings which are wound one over the other.

concurrent processing — The ability of a computer to work on more than one program at the same time.

condensed mercury temperature—The temperature of a mercury-vapor tube, measured on the outside of the tube envelope, in the region where the mercury is condensing in a glass tube or at a designated point on a metal tube.

condenser—*See* Capacitor.

condenser antenna—*See* Capacitor Antenna.

condenser microphone — *See* Electrostatic Microphone.

condenser speaker—*See* Electrostatic Speaker.

conditional—1. In a computer, subject to the result of a comparison made during computation. 2. Subject to human intervention.

conditional jump—Also called conditional transfer of control. An instruction to a computer which will cause the proper one of two (or more) addresses to be used in obtaining the next instruction, depending on some property of one or more numerical expressions or other conditions.

conditional transfer of control—*See* Conditional Jump.

Condor—A c-w navigational system, similar to benito, which automatically measures bearing and distance from a single ground station and displays them on a cathode-ray indicator. The distance is determined by phase comparison, and the bearing, by automatic direction finding.

conductance—Symbolized by G or g. 1. In an element, device, branch, network, or system, the physical property that is the factor by which the square of an instantaneous voltage must be multiplied to give the corresponding energy lost by dissipation as heat or other permanent radiation, or by loss of electromagnetic energy from the circuit. 2. The real part of admittance.

conducted heat—Thermal energy transferred by thermal conduction.

conducted interference—Any unwanted electrical signal conducted on the power lines supplying the equipment under test, or on lines supplying other equipment to which the one under test is connected.

conduction—The transmission of heat or electricity through or by means of a conductor.

conduction band—In a semiconductor, a partially filled energy band in which electrons can move freely, allowing the material to carry an electric current.

conduction current—The power flow parallel to the direction of propagation, expressed in mhos/meter.

conduction-current modulation—1. Periodic variations in the conduction current passing a point in a microwave tube. 2. The process of producing such variations.

conduction electrons—The electrons which are free to move under the influence of an electric field in the conduction band of a solid.

conduction error—The error in a temperature transducer due to heat conduction between the sensing element and the mounting to the transducer.

conduction field—Energy which surrounds a conductor when an electric current is passed through the conductor, and which, because of the difference in phase between the electrical field and magnetic field set up in the conductor, cannot be detached from the conductor.

conductive gasket—A special highly resilient gasket used to reduce r-f leakage in shielding which has one or more access openings.

conductive material—A material in which a relatively large conduction current flows when a potential is applied between any two points on or in a body constructed from the material. Metals and strong electrolytes are examples of conductors.

conductive pattern—The arrangement or design of the conductive lines on a printed-circuit board.

conductivity—1. The conductance between opposite faces of a unit cube of material. The volume conductivity of a material is the reciprocal of the volume resistivity. 2. The ability of a material to conduct current. The reciprocal of resistivity.

conductivity modulation — The change in conductivity of a semiconductor as the charge-carrier density is varied.

conductivity - modulation transistor — A transistor in which the active properties are derived from minority-carrier modulation of the bulk resistivity of the semiconductor.

conductor—1. A bare or insulated wire or combination of wires not insulated from one another, suitable for carrying an electric current. 2. A body of conductive material so constructed that it will serve as a carrier of electric current.

conduit — A tubular raceway for holding wires or cables designed and used expressly for this purpose.

conduit wiring—Wiring carried in conduits and conduit fittings.

cone—The cone-shaped paper or fiber diaphragm of a speaker.

cone of nulls—A conical surface formed by directions of negligible radiation.

cone of silence — An inverted cone-shaped space directly over the aerial towers of some radio beacons. Within the cone, signals cannot be heard or will be greatly reduced in volume.

confetti—Flecks or streaks of color caused by tube noise in the chrominance amplifier. Because of its colors, confetti is much more noticeable than snow in a black-and-white picture. The chrominance amplifier is therefore cut off during a monochrome program.

confidence—1. The likelihood, expressed in per cent, that a statement is true. 2. The degree of assurance that the stated failure rate has not been exceeded.

confidence factor—The percentage figure expressing confidence level.

confidence interval—A range of values believed to include, with a preassigned degree of confidence, the true characteristic of the lot.

confidence level—The probability (expressed as a percentage), that a given assertion is true or that it lies within certain limits calculated from the date. 2. A degree of certainty.

confidence limits—Extremes of a confidence interval within which there is a designated chance that the true value is included.

configuration—1. The relative arrangement of parts (or components) in a circuit. 2. A listing of the names and/or serial numbers of the assemblies which comprise an equipment.

conformance error — The deviation of a calibration curve from a specified curve line.

conical horn—A horn the cross-sectional area of which increases as the square of the axial length.

conical scanning—A form of scanning in which the beam of a radar unit describes a cone, the axis of which coincides with that of the reflector.

conjugate branches—Any two branches of a network in which a driving force impressed

on one branch does not produce a response in the other.

conjugate bridge—A bridge in which the detector circuit and the supply circuits are interchanged, compared with a normal bridge.

conjugate impedance — An impedance the value of which is the conjugate of a given impedance. For an impedance associated with an electric network, the conjugate is an impedance with the same resistance component and a negative reactive component of the original.

connected—A network is connected if, between every pair of nodes of the network, there exists at least one path composed of branches of the network.

connection — 1. The attachment of two or more component parts so that conduction can take place between them. 2. The point of such attachment.

connection diagram—A diagram showing the electrical connections between the parts comprising an apparatus.

connector—A coupling device which provides an electrical and/or mechanical junction between two cables, or between a cable and a chassis or enclosure.

connector assembly—The combination of a mated plug and receptacle.

conoscope—An instrument for determining the optical axis of a quartz crystal.

consequent poles—Additional magnetic poles present at other than the ends of a magnetic material.

consol—*See* Sonne.

console—1. A cabinet for a radio or television receiver that stands on the floor rather than on a table. 2. Main operating unit in which indicators and general controls of a radar or electronic group are installed.

console operator — A person who monitors and controls an electronic computer by means of a central control unit or console.

consonance — Electrical or acoustical resonance between bodies or circuits not connected directly together.

constant — An unvarying or fixed value or data item.

constant-amplitude recording—The disc-recording characteristic in which the groove displacement is directly proportionate to the signal amplitude.

constant current — A current that does not undergo a change greater than the required precision of the measurement when the impedance of the generator is halved.

constant-current characteristic—The relationship between the voltages of two electrodes, the current to one of them as well as all other voltages being maintained constant.

constant-current modulation — Also called Heising modulation. A system of amplitude modulation in which the output circuits of the signal amplifier and carrier-wave generator or amplifier are directly and conductively coupled by a common inductor. The inductor has an ideally infinite impedance to the signal frequencies and therefore maintains the common plate-supply current of the two devices constant. The signal-frequency voltage thus appearing across the common inductor also modulates the plate supply to the carrier generator or amplifier, with corresponding modulation of the carrier output.

constant-current power supply — A regulated power supply which acts to keep its output current constant in spite of changes in load, line, or temperature. Thus, for a change in load resistance, the output current remains constant to a first approximation, while the output voltage changes by whatever amount necessary to accomplish this.

constant-current transformer — A transformer that automatically maintains a constant current in its secondary circuit under varying conditions of load impedance when supplied from a constant-potential source.

constant-delay discriminator — *See* Pulse Demoder.

constant-K filter—An image-parameter filter comprising a tandem connection of a number of identical prototype L-section filters. Each adjacent pair of L-sections together forms either a T- or π-network. The product of the series and shunt impedances is a constant that is independent of frequency.

constant-K network—A ladder network in which the product of its series and shunt impedances is independent of frequency within the range of interest.

constant-luminance transmission—A type of transmission in which the transmission primaries are a luminance primary and two chrominance primaries.

constant-power-dissipation line — A line superimposed on the output static characteristic curves and representing the points of collector voltage and current, the product of which represents the maximum collector power rating of a particular transistor.

constant-resistance network — A network that will reflect a constant resistance to the output circuit of the driving amplifier when terminated in a resistive load. Loudspeakers do not reflect a constant impedance; therefore, an amplifier does not see a constant resistance. This disadvantage may be somewhat compensated for by the use of negative feedback by the amplifier.

constant-velocity recording — The disc-recording characteristic in which the groove displacement is inversely proportionate to the signal frequency.

constant voltage — A voltage that does not undergo a change greater than the required precision of the measurement when the impedance of the generator is doubled.

constant-voltage/constant-current (cv/cc) output characteristic—A regulated power supply which acts as a constant-voltage source for comparatively large load resistances and as a constant-current source for comparatively small load resistances.

constant-voltage power supply — A regulated power supply which acts to keep its output voltage constant in spite of changes in load, line, or temperature. Thus, for a change in load resistance, the output voltage of this type of supply remains constant to a first approximation, while the output current changes by whatever amount necessary to accomplish this.

constant-voltage transformer — A transformer delivering a fixed predetermined voltage over a limited range of input voltage variations (e.g. 95-125 volts).

consumer's reliability risk—The risk to the consumer that a product will be accepted by a reliability test when the reliability of the product is actually below the value specified for rejection.

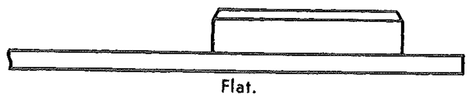

Flat.

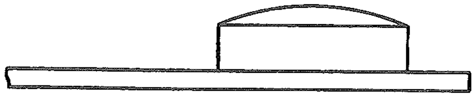

Convex.

Contacts (side view).

contact—1. One of the current-carrying parts of a relay, switch, or connector that are engaged or disengaged to open or close the associated electrical circuits. 2. (v) To join two conductors or conducting objects in order to provide a complete path for current flow. 3. (n) The juncture point to provide the complete path.

contact-actuation time—The time required for any specified contact on the relay to function. When not otherwise specified, it is the initial actuation time. For some purposes, it is preferable to use either the final or effective actuation time.

contact arc — The electrical (current) discharge that occurs between mating contacts when the circuit is being disestablished.

contact area — The common area between two conductors or a conductor and a connector through which the flow of electricity takes place.

contact arrangement—The combination of contact forms that make up the entire relay-switching structure.

contact bounce—The uncontrolled making and breaking of contact when the relay contacts are closed.

contact chatter—The undesired vibration of mating contacts during which there may or may not be actual physical contact opening. If there is no actual opening but only a change in resistance, it is referred to as dynamic resistance, and it appears as "grass" on the screen of an oscilloscope having adequate sensitivity and resolution.

contact combination—1. The total assembly of contacts on a relay. 2. Sometimes used for contact form.

contact emf — A small voltage established whenever two conductors of different materials are brought into contact.

contact follow—The displacement of a stated point on the contact-actuating member following the initial closure of a contact.

contact force—1. The amount of force exerted by one of a pair of closed contacts on the other. 2. The force exerted by the moving mercury on a stationary contact or electrode in a mercury switch.

contact gap—Also called contact separation. The distance between a pair of mating relay contacts when they are open.

contact length—The length of travel of one contact while touching another contact during the assembly or disassembly of a connector.

contact load—The electrical power demands encountered by a contact set in any particular application.

contact microphone—A microphone designed to pick up mechanical vibrations directly from the sound source and convert them into corresponding electrical currents or voltages.

contact modulator — A switch used to produce modified square waves having the same frequency as, and a definite phase relationship to, a driving sine wave. Also called electromechanical chopper.

contact noise—The random fluctuation of voltage across a junction through which current is flowing from one solid to another.

contactor—A heavy-duty relay used to control electrical circuits.

contactor alarm—A signal calling attention to lowered pressure in a cable gas-pressure system.

contact potential—Also called volta effect. The difference of potential that exists when two dissimilar, uncharged metals are placed in contact. One becomes positively charged and the other negatively charged, the amount of potential depending on the nature of the metals.

contact-potential difference—The difference between the work functions of two materials, divided by the electronic charge generated by them.

contact pressure — The amount of pressure holding a set of contacts together.

contact rating — The electrical power-handling capability of relay or switch contacts under specified environmental conditions and for a prescribed number of operations.

contact rectifier—A rectifier consisting of two different solids in contact. Rectification is due to the greater conductivity across the contact in one direction than in the other.

contact resistance—Total electrical resistance of a contact system, such as the resistance of a relay or a switch measured at the terminals. Usually this resistance is only a fraction of an ohm.

contact retention—The minimum axial load a contact in a connector can withstand in either direction while remaining firmly fixed in its normal position in the insert.

contacts — In a relay, the current-carrying parts that engage or disengage to open or close electrical circuits.

contact separation—The maximum distance between the stationary and movable contacts when the circuit is broken.

contact spring—1. A current-carrying spring to which the contacts are fastened. 2. A non-current-carrying spring that positions and tensions a contact-carrying member.

contact wetting—The coating of a contact surface with an adherent film of mercury.

contact wipe—The sliding or tangential motion between two touching contact surfaces.

contaminated—Made radioactive by addition of a radioactive material.

contents—The information stored in any part of the computer memory.

Continental code—*See* International Morse Code.

continuity—1. A continuous path for the flow of current in an electric circuit. 2. In radio broadcasting, the prepared copy from which the spoken material is presented.

continuity test—An electrical test for determining whether a connection is broken.

continuity writer—In radio broadcasting, the person who writes the copy from which the spoken material is presented.

continuous carrier — A carrier over which transmission of information is accomplished by means which do not interrupt the carrier.

continuous commercial service—*See* CCS.

continuous-duty rating—The rating applied to equipment if operated for an indefinite length of time.

continuously loaded cable — A submarine cable in which the conductors are continuously loaded.

continuous power—The power an amplifier is capable of delivering for at least 30 seconds with a sine-wave signal.

continuous rating—The rating that defines the load which can be carried for an indefinite length of time.

continuous recorder—A recorder that makes its record on a continuous sheet or web rather than on individual sheets.

continuous spectrum—The spectrum which exhibits no structure and appears to represent a continuous variation of wavelength from one to the other.

continuous variable—A variable that may assume any value within a defined range.

continuous-wave radar—A system in which a transmitter sends out a continuous flow of radio energy to the target, which reradiates (scatters) the energy intercepted and returns a small fraction to a receiving antenna.

continuous waves — Electromagnetic waves generated as a continuous train of identical oscillations. They can be interrupted according to a code, or modulated in amplitude, frequency, or phase in order to convey information. Abbreviated cw.

continuous-wave tracking system—A tracking system which operates by keeping a continuous radio beam on a target and determining its behavior from changes in the antenna necessary to keep the beam on the target.

contrahelical—In the wire and cable industry the term is used to mean the direction of a layer with respect to the previous layer. Thus it would mean a layer spiraling in an opposite direction from the preceding layer within a wire or cable.

Contran — A computer-programming language in which instructions are written at a compiler level, thereby eliminating the need for translation by a compiling routine.

contrast—1. The actual difference in density between the highlights and the shadows. Contrast is not concerned with the magnitude of density, but only with the difference in densities. 2. Amplitude ratio between picture white and picture black. 3. Ratio between the maximum and minimum brightness values in a picture. 4. In optical character recognition, the differences between the color or shading of the printed material on a document and the background on which it is printed.

contrast control—A method of adjusting the contrast in a television picture by changing the amplitude of the video signal.

contrast range — The ratio between the whitest and blackest portions of a television image.

contrast ratio—Ratio of the maximum to the minimum luminance values in a television picture or a portion thereof.

control—Also called a control circuit. 1. In a digital computer, those parts which carry out the instructions in proper sequence, interpret each instruction, and apply the proper signals to the arithmetic unit and other parts in accordance with the interpretation. 2. Sometimes called a manual control. In any mechanism, one or more components responsible for interpreting and carrying out manually initiated directions. 3. In some business applications, a mathematical check. 4. In electronics, a potentiometer or variable resistor.

control ampere turns—The magnitude and polarity of the control magnetomotive force required for operation of a magnetic amplifier at a specified output.

control card—In computer programming, a card containing input data or parameters for a specific application of a general routine.

control characteristic—1. A plot of the load current of a magnetic amplifier as a function of the control ampere turns for various loads and at the rated supply voltage and frequency. 2. The relationship, usually shown by a graph, between the critical grid voltage and the anode voltage of a tube.

control circuits—In a digital computer, the circuits which carry out the instruction in proper sequence, interpret each instruction, and apply the proper commands to the arithmetic element and other circuits in accordance with the interpretation.

control-circuit transformer — A voltage transformer utilized to supply a voltage suitable for the operation of control devices.

control-circuit voltage — The voltage provided for the operation of shunt-coil magnetic devices.

control compartment — A space within the base, frame, or column of a machine used for mounting the control panel.

control counter — In a computer, a device which records the storage location of the instruction word to be operated on following the instruction word in current use.

control data — In a computer, one or more items of data used to control the identification, selection, execution, or modification of another routine, record file, operation, data value, etc.

control electrode—An electrode on which a voltage is impressed to vary the current flowing between other electrodes.

control field—In a sequence of similar items of computer information, a constant location where control information is placed.

control grid—The electrode of a vacuum tube, other than a diode, upon which a signal voltage is impressed to regulate the plate current.

control-grid bias—The average direct-current voltage between the control grid and cathode of a vacuum tube.

control-grid plate transconductance—The ratio of the amplification factor of a vacuum tube to its plate resistance, combining the effect of both into one term.

controlled avalanche device — A semiconductor device that has very specific maximum and minimum avalanche-voltage characteristics and is also able to operate and absorb momentary power surges in this avalanche region indefinitely without damage.

controlled-carrier modulation — Also called variable-carrier or floating-carrier modulation. A modulation system in which the carrier is amplitude-modulated by the signal frequencies, and also in accordance with the envelope of the signal, so that the modulation factor remains constant regardless of the amplitude of the signal.

controlled rectifier—1. A rectifier employing grid-controlled devices such as thyratrons or ignitrons to regulate its own output current. 2. Also called an scr (silicon-controlled rectifier). A four-layer PNPN semiconductor which functions like a grid-controlled thyratron.

controller—An element or group of elements that take data proportional to the difference between input and output of a device or system and convert these data into power used to restore agreement between input and output.

control locus—A curve which shows the critical value of grid bias for a thyratron.

control panel — A panel having a systematic arrangement of terminals used with removable wires to direct the operation of a computer or punched-card equipment.

control-power disconnecting device—A disconnective device such as a knife switch, circuit breaker, or pullout fuse block used for the purpose of connecting and disconnecting, respectively, the source of control power to and from the control bus or equipment.

control ratio—1. The ratio of the change in anode voltage to the corresponding change in critical grid voltage of a gas tube, with all other operating conditions maintained constant. 2. Also called programming coefficient. The required range in control resistance of a regulated power supply to produce a 1-volt change in output voltage. Expressed in ohms per volt.

control register—In a digital computer, the register that stores the current instruction governing the operation of the computer for a cycle. Also called instruction register.

control section—*See* Control Unit.

control sequence—In a computer, the normal order of execution of instructions.

control tape—In a computer, a paper or plastic tape used to control the carriage operation of some printing output devices. Also called carriage tape.

control unit—That section of an automatic digital computer that directs the sequence of operations, interprets coded instructions, and sends the proper signals to the other computer circuits to carry out the instructions. Also called control section.

control-voltage winding—The motor winding which is excited by a varying voltage at a time phase difference from the voltage applied to the fixed voltage windings of a servomotor.

control winding—In a saturable reactor, the winding used for applying a controlling magnetomotive premagnetization force to the saturable-core material.

convection—The motion in a fluid as a result of differences in density and the action of gravity.

convection cooling—A method of heat transfer which depends on the natural upward movement of the air warmed by the heat dissipated from the device being cooled.

convection current—The amount of time required for a charge in an electron stream to be transported through a given surface.

convection-current modulation — 1. The time variation in the magnitude of the convection current passing through a surface. 2. The process of producing such a variation.

convergence—The condition in which the electron beams of a multibeam cathode-ray tube intersect at a specified point.

convergence coil—One of the two coils associated with an electromagnet, used to obtain dynamic beam convergence in a color-television receiver.

convergence control—A variable resistor in the high-voltage section of a color-television receiver. It controls the voltage applied to the three-gun picture tube.

convergence electrode — An electrode the electric field of which causes two or more electron beams to converge.

convergence magnet—A magnet assembly the magnetic field of which causes two or more electron beams to converge.

convergence phase control—A variable resistor or inductance for adjusting the phase of the dynamic convergence voltage in a color-tv receiver employing a three-gun picture tube.

convergence surface—The surface generated by the point at which two or more electron beams intersect during the scanning process in a multibeam cathode-ray tube.

conversion efficiency—1. The ratio of a-c output power to the d-c input power to the electrodes of an electron tube. 2. The ratio of the output voltage of a converter at one frequency to the input voltage at some other frequency. 3. In a rectifier, the ratio of d-c output power to a-c input-power.

conversion gain—1. The ratio of the intermediate-frequency output voltage to the input-signal voltage of the first detector of a superheterodyne receiver. 2. The ratio of the available intermediate-frequency power output of a converter or mixer to the available radio-frequency power input.

conversion loss—The ratio of available input power to available output power under specified test conditions.

conversion time — The length of time required by a computer to read out all the digits in a given coded word.

conversion transconductance — The magnitude of the desired output-frequency component of current divided by the magnitude of the input-frequency component of voltage when the impedance of the output external termination is negligible for all frequencies which may affect the result.

conversion transducer—One in which the signal undergoes frequency conversion. The gain or loss is specified in terms of the useful signal.

conversion voltage gain (of a conversion transducer)—With the transducer inserted between the input-frequency generator and the output termination, the ratio of the magnitude of the output-frequency voltage across the output termination to the magnitude of the input-frequency voltage across the input termination of the transducer.

convert — To change information from one form to another without changing the meaning, e.g., from one number base to another.

converter—1. In a superheterodyne radio receiver, the section which converts the desired incoming r-f signal into a lower carier frequency known as the intermediate frequency. 2. A rotating machine consisting of an electric motor driving an electric generator, used for changing alternating current to direct current. 3. A facsimile device that changes the type of modulation delivered by the scanner. 4. Generally called a remodulator. A facsimile device that changes amplitude modulation to audio-frequency-shift modulation. 5. Generally called a discriminator. A device that changes audio-frequency-shift modulation to amplitude modulation. 6. A conversion transducer in which the output frequency is the sum or difference of the input frequency and an integral multiple of the local-oscillator frequency. 7. A device that accepts an input that is a function of maximum voltage and time, and converts it to an output that is a function of maximum voltage only. 8. *See* Shaft Position Encaser. 9. A device capable of converting impulses from one mode to another, such as analog to digital, or parallel to serial, or one code to another.

converter tube — A multielement electron tube that combines the mixer and local-oscillator functions of a heterodyne conversion transducer.

converter unit—The unit of a radar system in which the mixer of a superheterodyne receiver and usually two stages of intermediate-frequency amplification are located. Performs a preamplifying operation.

converting—Changing data from one form to another to facilitate its transmission, storage, or manipulation of information.

convex—Curved outward.

Cook system—An early stereo-disc recording technique in which the two channels were recorded simultaneously with two cutters on different portions (bands) of a record as concentric spirals. The playback equipment consisted of two pickups mounted side by side so that each played at the correct spot on its own band.

Coolidge tube—An X-ray tube in which the electrons are produced by a hot cathode.

coordinated transpositions—Transpositions which are installed in either electric supply or communication circuits, or in both, for the purpose of reducing inductive coupling, and which are located effectively with respect to the discontinuities in both the electric supply and communication circuit.

coordinater indexing—1. In a computer, a system in which individual documents are indexed by descriptors of equal rank so that a library can be searched for a combination of one or more descriptors. 2. A computer indexing technique in which the coupling of individual words is used to show the interrelations of terms.

coordination—A term describing the ability of the lower rating of two breakers in series to trip before the higher-rating one trips.

copper loss—*See* I^2R Loss.

copper-oxide rectifier — A metallic rectifier in which the rectifying barrier is the junction between metallic copper and cuprous oxide. A disc of copper is coated with cuprous oxide on one side, and a soft lead washer is used to make contact with the oxide layer.

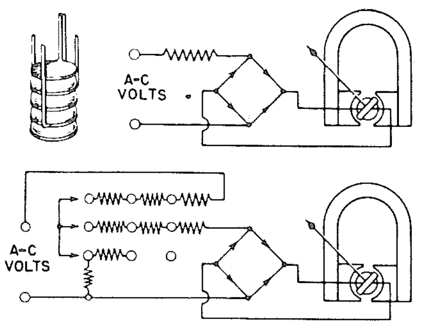

Copper-oxide rectifier.

copper-sulfide rectifier — A semiconductor rectifier in which the rectifying barrier is a junction between magnesium and copper sulfide.

copy—*See* Subject Copy.

copying telegraph—An absolute term for a facsimile system for the transmission of black-and-white copy only.

Corbino effect—A special case of the Hall effect that occurs when a disc carrying a radial current is placed perpendicularly into a magnetic field.

cord—One or a group of flexible insulated conductors covered by a flexible insulation and equipped with terminals.

cord sets—Portable cords fitted with any type of wiring device at one or both ends.

cordwood — A sandwich-type construction wherein components lie in a vertical "cordwood" pattern between horizontal layers.

cordwood module — A high-density circuit module in which discrete components are mounted between and perpendicular to two small, parallel printed-circuit boards to which their terminals are attached.

core—1. A magnetic material placed within a coil to intensify the magnetic field. 2. Magnetic material inside a relay or coil winding.

coreless-type induction heater—A device in which an object is heated by induction without being linked by a magnetic-core material.

core loss—Also called iron loss. Loss of energy in a magnetic core as the result of eddy currents, which circulate through the core and dissipate energy in the form of heat.

core memory — A computer memory device containing magnetic cores.

core plane — A horizontal network of magnetic cores that contains a core common to each storage position.

core storage—In a computer, a form of high-speed storage that uses magnetic cores.

core transformer—A transformer in which the windings are placed on the outside of the core.

core-type induction heater — A device in which an object is heated by induction. Unlike the coreless type, a magnetic core links the inducing winding to the object.

core wrap—Insulation placed over a core before the addition of windings.

corner—An abrupt change in direction of the axis of a waveguide.

corner cut — A corner removed, for orientation purposes, from a card to be used with a computer.

corner effect—The rounding off of the attenuation versus frequency characteristic of a filter at the extremes (or corners) of the passband.

corner frequency — 1. The frequency at which the open-loop gain-versus-frequency curve changes slope. For a servo motor, the product of the corner frequency in radians per second and the time constant of the motor is unity. 2. The frequency at which the two asymptotes of the gain-magnitude curves of an operational amplifier intersect.

corner reflector—A reflecting object consisting of two (dihedral) or three (trihedral) mutually intersecting conducting surfaces. Trihedral reflectors are often used as radar targets.

corner-reflector antenna—An antenna consisting of a primary radiating element and a dihedral corner reflector formed by the elements of the reflector.

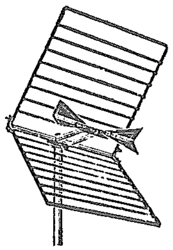

Corner-reflector antenna.

corona—A luminous discharge of electricity, due to ionization of the air, appearing on the surface of a conductor when the potential gradient exceeds a certain value.

corona effect (of alternating current)—The effect produced when two wires, or other conductors having a great difference of voltage, are placed near each other.

corona failure—Failure due to corona degradation at areas of high voltage stress.

corona loss—A loss or discharge which occurs when two electrodes having a great difference of pressure are placed near each other. The corona loss takes place at the critical voltage and increases very rapidly with increasing pressure.

corona resistance—That length of time that an insulation material withstands the action of a specified level of field-intensified ionization that does not result in the immediate, complete breakdown of the insulation. Also called ionization resistance, brush-discharge resistance, or slot-discharge resistance.

corona shield — A shield placed around a high-potential point to redistribute electrostatic lines of force and prevent corona.

corona voltmeter—A voltmeter in which the peak voltage value is indicated by the beginning of corona at a known and calibrated electrode spacing.

corrective equalization—*See* Frequency-Response Equalization.

corrective maintenance—The maintenance performed on a nonscheduled basis to restore equipment to satisfactory condition.

corrective network—Also called shaping network. An electric network designed to be inserted into a circuit to improve its transmission or impedance properties, or both.

correlated characteristic — A characteristic known to be reciprocally related to some other characteristic.

correlation—1. The relationship, expressed as a number between minus one and plus one, between two sets of data, etc. 2. A relationship between two variables; the strength of the linear relationship is indicated by the coefficient of correlation.

correlation detection—A method of detection in which a signal is compared, point to point, with an internally generated reference. The output of such a detector is a measure of the degree of similarity of the input and reference signal. The reference signal is constructed in such a way that it is at all times a prediction, or best guess, of what the input signal should be at that time.

correlation direction finder—A satellite station separated from a radar to receive a jamming signal. By correlating the signals received from several such stations, the range and azimuth of many jammers may be obtained.

correlation distance—A term used in tropospheric propagation. The minimum spatial separation between antennas which will give rise to independent fading of the received signals.

Correlation Orientation Tracking and Range System—A system generally using a parabolic antenna for the analysis of a narrow band of radar energy for tracking and ranging purposes.

correlation tracking and ranging—A nonambiguous short-base-line, single-station c-w phase comparison system measuring two direction cosines and a slant range from which space position can be computed.

correlation tracking and triangulation—A trajectory-measuring system composed of several antenna base lines separated by large distances and used to measure direction cosines to an object. From these measurements, the space position is computed by triangulation.

correlation tracking system—A system utilizing correlation techniques in which signals derived from the same source are correlated to derive the phase difference between the signals. This phase difference contains the system data.

corrosion—A chemical action which causes gradual destruction of the surface of a metal by oxidation or chemical contamination. Also caused by reduction of the electrical efficiency between the metal and a contiguous substance or to the disintegrating effect of strong electrical currents or ground-return currents in electrical systems. The latter is known as electrolytic corrosion.

cosecant - squared antenna — An antenna which emits a cosecant-squared beam. In the shaped-beam antenna used, the radiation intensity over part of its pattern in some specified plane (usually the vertical) is proportionate to the square of the cosecant of the angle measured from a specified direction in that plane (usually the horizontal).

cosecant-squared beam—A radar-beam pattern designed to give uniform signal intensity in echoes from distant and nearby objects. It is generated by a spun-barrel reflector. The beam intensity varies as the square of the cosecant of the elevation angle.

cosine law—The law which states that the brightness in any direction from a perfectly diffusing surface varies in proportion to the cosine of the angle between that direction and the normal to the surface.

cosmic noise—Radio static, the origin of which is due to sources outside the earth's atmosphere. The source may be similar to sunspots, or spots on other stars.

cosmic rays—Any rays of high penetrating power produced by transmutations of atoms in outer space. These particles continually enter the earth's upper atmosphere from interstellar space.

coulomb—The quantity of electricity which passes any point in an electric circuit in 1 second when the current is maintained constant at 1 ampere. The coulomb is the unit of electric charge in the mksa systems.

Coulomb's law—Also called law of electric charges or law of electrostatic attraction. The force of attraction or repulsion between two charges of electricity concentrated at two points in an isotropic medium is proportionate to the product of their magnitudes and is inversely proportionate to the square of the distance between them. The force between unlike charges is an attraction, and between like charges a repulsion.

coulometer—An electrolytic cell which measures a quantity of electricity by the amount of chemical action produced.

Coulter counter—An electronic-cell counting instrument operating on the ion-conductivity principle. Designed by J. R. Coulter. (*Also see* cell counter.)

count — In radiation counters, a single response of the counting system.

countdown—A decreasing tally that indicates the number of operations remaining in a series.

counter—1. A circuit which counts input pulses. One specific type produces one output pulse each time it receives some predetermined number of input pulses. The same term may also be applied to several such circuits connected in cascade to provide digital counting. 2. In mechanical analog computers, a means for measuring the angular displacement of a shaft. 3. Sometimes called accumulator. A device capable of changing from one to the next of a sequence of distinguishable states upon receipt of each discrete input signal.

counter circuit—A circuit which receives uniform pulses representing units to be counted and produces a voltage in proportion to their frequency.

counterclockwise polarized wave—*See* Left-Handed Polarized Wave.

counterelectromotive cell — Cells of practically no ampere-hour capacity used to oppose the line voltage.

counterelectromotive force — Abbreviated counter emf. A voltage developed in an inductive circuit by an alternating or pulsating current. The polarity of this voltage is at every instant opposite that of the applied voltage.

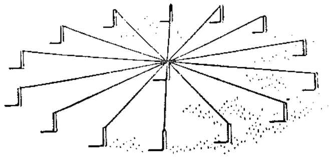

Counterpoise.

counterpoise—A system of wires or other conductors, elevated above and insulated from ground, forming a lower system of conductors of an antenna.

counter tube—Also called radiation counter tube. An electron tube that converts an incident particle or burst of incident radiation into a discrete electric pulse. This is generally done by utilizing the current flow through a gas that is ionized by the radiation.

counting efficiency—In a scintillation counter, the ratio, under specified conditions, of the average number of photons or particles of ionizing radiation that produce counts to the average number of photons or particles incident on the sensitive area.

counting-rate meter—A device for indicating the time rate of occurrence of input pulses averaged over a time interval.

counting-type frequency meter—An instrument for measuring frequency. Its operation depends on the use of pulse-counting techniques to indicate the number and/or rate of recurring electrical signals applied to its input circuits.

counts per turn—The total number of code positions per 360° of encoder shaft rotation.

couple — Two or more dissimilar metals or alloys in electrical contact with each other that act as the electrodes of an electrolytic cell when they are immersed in an electrolyte.

coupled circuit — Any network containing only resistors, inductors (self and mutual), and capacitors, and having more than one independent mesh.

coupler—An arrangement of coils or capacitors so placed with respect to each other that there is electromagnetic or electrostatic coupling between their circuits.

coupling—The association of two or more circuits or systems in such a way that power may be transferred from one to another.

coupling angle—In connection with synchronous motors, the mechanical-degree relationship between the rotor and the rotating field.

coupling aperture—Also called coupling hole or coupling slot. An aperture, in the wall of a waveguide or cavity resonator, designed to transfer energy to or from an external circuit.

coupling capacitor—Any capacitor used to couple two circuits together. Coupling is accomplished by means of the capacitive reactance common to both circuits.

coupling coefficient — Also called coefficient of coupling. The degree of coupling that exists between two circuits. It is equal to the ratio between the mutual impedance and the square root of the product of the total self-impedances of the coupled circuits, all impedances being of the same kind.

coupling hole—*See* Coupling Aperture.

coupling loop—A conducting loop projecting into a waveguide or cavity resonator and designed to transfer energy to or from an external circuit.

coupling probe—A probe projecting into a waveguide or cavity resonator and designed to transfer energy to or from an external circuit.

coupling slot—*See* Coupling Aperture.

coupling transformer—A transformer that couples two circuits together by means of its mutual inductance.

covalent bond—A type of linkage between atoms. Each atom contributes one electron to a shared pair that constitutes an ordinary chemical bond.

coverage—A percentage of the completeness with which a braid or shield covers the surface of underlying insulated conductors or conductor.

cpm—Abbreviation for cycles per minute.

C-power supply—A device connected in the circuit between the cathode and grid of a vacuum tube to apply grid bias.

cps—Abbreviation for cycles per second. Now replaced by hertz (Hz).

crash-locator beacon — Airborne equipment consisting of various transmitters, collapsible antennas, etc., designed to be ejected from a downed aircraft and to transmit beacon signals to help searching forces to locate the crashed aircraft.

crater lamp — 1. A glow-discharge type of vacuum tube the brightness of which is proportional to the current passing through the tube. The glow discharge takes place in a cup or crater rather than on a plate as in a neon lamp. 2. A gaseous lamp usually containing neon. Provides a point source of light that can be modulated with a signal.

crazing—Checking of an insulation material when it is stressed and in contact with certain solvents or their vapors.

creepage — The conduction of electricity across the surface of a dielectric.

creepage distance—The shortest distance between conductors of opposite polarities, or between a live part and ground measured over the surface of the supporting material.

creepage path—The path across the surface of a dielectric between two conductors. Lengthening the creepage path reduces the possibility of arc damage or tracking.

creepage surface—An insulating surface that provides physical separation between two electrical conductors of different potential.

creep recovery—The change in no-load output occurring with time after removal of a load which had been applied for a specific period of time.

crest factor—The ratio of the peak voltage to the rms voltage of a waveform (with the d-c component removed.)

crest factor of a pulse—The ratio of the peak amplitude to the rms amplitude of a pulse.

crest value—Also called peak value. The maximum absolute value of a function.

crest voltmeter—A peak-reading voltmeter.

crimp — To compress or deform a connector barrel around a cable so as to make an electrical connection.

crimp contact—A contact whose back portion is a hollow cylinder to allow it to accept a wire. After a bared wire is inserted, a swedging tool is applied to crimp the contact metal firmly against the wire. An excellent mechanical and electrical contact results. A crimp contact often is referred to as a solderless contact.

crippled leapfrog test—In a computer, a variation of the "leapfrog" test in which the test is repeated from a single set of storage locations rather than from a changing set of storage locations.

critical angle—The maximum angle at which a radio wave may be emitted from an antenna and still be returned to the earth by refraction in the ionosphere.

critical area—*See* elemental area.

critical characteristic—A characteristic not having the normal tolerance to variables.

critical coupling—Also called optimum coupling. Between two circuits independently resonant to the same frequency, the degree of coupling which transfers the maximum amount of energy at the resonant frequency.

critical current—That current, at a specified temperature and in the absence of external magnetic fields, above which a material is normal and below which it is superconducting.

critical damping—1. The value of damping which provides the most rapid transient response without overshoot. 2. Operation on the borderline between underdamping and overdamping.

critical dimension — The dimension of a waveguide cross section that determines the cutoff frequency.

critical failure—A failure that causes a system to operate outside designated limits.

critical field—Also called cutoff field. Of a magnetron, the smallest theoretical value of a steady magnetic-flux density, at a steady anode voltage, that would prevent an electron emitted from the cathode at zero velocity from reaching the anode.

critical frequency—Also called penetrating frequency. The limiting frequency below which a magneto-ionic wave component is reflected by an ionospheric layer and above which the component penetrates the layer at vertical incidence. (*Also see* Waveguide Cutoff Frequency.)

critical grid current—The instantaneous value of grid current in a gas tube when the anode current starts to flow.

critical grid voltage — The instantaneous value of grid voltage at which the anode current starts to flow in a gas tube.

critical high-power level—The radio-frequency power level at which ionization is produced in the absence of a control-electrode discharge.

critical item—*See* Critical Part.

critical magnetic field — That field intensity below which at a specified temperature and in the absence of current, a material is superconducting and above which it is normal.

critical part—A part whose failure to meet specified requirements results in the failure of the product to serve its intended purpose. Also called critical item.

critical potential—*See* Ionization Potential.

critical temperature—That temperature below which, in the absence of current and external magnetic fields, a material is superconducting and above which it is normal.

critical voltage—Also called cutoff voltage. In a magnetron, the highest theoretical value of steady anode voltage, at a given steady magnetic-flux density, at which electrons emitted from the cathode at zero velocity will fail to reach the anode.

critical wavelength — The free-space wavelength corresponding to the critical frequency.

cro — Abbreviation for cathode-ray oscilloscope.

Crookes dark space — *See* Cathode Dark Space.

Crosby system—A compatible multiplex f-m stereo broadcast technique in which the right and left signals are combined in phase (sum signal) and transmitted on the main carrier, and also combined out-of-phase (difference signal) and transmitted on the subcarrier. The two signals are combined (matrixed) in the receiving apparatus to restore the right and left channels.

crossbanding—The use of combinations of interrogation and reply frequencies such that either one interrogation frequency is used with several reply frequencies or one reply frequency is used with several interrogation frequencies.

crosscheck—To check a computation by two different methods.

cross color — The interference in a color-television receiver chrominance channel caused by cross talk from monochrome signals.

cross coupling—Unwanted coupling between two different communication channels or their components.

crossed-field amplifier—A high-power electron tube in which direct-current power is converted to microwave power by a combination of crossed electric and magnetic fields.

crossed-pointer indicator — A two-pointer indicator used with instrument landing systems to indicate the position of an aircraft with respect to the glide path.

crossfire — Interfering current in one telegraph or signaling channel resulting from telegraph or signaling currents in another channel.

crossfoot—1. In a computer, to add or subtract numbers in different fields of the same punch card and punch the result into another field of the same card. 2. To compare totals of the same numbers obtained by different methods.

cross-hatching — In a printed-circuit board, the breaking up of large conductive areas where shielding is required.

cross magnetostriction — Under specified conditions, the relative change of dimension in a specified direction perpendicular to the magnetization of a body of ferromagnetic material when the magnetization of the body is increased from zero to a specified value (usually saturation).

cross modulation—A type of intermodulation that occurs when the carrier of the desired signal is modulated by an undesired signal.

cross neutralization—A method of neutralization used in push-pull amplifiers. A portion of the plate-to-cathode a-c voltage of each tube is applied to the grid-to-cathode circuit of the other tube through a neutralizing capacitor.

crossover—The point where two conductors that are insulated from each other cross.

crossover distortion—Distortion that occurs in a push-pull amplifier at the points of operation where the input signals cross over (go through) the zero reference points.

crossover frequency—As applied to electrical dividing networks, the frequency at which equal power is delivered to each of the adjacent-frequency channels when all channels are terminated in the specified load. (*Also see* Transition Frequency.)

crossover network — A selective network which divides its audio input into two or more frequency bands for distribution to speakers.

crossover spiral—*See* Leadover Groove.

cross polarization—The component electric field vector normal to the desired polarization component.

cross-sectional area of a conductor—The summation of all cross-sectional areas of the individual strands in the conductor, expressed in square inches or more commonly in circular mils.

cross talk—Interference which appears in a given transmitting or recording channel but has its origin in another channel. (*Also see* Magnetic Printing.)

cross-talk coupling — Also called cross-talk loss. Cross coupling between speech communication channels or their component parts. Note: Cross-talk coupling is measured between specified points of the disturbing and disturbed circuit and is preferably expressed in decibels.

cross-talk level — The volume of cross-talk energy, measured in decibels, referred to a base.

cross-talk loss—*See* Cross-talk Coupling.

crowbar—A term describing the action which effectively creates a high overload on the actuating member of the protective device. This crowbar action may be triggered by a slight increase in current or voltage.

crowbar voltage protector—A separate circuit which monitors the output of a power supply and instantaneously throws a short circuit (or crowbar) across the output terminals of the power supply whenever a preset voltage limit is exceeded. An scr is usually used as the crowbar device.

crt—Abbreviation for cathode-ray tube.

cryogenic device—A device intended to function best at temperatures near absolute zero.

cryogenic motor—A motor that operates at a temperature below −129°C and that uses a cryogenic fluid or gas to cool its windings and bearings.

cryogenics—The study and use of devices utilizing properties of materials near absolute zero in temperature.

cryosar—A semiconductor device primarily intended for high-speed computer switching and memory applications. This device operates by the low-temperature avalanche breakdown produced by impact ionization of impurities.

cryostat—A refrigerating unit such as that for producing or utilizing liquid helium in establishing extremely low temperatures (approaching absolute zero).

cryotron — A superconductive four-terminal device in which a magnetic field, produced by passing a current through two input terminals, controls the superconducting-to-

normal transition—and thus the resistance—between the two output terminals.

cryotronics — A contraction of "cryogenic electronics."

crystal—A solid in which the constituent atoms are arranged with some degree of geometric regularity. In communication practice, a piezoelectric crystal, piezoelectric crystal plate, or crystal rectifier.

crystal anisotropy — A force which directs the magnetization of a single-domain particle along a direction of easy magnetization. To rotate the magnetization of the particle, an applied magnetic field must provide enough energy to rotate the magnetization through a difficult crystal direction.

crystal audio receiver—Similar to the crystal video receiver except for the path detection bandwidth, which is audio rather than video.

crystal calibrator—A crystal-controlled oscillator used as a reference to check and set the frequency tuning of a receiver or transmitter.

crystal can relay—A relay mounted in a can of a specific size and shape; called a crystal can because of its common usage as a mounting for quartz crystals used in frequency-control circuits.

crystal control—Control of the frequency of an oscillator by means of a specially designed and cut crystal.

crystal-controlled oscillator — *See* Crystal Oscillator.

crystal-controlled transmitter — A radio transmitter in which the carrier frequency is controlled directly by a crystal oscillator.

crystal cutter—A disc cutter in which the mechanical displacements of the recording stylus are derived from the deformities of a crystal having piezoelectric properties.

crystal detector—A mineral or crystalline material which allows electrical current to flow more easily in one direction than in the other. In this way, an alternating current can be converted to a pulsating current.

crystal diode—A rectifying element comprising a semiconductor crystal having two terminals. Its operation is analogous to that of electron-tube diodes.

crystal filter—A highly selective circuit capable of discriminating against all signals except those at the center frequency of a crystal, which serves as the selective element.

crystal headphones—Headphones using Rochelle-salt or other crystal elements to convert audio-frequency signals into sound waves.

crystal holder—A case of insulating material for mounting a crystal. External prongs allow the crystal to be plugged into a suitable socket.

crystal imperfection—Any deviation in lattice structure from that of a perfect single crystal.

crystal microphone—Also called piezoelectric microphone. A microphone which depends for its operation on the generation of an electric charge by the deformation of a body (usually crystalline) having piezoelectric properties.

crystal mixer—1. A mixer circuit with the frequency of the local oscillator being controlled by a crystal. Normally used in superheterodyne radio receivers. 2. A mixer that utilizes the nonlinear characteristic of a crystal diode to mix two frequencies. Frequently used in radar receivers to convert the received radar signal to a lower i-f signal by mixing it with a local-oscillator signal.

crystal operation—Operation using crystal-controlled oscillators.

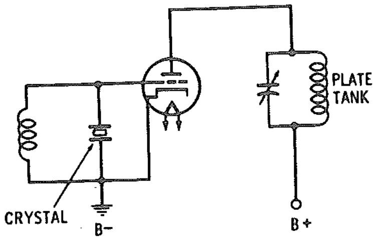

Crystal oscillator.

crystal oscillator — Also called crystal-controlled oscillator. An oscillator in which the frequency of oscillation is controlled by a piezoelectric crystal.

crystal oven—A container, maintained at a constant temperature, in which a crystal and its holder are enclosed in order to keep their temperature constant and thereby reduce frequency drift.

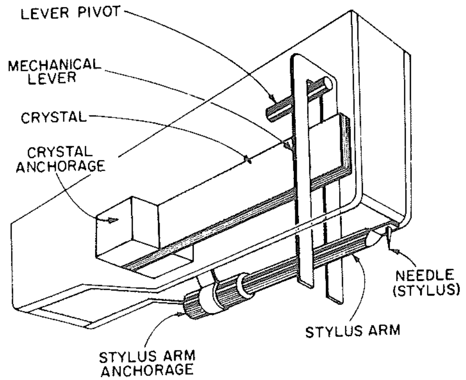

Crystal pickup.

crystal pickup — Also called piezoelectric pickup. A phonograph pickup which depends for its operation on the generation of an electric charge by the deformation of a body (usually crystalline) having piezoelectric properties.

crystal pulling—A method of growing crystals in which the developing crystal is gradually withdrawn from a melt.

crystal rectifier—An electrically conductive or semiconductive substance, natural or synethetic, which has the property of rectifying small radio-frequency voltages.

crystal set—A simple type of radio receiver having only a crystal-detector stage for demodulation of the received signal and no amplifier stages.

crystal slab—A relatively thick piece of crystal from which crystal blanks are then cut.

crystal speaker—Also called piezoelectric speaker. A speaker in which the mechanical displacements are produced by piezoelectric action.

crystal-stabilized transmitter—A transmitter employing automatic frequency control, in which the reference frequency is the same as the crystal-oscillator frequency.

crystal transducer—*See* Piezoelectric Transducer.

crystal video receiver—A radar receiver consisting only of a crystal detector and video amplifier.

C-scope—A rectangular radar display in which targets appear as bright spots with azimuth indicated by the horizontal coordinate and elevation angle by the vertical coordinate.

CT-cut crystal—A natural quartz crystal cut to vibrate below 500 kHz.

CTL—Abbreviation for complementary transistor logic.

cubical antenna—An antenna array, the elements of which are positioned to form a cube.

cue circuit—A one-way communication circuit for conveying program control information.

cup—A single mechanical section of a potentiometer which may contain one or more electrical resistance elements.

cup core—A core which forms a magnetic shield around an inductor. Usually a cylinder with one end closed. A center core inside the inductor is normally used and may or may not be part of the cup core.

cup washer—A washer formed with a recess in one side to retain compression springs or, on binding-post terminals, to prevent escape of connecting wire strands.

curie—A unit used for indicating the strength of radioactive sources in terms of the number of disintegrations per second in the source. One curie is equal to 3.7 × 10^{10} disintegrations per second.

Curie point—Also called magnetic transition temperature. The temperature above which a ferromagnetic material becomes substantially nonmagnetic.

curing temperature—Temperature at which a material undergoes a curing process.

Curpistor—A subminiature constant-current tube containing two electrodes and filled with radioactive nitrogen.

current—The movement of electrons through a conductor. Measured in amperes, and its symbol is I.

current amplification—1. The ratio of the current produced in the output circuit of an amplifier to the current supplied to the input circuit. 2. In photomultipliers, the ratio of the signal output current to the photoelectric signal current from the photocathode.

current amplifier—A device designed to deliver a greater output current than its input current.

current antinode—Also called current loop. The point at which current is a maximum along a transmission-line antenna or other circuit element having standing waves.

current attenuation—The ratio of the magnitude of the current in the input circuit of a transducer to the magnitude of the current in a specified load impedance connected to the transducer.

current-balance relay—A relay in which operation occurs when the magnitude of one current exceeds the magnitude of another current by a predetermined ratio.

current-carrying capacity—The maximum current a conductor (or braid) can carry without heating beyond a safe limit.

current-carrying rating—The current that can be carried continuously or for stated periodic intervals without impairment of the contact structure or interrupting capability.

current density—The amount of electric current passing through a given cross-sectional area of a conductor in amperes per square inch—i.e., the ratio of the current in amperes to the cross-sectional area of the conductor.

current echo—The signal which on a transmission line is reflected as the result of some discontinuity.

current feed—A method of feeding to a point where the current is a maximum (e.g., at the center of a half-wave antenna).

current flicker—Current surges resulting from momentary shorts that can occur within a solid-electrolytic capacitor. They may produce a leakage-current increase, a decrease, no change at all, or a catastrophic failure. Under certain conditions, current flicker can avalanche to cause a short, which under low-impedance circuit conditions results in catastrophic destruction of the capacitor.

current generator—A two-terminal circuit element with a terminal current independent of the voltage between its terminals.

current limiting (automatic)—An overload-protection mechanism which limits the maximum output current of a power supply to a preset value and automatically restores the output when the overload is removed. (*Also see* Short-Circuit Protection.)

current-limiting fuse—A protective device which anticipates a dangerous short-circuit current and opens the circuit, precluding the development of the peak available current.

current-limiting reactor—A form of reactor intended for limiting the current that can flow in a circuit under short-circuit conditions.

current-limiting resistor—A resistor inserted into an electric circuit to limit the flow of current to some predetermined value. Usually inserted in series with a fuse or circuit breaker to limit the current flow during a short circuit or other fault, to prevent excessive current from damaging other parts of the circuit.

current loop—*See* Current Antinode.

current margin—The difference between the steady-state currents flowing through a telegraph receiving instrument corresponding, respectively, to the two positions of the telegraph transmitter.

current-mode logic—Abbreviated CML. A nonsaturating logic circuit that employs the characteristics of a differential amplifier circuit in its design. Because it is nonsaturating, it is the fastest switching logic design. The gate input element is the base of a transistor with a separate transistor for each input.

current node—A point at which current is zero along a transmission line, antenna, or any other circuit element that has standing waves.

current noise—Also called excess noise. A low-frequency noise caused by current flowing in a resistor, particularly a carbon resistor. The amount varies widely with the type and construction of the resistor. This low-frequency noise is generally measurable only in the region below 100 kHz, and the noise power varies inversely with frequency.

current probe—A type of transformer usually having a snap-around configuration, used for measuring the current in a conductor.

current pump—A circuit that drives, through an external load circuit, an adjustable variable or constant value of current, regardless of the reaction of that load to the current, within rated limits of current, voltage, and load impedance.

current regulator—A device that functions to maintain the output current of a generator or other voltage source at a predetermined value, or varies the voltage according to a predetermined plan.

current relay—A relay that operates at a predetermined value of current. It can be an overcurrent relay, an undercurrent relay, or a combination of both.

current saturation—The condition in which the plate current of a vacuum tube cannot be further increased by increasing the plate voltage.

current-sensing resistor—A resistor of low value placed in series with the load to develop a voltage proportional to the output current. A regulated d-c power supply regulates the current in the load by regulating the voltage across this sensing resistor.

current sensitivity—The current required to give standard deflection on a galvanometer.

current-sinking logic—Also called input-coupled logic. A logic form that requires that current flow out of the input of a circuit and back into the output of the preceding stage, which serves as a current sink instead of a source.

current-sourcing logic—Also called collector-coupled logic. A logic form where current flows from the output of a circuit and is forced into the input of a similar circuit to activate the circuit which drives.

current-stability factor—In a transistor, the ratio of a change in emitter current to a change in reverse-bias current between the collector and base.

current-transfer ratio—*See* Beta.

current transformer—A transformer, intended for measuring or control purposes, designed to have its primary winding connected in series with a circuit carrying the current to be measured or controlled.

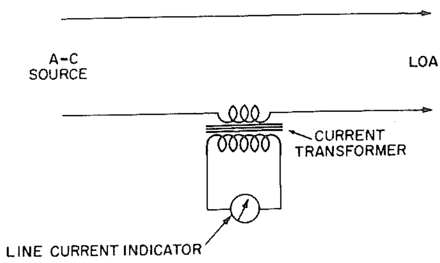

Current transformer.

current-type telemeter—A telemeter in which the magnitude of a single current is the translating means.

cursor—1. A mechanically or electronically generated line which moves back and forth over another surface to delineate accurate readings. 2. A mechanical bearing line on a plan position indicator-type display for reading target bearing.

customer set—*See* Subscriber Set.

Cutler feed—A resonant cavity, at the end of a waveguide, which feeds radio-frequency energy to the reflector of the antenna assembly of some airborne antennas.

cutoff—1. Minimum value of bias which cuts off, or stops, the flow of plate current in a tube. 2. The frequency above or below which a selective circuit fails to respond. 3. The frequency of transmission at which the loss exceeds by 10 decibels that observed at 1,000 hertz.

cutoff attenuator—A variable length of waveguide used below the cutoff frequency of the waveguide to introduce variable nondissipative attenuation.

cutoff current—Transistor collector current with no emitter current and normal collector-to-base bias.

cutoff field—*See* Critical Field.

cutoff frequency—The frequency at which the gain of an amplifier falls below .707 times the maximum gain.

cutoff frequency of a line—The upper frequency limit, usually of a loaded transmission circuit, beyond which the attenuation rises very rapidly.

cutoff limiting—Keeping the output of a vacuum tube below a certain point by driving the control grid beyond cutoff.

cutoff voltage—1. The electrode voltage which reduces the dependent variable of an electron-tube characteristic to a specified value. (*Also see* Critical Voltage.) 2. The voltage at which the discharge is considered complete. This need not be a very low voltage.

cutoff wavelength—The ratio of the velocity of electromagnetic waves in free space to the cutoff frequency of a waveguide.

cutout—1. An electrical device that interrupts the flow of current through any particular apparatus or instrument, either automatically or manually. 2. Pairs brought out of a cable and terminated at some place other than at the end of the cable.

cut-signal branch operation — In systems where radio reception continues without cutting off the carrier, the cut-signal branch operation technique disables a signal branch in one direction when it is enabled in the other to preclude unwanted signal reflections.

cutter — Also called mechanical recording head. An electromechanical transducer which transforms an electric input into a mechanical output (for example, the mechanical motions which a cutting stylus inscribes into a recording medium).

cut-through—The resistance of a solid material to penetration by an object under conditions of pressure, temperature, etc.

cut-through flow test—A test to measure the resistance to deformation of insulation subjected to heat and pressure.

cutting rate—The number of lines per inch the lead screw moves the cutting-head carriage across the face of a recording blank. Standard rates are 96, 104, 112, 128, 136, and greater in multiples of eight lines per inch. For microgroove recordings, 200 to 300 lines per inch are used.

cw—Abbreviation for: 1. Continuous wave. 2. Clockwise.

c-w reference signal—In color television, a sinusoidal signal used to control the conduction time of a synchronous demodulator.

cybernetics—The study of systems of control and communications in humans and animals, and in electrically operated devices such as calculating machines.

cycle—1. The change of an alternating wave from zero to a negative peak to zero to a positive peak and back to zero. The number of cycles per second (hertz) is called the frequency. (*Also see* Alternation.) 2. An off-on application of power.

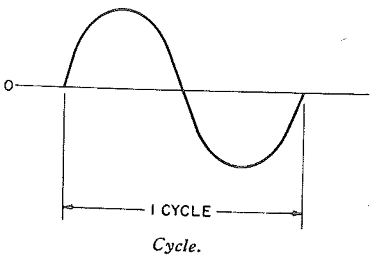

Cycle.

cycle counter—A mechanism or device used to record the number of times a specified cycle is repeated.

cycle index—1. In digital computer programming, the number of times a cycle has been executed. 2. The difference between the number of times a cycle has been executed and the number of times it is desired that the cycle be repeated.

cycle life—For rechargeable cells, the total number of discharge/charge cycles before the cell becomes inoperative.

cycle reset—To return a cycle index to its initial value.

cycle shift — In a computer, the removal of the digits of a number or characters from a word from one end of the number or word and their insertion, in the same sequence, at the other end.

cycle timer—A controlling mechanism which opens or closes contacts according to a preset cycle.

cyclically magnetized condition—The condition of a magnetic material after being under the influence of a magnetizing force varying between two specific limits until, for each increasing (or decreasing) value of the magnetizing force, the magnetic-flux density has the same value in successive cycles.

cyclic code—Positional notation, not necessarily binary, in which quantities differing by one unit are represented by expressions which are identical except for one place or column, and the digits in that place or column differ by only one unit. Cyclic codes are often used in mechanical devices because no ambiguity exists at the changeover point between adjacent quantities.

cyclic memory—A memory that continuously stores information but provides access to any piece of stored information only at multiples of a fixed time called the cycle time.

cyclic shift—A shift in which the data moved out of one end of the storing register are reentered into the other end, as in a closed loop.

cycling vibration—Sinusoidal vibration applied to an instrument and varied in such a way that the instrument is subjected to a specified range of vibrational frequencies.

cycloconverter—A step-down static frequency converter that produces a constant or a precisely controllable output frequency from a variable-frequency a-c power input. In general, the frequency ratio chosen is three to one, or greater.

cyclogram—An oscilloscope display obtained by monitoring two voltages having a direct cyclic relationship to each other.

cycloidotron—A fast-wave crossed-field microwave tube similar to the cyclotron resonance tube.

cyclometer register—A set of four or five wheels numbered from 0 to 9 inclusive on their edges, and enclosed and connected by gearing so that the register reading appears as a series of adjacent digits.

cyclotron—A device consisting of an evacuated tank in which positively charged particles (for example, protons, deuterons, etc.) are guided in spiral paths by a static magnetic field while being highly accelerated by a fixed-frequency electric field.

cyclotron frequency — The frequency at which an electron traverses an orbit in a steady, uniform magnetic field and zero electric field. Given by the product of the electronic charge and the magnetic-flux density, divided by 2π times the electron mass.

cyclotron-frequency magnetron oscillations—Those oscillations having substantially the same frequency as that of the cyclotron.

cyclotron radiation — The electromagnetic radiation emitted by charged particles orbiting in a magnetic field. It arises from the centripetal acceleration of the particle moving in a circular orbit.

cyclotron resonance—The effect characterized by the tendency of charge carriers to spiral around an axis in the same direction as an applied magnetic field, with an angular frequency determined by the value of the applied field and the ratio of the charge to the effective mass of the charge carrier.

cylindrical reflector—A reflector which is part of a cylinder, usually parabolic.

cylindrical wave—A wave the equiphase surfaces of which form a family of coaxial cylinders.

Czochralski technique—A method of growing large single crystals by pulling them from a molten state. Usually used for growing single crystals of germanium and silicon.

D

D—Symbol for electrostatic flux density, deuterium, or dissipation factor.

dag—Abbreviation for Aquadag.

Damon effect—The change in susceptibility of a ferrite, caused by a high r-f power input.

damped natural frequency—1. The frequency at which a system with a single degree of freedom will oscillate, in the presence of damping, after momentary displacement from the rest position by a transient force. 2. The rate of free oscillation of a sensing element in the presence of damping.

damped oscillation—The oscillation that occurs when the amplitude of the oscillating quantity decreases with time. If the rate of decrease can be expressed mathematically, the name of the function describes the damping. Thus, if the rate of decrease is expressed as a negative exponential, the system is said to be an exponentially damped system.

damped waves—Waves in which successive cycles at the source progressively diminish in amplitude.

dampen—1. To diminish progressively in amplitude; usually said of waves or oscillations. 2. To deaden vibrations.

damper—*See* Stabilizer.

damper tube—The tube which conducts in the horizontal-output circuit of a television receiver when the current in the horizontal-deflecting yoke reaches its negative peak. This causes the sawtooth deflection current to decrease smoothly to zero instead of continuing to oscillate.

damper winding—In electric motors, a permanently short-circuited winding, usually uninsulated, and arranged so that it opposes rotation or pulsation of the magnetic field with respect to the pole shoes.

damping—1. Reduction of energy in a mechanical or electrical system by absorption or radiation. 2. Act of reducing the amplitude of the oscillations of an oscillatory system, hindering or preventing oscillation or vibration, or diminishing the sharpness of resonance of the natural frequency of a system. 3. The dissipation of kinetic energy in a system by a controlled energy-absorbing medium. A system can be described as being either critically damped, overdamped, or underdamped. 4. The manner in which the pointer settles to its steady indication after a change in the value of the measured quantity. Two general classes of damped motion are: a. Periodic, in which the pointer oscillates about the final position before coming to rest; b. Aperiodic, in which the pointer comes to rest without overshooting the rest position. Sometimes referred to as overdamping.

damping coefficient—The ratio of actual damping to critical damping.

damping constant—The Napierian logarithm of the ratio of the first to the second of two values of an exponentially decreasing quantity separated by a unit of time.

damping diode — A vacuum tube which damps the positive or negative half-cycle of an a-c voltage.

damping factor — 1. For any underdamped motion during any complete oscillation, the quotient obtained by dividing the logarithmic decrement by the time required by the oscillation. 2. Numerical quantity indicating ability of an amplifier to operate a speaker properly. Values over 4 are usually considered satisfactory. 2. The ratio of rated load impedance to the internal impedance of an amplifier. 3. *See* Decrement.

damping magnet—A permanent magnet and a movable conductor such as a sector or disc arranged in such a way that a torque (or force) is produced which tends to oppose any relative motion.

damping of an instrument—The manner in which the pointer of an instrument settles to a steady indication after a change in the measured quantity has occurred.

damping ratio—1. The ratio of the degree of actual damping to the degree of damping required for critical damping. May be effected by changes in ambient temperature. 2. Of a galvanometer, the ratio, expressed as a positive number, of a given deflection to the next deflection in the opposite direction. The greater this ratio, the greater the degree of damping. The natural logarithm of this ratio is called the logarithmic decrement.

Daniell cell—A cell having a copper electrode in a copper-sulfate solution and a zinc electrode in a diluted sulfuric acid or zinc-sulfate solution, with the two solutions separated by a porous partition. Generates an essentially constant electromotive force of about 1.1 volts.

daraf—The unit of elastance. It equals the reciprocal of capacitance and is actually farad spelled backward.

dark conduction—Residual electrical conduction in a photosensitive substance in total darkness.

dark current—*See* Electrode Dark Current.

dark discharge—In a gas, an electric discharge that has no luminosity.

dark-field disc—A disc used in the optical-electronic type of cell counter that controls light transmission.

dark resistance—The resistance of a photoelectric device in total darkness.

dark spot—A phenomenon sometimes observed in a reproduced television image. It is caused by the formation of electron clouds in front of the mosaic screen in the transmitter camera tube.

dark spot signal—The signal existing in a television system while the television camera is scanning a dark spot.

dark-trace tube—A cathode-ray tube with a screen coated with a halide of sodium or potassium. The screen normally is nearly white, and whenever the electron beam strikes, it turns a magenta color which is of long persistence. The screen can be illuminated by a strong light source, so that the reflected image may be made intense enough to be projected.

Darlington amplifier—Also called Darlington pair, double-emitter follower, or B multiplier. A transistor circuit which, in its original form, consists of two transistors in which the collectors are tied together and the emitter of the first transistor is directly coupled to the base of the second transistor. Therefore, the emitter current of the first transistor equals the base current of the second transistor. This connection of two transistors can be regarded as a compound transistor with three terminals.

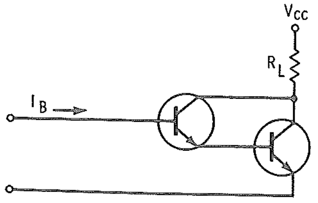

Darlington amplifier.

Darlington pair—*See* Darlington Amplifier.

D'Arsonval current—A high-frequency, low-voltage current of comparatively high amperage.

D'Arsonval galvanometer—A d-c galvanometer consisting of a narrow rectangular coil suspended between the poles of a permanent magnet.

D'Arsonval instrument—*See* Permanent-Magnet Moving-Coil Instrument.

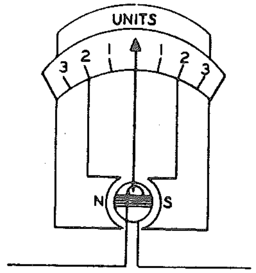

D'Arsonval movement.

D'Arsonval movement—A meter movement consisting essentially of a small, lightweight coil of wire supported on jeweled bearings between the poles of a permanent magnet. When the direct current to be measured is sent through the coil, its magnetic field interacts with that of the permanent magnet and causes the coil and attached pointer to rotate.

dash—Term used in radiotelegraphy. It consists of three units of a sustained transmitted signal followed by one unit during which no signal is transmitted.

dashpot—A device using a gas or liquid to absorb energy from or retard the movement of the moving parts of a circuit breaker or other electrical or mechanical device.

data—A general term used to denote any or all facts, numbers, letters, and symbols, or facts that refer to or describe an object, idea, condition, situation, or other factors. It connotes basic elements of information which can be processed or produced by a computer. Sometimes data are considered to be expressible only in numerical form, but information is not so limited.

data acquisition system—A system in which a computer at a central computing facility gathers data from multiple remote locations.

data collection—In a computer, the transferring of data from one or more points to a central point. Also called data gathering.

data communication—The transmission of information from one point to another.

data conversion—The changing of data from one form of representation to another.

data element—An element which converts data functions into a usable signal. (*Also see* Element.)

data-flow diagram—An illustration having a configuration such that it suggests a certain amount of circuit operation.

data gathering—*See* Data Collection.

data-handling capacity—The number of bits of information which may be stored in a computer system at one time. The rate at which these bits may be fed to the input either by hand or with automatic equipment.

data-handling system—Semiautomatic or automatic equipment used in the collection, transmission, reception, and storage of numerical data.

data link—Electronic equipment which permits automatic transmission of information in digital form.

data processing—The handling of information in a sequence of reasonable operations.

data-processing machines—A general name for a machine which can store and process numerical and alphabetical information.

data processor—An electronic or mechanical machine for handling information in a sequence of reasonable operations.

data reduction—The process of summarizing a large quantity of experimentally obtained data.

dataset—A circuit termination device used to provide interface between a circuit and terminal input/output equipment.

data synchronizer—A device that controls and synchronizes the transmission of data between an input/output (I/O) device and the computer system.

data transmission—The sending of information from one place to another or from one part of a system to another.

data-transmission equipment—The communications equipment used in direct support of data-processing equipment.

data transmission utilization measure—In a data-transmission system, the ratio of the useful data output to the total data input.

daylight visible range—The maximum range at which, under given conditions, a large black object can be seen against a white sky.

db—Abbreviation for decibel.

dba—Abbreviation for decibels adjusted. Used in conjunction with noise measurements. The reference level is −90 dbm, and the adjustment depends on the frequency-band weighting characteristics of the measuring device.

dbj—A unit used to express relative r-f signal levels. The reference level is zero dbj = 1,000 microvolts. (Originated by Jerrold Electronics.)

dbk—Decibels referred to 1 kilowatt.

dbm—Abbreviation for decibels above (or below) one milliwatt. A quantity of power expressed in terms of its ratio to 1 milliwatt.

db meter—A meter having a scale calibrated to read directly in decibel values at a specified reference level (usually one milliwatt equals zero db).

dbRAP—Decibels above reference acoustical power, which is defined as 10^{-16} watts.

dbRN—Decibels above reference noise. This is a unit used to show the relationship between the interfering effect of a noise frequency, or band of noise frequencies, and a fixed amount of noise power commonly called reference noise. A tone of 1000 hertz having a power level of −90 dbm was selected as the reference noise power because it appeared to have negligible interfering effect and would permit the measurement of interfering effect in positive numbers.

dbv—The increase or decrease in voltage independent of impedance levels.

dbw—Decibels referred to 1 watt.

dbx—Decibels above the reference coupling. Reference coupling is defined as the coupling between two circuits that would be required to give a reading of 0 dba on a two-type noise-measuring set connected to the disturbed circuit when a test tone of 90 dba (using the same weighting as that used on the disturbed circuit) is impressed on the disturbing circuit.

dc—Abbreviation for direct current.

D cable—Two-conductor cable, each conductor having the shape of the capital letter D, with insulation between the conductors and between the conductors and the sheath.

d-c amplifier—*See* Direct-Current Amplifier.

d-c balance—An adjustment of circuitry to avoid a change in d-c level when the gain is changed.

d-c beta—The d-c current gain of a transistor; the ratio of the collector current to the base current that caused it, measured at constant collector-to-emitter voltage.

d-c block—A coaxial component employed to prevent the flow of d-c or video along a transmission line while allowing the uninterrupted flow of r-f. The structure is of a short section of coaxial line, having a capacitance in series with the center and/or outer conductor. The r-f flows with negligible reflection or attenuation, while the video frequencies or d-c are blocked.

dcc—Abbreviation for double cotton-covered.

d-c circuit breaker—A device used to close and interrupt a d-c power circuit under normal conditions or to interrupt this circuit under fault or emergency conditions.

d-c component—The average value of a signal. In television it represents the average luminance of the picture being transmitted, and in radar it is the level from which the transmitted and the received pulses rise.

d-c coupled—The connection by a device which passes the steady-state characteristics of a signal and which largely eliminates the transient or oscillating characteristics of the signal.

d-c dump—The withdrawal of direct-current power from a computer. This may result in loss of the stored information.

d-c generator—A rotating electric device for converting mechanical power into d-c power.

d-c inserter stage—A television transmitter stage that adds a d-c component known as the pedestal level to the video signal.

d-c leakage current—The conduction current that flows through a capacitor when rated d-c voltage is applied.

d-c noise—The noise arising when reproducing a tape which has been nonuniformly magnetized by energizing the record head with direct current, either in the presence or absence of bias. This noise has pronounced long-wavelength components which can be as much as 20 db higher than those obtained from a bulk-erased tape.

d-c noise margin—The difference between the output voltage level of a driving gate and the input threshold voltage of a driven gate for both the "1" and the "0" states.

d-c operating point—The d-c values of collector voltage and current of a transistor with no signal applied.

d-c overcurrent relay—A device which functions when the current in a d-c circuit exceeds a given value.

d-c patch bay — Specific patch panels provided for termination of all direct-current circuits and equipment used in an installation.

d-c picture transmission—Transmission of the d-c component of the television picture signal. This component represents the background or average illumination of the over-all scene and varies only with the over-all illumination.

d-c plate resistance—The value or characteristic used in vacuum-tube computations. It is equal to the direct-current plate voltage divided by the direct-current plate current and is given the symbol r_p.

d-c reclosing relay—A device which controls the automatic closing and reclosing of a d-c circuit interrupter, generally in response to load-circuit conditions.

d-c resistivity—The resistance of a body of ferromagnetic material having a constant cross-sectional area, measured under stated conditions by means of direct voltage, multiplied by the cross-sectional area, and divided by the length of the body.

d-c restoration—The re-establishment, by a sampling process, of the d-c and the low-frequency component which, in a video signal, have been suppressed by a-c transmission.

d-c restorer—Also called clamper or restorer. A clamping circuit which holds either amplitude extreme of a signal waveform to a given reference level of potential.

d-c shift — An error in transient response, with a time constant approaching several seconds.

d-c short—A coaxial component that provides a d-c circuit between the center and outer conductors, while allowing the r-f signal to flow uninterrupted. The unit has a high-impedance line shunted across the main coax line. This consequently makes this device frequency dependent. However, it can be acquired in octave bandwidths.

d-c signaling—A transmission method which utilizes direct current.

DCTL — Abbreviation for direct-coupled transistor logic.

d-c transducer—A transducer capable of proper operation when excited with direct current. Its output is given in terms of direct current unless otherwise modified by the function of the stimulus.

d-c transmission—Transmission of a television signal in such a way that the d-c component of the picture signal is still present. This is done to maintain the true level of background illumination.

dcwv—Abbreviation for direct current working volts. The maximum continuous voltage which can be applied to a capacitor.

deac—A device employed in frequency-modulation receivers to de-emphasize the higher frequencies in the received signal so as to restore their proper relative amplitude. (*Also see* De-emphasis.)

deaccentuator — A network or circuit employed in frequency-modulated receivers to de-emphasize the higher frequencies in the received signal to store their proper relative amplitude.

dead—Free from any electric connection to a source of potential difference and from electric charge; having the same potential as that of earth. The term refers only to current-carrying parts which are sometimes alive, or charged.

dead band—1. In a control system, the range of values through which the measurand can be varied without initiating an effective response. 2. Also called dead space, dead zone, or switching blank. A specified range of values in which the incoming signal can be altered without also changing the outgoing response.

deadbeat—Coming to rest without vibration or oscillation — i.e., the pointer which a highly damped meter or galvanometer moves to a new position without overshooting and vibrating about its final position.

deadbeat instrument—A voltmeter, ammeter, or similar device in which the movement is highly damped to bring it to rest quickly.

dead break—An unreliable contact made near the trip point of a relay or switch at low contact pressure. As a result, the switch does not actuate, even though the circuit is interrupted.

dead end—1. In a sound studio, the end with the greater sound-absorbing characteristic. 2. In a tapped coil, the portion through which no current is flowing at a particular bandswitch position.

dead-end tower—An antenna or transmission-line tower designed to withstand unbalanced mechanical pull from all the conductors in one direction, together with the wind strain and vertical loads.

dead room—A room for testing the acoustic efficiency or range of electroacoustic devices such as speakers and microphones. The room is designed with an absolute minimum of sound reflection, and no two dimensions of the room are the same. A ratio of 3 to 4 to 5 is usually employed (e.g., 15′ × 20′ × 25′). The walls, floor, and ceiling are lined with a sound-absorbing material.

dead short—A short circuit having minimum resistance.

dead space—1. An area or zone, within the normal range of a radio transmitter, in which no signal is received. 2. *See* Dead Band.

dead spot—1. A geographic location in which signals from one or more radio stations are received poorly or not at all. 2. That portion of the tuning range of a receiver where stations are heard poorly or not at all because of poor sensitivity.

dead time—The minimum interval, following a pulse, during which a transponder or component circuit is incapable of repeating a specified performance.

dead volume—The total volume of the pressure port cavity of a pressure transducer at the rest position (i.e., with no stimulus applied).

dead zone—1. The range of values over which no corrective action will take place in an automatic control system. 2. *See* Dead Band.

deafness—*See* Hearing Loss.

debug—To examine or test a procedure, routine, or equipment for the purpose of detecting and correcting errors.

debugging — 1. Isolating and removing all malfunctions ("bugs") from a computer or other device to restore its operation. 2. A process undertaken to carry an equipment or a group of parts through the initial period during which catastrophic-type failures occur at an abnormally high rate.

debugging period—*See* Early Failure Period.

debunching—Space-charge effect that tends to destroy the electron bunching in a velocity-modulation vacuum tube by spreading the beam due to mutual repulsion of the electrons.

Debye length—A theoretical length that describes the maximum separation at which a given electron is influenced by the electric field of a given positive ion. Also called Debye shielding distance or plasma length.

Debye shielding—*See* Debye Length.

decade—1. The interval between any two quantities having a ratio of 10:1. 2. A group or assembly of 10 units (e.g., a counter which counts to 10 in one column, or a resistor box which inserts resistance quantities in multiples of powers of 10).

decade band—A band having frequency limits related by the equation $f_b - f_a = 10$.

decade box—A special assembly of precision resistors, coils, or capacitors. It contains two or more sections, each having 10 times the value of the preceding section. Each section is divided into 10 equal parts. By means of a 10-position selector switch or equivalent arrangement, the box can be set to any desired value in its range.

decade counter—*See* Decade Scaler.

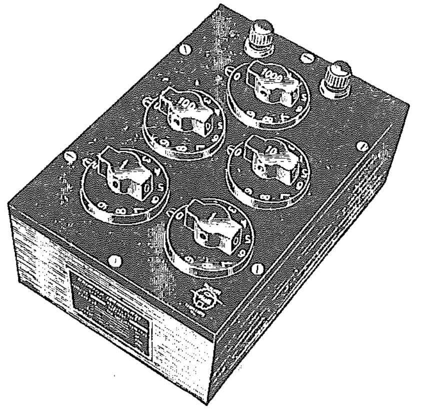

Decade resistance box.

decade resistance box—A resistance box containing two or more sets of 10 precision resistors.

decade scaler—Counter or scale-of-ten counter. A scaler with a factor of 10. It produces one output pulse for every 10 input pulses.

decalesent point—The temperature at which there is a sudden absorption of heat as the temperature of a metal is raised.

decametric waves—High-frequency band; 3 MHz to 30 MHz.

decay—1. Gradual reduction of a quantity. 2. The decrease in the radiation intensity of any radioactive material with respect to time. 3. In a storage tube, a change in magnitude or configuration of stored information by any cause other than erasing or writing.

decay characteristic—*See* Persistence Characteristic (of a luminescent screen).

decay distance — The distance between an area of wave generation and a point of passage of the resulting waves outside the area.

Decca—A British long-range hyperbolic navigational system that operates in the 70- to 130-kilohertz frequency band. It is a continuous-wave system in which the receiver measures and integrates the relative phase difference between the signal received from two or more synchronized ground stations. One master station and three slave stations are usually arranged in stat formation. Operational range is about 250 miles.

decelerated electrons — Electrons which, after traveling at a great rate of speed, strike a target, become quickly decelerated, and cause the target to emit X rays.

decelerating electrode—In an electron-beam tube, an electrode to which a potential is applied to slow down the electrons in the beam.

deceleration—The act or process of moving, or of causing to move, with decreasing speed; the state of so moving.

deceleration time—In a computer, the time interval between the completion of the reading or writing of a record on a magnetic tape and the time when the tape stops moving.

deception—Deliberate production of false or misleading echoes on enemy radar by the radiation of spurious signals synchronized to the radar or by the reradiation of the radar pulses from extraneous reflectors.

deception device – A device that works to make unfriendly signals either unusable or misleading.

deci—Prefix meaning one-tenth (10^{-1}).

decibel—Abbreviated db. The standard unit for expressing transmission gain or loss and relative power levels. The term "dbm" is used, when a power of one milliwatt is the reference level. Db indicates the ratio of power output to power input:

$$db = 10 \log_{10} \frac{P_1}{P_2}$$

One decibel is 1/10 of a bel.

decibel meter—Also called db meter. An instrument for measuring the electric power level, in decibels, above or below an arbitrary reference level.

decibels above or below 1 milliwatt – The unit used to describe the ratio of the power at any point in a transmission system to a reference level of 1 milliwatt. The ratio expresses decibels above or below this reference level of 1 milliwatt.

decibels above or below 1 watt—A measure of power expressed in decibels to a reference level of 1 watt.

decibels above reference noise—An expression used to describe the ratio of the circuit noise level in a transmission system, at any point, to some arbitrarily chosen reference noise. The expression signifies the reading of a noise meter. Where the circuit noise meter has been adjusted to represent effect under specified conditions, the expression is in adjusted decibels.

decilog—A division of the logarithmic scale used for measuring the logarithm of the ratio of two values of any quantity. The number of decilogs is equal to 10 times the logarithm to the base 10 of the ratio. One decilog therefore corresponds to a ratio of $10^{0.1}$ (i.e. 1.25892+).

decimal—1. Pertaining to a characteristic or property involving a selection, choice, or condition in which there are 10 possibilities. 2. Pertaining to the number representation system with a radix of 10.

decimal attenuator—A system of attenuators arranged so that a voltage or current can be reduced decimally.

decimal notation—The writing of quantities in the decimal numbering system.

decimal numbering system – The popular numbering system using the Arabic numerals 0 through 9 and thus having a base, or radix, of 10. For example, the decimal number 2,345 can be derived in this way:

$$2{,}000 + 300 + 40 + 5 = 2{,}345 \text{ or:}$$

$$2\,(10^3) + 3\,(10^2) + 4\,(10^1) + 5\,(10^0) = 2{,}345.$$

In the decimal system, all numbers are obtained by raising the radix (total number of marks, or 10 in this system) to various powers.

decimal point—In a decimal number, the point which marks the place between integral and fractional powers of 10.

decimal-to-binary conversion – The mathematical process of converting a number written in the scale of 10 into the same number written in the scale of 2.

decimetric waves—1. Electromagnetic waves having wavelengths between 0.1 and 1 meter. 2. Ultrahigh frequency band; 300 MHz to 3 GHz.

decineper—One-tenth of a neper.

decinormal calomel electrode – A calomel electrode containing a decinormal potassium chloride solution.

decision—In a computer, the process of determining further action on the basis of the relationship of two similar items of data.

decision box—On a flowchart, a rectangle or other symbol used to mark a choice or branching in the sequence of programming of a digital computer.

decision element – In computers or data-handling systems, a circuit which performs a logical operation—such as AND, OR, NOT, or EXCEPT on one or more binary digits of input information which represent "yes" or "no"—and expresses the result in its output.

declination—The angular difference between the position of a compass needle and the true position of geographical north and south.

declinometer—Also called a compass declinometer. A device for measuring the direction of a magnetic field relative to astronomical or survey coordinates.

decode—In a computer, to obtain a specific output when specific character-coded input lines are activated.

decoder—1. A device for decoding a series of coded signals. 2. In automatic telephone switching, a relay-type translator which determines from the office code of each call the information required for properly recording the call through the switching train. Each decoder has means, such as a cross-connecting field, for establishing the controls desired and readily changing them. 3. Sometimes called matrix. In an electronic computer, a network or system in which a combination of inputs is excited at one time to produce a single output.

decoding—1. The process of obtaining intelligence from a code signal. 2. In multiplex, a process of separating the subcarrier from the main carrier.

decommutation—The process of recovering a signal from the composite signal previously created by a commutation process.

decommutator – Equipment for separating, demodulating, or demultiplexing commutated signals.

decoupling—The reduction of coupling.

decoupling circuit—A circuit used to prevent interaction of one circuit with another.

decoupling network—A network of capacitors and chokes or resistors placed into leads which are common to two or more circuits, to prevent unwanted, harmful interstage coupling.

decoy—A reflecting object used in radar deception, having reflective characteristics of a target.

decrement—1. Progressive diminution in the value of a variable quantity; also the amount by which a variable decreases. When applied to damped oscillations, it is usually called damping factor. 2. A specific part of an instruction word in some binary computers, thus a set of digits.

dee – A hollow, D-shaped accelerating electrode in a cyclotron.

dee line—A structural member that supports the dee of a cyclotron and together with the dee forms the resonant circuit.

de-emphasis—Also called post-emphasis or post-equalization. Introduction of a frequency-response characteristic which is complementary to that introduced in pre-emphasis.

de-emphasis network—A network inserted into a system to restore the pre-emphasized frequency spectrum to its original form.

de-energize—To disconnect a device from its power source.

deep discharge—The withdrawal of all available electrical energy before recharging a cell or battery.

defect – A condition considered potentially hazardous or operationally unsatisfactory and therefore requiring attention.

defect conduction—Hole conduction in the valence band in a semiconductor.

defibrillator—A device used to eliminate fibrillation (irregular beating) of the heart muscle by the application of high-voltage impulses.

definite-purpose relay – A relay with some electrical or mechanical feature which distinguishes it from a general-purpose relay.

definition—1. The fidelity with which the detail of an image is reproduced. When the image is sharp (i.e., has definite lines and boundaries), the definition is said to be good. 2. The degree with which a communication system reproduces sound images or messages.

deflecting coil—An inductor used to produce a magnetic field that will bend the electron beam a desired amount in the cathode-ray tube of an oscilloscope, television receiver, or television camera.

deflecting electrode—An electrode to which a potential is applied in order to deflect an electron beam.

deflecting torque—*See* Torque of an Instrument.

deflection—Movement of the electron beam in a cathode-ray tube as electromagnetic or electrostatic fields are varied to cause the light spot to traverse the face of the tube in a predetermined pattern.

deflection coil—One of the coils in the deflecting yoke.

deflection factor—*See* Deflection Sensitivity.

deflection plane—A plane perpendicular to the cathode-ray-tube axis and containing the deflection center.

deflection plates—Two pairs of parallel electrodes, the pairs set one forward of the other and at right angles to each other, parallel to the axis of the electron stream within an electrostatic cathode-ray tube. An applied potential produces an electric field between each pair. By varying the applied potential, this field may be varied to cause a desired angular displacement of the electron stream.

deflection polarity – The relationship between the direction of displacement of an oscilloscope trace and the polarity of the applied signal wave.

deflection polarity of an oscilloscope—The relationship between the direction of a displacement of the cathode beam and the polarity of the applied signal wave.

deflection sensitivity—Also called deflection factor. The displacement of the electron beam at the target or screen of a cathode-ray tube divided by the change in magnitude of the deflecting field. Deflection sensitivity is usually expressed in millimeters (or inches) per volt applied between the deflecting electrodes, or in millimeters (or inches) per ampere in the deflection coil.

deflection voltage—The voltage applied to the electrostatic plates of a cathode-ray tube to control the movement of the electron beam.

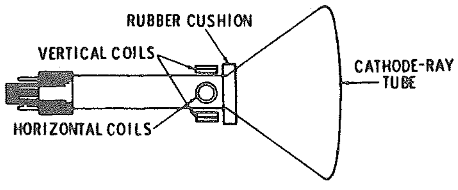

Deflection yoke.

deflection yoke—An assembly of one or more electromagnets for deflecting one or more electron beams.

degassing—The process of driving out and exhausting the gases of an electron tube occluded in its internal parts.

degausser—Also called automatic degausser and bulk eraser. A device to clarify the color picture by means of coils within the set. The coils deactivate the magnetization which builds up around a color tv set when it is moved around, or when other electrical devices are brought too close to the receiver.

degaussing—Girdling a ship's hull with a web of current-carrying cable that sets up a magnetic field equal in value and opposite in polarity to that induced by the earth's magnetic field, thus rendering the ship incapable of actuating the detonator of a magnetic mine.

degeneracy—The condition in which two or more modes have the same resonant frequency in a resonant device.

degenerate modes—A set of modes having the same resonance frequency (or propagation constant). The members of a set of degenerate modes are not unique.

degenerate parametric amplifier — An inverting parametric device for which the two signal frequencies are identical and equal to one-half the frequency of the pump. (This exact but restrictive definition is often relaxed to include cases where the signals occupy frequency bands which overlap.)

degeneration—*See* Negative Feedback.

degradation—A gradual decline of quality or loss of ability to perform within required limits. The synonym "drift" is often used for electronic devices.

degradation failure—Failure resulting from a gradual change in the performance characteristics of a device or part.

degree of current rectification—The ratio between the average unidirectional current output and the root-mean-square value of the alternating-current input from which it was derived.

degree of voltage rectification—The ratio between the average unidirectional voltage and the root-mean-square value of the alternating voltage from which it was derived.

deion circuit breaker — A circuit breaker built so that the arc that forms when the circuit is broken is magnetically blown into a stack of insulated copper plates, giving the effect of a large number of short arcs in series. Each arc becomes almost instantly deionized when the current drops to zero in the alternating-current cycle, and the arc cannot re-form.

deionization—The process by which an ionized gas returns to its neutral state after all sources of ionization have been removed.

deionization potential — The potential at which ionization of the gas within a gas-filled tube ceases and conduction stops.

deionization time—The time required for the grid of a gas tube to regain control after the anode current has been interrupted.

dekahexadecimal—*See* Sexidecimal Notation.

Dekatron—A cold-cathode counting tube.

delay—The time required for a signal to pass through a device or conductor.

delay circuit—A circuit which delays the passage of a pulse or signal from one part of a circuit to another.

delay coincidence circuit — A coincidence circuit actuated by two pulses, one of which is delayed a specific amount with respect to the other.

delay counter—In a computer, a device that can temporarily delay a program a sufficient length of time for the completion of an operation.

delay distortion — 1. Phase-delay distortion (i.e., departure from flatness in the phase delay of a circuit or system over the frequency range required for transmission, or the effect of such departure on a transmitted signal). 2. Also called envelope delay distortion or phase delay. Envelope distortion (i.e., departure from flatness in the envelope delay of a circuit or system over the frequency range required for transmission, or the effect of such departure on a transmitted signal). 3. The amount of variation in delay for various frequency components of the facsimile signal, usually expressed in microseconds from an average delay time.

delayed automatic volume control—Abbreviated delayed avc. An automatic volume-control circuit that acts only on signals above a certain strength. It thus permits reception of weak signals even though they may be fading, whereas normal automatic volume control would make the weak signals even weaker.

delayed avc—*See* Delayed Automatic Volume Control.

delayed contacts—Contacts which are actuated a predetermined time after the start of a (timing) cycle.

delayed ppi—A ppi (plan-position indicator) in which the initiation of the time base is delayed.

delayed repeater satellite—A satellite which stores information obtained from a ground terminal at one location and, upon interrogation by a terminal at a different location, transmits the stored message.

delayed sweep—1. In a cathode-ray tube, a type of sweep which is not allowed to begin for a while after being triggered by the initiating pulse. 2. A sweep that has been delayed either by a predetermined period or by a period determined by an additional independent variable.

delay equalizer—1. A device which adds delay at certain frequencies to a circuit in a way to reduce the delay distortion. 2. A corrective network which is designed to make the phase delay or envelope delay of a circuit or system substantially constant over a desired frequency range.

delay/frequency distortion—That form of distortion which occurs when the envelope delay of a circuit or system is not constant over the frequency range required for transmissions.

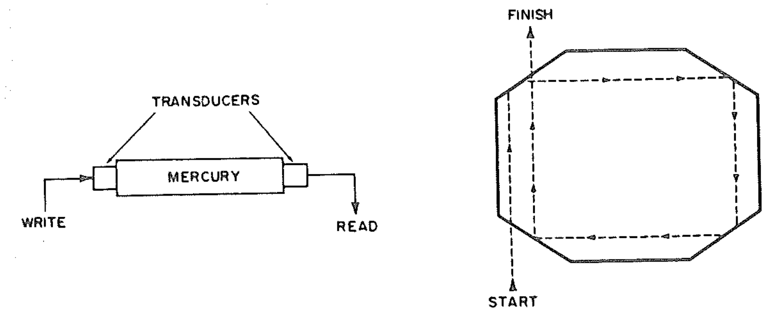

Mercury type.

Prism type.

Delay lines.

delay line—1. A real or artificial transmission line or equivalent device designed to delay a signal or wave for a predetermined length of time. 2. A specially constructed cable used in the luminance channel of a color receiver to delay the luminance signal.

delay-line memory—*See* Delay-Line Storage.

delay-line register—An acoustic or electric delay line in an electronic computer, usually one or an integral number of words in length, together with input, output, and circulation circuits.

delay-line storage—Also called delay-line memory. In an electronic computer, a storage or memory device consisting of a delay line and a means for regenerating and reinserting information into it.

delay multivibrator—A monostable multivibrator that produces an output pulse a predetermined time after it is triggered by an input pulse.

delay relay—Also called time-delay relay. A relay in which there is a delay between the time it is energized or de-energized and the contacts open or close.

delay time—The amount of time one signal is behind (lags) another.

delay timer—A term sometimes used to designate a timer which is primarily used for energizing (or de-energizing) a load at the end of a timed period. (*See* Time Delay Relay—preferred meaning.)

delay unit—The unit of a radar system in which pulses may be delayed a controllable amount.

deletion record—In a computer, a new record to replace or remove an existing record in a master file.

delimiter—1. In a computer, a character that limits a string of characters and therefore cannot be a member of the string. 2. Also called separator. A flag that separates and organizes items of data.

Dellinger effect—*See* Radio Fadeout.

delta—1. The Greek letter delta (Δ) represents any quantity which is much smaller than any other quantity of the same units appearing in the same problem. 2. In a magnetic cell, the difference between the partial-select outputs of the same cell in a one state and in a zero state.

delta circuit—A three-phase circuit in which the windings of the system are connected in the form of a closed ring, and the instantaneous voltages around the ring equal zero. There is no common or neutral wire, so the system is used only for three-wire systems or generators.

delta connection—In a three-phase system, the terminal connections. So called because they are triangular like the Greek letter delta.

delta match—*See* Y Match.

delta matched antenna—Also called Y antenna. A single-wire antenna (usually one half-wavelength long) to which the leads of an open-wire transmission line are connected in the shape of a Y. The flared part of the Y matches the transmission line to the antenna. The top of the Y is not cut, giving the matching section its triangular shape of the Greek letter *delta*, hence the name.

delta matching transformer—An impedance device used to match the impedance of an open-wire transmission line to an antenna. The two ends of the transmission line are fanned out so that the impedance of the line gradually increases. The ends of the transmission line are attached to the antenna at points of equal impedance, symmetrically located with respect to the center of the antenna.

delta modulation — A means of encoding analog signals in control and communication systems. The output of the delta encoder is a single weighed digital pulse train which may be decoded at the receiving end to reconstruct an original analog signal.

delta network—A set of three branches connected in series to form a mesh.

delta pulse code modulation—A modulation system that converts audio signals into corresponding trains of digital pulses to provide greater freedom from interference during transmission over wire or radio channels.

delta wave—A brain wave the frequency of which is below 9 hertz.

dem—Abbreviation for demodulator.

demagnetization—Partial or complete reduction of residual magnetism.

demagnetization curve—In the second quadrant of a hysteresis loop, the portion which lies between the residual induction point, B_r, and the coercive force point, H_c.

demagnetization effect—A decrease in internal magnetic field caused by uncompensated magnetic poles at the surface of a sample.

demagnetizer—A device for removing the magnetism which may build up in a recording or playback head.

demagnetizing force—A magnetizing force applied in such a direction that it reduces the residual induction in a magnetized body.

Dember effect—Also known as the photodiffusion effect. The production of a potential difference between two regions of a semiconductor specimen when one is illuminated. This phenomenon is related to the photoelectromagnetic effect, except there is no magnetic field. H. Dember discovered when an illuminated metal plate, producing electrons, is bombarded by other electrons from an outside source, the photoelectric emission increases because, in addition to photoelectrons, secondary electrons are also knocked out by bombardment.

demodulation — Also called detection. The operation on a previously modulated wave in such a way that it will have substantially the same characteristics as the original modulating wave.

demodulator—1. A device which operates on a carrier wave to recover the wave with which the carrier was originally modulated. 2. A facsimile device which detects an amplitude-modulated signal and produces the modulating frequency as a direct current of varying amplitude. This type of unit is used to provide a keying signal for a frequency-shift exciter unit for radio facsimile transmission. 3. A device which receives tones from a transmission circuit and converts them to electrical pulses, or bits, which may be accepted by a business machine.

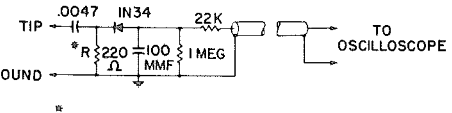

*R MAY BE INCREASED IN VALUE TO REDUCE CIRCUIT LOADING

Demodulator probe.

demodulator probe—A probe designed for use with an oscilloscope, for displaying modulated high-frequency signals.

demountable tube — A high-power electron tube having a metal envelope with porcelain insulation. Can be taken apart for inspection and for renewal of electrodes.

demultiplexer—A device used to separate two or more signals that were previously combined by a compatible multiplexer and are transmitted over a single channel.

demultiplexing circuit — A circuit that is used to separate the signals that have been combined for transmission by multiplex.

denary band — A band having frequency limits with the ratio of $f_b/f_a = 10$.

dendrite — A semiconductor crystal with a heavily-branched, tree-like structure which grows from the nucleus as the metal becomes solidified.

dendritic growth—A technique of producing semiconductor crystals in long, uniform ribbons with optically flat surfaces.

dense binary code—A binary code in which all the possible states of the pattern are used.

densitometer—An instrument for measuring the optical density (photographic transmission, photographic reflection, visual transmission, and so forth) of a material.

density—1. A measure of the light-reflecting or -transmitting properties of an area. 2. The mass per unit volume. The specific gravity of a body is the ratio of a density to the density of a standard substance. Water and air are commonly used as the standard substances. 3. Amount per unit cross-sectional area (e.g., current, magnetic flux, or electrons in a beam).

density modulation—Modulation of an electron beam by varying the density of the electrons in the beam with time.

dentaphone—A device by which deaf persons can hear sounds via the teeth.

dependent linearity—Nonlinearity errors expressed as a deviation from a desired straight line of fixed slope and/or position.

depletion field-effect transistor (FET) — An active semiconductor device in which the main current is controlled by the depletion width of a P-N junction.

depletion layer—Also called barrier layer. In a semiconductor, the region in which the mobile-carrier charge density is insufficient to neutralize the net fixed charge density of donors and acceptors.

depletion-layer capacitance — Also called barrier capacitance. Capacitance of the depletion layer of a semiconductor. It is a function of the reverse voltage.

depletion-layer rectification — Also called barrier-layer rectification. The rectification that appears at the contact between dissimilar materials, such as a metal-to-semiconductor contact or a PN junction, as the energy levels on each side of the discontinuity are readjusted.

depletion-layer transistor—Any of several types of transistors which rely directly for their operation on the motion of carriers through depletion layers (for example, a spacistor).

depletion region—Also referred to as the space charge, barrier, or intrinsic region. In a semiconductor, the region containing the acceptor and donor ions the excess holes or electrons of which have been removed.

depolarization—The process of preserving the activity of a primary cell by the addition of a substance to the electrolyte. This substance combines chemically with the hydrogen gas as it forms, thus preventing excessive buildup of hydrogen bubbles.

depolarizer—A chemical used in some primary cells to prevent formation of hydrogen bubbles at the positive electrode.

deposition — The application, usually in a vacuum, of a material to a substrate by means of an electrical, chemical, screening, or vapor method.

depth finder—*See* Fathometer.

depth of cut—The depth to which the recording stylus penetrates the lacquer of a recording disc.

depth of heating—The depth at which effective dielectric heating can be confined below the surface of a material when the applicator electrodes are placed adjacent to only one surface.

depth of modulation—In a radio-guidance system obtaining directive information from the two spaced lobes of a directional antenna, the ratio of the difference in total field strength of the two lobes to the field strength of the greater lobe at a given point in space.

depth of penetration—The thickness of a layer extending inward from the surface of a conductor and having the same resistance to direct current as the whole conductor has to alternating current of a given frequency.

depth sounder—*See* Fathometer.

de-Q—To reduce the Q of a tuned circuit, as generally applied to carrier-current transmission systems.

derate—To reduce the voltage, current, or power rating of a device to improve its reliability or to permit operation at high ambient temperatures.

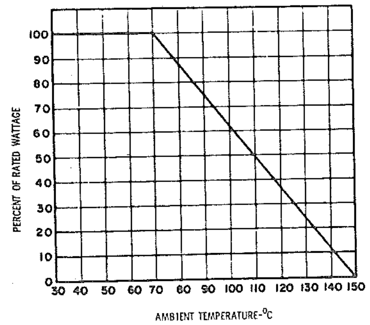

Derating curve.

derating—The reduction in rating of a device or component, especially the maximum power-dissipation rating at higher temperatures.

derating factor — The factor by which the ratings of component parts are reduced to provide additional safety margins in critical applications or when the parts are subjected to extreme environmental conditions for which their normal ratings do not apply.

derivative action—*See* Rate Action.

derivative control — Automatic control in which the rate of correction is determined by the rate at which the error producing it changes.

derived center channel — A monophonic composite signal derived from the sum or difference of the left and right stereo channels, often fed to an extra speaker to fill in an aural "hole" between the left and right speakers. The signal from a voltage-derived center-channel output must be fed to an external power amplifier before it can drive a speaker. A power-derived center channel can drive a speaker directly.

desensitization — 1. The saturation of one component (an amplifier, for instance) by another so that the first cannot perform its proper function. 2. The reduction in receiver sensitivity due to the presence of a high-level off-channel signal overloading the radio-frequency amplifier or mixer stages, causing automatic gain control action.

design-center rating—Values of operating and environmental conditions which should not be exceeded under normal conditions in a bogey electron device.

design for maintainability—Those features and characteristics of design of an item that reduce requirements for tools, test equipment, facilities, spares, highly skilled personnel, etc., and improve the capability of the item to accept maintenance actions.

design-maximum rating—Values of operating and environmental conditions which should not be exceeded under the worst possible conditions in a bogey electron device.

design proof test—A test used to verify that a design specification meets the overall functional requirements of the finished product.

Desk-Fax — Trade name of Western Union Telegraph Co. for a small facsimile transceiver employed principally in short-line telegraph service.

desk-top computer — A computer that requires no formally prepared stored program entered in a special language. Most desk-top computers are operated like adding machines or mechanical calculators.

destination register—In a computer, a register into which data is being placed.

Destriau effect—Sustained emission of light by suitable phosphor powders embedded in an insulator and subjected only to the action of an alternating electric field.

destructive read-out — In a computer, the destruction of data during the process of sensing.

destructive test — Any test resulting in the destruction or drastic deterioration of the test specimen.

detail—A measure of the sharpness of a recorded facsimile copy or reproduced image. Generally related to the number of lines scanned per inch. Defined as the square root of the ratio between the number of scanning lines per unit length and the definition in the direction of the scanning line.

detail contrast—The ratio of the amplitude of the high-frequency components of a video signal to the amplitude of the reference low-frequency component.

detection—*See* Demodulation.

detectophone—An instrument for secretly listening in on a conversation. A high-sensitivity, nondirectional microphone is concealed in the room and connected to an amplifier and headphones or recorder remotely located. Sometimes the microphone feeds into a wired-wireless transmitter that broadcasts over power lines, to permit the listener to be farther away.

detector — 1. A device for effecting the process of detection or demodulation. 2. A mixer or converter in a superheterodyne receiver; often referred to as a "first detector." 3. A device that produces an electrical output that is a measure of the radiation incident on the device. 4. A rectifier tube, crystal, or dry disc by which a modulation envelope on a carrier or the simple on-off state of a carrier may be made to drive a lower-frequency device.

detector balanced bias—A controlling circuit used in radar systems for anticlutter purposes.

detector circuit—That portion of a receiver which recovers the modulation signal from the r-f carrier wave.

detector probe—A probe containing a high-frequency rectifying element such as a crystal diode or a tube. Used with an oscilloscope, vacuum-tube voltmeter, or signal tracer for recovering the modulation from a carrier.

detent—1. A stop or other holding device, such as a pin, lever, etc., on a ratchet wheel. 2. Switch action typified by a gradual increase in force to a position at which there is an immediate and marked reduction in force.

detune—To change the inductance and/or capacitance of a tuned circuit and thereby cause it to be resonant at other than the desired frequency.

deuterium—Heavy hydrogen, so called because it weighs twice as much as ordinary hydrogen. The nucleus of heavy hydrogen is a deuteron.

deuteron—Also called deuton. The nucleus of an atom of heavy hydrogen (deuterium) containing one proton and one neutron. Deuterons are often used as atomic projectiles in atom smashers.

deviation—1. The difference between the actual and specified values of a quantity. 2. In frequency modulation, the difference between the instantaneous frequency during modulation and the carrier frequency. 3. A departure from specification requirements for which approval is obtained from the consumer prior to occurrence of the departure from specification requirements.

deviation absorption — Absorption that occurs at frequencies near the critical frequency. Occurs in conjunction with the slowing up of radio waves near the critical frequency, upon reflection from the ionosphere.

deviation distortion—Distortion caused by inadequate bandwidth, amplitude-modulation rejection, or discriminator linearity in an f-m receiver.

deviation ratio — In frequency modulation, the ratio of the maximum change in carrier frequency to the highest modulating frequency.

deviation sensitivity — The smallest frequency deviation that produces a specified output power in f-m receivers.

device—1. A single discrete conventional electronic device such as a resistor or transistor, or a microelectronic circuit. 2. Also called item. Any subdivision of a system.

device complexity—The number of circuit elements within an integrated circuit.

Dewar flask—A container with double walls. The space between the walls is evacuated, and the surfaces bounding this space are silvered.

dew point—The temperature at which condensation first occurs when a vapor is cooled.

DF—Abbreviation for direction finder or dissipation factor.

DF antenna—Any antenna combination included in a direction finder for obtaining the phase or amplitude reference of the received signal. May be a single or orthogonal loop, an adcock, or spaced differentially-connected dipoles.

DF antenna system—One or more DF antennas and their combining circuits and feeder systems, together with the shielding and all electrical and mechanical items up to the receiver input terminals.

Diac — Two-lead alternating-current switch semiconductor.

diagnostic routine—An electronic-computer routine designed to locate a malfunction in the computer, a mistake in coding, or both.

diagnotor — In a computer, combined diagnostic and edit routine that questions unusual situations and makes note of the implied results.

diagram—1. Schematics, prints, charts, etc., or any other graphical representation, the purpose of which is to explain rather than to represent. 2. A schematic representation of a sequence of subroutines designed to solve a problem. 3. A coarser and less symbolic representation than a flowchart, frequently including descriptions in English words. 4. A schematic or logical drawing showing the electrical circuit or logical arrangements within a component.

dial—1. A means for indicating the value to which a control knob has been set. 2. A calling device which generates the required number of pulses in a telephone set and thereby establishes contact with the party being called.

dial cable—*See* Dial Cord.

dial cord—Also called dial cable. A braided cord or flexible wire cable connected to a tuning knob so that turning the knob will move the pointer or dial which indicates the frequency to which a radio receiver is tuned. Also used for coupling two shafts together mechanically.

dial light—A small pilot lamp which illuminates the tuning dial of a radio receiver.

dial pulse—Interruption in the direct current flowing through the loop of a calling telephone, produced by the opening and closing of the dial pulse springs of a calling telephone in response to the dialing of a digit. The current in the calling-line loop is interrupted as many times as there are units in the digit dialed—i.e., dialing of the digit 7 generates seven dial pulses (interruptions) in current flowing through the loop of the calling telephone.

dial register — *See* Standard Register of a Motor Meter.

dial tone—A hum or other tone employed in a dial telephone system to indicate that the line is not busy and that the equipment is ready for dialing.

diamagnetic — Term applied to a substance with a negative magnetic susceptibility.

diamagnetic material—A material which is less magnetic than air, or in which the intensity of magnetization is negative. There is no known material in which this effect has more than a very feeble intensity. Bismuth is the leading example of materials of this class.

diamond antenna—Also called a rhombic antenna. A horizontal antenna having four conductors that form a diamond, or rhombus.

diamond lattice — The crystal structure of germanium and silicon (as well as a diamond).

diamond stylus—A phonograph pickup with a ground diamond as its point.

diapason — The unique fundamental tone color of organ music.

diaphragm—1. A flexible membrane used in various electroacoustic transducers for producing audio-frequency vibrations when actuated by electric impulses, or electric impulses when actuated by audio-frequency vibrations. (*Also see* Iris.) 2. In electrolytic cells, a porous or permeable membrane, usually flexible, separating the anode and cathode compartments. 3. In waveguide technique, a thin plate, or plates, placed transversely across the waveguide, not completely closing it, and usually introducing a reactance component.

diathermy — Therapeutic heating of tissues beneath the skin by applying an r-f current to some part of the body.

diathermy interference — A form of television interference caused by diathermy equipment, resulting in a horizontal herringbone pattern across the picture.

diathermy machine — A medical apparatus consisting of an r-f oscillator frequently followed by r-f amplifier stages, used to generate high-frequency currents that produce heat within some predetermined part of the body for therapeutic purposes.

dichroic mirror—A special mirror through which all light frequencies pass except those for the color which the mirror is designed to reflect.

dichroism—A property of an optical material which causes light of some wavelengths to be absorbed when the incident light has its electric-field vector in a particular orientation and not absorbed when the electric-field vector has other orientations.

dicing—The process of sawing a crystal wafer into blanks.

die—*See* Chip.

die bonding — A bonding process in which chips are soldered to the substrate.

dielectric — 1. The insulating (nonconducting) medium between the two plates of a capacitor. Typical dielectrics are air, wax-impregnated paper, plastic, mica, and ceramic. A vacuum is the only perfect dielectric. 2. A medium capable of recovering, as electrical energy, all or part of the energy required to establish an electric field (voltage stress). The field, or voltage stress, is accompanied by displacement or charging currents.

dielectric absorption—Also called dielectric hysteresis or dielectric soak. A characteristic of dielectrics which determines the length of time a capacitor takes to deliver the total amount of its stored energy. It manifests itself as the reappearance of a potential on the electrodes after the capacitor has been discharged. Its magnitude depends on the charge and discharge time of the capacitor.

dielectric amplifier—An amplifier employing a device similar to an ordinary capacitor, but with a polycrystalline dielectric which exhibits a ferromagnetic effect.

dielectric antenna—An antenna in which a dielectric is the major component producing the required radiation pattern.

dielectric breakdown — An abrupt increase in the flow of electric current through a dielectric material as the applied electric field strength exceeds a critical value.

dielectric breakdown voltage—The voltage between two electrodes at which electric breakdown of the specimen occurs under prescribed test conditions. Also called electric breakdown voltage, breakdown voltage, or hi-pot.

dielectric capacity—The inductivity or specific inductive capacity of a substance, being its ability to convey the influence of an electrified body.

dielectric constant—The ratio of the capacitance of a capacitor with the given dielectric to the capacitance of a capacitor having air for its dielectric but otherwise identical. Also called permittivity, specific inductive capacity, or capacitivity.

dielectric current—The current flowing at any instant through the surface of an isotropic dielectric which is in a changing electric field.

dielectric dissipation—*See* Loss Tangent.

dielectric dissipation factor—The cotangent of the dielectric phase angle of a material.

dielectric fatigue—The property of some dielectrics in which the insulating quality decreases after a voltage has been applied for a considerable length of time.

dielectric guide—A waveguide made of a solid dielectric material through which the waves travel.

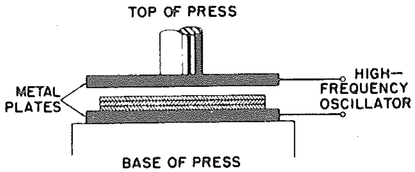

Dielectric heating.

dielectric heating—A method of raising the temperature of a nominally insulating material by sandwiching it between two plates to which an r-f voltage is applied. The material acts as a dielectric, and its internal losses cause it to heat up.

dielectric hysteresis—*See* Dielectric Absorption.

dielectric isolation—The electrical isolation of monolithic integrated circuit elements from each other by dielectric material rather than by reverse biased P-N junctions.

dielectric lens—A lens used with microwave antennas; it is made of dielectric material so that it refracts radio waves similar to the way an optical lens refracts light waves.

dielectric loss—The power dissipated by a dielectric as the friction of its molecules opposes the molecular motion produced by an alternating electric field.

dielectric loss angle — The complement of the dielectric phase angle (i.e., the dielectric phase angle minus 90°).

dielectric loss factor — The product of the dielectric constant of a material times the tangent of the dielectric loss angle. Also called dielectric loss index.

dielectric loss index — *See* Dielectric Loss Factor.

dielectric matching plate — In waveguide technique, a dielectric plate used as an impedance transformer for matching purposes.

dielectric mirror—A highly frequency-selective, multilayer dielectric reflector acting by partial reflection of light at the interface between materials of unequal refractive indices.

dielectric phase angle—The angular difference in phase between the sinusoidal alternating voltage applied to a dielectric and the component of the resultant alternating current having the same period.

dielectric phase difference — *See* Dielectric Loss Angle.

dielectric polarization—*See* Polarization, 3.

dielectric power factor—The cosine of the dielectric phase angle.

dielectric rating—Standard test voltages and frequencies above which failure occurs between specified points in a relay structure.

dielectric soak—*See* Dielectric Absorption.

dielectric strength—The maximum voltage a dielectric can withstand without rupturing. Usually expressed as volts per mil. Also called: electric strength, breakdown strength, and electric field strength.

dielectric susceptibility—The ratio of the polarization in a dielectric to the electric intensity responsible for it.

dielectric tests—1. Tests which consist of the application of a voltage higher than the rated voltage for a specified time for the purpose of determining the adequacy against breakdown of insulating materials and spacings under normal conditions. 2. The testing of insulating materials by the application of a constantly increasing voltage until failure occurs.

dielectric waveguide—A waveguide constructed from a dielectric (nonconductive) substance.

difference amplifier—*See* Differential Amplifier.

difference channel—In a stereophonic sound system, an audio channel that handles the difference between the signals in the left and right channels.

difference detector—A detector circuit in which the output is a function of the difference between the peak or rms amplitudes of the input waveforms.

difference frequency—1. A signal representing, in essence, the difference between the left and right sound channels of a stereophonic sound system. 2. One of the output frequencies of a converter. It is the difference between the two input frequencies.

difference in depth modulation—In directive systems employing overlapping lobes with modulated signals, a ratio obtained by subtracting from the percentage of modulation of the larger signal the percentage of modulation of the smaller signal and dividing by 100.

difference of potential—The voltage or electrical pressure existing between two points. It will result in a flow of electrons whenever a circuit is established between the two points.

differential — 1. A planetary gear system which adds or subtracts angular movements transmitted to two components and delivers the answer to a third. Widely used for adding and subtracting shaft movements in servo systems and for addition and subtraction in computing machines. 2. In electronics, the difference between two levels.

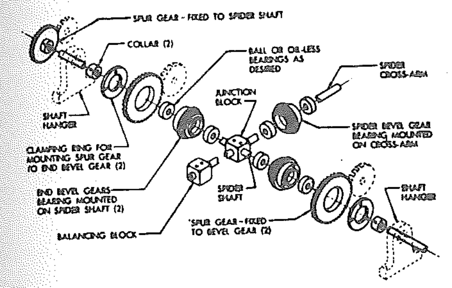
SPUR GEAR - FIXED TO SPIDER SHAFT
COLLAR (2)
BALL OR OIL-LESS BEARINGS AS DESIRED
SPIDER CROSS-ARM
JUNCTION BLOCK
SHAFT HANGER
SPIDER BEVEL GEAR BEARING MOUNTED ON CROSS-ARM
CLAMPING RING FOR MOUNTING SPUR GEAR TO END BEVEL GEAR (2)
END BEVEL GEARS BEARING MOUNTED ON SPIDER SHAFT (2)
SPIDER SHAFT
SHAFT HANGER
SPUR GEAR - FIXED TO BEVEL GEAR (2)
BALANCING BLOCK

Differential.

differential amplifier—An amplifier having two similar input circuits so connected that they respond to the difference between two voltages or currents, but effectively suppress like voltages or currents.

differential angle—The total angle from the operation to the releasing position in a mercury switch.

differential analyzer — 1. A mechanical or electrical device primarily designed and used to solve differential equations. 2. An analog computer designed primarily for solving or analyzing many types of differential equations.

differential capacitance — The derivative with respect to voltage of a capacitor charge characteristic at a given point on the curve.

differential capacitance characteristic — The function that relates differential capacitance to voltage.

differential capacitor—A variable capacitor having two similar sets of stator plates and one set of rotor plates. When the rotor is turned, the capacitance of one section is increased while the capacitance of the other section is decreased.

differential cooling—A lowering of temperature which takes place at a different rate at various points on an object or surface.

differential delay — The difference between the maximum and the minimum frequency delays occurring across a band.

differential discriminator—A discriminator that passes only pulses having amplitudes between two predetermined values, neither of which is zero.

differential duplex system—A duplex system in which the sent currents divide through two mutually inductive sections of the receiving apparatus. These sections are connected respectively to the line and to a balancing artificial line in opposite directions. Hence, there is substantially no net effect on the receiving apparatus. The received currents pass mainly through one section, or through the two sections in the same direction, and operate the apparatus.

differential flutter — Speed-change errors which occur at different magnitudes, frequencies, or phases across the width of a magnetic tape.

differential gain—The ratio of the differential output signal of a differential amplifier divided by the differential input signal causing that output.

differential gain control—Also called gain-sensitivity control. A device for altering the gain of a radio receiver in accordance with an expected change of signal level, in order to reduce the amplitude differential between the signals at the receiver output.

differential galvanometer—A galvanometer having two similar but opposed coils, so that their currents tend to neutralize each other. A zero reading is obtained when the currents are equal.

differential gap—The difference between two target values, one of which applies to an upswing of conditions and the other to a downswing.

differential generator—A synchro differential generator driven by a servo system.

differential impedance — The internal impedance observed between the input terminals of an operational amplifier.

differential input—An input circuit that rejects voltages which are the same at both input terminals and amplifies the voltage difference between the two input terminals. May be either balanced or floating and may also be guarded.

differential-input amplifier — An amplifier in which the output is ideally a function only of the difference between the signals applied to its two inputs, both signals being measured with respect to a common "low" or "ground" reference point.

differential-input capacitance—The capacitance between the inverting and noninverting input terminals.

differential-input impedance — The impedance between the inverting and noninverting input terminals of a differential amplifier.

differential-input measurement — Also called floating input. A measurement in which the two inputs to a differential amplifier are connected to two points in a circuit under test and the amplifier displays the difference voltage between the points. In this type of measurement, each input of the amplifier acts as a reference for the other, and ground connections are used only for safety reasons.

differential-input resistance — The resistance between the inverting and noninverting input terminals of a differential amplifier.

differential-input voltage — The maximum voltage that can be applied across the input terminals of a differential amplifier without damaging the amplifier.

differential-input voltage range—The range of voltages that may be applied between input terminals without forcing the circuit to operate outside its specifications.

differential-input voltage rating – The maximum allowable signal that may be applied between the inverting and noninverting inputs of a differential amplifier without damaging the amplifier.

differential instrument—A galvanometer or other measuring instrument having two circuits or coils, usually identical, through which currents flow in opposite directions. The difference or differential effect of these currents actuates the indicating pointer.

differential microphone—*See* Double-Button Carbon Microphone.

differential-mode signal—A signal that is applied between the two ungrounded terminals of a balanced three-terminal system.

differential modulation—A type of modulation in which the choice of the significant condition for any signal element is dependent on the choice for the previous signal element.

differential permeability—The ratio of the positive increase of normal induction to the positive increase of magnetizing force when these increases are minute.

differential phase—The difference in phase shift through a television system for a small, high-frequency sine-wave signal at two stated levels of a low-frequency signal on which the first signal is superimposed.

differential phase-shift keying—A modulation scheme in which the information is conveyed by changes in carrier phase during one interval relative to the preceding interval.

differential pressure – The difference in pressure between two pressure sources.

differential pressure transducer – A pressure transducer that accepts simultaneously two independent pressure sources, and the output of which is proportional to the pressure difference between the sources.

differential protective relay – A protective device which functions on a percentage or phase angle or other quantitative difference of two currents or of some other electrical quantities.

differential relay – A relay with multiple windings that functions when the voltage, current, or power difference between the windings reaches a predetermined value. The power difference may result from the algebraic addition of the multiple inputs.

differential selsyn—A selsyn in which both the rotor and the stator have similar windings that are spread 120° apart. The position of the rotor corresponds to the algebraic sum of the fields produced by the stator and rotor.

differential synchro—*See* Synchro Differential Generator *and* Synchro Differential Motor.

differential transducer—A device capable of simultaneously measuring two separate stimuli and providing an output proportionate to the difference between them.

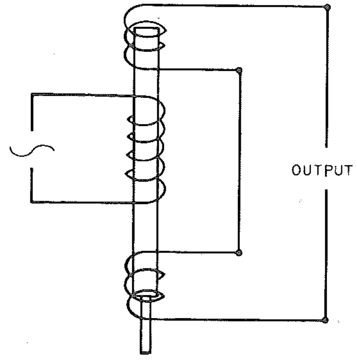

Differential transformer.

differential transformer—Also called linear variable-differential transformer. 1. A transformer used to join two or more sources of signals to a common transmission line. 2. An electromechanical device which continuously translates displacement or position change into a linear a-c voltage.

differential voltage gain—The ratio of the change in output signal voltage at either terminal or a differential device to the change in signal voltage applied to either input terminal, all voltages measured to ground.

differential winding—A coil winding so arranged that its magnetic field opposes that of a nearby coil.

differential-wound field—A type of motor or generator field having both series and shunt coils connected so they oppose each other.

differentiating circuit—Also called differentiator.

differentiating network—*See* Differentiator.

differentiator—1. An analog computer device which produces an output proportional to the derivative of the one variable with respect to another, usually time. 2. A transducer or circuit the output waveform of which is the time derivative of its input waveform. Such a transducer or circuit preceding a frequency modulator makes the combination a phase modulator, or following a phase detector makes the combination a frequency detector. Its ratio of output amplitude to input amplitude is proportional to the frequency, and its output phase leads its input phase by 90°. 3. A circuit which produces an output voltage substantially in proportion to the rate of change of the input voltage or current. Differentiating circuits employ time constants that are short compared to the duration of the pulse applied, thus differentiating the input pulse.

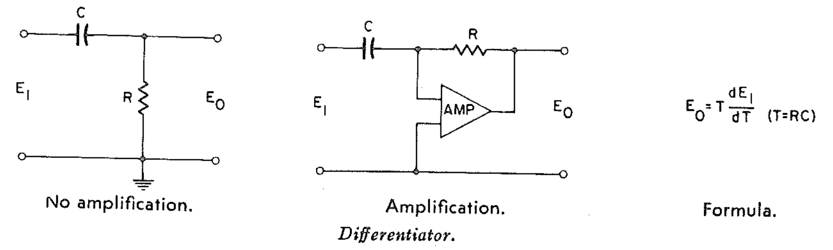

No amplification.

Amplification.

Formula.

Differentiator.

diffracted wave—A radio, sound, or light wave which has struck an object and been bent or deflected, other than by reflection or refraction.

diffraction—1. The bending of radio, sound, or light waves as they pass through an object or barrier, thereby producing a diffracted wave. 2. The phenomenon whereby waves traveling in straight paths bend around an obstacle.

diffused-alloy transistor—Also called drift transistor. A transistor in which the semiconductor wafer is subjected to gaseous diffusion to produce a nonuniform base region, after which alloy junctions are formed in the same manner as for an alloy-junction transistor. It may also have an intrinsic region to give a PNIP unit.

diffused-base transistor—Also called graded-base transistor. A type of transistor made by combining diffusion and alloy techniques. A nonuniform base region and the collector-to-base junction are formed by gaseous dissemination into a semiconductor wafer that constitutes the collector region. Then the emitter-to-base junction is formed by a conventional alloy process on the base side of the diffused wafer.

diffused device—A semiconductor device in which a base, usually of silicon, has successive layers of P and N characteristics diffused upon and into the base by means of a series of masks and around which P and N materials, usually phosphorus and boron, adhere to the base by gaseous diffusion in a high-temperature furnace. It is possible to build areas of resistance, capacitance, and "active" diodes and transistors into the base, creating an entire circuit. Performance is poor in the presence of radiation.

diffused-emitter-and-base transistor—Also called double-diffused transistor. A semiconductor wafer which has been subjected to gaseous dissemination of both N- and P-type impurities to form two PN junctions in the original semiconductor material.

diffused-emitter-and-collector transistor—A transistor in which both the emitter and collector are made by a dissemination process.

diffused junction—A junction formed by the dissemination of an impurity within a semiconductor crystal.

diffused-junction rectifier—A semiconductor diode in which the PN junction is produced by diffusion.

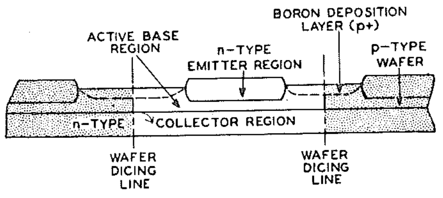

Diffused-junction silicon transistor wafer.

diffused-junction transistor—A transistor in which the emitter and collector electrodes have been formed by diffusion of an impurity into the semiconductor wafer without heating.

diffused-layer resistor – A resistor formed by including an appropriate pattern in the photomask to define diffusion areas.

diffused mesa transistor – A transistor in which the collector-base junction is formed by gaseous diffusion, and the emitter-base junction is formed either by gaseous diffusion or by an evaporated metal strip. The collector-base junction is then defined by etching away the undesired parts of the emitter and base regions, thus producing a mesa.

diffused planar transistor – A transistor made by two gaseous diffusions, but in which the collector-base junction is defined by oxide masking. Junctions are formed beneath this protective oxide layer with the result that the device has lower reverse currents and good d-c gain at low currents.

diffused sound—Sound which has uniform energy density, meaning the energy flux is equal in all parts of a given region.

diffused transistor – A transistor in which the emitter and collector junctions are both formed by diffusion.

diffused transmission—The total net transmission, by a medium or device, of light that is neither perfectly Lambertian nor parallel. Often used interchangeably with the term "gross transmission."

diffusion—1. The movement of carriers from a region of high concentration to regions of lower concentration. 2. Migration of atoms from one substance into another substance.

diffusion capacitance—The capacitance of a forward-biased PN junction.

diffusion constant—The quotient of diffusion-current density in a homogeneous semiconductor, divided by the charge-carrier concentration gradient. It is equal to the drift mobility times the average thermal energy per unit charge of carriers.

diffusion current – The current produced when charges move by diffusion.

diffusion length—In a homogeneous semiconductor, the average distance the minority carriers move between generation and recombination.

diffusion process—The method of producing junctions by disseminating acceptors or donors into a semiconductor at a high temperature.

diffusion transistor—A transistor relying on the dissemination of carriers, donors, or acceptors for carrying current.

digiralt — A system of high-resolution radar altimetry in which pulse-modulated radar and high-performance time-to-digital conversion techniques are combined.

digit—One of the symbols, 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9, used in numbering in the scale of 10. One of these symbols, when used in a scale of numbering to the base *n*, expresses integral values ranging from *0* to *n—1* inclusive.

digital — Using numbers expressed in digits and in a certain scale of notation to represent all the variables that occur in a problem.

digital circuit—A circuit which operates like a switch—i.e., it is either "on" or "off."

digital communications—A system of telecommunications employing a nominally discontinuous signal that changes in frequency, amplitude, or polarity.

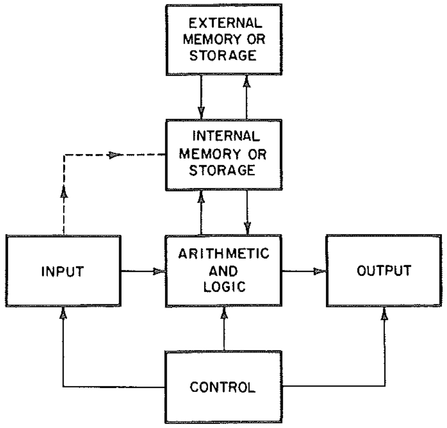

Block diagram of digital computer.

digital computer—An electronic calculator that operates with numbers expressed directly as digits, as opposed to the directly measurable quantities (voltage, resistance, etc.) in an analog computer. In other words, the digital computer counts (as does an adding machine); the analog computer measures a quantity (as does a voltmeter).

digital data — Data represented in discrete, discontinuous form, as contrasted with analog data represented in continuous form. Digital data are usually represented by means of coded characters (e.g., numbers, signs, symbols, etc.).

digital data-handling system — The electronic equipment which receives digital data, operates on it in a suitable manner, records it in a suitable manner, records it on a suitable medium, and presents it directly to a computer or a display.

digital information display—The presentation of digital information in tabular form on the face of a digital information display tube.

digital output—An output signal which represents the size of a stimulus or input signal in the form of a series of discrete quantities which are coded to represent digits in a system of numerical notation. This type of output is to be distinguished from one which provides a continuous rather than a discrete output signal.

digital phase shifter—A device which provides a signal phase shift by the application of a control pulse. A reversal of phase shift requires a control pulse of opposite polarity.

digital readout indicator—An indicator that reads directly in numerical form, as opposed to an analog indicator needle and scale.

digital rotary transducer—A rotating device utilizing an optical sensor that produces a serial binary output as a result of shaft rotation.

digital signals — Discrete or discontinuous signals whose various states are discrete intervals apart.

digital-to-analog converter — A computing device that changes digital quantities into physical motion or into a voltage (i.e., a number output into turns of a potentiometer).

digital voltmeter—An indicator which provides a digital readout of measured voltage rather than a pointer indication.

digit compression — In a computer, any of several techniques used to pack digits.

digitize—To convert an analog measurement of a physical variable into a number expressed in digits in a scale or notation.

digitizer—A device which converts analog data into numbers expressed in digits in a system of notation.

digit selector — In a computer, a device for separating a card column into individual pulses that correspond to punched row positions.

digit-transfer bus—The main wire or wires used to transfer information (but not control signals) among the various registers in a digital computer.

diheptal base—Also called diheptal socket. A vacuum-tube base having 14 pins (such as the base of a cathode-ray tube).

diheptal socket—*See* Diheptal Base.

dimensional stability—The ability of a body to maintain precise shape and size.

dimmer—A control for varying the intensity of illumination of a pilot, scale, or dial light.

D-indicator—A radar indicator which combines Types B and C indicators. The signal appears as a bright spot, with azimuth angle as the horizontal coordinate and elevation angle as the vertical coordinate. Each horizontal trace is expanded vertically by a compressed time sweep to facilitate separation of the signal from noise and to give a rough range indication.

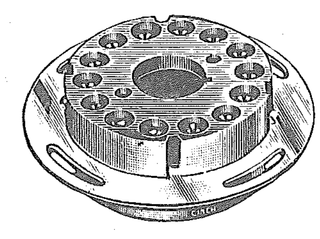

Diheptal socket.

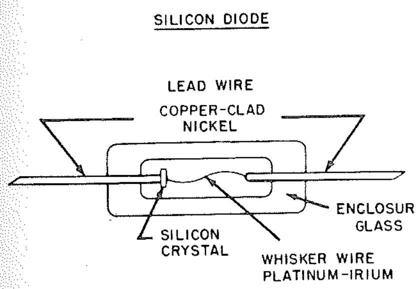

Diode.

diode—A two-terminal device which will conduct electricity more easily in one direction than in the other.

diode amplifier—A parametric amplifier that uses a special diode in a cavity. Used to amplify signals at frequencies as high as 6,000 MHz.

diode assembly—A single structure of more than one diode.

diode characteristic—The composite electrode characteristic of a multielectrode tube, taken with all electrodes except the cathode connected together.

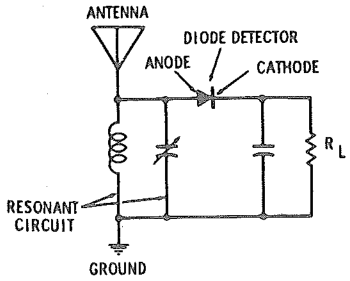

Diode detector.

diode detector—A diode used in a demodulation, or detection, circuit.

diode gate—An AND gate that uses diodes as switching elements.

diode laser — A PN junction semiconductor electron device which converts direct forward-bias electrical input (pump power) directly into coherent optical output power via a process of stimulated emission in the region near the junction. Called laser diodes, injected lasers, coherent electroluminescence devices, semiconductor lasers.

diode limiter—A circuit employing a diode and used to prevent signal peaks from exceeding a predetermined value.

diode mixer—A diode which mixes incoming radio-frequency and local-oscillator signals to produce an intermediate frequency.

diode peak detector—A diode used in a circuit to indicate when audio peaks exceed a predetermined value.

diode-pentode — A vacuum tube having a diode and a pentode combined in the same envelope.

diode rectification—The conversion of an alternating current into a unidirectional current by means of a two-element device such as a crystal, vacuum tube, etc.

diode switch—A diode in which positive and negative biasing voltages (with respect to the cathode) are applied in succession to the anode in order to pass and block, respectively, other applied waveforms within certain voltage limits. In this way, the diode acts as a switch.

diode transistor logic — Abbreviated DTL. A logic circuit that uses diodes at the input to perform the electronic logic function that activates the circuits transistor output. In monolithic circuits, the DTL diodes are a positive level logic AND function or a negative level OR function. The output transistor acts as an inverter to result in the circuit becoming a positive NAND or a negative NOR function.

diode-triode—A vacuum tube having a diode and triode combined in the same envelope.

diopter—The unit of optical measurement which expresses the refractive power of a lens or prism.

dip—1. A drop in the plate current of a class-C amplifier as its tuned circuits are being adjusted to resonance. 2. The angle between the direction of the earth's magnetic field and the horizontal as measured in a vertical plane.

dip coating—1. A method of applying an insulating coating to a conductor by passing it through an applicator containing the insulating medium in liquid form. The insulation is then sized and passed through ovens to solidify. 2. The insulation so applied.

dip encapsulation — A type of conformal coating—An embedding process in which the insulating material is applied by immersion and without the use of an outer container. The coating conforms generally with the contour of the embedding part or assembly.

diplexer—A coupling unit which allows more than one transmitter to operate together on the same antenna.

diplex operation—The simultaneous transmission or reception of two messages from a single antenna or on a single carrier.

diplex radio transmission — Simultaneous

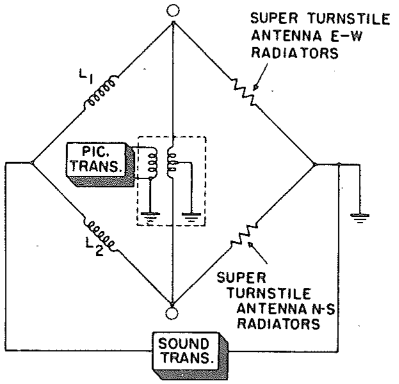
SUPER TURNSTILE ANTENNA E-W RADIATORS
L1
L1
PIC. TRANS.
L2
L2
SUPER TURNSTILE ANTENNA N-S RADIATORS
SOUND TRANS.

Diplexer circuit.

transmission of two signals by using a common carrier wave.

diplex reception—The simultaneous reception of two signals having some feature in common—i.e., a single receiving antenna or a single carrier frequency.

dipole—1. *See* Dipole Antenna. 2. A molecule which has an electric moment. For a molecule to be a dipole, the effective center of the positive charges must be at a different point from the center of the negative charges.

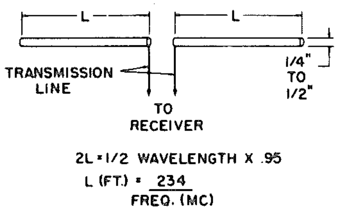

Dipole antenna.

dipole antenna—Also called dipole. A straight radiator usually fed in the center. Maximum radiation is produced in the plane normal to its axis. The length specified is the overall length.

dipole disc feed—An antenna, consisting of a dipole near a disc, used to reflect energy to the disc.

dipping — The process of impregnating or coating insulating materials or windings by the simple method of immersion in the liquid insulating material. A step in the process of treating insulating materials or electrical components by immersion in a liquid insulation, followed by draining and curing to provide increased electrical and mechanical protection.

dip soldering—The process of soldering component leads, terminals, and hardware to the conductive pattern on the "bottom" of a printed-circuit board by dipping that side into molten solder or floating it on the surface.

direct-acting recording instrument—An instrument in which the marking device is mechanically connected to or directly operated by the primary detector.

direct capacitance—The capacitance between two conductors excluding stray capacitance that may exist between the two conductors and other conducting elements.

direct-coupled amplifier — A direct-current amplifier in which the plate of one stage is coupled to the grid of the next stage by a direct connection or a low-value resistor.

direct coupling—The association of two or more circuits by means of an inductance, a resistance, a wire or a combination of these so that both direct and alternating currents can be coupled.

direct current—Abbreviated dc. An essentially constant-value current that flows in only one direction.

direct-current amplifier—Also called d-c amplifier. An amplifier capable of boosting d-c voltages. Resistive coupling only is generally employed between stages, but sometimes will be combined with other forms.

direct-current erasing head—A head which uses direct current in magnetic recording to produce the magnetic field required for erasure. Direct-current erasing is achieved by subjecting the medium to a unidirectional field. Such a medium is therefore in a different magnetic state than one erased by alternating current.

direct-current generator — A rotating machine that changes mechanical into electrical energy in the form of direct current. This is accomplished by commutating bars on the armature. The bars make contact with stationary brushes from which the direct-current is taken.

direct-current restorer — The means by which a direct current or low-frequency component is reinserted after transmission. Used in a circuit incapable of transmitting slow variations, but capable of transmitting components of higher frequency.

direct-current transmission—Of television, that form of transmission in which a fixed setting of the controls makes any instantaneous value of signal correspond to the same value of brightness at all times.

direct digital control — Time sharing, or multiplexing, of a computer among many controlled loops.

direct-drive torque motor—A servoactuator which can be directly attached to the load it is to drive. It converts electrical signals directly into sufficient torque to maintain the desired accuracy in a positioning or speed control system.

direct electromotive force—A unidirectional electromotive force in which the changes in values are either zero or so small that they may be neglected.

direct grid bias—The d-c component of grid bias (voltage). Commonly called grid bias.

direct-insert subroutine—*See* Open Subroutine.

direction—The position of one point in space with respect to another.

directional antenna—An antenna which radiates radio waves more effectively in some directions than in others. (The term is usually applied to an antenna the directivity of which is larger than that of a half-wave dipole.)

directional beam—An electromagnetic wave that is concentrated in a given direction.

directional coupler—A junction consisting of two waveguides coupled together in such a manner that a traveling wave in either guide will induce a traveling wave in the same direction in the other guide.

directional filter — A filter used to separate the two frequency ranges in a carrier system where one range of frequencies is used for transmission in one direction and another range of frequencies for transmission in the opposite direction. Also called directional separation filter.

directional gain—*See* Directivity Index.

directional homing—The procedure of following a path in such a way that the target is maintained at a constant relative bearing.

directional hydrophone—A hydrophone having a response that varies significantly with the direction of incidence of sound.

directional lobe—*See* Lobe.

directional microphone—A microphone the response of which varies significantly with the direction of sound. (*Also see* Unidirectional, Bidirectional, *and* Semidirectional Microphone.)

directional pattern—Also called radiation pattern. A graphical representation of the radiation or reception of an antenna as a function of direction. Cross sections are frequently given as vertical and horizontal planes, and principal electric and magnetic polarization planes.

directional power relay — A device which functions on a desired value of power flow in a given direction, or upon reverse power resulting from arcback in the anode or cathode circuits of a power rectifier.

directional relay—A relay which functions in conformance with the direction of power, voltage, current, pulse rotation, etc. *Also see* Polarized Relay.

directional separation filter—*See* Directional Filter.

direction angle—The angle between the antenna base line and a line connecting the center of the base line with the target.

direction cosine — The cosine of the angle between the base line and the line from the center of the base line to the target.

direction finder — Abbreviated DF. Also called radio compass. Apparatus for receiving radio signals and taking their bearings in order to determine their point of origin.

direction finding—The principle and practice of determining a bearing by radio means, using a discriminating antenna system and a radio receiver so that the direction of an arriving wave, and ostensibly the direction or bearing of a distant transmitter, can be determined.

direction of lay—The lateral direction in which strands or the elements of a cable are wound over the top of the cable. Expressed as right- or left-hand lay, viewed as they recede from the observer.

direction of polarization—For a linearly polarized wave, the direction of the electrostatic field.

direction of propagation—At any point in a homogeneous, isotropic medium, the direction of the time-average energy flow. In a uniform waveguide, the direction of propagation is often taken along the axis. In a uniform lossless waveguide, the direction of propagation at every point is parallel to the axis and in the direction of time-average energy flow.

direction rectifier — A rectifier which supplies a direct-current voltage, the magnitude and polarity of which are determined by the magnitude and relative polarity of an alternating-current selsyn error voltage.

directive gain—In a given direction, 4π times the ratio of the radiation intensity to the total power radiated by the antenna.

directivity diagram of an antenna — The graphical representation of the gain of an antenna in the different directions of space.

directivity factor—1. In acoustics, the directivity factor is equivalent to directivity, as applied to an antenna. 2. Of a transducer used for sound emission, the ratio of the intensity of the radiated sound at a remote point in a free field on the principal axis to the average intensity of the sound transmitted through a sphere passing through the remote point and concentric with the transducer. 3. Of a transducer used for sound reception, the ratio of the square of the electromotive force produced in response to sound waves arriving in a direction parallel to the principal axis of the transducer to the mean square of the electromotive force that would be produced if sound waves having the same frequency and mean square pressure were arriving at the transducer simul-

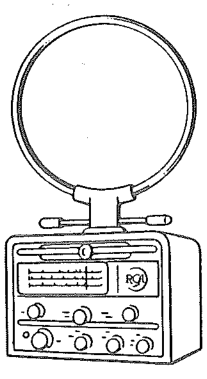

Direction finder.

taneously from all directions with random phase. The frequency should be specified in both cases.

directivity index—Also called directional gain. A measure of the directional properties of a transducer. It is the ratio, in decibels, of the average intensity or response over the whole sphere surrounding the projector or hydrophone to the intensity or response on the acoustic axis.

directivity of a directional coupler—Ratio of the power measured at the forward-wave sampling terminals with only a forward wave present in the transmission line, to the power measured at the same terminals when the forward wave reverses direction. This ratio is usually expressed in db and would be infinite for a perfect coupler.

directivity of an antenna—The ratio of the maximum field intensity to the average field intensity at a given distance, implying a maximum value.

directivity signal—A spurious signal present in the output of any coupler because its directivity is not infinite.

direct light—Light from a luminous object such as the sun or an incandescent lamp, as opposed to reflected light.

directly grounded—*See* Solidly Grounded.

directly heated cathode—A wire, or filament, designed to emit the electrons that flow from cathode to plate. This is done by passing a current through the filament; the current heats the filament to the point where electrons are emitted. In an indirectly-heated cathode, the hot filament raises the temperature of a sleeve around the filament; the sleeve then becomes the electron emitter.

director—1. A parasitic antenna element located in the general direction of the major lobe of radiation for the purpose of increasing the radiation in that direction. 2. Equipment in common-carrier telegraph-message switching systems, used to make cross-office selection and connection from an input-line to an output-line equipment in accordance with addresses in the message. 3. A telephone switch which translates the digits dialed into the directing digits actually used to switch the call. 4. Electromechanical equipment which is used to track a moving target in azimuth and angular height, and which, with the addition of other necessary information from an outside source such as a radar set or a range finder, continuously computes firing data and transmits them to the guns.

direct pickup—Transmission of television images without resorting to an intermediate magnetic or photographic recording.

direct piezoelectricity—A name sometimes given to the piezoelectric effect in which an electric charge is developed on a crystal by the application of mechanical stress.

direct-radiator speaker—A speaker in which the radiating element acts directly on the air instead of relying on any other element such as a horn.

direct recording—The production of a visible record without subsequent processing, in response to received signals.

direct-recording magnetic tape—A method of recording using a high-frequency bias in which the electrical input signal is applied to the recording head without alteration.

direct resistance-coupled amplifier—An amplifier in which the plate of one stage is connected either directly or through a resistor to the control grid of the next stage, with the plate-load resistor being common to both stages. Used to amplify small changes in direct current.

direct scanning—In this method, the entire subject is illuminated continuously but the television camera views only one portion of it at a time.

direct sound wave—A wave in an enclosure emitted from a source prior to the time it has undergone its first reflection from a boundary of the enclosure. Frequently a sound wave is said to be direct if it contains reflections that have occurred from surfaces within about 0.05 second after the sound was first emitted.

direct voltage—Also called d-c voltage. A voltage that forces electrons to move through a circuit in the same direction and thereby produce a direct current.

direct wave—A wave that is propagated directly through space, as opposed to one that is reflected from the sky or ground.

direct Wiedemann effect—*See* Wiedemann Effect.

direct-wire circuit—A supervised protective signaling circuit usually consisting of one metallic conductor and a ground return and having signal-receiving equipment responsive to either an increase or a decrease in current.

direct writing recorder—A strip-chart recorder which produces a readable record without further processing.

disc—1. A phonograph record. 2. The blank used in a recorder.

disc capacitor—A small disc-shaped capacitor with a ceramic dielectric, generally used for bypassing or for temperature compensation in tuned circuits.

disc files—A type of storage medium consisting of numbers of discs that rotate; each disc has a special coating for retaining stored information.

disc generator—A capacitive-charge type of voltage generator.

discharge—1. In a storage battery, the conversion of chemical energy into electric energy. 2. The release of energy stored in a capacitor when a circuit is connected between its terminals.

discharge breakdown—Breakdown of a material as a result of degradation due to gas discharges.

discharge key—A device for switching a capacitor suddenly from a charging circuit to a load through which it can discharge.

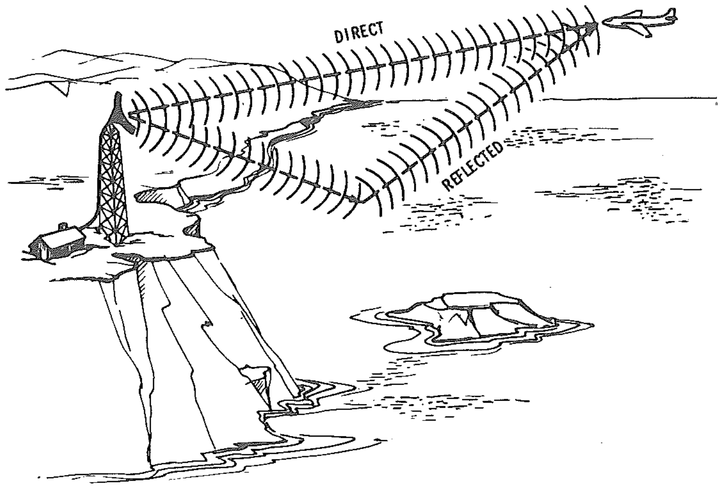

Direct waves.

discharge lamp—A lamp containing a low-pressure gas or vapor which ionizes and emits light when an electric discharge is passed through it. Fluorescent materials are sometimes used on the inside of the glass envelope to increase the illumination, as in an ordinary fluorescent lamp.

discharge rate—The amount of current a battery will deliver over a given period of time. A slower discharge rate generally results in more efficient use of a battery.

discharge tube—A tube containing a low-pressure gas which passes a current whenever sufficient voltage is applied.

discone antenna—A special form of biconical antenna in which the vertex angle of one cone is 180°.

disconnect—1. To break an electric circuit. 2. To remove the power from an electrical device (colloquially, "to unplug the device"). 3. Also called release. To disengage the apparatus used in a connection and to restore it to its ready condition when not in use.

disconnecting means—A device whereby the current-carrying conductors of a circuit can be disconnected from their source of supply.

disconnect switch (motor circuit switch)—A switch intended for use in a motor branch circuit. It is rated in horsepower, and is capable of interrupting the maximum operating overload current of a motor of the same rating at the rated voltage.

disc pack—A set of magnetic discs that can be removed from a disc storage as one unit.

disc recorder—A recording device in which the sounds are mechanically impressed onto a disc—as opposed to a tape recorder, which impresses the sound magnetically on a tape.

discrete component—A component which has been fabricated prior to its installation (e.g., resistors, capacitors, diodes, and transistors).

discrete part—A separately packaged single circuit element supplying one fundamental property as a lumped characteristic in a given application. Examples: resistor, transistor, diode.

discrete sampling—The lengthening of individual samples so that the sampling process does not deteriorate the intelligence frequency response of the channel.

discrete thin-film component—An individually packaged electronic component having one or more thin films serving as resistive, conductive, and/or insulating elements. Resistors and potentiometers having thin-film metallic resistance elements are examples.

discrimination—1. The difference between losses at specified frequencies, with the system or transducer terminated in specified impedances. 2. In a frequency-modulated system, the detection or demodulation of the imposed variations in the frequency of the carriers. 3. In a tuned circuit, the degree of rejection of unwanted signals.

discriminator—1. A device in which amplitude variations are derived in response to frequency or phase variations. 2. Part of a receiver circuit which removes the desired signal from an incoming frequency-modulated carrier wave by changing modulations in terms of frequency variations into amplitude variations. 3. A facsimile auxiliary device between the radio receiver and the recorder which converts an audio-frequency shift facsimile signal to an amplitude-modulated facsimile signal. 4. A circuit, the output voltage of which varies in amplitude and polarity according to the frequency of the applied signal. It is used as a detector in a frequency-modulated receiver.

discriminator transformer—A transformer used in f-m receivers to convert frequency changes directly to a-f signals.

discriminator tuning unit—A device which tunes the discriminator to a particular subcarrier.

disc-seal tube—Also called lighthouse tube or megatron. An electron tube with disc-shaped electrodes arranged in closely spaced parallel layers, to give a low interelectrode capacitance along with a high power output in the uhf region.

disc storage—1. A storage device that uses magnetic recording on flat rotating discs. 2. The storage of data on the surface of magnetic discs.

dish—A microwave antenna, usually shaped like a parabola, which reflects the radio energy leaving or entering the system.

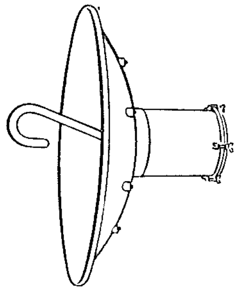

Dish antenna.

dislocation—In a crystal, a region in which the atoms are not arranged in the perfect crystal-lattice structure.

dispatcher—In a digital computer, the section which transfers the "words" to their proper destinations.

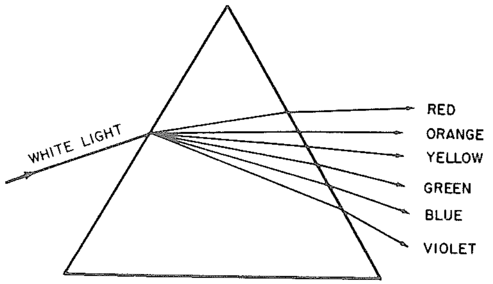

Dispersion by a prism.

dispersion — 1. Separation of a wave into its component frequencies. 2. Scattering of a microwave beam as it strikes an obstruction. 3. The property of an optical material which causes some wavelengths of light to be transmited through the material at different velocities, and the velocity is a function of the wavelength. (This causes each wavelength of light to have a different refractive index.) 4. In a magnetostrictive delay line, the variation of a delay as a function of frequency. 5. The frequency difference that can be analyzed in one sweep by a spectrum analyzer. Dispersion can be considered as that frequency width over which sampling can be performed, and is always equal to or less than the frequency range.

dispersive medium—A medium in which the phase velocity of a wave is related to the frequency.

displacement — The vector quantity representing change of position of a particle.

displacement current—A current which exists in addition to ordinary conduction current in a-c circuits. It is proportional to the rate of change of the electric field. The current at right angles to the direction of propagation determined by the rate at which the field energy changes.

displacement of porches—The difference in level between the front and back porch of a television signal.

displacement transducer—A device which converts mechanical energy into electrical energy, usually by the movement of a rod or an armature. The amount of output voltage is determined by the amount the rod or armature is moved.

display—Visual presentation of a received signal on a cathode-ray tube.

display loss—*See* Visibility Factor.

display modes—Each display mode, such as vector, increment, character, point, vector continue, or short vector specifies the manner in which points are to be displayed on the screen.

display primaries—Also called receiver primaries. The red, green, and blue colors produced by a color television receiver and mixed in proper proportions to produce other colors.

display unit — A device used to provide a visual representation of data.

display window—The width of the portion of the frequency spectrum presented on panoramic presentation, expressed in frequency units, usually megahertz.

disruptive discharge — The sudden, large current through an insulating medium when electrostatic stress ruptures the medium and thus destroys its insulating ability.

dissector—In optical character recognition, a mechanical or electronic transducer that sequentially detects the level of light in different areas of a completely illuminated sample space.

dissector tube—A camera tube having a continuous photocathode on which a photoelectric-emission pattern is formed. Scanning is done by moving the electron optical image of the pattern over an aperture. (*Also see* Image Dissector.)

dissipation—Loss of electrical energy as heat.

dissipation constant—A constant of proportionality between the power dissipated and the resultant temperature rise in a thermistor at a specified temperature.

dissipation factor—Symbolized by D. Ratio between the permittivity and conductivity of a dielectric. The reciprocal of the dissipation factor is the storage factor, sometimes called the quality factor. Abbreviated DF.

dissonance—The formation of maxima and minima by the superposition of two sets of interference fringes from light of two different wavelengths.

dissymmetrical transducer—Also called dissymmetrical network. A transducer with unequal input and output image impedances.

dissymmetrical network—*See* Dissymmetrical Transducer.

distance mark—Also called range mark. A mark which indicates, on a cathode-ray screen, the distance from the radar set to a target.

distance-measuring equipment — Abbreviated DME. A radio navigational aid for determining the distance from a transponder beacon by measuring the time of transmission to and from it.

distance protection—The effect of a device operative within a predetermined electrical distance on the protected circuit to cause and maintain an interruption of power in a faulty circuit.

distance relay—1. A protective relay, the operation of which is a function of the distance between the relay and the point of fault. 2. A device which functions when the circuit admittance, impedance, or reactance increases or decreases beyond predetermined limits.

distance resolution—The ability of a radar to differentiate targets solely by distance measurement. Generally expressed as the minimum distance the targets can be separated and still be distinguishable.

distortion — An undesired change in the waveform of the original signal, resulting in unfaithful reproduction of audio or video signals. Distortion is classified as nonlinear, frequency, or phase, depending on how the waveform is affected.

distortion factor of a wave—The ratio of the effective value of the residue after the elimination of the fundamental to the effective value of the original wave.

distortion meter — An instrument which measures the deviation of a complex wave from a pure sine wave.

distortion tolerance — Of a telegraph receiver, the maximum signal distortion which can be tolerated without error in reception.

distress frequency—A frequency reserved for distress calls, by international agreement. It is 500 kHz for ships at sea and aircraft over the sea.

distributed—Spread out over an electrically significant length, area, or time.

distributed capacitance — Also called self-capacitance. Any capacitance not concentrated within a capacitor, such as the capacitance between the turns in a coil or choke, or between adjacent conductors of a circuit.

distributed constants—Constants such as resistance, inductance, or capacitance that exist along the entire length or area of a circuit, instead of being concentrated within circuit components.

distributed inductance — The inductance along the entire length of a conductor, as distinguished from the inductance concentrated within a coil.

distributed network—An electrical-electronic device which for proper operation depends on physical size in comparison to a wavelength and physical configuration.

distributed parameter network—A network in which the parameters of resistance, capacitance, and inductance cannot be taken as being concentrated at any one point in space. Rather, the network must be described in terms of its magnetic and electric fields and the quantities related to the distributed constants of the network.

distributed pole—A motor has distributed poles when its stator or field windings are distributed in a series of slots located within the arc of the pole.

distributing amplifier—An amplifier, either radio-frequency or audio-frequency, having one input and two or more isolated outputs.

distributing cable—*See* Distribution Cable.

distributing frame—A structure for terminating permanent wires of a central office, private branch exchange, or private exchange, and for permitting the easy change of connections between them by means of cross-connecting wires.

distributing terminal assembly — A frame situated between each pair of selector bays to provide terminal facilities for the selector bank wiring and facilities for cross connection to trunks running to succeeding switches.

distribution—Also called frequency distribution. The number of occurrences of the particular values of a variable as a function of those values.

distribution amplifier—A power amplifier designed to energize a speech, music, or antenna distribution system. Its output impedance is sufficiently low that changes in the load do not appreciably affect the output voltage.

distribution cable—1. Also called distributing cable. A cable extended from a feeder cable for the purpose of providing service to a specific area. 2. In a system, the transmission cable from the distribution amplifier to the drop cable.

distribution center—In an alternating-current power system, the point at which control and rotating equipment is installed.

distribution coefficients — Equal-powered tristimulus values of monochromatic radiations.

distribution switchboard—A power switchboard used for the distribution of electrical energy at the voltage common for each distribution within a building.

distributor—1. *See* Memory Register. 2. The electronic circuitry which acts as an intermediate link between the accumulator and drum storage.

disturbance — An irregular phenomenon which interferes with the interchange of intelligence during transmission of a signal.

disturbed-one output—A "one" output of a magnetic core to which partial-write pulses have been applied since that core was last selected for writing.

disturbed-zero output—A "zero" output of a magnetic core to which partial-write pulses have been applied since that core was last selected for reading.

dither — An oscillation introduced for the purpose of overcoming the effects of friction, hysteresis, or clogging.

dithering — The application of intermittent or periodic acceleration forces sufficient to minimize the effect of static friction with a transducer, without introducing other errors.

divergence loss—The part of the transmission loss that is caused by the spreading of sound energy.

diverging lens—A lens that is thinner in the center than at the edges. Such a lens causes light passing through to spread out, or diverge.

diversity—A form of transmission using several modes, usually in space or in time, to compensate for fading or outages in any one of the modes. In the space diversity system, the same signal is sent simultaneously over several different transmission paths, which are separated enough so that independent propagation conditions can be expected. With time diversity, the same path may be used, but the signal is transmitted more than once, at different times. There are other forms of diversity, using different frequencies or different polarizations to provide the separate transmission modes.

diversity gain—The gain in reception as a result of the use of two or more receiving antennas.

diversity reception—A method of minimizing the effects of fading during reception of a radio signal. This is done by combining and/or selecting two or more sources of received-signal energy which carry the same intelligence but differ in strength or signal-to-noise ratio in order to produce a usable signal.

diverter pole generator — A compound-wound direct-current generator with the series winding of the diverter pole opposing the flux generated by the shunt-wound main pole; provides a close voltage regulation.

divided-carrier modulation—The process by which two signals are added so that they can modulate two carriers of the same frequency but 90° out of phase. The resultant signal will have the same frequency as the carriers, but its amplitude and phase will vary in step with the variations in amplitude of the two modulating signals.

divider—*See* Frequency Divider.

dividing network—Also called speaker dividing network and crossover network. A frequency-selective network which divides the audio-frequency spectrum into two or more parts to be fed to separate devices such as amplifiers or speakers.

D layer—The lowest ionospheric layer. Its intensity is proportional to the height of the sun and is greatest at noon. Waves below approximately 3 MHz are absorbed by the D layer when it is present. High-angle radiation may penetrate the D layer and be reflected by the E layer.

Doherty amplifier—A radio-frequency linear power amplifier divided into two sections, the inputs and outputs of which are connected by quarter-wave (90°) networks. As long as the input-signal voltage is less than half the maximum amplitude, section No. 2 is inoperative and section No. 1 delivers all the power to the load. The load presents twice the optimum impedance required for maximum output. At one-half the maximum input, section No. 1 is operating at peak efficiency but is beginning to saturate. Above this level, section No. 2 comes into operation and decreases the impedance presented to section No. 1. As a result, section No. 2 delivers more and more power to the load until, at maximum signal input, both sections are operating at peak efficiency and each section is delivering one-half the total output power.

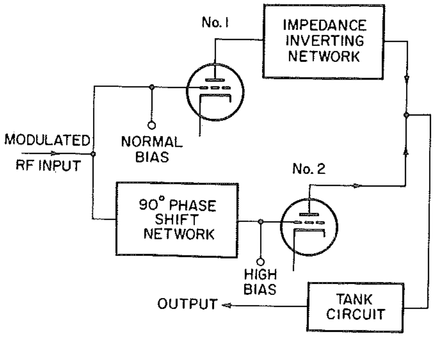

Doherty amplifier.

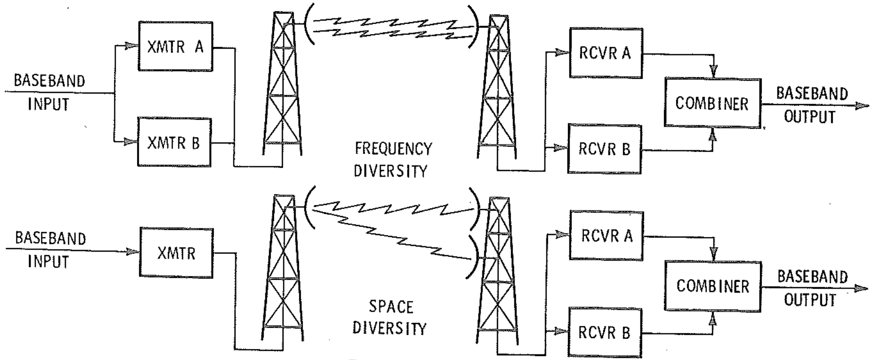

Diversity reception.

dolly—A wheeled platform on which a television camera or other apparatus is mounted to give it wider mobility.

domain—In magnetic theory, that region of a magnetic material in which the spontaneous magnetization is all in one direction. In conventional magnetic-tape coatings, this corresponds to one oxide particle.

domestic induction heater—A home cooking utensil which is heated by induced currents within it. The unit contains a primary inductor, with the utensil itself acting as the secondary.

dominant mode — Also called fundamental mode or principal mode. In waveguide transmission, the mode with the lowest cutoff frequency. Designations for this mode are $TE_{0,1}$ and $TE_{1,1}$ for rectangular and circular wave guides, respectively.

dominant wave—The guided wave which has the lowest cutoff frequency. It is the only wave which will carry energy when the excitation frequency is between the lowest and the next higher cutoff.

dominant wavelength of a color—The predominant wavelength of light in a color.

donor—Also called donor impurity. An impurity atom which tends to give up an electron and thereby affects the electrical conductivity of a crystal. Used to produce N-type semiconductors.

donor impurity—*See* Donor.

donut—*See* Land.

doorknob tube—So called because of its shape. It is a vacuum tube designed for uhf transmiter circuits. It has a low electron-transit time and low interlectrode capacitance because of the close spacing and small size, respectively, of its electrodes.

dopant—An impurity added to a semiconductor to improve its electrical conductivity.

doped junction — A semiconductor junction produced by the addition of an impurity to the melt during crystal growth.

doping—Adding impurities to change the resistivity of semiconductors in order to make N- or P-types.

doping agent—An impurity element added to semiconductor materials used in crystal diodes and transistors. Common doping agents for germanium and silicon include aluminum, antimony, arsenic, gallium, and indium.

doping compensation—The addition of donor impurities to a P-type semiconductor or of acceptor impurities to an N-type semiconductor.

Doppler cabinet—A speaker cabinet in which either the speaker or a baffle board is rotated or moved to change the length of the sound path cyclically and thereby produce a vibrato effect mechanically.

Doppler effect—The observed change of frequency of a wave caused by a time rate of change of the effective distance traveled by the wave between the source and the point of observation. As the distance between a source of constant vibration and an observer diminishes or increases, the received frequencies are greater or less.

Doppler radar—A radar unit that measures the velocity of a moving object by the shift in carrier frequency of the returned signal. The shift is proportionate to the velocity of the object as it approaches or recedes.

Doppler *ranging* (doran)—A c-w trajectory-measuring system which utilizes the Doppler shift to measure the distance between a transmitter, missile, transponder, and several receiving stations. From these measurements trajectory data are computed. In contrast to a similar system, doran circumvents the necessity of continuously recording the Doppler signal by performing the distance measurements with four different frequencies simultaneously.

Doppler shift—The change in frequency of a wave reaching an observer or a system, caused by a change in distance or range between the source and the observer or the system during the interval of reception. It is due to the Doppler effect.

dosage meter—*See* Dosimeter *and* Intensitometer.

dosimeter—Also called intensitometer or dosage meter. An instrument that measures the amount of exposure to nuclear or X-ray radiation utilizing the ability of such radiation to produce ionization of a gas.

dot—*See* Button.

dot cycle—One cycle of a periodic alternation between two signaling conditions, each condition having unit duration. Thus, two-condition signaling consists of a dot, or marking element, followed by a spacing element.

dot encapsulation — A packaging process in which cylindrical components are inserted into a perforated wafer to form a solid block with interconnecting conductors on both surfaces joining the components.

dot generator—An instrument used in servicing color television receivers. It produces a pattern of white dots so that convergence adjustments can be made on the picture tube.

dot pattern—Small dots of light produced on the screen of a color picture tube by the signal from a dot generator. If overall beam convergence has been obtained, the three color-dot patterns will merge into one white-dot pattern.

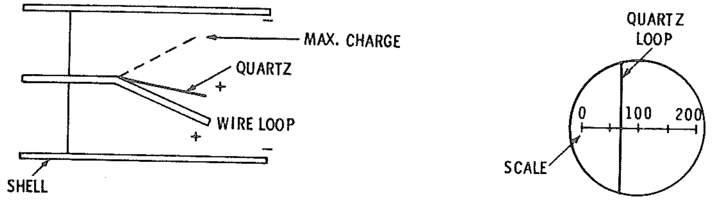

Expanded view of dosage indicator.

Dosage scale.

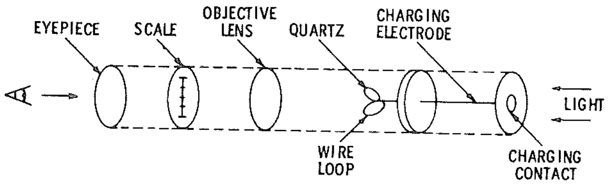

Dosimeter viewer.

Pocket dosimeter.

dot sequential—Pertaining to the association of the primary colors in sequence with successive picture elements of a color television system. (Examples: dot-sequential pickup, dot-sequential display, dot-sequential system, dot-sequential transmission.)

double armature—An armature having two windings and commutators but only one core.

double-base diode—*See* Unijunction Transistor.

double-base junction transistor — Also called tetrode junction transistor. Essentially a junction triode transistor with two base connections on opposite sides of the central region of the transistor.

double-beam cathode-ray tube—A cathode-ray tube having two electron beams capable of producing on the screen two independent traces that may overlap. The beams may be produced by splitting the beam of one gun or by using two guns.

double-bounce calibration — A method of calibration used to determine the zero set error by using round-trip echoes. The correct range is the difference between the first and second echoes.

double-break contacts—A set of contacts in which one contact is normally closed and makes simultaneous connection with two other contacts.

double-break switch—A switch which opens the connected circuit at two points.

double bridge—*See* Kelvin Bridge.

double-button carbon microphone — Also called differential microphone. A microphone with two carbon-resistance elements or buttons, one on each side of a central diaphragm. They are connected in parallel to the current source in order to give twice the resistance change obtainable with a single button.

double-checkerboard pattern — *See* Worst-Case Noise Pattern.

double-current generator — A machine which supplies both direct and alternating current from the same armature winding.

double-diffused epitaxial mesa transistor —*See* Epitaxial-Growth Mesa Transistor.

double-diffused transistor—A transistor in which two PN junctions are formed in the semiconductor wafer by gaseous diffusion of both P-type and N-type impurities. An intrinsic region can also be formed.

double diode — Also called duodiode. A vacuum tube or semiconductor having two diodes in the same envelope.

double-diode limiter—A type of limiter used to remove all the positive signals from a combination of positive and negative pulses or to remove all the negative signals from such a combination of positive and negative pulses.

double-doped transistor — A transistor formed by growing a crystal and successively adding P- and N-type impurities to the melt while the crystal is being grown.

double emitter follower — *See* Darlington Amplifier.

double image—A television picture consisting of two overlapping images due to reception of the signal over two paths that differ in length so signals arrive at slightly different times.

double-junction photosensitive semiconductor—A semiconductor in which the current flow is controlled by light energy. It consists of three layers of a semiconductor material with electrodes connected to the ends of each.

double-length number—Also called double-precision number. An electronic computer number having twice the normal number of digits.

double local oscillator—An oscillator mixing system which generates two r-f signals accurately spaced a few hundred hertz apart and mixes these signals to give the difference frequency which is used as the reference. This equipment is used in an interferometer system to obtain a detectable signal containing the phase information of an antenna pair, and the reference signal to allow removal of the phase data for use.

double-make contacts—A set of contacts in which one contact is normally open and makes simultaneous connection with two other independent contacts when closed.

double moding — Changing from one frequency to another abruptly and at irregular intervals.

double modulation—The process of modulation in which a carrier wave of one frequency is first modulated by a signal wave, and the resultant wave is then made to modulate a second carrier wave of another frequency.

double pole—A term applied to a contact arrangement to denote that it includes two separate contact forms (i.e., two single-pole contact assemblies).

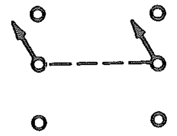

Double-pole, double-throw switch.

double-pole, double-throw switch—Abbreviated dpdt. A switch that has six terminals and is used to connect one pair of terminals to either of the other two pairs.

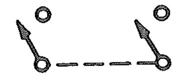

Double-pole, single-throw switch.

double-pole, single-throw switch—Abbreviated dpst. A switch that has four terminals and is used to connect or disconnect two pairs of terminals simultaneously.

double-pole switch—A switch that operates simultaneously in two separate electric circuits or in both lines of a single circuit.

double-precision arithmetic—The use of two computer words to represent a number. This is done where it is necessary to obtain greater accuracy than a single word of computer storage will provide.

double-precision number — *See* Double-Length Number.

double-pulsing station — A Loran station that receives two pairs of pulses and emits pulses at two pulse rates.

double pumping—A technique of pumping a laser for a relatively long time to store energy in subthreshold-level excited states, followed immediately by a very brief second pumping in which the threshold condition is exceeded in some region. This triggers laser oscillation throughout the entire active region and produces peak output powers several times larger than normally seen.

double screen—A three-layer screen consisting of a two-layer screen with the additional second long-persistence coating having a different color and different persistence from the first.

double sideband—Amplitude-modulated intelligence which is transmitted at frequencies both above and below the carrier frequency by the audio-frequency value of the intelligence.

double-sideband transmitter — One which transmits not only the carrier frequency, but also the two sidebands resulting from modulation of carrier.

double-spot tuning—Superheterodyne reception of a given station at two different local-oscillator frequencies. The local oscillator is adjusted either above or below the incoming signal frequency by the intermediate-frequency value.

double-stream amplifier — A microwave traveling-wave amplifier in which amplification occurs through interaction of two electron beams having different average velocities.

double-stub tuner—An impedance-matching device consisting of two stubs, usually fixed three-eighths of a wavelength apart, in parallel with the main transmission lines.

double superheterodyne reception — Also called triple detection. The method of reception in which two frequency converters are employed before final detection.

double-surface transistor—A point-contact transistor, the emitter and collector whiskers of which are in contact with opposite sides of the base.

doublet—The output voltage waveform of a delay line under linear operating conditions when the input to the line is a current step function.

doublet antenna—An antenna consisting of two elevated conductors substantially in the same straight line and of substantially equal length, with the power delivered at the center.

double throw—A term applied to a contact arrangement to denote that each contact form included is a breakmate.

double-throw circuit breaker — A circuit breaker by means of which a change in the circuit connections can be obtained by closing either of two sets of contacts.

double-throw switch—A switch which alternately completes a circuit at either of its two extreme positions. It is both normally open and normally closed.

double-track recorder—*See* Dual-Track Recorder.

double triode—*See* Duotriode.

doublet trigger—A trigger signal consisting of two pulses spaced by a fixed amount for coding.

double-tuned amplifier — An amplifier in which each stage utilizes coupled circuits having two frequencies of resonance for the purpose of obtaining wider bands than are possible with single tuning.

double-tuned circuit—A circuit in which two circuit elements are available for tuning.

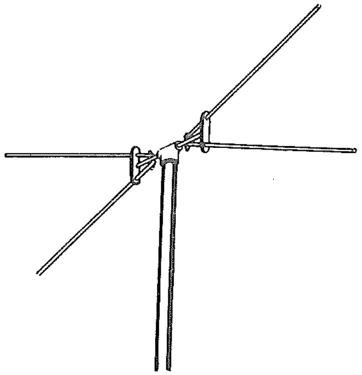

Double-V antenna.

double-V antenna—Also called fan antenna. A modified single dipole which has a higher input impedance and broader bandwidth than an ordinary dipole.

double-winding synchronous generator—A synchronous generator which has two similar windings in phase with one another, mounted on the same magnetic structure but not connected electrically, designed to supply power to two independent external circuits.

doubly-balanced modulator—A modulator circuit in which two Class-A amplifiers are supplied with modulating and carrier signals of equal amplitudes and opposite polarities. Carrier suppression takes place because the two amplifiers share a common plate circuit and only the sidebands appear at the output.

down lead—The wire that connects an antenna to a transmitter or receiver.

down time—The length of time a device is not usable because of a malfunction.

downward modulation — Modulation in which the instantaneous amplitude of the modulated wave is never greater than that of the unmodulated carrier.

dpdt—Abbreviation for double-pole, double-throw switch.

dpst—Abbreviation for double-pole, single-throw switch.

drag angle—A stylus cutting angle of less than 90° to the surface of the record. So called because the stylus drags over the surface instead of digging in. It is the opposite of dig-in angle.

drag cup—A nonmagnetic metal rotated in a magnetic field to generate a torque or voltage proportional to its speed.

drag-cup motor—A small, high-speed, two-phase, alternating-current electric motor having a two-pole, two-phase stator. The rotating element consists only of an extremely light metal cup attached to a shaft rotating on ball bearings. Reversal is accomplished by reversing the connections to one phase. Used in applications requiring quick starting, stopping, and reversal characteristics.

drag magnet—*See* Retarding Magnet.

drain—1. The current taken from a voltage source. 2. In a field-effect transistor, the element that corresponds to the plate of a vacuum tube.

drainage equipment — Equipment used to protect connected circuits from transients produced by the operation of protection equipment.

drain wire—A metallic conductor frequently used in contact with foil-type signal-cable shielding to provide a low-resistance ground return at any point along the shield.

D-region—The region of the ionosphere up to about 90 kilometers above the earth's surface. It is below the E-region.

dress—The exact placement of leads and components in a circuit to minimize or eliminate undesirable feedback and other troubles.

drift—1. Movement of carriers in a semiconductor as voltage is applied. 2. A change in either absolute level or slope of an input-output characteristic. 3. *See* Flutter. 4. *See* Degradation. 5. An undesired change in one of the output parameters of a power supply (voltage, current, frequency, etc.) over a period of time. The change is unrelated to all other variables such as load, line, and environment. Drift is measured over a period of time by keeping all variables (such as line, load, and environment) constant. Specifications usually apply only after a warm-up period. 6. The angular displacement of an aircraft by the wind, generally expressed in degrees. 7. In a d-c amplifier, the change in output with constant input, usually measured in terms of the d-c input signal required to restore normal output; may be called out as microvolts or millivolts per hour. 8. A change in output attributable to any cause.

drift mobility—The average drift velocity of carriers in a semiconductor per unit electric field. In general, the mobilities of electrons and holes are not the same.

drift space — In an electron tube, a region substantially free of alternating fields from external sources, in which relative repositioning of the electrons depends on their velocity distributions and the space-charge forces.

drift speed—Average speed at which electrons or ions progress through a medium.

drift transistor—A type of transistor manufactured with a variable-conductivity base region. Such a base sets up an electric field which speeds up the carriers, thus reducing the transit time and improving high-frequency operation.

drift velocity—Net velocity of charged particles in the direction of the applied field.

drip loop—A loop formed in a transmission line at a point where it enters a building. Condensation of moisture and water that may form on the line will drip off at the loop and thus will not enter the building.

drip-proof motor — A motor in which the ventilating openings are such that foreign matter falling on the motor at any angle not exceeding 15° from the vertical cannot enter the motor either directly or indirectly.

driptight enclosure—An enclosure that is intended to prevent accidental contact with the enclosed apparatus and, in addition, is so constructed as to exclude falling moisture or dirt.

drive—*See* Excitation.

drive belt — A belt used to transmit power from a motor to a driven device.

drive circuit—A circuit, usually a printed-circuit card or an encapsulated module, that converts an input pulse to the appropriate winding excitation sequence to produce one step of the motor shaft.

drive control—*See* Horizontal-Drive Control.

driven element—An antenna element connected directly to the transmission line.

driven sweep—A sweep signal triggered by an incoming signal only.

drive pin—In disc recording, a pin similar to the center pin but located at one side of it and used to prevent a disc record from slipping on the turntable.

drive pulse—A pulsed magnetomotive force applied to a magnetic cell from one or more sources.

driver—1. An electronic circuit which supplies input to another electric circuit. 2. A stage of amplification which precedes the power output stage. 3. In a radar transmitter, a circuit which produces a pulse to be delivered to the control grid of the modulator tube. 4. A unit or stage driving another unit or stage.

driver element—An antenna array element that receives power directly from the transmitter.

driver stage—The amplifier stage preceding the power-output stage.

driving-point admittance—The complex ratio of the alternating current to the applied alternating voltage for an electron tube, network, or other transducer.

driving-point impedance—At any pair of terminals in a network, the driving-point impedance is the ratio of an applied potential difference to the resultant current at these terminals, all terminals being terminated in any specified manner.

driving power—The power supplied to the grid circuit of a tube where the grid swings positive and draws current for part of each cycle of the input signal.

driving-range potential — The voltage difference between the potential of the electrochemically more active anode and the less active protected metal or cathode. One example of driving potential is the electromotive force in a cathodic protection system that causes current between the protected structure (cathode) and the anode. The driving potential decreases as the electrodes become polarized.

driving signal—Television signals that time the scanning at the pickup point. Two kinds of driving signals are usually available from a central sync generator, one composed of pulses at the line frequency and the other of pulses at the field frequency.

driving spring—The spring driving the wipers of a stepping relay.

drone cone — An undriven speaker cone mounted in a bass-reflex enclosure.

droop — The decrease in mean pulse amplitude, expressed as a percentage of the 100% amplitude, at a specified time following the initial attainment of 100% amplitude.

drop bar—A protective device used to ground a high-voltage capacitor when opening a door.

drop bracket transposition—Reversal of the relative positions of two parallel wire conductors while depressing one, so that the crossover is in a vertical plane.

drop cable—In a catv system, the transmission cable from the distribution cable to a dwelling.

drop channel—A type of operation in which one or more channels of a multichannel system are terminated (dropped) at some intermediate point between the end terminals of the system.

drop indicator—An indicator for signaling, consisting of a hinged flap normally held up by a catch. The catch is released by an electromagnet, allowing the flap to drop when a signal is received.

dropout—1. Separate and distinct variations in signal level during reproduction of recorded data, the result of which causes an error in data reduction. 2. Any discrete variation in signal level during the reproduction of magnetic-tape recorded data which results in a data-reduction error.

dropouts—Also called keys. Special images inserted at certain points in the array on a photomask used in the production of monolithic circuits.

drop-out value—The maximum value of current, voltage, or power which will de-energize a previously energized relay. (*Also see* Hold Current *and* Pickup Value.)

dropping resistor—A resistor used to decrease a given voltage by an amount equal to the potential drop across the resistor.

drop repeater—A microwave repeater station equipped for local termination of one or more circuits.

drop wire—A wire suitable for extending an open wire or cable pair from a pole or cable terminal to a building.

drum memory — A rotating cylinder or disc coated with magnetic material so that information can be stored in the form of magnetic spots.

drum recorder—A facsimile recorder in which the record sheet is mounted on a rotating cylinder.

drum speed—The number of revolutions per minute made by the transmitting or receiving drum of a facsimile transmitter or recorder.

drum switch—A switch in which the electrical contacts are made on pins, segments, or surfaces on the periphery of a rotating cylinder or sector, or by the operation of a rotating cam.

drum transmitter—A facsimile transmitter in which the copy is mounted on a rotating cylinder.

drum-type controller—1. A multicircuit timing device, with or without a motor, using a cylindrical carriage into which pins are inserted to program events. 2. A multicircuit timing device intended to be driven from an external rotary power source.

drunkometer—A device measuring the degree of alcoholic intoxication by analyzing the subject's breath.

dry—A condition in which the electrolyte in a cell is immobilized. The electrolyte may be either in the form of a gel or paste or absorbed in the separator material.

dry battery—Two or more dry cells arranged in series, parallel, or series-parallel within a single housing to provide desired voltage and current values.

dry cell—A voltage-generating cell having an immobilized electrolyte. The commonest form has a positive electrode of carbon and a negative electrode of zinc in an electrolyte of sal ammoniac paste.

dry circuit—A circuit in which the open-circuit voltage is .03V or less and the current 200ma or less. (The voltage is most important because at such a low level it is not able to break through most oxides, sulfides, or other films which can build up on contacting surfaces.)

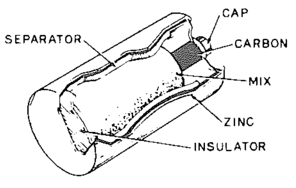

Dry cell.

dry contacts—Contacts which neither break nor make a circuit.

dry-disc rectifier—A rectifier consisting of discs of metal and other materials in contact under pressure. Examples are the copper-oxide and the selenium rectifier.

dry-electrolytic capacitor—An electrolytic capacitor with a paste rather than liquid electrolyte. By eliminating the danger of leakage, the paste electrolyte permits the capacitor to be used in any position.

dry flashover voltage—The voltage at which the air surrounding a clean, dry insulator or shell completely breaks down between electrodes.

dry-reed relay—A relay that consists of one or more capsules containing contact mechanisms that are generally surrounded by an electromagnetic coil for actuation. The capsule consists of a glass tube with a flattened ferromagnetic reed sealed in each end. These reeds, which are separated by an air gap, extend into the tube so as to overlap. When placed in a magnetic field, they are brought together and close a circuit.

dry shelf life—The length of time that a cell can stand without electrolyte before it deteriorates to a point where a specified output cannot be obtained.

dry-type force-air-cooled transformer (class AFA)—A transformer which is not immersed in oil and which derives its cooling by the forced circulation of air.

dry-type self-cooled/forced-air-cooled transformer (class AA/FA)—A transformer which is not immersed in oil, and which has a self-cooled rating with cooling obtained by the forced circulation of air.

dry-type self-cooled transformer (class AA)—A transformer which is cooled by the natural circulation of air and which is not immersed in oil.

dry-type transformer—A transformer which is cooled by the circulation of air and which is not immersed in oil.

dsc—Abbreviation for double silk-covered.

D-scope—A radar display similar to a C-scope except that the blips extend vertically to give a rough estimate of the distance.

"D" service—FAA service pertaining to radio broadcast of meteorological information, advisory messages, and notices to airmen.

DT-cut crystal—A crystal cut to vibrate below 500 kHz.

dual-beam oscilloscope—An oscilloscope in which the cathode-ray tube produces two separate electron beams that may be individually or jointly controlled.

dual capacitor—Two capacitors within a single housing.

dual-channel amplifier—An amplifier which has two channels independent of each other, but similar in design, construction, and output.

dual-diversity receiver—A radio receiver that receives signals from two different receiving antennas and uses whichever signal is the stronger at each instant to offset fading. In one arrangement, two identical radio-frequency systems, each with its own antenna, feed a common audio-frequency channel. In another arrangement, a single receiver is changed over from one antenna to the other by electronic switching at a rate fast enough to prevent loss of intelligibility.

dual-emitter transistor—A passivated PNP silicon planer epitaxial transistor with two emitters; used as a low-level chopper.

dual-frequency induction heater—A type of induction heater in which work coils operating at two different frequencies induce energy, either simultaneously or successively, to material within the heater.

dual-groove record—*See* Cook System.

dual meter—A meter constructed so that two aspects of a circuit may be read simultaneously.

dual modulation—The use of two different types of modulation, each conveying separate intelligence, to modulate a common carrier or subcarrier wave.

dual pickup—*See* Turnover Pickup.

dual trace—A mode of operation in which a single beam in a cathode-ray tube is shared by two signal channels. *Also see* Alternate Mode *and* Chopped Mode.

dual-track recorder—Sometimes called a half-track recorder. Its recording head covers only half the width of the tape, making it possible to record on one half of the tape and then, by turning the reels over, to record on the unused half.

dubbing—In radio broadcasting, the addition of sound to a prerecorded tape or disc.

duct—1. A protective pipe through which conductors or cables are run. 2. In microwave transmission, atmospheric conditions may cause radio waves to follow a narrower path than usual. The narrower path is called a duct. The presence of ducting sometimes causes unusual transmission because the transmission waves do not follow the intended path.

ducting—The trapping of an electromagnetic wave, in a waveguide action, between two layers of the earth's atmosphere, or between a layer of the atmosphere and the earth's surface.

dummy antenna—*See* Artificial Antenna.

dummy instruction—An artificial instruction or address inserted in a list of instructions to a computer solely to fulfill prescribed conditions (such as word length or block length) without affecting the operation.

dummy load—*See* Artificial Load.

dump—1. Also called power dump. To withdraw all power from a computer, either accidentally or intentionally. 2. To transfer all or part of the contents of one section of a digital-computer memory into another section.

dump check—Checking a computer by adding all digits as they are dumped (transferred) to verify the sum to make sure no errors exist as the digits are retransferred.

dumping resistor—A resistor whose function is to discharge a capacitor or network for safety purposes.

duodecal socket—A vacuum-tube socket having 12 pins. Used for cathode-ray tubes.

duodecimal—1. Pertaining to a characteristic or property involving a selection, choice, or condition in which there are 12 possibilities. 2. Pertaining to the numbering system with a radix of 12.

duodiode—Also called dual diode. A vacuum tube or semiconductor having two diodes within the same envelope.

duodiode-pentode—An electron tube containing two diodes and a pentode in the same envelope.

duodiode-triode—An electron tube containing two diodes and a triode in the same envelope.

duolateral coil—*See* Honeycomb Coil.

duopole—An all-pass section with two poles and two zeros.

duotriode—An electron tube containing two triodes in the same envelope. Also called double triode.

duplex—1. The method of operation of a communication circuit in which each end can simultaneously transmit and receive. (Ordinary telephones are duplex. When used on a radio circuit, duplex operation requires two frequencies.) 2. Two-in-one, as two conductors with a common overall insulation or two telegraph transmission channels over one wire.

duplex artificial line—A balancing network simulating the impedance of the real line and distant terminal apparatus; it is employed in a duplex circuit for the purpose of making the receiving device unresponsive to outgoing signal currents.

duplex cable—A cable composed of two insulated stranded conductors twisted together. They may or may not have a common insulating covering.

duplex channel—A communication channel providing simultaneous transmission in both directions.

duplexer—A radar device which, by using the transmitted pulse, automatically switches the antenna from receive to transmit at the proper time.

duplexing assembly (radar)—*See* TR Switch.

duplex operation—Simultaneous operation of transmitting and receiving apparatus at two locations.

duplex system—A system with two distinct and separate sets of facilities, each of which is capable of assuming the system function while the other assumes a standby status. Usually both sets are identical in nature.

duplex tube—A combination of two vacuum tubes in one envelope.

duplication check—A computer check in which the same operation or program is checked twice to make sure the same result is obtained both times.

duration control—A control for adjusting

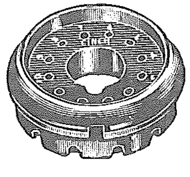

Duodecal socket.

the time duration of reduced gain in a sensitivity-time control circuit.

during cycle—The interval while a timer is operating for its present time period.

dust core—A pulverized iron core consisting of extremely fine iron particles mixed with a binding material for use in radio-frequency coils.

dust-ignitionproof motor — A totally enclosed motor, the enclosure of which is designed and constructed in a manner that will exclude ignitable amounts of dust or amounts which might affect the performance rating, and will not permit arcs, sparks, or heat otherwise generated or liberated inside the enclosure to cause ignition of exterior accumulations or atmospheric suspensions of a specific dust on or in the vicinity of the enclosure.

duty cycle—The amount of time a device operates, as opposed to its idle time. Applied to a device that normally runs intermittently rather than continuously.

duty cyclometer—A test meter which gives a direct reading of duty cycle.

duty factor—1. In a carrier composed of regularly recurring pulses, the product of their duration and repetition frequency. 2. Ratio of average to peak power.

duty ratio—In a pulsed system, such as radar, the ratio of average to peak power.

DX—1. Abbreviation for distance. 2. Reception of distant stations.

DX hound—An amateur who specializes in making distant contacts.

dynamic acceleration — Acceleration in a constantly changing magnitude and direction, either simple or complex motion usually called vibration. Also measured in gravity units.

dynamic braking—A system of braking of an electric drive in which the motor is used as a generator, and the kinetic energy of the motor and driven machinery is employed as the actuating means of exerting a retarding force.

dynamic characteristics — Relationship between the instantaneous plate voltage and plate current of a vacuum tube as the voltage applied to the grid is varied.

dynamic check — A check used to ascertain the correct performance of some or all components of equipment or a system under dynamic or operating conditions.

dynamic contact resistance—In a relay, a change in contact electrical resistance due to a variation in contact pressure on mechanically closed contacts.

dynamic convergence—The condition where the three beams of a color picture tube come together at the aperture mask as they are reflected both vertically and horizontally. (*Also see* Vertical Dynamic Convergence *and* Horizontal Dynamic Convergence.)

dynamic decay — In a storage tube, decay caused by an action such as that of ion charging.

dynamic demonstrator—A three-dimensional schematic diagram in which the components of the radio, television receiver, etc., are mounted directly on the diagram.

dynamic deviation — The difference between the ideal output value and the actual output value of a device or circuit when the reference input is changing at a specified constant rate and all other transients have expired.

dynamic equilibrium of an electromagnetic system—1. The tendency of any electromagnetic system to change its configuration so that the flux of magnetic induction will be maximum. 2. The tendency of any two current-carrying circuits to maintain the flux of magnetic induction linking the two at maximum.

dynamic error—An error in a time-varying signal resulting from inadequate dynamic response of a transducer.

dynamic focus—The application of an a-c voltage to the focus electrode of a color picture tube to compensate for the defocusing caused by the flatness of the screen.

dynamic magnetic field—A field the intensity of which is changing and the lines of force of which are expanding or contracting. Such change can be periodic or random. Unlike the static field, the dynamic field can transfer energy from one point to another without relative motion between the points.

dynamic microphone—*See* Moving-Coil Microphone.

dynamic mutual-conductance tube tester—*See* Transconductance Tube Tester.

dynamic noise suppressor—An audio filter the bandpass of which is adjusted automatically to the signal level. At low signal levels filtering is highest, and at high signal levels all filter action is removed.

dynamic output impedance—*See* Output Impedance.

dynamic pickup—*See* Moving-Coil Pickup.

dynamic plate impedance—The internal resistance to the flow of alternating current between the cathode and plate of a tube.

dynamic plate resistance—*See* A-C Plate Resistance.

dynamic printout—In a computer, a printout of data which occurs as one sequential operation during the machine run.

dynamic range—1. The difference, in decibels, between the overload level and the minimum acceptable signal level in a system or transducer. 2. The ratio of the specified maximum signal-level capability of a system or component to its noise or resolution level, usually expressed in decibels.

dynamic regulator—A transmission regulator in which the adjusting mechanism is in self-equilibrium at only one or a few settings and requires control power to maintain it at any other setting.

dynamic reproducer — *See* Moving-Coil Pickup.

dynamic resistance — Incremental resistance

measured over a relatively small portion of the operating characteristic of a device.

dynamic run—Also called dynamic test. The test performed on an instrument to obtain the overall behavior and to establish or corroborate specifications such as frequency response, natural frequency of the device, etc.

dynamic sequential control—A method of operation in which a digital computer can alter instructions or their sequence, as the computation proceeds.

dynamic speaker—*See* Moving-Coil Speaker.

dynamic storage — Stored computer data which remain in motion on the sensing device (i.e., acoustic delay line, magnetic drum, etc.), as opposed to static storage.

dynamic subroutine — In digital-computer programming, a subroutine which involves parameters (such as decimal point position) from which a properly coded subroutine is derived. The computer itself adjusts or generates the subroutine according to the parametric values chosen.

dynamic test—*See* Dynamic Run.

dynamic transfer-characteristic curve — A curve showing the variation in output current as the input current changes.

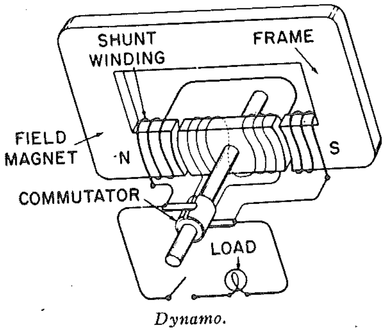

Dynamo.

dynamo—1. Normally called a generator. A machine that converts mechanical energy into electrical energy by electromagnetic induction. 2. In precise terminology, a generator of direct current—as opposed to an alternator, which generates alternating current.

dynamoelectric—Pertaining to the relationship between mechanical force and electrical energy or vice versa.

dynamometer—1. An instrument in which the force between a fixed and a moving coil provides a measure of current, voltage, or power. 2. Equipment designed to measure the power output of a rotating machine by determining the friction absorbed by a hand brake opposing the rotation.

dynamotor—Also called a rotary converter or synchronous inverter. A rotating device for changing a d-c voltage to another value. It is a combination electric motor and d-c generator with two or more armature windings and a common set of field poles. One armature winding receives the direct current and rotates (thus operating as a motor), while the others generate the required voltage (and thus operate as dynamos or generators).

dynaquad—A germanium PNPN semiconductor switching device which is base-controlled and has three terminals. Its operation is similar to that of a flip-flop circuit or latching relay.

dynatron—*See* Tetrode.

dynatron oscillator — A negative-resistance oscillator with negative resistance derived between the plate and cathode of a screen-grid tube operating such that secondary electrons produced at the plate are attracted to the higher-potential screen grid.

dyne—The force that produces an acceleration of one centimeter per second per second on a one-gram mass.

dyne per square centimeter — The unit of sound pressure. One dyne per square centimeter was originally called a bar in acoustics, but the full expression is used in this field now because the bar is defined differently in other applications. Also called microbar.

dynistor—A nonlinear semiconductor having the characteristics of a small current flow as voltage is applied. As the applied voltage is increased a point is reached at which the current flow suddenly increases radically and will continue at this rate even though the applied voltage is reduced.

dynode—An electrode having the primary function of supplying secondary-electron emission in an electron tube.

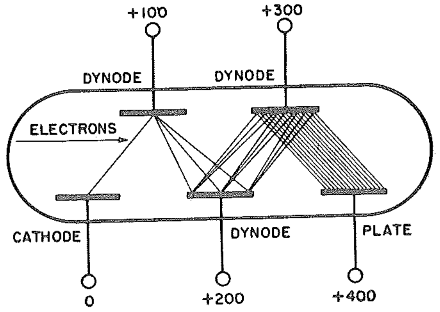

Dynode.

E

E—1. Symbol for voltage or emitter. 2. Abbreviation for illumination.

E and M leads—In a signaling system, the output and input leads, respectively.

early-failure period—The period of equipment life, starting immediately after final assembly, during which equipment failures initially occur at a higher than normal rate due to the presence of defective parts and abnormal operating procedures. Also called debugging period, burn-in period, or infant-mortality period.

early-warning radar—A radar which usually scans the sky in all directions, in order to detect approaching enemy planes and/or missiles at distances far enough away that interceptor planes can be in the air to meet their approach before they are near their target.

earphone—Also called receiver. An electroacoustic transducer intended to be placed in or over the ear.

earth—Term used in Great Britain for ground.

earthed—A British term meaning grounded.

earth inductor—*See* Generating Magnetometer.

earth oblateness—The slight departure from a perfect spherical shape of the shape of the earth and the form of its gravity field.

E-bend—*See* E-Plane Bend.

ebiconductivity—Conductivity induced as the result of electron bombardment.

ebmd—Abbreviation for electron-beam mode discharge.

ec—Abbreviation for enamel covered.

eccentric circle—*See* Eccentric Groove.

eccentric groove—Also called eccentric circle. An off-center locked groove for actuating the trip mechanism of an automatic record changer at the end of a recording.

eccentricity—In disc recording, the displacement of the center of the recording-groove spiral with respect to the record center hole.

Eccles-Jordan circuit—A flip-flop circuit consisting of a two-stage, resistance-coupled electron-tube amplifier. Its output is coupled back to its input, two separate conditions of stability being achieved by alternately biasing the two stages beyond cutoff.

ECCM (electronic counter-countermeasures)—Retaliatory tactics used to reduce the effectiveness of electronic countermeasures.

ECDC (abbreviation for ElectroChemical Diffused-Collector transistor)—A PNP transistor in which all the mass of P material is etched off and replaced with metal, which acts as a heat sink. It is suitable for high-current, high-speed core driver and computer-memory applications.

echelon—One of a series of levels of accuracy of calibration, the highest of which is represented by an accepted national standard. There may be auxiliary levels between two successive echelons.

echo—1. In radar, that portion of the energy reflected to the receiver from a target. 2. A wave which has been reflected or otherwise returned with sufficient magnitude and delay to be distinguishable from the directly transmitted wave. 3. In facsimile, a multiple reproduction on the record sheet caused by the arrival of the same original facsimile signal at different times over transmission paths of different lengths. 4. In a radio system, an electronic condition which causes a signal such as a voice signal to be reflected from some point or points in the circuit back to the point of origination of the signal. 5. A delayed repetition (sometimes several rapid repetitions) of the original sound or signal.

echo area—Equivalent echoing area of a radar target (i.e., the relative amount of radar energy the target will reflect).

echo attenuation—In a four-wire (or two-wire) circuit equipped with repeater or multiplex equipment in which the two directions of transmission can be separated from each other, the attenuation of the echo currents (which return to the input of the circuit under consideration) is determined by the ratio of the transmitted power P1 to the echo power received P2.

echo box—Also called phantom target. A device for checking the over-all performance of a radar system. It comprises a resonant cavity which receives a portion of the pulse energy from the transmitter and retransmits it to the receiver as a slowly decaying transient. The time required for this transient response to decay below the minimum detectable level on the radar indicator is known as the ring time and is indicative of the over-all performance of the radar set.

echo chamber—A reverberant room or enclosure used for adding hollow effects or actual echoes to radio or television programs.

echo check—A method of checking the accuracy of transmission of data in which the received data are returned to the sending end for comparison with the original data.

echo depth sounder—*See* Fathometer.

echo depth sounding—A system of determining the ocean depth by producing a sound just below the water's surface and measuring the amount of time before the echo is reflected from the floor of the ocean.

echo intensifier—A device, located at the target, that is used to increase the amplitude of the reflected energy to an abnormal level.

echo matching—Rotating an antenna to a position in which the pulse indications of an echo-splitting radar are equal.

echo ranging—Determination of both direction and distance of an underwater object from a vessel by ultrasonic radiation.

echo sounder—A sounding device used by ships to determine the depth of water. (*Also see* Echo Depth Sounding.)

echo splitting—In certain radar equipment, the echo return is split and appears as a double indication on the screen of the radar indicator. This splitting is accomplished by special electronic circuits associated with the antenna lobe switching mechanism. When the two echo indications are of equal height, the target bearing is read from a calibrated scale.

echo suppressor—1. A voice-operated device that is connected to a two-way telephone circuit to attenuate echo currents in one direction caused by telephone currents in the other direction. 2. In navigation, a circuit which desensitizes the equipment for a fixed period after the reception of one pulse, for the purpose of rejecting delayed pulses arriving from indirect reflection.

echo talker—A portion of the transmitted signal returned from a distant point to the transmitting source with sufficient time delay to be received as interference.

ECM (electronic countermeasures)—Methods of jamming or otherwise hindering the operation of enemy electronic equipment.

ECO—Abbreviation for electron-coupled oscillator.

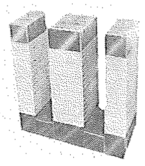

E-core.

E-core—The laminated configuration resembling the capital letter *E* in some transformers and inductive transducers.

eddy-current heating—Synonym for induction heating.

eddy-current loss—The core loss which results when a varying induction produces electromotive forces which cause a current to circulate within a magnetic material.

eddy currents—Also called Foucault currents. Those currents induced in the body of a conducting mass by a variation in magnetic flux.

edge effect—1. *See* Following or Leading White *and* Following or Leading Black. 2. Nonuniformity of electric fields between two parallel plates caused by an outward bulging of electric flux lines at the edges of the plates.

edging—Undesired coloring around the edges of different-colored objects in a color television picture.

Edison base—Standard screw-thread base used for ordinary electric lamps.

Edison distribution system—A three-wire direct-current distribution system, usually 120 to 240 volts, for combined light and power service from a single set of mains.

Edison effect—Also called Richardson effect. The phenomenon wherein electrons emitted from a heated element within a vacuum tube will flow to a second element that is connected to a positive potential.

Edison storage cell—A storage cell having negative plates of iron oxide and positive plates of nickel oxide immersed in an alkaline solution. An open-circuit voltage of 1.2 volts per cell is produced.

E-display—In radar, a rectangular display in which targets appear as blips with distance indicated by the horizontal coordinate and elevation by the vertical coordinate.

EDP—Abbreviation for electronic data processing.

EDP center—*See* Electronic Data Processing Center.

EDPM—Abbreviation for electronic data-processing machine.

eeg—Abbreviation for electroencephalograph.

effective acoustic center—The point from which the spherically divergent sound waves from an acoustic generator appear to diverge. Also called apparent source.

effective actuation time—The sum of the initial actuation time and the contact chatter intervals of a relay following such actuation.

effective ampere—That alternating current which, when flowing through a standard resistance, produces heat at the same average rate as 1 ampere of direct current flowing in the same resistance.

effective antenna length—The length which, when multiplied by the maximum current, will give the same product as the length and uniform current of an elementary electric dipole at the same location, and the same radio field intensity in the direction of maximum radiation.

effective area—The effective area of an antenna in any specified direction is equal to the square of the wavelength multiplied by the power gain (or directive gain) in that direction, divided by 4π.

effective bandwidth—For a bandpass filter, the width of an assumed rectangular bandpass filter having the same transfer ratio at a reference frequency and passing the same mean-square value of a hypothetical current and voltage having even distribution of energy over all frequencies.

effective capacitance—The total capacitance existing between any two given points of an electric circuit.

effective conductivity—The conductance between the opposite parallel faces of a portion of a material having unit length and unit cross section.

effective current—That value of alternating current which will give the same heating effect as the corresponding value of direct current. For sine-wave alternating currents, the effective value is 0.707 times the peak value.

effective cutoff—*See* Effective Cutoff Frequency.

effective cutoff frequency—Also called effective cutoff. The frequency at which the insertion loss of an electric structure between specified terminating impedances exceeds the loss at some reference point in the transmission band.

effective field intensity—Root-mean-square value of the inverse distance fields one mile from the transmitting antenna in all directions horizontally.

effective height—The height of the antenna center of radiation above the effective ground level.

effectively grounded—Grounded through a ground connection of sufficiently low impedance (inherent and/or intentionally added) so that fault grounds which may occur cannot build up voltages that are dangerous to connected equipment.

effective parallel resistance—The resistance leakage considered to be in parallel with a pure dielectric.

effective percentage modulation—For a single, sinusoidal input component, the ratio between the peak value of the fundamental component of the envelope and the direct-current component in the modulated conditions, expressed in percent.

effective radiated power—Antenna input times gain, expressed in watts. If specified for a particular direction, it is the antenna power gain in that direction only.

effective radius of the earth—A value used in place of the geometrical radius to correct the atmospheric refraction when the index of refraction in the atmosphere changes linearly with height. Under conditions of standard refraction, the effective radius is one and one-third the geometrical radius.

effective resistance—1. The average rate of dissipation of electric energy during a cycle divided by the square of the effective current. 2. The equivalent pure d-c resistance which, when substituted for the winding of a motor being checked, will draw the same power. It is also equivalent to the impedance of a circuit having a capacitor connected in parallel with the winding and the capacitor adjusted to unity power factor for the circuit.

effective series resistance—A dropping resistance considered to be in series with an assumed pure capacitance.

effective sound pressure—The root-mean-square value of the instantaneous sound pressure at one point over a complete cycle. The unit is the dyne per square centimeter.

effective speed—The speed (less than rated) which can be sustained over a significant period of time and which reflects slowing effects of control codes, timing codes, error detection, retransmission, tabbing, hand keying, etc.

effective speed of transmission—The average rate over some specified time interval at which information is processed by a transmission facility. Usually expressed as average characters or average bits per unit time. Also called average rate of transmission.

effective thermal resistance—Of a semiconductor device, the effective temperature rise per unit power dissipation of a designated junction above the temperature of a stated external reference point under conditions of thermal equilibrium.

effective value—Also called the rms (root mean square) value. The value of alternating current that will produce the same amount of heat in a resistance as the corresponding value of direct current. The effective value is 0.707 times the peak value.

effective wavelength—The wavelength corresponding to the effective propagation velocity and the observed frequency.

efficiency—1. Ratio of the useful output of a physical quantity which may be stored, transferred, or transformed by a device, to the total input of the device. 2. The ratio of the output power to the input power of a regulated power supply, expressed in percent. In the absence of statements to the contrary, it is assumed to be taken at nominal input and output levels and full load conditions.

efficiency of a source of light—The ratio of the total luminous flux to the total power consumed. In the case of an electric lamp, it is expressed in lumens per watt.

efficiency of rectification—Ratio of direct-current power output to alternating-current power input of a rectifier.

EGO—Acronym for Eccentric orbit Geophysical Observatory.

ehf—Abbreviation for extremely high frequency.

E-H tee—A waveguide junction composed of a combination of E and H plane tee junctions which intersect the main guide at a common point.

E-H tuner—An E-H tee having two arms terminated in adjustable plungers. It is used for impedance transformation.

EIA—Abbreviation for Electronic Industries Association.

Einthoven string galvanometer—A moving-coil type of galvanometer in which the coil is a single wire suspended between the poles of a powerful electromagnet.

E-I pick-off—An assembly of transformerlike laminations, the output coils of which develop a voltage proportional to the displacement of a magnetic element from the neutral position for limited rotary as well as angular travel.

ekg—Abbreviation for electrocardiograph.

elastance—Symbolized by S. In a capacitor, the ratio of potential difference between its electrodes to the charge in the capacitor. It is the reciprocal of capacitance. The unit of measure is the daraf.

$$S\ (\text{daraf}) = \frac{V}{Q}$$

elastic collision—Collision resulting in no molecular excitation when the conservation of momentum and kinetic energy governs the energy transfer.

elastic limit—The maximum stress a solid can endure and still return to its unstrained state when the stress is removed.

elastic wave—A pure acoustic wave; a moving lattice distortion without a magnetic component.

elastivity—The resistance of an electrostatic field. It is the reciprocal of permittivity.

E-layer—One of the regular ionospheric layers with an average height of about 100 kilometers. This layer occurs during daylight hours, and its ionization is dependent on the sun's angle. The principal layer corresponds roughly to what was formerly called the Kennelly-Heaviside layer.

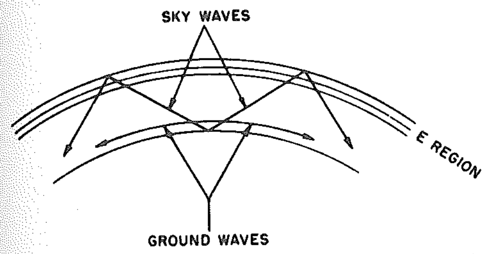

E-layer.

electra—A specific radionavigational aid that provides a number (usually 24) of equisignal zones. Electra is similar to sonne except that in sonne the equisignal zones as a group are periodically rotated in bearing.

electralloy—A nonmagnetic alloy frequently used in radio chassis.

electret—A permanently polarized piece of dielectric material produced by heating the material and placing it in a strong electric field during cooling. Some barium titanate ceramics can be polarized in this way, and so can carnauba waxes. The electric field of an electret corresponds somewhat to the magnetic field of a permanent magnet.

electric—Pertaining to, consisting of, containing, producing, derived from, or produced or operated by, electricity. Used interchangeably with electrical.

electrical—*See* Electric.

electrical angle—A means of specifying a particular instant in an a-c cycle. One complete cycle is considered equal to 360°; hence a half-cycle is 180° and a quarter-cycle is 90°. If one voltage reaches a peak value a quarter of a cycle after another, the electrical angle (phase difference) between the two voltages is 90°.

electrical bail—A switch action in which, upon actuation of one station, the switch changes the contact position, electrically locks the switch in that position, and releases any station previously actuated.

electrical bias—An electrically produced force tending to move the armature of a relay toward a given position.

electrical boresight—The tracking axis as determined by an electrical indication, such as the null direction of a conical scanning or monopulse antenna system or the beam maximum direction of a highly directive antenna.

electrical center—The point approximately midway between the ends of an inductor or resistor. This point divides the inductor or resistor into two equal electrical values (e.g., voltage, resistance, inductance, or number of turns).

electrical degree—A unit of time measurement applied to alternating current.

electrical discharge machining—Machining in which metal is removed by a controlled electrical spark in a dielectric.

electrical distance—The distance between two points, expressed as the length of time an electromagnetic wave in free space takes to travel between them.

electrical element—The concept in uncombined form of any of the individual building blocks from which electronic circuits are synthesized. Examples of basic electrical elements are insulation, conductance, resistance, capacitance, and inductance.

electrical erosion—The loss of contact material due to action of an electrical discharge.

electrical forming—The application of electric energy to a semiconductor device in order to permanently modify the electrical characteristics.

electrical gearing—A term used to describe the action of a system in which the output shaft rotates at a different speed from the input shaft, the ratio being established by electrical means.

electrical glass insulation—Insulating materials made from glass fibers of varying diameters, lengths, compositions, etc., including yarns, rovings, slivers, cords, and sheets or mats, bounded or treated only as necessary to their manufacture.

electrical inertia—Inductance that opposes any change in current flow through an inductor.

electrical initiation—Any source of electrical power used to start a function or sequence.

electrical interlocks—Switches mounted on contactors or other devices and operated by rods or levers. These interlocks open or close, depending on the open or closed position of the contactor or device with which they are associated, and are used to govern succeeding operations of the same or allied devices.

electrical length—Length expressed in wavelengths, radians, or degrees. Distance in wavelengths × 2π = radians; distance in wavelengths × 360 = degrees.

electrical load—A device (e.g., a speaker) comprising resistive and/or reactive components into which an amplifier, generator, etc., delivers power.

electrically operated rheostat—A rheostat used to vary the resistance of a circuit in response to some means of electrical control.

electrically operated valve—A solenoid- or motor-operated valve used in vacuum, air, gas, oil, water, or similar lines.

electrically variable inductor—An inductor in which the inductance can be controlled by a current or a voltage. It is usually made in the form of a saturable reactor with two windings. One is called the signal, or tuned winding, corresponding to the a-c or load winding of a power-handling saturable reactor; the other is the control winding and corresponds to the d-c winding of the saturable reactor.

electrical noise—Unwanted electrical energy other than cross talk in a transmission system.

electrical radian—57.296°, or $1/2\pi$ (1/6.28) of a cycle of alternating current or voltage.

electrical reset—A term applied to a relay to indicate that it is capable of being electrically reset after an operation.

electrical resistivity—The resistance of a material to passage of an electric current through it. Expressed as ohms (units of resistance) per mil foot or as microhms (millionths of an ohm) per centimeter cubed (cm^3) at a specified temperature.

electrical resolver—Special type of synchro having a single winding on the stator and two windings the axes of which are 90° apart on the rotor.

electrical sheet—Iron or steel sheets from which laminations for electric motors are punched.

electrical system—The organized arrangement of all electrical and electromechanical components and devices in a way that will properly control the particular machine tool or industrial equipment.

electrical twinning—*See* Twinning.

electrical zero—A standard synchro position at which electrical outputs have defined amplitudes and time phase.

electric arc—A discharge of energy through a gas.

electric bell—An audible signaling device consisting of one or more gongs and an electromagnetically actuated striking mechanism.

electric brazing—A brazing (alloying) process in which the heat is furnished by an electric current.

electric breakdown voltage—*See* Dielectric Breakdown Voltage.

electric breeze or wind—The emission of electrons from a sharp point of a conductor which carries a high negative potential.

electric charge—Electric energy stored on the surface of an object.

electric chronograph—A highly accurate apparatus for measuring and recording time intervals.

electric circuit—A continuous path consisting of wires and/or circuit elements over or through which an electric current can flow. If the path is broken at any point, current can no longer flow and there is no circuit.

electric contact—A separable junction between two conductors which is designed to make, carry, or break (in any sequence or singly) an electric circuit.

electric controller—A device which governs the amount of electric power delivered to an apparatus.

electric delay line—A delay line using properties of lumped or distributed capacitive and inductive elements. Can be used as a storage medium by recirculating the information-carrying signal.

electric dipole—Also called a doublet. A simple antenna comprising a pair of oppositely charged conductors capable of radiating an electromagnetic wave in response to the movement of an electric charge from one conductor to the other.

electric-discharge lamp—A sealed glass enclosure containing a metallic vapor or an inert gas through which electricity is passed to produce a bright glow.

electric displacement—*See* Electric-Flux Density.

electric-displacement density—*See* Electric-Flux Density.

electric eye—1. The layman's term for a photoelectric cell. 2. The cathode-ray, tuning-indicator tube used in some radio receivers.

electric field—The region about a charged body. Its intensity at any point is the force which would be exerted on a unit positive charge at that point.

electric-field intensity—A measure of the force exerted at a point by a unit charge at that point.

electric-field strength—The magnitude of the electric field in an electromagnetic wave. Usually stated in volts per meter. (*Also see* Dielectric Strength.)

electric-field vector—At a point in an electric field, the force on a stationary positive charge per unit charge. May be measured in either newtons per coulomb or volts per meter. This term is sometimes called the electric-field intensity, but such use of the word "intensity" is deprecated in favor of "field strength," since intensity denotes power in optics and radiation.

electric-filament lamp—A glass bulb either evacuated or filled with an inert gas and having a resistance element electrically heated to, and maintained at, the temperature necessary to produce incandescence.

electric filter—*See* Electric-Wave Filter.

electric-flux density—Also called electric-displacement density or electric displacement. At a point, the vector equal in magnitude to the maximum charge per unit area which would appear on one face of a thin metal plate introduced in the electric field at that point. The vector is normal to the plate from the negative to the positive face.

electric furnace—A furnace in which electric arcs provide the source of heat.

electric generator—A machine that transforms mechanical power into electrical power.

electric governor-controlled series-wound motor—A series-wound motor having an electric speed governor connected in series with the motor circuit. The governor is usually built into the motor.

electric hygrometer—An instrument for indicating humidity by electric means. Its operation depends on the relationship between the electric conductance and moisture content of a film of hygroscopic material.

electric hysteresis—Internal friction in a dielectric material when subjected to a varying electric field (e.g., the paper or mica dielectric of a capacitor in an a-c circuit). The resultant heat generated can eventually break down the dielectric and cause the capacitor to fail.

electrician—A person engaged in designing, making, or repairing electric instruments or machinery. Also, one who sets up an electrical installation.

electric image—The electrical counterpart of an object; i.e., the fictitious distribution of the same amount of electricity that is actually distributed on a nearby object.

electricity—The property of certain particles to possess a force field which is neither gravitational nor nuclear. The type of force field associated with electrons is defined as negative and that associated with protons and positrons as positive. The fundamental unit is the charge of an electron: 1.60203×10^{-19} coulomb. Electricity can be further classified as static electricity or dynamic electricity. Static electricity in its strictest sense refers to charges at rest, as opposed to dynamic electricity, or charges in motion. Static electricity is sometimes used as a synonym for triboelectricity or frictional electricity.

electric light—Light produced by an electric lamp.

electric lines of force—In an electric field, curves the tangents of which at any point give the direction of the fields at that point.

electric meter—A device that measures and registers the amount of electricity consumed over a certain period of time.

electric mirror—*See* Dynode.

electric moment—For two charges of equal magnitude but opposite polarities, a vector equal in magnitude to the product of the magnitude of either charge by the distance between the centers of the two charges. The direction of the vector is from the negative to the positive charge.

electric motor—A device which converts electrical energy into rotating mechanical energy.

electric network—A combination of any number of electric elements, having either lumped or distributed impedances, or both.

electric oscillations—The back-and-forth flow of electric charges whenever a circuit containing inductance and capacitance is electrically disturbed.

electric potential—A measure of the work required to bring a unit positive charge from an infinite distance or from one point to another (the difference of potential between two points).

electric precipitation—The collecting of dust or other fine particles floating in the air. This is done by inducing a charge in the particles, which are then attracted to highly oppositely charged collector plates.

electric probe—A rod inserted into an electric field during a test to detect d-c, audio, or r-f energy.

electric reset—A qualifying term indicating that the contacts of a relay must be reset electrically to their original positions following an operation.

electric shield—A housing, usually aluminum or copper, placed around a circuit to provide a low-resistance path to ground for high frequency radiations and thereby prevent interaction between circuits.

electric strain gage—A device which detects the change in shape of a structural member under load and causes a corresponding change in the flow of current through the device.

electric strength—The maximum electric charge a dielectric material can withstand without rupturing. (*Also see* Dielectric Strength.)

electric stroboscope—An instrument for observing or for measuring the speed of rotating or vibrating objects by electrically producing periodic changes in the intensity of light used to illuminate the object.

electric tachometer—A tachometer (rpm indicator) that utilizes voltage or electrical impulses.

electric telemeter—A system consisting of a meter which measures a quantity, a transmitter which sends the information to a distant station, and a receiver which indicates or records the quantity measured.

electric transcription—In broadcasting, a disc recording of a message or a complete program.

electric transducer—A device actuated by electric waves from one system and supplying power, also in the form of electric waves, to a second system.

electric tuning—A system by which a radio receiver is tuned to a station by pushing a button (instead of, say, turning a knob).

electric vector—A component of the electromagnetic field associated with electromagnetic radiation. The component is of the nature of an electric field. The electric vector is supposed to coexist with, but act at right angles to, the magnetic vector.

electric wave—Another term for the electromagnetic wave produced by the back-and-forth movement of electric charges in a conductor.

electric-wave filter—Also called electric filter. A device that separates electric waves of different frequencies.

electrification—1. The process of establishing an excess of positive or negative charges in a material. 2. The process of applying a voltage to a component or device.

electroacoustic—Pertaining to a device (e.g., a speaker and a microphone) which involves both electric current and sound-frequency pressures.

electroacoustic transducer—A device that receives excitations from an electric system and delivers an output to an acoustic system, or vice versa. A speaker is an example of the first, and a microphone is an example of the second.

electroanalysis—The process of determining the quantity of an element or compound in an electrolyte solution by depositing the element or compound on an electrode by electrolysis.

electrobiology—The science concerned with electrical phenomena of living creatures.

electrobioscopy—The application of a voltage to produce muscular contractions.

electrocardiogram — Essentially an electromyogram of the heart muscle. All muscular activity in the body is characterized by the discharge of polarized cells, the aggregate current from which causes a voltage drop that can be measured on the skin. A changing emf will appear between electrodes connected to the arms, legs, and chest, which rises and falls with heart action such that the period of the resulting waveform is the time between heartbeats. Various positive and negative peaks within one cycle of this waveform have been lettered P, Q, R, S, and T, a notation which aids in subsequent analysis and diagnosis.

electrocardiograph—A medical instrument for detecting irregularities in the action of a human heart. It measures the changes in voltage occurring in the human body with each heartbeat. Abbreviated ekg.

electrochemical deterioration—A process in which autocatalytic electrochemical reactions produce an increase in conductivity and in turn ultimate thermal failure.

electrochemical equivalent—The weight of an element, compound, radical, or ion involved in a specified electrochemical reaction during passage of a specified quantity of electricity such as a coulomb.

electrochemical junction transistor — A junction transistor produced by etching an N-type germanium wafer on opposite sides with jets of a salt solution such as indium chloride.

electrochemical recording — A recording made by passing a signal-controlled current through a sensitized sheet of paper. The paper reacts to the current and thereby produces a visual record.

electrochemical transducer—A device which uses a chemical change to measure the input parameter, and the output of which is a varying electrical signal proportional to the measurand.

electrochemistry—That branch of science concerned with reciprocal transformations of chemical and electrical energy. This includes electrolysis, electroplating, the charge and discharge of batteries, etc.

electrocoagulation—The process of solidifying tissue by means of a high frequency electrical current.

electrocution—Killing by means of an electric current.

electrode—1. In an electronic tube, the conducting element that does one or more of the following: emits or collects electrons or ions, or controls their movement by means of an electric field on it. 2. In semiconductors, the element that does one or more of the following: emits or collects electrons or holes, or controls their movements by means of an electric field on it. 3. In electroplating, the metal being plated. 4. A conductor by means of which a current passes into or out of a fluid or an organic material such as human skin; often one terminal of a lead. 5. A metallic conductor such as in an electrolytic cell, where conduction by electrons is changed to conduction by ions or other charged particles.

electrode admittance—The alternating component of the electrode current divided by that of the electrode voltage (all other electrode voltages maintained constant).

electrode capacitance—The capacitance between one electrode and all the other electrodes connected together.

electrode characteristic—The relationship, usually shown by a graph, between the electrode voltage and current, all other electrode voltages being maintained constant.

electrode conductance—The quotient of the in-phase component of the electrode alternating current divided by the electrode alternating voltage, all other electrode voltages being maintained constant. This is a variational and not a total conductance.

electrode current—Current passing into or out of an electrode.

electrode dark current—In phototubes, the component of electrode current that flows in the absence of ionizing radiation and optical photons. Also called dark current.

electrode dissipation—The power which an electrode dissipates as heat when bombarded by electrons and/or ions and radiation from nearby electrodes.

electrode drop—The voltage drop produced in an electrode by its resistance.

electrode impedance—The reciprocal of electrode admittance.

electrodeless discharge — A luminous discharge produced by means of a high-frequency electric field in a gas-filled glass tube that has no internal electrodes.

electrodeposition—Also called electrolytic deposition. The process of depositing a substance on an electrode by electrolysis, as in electroplating, electroforming, electrorefining, or electrotinning.

electrode potential—The instantaneous voltage on an electrode. Its value is usually given with respect to the cathode of a vacuum tube.

electrode reactance—The imaginary component of electrode impedance.

electrode resistance—The reciprocal of electrode conductance. It is the effective parallel resistance, not the real component of electrode impedance.

electrode voltage—The voltage between an electrode and the cathode or a specified point of a filamentary cathode. The terms "grid voltage," "anode voltage," "plate voltage," etc., designate the voltage between these electrodes and the cathode. Unless otherwise stated, electrode voltages are measured at the available terminals.

electrodialytic process — A process for producing fresh water by using a combination of electric current and two types of chemically treated membranes.

electrodynamic—Pertaining to electric current, electricity in motion, and the actions and effects of magnetism and induction.

electrodynamic braking—A method of stopping a tape-deck motor gently by the application of a predetermined voltage to the motors.

electrodynamic instrument—An instrument which depends for its operation on the reaction between the current in one or more moving coils and the current in one or more fixed coils.

electrodynamics—The science dealing with the various phenomena of electricity in motion, including interactions of currents with each other, with their associated magnetic fields, and with other magnetic fields.

electrodynamic speaker—A speaker consisting of an electromagnet called the field coil, through which a direct current flows.

electrodynamometer—An instrument for detecting or measuring an electric current by determining the mechanical reactions between two parts of the same circuit.

electroencephalogram — A waveform obtained by plotting brain voltages (available between two points on the scalp) against time. An electroencephalogram is not necessarily a periodic function, although it can be—particularly if the patient is unconscious. These voltages are of extremely low level and require recording apparatus which display excellent noise rejection.

electroencephalograph (eeg) — An instrument for measuring and recording the rhythmically varying potentials produced by the brain by the use of electrodes applied to the scalp.

electroencephaloscope — An instrument for detecting brain potentials at many different sections of the brain and displaying them on a cathode-ray tube.

electroforming—1. Also called electrodeposition and electroplating. Making a metal object by using electrolysis to deposit a metal on an electrode. 2. Creating a PN junction by passing a current through point contacts on a semiconductor.

electrographic recording—Also called electrostatography. The producing of a visible record by using a gaseous discharge between two or more electrodes to form electrostatically charged patterns on an insulator. (*Also see* Electrostatic Electography.)

electrokinetics—The branch of physics concerned with electricity in motion.

electroluminescence — Luminescence resulting from a high-frequency discharge through a gas or from application of an alternating current to a layer of phosphor.

electroluminescent lamp—A lamp in the shape of a panel which is decorative as well as illuminative. It consists primarily of a capacitor having a ceramic dielectric with electroluminescent phosphor. The amount of illumination is determined by the voltage across the layer and by the frequency applied to it.

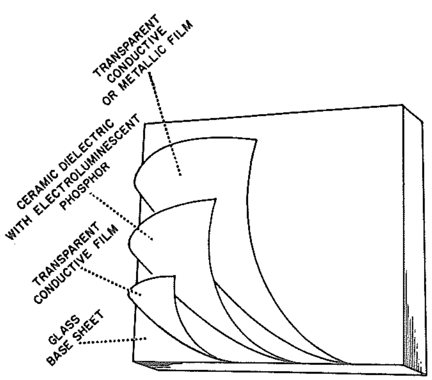
TRANSPARENT CONDUCTIVE OR METALLIC FILM
CERAMIC DIELECTRIC WITH ELECTROLUMINESCENT PHOSPHOR
TRANSPARENT CONDUCTIVE FILM
GLASS BASE SHEET

Electroluminescent lamp.

electrolysis—1. The process of changing the chemical composition of a material (called the electrolyte) by sending an electric current through it. 2. The decay of an underground structure by chemical action due to stray electrical currents.

electrolyte—1. A substance in which the conduction of electricity is accompanied by chemical action. 2. The paste which forms the conducting medium between the electrodes of a dry cell, storage cell, or electrolytic capacitor. 3. A substance which, when dissolved in a suitable liquid (often water), dissociates into ions, thus rendering the liquid electrically conducting.

electrolytic—Pertaining to or made by electrolysis; deposited by electrolysis; pertaining to or containing an electrolyte.

electrolytic capacitor—A capacitor consisting of two conducting electrodes, with the anode having a metal oxide film formed on it. The film acts as the dielectric or insulating medium. It is repairable in the presence of an operating electrolyte, usually an acid or salt. Generally used for filtering, bypassing, coupling, or decoupling.

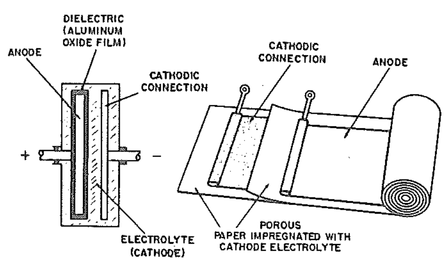
DIELECTRIC (ALUMINUM OXIDE FILM)
ANODE
CATHODIC CONNECTION
CATHODIC CONNECTION
ANODE
+
-
ELECTROLYTE (CATHODE)
POROUS PAPER IMPREGNATED WITH CATHODE ELECTROLYTE

Electrolytic capacitor.

electrolytic cell—In a battery, the container, two electrodes, and the electrolyte.

electrolytic conduction—The flow of current between electrodes immersed in an electrolyte. It is caused by the movement of ions from one electrode to the other when a voltage is applied between them.

electrolytic conductivity — Also called specific conductance. A measure of the ability of a solution to carry an electric current. Defined as the reciprocal of the resistance in ohms of a 1-cm cube of the liquid at a specified temperature. The units of specific conductance are the reciprocal ohm-cm (or mho/cm) and one millionth of this, micromho/cm. (High-quality condensed steam and distilled or demineralized water have specific conductances at room temperatures as low as or lower than 1 micromho/cm).

electrolytic corrosion—*See* Corrosion.

electrolytic deposition — *See* Electrodeposition.

electrolytic dissociation—The breaking up of molecules into ions in a solution.

electrolytic interrupter—A device which is tilted to change the current through it.

electrolytic iron—Iron obtained by an electrolytic process. The iron possesses good magnetic qualities and is exceptionally free of impurities.

electrolytic potential—The difference in potential between an electrode and the immediately adjacent electrolyte, expressed in terms of some standard electrode difference.

electrolytic recording—A form of facsimile recording in which ionization causes a chemically moistened paper to undergo a change.

electrolytic rectifier — A rectifier consisting of metal electrodes in an electrolyte, in which rectification of alternating current is accompanied by electrolytic action. A polarization film formed on one of the electrodes permits current in one direction but not in the other.

electrolytic refining—The refining or purifying of metals by electrolysis.

electrolytic switch — A switch having two electrodes projecting into a chamber containing a precisely measured quantity of a conductive electrolyte, leaving an air bubble of predetermined width. When the switch is tilted from true horizontal, the bubble shifts position and changes the amount of electrolyte in contact with the electrodes, thereby changing the amount of current passed by the switch. Used as a leveling switch in gyro systems.

electrolyzer—An electrolytic cell that produces alkalies, metals, chlorine, or other allied products.

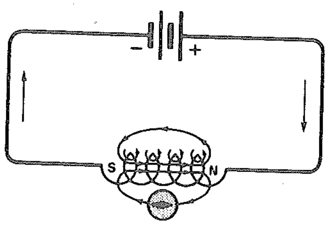

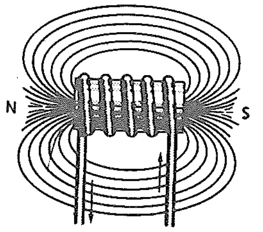

Electromagnet.

electromagnet—A temporary magnet consisting of a solenoid with an iron core. A magnetic field exists only while current flows through the solenoid.

electromagnetic—Having both magnetic and electric properties.

electromagnetic cathode-ray tube—A cathode-ray tube which uses electromagnetic deflection to deflect the electron beam.

electromagnetic communications — The electromagnetic wave conductor is space itself. The electromagnetic frequencies available today for communications fall into two categories: frequencies which form "wireless" communications (such as visual light of fairly high frequency), and frequencies man uses for wireless communications (such as radio, short wave, and microwave transmitting, of relatively lower frequencies). In communicating by radio, short wave, and microwave frequencies, translators similar in principle to those used in electrical communications are needed, although the equipment requirement increases.

electromagnetic coupling—The mutual relationship between two separate but adjacent wires when the magnetic field of one induces a voltage in the other.

electromagnetic crack detector—An instrument for detecting hidden cracks in iron or steel objects by magnetic means.

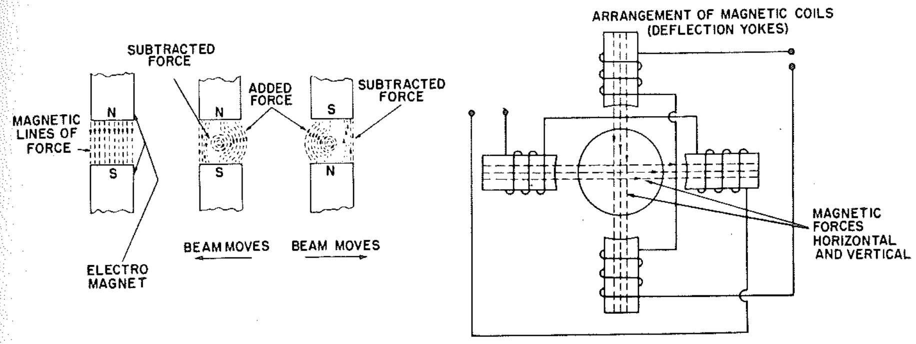

Electromagnetic deflection coil.

electromagnetic deflection—The deflection of an electron stream by means of a magnetic field. In a television receiver, the magnetic field for deflecting the electron beam horizontally and vertically is produced by two pairs of coils, called the deflection yoke, around the neck of the picture tube.

electromagnetic deflection coil — A coil, around the neck of a cathode ray tube, for deflecting the electron beam.

electromagnetic delay line—A delay line the operation of which is based on the time of propagation of electromagnetic waves through distributed or lumped capacitance and inductance.

electromagnetic energy — Forms of radiant energy such as radio waves, heat waves, light waves, X-rays, gamma rays, and cosmic rays.

electromagnetic environment—The r-f field or fields existing in an area or desired in an area to be shielded.

electromagnetic field—1. The field of influence produced around a conductor by the current flowing through it. 2. A rapidly moving electric field and its associated magnetic field. The latter is perpendicular to both the electric lines of force and their direction.

electromagnetic focusing—In a television picture tube, the focusing produced by a coil mounted on the neck. Direct current through the coil produces magnetic field lines parallel to the tube axis.

electromagnetic horn—A horn-shaped structure that provides highly directional radiation of radio waves in the 100-megahertz or higher frequency range.

electromagnetic induction — The voltage produced in a coil as the number of magnetic lines of force (flux linkages) passing through the coil changes.

electromagnetic interference—Electromagnetic phenomena which, either directly or indirectly, can contribute to a degradation in performance of an electronic receiver or system. (The phrases "radio interference," "radio-frequency interference," and "noise" have been employed at various times in the same context.)

electromagnetic lens—An electron lens in which the electron beams are focused electromagnetically.

electromagnetic mirror—A surface or region capable of reflecting radio waves, such as one of the ionized layers in the upper atmosphere.

electromagnetic oscillograph — An oscillograph in which a mechanical motion is derived from electromagnetic forces to produce a record.

electromagnetic radiation — That form of energy which is characterized by transversely oscillating electric and magnetic fields and which propagates at velocity "c" in free space. At a sufficient distance from the source, the electric-field vector and the magnetic-field vector are at right angles to each other, forming a right-handed (coordinate) system. In an ionized medium, a longitudinal component may be present.

electromagnetic relay—An electromagnetically operated switch composed of one or more coils and armatures. Whenever current flows through the coil, its magnetic field attracts the armature and completes a circuit.

electromagnetic repulsion — The repelling action between like poles of electromagnets.

electromagnetics — In physics, the branch concerned with the relationships between electric currents and their associated magnetic fields.

electromagnetic spectrum—A chart or graph showing the relationship between all known electromagnetic waveforms classified by wavelengths.

electromagnetic theory of light—The theory which states that electromagnetic and light waves have identical properties.

electromagnetic transduction — 1. Conversion of the measurand into the output induced in a conductor by a change in magnetic flux. 2. A wave produced by the oscillation of an electric charge. 3. A wave in which there are both electric and magnetic displacements. 4. A transverse wave associated with the transmission of electromagnetic energy.

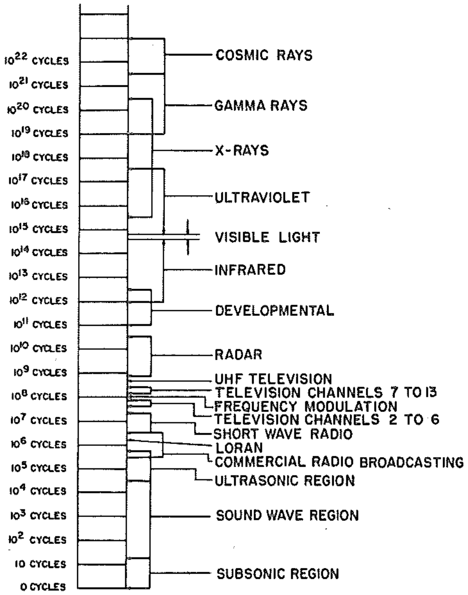

Electromagnetic radiation spectrum chart.

electromagnetic-type microphones—Microphones in which the voltages are varied by an electromagnet (namely, ribbon or velocity, dynamic or moving-coil, and reluctance or moving-vane microphones).

electromagnetic unit—Abbreviated emu. A unit .of electricity based primarily on the magnetic effect of an electric current. The fundamental centimeter-gram-second unit is the abampere. Now considered obsolete.

electromagnetic vibrator—A mechanical device for interrupting the flow of direct current and thereby making it a pulsating current. This is done where a circuit requires an alternating current to operate. A reed within the vibrator is alternately attracted to two electromagnets.

electromagnetic wave—The radiant energy produced by oscillation of an electric charge. It includes radio, infrared, visible and ultraviolet light waves, and X-, gamma, and cosmic rays.

electromagnetism — The magnetic field around a wire or other conductor when, and only when, current passes through it.

electromechanical—Any device using electrical energy to produce mechanical movement.

electromechanical bell—A bell with a prewound spring-driven clapper which is tripped electrically to ring the bell.

electromechanical breakdown—A mechanical runaway that occurs when the mechanical restoring force fails to balance the electrical compressive force.

electromechanical chopper — *See* Contact Modulator.

electromechanical frequency meter — A meter which uses the resonant properties of mechanical devices to indicate frequency.

electromechanical recorder—A device which transforms electrical signals into equivalent mechanical motion which is transferred to a medium by cutting, embossing, or writing.

electromechanical timer—Usually refers to a motor-driven timer, with or without an electrically operated clutch. Can also apply to pneumatic and thermal timers, or slow pull-in or drop-out relays.

electromechanical transducer — A device that transforms electrical energy into mechanical energy or vice versa. A speaker is an example of the first, and a microphone of the second.

electromechanics—That branch of electrical engineering concerned with machines producing or operated by electric currents.

electrometallurgy—That branch of science concerned with the application of electrochemistry to the extraction or treatment of metals.

electrometer—1. An electrostatic instrument that measures a potential difference or an electric charge by the mechanical force exerted between electrically charged surfaces. 2. A d-c vacuum-tube voltmeter with an extremely high input resistance, usually around 10^{10} megohms, as opposed to 10 megohms or less for a conventional type.

electrometer amplifier—An amplifier circuit having sufficiently low-current drift and other noise components, sufficiently low amplifier input-current offsets, and adequate power and current sensitivities to be usable for measuring current variations of considerably less than 10^{-12}A.

electrometer tube—A vacuum tube having a very low control-electrode conductance, to facilitate the measurement of extremely small direct currents and voltages.

electromotive force—Abbreviated emf. The force which causes electricity to flow when there is a difference of potential between two points. The unit of measurement is the volt.

electromotive series—A list of metals arranged in decreasing order of their tendency to pass into ionic form by losing electrons.

electromyogram—Classically, a waveform of the contraction of a muscle as a result of electrical stimulation. Usually the stimulation comes from the nervous system (normal muscular activity). The record of potential difference between two points on the surface of the skin resulting from the activity or action potential of a muscle.

electromyograph — An instrument for measuring and recording potentials generated by muscles.

electromyography—An electronic method of measuring and recording the electrical activity of muscles.

electron—Also called negatron. One of the natural elementary constituents of matter; it carries a negative electric charge of one electronic unit and has approximately 1/1840th the mass of a hydrogen atom or 9.107×10^{-28} gram. Electrons surround the positively charged nucleus and determine the chemical properties of the atom. Positive electrons, or positrons, also exist.

electron attachment—Process by which an electron is attached to a neutral molecule to form a negative ion. Often characterized by the attachment coefficient η, which is the number of attachments per centimeter of drift. Also characterized by the ratio $h = \sigma_a/\theta$, where σ_a is the attachment cross section and θ the total cross section.

electron avalanche — The chain reaction started when one free electron collides with one or more orbiting electrons and frees them. The free electrons then free others in the same manner, and so on.

electron beam—A narrow stream of electrons moving in the same direction under the influence of an electric or magnetic field.

electron-beam generator—A velocity-modulated generator, such as a klystron tube, used to generate extremely high frequencies.

electron-beam instrument — Also called a cathode-ray instrument. An instrument in which a beam of electrons is deflected by an electric or magnetic field (or both). Usually the beam is made to strike a fluorescent screen so the deflection can be observed.

electron-beam machining — A process in which controlled electron beams are used to weld or shape a piece of material.

electron-beam magnetometer — An instrument that measures the intensity and direction of magnetic forces by the immersion of an electron beam into the magnetic field.

electron-beam mode discharge—A form of discharge produced by a perforated-wall hollow cathode operating under conditions of pressure, voltage, and geometry usually associated with the abnormal glow discharge.

electron-beam tube—An electron tube which depends for its operation on the formation and control of one or more electron beams.

electron-beam welding—The process of using a focused beam of electrons to heat materials to the fusion point.

electron charge — Also called elementary charge. The charge of a single electron. Its value is 1.60203×10^{-19} coulomb. The fundamental unit of electrical charge.

electron-coupled oscillator — Abbreviated ECO. A circuit using a multigrid tube in which the cathode and two grids operate as a conventional oscillator and the electron stream couples the plate-circuit load to the oscillator.

electron coupling—In vacuum (principally multigrid) tubes, the transfer of energy between electrodes as electrons leave one and go to the other.

electron drift—The movement of electrons in a definite direction through a conductor, as opposed to the haphazard transfer of energy from one electron to another by collision.

electronegative—Having an electric polarity that is negative.

electron emission—The freeing of electrons into space from the surface of a body under the influence of heat, light, impact, chemical disintegration, or a potential difference.

electron flow—The movement of electrons from a negative to a positive point in a metal or other conductor, or from a negative to a positive electrode through a liquid, gas, or vacuum.

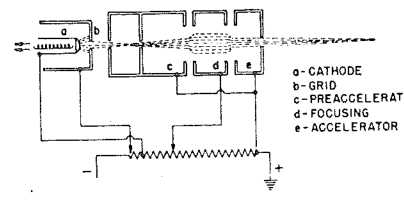

Electron gun.

electron gun — 1. An electrode structure which produces and may control focus and may deflect and converge one or more electron beams. 2. The portion of a cathode-ray tube or camera tube which emits a beam of controlled electrons.

electronic—1. Pertaining to that branch of science which deals with the motion, emission, and behavior of currents of free electrons, especially in vacuum, gas, or phototubes and special conductors or semiconductors. This is contrasted with electric, which pertains to the flow of large currents in metal conductors. 2. Of or pertaining to devices, circuits, or systems using the principle of electron flow through a conductor—for example, *electronic* control, equipment, instrument, circuit.

electronic autopilot — An arrangement of gyroscopes, electronic amplifiers, and servomotors for detecting deviations in the flight of an aircraft and applying the required corrections directly to its control cables.

electronic "bug"—A keying system which converts the Morse signals from a hand key into correctly proportioned and spaced dots and dashes.

electronic camouflage — Use of electronic means, or exploitation of electronic characteristics, to reduce, submerge, or eliminate the radar-echoing properties of a target.

electronic circuit—A circuit containing one or more electron tubes, transistors, magnetic amplifiers, etc.

electronic commutator — A type of switch which provides a continuous switching or sampling of a number of circuits by means of a radial-beam electronic tube or electronic switching circuit.

electronic control — Also called electronic regulation. The control of a machine or condition by electronic devices.

electronic counter—An instrument capable of counting up to several million electrical pulses per second.

electronic countermeasures—Reducing the effectiveness of enemy equipment employing or affected by electromagnetic radiations. This is done by using appropriate electronic devices.

electronic countermeasures control—1. Collection and sorting of large quantities of data for the purpose of measuring and defining radar signals. 2. Examination of the data received in order to determine selection and switching of countermeasure devices with little or no time delay.

electronic coupling—The method of coupling electrical energy from one circuit to another through the electron stream in a vacuum tube.

electronic data-processing center—A place in which is kept automatically operated equipment, including computers, designed to simplify the interpretation and use of data gathered by instrumentation installations or information-collection agencies. Abbreviated EDP center.

electronic data-processing machine—Abbreviated EDPM. A machine or its device and attachments used primarily in or with an electronic data-processing system.

electronic data-processing system — Any machine or group of automatically intercommunicating machines capable of entering, receiving, sorting, classifying, computing and/or recording alphabetical or numerical accounting or statistical data (or all three) without intermediate use of tabulating cards.

electronic deception—Radiation or reradiation of electromagnetic waves to mislead the enemy in the interpretation of data received by his electronic equipment.

electronic differential analyzer—A form of analog computer using interconnected electronic integrators to solve differential equations.

electronic digital computer — A machine which uses electronic circuitry in the main computing element to perform arithmetic and logical operations on digital data (i.e., data represented by numbers or alphabetic symbols) automatically, by means of an internally stored program of machine instruction. Such devices are distinguished from calculators, on which the sequence of instructions is externally stored and is impressed manually (desk calculators) or from tape or cards (card-programmed calculators).

electronic efficiency — The ratio of (a) the power at the desired frequency delivered by the electron stream to the oscillator or amplifier circuit to (b) the direct power supplied to the stream.

electronic flash—Also called strobe. The firing of special light-producing, high-voltage, gas-filled glass tubes with a high instantaneous surge of current furnished by a capacitor or bank of capacitors which have been charged from a high-voltage source (usually 450 volts or higher).

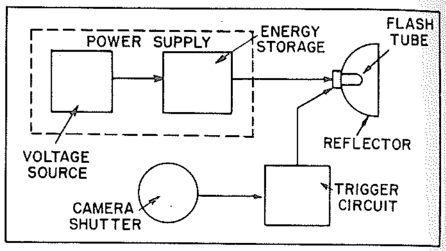

Electronic flash unit.

electronic flash tube—*See* Flash Tube.

electronic frequency synthesizer—A device which generates two or more selectable frequencies from one or more fixed-frequency sources.

electronic heating — Also called high-frequency heating. A method of heating a material by inducing a high-frequency current in it or having the material act as the dielectric between two plates charged with a high-frequency current.

electronic industries — Industrial organizations engaged in the manufacture, design, development, and/or substantial assembly of electronic equipment, systems, assemblies, or the components thereof.

Electronic Industries Association (EIA)—A trade association of the electronics industry. Some of its functions are the formulation of technical standards, dissemination of marketing data, and the maintenance of contact with government agencies in matters relating to the electronics industry.

electronic instrument — Any instrument which depends for its operation on the action of either one or more electron devices.

electronic intelligence — The technical and intelligence information derived from foreign noncommunications electromagnetic radiations emanating from other than nuclear detonations or radioactive sources.

electronic keying — A method of keying whereby the dots and dashes are produced solely by electronic means.

electronic line scanning—Facsimile scanning in which a spot on a cathode-ray tube moves across the copy electronically while the record sheet or subject copy is moved mechanically in a perpendicular direction.

electronic microphone—A device which depends for its operation on the generation of a voltage by the motion of one of the electrodes in a special electron tube.

electronic mine detector—*See* Mine Detector.

electronic multimeter—A device employing the characteristics of an electron-tube circuit for the measurement of electrical quantities, at least one of which is voltage or current, on a single calibrated scale.

electronic music—The electronic generation and processing of audio signals or the electronic processing of natural sound, and the manipulation and arrangement of these signals via tape recorders into a finished musical composition.

electronic organ—The electronic counterpart of the pipe organ. All tones and tone variations such as vibrato, tremolo, etc., are produced by electronic circuits instead of by pipes.

electronic packaging — The coating or surrounding of an electronic assembly with a dielectric compound.

electronic photometer—An electronic instrument for measuring intensity of light, brightness of paints, turbidity of solutions, etc. It comprises a phototube, an electronic direct-current amplifier, and an indicating instrument.

electronic power supply — A circuit which transforms electrical input energy, alternating or direct current—into output energy—alternating or direct current. (Sources operating on rotating machine principles, or deriving electrical power from other energy forms such as batteries and solar cells, are excluded.) Supplies covered by this definition fall into one of four groups: 1. A-c in, d-c out—most common supplies. 2. A-c in, a-c out—line regulators, variable frequency supplies. 3. D-c in, d-c out—converters. 4. D-c in, a-c out—inverters.

electronic products—Materials, parts, components, subassemblies, and equipment which employ the principles of electronics in performing their major functions. These products may be used as instruments and controls in communications, detection, amplification, computation, inspection, testing, measurement, operation, recording, analysis, and other functions employing electronic principles.

electronic profilometer—An electronic instrument for measuring surface roughness. The diamond-point stylus of a permanent-magnet dynamic pickup is moved over the surface being examined. The resultant variations in voltage are amplified, rectified, and measured with a meter calibrated to read directly in micro-inches of deviation from smoothness.

electronic reconnaissance—Search for electromagnetic radiations to determine their existence, source, and pertinent characteristics for electronic warfare purposes.

electronic rectifier—A rectifier using electron tubes or equivalent semiconductor elements as rectifying elements.

electronic regulation — *See* Electronic Control.

electronic relay—An electronic circuit that provides the functional equivalent of a relay, but has no moving parts.

electronics—1. The field of science and engineering concerned with the behavior of electrons in devices and the utilization of such devices. 2. Of or pertaining to the field of electronics, such as electronics engineer, course, laboratory, committee.

electronic stethoscope — An electronic amplifier of sounds within a body. Its selective controls permit tuning for low heart tones or high pulmonary tones. It has an auxiliary output for recording or viewing audio patterns.

electronic stimulator—A device for applying electronic pulses or signals to activate muscles, or to identify nerves, or for muscular therapy, etc.

electronic switch—1. A circuit element causing a start and stop action or a switching action electronically, usually at high speeds. 2. A test instrument for presenting two waveshapes on a single-gun cathode-ray tube.

electronic thermal conductivity—The part of the thermal conductivity due to the transfer of thermal energy by means of electrons and holes.

electronic timer—1. A synchronizer, pulse generator, modulator, or keyer that originates a series of continuous control pulses at an unvarying repetition rate known as the pulse-recurrence frequency. 2. A timer using electronic circuits (either tube or transistor type) to control a time period, in place of a motor or other means.

electronic tube relay—A relay that employs electronic tubes as components.

electronic tuning—1. Altering the frequency of a reflex klystron oscillator by changing the repeller voltage. 2. Frequency changing in a transmitter or receiver by changing a control voltage rather than circuit components.

electronic voltmeter—Also called vacuum-tube voltmeter. A voltmeter which utilizes the rectifying and amplifying properties of electron tubes and their circuits to secure such characteristics as high input impedance, wide frequency range, peak-to-peak indications, etc.

electronic volt-ohmmeter—A device employing the characteristics of an electron-tube circuit for the measurement of voltage and resistance on a single calibrated scale.

electronic warfare—Military usage of electronics to reduce an enemy's effective use of radiated electromagnetic energy and to insure our own effective use.

electronic waveform synthesizer—An instrument using electron devices to generate an electrical signal of a desired waveform.

electron image tube — A cathode-ray tube having a photoemissive mosaic upon which an optical image is projected, and an electron gun to scan the mosaic and convert the optical image into corresponding electrical current.

electron lens—The convergence of the electrons into a narrow beam in a cathode-ray tube by deflecting them electromagnetically or electrostatically. So called because its action is analogous to that of an optical lens.

electron microscope—An instrument which uses an electron beam to penetrate thin samples of a material. It is possible to magnify images of the material on a screen or film up to 350,000 times.

electron multiplier — A vacuum tube in which electrons liberated from a photosensitive cathode are attracted successively to a series of electrodes called dynodes. In doing so, each electron liberates others by secondary emission and thereby greatly increases the number of electrons flowing in the tube.

electron-multiplier section—A section of an electron tube in which an electron current is amplified by one or more successive dynode stages.

electron optics—The branch of electronics concerned with the behavior of the electron beam under the influence of electrostatic and electromagnetic forces.

electron-pair bond—A valence bond formed by two electrons, one from each of two adjacent atoms.

electron paramagnetic resonance—A condition in which a paramagnetic solid subjected to two magnetic fields, one of which is fixed and the other normal to the first and varying at the resonance frequency, emits electromagnetic radiation associated with changes in the magnetic quantum number of the electrons.

electron-ray tube—Also called a "magic eye." A tube which indicates visibly on a fluorescent target the effects of changes in control-grid voltage applied to the tube. Used as a tuning indicator in receivers.

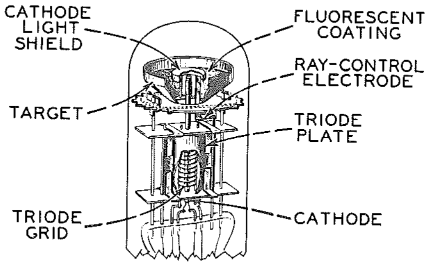

Electron-ray tube.

electron scanning—The moving of an electron beam back and forth and/or up and down by deflecting the beam electromagnetically or electrostatically.

electron spin—The twirling motion of an electron, independent of any orbital motion.

electron telescope—An apparatus for seeing through haze and fog. An infrared image is formed optically on the photoemissive mosaic of an electron-image tube and then made visible by the tube.

electron transit time—The time required for electrons to travel between two electrodes in a vacuum tube. This time is extremely important in tubes designed for ultrahigh frequencies.

electron tube—An electron device in which the electrons move through a vacuum or gaseous medium within a gas-tight envelope.

electron unit—The unit of charge (negative or positive) equal to the charge on an electron.

electron volt — Abbreviated ev or eV. The amount of kinetic energy gained by an electron when it is accelerated through an electric potential difference of 1 volt. It is equivalent to 1.603×10^{-12} erg. It is a unit of energy or work, not of voltage.

electron-wave tube — An electron tube in which streams of electrons having different velocities interact and cause a progressive change in signal modulation along their length.

electro-optical transistor—A transistor capable of responding in nanoseconds to both light and electrical signals.

electropad — The part of an electrocardiograph body electrode that makes contact with the skin.

electrophonic effect—The sensation of hearing produced when an alternating current of suitable frequency and magnitude from an external source is passed through an animal or human body.

electrophoresis—The movement of particles or ions in solution caused by applying an electric field, as reported by O. Lodge in 1886.

electrophoresis apparatus — An apparatus for causing migration of charged particles (ions) in solution in an electric field. Types include paper, cascading electrodes, high voltage, gel, and thin layer.

electrophoresis scanner—An instrument for reading bands on paper strips or gel, for the purpose of measuring particle movement due to electrophoresis.

electrophorus—An early type of static-electricity generator.

electrophotographic process — The process in which images are formed by various electrical and photographic means. Examples are processes employing selenium-coated drums or zinc-oxide–coated paper.

electrophotometer—An instrument using a photoelectric sensor for colorimetric determinations.

electroplaques — Individual electricity-producing cells in eels and other electric fishes connected in series-parallel arrays, like miniature elements of a battery. They are usually thin waferlike cells, the two surfaces of which differ markedly.

electroplate—To deposit a metal on the surface of certain materials by electrolysis.

electroplating—The electrodeposition of an adherent metal coating on a conductive object for protection, decoration, or other purposes. The object to be plated is placed in an electrolyte and connected to one terminal of a d-c voltage source. The metal to be deposited is similarly immersed and connected to the other terminal. Ions of the metal provide transfer to the metal as they make up the current between the electrodes.

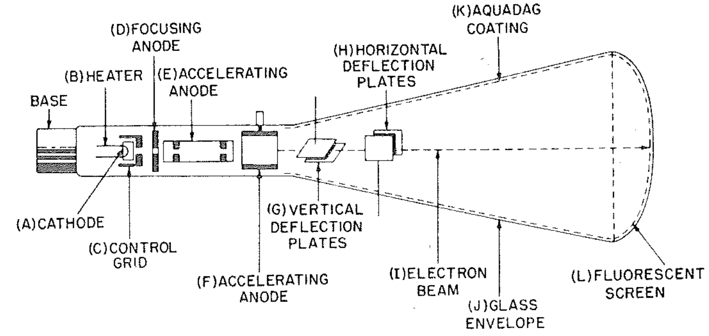

Electrostatic deflection.

electropolishing—The process of producing a smooth, lustrous surface on a metal by making it the anode in an electrolytic solution and preferentially dissolving the minute protuberances.

electrorefining—*See* Electroplating.

electroretinograph—An instrument for measuring the electrical response of the human retina to light stimulation.

electroscope—An electrostatic instrument for measuring a potential difference or an electric charge by means of the mechanical force exerted between electrically charged surfaces.

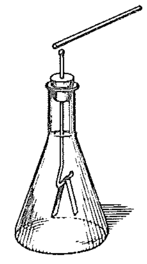

Electroscope.

electrosensitive recording—The passage of electric current into a sheet of sensitive paper to produce a permanent record.

electroshock—A state of shock produced by passing an electric current through the brain. It is useful in the treatment of certain mental disorders.

electrospinograph — A device for detecting and recording electric signals of the spinal cord.

electrostatic—Pertaining to static electricity —i.e., electricity, or an electric charge, at rest.

electrostatic actuator—An apparatus comprising an auxiliary external electrode which permits known electrostatic forces to be applied to the diaphragm of a microphone for the purpose of obtaining a primary calibration.

electrostatic capacitor — Two conducting electrodes separated by an insulating material such as air, ceramic, mica, gas, paper, plastic film, or glass. These are generally high-impedance devices.

electrostatic charge — An electric charge stored in a capacitor or on the surface of an insulated object.

electrostatic component — The portion of radiation due to electrostatic fields.

electrostatic-convergence principle — The principle of electron-beam convergence through use of an electrostatic field.

electrostatic coupling—Method of coupling by which charges on one surface influence those on another through capacitive action.

electrostatic deflection—The method of deflecting an electron beam by passing it between charged plates mounted inside a cathode-ray tube.

electrostatic electrography—The branch of electrostatography which produces a visible record by employing an insulating medium to form latent electrostatic patterns without the aid of electromagnetic radiation.

electrostatic electrophotography — That branch of electrostatography which produces a visible record by employing a photoresponsive medium to form latent electrostatic images with the aid of electromagnetic radiation.

electrostatic energy—The energy contained in electricity at rest, such as in the charge of a capacitor.

electrostatic field—The vector force field set up in the vicinity of nonmoving electrical charges. The strength of this static field at a point is defined as the force per unit charge on a stationary positive test charge, provided the test charge is so small that it does not disturb the original charge distribution.

electrostatic flux—The electrostatic lines of force existing between bodies at different potentials.

electrostatic focusing—The focusing of an electron beam by the action of an electric field.

electrostatic galvanometer — Galvanometer operated by the effects of two electric charges on each other.

electrostatic generator—A device for the production of electric charges by electrostatic action.

electrostatic induction—1. The process of charging an object electrically by bringing it near a charged object. 2. Capacitive induction of interfering signals over an air gap separating an instrument (e.g., from its wiring or housing).

electrostatic instrument — An instrument which depends for its operation on the attraction and repulsion between electrically charged bodies.

electrostatic memory—Also called electrostatic storage. A memory device in which information is retained by an electrostatic charge. A special type of cathode-ray tube is usually employed, together with associated circuitry.

electrostatic memory tube—Also called storage tube. An electron tube in which information is retained by electric charges.

electrostatic microphone—Also called capacitor or condenser microphone. A microphone which depends for its operation on variations in its electrostatic capacitance.

electrostatic precipitation—The process of removing smoke, dust, and other particles from the air by charging them so that they can be attracted to and collected by a properly polarized electrode.

electrostatic process — A reproduction method in which image formation depends on electrical rather than chemical changes induced by light.

electrostatic recording—Recording by means of a signal-controlled electrostatic field.

electrostatic relay—A relay in which two or more conductors that are separated by insulating material move because of the mutual attraction or repulsion produced by electric charges applied to the conductors.

electrostatics—The branch of physics concerned with electricity at rest.

electrostatic separator — An apparatus in which a finely pulverized mixture of the materials to be separated is passed through the powerful electrostatic field between two electrodes.

electrostatic series—*See* Triboelectric Series.

electrostatic shield—A shield which prevents electrostatic coupling between circuits, but permits electromagnetic coupling.

electrostatic speaker—Also called capacitor or condenser speaker. A speaker in which the mechanical forces are produced by the action of electrostatic fields.

electrostatic storage — The storage of changeable information in the form of charged or uncharged areas usually on the screen of a cathode-ray tube.

electrostatic transducer—A transducer that consists of a capacitor, at least one plate of which can be set into vibration. Its operation depends on the interaction between its

electric field and a change in its electrostatic capacity.

electrostatic tweeter — A speaker with a movable flat metal diaphragm and a nonmovable metal electrode capable of reproducing high audio frequencies. The diaphragm is driven by the varying high voltages applied across it and the electrode.

electrostatic unit—An electric unit based primarily on the dynamic interaction of electric charges. Defined as a charge which, if concentrated on a small sphere, would repel with a force of one dyne a similar charge one centimeter away in a vacuum.

electrostatic voltmeter — A voltmeter depending for its action on electrostatic forces. Its scale is usually graduated in volts or kilovolts.

electrostatography—The process of recording and reproducing visible patterns by the formation and utilization of latent electrostatic charge patterns.

electrostriction—A mechanical deformation caused by the application of an electric field to any dielectric material. The deformation is proportional to the square of the applied field. This phenomenon results from the induced dipole movement caused by the applied field, resulting in the mechanical distortion.

electrostrictive effect—The elastic deformation of a dielectric by an electrostatic field.

electrostrictive relay—A relay the operation of which is produced by an electrostrictive-dielectric actuator.

electrotape — An electronic distance-measuring device.

electrotherapeutics—*See* Electrotherapy.

electrotherapy—Also known as electrotherapeutics. The medical science or use of electricity to treat a disease or ailment.

electrotherapy apparatus — Equipment for applying electric current to the body for massage or heat treatment.

electrothermal—The heating effect of electric current, or the electric current produced by heat.

electrothermal expansion element—An actuating element consisting of a wire strip or

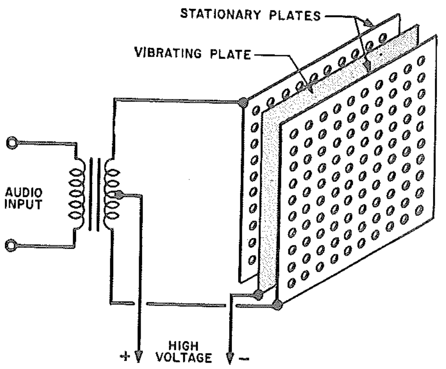

Electrostatic speaker.

other shape and having a high coefficient of thermal expansion.

electrothermal recorder — A recorder in which heat produces the image on the recording medium in response to the received signals.

electrothermal recording—*See* Electrothermal Recorder.

electrothermic instrument—An instrument which depends for its operation on the heating effect of a current. Examples are the thermocouple, bolometric, hot-wire, and hot-strip instruments.

electrothermics—The branch of science concerned with the direct transformation of electric energy into heat.

electrowinning—The process by which metals are recovered from a solution by electrolysis.

element—1. One of the 104 known chemical substances that cannot be divided into simpler substances by chemical means. A substance the atoms of which all have the same atomic number (e.g., hydrogen, lead, uranium). 2. In a computer, the portion or subassembly which constitutes the means of accomplishing one particular function, such as the arithmetic element. 3. Any electrical device (such as an inductor, resistor, capacitor generator, line, or electron tube) with terminals at which it may be connected directly to other electrical devices. 4. The dot or dash of an International Morse character. 5. A radiator, either active or parasitic, that is part of an antenna. 6. The smallest portion of a televised picture that still retains the characteristics of the picture. 7. A portion of a part that cannot be renewed without destruction of the part. 8. A constituent of a microcircuit or integrated circuit that contributes directly to its operation.

elemental area—*See* Picture Element.

elemental semiconductor—A semiconductor containing only one element in the undoped state.

elementary charge—A natural unit or quantum into which both positive and negative charges appear to be subdivided. It is the charge on a single electron and has a value of about 4.77×10^{-10} electrostatic units.

element error rate—The ratio of the number of elements incorrectly received to the total number of elements sent.

elevation—The angular position perpendicular to the earth's surface.

elevation-position indicator—A radar display which simultaneously shows angular elevation and slant range of detected objects.

elevator leveling control — A positioning control used to align the platform of an elevator with the floor level of the building. Metal vanes are mounted in the elevator shaft at each floor level, and an oscillator is mounted on the elevator car. When the elevator is properly leveled, the metal vane is between the plate and the grid coils of the oscillator. A relay connected in the oscillator circuit now energizes. The contacts of

this relay are connected in the motor-control circuit of the elevator so that the elevator stops in alignment with the floor level.

elf — Abbreviation for extremely low frequency.

eliminator—Also called a battery eliminator. A device operated from an a-c or d-c power line and used for supplying direct current and voltage to a battery-operated circuit.

elliptically polarized wave—An electromagnetic wave the electric intensity vector of which describes an ellipse at one point.

elliptical polarization—Polarization in which the wave vector rotates in an elliptical orbit about a point.

e/m—The ratio of the electric charge to the mass for particles such as electrons and positive rays. For slow-moving electrons, the value of e/m is 1.77×10^4 coulombs per gram. The value decreases with increasing velocity, however, because of an increase in effective mass.

embedding—A general term for all methods of surrounding or enclosing components and assemblies with a substantial thickness of electrically insulating solid or foam material that substantially fills the voids or interstices between parts.

embedment—The complete encasement of a part or assembly to some uniform external shape. A relatively large volume of a complete package consists of the embedment material.

embossed-foil printed circuits—A printed circuit formed by indenting the desired pattern of metal foil into an insulating base, and then mechanically removing the remaining unwanted raised portion.

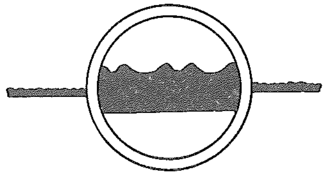

Embossed-groove recording.

embossed-groove recording—A method of recording sounds on discs or film strips by embossing sound grooves with a blunt stylus rather than by cutting into them with a sharp stylus. Embossing throws the material up in furrows on each side of the sound groove without actually removing any of the material in the disc or strip.

embossing stylus—A recording stylus with a rounded tip which forms a groove in the recording medium by merely displacing the material instead of removing it completely.

emergency communication — The transmission or reception of distress, alarm, urgent, or safety signals or messages relating to the safety of life or property, or the occasional operation of equipment to determine whether it is in working condition.

emergency radio channel—Any radio frequency reserved for emergency use, particularly for distress signals.

emergency service—The radio-communication service carrier used for emergency purposes.

emf—Abbreviation for electromotive force.

emission—1. The waves radiated into space by a transmitter. 2. The ejection of electrons from the surface of a material (under the influence of heat, for example).

emission characteristic — The relationship between the emission and the factor controlling it, such as temperature, voltage, or current of the filament or heater. This relationship is usually shown on a graph.

emission current—The current produced in the plate circuit of a tube when all the electrons emitted by the cathode pass to the plate.

emission efficiency—The rating of a hot cathode. Expressed in milliamperes per watt.

emission spectrum—The spectrum showing the radiation emitted by a substance, such as the light emitted by a metal when placed in an electric arc, or the light emitted by an incandescent filament.

emission types—The classification of modes of radio transmission adopted by international agreement. The a-m designations are:

Type A0. Unmodulated continuous-wave transmission.
Type A1. Telegraphy or pure continuous waves.
Type A2. Modulated telegraphy.
Type A3. Telephony.
Type A4. Facsimile.
Type A5. Television.

emission-type tube tester — Also called an English-reading tube tester. A tube tester for checking the electron emission from the filament or cathode. The indicating meter is generally calibrated to read "Good" or "Bad." The tester connects all elements such as the plate and screen, suppressor, and control grids together and uses them as an anode.

emission velocity—The initial velocity at which electrons emerge from the surface of a cathode, ranging from zero up to a few volts (attained by very few electrons). This effect accounts for the existence of virtual cathodes, and also for the shape of the cutoff region of plate current.

emissivity—The ratio of the radiant energy emitted by a radiation source to the radiant energy of a perfect (black-body) radiator having the same area and at the same temperature and conditions.

emitron — A cathode-ray tube developed in England.

emitron camera—A British television camera tube resembling an iconoscope.

emittance—The power per unit area radiated by a source of energy.

emitter—1. An electrode within a transistor from which carriers are usually minority carriers; when they are majority carriers, the emitter is referred to as a majority emitter. 2. Of a transistor, a region from which charge carriers that are minority carriers in the base are injected into the base. 3. A device used on a punched-card machine to give timed pulses at regular intervals during the machine cycle.

emitter bias—The bias voltage applied to the emitter of a transistor.

emitter current—The direct current flowing in the emitter circuit of a transistor.

emitter follower—A common-collector transistor amplifier similar in operation to a vacuum-tube cathode follower.

emitter-follower amplifier — *See* Common-Collector Amplifier.

emitter junction—A semiconductor junction normally biased in the low-resistance direction so that minority carriers are injected into the interelectrode region.

emitter (junction transistor)—In a double-junction (PNP or NPN) transistor, the end semiconductor material that is forward-biased with respect to the base and emits a flow of carriers into the interelectrode region. Roughly comparable to the cathode of a tube.

emitter resistance—The resistance in series with the emitter lead in the common-T equivalent circuit of a transistor.

emitter semiconductor — A junction normally biased in the low-resistance direction to inject minority carriers into an interelectrode region.

emitter voltage—The voltage between the emitter terminal and a reference point.

empire cloth—A cotton or linen cloth coated with varnish and used as insulation on coils and other parts of electrical equipment.

empirical—Based on actual measurement, observation, or experience, as opposed to theoretical determination.

emu—Abbreviation for electromagnetic unit.

enable—To permit a circuit to be activated by the removal of a suppression signal.

enabling gate—A circuit which determines the start and length of a generated pulse.

enabling pulse—A pulse which opens a normally closed electric gate, or otherwise permits an operation for which a pulse is a necessary condition.

enameled wire—Wire coated with a layer of baked-enamel insulation.

encapsulating material—A composition primarily adapted for use on or around an electrical device to provide protection from the surrounding environment.

encapsulation—A protective coating of cured plastic placed around delicate electronic components and assemblies. It is identical to potting, except the cured plastic is removed from the mold. The plastic therefore determines the color and surface hardness of the finished part. The molds may be made of any suitable material.

enclosed relay — A relay in which both the coil and the contacts are protected from the environment.

enclosure—A housing for any electrical or electronic device, normally a speaker.

encode—Also called encipher; the inverse of decode. Conversion of signals to coded form. A circuit that generally converts signals from decimal to binary form.

encoder—1. An electromechanical device that can be attached to a shaft to produce a series of pulses to indicate shaft position; when the output is differentiated, the device is an accurate tachometer. (It is fundamentally oriented to digital rather than analog techniques.) An encoder contains a disc with a printed pattern; as the disc rotates, it makes and breaks a circuit. The more make-and-break cycles per revolution, the better the resolution. 2. A digital-to-analog converter.

end-around shift—In a computer, the movement of characters from one end of the register to the other end of the same register.

end bell—An accessory which is similar to a cable clamp and attaches to the back of a plug or receptacle. It serves as an adapter for the rear of connectors. Some angular end bells have built-in cable clamps. Angular end bells up to 90° are available.

end bracket—*See* End Shield.

end effect—The capacitive effect at the ends of a half-wave antenna. To compensate for this effect, a dipole is cut slightly shorter than a half wave.

end-fire array—A linear or cylindrical antenna having its direction of maximum radiation parallel to the long axis of the array.

end instrument—A device connected to one terminal of a loop and capable of converting usable intelligence into electrical signals, or vice versa. Includes all generating, signal-converting, and loop-terminating devices at the transmitting and/or receiving location.

end item—A combination of products, parts, and/or materials that is ready for its intended use.

end mark—In a computer, a code or signal used to indicate the termination of a unit of information.

endodyne reception—A British term applying to reception of unmodulated code signals. A vacuum-tube circuit having a local oscillator whose frequency is slightly different from that of the carrier signal. Thus a beat signal in the audio range is produced.

end-of-file mark—In a computer, a code instruction indicating that the last record of a file has been read.

end-on armature—Of a relay, an armature which moves in the direction of the core axis, with the pole face at the end of the core and perpendicular to this axis.

end-on directional antenna—A directional antenna which radiates chiefly toward the line on which the antenna elements are arranged.

endoradiosonde—Also called radio pill. A device for detecting and transmitting physiological data from the gastrointestinal tract or other inaccessible body cavities.

endothermic—A term describing a chemical reaction in which heat is absorbed.

end point—The shaft positions immediately before the first and after the last measurable change(s) in output ratio after wiper continuity has been established, as the shaft of a precision potentiometer moves in a specified direction.

end-point sensitivity—The algebraic difference in electrical output between the maximum and minimum value of the measurand over which an instrument is calibrated.

end-point voltage—The terminal voltage of a cell below which equipment connected to it will not operate or should not be operated.

end resistance—The resistance of a precision potentiometer measured between the wiper terminal and an end terminal, with the shaft positioned at the corresponding end point.

end-resistance offset — In potentiometers, the residual resistance between a terminal and the moving contact, at a position corresponding to full rotation against that terminal.

end scale value—The value of the actuating electrical quantity that corresponds to end scale indication of an instrument. When zero is not at the end or at the electrical center of the scale, the higher value is taken. Certain instruments such as power-factor meters, ohmmeters, etc., are necessarily excepted from this definition.

end setting—In a potentiometer, the minimum resistance that is measured between one end of a potentiometer and the wiper, with the wiper mechanically positioned at that end.

end shield—1. Frequently called end bracket or end bell. In a motor housing, the part that supports the bearing and also guards the electrical and rotating parts inside the motor. 2. In a magnetron, the shield that confines the space charge to the interaction space.

end spaces—In a multicavity magnetron, the two cavities at either end of the anode block which terminate all the anode-block cavity resonators.

end use — The way the ultimate consumer uses a device.

energize—To apply the rated voltage to a circuit or device, such as to the coil of a relay, in order to activate it.

energized—Also called alive, hot, and live. Electrically connected to a voltage source.

energy—The capacity for performing work. A particle or piece of matter may have energy because it is moving or because of its position in relation to other particles or pieces of matter. A rolling ball is an example of the first; a ball at rest at the top of an incline, an example of the second.

energy conversion — The change of energy from one form to another, e.g., from chemical energy to electrical energy.

energy conversion devices—Devices including primary and secondary cells; fuel cells;

photovoltaic systems; electrochemical energy converters; radiation conversion devices; thermionic converters; converters using solar, ionic, or nuclear energy sources; devices for creating a plasma in an interaction space between an emitter and a collector; electrostatic generators for creating an electrical output; organic and inorganic ion exchange and membrane devices; electron-volt energy devices; devices for direct conversion of fuel to electricity; and electrical-energy storage-unit devices capable of delivering a power output.

energy density—The ratio of the energy available from a cell to the weight or volume of the cell.

energy gap—The energy range between the bottom of the conduction band and the top of the valence band of a semiconductor.

energy-level diagram—A line drawing that shows the increase or decrease in electrical power as current intensities rise and fall along a channel of signal communications.

energy-measuring equipment—Equipment used to measure energy in electrical, electronic, acoustical, or mechanical systems.

energy of a charge—Represented by $E = \frac{1}{2}QV$, given in ergs, when the charge Q and the potential V are in electrostatic units.

energy-product curve—A curve obtained by plotting the product of the value of magnetic induction B and demagnetizing force H for each point of the demagnetization curve of a permanent magnetic material. Usually shown together with the demagnetization curve.

energy redistribution—A method of finding the duration of an irregularly shaped pulse by considering it as a power curve. The area under the curve can be represented by an equivalent rectangle of the same area and peak amplitude. The original-pulse duration is equal to the rectangle width.

energy state—The position and speed of an electron relative to the position and speed of other electrons in the same atom or adjoining atoms.

engineering—A profession in which a knowledge of the natural sciences is applied with judgment to develop ways of utilizing the materials and forces of nature.

English-reading tube tester—*See* Emission-Type Tube Tester.

enhanced carrier demodulation—An amplitude-demodulation system in which a synchronized local carrier of the proper phase is added to the demodulator. This has the effect of materially reducing the distortion produced in the demodulation process.

entropy—1. A measure of the unavailable energy in a thermodynamic system. 2. The unavailable information in a set of documents. 3. An inactive or static condition (total entropy).

entry—Each statement in a computer programming system.

envelope—1. The glass or metal housing of a vacuum tube. 2. The curve passing through the peaks of a graph and showing the waveform of a modulated radio-frequency carrier signal.

envelope delay—The time which elapses as a transmitted wave passes any two points of a transmission circuit. Such delay is determined primarily by the constants of the circuit and is measurable in milliseconds or microseconds.

envelope-delay distortion—The distortion that occurs during transmission when the phase shift of a circuit or system is not constant over the frequency range.

environmental conditions—External conditions of heat, shock, vibration, pressure, moisture, etc.

environmentally sealed—Provided with gaskets, seals, potting, or other means to keep out contamination which might reduce performance.

environmental testing—The testing of a system or component under controlled environmental conditions, each of which tends to affect its operation or life.

environmentproof switch—A switch which is completely sealed to insure constant operating characteristics. Sealing normally includes an "O" ring on the actuator shaft and fused glass-to-metal terminal seals or complete potting and an elastomer plunger-case seal.

episcotister—A device consisting of alternate opaque and transparent discs which rotate at a speed which interrupts light beams at an audio-frequency rate. It modulates the light beam used to excite a photoelectric element.

epitaxial—Pertaining to a single crystal layer on a crystalline substrate, oriented the same as the substrate. In certain semiconductor processes, an epitaxial layer is grown on a silicon substrate during the fabrication of transistors and integrated circuits.

epitaxial device—A device constructed in such a manner that the crystalline structure of successive layers is oriented in the same direction as that of the original base material.

epitaxial film—1. A film of single-crystal semiconductor material that has been deposited onto a single-crystal substrate. 2. Any deposited film, provided the orientation of its crystal is the same as that of the substrate material.

epitaxial growth—A chemical reaction in which silicon is precipitated from a gaseous solution, and grows on the surface of a silicon wafer present in the gaseous solution.

epitaxial-growth mesa transistor—A transistor made by overlaying a thin mesa crystal over another mesa crystal.

epitaxial growth process—The process of growing a semiconductor material by depositing it in vaporized form on a semiconductor seed crystal. The deposited layer continues the single-crystal structure of the seed.

epitaxial layer—A grown or deposited crystal layer with the same crystal orientation as the parent material and, in the case of semiconductor circuits, is also the same basic material as the original substrate.

epitaxial planar transistor—A transistor in which a thin collector region is epitaxially deposited on a low-resistivity substrate, and the base and emitter regions are produced by gaseous diffusion with the edges of the junction under a protective oxide mask.

epitaxial process—The process of growing from the vapor phase a single-crystal semiconductor material with controlled resistivity and thickness.

epitaxial transistor—A transistor with one or more epitaxial layers.

epitaxy—The controlled growth of a layer of semiconductor material in a suitable substrate.

E plane—The plane of an antenna containing the electric field. The principal E plane also contains the direction of maximum radiation.

E-plane bend—Also called E-bend. The smooth change in direction of the axis of a waveguide. The axis remains parallel to the direction of polarization throughout the change.

E-plane T-junction—Also called series T-junction. A waveguide T-junction in which the structure changes in the plane of the electric field.

epoxy resin—An insulating plastic used for encapsulating or embedding electronic equipment.

epsilon—The greek letter E, or ε, frequently used to represent 2.71828, which is the base of the natural system of logarithms.

equal-energy source—A light source having a constant emission rate of energy per unit of wavelength throughout the visible spectrum.

equal-energy white—The light produced by a source which radiates equal energy at all visible wavelengths.

equalization—The process of reducing the frequency and/or phase distortion of a circuit by the introduction of networks to compensate for the difference in attenuation and/or time delay at the various frequencies in the transmission band.

equalizer—1. A passive device designed to compensate for an undesired amplitude-frequency and/or phase frequency characteristic of a system or component. 2. A network having an attenuation complementary to that of a telephone line and inserted for the purpose of correcting frequency distortion caused by the line. 3. A network which improves the frequency response of a radio system. Usually a combination of resistors, inductors, and capacitors. 4. A series of connections made in paralleled, cumulatively compound direct-current generators to give the system stability.

equalizer circuit breaker—A breaker which serves to control or to make and break the equalizer or the current-balancing connections for a machine field, or for regulating equipment, in a multiple-unit installation.

equalizing current—A current circulated between two parallel-connected compound generators to equalize their output.

equalizing network—A network connected to a line to correct or control its transmission frequency characteristics.

equalizing pulses—A series of pulses (usually six) occurring at twice the line frequency before and after the serrated vertical tv synchronizing pulse. Their purpose is to cause vertical retrace to occur at the correct instant for proper interlace.

equal-loudness contours—*See* Fletcher-Munson Curves.

equation solver—A computer, usually of the analog type, designed to solve systems of linear simultaneous (nondifferential) equations or to find the roots of polynomials.

equiphase surface—In a wave any surface over which the field vectors at the same instant are either in phase or 180° out of phase.

equiphase zone—In radionavigation, the region in space within which the difference in phase between two radio signals is indistinguishable.

equipment—An item having a complete function apart from being a substructure of a system. Sometimes called a set.

equipment life—The arithmetic mean of the cumulative operating times of identical pieces of equipment beginning with the time of acceptance by the ultimate consumer and ending when the equipment is no longer serviceable.

equipotential—A conductor having all parts at a single potential. The cathode of a heater-type tube is equipotential, whereas the filament is not because its voltage varies from one end to the other.

equipotential cathode—*See* Indirectly Heated Cathode.

equipotential line—An imaginary line in space having the same potential at all points.

equipotential surface—A surface or plane passing through all points having the same potential in a field of flow.

equisignal localizer—Also called tone localizer. A type of localizer in which lateral guidance is obtained by comparing the amplitudes of two modulation frequencies.

equisignal radio-range beacon—A radio-range beacon used for aircraft guidance. It transmits two distinctive signals, which are received with equal intensity only in certain directions called equisignal sectors.

equisignal surface—The surface formed around an antenna by all points which have a constant field strength (usually measured in volts per meter) during transmission.

equisignal zone—In radionavigation, the region in space within which the difference in amplitude between two radio signals is indistinguishable.

equivalence—A logic operator having the property that if P is a statement, Q is a statement, R is a statement, etc., then the equivalence of P,Q,R . . . is true if and only if all statements are true or all statements are false.

equivalent absorption—The rate at which a surface will absorb sound energy, expressed in sabins. Defined as the area of a perfect absorption surface that will absorb the same sound energy as the given object under the same conditions.

equivalent binary digits—The number of binary digits equivalent to a given number of decimal digits or other characters.

equivalent circuit—An arrangement of common circuit elements with electrical characteristics, equivalent to those of a more complicated circuit or device.

equivalent circuit of a piezoelectric crystal unit—The electric circuit which has the same impedance as the unit in the frequency region of resonance. It is usually represented by an inductance, capacitance, and resistance in series, shunted by the direct capacitance between the terminals of the crystal unit.

equivalent component density—In circuits where discrete components are not readily identifiable, the volume of the circuit divided by the number of discrete components necessary to perform the same function.

equivalent conductance—The normal conductance of an ATR tube in its mount, measured at its resonance frequency.

equivalent dark-current input—The incident luminous flux required to give an output current equal to the dark current.

equivalent differential input capacitance—The equivalent capacitance looking into the inverting or noninverting inputs of a differential amplifier with the opposite input grounded. *Also see* Equivalent Differential Input Impedance.

equivalent differential input impedance—The equivalent impedance looking into the inverting or noninverting input, with the opposite input grounded and the operational amplifier operated in the linear amplification region.

equivalent differential input resistance—The equivalent resistance looking into the inverting or noninverting input of a differential amplifier with the opposite input grounded. *Also see* Equivalent Differential Input Impedance.

equivalent diode—An imaginary diode consisting of the cathode of a triode or multigrid tube and a virtual anode to which is applied a composite controlling voltage of such a value that the cathode current would be the same as the current in the triode or multigrid tube.

equivalent four-wire system—A transmission system using frequency division to obtain full-duplex operation over only one pair of wires.

equivalent grid voltage—The grid voltage plus plate voltage, divided by the mu of the tube.

equivalent input offset current—The difference between the two currents flowing into the inverting and noninverting inputs of a differential amplifier when the output voltage is zero.

equivalent input offset voltage—The amount of voltage required at the input to bring the output to zero. Usually this voltage is adjustable to zero by using either a built-in or an external variable resistor (balance control).

equivalent input wideband noise voltage—The output noise voltage of a differential amplifier with the input shorted, divided by the d-c voltage gain of the amplifier. This voltage is measured with a true rms voltmeter and is limited to the combined bandwidth of the amplifier and meter.

equivalent height—The virtual height of an ionized layer of the ionosphere.

equivalent loudness—The intensity level of a sound relative to some arbitrary reference intensity, such as a 1000-hertz pure tone which is judged by the listeners to be equivalent in loudness.

equivalent network—A network which may replace another network without substantial change in the operation of the system.

equivalent noise input—In a photosensitive device, the value of incident luminous flux which produces an rms output current equal to the rms noise current within a specified bandwidth when the flux is modulated in a stated manner.

equivalent noise pressure—*See* Transducer Equivalent Noise Pressure.

equivalent noise resistance—A measure of the residual noise output of a potentiometer while the slider is being actuated. (The residual noise consists of active components in the form of self-generated voltages arising in the slider contact interface, and passive components in the form of ohmic contact resistance at the point of slider contact.)

equivalent noise temperature—The absolute temperature at which a perfect resistor with the same resistance as the component would generate the same noise as the component at room temperature.

equivalent permeability—The relative permeability which a component would have under specified conditions if it had the same reluctance as a component of the same shape and size but different materials.

equivalent plate voltage—The plate voltage plus mu times the grid voltage.

equivalent resistance—The concentrated or lumped resistance that would cause the same power loss as the actual small resistances distributed throughout a circuit.

equivalent series resistance—In a circuit, the square root of the difference between the impedance squared and the reactance squared.

equivocation—In a computer, the conditional information contained in an input symbol given an output symbol, averaged over all input-output pairs.

erasable storage—Storage media in a computer which holds information that can be changed.

erase—To replace all the binary digits in a storage device by binary zeros. In a binary computer, erasing is equivalent to clearing, while in a coded decimal computer where the pulse code for decimal zero may contain binary ones, clearing leaves decimal zero while erasing leaves all-zero pulse codes in all storage locations.

erase oscillator—In a magnetic recorder an oscillator that provides the high-frequency signal necessary for erasing a magnetic tape.

erasing head—A device for obliterating any previous recordings. It may also be used to precondition magnetic media for recording purposes.

erasing speed—In charge-storage tubes, the rate of erasing successive storage elements.

erasure—A process by which a signal recorded on a tape is removed and the tape made ready for rerecording. This may be accomplished by a-c erasure, where the tape is demagnetized by an alternating field which is reduced in amplitude from an initially high value, or by d-c erasure, where the tape is saturated by applying a primarily unidirectional field.

E-region—The region of the ionosphere about 50 to 100 miles above the earth's surface.

$E_r - E_y$—The resultant color television signal when E_y is subtracted from the original full red signal.

erg—The absolute centimeter-gram-second unit of energy and work. The work done when a force of one dyne is applied through a distance of one centimeter.

error—1. In mathematics, the difference between the true value and a calculated or observed value. A quantity (equal in absolute magnitude to the error) added to a calculated or observed value to obtain the true value is called a correction. 2. In a computer or data-processing system, any incorrect step, process, or result. In addition to the mathematical usage in the computer field, the term also commonly refers to machine malfunctions, or "machine errors," and to human mistakes, or "human errors."

error-correcting code—A code in which each acceptable expression conforms to specific rules of construction that also define one or more equivalent nonacceptable expressions, so that if certain errors occur, in an acceptable expression the result will be one of its equivalents and thus the error can be corrected.

error-correction routine—A series of computer instructions programmed to correct a detected error condition.

error detecting and feedback system—A system employing an error-detecting code and so arranged that a signal detected as being in error automatically initiates a request for retransmission of the signal detected as being in error.

error-detecting code—In a digital computer, a system of coding characters such that any single error produces a forbidden or impossible code combination.

error detector—That portion of an automatic control system that determines when the regulated quantity has deviated outside the dead zone.

error-rate damping—A damping method in which a signal proportional to the rate of change of error is added to the error signal for anticipatory purposes.

error signal—In an automatic control device, a signal whose magnitude and sign are used to correct the alignment between the controlling and the controlled elements.

error tape—A special tape developed and used for writing out errors in order to correct them by study and analysis after printing.

error voltage—A voltage which is present in a servo system when the input and output shafts are not in correspondence. The error voltage, which actuates the servo system, is proportional to the angular displacement between the two shafts.

Esaki diode—*See* Tunnel Diode.

E-scape—A radar display in which targets appear as blips, with distance indicated by the horizontal coordinate and elevation by the vertical coordinate.

escapement—A mechanical stepping device that stores energy over a predetermined time interval and delivers this stored energy at the end of said time interval to advance or move a camshaft a predetermined number of degrees.

escape velocity—The speed which a particle or larger body must attain in order to escape from the gravitational field of a planet or star.

escutcheon—A backing plate around an opening. Commonly, the ornamental metal, wood, plastic, or other framework around a radio tuning dial, control knob, or other panel-mounted part in a radio receiver or television receiver, audio-frequency amplifier, etc.

Estiatron—A special type of electrostatically focused traveling-wave tube.

etchant—In the preparation of a printed-circuit board, a solution used to remove the unwanted portion of the conductive material by means of a chemical reaction.

etched printed circuit—A type of printed circuit formed by chemically or electrolytically (or both) removing the unwanted portion of a layer of material bonded to a base.

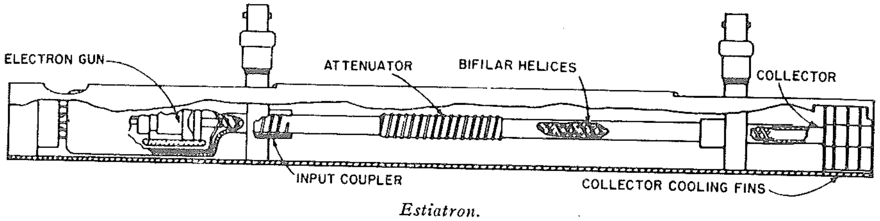

Estiatron.

etching to frequency—Finishing a crystal blank to its final frequency by etching it in hydrofluoric acid.

ether—A hypothetical medium that pervades all space (including vacuum) and all matter and is assumed to be the vehicle for propagation of electromagnetic radiations.

E transformer—A special form of differential transformer employing an E-shaped core. The secondaries of the transformer are wound on the outer legs of the E, and the primary is on the center leg.

Ettingshausen effect—Analogous to the Hall effect. The different temperatures found on opposite edges of a metal strip which is perpendicular to a magnetic field and through which an electric current flows longitudinally.

eureka—The ground transponder of secondary radar system rebecca-eureka.

ev or eV—Abbreviation for electron volt.

evaporation of electrons—The cooling which occurs on the surface of a cathode during emission. It is analogous to the cooling of a liquid or solid as it evaporates.

E-vector—A vector representing the electric field of an electromagnetic wave. In free space it is perpendicular to the direction of propagation.

even harmonic—Any harmonic that is an even multiple (2, 4, 6, etc.) of the fundamental frequency.

E-wave—Designation for TM (transverse magnetic) wave, one of the two classes of electromagnetic waves that can be sent through waveguides.

exalted-carrier receiver—A receiver that counteracts selective fading by maintaining the carrier at a high level at all times.

exalted-carrier reception—A method of receiving either amplitude- or phase-modulated signals in which the carrier is separated from the sidebands, filtered and amplified, and then recombined with the sidebands at a higher level prior to demodulation.

except gate—A gate in which the specified combination of pulses producing an output pulse is the presence of a pulse on one or more input lines and the absence of a pulse on one or more other input lines.

excess conduction—Conduction by excess electrons in a semiconductor.

excess electron—An electron introduced into a semiconductor by a donor impurity and available to promote conduction. An excess electron is not required to complete the bond structure of the semiconductor.

excess fifty—In a computer, a representation in which a number (N) is denoted by the equivalent of (N + 50).

excess meter—An electricity meter which measures and registers the integral, with respect to time, of those portions of the active power in excess of the predetermined value.

excess minority carriers—In a semiconductor, the number of minority carriers that exceed the normal equilibrium number.

excess modified index of refraction—*See* Refractive Modulus.

excess noise—Noise resulting from the passage of current through a semiconductor material. Also called current noise, bulk noise, and 1/f noise.

excess sound pressure—The total instantaneous pressure at a point in a medium containing sound waves, minus the static pressure when no sound waves are present. The unit is the dyne per square centimeter.

excess-three code—A number code in which the decimal digit N is represented by the four-bit binary equivalent of n + 3. (*Also see* Binary-Coded Decimal.)

exchange—To remove the contents of one storage unit of a computer and place it in a second, at the same time placing the contents of the second storage unit into the first.

exchange register—*See* Memory Register.

excitation—1. Also called stimulus. An external force or other input applied to a system to cause it to respond in some specified way. 2. Also called drive. A signal voltage applied to the control electrode of an electron tube. 3. In electric or electromagnetic equipment, supplying with a potential, a charge, or a magnetic field. 4. The addition of energy to a system so as to transfer the system from its ground state to an excited state.

excitation anode—An auxiliary anode of a pool-cathode tube, used to maintain a cathode spot when the output current is zero.

excitation current—The resultant current in the shunt field of a motor when voltage is applied across the field.

excitation energy—The external electrical energy required for proper operation of a transducer.

excitation purity—Also called purity. The ratio between the distance from the reference point to the point representing the sample and the distance along the same straight line from the reference point to the spectrum locus or to the purple boundary, both distances being measured (in the

same direction from the reference point) on the CIE chromaticity diagram.

excitation voltage—The voltage required for excitation of a circuit.

excited-field speaker—A speaker in which the steady magnetic field is produced by an electromagnet.

exciter—1. In a directional transmitting antenna system, the part connected directly to the source of power such as to the transmitter. 2. A crystal or self-excited oscillator that generates the carrier frequency of a transmitter. 3. A small, auxiliary generator that provides field current for an a-c generator.

exciter lamp—1. A high-intensity incandescent lamp having a concentrated filament. It is used in making variable-area, sound-on-film recording and in reproducing all types of sound tracks on film, as well as in some mechanical television systems. 2. A light source used in a facsimile transmitter to illuminate the subject copy being scanned.

exciter or d-c generator relay—A device which forces the d-c machine-field excitation to build up during starting, or which functions when the machine voltage has built up to a given value.

exciting current—1. The current which flows in the primary of a transformer when the secondary is open-circuited. This current produces a flux that generates a back emf equal to the applied voltage. 2. Also called magnetizing current. The current that passes through the field windings of a generator.

exciting current of a transformer—Also called no-load current. The current used in any winding to excite the transformer when all other windings are open-circuited, usually expressed in percent of the rated current of the windings in which it is measured.

exciton—A bound electron-hole pair in an insulator or a semiconductor. It may move through the crystal transporting excitation energy, but does not contribute to electric conductivity since it is neutral.

excitron—A single-anode pool tube with means for maintaining a continuous cathode spot.

exclusion principle—The principle which states that if particles are considered to occupy quantum states, then only one particle of a given kind can occupy any one state. Particles differ in kind due to their direction of spin, momentum, orbit, etc.

exclusive OR—A logic element with two inputs and one output. The output is true if either of the two inputs is true, but *not* if both are true.

excursion—A single movement away from the mean position in an oscillating or alternating motion.

execution—The performance of an operation or instruction.

executive routine—In digital-computer programming, a routine designed to process and control other routines.

exhaustion—The removal of gases from a space, such as the bulb of a vacuum tube, by means of vacuum pumps.

exit—In a computer, the means of halting a repeated cycle of operation in a program.

exosphere—The outermost region of the earth's atmosphere, where the atoms and molecules move in dynamic orbits under the influence of the gravitational field.

exothermic—A chemical reaction in which heat is produced.

expand—To spread out part or all the trace of a cathode-ray display.

expanded scale meter—A meter in which the ratio of deflection per unit of applied energy becomes greater as the energy approaches a specified value.

expanded scope—A magnified portion of a cathode-ray-tube presentation.

expanded sweep—A preselected portion of a sweep, during which time the electron beam is speeded up in a cathode-ray tube.

expander—A transducer which produces a larger range of output voltages for a given amplitude range of input voltages. One important type expands the volume range of speech signals by employing their envelope.

expansion—A process in which the effective gain applied to a signal is increased for larger signals and decreased for smaller signals.

experimental model—An equipment model that demonstrates the technical soundness of a basic idea but does not necessarily have the same form or parts as the final design.

experimental station—Any station (except amateur) utilizing electromagnetic waves between 10 kHz and 3000 GHz in experiments, with a view toward the development of a science or technique.

exploring coil—*See* Magnetic Test Coil.

explosion-proof motor—A motor designed and constructed so as to withstand an internal explosion of a specified gas or vapor and to prevent the ignition of the specified gas or vapor surrounding the motor by sparks, flashes, or explosions of the specified gas or vapor inside the motor casing.

explosive atmosphere—The condition where air is mixed with dust, metal particles, or inflammable gas in such proportion that it is capable of igniting or exploding.

exponential—Pertaining to exponents or to an expression having exponents. A quantity that varies in an exponential manner increases by the square or some other power of a factor, instead of linearly.

exponential curve—A curve representing the variation of an exponential function.

exponential damping—Damping which follows an exponential law.

exponential decay—The decay of signal strength, radiation, charge, or some other quantity at an exponential rate.

exponential horn—A horn the cross-sectional area of which increases exponentially with axial distance.

exponential quantity—A single quantity which increases or decreases at the same rate as the quantity itself (e.g., the discharge current of a capacitor through a noninductive resistor).

exponential sweep—An electron-beam sweep which starts rapidly and slows down exponentially.

exponential transmission line—A two-conductor transmission line the characteristic impedances of which vary exponentially with the electrical length of the line.

exponential waveform—A waveform which is characterized by smooth curves but which possesses pulse properties because it contains numerous constituent frequencies. The exponential waveform undergoes a rate of amplitude change which is either inversely or directly proportional to the instantaneous amplitude.

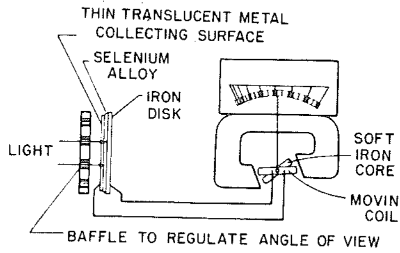
THIN TRANSLUCENT METAL COLLECTING SURFACE
SELENIUM ALLOY
IRON DISK
LIGHT
SOFT IRON CORE
MOVING COIL
BAFFLE TO REGULATE ANGLE OF VIEW

Exposure meter.

exposure meter—In photography, an instrument which measures scene brightness and indicates proper lens opening and exposure time.

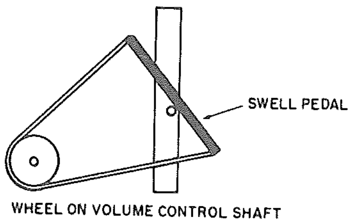

Expression control.

expression control—In an organ the control which regulates the over-all volume. Usually operated with the right foot.

extended-cutoff tube—*See* Remote-Cutoff Tube.

extended octaves—In an organ, tones above or below the notes on the regular keyboard which can be sounded only when certain couplers are on.

extended play—Abbreviated ep. A 45-rpm record on a seven-inch disc. It provides eight minutes of playing time, instead of the five minutes of a standard 45-rpm disc.

external armature—A ring-shaped armature that rotates around the field magnets of a generator or motor.

external circuit—All wires and other conductors which are outside the source.

external control devices—All control devices mounted external to the control panel.

external critical damping resistance—The value of resistance that must be placed in series or in parallel with a galvanometer in order to produce the critically damped condition.

external feedback—In a magnetic amplifier, the ampere-turns of auxiliary windings which assist in equalizing the load winding ampere-turns, thereby reducing the control ampere-turns required for control. External feedback may be degenerative to reduce the gain and improve the stability of an amplifier.

externally adjustable timer—A time delay which can be adjusted by varying the resistance of the timer resistor externally, usually by means of a screwdriver.

externally caused contact chatter—That chatter resulting from shock or vibration imposed on the relay by external action.

externally caused failure—A failure caused by an environment outside the design limitations, such as excessive loads, voltages, etc., resulting from operator error, accident, or failure of another part.

externally quenched counter tube—A radiation counter tube that requires an external circuit to prevent it from reigniting.

external memory—A storage unit outside the computer (e.g., a magnetic tape).

external Q—In a microwave tube, the reciprocal of the difference between the reciprocal of the loaded Q and the reciprocal of the unloaded Q.

external storage—Storage facilities separate from the computer itself but holding information in a form acceptable to the computer, e.g., magnetic tapes, magnetic wires, punched cards, etc.

extinction potential—The lowest value to which the plate voltage of a gaseous tube can be reduced without cutting off the flow of plate current.

extra-class license—The highest classification of United States amateur license. Requirements include a code sending and receiving ability of 20 words per minute, a knowledge of advanced theory, and the holding of a general- or conditional-class license for two years.

extract—To remove from a set of items of information all those items that meet some arbitrary criterion.

extract instruction—In a digital computer, the instruction to form a new word by placing selected segments of given words side by side.

extractor—*See* Filter.

extraneous response—Any undesired response of a receiver, recorder, or other susceptible device, due to the desired signals, undesired signals, or any combination or interaction among them.

extraordinary wave component—The magnetoionic wave component in which the electric vector rotates in the opposite sense from that for the ordinary wave component.

extrapolate—To estimate the value of a function for variables lying outside the range in which values of the function are known (e.g., to extend the graphs of the function beyond the plotted points).

extremely high frequency—Abbreviated ehf. The frequency band extending from 30 to 300 GHz.

extremely low frequency—A frequency below 300 hertz. Abbreviated elf.

extrinsic base resistance-collector capacitance product—The product of the base resistance and collector capacitance of a transistor. It is expressed in units of time, since it is an RC time constant and affects the high-frequency operation of a transistor.

extrinsic conductance—The conductance resulting from impurities or external causes.

extrinsic properties—The properties of a semiconductor, modified by impurities or imperfections within the crystal.

extrinsic semiconductor—A semiconductor the electrical properties of which depend on its impurities.

eyelet—A tubular metal piece having one end (and possibly the second) headed or rolled over at a right angle.

F

f—1. Abbreviation for farad, femto (10^{-15}). 2. Symbol for focal length, frequency.

F—1. Symbol for filament, fuse. 2. Abbreviation for Fahrenheit.

F−—*See* A−.

F+—*See* A+.

fa—Abbreviation for femtoampere (10^{-15} ampere).

fabrication holes—*See* Pilot Holes.

fabrication tolerance—In the construction or assembly of an equipment or portion of an equipment, the maximum variation in the characteristics of a part which, considering the defined variations of the other parts in the equipment, will permit the equipment to operate within specified performance limits.

Fabry-Perot interferometer—A resonant cavity bounded by two end mirrors separated so as to produce interference for certain allowed optical frequencies. The enhancements and cancellations produced by the internal reflections are, in fact, optical standing waves. The structure can be used as a laser cavity, mode filter, or frequency selector.

face—1. A plane surface on a crystal which stands in a particular and invariable relation to the axes and planes of reference and to other faces. 2. Front, or viewing, surface of a cathode-ray tube. 3. The portion of a meter bearing the scale markings.

face bonding—A method of attaching active devices to thin-film passive networks. The semiconductor chips are provided with small mounting pads turned face down and bonded directly to the end of the thin-film conductors on the passive substrate.

faced crystal—A single or twinned mass of quartz bounded in part or entirely by the original crystal growth faces.

face-parallel cut—A Y-cut for a quartz crystal.

face-perpendicular cut—An X-cut for a quartz crystal.

facom—A long-distance measuring or radio-navigational system which derives information of distances by comparing the phases of received and locally generated signals. It is a base-line system operating in the low-frequency band and will work under adverse propagation and noise conditions at ranges of up to 3000 miles from the signal source.

facsimile—A process or the result of a process by which fixed graphic material including pictures or images is scanned and the information converted into electrical signal waves, which are used either locally or remotely to produce in record form a likeness (facsimile) of the subject copy.

facsimile broadcast station—A station licensed to transmit images of still objects for reception by the general public.

facsimile posting—The process of transferring, by a duplicating process, a printed line on a report to a ledger or other recorded sheet. These may be posted from a transaction listing previously prepared on an accounting machine. A data-processing function.

facsimile receiver—An apparatus that translates the facsimile signal into a reproduced image.

facsimile recorder—Apparatus which reproduces on paper the image transmitted by a facsimile system.

facsimile signal—The signal resulting from the scanning in a facsimile system.

facsimile-signal level—An expression of the maximum signal power or voltage created by the scanning of the subject copy as measured at any point in a facsimile system. According to whether the system employs positive or negative modulation, this will correspond to picture white or black, respectively. It may be expressed in decibels with respect to some standard value, such as 1 milliwatt or 1 volt.

facsimile system—An integrated assembly of the elements used for facsimile transmission and reception.

facsimile transmission—The transmission of signal waves produced by the scanning of a

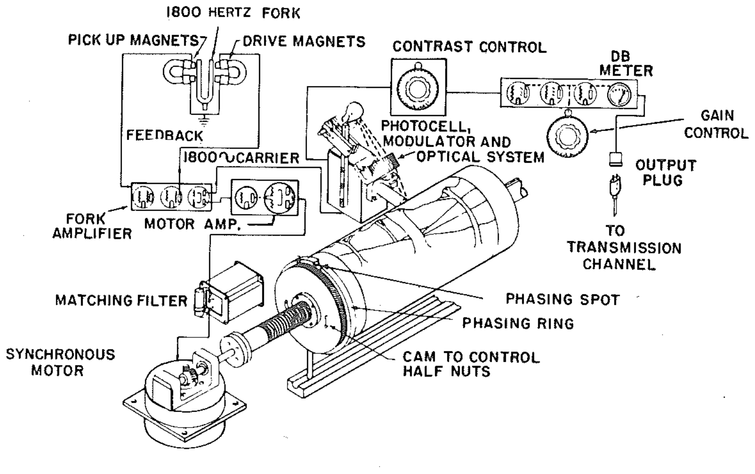

Facsimile transmitter.

picture on a revolving drum by a photoelectric cell for reproduction in permanent form at the receiver.

facsimile transmitter—An apparatus employed to convert the subject copy into suitable facsimile signals. Usually, the copy is wrapped around a revolving drum and scanned by a photoelectric cell, which converts the darks and lights into corresponding signal amplitudes.

fade—The gradual lowering in amplitude of a signal.

fade in—To increase the signal strength gradually in a sound or television channel.

fade out—1. The gradual decrease in signal strength in a sound or television channel. 2. The cessation or near cessation of radio-wave propagation through parts of the ionosphere due to a sudden atmospheric disturbance.

fader—A multiple-unit control used in radio for gradual changeover from one microphone or audio channel to another; in television, from one camera to another; and in motion-picture projection, from one projector to another.

fading—The variation of radio field strength caused by a gradual change in the transmission medium.

fading margin—The number of decibels of attenuation which can be added to a specified radio-frequency propagation path before the signal-to-noise ratio of the channel falls below a specified minimum.

Fahnestock clip—A spring-type terminal to which a temporary connection can readily be made.

Fahrenheit temperature scale—A temperature scale in which the freezing point of water is defined as 32° and the boiling point as 212° under normal atmospheric pressure (760 mm of mercury).

fail hardover — Failure which results in a steady-state maximum system output with no input signal.

fail-safe circuit—A circuit that has an output state which indicates that either a circuit input or the circuit itself has failed. Finds application in complex systems where self-healing subsystems exist. When a subsystem failure is detected, a back-up subsystem is automatically inserted in its place.

fail-safe control—A system of remote control for preventing improper operation of the controlled function in event of circuit failure.

fail-safe operation—An electrical system so designed that the failure of any component in the system will prevent unsafe operation of the controlled equipment.

fail soft—An active failure which can be compensated for by the system or one where the operator can safely assume control.

failure—The inability of a system, subsystem, component, or part to function in the required or specified manner.

failure indicator—The observed characteristic that shows that an item is defective.

failure mechanism—*See* Failure Mode.

failure mode—The manner in which a failure occurs, including the operating condition of the equipment or part at the time of the failure. Also called failure mechanism.

failure rate—Also called hazard. The probability of failure, normally expressed in per cent per 1000 hours, of the items still operating.

fall-in—In a synchronous motor, the point at which synchronous speed is reached.

fallouts—*See* Transistor Seconds.

fall time—1. The length of time during which a pulse is decreasing from 90% to 10% of its maximum amplitude. 2. A measure of the time required for a circuit to change its output from a high level (1) to a low level (0).

false course—In navigation normally providing one or more course lines, a spurious additional course-line indication due to undesired reflections or maladjusted equipment.

false-echo device — A device for producing an echo different from that normally observed.

fan antenna—*See* Double-V Antenna.

fan beam—A field pattern having an elliptically shaped cross section in which the ratio of major to minor axes usually exceeds 3 to 1.

fan-in—The number of inputs that can be connected to a logic circuit.

fan-in circuit—A circuit that has many inputs feeding to a common point.

fan marker—A radio signal having a vertically directed fan beam which tells the pilot the location of his aircraft while flying along a radio range.

fanned-beam antenna—A unidirectional antenna so designed that transverse cross sections of the major lobe are almost elliptical.

fan-out — The number of parallel loads within a given logic family that can be driven from one output mode of a logic circuit.

fan-out circuit—A circuit that has a single output point which feeds many branches.

farad — The capacitance of a capacitor in which a charge of 1 coulomb produces a change of 1 volt in the potential difference between its terminals. The farad is the unit of capacitance in the mksa systems.

faraday—1. A unit equal to the number of coulombs (96,500) required for an electrochemical reaction involving one electrochemical equivalent. 2. In an electrolytic process, 1-gram equivalent weight of matter is chemically altered at each electrode for 1 faraday of electricity passed through the electrolyte.

Faraday cage—*See* Faraday Shield.

Faraday dark space—The relatively nonluminous region between the negative glow and the positive column in a glow-discharge–cold-cathode tube. (*Also see* Glow Discharge.)

Faraday effect—The rotation of the plane of polarization of radio waves as they pass through the ionosphere in the earth's magnetic field. This effect produces a decoupling loss between linearly polarized antennas.

Faraday rotation—The rotation of the plane of polarization of radiation when it passes through a plasma containing a magnetic field having a component in the direction of propagation.

Faraday screen—*See* Faraday Shield.

Faraday shield—Also called Faraday screen or Faraday cage. A network of parallel wires connected to a common conductor at one end to provide electrostatic shielding without affecting electromagnetic waves. The common conductor is usually grounded.

Faraday's laws—1. The mass of a substance liberated in an electrolytic cell is proportionate to the quantity of electricity passing through the cell. 2. When the same quantity of electricity is passed through different electrolytic cells, the masses of the substances liberated are proportionate to their chemical equivalents. 3. Also called the law of electromagnetic induction. 4. When a magnetic field cuts a conductor, or when a conductor cuts a magnetic field, an electric current will flow through the conductor if a closed path is provided over which the current can circulate.

faradic current—An intermittent and nonsymmetrical alternating current like that obtained from the secondary winding of an induction coil.

faradmeter—An instrument for measuring electric capacitance.

far-end cross talk—Cross talk which travels along the disturbed circuit in the same direction as the signals in that circuit. To determine the far-end cross talk between two pairs, 1, and 2, signals are transmitted on pair 1 at station A and the cross-talk level is measured on pair 2 at station B.

far field—The space beyond the near field of an antenna in which radiation is essentially confined to a fixed pattern and power density along the axis of the pattern falls off inversely with the square of the distance.

far-field region—That region of the field of an antenna where the angular field distribution is essentially independent of the distance from the antenna.

Farnsworth image-dissector tube—A special cathode-ray tube for use in television cameras.

fast-access storage—In a computer memory or storage, the section from which information may be obtained most rapidly.

fastener—A device used to secure a conductor (or other object) to the structure which supports it.

fast forward—The provision on a tape recorder permitting tape to be run rapidly through the recorder in the play direction, usually for search or selection purposes.

fast-forward control—A tape-recorder control which permits running the tape through the machine rapidly in the forward direction.

fast groove—Also called fast spiral. In disc recording, an unmodulated spiral groove having a much greater pitch than the recorded grooves.

fast-operate–fast-release relay — A high-speed relay designed specifically for both short-operate and short-release times.

fast-operate relay—A high-speed relay designed specifically for short-operate but not short-release time.

fast-operate–slow-release relay—A relay designed specifically for short-operate and long-release times.

fast-release relay—A high-speed relay designed specifically for short-release but not short-operate time.

fast spiral—*See* Fast Groove.

fast time constant—An antijamming device used in radar video-amplifier circuits. It differentiates incoming pulses so that only the leading edges of the pulse are used.

fathometer—Also called depth finder, depth sounder, acoustic depth finder, echo depth sounder, or sonic depth sounder. A direct-reading device for determining the depth of water in fathoms or other units by reflecting sonic or ultrasonic waves from the ocean bottom. Also used to locate underwater bodies, such as schools of fish or sunken objects.

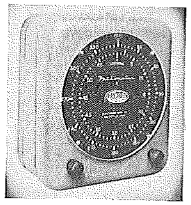

Fathometer.

fatigue—The weakening of a material under repeated stress.

fault—A defect in a wire circuit due to unintentional grounding, a break in the line, or a crossing or shorting of the wires.

fault current—The current that may flow in any part of a circuit or amplifier under specified abnormal conditions.

fault electrode current—The peak current that flows through an electrode during a fault, such as an arc back or a load short circuit.

fault finder—A test set for locating troubles in a telephone system.

fault indicator—Equipment that provides an instantaneous alarm, both visual and audible, of failures detected in the various components of its assorted equipment.

Faure plate—A storage-battery plate consisting of a conductive lead grid filled with active paste material.

fax—Abbreviation for facsimile.

fc—Abbreviation for footcandle.

FCC—Abbreviation for Federal Communications Commission.

F-display—Also called F-scan or F-scope. In radar, a rectangular display in which a target appears as a centralized blip when the radar antenna is aimed at it. Horizontal and vertical aiming errors are indicated by the horizontal and vertical displacement of the blip.

fdm — Abbreviation for frequency-division multiplex.

Federal Communications Commission—Abbreviated FCC. A board of seven commissioners appointed by the President under the Communications Act of 1934, having the power to regulate all interstate and foreign electrical communication systems originating in the United States.

feedback—1. In a transmission system or a section of it, the returning of a fraction of the output to the input. 2. In a magnetic amplifier, a circuit connection by which an additional magnetomotive force (which is a function of the output quantity) is used to influence the operating condition. 3. Part of a closed-loop system which brings back information about the condition under control, for comparison with the target value.

feedback amplifier—An amplifier that uses a passive network to return a portion of the output signal to modify the performance of the amplifier.

feedback attenuation—In the feedback loop of an operational amplifier, an attenuation factor by which the output voltage is attenuated to produce the input error voltage.

feedback control—A type of system control obtained when a portion of the output signal is operated upon and fed back to the input in order to obtain a desired effect.

feedback control loop—A closed transmission path which includes an active transducer and consists of a forward path, a feedback path, and one or more mixing points arranged to maintain a prescribed relationship between the loop input and output signals.

feedback control signal — That portion of the output signal which is returned to the input in order to achieve a desired effect, such as fast response.

feedback control system—A control system comprising one or more feedback control loops; it combines the functions of the controlled signals and commands, tending to maintain prescribed relationship between the two.

feedback cutter—An electromechanical transducer which performs like a disc cutter except it is equipped with an auxiliary feedback coil in the magnetic field. Signals exciting the cutter are induced into the feedback coil, the output of which is fed back in turn to the input of the cutter amplifier. The result is a substantially uniform frequency response.

feedback oscillator—An oscillating circuit, including an amplifier, in which the output is coupled in phase with the input. The oscillation is maintained at a frequency determined by the parameters of the amplifier and the feedback circuits, such as LC, RC, and other frequency-selective elements.

feedback path—In a feedback control loop, the transmission path from the loop output signal to the loop feedback signal.

feedback regulator—A feedback control system which tends to maintain a prescribed relationship between certain system signals and other predetermined quantities.

feedback transfer function—In a feedback control loop, the transfer function of the feedback path.

feedback winding—In a saturable reactor, the control winding to which a feedback connection is made.

feeder—A conductor or group of conductors connecting two generating stations, two substations, a generating station and a substation or feeding point, a substation and a feeding point, or a transmitter and antenna.

feeder cable—1. A communication cable extending from the central office along a primary route (main feeder cable), or from a main feeder cable along a secondary route (branch feeder cable) and providing connections to one or more distribution cables. 2. Also called trunk cable. In a catv system, the transmission cable from the head end (signal pickup) to the trunk amplifier.

feed holes—A series of small holes in perforated paper tape which convey no information, but are solely for the purpose of engaging the feed pawls or sprocket which transports the tape over the sensing pins of various reading devices.

feed pitch—The distance between the centers of adjacent feed holes in a tape.

feed reel—On a tape recorder, the reel which supplies the magnetic tape to the recording or playback head.

feedthrough—1. The accidental or unintentional transfer of a signal from one track to another on a multitrack tape. 2. The use of special connectors to pass conductors through bulkheads or panels. Contacts can be male on one side and female on the other, or they can be male on both sides or female on both sides. Feedthrough connectors differ from rack-and-panel types in that connection can be made on both sides of the panel.

feedthrough capacitor — A feedthrough insulator that provides a desired value of capacitance between the feedthrough conductor and the metal chassis or panel through which the conductor is passing. Used chiefly for bypass purposes in uhf.

feedthrough insulator—A type of insulator which permits wire or cable to be fed through walls, etc., with minimum current leakage.

feed-thru connection — *See* Thru-hole Connection.

female—The recessed portion of a device into which another part fits.

femto — Prefix meaning 10^{-15}. Abbreviated f.

femtoampere — A unit of current equal to 10^{-15} ampere. Abbreviated fa.

femtovolt—A unit of voltage equal to 10^{-15} volt. Abbreviated fv.

Fermi level—The value of electron energy at which the Fermi distribution function is one half.

ferreed—An electromechanical switch that combines the rapid switching of bistable magnetic material with metallic contacts to produce output indications that persist as long as desired without further application of power.

ferret reconnaissance—A form of reconnaissance which detects, locates, and analyzes enemy radars. It is a passive technique that listens for signals transmitted by enemy radars and thus cannot be jammed. The maximum effective range is limited only by the radio horizon. Ferret systems flown at altitudes of a few miles can accurately locate and analyze radars 300 miles away. Ferret is the only all-weather reconnaissance technique, and no camouflage system works against a properly designed ferret.

ferri—Prefix indicating a material having a net dipole moment.

ferric oxide (Fe_2O_3) — A red, iron oxide coating for magnetic recording tapes.

ferrimagnetic amplifier—A microwave amplifier utilizing ferrite material in the coupling inductors and transformers.

ferrimagnetic materials — Those materials in which spontaneous magnetic polarization occurs in nonequivalent sublattices; the polarization in one sublattice is aligned antiparallel to the other.

ferristor—A two-winding ferroresonant magnetic amplifier that operates on a high carrier frequency.

ferrite—Also called ferrospinel. A powdered, compressed, and sintered magnetic material having high resistivity, consisting chiefly of ferric oxide combined with one or more other metals. The high resistance makes eddy-current losses extremely low at high frequencies. Examples of ferrite compositions include nickel ferrite, nickel-cobalt ferrite, manganese-magnesium ferrite, yttrium-ion garnet, and single-crystal yttrium-ion garnet.

ferrite circulator — A nonreciprocal microwave network which transmits power from one terminal to another in sequence. Can replace a conventional duplexer, provide isolation of transmitter from receiver, eliminate the requirement for an ATR, and isolate the transmitter from antenna reflections.

ferrite core—A core made from iron and other oxides and usually shaped like a doughnut. It is used in circuits and magnetic memories and can be magnetized and demagnetized very rapidly.

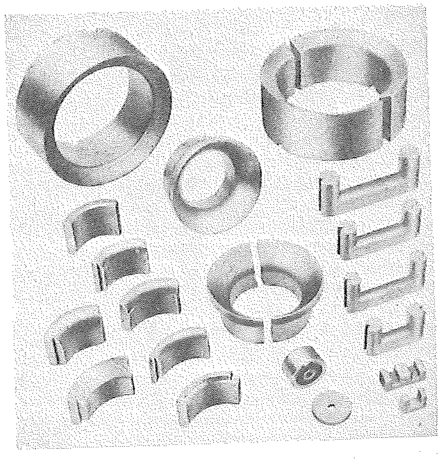

Ferrite cores.

ferrite isolator—A device either in a waveguide or coax which allows power to pass through in one direction with very little loss, while the r-f power in the reverse direction is absorbed. It is useful for maintaining signal source stability and eliminating long-line and frequency-pulling effects in all types of low-power microwave signal sources.

ferrite-rod antenna—Also called ferrod or loopstick antenna. An antenna used in place of a loop antenna in a radio receiver. It consists of a coil wound around a ferrite rod.

ferrites—Chemical compounds of iron oxide and other metallic oxides combined with ceramic material. They have ferromagnetic properties but are poor conductors of electricity. Hence they are useful where ordinary ferromagnetic materials (which are good electrical conductors) would cause too great a loss of electrical energy.

ferrod—*See* Ferrite-Rod Antenna.

ferrodynamic instrument — An electrodynamic instrument in which the measuring forces are materially increased by the presence of ferromagnetic material.

ferroelectric—Pertaining to a phenomenon exhibited by certain materials in which the material is polarized in one direction or the other, or reversed in direction by the application of a positive or negative electric field of magnitude greater than a certain amount. The material retains the electric polarization unless it is disturbed. The polarization can be sensed by the fact that a change in the field induces an electromotive force which can cause a current.

ferroelectric converter—A device which generates high voltage when heat is applied to it. Its operation is based on the change in the dielectric constant or the permittivity of certain materials such as barium titanate when heated. This change reaches maximum at the Curie point.

ferroelectric crystal—A crystal which can be polarized in the opposite direction by applying an electric field weaker than the breakdown strength of the material.

ferroelectricity—A property of a crystal having a spontaneously reversible polarization.

ferroelectric materials—Those materials in which the electric polarization is produced by cooperation action between groups or domains of collectively oriented molecules.

ferroelectrics—Pyroelectric materials the direction of polarization of which can be reversed by application of an electric field.

ferromagnetic—Pertaining to a phenomenon exhibited by certain materials in which the material is polarized in one direction or the other, or reversed in direction by the application of a positive or negative magnetic field of magnitude greater than a certain amount. The material retains the magnetic polarization unless it is disturbed. The polarization can be sensed by the fact that a change in the field induces an electromotive force, which can cause a current.

ferromagnetic amplifier—A parametric amplifier based on the nonlinear behavior of ferromagnetic resonance at high r-f power levels. In one version, microwave pumping power is supplied to a garnet or other ferromagnetic crystal mounted in a cavity containing a strip line. A permanent magnet provides sufficient field strength to produce gyromagnetic resonance in the garnet at the pumping frequency. The input signal is applied to the crystal through the strip line, and the amplified output signal is extracted from the other end of the strip line. Sometimes incorrectly called a garnet maser, but the operating principle differs from that of the maser.

ferromagnetic material—A material having a specific permeability greater than unity, the amount depending on the magnetizing force. A ferromagnetic material usually has relatively high values of specific permeability and it exhibits hysteresis. The principal ferromagnetic materials are iron, nickel, cobalt, and certain of their alloys.

ferromagnetic oxide parts — Parts, consisting primarily of oxides which display ferromagnetic properties.

ferromagnetism — A high degree of magnetism in ferrites and similar compounds. The magnetic moments of neighboring ions tend not to align parallel with each other. The moments are of different magnitudes, and the resultant magnetization can be large.

ferromanganese—An alloy of iron and manganese.

ferrometer—An instrument for making permeability and hysteresis tests of iron and steel.

ferroresonance — Resonance associated with circuits in which at least one of the circuit elements is nonlinear and contains iron.

ferroresonant circuit—A resonant circuit in which one of its elements is a saturable reactor.

ferrospinel—A ceramic-like material containing iron and other elements combined with oxygen. A poor conductor of electricity, it is used in transformers, antenna loops, and television deflecting yokes.

ferrous — Composed of and/or containing iron. A ferrous metal exhibits magnetic characteristics, as opposed to a nonferrous metal such as aluminum which does not.

ferrule resistor—A resistor having ferrule terminals for mounting in standard fuse clips.

FET—Abbreviation for (unipolar) field effect transistor.

fetal electrocardiograph — An instrument for detecting and recording fetal ecg.

fetal heart monitor — An instrument that combines fetal ecg with an oscilloscope display and/or alarm.

fetal phonocardiograph — An instrument which provides continuous instantaneous recording of beat-to-beat changes in the fetal heart rate.

fetch—That portion of a computer cycle during which the location of the next instruction is determined. The instruction is taken from memory and modified if necessary, and it is then entered into the control register.

"f" factor — The slope of the straight line from which the nonlinearity of a displacement transducer is calculated in microvolts output per volt excitation per unit stimulus.

fiber—A tough insulating material, generally of paper and cellulose, compressed into rods, sheets, or tubes.

fiber needle—A playback point or phonograph needle made from fiber. Being softer than a metal or diamond needle, it is less scratchy; however, it has an extremely short life.

fiber-optic bundles—Assemblies of optical fibers.

fiber optics—Also called optical fibers or optical fiber bundles. An assemblage of transparent glass fibers all bundled together parallel to one another. The length of each fiber is much greater than its diameter. This bundle of fibers has the ability to transmit a picture from one of its surfaces to the other around curves and into otherwise inaccessible places with an extremely low loss of definition and light, by a process of total reflection.

fiber optics probe—A flexible probe made up of a bundle of fine glass fibers optically aligned to transmit an image.

fiber optics scrambler—Similar to a fiberscope except that the middle section of loose fiber is deliberately disoriented as much as possible, then potted and sawed. Each half is then capable of coding a picture, which can be decoded by the other half.

fiberscope—Optical glass fibers, when systematically arranged in a bundle, transmit a full color image that remains undisturbed when the bundle is bent. By mounting an objective lens on one end of the bundle, and an eyepiece at the other, the assembly becomes a flexible fiberscope that can be used to view objects that would be inaccessible for direct viewings.

fidelity—1. The accuracy with which a system or portion of a system reproduces at its output the essential characteristics of the signal impressed on its input. 2. A measure of the exactness with which sound is duplicated or reproduced.

field—1. One half of a television image. With present U. S. standards, pictures are transmitted in two fields of 262½ lines each, which are interlaced to form 30 complete frames, or images, per second. 2. A general term referring to the region under the influence of some physical agency such as electricity, magnetism, or a combination produced by an electrically charged object, electrons in motion, or a magnet. 3. A group of characters in a computer which is treated as a single unit of information. 4. In each of a number of punch cards, a column or columns regularly used for a standard item of information.

field application relay—A device which automatically controls the application of the field excitation to an a-c motor at some predetermined point in the slip cycle.

field circuit breaker—A device which functions to apply, or to remove, the field excitation of a machine.

field coil—A coil of insulated wire wound around an iron core. Current flowing in the coil produces a magnetic field.

field density—*See* Magnetic Induction.

field-discharge protection — A control function or device to limit the induced voltage in the field when the field current attempts to change suddenly.

field distortion — Distortion between the north and south poles of a generator due to the counterelectromotive force in the armature winding.

field-effect tetrode—A semiconductor device consisting basically of a thin N region adjacent to a similarly thin P region. Two contacts are made to the N side and two to the P side so that currents can be passed through each thin region parallel to the single junction. The two currents remain separate because reverse bias is maintained on the junction. A current in either side affects the resistance of the other side and hence the current in the other side.

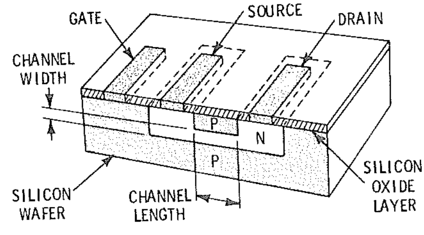

Field-effect transistor.

field-effect transistor—A transistor with a circuit application similar to that of a vacuum tube. The main conduction path is through a bar of N-type silicon, with control through depletion layers on each side of the N-type bar.

field-effect tube—A triode with its grid replaced by a nonintercepting control gate. A high positive voltage is applied to this gate in order to draw sufficient current from the cathode. The result is a strong concentration of the electric field at the gap between the gate and the cathode, producing an electron beam passing through the gate to the anode.

field emission—Also called cold emission. The liberation of electrons from a solid or liquid by application of a strong electric field at the surface.

field-enhanced photoelectric emission—Increased photoelectric emission resulting from

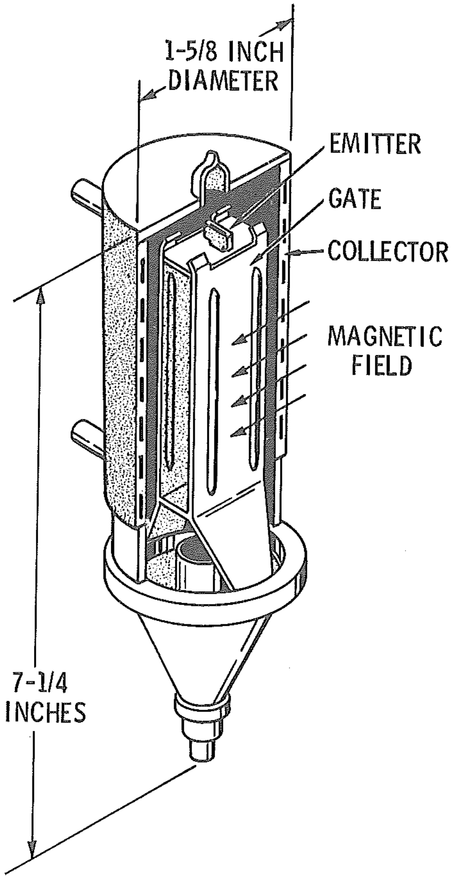

Field-effect tube.

the action of a strong electric field on the emitter.

field-enhanced secondary emission — Increased secondary emission resulting from the action of a strong electric field on the emitter.

field forcing—The effect of a control function or device which temporarily overexcites or underexcites the field of an electrical machine in order to increase the rate of change of flux.

field-free emission current — Also called zero-field emission. The electron current emitted by a cathode when the electric field at the surface of the cathode is zero.

field frequency—Also called field-repetition rate. In television, the frame frequency multiplied by the number of fields contained in one frame. In the United States the field frequency is 60 per second, or twice the frame frequency.

field intensity—*See* Field Strength.

field loss relay — *See* Motor Field Failure Relay.

field magnet—An electromagnet or permanent magnet which produces a strong magnetic field in a speaker, microphone, phonograph pickup, generator, motor, or other electrical device.

field-neutralizing coil—A coil encircling the faceplate of a color picture tube. The current through it produces a magnetic field which offsets any effects of the earth's and other stray magnetic fields on the electron beams.

field-neutralizing magnet—Also called rim magnet. A permanent magnet mounted near the edge of the faceplate of a color picture tube to prevent stray magnetic fields from affecting the path of the electron beams.

field of view—The solid angle from which objects can be acceptably viewed, photographed, or otherwise detected.

field period—The time required to transmit one television field. In the United States, it is 1/60th of a second.

field pickup—Also called a remote or nemo. A radio or television program originating outside the studio.

field relay — A device that functions on a given or abnormally low value or failure of machine field current, or on an excessive value of the reactive component of armature current in an a-c machine, indicating abnormally low field excitation.

field-repetition rate—*See* Field Frequency.

field resistor—A component in which the resistance element is a thin layer of conductive material on an insulated form. The conductive material does not contain either binding or insulating material.

field rheostat—A variable resistance connected to the field coils of a motor or generator and used for varying the field current.

field ring—The part which supports the field of a dc or series-wound motor housing. The motor end shields are attached to the ends of the field ring.

field selection—In a computer, the isolation of a particular data field with one computer word without isolating the word.

field sequential—Pertaining to the association of individual primary colors with successive fields in a color-television system (e.g., field-sequential pickup, display, system, transmission).

field-sequential color television—A color-television system in which the individual primary colors (red, blue, and green) are produced in successive fields.

field-simultaneous system — A color-television system in which a complete full-color field is presented simultaneously as a unit. The eyes then see a succession of full-color images rather than a succession of primary-color fields.

field strength—1. Also called field intensity. The value of the vector at a point in the region occupied by a vector field. In radio, it is the effective value of the electric-field intensity in microvolts or millivolts per meter produced at a point by radio waves from a particular station. Unless otherwise specified, the measurement is assumed to be in the direction of maximum field intensity. 2. The amount of magnetic flux produced at a particular point by an electromagnetic or permanent magnet.

field-strength meter—A calibrated measuring instrument for determining the strength of radiated energy (field strength) being received from a transmitter.

field telephone—A durable, portable telephone designed for use in the field.

field wire—A flexible insulated wire used in field-telephone and -telegraph systems.

figure of merit—1. The property or characteristic which makes a tube, coil, or other electronic device suitable for a particular application. It is a quality to look for in choosing a piece of equipment. 2. In a magnetic amplifier, the ratio of the power gain to the time constant. 3. For a thermoelectric material, the quotient of the square of the absolute Seebeck coefficient (α) divided by the product of the electrical resistivity (ρ) and the thermal conductivity.

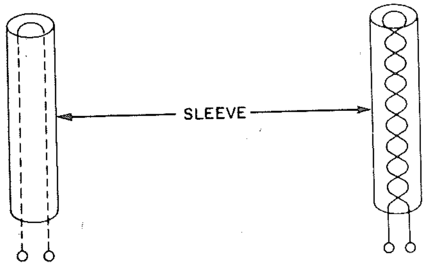

Filaments (indirectly-heated).

filament—1. Also known as a filamentary cathode. The cathode of a thermionic tube, usually a wire or ribbon, which is heated by passing a current through it. 2. In tubes employing a separate cathode, the heating element.

filamentary transistor — A conductivity-modulation transistor which is much longer than it is wide.

filament battery—The source of energy for heating the filament of a vacuum tube.

filament circuit — The complete circuit through which filament current flows.

filament current—The current supplied to a filament to heat it.

filament emission—The liberation of electrons when the filament in a vacuum tube is heated.

filament power supply—The source of power for the filament or heater of a vacuum tube.

filament resistance — The resistance (in ohms) of the filament of a vacuum tube or incandescent lamp.

filament rheostat — A variable resistance placed in series with the filament of a vacuum tube to regulate the filament current.

filament sag—The bending of a filament when it heats up and expands.

filament saturation—Also called temperature saturation. The condition whereby a further increase in filament voltage will no longer increase the plate current at a given value of plate voltage.

filament transformer—A transformer used exclusively to supply filament voltage and current for the vacuum tubes.

filament voltage—The voltage value which must be applied to the filament of a vacuum tube to obtain the rated filament current.

filament winding—A secondary winding provided on a power transformer to furnish alternating filament voltage for one or more vacuum tubes.

file—1. A collection of related records—e.g., in inventory control, one line of an invoice containing data on the material, the quantity, and the price forms an item; a complete invoice forms a record; and the complete set of such records forms a file. 2. To insert an item into such a set.

file gap—A space or interval of time, the purpose of which is to indicate or signal the end of a file.

file maintenance—The processing of a computer file in order to bring it up to date.

filler—1. In mechanical recording, the inert material of a record compound (as distinguished from the binder). 2. Nonconducting component cabled with insulated conductors to impart roundness, flexibility, tensile strength, or a combination of all three, to the cable.

film badge — A type of dosimeter consisting of a badge containing a sensitized film which, when developed, gives an indication of the total dose of ionizing radiation to which the badge has been subjected.

film integrated circuit—An integrated circuit the elements of which are films formed *in situ* upon an insulating substrate. (*Also see* Integrated Circuit.) To further define the nature of a film integrated circuit, additional modifiers may be prefixed. Examples are: thin-film integrated circuit; thick-film integrated circuit.

film pickup—A film projector combined with a television camera for telecasting scenes from a motion-picture film.

film reader—A computer input device which scans opaque and transparent patterns on photographic film and relays the corresponding information to the computer.

film recorder—A device that receives information from a computer and records it on photographic film.

film reproducer—An instrument which reproduces a recording on film.

film resistor—A fixed resistor the resistance element of which is a thin layer of conductive material on an insulated form. Some sort of mechanical protection is placed over this layer.

film scanning—The process of converting movie film into corresponding electrical signals that can be transmitted by a television system.

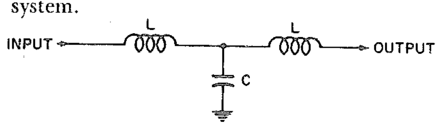

(A) T-section filter.

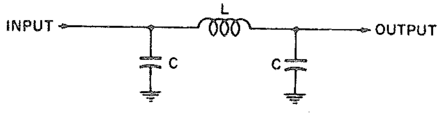

(B) Pi-section filter.

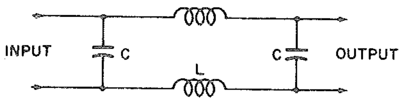

(C) Ladder-section filter.

Filters (wave filters).

filter—1. A selective network of resistors, inductors, or capacitors which offers comparatively little opposition to certain frequencies or to direct current, while blocking or attenuating other frequencies. 2. Also called a wave filter. A tuned circuit designed to pass a-c signals of a specified frequency. 3. A pattern of characters used to control the selection or elimination of portions of another pattern of characters. 4. A device or program that separates data, signals, or materials in accordance with specified criteria. 5. Also called extractor or mask. A machine word that specifies which parts of another machine word are to be operated on.

filter attenuation band—Also called a filter stop band. A frequency band in which the attenuation constant is not zero if dissipation is neglected. In other words, a frequency band of attenuation.

filter capacitor—A capacitor used in a filter circuit. The term is usually reserved for electrolytic capacitors in a power-supply filter circuit.

filter center—In an aircraft control and warning system, a location at which information from observation posts is filtered for further dissemination to air-direction centers.

filter choke — Normally, an iron-core coil which allows direct current to pass while opposing the passage of pulsating or alternating current.

filter-impedance compensator—An impedance compensator which is connected across the common terminals of electric-wave filters, when the latter are used in parallel, in order to compensate for the effects of the filters on each other.

filter passband — *See* Filter-Transmission Band.

filter slot—A choke, in the form of a slot, designed to suppress unwanted modes in a waveguide.

filter stop band — *See* Filter Attenuation Band.

filter-transmission band — Also called filter passband. A frequency band in which the attenuation constant is zero if dissipation is neglected. In other words, a frequency band of free transmission.

final actuation time—The time of termination of the chatter of a relay following contact actuation.

final amplifier—The stage which feeds the antenna in a transmitter.

final wrap — The outer layer of insulation around a coil, covering the saddle and splice insulation.

finder — In a telephone switching system, a name applied to the switch or relay group that selects the path which the call is to follow through the system.

fine-chrominance primary—Also called the I-signal. In the color television system presently standardized for broadcasting in the United States, the chrominance primary associated with the greater transmission bandwidth.

fine-tuning control—A receiver control which varies the frequency of the local oscillator over a small range to compensate for drift and permit fine adjustment to the carrier frequency of a station.

finished blank—A crystal product after completion of all processes. It may also include the electrodes adherent to the crystal blank.

finishing — The process of repeated hand lapping and electrical testing by which a finished crystal blank is brought up to specifications.

finishing rate—Expressed in amperes, the rate of charge to which the charging current of a battery is reduced near the end of the charge to prevent excessive gassing and temperature rise.

finish lead—The lead connected to the finish, or outer end, of a coil.

finite—Having fixed and definite limits.

fins—Radial sheets or discs of metal attached to metal parts of a power tube or other component for the purpose of dissipating heat.

fire-control equipment — Equipment that takes in target indications from optical or radar devices and, after calculating the motion of the target and firing vehicle, properties of air, etc., puts out directions of bearing, elevation, and timing for aiming and firing the guns.

fire-control radar—Radar employed for directing gunfire against the targets it observes.

fired tube (TR, ATR, and pre-TR tubes)—The condition of a tube while a radio-frequency glow discharge exists at the resonant gap, resonant window, or both.

firing—1. In any gas- or vapor-filled tube, the ionization of the gas and the start of current flow. 2. The excitation of a device during a brief pulse. 3. In a magnetic amplifier, the transition from the unsaturated to the saturated state of the saturable reactor during the conducting or gating alternation. 4. An adjective modifying phase or time, to designate when firing occurs.

firing angle—1. The electrical angle of the plate-supply voltage at which an ionization of a gaseous tube occurs. 2. In a magnetic amplifier, the point on a sine-wave control voltage at which the control ampere-turns are sufficient to saturate the cure. This is the point where the secondary winding (load) impedance drops to zero, and almost all of the supply voltage appears across the load.

firing point—The point at which the gas or vapor in a tube ionizes and current begins to flow.

firing potential—The controlled potential at which conduction through a gas-filled tube begins.

first audio stage—The first stage in an audio amplifier.

first detector—Now called the mixer. In a superheterodyne receiver, the stage where the local-oscillator signal is combined with the modulated incoming radio-frequency signal to produce the modulated intermediate-frequency signal.

first Fresnel zone—In optics and radio communications, the circular portion of a wave front intersecting the line between an emitter and a more distant point where the resultant disturbance is being observed. The center intersects the front with the direct ray, and the radius is such that the shortest path from the emitter through the periphery to the receiving point is one-half wave longer than the ray.

fishbone antenna—An antenna consisting of a series of coplanar elements arranged in collinear pairs and loosely coupled to a balanced transmission line.

fishpaper—A tough fiber used in sheet form for insulating transformer windings from the core, field coils from field poles, or conductors from the armature.

fission—Also called atomic fission or nuclear fission. The splitting of an atomic nucleus into two parts. Fission reactions occur only with heavy elements such as uranium and plutonium and are accompanied by large amounts of radioactivity and heat.

fissionable—Capable of undergoing fission.

fission products—The elements which result from atomic fission. They may consist of more than forty different radioactive elements such as arsenic, silver, cadmium, iodine, barium, tin, cerium, and others.

fitting—An accessory, such as a locknut or bushing, to a wiring system. Its function is primarily mechanical rather than electrical.

fix—A position determined without reference to any former position.

fixed bias—A constant value of bias voltage.

fixed capacitor—A capacitor designed with a definite capacitance that cannot be adjusted.

fixed crystal—A crystal detector with a nonadjustable contact position.

fixed-cycle operation—1. A type of computer performance whereby a fixed amount of time is allocated to an operation. 2. Synchronous or clock-type arrangement in a computer in which events occur as a function of measured time.

fixed echo—A stationary echo indication on a radar ppi display, indicating a fixed target.

fixed-frequency iff—A class of iff (identification friend or foe) equipment which responds immediately to every interrogation, thus permitting the response to be displayed on plan-position indicators.

fixed-frequency transmitter—A transmitter designed for operation on a single carrier frequency.

fixed-point system—A system of notation in which a number is represented by a single set of digits and the position of the radix point is not numerically expressed. (*Also see* Floating-Point System.)

fixed-program computer — *See* Wired-Program Computer.

fixed resistor—A resistor designed to introduce only a predetermined amount of resistance into an electrical circuit and not adjustable.

fixed screen—Application of a potential to a screen grid which is unaffected by other operating conditions within the tube.

fixed service—Any service communicating by radio between fixed points, except broadcasting and special services.

fixed station—1. A station in the fixed service. (A fixed station may, as a secondary service, transmit to mobile stations on its normal frequencies.) 2. A permanent station which communicates with other fixed stations.

fixed transmitter — A transmitter operated from a permanent location.

fixed-voltage winding—The motor winding which is excited by a fixed voltage.

fl—Abbreviation for footlambert.

flag—A large sheet of metal or fabric for shielding television camera lenses from light.

flame-failure control—A system which automatically stops the fuel supply to a furnace if the pilot burner accidentally goes out.

flame microphone—A microphone in which the action of sound waves on a flame changes the resistance between two electrodes in the flame.

flame resistance — The characteristic of a material that prevents it from flaming when the source of heat is removed.

flame-resistant—*See* Flame-Retardant.

flame-retardant—Retarding ignition and the spread of flames, either inherently or because of special treatment. Also called flame-resistant.

flammability — The ability of a material to support combustion.

flammable—Term applied to material which readily ignites and burns when exposed to flames or elevated temperatures. Also called combustible.

flange—Also called waveguide flange. A fitting used at the end of a waveguide to a microwave component or to another waveguide.

flange connector—A mechanical joint employing plane flanges bolted together in a waveguide.

flange coupling — A connection utilizing flanges not in mechanical contact between two parts of a waveguide, yet introducing no discontinuity in the flow of energy along the guide.

flanking effect—The effect on filter characteristics of connecting additional filters in parallel.

flap attenuator — A form of waveguide attenuator in which a variable amount of loss is introduced by insertion of a sheet of resistive material, usually through a nonradiating slot.

flare—An enlarged and distorted radar-screen target indication due to excessive brightness.

flare angle—The continuous change in cross section of a wave guide.

flare factor—A number expressing the degree of outward curvature of a speaker horn.

flash—Sometimes called hit. Momentary interference to a television picture, lasting approximately one field or less and of sufficient magnitude to totally distort the picture information. In general, this term is used only when the impairment is so short that the basic impairment cannot be recognized.

flashback voltage—The inverse peak voltage at which ionization takes place in a gas tube.

flasher — A device, generally a thermal or motor-driven switch, that rapidly and automatically lights and extinguishes electric lamps.

flasher relay—A self-interrupting relay.

flashing — The application of a high-frequency electromagnetic field to an electron tube through the envelope to flash its getter during evacuation.

flashing of a dynamo — The flashing or sparking which is likely to take place at the brushes of a commutator.

flash magnetization — Magnetization of a ferromagnetic object by an abrupt current impulse.

flashover — A disruptive discharge through air, around or over the surface of insulation, or between parts of different potential or polarity produced by the application of voltage, wherein the breakdown path becomes sufficiently ionized to maintain an electric arc.

flashover voltage—1. The highest value attained by any voltage impulse which caused a flashover. 2. The voltage at which insulation fails by discharge between electrodes across the insulation surface.

flash pulsing—Transmission of short bursts of radiation at irregular intervals by a mechanically controlled keyer.

flash test—A method of testing insulation by momentarily applying a voltage much higher than the working voltage.

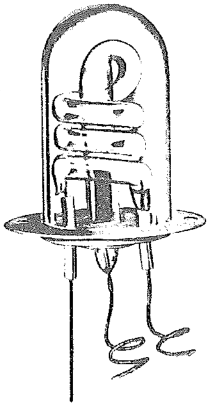

Flash tube.

flash tube—Also called electronic flash tube and photoflash tube. A gas-discharge tube for producing high-intensity, short-duration flashes of light. It consists of a glass tube bent in a U, a helix, or a combination of the two and filled with a rare gas. The tube has an anode, a cold cathode, and a trigger electrode. It is flashed by applying a high-voltage pulse to the trigger electrode.

flash welding—Welding in which an arc is first struck between the pieces to be welded. After the ends are thus heated, the weld is completed by bringing them together under pressure and cutting off the current.

flat fading—That type of fading in which all components of the received radio signal fluctuate in the same proportion simultaneously.

flat, flexible cable—*See* Tape Cable.

flat frequency response—The response of a system in a constant-amplitude function which varies in frequency. The response is flat if it varies within specified limits of amplitude, usually specified in decibels from a reference quantity.

flat leakage power (TR and pre-TR tubes) —The peak radio-frequency power transmitted through the tube after establishment of a steady-state radio-frequency discharge.

flat line—A radio-frequency transmission line or part of a line having a low standing-wave ratio.

flat pack—A flat, rectangular integrated-circuit or hybrid-circuit package with coplanar leads.

flat response—Ability of a sound system to reproduce all tones, from the lowest to the highest, in the proper proportion.

flat top—The horizontal portion of an antenna.

flat-top antenna—An antenna having two or more lengths of wire parallel to each other and to the ground.

flat-top response—Response characteristic in which a definite band of frequencies is transmitted uniformly.

flat-type armature — Of a relay, an armature which rotates about an axis perpendicular to that of the core, with the pole face on a side surface of the core.

flat-type relay — A relay having a flat-type armature.

flaw—In a material, any discontinuity that would be harmful to proper functioning of the material.

flaw detection—The process of using sonic or ultrasonic waves to locate imperfections in a solid material. This is done by transmitting the waves through the material and listening for reflections or variations in transmission when they strike an imperfection in the material.

F layer—An ionized layer in the F region, existing in the night hemisphere and in the weakly illuminated portion of the day hemisphere.

F$_1$ layer — One of the regular ionospheric layers at an average height of about 225 kilometers, which occurs during the daylight hours.

F$_2$ layer—The most useful of the ionospheric layers for radio-wave propagation. It is the most highly ionized and highest of the layers, having an average night height of 225 kilometers and a midday height of about 400 kilometers.

Fleming's rule—Also called the right-hand or left-hand rule. If the thumb and the first and second fingers are extended at right angles to one another, with the thumb representing the direction of the wire motion, the first finger representing the direction of magnetic lines of force (from the north pole to the south pole), and the second finger representing the direction of the current, then the right hand will give the correct relationships for a conductor in the armature of a generator, and the left hand will give the correct relationships for a conductor in the armature of a motor. This rule is applied to the so-called conventional current flow, which is the opposite of electron flow.

Fleming valve—An early name for a diode, or two-electrode thermionic vacuum tube used as a detector.

Fletcher-Munson curves—Also called equal-loudness contours. A group of sensitivity curves showing the characteristics of the human ear for different intensity levels between the threshold of hearing and the threshold of feeling. The reference frequency is 1,000 Hz.

Flewelling circuit—An early radio circuit in which one tube served as a detector, amplifier, and local oscillator.

flexible coupling—1. A device for connecting two shafts end to end so that they can be rotated even though not exactly aligned. 2. Mechanical connection between two lengths of waveguide normally lying in a straight line and designed to allow a limited angular movement between axes.

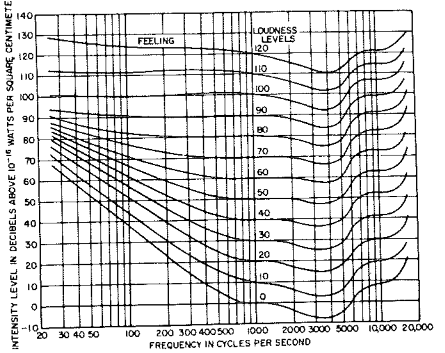

Fletcher-Munson curves.

flexible resistor—A wirewound resistor that looks like a flexible lead. It is made by winding *Nichrome* or any other type of resistance wire around asbestos or other heat-resistant cord. The wire is then covered with braided insulation, which is color coded to indicate the resistor value.

flexible shaft—A flexible core made up of layers of wire, which rotate inside a metal or rubber-covered flexible casing. The casing not only supports the core, but also acts as the bearing surface for the core.

flex life — A measure of the resistance of a conductor or other device to failure due to fatigue from repeated bending.

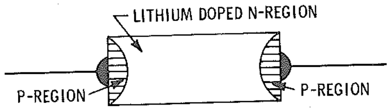

Flexode.

flexode—A flexible diode containing a junction that may be altered at will from a PN junction in one direction, to no junction at all, to a PN junction in the opposite direction. Thus the direction of easy current may be reversed without reversing the leads to the diode, and the resistance of the diode may be continuously varied from the back-resistance value to the forward-resistance value. It may be set to behave as a simple resistor, with the same value for both directions of current.

flicker—1. In television, the flickering produced in the picture when the field frequency is insufficient to completely synchronize the visual images. 2. In a regulated power supply, a phenomenon due primarily to sudden, minute changes of brief duration in the reference voltage or the input-stage junctions of the correction amplifier in the regulator. 3. Also called jitter or wobble. Noise in an amplifier, of higher frequency than drift, but lower than power-line or chopper-drive frequency noise.

flicker effect—Small variations in the plate current of a thermionic vacuum tube, believed to be due to random emission of positive ions by the cathode.

flicker photometer—A device for measuring the intensity of a light source. Illumination from the light source being measured and a standard light source are observed alternately in rapid succession. When the standard source is equal to the other, the flickering disappears.

flight control—Real-time calculations for the control of a vehicle in flight; includes stabilization, fuel monitoring, cruise control, etc.

flight path—A planned course for an airborne vehicle.

flight-path computer—A computer that includes all the functions of a course-line computer and also controls the altitude of an aircraft in accordance with a desired plan of flight.

flight-path deviation — The difference between the flight track of an aircraft and the actual flight path expressed in terms of either angular or linear measurement.

flight-path-deviation indicator—An instrument that provides a visual indication of deviation from a flight path.

flight track—The three-dimensional path in space actually traced by a vehicle.

Flinders bar—A bar of soft iron placed near a compass to correct errors due to variation of the vertical component of the earth's magnetism in different parts of the world.

flip-chip—An unencapsulated semiconductor device in which bead-type leads terminate on one face to permit flip (face-down) mounting of the device by contact of the leads to the required circuit interconnectors.

flip coil—A small coil used for measuring a magnetic field. When connected to a ballistic galvanometer or other instrument, it gives an indication whenever the magnetic field of the coil or its position in the field is suddenly reversed.

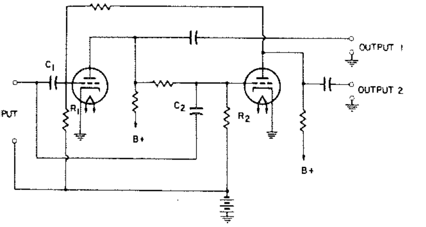

Flip-flop circuit.

flip-flop—Also called bistable multivibrator, Eccles-Jordan circuit, or trigger circuit. 1. A two-stage multivibrator circuit having two stable states. In one state, the first stage is conducting and the second is cut off. In the other state, the second stage is conducting and the first stage is cut off. A trigger signal changes the circuit from one state to the other, and the next trigger signal changes it back to the first state. For counting and scaling purposes, a flip-flop can be used to deliver one output pulse for each two input pulses. 2. A similar bistable device with an input which allows it to act as a single-stage binary counter.

flip-flop equipment—An electronic or electromechanical device which causes automatic alternation between two possible circuit paths. The same term is often applied to any mechanical operation which is analogous to the principles of the flip-flop.

flip-flop multivibrator—Also called start-stop multivibrator. A biased rectangular wave generator which operates for one cycle when a synchronizing trigger signal is applied.

flipover cartridge — A phonograph cartridge having separate needles for playing microgrove and standard records. It may be turned to bring the proper needle into playing position.

floated battery—A storage battery kept fully charged across the leads of a generator. The generator carries the normal load, and the battery assists during peaks.

floating — 1. Keeping a storage battery connected in parallel with an electric supply to serve as a standby in case of supply failure and to assist in handling peak loads. 2. The condition of a device or circuit that is not grounded and not tied to any established potential.

floating address—*See* Symbolic Address.

floating-carrier modulation—*See* Controlled-Carrier Modulation.

floating charge—Continuous charging of a storage battery with a low current to keep the battery fully charged while idle or on light duty.

floating grid—A vacuum-tube grid that is not connected to any circuit. It assumes a negative potential with respect to the cathode.

floating input—1. An isolated input circuit not connected to ground at any point (the maximum permissible voltage to ground is limited by electrical design parameters of the circuit involved). It is understood that in a floating input circuit, both conductors are equally free from any reference potential, a qualification which limits the types of signal sources which can be operated floating. 2. *See* Differential Input Measurement.

floating junction—A semiconductor junction through which no net current flows.

floating neutral—A circuit in which the voltage to ground is free to vary with circuit conditions.

floating-point calculation—In a computer, a calculation taking into account the varying location of the decimal point (if base 10) or binary point (if base 2). The sign and coefficient of each number are specified separately.

floating-point routine — Coded instructions in proper sequence to direct a computer to perform a calculation with floating-point operation.

floating-point system—A system of numbering in which an added set of digits is used to denote the location of the radix point. (*Also see* Fixed-Point System.)

floating potential—The d-c voltage between an open-circuited terminal of a circuit and a reference point when a d-c voltage is applied to the other circuit terminals as specified.

float switch—A switch actuated by a float on the surface of a liquid.

flock—Finely divided felt used on phonograph turntables, underneath microphone stands, or wherever a nonscratching surface is desired.

flow—1. The passage of electrons (a current) through a conductor or through the space between electrodes. 2. A general term to indicate a sequence of events.

flowchart—*See* Flow Diagram.

flow diagram—Also called flowchart. A chart showing all the logical steps of a computer program. A program is coded by writing down the successive instructions that will cause the computer to perform the logical operations necessary for solving the problem, as represented on a flowchart.

flowed wax—A mechanical recording disc prepared by melting and flowing wax onto a metal base.

flowmeter—A device for measuring the rate of flow of liquids or gases.

flow soldering—Also called wave soldering. A method of soldering printed-circuit boards by moving them over a flowing wave of molten solder in a solder bath.

fluctuating current—A direct current that changes in value, but not at a steady rate.

fluctuation noise—*See* Random Noise.

fluctuation voltage—Small voltage variations in a thermionic tube due to thermal agitation, shot effect, flicker effect, etc.

fluid damping—Damping obtained through displacement of a viscous fluid and the accompanying dissipation of heat.

fluorescence—Emission of light when a material is excited by electrons, ultraviolet radiation, or X rays.

fluorescent—Having the property of giving off light when activated by electronic bombardment or a source of radiant energy.

fluorescent lamp—An electric discharge lamp in which a gas ionizes and produces radiation that activates the fluorescent material inside the glass tubing. The phosphors in the fluorescent material transform the radiant energy from the electric discharge into wavelengths giving more light (higher luminosity).

fluorescent material—A material that fluoresces readily when exposed to electron beams, X rays, or other radiation.

fluorescent screen—The coating, on the face of a cathode-ray or television picture tube, which glows under electron bombardment.

fluorometer — An instrument for measuring fluorescence.

fluoroscope—An instrument with a fluorescent screen suitably mounted with respect to an X-ray tube, used for immediate indirect viewing of internal organs of the body, or internal structures in apparatus or masses of metals, by means of X rays. A fluorescent image (really a kind of X-ray shadow picture) is produced.

fluoroscopy—The use in diagnosis, testing, etc., of a fluorescent screen activated by X rays.

flush receptacle—A receptacle recessed into a wall, with only the plate extending beyond the surface.

flush-type instrument—An instrument designed to be mounted with its face projecting only slightly from the front of the panel.

flutter—1. Also called wow and drift. The frequency deviations produced by irregular motion of the turntable during recording, duplication, or reproduction. The term "flutter" usually refers to relatively high cyclic deviations (for example, 10 hertz), and the term "wow" to relatively low ones (for example, a variation of once per revolution). The term "drift" usually refers to a random rate close to zero hertz. 2. In communications, (a) distortion due to variations in loss resulting from simultaneous transmission of a signal at another frequency, or (b) a similar effect due to phase distortion.

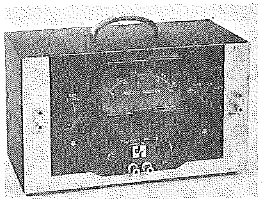

Flutter bridge.

flutter bridge—An instrument for measuring the irregularities in a constant-speed device such as a film, disc, or tape recorder.

flutter echo—1. A rapid succession of reflected pulses resulting from a single initial pulse. 2. A multiple echo in which the reflections occur in rapid succession. If periodic and audible, it is referred to as a musical echo.

flutter rate—The number of times per second the flutter varies.

flux—1. A material used to promote the joining of metals in soldering. Rosin is widely used in soldering electronic parts. 2. Number of particles crossing a unit area per unit time. The common unit of flux is particles/cm²/sec. Integrated flux, after an exposure of time T, is equal to the total number of particles which have traversed a unit area during time T.

flux density—Also called induction density. The number of magnetic lines of force passing through a unit area.

fluxgate—A magnetic azimuth-sensitive element of the fluxgate-compass system activated by the earth's magnetic field.

fluxgate compass—A gyrostabilized, remote-indicating compass and azimuth-control system used with automatic pilots.

fluxgraph — A machine that automatically plots on paper the magnetic field strength at various points in the vicinity of a coil.

flux guide—In induction heating, a magnetic material used for guiding the electromagnetic flux to the desired location or for confining it to definite regions.

flux linkage—Magnetic lines of force which link a coil of wire. Whenever the flux linkage changes, an emf is generated in the coil.

fluxmeter—An instrument used with a test coil for measuring magnetic flux. It consists usually of a moving-coil galvanometer in which the torsional control is either negligible or compensated.

flyback—1. The shorter of the two time intervals comprising a sawtooth wave. 2. Also called retrace. As applied to a cathode-ray tube, the return of the spot to its starting point after having reached the end of its trace. This portion of the wave is usually not seen because of blanking circuits or the shortage of time.

flyback checker — An instrument used to check flyback or other transformers or inductors for open windings or shorted turns.

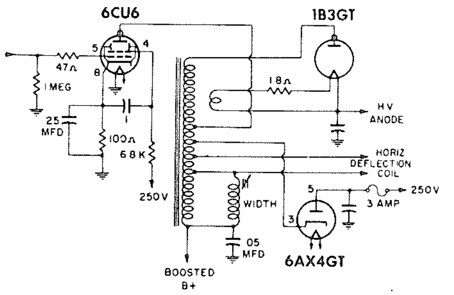

Flyback power supply.

flyback power supply—The power supply that generates the high d-c voltage required by the second anode of a picture tube. This voltage is produced during the flyback period, the current in the horizontal-deflecting coils reversing and inducing a sharp pulse in the primary of the transformer supplying the deflection circuit. This pulse is stepped up by an autotransformer and rectified. After suitable filtering, it becomes a very high d-c voltage.

flyback tester—An instrument that tests flyback transformers and sometimes also deflection yokes.

flyback time—The period during which the electron beam is returning from the end of a scanning line to begin the next line.

flyback transformer—Also called horizontal-output transformer. A transformer used in the horizontal-deflection circuit of a television receiver to provide the horizontal-scanning and accelerating-anode voltages for the cathode-ray tube. It also supplies the filament voltage for the high-voltage rectifier.

flycutter—An accessory used with a drill press to cut out large holes in metal or wood.

flying spot—A small, rapidly moving spot of light, usually generated by a cathode-ray tube, used to scan an image field for television transmission.

flying-spot scanner—1. Also called light-spot scanner. A television scanning device embodying a small beam which is moved over a scene or film and translates the highlights and shadows into electrical signals. 2. In optical character recognition, a device employing a moving sopt of light to scan a sample space, the intensity of the transmitted or reflected light being sensed by a photoelectric transducer.

flywheel effect—The maintaining of oscillations in a circuit in the intervals between pulses of excitation energy. The action is analogous to the rotation of a flywheel due to its stored mechanical energy.

flywheel synchronization—Automatic frequency control of a scanning circuit. The sweep oscillator responds to the average timing of the sync pulses, not to each pulse individually.

flywheel tuning—A tuning-dial mechanism which uses a heavy flywheel on the control shaft for added momentum, to obtain a smoother tuning action.

f-m—Abbreviation for frequency modulation.

fm/am—A system in which information subcarriers are frequency-modulated and are used to amplitude-modulate the carrier.

f-m broadcast band—The band of frequencies extending from 88 to 108 MHz, which include those assigned to noncommercial educational broadcasting.

f-m broadcast channel—A band of frequencies 200 kHz wide and designated by its center frequency. Channels for f-m broadcast stations begin at 88.1 MHz and continue in successive steps of 200 kHz to and including 107.9 MHz.

f-m broadcast station—A station employing frequency modulation in the f-m broadcast band and licensed primarily for the transmission of radio-telephone emissions intended to be received by the general public.

f-m discriminator—A device which converts frequency variations to proportional variations in the amplitude of an electrical signal. Discriminators may be of several basic types, such as pulse averaging, Foster-Seely, ratio detector, or phase-lock correlation detector.

f-m discriminator (subcarrier)—The same as an f-m discriminator except that it is used to convert subcarrier frequency variations into proportional voltage or current signals.

fm/pm—A system in which information subcarriers are frequency-modulated and used to phase-modulate the carrier.

f-m multiplex—*See* F-M Stereo.

f-m radar—*See* Frequency-Modulated Radar.

f-m recording (magnetic tape)—A method of recording in which the input signal modulates a voltage-controlled oscillator, the output of which is delivered to the recording head.

f-m stereo — Also called f-m multiplex. A means by which f-m radio stations are able to transmit stereophonic program material to specially designed receivers and which is at the same time compatible with monophonic equipment.

f-m stereophonic broadcast—The transmission of a stereophonic program by a single f-m broadcast stations utilizing the main channel and a subchannel to carry the signals required to produce the stereophonic effect.

focal length—Symbolized by *f*. The distance from the principal focus (focus of parallel rays of light) to the surface of a mirror or the optical center of a lens.

focometer—An instrument for measuring the focal length of a lens or an optical system.

focus—1. The convergence of light rays or an electron beam at a selected point. 2. The sharp definition of a scanning beam in television receivers or optical systems.

focus control—On a television receiver, a potentiometer control used for fine focusing of the electron beam. The control varies the first-anode voltage of an electrostatic tube or the focus-coil current of a magnetic tube.

focusing—The process of controlling the convergence and divergence of an electron or light beam.

focusing anode—One of the electrodes used to focus the electron beam in a cathode-ray tube. As its potential changes, so does the electric field, thereby altering the path of the electrons.

focusing coil—The coil around the neck of a cathode-ray tube. It provides a magnetic field, parallel to the electron beam, for controlling the cross-sectional area of the beam on the screen.

focusing electrode—An electrode to which a potential is applied to control the cross-sectional area of the electron beam.

focusing magnet—A permanent magnet assembly that produces a magnetic field for focusing the electron beam in a cathode-ray tube.

fold-back characteristic—*See* Current Limiting.

folded cavity—An arrangement used for producing a cumulative effect in a klystron repeater. This is done by making the incoming wave act in several places on the electron stream from the cathode.

folded-dipole antenna—An antenna comprising two parallel, closely spaced dipole antennas. Both are connected together at their ends, and one is fed at its center.

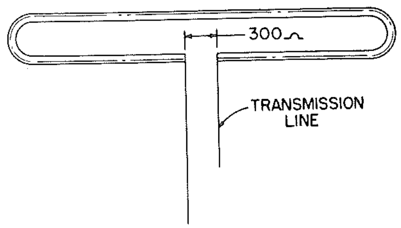

Folded-dipole antenna.

folded heater—A strand of bent, coated wire inserted into a cathode sleeve.

folded horn—An acoustic horn which is

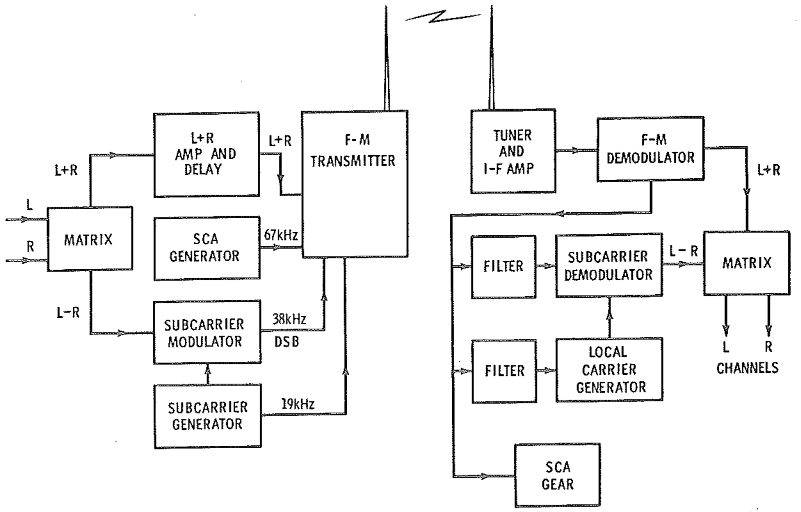

F-m stereophonic broadcast system.

curled to permit more efficient use of the space it occupies.

foldover—A distorted television picture, which appears to overlap horizontally or vertically. It is due to nonlinear horizontal- or vertical-sweep circuits.

follow current—That line current which tends to follow a lightning discharge through an arrester to ground.

follower—A circuit in which the output of a high-gain amplifier is fed directly back to its negative input. The input signal is reproduced without polarity reversal.

follower drive—Also called slave drive. A drive in which the reference input and operation are direct functions of a master drive.

follower with gain—A follower in which only a part of the output voltage is fed back in series opposition to the input signal. Hence, closed-loop gain greater than unity is obtained over the rated range of operation.

following blacks—Also called edge effect, trailing reversal, or trailing blacks. A picture condition in which the edge following a white object is overshaded toward black (i.e., the object appears to have a trailing black border).

following whites—Also called edge effect, trailing reversal, or trailing whites. A picture condition in which the edge following a black or dark gray object is shaded toward white (i.e., the object appears to have a trailing white border).

footcandle—Abbreviated fc. A unit of brilliance of a light. It is the luminance on a surface of one square foot on which a flux of one lumen is uniformly distributed, or the luminance at a surface one foot from a uniform source of one candlepower.

footlambert—A unit of luminance equal to $1/\pi$ candle per square foot, or to the uniform luminance of a perfect diffusing surface emitting or reflecting light at the rate of one lumen per square foot.

foot-pound—A unit of measurement equivalent to the work of raising one pound vertically a distance of one foot.

forbidden band—The energy band lying between the conduction and valence bands. The energy difference across it determines whether a solid acts as a conductor, semiconductor, or insulator.

forbidden energy gap—The energy range of a semiconductor between the bottom of the conduction band and the top of the valence band.

force—1. Any physical action capable of moving a body or modifying its motion. 2. In computer programming, manual intervention which directs the computer to execute a jump instruction.

force-balance transducer—A transducer in which the output from the sensing member is amplified and fed back to an element which causes a force-summing member to return it to its rest position. The magnitude of the signal fed back determines the output of the device, like the error signal in a servo signal.

forced coding—*See* Minimum-Access Programming.

force differential—The difference between the operating force and the release force of a momentary contact switch.

forced oscillation—In a linear constant-parameter system, the response to an applied driving force, excluding the transient which results from energy at the time the driving force is applied.

force factor (of an electromechanical or electroacoustic transducer)—1. The complex quotient of the force required to block the mechanical or acoustic system, divided by the corresponding current in the electrical system. 2. The complex quotient of the resultant open-circuit voltage in the electric system, divided by the velocity in the mechanical or acoustic system.

force-summing device—In a transducer the element directly displaced by the applied stimulus.

fore pump—An auxiliary vacuum pump used as the first stage in evacuating vacuum systems.

fork oscillator—An oscillator in which a tuning fork is the frequency-determining element.

fork tines—The projecting ends of a tuning fork. When vibrated, they produce a constant frequency.

formant—1. The particular frequency region in which the energy of a vowel sound is concentrated most strongly. 2. In an organ, an electrical circuit, the purpose of which is to alter the tone quality of sound amplified by it. A formant filter is applied to the entire output from a manual, rather than to individual tones.

formant filter—A waveshaping network used in an organ to modify the signal from the tone generator so it will assume the waveshape of the desired tone.

form factor—1. Ratio of the effective value of a current to its average value. 2. Shape (diameter/length) of a coil. 3. Ratio of the effective value of a symmetrical alternating quantity to its half-period average value.

Formica—Trade name for a phenolic compound having good insulating qualities.

forming—The application of voltage to an electrolytic capacitor, electrolytic rectifier, or semiconductor device to produce a desired permanent change in electrical characteristics as a part of the manufacturing process.

form-wound coil—An armature coil that is formed or shaped over a fixture before being placed on the armature of a motor or generator. Any coil wound on a fixture or dummy form.

FORTRAN—A computer programming system which enables the programmer to state,

in relatively simple language, the steps of a procedure to be carried out by the computer and to automatically obtain an efficient machine language program.

fortuitous telegraph distortion—Distortion other than bias or characteristic. It occurs when a signal pulse departs from the average combined effects of bias and characteristic distortion for one occurrence. Since fortuitous distortion varies from one signal to another it must be measured by a process of elimination over a long period. It is expressed in a percentage of unit pulse.

forty-five/forty-five—Also called the Westrex system. A system of disc recording in which signals originating from two microphones are impressed on each side of a groove. The two sides are cut 45° from the surface of the record.

forty-five record—A 7-inch record with a 1½-inch center hole. It is recorded at 45 rpm and played at the same speed.

forty-four-type repeater—Type of telephone repeater used in a four-wire system. It employs two amplifiers and no hybrid arrangements.

forward-acting regulator—A transmission regulator which makes an adjustment without affecting the quantity that caused the adjustment.

forward-backward counter—A counter having both an add and subtract input and thus capable of counting in either an increasing or a decreasing direction.

forward bias—An external voltage applied in the conducting direction of a PN junction. The positive terminal is connected to the P-type region, and the negative terminal to the N-type region.

forward coupler—A directional coupler used for sampling incident power.

forward current—The current which flows across a semiconductor junction when a forward-bias voltage is applied.

forward direction—The direction of easy current flow through a semiconductor device when a given voltage within the ratings of the device is applied. In a conventional rectifier or diode, the forward direction is from anode to cathode when the anode is at a positive voltage with respect to the cathode.

forward path—In a feedback control loop, the transmission path from the loop-actuating to the loop-output signal.

forward propagation ionospheric scatter—A radiocommunication technique using the scattering phenomenon exhibited by electromagnetic waves in the 30- to 100-megahertz region when passing through the ionosphere at an elevation of about 85 kilometers.

forward-propagation tropospheric scatter—A method of communication by means of ultrahigh frequency f-m radio. It provides reliable multichannel telephone, teletype, and data transmission without line-of-sight restrictions or the necessity of using wire or cables.

forward recovery time—In a semiconductor diode, the time required for the current or voltage to arrive at a specified condition after instantaneous switching from zero or a specified reverse voltage to a specified forward voltage.

forward resistance—The resistance measured at a specified forward voltage drop or forward current in a rectifier.

forward scatter—1. Propagation of electromagnetic waves at frequencies above the maximum usable high frequency through use of the scattering of a small portion of the transmitted energy when the signal passes from a nonionized medium into a layer of the ionosphere. 2. A term referring collectively to very-high-frequency forward propagation by ionospheric scatter and ultrahigh-frequency forward propagation by tropospheric scatter communication techniques.

forward scattering—The reflected radiation of energy from a target away from the illuminating radar.

forward short-circuit-current-amplification factor—In a transistor, the ratio of incremental values of output to input current when the output circuit is a-c short-circuited.

forward-transfer function—In a feedback control loop, the transfer function of the forward path.

forward voltage—Voltage of the polarity that produces the larger current.

forward voltage drop—The resultant voltage drop when current flows through a rectifier in the forward direction.

forward wave—In a traveling-wave tube, a wave with a group velocity in the same direction the electron stream moves.

Foster-Seeley discriminator—A type of frequency discriminator that converts a frequency-modulated signal into an audio signal. It requires a limiter to prevent random amplitude variations of the f-m signal from appearing in its output.

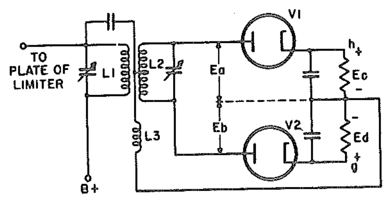

Foster-Seeley discriminator.

Foster's reactance theorem—The driving-point impedance of a finite two-terminal network composed of pure reactances is a reactance which is an odd rational function of frequency and which is completely determined, except for a constant factor, by assigning the resonant and antiresonant frequencies. In other words, the driving-point impedance consists of segments going from minus infinity to plus infinity (except that at zero, or infinite, frequency, a segment may start or stop at zero impedance). The frequencies at which the impedance is infinite are termed poles, and those at which it is zero are termed zeros.

Foucault currents—*See* Eddy Currents.

four-address code—An artificial language for describing or expressing the instructions carried out by a digital computer. In automatically sequenced computers, the instruction code is used for describing or expressing sequences of instructions. Each instruction word then contains a part specifying the operation to be performed, plus one or more addresses which identify a particular location in storage.

four-horn feed—A cluster of four rectangular horn antennas used as the radiating and receiving elements of parabolic or lens-type radar antennas. The four segments of the horn assembly define the four quadrants of information for direction to target sensing. Used on monopulse-type radar systems such as the AN/FPS-16.

Fourier series—A mathematical analysis that permits any complex waveform to be resolved into a fundamental, plus a finite number of terms involving its harmonics.

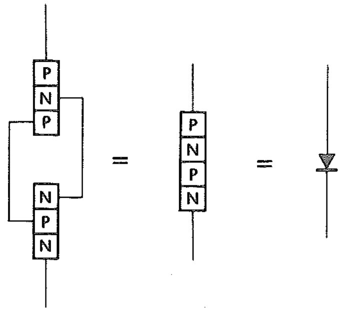

Four-layer diode.

four-layer diode—A semiconductor diode that has three junctions with connections made only to the two outer layers that form the junctions. A Shockley diode is an example.

four-layer transistor—A junction transistor that has four conductivity regions, but only three terminals. A thyristor is an example.

four-level system—A laser involving four electronic energy levels. The ground state (level 1) is pumped to level 4, from which the excited electrons make a downward transition to the upper laser level 3 (or metastable level 3). Then, stimulated transition to the lower laser level 2 occurs, followed by rapid decay to the ground state. The four-level system has the advantage that the pump level and ground state are isolated from the laser action.

four-pole network—*See* Two-Terminal-Pair Network.

four-track recording—Also called quarter-track. On quarter-inch-wide tape, the arrangement by which four different channels of sound may be recorded on adjacent tracks. These may be recorded as four separate and distinct tracks (monophonic), or two related (stereo) pairs of tracks. By convention, tracks 1 and 3 are recorded in the forward direction of a given reel, and tracks 2 and 4 are recorded in the reverse direction.

four-track tape—A tape with four recording strips which permits it to be played in both directions for stereo recordings, like a two-track monophonic tape. Thus, twice as much material can be recorded than on a two-track stereo tape.

four-wire circuit—A two-way circuit with two paths. Each path transmits the electric waves in one direction only. The transmission paths may or may not employ four wires.

four-wire repeater—A telephone repeater used in a four-wire circuit. It has two amplifiers; one amplifies the telephone currents in one side of the four-wire circuit, and the other in the other side.

four-wire terminating set—A hybrid arrangement involving termination of four-wire circuits on a two-wire basis for interconnection with two-wire circuits.

fox message—A standard message which is used for testing teletypewriter circuits and machines because it includes all the alphanumerics on a teletypewriter, as well as most of the function characteristics such as space, figures shift, letters shift, etc. It is: THE QUICK BROWN FOX JUMPED OVER A LAZY DOG'S BACK 1234567890——SENDING. The sending station's identification is inserted in the three space blanks which precede the word "sending."

fractional arithmetic units—Arithmetic units in a computer that is operated with the decimal point at the extreme left so that all numbers have a value less than 1.

fractional-horsepower motor—Any motor having a continuous rating of less than one horsepower.

Frahm frequency meter—A meter that measures the frequency of an alternating current. It consists of a row of steel reeds, each with a different natural frequency. All are excited by an electromagnet fed with the current to be measured. The reed that vibrates is the one with a frequency corresponding most nearly to that of the current.

frame—1. In television, the total area occupied by the picture. In the United States, each frame contains 525 horizontal scanning lines, and 30 complete frames are shown per second. 2. One cycle of a recurring number of pulses. 3. In pam and pdm, one complete commutator revolution or sweep. In pcm, a recurring group of words which includes a single synchronizing signal. 4. The array of binary digits across the width of magnetic or paper tape.

frame frequency—The number of times per second the picture area is completely scanned (30 per second in the United States television system).

frame grid—The grid of a vacuum tube consisting of a rigid welded frame on which tungsten wire is wound under tension, resulting in a firm precision structure that can be positioned accurately. It also allows the use of much finer grid wire, which reduces electron interception and power dissipation in the grid.

frame-grounding circuit—A conductor which is electrically bonded to the machine frame and/or to any conduction parts which are normally exposed to operating personnel. This circuit may further be connected to external grounds as may be required by applicable Underwriters code.

frame of reference—A set of points, lines, or planes used for defining space coordinates.

frame pulse synchronization—Synchronization of the local-channel rate oscillator by comparison and phase lock with the separate frame synchronizing pulses.

framer—A device for adjusting facsimile equipment so that the recorded elemental area bears the same relationship to the record sheet as the corresponding transmitted elemental area bears to the subject copy as the line progresses.

frame rate—*See* Frame Frequency.

frame roll—A momentary roll, or "flip-flop," of a television picture.

frame synchronization signal—In pam, a coded pulse or interval to indicate the start of the commutation frame period. In pcm, any signal used to identify a frame of data.

frame-synchronizing pulse—A recurrent signal that establishes each frame.

frame-synchronizing pulse separator—A circuit for separating frame-synchronizing pulses or intervals from commutated signals.

framing—1. Adjusting the picture to a desired position in the direction of line progression. 2. The process of selecting the bit groupings representing one or more characters from a continuous stream of bits.

framing bits—Also called sync bits. Noninformation-carrying bits used to make possible the separation of characters in a bit stream.

framing control—More often called centering control. A knob (or knobs) for centering and adjusting the height and width of a television picture.

framing magnet—More often called centering magnet. A magnet which centers the televised picture on the face of the tube.

Franklin antenna—An antenna which is several half-wavelengths long and has nonradiating phasing coils between half-wave sections.

Franklin oscillator—A two-terminal feedback oscillator using two tubes or transistors and having sufficient loop gain to permit extremely loose coupling to the resonant circuit.

Fraunhofer region—The region in which the energy from an antenna proceeds essentially as though coming from a source located in the vicinity of the antenna.

free electrons—Electrons which are not bound to a particular atom, but circulate among the atoms of a substance.

free energy—The available energy in a thermodynamic system.

free field—1. Theoretically, a field (wave or potential) which is free from boundaries in a homogeneous, isotropic medium. In practice, a field in which the effects of the boundaries are negligible over the region of interest. 2. A property of information-processing recording media which permits recording of information without regard to a preassigned or fixed field; e.g., information-retrieval-devices information may be dispersed in the record in any sequence or location.

free-field emission—Electron emission that occurs when the electric field at the surface of an emitter is zero.

free grid—A grid electrode that is left unconnected in a vacuum tube. Its potential exerts a control over the plate current.

free impedance—Also called normal impedance. The input impedance of a transducer when the load impedance is zero.

free magnetic pole—A magnetic pole so far from an opposite pole that it is free from the effect of the other pole.

free motional impedance—The complex remainder after the blocked impedance of a transducer has been subtracted from the free impedance.

free net—A net in which any station may communicate with any other station in the same net without first obtaining permission from the control station.

free oscillations—Commonly referred to as shock-excited oscillations. Oscillations that continue in a circuit or system after the applied force has been removed. The frequency of the oscillations is determined by the parameters of the system or circuit.

free-point tube tester—A tester instrument that permits transferring a tube from a circuit to a test panel at which either voltage or current measurement for any electrode of the tube is readily made by plugging a meter into appropriate jacks. Connections to the receiver are made by means of a cord and plug inserted into the socket from which the tube was removed.

free position—The initial position of the actuator of a momentary-contact switch when there is no external force (other than gravity) applied on the actuator, and the switch is in the specified position.

free progressive wave—Also called free wave. A wave free from boundary effects in a medium. In other words, there are no reflections from nearby surfaces. A free wave can only be approximated in practice.

free radicals—Atoms, ionized fragments of atoms, or molecules which combine and release enormous amounts of energy.

free reel—The reel which supplies the magnetic tape on a recorder.

free-rotor gyro — A gyro the rotor of which is supported by a gas-lubricated spherical bearing.

free-running frequency—The frequency at which a normally synchronized oscillator operates in the absence of a synchronizing signal.

free-running local synchronizer oscillator —A free-running oscillator circuit in the decommutator normally triggered by separated channel synchronizing pulses. It supplies substitute pulses for missing channel pulses.

free-running multivibrator — *See* Astable Multivibrator.

free-running sweep — A sweep operating without synchronizing pulses.

free sound field—A field in a medium free from discontinuities or boundaries. In practice it is a field in which the boundaries cause negligible effects over the region of interest.

free space—Empty space, or space with no free electrons or ions. It has approximately the electrical constants of air.

free-space field intensity—The radio-field intensity that would exist at a point in a uniform medium in the absence of waves reflected from the earth or other objects.

free-space loss—The theoretical radiation loss which would occur in radio transmission if all variable factors were disregarded.

free-space propagation—Electromagnetic radiation over a straight-line path in a vacuum or ideal atmosphere, sufficiently removed from all objects that affect the wave in any way.

free-space radar equation—The equation for determining the characteristic of a radar signal propagated between the radar set and a reflected target in free space.

free-space radiation pattern—The radiation pattern of an antenna in free space, where there is nothing to reflect, refract, or absorb the radiated waves.

free-space transmission — Electromagnetic radiation over a straight line in a vacuum or ideal atmosphere sufficiently removed from all objects that affect the wave.

free speed—The angular speed of an energized motor under no-load conditions. (*Also see* Angular Velocity.)

free wave—*See* Free Progressive Wave.

freewheeling circuit—A motor arrangement in which the field is shunted by a half-wave rectifier which discharges the energy stored in the field during the negative half cycles.

freeze-out—A short-time denial of a telephone circuit to a subscriber by a speech-interpolation system.

F-region — The region of the ionosphere above 100 miles.

freq—Abbreviation for frequency.

frequency—Symbolized by f. 1. The number of recurrences of a periodic phenomenon in a unit of time. Electrical frequency is specified as so many hertz. Radio frequencies are normally expressed in kilohertz at and below 30,000 kilohertz and in megahertz above this frequency. 2. Number of complete hertz of a wave motion (e.g., of an alternating current or sound wave).

frequency agility—The rapid and continual shifting of a radar frequency to avoid jamming by the enemy, reduce mutual interference with friendly sources, enhance echoes from targets, or provide necessary patterns of ecm (electronic countermeasures) or eccm (electronic counter-countermeasures) radiation.

frequency allocation — The assignment of available frequencies in the radio spectrum to specific stations, for specific purposes. This is done to yield maximum utilization of frequencies with minimum interference between stations. Allocations in the United States are made by the Federal Communications Commission.

frequency authorization—The document of power which legalizes the assignment of a frequency or a frequency band.

frequency band—A continuous and specific range of frequencies.

frequency band of emission—The frequency band required for a specific type of transmission and speed of signaling.

frequency bias—A constant frequency purposely added to the frequency of a signal.

frequency-change signaling — A telegraph signaling method in which one or more particular frequencies correspond to each desired signaling condition of a telegraph code. The transition from one set of frequencies to the other may be either a continuous or a discontinuous change in frequency or in phase.

frequency changer — *See* Frequency Converter.

frequency-changing circuit—A circuit comprising an oscillator and a mixer and delivering an output at one or more frequencies other than the input frequency.

frequency channel—A continuous portion of the appropriate frequency spectrum for a specified class of emission.

frequency compensation—A method of extending the higher or lower (or both) frequency ranges within which an amplifier has nearly uniform gain.

frequency constant—The number relating the natural vibration frequency of a piezoid (finished crystal blank) to its linear dimension.

frequency conversion—1. The process of converting a signal to some other frequency by combining it with another frequency. 2. Of a heterodyne receiving system, converting the carrier frequency of a received signal from its original value to the intermediate frequency value in a superheterodyne receiver.

frequency converter—Also called frequency changer. A circuit, device, or machine that changes an alternating current from one frequency to another, with or without a change in voltage or number of phases. In a superheterodyne receiver, the oscillator and mixer first-detector stages together serve as a frequency converter.

frequency correction — Compensation, by means of an attenuation equalizer, for unequal transmission of various frequencies in a line.

frequency cutoff—The frequency at which the current gain of a transistor drops 3 db below the low-frequency gain.

frequency demodulation—Removal of the intelligence from a modulated carrier.

frequency departure—The amount a carrier or center frequency deviates from its assigned value.

frequency deviation—In frequency modulation, the peak difference between the instantaneous frequency of the modulated wave and its carrier frequency.

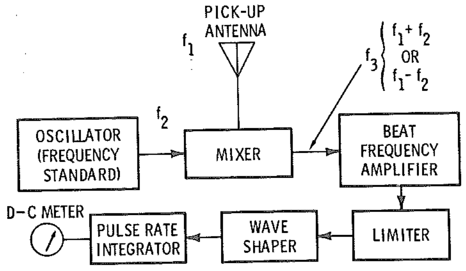

Frequency-deviation meter.

frequency-deviation meter—An instrument that indicates the number of hertz a transmitter has drifted from its assigned carrier frequency.

frequency discrimination — A term applied to the operation of selecting a desired frequency or frequencies from a spectrum of frequencies.

frequency discriminator—A circuit that converts a frequency-modulated signal into an audio signal.

frequency distortion—1. The distortion that results when all frequencies in a complex wave are not amplified or attenuated equally. 2. The impairment of fidelity introduced by a transducer as a result of the unequal transfer of frequencies—e.g., unequal amplification of frequencies with the passband of an amplifier.

frequency distribution—The number of occurrences of particular values plotted against those values.

frequency diversity—*See* Frequency-Diversity Reception.

frequency-diversity reception—Also called frequency diversity. The form of diversity reception which utilizes transmission at different frequencies.

frequency divider—A device delivering an output voltage which is an integral submultiple or proper fraction of the input frequency.

frequency-division multiplex — Abbreviated fdm. A device or process for transmitting two or more signals over a common path by sending each one over a different frequency band.

frequency doubler—An electronic stage having a resonant plate circuit tuned to the second harmonic of the input frequency. The output signal will then have twice the frequency of the input signal.

frequency-doubling transponder — A transponder that doubles the frequency of the interrogating signal before retransmission.

frequency drift—Any undesired change in the frequency of an oscillator, transmitter, or receiver.

frequency-exchange signaling — The method in which the change from one signaling condition to another is accompanied by a decay in amplitude of one or more frequencies and by a build-up in amplitude of one or more other frequencies.

frequency frogging — The interchanging of the frequency allocations of carrier channels to prevent singing, reduce cross talk, and correct for line slope. It is accomplished by having the modulators in a repeater translate a low-frequency group to a high-frequency group and vice versa.

frequency indicator—A device which shows when two alternating currents have the same phase or frequency.

frequency interlace—In television, the relationship of intermeshing between the frequency spectrum of an essentially periodic interfering signal and the spectrum of harmonics of the scanning frequencies. Such relationship minimizes the visibility of the interfering pattern by altering its appearance on successive scans.

frequency keying—A method of keying in which the carrier frequency is shifted between two predetermined frequencies.

frequency-measuring equipment — Equipment for indicating or measuring the frequency or pulse-repetition rate of an electrical signal.

frequency meter—An instrument for measuring the frequency of an alternating current.

frequency-modulated broadcast band—The band from 88.1 to 107.9 megahertz. It is divided into 100 channels, each 200 kilohertz in width, and set aside for frequency-modulated broadcasting.

frequency-modulated, carrier-current telephony—A form of telephony in which a frequency-modulated carrier signal is transmitted over power lines or other wires.

frequency-modulated cyclotron—A cyclotron in which the frequency of the accelerating electric field is modulated in order to hold the positively charged particles in synchronism with the accelerating field despite their much greater mass at very high speeds.

frequency-modulated output—A transducer output which is obtained in the form of a deviation from a center frequency, where the deviation is proportional to the applied stimulus.

frequency-modulated radar—Also called f-m radar. A form of radar in which the radiated wave is frequency-modulated. The range is measured by beating the returning wave with the one being radiated.

frequency-modulated transmitter—One in which the frequency of the wave is modulated.

frequency-modulated wave—A carrier wave whose frequency is varied by an amount proportionate to the amplitude of the modulating signal.

frequency modulation — Modulation of a sine-wave carrier so that its instantaneous frequency differs from the carrier frequency by an amount proportionate to the instantaneous amplitude of the modulating wave. Combinations of phase and frequency modulation also are commonly referred to as frequency modulation. (*Also see* Frequency-Modulated Wave.)

frequency-modulation deviation—The peak difference between the instantaneous frequency of a modulated wave and the carrier or reference frequency.

frequency modulation-frequency modulation—A system in which the frequency-modulated subcarriers are used to frequency modulate a second carrier.

frequency modulation-phase modulation—A system in which the frequency-modulated subcarriers are used to phase-modulate a second carrier.

frequency monitor—An instrument for indicating the amount a frequency deviates from its assigned value.

frequency multiplex—A technique for the transmission of two or more signals over a common path. Each signal is characterized by a distinctive reference frequency or band of frequencies.

frequency multiplier—A device for delivering an output wave whose frequency is a multiple of the input frequency (e.g., frequency doublers and triplers).

frequency offset—The amount by which a frequency lies above or below a reference frequency. For example, if a frequency measures 1.000 001 MHz when compared against a reference frequency of 1.000 000 MHz, then its fractional frequency offset is 1 Hz/1 MHz, or 1 part in 10^6.

frequency offset transponder—A transponder which changes the interrogating signal frequency by a fixed amount before retransmission.

frequency output (transducer)—An output in the form of frequency which is a function of the applied measurand (e.g., angular speed and flow rate).

frequency overlap—1. In a color tv system, that part of the frequency spectrum occupied by both the monochrome and chrominance signals. 2. That part of the frequency band which is shared as a result of interleaving.

frequency pulling—A change in oscillator frequency due to a change in the load impedance.

frequency pushing — A change in the frequency of an oscillator due to a change in anode current or voltage.

frequency range—1. In a transmission system, those frequencies at which the system is able to transmit power without attenuating it more than an arbitrary amount. 2. In a receiver, the frequency band over which the receiver is designed to operate, covering those frequencies the receiver will readily accept and amplify. 3. A designated portion of the frequency spectrum.

frequency record—A recording of various known frequencies at known amplitudes, usually for testing purposes.

frequency regulator—A regulator that maintains the frequency of the frequency-generating equipment at a predetermined value or varies it according to a predetermined plan.

frequency relay—A relay which functions at a predetermined value of frequency. It may be an overfrequency or underfrequency relay, or a combination of both.

frequency response—1. A measure of how effectively a circuit or device transmits the different frequencies applied to it. 2. The portion of the frequency spectrum which can be sensed by a device within specified limits of amplitude error.

frequency-response characteristic — The amount by which the gain or loss of a device varies with the frequency.

frequency-response curve—A graphical representation of the way a circuit responds to different frequencies within its operating range.

frequency-response equalization — Also called equalization or corrective equalization. The effect of all frequency-discriminative means employed in a transmission system to obtain the desired overall frequency response.

frequency run—A series of tests for determining the frequency-response characteristics of a transmission line, circuit, or device.

frequency selectivity—The degree to which a transducer is capable of differentiating between the desired signal and between signals or interference at other frequencies.

frequency-sensitive relay—A relay that operates only when energized with voltage, current, or power within specific frequency limits.

frequency separator — The circuit which separates the horizontal-scanning from the vertical-scanning synchronizing pulses in a television receiver.

frequency shift—A system of telegraph operation in which the space signal is 70 hertz lower than the mark signal.

frequency-shift converter—A device which limits the amplitude of the received frequency-shift signal and then changes it to an amplitude-modulated signal.

frequency-shift indicator—In automatic code transmission, a device which designates mark and space by shifting the carrier back and forth between two frequencies instead of keying it on and off.

frequency-shift keying—Abbreviated fsk. A form of frequency modulation in which the modulating wave shifts the output frequency between predetermined values and the output wave has no phase discontinuity.

frequency-shift telegraphy—Telegraphy by frequency modulation in which the telegraph signal shifts the frequency of the carrier between predetermined values. There is phase continuity during the shift from one frequency to the other.

frequency-shift transmission—A method of transmitting the mark and space elements of a telegraph code by shifting the carrier frequency slightly, usually about 800 Hz.

frequency splitting—One condition of magnetron operation in which rapid alternation occurs from one mode of operation to another. This results in a similar rapid change in oscillatory frequency and consequent loss of power at the desired frequency.

frequency stability—The ability of electronic equipment to maintain the desired operating frequency.

frequency stabilization—The controlling of the center or carrier frequency so that it does not differ more than a prescribed amount from the reference frequency.

frequency standard—A stable low-frequency oscillator used for frequency calibration. It can generate a fundamental frequency of 50 to 100 kilohertz with a high degree of accuracy. Harmonics of this fundamental are then used as reference points for checking throughout the radio spectrum at 50- or 100-kilohertz intervals.

frequency swing—The instantaneous departure of the emitted wave from the center frequency when its frequency is modulated.

frequency synthesizer—A frequency source of high accuracy generally characterized by the fact that the output frequency is composed of two components. The frequency steps, mostly decadic, are derived from the crystal-stabilized frequency standard and the variable frequency of a free-running oscillator which fills in between these steps. The simplest method of frequency synthesis is to derive standard frequencies from a crystal-controlled frequency by harmonic generation and frequency division.

frequency tolerance—The maximum permissible deviation with respect to the reference frequency of the corresponding characteristic frequency of an emission. Expressed in percent or in hertz.

frequency translation—The transfer *en bloc* of signals occupying a definite frequency band, such as a channel or group of channels, from one position in the frequency spectrum to another in such a way that the arithmetic frequency difference of the signals within the band is unaltered.

frequency tripler—An amplifier whose output circuit is resonant to the third harmonic of the input signal. The output frequency is three times the input frequency.

frequency-type telemeter — A telemeter which employs the frequency of a periodically recurring electric signal as the translating means.

frequency-wavelength relation — For radio waves, the frequency in hertz is equal to approximately 300,000,000 divided by the wavelength in meters. The wavelength in meters is equal to approximately 300,000-000 divided by frequency in hertz or 300 divided by frequency in megahertz.

Fresnel—A little-used unit of frequency equal to 10^{12} hertz.

Fresnel lens—A thin lens constructed to have the optical properties of a much thicker lens.

Fresnel loss—*See* Surface Reflection.

Fresnel region—The region between an antenna and the Fraunhofer region.

Fresnel zone—1. An area selected in the aperture of a radiating system so that radiation from all parts of the system reaches some point at which it is desired at a common phase within 180°. 2. A circular zone about the direct path between a transmitter and a receiver at each radius that the distance from a point on this circle to the receiving point has a path length that is some multiple of a half-wavelength longer than the direct path.

frictional electricity—Electric charges produced by rubbing one material against another.

frictional error—1. Applied to pickups, the difference in values measured in percent of full scale before and after tapping, with the measurand constant.

frictional loss—The loss of energy due to friction between moving parts.

frictional machine—Also called a static machine. A device for producing frictional electricity.

friction effects—The difference in resistance or output between readings obtained prior to and immediately after tapping an instrument while applying a constant stimulus. Particularly applicable to potentiometric transducers.

friction error—A change in a reading originally taken in the absence of vibration that occurs after a transducer is tapped or dithered to remove internal friction.

friction-free calibration (transducer) — Calibration under conditions minimizing the effect of static friction often obtained by dithering.

friction tape—A fibrous tape impregnated with a sticky, moisture-resistant compound which provides a protective covering or insulation.

fringe area—The area just beyond the limits of the reliable service area of a television transmitter. Signals are weak and erratic, requiring the use of high-gain directional receiving antennas and sensitive receivers for satisfactory reception.

fringe howl—A squeal or howl heard when some circuit in a receiver is on the verge of oscillation.

fritting—A type of contact erosion in which an electrical discharge makes a hole through the contact film and produces molten matter that is drawn through the hole by electrostatic forces and then solidifies and forms a conducting bridge.

Frohlich high-temperature breakdown theory—A thermal mechanism for breakdown in which electrons rather than ions carry the current. The necessary number of conduction electrons is produced by thermal excitation of the electrons in impurity and imperfection levels.

Frohlich low-temperature breakdown theory—Also called the high-energy criterion. Similar to the Von Hippel theory, except that an electron energy distribution is assumed. Only a few electrons on the high-energy tail of the distribution must gain the necessary critical energy.

front contact—A movable relay contact which closes a circuit when the associated device is operated.

front end—The section of a tuner or receiver which is used to select the desired station from either the a-m, f-m, or tv band, and to convert the r-f signal to i-f. To do its job properly, a front end requires a high-gain, low-noise r-f stage, a mixer, and an oscillator. The degrees to which a desired station can be received without interference and without adding noise is expressed by sensitivity, selectivity, and signal-to-noise ratio.

front-end rejection—A db expression of the relative ability of a receiver network to reject signals outside the tuned bandwidth.

front porch—In the composite television signal, that portion of the synchronizing signal (at the blanking or black level) preceding the horizontal-sync pulse at the end of each active horizontal line. The standard EIA signal is 1.27 microseconds in duration.

front projection—A system of picture enlargement using an opaque reflective screen. The projector and viewers are on the same side of the screen.

front-surface mirror—An optical mirror on which the reflecting surface is applied to the front of the mirror instead of the back.

front-to-back ratio—Also called front-to-rear ratio. 1. The ratio of power gain between the front and rear of a directional antenna. 2. Ratio of signal strength transmitted in a forward direction to that transmitted in a backward direction. For receiving antennas, the ratio of received-signal strength when the source is in the front of the antenna to the received-signal strength when the antenna is rotated 180°.

front-to-rear ratio—*See* Front-to-Back Ratio.

fruit—*See* Fruit Pulse.

fruit pulse—Also called fruit. Radar beacon system video display of a synchronous beacon return which results when several interrogator stations are located within the same general area. Each interrogator receives its own interrogated reply as well as many synchronous replies resulting from interrogation of the airborne transponders by other ground stations.

F-scan—*See* F-Display.

F-scope—*See* F-Display.

fsk—*See* Frequency-Shift Keying.

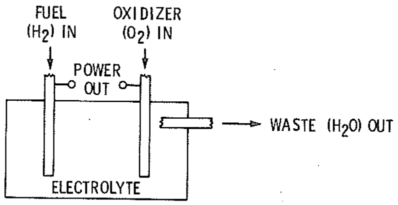

Fuel cell.

fuel cell—An electrochemical generator in which the chemical energy from the reaction of air (oxygen) and a conventional fuel is converted directly into electricity. A fuel cell differs from a battery in that it uses hydrocarbons (or some derivative such as hydrogen) for fuel, and it operates continuously as long as fuel and air are available.

full duplex—A method of operation of a communication circuit in which each end can simultaneously transmit and receive. When used on a radio circuit, duplex operation requires two frequencies.

full-duplex operation—Simultaneous operation in opposite directions in a telegraph system.

full excursion—The application of a measurand, in a controlled manner, over the entire range of a transducer.

fullhouse—A multichannel radio control system in which all controls work to allow the model to fly a complete flight pattern.

full load—The greatest load a piece of equipment is designed to carry under specified conditions.

full-pitch winding—A type of armature winding in which the number of slots between the sides of the coil equals the pole-pitch measure in the slots.

full scale—The total interval over which an instrument is intended to be operated. Also, the output from a transducer when the maximum rated stimulus is applied to the input.

full-scale cycle—The complete range of an instrument, from minimum reading to full scale and back to minimum reading.

full-scale output—The algebraic difference in electrical output between the maximum and minimum values of measurand over which an instrument is calibrated. When the sensitivity slope is given by any other line than the end-point sensitivity, full scale expresses the algebraic difference, for the span of the instrument, which is calculated from the slope of the straight line from which nonlinearity is determined.

full-scale sensitivity—*See* Full-Scale Output.

full-scale value of an instrument—The largest actuating electrical quantity which can be indicated on the scale; or, for an instrument having its zero between the ends of the scale, the sum of the values of the actuating electrical quantity corresponding to the two ends.

full-track recording—Defines the track width as essentially equal to the tape width. Applies to quarter-inch-wide (or less) tape only.

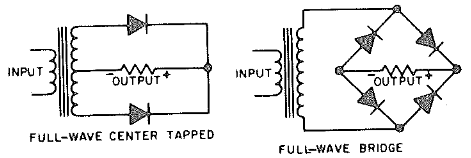

Full-wave rectifier.

full-wave rectifier—A circuit electron tube or other device which uses both positive and negative alternations in an alternating current to produce direct current.

full-wave rectifier tube—A tube containing two sets of rectifying elements, to provide full-wave rectification.

full-wave vibrator—A vibrator having an armature that moves back and forth between two fixed contacts, so as to change the direction of the direct-current flow through a transformer at regular intervals and thereby permit voltage step-up by the transformer. Used in battery-operated power supplies for mobile and marine radio equipment.

function—1. A quantity the value of which depends on the value of one or more other quantities. 2. A specific purpose of an entity, or its characteristic action.

functional device—*See* Integrated Circuit.

functional electronic block—A fabricated device serving a complete electronic function, such as amplification, without other individual components or conducting wires except those required for input, power, and output.

functional interface—An interface between the operating characteristics of equipment such as electrical power and signal characteristics, signal timing, and environmental coupling.

functional parts—Discrete items defined by functional characteristics and dimensions which are not repairable with the use of spare parts (e.g., resistors, capacitors, diodes, potted transformers, permanently sealed batteries, etc.).

function characteristic—The relationship between the output ratio and the shaft position of a precision potentiometer.

function codes—Codes which appear in tape or cards to operate machine functions, such as carriage returns, space, shift, skip, tabulate, etc.

function digit—A coded instruction used in a computer for setting a branch order to link subroutines into the main program.

function generator—A device capable of generating one or more desired waveforms.

functioning time—In a relay, the time that elapses between energization and operation or between de-energization and release.

functioning value—In a relay, the value of applied voltage, current, or power at which operation or release occurs.

function switch—1. A network or system having a number of inputs and outputs. When signals representing information expressed in a certain code are applied to the inputs, the output signals will represent the input information in a different code. 2. In adapters or control units, the switch which determines whether the system plays as a monophonic or stereophonic unit; it may parallel the speakers or cut out one or the other, switch amplifiers from one speaker to the other, reverse channels, etc.

function table—1. A table of values for mathematical function. 2. A hardware device or a computer program which translates one representation of information into another.

function unit—A device which can store a functional relationship and release it continuously or sporadically.

fundamental—*See* Fundamental Frequency.

fundamental component—The fundamental frequency component in the harmonic analysis of a wave.

fundamental frequency—The principal component of a wave; i.e., the component with the lowest frequency or greatest amplitude. It is usually taken as a reference.

fundamental harmonic—The harmonic component with the lowest frequency.

fundamental mode—1. *See* Dominant Mode. 2. Of vibration, the mode having the lowest natural frequency.

fundamental piezoelectric crystal unit—A unit designed to use the lowest resonant frequency for a particular mode of vibration.

fundamental tone—1. In a periodic wave, the component corresponding to the fundamental frequency. 2. In a complex tone, the component tone of lowest pitch.

fundamental units—Units arbitrarily selected to serve as the basis of an absolute system of units.

fundamental wavelength—The wavelength corresponding to the fundamental frequency.

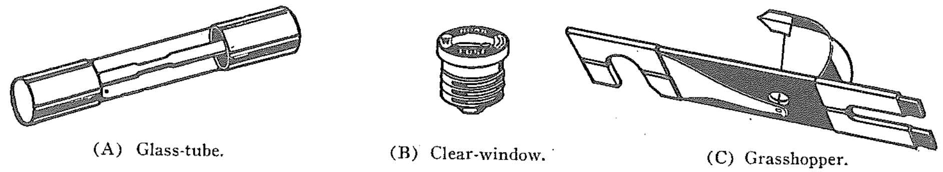

(A) Glass-tube. (B) Clear-window. (C) Grasshopper.

Fuses.

In an antenna, the lowest resonant frequency of the antenna alone, without inductance or capacitance.

funguspro of—To chemically treat a material, component, or unit to prevent the growth of fungus spores.

fuse—A protective device, usually a short piece of wire but sometimes a chemical compound, which melts and breaks the circuit when the current exceeds the rated value.

fuse alarm—A circuit which produces a visual and/or audible signal to indicate a blown fuse.

fuse block—An insulating base on which fuse clips or other contacts for holding fuses are mounted.

fuse box—An enclosed box containing fuse blocks and fuses.

fuse clips—Contacts on the fuse support for connecting the fuse holder into the circuit.

fused disconnect — Generally an air-break switch with a fusing unit in the blade. Used for opening and closing high-voltage circuits.

fused junction—*See* Alloyed Junction.

fused quartz—A glasslike insulating material having exceptional heat and acid resistance.

fused semiconductor—The junction formed by recrystallization on a base crystal from a liquid phase of one or more components and the semiconductor.

fuse filler—Material placed within the fuse tube to aid circuit interruption.

fuse holder—A device for supporting a fuse and providing connections for its terminals.

fuse link—In a fuse, the current-carrying portion which melts when the current exceeds a predetermined value.

Fusestat—Trade name for a time-delay fuse similar to a Fusetron. It has a sized base requiring a permanent socket adapter, which prevents insertion of a fuse or Fusetron of an incorrect rating.

Fusetron—A screw-plug fuse that permits up to 50% overload for short periods of time without blowing.

fuse tube—The insulated tube enclosing a fuse link.

fuse unit—An assembly comprised of a fuse link mounted in a fuse holder, which contains parts and materials essential to the operation of the fuse link.

fuse wire—A wire made from an alloy that melts at a relatively low temperature.

fusible resistor—A resistor designed to protect a circuit against overload by opening when the current drain exceeds the design limits.

fusible wire—A wire used in fire-alarm circuits. It is made of an alloy with a low melting point.

fusion—Also called atomic fusion or nuclear fusion. The merging of atomic nuclei, under extreme heat (millions of degrees), to form a heavier nucleus. The fusion of two nuclei of light atoms is accompanied by a tremendous release of energy.

fv—Abbreviation for femtovolt (10^{-15} volt).

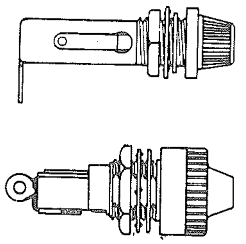

Fuse holders.

G

G—1. Symbol for conductance, the grid of a vacuum tube, a generator, or ground. 2. Abbreviation for giga (10^9).

g—Also called G-force. Symbol for the acceleration of a free-falling body due to the earths gravitational pull. Equal to 32.17 feet per second per second.

G/A (ground-to-air)—Communication with airborne objects from the ground.

GA coil—A coil wound with air spaces between its turns and layers to reduce the capacitance.

gage—Also spelled gauge. An instrument or means for measuring or testing. By extension, the term is often used synonymously with transducer.

gage pressure—1. A differential pressure measurement using the ambient pressure as a reference. 2. A pressure in excess of a standard atmosphere at sea level (i.e., 14.7 pounds per square inch).

gage pressure transducer — A pressure transducer that uses ambient pressure as the reference pressure. The sensing element is normally vented to the ambient pressure.

gain—Also called transmission gain. 1. Any increase in power when a signal is transmitted from one point to another. Usually expressed in decibels. Widely used for denoting transducer gain. 2. The ratios of voltage, power, or current with respect to a standard or previous reading.

gain bandwidth product — The product of the closed-loop gain of an operational amplifier and its corresponding closed-loop bandwidth. This product is often constant in operational amplifiers.

gain control—A device for varying the gain of a system or component.

gain margin—The amount of gain change of an operational amplifier at 180° phase-shift angle frequency that would produce instability.

gain-sensitivity control — *See* Differential Gain Control.

gain stability—The extent to which the sensitivity of an instrument remains constant with time. (The property reported in specifications should be instability, which is the maximum change in sensitivity from the initial value over a stated period of time under stated conditions.)

gain-time control—*See* Sensitivity Time Control.

galena—A bluish-gray, crystalline form of lead sulphide often used as the crystal in a variable crystal detector.

galloping ghost — *See* Proportional Control.

galvanic—An early term for current resulting from chemical action, as distinguished from electrostatic phenomena.

galvanic anode—A source of emf for cathodic protection provided by a metal less noble than the one to be protected (i.e., magnesium, zinc, or aluminum as used for cathodic protection of steel).

galvanic cell—An electrolytic cell capable of producing electric energy by electrochemical action.

galvanic corrosion — Accelerated electrochemical corrosion produced when one metal is in electrical contact with another more noble metal, both being in the same corroding medium, or electrolyte, with a current between them. (Corrosion of this type usually results in a higher rate of solution of the less noble metal and protection of the more noble metal.)

galvanic current—An electrobiological term for unidirectional current such as ordinary direct current.

galvanic series—A list of metals and alloys arranged in the order of their relative potentials (ability to go into solution) in a given environment. The table of potentials is arranged with the anodic, or least noble, metals at one end and the cathodic, or more noble, metals at the other. (For marine use, the potentials listed are related to a seawater environment.)

galvanizing—The coating of steel with zinc to retard corrosion.

galvanometer—An instrument for measuring an electric current. This is done by measuring the mechanical motion produced by the electromagnetic or electrodynamic forces set up by the current.

galvanometer constant—The factor by which a certain function of a galvanometer reading must be multiplied to obtain the current in ordinary units.

galvanometer lamp—The lamp that illuminates a movable mirror of the galvanometer in some spectrophotometers. The angle at which the light is reflected to the galvanometer scale depends on the amount of current through the galvanometer coil.

galvanometer recorder (for photographic recording)—A combination of a mirror and coil suspended in a magnetic field. A signal voltage, applied to the coil, causes a light beam from the mirror to be reflected across a slit in front of a moving photographic film.

galvanometer shunt—A resistor connected in parallel with a galvanometer to increase the range of the instrument. The resistor limits the current to a known fraction and thus prevents excess current from damaging the galvanometer.

gamma—A numerical indication of the degree of contrast in a televised image. It is obtained by making a plot of the log of the output magnitude (ordinate) against the log of the input magnitude (abscissa), measured from a point corresponding to some reference black level. Then, take a straight line which approximates this plot over the region of interest and take its slope.

gamma correction—Introduction of a nonlinear output-input characteristic for the purpose of changing the effective value of gamma.

gamma ferric oxide—The magnetic constituent of practically all present-day tapes, in the form of a dispersion of fine acicular particles within the coating.

gamma radiation—The radiation of gamma rays.

gamma rays—The emission from certain radioactive substances. They are electromagnetic radiations similar to X-rays, but with a shorter wavelength.

gang—To mechanically couple two or more variable capacitors, switches, potentiometers, or other components together so they can be operated from a single control knob.

gang capacitor—Also called gang tuning capacitor. Two or more variable tuning capacitors mounted on the same shaft and controlled by a single knob, but each capacitor tuning a different circuit. Thus, more than one circuit can be tuned simultaneously by a single control.

gang control—Simultaneous control of several

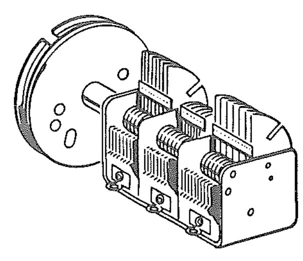

Gang capacitor.

similar pieces of apparatus with one adjustment knob or other control.

ganged tuning—Simultaneous tuning of two or more circuits with a single mechanical control.

gang punch—To punch information that is identical or constant into all of a group of punch cards.

gang switch—A number of switches mechanically coupled for simultaneous operation, but electrically connected to different circuits. In one common form, two or more rotary switches are mounted on the same shaft and operated by a single control.

gang tuning capacitor—*See* Gang Capacitor.

gap—1. In a magnetic circuit, the portion that does not contain ferromagnetic material (e.g., an air space). 2. The space between two electrodes in a spark gap. 3. The distance between the poles of a recording head.

gap arrester—A type of lightning arrester comprising a number of air gaps in series between metal cylinders or cones.

gap coding—In navigation, a process of communicating by interrupting the transmission of an otherwise continuous signal so that the interruptions form a telegraphic message.

gap depth—The dimension of the gap in a recording head, measured perpendicular to the surface of the head.

gap filler—1. A lightweight radar set used to fill in gaps in the coverage pattern of an early-warning radar net. 2. An auxiliary radar antenna used to fill in gaps in the pattern of the main radar antenna.

gap filling—Electrical or mechanical rearrangement of an antenna array, or the use of a supplementary array, to produce lobes where gaps previously occurred.

gap insulation—Insulation wound in a gap.

gap length—The dimension of the gap of a recording head, measured from one pole face to the other. In longitudinal recording, the gap length is the dimension of the gap in the direction of tape travel.

gap loss—The loss in output attributable to the finite gap length of the reproduce head. The loss increases as the wavelength decreases, amounting to approximately 4 db when the wavelength is equal to twice the gap length, and subsequently increases rapidly toward a complete extinction of output when the wavelength is approximately equal to 1.15 times the gap length.

gap motor—A spark-gap drive motor.

gap scatter—In a computer, the deviation from true vertical alignment of the gaps of the magnetic readout heads for the several parallel tracks.

gap width—The dimension of the gap, measured in the direction parallel to the head surface and pole faces. The gap width of the record head governs the track width. The gap widths of reproduce heads are sometimes made appreciably less than those of the record heads to minimize tracking errors.

garbage—In a computer, a slang term for unwanted and meaningless information carried along in storage. Sometimes called hash.

garble—Faulty transmission, reception, or encoding which renders the message incorrect or unreadable.

garnet maser—*See* Ferromagnetic Amplifier.

garter spring—In facsimile, the spring fastened around the drum to hold the record sheet or copy in place.

gas—One of the three states of matter. An aeriform fluid having neither independent shape nor volume, but tending to spread out and occupy the entire enclosure in which it is placed. Gases are formed by heating a liquid above its boiling point.

gas amplification—Ratio of the charge collected to the charge liberated by the initial ionizing of the gas in a radiation counter.

gas-amplification factor—Ratio of radiant or luminous sensitivities with and without ionization of the gas contained in a gas phototube.

gas cleanup—The tendency of many gas-filled tubes to lose their gas pressure and hence become inoperable. This occurs when the ions of gas are driven at high velocity into the metal parts or the glass envelope of the tube where they form stable compounds and are lost as far as the tube is concerned.

gas current—The current which flows in the grid circuit of a vacuum tube when gas ions within the tube are attracted by the grid.

gas detector—An instrument used to indicate the concentration of harmful gases in the air.

gas diode—A tube having a hot cathode and an anode in an envelope containing a small amount of an inert gas or vapor. When the anode is made sufficiently positive, the electrons flowing to it collide with gas atoms and ionize them. As a result, the anode current is much greater than that for a comparable vacuum diode.

gas-electric drive—A self-contained power-conversion system comprising an electric generator driven by a gasoline engine. The generator in turn supplies power to the driving motor or motors.

gaseous electronics—The field of study involving the conduction of electricity through gases and a study of all atomic-scale collision phenomena.

gaseous tube—An electronic tube into which a small amount of gas or vapor is introduced after the tube has been evacuated. Ionization of the gas molecules during operation of the tube affects its operating characteristics.

gaseous-tube generator—A power source comprising a gas-filled electron-tube oscillator and a power supply, plus associated controls.

gas-filled cable—A coaxial or other type of cable containing gas under pressure which serves as insulation and keeps out moisture.

gas-filled lamp—A tungsten-filament lamp containing nitrogen or an inert gas such as argon.

gas-filled radiation-counter tube—A gas tube used for the detection of radiation. It operates on the principle that radiation will ionize a gas.

gas-filled tube rectifier—A rectifier tube in which a unidirectional flow of electrons from a heated electrode ionizes the inert gas within the tube. In this way, rectification is accomplished.

gas focusing—Also called ionic focusing. The use of an inert gas to focus the electron beam in a cathode-ray tube. Beam electrons ionize the gas molecules, forming a core of positive ions along the beam path which tends to attract beam electrons, making the beam more compact.

gasket-sealed relay—A relay in an enclosure sealed with a gasket.

gas laser—A laser where the active medium is a discharge in a gas within a glass or quartz tube that has a Brewster window at each end. This gas can be excited by a high-frequency oscillator or by a direct current flowing between electrodes inside the tube to pump the medium so as to obtain a population inversion.

gas magnification—The increase in current through a phototube due to ionization of the gas within the tube.

gas maser—A maser in which the microwave electromagnetic radiation interacts with the molecules of a gas such as ammonia. Use is limited chiefly to highly stable oscillator applications, as in atomic clocks.

gas noise—Electrical noise produced by erratic motion of gas molecules in gas or partially evacuated vacuum tubes.

gas-phase laser—A continuous-wave device for general experimental work with coherent light. It employs a resonator made up of a fused-silica plasma tube 60 cm long having internal, multilayer, dielectric-coated confocal reflectors of optical-grade fused silica.

gas phototube—A phototube into which a quantity of gas has been introduced, usually to increase its sensitivity.

gas ratio—The ratio of the ion current in a tube to the electron current that produces it.

gassiness—The presence of unwanted gas in a vacuum tube, usually in relatively small amounts. It is caused by leakage from outside the tube or by evolution from its inside walls or elements.

gassing—1. Evolution of a gas from one or more electrodes during electrolysis. 2. The production of gas in a storage battery when the charging current is continued after the battery has been completely charged.

gassy tube—*See* Soft Tube.

gaston—A modulator that produces a random-noise modulation signal from a gas tube. It may be attached to any standard aircraft communications transmitter to provide a counterjamming modulation.

gas tube—A partially evacuated electron tube containing a small amount of gas. Ionization of the gas molecules is responsible for the current flow.

gas-tube relaxation oscillator—A relaxation oscillator in which the abrupt discharge is provided by the breakdown of the gas in the tube.

gas X-ray tube—An X-ray tube in which electron emission from the cathode is produced by bombarding it with positive ions.

gate—1. A circuit having two or more inputs and one output, the output depending upon the combination of logic signal at the inputs. There are four gates, called AND, OR, NAND, and NOR. The definitions assume postive logic is used. In computer work, a gate is often called an AND circuit. 2. A signal used to trigger the passage of other signals through a circuit. 3. An electrode in a field-effect transistor. 4. An output element of a cryotron.

gate circuit—A circuit that passes a signal only when a gating pulse is present.

gate-controlled switch—Also called gate turn-off switch. A three-junction, three-terminal, solid-state device, constructed very much like a silicon controlled rectifier except that it has a turn-off ability, which is controlled by a negative-current pulse applied to the gate.

gate current—Instantaneous current flowing between the gate and cathode of a silicon controlled rectifier.

gate current for firing—Gate current required to fire a silicon controlled rectifier when the anode is at +6 volts with respect to the cathode and with the device at stated temperature conditions.

gated-beam detector—A single-stage f-m detector using a gated-beam tube.

gated-beam tube—A five-element tube in which the electrons flow in a beam between the cathode and plate. A small increase in voltage on the limiter grid will cut off the plate current, and further increases will have a negligible effect on it.

gated buffer—A low-impedance inverting driver circuit that may be used as a line driver for pulse differentiation or in multivibrators.

gated flip-flop—One that has a steering circuit that prevents both flip-flop outputs from becoming "0" at the same time. Alternating-current input pulses must be used to prevent the flip-flop from oscillating. This situation can arise if both input lines are made high simultaneously.

gated sweep—Sweep in which the duration as well as the starting time is controlled to exclude undesired echoes from the indicator screen.

gated transistor—A transistor in which a gate electrode covers the emitter and collector junctions. This allows the application of an electric field at the surface of the base region.

gate electrode—The control electrode of a FET, MOST controlled rectifier, or triac.

gate generator—A circuit or device used to produce one or more gate pulses.

gate impedance—The impedance of a gate winding in a magnetic amplifier.

gate power dissipation — The power dissipated between the gate and cathode terminals of a silicon controlled rectifier.

gate-producing multivibrator—A rectangular-wave generator that produces a single positive or negative gate voltage only when triggered by a pulse.

gate pulse—A pulse that enables a gate circuit to pass a signal. The gate pulse generally has a longer duration than the signal to insure time coincidence.

gate signal—That signal generated by some form of delay circuit required in connection with beam switching, automatic following, the application of agc to a selected echo, and many other purposes.

gate trigger current—In a controlled rectifier, the minimum gate current, for a given anode-to-cathode voltage, required to switch the rectifier on.

gate trigger voltage—In a controlled rectifier, the gate voltage that produces the gate trigger current.

gate tube—A tube which does not operate unless two signal voltages, derived from two independent circuits, are applied simultaneously to two separate electrodes.

gate turn-off current—In a controlled rectifier, the minimum gate current, for a given collector current in the on state, required to cause the rectifier to switch off.

gate turn-off switch — *See* Gate-Controlled Switch.

gate turn-off voltage—In a controlled rectifier, the gate voltage required to produce the gate turn-off current.

gate voltage—1. The voltage across the gate-winding terminals of a magnetic amplifier. 2. The instantaneous voltage between gate and cathode of a silicon controlled rectifier with anode open.

gate winding—The reactor winding that produces the gating action in a magnetic amplifier.

gating—1. Selecting those portions of a wave which exist during certain intervals or which have certain magnitudes. 2. Applying a rectangular voltage to the grid or cathode of a cathode-ray tube, to sensitize it during the sweep time only. 3. Application of a specific waveform to perform electronic switching. *Also see* Blanking.

gating circuit—A circuit that operates as a selective switch and allows conduction only during selected time intervals or when the signal magnitude is within specified limits.

gating pulse—A pulse which modifies the operation of a gate circuit.

gauge—*See* Gage.

gauge pressure—*See* Gage Pressure.

gauss — The centimeter-gram-second electromagnetic unit of magnetic induction. One gauss represents one line of flux (one maxwell) per square centimeter.

Gaussian distribution — Also called normal distribution. A density function of a population which is bell-shaped and symmetrical and which is completely defined by two independent parameters, the mean and the standard deviation.

Gaussian noise — Unwanted electrical disturbances or perturbations described by a probability density function that follows a normal law of statistics.

Gaussian random vibration — *See* Random Vibration.

gaussmeter—An instrument that provides direct readings of magnetic field density (flux density) by virtue of the interaction with an internal magnetic field.

Gauss's theorem—The summation of the normal component of the electric displacement over any closed surface is equal to the electric charge within the surface.

GCA—Abbreviation for ground-controlled approach.

GCI—Abbreviation for ground-controlled interception.

GCT or Gct — Abbreviation for Greenwich civil time.

G-display — Also called G-scan or G-scope. In radar, a rectangular display in which a target appears as a laterally-centralized blip on which wings appear to grow as the target approaches. Horizontal and vertical aiming errors are indicated by horizontal and vertical displacement of the blip.

gear—An element shaped like a toothed wheel which engages one or more similar wheels. The energy transmitted can be stepped up or down by making the driven gears of different sizes.

geared synchro system—A system in which the transmitting and receiving synchros turn at a higher speed than the input and output shafts. Geared systems are generally used when a high degree of accuracy is required.

gearmotor—A train of gears and a motor used for reducing or increasing the speed of the driven object.

Geiger counter—Also called Geiger-Mueller or G-M counter. A radiation detector that uses a Geiger-Mueller counter tube, an amplifier, and an indicating device. The tube consists of a thin-walled gas-filled metal cylinder with a projecting electrode. Nuclear particles enter a window in the metal cylinder and temporarily ionize the gas, causing a brief pulse discharge. These pulses, which appear at the projecting electrode, are amplified and indicated visibly or audibly.

Geiger-Mueller counter—*See* Geiger Counter.

Geiger-Mueller counter tube—A radiation-counter tube designed to operate in the Geiger-Mueller region.

Geiger-Mueller region—Also called Geiger region. The voltage interval in which the pulse size is independent of the number of primary ions produced in the initial ionizing event.

Geiger-Mueller threshold — Also called Geiger threshold. The lowest voltage at which all pulses produced in the tube by any ionizing event are of the same size regardless of the size of the primary ionizing event. This threshold is the start of the Geiger region where the counting rate does not substantially change with applied voltage.

Geiger region—*See* Geiger-Mueller Region.

Geiger threshold — *See* Geiger-Mueller Threshold.

Geissler tube—A gas filled dual-electrode discharge tube that glows when electric current passes through the gas.

gel—A material composed of a solid held in a liquid.

genemotor—A type of dynamotor having two armature windings. One winding serves as the driving motor and operates from the vehicle battery. The other winding functions as a high-voltage d-c generator for operation of mobile equipment.

general-purpose computer—A computer designed to solve a wide variety of problems, the exact nature of which may have been unknown before the computer was designed.

general-purpose motor—A motor of 200 hp or less and 450 rpm or more, rated for continuous operation, having standard ratings, and suitable for use without restriction to a particular application.

general-purpose relay — A relay that is adaptable to a variety of applications.

general rate—The amount of time taken by the creation of electron-hole pairs in a semiconductor.

general routine — A computer routine designed to solve a general class of problems, but when appropriate parametric values are supplied, it specializes in a specific problem.

generated noise—1. In potentiometric transducers, the noise which is attributable to causes such as the generation of emf when dissimilar metals are rubbed against each other, or the emf resulting from the thermocouple effects at points where dissimilar metals are joined.

generating electric field meter—Also called a gradient meter. A device for measuring the potential gradient at the surface of a conductor. A flat conductor is alternately exposed and then shielded from the electric field to be measured. The resultant current in the conductor is then rectified and used as the measure.

generating magnetometer—Also called earth inductor. A magnetometer which measures an electric field by the amount of emf generated in a coil rotated in the field.

generating station—An installation that produces electric energy from chemical, mechanical, hydraulic, or some other form of energy.

generating voltmeter—Also called a rotary voltmeter. A device which measures voltage. A capacitor is connected across the voltage, and its capacitance is varied cyclically. The resultant current in the capacitor is then rectified and used as a measure.

generation rate—The time rate of creation of electron-hole pairs in a semiconductor.

generator—1. Symbolized by G. A rotating machine which converts mechanical energy into electrical energy. 2. An electronic device which converts d-c voltage to alternating current of the desired frequency and waveshape.

generator efficiency—1. In a generator, the ratio between the power required to drive the generator and the output power obtained from it. 2. In a thermoelectric couple, the ratio of the electrical power output to the thermal power input. It is an idealized efficiency assuming perfect thermal insulation of the thermoelectric arms.

generator voltage regulator—A regulator which maintains or varies the voltage of a synchronous generator, capacitor motor, or direct-current generator at or within a predetermined value.

geodesic—The shortest line between two points on a given surface.

geometric distortion—In television, any geometric dissimilarity between the original scene and reproduced image.

geometric mean—The square root of the product of two quantities.

george box — An amplitude-sensitive device employed in an i-f amplifier. It rejects jamming signals of insufficient amplitude to operate its circuits; however, jamming signals having sufficient amplitude are not affected.

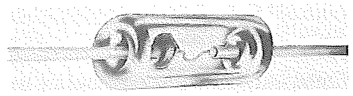

Germanium diode.

germanium diode—A rectifier or detector made from a germanium crystal.

germanium (Ge) — A brittle, grayish-white metallic element having semiconductor properties. Widely used in transistors and crystal diodes. Its atomic number is 32.

germanium transistor — A transistor in which germanium is the semiconducting material.

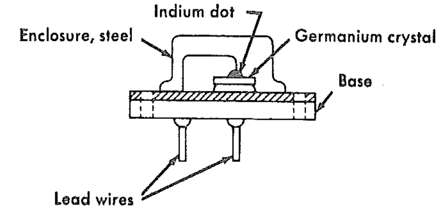

Germanium (power) transistor.

German silver—Usually called nickel silver. A silverish alloy of copper, zinc, and nickel.

getter—An alkali metal introduced into a vacuum tube during manufacture. It is fired after the tube has been evacuated, to react chemically with and eliminate any remaining gases. The getter then remains inactive inside the tube. The silvery deposit sometimes seen on the inside of the glass envelope is due to getter firing.

G force—*See* g.

G/G (ground-to-ground) — Communication between two points on the ground.

ghost—*See* Ghost Image.

ghost image—Also called ghost. An undesired duplicate image offset somewhat from the desired image as viewed on a television screen. It is due to a reflected signal traveling over a longer path and, hence, arriving later than the desired signal. It may be eliminated by the use of a directional antenna array which receives signals over only one path.

ghost pulse—*See* Ghost Signal.

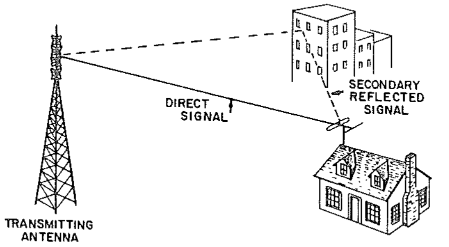

Ghost signal.

ghost signal—Also called ghost pulse. An unwanted signal on the screen of a radar indicator. Echoes which experience multiple reflections before reaching the receiver are an example.

giant grid—An extensive regional or national system of backbones and networks.

giant ties—*See* Interconnection.

Gibson girl—A portable, hand-operated transmitter used by pilots forced down at sea.

giga—A prefix meaning one billion, or 10^9.

gigacycle—One kilomegacycle, or one billion cycles. An obsolete term. The currently preferred term is gigahertz.

gigahertz—A term for 10^9 cycles per second. Used to replace the more cumbersome and obsolete term kilomegacycle. Abbreviated GHz.

gigawatt—Abbreviated GW. One thousand megawatts (10^9 watts).

gigohm—One thousand megohms (10^9 ohms).

gilbert—A cgs unit of the magnetomotive force required to produce one maxwell of magnetic flux in a magnetic circuit of unit reluctance. Magnetomotive force in gilberts $= 10/4\pi$ ampere-turns.

gilbert per centimeter—The practical cgs unit of magnetic intensity. Gilberts per centimeter are the same as oersteds.

Gill-Morrell oscillator—A retarding-field oscillator in which the oscillation frequency depends not only on the electron-transit time within the tube, but also on the associated circuit parameters.

gill selector—A slow-acting telegraph sender and calling key for selective signaling.

gimbal—A mechanical frame having two perpendicularly intersecting axes of rotation.

gimmick—A capacitance formed by twisting two insulated wires together or bringing two conductors into close proximity.

gimp—A slang name given to the extremely flexible wire used in telephone cords and similar equipment. This wire cannot be directly soldered to, as it is a metallic cloth-type material.

Giorgi system—*See* Mksa System of Units.

glass electrode — In electronic pH measurement, an electrode used for determining the potential of a solution with respect to a reference electrode. The calomel type is the most common.

glassiviation—A method of transistor passivation by a pyrolytic glass-deposition technique, whereby silicon semiconductor devices, complete with metal contact systems, are fully encapsulated in glass.

glass-plate capacitor—A high-voltage capacitor in which the metal plates are separated by sheets of glass for the dielectric. The complete assembly is generally immersed in oil.

glass-to-metal seal — An airtight seal between glass and metal parts of an electron tube, made by fusing together a special glass and special metal alloy having nearly the same temperature coefficients of expansion.

glass tube—A vacuum or gaseous tube that has a glass envelope.

glide path—The approach path used by an aircraft making an instrument landing.

glide-path localizer—In an aircraft instrument-landing system, the part which indicates the altitude of the plane and creates a glide path for a blind landing.

glide slope—A radio beam used by pilots to determine their altitude when landing.

glide-slope facility — A radio transmitting facility which provides the glide-slope signals.

G-line—A round wire coated with a dielectric and used to transmit microwave energy.

glint—Also called glitter. A distorted radar-signal echo, which varies in amplitude from pulse to pulse because the beam is being reflected from a rapidly moving object such as an airplane propeller.

glissando — A tone that changes smoothly from one pitch to another.

glitch—A form of low-frequency interference appearing as a narrow horizontal bar moving vertically through the television picture. This is also observed on an oscilloscope at the field or frame rate as an extraneous voltage pip moving along the signal at approximately the reference-black level.

glitter—*See* Glint.

glossmeter—A photoelectric instrument for determining the gloss factor of a surface (i.e., the ratio of light reflected in one direction to the light reflected in all directions).

glow discharge—A discharge of electricity through a gas in an electron tube. It is characterized by a cathode glow resulting from a space potential, much higher than the ionization potential of the gas in the vicinity of the cathode.

glow-discharge microphone—A microphone in which the sound waves cause corresponding variations in the current forming a glow discharge between two electrodes.

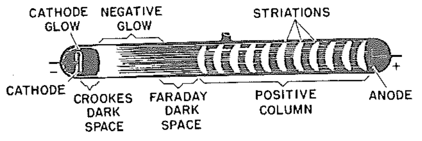

Glow-discharge tube.

glow-discharge tube—A gas tube that depends for its operation on the properties of a glow discharge.

glow-discharge voltage regulator — A gas tube used for voltage regulation. The resistance of the gas within the tube varies in step with the voltage applied across the tube.

glow lamp — A lamp containing a small amount of gas or vapor. Current between the two electrodes ionizes the gas and causes the lamp to glow but does not provide rectification. Neon gives a red-orange glow, mercury vapor a blue glow, and argon a purple glow.

glow potential—The voltage at which a glow discharge begins in a gas-filled electronic tube as the voltage is gradually increased.

glow switch—An electron tube used in some fluorescent-lamp circuits. It contains two bimetal strips which are closed when heated by the glow discharge.

glow-tube rectifier — Also called a point-plane rectifier. A cold-cathode gas-discharge tube which provides a unidirectional current flow.

glue-line heating—An arrangement of electrodes designed to heat a thin film of material having a high loss factor between alternate layers of materials having a low loss factor.

g_m—Symbol for the mutual conductance or transconductance of a vacuum tube.

G-M counter — Abbreviation for Geiger-Mueller counter.

GMT or Gmt—Abbreviation for Greenwich mean time.

gobo—A dark mat used to shield the lens of a television camera from stray lights.

gold-bonded diode — A semiconductor diode in which a preformed whisker of gold contacts an N-germanium substrate as the junction is formed by millisecond electrical pulses.

gold doping—A technique used to control the lifetime of minority carriers in a diffused-mesa transistor. Gold is diffused into the base and collector regions to reduce the storage time.

gold-leaf electroscope—An apparatus comprising two pieces of gold leaf joined at their upper ends and suspended inside a glass jar. When a charge is applied to the terminal connected to the leaves, they spread apart due to repulsion of the like charges on them.

Goldschmidt alternator — An early radio transmitter. It is a rotating machine employing oscillating circuits in connection with the field and the armature to introduce harmonics in the generated fundamental frequency. Interaction between the stator and rotor harmonics gives a cumulative effect and thereby provides very high radio frequencies.

goniometer—1. In a radio-range system, a device for electrically shifting the directional characteristics of an antenna. 2. An electrical device for determining the azimuth of a received signal by combining the outputs of individual elements of an antenna array in certain phase relationships.

googol — In mathematics, the figure 1 followed by 100 zeros.

goto circuit—A circuit capable of sensing the direction of current. It can be used in majority logic circuits in which the output is either positive or negative, depending on whether the majority of its inputs is positive or negative.

governed series motor—A motor used with teletypewriter equipment. It has a governor for regulating the speed.

governor—1. A motor attachment that automatically controls the speed at which the motor rotates. 2. The equipment which controls the gate or valve opening of a prime mover.

GPI—Abbreviation for ground-position indicator.

graded base transistor — *See* Diffused Base Transistor.

graded filter—A power-supply filter in which the output stage of a receiver or audio amplifier is connected at or near the filter input so that the maximum available d-c voltage will be obtained. The output stage has low gain; therefore, ripple is not too important.

graded-junction transistor—*See* Rate-Grown Transistor.

graded thermoelectric arm — A thermoelectric arm having a composition that changes continuously in the direction of the current.

gradient — The rate at which a variable quantity increases or decreases. For example, potential gradient is the difference of potential along a conductor or through a dielectric.

gradient meter — *See* Generating Electric Field Meter.

gradient microphone — A microphone in which the output rises and falls with the sound pressure. (*Also see* Pressure Microphone.)

gram—A unit of mass and weight in the metric system.

gramme ring—A ring-shaped iron armature around which the coils are wound. Each turn is tapped from the inside diameter of the ring to a commutator segment.

grandfather cycle—The period during which magnetic-tape records are retained before reusing so that records can be reconstructed in the event of loss of information stored on a magnetic tape.

granular carbon—Small particles of carbon used in carbon microphones.

granularity — A characteristic of the output data of a measuring instrument. The measure of granularity is the smallest increment of the output data when it is in a digital form. The smallest increment is also called least count.

graph—A pictorial presentation of the relationship between two or more variables.

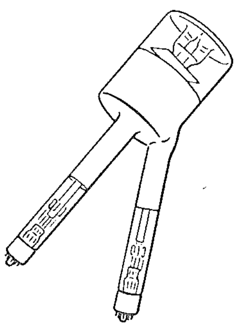

Graphechon.

graphechon—An electron tube which utilizes camera-tube principles for storing and recovering electrical signals.

graphical analysis—The use of diagrams and other graphic methods to obtain operating data and answers to scientific or mathematical problems.

graphic instrument—*See* Recording Instrument.

graphite—A finely divided carbon used as a lubricant and in the construction of some carbon resistance elements. The most common use is in so-called lead pencils.

grass — The pattern produced by random noise on an A-scope; it appears as closely spaced, sharp, constantly moving pulses on the base line. *Also see* Random Noise.

graticule—A calibrated screen placed in front of a cathode-ray tube for measurement purposes.

grating—A device for spreading out light or other radiation. It consists of narrow parallel slits in a plate or narrow parallel reflecting surfaces made by ruling grooves on polished metal. The slits or grooves break up the waves as they emerge. (*Also see* Ultrasonic Space Grating *and* Ultrasonic Cross Grating.)

grating reflector—An antenna reflector consisting of an openwork metal structure that resembles a grating.

Gratz rectifier—An arrangement of two rectifiers per phase connected into a three-phase bridge circuit to provide full-wave rectification.

gravity—The force which tends to pull bodies toward the center of the earth, thereby giving them weight. (*Also see* g.)

gravity cell—A primary cell in which two electrolytes are kept separated by differences in specific gravity. It is a modification of the Daniell cell and is now obsolete.

gray body—A radiating body whose spectral emissivity remains the same at all wavelengths. It is in constant ratio of less than unity to the ratio of a black-body radiator at the same temperature.

Gray code—A positional binary number notation in which any two numbers whose difference is one are represented by expressions that are the same except in one place or column and differ by only one unit in that place or column.

gray scale—In a television system, a scale of brightness values ranging from maximum to minimum.

great manual — In an organ, the keyboard normally used for playing the accompaniment to the melody. Also called the accompaniment manual or lower manual.

green-gain control—A variable resistor used in the matrix of a three-gun color television receiver to adjust the intensity of the green primary signal.

green gun—The electron gun the beam of which, when properly adjusted, strikes only the green phosphor dots in the color picture tube.

green restorer—A d-c restorer used in the green channel of a three-gun color-television picture-tube circuit.

green video voltage—The signal voltage that controls the grid of the green gun in a three-gun picture tube.

Greenwich mean time—Abbreviated GMT or Gmt. The mean solar time at the meridian of Greenwich (zero longitude).

grid—1. An electrode having one or more openings for the passage of electrons or ions. (*Also see* Control Grid, Screen Grid, Shield Grid, Space-Charge Grid, *and* Suppressor Grid.) 2. An interconnected system in which high-voltage, high-capacity backbone lines overlay and are connected with networks of lower voltages. 3. A two-dimensional network consisting of a set of equally spaced parallel lines superimposed upon another set of equally spaced parallel lines so that the lines of one set are perpendicular to the lines of the other, thereby forming square areas. The intersections of the lines provide the basis for an incremental location system.

grid battery—Sometimes called a C-battery. A source of energy for supplying a bias voltage to the grid of a vacuum tube.

grid bearing—A bearing made with the reference line to grid north.

grid bias—Also called C-bias. A constant potential applied between the grid and cathode of a vacuum tube to establish an operating point.

grid-bias cell—A small cell used in a vacuum tube circuit to make the grid more negative than the cathode. It provides a voltage, but cannot supply an appreciable amount of current.

grid blocking—Blocking of capacitance-coupled stages in an amplifier because of an accumulated charge of the coupling capacitor as the result of current flow during the reception of large signals.

grid cap—At the top of some vacuum tubes, the terminal which connects to the control grid.

grid capacitor—A capacitor in parallel with the grid resistor or in series with the grid lead of a tube.

grid-cathode capacitance — Capacitance between the grid and the cathode in a vacuum tube.

grid characteristic—The curve obtained by plotting grid-voltage values of a vacuum tube as abscissas against grid-current values as ordinates on a graph.

grid circuit—The circuit connected between the grid and cathode, and forming the input circuit of a vacuum tube.

grid-circuit tester — A tester designed to measure the grid resistance of vacuum tubes without discriminating between the type or polarity of impedance.

grid clip—A spring clip used for making a connection to the top-cap terminal of some vacuum tubes.

grid conductance—The in-phase component of the alternating grid current divided by the alternating grid voltage, all other electrode voltages being maintained constant.

grid-controlled mercury-arc rectifier — A mercury-arc rectifier employing one or more electrodes exclusively for controlling start of the discharge.

grid-controlled rectifier—A triode mercury-vapor rectifier tube in which the grid determines the instant at which plate current starts to flow during each cycle, but does not determine how much current will flow.

grid-control tube — A mercury-vapor–filled thermionic vacuum tube with an external grid control.

grid current—The current which flows in the grid-to-cathode circuit of a vacuum tube. It is usually a complex current made up of several currents having a variety of polarities and impedances.

grid detection—Detection by rectification in the grid circuit of a vacuum tube.

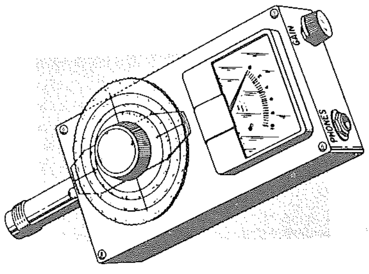

Grid-dip meter.

grid-dip meter—A multiple-range oscillator incorporating a meter in the grid circuit to indicate grid current. The meter is so named because its reading dips (reads a lower grid current) whenever an external resonant circuit is tuned to the oscillator frequency.

grid dissipation—The power lost as heat at the grid of a tube.

grid-drive characteristic—The relationship between the electrical or light output of an electron tube and the control-electrode voltage measured from cutoff.

grid driving power—The average product of the instantaneous value of grid current and the alternating component of grid voltage over a complete cycle.

grid emission—Electron or ion emission from the grid of an electron tube.

grid-glow tube — A glow-discharge, cold-cathode tube in which one or more control electrodes initiate the anode current, but do not limit it except under certain conditions.

gridistor—A field-effect transistor which uses the principle of centripetal striction and has a multichannel structure, combining advantages of both field-effect transistors and minority-carrier injection transistors.

grid leak—A high resistance connected across the grid capacitor or between the grid and cathode. It provides a direct-current path, to limit the accumulation of a charge on the grid.

grid-leak capacitor—A small capacitor connected in a vacuum-tube grid circuit, together with a resistor, to produce grid bias.

grid-leak detector—A triode or multielectrode tube in which rectification occurs because of electron current through a high resistance in the grid circuit. The voltage associated with this flow appears in amplified form in the plate circuit.

grid limiting—The use of grid-current bias derived from the signal, through a large series grid resistor, in order to cut off the plate current and consequently level the output wave for all input signals above a critical value.

grid locking—Faulty tube operation in which excessive grid emission causes the grid potential to become continuously positive.

grid modulation—Modulation produced by application of the modulating voltage to the control grid of any tube in which the carrier is present. Modulation in which the grid voltage contains externally generated pulses is called grid-pulse modulation.

grid neutralization—A method of neutralizing an amplifier. A portion of the grid-to-cathode alternating-current voltage is shifted 180° and applied to the plate-to-cathode circuit through a neutralizing capacitor.

grid north—An arbitrary reference direction used with the grid system of navigation.

grid-pool tank—A grid-pool tube having a heavy metal envelope somewhat resembling a tank in appearance.

grid-pool tube—A gas discharge tube having a mercury-pool cathode, one or more anodes, and a control electrode or grid to control the start of current. *Also see* Excitron *and* Ignitron.

grid-pulse modulation—Modulation produced in an amplifier or oscillator by application of one or more pulses to a grid circuit.

grid pulsing—Method of controlling the operation of a radio-frequency oscillator. The oscillator-tube grid is biased so negatively that no oscillation occurs, even at full plate voltage, except when this negative bias is removed by application of a positive voltage pulse to the grid.

grid resistor—A general term that denotes any resistor in the grid circuit.

grid return—An external conducting path for the return of grid current to the cathode.

grid suppressor—A resistor, sometimes connected between the control grid and the external circuit of an amplifier, to prevent parasitic oscillations caused by stray-capacitance feedback.

grid swing—The total variation in grid-to-cathode voltage from the positive peak to the negative peak of the applied-signal voltage.

grid-to-cathode capacitance—The direct capacitance between the grid and cathode of a vacuum tube.

grid-to-plate capacitance—Designated C_{gp}. The direct capacitance between the grid and plate in a vacuum tube.

grid-to-plate transconductance—The mutual conductance, or ratio of plate-current to grid-voltage changes, in a vacuum tube.

grid voltage—The voltage between the grid and cathode of a tube.

grid-voltage supply—The means for supplying, to the grid of an electron tube, a potential which is usually negative with respect to the cathode.

grommet—An insulating washer, usually of rubber or plastic, inserted through a hole in a chassis or panel to prevent a wire from touching the sides.

groove—In mechanical recording, the track inscribed in the record by a cutting or embossing stylus, including undulations or modulations caused by vibration of the stylus. In stereo discs, its cross section is a right-angled triangle, with each side at a 45° angle to the surface of the record; information is cut on both sides of the groove.

groove angle—In disc recording, the angle between the two walls of an unmodulated groove in a radial plane perpendicular to the surface of the recording medium.

groove shape—In disc recording, the contour of the groove in a radial plane perpendicular to the surface of the recording medium.

groove speed—In disc recording, the linear speed of the groove with respect to the stylus.

ground—1. A metallic connection with the earth to establish ground potential. 2. The voltage reference point in a circuit. There may or may not be an actual connection to earth, but it is understood that a point in the circuit said to be at ground potential could be connected to earth without disturbing the operation of the circuit in any way.

ground absorption—The loss of energy during transmission because of the radio waves dissipated to ground.

ground bus—A conductor, usually large-diameter wire, that connects a number of points to one or more grounding electrodes.

ground check (base-line check)—1. A procedure followed prior to the release of a radiosonde in order to obtain the temperature and humidity correction for the radiosonde system. 2. Any instrumental check prior to the ground launch of an airborne experiment.

ground clamp—A clamp used for connecting a grounding conductor (ground wire) to a grounded object such as a water pipe.

ground clutter—The pattern produced on the screen of a radar indicator by undesired ground return.

ground conduit—A conduit used solely to contain one or more grounding conductors.

ground-controlled approach — Abbreviated GCA. A ground radar system providing information with which aircraft approaches can be directed by radio.

ground-controlled interception — Abbreviated GCI. A radar system used for directing an aircraft to intercept enemy aircraft.

ground current—*See* Earth Current.

ground dielectric constant—The dielectric constant of the earth at a given location.

ground distance—The great-circle component of distance from one point to another at mean sea level.

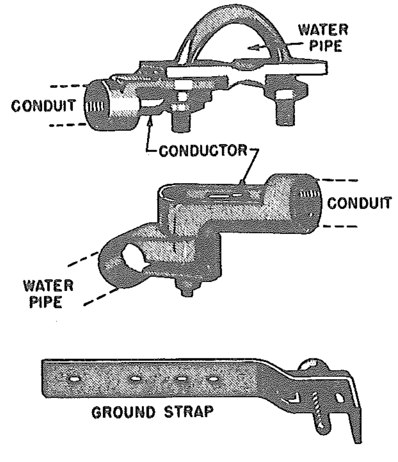

Ground clamps.

grounded—Connected to the earth, or to some conducting body in place of the earth.

grounded-base amplifier—*See* Common-Base Amplifier.

grounded cable bond—A cable bond used for grounding the armor and/or sheaths of cables.

grounded capacitance—In a system having several conductors, the capacitance between a given conductor and the other conductors when they are connected together and to ground.

grounded-cathode amplifier—The conventional amplifier circuit. It consists of a tube amplifier in which the cathode is at ground potential at the operating frequency. The input is applied between the control grid and ground, and the output load is between the plate and ground.

grounded circuit—A circuit in which one conductor or point (usually the neutral conductor or neutral point of transformer or generator windings) is intentionally grounded (earthed) either solidly or through a grounding device.

grounded-collector amplifier—*See* Common-Collector Amplifier.

grounded conductor—A conductor which is intentionally grounded, either directly or through a current-limiting device.

grounded-emitter amplifier—*See* Common-Emitter Amplifier.

grounded-grid amplifier—An electron-tube amplifier circuit in which the control grid is at ground potential at the operating frequency. The input is applied between the cathode and ground, and the output load is between the plate and ground. The grid-to-plate impedance of the tube is in parallel with the load, instead of acting as a feedback path.

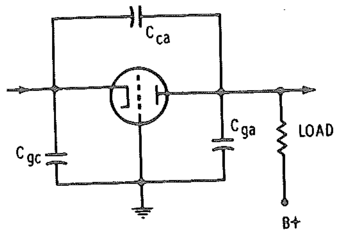

Grounded-grid amplifier.

grounded-grid triode—A type of triode designed for use in a grounded-grid circuit.

grounded-grid triode circuit—A circuit in which the input signal is applied to the cathode and the output is taken from the plate. The grid is at r-f ground and serves as a screen between the input and output circuits.

grounded-grid triode mixer—A triode in which the grid forms part of a grounded electrostatic screen between the anode and cathode. It is used as a mixer for centimeter wavelengths.

grounded outlet—An outlet equipped with a receptacle of the polarity type having, in addition to the current-carrying contacts, one ground contact which can be used for the connection of an equipment-grounding conductor.

grounded parts—Parts of a completed installation that are so connected that they are substantially at the same potential as the earth.

grounded-plate amplifier—Also called cathode follower. An electron-tube amplifier circuit in which the plate is at ground potential at the operating frequency. The input is applied between the control grid and ground, and the output load is between the cathode and ground.

grounded system—A system of conductors in which at least one conductor or point (usually the middle wire or neutral point of transformer or generator windings) is intentionally grounded, either solidly or through a current-limiting device.

ground gating—The conversion of pam signals at a telemetry ground station to 50% duty-cycle signals.

ground indication—An indication of the presence of a ground on one or more of the normally ungrounded conductors of a system.

grounding—1. Connecting to ground, or to a conductor which is grounded. 2. A means of referencing electrical circuits to the well-bonded equipotential surface.

grounding conductor—A conductor which, under normal conditions, carries no current, but serves to connect exposed metal surfaces to an earth ground, to prevent hazards in case of breakdown between current-carrying parts and the exposed surfaces. The conductor, if insulated, is colored green with or without a yellow stripe.

grounding connection—A connection used in establishing a ground.

grounding electrode—A conductor embedded in the earth and used for maintaining ground potential on conductors connected to it, or for dissipating into the earth any current conducted to it.

grounding plate — An electrically grounded metal plate on which a person stands in order to discharge any static electricity that may be picked up by his body.

grounding switch—A form of air switch for connecting a circuit or apparatus to ground.

grounding transformer—A transformer intended primarily for the purpose of providing a neutral point for grounding purposes.

ground insulation — The major insulation used between a winding and structural parts at ground potential.

ground loop—The generation of undesirable signals within a ground conductor due to circulation current within the conductor which originates from a second source of potential—frequently as a result of connecting two separate grounds to a signal circuit.

ground lug—A lug for connecting a grounding conductor to a grounding electrode.

ground noise—In recording and reproducing, the residual noise in the absence of a signal. It is usually caused by dissimilarities between the recording and reproducing media, but may also include amplifier noise such as from a tube or noise generated in resistive elements at the input of the reproducer amplifier system.

ground-noise margin—The voltage that may be applied at the ground connection of a logic circuit without causing the circuit to malfunction. It is usually measured by increasing the static ground voltage on a single gate until the logic fails to operate properly.

ground plane—Copper or brass sheet used in interference testing to simulate missile, aircraft, or vehicle frame or skin so that actual installation and grounding conditions may be approximated.

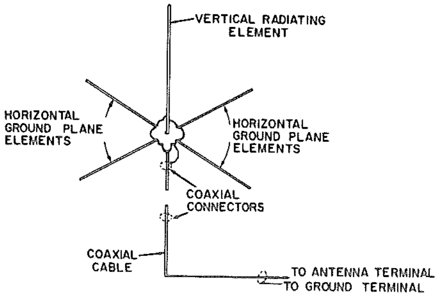
VERTICAL RADIATING ELEMENT
HORIZONTAL GROUND PLANE ELEMENTS
HORIZONTAL GROUND PLANE ELEMENTS
COAXIAL CONNECTORS
COAXIAL CABLE
TO ANTENNA TERMINAL
TO GROUND TERMINAL

Ground-plane antenna.

ground-plane antenna—A vertical antenna combined with a turnstile element to lower the angle of radiation. It has a concentric base support and a center conductor that place the antenna at ground potential, even though located several wavelengths above ground.

ground plate—A plate of conductive material buried in the earth to serve as a grounding electrode.

ground-position indicator — Abbreviated GPI. A dead-reckoning computer, similar to an air-position indicator, with provision for taking drift into account.

ground potential—Zero potential with respect to ground or the earth.

ground protective relay — A device which functions on failure of the insulation of a machine, transformer, or other apparatus to ground, or on flashover of a d-c machine to ground.

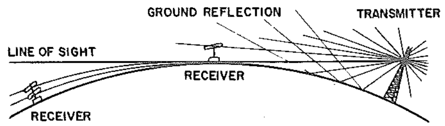

Ground reflection.

ground-reflected wave—In a ground wave, the component reflected from the earth.

ground resistance — The opposition of the earth to the flow of current through it. Its value depends on the nature and moisture content of the soil. The material, composition, and physical dimensions of the connections to earth, and the electrolytic action present.

ground return—1. In radar, the echoes reflected from the earth's surface and fixed objects on it. 2. A lead from an electronic circuit, antenna, or power line to ground.

ground-return circuit—A circuit which has a conductor (or two or more in parallel) between two points and which is completed through ground or the earth.

ground rod—A steel or copper rod driven into the earth to make an electrical contact with it.

ground shift—The variation in signal amplitude at different grounding points because of a voltage drop along a ground line.

ground speed—In navigation, the speed of a vehicle with reference to ground.

ground-support equipment — All ground equipment that is part of a complete weapons system and that must be furnished to insure complete support of the weapon system.

ground-to-air communication — One-way communication from ground stations to aircraft.

ground wave—A radio wave which travels along the earth's surface rather than through the upper atmosphere.

ground wire—A conductor leading to an electric connection with the earth.

group frequency—The number of sets or groups of waves passing a given point in one second.

grouping—Nonuniform spacing between the grooves of a disc recording.

group loop—A source of interference when a system is grounded improperly at several points.

group modulation—The process by which a number of channels, already separately modulated to a specific frequency range, are again modulated to shift the group to another range.

group velocity—1. Of a traveling plane wave, the velocity of propagation of the envelope delay is approximately constant. It is equal to the reciprocal of the envelope delay per unit length. (Group velocity differs from phase velocity in a medium in which the phase velocity varies with the frequency.) 2. The velocity of propagation of a pulse or a group of waves.

Grove cell—A primary cell with a platinum electrode submerged in an electrolyte of nitric acid within a porous cup, surrounded by a zinc electrode in an electrolyte of sulphuric acid. This cell normally operates on a closed circuit.

growler—An electromagnetic device for locating short-circuited coils and for magnetizing or demagnetizing objects. So called because of the growling noise it makes when indicating a short circuit. It consists essentially of two field poles arranged as in a motor.

grown-diffused transistor — A transistor made by combining the diffusion and double-doped techniques. Suitable N- and P-type impurities are added simultaneously to the melt while the crystal is being grown. Subsequently, the base region is formed by diffusion as the crystal grows.

grown junction—The boundary between P- and N-type semiconducting materials. It is produced by varying the impurities during the growth of a crystal from the melt. Such junctions have strong rectifying properties, the forward current being obtained when P is positive to N.

grown-junction photocell — A photodiode made of a small bar of semiconductor material that has a PN junction at right angles to its length and an ohmic contact at each end of the bar.

grown-junction transistor—A transistor in which junctions are formed by adding impurities to the melt while the crystal is being grown.

G-scan—*See* G-Display.

G-scope—*See* G-Display.

guard band—1. Also called interference-guard band. A frequency band left vacant between two channels to safeguard against mutual interference. 2. The unused chip surface area which by virtue of physical spacing serves to isolate functional elements in a printed circuit or an integrated circuit. Also refers to the consideration given instrumentation precision in electrical testing.

guard circle—An inner concentric groove on disc records. It prevents the pickup from being thrown to the center of the record and possibly damaged.

guarded input—An input that has a third terminal which is maintained at a potential near the input-terminal potential for a single-ended input, or near the mean input potential for the differential input. It is used to shield the entire input circuit.

guarded motor — An open motor in which all openings given direct access to live or rotating parts (except smooth shafts) are limited in size by the structural parts or by screens, grilles, expanded metal, etc., to prevent accidental contact with such parts. Such openings shall not permit the passage of a cylindrical rod ½ inch in diameter.

guard ring—A metal ring placed around a charged terminal or object to distribute the charge uniformly over the surface of the object.

guard shield—A shield surrounding the input circuit of an amplifier.

guard wire—A grounded wire used frequently where high-tension lines cross a thoroughfare. Should a line break, it will contact the guard wire and be grounded.

Gudden-Pohl effect—The momentary illumination produced when an electric field is applied to a phosphor previously excited by ultraviolet radiation.

guidance—Control of a missile or vehicle from within by a person, a preset or self-reacting automatic device, or a device that reacts to outside signals.

guidance system—A system which measures and evaluates flight information, correlates it with target data, and converts the resultant into the parameters necessary to achieve the desired flight path.

guided ballistic missile—A ballistic missile which is guided during the powered portion of the trajectory and follows a free ballistic path during the remainder.

guided missile—An unmanned vehicle moving above the surface of the earth, the trajectory or flight path of which is capable of being altered by an external or internal mechanism.

guided propagation — A type of radio-wave propagation in which radiated rays are bent excessively by refraction in the lower layers of the atmosphere. This bending creates an effect much as if a duct or a wave guide has been formed in the atmosphere to guide part of the radiated energy over distances far beyond the normal range.

guided wave—A wave in which the energy is concentrated near a boundary (or between substantially parallel boundaries) separating materials of different properties. The direction of propagation is parallel to the boundary.

guide wavelength—*See* Waveguide Wavelength.

Guillemin line—A special type of artificial transmission line or pulse-forming network used in radar sets to control the duration of the pulses. It generates a nearly square pulse for use in high-level pulse modulation.

guillotine capacitor—A translatory-motion tuning capacitor consisting of a pair of stators and a sliding plunger in place of a rotor.

gun-directing radar—Radar used for directing antiaircraft or similar artillery fire.

Gunn effect—Current oscillations that occur at an r-f rate when an electric field of about 3000 v/cm is applied to a short (0.005 inch or less) specimen of N-type gallium arsenide. This effect takes place because electrons under the influence of sufficiently high fields are transferred from high- to low-mobility valleys in the conduction band of GaAs.

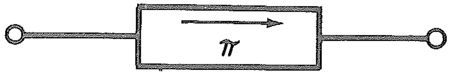

Gyrator symbol.

gutta-percha—A natural vegetable gum, similar to rubber, used principally as insulation for wires and cables.

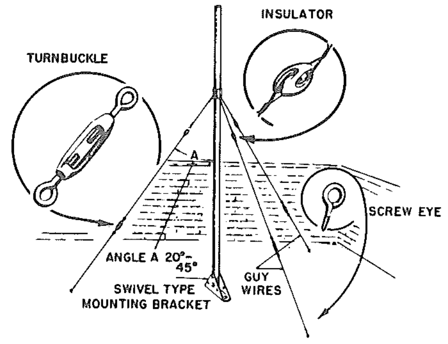

Guy wire.

guy wire—A wire used to brace the mast or tower of a transmitting or receiving antenna system.

gyrator—A two-port circuit element that exhibits a 180° differential, or nonreciprocal, phase shift. The gyrator circuit symbol indicates that an r-f signal transmitted from port 1 to port 2 will undergo a 180° phase shift (arrow direction) relative to an r-f signal transmitted in the reverse direction.

gyro—Abbreviation for gyroscope.

gyrofrequency—The natural frequency at which charged particles rotate around the lines of force of the earth's magnetic field. For electrons, it is 700 to 1,600 kilohertz; for ions, it is in the audio-frequency range.

gyromagnetic—The magnetic properties of rotating electric charges, such as electrons spinning within atoms.

gyropilot—*See* Autopilot.

gyroscope—A rotating device the axle of which will maintain a constant direction, even though the earth is turning under it. It consists of a wheel mounted so that its spinning axis is free to rotate around either of two other axes perpendicular to itself and to each other. When its axle is pointed north, it can be used as a gyrocompass. Abbreviated gyro.

gyroscopic action—An action that causes a mass to turn on an axis perpendicular to the applied torque and to the axis of spin.

gyrostabilized platform—*See* Stable Platform.

G−Y signal—In color television, the green-minus-luminance signal representing primary green minus the luminance, or Y− signal. It is combined with a luminance, or Y− signal outside or inside the picture tube to yield a primary green signal.

H

h—Abbreviation for henry.

H—1. A radar air-navigation system using an airborne interrogator to measure the distance from two ground responder beacons. (*Also see* Shoran.) 2. Symbol for heater or magnetic field strength.

hairpin pickup coil—A hairpin-shaped, single-turn coil for transferring uhf energy.

hairpin tuning bar—A sliding hairpin-shaped metal bar inserted between the two halves of a doublet antenna to vary its electrical length.

halation—Distortion seen as blurred images and caused by reflection of the image rays off the back of a fluorescent screen that is too thick.

half adder—A circuit that will accept two binary input signals and produce corresponding sum and carry outputs. So called because, above the first order, two half adders per order are required when adding two quantities.

half cell—An electrode, submerged in an electrolyte, for measuring single electrode potentials.

half cycle—The time interval required for the operating frequency to complete one half, or 180°, of its cycle.

half-duplex circuit—A circuit which permits one-direction electrical communications between stations. Technical arrangements may permit operation in either direction, but not simultaneously. Therefore, this term is qualified by one of the following suffixes: S/O for send only; R/O for receive only; S/R for send or receive. Abbreviated hdx.

half-duplex operation—A duplex telegraph system capable of operating in either direction, but not in both simultaneously.

half-duplex repeater—A duplex telegraph repeater provided with interlocking arrangements which restrict the transmission of signals to one direction at a time.

half life—The time interval used to measure the rate of decay of radioactive material. In the first half-life, the amount of radioactive material left unchanged is one-half the original amount; in the next half-life interval, half of the remaining amount, or one-fourth of the original amount remains. Thus, by determining the remaining radioactivity of a fossil and comparing it with the half-life of the material, scientists can estimate the age of the fossil. The half-life of various materials varies greatly—from millionths of a second to billions of years.

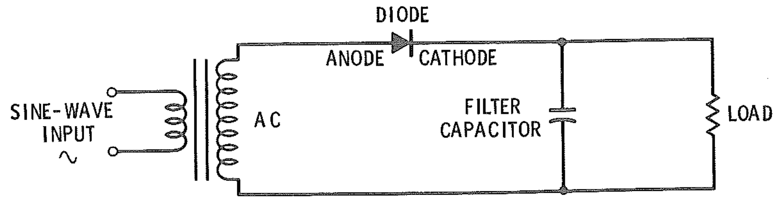

Half-wave rectifier.

half-nut—A feed nut which engages half the circumference or less of a lead screw, so that it can be withdrawn from the lead screw to stop the lateral scanning movement.

half-power point—On an amplitude response characteristic or other curve of the magnitude of a network quantity versus frequency, distance, angle, or other variable, the point that corresponds to half the power of a neighboring point having maximum power.

half-power width of a radiation lobe—In a plane containing the direction of the maximum of the lobe, the full angle between the two directions in that plane in which the radiation intensity is one half the maximum value of the lobe.

half-shift register—A logic circuit that consists of a gated input storage element with or without an inverter.

half step—*See* Semitone.

half tap—A bridge that can be placed across conductors without disturbing their continuity.

half-time emitter—A device that produces synchronous pulses midway between the row pulses of a punched card.

half-tone characteristic—In facsimile, the fidelity of the recorded density shadings in comparison with the original transmitted subject copy. Also used to express the relationship between the facsimile signal and the subject or recorded copy.

half-track recorder—*See* Dual-Track Recorder.

half-track tape—Also called two-track tape. Quarter-inch magnetic tape on which half the width of the tape is used for one sound path. Such a tape provides stereo in one direction of tape travel, or monosound in both directions.

half wave—A wave with an electrical length of half a wavelength.

half-wave antenna—An antenna having an electrical length equal to half the wavelength of the signal being transmitted or received.

half-wave dipole—A straight, ungrounded antenna measuring substantially one-half wavelength.

half-wave rectification—The production of a pulsating direct current by passing only half the input cycle of an alternating current. The other half is blocked by the rectifier.

half-wave rectifier—A rectifier utilizing only one half of each cycle to change alternating current into pulsating direct current.

half-wave transmission line—A piece of transmission line having an electrical length equal to half the wavelength of the signal being transmitted or received.

half-wave vibrator—A vibrator used mainly in battery-operated mobile power supplies. It has only one pair of contacts, and supplies an intermittent unidirectional current at its output (usually connected to a half-wave rectifier).

Hall constant—The constant of proportionality in the equation for a current-carrying conductor in a magnetic field. The constant is equal to the transverse electric field (Hall field) divided by the product of the current density and the magnetic field strength. The sign of the majority carrier can be inferred from the sign of the Hall constant.

Hall effect—When a metal strip is placed with its plane perpendicular to a magnetic field and an electric current flows longitudinally through the strip, a potential difference is developed across the strip at right angles to the current flow and to the magnetic field.

Hall generator—A thin wafer of semiconductor material used for measuring a-c power and magnetic field strength. Its output volt-

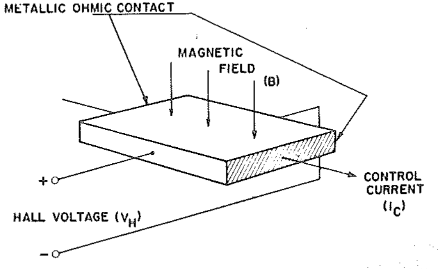

Hall generator.

age is proportional to the current passing through it times the magnetic field perpendicular to it.

Hall mobility—The product of conductivity and the Hall constant for a conductor or semiconductor. It is a measure of the mobility of the electrons or holes in a semiconductor.

halo—The undesirable ring of light around a spot on the fluorescent screen of a cathode-ray tube.

halogen — A general name applied to four chemical elements, flourine, chlorine, bromine, and iodine, that have similar chemical properties.

halogen quenching—A method of quenching the discharge in a counter tube by the introduction of a small quantity of one of the halogens.

ham — Also called amateur. Slang for a licensed radio operator who operates a station as a hobby rather than a business.

Hamming code—One of the error correction code systems used in data transmission.

hand capacitance—The capacitance introduced when one's hand is brought near a tuning capacitor or other insufficiently shielded part of a receiver.

Handie-Talkie—Trade name of the Motorola Communications Div. for a two-way radio small enough to be carried in one's hand.

hand receiver—An earphone held to the ear by hand.

hand reset—A relay in which the contacts must be reset manually to their original positions after normal conditions are resumed.

handset—A telephone-type receiver and transmitter mounted on a single frame.

handset telephone—*See* Hand Telephone Set.

hand telephone set—Also called a handset telephone. A telephone set having a handset and a mounting which supports the handset when not in use.

hangover—Also called tailing. The smeared or blurred bass notes reproduced by a poorly damped speaker or one mounted in an improperly vented enclosure.

hangup — A condition in which the central processor of a computer is trying to perform an illegal or forbidden operation or in which it is continually repeating the same routine.

hard copy—Typewritten or printed characters on paper produced by a computer at the same time information is copied or converted into machine language that is not easily read by a human.

hard-drawn copper wire—Copper wire that is not annealed after work hardening during drawing, thus providing increased tensile strength.

hard firing—A condition in which the gate signal of an scr is several times the d-c triggering current and in which the rise time of the gate current is short relative to the turn-on time.

hardline—The intelligence link between two objects, consisting of a wire or wires, as opposed to a radio or radar link.

hard magnetic materials—Magnetic materials that are not easily demagnetized.

hardness—Referring to X rays, the quality which determines their penetrating ability. The shorter the wavelength, the harder and hence more penetrating they are.

hardness tester—Equipment for determining the force required to penetrate the surface of a solid.

hard rubber — A material formerly widely used for insulation. It is formed by vulcanizing rubber at high temperature and pressure to give it the desired hardness.

hard solder—Solder composed principally of copper and zinc. It must be red-hot before it will melt. Hard soldering is practically equivalent to brazing.

hard tube—A high-vacuum electronic tube.

hardware—1. Slang for the individual components in a circuit. 2. Particular circuits or functions built into a system.

hard X-rays—Highly penetrating X-rays as distinguished from less penetrating, or soft, X-rays.

harmful interference—Any radiation or any induction which disrupts the proper functioning of an electromagnetic system.

harmonic—A sinusoidal wave having a frequency that is an integral multiple of the fundamental frequency. For example, a wave with twice the frequency of the fundamental is called the second harmonic.

harmonic analysis—1. A method of identifying and evaluating the harmonics that make up a complex waveform of voltage, current, or some other varying quantity. 2. The expression of a given function as a series of sine and cosine terms that are approximately equal to the given function, such as a Fourier series.

harmonic analyzer—Also called harmonic-wave analyzer. A mechanical or electronic device for measuring the amplitude and phase of the various harmonic components of a wave from its graph.

harmonic antenna—An antenna the electrical length of which is an integral multiple of a half wavelength.

harmonic attenuation—Elimination of a harmonic frequency by using a pi network and tuning its shunt resistances to zero for the frequency to be eliminated.

harmonic component—Of a periodic quantity, any one of the simple sinusoidal quantities of the Fourier series into which the periodic quantity may be resolved.

harmonic content—1. The degree of distortion in the output signal of an amplifier. 2. The components remaining after the fundamental frequency has been removed from a complex wave.

harmonic conversion transducer—A conversion transducer in which the useful output frequency is a multiple or submultiple of the input frequency.

harmonic detector—A voltmeter circuit that measures only a particular harmonic of the fundamental frequency.

harmonic distortion—1. The production of harmonic frequencies at the output by the nonlinearity of a transducer when a sinusoidal voltage is applied to the input. The amplitude of the distortion is usually a function of the amplitude of the input signal. 2. A condition that exists in the output of an amplifying circuit when harmonics, added during the process, alter the signal waveform.

harmonic filter—A combination of inductance and capacitance tuned to an undesired harmonic to suppress it.

harmonic generator—A vacuum tube or other generator operated so that it generates strong harmonics in the output.

harmonic interference—Interference between radio stations because harmonics of the carrier frequency are present in the output of one or more stations.

harmonic-leakage power (TR and Pre-TR tubes) — The total radio-frequency power transmitted, through the fired tube in its mount, at other than the fundamental frequencies generated by the transmitter.

harmonic motion—Back and forth motion, such as that of a pendulum, in which the distance on one side of equilibrium always equals the distance on the other side; the acceleration is toward the point of equilibrium and directly proportional to the distance from it. Graphically, harmonic motion is represented by a sine wave.

harmonic producer — A tuning-fork-controlled oscillator used to provide carrier frequencies for broad-band carrier systems. It is capable of producing odd and even harmonics of the fundamental tuning-fork frequency.

harmonic ringing—A system of selectively signaling several parties on a subscriber's line. The different rings are produced by currents which are harmonics of several fundamental frequencies.

harmonic series of sounds — A series in which each basic frequency in it is an integral multiple of a fundamental frequency.

harmonic telephone ringer—A ringer which responds only to alternating current within a very narrow frequency band. A number of such ringers, each responding to a different frequency, are used in one type of selective ringing.

harmonic-wave analyzer — *See* Harmonic Analyzer.

harness—Wires and cables arranged and tied together so they can be connected or disconnected as a unit.

hartley—In computers, a unit of information content equal to one decimal decision, or the designation of one of ten possible and equally likely values or states of anything used to store or convey information. One hartley equals $\log_2 10$ (3.323) bits.

Hartley oscillator—An oscillator in which a parallel-tuned tank circuit is connected between the grid and plate of an electron tube or between the base and collector of a junction transistor, the inductive element of the tank having an intermediate tap at the cathode or emitter potential.

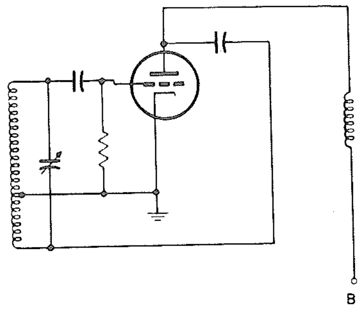

Hartley oscillator.

hash—1. Electrical noise generated within a receiver by a vibrator or a mercury-vapor rectifier. (*Also see* Grass, Garbage.) 2. A completely random interfering signal usually caused by arcing and occasionally by natural environmental disturbances.

hash total—In a computer, a total for checking purposes. It is determined by adding all the digits or all the numbers in a particular field in a batch of unit records, with no attention paid to the meaning or significance of the total.

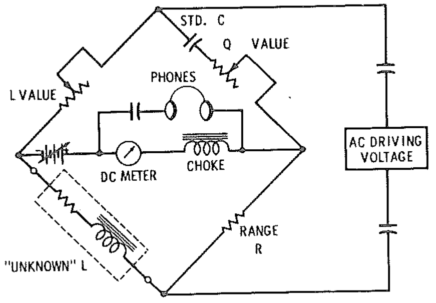

Hay bridge.

Hay bridge—A four-arm, alternating-current bridge used for measuring inductance in terms of capacitance, resistance, and frequency. The arms adjacent to the unknown impedance are nonreactive resistors, and the opposite arm is comprised of a capacitor in series with a resistor (unlike the Maxwell bridge, where it is in parallel). Usually the bridge is balanced by adjustment of the resistor, which is also in series with the capacitor and one of the nonreactive arms. The balance depends upon the frequency.

hazard—*See* Failure Rate.

Hazeltine neutralizing circuit — An early form of neutralized radio-frequency amplifier circuit.

H-bend—Also called H-plane bend. In waveguide technique, a smooth change in the direction of the axis of the waveguide. Throughout the change, the axis remains perpendicular to the direction of polarization.

H-display—Also called H-scan. In radar, a B-display modified to indicate the angle of elevation. The target appears as two closely spaced blips which approximate a short, bright line that slopes in proportion to the sine of the angle of target elevation.

head—A device which reads, records, or erases information in a storage medium; e.g., a small electromagnet used to read, write, or erase information on a magnetic drum or tape.

head alignment — Positioning the record-playback head on a tape recorder so that its gap is perpendicular to the tape.

head amplifier—An audio-frequency amplifier mounted on or near the sound head of a motion-picture projector to amplify the extremely weak output of the phototube.

head demagnetizer—Device for eliminating any magnetism built up in a recording head.

header—1. The part of a sealed component or assembly that provides support and insulation for the leads passing through the walls. 2. A preformed part made of an insulating material and containing the input and output leads of a module or component. May be part of, or attached to, the case by an adhesive bond or embedment material. Forms a portion of the other case of the module or component.

head guy—A messenger cable and attachments placed so they pull toward the pole line.

heading—The direction of a ship, aircraft, or other object with reference to true, magnetic, compass, or grid north.

headlight—An aircraft radar antenna small enough to be housed in the wing, like an automobile headlight. The beam operates like a searchlight.

headphone—Also called a head receiver or phone. A device held against the ear and having a diaphragm which vibrates according to current variations. It reproduces the incoming electrical signals as sound. Thus, the headphone permits private listening to a receiver, amplifier, or other device.

head receiver—*See* Headphone.

headset—A headphone (or a pair of headphones) and its associated headband and connecting cord.

head stack—A group of two or more heads mounted in a single unit, used to provide multiple-track recording or reproduction.

head-to-tape contact—The degree to which the surface of the magnetic coating approaches the surface of the record or replay heads during normal operation of a recorder. Good head-to-tape contact minimizes separation loss and is essential in obtaining high resolution.

hearing aid—A small audio reproducing system for the hard of hearing. It consists of a microphone, amplifier, battery, and earphone and is used to increase the sound level normally received by the ear.

hearing loss—Also called deafness. The hearing loss of an ear at a specified frequency—i.e., the ratio, expressed in decibels, of its threshold of audibility to the normal threshold.

hearing loss for speech—The difference in decibels between the speech levels at which the average normal ear and the defective ear, respectively, reach the same intelligibility. It is often arbitrarily set at 50%.

heat aging—A test used to indicate the relative resistance of various insulating materials to heat degradation.

heat coil—A protective device that grounds or opens a circuit, or both, when the current rises above a predetermined value. A mechanical element moves when the fusible substance that holds it in place is heated above a certain point by current through the circuit.

heater—1. Also called filament. An element that supplies the heat to an indirectly heated cathode. 2. A resistor that converts electrical energy into heat.

heater biasing—Application of a d-c potential to the heater to eliminate diode conduction between it and some other element within the tube.

heater current—The current flowing through a heater in a vacuum tube.

heater voltage—The voltage between the terminals of a heater.

heater-voltage coefficient—In a klystron, the frequency change per volt of heater voltage change when the reflector voltage is adjusted for the peak of a reflector voltage mode.

heat-eye tube—A cathode-ray tube powered by a midget generator. It is used as an infrared instrument that can "see" in the dark.

heat gradient—The difference in temperature between two parts of the same object.

heating effect of a current—Assuming a constant resistance, the amount of heat produced by the current through it. It is proportionate to the square of the current.

heating element—The wirewound resistor, terminals, and insulating supports used in electric cooking and heating devices.

heating pattern—In induction or dielectric heating, the distribution of temperature in a load or charge.

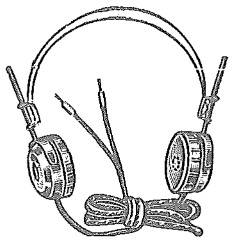

Headset.

heating station—In induction or dielectric heating, the work coil or applicator and its associated production equipment.

heat loss—The loss due to conversion of part of the electric energy into heat.

heat of emission—Additional heat energy that must be supplied to an electron-emitting surface to keep its temperature constant.

heat of radioactivity—Heat generated by radioactive disintegration.

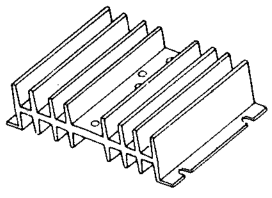

Heat sink.

heat sink — Any device that absorbs and draws off heat from a hot object, reradiating it into the surrounding atmosphere.

heat waves—Infrared radiation similar to radiowaves but of a higher frequency.

heat writer—A direct writer which records on thermally sensitive paper with a heated stylus.

Heaviside - Campbell mutual - inductance bridge — A Heaviside mutual - inductance bridge in which one inductive arm contains a separate inductor that is included in the bridge arm during the first of a pair of measurements and is short-circuited during the second. The balance is independent of frequency. (*Also see* Heaviside Mutual-Inductance Bridge.)

Heaviside layer—Also called the Kennelly-Heaviside layer. The region of the ionosphere that reflects radio waves back to earth.

Heaviside mutual-inductance bridge—An alternating-current bridge normally used for the comparison of self- and mutual-inductances. Each of the two adjacent arms contains self-inductance, and one or both of them have mutual inductance to the supply circuit. The other two arms normally are nonreactive resistors. The balance is independent of frequency.

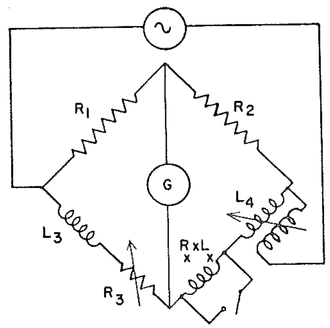

Heaviside-Campbell mutual-inductance bridge.

heavy hydrogen—Another term for deuterium (H^2) or tritium (H^3).

hecto—A prefix meaning 100.

hectometric wave—An electromagnetic wave between the wavelength limits of 100 and 1000 meters corresponding to the frequency range 3 MHz to 300 MHz.

heelpiece—Part of a relay magnetic structure at the end of a coil, opposite the armature. It generally supports the armature and completes the magnetic path between it and the core of the coil.

Hefner lamp—A standard source that gives a luminous intensity of 0.9 candlepower.

height control—In a television receiver, the adjustment which determines the amplitude of the vertical-scanning pulses and hence the height of the picture.

height finder—A radar which measures the altitude of an airborne object.

height-position indicator—A radar display which simultaneously shows the angular-elevation slant range and height of objects.

height-range indicator—A cathode-ray tube from which altitude and range measurements of airborne objects may be viewed.

Heising modulation—*See* Constant-Current Modulation.

helical—Spiral-shaped.

helical antenna—Also called a helical-beam antenna. A spiral conductor wound around a circular or polygonal cross section. The axis of the spiral normally is mounted parallel to the ground and fed at the adjacent end. The radiation produced has approximately a circular polarization and is confined mainly to a single lobe located along the axis of the spiral.

helical-beam antenna—*See* Helical Antenna.

helical potentiometer—A precision potentiometer which requires several turns of the control knob to move the contact arm from one end of the spiral-wound resistance element to the other end.

helical scanning — 1. Radar scanning in which the r-f beam describes a distorted spiral motion. The antenna rotates about the vertical axis while the elevation angle rises slowly from zero to 90°. 2. Method of facsimile scanning in which the elemental area sweeps across the copy in a spiral motion.

helionics—The conversion of solar heat to electric energy.

helix—A spiral.

helix recorder—A recorder in which helical scanning is used.

Helmholtz coil — A phase - shifting network used for determining the range in certain types of radar equipment. It consists of fixed and movable coils. The phase is kept constant at the input, but may be continually shifted from 0° to 360° at the output.

Helmholtz resonator—An acoustic enclosure with a small opening which causes the enclosure to resonate. The frequency at which it does depends on the geometry of the resonator.

hemimorphic—Terminated at the two ends by dissimilar sets of faces.

HEM wave—*See* Hybrid Electromagnetic Wave.

henry—Abbreviated h. The measure of inductance. In a closed circuit one henry of inductance is present, when a current variation of 1 ampere per second induces 1 volt.

heptode—A seven-electrode electron tube containing an anode, a cathode, a control electrode, and four grids.

hermaphroditic connector—A connector in which both mating contacts are exactly alike at their mating face.

hermetically sealed relay—A relay in a gas-tight enclosure which has been completely sealed by fusion or comparable means to ensure a low rate of gas leakage over a long period of time.

hermetic sealing—The evacuating and sealing of an enclosure to make it airtight.

herringbone pattern—Television interference seen as one or more horizontal bands of closely spaced V- or S-shaped lines.

hertz—Abbreviated Hz. A unit of frequency equal to one cycle per second.

Hertz antenna—An antenna system which does not depend for its operation on the presence of ground. Its resonant frequency is determined by its distributed capacitance, which varies according to its physical length.

Hertz effect—The ionization and spark discharge produced by ultraviolet radiation.

Hertzian oscillator—A type of oscillator for producing ultrahigh-frequency oscillations. It consists of two metal plates or other conductors separated by an air gap. The capacitor formed has such a small capacitance that ultrahigh-frequency oscillations can occur.

Hertzian waves—Electromagnetic waves of frequencies between 10 kHz and 30,000 GHz. A radio wave.

Hertz vector—A vector which specifies the electromagnetic field of a radio wave. Both the electric and the magnetic intensities can be specified in terms of it.

heterodyne—Also called beat. To mix two frequencies together in order to produce two other frequencies equal to the sum and difference of the first two. For example, heterodyning a 100-kHz and a 10-kHz signal will produce a 110-kHz (sum frequency) and a 90-kHz (difference frequency) signal.

heterodyne conversion transducer (converter)—A conversion transducer in which the output frequency is the sum or difference of the input frequency and an integral multiple of the frequency of another wave.

heterodyne detection—Detection (or conversion) by mixing two signals together to generate the intermediate frequency in a superheterodyne receiver or to make c-w signals audible.

heterodyne detector—A detector that converts an incoming r-f signal to an audible tone by heterodyning. It incorporates a local oscillator (called a beat-frequency oscillator).

heterodyne frequency—The sum or difference frequency produced by combining two other frequencies.

heterodyne frequency meter—*See* Heterodyne Wavemeter.

heterodyne oscillator—An oscillator which produces a desired frequency by combining two other frequencies (e.g., two radio frequencies to produce an audio frequency, or the incoming and local-oscillator frequencies to produce the intermediate frequency of a superheterodyne receiver).

heterodyne principle—*See* Heterodyne.

heterodyne reception—Also called beat reception. Reception by combining a received high-frequency wave with a locally generated wave in a nonlinear device to produce sum and difference frequencies at the output.

heterodyne-type frequency meter—An instrument for measuring frequency by producing a zero difference frequency (zero beat) between the signal under test and an internally generated signal.

heterodyne wavemeter—A wavemeter employing the heterodyne principle to compare the frequency being measured with a frequency being generated in a calibrated oscillator circuit.

heterodyne whistle—A steady squeal heard in a radio receiver when the signals from stations having nearly equal frequencies beat together.

heterodyning—*See* Heterodyne.

heterogeneity—A state or condition being unlike in nature, kind, or degree.

heterogeneous—Composed of different materials (opposite of homogeneous).

heterosphere—The portion of the upper atmosphere in which the relative proportions of oxygen, nitrogen, and other gases are unfixed and radiation particles and micrometeroids are mixed with the air particles.

heuristic program—A set of computer instructions that simulate the behavior of human operators in approaching similar problems.

hexadecimal number system—A number system having as its base the equivalent of the decimal number sixteen.

hexode—A six-electrode electron tube containing an anode, a cathode, a control electrode, and three grids.

hf—Abbreviation for high frequency.

HH beacon—A nondirectional radio homing beacon with a power output of 2,000 watts or more.

hierarchy—A series of items classified according to rank or order.

hi-fi—*See* High Fidelity.

high band—Television channels 7 - 13, covering a frequency range of 174-216 MHz.

high boost—*See* High-Frequency Compensation.

high-contrast image—A picture in which strong contrast between light and dark areas is visible. Intermediate values, however, may be missing.

high definition—Television or facsimile equivalent of high fidelity.

high-energy materials—Also called hard magnetic materials. Magnetic materials having a comparatively high energy product, e.g., materials used for permanent-magnets.

highest probable frequency—Abbreviated HPF. An arbitrarily chosen frequency value 15% above the F_2-layer MUF (maximum usable frequency) for the radio circuit. For the E-layer, the HPF is equal to the MUF.

high fidelity—Popularly called hi-fi. The characteristic which enables a system to reproduce sound as nearly like the original as possible.

high-fidelity receiver—A radio receiver capable of receiving and reproducing, without noticeable distortion, the original modulation impressed on the carrier waves.

high filter—An audio circuit designed to remove undesired high-frequency noise from the program material. Such noise includes record scratch, tape hiss, a-m whistles, etc.

high frequency—Abbreviated hf. The frequency band from 3 to 30 MHz (10 meters to 100 meters).

high-frequency alternator—An alternator capable of generating radio-frequency carrier waves.

high-frequency band—The band of frequencies extending from 3 to 30 MHz.

high-frequency bias—In a tape recorder, a sinusoidal voltage that is mixed with the signal being recorded to better the linearity and dynamic range of the recorded signal. In practice, the bias frequency is three to four times the highest information frequency to be recorded.

high-frequency broadcast station—A commercial station licensed for transmission to the general public and operated in the high-frequency broadcast band, using frequency modulation exclusively.

high-frequency carrier telegraphy—Carrier telegraphy with the carrier currents above the frequencies transmitted over a voice telephone channel.

high-frequency compensation—Also called high boost. An increase in the amplification of the high frequencies with respect to the low and middle frequencies within a given band of frequencies.

high-frequency heating—*See* Electronic Heating.

high-frequency induction heater or furnace—An induction heater or furnace using frequencies much higher than the standard 60-hertz.

high-frequency resistance—Also called r-f or a-c resistance. The total resistance offered by a device in a high-frequency a-c circuit. This includes the d-c and all other resistances due to the effects of the alternating current.

high-frequency treatment—Therapeutic use of intermittent and isolated trains of heavily damped oscillations having a high frequency and voltage and a relatively low current.

high-frequency trimmer—A trimmer capacitor that is used to calibrate the high-frequency end of the tuning range in a superheterodyne receiver.

high-frequency unit—*See* Tweeter.

high-frequency welding—*See* Radio-Frequency Welding.

high-level detector—A linear power detector with a voltage-current characteristic that may be treated as a straight line or two intersecting lines.

high-level firing time—The time required to establish a radio-frequency discharge in a switching tube after radio-frequency power is applied.

high-level modulation—A system in which the modulation is introduced at a point where the power level approximates the output power.

high-level, radio-frequency signal (TR, ATR, and pre-TR tubes)—A radio-frequency signal with sufficient power to fire the tube.

high-level vswr (switching tubes)—The voltage standing-wave ratio due to a fired tube in its mount, located between a generator and the matched termination in a wave guide.

highlight—The brightest portion of a reproduced image.

high-mu tube—A vacuum tube with a high amplification factor.

high-pass filter—A wave filter having a single transmission band extending from some critical, or cutoff, frequency other than zero, up to infinite frequency.

high-performance equipment—Equipment having sufficiently exacting characteristics to permit their use in trunk or link circuits.

high-potential test—A test for determining the breakdown point of insulating materials and spacings. It consists of applying a voltage higher than the rated voltage between two points or between two or more windings. However, it is not a test of conductor insulation.

high-power silicon rectifiers—A group of rectifiers with continuous ratings exceeding 50 average amperes per section in a single-phase, half-wave circuit.

high Q—Having a high ratio of reactance to effective resistance. The factor determining the efficiency of a reactive component.

high-rate discharge—The storage-battery discharge equivalent to the heaviest possible duty in service.

high-recombination-rate contact—A semiconductor - to - semiconductor or metal-to-semiconductor contact at which thermal equilibrium charge-carrier concentrations are maintained substantially independent of current density.

high-resistance joint—A faulty union of conductors or conductor and terminal. The result is less current flow and a drop in voltage at the union.

high-resistance voltmeter—A voltmeter having a resistance considerably higher than 1,000 ohms per volt. As a result, it draws very little current from the circuit being measured.

high-speed bus—*See* Memory Register.

high-speed carry—In a computer, a type of carry in which: 1. A carry into a column results in a carry out of that column, because the sum without carry in that column is 9. 2. Instead of a normal adding process, a special process is used which takes the carry at high speed to the actual column where it is added. Also called standing-on-nines carry.

high-speed d-c circuit breaker—A device which starts to reduce the current in the main circuit in 0.01 second or less, after the occurrence of the d-c overcurrent or the excessive rate of current rise.

high-speed relay—A relay designed specifically for short-operate or short-release time, or both.

high-speed telegraph transmission—Transmission of code at higher speeds than are possible with hand-operated keys.

high tension—Lethal voltages, on the order of thousands of volts.

high-tension magneto—A self-contained generator in which the required high potential is generated directly; no induction coil is needed.

high-vacuum phototube—A phototube that is highly evacuated so that its electrical characteristics are essentially unaffected by gaseous ionization. In a gas phototube, some gas is intentionally introduced.

high-vacuum rectifier—A vacuum-tube rectifier in which conduction is entirely by electrons emitted from the cathode.

high-vacuum tube—An electron tube that is highly evacuated so that its electrical characteristics are essentially unaffected by gaseous ionization.

high-velocity scanning—The scanning of a target with electrons of such velocity that the secondary-emission ratio is greater than unity.

high voltage—The accelerating potential that speeds up the electrons in a beam of a cathode-ray tube.

high-voltage probe—A probe with a high internal resistance, for measuring extremely high voltages. It is used with a voltmeter having an internal resistance of 20,000 ohms per volt or more.

hill-and-dale recording—*See* Vertical Recording.

hinge—A joint in a relay that permits movement of the armature relative to the stationary parts of the relay structure.

hinged-iron ammeter—A moving-iron ammeter in which the fixed portion of the magnetic circuit is placed around the conductor to measure the current through it.

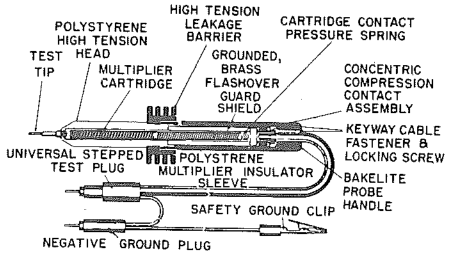
HIGH TENSION LEAKAGE BARRIER
POLYSTYRENE HIGH TENSION HEAD
CARTRIDGE CONTACT PRESSURE SPRING
TEST TIP
MULTIPLIER CARTRIDGE
GROUNDED, BRASS FLASHOVER GUARD SHIELD
CONCENTRIC COMPRESSION CONTACT ASSEMBLY
KEYWAY CABLE FASTENER & LOCKING SCREW
UNIVERSAL STEPPED TEST PLUG
POLYSTRENE MULTIPLIER INSULATOR SLEEVE
BAKELITE PROBE HANDLE
SAFETY GROUND CLIP
NEGATIVE GROUND PLUG

High-voltage probe.

HIPERNAS—Acronym for HIgh PERformance NAvigation System. A self-compensated, pure-inertial guidance system.

hi-pot—1. *See* Dielectric Breakdown Voltage. 2. Contraction of high potential. *Commonly* refers to a device used for testing insulation breakdown or leakage, with high voltages. High potting is the verb.

hiss—Random noise characterized by prolonged sibilant sounds in the audio-frequency range.

histogram—A graphical representation of a frequency distribution by a series of rectangles which have for one dimension a distance proportional to a definite range of frequencies, and for the other dimension a distance proportional to the number of frequencies appearing within the range.

hit—*See* Flash.

hit-on-the-fly printer—A mechanical printer in which the printing head is in continual motion.

hits—Momentary line disturbances which could result in mutilation of characters being transmitted.

H-lines—Imaginary lines that represent the direction and strength of magnetic flux on a diagram.

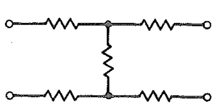

H-network.

H-network—A network composed of five impedance branches. Two are connected in series between an input terminal and an output terminal and two are connected between another input terminal and another output terminal. The fifth is connected from the junction points of the two branches.

hog horn—A microwave feed horn shaped so that the input energy from the waveguide approaches from the same direction as the horn opening.

hold—Opposite of clear. 1. To maintain storage elements in charge storage tubes at equilibrium potentials by electron bombardment. 2. In computers, to retain information contained in one storage device after copying it into a second storage device.

hold control—In a television receiver, the adjustment which controls the frequency of the vertical- or horizontal-scanning pulses and hence the stability of the picture.

hold current—Also called the electrical hold value. The minimum current which will keep the contact springs energized in a relay.

hold electrode—In a mercury switch, the electrode that remains in contact with the mercury pool while the circuit is being closed or opened.

holding anode—In a mercury-arc rectifier, a small auxiliary anode that maintains the ionization while the main anode current is zero.

holding beam—A diffused beam of electrons for regenerating the charges retained on the dielectric surface of an electrostatic memory or storage tube.

holding circuit—Also called a locking circuit. An alternate operating circuit which, when completed, maintains sufficient current in a relay winding to keep the relay energized after the initial current has ceased.

holding coil—A separate relay coil which keeps the relay energized after the original current has been removed.

holding current—That value of average forward current (with the gate open) below which a silicon-controlled rectifier returns to the forward blocking state after having been in forward conduction.

holding gun—In a storage tube, the source of electrons constituting the holding beam.

holding time—The total time a trunk or circuit is in use on a call, including both operator's and user's time.

hold mode—In integrators or other charge-storage circuits, a condition or time interval in which input (s) are removed and the circuit is commanded (or expected) to maintain a constant output.

hold-off voltage—The maximum voltage an electronic flash tube will stand without self-flashing. Normal hold-off voltage is reduced at the end of lamp life and in the presence of high temperatures or r-f fields.

hold time—In resistance welding, the time that is allowed for the weld to harden.

hole—In the electronic valence structure of a semiconductor, a mobile vacancy which acts like a positive electronic charge with a positive mass.

hole conduction—Conduction occurring in a semiconductor when electrons move into holes under the influence of an applied voltage and thereby create new holes. The apparent movement of such holes is toward the more negative terminal, and is hence equivalent to a flow of positive charges in that direction.

hole current—Conduction in a semiconductor when electrons move into holes, creating new holes. The holes appear to move toward the negative terminal, giving the equivalent of positive charges flowing to the terminal.

hole density—In a semiconductor, the density of holes in an otherwise full band.

hole injection—The production of mobile vacancies in an N-type semiconductor when a voltage is applied to a sharp metal point in contact with the surface of the material.

hole injector—A pointed metallic device for injecting holes into an N-type semiconductor.

hole-in-the-center effect—Also called hole-in-the-middle effect. The lower volume or absence of sound between the left and right speakers of a stereo system.

hole mobility—The ability of a hole to travel easily through a semiconductor.

hole site—The area on a computer punch card or paper tape where a hole may or may not be punched. It can be a form of binary storage in which a hole represents a 1 and the absence of a hole represents a 0.

hole storage factor (K'_s)—In a transistor, the excess stored charge (when the transistor is in saturation) per unit excess base current. Excess base current is defined as the amount of current supplied to the base in excess of the current required to just keep the transistor in saturation.

hole trap—A semiconductor impurity which can trap holes by releasing electrons into the conduction or valence bands.

Hollerith code—A code based on the punching of holes in cards at specified locations. From one to three punches may be made in each column of the card and up to 80 columns may be punched in each card. Each column corresponds to one character; the specific character is determined by the number and location of the punches in that column.

hollow-cathode tube—A gas discharge tube with a hollow cathode closed at one end. Almost all the radiation is from the cathode glow within the hollow cathode.

hollow core—A plain ferrite core having a center hole for mounting purposes.

home loop—An operation involving only those input and output units associated with the hole terminal.

homing—1. Approaching a desired point by maintaining some indicated navigational parameter constant (other than altitude). 2. In missile guidance, the use of radiation from a target to establish a collision course.

homing adapter—A device used with an aircraft radio receiver to produce aural and/or visual signals which indicate the direction of a transmitting radio station.

homing antenna—A type of directional-antenna array used for pinpointing a target.

homing beacon—A radio transmitter which emits a distinctive signal for determining bearing, course, or location.

homing device—1. An automatic device that moves or rotates in the correct direction without first having to go to the end of its travel in the opposite direction. 2. A radio device that guides an aircraft to an airport or transmitter site.

homing guidance—A missile-guidance system in which the missile steers itself toward a target by means of a self-contained mechanism (infrared detectors, radar, etc.). Activated by some distinguishing characterism of the target. Homing guidance may be active, semiactive, or passive.

homing relay—A stepping relay that returns to a specified starting position prior to each operating cycle.

homing station—A radionavigational aid incorporating direction-finding facilities.

homodyne reception—Also called zero-beat reception. A system of reception using a locally generated voltage at the carrier frequency.

homogeneity—The state or condition of being similar in nature, kind, or degree.

homogeneous—Of the same nature (the opposite of heterogeneous).

homologous field—A field in which the lines of force in a given plane all pass through one point (e.g., the electric field between two coaxial charged cylinders).

homopolar — Electrically symmetrical — i.e., having equally distributed charges.

homopolar generator—A d-c generator in which all the poles presented to the armature are of the same polarity, so that the armature conductor always cuts the magnetic lines of force in the same direction. A pure direct current can thus be produced without commutation.

homopolar magnet—A magnet with concentric pole pieces.

homosphere—That part of the atmosphere which is made up mostly of atoms and molecules found near the earth's surface, and retaining the same relative proportions of oxygen, nitrogen, and other gases throughout.

homotaxial—A term coined by RCA from "homogeneous" and "axial" to describe a single-diffused transistor with a base region of homogeneous resistivity silicon in the axial direction (emitter-to-collector).

honeycomb coil—Also called duolateral or lattice-wound coil. A coil with the turns wound crisscross to reduce the distributed capacitance.

hood—A shield placed over a cathode-ray tube to eliminate extraneous light and thus make the image on the screen appear more clearly.

hookswitch — A hook or plunger-operated switch used in a telephone set to open the circuit when the instrument is not in use.

hook transistor — A transistor having four alternating P-type and N-type layers, with one layer floating between the base layer and the collector layer. This arrangement gives high emitter-input current gains. The PNPN transistor has a P-type floating layer, while an NPNP transistor has an N-type floating layer.

hookup—1. Method of connection between

the various units in a circuit. 2. The diagram of connections used.

hookup wire—The wire used in coupling circuits together. It may be solid or stranded, and is usually tinned and insulated No. 18 or 20 soft-drawn copper.

hop—An excursion of a radio wave from the earth to the ionosphere and back. It is usually expressed as single-, double-, and multihop. The number of hops is called the order of reflection.

hopoff—In a potentiometer, the sudden jump in resistance as the contact is rotated over the junction of two resistance slopes or over the junction of two resistance slopes or over resistance slope. The magnitude of the hopoff is dependent on the ratio of the slopes and on the junction blending characteristic.

horizon—An apparent or visible junction of earth and sky as seen on or above the earth. It bounds the part of the earth's surface that can be reached by the direct wave of a radio station. The distance to the horizon is affected by atmospheric refraction.

horizon distance—The space between the farthest visible point and the transmitter antenna. It is the distance over which ultra-high-frequency transmission can be received under ordinary conditions with an unelevated receiving antenna.

horizontal—1. Perpendicular to the direction of gravity. 2. In the direction of or parallel to the horizon. 3. On a level.

horizontal axes—The three horizontal axes of crystallographic reference.

horizontal blanking—Cutting off the electron beam between successive active horizontal lines during retrace.

horizontal blanking pulse—A rectangular pedestal in the composite television signal. It occurs between active horizontal lines and cuts off the beam current of the picture tube during retrace.

horizontal centering control—In a television receiver or cathode-ray oscilloscope, the adjustment for moving the entire display back and forth.

horizontal-convergence control—In a color television receiver, the control which adjusts the amplitude of the horizontal dynamic convergence voltage.

horizontal-deflecting electrode—A pair of electrodes that move the electron beam from side to side on the screen of a cathode-ray tube employing electrostatic deflection.

horizontal-discharge tube—A vacuum tube used in the horizontal-deflection circuit to discharge a capacitor and thereby form the sawtooth scanning wave. (*Also see* Discharge Tube.)

horizontal-drive control—In an electromagnetically deflected television receiver, the control which adjusts the ratio of the pulse amplitude to the linear portion of the scanning-current wave.

horizontal dynamic convergence—Convergence of the three electron beams in a color

picture tube at the aperture mask during scanning of a horizontal line.

horizontal frequency—*See* Line Frequency.

horizontal hold control—A synchronization control which varies the free-running frequency of the horizontal deflection oscillator so it will be in step with the scanning frequency at the transmitter.

horizontal hum bars — Broad, horizontal, moving or stationary bars, alternately dark and light, that extend over an entire television picture. They are caused by interference at approximately 60 Hz or a harmonic of 60 Hz.

horizontal-linearity control—In a television receiver, the control for adjusting the width at the left side of the screen.

horizontal line frequency—*See* Line Frequency.

horizontal lock—The circuit that maintains horizontal synchronization in a television receiver.

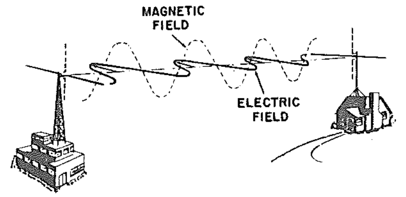

Horizontally polarized wave.

horizontally polarized wave — A linearly polarized wave with a horizontal electric-field vector.

horizontal-output transformer—*See* Flyback Transformer.

horizontal parabola control—*See* Phase Control.

horizontal polarization — Transmission in which the electrostatic field leaves the antenna in a horizontal plane. Elements of the transmitting and receiving antennas likewise are horizontal. Horizontal polarization is standard for television in the United States.

horizontal repetition rate—Also called horizontal scanning frequency. The number of horizontal lines per second (15,750 hertz in the United States).

horizontal resolution—The number of picture elements in a horizontal scanning line that can be distinguished.

horizontal retrace—The line that would be seen on the screen, while the spot is returning from right to left, if retrace blanking were not used.

horizontal ring-induction furnace—A furnace for melting metal. It comprises an open trough or melting channel, a primary inductor winding, and a magnetic core which links the melting channel to the primary winding.

horizontal-scanning frequency—*See* Horizontal-Repetition Rate.

horizontal sweep—Movement of the electron beam from left to right across the screen or the scene being televised.

horizontal-sync discriminator — A circuit employed in the flywheel method of synchronization to compare the phase of the horizontal-sync pulses with that of the horizontal-scanning oscillator.

horizontal-sync pulse — The rectangular pulses which occur above the pedestal level between each active horizontal line. They keep the horizontal scanning at the receiver in step with the horizontal scanning at the transmitter.

horn—Also called an acoustic horn. A tubular or rectangular enclosure for radiating or receiving acoustic waves. (*Also see* Horn Antenna.)

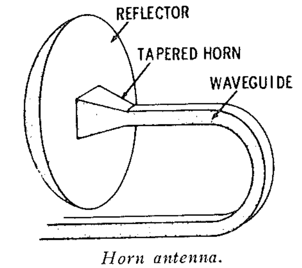

Horn antenna.

horn antenna—Also called a horn. A tubular or rectangular microwave antenna which is wider at the open end and through which radio waves are radiated into space.

horn arrester—A lightning arrester that has a spark gap with upward-projecting diversion horns of thick wire. When the arc is formed, it travels up the gap, and is extinguished upon reaching the widest part of the gap.

horn cutoff frequency—A frequency below which an exponential horn will not function correctly, because it fails to provide for proper expansion of the sound waves.

horn loading — A method of coupling a speaker diaphragm to the listening space by an expanding air column having a small throat and large mouth.

horn mouth—The wide end of a horn.

horn radiator—An open-ended, metallic device for concentrating energy from a waveguide and directing this energy into space.

horn speaker—A speaker in which a horn couples the radiating element to the medium.

horn throat—The narrow end of a horn.

horsepower — Abbreviated hp. A unit of power, or the capacity of a mechanism to do work. It is the equivalent of raising 33,000 pounds one foot in one minute, or 550 pounds one foot in one second. One horsepower equals 746 watts.

horseshoe magnet—A permanent magnet or electromagnet shaped like a horseshoe or U to bring the two poles close together.

hot—1. Connected, alive, energized; pertains to a terminal or any ungrounded conductor. 2. Not grounded.

hot-carrier diode — A semiconductor diode

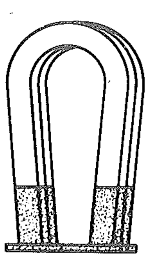

Horseshoe magnet.

which uses the thermoelectric effect of radio-frequency heating in bulk semiconductor material to provide current carriers. Characterized by relatively high power and low internal noise generation.

hot carriers—In barrier diodes, carriers that have energies greater than those that are in thermal equilibrium with the metal. Thus, electrons that cross the junction from semiconductor to metal must be energetic enough to surmount the barrier. Therefore electrons that are energetic enough to cross the junction are called "hot electrons."

hot cathode—Also called thermionic cathode. A cathode which supplies electrons by thermionic emission. (As opposed to a cold cathode, which has no heater.)

hot-cathode tube—Also called thermionic tube. Any electron tube containing a hot cathode.

hot-cathode X-ray tube—A high-vacuum X-ray tube in which a hot rather than cold cathode is used.

hot electrons—*See* Hot Carriers.

hot spot—The point of maximum temperature on the outside of a device or component.

hot tin dip—A process of passing a bare wire through a bath of molten tin to provide a coating.

hot-wire ammeter—Also called thermal ammeter. An ammeter in which the expansion of a wire moves a pointer to indicate the amount of current being measured. The current flows through the wire and changes its length in proportion to I^2. Instability because of wire stretching, and the lack of ambient temperature compensation make the hot-wire ammeter commercially unsatisfactory.

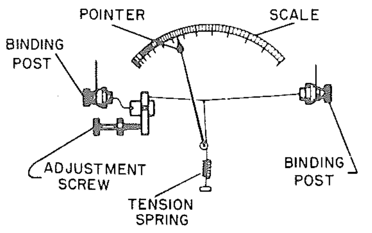

Hot-wire ammeter.

hot-wire anemometer—An instrument that measures the velocity of wind or a gas by its cooling effect on an electrically heated wire.

hot-wire instrument—An electrothermic instrument operated by expansion of a wire heated by the current it is carrying.

hot-wire microphone—A microphone in which the cooling or heating effect of a sound wave changes the resistance of a hot wire and thus the current through it.

hot-wire relay—A form of linear-expansion time-delay relay in which the longitudinal expansion of a wire, when heated, provides the mechanical motion to open or close contacts. The time required to heat the wire constitutes the delay.

hot-wire transducer—A unilateral transducer in which the cooling or heating effect of a sound wave changes the resistance of a hot wire and thus the current.

housekeeping—In a computer routine, those operations, such as setting up constants and variables for use in the program, that contribute directly ot the proper operation of the computer but not to the solution of the problem.

howl—An undesirable prolonged wail produced in a speaker by electric or acoustic feedback.

howler—An electromechanical device which produces an audio-frequency tone.

hp—Abbreviation for horsepower.

H-pad—An attenuation network in which the elements are arranged in the form of the letter *H*.

h-parameters—*See* Hybrid Parameters.

h-particle—The positive hydrogen ion or proton resulting from bombardment of the hydrogen atom by alpha rays or fast-moving positive ions.

HPF—Abbreviation for highest probable frequency.

H plane—The plane in which the magnetic field of an antenna lies. It is perpendicular to the E plane. The principal H plane of an antenna is the H plane that also contains the direction of maximum radiation.

H-plane bend—*See* H-Bend.

H-plane T-junction—Also called shunt T-junction. A waveguide T-junction in which the structure changes in the plane of the magnetic field.

H-scan—*See* H-Display.

hub—On a control panel or plugboard, a socket or receptacle into which an electrical lead or plug wire may be connected for the purpose of carrying signals.

hue—Often used synonymously with the term "color," but does not include gray. It is the dominant wavelength—i.e., the one which distinguishes a color as red, yellow, etc. Varying saturations may have the same hue.

hue control—On a color television receiver, the operating control that changes the hue (color) of the picture.

hull potential—The voltage difference between a reference electrode and an immersed metallic hull, or the bonded underwater metallic appendages of a nonmetallic hull.

hum—In audio-frequency systems, a low-pitched droning noise consisting of several harmonically related frequencies. It results from an alternating-current power supply, ripple from a direct-current power supply, or induction from exposure to a power system. By extension, the term is applied in visual systems to interference from similar sources.

human engineering—The science and art of developing machines for human use, giving consideration to the abilities, limitations, habits, and preferences of the human operator.

hum-balancing pot—A potentiometer usually placed across the heater circuit. Its arm is grounded so that the heater voltage is balanced with respect to ground.

hum bar—A dark band extending across the picture. It is caused by excessive 60-hertz hum (or harmonics) in the signal applied to the picture-tube input.

hum bucking—The introduction of a small amount of voltage, at the power-line frequency, into a circuit to cancel unwanted power.

hum-bucking coil—A coil wound around the field coil of a dynamic speaker and connected in series opposition with the voice coil. In this way, any hum voltage induced in the field coil will be induced in the voice coil in the opposite direction and buck, or cancel, the effects of the hum.

humidity transducer—A layer of hygroscopic (moisture-absorbing) substance deposited between two metal electrodes. These electrodes establish electrical contact with the hygroscopic chemical, which serves as a resistance element. Since the chemical coating tends to absorb moisture from the surrounding air, its resistance decreases as the humidity increases. In this manner, humidity variations are converted to resistance variations.

humming—A sound produced by transformers having loose laminations or by magnetostriction effects in iron cores. The frequency of the sound is twice the power-line frequency.

hum modulation—Modulation of a radio-frequency or detected signal by using hum.

hunting—An undesirable oscillation of an automatic control system where the controlled variable swings on both sides of the predetermined reference value without settling on it.

hv or HV—Abbreviation for high voltage.

H-vector—A vector which represents the magnetic field of an electromagnetic wave. In free space, it is perpendicular to the E-vector and the direction of propagation.

H-wave—A mode in which electromagnetic energy can be transmitted in a waveguide. An H-wave has an electric field perpendicular to the length of the waveguide, and a magnetic field parallel as well as perpendicular to the length.

hybrid—*See* Hybrid Junction.

hybrid balance—A measure of the degree of balance between two impedances connected to two conjugate sides of a hybrid set. It is given by the formula for return loss.

hybrid circuit—1. A circuit which combines the thin-film and semiconductor technologies. Generally, the passive components are made by thin-film techniques, and the active components by semiconductor techniques. The active devices are attached to the thin-film passive components by a suitable bonding process. 2. Also called two-wire–four-wire terminating set. In telephone transmission circuits, a circuit for interconnecting two-wire and four-wire circuits through a differential balance or bridge circuit in which the two sides of the four-wire circuit form conjugate arms.

hybrid coil—Also called bridge transformer. A single transformer which has, effectively, three windings and which is designed to be connected to four branches of a circuit so as to render these branches conjugate in pairs.

hybrid computer—A computer that results from the interconnection of an analog computer and a digital computer, plus conversion equipment, each contributing its special advantages to an assigned part of the solution of a class of complex problems.

hybrid electromagnetic wave—Abbreviated HEM wave. An electromagnetic wave having components of both electric- and magnetic-field vectors in the direction of propagation.

hybrid integrated circuit—1. An arrangement that consists of one or more integrated circuits combined with one or more discrete component parts. 2. The combination of more than one type of integrated circuit into one integrated component.

hybrid junction—1. Also called hybrid-t or magic-t. A waveguide arrangement with four branches. When they are properly terminated, energy is transferred from any one branch into two of the remaining three branches. In common usage, this energy is divided equally between the two. 2. Also called bridge hybrid, and hybrid. A transformer, resistor, or waveguide circuit or device that has four pairs of terminals so arranged that a signal entering at one terminal pair will divide and emerge from the two adjacent terminal pairs, but will be unable to reach the opposite terminal pair.

hybrid microcircuit—A microcircuit in which thin-film or diffusion techniques are combined with separately attached semiconductor chips to form the circuit.

hybrid network—A nonhomogeneous communication network required to operate with signals of dissimilar characteristics (such as analog and digital modes).

hybrid parameters—Also called h-parameters. The resultant parameters of an equivalent transistor circuit when the input current and output voltage are selected as independent variables.

hybrid ring—Also called a "rat race." A hybrid junction commonly used as an equal power divider. It consists of a re-entrant line (waveguide) to which four side arms are connected. The line is of the proper electrical length to sustain standing waves.

hybrid set—Two or more transformers interconnected to form a network having four pairs of accessible terminals. Four impedances may be connected to the four terminals, so that the branches containing them may be made conjugate in pairs.

hybrid-t—*See* Hybrid Junction.

hybrid thin-film circuit — A microcircuit formed by attaching discrete components and semiconductor devices to networks of passive components and conductors that have been vacuum deposited on glazed ceramic, sapphire, or glass substrates.

hybrid transformer—*See* Hybrid Coil.

hydroacoustics—The generation of acoustic energy from the flow of fluids under pressure.

hydroacoustic transducer — A transducer that produces high-level acoustic energy from the flow of high-pressure fluid.

hydroelectric—The production of electricity by water power.

hydrogen electrode—A platinum electrode covered with platinum black, around which a stream of hydrogen is bubbled. The hydrogen electrode furnishes a standard against which other electrode potentials can be compared.

hydrogen lamp—A special light source, used in some spectrophotometers, which produces invisible light energy. It is used in finding the light-energy frequency of test solutions.

hydrogen thyratron — A thyratron containing hydrogen.

hydromagnetics — *See* Magnetohydrodynamics.

hydromagnetic waves—Waves in which the energy oscillates between the magnetic field energy and kinetic energy of the hydrodynamic motion, the reservoirs being the self-inductance of the conductive matter and the mass inertia of the moving fluid.

hydrometer—An instrument for measuring the specific gravity of a liquid such as the electrolyte of a storage battery. It contains a graduated float which indicates the specific gravity by the amount of liquid displaced.

hydrophone — An electroacoustic transducer which responds to waterborne sound waves and delivers essentially equivalent electric waves.

hydrostatic pressure—*See* Static Pressure.

hygrometer—An instrument that measures the relative humidity of the atmosphere.

hygroscopic—Readily absorbing and retaining moisture from the atmosphere. The opposite term is nonhygroscopic.

hygrostat—A device that closes a pair of contacts when the humidity reaches a prescribed level.

hyperacoustic zone—The region in the upper atmosphere, above about 60 miles, in which the distance between the air molecules is roughly equal to the wavelength of sound, so that sound is transmitted with less efficiency than at lower levels. Sound waves cannot be propagated above this zone.

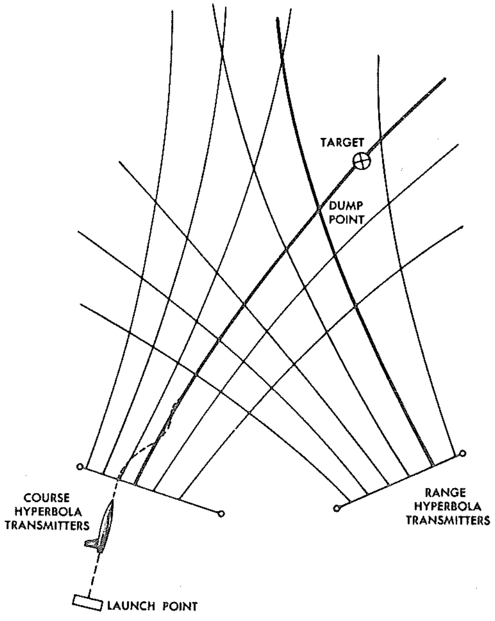

Hyperbolic guidance.

hyperbola — 1. A curve that is the focus of points having a constant difference of distance from two fixed points. 2. In hyperbolic guidance systems, a path along which the difference between the arrival times of pulses from two transmitters is constant. (*Also see* Hyperbolic Guidance System).

hyperbolic error — The error in an interferometer system arising from the assumption that the directions of the wavefronts incident at two antennas of a base line are parallel, whereby the equiphase path is a cone. Mathematically the equiphase path is a hyperbola.

hyperbolic grind—A shape of tape playback and record heads. It permits good head contact and better response at high frequencies.

hyperbolic guidance system — A method of guidance in which sets of ground stations transmit pulses from which a hyperbolic path can be derived to give range and course information for steering. (*Also see* Hyperbola.)

hyperbolic horn—A horn in which the equivalent cross-sectional radius increases according to a hyperbolic law.

hyperbolic navigation system—A method of radionavigation (e.g., loran) in which pulses transmitted by two ground stations are received by an aircraft or ship. The difference in arrival time from each station is a measure of the difference in distance between the aircraft or ship and each station. This distance is plotted on one of many hyperbolic curves on a map. A second reading from another pair of stations (or from the same master and a different slave) establishes another point on a different hyperbolic curve. The intersection of the two curves gives the position of the aircraft or ship.

hyperfrequency waves — Microwaves having wavelengths in the range from 1 centimeter to 1 meter.

hypersonic—Having five or more times the speed of sound.

hyspersyn motor — A synchronous motor which combines the desirable features of the induction, hysteresis, and d-c excited synchronous motor, resulting in high efficiency and power factor. It possesses the vigorous starting torque of an induction motor, the synchronizating torque of a hysteresis motor, and the stiffness of an externally d-c excited synchronous motor.

hysteresigraph—A device for experimentally presenting or recording the hysteresis loop of a magnetic specimen.

hysteresis—1. The amount the magnetization of a ferrous substance lags the magnetizing force because of molecular friction. 2. A type of oscillator behavior where multiple values of the output power and/or frequency correspond to given values of an operating parameter. 3. The temporary change in the counting-rate-vs-voltage characteristic of a radiation-counter tube (caused by previous operation). 4. The difference between the response of a unit or system to an increasing and a decreasing signal. 5. A form of nonlinearity in which the response of a circuit of a particular set of input conditions depends, not only on the instantaneous values of those conditions, but also on the immediate past (recent history) of the input and output signal. Hysteretical behavior is characterized by inability to "retrace" exactly on the reverse swing a particular locus of input/output conditions.

hysteresis brake—*See* Hysteresis Clutch.

hysteresis clutch — Also called hysteresis brake. A proportional torque-control device which employs the hysteresis effect in a permanent-magnet rotor to develop its output torque. It is capable of synchronous driving or continuous slip, provided heat can be removed, with almost no torque variation at any slip differential. Its control-power requirement is small enough for vacuum-tube or transistorized drive.

hysteresiscope—An instrument used to obtain hysteresis loops on a cathode-ray oscilloscope screen without the need for specially prepared ring samples. It is used in the inspection of magnetic material.

hysteresis curve—A curve showing the relationship between a magnetizing force and the resultant magnetic flux.

hysteresis distortion — Distortion of waveforms in circuits containing magnetic components. It is due to the hysteresis of the magnetic cores.

hysteresis error—The difference in the reading obtained on a measuring instrument containing iron when the current is increased to a definite value and when the current is reduced from a higher value to the same definite value.

hysteresis heater—An induction device in which a charge (or a muffle around the charge) is heated by hysteresis losses due to the magnetic flux produced in the charge.

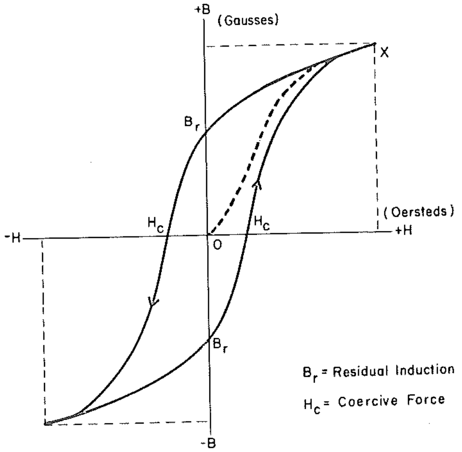

Hysteresis loop.

hysteresis loop—For a magnetic material, a curve showing two values of magnetic flux density—one when the magnetizing force is increasing and the other when it is decreasing—for each value of the magnetizing force.

hysteresis loss—The power expended in a magnetic material as a result of magnetic hysteresis.

hysteresis meter—An instrument for determining the hysteresis loss in a ferromagnetic material. It measures the torque produced when the test specimen is placed in a rotating magnetic field or is rotated in a stationary magnetic field.

hysteresis motor — A synchronous motor without salient poles or direct-current excitation. It is started by the hysteresis losses induced in its secondary by the revolving field.

hystoroscope — An instrument used to observe, measure, and record the magnetic characteristics of both easy and hard axes of magnetic materials.

Hz—Abbreviation for hertz — the new term used in place of cycles per second.

I

I—1. Symbol for current. 2. Abbreviation for luminous intensity.

IC—1. Abbreviation for internal connection. 2. Abbreviation for integrated circuit.

ICBM—Abbreviation for intercontinental ballistic missile.

I_{CBO}—The reverse current that occurs when a specific d-c voltage is applied in the nonconducting directon to the collector junction of a transistor while the emitter is open-circuited.

iconoscope—A camera tube in which a beam of high-velocity electrons scans a photoemissive mosaic capable of storing an electrical charge pattern.

icw—Abbreviation for interrupted continuous wave.

ideal bunching — A theoretical condition where bunching of the electrons in a velocity-modulated tube would give an infinitely large current peak during each cycle.

ideal capacitor—A capacitor having a single-valued transferred-charge characteristic.

ideal crystal—A crystal having no mosaic structure and capable of X-ray reflection in accordance with the Darwin-Ewald-Prins law.

ideal dielectric—A dielectric in which all the energy required to establish an electric field in the dielectric is returned to the source when the field is removed. (A perfect dielectric must have zero conductivity. Also, all absorption phenomena must be lacking. A vacuum is the only known perfect dielectric.)

ideal-noise diode—A diode that has an infinite internal impedance and in which the current exhibits full shot-noise fluctuations.

ideal transducer—Theoretically, any linear passive transducer which—if it dissipated no energy and, when connected to a source and load, presented its combined impedance to each—would transfer maximum power from source to load.

ideal transformer — A hypothetical transformer which would neither store nor dissipate energy. Its self-inductances would have a finite ratio and unity coefficient of coupling, and its self- and mutual impedances would be pure inductances of infinitely great value.

I demodulator—A demodulator circuit the inputs of which are the chrominance signal and the signal from the local 3.58-MHz oscillator. The output of this demodulator is a video signal representing color in the televised scene. The Q demodulator is similar except that its input from the local oscillator is shifted 90°.

identification—1. In radar, determining the identity of a displayed target (i.e., which one of the blips in the display represents the target). 2. In a computer, a code number or code name which uniquely identifies a record, block, file, or other unit of information.

identification beacon—A code beacon used for positively identifying a particular point on the earth's surface.

identification, friend or foe—A system using radar transmissions to which equipment carried by friendly forces automatically responds, for example, by emitting pulses, thereby distinguishing themselves from enemy forces. It is the primary method of determining the friendly or unfriendly character of aircraft and ships by other aircraft or ships and by ground forces employing radar-detection equipment and associated identification, friend or foe units. Abbreviated iff.

identify—In a computer, to attach a unique code or code name to a specific unit of information.

idiochromatic — Having photoelectric properties characteristic of the pure crystal itself and not due to foreign matter.

I-display—In radar, a display in which a target appears as a circle when the radar antenna is pointed directly at it. The radius of the circle is proportionate to the target distance. When the antenna is not pointing at the target, only a segment of the circle appears. Its length is inversely proportionate to the magnitude of the pointing error, and the segment points away from the direction of error.

idle characters — Control characters interchanged by a synchronized transmitter and receiver to maintain synchronization during a nondata period.

idler—A rubber-tired wheel that transfers

power from a phonograph motor to the turntable rim.

idler pulley—A pulley used only for tightening a belt or changing its direction. The shaft does not drive any other part.

idling current—Also called quiescent current. The zero-signal power supply current drawn by a circuit or by a complete amplifier.

IEC's — *See* Integrated Equipment Components.

IEEE—A professional organization of scientists and engineers whose purpose is the advancement of electrical engineering, electronics, and allied branches of engineering and science. (The IEEE resulted from the merger of the IRE and the AIEE.)

if—Abbreviation for intermediate frequency.

i-f amplifier — *See* Intermediate-Frequency Amplifier.

iff—*See* Identification, Friend or Foe.

IFRU—*See* Interference-Rejection Unit.

i-f strip—*See* Intermediate-Frequency Strip.

ignistor—A transistor and a matched zener diode integrated into one power package.

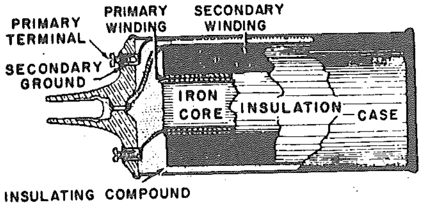

Ignition coil.

ignition coil—An iron-core transformer which converts a low direct voltage to the 20,000 volts or so required to produce an ignition spark in gasoline engines. It has an open core, a heavy primary winding connected to the battery or other source through a vibrating armature contact, and a secondary winding with many turns of fine wire.

ignition control—Control of the instant that static current begins to flow in the anode circuit of a gas tube.

ignition interference — Noise produced by sparks or other ignition discharges in a car, motor, or furnace ignition, or by equipment with loose contacts or connections.

ignition reserve—In a gasoline engine, the difference between the available voltage and the required ignition voltage.

ignition voltage — In a gasoline engine, the peak voltage required to produce a spark across the plug electrodes.

ignitor discharge—In switching tubes, a d-c glow discharge between the ignitor electrode and a suitably located electrode. It is used to facilitate radio-frequency ionization.

ignitor electrode—An electrode (which is partly immersed in the mercury-pool cathode of an ignitron) used to initiate conduction at the desired points in each cycle.

ignitor firing time—In switching tubes, the interval between application of a d-c voltage to the ignitor electrode and start of current flow.

ignitor interaction—In a TR, pre-TR or attenuator tube, the difference between the insertion loss measured at a specified level of ignitor current and that measured at zero ignitor current.

ignitor leakage resistance—In a switching tube, the insulation resistance measured between the ignitor electrode terminal and the adjacent r-f electrode in the absence of an ignitor discharge.

ignitor oscillation—A relaxation type of oscillation in the ignitor circuit of a TR, pre-TR, or attenuator tube.

ignitor voltage drop—In switching tubes, the d-c voltage between the cathode and anode at a specified ignitor current.

ignitron — A type of mercury-pool rectifier which has only one anode. The arc is started for each cycle of operation by an ignitor which dips into the mercury pool. The mercury pool serves as the cathode of the rectifier. The ignitron is characterized by the ability to withstand tube currents several times as high as rated values for a few cycles.

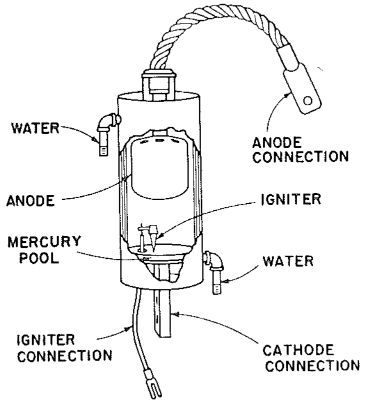

Ignitron.

ignore—In a computer, a character code indicating that no action is to be taken.

IGY — Abbreviation for international geophysical year.

IHFM—Abbreviation for Institute of High-Fidelity Manufacturers, an association of manufacturers which publishes ratings and standards for high-fidelity equipment.

illegal character—A character which a computer system cannot properly use because the structure of the character does not agree with the machine logic. Such characters can be generated by machine failure or improper programming.

illuminance—The density of the luminous flux on a surface. It equals the flux divided by the area of the surface when uniformly illuminated.

illuminant-C—The reference white of color television — i.e., light which most nearly matches average daylight.

illuminate—1. To expose to light. 2. In radar, to strike with a radar signal so that reflection returns the signal to the source for interpretation.

illumination—The density of luminous flux (amount of light) striking a surface.

illumination control—A photo-relay circuit that turns on artificial lighting when natural illumination decreases below a predetermined level.

illumination sensitivity—The output current of a photosensitive device divided by the incident illumination at constant electrode voltages.

illuminometer—A portable photometer for measuring the illumination on a surface.

ILS — Abbreviation for instrument landing system.

image—1. The instantaneous illusion of a picture on a flat surface. 2. The unused one of the two groups of sidebands generated in amplitude modulation.

image admittance—Reciprocal of image impedance.

image antenna—The imaginary counterpart of an actual antenna. For mathematical purposes it is assumed to be located below the ground and symmetrical with the actual antenna.

image-attenuation constant—The real part of the transfer constant.

image converter—A solid-state optoelectric device capable of changing the spectral characteristics of a radiant image. Examples of such changes are infrared-to-visible and X-ray-to-visible.

image-converter tube—*See* Image Tube.

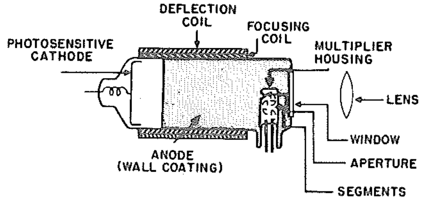

Image dissector.

image dissector—A television camera tube in which the image is swept past an aperture in a series of 525 interlaced lines thirty times per second. Instead of a beam scanning the image, the entire image is scanned past the aperture, which "dissects" the image—hence, the name. (*Also see* Dissector Tube.)

image distortion—Failure of the reproduced image in a television receiver to resemble the original scene scanned by the camera.

image effect—An effect produced on the field of an antenna as the electromagnetic waves are reflected from the earth's surface.

image force—The force on a charge due to that charge or polarization which it induces on neighboring conductors or dielectrics.

image frequency — In heterodyne frequency converters, an undesired input frequency capable of producing the selected frequency by selecting one of the two sidebands produced by beating. The word "image" implies the mirrorlike symmetry of signal and image frequencies about the beating oscillator frequency or intermediate frequency, whichever is higher.

image-frequency rejection ratio—Of a superheterodyne receiver, the ratio of the response at the desired frequency to the response at the image frequency.

image iconoscope—An iconoscope in which greater sensitivity is obtained by separating the function of charge storage from that of photoelectric emission. An optical image is projected on a continuous photosensitive screen, and the electron emission from the back of this screen is focused electromagnetically onto a mosaic screen that is scanned by an electron beam as in the original emitron cathode-ray tube. The British term is super-emitron.

image impedance—The impedances which will simultaneously terminate all inputs and outputs of a transducer in such a way that at each of its inputs and outputs the impedances in both directions are equal.

image intensifier — A system for increasing the sensor response to a radiation pattern or image by interposing active elements between the sensor and the image, and supplying power to the active element. This is normally done by focusing the scene to be imaged on the photocathode of the tube, giving rise to a photoelectron pattern corresponding to the optical image. This pattern is accelerated and focused onto a phosphor which emits light to reproduce a visual image of the scene.

image interference—In a receiver, a response due to signals of a frequency removed from the desired signal by twice the intermediate frequency.

image-interference ratio—In a superheterodyne receiver, the effectiveness of the preselector in rejecting signals at the image frequency.

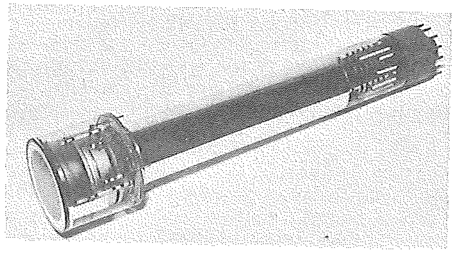

Image orthicon.

image orthicon—A camera tube in which a photoemitting surface produces an electron image and focuses it on one side of a separate storage target. The opposite side of the target is then scanned by low-velocity electrons to produce the output.

image phase constant—The imaginary part of the transfer constant.

image ratio—In a heterodyne receiver, the ratio of the image-frequency signal input to the desired signal input for identical amplitude outputs.

image rejection—The suppression of image-frequency signals in a superheterodyne receiver.

image response—The response of a heterodyne receiver to a signal that is separated by twice the intermediate frequency from the frequency to which the receiver is tuned. Unless there is some preselection, images will cause spurious unwanted responses when the spectrum occupied by a signal is greater than twice the frequency of the first i-f stage of the receiver.

image-transfer constant—*See* Transfer Constant.

image tube—Also called an image-converter tube. An electron tube which reproduces on its fluorescent screen an image of an irradiation pattern incident on its photosensitive surface.

immersion pyrometer — An instrument for determining molten-steel temperature and normally consisting of a platinum-platinum rhodium bimetal thermocouple junction and a recording device for transposing the millivoltage into degrees of temperature.

impact excitation—The starting of damped oscillations by a sudden surge, such as by a spark discharge.

impact predictor—A device which can determine, in real time, the point on the earth's surface where a ballistic missile will impact if thrust is instantaneously terminated.

impedance—The total opposition (i.e., resistance and reactance) a circuit offers to the flow of alternating current. It is measured in ohms, and its reciprocal is called admittance.

impedance angle — Angle of the impedance vector with respect to the resistance vector. Represents the phase angle between voltage and current.

impedance bridge — A device for measuring the combined resistance and reactance of a component part of a circuit.

impedance characteristic—A graph of impedance versus frequency of a circuit or component.

impedance coil—A coil in which its inductive reactance is used to hinder the flow of alternating current in or between circuits.

impedance compensator—An electric network used with a line or another network to give the impedance of the combination a certain characteristic over a desired frequency range.

impedance coupling—A method of coupling using an impedance as the coupling device, common to both the primary and secondary circuits. This type of coupling is usually limited to audio systems, where high gain and limited bandpass are required.

impedance drop—The vector sum of the resistance drop and the reactance drop. (For transformers, the resistance drop, the reactance drop, and the impedance drop are, respectively, the sum of the primary and secondary drops reduced to the same terms. They are usually expressed in percent of the secondary-terminal voltage.)

impedance irregularities—Breaks or abrupt changes which occur in an impedance-frequency curve when unlike sections of a transmission line are joined together or when there are irregularities on the line.

impedance matching—The connection across a source impedance of another impedance having the same magnitude and phase angle. If the source is a transmission line, reflection is thereby avoided.

impedance-matching transformer—A transformer used to match the impedance of a source and load.

impedance transformer—A transformer that transfers maximum energy from one circuit to another.

impedance triangle—A diagram consisting of a right triangle. The sides are proportional to the resistance and reactance in an a-c circuit, with the hypotenuse representing the impedance.

imperfect dielectric—A dielectric in which part of the energy required to establish its electric field is converted into heat instead of being returned to the electric system when the field is removed.

imperfection (of a crystalline solid)—Any deviation in structure from an ideal crystal (one which is perfectly periodic in structure and contains no foreign atoms).

implode—The inward bursting of a picture tube due to its high vacuum.

impregnated coils—Coils that have been permeated with an electric grade varnish or other protective material to protect them from mechanical vibration, handling, fungus, and moisture.

impregnation—The process of coating the insides of coils and closely packed electronic assemblies by dipping them into a liquid and letting it solidify.

impressed voltage—The voltage applied to a circuit or device.

improvement threshold—A characteristic of f-m radio receivers which determines the minimum r-f signal power required to overcome the inherent thermal noise. For increasing values of r-f power above this point, an improvement of signal-to-noise ratio is obtained.

impulse—A pulse that begins and ends within so short a time that it may be regarded mathematically as infinitesimal. The change produced in the medium, however, is generally of a finite amount. (*Also see* Pulse.)

impulse bandwidth — The area divided by the height of the voltage-response selectivity curve as a function of frequency. It is used in the calculation of broad-band interference.

impulse-driven clock—An electric clock in which the hands are moved forward at regular intervals by current impulses from a master clock.

impulse excitation—Also called shock excitation. A method of producing oscillatory current in which the duration of the impressed voltage is relatively short compared with that of the current produced.

impulse frequency—The number of pulse periods per second generated by the dial-pulse springs in a telephone as they rapidly open and close in response to the dialing of a digit.

impulse generator—1. Also called surge generator. An electric apparatus that produces high-voltage surges for testing insulators and for other purposes. 2. A device that generates a broad energy spectrum by means of a very narrow impulse. Usually generated by discharging a short coaxial or waveguide transmission line. The pulses are discrete and regularly spaced, and are generally variable at a repetition rate from a few pulses per second to a few thousand pulses per second. The output of an impulse generator is specified as the rms equivalent of the peak voltage in db above 1 microvolt per megahertz. 3. An oscillator circuit that generates electric impulses for synchronizing purposes in a television system.

impulse noise — Noise due to disturbances having abrupt changes and of short duration. (These noise impulses may or may not have systematic phase relationships. The noise is characterized by nonoverlapping transient disturbance. The same source may produce impulse noise in one system and random noise in a different system.)

impulse-noise generator — Equipment for generating repetitive pulses which provide random noise signals uniformly spread over a wide band of frequencies.

impulse ratio — The ratio of the flashover, sparkover, or breakdown voltage of an impulse to the crest value of the power-frequency flashover, sparkover, or breakdown voltage.

impulse relay—1. A relay that stores enough energy from a brief impulse to complete its operation after the impulse ends. 2. A relay that can distinguish between different types of impulses, operating on long or strong impulses, and not operating on short or weak ones. 3. An integrating relay.

impulse sealing—A heat-sealing technique in which a pulse of intense thermal energy is applied to the sealing area for a very short time, followed immediately by cooling. It is usually accomplished by using an r-f heated metal bar which is cored for water cooling, or is of such a mass that it will cool rapidly at ambient temperatures.

impulse separator — Normally called sync separator. In a television receiver, the circuit that separates the synchronizing impulses from the video information in the received signal.

impulse speed—The rate at which a telephone dial mechanism makes and breaks the circuit to transmit pulses.

impulse timer—A timing device electrically powered by a synchronous motor, featuring a mechanical stepping device which enables it to advance a predetermined number of degrees within a predetermined time interval, controlling a multiple number of circuits. Said circuits are controlled by individual cams which program their activity.

impulse train—*See* Pulse Train.

impulse transmission—The form of signaling used principally to reduce the effects of low-frequency interference. Impulses of either or both polarities are employed for transmission, to indicate the occurrence of transitions in the signals.

impulse-transmitting relay — A relay in which a set of contacts closes briefly when the relay changes from the energized to the de-energized position, or vice-versa.

impulse-type telemeter—A telemeter which employs the characteristics of intermittent electric signals, other than their frequency, as the translating means.

impurity—An atom that is foreign to the crystal in which it exists. Material added to a semiconductor crystal to produce excess electrons (a donor impurity) or excess holes (an acceptor impurity).

impurity density—The amount of impurity material diffused into a certain volume of semiconductor material used in manufacturing semiconductor devices.

impurity level—The energy level existing in a substance because of impurity atoms.

inaccuracy — 1. The difference between the input quantity applied to a measuring instrument and the output quantity indicated by that instrument. The inaccuracy of an instrument is equal to the sum of its instrument error and its uncertainty. 2. The term sometimes used to indicate the deviation from an indicated or recorded value or the measure of conformity to an accepted standard.

inactive leg—Within a transducer, an electrical element which does not change its electrical characteristics with the applied stimulus. Applied specifically to elements that complete a Wheatstone bridge in certain transducers.

in-band signaling—The transmission of signaling tones at some frequency or frequencies within the channel normally used for voice transmission.

incandescence—The state of a body with such a high temperature that it gives off light.

incandescent lamp — An electric lamp in which electric current flowing through a filament of resistance material heats the filament until it glows.

INCH—Acronym for INtegrated CHopper. It is a device designed to operate as a chopper, commutator, modulator, demodulator, or mixer, depending on circuit requirements.

inching—*See* Jogging.

incidence angle—The angle between an approaching light ray or emission and the perpendicular (normal) to the surface in the path of the ray.

incidental f-m—Also called residual f-m. The short-term jitter or undesired f-m deviation of a local oscillator. It limits resolution when it approaches the i-f bandwidth in magnitude.

incident field intensity—The field strength of a down-coming sky wave, not including the effects of earth reflections at the receiving location.

incident wave—In a medium of certain propagation characteristics, a wave which strikes a gap in the medium or strikes a medium having different propagation characteristics.

incipient failure — A degradation failure in its beginning stages.

inclination—The angle which a line, surface, or vector makes with the horizontal.

inclinometer—An instrument for measuring the magnetic inclination of the earth's magnetic field. It uses a magnetic needle that pivots vertically to indicate the inclination.

incoherent detection—Detection wherein the information contained in the phase of the carrier is discarded.

incoherent scattering — The disordered change in their direction of propagation when radio waves encounter matter.

incomplete sequencer relay — A device which returns the equipment to the normal, or off, position and locks it out if the normal starting, operating, or stopping sequence is not properly completed within a predetermined time.

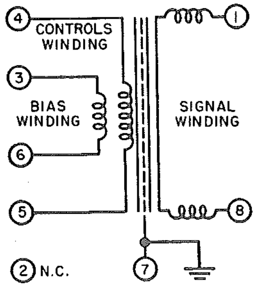

Increductor.

increductor—A controllable inductor similar to a saturable reactor, except that it is capable of operating at high frequencies (e.g., up to 400 MHz).

increment—A small change in value.

incremental hysteresis loss — Losses in a magnetic material that has been subjected to a pulsating magnetizing force.

incremental induction—One half the algebraic difference between the maximum and minimum magnetic induction at a point in a material which has been subjected simultaneously to a polarizing and a varying magnetizing force.

incremental permeability — Ratio of the cyclic change in magnetic induction to the corresponding cyclic change in magnetizing force when the mean induction is other than zero.

incremental sensitivity—The smallest change that can be detected by a particular instrument in a quantity under observation.

incremental tuner—A television tuner in which antenna, r-f-amplifier and r-f-oscillator inductors are continuous or in small sections connected in series. Rotary switches, connected to taps on the inductors, provide the portion of total inductance required for a channel, or short-circuit all remaining inductance except that required for the channel.

independent failure—A failure which has no significant relationship to other failures in a given device and can occur without interaction with other component parts in the equipment.

independent load contacts—Contacts which can control electrical loads that must be isolated from the timer clutch solenoid and motor circuit.

independent variable—One of several voltages and currents chosen arbitrarily and considered to vary independently.

index counter—An odometer-type counter which shows the amount of tape that has been unreeled on a recorder. In this way, the operator is able to note the location of a particular selection on the tape.

indexing—In a computer, a technique of address modification that is often implemented by means of index registers.

indexing slots—*See* Polarizing Slots.

index of cooperation—In facsimile, the product of the diameter of the scanning drum multiplied by the number of scanning lines per inch.

index of modulation—The modulation factor.

index of refraction—Ratio of the speeds of light or other radiation in two different materials. This determines the amount the ray will be refracted or bent when passing from one material to the other, such as from air to water.

index register — In a computer, a register that holds a quantity which may be used for modifying addresses or for other purposes, as directed by the program.

indicating demand meter—A meter equipped with a scale over which a pointer is advanced to indicate maximum demand.

indicating fuse—A protective device placed in a telephone circuit to provide visual and audible indication of a fault in the line. It consists of a fuse, pilot lamp, relay, and buzzer. When a line fault blows the fuse, the lamp lights and the buzzer sounds.

indicating instrument—An instrument which visually indicates only the present value of the quantity being measured.

indicating lamp—A lamp which indicates the position of a device or the condition of a circuit.

indicating meter — A meter which gives a visual indication of only the present or short-time average value of the measured quantity.

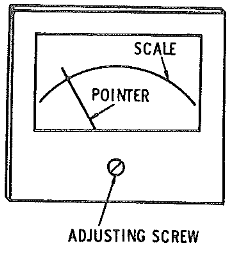

Indicating meter.

indication—The display to the human senses of information concerning a quantity being measured.

indicator—An instrument that makes information available, but does not store it.

indicator gate—A rectangular voltage waveform applied to the grid or cathode circuit of an indicator cathode-ray tube to sensitize or desensitize the tube during the desired portion of the operating cycle.

indicator tube—An electron-beam tube which conveys useful information by the variation in cross section of the beam at a luminescent target.

indirect-acting recording instrument — An instrument in which the marking device is actuated by raising the level of measurement energy of the primary detector. This is done mechanically, electrically, electronically, photoelectrically, or by some other intermediate means.

indirect addressing—A method of computer cross reference in which one memory location indicates where the correct address of the main fact can be found.

indirect light—Light from an object which has no self-luminous properties. Instead, it reflects light from another source.

indirectly controlled variable — A variable that is related to and influenced by the directly controlled variable but is not directly measured for control.

indirectly heated cathode—Also called equipotential or unipotential cathode. A cathode which is heated by an independent heater.

indirectly heated thermistor—A thermistor which incorporates, as part of its composite structure, an electrical heater. A thermistor the body temperature of which in use is significantly higher than the temperature of its surrounding medium as a result of current passing through its heater.

indirect piezoelectricity — The production of a mechanical strain in a crystal by applying a voltage to it (as opposed to the more common piezoelectric effect of applying a strain to the crystal in order to produce a voltage).

indirect scanning — A television technique used in early mechanical systems, and today in the flying-spot scanning of films. A small beam of light is moved across the subject and then reflected to a battery of phototubes.

indirect wave—A wave reaching a given reception point by a path from the transmitting point other than the direct-line path between the two (e.g., a sky wave received after deflection from the ionosphere layers).

individual gap azimuth—In a magnetic-tape record or reproduce head stack, the angle of an individual gap relative to a line perpendicular to the precision milled mounting pads in a plane parallel to the surface of the tape.

indoor antenna—Any receiving antenna located inside a building but outside the receiver.

indoor transformer—A transformer which must be protected from the weather.

induced—Produced by the influence of an electric or magnetic field.

induced charge—An electrostatic charge produced in one object by the electric field surrounding a nearby object.

induced current—1. The current that flows in a conductor which is moved perpendicularly to a magnetic field, or which is subjected to a magnetic field of varying intensity. The former takes place in an induction-motor rotor; the latter, in the secondary winding of a transformer. 2. In induction heating, the current that flows in a conductor when a varying electromagnetic field is applied.

induced electromotive force — Represented by E and is proportional to the rate of change of magnetic flux through the circuit ($d\phi/dt$).

induced environment — The temperatures, vibrations, shocks, accelerations, pressures, and other conditions imposed on a system due to the operation or handling of the system.

induced failure — A failure that is basically caused by a condition or phenomenon external to the item that fails.

induced voltage—The voltage produced in a conductor when the conductor is moved up and down through the magnetic field of a second conductor, or when the field varies in intensity and cuts across the first conductor. Even though there is no mechanical coupling between the two conductors, the one producing the field will produce a voltage in the other.

inductance—In a circuit, the property which opposes any change in the existing current. Inductance is present only when the current is changing.

inductance bridge—An instrument, similar to a Wheatstone bridge, for measuring an unknown inductance by comparing it with a known inductance.

inductance coil—*See* Inductor.

inductance-tube modulation—A method of modulation employed in frequency-modulated transmitters. An oscillator control tube acts as a variable inductance in parallel with the tank circuit of the radio-frequency oscillator tube. As a result, the oscillator frequency varies in step with the audio-frequency voltage applied to the grid of the oscillator control tube.

induction—1. The establishment of an electric charge of a magnetic field in a substance by the proximity of an electrified source, a magnet, or a magnetic field. 2. The setup of an electromotive force and current in a conductor by variation of the magnetic field affecting the conductor.

induction brazing — The electric brazing process in which heat is produced by an induced current.

induction coil—1. A device for changing direct current into high-voltage alternating current. Its primary coil contains relatively few turns of heavy wire; and its secondary coil, wound over the primary, contains many turns of fine wire. Interruption of the direct current in the primary by a vibrating-contact arrangement induces a high voltage in the secondary.

induction compass—A compass in which the indications are produced by the current generated in a coil revolving in the magnetic field of the earth.

induction-conduction heater—A heating device through which electric current is conducted but is restricted by induction to a preferred path.

induction density—*See* Flux Density.

induction factor—In an alternating current circuit, the ratio between that element of the current which does no work and the total strength of the current.

induction field—1. That portion of the electromagnetic field of a transmitting antenna which acts as if it were permanently associated with the antenna, and into which energy is alternately stored and removed. 2. The electromagnetic field of a coil carrying alternating current, responsible for the voltage induced by that coil in itself or in a nearby coil.

induction frequency converter—A slip-ring induction machine driven by an external source of mechanical power. Its primary circuits are connected to a source of electric energy having a fixed frequency. The energy delivered by its secondary circuits is proportionate in frequency to the relative speed of the primary magnetic field and the secondary member.

induction furnace—A furnace heated by electromagnetic induction.

induction hardening—The process of hardening the surface of a casting by heating it above the transformation range by electrical induction, followed by rapid cooling.

induction heating—The method of producing heat by subjecting a material to a variable electromagnetic field. Internal losses in the material then cause it to heat up.

induction instrument—An instrument operated by the reaction between the magnetic flux set up by one or more currents in fixed windings and the currents set up by electromagnetic induction in movable conductive parts.

induction motor — An alternating-current motor in which the primary winding (usually the stator) is connected to the power source and induces a current into a polyphase secondary or squirrel-cage secondary winding (usually the rotor).

induction-motor meter—A meter containing a rotor that moves in reaction to a magnetic field and the currents induced into it.

induction noise — The noise—other than thump, flutter, cross fire or cross talk—produced when two circuits are inductively coupled together.

induction-resistance welding—Welding in which electromagnetic induction alone causes the heating current to flow in the parts being welded.

induction-ring heater—A core-type induction heater adapted principally for heating round objects. The core is open or can be taken off to facilitate linking the charge.

induction speaker—A speaker in which the current that reacts with the steady magnetic field is induced into the moving member.

induction-voltage regulator—A device having a primary winding in shunt, and a secondary winding in series with a circuit for gradually adjusting the voltage or the phase relation of the circuit by changing the relative position of the exciting and series windings of the regulator.

inductive—Pertaining to inductance or to the inducing of a voltage through mutual or electrostatic induction.

inductive circuit—A circuit with more inductive than capacitive reactance.

inductive coupling—The association of two or more circuits solely by the mutual inductance between them.

inductive feedback—1. The transfer of energy from the plate circuit to the grid circuit of a vacuum tube by induction. 2. The transfer of energy from the output circuit to the input circuit of an amplifying device through an inductor or inductive coupling.

inductive interference — Interference produced in communication systems by induced voltages within the system.

inductive kick — The voltage, many times higher than the impressed voltage, produced by the collapsing field in a coil when the current through it is abruptly cut off.

inductive load—Also called lagging load. A load that is predominantly inductive, so that the alternating load current lags behind the alternating voltage of the load.

inductively coupled circuit—A network with two meshes having only mutual inductance in common.

inductive microphone—*See* Inductor Microphone.

inductive neutralization—Also called shunt or coil neutralization. A method of neutralizing an amplifier, whereby the equal and opposite susceptance of an inductor cancels the feedback susceptance caused by interelement capacitance.

inductive pickup—Signals generated in a circuit or conductor due to mutual inductance between it and a disturbing source.

inductive reactance—The opposition to the flow of alternating or pulsating current by the inductance of a circuit. It is measured in ohms, and its symbol is X_L. It is equal to 2π times the frequency in hertz times the inductance in henrys.

inductive transducer—A transducer in which changes in inductance convey the stimulus information.

inductive transduction—The conversion of the measurand into a change in the self-inductance of a single coil.

inductive tuning—A method of tuning a radio by moving a core into and out of a coil to vary the inductance.

inductive winding—A coil through which a varying current is sent to give it an inductance.

inductor—Also called inductance or retardation coil. A conductor used for introducing inductance into an electric circuit. The conductor is wound into a spiral, or coil, to increase its inductive intensity.

inductor microphone—Also called inductive microphone. A microphone in which the sound waves move a conductor back and forth, cutting magnetic lines of force and producing an electrical output of the same frequency and proportional to the amplitude of the sound waves.

inductor-type synchronous motor — A type of synchronous motor having field magnets that are fixed in magnetic position relative to the armature conductors, the torques being produced by forces between the stationary poles and salient rotor teeth. Such motors usually have permanent-magnet field excitation, are built in fractional-horsepower frames, and operate at low speeds (300 revolutions per minute or less).

inductosyn—An extremely precise transducer based upon the magnetic circuit of a conductor deposited on glass for stability, and operated at a relatively high frequency. Extremely accurate, but requires much auxiliary equipment (Farrand Controls Inc.).

industrial radio services — Radiocommunication services essential to, operated by, and for the sole use of those enterprises which require radiocommunications in order to function efficiently.

industrial television — Television used for remote viewing of manufacturing or assembling processes, usually over cables rather than through the air.

industrial timer—A timing device, impulse or constant-speed type, used in industrial applications other than the appliance industry.

industrial tube—A vacuum tube designed for industrial electronic equipment.

inelastic collision—Collision resulting in excitation of a molecule.

inertance—Acoustical equivalent of inductance.

inert gas—*See* Noble Gas.

inertia—The tendency of an object at rest to remain at rest, or of a moving object to continue moving in the same direction and at the same speed, unless disturbed by an outside force.

inertial guidance—The automatic guidance system used in missiles and satellites to make them follow a predetermined trajectory without assistance from the ground. Gyroscopes, accelerometers, and possibly a gyro-stabilized platform are used.

inertial navigation—A guidance technique in which air-frame acceleration is first measured and then integrated twice with respect to time in order to determine the distance traveled. External aids such as radio and radar are not necessary. The acceleration or deceleration of the air frame is measured continuously with accelerometers oriented in some convenient frame of reference, usually corresponding to the earth's north-south, east-west coordinates.

inertia relay—A relay having added weights or other modifications that increase its moment of inertia and either slow it or cause it to continue in motion after the energizing force is removed.

inertia switch—A switch capable of sensing acceleration, shock, or vibration.

infant mortality—The occurrence of premature catastrophic-type failures at a rate substantially greater than that observed during life prior to wearout.

infant-mortality period — *See* Early-Failure Period.

inferential—The kind of instrumentation, especially its signal source, in which there is sampling of an entirely different quantity from the one of interest, upon assuming that they vary in perfect proportion. Linearity or a perfectly repeatable relationship between the two is inferred, for the sake of a more convenient signal-source arrangement.

infinite—Boundless; having no limits whatsoever.

infinite baffle—A speaker enclosure with no openings through which sound waves can travel from the front of the cone to the back.

infinite-impedance detector—A detector circuit in which the load is a resistor connected in parallel with an r-f bypass capacitor between the cathode and ground. Since the grid is always negative with respect to the cathode, the tube presents an infinite impedance to the input.

infinite line—A transmission line with the same characteristics as an ordinary line that is infinitely long.

infinite resolution—The capability of a device to provide continuous output over its entire range.

infinitesimal — Immeasurably small; approaching zero.

infinity—1. A hypothetical amount larger than any assignable amount. 2. A number larger than any number a computer can store in any register.

inflection point—The point where a curve changes direction.

information—In computing, the basic data and/or program entered into the system.

information channel—The transmission and intervening equipment involved in the transfer of information in a given direction between two terminals. An information channel includes the modulator and demodulator and any error-control equipment irrespective of its location, as well as the backward channel, when provided.

information gate — A circuit that permits information or data pulses from passing when the circuit is triggered by an external source.

information handling — The storing and processing of information and its transmission from the source to the user. Information handling excludes the creation and use of information.

information rate—In computers, the minimum number of binary digits per second required to specify the source messages.

information rate changer — A device that speeds up the playback of tape-recorded speech without pitch change or deterioration of characteristic resonances. This is accomplished by rotating the playback head in the direction of tape travel.

information retrieval — A method for cataloging vast amounts of data related to one field of interest so that any part or all of this data can be called out at any time with accuracy and speed.

infrared—A form of electromagnetic radiation similar to visible light and radio waves. It is generated by thermal agitation and radiated by everything with a temperature above absolute zero (−273°C). High temperatures increase the thermal agitation. Hence, the hotter the object, the greater the infrared radiation. Abbreviated IR.

infrared counter-countermeasures—Action taken to employ infrared radiation equipment and systems in spite of enemy measures to counter their use.

infrared countermeasures—Action taken to reduce the effectiveness of enemy equipment employing infrared radiation.

infrared detector — A transducer which is sensitive to invisible infrared radiation (wavelength between 0.75 and 1000 microns), usually using a semiconductor, thermocouple bolometer, or pneumatic (pressure) device to detect the radiation.

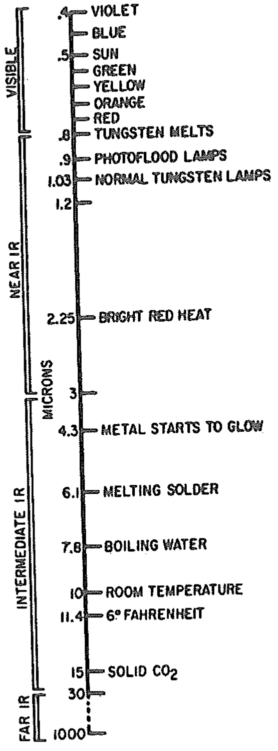

Infrared spectrum.

infrared guidance—A system using infrared heat resources for reconnaissance of targets or for navigation.

infrared homing—A type of missile homing in which the guidance system tracks the target from the infrared radiation it emits.

infrared light—Light rays just below the red end of the visible spectrum.

infrared optics — Lenses, prisms, and other optical elements for use with infrared radiation (radiation with a wavelength between 0.75 and 1000 microns).

infrared radiation — Heat radiation with wavelengths slightly longer than those of visible red (i.e., 0.75 to 1000 microns).

infrared sources—Emitters of radiation with a wavelength between 0.75 and 1000 microns.

infrared waves—Also called black light. Invisible waves longer than the longest visible red light waves but shorter than radio-frequency waves.

infrared window—A region of relatively high transmission in the infrared-frequency range.

infrasonic—Pertaining to frequencies below the range of human hearing, hence below about 15 hertz. Formerly called subsonic.

infrasonic frequency—Also called subsonic

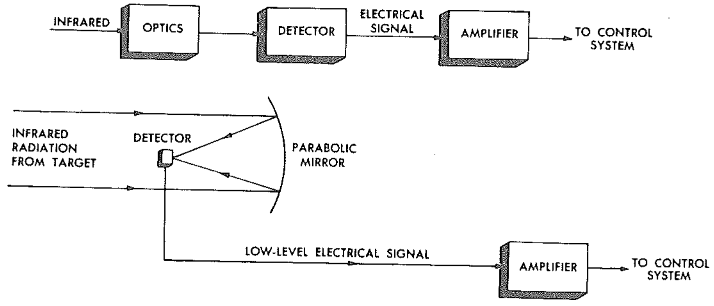

Infrared guidance.

frequency. A frequency below the audio range. Infrasonic vibrations can be felt but not heard.

inharmonic frequency—A frequency which is not a rational multiple of another frequency.

inherent delay—Delay between the insertion of information into a unit and presentation of the information at the output. Delay is usually inserted into the crt vertical amplifier of pulse analyzers to allow the leading edge of the signal triggering the sweep to be seen.

inherent interference—A type of electromagnetic interference generated within a receiver by thermal agitation, shot effect, and nonlinear impedance.

inherited error—In a computer, the error in the initial values, especially that error accumulated from prior steps in a step-by-step integration.

inhibit gate—A circuit that provides an output when the inputs differ and only for one of these combinations. This function can be implemented directly or by using any previous gate, plus an inverter.

inhibiting input—A computer gate input which, if in its prescribed state, prevents any output which might otherwise occur.

inhibition gate—A gate circuit used as a switch and placed in parallel with the circuit it is controlling.

inhibit pulse—A computer drive pulse that tends to prevent certain drive pulses from reversing the flux of a magnetic cell.

initial actuation time—The time of the first closing of a previously open contact of a relay or the first opening of a previously closed contact.

initial contact chatter—That chatter caused by vibration produced by opening or closing the contacts in a relay themselves, as by contact impact in closure.

initial differential capacitance—The differential capacitance of a nonlinear capacitor when the capacitor voltage is zero.

initial drain—The current supplied at nominal voltage by a cell or battery.

initial element—*See* Primary Detector.

initial erection—The mode of operation of a vertical gyro in which the gyro is being erected or slaved initially. The initial erection rate is usually relatively fast.

initial failure—The first failure that occurs in use.

initial inverse voltage—Of a rectifier tube, the peak-inverse anode voltage immediately following the conducting period.

initial ionizing event—Also called primary ionizing event. An ionizing event that initiates a tube count.

initializing—The preliminary steps in arranging those instructions and data in a computer memory that are not to be repeated.

initial permeability—Permeability at a field density approaching zero.

initial reversible capacitance—In a nonlinear capacitor, the reversible capacitance at a constant bias voltage of zero.

initial-velocity current—A current which flows between an electrode, such as the grid of a vacuum tube, and its cathode as a result of electrons thrown off from the cathode because of heat alone. Their velocity is sufficient to allow the electrons to reach the grid unaided by an accelerating field.

injected lasers—*See* Diode Laser.

injection grid—A vacuum-tube grid that controls the electron stream without causing interaction between the screen and control grids. In some superheterodyne receivers, the injection grid introduces the oscillator signal into the mixer stage.

injection laser—A semiconductor diode in which radiation is produced as electrons recombine with holes in the junction region. For coherent emission, the current density must exceed a certain threshold value, commonly about 10,000 amperes per square centimeter for gallium-arsenide diodes.

injector—An electrode on a spacistor.

ink-mist recording—Also called ink-vapor recording. In facsimile, electromechanical recording in which particles of an ink mist are deposited directly onto the record sheet.

ink recorder—The ink-filled pen or capillary tube that produces a graphic record.

ink recording—A type of mechanical facsimile recording in which an inked helix marks the record sheet.

ink-vapor recording—*See* Ink-Mist Recording.

inleads—Those portions of the electrodes of a device that pass through an envelope or housing.

in-line heads—*See* Stacked Heads.

in-line tuning—The method of tuning the intermediate-frequency strip of a superheterodyne receiver in which all the intermediate-frequency amplifier stages are made resonant to the same frequency.

inorganic electrolyte—A solution that conducts electricity due to the presence of ions of substances not of organic origin.

in phase—Two waves of the same frequency that pass through their maximum and minimum values of like polarity at the same instant, are said to be in phase.

in-phase portion of the chrominance signal.—That portion of the chrominance signal having the same phase as, or exactly the opposite phase from, that of the subcarrier modulated by the I-signal. This portion of the chrominance signal may lead or lag the quadrature portion by 90 electrical degrees.

input—1. The current, voltage, power, or other driving force applied to a circuit or device. 2. The terminals or other places where current, voltage, power, or driving force may be applied to a circuit or device.

input admittance—1. The reciprocal of input impedance. 2. The admittance between the input terminals with the outputs shorted together.

input area—In a computer, the area of internal storage into which data from external storage is transferred.

input block—In a computer, a section of the internal storage reserved for receiving and processing input data.

input capacitance—The sum of the direct capacitances between the control grid and cathode (and any other electrodes operated at the alternating potential of the cathode). This is not the effective capacitance, which is a function of the impedances of the associated circuits.

input common-mode rejection ratio—1. The ratio of the change in input voltage to the corresponding change in output voltage, divided by the open-loop voltage gain. 2. The ratio of the full differential voltage gain to the common-mode voltage gain.

input equipment—The equipment that introduces information into a computer.

input error voltage—The error voltage appearing across the input terminals of an operational amplifier when a feedback loop is applied around the amplifier.

input extender—A high-speed diode array used in a logic circuit when increased fan-in capability is required.

input gap—Also called buncher gap. In a microwave tube, the gap where the initial velocity modulation of the electron stream occurs.

input impedance—The impedance a transducer presents to a source.

input impedance of a transmission line—The impedance between the input terminals with the generator disconnected.

input-output devices—Computer hardware by which data is entered into a computer or by which the finished computations are recorded for immediate or future use. Abbreviated I/O.

input-power rating—Also called coil rating. A statement of the allowable voltage, current, or power to the actuating element of a relay beyond which unsatisfactory performances will occur.

input recorder—Any device which makes a record of an input electrical signal.

input register—In a computer, the register of internal storage able to accept information from outside the computer at one speed and supply the information to the computer calculating unit at another, usually much greater, speed.

input resonator—The buncher resonator in a velocity-modulated tube. It modifies the velocity of the electrons in the beam.

input transformer—A transformer that transfers energy from an alternating-voltage source to the input of a circuit or device. It usually provides the correct impedance match, too.

input uncertainty—In an operational amplifier, the algebraic sum of all the factors, including environmental and time effects, that contribute to the nonideal behavior of the input circuit.

input unit—In a computer, the unit that takes information from outside the computer into the computer.

input voltage offset—The d-c potential difference between the two inputs of a differential amplifier when the potential between the output terminals is zero.

input winding—*See* Signal Winding.

inquiry station—In a computer, a direct method of reading data from or entering data into storage.

inquiry unit—A device used to extract a quick reply to a random question regarding information in a computer storage.

inrush current—In a solenoid or coil, the steady-state current drawn from the line with the armature in its maximum open position.

insert core—An iron core used generally for adjusting an inductor to a fixed frequency. It consists of a threaded metal insert molded or cemented into one or both ends of the core.

insert earphones—Small earphones which fit partially inside the ear.

insertion gain—The gain resulting from the insertion of a transducer in a transmission system is the ratio of the power delivered to that part of the system following the transducer to the power delivered to that same part before insertion. (If more than one component is involved in the input or output, the particular component used must be specified. This ratio is usually expressed in decibels.)

insertion loss—1. The difference between the power received at the load before and after the insertion of apparatus at some point in the line. 2. Ratio in decibels of the power received at the load before and after insertion of the apparatus.

insertion phase shift—The change in phase of an electric structure when inserted into a transmission system.

inside lead—*See* Start Lead.

inside spider—A flexible device placed inside a voice coil to center it with the pole pieces of a speaker.

inspection chamber—In a spectrophotometer, the part in which the solution to be tested is placed for analysis.

inspectoscope—An instrument for viewing quartz crystals, while they are immersed in oil, to determine mechanical faults, the approximate direction of the optical axis, and regions of optical twinning.

instability—The measure of the fluctuations or irregularities in the performance of a device, system, or parameter.

instantaneous automatic gain control—In radar, an automatic gain control which instantly responds to variations in the mean clutter level and thereby reduces it.

instantaneous automatic volume control—A fast-acting automatic volume-control circuit installed in some radar receivers as an anticlutter device.

instantaneous companding — Companding which varies the effective gain in response to instantaneous values of the signal wave.

instantaneous contacts—Contacts which are actuated immediately when a starting signal is applied to a timer.

instantaneous disc—A blank recording disc that can be played back on a phonograph immediately after being cut on a recorder.

instantaneous frequency—The rate at which the angle of a wave changes when the wave is a function of time. If the angle is measured in radians, the frequency in hertz is the rate of change of the angle divided by 2π.

instantaneous overcurrent relay — Also called rate-of-rise relay. A device which functions instantaneously on an excessive value of current, or on an excessive rate of current rise, thus indicating a fault in the apparatus of the circuit being protected.

instantaneous power—The power at the points where an electric circuit enters a region. It is equal to the rate at which the circuit is transmitting electrical energy into the region.

instantaneous power output—The rate at which energy is delivered to a load at a particular instant.

instantaneous recording—A recording intended for direct reproduction without further processing.

instantaneous sampling—The process of obtaining a sequence of instantaneous values of a wave. These values are called instantaneous samples.

instantaneous sound pressure—The total instantaneous pressure at a certain point, minus the static pressure at that point. The most common unit is the microbar.

instantaneous speech power—The rate at which the speaker is radiating sound energy at any given instant.

instantaneous value—The magnitude, at any particular instant, of a varying value.

instruction—Information which, when properly coded and introduced as a unit into a digital computer, causes the computer to perform one or more of its operations. An instruction commonly includes one or more addresses.

instructional constant — Also called pseudo instruction. In a computer, data stored in the program or instructional area which will be used only as a test constant.

instruction code—The list of symbols, names, and definitions of the instructions which are intelligible to a given computer or computing system.

instruction counter—*See* Control Counter.

instruction modification — A change in the operation-code portion of a computer instruction or command such that, if the routine containing the instruction or command is repeated, the computer will perform a different operation.

instruction register — In a computer, the register that temporarily stores the instruction currently being performed by the control unit of the computer.

instrument—A device capable of measuring, recording, and/or controlling.

instrument approach—A blind landing—i.e., solely by navigational instruments, without visual reference to the terrain.

instrument-approach system — In navigation, a system furnishing vertical and horizontal guidance to aircraft during descent. Touchdown requires some other guidance.

instrumentation—The introduction and use of devices of any kind into a process for the purpose of obtaining more precise data (improving quality) or for obtaining information not otherwise available. Typically accomplished by generating, transmitting, displaying, controlling, correlating, and recording signals.

instrument chopper—A vibrating switch used for modulating, demodulating, and switching d-c or low-frequency a-c information in instrumentation. It is driven synchronously from an a-c or pulsating d-c source. The driven switching circuit is designed for low-level (0- to 10-volt) signal information.

instrument error—The inaccuracy of an instrument.

instrument flight—A blind flight—i.e., one in which the pilot controls the path and altitude of the aircraft solely by instrument.

instrument lamp—A light that illuminates or irradiates an instrument.

instrument-landing station—A special radio station for aiding in landing aircraft.

instrument-landing system — Abbreviated ILS. A radionavigation system intended to aid aircraft in landing. It provides lateral and vertical guidance, including distance from the landing point.

instrument multiplier—*See* Voltage-Range Multiplier.

instrument relay—A relay which operates on the principles employed in such electrical measuring instruments as the electrodynamometer, iron-vane, and D'Arsonval meters.

instrument shunt—An internal or external resistor connected in parallel with the circuit of an instrument to extend its current range.

instrument switch—A switch disconnecting an instrument or transferring it from one circuit or phase to another.

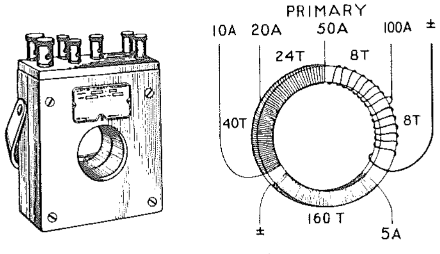

Instrument transformers.

instrument transformer — A transformer which reproduces, in its secondary circuit, the primary current (or voltage) with its phase relationship substantially preserved, suitable for utilization in measurement, control, or protective devices.

instrument-transformer correction factor —The factor by which a wattmeter reading must be multiplied to correct for the effect of the instrument-transformer ratio correction factor and phase angle.

insulated—Separated from other conducting surfaces by a nonconductive material offering a high, permanent resistance to the passage of current and disruptive discharge.

insulated carbon resistor—A carbon resistor encased in fiber, plastic, or other insulation.

insulated clip—A clip terminating in an insulated eye through which flexible cords or wires may be run and supported.

insulated enclosure—A special shielded enclosure design providing insulation against weather or providing maximum temperature stability. Usually prefabricated as an exterior building panel in modular construction.

insulated-substrate monolithic circuit—An intergrated circuit which may be either an all-diffused device or a compatible structure so constructed that the components within the silicon substrate are insulated from one another by a layer of silicon dioxide, instead of the reverse-biased PN junctions used for isolation in other techniques.

insulated wire—A conductor covered with a nonconductive material.

insulating material — A material on or through which essentially no current will flow. It is used to confine the flow of current within a conductor or to eliminate the shock hazard of a bare conductor.

insulating tape — Tape that is wrapped around joints in insulated wires or cables. It is impregnated with an insulating material and covered with adhesive on one side.

insulating varnish—A varnish applied to coils and windings to improve their insulation (and, at times, their mechanical rigidity).

insulation—A nonconductive material that prevents the leakage of electricity from a conductor, provides mechanical spacing or support, or protects against accidental contact.

insulation rating — The dielectric-strength and insulation-resistance values required to ensure satisfactory performance.

insulation resistance—1. The ratio of the voltage applied between two electrodes in contact with a specific insulator to the total current between the electrodes. 2. The ratio of the direct-current voltage applied to the terminals of a capacitor and the resulting leakage current after the initial charging current has ceased. It includes both the volume and surface resistance. Industrial specifications usually call for a certain minimum value (several thousand megohms) determined with a specific voltage applied.

insulation system—All of the insulation materials used to insulate a particular electrical or electronic product.

insulator—1. A material of such extremely low conductivity that, in effect, no current flows through it. 2. A high-resistance device that supports or separates conductors to prevent a flow of current between them or to other objects.

insulator arcing ring—A circular or oval metal part placed at one or both ends of an insulator to prevent current from arcing over and damaging it and/or the conductor.

insulator arc-over—The flow of power current over an insulator in the form of an arc following a surface discharge.

insulazing—*See* Surface Insulation.

insulectrics — The science encompassing insulating materials in electrical insulation.

integer—A whole number, not fractional or mixed (e.g., 1, 2, 9, 100, etc.—not 1.1 or ½).

integral-cavity, reflex-klystron oscillator—A reflex-klystron oscillator in which tuning is accomplished by changing the physical dimensions of the resonant cavity. It is usually referred to as a diaphragm- or grid-gap-tuned klystron, since a flexible diaphragm is used to change the cavity dimension, i.e., the gap between the cavity grids.

integral circuit packages—Microcircuits assembled from discrete components and all circuits created essentially in an active or passive substrate.

integral-external-cavity reflex oscillator—A reflex-klystron oscillator in which a fixed internal cavity is tightly coupled to a permanently attached external cavity. Tuning is achieved by varying a reactance probe in the external cavity.

integral-horsepower motor—A motor which is built into a frame and has a continuous rating of one horsepower.

integral resistor—An internal or external resistor preconnected to the electrical element and forming an integral part of the cup assembly to provide a desired electrical characteristic of a precision potentiometer.

integrated amplifier—An audio unit containing both a preamplifier and a power amplifier on one chassis.

integrated circuit—A combination of interconnected circuit elements inseparably associated on or within a continuous substrate. (*Also see* Monolithic Integrated Circuit, Multichip Integrated Circuit, Film Integrated Circuit *and* Hybrid Integrated Circuit.) To further define the nature of an integrated circuit, additional modifiers may be prefixed. Examples are: dielectric isolated monolithic integrated circuit, beam-lead monolithic integrated circuit, or silicon chip-tantalum thin-film hybrid integrated circuit.

integrated-circuit array—Multiple integrated circuits formed on a common substrate and electrically interconnected during fabrication.

integrated component—A number of electrical elements comprising a single structure which cannot be divided without destroying its stated electronic function.

integrated data processing—A method of transforming disjointed and repetitive paperwork tasks into a correlated and mechanized production of information for any purpose.

integrated electronics—That portion of electronic art and technology in which the interdependence of material, device, circuit, and system-design considerations is especially significant; more specifically, that portion of the art dealing with integrated circuits.

integrated equipment components—Integrated-circuit chips which contain a complete logic function. Abbreviated IEC's.

integrated morphology—The structural characterization of an electronic component in which the identity of the current or signal modifying areas, patterns, or volumes has become lost in the integration of electronic materials, in contrast to an assembly of devices performing the same function.

integrating circuit—*See* Integrator.

integrating meter—A meter which adds up (integrates) the electrical energy used over a period of time. An ordinary electric watt-hour meter is an example.

integrating photometer—A photometer which, with a single reading, indicates the average candlepower from a source in all directions or at all angles in a single plane.

integrating relay—A relay that sums up the inputs of voltage or current supplied to it and opens or closes its contacts in response to the input so integrated.

integrating-sphere densitometer—A photoelectric instrument that measures the density of motion-picture film or its sound track.

integrator (in electronic computers)—1. A device with an output proportionate to the integral of the input signal. 2. In certain digital machines, a device that numerically approximates the mathematical process of integration. 3. Also called an integrating network or circuit. A transducer with an output waveform that is the time integral of its input waveform.

intelligence bandwidth—The total audio (or video)frequency bandwidths of one or more channels.

intelligence sample—Part of a signal taken as evidence of the quality of the whole.

intelligence signal—Any signal which conveys information (e.g., voice, music, code, or television).

intelligibility—*See* Articulation.

intensifier electrode—Also called post-accelerating electrode. In some types of electrostatic cathode-ray tubes, an electrode which permits additional acceleration of the electron beam after it has been deflected. This electrode permits greater intensity of the trace without materially reducing the deflection sensitivity of the tube.

intensifying screen—A thin fluorescent screen placed next to a photographic plate to increase the effect of radiation on the plate.

intensitometer—Also called a dosage meter or dosimeter. An instrument that estimates the amount of X-ray radiation, for determining the duration of exposure during X-ray pictures or therapy.

intensity—1. The strength of a quantity. 2. The relative strength, or amplitude, of electric, magnetic, or vibrational energy. 3. The brilliance of an image on the screen of a cathode-ray tube.

intensity control—Used with cathode-ray tubes to control the intensity of the electron beam and hence the amount of light generated by the fluorescent screen. Generally, the grid bias of the tube is regulated.

intensity level—Ratio of the intensity of the sound to a reference intensity of a free plane wave of 1 microwatt per square centimeter under normal conditions. Commonly expressed in decibels.

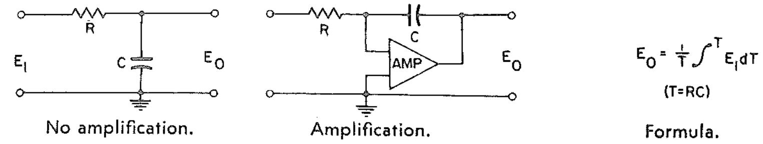

No amplification. Amplification. Formula.

Integrating circuits.

intensity modulation—The process and/or effect of varying the electron-beam current in a cathode-ray tube, resulting in varying brightness or luminance of the trace. *Also see* Z-Axis Modulation.

intensity of radiation—The radiant energy emitted in a specified direction per unit time, per unit area of surface, per unit solid angle.

interaction—The effects two or more parts, components, etc., have on each other while each is performing a function.

interaction cross talk—Cross talk resulting from mutual coupling between two paths by means of a third path. For example, if a signal on pair 1 is coupled to pair 2, and then coupled from pair 2 to pair 3, where it is measured it is known as interaction cross talk.

interaction loss (of a transducer)—Expressed in decibels, it equals 20 times the logarithm (to the base *10*) of the scalar value of the reciprocal of the interaction factor.

interaction space—In an electronic tube, the region where the electrons interact with an alternating electromagnetic field.

interaxis error—The deviation from 90° perpendicularity of one set of resolver windings when excitation is applied to one of the other windings. For rotor interaxis error one stator winding is excited; for stator interaxis error one rotor winding is excited.

interbase current—In a junction tetrode transistor, the current that flows from one base connection to the other through the base region.

interblock space—*See* Inter-Record Gap.

intercarrier noise suppression—The means of suppressing the noise resulting from increased gain when a high-gain receiver with automatic volume control is tuned between stations. The suppression circuit automatically blocks the audio-frequency output of the receiver when there is no signal at the second detector.

intercarrier sound system—A television receiving system in which use of the picture carrier and the associated sound-channel carrier produces an intermediate frequency equal to the difference between the two carrier frequencies. This intermediate frequency is frequency-modulated in accordance with the sound signal.

intercellular massage—The ultrasonic stimulation of body cells. Sometimes called micromassage.

intercept receiver—Also called search receiver. A specially calibrated receiver which can be tuned over a wide frequency range in order to detect and measure enemy r-f signals.

intercept service—In a telephone system, a service provided to subscribers whereby calls to disconnected stations or dead lines are either routed to an intercept operator for explanation or the calling party receives a distinctive tone signal to indicate that he has made such a call.

intercharacter space—In telegraphy, the space between characters of a word. It is equal to three unit lengths.

intercom—*See* Intercommunication System.

intercommunicating system—Also called intercom. A two-way communication system without a central switchboard, usually limited to a single vehicle, building, or plant area. Stations may or may not be equipped to originate a call, but can answer any call.

interconnecting wire—Wires used for connections between subassemblies, panels, chassis, and remotely mounted devices. Does not necessarily apply to the internal connections of these units.

interconnection—1. Also called tie line. A transmission line connecting two electric systems or networks and permitting energy to be transferred in either direction. Larger interconnections are often called interties, giant ties, or regional interconnections. 2. Also called intraconnection. The physical wiring between components (outside a module), between modules, between units, or between larger portions of a system or systems.

interconnection diagram—Diagram showing the identity of all units in a piece of electronic equipment and the connections between them.

interconnections (microelectronic)—Those conductors and connections which are not in continuous integral contact with the substrate or circuit elements of an integrated circuit.

interconnection system—The electrical and mechanical interconnection of any one or all of the six levels of interconnections generally common to electronic equipment. The six levels of interconnection are: intramodule, module to motherboard, intramotherboard, motherboard to back panel, back-panel wiring, and input/output.

interdigital magnetron—A magnetron with anode segments around the cathode. Alternate segments are connected together at one end, and remaining segments at the opposite end.

interelectrode capacitance—The capacitance between certain electrodes of an electron tube.

interelectrode coupling—Capacitive feedback from the plate of a tube to the grid. In triodes, this limits the maximum amplification possible without starting oscillation.

interelectrode leakage—The undesired current which flows between elements not normally connected in any way.

interelectrode transit time—The time required for an electron to travel between two electrodes.

interelement capacitance—The capacitance caused by the PN junctions between the regions of a transistor and measured between the external leads of the transistor.

interexchange channel—A channel connecting two different exchange areas.

interface—The junction or the points within or between systems or subsystems where matching or accommodation must be properly achieved in order to make their operation compatible with the successful operation of all other functional entities.

interface resistance—*See* Cathode Interface.

interfacial connection—In a printed-circuit board, a conductor that connects conductive patterns on opposite faces of the base.

interference — Any electrical or electromagnetic disturbance, phenomenon, signal, or emission, man-made or natural, which causes or can cause undesired response, malfunctioning, or degradation of the electrical performance of electrical and electronic equipment.

interference eliminator—A device designed for the purpose of reducing or eliminating interference.

interference fading — Fading produced by different wave components traveling slightly different paths in arriving at the receiver.

interference filter—A device added between a source of man-made interference and a radio receiver to attenuate or eliminate noise signals. It generally contains a combination of capacitance and inductance.

interference guard band—*See* Guard Band.

interference pattern—1. The resultant space distribution of pressure, particle velocity, or energy flux when progressive waves of the same frequency and kind are superimposed. 2. The pattern produced on a radar scope by interference signals.

interference-rejection unit — A tunable filter or wave trap capable of being adjusted to reject any frequency within the i-f passband of a receiver while allowing the remainder of the passband curve to remain intact. It is adjusted to reject an interference signal and thus constitutes a form of antijamming. Abbreviated IFRU.

interference spectrum—The frequency distribution of the jamming interference.

interferometer—An apparatus that shows interference between two or more wave trains coming from the same luminous area, and also compares wavelengths with observable displacements of reflectors or other parts.

interferometer homing—A homing guidance system in which the direction of the target is determined by comparing the phase of the echo signal as received at more than one antenna.

interior label—In a computer, a magnetically recorded sequence added to a tape to identify the contents.

interior-wiring-system ground—The ground connection to one of the current-carrying conductors of an interior wiring system.

interlace—1. In a computer, to assign successive storage location numbers to physically separated storage locations on a magnetic drum. This serves to reduce access time. 2. To transmit different interrogation modes on successive sweeps.

interlaced scanning—Also called line interlace. A system of scanning whereby the odd- and even-numbered lines of a picture are transmitted consecutively as two separate fields. These are superimposed to create one frame, or complete picture, at the receiver. The effect is to double the apparent number of pictures and thus reduce flicker.

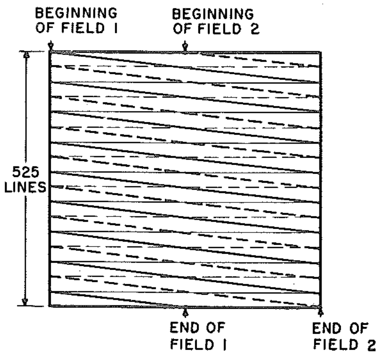

Interlaced scanning.

interlace factor—A measure of the degree of interlace of normally interlaced fields.

interlacing—A method of scanning used in television, in which each picture is divided into two or more complete sets of interlacing lines to reduce flicker.

interleave — 1. In a computer, to insert segments of one program into another program so that the two programs can be executed essentially simultaneously. 2. *See* interlace.

interleaving—Placing between. For example, in the transmission of a composite color signal, the bands of energy of the chrominance signal are interleaved with, or placed between, those of the luminance signal.

interlock—A device actuated by the operation of some other device with which it is directly associated, to govern succeeding operations of the same or allied devices. Interlocks may be either electrical or mechanical.

interlocking—The forcing of a voltage of one frequency to be in step with a voltage of another frequency.

interlock relay—1. A relay in which one armature cannot move or its coil be energized unless the other armature is in a certain position. 2. *See* Checking Relay.

intermediate frequency—Abbreviated if. A frequency to which a signal wave is shifted locally as an intermediate step in transmission or reception.

intermediate-frequency amplifier—1. An amplifier tuned to a fixed frequency, or capable of single-control tuning over a range of frequencies, for the purpose of selecting one of the frequency components generated in a mixer circuit. 2. The section of a superheterodyne receiver which amplifies signals with high efficiency at a predetermined frequency called the intermediate frequency of the receiver.

intermediate-frequency harmonic interference—Interference caused in superheterodyne receivers by the radio-frequency circuit accepting harmonics of the intermediate-frequency signal.

intermediate-frequency interference ratio — *See* Intermediate-Frequency Response Ratio.

intermediate-frequency response ratio — Also called intermediate-frequency interference ratio. In a heterodyne receiver, the ratio of intermediate-frequency signal input at the antenna to the desired signal input for identical outputs.

intermediate-frequency strip—A subassembly containing the intermediate-frequency stages in a receiver.

intermediate-frequency transformer — A transformer designed for use in the intermediate-frequency amplifier of a superheterodyne receiver.

intermediate-frequency-transformer-lead color code — Transformer leads in many radio receivers are identified by the following standard EIA colors:

Blue—plate. Green—grid or diode.
Red—B+. Black—grid return.
Green-black—second diode (full-wave transformers only).

intermediate means—All system elements needed to perform distinct operations in the measurement sequence between the primary detector and the end device.

intermediate-range ballistic missile — A tactical missile or rocket weapon with a range of 200 to 1,500 miles.

intermediate repeater — A repeater used other than at the end of a trunk or line.

intermediate state—The partial superconductivity that occurs when a magnetic field of appropriate strength is applied to a sphere of material below its critical temperature (i.e., the temperature below which the material superconducts if no magnetic field were present).

intermediate subcarrier—A carrier used for modulating a carrier or another intermediate subcarrier. It also may have been modulated by one or more subcarriers.

intermediate switching region—In a relay, an area between low level (including dry circuit) and power switching (including full rated load) where the contact arc does not destroy deposits which are by-products of the switching function.

intermittent—Occurring at intervals.

intermittent current—A unidirectional current that is interrupted at intervals.

intermittent defect—A defect that depends on variable conditions in a circuit. Hence it is not present at all times.

intermittent duty—Operation for specified alternate intervals of load and no-load; load and rest; or load, no-load, and rest.

intermittent-duty rating—The output rating of a device operated for specified intervals, rather than continuously.

intermittent-duty relay — A relay which must be de-energized at intervals to avoid excessive temperature, or a relay that is energized at regular or irregular intervals, as in pulsing.

intermittent pulsing — The transmission of short bursts of radiation at irregular intervals.

intermittent reception—A defect where the receiver operates normally for a while, at regular or irregular intervals.

intermittent scanning—One or two 360° scans of an antenna beam at irregular intervals to make detection by intercept receivers more difficult.

intermittent-service area—An area still receiving the ground wave of a broadcast station, but subject to interference and fading.

intermodulation—1. In a nonlinear transducer element, the production of frequencies corresponding to the sums and differences of the fundamentals and harmonics of two or more frequencies transmitted through the transducer. 2. An error form that occurs in chopper-stabilized amplifiers when a beat component forms between the chopper drive frequency and normal signals which have frequencies near that of the chopper.

intermodulation distortion — Nonlinearity characterized by the appearance of frequencies in the output equal to the sums and differences of integral multiples of the component frequencies present in the input signal. (Harmonics are usually not included.)

intermodulation frequencies—The sum and difference frequencies generated in a nonlinear element.

intermodulation interference—The combination-frequency tones produced at the output by a nonlinear amplifier or network when two or more sinusoidal voltages are applied at the input. Generally expressed as the ratio of the root-mean-square voltage of one or more combination frequencies to that of one of the parent frequencies measured at the output.

intermodulation noise—Noise introduced in the channel of interest by signals being transmitted in other channels.

internal arithmetic—Any computations performed by the arithmetic unit of a computer, as distinguished from those performed by peripheral equipment.

internal calibration—Calibration by an internal voltage source (provided with the instrument) rather than an external standard.

internal connection—Abbreviated IC. In a vacuum tube, a base-pin connection designed not to be used for any circuit connections.

internal correction voltage — The voltage added to the composite controlling voltage of an electron tube. It is the voltage equivalent of those effects produced by initial electron velocity, contact potential, etc.

internal graticule—A graticule the rulings of which are a permanent part of the inner surface of the cathode-ray tube faceplate.

internal input impedance—The actual impedance at the input terminals of a device.

internally caused contact chatter — That chatter resulting from the operation or release of the relay. It may be classified as initial, armature-impact, armature-rebound, or armature-hesitation chatter.

internally stored program — A sequence of instructions (program) stored inside a computer in the same storage facilities as the computer data, as opposed to being stored externally on punched paper tape, pin boards, etc.

internal magnetic recording—Storage of information within the material itself, such as used in magnetic cores.

internal memory—Also called internal storage. The total memory or storage which is automatically accessible to a computer. It is an integral physical part of the computer and is directly controlled by it.

internal output impedance—The actual impedance at the output terminals of a device.

internal resistance—The inherent resistance of a battery, generator, or circuit component.

internal storage — 1. Storage facilities in a computer forming an integral physical part of and directly controlled by the computer. 2. The total storage automatically available to the computer.

international broadcast station—A station licensed for transmission of broadcast programs for international public reception. By international agreement, frequencies are allocated between 6,000 and 26,600 kHz.

international call sign—The identifying letters and numbers assigned to a radio station in accordance with the International Telecommunications Union. The first character, or the first two, identify the nationality of the station.

international communication service — A telecommunication service between offices or stations (including mobile) belonging to different countries.

international coulomb — The quantity of electricity passing any section of an electric circuit in one second when the current is one international ampere. One international coulomb equals 0.99985 absolute coulomb.

international farad—The capacitance of a capacitor when a charge of one international coulomb produces a potential difference of one international volt between the terminals. One international farad equals 0.99952 absolute farad.

international henry—The inductance which produces an electromotive force of one international volt when the current is changing at a rate of one international ampere per second. One international henry equals 1.00018 absolute henrys.

international joule—The energy required to transfer one international coulomb between two points having a potential difference of one international volt. One international joule equals 1.00018 absolute joules.

International Morse code—Also called Continental code. A system of dot-and-dash signals used chiefly for international radio and wire telegraphy. It differs from American Morse code in certain code combinations only.

international ohm—The resistance at 0°C. of a column of mercury of uniform cross section 106.300 centimeters in length and with a mass of 14.4521 grams. One international ohm equals 1.00048 absolute ohms.

International Radio Consultative Committee — Abbreviated CCIR. An international committee which studies technical operating and tariff questions pertaining to radio, broadcast television, and multichannel video transmissions and issues recommendations. It reports to the International Telecommunications Union.

international radio silence — Three-minute periods of radio silence, commencing 15 and 45 minutes after each hour, on a frequency of 500 kHz only. During this time all radio stations are supposed to listen on that frequency for distress signals of ships and aircraft.

international radium standard—A standard of radioactivity, consisting of 21.99 milligrams of pure radium chloride.

international system (of electrical and magnetic units)—A system for measuring electrical and magnetic quantities by using four fundamental quantities. Resistance and current are arbitrary values that correspond approximately to the absolute ohm and the absolute ampere. Length and time are arbitrarily called centimeter and second. The international system of electrical units was used between 1893 and 1947. By international agreement, it was discarded on January 1, 1948, in favor of the mksa (Giorgi) system.

International Telecommunications Union—The international counterpart of the Federal Communications Commission. It provides standardized communications procedures and practices, including frequency allocations and radio regulations, on a world-wide basis.

International Telegraph Consultative Committee — Abbreviated CCIT. An international committee responsible for studying technical operating and tariff questions pertaining to telegraph and facsimile and issuing recommendations. It reports to the International Telecommunications Union.

International Telephone Consultative Committee — Abbreviated CCIF. An international committee responsible for studying and issuing recommendations regarding technical operations and tariff questions pertaining to ordinary telephones; carrier telephones; and music, picture, television, and multichannel telegraph transmission over wire line. It reports to the International Telecommunications Union.

international temperature scale — A temperature scale adopted in 1948 by international agreement. Between the boiling point of oxygen (−182.97°C) and 630.5°C it is based upon the platinum resistance thermometer. From 630.5°C to 1063.0°C it is based on the platinum rhodium thermocouple, and above 1063.0°C on the optical pyrometer.

international volt—The voltage that will produce a current of one international ampere through a resistance of one international ohm. One international volt equals 1.00033 absolute volts.

international watt—The power expended when one international ampere flows between two points having a potential difference of one international volt. One international watt equals 1.00018 absolute watts.

interphase transformer — An autotransformer or a set of mutually coupled reactors used with three-phase rectifier transformers to modify current relationships in the rectifier system and thereby cause a greater number of rectifier tubes to carry current at any instant.

interphone system—An intercommunication system like that in an aircraft or other mobile unit.

interpolation—The process of finding a value between two known values on a chart or graph.

interpole—A small auxiliary pole placed between the main poles of a direct-current generator or motor to reduce sparking at the commutator.

interpreter — A punch-card machine which will read the information conveyed by holes punched in a card and print its translation in characters arranged in specified rows and columns on the card.

interpreter code — A computer code which an interpretive routine can use.

interpretive programming — The writing of computer programs in a "pretend" machine language, which the computer precisely converts into actual machine-language instructions before performing them.

interpretive routine—Computer routine designed to transfer each pseudocode and, using function digits, to set a branch order that links the appropriate subroutine into the main program.

interrecord gap — Also called interblock space. The space between records on magnetic tape caused by delays involved in starting and stopping the tape motion. This gap is used to signal that the end of a record has been reached.

interrogation—The triggering of one or more transponders by transmitting a radio signal or combination of signals.

interrogation signal—A pulsed or c-w signal emitted to initiate a reply signal from a transponder or responder.

interrogator—Also called challenger. A radio transmitter used to trigger a transponder.

interrogator-responser—A combined radio transmitter and receiver for interrogating a transponder and displaying the replies.

interrupt—In a computer, a break in the normal flow of a system or routine such that the flow can be resumed from that point at a later time. An interrupt is usually caused by a signal from an external source.

interrupted continuous waves—Continuous waves that are interrupted at an audio-frequency rate.

interrupter—A magnetically operated device used for rapidly and periodically opening and closing an electric circuit in doorbells and buzzers and in the primary circuit of a transformers supplied from a d-c source.

interrupter contacts — On a stepping relay, an additional set of contacts operated directly by the armature.

interrupting capacity—The highest current at which the device can interrupt at the rated voltage.

interrupting rating — Conditions under which the contact of a relay must interrupt with a prescribed duty cycle and contact life.

interrupting time—In a circuit breaker, the interval between the energizing of the trip coil and the interruption of the circuit, at the rated voltage.

interstage—Between stages.

interstage coupling — Coupling between stages.

interstage punching — A system in which only the odd-numbered rows are punched in the British standard card.

interstage transformer—A transformer that couples two stages together.

interstation noise suppression—Canceling of the noise which occurs when a high-gain radio receiver with automatic volume control is tuned between stations.

interties—*See* Interconnection.

interval calibration—*See* Step Calibration.

interval circuit—A circuit which is energized during timing only. This can be accomplished by using a timer with interval contacts; or by using a timer with delayed contacts in series with the start switch, or one with instantaneous contacts in series with delayed contacts.

interval contacts—In a timer, contacts which are actuated only for the duration of the present time interval.

interval timer—1. A device for measuring the time interval between two actions. 2. A timer which switches electrical circuits on or off for the duration of the preset time interval.

interword space—In telegraphy, the space between words or coded groups. It is equal to seven unit lengths.

intonation—The slight modification of pitch, or frequency, that makes a note sound flat or sharp compared with the natural frequency of the note played.

intraconnections (microelectronics)—Those conductors and connections which are in continuous integral contact with the substrate or circuit elements of an integrated circuit.

intrinsic-barrier diode—A PIN diode in which a thin region of intrinsic material separates the N-type and P-type region.

intrinsic-barrier transistor—A PNIP or NPIN transistor in which a thin region of intrinsic material separates the collector from the base.

intrinsic brightness—The luminous intensity measured in a given direction per unit of apparent (projected) area when viewed from that direction.

intrinsic coercive force—The magnetizing force that, when applied to a magnetic material in a direction opposite to that of the residual induction, reduces the intrinsic induction to zero.

intrinsic coercivity—The measurement (in oersteds) of the force required to reduce the intrinsic induction of the magnetized material to zero.

intrinsic concentration—In a semiconductor, the number of minority carriers that exceeds the normal equilibrium number.

intrinsic conduction—In an intrinsic semiconductor, the conduction associated with the directed movement of electron-hole pairs under the influence of an electric field.

intrinsic-contact potential difference—The true potential difference between two spotlessly clean metals in contact.

intrinsic electric strength—The characteristic electric strength of a material.

intrinsic induction—The excess magnetic induction produced in a magnetic material by a given magnetizing force, over the induction that would be produced by the same magnetizing force in a vacuum.

intrinsic-junction transistor—*See* Intrinsic-Region Transistor.

intrinsic layering—The method of separating two conductive semiconductor regions by a region of near-intrinsic semiconductor material.

intrinsic mobility—The mobility of electrons in an intrinsic semiconductor, or in a semiconductor having a very low concentration of impurities.

intrinsic noise—Noise that is due to the device or transmission path and is independent of modulation.

intrinsic permeability—Ratio of intrinsic normal induction to the corresponding magnetizing force.

intrinsic properties—The semiconductor properties which are characteristic of the pure, ideal crystal.

intrinsic Q—*See* Unloaded Q.

intrinsic region—*See* Depletion Region.

intrinsic-region transistor—Also called intrinsic-junction transistor. A four-layer transistor with an intrinsic region between the base and collector. Examples are NPIN, PNIP, NPIP, and PNIN transistors.

intrinsic semiconductor—1. A semiconductor in which some hole and electron pairs are created by thermal energy at room temperature, even though there are no impurities in it. 2. A semiconductor with substantially the same electrical properties as those of the ideal crystal.

intrinsic temperature range—The temperature range at which impurities or imperfections within the crystal do not modify the electrical properties of a semiconductor.

Invar—An alloy containing 63.8% iron, 36% nickel, and 0.2% carbon. Has a very low thermal coefficient of expansion. Used primarily as resistance wire in wirewound resistors.

inverse electrical characteristics—In a transistor, those characteristics obtained when the collector and emitter terminals are interchanged and the transistor is then tested in the normal manner.

inverse electrode current—The current flowing through an electrode in the opposite direction from that for which the tube was designed.

inverse feedback—Returning part of the output power of an amplifying device to its input circuit in order to cancel a portion of the input.

inverse-feedback filter—A tuned circuit at the output of a highly selective amplifier having negative feedback. The feedback output is zero for the resonant frequency, but increases rapidly as the frequency deviates.

inverse limiter—A transducer with a constant output for inputs of instantaneous values within a specified range. Above and below that range, the output is linear or some other prescribed function of the input.

inverse networks—Any two two-terminal networks in which the product of their impedances is independent of frequency within the range of interest.

inverse-parallel connection—*See* Back-to-Back Circuit.

inverse peak voltage—1. The peak instantaneous voltage across a rectifier tube during the nonconducting half-cycle. 2. The highest negative voltage reached between the plate and cathode of a rectifier tube.

inverse photoelectric effect—The transformation of the kinetic energy of a moving electron into radiant energy, as in the production of X-rays.

inverse piezoelectric effect—Contraction or expansion of a piezoelectric crystal under the influence of an electric field.

inverse ratio—The seesaw effect whereby one value increases as the other decreases or vice versa.

inverse-square law—The strength of a field, or the intensity of radiation, decreases in proportion to the square of the distance from its source.

inverse time—A qualifying term applied to a relay, indicating that its time of operation decreases as the magnitude of the operating quantity increases.

inverse voltage—The effective voltage across a rectifier tube during the half-cycle when current does not flow.

inverse Wiedeman effect—*See* Wiedeman Effect.

inversion—1. The bending of a radio wave because the upper part of the beam is slowed down as it travels through denser air. This may occur when a body of cold air moves in under a moisture-laden body of air. 2. The producing of inverted or scrambled speech by beating an audio-frequency signal with a fixed band of the resultant beat frequencies. The original low audio frequencies then become high frequencies and vice versa.

inverted amplifier—An amplifier stage containing two vacuum tubes. The control grids are grounded, and the driving excitation is applied between the cathodes. The grids then serve as a shield between the input and output circuits. Thus, the output-circuit capacitance is greatly reduced.

inverted-L antenna—An antenna consisting of one or more horizontal wires with a vertical wire connected at one end.

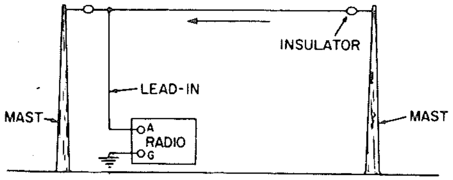

Inverted-L antenna.

inverted speech—*See* Scrambled Speech.

inverter—1. A circuit which takes in a positive signal and puts out a negative one, or vice versa. 2. A device that changes alternating current to direct current or vice versa. It frequently is used to change 6-volt or 12-volt direct current to 110-volt alternating current. 3. A device that accepts an input that is a function of the maximum voltage and changes it into an output that is a function of both the maximum voltage and time. 4. A circuit with one input and one output, and its function is to invert the input. When the input is high, the output is low, and vice versa. The inverter is sometimes called a NOT circuit, since it produces the reverse of the input.

inverting connection—The closed-loop connection of an operational amplifier when the forward gain is negative for d-c signals. A 180° phase shift.

inverting parametric device—A parametric device the operation of which depends essentially upon three frequencies—a harmonic of the pump frequency, and two signal frequencies—of which the higher signal frequency is the difference between the pump harmonic and the lower signal frequency.

I/O device—Abbreviation for input/output device. A card reader, magnetic tape unit, printer, or similar device that transmits data to or receives data from a computer or secondary storage device.

ion—An electrically charged atom or group of atoms. Positively charged ions have a deficiency of electrons, and negatively charged ions have surplus electrons.

ion charging—In a storage tube, spurious charging or discharging caused by ions striking the storage surface.

ion counter—A tubular chamber for measuring the ionization of air.

ionic focusing—Focusing the electron beam in a cathode-ray tube by varying the filament voltage and temperature to change the electrostatic focusing field automatically produced by the accumulation of positive ions in the tube.

ionic-heated cathode—A cathode that is heated primarily by bombardment with ions.

ionic-heated cathode tube—An electron tube containing an ionic-heated cathode.

ionization—1. A process by which electrons are detached from atoms or molecules, thus originating ions and free electrons. It can be done by high-energy radiation or by collisions. 2. The electrically charged particles produced by high-energy radiation (such as light or ultra-violet rays) or by the collision of particles during thermal agitation.

ionization chamber—An enclosure containing two or more electrodes between which an electric current may pass when the gas within is ionized. The current is a measure of the total number of ions produced in the gas by externally induced radiations.

ionization current—The current resulting from the movement of electric charges, under the influence of an applied electric field, in an ionized medium.

ionization energy—Sometimes called ionization potential. The minimum amount of energy (usually expressed in electron-volts) required to eject an electron from a molecule.

ionization gage—A gage that measures the degree of a vacuum in an electron tube by the amount of ionization current in the tube.

ionization-gage tube—An electron tube that measures low gas pressure by the amount of ionization current produced.

ionization potential—Also called critical potential. The minimum energy that will produce ionization in a substance. Expressed as volts acting on one electronic charge.

ionization pressure—An increase in the pressure in a gaseous discharge tube due to ionization of the gas.

ionization resistance — *See* Corona Resistance.

ionization time—The time interval between the initiation and the establishment of conduction in a gas tube at some stated voltage drop for the tube.

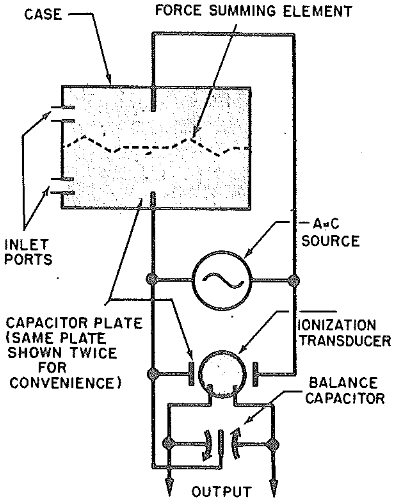

Ionization transducer.

ionization transducer — A transducer in which displacement of the force-summing member is sensed by the induced changes in differential ion conductivity.

ionization vacuum gage—A gage in which the operation depends on the positive ions produced in a gas by electrons as they accelerate between a hot cathode and another electrode in a vacuum. It ordinarily covers a pressure range of 10^{-4} to 10^{-10} mm of mercury.

ionize—To free an electron from an atom or molecule (e.g., by X-ray bombardment) and thus transform the atom or molecule into a positive ion. The freed electron attaches itself to another atom or molecule, which then becomes a negative ion.

ionized layers—Layers of increased ionization within the ionosphere. They are responsible for absorption and reflection of radio waves and are important for communication and for tracking satellites and other space vehicles.

ionizing event—Any interaction by which one or more ions are produced.

ionizing radiation — Electromagnetic radiation or particle radiation having sufficient energy to dislodge electrons from atoms or molecules, thereby producing ions.

ion migration—Movement of the ions produced in an electrolyte by application of an electric potential between electrodes.

ionophone — A high-frequency speaker in which the audio-frequency signal modulates an r-f supply to maintain an arc in the mouth of a quartz tube. The resultant modulated wave acts directly on the ionized air under pressure and thus creates sound waves.

ionosphere—The part of the earth's outer atmosphere where sufficient ions and electrons are present to affect the propagation of radio waves.

ionospheric disturbance—The variation in the state of ionization of the ionosphere beyond the normal observed random day-to-day variation from average value for the location, data, and time of day under consideration.

ionospheric error—Also called sky error. In navigation, the total systematic and random error resulting from reception of the navigational signal after it has been reflected from the ionosphere. It may be due to variations in the transmission path, uneven height of the ionosphere, or uneven propagation within the ionosphere.

ionospheric prediction—The forecasting of ionospheric conditions and the preparation of radio propagation data derived from it.

ionospheric scatter—*See* Forward Scatter.

ionospheric storm—An ionospheric disturbance associated with abnormal solar activity and characterized by wide variations from normal, including turbulence in the F-region and increases in absorption. Often the ionization density is decreased and the virtual height is increased. The effects are most marked in high magnetic latitudes.

ionospheric wave—Also called a sky wave. A radio wave that is propagated by way of the ionosphere.

ion sheath—A positive-ion film which forms on or near the grid of a gas tube and limits its control action.

ion spot—1. In camera or image tubes, the spurious signal resulting from bombardment of the target or photocathode by ions. 2. On a cathode-ray-tube screen, an area where the luminescence has been deteriorated by prolonged bombardment with negative ions.

ion trap—Also called a beam bender. An electron-gun structure and magnetic field which diverts negative ions to prevent their burning a spot in the screen, but permits electrons to flow toward the screen.

I_p—Symbol for the plate current of a vacuum tube.

I-phase carrier—Also called in-phase carrier. A carrier separated in phase by 57° from the color subcarrier.

ips—Abbreviation for inches per second.

IR—Abbreviation for interrogator-responser or infrared.

I^2R—Power in watts expressed in terms of the current (I) and resistance (R).

IRAC—Acronym for Interdepartmental Radio Advisory Committee. It is composed of representatives of eleven government agencies: The FCC; Army; Navy; Air Force; Maritime Commission; and the Treasury, State, Commerce, Agriculture, Interior, and Justice Departments.

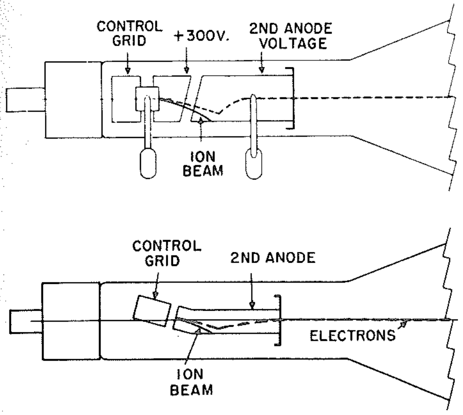

Ion trap.

IR compensation — A control device that compensates for voltage drop due to current flow.

IR drop—The voltage drop produced across a resistance (R) by the flow of current (I) through it.

IRE—Abbreviation for Institute of Radio Engineers, an organization now merged with AIEE. *See* IEEE.

i (intrinsic) region — In silicon, a pure region of the group-IV element (i.e., having neither excess holes or excess electrons, and therefore having very high resistivity).

iris—Also called diaphragm. In a waveguide, a conducting plate which is very thin (compared with the wavelength) and occupies part of the cross section of a waveguide.

I^2R loss—The power lost in transformers, generators, connecting wires, and other parts of a circuit because of the current flow I through the resistance R of the conductors.

iron-constantan — A combination of metals used in thermocouples, thermocouple wires, and thermocouple lead wires. Constantan is an alloy of copper, nickel, manganese, and iron.

iron-core coil—A coil in which iron forms part or all of the magnetic circuit, linking its winding. In a choke coil, the core is usually built up of laminations of sheet iron.

iron-core transformer — A transformer in which iron forms part or all of the magnetic circuit, linking the transformer windings.

iron loss—*See* Core Loss.

iron-vane instrument—A measuring instrument in which the movable element is an iron vane.

irradiation—The application of X-rays, radium rays, or other radiation.

irregularity—A change from normal in the impedance characteristics.

I-scan—A radar display in which a target appears as a complete circle when the radar antenna is correctly pointed at it, the radius of the circle being proportional to the target distance.

I-scope—*See* I-scan.

I-signal—Also called the fine-chrominance primary. A signal formed by the combination of an R—Y signal having a +.74 polarity and a B—Y signal having a —.27 polarity. One of the two signals used to modulate the chrominance subcarrier, the other being the Q signal.

ISM equipment—A Federal Communications Commission designation for *i*ndustrial, *s*cientific, and *m*edical equipment capable of causing interference.

isobar—1. On meteorological maps, a line denoting places having the same atmospheric pressure at a given time. 2. One of a group of atoms or elements having the same atomic weights but different atomic numbers.

isochrone—On a map or chart, a line joining points associated with a constant time difference in the reception of radio signals.

isochrone determination—A radio location in which a position line is determined by the difference in transit times of signals along two paths.

isochronous circuits — Circuits having the same resonant frequency.

isoclinic line—*See* Aclinic Line.

isodynamic lines—On a magnetic map, lines passing through points of equal strength of the earth's magnetic field.

isoelectronic—Having the same number of electrons outside the nucleus of the atom.

isolated—Utterly cut off from; refers to that condition in which a conductor, circuit, or device is not only insulated from another (or others), but the two are mutually unable to engender current, emf, or magnetic flux in each other. As commonly used, insulation is associated predominantly with direct current, whereas isolation implies additionally a bulwark against a-c fields.

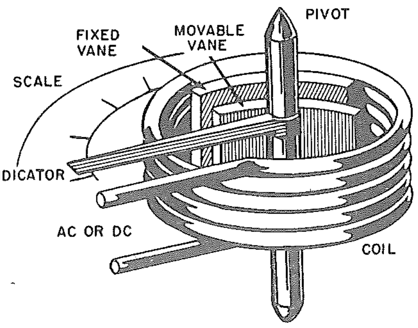

Iron-vane instrument (movable).

isolated amplifier—A differential amplifier in which the input-signal lines are conductively isolated from the output-signal lines and chassis ground.

isolating diode—A diode that passes signals in one direction through a circuit but blocks signals and voltages in the opposite direction.

isolating switch—A switch intended for isolating an electric circuit from the source of power. It has no interrupting rating and is intended to be operated only after the circuit has been opened by some other means.

isolation—1. Electrical or acoustical separation between two locations. 2. The technique for producing a high electrical resistance between an integrated-circuit component and the substrate in which it is formed.

isolation amplifier—An amplifier employed to minimize the effects of the following circuit on the preceding circuit.

isolation network—A network inserted into a circuit or transmission line to prevent interaction between circuits on each side of the insertion point.

isolation transformer — A transformer designed to provide magnetic coupling (flux coupling) between one or more pairs of isolated circuits, without introducing significant coupling of any other kind between them—i.e., without introducing either significant conductive (ohmic) or significant electrostatic (capacitive) coupling.

isolator ferrite—A microwave device which allows r-f energy to pass through in one direction with very little loss but absorbs r-f power in the other direction.

isolith—An integrated circuit of components formed on a single silicon slice, but with the various components interconnected by beam leads and with circuit parts isolated by removal of the silicon between them.

isomer—One of two or more substances composed of molecules having the same kinds of atoms in the same proportions but arranged differently. Hence, the physical and chemical properties are different. Isomers which do not have the same molecular weights are called polymers.

isostatic — Being subjected to equal pressure from every side.

isothermal region—The stratosphere considered as a region of uniform temperature.

isotones — A group of atoms the nuclei of which have the same number of neutrons.

isotope—A species of matter the atoms of which contain the same number of protons as some other species, but a different number of neutrons. The atomic numbers of isotopes are identical, but the mass numbers (atomic weights) differ.

isotropic—Having the same properties in all directions.

isotropic antenna — Also called unipole. A hypothetical antenna radiating or receiving equally in all directions. A pulsating sphere is a unipole for sound waves. In the case of electromagnetic waves, unipoles do not exist physically, but represent convenient reference antennas for expressing directive properties of actual antennas.

isotropic gain of an antenna—*See* Absolute Gain of an Antenna.

isotropic magnet—A magnetic material having no preferred axis of magnetic characteristics.

isotropic radiator—A radiator which sends out equal amounts of energy in all directions.

item—A general term denoting one of a number of similar units, assemblies, objects, etc.

iterative impedance—An impedance used to terminate one of two pairs so that the other pair of terminals will have the same input impedance.

iterative process—The calculating of a desired result by means of a repeating cycle of operations which comes closer and closer to the desired result.

iterative routine—A computer routine composed of repetitive computations, so that the output of every step becomes the input of the succeeding step.

ITU—Abbreviation for International Telecommunications Union, an international organization established to provide standardized communications procedures and practices, including frequency allocation and radio regulations on a world-wide basis. It is the international equivalent of the Federal Communications Commission.

ITV—Abbreviation for industrial television.

I-type—Intrinsic semiconductor.

J

J—Abbreviation for joule.

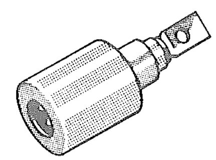

Jack.

jack—A socket to which, the wires of a circuit are connected at one end, and into which a plug is inserted at the other end.

jacket — Pertaining to wire and cable, the outer sheath of which protects against environment and may also provide additional insulation.

jack panel—An assembly composed of a number of jacks mounted on a board or panel.

jackscrew—A screw attached to one-half of a two-piece, multiple-contact connector and used to draw both halves together and to separate them.

jaff—Slang for the combination of electronic and chaff jamming.

jag—In facsimile, distortion caused in the received copy by a momentary lapse in synchronism between the scanner and recorder.

jam—In punch-card machines, a condition in the card feed which interferes with the normal travel of the punch cards through the machine.

jammer—An electronic device for intentionally introducing unwanted signals into radar sets to render them ineffective.

jammer band — The radio-frequency band where the jammer output is concentrated. It is usually the band between the points where the intensity is 3 db down from maximum.

jammer finder — Also called burnthrough. Radar which attempts to obtain the range of the target by training a highly directional pencil beam on a jamming source.

jamming—The intentional transmission of radio signals in order to interfere with the reception of signals from another station.

jamming effectiveness—The jamming-to-signal ratio—i.e., the percentage of information incorrectly received in a test message.

JAN specification — Joint Army-Navy specification. The forerunner of present Military Specifications, which are generally superseded by the designation MIL.

J-antenna—A half-wave antenna fed at one end by a parallel-wire, quarter-wave section having the configuration of a *J*.

J-carrier system—A broad-band carrier system which provides 12 telephone channels and utilizes frequencies up to about 140 kilohertz by means of four-wire transmission on a single open-wire pair.

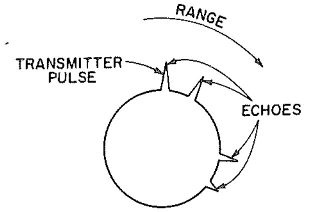

J-display.

J-display—Also called J-scan. In radar, a modified A-display in which the time base is a circle and the target signal appears as a radial deflection from it.

JEDEC—Acronym for Joint Electron Device Engineering Council.

JETEC — Acronym for Joint Electron Tube Engineering Council.

jewel bearing—A natural or synthetic jewel, usually sapphire, used as a bearing for a pivot or other moving parts of a delicate instrument.

jezebel—A system for the detection and classification of submarines.

jitter—1. Instability of a signal in either its amplitude, its phase, or both. The term is applied especially to signals reproduced on the screen of a cathode-ray tube. The term "tracking jitter" describes minor variations in the pointing of an automatic tracking radar. 2. In facsimile, raggedness in the received copy caused by erroneous displacement of recorded spots in the scanning direction. 3. An aberration of a repetitive display indicating instability of the signal or of the oscilloscope. May be random or periodic, and is usually associated with the time axis. 4. *See* Flicker *and* Fortuitous distortion.

jittered pulse recurrence frequency—The random variation of the pulse-repetition period. Provides a discrimination capability against repeater-type jammers.

jogging—Also called inching. Quick and repeated opening and closing of a motor starting circuit to produce slight movements of the motor.

Johnson noise—Also called thermal noise. The noise generated by any resistor at a temperature above absolute zero. It is proportionate to the absolute temperature and the bandwidth, according to the following formula:

$$N = KTB$$

where,

N is the noise power in watts,
K is Boltzmann's constant, or 1.38047 $\times 10^{-23}$
T is the absolute temperature in degrees Kelvin,
B is the bandwidth in hertz.

joined actuator — A multipole breaker such that when one pole trips, all trip, but whereas the faulted pole is trip-free, the other poles may be kept maintained by a restraining actuator.

joint — A connection between two or more conductors.

joint use—The simultaneous use of pole, line, or plant facilities by two or more kinds of utilities.

Jones plug—A type of polarized connector designed in the form of a receptacle and having several contacts.

Josephson effect — The phenomenon described by Brian Josephson to explain the action of currents through and voltages across hairlike gaps in superconductors. On the basis of theoretical considerations, it is predicted that if two superconductors could be brought close enough together, a current could be made to flow across the gap between them. Under certain conditions, a voltage appears across the gap, and high-frequency radiation emanates from it. This predicted radiation would have a frequency precisely equal to 2eV/h, where V is the measured voltage across the gap.

joule—1. The work done by a force of 1 newton acting through a distance of 1 meter. The joule is the unit of work and energy in the MKSA system. 2. The energy required to transport 1 coulomb between two points having a potential difference of 1 volt. The joule is 10^7 ergs. The kilowatt-hour is 3.6×10^6 joules.

Joule effect—In a circuit, electrical energy is converted into heat by an amount equal to I^2R. Half of this heat flows to the hot junction and the other half to the cold junction.

Joule heat—The thermal energy produced as a result of the Joule effect.

Joule's law of electric heating—The amount of heat produced in a conductor is proportional to the resistance of the conductor, the square of the current, and the time.

J-scan—*See* J-Display.

J-scope—Also called Class-J oscilloscope. A cathode-ray oscilloscope that presents a J-display.

J/S ratio—A ratio, normally expressed in db, of the total interference power to the signal-carrier power in the transmission medium at the receiver.

juice—Slang for electric current.

jump—To cause the next instruction to be selected from a specified storage location in a computer.

jumper—1. A short length of wire used to complete a circuit temporarily or to bypass a circuit. 2. A direct electrical connection, which is not a portion of the conductive pattern, between two points in a printed circuit.

junction—1. A connection between two or more conductors or two or more sections of a transmission line. 2. A contact between two dissimilar metals or materials (e.g., in a rectifier or thermocouple). 3. A region of transition between P- and N-type semiconductor material. The controllable resultant asymmetrical properties are exploited in semiconductor devices. There are diffused, alloy, grown, and electrochemical junctions.

junction barrier—The opposition to the diffusion of majority carriers across a PN junction due to the charge of the fixed donor and acceptor ions.

junction box—A box for joining different runs of raceway or cable, plus space for connecting and branching the enclosed conductors.

junction diode—A two-terminal device containing a single crystal of semiconducting material which ranges from P-type at one terminal to N-type at the other. It conducts current more easily in one direction than in the other and is a basic element of the junction transistor. When fabricated in a suitable geometrical form, it can be used as a solar cell.

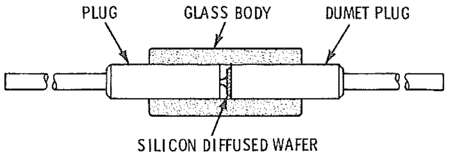

Junction-diode assembly.

junction filter—A combination of a high-pass and a low-pass filter which is used to separate frequency bands for transmission over separate paths. For example, junction filters are used to separate voice and carrier frequencies at the junction between open-wire and cable so that the carrier frequencies and voice frequencies can be sent over nonloaded and v-f loaded cable pairs, respectively.

junction point—*See* Node.

junction station—A microwave relay station that joins a microwave radio leg or legs to the main, or through, route.

junction transistor — A transistor having three alternate sections of P-type or N-type semiconductor material. (*Also see* PNP Transistor *and* NPN Transistor.)

junctor—In crossbar systems, a circuit extending between frames of a switching unit and terminated in a switching device on each frame.

just-operate value — Also called drop-out value. The measured functioning value at which a particular relay releases.

just-release value—The measured functioning value at which a particular relay releases.

just scale—A musical scale formed by three consecutive triads (those in which the highest note of one is the lowest note of the other), each having the ratio 4:5:6 or 10:12:15.

jute—Cordage fiber (such as hemp) saturated with tar and used as a protective layer over cable.

jute-protected cable — A cable having its sheath covered by a wrapping of tarred jute or other fiber.

K

k—Abbreviation for kilo.

K—1. Symbol for cathode or dielectric constant. 2. Abbreviation for Kelvin or kilo. 3. Abbreviation for luminosity factor.

ka—Abbreviation for kiloampere.

Karnaugh map—A truth table that is rearranged to show a geometrical pattern of functional relationships for gating configurations. Essential gating requirements can then be recognized in their simplest form.

K-band—A radio frequency band extending from 11 to 36 GHz and having wavelengths of 2.73 to 0.83 cm.

kc—Abbreviation for kilocycle. Now obsolete.

K-carrier system—A broad-band carrier system which provides 12 telephone channels and utilizes frequencies up to about 60 kHz by means of four-wire transmission on cable facilities.

K-display—Also called K-scan. Modification of a type-A scan, used for aiming a double-lobe system in bearing or elevation. The entire range scale is displaced toward the antenna lobe in use. One signal appears as a double deflection from the range and relative scales. The relative amplitudes of these two pips indicate the amount of error in aiming the antenna.

keep-alive anode — An auxiliary electrode that maintains a d-c discharge in a mercury-pool tube. It has the disadvantage of reducing the peak inverse voltage rating.

keep-alive circuit — In a TR or anti-TR switch, a circuit for producing residual ionization in order to reduce the time for full ionization when the transmitter fires.

keep-alive voltage—A d-c voltage that maintains a small glow discharge within one of the gap electrodes of a TR tube. This allows the tube to ionize more rapidly when the transmitter fires thus preventing damage to the receiver.

keeper—A magnetic conductor placed over the ends of a permanent magnet to protect it against being demagnetized.

kel-f—Polymonochlorotrifluoroethylene—used as a high-temperature insulation (−55°C to +135°C).

Kelvin balance—An instrument for measuring current. This is done by sending it through a fixed and a movable coil attached to one arm of a balance. The resultant force between the coils is then compared with the force of gravity acting on a known weight at the other end of the balance arm.

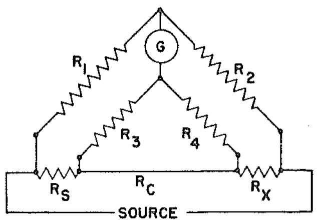

Kelvin bridge.

Kelvin bridge — Also called a double or Thomson bridge. A seven-arm bridge for comparing the resistances of two 4-terminal resistors or networks. Their adjacent potential terminals are spanned by a pair of auxiliary resistance arms of known ratio, and they are connected in series by a conductor joining their adjacent current terminals.

Kelvin scale—Also called absolute scale. A temperature scale using the same divisions as the Celsius scale, but with the zero point established at absolute zero (≅ −273°C), theoretically the lowest possible temperature.

Kendall effect—A spurious pattern or other distortion in a facsimile record. It is caused by unwanted modulation produced by transmission of a carrier signal. Such modulation appears as a rectified baseband that interferes with the lower sideband of the carrier.

Kennelly-Heaviside layer — *See* Heaviside Layer.

kenopliotron—A diode-triode vacuum tube within one envelope. The anode of the diode also serves as the cathode of the triode.

kenotron—Also called a valve tube. A term used primarily in industrial and X-ray fields for a hot-cathode vacuum tube.

keraunophone—A radio circuit device for audibly demonstrating the occurrence of distant lightning flashes.

kernel—A line within a current-carrying conductor along which the magnetic intensity due to the current is zero.

Kerr cell—A transparent enclosure used as a light valve in some mechanical television systems. It contains a transparent substance which exhibits electric double refraction. Hence, it can convert a varying voltage into corresponding variations in the intensity of a polarized light passing through the cell.

Kerr effect—1. An electro-optical effect in which certain transparent substances become double refracting when subjected to an electric field perpendicular to a beam of light. 2. The conversion of plane into elliptically polarized light when reflected from the polished end of a magnet.

kev—Abbreviation for kilo-electron-volt.

key—1. A hand-operated switching device for switching one or more parts of a circuit. It ordinarily consists of concealed spring contacts and an exposed handle or push button. 2. A projection which slides into a mating slot or groove so as to guide two parts being assembled and assure proper polarization.

keyboard computer—A computer whose input employs a keyboard, e.g., an electric typewriter.

keyboard perforator — A mechanism that punches a paper tape from which messages are automatically transmitted by a transmitter distributor. The keyboard is similar to that of a typewriter and can be operated by any trained typist after a few hours' instruction. As each key is depressed, the tape is punched with corresponding code symbols.

keyboard send/receive—A combination teletypewriter transmitter and receiver with transmission capability from a keyboard only.

key click—Undesired clicking noises caused by the opening and closing of the signaling key.

key-click filter—Also called a click filter. A filter that attenuates the surges produced each time the keying circuit of a transmitter is opened or closed.

keyed agc — Abbreviation for keyed automatic gain control.

keyed automatic gain control—Abbreviated keyed agc. A television automatic gain control in which the agc tube is kept cut off except when the peaks of the positive horizontal-sync pulse act on its grid. The agc voltage is therefore not affected by noise pulses occurring between the sync pulses.

keyed rainbow generator—A color television test instrument which displayes the individual colors of the spectrum, separated by black bars, on the picture tube.

keyed rainbow signal—A 3.563795-MHz (3.56-MHz) continuous sine-wave signal from a color-bar generator that is pulsed on and off. This signal creates a series of different color bars on the screen of the color picture tube. A typical pulse rate (for 10 color bars) is 12 times per 1 horizontal line.

keyer—In telegraphy a device which breaks up the output of a transmitter or other device into the dots and dashes that are used in the code.

keyer adapter—A device that detects a modulated signal and produces a d-c output signal whose amplitude varies in accordance with the modulation. In radio facsimile transmission, it is used to provide the keying signal for a frequency-shift exciter unit.

keying—The forming of signals, such as those employed in telegraph transmission, by an abrupt modulation of the output of a direct- or an alternating-current source (e.g., by interrupting it or by suddenly changing its amplitude, frequency, or some other characteristic).

keying chirps—Sounds accompanying code signals when the transmitter is unstable and shifts slightly in frequency each time the sending key is closed.

keying frequency—In facsimile, the maximum number of times a second a black-line signal occurs while scanning the subject copy.

keying wave—Also called marking wave. The emission that takes place in telegraphic communication while the information portion of the code characters is being transmitted.

key pulse—A telephone signaling system in which numbered keys are depressed instead of a dial being turned.

key pulsing—A switchboard arrangement using a nonlocking keyset for the transmission of pulse signals corresponding to the key depressions.

key punch—A keyboard machine for manually punching information into paper tape or cards.

key station—The master station from where a network radio or television program originates.

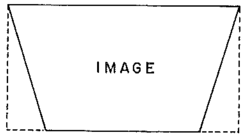

Keystone distortion.

keystone distortion—1. In a tv receiver, distortion seen as a trapezoidal instead of rectangular raster. 2. The distortion produced when a plane target area not normal to the average direction of the beam is scanned rectilinearly with constant-amplitude sawtooth waves. 3. Camera-tube distortion in which the slope or length of a horizontal line trace or scan line is linearly related to its vertical displacement.

keystone shaped—Wider at the top than at the bottom, or vice versa.

keystoning—The keystone-shaped scanning pattern produced when the electron beam in the television camera tube is at an angle with the principal axis of the tube. (*Also see* Keystone Shaped.)

keyswitch—In an organ, the switch which is closed to allow a tone from the tone generator to sound when a key is depressed.

keyway—The mating slot or groove in which a key slides.

kHz—Abbreviation for kilohertz.

kickback—The voltage developed across an inductance by the sudden collapse of the magnetic field when the current through the inductance is cut off.

kickback power supply—*See* Flyback Power Supply.

kick-sorter—British term for pulse-height analyzer.

Kikuchi lines—A series of spectral lines obtained by the scattering of electrons, when an electron beam is directed against a crystalline solid. The pattern may be interpreted to yield information on the structure of the crystal and its mechanical perfection.

killer circuit—1. The vacuum tube or tubes and associated circuits in which are generated the blanking pulses used to temporarily disable a radar set. 2. In a transponder, a logic circuit that kills replies to side-lobe interrogations.

killer pulse—*See* Suppression Pulse.

kilo—Abbreviated k. A prefix representing 10^3, or 1,000.

kiloampere—1,000 amperes. Abbreviated ka.

kilobit—1,000 bits.

kilocycle (kc)—1,000 cycles. Generally interpreted as meaning 1,000 cycles per second. Obsolete term, replaced by kilohertz (kHz).

kilo-electron-volt—1,000 electron-volts. The energy acquired by an electron that has been accelerated through a voltage difference of 1,000 volts. Abbreviated kev.

kilogauss—1,000 gausses.

kilogram (unit of mass)—The mass of a particular cylinder of platinum-iridium alloy, called the international prototype kilogram, which is preserved in a vault at Sevres, France, by the International Bureau of Weights and Measures.

kilohertz—1,000 hertz. Abbreviated kHz.

kilohm—One thousand ohms. Abbreviated k or kohm.

kilohmmeter—A meter designed for measuring resistance in kilohms.

kilomega (kM)—Obsolete prefix for giga (G), representing 10^9, or 1,000,000,000.

kilomegacycle—Now called gigahertz. One billion cycles per second.

kilometer—One thousand meters, or approximately 3,280 feet.

kilometric waves—British term for electromagnetic waves between 1,000 and 10,000 meters in length.

kilosecond—1,000 seconds.

kilovar—One reactive kilovolt-ampere, or 1,000 reactive volt-amperes. Abbreviated kvar.

kilovar-hour—1,000 reactive volt-ampere-hours.

kilovolt—1,000 volts. Abbreviated kv.

kilovolt-ampere—1,000 volt-amperes. Abbreviated kva.

kilovoltmeter—A voltmeter which reads thousands of volts.

kilowatt—1,000 watts. Abbreviated kw.

kilowatt-hour—The equivalent energy supplied by a power of 1,000 watts for one hour. Abbreviated kwhr.

kine—Slang term for kinescope recording.

kine-klydonograph—An instrument that records the current-time characteristics of a lightning stroke. The instrument records a series of Lichtenberg figures in a manner similar to the field gradient recorder.

kinescope—1. In television receivers, the cathode-ray tube in which the electrical signals are translated into a visible picture on a luminescent screen. 2. A film recording made from a television program on a picture tube and used as a permanent record or for subsequent rebroadcasting.

kinescope recorder—A camera which photographs television images directly from the picture tube onto motion-picture film.

kinetic energy—Energy which a system possesses by virtue of its motion.

Kirchhoff's laws—1. The current flowing to a given point in a circuit is equal to the current flowing away from that point. 2. The algebraic sum of the voltage drops in any closed path in a circuit is equal to the algebraic sum of the electromotive forces in that path. (Laws 1 and 2 are also called Laws of electric networks.) 3. At a given temperature, the emissive power of a body is the same as its radiation-absorbing power for all surfaces.

kit—A prepared package of parts with instructions for assembly and/or wiring a component or chassis (also a small accessory item).

klydonograph—A field gradient recording instrument that registers voltages on photographic film in the form of Lichtenberg figures.

klystron—1. An electron tube used as an oscillator or amplifier in ultrahigh frequencies. The electron beam is velocity-modulated (periodically bunched) to accomplish the desired results. 2. A type of oscillator or amplifier used for the generation or reception of microwave frequencies.

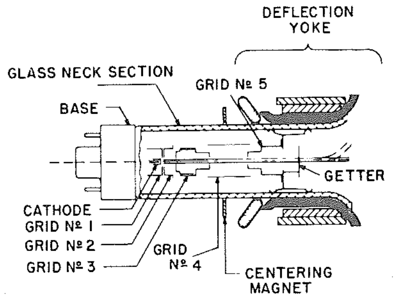

Kinescope.

klystron control grid—An electrode which controls the emission, or beam current, of a klystron or other velocity-modulated tube.

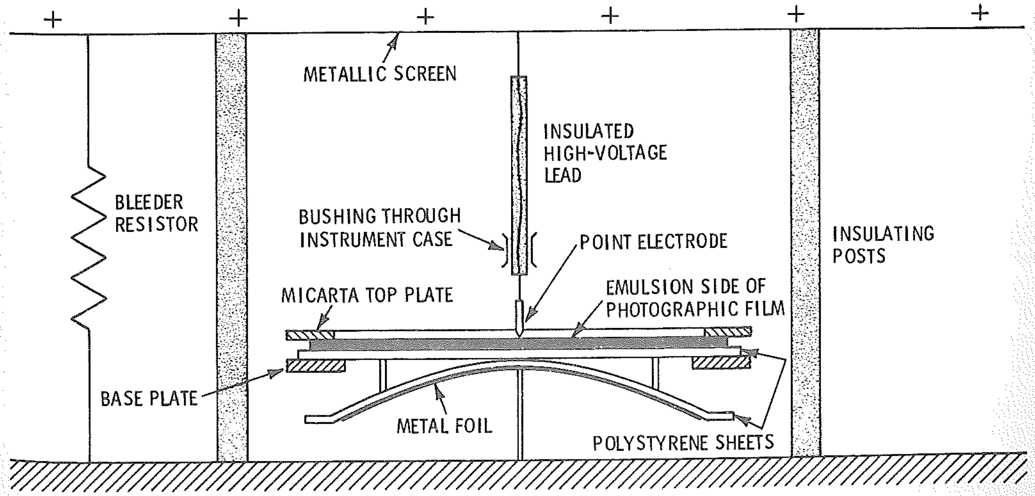

Klydonograph gradient recorder.

klystron frequency multiplier—A two-cavity klystron that has the output cavity tuned to a multiple of the fundamental frequency.

klystron generator—A klystron tube used as a generator. Its cavity feeds energy directly into a waveguide.

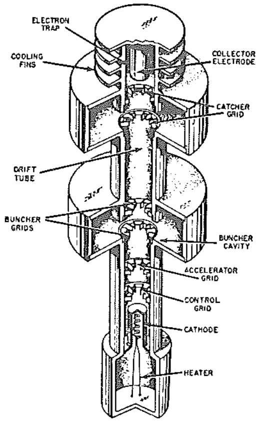

Cutaway view of klystron tube.

klystron oscillator—An oscillator employing a klystron tube to generate radio-frequency power.

klystron repeater—A klystron tube operated as an amplifier and inserted directly into a waveguide in such a way that incoming waves velocity-modulate the electron stream emitted from a heated cathode. A second cavity converts the energy of the electron clusters into waves of a much higher amplitude and feeds them into the outgoing guide.

kM—Abbreviation for kilomega (an obsolete term).

kmc—Abbreviation for kilomegacycle. Now replaced by gigahertz.

knee—An abrupt change in direction between two relatively straight segments of a curve, such as the region of a magnetization curve near saturation or the top bend of a vacuum-tube characteristic curve.

knife switch—A form of air switch in which a moving element is sandwiched between two contact clips. The moving element is usually a hinged blade, although when it is not, it is removable.

knob—A round, polygonal, or pointer-shaped part which is fastened to one end of a control shaft so that the shaft can be turned more easily. The knob sometimes indicates the degree of rotation also.

knocker—A term used with some fire-control radars to indicate a subassembly comprising synchronizing and triggering circuits. It drives the r-f pulse-generating equipment in the transmitter, and also synchronizes the cycle of operation with the transmitted pulse in range units and indicators.

knockout—A removable portion in the side of a box or cabinet. During installation it can be readily taken out with a hammer, screwdriver, or pliers so the raceway, cables, or fittings can be attached.

knot—One nautical mile (6,080.20 feet, or 1.15 statute miles) per hour.

kohm—Abbreviation for kilohm.

kraft paper—Relatively heavy, high-strength sulfate paper used for electrical insulating material. (Capacitor tissue is kraft paper of normal thickness equal to 0.002 inch or less.)

K-scan—*See* K-Display.

K-series — A series of frequencies in the X-ray spectrum of an element.

KSR—Keyboard send-receive set—A combination transmitter and receiver with transmission capability from the keyboard only (teletypewriter term).

kurtosis — The degree of curvature of the peak of a probability curve.

kv—Abbreviation for kilovolt.

kva—Abbreviation for kilovolt-ampere.

kvar—Abbreviation for kilovar.

kw—Abbreviation for kilowatt.

kwhr—Abbreviation for kilowatt-hour.

kymograph — An instrument for recording wavelike oscillations of varying quantities for medical studies.

L

L—Symbol for coil or inductance.

label—A code name used to identify or classify a name, term, phrase, or document.

labile oscillator—A local oscillator the frequency of which is remote-controlled by a signal received from a radio or over a wire.

laboratory power supply — A regulated d-c source having (a) less than 10-kV output at up to 500 W. (b) Output adjustable over a wide range, usually down to zero. (c) Regulation on the order of ±0.01% static line and load.

labyrinth—Speaker enclosure with absorbing air chambers at the rear to eliminate acoustic standing waves.

lacquer disc — Also called cellulose-nitrate discs. A mechanical recording disc, usually made of metal, glass, or paper and coated with a lacquer compound often containing cellulose nitrate.

lacquer master—*See* Lacquer Original.

lacquer original—Also called lacquer master. An original recording made on a lacquer surface to be used as a master.

lacquer recording—Any recording made on a lacquer medium.

ladder attenuator—A series of symmetrical sections used in signal generators and other devices where voltages and currents must be reduced in known ratios. They are designed so that the required ratio of voltage loss per section is obtained with image-impedance operation. The impedance between any junction point and common ground in a ladder attenuator is half the image impedance.

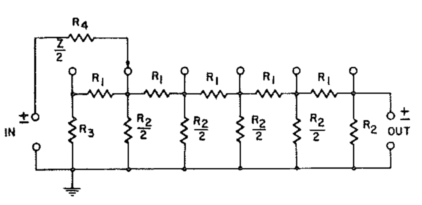

Ladder attenuator.

ladder network—Also called series-shunt network. A network composed of H-, L-, T-, or pi-networks connected in series.

LAFOT—Coded weather broadcasts issued by the U. S. Weather Bureau for the Great Lakes region. They are broadcast every six hours by Marine radiotelephone broadcasting stations on their assigned frequencies.

lag—1. The displacement in time, expressed in electrical degrees, between two waves of the same frequency. 2. The time between transmission and reception of a signal. 3. In a television camera tube, the persistence of the electrical-charge image for a time interval equal to a few frames.

lagged-demand meter—A meter in which there is a characteristic time lag, by either mechanical or thermal means, before maximum demand is indicated.

lagging current—The current flowing in a circuit that is mostly inductive. If the circuit contains only inductance, the current lags the applied voltage by 90°. Because of the characteristics of an inductance, the current does not change direction until after the corresponding voltage does.

lagging load — A predominantly inductive load—i.e., one in which the current lags the voltage.

lag-lead (lead-lag)—A circuit the response of which includes lag components and their derivatives.

lambda—Greek letter λ, used to designate wavelength measured in meters.

lambert—A unit of luminance equal to $1/\pi$ candle per square centimeter. It is equal to the uniform luminance of a perfectly diffusing surface emitting or reflecting light at the rate of one lumen per square centimeter.

Lambert's law of illumination—The illumination of a surface on which the light falls normally from a point source is inversely proportional to the square of the distance of the surface from the source. If the normal to the surface makes an angle with the direction of the rays, the illumination is proportional to the cosine of that angle.

laminated—Made of layers.

laminated contact—A switch contact made up of a number of laminations, each making contact with an opposite conducting surface.

laminated core—An iron core for a coil, transformer, armature, etc. It is built up from laminations to minimize the effect of eddy currents. The sheet iron or steel laminations are insulated from each other by surface oxides or by oxides and varnish.

laminated record—A mechanical recording medium composed of several layers of material (normally a thin face of material on each side of a core).

lamination—A single stamping of sheet material used in building up a laminated object such as the core of a power transformer.

Lamont's law—The permeability of steel at any flux density is proportional to the difference between the saturation value of the flux density and its value at the point in question. This law is only approximately accurate and is not true for the initial part of the magnetization curve.

lamp—A device for producing light.

lamp bank—An arrangement of incandescent lamps commonly used as a resistance load during electrical tests.

lamp cord—A twin conductor, either twisted or parallel, used for connecting floor lamps and other electric appliances to wall outlets.

lamp holder—A lamp socket.

Lampkin oscillator—A variation of the Hartley oscillator. Its distinguishing feature is that an approximate impedance match is effected between the tank and grid-cathode circuits.

lamp receptacle—A device that supports an electric lamp and connects it to a power line.

land — 1. The surface between two adjacent grooves of a recording disc. 2. Also called boss, pad, terminal point, blivet, tab, spot, donut. In a printed-circuit board, the conductive area to which components or separate circuits are attached. It usually surrounds a hole through the conductive material and the base material.

Landau damping—The damping of a space-charge wave by electrons moving at the phase velocity of the wave.

landing beacon—The radio transmitter that produces a landing beam for aircraft. (*Also see* Landing Beam.)

landing beam—A highly directive radio signal projected upward from an airport to guide aircraft in making a landing during poor visibility.

landline—A telegraph or telephone line passing over land, as opposed to submarine cables.

landline facilities — Domestic communications common-carrier's facilities that are within the continental United States.

landmark beacon—Any beacon other than an airport or airway beacon.

land mobile service — A radio service in which communication is between base stations and land mobile stations or between land mobile stations.

land mobile station—A two-way mobile station that operates solely on land.

land return—Radiation reflected from nearby land masses and returned to a radar set as an echo.

lands—Bonding points used in the manufacture of microelectronic circuits.

land station—A permanent, or fixed station.

Langevin ion—An electrified particle produced in a gas by an accumulation of ions on dust particles or other nuclei.

Langmuir dark space—The nonluminous region surrounding a negatively charged probe inserted into the positive column of a glow or arc discharge.

language—A set of computer symbols, with rules for their combination. They form a code to express information with fewer symbols and rules than there are distinct expressible meanings.

language converter—A data-processing device designed to change one form of data, i.e., microfilm, strip chart, etc., into another (punch card, paper tape, etc.).

language translation—The transformation of information from one language to another.

L-antenna—An antenna consisting of an elevated horizontal wire to which a vertical lead is connected at one end.

lanyard—A device which is attached to certain quick-disconnect connectors and which permits uncoupling and separation of connector halves by a pull on a wire or cable.

lap dissolve—In motion pictures or television, simultaneous transition in which one scene is faded down and out while the next scene is faded up and in.

lapel microphone—A microphone worn on the user's clothing.

Laplace's law—The strength of the magnetic field at any given point due to any element of a current-carrying conductor is directly proportional to the strength of the current and the projected length of the element, and is inversely proportional to the square of the distance of the element from the point in question.

Laplace transform—A mathematical substitution the use of which permits the solution of a certain type of differential equation by algebraic means.

lapping—Bringing quartz crystal plates up to their final frequency by moving them over a flat plate over which a liquid abrasive has been poured.

lap winding—An armature winding in which opposite ends of each coil are connected to adjoining segments of the commutator so that the windings overlap.

lap wrap—Tape wrapped around an object in an overlapping condition.

large-scale integrated circuit—An integrated circuit which contains 100 gates or more on a single chip, resulting in an increase in the scope of the function performed by a single device.

large-signal characteristics—The characteristics of an amplifier when rated (full) output signals are produced.

large-signal, d-c current gain—The d-c output current of a transistor with the d-c output circuit shorted, divided by the d-c input current producing the d-c output current.

large-signal power gain—The ratio of the a-c output power to the a-c input power under specified large-signal conditions. Usually expressed in decibels (db).

large-signal, short-circuit, forward-current transfer ratio—In a transistor, the ratio under specified test conditions of a change in output current to the corresponding change in input current.

Larmor orbit—The path of circular motion of a charged particle in a uniform magnetic field. The motion of the particle is unimpeded in the direction of the magnetic field, but motion perpendicular to the direction of the field is always accompanied by a force perpendicular to the direction of motion and the field.

laryngaphone—Also called a throat microphone. A microphone applied to the throat of a speaker to pick up voice vibrations directly. It is very useful in noisy locations because it picks up only the speaker's voice—no outside noises.

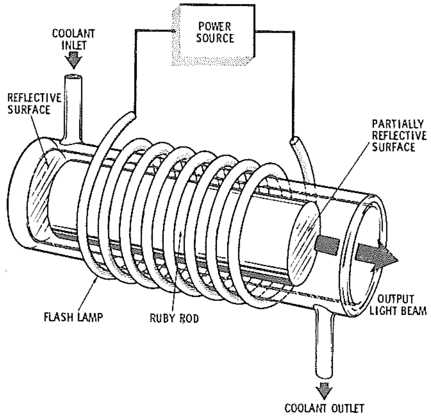

Laser.

laser—A device for transforming incoherent light of various frequencies of vibration into a very narrow, intense beam of coherent light. The name is derived from the initial letters of "Light Amplification by Stimulated Emission of Radiation." In the emission of ordinary light the molecules or atoms of the source emit their radiation independently of each other, and consequently there is no definite phase relationship among the vibrations in the resultant beam. The light is incoherent. The laser, by means of an optical resonator, forces the atoms of the material of the resonator to radiate in phase. The emitted radiation is stimulated by the excitation of atoms to a higher energy level by means of energy supplied to the device.

In the microwave region, the corresponding device is called a maser, and hence the laser is also known as a light maser.

laser cavity—An optical resonant and hence mode-selecting low-loss structure in which laser action occurs through the build-up of electromagnetic field intensity upon multiple reflection.

laser diode—*See* Diode Laser.

laser ranger—A device similar to conventional radar but using high intensity light rather than microwaves.

laser welding—A method of welding in which material heating is accomplished by concentration of a beam of coherent light on the area until fusion of the materials takes place.

lasing—In a laser, the generation of visible or infrared light waves that have very nearly a single frequency, by pumping or exciting electrons into high-energy states.

latching relay—A relay with contacts that lock in either the energized or de-energized positions, or both, until reset either manually or electrically.

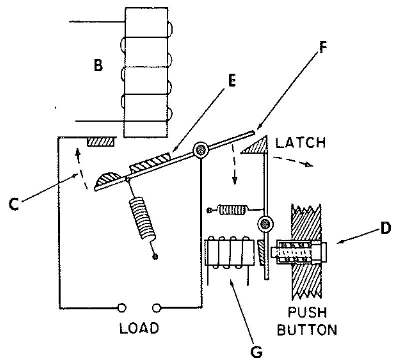

Latch-in relay.

latch-in relay—Also called locking relay. A relay with contacts which remain energized or de-energized until reset manually or electrically.

latch mode—A mode of operation for a storage circuit in which all encoder contact closures, even momentary ones, are latched "on."

latch-up—1. A condition in which the collector voltage in a given circuit does not return to the supply voltage when a transistor is switched from saturation to cutoff. Instead, the collector finds a stable operating point in the avalanche region of the collector characteristics. 2. An unintended stable circuit mode which will not revert to a previous intended circuit mode after removal of a stimulus such as a spurious signal or radiation. The effect is usually caused by parasitic circuit elements.

late contacts—In a relay, contacts that open or close after other contacts when the relay operates.

latency—1. The time required by a digital computer to deliver information from its memory. 2. In a serial-storage system, the access time minus the word time.

latent image—A stored image (e.g., the one contained in the charged mosaic capacitance in an iconoscope).

lateral chromatic aberration—Aberration which affects the sharpness of images off the axis. This occurs because different colors produce different magnifications.

lateral compliance—The force required to move the reproducing stylus from side to side as it follows the modulation on a laterally recorded record.

lateral-correction magnet—In a three-gun picture tube, an auxiliary component used for positioning the blue beam horizontally so that beam convergence will be obtained. It operates on the principle of magnetic convergence and is used in conjunction with a set of pole pieces mounted on the focus element of the blue gun.

lateral forced-air cooling—A method of heat transfer which employs a blower to produce side to side circulation of air through or across the heat dissipators.

lateral recording—A mechanical recording in which the groove modulation is perpendicular to the direction of motion of the recording medium and parallel to its surface.

lattice—1. In navigation, a pattern of identifiable intersection lines placed in fixed positions with respect to the transmitters that establish it. 2. The geometrical arrangement of atoms in a crystalline material.

lattice network—A network composed of four branches connected in series to form a mesh. Two nonadjacent junction points serve as input terminals, and the remaining two as output terminals.

lattice structure—In a crystal, a stable arrangement of atoms and their electron-pair bonds.

lattice-wound coil—*See* Honeycomb Coil.

launch complex—The entire launch, control, and support system required for launching rockets.

launching—The transferring of energy from a coaxial cable or shielded paired cable into a waveguide.

lavalier microphone—A small microphone suspended from the neck by a cord or wire. Used mainly in broadcasting and public-address applications, it frees the speaker's hands.

LAWEB—Weather bulletins issued every 6 hours. They are given in layman's language from ship and shore positions along the Great Lakes during the navigation season. Part 1 is from land stations, and Part 2 is from a ship four or more miles off shore with the ship's position given.

lawn mower—In facsimile, a term often used when referring to a helix-type recorder mechanism.

law of electric charges—Like charges repel; unlike charges attract. (*Also see* Coulomb's Law.)

law of electromagnetic induction—*See* Faraday's Laws.

law of electromagnetic systems—1. Any two circuits carrying currents tend to dispose themselves in such a way that the flux of magnetic induction linking the two will be a maximum. 2. Every electromagnetic system tends to change its configuration so that the flux of magnetic induction will be a maximum.

law of electrostatic attraction—*See* Coulomb's Law.

law of magnetism—Like poles repel; unlike poles attract.

law of normal distribution—The Gaussian law of the frequency distribution of any normal, repetitive function. It describes the probability of the occurrence of deviants from the average.

law of reflection—The angle of reflection is equal to the angle of incidence—i.e., the incident, reflected, and normal rays all lie in the same plane.

laws of electric networks—*See* Kirchhoff's Laws.

lay—Pertaining to wire and cable, the axial distance required for one cable conductor or conductor strand to complete one revolution about the axis around which it is cabled.

layer—The consecutive turns of a coil lying in a single plane.

layer-to-layer adhesion—The tendency for adjacent layers of recording tape in a roll to adhere, particularly after prolonged storage under conditions of high temperature and/or humidity.

layer-to-layer signal transfer—The magnetization of a layer of tape in a roll by the field from a nearby recorded layer. The magnitude of the induced signal tends to increase with storage time and temperature, and to decrease after the tape is unwound. These changes are a function of the magnetic instability of the oxide.

layer winding—A coil-winding method in which adjacent turns are placed side by side and touch each other. Additional layers may be wound over the first and are usually separated by sheets of insulation.

layout—1. Diagram indicating the positions of parts on a chassis or panel. 2. The actual positions of the parts themselves.

L-band—A radio-frequency band of 390 to 1,550 MHz and corresponding wavelengths of 77 to 19 cm.

LC product—Inductance (L) in henrys multiplied by capacitance (C) in farads.

L/C ratio—Inductance in henrys, divided by capacitance in farads.

LCS—Abbreviation for loudness-contour selector.

LCVD—*See* Least Voltage Coincidence Detection.

L-display—Also called L-scan. A radar display in which the target appears as two horizontal blips to the right and left of a central vertical time base.

lead—1. A wire to or from a circuit element. 2. To precede (the opposite of lag).

lead-acid cell—Also called lead cell. A cell in an ordinary storage battery. It consists of electrodes (plates) immersed in an electrolyte of dilute sulphuric acid. The electrodes contain certain lead oxides that change their composition as the cell is charged or discharged.

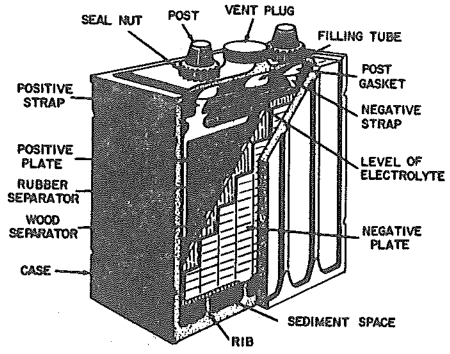

Lead-acid cell.

lead cell—*See* Lead-Acid Cell.

lead-covered cable—A cable with a lead sheath. The sheath offers protection from the weather and mechanical damage to wires contained.

leader cable—A navigational aid in which the path to be followed is defined by a magnetic field around a cable.

leader tape—Also called timing tape. Plain nonmagnetic tape for splicing to either end of magnetic tape to facilitate threading and preserve recorded material, or for splicing between recorded tapes to separate selections or provide pauses.

lead-in—The conductor that provides the path for r-f energy between the antenna and the radio/television receiver or transmitter.

leading blacks—Also called edge effect. In a television picture, the condition where the edge preceding a white object is overshaded toward black (i.e., the object appears to having a preceding, or leading, black border).

leading current—Current that reaches maximum before the voltage that produces it does. A leading current flows in any predominantly capacitive circuit.

leading edge—The major portion of the rise of a pulse.

leading-edge pulse time—The time required by a pulse to rise from its instantaneous amplitude to a stated fraction of its peak amplitude.

leading ghost—A twin image appearing to the left of the original in a televised picture.

leading load—A predominantly capacitive load—i.e., one in which the current leads the voltage.

lead-in groove—Also called a lead-in spiral. A blank spiral groove around the outside of the record. Its pitch is usually much greater than the other grooves and is used to quickly lead the needle into the beginning of the recorded groove.

leading whites—Also called edge effect. In a television picture, the condition where the edge preceding a black object is shaded toward white (i.e., the object appears to have a preceding, or leading, white border).

lead-in insulator—A tubular insulator through which cables or wires are brought inside a building.

lead-in spiral—*See* Lead-in Groove.

lead-in wires—Wires which carry current into a building (e.g., from an antenna).

lead network—A network, either a-c or d-c, designed to provide error-rate damping in the controlling device of a servo system.

lead-out groove—Also called a throw-out spiral. A blank spiral groove on the inside of a recording disc, next to the label. It is generally much deeper than the recording groove and is connected to either the locked or eccentric groove.

lead-over groove—Also called a crossover spiral. On disc records containing several selections, the groove in which the needle travels as it crosses from one selection to the next.

lead polarity of a transformer—Also called polarity. A designation of the relative instantaneous directions of currents in the leads of the transformer. Primary and secondary leads are said to have the same polarity when, at a given instant, the current enters the primary lead in question and leaves the secondary lead in question in the same direction as though the two leads formed a continuous circuit. The lead polarity of a single-phase distribution or power transformer may be either additive or subtractive. If one pair of adjacent leads from the two windings in question is connected together and voltage is applied to one of the windings, a. the lead polarity is additive if the voltage across the other two leads of the windings in question is greater than that of the higher-voltage winding alone. b. the lead polarity is subtractive if the voltage across the other two leads of the winding in question is less than that of the higher-voltage winding alone.

lead screw—1. In recording, a threaded rod which guides the cutter or reproducer across the surface of a disc. 2. In facsimile, a threaded shaft which moves the scanning mechanism or drum lengthwise.

leakage—1. Electrical loss from poor insulation. 2. Undesired flow of electricity over or through an insulator. 3. The portion not utilized most effectively in a magnetic field (e.g., at the end pieces of an electromagnet).

leakage coefficient—Ratio of total to useful flux produced in the neutral section of a magnet.

leakage current—1. An undesirable small value stray current which flows through (or across the surface of) an insulator or the dielectric of a capacitor. 2. A current which flows between two or more electrodes in a tube other than across the interelectrode space.

leakage flux—The flux which does not pass through the air gap, or useful part, of the magnetic circuit.

leakage inductance—A self-inductance due to the leakage flux generated in the winding of a transformer.

leakage power—In TR and pre-TR tubes, the radio-frequency power transmitted through a fired tube.

leakage radiation—Spurious radiation in a transmitting system—i.e., radiation from other than the system itself.

leakage reactance—The reactance represented by the difference in value between two mutually coupled inductances when their fields are aiding and then opposing.

leakage resistance—The normally high resistance of the path over which leakage current flows.

leakance—The reciprocal of insulation resistance.

leaky—Usually applied to a capacitor in which the resistance has dropped so far below normal that leakage current flows.

leaky waveguide—A waveguide with a narrow longitudinal slot, permitting a continuous energy leak.

leaky waveguide antenna—An antenna constructed from a long waveguide with radiating elements along its length. It has a very sharp pattern.

leapfrogging—The process of phasing, or delaying, the ranging pulse of a tracking radar in order to move, or shift (on the scope presentation), the tracking gate (at target blip) past the target blip from another radar.

"leapfrog" test—Slang for a program designed to uncover malfunctions in a digital computer.

least maximum deviation—A manner of expressing nonlinearity as a deviation from a straight line for which the deviations for proportional or normal linearity are minimized.

least mechanical equivalent of light—The radiant power that is contained in one lumen at the wavelength of maximum visibility. It is equal to 1.46 milliwatts at a wavelength of 555 millimicrons.

least voltage coincidence detection—A system that provides protection against interfering signals by blocking all signals except those having a pulse-repetition frequency the same as or some exact multiple of the radar-set pulse-repetition frequency. Abbreviated LCVD.

Lecher line—*See* Lecher Wire.

Lecher oscillator—A device for producing standing waves on two parallel wires called Lecher wires.

Lecher wire—Also called Lecher line. Two parallel wires on which standing waves are set up, usually for the measurement of wavelength.

Leclanche cell—An ordinary dry cell. It is a primary cell with a positive electrode of carbon and a negative electrode of zinc in an electrolyte of sal ammoniac and a depolarizer of manganese dioxide.

left-handed polarized wave—Also called counterclockwise-polarized wave. An elliptically polarized transverse-electromagnetic wave in which the electric intensity vector rotates counterclockwise (looking in the direction of propagation).

left-hand rule—*See* Fleming's Rule.

left-hand taper—The greater resistance in the counterclockwise half of the operating range of a rheostat or potentiometer than in the clockwise half (looking from the shaft end).

left (or right) signal—The electrical output of a microphone or combination of microphones placed so as to convey the intensity, time, and location of sound originating predominantly to the listener's left (or right) of the center of the performing area.

left (or right) stereophonic channel—The left (or right) signal as electrically reproduced in the reception of f-m stereophonic broadcasts.

leg—A section or branch of a component or system (e.g., one of the windings of a transformer).

legend—A table of symbols or other data placed on a map, chart, or diagram to assist the reader in interpreting it.

Lenard rays—Cathode rays that emerge from a special vacuum tube through a thin glass window or metallic foil.

Lenard tube—An electron tube in which the beam can be taken through a section of the wall of the evacuated enclosure.

length of a scanning line—1. The length of the path traced by the scanning or recording spot as it moves from line to line. 2. On drum-type equipment, the circumference of the drum. 3. The spot speed divided by the scanning-line frequency.

lens—1. An optical device which focuses light by refraction. 2. An electrical device which focuses microwaves by refraction or diffraction. 3. An acoustic device which concentrates sound waves by refraction. 4. An electronic optical device which focuses electrons.

lens disc—A television scanning disc having a number of openings arranged in a spiral, with a lens set into each opening.

lens speed—The amount of light a lens will pass. It is equal to the focal length divided by the diameter of the lens.

lens turret—On a camera, an arrangement which accommodates several lenses and can be rotated to facilitate their rapid interchange.

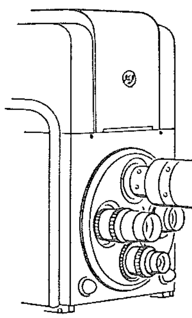

Lens turret.

Lenz's law—The current induced in a circuit due to a change in the magnetic flux through it or to its motion in a magnetic field is so directed as to oppose the change in flux or to exert a mechanical force opposing the motion. If a constant current flows in a primary circuit A and if by motion of A or the secondary circuit, B, a current is induced in B, the direction of the induced current will be such that, by its electromagnetic action on A, it tends to oppose the relative motion of the circuits.

Lepel discharger—A quenched spark gap used in early radio-telegraph transmitters employing shock excitation.

let-through current—The current that actually passes through a circuit breaker under short-circuit conditions.

level—The magnitude of a quantity in relation to an arbitrary reference value. Level normally is stated in the same units as the quantity being measured (e.g., volts, ohms, etc.). However, it may be expressed as the ratio to a reference value (e.g., db—as in blanking level, transmission level, etc.).

level above threshold—Also called sensation level. The pressure level of a sound in decibels above its threshold of audibility for the individual listener.

level compensator—1. An automatic gain control which minimizes the effect of amplitude variations in the received signal. 2. A device that automatically controls the gain in telegraph-receiving equipment.

level indicator—A sound-volume indicator.

lever switch—Commonly referred to as a key lever or lever key. A hand-operated switch for rapidly opening and closing a circuit.

Leyden jar—The original capacitor. It consists of metal foil sheets on the inside and outside of a glass jar. The foil serves as the plates and the glass as the dielectric.

lf—Abbreviation for low frequency—(i.e., between 30 and 300 kHz).

LFM—A vhf fan-type marker. It is low powered (5 watts) and has a range of only 10 miles or less.

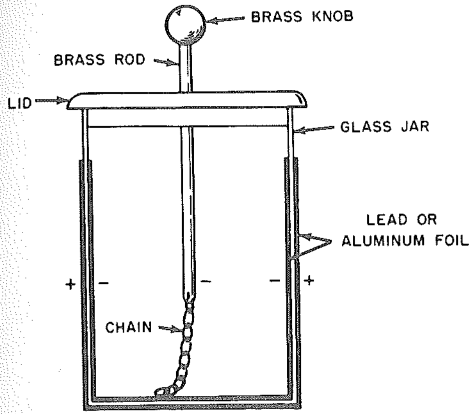

Leyden jar.

library—In computer programming, a collection of standard and fully tested routines and subroutines by which problems and parts of problems may be solved.

Lichtenberg figure camera—Also called klydonograph or surge-voltage recorder. A device for indicating the polarity and approximate crest value of a voltage surge by the appearance and dimensions of the Lichtenberg figure produced on a photographic plate or film. The emulsion coating of the plate or film contacts a small electrode coupled to the circuit in which the surge occurs. The film is backed by an extended plane electrode.

lie detector—Also called a polygraph. An electronic instrument which measures the blood pressure, temperature, heart action, breathing, and skin moisture of the human body. Abrupt or violent changes in these variables indicate the subject is not telling the truth.

life—The expected number of full excursions over which a transducer would operate within the limits of the applicable specification.

life test—The test of a component or unit under the conditions which approximate, or stimulate by acceleration, a normal lifetime of use. The test is performed to determine life expectancy or reliability throughout a predetermined life expectancy.

lifetime—The average time interval between the introduction and recombination of minority carriers in a semiconductor.

lifting magnet—A powerful electromagnet used on the end of a crane to lift iron and steel objects. They can be dropped instantly by merely cutting off the current.

light—1. Radiant energy within the wavelength limits perceptible by the average human eye (roughly, between 400 and 700 millimicrons). Although ultraviolet and infrared emissions will excite some types of photocells, they are usually not considered light. 2. In combination with other terms, a device used as a source of luminous energy (e.g., a pilot light).

light-activated silicon controlled rectifier—A silicon controlled rectifier with the ability to be triggered by current to the gate terminal, or by a light flashed on a light-sensitive part of the semiconductor pellet.

light-activated switch—A semiconductor diode which is triggered into conduction by light irradiation of a light-sensitive part of the semiconductor pellet.

light amplifier—A solid-state amplifier using photoconductive and electroluminescent films.

light-beam instrument—An instrument in which a beam of light is the indicator.

light-beam pickup—A phonograph pickup utilizing a beam of light as a coupling element of the transducer.

light chopper—A device for interrupting a light beam. It is frequently used to facilitate amplification of the output of a phototube on which the beam strikes.

light-dimming control—A circuit, often employing a saturable reactor, used to control the brightness of the lights in theaters, auditoriums, etc.

light-emitting diode—A PN junction that emits light when biased in the forward direction.

light flux—*See* Luminous Flux.

light gun—A photoelectric cell used by computer operators to take specific actions in assisting and directing computer operation. So called because of its gunlike case.

lighthouse tube—1. An ultrahigh-frequency

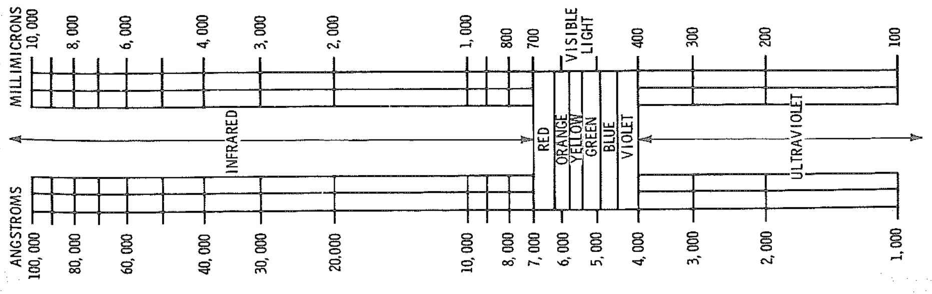

Light (spectrum).

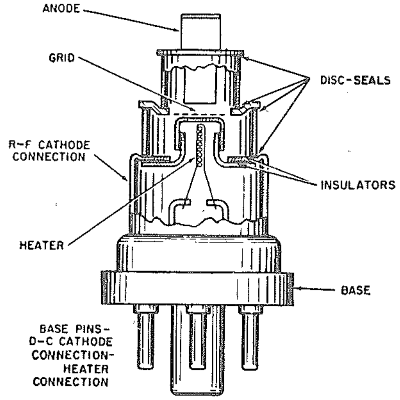

Cutaway view of lighthouse tube.

electron tube shaped like a lighthouse and having disc-sealed planar elements. 2. *See* Disc-Seal Tube.

lighting outlet—An outlet for direct connection of a lamp holder, lighting fixture, or pendant cord terminating in a lamp holder.

light load—A fraction of the total load the device is designed to handle.

light microsecond—The unit for expressing electrical distance. It is the distance over which light travels in free space in one microsecond (i.e., about 983 feet, or 300 meters).

light modulation—Variation in the intensity of light, usually at audio frequencies, for communications or motion-picture sound purposes.

light modulator—The device for producing the sound track on a motion-picture film. It consists of a source of light, an appropriate optical system, and a means for varying the resulting light beam (such as a galvanometer or light valve).

light negative—Having a negative photoconductivity—i.e., decreasing in conductivity when subjected to light.

lightning arrester—A device containing spark gaps, which allow lightning currents to flow into the earth.

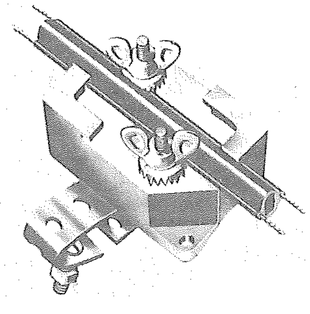

Lightning arrester.

lightning generator—A generator of high-voltage surges (e.g., for testing insulators).

lightning rod—A rod projecting above the highest point on a structure and connected to ground in such a way that it can carry a lightning discharge to ground.

lightning surge—A transient disturbance in an electric circuit caused by lightning.

lightning switch—A switch for connecting a radio antenna to ground during electrical storms.

light pen—A small photocell or photomultiplier in a pen-shaped housing, which is held against a cathode-ray screen so as to detect the very instant the electron beam goes through that particular location during its scanning sweep.

light pipe—An optical transmission element in which unfocused transmission and reflection are used to reduce photon losses.

light-positive—Having positive photoconductivity—i.e., increasing in conductivity when subjected to light.

light ray—A very thin beam of light.

light relay—A photoelectric device that opens or closes a relay when the intensity of a light beam changes.

light sensitive—Exhibiting a photoelectric effect when irradiated (e.g., photoelectric emission, photoconductivity, and photovoltaic action).

light-sensitive tube—A vacuum tube that changes its electrical characteristics with the amount of illumination.

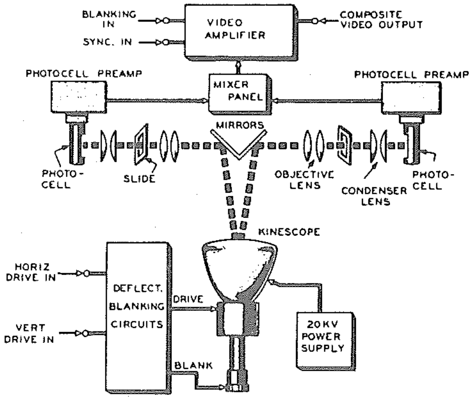

Light-spot scanner.

light-spot scanner—Also called a flying-spot scanner. A television camera in which the source of illumination is a spot of light that scans the scene to be televised. The picture signal is generated in a phototube, which picks up light either transmitted through the scene or reflected from it.

light valve—A device the light transmission of which can be varied in accordance with an externally applied electrical quantity such as voltage, current, an electric or magnetic field, or an electron beam.

light-year—The distance traveled by light in one year, or about 5,880,000,000,000 miles.

limit bridge—A form of Wheatstone bridge used for rapid routine production testing. Conformity with tolerance limits, rather than exact value, is determined.

limited signal—In radar, a signal that is intentionally limited in amplitude by the dynamic range of the system.

limited space-charge accumulation—A mode of oscillation for gallium arsenide diodes.

limited stability—The property of a system which remains stable only as long as the input signal falls within a particular range.

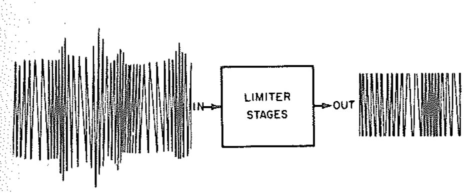

Limiter action.

limiter—1. A device in which some characteristic of the output is automatically prevented from exceeding a predetermined value—e.g., a transducer in which the output amplitude is substantially linear (with regard to the input) up to a predetermined value and substantially constant thereafter. 2. A radio-receiver stage or circuit that limits the amplitude of the signals and hence keeps interfering noise low by removing excessive amplitude variations from the signals.

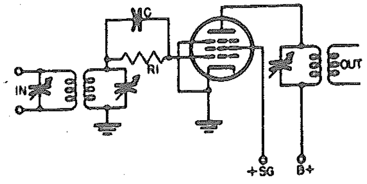

Limiter circuit.

limiting—The restricting of the amplitude of a signal so that interfering noise can be kept to a minimum.

limiting resolution—In television, a measure of the resolution usually expressed in terms of the maximum number of lines per picture height discriminated on a test chart.

limits—The minimum and maximum values specified for a quantity.

limit switch—A mechanically operated contact-making or -breaking device mounted in the path of a moving object and actuated by its passage.

line—1. In television, a single trace of the electron beam from left to right across the screen. The present United States standard is based on 525 lines to a complete picture. 2. A conductor of electrical energy. 3. The path of the moving spot in a cathode-ray tube 4. A term used interchangeably for maxwell.

line advance—Also called line feed. The distance between the centers of the scanning lines.

line amplifier—An amplifier that supplies a program transmission line or system with a signal at a specified level.

linear—Having an output that varies in direct proportion to the input.

linear acceleration—The rate of change in linear velocity.

linear accelerator—A device for speeding up charged particles such as protons. It differs from other accelerators in that the particles move in a straight line instead of in circles or spirals.

linear accelerometer—A transducer for measuring linear accelerations.

linear actuator—An actuator which produces mechanical motion from electrical energy.

linear amplification—Vacuum-tube amplification in which the plate current changes in direct proportion to the applied grid voltage.

linear amplifier—An amplifier that develops an output directly proportionate in amplitude to that of the input signal.

linear array—An antenna array in which the elements are equally spaced and in a straight line.

linear circuit—A circuit the output of which is an amplifier version of its input or a predetermined variation of its input.

linear control—A rheostat or potentiometer having uniform distribution of graduated resistance along the entire length of its resistance element.

linear detection—Detection in which the output voltage is substantially proportionate to the input voltage over the useful range of the detector.

linear detector—A detector that produces an output signal directly proportionate in amplitude to the variations in amplitude (for a-m transmission) or frequency (for f-m transmission) of the r-f input.

linear differential transformer—A type of electromechanical transducer that converts physical motion into an output voltage, the phase and amplitude of which are proportional to position. *Also see* Linear Motion Transducer.

linear distortion—Amplitude distortion in which the output- and input-signal envelopes are not proportionate, but no alien frequencies are involved.

linear electrical parameters of a uniform line—Frequently called the linear electrical constants. The series resistance and inductance, and the shunt conductance and capacitance, per length of a line.

linear electron accelerator—An evacuated metal tube in which electrons are accelerated through a series of small gaps (usually cavity resonators in the high-frequency range). The gaps are so spaced that, at a specific excitation frequency, the electrons gain additional energy from the electric field as they pass through successive gaps.

linear feedback-control system—A feedback-control system in which the relationship between the pertinent measures of the system signals are linear.

linearity—1. The relationship existing between two quantities when a change in a second quantity is directly proportionate to a change in the first quantity. 2. Deviation from a straight-line response to an input signal. 3. The ability of a meter to provide equal angular deflections proportional to the applied current. Usually expressed as a percent of the full-scale deflection.

linearity control—A manual control for adjusting the scanning waveshapes in television receivers.

linearity error—The deviation of a calibration curve from a specified straight line.

linear logarithmic intermediate-frequency amplifier—An amplifier used to avoid overload or saturation as a protection against jamming in a radar receiver.

linearly polarized wave—At a point in a homogeneous isotropic medium, a transverse electromagnetic wave the electric field vector of which lies along a fixed line.

linear magnetostriction—Under stated conditions, the relative change of length of a ferromagnetic object in the direction of magnetization when the magnetization of the object is increased from zero to a specified value (usually saturation).

linear mobility—The synchronized incremental mobility of functionally transitional electrons in a semiconductor.

linear modulation—Modulation in which the amplitude of the modulation envelope (or the deviation from the resting frequency) is directly proportionate to the amplitude of the modulating wave at all audio frequencies.

linear modulator—A modulator in which the modulated characteristic of the output wave is substantially linear with respect to the modulating wave for a given carrier magnitude.

linear motion transducer—An instrumentation component that translates straight-line (linear) mechanical motion into an a-c analog which is usable as a feedback signal for control or display. A transformer-type device in which a movable magnetic core is displaced axially by the moving component being monitored. When the core is moved in one direction from the center of its stroke, the output voltage is in phase with the excitation voltage, and when the core is moved in the opposite direction from the center, the output voltage is 180° out of phase. At the center, the output voltage is (virtually) zero. In either direction from the center, the voltage increases as a precise linear function of probe displacement. Thus the output signal has two basic analog components; phase relationship wth the excitation voltage, indicating the direction of travel; and voltage amplitude, indicating the length of travel.

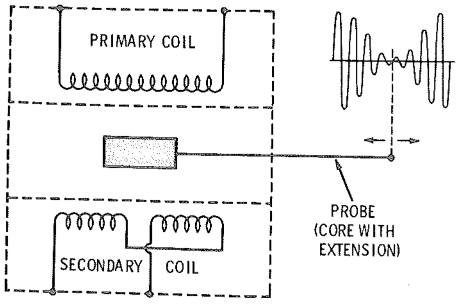

Linear motion transducer.

linear polarization—The polarization of a wave radiated by an electric vector that does not rotate but that alternates so as to describe a line. Normally the vector is oriented either horizontally or vertically.

linear power amplifier—A power amplifier in which the output voltage is directly proportionate to the input voltage.

linear programming—In computers, a mathematical method of sharing a group of limited resources among a number of competing demands. All decisions are interlocking because they must be made under a common set of fixed limitations.

linear pulse amplifier—A pulse amplifier which maintains the peak amplitudes of the input and output pulses in proportion.

linear rectification—The production, in the rectified current or voltage, of variations that are proportionate to variations in the input-wave amplitude.

linear rectifier—A rectifier with the same output-current or -voltage waveshape as that of the impressed signal.

linear scan—A radar beam which traverses only one arc or circle.

linear scanning—Scanning in which a radar beam generates only one arc or circle.

linear sweep—In a television receiver, the spot sweeps across the screen at a uniform velocity. This is due to a periodic voltage which, when applied to the horizontal-deflection plates of the cathode-ray tube, increases linearly to maximum and then falls abruptly to zero.

linearsyn—A linear displacement pick-off of the differential-transformer type consisting of a coil assembly and a moveable magnetic core. Linear velocity units of high-coercive-force permanent magnetic cores which induce sizable d-c voltages while moving concentrically within shielded coils; the voltage varies linearly with the core velocity. (Sanborn Co.)

linear taper—A potentiometer which changes the resistance linearly as it is rotated through its range.

linear time base—In a cathode-ray tube, the

time base in which the spot moves at a constant speed along the time scale. This type of time base is produced by application of a sawtooth waveform to the horizontal-deflection plates of a cathode-ray tube.

linear transducer—A transducer for which the pertinent measures of all the waves concerned are related by a linear function (e.g., a linear algebraic differential, or integral equation).

linear variable-differential transformer—*See* Differential Transformer.

linear varying parameter network—A linear network in which one or more parameters vary with time.

linear velocity transducer—A transducer which produces an output signal proportionate to the velocity of single-axis translational motion between two objects.

line balance—1. The degree to which the conductors of a transmission line are alike in their electrical characteristics with respect to each other, other conductors, and ground. 2. Impedance equal to that of the line at all frequencies (e.g., in terminating a two-wire line).

line-balance converter—A device used at the end of a coaxial line to isolate the outer conductor from ground.

line circuit—In a telephone system, the relay equipment associated with each station connected to a dial or manual switchboard. The term is also applied to a circuit for interconnecting an individual telephone and a channel terminal.

line coordinate—In a matrix, a symbol (normally at the side) identifying a specific row of cells and, in conjunction with a column coordinate, a specific cell.

line cord—Also called a power cord. A two-wire cord terminating in a two-prong plug at the end that goes to the supply, and connected permanently into a radio receiver or other appliance at the other end.

line-cord resistor—An asbestos-enclosed, wire-wound resistance element incorporated into a line cord along with the two regular wires. It lowers the line voltage to the correct value for the series-connected tube filaments and pilot lamps of a universal a-c/d-c receiver.

line diffuser—A circuit used to produce small vertical oscillations of the spot on the screen of a television monitor or receiver to make the line structure of the image less noticeable to an observer close to the screen.

line drop—A voltage loss occurring between any two points in a power or transmission line. Such a loss, or drop, is due to the resistance, reactance, or leakage of the line. An example is the voltage drop between a power source and load when the line supplying the power has excessive resistance to the amount of current.

line-drop signal—A signal associated with a subscriber line on a manual switchboard.

line-drop voltmeter compensator—A device using a voltmeter to enable it to indicate the voltage at some distant point in the circuit.

line equalizer—An inductance and/or capacitance inserted into a transmission line to correct its frequency-response characteristics.

line-equipment balancing network—A hybrid network designed to balance filters, composite sets, and other line equipment.

line feed—*See* Line Advance.

line filter—1. A device containing one or more inductors and capacitors. It is inserted between a transmitter, receiver, or appliance and the power line to block noise signals. In a radio receiver, it prevents power-line noise signals from entering the receiver. In other appliances, it prevents their own electrical noises from entering the power line. 2. A filter associated with a transmission line. In some applications, line filter may imply a filter used to separate the speech frequencies. In other applications, it may imply directional separation, etc.

line-filter balance—A network designed to maintain phantom-group balance when one side of the group is equipped with a carrier system. Since the network must balance the phantom group for voice frequencies only, its configuration is much simpler than the filter it balances.

line-focus tube—An X-ray tube in which the focal spot is roughly a line.

line frequency—1. Also called horizontal line frequency or horizontal frequency. In television, the number of times per second the scanning spot crosses a fixed vertical line in the picture in one direction, including vertical-return intervals. 2. The frequency of the supply voltage.

line-frequency regulation—The change in output (current, voltage, or power) of a regulated power supply for a specified change in line frequency.

line group—The frequency spectrum occupied by a group of carrier channels as they are applied to a transmission facility.

line hit—An electrical interference causing the introduction of spurious signals on a circuit.

line hydrophone—A directional hydrophone consisting of a single straightline element, an array of adjacent electroacoustic transducing elements in a straight line, or the acoustic equivalent of such an array.

line impedance—The impedance measured across the terminals of a transmission line.

line interlace—*See* Interlaced Scanning.

line leakage—Resistance existing through the insulation between the two wires of a telephone-line loop.

line lengthener—A device for altering the electrical length of a waveguide or transmission line, but not its physical length or other electrical characteristics.

line level—The intensity of the signal at a certain position on a transmission line.

line load—Usually a percentage of maximum circuit capability to reflect actual use during a period of time (e.g., peak-hour line load).

line loop—An operation performed over a communication line from an input unit at one terminal to output units at a remote terminal.

line loop resistance—Also called loop resistance. The metallic resistance of the local loop.

line loss—The total of the various energy losses in a transmission line.

line microphone—A directional microphone consisting of a single straight-line element, an array of adjacent electro-acoustic transducing elements in a straight line, or the acoustical equivalent of such an array.

line noise—Noise originating in a transmission line.

line of force—In an electric or magnetic field, an imaginary line in the same direction as the field intensity at each point. Sometimes called a maxwell when used as a unit of magnetic flux.

line of propagation—The path over which a radio wave travels through space.

line of sight—1. The distance to the horizon from an elevated point, including the effects of atmospheric refraction. The line-of-sight distance for an antenna at zero height is zero. 2. A straight line between an observer or radar antenna and a target. 3. An unobstructed, or optical, path between two points. 4. The radio-propagation characteristic of a microwave.

line-of-sight coverage—The maximum distance for transmission above the highest usable frequency. Radio waves at those frequencies do not follow the curvature of the earth and are not reflected from the ionosphere, but go off into space and are lost.

line-of-sight stabilization—In shipboard or airborne radar, compensating for the roll and pitch by automatically changing the elevation of the antenna in order to keep the beam pointed at the horizon.

line of travel—The path followed by an electromagnetic wave from one point to another.

line oscillator—An oscillator in which the resonant circuit is a section of transmission line an integral number of quarter wavelengths in electrical length.

line pad—In radio broadcasting, a pad inserted between the final program amplifier and the transmitter to insure a constant load on the amplifier.

line printer—In computers, a high-speed printer which produces an entire line at one time. All characters of the alphabet are contained around the rim of a continuously rotated disc, and there are as many discs as there are characters in the line. The computer momentarily stops the discs at the right characters for each line, and stamps out an impression in a fraction of a second.

line pulsing—A method of pulsing a transmitter by charging an artificial line over a relatively long period, and then discharging it through the transmitter tubes at a shorter interval determined by the line characteristics.

line regulation—The change in output (current, voltage, or power) of a regulated power supply for a specified change in line voltage. It may be stated as a percentage of the specified output and/or an absolute value.

line relay—A relay activated by the signals on a line.

line-sequential color-television system—A color-television system in which the individual lines of green, red, and blue are scanned in sequence rather than simultaneously.

line spectrum—The spectrum of a periodic, discrete signal consisting of one or more frequencies. For example, a square wave is characterized by a fundamental and odd-order harmonics.

line-stabilized oscillator—An oscillator in which the frequency is controlled by using one section of a line as a sharply selective circuit.

line stretcher—A section of rigid coaxial line with telescoping inner and outer conductors that permit the section to be conveniently lengthened or shortened.

line switching—Also called circuit switching. A communications switching system which completes a circuit from sender to receiver at the time of transmission, as opposed to message switching.

line transformer—A transformer inserted into a system for such purposes as isolation, impedance matching, or additional circuit derivation.

line triggering—Triggering from the power-line frequency.

line unit—An electric device used in sending, receiving, and controlling the impulses of a teletypewriter.

line voltage—The voltage level of the main power supply to the equipment.

line-voltage regulator—A device that counteracts variations in the power-line voltage and delivers a constant voltage to the connected load.

linguistic—Pertaining to language or its study, including its origin, structure, phonetics, etc.

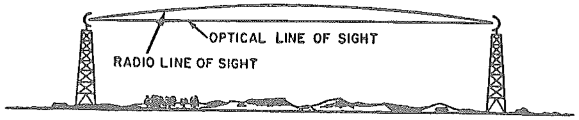

Line of sight.

link—1. A transmitter-receiver system connecting two locations. 2. In a digital computer, the part of a subprogram that connects it with the main program. 3. An interconnection. 4. In automatic switching, a path between two units of switching apparatus within a central office.

linkage—1. A measure of the voltage that will be induced in a circuit by magnetic flux. It is equal to the flux times the number of turns linked by the flux. 2. A mechanical arrangment for transferring motion in a desired manner. It consists of solid pieces with movable joints. 3. In a computer, a technique used to provide interconnections for entry and exit of a closed subroutine to or from the main routine.

link circuit—A closed loop used for coupling purposes. It generally consists of two coils, each having a few turns of wire, connected by a twisted pair of wires or by other means, with each coil placed over, near, or in one of the two coils to be coupled.

link coupling—Inductive coupling between circuits. A coil in one circuit acts as the primary, and a coil in the second circuit as the secondary.

link fuse—An unprotected fuse consisting of a short, bare wire between two fastenings.

link neutralization—Neutralization of a tuned radio-frequency amplifier by means of an inductive coupling loop from the output to the input.

link transmitter—In broadcasting, a booster for a remote pickup or from studio to main transmitter.

lin-log receiver—A radar receiver in which the amplitude response is linear for small-amplitude signals and logarithmic for large ones.

lip microphone—A sensitive microphone which is placed in contact with the lip.

liquid-borne noise—Undesired sound characterized by fluctuations of pressure of a liquid about the static pressure as a mean.

liquid cooling—The use of a circulating liquid to cool components or equipment that heat up during operation.

liquid-filled capacitor—A capacitor in which a liquid impregnant occupies substantially all of the case volume not required by the capacitor element and its connections. Space may be allowed for the expansion of the liquid under temperature variations.

liquid fuse unit—A fuse unit in which the fuse link is immersed in a liquid or the arc is drawn into the liquid when the fuse link melts.

liquid-impregnated capacitor—A capacitor in which a liquid impregnant is dominantly contained within the foil and paper winding, but does not occupy substantially all of the case volume not required by the capacitor element and its connections.

liquid rheostat—A rheostat consisting of metal plates immersed in a conductive liquid. The resistance is changed by raising or lowering the plates or the liquid level to vary the area of the plates contacting the liquid.

liser—A microwave oscillator of very high spectrum purity. Its emission consists of right circularly polarized waves of two different cavity resonant frequencies.

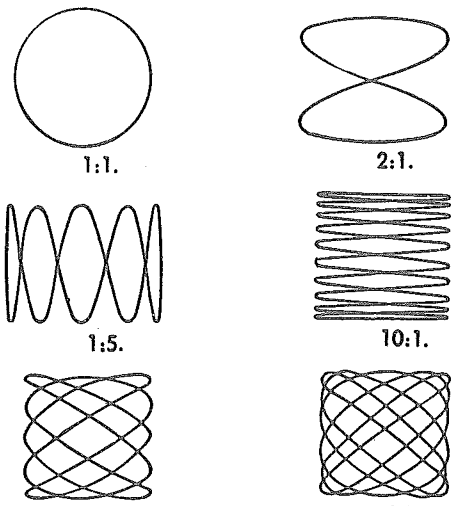

Lissajous figure.

Lissajous figures—Patterns produced on the screen of a cathode-ray tube when sine-wave signal voltages of various amplitude and phase relationships are applied to the horizontal- and vertical-deflection circuits simultaneously.

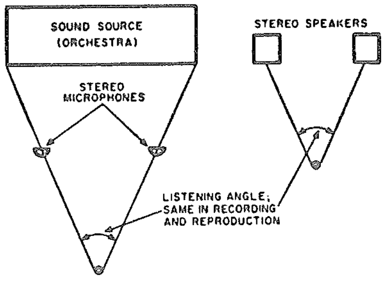

Listening angle.

listening angle—The enclosed angle between the listener and the two speakers of a stereo reproducing system.

listening sonar—*See* Sonar.

lithium—An alkali metal used in the construction of photocells.

litz wire—Also called Litzendraht wire. A conductor composed of a number of fine, separately insulated strands which are woven together so that each strand successively takes up all possible positions in the cross section of the entire conductor. Litz wire gives reduced skin effect; hence, lower resistance to high-frequency currents.

live—A term applied to a circuit through which current is flowing. *Also see* Energized.

live cable test cap—A protective cap placed over the end of a cable to insulate the cable and seal its sheath.

live end—The end of a radio studio where the reflection of sound is greatest.

live room—A room with a minimum of sound-absorptive material such as drapes, upholstered furniture, rugs, etc. Because of the many reflecting surfaces, any sound produced in the room will have a long reverberation time.

L-network—A network composed of two impedance branches in series. The free ends are connected to one pair of terminals, and the junction point and one free end are connected to another pair.

load—1. The power consumed by a machine or circuit in performing its function. 2. A resistor or other impedance which can replace some circuit element. 3. The power delivered by a machine. 4. A device that absorbs power and converts it into the desired form. 5. The impedance to which energy is being supplied. 6. Also called work. The material heated by a dielectric or induction heater. 7. In a computer, to fill the internal storage with information obtained from auxiliary or external storage.

load and go — In a computer, an operation and compiling technique in which the pseudo language is converted directly to machine language and the program is then run without the creation of an output machine-language program.

load balance—*See* Load Division.

load circuit—The complete circuit required to transfer power from a source to a load (e.g., an electron tube).

load-circuit efficiency—In a load circuit, the ratio between its input power and the power it delivers to the load.

load-circuit power input—The power delivered to the load circuit. It is the product of the alternating components of the voltage across the load circuit and the current passing through it (both root-mean-square values), times their power factors.

load coil—Also called a work coil. In induction heaters a coil which, when energized with an alternating current, induces energy into the item being heated.

load curve—A curve of power versus time—i.e., the value of a specified load for each unit of the period covered.

load divider—A device for distributing power.

load division — Also called load balance. A control function that divides the load in a prescribed manner between two or more power sources supplying the same load.

loaded antenna—An antenna to which extra inductance or capacitance has been added to change its electrical (but not its physical) length.

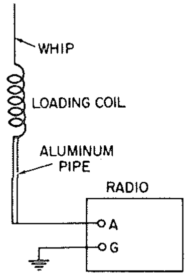

Center-loaded antenna.

loaded applicator impedance—In dielectric heating, the complex impedance measured at the point of application with the load material properly positioned for heating and at the specified frequency.

loaded impedance—In a transducer, the impedance at the input when the output is connected to its normal load.

loaded line — A transmission line that has lumped elements (inductance or capacitance) added at uniformly spaced intervals. Loading is used to provide a given set of characteristics to a transmission line.

loaded motional impedance—*See* Motional Impedance.

loaded *Q*—Also called the working *Q*. The *Q* of an electric impedance when coupled or connected under working conditions.

load factor—The ratio of average power demand, over a stipulated period of time, to the peak or maximum demand for that same interval.

load impedance—The impedance which the load presents to a transducer.

loading—1. In communication practice, the insertion of reactance into a circuit to improve its transmission characteristics. 2. Placing some material at the front or rear of a speaker so as to change its acoustic impedance and thus alter its radiation pattern.

loading coil—An inductor inserted into a circuit to increase its inductance and thereby improve its transmission characteristics.

loading-coil spacing—The line distance between the successive loading coils of a line.

loading disc—A metal disc placed on top of a vertical antenna to increase its natural wavelength.

loading error—The error introduced when more than negligible current is drawn from the output of a device.

loading noise—Any unwanted signal caused by fluctuating contact resistance between the slider and the wire or film in a potentiometric transducer when current is drawn from the instrument.

loading routine — In a computer, a routine which, when in the memory, is able to read other information into the memory from cards, tape, etc.

load isolator — A waveguide or coaxial device (usually ferrite) which provides a good energy path from a signal source to a load, but provides a poor energy path for reflections from a mismatched load back to the signal source.

load life—The ability of a device to withstand its full power rating over an extended period of time, usually expressed in hours.

load line—1. A straight line drawn across a series of plate-current-plate-voltage characteristic curves on a graph to show how plate current changes with grid voltage when a specified plate-load resistance is used. 2. A line drawn on the collector characteristic curves of a transistor on which the operating point of the transistor moves as the collector current changes. Called a load line because the slope of the line depends on the value of the collector load resistance.

load matching—In induction and dielectric heaters, adjustment of the load-circuit impedance so that the desired energy will be transferred from the power source to the load.

load-matching network—An electrical impedance network inserted between the source and the load to provide for maximum transfer of energy.

load-matching switch—In induction and dielectric heaters, a switch used in the load-matching network to alter its characteristics and thereby compensate for a sudden change in the load characteristics (such as in passing through the Curie point).

load point — The physical beginning of a magnetic tape recording as determined by a reflective spot attached to the tape at that point.

load regulation — In a power supply, the change in output for a specified change in load conditions. It may be stated as a percentage of the specified output or as an absolute value.

loadstone—*See* Lodestone.

load-transfer switch—A switch for connecting either a generator or a power source to one load circuit or another.

lobe — Also called directional, radiation, or antenna lobe. One of the areas of greater transmission in the pattern of a directional antenna. Its size and shape are determined by plotting the signal strength in various directions. The area with the greatest signal strength is known as the major lobe, and all others are called minor lobes.

lobe frequency—The number of times a lobing pattern is repeated per second.

lobe half-power width—In a plane containing the direction of the maximum energy of the lobe, the angle between the two directions in that plane about the maximum in which the radiation intensity is one-half the maximum value of the lobe.

lobe penetration — The penetration of the radar coverage of a station which is not limited by the pulse-repetition frequency, scope limitations, or the screening angle at the azimuth of penetration.

lobe switching—A form of scanning in which the maximum radiation or reception is periodically switched to each of two or more directions in turn.

lobing—The formation of maxima and minima at various angles of the vertical-plane antenna pattern by the reflection of energy from the surface surrounding the radar antenna. These reflections reinforce the main beam at some angles and detract from it at other angles, producing finders of energy.

local action—In a battery, the loss of otherwise usable chemical energy by currents which flow within regardless of its connections to an external circuit.

local battery—A battery made of single dry cells located at the subscriber's station.

local channel—1. In private line services, that portion of a through channel wthin an exchange which is provided to connect the main station with an interexchange channel. 2. A standard broadcast channel in which several stations, with powers not in excess of 250 watts, may operate.

local control—Control of a radio transmitter control directly at the transmitter, as opposed to remote control.

localizer—A radio which provides signals for guiding aircraft onto the center line of a runway.

localizer on-course line—A vertical line passing through a localizer. Indications of opposite sense are received on either side of the line.

local oscillator—Also called beating oscillator. An oscillator used in a superheterodyne circuit to produce a sum or difference frequency equal to the intermediate frequency of the receiver. This is done by mixing its output with the received signal.

local-oscillator injection — An adjustment used to vary the magnitude of the local-oscillator signal that is coupled into the mixer.

local-oscillator radiation—Radiation of the fundamental or harmonics of the local oscillator of a superheterodyne receiver.

local-oscillator tube — The vacuum tube which provides the local-oscillator signal in a superheterodyne receiver.

local program—A program originating at and released through only one broadcast station.

locating lug—The selection of two elements at specified attenuation points or in specified attenuation ranges so the ratio of the outputs of the elements at these points, calculated as a power ratio, do not exceed a specified limit.

location—A unit-storage position in the main or secondary storage of a computer.

location counter—In the control section of a computer, a register which contains the address of the instruction currently being executed.

locked groove—Also called concentric groove. The blank and continuous groove in the center, near the label, of a disc record. It prevents the needle from traveling farther inward, onto the label.

locked-rotor current—In a motor, the steady-state current taken from the line while the rotor is locked and the rated voltage (and frequency in alternating-current motors) is applied to the motor.

locked-rotor torque—Also called static torque. The minimum torque a motor will develop at rest, for all angular positions of the rotor, when the rated voltage is applied at rated frequency.

lock-in—1. The term used when a sweep oscillator is in synchronism with the applied sync pulses. 2. The shifting and automatic holding of one or both of the frequencies of two oscillating systems which are coupled together, so that the two frequencies have the ratio of two integral numbers.

locking—1. Controlling the frequency of an oscillator by means of an applied signal of constant frequency. 2. Automatic following of a target by a radar antenna.

locking circuit—*See* Holding Circuit.

locking-in—In two oscillators which are coupled together, the shifting and automatic holding of one or both of their frequencies so that the two frequencies are in synchronism (i.e., have the ratio of two integral numbers).

locking-out relay—An electrically operated hand or electrically reset device which functions to shut down and hold an equipment out of service on the occurrence of abnormal conditions.

locking relay—*See* Latch-in Relay.

lock-on—1. The instant at which radar begins to track a target automatically. 2. Signifies that a tracking or target-seeking system is continuously and automatically tracking a target in one or more coordinates.

lock-on range—The range from a radar to its target at the instant when lock-on occurs.

lockout—*See* Receiver Lockout System.

lock-up relay—A relay that is locked in the energized position magnetically or electrically rather than mechanically.

loctal base—*See* Loktal Base.

lodar—Also called lorad. A direction finder which compensates for night effect by observing the distinguishable ground- and sky-wave loran signals on a cathode-ray oscilloscope and positioning a loop antenna to obtain a null indication of the more suitable components.

lodestone—Also spelled loadstone. A natural magnet consisting chiefly of a magnetic oxide of iron called magnetite.

Loftin-White circuit—A type of direct-coupled amplifier circuit.

log—1. A listing of radio stations and their frequency, power, location, and other pertinent data. 2. A record of the station with which a radio station has been in communication. Amateur radio operators, as well as all commercial operators, are required by law to keep a log. 3. At a broadcast station, a detailed record of all programs broadcast by the station. 4. At a broadcast transmitter, a record of the meter readings and other measurements required by law to be taken at regular intervals. 5. Abbreviation for logarithm.

logarithmic amplifier—An amplifier the output of which is a logarithmic (as opposed to linear) function of its input.

logarithmic curve—A curve on which one coordinate of any point varies in accordance with the logarithm of the other coordinate of the point.

logarithmic decrement—1. For an exponentially damped alternating current, the natural logarithm of the ratio of the first to the second of two successive amplitudes having the same polarity. 2. *See* Damping Ratio.

logarithmic horn—A horn the diameter of which varies logarithmically with the length.

logarithmic scale—A scale on which the various points are plotted according to the logarithm of the number with which the point is labeled.

logarithmic scale meter—A meter having deflections proportional to the logarithms of the applied energies.

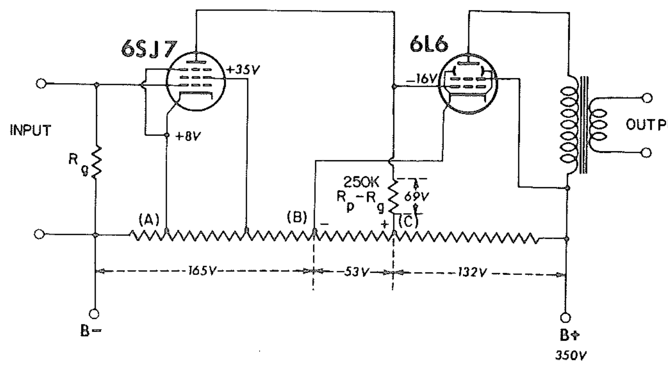

Loftin-White amplifier.

logger—An instrument which automatically scans conditions (temperature, pressure, humidity) and records (logs) the findings on a chart.

logic—1. The science dealing with the basic principles and applications of truth tables, switching, gating, etc. 2. *See* Logical Design. 3. Also called symbolic logic. A mathematical approach to the solution of complex situations by the use of symbols to define basic concepts. The three basic logic symbols are "and", "or", and "not." When used in Boolean algebra, these symbols are somewhat analogous to addition and multiplication.

logical choice—In a computer, the correct decision where alternatives or different possibilities are open.

logical design—1. The preplanning of a computer or data-processing system prior to its detailed engineering design. 2. The synthesizing of a network of logical elements to perform a specified function. 3. The result of 1 and 2 above, frequently called the logic of the system, machine, or network.

logical element—In a computer or data-processing system, the smallest building blocks which operators can represent in an appropriate system of symbolic logic. Typical logical elements are the AND gate and the "flip-flop."

logical operations—The operations in which logical quantities (yes-no decisions, quantities expressed as zeros and ones, etc.) make comparisons, decisions, and extractions.

logical threshold voltage—At the output of a logic device, the voltage level at which the following logic device switches states.

logic array—An integrated device in which 50 or more circuits are integral to a single silicon chip. In addition, the circuits are interconnected on the chip to form some electronic function at a higher level of organization than a single circuit. Logic arrays are constructed by the unit cell method in which a simple circuit (or function) is repeated many times on a slice. The interconnection pattern for converting groups of cells into large functions is determined after cell probe tests are completed. Each interconnection pattern may be unique to a single slice. In general, logic arrays are characterized by multiple levels of metallization to effect the large-scale function.

logic diagram—In logical design, a diagram representing the logical elements and their interconnections, but not necessarily their construction or engineering details.

logic symbol—A symbol used to represent a logical element on a graph.

log/linear preamplifier—A preamplifier the electrical output of which is proportional to the common logarithm of the input, but which can be switched to a different mode wherein the output is proportional to the input. Any amplifier the transfer characteristic or calibration curve of which can be arbitrarily switched from the form y — log x to the form y — x (y equals the output and x equals the input).

log-periodic antenna—A type of directional antenna that achieves its wideband properties by geometric iteration. The radiating element and the spacing between elements have dimensions that increase logarithmically from one end of the array to the other so that the ratio of element length to element spacing remains constant.

log receiver—A receiver in which the output amplitude is proportionate to the logarithm of the input amplitude.

log scan—A spectrum display in which the frequency axis is calibrated logarithmically.

loktal base—Also spelled loctal. An eight-pin base for small vacuum tubes. It is designed so that it locks the tube firmly in a corresponding eight-hole socket. Unlike in other tubes, the pins are sealed directly into the glass envelope.

loktal tube—Also spelled loctal. A vacuum tube with a loktal base.

long base-line system—A system in which the distance separating ground stations approximates the distance to the target being tracked.

long-distance navigational aid—A navigational aid usable beyond radio line of sight.

longevity—The period of time during which the failure rate of a group of components is essentially constant.

longitudinal chromatic aberration—A lack of sharpness in the image because different colors come to a focus at different distances from the lens.

longitudinal coils—The coils that are used in transmission lines to suppress longitudinal currents.

longitudinal current—A current which flows in the same direction in both wires of a parallel pair. The earth is its return path.

longitudinal magnetization—In magnetic recording, magnetization of the recording medium in a direction essentially parallel to the line of travel.

longitudinal redundance—In a computer, a condition in which the bits in each track or row of a record do not total an even (or odd) number. The term is generally used to refer to records on magnetic tape, and a system can have either odd or even longitudinal parity.

longitudinal wave—A wave in which the direction of displacement and propagation are the same at each point.

long-persistence screen—A fluorescent screen in which the light intensity of its spots does not immediately die out after the beam has moved on.

long-play record—Abbreviated LP record. Also called a microgroove record. A 10- or 12-inch record or transcription with finely-cut grooves which give it a long playing time.

long-pull magnet—An electromagnet designed to exert a practically uniform pull, for an extended range of armature movement. It consists of a conical plunger moving up and down inside a hollow core.

long-range radar—A radar installation capable of detecting targets 200 or more miles away.

long-range navigation — A long-range electronic navigation system which uses the time divergence of pulse-type transmission from two or more fixed stations.

long shunt—A shunt field connected across the series field and the armature, instead of directly across the armature alone, of a motor or generator.

long-tailed pair—A two-tube circuit in which decreased plate current through one tube results in increased plate current through the other tube, and vice versa.

long-term stability (or long-term instability)—The slow changes in average frequency arising from changes in an oscillator. Statements of long-term stability for quartz oscillators often term this characteristic "aging rate" and specify it as "parts per day" (fractional frequency change over 24 hours). For cesium standards, this term commonly refers to the total fractional frequency drift for the life of the cesium beam tube.

long throw—A method of speaker design in which the woofer moves freely through long excursions, providing excellent low-frequency response with low distortion.

long wave—Wavelengths longer than about 1,000 meters. They correspond to frequencies below 300 kHz.

long-wire antenna—A linear antenna which is much longer than the operating wavelength. For this reason, it can provide a directional pattern.

lookthrough—1. In jamming, sporadic interruption of the emission for extremely short periods in order to monitor the victim signal. 2. When a set is being jammed, the monitoring of the desired signal during lulls in the jamming signals.

loom—A flexible nonmetallic tubing placed around insulated wire for protection.

loop—*See* Mesh *and* Antinodes.

loop actuating signal—The signal derived from mixing the loop-input and loop-feedback signals.

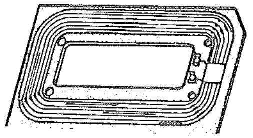

Loop antenna.

loop antenna—An antenna used in radio direction-finding apparatus and modern radio receivers. It consists of one or more loops of wire.

loopback test—A test in which signals are looped from a test center through a data set or loopback switch and back to the test for measurement. *Also see* Bussback.

loop control — The maintaining of a specified loop of material between two sections of a machine by automatically adjusting the speed of at least one of the driven sections.

loop difference signal—The output signal from a summing point of a feedback control loop. It is a specific type of loop actuating signal produced by a particular loop input signal applied to that summing point.

loop error—The desired value minus the actual value of the loop output signal.

loop error signal—The loop actuating signal, when it is the loop error.

loop feedback signal—The signal derived from the loop output signal and fed back to the mixing point for control purposes.

loop feeder—A feeder which follows along a circuit and distributes the voltage more evenly at different points.

loop gain—1. The total usable power gain of a carrier terminal or two-wire repeater. Because any closed system tends to sing (oscillate), its usable gain may be less than the sum of the enclosed amplifier gains. The maximum usable gain is determined by—and cannot exceed—the losses in the closed path. 2. The increase in gain observed when the feedback path of an amplifier is opened, but with all circuit loads intact. 3. In an operational amplifier circuit, the product of the transfer characteristics of all of the elements (active or passive) encountered in a complete trip around the loop, starting at any point and returning to that point.

looping — A computer operation in which a sequence of steps is repeated.

loop input signal—An external signal applied to a feedback control loop.

loop-mile—The length of wire in a mile of two-wire line.

loop output signal—The controlled signal extracted from a feedback control loop.

loop pulsing—Also called dial pulsing. The regular, momentary interruption of the direct-current path at the sending end of a transmission line.

loop resistance—*See* Line Loop Resistance.

loop return signal—The signal returned, via a feedback control loop, to a summing point in response to a loop input signal applied to that summing point. The loop return signal is a specified type of loop input signal and is subtracted from it.

loopstick antenna—*See* Ferrite-Rod Antenna.

loop test—A method of locating a fault in the insulation of a conductor when the conductor can be arranged to form part of a closed circuit.

loop transfer function—The transfer function of the transmission path. It is formed by opening and properly terminating a feedback loop. (*Also see* Transfer Function.)

loose coupler—An obsolete tuning system consisting of two coils, one inside the other. Coupling was varied over a wide range by sliding one coil over the other.

loose coupling — Also called weak coupling. A degree of coupling that is considerably less than critical coupling. Hence, there is very little transfer of energy.

Lorac—A trademark of Seismograph Service Corp. for a specific navigation system that determines a position fix by the intersection of lines of position. Each line is defined by the phase angle between two heterodyne beat-frequency waves. One wave is the beat frequency between two c-w signals from two widely spaced transmitters. The other is a reference wave of the same frequency and is obtained by deriving the heterodyne beat of the same two c-w signals at a fixed location and transmitting it, via a second radio-frequency channel, to the receiver being located.

lorad—*See* Lodar.

loran—A contraction of LOng-RAnge Navigation. A navigation aid sending out pulses at radio frequencies between 1,800 and 2,000 kHz. It defines lines of position which are based on the differences in travel time between radio waves from a master and a slave station. Airborne equipment utilizes a picture tube and measuring circuits to determine the time differences and relate them to time-difference lines drawn on a map. A navigation fix is established by the intersection of two or more lines of position.

loran-C — A hyperbolic navigation system that relies on the stability of electromagnetic propagational characteristics at a frequency of 100 kHz. Each loran-C station operates in a pulse mode in such a way that no two stations can be received simultaneously. During the transmission period of the master station, both slave stations remain silent. After a fixed delay time, slave X will transmit a similar pulse. After a second delay, slave Y will transmit a pulse.

loran line—A line of position on a loran chart where each line is the locus of points, the distances from two fixed stations of which differ by a constant amount.

loran station—A radionavigational land station transmitting synchronized pulses.

loran tables—Tables giving terrestrial coordinates (latitude and longitude) of loran lines of position and values of the sky-wave correction.

Lorentz force — The force exerted by an electric field and a magnetic field on a moving electric charge.

Lorentz force equation—An equation relating the force on a charged particle to its motion in an electromagnetic field.

lorhumb line—In navigation, a course line in a lattice; the derivation of one coordinate from the other is always equal to the ratio of the difference between coordinates at the beginning and ending of the course line.

loss—1. A decrease in power suffered by a signal as it is transmitted from one point to another. It is usually expressed in decibels. 2. Energy dissipated without accomplishing useful work.

loss angle—The complement of the phase angle of an insulating material.

losser—A circuit having less power in the output as compared to the power applied to the input. This term is particularly applicable to mixers; a crystal mixer is a losser.

losser circuit—A resonant circuit having sufficient high-frequency resistance to prevent sustained oscillation at the resonant frequency.

losser effect—Radiation resulting from the recombination of charge carriers injected in a PN or PIN junction biased in the forward direction.

Lossev effect—The resultánt radiation when charge carriers recombine after being injected into a forward-biased PN or PIN junction.

loss factor—The characteristic which determines the rate at which heat is generated in an insulating material. It is equal to the dielectric constant times the power factor.

loss index—The product of the power factor times the dielectric constant of a material.

lossless line—A theoretically perfect line—i.e., one that has no loss and hence transmits all the energy fed to it.

loss modulation—*See* Absorption Modulation.

loss tangent — Also called dielectric dissipation factor. The decimal ratio of the irrecoverable to the recoverable part of the electrical energy introduced into an insulating material by the establishment of an electric field in the material.

lossy—Insulating material which dissipates more than the usual energy.

lossy attenuator—A length of waveguide made from some dissipative material and used to deliberately introduce transmission loss.

lossy line—1. A transmission line designed with high attenuation. 2. Cable used in test measurements which has a large attenuation per unit length.

lot — A group of similar components which have been either all manufactured in a continuous production run from homogeneous raw materials under constant process conditions or assembled from more than one production run and submitted for random sampling and acceptance testing. Specifically, a quantity of material all of which was manufactured under identical conditions, and assigned an identifying lot number.

lot tolerance percent defective—A process average that will be rejected 90% of the time. Abbreviated LTPD.

loudness—The intensive attribute of an auditory sensation, in terms of which sounds may be ordered on a scale extending from soft to loud. Loudness depends primarily upon the sound pressure of the stimulus, but it also depends upon the frequency and waveform of the stimulus.

loudness contour — A curve showing the sound pressure required at each frequency to produce a given loudness sensation to a typical listener.

loudness-contour selector — Abbreviated as LCS. A circuit that alters the frequency response of an amplifier so that the characteristics of the amplifier will more closely match the requirements of the human ear.

loudness control—Also called compensated volume control. A combined volume and tone control which boosts the bass frequencies at low volume to compensate for the inability of the ear to respond to them. Some loudness controls provide similar compensation at the treble frequencies.

loudness level—The sound-pressure level of a 1,000-Hz tone judged by a listener to be as loud as the sound under consideration. It is measured in decibels relative to .0002 microbar.

loudspeaker—*See* Speaker.

loudspeaker dividing network—*See* Dividing Network.

loudspeaker impedance—*See* Speaker Impedance.

loudspeaker system—*See* Speaker System.

loudspeaker voice coil—*See* Speaker Voice Coil.

louver—The grill of a speaker.

low-angle radiation—Radiation that proceeds at low angles above ground.

low band—Television channels 2 through 6 covering frequencies between 54 and 88 MHz.

low-capacitance contacts—A type of contact construction providing small capacitance between contacts.

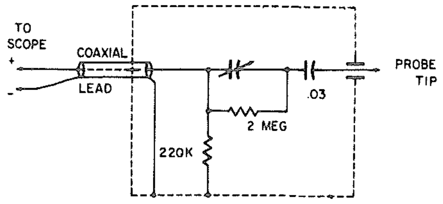

Low-capacitance probe circuit.

low-capacitance probe—A probe with very low capacitance. It is connected between the input of an oscilloscope and the circuit under observation.

low-corner frequency — The frequency at which the output of a resolver is 3 db below the mid-frequency value and the phase shift is 45°.

low-definition television—A television system employing less than 200 scanning lines per frame.

low-energy circuit—Also called a dry circuit. A circuit application that functions at low voltage (i.e., approximately 10 volts or less) and low current (i.e., approximately 1 ma or less).

low-energy criterion — *See* Von Hippel Breakdown Theory.

low-energy material — *See* Soft Magnetic Material.

lower manual—*See* Accompaniment Manual.

lower sideband—The lower of two frequencies or groups of frequencies produced by the amplitude-modulation process.

lowest effective power—The minimum product of the antenna input power in kilowatts and the antenna gain required for satisfactory communication over a particular radio route.

lowest useful high frequency—In radio transmission, the lowest frequency effective at a specified time for ionospheric propagation of radio waves between any two points, excluding frequencies below several megahertz. It is determined by such factors as absorption, transmitter power, antenna gain, receiver characteristics, type of service, and noise conditions.

low filter—An audio circuit designed to remove low-frequency noises from the program material. Such noises include turntable rumble, tone-arm resonance, etc.

low frequency—Abbreviated lf. The band of frequencies extending from 30 to 300 kHz (1,000 to 10,000 meters).

low-frequency compensation—A technique for extending the low-frequency response of a broad-band amplifier.

low-frequency impedance corrector — An electric network designed to be connected to a basic network, or to a basic network and a building network, so that the combination will simulate, at low frequencies, the sending-end impedance, including the dissipation, of a line.

low-frequency induction heater or furnace —A heater or furnace in which the charge is heated by inducing a current at the power-line frequency through it.

low-frequency padder—In a radio receiver, a small adjustable capacitor connected in series with the oscillator tuning coil. During alignment it is adjusted to obtain correct calibration of the circuit at the low-frequency end of the tuning range.

low-level contacts — Contacts that control only the flow of relatively small currents in relatively low-voltage circuits—e.g., alternating currents and voltages encountered in voice or tone circuits, direct current on the order of microamperes, and voltages below the softening voltages on record for various contact materials (that is, 0.080 volt for gold, 0.25 volt for platinum, and the like). Also defined as the range of contact electrical loading where there can be no electrical (arc transfer) or thermal effects and where only mechanical forces can change the conditions of the contact interface.

low-level modulation—The modulation produced in a system when the power level at a certain point is lower than it is at the output.

low-level radio-frequency signal (TR, ATR, and pre-TR tubes)—A radio-frequency signal with insufficient power to fire the tube.

low-level signal—Very small amplitudes serving to convey information or other intelligence. Variations in signal amplitude are frequently expressed in microvolts.

low-loss insulator—An insulator with high radio-frequency resistance and hence slight absorption of energy.

low-loss line—A transmission line with relatively low losses.

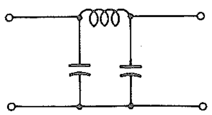

Low-pass filter.

low-pass filter—A filter network which passes all frequencies below a specified frequency with little or no loss but discriminates strongly against higher frequencies.

low-print tape — Special magnetic recording tape significantly less susceptible to print-through (the transfer of signal from one layer to another), which results when tape is stored for long periods of time. These tapes are especially useful for speech "master recording" (making an original recording from which copies will be made) on professional-quality equipment.

low-rate discharge — The withdrawal of a small current from a cell or battery for long periods of time.

low tension—British term for low voltage. Generally refers to the heater or filament voltage.

low-velocity scanning—The scanning of a target with electrons having a velocity below the minimum required to give a secondary-emission ratio of unity.

low-voltage protection — *See* Undervoltage Protection.

L-pad—A volume control that has practically the same impedance at all settings. It consists essentially of an L-network in which both elements are adjusted simultaneously.

LP record—Abbreviation for long-play record.

L + R, L − R—The sum and difference signals of the two stereo channels; the L + R signals combines the signals of both channels in phase; the L − R signal combines them out of phase. By combining in suitable circuitry, L + R and L − R can be added to obtain 2 L, the signal from one channel; L − R can be subtracted from L + R to obtain 2 R, the signal from the other channel.

LTPD—Abbreviation for lot tolerance percent defective.

L-scan—*See* L-Display.

lug—A device soldered or crimped to the end of a wire lead to provide an eye or fork that can be placed under the head of a binding screw.

lumen—A unit of luminous flux equal to the flux (illumination) on a surface, all points of which are the same distance from a uniform source of one candle.

lumen-hour—A unit of the quantity of light delivered in one hour by a flux of one lumen.

luminaire—A complete lighting unit consisting of a light source, together with a globe, reflector, refractor, housing, socket, and other parts integral with the housing.

luminance—1. The luminous flux emitted, reflected, or transmitted from the source. Usual units are the lumen per steradian per square meter, candle per square foot; meter-lambert, millilambert, and foot-lambert. 2. In color television, the photometric quantity of light radiation.

luminance channel—In a color television system, any path intended to carry the luminance signal. The luminance channel may also carry other signals such as the carrier color, which may or may not be used.

luminance channel bandwidth—The bandwidth of the path intended to carry the luminance signal.

luminance flicker—The flicker resulting from fluctuation of the luminance only.

luminance primary—The one of three transmission primaries whose amount determines the luminance of a color.

luminance signal—Also called Y signal. The signal in which the amplitude varies with the luminance values of a televised scene. It is part of the composite color signal and is made up of 0.30 red, 0.59 green, and 0.11 blue. The luminance signal is capable of producing a complete monochrome picture.

luminescence—1. The absorption of energy by matter, and its subsequent emission as light. If the light is emitted within 10^{-8} second after the energy is absorbed, the process is known as fluorescence. If the emission takes longer, the process is called phosphorescence. 2. The emission of light of certain wavelengths or limited regions of the spectrum in excess of that due to incandescence and the emissivity of the surface. This property is not exhibited by all materials. 3. Light emission without heat.

luminescence threshold—Also called threshold of luminescence. The lowest radiation frequency that will excite a luminescent material.

luminescent—Any material which will give off light but not heat when energized by an external source such as a stream of electrons or radiant energy.

luminescent screen—A screen which becomes luminous in those spots excited by an electron beam. The screen of a cathode-ray tube is the best-known example.

luminosity—Ratio of luminous flux to the corresponding radiant flux at a particular wavelength. It is expressed in lumens per watt.

luminosity coefficients—The constant multipliers for the respective tristimulus values of any color. The sum of the three products is the luminance of the color.

luminosity curve — A distribution curve showing luminous flux per element of wavelength as a function of wavelength.

luminosity factor—For radiation of a particular wavelength, the ratio of the luminous flux at that wavelength to the corresponding radiant flux. It is expressed in lumens per watt. Abbreviated K.

luminous—Emitting light (i.e., glowing).

luminous efficiency—The ratio of luminous flux to radiant flux. It is usually expressed in lumens per watt of radiant flux and should not be confused with the term "efficiency" when applied to a practical source of light. The latter is based on the power supplied to the source—not the radiant flux from the source. For energy radiated at a single wavelength, luminous efficiency is synonymous with luminosity.

luminous flux—Also called light flux. Time rate of flow of light (the total visible energy produced by a source per unit time). Usually measured in lumens.

luminous intensity—Light intensity in a specified direction. The standard unit is called the international candle. Originally an ordinary wax candle was selected as the standard, and even today it is rated as having a luminous intensity of approximately one candle.

luminous sensitivity—1. In a phototube, the output current divided by the incident luminous flux at a constant electrode voltage. (The term "output current" here does not include the dark current.) 2. The sensitivity of an object to light from a tungsten-filament lamp operating at a color temperature of 2870°K. This definition is permissible, since luminous sensitivity is not an absolute characteristic but depends on the spectral distribution of the incident flux.

lumped—Concentrated at a single point.

lumped-constant elements—Distinct electrical units smaller than a wavelength. They are calibrated and used, in conjunction with other electrical/electronic equipment, for controlling voltage and current.

lumped-constant - tuned heterodyne-frequency meter—A device for measuring frequency. Its operation depends on the use of a tuned electrical circuit consisting of lumped inductance and capacitance.

lumped impedance—Impedance concentrated in a component, as distinguished from impedance due to stray or distributed effects.

lumped inductance—Inductance concentrated in a component, as opposed to stray or distributed inductance.

lumped loading—Inserting uniformly spaced inductance coils along the line, since continuous loading is impractical.

lumped parameter—Any circuit parameter which—for purposes of analysis—can be considered to represent a single inductance, capacitance, resistance, etc., throughout the frequency range of interest.

lux—A practical unit of illumination in the metric system. It is equivalent to the meter-candle and is the illumination on a one square-meter area on which there is a uniformly distributed flux of one lumen, or the illumination produced at a surface one meter from a source of one candle.

Luxemburg effect—A nonlinear effect in the ionosphere. As a result, the modulation on a strong carrier wave will be transferred to another carrier passing through the same region.

luxmeter—A type of illumination photometer employing a variable aperture and the contrast principle.

M

m—Abbreviation for the prefix milli-.

M—1. Symbol for mutual inductance. 2. Abbreviation for the prefix mega or meg meaning one million.

ma—Abbreviation for milliampere.

Ma—Abbreviation for megampere.

machine-available time—In a computer, power-on time less maintenance time.

machine cycle—The shortest complete process or action that is repeated in order. The minimum length of time in which the foregoing can be performed.

machine hardware—Circuits contained in the five parts of the computer: input, output, control, storage, and arithmetic sections; any other circuit is "off-line."

machine language—Also called computer code. A set of symbols, characters, or signs and the rules for combining them to convey instructions or data to a computer.

machine run—In a computer, the performance of one or more machine routines which are linked to form one operating unit.

machine sensible—Information in a form such that it can be read by a specific machine.

machine thermal relay—A device which functions when the temperature of an a-c machine armature, or of the armature or other load-carrying winding or element of a d-c machine, or converter, power rectifier, or power transformer (including a power-rectifier transformer) exceeds a predetermined value.

machine word—A unit of information consisting of a standard number of characters which a computer regularly handles in each register.

macrocode—A coding system which assembles groups of computer instructions into single code words and which therefore requires interpretation or translation so that an automatic computer can follow it.

macroprogramming—In a computer, the writing of machine-procedure statements in terms of macrocode instructions.

macroscopic—Large enough to be observed by the unaided eye.

macrosonics—The utilization of high amplitude sound waves for performing functions such as cleaning, drilling, emulsification, etc.

madistor—A semiconductor device that makes use of the effects of a magnetic field on a plasma current. The strength of a magnetic field determines the conductivity of the madistor material. A small change in the magnetic field produces a larger change in the madistor current.

MADT transistor—A microalloy diffused-base transistor.

magamp—Abbreviation for magnetic amplifier.

magic eye—*See* Electron-Ray Tube.

magic-t—*See* Hybrid Junction.

magnaflux—A magnetic method of determining surface and subsurface defects in metals.

magnal base—An 11-pin base used on cathode-ray tubes.

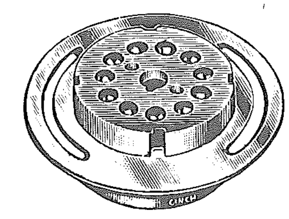

Magnal socket.

magnal socket—An 11-pin socket used with cathode-ray tubes.

magnesium—An alkaline metal the compounds of which are sometimes used for cathodes.

magnesium anode—A bar of magnesium, buried in the earth, connected to an underground cable to prevent cable corrosion due to electrolysis.

magnesium cell—A primary cell the negative electrode of which is made of magnesium or one of its alloys.

magnesium-copper-sulphide rectifier—A dry disc rectifier consisting of magnesium in contact with copper sulphide.

magnesium-silver chloride cell—A reserve primary cell that becomes activated when water is added.

Magnesyn—Trade name for a device made by Bendix Aviation Corp. It is a portion of a repeater unit consisting of a two-pole permanent-manet rotor within a three-phase, two-pole, delta-connected stator. The rotor carries the indicating pointer and is free to rotate in any direction.

magnet—A body that has the property of attracting or repelling magnetic materials. In its natural form it is called a lodestone. It may also be produced by permanently magnetizing a piece of iron or steel. A temporary magnet—called an electromagnet—is produced by passing a current through a coil surrounding a piece of iron or steel; the magnetism persists only while current is flowing. When suspended freely, a magnet will turn and align its poles with the north and south magnetic poles of the earth. (*Also See* Electromagnetism *and* Permanent Magnet.)

magnet brake—A friction brake controlled electromagnetically.

magnetic—Pertaining to magnetism.

magnetic aging—The normal change in the magnetic material resulting from gradual metallurgical change in the material. The term also applies when the metallurgical changes are accelerated by an increase or decrease in temperature. When used in reference to core loss, this term, unless otherwise modified, implies an increase in loss. When used in reference to permeability remanence, the term, in a positive sense, indicates a decrease in these quantities.

magnetic air gap—The air space, or nonmagnetic portion, of a magnetic circuit.

magnetically damped—A form of damping achieved by moving a metal vane through a magnetic field. This motion induces currents in the vane which set up magnetic fields opposing those of the stationary magnets, thus tending to bring the pointer to rest. This type of damping is found in many quality moving-iron and dynamometer-type instruments.

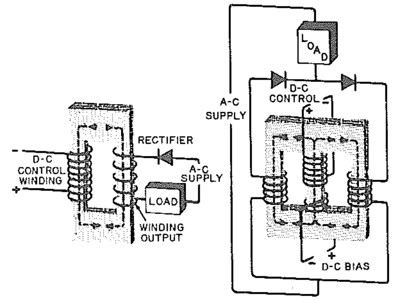

Magnetic amplifiers.

magnetic amplifier—Abbreviated magamp. A variable-impedance device by which load power and voltage can be controlled. Amplification occurs since the control source requires much less power than is supplied to the load.

magnetic analysis—The separation of a stream of electrified particles by a magnetic field, in accordance with their mass, charge, or speed. This is the principle of the mass spectrograph.

magnetic anisotropy—The dependence of the magnetic properties of some materials on direction.

magnetic-armature speaker—Also called a magnetic-armature loudspeaker or magnetic speaker. A speaker comprising a ferromagnetic armature actuated by magnetic attraction.

magnetic azimuth—Azimuth measured from magnetic north.

magnetic bearing—The position in which an object is pointing with respect to the earth's magnetic north pole. It is expressed in degrees clockwise from that pole.

magnetic bias—A steady magnetic field applied to the magnetic circuit of a relay.

magnetic biasing—The superimposing of another magnetic field on the signal magnetic field of a tape while a magnetic recording is being made.

magnetic blowout—A magnet for establishing a field where an electrical circuit is broken. The field lengthens the arc and thus helps to extinguish it.

magnetic blow-out switch — A switch that contains a small permanent magnet which provides a means of switching high d-c loads. The magnet deflects the arc to quench it.

magnetic braking—Application of the brakes by magnetic force. The current for exciting the electromagnets is derived either from the traction motors acting as generators, or from an independent source.

magnetic bremsstrahlung—*See* Synchrotron Radiation.

magnetic cartridge—*See* Variable-Reluctance Pickup.

magnetic character—A character imprinted with ink having magnetic properties. These characters are unique in that they can be read directly by both humans and machines.

magnetic circuit—The path of the flux as it travels from the north pole, through the circuit components, and back to the north pole. In a generator, the magnetic-circuit components include the field yoke, field pole pieces, air gap, and armature core. The magnetic circuit may be compared to an electrical circuit, with magnetomotive force corresponding to voltage, flux lines to current, and reluctance to resistance.

magnetic circuit breaker—A circuit breaker that depends on the response of an electromagnetic coil for operation.

magnetic coated disc — A magnetic disc-recording medium consisting of a coat of magnetizable material over a nonmagnetic base.

magnetic coil—The winding of an electromagnet.

magnetic compass—A device for indicating direction. It consists of a magnetic needle which pivots freely and points toward the earth's north magnetic pole.

magnetic conduction current—The rate of flow of magnetism through a magnetized body.

magnetic contactor — A contactor actuated electromagnetically.

magnetic-convergence principle — The obtaining of beam convergence through the use of a magnetic field.

magnetic core—1. A configuration of magnetic material which is, or is intended to be, placed in a rigid special relationship to current-carrying conductors and the magnetic properties of which are essential to its use. For example, it may be used to concentrate an induced magnetic field as in a transformer induction coil, or armature, to retain a magnetic polarization for the purpose of storing data, or for its nonlinear properties as in a logic element. It may be made of such material as iron, iron oxide, or ferrite in such shapes as wires, tapes, toroids, or thin film. 2. The ferrous material in the center of an electromagnet.

magnetic core storage—A type of computer storage which employs a core of magnetic material surrounded by a coil of wire. The core can be magnetized to represent a binary 1 or 0.

magnetic course—A course in which the reference line is magnetic north.

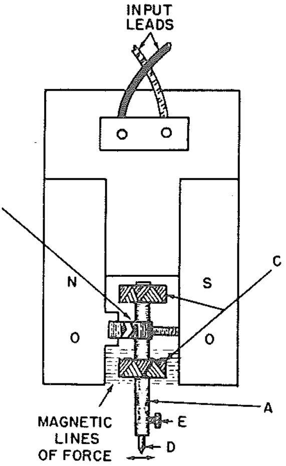

Magnetic cutter.

magnetic cutter—A cutter in which the mechanical displacement of the recording stylus is produced by the action of magnetic fields.

magnetic cycle—The sequence of changes in the magnetization of an object corresponding to one cycle of the alternating current producing the magnetization.

magnetic damping—The damping of a mechanical motion by means of the reaction between a magnetic field and the current generated in a conductor moving through that field. The resistance of this conductor converts excess kinetic energy to heat.

magnetic deflection — The moving of the electron beam by means of a magnetic field produced by a coil around the neck of the cathode-ray tube (i.e., a linear motion is produced when the current through the coil has a sawtooth waveform).

magnetic delay line—A computer delay line in which magnetic energy is propagated consisting in essence of a metallic medium along which the velocity of propagation of magnetic energy is small when compared to the speed of light. The storage of information, which is usually in a binary form, is accomplished by recirculation of wave patterns.

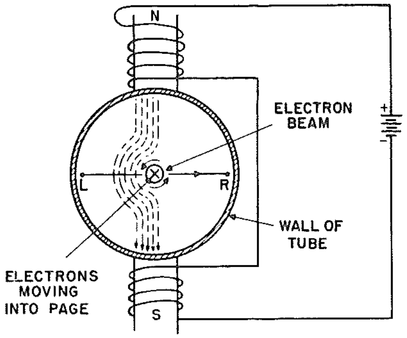

Magnetic deflection.

magnetic density—The number of lines of magnetic force passing through a magnet or magnetic field per unit area of cross section.

magnetic detecting device—A device for detecting cracks in iron or steel. This is done by introducing magnetic particles, which are attracted to the opposing magnetic poles created at the break.

magnetic device—Any device actuated electromagnetically.

magnetic dip—Also called magnetic inclination. The angle of the magnetic field of the earth with the horizontal at a particular location.

magnetic dipole—A pair of equal-strength north and south magnetic poles spaced close together, and so small that its directive properties are independent of its size and shape. It is the magnetic equivalent of an electrical dipole.

magnetic direction indicator — Abbreviated MDI. An instrument that provides a compass indication, which it obtains electrically from a remote gyrostabilized magnetic compass (or its equivalent).

magnetic disc—A flat circular plate with a magnetic surface on which data can be stored by selective magnetization of portions of the flat surface.

magnetic disc storage — Storage of digital information on magnetic discs.

magnetic displacement—Magnetic flux density or magnetic induction.

magnetic drag — Undesired torque usually caused by useless eddy currents and residual magnetism.

magnetic drum—A storage device consisting of a rapidly rotating cylinder, the surface of which can be easily magnetized and which will retain the magnetization. Information is stored in the form of magnetized spots (or no spots) on the drum surface.

magnetic-drum receiving equipment — Radar developed for detection of targets beyond line of sight using ionospheric reflection and very low power. To distinguish radar signals from ground noise, a noise-free standard system is employed whereby the amplitude and characteristics are reproduced and preserved for comparison with returned signals. Fluctuations in the amplitude of the returned signals indicate the position and velocity of the target accurately to within 10 miles.

magnetic electron multiplier—One type of electron multiplier in which the paths of the emitted secondary electrons are controlled by an externally applied magnetic field.

magnetic field — An area where magnetic forces can be detected around a permanent, natural, or electromagnet.

magnetic field strength — *See* Magnetizing Force.

magnetic figures—A pattern showing the distribution of a magnetic field. It is made by sprinkling iron filings on a nonmagnetic surface in the field.

magnetic flip-flop—A bistable amplifier using one or more magnetic amplifiers. The two stable output levels are determined by appropriate changes in the control voltage or current.

magnetic flux—The magnetic induction in a material. An electromotive force will be induced in a conductor placed in a magnetic field whenever the magnitude of the flux changes.

magnetic flux density—*See* Magnetic Induction.

magnetic focusing—The focusing of an electron beam by the action of a magnetic field.

magnetic freezing—In a relay, sticking of the armature to the core due to residual magnetism.

magnetic gap — The nonmagnetic part of a magnetic circuit.

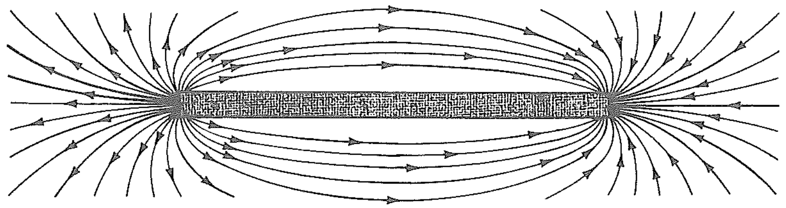

Magnetic field.

magnetic gate—A gate circuit used in a magnetic amplifier.

magnetic head—In magnetic recording, a transducer that converts electric variations into magnetic variations for storage on magnetic media, or for reconverting such stored energy into electric energy. Stored energy can also be erased by this method.

magnetic heading—The heading of an aircraft with reference to magnetic north.

magnetic hysteresis—In a magnetic material, the property by virtue of which the magnetic induction for a given magnetizing force depends on the previous conditions of magnetization.

magnetic hysteresis loss—The power expended in a magnetic material, as a result of magnetic hysteresis, when the magnetic induction is cyclic.

magnetic inclination—*See* Magnetic Dip.

magnetic induction—Also called magnetic flux density. The flux per unit area perpendicular to the direction of the flux. The cgs unit of induction is called the gauss (plural, gausses) and is defined by the equation:

$$B = \frac{d\phi}{dA}$$

magnetic ink—Visible ink containing magnetic particles. When printed on a document (e.g., a bank check), the ink can be read by a magnetic character sensor and also by humans.

magnetic integrated circuit—An integrated component in which magnetic elements perform all or a major portion of its intended function.

magnetic integrated circuitry—The physical realization of one or more magnetic elements inseparably associated to perform all or at least a major portion of its intended function.

magnetic-latch relay—*See* Polarized Double-Biased Relay.

magnetic leakage—Passage of magnetic flux outside the path along which it can do useful work.

magnetic lens—An apparatus that uses a nonuniform magnetic field to focus beams of rapidly moving electrons or ions in a cathode-ray or other tube.

magnetic line of force—In a magnetic field, an imaginary line which has the direction of the magnetic flux at every point.

magnetic materials—Materials that show magnetic properties. Ferromagnetic materials are more strongly magnetic than paramagnetic materials.

magnetic memory—A computer memory (or any portion) in which information is stored in the form of magnetism.

magnetic microphone—*See* Variable-Reluctance Microphone.

magnetic mine—An underwater mine that detonates when near the steel hull of a ship. This is accomplished by relays, which redistribute the magnetic field in the mine when it is near the ship.

magnetic modulator—Also called a magnettor. A modulator employing a magnetic amplifier as the modulating element.

magnetic moment—Ratio of the maximum torque exerted on a magnet, to the magnetizing force of the field in which the magnet is situated.

magnetic needle—The magnetized needle used in a compass. When freely suspended, it will point to the earth's magnetic north and south poles.

magnetic north—The direction indicated by the north-seeking end of the needle in a magnetic compass.

magnetic phase modulator—A ferrite-core delay line in which the delay is varied by an external magnetic field.

magnetic pickup—*See* Variable-Reluctance Pickup.

magnetic-plated wire—A wire with a nonmagnetic core and a plated surface of ferromagnetic material.

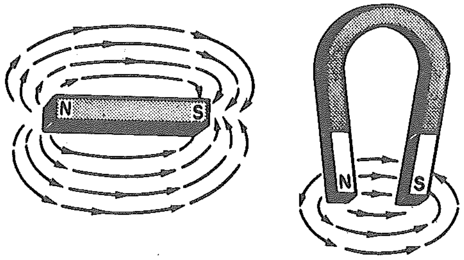

Magnetic poles.

magnetic poles—Those portions of a magnet toward which the lines of flux converge. All magnets have two poles, called north and south or north-seeking and south-seeking, poles.

magnetic pole strength—The magnetic moment of a magnetized body divided by the distance between the poles.

magnetic potential difference—The line integral of magnetizing force between two points in a magnetic field.

magnetic-powder-coated tape—Also called coated tape. A tape consisting of a coating of uniformly dispersed powdered ferromagnetic material on a nonmagnetic base.

magnetic printing—Also called magnetic transfer, and cross talk. The permanent transfer of a recorded signal from a section of one magnetic recording medium to another section of the same or a different medium when they are brought near each other.

magnetic probe—A loop-type conductor for detecing presence of static, audio, or r-f magnetic fields.

magnetic radio bearing—*See* Radio Bearing.

magnetic recorder—Equipment incorporating an electromagnetic transducer and a means for moving a magnetic recording medium past the transducer. Electric signals are recorded in the medium as magnetic variations. (*Also see* Magnetic Recording.)

magnetic recording—Recording audio frequencies by magnetizing areas of a tape or wire. The magnetized tape or wire is played back by passing it through a reproducing head. Here the magnetized areas are reconverted into electrical energy, which headphones or speakers then change back into sound.

magnetic recording head—In magnetic recording, a magnetic head that transforms electric variations into magnetic variations for storage on a magnetic medium such as tape or discs.

magnetic recording medium—A wire, tape, cylinder, disc, or other magnetizable material which retains the magnetic variations imparted to it during magnetic recording.

magnetic reproducer—Equipment which picks up the magnetic variations on magnetic recording media and converts them into electrical variations.

magnetic reproducing head—In magnetic recording, the head that converts the magnetic variations into electric variations.

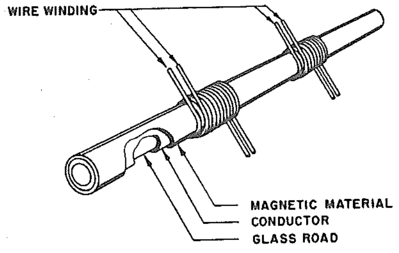

Magnetic rod.

magnetic rod—A square-loop switching and storage element for digital systems. It consists of a silver-coated glass rod upon which a thin layer of iron and nickel is deposited. Conductors are wound around the rod for the drive, sense, enable, and inhibit currents.

magnetics—The branch of science concerned with the laws of magnetic phenomena.

magnetic saturation—In an iron core, the point where application of a further increase in magnetizing force will produce little or no increase in the magnetic lines of force.

magnetic separator—An apparatus for separating powdered magnetic ores from nonmagnetic ores, or iron filings and other small iron objects from nonmagnetic materials. An electromagnet is employed which deflects the magnetic materials from the path taken by the nonmagnetic materials.

magnetic shield—A sheet or core of iron enclosing instruments or radio parts to protect them from stray magnetic fields. The shield provides a convenient path for the magnetic lines of force and thus diverts them from the component being protected.

magnetic shift register—A register in which magnetic cores are used as binary storage elements. By means of pulses, the pattern of binary digital information can be shifted one position to the left or right in the register.

magnetic shunt—A piece of iron used during instrument calibration to divert a portion of the magnetic lines of force passing through an air gap in the instrument.

magnetic speaker—A speaker in which acoustic waves are produced by mechanical forces resulting from magnetic reaction.

magnetic starter—A starter actuated electromagnetically.

magnetic storage—Any storage system in a computer which makes use of the magnetic properties of materials to store information.

magnetic storm—A disturbance in the earth's magnetic field. It is associated with abnormal solar activity and is capable of disrupting both radio and wire transmission.

magnetic strain gage—Also spelled gauge. An instrument for measuring strain in rails or other structural members that bend only microscopically under a normal load. It does this by determining the change in reluctance of a magnetic circuit having a movable armature.

magnetic susceptibility—Ratio of the intensity of magnetization to the corresponding value of the magnetizing force.

magnetic tape—The recording media used in tape recorders. A paper or plastic tape on which a magnetic emulsion (usually ferric oxide) has been deposited. The most common width for home recorders is one-quarter inch. Some tapes are made of a magnetic material and hence need no magnetic-emulsion coating.

magnetic-tape core—A toroidal core made by winding a strip of thin magnetic-core material around a form. A toroidal winder is then used to wind coils around the core.

magnetic tape recorder—A recorder in which magnetic tape is the recording medium.

magnetic test coil—Also called search or exploring coil. A coil which is connected to a suitable device to measure a change in the magnetic flux linked with it. The flux linkage may be changed by either moving the coil or varying the magnitude of the flux.

magnetic thin film—A layer of magnetic material, usually less than 10,000 angstroms thick. In electronic computers, magnetic thin films may be used for logic or storage elements.

magnetic transducer—*See* Variable-Reluctance Transducer.

magnetic transfer—*See* Magnetic Printing.

magnetic transition temperature—Also called the Curie point. In a ferromagnetic material, the point where its transition to paramagnetic seems to be complete as its temperature is raised.

magnetic tubes of flux—Regions in space the sides of which are everywhere tangent to the magnetic induction and the ends of which may meet to form closed rings.

magnetic units—Ampere-turn, gauss, gilbert, line of force, maxwell, oersted, and unit magnetic pole are examples of magnetic units—i.e., those used in measuring magnetic quantities.

magnetic-vane meter—Also called a moving-vane meter. A meter for measuring alternating current. It contains a metal vane which pivots inside a coil. Alternating current flows through the coil and sets up magnetic forces which rotate the vane and attached pointer in proportion to the value of the current.

magnetic variometer — An instrument for measuring the differences in a magnetic field with respect to space or time.

magnetism—A property possessed by certain materials by which these materials can exert mechanical force on neighboring masses of magnetic materials and can cause voltages to be inducted in conducting bodies moving relative to the magnetized bodies.

magnetite—A mineral which exists in a magnetized condition in its natural state. It consists chiefly of a magnetic oxide of iron.

magnetization curve—A curve plotted on a graph to show successive states during magnetization of a ferromagnetic material. A normal magnetization curve is a portion of a symmetrical hysteresis loop. A virgin magnetization curve shows what happens to the material the first time it is magnetized.

magnetization intensity—At any point in a magnetized body, the ratio between the magnetic moment of the element of volume surrounding the point, and an infinitesimal volume.

magnetize—To make magnetic.

magnetizing current—*See* Exciting Current.

magnetizing force — Also called magnetic field strength. The magnetomotive force per unit length at any given point in a magnetic circuit. In the cgs system the unit is the oersted. It is defined by the equation:

$$\frac{\text{Magnetomotive force in gilberts}}{\text{Length in centimeters}}$$

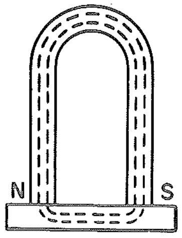

Magnet keeper.

magnet keeper—A bar of iron or steel placed across the poles of a horseshoe magnet when the magnet is not in use. The keeper prevents the magnet from becoming demagnetized, by completing the magnetic circuit to keep the flux from leaking off.

magnet meter—Also called a magnet tester. An instrument for measuring the magnetic flux produced by a permanent magnet. It usually comprises a torque-coil or moving-magnet magnetometer with a particular arrangement of pole pieces.

magneto—1. An a-c generator for producing ringing signals. 2. An a-c generator for producing the ignition voltage in some gasoline engines.

magnetoelastic coupling — Energy transfer between elastic and spin-wave modes of propagation. The strength of the effect varies with the material used.

magnetoelastic energy—In a crystal, energy of interaction between elastic (or lattice) strains and applied magnetization. If a ferromagnetic crystal is placed in a magnetic field, or if a magnetic field is varied about the crystal, the lattice structure of the crystal is distorted (magnetostriction occurs) and changes the amount of potential energy in the lattice. The energy associated with this change is magnetoelastic.

magnetoelastic wave—A hybrid of spin and elastic waves.

magnetoelectric generator—An electric generator with permanent-magnet field poles.

magnetoelectric transducer — A transducer which measures the emf generated by the movement of a conductor relative to a magnetic field.

magnetofluiddynamics—*See* Magnetohydrodynamics.

magnetofluidmechanics—*See* Magnetohydrodynamics.

magnetogasdynamics — *See* Magnetohydrodynamics.

magnetograph—A magnetometer that provides a continuous record of the changes occurring in the earth's magnetic field.

magnetohydrodynamic gyroscope — A gyroscope in which a rotating magnetic field drives a conducting fluid (e.g., mercury) around the closed path formed between the inner surface of a magnetic sleeve and the outer surface of a concentric magnetic cylinder.

magnetohydrodynamic power generation—The generating of electric current by the motion of an ionized gas.

magnetohydrodynamics—Abbreviated MHD. The study of the interaction between magnetic fields and electrically conductive fluids and gases. Also called magnetofluiddynamics, magnetofluidmechanics, magnetogasdynamics, hydromagnetics.

magnetoionic duct—A duct along the geomagnetic lines of force which exhibit wave-guide characteristics for radio-wave propagation between conjugate points on the earth's surface.

magnetoionic wave component—Either of the two elliptically polarized wave components into which a linearly polarized wave

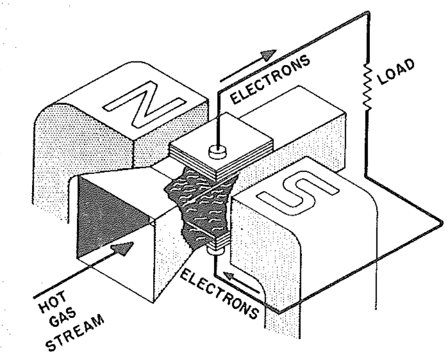

Magnetohydrodynamic generator.

in the ionosphere is separated by the earth's magnetic field.

magnetometer—An instrument for measuring the intensity or direction (or both) of a magnetic field (or component) in a particular direction.

magnetomotive force—The force by which a magnetic field is produced, either by a current flowing through a coil of wire or by the proximity of a magnetized body. The amount of magnetism produced in the first method is proportional to the current through the coil and the number of turns in it. The cgs unit of magnetomotive force is called the gilbert. Magnetomotive force may also result from a magnetized body.

magnetoresistance—The change in electrical conductivity of a material when a magnetic field is applied. This change is quite pronounced in materials with a high carrier mobility.

magnetoresistor—A semiconductor device in which the electrical resistance is a function of the applied magnetic field.

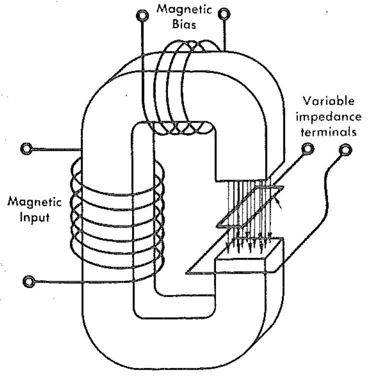

Magnetoresistor.

magnetosphere—A 900-mile-thick belt in the upper atmosphere, composed primarily of helium gas.

magnetostatic field—A magnetic field that is neither moving nor changing direction. Such a field could be produced by a stationary magnetic pole or by a constant current flowing in a stationary conductor.

magnetostriction—That property of certain ferromagnetic metals, such as nickel, iron, cobalt, and manganese alloys, which causes them to shrink or expand when placed in a magnetic field. Conversely, if subjected to compression or tension, the magnetic reluctance changes, thus making it possible for a magnetostrictive wire or rod to vary a magnetic field in which it may be placed. This is true for lateral as well as longitudinal strains.

magnetostriction microphone — A microphone in which the deformation of a magnetostrictive material generates the required voltages.

magnetostriction oscillator — An oscillator in which the frequency is determined by the characteristics of the magnetostrictive element that inductively couples the plate circuit to the grid circuit.

magnetostriction speaker — A speaker in which the mechanical displacement is derived from the deformation of a magnetostrictive material.

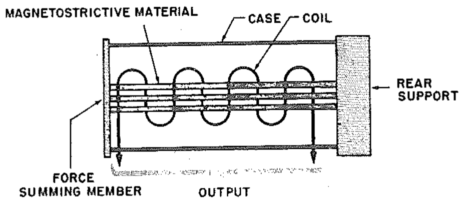

Magnetostriction transducer.

magnetostriction transducer—A transducer comprising an element of magnetostrictive material inside a coil, and a force-summing member attached to one end of the element. Current flows through the coil, and the magnetic field around it expands and contracts the element to move the member back and forth.

magnetostrictive delay line—A delay line made of nickel or certain other materials which become shorter when placed in a magnetic field.

magnetostrictive filter — A filter network which uses the magnetostrictive phenomena to form high-pass, low-pass, bandpass, or band-elimination filters. The impedance characteristic is the inverse of that of a crystal.

magnetostrictive relay — A relay that functions because of dimensional changes occurring in a magnetic material under the influence of a magnetic field.

magnetostrictive resonator — A ferromagnetic rod which can be excited magnetically so that it will resonate (vibrate) at one or more desired frequencies.

magneto telephone—A telephone equipped with a magneto (a hand-driven, two-pole, ringing-signal generator). Although obsolete for home and business telephones, it is still used in many field applications.

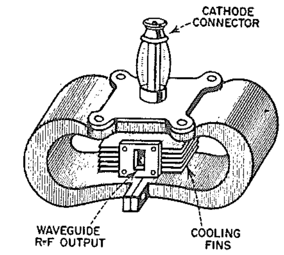

Magnetron.

magnetron—An electron tube which produces an a-c power output. The tube is surrounded by an electromagnet, which controls the flow of electrons from cathode to anode.

magnetron effect—In a thermionic vacuum tube, the reduced electron emission due to the magnetic field of the filament current.

magnetron oscillator—An electron tube in which electrons are accelerated by a radial electric field between the cathode and one or more anodes and by an axial magnetic field that provides a high-energy electron stream to excite the tank circuits.

magnetron pulling—A shift in the frequency of a magnetron due to a change in the standing waves on the r-f lines.

magnetron pushing—The shift in frequency of a magnetron caused by faulty modulator operation.

magnetron rectifier—A gas-tube rectifier in which the electron stream is controlled by an optical element or instrument instead of by heated electrodes.

magnet steel—A special steel used in permanent magnets because of its high retentivity. In addition to steel, it also contains tungsten, cobalt, chromium, and manganese.

magnet strip—Sheet or foil aluminum (either bare or insulated) used as a conductor in electric windings. Copper is also used sometimes.

magnet tester—*See* Magnet Meter.

magnettor—*See* Magnetic Modulator.

magnet wire—An insulated copper wire used for winding the coils of electromagnets.

magnified sweep—In an oscilloscope, a sweep the time per division of which has been decreased by amplification of the sweep waveform rather than by changing the time constants used to generate it.

Magnistor—A saturable reactor for controlling electrical pulses or sine waves having frequencies of 100 kHz to 30 MHz and power levels ranging from microwatts to tens of watts.

magnitude—Size; the quantity assigned to one unit so that it may be compared with other units of the same class (i.e., the ratio of one quantity to another).

magnitude-controlled rectifier—A type of rectifier circuit in which a thyratron is used as the rectifying element. The load current is controlled by varying the bias on the grid of the thyratron.

mag-slip—A British term for synchro (i.e., a synchronous device such as the selsyn, autosyn, motortorque and generator).

main anode—The anode which conducts the load current.

main bang—The transmitted pulse of a radar system.

main control unit—Transmitter or receiver controls for energizing, adjustment, etc., of the transmitter or receiver, but not for operating it while on the air.

main exciter—An exciter which supplies energy for the field excitation of another exciter.

main frame—1. Also called central processing unit. The central processor of the computer system. It contains the main storage, arithmetic unit, and special register groups. 2. All of a computer exclusive of the input, output, peripheral, and, in some instances, storage units.

main gap—The conduction path between the principal cathode and anode.

main power—Power supplied to a complete system from a line.

mains—Interior wires extending from the service switch, generator bus, or converter bus to the main distribution center.

main storage—Usually the fastest storage device of a computer and the one from which instructions are executed.

main sweep—The longest-range scale available on some fire-control radars.

maintained contact switch—A switch which remains in a given condition until actuated to another condition, which is also maintained until further actuation.

maintained switch—A switch which remains at the operated circuit condition when the actuating force is removed. It returns to the normal circuit condition, or moves to another position, when actuated a second time.

maintenability—The probability of restoring a system to its specified operating conditions within a specified total down time, when maintenance action is initiated under stated conditions.

maintenance—All procedures necessary to keep an item in, or restoring it to, a serviceable condition, including servicing, repair, modification, modernization, overhaul, inspection, etc.

major-apex face—In a natural quartz crystal, any one of the three larger sloping faces extending to the apex (pointed) end. The other three are called the minor-apex faces.

major cycle—1. In a memory device which provides serial access to storage positions, the time interval between successive appearances of a given storage position. 2. A number of minor cycles.

major defects—Those defects usually responsible for the failure of a component to function in its intended manner.

major face—Any one of the three larger sides of a hexagonal quartz crystal.

major failure—A noncritical failure that can degrade the system performance due to cumulative tolerance buildup.

majority carrier—The predominant carrier in a semiconductor. Electrons are the majority carrier in N-type semiconductors, since there are more electrons than holes. Likewise, holes are the majority carrier in P-types, since they outnumber the electrons.

majority-carrier contact—An electrical contact across which the ratio of majority-carrier current to applied voltage is substantially independent of the voltage polarity but the ratio of minority-carrier current to applied voltage is not.

majority emitter—The transistor electrode from which the majority carriers flow into the interelectrode region.

major lobe—The antenna lobe in the direction of maximum energy radiation or reception.

major loop—A continuous network composed of all the forward elements and the primary feedback elements of a feedback control system.

make—The closing of a relay, key, or other contact.

make-before-break contacts—Double-throw contacts so arranged that the moving contact establishes a new circuit before disrupting the old one.

make-before-break switch—*See* Shorting Switch.

make-break contacts—Also called continuity-transfer contact. A contact form of a relay, in which one contact closes connection to another contact and then opens its prior connection to a third contact.

make-break electrode—In a mercury switch, that electrode which serves the function of making and breaking contact with the mercury pool to close or open the electrical circuit.

make contact—The contact which closes a circuit when a normally open device is operated. *Also see* Normally Open Contact.

male—Adapted so as to fit into a matching hollow part.

malfunction—*See* Error.

mandrel test—A test used to determine the flexibility of insulation. In it a wire, with or without previous stretch, is wrapped around a mandrel.

manganin—An alloy wire used in precision wirewound resistors because of its low temperature coefficient of resistance.

manipulated variable—A quantity or condition that is varied as a function of the actuating signal in order to control the value of the directly controlled variable.

man-made interference—A type of electromagnetic interference generated by electric motors, communication and broadcast transmitters, fluorescent lighting, and other electrical and electronic systems that radiate spurious signals.

man-made static—High-frequency noise signals created by sparking in an electric circuit. When picked up by radio receivers, it causes buzzing and crashing sounds from the speaker.

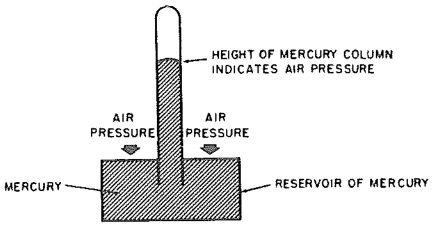

Manometer.

manometer—A gage for measuring the pressure of gases. It contains a column of incompressible liquid. The amount the liquid is displaced indicates the magnitude of the pressure causing the displacement.

manpack—Also called packset. A portable radio-transmitting/receiving device which can be carried easily in a harness by a man.

manual—1. Hand operated. 2. In an organ, a group of keys played with the hand. In two-manual organs the upper manual, also referred to as the solo or swell manual, is normally used to play the melody. The lower manual, also referred to as the accompaniment manual or great manual, is normally used to play the accompaniment.

manual control—1. The opening or closing of switches by hand. 2. The direction of a computer by means of manually operated switches.

manual controller—An electric controller in which all but its basic functions are performed by hand.

manual direction finder—A radio compass that uses a rotatable loop which is operated manually.

manual preset timer—A manual start timer, the cycle of which is initiated by turning a pointer to the desired setting.

manual reset—A qualifying term used to indicate that a relay may be reset manually after an operation.

manual ringing—A method of ringing a telephone. The key must be held down for the ringing to continue. Nor does it stop when the receiver is lifted off the hook, unless the caller releases the key.

manual start timer—An interval time on which each cycle must be manually started at the timer.

manual switch—A switch which is actuated by an operator.

manual switchboard—A telephone switchboard in which the operator makes connections manually with plugs and jacks or with keys.

manual switching—A characteristic of a circuit breaker permitting manual opening and closing of the circuit by operation of the actuator.

manual telegraphy—Telegraphy in which an operator forms the individual characters of the alphabet in code.

manual telephone set—A set not equipped with a dial for securing the number to be called. Instead, lifting the receiver alerts the switchboard operator, who then connects the caller to the person being called.

manual telephone system — A system in which telephone connections between customers are ordinarily established manually by telephone operators, in accordance with orders given verbally by the calling parties.

manual tuning—Rotation by hand of a knob on a radio receiver to tune in a desired station.

manufacturing holes—*See* Pilot Holes.

Marconi antenna — An antenna system in which one end of the signal source is connected to a radiating element and the other end is connected to ground.

margin—1. The difference between an actual operating point and the point or conditon where a failure to operate properly will occur. 2. Also called range or printing range. In telegraphy, the interval between limits on a scale, usually arbitrary, in which printing is error-free.

marginal checking — Also called marginal testing. Preventive maintenance in which certain operating conditions (e.g., supply voltage or frequency) are varied from normal in order to locate defects before they become serious.

marginal relay—A relay with a small margin between its nonoperative current value (maximum current applicable without operation) and its operative value (minimum current that operates the relay).

marginal testing—*See* Marginal Checking.

marine broadcast station—A coastal station which regularly broadcasts the time and meteorological and hydrographic information.

marine radiobeacon station — A radionavigation land station, the emissions from which are used to determine the bearing or direction of a ship in relation to the marine radiobeacon station.

maritime mobile service—The radio service in which ships communicate with each other or with coastal and other land stations on specified frequencies.

maritime radionavigation service—A radio service intended to be used for the navigation of ships.

mark — 1. In telegraphy, the closed-circuit condition—i.e., the signal that closes the circuit at the receiver to produce a click of the key or to print a character on a teletypewriter. 2. The presence of signal. In telegraph communications a mark represents the closed condition or current flowing. A mark impulse is equivalent to a binary 1. 3. *See* Flag.

marker—Also called marker beacon. A radio navigational aid consisting of a transmitter that sends a signal to designate the small area around and above it.

marker antenna—The transmitting antenna for a marker beacon.

marker beacon—*See* Marker.

marker generator—An r-f generator that injects one or more pips of a specific frequency onto the response curve of a tuned circuit being displayed on the screen of a cathode-ray oscilloscope.

marker pip—The inverted V (Λ) or spot of light used as a frequency index mark in cathode-ray oscilloscopes for alignment of tv sets. It is produced by coupling a fixed-frequency oscillator to the output of a sweep-driven signal generator.

marker thread—A colored thread laid parallel and adjacent to the strands of an insulated conductor which identifies the wire manufacturer and often the specification under which the wire is constructed.

marking-and-spacing intervals — In telegraphy, the intervals corresponding to the closed and open positions, respectively, of the originating transmitting contacts.

marking pulse—The signal interval during which the selector unit of a teletypewriter is operated.

marking wave—Also called keying wave. In telegraphy, the emission while the active portions of the code characters are being transmitted.

mark sense—To mark a position on a punch card, using a special pencil that leaves an electrically conductive deposit for later conversion to machine punching.

mark sensing—A technique for detecting special pencil marks entered in special places on a card and automatically translating the marks into punched holes.

marshalling sequence — *See* Collating Sequence.

maser—Acronym for Microwave Amplification by Stimulated Emission of Radiation. A low-noise microwave amplifier in which a signal is boosted by changing the energy level of a gas or crystal (commonly, ammonia or ruby, respectively).

mask—1. A frame mounted in front of a television picture tube to limit the viewing area of the screen. 2. A device (usually a thin sheet of metal which contains an open pattern) used to shield selected portions of a base during a deposition process. 3. A device used to shield selected portions of a photosensitive material during photographic processing.

masking—The process by which a sound is made audible by the addition of a second sound called the masking sound. The unit of measurement is usually the decibel.

masking audiogram—A graphical representation of the amount of masking by a noise. It is plotted in decibels as a function of the frequency of the masked tone.

mask microphone—A microphone designed for use inside an oxygen or other respiratory mask.

Masonite — Trade name of The Masonite Corp. Fiberboard made from steam-exploded wood fiber. Its highly compressed forms are used for panels in electrical equipment.

mass—The quantity of matter in an object. It is equal to the weight of a body divided by the acceleration due to gravity.

mass data—A larger amount of data than can be stored in the central processing unit of a computer at any one time.

mass radiator—A spark radiator which generates a low-level, broad-band signal extending into and above the ehf band. Arcing occurs between fine metal particles suspended in a liquid dielectric.

mass spectrometer—An instrument that permits rapid analysis of chemical compounds. It consists of a vacuum tube into which a small amount of the gas to be studied is admitted. The gas is ionized by the electrons emitted from the cathode and speeded up by an accelerating grid. An electric field draws the ions out of the ionizing chamber. They are then sent through electric and magnetic fields that sort them according to their ratios of mass to charge.

mass spectrum—The spectrum obtained by deflecting a beam of electrons with an electric or magnetic field as they emerge from a tube containing a small quantity of the gas being investigated. The amount a particle is deflected depends on the ratio of its mass to its atomic charge. Hence, every element has a characteristic mass-spectrum line.

mast—The pole on which an antenna is mounted.

master—The mold from which other disc recordings are cast. It is made by electroforming from a disc recording, and is a "negative" of the disc (i.e., has ridges instead of grooves).

master brightness control—In a color television receiver, a variable resistor that adjusts the bias level on all three guns in the picture tube at the same time.

master clock—1. In a computer, the primary source of timing signals. 2. A very accurate timer with an absolute time reference, providing controlled power to drive slave or auxiliary timers and display units. May also provide correcting pulses for slave devices.

master contactor — A device which is generally controlled by a master element or equivalent, and the necessary permissive and protective devices, and which serve to make and break the necessary control circuits to place an equipment into operation under the desired conditions, and to take it out of operation under other or abnormal conditions.

master control—1. In a studio, a central point from which sound or television programs are switched to one or more transmitters. 2. An application-oriented routine usually applied to the highest level of a subroutine hierarchy.

master drawing—A drawing showing the dimensional limits or grid location applicable to any or all parts of a printed circuit including the base.

master drive—A drive that determines the reference input for one or more follower drives.

master element—The initiating device, such as a control switch, voltage relay, float switch, etc., which serves either directly or through such permissive devices as protective and time-delay relays to place an equipment in or out of operation.

master file — In a computer, a file of relatively more permanent information that is usually updated periodically.

master gain control—An amplifier control that permits adjusting the gain of two or more channels simultaneously.

master oscillator—In an amplifier, the oscillator that establishes the carrier frequency of the output.

master oscillator-power amplifier—Abbreviated MOPA. An oscillator followed by a radio-frequency buffer-amplifier stage.

master pattern—An accurately scaled pattern which is used to produce the printed circuit within the accuracy specified in the master drawing.

master reticle — A reticle plate properly aligned in a reticle frame and sealed in place. The master reticle is inserted into a photorepeater to make the final-sized photomask. Used in the production of monolithic circuits.

master routine—*See* Subroutine.

master slice—Silicon wafers containing 30 or more groups of components. The elements can be interconnected with paths of aluminum to form desired circuits. The wafer is then diced into single devices.

master stamper — A master from which phonograph records are pressed.

master station—The radio station to which the emissions of other stations of a synchronized group are referred.

master switch—The switch that dominates the operation of contactors, relays, or other electromagnetic devices.

master tv system—A combination of components for providing multiple tv sets operation from one antenna.

master wafer — A processed semiconductor wafer with unconnected active and passive circuit elements located in a standardized pattern. Different integrated circuits may be synthesized by using various interconnection paths.

MAT—Acronym for microalloy transistor.

match—The similarity or equality of one thing to another.

matched filter—An optimum filter for separating a waveform of known shape from random perturbing noise.

matched impedance—The condition that exists when two coupled circuits are adjusted so that the impedance of one circuit equals the impedance of the other.

matched load—A device used to terminate a transmission line or waveguide so that all the energy from the signal source will be absorbed.

matched power gain — That power gain which is obtained when the impedance of the load is matched to the effective output impedance of the amplifier to which the load is connected.

matched symmetrical transistor—A special case of the bidirectional transistor in which not only are the requirements for a symmetrical transistor met, but actual matching specifications are also given. *Also See* Symmetrical Transistor.

matched termination — A termination that does not reflect the waves along the waveguide or transmission line.

matched transmission line—A transmission line along which there is no wave reflection.

matched waveguide—A waveguide along which there is no reflected wave.

matching—Connecting two circuits or parts together with a coupling device in such a way that the maximum transfer of energy occurs between the two circuits, and the impedance of either circuit will be terminated in its image.

matching diaphragm—A window consisting of an aperture (slit) in a thin piece of metal, placed transversely across a waveguide; used as a matching device. The orientation of the slit (whether parallel to the long or short dimensions of the waveguide) determines whether it is respectively capacitive or inductive.

matching impedance—The impedance value that must be connected to the terminals of a signal-voltage source for proper matching.

matching plate—In waveguides, a diaphragm used for matching.

matching stub—A device placed on a radio-frequency transmission line to vary its electrical length and hence its impedance.

matching transformer—A transformer used for matching impedances.

mate—Joining of two connector halves, or of a cable to a connector.

material system — The designation of the number of basic metals (e.g., silver-antimony-telluride) making up thermoelectric materials.

mathematical check — A check making use of mathematical identities or other properties.

mathematical logic — Also called symbolic logic. Exact reasoning concerning nonnumerical relations by using symbols that are efficient in calculation.

matrix—1. A coding network or system in a computer. When signals representing a certain code are applied to the inputs, the output signals are in a different code. 2. In electronic computers, any logical network the configuration of which is a rectangular array of intersections of its input-output leads, with elements connected at some of these intersections. The network usually functions as an encoder or decoder. 3. A computer network or system in which only one input is excited at a time and produces a combination of outputs. 4. In a color tv circuit, the section that combines the I, Q, and Y signals and transforms them into individual red, green, and blue signals which are applied to the picture-tube grids.

matrixer—Also called matrix unit. A device which transforms the color coordinates, usually by electrical or optical means.

matrix life test—A test in which each test condition has two components. For example in transistor-matrix life testing each life-test condition is represented by an ambient temperature and a power dissipation. At each stress level of temperature, there are several dissipations; and at each stress level of dissipation, there are several temperatures. The test conditions are placed into blocks or groups.

matrix unit—*See* Matrixer.

matter — Any physical entity—i.e., having mass.

Matteucci effect—The ability of a twisted ferromagnetic wire to generate a voltage as its magnetization changes.

max—Abbreviation for maximum.

maxima/minima—In radar, regions of maximum and minimum return from the transmitted pulse caused by additive and subtractive combinations of the direct and reflected wave. A plot of these data is usually known as a null pattern or fade chart.

maximum—Abbreviated max. The highest value occurring during a stated period.

maximum average power output — In television, the maximum radio-frequency output power, averaged over the longest repetitive modulation cycle.

maximum deviation sensitivity — Under maximum system deviation, the smallest signal input for which the output distortion does not exceed a specified limit.

maximum frequency of oscillation — The highest frequency at which a transistor or vacuum tube will oscillate.

maximum keying frequency—In a facsimile system, the frequency (in hertz) equal to half the number of critical areas of the subject copy scanned per second.

maximum modulating frequency—In a facsimile system, the maximum scanning frequency process that can be transmitted without degrading the recorded copy.

maximum output—The highest average output power into a rated load, regardless of distortion.

maximum overshoot—The maximum amplitude deviation from the average of the steady-state values that exist immediately before and after the transient.

maximum peak plate current—The highest instantaneous plate current a tube can safely carry.

maximum percentage modulation — The highest percentage of modulation permitted in a transmitter without producing excessive harmonics in the modulating frequency.

maximum permeability—The highest permeability reached as induction or magnetization is increased.

maximum power transfer theorem — The maximum power that will be absorbed by one network from another joined to it at two terminals, when the impedance of the receiving network is varied, if the impedances (looking into the two networks at the junction) are conjugates of each other.

maximum record level—In direct recording, the amount of record-head current required to produce 3-percent third-harmonic distortion of the reproduced signal at the record-level set frequency. Such distortion must result from magnetic-tape saturation, not from electronic circuitry.

maximum response speed — The maximum pulse frequency that can be applied to a stepper motor at random and result in synchronized steps. The motor must not miss a step while operating within this range.

maximum retention time — The maximum time interval between writing into a storage element of a charge storage tube, and reading an acceptable output.

maximum sensitivity — The smallest signal input that produces a specified output.

maximum signal level — In an amplitude-modulated system, the level corresponding to copy black or copy white—whichever has the higher amplitude.

maximum sound pressure — For any given cycle of a periodic wave, the maximum absolute value of the instantaneous sound pressure. The most common unit is a microbar.

maximum storage time—In a storage tube, the length of time after writing during which an acceptable output can be read. *Also See* Maximum Usable Viewing Time.

maximum system deviation—In a frequency-modulation system, the greatest permissible deviation in frequency.

maximum undistorted output—Also called maximum useful output. The maximum power an amplifier can deliver without producing excessive harmonics.

maximum usable frequency — Abbreviated MUF. In radio transmission by ionospheric reflection, the highest frequencies that can be transmitted by reflection from regular ionized layers.

maximum usable viewing time—Also called maximum storage time. The length of time during which the visible output of a storage tube can be viewed, without rewriting, before a specified decay occurs.

maximum useful output—*See* Maximum Undistorted Output.

maximum writing rate—The maximum spot speed which produces a line of specified density on a photographic negative or on the screen of a cathode-ray tube.

maxwell—10^{-8} weber, the cgs unit of magnetic flux.

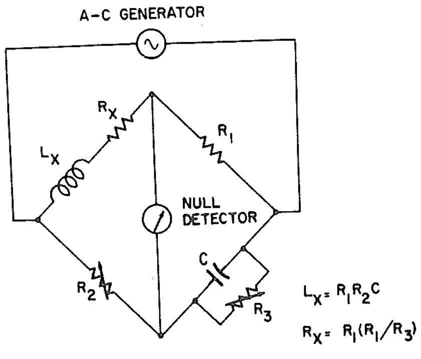

Maxwell bridge.

Maxwell bridge—A four-arm a-c bridge normally used for measuring inductance in terms of resistance and capacitance (or capacitance in terms of resistance and inductance). One arm has an inductor in series with a resistor, and the opposite arm has a capacitor in parallel with a resistor. The other two arms normally are nonreactive resistors. The balance is independent of frequency.

Maxwell inductance bridge—A four-arm a-c bridge for comparing inductances. Two adjacent arms have inductors, and the other two arms usually have nonreactive resistors. The balance is independent of frequency.

Maxwell mutual-inductance bridge—A four-arm a-c bridge for measuring mutual inductance in terms of self-inductance. Mutual inductance is present between the supply circuit and the arm which includes one coil of the mutual inductor. The other three arms normally are nonreactive resistors. The balance is independent of frequency.

Maxwell's equations — Fundamental equations, developed by J. C. Maxwell, for expressing radiation mathematically and describing the condition at any point under the influence of varying electric and magnetic fields.

Maxwell's law—A movable portion of a circuit will always travel in the direction that gives maximum flux linkages through the circuit.

Maxwell triangle—A graph that defines the chromaticity values of a color in terms of three coordinates.

Maxwell-turn—A unit of magnetic linkage equal to one magnetic line of force passing through one turn of a circuit.

mayday—International distress call for radiotelephone communication. It is derived from the French *m'aidez,* meaning "help me."

mc—Abbreviation for megacycle—an obsolete term superseded by MHz (megahertz).

MCM—*See* Monte Carlo Method.

McNally tube—A velocity-modulated vacuum tube that produces low-power uhf oscillations. It is used as a local oscillator in some radar receivers.

mcw—Abbreviation for modulated continuous wave.

m-derived filter—A filter derived by multiplying the values of a constant-K filter by a factor m, the value of which is less than unity.

m-derived L-section filter—A reactance network derived from the prototype L-section filter, so that the image-transfer coefficient and one image impedance are changed, but the other image impedance is left unchanged.

MDI — Abbreviation for magnetic direction indicator.

M-display—*See* M-Scan.

Meacham-bridge oscillator—A crystal oscillator in which the crystal forms one arm in a bridge so as to obtain effective multiplication of the actual Q of the crystal.

meaconing—The interception and rebroadcast of beacon signals. They are rebroadcast on the received frequency to confuse enemy navigation. As a result, aircraft or ground stations are given inaccurate bearings.

mean carrier frequency—The average carrier frequency of a transmitter (corresponding to the resting frequency in a frequency-modulated system).

mean charge — The arithmetic mean of the transferred charges corresponding to a given capacitor voltage, as determined from a specified alternating-charge characteristic curve.

mean charge characteristic — The function giving the relation of mean charge to capacitor voltage.

mean free path—1. The average distance which sound waves travel between successive reflections in an enclosure. 2. The average distance between collisions of atomic particles, which may be further specified according to type of collision (e.g., elastic, inelastic).

mean life — 1. Also called average life. In a semiconductor, the time taken by injected excess carriers to recombine with others of the opposite sign. 2. A measure of the probability that a part or equipment will function satisfactorily during its constant-failure-rate period. It is unrelated to longevity.

mean power of a radio transmitter—Power supplied to the antenna transmission line by a transmitter during normal operation, averaged over a time sufficiently long compared with the period of the lowest frequency encountered in the modulation. A time of 1/10 second during which the mean power is greatest will be selected normally.

mean pulse time—The arithmetic mean of the leading-edge and trailing-edge pulse times.

means of communications — The medium (i.e., electromagnetic or sound waves, visual, messenger) by which a message is conveyed from one person or place to another.

mean-time-between-failures — Abbreviated MTBF. The limit of the ratio of operating time in a device to the number of observed failures as the latter approaches infinity.

mean-time-to-failure — Abbreviated MTTF. In a piece of equipment, its measured operating time divided by its total number of failures during that time. Normally this measurement is made between the early-life and wearout failures.

mean-time-to-first-failure — Abbreviated MTTFF. A special case of MTBF, where T is the accumulated operating time to first failure of a number of devices (failures).

mean-time-to-repair — The total effective maintenance time during a given time interval divided by the total number of failures during the interval.

measurand—Also called stimulus. The physical quantity, force, property, or condition measured by an instrument.

measurement component—Those parts or subassemblies used primarily for the construction of measurement apparatus, excluding screws, nuts, insulated wire, or other stable materials.

measurement device—An assembly of one or more basic elements with other components needed to form a self-contained unit for performing one or more measurement operations. Included are the protecting, supporting, and connecting, as well as functioning parts.

measurement energy—The energy required to operate a measurement device or system. Normally it is obtained from the measurand or the primary detector.

measurement equipment — Any assemblage of measurement components, devices, apparatus, or systems.

measurement inverter—*See* Measuring Modulator.

measurement range of an instrument—The total range throughout which the instrument will provide an accurate measurement.

measurement voltage divider—Also called voltage-ratio or volt box. A combination of two or more resistors, capacitors, or other elements arranged in series so that the voltage across any one is a definite, known fraction of the voltage applied to the combination (provided the current drain at the tap point is negligible or taken into account). The term "volt box" is usually limited to resistance voltage dividers intended for extending the range of direct-current potentiometers.

measuring modulator—Also called measurement inverter or chopper. An intermediate means of modulating a direct-current or low-frequency alternating-current input in a measurement system to give a proportionate alternating-current output, usually as a preliminary to amplification.

mechanical bail — Switch action in which, upon actuation of one station, the switch changes the contact position, mechanically locks the switch in that position, and releases any station previously actuated.

mechanical bandspread — A vernier tuning dial or other mechanical means of lengthening the rotation of a control knob. This permits more precise tuning in crowded short-wave bands.

mechanical bias—A mechanical force tending to move the armature of a relay toward a given position.

mechanical compliance—The displacement of a mechanical element per unit of force, expressed in centimeters per dyne. It is the reciprocal of stiffness and is analogous to capacitance.

mechanical damping—Mechanical resistance, generally associated with the moving parts of a cutter or reproducer.

mechanical damping ring—A loose member mounted on a contact spring for the purpose of reducing contact chatter.

mechanical filter — *See* Mechanical Wave Filter.

mechanical impedance — The complex ratio of the effective force acting on a specified area of an acoustic medium or mechanical device to the resulting effective linear velocity through or of that area, respectively. The unit is the newton-sec/m or the mks mechanical ohm. (In the cgs system, the unit is the dyne-sec/cm or the mechanical ohm.)

mechanical joint—A joint made by clamping cables or other conductors together mechanically rather than by soldering them.

mechanical life—The maximum number of complete cycles through which a device may be actuated without electrical or mechanical failure.

mechanically timed relay — A relay that is mechanically timed by such means as a clockwork, escapement, bellows, or dashpot.

mechanical ohm—The magnitude of a mechanical resistance, reactance, or impedance for which a force of 1 dyne produces a linear velocity of 1 centimeter per second (dyne-sec/cm). When expressed in newton-sec/m it is called the mks mechanical ohm.

mechanical overtravel—The shaft travel of a precision potentiometer between each end point (or limit of theoretical electrical travel) and its adjacent corresponding limit of total mechanical travel.

mechanical phonograph—A phonograph the playback stylus of which drives the diaphragm of an acoustic pickup that radiates acoustic energy without any further amplification.

mechanical phonograph recorder — Also called mechanical recorder. Equipment that converts electric or acoustic signals into mechanical motion and cuts or embosses it into a medium.

mechanical reactance—The magnitude (size) of the imaginary component of mechanical impedance.

mechanical recorder—*See* Mechanical Phonograph Recorder.

mechanical recording head—*See* Cutter.

mechanical rectifier—A rectifier in which its action is done mechanically (e.g., by making and breaking the electrical circuit at the correct times with a rotating wheel or vibrating reed).

mechanical register—An electromechanical device which records or indicates a count.

mechanical reproducer — *See* Phonograph Pickup.

mechanical resistance—The real part of the mechanical impedance. The cgs unit is the mechanical ohm. The mks unit is the mks mechanical ohm.

mechanical scanning—An obsolete type of scanning in which a rotating device, such as a disc or mirror, breaks up a scene into a rapid succession of narrow lines for conversion into electrical impulses.

mechanical shock—Shock which occurs when the position of a system is significantly changed in a relatively short time in a nonperiodic manner. It is characterized by suddenness and large displacements which develop significant internal displacements within the system.

mechanical television system—A television system which uses mechanical scanning.

mechanical transducer—A device that transforms mechanical energy directly into acoustical energy.

mechanical transmission system — An assembly of gears, etc., for transmitting mechanical power.

mechanical tuning range — The frequency range of oscillation of a klystron that is obtainable by tuning mechanically while keeping the reflector voltage optimized for the peak of the reflector-voltage mode.

mechanical tuning rate—In a klystron, the frequency change per degree of rotation of the tuning apparatus while oscillation is maintained on the peak of the reflector-voltage mode.

mechanical wave filter—Also called mechanical filter. A filter that separates mechanical waves of different frequencies.

mechanical waveform synthesizer—A device which mechanically generates an electrical signal with the desired waveform.

mechanism of failure—*See* Failure Mode.

mechanized assembly—The joining together of elements by operators using semiautomatic equipment as contrasted to fully automatic assembly.

median—The middle, or average, value in a series (e.g., in the series 1, 2, 3, 4, and 5, the median is 3).

medical electronics—The branch of electronics concerned with its therapeutic or diagnostic applications in the field of medicine.

medium frequency — The band of frequencies between 300 kHz and 3000 kHz (3 MHz).

medium-power silicon rectifiers—Rectifiers with maximum continuous ratings of 1 to

50 average amperes per section in a single-phase, half-wave circuit.

meg—Abbreviation for megohm.

mega—Abbreviated M. Prefix denoting 10^6 (one million).

megabar—The absolute unit of pressure equal to one million bars.

megabit—A unit equal to one-million binary digits.

megacycle (mc)—One million cycles—Obsolete term replaced by megahertz. (MHz).

megahertz—One million hertz. Abbreviated MHz.

megampere—One million amperes. Abbreviated Ma.

megatron—A tube having a high power output at high frequencies, but very low interelectrode capacitances because its electrodes are arranged in parallel layers. *Also see* Disc-Seal Tube.

megavolt — One million volts. Abbreviated Mv.

megavolt-ampere — One million volt-amperes. Abbreviated Mva.

megawatt — One million watts. Abbreviated Mw.

megawatt-hour — One million watt-hours. Abbreviated Mwh.

megger—A high-range ohmmeter having a built-in, hand-driven generator as a direct voltage source, used for measuring insulation resistance values and other high resistances. Also used for continuity, ground, and short-circuit testing in general electrical power work.

megohm — Abbreviated meg. One million ohms.

megohm-farads—*See* Megohm-Microfarad.

megohm-microfarad—A term used to indicate the insulation resistance of capacitors. For larger high-voltage capacitors, megohm-farads are used.

megohm sensitivity—The resistance in megohms which must be placed in series with a galvanometer in order that an applied emf of 1 volt shall produce the standard deflection. If the resistance of the galvanometer coil itself is neglected, the number representing the megohm sensitivity is equal to the reciprocal of the number representing the current sensitivity.

Meissner effect—The sudden loss of magnetism in superconductors as they are cooled below the temperature required for superconductivity. As a result, they become diamagnetic—i.e., the self-induced magnetization opposes the applied magnetic field to such an extent that there is no longer a magnetic field.

Meissner oscillator—An oscillator in which the grid and plate circuits are inductively coupled through an independent tank circuit, which determines the frequency.

mel—A unit of pitch. A simple 1,000-hertz tone, 40 db above a listener's threshold, produces a pitch of 1,000 mels. The pitch of any sound that is judged by the listener to be *n* times that of a 1-mel tone is *n* mels.

meltback process—The making of junctions by melting a correctly doped semiconductor and allowing it to refreeze.

meltback transistor — A grown transistor produced by melting the tip of a double-doped pellet. Junctions are formed when the tip recrystallizes.

melting channel—The restricted portion of the charge in a submerged horizontal ring induction furnace. The induced currents are concentrated here to effect high energy absorption and thereby melt the charge.

melt-quench transistor—A junction transistor made by quickly cooling a melted-back region.

memory—*See* Storage.

memory buffer register — In a computer, a register in which a word is stored as it comes from memory (reading) or just prior to its entering memory (writing).

memory capacity—*See* Storage Capacity.

memory cycle—In a computer, an operation consisting of reading from and writing into memory.

memory dump—In a computer, a process of writing the contents of memory consecutively in such a form that it can be examined for computer or program errors.

memory fill—In a computer, the placing of a pattern of characters in the memory registers not in use in a particular problem to stop the computer if the program, through error, seeks instructions taken from forbidden registers.

memory register—Also called high-speed bus, distributor, or exchange register. In some computers, a register used in all data and instruction transfers between the memory, the arithmetic unit, and the control register.

memory relay—A relay in which each of two or more coils may operate independent sets of contacts, and another set of contacts remains in a position determined by the coil last energized. The term is sometimes erroneously used for polarized relay.

memory timer—A timing device wherein the cycle duration is infinitely variable within the specified overall cycle time and having the ability, once the cycle is selected, to repeat this cycle any number of times by a simple mechanical actuation of the time shaft.

memory unit—That part of a digital computer in which information is stored in machine language, using electrical or magnetic techniques.

mercuric-oxide-cadmium cell — A primary-cell electrochemical system. Its primary advantage is its long shelf life in the fully charged condition and its operation at low temperatures, far below 0°C. Nominal cell voltage is about 0.9 volt.

mercury—A silver-white metal that becomes a liquid above −38.87° Centigrade. In addition to thermometers, it is used in switches and many electronic tubes. When it is vaporized, mercury ionizes readily and conducts electricity.

mercury arc — A cold-cathode arc through ionized mercury vapor. A very bright bluish-green glow is given off.

mercury-arc converter — A frequency converter using a mercury vapor tube.

mercury-arc rectifier—Also called mercury-vapor or simply mercury rectifier. A diode rectifier tube containing mercury vapor. The mercury vapor is ionized by the voltage across the tube, and a much greater current can flow. There is only a small voltage drop in the tube during conduction.

mercury barometer — An instrument for measuring atmospheric pressure.

mercury battery—A type of battery characterized by extremely uniform output throughout its life. Employs a zinc-powder anode; the cathode is mercuric-oxide powder and graphite powder.

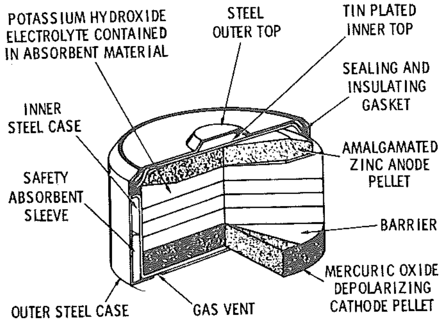
POTASSIUM HYDROXIDE ELECTROLYTE CONTAINED IN ABSORBENT MATERIAL
STEEL OUTER TOP
TIN PLATED INNER TOP
SEALING AND INSULATING GASKET
INNER STEEL CASE
AMALGAMATED ZINC ANODE PELLET
SAFETY ABSORBENT SLEEVE
BARRIER
MERCURIC OXIDE DEPOLARIZING CATHODE PELLET
OUTER STEEL CASE
GAS VENT

Mercury cell.

mercury cell—A primary cell with a zinc anode and a mercuric-oxide cathode in a potassium-hydroxide electrolyte. This cell offers a higher capacity than the alkaline or zinc-carbon cells and a much flatter voltage-discharge characteristic.

mercury-contact relay—A relay in which the contacts are mercury.

mercury delay line—*See* Mercury Memory.

mercury displacement relay — A relay in which the displacement of mercury, such as caused by a solenoid-actuated plunger, results in a mercury-to-mercury electrical contact.

mercury-hydrogen spark-gap converter—A spark-gap generator in which the source of radio-frequency power is the oscillatory discharge of a capacitor through an inductor and a spark gap. The latter comprises a solid electrode and a pool of mercury in hydrogen.

mercury-jet scanning switch—A commutating switch in which a stream of mercury performs the switching between the common circuit and those to be sampled.

mercury memory—Also called mercury storage and mercury delay line. Delay lines using mercury as the medium for storage of a circulating train of waves or pulses.

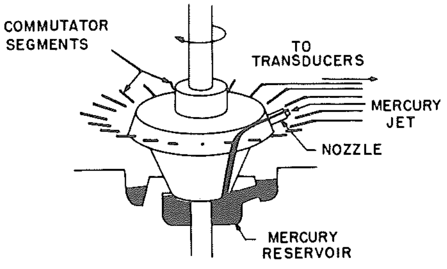

Mercury-jet scanning switch.

mercury-motor meter—A motor meter in which a portion of the rotor is immersed in mercury, which directs the current through conducting portions of the rotor.

mercury-pool cathode—The cathode of a gas tube consisting of a pool of mercury. An arc spot on the pool emits electrons.

mercury rectifier—*See* Mercury-Arc Rectifier.

mercury relay—A relay in which the energized coil pulls a magnetic plunger into a tube containing mercury. The plunger moves the mercury in order to make connection between the contacts.

mercury storage—*See* Mercury Memory.

mercury switch—An electric switch comprising a large globule of mercury in a metal or glass tube. Tilting the tube causes the mercury to move toward or away from the electrodes to make or break the circuit.

mercury tank—In a computer, a container of mercury holding one or more delay lines for storing information.

mercury-vapor lamp—A glow lamp in which mercury vapor is ionized by the electric current, producing a bluish-green luminous discharge.

mercury-vapor rectifier—*See* Mercury-Arc Rectifier.

mercury-vapor tube — A tube containing mercury vapor, which when ionized allows conduction and produces a luminous glow.

mercury-wetted contact relay — A special form of reed relay. It consists of a glass-encapsulated reed, with its base mounted in a pool of mercury and the other end arranged so as to move between two sets of stationary contacts. By capillary action, the mercury flows up the reed to coat the movable and stationary contact surrfaces, thus assuring mercury-to-mercury contact during make.

merge — To produce a single sequence of items ordered according to some rule (i.e., arranged in some orderly sequence), from two or more sequences previously ordered according to the same rule, without changing the size, structure, or total number of the items.

mesa isolation—A process for isolating IC's in which transistors, resistors, and other components are fabricated before isolation.

mesa transistor—A transistor that is produced by chemically etching away a transistor chip formed by either a double-diffused or diffused-alloy process. When the etching process is complete, the base and emitter regions appear as plateaus above the collector region. One result of the mesa construction is a reduction in the collector-base capacitance as a result of lowering the junction area.

mesh—Sometimes called a loop. A set of branches forming a closed path in a network—provided that if any branch is omitted, the remaining branches do not form a closed path.

mesh beat—*See* Moiré.

mesh current—The current assumed to exist over all cross sections of a closed path in a network. It may be the total current in a branch included in the path, or a partial current which, when combined with the others, forms the total current.

meson—A particle which weighs more than an electron but generally less than a proton. Mesons can be produced artificially as well as by cosmic radiation. They are so unstable that they disintegrate in millionths of a second.

message—1. An ordered selection of an agreed set of symbols for the purpose of communicating information. 2. The original modulating wave in a communication system.

metadyne—British term for amplidyne. A direct-current machine used for voltage regulation or transformation. It has more than two brushes for each pair of poles.

metal—A material that has high electrical and thermal conductivity at normal temperatures.

metal-base transistor—A transistor with a base of a thin metal film sandwiched between two N-type semiconductors, with the emitter doped more heavily than the base to give it a high electron-current-to-hole-current ratio.

metal detector—Also called metal locator. An electronic device for detecting concealed metal objects.

metal film resistor—An electronic component in which the resistive element is an extremely thin layer of metal alloy vacuum-deposited on a substrate.

metalized capacitor—A self-healing fixed capacitor. A thin film of metal is vacuum-deposited directly on the dielectric, which can be paper or plastic. When a breakdown occurs, the metal film around it immediately burns away.

metalized resistor — A fixed resistor in which the resistance element is a thin film of metal deposited on the surface of a glass or ceramic substrate.

metalizing—Applying a thin coating of metal to a nonmetallic surface. This may be done by chemical deposition or by exposing the surface to vaporized metal in a vacuum chamber.

metallic circuit—A circuit in which the earth itself is not used as ground.

metallic insulator—A shorted quarter-wave section of transmission line. It acts as an electrical insulator at a quarter-wave length of the transmitted frequency.

metallic noise—Weighted noise current in a metallic circuit at a given point when the circuit is terminated at that point in the nominal characteristic impedance of the circuit.

metallic rectifier—A rectifier in which the asymmetrical junction between dissimilar solid conductors presents a high resistance to current flow in one direction and a low resistance in the opposite direction.

metallic-rectifier cell—An elementary rectifying device having only one positive electrode, negative electrode, and rectifying junction.

metallic-rectifier stack—A single structure made up of one or more metallic-rectifier cells.

metal locator—*See* Metal Detector.

metal master—*See* Original Master.

metal negative—*See* Original Master.

metal-oxide resistor—A type of film resistor in which the material deposited on the substrate is tin oxide, which provides good stability.

metal-oxide semiconductor—The technique of insulating a metalized electrode, the "gate" from a semiconductor substrate material using an oxide (or nitride) dielectric to form a unipolar device. Abbreviated MOS.

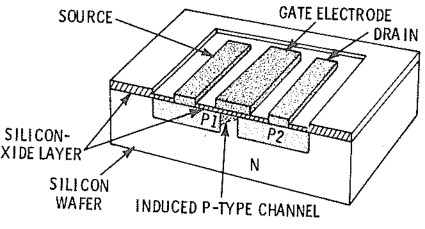

Metal-oxide semiconductor transistor.

metal-oxide semiconductor transistor (MOST)—An active semiconductor device in which conduction is controlled in a region between two electrodes by a voltage applied to an insulated electrode over the region.

metal-tank mercury-arc rectifier—A mercury-arc rectifier in which its anodes and mercury cathode are enclosed in a metal chamber.

metal tube—A vacuum tube with a metal envelope. The electrode leads pass through glass beads fused into the metal housing.

meteorological aids service—A radio service in which emissions consist of special signals intended solely for meteorological uses.

meteorological radar station—A station, operating in the meteorological aids service, that employs radar and is not intended for operation while in motion.

meter—1. Any measuring device; specifically, any electrical or electronic measuring instrument. 2. In the metric system of measurement, the unit of length equal to 39.37 inches, 3.281 feet, or 1.094 yards.

meter ampere—A measure of the strength of a radio transmitter. It is equal to the antenna height in meters times the maximum antenna current in amperes.

meter-candle—*See* Lux.

meter-correction factor — The factor by which a meter reading must be multiplied to compensate for meter errors in order to obtain the true reading.

meter display—A display in which one or more pointer instruments give the indications.

meter-kilogram-second-ampere system of units—*See* MKSA System of Units.

meter rating—A manufacturer's designation used to indicate the operating limitations of the meter. The full-scale marking on a meter scale does not necessarily correspond to the rating of the meter.

meter resistance—The resistance of a meter as measured at the terminals at a given reference temperature. When applied to rectifier-type meters, the frequency and waveshape of the applied energy, as well as the indicated value at which the measurement is to be made, must be specified.

meter-type relay—A meter movement in which the armature function is performed by a contact-bearing pointer.

metrechon—A storage tube used in scanning converters (e.g., in radar and industrial tv).

metric system—The decimal system of weights and measures, used extensively by scientists.

metric waves—British classification for wavelengths between 1 and 10 meters.

mev—Usually pronounced M-E-V. Abbreviation for million electron-volts. 1 mev equals 4.45 × 10^{-20} kilowatt hours, or 1.6020 × 10^{-6} erg.

MEW—Abbreviation for microwave early warning.

mf—1. Abbreviation for medium frequency. 2. Alternate abbreviation for microfarad.

mfd—Alternate abbreviation for microfarad.

mh—Abbreviation for millihenry.

MHD—Abbreviation for magnetohydrodynamics.

mho—The unit of conductance of a conductor when a potential difference of 1 volt between its ends maintains an unvarying current of 1 ampere. The mho is the unit of conductance in the mksa system. This term has now been replaced by siemens.

MHz—Abbreviation for megahertz.

mica—A transparent mineral which can be split into thin sheets. Because of its excellent insulating and heat-resisting qualities, it is used to separate the plates of capacitors and to insulate electrode elements in vacuum tubes.

mica capacitor—A fixed capacitor employing mica sheets as the dielectric material between adjacent plates. The complete units are usually encased in molded *Bakelite*.

micro—1. In the metric system, a prefix meaning one millionth (1/1,000,000). 2. A prefix meaning something very small.

microalloy diffused transistor—Abbreviated MADT. A microalloy transistor made by incorporating diffusion techniques. Prior to the electrochemical plating process, the semiconductor wafer is subjected to gaseous diffusion in order to provide a nonuniform base region.

microalloy transistor—A high-frequency transistor in which the emitter and collector are alloyed to a slight depth on opposite sides of the germanium base material by an electrochemical process. Abbreviated MAT.

microammeter—A meter, having a scale that reads in microamperes, for measuring extremely small currents.

microampere—One millionth of an ampere.

microbar—A unit of pressure commonly used in acoustics. One microbar is equal to one dyne per square centimeter. (Originally the term "bar" denoted 1 dyne per square centimeter. Therefore, to avoid confusion it is preferable to use microbar or "dynes per square centimeter" when speaking of sound pressures.)

micro B-display—A B-scope in which range and azimuth data are so expanded that only a small portion of the area under surveillance is presented, and, because of the degree of expansion, distortion of the presentation is negligible.

microcircuit—A packaged circuit made up of discrete components which are often indistinguishable to the naked eye, or of components fabricated with integrated-circuit techniques. *Also see* Microminiaturization.

microcircuitry—A small circuit having a high equivalent circuit element density which is considered as a single part and which is a combination of interconnection elements inseparably associated on or within a single substrate to perform an electronic circuit function. This excludes printed-wiring boards, circuit-card assemblies, and modules composed exclusively of discrete electronic parts.

microcircuit wafer—A microwafer carrying one or more circuit functions such as a flip-flop or gate. Integrated-circuit chips may be bonded to deposited conductors.

microcoding—In a computer, a system of coding that uses suboperations not ordinarily accessible in programming.

microcomponents—Those components smaller than existing components by several orders of magnitude.

microdensitometer—An instrument used in spectroscopy to measure lines in a spectrum by light-transmission measurement.

microelectronic circuit — Discrete electrical components assembled and connected in extremely small and compact form.

microelectronic device — An alternate term for "integrated circuit."

microelectronics — Also called microsystems electronics. The entire body of electronic art which is connected with or applied to the realization of electronic systems from extremely small electronic parts.

microelement—A resistor, capacitor, transistor, diode, inductor, transformer, or other electronic element or combination of elements mounted on a small ceramic wafer (0.01 inch thick and about 0.3 inch square). Individual microelements are stacked, interconnected, and potted to form micromodules.

microelement wafer—A microwafer carrying one or more components or a simple network. The network can consist, for example, of several thin-film resistors deposited directly on the wafer.

microfarad—One millionth of a farad. Abbreviated mfd, mf, or μf.

microfaradmeter—*See* Capacitance Meter.

microflash lamp — A lamp that emits radiation pulses having a duration of approximately one microsecond.

microgroove—In disc recording, the groove width of most long-play and 45-rpm records. Normally it is .001 inch, or about half as wide as the groove on a 78-rpm record.

microgroove record—*See* Long-Play Record.

microhenry—One millionth of a henry.

microhm—One millionth of an ohm.

microlock — A phase-lock-loop system for transmitting and receiving information. Because the system reduces bandwidth drastically, it is used as a radar beacon for tracking, or to provide telemetering data.

micrologic — A group of high-speed, low-power integrated logic building blocks primarily intended to be used in building the logic section of a digital computer.

micrologic elements—Semiconductor networks used in computer and other critical circuits.

micromassage—*See* Intercellular Massage.

micromho—One millionth of a mho.

micromicro — An obsolete prefix meaning one-millionth of a millionth, or 10^{-12}. Now called pico.

micromicrofarad — Abbreviated mmf or mmfd. One millionth of a microfarad, or one picofarad.

micromicrowatt — One picowatt, or 10^{-12} watt.

microminiature lamp — Any incandescent lamp, usually rated in the milliwatt range, that operates on 3 volts or less. Diameters range from 0.01 to 0.06 inches.

microminiaturization — 1. The producing of microminiature electronic circuits from individual miniature solid-state and other nonthermionic components. 2. A relative degree of miniaturization resulting in an equipment or assembly volume an order of magnitude smaller than that existing in subminiature equipment. 3. The technique of packaging a microminiature part of an assembly composed of elements radically different in shape and form. Electronic parts are replaced by active and passive elements, through use of fabrication processes such as screening, vapor-deposition diffusion, and photoetching.

micromodule—A tiny ceramic wafer made from semiconductive and insulative materials. It is capable of functioning as either a transistor, resistor, capacitor, or other basic component.

micromodules—The packaging approach developed by the U.S. Army and RCA wherein individual components or combined components are fabricated on small square wafers and made into larger modules by stacking.

micron—1. A unit of length equal to 10^{-6} meter. 2. A unit used in the measurement of very low pressures. It is equivalent to 0.001 mm (10^{-6} meter) of mercury at 32° F.

microphone — An electroacoustic transducer which responds to sound waves and delivers essentially equivalent electric waves.

microphone amplifier—Also called a microphone preamplifier. An audio-frequency amplifier that boosts the output of a microphone before the signal reaches the main audio-frequency amplifier.

microphone boom—A movable crane from which a microphone is suspended.

microphone button—The resistance element of a carbon microphone. It is button-shaped and filled with carbon particles.

microphone cable — A shielded cable for connecting a microphone to an amplifier.

microphone mixer — An audio mixer that feeds the output from two or more microphones into a single input to an audio amplifier. The output from each microphone is adjustable by individual controls on the mixer.

microphone preamplifier — *See* Microphone Amplifier.

microphone stand—A stand that holds a microphone the desired distance above the floor or a table.

microphone transformer — An iron-core transformer used for coupling certain microphones to an amplifier or transmission line.

microphonics—The mechanical translation, by a vacuum tube, of vibration or shock into an electrical signal.

microphonism—1. The production of noise as a result of magnetic shock or vibration. 2. The quasiperiodic voltage output of a tube produced by mechanical resonance of its elements as a result of mechanical impulse excitation. 3. The periodic voltage output of a tube produced by mechanical resonances of its elements as a result of sustained mechanical excitation. 4. The output voltage of a tube action as an electrical transducer or mechanical energy.

microphonograph—A device which amplifies

and records weak sounds; used in training the deaf to speak.

microphonoscope — A binaural stethoscope using a membrane in the chest piece to accentuate the sound.

microphotograph—A small picture of a large subject. The microfilming of a check or other document produces a microphotograph.

microprogramming—In a computer, a technique in which the coder builds his own machine instructions from the basic, elemental instructions built into the hardware.

microradiometer—Also called a radio micrometer. A thermosensitive detector of radiant power. It consists of a thermopile supported on and connected directly to the moving coil of a galvanometer.

microsecond—One millionth of a second.

microstrip—A microwave transmission component utilizing a single conductor supported above ground.

microsyn—A precise and sensitive pick-off device for converting angular displacement within a small range to an electrical signal.

microsystems electronics — *See* Microelectronics.

microvolt—One millionth of a volt.

microvoltmeter — A highly sensitive voltmeter, which measures millionths of a volt.

microvolts/meter/mile — One method of stating the field strength of a radiated field. Radiation from industrial heating equipment, for example, must be suppressed so that the radiated field strength does not exceed 10 microvolts per meter at a distance of 1 mile from the source.

microvolts per meter—The potential difference in microvolts developed between an antenna system and ground, divided by the distance in meters between the two points.

microwafer — A basic microcircuit building block generally made of beryllia, alumina, or glass. Terminations on the edges are usually of gold on top of chromium, with a heavy nickel overlay for welding.

microwatt—One millionth of a watt.

microwave—A term applied to radio waves in the frequency range of 1,000 megahertz and upward. Generally defines operations in the region where distributed constant circuits enclosed by conducting boundaries are used instead of the conventional lumped-constant circuit components.

microwave discriminator—A tuned cavity which converts a frequency-modulated microwave signal into an audio or video signal.

microwave early warning — Abbreviated MEW. A high-power, long-range, early-warning radar. It has numerous indicators that give high resolution and large traffic-handling capacity.

microwave filter — A filter built into a microwave transmission line to pass desired frequencies but reject or absorb all other frequencies.

microwave frequencies—Frequencies of 890 MHz and above.

microwave oscillator — An oscillator that generates frequencies higher than about 300 megahertz (wavelengths shorter than about one meter).

microwave power transmission—A method of transmitting power through space from a microwave transmitting antenna to a remotely located receiving antenna.

microwave radio relay — The relaying of long-distance telephone calls and television broadcast programs by means of highly directional high-frequency radio waves that are received and sent on from one booster station to another.

microwave refractometer — A device for measuring the refractive index of the atmosphere at microwave frequencies, usually in the 3-cm region.

microwave region—The portion of the electromagnetic spectrum between the far infrared and conventional radio-frequency portion. Commonly regarded as extending from 1,000 (30 cm) to 300,000 (1 mm) megahertz.

microwave relay system—A series of ultra-high-frequency radio transmitters and receivers comprising a system for handling communications (usually multichannel).

microwaves — Radio frequencies with such short wavelengths that they exhibit some of the properties of light. Their frequency range is from 1,000 MHz and up. (Microwaves are preferred in point-to-point communications because they are easily concentrated into a beam.)

middle marker — In an instrument-landing system, a marker located on a localizer course line, about 3,500 feet from the approach end of the runway.

middle-side system—*See* M-S Stereo System.

migration—The movement of ions through a medium.

mike—Slang for microphone.

mil—One thousandth of an inch. Used in the United States for measuring wire diameter.

MIL—Abbreviation for military. Pertains to a nation's armed forces, including its army, navy, and air force. Specifically, the armed forces of the United States.

Miller bridge—A type of bridge circuit for measuring the amplification factor of vacuum tubes.

Miller effect—The increase in the effective grid-to-cathode capacitance of a vacuum tube because the plate induces a charge electrostatically on the grid through grid-to-plate capacitance.

Miller oscillator—A crystal-controlled oscillator in which the crystal oscillates at its parallel resonant frequency due to the connection of negative resistance across its plates.

milli—Abbreviated m. Prefix meaning one thousandth (1/1000, or 10^{-3}).

milliammeter — An electric current meter calibrated in milliamperes.

milliampere — Abbreviated ma. One one-thousandth (.001) of an ampere.

millihenry—Abbreviated mh. One one-thousandth (.001) of a henry.

millilambert—A unit of brightness equal to one one-thousandth (.001) of a lambert.

millimaxwell—One-thousandth of a maxwell.

millimeter waves—Electromagnetic radiation in the frequency range of 30 to 500 gigahertz with corresponding wavelengths of 10 millimeters to 0.6 millimeter.

millimicro—Obsolete prefix for nano, representing 10^{-9}.

millimicron—A unit of length equal to one ten-millionth of a centimeter (10^{-7} cm), or one one-thousandth of a micron.

milliohm—One one-thousandth (.001) of an ohm.

milliroentgen — A unit of radioactive dose equal to one one-thousandth of a roentgen.

millisecond — One-thousandth of a second. Abbreviated msec.

millitorr—One-thousandth of a torr.

millivolt—One-thousandth of a volt. Abbreviated mv.

millivoltmeter—A sensitive voltmeter calibrated in millivolts.

millivolts per meter—The potential difference in millivolts developed between an antenna system and ground, divided by the distance in meters between the two points.

milliwatt—One one-thousandth of a watt. The reference level used for db measurements.

mine detector—Also called electronic mine detector. An electronic device that indicates the presence of metallic or nonmetallic explosive mines under the ground or under water.

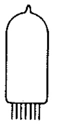

Miniature tube.

miniature tube—A small electron tube usually having a 7- or 9-pin base.

miniaturization — The process of reducing the minimum volume required by equipments or parts in order to perform their required functions.

minigroove—A recording having more lines per inch than the average 78-rpm phonograph record but not as many as the extended play, long playing, or microgroove records.

minimum-access programming—Also called minimum-latency programming, or forced coding. Programming a digital computer so information is obtained from the memory in the minimum waiting time.

minimum detectable signal — The signal level which just exceeds the threshold.

minimum discernible signal — Abbreviated MDS. In a receiver, the smallest input signal that will produce a discernible signal at the output. The smaller the signal required, the more sensitive the receiver.

minimum firing power—The lowest radio-frequency power that will initiate a radio-frequency discharge in a switching tube at a specified ignitor current.

minimum flashover voltage — The crest value of the lowest voltage impulse of a given waveshape and polarity which causes flashover.

minimum-latency programming—*See* Minimum-Access Programming.

minimum-phase network — A network for which the phase shift at each frequency equals the minimum value determined solely by the attenuation-frequency characteristic.

minimum reject number—A number that defines the maximum number of rejects allowed for each sample given for each LTPD. Abbreviated MRN.

minimum (reliable) current—As applied to relay contacts, the range at which there is insufficient energy under arcing conditions at a mating contact surface to insure good contacting for the kind of contact material, shape, and forces employed.

minimum resistance — The resistance of a precision potentiometer, measured between the wiper terminal and any terminal with the shaft positioned to give a minimum value.

minimum-signal level — In facsimile, the level corresponding to the copy white or copy black signal, whichever is the lower.

minimum starting voltage—The minimum voltage of rated frequency applied to the control-voltage winding of a servomotor necessary to start the rotor turning at no-load conditions with rated voltage and frequency on the fixed-voltage winding.

minitrack—A satellite tracking system which uses a miniature pulse-type telemeter and a precise directional antenna system, with phase-comparison tracking techniques.

minor apex face—In a quartz crystal, any one of the three smaller sloping faces near but not touching the apex (pointed end). The other three are called the major-apex faces.

minor bend—A rectangular waveguide bent so that one of its longitudinal axes is parallel to its narrow side throughout the length of the bend.

minor cycle—Also called word time. In a digital computer using serial transmission, the time required to transmit one word or the space between words.

minor face—One of the three smaller sides of a hexagonal quartz crystal.

minor failure—A failure which has no significant effect on the satisfactory performance of a system.

minority carrier — The less predominant carrier in a semiconductor. Electrons are the minority carrier in P-type semiconductors, since there are fewer electrons than holes. Likewise, holes are the minority carrier in N-types, since they are outnumbered by the electrons.

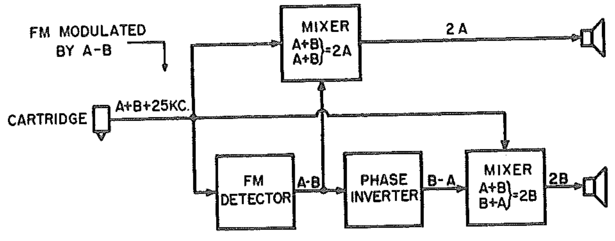

Minter stereo-disc playback.

minority emitter—An electrode from which minority carriers flow into the inter-electrode region.

minor lobe—Any lobe except the major lobe.

minor loop—A continuous network composed of both forward and feedback elements, which is only part of the overall feedback control system.

Minter stereo system—A stereo recording technique for producing the right and left channels. The two program channels are combined additively and recorded with a monophonic cutter. A 25-kHz note is also recorded and is frequency-modulated by the two channels combined subtractively. The sum and difference signals are then matrixed (combined) to produce the right and left channels.

mirror galvanometer—A suspended-coil instrument which, instead of using a pointer to indicate the reading, employs a light beam reflected from a mirror attached to the moving coil.

mirror galvanometer oscillograph—An instrument that photographs the deflection of a light spot from a mirror attached to a moving coil. Used for recording small current variations.

mirror-reflection echoes — Multiple-reflection echoes produced when a radar beam is reflected from a large, flat surface (such as the side of an aircraft carrier) and strikes nearby targets.

misfire—Failure of a mercury-pool-cathode tube to establish an arc between the main anode and cathode during a scheduled conducting period.

mismatch—The difference in impedance between a load and its source.

mismatch factor—*See* Reflection Factor.

mismatch loss—The ratio between the power a device would absorb if it were perfectly matched to the source, and the power it actually does absorb.

mistake—*See* Error.

mistor—A magnetic-field sensing device the resistance of which increases with an increase in magnetic-field intensity.

mitte-seite stereo system—German for middle-side system. A technique of stereo pickup in which two directional microphones, placed close together and at right angles to each other, are oriented so that one picks up sound from directly ahead, and the other from the two sides, with maximum intensity.

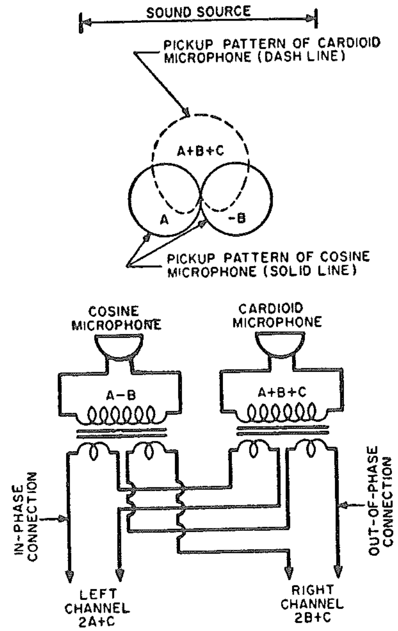

Mitte-seite stereo system.

mixed-base notation — A number system in which a single base, such as 10 in the decimal system, is replaced by two number bases, such as 2 and 5, used alternately.

mixed highs—In color television, the method of reproducing very fine picture detail by transmitting high-frequency components as part of the luminance signal for achromatic reproduction.

mixer—1. In a sound transmission, recording or reproducing system, a device having two or more inputs (usually adjustable) and a common output. The latter combines the separate input signals linearly in the desired proportion to produce an output signal. 2. Also called the first detector. In a superheterodyne receiver, the stage where the incoming signal is modulated with the signal from the local oscillator to produce the intermediate-frequency signal.

mixer tube—An electron tube which, when supplied with voltage or power from an external oscillator, performs the frequency-conversion function of a heterodyne-conversion transducer.

mixing—Combining two or more signals—e.g., the outputs of several microphones, or the received and local-oscillator signals in a superheterodyne receiver.

mixing amplifier—An amplifier which combines several signals, each with a different amplitude and waveshape, into a composite signal.

mixing point—In a block diagram of a feedback control loop, a symbol indicating that the output is a function of the inputs at any instant.

mksa electromagnetic system of units—Also called the Giorgi system. A system in which the fundamental units are the meter, kilogram, second, and ampere.

mm—Abbreviation for millimeter.

mmf—Abbreviation for micromicrofarad; replaced by picofarad.

mmfd—Abbreviation for micromicrofarad; replaced by picofarad.

mnemonic – A term describing something used to assist the human memory.

mnemonic code—Computer instructions written in a form easy for the programmer to remember (although they must later be converted into machine language).

mobile radio service—Radio service between a fixed location and one or more mobile radio stations, or between mobile stations.

mobile receiver—The radio receiver in an automobile, truck, or other vehicle.

mobile-relay station—A type of base station in which the base-station receiver automatically turns on the base-station transmitter which then retransmits all signals received by the base-station receiver. Such a station is used to extend the range of mobile units and requires two frequencies for operation.

mobile station—A radio station intended for use while in motion or during halts at unspecified points. Included are hand- and pack-carried units.

mobile telemetering—Electric telemetering between moving objects, where interconnecting wires cannot be used.

mobile transmitter—A radio transmitter installed and operated in a vessel, land vehicle, or aircraft.

mobility—Symbolized by the Greek letter *mu* (μ). The ease with which carriers move through a semiconductor when they are subjected to electric forces. In general, electrons and holes do not have the same mobility in a given semiconductor. Also, their mobility is higher in germanium than in silicon.

mod/demod—Abbreviated form of modulating and demodulating.

mode—1. One of several types of electromagnetic wave oscillations that may be sustained in a given resonant system. Each type of vibration is designed as a particular mode and has its own particular electric- and magnetic-field configurations. 2. One of several methods of exciting a resonant system. The term has also been used to describe the existence of a number of different input voltages which allow operation of a klystron at the same frequency. 3. A computer system of data representations (e.g., the binary mode). 4. The most frequent value in a series of measurements or operations.

mode coupling—The interaction or exchange of energy between similar modes.

mode filter—A waveguide filter designed to separate waves of the same frequency but of different transmission modes.

mode hopping—The random shifting of laser output energy from one mode to another for short durations, usually in the microsecond range. The modes are generally closely spaced.

mode jump—A change in the mode of magnetron operation from one pulse to the next. Each mode represents a different frequency and power level.

modem – Acronym for MOdulator DEModulator.

mode number – 1. In a reflex klystron, the number of whole cycles that a mean-speed electron remains in the drift space. 2. The number of radians of phase shift resulting from going once around the anode of a magnetron divided by 2π.

mode of failure—*See* Failure Mode.

mode of resonance—A form of natural electromagnetic oscillation in a resonator. It is characterized by an unvarying field pattern.

mode (of transmission propagation)—1. A form of guided-wave propagation characterized by a particular field pattern that intersects the direction of propagation. The field pattern is independent of its position along the waveguide. For a uniconductor waveguide, it also is independent of frequency. 2. In a vibrating system, the state which corresponds to one of the resonant frequencies.

mode of vibration—1. The pattern formed by the movement of the individual particles in a vibrating body (e.g., a piezoelectric crystal). This pattern is determined by the stresses applied to the body, the properties of the body, and the boundary conditions. The three common modes of vibration are flexural, extensional, and shear. 2. In a vibrating system, its characteristic pattern (i.e., the motion of every particle) is a simple harmonic, all of the same frequency. More than one mode may exist concurrently in a multiple-degree-of-freedom system.

mode purity—The freedom of an ATR tube from undesirable mode conversion while the tube is in its mount.

mode separation—In an oscillator, the difference in frequency between resonator modes of oscillation.

mode shift—In a magnetron, the change in mode during a pulse.

mode skip—Failure of a magnetron to fire during each successive pulse.

mode transducer—Also called mode transformer. A device that transforms an electromagnetic wave from one mode of propagation to another.

mode transformer—*See* Mode Transducer.

modification – The physical alteration of a system, subsystem, etc., for the purpose of changing its designed capabilities or characteristics.

modified constant-voltage charge – The charging of a storage battery in which the voltage of the charging circuit is held substantially constant, but a fixed resistance is inserted in the battery circuit, producing a rising voltage characteristic at the battery terminals as the charge progresses.

modified index of refraction – Also called modified refractive index. In the troposphere, the index of refraction at any height increased by h/a, where h is the height above sea level and a is the mean geometrical radius of the earth. When the index of refraction in the troposphere is horizontally stratified, propagation over a hypothetical flat earth through an atmosphere with the modified index of refraction is substantially equivalent to propagation over a curved earth through the real atmosphere.

modified refractive index—*See* Modified Index of Refraction.

moding—A defect of magnetron oscillation in which the magnetron oscillates in one or more undesired modes.

modular—1. Made up of modules. 2. Dimensioned according to a prescribed set of size increments.

modulate—To vary the amplitude, frequency, or phase of a wave by impressing one wave on another wave of constant properties.

modulated amplifier—In a transmitter, the amplifier stage where the modulating signal is introduced and modulates the carrier.

modulated-beam photoelectric system—An intrusion-detector system which provides reliable beam ranges of several thousand feet. This is done by interrupting the light beam at the source with a rotating punched or slotted disc. In this way, the phototube output is converted to an a-c signal, which can then be easily amplified.

modulated carrier—A radio-frequency carrier in which the amplitude or frequency has been varied by, and in accordance with, the intelligence to be conveyed.

modulated continuous wave – Abbreviated mcw. A wave in which the carrier is modulated by a constant audio-frequency tone. In telegraphy service, the carrier is keyed to produce the modulation.

modulated light – Light whose intensity has been made to vary in accordance with variations in an audio-frequency or code signal.

modulated oscillator—An oscillator the output frequency of which is varied by an input signal.

modulated signal generator—A device which produces an output signal that may be changed in amplitude and/or frequency according to a desired pattern. It is calibrated in units of both power (or voltage) and frequency.

modulated stage—The radio-frequency stage to which the modulator is coupled and in which the continuous wave (carrier wave) is modulated, in accordance with the system of modulation and the characteristics of the modulating wave.

modulated wave—A carrier wave in which the amplitude, frequency, or phase varies in accordance with the intelligence signal being transmitted.

modulating amplifier using variable reactance—A very-high-frequency electron-beam parametric amplifier with bandpass characteristics that are independent of gain, and which is unconditionally stable.

modulating electrode – In a cathode-ray tube, an electrode to which a potential is applied to control the magnitude of the beam current.

modulating signal—*See* Modulating Wave.

modulating wave—Also called modulating signal, or simply signal. A wave which varies some characteristic (i.e., frequency, amplitude, phase) of the carrier.

modulation—The process of modifying some characteristic of a wave (called a carrier) so that it varies in step with the instantaneous value of another wave (called a modulating wave or signal). The carrier can be a direct current, an alternating current (provided its frequency is above the highest frequency component in the modulating wave), or a series of regularly repeating, uniform pulses called a pulse chain (provided their repetition rate is at least twice that of the highest frequency to be transmitted).

modulation capability—The maximum percentage of modulation possible without objectionable distortion.

modulation code—A code used to cause variations in a signal in accordance with a predetermined scheme; normally used to alter or modulate a carrier wave to transmit data.

modulation distortion—Distortion occurring in the radio-frequency amplifier tube of a receiver when the operating point is at the bend of the grid-voltage–plate-current characteristic curve. As a result, the plate-current changes are greater on positive than on negative half-cycles. The effect is equivalent to an increase in the percentage of modulation.

modulation envelope – A curve, drawn through the peaks of a graph, showing how the waveform of a modulated carrier represents the waveform of the intelligence carried by the signal. The modulation envelope is the intelligence waveform.

modulation factor—In an amplitude-modulated wave, the ratio of half the difference

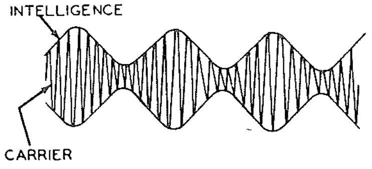

Modulation envelope.

between the maximum and minimum amplitudes to the average amplitude. This ratio is multiplied by 100 to obtain the percentage of modulation.

modulation frequency – That signal which causes the output frequency of an oscillator to be modulated.

modulation index—In frequency modulation with a sinusoidal modulating wave, the ratio of the frequency deviation to the frequency of the modulating wave.

modulation meter—Also called modulation monitor. An instrument for measuring the degree of modulation (modulation factor) of a modulated wave train. It is usually expressed in percent.

modulation monitor—*See* Modulation Meter.

modulation noise—Also called noise behind the signal. The noise caused by the signal, but not including the signal.

modulation percentage – The modulation factor multiplied by 100.

modulation plan – The arrangement by which the individual channels or groups of channels are modulated to a final frequency allocation.

modulation ratio—For an electrically modulated source, the number obtained by dividing the percentage of radiation modulation by the percentage of current modulation.

modulation rise—An increase in the modulation percentage. It is caused by a nonlinear tuned amplifier, usually the last intermediate-frequency stage of a receiver.

modulator—1. A device that effects the process of modulation. 2. In radar, a device for generating a succession of short energy pulses which cause a transmitter tube to oscillate during each pulse. 3. An electrode in a spacistor.

modulator crystal—A crystal which is used to modulate a polarized light beam by the use of Pockel's effect. Useful as a modulator in laser systems.

modulator driver—A transmitter circuit that produces a pulse to be delivered to the control grid of the modulator stage.

modulator glow tube—A cold-cathode recorder tube used for facsimile and sound-on-film recording. It provides a modulated, high-intensity, point-of-light source.

modulator stage—The last stage in which the modulated radio-frequency wave is amplified.

module—A unit in a packaging scheme displaying regularity and separable repetition. It may or may not be separable from other modules after initial assembly. Usually all major dimensions are in accordance with a prescribed series of dimensions.

modulus—An integer designating the number of states through which a counter sequences during each cycle. Abbreviated mod.

moiré—A wavy or satiny effect produced by the convergence of lines. Usually appears as a curving of the lines in the horizontal wedges of the test pattern and is most pronounced near the center, where the lines forming the wedges converge. A moiré pattern is a natural optical effect when converging lines in the picture are nearly parallel to the scanning lines. To a degree this effect is sometimes due to the characteristics of color picture tubes and of image-orthicon pickup tubes (in the latter, termed "mesh beat").

moisture absorption—The amount of moisture (in percentage) that an insulation will absorb. The figure should be as low as possible when the insulation is to be used in a moist environment.

moisture repellent—Having properties such that moisture will not penetrate.

moisture resistance—The ability of a material not to absorb moisture, either from the air or from being immersed in water.

moisture resistant – Having characteristics such that exposure to a moist atmosphere will not readily lead to a malfunction.

mol—Abbreviation for molecular weight.

mold—In disc recording, a metal part derived from a master by electroforming. It is a positive of the recording (i.e., it has grooves similar to those of a recording and can thus be played).

molded capacitor—A capacitor that has been encased in a molded plastic insulation.

molecular circuitry—*See* Morphological Circuitry.

molecular circuits – *See* Monolithic Integrated Circuits.

molecular electronics—Abbreviated moletronics. The science of making a single block of matter perform the function of a complete circuit. This is done by merging the function with a material, using solid-state functional blocks.

molecular technique—A practical method of causing a single piece of material or a crystal to provide a complete circuit function.

molecular weight—The sum of the weights of all the atoms in a molecule, expressed relative to the arbitrary weight of atomic oxygen ($O = 16$). Abbreviated mol.

molecule—In any substance, the smallest particle that still retains the physical and chemical characteristics of that substance. A molecule consists of one or more atoms of one or more elements. Sometimes two entirely different substances may have similar chemical elements, but their atoms will be arranged in a different order.

moletronics – Acronym for molecular electronics.

molybdenum—A metallic element (chemical symbol, *Mo;* atomic number, 42; atomic weight, 95.95) sometimes used for the grid and plate electrodes of vacuum tubes.

momentary-contact switch – A mercury switch designed to make contact for only a brief transitory interval while the mercury moves from one extreme position to another.

momentary switch—A switch which returns to its normal circuit condition when the actuating force is removed.

monaural—*See* Monophonic.

monaural recorder – Literally, a tape recorder intended for listening with one ear only; however, in popular usage refers to single-channel recorders, as distinguished from multichannel (stereophonic, binaural, etc.) types. More correctly but less universally called monophonic recorder.

monitor—1. "To listen . . ." to a communication service, without disturbing it, to determine its freedom from trouble or interference. 2. A device (e.g., a receiver, oscilloscope, teleprinter, etc.) used for checking signals.

monitor head—On some tape recorders an additional playback head which permits listening to the recorded material while it is being made.

monitoring—Observing the characteristics of transmitted signals as they are being transmitted.

monitoring amplifier – A power amplifier used primarily for evaluation and supervision of a program.

monitoring radio receiver—A radio receiver for checking the operation of a transmitting station.

monkey chatter—So called because of the garbled speech or music heard along with the desired program. This interference occurs when the side frequencies of an adjacent-channel station beat with the signal from the desired station.

monobrid—A method of manufacturing an integrated circuit by using more than one monolithic chip within the same package.

monobrid circuit—An integrated circuit using a combination of monolithic and multichip techniques by means of which a number of monolithic circuits or a monolithic device in combination with separate diffused or thin-film components are interconnected in a single package.

monochromatic—1. Pertaining to or consisting of a single color. 2. Radiation of a single wavelength.

monochromatic emissivity—*See* Total Emissivity.

monochromatic sensitivity—The response of a device to light of a given color only.

monochromator—An instrument used to isolate narrow portions of the spectrum by making use of the dispersion of light into its component colors.

monochrome—Also called black-and-white in referring to television. Having only one chromaticity—usually achromatic, or black and white and all shades of gray.

monochrome channel—In a color television system, any path intended to carry the monochrome signal (although it may carry other signals also).

monochrome channel bandwidth – The bandwidth of the path that carries the monochrome signal.

monochrome signal – 1. In a monochrome television transmission, the signal wave that controls the luminance values in the picture. 2. In a color-television transmission signal wave, the portion with major control of luminance—whether displayed in color or monochrome.

monochrome television—Also called black-and-white television. Television in which the final reproduced picture is monochrome. That is, it has only shades of gray between black and white.

monochrome transmission – Also called black-and-white transmission. In television, the transmission of a signal wave for controlling the luminance—but not the chromaticity—values in the picture.

monoclinic—A crystal structure in which two of the three axes are perpendicular to the third, but not to each other.

monoergic—A type of emission in which the particles or radiations are produced with a small energy spread (i.e., a "line spectrum").

monogroove stereo—Also called single-groove stereo. A stereo recording in which both channels are contained in one groove.

monolithic—1. Existing as one large, undifferentiated whole. 2. Formed within a single body of material.

monolithic circuit—A circuit consisting of a semiconductor substrate into which all of the parts are fabricated by diffusion and epitaxial processes.

monolithic integrated circuit – An integrated circuit the elements of which are formed *in situ* upon or within a semiconductor substrate with at least one of the elements formed within the substrate. (*Also see* Integrated Circuit.) To further define the nature of a monolithic integrated circuit, additional modifiers may be prefixed. Examples are: PN junction isolated monolithic integrated circuit, dielectric isolated monolithic integrated circuit, or beam-lead monolithic integrated circuit.

monophonic—Also called by the older term "monaural." Pertaining to audio information on one channel (e.g., as opposed to binaural or stereophonic). Monophonic and monaural are usually, although not necessarily, associated with a one-speaker system.

monophonic recorder – *See* Monaural Recorder.

monopinch – Antijam application of the monopulse technique in which the error signal is used to provide discrimination against jamming signals.

monopole – A stub antenna fed against a ground plane.

monopulse—A method of determining azimuth and elevation angles simultaneously. The antenna system is not mechanically lobed in this method.

monopulse radar — Radar using a receiving antenna system having two or more partially overlapping lobes in the radiation pattern. Sum and difference channels in the receiver compare the amplitudes or phases of the antenna outputs.

monopulse tracking—A form of telemetry tracking which compares the phases of the carrier received simultaneously at four points on a plane (two elevation and two azimuth references) and positions the antenna so that all four signals are in phase, or equidistant from the source of radiation. A variation of the monopulse tracking technique is the scan-coded or single-channel monopulse tracking system.

monorange speaker—A speaker that provides the full spectrum of audio frequencies.

monoscope — Also called phasmajector or monotron. An electron-beam tube in which the picture signal is generated by scanning an electrode, parts of which have different secondary-emission characteristics.

monostable — The term used to describe a circuit with one stable state and one quasi-stable state. The circuit requires an external trigger to perform one cycle.

monostable multivibrator—Also called one-shot multivibrator, single-shot multivibrator, or start-stop multivibrator. A circuit having only one stable state, from which it can be triggered to change the state, but only for a predetermined interval, after which it returns to the original state.

monostatic reflectivity — The characteristic of a reflector which reflects energy only along the line of the incident ray (e.g., a corner reflector).

monotron—*See* Monoscope.

Monte Carlo method—Abbreviated MCM. 1. A computer technique in which a number of possible models under study are mathematically constructed from constituents selected at random from representative populations. 2. Any procedure that involves statistical sampling techniques in order to obtain an approximate solution of a mathematical or physical problem.

MOPA—Acronym for master oscillator-power amplifier.

morphological circuitry—Also called molecular circuitry. A circuit made from a material in which the molecular structure has been arranged to perform a certain electrical function.

Morse code—A system of dot-and-dash signals developed by Samuel F. B. Morse and now used chiefly in wire telegraphy.

Morse sounder—A telegraph receiving instrument that produces a sound at the beginning and end of each dot and dash. From these sounds, a trained operator can interpret the message.

Morse telegraphy—Telegraphy in which the Morse code or its derivative is used—specifically, the International (also called Continental) or American Morse code.

MOS—Abbreviation for metal-oxide semiconductor.

mosaic—The light-sensitive surface of an iconoscope or other television camera tube. In one form, it consists of millions of tiny silver globules on a sheet of ruby mica. Each globule is treated with cesium vapor to make it sensitive to light.

mosaic detector—A device in which a number of active elements are arranged in an array. It is generally used as an imaging device.

MOS monolithic IC—A single-chip IC, consisting largely of interconnected unipolar active-device elements or MOS field-effect devices. This class of circuits results in higher equivalent parts density.

MOST—Abbreviation for metal-oxide semiconductor transistor.

mother—In disc recording, a mold electroformed from the master.

motherboard—A relatively large piece of insulating material on which components, modules, or other electronic subassemblies are mounted and interconnections made by welding, soldering, or other means, using point-to-point or matrix wire, or circuitry fabricated integrally with the board.

mother crystal—The quartz crystal found in nature. It has the characteristic geometric design of a crystal (i.e., flat faces at definite angles to each other), but all or some of the faces may be worn because of abrasion with stones or other objects.

motional impedance—Also called loaded motional impedance. The complex remainder after the blocked impedance of a transducer has been subtracted from the loaded impedance.

motion frequency — The natural frequency of a servo system; the frequency at which a given servo tends to oscillate.

motion-picture pickup—A television camera or technique for televising scenes directly from motion-picture film.

motor—A device that moves an object. Specifically, a machine that converts electric energy into mechanical energy.

motor board — Also called tape-transport mechanism. The platform or assembly of a tape recorder on which the motor (or motors), reels, heads, and controls are mounted. It includes parts of the recorder other than the amplifier, preamplifier, speaker, and case.

motorboating—1. Interference heard as the characteristic "putt-putt" made by a motorboat. It is due to self-oscillation, usually pulsating, in an amplifier below or at a low audio frequency. 2. A generally periodic, relatively low-frequency pulse disturbance of the output voltage of a regulated power supply, frequently line- or load-dependent, unstable, and significantly large, as in the case of oscillations.

motor circuit switch—*See* Disconnect Switch.

motor converter—A device for converting an alternating current to a pulsating direct current. It consists of an induction motor to which an a-c supply is connected. The armature of the induction motor is linked mechanically to the armature of a synchronous converter, which is connected to a d-c circuit.

motor-driven relay — A relay in which the contacts are actuated by the rotation of a motor shaft.

motor effect—The repulsion force exerted between adjacent conductors carrying currents in the opposite direction.

motor-field control—The method of controlling the speed of a motor by changing the magnitude of its field current.

motor-field failure relay (field loss relay) —A relay which functions to disconnect the motor armature from the line in the event of loss of field excitation.

motor-field induction heater—An induction heater in which the inducing winding typifies that of a rotary or linear induction motor.

motor-generator set — A motor-generator combination for converting one kind of electric power to another (e.g., alternating current to direct current). The two are mounted on a common base and their shafts are coupled together.

motor junction (conduit) box—An enclosure on a motor for the purpose of terminating a conduit run and joining the motor to power conductors.

motor meter—A meter comprising a rotor, one or more stators, and a retarding element which makes the speed of the rotor proportionate to the quantity being measured (e.g., power or current). A register, connected to the rotor by suitable gearing, counts the revolutions of the rotor in terms of the total.

motor-operated sequence switch—A multicontact switch which fixes the operating sequence (or the major devices) during starting and stopping, or during other sequential switching operations.

motor-run capacitor—A capacitor which is left in the auxiliary motor winding, and which is in parallel with the main winding, to obtain a higher power factor and efficiency.

motor-start capacitor — A capacitor which is in the circuit only during the starting period of a motor. The capacitor and its auxiliary winding are disconnected automatically by a centrifugal switch or other device when the motor reaches a predetermined speed, after which the motor runs as an induction motor.

mount—The flange or other means by which a switching tube, or a tube and cavity, are connected to a waveguide.

mount structure—The essential elements of a vacuum tube except the envelope.

mouth of a horn—The end having the larger cross section.

M-out-of-N code — A type of fixed-weight binary code in which M of the N digits are always in the same state.

movable contact—The one of a pair of contacts that is moved directly by the actuating system.

movement differential — The distance or angle from the operating position to the releasing position of a momentary contact switch.

moving-coil galvanometer—A galvanometer in which the moving element is a suspended or pivoted coil.

moving-coil meter—A meter in which a coil pivots between permanent magnets.

moving-coil microphone—Also called a dynamic microphone. A moving-conductor microphone in which the diaphragm is attached to a coil positioned in a fixed magnetic field. The sound waves strike the diaphragm moving it, and hence the coil, back and forth. An audio-frequency current is induced in the moving coil in the magnetic field and coupled to the amplifier.

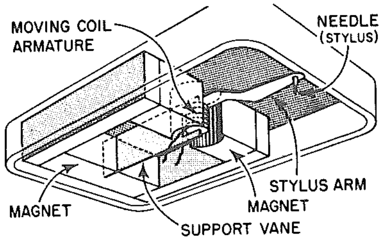

Moving-coil pickup.

moving-coil pickup—Also dynamic pickup or reproducer. A phonograph pickup in which a conductor or coil produces an electric output as it moves back and forth in a magnetic field.

moving-coil speaker—Also called a dynamic speaker. A speaker in which the moving diaphragm is attached to a coil, which is conductively connected to the source of electric energy and placed in a constant magnetic field. The current through the coil interacts with the magnetic field, causing the coil and diaphragm to move back and forth in step with the current variations through the coil.

moving-conductor microphone — A microphone which produces its electric output from the motion of a conductor in a magnetic field.

moving-conductor speaker — A speaker in which a conductor is moved back and forth in a steady magnetic field. The cone is moved by the reaction between the magnetic field and the current in the conductor.

moving element—The parts of a meter that change position as a result of a variation in the electrical quantity being measured by the meter.

moving-iron instrument—An instrument in which the current in one or more fixed coils acts on one or more pieces of soft iron or magnetically similar material, at least one of which is movable. The various forms of this instrument (plunger, vane, repulsion, attraction, repulsion-attraction) are distinguished chiefly by their mechanical construction. Otherwise, the action is the same.

moving-magnet instrument—An instrument in which a movable permanent magnet aligns itself in the field produced by another permanent magnet and an adjacent coil or coils carrying current, or by two or more current-carrying coils.

moving-magnet magnetometer — A magnetometer in which the torques act on one or more permanent magnets, which can turn in the field to be measured. Some types use auxiliary magnets (gaussian magnetometer); others use electric coils (sine or tangent galvanometer).

moving-target indicator—Abbreviated MTI. A device which limits the display of radar information primarily to moving targets.

moving-vane meter — *See* Magnetic-Vane Meter.

mr/m — Abbreviation for milliroentgens per minute.

MRN — Abbreviation for minimum reject number.

ms—Abbreviation for milliseconds.

M-scan—Also called M-display. An A-scan radar display in which the target distance is determined by moving a pedestal signal along the base line until it coincides with the horizontal position of the target-signal deflection.

msec—Abbreviation for millisecond.

M-S stereo system — *See* Mitte-Seite Stereo System.

MTBF—Abbreviation for mean-time-between-failures.

MTI—Abbreviation for moving-target indicator.

MTTF — Abbreviation for mean-time-to-failure.

MTTFF—Abbreviation for mean-time-to-first-failure.

M-type backward-wave oscillator—A cross-field injected-beam oscillator. The electrons in this device interact with an r-f wave traveling backward or opposite to the electron beam. It is efficient, broad-band, and can be voltage-tuned. It is also insensitive to load variations.

mu—Generally written as the Greek letter μ. 1. Symbol for amplification factor. 2. Symbol for permeability. 3. Abbreviation for prefix micro-. 4. Abbreviation for micron.

μ—Abbreviation for the prefix "micro-."

μa—Abbreviation for microampere.

μh—Abbreviation for microhenry.

μmho—Abbreviation for micromho.

μohm—Abbreviation for micro-ohm.

μsec—Abbreviation for microsecond.

μv—Abbreviation for microvolt.

μw—Abbreviation for microwatt.

mu-circuit—In a feedback amplifier, the circuit which amplifies the vector sum of the input signal and the feedback portion of the output signal in order to generate the output signal.

muf—Abbreviation for maximum usable frequency.

mu factor—Ratio of the changes between two electrode voltages, assuming the current and all other electrode voltages are maintained constant—i.e., it is a measure of the relative effect which the voltages on two electrodes have on the current in the circuit of a specified electrode.

Muller tube—A thermionic vacuum tube having an auxiliary cathode or grid connected internally to the main cathode through a high-value resistor.

multiaddress — Pertaining to computer instructions which specify two or more addresses.

multianode tank—*See* Multianode Tube.

multianode tube — Also called multianode tank. An electron tube having two or more main anodes and a single cathode.

multiband antenna—An antenna usable at more than one frequency band.

multicasting—Broadcasting a stereo program by using two f-m stations. Two f-m receivers are required.

multicavity magnetron — A magnetron in which the circuit has more than one cavity.

multicellular horn—A cluster of horns with juxtaposed mouths, which control the directional pattern of the radiated energy.

multichannel radio transmitter — A radio transmitter having two or more complete radio-frequency portions capable of operating on different frequencies, either individually or simultaneously.

multichannel R/C—A radio-control installation that employs tuned reeds to supply several control functions. The basic carrier frequency remains the same, but different tones make possible a number of control channels.

multichip circuit—A microcircuit in which discrete, miniature active electronic elements (transistors and/or diode chips) and thin-film or diffused passive components or component clusters are interconnected by thermocompression bonds, alloying, soldering, welding, chemical deposition, or metalization.

multichip integrated circuit—An integrated circuit the elements of which are formed on or within two or more semiconductor chips that are separately attached to a substrate. *Also see* Integrated Circuit.

multicoupler—A device for connecting several receivers to one antenna and properly matching their impedances.

multielectrode tube—An electron tube containing more than three electrodes associated with a single electron stream.

multielement parasitic array—An array of dipoles and parasitic reflectors arranged to produce a beam of desired directivity.

multiemitter transistor — A transistor that has more than one emitter. Used mainly in logic circuits.

multifrequency transmitter—A radio transmitter capable of operating on two or more selectable frequencies, one at a time, using preset adjustments of a single radio-frequency portion.

multifunction array radar — An electronic scanning radar which will perform target detection and identification, tracking, discrimination, and some interceptor missile tracking on a large number of targets simultaneously and as a single unit.

multigun tube—A cathode-ray tube having more than one electron gun. Used in color television receivers and multiple-presentation oscilloscopes.

multihop propagation — The bouncing of radio waves from the ionosphere to increase their range.

multilayer circuit—Three or more conductive patterns interspersed with layers of insulating base and laminated into sandwiches, with interconnection between layers provided by plate-through holes.

multilayer interconnection pattern — A technique used for the interconnection of arrays performing large complicated electronic functions. This technique involves the use of alternating films of insulating and conducting materials as a precondition for the realization of multiple interconnection planes.

multimeter—*See* Circuit Analyzer.

multimode operation—The operation of a laser sufficiently above threshold so as to stimulate more than one mode. If the modes are not degenerate, the output pattern will contain measurable power at more than one frequency.

multimoding—The simultaneous generation of many frequencies instead of one discrete frequency.

multipath cancellation—In effect, complete cancellation of signals because of the relative amplitude and phase differences of the components arriving over separate paths.

multipath delay—A form of phase distortion occurring most often in high-frequency layer-refracted or reflected signals, and also in vhf scattered signals. The existence of more than one signal path between transmitter and receiver causes the signal components to reach the receiver at slightly different times, causing echos or ghosting.

multipath effect—The arrival of radio waves at slightly different times because all components do not travel the same distance.

multipath reception—Reception in which the radio signal from the transmitter travels to a receiver antenna by more than one route, usually because it is reflected from obstacles. The result is seen as ghosts in a tv picture.

multipath transmission—The phenomenon where the signals reach the receiving antenna from two or more paths and usually have both amplitude and phase differences. It may cause jitter in facsimile (in Europe it is called echo).

multiple—1. A group of terminals arranged in parallel to make a circuit or group of circuits accessible at any of several points to which a connection can be made. 2. To render a circuit accessible as in (1) above by connecting it in parallel with several terminals.

multiple access—The use of a communication satellite by more than one pair of ground stations at a time. Ideally it lets many independent ground stations use an active repeater without interfering with each other.

multiple accumulating registers — In a computer, additional internal storage capacity which can contain loading, storing, adding, subtracting, and comparing factors up to four computer words in length.

multiple-address code—A computer instruction code that includes more than one address.

multiple-address instruction—In a computer, an instruction that contains more than one address.

multiple break—In an electrical circuit, an interruption at more than one point.

multiple-break contacts—A contact arrangement such that a circuit is opened in two or more places.

multiple-chip circuit—*See* Hybrid Integrated Circuit.

multiple-contact switch—A switch in which the movable contact can be set to any one of several fixed contacts.

multiple course—One of a number of lines of position defined by a navigational system. Any one of these lines may be selected as a course line.

multiple-length number—In a computer, a quantity or expression that occupies two or more registers.

multiple modulation—Also called compound modulation. A succession of modulation processes in which the modulated wave from one process becomes the modulating wave for the next.

multiple pileup—Also called multiple stack. An arrangement of contact springs composed of two or more pile-ups.

multiple programming — In computer programming, simultaneous execution of two or more arithmetical or logical operations.

multiple-purpose tester—A single test instrument having several ranges, for measuring voltage, current, and resistance.

multiple-reflection echoes—Returned echoes that have been reflected from an object in the radar beam. Such echoes give a false bearing and range.

multiple sound track—Two or more sound tracks printed side by side on the same medium, containing the same or different material, but meant to be played at the same time (e.g., those used for stereophonic recording).

multiple stack—*See* Multiple Pile-up.

multiple tube counts—Spurious counts induced in a radiation counter by previous tube counts.

multiple-tuned antenna — A low-frequency antenna with one horizontal section and several tuned vertical sections.

multiple-twin quad — Quad cable in which the four conductors are arranged in two twisted pairs, and the two pairs twisted together.

multiple-unit semiconductor device — A semiconductor device having two or more sets of electrodes associated with independent carrier systems. It is implied that the device has two or more output functions which are independently derived from separate inputs (e.g., a duotriode transistor).

multiple-unit steerable antenna — An antenna unit composed of a number of stationary antennas, the major composite lobe of which is electrically steerable.

multiple-unit tube—Also called a duodiode, duotriode, diode-pentode, duodiode-triode, duodiode-pentode and triode-pentode. An electron tube containing two or more groups of electrodes associated with independent electron streams within one envelope.

multiplex—*See* Multiplexing.

multiplex adapter—A circuit incorporated in an f-m tuner or receiver to permit two-channel, or stereophonic, reception from a station transmitting multiplex broadcasts. The other audio channel is produced by demodulating the transmitted subcarrier.

multiplex code transmission—The simultaneous transmission of two or more code messages in either or both directions over the same transmission path.

multiplexer—A device which simultaneously transmits two or more signals over a common carrier wave.

multiplexing—The process of combining several measurements for transmission over a signal wire or link. There are two widely used methods of multiplexing. Time-division multiplexing utilizes the principle of time sharing among measurement channels. The second method of multiplexing also utilizes the principle of frequency sharing among measurement channels; it is called frequency-division multiplexing, where the data from each channel are used to modulate sinusoidal signals called "subcarriers" so that the resultant signal representing each channel contains only frequencies in a restricted, narrow frequency range.

multiplex operation—Simultaneous transmission of two or more messages in either or both directions over the same transmission path.

multiplex printing telegraphy—The form of printing telegraphy that uses a line circuit to transmit one character (or one or more pulses of a character) for each of two or more independent channels.

multiplex radio transmission—The simultaneous transmission of two or more signals over a common carrier wave.

multiplex stereo—A system of broadcasting both channels of a stereo program on a single carrier. This is commonly done by modulating an ultrasonic subcarrier—either with the signals of one of the stereo channels, or by a difference signal composed of the two channels combined out of phase, or subtractively.

multiplex telegraphy—Telegraphy employing multiplex code transmission.

multiplex transmission—*See* Multiplexing.

multiple x-y recorder—A recorder that plots a number of independent charts simultaneously, each showing the relation of two variables, neither of which is time.

multiplication point—A mixing point the output of which is obtained by multiplication of its inputs.

multiplier—1. A device in which the output is the product of the signal magnitudes represented by one of the two or more input signals. 2. A series resistor placed in a voltmeter to increase the voltage range.

multiplier phototube—Also called photomultiplier tube. A phototube with one or more dynodes between its photocathode and the output electrode.

multiplier-quotient register — In a computer, a register in which is placed the multiplier for multiplication and in which is developed the quotient for division.

multiplier resistor—*See* Multiplier.

multiplying factor—The number by which the reading of a meter must be multiplied to obtain the true value.

multipoint circuit—A circuit interconnecting several locations, wherein information transmitted is available at all locations simultaneously.

multipolar—Having more than one pair of magnetic poles.

multiposition relay — A relay having more than one operate or nonoperate position, e.g., a stepping relay.

multiprocessor—A computer having multiple arithmetic and logic units that can be used simultaneously.

multirate meter—A meter which registers at different rates or on different dials at different hours of the day.

multi-r-f-channel transmitter—A radio transmitter having two or more complete radio-frequency portions that may be operated on different frequencies either individually or simultaneously.

multisegment magnetron — A magnetron with an anode divided into more than two segments, usually by parallel slots.

multispeed motor—A motor which can be operated at any one of two or more definite speeds, each practically independent of the load.

multistage tube—An X-ray tube in which the cathode rays are accelerated by multiple ring-shaped anodes, each at a progressively higher potential.

multistate noise—Noise consisting of erratic switching that is generated within a device at various sharply defined levels of applied current.

multitrack magnetic system—A magnetic-recording system in which its medium has two or more tracks.

multitrack recording system—A recording system in which the medium has two or more recording paths, which may carry the same or different material but are played at the same time.

multiturn potentiometer—A potentiometer that must be rotated more than one turn for the slider to travel the complete length of the resistive element.

multiunit tube—An electron tube containing within one glass or metal envelope, two or more groups of electrodes, each associated with separate electron streams.

multivibrator — A relaxation oscillator in which the in-phase feedback voltage is obtained from two electron tubes or transistors. Typically, their outputs are coupled through resistive-capacitive elements. The time constants of the coupling elements determine the fundamental frequency, which may be further controlled by an external voltage. When such a circuit is normally in a nonoscillating state and a trigger signal is required to start a cycle of operation, it is called a one-shot, flip-flop, or start-stop multivibrator.

multivoltage control—A method of controlling the voltage of an armature by successive impression of a number of substantially fixed voltages on the armature. The voltages are usually obtained from multicommutator generators common to a group of motors.

Mumetal—A metallic alloy with high permeability and a low hysteresis loss. It is excellent for magnetic shielding.

Mumetal shield — A cone-shaped covering made of Mumetal. Placed over the flared portion of a picture tube, it acts as a shield to prevent outside magnetic fields from affecting the alignment of the electron beams in the tube.

Munsell system—A color-specification system used principally in photography and color printing. It is based on sample cards containing the hue scale in five principal and five intermediate hues, and the brilliance scale in ten steps ranging from black to white. These represent visual, not physical, intervals.

Munsell value—In the Munsell system of object-color specification, the dimension which indicates the apparent luminous transmittance or reflectance of the object on a scale having approximately equal perceptual steps under the usual conditions of observation.

Murray loop test—A method of localizing a fault in a cable. This is done by replacing two arms of a Wheatstone bridge with a loop formed by the cable under test and a good cable connected to the far end of the defective cable.

MUSA antenna—Acronym for Multiple-Unit Steerable Antenna. It consists of several stationary antennas, the composite major lobe of which is electrically steerable.

mush winding—A type of winding in an a-c machine. The conductors are placed one by one in prepared slots and the end connections are separately insulated.

musical cushion—A musical selection added at the end of a program which is running short or over. By playing it slowly or cutting out portions, the program director can make the program come out on time.

musical echo—*See* Flutter Echo.

musical quality—*See* Timbre.

music power—The short-term power available from an amplifier for the reproduction of program material. The music power output exceeds the rms power rating to a greater or lesser extent. Its measurement is standardized by the Institute of High Fidelity (IHF) and represents a practical means of stating the actual capabilities of an amplifier for the reproduction of program material.

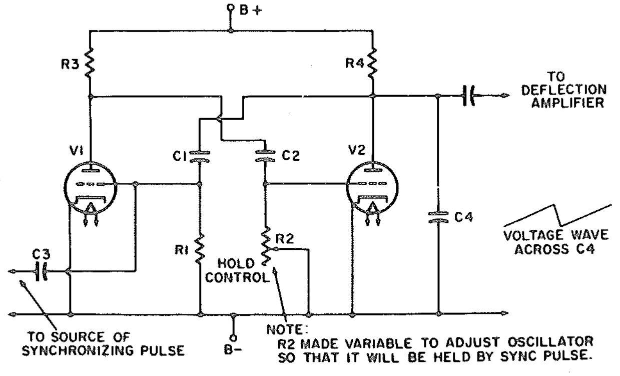
B+
R3
R4
TO DEFLECTION AMPLIFIER
V1
C1
C2
V2
C4
VOLTAGE WAVE ACROSS C4
C3
R1
R2
HOLD CONTROL
TO SOURCE OF SYNCHRONIZING PULSE
B-
NOTE:
R2 MADE VARIABLE TO ADJUST OSCILLATOR SO THAT IT WILL BE HELD BY SYNC PULSE.

Multivibrator (asymmetrical).

must operate valve—A specified functioning value at which all relays meeting the specification must release.

must-release value—The specified operating value at which all relays that meet a certain specification must release.

muting—Muffling or deadening a sound.

muting circuit—1. A circuit which cuts off the receiver output when the r-f carrier reaching the first detector is at or below a predetermined intensity. 2. A circuit for making a receiver insensitive while its associated transmitter is on.

muting switch—A switch used with automatic tuning systems to silence the receiver while it is being tuned.

mutual capacitance — Capacitance between two conductors with all other conductors connected together and to a grounded shield. (Does not apply to a shielded single conductor.)

mutual conductance—*See* Transconductance.

mutual-conductance meter—*See* Transconductance Meter.

mutual-conductance tube tester—*See* Transconductance Tube Tester.

mutual impedance—Between any two pairs of network terminals (all other terminals being open), the ratio of the open-circuit potential at either one of the two pairs, to the current applied to the other pair of network terminals.

mutual inductance—The property that exists between two current-carrying conductors when the magnetic lines of force from one link with those of the other. It determines, for a given rate of change of current in one circuit, the electromotive force induced in the other.

mutual induction — The production of a voltage in one circuit by a changing current in a neighboring circuit, even though no apparent connection exists between the two circuits.

mutual inductor—An inductor for changing the mutual inductance between two circuits.

mutual information—*See* Transinformation.

mutual interference — Man-made interference from two or more electrical or electronic systems which affects these systems on a reciprocal basis.

mv—Abbreviation for millivolt.

Mv—Abbreviation for megavolt.

Mva—Abbreviation for megavolt-ampere.

Mw—Abbreviation for megawatt.

Mwh—Abbreviation for megawatt-hour.

Mycalex—Trade name of the Mycalex Corp. for mica bonded with glass. It has a low power factor at high frequencies, and is a good insulator at all frequencies.

Mylar — Trade name of E. I. duPont de Nemours and Co., Inc., for a highly durable, transparent plastic film of outstanding strength. It is used as a base for magnetic tape and as a dielectric in capacitors.

Mylar capacitor—A capacitor in which Mylar film, either alone or in combination with paper, is the dielectric.

myoelectric signals—Complex pulse potentials of 10 to 1,000 microvolts with durations between 1 and 10 milliseconds. As generated by a muscular effort, recorded from the body surface.

myograph—A recorder of forces of muscular contraction. *Also see* Electromyograph.

myokinesimeter—An apparatus for measuring the response of a muscle to stimulation by electric current.

myophone—An instrument for making audible the sound of a muscle when it is contracting.

myriametric waves — Very-low frequency band; 3 kHz to 30 kHz.

N

n—Abbreviation for the prefix "nano" (10^{-9}).

N—Symbol for number of turns, or the north-seeking pole of a magnet.

na—Abbreviation for nanoampere.

NAB—Abbreviation for National Association of Broadcasters.

NAB curve—The standard playback equalization curve set by the National Association of Broadcasters.

nail-head bond—An alternate term for ball bond.

nano—Abbreviated n. Prefix meaning one-billionth (10^{-9}).

nanoampere—One-thousandth of a microampere (10^{-9} ampere). Abbreviated na.

nanocircuit — An integrated microelectronic circuit in which each component is fabricated on a separate chip or substrate so that independent optimization for performance can be achieved.

nanofarad—One-billionth of a farad, equal to 10^{-9} farad, 0.001 microfarad, and 1000 picofarads. Abbreviated nf.

nanohenry—One-thousandth of a microhenry, equal to 10^{-9} henry and 1000 picohenrys. Abbreviated nh.

nanosecond—One-billionth of a second (10^{-9} second). Light travels approximately 1 foot in 1 nanosecond. Abbreviated nsec.

nanovolt — One-thousandth of a microvolt (10^{-9} volt). Abbreviated nv.

nanovoltmeter—A voltmeter sufficiently sensitive to give readings in thousandths of microvolts.

Naperian logarithm—Also called hyperbolic or natural logarithm. A logarithm to the base 2.71828.

napier—*See* Neper.

narrow band—A band whose width is greater than 1% of the center frequency and less than one-third octave.

narrow-band amplifier — An amplifier designed for optimum operation over a narrow band of frequencies.

narrow-band axis—In phasor representation of the chrominance signal, the direction of the coarse-chrominance primary of a color tv system.

narrow-band fm—Abbreviation for narrow-band frequency modulation.

narrow-band f-m adapter—An attachment which converts an a-m communications receiver to f-m.

narrow-band frequency modulation — Abbreviated NBFM. Frequency modulation which occupies only a small portion of the conventional f-m bandwidth. Used mainly for two-way voice communication by police, fire, taxicabs, and amateurs. The FCC limits the deviation to 3 kHz for amateurs and 15 kHz for all others.

narrow-band interference—Sharply tunable interference, having a spectrum that is small compared with the bandwidth of the measuring instrument. (Sine-wave carriers — both modulated and unmodulated—are good examples.)

narrow-sector recorder — A radio direction finder with which atmospherics are received from a limited sector related to the position of the antenna. The antenna is usually rotated continuously, and the bearings of the atmospherics recorded automatically.

NARTB—Abbreviation for National Association of Radio and Television Broadcasters.

n-ary code—A code in which each element can be any one of *n* distinct kinds or values.

n-ary pulse-code modulation — A type of pulse-code modulation in which the code for each element of information can consist of any one of n distinct kinds or values.

NASA — Acronym for National Aeronautics and Space Administration. The federal agency charged with all scientific space missions.

National Association of Broadcasters—Abbreviated NAB. An association of radio and television broadcasters.

National Association of Radio and Television Broadcasters—Abbreviated NARTB. A name used for a number of years by an association of broadcasters. In 1958 the name was changed back to National Association of Broadcasters, an earlier title.

National Electrical Manufacturers Association—Abbreviated NEMA. An organization of manufacturers of electrical products.

National Electric Code — Regulations governing construction and installation of electrical wiring and apparatus in the United States. They were established by the American National Board of Fire Underwriters for safety purposes.

National Television System Committee—Abbreviated NTSC. A committee organized in 1940 and comprising all United States companies and organizations interested in television. Between 1940 and 1941, it formulated the black and white television standards and between 1950 and 1953, the color television standards that were approved by the Federal Communications Commission.

natural frequency — 1. The frequency at which a body will oscillate if disturbed from its equilibrium position. 2. The lowest resonant frequency of a circuit or component without adding inductance or capacitance.

natural frequency of an antenna—The lowest resonant frequency of an antenna with no added inductance or capacitance.

natural interference—A type of electromagnetic interference caused by atmosphere and galactic electrical disturbances and precipitation static.

natural logarithm—*See* Naperian Logarithm.

natural magnet—Magnetic ore (e.g., a lodestone) which exhibits the property of magnetism in its natural state.

natural period—The period of the free oscillation of a body or system. When the period varies with amplitude, the natural period is the period when the amplitude approaches zero.

natural radiation — *See* Background Radiation.

natural resonance—*See* Periodic Resonance.

natural wavelength—The wavelength corresponding to the natural frequency of an antenna or circuit.

Navaglobe—A long-distance navigational system of the continuous-wave, low-frequency type. Bearing information is provided by amplitude comparison.

navar—A coordinated series of radar air-navigation and traffic-control aids utilizing transmissions at wavelengths of 10 and 60 centimeters. In an aircraft, it provides distance and bearing from a given point, display of other aircraft in the vicinity, and commands from the ground. On the ground, it provides a display of all aircraft in the vicinity, as well as their altitudes and identities, plus means for transmitting certain commands.

navarho — A continuous-wave, low-frequency navigation system that provides simultaneous bearing and range information over long distances.

navigation—The process of directing an airplane or ship toward its destination by determining its position, direction, etc.

navigational parameter—In a navigational aid, a visual or aural output having a specific relationship to navigational coordinates.

navigation beacon—A light, radio, or radar beacon that provides navigational aid to ships and aircraft.

NBFM—Abbreviation for narrow-band frequency modulation.

NC—Abbreviation for no connection (in a vacuum-tube base) or normally closed relays or other contacts.

N-conductor concentric cable—A cable composed of an insulated central conducting core with (N-1) tubular stranded conductors laid over it concentrically and separated by layers of insulation.

n-cube—Also called n-dimensional cube or n-variable cube. In switching theory, a term used to indicate two n-1 cubes having corresponding points connected.

n-dimensional cube—*See* n-cube.

N-display—Also called N-scan, or N-scope. A radar display similar to the K-display. The target appears as a pair of vertical deflections (blips) from the horizontal time base. Direction is indicated by the related amplitude of the vertical deflections. Target distance is determined by moving a pedestal signal (the control of which is calibrated in distance) along the base line until it coincides with the horizontal position of the vertical deflections.

near-end cross talk—In a disturbed telephone channel, cross talk propagated in the opposite direction from the current in the disturbing channel. The terminal where the near-end cross talk is present is ordinarily near, or coincides with, the energized terminal of the disturbing channel.

near field—1. The acoustic radiation field close to the speaker or some other acoustic source. 2. The electromagnetic field within a distance of 1 wavelength from a transmitting antenna.

near infrared—Name applied to the spectral region primarily comprising wavelengths between 3 and 30 microns.

necessary bandwidth—For a given class of emission, the minimum value of the occupied bandwidth sufficient to ensure the transmission of information at the rate and the quality required for the system employed, under specified conditions. Emissions useful for the good functioning of the receiving equipment—for example, the emission corresponding to the carrier of reduced carrier systems—shall be included in the necessary bandwidth.

neck—The narrow tubular part of the envelope of a cathode-ray tube; it extends from the funnel to the base and houses the electron gun.

needle—1. A probe used on stacks of punched cards. 2. *See* Stylus.

needle chatter—*See* Needle Talk.

needle drag—*See* Stylus Drag.

needle gap—A spark gap having needle-point electrodes.

needle pressure—*See* Stylus Force.

needle scratch—*See* Surface Noise.

needle talk—Also called needle chatter. Sounds produced directly by the needle of a phonograph pickup.

needle test point—Sharp steel probes connected to test cords for making contacts with conductors.

neg—Abbreviation for negative.

negative—1. Abbreviated neg. Less than zero. 2. The opposite of positive (e.g., negative resistance, transmission, feedback, etc.). 3. A terminal or electrode having an excess of electrons. Electrons flow out of the negative terminal of a voltage source, toward a positive source.

negative acceleration—A relative term often used to indicate that, referenced to zero acceleration, a negative electrical output will be obtained from a linear accelerometer when its sensitive axis is oriented normal to the surface of the earth, and a given minus reference point along the sensitive axis is located upward from the center of the seismic mass.

negative bias—In a vacuum tube, the voltage which makes the control grid more negative than the cathode.

negative booster—A booster used with a ground-return system to reduce the difference of potential between two points to the grounded return. It is connected in series with a supplementary insulated feeder extending from the negative bus of the generating station or substation to a distant point on the grounded return.

negative charge—A condition in a circuit when the element in question retains more than its normal quantity of electrons.

negative conductor—A conductor connected to the negative terminal of a source of supply. Such a conductor is frequently used as an auxiliary return circuit in a system of electric traction.

negative electricity—A body that contains an excess of electrons (e.g., the electricity that predominates in a comb which has been run through one's hair).

negative electrode—The electrode from which the forward current flows.

negative electron—Also called negatron. An electron, distinguished from a positive electron or positron.

negative feedback—Also called degeneration, inverse feedback, or stabilized feedback. A process by which a part of the output signal of an amplifying circuit is fed back to the input circuit. The signal fed back is 180° out-of-phase with the input signal; therefore, the amplification is decreased, and distortion reduced.

negative-feedback amplifier—An amplifier in which negative feedback is employed to improve the stability or frequency response, or both.

negative ghost—A ghost which has the opposite shading from that of the original image (i.e., is black when the image is white, and vice versa).

negative glow—The luminous glow between the cathode and Faraday dark spaces in a glow-discharge, cold-cathode tube. (*Also see* Glow Discharge.)

negative-grid generator—A conventional oscillator circuit in which oscillation is produced by feedback from the plate circuit to a grid which is normally negative with respect to the cathode and which is designed to operate without drawing grid current at any time.

negative image—A televised picture in which the whites appear black and vice versa. It is due to the picture signal having a polarity opposite to that of a normal signal.

negative impedance—1. Also called negative resistance when there is no inductance or capacitance in the circuit. A characteristic of certain electrical devices or circuits—instead of increasing, the voltage decreases when the current is increased and vice versa. 2. In general, the driving-point impedance of a circuit in which a current increase produces a voltage decrease, and vice versa. For a negative admittance, a voltage increase produces a current decrease, and vice versa.

negative ion—An atom with more electrons than normal. Thus, it has a negative charge.

negative light modulation—In television, the process whereby a decrease in initial light intensity causes an increase in the transmitted power.

negative modulation—In an a-m television system, that modulation in which an increase in brightness corresponds to a decrease in transmitted power.

negative modulation factor—The maximum negative departure of the envelope of an a-m wave from its average value, expressed as a ratio. This rating is used whenever the modulation signal wave has unequal positive and negative peaks.

negative phase-sequence relay—A relay which functions in conformance with the negative phase-sequence component of the current, voltage, or power of the circuit.

negative picture phase—For a television signal, the condition in which an increase in brilliance makes the picture-signal voltage swing in a negative direction from the zero level.

negative plate—The grid and active material connected to the negative terminal of a storage battery. When the battery is discharging, electrons flow from this terminal, through the external circuit, to the positive terminal.

negative resistance—A resistance which exhibits the opposite characteristic from normal—i.e., when the voltage is increased across it, the current will decrease instead of increase.

negative-resistance magnetron—A magnetron operated so that it acts as a negative resistance.

negative-resistance oscillator—An oscillator produced by connecting a parallel-tuned resonant circuit to a two-terminal negative-resistance device (one in which an increase in voltage results in a decrease in current). Dynatron and transitron oscillators are examples.

negative-resistance region—That operating region in which an increase in applied voltage results in a current decrease.

negative-resistance repeater—A repeater in which gain is provided by a series or shunt negative resistance, or both.

negative temperature coefficient—The amount of reduction in the value of a quantity, such as capacitance or resistance, for each degree of increase in temperature. (*Also see* Temperature Coefficient.)

negative terminal—In a battery or other voltage source, the terminal having an excess of electrons. Electrons flow from it, through the external circuit, to the positive terminal.

negative-transconductance oscillator—An electron-tube oscillator the output of which is coupled back to the input without phase shift, the phase condition for oscillation being satisfied by the negative transconductance of the tube. A transitron oscillator is an example.

negative transmission—The modulation of the picture carrier by a picture signal with a polarity such that the sync pulses occur in the blacker-than-black region and a decrease in initial light intensity increases the transmitted power.

negatron—1. An electron. 2. A four-electrode vacuum tube having a negative-resistance characteristic. *Also see* Negative Electron.

NEMA—Abbreviation for National Electrical Manufacturers Association.

NEMA standards—Specifications adopted as standard by the National Electric Manufacturers Association.

nemo—*See* Field Pickup.

neon—An inert gas used in neon signs and in some electron tubes. It produces a bright red glow when ionized. Symbol, *Ne;* atomic number, 10; atomic weight, 20.183.

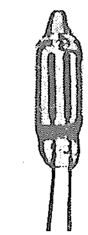

Neon bulb.

neon bulb—A glass envelope filled with neon gas and containing two or more insulated electrodes. The tube will not conduct until the potential difference between two electrodes reaches the firing, or ionization, potential, and will remain conductive until the voltage is reduced to the extinction level.

neon-bulb oscillator—A simple relaxation oscillator in which a capacitor charges to the ionization potential of the neon gas within a bulb. Ionization rapidly depletes the charge on the capacitor, and a new charge cycle begins.

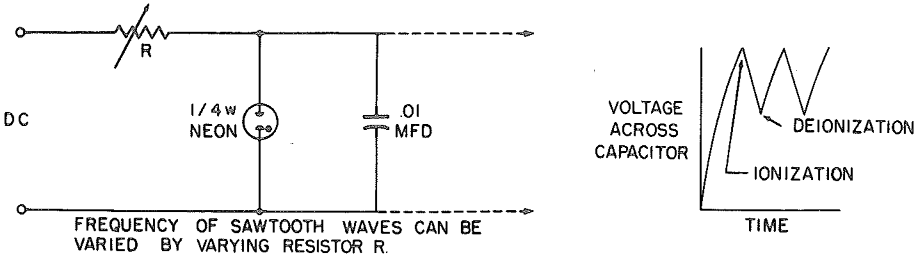

Neon-bulb oscillator.

neper—Also called napier. The fundamental division of a logarithmic scale for expressing the ratio between two currents, powers, or voltages. The number of nepers denoting such a ratio is the natural (Naperian) logarithm of the square root of this ratio. One neper equals 0.8686 bels, or 8.686 decibels. Expressed as a formula:

$$N = \log_{\epsilon} \sqrt{\frac{E_1}{E_2}}$$

where,

N is the number of nepers denoting their ratio,

E_1 and E_2 are the two voltages.

Nernst effect—The effect whereby a potential difference is developed across a heated metal strip placed perpendicular to a magnetic field.

Nernst-Ettinghausen effect — A thermomagnetic effect that occurs in certain pure crystals, where a temperature difference is produced in a direction perpendicular to an applied magnetic field and a longitudinal electric current.

Nernst lamp—An electric lamp consisting of a short, slender rod of zirconium oxide that is heated to incandescence by a current.

nesting—In a computer, the inclusion of a routine or block of data within a larger routine or block of data.

net—Organization of stations capable of direct communications on a common channel, often on a definite schedule.

net authentification—Identification used on a communications network to establish the authenticity of several stations.

net information content—A message of the essential information contained in a message. It is expressed as the minimum number of bits or hartleys required to transmit the message with specified accuracy over a noiseless medium.

net loss—The algebraic sum of the gains and losses between two terminals of a circuit. It is equal to the difference in the levels at these points.

net reactance—The difference between the capacitive and inductive reactance in an a-c circuit.

network—1. A combination of electrical elements. 2. An interconnected system of transmission lines that provides multiple connections between loads and sources of generation.

network analysis—The obtaining of the electrical properties of a network (e.g., its input and transfer impedances, responses, etc.) from its configuration, parameters, and driving forces.

network analyzer—1. A group of electric-circuit elements which can readily be connected to form models of electric networks. From corresponding measurements on the model, it is then possible to infer the electrical quantities at various points on the prototype system. 2. Also called network calculator. An analog device designed primarily for simulating electrical networks.

network calculator — An analog device designed primarily for simulating electric networks.

network constant—Any one of the resistance, inductance, mutual-inductance, or capacitance values in a circuit or network. When these values are constant, the network is said to be linear.

network filter—A transducer for separating waves on the basis of their frequency.

network master relay—A relay that closes and trips an alternating-current, low-voltage network protector.

network phasing relay—A relay which functions in conjunction with a master relay to limit closure of the network protector to a predetermined relationship between the voltage and the network voltage.

network relay—A form of relay (e.g., voltage, power, etc.) used in the protection and control of alternating-current, low-voltage networks.

network synthesis—The obtaining of a network from prescribed electrical properties such as input and transfer impedances, specified responses for a given driving force, etc.

network transfer function — A frequency-dependent function, the value of which is the ratio of the output to the input voltage.

network transformer — A transformer suitable for use in a vault to feed a variable-capacity system of interconnected secondaries.

Neuman's law—Mutual inductance is a constant for a given relative physical position of coils, and independent of the fact that the current flows in one or the other coil, and of frequency, current, and phase.

neuristor—A two-terminal active device with some of the properties of neurons (e.g., propagation that suffers no attenuation and has a uniform velocity, and a refractory period).

neuroelectricity—The minute electric voltage generated by the nervous system.

neuron — The basic building block of the human nervous system. A specialized body cell that can conduct and code an electrical pulse.

neutral — In a normal condition; hence, neither positive nor negative. A neutral object has its normal number of electrons—i.e., the same number of electrons as protons.

neutral circuit—A teletypewriter circuit in which current flows in only one direction. The circuit is closed during the marking condition and open during the spacing condition.

neutral conductor — That conductor of a polyphase circuit or of a single-phase three-wire circuit which is intended to have a potential such that the potential differences between it and each of the other conductors are approximately equal in magnitude and are also equally spaced in phase.

neutral ground—A ground connection to the neutral point or points of a circuit, transformer, rotating machine, or system.

neutralization — The nullifying of voltage feedback from the output to the input of an amplifier through the interelectrode impedance of the tube. Its principal use is in preventing oscillation in an amplifier. This is done by introducing, into the input, a voltage equal in magnitude but opposite in phase to the feedback through the interelectrode capacitance.

neutralize—In an amplifier stage, to balance out the feedback voltage due to grid-plate capacitance, thus preventing regeneration.

neutralizing capacitor—A capacitor, usually variable, employed in a radio receiving or transmitting circuit to feed a portion of the signal voltage from the plate circuit back to the grid circuit.

neutralizing circuit—In an amplifier circuit, the portion that provides an intentional feedback path from plate to grid. This is done to prevent regeneration.

neutralizing indicator—An auxiliary device (e.g., a lamp or detector coupled to the plate tank circuit of an amplifier) for indicating the degree of neutralization in an amplifier.

neutralizing tool—Also called a tuning wand. A small screwdriver or socket wrench, partly or entirely nonmetallic, for making neutralizing or aligning adjustments in electronic equipment. (*Also see* Alignment Tool.)

neutralizing voltage—The a-c voltage fed from the grid circuit to the plate circuit (or vice versa). It is deliberately made 180° out of phase with and equal in amplitude to the a-c voltage similarly transferred through undesired paths (usually the grid-to-plate interelectrode capacitance).

neutral operation — The system whereby marking signals are formed by current impulses of one polarity, either positive or negative, and spacing signals are formed by reducing the current to zero or nearly zero.

neutral point—The point which has the same potential as the point of junction of a group of equal nonreactive resistances connected at their free ends to the appropriate main terminals or lines of the system.

neutral relay—Also called a nonpolarized relay. A relay in which the armature movement does not depend on the direction of the current in the controlling circuit.

neutrino—An atomic particle with essentially no mass and no charge postulated to explain the conservation of energy and momentum when a radioactive atom emits an electron.

neutrodyne — An amplifier circuit used in early tuned radio-frequency receivers. It was neutralized by the voltage fed back by a capacitor.

neutron—One of the three elementary particles (the electron and proton are the other two) of an atom. It has approximately the same mass as the hydrogen atom, but no electric charge. It is one of the constituents of the nucleus (the proton is the other one).

neutron flux—A term used to express the intensity of neutron radiation, usually in connection with the operation of a reactor.

newton—In the mksa system, the unit of force that will impart an acceleration of 1 meter per second per second to a mass of 1 kilogram.

nf—Abbreviation for nanofarad.

Nichrome—Trade name of Driver-Harris Co. for an alloy of nickel and chromium used extensively in wirewound resistors and heating elements.

nickel-cadmium cell—The most widely used rechargeable sealed cell, nickel-cadmium gives a flat voltage discharge characteristic,

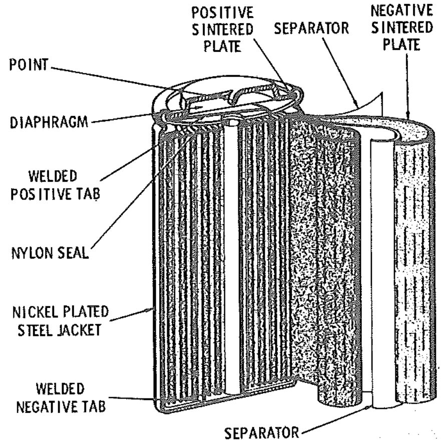

Nickel-cadmium cell.

nominal cell voltage of 1.25 volts, and good low-temperature operation. A cell with a nickel and oxide positive electrode and a cadmium negative electrode. The plates are wrapped with a separator between them, and are immersed in a potassium-hydroxide electrolyte.

nickel-oxide film diode—A solid-state diode made of nickel-oxide film. It may be switched from off to on by applying a 300-volt low-current pulse for 10 μsec, and may be switched from on to off by applying a 30-volt high-current pulse for 10 μsec. In effect it works like a diode flip-flop.

nickel silver—*See* German Silver.

nif—Abbreviation for noise-improvement factor.

NI junction—A semiconductor junction between N-type and intrinsic materials.

Nipkow disc—A round plate used for scanning small elementary areas of an image in correct sequence for a mechanical television system. It has one or more spirals of holes around its outer edge, and successive openings are positioned so that rotation of the disc provides the scanning.

NIPO—Acronym for Negative Input, Positive Output.

nit—In a computer, a choice among events that are equally probable. One nit equals 1.44 bits.

Nixie tube—A glow tube which converts a combination of electrical impulses into a visual display. (*Also see* Numerical-Readout Tube.)

n-level logic — Pertaining to a collection of gates connected in such a way that not more than n gates appear in series. (Computer term).

N-N junction—In an N-type semiconducting material, a region of transition between two regions having different properties.

no-address instruction—An instruction specifying an operation which the computer can perform without having to refer to its storage or memory unit.

noble gas—Also called inert gas or rare gas. One of the chemically inert gases, including helium, neon, argon, krypton, and xenon.

noctovision—A television system employing invisible rays (usually infrared) for scanning purposes at the transmitter. Hence, no visible light is necessary.

nodal point—1. *See* Node. 2. Two points on the axis of a lens such that a ray entering the lens in the direction of one, leaves as if from the other and parallel to the original direction.

nodal-point keying—Keying an arc transmitter at a point that is essentially at ground potential in the antenna circuit.

node—1. Also called junction point, branch point, or vertex. A terminal of any branch of a network, or a terminal common to two or more branches. 2. Also called nodal point. In a standing wave, a point, line, or surface where some characteristic of the wave field has essentially zero amplitude. 3. The point, line, or surface in a standing-wave system where some characteristic of the wave field has essentially zero amplitude. The appropriate modifier should be used with the word "node" to signify the type that is intended (pressure node, velocity node, etc.).

nodules—Clusters of oxide particles which protrude above the surface of magnetic tape.

noise—1. Any unwanted disturbance within a dynamic electrical or mechanical system (e.g., undesired electromagnetic radiation in a transmission channel or device). 2. Any unwanted electrical disturbance or spurious signal which modifies the transmitting, indicating, or recording of desired data. 3. In a computer, extra bits or words which have no meaning and must be ignored or removed from the data at the time it is used. 4. Random electrical variations generated internally in electronic components.

noise analysis—Determination of the frequency components that make up the noise being studied.

noise analyzer—An instrument used for determining the amplitude versus frequency characteristics of noise.

noise bandwidth—The width of the equivalent rectangular power gain versus frequency response of a network. It is the width of the rectangle the area of which equals that of the actual power gain versus frequency response and the height of which equals the maximum available power gain at a reference frequency f_0. It does not include any spurious responses. Any responses other than those in the useful channel are accounted for by a degraded noise figure.

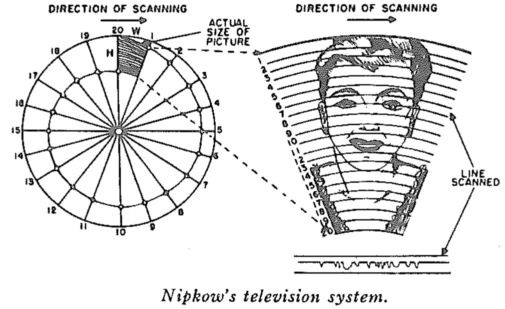

Nipkow's television system.

noise behind the signal — *See* Modulation Noise.

noise-canceling microphone — *See* Close-Talking Microphone.

noise clipper—A circuit that automatically or manually clips the noise from the output of a receiver.

noise-current generator—A current generator in which the output is a random function of time.

noise diode — A standard electrical-noise source consisting of a diode operated at saturation. The noise is due to random emission of electrons.

noise equivalent bandwidth — The useful bandwidth of a thermistor bolometer for various frequencies in the input radiation is ¼t (where t is the time constant of the bolometer).

noise equivalent power—The value of radiation which produces, in a detector, an rms signal-to-noise ratio of unity. It is usually measured with a black-body radiation source at 500 K and a bandwidth of 1 or 5 hertz. The modulation rate varies with the type of detector and is usually between 10 and 1,000 hertz.

noise factor—Also called noise figure. 1. For a given bandwidth, the ratio of total noise at the output, to the noise at the input. 2. A number expressing the amount by which a receiver falls short of equaling the theoretical optimum performance.

noise figure—*See* Noise Factor.

noise figure (of a network with a generator connected to its input terminal)—The ratio of the available signal-to-noise-power ratio at the signal-generator terminals (weighted by the network bandwidth) to the available signal-to-noise-power ratio at the output.

noise filter—A combination of electrical components which prevent extraneous signals from passing into or through an electronic circuit.

noise-improvement factor—In a receiver, the ratio of output signal-to-noise ratio to the input signal-to-noise ratio. (The term receiver is used in the broad sense and is taken to include pulse demodulators.) Abbreviated nif.

noise killer—An electric network inserted into a telegraph circuit (usually at the sending end) to reduce interference with other communication circuits.

noise level—1. The strength of extraneous audible sounds at a given location. 2. The strength of extraneous signals in a circuit. (Noise level is referred to a specified base and usually measured in decibels.)

noise limiter—A vacuum-tube circuit that cuts off all noise peaks stronger than the highest peak in the received signal. In this way, the effects of strong atmospheric or man-made interference are reduced.

noise-measuring set — *See* Circuit Noise Meter.

noise pulse—Also called noise spike. A spurious signal of short duration and of a magnitude considerably in excess of the average peak value of the ordinary system noise.

noise quieting — The ability, usually expressed in decibels, of a receiver to reduce background noise in the presence of a desired signal.

noise ratio—The ratio of the available noise power at the output of a circuit divided by the noise power at the input.

noise-reducing antenna system — A receiving-antenna system so designed that only the antenna proper can pick up signals. It is placed high enough to be out of the noise-interference zone, and is connected to the receiver with a shielded cable or twisted transmission line that is incapable of picking up signals.

noise source—A device employed to generate random noise (e.g., photomultiplier and gaseous-discharge tubes).

noise spike—*See* Noise Pulse.

noise suppression—1. The ability of a radio receiver to materially reduce the noise output when no carrier is being received. 2. A means of reducing surface noise during reproduction of a phonograph record.

noise suppressor—1. A circuit which reduces high-frequency hiss or noise. It is utilized primarily with old phonograph records. 2. In a receiver circuit, the portion which reduces noise automatically when no carrier is being received.

noise temperature—At a pair of terminals and at a specific frequency, the temperature of a passive system having an available noise power per unit bandwidth equal to that of the actual terminals.

noise-voltage generator—A voltage generator the output of which is a random function of time.

noisy mode—In a computer, a floating-point arithmetic procedure associated with normalization. In this procedure digits other than zero are introduced in the low-order positions during the left shift.

no-load current—*See* Exciting Current of a Transformer.

no-load losses—The losses in a transformer when it is excited at the rated voltage and frequency, but is not supplying a load.

nominal band—In a facsimile-signal wave, the frequency band equal to the width between the zero and the maximum modulating frequency.

nominal bandwidth—1. The difference between the nominal upper and lower cutoff frequencies of a filter. It may be expressed in octaves, in hertz, or as a percentage of the passband center frequency. 2. The maximum band of frequencies, inclusive of guard bands, assigned to a channel.

nominal horsepower—The rated power of a motor, engine, etc.

nominal impedance—The impedance of a circuit under normal conditions. Usually it is specified at the center or the operating-frequency range.

nominal line pitch—The average separation between the centers of adjacent lines in a raster.

nominal line width—1. In television, the reciprocal of the number of lines per unit length in the direction of line progression. 2. In facsimile transmission, the average separation between centers of adjacent scanning or recording lines.

nominal power rating of a resistor—The power which a resistor can dissipate continuously at a specified ambient temperature and for a stipulated length of time without excessive resistance drift.

nominal value—The stated or specified value as opposed to the actual value.

nominal voltage — The voltage of a fully charged storage cell when delivering rated current.

nomograph—*See* Alignment Chart.

nonbridging—A contact transfer in which the movable contact leaves one contact before touching the next.

nonbridging contacts — A contact arrangement in which the opening of one set of contacts occurs before the closing of another set.

noncombustible—*See* Nonflammable.

nonconductor—An insulating material—i.e., one through which no electric current can flow.

noncorrosive flux—A flux that does not contain acid and other substances which might corrode the surfaces being soldered.

nondestructive readout—In a computer, the copying of information from a storage device without altering the physical representation of the information in the device.

nondeviated absorption—1. Absorption that occurs without any appreciable slowing up of waves. 2. Normal sky-wave absorption.

nondirectional—*See* Omnidirectional.

nondirectional antenna — *See* Omnidirectional Antenna.

nondirectional microphone—*See* Omnidirectional Microphone.

nondissipative stub—A lossless length of waveguide or transmission line coupled into the sides of a waveguide.

nonequivalence element — A logic element having an action that represents the Boolean connective exclusive OR.

nonerasable storage—In a computer, a storage medium which cannot be erased and reused, e.g., punched cards or perforated paper tape.

nonferrous—Not made of or containing iron.

nonflammable — Also called noncombustible. Term applied to material which will not burn when exposed to flame or elevated temperatures, e.g., asbestos, ceramics, and structural metals.

nonhoming tuning system—A motor-driven automatic-tuning system in which the motor starts in the direction of the previous rotation. If this direction is incorrect for the new station, the motor reverses, after turning to the end of the dial, then proceeds to the desired station.

noninductive—Having practically no inductance.

noninductive capacitor — A capacitor in which the inductive effects at high frequencies are reduced to a minimum. Foil layers are offset during winding so that an entire layer of foil (projecting at either end) is connected together for contact-making purposes. A current then flows laterally rather than spirally around the capacitor, and the inductive effect is minimized.

noninductive circuit—A circuit in which the inductance is reduced to a minimum or negligible value.

noninductive load—A load that has no inductance. It may consist entirely of resistance or capacitance.

noninductive resistor—A wirewound resistor with little or no self-inductance.

noninductive winding—A winding in which the magnetic fields produced by its two parts cancel each other and thereby provide a noninductive resistance.

noninverting connection — The closed-loop connection of an operational amplifier when the forward gain is positive for a d-c signal (0° phase shift).

noninverting parametric device — A parametric device the operation of which depends essentially upon three frequencies, a harmonic of the pump frequency and two signal frequencies, of which one is the sum of the other plus the pump harmonics. Such a device can never provide gain at either of the signal frequencies. It is said to be noninverting because, if either of the two signals is moved upward in frequency, the other will move upward in frequency.

nonionizing radiation—Radiation which does not produce ionization (e.g., infrared, ultraviolet, and visible light).

nonlinear—Having an output that does not rise or fall in direct proportion to the input.

nonlinear capacitor — A capacitor that has a nonlinear mean-charge characteristic or peak-charge characteristic or a reversible capacitance that varies with bias voltage.

nonlinear coil—A coil with an easily saturable core. Its impedance is high at low or zero current, and is low when enough current flows to saturate the core.

nonlinear distortion—1. Distortion that occurs when the output does not rise and fall directly in proportion to the input. The input and output values need not be of the same quantity—e.g., in a linear detector, these values are the signal voltage at the output and the modulation envelope at the input. 2. Distortion which occurs because the transmission properties of a system are dependent upon the instantaneous magnitude of the transmitted signal. NOTE: Nonlinearity distortion gives rise to amplitude/amplitude and harmonic distortion, intermodulation, and flutter.

nonlinear network—A network (circuit) not specifiable by linear differential equations with time as the independent variable.

nonloaded Q—Also called basic Q. The value of the Q of an electric impedance without external coupling or connection.

nonmagnetic—Material which is not attracted by a magnet and cannot be magnetized (e.g., paper, plastic, tin, glass). In a strict sense, having a permeability equal to that of air or 1.

nonmagnetic armature shim — Also called nonmagnetic armature stop. A nonmagnetic member separating the pole faces of the core and armature in the operated position, used to reduce and stabilize the pull from residual magnetism in release.

nonmagnetic steel—A steel alloy that contains about 12% manganese, and sometimes a small quantity of nickel. It is practically nonmagnetic at ordinary temperatures.

nonmetallic sheathed cable—Two or more rubber- or plastic-covered conductors assembled in an outer sheath of nonconducting fibrous material that has been treated to make it flame- and moisture-resistant.

nonphysical primary—A color primary represented by a point outside the area of the chromaticity diagram enclosed by the spectrum locus and the purple boundary.

nonplanar network—A network that cannot be drawn on a plane without crossing branches.

nonpolar crystals—Crystals having the property that each lattice point is identical.

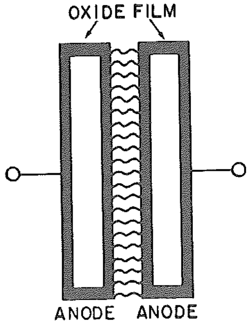

Nonpolarized electrolytic capacitor.

nonpolarized electrolytic capacitor — An electrolytic capacitor which can be connected without regard to polarity. This is possible because the dielectric film is formed on both electrodes.

nonpolarized relay—*See* Neutral Relay.

nonrenewable fuse unit—A fuse unit that cannot be readily restored for service after operation.

nonreset timer—A timer which cannot be reset by electrical means.

nonresonant line—A transmission line with a natural resonant frequency different from that of the transmitted signal.

nonreturn-to-zero—A method of writing information on a magnetic surface in which the current through the write-head winding does not return to zero after the write pulse.

nonsaturated color—A color that is not pure—i.e., one that has been mixed with white or its complementary color.

nonshorting switch—A switch in which contact is broken from one position before contact is made with the next. Classified as a "break before make" switch.

nonsinusoidal wave—Any wave that is not a sine wave. It therefore contains harmonics.

nonstorage camera tube—A television camera tube in which the picture signal is always in proportion to the intensity of the illumination on the corresponding area of the scene being televised.

nonsynchronous—Not related in frequency, speed, or phase to other quantities in a device or circuit.

nonsynchronous starting torque — The maximum load torque with which motors can start. However, the motor will not come to a synchronous speed.

nonsynchronous vibrator—A vibrator that interrupts a direct-current circuit at a frequency unrelated to the other circuit constants. It does not rectify the resulting stepped-up alternating voltage.

nontrip-free circuit breaker — A breaker that can be maintained closed by manual override action while a tripping condition persists.

nonuniform field—A field in which the scalar (or vector) at that instant does not have the same value at every point in a given region.

nonvolatile storage — Storage media which retain information in the absence of power and which will make the information available when power is restored.

no operation — In a computer, a specific instruction that causes the computer to perform no operation.

NOR element—A gate circuit having multiple inputs and one output that is energized only if all inputs are zero.

norm—1. The mean or average. 2. A customary condition or degree.

normal condition—The de-energized condition of a relay.

normal contact—A contact which in its normal position closes a circuit and permits current to flow.

normal distribution—The most common frequency distribution in statistics. The probability curve is bell-shaped, and the greatest probability occurs at the arithmetical average (i.e., at the top of the curve). The probability of occurrence of a particular value is shown by the areas between two abscissa values on the curve. *Also see* Gaussian Distribution.

normal electrode—A standard electrode used for measuring electrode potentials.

normal failure period—That period of time during which an essentially constant failure rate exists.

normal impedance—*See* Free Impedance.

normal induction—The limiting induction, either positive or negative, in a magnetic material that is under the influence of a magnetizing force that varies between two extremes.

normal induction curve—The curve obtained by plotting B (induction) against H (magnetizing force), starting from a totally demagnetized state.

normalization—The transforming of signals to a common basis—e.g., adjusting two signals, representing the same spoken word but differing in loudness, to the same loudness.

normalized admittance—The reciprocal of normalized impedance.

normalized impedance—An impedance divided by the characteristic impedance of a wave guide.

normalized plateau slope—The slope of the substantially straight portion of the counting-rate-versus-voltage characteristic of a radiation counter tube, divided by the quotient of the counting rate and the voltage at the Geiger-Mueller threshold.

normal linearity—A manner of expressing linearity as the deviation from a straight line in terms of a given percentage of the output at a certain stimulus value, usually the full-scale value.

normally closed—Symbolized by NC. Designation applied to the contacts of a switch or relay when they are connected so that the circuit will be completed when the switch is not activated or the relay coil is not energized.

normally closed contact—Also called break contact. A contact pair which is closed when the coil of a relay is not energized.

normally open—Symbolized by NO. Designation applied to the contacts of a switch or relay when they are connected so that the circuit will be broken when the switch is not activated or the relay coil is not energized.

normally open contact—A contact pair which is open when the coil of a relay is not energized.

normal mode—The expected or usual operating conditions, such as the voltage that occurs between the two input terminals of an amplifier.

normal mode of vibration—A characteristic distribution of vibration amplitudes among the parts of the system, each part of which is vibrating freely at the same frequency. Complex free vibrations are combinations of these simple vibration forms.

normal-mode voltage—1. The actual signal voltage developed by a transducer. 2. The difference voltage between two input-signal lines.

normal operating period—The time interval between debugging and wear-out.

normal permeability—Ratio of the normal induction to the corresponding magnetizing force. In the cgs system, the flux density in a vacuum is numerically equal to the magnetizing force.

normal position—The position of the relay contacts when the coil is not energized.

normal propagation—The phenomenon of passing radio waves through space when atmospheric and/or ionospheric conditions are such as to permit the passage with little or no difficulty.

normal record level—In direct recording, the amount of record-head current required to produce 1% third-harmonic distortion of the reproduced signal at the record-level set frequency. Such distortion must result from magnetic-tape saturation, not from electronic circuitry.

normal-stage punching—In a computer, a card-punching system in which only the even-numbered rows are punched on the British standard card.

north pole—In a magnet, the pole from where magnetic lines of force are considered to leave the magnet.

Norton's theorem—The current in any impedance Z_R, connected to two terminals of a network, is the same as though Z_R were connected to a constant-current generator the generated current of which is equal to the current that flows through the two terminals when these terminals are short-circuited, the constant-current generator being in shunt with an impedance equal to the impedance of the network, looking back from the terminals in question.

NOT AND circuit—An AND gating circuit which inverts the pulse phase.

notation—A manner of representing numbers. Some of the more important notation scales are:

Base	*Name*
2	binary
3	ternary
4	quaternary, tetral
5	quinary
10	decimal
12	duodecimal
16	hexadecimal, sexidecimal
32	duotricenary
2, 5	biquinary

notch antenna—An antenna that forms a pattern by means of a notch or slot in a radiating surface. Its characteristics are similar to those of a properly proportioned metal antenna and may be evaluated with similar techniques.

notch filter—An arrangement of electronic components designed to attenuate or reject a specific frequency band with sharp cutoff at either end.

notch gate—The early and late gate display on the range crt. It appears as a negative deflection of the range base line equal in width to the early and late gates.

notching—A term indicating that a predetermined number of separate impulses are required to complete operation of a relay.

notching circuit—A control circuit used in a cathode-ray oscilloscope to expand portions of the displayed image.

NOT circuit—In a computer, a circuit in which the output signal has a phase or polarity opposite from that of the input signal. *Also see* inverter.

note—1. The pitch, duration, or both of a tone sensation. 2. The sensation itself. 3. The vibration causing the sensation. 4. The general term when no distinction is desired between the symbol, the sensation, and the physical stimulus. 5. A single musical tone. The notes of the musical scale are referred to by letters running alphabetically from A to G for the white keys. The black keys may be called a sharp of the note below it or a flat of the note above it. Each note has a frequency exactly one-half that of the corresponding note in the next higher octave.

NOT gate—An inhibitory circuit equivalent to the logical operation of negation (mathematical complement). The output of the circuit is energized only when its single input is not energized and there will be no output if the input is energized.

noval base—A nine-pin glass base for a miniature electron tube. For orientation of the tube into its socket, the spacing between pins 1 and 9 is greater than between the other pins.

novice license—A class of amateur license issued only to those who have never held an FCC license of any kind. Its requirements are the easiest of all, but novice operation is limited. Transmitters must be crystal controlled, and the maximum permissible plate input power is 75 watts. Novice licensees are restricted to certain frequencies (3.70-3.75, 7.15-7.20, 21.10-21.25, and 145-147 MHz). A novice license is valid for two years only, and it is not renewable.

NPIN transistor—An intrinsic-region transistor with a P-type base and an N-type emitter and collector.

NPIP transistor—An intrinsic-region transistor in which the intrinsic region is located between two P-regions.

n-plus-one address instruction—In a computer, a multiple-address instruction in which one address serves to specify the location of the next instruction of the normal sequence to be executed.

NPN transistor—A junction transistor with a P-type base and an N-type collector and emitter.

NPNP transistor—A hook transistor with a P-type base, N-type emitter, and a hook collector.

npo body (negative positive zero)—A designation referring to a class of capacitors in the temperature-compensating series that have an essentially variant dielectric constant over a specified temperature range (commonly +25°C to +85°C).

N-quadrant—One of the two quadrants where the N-signal is heard in an A-N radio range.

N-region—1. Also called N-zone. The region where the conduction-electron density in a semiconductor exceeds the hole density. 2. The region created by diffusing into a silicon crystal a group-V element which creates an excess electron condition.

N-scan—*See* N-Display.

N-scope—*See* N-Display.

nsec—Abbreviation for nanosecond (10^{-9} sec).

n-signal—A dash-dot signal heard in either a bisignal zone or an N-quadrant of a radio range.

***n*-terminal network**—A network with *n* accessible terminals.

***n*th harmonic**—A harmonic whose frequency is *n* times the frequency of the fundamental.

NTSC—National Television Systems Committee, named after an all-industry engineering group who developed U.S. color-television specifications. Now describes the American system of color telecasting.

NTSC signal—A 3.579545-MHz (3.58-MHz) signal, the phase of which is varied with the instantaneous hue of the televised color, and the amplitude of which is varied with the instantaneous saturation of the color—as specified by the National Television Systems Committee.

NTSC triangle—On a chromaticity diagram, a triangle which defines the gamut of color obtainable through the use of phosphors.

N-type conductivity—The conductivity associated with conduction electrons in a semiconductor.

N-type crystal rectifier—A crystal rectifier in which forward current flows when the semiconductor is more negative than the metal.

N-type semiconductor—An extrinsic semiconductor in which the conduction-electron density exceeds the hole density. By implication, the net ionized impurity concentration is a donor type.

nuclear battery—Also called an atomic battery. A battery which converts nuclear energy into electrical energy.

nuclear bombardment—The shooting of atomic particles at nuclei, usually in an attempt to split the atom in order to form a new element.

nuclear energy—Also called atomic energy or power. The energy released in a nuclear reaction when a neutron splits the nucleus of an atom into smaller pieces (fission), or when two nuclei are joined together under millions of degrees of heat (fusion).

nuclear fission—*See* Fission.

nuclear fusion—*See* Fusion.

nuclear magnetic resonance—The flipping over of a particle such as a proton as the result of the application of an alternating magnetic field at right angles to a steady magnetic field in which the particle is placed.

nuclear pile—*See* Nuclear Reactor.

nuclear reaction—The reaction, accompanied by a tremendous release of energy, when the nucleus of an atom is split into smaller pieces (fission) or when two or more nuclei are joined together under millions of degrees of heat (fusion).

nuclear reactor—Also called an atomic reactor, nuclear pile, or reactor. A device which can sustain nuclear fission in a self-supporting chain reaction.

nucleation — The occurrence in an existing phase or state of a new phase or state.

nucleonics—The application of nuclear science in physics, chemistry, astronomy, biology, industry, and other fields.

nucleus—The core of an atom. It contains most of the mass and has a positive charge equal to the number of protons it contains. Its diameter is about one ten-thousandth that of the atom. Except for the ordinary hydrogen atom, the nuclei of all other atoms consist of protons and neutrons tightly locked together.

null—1. A balanced condition which results in zero output from a device or system. 2. To oppose an output which differs from zero so that it is returned to zero. 3. The minimum output amplitude (ideally, zero) in direction-finding systems where the amplitude is determined by the direction from which the signal arrives or by the rotation in bearing of the system's response pattern. 4. In a computer, a lack of information as opposed to a zero or blank for the presence of no information.

null balance—A condition in which two or more signals are summed and produce a result that is essentially zero.

null detection—A method of making DF measurements. The antenna is turned to the point where the received signal is weakest. The true bearing of the signal is then found by noting the antenna direction and using a correction factor.

null detector—An apparatus that senses the complete balance, or zero-output condition, of a system or device.

null-frequency indicator—A device that indicates frequency by heterodyning two electrical signals together to give a zero-beat indication.

null indicator—A device that indicates when current, voltage, or power is zero.

null method—Also called zero method. A method of measurement in which the circuit is balanced, to bring the pointer of the indicating instrument to zero, before a reading is taken (e.g., in a Wheatstone bridge, or in a laboratory balance for weighing purposes).

null-spacing error—In a resolver, the difference between 180° and the angle between null positions of the output winding with respect to one input winding.

number—An abstract mathematical symbol for expressing a quantity. In this sense, the manner of representing the number is immaterial. Take 26, for example. This is its decimal form—but it could be expressed as binary 011010 and still mean the same. Some common numbering systems are: binary (base *2*), quinary (base *5*), octonary or octal (base *8*), and decimal (base *10*).

number of overshoots—The total number of times that the transient response goes to a minimum or maximum value that is outside a specified range.

number of scanning lines—The ratio of the scanning frequency to the frame frequency.

number system—Any system for the representation of numbers. (*Also see* Positional Notation.)

numerical control—Descriptive of systems in which digital computers are used for the control of operations, particularly of automatic machines (e.g., drilling or boring machines, wherein the operation control is applied at a discrete point in the operation or process). Contrasted with process control, in which control is applied continuously.

numerical control system—A system controlled by direct insertion of numerical data at some point. The system must automatically interpret at least some portion of the data.

numerical-readout tube—A gas-filled, cold-cathode, digital-indicator tube having a common anode and containing stacked metallic elements in the form of numerals. When a negative voltage is applied to one of them, the element then becomes the cathode of a simple gas-discharge diode.

numeric coding — A system of abbreviation used in the preparation of information for machine acceptance in which all information is reduced to numerical quantities, in contrast to alphabetic coding.

nutating feed—In a tracking radar, an oscillating antenna feed that produces an oscillating deflection of the beam without changing the plane of polarization.

nutation field — The time-variant, three-dimensional field pattern of a directional or beam-producing antenna having a nutating feed.

nuvistor—A small vacuum tube with a cantilever-supported cylindrical electrode that eliminates the need for mica supports. Only metal and ceramic are used, and there is no getter.

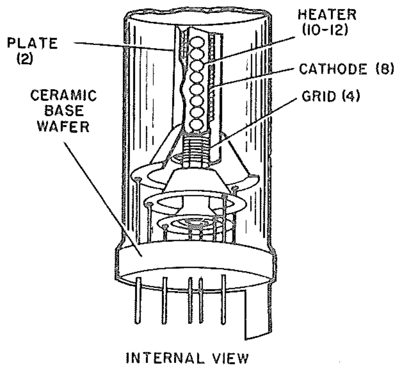

Nuvistor.

nv—Abbreviation for nanovolt.

n-variable cube—See n-Cube.

nw—Abbreviation for nanowatt (10^{-9} watt).

Nyquist diagram—A plot, in rectangular coordinates, of the real and imaginary parts of factor $\mu\beta$ for frequencies from zero to infinity in a feedback amplifier—where μ is the amplification in the absence of feedback, and β is the fraction of the output voltage superimposed on the amplifier input.

Nyquist interval—The maximum separation in time which can be given to regularly spaced instantaneous samples of a wave of a given bandwidth for complete determination of the waveform of the signal. Numerically it is equal in seconds to one-half the bandwidth.

Nyquist rate—Of a channel, the maximum rate at which independent signal values can be transmitted over the specified channel without exceeding a specified amount of mutual interference.

N-zone—*See* N-Region.

O

OAO—Abbreviation for Orbiting Astronomical Observatory.

object program—A computer program which is the output of an automatic coding system. The object program may be a machine-language program ready for execution, or it may be in an intermediate language.

oblique-incidence transmission—Transmission of a radio wave by reflection from the ionosphere.

oboe—A radar navigation system consisting of two ground stations, which measure the distance to an airborne transponder beacon and relay this information to the aircraft.

obsolescence-free—Not likely to become outdated within a reasonable time. Frequently applied to tube testers and other test instruments which have provisions for accommodating new developments.

occlude—To absorb—e.g., some metals will occlude gases, which must be driven out before the metals can be used in the electrodes or supports of a vacuum tube.

occluded gas—Gas that has been absorbed by a material (e.g., by the electrodes, supports, leads, and insulation of a vacuum tube).

occupied bandwidth—The frequency bandwidth such that, below its lower and above its upper frequency limits, the mean powers radiated are each equal to 0.5% of the total mean power radiated by a given emission.

octal—*See* Positional Notation.

octal base—An eight-pin tube base (although unneeded pins are often omitted without changing the position of the remaining pins). An aligning key in the center of the base assures correct insertion of the tube into the socket.

octal digit—One of the symbols 0, 1, 2, 3, 4, 5, 6, and 7 when used in numbering in the scale of 8.

octal number system—A number system in which the equivalent of the decimal integer 8 is used as a base.

octave—The interval between two sounds having a basic frequency ratio of two—or, by extension, the interval between any two frequencies having a ratio of 2:1. The interval, in octaves, between any two frequencies is the logarithm to the base *2* (or 3.322 times the logarithm to the base *10*) of the frequency ratio.

octave band—A band of frequencies the limits of which have the ratio 2 to 1.

octave-band pressure level—Also called octave pressure level. The pressure level of a sound for the frequency band corresponding to a specified octave.

octave pressure level — *See* Octave-Band Pressure Level.

octode—An eight-electrode electron tube containing an anode, a cathode, a control electrode, and five additional grids.

octonary—*See* Positional Notation.

odd-even check — An automatic computer check in which an extra digit is carried along with each word, to determine whether the total number of 1's in the word is odd or even, thus providing a check for proper operation.

odd harmonic—Any frequency that is an odd multiple of the fundamental frequency (e.g., 1, 3, 5, etc.).

odd-line interlace—The double-interlace system in which, since there are an odd number of lines per frame, each field contains a half line. In the 525-line television frame used in the United States, each field contains 262.5 lines.

odograph—Automatic electronic map tracer used in military vehicles for map making and land navigation. It automatically plots, on an existing map or on cross-sectional paper, the exact course taken by the vehicle. This is done by phototubes and thyratrons, which transfer the indication of a precision magnetic compass onto a plotting unit actuated by the speedometer drive cable. A pen then traces the course.

oersted—In the cgs electromagnetic system, the unit of magnetizing force equal to $1{,}000/4\pi$ ampere-turns per meter.

off-center display—A ppi display, the center of which does not correspond to the position of the radar antenna.

off-delay — A circuit that retains an output signal some definite time after the input signal is removed.

off-ground—The voltage above or below ground at which a device is operated.

off-limit contacts—Contacts on a stepping relay used to indicate that the wiper has reached the limiting position on its arc and must be returned to normal before the circuit can function again.

off-line operation—1. Computer operation that is independent of the time base of the actual inputs. 2. In a computer system, operation of peripheral equipment independent from the central processor, e.g., the transcribing of card information to magnetic tape, or of magnetic tape to printed form.

off-normal contacts—Relay contacts that assume one condition when the relay is in its normal position and the reverse condition for any other position of the relay.

off-period—That portion of an operating cycle during which an electron tube or semiconductor is nonconducting.

offset—The measure of unbalance between halves of a symmetrical circuit. Generally caused by differences in transistor betas or in values of biasing resistors.

offset angle—In lateral-disc reproduction, the smaller of the two angles between the projections, into the plane of the disc, of the vibration axis of the pickup stylus and the line connecting the vertical pivot (assuming a horizontal disc) of the pickup arm with the stylus point.

offset current—A d-c error current appearing at either input terminal of a d-c amplifier.

offset stacker—In a computer, a card stacker having the ability to stack cards selectively under machine control so that they protrude from the balance of the stack, thus giving physical identification.

offset voltage—A d-c error voltage appearing in series with either input terminal of a d-c amplifier.

off-target jamming—Employment of a jammer away from the main units of the force. This is done to prevent the enemy from monitoring the jamming signals and using them to pinpoint the location of the force.

OGO—Abbreviation for Orbiting Geophysical Observatory.

ohm—Symbolized by the Greek letter *omega* (Ω). The unit of resistance. It is defined as the resistance, at 0°C., of a uniform column of mercury 106.300 cm long and weighing 14.4521 grams. One ohm is the value of resistance through which a potential difference of one volt will maintain a current of one ampere.

ohmic contact—1. Between two materials, a contact across which the potential difference is proportionate to the current passing through it. 2. A contact that has linear voltage-current characteristics through its entire operating range.

ohmic heating—The energy imparted to charged particles as they respond to an electric field and make collisions with other particles. The name was chosen due to the similarity of this effect to the heat generated in an ohmic resistance due to the collisions of the charge carriers in their medium.

ohmic resistance—Resistance to direct current.

ohmic value—The resistance in ohms.

ohmmeter—A direct-reading instrument for measuring electric resistance. Its scale is usually graduated in ohms, megohms, or both. (If the scale is graduated in megohms, the instrument is called a megohmmeter; if the scale is calibrated in kilohms, the instrument is a kilohmmeter.)

ohmmeter zero adjustment—In an ohmmeter, a potentiometer or other means of compensating for the drop in battery voltage with age. Usually a knob is rotated until the meter pointer is at zero on the particular scale being used.

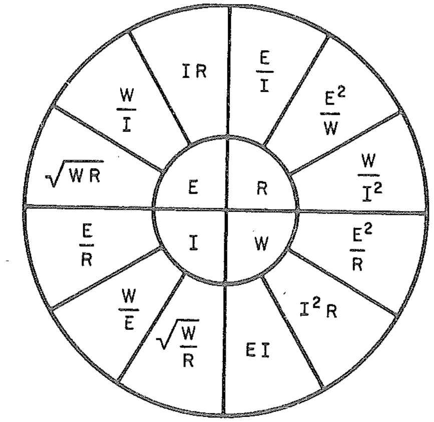

Ohm's law.

Ohm's law—The voltage across an element of a d-c circuit is equal to the current in amperes through the element, multiplied by the resistance of the element in ohms. Expressed mathematically as $E = I \times R$. The other two equations obtained by transposition are $I = E/R$ and $R = E/I$.

ohms per square—The resistance of any square area of thin-film resistive material as measured between two parallel sides.

ohms-per-volt—A sensitivity rating for voltage-measuring instruments (the higher the rating, the more sensitive the meter). On any particular range, it is obtained by dividing the resistance of the instrument (in ohms) by the full-scale voltage value of that range.

oil—A prefix used with a device which interrupts an electrical circuit; it indicates that the interruption occurs in oil.

oil circuit breaker—A circuit breaker in which the interruption occurs in oil to suppress the arc and prevent damage to the contacts.

oiled paper—Paper that has been treated with oil or varnish to improve its insulating qualities.

oil-filled cable—A cable having insulation impregnated with an oil which is fluid at all operating temperatures and provided with facilities such as longitudinal ducts or channels and with reservoirs, or their equivalent, by means of which positive oil pressure can be maintained within the cable at all times.

oil fuse cutout—An enclosed fuse cutout in which all or part of the fuse support is mounted in oil.

oil-immersed forced-oil-cooled transformer (class FOA or FOW)—A transformer having its core and coils immersed in oil and cooled by the forced circulation of its oil through an external heat exchanger. If the heat exchanger utilizes forced circulation of air, the transformer is class FOA; if it utilizes forced circulation of water, the transformer is class FOW.

oil-immersed self-cooled forced-air-cooled transformer (class OA/FA)—A transformer having its core and coils immersed in oil and having a self-cooled rating with cooling obtained by the natural circulation of air over the cooling surface and a forced-air-cooled rating with cooling obtained by the forced circulation of air over this same cooling surface.

oil-immersed self-cooled transformer (class OA)—A transformer having its core and coils immersed in oil, the cooling being effected by the natural circulation of air over the cooling surface. (Smooth tank surfaces are generally sufficient to cool small transformers, but for about 25 kva and larger, a supplementary heat-dissipating surface is provided in the form of external fins, tubes, or tubular radiators.)

oil-immersed transformer—A transformer the core and coils of which are immersed in oil.

oil-immersed water-cooled transformer (class OW)—A transformer having its core and coils immersed in oil, the cooling being effected by the natural circulation of oil over the water-cooled surface. The water-cooled surface is provided by water flowing through copper cooling coils immersed in the oil and located inside the main transformer tank, through which the oil flows by thermosiphon action.

oil switch—A switch in which the interruption of the circuit occurs in oil to suppress the arc and prevent damage to the contacts.

Omega—A long-range hyperbolic navigation system which transmits interrupted c-w signals from which phase differences are extracted. Transmissions occur from two or more locations on a single carrier frequency utilizing time-sharing techniques.

omnibearing—The bearing indicated by a navigational receiver on transmissions from an omnirange.

omnibearing converter—An electromechanical device which combines an omnirange signal with aircraft heading information to furnish electrical signals for operating the pointer of a radio magnetic indicator. (*Also see* Omnibearing Indicator.)

omnibearing indicator—An omnibearing converter to which a dial and pointer have been added.

omnibearing line—One of an infinite number of straight, imaginary lines radiating from the geographical location of a vhf omnirange.

omnibearing selector—Instrument capable of being set manually to any desired omnibearing or its reciprocal in order to control a coarse-line deviation indicator.

omnidirectional—Also called nondirectional. All-directional; not favoring any one direction.

omnidirectional antenna—Also called nondirectional antenna. An antenna producing essentially the same field strength in all horizontal directions and a directive vertical radiation pattern.

omnidirectional hydrophone—A hydrophone having a response that is essentially independent of the angle of arrival of the incident sound wave.

omnidirectional microphone—Also called nondirectional microphone. A microphone that responds to a sound wave from almost any angle of arrival.

omnidirectional radio range—Radio air-to-air navigation using a transmitter that radiates throughout 360° azimuth, providing aircraft with a direct indication of the bearing of the transmitter.

omnidirectional range—Also called omnirange. A radio facility providing bearing information to or from such facilities at all azimuths within its service area.

omnidirectional range station—In the aeronautical radionavigational service, a land station that provides a direct indication of the bearing (omnibearing) of that station from an aircraft.

omnigraph—An instrument that produces Morse-code messages for instruction purposes. It contains a buzzer circuit which is usually actuated by a perforated tape.

omnirange—*See* Omnidirectional Range.

on-call channels—Similar to allocated channels except that full-time exclusive use of the channel is not warranted.

on-course curvature—In navigation, the rate at which the indicated course changes with respect to the distance along the course path.

on-course signal—The monotone radio signal which indicates to the pilot that he is neither too far to the right nor to the left of the radio beam being followed.

on-delay—A circuit that produces an output signal some definite time after an input signal is applied.

on-demand system—A system from which the desired information or service is available at the time of request.

ondograph—An instrument for drawing alternating voltage waveform curves. It employs a capacitor which is momentarily charged to the amplitude at a particular point on the curve and then discharged through a recording galvanometer. This is repeated at intervals of about once every hundred cycles with the sample taken a little farther along on the waveform each time.

ondoscope—A glow-discharge tube used on an insulating rod to indicate the presence of high-frequency r-f near a transmitter. The radiation ionizes the gas in the tube, and the visible glow indicates the presence of r-f.

one-input terminal—Called the set terminal. The terminal which, when triggered, will put a flip-flop in the one (opposite of starting) condition.

one-many function switch — A function switch in which only one input is excited at a time and each input produces a combination of outputs.

one output—*See* One State.

one-output terminal—The terminal which produces an output of the correct polarity to trigger a following circuit when a flip-flop circuit is in the one condition.

one-shot—A circuit that produces an output signal of fixed duration when an input signal of any duration is applied.

one-shot multivibrator—*See* Monostable Multivibrator.

one state—Also called one output. In a magnetic cell, the positive value of the magnetic flux through a specified cross-sectional area, determined from an arbitrarily specified direction. (*Also see* Zero State).

one-third-octave band — A frequency band in which the ratio of the extreme frequencies is 1.2599.

one-to-partial-select ratio—In a computer, the ratio of a *1* output to a partial-select output.

one-to-zero ratio—In a computer, the ratio of a *1* output to the *0* output.

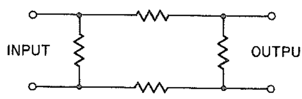

O-network.

O-network—A network composed of four impedance branches connected in series to form a closed circuit. Two adjacent junction points serve as input terminals, and the remaining two as output terminals.

one-way communication—Applied to certain radio-communication or intercommunications systems where a message is transmitted from one station to one or more receiving stations that have no transmitting apparatus.

one-way repeater—*See* Repeater.

on line—In a computer, having to do with the operation of input-output devices under direct control of the central processing unit.

on-line data reduction—The processing of information as rapidly as the information is received by the computing system or as rapidly as it is generated by the source.

on-line operation—An operation carried on within the main computer system (e.g., computing and writing results onto a magnetic tape, printed report, or paper tape).

on-off control—A simple control system in a switch in which the device being controlled is either fully on or fully off and no operating positions are available.

on-off keying—Keying in which the output of a source is alternately transmitted and suppressed to form signals.

on-off ratio—Ratio of the duration (on) of a pulse to the space (off) between successive pulses.

on-off switch—*See* Power Switch.

on period—That portion of an operating cycle during which an electron tube is conducting.

on the air—Transmitting.

opacimeter — Also called turbidimeter. A photoelectric instrument for measuring the turbidity (amount of sediment) of a liquid. It does this by determining the amount of light that passes through the liquid.

opacity—The degree of nontransparency of a substance—i.e., its ability to obstruct, by absorption, the transmission of radiant energy such as light. Opacity is the reciprocal of transmission.

open—A circuit interruption that results in an incomplete path for current flow (e.g., a broken wire, which opens the path of the current).

open-center display—A ppi display on which zero range corresponds to a ring around the center of the display.

open circuit—A circuit which does not provide a complete path for the flow of current. (*Also see* Open.)

open-circuit impedance—The driving-point impedance of a line or four-terminal network when the far end is open.

open-circuit jack — A jack that normally leaves its circuit open. The circuit can be closed only by inserting the plug into the jack.

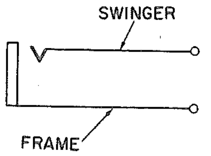

Open-circuit jack.

open-circuit signaling—Signaling in which no current flows under normal (i.e., inoperative) conditions.

open-circuit voltage—The voltage at the terminals of a battery or other voltage source when no appreciable current is flowing.

open core—A core fitting inside a coil but having no external return path. The magnetic circuit thus has a long path through air.

open-delta connection — Two single-phase transformers connected so that they form only two sides of a delta, instead of the three sides with three transformers in a regular delta connection.

open-fuse cutout—An enclosed fuse cutout in which the fuse support and fuse holder are exposed.

open loop—Implies an operational amplifier without feedback.

open-loop bandwidth — Without feedback, the frequency limits at which the voltage gain of the device drops off 3 db below the gain at some lower reference frequency.

open-loop control system—A control system in which there is no self-correcting action, as there is in a closed-loop system.

open-loop differential voltage gain — *See* Differential Voltage Gain.

open-loop gain—The ratio of the (loaded) output of an amplifier without any feedback to its net input at any frequency. Usually implies voltage gain.

open-loop output impedance—The complex impedance seen looking into the output terminals of an operational amplifier with no external feedback and in the linear-amplification region. In closed-loop operation, the output impedance is equal to the open-loop impedance divided by the loop gain. If the open-loop impedance is not more than a few hundred ohms and the loop gain is high enough for good gain accuracy and stability, the closed-loop impedance will be on the order of an ohm or less, which can be neglected in most applications.

open-loop output resistance—The resistance looking into the output terminal of an operational amplifier operating without feedback and in the linear-amplification region.

open-loop voltage gain—The ratio of the change in voltage between the differential input terminals of an operational amplifier to the corresponding change in output voltage, with no feedback.

open magnetic circuit—A magnet that has no closed external ferromagnetic circuit and does not form a complete conducting circuit itself (e.g., a permanent magnet ring interrupted by an air gap).

open motor—A motor having ventilating openings which permit passage of external cooling air over and around the windings of the machine. When such motors have an internal fan to aid the movement of ventilating air, they may be referred to as self-ventilated.

open-phase protection—The effect of a device operating on the loss of current in one phase of a polyphase circuit to cause and maintain the interruption of power in the circuit.

open-phase relay—A relay which functions when one or more phases of a polyphase circuit open and sufficient current is flowing in the remaining phase or phases.

open plug—A plug designed to hold jack springs in their open position.

open relay—A relay not having an enclosure.

open routine—In a computer, a routine that it is possible to insert directly into a larger routine without a linkage or calling sequence.

open subroutine—Also called direct-insert subroutine. A subroutine inserted directly into a larger sequence of instructions. Such a subroutine is not entered by a jump instruction; hence, it must be recopied at each point where it is needed.

open-temperature pickup — A temperature transducer in which its sensing element is directly in contact with the medium whose temperature is being measured.

open wire—A conductor separately supported above the earth's surface.

open-wire circuit—A circuit made up of conductors separately supported on insulators.

open-wire loop—The branch line on a main open-wire line.

open-wire transmission line—A transmission line formed by two parallel wires. The distance between them, and their diameters, determine the surge impedance of the transmission line.

operand—A computer word on which an operation is to be performed.

operate current—The minimum current required to trip all the contact springs of a relay.

operate time—The time that elapses, after power is applied to a relay coil, until the contacts being checked have operated (i.e., first opened in a normally closed contact, or first closed in a normally open contact).

operate-time characteristic—The relation between the operate time of an electromagnetic relay and the operate power.

operating angle—The electrical angle (portion of a cycle) during which plate current flows in an amplifier or an electronic tube. Class-A amplifiers have an operating angle of 360°; Class-B, 180° to 360°; and Class-C, less than 180°.

operating cycle—The complete sequence of operations required in the normal functioning of an item of equipment.

operating frequency—The rated a-c frequency of the supply voltage at which a relay is designed to operate.

operating life—The minimum length of time over which the specified continuous and intermittent rating of a device, system, or transducer applies without change in performance beyond the specified tolerance.

operating-mode factor—A failure-rate modifier determined by the type of equipment environment, e.g., laboratroy computer, nosecone compartment, etc.

operating overload—The overcurrent to which electric apparatus is subjected in the course of the normal operating conditions that it may encounter.

operating point—Also called the quiescent point. On a grid-voltage-plate-current characteristic curve of a vacuum tube, the point which corresponds to the direct-voltage value being used for the grid and plate.

operating position—The operator-attended terminal of a communications channel. It is usually used in its singular sense (e.g., a radio-operator's position, a telephone-operator's position), even when there is more than one operating position.

operating power—The power actually reaching a transmitting antenna.

operating ratio—*See* Availability.

operating temperature—The temperature or range of temperatures at or over which a device is expected to operate within specified limits of error.

operating temperature range—The interval of temperatures in which a component device or system is intended to be used, specified by the limits of this interval.

operating time — The time period between turn-on and turn-off of a system, subsystem, component, or part during which the system, etc., functions as specified. Total operating time is the summation of all operating-time periods.

operating voltages—The direct voltages applied to the electrodes of a vacuum tube under operating conditions.

operation—A specific action which a computer will perform whenever an instruction calls for it (e.g., addition, division).

operational amplifier—An amplifier that performs various mathematical operations. Application of negative feedback around a high-d-c amplifier produces a circuit with a precise gain characteristic that depends on the feedback used. By the proper selection of feedback components, operational amplifier circuits can be used to add, subtract, average, integrate, and differentiate. An operational amplifier can have a single-input-single-output, differential-input-single-output, or differential-input-differential-output configuration.

operational differential amplifier—An operational amplifier which utilizes a differential amplifier in the first stage.

operational readiness—The probability that, at any point in time and under stated conditions, a system will be either operating satisfactorily or ready to be placed in operation.

operational reliability—*See* Achieved Reliability.

operation code (computer)—1. Also called the operation part of an instruction. The list of operation parts occurring in an instruction code, together with the names of the corresponding operations (e.g., "add," "unconditional transfer," "add and clear," etc.). 2. The part of a computer instruction word which specifies, in coded form, the operation to be performed.

operation decoder—In a computer, circuitry which interprets the operation-code portion of the machine instruction to be executed and sets other circuitry for its execution.

operation number—In a computer program, a number indicating the position of a given operation or its equivalent subroutine.

operation part—In a computer instruction, the part that specifies the kind of operation, but not the location of the operands. (*Also see* Instruction Code.)

operation register—In a computer, the register that stores the operation-code portion of an instruction.

operation time—The amount of time required by a current to reach a stated fraction of its final value after voltages have been applied simultaneously to all electrodes. The final value is conventionally taken as that reached after a specified length of time.

operator—Any person who operates, adjusts, and maintains equipment—specifically, a computer, radio transmitter, or other communications equipment.

opposition—The phase relationship between two periodic functions of the same period when the phase difference between them is half of a period.

optical ammeter—An electrothermic instrument for measuring the current in the filament of an incandescent lamp. It normally uses a photoelectric cell and indicating instrument to compare the illumination with that produced by a current of known magnitude in the same filament.

optical axis—The Z-axis of a crystal.

optical character recognition — An automatic means of direct reading and recognition of data in printed or hand-lettered form.

optical damping—A damping ratio slightly less than unity in which the overshoot is less than the specified uncertainty of the instrument.

optical maser—*See* Laser.

optical mode—In a crystal lattice, a mode of vibration that produces an oscillating dipole.

optical pattern—Also called Christmas-tree pattern. In mechanical recording, the pattern observed when the surface of the record is illuminated by a light beam of essentially parallel rays.

optical pyrometer—A temperature-measuring device comprising a comparison source of illumination, together with some convenient arrangement for matching this source—either in brightness or color—against the source whose temperature is to be measured. The comparison is usually made by the eye.

optical sound recorder—Also called a photographic sound recorder. Equipment for optically recording the electric signals derived from sound signals. A light modulator is incorporated, plus means for moving a light-sensitive medium relative to the modulator.

optical sound reproducer — Also called a photographic sound reproducer. Equipment for converting optical sound into electrical signals. It consists of a light source, an optical system, a photoelectric cell (or other light-sensitive device such as a photoconductive cell), and a mechanism for moving the medium (usually film) carrying the optical sound record.

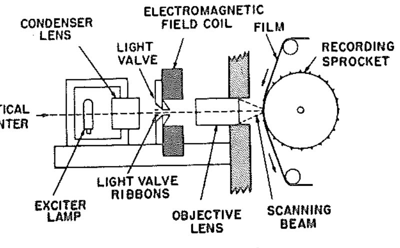

Optical sound recorder.

optical spectrometer—An instrument with an entrance slit, a dispersing device, and one or more exit slits, with which measurements are made at selected wavelengths within the spectral range, or by scanning over the range. The quantity detected is a function of radiant power.

optical twinning—A defect occurring in natural quartz crystals. The right quartz and left quartz both occur in the same crystal. This generally results in small regions of unusable material, which are discarded when the crystal is cut up.

optics—The branch of science concerned with vision—i.e., the nature and propagation of electromagnetic radiation in the infrared, visible, and ultraviolet regions.

optimization—The continual adjustment of a process for the best obtainable set of conditions.

optimum bunching—The bunching condition required for maximum output in a velocity-modulation tube.

optimum coupling—*See* Critical Coupling.

optimum load—The value of load impedance which will transfer maximum power from the source to the load.

optimum plate load—The ideal plate-load impedance for a given tube and set of operating conditions.

optimum reliability—The value of reliability that yields a minimum total successful mission cost.

optimum working frequency — The frequency at which transmission by ionospheric reflection can be expected to be most effectively maintained between two specified points and at a certain time of day. (For propagation by way of the F_2 layer, the optimum working frequency is often taken as being 15% below the monthly median value of the F_2 maximum usable frequency for a specified path and time of day.)

optoelectronic integrated circuit—An integrated component that uses a combination of electroluminescence and photoconductivity in the performance of all or at least a major portion of its intended function.

optoelectronics—Technology dealing with the coupling of functional electronic blocks by light beams.

optoelectronic transistor—A transistor that has an electroluminescent emitter, a transparent base, and a photoelectric collector.

optophone—A photoelectric device that converts light energy into sound energy. Thus, a blind person by using a selenium cell and a circuit for connecting the resulting signals into sounds of corresponding pitch can "read" by ear.

orange peel—A term applied to the surface of a recording blank which resembles an orange peel. Such a surface has a high background noise.

orbit—1. The path of a body or particle under the influence of a gravitational or other force. For instance, the orbit of a celestial body is its path relative to another body around which it revolves. 2. To go around the earth or other body in an orbit.

orbital electron—An electron which is visualized as moving in an orbit around the nucleus of an atom or molecule—as opposed to a free electron.

OR circuit—*See* OR Gate.

order—In computer terminology, the synonym for instruction, command, and—loosely—operation part. These three usages, however, are losing favor because of the ambiguity between them and the more common meanings in mathematics and business.

ordering—In a computer, the process of sorting and sequencing.

orders of logic—A measure of the speed with which a signal can propagate through a logic network (commonly referred to as orders-of-logic capability).

order tone—A tone sent over trunks to indicate the trunk is ready to receive an order, and to the receiving operator that it is about to arrive.

ordinary wave—Sometimes called the O-wave. One of the two components into which the magnetic field of the earth divides a radio wave in the ionosphere. The other component is called the extraordinary wave, or X-wave.

ordinate—1. The vertical line, or one of the coordinates drawn parallel to it, on a graph. 2. The vertical scale on a graph.

organ—In a computer subassembly, the portion which accomplishes some operation or function (e.g., arithmetic organ).

OR gate—Also called an OR circuit. A gate that performs the function of logical "inclusive or." It produces an output whenever any one (or more) of its inputs is energized.

orient—To position or otherwise adjust with respect to some reference point (e.g., to orient an antenna for best reception).

oriented — In crystallography, a crystal in which the axes of its individual grains are aligned so that they have directional magnetic properties.

orifice—An opening or window—specifically, in a side or end wall of a waveguide or cavity resonator, an opening through which energy is transmitted.

original lacquer—An original disc recording made on a lacquer surface for the purpose of producing a master.

original master—Also called metal master or metal negative. In disc recording, the master produced by electroforming from the face of a wax or lacquer recording.

orthicon—Also called an orthiconoscope. A camera tube in which a beam of low-velocity electrons scans a photoemissive mosaic capable of storing an electrical-charge pattern. It is more sensitive than the iconoscope.

orthiconoscope—*See* Orthicon.

orthocode—An arrangement of black and white bars that resemble a piano keyboard and that can be read by an electric-eye device.

orthocore—A completely closed flux memory device designed to almost duplicate the geometry of the ferrite core memory, eliminate the wiring of memory cores, and provide a plurality of wires through the memory element. The concept involves the formation of a group of plastic rods around a suitable wiring array.

orthogonal axes—Axes which are perpendicular to each other. In an instrument, these axes usually coincide with its axes of symmetry.

osciducer—A transducer in which information pertaining to the stimulus is provided in the form of deviation from the center frequency of an oscillator.

oscillating current—An alternating current—specifically, one that changes according to some definite law.

oscillating quantity—A quantity which alternately increases and decreases in value, but always remains within finite limits—e.g., the discharge of current from a capacitor through an inductive resistance (provided the inductance is greater than the capacitance times the resistance squared).

oscillating transducer—A transducer in which information pertaining to the stimulus is provided in the form of deviation from the center frequency of an oscillator.

oscillation—1. The state of a physical quantity when, in the time interval under consideration, the value of the quantity is continually changing in such a manner that it passes through maxima and minima (e.g., oscillating pendulum, oscillating electric current, and oscillating electromotive force). 2. Fluctuations in a system or circuit, especially those consisting of the flow of electric currents alternately in opposite directions; also, the corresponding changes in voltages.

oscillator—1. An electronic device which generates alternating-current power at a frequency determined by the values of certain constants in its circuits. An oscillator may be considered an amplifier with positive feedback, with circuit parameters that restrict the oscillations of the device to a single frequency. 2. A nonrotating device capable of setting up and maintaining oscillations at a frequency determined by the physical constants of the system, such as a vacuum-tube, sparks, or arc generator. 3. A circuit generally using a vacuum tube capable of converting direct current into alternating current of a frequency determined by the inductive and capacitive constant of the circuit. 4. A device used to generate and oscillate at radio frequencies.

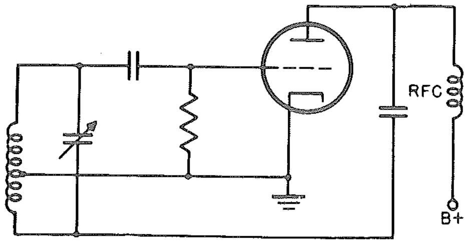

Oscillator (shunt-fed Hartley).

oscillator circuit—*See* Oscillator.

oscillator coil—A radio-frequency transformer that provides the feedback required for oscillation in the oscillator circuit of a superheterodyne receiver or in other oscillator circuits.

oscillator harmonic interference—Interference caused in a superheterodyne receiver by the interaction of incoming signals with harmonics (usually the second harmonic) of the local oscillator.

oscillator-mixer-first detector—A single stage which, in a superheterodyne receiver, combines the functions of the local oscillator and the mixer-first detector. It usually employs a pentagrid converter tube.

oscillator padder—An adjustable capacitor placed in series with the oscillator tank circuit of a superheterodyne receiver. It is used to adjust the tracking between the oscillator and preselector at the low-frequency end of the tuning dial.

oscillator radiation—The amount of voltage available across the antenna terminals of a receiver, (or at a distance) traceable to any oscillators incorporated in the receiver.

oscillatory circuit—A circuit containing inductance and/or capacitance and resistance, so arranged or connected that a voltage impulse will produce a current which periodically reverses.

oscillatory current—A current which periodically reverses its direction of flow.

oscillatory discharge—Alternating current of gradually decreasing amplitude which, under certain conditions, flows through a circuit containing inductance, capacitance, and resistance when a voltage is applied. *Also see* Damped Wave.

oscillatory surge—A surge which includes both positive and negative polarity values.

oscillistor—A semiconductor bar which is subjected to a magnetic field and a direct current, and which generates oscillations believed to be due to diffusion of ions toward the surface of the semiconductor as a result of the magnetic field.

oscillogram—The recorded trace produced by an oscillograph.

oscillograph—An instrument primarily for producing a record of the instantaneous values of one or more rapidly varying electrical quantities as a function of time, or of another electrical or mechanical quantity.

oscillograph recorder—A form of mechanical oscillograph in which the waveform is traced on a moving strip of paper by a pen.

oscillograph tube—*See* Oscilloscope Tube.

oscillography—The art and practice of utilizing the oscillograph.

oscilloscope—An instrument primarily for making visible the instantaneous values of one or more rapidly varying electrical quantities as a function of time, or of another electrical or mechanical quantity. It does not inherently produce a permanent record.

oscilloscope differential amplifier—A device that amplifies and displays the voltage difference that exists at every instant between signals applied to its two inputs.

oscilloscope tube—Also called oscillograph tube. A cathode-ray tube that produces a visible pattern which is the graphical representation of electric signals. The pattern is seen as a spot or spots, which change position in accordance with the signals.

OSO—Abbreviation for Orbiting Solar Observatory.

O-type backward-wave oscillator—A wideband, voltage-tunable microwave oscillator that uses a fundamental or space harmonic with phase and group velocity of different signs.

outdoor antenna—A receiving antenna located on an elevated site outside a building.

outdoor transformer—A transformer of weatherproof construction.

outer marker—In an instrument-landing system, a marker located on a localizer course line at a recommended distance (normally about 4½ miles) from the approach end of the runway.

outlet—The point where current is taken from a wiring system.

outline drawing—A drawing showing approximate over-all shape, but no detail.

out of phase—Two or more waveforms that have the same shape, but do not pass through corresponding values at the same instant.

outphasing—In electronic organs, a term applied to a method sometimes used for producing certain voices. Special circuitry, placed between the keying-system output and the formant filters, either adds or subtracts harmonics or subharmonics of the tone-generator signal.

output—1. The current, voltage, power, or driving force delivered by a circuit or device. 2. The terminals or other places where the circuit or device may deliver the current, voltage, power, or driving force. 3. Information transferred from the internal to the secondary or external storage of a computer.

output amplifier—A circuit that energizes high-power-level devices upon application of a low-power-level input signal.

output axis—The axis around which the spinning wheel of a gyroscope precesses after the wheel has received an input.

output block—In a computer, a portion of the internal storage reserved for holding data which is to be transferred out.

output capacitance—1. Of an n-terminal electron tube, the short-circuit transfer capacitance between the output terminal and all other terminals, except the input terminal, connected together. 2. The shunt capacitance at the output terminal of a device.

output capacitive loading—The maximum capacitance that can be placed on the output of an operational amplifier at unity gain without increasing the phase shift to the point of inducing oscillation. The limiting value increases in direct proportion to the closed-loop gain.

output capacity—The number of loads that can be driven by the output of a circuit.

output device—The part of a machine which translates the electrical impulses representing data processed by the machine into permanent results such as printed forms, punched cards, and magnetic writing on tape.

output equipment—Equipment that provides information in visible, audible, or printed form from a computer.

output gap—An interaction gap with which usable power can be extracted from an electron stream.

output impedance—1. The impedance measured at the output terminals of a transducer with the load disconnected and all impressed driving forces (including those connected to the input) taken as zero. 2. Also called dynamic output impedance. The impedance presented by power supply to the load. It is calculated from the ratio of the change in output voltage (at the prescribed terminals) to the change in load current causing the change. The impedance is specified from d-c to a stated maximum a-c.

output indicator—A meter or other device that indicates variations in the signal strength at the output circuits.

output limits—The maximum output signal available when an operational amplifier is operated in the saturation region.

output load current—The maximum current that the amplifier will deliver to, or accept from, a load. This rating includes the amount, however small, which is caused to flow in the feedback loop.

output meter—An alternating-current voltmeter that measures the signal strength at the output of a receiver or amplifier.

output-meter adapter—A device that can be slipped over the plate prong of the output tube of a radio receiver to provide a conventional terminal to which an output meter can be connected during alignment.

output power—The power which a system or component delivers to its load.

output saturation voltage—The lowest voltage level to which the collector of the output transistor can be reduced without degrading circuit performances.

output stage—The final stage in any electronic equipment. In a radio receiver, it feeds the speaker directly or through an output transformer. In an audio-frequency amplifier, it feeds one or more speakers, the cutting head of a sound recorder, a transmission line, or any other load. In a transmitter, it feeds the antenna.

output transformer—A transformer used for coupling the plate circuit of one or more power tubes to a speaker or other load.

output tube—A power-amplifier tube designed for use in an output stage.

output voltage—The maximum output voltage which an amplifier will develop in the linear operating region (i.e., before the onset of saturation).

output winding—The winding of a saturable reactor, other than a feedback winding, through which power is delivered to the load.

outside lead—*See* Finish Lead.

overall electrical efficiency (induction- and dielectric-heating usage)—Ratio of the power absorbed by the load material, to the total power drawn from the supply lines.

overall thermoelectric generator efficiency—The ratio of electrical power output to thermal power input to the thermoelectric generator.

overall ultrasonic system efficiency—The acoustical power output at the point of application, divided by the electrical power input into the generator.

overbunching—The condition where the buncher voltage of a velocity-modulation tube is higher than required for optimum bunching of the electrons.

overcompounding—In a compound-wound generator, use of sufficient series turns to raise the voltage as the load increases, in order to compensate for the increased line drop. In a motor, overcompounding makes it run faster as the load increases.

overcoupled circuit—A tuned circuit in which the coupling is greater than the critical coupling. The result is a broad-band response characteristic.

overcurrent—In a circuit, the current which will cause an excessive or even dangerous rise in temperature in the conductor or its insulation.

overcurrent protective device—A device operative on excessive current which causes and maintains the interruption of power in the circuit.

overcutting—In disc recording, the cutting through of one groove into an adjacent one when the level becomes excessive.

overdamping—1. Any periodic damping greater than the amount required for critical damping. 2. The absence of overshoot or oscillation as the output of a transducer reaches a final steady value in response to a step change in the measurand. 3. *See* Aperiodic Damping.

overdriven amplifier—An amplifier stage designed to distort the input-signal waveform by permitting the grid signal to drive it beyond cutoff or even into plate-current saturation.

overflow—1. The condition occurring whenever the result of an arithmetic operation exceeds the capacity of the number representation in a digital computer. 2. The carry digit arising from (1) above.

overflow position—In a computer, an extra register position in which the overflow digit is developed.

overhang—1. In terms of printed wiring, the inverted shelf of plating formed when conductor material is selectively removed from under the plating. The measurement between the base copper and the plating (generally refers to one side of the conductor only; thus, for a conductor width reduction, one must take two times the overhang). 2. The critical dimension by which the stylus overreaches the center spindle of a turntable when an offset tone arm is mounted for minimum tracking error.

overhead line—A conductor carried on elevated poles (e.g., telephone or telegraph wires).

overinsulation—In a coil, the insulating material placed over a wire which is brought from the center over the top or bottom wall.

overlap—The amount by which the effective height of the scanning facsimile spot exceeds the nominal width of the scanning line. When the spot is rectangular, overlap may be expressed as a percentage of the nominal width of the scanning line.

overlapping contacts—Combinations of two sets of contacts actuated by a common means, each set closing in one of two positions, and so arranged that the contacts of one set open after the contacts of the other set have been closed.

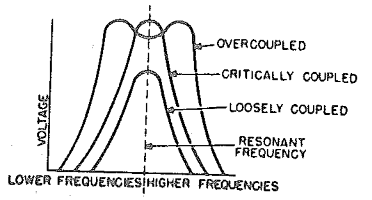

Overcoupled circuit.

overlap radar—Long-range radar which is located in one sector but also covers a portion of another sector.

overlay—In a computer, the technique of using the same blocks of internal storage for different routines during different stages of a problem, e.g., when one routine is no longer needed in internal storage, another routine can be placed in that storage location.

overlay transistor—A transistor containing a large number of emitters connected in parallel to provide maximum power amplification at extremely high frequencies.

overload—1. A load greater than that which an amplifier, other component, or a whole transmission system is designed to carry. It is characterized by waveform distortion or overheating. 2. In electronics, that quantity of power from an amplifier or other component or from a whole transmission system which is sufficient to produce unwanted waveform distortion.

overload capacity—The level of current, voltage, or power beyond which a device will be ruined. It is usually higher than the rated load capacity.

overload level—The level at which a system, component, etc., ceases to operate satisfactorily and produces signal distortion, overheating, damage, etc.

overload operating time—The length of time a system, component, etc., may be safely subjected to a specified overload current.

overload protection—A device which automatically disconnects the circuit whenever the current or voltage becomes excessive.

overload relay—A relay designed to operate when its coil current rises above a predetermined value.

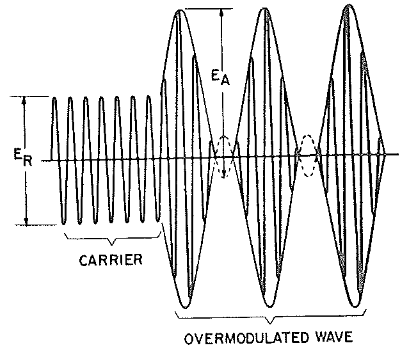

Overmodulation.

overmodulation—Modulation greater than 100%. Distortion occurs because the carrier is reduced to zero during certain portions of the modulating signal.

overpotential—Also called overvoltage. A voltage greater than the normal operating voltage of a device or circuit.

overpressure—Pressure greater than the full-scale rating of a pressure transducer.

overpunch—Also called zone punch. A hole punched in one of the three top rows of a punch card and which, in combination with a second hole in one of the nine lower rows, identifies an alphabetic or special character.

override—To manually or otherwise deliberately overrule an automatic control system or circuit and thereby render it ineffective.

overscanning—In a cathode-ray tube, the deflection of the electron beam beyond the normal limits of the screen.

overscan recovery—A characteristic of differential comparators that states the time required for the amplifier to recover to within some amount of voltage after a return to the screen of a crt.

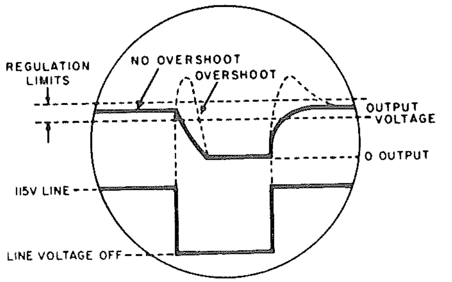

Overshoot.

overshoot—1. The initial transient response, which exceeds the steady-state response, to a unidirectional change in input. 2. Amplitude of the first maximum excursion of a pulse beyond the 100% amplitude level expressed as a percentage of this 100% amplitude. 3. A transient rise beyond regulated output limits, ocurring when the a-c power input is turned on or off, and for line or load step changes.

overshoot distortion—*See* Overthrow Distortion.

overshoot of an instrument—The amount which the indicator travels beyond its final steady deflection when a new constant value of the measured quantity is suddenly applied to the instrument. The overtravel and deflection are determined in angular measure, and the overshoot is expressed as a percentage of the change in steady deflection.

overtemperature protection—A thermal relay or other protective device which turns off the power automatically in the event of the occurrence of an overtemperature condition.

over-the-horizon transmission—*See* Scatter Propagation.

overthrow distortion—Also called overshoot distortion. The distortion that occurs in a signal wave when the maximum amplitude of the signal wavefront exceeds the steady-state amplitude.

overtone—1. A physical component of a complex sound having a frequency higher than that of the basic frequency. 2. A component of a complex tone having a pitch higher than that of the fundamental pitch. The term "overtone" has frequently been used in place of "harmonic," the *n*th harmonic being called the (*n*-1)st overtone. There is, however, ambiguity sometimes in the numbering of components of a complex sound when the word "overtone" is employed. Moreover, the word "tone" has many different meanings, so that it is preferable to employ terms which do not involve "tone" wherever possible.

overtone-type piezoelectric crystal unit—Also called harmonic-mode crystal unit. A crystal designed to be operated at an overtone of its resonant frequency.

overvoltage—1. The amount by which the applied voltage in a radiation-counter tube exceeds the Geiger-Mueller threshold. 2. *See* Overpotential.

overvoltage relay—A relay designed to operate when its coil voltage rises above a predetermined value.

O-wave—*See* Ordinary Wave.

Owen bridge—A four-arm, alternating-current bridge for measuring self-inductance in terms of capacitance and resistance. One arm, adjacent to the unknown inductor, comprises a capacitor and resistor in series. The arm opposite the unknown consists of a second capacitor, and the fourth arm is a resistor. Usually the bridge is balanced by adjusting the resistor in series with the first capacitor, and also the resistor in series with the inductor. The balance is independent of frequency.

oxalizing—*See* Surface Insulation.

oxidation—Commonly known as rust when ferrous material is involved. The increase in oxygen or in an acid-forming element or radical in a compound.

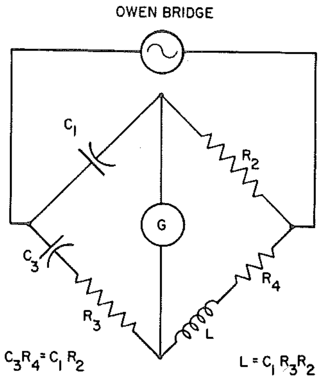

Owen bridge.

oxide (in magnetic recording)—Microscopic particles of ferric oxide dispersed in a liquid binder and coated on a recording-tape backing. These oxides are magnetically hard—i.e., once magnetized, they remain so permanently unless exposed to a strong magnetic field.

oxide-coated cathode—A cathode that has been coated with oxides of alkaline-earth metals to improve its electron emission at moderate temperatures.

oximeter—An instrument that determines the degree of oxygenation of the blood (e.g., in the ear lobe). It uses a photoelectric cell.

ozone—An extremely reactive form of oxygen, normally occurring around electrical discharges and present in the atmosphere in small but active quantities. It is faintly blue and has the odor of weak chlorine. In sufficient concentrations, it can break down certain rubber insulations under tension (such as a bent cable).

P

P—Symbol for power, primary winding, permeance, or the plate of an electron tube.

p—Abbreviation for the prefix "pico-" (10^{-12}).

pa—Abbreviation for picoampere.

PABX—Abbreviation for Private Automatic Branch Exchange. Has the same usage as a PBX except that calls within the system are completed automatically by dialing. An attendant at an attendant's board is required to route and complete incoming calls from the central office. Stations within the system are connected to the central office by dialing directly, or they are made to go through the attendant as company policy dictates.

pacemaker—Also called pacer. An electronic device for controlling rhythms of the heart or for initiating cardiac beatings, in event of cardiac arest. Includes implanted induction type and external type.

pacer—*See* Pacemaker.

pack—In computer programming, to combine several fields of information into one machine word.

packaged magnetron—An integral structure comprising a magnetron, its magnetic circuit, and its output matching device.

packaging—The physical process of locating, connecting, and protecting devices, components, etc.

packaging density—1. The number of devices or equivalent devices in a unit volume of a working system. 2. In a computer, the number of units of information per dimensional unit.

packing—Excessive crowding of carbon particles in a carbon microphone. The abnormal pressure of the particles lowers their resistance. As a result, the current increases excessively and fuses some of the particles together, further lowering the resistance and raising the current. Packing causes the sensitivity of the microphone to decrease.

packing density—In a digital computer, the number of units of desired information contained within a storage or recording medium.

packing factor—The number of pulses or bits of information that can be written on a given length of magnetic surface.

pack unit—A term applied to a compact, combination radio transmitter/receiver that can be carried or strapped on the back. Some pack units are popularly known as walkie-talkies.

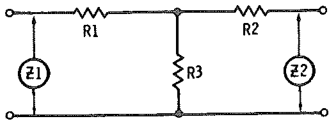

Pad (definition 2).

pad—1. A transducer capable of reducing the amplitude of a wave without introducing appreciable distortion. 2. An assembly of resistors that presents the proper input and output impedances to the circuits with which it is connected, and thereby maintains a fixed value of energy loss. 3. A device inserted into a circuit to introduce transmission loss or to match impedances. 4. *See* Land, 2.

padder—A series capacitor inserted into the oscillator tuning circuit of a superheterodyne receiver to control the calibration at the low-frequency end of the tuning range.

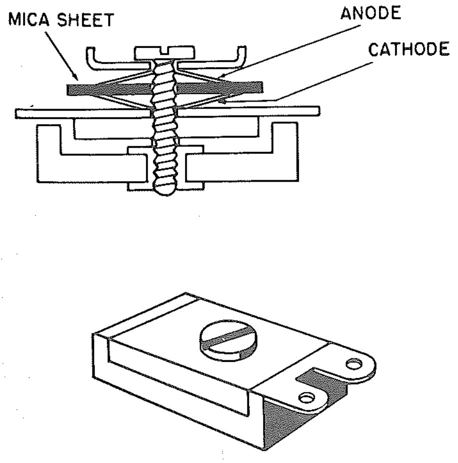

Padder capacitor.

pad electrode—One of a pair of electrode plates between which a load is placed for dielectric heating.

paint—Vernacular for a target image on a radarscope.

pair—In electric transmission, two like conductors employed to form an electric circuit.

paired cable—A cable in which all of the conductors are arranged in the form of twisted pairs, none of which is arranged with others to form quads.

pairing—In television, the imperfect interlace of lines comprising the two fields of one frame of the picture. Instead of being equally spaced, the lines appear in groups of two—hence the name.

Palmer scan—A combination of circular and conical scans. The beam is swung around the horizon at the same time the conical scan is performed.

PAM—Abbreviation for pulse-amplitude modulation.

pam/fm—Frequency modulation of a carrier by subcarriers modulated by pulses which, in turn, are modulated by data.

pam/fm/fm—Frequency modulation of a carrier by subcarriers modulated by pulses which, in turn, are modulated by data.

pan—1. To move a television or movie camera slowly up and down or across a scene to secure a panoramic effect. 2. To move the camera up and down, or back and forth, in order to keep it trained on a moving object.

pancake coil—A coil shaped like a pancake, usually with the turns arranged in a flat spiral.

panel—1. An electrical switchboard or instrument board. 2. A mounting plate of metal or insulation for the controls and/or other parts of equipment.

panel layout—The general physical layout of an electrical control panel with all relays, disconnect switch, control transformer, terminal strip, etc.

panning—*See* Pan.

panoramic adapter—An attachment used with a search receiver to provide, on an oscilloscope screen, a visual presentation of the band of frequencies extending above and below the center frequency to which the search receiver is tuned.

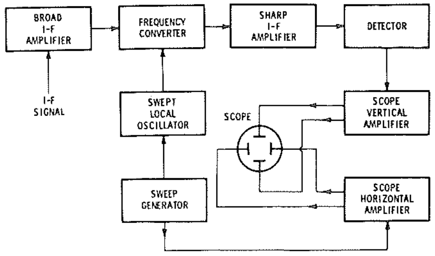

Panoramic adapter.

panoramic presentation—A presentation of signals as intensity pips (vertical deflections) along a line. The horizontal distance along the line represents frequency.

panoramic radar—A nonscanning radar which transmits signals omnidirectionally over a wide beam.

panoramic receiver—A radio receiver that displays, on the screen of a cathode-ray tube, the presence and relative strength of all signals within a wide frequency range. Used in communications for monitoring a wide band, locating open channels quickly, indicating intermittent signals or interference, and monitoring a frequency-modulated transmitter.

panoramic sonic analyzer—A heterodyne-type instrument which separates the frequency components of a complex waveform and displays them on an oscillographic screen, indicating both frequency and magnitude.

pan-range—Intensity-modulated A-type indication with slow vertical sweep applied to video. Stationary targets give solid vertical deflection, and moving targets give broken vertical deflection.

pantography—A system for transmitting and automatically recording radar data from an indicator to a remote point.

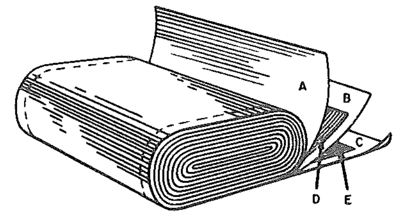

A-B-C PARAFINED PAPER
D-E METAL FOIL

Paper capacitor.

paper capacitor—A fixed capacitor consisting of two strips of metal foil separated by oiled or waxed paper or other insulating material, the whole rolled together into a compact roll. The foil strips can be staggered so that one strip projects from each end, or tabs can be added. The connecting wires are attached to the strips or tabs.

PAR—*See* Precision Approach Radar.

parabola—Locus of points equidistant from a fixed point and a straight line.

parabola controls—Sometimes called vertical-amplitude controls. The three controls in a color television receiver employing the magnetic-convergence principle. They are used for adjusting the amplitude of the parabolic voltages applied, at the vertical-scanning frequency, to the coils of the magnetic-convergence assembly.

parabolic antenna—An antenna with a radiating element and a parabolic reflector that concentrates the radiated power into a beam.

parabolic microphone—A microphone positioned at the focus of a parabolic sound reflector to give highly directional characteristics.

parabolic reflector—A metallic sheet formed so that its cross section is in the shape of a cylindrical parabola. The antenna elements are placed along the line that runs through the focal point of the parabola, parallel to the leading edge of the reflecting sheet.

parabolic-reflector microphone—A microphone employing a parabolic reflector for improved directivity and sensitivity.

paraboloid—The surface which a parabola generates when rotated about its axis of symmetry.

paraboloidal reflector—A hollow concave reflector which is a portion of a paraboloid of revolution.

paraffin—A vegetable wax having insulating properties.

parallax—An optical illusion which makes an object appear displaced when viewed from a different angle. Thus, a meter pointer will seem to be at different positions on the scale, depending from which angle it is read. To eliminate such errors, the eye should be directly above the meter pointer.

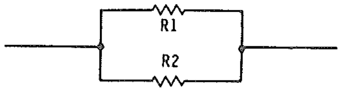

Parallel-connected resistors.

parallel—1. Also called shunt. Connected to the same pair of terminals, so that the current can branch out over two or more paths. 2. In electronic computers, the simultaneous transmission of, storage of, or logical operations on a character or other subdivision of a word, using separate facilities for the various parts.

parallel arithmetic unit—In a computer, a unit in which separate equipment operates (usually simultaneously) on the digits in each column.

parallel circuit—A circuit in which all positive terminals are connected to a common point, and all negative terminals are connected to a second common point. The voltage is the same across each element in the circuit.

parallel connection—Also called shunt connection. Connection of two or more parts of a circuit to the same pair of terminals, so that current divides between the parts—as contrasted with a series connection, where the parts are connected end to end so that the same current flows through all.

parallel cut—A Y-cut in a crystal.

parallel digital computer—A computer in which the digits are handled in parallel. Mixed serial and parallel machines are frequently called serial or parallel according to the way arithmetic processes are performed. For example, a parallel digital computer handles decimal digits in parallel, although the bits which comprise a digit might be handled either serially or in parallel.

parallel feed—Also called shunt feed. Application of a d-c voltage to the plate or grid of a tube in parallel with an a-c circuit, so that the d-c and a-c components flow in separate paths.

parallel gap welding—A method of resistance welding in which both electrode tips are in close proximity to each other, being separated by a small gap or insulating material, approach the work from the same direction, and contact only one of the two materials being welded.

paralleling reactor—A reactor for correcting the division of load between parallel-connected transformers with unequal impedance voltages.

parallelogram distortion—In camera or image tubes, a form of distortion which amounts to a skewing of the reproduced image laterally across the crt face.

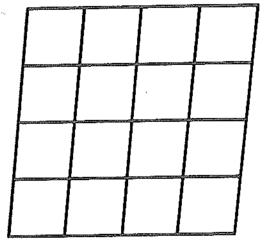

Parallelogram distortion.

parallel operation—1. Also called master/slave operation. The connecting of two or more power supplies so that their outputs are tied together, permitting the accumulated flow of current from all units to a common load. In regulated power supplies, interconnections other than the output terminals themselves may be required. For example, the amplifiers of all units but one may be made inoperable, and this single amplifier would control all regulating elements. Such operation is denoted. 2. Pertaining to the manipulation of information within the computer circuitry, in which the digits of a word are transmitted simultaneously on separate lines. Faster than serial operation, but requires more equipment.

parallel-plate oscillator—A push-pull, ultrahigh-frequency oscillator circuit that uses two parallel plates as the main frequency-determining elements.

parallel-plate waveguide—A pair of parallel conducting planes for propagating uniform cylindrical waves that have their axes normal to the plane.

parallel processing—In a computer, the processing of more than one program at a time through more than one active processor.

parallel recording—A technique in which the record heads in a head stack are energized simultaneously to record a specific set of bits.

parallel resonance—In a circuit comprising inductance and capacitance connected in parallel, the steady-state condition that exists when the current entering the circuit from the supply line is in phase with the voltage across the circuit.

parallel-resonant circuit—1. A resonant circuit in which the applied voltage is connected across a parallel circuit formed by a capacitor and an inductor. 2. An inductor and capacitor connected in parallel to furnish a high impedance at the frequency to which the circuit is resonant.

parallel-rod oscillator—An ultrahigh-frequency oscillator circuit in which the tank circuits are formed by parallel rods or wires.

parallel-rod tank circuit—A tank circuit consisting of two parallel rods connected at their far ends. This is done to provide the small values of inductance and capacitance in parallel required for ultrahigh-frequency circuits.

parallel-rod tuning—A tuning method sometimes used at ultrahigh frequencies. The transmitter, receiver, or oscillator is tuned by sliding a shorting bar back and forth on two parallel rods.

parallel-series circuit—Also called shunt-series circuit. Two or more parallel circuits connected together in series.

parallel storage—Computer storage where all bits, words, or characters are equally available.

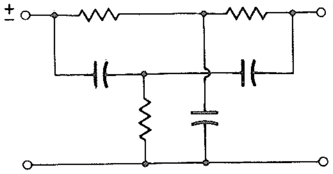

Parallel-T network.

parallel-T network—Also called twin-T network. A network composed of separate T-networks (usually two), the terminals of which are connected in parallel.

parallel-T oscillator—An RC sine-wave oscillator which provides phase inversion at one discrete frequency and is so connected that positive feedback results only when phase inversion occurs.

parallel transfer—Data transfer where all characters of a word are transferred simultaneously over a set of lines.

parallel transmission—In a computer, the system of information transmission in which the characters of a word are transmitted (usually simultaneously) over separate lines, as contrasted with serial transmission.

parallel-wire line—A transmission line consisting of two wires a fixed distance apart.

parallel-wire resonator—A resonator circuit consisting of two parallel wires connected at one end to the oscillator tube or transistor. The other end is short circuited and can be adjusted for the desired frequency.

Paraloc—A phase-shift amplifier the positive feedback of which results in an oscillation that is frequency-modulated by the change in a sensed variable. The sensor is one arm of a resistance bridge.

paramagnetic—Having a magnetic permeability greater than that of a vacuum but less than that of ferromagnetic materials. Unlike the latter, the permeability of paramagnetic materials is independent of the magnetizing force.

paramagnetic material—A material having a permeability which is slightly greater than that of a vacuum and which is approximately independent of the magnetizing force.

paramagnetism—Magnetism that involves a permeability somewhat greater than unity.

parameter—1. A constant or element, the value of which characterizes the behavior of one or more variables associated with a given system. 2. A measured value which expresses performance.

parameter extraction—A technique that reduces the bandwidth required to transmit a given data sample by means of an information-describing irreversible transformation. These transformations are considered irreversible in that, while they provide useful descriptions of the input signal, they so distort the signal that it is impossible to reconstruct the original waveform.

parametric amplifier—A low-noise device for amplifying signals in the uhf and microwave regions of the electromagnetic spectrum. The essential element of the amplifier is a semiconductor crystal (varactor). By its very nature the diode contributes very little noise, unlike an amplifier utilizing electron tubes, where the random nature of the electron emission from the hot cathode produces large fluctuations in the current, called shot noise. Thus the signals to be amplified can be of extremely low levels.

parametric converter—An inverting or noninverting parametric device used to convert an input signal at one frequency into an output signal at a different frequency.

parametric device—A device the operation of which depends essentially upon the time variation of a characteristic parameter usually understood to be a reactance.

parametric down-converter—A parametric converter in which the output signal is at a lower frequency than the input signal.

parametric excitation—A term referring to the method of exciting and maintaining oscillation in either an electrical or mechanical dynamic system, in which excitation results from a periodic variation in an energy storage element in a system such as a capacitor, inductor, or spring constant.

parametric frequency converter—A frequency converter that utilizes the variation of the reactance parameter of an energy-storage element for frequency conversion.

parametric modulator—A modulator that utilizes the variation in the reactance parameter of an energy-storage element to produce modulation.

parametric up-converter—A parametric converter in which the output signal is at a higher frequency than the input signal.

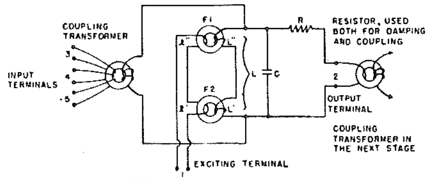

(A) Magnetic types.

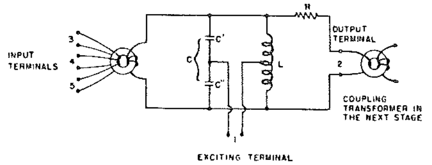

(B) Capacitive types.

Parametrons.

parametron—A digital circuit element utilizing the principle of parametric excitation. It is essentially a resonant circuit with a nonlinear reactive element which oscillates at half the driving frequency. The oscillation can be made to represent a binary digit by the choice between two stationary phases π radians apart.

paramistor—A digital logic-circuit module containing several parametron elements.

paraphase amplifier—An amplifier which converts a single input signal into two out-of-phase signals for driving a push-pull stage.

parasite—Current in a circuit, due to some unintentional cause such as inequalities of temperature or of composition; particularly troublesome in electrical measurement.

parasitic—An undesired low- or high-frequency signal in an electronic circuit.

parasitic antenna—An antenna that is excited by radiation from other antennas rather than by electrical connection with them.

parasitic arrays—An antenna array containing one or more elements not connected to the transmission line.

parasitic component—In a monolithic integrated circuit, the capacitors and diodes which are formed between the planned circuit elements and the substrate during processing. The circuit design must allow for the functional effects of these parasitic components.

parasitic element—Also called passive element. An antenna element (i.e., reflector, director, etc.) not connected to the transmission line or to any driven element. A parasitic element affects the gain and directivity pattern of an antenna, and also acts

on a driven element by absorbing and re-turning energy from it. In a dipole reflector combination, the reflector is the parasitic element.

parasitic oscillations—Also called parasitics. Unintended self-sustaining oscillations, or transient impulses.

parasitics—*See* Parasitic Oscillations.

parasitic suppressor—A parallel resistance, or a parallel combination of inductance and resistance, inserted into a grid or plate circuit to suppress parasitic oscillations.

pard—In an electronic power conversion unit, the Periodic And Random Deviation of the output d-c voltage, current, or power from its average value, with all external operational and environmental parameters maintained constant. The load impedance in particular must be held constant. Perturbations in the output which are induced by load impedance changes are dynamic load regulation or transient response. Pard may be defined as rms pard or peak-to-peak pard.

parent population—Prototype or initial group of the articles under consideration.

parity check—A computer checking method in which the total number of binary 1's (or 0's) is always even or always odd. Either an even-parity or odd-parity check can be made.

part—1. The smallest subdivision of a system. 2. An item which cannot ordinarily be disassembled without destruction.

part failure—A breakdown that cannot be repaired and which ends the life of a part.

part-failure rate—The number of occasions, during a specified time period, on which a given quantity of identical parts will not function properly.

partial—1. A physical component of a complex tone. 2. A component of a sound sensation that can be distinguished as a simple tone which cannot be further analyzed by the ear and which contributes to the character of the complex sound.

partial node—The place in a standing-wave system at which some characteristic of the wave field has a minimum amplitude other than zero.

partial-read pulse—In a computer, any one of the applied currents which cause selection of a core for reading.

partial-select output—In a computer, the voltage response of an unselected magnetic cell produced by the application of partial-read pulses or partial-write pulses.

partial-write pulse—In a computer, any one of the applied currents which cause a core to be selected for writing.

particle—An infinitesimal subdivision of matter—e.g., a molecule, atom, or electron.

particle accelerator—Any device for accelerating charged particles to high energies (e.g., cyclotron, betatron, Van de Graaff generator, linear accelerator, etc.).

particle velocity—The velocity of a given infinitesimal part of a sound wave. The most common unit is centimeter per second.

partitioning—Also called segmenting. In a computer, subdividing a large block into smaller, more conveniently handled subunits.

partition noise—A noise caused in an electron tube by random fluctuations as the electron stream divides between the electrodes. It is more pronounced in pentodes and tetrodes than in triodes.

part programmer—One who translates the physical operations for machining a part into a series of mathematical steps and then prepares the coded computer instructions for those steps.

parts density—The number of parts in a unit volume.

party line—A telephone line serving more than one subscriber, with discriminatory ringing for each.

Paschen's law—The sparking potential between two terminals in a gas is proportional to the pressure times the spark length. For a given voltage, this means the spark length is inversely proportionate to the pressure.

passband—The band of frequencies which will pass through a filter with essentially no attenuation.

passband ripple—In a filter, the difference, in db, between the minimum loss point and the maximum loss point in a specified bandwidth.

pass element—An automatic variable resistance device, either a vacuum tube or power transistor, in series with the source of d-c power. The pass element is driven by the amplifier error signal to increase its resistance when the output needs to be lowered or to decrease its resistance when the output must be raised. *Also see* Series Regulator.

passivation—The growth of an oxide layer on the surface of a semiconductor to provide electrical stability by isolating the transistor surface from electrical and chemical conditions in the environment. This reduces reverse-current leakage, increases breakdown voltages, and raises the power dissipation rating.

passive—An inert component which may control, but does not create or amplify energy.

passive communication satellite—A communication satellite which simply reflects a signal without amplification. In essence, it is a radio mirror. It requires a large reflecting surface and large, high-powered, complex ground stations. Abbreviated passive comsat.

passive component—A nonpowered component generally presenting some loss (expressed in db) to a system.

passive comsat—*See* Passive Communications Satellite.

passive decoder—A device that is set so that only one specific reply code will pass a decoder and give an output from one decoder for display.

passive detection—Detection of a target without revealing the position of the detector.

passive device—A device which exhibits no transistance. It has no gain or control and does not require any input other than a signal to perform its function. Examples of passive devices are conductors, resistors, and capacitors.

passive electric network—An electric network with no source of energy.

passive element—1. A parasitic element. 2. A circuit element with no source of energy (e.g., a resistor, capacitor, inductor, etc.).

passive film circuit—A thin- or thick-film circuit network consisting entirely of passive circuit elements and interconnections.

passive homing system—A guidance system based on the sensing of energy radiated by the target. (*Also see* Active Homing *and* Homing Guidance System.)

passive network—A network with no source of energy.

passive reflector—A reflector often used on microwave relay towers to change the direction of a microwave. This permits convenient location of transmitter, repeater, and receiver equipment on the ground rather than at the tops of towers.

passive sonar—*See* Sonar.

passive substrate—A substrate that may serve as a physical support and thermal sink for a thin-film integrated circuit but does not exhibit transistance. Examples of passive substrates are glass, ceramic, etc.

passive tracking system—Usually a system that tracks by reflected radiation from some external source, or by the jet emission of the vehicle (e.g., optical systems, use of commercial radio or television, reflection and infrared systems).

passive transducer—A transducer, the output waves of which are independent of any sources of power controlled by the actuating waves.

paste—In batteries, the medium in the form of a paste or jelly, containing an electrolyte. It is positioned adjacent to the negative electrode of a dry cell. In an electrolytic cell, the post serves as one of the conducting plates.

pa system—Abbreviation for public-address system.

patch—1. To connect circuits together temporarily with a special cord known as a patch cord. 2. In a computer, to make a change or correction in the coding at a particular location by inserting transfer instructions at that location and by adding elsewhere the new instructions and the replaced instructions. This procedure is usually used during checkout. 3. The section of coding so inserted.

patch board—A board or panel where circuits are terminated in jacks for patch cords.

patch cord—Sometimes called an attachment cord. A short cord with a plug or a pair of clips on one end, for conveniently connecting two pieces of sound equipment such as a phonograph and tape recorder, an amplifier and speaker, etc.

Patch cord.

patching—Connecting two lines or circuits together temporarily by means of a patch cord.

patching jack—A jack for interconnection of circuit elements.

path—In navigation, an imaginary line connecting a series of points in space and constituting a proposed or traveled route.

pattern—1. The means of specifying the character of a wave in a guide. This is done by showing the loops of force existing in the guide for that wave. The pattern identifies the order and mode of the wave and the cross-sectional shape of the guide. 2. A geometrical figure representing the directional qualities of an antenna array.

pattern recognition—In a computer, the examination of records for certain code-element combinations.

pause control—A feature of some tape recorders making it possible to temporarily stop the movement of the tape without switching the machine from the "play" or "record" position. (Essentially for a tape recorder used for dictation and generally helpful for editing purposes.)

PAX—Acronym for Private Automatic Exchange. An automatic system used exclusively for interoffice dial communications and having no trunks to the central office.

pay television—Also called subscription television. A system whereby viewers must insert coins or record cards into a decoding device in order to view a television program that has been deliberately scrambled to prevent unpaid viewing.

P-band—A radio-frequency band extending from 225 to 390 MHz and having a wavelength from 133.3 to 76.9 cm.

PBX—Abbreviation for Private Branch Exchange. A manual telephone system located on the premises of a business and requiring an attendant to complete all calls. It is usually owned by the telephone company and is equipped with trunks to a telephone-company central office.

pc—1. Abbreviation for picocoulomb. 2. Abbreviation for picocurie.

pcm—Abbreviation for pulse-code modulation.

pcm/fm—Frequency modulation of a carrier by pulse code modulated information.

pcm/fm/fm—Frequency modulation of a carrier by subcarrier(s) which is (are) frequency modulated by pulse code modulated information.

pcm level—The number by which identification of a given quantized-signal subrange may be made.

pcm/pm—Phase modulation of a carrier by pulse code modulated information.

P-display—*See* Plan-Position Indicator.

pdm—Abbreviation for pulse-duration modulation.

pdm/fm—Frequency modulation of a carrier by a subcarrier(s) which is (are) frequency modulated by pulses which are modulated in duration by information.

pdm/fm/fm—Frequency modulation of a carrier by subcarrier(s) which is (are) frequency modulated by pulses which are time duration modulated by information.

pdm/pm—Phase modulation of a carrier by pulses which are duration modulated by information.

peak—Also called crest. 1. A momentary high amplitude level. 2. The maximum instantaneous value of a quantity.

peak alternating gap voltage—In a microwave tube, the negative of the line integral of the peak alternating electric field, taken along a specified path across the gap.

peak amplitude—The maximum deviation (e.g., of a wave) from an average or mean position.

peak-anode current—The maximum instantaneous value of an anode current in an electron tube.

peak-cathode current (fault)—The highest instantaneous value of a nonrecurrent pulse of cathode current occurring under fault conditions.

peak-cathode current (steady-state)—The maximum instantaneous value of a periodically recurring cathode current.

peak-cathode current (surge)—The highest instantaneous value of a randomly recurring pulse of cathode current.

peak-charge characteristic—The function giving the relation of one half the peak-to-peak value of transferred charge in the steady state to one half the peak-to-peak value of a specified symmetrical alternating voltage applied to a nonlinear capacitor.

peak current—The maximum current which flows during a complete cycle.

peak distortion—The largest total distortion of signals noted during a period of observation.

peak electrode current—The maximum instantaneous current that flows through an electrode.

peak-envelope power of a radio transmitter—The average power supplied to the antenna transmission line by a transmitter during one radio-frequency cycle at the highest crest of the modulation envelope, taken under conditions of normal operation.

peak flux density—The maximum flux density in a magnetic material.

peak forward anode voltage—The maximum instantaneous anode voltage in the direction the tube is designed to pass current.

peak forward-blocking voltage—The maximum instantaneous value of repetititve positive voltage that may be applied to the anode of an scr with its gate circuit open.

peak forward drop—The maximum instantaneous voltage drop measured when a tube or rectifier cell is conducting forward current, either continuously or during transient operation.

peaking—Adjusting a component so as to increase the response of a circuit at a desired frequency or band of frequencies.

peaking circuit—A circuit capable of converting an input wave into a peaked waveform.

peaking control—In a television receiver, a fixed or variable resistor-capacitor circuit which controls the negative shape of the pulses originating at the horizontal oscillator. This is done to assure a linear sweep.

peaking network—A type of interstage coupling network used to increase the amplification at the upper end of the frequency range. It consists of an inductance effectively in series (series peaking network) or shunt (shunt peaking network) with a parasitic capacitance.

peaking resistor—A resistor placed in series with the charging capacitor of the vertical sawtooth generator. By adding a negative peaking pulse to the sawtooth voltage, it creates the waveform required to produce a linear sawtooth current in the yoke.

peaking transformer—A transformer having a core designed to saturate at relatively low values of primary current.

peak inverse anode voltage—The maximum instantaneous anode voltage in the direction opposite from that in which the tube is designed to pass current.

peak inverse voltage—The peak a-c voltage which a rectifying cell or PN junction will withstand in the reverse direction.

peak level—The maximum instantaneous level that occurs during a specific time interval (i.e., in acoustics, the peak sound pressure level).

peak limiter—A device which automatically limits the magnitude of its output signal to approximate a preset maximum value by reducing its amplification when the instantaneous signal magnitude exceeds a preset value.

peak load—The maximum electrical power load consumed or produced in a stated period of time. It may be the maximum instantaneous load or the maximum average load over a designated interval of time.

peak magnetizing force — The upper or lower limiting value of a magnetizing force.

peak-or-valley readout memory—A circuit in which the output remains at the condition corresponding to the most positive (least negative) or vice versa input signal since the circuit was set to initial conditions, until reset to those conditions.

peak plate current—The maximum instantaneous current passing through the plate circuit of a tube.

peak power—1. The mean power supplied to the antenna of a radio transmitter during one radio-frequency cycle at the highest crest of the modulation envelope. 2. The maximum power of the pulse from a radar transmitter. Since the resting time of a radar transmitter is longer than its operating time, the average power output is much lower than the peak power.

peak power output—The output power averaged over the radio-frequency cycle and having the maximum peak value that can occur under any combination of signals transmitted.

peak pulse amplitude—The maximum absolute peak value of the pulse, excluding unwanted portions such as spikes.

peak pulse power—The maximum power of a pulse, excluding spikes.

peak response—The maximum response of a system to an input.

peaks—1. A momentarily high amplitude occurring in electronic equipment. 2. A momentarily high volume level during a radio program. It causes the volume indicator at the studio or transmitter to swing upward.

peak signal level—An expression of the maximum instantaneous signal power or voltage as measured at any point in a facsimile transmission system. This includes auxiliary signals.

peak sound pressure—The maximum absolute value of instantaneous sound pressure for any specified time interval. The most common unit is the microbar.

peak speech power—The maximum instantaneous speech power over the time interval considered.

peak to peak—The algebraic difference between the positive and negative maximum values of a waveform.

peak-to-peak amplitude—The amplitude of an alternating quantity, measured from positive peak to negative peak.

peak-to-peak voltmeter—A voltmeter which indicates the over-all difference between the positive and negative voltage peaks.

peak value—Also called crest value. The maximum instantaneous value of a varying current, voltage, or power. For a sine wave, it is equal to 1.414 times the effective value of the sine wave.

peak voltage—The maximum value present in a varying or alternating voltage. This value may be either positive or negative.

peak voltmeter—A voltmeter that reads peak values of an alternating voltage.

pea lamp—An incandescent lamp with a bulb about the size of a pea. Its small size makes it ideal for use by doctors, on instrument panels, and in small flashlights.

pedal clavier — In an organ, the pedal keyboard which supplies the bass accompaniment for the other manuals.

pedal keyboard—*See* Pedal Clavier.

pedestal—1. A substantially flat-topped pulse which elevates the base level for another wave. 2. The base of a radar antenna.

pedestal level—*See* Blanking Level.

pedestal pulse—A square-wave pulse or gate on which a video signal or sweep voltage may be superimposed.

peek-a-boo—In a computer, a method of determining the presence or absence of holes in identical locations on punched cards by placing one card on top of another. (*Also see* Batten System.)

peel-strength adhesion—*See* Bond Strength.

Peltier coefficient—The quotient of the rate of Peltier heat absorption by the junction of two dissimilar conductors divided by the current through the junction. The Peltier coefficient of a couple is the algebraic difference between either the relative or absolute Peltier coefficients of the two conductors making up the couple.

Peltier effect—The production or absorption of heat at the junction of two metals when current is passed through the junction. Heat generated by current in one direction will be absorbed when the current is reversed.

Peltier electromotive force—The component of voltage produced by a thermocouple after being heated by the Peltier effect at the junction of the different metals. It adds to the Thomson electromotive force to produce the total voltage of the thermocouple.

Peltier heat — The thermal energy absorbed or produced as a result of the Peltier effect.

pen centering—An electrical or mechanical adjustment by which an oscillograph pen is positioned to channel center.

pencil beam—A radar beam in which the energy is confined to a narrow cone.

pencil-beam antenna—A unidirectional antenna in which those cross sections of the major lobe perpendicular to the maximum radiation are approximately circular.

pencil tube—A small tube designed for operation in the ultrahigh-frequency band and used as an oscillator or r-f amplifier.

pendant—The type of plug and/or receptacle that is not mounted in a fixed position or attached to a panel or side of equipment.

pendant station—A push-button station suspended from overhead and connected by means of flexible cord or conduit, but supported by a separate cable.

pendulous accelerometer—A device which measures linear accelerations by means of a restrained unbalanced mass. Two pivots and jewels support the unbalanced gimbal, and the torsion bar functions as the spring.

penetrating frequency — *See* Critical Frequency.

Penning discharge—A type of discharge in which electrons are forced to oscillate between two opposed cathodes and are prevented from going to the surrounding anode by the presence of a magnetic field. It is sometimes referred to as a pig discharge because the device producing it was first used as an ionization gage called the Penning Ionization Gage.

Penning Ionization Gage—*See* Penning Discharge.

pen position—An electrical or mechanical adjustment by which an oscillograph pen is positioned to any desired amplitude grid mark on the chart to represent zero signal.

pent—Abbreviation for pentode.

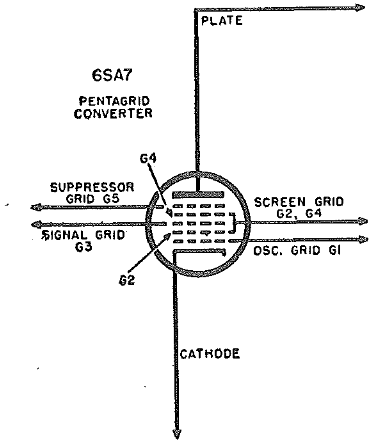

Pentagrid converter.

pentagrid converter—A pentagrid tube used as a combination oscillator, mixer, and first detector in a superheterodyne receiver.

pentagrid mixer—A pentagrid tube used to mix the r-f and local-oscillator signals in a superheterodyne receiver.

pentagrid tube—An electron tube having five grids, plus an anode and a cathode.

pentatron—A five-electrode vacuum tube that provides push-pull amplification with a single tube. It has one cathode, two grids, and two anodes. In effect, it is two tubes in one.

pentode—A five-electrode electron tube containing an anode, a cathode, a control electrode, and two grids.

pentode field-effect transistor—A five-lead transistor with three gates. It can be connected like a pentode if each of the gates is supplied from an independent bias source.

percentage-differential relay—A differential relay which functions when the difference between two quantities of the same nature exceeds a fixed percentage of the smaller quantity. This term includes relays formerly known as ratio-balance relays, biased relays, and ratio differential relays.

percentage of meter accuracy—The ratio of the actual meter reading to the true reading, expressed as a percent.

percentage sync—Ratio of the amplitude of the synchronizing signal to the peak-to-peak amplitude of the picture signal between blanking and reference white level, expressed in percent.

percentage timers—A repeat-cycle nonreset timer with fixed cycle length, having a dial adjustment of the percentage of the cycle time for which the contacts are operated.

percent modulation meter—An instrument which indicates the modulation percentage of an amplitude-modulated signal, either on a meter or a cathode-ray tube.

percent of deafness—*See* Percent of Hearing Loss.

percent of harmonic distortion—A measure of the harmonic distortion in a system or component. Each harmonic frequency is equal to 100 times the ratio of the square root of the sum of the squares of the root-mean-square voltages (or currents), to the root-mean-square voltage (or current) of the fundamental.

percent of hearing—At a given frequency, 100 minus the percent of hearing loss at that frequency.

percent of hearing loss—Also called percent of deafness. At a given frequency, 100 times the ratio of the hearing loss in decibels, to the number of decibels between the normal threshold levels of audibility and feeling.

percent of modulation—1. In am, the ratio of half the difference between the maximum and minimum amplitudes of a wave to the average amplitude, expressed in percentage. 2. In f-m and tv audio transmission, the ratio of the actual frequency swing, to the frequency swing defined as 100% modulation, expressed in percentage. For f-m broadcast stations, a frequency swing of ±75 kHz is defined as 100% modulation. For television, it is ±25 kHz.

percent of ripple voltage—Ratio of the effective (root-mean-square) value of the ripple voltage to the average value of the total voltage, expressed in percent.

percent of syllabic articulation—*See* Syllable Articulation.

percent ripple (rms)—The ratio of the effective (rms) value of the ripple voltage to the average value of the total voltage, expressed in percent. The new term is pard.

perceptron—A system capable of—either in theory or in practice—performing knowledgeable functions such as recognition, classification, and learning. These functions may exist as mathematical analyses, computer programs, or "hardware."

percussion—Musical sounds characterized by sudden or sharp transients. Organ percussion is achieved by causing the tone to start to decay the instant it is played rather than waiting until the key is released.

percussive welding — An electrical arc method of welding studs and cable ends to plates and terminals. The weld is accomplished by storing electrical energy in capacitors, then releasing it to produce a high current density at the stud or cable tip. This ionizes an area which in turn produces an arc that melts the tip and allows the stud to be pressed against the plate for a welded union.

perfect dielectric—Also called ideal dielectric. A dielectric in which all the energy required to establish the electric field in it is returned to the electric system when the field is removed. A perfect dielectric has zero conductivity and exhibits no absorption phenomena. A vacuum is the only known perfect dielectric.

perforated tape—*See* Punched Tape.

performance—Degree of effectiveness of operation.

performance characteristic—A characteristic measurable in terms of some useful denominator—e.g., gain, power output, etc.

period—The time required for one complete cycle of a regular, repeating series of events.

periodic—Repeating itself regularly in time and form.

periodic antenna—An antenna in which the input impedance varies as the frequency does (e.g., open-end wires and resonant antennas).

periodic current—Oscillating current, the values of which recur at equal time intervals.

periodic damping—Damping in which the pointer of an instrument oscillates about the final position before coming to rest. The point of change between periodic and aperiodic damping is called critical damping.

periodic duty—Intermittent duty where the load conditions recur at regular intervals.

periodic electromagnetic wave—A wave in which the electric field vector is repeated in detail—either at a fixed point, after a lapse of time known as the period; or at a fixed time, after the addition of a distance known as the wavelength.

periodicity—The variations in the insulation diameter of a transmission cable that result in reflections of a signal, when its wavelength or a multiple thereof is equal to the distance between two diameter variations.

periodic law—When chemical elements are arranged in the ascending or descending order of their atomic number, their properties will occur in cycles.

periodic line—A line consisting of identical, similarly oriented sections, each section having nonuniform electrical properties throughout.

periodic pard—Pertains to that portion of the total pard, in an electronic power supply, the frequency of which is identical or harmonically related to the input frequency and/or intentionally internally generated signal frequencies. This phenomenon is frequently referred to as "ripple."

periodic pulse train—A pulse train made up of identical groups of pulses repeated at regular intervals.

periodic quantity—An oscillating quantity in which any value it attains is repeated at equal time intervals.

periodic rating—The load which can be carried for the alternate periods of load and rest specified in the rating without exceeding the specified heating limits.

periodic resonance—Also called natural resonance. Resonance in which the applied agency maintaining the oscillation has the same frequency as the natural period of oscillation of a system.

periodic time — *See* Period of an Underdamped Instrument.

periodic vibration — A vibration having a regularly recurring waveform, e.g., sinusoidal vibration.

periodic wave—A wave in which each point is displaced at regular time intervals.

period of an underdamped instrument — Also called periodic time. The time required, following an abrupt change in the measurand, for the pointer or other indicating means to make two consecutive transits in the same direction through the rest position.

peripheral control unit—An intermediary control device which links a peripheral unit to the central processor, or, in the case of off-line operation, to another peripheral unit.

peripheral electron—Also called a valence electron. One of the outer electrons of an atom. Theoretically, it is responsible for visible light, thermal radiation, and chemical combination.

peripheral equipment—Units which work in conjunction with a computer but are not part of it (e.g., tape reader, analog-to-digital converter, typewriter, etc.).

permalloy — A high-permeability magnetic alloy composed mainly of iron and nickel.

permanent echo—A radar echo from a fixed target.

permanent-field synchronous motor — A type of synchronous motor in which the member carrying the secondary laminations and windings also carries permanent-magnet field poles that are shielded from the alternating magnetic flux by the laminations. It behaves as an induction motor when starting but runs at synchronous speed.

permanent magnet—A piece of hardened steel or other magnetic material which has

been so strongly magnetized that it retains the magnetism indefinitely.

permanent-magnet centering — Vertical or horizontal shifting of a television picture by means of magnetic fields from permanent magnets mounted around the neck of the picture tube.

permanent-magnet focusing — Focusing of the electron beam in a television picture tube by means of one or more permanent magnets located around the neck.

permanent-magnet material — Ferromagnetic material which, once having been magnetized, resists external demagnetizing forces (i.e., requires a high coercive force to remove the magnetism).

permanent-magnet, moving-coil instrument—Also called D'Arsonval instrument. An instrument in which a reading is produced by the reaction between the current in a movable coil or coils and the field of a fixed permanent magnet.

permanent-magnet, moving-iron instrument—Also called polarized-vane instrument. An instrument in which a reading is produced by an iron vane as it aligns itself in the magnetic field produced by a permanent magnet and by the current in an adjacent coil of the instrument.

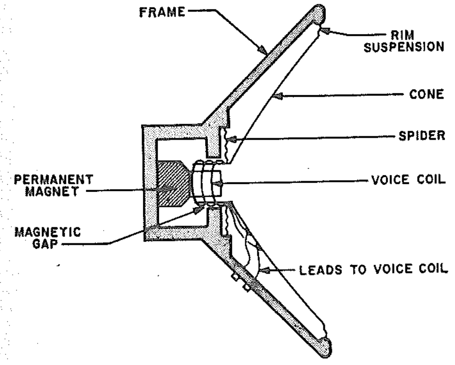

Permanent-magnet speaker.

permanent-magnet speaker—A moving-conductor speaker in which the steady magnetic field is produced by a permanent magnet.

permanent magnistor—A saturable reactor which has the properties of memory and the ability to handle appreciable power.

permanent-memory computer—A computer in which the stored information remains intact, even after the power has been turned off.

permanent storage—A computer storage device which retains the stored data indefinitely.

permatron—A thermionic gas diode, the discharge of which is controlled by an external magnetic field. It is used mainly as a controlled rectifier and functions like a thyratron.

permeability—Symbolized by the Greek letter *mu* (μ). The measure of how much better a given material is than air as a path for magnetic lines of force. (Air is assumed to have a permeability of 1.) It is equal to the magnetic induction (B) in gausses, divided by the magnetizing force (H) in oersteds.

permeability tuning—A method of tuning a circuit by moving a magnetic core into or out of a coil to vary its inductance.

permeameter—An apparatus for determining the magnetizing force and flux density in a test specimen. From these values, the normal induction curves or hysteresis loops can then be plotted and the magnetic permeability computed.

permeance — The reciprocal of reluctance. Through any cross section of a tubular portion of a magnetic circuit bounded by lines of force and by two equipotential surfaces, permeance is the ratio of the flux to the magnetic potential difference between the surfaces under consideration. In the cgs system, it is equal to the magnetic flux (in maxwells) divided by the magnetomotive force (in gilberts).

permissive control device—Generally a two-position, manually operated switch which in one position permits the closing of a circuit breaker or the placing of an equipment into operation, and in the other position prevents the circuit breaker or the equipment from being operated.

permittivity—*See* Dielectric Constant.

peroxide of lead—A lead compound that forms the principal part of the positive plate in a charged lead-acid cell.

perpendicular magnetization—In magnetic recording, magnetization that is perpendicular to the line of travel and parallel to the smallest cross-sectional dimension of the medium. Either single- or double-pole-piece magnetic heads may be used.

persistence—The length of time a phosphor dot glows on the screen of a cathode-ray tube before going out—i.e., the length of time it takes to decay from initial brightness (reached during fluorescence) until it can no longer be seen.

persistence characteristic (of a luminescent screen)—Also called the decay characteristic. The relationship (usually shown by a graph) between the time a luminescent screen is excited and the time it emits radiant power.

persistence of vision — The phenomenon whereby the eye retains an image for a short time after the field of vision has disappeared.

persistent current—A current that is magnetically induced and flows undiminished in a superconducting material or circuit.

persistor—A bimetallic circuit used for storage or readout in a computer. It is operated near absolute zero, and changes from a resistive to a superconductive state at a critical current value.

persistron—A device in which electroluminescence and photoconductivity are combined into a single panel capable of producing a steady or persistent display with pulsed signal input.

perveance—The space-charge-limited cathode current divided by the three-halves power of the anode voltage in a diode.

petticoat insulator—An insulator having an outward flaring lower part that is hollow to increase the length of the surface leakage path and keep part of the path dry at all times.

pf—Abbreviation for picofarad.

phanotron—A term used primarily in industrial electronics to mean a hot-cathode gas diode.

phantastron—An electronic circuit of the multivibrator type which is normally used in the monostable form. It is a stable trigger generator in this connection and is used in radar systems for gating functions and sweep-delay functions.

phantom channel—In a stereo system, an electrical combination of the left and right channels fed to a third, centrally located speaker.

phantom circuit—A superimposed circuit derived from two suitably arranged pairs of wires called side circuits. Each pair of wires is a circuit itself, and at the same time acts as one conductor of the phantom circuit.

phantom-circuit loading coil—A loading coil that introduces the desired amount of inductance into a phantom circuit and a minimum amount into the constituent side circuits.

phantom-circuit repeating coil—A repeating coil used at a terminal of a phantom circuit in the terminal circuit extending from the midpoints of the associated side-circuit repeating coils.

phantom coil—A coil originally used in a phantom circuit for impedance matching. Now, generally, any coil, side or phantom, in a phantom circuit. When the term is used, the meaning should be made clear.

phantom group—1. A group of four open-wire conductors suitable for the derivation of a phantom circuit. 2. Three circuits which are derived from simplexing two physical circuits to form a phantom circuit.

phantom repeating coil—A side-circuit repeating coil or a phantom-circuit repeating coil, when discrimination between these two types is not necessary.

phantom signals—Signals appearing on the screen of a cathode-ray tube indicator; their cause cannot readily be determined, and they may be due to circuit fault, interference, propagation and anomalies, jamming, etc.

phantom target—*See* Echo Box.

phase—1. The angular relationship between current and voltage in alternating-current circuits. 2. The number of separate voltage waves in a commercial alternating-current supply (e.g., single-phase, three-phase, etc.). Symbolized by the Greek letter *phi* (ϕ). 3. In a periodic function or wave, the fraction of the period which has elapsed, measured from some fixed origin. If the time for one period is represented as 360° along a time axis, the phase position is called phase angle.

phase advancer—A phase modifier which supplies leading reactive volt-amperes to the system to which it is connected. Phase advancers may be either synchronous or nonsynchronous.

phase angle—Of a periodic function, the angle obtained by multiplying the phase by 2π if the angle is to be expressed in radians, or 360 for degrees.

phase-angle correction factor—That factor by which the reading of a wattmeter or watt-hour meter operated from the secondary of a current or potential transformer, or both, must be multiplied to correct for the effective phase displacement of current and voltage due to the measuring apparatus.

phase-angle measuring relay—Also called out-of-step protective relay. A device which functions at a predetermined phase angle between two voltages or currents, or between voltage and current.

phase-angle meter—*See* Phase Meter.

phase angle of a current transformer—The angle between the primary-current vector and the secondary-current vector reversed. This angle is conveniently considered as positive when the reversed secondary-current vector leads the primary-current vector.

phase angle of a potential (voltage) transformer—The angle between the primary-voltage vector and the secondary-voltage vector reversed. This angle is conveniently considered as positive when the reversed secondary-voltage vector leads the primary-voltage vector.

phase balance—On a chopper, the phase-angle difference between positive and negative halves of the square wave; the difference in degrees between 180° and the measured angle between square-wave midpoints.

phase-balance current relay—*See* Reverse-Phase Current Relay.

phase-balance relay—A relay which functions by reason of a difference between two quantities associated with different phases of a polyphase circuit.

phase center (center of radiation)—Pertaining only to antenna types which have radiation characteristics such that while they are radiating energy, one can observe the antenna from a distance of many wavelengths and see the energy radiating from a point within the antenna array. The position of a point-source radiator which would replace the antenna and produce the same for field phase contour.

phase characteristic—A graph of phase shift versus frequency, assuming sinusoidal input and output.

phase-coefficient distortion—The difference between the maximum transit time for frequencies within a specified band.

phase-comparison tracking system—A system which provides target-trajectory information by the use of c-w phase-comparison techniques.

phase-compensation network—A network used to provide closed-loop stability in an operational amplifier. No greater than 12db/octave rolloff of the open-loop gain is allowed.

phase conductor—Those conductors other than the neutral conductor of a polyphase circuit.

phase constant—The imaginary component of the propagation constant. For a traveling plane wave at a given frequency, the rate in radians per unit length, at which the phase lag of a field component (for the voltage or current) increases linearly in the direction of propagation.

phase control—Also called horizontal-parabola control. One of three controls for adjusting the phase of a voltage or current in a color television receiver employing the magnetic-convergence principle. Each control varies the phases of the sinusoidal voltages applied, at the horizontal-scanning frequency, to the coils of the magnetic-convergence assembly.

phase-controlled rectifier — A rectifier circuit in which the rectifying element is a thyratron having a variable-phase, sine-wave grid bias.

phase correction—The process of keeping synchronous telegraph mechanisms in substantially correct phase relationship.

phase corrector—A network designed to correct for phase distortion.

phased array—An antenna term used to describe an array of dipoles on a radar antenna in which the signal feeding each dipole is varied in such a way that antenna beams can be formed in space and scanned very rapidly in azimuth and elevation.

phase delay—1. In the transfer of a single frequency wave from one point to another in a system, the delay of part of the wave identifying its phase. 2. The insertion phase shift divided by the frequency (both in cycles). *Also see* Delay Distortion.

phase-delay distortion—The difference between the phase delay at one frequency and the phase delay at a reference frequency.

phase detector—1. A tv circuit in which a d-c correction voltage is derived to maintain a receiver oscillator in sync with some characteristic of the transmitted signal. 2. A circuit which detects both the magnitude and the sign of the phase angle between two sine-wave voltages or currents.

phase deviation—In phase modulation, the peak difference between the instantaneous angle of the modulated wave and the angle of the carrier.

phase difference—1. The time in electrical degrees by which one wave leads or lags another. 2. Between two sinusoidal quantities that have the same period, the phase difference is the fraction of a period (one-half or less) through which the independent variable must be advanced or retarded so that the two quantities will coincide.

phase discriminator—A device in which amplitude variations are derived in response to phase variations.

phase distortion—*See* Phase-Frequency Distortion.

phase-distortion coefficient—In a transmission system, the difference between the maximum and minimum transit times for frequencies within a specified band.

phase equalizer—A circuit employed to neutralize the effect of phase-frequency distortion in a particular range of frequencies.

phase-frequency distortion — Also called phase distortion. Distortion that occurs when the phase shift is not directly proportionate to the frequency over the range required for transmission.

phase inversion — The condition whereby the output of a circuit produces a wave of the same shape and frequency but 180° out of phase with the input.

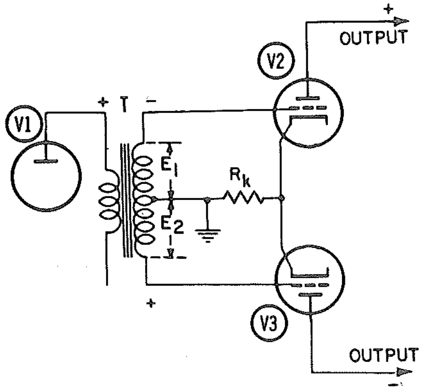

Phase inverter.

phase inverter—1. A stage that functions chiefly to change the phase of a signal by 180°, usually for feeding one side of a following push-pull amplifier. 2. *See* Vented Baffle.

phase localizer—An airfield runway localizer in which lateral guidance is obtained by comparing the phases of two signals.

phase lock—The technique of making the phase of an oscillator signal follow exactly the phase of a reference signal by comparing the phases between the two signals and using the resultant difference signal to adjust the frequency of the reference oscillator.

phase-locked loop—A narrow-band filter that continuously servos itself to the strongest signal frequency in the input spectrum. Phase-locked loops are used as highly selective, very narrow-band noise-rejection filters when an ordinary fixed filter might not stay aligned with the incoming signal frequency. Phase-locked loops differ from ordinary passive filters in that they utilize active components (they have gain) and a feedback loop. The feedback property is used for both the filtering and the frequency-servoing action.

phase magnet—Also called trip magnet. A magnetically operated latch used to phase a facsimile transmitter or recorder.

phase margin—1. A safety factor in phase shift. When the loop gain is 1.0 or more and the phase shifts total 180°, instability will occur. The amount that the total phase shift is less than 180° is called the phase margin. 2. The additional amount of phase shift of the output signals in an operational amplifier at the open-loop unity-gain crossover frequency that would produce instability.

phase meter—Also called phase-angle meter. An instrument for measuring the difference in phase between two alternating quantities of the same frequency.

phase modifier—A device that supplies leading or lagging volt-amperes to the system to which it is connected.

phase-modulated transmitter—A transmitter, the output of which is a phase-modulated wave.

phase-modulated wave—A wave whose phase angle has been caused to deviate from its original (no-signal) angle by an amount proportional to the modulating signal amplitude.

phase modulation—Abbreviated pm. Modulation in which the angle of a sine-wave carrier deviates from the original (no signal) angle by an amount proportional to the instantaneous value of the modulating wave. Phase and frequency modulation in combination are commonly referred to as "frequency modulation."

phase modulator—A circuit which modulates the phase of a carrier signal.

phase multiplier—A device that multiplies the frequency of signals used for phase comparison so that phase differences may be measured to a higher degree of resolution.

phase noise—A measure of the random phase instability of a signal.

phase of a wave—*See* Angle of a Wave.

phase-propagation ratio—In wave propagation, the propagation ratio divided by the magnitude of the wave.

phaser—A device for adjusting facsimile equipment so that the recorded area bears the same relationship to the record sheet as the corresponding transmitted area bears to the subject copy in the direction of the scanning line.

phase-recovery time (TR and pre-TR tubes)—The time required for a fired tube to deionize to such a level that a specified phase shift is produced in the low-level radio-frequency signal transmitted through the tube.

phase resonance—Also called velocity resonance. Resonance in which the angular phase difference between the fundamental components of the oscillation or vibration and the applied agency is 90°.

phase-response characteristic—The phase displacement versus frequency properties of a network or system.

phase reversal—A change of 180°, or one half of a cycle.

phase-reversal protection—In a polyphase circuit, the interruption of power whenever the phase sequence of the circuit is reversed.

phase-reversal switch—A switch used on a stereo amplifier or in a speaker system to shift the phase 180° on one channel.

phase-rotation relay—*See* Phase-Sequence Relay.

phase-sensitive amplifier—A servoamplifier the output signal polarity or phase of which is dependent upon the polarity or phase relationship between an error (input) voltage and a reference voltage.

phase-sequence indicator—A device that indicates the sequence in which the fundamental components of a polyphase set of potential differences, or currents, successively reach some particular value (e.g., their maximum positive value).

phase-sequence relay—Also called phase-rotation relay. A relay which functions according to the order in which the phase voltages successively reach their maximum positive values.

phase-sequence voltage relay—A device which functions upon a predetermined value of polyphase voltage in the desired phase sequence.

phase shift—1. The difference between corresponding points on input and output signals waveshapes (not affected by magnitude) expressed as degrees lead or lag. 2. A change in the phase of a periodic quantity.

phase-shift circuit—A network which shifts the phase of one voltage with respect to another voltage of the same frequency.

phase shifter—A device in which the output voltage (or current) may be adjusted to have some desired phase relationship with the input voltage (or current).

phase-shifting transformer—Also called a phasing transformer. A transformer connected across the phases of a polyphase circuit to provide voltages of the proper phase for energizing varmeters, varhour meters, or other instruments. (*Also see* Rotatable Phase-Adjusting Transformer.)

phase-shift keying—A form of phase modulation in which the modulating function shifts the instantaneous phase of the modulated wave between predetermined discrete values.

phase-shift microphone—A microphone, the

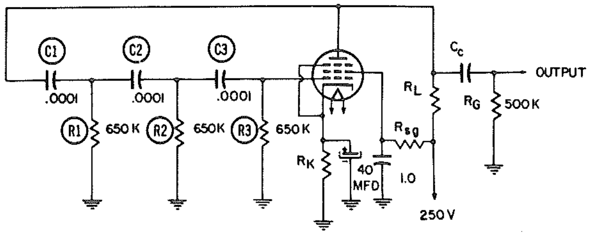

Phase-shift oscillator.

directional properties of which are provided by phase-shift networks.

phase-shift oscillator—An oscillator in which a network having a phase shift of an odd multiple of 180° (per stage) at the oscillation frequency is connected between the output and input of an amplifier. When the phase shift is obtained by resistance-capacitance elements, the circuit is called an RC phase-shift oscillator.

phase simulator—A precision test instrument which generates reference and data signals on the same frequency but precisely separated in phase. It is normally used to check out precision phase meters.

phase splitter—1. A device which produces, from a single input wave, two or more output waves that differ in phase from one another. 2. In color television, the stage which takes I and Q signals from demodulators, produces four signals, positive and negative I and Q, and feeds them to the matrix.

phase-tuned tube (TR tubes)—A fixed tuned broad-band TR tube in which the phase angle through it and the reflection it introduces are kept within limits.

phase undervoltage relay—A relay which is tripped by the reduction of one phase voltage in a polyphase circuit.

phase velocity—The velocity at which a point of constant phase is propagated in a progressive sinusoidal wave.

phase-versus-frequency response characteristic—A graph or other tabulation of the phase shifts occurring, in an electrical transducer, at several frequencies within a band.

phasing—1. Causing two systems or circuits to operate in phase or at some desired difference from the in-phase condition. 2. Adjusting a facsimile-picture position along the scanning line. 3. In stereo application, the establishment of the correct relative polarity in the connection between amplifier output and speakers so that one speaker tends to reinforce rather than cancel the output of the other (particularly evident at low frequencies).

phasing capacitor—A capacitor used in a crystal-filter circuit for neutralizing the capacity of the crystal holder.

phasing line—In facsimile, the portion of the scanning line set aside for the phasing signal.

phasing pulse—A short pulse or signal employed for phasing the recorder with the transmitter in a television or facsimile system.

phasing signal—In facsimile, a signal used for adjusting the position of the picture along the scanning line.

phasing transformer—*See* Phase-Shifting Transformer.

phasitron—A tube designed to produce a frequency-modulated audio signal, which is induced by a varying field from a magnet placed around the glass envelope of the tube.

phasmajector—*See* Monoscope.

phasor—An entity which includes the concepts of magnitude and direction in a reference plane.

phenolic material—Any one of several thermosetting plastic materials available which may be compounded with fillers and reinforcing agents to provide a broad range of physical, electrical, chemical, and molding properties.

Phillips gage—A vacuum gage in which gas pressure is determined by measuring the current in a glow discharge.

Phillips screw—A screw with an indented cross in its head, instead of the conventional slot. It must be removed or inserted with a special screwdriver, also called a Phillips.

phi polarization—In an electromagnetic wave, the state in which the E vector of the wave is tangential to the lines of latitude of some given spherical frame of reference.

pH meter—An instrument used with a probe to determine the alkalinity or acidity of a solution.

phon—The unit for measuring the apparent loudness level of a sound. Numerically equal to the sound-pressure level, in decibels relative to 0.0002 microbar, of a 1,000-hertz tone that is considered by listeners to be equivalent in loudness to the sound under consideration.

phone—*See* Headphone.

phone jack—Also called a telephone jack. A jack designed for use with phone plugs.

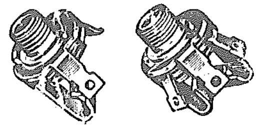

Phone jack.

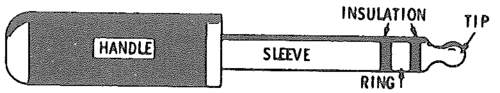

Phone plug.

phonemes—The minimal set of shortest segments of speech which, if substituted one for another, convert one word to another.

phone plug—Also called telephone plug. A plug used with headphones, microphones, and other audio equipment.

phonetic alphabet—A list of standard words, one for each letter in the alphabet. It is used for distinguishing the letters in a spoken radio or telephone message. The list reads:

ALFA
BRAVO
CHARLIE
DELTA
ECHO
FOXTROT
GOLF
HOTEL
INDIA
JULIETT
KILO
LIMA
MIKE
NOVEMBER
OSCAR
PAPA
QUEBEC
ROMEO
SIERRA
TANGO
UNIFORM
VICTOR
WHISKEY
XRAY
YANKEE
ZULU

phonocardiogram—A graphic recording of the sounds produced by the heart and its associated parts (e.g., its mitral or aortic valves).

phonocardiograph—An instrument for recording sounds of the heart on a strip chart.

phonocatheter—A catheter-microphone combination that is inserted through the artery into the heart. It picks up inner cardiac sounds.

phonoelectrocardioscope—A dual-beam oscilloscope which displays both ecg signals and heart-sound signals.

phonograph—An instrument for reproducing sound. It consists of a turntable on which the grooved medium containing the impressed sound is placed, a needle that rides in the groove, and an electrical (formerly mechanical) amplifying system for taking the minute vibrations of the needle and converting them into electrical (formerly mechanical) impulses that drive a speaker.

phonograph oscillator—An r-f oscillator circuit, the output of which is modulated by a phonograph pickup and sent through space to a receiver. Thus, no wires to the receiver are needed.

phonograph pickup—Also called mechanical reproducer, pickup, or phono pickup. A mechanoelectric transducer which is actuated by modulations present in the groove of a recording medium and transforms this mechanical input into an electric output.

phono jack—A jack designed to accept a phono plug.

phonon—A lattice vibration with which a discrete amount (quantum) of energy is associated. Some thermal and electrical properties of the lattice are theoretically treated in terms of electron-phonon interactions.

phono pickup—*See* Phonograph Pickup.

Phono plug.

phono plug—A plug used at the end of a shielded conductor for feeding a-f signals to a mating phono jack on an audio preamplifier or amplifier.

phonoselectoscope — A stethoscopic device which suppresses low frequencies (characteristic of the normal heart function) to permit detection of higher-frequency sounds.

phosphor—A layer of luminescent material applied to the inner face of a cathode-ray tube. During bombardment by electrons it fluoresces, and after the bombardment, it phosphoresces.

phosphor-dot faceplate—The glass viewing screen on which the trios of color phosphor dots are mounted in a three-gun picture tube.

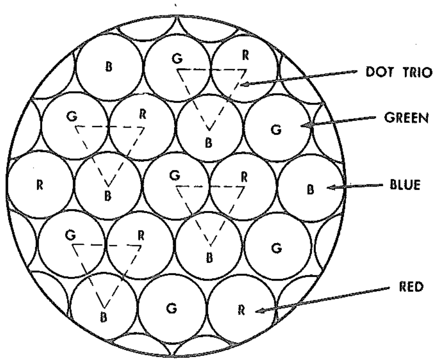

Phosphor dots.

phosphor dots—Minute particles of phosphor on the viewing screen of a picture tube. On a tricolor picture tube, the red, green, and blue phosphor dots are placed on the viewing screen in a pattern of dot triads—a phosphor dot of each color forming one-third of the triad.

phosphorescence — The emission of light from a substance after excitation has been removed. *Also see* Afterglow.

phosphors—Materials that emit visible or invisible (ultraviolet or infrared) radiation when irradiated with X-rays, cathode rays, and ultraviolet radiation, or when placed in an electric field. The quantity of visible light is proportional to the amount of excitation energy present. The emission is not due to incandescence and so is often called "cold light" to distinguish it from radiation due to heating, which occurs in an electric-light bulb. The conversion of electrical energy to light within these phosphors is called electroluminescence.

phosphor trio—In the phosphor screen of a tricolor kinescope, closely spaced triangular groups of three phosphor dots accurately deposited in interlaced positions.

photocathode—An electrode which releases electrons when exposed to light or other suitable radiation. Used in phototubes, television camera tubes, and other light-sensitive devices.

photocathode blue response — The photoemission current resulting from a specified luminous flux from a tungsten lamp filament at a color temperature of 2,870° K when the flux is filtered by a specified blue filter.

photocathode luminous sensitivity — The quotient of photoelectric emission current from the photocathode divided by the incident luminous flux. The measurement is made under specified conditions of illumination, usually with radiation from a tungsten-filament lamp operated at a color temperature of 2,870° K. The cathode is usually illuminated by a collimated beam at normal incidence.

photocathode radiant sensitivity—The quotient of the photoelectric emission current from the photocathode divided by the incident radiant flux. It is usually measured at a given wavelength under specified conditions of irradiation with a collimated beam at normal incidence.

photocell—*See* Photoelectric Cell.

photoconductive cell—A photoelectric cell, the electrical resistance of which varies inversely with the intensity of light that strikes its active material.

photoconductive effect—The change of electrical conductivity of a material when exposed to varying amounts of radiation.

photoconductive material—Material having a high resistance in the dark, and a low resistance when exposed to the light.

photoconductivity — The greater electrical conductivity shown by some solids when illuminated. The incoming radiation transfers energy to an electron, which then takes on a new energy level (in the conduction band) and contributes to the electrical conductivity.

photoconductor—A passive, high-impedance device composed of thin single-crystal or polycrystalline films of compound semiconductor materials. When the sensitive surface is illuminated its resistance decreases and hence its conductivity increases.

photodielectric effect — The change in the dielectric constant and loss of a material when illuminated. The effect is observed only in phosphors that show photoconductivity during luminescence.

photodiffusion effect—*See* Dember Effect.

photodiode—A semiconductor diode in which the reverse current increases whenever light falls on the diode.

photoelasticity—Changes in the optical properties of transparent isotropic dielectrics subject to stress.

photoelectric—Pertaining to the electrical effects of light or other radiation—i.e., emission of electrons, generation of a voltage, or a change in electrical resistance upon exposure to light.

photoelectric absorption—Conversion of radiant energy into photoelectric emission.

photoelectric cathode—A cathode the primary function of which is photoelectric emission.

photoelectric cell—Also called a photocell. A light-sensitive cell which translates variations in light into corresponding variations in electrical signals.

photoelectric color comparator—*See* Color Comparator.

photoelectric colorimeter — A colorimeter which uses a photoelectric cell and a set of color filters to determine, by the output current for each filter, the chromaticity coordinates of light of a given sample.

photoelectric conductivity — The increased conductivity exhibited by certain crystals when struck by light (e.g., a selenium cell).

photoelectric constant — A quantity which, when multiplied by the frequency of the radiation causing the emission, gives (in centimeter-gram-second units) the voltage absorbed by the escaping photoelectron. The constant is equal to h/e, where h is Planck's constant and e is the electronic charge.

photoelectric control — The control of a circuit or piece of equipment in response to a change in incident light impinging on a photosensitive device.

photoelectric counter—A device that registers a count whenever an object breaks the light beam shining on its phototube or photocell. An amplifier then boosts the minute energy to register on a mechanical or other type of counter.

photoelectric current—The stream of electrons emitted from the cathode of a phototube under the influence of light.

photoelectric cutoff control—A photorelay circuit used in machines for cutting long strips of paper, cloth, metal, or other material accurately into predetermined lengths or at predetermined positions.

photoelectric effect—The transfer of energy from incident radiation to electrons in a substance. This phenomenon includes photoelectric emission of electrons from the surface of a metal, the photovoltaic effect, and photoconductivity.

photoelectric electron-multiplier tube — A vacuum phototube that employs secondary emissions to amplify the electron stream emitted from the illuminated photocathode.

photoelectric emission — Electron emission due directly to the incidence of radiant energy on the emitter.

photoelectric flame-failure detector — An industrial electronic control employing a phototube and amplifier to actuate an electromagnetic or other valve that cuts off the fuel flow when the fuel-consuming flame is extinguished and light no longer falls on the phototube.

photoelectric inspection—Quality control of a product by means of a phototube, light-beam system, and associated electronic equipment.

photoelectric intrusion detector—A burglar-alarm system in which interruption of a light beam by an intruder reduces the illumination on a phototube and thereby closes an alarm circuit.

photoelectric liquid-level indicator—A level indicator in which the rising liquid interrupts the beam of light in a photoelectric control system.

photoelectric material—Any material that will emit electrons when illuminated in a vacuum (e.g., barium, caesium, lithium, potassium, rubidium, sodium, and strontium).

photoelectric phonograph pickup—A phonograph reproducing device consisting essentially of a light source, a jewel stylus to which a very thin mirror is attached, and a selenium cell that picks up light reflected from the mirror. Sidewise movements of the stylus in the record groove cause the amount of reflected light to vary, and accordingly the resistance of the selenium cell. The light source is fed by a radio-frequency oscillator rather than from the power line, to eliminate 60-hertz flicker from the light beam.

photoelectric photometer — A photometer which incorporates a phototube or photoelectric cell for measurements of light.

photoelectric pickup — A transducer that transforms a change in light into an electric signal.

photoelectric pyrometer—An instrument for measuring high temperatures from the intensity of the light given off by the heated object.

photoelectric reader—A device that reads information stored in the form of holes punched in paper tape or cards, by sensing light passed through the holes.

photoelectric recorder—An optical recording instrument employing a light source and phototube for the basic measuring element.

photoelectric register control — A photoelectric device used for controlling the position of a strip of paper, cloth, metal, etc., with respect to the machine through which it is being passed.

photoelectric relay—Also called light relay. A relay combined with a phototube (and amplifier if necessary), so arranged that changes in incident light on the phototube cause the relay contacts to open or close.

photoelectric scanner—A light source, lens system, and one or more phototubes in a single, compact housing. It is mounted a few inches above a moving surface, where it actuates control equipment when the amount of light reflected from the surface changes.

photoelectric sensitivity—Also called photoelectric yield. The rate at which electrons are emitted from a metal per unit radiant flux at a given frequency.

photoelectric smoke detector—A photoelectric instrument used to measure the density of smoke and to sound an alarm when a predetermined smoke density is exceeded.

photoelectric sorter — An industrial-electronic control employing a light beam, phototube, and amplifier to sort objects according to color, size, shape, or other characteristics.

photoelectric threshold—The quantum energy just sufficient to release photoelectrons from a given surface. The corresponding frequency is the critical, or threshold, frequency.

photoelectric timer—An electronic instrument that automatically turns off an X-ray machine or other photographic device when the film reaches the correct exposure.

photoelectric transducer—A transducer that converts changes in light energy into electrical changes.

photoelectric tube—*See* Phototube.

photoelectric work function—The energy required to transfer electrons from a given metal to a vacuum or other adjacent medium during photoelectric emission. It is sometimes expressed as energy in ergs or joules per unit of emitted charge, and sometimes as energy per electron in electron volts.

photoelectric yield—*See* Photoelectric Sensitivity.

photoelectromagnetic effect — *See* Dember Effect.

photoelectromotive force — Electromotive force caused by photovoltaic action.

photoelectrons—The electrons emitted from a metal in the photoelectric effect.

photoemissive—Capable of emitting electrons when under the influence of light or other radiant energy.

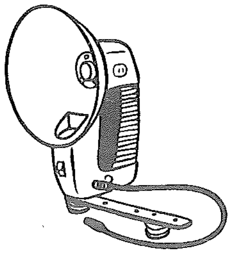

Photoflash.

photoflash—A means of firing expendable flashbulbs with an instantaneous surge of current supplied by two or more $1\frac{1}{2}$-volt, single-cell batteries, or by the discharge of a capacitor which has been charged to full capacity by a medium-voltage battery.

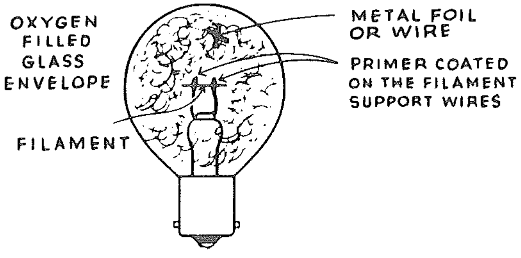
OXYGEN FILLED GLASS ENVELOPE
FILAMENT
METAL FOIL OR WIRE
PRIMER COATED ON THE FILAMENT SUPPORT WIRES

Photoflash bulb.

photoflash bulb—An oxygen-filled glass bulb containing a metal foil or wire. A surge of current heats the metal to incandescence, and a brilliant flash of light is produced when the wire burns in the oxygen.

photoflash tube—*See* Flash Tube.

photoflood lamp—An incandescent lamp employing excess voltage to give brilliant illumination. Used in television and photography, it has a life of only a few hours.

photogalvanic cell—A cell which generates an electromotive force when light falls on either of the electrodes immersed in an electrolyte.

photogenerator—A semiconductor-junction device that emits light when pulsed.

photoglow tube—A gas-filled phototube used as a relay. This is done by making the operating voltage so high that ionization and a glow discharge occur, accompanied by considerable current flow, when certain illumination is reached.

photographic sound recorder—*See* Optical Sound Recorder.

photographic sound reproducer — *See* Optical·Sound Reproducer.

photoionization—Ionization occurring in a gas as a result of visible light or ultraviolet radiation.

photo-island grid—A photosensitive surface in the storage-type Farnsworth dissector tube used with television cameras. It comprises a thin, finely perforated (about 400 holes per square inch) sheet of metal.

photojunction battery—A nuclear-type battery in which the radioactive material, promethium 147, irradiates a phosphor which converts nuclear energy into light. The light is then converted to electrical energy by a small silicon junction.

photoluminescence — Luminescence stimulated by visible light or ultraviolet radiation.

photomagnetic effect—The direct effect of light on the magnetic susceptibility of certain substances.

photomagnetoelectric effect — The production in a semiconductor of an electromotive force normal to both an applied magnetic field and to a photon flux of proper wavelength.

photometer—An instrument for measuring the intensity of a light source or the amount of illumination, usually by comparison with a standard light source.

photometry—The techniques for measuring luminous flux and related quantities (e.g., luminous intensity, illuminance, luminance, luminosity, etc.).

photomultiplier pulse-height resolution — A measure of the smallest change in the number of electrons emitted during a pulse from the photocathode that can be discerned as a change in output-pulse height.

photomultiplier tube—*See* Multiplier Phototube.

photon—A quantum of electromagnetic energy. The equation is hv, where h is Planck's constant and v is the frequency associated with the photon.

photon-coupled isolator—A circuit-coupling device consisting of an infrared emitter diode-coupled to a photon detector over a short shielded light path, which provides extremely high circuit isolation.

photon coupling—Coupling between circuits by a beam of light.

photonegative—Having a negative photoconductivity—hence, decreasing in conductivity (increasing in resistance) under the action of light. Selenium sometimes exhibits this property.

photoparametric diode—A pill-sized device for simultaneously detecting and amplifying optical energy modulated at microwave frequencies.

photophone—A device for converting variations in light intensity into sound.

photopositive—Having a positive photoconductivity—hence, increasing in conductivity (decreasing in resistance) under the action of light. Selenium ordinarily has this property.

photorelay circuit—A form of on-off control actuated by a change of illumination.

photoresistive or photoconductive transduction—Conversion of the measurand into a change in the resistance of a semiconductor material (by changing the illumination incident on the material).

photoresistor — A semiconductor resistor which, when illuminated, drops in resistance.

photosensitive—Capable of emitting electrons when struck by light rays.

photosensitive field-effect transistor — A special unipolar field-effect transistor (FET) structure that is positioned on a header to receive illumination transmitted through a lens in the top of the header can. It combines the circuit and device characteristics of a photodiode and a high-impedance low-noise amplifier.

photosensitive recording—Recording by the exposure of a photosensitive surface to a signal-controlled light beam or spot.

photosensitive semiconductor—A semiconductor material in which light energy controls the current-carrier movement.

photosphere—The outermost luminous layer of the gaseous body of the sun.

phototelegraphy—In facsimile, the process of sending photographs over a wire.

phototransistor—A thin wafer of germanium in which holes are generated by light absorption and produce a multiplied photocurrent by transistor action at the collector.

phototube—Also called photoelectric tube. An electron tube containing a photocathode. Its output depends on the total photoelectric emission from the irradiated area of the photocathode.

phototube bridge circuit—A circuit in which a phototube is one arm of a bridge circuit. With such a circuit, a balanced condition (no signal output) can be reached under either a black-signal or white-signal condition, depending on the impedance adjustments in the other arms.

phototube relay—An electrical relay in which the action of a beam of light on a phototube operates mechanical devices such as counters and safety controls.

photovaristor—A varistor in which the current-voltage relation may be modified by illumination. Cadmium sulphide and lead telluride exhibit such properties.

photovoltaic—Capable of generating a voltage when exposed to visible or other light radiation.

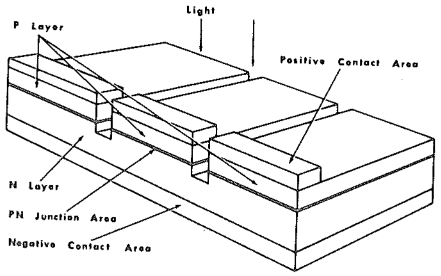

Photovoltaic cell.

photovoltaic cell—A self-generating semiconductor device which converts light energy into electrical energy when struck by light.

photovoltaic converter—A device for converting light to electric energy by means of the photovoltaic effect.

photovoltaic effect—The generation of a voltage (or an electric field) in a material that is illuminated with radiation of a suitable wavelength.

photovoltaic transduction — Conversion of the measurand into a change in the voltage generated when a junction of dissimilar material is illuminated.

photox cell—A type of photovoltaic cell in which a voltage is generated between a copper base and a film of cuprous oxide during exposure to visible or other radiation.

photran—A triode PNPN-type switch of the reverse-blocking PNPN-type switch class. The photran provides optical triggering in addition to standard gate-terminal triggering.

photronic cell—A type of photovoltaic cell in which a voltage is generated in a layer of selenium during exposure to visible or other radiation.

pi—The Greek letter π. It designates the value 3.14159+, which is approximately the ratio of the circumference of a circle to its diameter.

pickoff—A device which produces a signal output, generally a voltage as a function of the angle between two gimbals, or between a gimbal and a base.

pickup—1. A device that converts a sound, scene, or other form of intelligence into corresponding electric signals (e.g., a microphone, television camera, or phonograph pickup). 2. The minimum current, voltage, power, or other value which will trip a relay. 3. Interference from a nearby circuit or system.

pickup arm—Also called tone arm. A pivoted arm for holding a pickup.

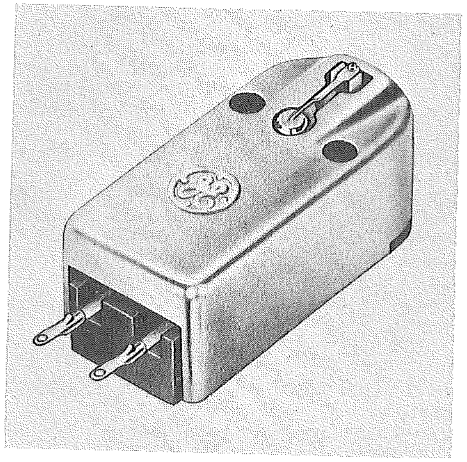

Pickup cartridge.

pickup cartridge—The removable portion of a pickup arm. It contains the electromechanical transducing elements and the reproducing stylus.

pickup current—Also called pull-in current. That current at which a magnetically operated device starts to operate.

pickup spectral characteristic—In television, the spectral response of the camera tube including the optical parts. It converts radiation into electric signals, which are measured at the output terminals of the pickup tube.

pickup value (voltage, current, or power)—The minimum value which will energize the contacts of a relay. (*Also see* Drop-out Value *and* Hold Current.)

pickup voltage—That voltage at which a magnetically operated device starts to operate.

pico—Prefix meaning 10^{-12}. (Formerly micromicro.) Abbreviated p.

picoammeter—A sensitive ammeter that indicates current values in picoamperes.

picoampere—One-millionth of a microampere (10^{-12} ampere). Abbreviated pa.

picocoulomb — One-millionth of a microcoulomb. (10^{-12} coulomb). Abbreviated pc.

picocurie—One-millionth of a microcurie (10^{-12} curie). Abbreviated pc.

picofarad—Abbreviated pf. 10^{-12} farad; formerly micromicrofarad. Called puff in England.

picosecond—A micromicrosecond (10^{-12} second). Abbreviated psec.

picowatt—One millionth of a microwatt (10^{-12} watt). Abbreviated pw. *Also see* Micromicrowatt.

pictorial wiring diagram—A wiring diagram containing actual sketches of components and clearly showing all connections between them.

picture black—Also called black signal. In facsimile, the signal produced at any point by the scanning of a selected area of subject copy having maximum density.

picture brightness—A measure of the brightness of the highlights in a television picture, usually measured in foot-lamberts.

picture carrier—Also called luminance carrier. The carrier frequency 1.25 MHz above the lower frequency limit of a standard NTSC television signal. In color television, this carrier is used for transmitting luminance information; and the chrominance subcarrier, which is 3.579545 MHz higher in frequency, transmits the color information.

picture element—1. In facsimile, that portion of the subject copy which is seen by the scanner at any instant. It can be considered a square area having dimensions equal to the width of the scanning line. 2. Also called elemental area, critical area, scanning spot, or recording spot. In television, any segment of a scanning line, the dimension of which along the line is exactly equal to the nominal line width. The area being explored at any instant in the scanning process.

picture frequency—The number of complete pictures scanned per second in a television system.

picture line standard—The total number of horizontal lines in a complete television image. The standard in the United States is 525 lines.

picture monitor—A cathode-ray tube and its associated circuits arranged for viewing a television picture.

picture signal—In television, the signal resulting from the scanning process.

picture-signal amplitude—The difference between the white peak and the blanking level of a video signal.

picture-signal polarity—The polarity of the signal voltage which represents a dark area of a scene given with respect to the signal voltage representing a light area. Expressed as black negative or black positive.

picture size—The usable viewing area on the screen of a television receiver, measured in square inches.

picture synchronizing pulse—*See* Vertical Synchronizing Pulse.

picture transmission—Electric transmission of a shaded (halftone) picture.

picture transmitter—*See* Visual Transmitter.

picture tube—*See* Kinescope.

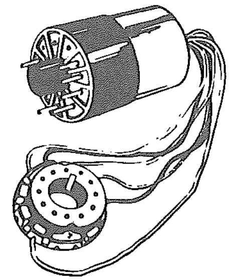

Picture-tube brightener.

picture-tube brightener—An accessory added to an aging picture tube to increase the image brightness and thereby extend its useful life. When the cathode emission is subnormal, the brightener raises the filament voltage and thereby increases the electron emission from the cathode.

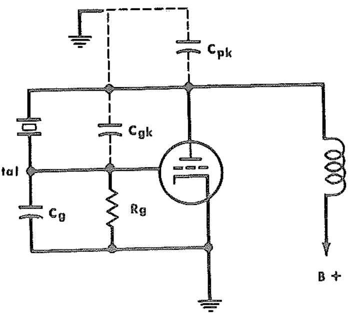

Pierce oscillator.

Pierce oscillator—Basically a Colpitts oscillator in which a piezoelectric crystal is connected between the plate and grid. Voltage division is provided by the grid-to-cathode and plate-to-cathode capacitances of the circuit.

pie winding—A winding constructed from individual washer-shaped coils called pies.

piezodielectric—Pertaining to a change in dielectric constant under mechanical stress.

piezoelectric—The property of certain crystals, which: 1. Produce a voltage when subjected to a mechanical stress. 2. Undergo mechanical stress when subjected to a voltage.

piezoelectric accelerometer — Basically a crystalline material which, when force is applied, generates a charge. Through the incorporation of a mass in direct contact with the crystal, an acceleration transducer is produced.

piezoelectric axis—In a crystal, one of the directions in which tension or compression will develop a piezoelectric charge.

piezoelectric crystal—A piece of natural quartz or other crystalline material capable of demonstrating the piezoelectric effect. A quartz crystal, when ground to certain dimensions, will vibrate at a desired radio frequency when placed in an appropriate electric circuit.

piezoelectric crystal cut—The orientation of a piezoelectric crystal plate with respect to the axes of the crystal. It is usually designated by symbols—e.g., GT, AT, BT, CT, and DT identify certain quartz-crystal cuts having very low temperature coefficients.

piezoelectric crystal element—A piece of piezoelectric material cut and finished to a specified shape and orientation with respect to the crystallographic axes of the material.

piezoelectric crystal plate—A piece of piezoelectric material cut and finished to specified dimensions and orientation with respect to the crystallographic axes of the material and having two essentially parallel major surfaces.

piezoelectric crystal unit—A complete assembly comprising a piezoelectric crystal element mounted, housed, and adjusted to the desired frequency, with means for connecting it into an electric circuit. Such a device is commonly employed for frequency control or measurement, electric wave filtering, or interconversion of electric and elastic waves.

piezoelectric device—1. A substance which generates an electric voltage when bent, squeezed, or twisted. 2. Conversely, when a voltage is applied, it will twist, bend, expand, or contract.

piezoelectric effect—1. The mechanical deformity of certain natural and synthetic crystals under the influence of an electric field. This effect is used in high-precision oscillators and certain high-frequency filters. 2. The property of certain natural and synthetic crystals to produce a voltage when subjected to mechanical stress (compression, expansion, twisting, etc.). 3. The production of a mechanical stress in a crystal by the application of a voltage.

piezoelectricity—The phenomenon whereby certain crystalline substances such as barium titanate generate electrical charges when subjected to mechanical deformation. The effect was first noticed by the Curie brothers in the 1880's as a result of extensive study of the symmetry of crystalline materials. The reverse effect also occurs.

piezoelectric material—A material which generates an electrical output when subjected to a mechanical stress. (The word is derived from the Greek *piezein*, meaning to squeeze or press.)

piezoelectric microphone — *See* Crystal Microphone.

piezoelectric oscillator—A crystal-oscillator circuit in which the frequency is controlled by a quartz crystal.

piezoelectric pickup—*See* Crystal Pickup.

piezoelectric pressure gage—An apparatus for measuring or recording very high pressures. The pressure is applied to quartz discs or other piezoelectric crystals. The resultant voltage, after amplification, is then measured or is recorded with an oscillograph.

piezoelectric speaker—*See* Crystal Speaker.

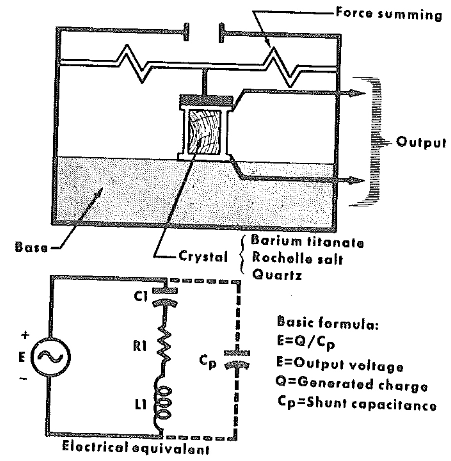

Piezoelectric transducer.

piezoelectric transducer — Also called ceramic or crystal transducer. A transducer that depends for its operation on the interaction between the electric charge and the deformation of certain asymmetric crystals having piezoelectric properties.

piezoelectric transduction—The conversion of the measurand into a change in the electrostatic charge or voltage generated by mechanically stressed crystals.

piezoid—The finished crystal product. It may include the electrodes making contact with the crystal blank.

piezoresistance — Resistance that changes with pressure.

pig discharge—*See* Penning discharge.

piggyback—*See* Voltage Corrector.

piggy-back control—*See* Cascade Control.

pigtail—Either a wire attached for terminating purposes to a shield, or a conductor extending from a small component.

pigtail splice—A splice made by tightly twisting the bared ends of parallel conductors together.

pile—A nuclear reactor. So called because early reactors were piles of graphite blocks and uranium slugs.

pileup—Also called stack. On a relay, a set of contact arms, assemblies, or springs fastened one on top of the other with layers of insulation separating them.

pillbox antenna—A cylindrical parabolic reflector enclosed by two plates perpendicular to the cylinder and spaced to permit propagation of only one mode in the desired direction of polarization. It is fed on the focal line.

pilot—1. In a transmission system, a signal wave, usually a single frequency, transmitted over the system to indicate or control its characteristics. 2. In a tape relay, instructions appearing in a routing line, relative to the transmission or handling of that message.

pilot cell—The storage-battery cell selected because its temperature, voltage, and specific gravity are assumed to be those of the entire battery.

pilot channel—A very narrow-band channel over which a single frequency, the pilot frequency, is transmitted to operate trouble alarms or automatic level regulators, or both.

pilot contacts—Contacts, the opening, closing, or transfer of which govern an operation of relays or similarly controlled devices.

pilot holes—Also called manufacturing holes or fabrication holes. Holes on a printed-circuit board used as guides during manufacturing operations. Mounting holes in a part are sometimes used as pilot holes.

pilot lamp—A light that indicates whether a circuit is energized.

pilot light—A light which, by means of position or color, indicates whether a control is functioning.

pilot regulator—A device for maintaining the receiving level of a carrier-derived circuit constant under varying attenuation conditions of the transmission line.

pilot spark—A low-power preliminary spark used in a gas-discharge tube to produce an ionized path for the large main spark discharge.

pilot subcarrier — A subcarrier used as a control signal in the reception of compatible f-m stereophonic broadcasts.

pilot wire—An auxiliary conductor used with remote measuring devices or for operating apparatus at a distance.

pilot-wire regulator—An automatic device for controlling adjustable gains or losses associated with transmission circuits to compensate for transmission changes caused by temperature variations, the control usually depending upon the resistance of a conductor or pilot wire having substantially the same temperature conditions as the conductors of the circuit being regulated.

pin—Also called a prong or base pin. A terminal on a connector, plug, or tube base.

pinch effect—1. The result of an electromechanical force that constricts, and sometimes momentarily ruptures, a molten conductor which is carrying a high-density current. 2. In the reproduction of lateral recordings, the pinching of the reproducing stylus tip twice each cycle. This is due to a decrease in the groove angle cut by the recording stylus during the swing from a negative to a positive peak. 3. The self-contraction of a plasma column carrying a large current due to the interaction of this current with the magnetic field it produces. The current required for such an effect is on the order of 10^5 amperes. If the current is in the form of a short pulse, a radially imploding shock wave is generated.

pinch-off—In a field-effect transistor, the reduction of source-to-drain current as much as possible.

pinch-off voltage—The gate voltage of a field-effect transistor that blocks the current for all source-drain voltages below the junction breakdown value. Pinch-off occurs when the depletion zone completely fills the area of the device.

pinch roller—*See* Pressure Roller.

pin connection—Connections made to the base of pins in a vacuum tube. They are identified by the following abbreviations: NC, no connection; IS, internal shield; IC, internal connection (not an electrode connection); P, plate; G, grid; SG, screen grid; SU, suppressor; K, cathode; H, heater; F, filament; RC, ray-control electrode; TA, target.

pin contact—A male-type contact, usually designed to mate with a socket or female contact. It is normally connected to the "dead" side of a circuit.

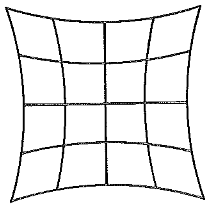

Pincushion distortion.

pincushion distortion—Distortion which results in a monotonic increase in radial magnification in the reproduced image away from the axis of symmetry of the electron optical system. The four sides of the raster are curved inward, leaving the corners extending outward.

pine-tree array—An array of dipole antennas aligned vertically (termed the radiating curtain), behind which and approximately a quarter wavelength away is a parallel array of dipole antennas forming the reflecting curtain.

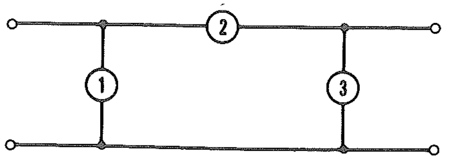

Pi network.

pi network—A network composed of three branches, all connected in series with each other to form a mesh. The three junction points form an input terminal, an output terminal, and a common input and output terminal, respectively.

pinfeed platen—In a computer, a cylindrical platen having integral rings of pins that engage perforated holes in the paper, thus permitting feeding of the paper.

ping—A sonic or ultrasonic pulse of predetermined width.

ping-pong-ball effect—The bouncing of sound back and forth between the two sides of a stereophonic reproducing system.

pinhole detector—A photoelectric device that detects extremely small holes and other defects in moving sheets of material, and often actuates sorting equipment that automatically rejects defective sheets.

pinholes—Small holes, occurring as imperfections, which penetrate entirely through an applied material to the substrate (e.g., holes in semiconductor insulating oxides, screened resistors, thin-film elements, etc.).

pinion—Of two gears that mesh, the one with the fewer teeth.

pin jack—A single-conductor jack having a small opening into which a plug tipped with a metal pin can be inserted.

PINO—Acronym for Positive Input, Negative Output.

pinouts—The external wires or pins on a module (generally having a circuit function).

pip—*See* Blip.

piped program—A program transmitted over telephone wires, usually from one studio to another.

pip-matching display—A navigational display in which the received signal appears as a pair of blips. The desired quantity is measured by comparing the characteristics.

pi (π) point—The frequency at which the insertion phase shift of an electric structure is 180° or an integral multiple of 180°.

Pirani gage—A bolometric vacuum gage for measuring pressure. Its operation depends on the thermal conduction of the gas present. The pressure being measured is a function of the resistance of a heated filament, ordinarily over a range of 10^{-1} to 10^{-4} mm Hg.

piston—Also called a plunger. In high-frequency communications a conducting plate that can be moved along the inside of an enclosed transmission path to short out high-frequency currents.

piston action—The movement of a speaker cone or diaphragm when driven at the bass audio frequencies.

piston attenuator—An attenuator generally used at microwave frequencies, the amount of attenuation of which can be varied by moving an output coupling device along its longitudinal axis.

pistonphone—A small chamber equipped with a reciprocating piston of measurable displacement. In this way, a known sound pressure can be established in the chamber.

pitch—1. That attribute of auditory sensation by which sounds may be ordered on a scale extending from low to high (e.g., a musical scale). 2. The distance between two adjacent corresponding threads of a screw measured parallel to the axis. 3. The distance between the peaks of two successive grooves of a disc recording.

pits—Small holes occurring as imperfections which do not penetrate entirely through the printed element.

piv—Abbreviation for peak inverse voltage.

place—*See* Column.

planar—Lying essentially in a single plane.

planar devices—*See* Planar Process.

planar mask—A shadow or aperture mask which has no curvature; one which is perfectly flat.

planar module—A packaged module wherein the individual components are positioned and terminated flat or parallel with the plane of the substrate.

planar network—A network in which no branches cross when drawn on the same plane.

planar process—The technology used in fabricating semiconductor devices wherein all PN junctions terminate in the same geometric plane. An oxide is formed at the surface for the purpose of stabilizing the parameters (passivating).

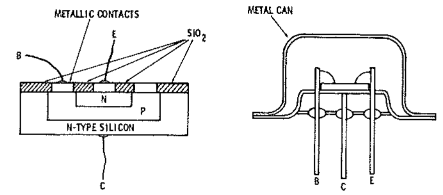

Planar transistor.

planar transistor—A diffused transistor in which the emitter, base, and collector regions come to the same plane surface. Their junctions are protected by a material such as silicon oxide. The manufacturing process consists of an oxide-masking technique in which the silicon oxide is formed by adding oxygen or water vapor to the atmosphere of a diffusion furnace. The thickness of the oxide layer is a function of time, temperature, and the amount of oxidizing agent.

planchet—A small metal container or sample holder for radioactive materials undergoing radiation measurements in a proportional counter or scintillation detector.

Planckian locus—A line drawn through all points on a chromaticity diagram to represent light radiation from a reference black body at 2,000° to 10,000° Kelvin (K).

Planck's constant—Symbolized by h. The constant representing the ratio of the energy of any radiation quantum to its frequency. It has the dimension of action (energy × time) and a numerical value of 6.547×10^{-27} erg-second. Its significance was first recognized by the German physicist Max Planck in 1900.

plane earth—Earth that is considered to be a plane surface. Used in ground-wave calculations.

plane-earth attenuation—Attenuation of an electromagnetic wave over an imperfectly conducting plane earth in excess of that over a perfectly conducting plane.

plane-earth factor—Ratio of the electric field strength that would result from propagation over an imperfectly conducting plane earth, to that over a perfectly conducting plane.

plane of a loop—An infinite imaginary plane which passes through the center of a loop and is parallel to its wires.

plane of polarization—For a plane-polarized wave, the plane containing the electric field vector and the direction of propagation.

plane-polarized wave—At any point in a homogeneous isotropic medium, an electromagnetic wave with an electric field vector that at all times lies in a fixed plane containing the direction of propagation.

planetary electron—One of the electrons moving in an orbit or shell around the nucleus of an atom.

plane wave—A wave in which the wavefronts are everywhere parallel planes normal to the direction of propagation.

plan-position indicator—Abbreviated ppi. Also called P-display. A type of presentation on a radar indicator. The signal appears as a bright spot, with range indicated by the distance of the spot from the center of the screen, and the bearing by the radial angle of the spot.

plasma—1. A wholly or partially ionized gas in which the positive ions and negative electrons are roughly equal in number. Hence, the space charge is essentially zero. 2. The region in which gaseous conduction takes place between the cathode and anode of an electric arc.

plasma diode—A thermodynamic engine with electrons as the working fluid, the potential energy of which is converted to a useful output.

plasma engine—A reaction engine using electrically accelerated plasma as the propellant.

plasma frequency—A natural frequency for coherent electron motion in a plasma.

plasma jet—A high-temperature stream of electrons and positive ions produced by the magnetohydrodynamic effect of a strong electrical discharge.

plasma length—*See* Debye Length.

plasma oscillation—Electrostatic or space-charge oscillations in a plasma which are closely related to the plasma frequency. There is usually enough damping due to electron collisions to prevent self-generation of the oscillations. They can be excited, however, by such techniques as shooting a modulated electron beam through the plasma.

plasma thermocouple—An electronic device in which the heat from nuclear fission is converted directly into electric power.

plasmatron—A helium-filled current amplifier that combines the grid-control characteristics and linearity of a vacuum triode with the extremely low internal impedance of a thyratron.

plasticizer—A substance added to a plastic to produce softness and adhesiveness in the finished product.

PLAT—Acronym for Pilot Landing Aid Television. A system in which television cameras cover aircraft landings on a carrier from several angles, allowing the landing personnel to "talk" the pilot down with increased precision. Recordings can be made for future reference.

plate—1. Preferably called the anode. The principal electrode to which the electron stream is attracted in an electron tube. 2. One of the conductive electrodes in a capacitor. 3. One of the electrodes in a storage battery. 4. *See* Printed-Circuit Board.

plateau—In the counting rate-versus-voltage characteristic of a radiation counter tube, that portion in which the counting rate is substantially independent of the applied voltage.

plateau length—The applied-voltage range over which the plateau of a radiation counter tube extends.

plate bypass capacitor—A capacitor connected between the plate and cathode of a vacuum tube to bypass high-frequency currents and thus keep them out of the load. *Also see* Anode Bypass Capacitor.

plate characteristic—A graph showing how changes in plate voltage affect the plate current of a vacuum tube.

plate circuit—The complete external electrical circuit between the plate and cathode of an electron tube.

plate-circuit detector—A detector that functions by virtue of its nonlinear plate-current characteristic.

plate conductance—The in-phase component of the alternating plate current divided by

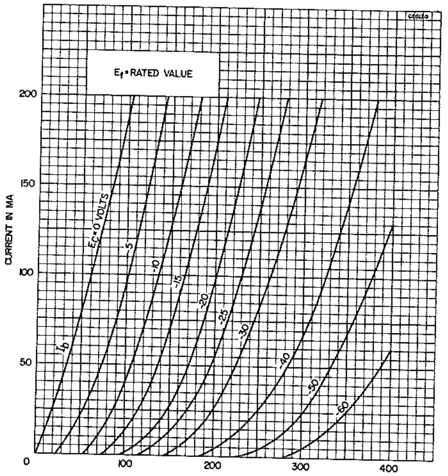

Plate characteristic curves.

the alternating plate voltage, all other electrode voltages being maintained constant.

plate current—Electron flow from the cathode to the plate inside an electron tube.

plate detection—The operation of a vacuum-tube detector at or near plate-current cutoff so that the input signal is rectified in the plate circuit.

plate dissipation—The amount of power lost as heat in the plate of a vacuum tube.

plated thru-hole—*See* Thru-Hole Connection.

plated-resist—A material electroplated on conductive areas to make them impervious to etching.

plate efficiency—Also called the anode efficiency. Ratio of load-circuit power (alternating current) to plate-power input (direct current).

plate impedance—Also called plate-load impedance. The total impedance between the anode and cathode, exclusive of the electron stream.

plate-input power—In the last stage of a transmitter, the direct plate voltage applied to the tubes times the total direct current flowing to their plates, measured without modulation.

plate keying—Keying done by interrupting the plate-supply circuit.

plate-load impedance—Also called the anode-load impedance. The total impedance between the anode and cathode of a vacuum tube, exclusive of the electron stream. *Also see* Plate Impedance.

plate modulation—Also called anode modulation. Modulation produced by applying the modulating voltage to the plate of any tube in which the carrier is present.

plate neutralization—Neutralizing an amplifier by shifting a portion of the plate-to-cathode a-c voltage 180° and applying it to the grid-to-cathode circuit through a capacitor.

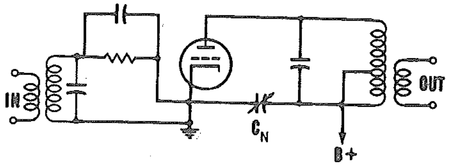

Plate neutralization.

plate neutralizer—*See* Anode Neutralization.

plate power input—Also called the anode power input. The d-c power (mean anode voltage times current) delivered to the plate (anode) of a vacuum tube.

plate power supply—*See* Anode Power Supply.

plate pulse modulation—Also called anode pulse modulation. Modulation produced in an amplifier or oscillator by applying externally generated pulses to the plate circuit.

plate resistance—The plate-voltage change divided by the resultant plate-current change in a vacuum tube, all other conditions being fixed.

plate saturation—Also called anode or voltage saturation. The point at which the plate current of a vacuum tube no longer increases as the plate voltage does.

plate supply—*See* Anode Supply.

plate-to-plate impedance—The load impedance between the two plates in a push-pull amplifier stage.

plate voltage—The d-c voltage between the plate and cathode of a vacuum tube.

plate winding—A transformer winding connected to the plate circuit of a vacuum tube.

plating resist—Any material which, when deposited on conductive areas, prevents the plating of the areas it covers.

platinotron—A cross-field vacuum tube used to generate and amplify microwave energy. It resembles the magnetron, except that it has no resonant circuit and has two external r-f connections instead of only one.

platinum—A heavy, almost white metal that resists practically all acids and is capable of withstanding high temperatures.

platinum contacts—Used where currents must be broken frequently (e.g., in induction coils and electric bells). Sparking does not damage platinum as much as it does other metals. Hence a cleaner contact is assured with minimum attention.

platter—A popular term for phonograph records and transcriptions.

playback—Reproduction of a recording.

playback head—A magnetic head used to pick up a signal from a magnetic tape.

playback loss—*See* Translation Loss.

plethysmogram—A graph of the changes in size or volume of an organ or of the amount of blood present or passing through a vessel. The transducer for a plethysmogram can be either a pressure transducer or a pair of surface electrodes.

pliodynatron—A four-element vacuum tube with an additional grid, which is maintained

at a higher voltage than the plate to obtain negative-resistance characteristics.

pliotron—An industrial-electronic term for a hot-cathode vacuum tube having one or more grids.

plotter—A device that produces an inscribed visual display of the variation of a dependent variable as a function of one or more other variables.

plotting board—A device which plots one or more variables against one or more other variables.

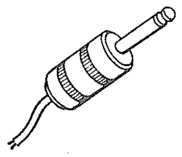

Plug.

plug—1. The part of the two mating halves of a connector which is movable when not fastened to the other mating half. The plug is usually thought of as the male portion of the connector. This is not always the case; the plug comprises contacts, insert, shell, and any special sealing or holding accessories. 2. A connector free to move so as to be plugged into a receptacle. (*Also see* Connector.)

plugboard—In a computer, a removable board having many electric terminals into which connecting cords may be plugged in patterns varying for different programs. To change the program, one wired plugboard is replaced by another.

plugboard computer—A computer which has a punch-card input and output, and to which program instructions are delivered by means of interconnecting patch cords on a removable plugboard.

pluggable unit—A chassis which can be removed from or inserted into the rest of the equipment by merely plugging in or pulling out a plug.

plugging—A control function which provides braking by reversing the motor line-voltage polarity or phase sequence so that the motor develops a countertorque which exerts a retarding force.

plug-in—Any device to which connections can be completed through pins, plugs, jacks, sockets, receptacles, or other ready connectors.

plug-in coil—A coil that can be easily interchanged and used for varying the tuning range of a receiver or transmitter. It is wound around a form often resembling an elongated tube base, with the coil leads connected to pins on the base.

plug-in device—A component or group of components and their circuitry which can be easily installed or removed from the equipment. Electrical connections are made by mating contacts.

plug-in unit—A standard subassembly of components that can be readily plugged into or pulled out of a circuit as a unit.

plumbicon—A vidicon with a lead-oxide target; its major advantage is its lack of image retention. It is a tube with the simplicity of a vidicon, and the sensitivity and lag of a glass target image orthicon. The tube is used for live black-and-white and color broadcasting.

plumbing—Waveguides and accessory equipment for radio-frequency transmission.

plume—The hot gaseous material ejected briefly from a highly absorbent material after bombardment by an intense laser pulse. The plume emits broad-band white light and is the most prominent feature in most irradiation experiments.

plunger—*See* Piston.

plunger relay—A relay consisting of a movable core or plunger surrounded by a coil. Solenoid action causes the plunger or core to move and thus energize the relay whenever current flows through the coil.

plunger-type instrument—A moving-iron instrument for measuring current. It consists of a pointer attached to a plunger inside a coil. The current being measured flows through the coil and pulls the plunger down. How far it goes into the coil depends on the magnitude of the current.

plutonium—A heavy element which undergoes fission when bombarded by neutrons. It is a useful fuel in nuclear reactors. Its symbol is *Pu;* its atomic number, 94.

pm—Abbreviation for phase modulation or permanent magnet.

pm erasing head—A head which uses the fields of one or more permanent magnets for erasing.

pm speaker—Abbreviation for permanent-magnet speaker.

PN boundary—The surface where the donor and acceptor concentrations are equal in the transition region between P- and N-type materials.

pneumatic bellows—A gas-filled bellows sometimes used to provide delay time in plunger-type relays.

pneumatic speaker—A speaker in which the acoustic output is produced by controlled variation of an air stream.

pneumogram—A graphic recording of respiratory movements and/or forces.

PN hook transistor—Also called hook transistor or hook collector transistor. A junction transistor which secures increased current amplification by means of an extra PN junction.

PNIN—A transistor in which its intrinsic region is between two N-regions.

PNIP transistor—An intrinsic-region transistor with an N-type base and a P-type emitter and collector.

PN junction—The region of transition between P-type and N-type material in a single semiconductor crystal.

PNPN (four-layer) diode—A semiconductor device which may be regarded as a two-transistor structure with two separate emitters feeding a common collector. This combination constitutes a feedback loop which is unstable for loop gains greater than unity. The instability results in a current which increases until ohmic circuit resistances limit the maximum value. This gives rise to a negative-resistance region which may be utilized for switching or for waveform generation.

PNPN transistor—A hook transistor with an N-type base, P-type emitter, and a hook collector. The electrodes are connected to the four end layers of the N- and P-type semiconductor materials.

PNPN-type switch—A bistable semiconductor device made up of three or more junctions, at least one of which is able to switch between reverse and forward voltage polarity within a single quadrant of the anode-to-cathode voltage-current characteristic.

PNP transistor—A transistor consisting of two P-type regions separated by an N-type region. When a small forward bias is applied to the first junction and a large reverse bias to the second junction, the system behaves much like a vacuum-tube triode.

Pockel's effect—The alternation in the refractive properties of a transparent piezoelectric crystal by the application of an electric field. *Also see* Modulator Crystal.

Pockel's-effect modulation—A phenomenon that occurs when a transparent dielectric is a piezoelectric crystal. The crystal tends to strain whenever an electric field is applied, rotating the plane of polarization of the incident wave. Some 7500 V/m cause a 90-degree rotation of light.

POGO — Acronym for Polar Orbiting Geophysical Observatory.

poid—The curve traced by the center of a sphere when it rolls or slides over a surface having a sinusoidal profile.

point—Called the binary point in binary notation, and the decimal point in decimal notation. In positional notation, the character or location of an implied symbol which separates the integral part of a number from its fractional part.

point availability—The percent of time an equipment is available for use when an operator requires it.

point-based linearity — Nonlinearity expressed as the deviation from a straight line which passes through a given point or points.

point contact—A pressure contact between a semiconductor body and a metallic point.

point-contact diode—A diode which obtains its rectifying characteristic from a point contact.

point-contact transistor—A transistor having a base electrode and two or more point-contact electrodes.

point defect—An imperfection caused by the presence of an extra atom or the absence of an atom from its proper place in the crystal.

point effect—The phenomenon whereby a discharge will occur more readily at sharp points than elsewhere on an object or electrode.

pointer—Also called a needle. A slender rod that moves over the scale of a meter.

point impedance—Ratio of the maximum E-field to the maximum H-field observed at a given point in a waveguide or transmission line.

point-junction transistor—A transistor having a base electrode and both point-contact and junction electrodes.

point-plane rectifier — *See* Glow-tube Rectifier.

point source—A radiation source whose dimensions are small compared with the distance from which it is observed.

point-to-point radio communication—Radio communication between two fixed stations.

point-to-point wiring—A method of forming circuit paths by connecting the various devices, components, modules, etc., with individual pieces of wire or ribbon. May be soldered, welded, or attached by other means.

point transposition—Transposition, usually in an open-wire line, which is executed within a distance comparable to the wire separation, without material distortion of the normal wire configuration outside this distance.

Poisson's ratio — The ratio of the lateral strain to the longitudinal strain in a specimen subjected to a longitudinal stress.

polar—Pertaining to, measured from, or having a pole (e.g., the poles of the earth or a magnet).

polar circuit—A teletypewriter circuit in which current flows in one direction on a marking impulse and in the opposite direction during a spacing impulse.

polar coordinates—A system of coordinates in which a point is located by its distance and direction (angle) from a fixed point on a reference line (called the polar axis).

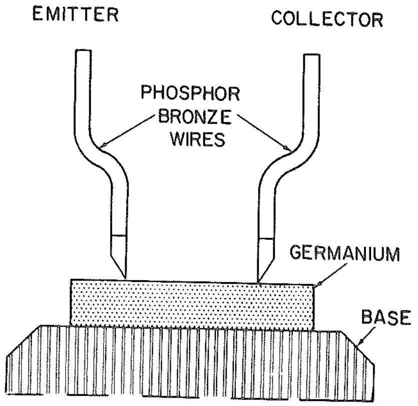

Point-contact transistor.

polar crystals—Crystals having a lattice composed of alternate positive and negative ions.

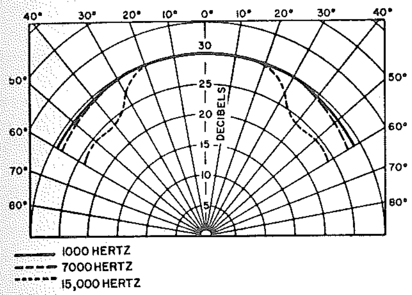

Polar diagram of a speaker enclosure.

polar diagram—A diagram in which the magnitude of a quantity is shown by polar coordinates.

polar grid—A type of circular grid on which range and azimuth are represented from a central reference point.

polarity—1. A condition by which the direction of the flow of current can be determined in an electrical circuit (usually batteries and other direct-voltage sources). 2. Having two opposite charges, one positive and the other negative. 3. Having two opposite magnetic poles, one north and the other south.

polarity of picture signal—Stated as black negative or black positive. The particular potential state of a portion of the signal representing a dark area of a scene relative to the potential representing a light area.

polarization—1. The process of making light or other radiation vibrate perpendicular to the ray. The vibrations are straight lines, circles, or ellipses—giving plane, circular, or elliptical polarization, respectively. 2. The increased resistance of an electrolytic cell as the potential of an electrode changes during electrolysis. In dry cells, this shortens their useful life. 3. The slight displacement of the positive charge in each atom whenever a dielectric is placed into an electric field. 4. The magnetic orientation of molecules in a piece of iron or other magnetizable material placed in a magnetic field, whereby the tiny internal magnets tend to line up with the magnetic lines of force. 5. The direction of the electric vector in a linearly polarized wave radiated from an antenna. 6. A mechanical arrangement of inserts and shell configuration (referred to as clocking in some instances) which prohibits the mating of mismatched plugs and receptacles.

polarization diversity—A term that designates a method of transmission and reception used to minimize the effects of selective fading of the horizontal and vertical components of a radio signal. It is usually accomplished through the use of separate vertically and horizontally polarized receiving antennas.

polarization-diversity antenna—An antenna in which any of a number of types of polarization can be readily selected. The polarization can be horizontal, vertical, right-hand circular, left-hand circular, or any combination of these four.

polarization-diversity reception — Diversity reception which uses separate vertically and horizontally polarized receiving antennas.

polarization error—In navigation, the error arising from the transmission or reception of radiation having other than the intended polarization for the system.

polarization fading—Fading as the result of changes in the direction of polarization in one or more of the propagation paths of waves arriving at a receiving point. *Also see* Faraday Effect.

polarization in a dielectric—The slight displacement of the positive charge in each atom whenever a dielectric is placed into an electric field.

polarization modulation — A technique in which modulation is produced by changing the direction of polarization of circularly polarized layer energy.

polarization receiving factor—Ratio of the power received by an antenna from a given plane wave of arbitrary polarization, to the power received by the same antenna from a plane wave of the same power density and direction of propagation whose state of polarization has been adjusted for the maximum received power.

polarization unit vector (for a field vector)—A complex field vector at a point, divided by the magnitude of the vector.

polarize—To cause to be polarized.

polarized double-biased relay — Also called magnetic-latch relay. A relay whose operation depends on the polarity of the energizing current and which is magnetically biased, or latched, in either of two positions. Its coil symbol is usually marked + and DB.

polarized light—Light which has the electric-field vector of all the energy vibrating in the same plane. Looking into the end of a beam of polarized light, one would see the electric-field vectors as parallel or coincident lines.

polarized no-bias relay—A three-position or a center-stable polarized relay. Its coil symbol is usually + and NB.

polarized plug—A plug so constructed that it may be inserted in its receptacle only in a predetermined position.

polarized receptacle — A receptacle into which a polarized plug can be inserted only in a predetermined position.

polarized relay—Also called a polar relay. A relay in which the armature movement depends on the direction of the current. Its coil symbol is sometimes marked +.

polarized-vane instrument—*See* Permanent-Magnet, Moving-Iron Instrument.

polarizer—A substance which, when added to an electrolyte, increases the polarization.

polarizing slots—Also called indexing slots. One or more slots placed in the edge of a printed-circuit board to accommodate and align certain types of connectors.

polar radiation pattern—The characteristic lobed pattern of varying field strength of radiation from a source.

polar relay—A relay containing a permanent magnet that centers the armature. The direction of movement of the armature is governed by the direction of the current. *Also see* Polarized Relay.

pole—1. One end of a magnet. 2. One electrode of a battery. 3. An output terminal on a switch.

pole face—In a relay, the end of the magnetic core nearest the armature.

pole piece—One or more pieces of ferromagnetic material forming one end of a magnet and so shaped that the distribution of the magnetic flux in the adjacent medium is appreciably controlled.

pole shoe—The portion of a field pole facing the armature of the machine. It may be separable from the body of the pole.

poles of a network function—Those real or complex values of p, for which the network function is infinite.

pole-type transformer—A transformer suitable for mounting on a pole or similar structure.

police calls—Broadcasts (usually orders) issued by police radio stations.

poling—1. The adjustment of polarity. Specifically, in wire-line practice, it signifies the use of transpositions between transposition sections of open wire or between lengths of cable, to cause the residual cross-talk couplings in individual sections or lengths to oppose one another. 2. A step in the production of ceramic piezoelectric bodies which orients the axes of the crystallites in the preferred direction. In general, a process similar to magnetizing ferromagnetic materials.

Polish notation—A system for writing logical and arithmetic expressions without the use of parentheses. So called because it was originated by Polish logician, J. Lukasiewicz.

polycrystalline structure—The granular structure of crystals which have nonuniform shapes and arrangements.

polyergic—A type of emission in which the groups of energies or velocities are produced simultaneously (e.g., simulated micrometeoroids in varying charge states, separated by velocity where accelerated by the same potential).

polyester backing—A plastic-film backing added to magnetic tape to make it stronger and more resistant to changes in humidity

polyesters—A class of thermosetting synthetic resins having great strength and good resistance to moisture and chemicals.

polyethylene—Short for polymerized ethylene, a tough, white plastic insulator with low moisture absorption. It is often used as a dielectric.

polygraph—Also called a lie detector. A recorder of several signals simultaneously, such as blood pressure, respiratory motion, galvanic skin resistance, etc., commonly used for study of emotional reactions involving deception (lie detection).

polyphase—Having or utilizing several phases. Thus, a polyphase motor operates from a power line having several phases of alternating current.

polyphase circuit—A group of alternating-current circuits (usually interconnected) which enter (or leave) a delimited region at more than two points of entry. They are intended to be so energized that, in the steady state, the alternating potential differences between them all have exactly equal periods, but have differences in phase, and may have differences in waveform.

polyphase motor—An induction motor wound for operation on two- or three-phase alternating current.

polyphase synchronous generator—A generator with its a-c circuits so arranged that two or more symmetrical alternating electromotive forces with definite phase relationships to each other are produced at its terminals.

polyphase transformer—A transformer designed for use in polyphase circuits.

polyphase voltages—In an a-c electrical system, voltages having a definite phase relationship to each other.

polyplexer—Radar equipment that combines the functions of duplexing and lobe switching.

polyrod antenna—An end-fire, dielectric, microwave antenna made of tapered polystyrene rods.

polystyrene—A clear thermoplastic material having excellent dielectric properties, especially at ultrahigh frequencies.

polystyrene capacitor—A low-loss precision capacitor with a polystyrene dielectric.

polyvinyl chloride—A general-purpose thermoplastic used for insulations and jackets on components, wire, and cable.

pool cathode—A cathode at which the principal source of electron emission is a cathode spot on a metallic-pool electrode.

pool tube—A gas tube with a pool cathode.

popi (Post Office Position Indicator)—A British long-distance navigational system for providing bearing information. It is a continuous-wave, low-frequency system in which the phase difference between sequential transmissions on a single frequency is measured.

population—The entire group of items being studied, from which samples are drawn. Sometimes called universe.

porcelain—A glazed ceramic insulating material made from clay, quartz, and feldspar.

porcelain capacitor—A fixed electrostatic capacitor, the dielectric of which is a high grade of porcelain, molecularly fused to alternate layers of fine silver electrodes so as to form a monolithic capacitor.

port—1. A place of access to a system or circuit. Through it, energy can be selectively supplied or withdrawn or measurements can be made. Examples are the port in a waveguide or in a bass-reflex speaker enclosure. 2. A fluid connection to the servovalve (e.g., supply port, return port, control port). 3. Any two input or output signal terminals.

portable recorder—1. A sound recorder built into a case so that it can be easily carried from place to place. 2. Any tape recorder designed for easy mobility, but one which may require connection to a 110-volt a-c supply for operation. Also applied to self-powered units which do not require external power for operation.

portable standard meter—A portable meter used principally as a standard for testing other meters.

portable transmitter—A transmitter which can be readily carried on one's person and operated while in motion (e.g., Walkie-Talkies, Handy-Talkies, and similar personal transmitters). (*Also see* Transportable Transmitter.)

portamento—The continuous change of a tone from one pitch to another. (*Also see* Glissando.)

pos—Abbreviation for positive.

posistor—A thermally sensitive resistor which has a positive temperature characteristic of resistance.

position—The location of an object with respect to a specific reference point or points.

positional cross talk—In a multibeam cathode-ray tube, the deviation of an electron beam from its path under the influence of another electron beam within the tube.

positional notation—One of the schemes for representing numbers. It is characterized by the arrangement of digits in sequence, with successive digits forming coefficients of successive powers of an integer called the base of the number system.

position-changing mechanism—The mechanism used to move a removable circuit-breaker unit to and from the connected, disconnected, and test positions.

position feedback—A feedback signal which is proportional to the position or deflection of some object.

position of effective short—The distance between a specified reference plane and the apparent location of the short circuit of a fired switching tube in its mount.

position-type telemeter—*See* Ratio-type Telemeter.

positive—Any point to which electrons are attracted—as opposed to negative, from where they come.

positive bias—The condition in which the control grid of a vacuum tube is more positive than the cathode.

positive charge—An electrical charge with fewer electrons than normal.

positive column—The luminous glow, often striated, between the Faraday dark space and the anode in a glow-discharge, cold-cathode tube.

positive electricity—1. The electricity which predominates in a glass body after it has been electrified by rubbing with silk. 2. *See* Positive Charge.

positive electrode—The conductor that is connected to the positive terminal of a primary cell and serves as the anode when the cell is discharging. Electrons flow to it through the external circuit.

positive electron—*See* Positron.

positive feedback—Also see regeneration. The process by which the amplification is increased by having part of the power in the output circuit returned to the input circuit in order to reinforce the input power.

positive ghost—A television ghost-signal display with the same tonal variations as those of the image.

positive-going—Increasing toward a positive direction (e.g., a current or waveform).

positive grid—A grid with a more positive potential than the cathode in a vacuum tube.

positive-grid multivibrator—A multivibrator which has one or more grids connected to the plate-voltage supply, usually through a large resistance.

positive-grid oscillator—*See* Retarding-Field Oscillator.

positive-grid oscillator tube—Also called a Barkhausen tube. An oscillating triode in which the grid has a more positive quiescent voltage than either of the other electrodes.

positive ion—An atom which has lost one or more electrons and thus has an excess of protons, giving it a positive charge.

positive-ion emission—Thermionic emission of positive particles from the cathode of a vacuum tube. They either are made up of ions from the metal in the cathode, or are due to some impurity in it.

positive-ion sheath—A collection of positive ions on the control grid of a gas-filled triode tube. If too high a negative bias is applied to the grid, this positive sheath will block the plate current.

positive light modulation—Modulation in which the transmitted power increases as the light intensity does, and vice versa.

positive magnetostriction—Magnetostriction in which a material expands whenever a magnetic field is applied.

positive modulation—Also called positive picture modulation. In an a-m television system, modulation in which the brightness increases as the transmitted power does, and vice versa.

positive phase-sequence relay—A relay which is energized by the positive phase-sequence component of the current, voltage, or power of a circuit.

positive picture modulation—*See* Positive Modulation.

positive picture phase—The condition in which the picture-signal voltage goes positive above the zero level whenever a positive scene or picture increases in brilliance.

positive plate—1. A hollow lead grid filled with active material and connected to the positive terminal of a storage battery. When the battery is discharging, electrons flow toward it through the external circuit. 2. In a capacitor, the temporarily charged condition whereby one of the plates has fewer electrons.

positive ray—*See* Canal Ray.

positive temperature coefficient—The condition whereby the resistance, capacitance, length, or other characteristic of a substance increases as the temperature does.

positive terminal—In a battery or other voltage source, the terminal toward which electrons flow through the external circuit from the the negative terminal.

positive transmission—Transmission of television signals in such a way that the transmitted power increases whenever the initial light intensity does.

positron—Also called a positive electron. An unstable particle having the same weight as an electron but the opposite charge (positive rather than negative). Its existence was predicted years before it could be proved.

post—In a computer, to place a unit of information on a record.

post-accelerating electrode—*See* Intensifier Electrode.

post acceleration—Also called post-deflection acceleration. Acceleration of the beam electrons in a tube after they have been deflected.

post-conversion bandwidth—In a telemetry receiver, the bandwidth presented to the detector.

post-deflection accelerating electrode—*See* Post-Accelerating Electrode.

post-deflection acceleration—*See* Post-Acceleration.

post-edit—In a computer, to edit output data resulting from a previous computation.

post-emphasis—*See* De-emphasis.

post-equalization—*See* De-emphasis.

post-mortem—A diagnostic computer routine for locating a malfunction in the computer or an error in coding a problem. Should a problem tape come to a standstill, the computer will print out—either automatically or when called for—any information concerning the contents of all or part of the registers in the computer.

postregulator—A circuit that performs the functions of reference, comparison, and control in power supplies. So called because it follows the transformer, rectifier, and (usually) the ripple filter.

pot—Short for potentiometer.

potassium—An alkali metal having photosensitive characteristics, especially to blue light. It is used on the cathodes of phototubes whenever maximum response to blue light is desired.

pot core—A magnetic structure consisting of a rod and a sleeve arranged so that the rod fits inside a coil and the sleeve fits around the coil. The sleeve and rod are connected at one end by a plate. The open end (opposite the plate) is usually, by not necessarily, ground so that two pot cores or a pot core and a separate plate can be put together around a suitable coil to form a low-reluctance magnetic path and/or shield for the coil.

potential—The difference in voltage between two points of a circuit. Frequently one point is assumed to be ground, which has zero potential.

potential barrier—A semiconductor region through which electric charges attempting to pass will encounter opposition and may be turned back.

potential coil—The shunt coil in a measuring instrument or other device having series and shunt coils—i.e., the coil connected across the circuit and affected by changes in voltage.

potential difference—1. The algebraic difference between the individual potentials of two points. 2. A voltage existing between two points (e.g., the voltage drop across an impedance, from one end to another).

potential divider—*See* Voltage Divider.

potential drop—The difference in potential between the two ends of a resistance with a current flowing through it.

potential energy—Energy due to the position of one body with respect to another or to the relative parts of the same body.

potential galvanometer—A galvanometer with such a high resistance that it takes practically no current. It has been replaced by the vacuum-tube voltmeter.

potential gradient—1. The differences in value of the potential per unit length along a conductor or through a dielectric. 2. The rate of change of potential with distance. Units such as volts per meter or kilovolts per centimeter may be used.

potential transformer—Also called a voltage transformer. An instrument transformer, the primary winding of which is connected in parallel with the circuit whose voltage is to be measured or controlled.

potentiometer—1. An electromechanical device having a terminal connected to each end of the resistive element, and a third terminal connected to the wiper contact. The electrical input is divided as the contact moves over the element, thus making it possible to mechanically change the resistance. 2. A unit for measuring potential difference. It uses a calibrated voltage divider, a reference potential, and a null detector.

potentiometer circuit—A network arranged so that, when two or more electromotive forces (or potential differences) are present in as many branches, the response of a suitable detecting device in any branch can be made zero by adjusting the electrical constants of the network.

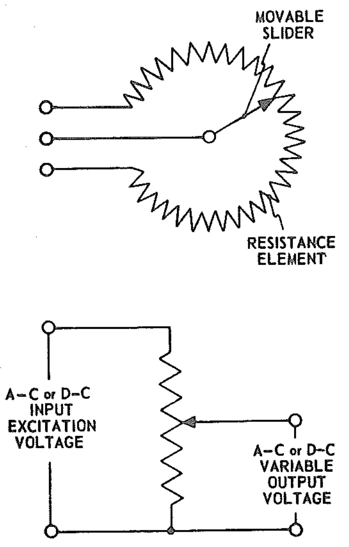
MOVABLE SLIDER
RESISTANCE ELEMENT
A-C or D-C INPUT EXCITATION VOLTAGE
A-C or D-C VARIABLE OUTPUT VOLTAGE

Potentiometer.

potentiometer recorder—A null-balance type of recorder using a servo-operated voltage-balancing device; the sliding contact of a precision measuring potentiometer is adjusted automatically by a servomechanism so that the difference in voltage of the circuit becomes zero. Main feature is high sensitivity.

potentiometric transducer—A transducer in which displacement of a force-summing member is transmitted to the slider in a potentiometer, thus changing the ratio of output resistance to total resistance. Transduction is accomplished by changing the ratios of a voltage divider.

potentiometric transduction—The conversion of the measurand into a change in the position of a contact on a resistance element across which excitation is applied, the output usually being given as a voltage ratio.

pothead—An insulator for making a sealed joint between an underground cable and an overhead line.

Potier diagram—A vector diagram showing the voltage and current relationships in an a-c generator.

pot life—The period after the addition of a catalyst to a potting compound during which the potting operation must be completed.

potted circuit—A circuit which has been encapsulated in a nonconductive material.

potted line—A pulse-forming network immersed in oil and enclosed in a metal container.

Potter oscillator—A cathode-coupled multivibrator.

potting—An embedding process for parts that are assembled in a container or can into which the insulating material is poured, with the container remaining an integral part as the outer surface of the finished unit.

potting compound—A sealing material used to fill the case or enclosure in which a component is contained.

powdered-iron core—A core consisting of fine particles of magnetic material mixed with a suitable bonding material and pressed into shape.

power—The rate at which work is done. Units of power are: the watt, the joule, and the kilowatt.

power amplification—*See* Power Gain.

power amplifier—An amplifier intended for driving one or more speakers or other transducers.

power-amplifier stage—1. An audio-frequency amplifier stage capable of handling considerable audio-frequency power without distortion. 2. A radio-frequency amplifier stage used in a transmitter primarily to increase the power of the carrier signal.

power attenuation—*See* Power Loss.

power cord—*See* Line Cord.

power density—The radiated field strength set up by a radiating source, expressed in microvolts per meter or in db above 1 μv/m.

power derating—Use of computed curves to determine the correct power rating of a device or component to be used above its reference ambient temperature.

power detection—Detection in which the power output of the detector is used for supplying a substantial amount of power directly to a device such as a speaker or recorder.

power detector—A vacuum tube detector operating with such a high plate voltage that strong input signals can be handled without appreciable distortion.

power dissipation—The dispersion of the heat generated within a device or component when a current flows through it. This is accomplished by convection to the air, radiation to the surroundings, or conduction.

power dissipation rating—The maximum average power that can be continuously dissipated under stated temperature conditions.

power dump—The removal of all power accidentally or intentionally. *Also see* Dump.

power factor—1. Ratio of the actual power of an alternating or pulsating current, as measured by a wattmeter, to the apparent power, as indicated by an ammeter and voltmeter. 2. Ratio of resistance to impedance—therefore, a measure of the loss in an inductor, capacitor, or insulator. 3. The cosine of the phase angle between the voltage applied to a load and the current passing through it. (Sometimes the cosine is multiplied by 100 and expressed as a percentage.)

power-factor correction—Adding capacitors to an inductive circuit in order to increase the power factor by making the total current more nearly in phase with the applied voltage.

power-factor meter—A direct-reading instrument for measuring power factor. Its scale is graduated directly in power factor.

power-factor regulator—A regulator which maintains the power factor of a line or apparatus at a predetermined value, or varies it according to a predetermined plan.

power-factor relay—A device which operates when the power factor in an a-c circuit becomes above or below a predetermined value.

power frequency—The frequency at which electric power is generated and distributed. Throughout most of the United States, this frequency is 60 Hz.

power gain—1. Also called power amplification. Of an amplifying device, the ratio of power delivered to a specified load impedance to the power absorbed by its input. 2. Of an antenna in a given direction, 4π times the ratio of the radiation intensity to the total power delivered to the antenna. (The term is also applied to receiving antennas.) 3. Of a transistor, the ratio of output power to signal input power. (Not to be confused with collector efficiency.)

power ground—1. The ground between units which is part of the circuit for the main source of power to, or from, these units. 2. The potential of the terminal or circuit point to which the output of a power supply and often an amplifier output load is returned (i.e., power-supply "zero").

power level—At any point in a transmission system, the difference between the measure of the steady-state power at that point, and the measure of an arbitrarily specified amount of power chosen as a reference.

power-level indicator — An a-c voltmeter calibrated to read the audio power level.

power line—Two or more wires conducting electric power from one location to another.

power loss—Also called power attenuation. Ratio of the power absorbed by the input circuit of a transducer, to the power delivered to a specified load under specified operating conditions.

power modulation factor—Ratio of the maximum positive departure of the envelope of an amplitude-modulation wave from its average value to its average value. This rating is used when the modulating signal wave has unequal positive and negative peaks.

power output—The power in watts delivered by a power amplifier to a load such as a speaker.

power pack—A unit for converting power from an alternating- or direct-current supply into alternating- or direct-current power at voltages suitable for supplying an electronic device.

power programmer—A device for controlling the output power of a radar automatically as a function of the target range.

power rating—The maximum power that can be dissipated in a component or device for a specified period.

power ratio—Ratio of the power output to the power input of a device. Usually expressed as the number of decibels loss or gain.

power-rectifier misfire relay—A device which functions if one or more of the power-rectifier anodes fails to fire.

power relay—A relay that functions at a predetermined value of power. It may be an overpower relay, an underpower relay, or a combination of both.

power response—The frequency-response capabilities of an amplifier running at or near its full rated power.

power supply—A unit that supplies electrical power to another unit. It changes ac to dc and maintains a constant voltage output within limits.

power switch—Often called an on-off switch. The switch that connects or disconnects a radio receiver, transmitter, or other equipment from its power line.

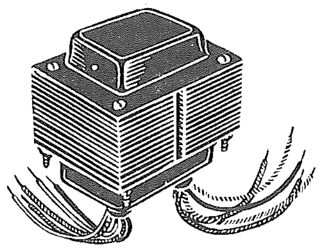

Power transformer.

power transformer—A transformer used for raising or lowering the supply voltage to the various values required by vacuum-tube plate, heater, and bias circuits.

power transistor—A transistor, usually an alloy-junction type, capable of handling high current and power.

power tube—An electron tube designed to handle more current and power than a voltage-amplifier tube.

power winding — A saturable-reactor winding to which the power to be controlled is supplied. Commonly, the output and power are furnished by the same winding, then termed the output winding.

Poynting's law—The transfer of energy can be expressed as the product of the values of the magnetic field and of the components of the electric field which are perpendicular to the magnetic field, and the flow of energy at any point is perpendicular to both fields.

Poynting's theorem—The rate of flow of electromagnetic energy into or out of a closed region is at any instant proportional to the surface integral of the vector product of the electric and magnetic intensities.

Poynting's vector—The vector product of the electric and magnetic intensities at one point and at a given instant in a wave.

ppi—Abbreviation for plan-position indicator.

ppi scope—A cathode-ray oscilloscope arranged to present a ppi display.

PP junction—A region of transition between two regions having different properties in a P-type semiconducting material.

ppm—1. Abbreviation for pulse-position modulation. 2. Abbreviation for parts per million.

ppm/am—Amplitude modulation of a carrier by pulses which are position modulated by data.

pps—Abbreviation for pulses per second.

practical system of electrical units—A system in which the units are multiples or submultiples of the units of the centimeter-gram-second electromagnetic system.

praetersonic—The higher region of the sonic spectrum.

preamplifier—An amplifier which primarily raises the output of a low-level source so that the signal may be further processed without appreciable degradation in the signal-to-noise ratio. A preamplifier may also include provision for equalizing and/or mixing.

preburning—Stabilizing tubes by operating their heaters continuously for a given number of hours. Cathode current may be drawn and the tubes vibrated at the same time.

precession — The effect resulting when a torque is applied to a rotating body, such as a gyroscope, causing it to wobble. The wobbling frequency is determined by the gravitational field strength and the mass of the body.

precipitation noise—Noise generated in an antenna circuit, generally in the form of a relaxation oscillation, caused by the periodic discharge of the antenna or conductors in the vicinity of the antenna into the atmosphere.

precipitation static—A type of interference experienced in a receiver during snow, rain, and dust storms. Often caused by the impact of dust particles against the antenna or the creation of induction fields by nearby corona discharges.

precipitator—Sometimes called a precipitron. An apparatus for removing small particles of smoke, dust, oil, mist, etc., from the air by electrostatic precipitation.

precipitron—*See* Precipitator.

precision—The quality of being sharply or exactly defined—i.e., the number of distinguishable alternatives from which a representation was selected. This is sometimes indicated by the number of significant digits the representation contains. (*Also see* Accuracy.)

precision approach radar — Abbreviated PAR. An airfield radar system aimed along the approach path of an aircraft and used for its guidance during approach.

precision comparator—A high-gain amplifier circuit the output of which changes decisively between two definite levels whenever the sum of the input voltages changes sign.

precision device — A device that operates within prescribed limits and will consistently repeat operations within those limits.

precision gate—A circuit that may be switched from closed to open circuit or vice versa without error (time, bias, impedance) in response to a command signal (voltage or current).

precision net—In a four-wire terminating set or similar device employing a hybrid coil, an artificial line designed and adjusted to provide an accurate balance for the loop and subscriber's set or line impedance.

precision potentiometer—A mechanical electrical transducer dependent upon the relative position of a moving contact (wiper) and a resistance element for its operation. It delivers a voltage output that is some specified function of the applied voltage and shaft position, to a high degree of accuracy.

precision snap-acting switch—An electromechanical switch having predetermined and accurately controlled characteristics and having a spring-loaded quick make-and-break contact action.

precision sweep—A delayed expanded radar sweep for high resolution and range accuracy.

preconduction current—The low value of plate current that flows in a thyratron or other grid-controlled gas tube prior to conduction.

precursor—Also called undershoot. The initial transient response to a unidirectional change in input. It precedes the main transition and is opposite in sense.

predetection recording—The recording of a telemetry receiver intermediate-frequency signals.

predictive control—A type of computer con-

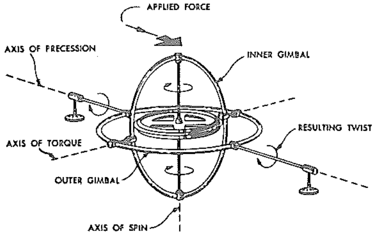

Precession of a gyroscope.

trol which allows a digital computer to include a dynamic control loop for repetitive comparison of pertinent factors.

predissociation—The dissociation that occurs in a molecule that has absorbed energy before it has had an opportunity to lose energy by radiation.

predistortion—*See* Pre-emphasis.

pre-emphasis—1. In a system, a process designed to emphasize the magnitude of some of the frequency components. Pre-emphasis is applied at the transmitting end (with de-emphasis at the receiving end) in order to improve the signal-to-noise ratio. 2. Also called pre-equalization or predistortion. In recording, an arbitrary change in the frequency response from its basic response (e.g., constant velocity or amplitude) in order to improve the signal-to-noise ratio or to reduce distortion. (*Also see* Accentuation.)

pre-emphasis network—A network inserted into a system to emphasize one range of frequencies.

pre-equalization—*See* Pre-emphasis.

preferred tube types—Tube types recommended to designers of electronic equipment, to minimize the number of tube types that must be stocked by the manufacturer or by service agencies.

preferred values—A series of resistor and capacitor values adopted by the EIA and military. In this system, the increase between any two steps is the same percentage as between all other steps. Increases may be in steps of 20%, 10%, or 5% each.

prefix multipliers—Prefixes which designate a greater or smaller unit than the original, by the factor indicated. These prefixes are:

Prefix	*Symbol*	*Factor*
tera	T	10^{12}
giga	G	10^{9}
mega	M	10^{6}
kilo	k	10^{3}
hecto	h	10^{2}
deka	da	10
deci	d	10^{-1}
centi	c	10^{-2}
milli	m	10^{-3}
micro	μ	10^{-6}
nano	n	10^{-9}
pico	p	10^{-12}
femto	f	10^{-15}
atto	a	10^{-18}

preform—Also called a biscuit. In disc recording, the small slab of record material used in the presses.

p+ region—The region created by diffusing into a silicon crystal a group-III element, which creates a deficiency of electrons or an excess of holes.

preliminary contacts — In a relay, contacts which open or close before other contacts when the relay is actuated.

prerecorded tape—*See* Recorded Tape.

preselection — 1. The use of a preselector. 2. In buffered computers, a time-saving technique in which a block of information is read into the computer memory ahead of time from whichever input tape will next be called on.

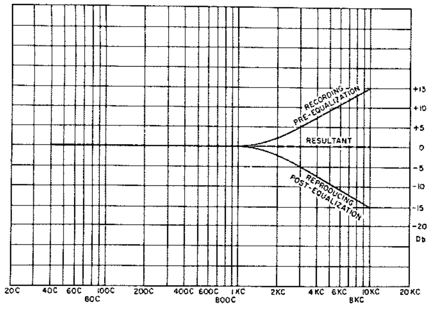

Pre- and post-emphasis curves.

preselector—1. A device placed ahead of a frequency converter or other device to pass signals of desired frequencies but reduce all others. 2. In automatic switching, a device which makes its selection before seizing an idle trunk.

preselector stage—A radio-frequency amplifier stage in the input of a superheterodyne receiver.

presence—The quality of naturalness in sound reproduction. When the presence of a system is good, the illusion is that the sounds are being produced intimately at the speaker.

presence control—A potentiometer used in a three-way speaker system for controlling the volume of the middle-range speaker.

presentation—The form which the radar echo signals take on the screen, depending on the nature of the sweep circuit.

presenting—Displaying data in a form which human intelligence can comprehend and use.

preset guidance system—A guidance system in which the flight path is determined before the missile is launched and cannot be altered after launch.

preshoot—The initial transient response to a unidirectional change in input which precedes the main transmission and may be of the same or opposite polarity.

pressed stem—An obsolete method of vacuum-tube construction in which all support wires are formed into a flattened piece of glass tubing (actually a relic from the lampmaker's era). (*Also see* Button Stem.)

pressing—A disc recording produced in a record-molding press from a master or stamper.

press-to-talk switch—Also called a push-to-talk switch. A spring-loaded switch that must be held down as long as the operator talks. Releasing the switch deactivates the microphone. It is used on transmitter and dictating-machine microphones.

pressure—1. Force per unit area. Measured in pounds per square inch (psi), or by the height (in feet, inches, or centimeters) of a column of water or mercury which the force will support. 2. Absolute pressure is measured with respect to zero pressure. Gage pressure is measured with respect to atmospheric pressure.

pressure amplitude—For a sinusoidal sound wave, the maximum absolute value of the instantaneous sound pressure at a point during any given cycle. The unit is the dyne per square centimeter.

pressure connector—A conductor terminal applied under pressure to make the connection mechanically and electrically more secure.

pressure-gradient hydrophone — A type of hydrophone in which the electric output is essentially determined by a component of the gradient (space derivative) of the sound pressure.

pressure hydrophone—A type of hydrophone in which the electric output is essentially determined by the instantaneous sound pressure of the impressed sound wave.

pressure microphone — A microphone in which the electric output corresponds substantially to the instantaneous sound pressure of the impressed sound waves. It is a gradient microphone of zero order, and is nondirectional when its dimensions are smaller than a wavelength.

pressure pad—In single-motor tape recorders, a device which forces the tape into intimate contact with the head gap, usually by direct pressure at the head assembly. Felt or similar material, occasionally protected with self-lubricating plastic, is used to apply pressure uniformly and with a minimum of drag to the backing side of the tape.

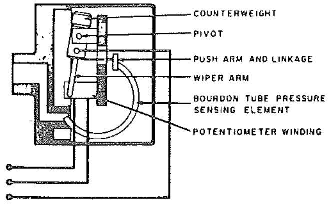

Pressure potentiometer.

pressure potentiometer—A pressure transducer in which the electrical output is derived by moving a contact arm along a resistance element.

pressure roller—Also called pinch roller, puck, or capstan idler. A spring-loaded rubber-tired roller which holds the magnetic tape tightly against the capstan, permitting the latter to draw the tape off the stock reel and past the heads at a constant speed.

pressure-sensing element — In a pressure transducer, the part which converts the measured pressure into mechanical motion.

pressure spectrum level — The effective sound-pressure level for the sound energy contained within a band one hertz wide and centered at a specified frequency. Ordinarily this level is not significant except for sound having a continuous distribution of energy within the frequency range under consideration.

pressure switch—A switch actuated by a change in the pressure of a gas or liquid.

pressure transducer—An instrument which converts a static or dynamic pressure input into the proportionate electrical output.

pressure-type capacitor—A fixed or variable capacitor used chiefly in transmitters. It is mounted inside a metal tank filled with nitrogen at a pressure that may be as great as 300 pounds per square inch. The high pressure permits a voltage rating several times that of air.

pressurization—The process by which the critical parts of equipment designed for high-altitude operation is surrounded with dry air or an inert gas under pressure (about five pounds per square inch at sea level). Thus, breakdowns from the impaired insulating properties of air at reduced pressure are prevented.

prestore—To store a quantity in an available or convenient location in a computer before it is required in a routine.

pretersonics—Sonic waves of frequencies higher than 500 MHz.

pretravel—The distance or angle through which the actuator moves from the actuator free position to the actuator operating position.

pretrigger—A timed pulse used to start a sequence of operation prior to the main trigger.

pre-TR tube—A gas-filled radio-frequency switching tube used to protect the TR tube from excessive power and the receiver from frequencies other than the fundamental.

preventive maintenance — Precautionary measures taken on a system to forestall failures rather than to eliminate them after they have occurred.

prf—Abbreviation for pulse-repetition frequency.

pri—Abbreviation for primary.

primaries—*See* Primary Colors.

primary—Abbreviated pri and symbolized by P. Also called a primary winding. A transformer winding that carries current and normally sets up a current in one or more secondary windings.

primary area—*See* Primary Service Area.

primary battery—A battery consisting of primary cells.

primary breakdown—Also called avalanche breakdown. The sustaining mode of a transistor—unlike second breakdown, not a failure mode. The transistor collector-to-emitter voltage is relatively constant for different collector supply voltages.

primary calibration—Calibration in which the transducer output is observed or recorded, while direct stimulus is applied under controlled conditions.

primary-carrier flow—Also called primary flow. The current flow responsible for the major properties of a semiconductor device.

primary cell—A cell that produces electric current through an electrochemical reaction but is not rechargeable. Once discharged, it must be discarded.

primary circuit—The first, in electrical order, of two or more coupled circuits, wherein a change in current will induce a voltage in the other, or secondary, circuit.

primary colors—Also called primaries. A set of colors from which all other colors are derived; hence, any set of stimuli from which all colors may be produced by mixture. A primary color cannot be matched by any combination of other primaries. In color television, the primary colors are red, blue, and green.

primary current—The current flowing through the primary winding of a transformer. Changes in this current cause a voltage to be induced in the secondary winding of the transformer.

primary detector—Also called a sensing, primary, or initial element. The first system element or group of elements that respond quantitatively to the measurand and perform the initial measurement operation. A primary detector performs the initial conversion or control of measurement energy. It does not include those transformers, amplifiers, shunts, resistors, etc., used as auxiliary means.

primary electron—1. After a collision between two electrons, the one with the greater energy. The other is called the secondary electron. 2. The electron produced in a detector or counter tube after ionization.

primary element—*See* Primary Detector.

primary emission—Emission of electrons due to primary causes (e.g., heating of a cathode) rather than secondary effects (e.g., electron bombardment).

primary failure—A failure occurring under normal environmental conditions and having no significant relationship to a previous failure but whose occurrence imposes abnormal stress on some other part or parts which may then undergo a secondary failure.

primary fault—In an electric circuit, initial breakdown of the insulation of a conductor, usually followed by a flow of power current.

primary flow—*See* Primary-Carrier Flow.

primary frequency—The frequency assigned for normal use on a particular circuit or communications channel.

primary grid emission—*See* Thermionic Grid Emission.

primary ionizing event—*See* Initial Ionizing Event.

primary power cable—Power service cables connecting the outside power source to the main-office switch and metering equipment.

primary radar—*See* Radar.

primary radiation—Radiation direct from the source without interaction.

primary radiator—The antenna element from which the radiated energy leaves the transmission system.

primary relay—A relay that produces the initial action in a sequence of operations.

primary service area—Also called the primary area. The area within which radio or tv reception is not normally subject to objectionable interference or fading.

primary skip zone—The area beyond the ground-wave range around a transmitter, but within the skip distance. Radio reception is possible in this zone by sporadic and zigzag reflections.

primary standard—A unit directly defined and established by some authority, and against which all secondary standards are calibrated.

primary storage—In a computer, the main internal storage.

primary voltage—The voltage applied to the terminals of the primary winding in a transformer.

primary winding—*See* Primary.

priming illumination—A small, steady illumination applied to a phototube or photoelectric cell to make it more sensitive to variations in the illumination being measured.

primitive period—The smallest increment of time during which a quantity repeats itself.

principal axis—A reference direction for angular coordinates, used in describing the directional characteristics of a transducer employed for sound emission or reception. It is usually an axis of structural symmetry or the direction of maximum response. If these two do not coincide, however, the reference direction must be described explicitly.

principal E-plane—The plane containing the direction of maximum radiation and in which the electric vector lies.

principal focus—For a lens or spherical mirror, the point of convergence of light coming from a source at an infinite distance.

principal H-plane—A plane containing the direction of maximum radiation; the electric vector is everywhere normal to the plane, and the magnetic vector lies in it.

principal mode—*See* Dominant Mode.

printed—Reproduced on a surface by some process (e.g., letterpress, lithography, silk screen, etching).

printed circuit—A circuit in which the interconnecting wires have been replaced by conductive strips printed, etched, etc., onto an insulating board. It may also include similarly formed components on the baseboard.

printed-circuit assembly—1. A printed-circuit board to which separable components have been attached. 2. An assembly of one or more printed-circuit boards, which may include several components.

printed-circuit board—Also called a card, chassis, or plate. An insulating board onto which a circuit has been printed. (*Also see* Printed Circuit.)

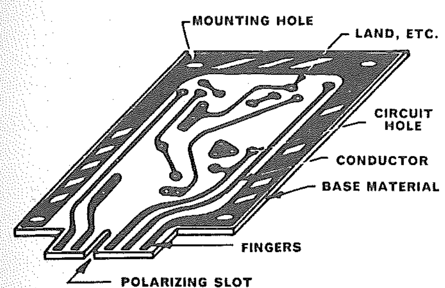

Printed circuit.

printed-circuit switch—A special rotary switch which can be connected directly to a mating printed-circuit board without wires.

printed component—A type of printed circuit intended primarily for electrical and/or magnetic functions other than point-to-point connections or shielding (e.g., printed inductor, resistor, capacitor, transmission line, etc.).

printed contact—The portion of a printed circuit that connects the circuit to a plug-in receptacle and performs the function of a plug pin.

printed element—An element, such as a resistor, capacitor, or transmission line, that is formed on a circuit board by deposition, etching, etc.

printed wiring—A conductive pattern formed on the surface of an insulating baseboard by plating or etching.

printed-wiring substrate—A conductive pattern printed on a substrate.

printer—1. Also called a teleprinter and teletypewriter. A telegraph instrument with a signal-actuated mechanism for automatically typing received messages. It may have a keyboard similar to that of a typewriter for sending messages. The term "receiving-only" is applied to a printer with no keyboard. 2. A device that prints the output from a computer. It ranges from a conventional "single-stick" or "flying typebar" printer (as on a typewriter), to a high-speed unit that prints up to 1,000 lines per minute!

printer telegraph code—A five- or seven-unit code used for operation of a teleprinter, teletypewriter, or similar telegraph printer.

printing—The reproduction of a pattern on a surface by any of various processes, such as vapor deposition, photo etching, embossing, or diffusion.

printing demand meter—An integrated demand meter which prints on a paper tape the demand for each interval and indicates the time it occurred.

printing telegraphy—Telegraph operation in which the received signals are automatically recorded as printed characters.

print-through—The transfer of the magnetic field from layer to layer of tape on the take-up reel.

print wheel—In a wheel printer, the single element providing the character set at one printing position.

privacy system—In radio transmission, a system designed to make unauthorized reception difficult.

private-aircraft station—A mobile radio station on board an aircraft not operated as an air carrier.

probability distribution—A mathematical model showing a representation of the probabilities for all possible values of a given random variable.

probability of success—The likelihood that an article will function satisfactorily for a stated period of time when subjected to a specified environment.

probable error—The amount of error which, according to the laws of probability, is most likely to occur during a measurement.

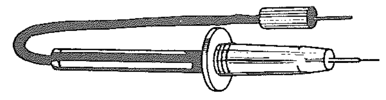

Probe (high voltage).

probe—1. A resonant conductor which can be placed into a waveguide or cavity resonator to insert or withdraw electromagnetic energy. 2. A test lead which contains an active or passive network and is used with certain types of test equipment. 3. A rod placed into the slotted section of a transmission line to measure the standing-wave ratio or to inject or extract a signal.

probing (near field)—The determination of radio interference by obtaining the relative interference level in the immediate area of a source by the use of a small insensitive antenna in conjunction with a receiving device.

problem language—The language a computer programmer uses in stating the definition of a problem.

problem-oriented language—In a computer, a source language suited to the description of a specific class of problems.

procedure-oriented language—In a computer, a source language suited to describing procedural steps in machine computing.

process control—Automatic control of industrial processes in which continuous material or energy is produced.

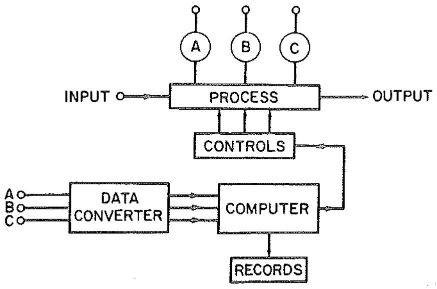

Process-control system.

processing section—The portion of a computer that does the actual changing of input into output. This includes the arithmetic and logic sections.

processor—1. A generalized program to process a group of computer programs. 2. Electronic equipment which is used to reformat, convert, translate, edit, or pulse-shape signals or data to satisfy the requirements of other equipment such as a computer.

producer's reliability risk—The risk faced by the producer (usually set at 10%) that a product will be rejected by a reliability-acceptance test even though the product is actually equal to or better than a specified value of reliability.

production sampling tests—Those tests normally made by either the vendor or the purchaser on a portion of a production lot for the purpose of determining the general performance level.

production tests—Those tests normally made on 100% of the items in a production lot by the vendor and normally on a sampling basis by the purchaser.

product modulator—A modulator, the output of which is substantially equal to the carrier times the modulating wave.

profile chart—A vertical cross-sectional drawing of the microwave path between two stations. Terrain, obstructions, antenna-height requirements, etc., are indicated on the drawing.

program—1. A set of instructions arranged in proper sequence for directing a digital computer in performing a desired operation or operations (e.g., the solution of a mathematical problem or the collation of a set of data). 2. To prepare a program (as contrasted with "to code"). 3. A sequence of audio signals transmitted for entertainment or information.

program assembly—Also called a translator. A process which translates a symbolic program into a machine-language program before the working program is executed. It can also integrate several sections or different programs.

program circuit—A telephone circuit that has been equalized to handle a wider range of frequencies that ordinary speech signals require. In this way, musical programs can be transmitted over telephone wires.

program control—1. A control system which automatically holds or changes its target value on the basis of time, to follow a prescribed program for the process. 2. A system in which a computer is used to direct an operation or process and automatically to hold or to make changes in the operation or process on the basis of a prescribed sequence of events.

program counter—*See* Program Register.

program-distribution amplifier—A group of amplifiers fed by a bridging bus from a single source. Each amplifier then feeds a separate line or other service.

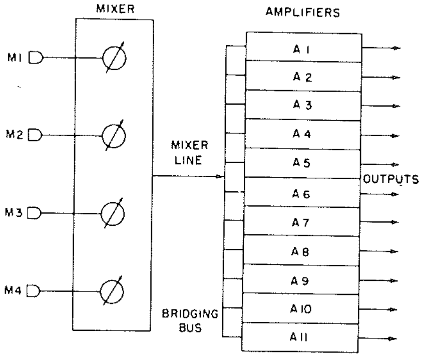

Program distribution amplifier.

program failure alarm—In broadcasting stations, a relay circuit that gives a visual and aural alarm when a program fails. A delay prevents the relay from giving a false alarm during the silence before and after station-identification or other short breaks.

program level—The measure of the program signal in an audio system. It is expressed in volume units (VU).

programmed check—A means of testing for the correctness of a computer program and machine functioning, either by running a similarly programmed sample problem with a known answer (including mathematical or logical checks) or by building a checking system into the actual program being run.

programmer—A person who prepares the sequences of instructions for a computer or other data-handling system. He may or may not convert them into detailed codes.

programming—1. Definition of a computer problem resulting in a flow diagram. 2. Preparing a list of instructions for the computer to use in the solution of a problem. 3. Selecting various circuit patterns by interconnecting or "jumping" the appropriate contacts on one side of a connector plug. 4. The control of a power-supply parameter, such as output voltage, by means of a remotely located or internally located control element (usually resistance) or signal (voltage).

programming device—A device by which a series of mechanical or electrical operations or events may be present to be performed automatically in a predetermined sequence and at specified time intervals.

program parameter—In a subroutine of a computing or other data-handling system, an adjustable parameter which can be given different values on several occasions when the subroutine is used.

program register—Also called program counter, or control register. The computer control-unit register into which is stored the program instruction being executed, hence controlling the computer operation during the cycles required to execute that instruction.

program-sensitive error—In a computer, an error arising from unforeseen behavior of some circuits, discovered when a comparatively unusual combination of program steps occurs.

program signal—In audio systems and components, the complex electric wave—corresponding to speech, music, and associated sounds—destined for audible reproduction.

program step—An increment, usually one instruction, of a computer program.

program tape—In a computer, a magnetic or punched paper tape which contains the sequence of instructions for solving a problem.

program timer—1. A loosely used term sometimes referring to a complex multicircuit timing device in which the program is readily changed, such as tape-controlled timers, can timers, time switches, or to any other type of timer. 2. A multiple-circuit repeat-cycle timer which repeats a preset program continuously as long as power is applied.

progressive scanning—A rectilinear process in which adjacent lines are scanned in succession. In television, the scanning process in which the distance from center to center of successively scanned lines is equal to the nominal line width.

projection cathode-ray tube—A cathode-ray tube that produces an intense but relatively small image, which can be projected onto a large viewing screen by an optical system consisting of lenses or a combination of lenses and mirrors.

projection ppi—A unit in which the image of a 4-inch dark-trace cathode-ray tube is projected on a 24-inch horizontal plotting surface. The echoes appear as magenta-colored arcs on a white background. *Also see* Dark-Trace Tube.

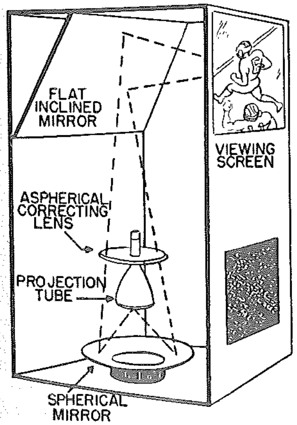

Projection television.

projection television—A combination of lenses and mirrors for projecting an enlarged television picture onto a screen.

projector—1. A device used in an underwater sound system to radiate sound pulses through the water from the bottom of a ship. 2. A horn designed to direct sound chiefly in one direction from a speaker.

promethium cell—A low-power cell containing a radioactive isotope called promethium 147, which emits beta particles that strike a phosphor. Two photocells then convert the light output from the phosphor into electrical energy.

prong—*See* Pin.

proof pressure — The maximum pressure which may be applied to the sensing element of a transducer without changing the transducer performance beyond specified tolerances.

propagation—1. In electrical practice, the travel of waves through or along a medium. 2. The traveling of a wave along a transmission path. 3. Also called wave propagation. The travel of electromagnetic waves or sound waves through a medium. Propagation does not refer to the flow of current in the ordinary sense.

propagation anomaly—An irregularity introduced into an electromagnetic or other sensing device by discontinuities in the propagation medium.

propagation constant—The transmission characteristic which indicates the effect of a line on a wave being transmitted along the line. It is a complex quantity having a real term called the attenuation constant and an imaginary term called the phase constant.

propagation delay—1. A measure of the time required for a change in logic level to propagate through a chain of circuit elements. 2. A measure of the time required for a logic signal to travel through a device or a series of devices forming a logic string. It occurs as the result of four types of circuit delays—storage, rise, fall, and turn-on delay—and is the time between when the input signal crosses the threshold-voltage point and when the responding voltage at the output crosses the same voltage point.

propagation error—For ranging systems, the algebraic sum of propagation-velocity error and curved-path error. Except at long ranges and low angles, the curved-path component of propagation error is generally negligible.

propagation factor—*See* Propagation Ratio.

propagation loss—The loss of energy suffered by a signal while passing between two points.

propagation ratio—Also called the propagation factor. For a wave that has been propagated from one point to another, the ratio of the complex electric-field strength at the second point to that at the first point.

propagation time—The time it takes for a signal to travel from point to point. In a communications channel, the velocity of signal propagation is less than that of radio. A signal delay of 20 msec per thousand miles is a reasonable maximum.

propagation time delay—The time required for a wave to travel from one point to another along a transmission line. It varies according to the type of line.

propagation velocity—*See* Velocity of Propagation.

property sort — In a computer, a technique for the selection from a file of records meeting a certain criterion.

proportional band—The range of the controlled variable corresponding to the full range of operation of the final control element.

proportional control—1. A method of control in which the intensity of action varies linearly as the condition being regulated deviates from the conditions prescribed. 2. Also called galloping ghost. An advanced type of radio-control system in which the rudder (and sometimes the elevator) can move as much (or as little) as the operator wishes.

proportional counter tube—A sealed tube containing an inert gas such as argon, krypton, xenon, methyl bromide, etc. It is used like a Geiger-Muller counter and operated at about 100 volts in the proportional region.

proportional linearity—A manner of expressing nonlinearity as the deviation from a straight line in terms of a given percentage of the transducer output at the stimulus point under consideration (i.e., as a percentage of the reading).

proportional region—In a radiation counter tube, the applied-voltage range in which the gas amplification is greater than unity and is independent of the charge liberated by the initial ionizing event.

protective cable—Small-gage quadded cable used in toll cables to serve as fuses, usually at building entrances.

protective device—Any device for keeping an undesirably large current, voltage, or power out of a given part of an electric circuit.

protective gap—A spark gap provided between a conductor and the earth by suitable electrodes. High-voltage surges due to lightning are thus permitted to pass harmlessly to earth through the gap.

protective relay—A relay, the principal function of which is to protect services from interruption or to prevent or limit damage to apparatus.

protective resistance—A resistance placed in series with a device (e.g., a gas tube) to limit the current to a safe value.

protector—1. A device to protect equipment or personnel from high voltage or current. 2. A device to protect a telephone circuit from the effects of lightning or other high voltages accidentally impressed on the line.

protector block—A rectangular piece of carbon, with an insulated metal insert, or porcelain with a carbon insert which, in combination with each other, makes one element of a protector. It forms a gap which will break down and provide a path to ground for voltages over 350 volts.

protector tube—A glow-discharge, cold-cathode tube in which a low-voltage breakdown is employed between two or more electrodes to protect the circuit against overvoltage.

proton—An elementary particle with a positive charge equivalent to the negative charge of the electron, but with approximately 1,845 times the mass. The proton is the positive nucleus of the hydrogen atom.

prototype model—A working model, usually hand-assembled, and suitable for complete evaluation of mechanical and electrical form, design, and performance. Approved parts are employed throughout, so that it will be completely representative of the final, mass-produced equipment.

proximity detector—A sensing device which gives an indication when approaching or being approached by another object (e.g., a burglar alarm).

proximity effect—The redistribution of current brought about in a conductor by the presence of another current-carrying conductor.

proximity fuse—A fuse designed to detonate a projectile, bomb, mine, or charge when activated by an external influence in the vicinity of a target. The variable time fuse is one type of proximity fuse.

prr—Abbreviation for pulse-repetition rate.

psec—Abbreviation for picosecond (10^{-12} second).

pseudocode—An instruction that is not meant to be followed directly by a computer. Instead, it initiates the linking of a subroutine into the main program.

pseudoinstruction — *See* Instructional Constant.

pseudoprogram—A program that is written in a pseudocode and may include short coded logical routines.

pseudorandom—Having the property of being produced by a definite calculation process while simultaneously satisfying one or more of the standard tests for statistical-randomness.

pseudostereo—Devices and techniques for obtaining stereo qualities from one channel.

psophometric emf—The electromotive force (or voltage) generated by a source having an internal resistance of 600 ohms and no internal reactance, which, when connected across a standard receiver having 600 ohms of resistance and no reactance, produces the same sinusoidal current as an 800-hertz generator of the same impedance. *Also see* Psophometric Voltage.

psophometric voltage—The voltage which would appear across a 600-ohm resistance connected between any two points in a telephone circuit. (This value is one-half the psophometric emf since the latter is essentially the open-circuit potential necessary from a source to produce the psophometric voltage if the source has a 600-ohm internal resistance.)

ptm—Abbreviation for pulse-time modulation.

ptm-ptm-am—A system in which a number of pulse positions or pulse-time modulated subcarriers are used to amplitude-modulate the carrier.

P-type conductivity—The conductivity associated with the holes in a semiconductor.

P-type conductor—A positive-type conductor, one with electron holes as the principal carriers. This implies the presence of acceptors.

P-type crystal rectifier—A crystal rectifier in which forward current flows whenever the semiconductor is more positive than the metal.

P-type semiconductor—An extrinsic semiconductor in which the hole density exceeds the conduction-electron density. By implication, the net ionized impurity concentration is an acceptor type.

public-address system—Also called a p-a system. One or more microphones, an audio-frequency system, and one or more speakers used for picking up and amplifying sounds to a large audience, either indoors or out.

public radiocommunication services—Land mobile and fixed radio services, the stations of which are open to public correspondence.

public-safety radio service—Any radiocommunication service essential either to the discharge of nonfederal governmental functions relating to public-safety responsibilities or to the alleviation of an emergency endangering life or property. The radio transmitting facilities in this service may be fixed, land, or mobile stations.

puck—*See* Pressure Roller.

puff—British abbreviation for picofarad.

pull curves—The characteristics relating force to displacement in the actuating system of a relay.

pull-down resistor—A resistor connected across the output of a device or circuit to hold the output equal to or less than the 0 input level of the following digital device. Also used to lower the output impedance of a device.

pull-in current—*See* Pickup Current.

pull-in current (or voltage)—The maximum current (or voltage) required to operate a relay.

pulling—1. In an oscillator, the undesired change from the desired frequency. It is caused either by coupling from another source of frequency, or by the influence of the load impedance. 2. In television, partial loss of synchronization.

pulling figure—The difference between the maximum and minimum frequencies of an oscillator whenever the phase angle of the load-impedance reflection coefficient varies through 360°. The absolute value of this coefficient is constant and equal to .20.

pull-up torque—The minimum torque developed by an alternating-current motor during the period of acceleration from rest to the speed at which breakdown torque occurs. For motors which do not have a definite breakdown torque, the pull-up torque is the minimum torque developed up to rated speed.

pulsating current—A nonuniform electron flow, which varies periodically but does not reverse its direction.

pulsating direct current—A direct current that changes its value at regular or irregular intervals but flows in the same direction at all times.

pulsating quantity—A periodic quantity that can be considered the sum of a continuous component and an alternating component in the quantity.

pulsation welding — A form of resistance welding in which the power is alternately applied and removed.

pulse—1. The variation of a quantity having a normally constant value. This variation is characterized by a rise and a decay of a finite duration. 2. An abrupt change in voltage, either positive or negative, which conveys information to a circuit. (*Also see* Impulse.)

pulse amplifier—A wideband amplifier used to amplify square waves without appreciably changing their shape.

pulse amplitude—A general term for the magnitude of a pulse. For more specific designation, adjectives such as average, instantaneous, peak, rms (effective), etc., should also be used.

pulse-amplitude modulation — Modulation in which the modulating wave is caused to amplitude-modulate a pulse carrier. Abbreviated pam.

pulse analyzer — Equipment for analyzing pulses to determine their time, amplitude, duration, shape, etc.

pulse arc welding—A type of welding in which the material to be welded is positioned together, forming one electrode. The other electrode is positioned to form a gap with one of the workpieces. An arc is struck and the current heats the workpieces to the melting point at their interface. *Also see* Arc Percussive Welding.

pulse-average time—The duration of a pulse, measured between two points at 50% of the maximum amplitude on the leading and trailing edges.

pulse bandwidth—The smallest continuous frequency interval outside of which the amplitude of the spectrum does not exceed a prescribed fraction of the amplitude at a specified frequency.

pulse carrier—A carrier consisting of a series of pulses. Usually employed as a subcarrier.

pulse code—1. A pulse or series of pulses which, by means of waveform, pulse width, pulse time, pulse numbers, or pulse sequences, may be used to convey information. 2. Loosely, a code consisting of pulses—e.g., Morse, Baudot, binary.

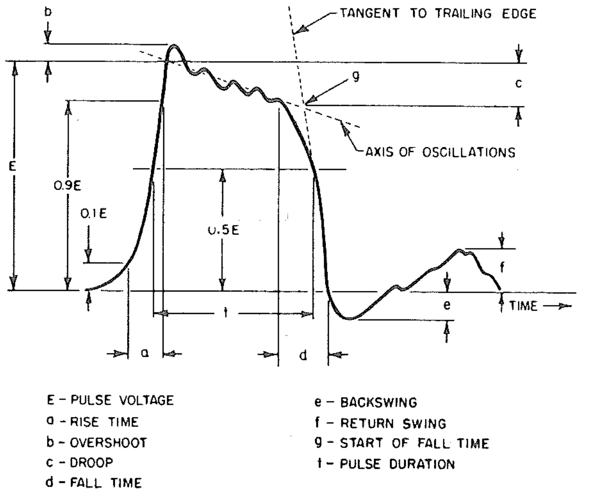

Pulse.

pulse-code modulation — Abbreviated pcm. Pulsed modulation in which the signal is sampled periodically and each sample is quantized and transmitted as a digital binary code.

pulse code modulation-frequency modulation—A system in which pulse code modulated subcarriers are used to frequency-modulate a second carrier. Binary digits are formed by the absence or presence of a pulse in an assigned position.

pulse coder—A circuit which sets up pulses in an identifiable pattern.

pulse coding and correlation—A general technique concerning a variety of methods used to change the transmitted waveform and then decode upon its reception. Pulse compression is a special form of pulse coding and correlation.

pulse compression—A matched filter technique used to discriminate against signals which do not correspond to the transmitted signal.

pulse counter — A device that gives an indication or records of the total number of pulses that it has received during a given time interval.

pulsed Doppler system—A pulsed radar system which utilizes the Doppler effect to obtain information about the target (not including simple resolution from fixed targets).

pulse decay time—The amount of time required for the trailing edge of a pulse to decay from 90% to 10% of the peak pulse amplitude.

pulse delay time—The time interval between the leading edges of the input and output pulses, measured at 10% of their maximum amplitude.

pulse demoder—Also called a constant-delay discriminator. A circuit which responds only to pulse signals with a specified spacing between them.

pulse digit—A code element comprising the immediately associated train of pulses.

pulse-digit spacing—The time interval between the end of one pulse digit and the start of the next.

pulse discriminator—A device that responds only to pulses having a particular characteristic (e.g., duration, amplitude, period). One that responds to period is also called a time discriminator.

pulsed oscillator—1. An oscillator in which oscillations are sustained by either self-generated or external pulses. 2. An oscillator which generates a carrier-frequency pulse or a train of pulses.

pulsed-oscillator starting time—The interval between the leading-edge times of the pulse at the oscillator control terminals and the related output pulse.

pulse droop—Distortion characterized by a slanting of the top of an otherwise essentially flat-topped rectangular pulse.

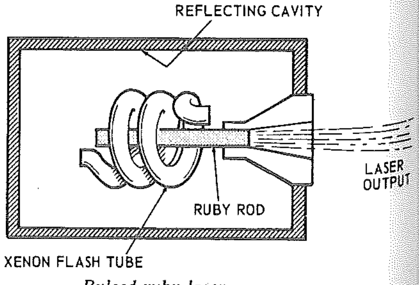

Pulsed ruby laser.

pulsed ruby laser—A laser that uses ruby as the active material. The extremely high pumping power required is obtained by discharging a bank of energy storage capacitors through a special high-intensity flash tube.

pulse duration—Also called pulse length or width. The time interval between the points at which the instantaneous value on the leading and trailing edges bears a specified relationship to the peak pulse amplitude.

pulse-duration discriminator—A circuit in which the sense and magnitude of the output is a function of the deviation of the pulse length from a reference.

pulse-duration modulation — Abbreviated pdm. Also called by the less preferred terms "pulse-width modulation" and "pulse-length modulation." Pulse-time modulation in which the duration of a pulse is varied.

pulse - duration modulation - frequency modulation—Also called pulse-width modulation-frequency modulation. A system in which pulse-duration modulation subcarriers are used to frequency-modulate a second carrier.

pulse duty factor—Ratio of the average pulse duration to the average pulse spacing. Equivalent to the average pulse duration times the pulse-repetition rate.

pulse emission—Emission drawn for short periods; it may or may not follow a regular repetition rate.

pulse emitter load — The load seen by the collector of an inverter that drives the pulse input to a flip-flop, pulse amplifier, or delay.

pulse equalizer—A circuit that produces output pulses of uniform size and shape when driven by input pulses that vary in size and shape.

pulse fall time—That time during which the trailing edge of a pulse is decreasing from 90% to 10% of its maximum amplitude.

pulse-forming line—A combination of circuit components used to produce a square pulse of controlled duration.

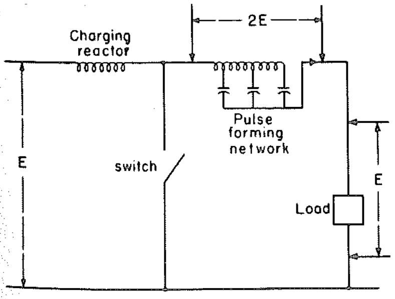

Pulse-forming network.

pulse-forming network — A network which converts either an a-c or d-c charging source into an approximately rectangular wave output. By means of a high-speed switch, it alternately stores energy from the charging source and releases the energy through the load to which it is connected. This network supplies the accurately shaped pulse required by the magnetron or klystron oscillator of a radar modulator.

pulse-frequency modulation — Abbreviated pfm. More precisely called pulse-repetition-rate modulation. Pulse-time modulation in which the pulse repetition rate is varied.

pulse-frequency spectrum—*See* Pulse Spectrum.

pulse generator—A device for generating a controlled series of electrical pulses.

pulse group—*See* Pulse Train.

pulse-height analyzer — Also called kicksorter. An instrument that indicates the number or rate of occurrence of pulses within each of one or more specified amplitude ranges.

pulse-height discriminator — Also called a pulse-height selector. A circuit that selects and passes only those pulses which exceed a certain minimum amplitude.

pulse-height selector—*See* Pulse-Height Discriminator.

pulse-improvement threshold — In a constant-amplitude pulse-modulation system the condition existing when the peak pulse voltage is at least twice the peak noise voltage after selection and before any nonlinear process such as amplitude clipping and limiting. The ratio of peak to rms noise voltage is ordinarily assumed to be 4 to 1.

pulse interference eliminator—A device that removes pulse signals which are not precisely on the radar operating frequency.

pulse interleaving—A process in which pulses from two or more time-division multiplexers are systematically combined in time division for transmission over a common path.

pulse interrogation — The triggering of a transponder by a pulse or pulse mode. Interrogations by the latter may be employed to trigger one or more transponders.

pulse interval—*See* Pulse Spacing.

pulse-interval modulation—Pulse-time modulation in which the pulse spacing is varied. *Also see* Pulse-Spacing Modulation.

pulse jitter—A relatively slight variation of the pulse spacing in a pulse train. It may be random or systematic, depending on its origin, and is generally not coherent with any imposed pulse modulation.

pulse length—*See* Pulse Duration.

pulse-length modulation—*See* Pulse-Duration Modulation.

pulse load — The load presented to a pulse source.

pulse mode—1. A finite sequence of pulses, in a prearranged pattern, used for selecting and isolating a communication channel. 2. The prearranged pattern in (1) above.

pulse-mode multiplex—A process or device for selecting channels by means of pulse modes. In this way, two or more channels can use the same carrier frequency.

pulse moder—A device for producing a pulse mode. (*Also see* Pulse Demoder.)

pulse-modulated radar—Radar in which the radiation consists of a series of discrete pulses.

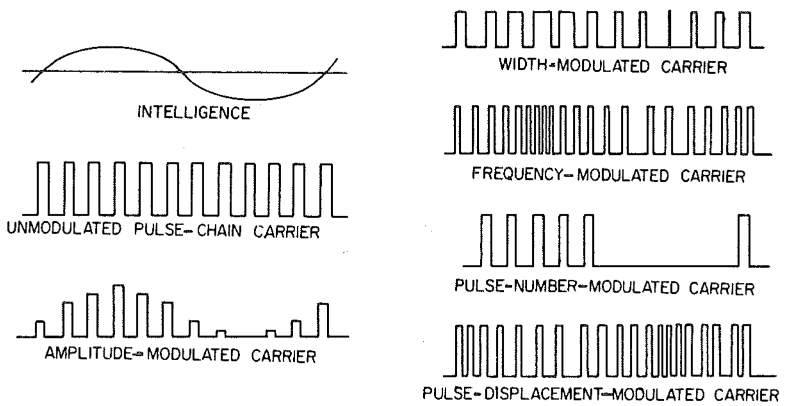
INTELLIGENCE
UNMODULATED PULSE-CHAIN CARRIER
AMPLITUDE-MODULATED CARRIER
WIDTH-MODULATED CARRIER
FREQUENCY-MODULATED CARRIER
PULSE-NUMBER-MODULATED CARRIER
PULSE-DISPLACEMENT-MODULATED CARRIER

Pulse modulation (definition 2).

pulse-modulated waves—Recurrent wave trains used extensively in radar. In general, their duration is shorter than the interval between them.

pulse modulation—1. Modulation of a carrier by pulses. 2. Modulation of a pulse carrier.

pulse modulator—A device which applies pulses to the element being modulated.

pulse operation—The method whereby the energy is delivered in pulses. Usually described in terms of the shape and the frequency of the pulses.

pulse oscillator—An oscillator in which the oscillations are sustained by self-generated or external pulses.

pulse packet—In radar, the volume of space occupied by the pulse energy.

pulse-position modulation—Abbreviated ppm. Pulse-time modulation in which the value of each instantaneous sample of the wave modulates the position in time of a pulse.

pulse-position modulator—A device which converts analog information to variations in pulse position.

pulser—A generator which produces extremely short, high-voltage pulses at definite recurrence rates for use in radar transmitters and similar pulsed systems.

pulse rate—*See* Pulse-Repetition Rate.

pulse ratio—Ratio of the length of any pulse to its total period.

pulse recovery—The time, usually in microseconds, required for electron flow in a diode to start or stop when voltage is suddenly applied or removed.

pulse-recurrence, counting-type frequency meter—A device for measuring frequency. It uses a direct-current ammeter calibrated in pulses per second.

pulse-recurrence time—The time elapsing between the start of one transmitted pulse and the next pulse. The reciprocal of the pulse-recurrence frequency.

pulse regeneration—The restoring of a series of pulses to their original timing, form, and relative magnitude.

pulse repeater—Also called a transponder. A device that receives pulses from one circuit and transmits corresponding pulses at another frequency, waveshape, etc., into another circuit.

pulse-repetition frequency—The rate (usually given in hertz or pulses per second) at which pulses or pulse groups are transmitted from a radar set. Abbreviated prf.

pulse-repetition period—The reciprocal of the pulse-repetition frequency.

pulse-repetition rate—Abbreviated prr. Also called pulse rate. The average number of pulses per unit of time.

pulse-repetition-rate modulation—*See* Pulse-Frequency Modulation.

pulse reply—The transmission of a pulse or pulse mode by a transponder as the result of an interrogation.

pulse resolution—The minimum time separation, usually in microseconds or milliseconds, between input pulses that permits proper circuit or component response.

pulse rise time—The interval of time required for the leading edge of a pulse to rise from 10% to 90% of its peak amplitude, unless some other percentage is stated.

pulse sample-and-hold circuit—A circuit which holds the final amplitude of an integrated pulse until the final amplitude of the succeeding integrated pulse is reached. A less desirable sample-and-hold circuit resets after each hold period to a fixed level before integrating a succeeding pulse.

pulse scaler—A device capable of producing an output signal whenever a prescribed number of input pulses has been received. It frequently includes indicating devices that facilitate interpolation.

pulse selector—A circuit or device that selects the proper pulse from a sequence of (telemetering) pulses.

pulse separation—The interval between the trailing-edge pulse time of a pulse and the leading-edge pulse time of the succeeding pulse.

pulse shaper—Any transducer (including pulse regenerators) used for changing one or more characteristics of a pulse.

pulse shaping—Intentionally changing the shape of a pulse.

pulse spacing—The time interval from one pulse to the next—i.e., between the corresponding times of two consecutive pulses. (The term "pulse interval" is ambiguous—it may be taken to mean the duration of a pulse instead of the space or interval between pulses.)

pulse-spacing modulation—Formerly called pulse-interval modulation. A form of pulse-time modulation in which the pulse spacing is varied.

pulse spectrum—Also called pulse-frequency spectrum. The frequency distribution, in relative amplitude and phase, of the sinusoidal components of a pulse.

pulse spike—A relatively short duration pulse superimposed on the main pulse.

pulse-spike amplitude—The peak amplitude of a pulse spike.

pulse-storage time—The time interval from a point at 90% of the maximum amplitude on the trailing edge of the input pulse to the same 90% point on the trailing edge of the output pulse.

pulse stretcher—1. A circuit designed to extend the duration of a pulse—primarily so that its pulse modulation will be more readily discernible in an audio presentation. 2. In a computer, a circuit which generates a long pulse when triggered by a short pulse. The width of the output pulse is determined by the value of the coupling capacitor. The maximum width of the output pulse cannot exceed 50% of the clock rate.

pulse tilt—A distortion characterized in an otherwise essentially flat-topped rectangular pulse by either a decline or a rise of the pulse top.

pulse time—The time interval from a point at 90% of the maximum amplitude on the leading edge of a pulse, to the 90% point on the trailing edge.

pulse-time-modulated radiosonde—Also called time-interval radiosonde. A radiosonde which transmits the indications of the meterological sensing elements in the form of pulses spaced in time. The meteorological data are evaluated from the intervals between the pulses.

pulse-time modulation—Abbreviated ptm. Modulation (e.g., pulse-duration and pulse-position) in which the values of instantaneous samples of the modulating wave are made to modulate the occurrence time of some characteristic of a pulse carrier.

pulse timer—A timer the time cycle of which is started with a continuous input voltage and application or removal of an additional positive input voltage pulse. At the end of the time-delay period, the contacts transfer.

pulse train—Also called pulse group or impulse train. A group or sequence of pulses of similar characteristics.

pulse-train frequency spectrum—*See* Pulse-train Spectrum.

pulse-train spectrum—Also called pulse-train frequency spectrum. The frequency distribution, in amplitude and in phase angle, of the sinusoidal components of the pulse train.

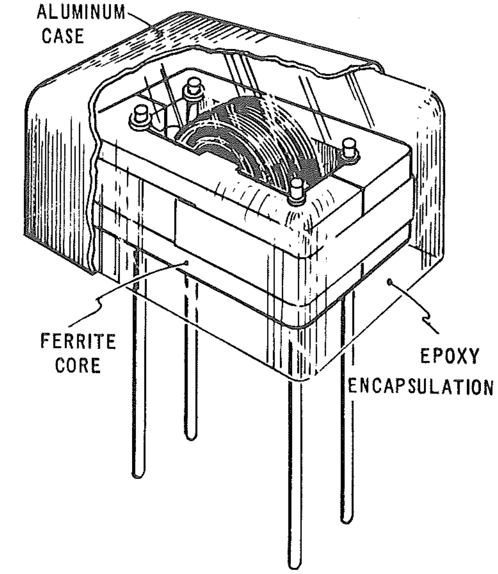

Pulse transformer.

pulse transformer—A transformer designed to pass pulse currents and voltages with the least possible change in waveshape, or to be used for generating pulses in a blocking oscillator.

pulse transmitter—1. A pulse-modulated transmitter in which the peak power-output capabilities are usually larger than the average power-output rating. 2. A transmitter used to generate and transmit pulses over a telemetering or pilot-wire circuit to the remote indicating or receiving device.

pulse valley—In a pulse, the portion between two specified maxima.

pulse width—*See* Pulse Duration.

pulse-width discriminator—A device that measures the pulse length of video signals and passes only those the time duration of which falls into some predetermined design tolerance.

pulse-width modulation—*See* Pulse-Duration Modulation.

pulse-width modulation-frequency modulation—*See* Pulse-Duration Modulation-Frequency Modulation.

pulsing transformer—A transformer designed to supply pulses of voltage or current.

pump—1. An external source used to increase the electron population of excited energy states. 2. Of a parametric device, the source of alternating-current power which causes the nonlinear reactor to behave as a time-varying reactance.

pumped tube—An electron tube (chiefly a pool-cathode) which is continuously connected to evacuating equipment during operation.

pumping—Of a laser, the application of radiation of appropriate frequency to invert the distribution of systems of electrons of the laser media so that the levels of higher energy states are more populated.

pumping frequency — The frequency at which pumping is provided in a maser, quadrupole amplifier, or other amplifier requiring high-frequency excitation.

pump oscillator—An alternating-current generator that supplies pumping energy for maser and parametric amplifiers. Operates at twice or some higher multiple of the signal frequency.

punched card—A heavy, stiff paper of constant size and shape, suitable for punching in a pattern that has meaning, and for being handled mechanically. The punched holes are sensed electrically by wire brushes, mechanically by metal fingers, or photoelectrically by photocells.

punch-card machine—*See* Key Punch.

punched tape—Also called tape or perforated tape. Paper tape punched in a coded pattern of holes, which convey information.

punch-through voltage—*See* Reach-Through Voltage.

puncture—A disruptive discharge of current through insulation, which breaks down under electrostatic stress and permits the flow of a sudden, large current through the opening. *Also see* Breakdown.

puncture voltage—1. The voltage at which an insulator or shell is electrically punctured when subjected to a gradually increasing voltage. 2. The voltage at which insulation fails by disruptive discharge through the insulation sample. It is assumed that the sample area is large enough to prevent flashover.

Pupin coil—An iron-core loading coil inserted into telephone lines at regular intervals to balance out the effect of capacitance between the lines.

pup jack—*See* Tip Jack.

pure tone—*See* Simple Tone.

purity—Physically complete saturation of a hue—i.e., uncontaminated by white and other colors. (*Also see* Excitation Purity.)

purity coil—A coil consisting of two current-carrying windings. In a color television receiver, they produce a magnetic field which directs the three electron beams so that each one will strike only the proper set of phosphor dots.

purity control—A variable resistor that controls the current through the purity coil mounted around the neck of a color picture tube.

purity magnet — Two adjustable magnetic rings used in place of a purity coil.

purple boundary—The straight line drawn between the ends of the spectrum locus on a chromaticity diagram.

pushback hookup wire—Tinned copper wire covered with loosely braided insulation, which can be pushed back with the fingers to expose enough bare wire for making a connection.

push-button control—Control of equipment by means of push-buttons, which in turn operate relays, etc.

push-button switch—A switch in which a button must be depressed each time the contacts are to be opened or closed.

push-button tuner—A series of interlocked push-button switches which connect into a circuit the correct tuning elements for the frequency corresponding to the depressed button.

pushdown list—A list of items in which the last item entered becomes the first item of the list and the relative position of each of the other items is pushed back one.

pushing figure—The change in oscillator frequency due to a specified change in plate current (excluding thermal effects).

push-pull amplifier—*See* Balanced Amplifier.

push-pull circuit—A circuit containing two like elements which operate in 180° phase relationship to produce additive output components of the desired wave and cancellation of certain unwanted products. Push-pull amplifiers and oscillators use such a circuit.

push-pull currents—Balanced currents.

push-pull doubler—An amplifier used for frequency doubling. It consists of two transistors or vacuum tubes; the latter have their grids (input) connected in push-pull and their plates (output) in push-push or parallel.

push-pull microphone—A microphone comprising two like elements actuated by the same sound waves and operated 180° out of phase.

push-pull oscillator—A balanced oscillator employing two similar tubes or transistors in phase opposition.

push-pull transformer—An audio-frequency transformer that has a center-tapped winding and is used in a push-pull amplifier circuit.

push-pull voltages—Balanced voltages.

push-push circuit—A circuit usually used as a frequency multiplier to emphasize even-order harmonics. Two similar transistors are employed, or two tubes with their grids connected in phase opposition and their plates in parallel to a common load.

push-push currents — Currents which are equal in magnitude and which flow in the same direction at every point in the two conductors of a balanced line.

push-push voltages — Voltages which are equal in magnitude and have the same polarity (relative to ground) at every point on the two conductors of a balanced line.

push rod—A shaft that connects a servo or other actuator with the part of the controlled device.

push-to-talk switch — *See* Press-to-Talk Switch.

pushup list — A list of items in which each item is entered at the end of the list and the previous items maintain their same relative position in the list.

pw—Abbreviation for picowatt.

pwm-fm—A system in which a number of pulse-width-modulated subcarriers are used to frequency-modulate the carrier.

pylon antenna — A vertical antenna constructed of one or more sheet-metal cylinders with a lengthwise slot. The gain depends on the number of sections.

pyramidal horn—An electromagnetic horn, the sides of which form a pyramid. The electromagnetic field in such a horn would be expressed basically in a family of spherical coordinates.

pyramid wave—A triangular wave, the sides of which are approximately equal in length.

pyrheliometer — A device for the measurement of infrared radiation.

pyroelectric effect—Also called pyroelectricity. The redistribution of the charge in a crystal that has been heated. The crystal is left with a net electric dipole moment—i.e., the centers of the positive and negative charges are separated.

pyroelectricity—*See* Pyroelectric Effect.

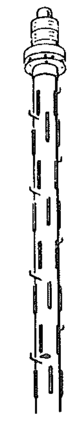

Pylon antenna.

pyroelectric material—A material which produces an electrical output when subjected to a change in temperature.

pyromagnetic—Pertaining to the effect of heat and magnetism on each other.

pyrometer—An instrument used to measure elevated temperatures (beyond the range of mercury thermometers) by electric means. These include immersion, optical, radiation, resistance, and thermoelectric pyrometers.

pyron detector—A crystal detector in which rectification occurs between iron pyrites and copper (or other metallic points).

Pythagorean scale—A musical scale in which the frequency intervals are represented by the ratios of integral powers of 2 and 3.

Q

Q—1. Symbol for quantity of electric charge. 2. A measure of the relationship between stored energy and rate of dissipation in certain electric elements, structures or materials. 3. In an inductor, the ratio of its reactance to its effective series resistance at a given frequency. 4. Also called quality factor or Q factor. In a capacitor, the ratio of its susceptance to its effective shunt conductance at a given frequency. 5. A measure of the sharpness of resonance or frequency selectivity of a mechanical or electrical system.

Q antenna—A dipole matched to its transmission line by stub matching.

Q band—A band of frequencies extending from 36 to 46 GHz, corresponding to wavelengths of 0.834 to 0.652 cm.

Q channel—The 0.5-MHz-wide band used in the American NTSC color-television system to transmit green-magenta color information.

QCW—A 3.58-MHz continuous-wave signal having Q phase. The term is generally limited to reference to the color-television receiver local oscillator and associated circuits.

QCW signal—*See* Quadrature-Phase Subcarrier Signal.

Q demodulator—A demodulator circuit the inputs of which are the chrominance signal and the signal from the local 3.58-MHz oscillator after it has been shifted 90°. This phase shift is necessary so that the local signal will be an accurate representation of the Q subcarrier that was suppressed at the transmitter. The output of the Q demodulator is a color-video signal representing colors in the televised scene.

Q-factor—*See* Q, 4.

Q-meter—Also called a quality-factor meter. An instrument for measuring the Q, or quality factor, of a circuit or circuit element.

Q-multiplier—A special filter which has a sharply-peaked response curve or a deep rejection notch at a particular frequency.

Q-phase—Also called quadrature carrier. A color-television signal carrier having a phase difference of 147° from the color subcarrier.

qrm—An obsolete term for any type of man-made interference.

qrs complex—That portion of the waveform in an electrocardiogram extending from point Q to point S; it includes the maximum amplitude shown in an ecg trace.

Q-signal—In color television, the signal formed by the combination of R—Y and B—Y color-difference signals having positive polarities of 0.48 and 0.41, respectively. It is one of the two signals used to modulate the chrominance subcarrier, the other being the I signal. (*Also see* Coarse-Chrominance Primary.)

Q-signals—Abbreviations used in radiocommunications.

QSL card—A card exchanged by radio amateurs to confirm radiocommunications with each other.

quad—1. A structural unit employed in cables. A quad consists of four separately insulated conductors twisted together. These conductors may take the form of two twisted pairs. 2. A combination of four elements, either electronic components or complete circuits, in a series-parallel or parallel-series arrangement. 3. A (series-parallel) combination of four transistors.

quadded cable—A cable in which some or all of the conductors are in the form of quads.

quadding—Connecting transistors in a series-parallel configuration, to achieve greater reliability.

quadrant—1. A sector, arc, or angle of 90°. 2. An instrument for measuring or setting vertical angles.

quadrantal error—The error caused in magnetic-compass readings by the magnetic field of the steel hull of a ship, or by metal structures near the loop antenna of radio direction finders aboard a vessel or aircraft.

quadrant electrometer—An electrometer for measuring voltages and charges by means of electrostatic forces. A metal plate or needle is suspended horizontally inside a vertical metal cylinder that is divided into four insulated parts, each connected electrically to the one opposite it. The two pairs of quadrants are connected to the two terminals between which the potential difference is to be measured. The resultant electrostatic forces displace the suspended indicator a certain amount, depending on the voltage.

quadrature—The state or condition of two related periodic functions or two related points separated by a quarter of a cycle, or 90 electrical degrees.

quadrature amplifier—A stage used to supply two signals of the same frequency but with phase angles that differ by 90 electrical degrees.

quadrature carrier—*See* Q-phase.

quadrature component—1. The reactive current or voltage component due to inductive or capacitive reactance in a circuit. 2. A vector representing an alternating quantity which is in quadrature (at 90°) with some reference vector.

quadrature modulation—The modulation of two carrier components 90° apart in phase by separate modulating functions.

quadrature-phase subcarrier signal—That portion of the chrominance signal which leads or lags the in-phase portion by 90°. Abbreviated qcw signal.

quadrature portion—In the chrominance signal, the portion with the same or opposite phase from that of the subcarrier modulated by the Q-signal. This portion of the chrominance signal may lead or lag the in-phase portion by 90 electrical degrees.

quadrature sensitivity—Also called side sensitivity, lateral sensitivity, or cross-talk sensitivity. The sensitivity of a transducer to motion normal to the principal axis. Commonly expressed in percent of the sensitivity in the principal axis.

quadripole network—*See* Two-Terminal–Pair Network.

quadrupole—A combination of two dipoles that produces a force varying in inverse proportion to the fourth power of the distance from the generating charge.

quadrupole network—*See* Two-Terminal–Pair Network.

quality—The extent of conformance to specifications of a device or the proportion of satisfactory devices in a lot.

quality assurance—A planned, systematic pattern of actions necessary to provide suitable confidence that an item will perform satisfactorily in actual operation.

quality control—The control of variation of workmanship, processes, and materials in order to produce a consistent, uniform product.

quality engineering—An engineering program the purposes of which are to establish suitable quality tests and quality acceptance criteria and to interpret quality data.

quality factor—*See* Q, 4.

quality-factor meter—*See* Q-meter.

quantity—Any positive or negative number. It may be a whole number, a fraction, or a whole number and a fraction.

quantity of electricity — *See* Electrical Charge.

quantization—The process whereby the range of values of a wave is divided into a finite number of subranges, each represented by an assigned (quantized) value.

quantization distortion—Also called quantization noise. Inherent distortion introduced during quantization.

quantization error—The difference between the actual values of data and corresponding discrete values resulting from quantization.

quantization level—1. A particular subrange in quantization. 2. The symbol designating the subrange of (1) above.

quantization noise—*See* Quantization Distortion.

quantize—To convert a continuous variable, such as a waveform, into a series of levels or steps. There are no "in-between" values in such a quantized waveform. All values of signal are represented by the nearest standard value or code position.

quantized pulse modulation—Pulse modulation which involves quantization (e.g., pulse numbers or pulse code modulation).

quantizer—A device which decides in what particular digital subdivision an analog quantity will be placed.

quantizing—Expressing an analog value as the nearest one of a discrete set of prechosen values.

quantum—1. A discrete portion of energy of a definite amount. It was first associated with intra-atomic or intramolecular processes involving changes among the electrons, and the corresponding radiation. 2. If the magnitude of a quantity is always an integral multiple of a definite unit, then that unit is called the quantum of the quantity. 3. The angular increment of input-shaft rotation of an encoder, subtended by one code position.

quantum efficiency (of a phototube) — The average number of electrons photoelectrically emitted from the photocathode per incident photon of a given wavelength.

quantum mechanics—The study of atomic structure and other related problems in terms of quantities that can actually be measured.

quantum noise—A random variation or noise signal due to fluctuations in the average rate of incidence of quanta on a detector. The basic electromagnetic quantum of noise power is just 1 photon per electromagnetic mode.

quantum theory—The theory that an atom or molecule does not emit or absorb energy continuously. Rather, it does so in a series of steps, each step being the emission or absorption of an amount of energy called the quantum. The energy in each quantum is directly proportionate to the frequency.

quarter phase—*See* Two-Phase.

quarter-square multiplier—A circuit that achieves true four-quadrant multiplication by taking advantage of the mathematical relationship that the product of two variables is equal to one-quarter of the difference of the squares of the sum and the difference of the variables.

quarter-wave antenna—An antenna, the electrical length of which is one-quarter the wavelength of the transmitted or received signal.

quarter-wave attenuator—Two energy-absorbing grids or other structures placed in a transmission line and separated by an odd number of quarter wavelengths. As a result, the wave reflected from the first grid annuls the wave reflected from the second grid.

quarter-wavelength—That distance which corresponds to an electrical length of a quarter of a wavelength of the frequency under consideration.

quarter-wave line—*See* Quarter-Wave Stub.

quarter-wave plate—A mica or other double-refracting crystal plate of such thickness that a phase difference of one-quarter cycle is introduced between the ordinary and extraordinary components of light passing through.

quarter-wave resonance—In a quarter-wave antenna, the condition in which its resonant frequency is equal to the frequency at which it is to be used.

quarter-wave stub—Also called a quarter-wave line or quarter-wave transmission line. A section of transmission line equal to one-quarter of a wavelength at the fundamental frequency. It is commonly used to suppress even harmonics. This is done by shorting the far end so that the open end presents a high impedance to the fundamental frequency and all odd harmonics, but not to the even-order harmonics.

quarter-wave support—A quarter-wave metallic stub, used in place of dielectric insulators between the inner and outer conductors of a coaxial transmission line.

quarter-wave termination—A waveguide termination consisting of a metal plate and a wire grating (or semiconducting film) spaced one-quarter wavelength apart. The plate is the terminating element. The wave reflected by the grating (or film) is cancelled by the wave reflected by the plate.

quarter-wave transformer—A one-quarter-

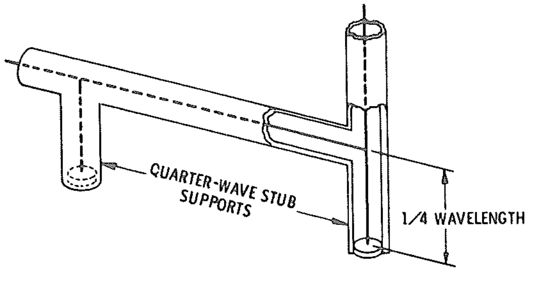

Quarter-wave supports.

wavelength section of transmission line used for impedance matching.

quarter-wave transmission line—*See* Quarter-Wave Stub.

quartz—A mineral occurring in hexagonal crystals in nature and having piezoelectric properties that are highly useful in radio and carrier communication. The crystals from which slabs are cut for oscillators are transparent and almost colorless. When excited electrically, they vibrate and maintain extremely accurate and stable frequencies.

quartz crystal—Also called a crystal. A thin slab cut from quartz and ground to the thickness at which it will vibrate at the desired frequency when supplied with energy. It is used to accurately control the frequency of an oscillator.

quartz delay line—A delay line in which fused quartz is the medium for delaying sound transmission or a train of waves.

quartz lamp—A mercury-vapor lamp having a transparent envelope made from quartz instead of glass. Quartz resists heat (permitting a higher current) and passes ultraviolet rays, which glass will absorb.

quartz plate — A crystalline-quartz section completely finished to specifications, with its two major faces essentially parallel.

quartz resonator—A piezoelectric resonator with a quartz plate.

quasi-linear feedback-control system—A feedback-control system in which the relationships between the pertinent measures of the system input and output signals are substantially linear despite the existence of nonlinear elements.

quasi-optical—Having properties similar to those of light waves. The propagation of waves in the television spectrum is said to be quasi-optical (i.e., cut off by the horizon).

quasi-rectangular wave — A wave nearly, but not, rectangular in shape.

quasi-single-sideband — Simulated single-sideband transmission done by transmitting parts of both sidebands.

quasi-steady-state vibration—A nearly periodic vibration in which the amplitude and phase relationships of the component sinusoids vary slowly with time.

quaternary signaling—The communications mode in which information is passed by the presence and absence, or plus and minus variations, of four discrete levels of one parameter of the signaling medium.

quenched spark—A spark consisting of only a few sharply defined oscillations, because the gap is deionized almost immediately after the initial spark has passed.

quenched spark gap—A spark gap with provision for producing a quenched spark. One form consists of many small gaps between electrodes that have a relatively large mass and thus are good radiators of heat. As a result, they cool the gaps rapidly and thereby stop conduction.

quenched spark-gap converter—A spark-gap generator or other power source in which the oscillatory discharge of a capacitor through an inductor and a spark gap provides the radio-frequency power. The spark gap comprises one or more closely spaced gaps in series.

quench frequency—1. An a-c voltage applied to an electrode of a tube used as a superregenerative detector to alternately vary its sensitivity and thereby prevent sustained oscillations. The quench frequency is usually lower than the signal frequency to be received. 2. The number of times per second a circuit goes in and out of oscillation.

quenching—1. The terminating of a discharge in a radiation-counter tube by inhibiting the re-ignition. 2. A process of rapid cooling from an elevated temperature, in contact with liquids, gases, or solids.

quenching circuit—A circuit which inhibits multiple discharges from an ionizing event by suppressing or reversing the voltage applied to a counter tube.

quenching frequency—That frequency at which oscillations in a superregenerative receiver are suppressed (quenched).

quench oscillator—A superregenerative receiver circuit which produces the quench-frequency signal.

quick-break—A characteristic of a switch or circuit breaker, whereby it has a fast contact-opening speed that is independent of the operator.

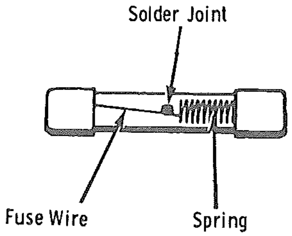

Quick-break fuse.

quick-break fuse—A fuse that draws out the arc and rapidly breaks the circuit when its wire melts. Usually a spring or weight is used to quickly separate the broken ends.

quick-break switch—A switch that minimizes arcing by breaking a circuit rapidly, independent of the rate at which the switch handle is moved.

quick disconnect — A type of connector designed to facilitate rapid locking and unlocking of two contacts or connector halves.

quick-make—A characteristic of a switch or circuit breaker, whereby it has a fast contact-closing speed that is independent of the operator.

quick-make switch—A switch or circuit breaker which has a high contact-closing speed independent of the operator.

quick-stop control—On some tape recorders, and on all recorders used for dictation, a control with which the operator can stop the tape without taking the machine out of the play or record position.

quiescence—The state of a transistor amplifier with no signal applied.

quiescent—At rest—specifically, the condition of a circuit when no input signal is being applied to it.

quiescent-carrier modulation—A modulation system in which the carrier is suppressed during intervals when there is no modulation.

quiescent-carrier telephony—Telephony in which the carrier is suppressed whenever no modulating signals are to be transmitted.

quiescent current—*See* Idling Current.

quiescent period—The resting period—e.g., the period between pulses in pulse transmissions.

quiescent point—On the characteristic curve of an amplifier, the point representing the conditions existing when there is no input signal. (*Also see* Operating Point.)

quiescent push-pull—In a radio receiver, a push-pull output stage in which practically no current flows when no signal is being received. Thus, there is no noise while the radio is being tuned between stations.

quiescent state—The time during which a tube or other circuit element is not performing its active function in the circuit.

quiescent value — The voltage or current value of a vacuum-tube electrode when no signals are present.

quiet automatic volume control—*See* Delayed Automatic Volume Control.

quiet avc—*See* Delayed Automatic Volume Control.

quiet battery—Also called talking battery. A source of energy of special design, or with added filters, which is sufficiently quiet and free from interference that it may be used for speech transmission.

quieting—The decrease in noise voltage at the output of an f-m receiver in the presence of an unmodulated carrier.

quieting sensitivity—In an f-m receiver, the minimum input signal that will give a specified output signal-to-noise ratio.

quiet tuning—In a radio receiver, a form of tuning in which the output is silenced except when the receiver is tuned to the precise frequency of the incoming carrier wave.

R

R—Symbol for resistor, resistance, or reluctance.

r—Abbreviation for roentgen.

RACES—Acronym for Radio Amateur Civil Emergency Service.

raceway—Any channel designed and used solely for holding wires, cables, or bus bars.

rack—1. A bar with teeth, which engage a pinion, worm, etc., to provide straight-line motion. 2. A vertical frame on which equipment, relays, etc., are mounted.

rack and panel connector—A connector that is attached to a panel or side of equipment so that when these members are brought together, the connector is engaged.

rack and pinion—A toothed bar (rack) which engages a gear (pinion) to convert the back-and-forth motion of the rack into rotary motion, or the rotary motion of the pinion into back-and-forth motion.

racon—*See* Radar Beacon.

rad—The unit of absorbed radiation dose equivalent to an energy deposition of 100 ergs/gm—i.e., a measure of the energy which the ionizing radiation imparts to matter per unit mass of irradiated material.

radar—Acronym for RAdio Detecting And Ranging. 1. A system that measures distance (and usually the direction) to an object by determining the amount of time required by electromagnetic energy to travel to and return from an object. 2. Also called primary radar when the signals are returned by reflection. 3. Also called secondary radar when the incident signal triggers a responder beacon and causes it to transmit a second signal.

radar altitude—Also called radio altitude. Absolute altitude measured by a radar altimeter.

radar and television aid to navigation—A device which converts a circular scan radar presentation to a horizontal scanned television presentation. It provides a continuous bright display with target trails for course and speed indications of moving targets.

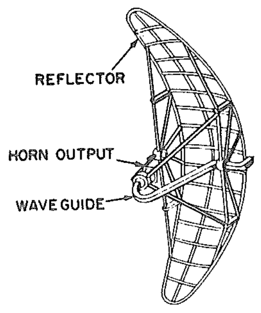

Radar antenna.

radar antenna—Any of the many types of antennas used in radar.

radar attenuation—Ratio of the transmitted power to the reflected (received) power—

specifically, the ratio of the power which the transmitter delivers to the transmission line connected to the transmitting antenna, to the power reflected from the target and delivered to the transmission line connected to the receiving antenna.

radar beacon—Also called a racon. An automatic transmitter-receiver which receives signals from a radar transmitter, and retransmits coded signals that enable the radar operator to determine his position.

radar beam—The space where a target can be effectively detected and/or tracked in front of a radar. Its boundary is defined as the loci of points measured radially from the beam center at which the power has decreased to one-half.

radar calibration—Taking measurements on various parts of electronic equipment (e.g., radar, iff, communications) to determine its performance level.

radar camouflage—The use of coverings or surfaces on an object to considerably reduce the reflected radio energy and thus "conceal" the object from the radar beam.

radar clutter—The image produced on a radar indicator screen by sea or ground return. If not of particular interest, it tends to obscure the target indication.

radar-confusion reflectors—Metallic devices (e.g., chaff, corner reflectors) employed to return false signals in order to confuse enemy radar receivers. The use of radar-confusion reflectors is termed reflective jamming.

radar control area—The designated space within which aircraft approach, holding, stacking, and similar operations are performed under guidance of a surveillance-radar system.

radar countermeasures—Interception, jamming, deception, and evasion of enemy radar signals to obtain information about the enemy from his radar and to prevent him from obtaining accurate, usable information from his radar.

radar-coverage indicator—A device showing how far a radar station should track a given aircraft. It also provides a reference (detection) range for quality control. Aircraft size and altitude, screening angle, site elevation, type of radar, antenna radiation pattern, and antenna tilt are all taken into account.

radar cross section—That portion of the back-scattering cross section of a target associated with a specified polarization component of the scattered wave.

radar deception—Radiation or reradiation of radar signals in order to confuse or mislead an enemy operator when he interprets the data shown on his scope.

radar distribution switchboard—A switching panel for connecting video, trigger, and bearing from any 1 or 5 system, to any or all of 20 repeaters. Also contains order lights, bearing cutouts, alarms, test equipment, etc.

radar echo—1. The radio-frequency energy received after it has been reflected from an object. 2. The deflection or change of intensity which a radar echo produces in the display of a cathode-ray tube.

radar equation—A mathematical expression relating the transmitted and received powers and antenna gains of a primary-radar system to the echo area and distance of the target.

radar fence—A network of radar warning stations which maintain constant watch against surprise attack (e.g., the DEW line).

radar homing—Missile guidance, the intelligence for which is provided by a radar aboard the missile.

radar horizon—The most distant point (from the radar antenna along a given azimuth) on the earth's surface illuminated by the radar on purely geometric conditions. The conditions are that the illumination occurs along a straight-line path, where the path is taken over an effective earth's radius of 4/3 its true radius and where the illuminating power of the radar is considered unlimited.

radar illumination—The subjection of an object (target) to electromagnetic radiation from a radar.

radar indicator—A cathode-ray tube with its associated equipment that provides a visual indication of the echo signals picked up by the radar set.

radar marker—A fixed facility which continuously emits a radar signal so that a bearing indication appears on a radar display.

radar nautical mile—The time interval of approximately 12.367 microseconds required for radio-frequency energy to travel one nautical mile and return, a total of two nautical miles.

radar paint—A radar-energy–absorbent material that can be applied to an object to reduce the possibility of detection.

radar-performance figure—Ratio of the pulse power of the radar transmitter, to the power of the minimum signal detectable by the receiver.

radar picket—Early-warning aircraft which flies at a distance from a ship or other force being protected, to increase the radar detection range.

radar pulse modulator—A modulator that turns an r-f energy source off and on in a precise, known manner. In essence, the modulator supplies the energy which causes an r-f source to oscillate or amplify, thus creating a burst or "pulse" or r-f energy.

radar range—The maximum range at which a radar can ordinarily detect objects.

radar receiver—The receiver which amplifies the returned radar signal and demodulates the r-f carrier before further amplifying the desired signal and delivering it—in a form suitable for presentation—to the indicator. Unlike a radio receiver, it is more sensitive, has a lower noise level, and is designed to pass a pulse signal.

radar-reflection interval—The length of time required for a radar pulse to reach a target and return.

radar relay—Equipment for relaying radar video and appropriate synchronizing signals to a remote location.

radar resolution—The ability of a radar to distinguish between the desired target and its surroundings.

radarscope—A cathode-ray tube serving as an oscilloscope, the face of which is the radar viewing screen.

radar selector switch—A manual or motor-driven switch which transfers a plan-position-indicator repeater from one system to another, switching video, trigger, and bearing data.

radar shadow—An area shielded from radar illumination by an intervening reflecting or absorbing medium. This region appears as an area that is void of targets on a radar display.

radar target—Any reflecting object of particular interest in the path of the radar beam (usually, but not necessarily, the object being tracked).

radar trace—The pattern produced on the screen of the cathode-ray tube in a radar unit.

radar transmitter—The transmitter portion of a radar system. The unit of the radar system in which the r-f power is generated and keyed.

radiac—Acronym for RAdioactive Detection, Identification, And Computation, a descriptive term referring to the detection, identification and measurement of nuclear radiation.

radiac instrument—*See* Radiac Set.

radiac meter—*See* Radiac Set.

radiac set—Also called radiac meter or radiac instrument. Equipment for detecting, identifying, and measuring the intensity of nuclear radiations.

radiac test equipment—Equipment for testing radiac sets.

radial—Pertaining to or placed like a radius (i.e., extending or moving outward from a central point, like the spokes of a wagon wheel).

radial-beam tube—A vacuum tube producing a flat, radial electron beam which can be rotated about the axis of the tube by an external magnetic field.

radial component—A component that acts along (parallel to) a radius—as contrasted to a tangential component, which acts at right angles (perpendicular) to a radius.

radial component of the electric field—The component of an electric field in the direction of the slant-range vector or the radius vector at which an antenna pattern is measured. The radial component of the electric field is relatively small in the far field and is normally neglected.

radial field—A field of force that is directed toward or away from a point in space.

radial field cathode-ray tube—A crt in which a fine-grained, curved, high-transmission metallic mesh is placed on the exit side of the deflection area. The mesh establishes a ground plane for the post-accelerating field so that the resulting equipotential surfaces are truly spherical, creating a radial electrostatic field.

radial grating—A conformal grating consisting of wires arranged radially in a circular frame, like the spokes of a wagon wheel. The radial grating is placed inside a circular waveguide to obstruct E waves of zero order, but not the corresponding H waves.

radial lead—A lead extending out the side of a component, rather than from the end. (The latter is called an axial lead.) Some resistors have radial leads.

radial transmission line—A pair of parallel conducting planes used for propagating uniform cylindrical waves whose axes are normal to the planes.

radian—In a circle, the angle included within an arc equal to the radius of the circle. Numerically it is equal to 57°, 17′, 44.8″. A complete circle contains 2π radians. One radian equals 57.3 degrees, and 1 degree equals 0.01745 radian.

radiance—The apparent radiation of a surface. It is the same as luminance except radiance applies to all kinds of radiation instead of only light flux.

radian frequency—*See* Angular Velocity.

radian length—The distance, in a sinusoidal wave, between phases differing by an angle of one radian. It is equal to the wavelength divided by 2π.

radian per second—A unit of angular velocity.

radiant energy—Energy transmitted in the form of electromagnetic radiation (e.g., radio, heat, or light waves). It is measured in units of energy such as kilowatt-hours, ergs, joules, or calories.

radiant-energy detecting device—A device employing radiant energy to detect flaws in the surface and/or volume of solids.

radiant flux—Time rate of flow of radiant energy, expressed in watts or in ergs per second.

radiant heat—Infrared radiation from a body not hot enough to emit visible radiation.

radiant heater—An electric heating appliance with an exposed incandescent heating element.

radiant intensity—The energy emitted within a certain length of time, per unit solid angle about the direction considered.

radiant power—The rate of transfer of radiant energy.

radiant sensitivity—The output current of a phototube or camera tube divided by the incident radiant flux of a given wavelength at constant electrode voltages. The term "output current" as here used does not include the dark current.

radiant temperature—Energy traveling in the form of electromagnetic waves.

radiant transmittance—The ratio of transmitted radiant power to incident radiant power.

radiate—To emit rays from a center source—e.g., electromagnetic waves emanating from an antenna.

radiated—Energy transfer by propagation of electromagnetic fields.

radiated interference—Interference which is transmitted through the atmosphere according to the laws of electromagnetic wave propagation. (The term "radiates interference" is generally considered to include the transfer of interfering energy by inductive or capacitive coupling.)

radiated power—The total energy, in the form of Hertzian waves, radiated from an antenna.

radiating curtain—An array of dipoles in a vertical plane, positioned to reinforce each other. Usually placed one-quarter wavelength ahead of a reflecting curtain of corresponding half-wave reflecting antennas.

radiating element—Also called radiator. A basic subdivision of an antenna. It, by itself, is capable of radiating or receiving radio-frequency energy.

radiating guide—A waveguide designed to radiate energy into free space. The waves may emerge through slots or gaps in the guide, or through horns inserted into its wall.

radiation — The propagation of energy through space or through a material. It may be in the form of electromagnetic waves or corpuscular emissions. The former is usually classified according to frequency—e.g., Hertzian, infrared, (visible) light, ultraviolet, X-rays, gamma rays, etc. Corpuscular emissions are classified as alpha, beta, or cosmic.

radiation belt—*See* Van Allen Belt.

radiation counter—A device for counting radiation particles (alpha, beta, gamma neutrons, etc.) or photons of energy (X rays, etc.) usually using either scintillation or ionization resulting from the presence of the particle or photon to be measured.

radiation counter tube—*See* Counter Tube.

radiation-detector tube—A tube in which current passes between its electrodes whenever the tube is exposed to penetrating radiation. The amount of this current corresponds to the intensity of radiation.

radiation efficiency—In an antenna, the ratio of the radiated power to the total power supplied to the antenna at a given frequency.

radiation field — The electromagnetic field that breaks away from a transmitting antenna and radiates outward into space as electromagnetic waves.

radiation filter—A transparent body which transmits only selected wavelengths.

radiation intensity—In a given direction, the power radiated from an antenna per unit solid angle in that direction.

radiation lobe—*See* Lobe.

radiation loss—In a transmission system, the portion of the transmission loss due to radiation of the radio-frequency power.

radiation pattern—*See* Directional Pattern.

radiation potential—The voltage required to excite an atom or molecule and cause the emission of one of its characteristic radiation frequencies.

radiation pyrometer—Also called a radiation thermometer. A pyrometer which uses the radiant power from the object or source whose temperature is being measured. Within wide- or narrow-wavelength bands filling a definite solid angle, the radiant power impinges on a suitable detector—usually a thermocouple, thermopile, or a bolometer responsive to the heating effect of the radiant power, or a photosensitive device connected to a sensitive electric instrument.

radiation report—A formal report of radiation measurements made by an engineer skilled in interference control techniques. Usually required by the FCC prior to certification of industrial heating equipment.

radiation resistance—1. The power radiated by an antenna, divided by the square of the effective antenna current referred to a specified point. 2. The resistance which, if inserted in place of the antenna, would consume the same amount of power radiated by the antenna. 3. The characteristic of a material that enables it to retain useful properties during or after exposure to nuclear radiation.

radiation sickness—An illness resulting from exposure to radiation.

radiation survey meter—An instrument that measures instantaneous radiation.

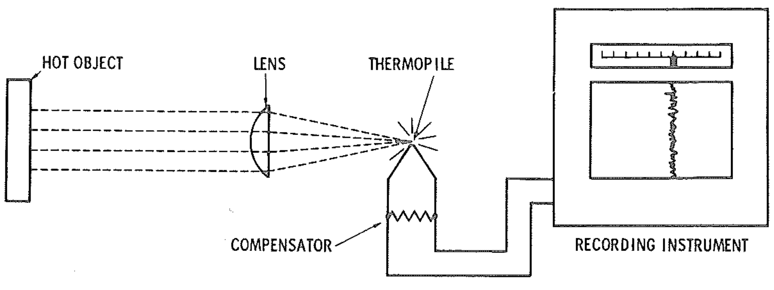

Radiation pyrometer.

radiation temperature—The temperature to which an ideal black body must be heated so it will have the same emissive power as a given source of thermal radiation.

radiation thermometer—*See* Radiation Pyrometer.

radiative equilibrium — The constant-temperature condition that exists in a material when the radiant energy absorbed and emitted are equal.

radiator—Any device which emits radiation. (*Also see* Radiating Element.)

radio—1. Communication by electromagnetic waves transmitted through space. 2. A general term, principally an adjective, applied to the use of electromagnetic waves between 10 kHz and 3,000 GHz.

radioacoustic position finding—A method of determining distance through water. This is done by closing a circuit at the same instant a charge is exploded under water. The distance to the observing station can then be calculated from the difference in arrival times between the radio signal and the sound of the explosion.

radioacoustics—A study of the production, transmission, and reproduction of sounds carried from one place to another by radiotelephony.

radioactive — Pertaining to or exhibiting radioactivity.

radioactive isotope—*See* Radioisotope.

radioactive series—A succession of radioactive elements, each derived from the disintegration of the preceding element in the series. The final element, known as the end product, is not radioactive.

radioactivity—A property exhibited by certain elements, the atomic nuclei of which spontaneously disintegrate and gradually transmute the original element into stable isotopes of that element or into another element with different chemical properties. The process is accompanied by the emission of alpha particles, beta particles, gamma rays, positrons, or similar radiations.

radio altitude—*See* Radar Altitude.

radio approach aids—Equipment making use of radio to determine the position of an aircraft with considerable accuracy from the time it is in the vicinity of an airfield or carrier until it reaches a position from which a landing can be carried out.

radioastronomy—The branch of astronomy where the radio waves emitted by certain celestial bodies are used for obtaining data about them.

radio attenuation—For one-way propagation, the ratio of the power delivered by the transmitter to the transmission line connecting it with the transmitting antenna, to the power delivered to the receiver by the transmission line connecting it with the receiving antenna.

radio beacon—Also called a radiophone or, in air operations, an aerophare. A radio transmitter, usually nondirectional, which emits identifiable signals for direction finding.

radio-beacon station—In the radionavigation service, a station the emissions of which are intended to enable a mobile station to determine its bearing or direction in relation to the radio-beacon station.

radio beam—A radio wave in which most of the energy is confined within a relatively small angle and in at least one plane.

radio bearing—The angle between the apparent direction of a source of electromagnetic waves, and a reference direction determined at a radio direction-finding station. In a true radio bearing, this reference direction is true north. Likewise, in a magnetic radio bearing, it is magnetic north.

radio broadcast—A program of music, voice, and/or other sounds broadcast from a radio transmitter for reception by the general public.

radio broadcasting—*See* Radio Broadcast.

radio channel—A band of frequencies wide enough to be used for radiocommunication. The width of a channel depends on the type of transmission and on the tolerance for the frequency of emission.

radio circuit—A means for carrying out one radiocommunication at a time in either direction between two points.

radiocommunication—An over-all term for transmission by radio of writing, signs, signals, pictures, and sounds of all kinds.

radiocommunication circuit—A radio system for carrying out one communication at a time in either direction between two points.

radiocommunication guard—A communication station designated to listen for and record transmission and to handle traffic on a designated frequency for a certain unit or units.

radio compass—*See* Direction Finder.

radio control—Remote control of apparatus by radio waves (e.g., model airplanes, boats).

radio deception — Sending false dispatches, using deceptive headings or enemy call signs, etc., by radio to deceive the enemy.

radio detection—Also called radio warning. Determining the presence of an object by radiolocation, but not its precise position.

radio detection and ranging—1. Any of certain methods or systems of using beamed and reflected electromagnetic energy for detecting and locating objects; for measuring distance, velocity, or altitude; or for other purposes such as navigating, homing, bombing, missile tracking, mapping, etc. 2. In Federal Communications Commission regulations, a radiodetermination system based on the comparison of reference signals with radio signals reflected or retransmitted from the position to be determined.

radio direction finder—A radio receiver which pinpoints the line of travel of the received waves.

radio direction finding — Radiolocation in which only the direction, not the precise location, of a source of radio emission is determined by means of a directive receiving antenna.

radio direction-finding station—A radiolocation station that determines only the direction of other stations, not their location, by monitoring their transmission.

radio doppler—A device for determining the radial component of the relative velocity of an object by observing the frequency change due to such velocity.

radioelectrocardiograph—A transmitter for broadcasting electrocardiograph signals from the subject to a remote receiver. It makes an ecg practical while the subject is execising or during natural work or home activities.

radio engineering—The branch of engineering concerned with the generation, transmission, and reception of radio (and now, television) waves, and with the design, manufacture, and testing of associated equipment.

radio fadeout—Also called the Dellinger effect. The partial or complete absorption of substantially all radio waves normally reflected by the ionospheric layers in or above the E-region.

radio field intensity—Also called radio field strength. The maximum (unless otherwise stated) electric or magnetic field intensity at a given location associated with the passage of radio waves. It is commonly expressed as the electric field intensity in microvolts, millivolts, or volts per meter. For a sinusoidal wave, its root-mean-square value is commonly stated instead.

radio field strength—*See* Radio Field Intensity.

radio field-to-noise ratio—The ratio of the radio field intensity of the desired wave, to the noise field intensity at a given location.

radio fix—1. A method by which the position source of radio signals can be determined. Two or more radio direction finders monitor the transmissions and obtain cross bearings. The position can then be pinpointed by triangulation. 2. The method by which a ship, aircraft, etc., equipped with direction finding equipment can determine its own position. This it does by obtaining radio bearings from two or more transmitting stations of known location. The position can then be pinpointed by triangulation as in (1) above.

radio-fixing aids—Equipment making use of radio to assist a user in determining his geographical position.

radio frequency—Abbreviated r-f. Any frequency at which coherent electromagnetic radiation of energy is possible.

radio-frequency alternator—A rotating generator that produces radio-frequency power.

radio-frequency amplification—Amplification of a signal by a receiver before detection, or by a transmitter before radiation.

radio-frequency choke—An inductor used to impede the flow of radio-frequency currents. Its core is generally air or pulverized iron.

radio-frequency component—In a signal or wave, the portion consisting of the r-f alternations only—not its audio rate of change in amplitude or frequency.

radio-frequency converter—A power source for producing electrical power at frequencies of 10 kHz and above.

radio-frequency generator—In industrial and dielectric heaters, a power source comprising an electron-tube oscillator, an amplifier (if used), a power supply, and associated control equipment.

radio-frequency heating—The process of heating a substance by subjecting it to a high-frequency energy field. *Also see* Dielectric Heating.

radio-frequency interference—Any electrical signal capable of being propagated into and interfering with the proper operation of electrical or electronic equipment. The frequency range of such interference may be taken to include the entire electromagnetic spectrum.

radio-frequency oscillator—Abbreviated r-f oscillator. An oscillator that generates alternating current at radio frequencies.

radio-frequency pulse—A radio-frequency carrier which is amplitude-modulated by a pulse. Between pulses, the modulated carrier has zero amplitude. (The coherence of the carrier with itself is not implied.)

radio-frequency resistance—*See* Skin Effect.

radio-frequency signal generator—*See* Service Oscillator.

radio-frequency suppressor—A device that absorbs radiated energy which might interfere with radio reception.

radio-frequency transformer—Abbreviated r-f transformer. A transformer used with radio-frequency currents.

radio-frequency welding—Also called high-frequency welding. A method of welding thermoplastics using a radio-frequency field to apply the necessary heat.

radiogoniometer—In a radio direction finder, the part that determines the phase difference between the two received signals. The Bellini-Tose system has two loop antennas, both at right angles to each other and connected to two field coils in the radiogoniometer. Bearings are obtained by rotating a search coil inductively coupled to the field coils.

radiogram—A message sent via radio telegraphy.

radiograph—*See* Radiophoto.

radiography—Any nondestructive method of internal examination in which metal objects are exposed to a beam of X-ray or gamma radiation. Differences in thickness, density, or absorption caused by internal defects or inclusions are apparent in the shadow image, either on a fluorescent screen or on photographic film placed behind the object.

radio guard—A ship, aircraft, or radio station designated to listen for and record transmissions, and to handle traffic on a designated frequency for a certain unit or units.

radio guidance system—A guidance system using radio signals to direct aircraft in flight.

radio horizon—The boundary line beyond which direct rays of the radio waves cannot be propagated over the earth's surface. This distance is not a constant; rather, it is affected by atmospheric refraction of the waves.

radio inertial-guidance system—A command type of guidance system consisting essentially of: a. A radar tracking unit, comprising radar equipment on the ground, one or more transponders in the missile, and necessary communications links to the guidance station. b. A computer that accepts missile position and velocity information from the tracking system and furnishes to the command link appropriate signals to steer the missile. c. The command link, which consists of a transmitter on the ground and an antenna and receiver on the missile; actually the command link is built into the tracking unit. d. An inertial system for partial guidance in case of radio guidance failure.

radio intelligence—Interception and interpretation of enemy radio transmissions.

radio interference—Undesired conducted or radiated electrical disturbances, including transients, which can interfere with the operation of electrical or electronic equipment. These disturbances fall between 14 kHz and 10 GHz.

radioisotope—Also called a radioactive isotope. It is the isotope produced when an element is placed into a nuclear reactor and bombarded with neutrons. Radioisotopes are used as tracers in many areas of science and industry. Like all isotopes, they decay spontaneously with the emission of their radiation, at a definite rate measured by their half-lives.

radio jamming—Blocking communications by sending overpowering interference signals.

radio knife—A form of surgical knife that uses a high-frequency electric arc at its tip to cut tissue and, at the same time, also sterilizes the edges of the wound.

radio landing beam—A distribution of vertical, directional radio waves used for guiding aircraft into a landing.

radiolocation—Use of the constant-velocity or rectilinear-propagation characteristics of radio waves to detect an object or to determine its direction, position, or motion.

radiolocation service—A radio service in which radiolocation is used.

radiolocation station—A radio station in the radiolocation service.

radio log—A record of all messages sent and received, transmitter tests made, and other important information pertaining to the operation of a particular station.

radiology—The branch of physics concerned with X-rays, gamma rays, and other penetrating radiations.

radioluminescence—Luminescence produced by radiant energy (e.g., by X-rays, radioactive emissions, alpha particles, or electrons).

radiomagnetic indicator—Abbreviated RMI. A navigational instrument used by land vehicles. It presents a display combining the heading and the relative and magnetic bearings of the vehicle with the relative bearing of a radio station whose location is known.

radio marker beacon—In the aeronautical radionavigation service, a land station which provides a signal to designate a small area above the station.

radiometallography—X-ray examination of the crystalline structure and other characteristics of metals and alloys.

radiometeorograph—A meteorograph which, when carried into the stratosphere by an unmanned gas-filled rubber balloon, automatically reports atmospheric conditions by radio as it ascends into the stratosphere. The ultrahigh-frequency signals are transmitted so that they can be recorded and interpreted in terms of pressure, temperature, and humidity. *Also see* Radiosonde.

radiometer—An instrument for measuring the intensity of thermal (especially infrared) radiation.

radiomicrometer—*See* Microradiometer.

radionavigation—Navigational use of radiolocation for determining position or direction, or for providing a warning of obstructions.

radionavigation land station—A fixed station in the radionavigation service—i.e., one not intended for mobile operation.

radionavigation mobile station—A radionavigation station operated from a vehicle.

radionavigation service—A radiolocation service used for radionavigation.

radio net—A system of radio stations.

radio-noise field intensity—A measure of the field intensity of interfering electromagnetic waves at some point (e.g., a radio receiving station). In practice, the field intensity itself is not measured, but some proportionate quantity.

radiopaque—Not penetrable by X-rays or other radiation. (The opposite of radioparent.)

radioparent—Penetrable by X-rays or other radiation. (The opposite of radiopaque.)

radiophone—*See* Radiotelephone and Radio Beacon.

radiophoto—Also called radiophotography, radiophotograph, radiophotogram, or facsimile. The transmission of photographs and other illustrations by radio.

radiophotogram—*See* Radiophoto.

radiophotoluminescence—The property whereby the previous exposure of certain

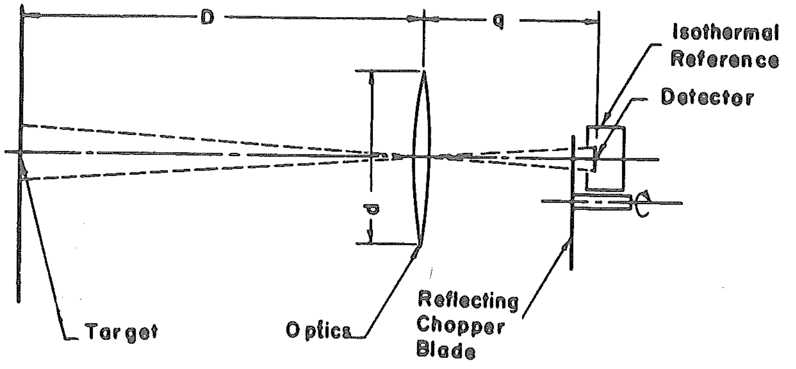

Simple radiometer.

materials to nuclear radiation enables them to give off visible light when irradiated with ultraviolet light.

radio prospecting—The use of radio equipment to locate mineral or oil deposits.

radio proximity fuse—A radio device which detonates a missile by electromagnetic interaction within a predetermined distance from the target.

radio pulse—An intense, split-second burst of electromagnetic energy.

radio range—A radionavigational facility, the emissions of which provide radial lines of position by having special characteristics that are recognizable as bearing information and useful in lateral guidance of aircraft.

radio range beacon—A radionavigation land station in the aeronautical radionavigation service providing radio equisignal zones.

radio range finding—Determination of range by means of radio waves.

radio range leg—The space within which an aircraft will receive an on-course signal from a radio-range station.

radio-range monitor — An instrument that automatically monitors the signal from a radio-range beacon and warns attending personnel when the transmitter deviates from its specified current bearings. It also transmits a distinctive warning to approaching planes whenever trouble exists at the beacon.

radio-range station—A land station that operates in the aeronautical radionavigation service and provides radial equisignal zones.

radio receiver—A device for converting radio waves into signals perceptible to humans.

radio reception—Reception of radioed messages, programs, or other intelligence.

radio relay—*See* Radio Relay System.

radio-relay system—Also called radio relay. A radio transmission system in which the signals are received and retransmitted from point to point by intermediate radio stations.

radio set—A radio transmitter, radio receiver, or a combination of the two.

radio shielding—A metallic covering over all electric wiring and ignition apparatus which is grounded at frequent intervals for the purpose of eliminating electric interference with radiocommunications.

radio signal—A signal that is transmitted by radio.

radio silence—A period during which all or certain radio equipment capable of radiation is kept inoperative.

radiosonde—A balloonborne instrument for the simultaneous measurement and transmission of meteorological data. The instrument consists of transducers for the measurement of pressure, temperature, and humidity; a modulator for the conversion of the output of the transducers to a quantity which controls a property of the radio-frequency signal; a selector switch which determines the sequence in which the parameters are to be transmitted; and a transmitter which generates the radio-frequency carrier.

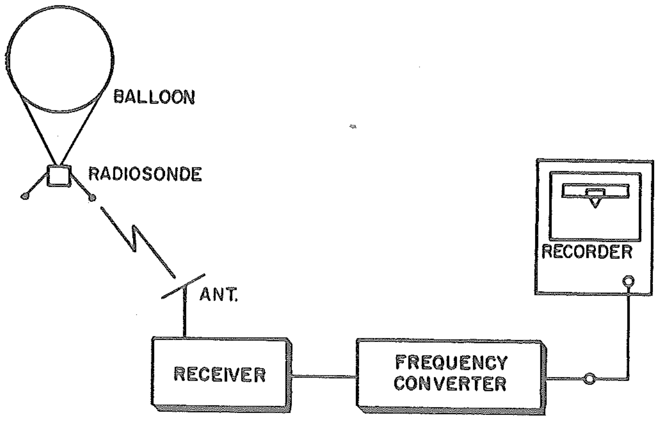

Radiosonde system.

radiosonde recorder—An instrument which is located at the surface observing station and is used to record the data presented by a radiosonde aloft. The mechanism of the recorder depends upon the type of radiosonde system used.

radiosonde transmitter—The component of the radiosonde which includes the modulating blocking oscillator and the radio-frequency carrier oscillator.

radiosonobuoy—*See* Sonobuoy.

radio spectrum—The range of frequencies of electromagnetic radiation usable for radiocommunication. The radio spectrum may range from about 10 kHz to over 300 GHz. Corresponding wavelengths are 30 km to 1 mm.

radio station—An assemblage of equipment for radio transmission, reception, or both.

radio-station interference — Interference caused by reception of radio waves from other than the desired station.

radiotelegraph transmitter—A radio transmitter capable of handling code signals.

radiotelegraphy — Radiocommunication by means of the International Morse Code or some other, similarly coded signal.

radiotelephone—Also called radiophone. The complete radio transmitter, receiver, and associated equipment required at one station for radiotelephony.

radiotelephone distress signal—The spoken word MAYDAY (phonetic spelling of the French expression *m' aidez,* or "help me"). It corresponds to S O S in radiotelegraphy and is used by aircraft, ships, etc., needing help.

radiotelephone transmitter—A radio transmitter capable of handling audio-frequency modulation (e.g., voice and music).

radiotelephony—Two-way transmission and reception of sounds by radio.

radio telescope—A very sensitive radio receiver that is used in conjunction with a large and highly directional antenna to receive signals from radio stars.

radiotherapy—The treatment of diseases by the application of rays from radioactive substances (e.g., X-rays).

radiothermics—The application of heat generated by radio waves (e.g., in diathermy or electronic heating).

radio transmission—Transmission of signals by electromagnetic waves other than light or heat waves.

radio transmitter—A device capable of producing radio-frequency power and used for radio transmission.

radiotransparent—Permitting the passage of X-rays or similar radiation.

radio tube—A general term for any type of electron tube used in electronic equipment.

radiovision—An early name for television.

radio warning—*See* Radio Detection.

radio watch—Also called watch. The vigil maintained by an operator when on duty in the radio room of a vessel and listening for signals—especially on the international distress frequencies.

radio wave—Also called a Hertzian wave, or a wave. An electromagnetic wave produced by rapid reversals of current in a conductor. Such a wave travels through space at approximately the speed of light.

radio wavefront distortion—A change in the direction or advance of a radio wave.

radio-wave propagation—The transfer of energy by electromagnetic radiation at frequencies below approximately 3×10^{12} hertz.

radix—Also called the base. The total number of distinct marks or symbols used in a numbering system. For example, since the decimal numbering system uses ten symbols (0, 1, 2, 3, 4, 5, 6, 7, 8, 9), the radix is 10. In the binary numbering system the radix is 2, because there are only two marks or symbols (0, 1).

radix point—Also called base point, and binary point, decimal point, etc., depending on the numbering system. The index which separates the integral and fractional digits of the numbering system in which the quantity is represented.

radome—Also called a radar dome. The housing which protects a radar antenna from the elements, but does not block radio frequencies.

radux—A long-distance, low-frequency navigational system which provides hyperbolic lines of position. It is of the continuous-wave, phase-comparison type.

railing—Radar pulse jamming at high recurrence rates (50 to 150 kHz). On a radar indicator it results in an image resembling fence railing.

railings—The pattern produced on an A-scope by c-w modulated with a high-frequency signal. It appears as a series of vertical lines resembling target echoes along the baseline.

rainbow—The technique which applies pulse-to-pulse frequency changing to identifying and discriminating against decoys and chaff.

rainbow generator—A signal generator with which the entire color spectrum can be produced on the screen of a color television receiver. The colors merge together as in a rainbow. (*Also see* Keyed Rainbow Generator.)

ramark—Also called a radar marker. A fixed radar transmitter which emits a continuous signal that is used as a bearing indication on a radar display.

ramp input—A change, in an input signal, which varies at a constant rate.

Ramsauer effect—The absorption of slow-moving electrons by intervening matter.

random—Irregular; having no set pattern.

random access—Access to a computer storage under conditions whereby there is no rule for predetermining the position from where the next item of information is to be obtained.

random-access memory—A memory system for which the access time is independent of the location being addressed.

random-access programming — Programming a problem for a computer without regard to the access time to the information in the registers called for in the program.

random error—The inherent imprecision of a given process of measurement; the unpredictable component of repeated independent measurements on the same object under sensible uniform conditions, usually of an approximately normal (Gaussian) frequency distribution, other than systematic or erratic errors and mistakes, sometimes called short-period errors.

random experiment — An experiment that can be repeated a large number of times but may yield different results each time, even when performed under similar circumstances.

random failure—Also called chance failure. Any chance failure the occurrence of which at a given time is unpredictable.

random-function generator—A device which generates nonrepetitive signals which are distributed over a broad frequency range.

random-impulse generator—A generator of electrical impulses which occur at random rather than at specific intervals.

random interlace — A condition in which there is no fixed relationship between adjacent scanning lines and successive fields.

randomness — A condition of equal chance for the occurrence of any of the possible outcomes.

random noise—Also called fluctuation noise. Large numbers of random, overlapping transient disturbances. (*Also see* Broad-Band Electrical Noise.)

random-noise generator—A generator of a succession of random signals which are distributed over a wide frequency spectrum.

random number—1. A set of digits such that each successive digit is equally likely to be any of n digits to the base n of the number. 2. A number composed of digits selected from an orderless sequence of digits.

random-number generator — A special machine routine or hardware that produces a random number or series of random numbers in accordance with specified limitations.

random pard—Pertains to that portion of the total pard in an electronic power supply which is not periodic. This phenomenon is frequently referred to as noise.

random pulsing—Varying the repetition rate of pulses by noise modulation or continuous frequency change.

random sample — A sample in which every item in the lot is equally likely to be selected in the sample.

random signals—Waveforms having at least one parameter (usually amplitude) that is a random function of time (e.g., thermal noise or shot noise).

random variable—1. Also called variate. The result of a random experiment. 2. A discrete or continuous variable which may assume any one of a number of values, each having the same probability of occurrence. 3. Also called stochastic variable. Any signal the amplitude or phase of which cannot be predicted by a study of previous values of the signal.

random velocity—The instantaneous velocity of a particle without regard to direction. It may be characterized by its distribution function or by its average, root-mean-square, or most probable value.

random vibration—A vibration generally composed of a broad, continuous spectrum of frequencies, the instantaneous magnitude of which cannot be specified to any given moment of time. (If random vibration has instantaneous magnitudes distributed according to the Gaussian distribution, it is called Gaussian random vibration.) *Also see* White Noise.

random winding—A coil winding in which the turns and layers are not regularly positioned or spaced but are positioned haphazardly.

range—1. The maximum useful distance of a radar or radio transmitter. 2. The difference between the maximum and the minimum value of a variable.

range-amplitude display—A radar display in which a time base provides the range scale from which echoes appear as deflections normal to the base.

range calibration—Adjustment of radar-range indications by use of known range targets or delayed signals so that, when on target, the radar set will indicate the correct range.

range coding—A method of coding a beacon response so that the response appears as a series of pulses on a radarscope. The coding provides identification.

range gate—A gate voltage used to select radar echoes from a very short-range interval.

range-height indicator—A radar display on which an echo appears as a bright spot on a rectangular field. The slant range is indicated along the X-axis, and the height above the horizontal plane (on a magnified scale) along the Y-axis. A cursor shows the height above the earth.

range mark—*See* Distance Mark.

range marker—A variable or movable discontinuity in the range time base of a radar display (in the case of a ppi, a ring). It is used for measuring the range of an echo or calibrating the range scale.

range of an instrument—*See* Total Range of an Instrument.

range ring—An accurate, adjustable ranging mark on a plan-position indicator corresponding to a range step on a type-M indicator.

range step—The vertical displacement on an M-indicator sweep to measure range.

range unit—A radar-system component used for control and indication (usually counters) of range measurements.

range zero—Alignment of the start of a sweep trace with zero range.

ranging oscillator—An oscillator circuit containing an LC resonant combination in the cathode circuit, usually used in radar equipment to provide range marks.

rapid memory—In a computer, that selection of the whole memory from which information may be obtained in the shortest possible time.

rare gas—*See* Noble Gas.

raster—1. On the screen of a cathode-ray tube, a predetermined pattern of scanning lines which provide substantially uniform coverage of an area. 2. The illuminated area produced by the scanning lines on a television picture tube when no signal is being received.

raster burn—In camera tubes, a change in the characteristics of the area that has been scanned. As a result, a spurious signal corresponding to that area will be produced when a larger or tilted raster is scanned.

ratchet relay—A stepping relay actuated by an armature-driven ratchet.

rate action—Also called derivative action. Corrective action, the rate of which is determined by how fast the error being corrected is increasing.

rate effect—The anode-voltage transients that cause PNPN devices to switch into high conduction.

rated operational voltage — The voltage on which a specific application rating of a device is based.

rated output—The output power, voltage, current, etc., at which a machine, device, or apparatus is designed to operate under normal conditions.

rated power output—The normal radio-frequency, power-output capability (peak or average) of a transmitter under optimum adjustment and operation conditions.

rated thermal current—In a contactor, that current at which the permissible temperature rise is reached.

rated voltage—The voltage at which a device or component is designed to operate under normal conditions.

rate generator — A proportional element which converts angular speed into a constant-frequency output voltage. (*Also see* Angular Velocity.)

rate grown junction—A grown junction, in a semiconductor, produced by varying the rate of crystal growth periodically so that N-type impurities alternately predominate.

rate-grown transistor—Also called graded-junction transistor. A variation of the double-doped transistor, in which N- and P-type impurities are added to the melt.

rate gyro—A particular kind of gyroscope used for measuring angular rates. A system of three rate gyros, each oriented to one of three mutually perpendicular axes — roll, pitch, and yaw—can control a missile or aircraft by detecting angular rates and then generating proportional corrective signals.

rate limit—*See* Slewing Rate.

rate limiting—Nonlinear behavior in an amplifier due to its limited ability to produce large, rapid changes in output voltage (slewing), restricting it to rates of change of voltage lower than might be predicted by observing the small signal frequency response.

ratemeter—An instrument for measuring the rate at which counts are received, usually in counts per minute.

rate of decay—The rate at which the sound-pressure level (velocity level, or sound-energy density level) is decreasing at a given point

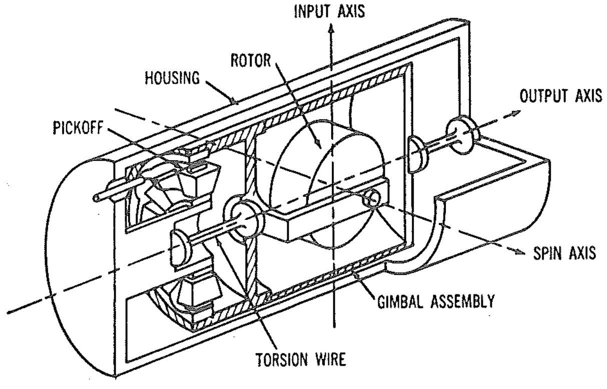

Rate gyro.

and at a given time. The practical unit is the decibel per second.

rate-of-rise relay—*See* Instantaneous Overcurrent Relay.

rate of transmission—*See* Speed Of Transmission.

rate receiver—A device for receiving a signal giving the rate of speed of a launched missile.

rate signal — A signal proportional to the time derivative of a specified variable.

rating—A value which establishes either a limiting capability or a limiting condition for an electron device. It is determined for specified values of environment and operation, and may be stated in any suitable terms. (Limiting conditions may be either maxima or minima.)

rating system—The set of principles upon which ratings are established and by which the interpretation of the ratings is determined.

ratio—The value obtained by dividing one number by another. This value indicates their relationship to each other.

ratio arms—Two adjacent arms of a Wheatstone bridge, both having an adjustable resistance and so arranged that they can be set to have any of several fixed ratios to each other.

ratio calibration—1. The calibration of a dimensionless quantity that represents the ratio of one of its values to another. 2. A method by which potentiometric transducers may be calibrated, in which the value of the measurand is expressed in terms of decimal fractions representing the ratio of output resistance to total resistance.

ratio detector—An f-m detector which inherently discriminates against amplitude modulation. A pair of diodes are connected in such a manner that the audio output is proportionate to the ratio of the f-m voltages applied to them.

ratio meter—An instrument which measures electrically the quotient of two quantities. It generally has no mechanical control means such as springs. Instead, it operates by the balancing of electromagnetic forces, which are a function of the position of the moving element.

ratio of transformation—The ratio of the secondary voltage of a transformer to the primary voltage under no-load conditions, or the corresponding ratio of currents in a current transformer.

ratio squelch—*See* Squelch Circuit.

ratio-type telemeter—Also called a position-type telemeter. A telemeter in which the relative phase position between, or magnitude relation of, two or more electrical quantities is the translating means.

rat race—A magic tee modification for the acceptance of higher power. A circular loop of coaxial line closed upon itself and having four branching connections. *Also see* Hybrid Ring.

rat race—*See* Hybrid Ring.

raw data — Information that has not been processed or reduced.

raw tape—Also called virgin tape or blank tape. A term sometimes used to describe tape that has not been recorded.

ray-control electrode — An electrode that controls the position of the electron beam on the screen of a cathode-ray tuning-indicator tube.

raydist—A system using c-w transmission to provide hyperbolic lines of position through r-f phase comparison techniques. The system is used for surveying or ship positioning in a two-dimensional array. The frequency band is 1.7 to 2.5 MHz. Similar to lorac in principle.

Rayleigh disc—A special acoustic radiometer used for the fundamental measurement of particle velocity.

Rayleigh distribution—Frequency distribution for an infinite number of quantities of the same magnitude, but of random phase relationships. Sky-wave field intensities follow the Rayleigh distribution for intervals of one minute or less.

Rayleigh distribution (fading)—A statistical distribution describing the magnitude of a phaser composed of the sum of many component phasers randomly distributed in amplitude and phase. Fading of signals caused by cancellation and reinforcement of contributions received over separate paths often exhibits a Rayleigh distribution.

Rayleigh line—In scattered radiation, a spectrum line which has the same frequency as the corresponding incident radiation.

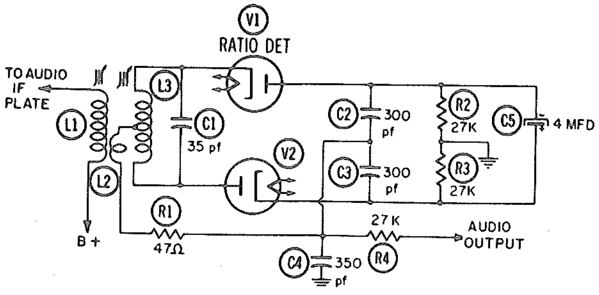

Ratio detector.

Rayleigh reciprocity theorem—The reciprocal relationship for an antenna when transmitting or receiving. The effective heights and the radiation resistance and pattern are alike, whether the antenna is transmitting or receiving.

Rayleigh scattering—Scattering of radiation by minute particles suspended in air (e.g., by dust).

Rayleigh wave—A surface wave associated with the free boundary of a solid. The wave is of maximum intensity at the surface, but diminishes quite rapidly as it proceeds into the solid.

ray path—An imaginary line, perpendicular to the wavefront, which describes the path along which the energy associated with a point on a wavefront moves.

Raysistor—Raytheon trade name for a four-terminal optoelectronic component designed for applications in a variety of control functions, providing comparatively noise-free control of a-c or d-c signals over a wide dynamic range, without transients or contact chatter, and with a high insulation and electrical isolation between the signal and control circuit. It consists of a light source and a photoresistive element assembled in a light-tight case.

RC—Symbol for resistance-capacitance, resistance-coupled, or ray-control electrode.

RC amplifier—Abbreviation for resistance-capacitance coupled amplifier.

RC circuit—A time-determining network of resistors and capacitors in which the time constant is defined as resistance times capacitance.

RC constant—The time constant of a resistor-capacitor circuit, equal in seconds to the resistance value in ohms multiplied by the capacitance value in farads.

RC coupling — *See* Resistance-Capacitance Coupling.

RC differentiator—*See* Resistance-Capacitance differentiator.

RC filter—*See* Resistance-Capacitance filter.

RC network—A circuit containing resistances and capacitances arranged in a particular manner to perform a specific function.

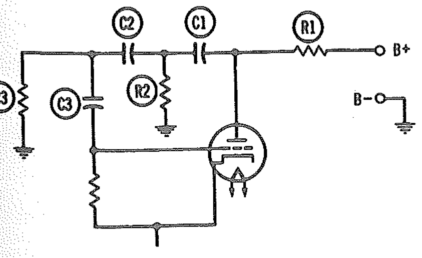

RC oscillator.

RC oscillator—An oscillator in which the frequency is determined by resistance-capacitance elements. *Also see* Resistance-Capacitance Oscillator.

RCTL—Abbreviation for resistor capacitor transistor logic.

rd—Abbreviation for rutherford.

rdf—Abbreviation for radio direction finder (or finding).

reach-through voltage — That value of reverse voltage for which the depletion layer in a reverse-biased PN junction spreads sufficiently to make electrical contact with another junction.

reacquisition time—The time required for a tracking radar to relock on the target after the radar's automatic tracking mechanism has been disengaged.

reactance—Symbolized by X. Opposition to the flow of alternating current. Capacitive reactance (X_C) is the opposition offered by capacitors, and inductive reactance (X_L) is the opposition offered by a coil or other inductance. Both reactances are measured in ohms.

reactance drop—The voltage drop in quadrature with the current.

reactance factor—In a conductor, the ratio of its a-c resistance to its ohmic resistance.

reactance frequency multiplier — A frequency multiplier the essential element of which is a nonlinear reactor, where the nonlinearity of the reactor is used to generate harmonics of a sinusoidal source.

reactance grounded—Grounded through a reactance.

reactance modulator—A modulator, the reactance of which can be varied in accordance with the instantaneous amplitude of the modulating electromotive force applied. It is normally an electron-tube circuit that usually modulates the phase or frequency.

reactance relay—A form of impedance relay, the operation of which is a function of the reactance of the circuit.

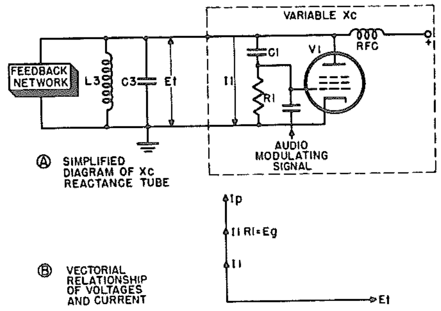
VARIABLE Xc
RFC
+
C1
V1
FEEDBACK NETWORK
L3
C3
Et
I1
R1
AUDIO MODULATING SIGNAL
Ⓐ SIMPLIFIED DIAGRAM OF Xc REACTANCE TUBE
Ip
I1 R1=Eg
I1
Et
Ⓑ VECTORIAL RELATIONSHIP OF VOLTAGES AND CURRENT

Reactance-tube modulator.

reactance tube—A stage connected in parallel with the tank circuit of an oscillator to produce a signal that will either lead or lag the signal produced by the tank circuit. The reactance tube is said to be inductive if this signal leads the one produced by the tank circuit, and is capacitive if the signal lags the tank-circuit signal. The resultant oscillator signal will be a combination of the two, and its phase or frequency can be advanced or retarded as the bias on the reactance tube is increased or decreased.

reactivation—Application of an above-normal voltage to a thoriated filament for a few seconds, to bring a fresh layer of thorium atoms to the filament surface and thereby improve electron emission.

reactive—Pertaining to either inductive or capacitive reactance. A reactive circuit has a higher reactance than resistance.

reactive attenuator—An attenuator that dissipates almost no energy.

reactive balance—1. The capacitive or inductive balance which is often required to null the output of certain transducers or systems when the excitation and/or the output is given in terms of alternating current. 2. The condition of an a-c circuit where the phase angle between the voltage and current is zero. 3. The amount of corrective capacitance or inductance required to null the output of certain transducers or systems having a-c excitation.

reactive factor—The ratio of reactive power to total power in a circuit.

reactive-factor meter—An instrument for measuring reactive factor.

reactive load—A load having reactance (i.e., a capacitive or inductive load), as opposed to a resistive load.

reactive near-field region—That region of the field of an antenna immediately surrounding the antenna wherein the reactive field predominates.

reactive power—Also called wattless power. The reactive voltage times the current, or the voltage times the reactive current, in an a-c circuit. Unit measurement is the var.

reactive volt-ampere—*See* Volt-Ampere Reactive.

reactive volt-ampere-hour meter—*See* Varhour Meter.

reactive volt-ampere meter—*See* Varmeter.

reactor—A physical device used primarily to introduce reactance or susceptance into a branch. (*Also see* Nuclear Reactor.)

reactor-start motor—A form of split-phase motor designed for starting with a reactor in series with the main winding. The reactor is short-circuited (or otherwise made ineffective) and the auxiliary circuit is opened as soon as the motor attains a predetermined speed.

read—In a computer: 1. To copy, usually from one form of storage to another. 2. To sense the meaning of an arrangement of hardware representing information. 3. To extract information.

readability—The ability to be understood—specifically, the understandability of signals sent by any means of telecommunications.

read-around—*See* Read-around Ratio.

read-around ratio—Also called read-around. In electrostatic storage tubes, the number of successive times information can be recorded as an electrostatic charge on a single spot in the array without producing the necessity for restoring the charge on surrounding spots.

reader—In a computer a device which converts information in one form of storage to information in another form of storage.

read head—In a computer, a device that converts digital information stored on a magnetic tape or on a drum into electrical signals usable by computer arithmetic.

reading access time—In a computer, the time before a word may be used during the reading cycle.

reading rate—In a computer, the number of characters, cards, etc. that can be sensed by an input unit in a given time.

readout—The manner in which a computer displays the processed information—e.g., digital visual display, punched tape, punched cards, automatic typewriter, etc.

readout device—In a computer, a device, consisting usually of physical equipment, that records the computer output either as a curve or as a set of printed numbers or letters.

readout equipment—The electronic apparatus that provides indications and/or recordings of transducer output.

read pulse—A pulse applied to one or more binary cells to determine whether a bit of information is stored there.

readthrough—The continuous recovery in an audio channel, of the target modulation, making possible rapid evaluation of the effectiveness of a jamming effort.

read-write head—The device that reads and writes information on tape, drum, or disc storage devices.

real power—The component of apparent power that represents true work in an a-c circuit. It is expressed in watts and is equal to the apparent power times the power factor.

real time—In solving a problem, a speed sufficient to give an answer within the actual time the problem must be solved.

real-time operation—Operations performed on a computer in time with a physical process so that the answers obtained are useful in controlling that process.

rear projection—A projection television system in which the picture is projected on a ground-glass screen to be viewed from the opposite side of that screen.

rebatron—A relativistic electron-bunching accelerator which produces a very tightly bunched beam with little velocity modulation, but high harmonic content. The beam can be used to excite structures which are large compared to a wavelength.

rebecca—An airborne interrogator-responsor of the British rebecca-eureka navigation system. It can also be used with a special ground beacon known as babs to provide low-approach facilities.

rebecca-eureka system—A British radar navigational system employing an airborne interrogator (rebecca) and a ground transponder beacon (eureka). It provides homing to an airfield from distances of up to 90 miles.

rebroadcast—The reception and the simultaneous or subsequent retransmission of a radio or television program by a broadcast station.

recalescent point—The temperature at which heat is suddenly liberated as the temperature of a heated metal drops.

receiver—1. Any device equipped for reception of incoming electrically transmitted signals. (*Also see* Earphone.) 2. An electromechanical device for converting electrical energy into sound waves.

receiver bandwidth—The spread in frequency between the half-power points on the response curve of a receiver.

receiver gating—Application of operating voltages to one or more stages of a receiver, only during the part of a cycle when reception is desired.

receiver incremental tuning—A control feature to permit receiver tuning (of a transceiver) up to 3 kHz to either side of the transmitter frequency.

receiver lockout system—In mobile communications, an arrangement of control circuits whereby only one receiver can feed the system at one time, to avoid distortion.

receiver noise figure—The ratio of noise voltage in a given receiver to that of a theoretically perfect receiver.

receiver primaries—*See* Display Primaries.

receiver radiation—Radiation of interfering electromagnetic signals by any oscillator of a receiver.

receiver sensitivity—The lower limit of useful signal input to the receiver. It is set by the signal-to-noise ratio at the output.

receiving amplifier—The amplifier used at the receiving end of a system to raise the level of the signal.

receiving antenna—A device for converting received space-propagated electromagnetic energy into electrical energy.

receiving circuit—An apparatus and connections used exclusively for the reception of messages at a radiotelephone or radiotelegraph station.

receiving equipment—The equipment (amplifiers, filters, oscillator, demodulator, etc.) associated with incoming signals.

receiving-loop loss—That part of the repetition equivalent assignable to the station set, subscriber line, and battery-supply circuit on the receiving end of a telephone line.

receiving perforator—In printing telegraph systems, an apparatus that punches a paper strip automatically, in accordance with the arriving signals. When the paper strip is later passed through a printing telegraph machine, the signals will be reproduced as printed messages, ready for delivery to the customer.

receptacle—Usually the fixed or stationary half of a two-piece multiple-contact connector. Also, the connector half usually mounted on a panel and containing socket contacts.

reception—Listening to, copying, recording, or viewing any form of emission.

rechargeable—Capable of being recharged. Usually used in reference to secondary cells or batteries.

rechargeable primary cell—A cell which is ordinarily used in "one-shot" service, but which is capable of a limited number of charge/discharge cycles (frequently five to ten). Applied to silver-zinc missile batteries.

reciprocal—The number *1* (unity) divided by a quantity—e.g., the reciprocal of 2 is ½; of 4, ¼, etc.

reciprocal-energy theorem—If an electromotive force E_1 in one branch of a circuit produces a current I_2 in any other branch, and if an electromotive force E_2 inserted into this other branch produces a current I_1 in the first branch, then $I_1E_1 = I_2E_2$.

reciprocal impedance—Two impedances Z_1 and Z_2 are said to be reciprocal impedances with respect to an impedance Z (invariably a resistance) if they are so related as to satisfy the equation $Z_1Z_2 = Z^2$.

reciprocal transducer—A transducer that satisfies the principles of reciprocity—i.e., if the roles of excitation and response are interchanged, the ratio of excitation to response will remain the same.

reciprocation—The process of deriving a reciprocal impedance from a given impedance or finding a reciprocal network for a given network.

reciprocity theorem—In any system composed of linear bilateral impedances, if an electromotive force E is applied between any two terminals and the current I is measured in any branch, their ratio (called the "transfer impedance") will be equal to the ratio obtained if the positions of E and I are interchanged.

reclosing relay—Any voltage, current, power, etc., relay which recloses a circuit automatically.

recognition device—A device which can identify any number of a set of distinguishable entities.

recognition differential—For a specified listening system, the amount by which the signal level exceeds the noise level that is presented to the ear when a 50% probability of detection of the signal exists.

recombination—The simultaneous elimination of both an electron and a hole in a semiconductor.

recombination velocity (on a semiconductor surface)—The normal component of the electron (hole)-current density at the surface, divided by the excess electron (hole) -charge density at the surface.

reconditioned-carrier reception—Also called exalted-carrier reception. Reception in which the carrier is separated from the sidebands in order to eliminate amplitude variations and noise, and then is increased and added to the sideband in order to provide a relatively undistorted output. This method is frequently employed with a reduced-carrier single-sideband transmitter.

record—1. A unit of information comprised of smaller units (e.g., fields, characters, words). 2. A group of related facts or fields of information treated as a unit. 3. To preserve for later reproduction.

record changer—A device which will automatically play a number of phonograph records in succession.

record compensator—Also called a record equalizer. An electrical network that compensates for different frequency-response curves in various recording techniques.

recorded tape—Also called a prerecorded tape. A tape that contains music, dialogue, etc., and is sold to audiophiles and others for their listening pleasure.

recorded value—The value recorded by the marking device on a chart with reference to the division lines marked on the chart.

record equalizer—*See* Record Compensator.

recorder—Also called a recording instrument. An instrument that makes a permanent record of varying electrical impulses—e.g., a code recorder, which punches code messages into a paper tape; a sound recorder, which preserves music and voices on disc, film, tape, or wire; a facsimile recorder, which reproduces pictures and text on paper; and a video recorder, which records television pictures on film or tape.

record gap—In a computer, a space between records on a tape. It is usually produced by acceleration or deceleration of the tape during the write operation.

recording ammeter—An ammeter that provides a permanent recording of the value of either an alternating or a direct current.

recording blank—*See* Recording Disc.

recording channel—One of several independent recorders in a recording system, or independent recording tracks on a recording medium.

recording demand meter—Also called demand recorder. An instrument that records the average value of the load in a circuit during successive short periods.

recording disc — Also called a recording blank. A blank (unrecorded) disc made for recording purposes.

recording head—A magnetic head that transforms electrical variations into magnetic variations for storage on magnetic media. (*Also see* Cutter.)

recording instrument—Also called a recorder or graphic instrument. An instrument which makes a graphic record of the value of one or more quantities as a function of another variable (usually time).

recording lamp—A light source used in the variable-density system of sound recording on movie film. Its intensity varies in step with the variations of the audio-frequency signal sent through it.

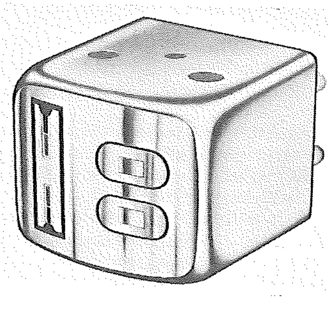

Recording head.

recording level—The amplifier output required to provide a satisfactory recording.

recording loss—In mechanical recording, the loss that occurs in the recorded level because the amplitude executed by the recording stylus differs from the amplitude of the wave in the recording medium.

recording noise—Noise induced by the amplifier and other components of a recorder.

recording-reproducing head—A dual-purpose head used in magnetic recording.

recording spot—1. An instantaneous area acted on by the registering system of a facsimile recorder. 2. The elemental area at the facsimile recorder.

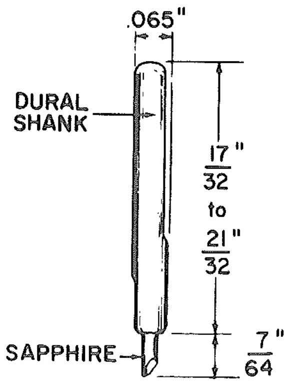

Recording stylus.

recording stylus—The tool which inscribes a groove into the recording medium.

recording voltmeter—A voltmeter that provides a permanent record of the value of either alternating or direct voltage.

record medium—In a facsimile recorder, the physical medium onto which the image of the subject copy is formed.

record player — A motor-driven turntable, pickup arm, and stylus, for converting the signals impressed onto a phonograph record into a corresponding a-f voltage. This voltage is then applied to an amplifier (usually contained within the record player cabinet) for amplification and conversion to sound waves.

record sheet—In a facsimile recorder, a sheet or medium upon which the image of the subject copy is recorded.

recovery time—1. The time required for a fired ATR tube in its mount to deionize to the level where its normalized conductance and susceptance are within the specified ranges. 2. The time required for the control electrode in a gas tube to regain control after the anode current has been interrupted. 3. In Geiger-Mueller counters, the minimum time from the start of a counted pulse to the instant a succeeding pulse can attain a specified percentage of the maximum amplitude of the counted pulse. 4. The time required for a fired TR or pre-TR tube to deionize to the level where the attenuation of a low-level radio-frequency signal transmitted through the tube drops to the specified value. 5. In a radar or its component, the time required—after the end of the transmitted pulse—for recovery to a specified relation between receiving sensitivity or received signal and the normal value. 6. The interval required, after a sudden decrease in input-signal amplitude to a system or component, for a specified percentage (usually 63%) of the ultimate change in amplification or attenuation to be attained. 7. In a thermal time-delay relay, the cooling time required from heater de-energization to re-energization such that the new time delay is 85% of that exhibited from a cold start. 8. In a power supply, the time required for recovery of the load voltage from a step change in load current or line voltage. Also called response time.

rectangular scanning — A two-dimensional sector scan in which a slow sector scan in one direction is superimposed perpendicularly onto a rapid sector scan.

rectangular wave — A periodic wave which alternately assumes one of two fixed values, the time of transition being negligible in comparison with the duration of each fixed value.

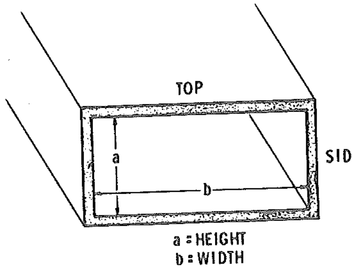

Rectangular waveguide.

rectangular waveguide—A waveguide with a rectangular cross section.

rectification—The conversion of alternating current into unidirectional or direct current.

rectification efficiency—Ratio of the direct-current power output to the alternating-current power input of a rectifier.

rectification factor—The change in average current of an electrode divided by the change in amplitude of the alternating sinusoidal voltage applied to the same electrode, the direct voltages of all electrodes being maintained constant.

rectifier—A device which, by virtue of its asymmetrical conduction characteristic, converts an alternating current into a unidirectional current.

rectifier instrument—The combination of an instrument sensitive to direct current, and a rectifying device whereby alternating currents or voltages can be measured.

rectifier meter—*See* Rectifier Instrument.

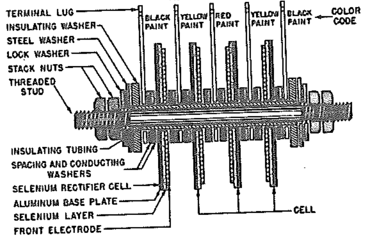

Rectifier stack.

rectifier stack—A dry-disc rectifier made up of layers of individual rectifier discs (e.g., a selenium or copper-oxide rectifier).

rectifier transformer—A transformer the primary of which operates at the fundamental frequency of the a-c system and the secondary of which has one or more windings conductively connected to the main electrodes of the rectifier.

rectifying element—A circuit element which conducts current in one direction only.

rectigon—A hot-cathode gas-filled diode that operates at a high pressure. Used most frequently in battery-charging circuits.

rectilineal compliance—A mechanical element that opposes a change in the applied force (e.g., the springiness that opposes a force on the diaphragm of a speaker or microphone).

rectilinear — In a straight line—specifically, moving, forming, or bounded by a straight line.

rectilinear scanning—The scanning of an area in a predetermined sequence of narrow, straight, parallel strips.

rectilinear writing recorder—An oscillograph that records in rectilinear coordinates.

recurrence rate—*See* Repetition Rate.

recursion — The continued repeating of the same operation or group of operations.

recursive—Capable of being repeated.

recyclability — The capability of a battery system to be recharged after it has been discharged.

red gun—In a three-gun color-television pic-

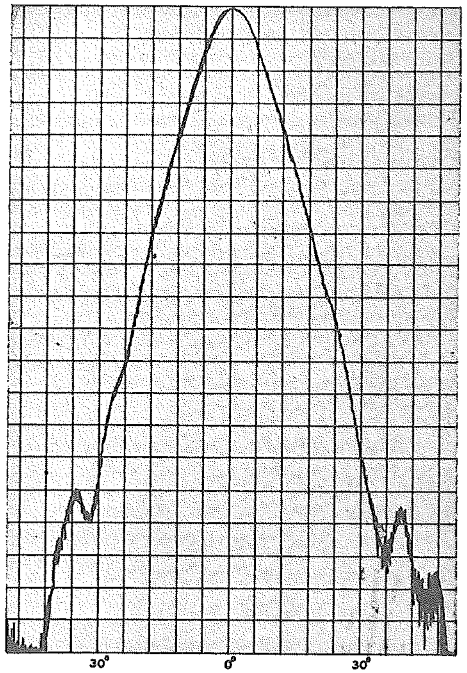

Rectilinear recording.

ture tube, the electron gun whose beam strikes only the phosphor dots that emit the red primary.

redistribution—In a charge-storage tube or television camera tube, the alteration of charges on one area of a storage surface by secondary electrons from any other area of the same surface.

redox cell—A cell designed to convert the energy of the reactants to electrical energy. An intermediate reductant, in the form of a liquid electrolyte, reacts at the anode in a conventional manner; it is then regenerated by reaction with a primary fuel.

red-tape operation — In a computer, operations which do not directly contribute to the results, i.e., those internal operations which are necessary to process data, but do not in themselves contribute to any final answer.

reduced coefficient of performance — The ratio of a given coefficient of performance to the corresponding coefficient of performance of a Carnot cycle.

reduced generator efficiency—The ratio of a given thermoelectric-generator efficiency to the corresponding Carnot efficiency.

redundancy—1. The employment of several devices, each performing the same function, in order to improve the reliability of a particular function. 2. Added or repeated information employed to reduce ambiguity or error in a transmission of information. As the signal-to-noise ratio decreases, redundancy may be employed to prevent an increase in transmission error. 3. As used in reliability, the concept of providing duplicate functional units in electronic equipment which allow continued performance if a failure occurs.

redundancy check—In a computer, the use of extra information for checking purposes only.

red video voltage—The signal voltage that controls the grid of the red gun in a three-gun picture tube. This signal is a reproduction of the output from the red camera at the transmitter.

reed—A thin bar located in a narrow gap and made to vibrate electrically, magnetically, or mechanically by forcing air through the gap.

reed relay—A relay in which two flat magnetic strips mounted inside a coil are attracted to each other when the coil is energized. The relay contacts are mounted on the strips.

re-entrancy—A type of feedback employed in microwave oscillators. In most magnetrons a circuit is used that can be described as a slow-wave structure that feeds back into itself. In beam re-entrancy the beam may be circulated repeatedly through the interaction space.

re-entrant cavity — A resonant cavity in which one or more sections are directed inward with the result that the electric field is confined to a small area or volume.

re-entrant winding—An armature winding that returns to its starting point, thus forming a closed circuit.

reference address—An address used in digital-computer programming as a reference for a group of relative addresses.

reference angle—The angle formed between the center line of a radar beam as it strikes a reflecting surface and the perpendicular drawn to that reflecting surface.

reference black level — The picture-signal level corresponding to a specified maximum limit for black peaks.

reference boresight—A direction defined by an optical, mechanical, or electrical axis of an antenna, established as a reference for the purpose of beam direction or tracking axis alignment.

reference burst—*See* Color-Burst.

reference dipole—A half-wave straight dipole tuned and matched for a given frequency and used as a unit of comparison in antenna measurement work.

reference electrode — In pH measurements, an electrode, usually hydrogen-filled, used to provide a reference potential. (*Also see* Glass Electrode.)

reference frequency—A frequency coinciding with, or having a fixed and specified relation to, the assigned frequency. It does not necessarily correspond to any frequency in an emission.

reference level—The starting point for designating the value of an alternating quantity or a change in it by means of decibel units. For sound loudness, the reference level is usually the threshold of hearing. For communications receivers, 60 microwatts is normally used. A common reference in electronics is one milliwatt, and power is stated as so many decibels above or below this figure.

reference line—A line from which angular measurements are made.

reference monitor—A receiver (or other similar device of known performance capabilities) used for judging the transmission quality.

reference noise—The magnitude of circuit noise that will produce a noise-meter reading equal to that produced by 10^{-12} watt of electric power at 1,000 hertz.

reference phase—The phase of the color burst transmitted with color-television carriers. It is used in synchronizing the receiver reference oscillator with the transmitted color signals.

reference point—A terminal that is common to both the input and the output circuits.

reference record—In digital-computer programming, a compiler output that lists the operations and their positions in the final specific routine, plus information describing the segmentation and storage allocation of the routine.

reference recording—A recording of a radio program for future reference or checking.

reference time—In a computer, an instant chosen near the beginning of switching as an origin for time measurements. It is taken as the first instant at which either the instantaneous value of the drive pulse, the voltage response of the magnetic cell, or the integrated voltage response reaches a specified fraction of its peak pulse amplitude.

reference tone—A stable tone of known frequency continuously recorder on one track of multitrack signal recordings and intermittently recorded on signal-track recordings by the collection-equipment operators for subsequent use by the data analysts as a frequency reference.

reference voltage—Alternating-current voltage in a synchro servosystem used to determine the in-phase or 180° out-of-phase condition to provide directional sense.

reference volume—The magnitude of complex electric wave, such as that corresponding to speech or music, which gives a reading of zero VU on a standard volume indicator. The sensitivity of the volume indicator is adjusted so that the reference volume or zero VU is read when the instrument is connected across a 600-ohm resistance to which there is delivered a power of 1 milliwatt at 1,000 hertz.

reference white—The light from a nonselective diffuse reflector as a result of the normal illumination of the scene to be televised.

reference white level — In television, the picture-signal level corresponding to a specified maximum limit for white peaks.

reflectance—*See* Reflection Factor.

reflected impedance—1. The apparent impedance across the primary of a transformer when current flows in the secondary. 2. The impedance at the input terminals of a transducer as a result of the impedance characteristics at the output terminals.

reflected power or signal—The power flowing back to the generator from the load.

reflected resistance—The apparent resistance across the primary of a transformer when a resistive load is across the secondary.

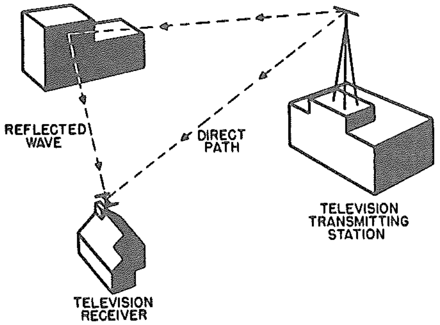

Reflected and direct wave.

reflected wave—The wave which has been reflected from a surface, a junction of two different media, or a discontinuity in the medium it is traveling in (e.g., the echo from a target in radar, the sky wave in radio, the wave traveling toward the source from the termination of a transmission line).

reflecting curtain—A vertical array of half-wave reflecting antennas placed one-quarter wavelength behind a radiating curtain of dipoles to form a pine-tree array.

reflecting electrode—A tubular outer electrode or the repeller plate in a microwave oscillator tube corresponding in construction but not function to the plate of an ordinary triode. It is capable of generating extremely high frequencies.

reflecting galvanometer — A galvanometer with a small mirror attached to the moving element. The mirror reflects a beam of light onto a scale.

reflecting grating—An arrangement of wires placed in a waveguide to reflect the desired wave while freely passing one or more other waves.

reflection — The phenomenon in which a wave that strikes a medium of different characteristics is returned to the original medium with the angles of incidence and reflection equal and lying in the same plane.

reflection altimeter — An aircraft altimeter that determines altitude by the reflection of sound, supersonic, or radio waves from the earth.

reflection coefficient—1. At the junction of a uniform transmission line and a mismatched terminating impedance, the vector ratios between the electric fields associated with the reflected and the incident waves. 2. At any specified plane in a uniform transmission medium, the vector ratios between the electric fields associated with the reflected and the incident waves. 3. At any specified plane in a uniform transmission line between a source and an absorber of power, the vector ratio between the electric fields associated with the reflected and the incident waves. It is given by the formula $(Z_2 - Z_1)/(Z_2 + Z_1)$, where Z_1 and Z_2 are the impedances of the source and load, respectively.

reflection color tube—A color picture tube that produces an image by electron reflection in the screen region.

reflection Doppler—A system utilizing the Doppler frequency shift to measure the position and/or velocity of an object not carrying a transponder.

reflection error—In navigation, the error due to the wave energy that reaches the receiver as a result of undesired reflections.

reflection factor—Also called mismatch factor, reflectance, reflectivity, or transition factor. The ratio of the current delivered to a load whose impedance is not matched to the source, to the current that would be delivered to a load of matched impedance. Expressed as a formula:

$$\frac{\sqrt{4Z_1Z_2}}{Z_1 + Z_2}$$

where Z_1 and Z_2 are the unmatched and the matched impedances, respectively.

reflection grating—A wire grating that is placed inside a waveguide so as to reflect a desired wave while at the same time allowing one or more other waves to pass freely.

reflection law—For any reflected object, the angle of incidence is equal to the angle of reflection.

reflection loss—1. That part of transition loss due to the reflection of power at the discontinuity. 2. The ratio (in decibels) of the power incident upon the discontinuity, to the difference between the powers incident upon and reflected from the discontinuity.

reflection sounding—Echo depth sounding in which the depth is measured by reflecting sound or supersonic waves off the bottom of the ocean.

reflective code—*See* Gray Code.

reflective jamming — *See* Radar-Confusion Reflectors.

reflective optics—A system of mirrors and lenses used in projection television.

reflectivity—*See* Reflection Factor.

reflectometer—A microwave system arranged to measure the incidental and reflected voltages and indicate their ratio.

reflector—Also called a reflector element. 1. One or more conductors or conducting surfaces for reflecting radiant energy—specifically, a parasitic antenna element located in other than the general direction of the major lobe of radiation. 2. (*Also see* Repeller.)

reflector satellite—A satellite so designed that radio or other waves bounce off its surface.

reflector element—*See* Reflector.

reflector voltage—The voltage between the reflector electrode and the cathode in a reflex klystron.

reflex baffle—A speaker baffle in which a portion of the radiation from the rear of the diaphragm is propagated forward after a controlled phase shift or other modification. This is done to increase the over-all radiation in some portion of the frequency spectrum.

reflex bunching — In a microwave tube, a type of bunching that is brought about when the velocity-modulated electron stream is made to reverse its direction by means of an opposing d-c field.

reflex circuit—A circuit through which the signal passes for amplification both before and after detection.

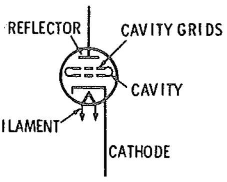

Reflex-klystron schematic.

reflex klystron—A klystron with a reflector (repeller) electrode in place of a second resonant cavity, to redirect the velocity-modulated electrons through the resonant cavity which produced the modulation. Such klystrons are well suited for use as oscillators, because the frequency is easily controlled by repositioning the reflector.

refracted wave—Also called the transmitted wave. In an incident wave, the portion that travels from one medium into a second medium.

refraction—The change in direction of propagation of a wavefront due to its passing obliquely from one medium into another in which its speed is different. Refraction may also occur in a single medium of varying characteristics.

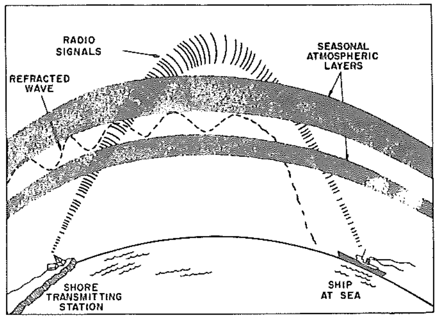

Refraction (of radio waves).

refraction error—In navigation, the error due to the bending of one or more wave paths by undesired refraction.

refraction loss—That part of the transmission loss due to refraction resulting from nonuniformity of the medium.

refractive index — Of a wave-transmission medium, the ratio between the phase velocity in free space and in the medium.

refractive modulus—Also called the excess modified index of refraction. The excess over unity of the modified index of refraction. It is expressed in millionths and is given by the equation:

$$M = (n + h/a - 1)\,10^6$$

where M is the refractive modulus, n is the index of refraction at a height h above sea level, and a is the radius of the earth.

refractivity—Ratio of phase velocity in free space to that in the medium, minus 1.

refractometer—An instrument for measuring the refractive index of a liquid or solid, usually from the critical angle at which total reflection occurs.

refrangible—Capable of being refracted.

regeneration—1. The gain in power obtained by coupling from a high-level point back to a lower-level point in an amplifier or in a system which encloses devices having a power-level gain. (*Also see* Positive Feedback *and* Regenerative Feedback.) 2. In a computer storage device whose information storing state may deteriorate, the process of restoring the device to its latest undeteriorated state. (*Also see* Rewrite.) 3. The replacement of a charge in a charge-storage tube to overcome decay effects, including a loss of charge by reading.

regeneration control—A variable capacitor, variable inductor, potentiometer, or rheostat used in a regenerative receiver to control the amount of feedback and thereby keep regeneration within useful limits.

regenerative amplification — Amplification where the increased gain and selectivity are given by a feedback arrangement similar to that in a regenerative detector. However, the operation is always kept just below the point of oscillation.

regenerative braking—Dynamic braking in which the momentum, as the equipment is being braked, causes the traction motors to act as generators. A retarding force is then exerted by the return energy to the power-supply system.

regenerative detector—A detector circuit in which regeneration is produced by positive feedback from the output to the input circuit. In this way, the amplification and sensitivity of the circuit are greatly increased.

regenerative divider—Also called a regenerative modulator. A frequency divider in which the output wave is produced by modulation, amplification, and selective feedback.

regenerative feedback—*See* Regeneration.

regenerative modulator — *See* Regenerative Divider.

regenerative receiver—A receiver in which controlled regeneration is used to increase the amplification provided by the detector stage.

regenerative repeater—A repeater that regenerates pulses to restore the original shape. Used in teletypewriter and other code circuits, each code element is replaced by a new code element with specified timing, waveform, and magnitude.

regional channel — A standard broadcast channel within which several stations may operate at five kilowatts or less. However, interference may limit the primary service area of such stations to a given field-intensity contour.

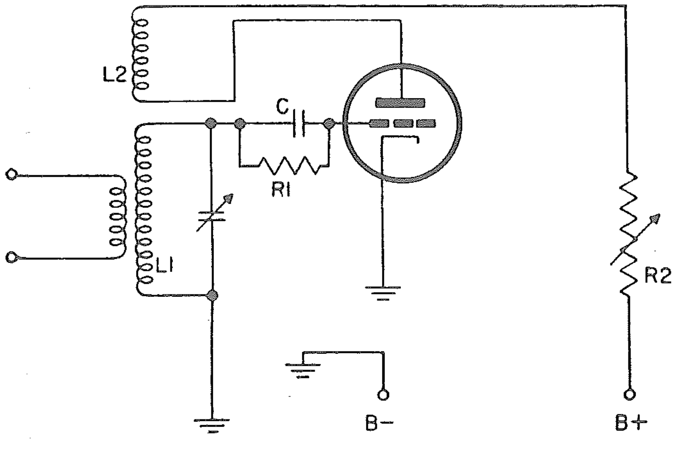

Regenerative detector.

regional interconnections—*See* Interconnection.

region of limited proportionality—The range of applied voltage, below the Geiger-Mueller threshold, where the gas amplification depends on the charge liberated by the initial ionizing event.

register—1. A digital-computer device capable of retaining information, often that contained in a small subset (e.g., one word) of the aggregate information. (*Also see* Storage.) 2. The relative position of all or part of the conductive pattern with respect to a mechanical feature of the board or to another pattern on the obverse side of the printed-circuit board (e.g., pattern-to-hole register or pattern-front-to-pattern-back register). 3. Also called registration. The accurate matching of two or more patterns such as the three images in color television. 4. A range of notes used for playing a particular piece or part of it (e.g., melody or harmony), particularly the range covered by a clavier or manual.

register constant—Symbolized by Kr. The factor by which the register reading must be multiplied in order to provide proper consideration of the register (gear) ratio and the instrument transformer ratios, to obtain the registration in the desired units.

register control—Any device that provides automatic register. In photoelectric register control, a light source and phototube form a scanning head. Whenever a special mark or a part of the design printed on a continuous web of paper arrives at the scanning head, the amount of light reaching the phototube changes. If necessary, the web is then moved slightly to bring it back into register.

register length—The number of digits, characters, or bits which a computer register can store.

register mark—In printed circuits, a mark used to establish the relative position of one or more printed-wiring patterns or portions of patterns with respect to their desired locations on the base.

register of a meter—In a meter, the part which registers the revolutions of the rotor, or the number of impulses received from or transmitted to the meter and gives the answer in units of electric energy or other quantity measured.

registration, conductive pattern-to-board-outline—The location of the printed pattern relative to the overall outline dimensions of the printed-circuit board.

registration, front-to-back—On a printed-circuit board, the location of the printed pattern on one side relative to the printed pattern on the opposite side.

registration of a meter—The apparent amount of electric energy (or other quantity being measured) that has passed through the meter, as shown by the register reading. It is equal to the register reading times the register constant. During a given period, it is equal to the register constant times the difference between the register readings at the beginning and end of the period.

registry—The superposition of one image onto another (e.g., in the formation of an interlaced scanning raster).

regulated power supply—A unit which maintains a constant output voltage or current for changes in line voltage, output load, ambient temperature, or time.

regulating device—A device that functions to regulate a quantity or quantities such as voltage, current, power, speed, frequency, temperature, and load, at a certain value or between certain limits for machines, tie lines, or other apparatus.

regulating transformer—A transformer for adjusting the voltage or the phase relation (or both) in steps, usually without interrupting the load. It comprises one or more windings excited from the system circuit or a separate source, and one or more windings connected in series with the system circuit.

regulating winding—A supplementary transformer winding connected in series with one of the main windings and used for changing the ratio of transformation or the phase relationship, or both, between circuits.

regulation—1. The difference between the maximum and minimum voltage drops within a specified anode-current range in a gas tube. 2. The holding constant of some condition (e.g., voltage, current, power, or position). 3. In a power supply, the ability to maintain a constant load voltage or current despite changes in line voltage or load impedance.

regulation of a constant-current transformer—The maximum departure of the secondary current from its rated value, expressed in percent of the rated secondary current, with the rated primary voltage and frequency applied and at the rated secondary power factor and with the current variation taken between the limits of a short circuit and rated load.

regulation of a constant-potential transformer—The change in secondary voltage, expressed in percent of rated secondary voltage, which occurs when the rated kva output at a specified power factor is reduced to zero, with the primary impressed terminal voltage maintained constant. In the case of a multiwinding transformer, the loads on all windings at specified power factors are to be reduced from the rated kva to zero simultaneously.

regulator—A device, the function of which is to maintain a designated characteristic at a predetermined value or to vary it according to a predetermined plan.

regulator tube—A two-electrode glow-discharge gas tube that has an essentially constant voltage drop. When series-connected with a resistance across a d-c source, the tube will maintain a constant d-c voltage across its terminal, with wide variations in the d-c source voltage.

reignition—The generation of multiple counts, within a radiation-counter tube, by the atoms or molecules excited or ionized in the discharge accompanying a tube count.

reignition voltage—Also called restriking voltage. That voltage which is just sufficient to re-establish conduction of a gas tube if applied during the deionization period. It varies inversely with time during the deionization period.

Reike diagram—A polar-coordinate load diagram for microwave oscillators, particularly for klystrons and magnetrons.

Reinartz crystal oscillator—A crystal-controlled vacuum-tube oscillator in which the crystal current is kept low by placing, in the cathode lead, a resonant circuit tuned to half the crystal frequency. The resultant regeneration at the crystal frequency improves the efficiency, but without the problem of uncontrollable oscillations at other frequencies.

reinsertion of a carrier—Combining a locally generated carrier signal in a receiver with an incoming suppressed-carrier signal.

rejection band—The frequency range below the cutoff frequency of a uniconductor waveguide.

rejector circuit—A circuit that suppresses or eliminates signals of the frequency to which it is tuned.

rel—A unit of reluctance, equal to one ampere-turn per magnetic line of force.

relative accuracy—The possible deviation among the standards in a group.

relative address—1. A designation used to identify the position of a memory location in a computer routine or subroutine. 2. A label used to identify a word in a routine or subroutine with respect to its position in that routine or subroutine.

relative bearing—The bearing in which the direction of the reference line is the heading of the vehicle.

relative coding—In a computer, coding in which all addresses refer to an arbitrarily selected position, or in which all addresses are represented symbolically.

relative damping of an instrument—Also called specific damping. Ratio of the actual damping torque at a given angular velocity of a moving element, to the damping torque which would produce critical damping at this same angular velocity.

relative dielectric constant—Ratio of the dielectric constant of a material to that of a vacuum. The latter is arbitrarily given a value of 1.

relative gain of an antenna—The gain of an antenna in a given direction when the reference antenna is a half-wave loss-free dipole isolated in space and the equatorial plane of which contains the given direction.

relative humidity—1. Ratio of the quantity of water vapor in the atmosphere, to the quantity which would saturate at the existing temperature. 2. Ratio of the pressure of water vapor to that of saturated water vapor at the same temperature.

relative luminosity—Ratio of the actual luminosity at a particular wavelength, to the maximum luminosity at the same wavelength.

relative Peltier coefficient—The Peltier coefficient of a couple made up of the given material as the first-named conductor and a specified standard conductor, commonly platinum, lead, or copper.

relative permeability—Ratio of the magnetic permeability of one material to that of another, or of the same material under different conditions.

relative plateau slope—The average percentage change in the counting rate of a radiation-counter tube, near the midpoint of the plateau, per increment of applied voltage. It is usually expressed as the percentage change in counting rate per 100-volt change in applied voltage.

relative power gain (of one transmitting or receiving antenna over another)—The measured ratio of the signal power one antenna produces at the receiver input terminals, to the signal power produced by the other, the transmitting power level remaining fixed.

relative refractive index—Ratio of the refractive indices of two media.

relative response—The ratio, usually expressed in decibels, of the response under some particular conditions to the response under reference conditions (which should be stated explicitly).

relative Seebeck coefficient—The Seebeck coefficient of a couple made up of the given material as the first-named conductor and a specified standard conductor such as platinum, lead, or copper.

relative velocity (of a point with respect to a reference frame)—The rate at which a position vector of that point changes with respect to the reference frame.

relaxation—An action requiring an observable length of time for initiation in response to a sudden change in conditions.

relaxation circuit—A circuit arrangement, usually of vacuum tubes, reactances, and resistances, which has two states or conditions, one, both, or neither of which may be stable. The transient voltage produced by passing from one to the other, or the voltage in the state of rest, can be used in other circuits.

relaxation inverter—An inverter that uses a relaxation-oscillator circuit to convert d-c power to a-c power.

relaxation oscillator—An oscillator which generates a nonsinusoidal wave by gradually charging and quickly discharging a capacitor or an inductor through a resistor. The frequency of a relaxation oscillator may be

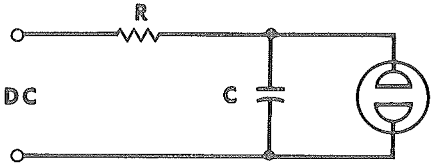

Relaxation oscillator.

self-determined, or determined by a synchronizing voltage derived from an external source.

relaxation time—1. The time an exponentially decaying quantity takes to decrease in amplitude by a factor of 0.3679. 2. The average time between collisions of an electron with the lattice.

relay—An electromechanical device in which contacts are opened and/or closed by variations in the conditions of one electric circuit and thereby affect the operation of other devices in the same or other electric circuits.

relay bias—Bias produced by a spring on an electromagnet. By acting on the relay armature, the spring tends to hold it in a given position.

relay broadcast station—A station licensed to retransmit, from points where wire facilities are not available, the programs from one or more broadcast stations.

relay contacts—Contacts that are closed or opened by the movement of a relay armature.

relay drop—A relay activated by an incoming ringing current, to call an operator's attention to a telephone subscriber's line.

relay flutter—Erratic rather than positive operation and release of a relay.

relay magnet—A coil and iron core forming an electromagnet which, when energized, attracts the armature of a relay and thereby opens or closes the relay contacts.

relay receiver—A specific assembly of apparatus which accepts a sound or television relay signal at its input terminals, and delivers the amplified signal at its output terminals.

relay selector—A relay circuit associated with a selector, consisting of a magnetic impulse counter, for registering digits and holding a circuit.

relay station—*See* Relay Transmitter.

relay system—Dial switching equipment made up principally of relays instead of mechanical switches.

relay transmitter—Also called a repeater or relay station (but only if the signal is reduced to a composite picture signal at a standard impedance level, and polarity between the receiver and transmitter). The specific assembly of apparatus which accepts a sound or television-relay input signal from the relay receiver, and rebroadcasts it to another station outside the range of the operating station.

release—1. An electromagnetic device that opens a circuit breaker automatically or allows a motor starter to return to its off position when tripped by hand, by an interruption of power-supply operation, or by an excessive current. 2. The condition attained by a relay when it has been deenergized, all contacts have functioned, and the armature (if applicable) has attained a fully opened position.

release current—The maximum current needed to fully release a relay, after it has been fully closed.

release force—The value to which the force on the actuator of a momentary-contact switch must be reduced to allow the contacts to snap from the operated contact position to the normal contact position.

release time—The interval between the time power is removed from a relay coil and the contacts being checked are released.

releasing position—That position of the actuator of a momentary-contact switch at which the contacts snap from the operated contact position to the normal contact position.

reliability—The probability that a device will perform adequately for the length of time intended and under the operating environment encountered.

reliability control—The scientific coordination and direction of technical reliability activities from a system viewpoint.

reliability data—1. Data related to the frequency of failure of an item, equipment, or system. These data may be expressed in terms of "failure rate," "mean time between failure" (mtbf), or "probability of success." 2. Data contained in comprehensive documents that provide a detailed history of the reliability evaluation of component parts, component assemblies, etc., or the entire program during the design, development, production, and major product improvement phases of an equipment, weapon, or weapon system in which engineering studies have been performed to select the most reliable product for the intended application.

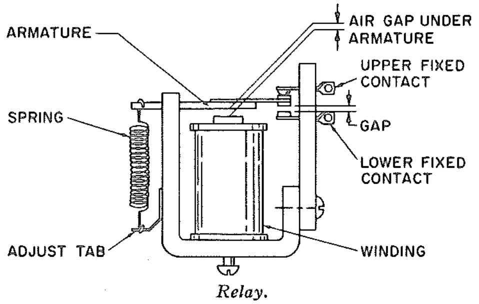

Relay.

reliability engineering—The establishment, during design, of an inherently high reliability in a product.

reliability test—Tests and analyses carried out in addition to other type tests and designed to evaluate the level of reliability in a product, etc. as well as the dependability, or stability, of this level relative to time and use under various environmental conditions.

relieving anode—In a pool-cathode tube, an auxiliary anode which provides an alternative conducting path to reduce the current to another electrode.

reluctance—The resistance of a magnetic path to the flow of magnetic lines of force through it. It is the reciprocal of permeance and is equal to the magnetomotive force divided by the magnetic flux.

reluctance motor—A synchronous motor similar in construction to an induction motor. The member carrying the secondary circuit has salient poles, but no d-c excitation. It starts as an induction motor, but operates normally at synchronous speed.

reluctive transduction—The conversion of the measurand into a change in a-c voltage by changing the reluctance path between two or more coils when a-c excitation is applied.

reluctivity—The ability of a magnetic material to conduct magnetic flux. It is the reciprocal of permeability.

REM—*See* Roentgen Equivalent Man.

remanence—The extent to which a body remains magnetized after removal of a magnetizing field that has brought the body to its saturation (maximum) magnetization. A substance with remanence is known as ferromagnetic.

remanent magnetization—The magnetization retained by a substance after the magnetizing force has been removed.

remodulator—Formerly called SCFM. A device for converting amplitude modulation to audio-frequency-shift modulation for transmission over a voice radio-frequency channel.

remote—*See* Field Pickup.

remote control—Any system of control performed from a distance. The control signal may be conveyed by intervening wires, sound (ultrasonics), light, or radio.

remote-control equipment—*See* Remote Control.

remote-cutoff tube—Also called a variable-mu or extended-cutoff tube. An electron tube used mainly in r-f amplifiers. The control-grid wires are farther apart at the center than at the ends. Therefore, the amplification of the tube does not vary in direct proportion to the bias. Also, some plate current will flow, regardless of the negative bias on the grid.

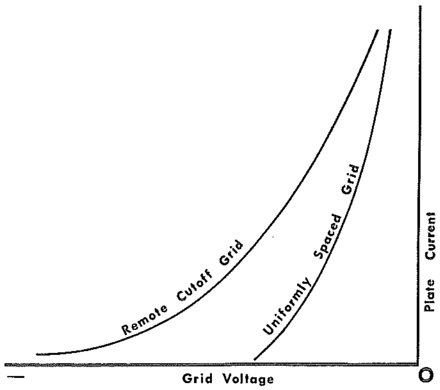

Remote-cutoff tube characteristic.

remote error sensing—A means by which the regulator circuit of a related power supply senses the voltage directly at the load. This connection is used to compensate for a voltage drop in the connecting wires.

remote indicator—1. A radar indicator which is connected in parallel with a primary indicator. 2. An indicator located some distance away from the data-gathering element.

remote line—A program transmission line between a remote-pickup point and the studio or transmitter site.

remotely adjustable timer—A time delay having external leads to which a variable or fixed resistor can be attached.

remote metering—*See* Telemetering.

remote pickup—A program which originates away from the studio and is transmitted to the studio or transmitter over telephone lines or a radio link.

remote programming—Power-supply operation which permits control of the regulated output current or voltage by means of a remotely varied resistance or voltage.

remote sensing—The process by which a power supply senses voltage directly at the load. This method eliminated the variations in load lead drop from affecting the load regulation.

renewable fuse—A fuse which may be readily restored to operation by replacing the fused link.

reoperate time—The release time of a thermal relay.

repeatability—1. The ability of a device or an instrument to come back to the same reading after a certain length of time. 2. The ability of an instrument to repeat its readings taken when deflecting the pointer upscale, compared to the readings taken when deflecting the pointer downscale, expressed as a percentage of the full-scale deflection.

repeatability error—The inability to reproduce the same output readings when a given level of measurand is applied repeatedly in the same direction.

repeat-cycle timer—A timer which has any number of load contacts and which continues to repeat its time program as long as power is applied.

repeater—Also called a one-way or two-way repeater. A combination of apparatus for receiving signals from either one or both directions and delivering corresponding signals which have been amplified or reshaped, or both.

repeater jammer—A jamming transmitter used to confuse or deceive the enemy by causing his equipment to present erroneous azimuth, range, or number of targets. This is accomplished by intercepting and reradiating a modified signal on his frequency.

repeater station—*See* Relay Transmitter.

repeating coil—An audio-frequency transformer, usually with a 1:1 ratio, for connecting two sections of telephone line inductively, to permit the formation of simplex and phantom circuits.

repeating flash tube—A flash tube which, by producing rapid, brilliant flashes, permits night aerial photographs to be taken from as high as two miles.

repeating timer—A timer which repeats each operating cycle automatically until excitation is removed.

repeller—Sometimes called a reflector. An electrode, the primary function of which is to reverse the direction of an electron stream.

reperforator—1. A device that converts teletypewriter signals into perforations on tape instead of the usual typed copy on a roll of paper or ticker tape. 2. A machine that reads one punched paper tape or card and punches the same information into another paper tape or card.

repetition frequency—*See* Repetition Rate.

repetition rate—Also called recurrence rate or repetition frequency. The rate at which signals recur or are transmitted.

repetitive peak-inverse voltage—The maximum, allowable instantaneous value of reverse (negative) voltage that may be repeatedly applied to the anode of an scr with the gate open. This value of peak-inverse voltage does not represent a breakdown voltage, but it should never be exceeded (except by the transient rating if the device has such a rating).

reply—In transponder operation, the radio-frequency signal or signals transmitted as a result of an interrogation.

report generation—In a computer, production of complete output reports from only a specification of the desired content and arrangement and from specifications regarding the input file.

report generator — A special computer routine designed to prepare an object routine that, when later run on the computer, produces the desired report.

reproduce—In a computer, to prepare a duplicate of stored information.

reproduce head—An electromagnetic transducer which converts the remnant flux pattern in a magnetic tape into electrical signals during the reproduce (playback) process.

reproducer — A device used to translate electrical signals into sound waves.

reproducibility—The exactness with which measurement of a given value can be duplicated.

reproducing stylus—A mechanical element that follows the modulations of a record groove and transmits the mechanical motion thus derived to the pickup mechanism.

reproduction speed—The area of copy recorded per unit time.

repulsion—A mechanical force tending to separate bodies having a like electric charge or magnetic polarity—or in the case of adjacent conductors, currents flowing in opposite directions.

repulsion-induction motor—A constant or variable-speed repulsion motor with a squirrel-cage winding in the rotor, in addition to the regular winding.

repulsion motor—A single-phase motor in which the stator winding is connected to the source of power and the rotor winding to the commutator. Brushes on the commutator are short-circuited and are placed so that the magnetic axis of the rotor winding is inclined to that of the stator winding. This type of motor has a varying speed characteristic.

repulsion-start, induction-run motor — A single-phase motor which has the same windings as a repulsion motor but operates at a constant speed. The rotor winding is short-circuited (or otherwise connected) to give the equivalent of a squirrel-cage winding. It starts as a repulsion motor, but operates as an induction motor with constant-speed characteristics.

reradiation—1. Scattering of incident radiation. 2. Radiation of the signals amplified in a radio receiver.

rerecording—The process of making a recording by reproducing a recorded sound source and recording this reproduction. (*Also see* Dubbing.)

rerecording system—An association of reproducers, mixers, amplifiers, and recorders capable of being used for combining or modifying various sound recordings to provide a final sound record. Recording of speech, music, and sound effects may be so combined.

rerun—Also called rollback. To run a computer program (or a portion of it) over again.

rerun point—In a computer program, one of a set of preselected points located in a computer program such that if an error is detected between two such points, the problem may be rerun by returning to the last such point instead of returning to the start of the problem.

rerun routine—A computer routine designed to be used, in the event of a malfunction or mistake, to reconstitute a routine from the previous rerun point.

reset—1. To restore a storage device to a prescribed state. 2. To place a binary cell in the initial, or zero, state. (*Also see* Clear.)

reset pulse—A drive pulse which tends to reset a magnetic cell.

reset terminal—*See* Zero-Input Terminal.

reset time—The period required from the time of the reset command until a timer is fully returned to the before-start conditions ready for the next cycle.

reset timer—A timer which can be reset by electrical means. May be either an "on delay" or "off delay" type. 2. A timer with one or more circuits which spring-reset to zero when the clutch is disengaged.

residual—1. The difference between any value and some estimate of the means of such values; e.g., residuals from a curve of regression. (Residual unbalance.) 2. The complement of a set is called a residual set. 3. (adjective) Pertaining to a measure of the output of a transducer under static conditions and with no stimulus applied.

residual charge—The charge remaining on the plates after an initial discharge of the capacitor.

residual current—The vector sum of the currents in the several wires of an electric supply circuit.

residual discharge—A discharge of the residual charge of a capacitor remaining after the initial discharge.

residual deviation — Apparent modulation due to noise and/or distortion in the transmitter.

residual error—The direction-finding errors remaining after errors due to site and antenna effects have been minimized.

residual field—The magnetic field left in an iron-field structure after excitation has been removed.

residual flux—The value of magnetic induction that remains in a magnetic circuit when the magnetomotive force is reduced to zero.

residual f-m—*See* Incidental F-M.

residual frequency modulation — In a klystron, frequency modulation of the fundamental frequency due to shot and ion noises, a-c heater voltage, etc.

residual gases—The small amounts of gases remaining in a vacuum tube despite the best possible exhaustion by vacuum pumps.

residual induction—The magnetic induction which remains in a magnetized material when the effective magnetizing force has been reduced to zero. When the material is a symmetrically cyclic magnetic condition, the residual induction is termed the "normal residual induction." Abbreviated B_r.

residual ionization—Ionization of air or other gas not accounted for in a closed chamber by recognizable neighboring agencies.

residual losses—In a magnetic core, the difference between the total losses and the sum of the eddy-current and hysteresis losses.

residual magnetic induction—Magnetic induction remaining in a ferromagnetic object after the magnetizing force has been removed. The amount depends on the material, shape, and previous magnetic history.

residual magnetism—The magnetism which remains in the core of an electromagnet after the operating circuit has been opened.

residual modulation — *See* Carrier Noise Level.

residual screw—A brass screw in the center of a relay armature. It is used to adjust the residual air gap between the armature and the coil core, to prevent residual magnetism from holding the armature operated after the relay operating circuit has opened.

residual voltage—The vector sum of the voltages to ground of the several phase wires in an electric supply circuit.

resist—In the preparation of a printed-circuit board, a material, such as ink, paint, metallic plating, etc., used to protect the desired portions of the conductive material from the action of the etchant, solder, or plating.

resistance—1. A property of conductors which—depending on their dimensions, material, and temperature—determines the current produced by a given difference of potential. The practical unit of resistance is the ohm. It is defined as the resistance through which a difference of potential of one volt will produce a current of one ampere. 2. Opposition which a device or material offers to current. 3. A circuit element designed to offer a predetermined resistance to current. 4. That property of a substance which impedes the current and results in the dissipation of the power in the form of heat.

resistance balance—The amount of resistance which is required to null the output of certain transducers or input systems.

resistance box—An assembly of resistors and the necessary switching or other means for changing the resistance connected across its output terminals by known, fixed amounts. (*Also see* Decade Box.)

resistance brazing—Brazing by resistance heating, the joint being part of the electrical circuit.

resistance bridge — A common form of Wheatstone bridge employing resistances in three arms. *Also see* Wheatstone Bridge.

resistance-bridge pressure pickup—A pressure transducer in which the electrical output is derived from the unbalance of a resistance bridge, which is varied according to the applied pressure.

resistance-capacitance-coupled amplifier—An amplifier, the stages of which are connected by a suitable arrangement of resistors and capacitors.

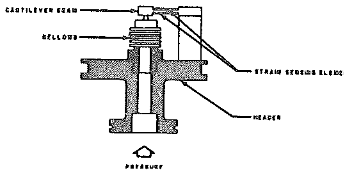

Resistance-bridge pressure pickup.

resistance-capacitance coupling—Also called RC coupling. Coupling between two or more circuits, usually amplifier stages, by a combination of resistive and capacitive elements.

resistance-capacitance filter—A filter made up only of resistive and capacitive elements. Abbreviated RC filter.

resistance-capacitance oscillator—An oscillator the output frequency of which is determined by resistance and capacitance elements. Abbreviated RC oscillator.

resistance coupling — Also called resistive coupling. The association of circuits with one another by means of the mutual resistance between circuits.

resistance drop—The voltage drop occurring across two points on a conductor when current flows through the resistance between those points. Multiplying the resistance in ohms by the current in amperes gives the voltage drop in volts.

resistance furnace—An electric furnace in which the heat is developed by the passage of current through a suitable resistor, which may be the charge itself or a resistor imbedded in or surrounding the charge.

resistance grounded—Grounded through a resistance.

resistance lamp—An electric lamp used as a resistance to limit the amount of current in a circuit.

resistance loss—The power lost when current flows through a resistance. Its value in watts is equal to the resistance in ohms multiplied by the square of the current in amperes ($W = R \times I^2$).

resistance magnetometer—A magnetometer which depends for its operation on the variation in the electrical resistance of a material immersed in the field to be measured.

resistance material—A material having sufficiently high resistance per unit length or volume to permit its use in the construction of resistors.

resistance noise—*See* Thermal Noise.

resistance pad—A network employing only resistances. It is used to provide a fixed amount of attenuation without altering the frequency response.

resistance ratio — In a thermistor, the ratio of the resistances measured at two specified reference temperatures with zero power in the thermistor.

resistance standard—*See* Standard Resistor.

resistance-start motor—A form of split-phase motor having a resistance connected in series with the auxiliary winding. The auxiliary circuit opens whenever the motor attains a predetermined speed.

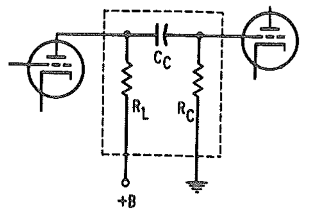

Resistance-capacitance coupling.

resistance strain gage—A strain gage consisting of a small strip of resistance material cemented to the part under test. Its resistance changes when the strip is compressed or stretched.

resistance temperature detector — Also called resistance-thermometer detector or resistance-thermometer resistor. A resistor made of some material for which the electrical resistivity is a known function of the temperature. It is intended for use with a resistance thermometer and is usually in such a form that it can be placed in the region where the temperature is to be determined.

resistance temperature meter—*See* Resistance Thermometer.

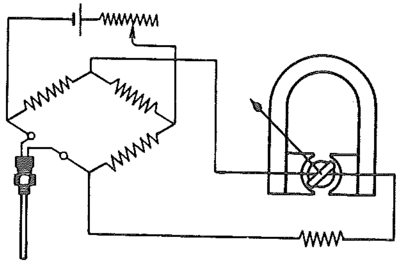

Resistance thermometer.

resistance thermometer—Also called resistance temperature meter. An electric thermometer which has a temperature-responsive element called a resistance temperature detector. Since the resistance is a known function of the temperature, the latter can be readily determined by measuring the electrical resistance of the resistor.

resistance-thermometer detector—*See* Resistance Temperature Detector.

resistance-thermometer resistor — *See* Resistance Temperature Detector.

resistance thermometry — A temperature-measuring technique that utilizes the temperature coefficient of a wirewound resistor. Known as a resistance thermometer, this resistor consists of a spiral of nickel or platinum wire. Since the ohmic value of the wire varies with temperature, the resistance of the spiral is thus an indication of temperature.

resistance weld—The junction produced by heat obtained from the resistance of the work to the flow of electric current in a circuit of which the work is a part, and by the application of pressure before and during the flow of current. The term includes all types of bonds produced by the process, which may or may not be classified metallurgically as welds.

resistance wire—A wire made from a metal or alloy having a high resistance per unit length (e.g., nichrome). It is used in wirewound resistors, heating elements, and other high-resistance circuits.

resist-etchant—Any material deposited onto a copper-clad base material to prevent the conductive area underneath from being etched away.

resistive conductor—A conductor used primarily because of its high electrical resistance.

resistive coupling—*See* Resistance Coupling.

resistive load—A load in which the voltage is in phase with the current.

resistive transduction—The conversion of the measurand into a change in resistance.

resistive unbalance—Unequal resistance in the two wires of a transmission line.

resistivity—1. The resistance of a semiconductor. It is the reciprocal of conductivity. 2. The resistance of a sample of material having specified dimensions. (*Also see* Specific Resistance.)

resistor—1. A device which resists the flow of electric current in a circuit. There are two types—fixed and variable. 2. A device connected into an electrical circuit to introduce a specified resistance.

resistor capacitor transistor logic—Abbreviated RCTL. A logic circuit design that employs a resistor and a speedup capacitor in parallel for each input of the gate. A transistor's base is connected to one end of the RC network. A positive voltage on the RC input will energize the transistor and turn it on, so that the output voltage is nearly zero volts. This circuit is a positive NOR or negative NAND when NPN transistors are used in the circuit.

resistor color code—A code adopted by the Electronic Industries Association to mark the values of resistance on resistors in a readily recognizable manner. The first color represents the first significant figure of the resistor value, the second color the second significant figure, and the third color represents the number of zeros following the first two figures. A fourth color is sometimes added to indicate the tolerance of the resistor.

resistor core—An insulating support around which a resistor element is wound or otherwise placed.

resistor element—1. That portion of a resistor which possesses the property of electric resistance. 2. That portion of a potentiometer which provides the change in resistance as the shaft is rotated.

resistor housing—The enclosure around the resistance element and the core of a resistor.

resistor transistor logic — Abbreviated RTL. A form of logic that has a resistor as the input component that is coupled to the base of an NPN transistor. As in RCTL the transistor is an inverting element that produces the positive NOR gate or the negative NAND gate junction.

resist plating—Any material which, when deposited on a conductive area, prevents the areas underneath from being plated.

resnatron—Acronym for RESoNAtor-TRON. A high-power, high-efficiency, cavity-resonator tetrode designed to operate in the vhf region.

resolution—1. The deriving of a series of discrete elements from a sound, scene, or other form of intelligence so that the original may subsequently be synthesized. 2. The degree to which nearly equal values of a quantity can be discriminated. 3. The fineness of detail in a reproduced spatial pattern. 4. The degree to which a system or a device distinguishes fineness of detail in a spatial pattern. 5. In facsimile, a measure of the narrowest line width that may be transmitted and reproduced. 6. A measure of the smallest possible increment of change in the variable output of a device. 7. The minimum output voltage (or current) change of a regulated power supply that can be made using the front-panel controls of a power supply. The resolution of "continuously variable" controls shall be defined as 2° of shaft rotation. 8. The minimum output voltage (or current) change that can be made using the front-panel controls of a power supply. The resolution of "continuously variable" controls is defined as 1% of shaft rotation. 9. In a potentiometer, the smallest possible incremental resistance change. 10. The reciprocal number of steps per revolution of a motor shaft, expressed in degrees per step.

resolution chart—A pattern of black and white lines used to determine the resolution capabilities of equipment.

resolution noise—The noise due to the stepped character of the resistance element in wirewound potentiometric transducers.

resolution wedge—A narrow-angled, wedge-shaped pattern calibrated for the measurement of resolution. It is composed of alternate contrasting strips which gradually converge and taper individually to preserve equal widths along a line drawn perpendicular to the axis of the wedge.

resolver—1. A means for resolving a vector into two mutually perpendicular components. 2. A transformer, the coupling between primary and secondary of which can be varied. 3. A small section with a faster access than the remainder of the magnetic-drum memory in a computer. 4. A device which separates or breaks up a quantity into constituent parts or elements. 5. An electromechanical transducing device which develops an output voltage proportional to the product of an input voltage and the sine of the shaft angle.

resolving power—The reciprocal of the beam width in a unidirectional antenna, measured in degrees. It may differ from the resolution of a directional-radio system, since the latter is affected by other factors as well.

resolving time—The minimum time interval by which two events must be separated to be distinguishable.

resonance—1. A circuit condition whereby the inductive- and capacitive-reactance (or -impedance) components of a circuit have been balanced. In usual circuits, resonance can be obtained for only a comparatively narrow frequency band or range. 2. In a mechanical system, the frequency at which the maximum displacement occurs.

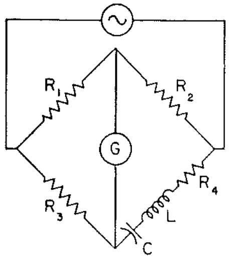

$$R_1R_4 = R_2R_3$$

$$\omega^2LC = 1$$

Resonance bridge (series).

resonance bridge—A four-arm, alternating-current bridge normally used for measuring inductance, capacitance, or frequency. An inductor and a capacitor are both present in one arm, the other three arms being (usually) nonreactive resistors. The adjustment for balance includes the etablishment of resonance for the applied frequency. Two general types—series or parallel—can be distinguished, depending on how the inductor and capacitor are connected.

resonance characteristic—*See* Resonance Curve.

resonance curve—Also called a resonant curve or resonance characteristic. A graphical representation of how a tuned circuit responds to the various frequencies in and near the resonant frequency.

resonance indicator—A meter, neon lamp, headphone, etc., that indicates when a circuit is at resonance.

resonance radiation—Radiation from a gas or vapor due to excitation and having the same frequency as the exciting source (e.g., sodium vapor irradiated with sodium light).

resonant capacitor—A tubular capacitor that is purposely wound so as to have inductance in series with its capacitance to resonate at a predetermined (i-f) frequency. Used as a bypass capacitor in i-f amplifiers for more effective bypassing.

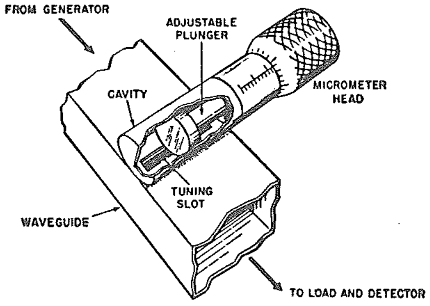

Resonant cavity.

resonant cavity—A form of resonant circuit in which the current is distributed on the inner surface of an enclosed chamber. By making the chamber of the proper dimensions, it is possible to give the circuit a high *Q* at microwave frequencies. The resonant frequency can be changed by adjusting screws which protrude into the cavity, or by changing the shape of the cavity.

resonant charging choke—A modulator inductor which sets up an oscillation of a given charging frequency with the effective capacitance of a pulse-forming network in order to charge a line to a high voltage.

resonant circuit—A circuit which contains both inductance and capacitance and is therefore tuned to resonance at a certain frequency. The resonant frequency can be raised or lowered by changing the inductance and/or capacitance values.

resonant-circuit–type frequency indicator—A frequency-indicating device which depends for its operation on the frequency-versus-reactance characteristics of two series-resonant circuits. The circuit is arranged so that the deflecting torque is independent of the amplitude of the signal to be measured.

resonant-current step-up—The ability of a parallel-resonant circuit to circulate a much higher current through its inductor and capacitor than the current fed into the circuit.

resonant curve—*See* Resonance Curve.

resonant diaphragm—In waveguide technique, a diaphragm so proportioned that it does not introduce reactive impedance at the design frequency.

resonant frequency—1. The frequency at which a given system or object will respond with maximum amplitude when driven by an external sinusoidal force of constant amplitude. For an LC circuit, the resonant frequency is determined by the formula:

$$f = \frac{1}{2\pi\sqrt{LC}}$$

where, f is in hertz, L is in henrys, and C is in farads. 2. The frequency of a crystal unit for a particular mode of vibration to which, discounting dissipation, the effective impedance of the crystal unit is zero.

resonant frequency of a crystal unit—For a particular mode of vibration, the frequency to which (discounting dissipation) the effective impedance of the crystal unit is zero.

resonant gap—The small region where the electric field is concentrated in the resonant structure inside a TR tube.

resonant gate transistor—A surface field-effect transistor incorporating a cantilevered beam which resonates at a specific frequency to provide high-Q frequency discrimination.

resonant line—A transmission line in which the distributed inductance and capacitance are such that the line is resonant at the frequency it is handling.

resonant-line oscillator—An oscillator in which one or more sections of transmission line are employed as tanks.

resonant mode—In the response of a linear device, a component characterized by a certain field pattern and, when not coupled to other modes, representable as a single-tuned circuit. When modes are coupled together, the combined behavior is similar to that of single-tuned circuits that have been correspondingly coupled.

resonant-reed relay—A relay with multiple contacts, each actuated by an a-c voltage of the frequency at which the reeds resonate.

resonant resistance—The resistance value to which a resonant circuit is equivalent.

resonant voltage step-up—The ability of an inductor and a capacitor in a series-resonant circuit to deliver a voltage several times greater than the input voltage.

resonant window (switching tubes)—A resonant iris sealed with a suitable dielectric material and constituting a portion of the vacuum envelope of the tube.

resonate—To bring to resonance—i.e., to maximize or minimize the amplitude or other characteristic of a steady-state quantity.

resonating cavity—A waveguide that is adjustable in length and terminates in a metal piston, diaphragm, or other wave-reflecting device at either or both ends. It is used as a filter, a means of coupling between guides of different sizes, or an impedance network.

resonator—An apparatus for setting up and maintaining oscillations of a frequency determined by the physical constants of the system.

resonator cavity—A section of coaxial line or waveguide completely enclosed by conductive walls.

resonator grid—One of the grids attached to a cavity resonator in a velocity-modulation tube.

resonator mode—A condition of operation corresponding to a particular field configuration for which the electron stream introduces negative conductance into the coupled circuit.

resonator wavemeter—A resonant circuit for determining wavelength (e.g., a cavity-resonator wavemeter).

resonistor—An experimental resonating device that is essentially a cantilevered chip of silicon mounted, like a tiny diving board, on a substrate. It measures 0.0350 inch long, 0.090 inch wide, and 0.008 inch thick, and contains two special diffused resistors. When an input signal is applied, it causes the silicon chip to vibrate, and the resonistor then delivers an output voltage with a specific stable frequency.

responder—The part of a transponder that automatically transmits a reply to the interrogator-responsor. By contrast, the responsor is the receiver which accepts and interprets the signals from the transponder.

response—A quantitative expression of the output of a device or system as a function of the input, under conditions which must be explicitly stated. The response characteristic, often presented graphically, gives the response as a function of some independent variable such as frequency or direction.

response curve—1. A plot of output versus frequency for a specific device. 2. A plot of stimulus versus output.

responser—*See* Responsor.

response speed—The time for a control action to start after a temperature change has occurred at the sensor.

response time—1. The time (usually expressed in cycles of the power frequency) required for the output voltage of a magnetic amplifier to reach 63% of its final average value in response to a step-function change of signal voltage. 2. The time required for the pointer of an instrument to come to apparent rest in its new position after the measured quantity abruptly changes to a new, constant value. 3. *See* Recovery Time (6). 4. The time required to reach a specified percentage (e.g., 90, 98, 99%) of the final output value). 5. *See* Transient Recovery Time.

responsor—1. Also spelled responser. The receiver used to receive and interpret the signals from a transponder. 2. An electronic device used to receive an electronic challenge and to display a reply thereto.

resting frequency—*See* Center Frequency.

restore—To return a variable address or other computer word to its initial value.

restorer—*See* D-C Restorer.

restoring spring—Also called return spring. A spring that moves the armature to the normal position and holds it there when the relay is de-energized.

rest potential—The residual potential difference remaining between an electrode and an electrolyte after the electrode has become polarized.

restriking voltage—*See* Reignition Voltage.

resultant—The effect produced by two or more forces or vectors.

retained image—Also called image burn. A change that is produced on the target of a television camera tube by a stationary light image and that results in the production of a spurious electrical signal corresponding to the light image for a large number of frames after the image is removed.

retardation coil—A high inductance coil used in telephone circuits to permit passage of d-c or low-frequency current while blocking audio-frequency currents. *Also see* Inductor.

retarding-field oscillator—Also called a positive-grid oscillator. An oscillator tube in which the electrons move back and forth through a grid which is more positive than the cathode and plate. The frequency depends on the electron-transit time and sometimes on the associated circuit parameters. The field around the grid retards the electrons and draws them back as they pass through it in either direction. Barkhausen-Kurz and Gill-Morell oscillators are examples of a retarding-field oscillator.

retarding magnet—Also called a braking magnet or drag magnet. A magnet used for limiting the speed of the rotor in a motor-type meter.

retard transmitter—A transmitter in which a delay is introduced between the time it is actuated and the time transmission begins.

retentivity—That property of a material measured by the normal residual induction remaining after the removal of an applied magnetizing force corresponding to the saturation induction for the material.

retentivity of vision—The image retained momentarily by the mind after the view has left the field of vision. (*Also see* Persistence of Vision.)

RETMA—Abbreviation for Radio-Electronics-Television Manufacturers Association, now the Electronic Industries Association (EIA).

retrace—*See* Flyback.

retrace blanking—The blanking of a television picture tube during vertical-retrace intervals to prevent the retrace lines from being visible on the screen.

retrace interval—*See* Return Interval.

retrace line—Also called the return line. The line traced by the electron beam in a cathode-ray tube as it travels from the end of one line or field to the start of the next line or field.

retrace time—*See* Return Interval.

retractile spring—The spring which tends to open the armature of an electromagnetic device and which holds the armature open when its force is not overcome by magnetic attraction.

retrieve—In a computer, to select specific information.

retrodirective reflector—A reflector which redirects incident flux back toward the point of origin of the flux.

return interval—Also called retrace interval, retrace time, or return time. The interval corresponding to the direction of sweep not used for delineation.

return line—*See* Retrace Line.

return lines—Conductors connecting the loads to the lowest-potential power-supply terminal. Lowest potential means most nearly zero with respect to the ground point, regardless of polarity. The term "ground" is reserved for a single point; return lines include all means of connecting the low-potential terminal of the power supply to the load such as ground busses and chassis.

return loss—1. At a discontinuity in a transmission system, the difference between the power incident upon, and the power reflected from, the discontinuity. 2. The ratio in decibels of (1) above.

return spring—*See* Restoring Spring.

return time—*See* Return Interval.

return trace—The path of the scanning spot during the return interval.

return transfer function—In a feedback control device, the transfer function that relates a loop return signal to the corresponding loop input signal.

return wire—The ground, common, or negative wire of a circuit.

reverberation—The persistence of sound due to the repeated reflections from walls, ceiling, floor, furniture, and occupants in a room or auditorium.

reverberation chamber—An enclosure in which all surfaces have been made as sound-reflective as possible. It is used for certain acoustic measurements.

reverberation period—The time required for the sound in an enclosure to die down to one millionth (60 db) of its original intensity.

reverberation strength—The difference between (a.) the level of a plane wave that produces in a nondirectional transducer a response equal to that produced by the reverberation corresponding to a 1-yard range from the effective center of the transducer and (b.) the index level of the pulse transmitted by the same transducer on any bearing.

reverberation time—For a given frequency, the time required for the average sound-energy density, originally in a steady state, to decrease to one-millionth (60 db) of its initial value after the source is stopped.

reverberation-time meter—An instrument for measuring the reverberation time of an enclosure.

reverberation unit—A circuit or device that

adds an artificial echo to a sound being reproduced or transmitted.

reversal—A change in the direction of transmission or polarity.

reverse bias—Also called back bias. An external voltage applied to a semiconductor PN junction to reduce the flow of current across the junction and thereby widen the depletion region. It is the opposite of forward bias.

reverse-blocking PNPN-type switch—A PNPN-type switch which exhibits a reverse-blocking state when its anode-to-cathode voltage is negative and does not switch in the normal manner of a PNPN-type switch.

reverse-breakdown voltage—The voltage that produces a sharp increase in reverse current in a semiconductor, without a significant increase in voltage.

reverse coupler—A directional coupler used for sampling reflected power.

reverse current—*See* Back Current.

reverse-current relay—A relay that operates whenever current flows in the reverse direction.

reversed feedback amplifier—An amplifier in which inverse feedback is employed to reduce harmonic distortion and otherwise improve fidelity.

reverse direction—Also called inverse direction. The direction of greater resistance to current flow through a diode or rectifier—i.e., from the positive to the negative electrode.

reverse emission—*See* Back Emission.

reverse key—A key used in a circuit to reverse the polarity of that circuit.

reverse-phase current relay—Also called phase-balance current relay. A device which functions when the polyphase currents are of reverse-phase sequence, or when the polyphase currents are unbalanced or contain negative phase-sequence components above a given amount.

reverse recovery time—In a semiconductor diode, the time required for the current or voltage to reach a specified state after being switched instantaneously from a specified forward-current condition to a specified reversed bias condition.

reverse resistance—The resistance measured at a specified reverse voltage or current in a diode or rectifier.

reverse saturation current—The reverse current that flows in a semiconductor because of a specified reverse voltage.

reverse voltage—The voltage applied in the reverse direction to a diode or rectifier.

reversible booster—A booster capable of adding to or subtracting from the voltage of a circuit.

reversible capacitance—For a capacitor, the limit, as the amplitude of the applied sinusoidal voltage approaches zero, of the ratio of the amplitude of the in-phase, fundamental-frequency component of transferred charge to the amplitude of the applied voltage, with a specified constant bias voltage superimposed on the sinusoidal voltage.

reversible capacitance characteristic—The function giving the relation of reversible capacitance to bias voltage.

reversible motor—A motor in which the rotation can be reversed by a switch that changes the motor connections.

reversible permeability—The limit approached by the incremental permeability as the alternating field strength approaches zero.

reversible transducer—*See* Bilateral Transducer.

reversing switch—A switch used for changing the direction of any form of motion—specifically, the direction of motor rotation or the polarity of circuit connections.

reverting call—In telephony, a call made by one party on a line to another party on the same line.

rewind—To return the tape to its starting point in a magnetic recorder.

rewind control—A button or lever for rapidly rewinding magnetic recording tape from the take-up reel to the feed reel.

rewrite—Also called regeneration. In a storage device where the information is destroyed by being read, the restoring of information into the storage.

rf—Abbreviation for radio frequency.

r-f amplifier—1. An amplifier used to increase the voltage or power at the carrier frequency. 2. Any amplifier capable of operation in the radio-frequency portion of the spectrum.

r-f bandwidth—The band of frequencies comprising 99% of the total radiated power extended to include any discrete frequency on which the power is at least 0.25% of the total radiated power.

rfc—Abbreviation for radio-frequency choke.

r-f cavity preselector—An ultrahigh-frequency circuit component which is similar in function to a tuned resonant circuit. A tunable cavity.

r-f choke—A coil used to prevent r-f currents from passing from one circuit to another.

r-f component—The portion of a signal or wave which consists only of the radio-frequency alternations, and not including its audio rate of change in amplitude frequency.

r-f current—Alternating current having a frequency higher than 10,000 hertz.

r-f energy—Alternating-current energy generated at radio frequencies.

r-f generator—A generator that produces sufficient r-f energy at its assigned frequency for induction or dielectric heating.

r-f head—A unit consisting of a radar transmitter and part of a radar receiver, the two contained in a package for ready removal and installation.

RFI (radio-frequency interference)—Radio-frequency energy of sufficient magnitude to possibly influence the operation of other electronic equipment.

r-f indicator—A device that shows the presence of r-f energy. It may consist of a tuned circuit or parallel line connected to an incandescent lamp or other indicator.

r-f interference-shield ground — The grounding technique for all shields that are used to suppress the radiation of interference from leads.

r-f intermodulation distortion—Intermodulation distortion that has its origin in the r-f stages of a receiver.

r-f line—1. A system of metallic tubes (waveguides and/or coaxial lines) which conduct radio-frequency energy from one point to another. 2. A metallic conductor used to transmit radio-frequency energy from one point to another.

r-f oscillator—*See* Radio-Frequency Oscillator.

r-f plumbing—Radio-frequency transmission lines and associated equipment in the form of waveguides.

r-f power supply—A high-voltage power supply consisting of an r-f oscillator the output voltage of which is stepped up and then rectified. Used in television receivers or other equipment to supply the high d-c voltage required by the second anode of cathode-ray tubes.

r-f preselectors—Bandpass filters which improve the selectivity by rejecting unwanted frequencies at the radio-frequency input stage.

r-f probe—1. A resonant conductor which is placed in a waveguide or cavity resonator for the purpose of inserting or withdrawing electromagnetic energy. 2. A detecting device used with a vacuum-tube voltmeter to measure r-f voltages.

r-f pulse—A radio-frequency carrier amplitude-modulated by a pulse. The amplitude of the modulated carrier is zero before and after the pulse. Coherence of the carrier with itself is not applied.

r-f resistance—*See* High-Frequency Resistance.

r-f shift—*See* Frequency Shift.

r-f signal generator—Also called service oscillator. A test instrument that generates several bands of radio frequencies necessary for the alignment and servicing of radios, television, and other electronic equipment.

r-f tolerance—The amount of r-f energy the human body can receive without injury.

r-f transformer — *See* Radio-Frequency Transformer.

R/h—Abbreviation for roentgens per hour.

rheo—Abbreviation for rheostat.

rheostat—A variable resistor which has one fixed terminal and a movable contact (often erroneously referred to as a "two-terminal potentiometer"). Potentiometers may be used as rheostats, but a rheostat cannot be used as a potentiometer because connections cannot be made to both ends of the resistance element.

RHI — Abbreviation for range-height indicator. A radar display in which the abscissa represents the range to the target, and the ordinate indicates height.

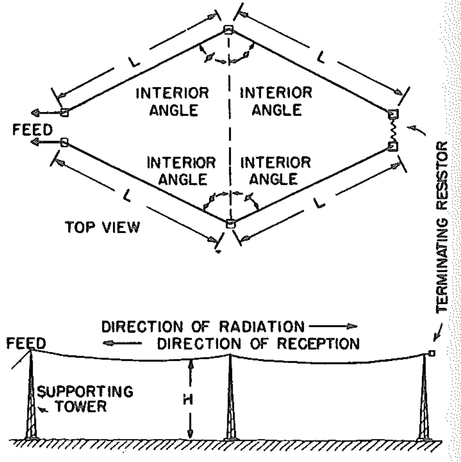

Rhombic antenna.

rhombic antenna—An antenna composed of long-wire radiators comprising the sides of a rhombus. The antenna usually is terminated in an impedance. The sides of the rhombus and the angle between them, the elevation, and the termination are proportioned to give the desired directivity.

rho-theta — A polar-coordinate navigational system providing sufficiently accurate data that a computer can be used which will provide arbitrary course lines anywhere within the coverage area of the system.

rhumbatron—A resonant cavity consisting of lumped inductance and capacitance. It is used, instead of circuits, to act as an oscillator capable of giving an output of several kilowatts at frequencies of several thousand megahertz.

rhythm bar — In some organs, a bar used to permit rhythmic playing of a chord without interrupting the pressure on the keys or chord button.

RIAA curve—1. A standard recording characteristic curve approved for long-playing records by the Record Industry Association of America. 2. The equalization curve for playback of records recorded as in (1).

ribbon—A common intraconnecting material, usually nickel, of rectangular cross section, which is used to connect electronic component parts or modules together into the form of a functioning circuit.

ribbon microphone—A microphone in which the moving conductor is in the form of a ribbon driven directly by the sound waves.

Rice neutralizing circuit—A radio-frequency amplifier circuit that neutralizes the grid-to-plate capacitance of the amplifier tube.

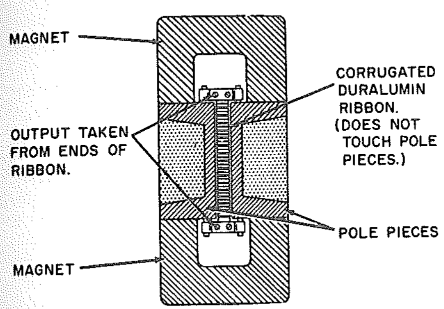
MAGNET
CORRUGATED DURALUMIN RIBBON. (DOES NOT TOUCH POLE PIECES.)
OUTPUT TAKEN FROM ENDS OF RIBBON.
POLE PIECES
MAGNET

Ribbon microphone.

Richardson effect—*See* Edison Effect.

Richardson equation—An expression for the density of thermionic emission at saturation current, in terms of the absolute temperature of the filament.

ride gain—To continually adjust the volume level of a program while observing a volume indicator so that the resulting audio-frequency signal will have the necessary magnitude for proper operation of the transmission equipment.

ridge waveguide—A circular or rectangular waveguide with one or more longitudinal ridges projecting inwardly from one or both sides. The ridges increase the transmission bandwidth by lowering the cutoff frequency.

Rieke diagram—A special polar-coordinate chart where load conditions can be determined for oscillators such as klystrons, and magetrons. Information from the chart indicates where optimum operation is located, as well as the limitations oscillators have with various loads.

rig—1. A system of components. 2. An amateur station consisting of receiver, transmitter, and all the necessary accessory equipment.

Righi-Leduc effect — The phenomenon whereby, when a metal strip is placed with its plane perpendicular to a magnetic field and heat flows through the strip, a temperature difference is developed across the strip.

right-handed polarized wave—Also called clockwise polarized wave. An elliptically polarized transverse electromagnetic wave in which the electric intensity vector rotates clockwise as an observer looks in the direction of propagation.

right-hand rule—*See* Fleming's Rule.

right-hand taper—The characteristic whereby a potentiometer or rheostat has a higher resistance in the clockwise half of its rotational range than in its counter-clockwise half (looking at the shaft end).

rigid metal conduit—A raceway specially constructed for the purpose of the pulling in or the withdrawing of wires or cables after the conduit is in place. It is made of metal pipes of standard weight and thickness, permitting the cutting of standard threads.

rim drive—The method of driving a phonograph or sound-recorder turntable by means of a small, rubber-covered wheel that is on the shaft of an electric motor and contacts the rim of the turntable.

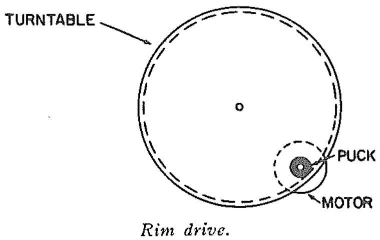

Rim drive.

rim magnet—*See* Field-Neutralizing Magnet.

ring—1. A ring-shaped contacting part of a plug usually placed in back of but insulated from the tip. 2. An audible alerting signal on a telephone line.

ring-around—In a secondary radar: 1. The undesired triggering of a transponder by its own transmitter. 2. The triggering of a transponder at all bearings, causing a ring presentation on a ppi.

ring circuit—In waveguide practice, a hybrid-T having the physical configuration of a ring with radial branches.

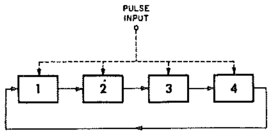

(A) Block diagram.

PULSES	STAGE			
	1	2	3	4
0	ON			
1		ON		
2			ON	
3				ON
4	ON			
5		ON		
6			ON	
7				ON
8	ON			

(B) Switching sequence.

Ring counter.

ring counter—1. A loop of interconnected bistable elements arranged so that only one is in a specified state at any given time. As input signals are counted, the specified state moves in an ordered sequence around the loop. 2. A loop or circuit of interconnected flip-flops so arranged that only one is on at any given time and that as input signals are received, the position of the "on" state moves in sequence from one flip-flop to another around the loop.

ringdown—A method of signaling subscribers and operators using either a 20-hertz a-c signal, a 135-hertz a-c signal, or a 1,000-hertz signal interrupted 20 times per second.

ring head—A magnetic head in which the magnetic material forms an enclosure with one or more air gaps. The magnetic-recording medium bridges one of these gaps and contacts or is close to the pole pieces on one side only.

ringing—1. The production of an audible or visible signal at a station or switchboard by means of an alternating or pulsating current. 2. A damped oscillation in the output signal of a system, as a result of a sudden change in the input signal. 3. High-frequency damped oscillations caused by shock excitation of high-frequency resonances.

ringing current—An alternating current which may or may not be superimposed onto a direct current for telephone ringing.

ringing key—A key which, when operated, sends a ringing current over its circuit.

ring magnet—A ceramic permanent magnet in which the axial length is no greater than the wall thickness and the wall thickness is no less than 15% of the outside diameter.

ring modulator—A modulator used as a balanced modulator, demodulator, or phase detector that has four diodes connected in series to form a ring around which current can easily flow in one direction. Input and output connections are made at the four nodal points of the ring.

ring oscillator—A circuit configuration in which two or more pairs of tubes are operated as push-pull oscillators in a ring-like arrangement. Usually alternate successive pairs of plates and grids are connected to tank circuits, and the load is coupled to the plate circuits.

ring retard—Also called slug retard. A heavy conductor surrounding the iron-flux path in a relay to retard the establishment or the decay of flux in the path.

ring time—In radar, the time during which the output of an echo box remains above a specified level. It is used in measuring the performance of radar equipment.

RIOMETER—Acronym for Relative Ionospheric Opacity METER. An instrument for recording the level of extra-terrestrial cosmic noise at selected frequencies in the hf and vhf regions.

ripple—1. The a-c component arising from sources within a d-c power supply. Unless otherwise specified, ripple is the ratio, expressed in per cent, of the root-mean-square value of the ripple voltage to the absolute value of the total voltage. 2. The excursions above and below the average peak ampressed in percent, of the root-mean-square plitude. 3. That portion of the output voltage of a power supply harmonically related in frequency to the input power, and to any internally generated switching frequency. Normally ripple is expressed as an rms percentage, but can also be expressed as peak to peak. Ripple is replaced by a new term known as pard, which includes hum and noise as well as spikes in the output.

ripple current—The alternating component of a substantially steady current.

ripple-current rating—The rms value of the a-c component of the current through a capacitor.

ripple filter—A low-pass filter designed to reduce the ripple current while freely passing the direct current from a rectifier or generator.

ripple frequency—The frequency of the ripple current. In a full-wave rectifier it is twice the supply frequency. In a generator it is a function of the speed and the number of poles.

ripple quantity—The alternating component of a pulsating quantity when this component is small relative to the continuous component.

ripple voltage—The alternating component of the unidirectional voltage from a rectifier or generator.

rise cable—In communication practice: 1. The vertical portion of a house cable extending from one floor to another. 2. Sometimes, any other vertical sections of cable.

rise time—The time required for the leading edge of a pulse to rise from 10% to 90% of its final value. It is proportionate to the time constant and is a measure of the steepness of the wavefront.

rising-sun magnetron—A multicavity vane-type magnetron in which resonators of two different resonant frequencies are arranged alternately for the purpose of mode separation.

rising-sun resonator—A magnetron anode structure in which large and small cavities alternate around the perimeter of the structure.

R/m—Abbreviation for roentgens per minute.

RMA—Abbreviation for Radio Manufacturers Association, now Electronic Industries Association (EIA).

RMA color codes—A term formerly used to designate the EIA Color Codes.

RMI—Abbreviation for radio-magnetic indicator.

rms—Abbreviation for root-mean-square.

rms amplitude—Also called effective value. The amplitude of an alternating-current wave which may be used for an accurate computation of power in watts. In communications particularly, with complex waves involved, the root-mean-square amplitude of a current is said to be that current which is computed from a measurement of power dissipated in a known resistance as a result of that current.

rms (effective) pulse amplitude—The square root of the average of the squares of the instantaneous amplitudes taken over the duration of the pulse.

rms pard—The value of the output waveform of a regulated power supply, omitting the d-c component value.

rms voltage—The effective value of a varying or alternating voltage. That value which would produce the same power loss as if a continuous voltage were applied to a pure resistance. In sine-wave voltages, the rms voltage is equal to 0.707 times the peak voltage.

Roberts rumble—The nickname given to a phenomenon whereby certain radio disturbances appear to be connected with the passage of satellites through the earth's ionosphere. Investigations have been made to determine the feasibility of exploiting this effect for the tracking of satellites.

robot pilot—*See* Autopilot.

Rochelle-salt crystal—A crystal made of sodium potassium tartrate. Because of its pronounced piezoelectric effect, it is used extensively in crystal microphones and phonograph pickups. Perfect Rochelle-salt crystals up to four inches and even more in length can be grown artificially.

rocking—Rotating the tuning control in a superheterodyne receiver back and forth while adjusting the oscillator padder near the low-frequency end of the tuning dial, to obtain more accurate alignment.

rocky-point effect—Transient but violent discharges between electrodes in high-voltage transmitting tubes.

rod gap—A spark gap in which the electrodes are two coaxial rods, with ends between which the discharge takes place, cut perpendicularly to the axis.

roentgen—The unit of radioactive dose of exposure. It is the amount of gamma radiation that will produce one electrostatic unit of charge in one cubic centimeter of air which is surrounded by an infinite mass of air at standard temperature and pressure conditions.

roentgen equivalent man—A radiation-exposure dose which produces the same effects on human tissue as one roentgen of X-ray radiation. Abbreviated REM.

roentgen meter—Also called a roentgenometer. An instrument for measuring the quantity or intensity of roentgen rays (X-rays or gamma rays).

roentgenogram—Also called an X-ray photograph or an X-ray. A photograph taken by showering an object or the human body with X-rays (roentgen rays). Depending on the transparency of the object or body, the interior can thus be seen and recorded.

roentgenology—That branch of science related to the application of roentgen rays (X-rays) for diagnostic or therapeutic purposes.

roentgenometer—*See* Roentgen Meter.

roentgen rays—*See* X-Rays.

roger—A code word used in communications to mean: 1. Your message has been received and is understood. 2. OK—an expression of agreement.

Roget spiral—A helix of wire that contracts in length, when a current is sent through, as a result of the mutual attraction between adjacent turns.

roll—Also called flip-flop, especially when intermittent. The upward or downward movement of a television picture due to lack of vertical synchronization.

rollback—*See* Rerun.

rolling transposition—The method by which two or more conductors of an open-wire circuit are spiral-wound. With two wires, a complete transposition can be executed utilizing two consecutive suspension points.

rolloff—The particular manner in which the amplitude frequency characteristic behaves as it approaches its frequency limits.

roll out—To read out of a computer storage by simultaneously increasing by 1 the value of the digit in each column, repeating this r times (where r is the radix) and, at the instant the representation changes from (r—1) to zero, either generating a particular signal, terminating a sequence of signals, or originating a sequence of signals.

roof filter—A low-pass filter used in repeater and terminal equipment to reduce the unwanted higher frequencies induced into the circuit from external sources.

room noise—*See* Ambient Noise.

root mean square—The square root of the average of the square of the values of a periodic quantity taken throughout one complete period. It is the effective value of a periodic quantity. Abbreviated rms.

root-mean-square value—Of alternating currents and voltages, the effective current or voltage applied. It is that value of alternating current or voltage that produces the same heating effect as would be produced by an equal value of direct current or voltage. It is equal to 0.707 times the maximum alternating-current value.

root-sum square—The square root of the sum of the squares. A common expression of the total harmonic distortion.

rope—Similar to chaff, but longer. Electromagnetic-wave reflectors used to confuse enemy radar. They consist of long strips of metal foil, to which small parachutes may be attached to reduce their rate of fall.

rope-lay conductor or cable—A cable consisting of one or more layers of helically laid groups of wires surrounding a central core.

rosin connection—Also called a rosin joint. A defective connection of a conductor to a piece of equipment or to another conductor. Supposedly the joint is tightly soldered, but actually it is held together only by unburnt rosin flux.

rosin-core solder—Self-fluxing solder consisting of a hollow center filled with rosin.

rosin joint—*See* Rosin Connection.

rotary-beam antenna—A highly directional short-wave antenna system. It is mounted on a mast and can be rotated manually or by an electric-motor drive to any desired position.

rotary converter—*See* Dynamotor.

rotary coupler—*See* Rotating Coupler.

rotary generator (induction-heating usage)—An alternating-current generator adapted to be rotated by a motor or other prime mover.

rotary joint—*See* Rotating Coupler.

rotary phase converter—A machine which converts power from an alternating-current system of one or more phases to an alternating-current system of a different number of phases, but of the same frequency.

rotary plunger relay—A relay in which the linear motion of the plunger is converted mechanically into rotary motion.

rotary relay—1. A relay in which the armature rotates to close the gap between two or more pole faces (usually with a balanced armature). 2. A term sometimes used for stepping relay.

rotary-solenoid relay—A relay in which the linear motion of the plunger is converted into rotary motion by mechanical means.

rotary spark gap—A device used to produce periodic spark discharges. It consists of several electrodes which are mounted on a wheel and rotate past a fixed electrode.

rotary stepping relay—*See* Stepping Relay.

rotary stepping switch—*See* Stepping Relay.

rotary switch—An electromechanical device which is capable of selecting, making, or breaking an electrical circuit. It is actuated by a rotational torque applied to its shaft.

rotary transformer—A term sometimes applied to a rotating machine used to transform direct-current power from one voltage to another.

rotary voltmeter—*See* Generating Voltmeter.

rotatable phase-adjusting transformer—A transformer in which the secondary voltage may be adjusted to have any desired phase relation with the primary voltage by mechanically orienting the secondary winding with respect to the primary. The latter winding consists usually of a distributed symmetrical polyphase winding and is energized from a polyphase circuit. (*Also see* Phase-Shifting Transformer.)

rotating-anode tube — An X-ray tube in which the anode rotates continually to bring a fresh area of its surface into the beam of electrons. This procedure allows a greater output without melting the target.

rotating coupler—Also called rotary coupler and rotary joint. A joint that permits one section of a waveguide to rotate while passing r-f energy.

rotating disc—*See* Drum Memory.

rotating element of a meter—*See* Rotor of a Meter.

rotating field—The magnetic field in the stator of induction motors. Because of excitation from a polyphase source, the field appears to rotate around the stator from pole to pole.

rotating radio beacon—A radio transmitter that rotates a concentrated beam horizontally at a constant speed. Different signals are transmitted in each direction so that ships and aircraft without directional receiving equipment can determine their bearings.

rotational life—The ability of a potentiometer to be operational after a defined number of wiper sweeps across the element.

rotational wave—*See* Shear Wave.

rotation spectrum—An X-ray spectrum of the diffraction pattern obtained when X-rays are sent through a rotating crystal.

rotator—1. A motor-driven assembly which turns an antenna so that it can be aimed in the direction of best reception. 2. In waveguides, a means of rotating the plane of polarization. In rectangular waveguides it is done by simply twisting the guide itself.

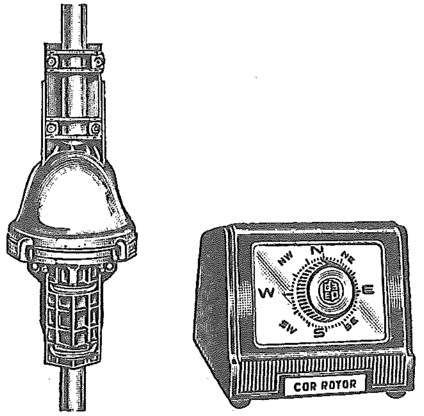

Rotator (antenna).

rotoflector—In radar, an elliptically shaped rotating reflector used to divert a vertically directed radar beam at right angles so that it radiates horizontally.

rotor—1. The rotating member of an electric machine. In a motor, it is connected to and turns the drive shaft. In a generator, the rotor is turned to produce electricity by cutting magnetic lines of force. 2. The movable plates of a variable capacitor.

rotor of a meter—Also called the rotating element of a meter. The portion driven directly by electromagnetic action.

rotor plates—The movable plates of a variable capacitor.

round conductor—A solid or stranded conductor with a substantially circular cross section.

rounding—A lack of a sharp corner of a waveform, or a smooth transition from the leading or trailing edge to the limiting final value.

rounding error—The error that results when the less significant digits of a number are dropped and the most significant digits are then adjusted.

round off—1. To delete less significant digits from a number and possibly apply some rule of correction to the part retained. 2. To approximate a number by discarding one or more final digits. If it is 5, 6, 7, 8, or 9 (or 50 . . . , etc.), the new final digit is raised by 1 (e.g., 30.7 would be rounded off to 31, and 519.2 to 519). This is done for ease of calculation, where estimate will suffice.

round-off error—*See* Rounding Error.

round-trip echoes—Multiple-reflection echoes produced when the radar pulse is reflected from a target strongly enough that the echo is reflected back to the target, where it produces a second echo.

routine—A set of computer instructions arranged in a correct sequence and used to direct a computer in performing one or more desired operations.

rpm — Abbreviation for revolutions per minute.

rps—Abbreviation for revolutions per second.

R-S flip-flop—A flip-flop having two inputs, designated R and S. At the application of a clock pulse, a 1 on the S input will set the flip-flop to the 1 or "on" state, and 1 on the R input will reset it to the 0 or "off" state. It is assumed that 1's will never appear simultaneously at both inputs.

R-S-T flip-flop—A flip-flop having three inputs, R, S, and T. The R and S inputs produce states as described for the R-S flip-flop; the T causes the flip-flop to change states.

RTMA — Abbreviation for Radio-Television-Manufacturers Association, now Electronic Industries Association (EIA).

rubber-covered wire—A wire with rubber insulation.

ruby—A type of aluminum-oxide crystal used to produce one form of solid-state laser.

ruby laser—An optically pumped solid-state laser in which a ruby crystal produces an extremely narrow and intense beam of coherent red light. Used for localized heating and for light-beam communication.

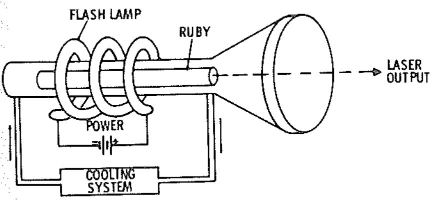

Ruby laser.

ruby maser—A maser that has a ruby crystal in the cavity resonator.

ruggedization—The redesign of a piece of equipment or its components, to make them able to withstand prolonged vibration and mechanical shock.

Ruhmkorff coil—An induction coil having a magnetic interrupter. It is used to produce a spark discharge across an air gap.

rumble—Also called turntable rumble. A descriptive term for a low-frequency vibration, which is mechanically transmitted to the recording or reproducing turntable and superimposed onto the reproduction.

runaround cross talk—Cross talk resulting from the coupling of the high-level end of one repeater to the low-level end of another repeater. Often a third repeater or line is the means of coupling; therefore, runaround cross talk may be a form of interaction cross talk.

runaway—Any additive condition to which continued exposure will eventually destroy a device.

run book — All the material needed to document a run of a program on a computer.

run motor—In facsimile equipment, a motor which supplies the power to drive the scanning or recording mechanisms. A synchronous motor is used to limit the speed.

running circuit breaker—A device the principal function of which is to connect a machine to its source of running voltage after having been brought up to the desired speed on the starting connection.

running torque—The turning power of a motor when running at its rated speed.

runway-localizing beacon — A small radio-range beacon that provides accurate directional guidance along the runway of an airport and for some distance beyond it.

rupture—The ability of contacts to break apart or rupture the electrical flow without welding under excessive currents.

rupture (or interrupting) capacity—The maximum current that a protective device will interrupt. Specified as the number of interruptions in amperes (adjusted circuit) without a change in calibration or a failure of dielectric strength.

rush-box—A superregenerative receiver.

rutherford—A quantity of radioactive material that produces one million disintegrations per second. Abbreviated rd.

ryotron—A thin-film inductive superconductive device. An inductive switch capable of inductance variation of better than three orders of magnitude.

R–Y signal—In color television, the red-minus-luminance color-difference signal. When combined with the luminance (Y) signal, it produces the red primary signal.

S

S—Symbol for secondary.

sabin (square-foot unit of absorption)—A measure of the sound absorption of a surface. It is equivalent to 1 square foot of a perfectly absorptive surface. (*Also see* Equivalent Absorption.)

saddle—Insulation placed under a splice in a coil lead.

safety factor — The amount by which the normal operating rating of a device can be exceeded without causing failure of the device.

safety service—A radiocommunication service used permanently or temporarily for the safeguarding of human life and property.

St. Elmo's fire—A visible electric discharge sometimes seen at the tips of aircraft propellers or wings, the mast of a ship, or any other metal point where there is considerable atmospheric difference of potential due to concentration of the electric field at the points of the conductor.

sal-ammoniac cell—A cell in which the electrolyte consists primarily of a solution of ammonium chloride.

salient pole—A pole consisting of a separate radial projection having its own iron pole piece and its own field coil, used in the field system of a generator or motor.

Salisbury darkbox—An isolating chamber used for test work in connection with radar equipment. The walls of the chamber are specially constructed to absorb all impinging microwave energy at a certain frequency.

sample—1. A specimen of a product selected to represent all units in a batch for inspection purposes. 2. One or more units of product drawn from a lot, the units being selected at random without regard to their quality.

sample and hold—A circuit used in an analog-to-digital converter whenever it is desirable to make a measurement of a signal and to know precisely when the input signal corresponds to the results of the measurement. It is also used to increase the duration of a signal.

sample pulse—*See* Strobe Pulse.

sampler—*See* Sampling Circuit.

sampling circuit—Also called a sampler. A circuit the output of which is a series of discrete values representative of the values of the input at a series of points in time.

sampling gate—A device which must be activated by a selector pulse before it will extract information from the input waveform.

sampling oscilloscope—An oscilloscope technique very similar in principle to the use of stroboscopic light to study fast mechanical motion. Progressive samples of adjacent portions of successive waveforms are taken; then they are "stretched" in time, amplified by relatively low-bandwidth amplifiers, and finally shown, one sample at a time, on the screen of a cathode-ray tube. The graph produced is a replica of the sampled waveforms. The principal difference in appearance between displays made by sampling techniques and conventional displays is that those made by sampling comprise separate segments or dots. This technique is limited to depicting repetitive signals, since no more than one sample is taken and displayed each time the signal occurs and provides a means for examining fast-changing signals of low amplitude that cannot be examined in any other way.

sampling plan—A program for the acceptance or rejection of a lot based on tests or inspections indicating the quality of predetermined sample sizes.

sampling rate—The number of times that a particular data channel is sampled by a commutator in one second.

sampling theory—A sampling theorem (developed by Nyquist in 1928) which states that two samples per cycle will completely characterize a band-limited signal; that is, the sampling rate must be twice the highest-frequency component. (In practice, the sampling rate is ordinarily from five to ten times the highest frequency.)

sand load—An attenuator used as a terminating section on a transmission line to dissipate power. The space between inner and outer conductors is filled with a sand and carbon mixture which acts as the dissipative element.

sandwich—A packaging method in which components are placed between boards or layers.

sapphire—A gem used on the tip of quality phonograph needles, and also for bearings in precision instruments.

satellite—A tv station licensed to rebroadcast the programming of a parent station. It differs from a translator in that satellite power limits are much higher, and satellites may also originate some programming.

saturable-core magnetometer—A magnetometer in which the change in permeability of a ferromagnetic core provides a measure of the field.

saturable-core oscillator—A relaxation oscillator in which the occurrence of saturation in a magnetic core initiates a change in the conductive state of amplifying or switching elements.

saturable-core reactor—*See* Saturable Reactor.

saturable reactor—Also called a saturable-core reactor. A magnetic-core reactor, the reactance of which is controlled by changing the saturation of the core by varying a superimposed unidirectional flux.

saturable transformer—A saturable reactor with an additional winding to provide voltage transformation or isolation from the a-c supply.

saturated color—A pure color—i.e., one not contaminated by white.

saturated recovery time—The recovery time of a thermal relay measured when the relay is de-energized after temperature saturation (equilibrium) has been reached.

saturated reoperate time—Reoperate time of a thermal relay when temperature saturation (equilibrium) is reached before the relay is de-energized.

saturating reactor—A magnetic-core reactor capable of operating in the region of saturation without independent control means.

saturating signal—In radar, a signal of greater amplitude than the dynamic range of the receiving system.

saturation—1. The attribute of any color perception possessing a hue that determines the degree of its difference from the achromatic color perception most resembling it. 2. The operating condition of a transistor when an increase in base current produces no further increase in collector current. 3. The state of magnetism beyond which a metal or alloy

is incapable of further magnetization—i.e., the point beyond which the BH curve is a straight line. 4. A circuit condition whereby an increase in the driving or input signal no longer produces a change in the output. 5. The condition when a transistor is driven so hard that it becomes biased in the forward direction. In a switching application, the charge stored in the base region prevents the transistor from turning off quickly under saturation conditions.

saturation absorption—The condition obtainable between pairs of electron energy levels in which, under very high radiation intensity, the absorption coefficient gradually decreases to zero because of the absence of empty levels to which transition can occur.

saturation control—In a color television receiver, a control which regulates the amplitude of the chrominance signal. The latter, in turn, determines the color saturation.

saturation current—1. The current in the plate circuit of a vacuum tube when all electrons emitted by the cathode pass on to the plate. 2. The current that flows between the base and collector of a transistor when an increase in the emitter-to-base voltage causes no further increase in the collector current.

saturation curve—A magnetization curve for a ferromagnetic material.

saturation flux density—*See* Saturation Induction.

saturation induction—Sometimes loosely referred to as saturation flux density. The maximum intrinsic induction possible in a material.

saturation limiting—Limiting the minimum output voltage of a vacuum-tube circuit by operating the tube in the region of plate-current saturation.

saturation magnetization—The magnetic condition of a body when an increase in the magnetizing force produces practically no change in the intensity of magnetization.

saturation noise—The noise arising when a uniformly saturated tape is reproduced. This is often some 15 db higher than the bulk-erased noise and is associated with imperfect particle dispersion.

saturation point—The point beyond which an increase in one of two quantities produces no increase in the other.

saturation resistance—Ratio of voltage to current in a saturated semiconductor.

saturation value—1. The highest value that can be obtained under given conditions. 2. The value of magnetic-flux density beyond which increases in the magnetizing force have no appreciable effect on the flux density in a particular sample of magnetic material.

saturation voltage—1. The voltage drop appearing across a switching transistor that is fully turned on. 2. Generally, the voltage excursion at which a circuit self-limits (i.e., is unable to respond to excitation in a proportional manner). In operational amplifiers, the output-voltage saturation limits may be imposed by any stage, from the input to the output, depending in part on the external loading and feedback parameters.

sawtooth current—A current that has a sawtooth waveform.

sawtooth generator—An oscillator providing an alternating voltage with a sawtooth waveform.

sawtooth voltage—A voltage that varies between two values in such a manner that the waveshape resembles the teeth of a saw.

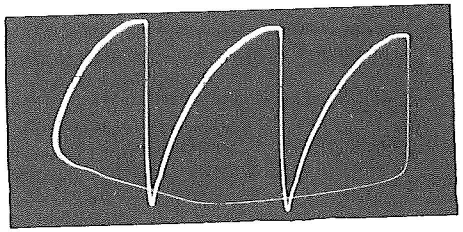

Sawtooth wave.

sawtooth wave—A periodic wave, the amplitude of which varies linearly between two values. A longer interval is required for one direction of progress than for the other.

saxophone—A linear-array antenna with a cosecant-squared radiation pattern.

SBA—Abbreviation for standard beam approach.

S-band—A radio-frequency band of 1,550 to 5,200 MHz, with wavelengths of 19.35 to 5.77 cm.

SBT—Abbreviation for surface-barrier transistor.

scalar—A quantity that has magnitude but no direction (e.g., real numbers).

scalar frequency meter—A frequency meter in which electronic circuits are used for counting and gating electrical signals to indicate their number and/or rate.

scalar function—A function which has magnitude only. Thus, the scalar product of two vectors is a scalar function, as is a real function of a real variable.

scalar quantity—Any quantity which has magnitude only—e.g., time, temperature, quantity of electricity.

scale—1. A series of musical notes, symbols, sensations, or stimuli arranged from low to high by a specified scheme of intervals suitable for musical purposes. 2. The theoretical basis of a numerical system. 3. A series of markings used for measurement or computation. 4. A defined set of values, in terms of which different quantities of the same nature can be measured. 5. In a computer, to change the units of a variable so that the problem is within its capacity.

scale division—The space between two adjacent markings on a scale.

scale factor—1. In analog computing, a proportionality factor which relates the magnitude of a variable to its representation within a computer. 2. In digital computing, the arbitrary factor which may be associated with numbers in a computer to adjust the position of the radix point so that the significant digits occupy specified columns. 3. The factor by which the number of scale divisions indicated or recorded by an instrument must be multiplied to compute the value of the measurand. 4. A value used to convert a quantity from one notation to another.

scale length of an indicating instrument—The length of the path described by the tip of the pointer (or other indicating means) in moving from one end of the scale to the other.

scale-of-ten circuit—*See* Decade Scaler.

scale-of-two counter—A flip-flop circuit in which successive similar pulses are applied at a common point, causing the circuit to alternate between its two conditions of permanent stability.

scaler—Also called a scaling circuit. A circuit which produces an output after a predetermined number of input pulses have been received.

scale span—The algebraic difference between the values of the actuating electrical quantity corresponding to the two ends of the scale of an instrument.

scaling—1. An electronic method of counting electrical pulses occurring too fast to be handled by mechanical recorders. 2. The changing of a quantity from one notation to another. 3. Adjusting the coefficient of a circuit to each of its one or more input-signal terminals. The relative scaling of one input to another is called weighting. 4. In computing, relating problem variables to machine variables.

scaling circuit—*See* Scaler.

scaling factor—Also called the scaling ratio. The number of input pulses per output pulse required by a scaler.

scaling ratio—*See* Scaling Factor.

scalloping distortion—In video tape recording, a series of small, vertical curves in the recorded image. Caused by unequal stretching across the width of the videotape.

scan—1. In facsimile, to analyze the density of successive elemental areas of the subject copy in a predetermined pattern at the transmitter, or to record these areas at the recorder. 2. To examine point by point—e.g., in converting a televised scene or image into a methodical sequence of elemental areas. 3. One sweep of the mosaic in a camera tube or of the screen in a picture tube.

scan-coded tracking system—*See* Monopulse Tracking.

scan converter—A unit for converting electronic-scan data to synchro-bearing data.

scan-converter tube—A device consisting of a cathode-ray tube and a vidicon imaging tube assembled face-to-face in the same envelope.

scanistor—An integrated semiconductor optical-scanning device that converts images into electrical signals. The output analog signal represents both the amount and the position of the light shining on its surface.

scanner—1. An instrument which automatically samples or interrogates the state of various processes, conditions, or physical states and initiates action in accordance with the information obtained. 2. In a facsimile transmitter, the part which systematically translates the densities of the subject copy into the signal waveform.

scanner amplifier—A vacuum-tube amplifier in a facsimile transmitter used to amplify the output-signal voltage of the scanner.

scanning—1. In television, facsimile, or picture transmission, the successive analyzing or synthesizing, according to a predetermined method, of the light values or equivalent characteristics of elements constituting a picture area. 2. In radar, the directing of a beam of radio-frequency energy successively over the elements of a given region, or the corresponding process in reception.

scanning-antenna mount—An antenna support which provides a mechanical means for scanning or tracking with the antenna, and a means for taking off information and using it for indication and control.

scanning beam—A beam of light, a radar beam, or an electron beam that is used in scanning.

scanning circuit – 1. A circuit which produces a linear, circular, or other movement of the beam in a cathode-ray tube at regular intervals. 2. Also called sweep circuit. That part of a cathode-ray oscilloscope which provides a time-reference base.

scanning disc—1. A rotating tricolor wheel used between the camera lens and subject, or picture tube and viewer, in field-sequential color television. 2. A Nipkow disc.

scanning frequency—*See* Stroke Speed.

scanning head—A light source and a phototube combined as a single unit for scanning a moving strip of paper, cloth, or metal in photoelectric side-register control systems.

scanning line—1. In television, the single, continuous narrow strip determined by the scanning process. In most systems, the scanning lines are blanked during the return intervals. The number of scanning lines is equal to the ratio of the line to the frame frequency. 2. A single, narrow, continuous strip containing highlights, shadows, and halftones of the picture area, as determined by the scanning process.

scanning linearity—In television, the uniformity of the scanning speed during the trace interval.

scanning-line frequency—The number of scanning lines per second. *Also see* Stroke Speed.

scanning loss—In a radar system employing a scanning antenna, the reduced sensitivity that occurs in scanning across a target, compared with the sensitivity when a constant beam is directed at the target. This loss is expressed in decibels.

scanning sonar—An echo-ranging system in which the sound pulse is transmitted simultaneously through the entire angle to be searched and a rapidly rotating transducer having a narrow beam angle scans for the returning echoes.

scanning speed—The number of inches per second explored by the spot of light or other source of energy in television, facsimile, radar, etc.

scanning spot—The immediate area being explored at any instant by a spot of light or other energy source in television, facsimile, radar, etc. *Also see* Picture Element.

scanning yoke—A yoke-shaped iron core that supports the electromagnetic deflecting coils around the neck of some cathode-ray tubes.

scannogram—The recording made on paper by a scanner.

scan rate—The rate at which a control computer periodically checks a controlled quantity.

scatter—1. A disordered change in the direction of propagation when radio waves encounter matter. 2. Spurious radar echoes due to reflections from layers of the ionosphere.

scatterband—In pulse systems, the total bandwidth occupied by the frequency spread of numerous interrogations operating on the same nominal radio frequency.

scattered reflections—Reflections from portions of the ionosphere at different virtual heights. These reflections interfere with each other and cause rapid fading of the signal.

scattering—1. The change in direction, frequency, or polarization of radio waves when they encounter matter. 2. In a narrower sense, a disordered change in the incident energy of (1) above. 3. The change of direction of particles or photons colliding with other particles or systems. 4. The diffusion of a sound or light beam due to discontinuities in the transmitting medium.

scattering loss—That part of the transmission lost because of scattering within the medium or the roughness of the reflecting surface.

scatter propagation—Also called beyond-the-horizon propagation, or over-the-horizon transmission. Transmission of high-power radio waves beyond line-of-sight distances by reflecting them from the troposphere or ionosphere.

scatter read—The ability of a computer to distribute data into several memory areas as it is being entered into the system from magnetic tape.

scc wire – Abbreviation for single-cotton-covered wire.

SCEPTRON—Acronym for Spectral ComparativE PaTtern RecOgNizer. A device which automatically classifies complex signals derived from any type of information that can be changed into an electrical signal.

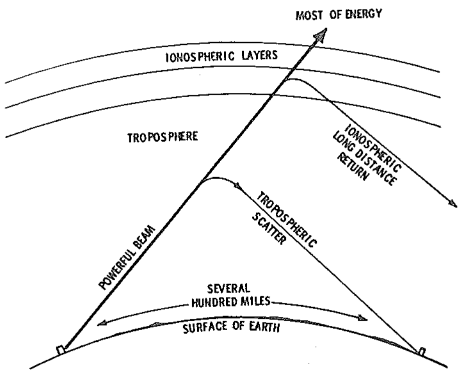

Scatter propagation.

sce wire—Abbreviation for single-cotton covering over enamel insulation on a wire.

schematic circuit diagram – *See* Schematic Diagram.

schematic diagram—Also called a schematic circuit diagram, diagram, or schematic. A diagram of the electrical scheme of a circuit, with components represented by graphical symbols.

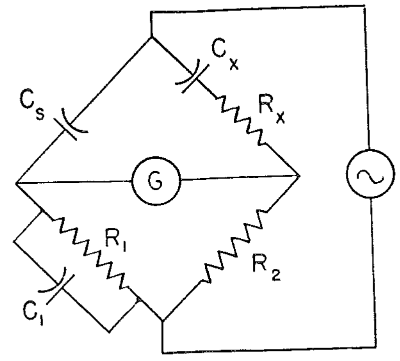

$$C_xR_2 = C_SR_1$$
$$C_xR_X = C_1R_1$$

Schering bridge.

Schering bridge—A four-arm alternating-current bridge used for measuring capacitance and dissipation factor. The unknown capacitor and a standard loss-free capacitor form two adjacent arms, the arm adjacent to the standard capacitor consists of a resistor and capacitor in parallel, and the fourth arm is a nonreactive resistor.

Schmidt optical system—An optical system for magnifying and projecting a small, brilliant image from a projection-type cathode-ray tube onto a screen.

Schmitt limiter—*See* Schmitt Trigger.

Schmitt trigger—Also called Schmitt limiter. A bistable pulse generator in which an out-

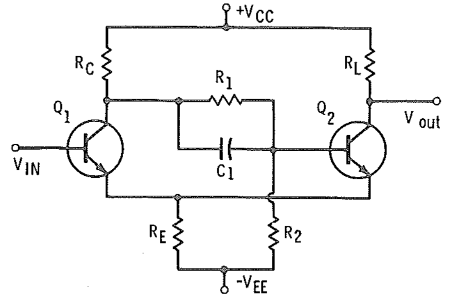

Schmitt trigger.

put pulse of constant amplitude exists only as long as the input voltage exceeds a certain d-c value.

scintillation—1. In radio propagation, a random and usually relatively small fluctuation of the received field about its mean value. 2. Also called target glint or wander. On a radar display, a rapid apparent displacement of the target from its mean position. 3. The flash of light produced by an ionic action.

scintillation conversion efficiency—In a scintillator, the ratio of the optical photon energy emitted to the energy of the incident particle or photon of ionizing radiation.

scintillation counter—A device that indirectly detects charged particles and gamma rays and neutrons by using a photomultiplier tube to convert the short flashes of light, produced as the particle passes through a transparent scintillating material, into electric signals which can be recorded. One advantage of scintillation counters is that they are very fast; i.e., they have a very small resolving time.

scintillation-counter cesium resolution—The scintillation-counter energy resolution for the gamma ray or conversion electron emitted from cesium-137.

scintillation-counter energy resolution—In a scintillation counter, a measure of the smallest discernible difference in energy between two particles or photons of ionizing radiation.

scintillation-counter energy-resolution constant—The product of the square of the scintillation-counter energy resolution times the specified energy.

scintillation-counter head—The combination of scintillators and photosensitive devices that produces electrical signals in response to ionizing radiation.

scintillation-counter time discrimination—In a scintillation counter, a measure of the smallest time interval between two successive individually discernible events. Quantitatively, the standard deviation of the time-interval curve.

scintillation decay time—The time required for the decrease of the rate of emission of optical photons in a scintillation from 90% to 10% of the maximum value.

scintillation duration—The interval from the time of emission of the first optical photon of a scintillation to the time when 90% of the optical photons of the scintillation have been emitted.

scintillation rise time—The time interval occupied by the increase of the rate of emission of optical photons of a scintillation from 10% to 90% of the maximum value.

scintillator—The combination of the body of scintillator material and its container.

scintillator material—A material that exhibits the property of emitting optical photons in response to ionizing radiation.

scintillator-material total-conversion efficiency—In a scintillator material, the ratio of the produced optical photon energy to the energy of a particle or photon of ionizing radiation that is entirely absorbed in the scintillator material.

scope—Slang for a cathode-ray oscilloscope.

scophony television system—A mechanical television projection system developed in England. The light-storage phenomenon of a supersonic light valve is utilized, and ingenious optical and mechanical methods provide large, bright images suitable for theater installation as well as home television receivers. The apparent screen brightness is multiplied several hundred times because several hundred picture elements are projected simultaneously.

scoring system—In motion-picture production, a system for recording music in time with the action on the film.

Scott connection—A method of connecting transformers to convert two-phase power to three-phase or vice versa.

scr—Abbreviation for silicon controlled rectifier.

scramble—1. To transpose and/or invert bands of frequencies, or otherwise modify the form of the intelligence at the transmitting end, according to a prearranged scheme, to obtain secrecy. 2. To make telephony unintelligible to casual interception. 3. To mix, in cryptography, in a random or quasi-random fashion.

scrambled speech—Also called inverted speech. Speech that has been made unintelligible (e.g., for secret transmission) by inverting its frequency. At the receiving end, it can then be converted back into intelligible speech by reinverting the frequency.

scrambler circuit—Also called a speech scrambler. A circuit where essential speech frequencies are divided into several ranges by filters and then inverted to produce scrambled speech. (*Also see* Speech Inverter.)

scratch filter—A low-pass filter which minimizes the needle-scratch output of a phonograph pickup by suppressing the higher audio frequencies.

screen—1. The surface upon which the visible pattern is produced in a cathode-ray tube. 2. A metal partition which isolates a device from external electric or magnetic fields. 3. *See* Screen Grid.

screen angle—A vertical angle bounded by a straight line from the radar antenna to the horizon and the horizontal at the antenna, assuming a 4/3 earth's radius.

screen dissipation—The power which the screen grid dissipates as heat after bombardment by the electron stream.

screen grid—Also called a screen. A grid placed between a control grid and an anode and usually maintained at a fixed positive potential. By reducing the electrostatic influence of the anode, it prevents the electrons from bunching in the space between the screen grid and the cathode.

screen-grid modulation—Modulation produced by introducing the signal into the screen-grid circuit of any multigrid tube where the carrier is present.

screen-grid tube—A vacuum tube in which a grid is placed between the control grid and the anode to prevent the latter from reacting with the control grid.

screen-grid voltage—The direct-voltage value applied between the screen grid and the cathode of a vacuum tube.

S-curve—An S-shaped frequency-response curve showing how the output of a frequency-modulation detector or circuit varies with frequency.

sea clutter—*See* Sea Return.

seal—A structure in which a metal part is embedded into glass or ceramic in such a way that the two are tightly bonded together.

sealed contacts—A contact assembly enclosed in a sealed compartment separate from the other parts of the relay.

sealed-gage pressure transducer—A pressure transducer that has the sensing element sealed in its case at room ambient pressure. The sealing method holds the original internal pressure for long periods of time.

sealed meter—A meter constructed so that moisture or vapor cannot enter the meter under specified test conditions.

sealed tube—Used chiefly for pool-cathode tubes. A hermetically sealed electron tube.

sealing compound—A type of wax or pitch compound used in dry batteries, capacitor blocks, transformers, or circuit units to keep out air and moisture.

sealing off—The final closing of the bulb of a vacuum tube or lamp after evacuation.

seam welding—A resistance welding process in which overlapping spot welds are made progressively along a joint by means of circular electrodes. The circular seam-welding wheels roll along the overlapping edges to be welded, and the control circuit is arranged to pass current at sufficiently close intervals to produce the desired degree of overlapping of the spot welds. The primary purpose of a seam-welding joint is to produce liquidtight or airtight containers from comparatively thin sheet metal.

search—1. In radar operation, the directing of the lobe (beam of radiated energy) in order to cover a large area. A broad-beam antenna may be used, or a rotating or scanning antenna. 2. A systematic examination of the available information in a specific field of interest. 3. To scan available stored information.

search coil—*See* Magnetic Test Coil.

search gate—A gate pulse which is made to search back and forth over a certain range.

searchlighting—In radar, the opposite of scanning. Instead, the beam is projected continuously at an object.

searchlight-type sonar—An echo-ranging system employing the same narrow beam pattern for both transmission and reception.

search radar—A radar intended primarily for displaying targets as soon as possible after their entrance into the coverage area.

search receiver—*See* Intercept Receiver.

sea return—Also called sea clutter. In radar, the aggregate received echoes reflected from the sea.

seasonal factors—Factors which are used to adjust sky-wave absorption data for seasonal variations. These variations are due primarily to seasonal fluctuations in the heights of the ionospheric layers.

seasoning—Overcoming a temporary unsteadiness of a component which may appear when the component is first installed.

seating time—The elapsed time after the coil of a relay has been energized to the time required to seat the armature of the relay.

sec—Abbreviation for second or for secondary winding of a transformer.

secondary—1. The transformer output winding where the current flow is due to inductive coupling with another coil called the primary. 2. Low-voltage conductors of a power distributing system.

secondary area—*See* Secondary Service Area.

secondary calibration—Also called sense step. Calibration of accessory equipment in which a transducer is deliberately unbalanced electrically to change the output voltage, current, or impedance. Generally performed by means of a calibration resistor which is placed across one leg of the bridge.

secondary cell—A voltaic cell which, after being discharged, may be restored to a charged condition by an electric current sent through the cell in a direction opposite that of the discharge current. *Also see* Storage Cell.

secondary color—A color produced by combining any two primary colors in equal proportions. In the light-additive process, the three secondary colors are cyan, magenta, and yellow.

secondary electron—The electron knocked loose from a molecule by another electron.

secondary-electron multiplier—An amplifier tube in which the electron stream is focused onto a succession of targets, each of which adds its secondary electrons to the stream. In this way, considerable amplification is provided.

secondary emission—The liberation of electrons from an element, other than the cathode, as a result of being struck by other high-velocity electrons. In a vacuum tube there are usually more secondary than primary electrons—a desirable phenomenon in electron-multiplier or dynatron-oscillator tubes. However, pentodes have a suppressor grid to nullify the undesirable effect of secondary emission.

secondary-emission ratio — The average number of secondary electrons emitted from a surface per incident primary electron.

secondary-emission tube — A tube which makes use of secondary emission to achieve a useful end. The photomultiplier is an example.

secondary failure—1. A failure occurring as a direct result of the abnormal stress on a component brought about by the failure of another part or parts. 2. Any failure that is the direct or indirect result of a primary failure.

secondary grid emission—Emission from the grid of a tube as a result of high-velocity electrons being driven against it and knocking off additional electrons. The effect is the same as for primary grid emission.

secondary line—The conductors connected between the secondaries of distribution transformers and the consumer service entrances.

secondary radar—*See* Radar.

secondary radiation—Random reradiation of electromagnetic waves.

secondary service area—Also called the secondary area. The service area of a radio or television broadcast station within which satisfactory reception can be obtained only under favorable conditions.

secondary standard—A unit (e.g., length, capacitance, weight) used as a standard of comparison in individual countries or localities, but checked against the one primary standard in existence somewhere in the world.

secondary storage—Storage which is not an integral part of a computer, but which is directly linked to and controlled by it.

secondary voltage—The voltage across the secondary winding of a transformer.

secondary winding—The winding on the output side of a transformer.

secondary X-rays—X-rays given off by an object irradiated with X-rays. Their frequency depends on the material in the object.

second breakdown—A condition in which the output impedance of a transistor changes almost instantaneously from a large value to a small limiting value. It may be distinguished from normal transistor operation by the fact that once it occurs, the base no longer controls normal collector characteristics. Second breakdown is associated with imperfections in the device structure, usually being more severe in multiple-diffused, high-speed devices.

second-channel attenuation—*See* Selectance.

second-channel interference—Also called alternate-channel interference. Interference in which the extraneous power originates from an assigned (authorized) signal two channels away from the desired channel.

second detector—Also called a demodulator. In a superheterodyne receiver, the portion that separates the audio component from the modulated intermediate frequency.

section—A four-terminal network which cannot be divided into a cascade of two simpler four-terminal networks.

sectionalized vertical antenna — A vertical antenna in which the continuity is broken at one or more points by the insertion of reactances or driving voltages.

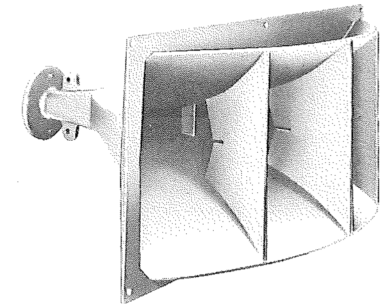

Sectoral horn.

sectoral horn—A horn with two parallel and two diverging sides.

sector cable — A multiconductor cable in which the cross section of each conductor is essentially a sector of a circle, an ellipse, or some figure intermediate between them. Sector cables are used in order to make possible the use of larger conductors in a cable of given diameter.

sector display—A range-amplitude display used with a radar set. The antenna system rotates continuously, and the screen (of the long-persistence type) is excited only while the beam is within a narrow sector centered on the object.

sector scan—A scan in which the antenna oscillates through a selected angle.

sector scanning—Modified circular scanning in which only a portion of the plane or flat cone is generated.

secular variation—A slow variation in the strength of the earth's magnetic field.

Seebeck coefficient of a couple—For homogeneous conductors, the limit, as the difference in temperature approaches zero, of the quotient of the Seebeck emf divided by the temperature difference between the junctions. By convention, the Seebeck coefficient of a couple is considered positive if, at the cold junction, the first-named conductor has a positive potential with respect to the second. It is the algebraic difference between either the relative or absolute Seebeck coefficients of the two conductors.

Seebeck effect—The production of an emf when two properly chosen materials form a closed circuit and one junction is hotter than the other.

Seebeck emf—Also called thermal emf. The emf produced by the Seebeck effect.

seed—A special single crystal from which large single crystals are grown by the Czochralski technique.

segment — 1. In a routine, the part short enough to be stored entirely in the internal storage of a computer, yet containing all the coding necessary to call in and jump automatically to other segments. 2. To divide a program into an integral number of parts, each of which performs a part of the total program and is short enough to be completely stored in internal memory.

segmented thermoelectric arm—A thermoelectric arm made up of two or more materials that have different compositions.

segmenting—*See* Partitioning.

seismic mass—The force-summing member for applying acceleration and/or gravitational force in an accelerometer.

seismograph—An instrument for recording the time, direction, and intensity of earthquakes or of earth shocks produced by explosions.

Seitz breakdown theory—Breakdown is due to the attainment of a critical avalanche size which leads to a conducting path.

selectance—1. A measure of the drop in response as a resonant device loses its resonance. It is the ratio of the amplitude of response at the resonant frequency, to the response at some other, specified frequency. 2. Often expressed as adjacent-channel attenuation (aca) or second-channel attenuation (2aca). The reciprocal of the ratio of the sensitivity of a receiver tuned to a specified channel, to its sensitivity at another channel a specified number of channels away.

selected mode—A mode of operation for an encoder selector circuit in which one set of brushes is selected to be read and another inhibited from being read; also a mode of operation for a system controlling several encoder outputs in which the encoder is selected to be read and all others inhibited from being read.

selection check—A verification of a computer instruction, usually automatic, to insure that the correct register or device has been chosen.

selection ratio—The ratio of the least magnetomotive force used to select a cell or core to the maximum magnetomotive force used which is not intended to select a cell or core.

selective—The characteristic of responding to a desired frequency to a greater degree than to other frequencies.

selective absorption—Absorption of rays of a certain group of frequencies only.

selective calling — A means of calling in which code signals are transmitted for the purpose of activating the automatic attention device at the station being called.

selective dump—A dump of a selected area of internal storage.

selective fading—Fading in which the received signal does not have the same variation in strength for all frequencies in the band. Selective fading usually occurs during multipath transmission.

selective interference—Interference the energy of which is concentrated within narrow frequency bands.

selective squelch—*See* Squelch Circuit.

selectivity—1. The characteristic which determines the extent to which the desired signal can be differentiated from disturbances of other frequencies. 2. The ability of a receiver to reject transmissions other than the one to which it is tuned. Usually shown by a curve on which the input-signal voltage required to produce a constant power output is plotted against frequency. 3. The degree to which a radio receiver can accept the signals of one station, while rejecting those of all other stations on adjacent channels.

selectivity control—The control for making a receiver more selective.

select lines — In a core memory array, the wires which pass through magnetic cores and carry the selecting coincident currents.

selector — 1. On a punch-card machine, a mechanism which reports a condition and accordingly causes a card or an operation to be selected. 2. In a telephone system, the switch or relay-group switching systems that select the path the call is to take through the system. It operates under the control of the dial at the calling station. 3. A sequential switch, usually multicontact or motor driven.

selector pulse—A pulse used to identify one event of a series.

selector relay—A relay capable of automatically selecting one or more circuits.

selector switch—A multiposition switch that permits one or more conductors to be connected to any of several other conductors.

selectron—A computer-memory tube capable of storing 256 binary digits and permitting very rapid selection and access.

selenium—A chemical element with marked photosensitive properties and a resistance that varies inversely with illumination. It is used as a rectifier layer in metallic rectifiers.

selenium cell—A photoconductive cell consisting of selenium between suitable electrodes.

selenium rectifier—A metallic rectifier in which a thin layer of selenium is deposited on one side of an aluminum plate and a highly conductive metal is coated over it.

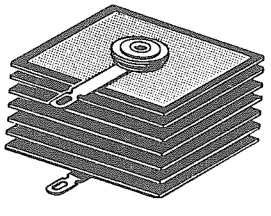

Selenium rectifier.

Electrons flow more freely from the coating to the selenium than in the opposite direction, thereby providing rectification.

self-adaptive system—A system which can exhibit the qualities of reorganization and/or learning.

self and systems testing and checkout—Logical and numerical processing for the purpose of exercising and monitoring responses of the system and the functioning of the computer itself.

self-balancing recorder—A recording device operating on the servomechanism principle.

self-bias—1. The voltage developed by the flow of vacuum-tube current through a resistor in a grid or cathode lead. 2. Also called automatic bias. A form of tube bias utilizing the voltage drop developed across a resistor through which either cathode or grid current flows.

self-capacitance—*See* Distributed Capacitance.

self-checking code—In computers, a code in which errors produce forbidden combinations. A single error-detecting code produces a forbidden combination if a digit gains or loses a single bit. A double error-detecting code produces a forbidden combination if a digit gains or loses either one or two bits, and so forth.

self-cleaning contacts—*See* Wiping Contacts.

self-complementing code—A machine language in which the code of the complement of a digit is the complement of the code of the digit.

self-contained instrument—An instrument which has all the necessary equipment built into it or the case.

self-demagnetization—The process by which a magnetized sample of magnetic material tends to demagnetize itself by virtue of the opposing fields created within it by its own magnetization.

self-excitation—The supplying of field current to a generator from its own armature.

self-excited oscillator—An oscillator that operates without external excitation and solely by the direct voltages applied to the electrode. It depends on its resonant circuits for frequency determination (i.e., not crystal controlled).

self-extinguishing—Material which ignites and burns when exposed to flame or elevated temperature, but which stops burning when the flame or high temperature is removed.

self-focused picture tube—A television picture tube with an automatic electrostatic focus designed into the electron gun.

self-generating transducer—A transducer which requires no external electrical excitation to provide an output.

self-healing capacitor—A capacitor that restores itself to operation after a breakdown caused by excessive voltage.

self-heated thermistor—A thermistor the body temperature of which is significantly higher than the temperature of its ambient medium as a result of the power being dissipated in it.

self-impedance—At any pair of terminals of a network, the ratio of an applied potential difference to the resultant current at these terminals (all other terminals open).

self-inductance—The property which determines the amount of electromotive force induced in a circuit whenever the current changes in the circuit.

self-induction—The property whereby a varying current will produce a voltage in a circuit.

self-inductor—An inductor used for changing the self-inductance of a circuit.

self-instructed carry—A system of executing the carry process in a computer by allowing information to propagate to succeeding places as soon as it is generated, without receipt of a specific signal.

self-latching relay—A relay in which the armature remains mechanically locked in the energized position until deliberately reset.

self-locking nut—A nut with an inherent locking action, so it cannot readily be loosened by vibration.

self-organizing machines—Machines that can recognize, or learn to recognize, such stimuli as patterns, characters, and sound, and which can then adapt to a changing environment.

self-powered—Equipment containing its own power supply. It may be either a combination of wet and dry cells, or dry cells in conjunction with a spring-driven motor.

self-pulse modulation—Modulation accomplished by using an internally generated pulse. (*Also see* Blocking Oscillator.)

self-pulsing—A special type of grid-pulsing circuit which automatically stops and starts the oscillations at the pulsing rate.

self-quenched counter tube—A radiation-counter tube in which reignition of the discharge is inhibited.

self-quenched detector—A superregenerative detector in which the grid-leak–grid-capacitor time constant is sufficiently large to cause intermittent oscillation above audio frequencies. As a result, normal regeneration is

stopped just before it spills over into a squealing condition.

self-quenching oscillator—An intermittent self-oscillator producing a series of short trains of r-f oscillations separated by intervals of quiescence. The quiescence is caused by rectified oscillatory currents, which build up to the point where they cut off the oscillations.

self-rectifying X-ray tube—An X-ray tube operating with an alternating anode potential.

self regulation—The tendency of a component or system to resist change in its condition or state of operation.

self-repeating timer—A form of time-delay circuit in which relay contacts are used to restart the time delay.

self-reset—Automatically returning to the original position when normal conditions are resumed (applied chiefly to relays and circuit breakers).

self-saturating rectifier—A half-wave rectifying-circuit element connected in series with the output windings of a saturable reactor in a self-saturating magnetic-amplifier circuit.

self-saturation—The saturation obtained in a magnetic amplifier by rectifying the output current of a saturable reactor.

self-screening range—The range at which a target can be detected by a radar in the midst of its jamming mask, with a certain specified probability.

self-selecting V scan—The V-scan method of reading a polystrophic code (primarily the binary code) in which diode logic circuits are used internally in the encoder to perform the necessary bit-to-bit selection to prevent ambiguity in the encoder output data.

self-starting synchronous motor—A synchronous motor provided with the equivalent of a squirrel-cage winding so it can be started like an induction motor.

self-sustained oscillations—Oscillations maintained by the energy fed back from the output to the input circuit.

self-testing—The ability of a piece of equipment to automatically verify the proper operation of its components or subsystems.

self-threading reel—A device for spooling and storing tape; in particular, one which does not require external aid to affix or start the first turn of tape on its winding surface or hub. Flanges on such reels can be continuous and free of windows; winding surfaces are continuous and free of distortion-producing threading slots.

self-wiping contacts—*See* Wiping Contacts.

selsyn—*See* Synchro.

semiactive repeater—A communications satellite which uses a minimum of on-board electronics to take a modulated signal beamed at it from a ground station and transfer its information (modulation) into an unmodulated beam (on a different frequency) sent up by the receiving station. For this transfer it uses Van Atta or other directive arrays and nonlinear elements.

semiautomatic tape relay—A method of communication whereby messages are received and retransmitted in teletypewriter tape form involving manual intervention in the transfer of the tape from the receiving reperforator to the automatic transmitter.

semiconducting material—A solid or liquid having a resistivity midway between that of an insulator and a metal.

semiconductor—A solid or liquid electronic conductor, with resistivity between that of metals and that of insulators, in which the electrical charge carrier concentration increases with increasing temperature over some temperature range. Over most of the practical temperature range, the resistance has a negative temperature coefficient. Certain semiconductors possess two types of carriers, negative electrons and positive holes. The charge carriers are usually electrons, but there may be also some ionic conductivity.

semiconductor chip—A single piece of semiconductor material of any dimension.

semiconductor device—A device in which the characteristic that distinguishes electronic conduction takes place within a semiconducting material.

semiconductor diode—1. An inactive semiconductor device with a nonlinear voltage-characteristic and two terminals. 2. A light-emitting diode that emits coherent light by suitably arranged geometry. Gallium arsenide is used for lasers of this type.

semiconductor-diode parametric amplifier—A parametric amplifier using one or more varactors.

semiconductor integrated circuit—A number of electrical elements inseparably associated on or within a continuous body of semiconductor material to perform a circuit function. *See* Monolithic IC.

semiconductor integrated circuits—Complex circuits fabricated by suitable and selective modification of areas on and in wafers of semiconductor material to produce patterns of interconnected passive and active elements. The circuit may be assembled from several chips and use thin-film elements or discrete components.

semiconductor intrinsic properties—Properties of a semiconductor that are characteristic to the ideal crystal.

semiconductor junction—The region of transition between semiconducting regions of different electrical properties, usually between P-type and N-type materials.

semiconductor laser—A light-emitting diode that uses stimulated emission to produce a coherent-light output. *Also see* Diode Laser.

semiconductor material—A material in which the conductivity ranges between that of a conductor and an insulator. The electrical characteristics of semiconductor materials such as silicon are dependent upon the small amounts of added impurities or dopants.

semidirectional microphone—A microphone, the field response of which is determined by the angle of incidence in part of the frequency range but is substantially independent of the angle of incidence in the remaining part.

semiduplex — In a communications circuit, a method of operation in which one end is duplex and one end simplex. This type of operation is sometimes used in mobile systems with the base station duplex and the mobile station or stations simplex. A semiduplex system requires two operating frequencies.

semimagnetic controller—An electrical controller in which not all its basic functions are performed by electromagnets.

semimetals—Materials such as bismuth, antimony, and arsenic having characteristics that class them between semiconductors and metals.

semiremote control—Radio-transmitter control performed near the transmitter by devices connected to but not an integral part of the transmitter.

semitone—Also called half step. The interval between two sounds. Its basic frequency ratio is equal to approximately the twelfth root of 2.

semitransparent photocathode — A photocathode in which radiant flux incident on one side produces photoelectric emission from the opposite side.

sending-end impedance — Also called the driving-point impedance. The ratio of an applied potential difference of a transmission line, to the resultant current at the point where the potential difference is applied.

sending filter—A filter used at the transmitting terminal to restrict the transmitted frequency band.

sensation level—*See* Level Above Threshold.

sense—1. In navigation, the relationship between the change in indication of a radio-navigational facility and the change in the navigational parameter being indicated. 2. In some navigational equipment, the property of permitting the resolution of 180° ambiguities. 3. To examine or determine the status of some system components. 4. To read holes in punched tape or cards.

sense finder—In a direction finder, that portion which permits determination of direction without 180° ambiguity.

sense-reversing reflectivity—The characteristic of a reflector that reverses the sense of an incident ray. (For example, a perfect corner reflector is invisible to a circularly polarized radar because it reverses the sense.)

sense step—*See* Secondary Calibration.

sensing—The process of determining the sense of an indication.

sensing element—*See* Primary Detector.

Sensistor—A silicon resistor the resistance of which varies with temperature, power, and time.

sensitive relay—A relay requiring only a small current. It is used extensively in photoelectric circuits.

sensitive volume—In a radiation-counter tube, the portion responding to a specific radiation.

sensitivity—1. The minimum input signal required in a radio receiver or similar device, to produce a specified output signal having a specified signal-to-noise ratio. This signal input may be expressed as power or voltage at a stipulated input network impedance. 2. Ratio of the response of a measuring device, to the magnitude of the measured quantity. It may be expressed directly in divisions per volt, milliradians per microampere, etc., or indirectly by stating a property from which sensitivity can be computed (e.g., ohms per volt for a stated deflection). 3. The signal current developed in a camera tube per unit incident radiation density (i.e., per watt per unit area).

sensitivity adjustment—Also called span adjustment. The control of the ratio of output signal to excitation voltage per unit measurand. Generally accomplished in a system by changing the gain of one or more amplifiers. The practice of placing excitation control components (such as potentiometers or rheostats) in series with the excitation to a transducer is a sensitivity adjustment for the system. However, in the latter case no significant change is introduced in the output-to-input ratio of the transducer.

sensitivity control—The control that adjusts the amplification of the radio-frequency amplifier stages and thereby makes the receiver more sensitive.

sensitivity-time control—Also called gain-time control or time gain. The portion of a system which varies the amplification of a radio receiver in a predetermined manner.

sensitizing (electrostatography) — The establishing of an electrostatic surface charge of uniform density on an insulating medium.

sensitometer—An instrument used to measure the sensitivity of light-sensitive materials.

sensitometry — Measurement of the light-response characteristics of photographic film.

sensor—1. In a navigational system, the portion which perceives deviations from a reference and converts them into signals. 2. A sensing element.

sentinel—1. A symbol marking the beginning or end of some piece of information in digital-computer programming. 2. *See* Tag.

separate excitation — Excitation in which generator field current is provided by an independent source, or motor field current is provided from a source other than the one connected across the armature.

separately instructed carry—Executing the carry process in a computer by allowing

carry information to propagate to succeeding places only when a specific signal is received.

separate parts of a network—The unconnected parts.

separation—The degree to which two stereo signals are kept apart. Stereo realism is dependent on the successful prevention of their mixture before reaching the output terminals of the power amplifier. Tape systems have a separation capability inherently far superior to that of the disc systems.

separation circuit—A circuit that separates signals according to their amplitude, frequency, or some other selected characteristic.

separation filter—A combination of filters used to separate one band of frequencies from another—often, to separate carrier and voice frequencies for transmission over individual paths.

separation loss—The loss that occurs in output when the surface coating of a tape fails to make perfect contact with the surfaces of either the record or reproduce head.

separator—1. An insulating sheet or other device employed in a storage battery to prevent metallic contact between plates of opposite polarity within a cell. 2. An insulator used in the construction of convolutely wound capacitors. 3. *See* Delimiter.

septate coaxial cavity—A coaxial cavity with a vane or septum added between the inner and outer conductors. The result is a cavity that acts as if it had a rectangular cross section bent transversely.

septate waveguide—A waveguide with one or more septa placed across it to control microwave power transmission.

septum—A thin metal vane which has been perforated with an appropriate wave pattern. It is inserted into a waveguide to reflect the wave.

sequence — 1. The order in which objects or items are arranged. 2. To place in order. 3. A succession of terms so related that each may be derived from one or more of the preceding terms in accordance with some fixed law.

sequence checking routine—A checking routine which examines every instruction executed and prints certain data concerning this check.

sequence control—Automatic control of a series of operations in a predetermined order.

sequencer—A mechanical or electronic device that may be set to initiate a series of events and to make the events follow in sequence.

sequence relay—A relay which controls two or more sets of contacts in a predetermined sequence.

sequencer register—In a computer, a counter which is pulsed or reset following the execution of an instruction to form the new memory address which locates the next instruction.

sequence timer — A succession of time-delay circuits arranged so that completion of the delay in one circuit initiates the delay in the following circuit.

sequency of operation—A detailed written description of the order in which electrical devices and other parts of the equipment should function.

sequential-access storage — A form of digital computer storage in which the items of stored information become available only in a one-after-another sequence regardless of whether all or only part of the information is desired.

sequential color television—A color television system in which the three primary colors are transmitted in succession and reproduced on the receiver screen in the same manner.

sequential control—Digital-computer operation in which the instructions are set up in sequence and fed to the computer consecutively during the solution of a problem.

sequential element—A device having at least one output channel and one or more input channels, all characterized by discrete states, such that the state of each output channel is determined by the previous states of the input channels.

sequential interlace—A method of interlacing in which the lines of one field are placed directly under the corresponding lines of the preceding field.

sequential lobing—A direction-determining technique utilizing the signals of overlapping lobes existing at the same time.

sequential relay—A relay that controls two or more sets of contacts in a predetermined sequence.

sequential sampling — Sampling inspection in which the decision to accept, reject, or inspect another unit is made following the inspection of each unit.

sequential scanning — In television, rectilinear scanning in which the distance from center to center of successively scanned lines is equal to the nominal line width.

sequential timer—A timer in which each interval is initiated by the completion of the preceding interval. All intervals may be independently adjusted.

serial—Pertaining to time-sequential transmission of storage of, or logical operations on, the parts of a word in a computer—the same facilities being used for successive parts.

serial arithmetic unit—One in which the digits are operated on sequentially. (*Also see* Parallel Arithmetic Unit.)

serial digital computer—One in which the digits are handled serially. Mixed serial and parallel machines are frequently called serial or parallel, according to the way the arithmetic processes are performed. An example of a serial digital computer is one which handles decimal digits serially, although the bits which comprise a digit might be handled either serially or in parallel. (*Also see* Parallel Digital Computer).

serial operation—Computer operation in which numbers are processed one character at a time—as opposed to parallel operation, where several numbers are processed simultaneously.

serial-parallel—Having the property of being partially serial and partially parallel.

serial printed—A device capable of printing characters one at a time across a page.

serial programming—Programming of a digital computer in such a manner that only one arithmetical or logical operation can be executed at one time.

serial storage—In a computer, storage in which time is one of the coordinates used in the location of any given bit, character, or word.

serial transfer—Data transfer in which the characters of an element of information are transferred in sequence over a single path.

serial transmission—Information transmission in which the characters of a word are transmitted in sequence over a single line.

series—1. The connecting of components end to end in a circuit, to provide a single path for the current. 2. An indicated sum of a set of terms in a mathematical expression (e.g., in an alternating or arithmetic series).

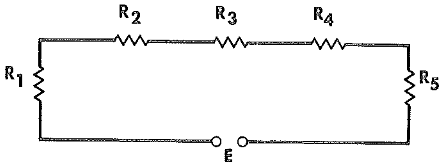

Series circuit.

series circuit—A circuit in which resistances or other components are connected end to end so that the same current flows throughout the circuit.

series connection—A way of making connections so as to form a series circuit.

series excitation—The field excitation obtained in a motor or generator by allowing the armature current to flow through the field winding.

series-fed vertical antenna—A vertical antenna which is insulated from ground and energized at the base.

series feed—The method by which the d-c voltage to the plate or grid of a vacuum tube is applied through the same impedance in which the alternating current flows.

series field—In a machine, the part of the total magnetic flux due to the series winding.

series loading—Loading in which reactances are inserted in series with the conductors of a transmission circuit.

series modulation—Modulation in which the plate circuits of a modulating tube and a modulated amplifier tube are in series with the same plate-voltage supply.

series motor—A motor in which the field and armature circuits are connected in series. In small motors with laminated field frames, the performance will be similar when operated on direct current or alternating current. For this reason, the series motor is frequently called a "universal" motor.

series operation—The connection of two or more power supplies together to obtain an output voltage of the combination equal to the sum of the individual supplies. A common current passes through all the supplies.

series-parallel network—Any network which contains only resistors, inductors, and capacitors and in which successive branches are connected in series and/or in parallel.

series-parallel switch—A switch which changes the connections of lamps or other devices from series to parallel or vice versa.

series peaking—*See* Peaking Network.

series regulator—A device that is placed in series with a source of power and is able to automatically vary its series resistance, thereby controlling the voltage or current output.

series resistor—A resistor generally used for adapting an instrument so that it will operate on some designated voltage or voltages. It forms an essential part of the voltage circuit and may be either internal or external to the instrument.

series resonance—1. The steady-state condition which exists when the current in a circuit comprising inductance and capacitance in series is in phase with the voltage across the circuit. 2. Resonance exhibited by a capacitor and inductor connected in series, characterized by low series attenuation at the resonant or tuned frequency, and high series attenuation at all others. 3. The condition existing in a circuit when the source of electromotive force is in series with an inductance and capacitance, the reactances of which cancel each other at the applied frequency, thereby reducing the impedance to minimum.

series-resonant circuit—A circuit in which an inductor and capacitor are connected in series and have values such that the inductive reactance of the inductor will be equal to the capacitive reactance of the capacitor at the desired resonant frequency. At resonance, the current through a series-resonant circuit is at maximum.

series-shunt network—*See* Ladder Network.

series T-junction—*See* E-plane T-junction.

series winding—In a motor or generator, a field winding that carries the same current as the armature; i.e., this winding is in series with the armature rather than in parallel with it. Series-wound motors are used in fractional horsepower, a-c–d-c applications, such as fans and electric mixers. Their other chief use is in heavy-duty d-c traction equipment such as electric locomotives, because of their extremely high starting torque.

series-wound motor—A commutator motor in which the field and armature circuits are in series.

serrated pulse—A vertical synchronizing pulse divided into a number of small pulses, each acting for the duration of half a line in a television system.

serrated rotor plate—Also called a slotted or split-rotor plate. A rotor plate with radial slots which permit different sections of the plate to be bent inward or outward so that the total capacitance of a variable-capacitor section can be adjusted during alignment.

serration—1. The sawtooth appearance of vertical and near-vertical lines in a television picture. This is caused by their starting at different points during the horizontal scan. 2. A designed irregular surface used as a reservoir to retain excess infiltrating material and/or multiple points to obtain high-current-density resistance welding or resistance brazing.

serrodyne—A frequency translator or frequency converter, based on linear sawtooth modulation of phase shift or time delay. One convenient modulable device for serrodyne use is a traveling-wave tube, which provides gain as well as frequency translation.

serve—With reference to cable construction, a type of separator applied directly over the conductor or conductors. The serve may consist of one or more materials such as paper, cotton, silk, nylon, or rayon. These materials may be applied spirally or laterally.

serviceability—Those properties of an equipment design that facilitate service and repair in operation.

service area—1. The area within which a navigational aid is of use. 2. The area, surrounding a broadcasting station, where the signal is strong enough for satisfactory reception at all times (i.e., not subject to objectionable interference or fading).

service band—The band of frequencies allocated to a class of radio service.

service channel—A band of frequencies, usually including a voice channel, utilized for maintenance and fault indication on a communication system.

service life—The length of time a primary cell or battery needs to reach a specified final electrical condition on a service test that duplicates normal usage.

service oscillator—*See* R-F Signal Generator.

service routine—In digital computer programming, a routine designed to assist in the actual operation of the computer.

service switch—A switch, usually in a box, for disconnecting the line voltage from the circuits it services.

service unit—In a microwave system, the equipment or facilities used for maintenance communications and transmission of fault indications.

serving (of a cable)—A wrapping applied around the core before a cable is leaded, or over the lead if the cable is armored. Some common materials are jute, cotton, or duck tape.

servo—Short for servomotor. A device that contains and delivers power to move a control or controls.

servoamplifier—A servo unit in which information from a synchro is amplified to control the speed and direction of the servomotor output.

servo loop—In a servoamplifier, the entire closed loop formed by feedback from output to input. In a position servo, the output position is compared to a command signal at the input.

servomechanism—1. An automatic feedback-control system in which one or more of its signals represent mechanical motion. 2. A system in which the output is mechanically driven by the difference between the input and the output for the purpose of making the output agree with the input.

servomotor—A motor used in a servo system. Its rotation or speed (or both) are controlled by a corrective electric signal that has been amplified and fed into the motor circuit.

servo noise—The hunting of the tracking servomechanism of a radar as a result of backlash and compliance in the gears, shafts, and structures of the mount.

servo oscillation—An unstable condition in which the load tends to hunt back and forth about the ordered position.

servo system—An automatic control system for maintaining a condition at or near a predetermined value by activation of an element such as a control rod. It compares the required condition (desired value) with the actual condition and adjusts the control element in accordance with the difference (and sometimes the rate of change of the difference).

servo techniques—Methods of studying the performance of servomechanisms or other control systems.

servovalve—(Electrohydraulic flow control.) An electrical-input, fluid-control valve capable of continuous control.

set—1. To place a storage device in a prescribed state. 2. To place a binary cell in the *1* state. 3. A permanent change, attributable to any cause, in a given parameter. 4. *See* Equipment.

set analyzer—A test instrument designed to permit convenient measurement of voltages and currents.

set composite—A signaling circuit in which two signaling or telegraph legs may be superimposed on a two-wire interoffice trunk by means of one of the balanced pairs of high-impedance coils connected to each side of the line with an associated capacitor network.

set noise—Inherent random noise caused in a receiver by thermal currents in resistors and by variations in the emission currents of vacuum tubes.

set point—In a feedback control loop, the point which determines the desired value of the quantity being controlled.

set pulse—A drive pulse which tends to set a magnetic cell.

set-reset flip-flop—A standard flip-flop except that if both the set and reset inputs are a 1 at the same time, the flip-flop will assume a prescribed state.

set terminal—*See* One-Input Terminal.

settling time—1. The time interval, following the initiation of a specified stimulus to a system, required for a specified variable to enter and remain within a specified narrow band centered on the final value of the variable. 2. The time required for the output of an amplifier to become stable after the application of a step-function input.

setup—In television, the ratio between the reference black and reference white levels, both measured from the blanking level. It is usually expressed in per cent.

setup time—The time, measured from the point of 10% input change, required for a capacitor-diode gate to open or close after the occurrence of a change of input level.

seven-unit teleprinter code — Frequently called Teletype code. A code that represents the letters of the alphabet, the numerals, the punctuation marks, and the various control functions necessary on a teleprinter (such as line feed, carriage return, upshift, and downshift) by five-unit combinations of *mark* and *space* conditions. (In addition to the five units which indicate the letter or other data, the code contains a "start" unit, which is always a "space," and a "stop" unit, which is always a "mark," to indicate the beginning and ending of each character.

sexadecimal notation—Also called dekahexadecimal notation. A scale of notation for numbers in which the base is 16.

sferics—Contraction of the term "atmospherics," meaning interference.

sferics receiver—Also called lightning recorder. A type of radio direction finder which measures electronically the direction of arrival, intensity, and rate of occurrence of atmospherics. In its simplest form the instrument consists of two orthogonally crossed antennas. Their output signals are connected to an oscillograph so that one loop measures the east-west components. These are combined vertically to give the azimuth.

sg—Abbreviation for screen-grid electrode (of a vacuum tube).

shaded-pole motor—A single-phase induction motor provided with one or more auxiliary short-circuited stator windings that are displaced magnetically from the main winding.

shading—1. A brightness gradient in the reproduced picture not present in the original scene, but caused by the camera tube. 2. Compensating for the spurious signal generated in a camera tube during the trace intervals. 3. Controlling the directivity pattern of a transducer through the distribution of phase and amplitude of the transducer action over the active face.

shading coil—*See* Shading Ring.

shading ring—A shorted turn placed around part of the pole of an alternating-current magnet to delay the change of magnetic flux in that part.

shading signal—A signal that increases the gain of the amplifier in a television camera while the electron beam is scanning a dark portion.

shadow attenuation—Attenuation of radio waves over a sphere in excess of that over a plane when the distance over the surface and other factors are the same.

shadow factor—The ratio of the electric field strength which would result from propagation over a sphere, to that which would result from propagation over a plane (other factors being the same).

shadow mask—*See* Aperture Mask.

shadow region—A region in which, under normal propagation conditions, an obstruction reduces the field strength from a transmitter to the point where radio reception or radar detection is ineffective or virtually so.

shadow tuning indicator—A vacuum tube in which a moving shadow shows how accurately a radio receiver is tuned.

shaft—The axial member to which torque is applied to cause rotation of an adjustable component.

shaft position encoder—Also called converter or coder. An analog-to-digital converter which transduces a mechanical analog shaft rotation to an electrical digital representation.

shaker—An electromagnetic device capable of imparting known and/or controlled vibratory acceleration to a given object.

shake-table test—A laboratory test in which a device or component is placed in a vibrator to determine the reliability of the device or component when subjected to vibration.

shank—1. The part of a phonograph needle which is clamped into position in the pickup or cutting head. 2. The cylindrical or rodlike portion of a connector or contact. 3. That part of a fastener lying between the head and the extreme opposite end.

Shannon limit—The maximum signal-to-noise ratio improvement which can be achieved by the best modulation technique, as implied by Shannon's theorem relating channel capacity to signal-to-noise ratio.

shaped-beam antenna—An antenna whose directional pattern over a certain angular range is designed to a special shape for some particular use.

shaped-beam display tube—A cathode-ray tube in which the beam is first deflected through a matrix, then repositioned along the axis of the tube, and finally deflected into the desired position on the faceplate. A typical tube is the Charactron.

shape factor—1. For a filter, the ratio (usually maximum) comparing a high-attenuation-level bandwidth and a low-attenuation-level bandwidth. 2. The ratio of the 60-db bandwidth to the 6-db bandwidth. Defines the selectivity of an amplifier stage.

shaping—Adjustment of a plan-position-indicator pattern set up by a rotating magnetic field.

shaping network—An electrical network designed to be inserted into a circuit to improve its transmission or impedance properties, or both. *Also see* Corrective Network.

sharp-cutoff tube—The opposite of a remote-cutoff tube. A tube in which the control-grid spirals are uniformly and closely spaced. As the grid voltage is made more and more negative, the plate current decreases steadily to cutoff.

sharp tuning—Responsive to a limited range of frequencies.

shaving—In mechanical recording, the removal of material from the surface of a recording medium for the purpose of obtaining a new surface.

shear wave—1. Also called a rotational wave. A wave, usually in an elastic solid, which causes an element of the solid to change its shape but not its volume. 2. A wave in which particle displacement is at right angles to the direction of propagation.

sheath—1. The external conducting surface of a shielded transmission line. 2. A metal wall of a waveguide. 3. Part of a discharge in a rarefied gas, in which there is a space charge due to an accumulation of electrons or ions.

sheath-reshaping converter—A converter in which the pattern of the wave is changed by gradual reshaping of the waveguide sheath and the metal sheets mounted longitudinally in the guide.

sheet grating—A three-dimensional grating consisting of thin metal sheets extending along the inside of a waveguide for about one wavelength. It is used to stop all but the predetermined wave, which passes unimpeded.

sheet resistivity—Also called ohms per square. The electrical resistance measured across the opposite side of a square pattern of deposited film material.

Sheffer-stroke function—The Boolean operator that gives a truth-table value of true only when both of the variables that the operator connects are not true.

shelf corrosion—Consumption of the negative electrode of a dry cell as a result of local action.

shelf life—The length of time under specified conditions that a material or component retains its usability.

shell—1. A group of electrons having a common energy level that forms part of the outer structure of an atom. 2. The outer section of a plug or receptacle that mechanically supports the assembly and in some cases provides coupling and locking.

shell-type transformer—A transformer in which the magnetic circuit completely surrounds the windings.

shf—Abbreviation for super-high frequency.

shield—1. Also called shielding. A screen or other housing (usually conducting) placed around devices or circuits to reduce the effect of electric or magnetic fields on them. 2. *See* Braid.

shield coverage—*See* Shield Percentage.

shielded building—In modern practice, a building in which shielding was incorporated in the basic architectural design. Shielded buildings often employ structural steel members as an integral part of RFI shielding.

shielded cable—A single- or multiple-conductor cable surrounded by a separate conductor (the shield) intended to minimize the effects of adjacent electrical circuits.

shielded-conductor cable—A cable in which the insulated conductor or conductors are enclosed in a conducting envelope or envelopes, almost every point on the surface of which is at ground potential or at some predetermined potential with respect to ground.

shielded enclosure—A room, hangar, or box, shielded or screened so as to provide a controlled electromagnetic environment.

shielded joint—A cable joint having its insulation so enveloped by a conducting shield that substantially every point on the surface of the insulation is at ground potential, or at some predetermined potential with respect to ground.

shielded line—A transmission line the elements of which confine propagated radio waves to an essentially finite space inside a tubular conducting surface called the sheath, thus preventing the line from radiating radio waves.

shielded pair—A two-wire transmission line surrounded by a metallic sheath.

shielded transmission line—A transmission line, the elements of which confine the propagated electrical energy inside a conducting sheath.

shielded wire—An insulated wire covered with a metal shield—usually of tinned, braided copper wire.

shielded X-ray tube—An X-ray tube enclosed in a grounded metal container, except for a small opening through which the X-rays emerge.

shield effectiveness—The relative ability of a shield to screen out undesirable radiation.

shield factor—Ratio between the noise (or induced current or voltage) in a telephone circuit when a source of shielding is present and when it is not.

shield grid (of a gas tube)—A structure which shields the control electrode from the anode or cathode, or both. It prevents the radiation of heat from and the depositing of thermionic activating material on them. It also reduces the electrostatic influence of the anode, and may be used as a control electrode in some applications.

shielding—The practice of confining the dielectric field of an electric cable to the inside of the cable insulation or insulated conductor assembly by surrounding the insulation or assembly with a grounded conducting medium called a shield.

shielding effectiveness—The relative reduction of radiated electromagnetic energy levels occasioned by the use of an enclosure either to contain or exclude the energy.

shield percentage—Also called shield coverage. The physical area of a circuit or cable actually covered by shielding material, expressed in percent.

shield wire—A wire employed for reducing the effects of extraneous electromagnetic fields on electric supply or communication circuits.

shift—Displacement of an ordered set of computer characters one or more places to the left or right. If the characters are the digits of a numerical expression, a shift is equivalent to multiplying by a power of the base.

shift counter—A shift register in which the first stage, through logic feedback, produces a pattern of ones or zeros as a function of the state of other states in the register. The pattern of ones and zeros so produced is termed a ring code.

shift-frequency modulation—A form of frequency modulation in which the modulating wave shifts the output frequency between predetermined values, and the output wave is coherent, with no phase discontinuity.

shift out—In a computer, to move information within a register toward one end so that, as the information leaves this one end, 0's are entered into the other end.

shift pulse—A drive pulse which initiates the shifting of characters in a register.

shift register—An element in the digital family which utilizes flip-flops to perform a displacement or movement of a set of digits one or more places to the right or left. If the digits are those of a numerical expression, a shift may be the equivalent of multiplying the number by a power of the base.

ship error—A radio direction-finder error that occurs when radio waves are reradiated by the metal structure of a ship.

ship-heading marker—On a ppi scope, an electronic radial sweep line indicating the heading of the ship on which the equipment is installed.

ship's emergency transmitter—A ship's transmitter to be used exclusively on a distress frequency for distress, urgency, or safety purposes.

ship station—A radio station operated in the maritime-mobile service and located on board a vessel which is not permanently moored.

ship-to-shore communication—Communication by radio between a ship at sea and a shore station.

shock—1. An abrupt impact applied to a stationary object. It is usually expressed in gravities (g). 2. An acceleration transient of short duration and nonrepetitive occurrence.

shock absorber—A device for dissipating vibratory energy, to modify the response of a mechanical system to an applied shock.

shock excitation—*See* Impulse Excitation.

shock-excited oscillations—*See* Free Oscillations.

shock isolator—Also called a shock mount. A resilient support which tends to isolate a system from applied shock.

Shockley diode — A four-layer controlled semiconductor rectifier diode without a base connection, used as a trigger or switching diode.

shock motion—In a mechanical system, transient motion characterized by suddenness and by significant relative displacements.

shock mount—*See* Shock Isolator.

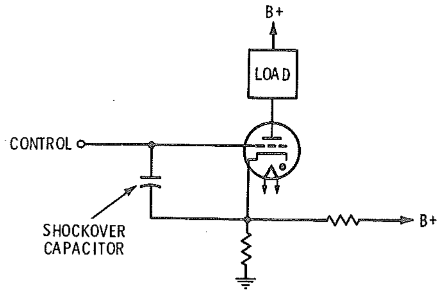

Shockover capacitor.

shockover capacitor—A capacitor connected between the grid and cathode of a thyratron to prevent premature firing.

shock pulse—Usually a single disturbance that is characterized by an increase and a decrease of acceleration in a relatively short period of time.

shodop—Acronym for Short-range Doppler.

shoran—Acronym for Short Range Air Navigation. A precision distance-measuring system employing the pulse timing principle to measure the distance from an aircraft to one or more fixed responder ground stations. Fundamentally, the system consists of a mobile transmitter-receiver-indicator unit and a fixed receiver-transmitter unit (transponder). Pulses are sent from the mobile transmitter and returned to the originating point by the transmitter. The indicator measures the time interval required for the travel of a pulse between stations and converts this information into distance to the nearest thousandth of a mile.

shore effect—The bending of radio waves toward the shore line when traveling over water, due presumably to the slightly greater velocity of radio waves over water than over land. This effect causes errors in radio direction-finder indications.

shore-to-ship communication—Communication by radio between a shore station and a ship at sea.

short—*See* Short Circuit.

short base-line system—A system which uses continuous waves and which has a base-line length that is short in comparison to the target signal.

short circuit—Also called a short. An abnormal connection of relatively low resistance between two points of a circuit. The result is a flow of excess (often damaging) current between these points.

short-circuit driving-point admittance—The driving-point admittance between the j terminal of an n-terminal network and the reference terminal when all other terminals have zero alternating components of voltage with respect to the reference point.

short-circuit feedback admittance (of an electron-device transducer) — The short-circuit transfer admittance from the output terminals to the input terminals of a specified socket, the associated filters, and the electron device.

short-circuit forward admittance (of an electron-device transducer) — The short-circuit transfer admittance from the input terminals to the output terminals of a specified socket, the associated filters and the electron device.

short-circuit impedance—The driving-point impedance of a line or four-terminal network when its far end is short-circuited.

short-circuiting or grounding device—A power or stored-energy operated device which functions to short-circuit or ground a circuit in response to automatic or manual means.

short-circuit input admittance (of an electron-device transducer)—The short-circuit driving-point admittance at the input terminals of a specified socket, the associated filters, and the electron device.

short-circuit output admittance (of an electron-device transducer) — The short-circuit driving-point admittance at the output terminals of a specified socket, the associated filters, and the electron device.

short-circuit output capacitance (of an n-terminal electron device) — The effective capacitance determined from the short-circuit output admittance.

short-circuit parameters—In an equivalent circuit of a transistor, the resultant parameter when independent variables are selected for the input and output voltages.

short-circuit protection (automatic)—A current-limiting system which enables a power supply to continue operating without damage into any output overload, including short circuits. The output voltage must be restored to normal when the overload is removed, as distinguished from a fuse or circuit-breaker system, which opens at overload and must be reclosed to restore power.

short-circuit transfer admittance — The transfer admittance from terminal j to terminal l of an n-terminal network when all terminals except j have zero complex alternating components of voltage with respect to the reference point.

short-circuit transfer capacitance (of an electron device)—The effective capacitance determined from the short-circuit transfer admittance.

short code—A system of instructions that causes an automaton to behave as if it were another, specified automaton.

short-contact switch—A selector switch in which the movable contact is wider than the distance between its clips, so that the new circuit is made before the old one is broken.

short-distance navigational aid—An equipment or system which provides navigational assistance to a range not exceeding 200 miles.

shorted out—Made inactive by connecting a heavy wire or other low-resistance path around a device or portion of a circuit.

short-gate gain—Video gain on a short-range gate.

shorting noise—A noise which occurs in wirewound potentiometric transducers, even when no current is drawn from the device. It is due to the shorting out of adjacent turns of the wire as the slider traverses the winding. The portion of the interturn current which flows through the slider appears as noise.

shorting switch—A switch type in which contact is made for a new position before breaking contact with the previous position. Classified as a "make-before-break" switch.

short plug—A plug designed to connect the springs of a jack together or to short them.

short-range navigation—A precision position-fixing system using a pulse transmitter and receiver in connection with two transponder beacons at fixed points.

short-slot coupler—*See* Three-db Coupler.

short-time duty—A service requirement that demands operation at a substantially constant load for a short, specified time.

short time limits—Values of minimum and maximum trip time measured at various percentages of overload.

short-time rating—The rating that defines the load which a machine, apparatus, or device can carry at approximately the room temperature for a short, specified time.

short wave—Radio frequencies which fall above the commercial broadcasting band from 1.5 to 30 MHz) and are used for sky-wave communication over long distances.

short-wave converter—An electronic unit designed to be connected between a receiver and its antenna system to permit reception of frequencies higher than those the receiver ordinarily handles.

short-wave transmitter—A radio transmitter that radiates short waves, which ordinarily are shorter than 200 meters.

shot effect—Noise voltages developed by the random travel of electrons within a tube. The effect is characterized by a steady hiss from a radio, and by snow or grass in a television picture.

shot noise—1. Noise generated due to the random passage of discrete current carriers across a barrier or discontinuity, e.g., a semiconductor junction. Shot noise is characteristic of all transistors and diodes and is directly proportional to the square root of the applied current. 2. Noise caused by thermal agitation in vacuum tubes.

short code — A system of instructions that causes an automaton to behave as if it were another, specified automaton.

shunt — 1. A precision low-value resistor placed across the terminals of an ammeter to increase its range. 2. Any part connected, or the act of connecting any part, in parallel with some other part. 3. In an electric circuit, a branch the winding of which is in parallel with the external or line circuit.

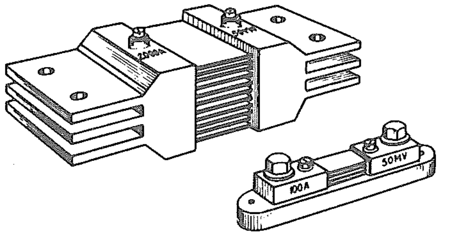

Shunt.

shunt calibration—A procedure in which a parallel resistance is placed across a (similar) element to obtain a known and deliberate electrical change.

shunt-fed vertical antenna—A vertical antenna connected to the ground at the base, and energized at a suitable point above the grounding point.

shunt feed—*See* Parallel Feed.

shunt field—Part of the magnetic flux produced in a machine by the shunt winding connected across the voltage source.

shunt-field relay—A special polarized relay with two coils on opposite sides of a closed magnetic circuit. The relay operates only when the currents in its two windings flow in the same direction.

shunting effect—A reduction in signal amplitude caused by the load which an amplifier or measuring instrument imposes on the signal source. For d-c signals the shunting effect is directly proportional to the output impedance of the signal source and inversely proportional to the input impedance of the amplifier.

shunting or discharge switch—A switch which serves to open or to close a shunting circuit around any piece of apparatus (except a resistor) such as a machine field, a machine armature, a capacitor, or a reactor.

shunt leads—Those leads which connect the circuit of an instrument to an external shunt. The resistance of these leads must be taken into account when the instrument is adjusted.

shunt loading—Loading in which reactances are applied in parallel across the conductors of a transmission circuit.

shunt neutralization—*See* Inductive Neutralization.

shunt peaking—*See* Peaking Network.

shunt regulator—A device placed across the output of a regulated power supply to control the current through a series-dropping resistance in order to maintain a constant output voltage or current.

shunt T-junction—*See* H-plane T-junction.

shunt-wound generator — A direct-current generator in which the field coils and armature are wound in parallel.

shunt-wound motor—A direct-current motor in which the field circuit and armature circuit are connected in parallel.

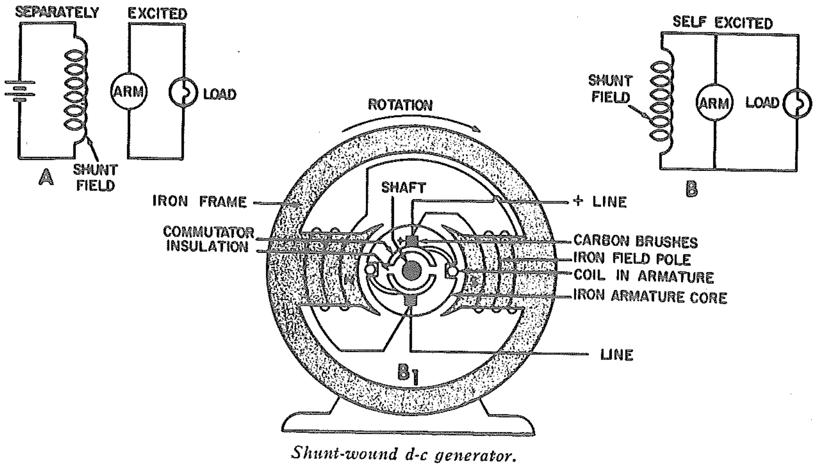

Shunt-wound d-c generator.

shutter—A movable cover that prevents light from reaching the film or other light-sensitive surface in a still, movie, or television camera except during the exposure time.

side armature—An armature which rotates about an axis parallel to that of the core, with the pole face on a side surface of the core of a relay.

sideband attenuation—Attenuation in which the relative transmitted amplitude of one or more components of a modulated signal (excluding the carrier) is smaller than the amplitude produced by the modulation process.

sideband power—The power contained in the sidebands. This is the power to which a receiver responds when receiving a modulated wave, not the carrier power.

sidebands—1. The frequency bands on both sides of the carrier frequency. The frequencies of the wave produced by modulation fall within these bands. 2. The wave components lying within such bands. During amplitude modulation with a sine-wave carrier, the upper sideband includes the sum (carrier plus modulating) frequencies, and the lower sideband includes the difference (carrier minus modulating) frequencies.

side circuit—A circuit arrangement for deriving a phantom circuit. In four-wire circuits, the two wires associated with the "go" channel form one "side circuit" and those associated with the return channel form another. (*Also see* Phantom Circuit.)

side echo—An echo due to a side lobe of an antenna.

side frequency—One of the frequencies of a sideband.

side lobe—A portion of the beam from an antenna, other than the main lobe. It is usually much smaller than the main lobe.

side-lobe blanking—A technique which compares relative signal strengths between an omnidirectional antenna and the radar antenna.

sideswiper—A telegraph key which operates from side to side rather than up and down.

side thrust—In disc recording, the radial component of force on a pickup arm caused by the stylus drag.

sidetone—The reproduction, in a telephone receiver, of sounds received by the transmitter of the same telephone set (e.g., hearing one's own voice in the receiver of a telephone set when speaking into the mouthpiece).

sidetone telephone set—A telephone set with no balancing network for reducing sidetone.

siemens—The new and preferred term for mho.

sight check—To verify the sorting or punching of punched cards by looking through the pattern of punched holes.

sign—1. A symbol which distinguishes negative from positive quantities. 2. A symbol which indicates whether a quantity is greater or less than zero. 3. A binary indicator of the position of the magnitude of a number relative to zero.

signal—1. A visible, audible, or other conveyor of information. 2. The intelligence, message, or effect to be conveyed over a communication system. 3. A signal wave. 4. The physical embodiment of a message.

signal attenuation—The reduction in the strength of electrical signals.

signal bias—A form of teletypewriter signal distortion brought about by the lengthening or shortening of pulses during transmission. When marking pulses are all lengthened, a marking-signal bias results; when marking pulses are all shortened, a spacing-signal bias results.

signal-carrier f-m recording—A method of recording in which the input signal is frequency modulated onto a carrier, and the carrier is recorded on a single track at saturation and without bias.

signal conditioning—To process the form or mode of a signal so as to make it intelligible to, or compatible with, a given device, including such manipulation as pulse shaping, pulse clipping, digitizing, and linearizing.

signal converter—A circuit that reduces, filters, and (if necessary) rectifies incoming signals to logic system levels.

signal delay—The transmission time of a signal through a network. The time is always finite, and it may be undesired or purposely introduced.

signal-distortion generator—An instrument furnished and designed to apply distortion on a signal for the purpose of ranging and adjusting teletypewriter equipment or for furnishing a clear signal.

signal electrode—The electrode from which the signal output of a camera tube is taken.

signal element—Also called a unit interval. That part of a signal which occupies the shortest interval of the signaling code. It is considered to be of unit duration in building up signal combinations.

signal-frequency shift—In a frequency-shift facsimile system, the numerical difference between the frequencies corresponding to the white and black signals at any point in the system.

signal generator—Also called a standard voltage generator. A device which supplies a standard voltage of known amplitude, frequency, and waveform for measuring purposes.

signal ground—The ground return for low-level signals such as inputs to audio amplifiers or other circuits that are susceptible to coupling through ground-loop currents.

signaling channel—A tone channel used for signaling purposes.

signaling key—A key used in wire or radiotelegraphy to control the sequence of current impulses that form the code signals.

signal injector—A test instrument, usually small, which contains an audio-frequency pulse oscillator. A signal injected at points in the circuitry aids in troubleshooting.

signal intelligence—A generic term which includes both communications intelligence and electronics intelligence.

signal interpolation—*See* Interpolation.

signal lamp—A lamp that indicates, when lit or out, the existence of certain conditions in a circuit (e.g., signal lamps on switchboards, or pilot lamps in radio sets).

signal leakage—Interference in a given playback channel which has its origin in the recording system. Such interference occurs during simultaneous record/reproduce, and has a leading time displacement with reference to the signal on the tape.

signal level—The difference between the measure of the signal at any point in a transmission system, and the measure of an arbitrary reference signal. (Audio signals are often stated in decibels—thus their difference can be conveniently expressed as a ratio.)

signal-muting switch—A switch used on a record changer to ground (mute) the signal from the pickup during a change cycle.

signal-noise ratio—*See* Signal-to-Noise Ratio.

signal plate—A metal plate that backs up the mica sheet containing the mosaic in one type of cathode-ray television camera tube. The electron beam acts on the capacitance between this plate and each globule of the mosaic to produce the television signal.

signal plus noise and distortion—A radio-receiver sensitivity measurement based on the signal input required to produce 50% of the rated output at a 12-decibel ratio (4:1 voltage ratio) of signal plus noise and distortion to noise and distortion alone.

signal-separation filter—A bandpass filter which selects the desired signal or channel from a composite signal.

signal-shaping network—An electric network inserted into a telegraph circuit, usually at the receiving end, to improve the waveshape of the signals.

signal-shield ground — A ground technique for all shields used for the protection from stray pickup of leads carrying low-level, low-frequency signals.

signal shifter—A variable-frequency oscillator for shifting amateur transmitters to a less crowded frequency within a given band.

signal strength—The strength of the signal produced by a transmitter at a particular location. Usually it is expressed as so many millivolts per meter of the effective receiving-antenna length.

signal-strength meter—Also called an S meter. A meter connected in the avc circuit of a receiver and calibrated in db or arbitrary "S" units to read the strength of a received signal.

signal-to-noise ratio — Also called signal-noise ratio. 1. Ratio of the magnitude of the signal to that of the noise (often expressed in decibels). 2. In television transmission, the ratio in decibels of the maximum peak-to-peak voltage of the video television signal (including the synchronizing pulse), to the rms voltage of the noise at any point. 3. The ratio of the amplitude of a signal after detection to the amplitude of the noise accompanying the signal. It may also be considered as the ratio, at any specific point of a circuit, of signal power to total circuit-noise power. Abbreviated s/n ratio or snr.

signal tracer—A test instrument used for tracing a signal through the circuit in order to find faulty wiring or components.

signal tracing — The tracing of a signal through each stage in order to locate a fault.

signal voltage — The effective (root-mean-square) voltage value of a signal.

signal wave—A wave with characteristics that permit it to carry intelligence.

signal-wave envelope—The contour of a signal wave which is composed of a series of wave cycles.

signal winding—Also called an input winding. In a saturable reactor, the control winding to which the independent variable (signal wave) is applied.

signature (target)—The characteristic pattern of the target displayed by detection and classification equipment.

sign-control flip-flop—In computers, a flip-flop in the arithmetic unit used for storing the sign of the result of an operation.

sign digit—A character (+ or −) used to designate the algebraic sign of a number.

significant digits (of a number)—A set of digits from consecutive columns, beginning with the most significant digit other than zero, and ending with the least significant digit the value of which is known or assumed to be relevant. The digits of a number can be ordered according to their significance, which is greater when occupying a column corresponding to a higher power of the radix.

silent discharge—The gradual and nondisruptive discharge of electricity from a conductor into the atmosphere. It is sometimes accompanied by the production of ozone.

silent period — An hourly period during which ship and shore radio stations must remain silent and listen for distress calls.

silica gel — A moisture-absorbent chemical used for dehydrating waveguides, coaxial lines, pressurized components, shipping containers, etc.

silicon—A metallic element often mixed with iron or steel during smelting to provide desirable magnetic properties for transformer-core materials. In its pure state, it is used as a semiconductor.

silicon capacitor—*See* Varactor.

silicon controlled rectifier—A three-junction semiconductor device that is normally an open circuit until an appropriate gate signal is applied to the gate terminal, at which time it rapidly switches to the conducting state. Its operation is equivalent to that of a thyratron.

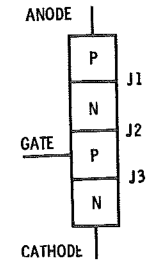

Silicon controlled rectifier.

silicon controlled switch—A PNPN semiconductor device with leads to all four semiconductor regions.

silicon detector—*See* Silicon Diode.

silicon diffused epitaxial mesa transistor —A silicon transistor that has high voltage and power ratings and low storage time and saturation voltage.

silicon diode—Also called a silicon detector. A crystal detector used for rectifying or detecting uhf and shf signals. It consists of a metal contact held against a piece of silicon in a particular crystalline state.

silicon double-base diode—*See* Unijunction Transistor.

silicone — A member of the family of polymeric materials characterized by a recurring chemical group that contains silicon and oxygen atoms as links in the main chain. These compounds are presently derived from silica (sand) and methyl chloride. One of their important properties is resistance to heat.

silicon monoxide—A dielectric material often used in the fabrication of a microelectronic device to form an insulator, substrate, or a thin-film capacitor dielectric.

silicon oxide—A dielectric material commonly used in the surface passivation of microelectronic circuits. Silicon oxide contains various combinations of silicon monoxide and silicon dioxide.

silicon planar transistor—A silicon transistor produced by the planar process and consisting of a series of etchings and diffusions to produce a transistor with a thin oxide layer within the planes of a silicon substrate.

silicon rectifier—One or more silicon rectifying cells or cell assemblies.

silicon rectifying cell—An elementary two-terminal silicon device which consists of a positive and a negative electrode and conducts current effectively in only one direction.

silicon solar cell—A photovoltaic cell designed to convert light energy into power for electronic and communication equipment. It consists essentially of a thin wafer of specially processed silicon.

silicon steel—Steel containing 3% to 5% silicon. Its magnetic qualities make it desirable for use in the iron cores of transformers and in other a-c devices.

silicon transistor—A transistor in which silicon is used as the semiconducting material.

silicon unijunction transistor — *See* Unijunction Transistor.

silk-covered wire—A wire covered with one or more layers of fine floss silk. It is a better insulator than cotton. Also, it is more moisture-resistant and permits more turns of wire within a given space.

silver—A precious metal which is more conductive than copper. Because it does not readily corrode, it is used for contact points of relays and switches. Its chemical symbol is *Ag*.

silvered-mica capacitor—A mica capacitor that has a coating of silver deposited directly on the mica sheets instead of using conducting metal foil.

silvering—*See* Silver Spraying.

silver migration—The ionic displacement of metallic silver through an insulating medium. Usually it is caused by a combination of extended time, high humidity, temperature variations, and d-c potential.

silver solder—A solder that is composed of copper, silver, and zinc. It has a melting point lower than that of silver, but higher than that of lead-tin solder.

silver spraying—Also calling silvering. Metallizing the surface of an original master disc recording by using a dual spray nozzle in which ammoniated silver nitrate and a reducer are combined in an atomized spray to precipitate the metallic silver.

silverstat—An arrangement of closely spaced contactors. Sometimes used as a step-by-step device to unbalance the arms of a resistance bridge.

silverstat regulator—A multitapped resistor, the taps of which are connected to single-leaf silver contacts. Variation of voltage causes a solenoid to open or close these contacts, shorting out more or less of the resistance in the exciter circuit as a means of regulating the output voltage to the desired value.

simple harmonic current—Also called sinusoidal current. A symmetrical alternating

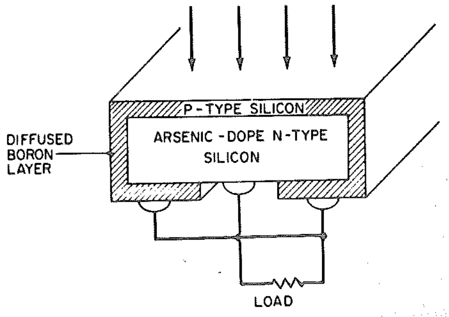

Silicon solar cell.

current, the instantaneous value of which is equal to the product of a constant, and the sine or cosine of an angle having a value varying linearly with time.

simple harmonic electromotive force—Instantaneous values of a symmetrical alternating electromotive force which are equal to the product of a constant and the cosine or sine of an angle, the values of which vary linearly with time.

simple quad—*See* S-quad.

simple scanning—Scanning of only one scanning spot at a time.

simple sound source—A source which radiates sound uniformly in all directions under free-field conditions.

simple steady-state vibration—A periodic motion made up of a single sinusoid.

simple target—In radar, a target the reflecting surface of which does not cause the amplitude of the reflected signal to vary with the aspect of the target (e.g., a metal sphere).

simple tone—1. A sound wave, the instantaneous sound pressure of which is a simple sinusoidal function of time. 2. Also called a pure tone. A sound sensation characterized by its singleness of pitch.

simplex—A form of communication satellite operation which involves communication in only one direction at a time (mainly for facsimile, television, and some data).

simplex channel — A path for electrical transmission of information in one direction between two or more terminals.

simplex coil—A repeating coil used on a pair of wires to derive a commercial simplex circuit.

simplexed circuit—A two-wire metallic circuit from which a simplex circuit is derived, the metallic and simplex circuits being capable of simultaneous use.

simplex operation — Communication that takes place in only one direction at a time between two stations. Included in this classification are ordinary transmit-receive or press-to-talk operation, voice-operated carrier, and other forms of manual or automatic switching from transmit to receive.

simulation—A type of problem in which a physical model and the conditions to which the model may be subjected are all represented by mathematical formulas.

simulator—A device which represents a system or phenomenon and which reflects the effects of changes in the original so that it may be studied, analyzed, and understood from the behavior of that device.

simulcast—1. To broadcast a program simultaneously over more than one type of broadcast station, e.g., to broadcast a stereophonic program over an a-m and f-m station. 2. A program so broadcast.

simulcasting—Broadcasting a stereo program over an a-m and f-m station. An a-m and f-m tuner are required for stereo reception.

simultaneous lobing—In radar, a direction-determining technique utilizing the received energy of two concurrent and partially overlapped signal lobes. The relative phase or power of the two signals received from a target is a measure of the angular displacement of the target from the equiphase or equisignal direction.

sine—The sine of an angle of a right triangle is equal to the side opposite that angle, divided by the hypotenuse (the long side opposite the right angle).

sine galvanometer—An instrument resembling a tangent galvanometer except that its coil is in the plane of the deflecting needle. The sine of the angle of deflection will then be proportionate to the current.

sine law—The law which states that the intensity of radiation in any direction from a linear source varies in proportion to the sine of the angle between a given direction and the axis of the source.

sine potentiometer — A d-c voltage divider (potentiometer), the output of which is proportionate to the sine of the shaft-angle position.

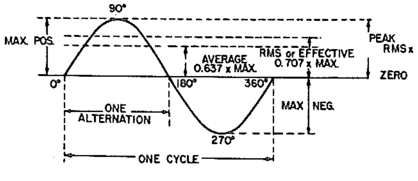

Sine wave.

sine wave—A wave which can be expressed as the sine of a linear function of time, space, or both.

singing—An undesired self-sustained oscillation at a frequency in or above the passband of a system or component.

singing margin—Also called gain margin. The excess of loss over gain around a possible singing path at any frequency, or the minimum value of such excess over a range of frequencies.

singing point—1. The condition of a circuit or transmission path where the sum of the gains exceeds the sum of the losses. When expressed in decibels, it is the gain that can be added to the circuit equivalent before singing will begin. 2. The amount of total gain in a transmission system (most commonly used in connection with two-wire repeaters) which causes the system to begin to lose efficiency of performance because the self-oscillating point is too closely approached. 3. The singing point of a circuit which is coupled back to itself is the point at which the gain is just sufficient to make the circuit break into oscillation.

single-address code—An instruction which contains the location of the data and the operation or sequence of operations to be performed on these data.

single amplitude—With reference to vibratory conditions, the peak displacement of an oscillating structure from its average or mean position.

single-anode tank—*See* Single-Anode Tube.

single-anode tube—Also called a single-anode tank. An electron tube with one anode (used chiefly for pool-cathode tubes).

single-axis gyro—A type of gyro in which the spinning rotor is mounted in a gimbal arranged so as to tilt about only one axis relative to the stable element.

single-button carbon microphone—A microphone having a carbon-filled buttonlike container on one side of its flexible diaphragm. As the sound waves move the diaphragm, the resistance of the carbon changes and the microphone current constitutes the desired audio-frequency signal.

single-carrier f-m recording—The method of recording in which the input signal is frequency-modulated onto a carrier and the carrier is recorded on a single track at saturation and without bias.

single-channel—A carrier-only or single-tone modulated radio control transmitter and matching receiver installation.

single-channel monopulse tracking system—*See* Monopulse Tracking.

single-channel simplex — Nonsimultaneous communication between stations over the same frequency channel.

single circuit—A telegraph circuit capable of nonsimultaneous two-way communication.

single-degree-of-freedom system—A system for which only one co-ordinate is required to define the configuration of the system.

single-dial control—Control of a number of different devices or circuits by means of a single adjustment (e.g., in tuning all variable-capacitor sections of a radio receiver).

single-ended—Unbalanced, such as grounding one side of a circuit or transmission line.

single-ended amplifier — An amplifier in which only one tube is normally employed in each stage—or if more than one tube is used, they are connected in parallel so that operation is asymmetric with respect to ground.

single-ended push-pull amplifier circuit—An amplifier circuit having two transmission paths designed to operate in a complementary manner and connected to provide a single unbalanced output. (No transformer is used.)

single-ended tube—A metal tube in which all electrodes—including the control grid—are connected to base pins and there is no top connection. The letter *S* after the first numerals in a receiving-tube designation (e.g., 6SN7) indicates a single-ended tube.

single-groove stereo — *See* Monogroove Stereo.

single-gun color tube—A color picture tube with a single electron gun that produces only one beam, which is sequentially deflected across the phosphor dots.

single harmonic distortion—The ratio of the power at the fundamental frequency measured at the output of the transmission system considered, to the power of any single harmonic observed at the output of the system because of its nonlinearity, when a single-frequency signal of specified power is applied to the input of the system. It is expressed in db.

single-hop propagation — Transmission in which the radio waves are reflected only once in the ionosphere.

single-junction photosensitive semiconductor—Two layers of semiconductor materials with an electrode connection to each material. Light energy controls the amount of current flow.

single-line diagram — A form of schematic diagram in which single lines are used to show component interconnections even though two or more conductors are required in the actual circuit.

single-loop feedback — A loop in which feedback may occur only through one electrical path.

single-phase circuit—Either an alternating-current circuit with only two points of entry, or one with more than two points of entry but energized in such a way that the potential differences between all pairs of points of entry are either in phase or 180° out of phase. A single-phase circuit with only two points of entry is called a single-phase, two-wire circuit.

single-phase synchronous generator — A generator which produces a single alternating electromotive force at its terminals.

single phasing—The tendency of the rotor (of a motor tach generator) to continue to rotate when one winding is opened and the other winding remains excited.

single-point grounding—A grounding system that attempts to confine all return currents to a network which serves as the circuit reference. It does not imply that the grounding system is limited to one earth connection. To be effective, no appreciable current is allowed to flow in the circuit reference; i.e., the sum of the above return currents is zero.

single-polarity pulse—A pulse which departs from normal in one direction only.

single pole—A contact arrangement in which all contacts in the arrangement connect, in one position or another, to a common contact.

single-pole, double-throw—Abbreviated spdt. A three-terminal switch or relay contact for connecting one terminal to either of two other terminals.

single-pole-piece magnetic head — A magnetic head with only one pole piece on one side of the recording medium.

single-pole, single-throw—Abbreviated spst. A two-terminal switch or relay contact which either opens or closes one circuit.

single sampling plan—The plan which consists of a single sample size with associated acceptance and rejection criteria.

single-shield solid enclosure—An all-metal enclosure providing higher attenuation than cell-type units. It is usually a rigid, free-standing enclosure.

single-shot blocking oscillator—A blocking oscillator modified to operate as a single-shot trigger circuit.

single-shot multivibrator — Also called a single-trip multivibrator. A multivibrator modified to operate as a single-shot trigger circuit. *Also see* Monostable Multivibrator.

single-shot trigger circuit—Also called a single-trip trigger circuit. A trigger circuit in which the pulse initiates one complete cycle of conditions ending with a stable condition.

single-sideband filter—A bandpass filter in which the slope on one side of the response curve is greater than on the other side. So-called because it is used in systems to suppress a carrier frequency and transmit one or both sidebands.

single-sideband modulation—Abbreviated ss or ssb. Modulation whereby the spectrum of the modulating wave is translated in frequency by a specified amount, either with or without inversion.

single-sideband suppressed carrier—Modulation resulting from the partial or complete elimination of the carrier and all components of one sideband from an amplitude-modulated wave.

single-sideband system — A type of radiotelephone service in which one set of sidebands (either the upper or lower) is completely suppressed and the transmitted carrier is partially suppressed.

single-sideband transmission—Transmission of only one sideband, the other sideband being suppressed. The carrier wave may be transmitted or suppressed.

single-sideband transmitter—A transmitter in which only one sideband is transmitted.

single-signal receiver—A superheterodyne receiver equipped for single-signal reception. A crystal filter, usually in the intermediate-frequency amplifier, can be shorted out by a switch when high selectivity is not needed.

single-signal reception—Use of a piezoelectric quartz crystal and associated coupling circuits as a crystal filter, to provide the high degree of selectivity required for reception in a crowded band.

single sweep—The operating mode for a triggering-sweep oscilloscope in which the sweep must be reset for each operation, thus preventing unwanted multiple display; it is particularly useful for trace photography. In the interval after the sweep is reset and before it is triggered, the oscilloscope is said to be armed.

single throw — A contact arrangement in which each contact form included is a single contact pair.

single-throw circuit breaker — A circuit breaker in which only one set of contacts need be moved to open or close the circuit.

single-throw switch—A switch in which only one set of contacts need be moved to open or close the circuit.

single-tone keying—Keying in which the carrier is modulated with a single tone for one condition, which may be either marking or spacing, but not for the other.

single-track magnetic system—A magnetic-recording system, the medium of which has only one track.

single-track recorder — A tape recorder which records or plays only one track at a time on or from the tape. *Also see* Monaural Recorder.

single-trip multivibrator — *See* Single-Shot Multivibrator.

single-trip trigger circuit—*See* Single-Shot Trigger Circuit.

single-tuned amplifier—An amplifier characterized by resonance at a single frequency.

single-tuned circuit—A circuit which may be represented by a single inductance and capacitance, together with associated resistances.

single-turn potentiometer—A potentiometer in which the slider travels the complete length of the resistive element with only one revolution of the shaft.

single-unit semiconductor device—A semiconductor device having one set of electrodes associated with a single carrier stream.

single-wire line—A transmission line that uses the ground as one side of the circuit.

single-wound resistor—A resistor in which only one layer of resistance wire or ribbon is wound around the base or core.

sink—In communication practice: 1. A device which drains off energy from a system. 2. A place where energy from several sources is collected or drained away.

sins—Acronym for a Ship's Inertial marine Navigational System especially applicable to submarine use.

sintered plate—A powder which holds the active plate material used for both the anode and cathode in secondary cells. This provides a large surface area for the active material, allowing better cycle life, higher discharge rates, and better efficiency than the pocket-type plate design.

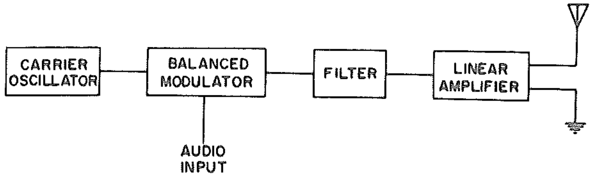

Single-sideband transmitter.

sinusoid — A curve having ordinates proportional to the sine of the abscissa.

sinusoidal—Varying in proportion to the sine of an angle or time function (e.g., ordinary alternating current).

sinusoidal current—*See* Simple Harmonic Current.

sinusoidal electromagnetic wave — In a homogeneous medium, a wave with an electric field strength proportionate to the sine (or cosine) of an angle that is a linear function of time, distance, or both.

sinusoidal field—A field in which the magnitude of the quantity at any point varies as the sine or cosine or an independent variable such as time, displacement, or temperature.

sinusoidal quantity—A quantity that varies in the manner of a sinusoid.

sinusoidal vibration—A cyclical motion in which the object moves linearly. The instantaneous position is a sinusoidal function of time.

sinusoidal wave—A wave the displacement of which varies as the sine (or cosine) of an angle that is proportional to time, distance, or both.

site error—In navigation, the error that occurs when the radiated field is distorted by objects near navigational equipment.

six-phase circuit—A combination of circuits energized by alternating electromotive forces which differ in phase by one-sixth of a cycle (60°).

size control—On a television receiver, a control which varies the size of the picture either horizontally or vertically.

skew—1. In facsimile, the nonrectangular received frame due to asynchronism between the scanner and recorder. Numerically it is the tangent of the angle of this deviation. 2. The motion characterized on a magnetic tape by an angular velocity between the gap center line and a line perpendicular to the tape center line. 3. In magnetic thin film, the deviation of the easy axis during fabrication.

skewed distribution—A frequency distribution of any natural phenomenon in which zero or infinity is one of its limits.

skewness—A statistical measure of the asymmetry existing in a distribution.

skiatron—1. A dark-trace oscilloscope tube. (*Also see* Darktrace Tube.) 2. A display employing an optical system with a dark-trace tube.

skin depth—In a current-carrying conductor, the depth below the surface at which the current density has decreased one neper below the current density at the surface.

skin effect—Also called radio-frequency resistance. The tendency of r-f currents to flow near the surface of a conductor. Thus they are restricted to a small part of the total sectional area, which has the effect of increasing the resistance.

skinner—A wire brought out at the end of a cable prepared for soldering to a terminal.

skinning—Peeling the insulation from a wire.

skip — 1. A digital-computer instruction to proceed to the next instruction. 2. In a computer, a "blank" instruction.

skip distance — The distance separating two points on the earth between which radio waves are transmitted by reflection from the ionized layers of the ionosphere.

skip fading—Fading due to fluctuations of ionization density at the place in the ionosphere where the wave is reflected, which causes the skip distance to increase or decrease.

skip keying—The reduction of the radar pulse-repetition frequency to a submultiple of that normally used, to reduce the mutual interference between radars or to increase the length of the radar time base.

skip zone—Also called zone of silence. A ring-shaped space or region within the transmission range wherein signals from a transmitter are not received. It is the distance between the farthest point reached by the ground wave and nearest point at which the refracted sky waves come back to earth.

sky error—*See* Ionospheric Error.

sky hook—Amateur term for antenna.

sky noise—Noise produced by radio energy from stars.

sky wave—*See* Ionospheric Wave *and* Indirect Wave.

sky-wave correction—In navigation, a correction for sky-wave propagation errors applied to measured positional data. The amount of the correction is established on the basis of an assumed position and on the height of the ionosphere.

sky-wave station error—In sky-wave-synchronized loran, the station-synchronization error due to the effect of the ionosphere on the synchronizing signal transmitted from one station to the other.

sky-wave-synchronization loran — A loran system in which the range is extended by using ionosphere-reflected signals for synchronizing the two ground stations.

sky-wave transmission delay—The longer time taken by a transmitted pulse when carried by sky waves reflected once from the E-layer, compared with the same pulse carried by ground waves.

slab—A relatively thick crystal from which blanks are cut.

slant range—In radar, the line-of-sight distance from the measuring point to the target, particularly an aerial target.

slave antenna—A directional antenna that is positioned in azimuth and elevation by a servo system. The information controlling the servo system is supplied by a tracking or positioning system.

slave drive—*See* Follower Drive.

slaved tracking—A method of interconnecting two or more regulated power supplies so that the master supply operates to control other power supplies called slaves.

slave relay—*See* Auxiliary Relay.

slave station—A radionavigational station, the emissions of which are controlled by a master station.

slave sweep—A time base which is synchronized or triggered by a waveform from an external source. It is used in navigational systems for displaying or utilizing the same information at different locations, or in displaying or utilizing different information with a common or related time base.

slaving—The use of a torque to maintain the orientation of the spin axis of a gyro relative to an external reference such as a pendulum or magnetic compass.

sleeping sickness — In transistors, the gradual appearance of leakage.

sleeve—1. A cylindrical contacting part usually placed in back of the tip or ring of a plug and insulated from it. 2. An iron core (usually a thin-walled cylinder) used as an electromagnetic shield around an inductor.

sleeve antenna—A vertical half-wave antenna the lower half of which is a metallic sleeve through which the concentric feed line runs. The upper radiating portion, which is one-quarter wavelength, is connected to the center of the line.

sleeve-dipole antenna—A dipole antenna with a coaxial sleeve around the center.

sleeve-stub antenna—An antenna consisting of half of a sleeve-dipole antenna projecting from an extended conducting surface.

sleeve wire—1. A third conductor when associated with a pair. 2. Wire which connects to the sleeve of a plug or jack. By extension, it is common practice to designate by this term the conductors having similar functions or arrangements in circuits where plugs or jacks may not be involved.

slewing—Rapid change of a mechanism associated with either end of a data-transmission system when it stops following one target and takes up another.

slewing rate—Also called rate limit or voltage velocity limit. The rate at which an output can be driven from limit to limit over the dynamic range. The time-rate-of-voltage change.

slew range—The high-speed range in which a motor can run continuously but cannot stop, start, or reverse without losing step count.

slice—An alternate term for "wafer."

slicer—Also called an amplitude gate or a clipper-limiter. A transducer which transmits only portions of an input wave lying between two amplitude boundaries.

slicked switch — An alacritized mercury switch in which the rolling surface has been treated with an oily material.

slideback—The technique of applying a d-c voltage to one input of a differential amplifier in order to change the vertical position on the crt screen of the signal applied to the other input.

slideback voltmeter—A vacuum-tube voltmeter that measures effective voltage values indirectly by measuring the change in grid-bias voltage required to restore the plate current of the vacuum tube to the value it had before the unknown voltage was applied to the grid circuit.

slider—A sliding contact.

slide-rule dial—A tuning dial in which a pointer moves in a straight line over a straight scale. So called because it resembles a slide rule.

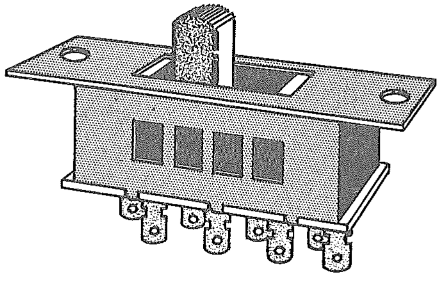

Slide switch.

slide switch—A switch which is positioned by sliding it back and forth.

slide wire—A bare resistance wire and a slider that can be set anywhere along the wire to provide a continuously variable resistance.

slide-wire bridge—A simplified Wheatstone bridge in which the resistance ratio is determined by the position of a slider on a resistance wire.

slide-wire rheostat—A long single-layer coil of resistance wire with a sliding contact. The resistance is varied by moving the slider.

sliding contact—*See* Wiping Contact.

slip—1. In facsimile, a distorted image similar to skew but caused by slippage in the mechanical drive system. 2. The difference between the synchronous and operating speed of a motor.

slip clutch—A protective device used in gear trains to disengage the load when it exceeds a specified value.

slip ring—A device for making electrical connections between stationary and rotating contacts.

slip rings—*See* Collector Rings.

slope — The essentially linear portion of the grid-voltage, plate-current characteristic curve of a vacuum tube. This is where the operating point is chosen when linear amplification is desired.

slope-based linearity—A manner of expressing nonlinearity as the deviation from a straight line for which only the slope is specified.

slope detection—A discriminator operation on one of the slopes of the response curve for a tuned circuit. It is rarely used in f-m receivers because the linear portion of the response curve is too narrow for large-signal operation.

slope detector—A detector in which slope detection is employed.

slot—One of the grooves formed in the iron core of a motor or generator armature for the conductors forming the armature winding.

slot antenna—A radiating element formed by a slot in a conducting surface.

slot armor — An insulator in the slot of a magnetic core of a machine; it may be on the coil or separate from it.

slot cell—A formed sheet of insulation that is separate from the coil and placed in the slot of a magnetic core.

slot coupling—A method of transferring energy between a coaxial cable and a waveguide by means of two coincident narrow slots, one in the sheath of the guide and the other in the sheath of the coaxial cable. E- or H-waves are launched into the guide, depending on whether the cable and guide are parallel or perpendicular to each other.

slot-discharge resistance — *See* Corona Resistance.

slot effect—The minimum voltage of rated frequency applied to the control-voltage winding of a motor tach generator necessary to start the rotor turning at no-load conditions with rated voltage and frequency on the fixed-voltage winding.

slot insulation—Flexible sheet-type insulation inserted into the slots of armatures and stators to insulate the windings from the core.

slot radiator—A primary radiating element in the form of a slot cut in the walls of a metal waveguide or cavity resonator or in a metal plate.

slotted line—*See* Slotted Section.

slotted rotor plate — *See* Serrated Rotor Plate.

slotted section—Also called a slotted line or slotted waveguide. A section of a waveguide or shielded transmission line, the shield of which is slotted to permit examination of the standing waves with a traveling probe.

slotted swr measuring equipment—A device in which standing and/or reflected waves are measured with a slotted line and a detecting probe.

sloted waveguide—*See* Slotted Section.

slow-acting relay—*See* Slow-Operating Relay.

slow-action relay—*See* Time-Delay Relay.

slow death—The gradual change of transistor characteristics with time. This change is attributed to ions which collect on the surface of the transistor.

slowed-down video—A technique of transmitting radar data over narrow-bandwidth circuits. The radar video is stored over the time required for the antenna to move through one beamwidth, and is subsequently sampled at such a rate that all range intervals of interest are sampled at least once each beam width or once per azimuth quantum.

slow memory—A computer memory, or storage, with a relatively long access time.

slow-operate, fast-release relay—A relay designed specifically for a long make and short release time.

slow-operate, slow-release relay—A slow-speed relay designed specifically for both a long make and a long release time.

slow-operating relay—Also called a slow-acting relay. One which is slow to attract its armature after its winding is energized.

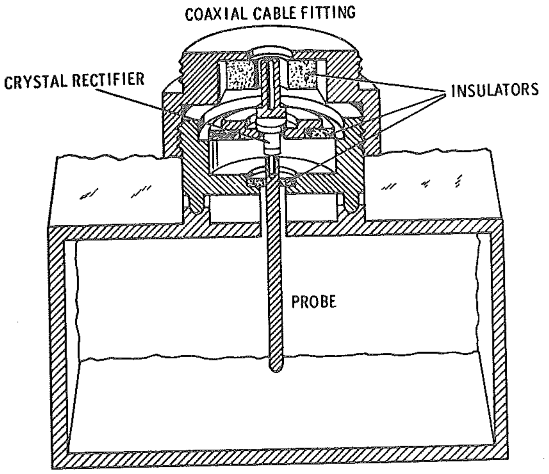

Slotted section.

A copper slug, or collar, at the armature end of the core delays the operation momentarily after the operating circuit is completed. Such a relay is often marked SO on circuit diagrams.

slow-release relay—*See* Slow-Releasing Relay.

slow-releasing relay—Also called a slow-release relay. A slow-acting relay in which a copper slug, or collar, at the heelpiece end of the core delays the restoration momentarily after the operating circuit is opened. Such a relay is often marked SR on circuit diagrams.

slow-speed relay—A relay designed specifically for long operate or release time, or both.

slow-wave circuit—A microwave circuit in which the phase velocity of the waves is considerably below the speed of light. Such waves are used in traveling-wave tubes.

slow-wave structure—A circuit composed of selected inductance and capacitance that causes a wave to be propagated at a speed slower than the speed of light.

slug—1. A heavy metal ring or short-circuited winding used on a relay core to delay operation of the relay. 2. A metallic core which can be moved along the axis of a coil for tuning purposes.

slug tuner—A waveguide tuner containing one or more longitudinally adjustable pieces of metal or dielectric.

slug tuning — Varying the frequency of a resonant circuit by introducing a slug of material into the electric or magnetic fields, or both.

small signal—That value of an a-c voltage or current which, when halved or doubled, will not affect the characteristic being measured beyond the normal accuracy of the measurement of that characteristic.

small-signal analysis—Consideration of only small excursions from the no-signal bias, so that a vacuum tube or transistor can be represented by a linear equivalent circuit.

small-signal characteristics—The characteristics of an amplifier operating in the linear amplification region.

small-signal current gain (current-transfer ratio)—The output current of a transistor with the output circuit shorted, divided by the input current. The current components are understood to be small enough that linear relationships hold between them.

small-signal, open-circuit forward-transfer impedance—In a transistor, the ratio of the a-c output voltage to the a-c input current when the a-c output current is zero.

small-signal, open-circuit input impedance—In a transistor, the ratio of the a-c input voltage to the a-c input current when the a-c output current is zero.

small-signal, open-circuit output admittance—In a transistor, the ratio of the a-c output current to the a-c voltage applied to the output terminals when the a-c input current is zero.

small-signal, open-circuit output impedance—In a transistor, the ratio of the a-c voltage applied to the output terminals to the a-c output current when the a-c input current is zero.

small-signal, open-circuit reverse-transfer impedance—In a transistor, the ratio of the a-c input voltage to the a-c output current when the a-c input current is zero.

small-signal, open-circuit, reverse-voltage transfer ratio—In a transistor, the ratio of the a-c input voltage to the a-c output voltage when the a-c input current is zero.

small-signal power gain — In a transistor, the ratio of the a-c output power to the a-c input power under specified small-signal conditions. Usually expressed in db.

small-signal, short-circuit, forward-current transfer ratio — In a transistor, the ratio of the a-c output current to the a-c input current when the a-c output voltage is zero.

small-signal, short-circuit forward-transfer admittance—In a transistor, the ratio of the a-c output current to the a-c input voltage when the a-c output voltage is zero.

small-signal, short-circuit input admittance—In a transistor, the ratio of the a-c input current to the a-c input voltage when the a-c output voltage is zero.

small-signal, short-circuit input impedance — In a transistor, the ratio of the a-c input voltage to the a-c input current when the a-c output voltage is zero.

small-signal, short-circuit output admittance—In a transistor, the ratio of the a-c output current to the a-c output voltage when the a-c input voltage is zero.

small-signal, short-circuit reverse-transfer admittance—In a transistor, the ratio of the a-c input current to the a-c output voltage when the a-c input voltage is zero.

small-signal transconductance—In a transistor, the ratio of the a-c output current to the a-c input voltage when the a-c output voltage is zero.

smear—Television-picture distortion in which objects appear stretched out horizontally and are blurred.

S-meter—A built-in meter used in some communications receivers to indicate the strength of the received signal. *See* Signal-Strength Meter.

Smith chart—1. A special polar coordinate diagram designed to solve problems concerning transmission lines. It was developed by P. H. Smith. 2. A special graph laid out in such a way that transmission-line impedances can conveniently be found on it.

smoothing choke—An iron-core choke coil that filters out fluctuations in the output current of a vacuum-tube rectifier or direct-current generator.

smoothing circuit—A combination of inductance and capacitance employed as a filter circuit to remove fluctuations in the output current of a vacuum-tube or semiconductor rectifier or direct-current generator.

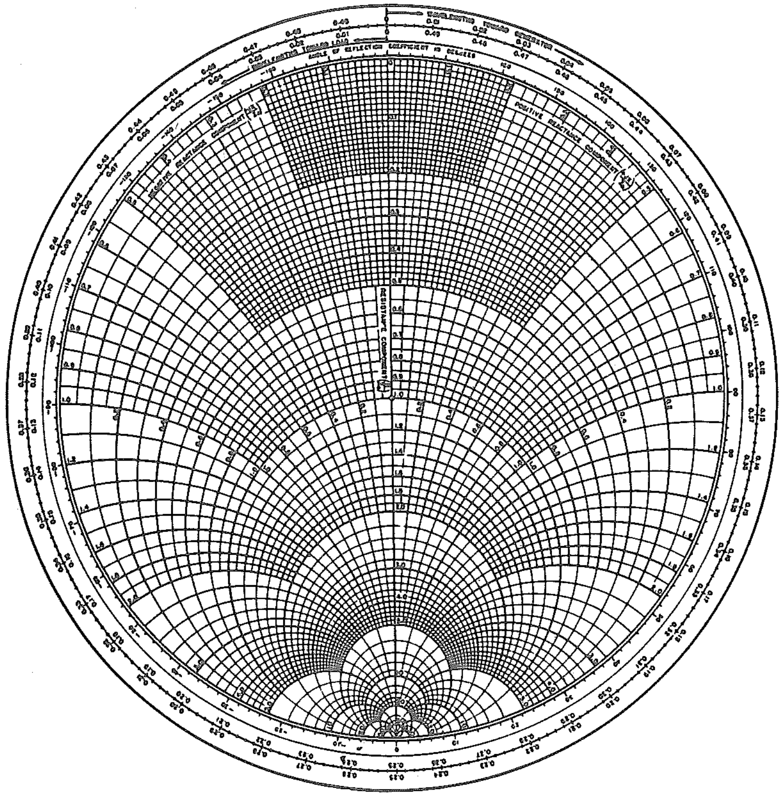

Smith chart.

smoothing factor—The factor expressing the effectiveness of a filter in smoothing out ripple voltages.

smoothing filter—1. A filter used to remove fluctuations in the output current of a vacuum-tube or semiconductor rectifier or direct-current generator. 2. A low-pass filter in the vertical-deflection amplifier of a spectrum analyzer. It is used to smooth amplitude fluctuation, in order to display spectral density and the average level of random signals as single lines. (The time constant of a smoothing filter is generally variable.)

SMPTE — Abbreviation for the Society of Motion Picture and Television Engineers.

snake—A tempered steel wire, usually of rectangular cross-section. The snake is pushed through a run of conduit or through an inaccessible space such as a partition and used for drawing in wires.

snap-action—1. In a mercury switch, the rapid motion of the mercury pool from one position to another. 2. A rapid motion of the contacts from one position to another position, or their return. This action is relatively independent of the rate of travel of the actuator.

snap-action contacts — A contact assembly such that the contacts remain in one of two positions of equilibrium with substantially constant contact pressure during the initial motion of the actuating member until a point is reached at which stored energy causes the contacts to move abruptly to a new position of equilibrium.

snap magnet—A permanent magnet used in thermostatic, pressure, and other control instruments to provide quick make-and-break action at the contact and thereby minimize sparking. The magnet pulls the armature in suddenly against the spring to close the contacts and hold them closed until the spring is compressed enough to make them fly apart.

snap-off diode — A modified very high-speed, planar, epitaxial, passivated diode in which the charge stored during conduction of the diode is held close to the junction.

snapshot—In a computer, a dynamic printout of selected data in storage that occurs at breakpoints and checkpoints during the computing operations as opposed to a static printout.

snap switch—A switch (e.g., a light switch) in which the contacts are separated or brought together suddenly as the operating knob or lever compresses or releases a spring.

sneak circuit—That part of a complete electrical circuit which carries an unintentional (sneak) current. Sneak currents may prevent proper operation of interconnected equipment.

Snell's law—The sine of the angle of incidence, divided by the sine of the angle of refraction, equals a constant called the index of refraction when one of the mediums is air.

snivet—A straight, jagged, or broken vertical black line that appears near the right edge of the screen of a television receiver.

snooperscope—A night viewing device to permit a user to see objects in total darkness. It consists of an infrared source, an infrared image converter, and a battery-operated high-voltage d-c–to–d-c converter. Infrared radiation sent out is reflected back to the snooperscope, where it is then converted into a visible image on the fluorescent screen of the image converter tube.

snow—1. A speckled background caused by random noise on an intensity-modulated display. 2. White specks in a television picture (usually indicative of a weak signal).

snowflake transistor — A high-speed medium-power switching transistor for use as a thin-film or core driver. It employs a six-pointed emitter geometry that permits the optimum ratio of emitter periphery to emitter area.

snr—Abbreviation for signal-to-noise ratio.

s/n ratio—Abbreviation for signal-to-noise ratio.

soak—In an electromagnetic relay, the condition that exists when the core is approximately saturated.

soakage—The disability of a capacitor to come up to voltage instantaneously without voltage lag or creep during or after charging. The lower the soakage, the lower the lag and creep.

soak time—The period of time required following activation for the electrolyte in a cell or battery to be sufficiently absorbed into the active materials.

soak timer—A reset timer, usually dial-adjustable, as applied in a temperature-control system for controlling the length of time the temperature is held at a predetermined level.

soak value—The voltage, current, or power applied to the coil of the relay coil to insure that a condition approximating magnetic saturation exists.

socket—An opening that supports and electrically connects to vacuum tubes, bulbs or other devices or components when they are inserted into it.

socket adapter—A device placed between a tube and its socket so that the tube can be used in a socket designed for some other base, or so that current or voltage can be measured at the electrodes while the tube is in use.

socket contact—A hollow female contact designed to mate with a male contact. It is normally connected to the "live" side of a circuit.

sodar — Acronym for SOund Detecting And Ranging. A device which detects large changes in temperature overhead by the amount of sound returned as echoes (the colder the atmosphere, the louder the echoes). The sound, which is within the range of human hearing, is launched upward and the echoes are changed into oscilloscope patterns.

sodium amalgam-oxygen cell—A fuel-cell system in which materials functioning in the dual capacity of fuel and anode are consumed continuously. Low operating temperatures and high power-to-weight ratios are significant characteristics of the system.

sodium-vapor lamp—A gas-discharge lamp containing sodium vapor. It is used chiefly for highway illumination.

sofar—Acronym for SOund Fixing And Ranging. An underwater sound system with which air and ship survivors can be located within a square mile and as far as 2,000 miles away. Survivors drop a TNT charge into the water. The charge, which is timed to explode at 3,000 to 4,000 feet, sets up underwater sound waves that can be picked up by hydrophones at shore stations.

soft magnetic material—Also called a low-energy material. Ferromagnetic material which, once having been magnetized, is very easily demagnetized (i.e., requires only a slight coercive force to remove the resultant magnetism).

soft phototube—A gas phototube.

soft tube—1. A high-vacuum tube which has become defective because of the entry of a small amount of gas. 2. An electronic tube into which a small amount of gas has purposely been put to obtain the desired characteristics. 3. Also called gassy tube.

software — A program package available for work on general-purpose, digital-computer hardware.

soft X-rays—X-rays with comparatively long wavelengths and hence poor penetrating power.

solar absorber—A surface that has the property of converting solar radiation into thermal energy.

solar cell—A silicon photovoltaic cell that converts solar energy directly into electrical energy.

solar concentrator—A device that increases the intensity of solar energy by optical means.

solar-energy conversion — The process of changing solar radiation into electrical or mechanical power, either directly or by using a heat engine.

solar noise—Electromagnetic radiation from the sun at radio frequencies.

solder—A readily meltable metal or alloy that produces a bond at a junction of two metal surfaces. True solder must have a lower melting point than the metals being joined and must also be capable of uniting with the metals to be joined.

solderability—1. In a printed-circuit board, a measure of the ability of the conductive pattern to be wet by solder. 2. The property of a metal surface which allows it to be wetted by solder.

solder eye—A solder-type terminal provided with a hole at its end through which a wire can be inserted prior to being soldered. A ring-shaped contact termination of a printed-circuit connector for the same purpose.

solder ground—A conducting path to ground due to dripping or overhanging solder.

soldering—The joining of metallic surfaces (e.g., electrical contacts) by melting a metal or an alloy (usually tin and lead) over them.

soldering iron—A soldering tool consisting of a heating element to heat the tip and melt the solder, plus a heat-insulated handle.

solderless connector—A device for clamping two wires firmly together to provide a good connection without solder. A common form is a cap with tapered internal threads, which are twisted over the exposed ends of the wires.

solderless wrap—Also called wire wrap. A method of connection in which a solid wire is tightly wrapped around a rectangular, square, or V-shaped terminal by means of a special tool.

solder short — A defect which occurs when solder forms a short-circuit path between two or more conductors.

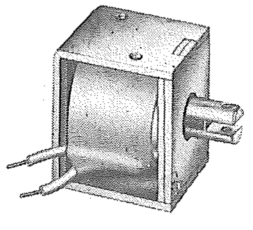

Solenoid.

solenoid—1. A coil of wire which, when a current flows through it, will act as a magnet and tend to pull a movable iron core to a central position. 2. An electric conductor wound as a spiral with a small pitch, or as two or more coaxial spirals. 3. An electromagnet having an energized coil approximately cylindrical in form and an armature the motion of which is reciprocating within and along the axis of the coil.

solenoid valve—A combination of an electromagnet plunger or an orifice to which a disc or plug can be positioned to either restrict or completely shut off a flow. (Orifice closure or restriction occurs when the electromagnet actuates a magnet plunger.)

solid—A state of matter in which the motion of the molecules is restricted. They tend to remain in one position, giving rise to a crystal structure. Unlike a liquid or gas, a solid has a definite shape and volume.

solid circuit—A semiconductor network fabricated in one piece of material by alloying, diffusing, doping, etching, cutting, and the use of necessary jumper wires.

solid conductor—A wire—i.e., a conductor composed of a single strand.

solid-electrolyte fuel cell—A self-contained fuel cell in which oxygen is the oxidant and hydrogen is the fuel. The oxidant and fuel are kept separated by a solid electrolyte which has a crystalline structure and a low conductivity.

solid-electrolyte tantalum capacitor—Also called solid tantalum capacitor. A tantalum capacitor with a solid semiconductor electrolyte instead of a liquid. A wire anode is used for low capacitance values and a sintered pellet for higher values.

solidly grounded — Also called directly grounded. Grounded through an adequate grounded connection in which no impedance has been inserted intentionally.

solid-state—1. Pertaining to circuits and components using semiconductors. 2. The physics of materials in their solid form. Examples of solid-state materials are: transistors, diodes, solid-state lasers, metals and alloys, etc.

solid-state atomic battery — A device in which a radioactive material and a solar cell are combined. The radioactive material emits particles that enter the solar cell, which in turn produces electrical energy.

solid-state circuit — A complete circuit formed from a single block of semiconductor material. *Also see* Monolithic Integrated Circuit.

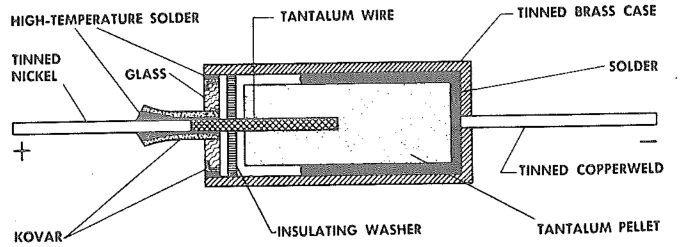

Solid-electrolyte tantalum capacitor.

solid-state component — A component the operation of which depends on the control of electric or magnetic phenomena in solids (e.g., a transistor, crystal diode, or ferrite).

solid-state computer — A computer built primarily from solid-state electronic circuit elements.

solid-state device—Any element that can control current without moving parts, heated filaments, or vacuum gaps. All semiconductors are solid-state devices, although not all solid-state devices (e.g., transformers) are semiconductors.

solid-state integrated circuit—The class of integrated componets in which only solid-state materials are used.

solid-state physics—The branch of physics that deals with the structure and properties of solids, including semiconductors (i.e., a material the electrical resistivity of which is between that of insulators and conductors). Generally used semiconductors are silicon and germanium.

solid-state relay — A relay that employs solid-state semiconductor devices as components.

solid tantalum capacitor—*See* Solid-Electrolyte Tantalum Capacitor.

solion—Contraction of SOLution ION. An electrochemical sensing and control device in which ions in solution carry electric charges to give amplification corresponding to that of vacuum tubes and transistors.

solion integrator—A precision electrochemical cell housed in glass and containing four small platinum electrodes in a solution of potassium iodide and iodine. The integrator anode and cathode make up the covers of a small cylindrical volume (less than 0.00025 cubic inch) for storing electrical information in the form of ions. The integrator cathode contains a fixed amount of hydraulic porosity for completing the internal-solution path to the other two electrodes.

solo manual—*See* Swell Manual.

Sommerfeld formula—An approximate wave-propagation relationship that may be used when distances are short enough that the curvature of the earth may be neglected in the computations.

Sonalert—A solid-state tone-emitting device.

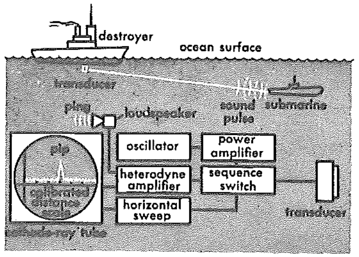

Sonar.

sonar—Acronym for SOund Navigation And Ranging. Also called active sonar if it radiates underwater acoustic energy, or passive (listening) sonar if it merely receives the energy generated from a distant source. Apparatus or technique of obtaining information regarding objects or events underwater through the transmission and reception of acoustic energy. Two well-known uses are to detect submarines and fish.

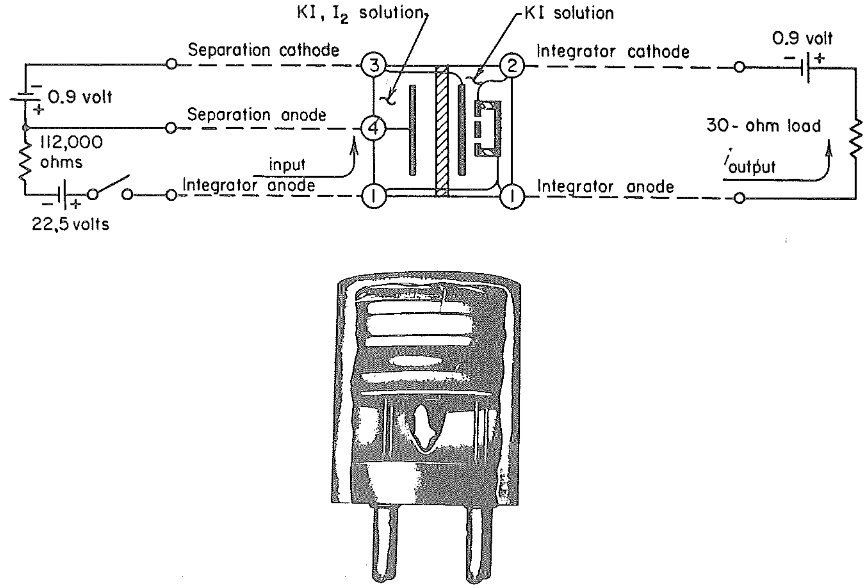

Solion integrator.

sonar background noise—In sonar, the total noise, presented to the final receiving element, that interferes with the reception of the desired signal.

sone—A unit of loudness. A simple tone of frequency 1,000 hertz, 40 db above a listener's threshold, produces a loudness of 1 sone. The loudness of any sound that is judged by the listener to be *n* times that of the 1-sone tone is *n* sones.

sonic—1. Pertaining to the speed of sound. 2. Utilizing sound waves.

sonic altimeter—An altimeter that determines the height of an aircraft above the earth by measuring the time the sound waves take to travel from the aircraft to the ground and back, based on the fact that the velocity of sound at sea level is 1,080 feet per second through dry air at 0°C. (32°F.).

sonic applicator—A self-contained electromechanical transducer for local application of sound for therapeutic purposes.

sonic boom—An explosionlike sound heard and felt when the shock wave that is generated by aircraft flying at supersonic speed reaches the ground.

sonic cleaning—The cleaning of contaminated materials by the action of intense sound waves produced in the liquid into which the material is immersed.

sonic delay line — A device in which electroacoustic transducers and the propagation of an elastic wave through a medium are used to produce the delay of an electrical signal.

sonic depth sounder—*See* Fathometer.

sonic drilling—The cutting or shaping of materials with an abrasive slurry driven by a reciprocating tool attached to an electromechanical transducer.

sonic frequencies—Vibrations which can be heard by the human ear (from about 15 hertz to approximately 20,000 hertz).

sonic soldering — The method of joining metals by the use of mechanical vibration to break up the surface oxides.

sonic speed—*See* Speed of Sound.

sonic thermocouple—A thermocouple so designed that gas moves past the junction with a velocity of mach 1 or greater, resulting in maximum heat transfer to the junction.

sonne — Also called consol. A radionavigational aid that provides a number of rotating characteristic signal zones. A bearing may be determined by observation (and interpolation) of the instant when transition occurs from one zone to the following zone.

sonobuoy—Also called a radiosonobuoy. A device used to locate a submerged target (e.g., a submarine). By means of a hydrophone system in the water, a sonobuoy detects the noises and converts them into radio signals, which are transmitted to a receiver in an airplane. Each sonobuoy transmits on one of several possible frequencies, and the receiver in the airplane has a channel selector so the operator can switch from one to another.

sonoluminescence—The creation of light in liquids by sonically-induced cavitation.

sonometer—A frequency meter that depends for its operation on mechanical resonance with the vibrations of a variable length of stretched wire.

sonoptography—The use of sound waves to obtain a 3-D image of an object. No lenses are required. In essence, a two stage process in which the diffraction pattern of an object irradiated by sound waves is biased by a coherent sound wave and recorded. The resultant pattern, like the hologram, then is interrogated with a suitable coherent light source to obtain a three-dimensional image.

sophisticated—A piece of equipment, system, etc., which is complex and intricate, or requires special skills to operate.

sophisticated vocabulary—An advanced and elaborate set of computer instructions, enabling the computer to perform such intricate operations as linearizing, extracting square roots, selecting the highest number, etc.

sort—To arrange items of information according to rules which depend on a key or field contained by the items.

sorter—A machine which sorts cards according to the position of coded holes.

SOS — A distress signal used in radiotelegraphy.

sound—1. Also called a sound wave. An alteration in pressure, stress, particle displacement or velocity, etc., propagated in an elastic material, or the superposition of such propagated alterations. 2. Also called a sound sensation. The auditory sensation usually evoked by the alterations described in (1) above.

sound absorption—The conversion of sound energy into some other form (usually heat) in passing through a medium or on striking a surface.

sound-absorption coefficient—The incident sound energy absorbed by a surface or medium, expressed in the form of a fraction.

sound analyzer—A device for measuring the amplitude and frequency of the components of a complex sound. It usually consists of a microphone, an amplifier, and a wave analyzer.

sound articulation—The per cent of articulation obtained when the speech units are fundamental sounds (usually combined into meaningless syllables).

sound bars—Alternate dark and light horizontal bars caused in a television picture by a-f voltage reaching the video-input circuit of the picture tube.

sound carrier—The frequency-modulated carrier which transmits the sound portion of television programs.

sound concentrator — A parabolic reflector used with a microphone at its focus to obtain a highly directive pickup response.

sound-effect filter—A filter, usually adjustable, designed to reduce the passband of a system at low and/or high frequencies in order to produce special effects.

sound energy—The total energy in a given part of a medium, minus the energy which would exist there if no sound waves were present.

sound-energy density—At a point in a sound field, the sound energy contained in a given infinitesimal part of the medium, divided by the volume there. The commonly used unit is the erg per cubic centimeter.

sound-energy flux — The average rate at which sound energy flows through any specified area for a given period. The commonly used unit is the erg per second.

sound-energy flux density—*See* Sound Intensity.

sounder—*See* Telegraph Sounder.

sound field—A region in any medium containing sound waves.

sound film—Motion-picture film having a sound track along one side of the picture frames, for simultaneous reproduction of the sounds that accompany the film. A beam of light is projected through the sound track and is modulated at an audio rate by the variations in the width or density of the track. A phototube and amplifier then convert these modulations into sound.

sound gate—A mechanical device through which film is passed in a projector, to convert the sound track into audio signals that can be amplified and reproduced. In a television camera used for pickups, a sound gate provides the sound accompaniment for the motion picture being televised. Associated with the sound gate are an exciter lamp, a lens assembly, and a phototube.

sound head—The part of a sound motion-picture projector which converts the photographic or magnetic sound track on the film into audible sound signals.

sounding—Determination of the depth of water or the altitude above the earth.

sound intensity—Also called specific sound-energy flux or sound-energy flux density. The average rate of sound energy transmitted in a specified direction through a unit area normal to this direction at the point considered. The common unit is the erg per second per square centimeter, although sound intensity expressed in watts per square centimeter may occasionally be used.

sound-level meter—An instrument—including a microphone, amplifier, output meter, and frequency-weighting networks—for the measurement of noise and sound levels. The measurements approximate the loudness level obtained for pure tones by the more elaborate ear-balance method.

sound-on-sound recording — A method by which material previously recorded on one track of a tape may be recorded on another track while simultaneously adding new material to it.

sound-powered telephone set—A telephone set in which the transmitter and receiver are passive transducers; operating power is obtained from the speech input only.

sound-power level—The ratio, expressed in db, of the sound power emitted by a source to a standard reference power of 10^{-13} watts.

sound power of a source—The total sound energy radiated by the source per unit of time. The common unit is the erg per second, but the power may also be expressed in watts.

sound pressure—The instantaneous pressure minus the static pressure at some point in a medium when a sound wave is present.

sound-pressure level — Abbreviated spl. In decibels, 20 times the logarithm of the ratio of the pressure of a sound to the reference pressure, which must be explicitly stated (usually, either 2×10^{-4} or 1 dyne per square centimeter).

sound probe—A small microphone (or tube added to a conventional microphone) for exploring a sound field without significantly disturbing it.

sound recordings—Records, tapes, or other sonic components upon which audio intelligence is inscribed or recorded or can be reproduced.

sound-recording system—A combination of transducing devices and associated equipment for storing sound in a reproducible form.

sound-reflection coefficient — Also called acoustical reflectivity. Ratio at which the sound energy reflected from a surface flows on the side of incidence, to the incident rate of flow.

sound-reproducing system—A combination of transducers and associated equipment for reproducing prerecorded sound.

sound sensation—*See* Sound.

sound spectrum—The frequency components included within the range of audible sound.

sound takeoff—The connection or coupling at which the 4.5-MHz frequency-modulated sound signal in a television receiver is obtained.

sound track—The narrow band which carries the sound in a movie film. It is usually along the margin of the film, and more than one band may be used (e.g., for stereophonic sound).

sound-transmission coefficient (of an interface or septum)—Also called acoustical transmittivity. The ratio of the transmitted to the incident sound energy. Its value is a function of the angle of incidence of the sound.

sound wave—*See* Sound.

source—1. The device which supplies signal power to a transducer. 2. In a field-effect transistor, the electrode that corresponds to the cathode of a vacuum tube.

source data automation—The methods of recording information in coded forms on paper tapes, punched cards, or tags that can be used repeatedly to produce many other records without rewriting.

source impedance—The impedance which a source of energy presents to the input terminals of a device.

source language—The language used to prepare a problem as the input for a computer operation.

source program — A computer program written in a language designed for ease of expression of a class of problems or procedures by humans. A generator, assembler, translator, or compiler routine translates the source program into an object program in machine language.

sourcing—Redesign or modification of existing equipment to eliminate a source of emi. When sourcing is not feasible, engineers are forced to resort to suppression, filtering, or shielding.

south pole—In a magnet, the pole into which magnetic lines of force are assumed to enter after emerging from the north pole.

space—1. An impulse which, in a neutral circuit, causes the loop to open or causes the absence of a signal, while in a polar circuit it causes the loop current to flow in a direction opposite to that for a mark impulse. A space impulse is equivalent to a binary 0. 2. In some codes, a character which causes a printer to leave a character width with no printed symbol.

space attenuation—The loss of energy, expressed in decibels, of a signal in free air caused by such factors as absorption, reflection, scattering, and dispersion.

space charge—1. The electrical charge caused in space by the presence of electrons or ions. 2. The electron cloud around the hot cathode of a vacuum tube.

space-charge debunching—In a microwave tube, a process in which the bunched electrons are dispersed due to the mutual interactions between electrons in the stream.

space-charge effect—Repulsion of electrons emitted from the cathode of a thermionic vacuum tube by electrons accumulated in the space charge near the cathode.

space-charge field — The electric field that occurs inside a plasma due to the net space charge in the volume of the plasma.

space-charge grid—A grid, usually positive, that controls the position, area, and magnitude of a potential minimum, or of a virtual cathode adjacent to the grid.

space-charge region — 1. The region where the net charge density of a semiconductor differs significantly from zero. (*Also see* Depletion Layer.) 2. The region in a semiconductor in which the mobile carrier charge density is insufficient to neutralize the net charge density of the fixed donor and acceptor ions.

space-charge tube—A tube in which the space charge is used to greatly increase the transconductance. A positively charged grid is placed next to the cathode, in front of the control grid. This enlarges the space charge, moving it out to where the control grid can have a greater effect on it and hence on the plate current.

space coordinates—A three dimensional system of rectangular co-ordinates. The x and y co-ordinates lie in a reference plane tangent to the earth, and the z co-ordinate is perpendicular.

space current—The total current flowing between the cathode and all other electrodes in a tube.

spaced antenna—An antenna system used for minimizing local effects of fading at short-wave receiving stations. So called because it consists of several antennas spaced a considerable distance apart.

spaced-antenna direction finder—A direction finder comprising two or more similar but separate antennas coupled to a common receiver.

space diversity—*See* Space-Diversity Reception.

space-diversity reception—Also called space diversity. Diversity reception from receiving antennas placed in different locations.

spaced-loop direction finder—A spaced-antenna direction finder in which the individual antennas are loops.

space factor — Ratio of the effective area utilized, to the total area in a winding section.

space harmonics—Harmonics in the distribution of flux in the air gap of a resolver. They may be determined as a percent of fundamental by a Fourier analysis of the flux distribution pattern. Space harmonics cause angular inaccuracy.

space pattern—On a test chart, a pattern designed for the measurement of geometric distortion. The EIA ball chart is an example.

space permeability — The factor that expresses the ratio of magnetic induction to magnetizing force in a vacuum. In the cgs electromagnetic system of units, the permeability of a vacuum is arbitrarily taken as unity.

space phase—Reaching corresponding peak values at the same point in space.

space quadrature—The difference in the position of corresponding points of a wave in space, the points being separated by one quarter of the wavelength in question.

space-to-mark transition—The transition, or switching, from a spacing impulse to a marking impulse. (Teletypewriter term.)

space wave—The radiated energy consisting of the direct and ground waves.

spacing—The distance between microphones or speakers in stereo recording or reproduction.

spacing end distortion — End distortion which lengthens the spacing impulse by advancing the mark-to-space transition. (Teletypewriter term.)

spacing interval—The interval between successive telegraph signal pulses. During this interval, either no current flows or the current has the opposite polarity from that of the signal pulses.

spacing pulse—In teletypewriter operation, the signal interval during which the selector unit does not operate.

spacing wave—Also called back wave. In telegraphic communication, the emission which takes place between the active portions of the code characters or while no code characters are being transmitted.

spacistor—A semiconductor device consisting of one PN junction and four electrode connections. It is characterized by a low transient time for carriers to flow from the input to the output.

spade bolt—A bolt with a threaded section and one spade-shaped flat end through which there is a hole for a screw or rivet. It is used for fastening shielded coils, capacitors, and other components to the chassis.

spade contact—A contact with fork-shaped female members designed to dovetail with spade-shaped male members. Alignment in this type of connection is very critical if good conductivity is to be achieved.

spade tips—Notched, flat metal strips connected to the end of a cord or wire so that it can be fastened under a binding screw.

spaghetti — Heavily varnished cloth tubing sometimes used to provide insulation for circuit wiring.

span—1. The part or space between two consecutive points of support in a conductor, cable, suspension strand, or pole line. 2. The reach or spread between two established limits such as the difference between high and low values in a given range of physical measurements.

spark—1. The abrupt, brilliant phenomenon which characterizes a disruptive discharge. 2. A single, short electrical discharge between two electrodes.

spark capacitor — A capacitor connected across a pair of contact points, or across the inductance which causes the spark, for the purpose of diminishing sparking at these points.

spark coil—An induction coil used to produce spark discharges.

spark frequency—The total number of sparks occurring per second in a spark transmitter (not the frequency of the individual waves).

spark gap—The arrangement of two electrodes between which a disruptive discharge of electricity may occur, and such that the insulation is self-restoring after the passage of a discharge.

spark-gap modulation—Modulation in which a controlled spark-gap breakdown produces one or more pulses of energy for application to the element in which the modulation is to take place.

spark-gap modulator — A modulator employed in certain radar transmitters. A pulse-forming line is discharged across either a stationary or a rotary spark gap.

spark-gap oscillator — A type of oscillator consisting essentially of an interrupted high-voltage discharge and a resonant circuit.

sparking — Intentional or accidental spark discharge, as between the brushes and commutator of a rotating machine, between the contacts of a relay or switch, in a solid tantalum capacitor,. or at any other point at which a circuit is broken.

sparking voltage—The minimum voltage at which a spark discharge occurs between electrodes of a given shape, at a given distance apart, under given conditions.

spark killer—An electric network, usually a capacitor and resistor in series, connected across a pair of contact points (or across the inductance which causes the spark) to diminish sparking at these points.

spark lag—The interval between attainment of the sparking voltage and passage of the spark.

sparkover—Breakdown of the air between two electrical conductors, permitting the passage of a spark.

spark plate—In an automobile radio, a metal plate insulated from the chassis by a thin sheet of mica. It bypasses the noise signals picked up by the wiring under the hood.

spark-quenching device — *See* Spark Suppressor.

spark recorder—A recorder in which the recording paper passes through a spark gap formed by a metal plate underneath and a moving metal pointer above the paper. Sparks from an induction coil pass through the paper, periodically burning small holes that form the record trace.

spark spectrum—The spectrum produced in a substance when the light from a spark passes between terminals made of that substance or through an atmosphere of that substance.

spark suppressor — Also called a spark-quenching device or an arc suppressor. An electric network, such as a capacitance and resistance in series, or a diode connected across a pair of contacts to diminish sparking (arcing) at these contacts.

spark test—A test performed on wire and cable to determine the amount of detrimental porosity or defects in the insulation.

spark transmitter—A radio transmitter in which the source of radio-frequency power is the oscillatory discharge of a capacitor through an inductor and a spark gap.

spatial distribution—The directional properties of a speaker, transmitting antenna, or other radiator.

spdt — Abbreviation for single-pole, double-throw.

speaker—Abbreviated spkr. Also called a loudspeaker. An electroacoustic transducer that radiates acoustic power into the air with essentially the same waveform as that of the electrical input.

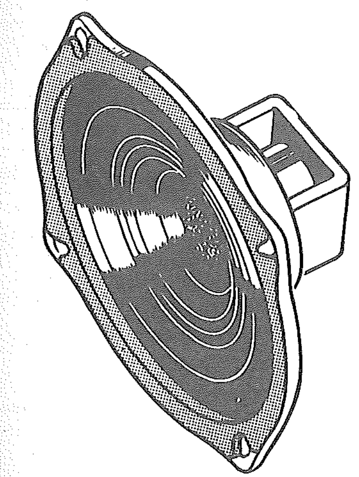

Speaker.

speaker efficiency—Ratio of the total useful sound radiated from a speaker at any frequency, to the electrical power applied to the voice coil.

speaker impedance—The rated impedance of the voice coil of a speaker.

speaker-reversal switch—A switch for connecting the left channel to the right speaker and vice versa on a stereo amplifier. It is a means of correcting for improper left-right orientation in the program source.

speaker system—A combination of one or more speakers and all associated baffles, horns, and dividing networks used to couple the driving electric circuit and the acoustic medium together.

speaker voice coil—In a moving-coil speaker, the part which is moved back and forth by electric impulses and is fastened to the cone in order to produce sound waves.

speaking arc — A d-c arc on which audio-frequency currents have been superimposed. As a result, the arc reproduces sounds in a manner similar to a speaker, and its light output will vary at the audio rate required for sound-film recording.

special-purpose computer—A computer designed to solve a restricted class of problems, as contrasted with a general-purpose computer.

special-purpose motor—A motor possessing special operating characteristics and/or special mechanical construction, designed for a particular application, and not included in the definition of a general-purpose motor.

special-purpose relay—A relay the application of which requires special features not characteristic of general-purpose or definite-purpose relays.

specific acoustic impedance — Also called unit-area acoustic impedance. The complex ratio of sound pressure to particle velocity at a point in a medium.

specific acoustic reactance—The imaginary component of the specific acoustic impedance.

specific acoustic resistance—The real component of the specific acoustic impedance.

specific coding—Digital-computer coding in which all addresses refer to specific registers and locations.

specific conductance—*See* Electrolytic Conductivity.

specific conductivity—The conducting ability of a material in mhos per cubic centimeter. It is the reciprocal of resistivity.

specific damping of an instrument — *See* Relative Damping of an Instrument.

specific dielectric strength—The dielectric strength per millimeter of thickness of an insulating material.

specific gravity—The weight of a substance compared with the weight of the same volume of water at the same temperature.

specific heat—The capacity of a material to be heated at a given temperature (expressed as calories per degree C. per gram), compared to water, which has a specific heat of 1.

specific inductive capacity — *See* Dielectric Constant.

specific program—Digital-computer programming for solving a specific problem.

specific repetition rate—In loran, one of a set of closely-spaced repetition rates derived from the basic rate and associated with a specific set of synchronized stations.

specific resistance—The resistance of a conductor. It is expressed in ohms per unit length per unit area, usually circular mil feet. (*Also see* Resistivity.)

specific routine—A digital-computer routine expressed in specific computer coding and used to solve a specific mathematical, logical, or data-handling problem.

specific sound-energy flux—*See* Sound Intensity.

spectral characteristic — The relationship between the radiant sensitivity of a phototube and the wavelength of the incident radiant flux. It is usually shown by a graph.

spectral contour plotter—A spectrum analyzer that presents a three-dimensional contour plot of analog signals, heart sounds, brain waves, etc.

spectral response—Also called spectral sensitivity characteristic. The relative amount of visual sensation produced by one unit of radiant flux of any one wavelength. The human eye or a photocell exhibits greatest spectral response to the wavelengths producing yellow-green light.

spectral sensitivity—The color response of a photosensitive device.

spectral sensitivity characteristic — *See* Spectral Response.

spectral voltage density — The rms voltage corresponding to the energy contained in a frequency band having a width of one hertz. For the spectral voltage density at a given frequency, the band is centered on the given frequency.

spectrograph—An instrument with an entrance slit and dispersing device that uses photography to obtain a record of the spectral range. The radiant power passing through the optical system is integrated over time, and the quantity recorded is a function of the radiant energy.

spectrometer—A test instrument that determines the frequency distribution of the energy generated by any source and displays all components simultaneously.

spectrophotoelectric—Pertaining to the dependence of photoelectric phenomena on the wavelength of the incident radiation.

spectroscope — An instrument used to disperse radiation into its component wavelengths and to observe or measure the resultant spectrum.

spectroscopy — The branch of optics that deals with radiations in the infrared, visible, and ultraviolet regions of the spectrum.

spectrum—A continuous range of electromagnetic radiations, from the longest known radio waves to the shortest known cosmic rays. Light, which is the visible portion of the spectrum, lies about midway between these two extremes.

spectrum analysis — The study of energy distribution across the frequency spectrum for a given electrical signal.

spectrum analyzer—A scanning receiver that automatically tunes through a selected frequency spectrum and displays on a crt or a chart a plot of amplitude versus frequency of the signals present at its input. A spectrum analyzer is, in effect, an automatic Fourier analysis plotter.

spectrum intervals—Frequency bands represented as intervals on a frequency scale.

spectrum level—For a specified signal at a particular frequency, the level of that part contained within a band one hertz wide, centered at the particular frequency.

spectrum-selectivity characteristic — A measure of the increase in the minimum input-signal power over the minimum detectable signal required to produce an indication on a radar indicator, if the received signal has a spectrum different from that of the normally received signal.

spectrum signature analysis—The evaluation of emi from transmitting and receiving equipment in order to determine operational and environmental compatibility.

spectrum utilization characteristics—The compiled data of either transmitters or receivers which describe operational parameters such as bandwidth, sensitivity, stability, antenna pattern, power output, etc. Also describes the capabilities of the equipment to either reject or suppress unwanted electromagnetic energy.

specular reflection — Reflection of light, sound, or radio waves from a surface so smooth that its inequalities are small in comparison with the wavelength of the incident rays. As a result, each incident ray produces a reflected ray in the same plane.

SPEDAC—Acronym for Solid-state, Parallel, Expandable, Differential-Analyzer Computer. A high-speed digital differential analyzer using parallel logic and arithmetic, solid-state circuitry, and modular construction and capable of being expanded in computing capacity, precision, and operating speed.

speech amplifier—A voltage amplifier made specifically for a microphone.

speech audiometer—An audiometer for measuring either live or recorded speech signals.

speech clipper—A speech-amplitude-limiting circuit which permits the average modulation percentage of an amplitude-modulated transmitter to be increased.

speech frequency—*See* Voice Frequency.

speech interpolation—The method of obtaining more than one voice channel per voice circuit by giving each subscriber a speech path in the proper direction only at the times when his speech requires it.

speech inverter—An apparatus that interchanges high and low speech frequencies by removing the carrier wave and transmission of only one sideband in a radiotelephone. This renders the speech unintelligible unless picked up by apparatus capable of replacing the carrier wave in the correct manner. (*Also see* Scrambler Circuit.)

speech level — The energy of speech (or music), measured in volume units on a volume indicator.

speech scrambler—*See* Scrambler Circuit.

speech synthesizer—A system used in research to generate speech from electrical signals in order to study human vocal patterns.

speed limit—A control function that prevents the controlled speed from exceeding prescribed limits.

speed of light—The speed at which light travels, or 186,284 miles per second.

speed of sound—Also called sonic speed. The speed at which sound waves travel through a medium (in air and at standard sea-level conditions, about 750 miles per hour or 1,080 feet per second).

speed of transmission — Also called rate of transmission. The instantaneous rate of processing information by a transmission facility. Usually measured in characters or bits per unit time.

speed-ratio control — A control function which maintains a preset ratio of the speeds of two drives.

speed regulation—The speed change of a motor between full-load and no-load, expressed in per cent of full-load speed.

speed regulator—A regulator which maintains or varies the speed of a motor at a predetermined rate.

speed-up capacitor—A capacitor used in RCTL to permit faster turn-on of the transistor in response to a change in input; it also helps overcome the storage delay of the transistor itself.

sphere gap—A spark gap with spherical electrodes. It is used as an excess-voltage protective device.

sphere-gap voltmeter—An instrument for measuring high voltages. It consists of a sphere gap, and the electrodes are moved together until the spark will just barely pass. The voltage can be calculated from the gap spacing and the electrode diameter, or read directly from a calibrated scale.

spherical aberration—Image defects (e.g., blurring) due to the spherical form of a lens or mirror. These defects cause a blurred image because the lens or mirror brings the central and marginal rays to different focuses. Common types of spherical aberration are astigmatism and curvature of the field.

spherical candlepower—In a lamp, the average candlepower in all directions in space. It is equal to the total luminous flux of the lamp, measured in lumens, divided by 4π.

spherical coordinates—A system of polar coordinates which originate in the center of a sphere. All points lie on the surface of the sphere, and the polar axis cuts the sphere at its two poles.

spherical-earth factor—Ratio between the electric field strengths that would result from propagation over an imperfectly conducting spherical earth and a perfectly conducting plane.

spherical wave—A wave in which the wavefronts are concentric spheres.

spider—A highly flexible ring, washer, or punched flat member used in a dynamic speaker to center the voice coil on the pole piece without appreciably hindering the in-and-out motion of the voice coil and its attached diaphragm.

spider-web antenna—An all-wave receiving antenna having several lengths of doublets connected somewhat like the web of a spider, to give favorable pickup characteristics over a wide range of frequencies.

spider-web coil—A flat coil having an open weave somewhat like the bottom of a woven basket. It was used in older radio receivers.

spike—An abrupt transient which comprises part of a pulse but exceeds its average amplitude considerably.

spike discriminator—A circuit used in a transponder to discriminate against pulses of extremely short duration, such as might be caused by ignition noise.

spike-leakage energy—The radio-frequency energy per pulse transmitted through TR and pre-TR tubes before and during the establishment of the steady-state radio-frequency discharge.

spike noise—In a chopper, the static field noise caused by insulating material; can be observed if the chopper is followed by a wideband amplifier.

spiking—Short, multiple, irregular bursts of laser-output radiation. Spiking is characteristic of pulse lasers, especially flash-pumped solid-dielectric types (e.g. ruby, neodymium in glass). The spike duration is typically 0.2 to 2 μsec.

spill—The redistribution and hence loss of information from a storage element of a charge-storage tube.

spindle—The upward-projecting shaft used on a phonograph turntable for positioning and centering the record.

spinner — An automatically rotatable radar antenna, together with its associated equipment.

spinning electron—An electron that spins with an angular momentum.

spinthariscope—An instrument for viewing the scintillations of alpha particles on a luminescent screen.

spin wave—A moving magnetic disturbance in a ferrite. The presence of an alternating magnetic field with its frequency near that of the natural precessional frequency of a spin tends to increase the precession angle. This increase in precession angle is passed from atom to atom by dipole and exchange interactions.

spin-wave amplitude—The difference between the precession angles of two spins.

spiral distortion—In camera tubes or image tubes using magnetic focusing, a form of distortion in which image rotation varies with distance from the axis of symmetry of the electron optical system.

spiral four—A quad in which the four conductors are twisted about a common axis, the two sets of opposite conductors being used as pairs.

spiral scanning—Scanning in which the maximum radiation describes a portion of a spiral, with the rotation always in one direction.

spkr—Abbreviation for speaker.

spl—Abbreviation for sound-pressure level.

splashproof—A device or machine so constructed and protected that external splashing will not interfere with its operation.

splashproof motor—An open motor in which the ventilating openings are so constructed that drops of liquid or solid particles falling on it, or coming toward it in a straight line at any angle not greater than 100° from the vertical, cannot enter either directly or by striking and running along a surface of the motor.

splatter—Adjacent-channel interference due to overmodulation of a transmitter by abrupt peak audio signals. It is particularly noticeable for sounds containing high-frequency harmonics.

splice insulation — Insulation used over a splice.

splicing block—A grooved metal or plastic device for holding magnetic recording tape while being spliced. The ends of the tape are inserted into the groove.

splicing tape — A pressure-sensitive, nonmagnetic tape used for splicing magnetic tape together.

split—1. Of radar tracks, the separation of radar data from a single track to such a degree that one or more additional tracks can be initiated manually or automatically; a similar condition exists when a raid is separated to such an extent that it can be represented on a situation display as two or more raids. 2. Initiation of split tracks, raids, or groups by a direction center-track monitor or by programming means.

split-anode magnetron—A magnetron with an anode divided into two segments, usually by parallel slots.

split-conductor cable—A cable in which each conductor is composed of two or more insulated conductors normally connected in parallel.

split fitting—A conduit fitting, bend, elbow, or tee split longitudinally so that it can be positioned after the wires have been drawn into the conduit. The two parts are held together usually by screws.

split gear—A type of gear designed to minimize backlash. The method consists of splitting one gear of a meshing pair and so connecting a spring between the two halves that pressure is exerted on both sides of the teeth of the other gear.

split hydrophone—A directional hydrophone in which the electroacoustic transducers are divided and arranged so that each division can induce a separate electromotive force between its own terminals.

split image—Two or more scenes appearing on a television screen as a result of trick "photography" at the studio.

split-phase motor—A single-phase induction motor having an auxiliary winding connected in parallel with the main winding, but displaced in a magnetic position from the main winding so as to produce the required rotating magnetic field for starting. The auxiliary circuit is generally opened when the motor has reached a predetermined speed.

split projector—A directional projector in which electroacoustic transducing elements are divided and arranged so that each division can be energized separately through its own terminals.

split-rotor plate—*See* Serrated Rotor Plate.

split-sound system—An early television receiver i-f system in which the audio and video i-f signals are separated right after the mixer stage and are amplified in separate i-f stages. Replaced by the more current intercarrier sound system.

split-stator variable capacitor—A variable capacitor with a rotor section common to two separate stator sections. Used for balancing in the grid and plate tank circuits of transmitters.

split transducer—A directional transducer in which electroacoustic transducing elements are divided and arranged so that each division is electrically separate.

split winding—An equal division of a winding which will allow series or parallel external connection of the divided winding (four external leads) of a servomotor.

spontaneous emission — Emission occurring without stimulation or quenching after excitation.

spool—A flanged form serving as the foundation on which a coil is wound.

sporadic E-layer—A portion that sometimes breaks away from the normal E-layer in the ionosphere and exhibits unusual erratic characteristics.

sporadic reflections—Also called abnormal reflections. Sharply defined, intense reflections from the sporadic E-layer. Their frequencies are higher than the critical frequency of the layer, and they occur anytime, anywhere, and at any frequency.

spot—1. The area instantaneously affected by the impact of an electron beam of a cathode-ray tube. 2. *See* Land.

spot noise factor—*See* Spot Noise Figure.

spot noise figure—Also called spot noise factor. Ratio of the output noise of a transducer, to the portion attributable to the thermal noise in the input termination when the termination has a standard noise temperature (290°K). The spot noise figure is a point function of input frequency.

spot projection—In facsimile: 1. An optical method in which the scanning or recording spot is delineated by an aperture between the light source and the subject copy or record sheet. 2. The optical system in which the scanning or recording spot is the size of the area being scanned or reproduced.

spot speed—1. In facsimile, the length of the scanning line times the number of lines per second. 2. In television, the product of the length (in units of elemental area, i.e., in spots) of the scanning line and the number of scanning lines per second.

spottiness—Bright spots scattered irregularly over the reproduced image in a television receiver, due to man-made or static interference entering the television system at some point.

spot welding—A resistance welding process whereby welds are made between two or more overlapping sheets of metal by pressing them together between two electrodes arranged to conduct current to the outer surfaces of the overlapped sheets. The tips of one or both of the electrodes are restricted in area to approximately the diameter of the spot weld desired.

spot wobble—An externally produced oscillating movement of an electron beam and its resultant spot. Spot wobble is used to eliminate the horizontal lines across the screen and thus make the picture more pleasing.

spreader—1. An insulating crossarm used to hold the wires of a transmission line apart. 2. The crossarm separating the parallel wire elements of an antenna.

spread groove—A groove cut between recordings. The groove, which has an abnormally high pitch, separates the recorded material

but still enables the stylus to travel from one to the next.

spreading anomaly—That part of the propagation anomaly that is identifiable with the geometry of the ray pattern.

spreading loss—The transmission loss suffered by radiant energy. The effect of spreading, or divergence, is measured by this loss.

spread spectrum transmission—A communications technique in which many different signal waveforms are transmitted in a wide band. Power is spread thinly over the band so narrow-band radios can operate within the wide band without interference.

spring—A resilient, flat piece of metal forming or supporting a contact member in a jack or a key.

spring-actuated stepping relay—A stepping relay in which cocking is done electrically and operation is produced by spring action.

spring contact—A relay or switch contact, usually of phosphor bronze and mounted on a flat spring.

spring curve—A plot of the spring force on the armature of a relay versus armature travel.

spring-finger action—The design of a contact, as used in a printed-circuit connector or a socket contact, permitting easy stress-free spring action to provide contact pressure and/or retention.

spring pile-up—An assembly of all contact springs operated by one armature lever.

spring-return switch—A switch which returns to its normal position when the operating pressure is released.

spring stop—In a relay, the member used to control the position of a pretensioned spring.

spring stud—In a relay, an insulating member that transmits the armature motion from one movable contact to another in the same pileup.

spst — Abbreviation for single-pole, single-throw.

spurious emanations — Unintentional and undesired emissions from a transmitting circuit.

spurious count—*See* Spurious Tube Counts.

spurious emission—*See* Spurious Radiation.

spurious pulse—In a scintillation counter, a pulse not purposely generated or directly due to ionizing radiation.

spurious pulse mode—An unwanted pulse mode which is formed by the chance combination of two or more pulse modes and is indistinguishable from a pulse interrogation or reply.

spurious radiation — Also called spurious emission. Emissions from a radio transmitter at frequencies outside its assigned or intended emission frequency. Spurious emissions include harmonic emissions, parasitic emissions, and intermodulation products, but exclude emissions in the immediate vicinity of the necessary band, which are a result of the modulation process for the transmission of information.

spurious response — 1. Any undesired response from an electric transducer or similar device. 2. The sensitivity of a circuit to signals of frequencies other than the frequency to which the circuit is tuned.

spurious-response attenuation—The ability of a receiver to discriminate between a desired signal to which it is resonant and an undesired signal at any other frequency to which it is simultaneously responsive.

spurious-response ratio—Ratio of the field strength at the frequency which produces a spurious response, to the field strength at the desired frequency, each field being applied in turn to produce equal outputs. Image ratio and intermediate-frequency response ratio are special forms of spurious-response ratio.

spurious signal—An unwanted signal generated either in the equipment itself or externally and heard (or seen) as noise.

spurious transmitter output—Any component of the radio-frequency output that is not implied by the type of modulation and the specified bandwidth.

spurious transmitter output, conducted—A spurious output of a radio transmitter that is conducted over a tangible transmission path such as a power line, control circuit, radio-frequency transmission line, waveguide, etc.

spurious transmitter output, extraband—A spurious transmitter output that lies outside the specified band of transmission.

spurious transmitter output, inband—A spurious transmitter output that lies within the specified band of transmission.

spurious transmitter output, radiated—A spurious output radiated from a radio transmitter. (The associated antenna and transmission lines are not considered part of the transmitter.)

spurious tube counts—Also called spurious counts. The counts in radiation-counter tubes, other than background counts and those caused directly by the radiation to be measured. They are caused by electrical leakage, failure of the quenching process, etc.

sputtering—1. Also called cathode sputtering. A process sometimes used in the production of the metal master disc. In this process the original is coated with an electric conducting layer by means of an electric discharge in a vacuum. 2. Ejection of atoms from a surface when bombarded by atomic particles.

spygmogram—A graphic recording of the movements, forms, and forces of an arterial pulse.

s-quad—Also called simple quad. An arrangement of two parallel paths, each of which contains two elements in series.

square-law demodulator — *See* Square-law Detector.

square-law detection—Detection in which the output voltage is substantially proportional to the square of the input voltage over the useful range of the detector.

square-law detector—Also called square-law demodulator. A detector in which the output voltage is essentially proportional to the square of the r-f input voltage.

square-law scale meter—A meter in which the deflection is proportional to the square of the applied energies.

square-loop ferrite—A ferrite with a rectangular hysteresis loop.

squareness ratio—1. For a magnetic material in a symmetrically cyclically magnetized condition, the ratio of the flux density at zero magnetizing force, to the maximum flux density. 2. The ratio of the flux density to the maximum flux density when the magnetizing force has changed halfway from zero toward its negative limiting value.

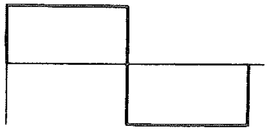

Square wave (ideal).

square wave—A square- or rectangular-shaped periodic wave which alternately assumes two fixed values for equal lengths of time, the transition time being negligible in comparison with the duration of each fixed value.

square-wave amplifier—A resistance-coupled amplifier (in effect, a wideband video amplifier) which amplifies a square wave with a minimum of distortion.

square-wave generator—A signal generator for producing square or rectangular waves.

square-wave response—In camera tubes, the ratio of the peak-to-peak signal amplitude given by a test pattern consisting of alternate black and white bars of equal widths, to the difference in signal between large-area blacks and whites having the same illuminations as the bars. Horizontal square-wave response is measured if the bars are perpendicular to the horizontal scan, and vertical square-wave response is measured if they are parallel.

squaring circuit—1. A circuit which changes a sine wave or other wave into a square wave. 2. A circuit which contains nonlinear elements and which produces an output voltage proportional to the square of the input voltage.

squawker — The midrange speaker of a three-way system.

squeal—Audible tape vibrations, primarily in the longitudinal node, caused by a frictional excitation at the heads and guides.

squealing—The high-pitched noise heard along with the desired intelligence in a radio receiver. It is due to interference between stations or to oscillation in one of the receiver circuits.

squeezable waveguide—In radar, a variable-width waveguide for shifting the phase of the radio-frequency wave traveling through it.

squeeze section—A length of waveguide, the critical dimension of which can be altered to correspond to changes in the electrical length.

squeeze track—A variable-density sound track in which variable width with greater signal-to-noise ratio is obtained by means of adjustable masking of the recording light beam and simultaneous increase of the electric signal applied to the light modulator.

squegger—A self-quenching oscillator in which the suppression occurs in the grid circuit.

squegging—A self-blocking condition in an oscillator circuit.

squegging oscillator—*See* Blocking Oscillator.

squelch—To automatically quiet a receiver by reducing its gain in response to a specified characteristic of the input.

squelch circuit—A circuit for preventing a radio receiver from producing an audio-frequency output in the absence of a signal having predetermined characteristics. A squelch circuit may be operated by signal energy in the receiver passband, by noise quieting, or by a combination of the two (ratio squelch). It may also be operated by a signal having special modulation characteristics (selective squelch).

squint—In radar, an ambiguous term, meaning either the angle between the two major-lobe axes in a lobe-switching antenna, or the angular difference between the axis of antenna radiation and a selected geometric axis such as the axis of the reflector.

squint angle—The angle between the physical axis of the antenna center and the axis of the radiated beam.

squirrel-cage induction motor—An induction motor in which the secondary circuit, usually the rotor, consists of a squirrel-cage winding (two discs connected along their circumference with copper bars) arranged in slots in the iron core.

squirrel-cage winding—A permanently short-circuited winding which is usually uninsulated, has its conductors uniformly distributed around the periphery of the machine, and is joined by continuous end rings.

squitter—In radar, random firing (intentional or otherwise) of the transponder transmitter in the absence of interrogation.

ss—*See* Single-Sideband Modulation.

ssb—*See* Single-Sideband Modulation.

ssfm—A system of multiplex in which the single-sideband subcarriers are used to frequency-modulate a second carrier.

sspm—A system of multiplex in which the single-sideband subcarriers are used to phase-modulate a second carrier.

stability—1. The ability of a component or device to maintain its nominal operating characteristics after being subjected to changes in temperature, environment, current, and time. It is usually expressed in either percent or parts per million for a given period of time. 2. The ability of a power supply to maintain a constant output voltage (or current) over a period of time under fixed conditions of input, load and temperature. Usually expressed in terms of a voltage (or current) change over a fixed length of time. 3. The ability to maintain effectiveness within reasonable bounds in spite of large changes in environment. *Also see* Drift. 4. For a feedback-control system or element, the property such that its output is asymptomatic (i.e., will ultimately attain a steady-state d-c level within the linear range and without continuing external stimuli).

stability factor—The measure of the bias stability of a transistor amplifier. It is defined as the change in collector current, I_c, per change in cutoff current, I_{co}.

stabilivolt—A gas-filled tube containing a number of concentric, coated iron electrodes. It is used as a source of practically constant voltage for apparatus drawing only small currents.

stabilization—1. The introducing of stability into a circuit. 2. A treatment of a magnetic material designed to increase its magnetic permanency.

stabilized feedback—*See* Negative Feedback.

stabilized flight—A type of flight in which control information is obtained from inertia-stabilized references such as gyroscopes.

stabilized local oscillator — An extremely stable radio-frequency oscillator used as a local oscillator in the superheterodyne radar receiver in a moving-target indicator system.

stabilized shunt-wound motor—A shunt-wound motor to which a light series winding has been added to prevent a rise in speed, or to reduce the speed when the load increases.

stabilized winding — Also called tertiary winding. An auxiliary winding used particularly in star-connected transformers: a. To stabilize the neutral point of the fundamental frequency voltage. b. To protect the transformer and the system from excessive third-harmonic voltages. c. To prevent telephone interference caused by third-harmonic currents and voltages in the lines and earth.

stabistor—A voltage-limiting semiconductor. A diode designed to break over and conduct at a certain voltage. This is the normal forward conduction of a diode and is also characteristic of zener diodes, which avalanche into conduction when breakdown (backward) voltage is exceeded.

stable element—In navigation, an instrument or device which maintains a desired orientation independently of the vehicle motion.

stable oscillation—A response which does not increase indefinitely with time; the opposite of an unstable oscillation.

stable platform—Also called a gyrostabilized platform. A gyro instrument which provides accurate azimuth, pitch, and roll attitude information. In addition to serving as reference elements, they are used for stabilizing accelerometers, star trackers, and similar devices in space.

stack—1. That portion of a computer memory and/or registers used to temporarily hold information. 2. *See* Pileup.

stacked array—An antenna consisting of elements placed one above the other. This is done to increase the sensitivity.

stacked dipole antenna—Antennas in which the antenna directivity is increased by providing a number of identical dipole elements, excited either directly or parasitically. The resultant radiation pattern will depend on the number of dipole elements used, the spacing and phase difference between the elements, and the relative magnitudes of the currents.

stacked heads—Also called in-line heads. An arrangement of magnetic recording heads used for stereophonic sound. The two heads are directly in line, one above the other.

stage—1. A term usually applied to an amplifier to mean one step, especially if part of a multistep process; or the apparatus employed in such a step. 2. A hydraulic amplifier used in a servovalve. Servovalves may be single-stage, two-stage, three-stage, etc.

stage-by-stage elimination—A method of locating trouble in electronic equipment by using a signal generator to introduce a test signal into each stage, one at a time, until the defective stage is found.

stage efficiency—Ratio of useful power (alternating current) delivered to the load, to the power at the input (direct current).

staggered heads—An infrequently-used arrangement of magnetic recording heads for stereophonic sound. The heads are $1\frac{7}{32}$ inch apart. Stereo tapes recorded with staggered heads cannot be played on recorders using stacked heads, and vice versa.

staggered tuning—A means of producing a wide bandwidth in a multistage i-f amplifier by tuning to different frequencies by a specified amount.

staggering—The offsetting of two channels of different carrier systems from exact side-band-frequency coincidence, in order to avoid mutual interference.

staggering advantage—A reduction in intelligible cross talk between identical channels of adjacent carrier systems as a result of using slightly different frequency allocations for the different systems.

stagger time—The interval between the times of actuation of any two contact sets.

stagger-tuned amplifier—An amplifier consisting of two or more stages, each tuned to a different frequency.

stagnation thermocouple—A type of thermocouple in which a high recovery factor is achieved by stagnating the flow in a space

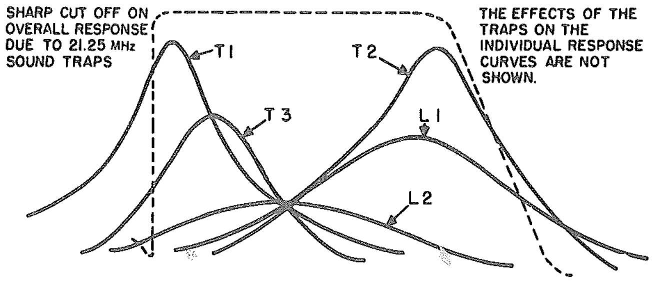

Staggered tuning.

surrounding the junction. This results in a high response time as compared with an exposed junction.

staircase generator — A special-purpose signal generator that produces an output which increases in steps; thus its output waveform has the appearance of a staircase.

staircase signal—A waveform consisting of a series of discrete steps resembling a staircase.

stalled-torque control — A control function used to control the drive torque at zero speed.

stall torque—1. The torque which the rotor of an energized motor produces when restrained from motion. 2. The torque developed by a servomotor at speed in excess of 1 rpm but less than ½% of the synchronous speed with a rated voltage and frequency of the proper phase relationship applied to both windings.

stalo—1. Acronym for STAbilized Local Oscillator used as part of a moving-target indication device in conjunction with a radar. 2. A highly stable oscillator, usually stabilized by feedback from a very high-Q LC circuit such as a high-Q cavity.

stamper—A negative (generally made of metal by electroforming), from which finished records are molded.

standard—An exact value, or a concept established by authority, custom, or agreement, to serve as a model or rule in the measurement of a quantity or in the establishment of a procedure.

standard antenna—An open single-wire antenna, including the lead-in wire, having an effective height of four meters.

standard beam approach—Abbreviated sba. A vhf 40-MHz, continuous-wave, low-approach system using a localizer and markers. The two main-signal lobes are tone-modulated with the Morse-code letters *E* and *T* (• and —). These modulations form a continuous tone when the aircraft is on its course. The airborne equipment is usually instrumented for visual reference, but may be used aurally in some applications.

standard broadcast band—The band of frequencies extending from 535 to 1605 kilohertz.

standard broadcast channel—The band of frequencies occupied by the carrier and two sidebands of a broadcast signal. The carrier frequency is at the center, with the sidebands extending 5 kHz on either side.

standard broadcast station—A radio station operated on a frequency between 535 and 1,605 kHz for the purpose of transmitting programs intended for reception by the general public.

standard candle — A unit of candlepower equal to a specified fraction of the visible light radiated by a group of 45 carbon-filament lamps preserved at the National Bureau of Standards, the lamps being operated at a specified voltage. The standard candle was originally the amount of light radiated by a tallow candle of specified composition and shape.

standard capacitor—A capacitor in which its capacitance is not likely to vary. It is used chiefly in capacitance bridges.

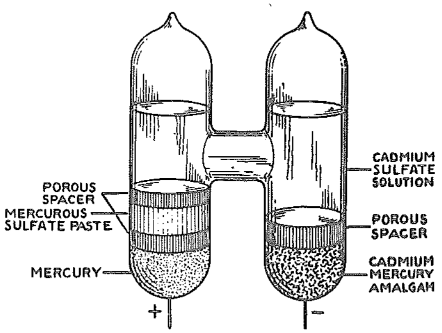

Standard cell.

standard cell—A primary cell which serves as a standard of voltage.

standard component—A component which is regularly produced by some manufacturer and is carried in stock by one or more distributors.

standard deflection — 1. In a galvanometer having an attached scale, one scale division. 2. In a galvanometer without an attached scale, 1 millimeter when the scale distance is 1 meter.

standard deviation—A measure of the variation of data from the average. It is equal to the root mean square of the individual deviations from the average.

standard eye—An observer that has red and infrared luminosity functions.

standard facility—In programming actions, a basic communications electronics functional entity which is engineered to satisfy a specific communications electronics operational requirement. An associated standard facility equipment list describes the facility functionally and indicates the material required for the standard facility.

standard-frequency service — A radiocommunication service that transmits for general reception specified standard frequencies of known high accuracy.

standard-frequency signal — One of the highly accurate signals broadcast by the National Bureau of Standards radio station WWV on 2.5, 5, 10, 15, 20, 25, 30, and 35 MHz at specific times. Used for testing and calibrating radio equipment all over the world.

standard-gain horn — A waveguide device that has essentially flared out its waveguide dimensions to specific lengths that match (with certain gain) the incoming energy to the atmosphere. The applications include using these horns for reflectors and lenses, pickup horns for sampling power, and receiving and transmitting antennas.

standardization—The process of establishing by common agreement engineering criteria, terms, principles, practices, materials, items, processes, equipment, parts, subassemblies, and assemblies to achieve the greatest practicable uniformity of items of supply and engineering practices, to insure the minimum feasible variety of such items and practices, and to effect optimum interchangeability of equipment parts and components.

standard microphone—A microphone, the response of which is known for the condition under which it is to be used.

standard noise temperature — A standard reference temperature (T) for noise measurements, taken as 290° K.

standard observer—A hypothetical observer who requires standard amounts of primaries in a color mixture to match every color.

standard pitch—The tone *A* at 440 hertz.

standard propagation—Propagation of radio waves over a smooth, spherical earth of uniform dielectric constant and conductivity, under standard atmospheric refraction.

standard reference temperature—In a thermistor, the body temperature for which the nominal zero-power resistance is specified.

standard refraction—The refraction which would occur in an idealized atmosphere; the index of refraction decreases uniformly with height at the rate of 39×10^{-6} per kilometer.

standard register of a motor meter—Also called a dial register. A four- or five-dial register, each dial being divided into ten equal parts numbered from zero to nine. The dial pointers are geared so that adjacent ones move in opposite directions at a 10-to-1 ratio.

standard resistor—Also called a resistance standard. A resistor which is adjusted to a specified value, is only slightly affected by variations in temperature, and is substantially constant over long periods of time.

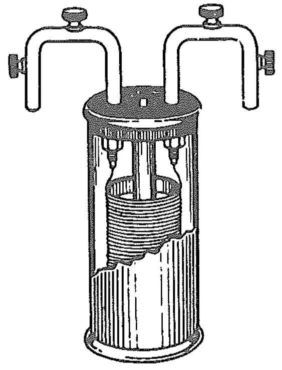

Standard resistor.

standard rod gap—A gap between the ends of two half-inch square rods. Each rod is cut off squarely and mounted on supports so that it overhangs the inner edge of each support by a length equal to or greater than half the gap spacing. It is used for approximate measurements of crest voltages.

standard sea-water conditions—Sea water with a static pressure of 1 atmosphere, a temperature of 15° C., and a salinity such that the velocity of sound propagation is exactly 1,500 meters per second.

standard sphere gap—A gap between two metal spheres of standard dimensions. It is used for measuring the crest value of a voltage, by observing the maximum gap spacing at which sparkover occurs when a voltage is applied under known atmospheric conditions.

standard subroutine—In a computer, a subroutine which is applicable to a class of problems.

standard television signal—A signal which conforms to accepted specifications.

standard test conditions—The environmental conditions under which measurements should be made when disagreement of data obtained by various observers at different times and places may result from making measurements under other conditions.

standard test-tone power—One milliwatt (0 dbm) at 1,000 hertz.

standard voltage generator — *See* Signal Generator.

standard volume indicator—A volume indicator with the characteristics prescribed by the American Standards Association.

standby — 1. The condition of equipment which will permit complete resumption of stable operation within a short period of time. 2. A duplicate set of equipment to be used if the primary unit becomes unusable because of malfunction.

standby battery—A storage battery held in reserve to serve as an emergency power source in event the regular power facilities at a radio station, hospital, etc., fail.

standby register—A register in which accepted or verified information can be stored to be available for a rerun in the event of a mistake in the program or a malfunction in the computer.

standby transmitter—A transmitter installed and maintained ready for use whenever the main transmitter is out of service.

standing-on-nines carry—A system of executing the carry process in a computer. If a carry into a given place produces a carry from there, the incoming carry information is routed around that place.

standing wave—1. The distribution of current and voltage on a transmission line formed by two sets of waves traveling in opposite directions, and characterized by the presence of a number of points of successive maxima and minima in the distribution curves. Also called stationary wave. 2. The sinusoidal distribution of current and voltage amplitudes along a transmission line as a result of the reflection of energy from a point where a mismatch of impedance occurs.

standing-wave detector—*See* Standing-Wave Meter.

standing-wave indicator—*See* Standing-Wave Meter.

standing-wave loss factor—Ratio of the transmission loss in a waveguide when it is unmatched, compared with the loss when it is matched.

standing-wave meter—Also called a standing-wave indicator or detector. An instrument for measuring the standing-wave ratio in a transmission line. It may also include means for finding the location of maximum and minimum amplitudes.

standing-wave ratio — Abbreviated swr. 1. The ratio of current (or voltage) at a loop (maximum) in the transmission line to the value at a (minimum) node. It is equal to the ratio of the characteristic impedance of the line to the impedance of the load connected to the output end of the line. 2. The ratio of the amplitude of a standing wave at an antinode to the amplitude at a node.

standing-wave-ratio bridge — Abbreviated swr bridge. A bridge for measuring the standing-wave ratio on a transmission line to check the impedance match.

standoff insulator—An insulator used to hold a wire or other radio component away from the structure on which it is mounted.

star chain—A group of navigational radio transmitting stations comprising a master station about which three or more slave stations are symmetrically located.

star-connected circuit—A polyphase circuit in which all current paths within the region that limits the circuit extend from each of the points of entry of the phase conductors to a common conductor (which may be the neutral conductor).

star connection—*See* Wye Connection.

stark effect—The splitting of spectral lines due to an applied electric field.

star network—A set of three or more branches with one terminal of each connected at a common node.

star-quad cable—Four wires laid together and twisted as a group.

starter—1. An auxiliary electrode used to initiate conduction in a glow-discharge, cold-cathode tube. 2. Sometimes referred to as a trigger electrode. A control electrode, the principal function of which is to establish sufficient ionization to reduce the anode breakdown voltage in a gas tube. 3. An electric controller for accelerating a motor from rest to normal speed.

starter breakdown voltage—The voltage required to initiate conduction across the starter gap of a glow-discharge, cold-cathode tube, all other tube elements being held at cathode potential before breakdown.

starter gap—The conduction path between a starter and the other electrode to which the starting voltage is applied in a glow-discharge, cold-cathode tube.

starter voltage drop—The voltage drop across the starter gap after conduction is established there in a glow-discharge, cold-cathode tube.

starting anode—The anode that establishes the initial arc in a mercury-arc rectifier tube.

starting circuit breaker — A device the principal function of which is to connect a machine to its source of starting voltage.

starting current of an oscillator—The value of oscillator current at which self-sustaining oscillations will start under specified loading.

starting electrode—The electrode that establishes the cathode spot in a pool-cathode tube.

starting reactor—A reactor for decreasing the starting current of a machine or device.

starting torque—Also called pull-in torque. The maximum load torque with which motors can start and come to synchronous speed.

starting-to-running transition contactor—A device which operates to initiate or cause the automatic transfer of a machine from the starting to the running power connection.

starting voltage—The voltage necessary for a gaseous voltage regulator to become ionized or to start conducting. As soon as this happens, the voltage drops to the operating value.

start lead—Also called inside lead. The inner termination of a winding.

startover—A program function that causes an inactive computer to become active.

start-record signal—In facsimile transmission, the signal that starts the converting of the electrical signal to an image on the record sheet.

start signal—The signal that converts facsimile-transmission equipment from standby to active.

start-stop multivibrator — *See* Monostable Multivibrator *and* Flip-Flop Multivibrator.

start-stop-printing telegraphy—Printing telegraphy in which the signal-receiving mechanisms are started and stopped at the beginning and end of each transmitted character.

starved amplifier—An amplifier employing pentode tubes in which the screen voltage is set 10% below the plate voltage and the plate-load resistance is increased to 10 times the normal value. Thus, the amplification factor is greatly increased—often a stage gain of 2,000 is achieved.

stat—A prefix used to identify electrostatic units in the cgs system. (*Also see* Statampere, Statcoulomb, Statfarad, Stathenry, Statmho, Statohm, *and* Statvolt.)

statampere—The cgs electrostatic unit of current, equal to 3.3356 × 10^{-10} ampere (absolute).

statcoulomb—The cgs electrostatic unit of charge equal to 3.3356 × 10^{-10} coulomb (absolute).

state—The condition of a circuit, system, etc.

state of charge—The condition of a storage cell or battery in terms of the remaining capacity.

statfarad—The cgs electrostatic unit of charge equal to 1.11263 × 10^{-12} farad (absolute).

stathenry—The cgs electrostatic unit of inductance equal to 8.98766 × 10^{11} henrys (absolute).

static—*See* Atmospherics.

static behavior—The behavior of a control system or an individual unit under fixed conditions (as contrasted to dynamic behavior, under changing conditions).

static characteristic—The relationship between a pair of variables such as electrode voltage and electrode current, all other voltages being maintained constant. This relationship is usually represented by a graph.

static charge—The accumulated electric charge on an object.

static check—Of a computer, consists of one or more tests of computing elements, their interconnections, or both, performed under static conditions.

static convergence—Convergence of the three electron beams at the center of the aperture mask in a color picture tube. The term "static" applies to the theoretical paths the beams would follow if no scanning forces were present.

static decay—In a storage tube, decay that is a function only of storage surface properties such as lateral or transverse leakage.

static device—As associated with electronic and other control or information-handling circuits, the term "static" refers to devices with switching functions that have no moving parts.

static electricity—Stationary electricity—i.e., in the form of a charge in equilibrium, or considered independently of the effects of its motion.

static eliminator—A device for reducing atmospheric static interference in a radio receiver.

static field—A field that is present between the poles of either a permanent magnet or an electromagnet that has a direct current passing through its coils.

static focus—The focus attained when the electron beam is theoretically at rest or is at the position it would occupy if scanning energy were not applied.

static forward-current transfer ratio—In a transistor, the ratio, under specified test conditions, of the d-c output current to the d-c input current.

static input resistance—In a transistor, the ratio of the d-c input voltage to the d-c input current.

staticizer—A storage device which is able to take information sequentially in time and put it out in parallel.

static machine—A machine for generating an electric charge, usually by induction.

static measurement—A measurement taken under conditions where neither the stimulus nor the environmental conditions fluctuate.

static pressure—Also called hydrostatic pressure. The pressure that would exist at a certain point in a medium with no sound waves present. In acoustics, the commonly used unit is the microbar.

static printout—In a computer, a printout of data that is not one of the sequential operations and occurs after conclusion of the machine run.

static register—A computer register which retains its information in static form.

static regulator—A transmission regulator in which the adjusting mechanism is in self-equilibrium at any setting and control power must be applied to change the setting.

static sensitivity—In phototubes, the direct anode current divided by the incident radiant flux of constant value.

static skew—A measure of the distance that the output from one track is ahead or behind (i.e., leading or lagging) the output of another track as a tape is transported over the read head.

static storage — In computers, storage in which the information does not change position (e.g., electrostatic storage, flip-flop storage, binary magnetic-core storage, etc.). The opposite of dynamic storage.

static subroutine—A digital-computer subroutine involving no parameters, other than the addresses of the operands.

static torque—*See* Locked-Rotor Torque.

static transconductance—In a transistor, the ratio of the d-c output current to the d-c input voltage.

station—One or more transmitters, receivers, and accessory equipment required to carry on a definite radio communication service. The station assumes the classification of the service in which it operates.

stationary battery—A storage battery designed for service in a permanent location.

stationary contacts—Those members of contact pairs that are not moved directly by the actuating system.

stationary field—A constant field—i.e., one where the scalar (or vector) at any point does not change during the time interval under consideration.

stationary wave—*See* Standing Wave.

station break—1. A cue given by the station originating a program, to notify network stations that they may identify themselves to their audiences, broadcast local items, etc. 2. The actual time taken in (1) above.

statmho—The cgs electrostatic unit of conductance equal to 1.1126×10^{-12} mho (absolute).

statohm—The cgs electrostatic unit of resistance equal to 8.98766×10^{11} ohms (absolute).

stator—1. The nonrotating part of the magnetic structure in an induction motor. It usually contains the primary winding. 2. The stationary plates of a variable capacitor. 3. The conducting surfaces of a switch. Similar to the commutator in electrical rotating mechanisms.

stator of an induction watthour meter—A voltage circuit, one or more current circuits, and a magnetic circuit combined so that the reaction with currents induced in an individual, or a common, conducting disc exerts a driving torque on the rotor.

stator plates—The fixed plates of a variable capacitor.

statvolt—The cgs electrostatic unit of voltage equal to 299.796 volts (absolute).

stave—One of the number of individual longitudinal elements which comprise a sonar transducer.

stay cord—A component of a cable, usually a high-tensile textile used to anchor the cable ends at their points of termination, to keep any pull on the cable from being transferred to the electrical connection.

steady state—A condition in which circuit values remain essentially constant, occurring after all initial transients or fluctuating conditions have settled down.

steady-state deviation—The difference between the final value assumed by a specified variable after the expiration of transients and its ideal value.

steady-state oscillation—Also called steady-state vibration. Oscillation in which the motion at each point is a periodic quantity.

steady-state regulation—Slow changes in the output voltage of a power supply following an input-voltage and/or load-impedance variation; it is usually expressed in volts (ΔV) or as a percentage of the nominal output.

steady-state vibration—*See* Steady-State Oscillation.

steatite—A ceramic consisting chiefly of a silicate of magnesium. Because of its excellent insulating properties—even at high frequencies—it is used extensively in insulators.

steerable antenna—An antenna the major lobe of which can be readily shifted in direction.

Stefan-Boltzmann law—The total emitted radiant energy per unit of a black body is proportionate to the fourth power of its absolute temperature.

Steinmetz coefficient—A factor by which the 1.6th power of the magnetic flux density must be multiplied to give the approximate hysteresis loss of an iron or steel sample in ergs per cubic centimeter per cycle when that sample is undergoing successive magnetization cycles having the same maximum flux density.

stenode circuit—A superheterodyne receiving circuit in which a piezoelectric unit is used in the intermediate-frequency amplifier to balance out all frequencies except signals at the crystal frequency, thereby giving very high selectivity.

step-by-step switch—A bank-and-wiper switch in which the wipers are moved by individual electromagnet ratchet mechanisms.

step calibration—Also called interval calibration. Often confused with sense step, that is, the application of a calibration resistor to produce a deliberate electrical unbalance.

step counter—In a computer, a counter used in the arithmetical unit to count the steps in multiplication, division, and shift operation.

step-down transformer—A transformer in which the voltage is reduced as the energy is transferred from its primary to its secondary winding.

step function—A signal having one or more sudden discontinuities.

step-function response—*See* Transient Response.

step generator—A device for testing the linearity of an amplifier. A step wave is applied to the amplifier input and the step waveform observed, on an oscilloscope, at the output.

step input—A sudden but sustained change in an input signal.

step load change—An instantaneous change in the magnitude of the load current.

stepper motor—A motor the normal operation of which consists of discrete angular motions of essentially uniform magnitude, rather than continuous rotation.

stepping—*See* Zoning.

stepping relay—Also called rotary stepping switch (or relay) or stepping switch. A multiposition relay in which moving wiper contacts mate with successive sets of fixed contacts in a series of steps, moving from one step to the next in successive operations of the relay.

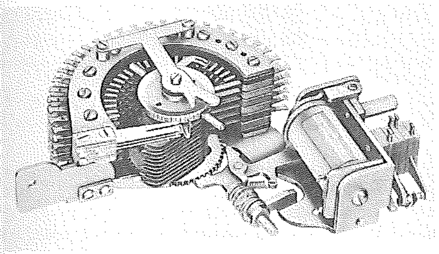

Stepping relay.

stepping switch—*See* Stepping Relay.

step-recovery diode—A varactor in which forward voltage injects carriers across the junction, but before the carriers can combine, the voltage reverses and carriers return to their origin in a group. The result is the abrupt cessation of reverse current and a harmonic-rich waveform.

step-servo motor—A device that, when properly energized by d-c voltage, indexes in definite angular increments.

step-strobe marker—A form of strobe marker in which the discontinuity is in the form of a step in the time base.

step-up transformer—A transformer in which the voltage is increased as the energy is transferred from the primary to the secondary winding.

step-voltage regulator—A device consisting of a regulating transformer and a means for adjusting the voltage or the phase relation. of the system circuit in steps, usually without interrupting the load.

steradian—A solid spherical angle which encloses a surface equal, on a sphere, to the square of the radius of the sphere.

stereo—A prefix meaning three-dimensional—specifically (especially without the hyphen), stereophonic.

stereo adapter—Also called a stereo control unit. A device used with two sets of monophonic equipment to make them act as a single stereo system.

stereo amplifier—An audio-frequency amplifier, with two or more channels, for a stereo sound system.

stereo broadcasting—*See* Stereocasting.

stereo cartridge—A phonograph pickup for reproduction of stereophonic recordings. Its high-compliance needle is coupled to two independent voltage-producing elements.

stereocasting—Also called stereo broadcasting. Broadcasting over two sound channels to provide stereo reproduction. This may be done by simulcasting, multicasting, or multiplexing.

stereocephaloid microphone—Two or more microphones arranged to simulate the acoustical patterns of human hearing.

stereo control unit—*See* Stereo Adapter.

stereo microphones—Two or more microphones spaced as required for stereo recording.

stereophonic—Multichannel sound, creating a sense of depth due to the separation of or variation between the signals received by each ear.

stereophonic reception—Reception involving the use of two receivers having a phase difference in their reproduced sounds. The sense of depth given to the received program is analogous to the listener's being in the same room as the orchestra or other medium.

stereophonic separation—The ratio of the electrical signal caused in the right (or left) stereophonic channel to the electrical signal caused in the left (or right) stereophonic channel by the transmission of only a right (or left) signal.

stereophonic sound system—A sound system with two or more microphones, transmission channels, and speakers arranged to give depth to the reproduced sound.

stereophonic subcarrier—A subcarrier having a frequency which is the second harmonic of the pilot subcarrier frequency and which is employed in f-m stereophonic broadcasting.

stereophonic system—A sound-reproducing system in which a plurality of microphones, transmission channels, and speakers (or earphones) are arranged to afford a listener a sense of the spatial distribution of the sound sources.

stereophonic subchannel—The band of frequencies from 23 to 53 kilohertz containing the stereophonic subcarrier and its associated sidebands.

stereo pickup—A phonograph pickup used with single-groove, two-channel stereo records.

stereo recording—The impressing of signals from two channels onto a tape or disc in such a way that the channels are heard separately on playback. The result is a directional, three-dimensional effect.

stereoscopic television—A system of television broadcasting in which the images appear to be three-dimensional.

stereosonic system—A recording technique using two closely-spaced directional microphones with their maximum directions of reception 45° from each other. In this way, one picks up sound largely from the right and the other from the left, similar to mid-side recording.

stick circuit—A circuit used to maintain energization of a relay or similar unit through its own contacts.

sticking—In computers, the tendency of a flip-flop to remain in, or to spontaneously switch to, one of its two stable states.

stiffness factor—The angular lag between the input and output of a servosystem.

still—Photographic or other stationary illustrative material used in a television broadcast.

stimularity—An arbitrary measure of sensitivity to stimulation. It is proportional to the quantum efficiency relative to incident radiation.

stimulated emission—The emission of radiation by a system going from an excited electron energy level to a lower energy level under the influence of a radiation field. The emitted radiation is in phase with the stimulating radiation and produces a negative-absorption condition.

stimulus—*See* Excitation *and* Measurand.

stirring effect—The circulation in a molten conductive charge due to the combined motor and pinch effects.

stochastic — The characteristic of events changing the probabilities of various responses.

stochastic process—A group of random variables.

stock reel—Also called supply reel or storage reel. On a tape recorder, the reel from which unrecorded or unplayed tape is taken as the machine records on or plays it.

stoichiometric impurity—A crystalline imperfection caused in a semiconductor by a deviation from the stoichiometric composition.

stop-cycle timer — A timer which runs through a single cycle and then stops until the starting signal is reinitiated.

stop opening — In a camera, the size of the aperature that controls the amount of light passing through the lens.

stopping potential—The voltage required to stop the outward movement of electrons emitted by photoelectric or thermionic action.

stop-record signal—A facsimile signal used for stopping the conversion of the electric signal into an image on the record sheet.

stop signal—The signal that transfers facsimile equipment from active to standby.

storage—1. The act of storing information (*Also see* Store.) 2. Sometimes called a memory. Any device in which information can be stored. 3. A computer section used primarily for storing information in electrostatic, ferroelectric, magnetic, acoustic, optical, chemical, electronic, electrical, mechanical, etc., form. Such a section is sometimes called a memory, or a store, in British terminology. 4. A device into which data can be entered, held, and retrieved at a later time.

storage access time—In a computer, the time required to transfer information from a storage location to the local storage register or other location, where the information then becomes available for processing.

storage battery—Two or more storage cells connected in series and used as a unit.

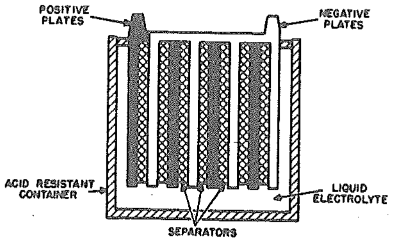

Storage battery.

storage capacity—Also called memory capacity. The amount of information that can be retained in a storage (memory) device. It is often expressed as the number of words (given the number of digits and the base of the standard word).

storage cell—1. Also called a secondary cell. A cell which, after being discharged, can be recharged by sending an electric current through it in the opposite direction from the discharging current. 2. An elementary unit of storage (e.g., binary cell, decimal cell).

storage device—A device into which data can be inserted, in which it can be retained, and from which it can be retrieved.

storage element—An area which retains information distinguishable from that of adjacent areas on the storage area of charge-storage tubes.

storage integrator—In an analog computer, an integrator used to store a voltage in the hold condition for future use while the rest of the computer assumes another computer control state.

storage laser—Any laser which stores unusually high energy prior to discharge. For example, a storage diode laser is a laser in which some carriers are electrically excited for a time longer than the lasing period.

storage life—The minimum length of time over which a device, system, or transducer can be exposed to specified environmental storage conditions without changing the performance beyond a specified tolerance.

storage location—A computer storage position that holds one machine word and usually has a specific address.

storage medium—Any recording device or medium into which data can be copied and held until some later date, and from which the entire original data can be obtained.

storage oscilloscope — An instrument that has the ability to store a crt display in order that it may be observed for any required time. This stored display may be instantly erased to make way for storage of a later event.

storage temperature—The temperature of the medium, immediately adjacent to the device, at which the device, without power applied, may be stored indefinitely without deterioration.

storage time—1. The time during which the output current or voltage of a pulse is falling from maximum to zero after the input current or voltage has been removed. 2. In a transistor, the time required to sweep current carriers from the collector region when the switch is turned off.

storage tube — *See* Electrostatic Memory Tube.

store—1. To retain information in a device from which the information can later be withdrawn. 2. To introduce information into the device in (1) above. 3. A British synonym for storage.

stored-energy welding—A method of welding in which electric energy is accumulated (stored) electrostatically, electromagnetically, or electrochemically at a relatively slow rate and is then released at the required rate for welding.

stored program—A set of instructions in the computer memory specifying the operations to be performed and the location of the data on which these operations are to be performed.

stored-program computer — 1. Also called general-purpose computer. A computer in which the instructions specifying the program to be performed are stored in the memory section along with the data to be operated on. 2. A digital computer which, under control of its own instructions, can synthesize, alter, and store instructions as though they were data and can subsequently execute these new instructions.

store transmission bridge—A transmission bridge which consists of four identical impedance coils (the two windings of the back-bridge relay and the live relay of a connector, respectively) separated by two capacitors. It couples the calling and called telephones together electrostatically for the transmission of voice-frequency (alternating) currents, but separates the two lines for the transmission of direct current for talking purposes (talking current).

storm loading—The mechanical loading imposed on the components of a pole line by wind, ice, etc., and by the weight of the components themselves.

straight dipole—A half-wave antenna consisting of one conductor, usually center-fed.

straight-line capacitance — The variable-capacitor characteristic obtained when the rotor plates are shaped so that the capacitance varies directly with the angle of rotation.

straight-line code—The repetition of a sequence of instructions, with or without address modification, by explicitly writing the instructions for each repetition. Generally straight-line coding will require less execution time and more space than equivalent loop coding. If the number of repetitions is large, this type of coding is tedious unless a generator is used. The feasibility of straight-line coding is limited by the space required as well as the difficulty of coding a variable number of repetitions.

straight-line frequency—The variable-capacitor characteristic obtained when the rotor plates are shaped so that the resonant frequency of the tuned circuit containing the capacitor varies directly with the angle of rotation.

straight-line wavelength — The variable-capacitor characteristic obtained when the rotor plates are shaped so that the wavelength of resonance in the tuned circuit containing the capacitor varies directly with the angle of rotation.

strain—The physical deformation, deflection, or change in length resulting from stress (force per unit area). The magnitude of strain is normally expressed in microinches per inch.

strain anisotropy—A force that directs the magnetization of a particle along a preferred direction relative to the strain.

strain gage—A measuring element for converting force, pressure, tension, etc., into an electrical signal.

strain-gage based—An instrument or transducer with a sensing element composed of bonded or unbonded strain gages.

strain-gage transduction—The conversion of the measurand into a change in resistance caused by strain in four or, more rarely, two arms of a Wheatstone bridge.

strain insulator—A single insulator, an insulator string, or two or more strings in parallel designed to transmit the entire pull of the conductor to, and insulate the conductor from, the tower or other support.

strain pickup—A phonograph pickup cartridge using the principle of the strain gage.

strain wire—Wire having a composition such that it exhibits favorable strain-gage performance. Also the strain-sensitive filaments which constitute the transfer electrical elements in certain transducers.

strand—One of the wires or groups of wires of any stranded conductor.

stranded conductor—*See* Stranded Wire.

stranded wire—Also called stranded conductor. A conductor composed of a group of wires or any combination of groups of wires. The wires in a stranded conductor are usually twisted or braided together.

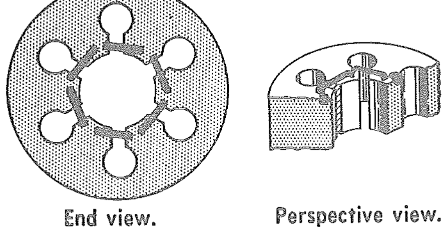

Straps on cavity-magnetron anode.

strap—A wire or strip connected between the ends of the segments in the anode of a cavity magnetron to promote operation in the desired mode.

strapping—The suppressing of undesired modes of oscillation in a magnetron.

stratosphere—A calm region of the upper atmosphere characterized by little or no temperature change throughout. It is separated from the lower atmosphere (troposphere) by a region called the tropopause.

stray capacitance—The capacitance introduced into a circuit by the leads and wires connecting the circuit components.

stray current—A portion of the total current that flows over paths other than the intended circuit.

stray-current corrosion—Corrosion that results when a direct current from a battery or other external source causes a metal in contact with an electrolyte to become anodic with respect to another metal in contact with the same electrolyte. Accelerated corrosion will occur at the electrode where the current direction is from the metal to the electrolyte and will generally be in proportion to the total current.

stray field—The leakage magnetic flux that spreads outward from an inductor and does no useful work.

strays—*See* Atmospherics.

streaking—Distortion in which televised objects appear stretched horizontally beyond their normal boundaries. It is most apparent at the vertical edges, where there is a large transition from black to white or white to black, and is usually expressed as short, medium, or long streaking. Long streaking may extend as far as the right edge of the picture, and in extreme low-frequency distortion, even over a whole line interval.

streamer breakdown—Breakdown caused by an increase in the field due to the accumulation of positive ions produced during electron avalanches.

streaming—The production of a unidirectional flow of currents in a medium where sound waves are present.

strength of a simple sound source—The rms magnitude of the total air flow at the surface of a simple source in cubic meters per second (or cubic centimeters per second), where a simple source is taken to be a spherical source the radius of which is small compared with one-sixth wavelength.

strength of a sound source—The maximum instantaneous rate of volume displacement produced by the source when emitting a sinusoidal wave.

stress—The force producing strain in a solid.

stretched dislay—A ppi display having the polar plot expanded in one rectangular dimension. The equal-range circles of the normal ppi display becomes ellipses.

striation technique—Rendering sound waves visible by using their individual ability to refract light waves.

striking an arc—Starting an electric arc by touching two electrodes together momentarily.

striking distance—The effective separation of two conductors having an insulating fluid between them.

striking potential—1. The voltage required to start an electric arc. 2. The lowest grid-to-cathode potential at which plate current begins flowing in a gas-filled triode.

string—In a list of items, a group of items that are already in sequence according to a rule.

string electrometer—An electrostatic voltage-measuring instrument consisting of a conducting fiber stretched midway between and parallel to two conducting plates. The electrostatic field between the plates displaces the fiber laterally in proportion to the voltage between the plates.

string-shadow instrument—An instrument in which the indicating means is the shadow (projected or viewed through an optical system) of a filamentary conductor, the position of which in a magnetic or an electric field depends on the measured quantity.

strip—To remove insulation from a wire or cable.

stripper—A hand-operated or motor-driven tool for removing insulation from wires.

strip transmission line—A transmission component that is similar to a microstrip except that it has a second ground plane placed above the conductor strip.

strobe—1. An intensified spot in the sweep of a deflection-type indicator, used as a reference mark for ranging or expanding the presentation. 2. An intensified sweep on a plan-position indicator or B-scope. Such a strobe may result from certain types of interference or it may be purposely applied as a bearing or heading marker. 3. On a console oscillator, a line representing the azimuth data generated by a jammed radar site. 4. *See* Electronic Flash.

strobe marker—A small birght spot, a short gap, or other discontinuity produced on the line trace of a radar display to indicate the part of the time base which is receiving attention.

strobe pulse—1. Also called sample pulse. A pulse used to gate the output of a core-memory sense amplifier into a trigger in a register. 2. A pulse of a duration less than the time period of a recurrent phenomenon used for making a close investigation of that phenomenon. The frequency of the strobe pulse bears a simple relation to that of the phenomenon, and the relative timing is usually adjustable.

strobolume—A trade name used to describe a high-intensity electronic stroboscope manufactured by General Radio Company.

stroboscope—A device that indicates frequency of operation by creating the optical illusion of slowing down or stopping a moving pattern, which is illuminated by a light that flashes at a known frequency.

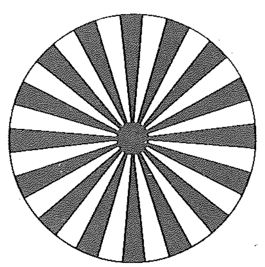

Stroboscope disc.

stroboscopic disc—A printed disc having several rings, each with a different number of dark segments. The pattern is placed on a rotating phonograph turntable and illuminated at a known frequency by a flashing discharge tube. The speed can then readily be determined by noting which pattern appears to stand still or rotate the slowest.

stroboscopic tachometer—A stroboscope with a scale calibrated in flashes or in revolutions per minute. The stroboscopic lamp is directed onto the rotating device being measured, and the flashing rate is adjusted until the device appears to be standing still. The speed can then be read directly from the scale.

strobotron—A type of glow lamp that produces intense flashes of light when fed with accurately timed voltage pulses. It is used in electronic stroboscopes for visual inspection of high-speed moving parts.

stroke speed—Also called scanning line frequency or scanning frequency. The number of times per minute that a fixed line, perpendicular to the direction of scanning, is crossed in one direction by a scanning or recording spot in a facsimile system.

structureborne noise—Undesired vibration in, or of, a solid body.

stub—1. A short length of transmission line or cable joined as a branch to another transmission line or cable. 2. A short section of transmission line, open or shorted at the far end, connected in parallel with a transmission line to match the impedance of the line to that of an antenna or transmitter.

stub angle—A right-angle elbow for a coaxial r-f transmission line, the inner conductor being supported by a quarter-wave stub.

stub cable—A short branch from a principal cable. The end is often sealed until used. Pairs in the stub are referred to as stubbed-out pairs.

stub matching—Using a stub to match a transmission line to an antenna or load.

stub-supported coaxial—A coaxial cable the inner conductor of which is supported by short-circuited coaxial stubs.

stub tuner—A stub terminated by movable short-circuiting means and used for matching impedance in the line to which it is joined as a branch.

stunt box—A device for controlling the nonprinting functions of a teletype terminal.

stutter—In facsimile, a series of undesired black and white lines sometimes produced when the signal amplitude changes sharply.

stylus—1. Also called a needle. The needlelike object used in a sound recorder to cut or emboss the record grooves. Generally it is made of sapphire, stellite, or steel. The plural is styli. 2. The pointed element that contacts the record sheet in a facsimile recorder.

stylus alignment—The position of the stylus with respect to the record. The correct position is perpendicular.

stylus drag—Also called needle drag. The friction between the reproducing stylus and the surface of the recording medium.

stylus force—Also called vertical stylus force, tracking force, and formerly called needle pressure or stylus pressure. The downward force, in grams or ounces, exerted on the disc by the reproducing stylus.

stylus oscillograph—An instrument in which a pen or stylus records, on paper or another suitable medium, the value of an electrical quantity as a function of time.

stylus pressure—*See* Stylus Force.

subassembly—Parts and components combined into a unit for convenience in assembling or servicing. A subassembly is only part of an operating unit; it is not complete in itself.

subatomic—Smaller than atoms—i.e., electrons or protons.

subatomic particles—The particles that make up the atom—i.e., protons, electrons, and neutrons.

subcarrier—A carrier used to generate a modulated wave which is applied, in turn, as a modulating wave to modulate another carrier.

subcarrier band—A band associated with a given subcarrier and specified in terms of maximum subcarrier deviation.

subcarrier channel—The channel required to convey the telemetric information of a subcarrier band.

subcarrier discriminator tuning unit—A device which tunes the discriminator to a particular subcarrier.

subcarrier frequency shift—The use of an audio-frequency shift signal to modulate a radio transmitter.

subcarrier oscillator—1. In a telemetry system, the oscillator which is directly modulated by the measurand or its equivalent in terms of changes in the transfer elements of a transducer. 2. In a color television receiver, the crystal oscillator operating at the chrominance subcarrier frequency of 3.58 MHz.

subchassis—The chassis on which closely associated components such as those of an amplifier or power supply are mounted. A subchassis is a building block, easily changed and usable in a variety of systems.

subcommutation—In a computer, the act of connecting one data source to a sampled data system less frequently than other data sources.

subcommutation frame—In pcm systems, a recurring integral number of subcommutator words which include a single subcommutation frame synchronization word. The number of subcommutator words in a subcommutation frame is equal to an integral number of primary commutator frames. The length of a subcommutation frame is equal to the total number of words or bits generated as a direct output of the subcommutator.

subcycle generator—A frequency-reducing device used in telephone equipment to furnish ringing power at a submultiple of the power-supply frequency.

subdivided capacitor—A capacitor in which several capacitors known as sections are mounted so that they may be used individually or in combination.

subelement—A distinguishable portion of a circuit element (e.g., the emitter, collector, and base are subelements of an integrated bipolar transistor).

subframe—A complete sequence of frames during which all subchannels of a specific channel are sampled once.

subharmonic—A sinusoidal quantity, the frequency of which is an integral submultiple of the fundamental frequency of its related periodic quantity. A wave with half the frequency of the fundamental of another wave is called the second subharmonic of that wave; one with a third of the fundamental frequency is called a third subharmonic, etc.

subject copy—Also called copy. In facsimile, the material in graphic form to be transmitted for reproduction by the recorder.

submarine cable—A cable designed for service underwater; usually a lead-covered cable with a steel armor applied between layers of jute.

submerged-resistor induction furnace — A furnace for melting metal. It comprises a melting hearth, a depending melting channel closed through the hearth, a primary induction winding, and a magnetic core which links the melting channel and primary winding.

submersible transformer—A transformer so constructed that it will operate when submerged in water under predetermined conditions of pressure and time—e.g., a subway transformer.

subminiature tube — A small electron tube used generally in miniaturized equipment.

subminiaturization—The technique of packaging discrete miniaturized parts, using assembly techniques which result in increased volumetric efficiency (e.g., a hybrid circuit).

submultiple resonance—Resonance at a frequency that is a submultiple of the frequency of the exciting impulses.

subnanosecond—Less than a nanosecond.

subpanel—An assembly of electrical devices connected together which forms a simple functional unit in itself.

subprogram—A part of a computer program.

subroutine—1. In computer technology, the portion of a routine that causes a computer to carry out a well-defined mathematical or logical operation. 2. Usually called a closed subroutine. One to which control may be transferred from a master routine, and returned to the master routine at the conclusion of the subroutine.

subscriber's drop—The line from a telephone cable to a subscriber's building.

subscriber set—Also called a customer set. An assembly of apparatus for originating or receiving calls on the premises of a subscriber to a communication or signaling service.

subscription television—*See* Pay Television.

subscription television broadcast program—A television broadcast program intended to be received in intelligible form by members of the public only for a fee or charge.

subset—1. In a telephone system, the handset or deskset at the station location. 2. Also known as a modem, data set, or subscriber set. A modulation/demodulation device designed to make business-machine signals compatible with communications facilities.

subsidiary conduit—The terminating branch of an underground conduit run extending from a manhole or handhole to a nearby building, handhole, or pole.

subsonic frequency — *See* Infrasonic Frequency.

subsonic speed—A speed less than the speed of sound.

substation—Any building or outdoor location at which electric energy in a power system is transformed, converted, or controlled.

substep—A part of a computer step.

substitute — In a computer, to replace one element of information by another.

substitution interference measurements—A measurement in which the noise level of the source being measured is compared to a known level from a calibrated source: impulse generator for broad-band noise or sine-wave generator for cw.

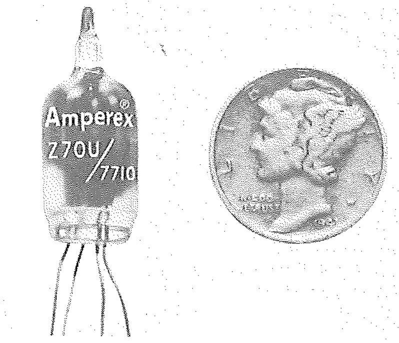

Subminiature tube.

substitution method—A three-step method of measuring an unknown quantity in a circuit. First, some circuit effect dependent on the unknown quantity is measured or observed. Then a similar but measurable quantity is substituted in the circuit. Finally, the latter quantity is adjusted to produce a like effect. The unknown value is then assumed to be equal to the adjusted known value.

substrate (of an integrated circuit)—1. The supporting material upon or within which an integrated circuit is fabricated or to which an integrated circuit is attached. 2. A material on the surface of which an adhesive substance is spread for bonding or coating; any material which provides a supporting surface for other materials, especially materials used to support printed-circuit patterns.

subsurface wave—An electromagnetic wave propagated through water or land. Operating frequencies for communications may be limited to approximately 35 kHz due to attenuation of high frequencies.

subsynchronous—Having a frequency that is a submultiple of the driving frequency.

subsynchronous reluctance motor—A form of reluctance motor with more salient poles than the number of electrical poles in the primary winding. As a result, the motor operates at a constant average speed which is a submultiple of its apparent synchronous speed.

subsystem—1. A major, essential, functional part of a system. The subsystem usually consists of several components. 2. An interconnected combination of a set of related circuits or integrated circuits to form a logical subdivision of an equipment or operational system.

subtractive filter—An optical filter which is of a certain color and eliminates that color when placed in the path of white light.

subtractor—An operational amplifier circuit in which the output is proportional to the difference between its two input voltages or between the net sums of its positive and negative inputs.

subway transformer—A transformer of submersible construction.

success ratio—The ratio of the number of successful attempts to the total number of trials. It is frequently used as a reliability index.

suckout—A hole in the response pattern of a tuned circuit due to the self-resonance of components at certain frequencies.

sudden commencement—Magnetic storms which start suddenly (within a few seconds) and simultaneously all over the earth.

sudden ionospheric disturbances—The sudden increase in ionization density in lower parts of the ionosphere, caused by a bright solar chromospheric eruption. It gives rise to a sudden increase of absorption in radio waves propagated through the low parts of the ionosphere, and sometimes to simultaneous disturbances of terrestrial magnetism and earth current. The change takes place within one or a few minutes, and conditions usually return to normal within one or a few hours.

Suhl effect—When a strong transverse magnetic field is applied to an N-type semiconducting filament, the holes injected into the filament are deflected to the surface. Here they may recombine rapidly with electrons and thus have a much shorter life, or they may be withdrawn by a probe as though the conductance had increased.

suicide control — A control function which uses negative feedback to reduce and automatically maintain the generator voltage at approximately zero.

sulfating—The accumulation of lead sulfate on the plates of a lead-acid storage battery. This reduces the energy-storing ability of the battery and causes it to fail prematurely.

sum channel—A combination of left and right stereo channels identical to the program, which may be recorded or transmitted monophonically.

summary punch — A punch-card machine which may be attached to another machine in such a way that it will punch information produced, calculated, or summarized by the other machine.

summary recorder—In computers, output equipment which records a summary of the information handled.

summation check—A redundant computer check in which groups of digits are summed, usually without regard to overflow. The sum is then checked against a previously computed sum to verify the accuracy of the computation.

summation frequency—A frequency which is the sum of two other frequencies that are being produced simultaneously.

summation tone—A combination tone, heard under certain circumstances, the pitch of which corresponds to a frequency equal to the sum of the frequencies of the two components.

summing point—1. A mixing point, the output of which is obtained by adding its inputs (with the prescribed signs). 2. The input terminal to which the feedback loop from the output of an amplifier is returned.

supercommutation — Commutation at a higher rate by connection of a single data input source to equally spaced contacts of the commutator (cross patching). Corresponding cross patching is required at the decommutator.

superconducting — Exhibiting superconductivity.

superconductivity—The decrease in resist-

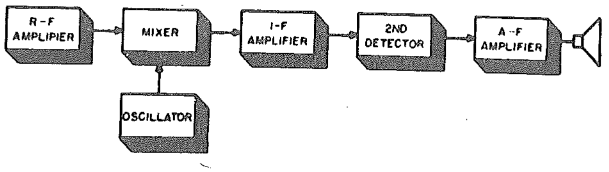

Superheterodyne receiver.

ance of certain materials (lead, tin, thallium, etc.) as their temperature is reduced to nearly absolute zero. When the critical (transition) temperature is reached, the resistance will be almost zero.

superconductor—A material that exhibits superconductivity.

super-emitron—*See* Image Iconoscope.

supergroup—A group of 60 channels occupying the range of frequencies between 312 kHz and 552 kHz.

superhet—Slang for a superheterodyne receiver.

superheterodyne receiver—A receiver in which the incoming modulated r-f signals are usually amplified in a preamplifier and then fed into the mixer for conversion into a fixed, lower carrier frequency (called the intermediate frequency). The modulated i-f signals undergo very high amplification in the i-f-amplifier stages and are then fed into the detector for demodulation. The resultant audio or video signals are usually further amplified before being sent to the output.

superheterodyne reception—A method of receiving radio waves in which heterodyne reception converts the voltage of the received wave into a voltage having an intermediate, but usually superaudible, frequency which is then detected.

superhigh frequency—Abbreviated shf. The frequency band extending from 3,000 to 30,000 megahertz.

Supermalloy—Trade name of Arnold Engineering Company for a magnetic alloy with a maximum permeability greater than 1,000,000.

supermode laser—A frequency-modulated laser, the output of which is passed through a second phase modulator driven 180° out of phase and with the same modulation index as the first modulator. All of the energy of the previously existing laser modes is compressed into a single frequency, with nearly the full power of the laser concentrated in that signal.

superposed circuit—An additional channel obtained in such a manner from one or more circuits normally provided for other channels that all channels can be used simultaneously, without mutual interference.

superposition theorem—When a number of voltages (distributed in any manner throughout a linear network) are applied to the network simultaneously, the current that flows is the sum of the component currents that would flow if the same voltages had acted individually. Likewise, the potential difference that exists between any two points is the component potential difference that would exist there under the same conditions.

superpower—The comparatively large power (sometimes over 1,000,000 watts) used by a broadcasting station in its antenna.

superradiance—A rapid increase in intensity of fluorescent-line emission with increasing excitation (pump) power. This intensity increase and associated line narrowing are attributed to coherent reinforcement of spontaneously emitted photons during a single pass through the active region. (Laser term.)

superrefraction—Abnormally large refraction of radio waves in the lower layers of the atmosphere, leading to abnormal ranges of operation.

superregeneration—A form of regenerative amplification frequently used in radio-receiver detecting circuits. Oscillations are alternately allowed to build up and are quenched at a superaudible rate.

superregenerative detector—A detector that functions on superregeneration to achieve extremely high sensitivity with a minimum number of amplifier stages.

superregenerative receiver—A receiver in which the regeneration is varied in such a manner that the circuit is periodically rendered oscillatory and nonoscillatory.

supersensitive relay—A relay that operates on extremely small currents (usually less than 250 microamperes).

supersonic—Faster than the speed of sound (approximately 750 mph). These speeds are usually referred to by the term "mach" or "mach number." Mach 1 equals the speed of sound; mach 2, twice the speed of sound, etc.

supersonic communication—Communication through water by manually keying the sound output of echo-ranging equipment used on ships.

supersonic frequency—*See* Ultrasonic Frequency.

supersonics—1. The general term covering phenomena associated with speeds higher than that of sound (e.g., aircraft and projectiles which travel faster than sound). 2. The general term covering the use of frequencies above the range of normal hearing.

supersonic sounding—A system of determining ocean depths by measuring the time interval between the production of a supersonic wave just below the surface of the water, and the arrival of the echo reflected from the bottom. The sounds are transmitted and received by either magnetostriction or piezoelectric units, and electronic equipment is employed to provide a continuous indication of depth (sometimes with a permanent recording).

supersync signal—A combination horizontal- and vertical-sync signal transmitted at the end of each scanning line in commercial television.

superturnstile antenna—A stacked antenna array in which each element is a batwing antenna.

supervisory control—A system by which selective control and automatic indication of remote units is provided by electrical means over a relatively small number of common transmission lines. (Carrier-current channels on power lines can be used for this purpose.)

supervisory signal—A signal for attracting the attention of an attendant in connection with switching apparatus, etc.

supervoltage—A voltage applied to X-ray tubes operating between 500 and 2,000 kilovolts.

supply voltage—The voltage obtained from a power supply to operate a circuit.

suppressed carrier—That type of system which results in the suppression of the carrier frequency from the transmission medium. (The intelligence of a carrier wave after modulation is contained in either sideband, and normally only one sideband is transmitted; the other sideband and carrier frequency are suppressed. The intelligence is recovered at the receiving end by inserting a carrier frequency from a local source which, when combined with the incoming signal, produces the original frequencies with which the transmitting carrier was modulated.)

suppressed-carrier operation—*See* Suppressed-Carrier Transmission.

suppressed-carrier transmission — Transmission in which the carrier frequency is either partially or totally suppressed. One or both sidebands may be transmitted.

suppressed time delay—Deliberate displacement of the zero of the time scale with respect to the time of emission of a pulse, in order to simulate electrically a geographical displacement of the true position of a transponder.

suppressed-zero instrument—An indicating or recording instrument in which the zero position is below the end of the scale markings.

suppression—1. Elimination of any component of an emission—e.g., a particular frequency or group of frequencies in an audio- or radio-frequency signal. 2. Reduction or elimination of noise pulses generated by a motor or motor generator. 3. Elimination of unwanted signals or interference by means of shielding, filtering, grounding, component relocation, or sometimes redesign.

suppression pulse—The pulse generated in an airborne transponder by coincidence of the first interrogation pulse and the control pulse. This pulse, also known as a killer pulse, is used to suppress unwanted interrogations from the side lobes.

suppressor—1. A resistor used in an electron-tube circuit to reduce or prevent oscillation or the generation of unwanted r-f signals. 2. A resistor in the high-tension lead of the ignition system in a gasoline engine.

suppressor grid—A grid interposed between two positive electrodes (usually the screen grid and the plate) primarily to reduce the flow of secondary electrons from one to the other.

suppressor pulse—The pulse used to disable an ionized flow field or beacon transponder during intervals when interference would be encountered.

surface analyzer—An instrument that measures or records irregularities in a surface. As a crystal-pickup stylus or similar device moves over the surface, the resulting voltage is amplified and fed to an indicator or recorder that magnifies the surface irregularities as high as 50,000 times.

surface asperities—Small projecting imperfections on the surface coating of a tape that limit and cause variations in head-to-tape contact. A term useful in discussions of friction and modulation noise.

surface barrier—A barrier formed automatically at a surface by the electrons trapped there.

surface-barrier transistor — Abbreviated SBT. A wafer of semiconductor material into which depressions have been etched electrochemically on opposite sides. The emitter- and collector-base junction or metal-to-semiconductor contacts are then formed by electroplating a suitable metal onto the semiconductor in the etched depressions. The original wafer constitutes the base region.

surface-controlled avalanche transistor—A transistor in which the avalanche breakdown voltage is controlled by an external field applied through surface insulating layers, and which permits operation at frequencies up to the 10-GHz range.

surface duct—An atmospheric duct for which the lower boundary is the surface of the earth.

surface insulation—Also called oxalizing or insulazing. A coating applied to magnetic-core laminations to retard the passage of current from one lamination to another.

surface leakage—The passage of current over the boundary surfaces of an insulator, as distinguished from passage through its volume.

surface noise—Also called needle scratch. In mechanical recording, the noise caused in the electric output of a pickup by irregular contact surfaces in the groove.

surface of position—Any surface defined by a constant value of some navigational coordinate.

surface recombination rate—The rate at which free electrons and holes recombine at the surface of a semiconductor.

surface recording — Storage of information on a coating of magnetic material such as that which is used on magnetic tape, magnetic drums, etc.

surface reflection — Also called Fresnel loss. The part of the incident radiation that is reflected from the surface of a refractive material. It is directly proportional to the refractive index of the material and is reduced for a given wavelength by application of an appropriate surface coating.

surface resistance—Ratio of the d-c voltage applied to two electrodes on the surface of a specimen, to that portion of the current between the electrodes and in a thin layer of moisture or other semiconducting material that may be native to or deposited on the surface.

surface resistivity—The resistance between opposite edges of a surface film 1 cm square. It is measured by determining the resistance between two straight conductors 1 cm apart, pressed upon the surface of a slab of the material. Water-absorbent materials usually show a lower resistivity than nonabsorbent ones.

surface states—Discontinuities and contaminants at the surface of a semiconductor device which tend to change the surface characteristics and promote device parameter instability.

surface-temperature resistor—A platinum-resistance thermometer designed for installation directly on the surface whose temperature is being measured.

surface wave — A subclassification of the ground wave. So called because it travels along the surface of the earth.

surface-wave transmission line—A single-conductor line having a relatively thick dielectric sheath, often 3 or more times the diameter of the conductor, and a conical horn at each end exciting a mode of propagation on the line that is practically nonradiating.

surge — Sudden current or voltage changes in a circuit.

surge admittance—Reciprocal of surge impedance.

surge-crest ammeter—A special magnetometer used with magnetizable links to measure the crest value of transient electric currents.

surge generator—*See* Impulse Generator.

surge impedance—*See* Characteristic Impedance.

surge voltage (or current)—A large, sudden change of voltage (or current), usually caused by the collapse of a magnetic field or by a shorted or open circuit element.

surge-voltage recorder — *See* Lichtenberg Figure Camera.

surround—The part of a speaker cone by which its outside edge is anchored, usually corrugated.

surveillance—Systematic observation of air, surface, or subsurface areas by visual, electronic, photographic, or other means.

surveillance radar—In air-traffic control systems, a radar set or system used in a ground-controlled approach system to detect aircraft within a certain radius of an airport and to present continuously to the radar operator information as to the position, in distance and azimuth, of these aircraft.

surveillance radar station — In the aeronautical radionavigation service, a land station employing radar to detect the presence of aircraft.

susceptance—The reciprocal of reactance, and the imaginary part of admittance. It is measured in mhos.

susceptibility—1. Ratio of the induced magnetization to the inducing magnetic force. 2. The undesired response of an equipment to emissions, interference, or transients, or to signals other than those to which the equipment is intended to be responsive.

susceptibility meter—A device for measuring low values of magnetic susceptibility.

susceptiveness—The tendency of a telephone system to pick up noise and low-frequency induction from a power system. It is determined by telephone-circuit balance, transpositions, wire spacing, and isolation from ground.

suspension—A wire that supports the moving coil of a galvanometer or similar instrument.

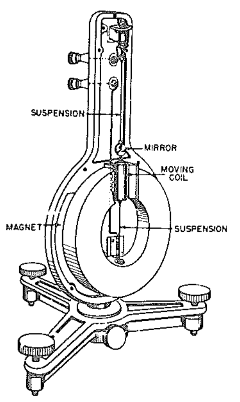

Suspension galvanometer.

suspension galvanometer—An early type of moving-coil instrument in which a coil of wire was suspended in a magnetic field and would rotate when it carried an electric current. A mirror attached to the coil deflected a beam of light, causing a spot of light to travel on a scale some distance from the instrument. The effect was a pointer of greater length but no mass.

suspension lines—The main lines leading from the canopy to the suspended load. Often mistakenly called "shroud" lines.

sustain—In an organ, the effect produced by circuitry which causes a note to diminish gradually after the key controlling the note has been released.

sustained oscillation—1. Oscillation in which forces outside the system but controlled by it maintain a periodic oscillation at a period or frequency that is nearly the natural period of the system. 2. Continued oscillation due to insufficient attenuation in the feedback path.

sustained start—An electrical signal for starting a timer which is of any duration longer than the timer setting.

sw—Abbreviation for short wave.

swamping resistor—In transistor circuits, a resistor placed in the emitter lead to mask (minimize the effects of) variations caused in the emitter-base junction resistance by temperature variations.

sweep—The crossing of a range of values of a quantity for the purpose of delineating, sampling, or controlling another quantity. Examples of swept quantities are the displacement of a scanning spot on the screen of a cathode-ray tube, and the frequency of a wave.

sweep accuracy—The accuracy of the trace horizontal displacement in an oscilloscope compared with the reference independent variable, usually expressed in terms of average rate error as a percent of full scale.

sweep amplifier—An amplifier stage designed to increase the amplitude of the sweep voltage.

sweep circuit—1. A circuit which produces, at regular intervals, an approximately linear, circular, or other movement of the beam in a cathode-ray tube. 2. The part of a cathode-ray oscilloscope that provides a time-reference base.

sweep delay—The time between the application of a pulse to the sweep-trigger input of an oscilloscope and the start of the sweep.

sweep-delay accuracy—The accuracy of an indicated sweep delay in an oscilloscope, usually specified in error terms.

sweep expander—*See* Sweep Magnifier.

sweep-frequency generator—A signal source capable of changing frequency automatically and in synchronism with a display device. The frequency sweep can be obtained by either mechanical or electronic means.

sweep-frequency record—A test record on which a series of constant-amplitude frequencies have been recorded. Each frequency is typically repeated 20 times per second, starting at 50 Hz and continuing up to 10 kHz or higher.

sweep generator — Also called timing-axis oscillator. A circuit which applies voltages or currents to the deflection elements in a cathode-ray tube in such a way that the deflection of the electron beam is a known function of time, against which other periodic electrical phenomena may be examined, compared, and measured.

sweeping receivers—Automatically and continuously tuned receivers designed to stop and lock on when a signal is found or to continually plot band occupancy.

sweep jammer—An electronic jammer which sweeps a narrow band of electronic energy over a broad bandwidth.

sweep linearity—The maximum displacement error of the independent variable between specified points on the display area in an oscilloscope.

sweep lockout—A means for preventing multiple sweeps when operating an oscilloscope in a single-sweep mode.

sweep magnifier—Also called sweep expander. A circuit or control for expanding part of the sweep display of an oscilloscope.

sweep oscillator—An oscillator used to develop a sawtooth voltage which can be amplified to deflect the electron beam of a cathode-ray tube. *Also see* Sweep Generator.

sweep switching—The alternate display of two or more time bases or other sweeps using a single-beam crt. Comparable to dual- or multiple-trace operation of a deflection amplifier.

sweep test—Pertaining to cable, checking the frequency response by generating an r-f voltage, the frequency of which is varied back and forth through a given frequency range at a rapid constant rate while observing the results on an oscilloscope.

sweep-through—A jamming transmitter that sweeps through a radio-frequency band and jams each frequency briefly, producing a sound like that of an aircraft engine.

sweep time—The reciprocal of sweep rate.

sweep voltage—The voltage used for deflecting an electron beam. It may be applied to either the magnetic deflecting coils or the electrostatic plates.

swell manual—Also called solo manual. In an organ, the upper manual normally used to play the melody. (*Also see* Manual.)

swept resistance—The portion of the total resistance of a potentiometric transducer over which the slider travels when the device is operated through its total range.

swing—The variation in frequency or amplitude of an electrical quantity.

swinging—Momentary variations in frequency of a received wave.

swinging arm—A type of mounting and feed used to move the cutting head at a uniform rate across the recording disc in some recorders. All phonograph pickups are of the swinging-arm type.

swinging choke—An iron-core choke coil designed so that its effective inductance varies with the current passing through it. It is used as the input choke in some power-supply filter circuits.

"swiss-cheese" packaging—Also called imitation 2-D. A high-density packaging technique in which passive and active components are inserted into holes punched in printed-circuit-board substrates and attached by soldering or thermocompression bonding or by means of conductive epoxy adhesive.

switch—A mechanical or electrical device that completes or breaks the path of the current or sends it over a different path.

switchboard—1. A manually operated apparatus at a telephone exchange. The various circuits from subscribers and other exchanges terminate here, so that operators can establish communications between two subscribers on the same exchange, or on different exchanges. 2. A single large panel or an assembly of panels on which are mounted the switches, circuit breakers, meters, fuses, and terminals essential to the operation of electrical equipment.

switch detector—A detector which extracts information from the input waveform only at instants determined by a selector pulse.

switch gear—A general term covering switching interrupting, control, metering, protective, and regulating devices; also assemblies of these devices and associated interconnections, accessories, and supporting structures, used primarily in connection with the generation, transmission, and distribution of electric power.

switching—Making, breaking, or changing the connections in an electrical circuit.

switching center — 1. A location at which data from an incoming circuit are routed to the proper outgoing circuit. 2. A group of equipment within a relay station for automatically or semiautomatically relaying communications traffic.

switching characteristics—An indication of how a device responds to an input pulse under specified driving conditions.

switching circuit—A circuit which performs a switching function. In computers, this is performed automatically by the presence of a certain signal (usually a pulse signal). When combined, switching circuits can perform a logical operation.

switching coefficient—The derivative of applied magnetizing force with respect to the reciprocal of the resultant switching time. It is usually determined as the reciprocal of the slope of a curve of reciprocals of switching times versus the values of applied magnetizing forces, which are applied as step functions.

switching device—Any device or mechanism, either electrical or mechanical, which can place another device or circuit in an operating or nonoperating state.

switching time—1. The interval between the reference time and the last instant at which the instantaneous-voltage response of a magnetic cell reaches a stated fraction of its peak value. 2. The interval between the reference time and the first instant at which the instantaneous integrated-voltage response reaches a stated fraction of its peak value.

switchplate—A small plate attached to a wall to cover a push-button or other type of switch.

swr—Abbreviation for standing-wave ratio.

swr bridge—*See* Standing-Wave-Ratio Bridge.

syllabic companding—Companding in which the effective gain variations are made at speeds allowing response to the syllables of speech but not to individual cycles of the signal wave.

syllable articulation—Also called percent-of-syllable articulation. The percent of articulation obtained when the speech units considered are syllables (usually meaningless and usually of the consonant-vowel-consonant type).

symbol—1. A simplified design representing a part in a schematic circuit diagram. 2. A letter representing a particular quantity in formulas. 3. A sign, mark, or drawing agreed upon to represent an electrical device or component part thereof.

symbolic—Having to do with the representation of something by a conventional sign.

symbolic address—Also called a floating address. In digital-computer programming, a label chosen in a routine to identify a particular word, function, or other information independent of the location of the information within the routine.

symbolic coding — In digital computer programming, any coding system using symbolic rather than actual computer addresses.

symbolic logic—A form of logic in which nonnumerical relationships for a computer are expressed by symbols suitable for calculation.

symbolic programming—A program using symbols instead of numbers for the operations and locations in a computer. Although the writing of a program is easier and faster, an assembly program must be used to decode the symbol into machine language and assign instruction locations.

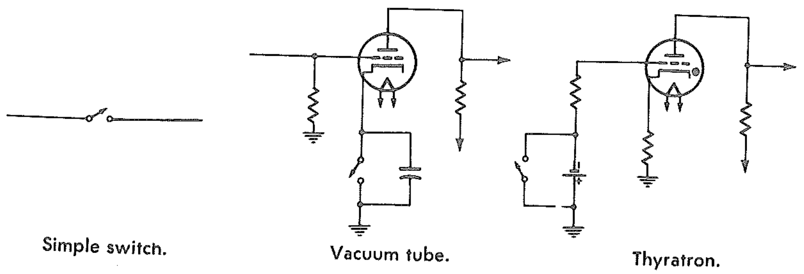
Simple switch.
Vacuum tube.
Thyratron.

Switching circuits.

symmetrical — Balanced—i.e., having equal characteristics on each side of a central line, position, or value.

symmetrical alternating quantity—An alternating quantity for which all values separated by a half period have the same magnitude but opposite sign.

symmetrical avalanche rectifier—An avalanche rectifier which can be triggered in either direction. After triggering, it presents a low impedance in the triggered direction.

symmetrically, cyclically magnetized condition—The condition of a cyclically magnetized material when the limits of the applied magnetizing forces are equal and of opposite sign.

symmetrical transducer (with respect to specified terminations)—A transducer in which all possible pairs of specified terminations can be interchanged without affecting the transmission.

symmetrical transistor — A transistor in which the collector and emitter are made identical, so either can be used interchangeably.

sync—Short for synchronous, synchronization, synchronizing, etc.

sync compression—The reduction in gain applied to the sync signal over any part of its amplitude range with respect to the gain at a specified reference level.

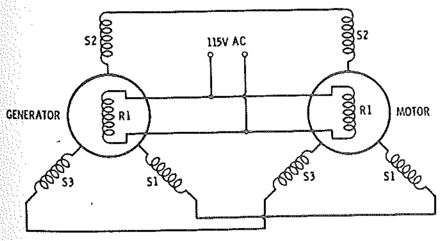

Synchro.

synchro—1. A small motorlike device containing a stator and a rotor. When several synchros are correctly connected together, all rotors will line up at the same angle of rotation. 2. An inductive device capable of transforming an angular-position input into an electrical output or an electrical input into an angular output.

synchro-control differential generator—A rotary component for modifying the synchro-control generator output signal to correspond to the addition or subtraction from the generator shaft angle. Usually used with a synchro-control generator and synchro-control transducer.

synchro-control generator—A rotary component for transforming the shaft angle to a corresponding set of electrical signals for ultimate retransformation to the shaft position in a remote location.

synchro-control transformer — A rotary component which accepts signals from a generator or differential generator for reconversion to the shaft angle with the aid of a servomechanism. Often used by itself as an angle-to-signal transducer, but usually with a synchro-control generator and synchro-control differential generator.

synchro differential generator—A synchro unit which receives an order from a synchro generator at its primary terminals, modifies this order mechanically by any desired amount according to the angular position of the rotor, and transmits the modified order from its secondary terminals to other synchro units.

synchro differential motor—A motor which is electrically similar to the synchro differential generator except that a damping device is added to prevent oscillation. Its rotor and stator are both connected to synchro generators, and its function is to indicate the sum or difference between the two signals transmitted by the generators.

synchro generator—A synchro which has an electrical output proportional to the angular position of its rotor.

Synchroguide—A type of control circuit for horizontal scanning in which the sync signal, oscillator voltage pulse, and scanning voltage are compared and kept in synchronism.

synchro motor—A synchro in which the rotor-shaft position is dependent on the electrical input.

synchronism—1. The phase relationship between two or more quantities of the same period when their phase difference is zero. 2. Applied to the synchronous motor, the condition under which the motor runs at a speed which is directly related to the frequency of the power applied to the motor and is not dependent upon other variables.

synchronization—1. The precise matching of two waves or functions. 2. The process of keeping the electron beam on the television screen in the same position as the scanning beam at the transmitter. 3. In a carrier, that degree of matching, in frequency, between the carrier used for modulation and the carrier used for demodulation which is sufficiently accurate to permit efficient functioning of the system.

synchronization error — In navigation, the error due to imperfect timing of two operations (may or may not include the signal transmission time).

synchronize—1. To adjust the periodicity of an electrical system so that it bears an integral relationship to the frequency of the periodic phenomenon under investigation. 2. To lock one element of a system into step with another. The term usually refers to locking a receiver to a transmitter, but it can refer to locking the data-terminal equipment bit rate to the data set frequency.

synchronized sweep—A sweep which would free-run in the absence of an applied signal, but in the presence of the signal is synchronized by it.

synchronizer—1. The component of a radar set which generates the timing voltage for the complete set. *Also see* Timer. 2. A computer storage device used to compensate for a difference in a rate of flow of information or time of occurrence of events when information is being transmitted from one device to another.

synchronizing (in television)—Maintaining two or more scanning processes in phase.

synchronizing-pulse selector — A circuit used to separate synchronizing pulses from commutated pulse trains.

synchronizing reactor—A current-limiting reactor that is connected momentarily across the open contacts of a circuit-interrupting device for synchronizing purposes.

synchronizing relay—A relay which functions when two alternating-current sources are in agreement within predetermined limits of phase angle and frequency.

synchronizing separator—*See* Amplitude Separator.

synchronizing signal—*See* Sync Signal.

synchronous—1. In step or in phase, as applied to two devices or machines. 2. A term applied to a computer, in which the performance of a sequence of operations is controlled by equally spaced clock signals or pulses.

synchronous booster converter — A synchronous converter connected in series with an a-c generator and mounted on the same shaft. It is used for adjusting the voltage at the commutator of the converter.

synchronous capacitor—A rotating machine running without mechanical load and designed so that its field excitation can be varied in order to draw a leading current (like a capacitor) and thereby modify the power factor of the a-c system, or influence the load voltage through such change in power factor.

synchronous clock—An electric clock driven by a synchronous motor, for operation on an a-c power system in which the frequency is accurately controlled.

synchronous computer—A digital computer in which all ordinary operations are controlled by equally spaced signals from a master clock.

synchronous converter—A synchronous machine which converts alternating current to direct current or vice versa. The armature winding is connected to the collector rings and commutator.

synchronous demodulator — Also called a synchronous detector. A demodulator in which the reference signal has the same frequency as the carrier or subcarrier to be demodulated. It is used in color television receivers to recover either the I or the Q signals from the chrominance sidebands.

synchronous detector—*See* Synchronous Demodulator.

synchronous gate—A time gate in which the output intervals are synchronized with the incoming signal.

synchronous generator—A circuit designed to synchronize an externally generated signal with a train of clock pulses. The generator produces precisely one output pulse for each cycle of the input signal. The output pulse thus has a width equal to that of the period of the clock-pulse train.

synchronous inverter—*See* Dynamotor.

synchronous machine—A machine which has an average speed exactly proportionate to the frequency of the system to which it is connected.

synchronous motor — An induction motor which runs at synchronous speed. Its stator windings have the same arrangement as in nonsynchronous induction motors, but the rotor does not slip behind the rotating magnetic stator field.

synchronous rectifier—A rectifier in which contacts are opened and closed at the correct instant by either a synchronous vibrator or a commutator driven by a synchronous motor.

synchronous speed—A speed value related to the frequency of an a-c power line and the number of poles in the rotating equipment. Synchronous speed in revolutions per minute is equal to the frequency in hertz divided by the number of poles, with the result multiplied by 120.

synchronous torque—The maximum load torque with which a motor can be loaded after it comes to synchronous speed. These torques are usually higher than starting torques.

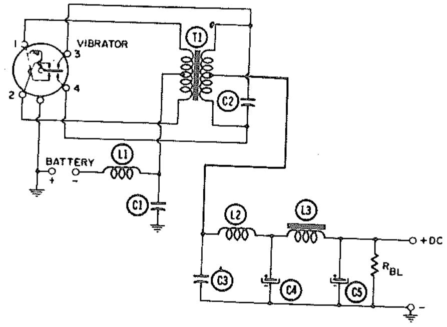

Synchronous-vibrator power supply.

synchronous vibrator—An electromagnetic vibrator that simultaneously converts a low d-c voltage to a low alternating voltage and applies it to a power transformer, from which a high alternating voltage is obtained and rectified. In power packs, it eliminates the need for a rectifier tube.

synchroscope—1. An instrument used to determine the phase difference or degree of

synchronism of two alternating-current generators or quantities. 2. An oscilloscope on which recurrent pulses or waveforms may be observed, and which incorporates a sweep generator that produces one sweep for each pulse.

synchro system—A system for obtaining remote indication or control by means of self-synchronizing motors such as selsyns and equivalent types.

synchro-torque receiver—A relatively low-impedance positioning device that generates its own torque when driven by a suitable synchro-torque transmitter.

synchro-torque transmitter—A positioning device that generates electrical information of sufficient power to drive a suitable torque receiver.

synchrotron — A device for accelerating charged particles (e.g., electrons) in a vacuum. The particles are guided by a changing magnetic field while being accelerated many times in a closed path by a radio-frequency electric field.

synchrotron radiation—Also called magnetic *bremsstrahlung*. The radiation produced by relativistic electrons as they travel in a region of space containing magnetic fields.

sync level—The level of the peaks of the sync signal.

sync limiter—A circuit used in television circuits to prevent sync pulses from exceeding a predetermined amplitude.

sync pulse—Part of the sync signal in a television system.

sync section—A color tv circuit comprising a keyer, burst amplifier, phase detector, reactance tube, subcarrier oscillator, and quadrature amplifier.

sync separator—The circuit which separates the picture signals from the control pulses in a television system.

sync signal—Also called a synchronizing signal. The signal employed for synchronizing the scanning. In television it is composed of pulses at rates related to the line and field frequencies.

sync-signal generator—A synchronizing signal generator for a television receiver or transmitter.

synthesizer frequency meter—A device for measuring frequency by utilizing a synthesized crystal-based signal for the internally generated signal.

synthetic display generation — Logical and numerical processing to display collected or calculated data in symbolic form.

syntony—The condition in which two oscillating circuits have the same resonant frequency.

system—1. An assembly of component parts linked together by some form of regulated interaction into an organized whole. 2. A collection of consecutive operations and procedures required to accomplish a specific objective.

systematic errors—Those errors which have an orderly character and can be corrected by calibration.

systematic inaccuracies—Those inaccuracies due to inherent limitations in the equipment.

system deviation—The instantaneous difference between the value of a specified system variable and the ideal value of the variable.

system element—One or more basic elements, together with other components necessary to form all or a significant part of one of the general functional groups into which a measurement system can be classified.

system engineering—A method of engineering approach whereby all the elements in the control system are considered, even the smallest value and the process itself.

system failure rate — The number of occasions during a given time period on which a given number of identical systems do not function properly.

system layout — In a microwave system, a chart or diagram showing the number, type, and terminations of circuits used in the system.

system noise—The output of a system when it is operating with zero input signal.

system of beams—The three electron beams emitted by the triple electron-gun assembly in a color tube. They occupy positions equidistant from a common axis and are spaced 120° apart around the axis.

system of units—An assemblage of units for expressing the magnitudes of physical quantities.

system overshoot—The largest value of system deviation following the dynamic crossing of the ideal value as a result of a specified stimulus.

system reliability — The probability that a system will perform its specified task properly under stated conditions of environment.

T

T—1. Symbol for transformer or absolute temperature. 2. Abbreviation for the prefix "tera" (10^{12}).

tab—*See* Land.

table lockup — In a computer, a method of controlling the location to which a jump or transfer is made. It is used especially when there is a large number of alternatives, as in function evaluation in scientific computations.

tabulator—A machine (e.g., a punch-card machine) which reads information from one medium and produces lists, totals, or tabulations on separate forms or continuous paper strips.

tachometer—An instrument used to measure

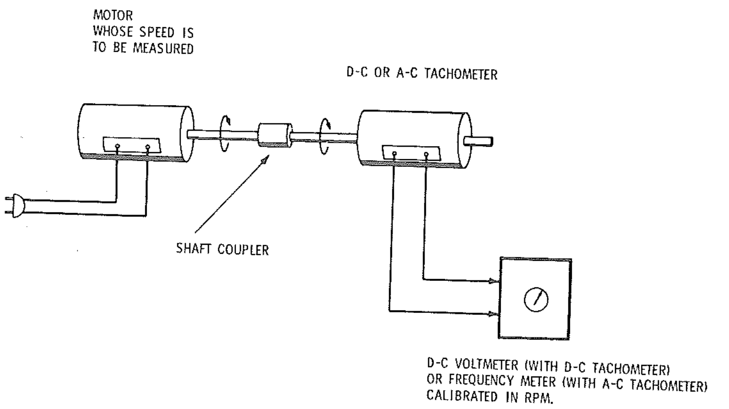

Tachometer.

the frequency of mechanical systems by the determination of angular velocity.

tachometer generator—A small generator attached to a rotating shaft for the purpose of generating a voltage proportional to the shaft speed. In speed-control circuits, a tachometer may be coupled to the shaft of the motor whose speed is to be controlled. A change of motor speed will produce a change of tachometer output voltage. This change of voltage can be used as an error signal to restore the speed to the desired value.

tag—Also called a sentinel. In digital-computer programming, a unit of information the composition of which differs from that of other members of the set so that it can be used as a marker or label.

tail—A small pulse following the main pulse and in the same direction, or the slow decay following the main body of the pulse.

tail clipping—A method of sharpening the trailing edge of a pulse.

tailing—*See* Hangover.

tail-warning radar sets—A radar set placed in the tail of an aircraft to warn of aircraft approaching from the rear.

take-up reel—The reel which accumulates the tape as it is recorded or played on a tape recorder.

talbot—A unit of luminous energy in the mksa system equal to 1 lumen-second.

talker echo—An echo which reaches the ear of the person who originated the sound.

talk-listen switch—A switch on an intercommunication unit to switch the speaker as required to function either as a reproducer or as a microphone.

tandem—Also called cascade. Two terminal-pair networks are in tandem when the output terminals of one network are connected directly to the input terminals of the other network.

tandem transistor—Two transistors in one package and internally connected together.

tangent—A straight line which touches the circumference of a circle at one point.

tangent galvanometer—A galvanometer consisting of a small compass mounted horizontally in the center of a large vertical coil of wire. The current through the coil is proportional to the tangent of the angle at which the compass needle is deflected.

tangential component—A component acting at right angles to a radius.

tangential sensitivity—A term generally applied as an indication of quality in a receiving system. This term can be used to define the minimum signal level that can be detected above the background noise. However, it is usually expressed as that signal power level which causes a 3-db rise above the noise-level reading.

tangential sensitivity on look-through—The strength of the target signal, measured at the receiver terminals, required to produce a signal pulse having twice the apparent height of the noise.

tangential wave path—In radio-wave propagation of a direct wave over the earth, a path which is tangential to the surface of the earth. The tangential wave path is curved by atmospheric refraction.

tangent sensitivity—The slope of the line tangent to the response curve at the point being measured.

tank—1. A unit of acoustically operating delay-line storage containing a set of channels. Each channel forms a separate recirculation path. 2. (*Also see* Tank Circuit.)

tank circuit—1. A circuit capable of storing electrical energy over a band of frequencies continuously distributed about a single frequency at which the circuit is said to be resonant or tuned. The selectivity of the circuit is proportionate to the ratio between the energy stored in the circuit and the energy dissipated. This ratio is often called the Q of the circuit. 2. Also called a tank. A parallel-resonant circuit connected in the plate circuit of an electron-tube generator.

tantalum (electrolytic) capacitor—An electrolytic capacitor with a tantalum foil or sintered-slug anode.

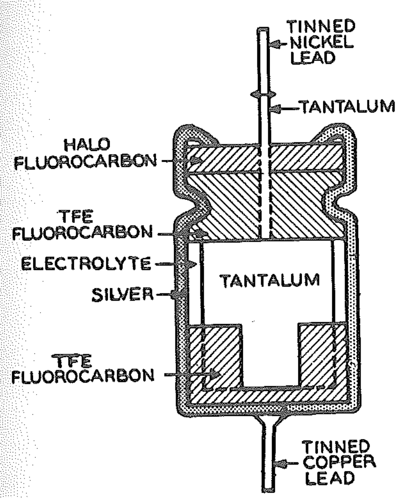

Tantalum capacitor.

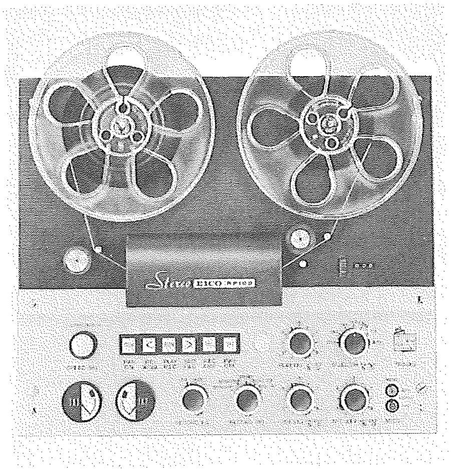

Tape deck.

tantalum-foil electrolytic capacitor—A capacitor which consists of two tantalum foil electrodes with an oxide on the anode, and separated by layers of absorbent paper saturated with an operating electrolyte.

T-antenna—Any antenna consisting of one or more horizontal wires, with the lead-in connected approximately in the center.

tap—1. A fixed electrical connection to a specified position on the element of a potentiometer, transformer, etc. 2. A branch. Applies to conductors such as a battery tap, and to miscellaneous general use.

tape—1. Magnetic recording tape with one or more recording strips (tracks). 2. (*Also see* Punched Tape.) 3. A ribbon of flexible material—e.g., friction, magnetic, punched, etc. 4. A linear medium for storing information which can be used as the input or output to a computer (e.g., magnetic tape).

tape cable—Also called flat flexible cable. A form of flexible multiple conductor in which parallel strips of metal are imbedded in an insulating material.

tape cartridge—A magazine or holder for a length of magnetic tape which by its design avoids the necessity for manual threading or handling. Usually compatible only with one specific type of machine. *Also see* Tape Magazine.

tape character—Information consisting of bits stored across the several longitudinal channels of a tape.

tape-controlled carriage—A paper-feeding device automatically controlled by a punched paper tape.

tape copy—A message received in tape form as the result of a transmission.

tape deck—A tape recorder designed for use in a high-fidelity music system. Usually it consists only of the motorboard mechanism and does not include a preamplifier, power amplifier, speaker, or case.

tape feed—A mechanism which feeds the tape to be read or sensed by a computer or other data-handling system.

tape guides—Grooved pins of nonmagnetic material mounted at either side of the recording-head assembly. Their function is to position the magnetic tape on the head as it is being recorded or played.

tape lifters—A system of movable guides which automatically divert tape from contact with the recorder heads during the fast forward or rewinding mode of operation.

tape loop—A length of magnetic tape with the ends joined together to form an endless loop. Used either on a standard recorder, a special message-repeater unit, or in conjunction with a cartridge device, it enables a recorded message to be played back repetitively; there is no need to rewind the tape.

tape magazine—Also called a tape cartridge. A container holding a reel of magnetic recording tape, which can be played without being threaded manually.

tape mark—A special record indicating end of file.

tape monitor—A circuit which permits the checking of recordings by taking the signal directly from the tape a moment after the recording is made. This is only possible on three-head recorders.

tape-on surface-temperature resistor—A surface-temperature resistor installed by adhering the sensing element to the surface with a piece of pressure-sensitive tape.

tape-path center line—The locus or path traced by an imaginary point, located on the recording tape midway between its edges, as it travels from reel to reel through guides, past heads, and between capstan and pressure rollers. For correct tracking and minimum tape distortion, the entire path should lie in a single plane located above the motor board at a height compatible with head-gap locations.

tape phonograph—*See* Tape Player.

tape player—Sometimes called a tape phonograph or a tape reproducer. A unit for playing recorded tapes. It has no facilities for recording.

taper—In communication practice, a continuous or gradual change in electrical properties with length—e.g., as obtained by a continuous change of cross section of a waveguide or by the distribution and change of resistance of a potentiometer or rheostat.

tape recorder—A mechanical-electronic device for recording voice, music, and other audio-frequency material. Sound is converted to electrical energy, which in turn sets up a corresponding magnetic pattern on iron-oxide particles suspended on paper or plastic tape. During playback, this magnetic pattern is reconverted into electrical energy and then changed back to sound through the medium of headphones or a speaker. The recorded material may be converted to a visual display by the use of an oscilloscope, other visual indicator, or a graphic recorder. Also used for magnetically recording the video material for television programs.

tapered potentiometer—A continuously adjustable potentiometer, the resistance of which varies nonuniformly along the element—being greater or less for equal slider movement at various points along the resistance element.

tapered transmission line — *See* Tapered Waveguide.

tapered waveguide—Also called a tapered transmission line. A waveguide in which a physical or electrical characteristic changes continuously with distance along the axis of the guide.

tape reproducer—*See* Tape Player.

tape reservoir—That part of a magnetic tape system used to isolate the tape storage inertia (i.e., tape reels, etc.) from the drive system.

tape skew—The deviation of a tape from following a linear path when transported across the heads, causing a time displacement between signals recorded on different tracks and amplitude differences between the outputs from individual tracks owing to variations in azimuth alignment. The adjectives "static" and "dynamic" are used to distinguish between the steady and fluctuating components of the tape skew.

tape speed—The speed at which tape moves past the head in the recording or playback mode. The standard tape speed for home use is 7½ ips or half this speed (3¾ ips). One-fourth or even one-eighth this speed is also used, but usually only for music when special high-quality tape is utilized. The professional recording speed for music mastering is usually 15 ips.

tape-speed errors—Any variation in tape speed from the normal speed over the record or reproduce head, regardless of cause.

tape-speed variations—*See* Flutter.

tape splicer—A device for splicing magnetic tape automatically or semiautomatically.

tape threader—A device that makes easier the threading of magnetic recording tape onto the reel.

tape-to-head speed—The relative speed of the tape and head during normal recording or replay. (The tape-to-head speed coincides with the tape speed in conventional longitudinal recording, but it is considerably greater than the tape speed in systems where the heads are scanned across or along the tape.)

tape-to-tape converter—A device for changing from one form of input/output medium or code to another, i.e., magnetic tape to paper tape (or vice versa) or eight-channel code to five-channel code.

tape transmitter—A machine for high-speed keying of telegraph code signals previously recorded on tape.

tape-transport mechanism — *See* Motor Board.

tape-wound core—A magnetic core consisting of a plastic or ceramic toroid around which is wound a strip of thin magnetic tape possessing a square-hysteresis-loop characteristic. Also known as bimag, a tape-wound core is used principally as a shift-register element.

tap lead—The lead connected to a tap on a coil winding.

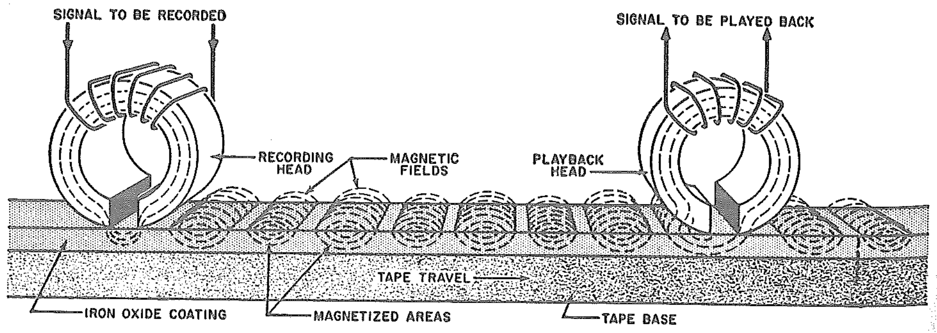

Tape recording.

tapped control—A rheostat or potentiometer having a fixed tap at some point along the resistance element, usually to provide fixed grid bias or automatic tone compensation.

tapped line — A delay line in which more than two terminal pairs are associated with a single sonic-delay channel.

tapped resistor—A wirewound fixed resistor having one or more additional terminals along its length, generally for voltage-divider applications.

tapped winding—A coil winding with connections brought out from turns at various points.

tap switch — A multicontact switch used chiefly for connecting a load to any one of a number of taps on a resistor or coil.

target—1. In a camera tube, a structure employing a storage surface which is scanned by an electron beam to generate an output-signal current corresponding to the charge-density pattern stored thereon. 2. Also called an anticathode. In an X-ray tube, an electrode or part of an electrode on which a beam of electrons is focused and from which X-rays are emitted. 3. In radar, a specific object of radar search or surveillance. 4. Any object which reflects energy back to the radar receiver.

target capacitance—In camera tubes, the capacitance between the scanned area of the target and backplate.

target cutoff voltage—In camera tubes, the lowest target voltage at which any detectable electrical signal, corresponding to a light image on the sensitive surface of the tube, can be obtained.

target discrimination—The characteristic of a guidance system that permits it to distinguish between two or more targets in close proximity.

target fade — The loss or decrease of signal from the target due to interference or other phenomena.

target glint—*See* Scintillation.

target identification—A visual procedure by which a radar target is positively identified as either hostile or friendly.

target language — The language into which some other language is to be properly translated.

target noise — Reflections of a transmitted radar signal from a target that has a number of reflecting elements randomly oriented in space.

target reflectivity—The degree to which a target reflects electromagnetic energy.

target scintillation — The apparent random movement of the center of reflectivity of a target observed during the course of an operation.

target seeker — In a missile homing system, the element that senses some feature of the target so that the resulting information can be used to direct appropriate maneuvers to maintain a collision course.

target voltage—In a camera tube with low-velocity scanning, the potential difference between the thermionic cathode and the backplate.

taut-band meter—An instrument the moving element of which is supported by two thin metal ribbons kept under tension by elastic members at each end of the assembly. This construction eliminates physical contact between the moving parts, and thus does away with friction.

tearing—Distortion observed on the television screen when the horizontal synchronization is unstable.

teaser transformer—A transformer of two T-connected, single-phase units for three-phase to two-phase or two-phase to three-phase operation; it is connected between the midpoint of the main transformer and the third wire of the three-phase system.

teasing — In the life-testing of switches, the slow movement of rotor contacts making and breaking with stator contacts.

technician license—A class of amateur radio license issued in the United States by the FCC for the primary purpose of operation and experimentation on frequencies above 50 MHz.

tecnetron—A high-power multichannel field-effect transistor similar to a triode tube in that it has anode and cathode connections (and a grid connection between) on opposite ends of a small germanium rod.

tee junction — Also spelled T-junction. A junction of waveguides in which the longitudinal guide axes form a T. The guide which continues through the junction is called the main guide; the one which terminates at a junction, the branch guide.

telautograph — Also called a telewriter. A writing telegraph instrument in which the movement of a pen in the transmitting apparatus varies the current and thereby causes the corresponding movement of a pen at the remote receiving instrument.

telecamera — Acronym for a TELEvision CAMERA.

telecardiophone—An amplifying stethoscope which permits heart sounds to be heard at a distance.

telecast — Acronym for TELEvision broadCASTing—specifically, a television program, or the act of broadcasting a television program.

telecasting—The broadcasting of a television program.

telecommunication—1. Any communication at a distance. Broadly, all forms of electrical transmission of intelligence. This includes the telegraph, telephone, radio, telephotograph, and television, listed in their chronological order of development. (*Also see* Communication.) 2. Data transmission between a computing system and remotely located devices via a unit that performs the necessary format conversion and controls the rate of transmission.

teleconference — A conference between persons who are remote from one another but linked together by a telecommunications system.

telegenic—The suitability of a subject or model for televising.

telegraph central office—A telephone switching unit installed in a telephone system providing service to the general public, having the necessary equipment and operating arrangements for terminating and interconnecting lines and trunks.

telegraph channel—The transmission media and intervening apparatus involved in the transmission of telegraph signals in a given direction between two intermediate telegraph installations. A means of one-way transmission of telegraph signals.

telegraph circuit—A complete circuit over which signal currents flow between transmitting and receiving apparatus in a telegraph system. It sometimes consists of an overhead wire or cable and a return path through the ground.

telegraph concentrator — A switching arrangement by means of which a number of branch or subscriber lines or station sets may be connected to a lesser number of trunk lines, operating positions, or instruments through the medium of manual or automatic switching devices to obtain more efficient use of facilities.

telegraph distributor—A device which effectively associates one direct current or carrier telegraph channel in rapid succession with the elements of one or more signal-sending or -receiving devices.

telegraph grade circuit—A circuit suitable for transmission by teletypewriter equipment. Normally the circuit is considered to employ d-c signalling at a maximum speed of 75 bauds.

telegraph key—A hand-operated device for opening and closing contacts to modulate the current with telegraph signals.

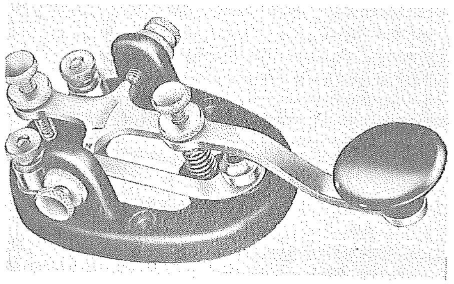

Telegraph key.

telegraph-modulated waves — Continuous waves, the amplitude or frequency of which is varied by means of telegraphic keying.

telegraph selector — A device which performs a switching operation in response to a definite signal or group of successive signals received over a controlling circuit.

telegraph signal distortion—The time displacement of transitions between conditions such as marking and spacing, with respect to their proper relative positions in perfectly timed signals. The total distortion is the algebraic sum of the bias and the characteristic and fortuitous distortions.

telegraph sounder—A telegraph receiving instrument in which an electromagnet attracts an armature each time a pulse arrives. This armature makes an audible sound as it hits against its stops at the beginning and end of each current impulse, and the intervals between these sounds are translated from code into the received message by the operator.

telegraph transmission speed—The rate at which signals are transmitted. This may be measured by the equivalent number of dot-cycles per second or by the average number of letters or words transmitted and received per minute.

telegraph transmitter—A device for controlling a source of electric power in order to form telegraph signals.

telegraphy—1. A system of telecommunication for the transmission of graphic symbols, usually letters or numerals, by the use of a signal code. It is used primarily for record communication. 2. Any system of telecommunication for the transmission of graphic symbols or images for reception in record form, usually without gradation of shade values.

telelectrocardiograph—A device for transmission and remote reception of electrocardiograph signals.

telemeter—1. To transmit analog or digital reports of measurements and observations over a distance (e.g., by radio transmission from a guided missile to a control or recording station on the ground). 2. A complete measuring, transmitting, and receiving apparatus for indicating, recording, or integrating the value of a quantity at a distance by electric translating means.

telemetering—Also called telemetry or remote metering. Measurement which, through intermediate means, can be interpreted at a distance from the primary detector. A receiving instrument converts the transmitted electrical signals into units of data, which can then be translated by data reduction into appropriate units.

telemeter service—Metered telegraph transmission between paired telegraph instruments over an intervening circuit adapted to serve a number of such pairs on a shared-time basis.

telemetry—The science of sensing and measuring information at some remote location and transmitting the data to a convenient location to be read and recorded.

telemetry beacon—A system whereby two or more reply pulses are transmitted by the beacon for the transmission of data from the test vehicle to the ground station.

telemetry frame—In pcm systems, one complete sampling of words or channels of information at a given rate; in time-division multiplexing, one complete commutator revolution.

telemetry-frame rate—The frequency derived from the period of one frame.

telephone — Combination of apparatus for converting speech energy to electrical waves, transmitting the electrical energy to a distant point, and there reconverting the electrical energy to audible sounds.

telephone capacitor—A fixed capacitor connected in parallel with a telephone receiver to bypass r-f and higher audio frequencies and thereby reduce noise.

telephone carrier current—A carrier current used for telephone communication so that more than one channel can be obtained on a single pair of wires.

telephone channel—A channel suitable for the transmission of telephone signals.

telephone circuit—A complete circuit over which audio and signaling currents travel between two telephone subscribers communicating with each other in a telephone system.

telephone current—An electric current produced or controlled by the operation of a telephone transmitter.

telephone jack—*See* Phone Jack.

telephone pickup—Any of several devices used to monitor telephone conversations, usually without direct connection to the telephone line and operating on the principle of mutual magnetic coupling.

telephone plug—*See* Phone Plug.

telephone receiver—The earphone used in a telephone system.

telephone repeater—An assemblage of amplifiers and other equipment employed at points along the line to rebuild the signal strength in a telephone circuit.

telephone ringer—An electric bell that operates on low-frequency alternating or pulsating current and is used for indicating a telephone call to the station being alerted.

telephone system—A group of telephones plus the lines, trunks, switching mechanisms, and all other accessories required to interconnect the telephones.

telephone transmitter—A microphone used in a telephone system. (*Also see* Microphone.)

telephony—The transmission of speech current over wires, enabling two persons to converse over almost any distance.

telephoto — Also called telephotography. A photoelectrical transmission system for point-to-point or air-to-ground transmission of high-definition pictorial information.

telephotography—*See* Telephoto.

telephoto lens—A lens system that is physically shorter than its rated focal length. It is used in still, movie, and television cameras to enlarge images of objects photographed at comparatively great distances.

teleprinter — 1. *See* Printer. 2. Trade name used by Western Union for its telegraph terminal equipment.

teleran—A navigational system in which radar and television transmitting equipment are employed on the ground, with television receiving equipment in the aircraft, to televise the image of the ground radar ppi scope to the aircraft along with map and weather data.

telering—A frequency-selector device for the production of ringing power.

telesynd—Telemeter or remote-control equipment which is synchronous in both speed and position.

teletype—A system used for transmitting messages over some distance, employing keyboard or paper-tape sending and printed receiving.

teletypewriter—1. *See* Printer. 2. A generic term referring to the basic equipment made by Teletype Corporation and to teleprinter equipment. The teletypewriter uses electromechanical functions to generate codes (Baudot) in response to a human input to a manual keyboard.

teletypewriter code—A special code in which each code group is made up of five units, or elements, of equal length which are known as marking or spacing impulses. The five-unit start-stop code consists of five signal impulses preceded by a start impulse and followed by a stop impulse. Each impulse except the stop impulse is 22 milliseconds in length; the stop impulse is 32 milliseconds (based on 60-word-per-minute operation).

televise—The act of converting a scene or image field into a television signal.

television—Abbreviated tv. A telecommunication system for transmission of transient images of fixed or moving objects.

television and radar navigation—A navigational system which; a. employs ground-based search radar equipment along an airway to locate aircraft flying near that airway, b. transmits, by television means, information pertaining to these aircraft and other information to the pilots of properly equipped aircraft, and c. provides information to the pilots appropriate for use in the landing approach.

television broadcast band—The frequencies assignable to television broadcast stations in the band extending from 54 to 890 MHz. These frequencies are grouped into channels, as follows: Channels 2 through 4, 54 to 72 MHz; Channels 5 and 6, 76 to 88 MHz; Channels 7 through 13, 174 to 216 MHz; and Channels 14 through 83, 470 to 890 MHz.

television broadcast station—A radio station for transmitting visual signals, and usually simultaneous aural signals, for general reception.

television camera—A pickup unit used in a television system to convert into electric signals the optical image which is formed by a lens.

television channel—A channel suitable for the transmission of television signals. The channel for associated sound signals may or may not be considered part of the television channel.

television engineering—*See* Radio Engineering.

television interference—Interference in the reception of the sound and/or video portion of a television program by a transmitter or another device.

television pickup station—A land mobile station used for the transmission of television program material and related communications from the scenes of events ocurring at points removed from television broadcast station studios, to the television broadcast stations.

television receiver—A radio receiver for converting incoming electric signals into television pictures and the associated sound.

television reconnaissance — Air reconnaissance by optical or electronic means to supplement photographic and visual reconnaissance.

television relay system—A system of two or more stations for transmitting television relay signals from point to point, using radio waves in free space as a medium. Such transmission is not intended for direct reception by the public.

television repeater—A repeater used in a television circuit.

television screen—In a television receiver, the fluorescent screen of the picture tube.

television signal—The audio signal and video signal that are broadcast simultaneously to produce the sound and picture portions of a televised scene.

television transmitter—The aggregate radio-frequency and modulating equipment necessary to supply, to an antenna system, the modulated radio-frequency power by which all component parts of a complete television signal (including audio, video, and synchronizing signals) are concurrently transmitted.

televoltmeter — A telemeter that measures voltage.

telewattmeter — A telemeter that measures power.

telewriter—*See* Telautograph.

telex—An audio-frequency teleprinter system used in Great Britain to provide teletypewriter service over telephone lines.

telluric current—*See* Earth Current.

Telstar—A low-altitude active communications satellite used for microwave communication and satellite tracking.

temperature coefficient—1. A factor used to calculate the change in the characteristics of a substance, device, or circuit element with changes in its temperature. 2. The percentage change in the output voltage (or current) of a regulated power supply due to a variation of ambient temperature. The values are usually expressed as a percentage per degree Celsius and restricted to the specified ambient range of the unit.

temperature coefficient of frequency—The rate at which the frequency changes with temperature, expressed in hertz per megahertz per degree Celsius at a given temperature.

temperature coefficient of permeability—A coefficient expressing the change in permeability as the temperature rises or falls. It is expressed as the rate of change in permeability per degree.

temperature coefficient of resistance—The ratio of the change in resistance (or resistivity) to the original value for a unit change in temperature. The temperature coefficient over the temperature range from t to t_1, referred to the resistance R_t at temperature t, is determined by the following equation:

$$\frac{R_{t1} - R_t}{R_t\,(t_1 - t)}$$

where t is the temperature, preferably in degrees C. The value will be positive unless otherwise indicated by a negative sign.

temperature coefficient of voltage drop — The change in the voltage drop of a glow-discharge tube, divided by the change in ambient temperature or in the temperature of the envelope.

temperature coefficient value—The expected percentage change per degree of temperature difference from a specified temperature.

temperature-compensated zener diode—A positive temperature-coefficient reversed-bias zener diode (PN junction) connected in series with one or more negative-temperature forward-biased diodes within a single package.

temperature-compensating capacitor—A capacitor the capacitance of which varies with temperature in a known and predictable manner. It is used extensively in oscillator circuits, to compensate for changes due to temperature variations in the values of other components.

temperature compensation—The process whereby the effects of an increase or decrease in ambient temperature are canceled (e.g., as in the case of an oscillator that is required to maintain a stable output frequency regardless of ambient temperature changes).

temperature control — 1. A switch actuated by a thermostat responsive to changes in temperature, and used to maintain temperature within certain limits. 2. A control device responsive to temperature.

temperature derating—Lowering the voltage, current, or power rating of a device or component when it is used at elevated temperatures.

temperature detector—An instrument used to measure the temperature of a body. Any physical property that is dependent on temperature may be employed, such as the differential expansion of two bodies, thermoelectromotive force at the junction of two metals, change of resistance of a metal, or the radiation from a hot body.

temperature-limited — The condition of a cathode when all the electrons emitted from it are drawn away by a strong positive field. The only way to increase the flow of electrons is to raise the cathode temperature.

temperature relay—A relay which functions at a predetermined temperature.

temperature rise—The difference between the initial and final temperature of a component or device. Temperature rise is expressed in degrees C or F, usually referred to an ambient temperature, and equals the hot-spot temperature minus the ambient temperature.

temperature saturation—*See* Filament Saturation.

temperature shock—A rapid change from one temperature extreme to another.

temperature-wattage characteristic — In a thermistor, the relationship, for a specific ambient temperature, between the temperature of the thermistor and the applied steady-state power.

temporary magnet—A magnetized material having a high permeability and low retentivity.

temporary storage — Internal storage locations in a computer reserved for intermediate and partial results.

TEM wave—Abbreviation for transverse electromagnetic wave.

tensiometer—A device for determining the tautness of a supporting wire or cable.

tension — 1. Mechanical—the condition of strain which tends to stretch. 2. Electrical—the potential or electrostatic voltage.

tenth-power width—In a plane containing the direction of the maximum of a lobe, the full angle between the two directions in that plane, about the maximum in which the radiation intensity is one-tenth the maximum value of the lobe.

tera—Prefix for the numerical quantity of 10^{12}. Abbreviated T.

terahertz—One million megahertz, or 10^{12} hertz. Abbreviated THz.

teraohm — One million megohms, or 10^{12} ohms.

teraohmmeter — An instrument used to measure extremely high resistance.

terminal—1. A point of connection for two or more conductors in an electrical circuit. 2. A device attached to a conductor to facilitate connection with another conductor.

terminal area—1. A portion of a microelectronic circuit used for making electrical connections to the conductor pattern—e.g., an enlarged pad (area) on a semiconductor die. *Also see* Pad. 2. A portion of a printed circuit used for making electrical connections to the conductive pattern, such as the enlarged portion of conductor material surrounding a component mounting hole.

terminal block—An insulating base or slab equipped with one or more terminal connectors for the purpose of making electrical connections thereto.

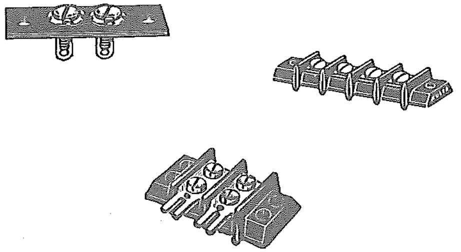

Terminal boards.

terminal board—Also called a terminal strip. An insulating base or slab equipped with terminals for connecting wiring.

terminal box—A housing where cable pairs are brought out to terminations for connections.

terminal brush—A brush with long bristles for cleaning fuses and terminals in a terminal box.

terminal cutout pairs—Numbered, designated pairs brought out of a cable at a terminal.

terminal equipment—1. At the end of a communications channel, the equipment essential for controlling the transmission and/or reception of messages. 2. Telephone and teletypewriter switchboards and other centrally located equipment to which wire circuits are terminated.

terminal guidance—1. Guidance applied to a guided missile between midcourse guidance and arrival at the target. 2. Electronic, mechanical, visual, or other assistance given an aircraft pilot to facilitate arrival at, landing upon, or departure from an air landing or air-drop facility.

terminal impedance—The complex impedance seen at the unloaded output or input terminals of transmission equipment or a line in otherwise normal operating condition.

terminal leg—*See* Terminal Stub.

terminal lug—1. A threaded lug to which a wire may be fastened in a terminal box. 2. A cylindrical piece of metal, either solid or hollow and of two or more diameters, which can be stacked, flared, swaged, or pressed into a hole for the purpose of connecting leads or external wires to the conductive pattern.

terminal pad—An alternate term for "terminal area" or "pad."

terminal pair—An associated pair of accessible terminals (e.g., the input or output terminals of a device or network).

terminal repeater — 1. An assemblage of equipment designed specifically for use at the end of a communication circuit—as contrasted with the repeater, which is designed for an intermediate point. 2. Two microwave terminals arranged to provide for the interconnection of separate systems, or separate sections of a system.

terminal station — The microwave equipment and associated multiplex equipment employed at the ends of a microwave system.

terminal strip—*See* Terminal Board.

terminal stub—Also called terminal leg. A piece of cable which comes with a cable terminal for splicing into the main cable.

terminal unit—Equipment usable on a communication channel for either input or output.

terminal vhf omnirange — Very-high-frequency omnirange, normally low powered, complete with a local monitoring device which will automatically shut down the facility if it is not operating properly.

terminated line—A transmission line terminated in a resistance equal to the characteristic impedance of the line, so that there is no reflection or standing waves.

terminating—The closing of the circuit at either end of a line or transducer by connection of some device. Terminating does not imply any special condition, such as the elimination of reflection.

termination—1. A load connected to a transmission line or other device. To avoid wave reflections, it must match the characteristic of the line or device. 2. A waveguide technique; the point at which energy flowing along a waveguide continues in a nonwave guide mode of propagation.

ternary — 1. A numerical system of notation using the base 3 and employing the characters 0, 1, and 2. 2. Able to assume three distinct states.

ternary code—A code in which each element may be any one of three distinct kinds or values.

ternary pulse-code modulation—A form of pulse-code modulation in which each element of information is represented by one of three distinct values, e.g., positive pulses, negative pulses, and spaces.

terrain-avoidance radar — Airborne radar which provides a display of terrain ahead of a low-flying airplane to permit horizontal avoidance of obstacles.

terrain-clearance indicator—A device for measuring the distance from an aircraft to the surface of the sea or earth.

terrain error—In navigation, the error resulting from distortion of the radiated field by the nonhomogeneous characteristics of the terrain over which the radiation in question has been propagated.

terrain-following radar — Airborne radar which provides a display of terrain ahead of a low-flying aircraft to permit manual control, or signals for automatic control, to maintain constant altitude above the ground.

terrestrial-reference flight—Stabilized flight in which control information is obtained from terrestrial phenomena (e.g., flight in which basic information derived from the magnetic field of the earth, atmospheric pressure, and the like is fed into a conventional automatic pilot).

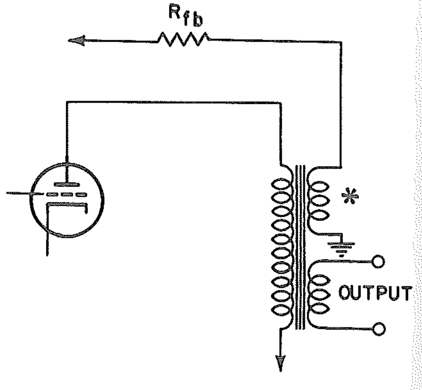

Tertiary coil.

tertiary coil—A third coil used in the output transformer of an audio amplifier to supply a feedback voltage.

tertiary winding—1. A winding added to a transformer, in addition to the conventional primary and secondary windings, to suppress third harmonics or to make connections to a power-factor–correcting device. 2. *See* Stabilized Winding.

tesla—A unit of magnetic induction equal to 1 weber per square meter in the mksa system.

Tesla coil—An air-core transformer used for developing high-voltage discharge at a very high frequency. It has a few turns of heavy wire as the primary and many turns of fine wire as the secondary.

test—A procedure or sequence of operations for determining the manner in which equipment is functioning or the existence, type, and location of any trouble.

test bench—Equipment designed specifically for making over-all bench tests on equipment in a particular test set-up under controlled conditions.

test board—A switchboard equipped with testing apparatus, arranged so that connections can be made from it to telephone lines or central office equipment for testing purposes.

test clip—A spring clip fastened to the end of an insulated wire to enable quick temporary connections when circuits or devices are being tested.

testing level—The value of power used for reference represented by 0.001 watt working into 600 ohms.

test lead — A flexible insulated lead used chiefly for connecting meters and test instruments to a circuit under test.

test oscillator—A test instrument that can be set to generate an unmodulated or tone-modulated radio-frequency signal at any frequency needed for aligning or servicing receivers and/or amplifiers.

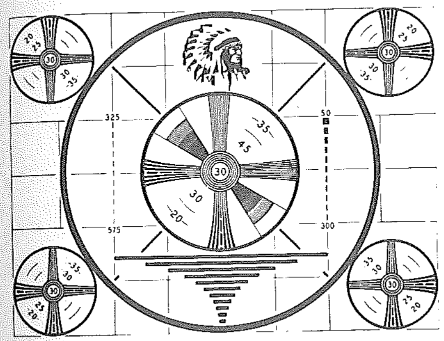

Test pattern.

test pattern—A geometric pattern containing a group of lines and circles, and used for testing the performance of a television receiver or transmitter by revealing the following video-signal characteristics: horizontal linearity, vertical linearity, contrast, aspect ratio, interlace, streaking, ringing, vertical resolution, and horizontal resolution.

test prod—A sharp metal point used for making a touch connection to a circuit terminal. It has an insulated handle and a means for electrically connecting the point to a test lead.

test record—A phonograph disc designed to test the quality and characteristics of turntables, pickups, amplifiers, etc.

test routine (in a computer)—1. A synonym for check routine. 2. Generally both the check and the diagnostic routines.

test set—One or more instruments required for servicing of a particular type of equipment.

tetrad—A group of four, especially a group of four pulses used to express a digit in the scale of 10 or 16.

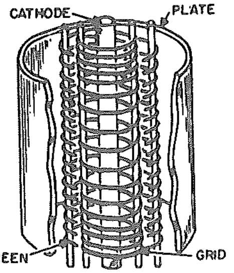

Tetrode.

tetrode—A four-electrode electron tube containing an anode, a cathode, a control electrode (grid), and one additional electrode that is ordinarily a screen grid.

tetrode junction transistor—*See* Double-Base Junction Transistor.

tetrode transistor — A junction transistor with two electrode connections to the base to reduce interelement capacitance (in addition to the normal emitter and collector elements, each having one connection).

Te value—The temperature at which the resistance of a centimeter cube is of 1 megohm.

TE wave—Abbreviation for transverse electric wave.

T flip-flop—Also called binary. A type of flip-flop the outputs of which change state each time the input-signal voltage falls from 1 to 0 and remain unchanged when the input-signal voltage rises from 0 to 1. Thus, there is one change in output state for every two changes in input signal. Thus the frequency of the output is half the frequency of the input.

thallofide cell — A photoconducting cell which has thallium oxysulfide as the light-sensitive agent.

theoretical acceleration at stall—A figure of merit derived from the stall-torque to rotor-inertia ratio, which indicates how rapidly a motor will accelerate from stall.

theoretical cutoff — *See* Theoretical Cutoff Frequency.

theoretical cutoff frequency—Also called theoretical cutoff. The frequency at which, disregarding the effects of dissipation, the attenuation constant of an electric structure changes from zero to a positive value or vice versa.

theoretical electrical travel—The shaft travel over which the theoretical function characteristic of a precision potentiometer extends, as determined from the index point.

theremin—An electronic musical instrument consisting of two radio-frequency oscillators which beat against each other to produce an audio-frequency tone, in a manner similar to a beat-frequency audio oscillator. The pitch and volume are varied by hand capacitance.

thermal—A general term for all forms of thermoelectric thermometers, including a series of couples, thermopiles, and single thermocouples.

thermal agitation—1. Movement of the free electrons in a material. In a conductor they produce minute pulses of current. When these pulses occur at the input of a high-gain amplifier in the conductors of a resonant circuit, the fluctuations are amplified together with the signal currents and heard as noise. 2. Also called thermal effect. Minute voltages arising from random electron motion, which is a function of absolute temperature expressed in degrees Kelvin.

thermal-agitation voltage — The potential difference produced in circuits by thermal agitation of the electrons in the conductor.

thermal ammeter—*See* Hot-Wire Ammeter.

thermal breakdown — 1. A form of breakdown in which decomposition or melting occurs due to the temperature rise resulting from the applied electric stress. 2. A runaway condition in a dielectric, the loss factor of which increases with temperature. Dielectric loss heats the material, producing an increase in temperature. Therefore, the dielectric loss increases still more, producing a further increase in temperature, and so on.

thermal circuit breaker—A circuit breaker the operation of which depends on temperature expansion due to electrical heating.

thermal coefficient of resistance – The change in the resistivity of a substance due to the effects of temperature only. Usually expressed in ohms per ohm per degree change in temperature.

thermal compensation—A method employed to reduce or eliminate the thermal effects on one or more of the performance parameters of a transducer.

thermal conduction—The transfer of thermal energy by processes having no net movement of mass and having rates proportional to the temperature gradient.

thermal conductivity—A measure of the ability of a substance to conduct heat. Expressed in terms of calories of heat conducted per second per sq. cm. per cm. of thickness per degree C difference in temperature from one surface to the other.

thermal conductor—A material which readily transmits heat by conduction.

thermal contraction—The shrinkage exhibited by most metals when cooled.

thermal converter—Also called thermocouple converter, thermoelectric converter, thermoelectric generator, or thermoelement. One or more thermojunctions in thermal contact with an electric heater or integral, so that the electromotive force developed by thermoelectric action at its output terminals gives a measure of the input current in its heater.

thermal cutout—An overcurrent protective device which contains a heater element that affects a fusible member and thereby opens the circuit.

thermal detector—*See* Bolometer.

thermal drift—A change in the output of a regulated power supply over a period of time, due to changes in internal ambient temperatures not normally related to environmental changes. Thermal drift is usually associated with changes in line voltage and/or load changes.

thermal effect—*See* Thermal Agitation.

thermal emf—The electromotive force generated when the junction of two dissimilar metals is heated. *Also see* Seebeck EMF.

thermal endurance—1. An indication of the relative life expectancy of a product when exposed to operating temperatures much higher than normal room temperature. 2. *See* Heat Endurance.

thermal expansion—The increase in length caused by heating, generally expressed in inches of increase per inch of original length per degrees of temperature.

thermal flasher—An electric device that automatically opens and closes a circuit at regular intervals, owing to alternate heating and cooling of a bimetallic strip heated by a resistance element in series with the circuit being controlled.

thermal generation—The creation of a hole and a free electron by freeing a bound electron through the addition of heat energy.

thermal instrument—An instrument that depends on the heating effect of an electric current for its operation (e.g., thermocouple and hot-wire instruments).

thermal ionization—Ionization due to high temperature (e.g., in the electrically conducting gases of a flame).

thermal junction—*See* Thermocouple.

thermal lag—The time expended in raising the entire mass of a cathode structure to the temperature of the heater.

thermal life—The operating life of a device under varying ambient temperatures.

thermal microphone—A microphone depending for its action on the variation in the resistance of an electrically heated conductor that is being alternately increased and decreased in temperature by sound waves.

thermal noise—1. Also called resistance noise. Random circuit noise associated with the thermodynamic interchange of energy necessary to maintain thermal equilibrium between the circuit and its surroundings. (*Also see* Johnson Noise.) 2. Noise generated by the random thermal motion of charged particles.

thermal protector—A current- and temperature-responsive device used to protect another device against overheating due to overload.

thermal radiation – Commonly known as heat. Radiation produced by the action of heat on molecules or atoms. Its frequency extends between the extremes of infrared and ultraviolet.

thermal rating—A statement of the permissible temperature rating beyond which unsatisfactory performance occurs.

thermal relay—A relay that responds to the heating effect of an energizing current, rather than to the electromagnetic effect.

thermal resistance—1. Ratio of the temperature rise to the rate at which heat is generated within a device under steady-state conditions. 2. The resistance of a substance to the conductivity of heat.

thermal resistor—An electronic device which makes use of the change in resistivity of a semiconductor with changes in temperature.

thermal response time—The time from the occurrence of a step change in power dissipation until the junction temperature reaches 90% of the final value of junction-temperature change, when the device-case or ambient temperature is held constant.

thermal runaway—A condition in which the dissipation in a transistor or other device increases so rapidly with higher temperature that the temperature keeps on rising.

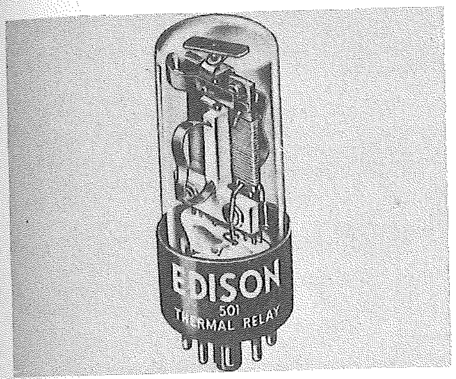

Thermal relay.

thermal sensitivity set – A permanent change in sensitivity due to temperature effects only. Usually expressed as the difference in sensitivity at room temperature before and after a temperature cycle over the operating temperature range of the transducer.

thermal shock—1. A sudden, marked change in the temperature of the medium in which a component or device operates. 2. The effect of heat or cold applied at such a rate that nonuniform thermal expansion or contraction occurs within a given material or combination of materials.

thermal telephone receiver – Also called thermophone. An electroacoustic transducer, such as a telephone receiver in which the temperature of a conductor is caused to vary in response to the current input, thereby producing sound waves as a result of the expansion and contraction of the adjacent air.

thermal time constant – 1. The time from the occurrence of a step change in power dissipation until the junction temperature reaches 63.2% of the final value of junction-temperature change, when the device-case or ambient temperature remains constant. 2. In a thermistor, the time required for 63.2% of the change from initial to final body temperature after the application of a step change in temperature under zero-power conditions.

thermal time-delay relay—A type of relay in which the time interval between energization and actuation is determined by the thermal storage capacity of the actuator critical operating temperature, power input, and thermal insulation.

thermal time-delay switch—An overcurrent-protective device containing a heater element and thermal delay.

thermal tuning—Adjusting the frequency of a cavity resonator by using thermal expansion to vary its shape.

thermal tuning time (cooling)—In microwave tubes, the time required to tune through a specified frequency range when the tuner power is instantaneously changed from the specified maximum to zero. The initial condition must be one of equilibrium.

thermal tuning time (heating) – In microwave tubes the time required to tune through a specified frequency range when the tuner power is instantaneously changed from zero to the specified maximum. The initial condition must be one of equilibrium.

thermion—An ion, either positive or negative, which has been emitted from a heated body. Negative thermions are electrons (thermoelectrons).

thermic—Pertaining to heat.

thermionic—Pertaining to the emission of electrons by heat.

thermionic cathode—*See* Hot Cathode.

thermionic converter—Also called thermionic generator or thermoelectron engine. A device which produces electrical power directly from heat. One type contains a heated cathode to emit electrons and a cold anode to collect them, thereby causing a flow of current. Both electrodes are enclosed in a vacuum or gas-filled envelope.

thermionic current—Current due to directed movements of thermions (e.g., the flow of electrons from the cathode to the plate in a thermionic vacuum tube).

thermionic detector—A detector circuit in which a thermionic vacuum tube delivers an audio-frequency signal when fed with a modulated radio-frequency signal.

thermionic diode – A diode electron tube which has a heated cathode.

thermionic emission—Liberation of electrons due to the temperature rise of the cathode alone, independent of any other element within the tube.

thermionic energy conversion—The direct production of electricity by means of the electron emission from a heated substance.

thermionic generator—*See* Thermionic Converter.

thermionic grid emission—Also called primary grid emission. The current produced by the electrons thermionically emitted from a grid. Generally it is due to excessive grid temperatures or to contamination of the grid wires by cathode-coating material.

thermionic rectifier—A rectifier utilizing a thermionic vacuum tube to convert alternating current into unidirectional current.

thermionic tube—*See* Hot-Cathode Tube.

thermionic work function—The energy required to transfer electrons from a given metal to a vacuum or some other adjacent medium during thermionic emission.

thermistor—A solid-state semiconducting device, the electrical resistance of which varies with the temperature. Its temperature coefficient of resistance is high, nonlinear, and negative.

thermoammeter—Also called a thermocouple

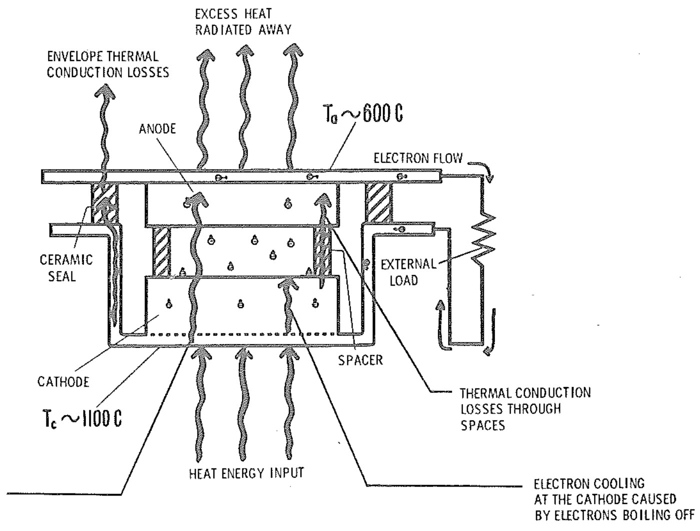
EXCESS HEAT RADIATED AWAY
ENVELOPE THERMAL CONDUCTION LOSSES
ANODE
$T_a \sim 600$ C
ELECTRON FLOW
CERAMIC SEAL
EXTERNAL LOAD
CATHODE
$T_c \sim 1100$ C
SPACER
THERMAL CONDUCTION LOSSES THROUGH SPACES
HEAT ENERGY INPUT
RADIATION LOSSES FROM CATHODE TO ANODE
ELECTRON COOLING AT THE CATHODE CAUSED BY ELECTRONS BOILING OFF

Thermionic converter.

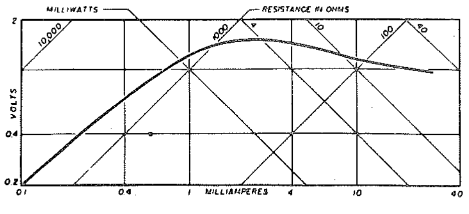

TYPICAL E-I CURVE OF THERMISTOR

Thermistor characteristics.

ammeter. An ammeter that is actuated by the voltage generated in a thermocouple through which the current to be measured is sent. It is used chiefly for measuring radio-frequency currents.

thermocompensator—In pH meters, a temperature-sensitive device sometimes used to make electronic adjustments in the circuit that are required due to changes in the temperature of the solution.

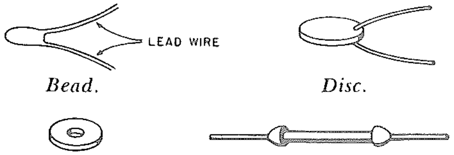

Bead. *Disc.*

Washer. *Rod.*

Thermistors.

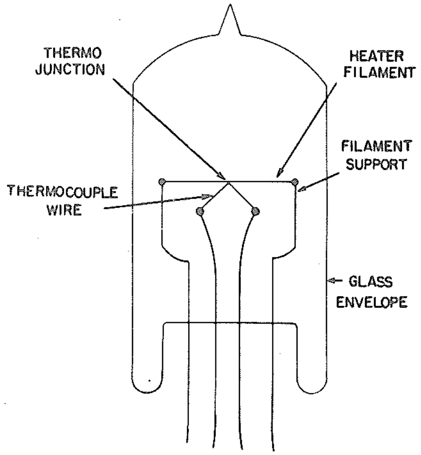

Thermocouple.

thermocouple—Also called thermal junction. A device for measuring temperature where two electrical conductors of dissimilar metals are joined at the point of heat application and a resulting voltage difference, directly proportional to the temperature, is developed across the free ends and is measured potentiometrically. Reference to standard tables is made to determine the temperature.

thermocouple ammeter—*See* Thermoammeter.

thermocouple contact—A contact of special material used in connectors employed in thermocouple applications. Materials often used are iron, constantan, copper, chromel, alumel, and others.

thermocouple converter—*See* Thermal Converter.

thermocouple instrument—An electrothermic instrument in which one or more thermojunctions are heated by an electric current, causing a direct current to flow through the coil of a suitable direct-current mechanism such as one of the permanent-magnet, moving-coil type.

thermocouple thermometer—*See* Thermoelectric Thermometer.

thermocouple vacuum gage — A vacuum gage which depends for its operation on the thermal conduction of the gas present. The pressure being measured is a function of the electromotive force of a thermocouple, the measuring junction of which is in thermal contact with a heater carrying a constant current. Thermocouple vacuum gages ordinarily are used over a pressure range of 10^{-1} to 10^{-3} mm Hg.

thermocouple wire—A wire drawn from special metals or alloys and calibrated to established specifications for use as a thermocouple pair (e.g., iron, constantan, alumel, etc.).

thermodynamics—The study of the relationship between heat and other forms of energy.

thermoelectric arm — Also called thermoelectric leg. The portion of a thermoelectric device having the electric current density and the temperature gradient approximately parallel or antiparrallel and having electrical connections made at its extremities to a part in which the opposite relation between the direction of the temperature gradient and the electric current density exists.

thermoelectric converter—A device capable of converting heat energy directly into electrical energy (bismuth telluride). *Also see* Thermal Converter.

thermoelectric cooler—A device utilizing the Peltier phenomenon to provide a silent, nonmoving cooler having a controllable cooling rate.

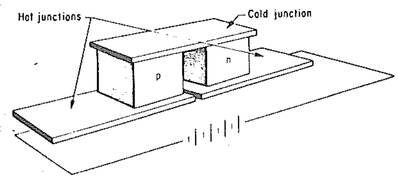

Thermoelectric couple.

thermoelectric couple—A thermoelectric device in which there are two arms of unlike composition.

thermoelectric device — A general term for thermoelectric heat pumps and generators.

thermoelectric effect — The electromotive force produced by the difference in temperature between two junctions of dissimilar metals in the same circuit.

thermoelectric generator — *See* Thermal Converter.

thermoelectric heating device—A thermoelectric heat pump used for adding thermal energy to a body.

thermoelectric heat pump — A device in which the direct interaction of an electrical current and heat flow is used to transfer thermal energy between bodies.

thermoelectricity—1. The direct conversion of heat into electricity. 2. The reciprocal use of electricity to create heat or cold. (*Also see* Seebeck Effect, Peltier Effect, *and* Thomson Effect.)

thermoelectric junction—A thermojunction, as in a thermocouple.

thermoelectric leg—*See* Thermoelectric Arm.

thermoelectric manometer — A manometer (pressure-measuring instrument) that depends on the variation of thermoelectromotive force (voltage due to heat) with pressure.

thermoelectric series—A series of metals arranged in the order of their thermoelectric powers.

thermoelectric thermometer—Also called a thermocouple thermometer. A thermometer employing one or more thermocouples, of which one set of measuring junctions is in thermal contact with the body whose temperature is to be measured, while the temperature of the reference junctions is either known or otherwise taken into account.

thermoelectromotive force—The voltage developed due to the differences in temperature between parts of a circuit containing two or more different metals.

thermoelectron—The electron emitted from a heated body.

thermoelectron engine — *See* Thermionic Converter.

thermoelement—A device consisting of a thermocouple and a heating element arranged for measuring small currents. *Also see* Thermal Converter.

thermogalvanometer — An instrument for measuring small high-frequency currents from their heating effect. Generally it consists of a d-c galvanometer connected to a thermocouple that is heated by a filament carrying the current to be measured.

thermojunction—One of the contact surfaces between the two conductors of a thermocouple. The thermojunction in thermal contact with the body under measurement is called the measuring junction, and the other thermojunction is called the reference junction.

thermojunction battery—A nuclear-type battery which converts heat into electrical energy directly by the thermoelectric or the Seebeck effect.

thermoluminescence—The production of light in a material by moderate heat.

thermomagnetic—1. Pertaining to the effect of temperature on the magnetic properties of a substance. 2. Pertaining to the effect of a magnetic field on the temperature distribution in a conductor.

thermometer—An instrument for measuring temperature. Electrical versions depend on the change in resistance of a material with temperature, the voltage produced in a thermocouple, or various other effects of temperature.

thermophone—An electroacoustic transducer

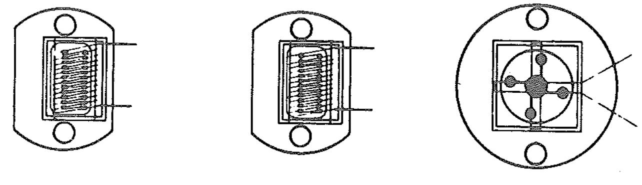

Thermopiles.

in which sound waves of calculable magnitude are produced by the expansion and contraction of the air adjacent to a conductor, the temperature of which varies in response to a current input.

thermopile—A group of thermocouples connected in series-aiding. Specifically: 1. A device used to measure radiant power or energy. 2. A source of electric energy.

thermoplastic flow test—For an insulating material, a measure of the resistance to deformation when subjected to heat and pressure.

thermoplastic material—A plastic material that can be softened by heat and rehardened into a solid state by cooling. This remelting and remolding can be done many times.

thermoplastic recording—A recording process in which information is placed onto plastic tape electronically. A special electron gun, fed by a digital or scanner input, writes a charge in narrow bands on a moving film coated with a plastic that has a low melting point. The film is heated by an r-f heater that melts the plastic coating, permitting it to be deformed by electrostatic and surface-tension forces in proportion to the charge laid down by the beam of the electron gun. The ridges cool quickly and form a diffraction grating that can then be viewed, projected by suitable optics, or read out by a flying-spot scanner. The system operates in a high vacuum.

thermosetting material—Plastic which hardens when heat and pressure are applied. Unlike a thermoplastic, it cannot be remelted or remolded.

thermostat—A mechanism that can be set to operate at definite temperatures and can convert the expansion of heated metal or fluid into sufficient movement and power to operate small devices, control electric circuits or small valves, etc.

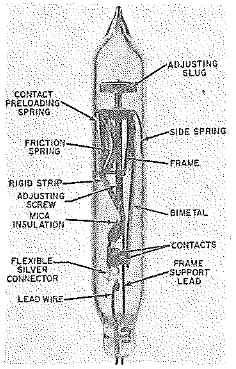

Thermostat.

thermostatic switch—A temperature-operated switch that receives its operating energy by thermal conduction or convection from the device being controlled or operated.

theta polarization—The state of the wave whereby the E vector is tangential to the meridian lines of some given spherical frame of reference.

Thevenin's theorem—The current that will flow through an impedance Z_1 when connected to any two terminals of a linear network between which a voltage E and an impedance Z previously existed, is equal to the voltage E divided by the sum of Z and Z_1.

thick-film circuit—A microcircuit in which passive components of a ceramic-metal composition are formed on a suitable substrate by screening and firing. Discrete active elements are attached separately.

thick-film process—A method in which electronic circuit elements—resistors, conductors, capacitors, etc.—are produced by applying specially formulated pastes to a ceramic substrate in a defined pattern and sequence. The substrate containing the "green pattern" is fired at a relatively high temperature to mature the circuit elements and bond them integrally to the substrate. The term "thick film" distinguishes this technology from thin-film practice, in which circuit elements are made by evaporation or sputtering in a high-vacuum environment. Circuits produced by this technology are identified by several descriptive phrases: thick-film passive circuits or thick-film hybrid circuits when active devices are included.

thick films—Layers of resistive, dielectric, and conductive inks that are deposited on a substrate. The deposition process, similar to graphic silk screening, employs a fine-mesh screen to hold the pattern for the components that are to be deposited. The pattern is produced by photographic means, and wherever the inks are not to be deposited, the holes in the mesh are blocked by an emulsion.

thickness vibration—Vibration of a piezoelectric crystal in the direction of its thickness.

thin film—1. Having to do with the branch of microelectronics in which thin layers of material are deposited on an insulating base in a vacuum. 2. A coating up to a thickness of about 10^4 angstrom.

thin-film capacitor—A capacitor utilizing a metal oxide as the dielectric or insulating material. Both the electrodes and the dielectric are deposited in layers on a substrate. This device is usually associated with microelectronics and integrated and thin-film circuits.

thin-film circuit—A circuit that is formed on a substrate by sputtering or vacuum deposition to make up the passive circuit elements such as resistors and interconnections. Active components are added and their connections made after the thin films have been laid down.

thin-film formation—A process that is either additive by pattern formation through masks, or subtractive by selective etching of predeposited films from a substrate.

thin-film integrated circuit—An integrated circuit consisting of a passive substrate on which the various passive elements (resistors and capacitors) are deposited in the form of thin-patterned films of conductive or nonconductive material. Active components (transistors and diodes) are attached separately as individually packaged devices or in unpacked (chip) form, or may be formed integrally by thin-film techniques. *Also see* Hybrid Thin-Film Circuit.

thin-film memory—In a computer, a storage device made of thin discs of magnetic material deposited on a nonmagnetic base. Its operation is similar to the core memory. (*Also see* Storage *and* Core Memory.)

thin-film microelectronics—Circuits made up of two-dimensional passive and essentially two-dimensional active elements mounted or deposited on thin wafers of an insulating substrate material.

thin-film semiconductor—A semiconductor produced by the deposition of an appropriate single-crystal layer on a suitable insulator.

thin-wall ring magnet—A type of ceramic permanent magnet in which the axial length is greater than the wall thickness or the wall thickness is less than 15% of the outside diameter.

third harmonic—A sine-wave component having three times the fundamental frequency of a complex wave.

Thomson bridge—*See* Kelvin Bridge.

Thomson effect—The net absorption or generation of heat that results, in addition to the joule heat, when an electric current is passed between two points of a homogeneous conductor having a temperature gradient.

Thomson electromotive force—The voltage that exists between two points that are at different temperatures in a conductor.

Thomson heat—The thermal energy absorbed or produced due to the Thomson effect.

thoriated filament—A tungsten vacuum-tube filament to which a small amount of thorium has been added to improve emission. The thorium comes to the surface and is primarily responsible for the electron emission.

thread—*See* Chip.

threaded core—A core with a threaded body.

threading slot—A slot, in the cover plate of a recording-head assembly, into which the tape is slipped in threading the reels of a recorder.

three-address code—Also called instruction code. In computers, a multiple-address code which includes three addresses, usually two addresses from which data are taken and one address where the result is entered. Location of the next instruction is not specified, and instructions are taken from storage in preassigned order.

three-address instruction—In computers, an instruction which includes an operation and specifies the location of three registers. *Also see* Three-Address Code.

three-channel stereo—A stereo recording or reproduction system which uses three spaced microphones for recording and three sound reproducers for playback.

three-db coupler—The junction of two waveguides having a common H wall. The two guides are coupled together by an H-type aperture coupling. Also called the Riblet coupler or short-slot coupler.

three-gun color picture tube—Also called a tri-gun color picture tube. A color-television picture tube with three individually controlled electron guns, one for each of the three primary colors (blue, green, and red).

three-level laser—A laser involving three electronic energy levels. The ground state (level 1) is pumped to level 3. This is followed by a transition to the upper laser level 2, which in turn is followed by stimulated emission back to the ground state.

three-phase circuit—A combination of circuits energized by alternating electromotive forces which differ in phase by one-third of a cycle, or 120 electrical degrees. In practice, the phases may vary several degrees from the specified angle.

three-phase current—A current delivered through three wires—each wire serving as the return for the other two, and the three current components differing in phase successively by one-third of a cycle, or 120 electrical degrees.

three-phase, four-wire system—An a-c supply system comprising four conductors—three connected as in a three-phase, three-wire system and the fourth to the neutral point of the supply, which may be grounded.

three-phase motor—An a-c motor operated from a three-phase circuit.

three-phase, seven-wire system—A system of a-c supply from groups of three single-phase transformers connected in a Y. Thus, a three-phase, four-wire, grounded-neutral system of a higher voltage for power is obtained, the neutral wire being common to both systems.

three-phase, three-wire system—An a-c supply system comprising three conductors, between successive pairs of which are maintained alternating differences of potential successively displaced in phase by one-third of a cycle.

three-plus-one instruction—In digital computer programming, a four-address instruction in which one of the addresses always specifies the location of the next instruction to be performed.

three-pole switch—An arrangement of three single-pole switches coupled together to operate three contacts simultaneously.

three-position relay—Sometimes called a center-stable polar relay.

three-pulse cascade canceler—A moving-target indicator technique in which two "two-pulse cancelers" are cascaded together. This improves the velocity response.

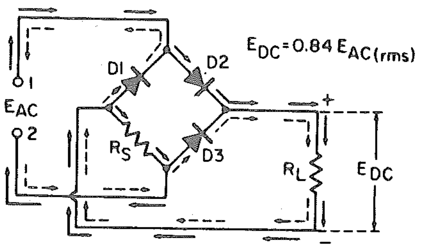

Three-quarter bridge.

three-quarter bridge—A bridge connection in which one of the diode rectifiers has been replaced by a resistor.

three-way speaker system—A sound-reproducing system using three separate speakers, each designed for a specific portion of the audio spectrum (high, low, and middle frequencies). The high- and low-frequency speakers are known as the tweeter and woofer, respectively.

three-way switch—A switch which can connect one conductor to any one of two other conductors.

three-wire system—A system of electric supply comprising three conductors, one of which (known as the neutral wire) is maintained at a potential midway between the potential of the other two (referred to as the outer conductor). Part of the load may be connected directly between the outer conductors, the remainder being divided as evenly as possible into two parts, each of which is connected between the neutral and one outer conductor. There are thus two distinct supply voltages, one being twice the voltage of the other.

threshold—1. The point at which an effect is first produced, observed, or otherwise indicated. 2. In a modulation system, the smallest value of carrier-to-noise ratio at the input of the demodulator for all values above which a small percentage change in the input carrier-to-noise ratio produces a substantially equal or smaller percentage change in the output signal-to-noise ratio. 3. That point at which an indication exceeds the background or ambient.

threshold current—The minimum current at which a gas discharge becomes self-sustaining.

threshold decoding—A decoding procedure so arranged that the decision on the symbol that was transmitted is based on a majority count of the parity-check equations involving that symbol.

threshold field—The least magnetizing force in a direction which tends to decrease the remanence, which, when applied either as a steady field of long duration or as a pulsed field appearing many times, will cause a stated fractional change of remanence.

threshold frequency—The frequency at which the quantum energy is just sufficient to release photoelectrons from a given surface.

threshold of audibility—Also called threshold of detectability or threshold of hearing. For a specified signal, the minimum effective sound pressure of a signal capable of evoking an auditory sensation in a specified fraction of the trials. The characteristics of the signal, the manner in which it is presented to the listener, and the point at which the sound pressure is measured must all be specified. This threshold is usually expressed in decibels relative to 0.0002 μb.

threshold of detectability—*See* Threshold of Audibility.

threshold of discomfort—Also called threshold of feeling. For a specified signal, the minimum effective sound-pressure level which, in a specified fraction of the trials by a battery of listeners, will stimulate the ear to the point where the sensation of feeling becomes uncomfortable. This threshold is customarily expressed in decibels relative to 0.0002 μb.

threshold of feeling—*See* Threshold of Discomfort.

threshold of hearing—*See* Threshold of Audibility.

threshold of luminescence—*See* Luminescence Threshold.

threshold of sensitivity—The smallest stimulus or signal that will result in a detectable output. This phrase is frequently used to describe the voltage point at which an operations monitor or event marker will trigger.

threshold signal—In navigation, the smallest signal capable of producing a recognizable change in the positional information.

threshold value—1. The minimum input that produces a corrective action in an automatic control system. 2. The minimum level for which there is a measurable output.

throat—1. Part of the flare or tapered parallel-plate guide immediately adjacent to and connected to the main run of a waveguide. 2. The smaller cross-sectional area of a horn.

throat microphone—A microphone worn around the throat and actuated by vibrations of the larynx as the user talks. It is used in jet airplanes, tanks, and other places where background noise would drown out the conversation.

through path—The transmission path from the loop input signal to the loop output signal in a feedback control loop.

through repeater—A microwave repeater that is not equipped to provide for connections to any local facilities other than the service channel.

through transfer function—The transfer function of the through path in a feedback control loop.

throw—1. In an electric motor or generator, the number of core slots spanned between the bottom leg of a coil and the top leg of the same coil. 2. The span between the bottom leg of a coil and the top leg of the same coil, expressed in the number of core slots spanned.

throwing power—Referring to the ability of an anode used in an impressed-current cathodic-protection system to distribute its current over a large surface. It is principally dependent upon the anode voltage, current, surface area, position of the anode with respect to the cathode, and the salinity and flow velocity of the water.

throw-out spiral—*See* Lead-Out Groove.

thru-hole connection—Also called feed-thru connection and plated thru hole. A conductive material used to make electrical and mechanical connection between the conductive patterns on opposite sides of a printed-circuit board.

thru repeater—In a microwave system, a repeater station that is not equipped to be connected to any local facilities other than the service channel.

thump—1. A low-frequency transient disturbance in a system or component. 2. The noise caused in a receiver by telegraph currents when the receiver is connected to a telephone circuit on which a direct-current telegraph channel is superimposed.

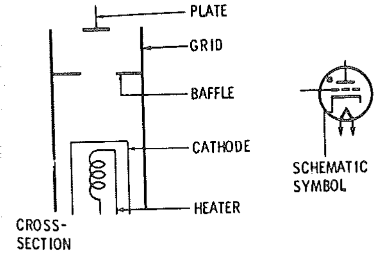

Thyratron.

thyratron—A hot-cathode gas tube in which one or more control electrodes initiate the anode current, but do not limit it except under certain operating conditions.

thyratron gate—In computers, an AND gate consisting of a multielement gas-filled tube in which conduction is initiated by the coincident application of two or more signals. Conduction may continue after one or more of the initiating signals have been removed.

thyratron inverter—An inverter circuit in which thyratron tubes convert the d-c power to a-c power.

thyristor—A bistable device comprising three or more junctions. At least one of the junctions can switch between reverse- and forward-voltage polarity within a single quadrant of the anode-to-cathode voltage–current characteristics. Used in a generic sense to include silicon controlled rectifiers and gate-control switches, as well as multilayer two-terminal devices. (Many countries use the term "thyristor" in place of the term "PNPN-type switch.")

thyrite—A silicon-carbide ceramic material with nonlinear resistance characteristics. Above a critical voltage, the resistance falls considerably.

THz—Abbreviation for terahertz (10^{12} hertz).

tickler—A small coil connected in series with the anode circuit of an electron tube and inductively coupled to a grid-circuit coil. The tickler coil is used chiefly in regenerative detector circuits, to establish feedback or regeneration.

tie cable—1. A cable between two distributing frames or distributing points. 2. A cable between two private branch exchanges. 3. A cable between a private branch exchange switchboard and the main office. 4. A cable connecting two other cables.

tie line—1. *See* Interconnection. 2. A leased communications circuit or channel.

tie point—An insulated distributing point (other than an active terminal connection) where junctions of component leads are made in circuit wiring.

tier array—An array of antenna elements, one above the other.

tie trunk—A telephone line or channel directly connecting two private branch exchanges.

tie wires—Short pieces of wire used to tie open-line wires to insulators.

tight coupling—*See* Close Coupling.

tilt—In radar, the angle between the axis of radiation in the vertical plane, and a reference axis which is normally the horizontal. *Also see* Wave Tilt.

tilt angle—In radar, the angle between the vertical axis of radiation and a reference axis (normally the horizontal).

tilt controls—In a color television receiver employing the magnetic-convergence principle, the three controls used to tilt the vertical center rows in the three colored patterns produced by a dot-generator signal.

tilt error—1. In navigation, the ionospheric error component due to nonuniform height. 2. The difference between the true tilt and the mechanical tilt. The sign is such that when it is algebraically added to the mechanical tilt, the result is the true tilt. *Also see* Antenna Tilt Error.

tilting—1. Forward inclination of the wave front of radio waves traveling along the ground. The amount of tilt depends on the electrical constants of the ground. 2. Changing the angle of a television camera to follow a moving object being televised. 3. Changing the vertical angle of a directional antenna.

timbre—Also called tone color or musical quality. The character of a musical tone that distinguishes one musical instrument from another playing the same note. The difference between two steady tones having the same pitch and degree of volume is called the difference in timbre. Timbre depends mostly on the relative intensity of the different harmonics and the frequencies of the most prominent harmonics.

time—The measure of the duration of an event. The fundamental unit of time is the second.

time assignment speech interpolation—Telephone switching equipment whereby a person is connected to idle circuits when he starts talking, and is disconnected when he stops talking.

time base—A voltage generated by the sweep circuit of a cathode-ray-tube indicator. Its waveshape is such that the trace is either linear with respect to time or, if nonlinear, is still at a known timing.

time clock—A loosely used term sometimes referring to time switches, sometimes to interval timers, or to any type of timer.

time-code generator—A timer with an absolute time reference having digital outputs suitable for machine interpretation. Dial or numeric display may be included.

time constant—The time required for an exponential quantity to change by an amount equal to 0.632 times the total change that will occur. Specifically: 1. In a capacitor-resistor circuit, the number of seconds required for the capacitor to reach 63.2% of its full charge after a voltage is applied. 2. In an inductor-resistor circuit, the number of seconds required for the current to reach 63.2% of its final value. 3. The time constant, in seconds, of an inductor having an inductance L in henrys and resistance R in ohms is equal to L/R. 4. The time constant of a capacitor having a capacitance C in farads in series with a resistance R in ohms is equal to R × C. 5. The time required for the motor to accelerate from 0 to 63.2% of its final no-load speed when the rated voltage, with the proper phase relationship, is applied.

time constant of fall—The time required for a pulse to fall from 70.7% to 26.0% of its maximum amplitude, excluding spikes.

time constant of rise—The time required for a pulse to rise from 26.0% to 70.7% of its maximum amplitude, excluding spikes.

time-current characteristics of a fuse—The relation between the root-mean-square alternating current or direct current and the time for the fuse to perform the whole or some specified part of its interrupting function. Usually shown as a curve.

timed acceleration—A control function that automatically controls the speed increase of a drive as a function of time.

timed deceleration—A control function that automatically controls the speed decrease of a drive as a function of time.

time delay—1. The time required for a signal to travel between two points in a circuit. 2. The time required for a wave to travel between two points in space. 3. The total elapsed time or lag required for a given command to be effected after the command is given.

time-delay circuit—A circuit that delays the transmission of an impulse signal, or the performance of a transducer, for a definite desired length of time.

time-delay closing relay — *See* Time-Delay Starting Relay.

time-delay generator—A device which accepts an input signal and provides a delay in time before the initiation of an output signal.

time-delay relay—Also called slow-action relay. A relay in which there is an appreciable interval of time between the energizing or de-energizing of the coil and the movement of the armature. (e.g., slow-operating relays and slow-release relays).

time-delay starting (or closing) relay—Also called time-delay closing relay. A device which functions to give a desired amount of time delay before or after any point or operation in a switching sequence or protective relay system.

time-delay stopping or opening relay—A time-delay device which serves in conjunction with the device that initiates the shutdown, stopping, or opening operation in an automatic sequence.

time-derived channel—Any of the channels obtained by time-division multiplexing of a channel.

time discriminator—A circuit in which the sense and magnitude of the output is a function of the time difference of, and relative time sequence between, two pulses.

time-distribution analyzer—Also called time sorter. An instrument that indicates the number or rate of occurrence of time intervals falling within one or more specified ranges. The time interval is defined by the separation between members of a pulse pair.

time-division data link—Radiocommunications which use time-division techniques for channel separation.

time-division multiplex—A system for transmitting information about two or more quantities (measurands) over a common channel, by dividing available time intervals among the measurands to form a composite train. Information may be transmitted by a

variation of pulse duration, pulse amplitude, or pulse position, or by a pulse code.

time flutter—A variation in the synchronization of components of a radar system, leading to variations in the position of the observed pulse along the time base, and reducing the accuracy with which the time of arrival of a pulse may be determined.

time frame—In telemetry, the time period containing all elements between corresponding points of two successive reference markers.

time gain—*See* Sensitivity-Time Control.

time gate—A transducer which has an output during chosen time intervals only.

time harmonics — High-frequency components generated by nonlinearity in a resolver magnetic circuit. In a-c resolvers with no d-c components present, time harmonics are odd multiples of the fundamental frequency. Time harmonics appear as a component of null voltage.

time-interval selector—A circuit that functions to produce a specified output pulse when and only when the time interval between two input pulses is between set limits.

time lag—1. The interval between application of any force and full attainment of the resultant effect. 2. The interval between two phenomena.

time-mark generator — A circuit that produces accurately spaced pulses for display on the screen of an oscilloscope.

time pattern—A picture-tube presentation of horizontal and vertical lines or rows of dots generated by two stable frequency sources operating at multiples of the line and field frequencies.

time phase—Reaching corresponding peak values at the same instants of time, though not necessarily at the same points in space.

time pulse distributor—A device or circuit for allocating timing pulses or clock pulses to one or more conducting paths or control lines in a specified sequence.

time quadrature—Differing by a time interval corresponding to one-fourth the time of one cycle of the frequency in question.

timer—1. A part of an electronic circuit which starts pulse transmission and synchronizes it with the beginning of an indicator sweep or the timing of gates, range markers, etc. 2. A special clock mechanism or motor-operated device used to perform switching operations at predetermined time intervals. 3. An assembly of electric circuits and associated equipment which provides the following: trigger pulses, sweep circuits, intensifier pulses, gate voltages, blanking voltages, and power supplies. 4. The part of the radar set that initiates pulse transmission and synchronizes this with the beginning of indicator sweeps, timing of gates, range markers, etc. *Also see* Synchronizer.

time response — An output, expressed as a function of time, that results from a specified input applied under specified operating conditions.

timer motor—A synchronous clock motor having a definite output speed determined by the number of poles in the stator and the reduction of the associated gear train.

time-sequencing—In a computer, switching signals generated by a program purely as a function of accurately measured elapsed time.

time sharing—1. A computer-use technique whereby several operators may use a computer at the same time. *Also see* Interlace Operation *and* Interrupt. 2. The use of a device for two or more purposes during the same overall time interval, accomplished by interspersing component actions in time.

time signals—Time-controlled radio signals broadcast by government-operated radio station WWV at regular intervals each day on several frequencies.

time sorter — *See* Time Distribution Analyzer.

time switch—A clock-controlled switch used to open or close a circuit at one or more predetermined times. Usually refers to repeat cycle timers with a dial graduated with time of day (in 24-hour type) or with time and day in 7-day type. Has means of adjusting the time when circuits are turned on or off, usually with pins or clips. Used for turning lights on at night, ringing bells in offices, schools, etc.

time-to-digital conversion—The process of converting an interval of time into a digital number.

timing axis oscillator—*See* Sweep Generator.

timing-pulse distributor—Also called waveform generator. A computer circuit driven by pulses from the master clock. It operates in conjunction with the operation decoder to generate timed pulses needed by other machine circuits to perform the various operations.

timing relay—A form of auxiliary relay used to introduce a definite time delay in the performance of a function.

timing signal—Any signal recorded simultaneously with data to provide a time index.

tinkertoy—An attempt at modularization where wafers, with one or more component parts printed or mounted on them, are stacked vertically, with interconnecting wiring stiff enough to provide support running through holes around the periphery of the wafer.

tinned—Covered with metallic tin to permit easy soldering.

tinned wire—A copper wire that has been coated with a layer of tin or solder to prevent corrosion and to simplify soldering.

tinsel conductor—A type of electrical conductor comprising a number of tiny threads, each thread having a fine, flat ribbon of copper or cadmium bronze closely spiraled about it. Used for small cables requiring limpness and extra-long flex life (e.g., used with headsets or handsets).

tip—1. The contacting part at the end of a plug or probe. 2. Also called tip side. The end of the plug used to make circuit connections in a manual switchboard. The tip is the connector attached to the positive side of the common battery which powers the station equipment. By extension, it is the positive battery side of a communications line.

tip jack—Also called a pup jack. A small single-hole jack for a single-pin contact plug.

tipoff—The last portion of a vacuum-tube bulb to be melted and sealed after evacuation of the bulb.

tip side—Also called a tip wire. The conductor of a circuit associated with the tip of a plug or the tip spring of a jack. *Also see* Tip.

tip wire—*See* Tip Side.

T-junction—*See* Tee Junction.

TM$_{m,n}$ wave—In a rectangular waveguide, the transverse magnetic wave for which m and n are the number of half-period variations of the magnetic field along the longer and shorter transverse dimensions, respectively.

TM wave—Abbreviation for transverse magnetic wave.

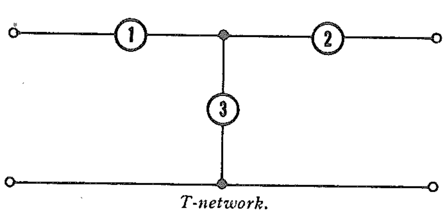

T-network.

T-network—A network composed of three branches. One end of each branch is connected to a common junction point. The three remaining ends are connected to an input terminal, an output terminal, and a common input and output terminal, respectively.

to-from indicator—An instrument that forms part of the omnirange facilities and is used for resolving the 180° ambiguity.

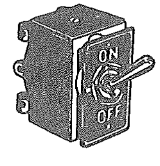

Toggle switch.

toggle switch—A two-position snap switch operated by a projecting lever to open or close circuits.

tolerance—A permissible deviation from a specified value. A frequency tolerance is expressed in hertz or as a percentage of the nominal frequency; an orientation tolerance, in minutes of arc; a temperature tolerance, in degrees Celsius, and a dimensional tolerance, in decimals or fractions.

toll center—Also called toll office and toll point. The basic toll switching entity; a central office where channels and toll-message circuits terminate. While this is usually one particular central office in a city, larger cities may have several central offices where toll-message circuits terminate.

toll-terminal loss—On a toll connection, that part of the over-all transmission loss attributable to the facilities from the toll center, through the tributary office to and including the subscriber's equipment.

tone—1. A sound wave capable of exciting an auditory sensation having pitch. 2. A sound sensation having pitch.

tone arm—*See* Pickup Arm.

tone burst—A single sine-wave frequency, 50 to 500 microseconds long having a rectangular envelope, used for testing the transient response of speakers.

tone channel—An intelligence or signalling circuit in which on-off or frequency-shift modulation of a frequency (usually an audio frequency) is used as a means of transmission.

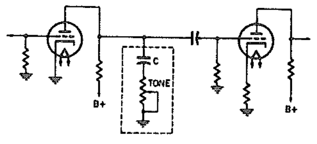

Tone control.

tone control—A control, usually part of a resistance-capacitance network, used to alter the frequency response of an amplifier so that the listener can obtain the most pleasing sound. In effect, a tone control accentuates or attenuates the bass or treble portion of the audio-frequency spectrum.

tone generator—A device for providing an audio-frequency current suitable for testing audio-frequency equipment or for signaling.

tone keyer—An instrument device which converts direct-current impulses to audio tones, for line transmission or for keying a transmitter.

tone localizer—*See* Equisignal Localizer.

tone-modulated waves—Waves obtained from continuous waves by amplitude-modulating them at an audio frequency in a substantially periodic manner.

tone modulation—A type of code-signal transmission obtained by causing the radio-frequency carrier amplitude to vary at a fixed audio frequency.

tone reversal—Distortion of the recorder copy in facsimile. It causes the various shades of black and white not to be in the proper order.

Tonotron—A multimode, selective-erasure storage tube.

top cap—A terminal in the form of a metal cap at the top of some vacuum tubes and connected to one of the electrodes.

top-loaded vertical antenna—A vertical antenna that is larger at the top, resulting in a modified current distribution, which gives a more desirable radiation pattern vertically. A series reactor may be connected between the enlarged portion of the antenna and the remaining structure.

topology—The surface layout of the elements comprising an IC.

tornadotron — A millimeter wave device which generates radio-frequency power from an enclosed, orbiting electron cloud excited by a radio-frequency field when subjected to a strong, pulsed magnetic field.

torn-tape relay—A method of receiving messages in tape form, breaking the tape, and retransmitting the message in tape form.

toroid—1. A surface, or its closed solid, generated by any closed plane rotating about a straight line in its own plane—the resulting configuration being doughnut shaped. 2. A highly efficient type of coil wound upon a ring or "doughnut" type of core. The toroid provides for a highly concentrated magnetic field within itself, and has a minimum magnetic flux leakage (external field).

toroidal coil—A coil wound in the form of a toroidal helix.

toroidal core—A ring-shaped core.

toroidal permeability—Under stated conditions, the relative permeability of a toroidal body of the given material. The permeability is determined from measurements of a coil wound on the toroid such that stray fields are minimized or can be neglected.

torque—1. A force that tends to produce rotation or twisting. 2. Force applied rotationally as around the axis of a nut, bolt, or screw so that these elements can assume their share of stress. This tightening operation is measured in inch-pounds, foot-pounds, or inch-ounces.

torque amplifier—A device with input and output shafts and supplying work to rotate the output shaft so that its position corresponds to that of the input shaft but does not impose any significant torque on the latter.

torque-coil magnetometer—A magnetometer which depends for its operation on the torque developed by a known current in a coil capable of turning in the field to be measured.

torque gradient—The torque required in inch-ounces to pull a specific energized synchro 1° away from its normal position.

torque of an instrument—Also called deflecting torque. The turning moment produced on the moving element by the quantity to be measured or by some quantity dependent thereon acting through the mechanism.

torquer—A device which produces torque about an axis of freedom in response to a signal input.

torque-to-inertia ratio—*See* Acceleration At Stall.

torsiometer—An instrument for measuring the amount of power which a rotating shaft is transmitting.

torsion galvanometer—A galvanometer in which the force between the fixed and moving systems is measured by the angle through which the supporting head of the moving system must be rotated to return the moving system to zero.

torsion-string galvanometer—A sensitive galvanometer in which the moving system is suspended by two parallel fibers that tend to twist around each other.

total capacitance—The capacitance between a given conductor and all other conductors in a system when all other conductors are connected together. The total capacitance of a conductor equals the ratio of the sum of its direct capacitances to the other conductors.

total combined regulation—The change in output of a regulated power supply arising from simultaneous changes in all of the specified operating conditions, when the direction of such changes is such as to make their effects additive. It may be stated as a percentage of the specified output and/or an absolute value.

total distortion—The sum total of all forms of signal distortions.

total emission—The magnitude of the current produced when electrons are emitted from a cathode under the influence of a voltage such that all the electrons emitted are drawn away from the cathode.

total emissivity—The ratio of radiation emitted by a surface to the radiation emitted by the surface of a black body under identical conditions. Important conditions which affect emissivity of a material are surface finish, color, temperature, and wavelength of radiation. Emissivity may be expressed for radiation of a single wavelength (monochromatic emissivity), for total radiation of a specified range of wavelengths (total spectral emissivity), or for total radiation of all wavelengths (total emissivity).

total excursion—The application of a stimulus, in a controlled manner, over the span of an instrument.

total harmonic distortion—The ratio of the power at the fundamental frequency measured at the output of the transmission system considered, to the power of all harmonics observed at the output of the system because of its nonlinearity when a single frequency signal of specified power is applied to the input of the system. It is expressed in decibels.

total internal reflection—When light passes from one medium to another which is optically less dense (e.g., from glass to air), the ray is bent away from the normal. If the incident ray meets the surface at such an angle that the refracted ray must be bent away at an angle of more than 90°, the light cannot emerge at all, and is totally internally reflected.

totalizing—To register a precise total count from mechanical, photoelectric, electromagnetic, or electronic inputs or detectors.

total losses of a ferromagnetic part—Under stated conditions, the power absorbed and then dissipated as heat when a body of ferromagnetic material is placed in a time-varying magnetic field.

total losses of a transformer—The losses represented by the sum of the no-load and load losses.

total luminous flux—The total light emitted in all directions by a light source.

totally enclosed motor—A motor so enclosed as to prevent the free exchange of air between the inside and the outside of the case, but not sufficiently enclosed to be termed airtight.

totally unbalanced currents—*See* Push-Push Currents.

total range of an instrument—Also called the range of an instrument. The region between the limits within which the quantity measured is to be indicated or recorded.

total regulation—The arithmetic sum of changes in output of a regulated power supply arising from changes in each of the specified operating conditions (current, voltage, or power) when such changes are applied individually, and in a manner to make their effects additive. It may be stated as a percentage of the specified output and/or absolute value.

total resistance—The d-c resistance of a (precision) potentiometer between the input terminals with the shaft positioned so as to give a maximum resistance value.

total spectral emissivity—*See* Total Emissivity.

total telegraph distortion—Telegraph transmission impairment expressed in terms of time displacement of mark-space and space-mark transitions from their proper positions and given in per cent of the shortest perfect pulses called the unit pulse.

total transition time—In a circuit, the time interval between the point of 10% input change and the point of 90% output change. It is equal to the sum of the delay time and rise (or fall) time.

touch control—A control circuit that actuates a circuit when two metal areas or a preselected area are bridged by one's finger or hand.

touch-tone dialing—Push-button telephone dialing in which the direct-current pulsing technique of the rotating dial is replaced by a combination of tones to provide the proper automatic switching.

tourmaline—A strongly piezoelectric natural or synthetic crystal.

tower—A structure usually used when an antenna must be mounted higher than 50 feet.

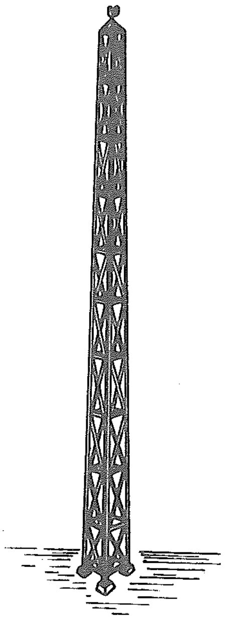

Tower radiator.

tower radiator—A metal tower structure used as a transmitting antenna.

Townsend criterion—The relationship expressing the minimum requirement for breakdown in terms of the ionization coefficients.

Townsend discharge—An electrical discharge in a gas at moderate pressure (above about 0.1 millimeter of mercury). It corresponds to corona, and is free from space charges.

Townsend Ionization Coefficient—The average number of ionizing collisions made by an electron as it drifts a unit distance in the direction of an applied electric force.

T-pad—A pad made up of resistance elements arranged in a T-network (two resistors inserted in one line, with a third between their junction and the other line).

trace—1. The pattern on the screen of a cathode-ray tube. 2. The visible line or lines produced on the screen of a cathode-ray tube by the deflection of the electron stream.

trace interval—The interval corresponding to the direction of sweep used for delineation.

tracer—1. A radioisotope which is mixed with a stable material to trace the material as it undergoes chemical and physical changes. 2. A thread of contrasting color woven into the insulation of a wire for identification purposes. 3. *See* Signal Tracer.

trace width—The distance between two points on opposite sides of a trace on an oscilloscope at which luminance is 50% of

maximum. If the trace departs from a well-behaved (approximately Gaussian) form, it should be smoothed for the purpose of measurement.

tracing distortion—The nonlinear distortion introduced in the reproduction of mechanical recording when the curve traced by the reproducing stylus is not an exact replica of the modulated groove. For example, in the case of sine-wave modulation in vertical recording, the curve traced will be a poid.

track—1. A path which contains reproducible information left on a medium by recording means energized from a single channel. 2. In electronic computers, that portion of a moving-type storage medium accessible to a given reading station (e.g., film, drum, tapes, discs). (*Also see* Band *and* Channel.) 3. An area of tape surface that coincides with the location of the recorded magnetization produced by one record gap.

track-command guidance—A missile-guidance method in which both target and missile are tracked by separate radars and commands are sent to the missile to correct its course.

track hold memory—A circuit that, in its "track" mode, develops an output that follows (ideally) the input exactly, or is proportional to it; and then, in its "hold" mode, maintains the output constant (ideally) at the value it had at the instant the circuit was commanded to change from "track" to "hold."

track homing—The process of following a line of position known to pass through an object.

tracking—1. The process of keeping a radio beam, or the cross hairs of an optical system, set on a target—usually while determining the range of the target. 2. The maintenance of proper frequency relationships in circuits designed to be simultaneously varied by ganged operation. 3. The accuracy with which the stylus of a phonograph pickup follows a prescribed path over the surface of the record. 4. The ability of a multigun tube to superimpose simultaneously information from each gun. Tracking error is the maximum allowable distance between the displays of any two guns. 5. The ability of an instrument to indicate at the division line being checked when energized by corresponding proportional values of actual end-scale excitation. 6. The interconnecting of power supplies in such a manner that one unit serves to control all units in series operation. In this manner, the output voltage of the slave units follows the variations of the master. 7. The difference at any shaft position between the output ratios of any two commonly actuated similar electrical elements expressed as a percentage of the single total voltage applied to them. 8. Motion given to the major lobe of an antenna so that a preassigned moving target in space is always contained within the major lobe. 9. The process of following the movement of a satellite or rocket by telescope, radar, radio, or photographic observations.

tracking accuracy—The ability to indicate at a division line being checked when a meter is energized by corresponding proportional values of actual end-scale excitation, expressed as a percentage of actual end-scale value. Tracking accuracy is usually within one and one-half times the rated full-scale accuracy of the meter.

tracking antenna—A directional antenna system which automatically changes in position or characteristics to follow the motion of a moving signal source.

tracking error—In lateral mechanical recording, the angle between the vibration axis of the mechanical system of the pickup and a plane which contains the tangent to the unmodulated record groove and is perpendicular to the recording surface at the point of needle contact.

tracking force—*See* Stylus Force.

tracking jitter—*See* Jitter.

tracking resistance—*See* Arc Resistance.

tracking spacing—The distance between the center line of adjacent tracks on a recording tape. In the typical case of longitudinal tracks 50 mils wide, the tracking spacing is 70 mils.

tracking spot—A moving spot used for target indication on a radar.

track-while-scan—A radar system utilizing electronic-computer techniques whereby raw data are used to track an assigned target, compute target velocity, and predict its future position without interfering with the scanning rate.

track width—The width of the track on a recording tape, corresponding to a given record gap. The most common track widths encountered in longitudinal recording are 48 and 50 mils, several such tracks being accommodated on a half-inch-wide tape.

tractive force—The force which a permanent magnet exerts on a ferromagnetic object.

traffic—Messages handled by communication or amateur stations.

trailers—Bright streaks at the right of large dark areas, or dark areas or streaks at the right of bright areas, in the televised picture. The usual cause is insufficient gain at low video frequencies.

trailing blacks—*See* Following Blacks.

trailing edge—The major portion of the decay of a pulse.

trailing-edge pulse time—The time at which the instantaneous amplitude last reaches a stated fraction of the peak pulse amplitude.

trailing reversal—*See* Following Whites *and* Following Blacks.

trailing whites—*See* Following Whites.

trainer—1. The representation of an operating system by computers and its associated equipment and personnel. 2. Electronic equipment, used for training operators of radar or sonar apparatus by simulating signals received under operating conditions in the field.

tramlines—A pattern on an *A* scope appearing as a number of horizontal lines above the baseline. The effect is produced by cw modulated with a low-frequency signal.

trans—Abbreviation for transmitter. Also abbreviated xmtr or xmitter.

transaction file — In a computer, a file containing information relating to current activities or transactions.

transadmittance—From one electrode to another, the alternating components of the current of the second electrode divided by the alternating component of the voltage of the first electrode, all other electrode voltages being maintained constant. As most precisely used, the term refers to infinitesimal amplitudes.

transadmittance compression ratio—Ratio of the magnitude of the small-signal forward transadmittance of the tube to the magnitude of the forward transadmittance at a given input-signal level.

transceiver — 1. The combination of radio transmitting and receiving equipment in a common housing, usually for portable or mobile use and employing some common circuit components for both transmitting and receiving. 2. A device which transmits and receives data from punched card to punched card. It is essentially a conversion device which, at the sending end, reads the card and transmits the data over the wire. At the receiving end, it punches the data into a card.

transceiver data link—Integrated data processing by means of punched cards, using transceivers as terminal equipment. The transmission path can be wire or radio.

transconductance — Symbolized g_m. Also called mutual conductance. An electron-tube rating equal to the change in anode current divided by the change in grid voltage causing the anode-current change. The unit of transconductance is the mho; a more commonly used unit is the micromho (10^{-6}mho).

transconductance amplifier—An amplifier that supplies an output current proportional to its input voltage. From its output terminals the amplifier appears to be a current source with a high output impedance R_0. Its input impedance $R_{in} > > R_g$ and the output impedance $R_o > > R_L$, where R_g and R_L are the source and load impedances.

transconductance meter—Also called a mutual-conductance meter. An instrument for indicating the transconductance of a grid-controlled electron tube.

transconductance tube tester—Also called a mutual-conductance or dynamic-mutual-conductance tube tester. A tube tester with circuits set up in such a manner that the test is made by applying an a-c signal of known voltage to the control grid and the tube amplification factor or mutual conductance is measured under dynamic operating conditions.

transconductor—An active or passive network, the short-circuit output current of which is a specific, accurately known, often nonlinear function of the input voltage.

transcribe—1. To copy, with or without translating, from one external storage medium of a computer to another. 2. To copy from a computer into a storage medium or vice versa.

transcriber—Equipment associated with a computing machine for the purpose of transferring input or output data from a record of information in a given language to the medium and language used by a digital computing machine, or from a computing machine to a record of information.

transcription—An electrical recording (e.g., a high-fidelity, 33⅓-rpm record) containing part or all of a (radio) program. It may be either an instantaneous recording disc or a pressing.

transdiode—A transistor so connected that the base and collector are actively maintained at equal potentials, although not connected together. The logarithmic transfer relationship between collector current and base-emitter voltage very closely approximates that of an ideal diode.

transducer—A device by means of which energy can be made to flow from one or more transmission systems to one or more other transmission systems. The energy may be in any form, such as electrical, mechanical, acoustical, etc. The term "transducer" is often restricted to a device in which the magnitude of an applied stimulus is converted into an electrical signal proportionate to the quantity of the stimulus. Usually, the variations of the phenomenon being measured are referenced to time.

transducer-coupling system efficiency—The power output at the point of application, divided by the electrical power input into the transducer.

transducer efficiency—Ratio of the power output to the electrical power input at the rated power.

transducer equivalent noise pressure (of an electroacoustic transducer or system used for sound reception) — Also called equivalent noise pressure. For a sinusoidal plane-progressive wave, the rms sound pressure which, if propagated parallel to the principal axis of the transducer, would produce an open-circuit signal voltage equal to the rms of the inherent open-circuit noise voltage of the transducer in a transmission band having a bandwidth of 1 hertz and centered on the frequency of the plane sound wave.

transducer gain—Ratio of the power that the transducer delivers to the load under specified operating conditions, to the available power of the source.

transducer insertion loss — *See* Insertion Loss.

transducer loss — Ratio of the available power of the source, to the power that the transducer delivers to the load under specified operating conditions.

transducer pulse delay — The interval of time between a specified point on the input pulse and on its related output pulse.

transfer—1. To transmit, or copy, information from one device to another. 2. To jump. 3. The act of transferring. 4. In electrostatography, the act of moving a developed image, or a portion thereof, from one surface to another without altering its geometrical configuration (e.g., by electrostatic forces or by contact with an adhesive-coated surface).

transfer admittance—The complex ratio of the current at the second pair of terminals of an electrical transducer, to the electromotive force applied between the first pair, all pairs of terminals being terminated in any specified manner.

transfer characteristic—1. The relationship, usually shown by a graph, between the voltage of one electrode and the current to another electrode, all other electrode voltages being maintained constant. 2. The characteristics pertaining to the degree of output activity for an input stimulus, or vice versa.

transfer check—In a computer, verification of transmitted data by temporarily storing, retransmitting, and comparing. Also a check to see if the transfer or jump instruction was properly performed.

transfer circuit—A circuit which connects communication centers of two or more separate networks in order to transfer the traffic between the networks.

transfer constant—Also called the image-transfer constant. One-half the natural logarithm of the complex ratio of the voltage times current entering a transducer, to that leaving the transducer when it is terminated in its image impedances.

transfer contact—*See* Break-Make Contact.

transfer control—*See* Jump.

transfer current—The starter-gap current required to cause conduction across the main gap of a glow-discharge, cold-cathode tube.

transfer function—1. The relationship between two system variables that enables the second variable to be determined from the first. 2. A mathematical expression frequently used by control engineers which expresses the relationship between the outgoing and the incoming signals of a process or control element. The transfer function is useful in studies of control problems. 3. A mathematical expression which describes the relationship between physical conditions at two different points in time or space in a given time, and perhaps, also describes the role played by the intervening time or space.

transfer impedance—Ratio of the potential difference between any two pairs of terminals of a network applied at one pair of terminals, to the resultant current at the other pair of terminals (all terminals being terminated in any specified manner).

transfer instruction—An instruction, signal, or operation which conditionally or unconditionally specifies the location of the next instruction and directs a computer to that instruction.

transfer operation — An operation which moves data from one storage location or one storage medium to another. Transfer is sometimes taken to refer specifically to movement between different media; storage to movement within the same medium.

transfer ratio—From one point to another in a transducer at a specified frequency, the complex ratio of the generalized force or velocity at the second point to that applied at the first point.

transferred charge—The net electric charge moved from one terminal of a capacitor to another through an external circuit.

transferred charge characteristic — The mathematical function that relates transferred charge to capacitor voltage.

transfer switch—A form of air switch arranged so that a conductor connection can be transferred from one circuit to another without interrupting the current.

transfer time—1. The time required for a transfer to be made in a digital computer. 2. In a relay, the total time—after contact bounce has ceased—between the breaking of one set of contacts and the making of another.

Transfluxor — A trade name (RCA) for a binary magnetic core having two or more openings. Control of the transfer of magnetic flux between the three or more legs of the magnetic circuits provides ways to store information, gate electrical signals, etc.

transform—1. To convert a current or voltage from one magnitude to another, or from one type to another. 2. In digital-computer programming, to change information in structure or composition without significantly altering the meaning or value at the same time.

transformation point—The temperature at which an alloy or metal changes from one crystal state to another as it heats or cools.

transformation ratio — The ratio between electrical output and input under certain specified conditions.

transformer—An electrical device which, by electromagnetic induction, transforms electric energy from one or more circuits to one or more other circuits at the same frequency, but usually at a different voltage and current value.

transformer build—The amount of window area used in constructing a transformer.

transformer-coupled amplifier—An amplifier, the stages of which are coupled together by transformers.

transformer coupling—Use of a transformer, between stages of an amplifier, i.e., to connect the anode circuit of one stage to the grid circuit of the following stage.

transformerless receiver — A receiver in which the power-line voltage is applied directly to series-connected tube heaters or filaments and to a rectifier circuit instead of first being stepped up or down in voltage by a power transformer.

transformer load loss—Losses in a transformer which are incident to the carrying of the load. Load losses include I^2R loss in the windings due to load current, stray loss due to stray fluxes in the winding core, clamps, etc., and to circulating current, if any, in parallel windings.

transformer loss—Expressed in decibels, the ratio of the signal power that an ideal transformer of the same impedance ratio would deliver to the load impedance, to the signal power delivered by the actual transformer.

transformer oil—A high-quality insulating oil in which windings of large power transformers are sometimes immersed to provide high dielectric strength, insulation resistance, and flash point, plus freedom from moisture and oxidation.

transformer vault — An isolated enclosure, either above or below ground, with fire-resistant walls, ceiling, and floor for unattended transformers and their auxiliaries.

transformer voltage ratio—The ratio of the root-mean-square primary terminal voltage to the root-mean-square secondary terminal voltage under specified conditions of load.

transforming section—A length of waveguide or transmission line of modified cross section, or with a metallic or dielectric insert, used for impedance transformation.

transhybrid loss—The transmission loss at a given frequency measured across a hybrid circuit when connected to a given two-wire termination and balancing network.

transient—1. The instantaneous surge of voltage or current produced by a change from one steady-state condition to another. 2. A phenomenon caused in a system by a sudden change in conditions, and which persists for a relatively short time after the change. 3. A distinct line or series of lines perpendicular to the direction of scanning produced in the recorded copy immediately following a sudden change in density. 4. A damped oscillatory quantity occurring in the output of a system as a result of a sudden change in input. 5. In a power supply, a variable observed during transition from one steady-state operating condition to another (or between two successive states in which the response is constant or identically repetitive). 6. A momentary surge on a signal or power line. It may produce false signals or triggering impulses and cause insulation or component breakdowns and failures.

transient analyzer—An electronic device for repeatedly producing a succession of equal electric surges of small amplitude and of adjustable waveform in a test circuit and presenting this waveform on the screen of an oscilloscope.

transient behavior—In a power supply, the general attitude of response in terms of amplitude and time.

transient distortion—Distortion due to the inability of a system to reproduce or amplify transients linearly.

transient magnistor—A high-speed saturable reactor in which an alternating electric current in the form of a sine-wave carrier or pulses is passed through a signal winding and modulated by variations of current passing through a control coil.

transient motion—Any motion which has not reached, or has ceased to be, a steady state.

transient oscillation—A momentary oscillation which occurs in a circuit during switching.

transient overshoot — The largest transient deviation of the measured quantity following the dynamic crossing of the final value as a result of the application of a specified stimulus.

transient peak-inverse voltage—Under specified conditions, the maximum allowable instantaneous value of nonrecurrent reverse (negative) voltage that may be applied to the anode of an scr with gate open.

transient phenomena—Rapidly changing actions occurring in a circuit during the interval between closing of a switch and settling to steady-state conditions, or between any other temporary actions occurring after some change in a circuit or its constants.

transient recovery time—Also called recovery time, transient response time, or response time. 1. The intervals between the time a transient deviated from a specified amplitude range and the time it returns and remains within the specified amplitude range. The amplitude range centered about the average of the steady-state values that exist immediately before and after the transient. 2. The time required for the output voltage of a power supply to come back to within a level approximating the normal d-c output following a sudden change in load current.

transient response—1. The transient of a dependent variable resulting from an abrupt change of an independent variable with all other variables constant. 2. Also called step-function response. Commonly, the characteristic response of a system to a unit sweep or unit impulse. The elements of transient response most commonly specified are: rise time, fall time, overshoot, undershoot, preshoot, and ringing. 3. (input line changes) The transient of an output variable (voltage, current, or power) of a electronic power supply resulting from an abrupt specified change of input line voltage, with all other conditions held constant. 4. (load changes) The transient of an output variable (voltage current, or power) of an electronic power supply resulting from an abrupt specified change of load, with all other conditions held constant.

transient response time—*See* Transient Recovery Time.

transient state—A temporarily abnormal condition of a variable.

transinformation (of an output symbol about an input symbol)—Also called mutual information. The difference between the information content of the input symbol and the conditional information content of the input symbol which is given the output symbol.

transistance — The characteristic of an electrical element which makes possible the control of voltages, currents, or flux so as to produce gain or switching action in a circuit. Examples of the physical realization of transistance occur in transistors, diodes, saturable reactors, etc.

transistor—An active semiconductor device, usually made of silicon or germanium, having three or more electrodes. The three main electrodes used are the emitter, base, and collector. Conduction is by means of electrons (elementary particles having the smallest negative electrical charge that can exist) and holes (mobile electron vacancies equivalent to a positive charge).

transistor action—The physical mechanism of amplification in a junction transistor.

transistor amplifier—An amplifier in which the required amplification is produced by one or more transistors.

transistor base — The region between an emitter and a collector of a transistor, into which minority carriers are injected.

transistor chip—An unencapsulated transistor element of very small size used in microcircuits.

transistor digital computer—A computer designed to operate satisfactorily in aircraft flying at supersonic speeds. Vacuum-tube failure and heat are eliminated by use of approximately 800 transistors. The power required for operation is less than 100 watts.

transistor dissipation — The power dissipated in the form of heat by the collector. The difference between the power supplied to the collector and the power delivered by the transistor to the load.

transistorized—Pertaining to equipment or a design in which transistors instead of vacuum tubes are used.

transistor oscillator—An oscillator which uses a transistor in place of an electron tube.

transistor pentode—A transistor designed for mixing, modulating, or switching, and containing the equivalent of three emitters, a base, and a collector.

transistor radio—A radio receiver which uses transistors in place of electron tubes.

transistor resistor logic—Abbreviated TRL An early form of logic-circuit design that was employed prior to the advent of monolithic circuits. The input elements are resistors; however, unlike RTL there is only one active output transistor. In discrete circuit designs, this logic form was the least expensive since it required a minimum number (one) of active devices.

transistor seconds — Also called fallouts. Those transistors that remain after the firsts (units meeting rigid specifications for a specific application) have been removed from the production line.

transistor transistor logic — Abbreviated TTL or T^2L. Also called multiemitter transistor logic. A logic circuit design similar to DTL, with the diode inputs replaced by a multiple emitter transistor. In a four-input DTL gate, there are four diodes at the input. A four-input TTL gate will have four emitters of a single transistor as the input element. TTL gates using NPN transistors are positive-level NAND gates or negative-level NOR gates.

transit angle—The product of angular frequency and the time taken for an electron to cross a given path.

transition—A change in the state of a teletypewriter line—the act of a line going from a marking state to a spacing state, or vice versa.

transition card — A card which signals the computer that the reading-in of a program has ended and that the carrying out of the program has started.

transition error—In an encoder, the difference between the shaft angle at which a code-position change should occur and the angle at which it actually does occur.

transition factor—*See* Reflection Factor.

transition frequency — 1. Also called crossover frequency. In a disc-recording system, the frequency corresponding to the point of intersection of the asymptotes, to the constant-amplitude and the constant-velocity portions of its frequency-response curve. This curve is plotted with output-voltage ratio in decibels as the ordinate, and the logarithm of the frequency as the abscissa. 2. In a transistor, the product of the magnitude of the small-signal, common-emitter, forward-current transfer ratio times the frequency of measurement when this frequency is high enough that the magnitude is decreasing with a slope of approximately 6 db per octave.

transition layer—*See* Transition Region.

transition-layer capacitance—*See* Depletion-Layer Capacitance.

transition loss—At any point in a transmission system, the ratio of the available power from that part of the system ahead of the point under consideration, to the power delivered to that part of the system beyond the point under consideration.

transition point—A point at which the circuit constants change in such a way that a wave being propagated along the circuit is reflected.

transition region—Also called transition layer. The region, between two homogeneous semiconductor regions, in which the impurity concentration changes.

transition temperature—The temperature below which the electrical resistance of a material becomes too small to be measured.

transitron—A thermionic tube circuit the action of which depends on the negative transconductance of the suppressor grid of a pentode with respect to the screen grid.

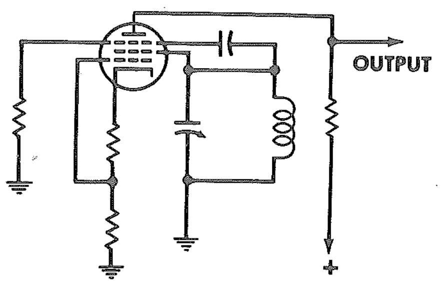

Transitron oscillator.

transitron oscillator—A negative-transconductance oscillator employing a pentode tube with a capacitor connected between the screen and suppressor grids. The suppressor grid periodically divides the current between the screen grid and anode, thereby producing oscillations.

transit time—1. The time taken for a charge carrier to cross a given path. 2. The average time a minority carrier takes to diffuse from emitter to collector in a junction transistor. 3. The time an electron takes to cross the distance between the cathode and anode.

transit-time mode—A condition of oscillator operation, corresponding to a limited range of drift-space transit angle, for which the electron stream introduces a negative conductance into the coupled circuit.

translate—To change computer information from one language to another without significantly affecting the meaning.

translational morphology—The structural characterization of an electronic component in which it is possible to identify the areas of patterns of resistive, conductive, dielectric, and active materials in or on the surface of the structure with specific corresponding devices assembled to perform an equivalent function.

translation loss—Also called playback loss. The loss in the reproduction of a mechanical recording, whereby the amplitude of motion of the reproducing stylus differs from the recorded amplitude in the medium.

translator—1. A device which transforms signals from the form in which they were generated, into a useful form for the purpose at hand. (*Also see* Program Assembly.) 2. A television receiver and low-power transmitter which receives television signals on one channel and retransmits them on another channel (usually any uhf channel from 70 to 83) to valleys and like areas which cannot receive the direct signals. 3. In telephone equipment, the device that converts dialed digits into call-routing information.

transmissibility—Ratio of the response amplitude of the system in steady-state forced vibration, to the excitation amplitude. The ratio may be between forces, displacements, velocities, or accelerations.

transmission—Conveying electrical energy from point to point along a path.

transmission anomaly—The difference, in decibels, between the total loss in intensity and the reduction in intensity that would be due to an inverse-square divergence.

transmission band—The frequency range above the cutoff frequency of a waveguide.

transmission band filter—A frequency band of free transmission; i.e., a frequency band in which, if dissipation is disregarded, the attenuation constant is zero.

transmission coefficient—For a transition or discontinuity between two transmission media at a given frequency, the ratio of some quantity associated with the transmitted wave at a specified point in the second medium, to the same quantity associated with the incident wave at a specified point in the first medium.

transmission facility—The transmission medium and all the associated equipment required to transmit a message.

transmission gain—*See* Gain.

transmission level—The level of signal power at any point in a transmission system. It is equal to the ratio of the power at that point to the power at some point in the system chosen as a reference point. This ratio is usually expressed in decibels.

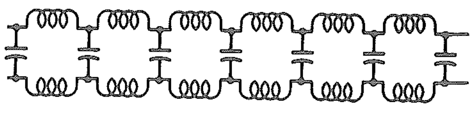

Transmission lines.

transmission line—A material structure forming a continuous path from one place to another, and used for directing the transmission of electromagnetic energy along this path.

transmission-line coupler—A coupler that allows the passage of electric energy in either direction between balanced and unbalanced transmission lines.

transmission-line-tuned, absorption-type frequency meter—A frequency meter using a tuned length of wire or a coaxial cavity as the frequency-determining element.

transmission-line-tuned, transmission-type frequency meter—A frequency meter using a tuned length of wire or a coaxial cavity as the frequency-determining element.

transmission loss—In communication, a general term denoting a decrease in power in transmission from one point to another. Usually expressed in decibels.

transmission measuring set—A measuring instrument comprising a signal source and a signal receiver having known impedances, for measuring the insertion loss or gain of a network or transmission path connected between those impedances.

transmission mode—A form of propagation characterized by the presence of any one of the elemental types of TE, TM, or TEM waves along a transmission line.

transmission primaries—The set of three primaries, either physical or nonphysical, each chosen to correspond in amount to one of the three independent signals contained in the color-picture signal. The I, Q, and Y signals.

transmission regulator—A device which maintains the transmission substantially constant over a system.

transmission response—The conversion of electrical energy traveling at the speed of light to sonic energy traveling at the speed of sound.

transmission speed—The number of information elements sent per unit time, usually expressed as bits, characters, word groups, or records per second or per minute.

transmission system—An assembly of elements capable of functioning together to transmit signal waves.

transmission time—The absolute time interval between transmission and reception of a signal.

transmission-type frequency meter—A frequency meter in which a tuned electrical circuit or a cavity is used to transmit the energy from the signal source under test to a detecting load.

transmit—To send a program, message, or other information from one location to another.

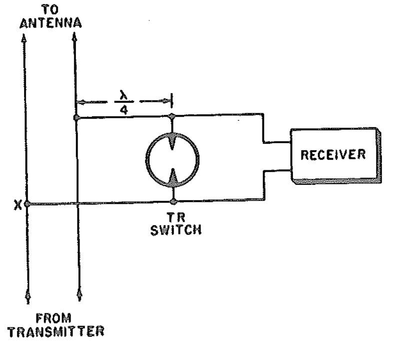

Transmit-receive switch.

transmit-receive switch—Also called a TR switch, TR box, TR tube, or duplex assembly. An automatic device employed in a radar to prevent the transmitted energy from reaching the receiver, but allowing the received energy to do so without appreciable loss.

transmit-receive tube—*See* TR Tube.

transmittance—1. Of a material, the ratio of the radiant power transmitted through the material, to the incident radiant power. 2. As a transfer function, a response function for which the variables are measured at different ports (terminal pairs).

transmitted-carrier operation—Amplitude-modulated carrier transmission in which the carrier wave is transmitted.

transmitted wave—*See* Refracted Wave.

transmitter—1. Equipment used to generate and amplify an r-f carrier signal, modulate this carrier with intelligence, and radiate the modulated r-f carrier into space. 2. In telephony, the microphone that converts sound waves into electrical signals at an audio-frequency rate.

transmitter distributor—In teletypewriter operations, a motor-driven device which translates teletypewriter code combinations from perforated tape into electrical impulses, and transmits these impulses to one or more receiving stations.

transmitter frequency tolerance—The extent to which the carrier frequency of a transmitter may legally depart from the assigned frequency.

transmitter start code—Usually a two-letter call that is sent to an outlying machine to automatically turn on its tape transmitter.

transmitting antenna—A device for converting electrical energy into electromagnetic radiation capable of being propagated through space.

transmitting current response—Of an electroacoustic transducer used for sound emission, the ratio of the sound pressure apparent at a distance of 1 meter in a specified direction from the effective acoustic center of the transducer, to the current flowing at the electric input terminals. Usually expressed in decibels above a reference-current response of 1 microbar per ampere.

transmitting efficiency (projector efficiency)—Ratio of the total acoustic power output of an electroacoustic transducer, to the electric power input.

transmitting loop loss—That part of the repetition equivalent assignable to the station set, subscriber line, and battery supply circuit which are on the transmitting end.

transmitting power response (projector power response)—Of an electroacoustic transducer used for sound emission, the ratio of the effective sound pressure apparent at a distance of 1 meter in a specified direction from the effective acoustic center of the transducer, to the electric power input. Usually expressed in decibels above a reference response of 1 microbar square per watt of electric power input.

transmitting station—A location at which the transmitter of a radio system and its antenna and associated equipment are grouped.

transmitting voltage response—Of an electroacoustic transducer used for sound emission, the ratio of the sound pressure apparent at a distance of 1 meter in a specified direction from the effective acoustic center of the transducer, to the signal voltage applied at the electric input terminals. Usually expressed in decibels above a reference-voltage response of 1 microbar per volt.

transonic speed—A speed of 600 to 900 miles per hour, corresponding to about Mach 0.8 to Mach 1.2.

transponder — A radio transmitter-receiver which transmits identifiable signals automatically when the proper interrogation is received. (*Also see* Pulse Repeater.)

transponder dead time—The time interval between the start of a pulse and the earliest instant at which a new pulse can be received or produced by a transponder.

transponder efficiency—A ratio expressed as a percentage of the number of replies to the number of interrogations from a transponder.

transponder suppressed time delay — The over-all fixed delay between reception of an interrogation and the transmission of a reply.

transportable transmitter—Sometimes called a portable transmitter. A transmitter designed to be readily carried from place to place, but normally not operated while in motion.

transport time—In an automatic system, the time required to move an object, element, or information between two predetermined positions.

transposition—An interchange of the positions of the conductors of a circuit. The term is most frequently applied to open-wire circuits.

transposition blocks — Spreaders used to space and reverse the relative positions of two conductors at fixed intervals.

transposition section—A length of open-wire line to which a fundamental transposition design or pattern is applied as a unit.

transrectification — The rectification that occurs in one circuit when an alternating voltage is applied to another circuit.

transrectification characteristic — The graph obtained by plotting the direct voltage values for one electrode of a vacuum tube as abscissas against the average current values in the circuit of that electrode as ordinates, for various values of alternating voltage applied to another electrode as a parameter. The alternating voltage is held constant for each curve, and the voltages on other electrodes are maintained constant.

transrectification factor — The change in average current of an electrode, divided by the change in the amplitude of the alternating sinusoidal voltage applied to another electrode (the direct voltages of this and other electrodes being maintained constant).

transrectifier—A device, ordinarily a vacuum tube, in which rectification occurs in one electrode circuit when an alternating voltage is applied to another electrode.

transresistance amplifier—An amplifier that supplies an output voltage proportional to an input current. The transfer function of the amplifier is $e_o/I_m = R_m$, where R_m is the transresistance.

transverse-beam traveling-wave tube — A traveling-wave tube in which the electron beam intersects the signal wave rather than moving in the same direction.

transverse cross-talk coupling—Between a disturbing and a disturbed circuit in any given section, the vector summation of the direct couplings between adjacent short lengths of the two circuits, without dependence on intermediate flow in nearby circuits.

transverse electric wave — Abbreviated TE wave. In a homogeneous isotropic medium, an electromagnetic wave in which the electric field vector is everywhere perpendicular to the direction of propagation.

transverse electromagnetic wave — Abbreviated TEM wave. In a homogeneous isotropic medium, an electromagnetic wave in which the electric and magnetic field vectors both are everywhere perpendicular to the direction of propagation.

transverse-field traveling-wave tube — A traveling-wave tube in which the traveling electric fields which interact with the electrons are essentially transverse to the average motion of the electrons.

transverse interference — Interference occurring across terminals or between signal leads.

transverse magnetic E-mode — A type of mode in which the longitudinal component of the magnetic field is zero and the longitudinal component of the electric field is not zero.

transverse magnetic wave—Abbreviated TM wave. In a homogeneous isotropic medium, an electromagnetic wave in which the magnetic field vector is everywhere perpendicular to the direction of propagation.

transverse magnetization—Magnetization of a recording medium in a direction perpendicular to the line of travel and parallel to the greatest cross-sectional dimension.

transverse recording—The technique of recording with rotating heads which are oriented perpendicular to the edge and surface of the tape.

transverse wave — A wave in which the direction of displacement is perpendicular to the direction of propagation at each point of the medium.

trap—1. A selective circuit that attenuates undesired signals but does not affect the desired ones. (*Also see* Wave Trap.) 2. A crystal imperfection which can trap carriers.

trapezoidal distortion—Distortion in which a televised picture has the shape of a trapezoid (wide at top or bottom) instead of a rectangle. It is due to the interaction between the vertical- and horizontal-deflection coils (or plates) of the cathode-ray tube.

trapezoidal generator—An electronic circuit that produces a trapezoidal voltage wave.

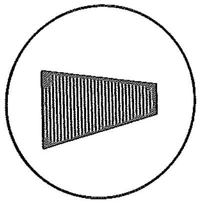

Trapezoidal pattern.

trapezoidal pattern—An oscilloscope pattern which indicates the percentage of modulation in an amplitude-modulated system.

trapezoidal wave—1. A trapezoidal-shaped waveform. 2. A square wave onto which a sawtooth has been superimposed. It is the voltage wave necessary to give a linear deflection current through the coils of a magnetic cathode-ray tube.

trapped flux — In a material in the superconducting state, magnetic flux linked with a closed superconducting loop.

trapping—The holding of electrons or holes by any of several mechanisms in a crystal, thereby preventing them from moving.

traveling detector—A probe mounted on a slider and free to move along a longitudinal slot cut into a waveguide or coaxial transmission line. The traveling detector is connected to auxiliary measuring apparatus and used for examining the relative magnitude of any standing-wave system.

traveling plane wave—A plane wave in which each frequency component has an exponential variation of amplitude and a linear variation of phase in the direction of propagation.

traveling wave—The resulting wave when the electric variation in a circuit takes the form of translation of energy along a conductor, such energy being always equally divided between current and potential forms.

traveling-wave magnetron — A traveling-wave tube in which the electrons move in crossed static electric and magnetic fields that are substantially normal to the direction of wave propagation.

traveling-wave magnetron oscillations — Oscillations sustained by the interaction between the space-charge cloud of a magnetron and a traveling electromagnetic field with approximately the same phase velocity as the mean velocity of the cloud.

traveling-wave parametric amplifier — A parametric amplifier which has a continuous iterated structure incorporating nonlinear reactors and in which the signal, pump, and difference-frequency waves are propagated along the structure.

traveling-wave phototube — A traveling-wave tube that has a photocathode and a window that admits a modulated laser beam, which causes the emission of a current-modulated photoelectron beam. This beam is then accelerated by an electron gun and directed into the helical slow-wave structure of the tube.

traveling-wave tube — Abbreviated twt. A tube in which a stream of electrons interacts continuously or repeatedly with a guided electromagnetic wave moving substantially in synchronism with it, and in such a way that there is a net transfer of energy from the stream to the wave.

traveling-wave tube amplifier — A power-amplifying piece of equipment yielding about 30 db of gain over broad bandwidths. The units are built around certain twt's in the most commonly used frequency bands, and they have external modulation circuitry incorporated.

traveling-wave-tube interaction circuit — An extended electrode arrangement used in a traveling-wave tube to propagate an electromagnetic wave in such a manner that the traveling electromagnetic fields are retarded to the point where they extend into the space occupied by the electron stream.

TR box—*See* Transmit-Receive Switch.

TR cavity—The resonant portion of a TR switch.

treating—Any of the processes of applying

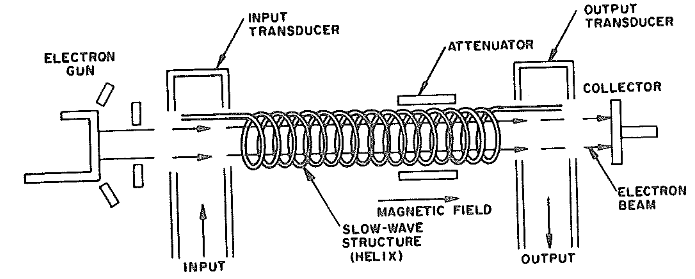

SCHEMATIC DIAGRAM SHOWING BASIC COMPONENTS OF A TRAVELING-WAVE-TUBE AMPLIFIER

Traveling-wave tube.

varnishes or other insulating compounds to insulation, coils, or windings. This includes the processes of impregnation, soaking, and surface coating by any of the various methods.

treble—The higher part in harmonic music or voice; of high or acute pitch. In music, the frequencies from middle C (256 hertz) and upward.

treble boost—Deliberate adjustment of the amplitude-frequency response of a system or component to accentuate the higher audio frequencies.

tree—A set of connected branches without meshes.

treeing — A progressive type of insulation failure in which branching hollow channels slowly penetrate the insulation at rather low applied voltages.

tremolo—1. The amplitude modulation of an audio tone. 2. A warbling or fluctuating effect approximately seven times per second, to the tone of an instrument characterized by a variation in intensity rather than pitch. (*Also see* Vibrato).

trf—Abbreviation for tuned radio frequency.

tri—Abbreviation for triode.

Triac—A General Electric trade name for a gate-controlled full-wave a-c silicon switch designed to switch from a blocking state to a conducting state for either polarity of applied voltage with positive or negative gate triggering.

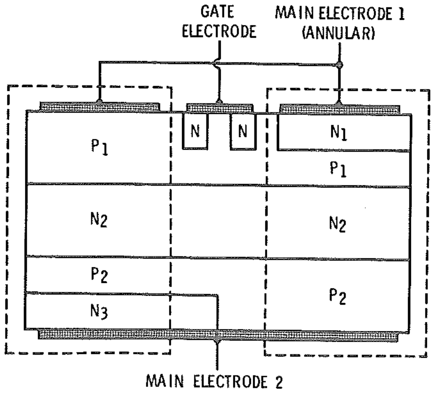

Triac.

triad—1. Three radio stations operated as a group for determining the position of aircraft or ships. 2. A group of three dots, one of each color-emitting phosphor, on the screen of a color picture tube. 3. A group of three bits or three pulses, usually in sequence on one wire or simultaneously on three wires.

triangulation—A method of finding the location of a third point by taking bearings from two fixed points a known distance apart. The third point will be at the intersection of the two bearing lines.

triax—A type of shielded conductor that employs a shield and jacket over the primary insulation, plus a second shield and jacket over all. Aside from applications requiring maximum attenuation of radiated signals or minimum pickup of external interference, this cable can also be used to carry two separate signals.

triaxial cable — A special form of coaxial cable containing three conductors.

triaxial speaker—A dynamic speaker unit consisting of three independently driven units combined into a single speaker.

tribo—A prefix meaning due to or pertaining to friction.

triboelectric—Pertaining to electricity generated by friction.

triboelectricity — Electrostatic charges generated due to friction between different materials.

triboelectric series—A list of substances arranged so that any of them can become positively electrified when rubbed with one farther down the list, or negatively charged when rubbed with one farther up the list.

tributary circuit—A circuit which connects an individual drop, or drops, to a switching center.

tributary station—A communications terminal consisting of equipment compatible for the introduction of messages into, or reception from, its associated relay station.

trickle charge—1. A continuous charge of a storage battery at a slow rate approximately equal to the internal losses and suitable to maintain the battery in a fully charged condition. 2. Very slow rates of charge suitable not only in compensating for internal losses, but in restoring small, intermittent discharges to the load circuit delivered from time to time.

trickle charger—A device for charging a storage battery at a low rate continuously, or for several hours at one time.

tricolor camera—A television camera designed to separate reflected light into three frequency groups, each corresponding to the light energies of the three primary colors. The camera transforms the intensity variations of each primary into amplitude variations of an electrical signal.

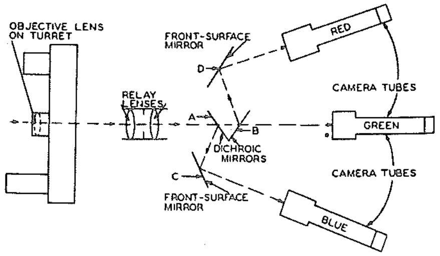

Tricolor camera.

tricolor picture tube—A picture tube which reproduces a scene in terms of the three light primaries.

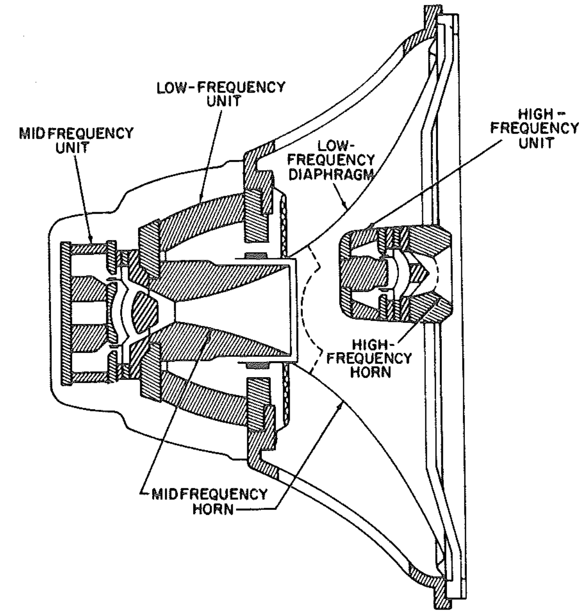

Triaxial speaker.

tricon—A radionavigational system in which the airborne receiver accepts pulses from a triplet (chain of three stations) in a variable time sequence so that the pulses arrive at the same time even though traveling over paths of various lengths.

trigatron — An electronic switch in which breakdown of an auxiliary gap initiates conduction.

trigger—1. To start action in another circuit, which then functions for a certain length of time under its own control. 2. A pulse that starts an action.

trigger action—The instantaneous initiation of main current flow by a weak controlling impulse in a device.

trigger circuit—1. A circuit with two conditions of stability, with means for passing spontaneously or through application of an external stimulus from one to the other when certain conditions are satisfied. 2. A circuit in which an electron tube performs the function of a relay. Impulses applied to the input of the tube produce corresponding impulses in the output circuit, starting a chain of events. 3. *See* Flip-Flop Circuit.

trigger countdown—A process that reduces the repetition rate of a triggering signal.

trigger diode — A symmetrical three-layer avalanche diode used to control scr's and Triacs. It has a symmetrical switching mode, and therefore fires whenever the breakover voltage is exceeded in either polarity of applied voltage.

triggered blocking oscillator—A blocking oscillator that can be reset to its starting condition by the application of a trigger voltage.

triggered spark gap—A fixed spark gap in which the discharge passes between two electrodes and is struck (started) by an auxiliary electrode called the trigger, to which low-power pulses are applied.

triggered sweep — In a cathode-ray oscilloscope, a sweep initiated by a signal pulse.

trigger electrode—*See* Starter.

triggering—The starting of circuit action, which then continues for a predetermined time under its own control.

triggering signal—The signal from which a trigger is derived.

trigger level—1. In a transponder, the minimum receiver input capable of causing the transmitter to emit a reply. 2. The instantaneous level of a triggering signal at which a trigger is to be generated. Also the name of the control which selects the level.

trigger point—The amplitude point on the input pulse at which triggering of the sweep of a cathode-ray oscilloscope occurs.

trigger pulse—A pulse used for triggering.

trigger-pulse steering—In transistors, the routing or directing of trigger signals (usually pulses) through diodes or transistors (called steering diodes or transistors) so that the signals affect only one of several associated circuits.

trigistor—A bistable PNPN semiconductor component with characteristics comparable to those of a flip-flop or bistable multivibrator.

tri-gun color picture tube—*See* Three-Gun Color Picture Tube.

trim control—On some regulated power supplies, a control used to make minor adjustments of output voltage.

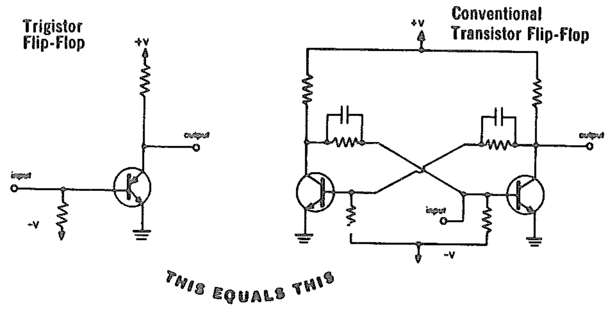

Trigistor.

trimmer—*See* Trimmer Capacitor.

trimmer capacitor—Also called a trimmer. A small variable capacitor associated with another capacitor and used for fine adjustment of the total capacitance of the combination.

trimmer potentiometer—A lead-screw–actuated potentiometer.

trimmer resistor—A small rheostat used in place of a fixed resistor to permit adjustment of resistance values in a circuit during initial calibration of an equipment or when recalibration is required.

trimming—The fine adjustment of capacitance, resistance, or inductance in a circuit.

trinistor—A three-terminal silicon semiconductor device with characteristics similar to those of a thyratron and used for controlling large amounts of power.

trinoscope—Any assembly of three kinescopes producing the red, green, and blue images required for tricolor television optical projection (e.g., for theater tv).

triode—A three-electrode electron tube containing an anode, a cathode, and a control electrode (grid).

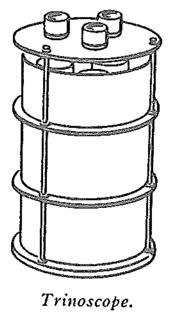

Trinoscope.

triode amplifier—An amplifier in which only triode tubes are used.

triode-heptode converter — A superheterodyne converter circuit which uses a triode local oscillator and a heptode converter, both contained in one envelope.

triode-hexode converter—A superheterodyne converter circuit which uses a triode local oscillator and a hexode converter, both contained in one envelope.

triode-pentode — A dual-purpose vacuum tube containing a triode and a pentode in the same envelope.

triode PNPN-type switch — A PNPN-type switch having an anode, cathode, and gate terminal.

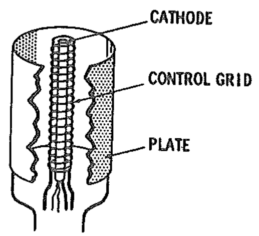

Triode.

trip coil—An electromagnet in which a moving armature trips a circuit breaker or other protective device and thereby opens a circuit under abnormal conditions.

triple detection—*See* Double Superheterodyne Reception.

triplet — 1. Three radionavigational stations operated as a group for the determination of position. 2. The waveform of the output voltage of a delay line when the input pulse has a width approximately equal to the resolution of the delay line.

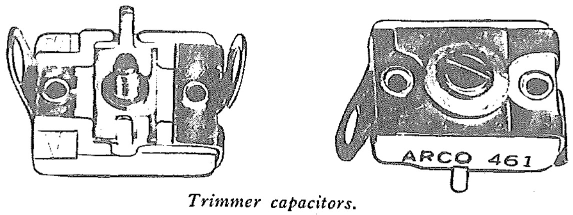

Trimmer capacitors.

triplex cable—A cable made up of three insulated single-conductor cables twisted together with or without a common insulating covering.

triplexer—A dual duplexer which permits the use of two receivers simultaneously and independently in a radar system by disconnecting the receivers during the transmitted pulse.

tripod—A three-legged camera support.

tripping device—A mechanical or electromagnetic device used for opening (turning off) a circuit breaker or starter, either when certain abnormal electrical conditions occur or when a catch is actuated manually.

tripping relay—Also called trip-free relay. A device which functions to trip a circuit breaker, contactor, or equipment or to permit immediate tripping by other devices; or to prevent immediate reclosure of a circuit interrupter, in case it should open automatically even though its closing circuit is maintained closed.

trisistor — A fast-switching semiconductor consisting of an alloyed junction PNP device in which the collector is capable of electron injection into the base. Its characteristics resemble those of a thyratron electron tube, and the switching time is in the nanosecond range.

tristimulus values—The amounts of each of the three primary colors that must be combined to match a sample.

tri-tet oscillator—A crystal-controlled, electron-coupled vacuum-tube oscillator which is isolated from the output circuit through use of the screen-grid electrode as the oscillator anode. Used for multiband operation because it generates strong harmonics of the crystal frequency.

tritium — A radioactive isotope of hydrogen with an atomic number of 3.

TRL — Abbreviation for transistor resistor logic.

trombone—An adjustable U-shaped coaxial-line matching assembly.

tropicalization—A chemical treatment developed to combat the fungi that ruin electronic equipment in hot, humid jungle regions.

tropo—*See* Tropospheric Scatter Communication.

troposphere—The lower layer of the earth's atmosphere, extending to about 60,000 feet at the equator and 30,000 feet at the poles. In this area, the temperature generally decreases with altitude, clouds form, and convection is active.

tropospheric scatter — The propagation of radio waves by scattering as a result of irregularities or discontinuities in the physical properties of the troposphere.

tropospheric scatter communication—Also called tropo. A method or system of transmitting, within the troposphere, microwaves in the uhf or shf bands to effect radiocommunication between two points on the earth's surface separated by moderate distances of from 70 to 600 miles. Such a span or hop may be augmented by other spans in tandem to permit end-to-end or through circuits up to many thousands of miles. More specifically, this method of communication is now generally understood to embrace a radio system that permits communication over the distances indicated, with excellent reliability and good information capacity, using relatively high transmitted power, frequency modulation, and highly sensitive receiving apparatus.

tropospheric superrefraction — The phenomenon occurring in the troposphere whereby radio waves are bent sufficiently to be returned to the earth.

tropospheric wave—A radio wave that is propagated by reflection from a place of abrupt change in the dielectric constant or its gradient in the troposphere. In some cases, the ground wave may be so altered that new components appear to arise from reflections in regions of rapidly changing dielectric constants; when these components are distinguishable from the other components, they are called tropospheric waves.

trouble—Failure of a circuit or element to perform in a standard manner.

trouble-location problem—A computer test problem, the incorrect solution of which supplies information on the location of faulty equipment. It is used after a check problem has shown that a fault exists.

troubleshoot—To look for, locate, and repair malfunctions in equipment.

TR switch—*See* Transmit-Receive Switch.

TR tube—A gas-filled radio-frequency switching tube used to protect the receiver in pulsed radio-frequency systems. (*Also see* Transmit-Receive Switch.)

true bearing—A bearing given in relation to geographic north, as opposed to a magnetic bearing.

true course—A course in which the direction of the reference line is true rather than magnetic north.

true homing—The following of a course in such a way that the true bearing of an aircraft or other vehicle is held constant.

true north—Geographic north.

true ohm—The actual value of the practical unit of resistance. It is equal to 10^9 absolute electromagnetic units of resistance.

true power—The average power consumed by a circuit during one complete cycle of alternating current.

true radio bearing—*See* Radio Bearing.

true random noise—A noise characterized by a normal or Gaussian distribution of amplitudes.

truncate—To drop digits of a number of terms in a series, thereby lessening precision. For example, the value 3.14159265 (π) when truncated to five figures is 3.1415, whereas it could be rounded off to 3.1416.

truncated paraboloid—A paraboloid reflector in which a portion of the top and bottom have been cut away to broaden the main radiated lobe in the vertical plane.

truncation error—The error resulting from the use of only a finite number of terms of an infinite series, or from the approximation of operations in the infinitesimal calculus by operations in the calculus of finite differences.

trunk—1. A single message circuit between two points, both of which are switching centers and/or individual message distribution points. 2. A communications channel between two different offices, or between groups of equipment within the same office. 3. A telephone line or channel between two central offices or switching devices, which is used in providing telephone connections between subscribers.

trunk circuit—A circuit which connects two switching centers.

trunk loss—That part of the repetition equivalent assignable to the trunk used in the telephone connection.

truth table—A logic function listing all possible combinations of input values and indicating, for each combination, the true output values.

T_s—Abbreviation for color temperature.

Tschebyscheff filters—Filter networks that are designed to exhibit a predetermined ripple in the passband (ripple amplitudes from 0.01 db to 3 db are common) in exchange for which they provide a more rapid attenuation above cutoff—which, unlike their passband response, is monotonic.

TSC—Abbreviation for transmitter start code.

T^2L or TTL—Abbreviation for transistor transistor logic.

tuba—A powerful land-based radar jamming transmitter operated between 480 to 500 MHz. It was developed during World War II for use against night fighter planes.

tube—A hermetically sealed glass or metal envelope in which conduction of electrons takes place through a vacuum or gas.

tube bridge—An instrument used in the precise measurement of vacuum-tube characteristics. It contains one or more bridge-type measuring circuits, plus power supplies and signal sources for all possible electrode combinations.

tube coefficients—Constants that describe the characteristics of a thermionic vacuum tube (e.g., amplification factor, mutual conductance, a-c plate resistance, etc.).

tube complement—The number and types of electron tubes required in an electronic equipment.

tube count—A terminated discharge produced by an ionizing event in a radiation-counter tube.

tube drop—The voltage measured across a tube, from plate to cathode, when the tube is conducting at its normal current rating.

tube electrometer—A thermionic vacuum tube adapted for use as an electrometer, to measure potential difference.

tube heating time—The time required for the coolest portion of a mercury-vapor tube to attain its operating temperature.

tube noise—Noise originating in a vacuum tube (e.g., from shot effect, thermal agitation, etc.).

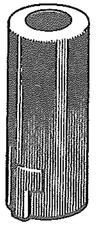

Tube shield.

tube shield—A metallic enclosure placed over a vacuum tube to prevent external fields from interfering with the function of the tube.

tube socket—A receptacle which provides mechanical support and electrical connection for a vacuum tube.

tube tester—A test instrument for indicating the condition of vacuum tubes used in electronic equipment.

tube voltage drop—1. The anode voltage in an electron tube during conduction. 2. In a gas tube, the anode voltage during the conducting period.

tubing—Extruded nonsupported plastic materials designed as protection and electrical insulation for exposed components of electrical and electronic assemblies, as opposed to coated, braided, woven tube termed sleeving.

tubular capacitor—A paper, ceramic, or electrolytic capacitor shaped like a cylinder; leads or lugs project from one or both ends.

tunable echo box—An echo box consisting of an adjustable cavity operating in a single mode. When the echo box is calibrated, the setting of the plunger at resonance will indicate the wavelength.

tunable magnetron—A magnetron that can be tuned mechanically or electronically over a limited band of frequencies.

tuned—Adjusted to resonate or operate at a specified frequency.

tuned amplifier—An amplifier in which the load is a tuned circuit. Thus, the load impedance and amplifier gain vary with the frequency.

tuned-anode oscillator — *See* Tuned-Plate Oscillator.

tuned antenna—An antenna designed, by means of its own inductance and capacitance, to provide resonance at the desired operating frequency.

tuned-base oscillator—A transistor oscillator comparable to a tuned-grid electron-tube oscillator. The frequency-determining device (resonant circuit) is located in the base circuit.

tuned circuit—A circuit consisting of inductance and capacitance which can be adjusted for resonance at the desired frequency.

tuned-collector oscillator—A transistor oscillator comparable to the tuned-plate electron-tube oscillator. The frequency-determining device is located in the collector circuit.

tuned dipole—A dipole antenna that resonates at its operating frequency.

tuned filter—A resonant arrangement of electronic components which either attenuates signals at a particular frequency and passes signals at other frequencies, or vice versa.

tuned filter oscillator—An oscillator in which a tuned filter is used.

tuned-grid oscillator—An oscillator, the frequency of which is determined by a parallel-tuned tank in the grid circuit. The tank is coupled to the plate to provide the required feedback.

tuned-grid-tuned-anode oscillator — *See* Tuned-Grid Tuned-Plate Oscillator.

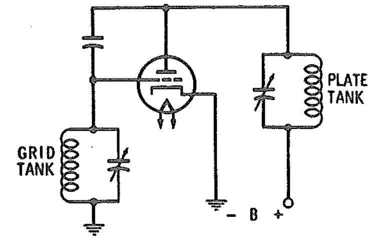

Tuned-grid, tuned-plate oscillator.

tuned-grid, tuned-plate oscillator — Also called tuned-grid–tuned-anode oscillator. An oscillator having parallel-resonant circuits in both the plate and the grid circuits. The necessary feedback is provided by the plate-to-grid interelectrode capacitance.

tuned-plate oscillator — Also called tuned-anode oscillator. An oscillator, the frequency of which is determined by a parallel-tuned tank in the plate circuit. The tank is coupled to the grid to provide the required feedback.

tuned radio-frequency amplifier—A tuned amplifier using resonant-circuit coupling and designed to operate at radio frequencies.

tuned radio-frequency receiver—A radio receiver consisting of several amplifier stages, which are tuned to resonance at the carrier frequency of the desired signal by a ganged variable-tuning capacitor. The amplified signals at the original carrier frequency are fed directly into the detector for demodulation. The resultant audio-frequency signals are again amplified, and are then reproduced by a speaker.

tuned radio-frequency transformer — A transformer used for selective coupling in radio-frequency stages.

tuned-reed frequency meter—A vibrating-reed instrument for measuring the frequency of an alternating current.

tuned relay—A relay having mechanical or other resonating arrangements that limit the response to currents at one particular frequency.

tuned resonating cavity—A resonating cavity half a wavelength long or some multiple of a half wavelength, used in connection with a waveguide to produce a resultant wave with the amplitude in the cavity greatly exceeding that of the wave in the guide. For reception of waves, a detecting grating can be placed at the point of maximum amplitude in the cavity, to convert the energy to a form suitable for amplification in a telephone or television circuit. A tuned cavity is a nonreflecting termination for a guide.

tuned rope—Long lengths of chaff cut to the various lengths necessary for tuning to different wavelengths.

tuned transformer—A transformer, the associated circuit elements of which are adjusted as a whole to be resonant at the frequency of the alternating current supplied to the primary, thereby causing the secondary voltage to build up to higher values than would otherwise be obtained.

tuner—In the broad sense, a device for tuning. Specifically, in radio-receiver practice: 1. A packaged unit capable of producing only the first portion of the functions of a receiver and delivering either r-f, i-f, or demodulated information to some other equipment. 2. That portion of a receiver which contains the circuits that are tuned to resonance at the received-signal frequency and those which are tuned to the local-oscillator frequency.

tungar rectifier—A gaseous rectifier containing argon gas. It is employed in battery chargers and low-voltage power supplies.

tungar tube—A phanatron (hot-cathode, gas-filled rectifier tube) having a heated filament serving as the cathode and a graphite disc as the anode in a bulb filled with low-pressure argon. Used chiefly in battery chargers.

tungsten—A metal used in the manufacture of filaments for vacuum tubes and in making contact points for switches and other parts where sparking may occur. After the tungsten is made ductile by rolling, swaging, and hammering, it is very tough.

tungsten filament—A filament used in incandescent lamps, and in thermionic vacuum tubes and other tubes requiring an incandescent cathode. Smaller tungsten filaments are operated in a vacuum, while those for larger lamps are used in an inert gas at about ordinary atmospheric pressure.

tuning—The adjustment relating to frequency of a circuit or system to secure optimum performance. Commonly, the adjustment of a circuit or circuits to resonance.

tuning capacitor—A variable capacitor for adjusting the natural frequency of an oscillatory or resonant circuit.

tuning circuit—A circuit containing inductance and capacitance, either or both of which may be adjusted to make the circuit responsive to a particular frequency.

tuning coil—A variable inductance for adjusting the natural frequency of an oscillatory or resonant circuit.

tuning control—A control knob that adjusts all tuned circuits simultaneously.

tuning core—Normally a molded iron core for permeability tuning, into which an adjusting screw has been cemented or molded.

tuning eye—Slang for a cathode-ray tuning indicator.

tuning fork—A two-pronged hard-steel device that vibrates at a definite natural frequency when struck or when set in motion by electromagnetic means. Used in some electronic equipment as an accurately controllable source of signals, because its vibrations can be formed readily into audio-frequency signals by means of pickup coils.

tuning-fork drive—Control of an oscillator by continuous vibrations of a tuning fork. A high harmonic of the oscillating signal obtained from the fork is selected by filter circuits and is strongly amplified to determine the main-oscillator frequency in a transmitter or other equipment.

tuning in—Adjusting the tuning controls of a receiver to obtain maximum response to the signals of the station it is desired to receive.

tuning indicator—A device that indicates whether or not a receiver is tuned accurately. It is connected to some circuit in which current or voltage is maximum or minimum when the receiver is accurately tuned to give the strongest output signal.

tuning meter—A direct-current meter connected to a receiver circuit and used for determining whether the receiver is accurately tuned to a station.

tuning probe—An essentially lossless probe of adjustable penetration extending through the wall of a waveguide or cavity resonator.

tuning range—The frequency range over which a tuned circuit can be adjusted.

tuning screw—In waveguide technique, an impedance-adjusting element in the form of a rod whose depth of penetration through the wall into a waveguide or cavity is adjustable by rotating a screw.

tuning stubs—Inductor elements, usually adjustable, which are connected to transmission lines at intervals to improve the voltage distribution.

tuning susceptance—The normalized susceptance of an ATR tube in its mount due to its deviation from the desired resonance frequency.

tuning wand—*See* Neutralizing Tool.

tunnel cathode — A metal insulator-metal sandwich. Electrons tunnel from the metal substrate and appear in the metal film as hot electrons. Some of the hot electrons have sufficient energy to pass over the cathode surface barrier into the vacuum.

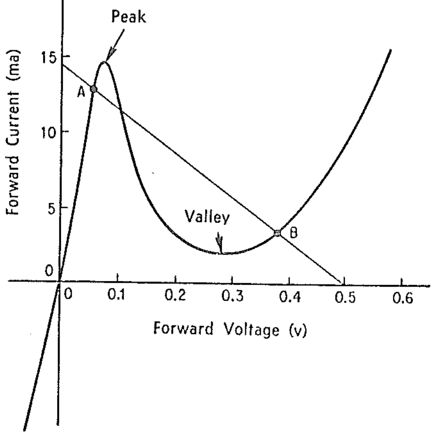

Tunnel diode characteristics.

tunnel diode—Also called Esaki diode. A semiconductor consisting of a PN junction which has been especially constructed to optimize the characteristics desired. It is designated P_tN and has an extremely narrow junction and large impurity concentrations in both regions. It is extremely stable and has excellent environmental characteristics, withstanding large temperature changes and radiation bombardment. The P_tN junction may be combined with other junctions. In operation, the tunnel diode differs from other types of PN junctions, in that it exhibits negative resistance and the current is carried by electrons that travel across the PN barrier by quantum mechanical tunneling. It is so stable that it could be operated in an unencapsulated state under severe moisture and contamination conditions.

tunnel effect—1. The probability that a particle of given potential energy can penetrate a finite barrier of higher potential. 2. Piercing of a rectangular potential barrier in a semiconductor by a particle that does not have sufficient energy to go over the barrier.

tunneling—*See* Tunnel Diode.

tunnel triode — A transistorlike device in which the emitter-base junction is a tunnel diode and the collector-base junction is a conventional diode.

tunneluminescence—The emission of light from a phosphor film deposited on the surface of a thin-film metal-oxide-metal sandwich.

turbidimeter—*See* Opacimeter.

turn—One complete loop of wire.

turn factor—Under stated conditions, the number of turns that a coil of given shape and dimensions placed on a core in a given position must have for a coefficient of self-inductance of 1 henry to be obtained for the core.

turnoff delay time—The time interval between occurrence of the trailing edge of a fast input pulse and the occurrence of the 90% point of the negative-going output waveform.

turn-off reversal—A polarity reversal of the output of the electronic power supply occurring from the turning off of the power of a regulated power supply.

turnoff time—The time that a switching circuit (gate) takes to stop the flow of current in the circuit it is controlling.

turnon delay time—The time interval from the occurrence of the leading edge of a fast input pulse to the occurrence of the 10% point of the positive-going output waveform, assuming that the rise time of the incoming pulses is 1/10 of the rise time of the element to be measured under loaded conditions.

turn-on overshoot — Overshoots occurring from the turning on of power of a regulated power supply.

turn-on reversal—A polarity reversal of the output of the power supply occurring from the turning on of the power supply of a regulated power supply.

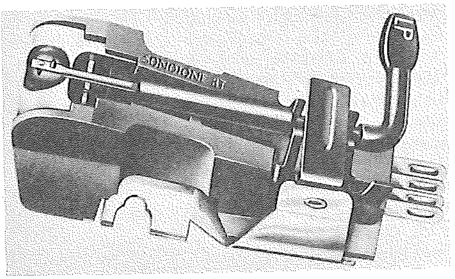

Turnover cartridge.

turnover cartridge—A phonograph cartridge adapted, by the use of two styli, to play both large- and fine-groove records.

turnover frequency—In disc recording, the frequency below which constant-amplitude recording is used and above which constant-velocity recording is employed.

turnover pickup—Also called dual pickup. A pickup designed for playing both standard and microgroove records, using a single magnetic structure. (*Also see* Turnover Cartridge.)

turns ratio—The ratio of the number of turns in the primary winding to the number in the secondary winding of a transformer.

turnstile antenna—An antenna composed of two dipole antennas normal to each other and with their axes intersecting at their midpoints. Usually the currents are equal and in phase quadrature.

turntable—A revolving platform used in recording or playing phonograph records.

turntable rumble—*See* Rumble.

turret—A revolving plate mounted at the front of some television cameras and carrying two or more lenses of different types, to permit rapid interchange of lenses.

turret tuner—A television-receiver tuner containing a separate set of resonating circuit elements for each channel. Each set is mounted on an insulating strip or strips placed on a drum rotated from the channel-selector control, which selects the drum position for the desired channel.

tv—Abbreviation for television.

tvi—Abbreviation for television interference.

tv recording—A permanent record of video signals recorded photographically, electronically, or by other means, and which may be displayed through a television system or projected as a motion-picture film.

tweeter—Also called a high-frequency unit. A speaker intended to reproduce the very high frequencies, usually those above 3,000 Hz, in a high-fidelity audio system.

twenty-one type repeater—A two-wire telephone repeater in which one amplifier serves to amplify the telephone currents in both directions. The circuit is arranged so that the input and output terminals of the amplifier are in one pair of conjugate branches, while the lines in the two directions are in another pair.

twenty-two type repeater—A two-wire telephone repeater with two amplifiers. One amplifies the telephone currents being transmitted in one direction, and the other the telephone currents being transmitted in the other direction.

twin cable—A cable composed of two parallel insulated, stranded conductors having a common covering.

twin check—A continuous check of computer operations accomplished by duplication of equipment and automatic comparison of results.

Twin lead.

twin lead—Also called twin line. A type of transmission line covered by a solid insulation and comprising two parallel conductors, the impedance of which is determined by their diameter and spacing. The three most common impedance values are 75, 150, and 300 ohms.

twin line—*See* Twin Lead.

twinning—The intergrowth of two crystal regions having opposite oriented axes. Two types of twinning may occur; electrical and optical. In electrical twinning the electrical sense of the crystal axes are reversed and the twinned regions will interfere with one another piezoelectrically. It is this effect which limits the high-temperature utility of quartz as a piezoelectric material.

twin-T network—*See* Parallel-T Network.

twin triode—Two triode vacuum tubes in a single envelope.

twin wire—A cable composed of two small, parallel insulated conductors having a common covering.

twist—1. The progressive rotation of the cross section of a waveguide about the longitudinal axis. 2. The deviation from a plane surface measured from one corner to the corner diagonally opposite.

twisted joint—A union of two conductors wound tightly around each other. A sleeve may be used, and it and the conductors twisted.

twisted pair—A cable composed of two small insulated conductors twisted together without a common covering. The two conductors of a twisted pair are usually substantially insulated, so that the combination is a special case of a cord.

twister—A piezoelectric crystal that generates a voltage when twisted.

twistor—A computer memory element containing inclined helical windings of magnet wire on a nonmagnetic wire, with another winding over the helix. Information is stored in the form of polarized helical magnetization.

two-address — In a computer, having the property that each complete instruction includes an operation and specifies the location of two registers, usually one containing an operand and the other the result of the operation.

two-dimensional circuitry — *See* Thin-Film Integrated Circuitry.

two-fluid cell—A cell having unlike electrolytes at the positive and negative electrodes.

two-level system—A laser which uses only two electron energy levels. Electrons in the ground state (level 1) are pumped to the excited state (level 2). The electrons then surrender their energy by stimulated emission and return to the ground state.

two-phase—Also called quarter phase. Having a phase difference of 90 electrical degrees, or one quarter-cycle.

two-phase current—Two currents delivered through two pairs of wires at a phase difference of one quarter-cycle (90°) between them.

two-phase, five-wire system — An alternating-current supply in which four of its conductors are connected as in a four-wire, two-phase system and the fifth is connected to the neutral points of each phase and usually grounded. Despite its name, it is strictly a four-phase, five-wire system.

two-phase, four-wire system—A system of alternating-current supply comprising two pairs of conductors, between one pair of which is maintained an alternating difference of potential displaced in phase by one quarter of a period from an alternating difference of potential of the same frequency maintained between the other pair.

two-phase, three-wire system—An alternating-current supply consisting of three conductors. Between one conductor (known as the common return) and each of the other two, alternating differences of potential which are 90° out of phase with each other are maintained.

two-pilot regulation—The use of two pilot frequencies within a transmitted band so that the change in attenuation due to twist can be detected and compensated for by a regulator.

two-port network — A network with two ports.

two-pulse canceler—A moving-target indicator canceler which compares the phase variation of two successive pulses received from a target. It discriminates against signals with radial velocities which produce a Doppler frequency equal to a multiple of the pulse-repetition frequency.

two-quadrant multiplier — Of an analog computer, a multiplier in which operation is restricted to a single sign of one input variable only.

two-source frequency keying—Keying in which the modulating wave shifts the output frequency between predetermined values derived from independent sources.

two-terminal network—A network that is connected by only two terminals to an external system.

two-terminal-pair network—Also called a four-pole, quadripole, or quadrupole network. A network with four accessible terminals grouped in pairs. One terminal of each pair may coincide with a network node.

two-tone keying—Keying in which the modulating wave causes the carrier to be modulated with one frequency for the marking condition, and a different frequency for the spacing condition.

two-tone modulation — 1. A method of modulation in which two different carrier frequencies are used for the two signaling conditions. 2. In teletypewriter operation, a method of modulation in which two different carrier frequencies are employed for the two signaling conditions. The transition from one frequency to the other is abrupt, with resultant phase discontinuities.

two-track recording—On quarter-inch-wide tape, the arrangement by which only two channels of sound may be recorded, either as a stereo pair in one direction or as separate monophonic tracks (usually in opposite directions).

two-value capacitor motor — A capacitor motor that uses different values of effective capacitance for starting and running.

two-wattmeter method—A method of measuring total power in a balanced or unbalanced three-phase system by adding the readings of two wattmeters, each with its current coil in one phase and its voltage coil connected between it and the third phase.

two-way amplifier—An amplifier in which the right and left channels of a stereo system are both amplified simultaneously by the same tubes, using push-pull circuitry but feeding one signal to the input grids in parallel instead of push-pull. The parallel and push-pull signals are then separated by two output transformers in a matrixing circuit.

two-way communication — Communication between radio stations, each having both transmitting and receiving equipment.

two-way repeater—*See* Repeater.

two-way switch—A switch used for controlling electrical or electronic equipment, components, or circuits from either of two positions.

two-way system—A speaker in which the low and the high frequencies are reproduced separately by two electrically independent speaker elements, each of which is provided with a suitable sound-radiating system.

two-wire channel — A two-way circuit for transmission in either direction.

two-wire circuit—A metallic circuit formed by two conductors insulated from each other. It is possible to use the two conductors as either a one-way transmission path, a half-duplex path, or a duplex path. Also used in contrast with a four-wire circuit to indicate a circuit using one line or channel for transmission of electric waves in both directions.

two-wire repeater — A telephone repeater which provides for transmission in both directions over a two-wire telephone circuit.

two-wire system—A system of electric supply comprising two conductors, with the load connected between them.

twt—Abbreviation for traveling-wave tube.

Twystron — A very high power, hybrid microwave tube combining the input section of a high-power klystron with the output section of a traveling-wave tube. It is characterized by high operating efficiency and wide bandwidths.

type-A facsimile—Facsimile communication in which the images are built up of lines of constant-intensity dots.

type-A waves—Continuous waves.

type A_1 waves—Unmodulated, keyed, continuous waves.

type A_2 waves—Modulated, keyed, continuous waves.

type A_3 waves—Continuous waves modulated by music, speech, or other sounds.

type A_4 waves—Superaudio frequency-modulated continuous waves, as used in a facsimile system.

type A_5 waves—Superaudio frequency-modulated continuous waves as used in television.

type A_0 waves—Composite transmissions and cases not covered by type-A thru type-A_5 waves.

type-B facsimile—Facsimile communication in which the images are built up of lines of dots having a varying intensity (e.g., in telephotography and photoradio).

type-B waves—Keyed, damped waves.

type-printed telegraphy — Telegraphy in which the message is automatically printed at the receiving station.

U

ubitron—An amplifier or oscillator in which an undulating electron beam interacts with an r-f wave. The kinetic energy of the beam is converted into r-f energy (0-type interaction). The undulation of the beam is produced by a periodic magnetic field. This field gives the beam a transverse-velocity component which interacts with the r-f wave.

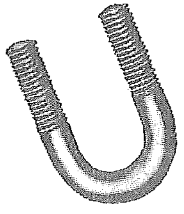

U bolt.

U-bolt—A U-shaped bolt threaded on both ends, for fastening antennas to masts.

uhf—Abbreviation for ultrahigh frequency.

UL—Abbreviation for Underwriters Laboratories, a corporation supported by some underwriters for the purpose of establishing safety standards on types of equipment or components.

ultimately controlled variable—The quantity whose control is the end purpose of an automatic control system.

ultimate sensitivity or threshold — One-half the deadband in a graphic recorder. When the instrument is balanced at the center of the deadband, it denotes the minimum change in measured quantity required to initiate pen response.

ultimate trip current—The smallest value of current that will cause tripping of a circuit breaker under a given set of ambient conditions.

ultimate trip limits—The values of overload of a circuit breaker at which the minimum and maximum limits of the time-current curve become asymptotic; i.e., the limits of current that will trip or not trip the breaker "ultimately." Minimum and maximum limits of ultimate trip are often called calibration.

ultor—An adjective used to identify the picture-tube anode or element farthest from the cathode, the anode to which the highest voltage is applied, or the voltage itself (e.g., the ultor anode is the second anode of the picture tube, and the ultor voltage is the voltage applied to it).

ultor element—The element which receives the highest d-c voltage in a cathode-ray tube.

ultra-audible frequency—*See* Ultrasonic Frequency.

ultra-audion—Any of several special vacuum-tube circuits employing regeneration.

ultra-audion circuit—A regenerative detector circuit in which a parallel-resonant circuit is connected between the grid and the plate of a vacuum tube, and a variable capacitor is connected between the plate and cathode to control the amount of regeneration.

ultra-audion oscillator—*See* Colpitts Oscillator.

Ultrafax—A trade name of RCA for a system in which printed information is transmitted by radio, facsimile, and television at high speeds.

ultrahigh frequency—Frequency band; 300 to 3,000 MHz. Wavelength: 10 to 100 centimeters. Abbreviated uhf.

ultrahigh-frequency converter—A circuit used to convert uhf television signals to vhf, to permit uhf television reception on a vhf receiver.

ultrahigh-frequency generator—Any device for generating ultrahigh-frequency alternating currents (e.g., a conventional negative-grid generator; a positive-grid, or Barkhausen, generator; a magnetron; and a velocity-modulation, or electron-beam generator such as the klystron).

ultrahigh-frequency loop—Generally a single-loop antenna used in ultrahigh-frequency work to secure a nondirectional radiation pattern in the plane of the loop. The doughnut-shaped pattern is perpendicular to the loop.

ultrahigh-frequency translator — A television-broadcast translator station that transmits on a uhf television-broadcast channel.

ultralinear amplifier—A high-power output, low-distortion amplifier used in high-fidelity systems. It requires a specially wound output transformer.

ultramicrometer—An instrument for measuring very small displacements by electrical means (e.g., by the variation in capacitance produced by the movement being measured).

ultramicroscope — A dark-field microscope with which very small particles are illuminated by refracted light so that diffraction rings may be observed.

ultramicrowave—Having wavelengths of about 10^{-1} to 10^{-4} cm.

ultrashort waves—Radio waves shorter than 10 meters in wavelength (about 30 MHz in frequency). Waves shorter than 1 meter are called microwaves.

ultrasonic—Having a frequency above that of audible sound—i.e., between sonic and hypersonic.

ultrasonic brazing—*See* Ultrasonic Soldering.

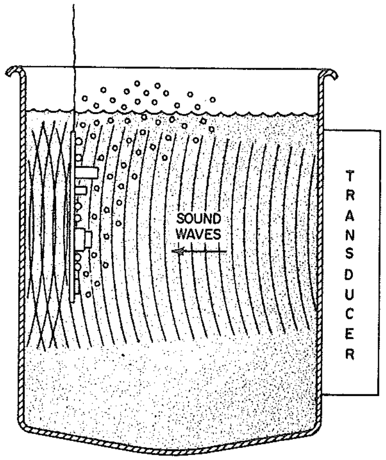

Ultrasonic cleaning tank.

ultrasonic cleaning tank — A heavy-gage, polished stainless-steel tank with transducers mounted on the bottom or sides.

ultrasonic coagulation—The bonding together of small particles by the action of ultrasonic waves.

ultrasonic cross grating—Also called grating. The two- or three-dimensional space grating produced when ultrasonic beams having different directions of propagation intersect.

ultrasonic delay line—Also called an ultrasonic storage cell. A contained medium (usually a liquid such as mercury) in which the signal is delayed because of the longer propagation time of the sound waves in the medium.

ultrasonic densitometer—An instrument device for determining the thickness or density of an object or material, based on the time required for an ultrasonic signal to penetrate to a receiver and/or "echo" back to a receiver adjacent to another transmitter.

ultrasonic detector—A device—either mechanical, electrical, thermal, or optical—for detecting and measuring ultrasonic waves.

ultrasonic diagnosis—A method of obtaining information from within the body in a visual presentation without employing ionizing radiation. It differs from X-ray techniques in that the form of energy used is high-frequency sound. Also, in contrast to X-ray techniques, in which the film is placed behind the tissue being examined, ultrasonic information is picked up at the original point of transmission in the form of echoes from internal structures.

ultrasonic disintegrator—An apparatus for using the pressure wave produced by an ultrasonic generator to tear cells apart.

ultrasonic drill—A special type of drill that has a magnetostrictive transducer attached to a tapered cone that serves as a velocity transformer. With an appropriate tool, practically any shape of hole can be drilled in brittle and hard material.

ultrasonic flaw detector—Equipment comprising an ultrasonic generator, transducer, detector, and display and used to detect flaws or cracks in solids from the reflection pattern of ultrasonic signals observed on a cathode-ray tube.

ultrasonic frequency—Also called an ultra-audible frequency. Any frequency above the audio range, but commonly applied to elastic waves propagated in gases, liquids, or solids.

ultrasonic generator—An electrical device that converts electrical power to an appropriate frequency for use with an electromechanical transducer.

ultrasonic grating constant—The distance between diffracting centers of the sound wave producing particular light-diffraction spectra.

ultrasonic inspection — A nondestructive testing method of locating internal defects in a part by sending ultrasonic impulses (inaudible high-frequency sound waves of 0.5 to 11 megahertz) into the part and measuring the time required for these impulses to penetrate the material, be reflected from the opposite side or from the defect, and return to the sending point.

ultrasonic level detector—A level detector consisting of an ultrasonic receiver and transmitter located in one wall of a container or vessel. With nothing to obstruct the beam, it is reflected from the opposite wall. When the level of the liquid or other material in the container reaches the beam, the liquid or material acts as a reflector, thus reducing the reflection time and indicating that a given level has been reached.

ultrasonic light diffraction—The formation of optical diffraction spectra when a beam of light is passed through a longitudinal sound-wave field. The diffraction results from the periodic variation of the light refraction in the sound field.

ultrasonic light modulator—A device containing a fluid which, by action of ultrasonic waves passing through the fluid, modulates a beam of light passed transversely through the fluid.

ultrasonic material dispersion—The production of suspensions or emulsions of one material in another by the action of high-intensity ultrasonic waves.

ultrasonic plating—The chemical or electrochemical deposition and bonding of one or more solid materials to the surface of another material by the use of vibrational wave energy.

ultrasonic probe—A rod for directing ultrasonic force, used in a disintegration or foreign-body location application.

ultrasonics—The general subject of sound in the frequency range above 15 kilohertz.

ultrasonic soldering—A method of forming a nonporous, continuously metallic connection between metal or alloy parts without necessarily employing chemicals or mechanical abrasives. Instead, vibrational wave energy, heat, and a separate alloy or metal having a melting point below 800° F. and also below that of the metals or alloys being joined is used.

ultrasonic space grating—Also called grating. A periodic spatial variation in the index of refraction caused by the presence of acoustic waves within the medium.

ultrasonic storage cell—*See* Ultrasonic Delay Line.

ultrasonic stroboscope—A light interrupter in which the light beam is modulated by an ultrasonic field.

ultrasonic therapy—The use of ultrasonic vibrations for therapeutic purposes.

ultrasonic thickness gage—A thickness gage in which the propagation time of an ultrasonic beam through a sheet of material is translated into a measure of the thickness of the material.

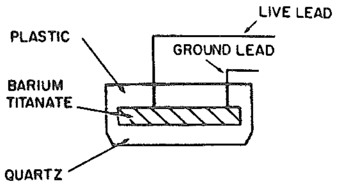

Ultrasonic transducer.

ultrasonic transducer—A device which takes the electrical oscillations produced by the ultrasonic generator and transforms them into mechanical oscillations. Typical transducer materials are piezoelectric (e.g., quartz or barium titanate) or magnetostrictive (e.g., nickel).

ultrasonic waves—Waves having a frequency in the ultrasonic range.

ultrasonic welding — A solid-state bonding process in which metals being joined never reach melting temperatures. They are joined by clamping them together and applying the ultrasonic energy in the plane of the weld without the application of heat.

ultraviolet—Electromagnetic radiation at frequencies higher than those of visible light and with wavelengths of about 200 to 4,000 angstrom units.

ultraviolet lamp—A lamp providing a high proportion of ultraviolet radiation (e.g., arc lamps, mercury-vapor lamps, or incandescent lamps in bulbs of a special glass that is transparent to ultraviolet rays).

ultraviolet rays—Radiation in the ultraviolet region.

umbilical cord—A quickly detachable cable through which missiles or rockets are powered and controlled until the moment of launching.

umbrella antenna—An antenna in which the wires are guyed downward in all directions from a central pole or tower to the ground, somewhat like the ribs of an open umbrella.

unamplified back bias—A degenerative voltage developed across a fast time-constant circuit within an amplifier stage itself.

unbalanced—1. Lacking the conditions for balance. 2. Frequently, a circuit having one side grounded. 3. Differential mutual impedance or mutual admittance between two circuits which ideally would have no coupling.

unbalanced circuit—A circuit, the two sides of which are electrically unlike.

unbalanced line—A transmission line in which the voltages on the two conductors are not equal with respect to ground (e.g., a coaxial line).

unbalanced output—An output where one of the two output terminals is substantially at ground potential.

unbalanced wire circuit—A circuit, the two sides of which are electrically unlike.

unblanking—The turning on of the crt beam.

unblanking generator—A circuit for producing pulses that turn on the beam of a cathode-ray tube.

unblanking pulse—A pulse that turns on the beam of a cathode-ray tube.

unbonded strain gage — A pressure-sensing element made up of resistance strain-gage wire elements arranged in a Wheatstone bridge. It can have two or four active arms, which respond to a pressure applied to the transducer. The unbonded strain-gage wires are suspended in air and are activated by a mechanism attached to a diaphragm or other pressure-responding element.

uncharged — Having a normal number of electrons and hence no electrical charge.

unconditional—In a computer, not subject to conditions external to the specific instruction.

unconditional jump—A computer instruction which interrupts the normal process of obtaining the instructions in an ordered sequence and specifies the address from which the next instruction must be taken.

unconditional transfer of control—In a digital computer which obtains its instructions serially from an ordered sequence of addresses, an instruction which causes the following instruction to be taken from an address that becomes the first of a new sequence.

undamped natural frequency — The frequency at which a system with a single degree of freedom will oscillate, in the absence of damping, upon momentary displacement from the rest position by a transient force.

undamped oscillations — Oscillations that have a constant amplitude for their duration.

undamped wave—A constant-amplitude wave.

underbunching—The condition whereby the buncher voltage of a velocity-modulation tube is lower than the value required for optimum bunching of the electrons.

undercompounded—A generator in which the output voltage drops as the load is increased.

undercurrent relay—A relay that functions when its coil current falls below a predetermined value.

undercut—In a printed-circuit board, the reduction of the cross section of a metal-foil conductor due to the removal of metal from beneath the edge of the resist by the etchant.

undercutting—A cutting with too shallow a groove or with insufficient lateral movement of the stylus during sound disc recordings.

underdamped—A degree of damping that is not sufficient to prevent oscillation in the output of a system following application of an abrupt stimulus.

underdamping—1. In a system, the condition whereby the amount of damping is so small that the system executes one or more oscillations when subjected to a single disturbance (either constant or instantaneous). 2. Oscillation of the transducer output about a final steady value in response to a step change in the measurand. After an initial overshoot, the oscillation amplitude decreases.

underflow—In a computer, the generation of a quantity smaller than the accepted minimum (e.g., floating-point underflow).

underground cable—A cable installed below the surface of the earth.

under insulation—The insulation under wire that is brought from the center of a coil over the top or bottom wall.

underlap—Recorded elemental areas that are smaller than normal—specifically, the space between the recorded elemental area in one recording line of a facsimile system and the adjacent elemental area in the next recording line, or the elemental areas in the direction of the recording line.

underload relay—A relay that operates when the load in a circuit drops below a certain value.

undermodulation—Insufficient modulation of a transmitter, due to maladjustment or to insufficient modulation signal.

underpower relay—A relay which functions when the power decreases below a predetermined value.

undershoot—The initial transient response to a unidirectional change in input which precedes the main transition and is opposite in sense. (*Also see* Precursor.)

underthrow distortion—Distortion resulting when the maximum amplitude of the signal wavefront is less than the steady-state amplitude which would be attained by a prolonged signal wave.

undervoltage protection—Also called low-voltage protection. The effect of a device to cause and maintain the interruption of power to the main circuit upon the reduction or failure of voltage.

undervoltage relay—A relay which operates when its coil voltage falls below a predetermined value.

underwater sound projector—An electroacoustic transducer designed to convert electric waves into sound waves, which are radiated in water for reception at a distance.

undistorted wave—A periodic wave in which both the attenuation and the velocity of propagation are the same for all sinusoidal components, and in which the same sinusoidal component is present at all points.

undisturbed-one output — A 1 output of a magnetic cell to which no partial-read pulses have been applied since that cell was last selected for writing.

undisturbed-zero output — A 0 output of a magnetic cell to which no partial-write pulses have been applied since that cell was last selected for reading.

unfired tube—The condition of TR, ATR, and pre-TR tubes when there is no radio-frequency glow discharge at either the resonant gap or the resonant window.

unfurlable antenna—A device which can be unfolded to form a larger antenna.

ungrounded—Not intentionally connected to ground except through high-impedance devices.

ungrounded system—A system in which no point is directly connected to earth except through potential or ground-detecting transformers or other very-high-impedance devices.

uniaxial magnetic anistropy—A property of magnetic thin film in which the direction of magnetization is always parallel to the easy axis unless an external force acts upon it.

uniconductor waveguide—A waveguide consisting of a rectangular or cylindrical metallic surface surrounding a uniform dielectric medium.

unidirectional—Flowing in only one direction (e.g., direct current).

unidirectional antenna—An antenna with a single, well-defined direction of maximum gain.

unidirectional coupler—A directional coupler which samples only one direction of transmission.

unidirectional current—A direct current—i.e., one that is always positive or always negative—never alternating.

unidirectional microphone — A microphone that responds predominantly to sound from a single solid angle of one hemisphere or less.

unidirectional pulses—Single-polarity pulses which all rise in the same direction.

unidirectional pulse train—A pulse train in which all pulses rise in the same direction.

unidirectional transducer — Also called a unilateral transducer. A transducer, the output of which cannot be actuated by waves to supply related waves at its input.

uniform field—A field in which the scalar (or vector) has the same value at every point in the region under consideration at that instant.

uniformity—In terms of magnetic tape properties, a figure of merit relating to the ability of the tape to deliver a steady and consistent output level upon being recorded with a constant input. Usually expressed in db variation from average at a midrange frequency.

uniform line—A line with substantially identical electrical properties throughout its length.

uniform plane wave—A plane wave with constant-amplitude electric and magnetic field vectors over the equiphase surfaces. Such a wave can only be found in free space, at an infinite distance from the source.

uniform precession—The condition in which the magnetic moments of all atoms in a sample are parallel and precess in phase about the magnetic field. Uniform precession occurs in regions where the magnetic field is uniform. A spin wave is a phase distortion of this condition.

uniform waveguide—A waveguide in which the physical and electrical characteristics do not change with distance along its axis.

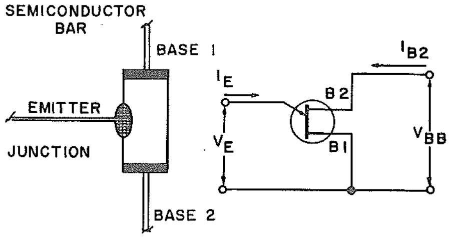

Unijunction transistor.

unijunction transistor—Also called a silicon double-base diode or silicon unijunction transistor. A three-terminal semiconductor device exhibiting stable open-circuit, negative-resistance characteristics. The internal construction consists of a uniformly doped N-type, single-crystal semiconductor with ohmic contacts at each end and the emitter wire attached between them.

unilateral area track—A sound track in which only one edge of the opaque area is modulated in accordance with the recorded signal. However, there may be a second edge modulated by a noise-reduction device.

unilateral bearing—A bearing obtained with a radio direction finder having a unilateral response, eliminating the chance of a 180° error.

unilateral conductivity—Conductivity in only one direction (e.g., in a perfect rectifier).

unilateral element—A two-terminal element with a zero voltage-to-current characteristic (or the equivalent) on one side of the origin.

unilateralization—A special case of neutralization in which the feedback parameters are completely balanced out. In transistors, these feedback parameters include a resistive in addition to a capacitive component. Unilateralization changes a network from bilateral to unilateral.

unilateral network—A network in which any driving force applied at one pair of terminals produces a response at a second pair, but yields no response when the driving force is applied in the other direction.

unilateral transducer — *See* Unidirectional Transducer.

unipolar field-effect transistor — A field-controlled majority-carrier device wherein the conductance of a semiconductor channel is modulated by a transverse electric field. The field is controlled by the combination of gate bias voltage V_{gs} and the net voltage V_{ds} between channel drain and source.

unipolar transistor—A transistor in which charge carriers are of only one polarity. *Also see* Field-Effect Transistor.

unipole—1. An all-pass filter section with one pole and one zero. 2. A hypothetical antenna which radiates and receives equally in all directions. *Also see* Isotropic Antenna.

unipotential cathode—*See* Indirectly-Heated Cathode.

unit—1. A computer portion or subassembly which constitutes the means of accomplishing some inclusive operation or function (e.g., an arithmetic unit). 2. The specific magnitude of a quantity set apart by appropriate definition and serving as a basis for the comparison or measurement of like quantities.

unit-area acoustic impedance—*See* Specific Acoustic Impedance.

unitary code—A code having only one digit, the number of times it is repeated determining the quantity it represents.

unit charge—The electrical charge which will repel a force of one dyne on an equal and like charge one centimeter away in a vacuum, assuming each charge is concentrated at a point.

unit interval—*See* Signal Element.

unit length—The basic element of time for determining code speeds in message transmission.

unit magnetic pole—A pole with a strength such that when it is placed 1 cm away from a like pole, the force between the two is 1 dyne.

unitor — In computers, a device or circuit which performs a function corresponding to the Boolean operation of union. *Also see* OR Circuit.

unit pulse—*See* Baud.

unit record equipment — Equipment using punched cards as input data, such as collators, tabulating machines, etc.

unit sequence starting relay — A device which functions to start the next available unit in a multiple-unit equipment on the failure or on the nonavailability of the normally preceding unit.

unit sequence switch — A switch used to change the sequence in which units may be placed in and out of service in multiple-unit equipment.

unit step current (or voltage)—A current (or voltage) which undergoes an instantaneous change in magnitude from one constant level to another.

unit substation transformer—A transformer which is mechanically and electrically connected to and coordinated in design with one or more switch-gear or motor-controlled assemblies or combinations thereof.

unit troque gradient—The torque gradient of a synchro, measured when the synchro is electrically connected to another synchro of the same size.

unitunnel diode—A diode similar to a tunnel diode, but specially treated to give peak reverse currents in the microampere region while providing high forward conductance at low voltage levels.

unity coupling—Perfect magnetic coupling between two coils, so that all the magnetic flux produced by the primary winding passes through the entire secondary winding.

unity-gain crossover frequency—The frequency at which the open-loop voltage gain of an amplifier curve crosses through unity gain, or zero db.

unity power factor—A power factor of 1.0. It is obtained only when current and voltage are in phase (e.g., in a circuit containing only resistance, or in a reactive circuit at resonance).

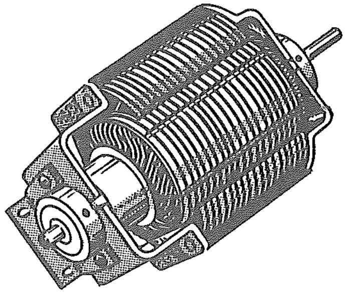

Universal motor.

universal motor—A series-wound motor designed to operate at approximately the same speed and output on direct current or on a single-phase alternating current of not more than 60 hertz and approximately the same rms voltage.

universal output transformer—An output transformer having a number of taps on its winding. By proper choice of connections, it can be used between the audio-frequency output stage and the speaker of practically any radio receiver or audio amplifier.

universal receiver—Also called an a-c/d-c receiver. A receiver with no power transformer and thus capable of operating from either a-c or d-c power lines, without changes in its internal connections.

universal shunt—*See* Ayrton Shunt.

universe — *See* Population.

unload—In a computer. 1. To remove the tape from the columns of a recorder by raising or lowering the recording head. 2. To remove a portion of the address part of an instruction. 3. *Also see* Dump.

unloaded antenna — An antenna with no added inductance or capacitance.

unloaded applicator impedance (dielectric heaters)—The complex impedance measured at the point of application and at a specified frequency without the load material in position.

unloaded *Q* (switching tubes)—Also called the intrinsic *Q*. The *Q* of a tube unloaded by either the generator or termination.

unloading amplifier—An amplifier capable of reproducing or amplifying a given voltage signal while drawing negligible current from the voltage source.

unloading circuit—In an analog computer, a computing element or combination of computing elements capable of reproducing or amplifying a given voltage signal while drawing negligible current from the voltage source, thus decreasing the load errors.

unmodulated—Having no modulation—e.g. a carrier that is transmitted during moments of silence in radio programs, or a silent groove in a disc recording.

unmodulated groove—Also called a blank groove. In mechanical recording, the groove made in the medium with no signal applied to the cutter.

unoriented—A structure in which the crystallographic axes of the grains of a metal are not aligned to give directional magnetic properties.

unpack—In a computer, to separate combined items of information, each into a separate machine word.

untuned—Not resonant at any of the frequencies being handled.

unwind—In a computer, to code all the operations of a cycle, at length and in full, for the express purpose of eliminating all red-tape operations.

up-converter—A type of parametric amplifier which is characterized by the frequency of the output signal being greater than the frequency of the input signal.

update—To search the file (such as a particular record in a computer tape) and select one entry, then perform some operation to bring the entry up-to-date.

updating—The act of bringing information up to the current value.

upper operating temperature — The maximum temperature to which a material can be subjected and still maintain specified operating characteristics within limits.

upper sideband—The higher frequency or group of frequencies produced by an amplitude-modulation process.

upset-duplex system—A direct-current telegraph system in which a station between any two pieces of duplex equipment may transmit signals by opening and closing the line circuit and thereby upsetting the duplex balance.

upset welding—A resistance-welding process wherein the weld is made simultaneously over the entire area of abuting surfaces or progressively along the joint with the aid of rolls or clamps which force the abutting surfaces together. The pressure is applied before heating starts and is maintained throughout the heating period.

up time — The time during which an equipment is either operating or available for operation as opposed to down time when no productive work can be accomplished.

urea plastic material—A thermosetting plastic material, with good dielectric, used for radio-receiver cabinets, instrument housings, etc.

USASI—Abbreviation for United States of America Standards Institute, the successor to ASA (American Standards Association).

"U" scan—A parallel reading method to prevent ambiguity in the readout of polystrophic codes at the code-position transitions by reading one of two sets of brushes, depending on the state of a control or selector bit.

useful life—The total time a device operates between debugging and wearout.

utilization factor—In electrical power distribution, the ratio of the maximum demand of a system (or part of a system) to the rated capacity of the system (or part) under consideration.

utilogic—A line of digital IC's built around a basic AND and a basic NOR circuit. The AND has multiple-emitter inputs; the NOR has emitter-follower inputs. The output for the AND is an emitter follower, and for the NOR, a totem-pole arrangement. J-K binary element is also included in the line.

V

V—1. Symbol for volt or voltmeter. 2. Schematic symbol for vacuum tube.

VA—Abbreviation for volt-ampere.

vac—Abbreviation for vacuum.

vacuum—Abbreviated vac. Theoretically, an enclosed space from which all air and gases have been removed. However, since such a perfect vacuum is never attained, the term is taken to mean a condition whereby sufficient air has been removed so that any remaining gas will not affect the characteristics beyond an allowable amount.

vacuum capacitor—A capacitor consisting usually of two concentric cylinders enclosed in a vacuum to raise the breakdown voltage.

vacuum deposition — A process in which a substance is deposited on a surface by heating the substance in a vacuum until vaporization occurs.

vacuum envelope — The airtight envelope which contains the electrodes of an electron tube.

vacuum gage—A device that indicates the absolute gas pressure in a vacuum system (e.g., in the evacuated parts of a mercury-arc rectifier).

vacuum impregnation—Filling the spaces between electric parts or turns of a coil with an insulating compound while the coil or parts are in a vacuum.

vacuum phototube—A phototube which is evacuated to such a degree that its electrical characteristics are essentially unaffected by gaseous ionization.

vacuum range — For a communications system, the maximum range computed for an atmospheric attenuation of zero.

vacuum seal—An airtight junction between component parts of an evacuated system.

vacuum switch—A switch in which the contacts are enclosed in an evacuated bulb, usually to minimize sparking.

vacuum tank—An airtight metal chamber which contains the electrodes and in which the rectifying action takes place in a mercury-arc rectifier.

vacuum tube—An electron tube evacuated to such a degree that its electrical characteristics are essentially unaffected by the presence of residual gas or vapor.

vacuum-tube amplifier — An amplifier in which electron tubes are used to control the power from the local source.

vacuum-tube characteristics — Data that show how a vacuum tube will operate under various electrical conditions.

vacuum-tube keying — A code-transmitter keying system in which a vacuum tube is connected in series with the plate-supply lead going to the winding in the plate circuit of the final stage. The grid of the tube is connected to its filament through the transmitting key so that when the key is open, the tube is blocked, interrupting the plate supply to the output stage. Closing the key allows plate current once more to flow through the keying tube and the output tubes.

vacuum-tube modulator—A modulator in which a vacuum tube is the modulating element.

vacuum-tube oscillator—A circuit in which a vacuum tube is used to convert d-c power into a-c power at the desired frequency.

vacuum-tube rectifier — A tube which changes an alternating current to a unidirectional pulsating direct current.

vacuum-tube transmitter—A radio transmitter in which electron tubes are utilized to convert the applied electric power into radio-frequency power.

vacuum-tube voltmeter—Abbreviated vtvm. *See* Electronic Voltmeter.

valence—A number representing the proportion in which the atom is able to combine with other atoms. It generally depends on the number and arrangement of electrons in the outermost shell of each type of atom.

valence band—In the spectrum of a solid crystal, the range of energy states containing the energies of the valence electrons which bind the crystal together. In a semiconductive material, it is just below the conduction band.

valence bond—Also called a bond. The bond formed between the electrons of two or more atoms.

valence electrons—The electrons of an atom in the outer shell that determine the chemical valency of the atom.

valence shell—The electrons which form the outermost shell of an atom.

validity—Correctness—specifically, how closely repeated approximations approach the desired (i.e., correct) result.

valley—A dip between two peaks in a curve.

valley current—In a tunnel diode, the current measured at the positive voltage for which the current has a minimum value from which it will increase if the voltage is further increased.

valley voltage—In a tunnel diode, the voltage corresponding to the valley current.

value—The magnitude of a physical quantity.

valve—1. A British term for a vacuum tube. 2. A device permitting current flow in one direction only (e.g., a rectifier). 3. A device or system which is capable of flow diversion, cutoff, or modulation.

valve tube—*See* Kenotron.

Van Allen radiation belts—Two doughnut-shaped belts of high-energy particles which surround the earth and are trapped in its magnetic field. They were first discovered by Dr. James A. Van Allen of Iowa State University.

Van de Graaff accelerator—An electrostatic-generator type of particle accelerator from which the voltage is obtained by picking up static electricity at one end of the machine (on a rubber belt) and carrying it to the other end, where it is stored.

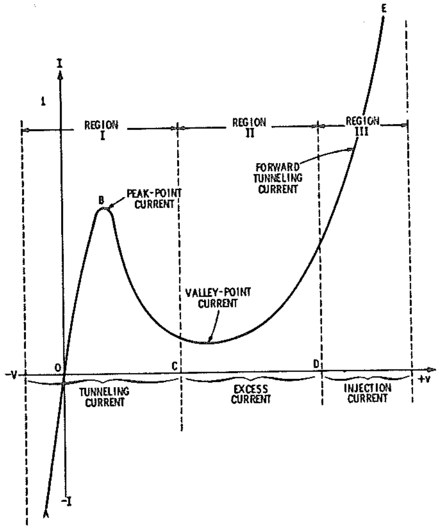

Valley current.

vane-anode magnetron—A cavity magnetron in which the walls between adjacent cavities have parallel plane surfaces.

vane attenuator — A waveguide device designed to present attenuation in a circuit by sliding a resistive element from the side wall of the waveguide to the center for maximum attenuation. This method of attenuation is used in precision calibrated attenuation readings, and resetting must be made. Some of its countless applications are calibrations of other attenuators, directional couplers, filters, and other lossy components, in antenna-pattern measurements, noise-level measurements, and for setting power levels to desired values.

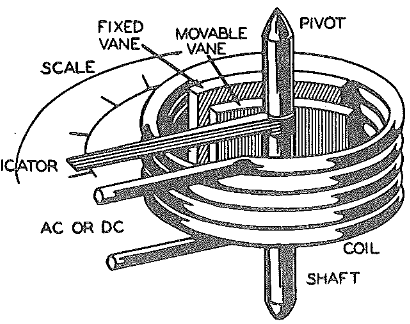

Vane-type instrument.

vane-type instrument—A measuring instrument in which the pointer is moved by the force of repulsion between fixed and movable magnetized iron vanes, or by the force between a coil and a pivoted vane-shaped piece of soft iron.

vane-type magnetron—A cavity magnetron in which the walls between adjacent cavities have plane surfaces.

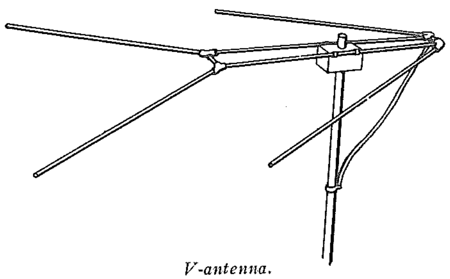

V-antenna.

V antenna—A V-shaped arrangement of conductors, the two branches being fed equally in opposite phase at the apex.

vapor pressure—The pressure of the vapor accumulated above a confined liquid (e.g., in a mercury-vapor rectifier tube).

var—1. Abbreviation for volt ampere reactive. The unit of reactive power, as opposed to real power in watts. One var is equal to one reactive volt-ampere. 2. Abbreviation for visual aural range.

varactor — 1. A PN junction semiconductor diode designed for low losses at high frequencies. Its capacitance varies with the applied voltage. 2. Also called varactor diode, silicon capacitor, voltage-controlled capacitor, and voltage-variable capacitor. A two-terminal solid-state device that utilizes the voltage-variable capacitance of a PN junction. In the normal semiconductor diode, efforts are made to minimize inherent capacitance, while in the varactor, this capacitance is emphasized. Since the capacitance varies with the applied voltage, it is possible to amplify, multiply, and switch with this device.

varhour meter—Also called a reactive volt-ampere-hour meter. An electricity meter which measures and registers the integral (usually in kilovarhours) of the reactive power of the circuit into which the meter is connected.

variable—Any factor or condition which can be measured, altered, or controlled (e.g., temperature, pressure, flow, liquid level, humidity, weight, chemical composition, color, etc.).

variable-area track—A sound track divided laterally into opaque and transparent areas. A sharp line of demarcation between these areas forms an oscillographic trace of the waveshape of the recorded signal.

variable-capacitance diode — Abbreviated vcd. A semiconductor diode in which the junction capacitance present in all semiconductor diodes has been accentuated. An appreciable change in the thickness of the junction-depletion layer and a corresponding change in the capacitance occur when the d-c voltage applied to the diode is changed.

variable-capacitance transducer—A transducer which measures a parameter or a change in a parameter by means of a change in capacitance.

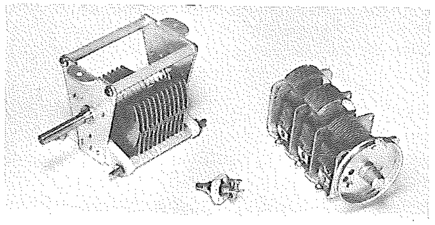

Variable capacitor.

variable capacitor—A capacitor which can be changed in capacitance by varying the useful area of its plates, as in a rotary capacitor, or by altering the distance between them, as in some trimmer capacitors.

variable-carrier modulation — *See* Controlled-Carrier Modulation.

variable connector — 1. A flowchart symbol representing a sequence connection which is not fixed, but which can be varied by the flowcharted procedure itself. 2. The device which inserts instructions in a program corresponding to the selection of paths appearing in a flowchart. 3. The computer instructions which cause a logical chain to take one of several alternative paths.

variable coupling—Inductive coupling that can be varied by moving the windings.

variable-cycle operation—Computer operation in which any cycle is started at the completion of the previous cycle, instead of at specified clock times.

variable-density track—A sound track of constant width and usually, but not necessarily, of uniform light transmission on any instantaneous transverse axis. The average light transmission varies along the longitudinal axis in proportion to some characteristic of the applied signal.

variable-depth sonar—A sensor that can be lowered by cable from a ship or helicopter, through thermal layers, to detect submarines operating at deeper levels.

variable-erase recording—The method of recording on magnetic tape by selective erasure of a prerecorded signal.

variable field—A field in which the scalar (or vector) at any point changes during the time under consideration.

variable-frequency oscillator — Abbreviated vfo. A stable oscillator, the frequency of which can be adjusted over a given range.

variable inductance—A coil the inductance of which can be varied.

variable-inductance pickup—A phonograph pickup in which the movement of a stylus causes the inductance to vary accordingly.

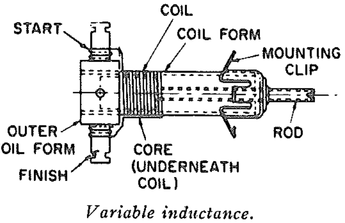

Variable inductance.

variable-inductance transducer — A transducer in which the output voltage is a function of the change in a variable-inductance element.

variable monoergic—A type of emission in which the magnitude of the homogeneous particle or radiation energy is continuously variable over broad limits; e.g., the proton energy from some types of accelerators can be controlled by varying the high voltage.

variable-mu pressure transducer—A device which converts mechanical input pressure to a proportional electrical output based on the change of the mu of its magnetic circuit due to the applied pressure (Villari effect).

variable-mu tube—*See* Remote-Cutoff Tube.

variable radio-frequency radiosonde — A radiosonde the carrier frequency of which is modulated by the magnitude of the meteorological variables being sensed.

variable-reluctance microphone—Also called a magnetic microphone. A microphone which depends for its operation on the variations in reluctance of a magnetic circuit.

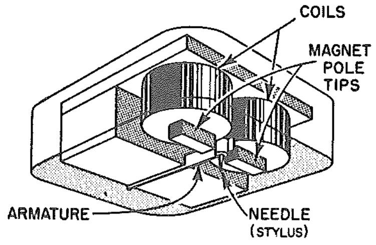

Variable-reluctance pickup.

variable-reluctance pickup — Also called a magnetic pickup or cartridge. A phonograph pickup which depends for its operation on the variations in reluctance of a magnetic circuit.

variable-reluctance transducer—Also called a magnetic transducer. A transducer that depends for its operation on the variations in reluctance of a magnetic circuit.

variable-resistance pickup—A phonograph pickup which depends for its operation on the variation of a resistance.

variable-resistance transducer — A transducer in which the signal output depends on the change in a resistance element.

variable resistor—A wirewound or composition resistor, the resistance of which may be changed. (*Also see* Rheostat *and* Potentiometer.)

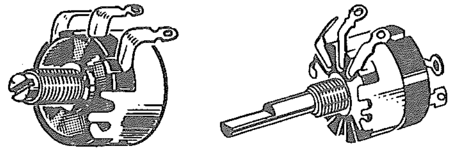

Variable resistors.

variable-speed motor—A motor the speed of which can be adjusted within certain limitations, regardless of load.

variable-speed scanning—A scanning method whereby the optical density of the film being scanned determines the speed at which the scanning beam in the cathode-ray tube of a television camera is deflected.

variable transformer—An iron-core transformer with provision for varying its output voltage over a limited range, or continuously from zero to maximum—generally by the movement of a contact arm along exposed turns of the secondary winding.

Variac—Trade name of General Radio Co. for a variable transformer.

variate — *See* Random Variable.

variations—The angular difference between a true and a magnetic bearing or heading.

varicap—*See* Varactor.

varindor—An inductor, the inductance of which varies markedly with the current in the winding.

variocoupler — A radio transformer with windings that have an essentially constant self-impedance, but the mutual impedance between them is adjustable.

variolosser—A device whose loss can be controlled by a voltage or current.

variometer—A variable inductor in which the inductance is changed by the relative position of two or more coils.

varistor—A two-electrode semiconductor device with a voltage-dependent nonlinear resistance that drops markedly as the applied voltage is increased.

Varley loop—A type of Wheatstone-bridge circuit which gives, in one measurement, the difference in resistance between two wires of a loop.

varmeter—Also called a reactive volt-ampere meter. An instrument for measuring reactive power in either vars, kilovars, or megavars. If the scale is graduated in kilovars or megavars, the instrument is sometimes designated a kilovarmeter or megavarmeter.

varnished cambric—A linen or cotton fabric that has been impregnated with varnish or insulating oil and baked. It is used as insulation in coils and other radio parts.

varying duty—A requirement of service that demands operation at loads and for intervals of time, both of which may be subject to wide variation.

varying-speed motor—A motor which slows down as the load increases (e.g., a series motor, or an induction motor with a large amount of slip).

varying-voltage control—A form of armature-voltage control obtained by impressing, on the armature, a voltage which varies considerably with a change in load and consequently changes the speed of the motor (e.g., by using a differentially compound-wound generator or a resistance in the armature circuit).

V-beam system—A radar system for measuring elevation. The antenna emits two fan-shaped beams, one vertical and the other inclined, which intersect at ground level. Each beam rotates continuously about a vertical axis, and the time elapsing between the two echoes from the target provides a measure of its elevation.

vcd — Abbreviation for variable-capacitance diode.

V-cut—A type of crystal-oscillator cut in which the major plane surfaces are not parallel to the X, Y, or Z planes.

vector—The term for a symbol which denotes a directed quantity—i.e., one which cannot be completely described except in terms of both magnitude and direction (e.g., wind velocities, voltages and currents of electricity, and forces of all kinds).

vector admittance—The ratio for a single sinusoidal current and potential difference in a portion of a circuit of the corresponding complex harmonic current to the corresponding complex potential difference.

vector-ampere—The unit of measurement of vector power.

vector cardiograph — An instrument that measures both the magnitude and the direction of heart signals by displaying cardiograph signals on any desired set of axes, usually X, Y, and Z.

vector diagram—An arrangement of vectors showing the relationships between alternating quantities having the same frequency.

vector field—In a given region of space, the total value of some vector quantity which has a definite value at each point of the region (e.g., the distribution of magnetic intensity in a region surrounding a current-carrying conductor.

vector function—A function which has both magnitude and direction (e.g., the magnetic intensity at a point near an electric circuit is a vector function of the current in that circuit).

vector impedance—The ratio for a simple sinusoidal current and potential difference in a portion of a circuit of the corresponding complex harmonic potential difference to the corresponding complex current.

vector power—A vector quantity equal to the square root of the sum of the squares of the active and reactive powers. The unit is the vector-ampere.

vector power factor—Ratio of the active power to the vector power. In sinusoidal quantities, it is the same as power factor.

vector quantity—A quantity with both magnitude and direction.

velocimeter—A c-w reflection Doppler system used to measure the radial velocity of an object.

velocity—1. A vector quantity that includes both magnitude (speed) and direction in relation to a given frame of reference. 2. Rate of motion in a given direction, employed in its higher magnitudes as a means of overcoming the force of gravity.

velocity error—The amount of angular displacement existing between the input and output shafts of a servomechanism when both are turning at the same speed.

velocity filter—A storage-tube device which blanks all targets that do not move more than one resolution cell in less than a predetermined number of antenna scans.

velocity hydrophone—A type of hydrophone in which the electric output is substantially proportional to the instantaneous particle velocity in the incident sound wave.

velocity-lag error—A lag, between the input and output of a device, that is proportional to the rate of variation of the input.

velocity level—In decibels of a sound, 20 times the logarithm to the base 10 of the ratio of the particle velocity of the sound to the reference particle velocity. The latter must be stated explicitly.

velocity microphone—A microphone in which the electric output corresponds substantially to the instantaneous particle velocity in the impressed sound wave. It is a gradient microphone of order one, and is inherently bidirectional.

velocity-modulated amplifier—Also called a velocity-variation amplifier. An amplifier in which velocity modulation is employed for amplifying radio frequencies.

velocity-modulated oscillator—Also called a velocity-variation oscillator. An electron-tube structure in which the velocity of an electron stream is varied (velocity-modulated) in passing through a resonant cavity called a buncher. Energy is extracted from the bunched electron stream at a higher level in passing through a second cavity resonator called the catcher. Oscillations are sustained by coupling energy from the catcher cavity back to the buncher cavity.

velocity modulation—Also called velocity variation. Modification of the velocity of an electron stream by the alternate acceleration and deceleration of the electrons with a period comparable to that of the transit time in the space concerned.

velocity of light—A physical constant equal to 2.99796 × 10^{10} centimeters per second. (More conveniently expressed as 186,280 statute miles per second, 161,750 nautical miles per second, or 328 yards per microsecond.)

velocity of propagation—1. The speed at which a disturbance (sound, radio, light, etc., waves) is radiated through a medium. 2. The ratio of the speed of the flow of an electric current in an insulated cable to the speed of light, expressed in percentage. In the case of coaxial cables, this ratio is 65-66%, where the insulation is polyethylene.

velocity resonance—*See* Phase Resonance.

velocity sorting—The selecting of electrons according to their velocity.

velocity spectrograph—An apparatus for separating an emission of electrically charged particles into distinct streams, in accordance with their speed, by means of magnetic or electric deflection.

velocity transducer—A transducer which generates an output proportionate to the imparted velocities.

velocity variation—*See* Velocity Modulation.

velocity-variation amplifier—*See* Velocity-Modulated Amplifier.

velocity-variation oscillator—*See* Velocity-Modulated Oscillator.

Venn diagrams—Diagrams in which circles or ellipses are used to give a graphic representation of basic logic relations. Logic relations between classes, operations on classes, and the terms of the propositions are illustrated and defined by the inclusion, exclusion, or intersection of these figures. Shading indicates empty areas, crosses indicate areas that are not empty, and blank spaces indicate areas that may be either. Named for English logician John Venn, who devised them.

vent—A controlled weakness somewhere in the enclosure of an aluminum electrolytic capacitor to permit pressure relief in case of failure due to shorting or improper installation of the capacitor.

vented baffle—An enclosure designed to properly couple a speaker to the air.

venturi tube—A short tube with flaring ends and a constricted throat. It is used for measuring flow velocity by measurement of the throat pressure, which decreases as the velocity increases.

verification—The process of checking the results of one data transcription against those of another, both transcriptions usually involving manual operations. (*Also see* Check.)

verifier—Of computers, a device on which a record can be compared or tested for identity character by character with a retranscription or copy as it is being prepared.

verify—1. To check, usually with an automatic machine, one recording of data against another in order to minimize the number of human errors in the data transcription. 2. To make certain that the information being prepared for a computer is correct.

vernier—1. An auxiliary scale comprising subdivisions of the main measuring scale and thus permitting more accurate measurements than are possible from the main scale alone. 2. An auxiliary device used for obtaining fine adjustments.

vernier capacitor—A variable capacitor placed in parallel with a larger tuning capacitor and used to provide a finer adjustment after the larger one has been set to the approximate desired position.

vernier dial—A type of tuning dial used chiefly for radio equipment. Each complete rotation of its control knob moves the main shaft only a fraction of a revolution and thereby permits fine adjustment.

vernitel—A precision device which makes possible the transmission of data with high accuracy over standard frequency-modulated–frequency-modulated telemetering systems.

Versatile Automatic Test Equipment—A computer-controlled tester for missile electronic systems. It troubleshoots faults by deductive logic and isolates them to the plug-in module or component level.

vertex—*See* Node.

vertex plate—A matching plate placed at the vertex of a reflector.

vertical—In 45/45 recording, the signal produced by a sound arriving at the two microphones simultaneously and 180° out of phase, causing the cutting stylus to move vertically.

vertical amplification—Signal gain in the circuits of an oscilloscope that produce vertical deflection on the screen.

vertical-amplitude controls—*See* Parabola Controls.

vertical antenna—A vertical metal tower, rod, or suspended wire that is used as a receiving and/or transmitting antenna.

vertical blanking—The blanking applied to a cathode-ray tube to eliminate the trace during flyback.

vertical-blanking pulse—In television, a pulse transmitted at the end of each field to cut off the cathode-ray beam while it returns to start the next field.

vertical-centering control—A control provided in a television receiver or cathode-ray oscilloscope to shift the entire image up or down on the screen.

vertical compliance—The ability of a reproducing stylus to move vertically while in the reproducing position on a record.

vertical-deflection electrodes—The pair of electrodes that move the electron beam up and down on the screen of a cathode-ray tube employing electrostatic deflection.

vertical dynamic convergence—Convergence of the three electron beams at the aperture mask of a color picture tube during the scanning of each point along a vertical line at the center of the tube.

vertical field-strength diagram—A representation of the field strength at a constant distance from, and in a vertical plane passing through, an antenna.

vertical-frequency response—In an oscilloscope, the band of frequencies passed, with amplification between specified limits, by the amplifiers that produce vertical deflection on the screen.

vertical-hold control—*See* Hold Control.

vertical-incidence transmission—The transmission of a radio wave vertically to the ionosphere and back. The transmission remains practically the same for a slight departure from the vertical (e.g., when the transmitter and receiver are a few kilometers apart).

vertical-lateral recording—A technique of making stereo phonograph discs by recording one signal laterally, as in monophonic records, and the other vertically, as in hill-and-dale transcriptions.

vertical-linearity control—A control that permits adjustment of the spacing of the horizontal lines on the upper portion of the picture to effect linear vertical reproduction of a televised scene.

vertically polarized wave—1. An electromagnetic wave with a vertical electric vector. 2. A linearly polarized wave with a horizontal magnetic field vector.

vertical polarization—Transmission in which the transmitting and receiving antennas are placed in a vertical plane, so that the electrostatic field also varies in a vertical plane.

vertical quarter-wave stub—An antenna with a vertical portion that is electrically one quarter-wavelength long. It is used generally with a ground plane at the base of the stub.

vertical radiator—A transmitting antenna perpendicular to the earth's surface.

vertical recording—Also called hill-and-dale recording. Mechanical recording in which the groove modulation is perpendicular to the surface of the recording medium.

vertical redundance—In a computer, an error condition which exists when a character fails a parity check, i.e., has an even number of bits in an odd-parity system, or vice versa.

vertical resolution—On a television test pattern, the number of horizontal wedge lines that can be clearly discerned by the eye before they merge together.

vertical retrace—The return of the electron beam from the bottom of the image to the top after each vertical sweep.

vertical speed transducer—An instrument which furnishes an electrical output that is proportionate to the vertical speed of the aircraft or missile in which it is installed.

vertical stylus force—*See* Stylus Force.

vertical sweep—The downward movement of the scanning beam from top to bottom of the televised picture.

vertical-synchronizing pulse—Also called picture-synchronizing pulse. One of the six pulses transmitted at the end of each field in a television system. It maintains the receiver in field-by-field synchronism with the transmitter.

very high frequency—Frequency band: 30 to 3000 MHz. Wavelength: 1 to 10 meters. Metric waves. Abbreviated vhf.

very long range—A classification of ground radar sets by slant range, applied to those with a maximum range exceeding 250 miles.

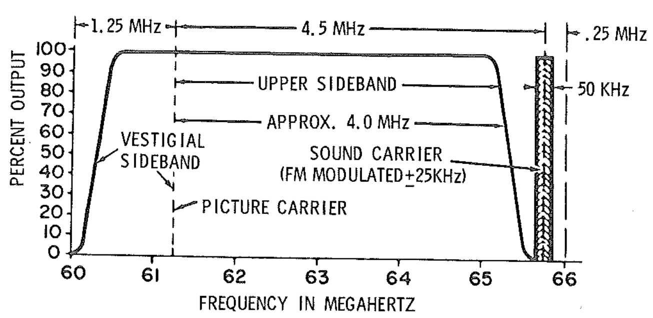

Vestigial-sideband television channel.

very low frequency—Frequency band: below 30 kHz. Wavelength: above 10,000 meters. Abbreviated vlf.

very short range—Classification of ground radar sets by slant range, applied to those with a maximum range of less than 25 miles.

vestigial—Pertaining to a remnant or remaining part.

vestigial sideband — Amplitude-modulated transmission in which a portion of one sideband has been largely suppressed by a transducer having a gradual cutoff in the neighborhood of the carrier frequency.

vestigial-sideband filter—A filter that is inserted between an a-m transmitter and its transmitting antenna to suppress part of one of the sidebands.

vestigial-sideband transmission — Also called asymmetrical sideband transmission. Signal transmission in which one normal sideband and the corresponding vestigial sideband are utilized.

vestigial-sideband transmitter — A transmitter in which one sideband and only a portion of the other are transmitted.

vfo—Abbreviation for variable-frequency oscillator.

vhf—Abbreviation for very high frequency.

vhf omnirange—Abbreviated vor. A specific type of range operating at vhf and providing radial lines of position in a direction determined by the bearing selection within the receiving equipment. A nondirectional reference modulation is emitted, along with a rotation pattern which develops a variable modulation of the same frequency as the reference modulation. Lines of position are determined by comparing the phase of the variable with that of the reference.

vibrating bell—A bell having a mechanism designed to strike repeatedly and as long as it is actuated.

vibrating-reed meter — A frequency meter consisting of a row of steel reeds, each having a different natural frequency. All are excited by an electromagnet fed with the alternating current whose frequency is to be measured. The reed whose frequency corresponds most nearly with that of the current vibrates, and the frequency is read on a scale beside the row of reeds.

vibrating-reed relay — A type of relay in which an alternating or a self-interrupted voltage is applied to the driving coil so as to produce an alternating or pulsating magnetic field that causes a reed to vibrate.

vibrating-wire transducer — A transducer which utilizes a thin wire suspended in a magnetic field; the change in tension of the wire reflects a frequency-modulating output.

vibration — 1. A continuously reversing change in the magnitude of a given force. 2. A mechanical oscillation or motion about a reference point of equilibrium.

vibration analyzer—A device used to analyze mechanical vibrations.

vibration galvanometer—An a-c galvanometer in which a reading is obtained by making the natural oscillation frequency of the moving element equal to the frequency of the current being measured.

vibration isolator—A resilient support that tends to isolate a system from steady-state excitation or vibration.

vibration meter—Also called a vibrometer. An apparatus comprising a vibration pickup, calibrated amplifier, and output meter, for the measurement of displacement, velocity, and acceleration of a vibrating body.

vibration pickup—A microphone that responds to mechanical vibrations rather than to sound waves. In one type a piezoelectric unit is employed; the twisting or bending of a Rochelle-salt crystal generates a voltage that varies with the vibration being analyzed.

vibration sensitivity — The peak instantaneous change in output at a given sinusoidal vibration level for any one stimulus value within the range of an instrument or equipment. It is usually expressed in percentage of full-scale output per vibratory "g" over a given frequency range. It may also be specified as a total error in percentage of full-scale output for a given vibratory acceleration level.

vibration survey—A method of determining the natural frequency of a transducer by observation of the output waveform upon

the application of a shock or tapping of sufficient magnitude to initiate oscillation of the instrument.

vibrato—A musical embellishment which depends primarily on periodic variations of frequency, often accompanied by variations in amplitude and waveform. The quantitative description of vibrato is usually in terms of the corresponding modulation of frequency (5-7 Hz), amplitude, or waveform, or all three.

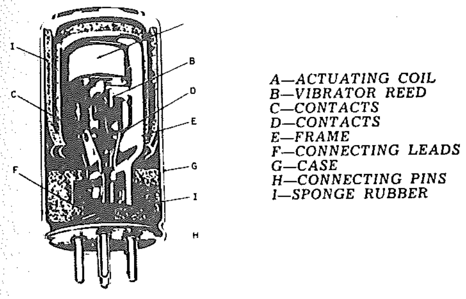

Vibrator.

vibrator—A vibrating reed which is driven like a buzzer and has contacts arranged to interrupt direct current to the winding(s) of a transformer resulting in an alternating current being supplied from another winding to the load.

vibrator power supply—A power supply incorporating a vibrator, step-up transformer, rectifier, and filters for changing a low d-c voltage to a high d-c voltage.

vibratron—A triode with an anode that can be moved or vibrated by an external force. Thus, the anode current will vary in proportion to the amplitude and frequency of the applied force.

vibrograph—An apparatus for recording mechanical vibrations.

vibrometer—*See* Vibration Meter.

video—1. Pertaining to the bandwidth and spectrum position of the signal resulting from radar or television scanning. In current usage, video means a bandwidth on the order of several megahertz. 2. A prefix to the name of television parts or circuits which carry picture signals. 3. Radar or television signals which actuate the cathode-ray tube; frequencies extend from approximately 60 hertz to several megahertz.

video amplifier—An amplifier which provides wideband operation in the frequency range of approximately 15 hertz to 5 megahertz.

video carrier—The television signal whose modulation sidebands contain the picture, sync, and blanking signals.

videocast — 1. To broadcast a program by means of television. 2. A program so broadcast.

video detector — The demodulator circuit which extracts the picture information from the amplitude-modulated intermediate frequency in a television receiver.

video discrimination—A radar circuit that reduces the frequency band of the video-amplifier stage in which it is used.

video frequency — 1. The frequency of the signal voltage containing the picture information which arises from the television scanning process. In the present United States television system, these frequencies are limited from approximately 30 Hz to 4 MHz. 2. A band of frequencies extending from approximately 100 hertz to several megahertz.

video-frequency amplifier—A device capable of amplifying those signals that comprise the periodic visual presentation.

video-gain control—A control for adjusting the amplitude of a video signal. Two such controls are provided in the matrix section of some color television receivers so that the proper ratios between the amplitudes of the three color signals can be obtained.

video integration—A method of improving the output signal-to-noise ratio by utilizing the redundancy of repetitive signals to sum the successive video signals.

video integrator — 1. An electric counter-countermeasure device used to reduce the response to nonsynchronous signals such as noise; it is useful against random pulse signals and noise. 2. A device which uses the redundancy of repetitive signals to improve the output signal-to-noise ratio by summing the successive video signals.

video mapping—The procedure whereby a chart of an area is electronically superimposed on a radar display.

video mixer—A circuit or device used to combine the signals from two or more television cameras.

video recording (magnetic tape) — The methods of recording data having a bandwidth in excess of 100 kHz on a single track.

video signal—The picture signal in a television system—generally applied to the signal itself and the required synchronizing and equalizing pulses.

video stretching—In navigation, a procedure whereby the duration of a video pulse is increased.

video tape—A wide magnetic tape designed for recording and playing back a composite black and white or color television signal.

video tape recording — Abbreviated vtr. A method of recording television picture and sound signals on tape for reproduction at some later time.

vidicon—A camera tube in which a charge-density pattern is formed by photoconduction and stored on that surface of the photoconductor which is scanned by an electron beam, usually of low-velocity electrons.

viewfinder—An auxiliary optical or electronic device attached to a television camera so the operator can see the scene as the camera sees it.

viewing mirror—A mirror used in some tele-

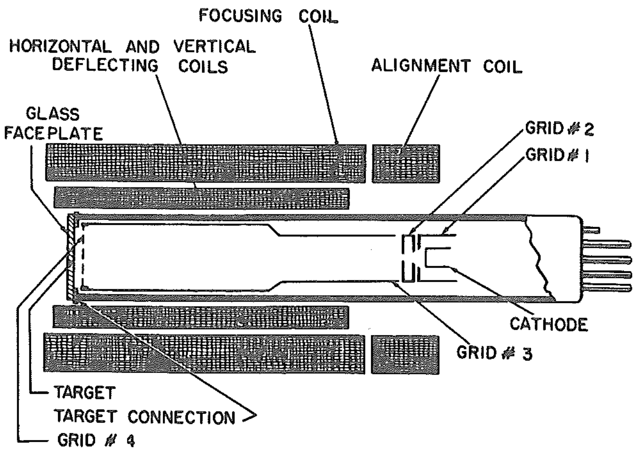

Vidicon.

vision receivers to reflect the image formed on the screen of the picture tube at an angle convenient to the viewer.

viewing screen—The face of a cathode-ray tube on which the image is produced.

Villari effect—A phenomenon in which a change in magnetic induction occurs when a mechanical stress is applied along a specified direction to a magnetic material having magnetostrictive properties.

vinyl resin—A soft plastic used for making phonograph records.

virgin tape—*See* Raw Tape.

virtual address—In a computer, an immediate, or real-time, address.

virtual cathode — An electron cloud that forms around the outer grid in a thermionic vacuum tube when the inner grid is maintained slightly more positive than the cathode.

virtual height—The height of the equivalent reflection point that will cause a wave to travel to the ionosphere and back in the same time required for an actual reflection. In determining the virtual height, the wave is assumed to travel at uniform speed and the height is determined by the time required to go to the ionosphere and back at the assumed velocity of light.

virtual image—The optical counterpart of an object, formed at imaginary focuses by prolongations of light rays (e.g., the image that appears to be behind an ordinary mirror).

virtual ppi reflectoscope — A device for superimposing a virtual image of a chart onto the ppi pattern. The chart is usually prepared with white lines on a black background to the scale of the ppi range scale.

viscometer—Also called a viscosimeter. A device for measuring the degree to which a liquid resists a change in shape.

viscosimeter—*See* Viscometer.

viscosity—The frictional resistance offered by one part or layer of a liquid as it moves past an adjacent part or layer of the same liquid.

viscous and magnetic damping — Damping by virtue of the viscosity of a fluid around the sensing element or of a magnetic field.

viscous-damped arm—A phonograph pickup arm mounted on a liquid cushion of oil, which provides high damping to eliminate arm resonances. It also protects the record groove and stylus; the arm does not fall on the record when dropped, but floats down gently.

visibility factor—Also called display loss. Ratio of the minimum input-signal power detectable by ideal instruments connected to the output of a receiver, to the minimum signal power detectable by a human operator through a display connected to the same receiver. The visibility factor may include the scanning loss.

visible radiation — Radiation with wavelengths ranging from about 4,000 to 8,000 angstrom units, corresponding to the visible spectrum of light.

visual aural range — Abbreviated var. A special type of vhf range providing a pair of radial lines of position which are reciprocal in bearing and are displayed to the pilot on a zero-center, left-right indicator. This facility also provides a pair of reciprocal radial lines of position located 90° from the above visually indicated lines. These are presented to the pilot as aural A-N radio-range signals, which provide a means for differentiating between the two visually indicated lines (and vice versa).

visual carrier frequency—The frequency of the television carrier which is modulated by the picture information.

visual communication — Communication by optical signs such as flags and lights.

visual radio range — Abbreviated vrr. A radio range, the course of which is followed by means of visual instruments.

visual transmitter — Also called a picture transmitter. In television, the radio equipment for transmission of the picture signals only.

visual transmitter power—The peak power output during transmission of a standard television signal.

vitreous—Having the nature of glass.

vlf — Abbreviation for very low frequency.

vocabulary—A list of operating codes or instructions available for writing the program for a given problem and for a specific computer.

vodas—Acronym for Voice-Operated Device, Anti-Sing. A system for preventing the overall voice-frequency singing of a two-way telephone circuit by disabling one direction of transmission at all times.

voder—Acronym for Voice Operation DEmonstratoR. An electronic device capable of artificially producing voice sounds. It uses vacuum tubes in connection with electrical filters controlled through a keyboard.

vogad—Acronym for Voice-Operated Gain-Adjusting Device. A voice-operated device used to give a substantially constant volume output for a wide range of inputs.

voice coil—Also called a speaker voice coil. A coil attached to the diaphragm of a dynamic speaker and moved through the air gap between the pole pieces.

voice filter—A parallel-resonant circuit connected in series with a line feeding several speakers. Its purpose is to remove the tubbiness of the male voice. The frequency of resonance is adjusted somewhere between 125 and 300 Hz.

voice frequency—Also called the speech frequency. The audible range of frequencies (32-16,000 Hz). In telephony, the voice range for speech is about 100-3,500 Hz.

voice-frequency carrier telegraphy—Carrier telegraphy in which the carrier currents have frequencies such that the modulated currents may be transmitted over a voice-frequency telephone channel.

voice-frequency dialing—A method of dialing by which the direct-current pulses from the dial are transformed into voice-frequency alternating-current pulses..

voice-frequency telephony — Telephony in which the frequencies of the components of the transmitted electric waves are substantially the same as the frequencies of corresponding components of the actuating acoustical waves.

voice-grade channel—A channel suitable for the transmission of speech, digital or analog data, or facsimile, generally with a frequency range of about 300 to 3,000 hertz.

voice-grade circuit—A circuit suitable for the transmission of speech, digital or analog data, or facsimile, generally with a frequency range of about 300 to 3,000 hertz.

voice-operated device—A device which permits the presence of voice or sound signals to effect a desired control.

voice-operated device antising—A system for preventing the overall voice-frequency singing of a two-way telephone circuit by disabling one direction of transmission at all times.

voice-operated gain-adjusting device — A voice-operated device used to give a substantially constant-volume output for a wide range of inputs.

voice-operated loss control and suppressor —A voice-operated device which switches the loss out of the transmitting branch and inserts the loss into the receiving branch under control of the subscriber's speech.

volatile—1. A computer storage medium in which information cannot be retained without continuous power dissipation. 2. Capable of evaporating.

volatile memory—In computers, any memory which can return information only as long as energizing power is applied. The opposite of nonvolatile memory.

volatile storage—A computer storage device in which the stored information is lost when the power is removed (e.g., acoustic delay lines, electrostatics, capacitors, etc.).

volatile store—A storage device in which stored data are lost when the applied power is removed (e.g., an acoustic delay line).

Voldicon—The trade name of Adage, Inc., for a family of high-speed, all-semiconductor, current-balancing devices that use digital logic and readout for high-speed precision measurement of analog signals.

volt—Abbreviated V. The unit of measurement of electromotive force. It is equivalent to the force required to produce a current of 1 ampere through a resistance of 1 ohm.

Volta effect—*See* Contact Potential.

voltage—1. Electrical pressure—i.e., the force which causes current to flow through an electrical conductor. 2. Symbolized by **E**. The greatest effective difference of potential between any two conductors of a circuit.

voltage amplification—Also called voltage gain. Ratio of the voltage across a specified load impedance connected to a transducer, to the voltage across the input of the transducer.

voltage amplifier—An amplifier used specifically to increase a voltage. It is usually capable of delivering only a small current.

voltage-amplifier tube—A tube that is designed primarily as a voltage amplifier. It has high gain, but delivers very little output power.

voltage and power directional relay — A device which permits or causes the connection of two circuits when the voltage difference between them exceeds a given value in a predetermined direction, and causes these two circuits to be disconnected from each other when the power flowing between them exceeds a given value in the opposite direction.

voltage attenuation—Ratio of the voltage across the input of a transducer, to the voltage delivered to a specified load impedance connected to the transducer.

voltage balance relay — A device which operates on a given difference in voltage between two circuits.

voltage breakdown—1. The voltage necessary to cause insulation failure. 2. A rapid increase of current flow, from a relatively low value to a relatively high value, upon the application of a voltage to a PN junction or dielectric.

voltage-breakdown test—A test whereby a specified voltage is applied between given points in a device, to ascertain that no breakdown will occur at that specified voltage.

voltage calibrator—Test equipment which supplies accurate a-c voltages for comparison on a scope screen with other waveforms to determine their voltage level.

voltage circuit of a meter—The combination of conductors and windings of the meter itself, excluding multipliers, shunts, or other external circuitry, to which is applied the voltage to be measured, a definite fraction of that voltage, or a voltage dependent on it.

voltage coefficient of capacitance — Also called voltage sensitivity. The quotient of the derivative with respect to voltage of a capacitance characteristic at a point divided by the capacitance at that point.

voltage comparator—An amplifying device with a differential input that will provide an output polarity reversal when one input signal exceeds the other. When operating with open loop and without phase compensation, operational amplifiers make fast and accurate voltage comparators.

voltage-controlled capacitor—*See* Varactor.

voltage-controlled crystal oscillator — A crystal oscillator the operating frequency of which can be changed by applying a controlling voltage to introduce a phase shift in the oscillator circuit.

voltage-controlled oscillator—An oscillator the frequency of oscillation of which can be varied by a changing voltage.

voltage corrector—An active source of regulated power placed in series with the output of an unregulated supply. The voltage corrector senses changes in the output voltage (or current) and corrects for these changes automatically by varying its own output in the opposite direction so as to maintain the total output voltage constant.

voltage/current crossover—The characteristic of a power supply that automatically converts the mode of operation of a power supply from voltage regulation to current regulation (or vice versa) as required by preset limits. The constant-current and constant-voltage settings are independently adjustable over specified limits. The region near the intersection of the constant-voltage and constant-current curves is described by the term "crossover characteristics."

voltage-directional relay—1. A relay which functions in conformance with the direction of an applied voltage. 2. A device which operates when the voltage across an open circuit breaker or contacter exceeds a given value in a given direction.

voltage divider—Also called a potential divider. A resistor or reactor connected across a voltage and tapped to make a fixed or variable fraction of the applied voltage available. (*Also see* Potentiometer and Rheostat.)

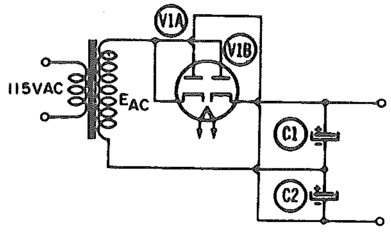

Voltage doubler.

voltage doubler—A voltage multiplier which rectifies each half cycle of the applied alternating voltage separately, and then adds the two rectified voltages to produce a direct voltage having approximately twice the peak amplitude of the applied alternating voltage.

voltage drop—The difference in voltage between two points, due to the loss of electrical pressure as a current flows through an impedance.

voltage feed—Excitation of a transmitting antenna by applying voltage at a point of maximum potential (at a voltage loop or antinode).

voltage feedback—A form of amplifier feedback in which the voltage drop across part of the load impedance is put in series with the input-signal voltage.

voltage gain—*See* Voltage Amplification.

voltage generator—A two-terminal circuit element with a terminal voltage independent of the current through the element.

voltage gradient — The voltage per unit length along a resistor or other conductive path.

voltage inverter—A circuit having a response (output) proportional to a constant (the gain) times the input signal, but opposite in sign to it. In a unit-gain inverter, the output is (−1) times the input.

voltage jump—An abrupt change or discontinuity in the tube voltage drop during operation of glow-discharge tubes.

voltage level—Ratio of the voltage at any point in a transmission system, to an arbitrary value of voltage used as a reference. In television and other systems where waveshapes are not sinusoidal or symmetrical about a zero axis and where the sum of the maximum positive and negative excursions of the wave is important in system performance, the two voltages are given as peak-to-peak values. This ratio is usually expressed in dbv, signifying decibels referred to 1 volt peak-to-peak.

voltage limit—A control function that maintains a voltage between predetermined values.

voltage loop—A point of maximum voltage in a stationary wave system. A voltage loop exists at the ends of a half-wave antenna.

voltage-measuring equipment — Equipment for measuring the magnitude of an alternating or direct voltage.

voltage multiplier — 1. A rectifying circuit which produces a direct voltage approximately equal to an integral multiple of the peak amplitude of the applied alternating voltage. 2. A series arrangement of capacitors charged by rapidly rotating brushes in sequence, giving a high direct voltage equal to the source voltage multiplied by the number of capacitors in series. 3. A precision resistor used in series with a voltmeter to extend its measuring range.

voltage node—A point having zero voltage in a stationary-wave system (e.g., at the center of a half-wave antenna).

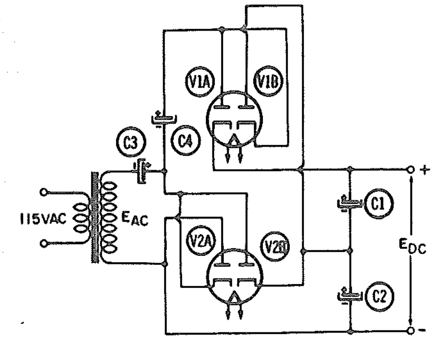

Voltage quadrupler.

voltage quadrupler—A rectifier circuit in which four diodes are employed to produce a d-c voltage of four times the peak value of the a-c input voltage.

voltage-range multiplier—Also called an instrument multiplier. A series resistor installed external to the measurement device to extend its voltage range.

voltage rating—Also called the working voltage. The maximum voltage which an electrical device or component can sustain without breaking down.

voltage ratio (of a transformer)—Ratio of the rms primary terminal voltage to the rms secondary terminal voltage under specified load conditions.

voltage-ratio box—*See* Measurement Voltage Divider.

voltage reference—A highly regulated voltage source used as a standard to which the output voltage of a power supply is continuously compared for purposes of regulation.

voltage-reference tube—A gas tube in which the voltage drop is essentially constant over the operating range of current, and is relatively stable at fixed values of current and temperature.

voltage reflection coefficient—The ratio of the complex electric field strength or voltage of a reflected wave to that of the incident wave.

voltage-regulating transformer — A saturated-core type of transformer which holds the output voltage to within a few percent, with input variations up to ±20%. Considerable harmonic distortion results unless extensive filters are employed.

voltage regulation—A measure of the degree to which a power source maintains its output-voltage stability under varying load conditions.

voltage regulator—1. A device that maintains or varies the terminal voltage of a generator or other machine at a predetermined value. 2. A gas-filled electronic tube which has the property of maintaining a nearly constant voltage across its terminals over a considerable range of current through the tube. It is used in an electronic voltage regulator.

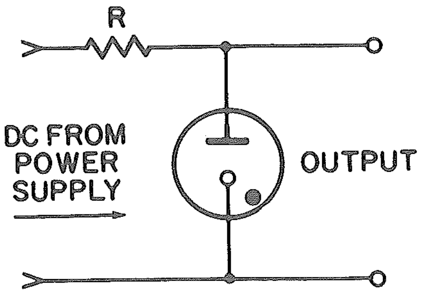

Voltage-regulator tube.

voltage-regulator tube—Also called a VR tube. A glow-discharge, cold-cathode tube in which the voltage drop is essentially constant over the operating range of current, and which is designed to provide a regulated direct-voltage output.

voltage relay—A relay that functions at a predetermined value of voltage.

voltage saturation—*See* Plate Saturation.

voltage-sensitive resistor—A resistor (e.g., a varistor), the resistance of which varies with the applied voltage.

voltage sensitivity—1. The voltage that produces standard deflection of a galvanometer when impressed on a circuit made up of the galvanometer coil and the external critical-damping resistance. The voltage sensitivity is equal to the product of the current sensitivity and the total circuit resistance. 2. *See* Voltage Coefficient of Capacitance.

voltage-stabilizing tube—A gas-filled glow-discharge tube normally working in that part of characteristic where the voltage is practically independent of current drop within a given range.

voltage standard — An accurately known voltage source (e.g., a standard cell) used for comparison with or calibration of other voltages.

voltage standing-wave ratio — Abbreviated vswr. Ratio between the sum and difference of the incident and reflected voltage waves.

voltage to ground—The voltage between any live conductor of a circuit and earth (or common reference plane).

voltage transformer—*See* Potential Transformer.

voltage tripler—A rectifier circuit in which three diodes are employed to produce a d-c voltage equal to approximately three times the peak a-c input voltage.

voltage-tunable magnetron — A high-frequency, continuous-wave oscillator operating in the microwave region. Power outputs begin in the milliwatts range and extend through hundreds of watts.

voltage-tunable tube — An oscillator tube the operating frequency of which can be changed by varying one or more of its electrode voltage (e.g., a backward-wave magnetron).

voltage-type telemeter — A telemeter in which the translating means is the magnitude of a single voltage.

voltage-variable capacitor—*See* Varactor.

voltage velocity limit—*See* Slewing Rate.

voltaic cell—Early name for a primary cell.

voltaic couple—Two dissimilar metals in contact, resulting in a contact potential difference.

voltaic pile—A voltage source consisting of alternate pairs of dissimilar metal discs separated by moistened pads, forming a number of elementary primary cells in series.

volt-ammeter—An instrument, calibrated to read both voltage and current.

volt-ampere—Abbreviated VA. A unit of apparent power in an a-c circuit containing reactance. It is equal to the potential in volts multiplied by the current in amperes, without taking phase into consideration.

volt-ampere-hour meter — An electricity meter which measures the integral, usually in kilovolt-ampere-hours, of the apparent power in the circuit where the meter is connected.

volt-ampere meter—An instrument for measuring the apparent power in an alternating-current circuit. Its scale is graduated in volt-amperes or kilovolt-amperes.

volt-ampere reactive—Abbreviated var. Also called reactive volt-ampere. The unit of reactive power.

Volta's law—When two dissimilar conductors are placed in contact, the same contact potential is developed between them, whether the contact is direct or through one or more intermediate conductors.

volt box—*See* Measurement Voltage Divider.

volt-electron — An obsolete expression for electron-volt.

voltmeter — An instrument for measuring potential difference. Its scale is usually graduated in volts. If graduated in millivolts or kilovolts, the instrument is usually designated as a millivoltmeter or a kilovoltmeter.

voltmeter-ammeter—A voltmeter and an ammeter combined into a single case, but with separate circuits.

voltmeter sensitivity—Expressed in ohms-per-volt, the ratio of the total resistance of a voltmeter, to its full-scale reading in volts.

volt-ohm-milliammeter—A test instrument with several ranges, for measuring voltage, current, and resistance.

volume—1. Also called power level. 2. The magnitude (measured on a standard volume indicator) of a complex audio-frequency wave, expressed in volume units. In addition, the term "volume" is used loosely to signify either the intensity of a sound or the magnitude of an audio-frequency wave. 3. The amount or a measure of energy in an electrical or acoustical train of waves.

volume compression—Also called automatic volume compression. The limiting of the volume range to about 30 to 40 decibels at the transmitter, to permit a higher average percentage modulation without overmodulation. Also used in recording to raise the signal-to-noise ratio.

volume compressor—A circuit which provides volume compression.

volume conductivity—*See* Conductivity.

volume control—A variable resistor for adjusting the loudness of a radio receiver or amplifying device.

volume equivalent—A measure of the loudness of speech reproduced over a complete telephone connection. It is expressed numerically in terms of the trunk loss of a working reference system which has been adjusted to give equal loudness.

volume expander—A circuit which provides volume expansion.

volume expansion—*See* Automatic Volume Expansion.

volume indicator—An instrument for indicating the volume of a complex electric wave such as that corresponding to speech.

volume lifetime—The average time interval between the generation and recombination of minority carriers in a homogeneous semiconductor.

volume limiter—An amplifier the gain of which is automatically reduced when the average input volume to the amplifier exceeds a predetermined value, so that the output for all inputs in excess of this value is substantially constant. A volume limiter differs from a peak limiter in that it is controlled by the average volume instead of by the instantaneous peaks.

volume-limiting amplifier — An amplifier which reduces the gain whenever the input volume exceeds a predetermined level, so that the output volume is maintained substantially constant. The normal gain is restored whenever the input volume drops below the predetermined limit.

volume magnetostriction—The relative volume change of a body of ferromagnetic material when the magnetization of the body is increased from zero to a specified value (usually saturation) under specified conditions.

volume range—1. Of a transmission system, the difference, expressed in db, between the maximum and minimum volumes which the system can satisfactorily handle. 2. Of a complex audio-frequency signal, the difference, expressed in db, between the maximum and minimum volumes occurring over a specified period.

volume recombination rate—The rate at which free electrons and holes recombine within the volume of a semiconductor.

volume resistance—Ratio of the d-c voltage applied to two electrodes in contact with or embedded in a specimen, to that portion of the current between them distributed through the specimen.

volume resistivity—1. Ratio of the potential gradient parallel to the current in a material, to the current density. (*Also see* Resistivity.) 2. Also called specific insulation resistance. The electrical resistance between opposite faces of a 1-cm cube of insulating material, commonly expressed in ohm-centimeters.

volumetric displacement — The change in volume required to displace the diaphragm of a pressure transducer from its rest position to a position corresponding to the application of a stimulus equal to the rated range of the transducer.

volumetric efficiency—Also called packing factor. The ratio of parts volume to total equipment volume, expressed in percent. In modules, it is usually taken as the volume of component bodies only (not including leads, other interconnecting media, insulators, heat sinks, and so on) and based on nominal sizes of components and nominal outside module dimensions.

volumetric radar—A radar capable of producing three-dimensional position data on several targets.

volume unit—1. The unit of transmission measurement for measuring the level of nonsteady-state currents. Zero level is the steady-state reference power of 1 milliwatt in a circuit of 600 ohms characteristic impedance. 2. A measure of the power level of the voice wave. Zero volume unit is equivalent to +4 dbm for simple electrical waves (single frequencies). 3. A standardized unit for specifying the power of voice or music.

volume-unit indicator—Also called a VU indicator or volume-unit meter. An instrument calibrated to read audio-frequency power levels directly in volume units.

volume-unit meter—Abbreviated VU meter. *See* Volume-Unit Indicator.

volume velocity—The rate at which a medium flows through a specified area due to a sound wave.

Von Hippel breakdown theory—Also called low-energy criterion. Breakdown occurs at fields for which the rate of recombination of electrons and positive holes is less than the rate of collisional ionization. Assumes no distribution in electron energies.

vor—Abbreviation for vhf omnirange.

vowel articulation—The percent of articulation obtained when the speech units considered are vowels, usually combined with consonants into meaningless syllables.

V ring — In commutator construction, a specially shaped insulating structure having one or more V-shaped sections.

vrr—Abbreviation for visual radio range.

vr tube — Abbreviation for voltage-regulator tube.

vswr — Abbreviation for voltage standing-wave ratio.

VTL — Abbreviation for variable threshold logic.

vtm—*See* voltage-tunable magnetron.

vtr — Abbreviation for video-tape recording.

vtvm — Abbreviation for vacuum-tube voltmeter.

VU—Abbreviation for volume unit.

VU indicator—Abbreviation for volume-unit indicator.

VU meter — Abbreviation for volume-unit meter. A volume indicator with a decibel scale and specified dynamic and other characteristics. It is used to obtain correlated readings of speech power necessitated by the rapid fluctuations in level of voice currents.

vulcanized fiber—A laminated plastic insulation made of paper and cellulose dried under heavy pressure.

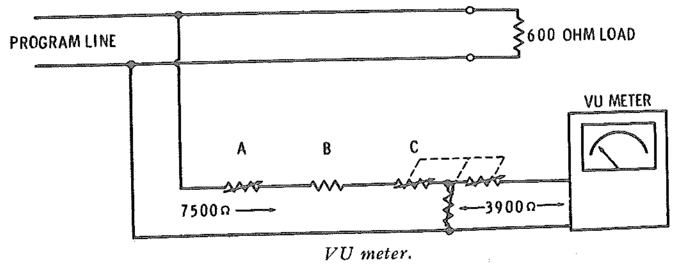

VU meter.

W

W—Symbol for energy, watt, or work.

wafer—1. A thin semiconductor slice of silicon or germanium with parallel faces on which matrices of microcircuits or individual semiconductors can be formed. After processing, the wafer is separated into dice or chips containing individual circuits. 2. A single section of a wafer switch.

wafer socket—A vacuum-tube socket that consists of two punched sheets or wafers of an insulating material, separated by spring-metal clips that grip the terminal pins of the inserted tube.

wafer switch—A rotary multiposition switch with fixed terminals on ceramic or Bakelite wafers. The rotor arm, in the center, is positioned by a shaft. Several decks may be stacked onto one switch and rotated by a common shaft.

Wagner ground—A bridge with an additional pair of ratio arms, onto which the ground connection to the bridge is moved in order to effect a perfect balance, free from error.

waiting time—In certain tubes (e.g., thyratrons), the time that must elapse between the turning on of their heaters and the application of plate voltage. (*Also see* Access Time.)

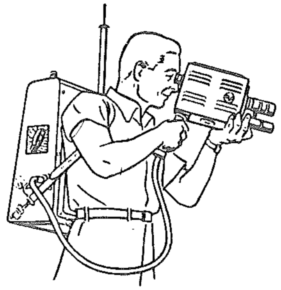

Walkie-lookie.

walkie-lookie—A compact, portable television camera used for remote broadcasts. The resultant electrical pulses are transmitted by microwave radio to a local control point for retransmission over a standard television station.

walkie-talkie—A two-way radio communication set designed to be carried by one person, usually strapped to his back, and capable of being operated while in motion.

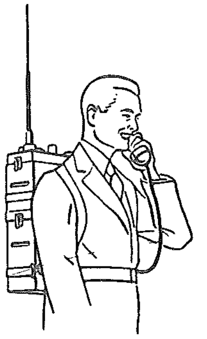

Walkie-talkie.

wall box—A metal box placed in the wall and containing switches, fuses, etc.

wall outlet—A spring-contact device to which a portable lamp or appliance is connected by means of a plug attached to a flexible cord. The wall outlet is installed in a box, and connected permanently to the power-line wiring of a home or building.

walls—The sides of the groove in a disc record.

wamoscope—Acronym for a WAve-MOdulated oscilloSCOPE. A cathode-ray tube which includes detection, amplification, and display of a microwave signal in a single envelope, thus eliminating the local oscillator, mixer, i-f amplifier, detector, video amplifier, and associated circuitry in a conventional radar receiver. Tubes are available for a range of 2,000 to 4,000 MHz.

wander—*See* Scintillation.

warble-tone generator—An oscillator, the frequency of which is varied cyclically at a subaudio rate over a fixed range. It is usually used with an integrating detector to obtain an averaged transmission or cross-talk measurement.

warm-up time—1. In an indirectly-heated tube, the time which elapses, after the heater is turned on, before the cathode reaches its optimum operating temperature. 2. The time, following power application to a device, required for the output to stabilize within specifications. 3. In a regulated power supply, the time, following power application to the supply, required for the output voltage (or current) to stabilize within specifications.

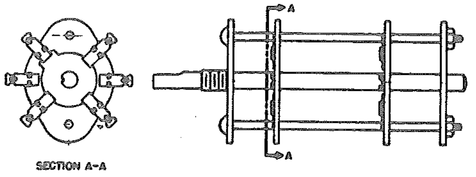

Wafer switch.

watch—*See* Radio Watch.

water-activated battery—A primary which contains the electrolyte but requires the addition of (or immersion in) water before it becomes usable.

water-cooled tube—A vacuum tube having an anode structure projecting through the glass envelope and constructed to permit circulation of water around the anode for cooling purposes during operation.

water load—A matched waveguide termination in which the electromagnetic energy is absorbed in water. The output power is calculated from the difference in temperature between the water at the input and output.

waterproof motor—A totally enclosed motor so constructed that it will exclude water applied in the form of a stream from a hose. Leakage may occur around the shaft, provided it is prevented from entering the oil reservoir and provision is made for automatically draining the motor. The means for automaitc draining may be a check valve, or a tapped hole at the lowest part of the frame which will serve for application of a drain pipe.

watt—1. Abbreviated W. A unit of the electric power required to do work at the rate of 1 joule per second. It is the power expended when 1 ampere of direct current flows through a resistance of 1 ohm. 2. Practical unit of electric power. In an alternating-current circuit, the true power in watts is effective volts multiplied by the circuit power factor (there are 746 watts in 1 horsepower).

wattage rating—The maximum power that a device can safely handle.

watt-hour—A unit of electrical work indicating the expenditure of 1 watt of electrical power for 1 hour. Equal to 3,600 joules.

watt-hour capacity—The number of watt-hours delivered by a storage battery at a specified temperature, rate of discharge, and final voltage.

watt-hour constant of a meter—The registration, expressed in watt-hours, corresponding to one revolution of the rotor.

watt-hour-demand meter—A combined watt-hour meter and demand meter.

watt-hour meter—An electricity meter which measures and registers the integral, usually in kilowatt-hours, of the active power of the circuit into which the meter is connected. This power integral is the energy delivered to the circuit during the integration interval.

wattless component—A reactive component.

wattless power—*See* Reactive Power.

wattmeter—An instrument for measuring the magnitude of the active power in an electric circuit. Its scale is usually graduated in watts. If graduated in kilowatts or megawatts, the instrument is usually designated as a kilowattmeter or megawattmeter.

watt-second—The amount of energy corresponding to 1 watt acting for 1 second. It is equal to 1 joule.

watt-second constant of a meter—The registration, in watt-seconds, corresponding to one revolution of the rotor.

wave—A physical activity that rises and falls, or advances and retreats, periodically as it travels through a medium.

wave amplitude—The maximum change from zero of the characteristic of a wave.

wave analyzer—An electric instrument for measuring the amplitude and frequency of the various components of a complex current or voltage wave.

wave angle—The angle at which a wave is propagated from one point to another.

wave antenna—Also called a Beverage antenna. A directional antenna composed of parallel horizontal conductors one-half to several wavelengths long, and terminated to ground in its characteristic impedance at the far end.

wave band—A band of frequencies, such as that assigned to a particular type of communication service.

wave-band switch—A multiposition switch for changing the frequency band tuned by a receiver or transmitter.

wave clutter—Clutter caused on a radar screen by echoes from sea waves.

wave converter—A device for changing a wave from one pattern to another (e.g., baffle-plate, grating, and sheath-reshaping converters for wave guides).

wave duct—1. A tubular waveguide capable of concentrating the propagation of waves within it. 2. A natural duct formed in air by atmospheric conditions. Waves of certain frequencies travel through it with more than average efficiency.

wave equation—An equation that gives a mathematical specification of a wave process, or describes the performance of a medium through which a wave is passing.

wave filter—A transducer for separating waves on the basis of their frequency. It introduces a relatively small insertion loss to waves in one or more frequency bands, and a relatively large insertion loss to waves of other frequencies. (*Also see* Filter.)

waveform—1. The shape of an electromagnetic wave. 2. The graphic representation of the wave in (1) above, showing the variations in amplitude with time.

waveform-amplitude distortion—Sometimes called amplitude distortion. Nonlinear waveform distortion caused by unequal attenuation or amplification between the input and output of a device.

waveform analyzer—An instrument that measures the amplitude and frequency of the components in a complex waveform.

waveform generator—*See* Timing-pulse Distributor.

waveform influence—The change in meter indication, caused solely by a change in waveform from a specified waveform, of the applied current and/or voltage.

waveform monitor — An oscilloscope designed especially for viewing the waveform of a video signal.

waveform synthesizer—Equipment for generating a signal of a desired waveform.

wavefront—1. In a progressive wave in space, a continuous surface which is a locus of points having the same phase at a given instant. 2. That part of a signal-wave envelope between the initial point of the envelope and the point at which the envelope reaches its crest.

wave function—In a wave equation, a point function that specifies the amplitude of a wave.

waveguide—1. A system of material boundaries capable of guiding electromagnetic waves. 2. A transmission line comprising a hollow conducting tube within which electromagnetic waves are propagated on a solid dielectric or dielectric-filled conductor.

waveguide attenuator — A waveguide device for producing attenuation by some means (e.g., by absorption and reflection).

waveguide connector—Also called a waveguide coupling. A mechanical device for electrically joining parts of a waveguide system together.

waveguide coupling—*See* Waveguide Connector.

waveguide critical dimension—The dimension of waveguide cross section which determines the cutoff frequency.

waveguide cutoff frequency — Also called the critical frequency. The frequency limit of propagation, along a waveguide, for waves of a given field configuration.

waveguide dummy load—Sections of waveguide for dissipating all the power entering the input flange.

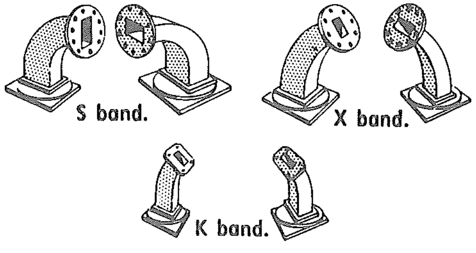

Waveguide elbows.

waveguide elbow—A bend in a waveguide.

waveguide flange—*See* Flange.

waveguide lens — A microwave device in which the required phase changes are produced by refraction through suitable waveguide elements acting as lenses.

waveguide mode suppressor—A filter used to suppress undesired modes of propagation in a waveguide.

waveguide phase shifter—A device for adjusting the phase of the output current or voltage relative to the phase at the input of a device.

waveguide plunger — In a waveguide, a plunger used for reflecting the incident energy.

waveguide post — In a waveguide, a rod placed across the waveguide and behaving substantially like a shunt susceptance.

waveguide propagation—Long-range communications in the 10-kHz to 35-kHz frequency range by the waveguide characteristics of the atmospheric duct formed by the ionospheric D-layer and the surface of the earth. *Also see* Atmospheric Duct.

waveguide resonator—A waveguide device intended primarily for storing oscillating electromagnetic energy.

waveguide shim — A thin metal sheet inserted between waveguide components to insure electrical contact.

waveguide shutter—A vane within a waveguide, used to protect the receiver system from adjacent radar power by establishing an electrical short across the waveguide when a companion transmitter is operating.

waveguide stub — An auxiliary section of wave guide that has an essentially nondissipative termination and is joined to the main section of the waveguide.

waveguide switch — A transmission-line switch for connecting a transmitter or receiver from one antenna to another or to a dummy load.

waveguide taper—A section of tapered wave guide.

waveguide tee — A junction for connecting a branch section of waveguide in series or parallel with the main transmission line.

waveguide transformer — A device, usually fixed, added to a waveguide for the purpose of impedance transformation.

waveguide tuner — An adjustable device added to a waveguide for the purpose of impedance transformation.

waveguide twist—A waveguide section in which the cross section rotates about the longitudinal axis.

waveguide wavelength—Also called a guide wavelength. For a traveling plane wave at a given frequency, the distance along the waveguide, between the points at which a field component (or the voltage or current) differs in phase by 2π radians.

wave heating—The heating of a material by energy absorption from a traveling electromagnetic wave.

wave impedance (of a transmission line)—At every point in a specified plane, the complex ratio between the transverse components of the electric and magnetic fields. (Incident and reflected waves may both be present.)

wave interference—The phenomenon which results when waves of the same or nearly same type and frequency are superimposed.

It is characterized by variations in the wave amplitude which differ from that of the individual superimposed waves.

wavelength—1. In a periodic wave, the distance between points of corresponding phase of two consecutive cycles. The wavelength (λ) is related to the phase velocity (v) and frequency (f) by the formula $\lambda = v/f$. 2. The quotient of the velocity of propagation divided by the frequency.

wavelength constant—The imaginary part of the propagation constant—i.e., the part that refers to the retardation in phase of an alternating current passing through a length of transmission line.

wavelength shifter—A photofluorescent compound employed with a scintillator material. Its purpose is to absorb photons and emit related photons of a longer wavelength, thus permitting more efficient use of the photons by the phototube or photocell.

wave mechanics—A general physical theory whereby wave characteristics are assigned to the components of atomic structure and all physical phenomena are interpreted in terms of hypothetical waveforms.

wavemeter—An instrument for measuring the wavelength of a radio-frequency wave. Resonant-cavity, resonant-circuit, and standing-wave meters are representative types.

wave normal—A unit vector normal to an equiphase surface, with its positive direction taken on the same side of the surface as the direction of propagation. In isotropic media, the wave normal is in the direction of propagation.

wave number—The number of waves per unit length. The usual unit of wave number is the reciprocal centimeter, cm-1. In terms of this unit, the wave number is the reciprocal of the wavelength when the latter is in centimeters in a vacuum.

wave packet—A short pulse of waves (e.g., spin waves.

wave propagation—*See* Propagation.

waveshape—A graph of a wave as a function of time or distance.

wave tail—That part of a signal-wave envelope between the steady-state value (or crest) of an envelope and the end.

wave tilt—The forward inclination of a radio wave due to its proximity to ground.

wave train—A limited series of wave cycles caused by periodic short-duration disturbances.

wave trap — 1. Also called a trap. A device used to exclude unwanted signals or interference from a receiver. Wave traps are usually tunable to enable the interfering signal to be rejected or the true frequency of a received signal to be determined. 2. A resonant circuit connected into the antenna system of a radio receiver to suppress undesired signals at a particular frequency.

wave velocity—A quantity which specifies the speed and direction at which a wave travels through a medium.

way-operated circuit—A circuit shared by three or more stations on a "party-line" basis. One of the stations may be a switching center. May be a single or duplex circuit.

way point—A selected point having some particular significance on a radionavigational course line.

way station—In telegraphy, one of the stations on a multipoint network.

wax—In mechanical recording, a blend of waxes with metallic soaps.

wax master—*See* Wax Original.

wax original—Also called a wax master. An original recording on a wax surface, from which the master is made.

weak coupling—Loose coupling in a radio-frequency transformer. *Also see* Loose Coupling.

wearout—The point at which the continued operation and repair of an item becomes uneconomical because of the increased frequency of failure. The end of the useful life of the item.

wearout failure—A failure that is predictable on the basis of known wearout characteristics. This type of failure is due to deterioration processes or mechanical wear, the probability of occurrence of which increases with time.

weather-protected motor—An open motor the ventilating passages of which are so designed as to minimize the entrance of rain, snow, and airborne particles to the electric parts.

weber—The practical unit of magnetic flux equal to the amount which, when linked at a uniform rate with a single-turn electric circuit during an interval of 1 second, will induce an electromotive force of 1 volt in the circuit. One weber equals 10^8 maxwells.

wedge—1. The fan-shaped pattern of equidistant black and white converging lines in a television test pattern. 2. A waveguide termination consisting of a tapered length of dissipative material, such as carbon, introduced into the guide.

wedge bonding—A type of thermocompression bonding used in intergrated-circuit manufacturing where a wedge-shaped tool is used to press a small section of the lead wire onto the bonding pad.

Wehnelt cathode—A hot cathode that consists of a metallic core coated with alkaline-earth oxides. It is widely used in vacuum tubes.

weight—The force with which a body is attracted toward the earth.

weight coefficient of a thermoelectric generator—The quotient of the electrical power output of the thermoelectric generator divided by the weight of the generator.

weight coefficient of a thermoelectric generator couple—The quotient of the electrical power output of the thermoelectric couple divided by the weight of the couple.

weighted distortion factor—The weighting of harmonics in proportion to their harmonic relationship.

weighted noise level — The noise level weighted in accordance with the 70-decibel equal-loudness contour of the human ear and expressed in dbm.

weighting—The artificial adjustment of measurements in order to account for factors which, during normal use of a device, would otherwise differ from the conditions during measurement. For example, background-noise measurements may be weighted by applying factors or introducing networks to reduce the measured values in inverse ratio to their interference.

weightlessness—1. A condition in which no acceleration, whether of gravity or other force, can be detected by an observer within the system in question. 2. A condition in which gravitational and other external forces acting on a body produce no stress, either internal or external, in the body. Weightlessness occurs when gravity forces are exactly balanced by other forms of acceleration (zero g).

weld—The consolidation of two metals, usually by application of heat to the proposed joint.

weldgate pulse — A waveform used in controlling the flow of welding current.

welding transformer—A power transformer with a secondary winding consisting of only a few turns of very heavy wire. It is used to produce high-value alternating currents at low voltages for welding purposes.

weld junction—A junction formed by heat or metallurgical fusion of conductors. It provides a strong electrical connection with good conductivity. It is widely used in microelectronic packaging. Wires, ribbons, or films as small as 0.0005-inch thick can be joined by resistance and electron-beam welding methods.

weld-on surface-temperature resistor — A surface-temperature resistor installed by welding the sensing element to the surface being measured.

weld polarity—Certain material combinations have a different resistance to a weld current, depending on the direction of the current. In d-c welding, a suitable weld may be possible in only one direction of current. A weld schedule must define the proper polarity for such cases.

weld time—The interval during which current is allowed to flow through the work during the performance of one weld. In pulsation welding, the weld period includes the "cool" time intervals.

Wertheim effect—When a wire placed in a longitudinal magnetic field is twisted, there will be a transient voltage difference between the ends of the wire.

Western Union joint—A strong, highly conductive splice made by crossing the cleaned ends of two wires, twisting them together, and soldering.

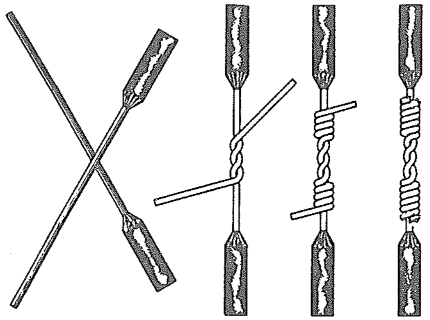

Western Union joint.

Weston normal cell—A standard cell of the saturated cadmium type in which the positive electrode is cadmium and the electrolyte is a cadmium-sulphate solution.

Westrex system—*See* Forty-five/Forty-five.

wet—Term describing the condition in which the liquid electrolyte in a cell is free-flowing.

wet cell—A cell the electrolyte of which is in liquid form and free to flow and move.

wet-charged stand—The period of time that a charged, wet secondary cell can stand before losing a specified, small percentage of its capacity.

wet contact—A contact through which direct current flows.

wet electrolytic capacitor—An electrolytic capacitor that has a liquid electrolyte.

wet flashover voltage—The voltage at which the air surrounding a clean, wet insulator shell breaks down completely between electrodes. This voltage will depend on the conditions under which the test is made.

wet-reed relay—A reed-type relay containing mercury at the relay contacts to reduce arcing and contact bounce.

wet shelf life—The period of time that a wet secondary cell can remain discharged without deteriorating to a point where it cannot be recharged.

wet tantalum capacitor—A polar capacitor the cathode of which is a liquid electrolyte (a highly ionized acid or salt solution). Characteristics: highest capacity per unit volume, low impedance, lowest d-c leakage, excellent shelf life.

wetting—The formation of a uniform, smooth, unbroken, and adherent film of solder to a base metal.

Wheatstone bridge—Also called resistance bridge. A null-type resistance-measuring circuit in which resistance is measured by direct comparison with a standard resistance.

wheel static—Auto-radio interference due to a static charge building up between the brake drum and the wheel spindle.

whiffletree switch—In computers, a multiposition electronic switch composed of gate tubes and flip-flops. It is so named because its circuit diagram resembles a whiffletree.

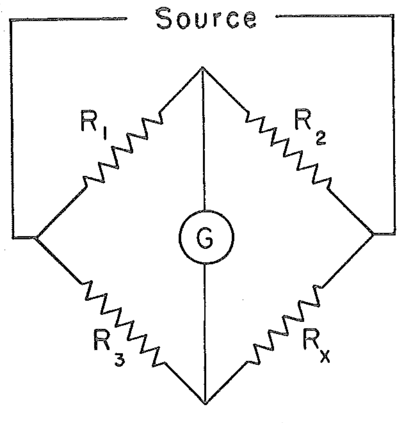

$$R_x = R_3R_2/R_1$$

Wheatstone bridge.

whip antenna — A simple vertical antenna consisting of a slender whiplike conductor supported on a base insulator.

whisker — 1. *See* Catwhisker. 2. Ultrapure elongated metal and ceramic filaments of extremely high tensile strengths.

whistler mode propagation—The transmission of radio waves between conjugate points with respect to the geomagnetic equator; i.e., points of opposite geomagnetic latitude and equal geomagnetic longitude by the apparent ducting of waves long the flux lines of the geomagnetic field.

whistlers — 1. High-frequency atmospherics that decrease in pitch and then tend to rise again. 2. Audio-frequency waves from lightning-stroke radiation that have penetrated the ionosphere.

white—1. For color tv, white is a mixture of red, green, and blue. In the picture this is produced by exciting all three dots in each phosphor trio. Since the eye cannot distinguish the individual dots, the mixture appears white. 2. The facsimile signal produced when an area of subject copy having minimum density is scanned.

white circuit—A cathode follower with another tube replacing the cathode resistor. By driving this additional tube with a signal that is out of phase with the original signal, low-impedance, broad-band characteristics are obtained.

white compression—Also called white saturation. The reduction in gain applied to a picture signal at those levels corresponding to light areas in a picture.

white-dot pattern—*See* Dot Pattern.

white level—The carrier-signal level which corresponds to maximum picture brightness in television and facsimile.

white light—Radiation producing the same color sensation as average sunlight at noon.

white noise—1. Random noise (e.g., shot and thermal noise) whose constant energy per unit bandwidth is independent of the central frequency at the band. The name is taken from the analogous definition of white light. 2. The electrical disturbance caused by the random movement of free electrons in a conductor or semiconductor. Since its electrical energy is evenly distributed throughout the entire frequency spectrum, it is useful for testing the frequency response of amplifiers, speakers, etc. 3. Continuous, random noise distributed uniformly across the frequency spectrum.

white object—An object which reflects all wavelengths of light with substantially equal high efficiencies and considerable diffusion.

white peak—A peak excursion of the picture signal in the white direction.

white raster—*See* Chroma-Clear Raster.

white recording—1. In an amplitude-modulation system, that form of recording in which the maximum received power corresponds to the minimum density of the record medium. 2. In a frequency-modulation system, that form of recording in which the lowest received frequency corresponds to the minimum density of the record medium.

white room—An area in which the atmosphere is controlled to eliminate dust, moisture, and bacteria. It is used in the production and assembly of components and systems, the reliability or functions of which might be adversely affected by the presence of foreign matter.

white saturation—*See* White Compression.

white signal—The facsimile signal produced when a minimum-density area of the subject copy is scanned.

white-to-black amplitude range—1. In a facsimile system employing positive amplitude modulation, the ratio of signal voltage (or current) for picture white to that for picture black at any point in the system. 2. In a facsimile system employing negative amplitude modulation, the ratio of the signal voltage (or current) for picture black to that for picture white. Both ratios are often expressed in decibels.

white-to-black frequency swing—In a facsimile system employing frequency modulation, the numerical difference between the signal frequencies corresponding to picture white and picture black at any point in the system.

white transmission—1. In an amplitude-modulation system, that form of transmission in which the maximum transmitted power corresponds to the minimum density of the subject copy. 2. In a frequency-modulation system, that form of transmission in which the lowest transmitted frequency corresponds to the minimum density of the subject copy.

whole step—*See* Whole Tone.

whole tone—Also called a whole step. The interval between two sounds with a basic frequency ratio approximately equal to the sixth root of 2.

wicking—1. The flow of solder up under the insulation on covered wire. 2. The act of drawing moisture through a fabric or thread, like the action of a wick in an oil lamp when it draws oil up to the flame.

wide-angle lens—A lens that picks up a wide area of a television stage setting at a short distance.

wideband—Capable of passing a broad range of frequencies (said of a tuner or amplifier). Especially vital to good multiplex reception, and for faithful audio reproduction.

wideband amplifier — An amplifier capable of passing a wide range of frequencies with equal gain.

wideband axis—In phasor representation of the chrominance signal, the direction of the phasor representing the fine-chrominance primary.

wideband communications system—A communications system which provides numerous channels of communications on a highly reliable and secure basis; the channels are relatively invulnerable to interruption by natural phenomenon or counter-measures. Included are multichannel telephone cable, tropospheric scatter, and multichannel line-of-sight radio systems such as microwave.

wideband improvement — Ratio of the signal-to-noise ratio of the system in question, to the signal-to-noise ratio of a reference system.

wideband ratio — Ratio of the occupied-frequency bandwidth to the intelligence bandwidth.

wide open—Refers to the untuned characteristic or lack of frequency selectivity.

wide-open receiver—A receiver that has essentially no tuned circuits and is designed to receive all frequencies simultaneously in the band of coverage.

width—1. The distance between two specified points of a pulse. 2. The horizontal dimension of a television or facsimile display.

width coding—Modifying the duration of the pulses emitted from the transponder according to a prearranged code for recognition in the display.

width control—A television-receiver or an oscilloscope control which varies the amplitude of the horizontal sweep and hence the width of the picture.

Wiedemann effect—The direct Wiedemann effect is the twist produced in a wire placed in a longitudinal magnetic field when a current flows through the wire; the twist is due to the helical resultant of the impressed longitudinal field and the circular field of the wire. The magnetic material expands (or contracts) parallel to the helical lines of force and hence the twist. The inverse Wiedemann effect is the axial magnetization of a current-carrying wire when twisted.

Wiedemann-Franz law—A theoretical result which states that the ratio of thermal conductivity to electrical conductivity is the same for all metals at the same temperature.

Wien bridge—An alternating-current bridge used to measure inductance or capacitance in terms of resistance and frequency. *Also see* Wien Capacitance Bridge *and* Wien Inductance Bridge.

Wien-bridge oscillator—An oscillator, the frequency of which is controlled by a Wien bridge.

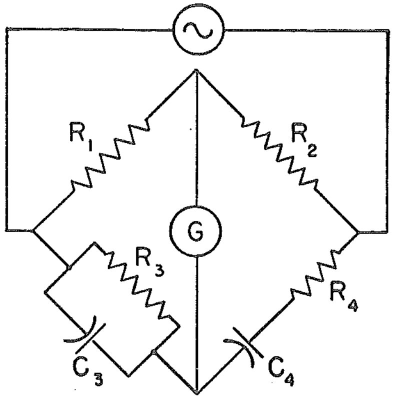

$$\frac{C_3}{C_4} = \frac{R_2}{R_1} - \frac{R_4}{R_3} \qquad C_3C_4 = \frac{1}{\omega^2 R_3 R_4}$$

Wien capacitance bridge.

Wien capacitance bridge—A four-arm, alternating-current capacitance bridge used for measuring capacitance in terms of resistance and frequency. Two adjacent arms contain capacitors—one in series and the other in parallel with a resistor—while the other two are normally nonreactive resistors. The balance depends on the frequency, but the capacitance of either or both capacitors can be computed from the resistances of all four arms and from the frequency.

Wien inductance bridge—A four-arm alternating-current inductance bridge used for measuring inductance in terms of resistance and frequency. Two adjacent arms contain inductors—one in series and the other in parallel with a resistor—while the other two are normally nonreactive resistors. The balance depends on the frequency, but the inductances of either or both inductors can be computed from the resistances of the four arms and from the frequency.

Wien's law—Better known as Wien's displacement law. The wavelength of maximum radiation intensity is inversely proportional to the absolute temperature of a black body, and the intensity of radiation at this maximum wavelength varies as the fifth power of the absolute temperature.

Williamson amplifier—A high-fidelity, push-pull, audio-frequency amplifier using triode-connected tetrodes. The circuit was developed by D. T. N. Williamson.

Williams-tube storage—A type of electrostatic storage using a cathode-ray tube.

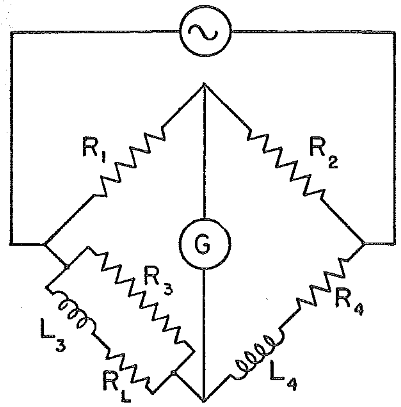

$$\frac{L_3}{L_4} = \frac{R_1(R_L + R_3)}{R_2R_3 - R_1R_4}$$

$$\omega^2 L_3 L_4 = R_4(R_L + R_3) - R_L R_3 \frac{R_2}{R_1}$$

Wien inductance bridge.

Wimshurst machine—A common static machine or electrostatic generator consisting of two coaxial insulating discs rotating in opposite directions. Sectors of tinfoil are arranged, with respect to a connecting rod and collecting combs, so that static electricity is produced for charging Leyden jars or discharging across a gap.

wind—The way in which recording tape is wound onto a reel. An A wind is one in which the tape is wound so that the coated surface faces toward the hub; a B wind is one in which the coated surface faces away from the hub. A uniform, as opposed to an uneven, wind is one giving a flat side tape pack, free from laterally displaced, protruding layers.

wind charger—A wind-driven d-c generator for changing batteries (e.g., 32-volt batteries formerly used on many farms).

winding—A conductive path, usually wire, inductively coupled to a magnetic core or cell. Windings may be designated according to function—e.g., sense, bias, drive, etc.

winding arc—In an electrical machine, the length of a winding stated in terms of degrees.

winding factor—The ratio of the total area of wire in the center hole of a toroid to the window area of a toroid or transformer core.

window—1. Strips of metal foil, wire, or bars dropped from aircraft or fired from shells or rockets as a radar countermeasure. 2. The small area through which beta rays enter a Geiger-Mueller tube.

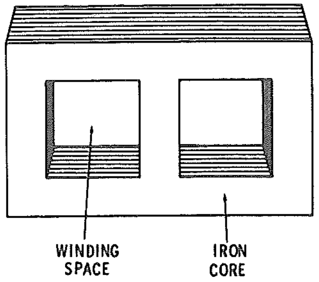

Window area.

window area—The opening in the laminations of a transformer.

window corridor—Also called the infected area or lane. An area where window has been sown.

window jamming — Reradiation of electromagnetic energy by reflecting it from a window to jam enemy electronic devices.

windshield—In radar, a streamlined cover placed in front of airborne paraboloid antennas to minimize wind resistance. The cover material is such as to present no appreciable attenuation to the radiation of the radar energy.

wing spot generator—An electronic circuit that grows wings on the video target signal of a type-G indicator. These wings are inversely proportional in size to the range.

wiper—1. The moving contact which makes contact with a terminal in a stepping relay or switch. 2. In a potentiometer, the contact that moves along the element, dividing the

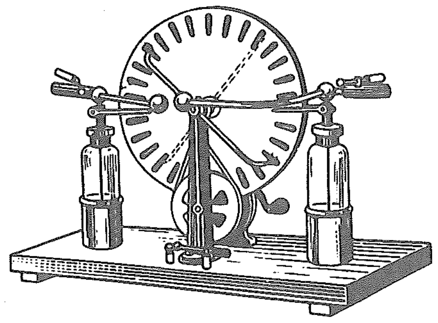

Wimshurst inductance static machine.

resistance according to its mechanical position.

wiper arm—In a pressure potentiometer, the movable electrical contact that is driven by the sensing element and moves along the coil.

wiping action—The action which occurs when contacts are mated with a sliding action. Wiping has the effect of removing small amounts of contamination from the contact surfaces, thus establishing better conductivity.

wiping contact—Also called self-cleaning contact, sliding contact, and self-wiping contact. A switch or relay contact designed to move laterally with a wiping motion when engaging with or disengaging from a mating contact.

wire—A solid or stranded group of solid cylindrical conductors having a low resistance to current flow, together with any associated insulation.

wire communication—Transmission of signs, signals, pictures, and sounds of all kinds over wire, cable, or other similar connections.

wired-program computer — A computer in which nearly all instructions are determined by the placement of interconnecting wires held in a removable plugboard. This arrangement allows for changes of operations by simply changing plugboards. If the wires are held in permanently soldered connections, the computer is called a fixed-program type.

wired radio—Communication whereby the radio waves travel over conductors.

wire drawing—The pulling of wire through dies made of tungsten carbide or diamond with a resultant reduction in the diameter of the wire.

wire dress—Arranging of wires or conductors in preparation for a mechanical hookup.

wire gage—Also called American Wire Gage (AWG), and formerly Brown and Sharpe (B&S) Gage. A system of numerical designations of wire sizes, starting with 0000 as the largest size and going to 000, 00, 1, 2, and beyond for the smaller sizes.

wire grating—An arrangement of wires set into a waveguide to pass one or more waves while obstructing all others.

wire-guided—In missile terminology, guided by electrical impulses sent over a closed wire circuit between the guidance point and the missile.

wireless—1. A British term for radio. 2. Used in the United States, in the sense of (1) above, when the word "radio" might be misinterpreted (e.g., wireless record player).

wireless device—Any apparatus (e.g., a wireless record player) that generates a radio-frequency electromagnetic field for operating associated apparatus not physically connected and at a distance in feet not greater than 157,000 divided by the frequency in kilohertz. Legally, the total electromagnetic field produced at the maximum operating distance cannot exceed 15 microvolts per meter.

wireless record player—*See* Wireless Device.

wire-link telemetry—Also called hard-wire telemetry. Telemetry in which a hard-wire link is used as the transmision path, no radio link being used.

wirephoto—Transmission of a photograph or other single image over a telegraph system. The image is scanned into elemental areas in orderly sequence, and each area is converted into proportional electric signals which are transmitted in sequence and reassembled in correct order at the receiver.

wire printer—A high-speed printer that prints characterlike configurations of dots through the proper selection of wire ends from a matrix of wire ends, rather than conventional characters through the selection of type faces.

wire recorder—A magnetic recorder in which the recording medium is a round stainless-steel wire about 0.004 inch in diameter rather than magnetic tape.

wire recording — A recording method in which the medium is a thin stainless-steel wire (instead of a tape or disc).

wiresonde—An atmospheric sounding instrument which is supported by a captive balloon and used to obtain temperature and humidity data from ground level to a height of a few thousand feet. Height is determined by means of a sensitive altimeter, or from the amount of cable released and the angle which the cable makes with the ground. The information is telemetered to the ground station through a wire cable.

wire splice—An electrically sound and mechanically strong junction of two or more conductors.

wireways—Sheet-metal troughs with hinged covers for housing and protecting electrical conductors and cable, and in which conductors are laid in place after the wireway has been installed as a complete system.

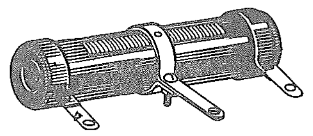

Wirewound resistor (adjustable).

wirewound resistor—A resistor in which the resistance element is a length of high-resistance wire or ribbon wound onto an insulating form.

wire wrap—*See* Solderless Wrap.

wire wrapping—A technique for terminating conductors.

wiring connector—A device for joining one or more wires together.

wiring diagram—A drawing that shows electrical equipment and/or components, together with all interconnecting wiring.

wobble—*See* Flicker.

wobble stick—A rod extending from a pendant station to operate the "stop" contacts; it will function when pushed in any direction.

wobbulator—More commonly called a sweep generator. A signal generator, the frequency of which is varied automatically and periodically over a definite range. It is used, together with a cathode-ray tube, for testing frequency response. One form consists of a motor-driven variable capacitor, which is used to vary the output frequency of a signal generator periodically between two limits.

woofer—A speaker designed to reproduce the bass frequencies. It is usually used with a crossover network and a tweeter.

word—A group of characters occupying one storage location in a computer. It is treated by the computer circuits as an entity, by the control unit as an instruction, and by the arithmetic unit as a quantity.

word length—1. The number of bits or characters handled as a physical unit by a computer or other device. 2. The size of a field.

word pattern—The smallest meaningful language unit recognized by a machine. It is usually composed of a group of syllables and/or words.

word rate—The frequency derived from the elapsed period between the beginning of the transmission of one word and the beginning of transmission of the next word.

word size — In computer terminology, the number of decimal or binary bits comprising a word.

word time—*See* Minor Cycle.

work—The magnitude of a force times the distance through which that force is applied. (*Also see* Load.)

work coil—*See* Load Coil.

work function — 1. The minimum energy (commonly expressed in electron volts) required to remove an electron from the Fermi level of a material and send it into field-free space. 2. A general term applied to the energy required to transfer electrons or other particles from the interior of one medium, across a boundary, into an adjacent medium.

work hardening — The strengthening of a crystalline material resulting from plastic deformation.

working memory—*See* Working Storage.

working ***Q***—*See* Loaded *Q*.

working storage—Also called the working memory. In a computer storage (internal), a portion reserved by the program for the data upon which the operations are being performed.

working voltage—1. *See* Voltage Rating. 2. The recommended maximum voltage of operation for an insulated conductor. It is usually set at approximately one-third of the breakdown voltage. 3. The voltage rating of a fixed capacitor. It is the recommended maximum voltage at which the unit should be operated.

worm—A gear with teeth in the form of screw threads.

worst-case circuit analysis—A type of circuit analysis used to determine the worst possible effect on the output parameters due to changes in the values of circuit elements. The circuit elements are set to the values within their anticipated ranges that produce the maximum detrimental changes in the output.

worst-case design—An extremely conservative design approach in which the circuit is designed to function normally, even though all component values have simultaneously assumed the worst possible condition that can be caused by initial tolerance, aging, etc. Worst-case techniques are also applied to obtain conservative derating of transient and speed specifications.

worst-case noise pattern—Maximum noise appearing when half of the half-selected cores are in a 1 state and the other half are in a 0 state. Sometimes called checkerboard or double-checkerboard pattern.

wound-rotor induction motor — An induction motor in which the secondary circuit consists of a polyphase winding or coils with either short-circuited terminals or ones closed through suitable circuits.

wow—Distortion caused in sound reproduction by variations in speed of the turntable or tape. (*Also see* Flutter.)

wow meter—An instrument that indicates the instantaneous speed variation of a turntable or similar equipment.

wrap—1. One winding of ferromagnetic tape. 2. The length of the path of a magnetic recording tape along which the tape and head are in intimate physical contact. It is sometimes measured as the angle of arrival and departure of the tape with respect to the head.

wrap-around—The amount of curvature exhibited by the magnetic tape or film in passing over the pole pieces of the magnetic heads.

wrapper—An insulating barrier applied to a coil by wrapping a sheet of insulating material around the coil periphery so as to form an integral part of the coil.

wrapping—A method of applying insulation to wire by serving insulating tapes around the conductor.

wratten filter—An optical filter used for filtering a given band of light. It is used in film-recorder optical systems when recording directly on color film (direct positive). It is also extensively used in photography.

wrinkle finish—An exterior paint that dries to a wrinkled surface when applied to cabinets or panels.

write—In a computer: 1. To copy, usually from internal to external storage. 2. To transfer elements of information to an output medium. 3. To record information in a register, location, or other storage device or medium. 4. To establish a charge pattern corresponding to the input (charge-storage tubes). (*Also see* Read.)

write pulse—In a computer, a pulse that is used to enter information into one or more magnetic cells for storage purposes.

writing rate—The maximum speed at which the spot on a cathode-ray tube can move and still produce a satisfactory image.

writing speed—The rate of writing on successive storage elements in a charge-storage tube.

wvdc—Abbreviation for working voltage, direct current. This is the maximum safe d-c operating voltage that can be applied across the terminals of a capacitor at its maximum operating temperature.

WWV—Call letters of the radio station of the National Bureau of Standards in Washington, D. C. WWV provides radio-broadcast technical services, including time signals, standard radio and audio frequencies, and radio-propagation disturbance warnings at 2.5, 5, 10, 15, 20, 25, 30, and 35 megahertz.

WWVH—Call letters of the National Bureau of Standards' radio station at Maui, Hawaii. It broadcasts on 5, 10, and 15 MHz for many locations not served by WWW.

wye connection—Also called a star connection. A Y-shaped winding connection.

wye junction—A Y-shaped junction of waveguides.

X

X—Symbol for reactance.

X and Z demodulation—A system of color-tv demodulation in which the two reinserted 3.58-MHz subcarrier signals differ by approximately 60° rather than the usual 90°. The R-Y, B-Y, and G-Y voltages are derived from the demodulated signals, and these voltages control the three guns of the picture tube. An important advantage of this system is that receiver circuitry is simpler than that required with I and Q demodulation.

X-axis—1. The reference axis in a quartz crystal. 2. The horizontal axis in a system of rectangular coordinates.

X-band—A radio-frequency band of 5,200 to 11,000 MHz, with wavelengths of 5.77 to 2.75 cm.

X-bar—A rectangular crystal bar, usually cut from a Z-section, elongated parallel to X and with its edges parallel to X, Y, and Z.

X_C—Symbol for capacitive reactance.

X-cut crystal—A crystal cut so that its major surfaces are perpendicular to an electrical (X) axis of the original quartz crystal.

xenon—A rare gas used in some thyratron and other gas tubes.

xerographic printer—A device for printing an optical image on paper; light and dark areas are represented by electrostatically charged and uncharged areas on the paper. Powdered ink, dusted on the paper, adheres to the charged areas and is subsequently melted into the paper by the application of heat.

xerographic recording — A recording produced by xerography.

xerography—That branch of electrostatic electrophotography in which images are formed onto a photoconductive insulating medium by infrared, visible, or ultraviolet radiation. The medium is then dusted with a powder, which adheres only to the electrostatically charged image. Heat is then applied in order to fuse the powder into a permanent image.

xeroprinting—Printing by xerography.

xeroradiography—Xerography in which X- or gamma rays are used.

xistor—Abbreviation for transistor.

X_L—Symbol for inductive reactance.

xmitter — Abbreviation for transmitter. Also abbreviated trans or xmtr.

xmsn—Abbreviation for transmission.

xmtr—Abbreviation for transmitter. Also abbreviated trans or xmitter.

X-particle—A particle having the same negative charge as an electron, but a mass between that of an electron and a proton. It is produced by cosmic radiation impinging on gas molecules or actually forming a part of cosmic rays.

X-ray apparatus—An X-ray tube and its accessories, including the X-ray machine.

X-ray crystallography—Use of X-rays in studying the arrangement of the atoms in a crystal.

X-ray detecting device—A device which detects surface and volume discontinuities in solids by means of X-rays.

X-ray diffraction camera—A camera that directs a beam of X-rays into a sample of unknown material and allows the resultant diffracted rays to act on a strip of film.

X-ray diffraction pattern—The pattern produced on film exposed in an X-ray diffraction camera. It is made up of portions of circles having various spacings, depending on the material being examined.

X-ray goniometer—An instrument that determines the position of the electrical axes of a quartz crystal by reflecting X-rays from the atomic planes of the crystal.

X-rays—Also called roentgen rays. Penetrating radiation similar to light, but having much shorter wavelengths (10^{-7} to 10^{-10} cm). They are usually generated by bombarding a metal target with a stream of high-speed electrons.

X-ray spectrometer—An instrument for producing an X-ray spectrum and measuring the wavelengths of its components.

X-ray spectrum—An arrangement of a beam of X-rays in the order of wavelength.

X-ray thickness gage—A contactless thickness gage used to measure and indicate the thickness of moving cold-rolled sheet steel during the rolling process. An X-ray beam directed through the sheet is absorbed in proportion to the thickness of the material and its atomic number, and a measure of the amount of absorption gives a continuous indication of sheet thickness.

X-ray tube—A vacuum tube in which X-rays are produced by bombarding a target with high-velocity electrons accelerated by an electrostatic field.

xtal—Abbreviation for crystal.

X-wave—One of the two components into which the magnetic field of the earth divides a radio wave in the ionosphere. The other component is the ordinary, or O-, wave.

XY-cut crystal—A crystal cut so that its characteristics fall between those of an X- and a Y-cut crystal.

XY plotter—A device used in conjuction with a computer to plot coordinate prints in the form of a graph.

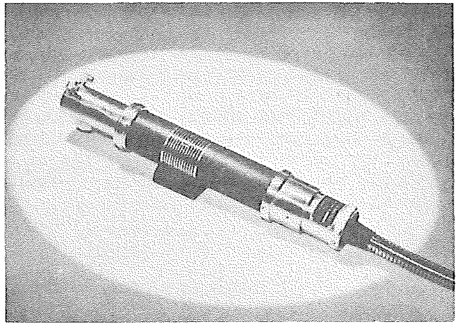

X-ray tube.

XY recorder—A recorder that traces, on a chart, the relationship between two variables, neither of which is time. Sometimes the chart moves and one of the variables is controlled so that the relationship does increase in proportion to time.

XY switch—A remote-controlled bank-and-wiper switch arranged so that the wipers move back and forth horizontally.

Y

Y—Symbol for admittance.

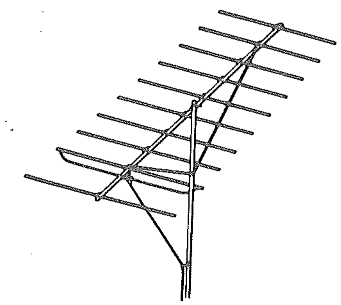

Yagi antenna.

Yagi antenna—A type of directional antenna array consisting usually of one driven half-wavelength dipole section, one parasitically excited reflector, and several parasitically excited directors.

Y antenna—*See* Delta Matched Antenna.

Y-axis — 1. A line perpendicular to two parallel faces of a quartz crystal. 2. The vertical axis in a system of rectangular coordinates.

Y-bar—A crystal bar cut in Z-sections, with its long direction parallel to Y.

Y-connected circuit — A star-connected, three-phase circuit.

Y-connection—*See* Y-network.

Y-cut crystal—A crystal cut in such a way that its major flat surfaces are perpendicular to the Y axis of the original quartz crystal.

yield map—A microcircuit or semiconductor wafer on which dots indicate those devices that failed the test criteria.

yield strength—*See* Yield Value.

yield value—Also called yield strength. The lowest stress at which a material undergoes plastic deformation. Below this stress, the material is elastic; above it, viscous.

Y-junction—A junction of waveguides in which their longitudinal axes form a Y.

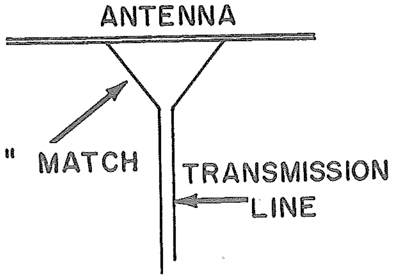

Y-match.

Y-match — Also called a delta match. A method of connecting to an unbroken dipole. The transmission line is fanned out and connected to the dipole at the points where the impedance is the same as that of the line.

Y-network—Also called a Y-connection. A star network of three branches.

yoke—A set of coils placed over the neck of a magnetically deflected cathode-ray tube to deflect the electron beam horizontally and vertically when suitable currents are passed through them.

Young's modulus—A constant that expresses the ratio of unit stress to unit deformation for all values within the proportional limit of the material.

Y-signal—A luminance transmission primary which is 1.5 to 4.2 MHz wide and equivalent to a monochrome signal. For color pictures, it contributes the finest details and brightness information. *Also see* Luminance Signal.

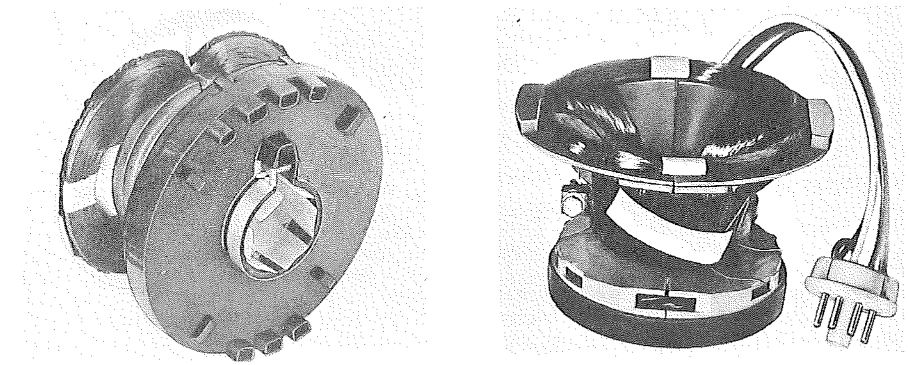

Yoke.

Z

Z—Symbol for impedance.

Zamboni pile—A primary electrochemical system capable of supplying very high electrical potentials in comparatively little space. The anode material in the pile is aluminum, and the cathode is manganese dioxide and carbon black. The electrolyte in the chemical system is aluminum chloride.

Z-angle meter—An electronic instrument for measuring impedance in ohms and phase angle in electrical degrees.

Z-axis modulation—Also called beam modulation or intensity modulation. Varying the intensity of the electron stream of a cathode-ray tube by applying a pulse or square wave to the control grid or cathode.

Z-bar—A rectangular crystal bar usually cut from X sections and elongated parallel to Z.

Zebra time—An alphabetic expression denoting Greenwich mean time.

Zeeman effect—If an electric discharge tube or other light source emitting a bright-line spectrum is placed between the poles of a powerful electromagnet, a very powerful spectroscope will show that the action of the magnetic field has split each spectrum line into three or more closely spaced but separate lines, the amount of splitting or the separation of the lines being directly proportional to the strength of the magnetic field.

zener breakdown—A breakdown caused in a semiconductor device by the field emission of charge carriers in the depletion layer.

Zener breakdown theory—A critical current density is attained by internal field emission, or tunneling, due to the high applied field.

zener diode—A two-layer device that, above a certain reverse voltage (the zener value), has a sudden rise in current. If forward-biased, the diode is an ordinary rectifier. But, when reversed-biased, the diode exhibits a typical knee, or sharp break, in its current-voltage graph. The voltage across the device remains essentially constant for any further increase of reverse current, up to the allowable dissipation rating. The zener diode is a good voltage regulator, overvoltage protector, and voltage reference.

Zener effect—A reverse-current breakdown due to the presence of a high electric field at the junction of a semiconductor or insulator.

zener impedance—*See* Breakdown Impedance.

zener voltage—*See* Breakdown Voltage.

zeppelin antenna—A horizontal antenna that is a multiple of a half-wavelength long. One end is fed by one lead of a two-wire transmission line that is also a multiple of a half-wavelength long.

zero (in a computer)—Positive binary zero is indicated by the absence of a digit or pulse in a word. In a coded-decimal computer, decimal zero and binary zero may not have the same configuration. In most computers, there are distinct representations for plus and minus zero conditions.

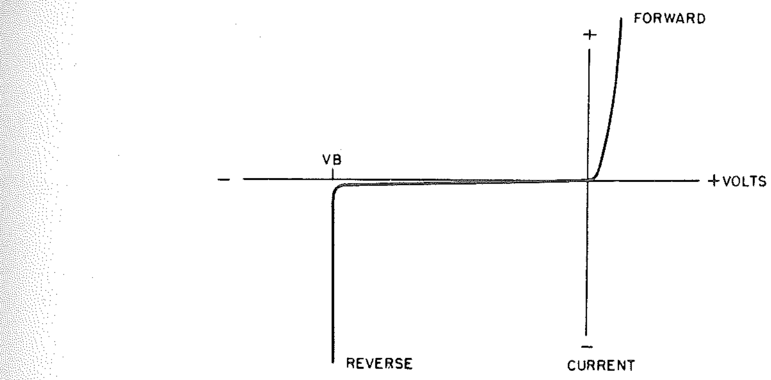

Zener diode characteristic.

zero-access storage—Computer storage for which the waiting time is negligible (e.g., flip-flop, trigger, or indicator storage).

zero-address instruction — A digital-computer instruction specifying an operation in which the locations of the operands are defined by the computer code; no explicit address is required.

zero adjuster—A device for adjusting a meter so that the pointer will rest exactly on zero when the electrical quantity is zero.

zero adjustment—1. The act of nulling out the output from a system or device. 2. The circuit or other means by which a no-output condition is obtained from an instrument when properly energized.

zero-axis symmetry—A type of symmetry in which a waveform is symmetrical about an axis and does not exhibit a net d-c component.

zero beat—The condition whereby two frequencies being mixed are exactly the same and therefore produce no beat note.

zero-beat reception—*See* Homodyne Reception.

zero bias—1. The absence of a potential difference between the control grid and the cathode. 2. When the received teletypewriter signal is equal to the transmitted signal (neither longer nor shorter), the circuit is said to have zero bias.

zero-bias tube—A vacuum tube designed to be operated as a class-B amplifier with no negative bias applied to its control grid.

zero compensation—A method by which, in certain transducers, the effects of temperature on the output at zero measurand may be minimized and maintained within known limits.

zero compression—In computers, any of several techniques used to eliminate the storage of non-significant leading zeros.

zero-cut crystal—A quartz crystal cut in such a direction that its temperature coefficient with respect to the frequency is essentially zero.

zero drift—*See* Zero Shift.

zero elimination—In a computer, the editing or deleting of nonsignificant zeros appearing to the left of the integral part of a quantity.

zero error—The delay time occurring within the transmitter and receiver circuits of a radar system. For accurate range data, this delay time must be compensated for when the range unit is calibrated.

zero-field emission — Thermionic emission from a hot conductor which is surrounded by a region of uniform electric potential.

zero-gravity switch—Also called weightlessness switch. A switch that closes when weightlessness or zero gravity is approached.

zero-input terminal—Also called a reset terminal. The terminal which, when triggered, will put a flip-flop in the zero (starting) condition—unless the flip-flop is already in a zero condition, in which event it will not change.

zero level—A reference level for comparing sound or signal intensities. In audio-frequency work, it is usually a power of .006 watt; and in sound, the threshold of hearing.

zero method—*See* Null Method.

zero-modulation noise—The noise arising when reproducing an erased tape with the erase and record heads energized as they would be in normal operation, but with zero input signal (usually 3 to 4 db higher than the bulk-erased noise).

zero-output terminal—The terminal which produces an output (of the correct polarity to trigger a following circuit) when a flip-flop is in the zero condition.

zero pole—1. A reference point for an open-wire poleline. 2. The dead-end pole at the origin of the line. 3. The lowest-numbered pole.

zero phase-sequence relay—A relay which functions in conformance with the zero phase-sequence component of the current, voltage, or power of the circuit.

zero potential—The potential of the earth, taken as a convenient reference for comparison.

zero-power resistance—In a thermistor, the resistance at a specified temperature when the electrical power dissipation is zero.

zero-power resistance-temperature characteristic—In a thermistor, the function relating the zero-power resistance and body temperature.

zero-power temperature coefficient of resistance — In a thermistor, the ratio, at a specified temperature, of the rate of change with temperature of zero-power resistance to the zero-power resistance.

zero set—1. A permanent change in the output of a device at zero measurand due to any cause. 2. A control on a vtvm to set the pointer to zero. 3. A control for adjusting a range counter to give the correct range.

zero shift—Also called zero drift. The amount by which the zero or minimum reading of an instrument deviates from the calibrated point as a result of aging or the application of an external condition to the instrument.

zeros of a network function—Those values of p (real or complex) for which the network function is zero.

zero stability—The ability of an instrument to withstand effects which might cause zero shift. Usually expressed as a percentage of full scale.

zero state—In a magnetic cell, the state wherein the magnetic flux through a specified cross-sectional area has a negative value, from an arbitrarily specified direction. (*Also see* One State.)

zero-subcarrier chromaticity—The chromaticity normally displayed when the subcarrier amplitude is zero in a color television system.

zero suppression — 1. In a recording system, the injection of a controllable voltage to balance out the steady-state component of the input signal. 2. In a computer, the elimination of zeros to the left of the significant integral part of a quantity.

zero time reference—In a radar, the time reference of the schedule of events during one cycle of operation.

Z factor—In thermoelectricity, an accepted figure of merit which denotes the quality of the material.

zigzag reflections—From a layer of the ionosphere, high-order multiple reflections which may be of abnormal intensity. They occur in waves which travel by multihop ionosphere reflections and finally turn back toward their starting point by repeated reflections from a slightly curved or sloping portion of an ionized layer.

zinc—A bluish-white metal which, in its pure form, is used in dry cells.

Z marker—Also called a zone marker. A marker beacon that radiates vertically and is used for defining a zone above a radio range station.

Z_o—Symbol for characteristic impedance—i.e., the ratio of the voltage to the current at every point along a transmission line on which there are no standing waves.

zone — 1. Any of the three top positions (12, 11 or 0) on a punch card. 2. A part of internal computer storage allocated for a particular purpose.

zone bits—1. The two leftmost binary digits in a digital computer in which six binary digits are used for characters and the four rightmost are used for decimal digits. 2. Any bit in a group of bit positions that are used to indicate a specific class of items (e.g., numbers, letters, special signs, and commands).

zone blanking—A method of turning off the cathode-ray tube during part of the sweep of the antenna.

zone leveling—In semiconductor processing, the passage of one or more molten zones along a semiconductor body, for the purpose of uniformly distributing impurities throughout the material.

zone marker—*See* Z-marker.

zone of silence—An area—between the points at which the ground wave becomes too weak to be detected and the sky wave first returns to earth—where normal radio signals cannot be heard. *Also see* Skip Zone.

zone plate antenna—A rapid-scanning antenna in which the reflector consists of arcs of confocal parabolas with focal lengths which differ by T/λ and which lie close to a circle centered at the focus, where λ is the wavelength.

zone-position indicator—An auxiliary radar set for indicating the general position of an object to another radar set with a narrower field.

zone punch—*See* Overpunch.

zone purification—In semiconductor processing, the passage of one or more molten zones along a semiconductor to reduce the impurity concentration of part of the ingot.

zoning—1. Also called stepping. Displacement of the various portions of the lens or surface of a microwave reflector, so that the resulting phase front in the near field remains unchanged. 2. Purifying a metal by passing it through an induction coil; the impurities are swept ahead of the heating effect. Specifically used in purifying semiconductor crystals.

zoom—To enlarge or reduce on a continuously variable basis the size of a televised image. It may be done electronically or optically.

zoom lens—An optical lens with some elements made movable so that the focal length or angle of view can be adjusted continuously without losing the focus.

PRONUNCIATION GUIDE

Here is an alphabetical listing of more than 1,100 often-mispronounced electronic terms taken from the dictionary proper. Turn to this section whenever you are in doubt about the pronunciation of a word.

You will note the use of accents, and also of diacritical marks over the vowels. For simplicity, no secondary accents are shown—only primary accents. The normal inflection of your voice will automatically stress the secondary accents, once you know where to place the primary accents. For example, the word "ther-mo-e-lec-TRIC-i-ty" would normally have a secondary accent on "ther," but you'll notice that your inflection tends to put it there.

The following pronunciation key, with diacritical markings, shows how the vowels are pronounced.

ā as in ace
ȧ as in chaotic
ä as in arm
ă as in map
â as in dare
ē as in me
ė as in event
ĕ as in met

ê as in here
ẽ as in her
ī as in ice
ĭ as in it
ō as in note
ȯ as in obey
ŏ as in not
ô as in or
oi as in oil

o͞o as in moon
o͝o as in foot
ou as in how
ū as in huge
u̇ as in unite
ŭ as in cup
û as in urn
ü as in blue

At first glance it may seem difficult to distinguish between ā and ȧ, ē and ė, ō and ȯ, or ū and u̇. The first are called long vowels, and the second, half-long vowels. Just remember that their use is relative—for example, in the word "photograph," the first *o* is a bit longer than the second. The slight differentiation is difficult to hear, but it is there.

Also remember that unaccented syllables are not given the full brunt of the pronunciation. For example, in speech we do not say "PLAT-IN-NUM"—we say "PLAT-i-num" and tend to "swallow" the *i* (although it can be heard). By the same token, the word "photography" is pronounced "fo-TOG-ruf-e," not "FO-TOG-RAFF-FEE."

By keeping these few pointers in mind, you will find the "Pronunciation Guide" an easy—as well as a handy—tool.

A

a′bac (ă′bäc)
ab′am•pere (ăb′ăm•pẽr)
ab•cou′lomb (ăb•ko͞o′lȯm)
ab•er•ra′tion (ăb•ẽr•ā′shŭn)
ab′far•ad (ăb′făr•ăd)
ab′hen•ry (ăb′hĕn•rē)
ab′mho (ăb′mō)
ab′ohm (ăb′ōm)
a•bort′ (ä•bôrt′)
ab•scis′sa (ăb•sĭs′ä)
ab′volt (ăb′vōlt)
ac•cel•er•om′e•ter (ăk•sĕl•ẽr•ŏm′ĕ•tẽr)
ac•cen•tu•a′tion (ăk•sĕn•tū•ā′shŭn)
ach•ro•mat′ic (ăk•rō•măt′ĭk)
a•clin′ic (ă•klĭn′ĭk)
a•cous′tic (ä•ko͞os′tĭk)
a•cous•to•e•lec′tric (ä•ko͞os•tō•ė-lĕk′trĭk)
ac•tin′ic (ăk•tĭn′ĭk)
ac•tin′i•um (ăk•tĭn′ĭ•ŭm)
ac′tu•a•tor (ăk′tū•ā•tẽr)
ad′dend (ăd′ĕnd)
ad′di•tron (ă′dĭ•trŏn)
ae′o•light (ȧ′ō•līt)
aer′i•al ((âr′ĭ•ŭl)
aer•o•dy•nam′ics (âr•ō•dī•năm′ĭks)
aer′o•phare (âr′ō•fâr)
a•gon′ic (ă•gŏn′ĭk)
a•lac′ri•tized (ă•lăk′rĭ•tīzd)
al•ge•bra′ic (ăl•jĕ•brā′ĭk)
al′gol (ăl′gŏl)
al′go•rithm (ăl′gȯ•rĭth′m)
al′ka•line (ăl′kä•līn)
al•lo•ca′tion (ăl•ō•kā′shŭn)
al′ni•co (ăl′nĭ•kō)
al′pha (ăl′fä)
al•tim′e•ter (ăl•tĭm′ĕ•tẽr)
al•to-trop′o•sphere (ăl•tō-trŏp′ȯ•sfėr)
a•lu′mi•nized (ä•lū′mĭ•nīzd)
am′bi•ent (ăm′bĭ•ĕnt)
am′me•ter (ăm′mē•tẽr)
am′pere (ăm′pẽr)
am′pli•dyne (ăm′plĭ•dīn)
am′pli•stat (ăm′plĭ•stăt)

an·a·cous'tic (ăn·ä·ko͞os'tĭc)
an·as'tig·mat (ăn·ăs'tĭg·măt)
an'chor (ăng'kẽr)
an·cil'lar·y (ăn·sĭ'lẽr·ē)
an·e·cho'ic (ăn·ĕ·kȯ'ĭk)
an·e·mom'e·ter (ăn·ĕ·mŏm'ĕ·tẽr)
an'i·on (ăn'ī·ŏn)
an·i·so·trop'ic (ăn·ī·sō·trŏp'ĭk)
an·i·sot'ro·py (ăn·ī·sŏt'rȯ·pē)
an·nealed' (ă·nēld')
an'nu·lar (ăn'ū·lẽr)
an·nun·ci·a'tion (ă·nŭn·sĭ·ā'shŭn)
an·nun'ci·a·tor (ă·nŭn'sĭ·ā·tẽr)
an'o·diz·ing (ăn'ō·dīz·ĭng)
a·nom'a·lous (ä·nŏm'ä·lŭs)
an·ti·air'craft (ăn·tĭ·âr'krăft)
an·ti·ca·pac'i·tance (ăn·tĭ·kä·pă'sĭ-
  tŭns)
an·ti·co·in'ci·dence (ăn·tĭ·kō·ĭn'sĭ-
  dĕns)
an·ti·fer·ro·e·lec·tric'i·ty (ăn·tĭ·fĕrō·ė-
  lĕk·trĭ'sĭ·tē)
an·ti·fer·ro·mag·net'ic (ăn·tĭ·fĕr·ō-
  măg·nĕt'ĭk)
an·ti·log'a·rithm (ăn·tĭ·lŏg'ä·rĭth'm)
an·ti·mi·cro·phon'ic (ăn·tĭ·mī·krō-
  fŏn'ĭk)
an'ti·nodes (ăn'tĭ·nōds)
an·ti·pro'ton (ăn·tĭ·prō'tŏn)
an·ti·side'tone (ăn·tĭ·sīd'tōn)
a·pe·ri·od'ic (ā·pė·rĭ·ŏd'ĭk)
a'per·ture (ă'pẽr·chẽr)
ap·pli·que' (ăp·lĭ·kā')
a·qua·dag (ăk'wä·dăg)
a·rith·met'ic [*adj.*] (ă·rĭth·mĕt'ĭk)
a·rith'me·tic [*noun*] (ä·rĭth'mĕ·tĭk)
as·bes'tos (ăs·bĕs'tŭs)
as'pect ra'tio (ăs'pĕkt rā'shō)
as·per'i·ties (ăs·pĕr'ĭ·tēs)
a·spher'ic (ā·sfêr'ĭc)
a·sym·met'ric (ă·sĭ·mĕt'rĭk)
a'sta·ble (ȧ'stā'b'l)
a·stig'ma·tism (ä·stĭg'mä·tĭzm)
a'symp·tote (ă'sĭm·tōt)
a·symp·tot'ic (ă·sĭm·tŏt'ĭk)
a·syn'chro·nous (ă·sĭng'krō·nŭs)
at'mos·phere (ăt'môs·fėr)
at·mos·pher'ic (ăt·môs·fėr'ĭk)
at·ten·u·a'tion (ă·tĕn·ū·ā'shŭn)
at'to (ăt'tō)
au·di·om'e·ter (ô·dĭ·ŏm'ĕ·tẽr)
au'di·on (ô'dĭ·ŏn)
au'di·o·phile (ô'dė·ō·fīl)
au'gend (ô'jĕnd)
au'ral (ô'rŭl)
au·ro'ra (ô·rô'rä)
au·ro'ral (ô·rô'răl)
au'to·dyne (ô'tō·dīn)
au·to·ma'tion (ô·tō·mā'shŭn)
au·to·ra·di·og'ra·phy (ô·tō·rā·dĭ-
  ŏg'rä·fē)
au·to·reg·u·la'tion (ô·tō·rĕg·ū·lā'
  shŭn)
Au'to·syn (ô'tō·sĭn)
au·to·trans'form·er (ô·tō·trăns'fôr-
  mẽr)
aux·il'ia·ry (ôg·zĭl'yû·rē)
a'va·lanche (ăv'ä·lănch)
a·vi·a'tion (ā·vĭ·ā'shŭn)
a·vi·on'ics (ā·vē·ŏn'ĭks)
A·vo·ga'dro's (ă·vȯ·gä'drōz)
ax'i·al (ăk'sĭ·ŭl)
ax'is (ăk'sĭs)
Ayr'ton-Per'ry (âr'tŭn) (pĕr'ē)

B

bab'ble (băb''l)
baf'fle (băf''l)
Ba'ke·lite (bā'kä·līt)
bal·lis'tic (bä·lĭs'tĭk)
bal'un (băl'ŭn)
ban'tam (băn'tŭm)
bar'i·um (bâr'ĭ·ŭm)
Bark'haus·en-Kurz (bärk'houz'n)
  (kẽrz)
ba·rom'e·ter (bä·rŏm'ĕ·tẽr)
bar'ret·ter (bâr'ĕt·ẽr)
bar'ri·er (bâr'ĭ·ẽr)
bass [*music*] (bās)
bath·y·con·duc'tor·graph (băth·ĭ·
  kŏn- dŭk'tẽr·grăf)
bath·y·ther'mo·graph (băth·ĭ·thẽr'
  mȯ- gráf)
baud (bôd)
bay·o·net' (bā·ȯ·nĕt')
ba·zoo'ka (bä·zo͞o'kä)
bel (bĕl)
Bel·li'ni-To'si (bĕl·lē'nė) (tō'sė)
Be·ni'to (bĕ·nē'tȯ)
be'ta (bā'tä)
be'ta·tron (bā'tä·trŏn)
bev'a·tron (bĕv'ä·trŏn)
Bev'er·age (bĕv'ẽr·ĕj)
bez'el (bĕz' 'l)
bi'as (bī'ŭs)
bi·con'i·cal (bī·kŏn'ĭ·k'l)
bi·di·rec'tion·al (bī·dĭ·rĕk'shŭn·'l)
bi·fi'lar (bī·fī'lẽr)
bi·lat'er·al (bī·lăt'ẽr·ôl)
bi·me·tal'lic (bī·mĕ·tăl'ĭk)
bi'na·ry (bī'nâ·rē)
bin·au'ral (bīn·ô'rôl)
bi·nis'tor (bī·nĭs'tẽr)
bi·no'mi·al (bī·nō'mĭ·ôl)
bi·o·met'rics (bī·ȯ·mĕt'rĭks)
bi·on'ics (bī·ŏn'ĭks)
bi·qui'na·ry (bī·kwī'nâ·rē)
bis'cuit (bĭs'kĭt)
bi·stab'le (bī·stā'b'l)
blip (blĭp)

bli'vet (blĭ'vĕt)
Bloch wall (blŏk) (wôl)
bof'fle (bŏf''l)
bo'gey (bō'gē)
bo·lom'e·ter (bȯ·lŏm'ĕ·tẽr)
Bool'e·an (bo͞ol'ė·ŭn)
byte (bīt)

C

cad'mi·um (kăd'mē·ŭm)
cal'i·brate (kăl'ī·brāt)
cal'o·mel (kăl'ȯ·mĕl)
cal·o·rim'e·ter (kăl·ô·rĭm'ĕ·tẽr)
ca·nal' ray (kä·năl') (rā)
can·do·lu·mi·nes'cence (kăn·dȯ·lo͞o-
  mĭ·nĕs'ĕns)
ca·pac'i·tance (kä·pă'sĭ·tŭns)
cap'stan (kăp'stăn)
Car·bo·run'dum (kär·bȯ·rŭn'dŭm)
car·ci'no·tron (kär·sĭ'nȯ·trŏn)
Car'dew (kär'do͞o)
car'di·oid (kär'dĭ·oid)
Car·not' (kär·nō')
cas·cade' (kăs·kād')
cas'code (kăs'kōd)
cat·a·stroph'ic (kăt·ä·strŏf'ĭk)
ca·te'na (kă·tē'nä)
cath·am'pli·fi·er (kăth·ămp'lĭ·fī·ẽr)
cath·o·do·lu·mi·nes'cence (kăth·ȯ·du·
  lo͞o·mĭ·nĕs'ĕns)
cat'i·on (kăt'ī·ŏn)
cav·i·ta'tion (kăv·ĭ·tā'shŭn)
cel'lu·lose (sĕl'u̇·lōs)
Cel'si·us (sĕl'sĭ·ŭs)
cen'ti· (sĕn'tĭ)
cen·trif'u·gal (sĕn·trĭf'u̇·gôl)
cen·trip'e·tal (sĕn·trĭp'ĕ·tôl)
ce·ram'ic (sẽ·răm'ĭk)
Cer'en·kov (sĕr'ĕn·kôf)
ce'si·um (sē'zė·ŭm)
chad (chăd)
chem·i·sorp'tion (kĕm·ĭ·sôrp'shŭn)
chro'ma (krō'mä)
chro·mat'ic (krō·măt'ĭk)
chro·ma·tic'i·ty (krō·mä·tĭ'sĭ·tē)
chro'mi·nance (krō'mĭ·năns)
chron'o·graph (krŏn'ȯ·grăf)
chron·om'e·ter (krŏn·ŏm'ĕ·tẽr)
chron'o·scope (krŏn'ȯ·skōp)
cir'clo·tron (sûr'klȯ·trŏn)
cla'vi·er (klă'vĭ·ẽr)
co·ax'i·al (kō·ăk'sĭ·ôl)
co'-chan·nel (kō'-chăn 'l)
co'dan (kō'dăn)
co·ef·fi'cient (kō·ĕ·fĭsh'ĕnt)
co·er'cive (kȯ·ẽr'sĭv)
co·er·civ'i·ty (kō·ẽr·sĭv'ĭ·tē)
cog'ging (kŏg'ĭng)
co·her'ent (kȯ·hĭr'ĕnt)
co·her'er (kȯ·hĭr'ẽr)
co'li·dar (kō'lĭ·där)

col·late' (kō·lāt')
col·li·ma'tion (kŏl·ĭ·mā'shŭn)
col'li·ma·tor (kŏl'ĭ·mā·tẽr)
col·lin'e·ar (kō·lĭn'ē·ẽr)
col·or·i·met'ric (kŭl·ẽr·ĭ·mĕt'rĭk)
col·or·i'me·try (kŭl·ẽr·ĭ'mĕ·trē)
Col'pitts (kōl'pĭts)
col'umn (kŏl'ŭm)
com'mu·ta·tor (kŏm'u̇·tā·tẽr)
com·pan'der (kŏm·păn'dẽr)
com·par'a·tor (kŏm·păr'ĕ·tẽr)
com·pat·i·bil'i·ty (kŏm·păt·ĕ·bĭl'ĭ·tē)
com·po'nent (kŏm·pō'nĕnt)
com·pos'ite (kŏm·pŏz'ĭt)
con·cat'e·nate (kŏn·kăt''n·āt)
con·cen'tric (kŏn·sĕn'trĭk)
Con'dor (kŏn'dôr)
con'duit (kŏn'do͞o·ĭt)
con'el·rad (kŏn'äl·răd)
con'i·cal (kŏn'ĭ·käl)
con'ju·gate [*adj.*] (kŏn'jo͝o·gȧt)
con'o·scope (kŏn'ŏ·skōp)
con'sol (kŏn'sŭl)
con'sole (kŏn'sōl)
con'so·nance (kŏn'sȯ·nŭns)
Cor·bi'no effect (kôr·bē'nȯ)
co·se'cant (kō·sē'kănt)
cou'lomb (ko͞o'lōm)
cou·lom'e·ter (ko͞o·lŏm'ĕ·tẽr)
Coul'ter (kōl'tẽr)
co·va'lent (kō·vā'lĕnt)
Crookes (kro͝oks)
cry·o·gen'ics (krī·ȯ·jĕn'iks)
cry'o·stat (krī'ȯ·stăt)
cry'o·tron (krī'ȯ·trŏn)
cu'rie (kû'rē)
Cur·pis'tor (kûr·pĭs'tẽr)
cy·ber·net'ics (sī·bẽr·nĕt'ĭks)
cy'clic (sī'klĭk)
cy'cli·cal·ly (sī'klĭ·kô·lē)
cy'clo·gram (sī'klȯ·grăm)
cy'clo·tron (sī'klȯ·trŏn)
cy·lin'dri·cal (sĭ·lĭn'drĭ·kôl)
Czo·chral'ski (zō·krôl'skē)

D

Da'mon (dā'mŭn)
Dan'iell (dăn'yĕl)
dar'af (dăr'ŭf)
D'Ar·son·val' (där'sŏn·vŭl)
da'ta (dā'tä)
De·bye' (dĕ·bī')
dec'ade (dĕk'ād)
de·ca·les'cent (dē·kä·lĕs'ĕnt)
dec'ca (dĕk'ä)
de·cel'er·at·ing (dē·sĕl'ẽr·āt·ĭng)
dec'i· (dĕs'ĭ-)
dec'i·bel (dĕs'ĭ·bĕl)
dec'i·log (dĕs'ĭ·lôg)
dec·i·met'ric (dĕs·ĭ·mĕt'rĭk)
dec'i·ne·per (dĕs'ĭ·nā·pẽr)

dec•li•na′tion (dĕk•lĭ•nā′shŭn)
dec•li•nom′e•ter (dĕk•lĭ•nŏm′ĕ•tẽr)
dee (dē)
de′fect (dḗ′fĕkt)
de•gaus′sing (dē•gous′ĭng)
de•gen′er•a•cy (dḗ•jĕn′ẽr•ĕ•sē)
de•i•on•i•za′tion (dḗ•ī•ŏn•ĭ•zā′shŭn)
dek′a (dĕk′ä)
dek•a•hex•a•dec′i•mal (dĕk•ä•hĕks•ä-dĕs′ĭ•môl)
de•lim′it•er (dḗ•lĭm′ĭ•tẽr)
Del′lin•ger (dĕl′ĭn•jẽr)
de•mag•net•i•za′tion (dē•măg•nĕt-ĭ•zā′shŭn)
Dem′ber (dĕm′bẽr)
de•mod•u•la′tion (dē•mŏd•ū́•lā′shŭn)
den′a•ry (dĕn′â•rē)
den′drite (dĕn′drīt)
den•drit′ic (dĕn•drĭt′ĭc)
den•si•tom′e•ter (dĕn•sĭ•tŏm′ĕ•tẽr)
de•ple′tion (dḗ•plē′shŭn)
de•po•lar•i•za′tion (dē•pō•lär•ĭ•zā′shŭn)
de•riv′a•tive (dḗ•rĭv′ä•tĭv)
Des′triau (dĕs′trĭô)
de•tec′to•phone (dē•tĕk′tṓ•fōn)
de′tent (dḗ′tĕnt)
deu•te′ri•um (dū́•târ′ḗ•ŭm)
deu′ter•on (dū′tẽr•ŏn)
de•vi•a′tion (dē•vē•ā′shŭn)
Dew′ar (dū′ẽr)
di•ag•nos′tic (dī•ăg•nŏs′tĭk)
di•ag′no•tor (dī́•ăg′nṓ•tẽr)
di•a•mag•net′ic (dī•ä•măg•nĕt′ĭk)
di•a•pa′son (dī•ä•pā′z’n)
di′a•phragm (dī′ä•frăm)
di′a•ther•my (dī′ä•thẽr•mē)
di•chro′ic (dī•krō′ĭk)
di′chro•ism (dī′krṓ•ĭzm)
di•e•lec′tric (dī•ḗ•lĕk′trĭk)
dig′i•ralt (dĭj′ĭ•rălt)
dig′i•tal (dĭj′ĭ•tôl)
di•hep′tal (dī•hĕp′tôl)
di•op′ter (dī•ŏp′tẽr)
di′plex•er (dī′plĕks•ẽr)
di•rec•tiv′i•ty (dē•rĕk•tĭv′ĭ•tē)
dis′cone (dĭs′kōn)
dis•crete′ (dĭs•krēt′)
dis•crim′i•na•tor (dĭs•krĭm′ĭ•nā•tẽr)
dis•per′sion (dĭs•pẽr′zhŭn)
dis•sec′tor (dĭ•sĕk′tẽr)
dis•si•pa′tion (dĭs•ĭ•pā′shŭn)
dis•so•ci•a′tion (dĭ•sō•sĭ•ā′shŭn)
dis′so•nance (dĭs′ṓ•nŭns)
dis•sym•met′ri•cal (dĭ•sĭ•mĕt′rĭ•kôl)
di•verg′ing (dī•vẽrj′ĭng)
di•ver′si•ty (dĭ•vẽr′sĭ•tē)
Doh′er•ty (dō′ẽr•tē)
do′nor (dō′nẽr)
do′pant (dō′pănt)
Dop′pler (dŏp′lẽr)
dos′age (dōs′âj)

do•sim′e•ter (dŏ•sĭm′ĕ•tẽr)
dou′blet (dŭb′lĕt)
drunk•om′e•ter (drŭnk•ŏm′ĕ•tẽr)
du•o•dec′al (dōō•ō•dĕk′ôl)
du•o•di′ode (dōō•ō•dī′ōd)
du•o•lat′er•al (dōō•ō•lăt′ẽr•ôl)
du′o•pole (dōō′ṓ•pōl)
dy′na•quad (dī′nä•kwäd)
dy′na•mo (dī′nä•mō)
dy•na•mo•e•lec′tric (dī•nä•mō•ḗ-lĕk′trĭk)
dy•na•mom′e•ter (dī•nä•mŏm′ĕ•tẽr)
dy′na•mo•tor (dī′nä•mō•tẽr)
dy′na•tron (dī′nä•trŏn)
dyne (dīn)
dy•nis′tor (dī•nĭs′tẽr)
dy′node (dī′nōd)

E

eb•i•con•duc•tiv′i•ty (ĕb•ĭ•cŏn•dŭk-tĭv′•ĭ•tē)
ec•cen′tric (ĕk•sĕn′trĭk)
ec•cen•tric′i•ty (ĕk•sĕn•trĭs′ĭ•tē)
Ec′cles-Jor′dan (ĕk′ls) (jôr′d’n)
ech′e•lon (ĕsh′ĕ•lŏn)
Eint′ho•ven (īnt′hō•vĕn)
e•las′tance (ḗ•lăs′tŭns)
e•las•tiv′i•ty (ē•lăs•tĭv′ĭ•tē)
e•lec′tra (ḗ•lĕk′trä)
e•lec′tral•loy (ḗ•lĕk′trä•loi)
e•lec′tret (ḗ•lĕk′trĕt)
e•lec•tro•a•cous′tic (ḗ•lĕk•trō•ă-kōōs′tĭk)
e•lec•tro•a•nal′y•sis (ḗ•lĕk•trō•ä-năl′ĭ•sĭs)
e•lec•tro•bi•os′co•py (ḗ•lĕk•trṓ•bī-ŏs′kŏ•pē)
e•lec•tro•car′di•o•gram (ḗ•lĕk•trō-kär′dĭ•ṓ•grăm)
e•lec•tro•co•ag•u•la′tion (ḗ•lĕk•trō-kō•ăg•ū́•lā′shŭn)
e•lec•tro•cu′tion (ḗ•lĕk•trō•kū′shŭn)
e•lec•tro•dep•o•si′tion (ḗ•lĕk•trō-dĕp•ṓ•zĭsh′ŭn)
e•lec•tro•di•a•lyt′ic (ḗ•lĕc•trō•dī•ă-lĭt′ĭk)
e•lec•tro•dy•nam′ic (ḗ•lĕk•trō•dī-năm′ĭk)
e•lec•tro•dy•na•mom′e•ter (ḗ•lĕk-trō•dī•nä•mŏm′ĕ•tẽr)
e•lec•tro•en•ceph′a•lo•graph (ḗ•lĕk-trō•ĕn•sĕf′ä•lṓ•grăf)
e•lec•tro•graph′ic (ḗ•lĕk•trō•grăf′ĭk)
e•lec•tro•ki•net′ics (ḗ•lĕk•trō•kĭ-nĕt′ĭks)
e•lec•tro•lu•mi•nes′cence (ḗ•lĕk•trō-lū•mĭ•nĕs′ ’ns)
e•lec•trol′y•sis (ḗ•lĕk•trŏl′ĭ•sĭs)
e•lec′tro•lyte (ḗ•lĕk′trō•līt)
e•lec•tro•lyt′ic (ḗ•lĕk•trō•lĭt′ĭk)

e•lec•tro•met′al•lur•gy (ḗ•lĕk•trō-mĕt′’l•ûr•jē)
e•lec•trom′e•ter (ḗ•lĕk•trŏm′ĕ•tẽr)
e•lec•tro•my•og′ra•phy (ḗ•lĕk•trō•mī-ŏg′rä•fē)
e•lec•troph′o•rus (ḗ•lĕk•trŏf′ô•rŭs)
e•lec•tro•pho•tog′ra•phy (ḗ•lĕk•trō-fṓ•tŏg′rä•fē)
e•lec•tro•sta•tog′ra•phy (ḗ•lĕk•trō•stă-tŏg′rä•fē)
e•lec•tro•stric′tion (ḗ•lĕk•trō-strĭk′shŭn)
e•lec•tro•ther′a•py (ḗ•lĕk•trō-thâr′ä•pē)
e•lec•tro•ther′mal (ḗ•lĕk•trō•thẽr′mŭl)
e•lec•tro•win′ning (ḗ•lĕk•trō•wĭn′ĭng)
em•bed′ment (ĕm•bĕd′mĕnt)
em′i•tron (ĕm′ĭ•trŏn)
em•pir′i•cal (ĕm•pêr′ĭ•käl)
en•cap•su•la′tion (ĕn•kăp•sŭ•lā′shŭn)
en′do•dyne (ĕn′dō•dīn)
en′tro•py (ĕn′trṓ•pē)
en•vi•ron•men′tal (ĕn•vīr•ŭn•mĕn′tôl)
e•pi•tax′i•al (ĕ•pĭ•tăks′ĭ•ôl)
e•pox′y (ḗ•pŏks′ē)
ep′si•lon (ĕp′sĭ•lŏn)
e•qui•lib′ri•um (ḗ•kwĭ•lĭb′rĭ•ŭm)
e′qui•phase (ḗ′kwĭ•fāz)
e•qui•po•ten′tial (ḗ•kwĭ•pṓ•tĕn′shŭl)
e•qui•sig′nal (ḗ•kwĭ•sĭg′nôl)
e•quiv′a•lent (ḗ•kwĭv′ä•lĕnt)
e•quiv•o•ca′tion (ḗ•kwĭv•ṓ•kā′shŭn)
E•sa′ki (ḗ•sä′kē)
es•cutch′eon (ĕs•kŭch′ŭn)
Es′ti•a•tron (ĕs′tḗ•ä•trŏn)
etch′ant (ĕch′ănt)
Et′ting•shaus•en (ĕt′ĭng•shouz ’n)
eu•re′ka (yōō•rē′kä)
ex′ci•ton (ĕks′ĭ•tŏn)
ex′ci•tron (ĕks′ĭ•trŏn)
ex′o•sphere (ĕks′ṓ•sfêr)
ex•o•ther′mic (ĕks•ṓ•ther′mĭk)
ex•po•nen′tial (ĕks•pṓ•nĕn′shôl)
ex•trin′sic (ĕks•trĭn′sĭk)

F

fa′com (fā′kŏm)
fac•sim′i•le (făk•sĭm′ĭ•lē)
Fahne′stock (fŏn′stŏk)
fa•rad′ic (fâ•răd′ĭk)
far′ad•me•ter (fâr′ăd•mē•tẽr)
fa•thom′e•ter (fă•thŏm′ĕ•tẽr)
fa•tigue′ (fä•tḗg′)
Faure (fôr)
fem′to- (fĕm′tō)
fem•to•am′pere (fĕm•tō•ăm′pẽr)
Fer′mi (fâr′mḗ)
fer′reed (fẽr′ēd)
fer•ri•mag•net′ic (fâr•ĭ•măg•nĕt′ĭk)
fer′ris•tor (fâr′ĭs•tẽr)

fer′rod (fâr′ŏd)
fer•ro•dy•nam′ic (fâr•ṓ•dī•năm′ĭk)
fer•ro•man′ga•nese (fâr•ṓ•măng′gä-nēz)
fer•rom′e•ter (fâr•ŏm′ĕ•tẽr)
fer•ro•spi•nel′ (fâr•ṓ•spĭ•nĕl′)
fer′rous (fâr′ŭs)
fer′rule (fâr′ŭl)
fil•a•men′ta•ry (fĭl•ä•mĕn′tä•rē)
fi′nite fī′nīt)
fis′sion (fĭsh′ŭn)
Fle•wel′ling (flü•wĕl′ĭng)
Flin′ders (flĭn′dẽrz)
flu•o•rem′e•ter (flōō•ŏr•rĕm′ĕ•tẽr)
flu•o•res′cence (flōō•ŏr•ĕs′äns)
flu•or•os′co•py (flōō•ŏr•ŏs′kŏ•pē)
fo•com′e•ter (fṓ•kŏm′ĕ•tẽr)
For•mi′ca (fôr•mī′kä)
FOR′TRAN (fôr′trăn)
for•tu′i•tous (fôr•tū′ĭ•tŭs)
Fou•cault′ (fōō•kō′)
Fou′rier (fōō•ryā′)
Frahm (främ)
Fraun′ho•fer (froun′hō•fẽr)
Fres•nel′ (frā́•nĕl′)
Fus′e•stat (fūz′ŭ•stăt)
Fus′e•tron ((fūz′ŭ•trŏn)

G

ga•le′na (gä•lē′nä)
gal•van′ic (găl•văn′ĭk)
gal•va•nom′e•ter (găl•vä•nŏm′ĕ•tẽr)
gam′ma (găm′ä)
gas′ton (găs′tŭn)
gauss (gous)
Gei′ger-Muel′ler (gī′gẽr) (mül′ẽr)
Geiss′ler (gīs′lẽr)
gen′e•mo•tor (jĕn′ĕ•mō•tẽr)
ge•o•des′ic (jē•ṓ•dĕs′ĭk)
ger•ma′ni•um (jẽr•mā′nĭ•ŭm)
gi′ga- (jĭ′gä)
gi′ga•cy•cle (jĭ′gä•sī•k’l)
gig′ohm (jĭg′ōm)
Gill-Mor•rell′ (gĭl) (môr•ĕl′)
gim′bal (gĭm′bôl)
gimp (gĭmp)
Gi•or′gi (jē•ôr′jḗ)
go′bo (gō′bō)
Gold′schmidt (gōld′smĭt)
go•ni•om′e•ter (gŏ•nĭ•ŏm′ĕ•tẽr)
goo′gol (gōō′gŏl)
gramme (grăm)
graph′e•chon (grăf′ĕ•kŏn)
grat′i•cule (grăt′ĭ•kūl)
Gratz (grătz)
Green′wich (grĕn′ĭch)
grom′met (grŏm′ĕt)
Gud′den-Pohl (gŏŏd′’n) (pōl)
Guil′le•min ((gĭl′ĕ•mĭn)
gut′ta-per′cha (gūt′ä-pẽr′chä)

gy·ro·fre′quen·cy (jī·rȯ·frē′kwĕn·sē)
gy·ro·scop′ic (jī·rȯ·skŏp′ĭk)

H

ha′lo (hā′lō)
hal′o·gen (hăl′ȯ·jĕn)
Heav′i·side (hĕv′ĭ·sīd)
hec′to- (hĕk′tȯ-)
Hef′ner (hĕf′nẽr)
Hei′sing (hī′sĭng)
hel′i·cal (hēl′ĭ·k'l)
he·li·on′ics (hē·lĭ·ŏn′ĭks)
he′lix (hē′lĭks)
Helm′holtz (hĕlm′hōlts)
hem·i·mor′phic (hĕm·ĭ·môr′fĭk)
her·maph·ro·dit′ic (hẽr·măf·rȯ·dĭt′ĭc)
her·met′ic (hẽr·mĕt′ĭk)
hertz (hẽrts)
Hertz′i·an (hẽrt′sĭ·ŭn)
het·er·o·ge·ne′i·ty (hĕt·ẽr·ȯ·gĕ·nē′ĭ·tē)
heu·ris′tic (hu̇·rĭs′tĭk)
hex·a·dec′i·mal (hĕks·ä·dĕs′ĭ·mäl)
HI·PER′NAS (hi·pẽr′năs)
his′to·gram (hĭs′tȯ·grăm)
ho′mo·dyne (hō′mȯ·dīn)
ho·mo·ge·ne′i·ty (hō·mō·gĕ·nē′ĭ·tē)
ho·mo·ge′ne·ous (hō·mȯ·jē′nė·ŭs)
ho·mol′o·gous (hō·mŏl′ŏ·gŭs)
ho·mo·po′lar (hō·mȯ·pō′lẽr)
ho·ri′zon (hô·rī′z'n)
hy′brid (hī′brĭd)
hy·dro·a·cous′tics (hī·drȯ·ä·k\u014d\u014ds′tĭks)
hy·drom′e·ter (hī·drŏm′ĕ·tẽr)
hy′dro·phone (hī′drȯ·fōn)
hy·grom′e·ter (hī·grŏm′ĕ·tẽr)
hy·gro·scop′ic (hī·grō·skŏp′ĭk)
hy′gro·stat (hī′grȯ·stăt)
hy·per′bo·la (hī·pẽr′bȯ·lä)
hy·per·bol′ic (hī·pẽr·bŏl′ĭk)
hy·per·son′ic (hī·pẽr·sŏn′ĭk)
hys·ter·e′si·graph (hĭs·tẽr·ē′sĭ·grăf)
hys·ter·e′sis (hĭs·tẽr·ē′sĭs)

I

i·con′o·scope (ī·kŏn′ŏ·skōp)
id·i·o·chro·mat′ic (īd·ĭ·ȯ·krō·măt′ĭk)
ig·ni′tion (ĭg·nĭ′shŭn)
ig·ni′tor (ĭg·nī′tẽr)
ig′ni·tron (ĭg′nĭ·trŏn)
il·lu′mi·nant (ĭ·lū′mĭ·nŭnt)
il·lu·mi·nom′e·ter (ĭ·lū·mĭ·nŏm′ĕ·tẽr)
im·preg·na′tion (ĭm·prĕg·nā′shŭn)
in·can·des′cence (ĭn·kăn·dĕs′ĕns)
in′ci·dence (ĭn′sĭ·dĕns)
in·cip′i·ent (ĭn·sĭp′ĭ·ĕnt)
in·cli·na′tion (ĭn·klĭ·nā′shŭn)
in·cli·nom′e·ter (ĭn·klĭ·nŏm′ĕ·tẽr)
in·co·her′ent (ĭn·kȯ·hêr′ĕnt)
in·cre·duc′tor (ĭn·krĕ·dŭk′tẽr)
in′cre·ment (ĭn′krä·mĕnt)
in·fin·i·tes′i·mal (ĭn·fĭn·ĭ·tĕs′ĭ·môl)
in·hib′it·ing (ĭn·hĭb′ĭ·tĭng)
in·spec′to·scope (ĭn·spĕk′tȯ·skōp)
in′su·laz·ing (ĭn′su̇·lāz·ĭng)
in′te·ger (ĭn′tĕ·jẽr)
in′teg·ral (ĭn′tĕg·rôl)
in·ten·si·tom′e·ter (ĭn·tĕn·sĭ·tŏm′ĕ·tẽr)
in·ter·dig′i·tal (ĭn·tẽr·dĭj′ĭ·tôl)
in·ter·fer·om′e·ter (ĭn·tẽr·fẽr-
ŏm′ĕ·tẽr)
in·ter·po·la′tion (ĭn·tẽr′pȯ·lā′shŭn)
in·ter′pre·tive (ĭn·tẽr′prĕ·tĭv)
in·ter·ro·ga′tion (ĭn·tĕr·ȯ·gā′shŭn)
in·ter′ro·ga·tor (ĭn·târ′ȯ·gā·tẽr)
in′ter·ties (ĭn′tẽr·tīz)
in·to·na′tion (ĭn·tō·nā′shŭn)
in·trin′sic (ĭn·trĭn′sĭk)
In′var (ĭn′vär)
i′on (ī′ŏn)
i·on′ic (ī·ŏn′ĭk)
i·on′o·sphere (ī·ŏn′ȯ·sfêr)
i′so·bar (ī′sȯ·bär)
i′so·chrone (ī′sȯ·krōn)
i·soch′ro·nous (ī·sŏk′rȯ·nŭs)
i·so·clin′ic (ī·sȯ·klĭn′ĭk)
i′so·mer (ī′sȯ·mẽr)
i·so·stat′ic (ī·sō·stăt′ĭc)
i·so·trop′ic (ī·sȯ·trŏp′ĭk)
it′er·a·tive (ĭt′ẽr·ä·tĭv)

J

jaff (jăf)
jez′a·bel (jĕz′ä·bĕl)
joule (jo͞ol)

K

ke·no·pli′o·tron (kē·nȯ·plī′ȯ·trŏn)
ke′no·tron (kē′nȯ·trŏn)
ke·rau′no·phone (kē·rä′nȯ·fōn)
Kerr (kẽr)
Ki·kuch′i (kĭ·ko͞o′chė)
kil′o- (kĭl′ō)
kil′ohm·me·ter (kĭl′ōm·mē·tẽr)
kil′o·me·ter (kĭl′ȯ·mē·tẽr)
kil·o·met′ric (kĭl·ȯ·mĕ′trĭk)
kil′o·var (kĭl′ȯ·vär)
kin′e·scope (kĭn′ĕ·skōp)
ki·net′ic (kĭ·nĕt′ĭk)
Kirch′hoff (kẽrk′ôff)
kly·don′o·graph (klī·dŏn′ȯ·grăf)
kly′stron (klī′strŏn)
kur·to′sis (kẽr·tō′sĭs)
ky′mo·graph (kī′mȯ·grăf)

L

la′bile (lā′bĭl)
lab′y·rinth (lăb′ĭ·rĭnth)
lac′quer (lăk′ẽr)
lamb′da (lăm′dä)
La·mont's′ (lä·mŏnts′)
Lan′ge·vin (lăn′gĕ·vĭn)
Lang′muir (lăng′mūr)
La·place′ (lä·pläs′)
la·ryn·ga·phone (lă·rĭng′gä·fōn)
la′ser (lā′sẽr)
la′ten·cy (lā′tĕn·sē)
la′tent (lā′tĕnt)
lat′er·al (lăt′ẽr·ôl)
la·va·lier′ (lă·vä·lêr′)
LA′WEB (lā′wĕb)
leak′ance (lēk′ŭns)
Lech′er (lĕch′ẽr)
Le·clan·che′ (lĕ·klän·shā′)
Le′nard (lā′närt)
Le·pel′ (lĕ·pĕl′)
Lich′ten·berg (lĭk′tĕn·bẽrg)
Lis′sa·jous ((lĭ′sä·zho͞o)
lith′i·um (lĭth′ĭ·ŭm)
litz (lĭtz)
lo′dar (lō′där)
log·a·rith′mic (lŏg·ä·rĭth′mĭk)
lok′tal (lŏk′tôl)
lo·rac′ (lȯ·răc′)
lo·rad′ (lȯ·răd′)
lo·ran′ (lô·răn′)
lor′humb (lôr′ŭmb)
Los′sev (lō′sĕf)
lou′ver (lo͞o′vẽr)
lu′men (lū′mĕn)
lu·min·aire′ (lū·mĭn·âr′)
lu·mi·nes′cence (lū·mĭ·nĕs′ĕns)
lu·mi·nos′i·ty (lū·mĭ·nŏs′ĭ·tē)

M

mac·ro·son′ics (măk·rȯ·sŏn′ĭks)
mag′amp (măg′ămp)
mag′nal (măg′nôl)
mag·ne′si·um (măg·nē′sĭ·ŭm)
Mag′ne·syn (măg′nĕ·sĭn)
mag′net·ite (măg′nĕ·tīt)
mag·ne′to (măg·nē′tō)
mag·ne·to·hy·dro·dy·nam′ic (măg·nē-
tȯ·hī·drȯ·dī·năm′ĭk)
mag·ne·to·i·on′ic (măg·nē·tȯ·ī·ŏn′ĭk)
mag·ne·tom′e·ter (măg·nĕ·tŏm′ĕ·tẽr)
mag·net′tor (măg·nĕt′ẽr)
Mag·nis′tor (măg·nĭs′tẽr)
man′ga·nin (măn′gä·nĭn)
ma·nom′e·ter (mă·nŏm′ĕ·tẽr)
Mar·co′ni (mär·kō′nĭ)
ma′trix (mā′trĭks)
ma′trix·er (mā′trĭk·sẽr)
Mat·te·u′cci (mät·tȧ·o͞ot′chė)
Mc·Nal′ly (mŭk·năl′ē)
Mea′cham (mē′chŭm)
mea′con·ing (mē′kŭn·ĭng)
me′di·an (mē′dĭ·ŭn)
meg′a- (mĕg′ä)
Meiss′ner (mīz′nẽr)
mel (mĕl)
me′sa (mā′sä)
mes′on (mĕs′ŏn)
met′a·dyne (mĕt′ä·dīn)
me·te·or·o·log′i·cal (mē·tĭ·ẽr·ŏ-
lŏj′ĭ·k'l)
met′re·chon (mĕt′rĕ·kŏn)
mho (mō)
mi′cro- (mī′krō)
mi·cro·cir′cui·try (mī·krō·sẽr′kĭ·trē)
mi·cro·log′ic (mī·krō·lŏj′ĭk)
mi·cro·mas·sage′ (mī·krȯ·mäs·säzh′)
mi·cro·min·i·a·tur·i·za′tion (mī·krȯ-
mĭn·ĭ·ä·chẽr·ĭ·zā′shŭn)
mi·cro·mod′ule (mī·krȯ·mŏd′u̇l)
mi·cro·ra·di·om′e·ter (mī·krȯ·rā·dė-
ŏm′ĕ·tẽr)
mil′li- (mĭl′ĭ)
mil·li·roent′gen (mĭl·ĭ·rĕnt′gĕn)
min·i·a·tur·i·za′tion (mĭn·ĭ·ä·chẽr·ĭ-
zā′shŭn)
Min′ter (mĭn′tẽr)
mitte-seite (mĭt-sīt)
mne·mon′ic (nė·mŏn′ĭk)
mod′ule (mŏd′ūl)
moi′re (mō′rā)
mol (mōl)
mol·ec·tron′ics (mŏl·ĕk·trŏn′ĭks)
mo·lec′u·lar (mŏ·lĕk′u̇·lẽr)
mol′e·cule (mŏl′ĕ·kūl)
mo·lyb′de·num (mȯ·lĭb′dĕ·nŭm)
mon·au′ral (mŏn·ô′rŭl)
mon·o·chro′ma·tor (mŏn·ȯ·krō′mȧ·tẽr)
mon′o·chrome (mŏn′ȯ·krōm)
mor·pho·log′i·cal (môr·fȯ·lŏj′ĭ·kŭl)
mor·phol′o·gy (môr·fŏl′ȯ·gē)
mo·sa′ic (mȯ·zā′ĭk)
mu (mū)
Mul′ler (mŭl′ẽr)
mul·ti·cel′lu·lar (mŭl·tĭ·sĕl′u̇·lẽr)
mul·ti·proc′es·sor (mŭl·tĭ·prŏs′ĕs·ẽr)
Mu′met·al (mū′mĕt·ŭl)
Mun′sell (mŭn′sŭl)
My′ca·lex (mī′kä·lĕks)
My′lar (mī′lär)

N

nan′o (năn′ō)
Na·per′i·an (nȧ·pêr′ĭ·ŭn)
na′pi·er (nā′pē·ẽr)
Nav′a·globe (năv′ä·glōb)
na′var (nā′vär)
nav′ar·ho (năv′är·hȯ)
neg′a·tron (nĕg′ä·trŏn)
ne′mo (nē′mō)

ne'per (nē'pẽr)
Nernst (nẽrnst)
neu·ris'tor (nu̍·rĭs'tẽr)
neu·ro·e·lec·tric'i·ty (nu̍·ro̍·ē̍·lĕktrĭs'ĭ·tē)
neu·tri'no (nōō·trē'nō)
neu'tron (nu̍'trŏn)
Ni'chrome (nī'krōm)
Nip'kow (nĭp'kou)
noc'to·vi·sion (nŏk'to̍·vĭ·zhŭn)
nod'al (nōd''l)
node (nōd)
nod'ules (nŏd'ūls)
no'mo·graph (nō'mo̍·grăf)
non·e·quiv'a·lence (nŏn·ē̍·kwĭv'ä·lĕns)
non·e·ras'a·ble (nŏn·ē̍·rās'ă·b'l)
nov'al (nōv''l)
nov'ice (nŏv'ĭs)
nu'cle·ar (nū'klē̍·ẽr)
nu·cle·a'tion (nōō·klĭ·ā'shŭn)
nu·cle·on'ics (nū·klē̍·ŏn'ĭks)
nu'cle·us (nū'klē̍·ŭs)
nu'tat·ing (nū'tāt·ĭng)
nu·vis'tor (nōō·vĭs'tẽr)
Ny'quist (nī'kwĭst)

O

o·blique' (ō·blēk')
o'boe (ō'bō)
ob·so·les'cence (ŏb·so̍·lĕs'ĕns)
oc·clude' (ŏ·klōōd')
oc'tal (ŏc'tŭl)
o'do·graph (ō'do̍·grăf)
oer'sted (ẽr'stĕd)
o·me'ga (ō·mĕ'gä)
om'ni·bear·ing (ŏm'nĭ·bâr·ĭng)
om·ni·di·rec'tion·al (ŏm·nĭ·dĭrĕk'shŭn·äl)
on'do·graph (ŏn'do̍·grăf)
o·pac'i·me·ter (ō·păs'ĭ·mē̍·tẽr)
op'er·and (ŏp'ẽr·ănd)
op·to·e·lec·tron'ic (ŏp·to̍·ē̍·lĕktrŏn'ĭk)
op'to·phone (ŏp'to̍·fōn)
or'i·fice (ŏr'ĭ·fĭs)
or'thi·con (ŏr'thĭ·kŏn)
or'tho·code (ŏr'thō·cōd)
or·thog'o·nal (ŏr·thŏg'ŏ·nôl)
os'cil·la·to·ry (ŏs'ĭ·lä·tôr·ē̍)
os·cil'lo·gram (ŏ·sĭl'o̍·grăm)
o'zone (ō'zōn)

P

pan·o·ram'ic (păn·ô·răm'ĭk)
pa·rab'o·la (pä·răb'o̍·lä)
par·a·bol'ic (pâr·ä·bŏl'ĭk)
pa·rab'o·loid (pä·răb'o̍·loid)
pa·rab·o·loi'dal (pä·răb·o̍·loi'd'l)
par'af·fin (pâr'ä·fĭn)

par'al·lax (pâr'ä·lăks)
pa·ram'e·ter (pä·răm'ĕ·tẽr)
par·a·me'tric (pâr·ă·mĕ'trĭk)
pa·ram'e·tron (pä·răm'ĕ·trŏn)
par·a·mis'tor (pâr·ä·mĭs'tẽr)
par·a·sit'ic (pâr·ä·sĭt'ĭk)
Pas'chen's (păs'kĕnz)
Pel·tier' (pĕl·tyā')
per·cep'tron (pẽr·sĕp'trŏn)
per·cus'sive (pẽr·cŭs'ĭv)
per·i·od'ic (pêr·ĭ·ŏd'ĭk)
pe·riph'er·al (pẽr·ĭf'ẽr·ŭl)
per·me·a·bil'i·ty (pẽr·mē̍·ä·bĭl'ĭ·tē)
per·me·am'e·ter (pẽr·mē̍·ăm'ĕ·tẽr)
per·mit·tiv'i·ty (pẽr·mĭt·tĭv'ĭ·tē)
per·sis'tor (pẽr·sĭs'tẽr)
phan'o·tron (făn'o̍·trŏn)
phan'tas·tron (făn'täs·trŏn)
phas'i·tron (fāz'ĭ·trŏn)
phas'ma·jec·tor (fāz'mä·jĕk·tẽr)
pha'sor (fā'zẽr)
phe·nol'ic (fē·nŏl'ĭk)
phi (fē)
phon (fŏn)
pho'nemes (fō'nēmz)
pho·net'ic (fo̍·nĕt'ĭk)
pho'non (fō'nŏn)
phos·pho·res'cence (fŏs·fo̍·rĕs'ĕns)
pho·to·e·las·tic·i·ty (fō·tō·ē̍·lăstĭ'sĭ·tē)
pho·to·mag·ne'to·e·lec'tric (fō·tōmăg·nē'tō·ē̍·lĕk'trĭk)
pho·tom'e·ter (fo̍·tŏm'ĕ·tẽr)
pho·tom'e·try (fo̍·tŏm'ĕ·trē)
pho'to·sphere (fō'tō·sfêr)
pho·to·vol·ta'ic (fō·tō·vŏl·tā'ĭk)
pi (pī)
pi'co (pī'kō)
pi·e·zo·di·e·lec'tric (pī·ē·zō·dī·ē̍lĕk'trĭk)
pi·e·zo·e·lec'tric (pī·ē·zō·ē̍·lĕk'trĭk)
pi'e·zo·id (pī'ē·zoid)
Pi·ra'ni (pĭ·rä'nē̍)
planch'et (plăn'chĭt)
Planck'i·an (plănk'ĭ·ŭn)
Planck's (plănks)
plas'ma·tron (plăz'mä·trŏn)
plas'ti·ciz·er (plăs'tĭ·sīz·ẽr)
pla·teau' (plă·tō')
pla·tin'o·tron (plă·tĭn'ŏ·trŏn)
plat'i·num (plăt'ĭ·nŭm)
pli·o·dy'na·tron (plī·o̍·dī'nä·trŏn)
pli'o·tron (plī'ŏ·trŏn)
plu·to'ni·um (plōō·tō'nĭ·ŭm)
poid (poid)
Pois·son's' (poi·sŏnz')
pol·y·es'ter (pŏl·ĭ·ĕs'tẽr)
pol·y·eth'yl·ene (pŏl·ĭ·ĕth'ĭ·lē̍n)
pol·y·sty'rene (pŏl·ĭ·stī'rēn)
por·ta·men'to (pôr·tä·mĕn'tō)

pos'i·tron (pŏz'ĭ·trŏn)
po·tas'si·um (po̍·tăs'ĭ·ŭm)
po·ten'tial (po̍·tĕn'shŭl)
po·ten·ti·om'e·ter (po̍·tĕn·shĭ·ŏm'ĕtẽr)
Po·tier' (pō·tyā')
Poyn'ting's (poin'tĭngz)
pre·ces'sion (prē̍·sĕsh'ŭn)
pre·cip'i·ta·tor (prē̍·sĭp'ĭ·tā·tẽr)
pre·cip'i·tron (prē̍·sĭp'ĭ·trŏn)
pre·cur'sor (prē̍·kẽr'sẽr)
pro·fi·lom'e·ter (pro̍·fĭ·lŏm'ĕ·tẽr)
pro·me'thi·um (pro̍·mē'thĭ·ŭm)
pseu'do·code (sōō'dō·cōd)
Pu'pin (pu̍'pĕn)
py'lon (pī'lŏn)
py·ram'i·dal (pĭ·răm'ĭ·dŭl)
pyr·he·li·om'e·ter (pīr·hē·lĭ·ŏm'ĕ·tẽr)
py·ro·e·lec'tric (pī·ro̍·ē̍·lĕk'trĭk)
py·rom'e·ter (pī·rŏm'ĕ·tẽr)
Py·thag·o·re'an (pĭ·thăg·o̍·rē'ŭn)

Q

quad (kwôd)
quad'rant (kwŏd'rănt)
quad·ran'tal (kwŏd·răn'tŭl)
quad'ri·pole (kwŏd'rĭ·pōl)
quad'ru·pole (kwŏd'rōō·pōl)
quan·ti·za'tion (kwŏn·tĭ·zā'shŭn)
quan'tize (kwŏn'tīz)
quan'tum (kwŏn'tŭm)
qua'si- (kwā'sī-)
qui·es'cent (kwī·ĕs'ĕnt)

R

ra'con (rā'kŏn)
rad (răd)
ra'dar (rā'där)
ra'di·ac (rā'dĭ·ăk)
ra'di·al (rā'dĭ·ŭl)
ra'di·a·tive (rā'dĭ·ā·tĭv)
ra·di·ol'o·gy (rā·dĭ·ŏl'ŏ·jĭ)
ra·di·o·paque' (rā·dĭ·o̍·pāk')
ra·di·o·son'o·buoy (rā·dĭ·o̍·sŏn'ŏ·boi)
ra·di·o·te·leg'ra·phy (rā·dĭ·o̍·tĕlĕg'rä·fē)
ra·di·o·te·leph'o·ny (rā·dĭ·o̍·tĕlĕf'o̍·nē)
ra·di·o·ther'a·py (rā·dĭ·o̍·thâr'ä·pē)
ra·di·o·ther'mics (rā·dĭ·o̍·thẽr'mĭks)
ra'dix (rā'dĭks)
ra'dome (rā'dōm)
ra'dux (rā'dŭks)
ra'mark (rā'märk)
Ram'sauer (răm'sour)
ras'ter (răs'tẽr)
ray'dist (rā'dĭst)
Ray'leigh (rā'lĭ)
re·ac·qui·si'tion (rē̍·ăk·wĭ·zĭ'shŭn)

re·bec'ca-eu·re'ka (rĕ·bĕk'ä-u̍·rē'kä)
re·ca·les'cent (rē·kä·lĕs'ĕnt)
re·cip'ro·cal (rē̍·sĭp'rŏ·kŭl)
rec·i·pros'i·ty (rĕs·ĭ·prŏs'ĭ·tē)
re·con'nais·sance (rē̍·kŏn'ĭ·sŭns)
rec'ti·gon (rĕk'tĭ·gŏn)
rec·ti·lin'e·al (rĕk·tĭ·lĭn'ē̍·ŭl)
rec·ti·lin'e·ar (rĕk·tĭ·lĭn'ē̍·ẽr)
re·cur'sion (rē̍·kûr'zhŭn)
re·dun'dan·cy (rē̍·dŭn'dŭn·sē)
re·flec·tom'e·ter (rē̍·flĕk·tŏm'ĕ·tẽr)
re·frac·tom'e·ter (rē̍·frăk·tŏm'ĕ·tẽr)
re·fran'gi·ble (rē̍·frăn'jĭ·b'l)
re·ig·ni'tion (rē̍·ĭg·nĭ'shŭn)
Rei'ke (rī'kē̍)
Rei'nartz (rī'närts)
rel (rĕl)
re·lax'or (rē̍·lăk'sẽr)
rem'a·nence (rĕm'ä·nĕns)
re·per'fo·ra·tor (rē̍·pẽr'fo̍·rā·tẽr)
re·sid'u·al (rē̍·zĭd'u̍·ŭl)
re·sis·tiv'i·ty (rē̍·zĭs·tĭv'ĭ·tē)
res'na·tron (rĕz'nä·trŏn)
ret·ro·di·rec'tive (rĕt·rō·dĭ·rĕk'tĭv)
re·ver·ber·a'tion (rē̍·vẽr·bẽr·ā'shŭn)
rhe'o·stat (rē'o̍·stăt)
rhom'bic (rŏm'bĭk)
rho-the'ta (rō-thā'tä)
rhum'ba·tron (rōōm'bä·trŏn)
Rig'hi-Le·duc' (rē'gē̍-lŭ·dūk')
RI·OM'E·TER (rī·ŏm'ĕ·tẽr)
Ro·chelle' (rō·shĕl')
roent'gen (rĕnt'gĕn)
roent·gen'o·gram (rĕnt·gĕn'o̍·grăm)
roent·gen·ol'o·gy (rĕnt·gĕn·ŏl'ŏ·jē)
roent·gen·om'e·ter (rĕnt·gĕn·ŏm'ĕ·tẽr)
Ruhm'korff (rŭm'kôrf)

S

scin·til·la'tion (sĭn·tĭ·lā'shŭn)
sel'en·ide (sĕl'ĕn·īd)
sep'tate (sĕp'tāt)
se·quen'tial (sē̍·kwĕn'shŭl)
ser'rat·ed (sĕr'āt·ĕd)
sex·i·dec'i·mal (sĕk·sĭ·dĕs'ĭ·mŭl)
sfer'ics (sfêr'ĭks)
sho·ran' (shō·răn')
sil'i·ca (sĭl'ĭ·kä)
sil'i·con (sĭl'ĭ·kŏn)
sil'i·cone (sĭl'ĭ·kōn)
si'mul·cast·ing (sī'mŭl·kăst·ĭng)
sine (sīn)
sins (sĭnz)
si·nus·oi'dal (sī·nu̍·soi'dŭl)
skew (skū)
ski'a·tron (skē'ä·trŏn)
sniv'et (snĭv'ĕt)
so'dar (sō'där)
so'di·um (sō'dĭ·ŭm)
so'far (sō'fär)

so′lar (sō′lẽr)
sol′der (sŏd′ẽr)
so′len•oid (sō′lĕn•oid)
so′li•on (sō′lĭ•ŏn)
so′nar (sō′när)
sone (sōn)
son′ic (sŏn′ĭk)
son′ne (sŏn′e̍)
son′o•buoy (sŏn′o̍•boi)
son•om′e•ter (sŏn•ŏm′ĕ•tẽr)
so•phis′ti•cat•ed (so̍•fĭs′tĭ•kāt•ĕd)
spa•cis′tor (spā•sĭs′tẽr)
spa•ghet′ti (spä•gĕt′ē)
spa′tial (spā′shŭl)
spec′tral (spĕk′trôl)
spec•trom′e•ter (spĕk•trŏm′ĕ•tẽr)
spec′u•lar (spĕk′u̍•lẽr)
sphere (sfêr)
spher′i•cal (sfêr′ĭ•kŭl)
spin•thar′i•scope (spĭn•thâr′ĭ•skōp)
spo•rad′ic (spo̍•răd′ĭk)
spur′i•ous (spẽr′ĭ•ŭs)
squeg′ger (skwĕg′ẽr)
squelch (skwĕlch)
squint (skwĭnt)
squit′ter (skwĭt′ẽr)
sta•bil′i•volt (stä•bĭl′ĭ•vo̍lt)
sta′lo (stā′lō)
stat (stăt)
ste′a•tite (ste̍′ä•tīt)
Stef′an-Boltz′mann (stĕf′ŭn-bōltz′mŭn)
Stein′metz (stīn′mĕts)
sten′ode (stĕn′ōd)
ste•ra′di•an (stĕ•rā′dĭ•ŭn)
ster•e•o•ceph′a•loid (stâr•e̍•o̍•sĕf′äloid)
stim′u•lus (stĭm′u̍•lŭs)
sto•chas′tic (stō•kăs′tĭk)
sto•chi•o•met′ric (stō•kĭ•ŏ•mĕt′rĭk)
strat′o•sphere (străt′o̍•sfêr)
stri•a′tion (strī•ā′shŭn)
strob′o•scope (strōb′o̍•skōp)
strob•o•scop′ic (strōb•o̍•skŏp′ĭk)
strob′o•tron (strōb′o̍•trŏn)
Suhl (so͞ol)
sur•veil′lance (sẽr•vā′lăns)
sus•cep•ti•bil′i•ty (sŭ•sĕp•tĭ•bĭl′ĭ•tē)
sym•bol′ic (sĭm•bŏl′ĭk)
syn′chro•tron (sĭn′kro̍•trŏn)
syn′the•siz•er (sĭn′thĕ•sīz•ẽr)

T

tach•om′e•ter (tăk•ŏm′ĕ•tẽr)
tan•gen′tial (tăn•jĕn′shŭl)
tan′ta•lum (tăn′tä•lŭm)
tec′ne•tron (tĕk′nĕ•trŏn)
tel•e•gen′ic (tĕl•ĕ•jĕn′ĭk)
te•leg′ra•phy (tĕ•lĕg′rä•fē)
te•lem′e•ter (tĕ•lĕm′ĕ•tẽr)
te•lem′e•try (tĕ•lĕm′ĕ•trē)
te•leph′o•ny (tĕ•lĕf′o̍•nē)
tel′e•ran (tĕl′ĕ•răn)
tel′e•synd (tĕl′ĕ•sĭnd)
tel′ex (tĕl′ĕks)
ten•si•om′e•ter (tĕn•sĭ•ŏm′ĕ•tẽr)
ten′sion (tĕn′shŭn)
ter′a (târ′ä)
ter′na•ry (tẽr′nä•rē)
ter•rain′ (tẽr•ān′)
ter•res′tri•al (tẽr•ĕs′trĭ•ŭl)
ter′ti•ar•y (tẽr′shĭ•âr•ē)
tes′la (tĕs′lä)
thal′lo•fide (thăl′o̍•fīd)
ther′e•min (thâr•ĕ•mĭn)
ther′mi•on (thẽr′mĭ•ŏn)
ther•mi•on′ic (thẽr•mĭ•ŏn′ĭk)
ther•mis′tor (thẽr•mĭs′tẽr)
the′ta (thā′tä)
Thev′e•nin′s (thĕv′ĕ•nĭns)
tho′ri•at•ed (thō′rĭ•āt•ĕd)
thy′ra•tron (thī′rä•trŏn)
thy′ris•tor (thī′rĭs•tẽr)
thy′rite (thī′rīt)
tim′bre (tĭm′bẽr)
Ton′o•tron (tŏn′ŏ•trŏn)
to•roi′dal (to̍•roi′dôl)
torque (tôrk)
tor•si•om′e•ter (tôr•sĭ•ŏm′ĕ•tẽr)
tor′sion (tôr′shŭn)
tour′ma•line (to͝or′mä•lĭn)
trans•hy′brid (trăns•hī′brĭd)
tran′sient (trăn′shĕnt)
tran•sis′tance (trăn•sĭs′tŭns)
tran•si′tion (trăn•zĭsh′ŭn)
tran′si•tron (trăn′sĭ•trŏn)
trans•mis•si•bil′i•ty (trăns•mĭs•ĭbĭl′ĭ•tē)
tran•son′ic (trăn•sŏn′ĭk)
trans•pon′der (trăns•pŏn′dẽr)
trap•e•zoi′dal (trăp•ĕ•zoi′d’l)
tri•ax′i•al (trī•ăx′ĭ•ôl)
tri•bo•e•lec′tric (trī•bo̍•e̍•lĕk′trĭk)
tri′con (trī′kŏn)
trig′a•tron (trĭg′ä•trŏn)
tri•gis′tor (trī•jĭs′tẽr)
tri•nis′tor (trī•nĭs′tẽr)
trin′o•scope (trĭn′ŏ•skōp)
tri•stim′u•lus (trī•stĭm′u̍•lŭs)
trit′i•um (trĭt′ĭ•ŭm)
trop′o•sphere (trŏp′o̍•sfêr)
trop•o•spher′ic (trŏp•o̍•sfĕr′ĭk)
trun′cate (trŭn′kāt)
tun′gar (tŭn′gär)
tung′sten (tŭng′stĕn)
tur•bid′i•me•ter (tẽr•bĭd′ĭ•me̍•tẽr)

U

ul′tor (ŭl′tôr)
ul•tra•mi•crom′e•ter (ŭl•trä•mīkrŏm′ĕ•tẽr)
um•bil′i•cal (ŭm•bĭl′ĭ•kŭl)

V

va′lence (vā′lĕns)
va•lid′i•ty (vä•lĭd′ĭ•tē)
Van′ de Graaff (văn′dĕ•grăf)
var′ac•tor (vâr′ăk•tẽr)
var′hour (vär′our)
Var′i•ac (vâr′ĭ•ăk)
var′in•dor (vâr′ĭn•dẽr)
var•i•om′e•ter (vâr•ĭ•ŏm′ĕ•tẽr)
var•is′tor (vâr•ĭs′tẽr)
ver′ni•er (vẽr′nĭ•ẽr)
ves•tig′i•al (vĕs•tĭj′ĭ•ŭl)
vi•bra′to (vĭ•brä′to̍)
vi′bra•tron (vī′brä•trŏn)
vi•brom′e•ter (vī•brŏm′ĕ•tẽr)
vi′di•con (vĭ′dĭ•kŏn)
Vil•la′ri (vĭ•lâr′e̍)
vi′nyl (vī′nŭl)
vis•com′e•ter (vĭs•kŏm′ĕ•tẽr)
vis•co•sim′e•ter (vĭs•ko̍•sĭm′ĕ•tẽr)
vis•cos′i•ty (vĭs•kŏs′ĭ•tē)
vis′cous (vĭs′kŭs)
vo′das (vō′dŭs)
vo′der (vō′dẽr)
vo′gad (vō′găd)
vol′a•tile (vŏl′ŭ•tĭl)
vol•ta′ic (vōl•tā′ĭk)
vol•u•met′ric (vŏl•u̍•mĕt′rĭk)

W

Wag′ner (wăg′nẽr)
wa′mo•scope (wā′mo̍•skōp)
web′er (wĕb′ẽr)
Weh′nelt (vā′nĕlt)
Wert′heim (wẽrt′hīm)
Wie′de•mann (wē′dŭ•mŭn)
Wie′de•mann-Franz (wē′dŭ•mŭn-fränz)
Wien (wēn)
Wims′hurst (wĭmz′hẽrst)
wob′bu•la•tor (wŏb′u̍•lā•tẽr)

X

xe′non (zē′nŏn)
xe•ro•graph′ic (zē•ro̍•grăf′ĭk)
xe•rog′ra•phy (zē•rŏg′rä•fē)

Y

Ya′gi (yä′gē)

Z

ze′ner (zē′nẽr)
zep′pe•lin (zĕp′ĕ•lĭn)
zinc (zĭngk)

SCHEMATIC SYMBOLS

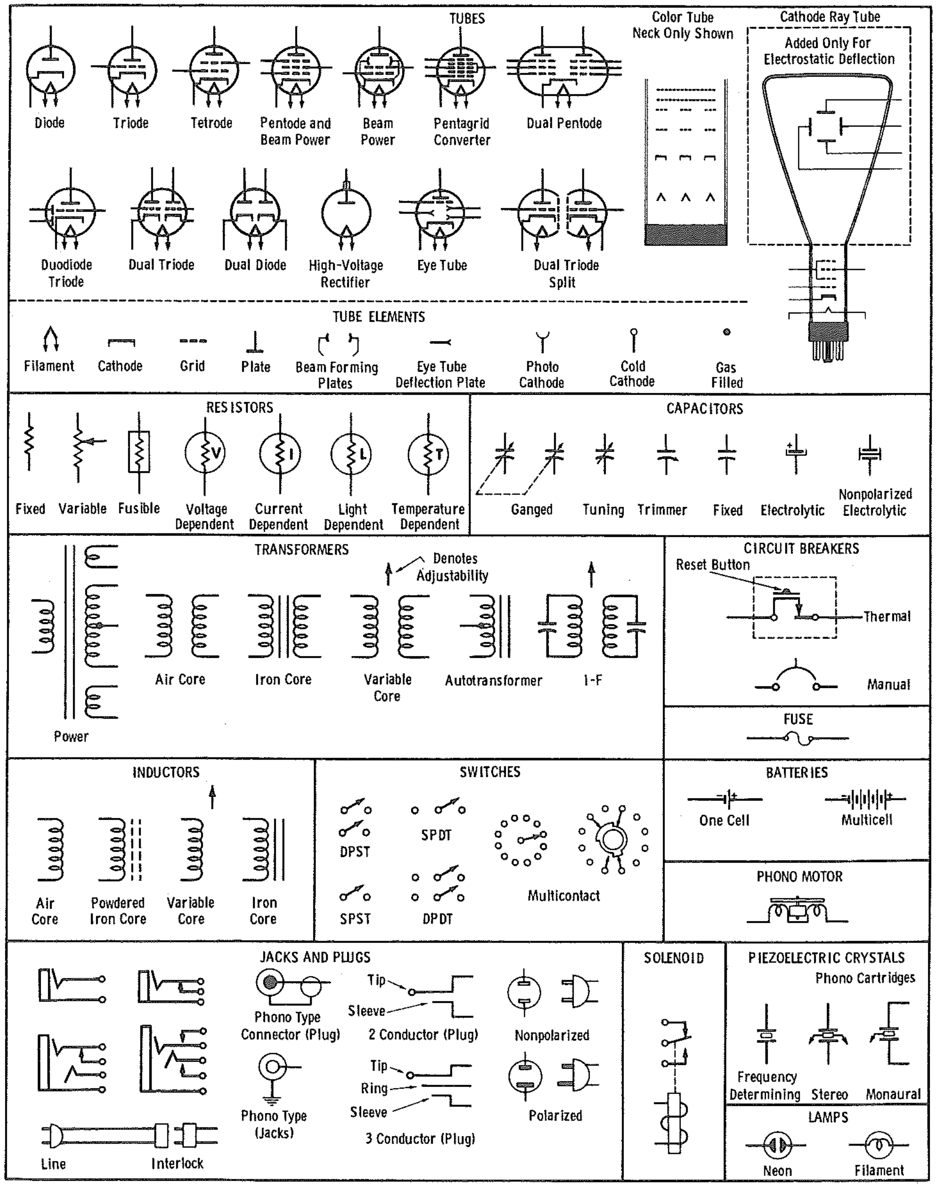
TUBES
Diode
Triode
Tetrode
Pentode and Beam Power
Beam Power
Pentagrid Converter
Dual Pentode
Duodiode Triode
Dual Triode
Dual Diode
High-Voltage Rectifier
Eye Tube
Dual Triode Split
Color Tube Neck Only Shown
Cathode Ray Tube
Added Only For Electrostatic Deflection
TUBE ELEMENTS
Filament
Cathode
Grid
Plate
Beam Forming Plates
Eye Tube Deflection Plate
Photo Cathode
Cold Cathode
Gas Filled
RESISTORS
Fixed
Variable
Fusible
Voltage Dependent
Current Dependent
Light Dependent
Temperature Dependent
V
I
L
T
CAPACITORS
Ganged
Tuning
Trimmer
Fixed
Electrolytic
Nonpolarized Electrolytic
TRANSFORMERS
Denotes Adjustability
Power
Air Core
Iron Core
Variable Core
Autotransformer
I-F
CIRCUIT BREAKERS
Reset Button
Thermal
Manual
FUSE
INDUCTORS
Air Core
Powdered Iron Core
Variable Core
Iron Core
SWITCHES
DPST
SPDT
SPST
DPDT
Multicontact
BATTERIES
One Cell
Multicell
PHONO MOTOR
JACKS AND PLUGS
Phono Type Connector (Plug)
Tip
Sleeve
2 Conductor (Plug)
Nonpolarized
Phono Type (Jacks)
Tip
Ring
Sleeve
3 Conductor (Plug)
Polarized
Line
Interlock
SOLENOID
PIEZOELECTRIC CRYSTALS
Phono Cartridges
Frequency Determining
Stereo
Monaural
LAMPS
Neon
Filament

SCHEMATIC SYMBOLS

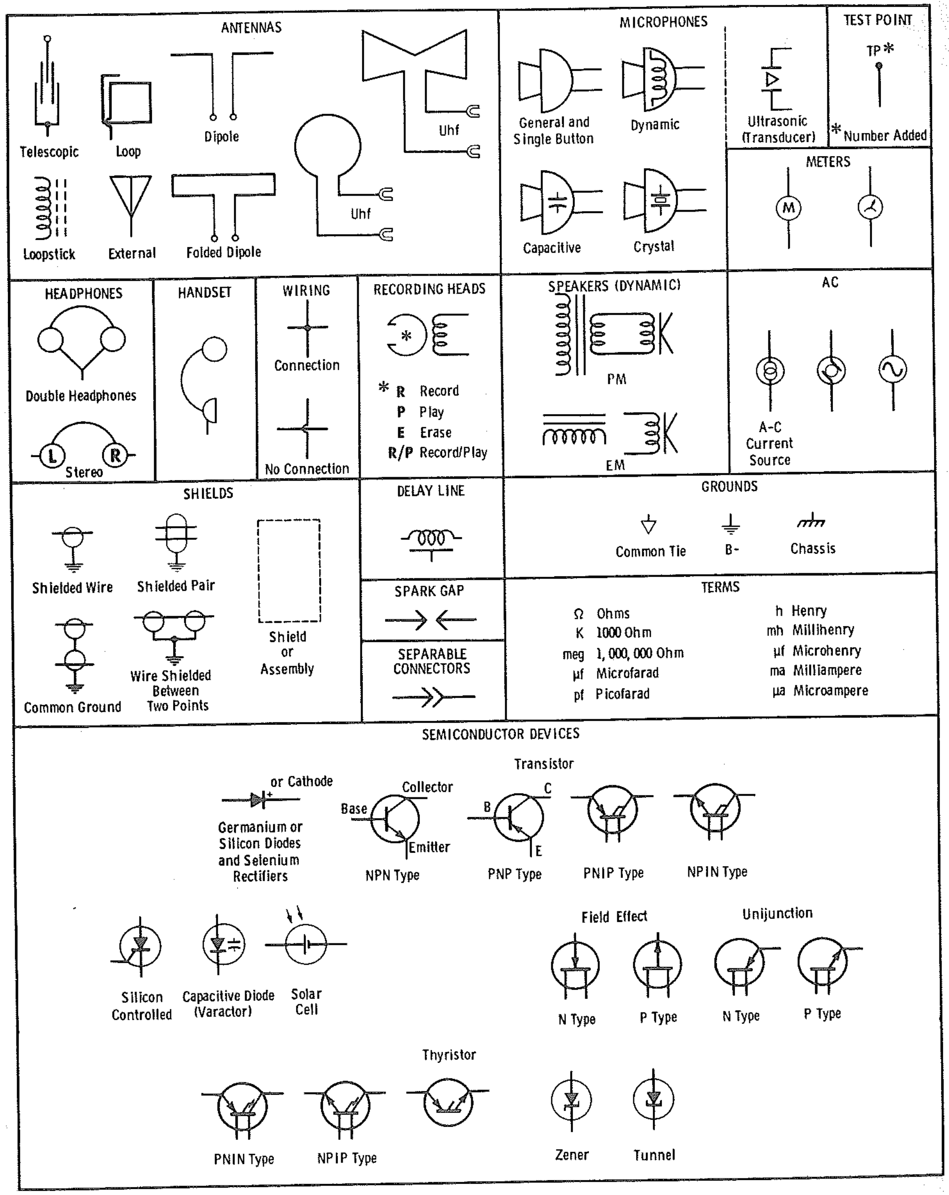
ANTENNAS
Telescopic
Loop
Dipole
Uhf
Uhf
Loopstick
External
Folded Dipole
MICROPHONES
General and Single Button
Dynamic
Capacitive
Crystal
Ultrasonic (Transducer)
TEST POINT
TP*
*Number Added
METERS
M
HEADPHONES
Double Headphones
L
R
Stereo
HANDSET
WIRING
Connection
No Connection
RECORDING HEADS
*
* R Record
P Play
E Erase
R/P Record/Play
SPEAKERS (DYNAMIC)
PM
EM
AC
A-C Current Source
SHIELDS
Shielded Wire
Shielded Pair
Shield or Assembly
Common Ground
Wire Shielded Between Two Points
DELAY LINE
SPARK GAP
SEPARABLE CONNECTORS
GROUNDS
Common Tie
B-
Chassis
TERMS
Ω Ohms
K 1000 Ohm
meg 1,000,000 Ohm
μf Microfarad
pf Picofarad
h Henry
mh Millihenry
μf Microhenry
ma Milliampere
μa Microampere
SEMICONDUCTOR DEVICES
Transistor
or Cathode
Germanium or Silicon Diodes and Selenium Rectifiers
Base
Collector
Emitter
NPN Type
B
C
E
PNP Type
PNIP Type
NPIN Type
Field Effect
Unijunction
Silicon Controlled
Capacitive Diode (Varactor)
Solar Cell
N Type
P Type
N Type
P Type
Thyristor
PNIN Type
NPIP Type
Zener
Tunnel

ABBREVIATIONS

A—Ammeter; ampere; area
a—Ampere
AC, a.c., a-c, ac—Alternating current
AF, a.f., a-f, af—Audio frequency
AFC, afc—Automatic frequency control
AGC, agc—Automatic gain control
AM, am—Amplitude modulation
Amp, amp., Amps, amps.—Ampere; amperes
Ant, ant.—Antenna
AVC, a.v.c., avc—Automatic volume control
B—Susceptance
b—Magnetic flux density
BC, bc—Broadcast
BFO, bfo—Beat-frequency oscillator
C—Capacitance; capacitor
°C—Degrees Celsius or centigrade
cm—Centimeter
cps—Cycles per second
CW, cw—Continuous wave
db—Decibels
DC, d.c., d-c, dc—Direct current
d.c.c, dcc—Double cotton-covered
DPDT, d.p.d.t., dpdt—Double-pole, double-throw
DPST, d.p.s.t., dpst—Double-pole single-throw
d.s.c, dsc—Double silk-covered
E, e—Voltage
e.c., ec—Enamel-covered
EMF, emf—Electromotive force
ERP—Effective radiated power
F, f—Farad
f—Frequency
°F—Degrees Fahrenheit
FM, f.m., fm—Frequency modulation
G—Conductance
G_m, gm, g_m—Mutual conductance
GCT—Greenwich Civil Time
gnd—Ground
H, h—Henry
HF, h.f., h-f, hf—High frequency
hp—Horsepower
hy—Henry
Hz—Hertz
I—Current
IF, i.f., i-f, if—Intermediate frequency
ips—Inches per second
j—Joule; an imaginary number; an operator to rotate a vector quantity 90° counterclockwise
K—× 1000; dielectric constant; a numerical value that does not change during a given period
k—Dielectric constant
KC, kc—Kilocycle
KHz—Kilohertz
kv—Kilovolt
kva—Kilovolt ampere
KW, kw—Kilowatt
KWH, kwh—Kilowatt hour
L—Inductance; inductor
l—Length
LF, l.f., l-f, lf—Low frequency
M—Mutual inductance; × 1000
m—Meter
ma—Milliampere
MC, Mc, mc—Megacycle
mcw—Modulated continuous wave
meg—Megohm
MF, m.f., m-f, mf—Medium frequency
mf, mfd—Microfarad
mh—Millihenry
MHz—Megahertz
mm—Millimeter
mmf, mmfd—Micromicrofarad
mv—Millivolt (sometimes microvolt)
mw—Milliwatt (sometimes microwatt)
NC—No connection
OD—Outside diameter
P—Power
pf—Power factor; picofarad
p-p—Peak-to-peak
Q—Merit of a coil or capacitor; quantity of electricity
R—Resistance; resistor
RC, R-C—Product of resistance time capacitance; resistor-capacitor
RF, r.f., r-f, rf—Radio frequency
RFC—Radio-frequency choke coil
rms—Root mean square
rpm—Revolutions per minute
s.c.c., scc—Single cotton-covered
s.c.e., sce—Single cotton enamel
sec—Second; secondary
SHF, s.h.f., shf—Super-high frequencies
s.s.c., ssc—Single silk-covered
SW, sw—Short wave
t—Time
T—Temperature
trf—Tuned radio frequency
UHF, uhf—Ultrahigh frequency
V, v—Volt; voltmeter
VHF, vhf—Very high frequencies
VOM, vom—Volt-ohm-milliammeter
VTVM, vtvm—Vacuum-tube voltmeter
VU—Volume unit
W—Watt; work
w—Watt
wh, whr—Watt-hour
X—Reactance
X_C—Capacitive reactance
X_L—Inductive reactance
Y—Admittance
Z—Impedance
μa—Microampere
μf—Microfarad
μh—Microhenry
$\mu\mu$f—Micromicrofarad
∿—Cycles per second

GREEK ALPHABET

The Greek alphabet is given below. The items for which each letter is a symbol are also listed. The small Greek letter is the symbol for all the items listed unless a capital letter is indicated (cap).

Letter		Name	Designates
Small	Capital		
α	A	Alpha	Angles, coefficients, attenuation constant, absorption factor, area.
β	B	Beta	Angles, coefficients, phase constant.
γ	Γ	Gamma	Specific quantity, angles, electrical conductivity, propagation constant, complex propagation constant (cap).
δ	Δ	Delta	Density, angles, increment or decrement (cap or small), determinant (cap), permittivity (cap).
ϵ	E	Epsilon	Dielectric constant, permittivity, base of natural (Napierian) logarithms, electric intensity.
ζ	Z	Zeta	Co-ordinate, coefficients.
η	H	Eta	Intrinsic impedance, efficiency, surface charge density, hysteresis, co-ordinates.
θ	Θ	Theta	Angular phase displacement, time constant, reluctance, angles.
ι	I	Iota	Unit vector.
κ	K	Kappa	Susceptibility, coupling coefficient.
λ	Λ	Lambda	Wavelength, attenuation constant, permeance (cap).
μ	M	Mu	Prefix *micro-*, permeability, amplification factor.
ν	N	Nu	Reluctivity, frequency.
ξ	Ξ	Xi	Co-ordinates.
o	O	Omicron	——
π	Π	Pi	3.1416 (circumference divided by diameter).
ρ	P	Rho	Resistivity, volume charge density, co-ordinates.
σ	Σ	Sigma	Surface charge density, complex propagation constant, electrical conductivity, leakage coefficient, sign of summation (cap).
τ	T	Tau	Time constant, volume resistivity, time-phase displacement, transmission factor, density.
υ	Υ	Upsilon	——
ϕ	Φ	Phi	Magnetic flux, angles, scalar potential (cap).
χ	X	Chi	Electric susceptibility, angles.
ψ	Ψ	Psi	Dielectric flux, phase difference, co-ordinates, angles.
ω	Ω	Omega	Angular velocity ($2\pi f$), resistance in ohms (cap), solid angles (cap).

SEMICONDUCTOR ABBREVIATIONS

The following symbols and abbreviations have been adopted as standard by the Electronic Industries Association (EIA) and the National Electrical Manufacturers Association (NEMA).

B, b—Base electrode for units employing a single base
b_1, b_2, etc.—Base electrodes for more than one base
BV_{CBO}—Breakdown voltage, collector to base, emitter open
BV_{CEO}—Breakdown voltage, collector to emitter, base open
BV_{CER}—Breakdown voltage, collector to emitter, with specified resistance between base and emitter
BV_{CES}—Breakdown voltage, collector to emitter, with base short-circuited to emitter
BV_{EBO}—Breakdown voltage, emitter to base, collector open
BV_R—Breakdown voltage, reverse
C, c—Collector electrode
C_{ib}—Input capacitance (common base)
C_{ic}—Input capacitance (common collector)
C_{ie}—Input capacitance (common emitter)
C_{ob}—Output capacitance (common base)
C_{oc}—Output capacitance (common collector)
C_{oe}—Output capacitance (common emitter)
E, e—Emitter electrode

h_{IB}—Static value of the input resistance (common base)
h_{ib}—Small-signal value of short-circuit input impedance (common base)
h_{IC}—Static value of the input resistance (common collector)
h_{ic}—Small-signal value of short-circuit input impedance (common emitter)
h_{IE}—Static value of the input resistance (common emitter)
h_{ie}—Small-signal value of short-circuit input impedance (common emitter)
h_{ie} (real)—Real part of small-signal value of short-circuit input impedance (common emitter)
h_{OB}—Static value of open-circuit output conductance (common base)
h_{ob}—Small-signal value of open-circuit output admittance (common base)
h_{OC}—Static value of open-circuit output conductance (common collector)
h_{oc}—Small-signal value of open-circuit output admittance (common collector)
h_{OE}—Static value of open-circuit output conductance (common emitter)
h_{oe}—Small-signal value of open-circuit output admittance (common emitter)

i_F—Forward current (instantaneous)
I_O—Average output (rectified) current
I_R—Reverse current (DC)
i_R—Reverse current (instantaneous)
K_θ—Thermal derating factor
L_c—Conversion loss
N, n—Region of a device where electrons are the majority carriers
NF—Noise figure
P, p—Region of a device where holes are the majority carriers
P_{BE}—Total power input (DC or average) to the base electrode with respect to the emitter electrode
p_{BE}—Total power input (instantaneous) to the base electrode with respect to the emitter electrode
P_{CB}—Total power input (DC or average) to the collector electrode with respect to the base electrode
p_{CB}—Total power input (instantaneous) to to the collector electrode with respect to the base electrode
P_{CE}—Total power input (DC or average) to the collector electrode with respect to the emitter electrode
p_{CE}—Total power input (instantaneous) to the collector electrode with respect to the emitter electrode

R_C—External collector resistance
r_{CE} (sat)—Collector-to-emitter saturation resistance
R_E—External emitter resistance
R_L—Load resistance
T—Temperature
T_A—Ambient temperature
T_C—Case temperature
t_d—Delay time
t_f—Fall time
t_{fr}—Forward recovery time
T_J—Junction temperature
T_{opr}—Operating temperature
t_p—Pulse time
t_r—Rise time
t_{rr}—Reverse recovery time
t_s—Storage time
T_{stg}—Storage temperature
t_w—Pulse average time
θ—Thermal resistance
$\theta_{J\text{-}A}$—Thermal resistance, junction-to-ambient
$\theta_{J\text{-}C}$—Thermal resistance, junction-to-case
V_{BB}—Base supply voltage (DC)
V_{BC}—Base-to-collector voltage (DC)
V_{bc}—Base-to-collector voltage (rms)
v_{bc}—Base-to-collector voltage (instantaneous)
V_{BE}—Base-to-emitter voltage (DC)

f_{hfb}—Small-signal, short-circuit, forward-current, transfer-ratio cutoff frequency (common base)
f_{hfc}—Small-signal, short-circuit, forward-current, transfer-ratio cutoff frequency (common collector)
f_{hfe}—Small-signal, short-circuit, forward-current, transfer-ratio cutoff frequency (common emitter)
f_{max}—Maximum frequency of oscillation
G_{PB}—Large-signal average power gain (common base)
G_{pb}—Small-signal average power gain (common base)
G_{PC}—Large-signal average power gain (common collector)
G_{pc}—Small-signal average power gain (common collector)
G_{PE}—Large-signal average power gain (common emitter)
G_{pe}—Small-signal average power gain (common emitter)
h_{FB}—Static value of the forward-current transfer ratio (common base)
h_{fb}—Small-signal, short-circuit, forward-current transfer ratio (common base)
h_{FC}—Static value of the forward-current transfer ratio (common collector)
h_{fc}—Small-signal, short-circuit, forward-current transfer ratio (common collector)
h_{FE}—Static value of the forward-current transfer ratio (common emitter)
h_{fe}—Small-signal, short-circuit, forward-current transfer ratio (common emitter)

h_{rb}—Small-signal value of open-circuit, reverse-voltage transfer ratio (common base)
h_{rc}—Small-signal value of open-circuit, reverse-voltage transfer ratio (common collector)
h_{re}—Small-signal value of open-circuit, reverse-voltage transfer ratio (common emitter)
I, i—Intrinsic region of a device (where neither holes nor electrons predominate)
I_B—Base current (DC)
I_b—Base current (rms)
i_b—Base current (instantaneous)
I_C—Collector current (DC)
I_c—Collector current (rms)
i_c—Collector current (instantaneous)
I_{CBO}—Collector cutoff current (DC), emitter open
I_{CEO}—Collector cutoff current (DC), base open
I_{CER}—Collector cutoff current (DC), with specified resistance between base and emitter
I_{CES}—Collector cutoff current (DC), with base short-circuited to emitter
I_{CEX}—Collector current (DC), with specified circuit between base and emitter
I_E—Emitter current (DC)
I_e—Emitter current (rms)
i_e—Emitter current (instantaneous)
I_{EBO}—Emitter cutoff current (DC), collector open
I_F—Forward current (DC)

P_{EB}—Total power input (DC or average) to the emitter electrode with respect to the base electrode
p_{EB}—Total power input (instantaneous) to the emitter electrode with respect to the base electrode
P_{IB}—Large-signal input power (common base)
P_{ib}—Small-signal input power (common base)
P_{IC}—Large-signal input power (common collector)
P_{ic}—Small-signal input power (common collector)
P_{IE}—Large-signal input power (common emitter)
P_{ie}—Small-signal input power (common emitter)
P_{OB}—Large-signal output power (common base)
P_{ob}—Small-signal output power (common base)
P_{OC}—Large-signal output power (common collector)
P_{oc}—Small-signal output power (common collector)
P_{OE}—Large-signal output power (common emitter)
P_{oe}—Small-signal output power (common emitter)
P_T—Total power input (DC or average) to all electrodes
p_T—Total power input (instantaneous) to all electrodes
R_B—External base resistance

V_{be}—Base-to-emitter voltage (rms)
v_{be}—Base-to-emitter voltage (instantaneous)
V_{CB}—Collector-to-base voltage (DC)
V_{cb}—Collector-to-base voltage (rms)
v_{cb}—collector-to-base voltage (instantaneous)
V_{CBF}—DC open-circuit voltage (floating potential) between collector and base, with emitter biased in reverse direction with respect to base
V_{CC}—Collector supply voltage (DC)
V_{CE}—Collector-to-emitter voltage (DC)
V_{ce}—Collector-to-emitter voltage (rms)
v_{ce}—Collector-to-emitter voltage (instantaneous)
V_{CE} (sat)—Collector-to-emitter saturation voltage
V_{EB}—Emitter-to-base voltage (DC)
V_{eb}—Emitter-to-base voltage (rms)
v_{eb}—Emitter-to-base voltage (instantaneous)
V_{EC}—Emitter-to-collector voltage (DC)
V_{ec}—Emitter-to-collector voltage (rms)
v_{ec}—Emitter-to-collector voltage (instantaneous)
V_{ECF}—DC open-circuit voltage (floating potential) between emitter and collector, with base biased in reverse direction with respect to collector
V_{EE}—Emitter supply voltage (DC)
V_F—Forward voltage (DC)
v_F—Forward voltage (instantaneous)
V_R—Reverse voltage (DC)
v_R—Reverse voltage (instantaneous)
V_{RT}—Reach-through voltage

Ohm's Law for A-C Circuits

The fundamental Ohm's law formulas for a-c circuits are given by

$$I = \frac{E}{Z}, \qquad Z = \frac{E}{I},$$

$$E = IZ, \qquad P = EI \cos \theta$$

where I = current in amperes,
Z = impedance in Ohms,
E = volts across Z,
P = power in watts,
θ = phase angle in degrees.

Phase Angle

The phase angle is defined as the difference in degrees by which current leads voltage in a capacitive circuit, or lags voltage in an inductive circuit, and in series circuits is equal to the angle whose tangent is given by the

ratio $\frac{X}{R}$ and is expressed by

$$\text{arc tan } \frac{X}{R}$$

where X = the inductive or capacitive reactance in ohms,
R = the non-reactive resistance in ohms,
of the combined resistive and reactive components of the circuit under consideration.

Therefore

in a purely resistive circuit, $\theta = 0°$
in a purely reactive circuit, $\theta = 90°$
and in a resonant circuit, $\theta = 0°$

also when

$\theta = 0°$, $\cos \theta = 1$ and $P = EI$,
$\theta = 90°$, $\cos \theta = 0$ and $P = 0$.

Degrees $\times$ 0.0175 = radians.
1 radian = 57.3°.

Power Factor

The power-factor of any a-c circuit is equal to the true power in watts divided by the apparent power in volt-amperes which is equal to the cosine of the phase angle, and is expressed by

$$p.f. = \frac{EI \cos \theta}{EI} = \cos \theta$$

where

$p.f.$ = the circuit load power factor,
$EI \cos \theta$ = the true power in watts,
EI = the apparent power in volt-amperes,
E = the applied potential in volts
I = load current in amperes.

Therefore

in a purely resistive circuit.

$$\theta = 0° \text{ and } p.f. = 1$$

and in a reactive circuit,

$$\theta = 90° \text{ and } p.f. = 0$$

and in a resonant circuit,

$$\theta = 0° \text{ and } p.f. = 1$$

Ohm's Law for D-C Circuits

The fundamental Ohm's law formulas for d-c circuits are given by,

$$I = \frac{E}{R}, \qquad R = \frac{E}{I},$$

$$E = IR, \qquad P = EI.$$

where I = current in amperes,
R = resistance in ohms,
E = potential across R in volts,
P = power in watts.

Ohm's Law Formulas for D-C Circuits

Known Values	Formulas for Determining Unknown Values of . . .			
	I	**R**	**E**	**P**
I & R			IR	I^2R
I & E		$\frac{E}{I}$		EI
I & P		$\frac{P}{I^2}$	$\frac{P}{I}$	
R & E	$\frac{E}{R}$			$\frac{E^2}{R}$
R & P	$\sqrt{\frac{P}{R}}$		$\sqrt{PR}$	
E & P	$\frac{P}{E}$	$\frac{E^2}{P}$		

Ohm's Law Formulas for A-C Circuits

Known Values	Formulas for Determining Unknown Values of . . .			
	I	**Z**	**E**	**P**
I & Z			IZ	$I^2Z \cos \theta$
I & E		$\frac{E}{I}$		$IE \cos \theta$
I & P		$\frac{P}{I^2 \cos \theta}$	$\frac{P}{I \cos \theta}$	
Z & E	$\frac{E}{Z}$			$\frac{E^2 \cos \theta}{Z}$
Z & P	$\sqrt{\frac{P}{Z \cos \theta}}$		$\sqrt{\frac{PZ}{\cos \theta}}$	
E & P	$\frac{P}{E \cos \theta}$	$\frac{E^2 \cos \theta}{P}$		

Trigonometric Relationships

In any right triangle, if we let

θ = the acute angle formed by the hypotenuse and the base leg,

ϕ = the acute angle formed by the hypotenuse and the altitude leg,

H = the hypotenuse,

A = the side adjacent θ and opposite ϕ,

O = the side opposite θ and adjacent ϕ,

then

$$\text{sine of } \theta = \sin\theta = \frac{O}{H}$$

$$\text{cosine of } \theta = \cos\theta = \frac{A}{H}$$

$$\text{tangent of } \theta = \tan\theta = \frac{O}{A}$$

$$\text{cosecant of } \theta = \csc\theta = \frac{H}{O}$$

$$\text{secant of } \theta = \sec\theta = \frac{H}{A}$$

$$\text{cotangent of } \theta = \cot\theta = \frac{A}{O}$$

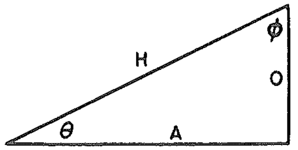

also

$$\sin\theta = \cos\phi \qquad \csc\theta = \sec\phi$$
$$\cos\theta = \sin\phi \qquad \sec\theta = \csc\phi$$
$$\tan\theta = \cot\phi \qquad \cot\theta = \tan\phi$$

and

$$\frac{1}{\sin\theta} = \csc\theta \qquad \frac{1}{\csc\theta} = \sin\theta$$
$$\frac{1}{\cos\theta} = \sec\theta \qquad \frac{1}{\sec\theta} = \cos\theta$$
$$\frac{1}{\tan\theta} = \cot\theta \qquad \frac{1}{\cot\theta} = \tan\theta$$

The expression "arc sin" indicates, "the angle whose sine is" . . .; likewise arc tan indicates, "the angle whose tangent is" . . . etc. See formulas in table below.

Known Values	Formulas for Determining Unknown Values of . . .				
	A	**O**	**H**	θ	ϕ
A & O			$\sqrt{A^2 + O^2}$	$\text{arc tan } \frac{O}{A}$	$\text{arc tan } \frac{A}{O}$
A & H		$\sqrt{H^2 - A^2}$		$\text{arc cos } \frac{A}{H}$	$\text{arc sin } \frac{A}{H}$
A & θ		$A \tan\theta$	$\frac{A}{\cos\theta}$		$90° - \theta$
A & ϕ		$\frac{A}{\tan\phi}$	$\frac{A}{\sin\phi}$	$90° - \phi$	
O & H	$\sqrt{H^2 - O^2}$			$\text{arc sin } \frac{O}{H}$	$\text{arc cos } \frac{O}{H}$
O & θ	$\frac{O}{\tan\theta}$		$\frac{O}{\sin\theta}$		$90° - \theta$
O & ϕ	$O \tan\phi$		$\frac{O}{\cos\phi}$	$90° - \phi$	
H & θ	$H \cos\theta$	$H \sin\theta$			$90° - \theta$
H & ϕ	$H \sin\phi$	$H \cos\phi$		$90° - \phi$	

Mathematical Symbols

$\times$ or $\cdot$	Multiplied by
$\div$ or :	Divided by
$+$	Positive. Plus. Add
$-$	Negative. Minus. Subtract
$\pm$	Positive or negative. Plus or minus
$\mp$	Negative or positive. Minus or plus
$=$ or ::	Equals
$\equiv$	Identity
$\cong$	Is approximately equal to
$\neq$	Does not equal
$>$	Is greater than
$\gg$	Is much greater than
$<$	Is less than
$\ll$	Is much less than
$\geqq$	Greater than or equal to
$\leqq$	Less than or equal to
$\therefore$	Therefore
$\angle$	Angle
Δ	Increment or Decrement
$\perp$	Perpendicular to
$\parallel$	Parallel to
$\lvert n \rvert$	Absolute value of n

Mathematical Constants

$\pi = 3.14$

$2\pi = 6.28$

$(2\pi)^2 = 39.5$

$4\pi = 12.6$

$\pi^2 = 9.87$

$\frac{\pi}{2} = 1.57$

$\frac{1}{\pi} = 0.318$

$\frac{1}{2\pi} = 0.159$

$\frac{1}{\pi^2} = 0.101$

$\frac{1}{\sqrt{\pi}} = 0.564$

$\sqrt{\pi} = 1.77$

$\sqrt{\frac{\pi}{2}} = 1.25$

$\sqrt{2} = 1.41$

$\sqrt{3} = 1.73$

$\frac{1}{\sqrt{2}} = 0.707$

$\frac{1}{\sqrt{3}} = 0.577$

$\log \pi = 0.497$

$\log \frac{\pi}{2} = 0.196$

$\log \pi^2 = 0.994$

$\log \sqrt{\pi} = 0.248$

Decimal Inches

Inches $\times$ 2.540 = Centimeters

Inches $\times$ 1.578×10^{-5} = Miles

Inches $\times$ 10^3 = Mils

Inches			Decimal Equivalent	Millimeter Equivalent
1/64			.0156	0.397
	1/32		.0313	0.794
3/64			.0469	1.191
		1/16	.0625	1.588
5/64			.0781	1.985
	3/32		.0938	2.381
7/64			.1094	2.778
		1/8	.1250	3.175
9/64			.1406	3.572
	5/32		.1563	3.969
11/64			.1719	4.366
		3/16	.1875	4.762
13/64			.2031	5.159
	7/32		.2188	5.556
15/64			.2344	5.953
		1/4	.2500	6.350
17/64			.2656	6.747
	9/32		.2813	7.144
19/64			.2969	7.541
		5/16	.3125	7.937
21/64			.3281	8.334
	11/32		.3438	8.731
23/64			.3594	9.128
		3/8	.3750	9.525
25/64			.3906	9.922
	13/32		.4063	10.319
27/64			.4219	10.716
		7/16	.4375	11.112
29/64			.4531	11.509
	15/32		.4688	11.906
31/64			.4844	12.303
		1/2	.5000	12.700
33/64			.5156	13.097
	17/32		.5313	13.494
35/64			.5469	13.891
		9/16	.5625	14.287
37/64			.5781	14.684
	19/32		.5938	15.081
39/64			.6094	15.478
		5/8	.6250	15.875
41/64			.6406	16.272
	21/32		.6563	16.669
43/64			.6719	17.067
		11/16	.6875	17.463
45/64			.7031	17.860
	23/32		.7188	18.238
47/64			.7344	18.635
		3/4	.7500	19.049
49/64			.7656	19.446
	25/32		.7813	19.842
51/64			.7969	20.239
		13/16	.8125	20.636
53/64			.8281	21.033
	27/32		.8438	21.430
55/64			.8594	21.827
		7/8	.8750	22.224
57/64			.8906	22.621
	29/32		.9063	23.018
59/64			.9219	23.415
		15/16	.9375	23.812
61/64			.9531	24.209
	31/32		.9688	24.606
63/64			.9844	25.004
		1.0	1.0000	25.400

MEMORANDA

MEMORANDA

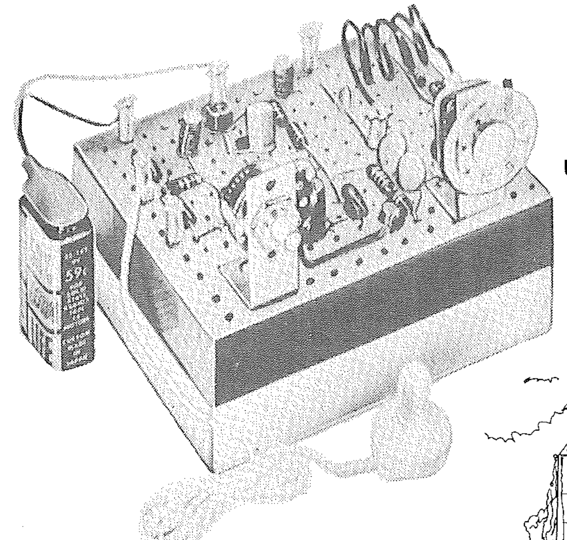

Science Fair®

America's Favorite Electronic Kits

Using the unique "P-Box"® concept, each kit is a useful device that demonstrates important electronic principles. Teacher-written instructions let you learn as you build! Top-quality components assure long-life and fine performance.

UNITED STATES

Each Kit with All Parts!

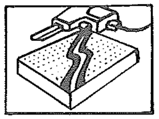

"P-Box" Material withstands soldering—remount parts or make wiring changes at will!

Quality Components!

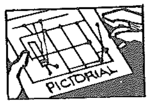

Detailed schematics and pictorial diagrams for easiest possible assembly!

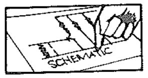

Waste-Free Package Design! Special pre-punched chassis "top" lets you mount parts easily; clear plastic bottom protects wiring!

22 P-BOX KITS AVAILABLE—PRICES LOW AS 1.49

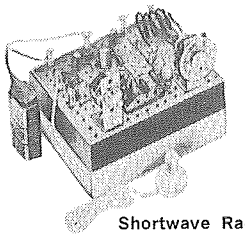

Shortwave Radio

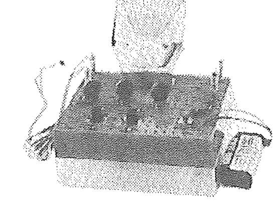

Electronic Thermometer

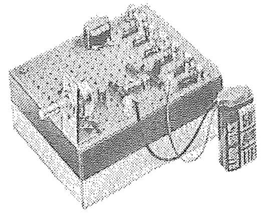

Electronic Organ

28-100. Single-Tube DC Radio.
28-101. Seven-Note Electronic Organ.
28-102. 2-Transistor Radio
28-103. Wireless AM Transmitter.
28-104. AC/DC Power Supply.
28-105. 3-Way Code Oscillator.
28-106. OTL Monophonic Amplifier.
28-107. "Rise-Fall' Siren Alarm.
28-110. 3-Transistor Shortwave Radio.
28-112. Bell/Buzzer Code Key.
28-116. Telephone Amp with Speaker.
28-118. Adujstable Metronome.
28-122. All-Purpose Electronic VOM.
28-123. 8-Ohm Extension Speaker.
28-126. Photoelectric Nite-Light.
28-128. Light-Operated Burglar Alarm.
28-129. Solar-Powered AM Radio.
28-130. "GoofyLite" Lamp Flasher.
28-131. Voice-Controlled Relay.
28-132. Moisture-Detector Relay.
28-133. WWV AM Radio Converter.
28-135. Indoor/Outdoor Thermometer.

The Professional Electronic Kits with the "Built-In Instructors"

Knight-Kit

"A staff of experts" guides your way to kit-building confidence. Detailed assembly manuals tell you what to do, when to do it, and a finished Knight-Kit compares to factory assembled units that cost twice as much!

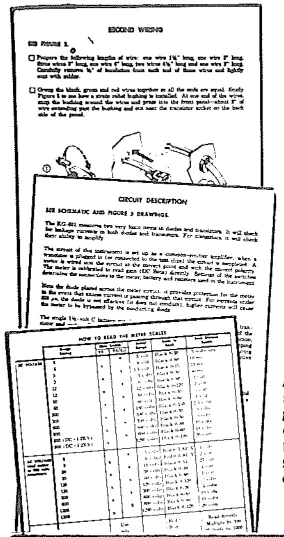

STEP-BY-STEP ASSEMBLY WITH CHECK-OFF SYSTEM
Oversize illustrations and diagrams cover each series of steps. Check box provided when step is completed so you always know where you are and what comes next.

OPERATING INSTRUCTIONS
Once you've assembled your Knight-Kit, the operating section of the manual shows you how to get the full benefit of every feature with detailed explanations and illustrations.

THEORY SECTION
An education in itself — this section provides the background in theory and electrical operation of your equipment. Really adds to the satisfaction of owning and operating your kit.

EXPERT SERVICE NATIONWIDE
In case you need it, service is as close as your neighborhood ARS store. We have factory-trained technicians and complete service facilities. Replacement parts are stocked at all 6 Regional Warehouses and available through all stores.

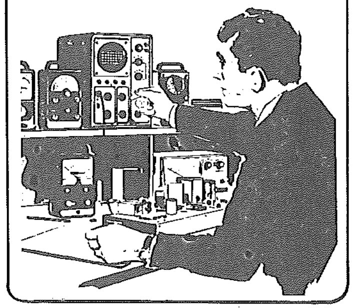

Knight-Kits for Every Need — You Build Them Better Because They're Designed Better!

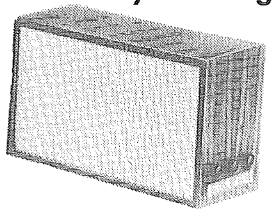

29-3387. Three-Channel Color Organ Kit. Reds, blues, greens dance to your music.

29-3383. Two-Station Wireless Intercom Kit Keeps You in Touch from Any Room in Your House.

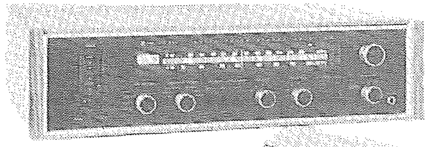

29-3950. 240-Watt AM/FM Stereo Receiver Kit. The most powerful you can buy! Every hi-fi feature.

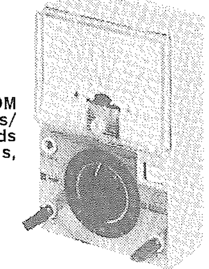

29-3986. Compact VOM Kit with 50,000 Ohms/Volt Sensitivity. Reads volts, current, Ohms, dB.

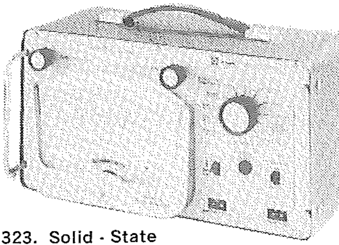

29-3323. Solid-State Auto Analyzer Kit for Tuneups, Troubleshooting.

Plus — 40 More!

Allied Radio Shack Will Give You Free Educational Material . . . You Buy Nothing!

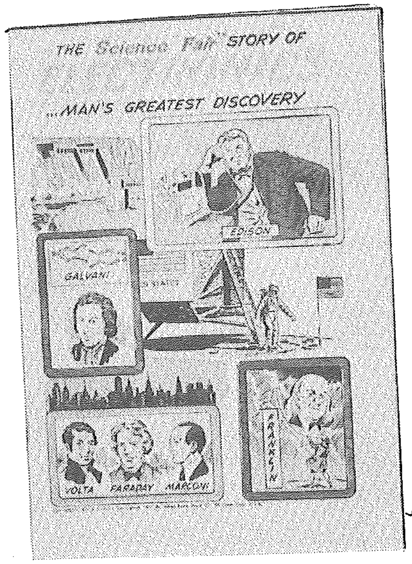
THE Science Fair STORY OF
...MAN'S GREATEST DISCOVERY
EDISON
GALVANI
FRANKLIN
VOLTA
FARADAY
MARCONI

TEACHERS & EDUCATORS

Special "No Strings" Offer!

"The Science Fair® Story of Electronics . . . Man's Greatest Discovery" is a unique, 24-page, full color "comic book" we're offering free—in any quantity—for your class or school! It touches on electricity in nature, the first electric light and telegraph, radio waves, the laser beam—and much more! It also covers the Science Fair Hall of Fame: Edison, Faraday, Franklin, Marconi, Morse—to mention a few! Here's the potential to spark a real interest in science for all students.

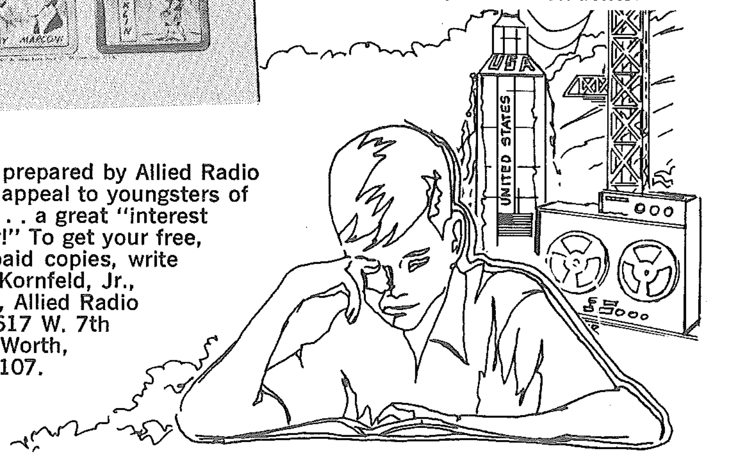
USA
UNITED STATES

Specially prepared by Allied Radio Shack to appeal to youngsters of all ages . . . a great "interest generator!" To get your free, postage-paid copies, write
Lewis F. Kornfeld, Jr.,
President, Allied Radio Shack, 2617 W. 7th St., Fort Worth,
Texas 76107.

Get Acquainted With Sixteen "Great Men In Electronics"

Biographies! Photos! Illustrations!

Now available at a very special price! An attractive 64-page reference book containing biographies of men who pioneered supreme achievements in electronics . . . Morse, Bell, Sarnoff, DeForest, Hammond, and others! Concisely written, factual — an ideal guide for teachers, students, hobbyists. Mail your order to: Electronics Digest — Great Men, P. O. Box 9108, Fort Worth, Texas 76107. Only $1 postpaid — and well worth it!

Reg. 1^{95}

Now 1^{00}

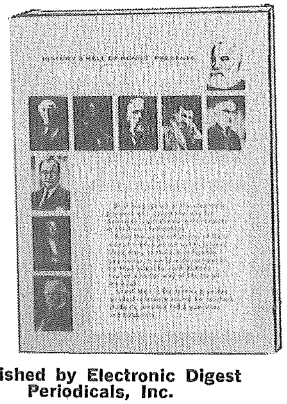

Published by Electronic Digest Periodicals, Inc.

ALLIED RADIO SHACK Technical Publications

A "Must" for Hobbyist and Pro

Learn more about the electronics field that interests you!

Authoritative publications written by industry leaders. All are written in easy-to-understand laymen's language. Expand your knowledge of the particular electronics field that interests you. Get yours today.

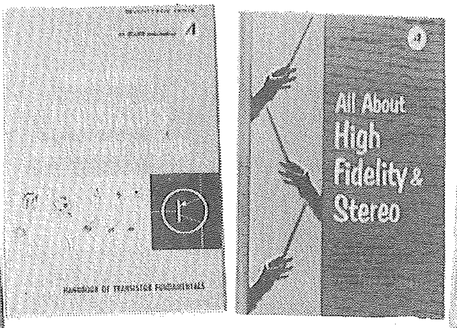
ALLIED
ELECTRONICS DATA HANDBOOK
FORMULAS AND DATA COMMONLY USED IN ELECTRONICS
HANDBOOK OF TRANSISTOR FUNDAMENTALS
All About High Fidelity & Stereo

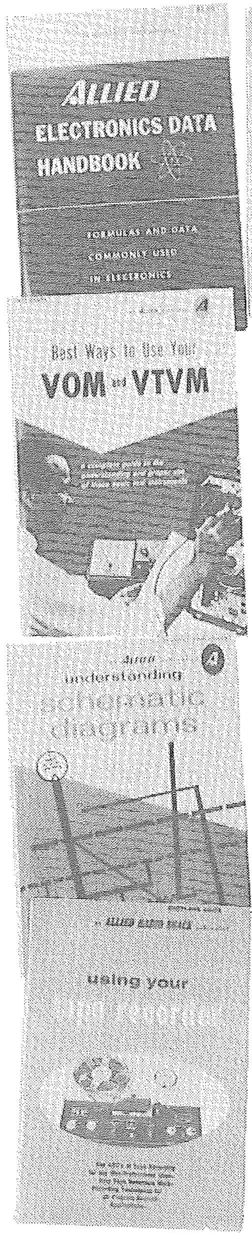
Best Ways to Use Your VOM and VTVM
understanding schematic diagrams
using your tape recorder

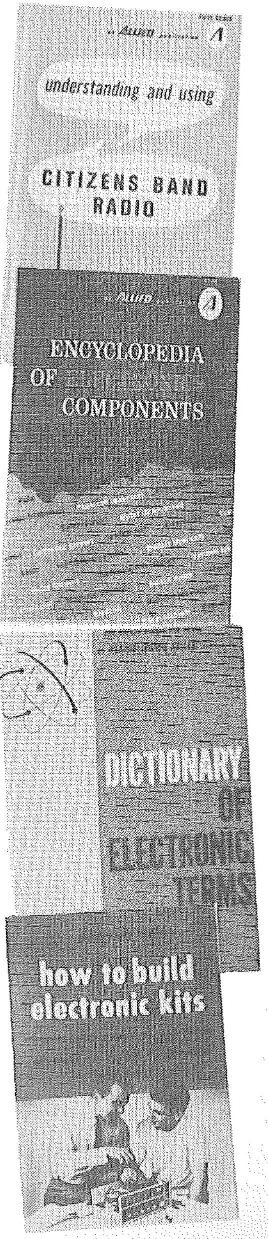
understanding and using CITIZENS BAND RADIO
ENCYCLOPEDIA OF ELECTRONIC COMPONENTS
DICTIONARY OF ELECTRONIC TERMS
how to build electronic kits

TUBE SUBSTITUTION HANDBOOK. 96 pages. 62-1326. ....1.75

TRANSISTOR SUBSTITUTION HANDBOOK. 160 pages. 62-1327. ....2.25

SHORTWAVE LISTENER'S GUIDE. 96 pages. 62-1328. ....1.95

RADIO LICENSE HANDBOOK. 136 pages. 62-1329. ....2.50

TV TUBE SYMPTOMS AND TROUBLESHOOTING. 96 pages. 62-1330. ....1.95

CB/HAM/SWL LOG BOOK. 80 pages. 62-1364. ....1.00

HOW TO BUILD ELECTRONIC KITS. 96 pages. 62-7066. ....95¢

BEST WAYS TO USE YOUR VOM AND VTVM. 96 pages. 62-7067. ....95¢

ELECTRONICS DATA HANDBOOK. 112 pages. 62-7398. ....1.25

UNDERSTANDING TRANSISTORS AND TRANSISTOR PROJECTS. 112 pages. 62-7602. . .95¢

INTEGRATED CIRCUIT FUNDAMENTALS AND PROJECTS. 96 pages. 62-7625. ....95¢

ALL ABOUT HIGH-FIDELITY AND STEREO. 96 pages. 62-7699. ....95¢

UNDERSTANDING AND USING CB RADIO. 112 pages. 62-7701. ....95¢

GETTING STARTED IN ELECTRONICS. 112 pages. 62-7705. ....95¢

USING YOUR TAPE RECORDER. 112 pages. 62-7730. ....95¢

DICTIONARY OF ELECTRONIC TERMS. 112 pages. 62-7756. ....1.25

ENCYCLOPEDIA OF ELECTRONIC COMPONENTS. 128 pages. 62-7930. ....1.25

UNDERSTANDING SCHEMATIC DIAGRAMS. 112 pages. 62-7931 ....95¢

UNDERSTANDING AND USING YOUR OSCILLOSCOPE. 112 pages. 62-7932. ....95¢

Prices Slightly Higher in Canada and Alaska

At Over 1000 Allied Radio Shack Stores Nationwide

Over 1000 ALLIED RADIO SHACK Stores in 49 States and Canada

Stop in at the Store Near You . . or Order by Mail!

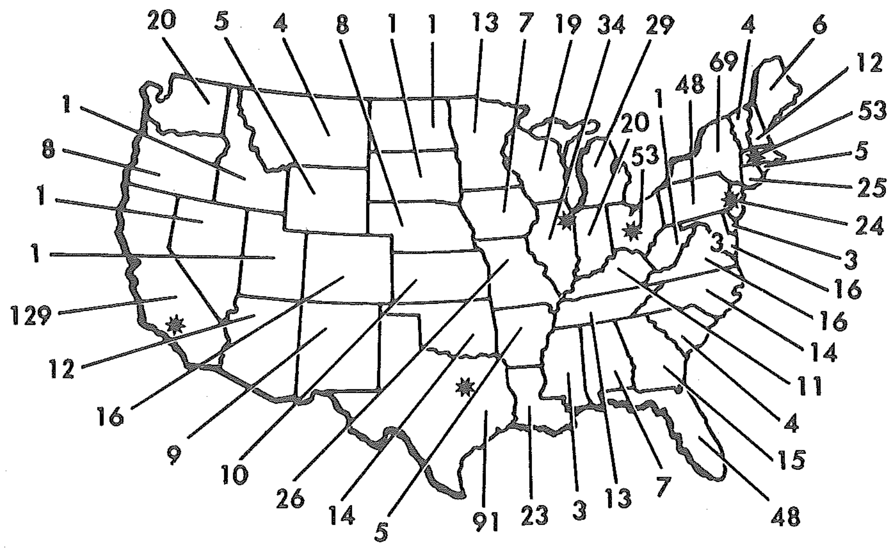
20
5
4
8
1
1
13
7
19
34
29
4
6
69
48
1
20
53
12
53
5
25
24
3
16
16
14
11
4
15
48
1
8
1
1
129
12
16
9
10
26
14
5
91
23
3
13
7
3

the most convenient shopping ever
from the world's largest electronic store chain

1000 STORES

- With managers who understand electronic needs!
- with quantity-buying power for the best values!
- with exclusive products sold nowhere else!
- with quality standards and quality controls!

*** 7 Giant Regional Warehouses strategically located to keep all our stores stocked with everything we sell. No wasted trips—You get what you want, when you want it, at any Allied Radio Shack Store!**

(Prices slightly higher in Canada and Alaska)

Four manufacturing plants in the U. S. and abroad—we make most of our own products—and pass the savings on to you!

ALLIED RADIO SHACK A TANDY CORPORATION COMPANY

At Over 1000 Allied Radio Shack Stores Nationwide

MODERN DICTIONARY OF ELECTRONICS

THIRD EDITION

Rudolf F. Graf

This edition of the *Modern Dictionary of Electronics* is the culmination of a large effort to bring the vocabulary of electronics into step with the recent rapid growth of the field and its allied areas.

New information has been analyzed and used in the updating of entries in the second edition of this dictionary. As with the previous editions, all new terms were carefully considered to determine common usage and meaning. The dictionary now contains approximately 16,500 terms used in electronics and related fields.

A unique guide which indicates correct pronunciations and syllabic divisions for over 1100 selected words is a special feature of the dictionary. Immediately following the pronunciation guide are the graphical symbols that are being used in electronic and electric schematic diagrams. Symbols and abbreviations for the more commonly used electronic terms and the Greek alphabet are given inside the front cover. Semiconductor symbols and abbreviations that have been adopted as standard by the Electronic Industries Association (EIA) and the National Electrical Manufacturers Association (NEMA) are included inside the back cover.

All terms are set in boldface, with marginal extension for ease in locating. Illustrations have been inserted where they will add greater meaning to definitions. Many entries have been cross-referenced to speed location of definitions.

The *Modern Dictionary of Electronics* was especially designed to be a comprehensive reference volume for engineers, technicians, instructors, editors, writers, and students. Since spelling, hyphenation, capitalization, etc., have been adopted from the style most consistently used by the industry, this dictionary also will be adopted as a manual of style by stenographers, secretaries, salesmen, report writers, and consultants.

Rudolf Graf has spent over 20 years in the electronics industry—holding positions ranging from instructor, consultant, magazine editor and design and development engineer, to sales engineer and director of engineering. He is holder of a first-class radiotelephone operator's license. He is also the author of *ABC's of Electronic Test Probes*, a popular SAMS book.

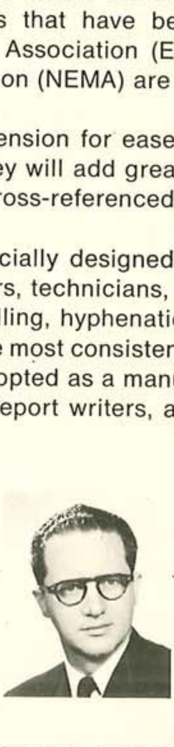

ALLIED RADIO SHACK A TANDY CORPORATION COMPANY

Almit Technology Ltd

- HAND SOLDERING TRAINING COURSE

Almit Technology Ltd

- Aim
  - Review the whole soldering process to improve soldering quality.
  - Aim for high first time pass rates.
- How to approach soldering
  - Regardless of what product an electronic assembler makes, we can say they really only make solder joints. It’s the process that brings everything electrical together.
  - Every joint should be made with care. If not then you may be able to ‘weed’ out the poorest during inspection/test but some will pass and they will be the weakest bit of the product and will fail in service, especially in automotive products where temperature & vibration can be considerable
  - In service failures are expensive to repair and always have other negative effects on the business.
- All customers know the solder joint is potentially the weakest part of the product.

Contents

1. What is soldering?
2. Cored Solder Wire
3. Soldering Irons
4. Soldering Workstation
5. Soldering Process
6. Inspection +Handling
7. Practical Training

1-1. What is soldering?

- Formation of a “metal to metal” joint using solder.
- The joint is made by alloy formation of base metal and solder.
- In soldering, there are three key elements :

Heat , Solder , Flux , they are all important.

Flux cored solder wire is used for hand soldering.

1-2. Flux Function

What does it do?

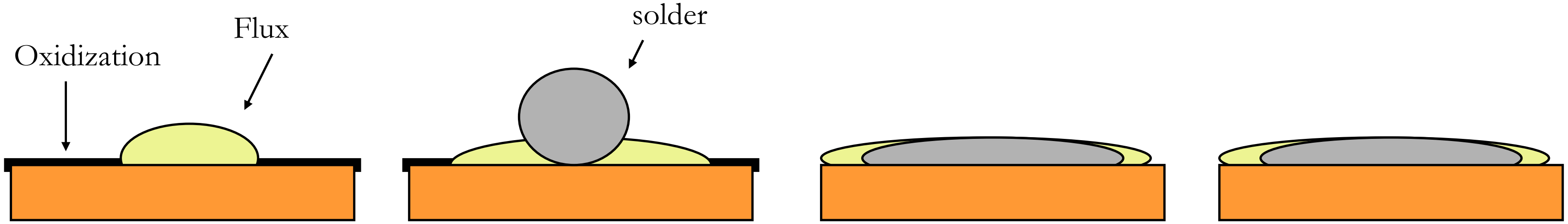

Remove oxide

Remove metal oxide to support wettability of solder.

Prevent from re-oxidation

Coat metal surface to prevent from re-oxidation.

Reduce surface tension

Reduce surface tension of solder to allow it to spread onto pad/lead.

Solder finish

Make smooth surface finish of solder and prevent solder bridges.

During soldering a chemical reaction takes place. The flux removes all the surface tarnish leaving clean metal underneath.

1-3. Solder Wetting

Check the wetting angle

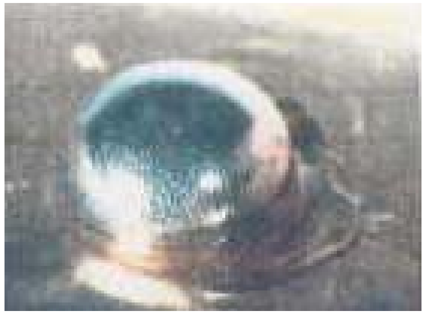

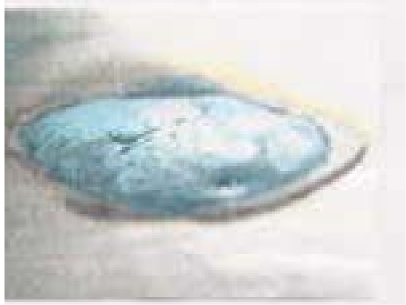

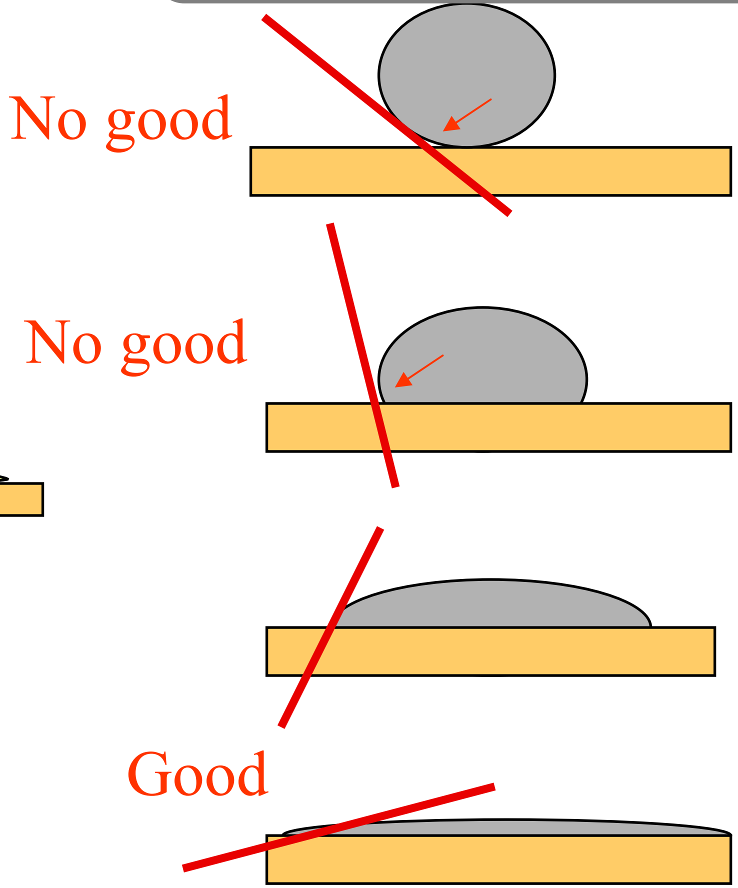
No good
No good
Good

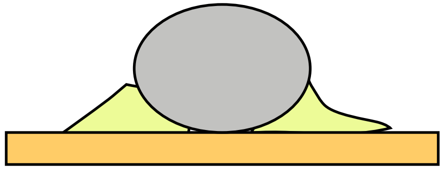

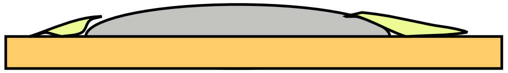

Not wetted

Not soldered

(dry joint)

Wetted

soldered

1-4. Melting and Diffusion

- Alloy is formed by melting and diffusion of metals
- Tin and copper alloy together

Alloy layer in good condition

Diffusion

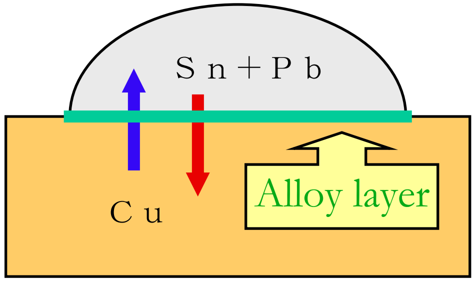

Melting

It takes a certain amount of time to create the perfect solder joint. Too slow a process can damage pcb/components and too fast will not create a ‘sound’ joint.

Thin and even thickness of alloy layer

2-1.Cored Solder Wire

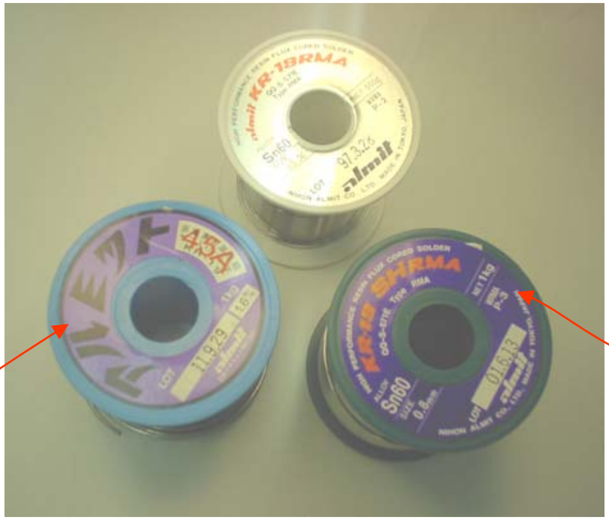

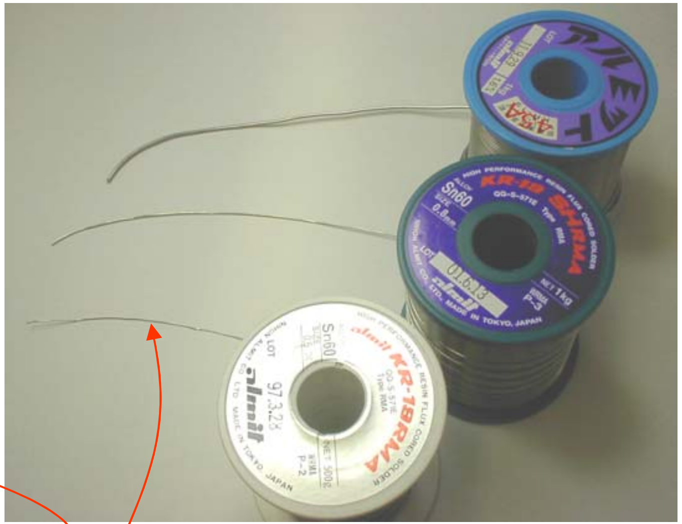

Solder type
Solder name

Diameter
0.5 ϕ ~1.6 ϕ

Alloy Composition
Sn60,Sn50,Sn45
PbFree Alloy(Sn-Ag-Cu)

Always confirm before use. (Use of different solders and diameters depends on the products being soldered)

2-2. Construction of Cored Solder

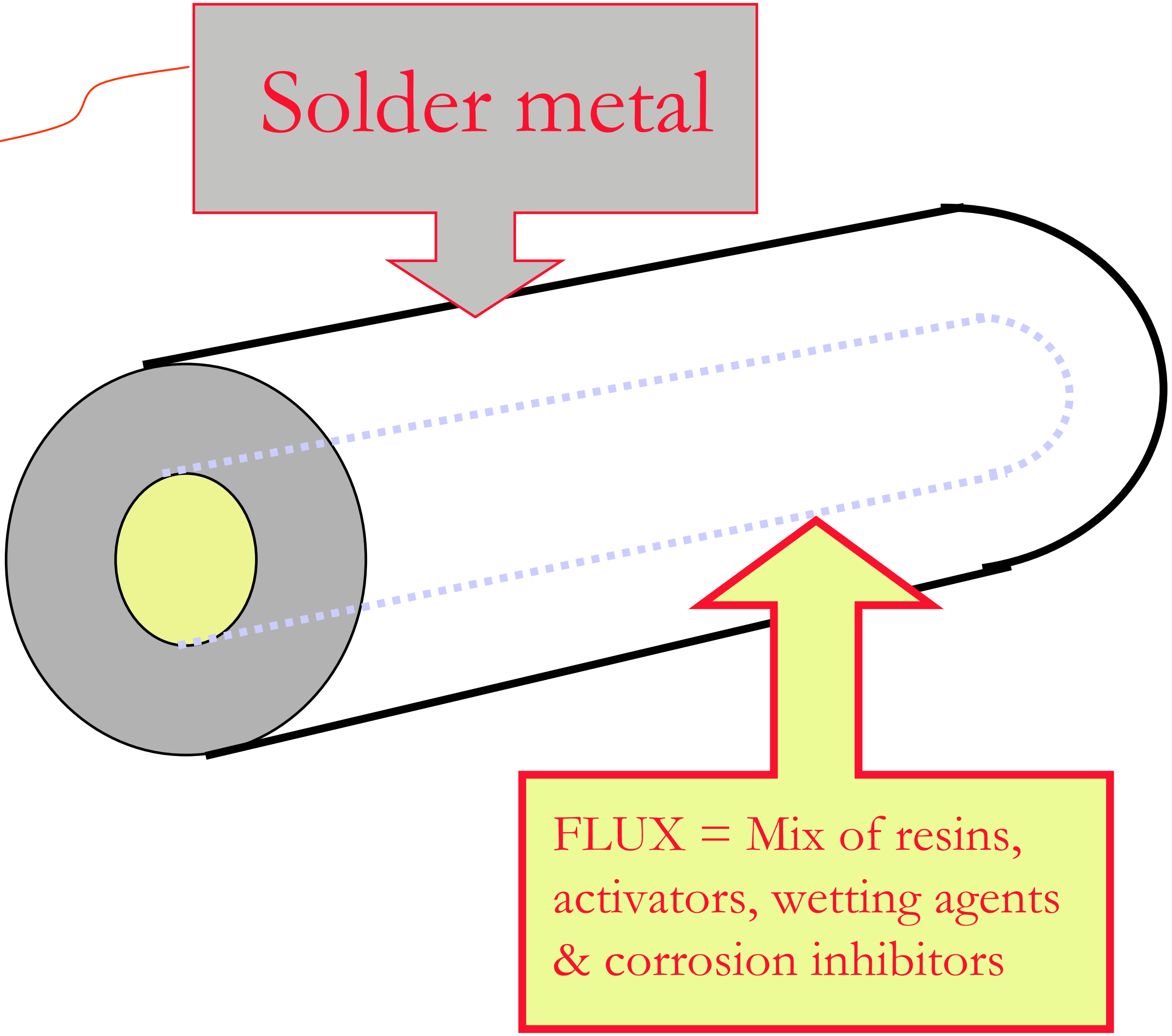

Attention
Sn-Pb series
PbFree series

Solder metal

FLUX = Mix of resins, activators, wetting agents & corrosion inhibitors

2-3. Solder Wire Variables

- Solder Alloy
  - Most use Sn60/Pb40 or Sn62/Pb36/Ag2 or LEAD FREE
- Flux Type.
  - No-Clean RMA (ALMIT KR19 SH)
    - Flux residues are safe to leave on pcb. They will not cause corrosion or electrical breakdown of cct during its' lifetime.
- Flux %
  - 1,2 or 3. Higher flux % makes soldering easier but can leave more flux residue which can make solder joint look cosmetically 'dirty'
- Wire Diameter
  - Select the right one for the job.
- Reel Size
  - Generally 500g

2-4. Handling method for cored solder wire

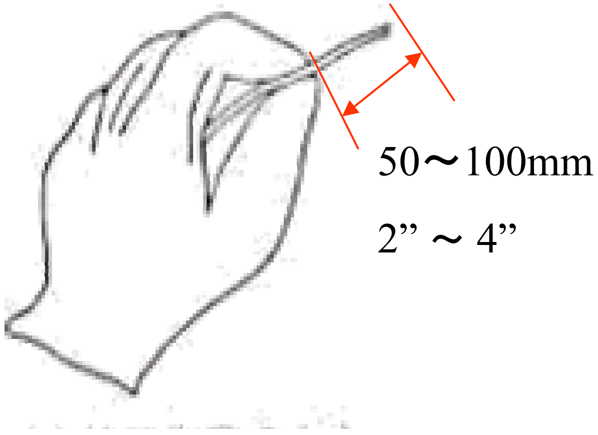

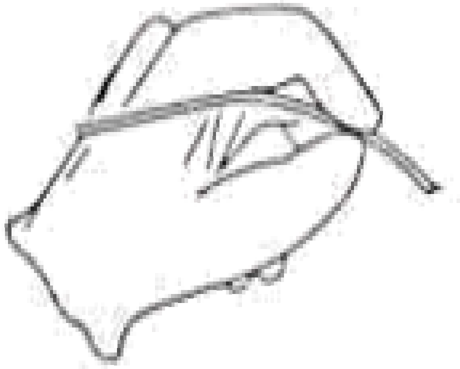

Easy feeding for continuous process.

Not suitable for continuous process.

3-1. Soldering Irons

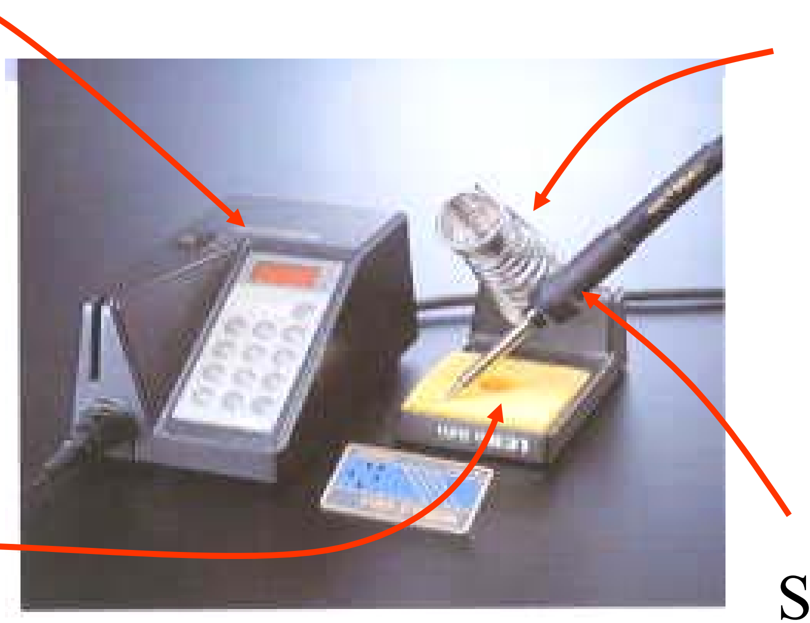

Soldering station

Temperature setting differs depending on application.

Necessary to check the temperature before starting the process.

Iron holder

Soldering Iron

Sponge for cleaning soldering iron tip.

The soldering iron is critical to good soldering. It must provide all the heat to heat up the joint which in turn must heat up the solder wire/flux

3-2. Soldering Irons

Differences in power capacity, size and ability.

Use one depending on the type of work . 80 W

Weller better for Pb Free or large joints

Iron tip

Iron tip shape is different for each process.
Choose correct iron tip for application.

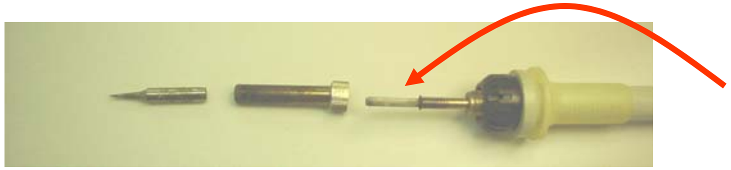

Internal part of soldering iron

Ceramic heater

Weak to impact / shock.

3-3. How to hold Soldering Iron

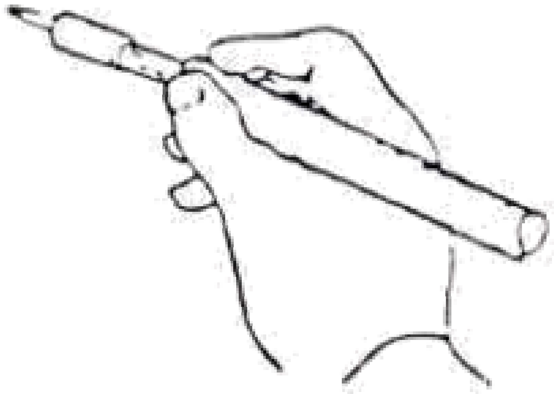

Pen-holder type
For normal operation

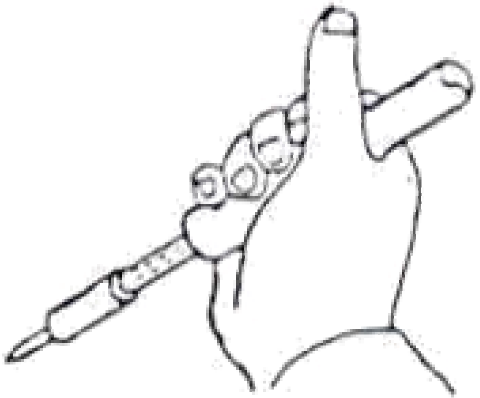

Grip type
For large component

3-4. How the tip is made

The most important part of the soldering iron is the TIP.
The main part of the tip is made of copper. To extend the lifetime of the tip, it is iron coated .
Before the iron is added the copper is nickel plated. The non wettable part of the tip is chrome plated. The chrome stops the solder flowing up the tip. The tip is pre-tinned with lead free solder.

Nickel coating: ca. 10 - 20μm
Iron coating: ca. 150 - 300μm
Chrome Coating:ca. 3 - 6μm

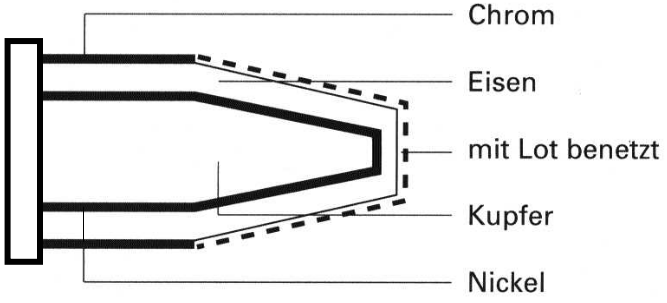

3-5. How to clean Soldering Iron

Cut sponge in V.
Use the cut part to clean

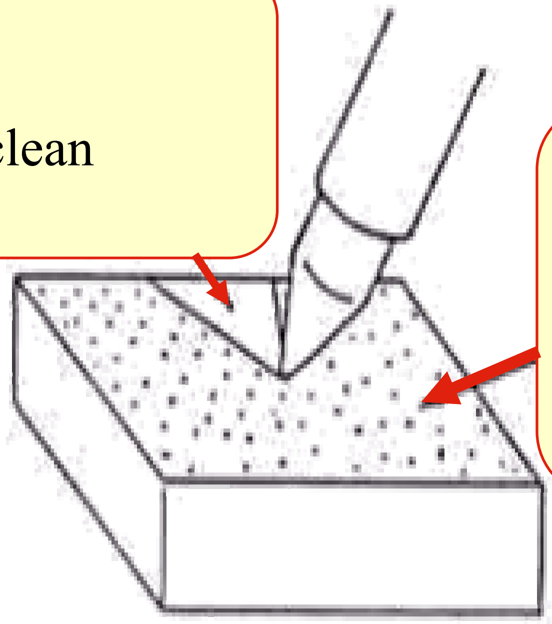

Use water or Tip Cleaning Fluid to clean.
Do not use to much or to little. Keep tip tinned with Tip Cleaning Tin.

If tip not properly cleaned this can cause flux splatter, solder waste and poor heat conduction to joint area which can cause defects.
May need to increase the number of cleaning times when using Pb Free solder.

3-6. Temperature when cleaning

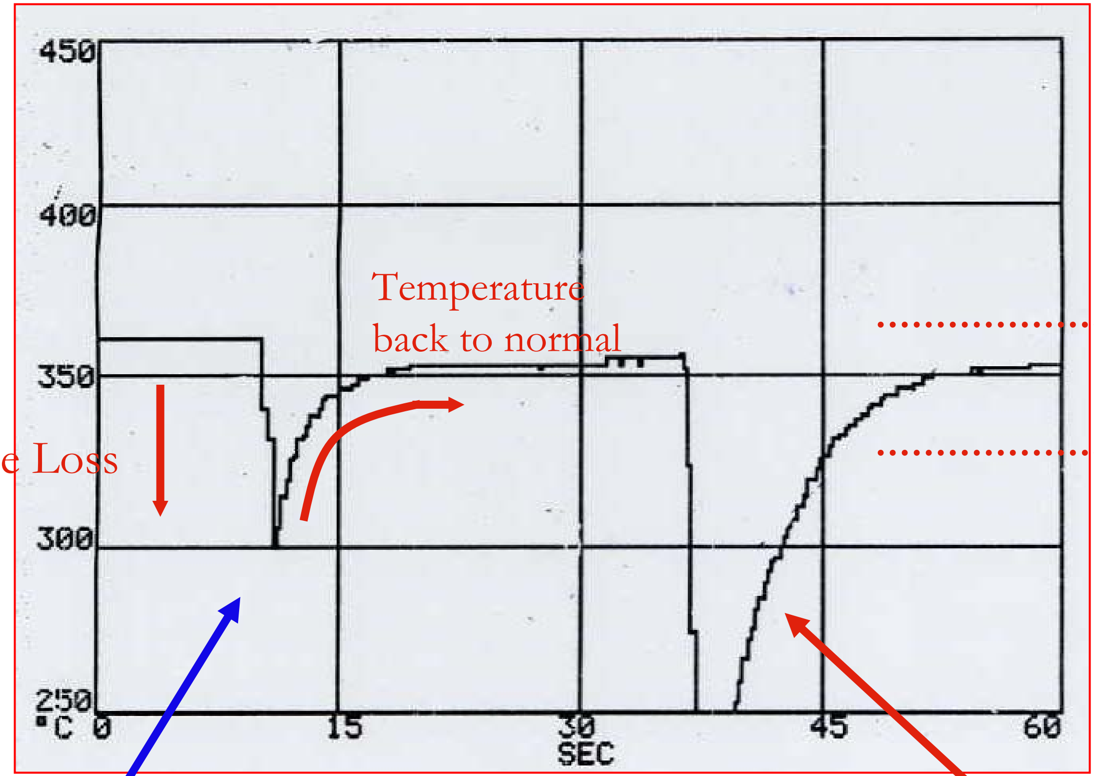

Temperature back to normal

Temperature Loss

(Ex)

Ideal range of operation

350℃±15℃

Good condition

To start operation after temperature back to normal and consistent..

Too much water

Need longer time if temperature reduce too much.

3-7. Temperature Stability

Each time a solder joint is made the heat in the soldering iron is depleted especially if joints are made ‘quickly’

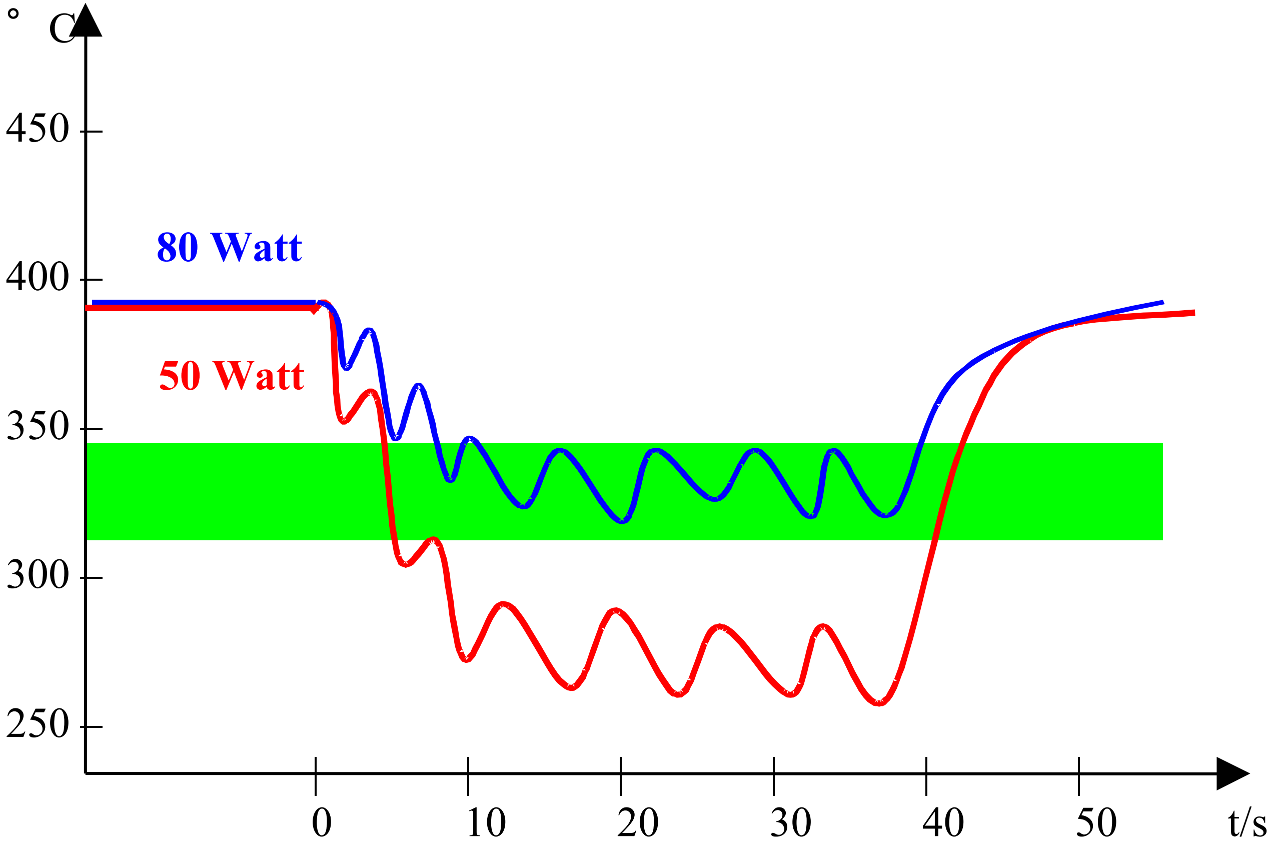

3-8. Solder Iron Tip Condition

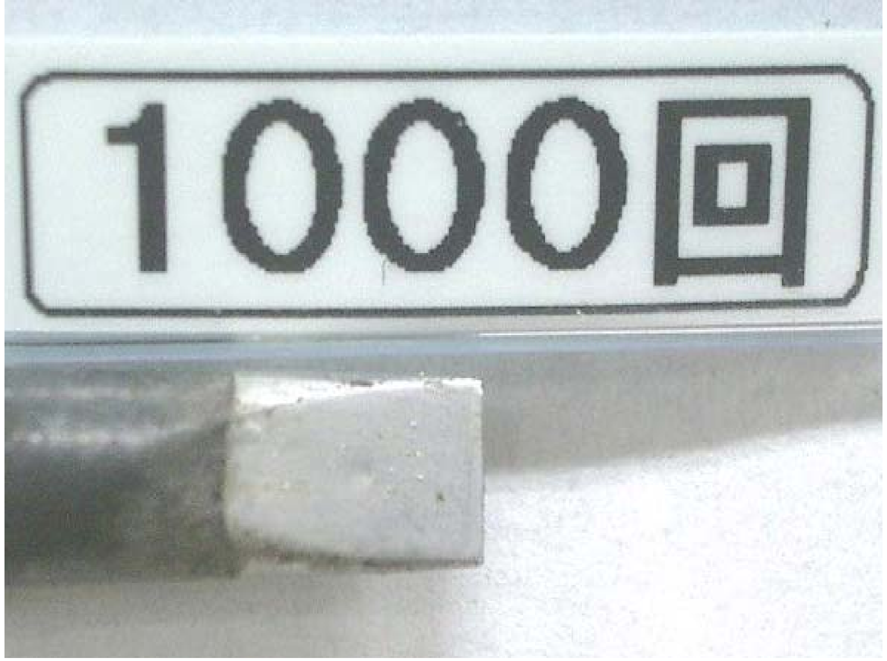

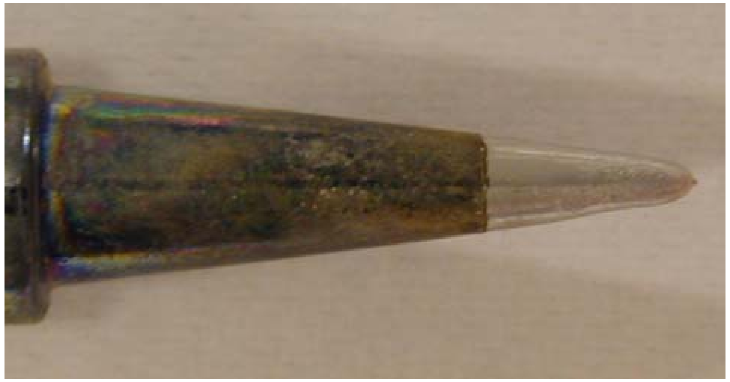

This tip has soldered 20000 joints. Still in perfect condition

This tip has soldered 1000 joints. The tinned area is still shiny and smooth. There is no charred flux adhering to it. Only use the tinned area to heat up joint area. It helps the heat to be transferred quickly/consistently. Using the none tinned area of the iron will cause soldering problems

3-9. Solder Iron Tip Condition 2

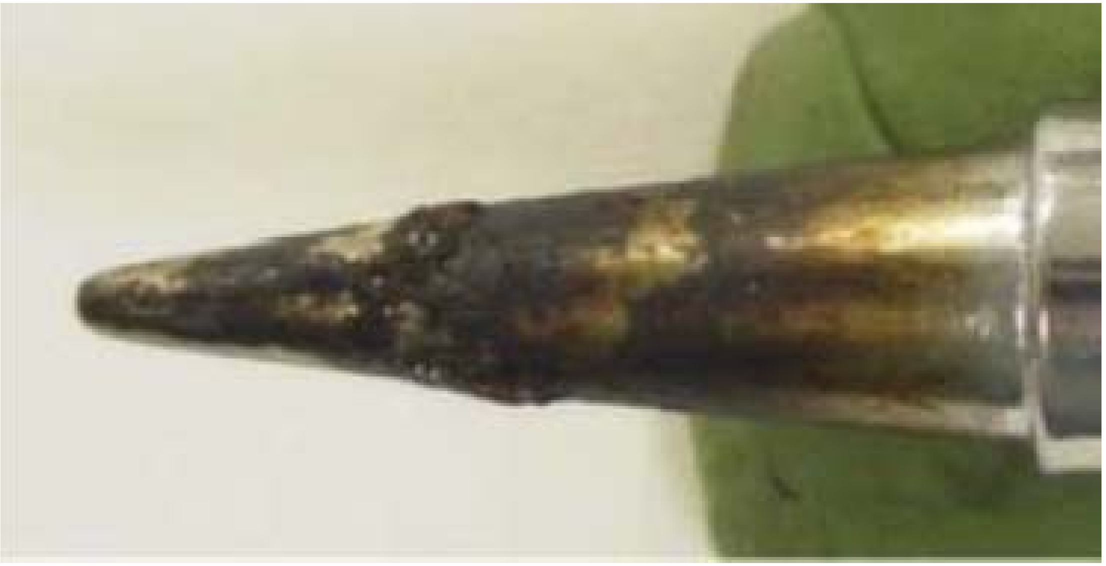

The tinned area is dull and inconsistent and there is lots of charred flux adhering to it. The ability of this tip to transfer heat to the solder joint is significantly impaired by its poor condition.

4-1. Soldering Workstation

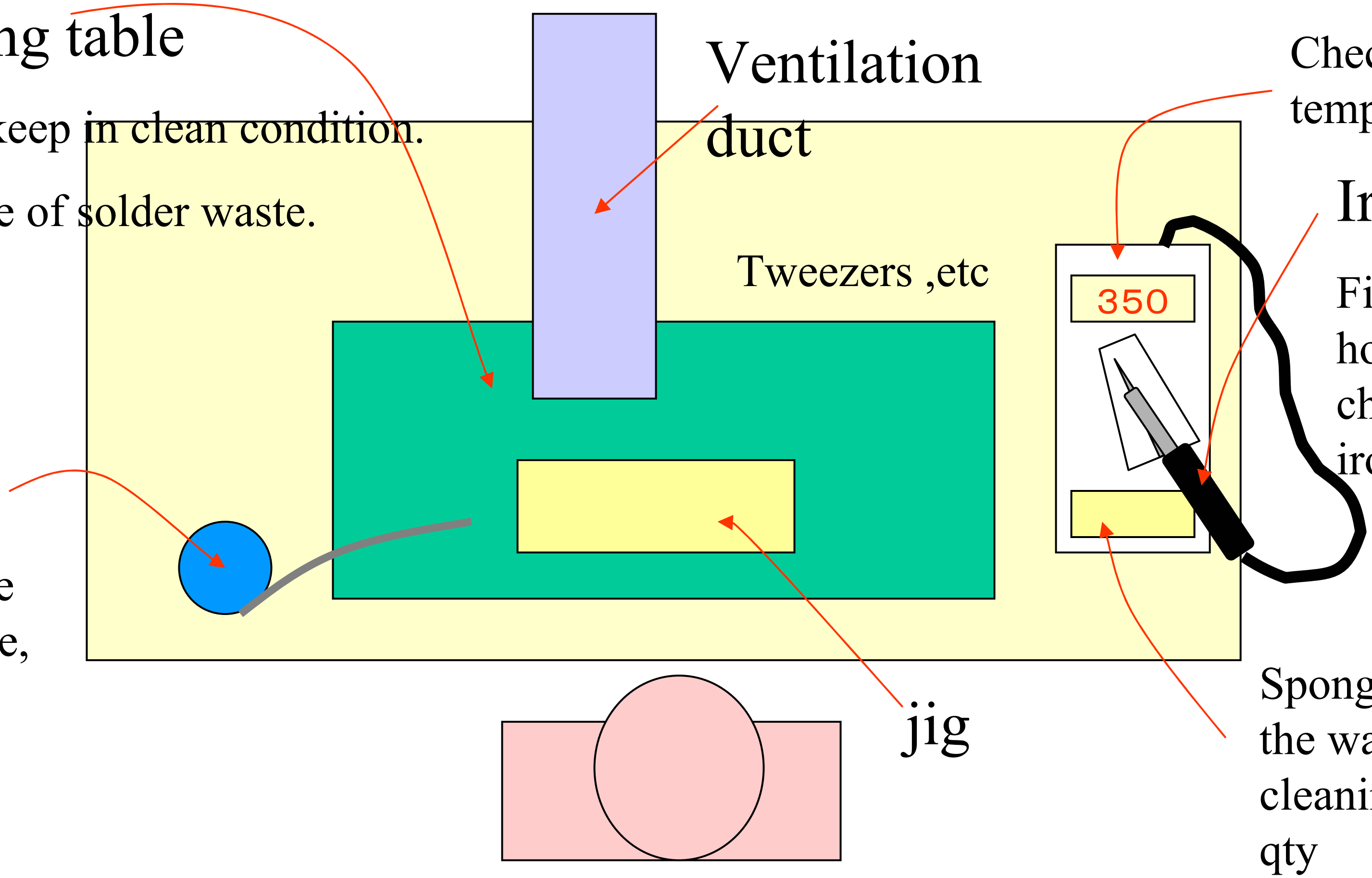

Working table

Always keep in clean condition.

Take note of solder waste.

Ventilation duct

Tweezers ,etc

Check temperature

Iron

Fix the holder and check the iron tip.

solder

Check the Sn%, type, diameter

jig

Sponge. Check the water or cleaning fluid qty

4-2. Correct posture when soldering

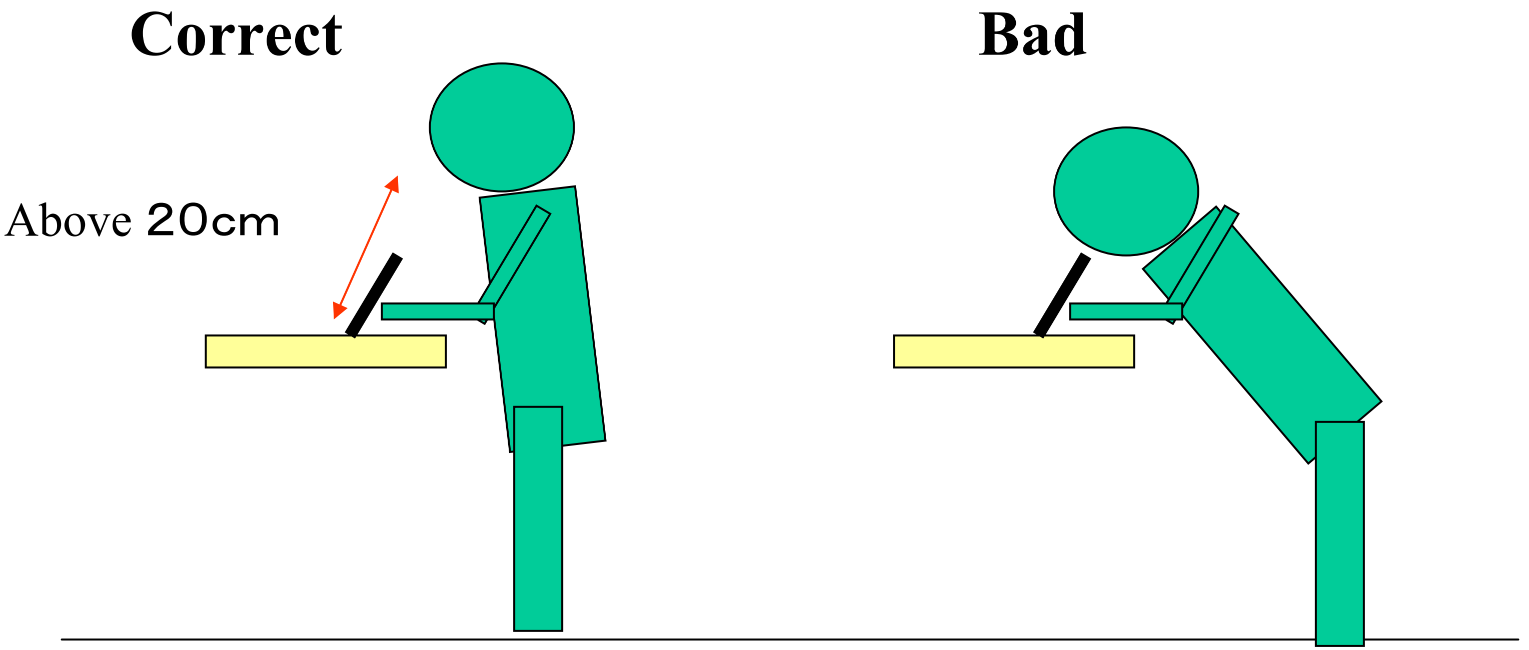

Must have clear view of work.
Feeding in wire and iron movement
must be smooth and easy.

Not suitable. Poor position for 'back pain'. Flux fumes could be inhaled.

4-3. Considerations to create best workstation

- ESD Bench
  - Must be ESD safe to eliminate static
  - Keep it clean/tidy
- Lighting
  - Keep it as natural as possible.
- Fume Extraction
  - Keep the filters clean
- Iron/wire Position
  - Must be positioned to allow smooth access to solder joint area.
- Work Jig
  - Must be free from flux residues etc

4-4. Summary so far..

- Check the solder type
- Check the solder iron
- Check the cleaning sponge
- Soldering posture

5-1. Soldering

5 step method

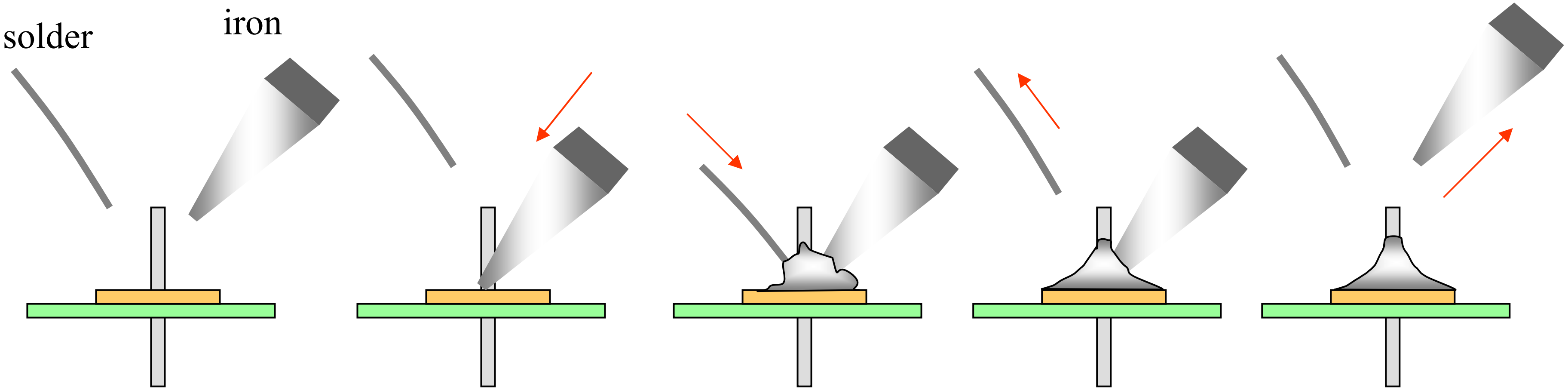

preparation

heat

Use iron tip to heat the soldering part.

Insert solder

Melt properly

Pull out solder

Pull out iron tip

Solder will not alloy if the parts and solder are not fully heated.

The soldering iron is not just to heat and melt the solder, but to heat the part being soldered.

5-2. Soldering

5 step method

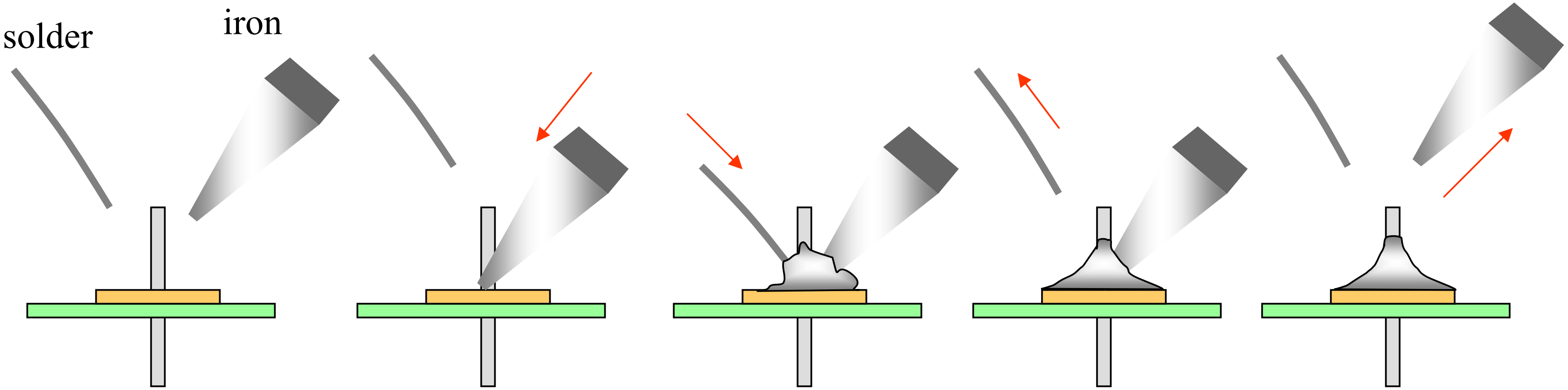

preparation

Heat up to @ over 100°C.

May take @ 2 secs

Insert solder to hot joint area. Flux melts and covers joint. Flux helps melt solder and it flows over clean pad/lead. Solder area @ 230-240°C.

Pull out solder

Pull out iron tip

5-3. Heating Large Joint

Solder large terminal/area by moving iron tip

Use iron tip body to heat
up the joint, but the area
in contact with the joint
must be tinned and clean.

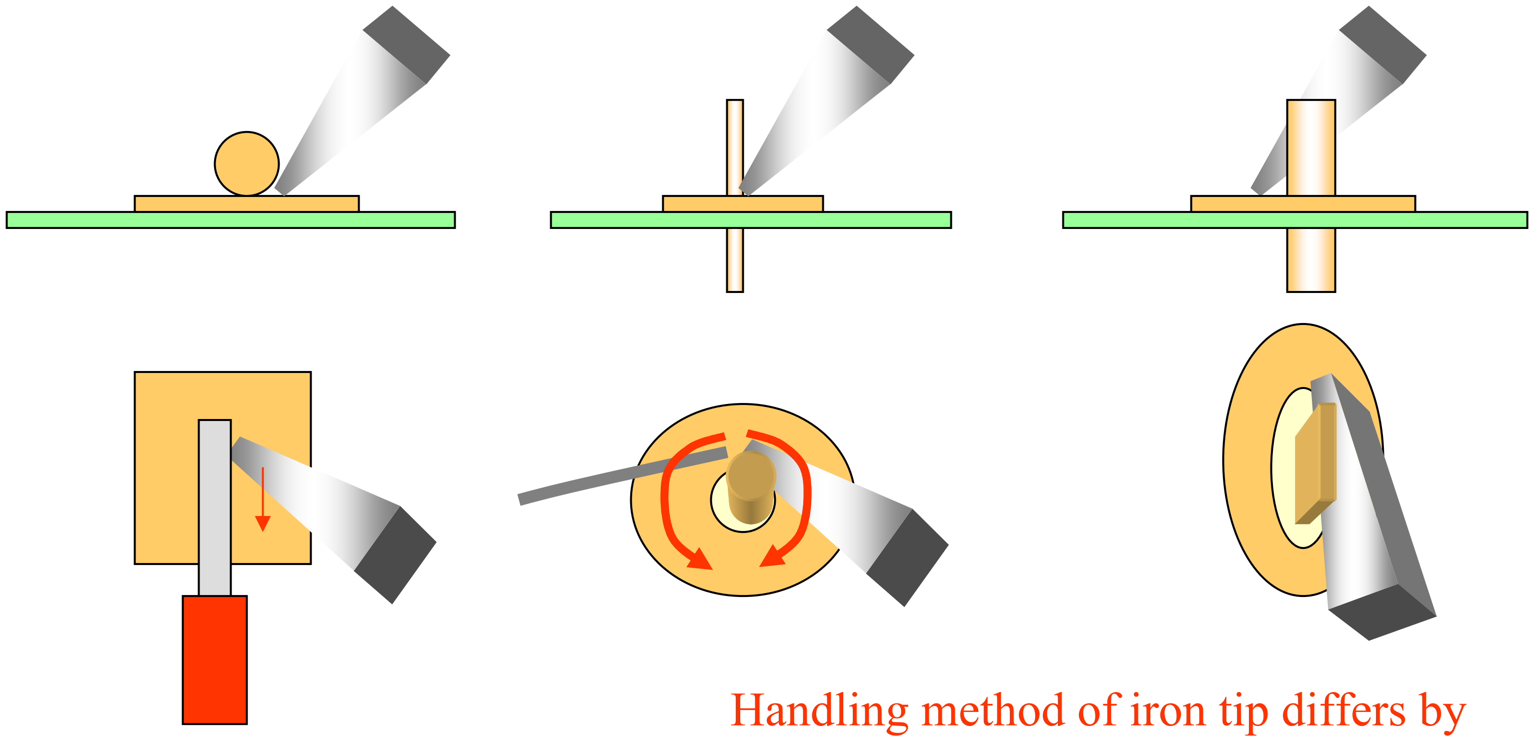

Handling method of iron tip differs by
shape of terminal.

5-4. Solder Wetting

Good

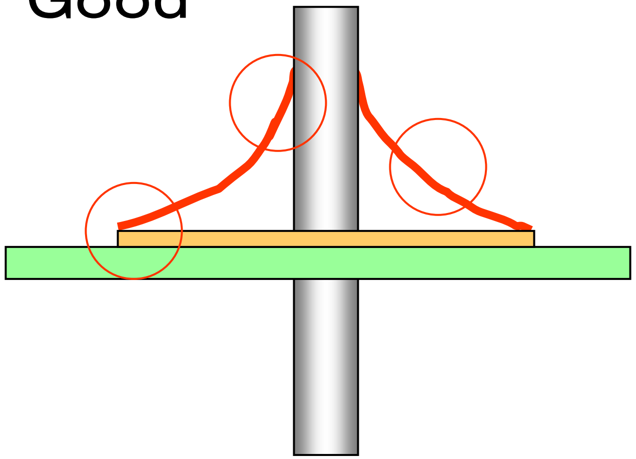

Solder wets to lead and land.

Smooth and shiny surface.

No holes, no spikes.

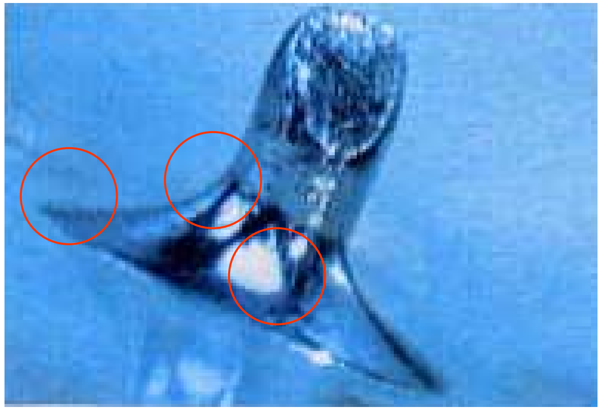

No good

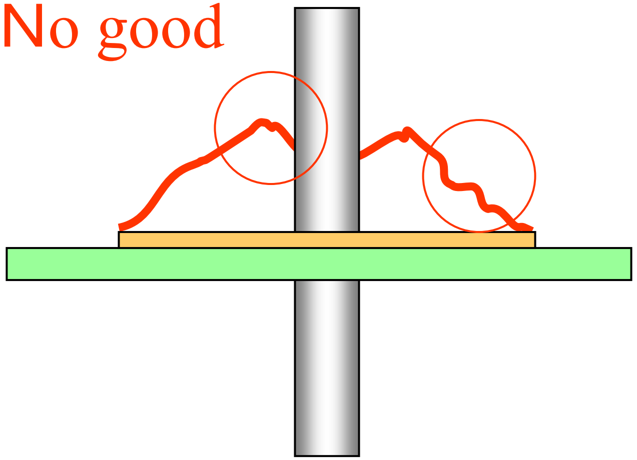

Poor wetting to lead.

Rough surface.

Holes.

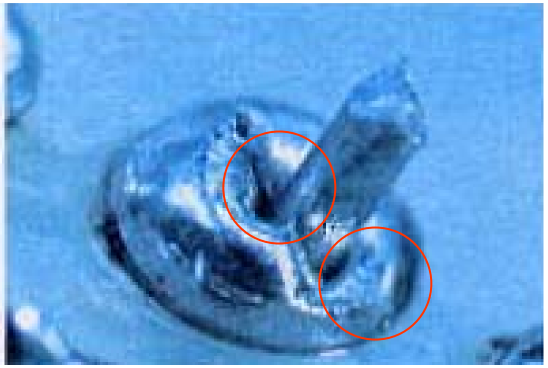

5-5. Solder Quantity

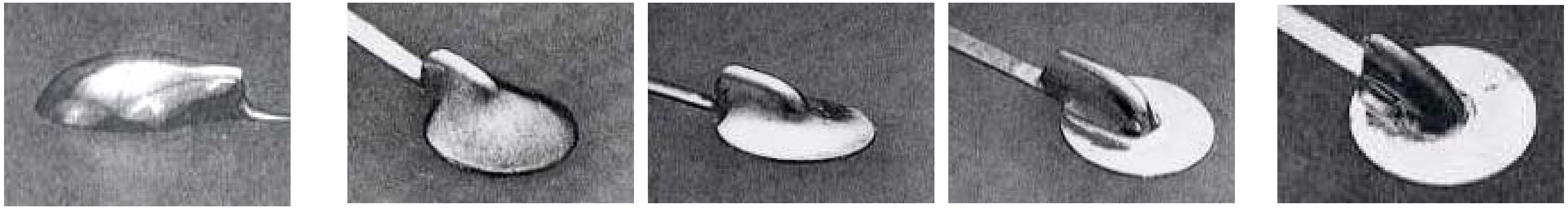

No good | OK | No good

Excess amount

Unable to
confirm whether
solder is wetted.

Apply correct solder amount.

No good if too much or too little.

Lack of solder

Solder joint is
weak, and
leads to crack.

5-6. Solder Quantity and Surface Condition 1

Sn60

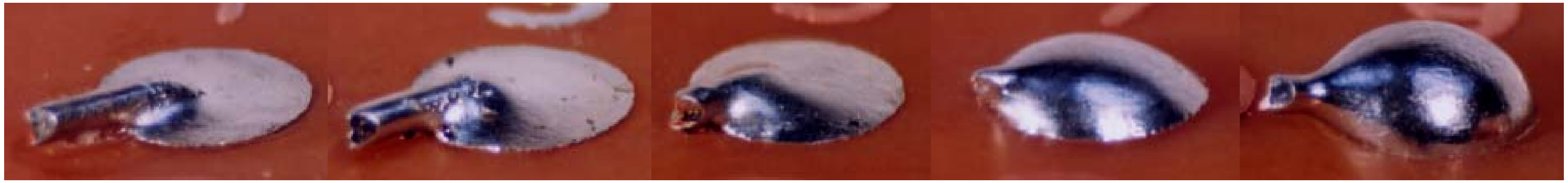

Pb Free

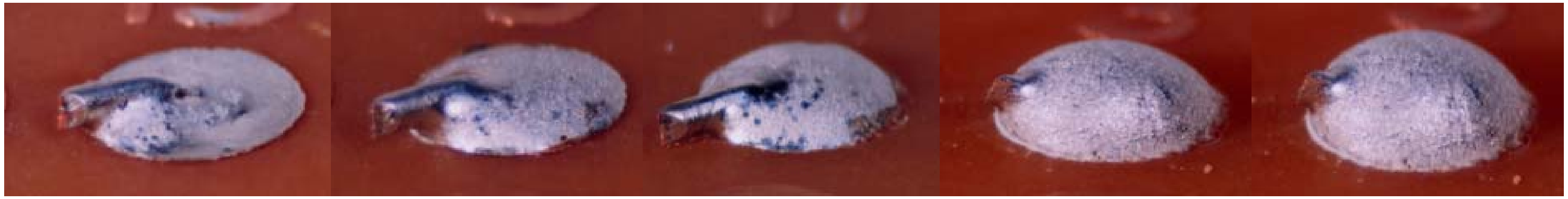

Small amount Large amount

5-7. Solder Quantity and Surface Condition 2

Sn60

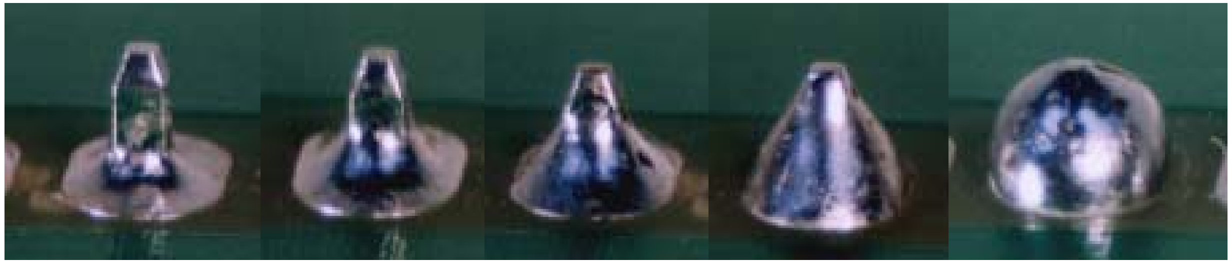

NG NG

Pb Free

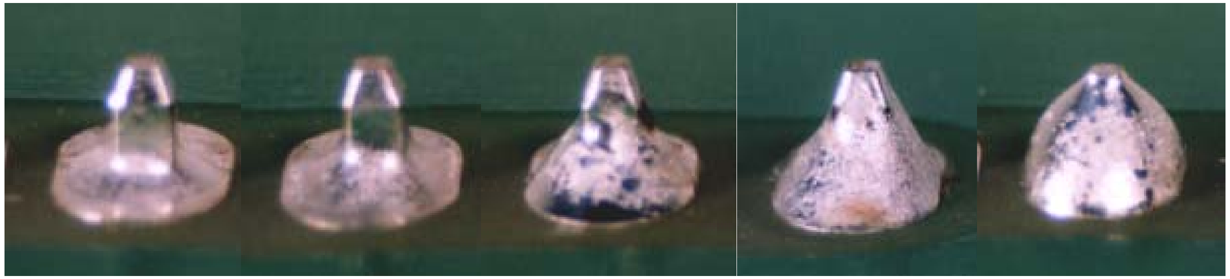

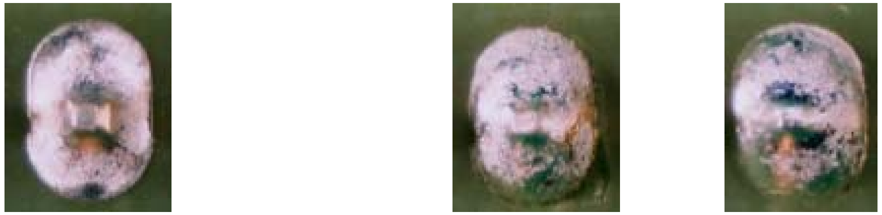

5-8. Key Points for Heating

- What is the appropriate temperature for soldering?

Joint area temperature = The M.P of solder alloy+40~50 Degree C

- Sn60% Solder (190Degree C) +40~50 Degree C =230~240 Degree C
- However the iron will be set much hotter @350-450 °C
  - **The hotter the iron the faster the joint can be made providing the heat transfer from it is good.**
- Must try to heat up both/all parts to be soldered **at the same time**.
- If it is difficult to do this then evaluate special solder iron tips.
- Heating up more than needed, will cause **overheating defects**.
  - Using a very hot iron will increase the risk of soldering defects and component damage. Soldering is a balance of time/temperature.

5-9. Incorrect Heating

No wetting

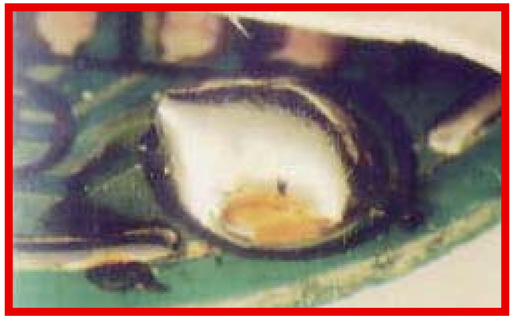

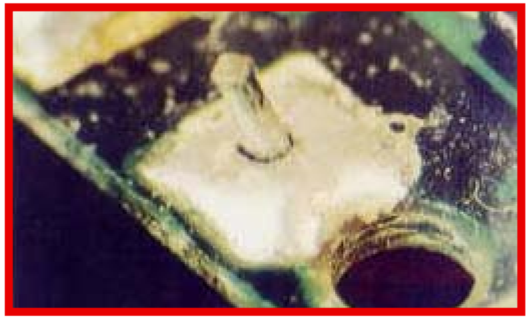

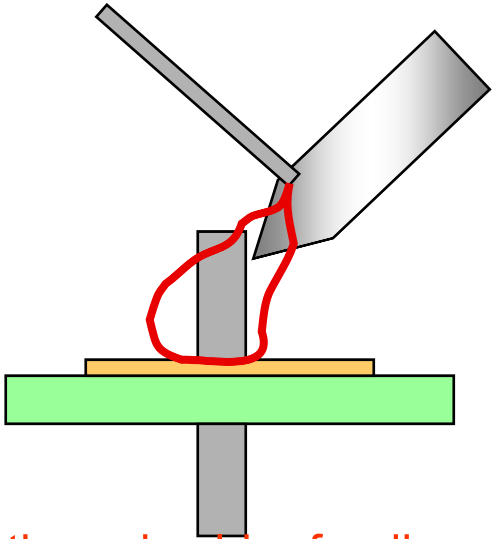

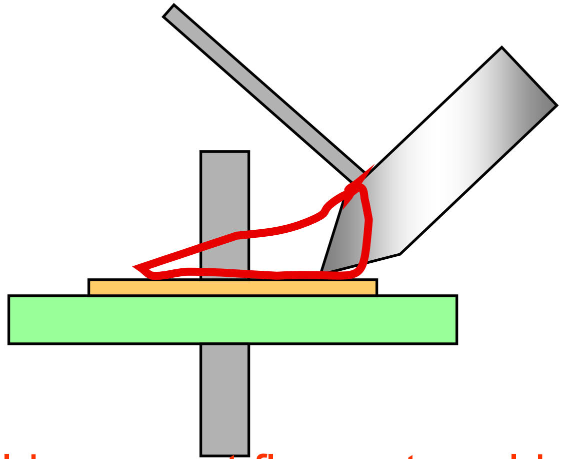

Locate iron tip and solder feeding spot carefully. Solder cannot flow onto cold areas or areas not cleaned by flux. Here flux has been used up on iron.

5-10. Movement before solder solidifies

- Crack occurs when soldering point moves.

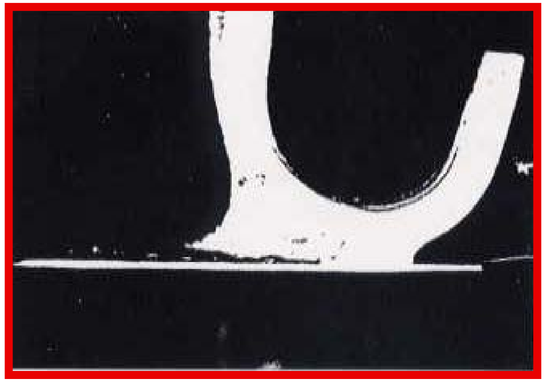

Crack due to vibration before solidification.

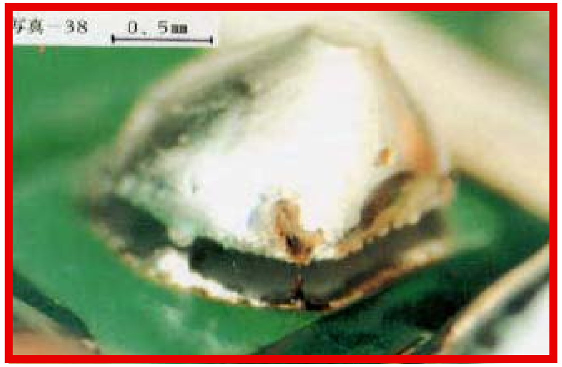

Crack due to movement before solidification.

Do not move parts until solder has solidified. Liquid solder has no strength. Large parts may take a few seconds to cool.

5-11. How to release the iron tip

- Common defects caused by iron tip releasing

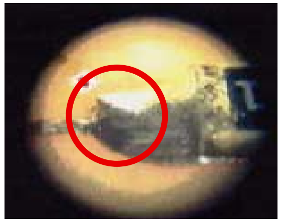

Projection/solder spike

Too slow release speed of iron tip

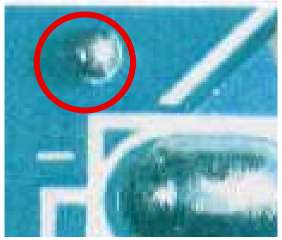

Solderball

Wrist twisting

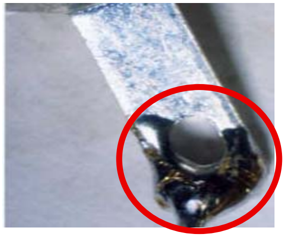

Adhesion of scrap/flux residue

Bad direction when releasing iron tip

5-12. Process Completion

Caring for the soldering iron tip

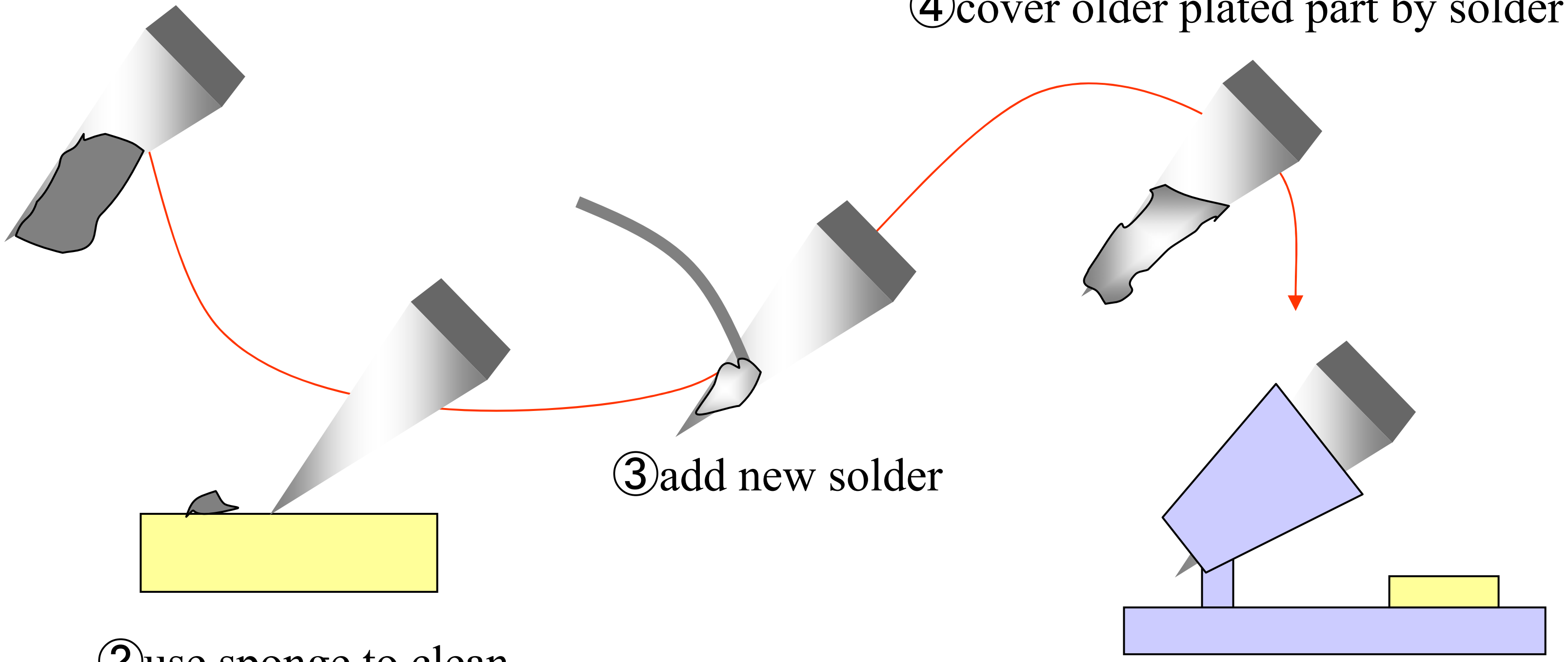

①some scrap on iron tip.

④cover older plated part by solder

③add new solder

②use sponge to clean

⑤put soldering iron back in iron holder .

5-13. Soldering Summary

- No success in soldering with poor heating,
- Heat all parts at once.
- Apply correct amount of solder
- Don’t move part until solder has solidified.
- Pay attention when releasing iron.
- Visual check to confirm soldering quality.

6-1. Inspection + Handling

Has soldering process been done well?
Take responsibility and do your own inspection.
Be able to make the correct judgment.

Inspection point

(1) Correct position (part, lead etc)

(2) Correct shape (part,lead etc)

(3) Wetting (lead, land)

(4) Solder amount (qty)

(5) Solder surface (smooth,shiny,no-hole,etc)

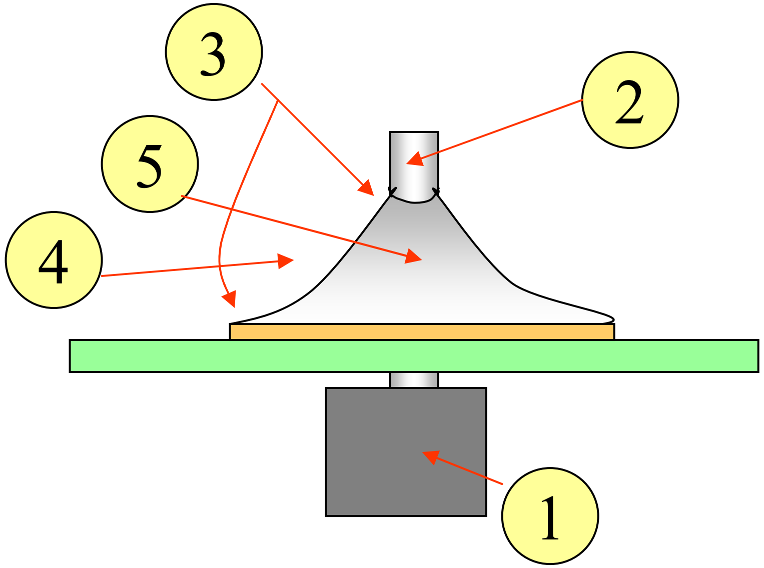

6-2. TRB Product

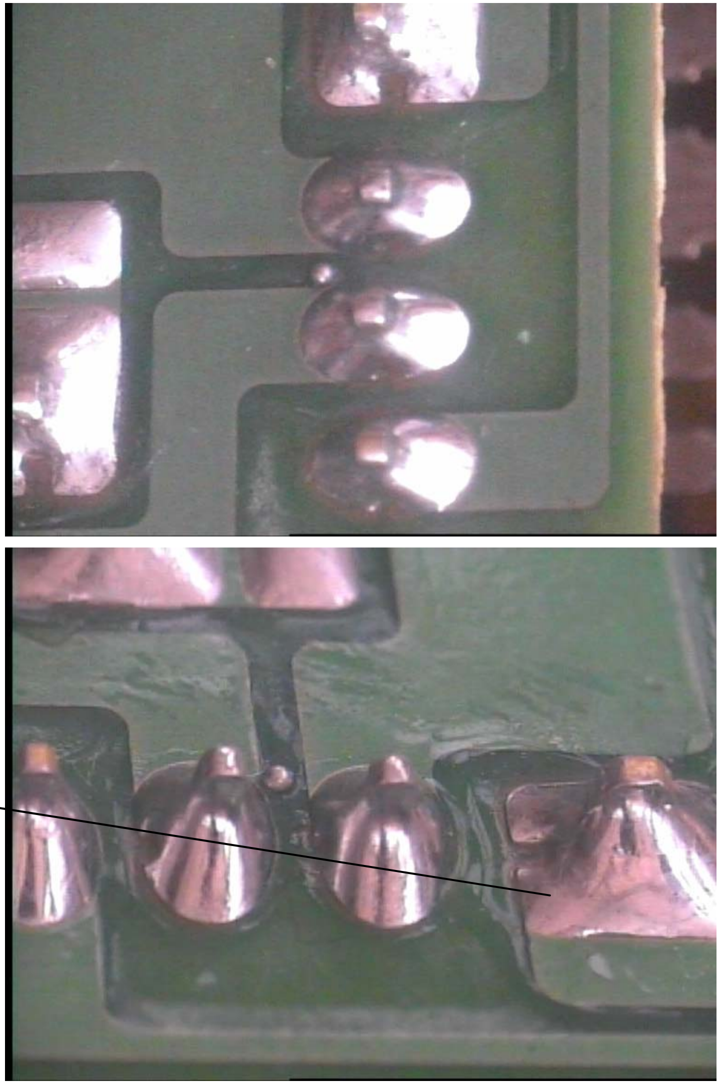

Inspection of this part; Solder has wet well to pcb pad and to pin. This shows the flux has cleaned the pad/lead well. The wetting angle is good. The solder surface is smooth/shiny. There is no charred flux or noticeable flux residue. The solderability of pcb and component seems good.

6-3. TRB Solder Defect

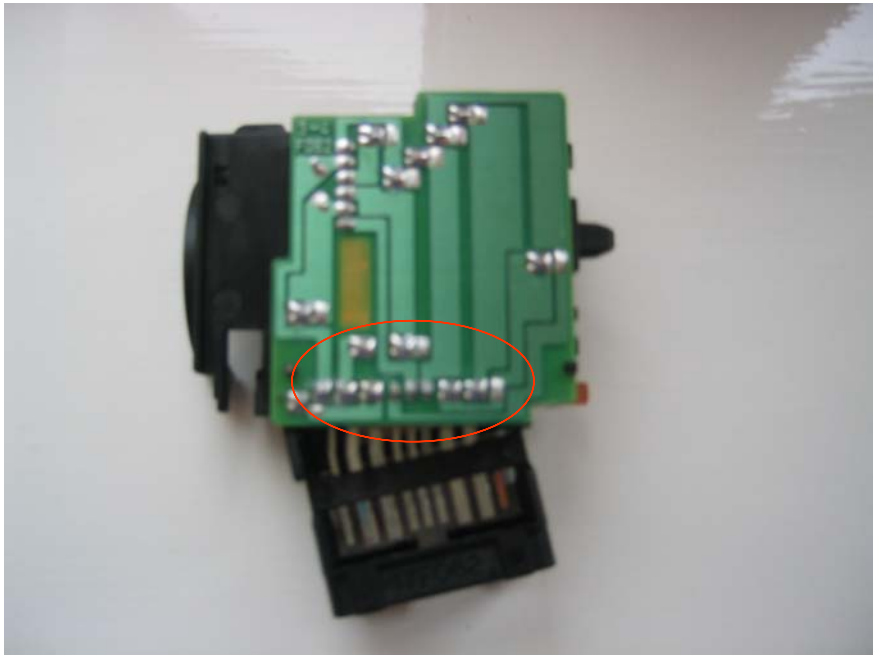

Pin 5 from right has solder slither on.
Solder ball on pcb resist.

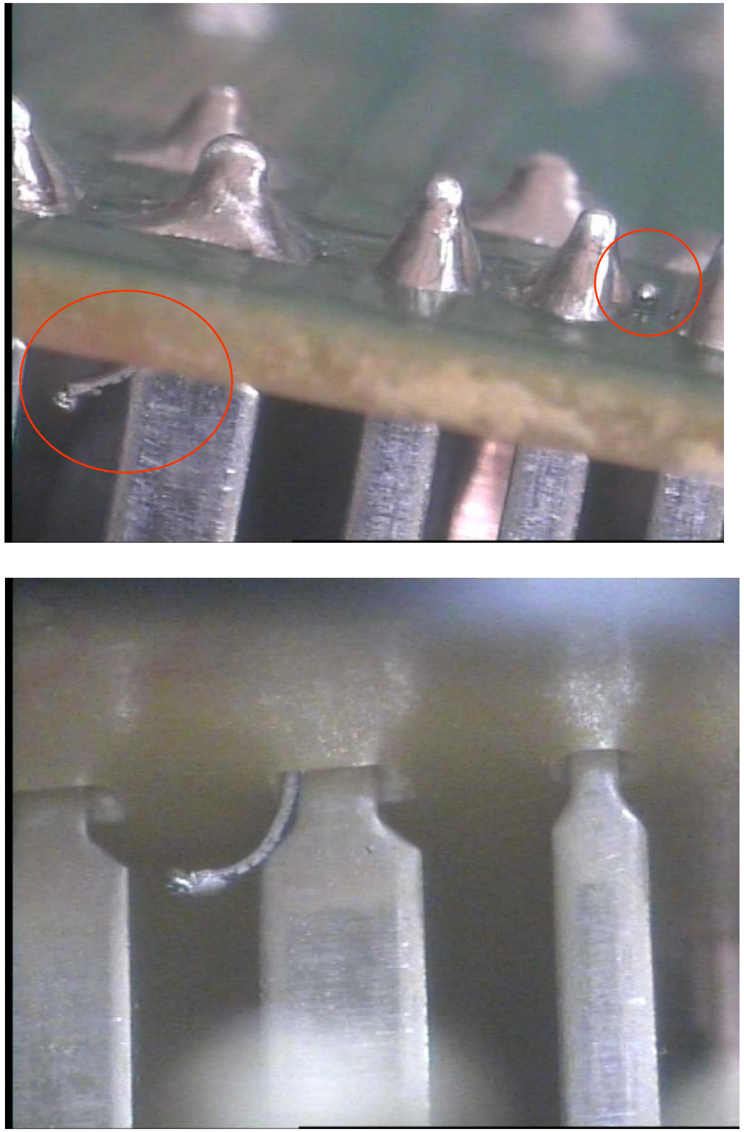

6-4. TRB Solder Defect

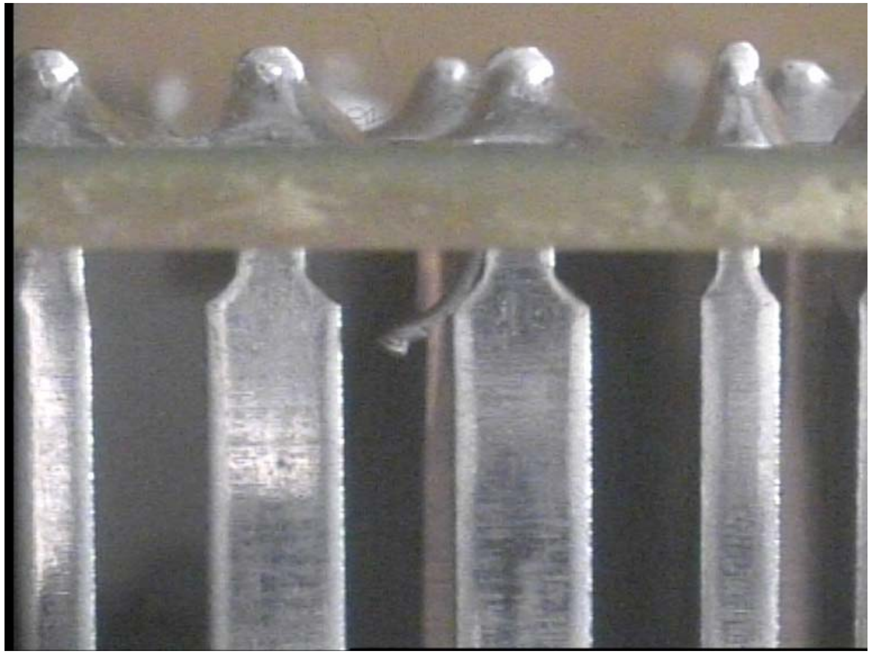

Solder process starts on right. First 2 pins are thick and soldering iron is at its Peak temp. As iron is moved onto pins 3,4,5, which are thinner pins, the iron temperature falls but is still hot enough to heat joint and solder. As it reaches pin 6, which is a thicker pin, the iron is quite a bit cooler than its' ideal temp. and does not heat up the joint area before the solder is fed onto joint. The 'warmed' solder wire is squeezed through pcb hole and onto pin shoulder.

6-5.Solder Ball

Solder ball and flux spattering

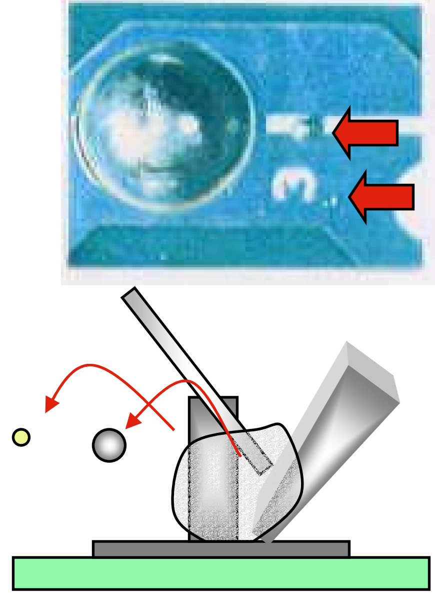

Apply solder material forcibly while the base metal is heated insufficiency.

Large solder ball

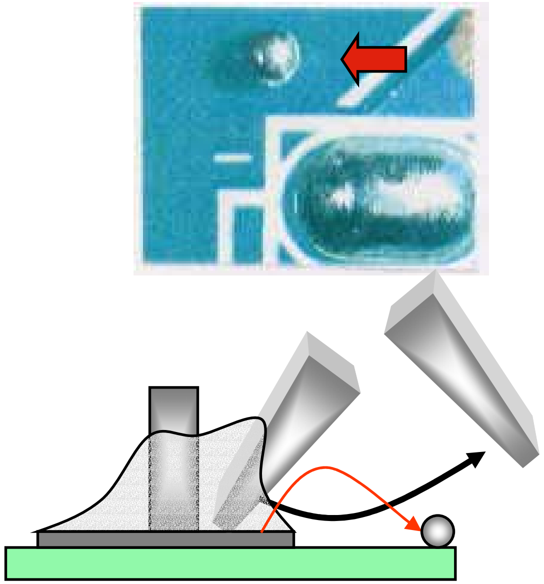

Speed of withdrawing solder-iron too fast..

Withdraw solder-iron with smooth sweeping movement.

Solder ball in flux

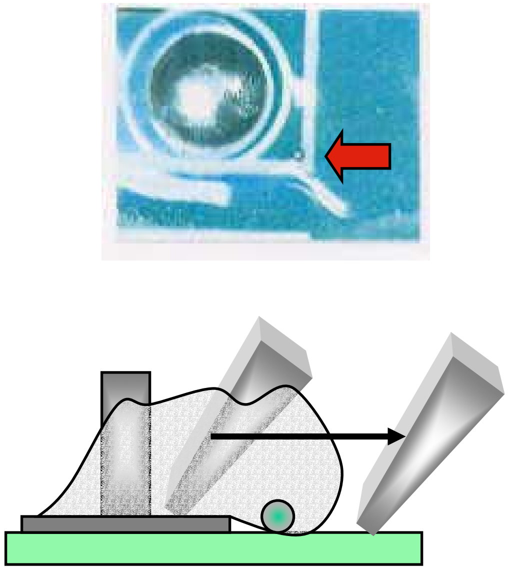

When iron-tip “slip” on to PCB

6-6. Other Defects that require attention

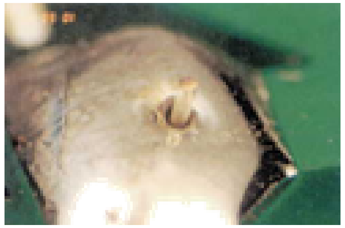
Non-wetting

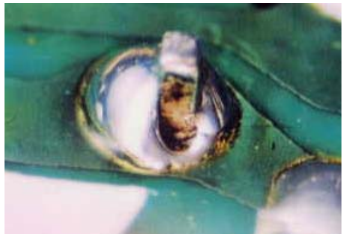
Non-wetting

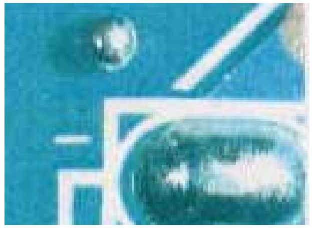
Solder ball

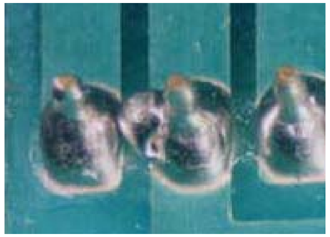
Solder scrap

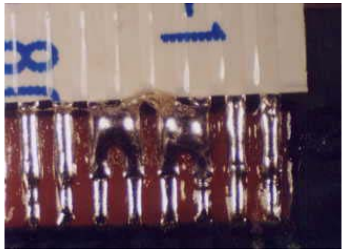
Bridge

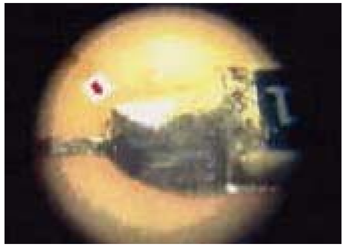
Projection

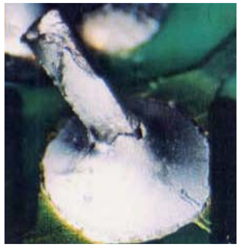
Crack

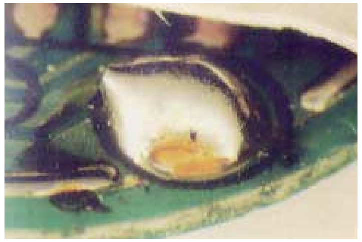
Poor wetting

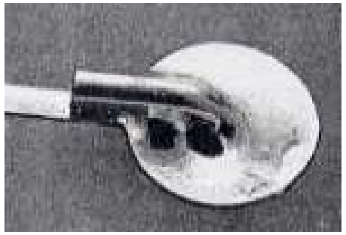
tunnel

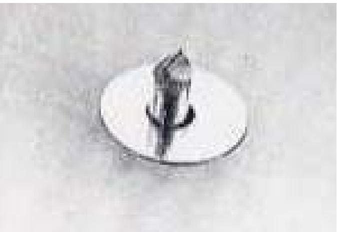
Forget to solder

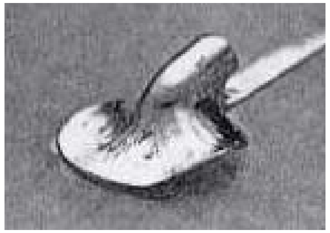
Land peeled off

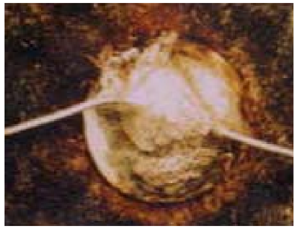
Over heat

6-7. Reminder ①

Handling PWB

Good

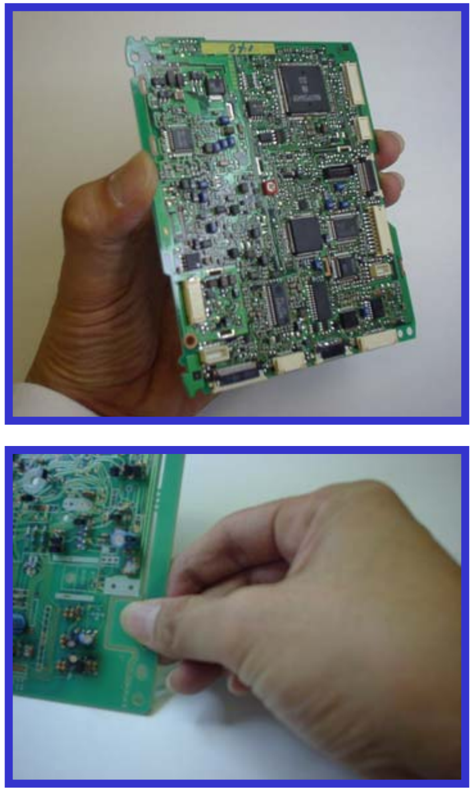

No good

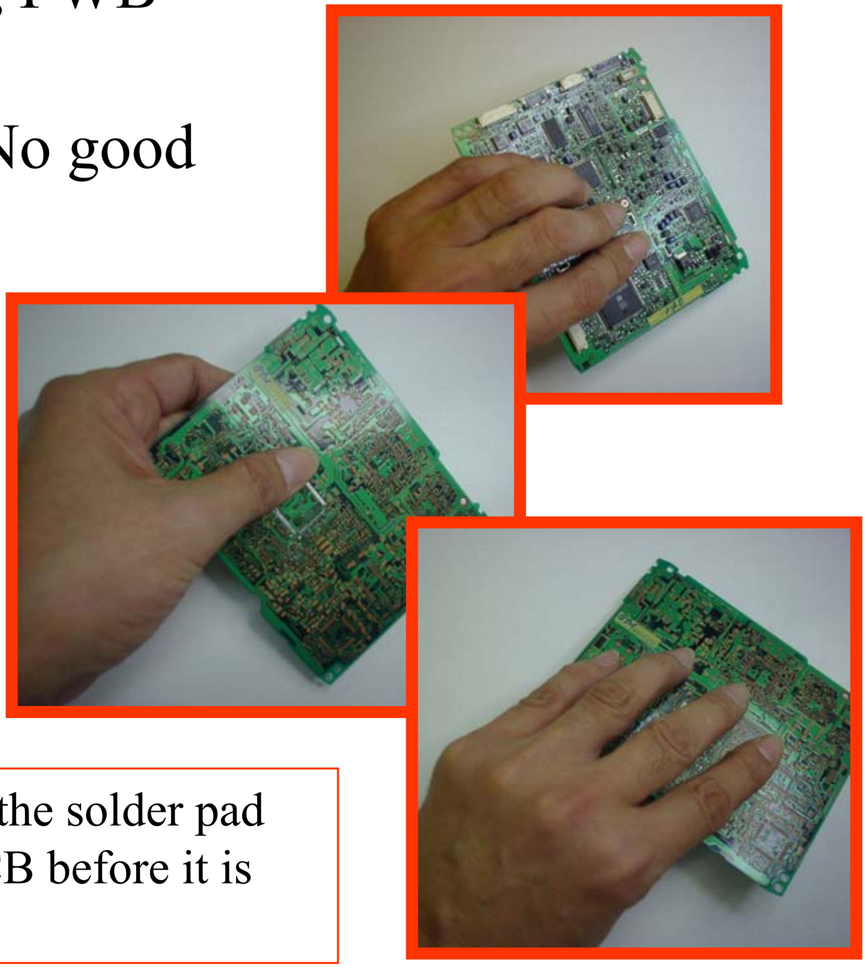

Do not touch the solder pad side of the PCB before it is soldered.

6-8. Reminder ②

Handling PWB

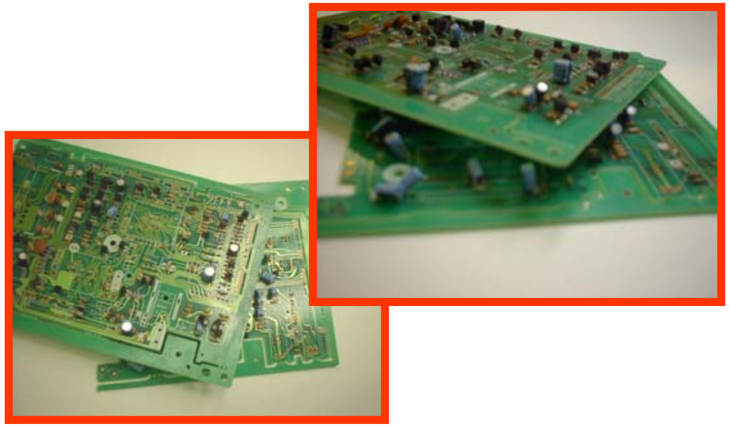

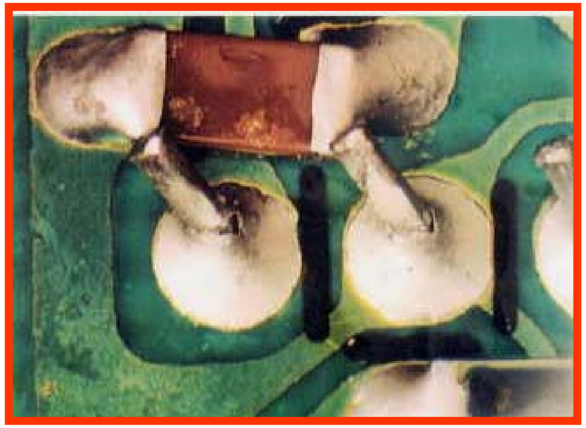

The crack occurs at the solder interface even with small impacts.

Do not ; pile up PWB

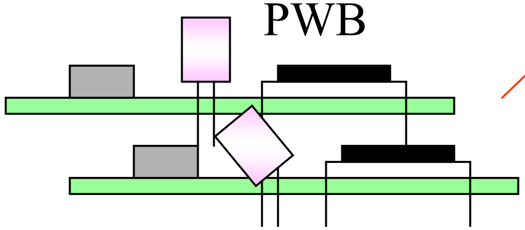

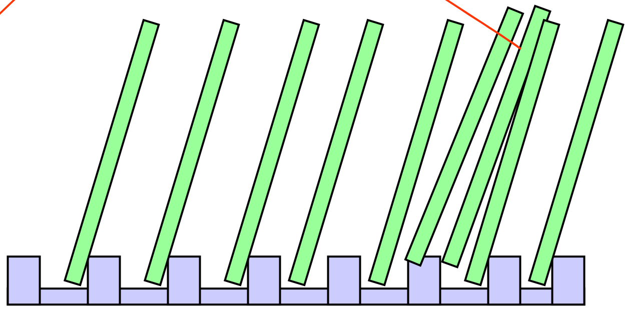

6-9. Cleanliness
Parts to be soldered must not be contaminated.

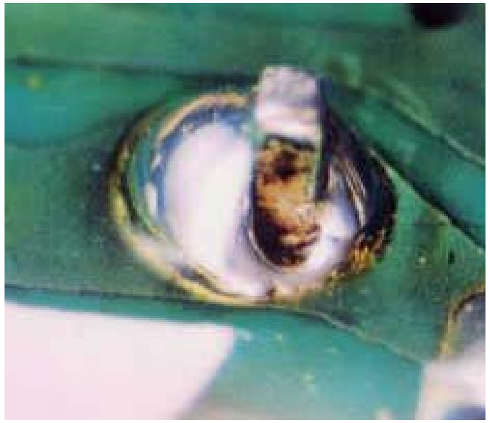

Contaminated lead

Didn’t wet on the pollution.

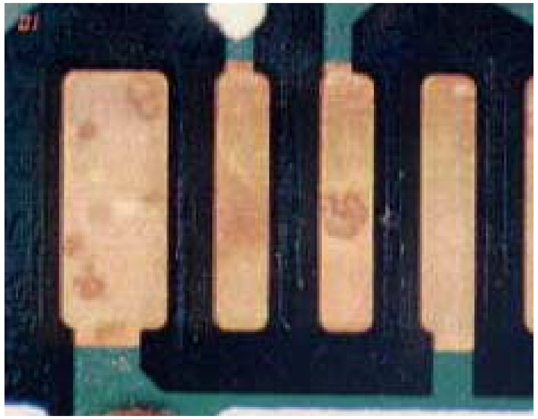

Dirty substrate

The pad on substrate was polluted by touching it with bare hands.

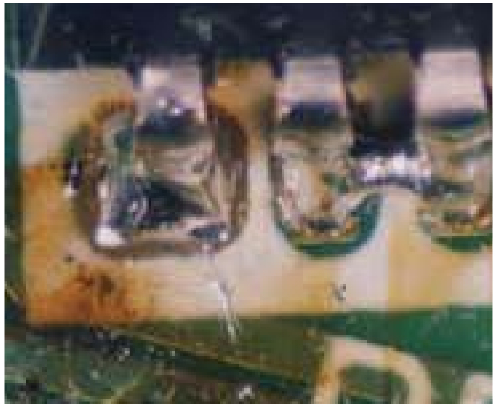

Poor cleaning of iron tip

Working with poorly cleaned iron tip causes soldering defects.

7. Practical training

Training 1

Soldering copper wire to pad
Inspection

7-1. Operation procedure

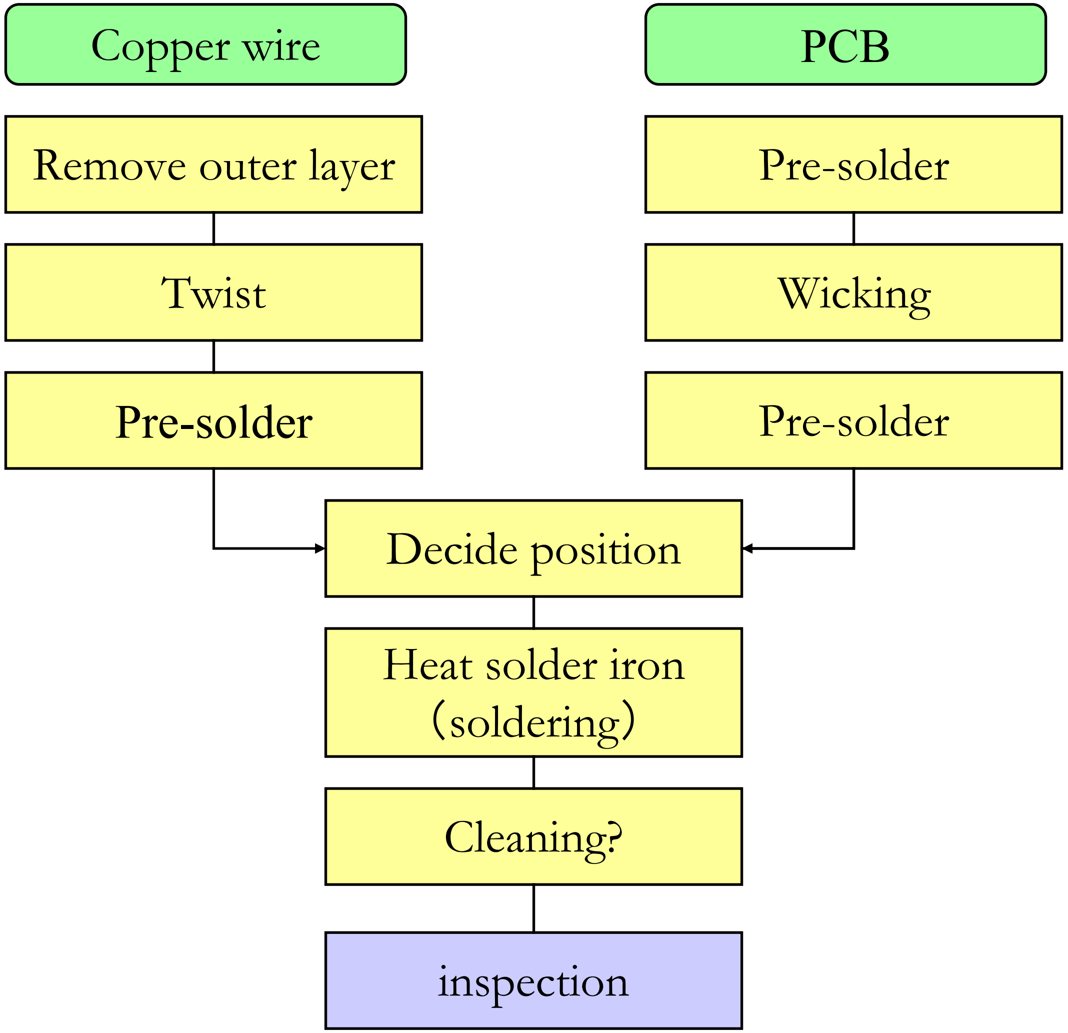
Copper wire
Remove outer layer
Twist
Pre-solder
PCB
Pre-solder
Wicking
Pre-solder
Decide position
Heat solder iron (soldering)
Cleaning?
inspection

7-2. Practical Training

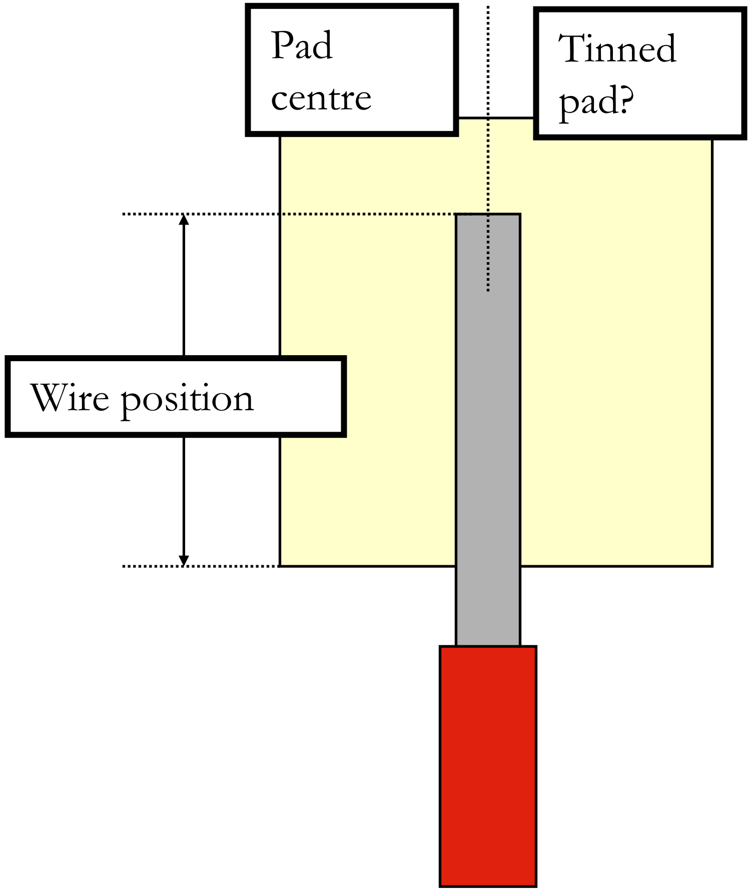

Pad centre

Tinned pad?

Wire position

①solder qty, must be able to see wire outline/shape

②fillet curve in shape of a bow

③smooth and shiny solder surface without holes and spikes etc

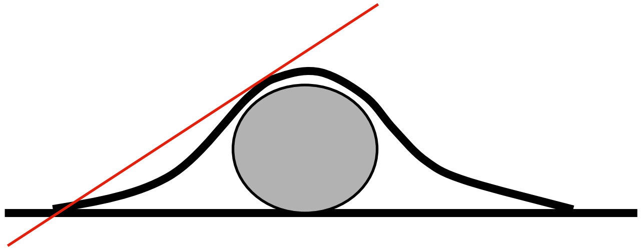

①correct position

②correct shape

7-3. Surface of copper soldering

Good sample of Sn60

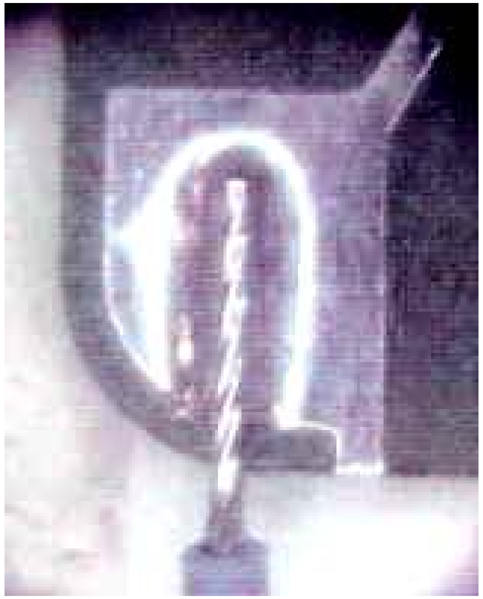

Condition of Pb Free

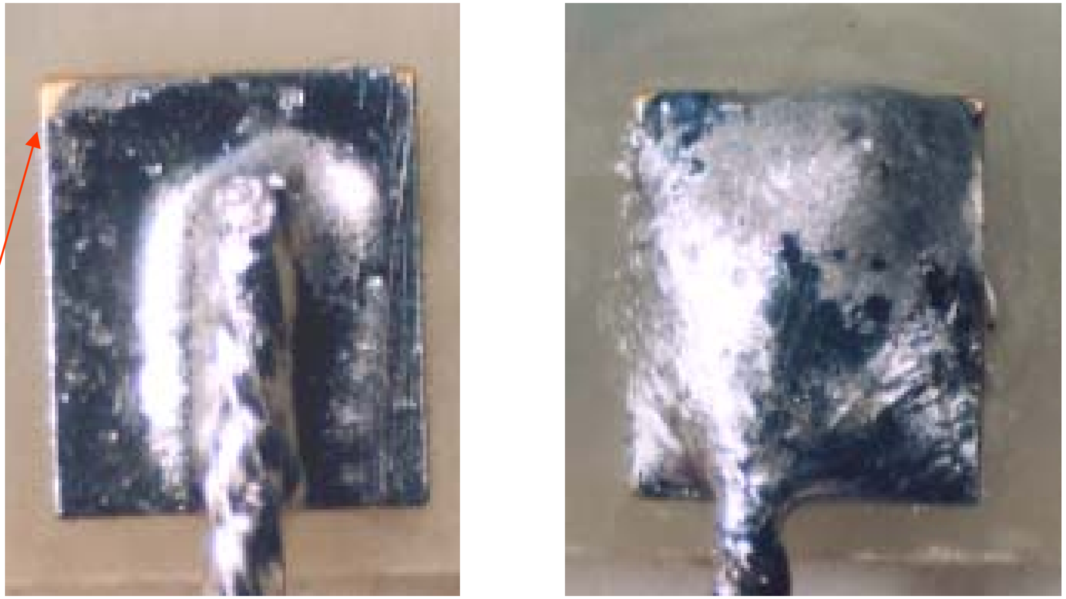

Poor wetted parts occur because spreadability of lead free is poor.

7-4. Soldering Wire to Metal Pad

Poor example of Sn60 solder

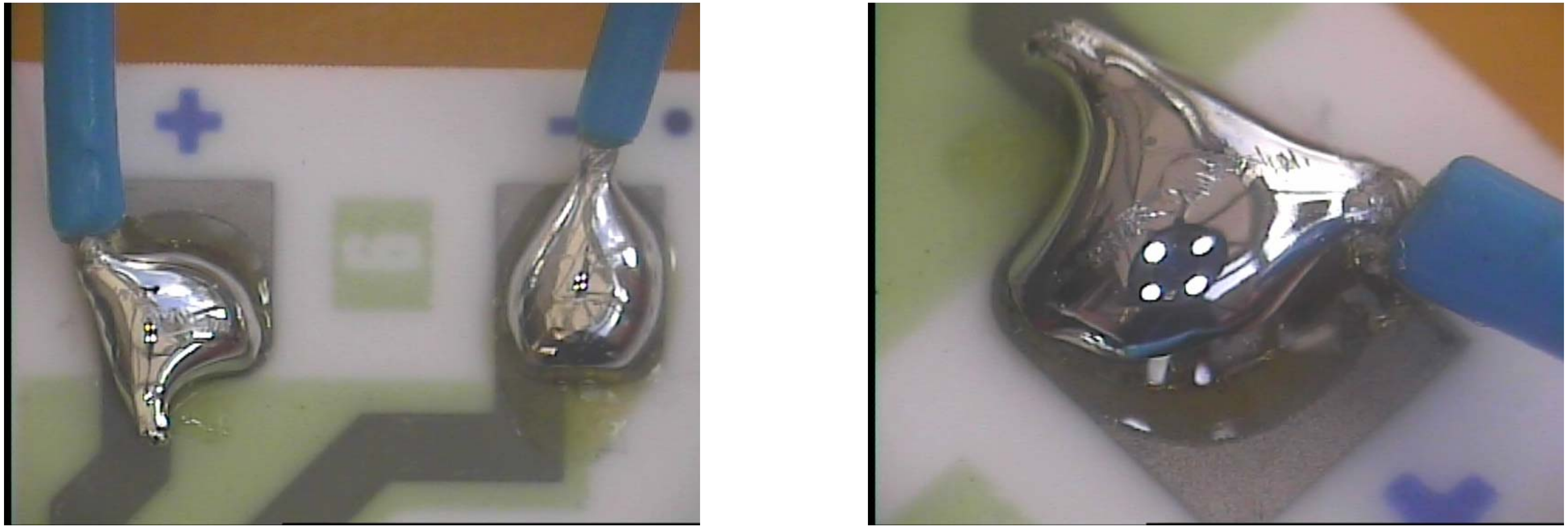

Poor wetting occurs because
insufficient heat on pad

Thank you

A Quick Guide to Welding and Weld Inspection

QG Publishing is a Matthews Engineering Training Ltd company

MATTHEWS
ENGINEERING TRAINING LTD

www.matthews-training.co.uk

Training courses for industry

- Plant in-service inspection training
- Pressure systems/PSSR/PED/PRVs
- Notified Body training
- Pressure equipment code design ASME/BS/EN
- API inspector training (UK) : API 510/570/653
- On-line training courses available

Matthews Engineering Training Ltd provides training in pressure equipment and inspection-related subjects, and the implementation of published codes and standards.

More than 500 classroom and hands-on courses have been presented to major clients from the power, process, petrochemical and oil/gas industries.

We specialise in in-company courses, tailored to the needs of individual clients.

Contact us at enquiries@matthews-training.co.uk
Tel: +44(0) 7732 799351

Matthews Engineering Training Ltd is an Authorized Global Training provider to The American Society of Mechanical Engineers (ASME)

www.matthews-training.co.uk

A Quick Guide to

Welding and Weld Inspection

Steven E. Hughes

Series editor: Clifford Matthews

Matthews Engineering Training Limited
www.matthews-training.co.uk

ASME PRESS

WOODHEAD PUBLISHING LIMITED
Oxford Cambridge New Delhi

Published by Woodhead Publishing Limited, Abington Hall, Granta Park, Great Abington, Cambridge CB21 6AH, UK
www.woodheadpublishing.com
and
Matthews Engineering Training Limited
www.matthews-training.co.uk

Woodhead Publishing India Private Limited, G-2, Vardaan House, 7/28 Ansari Road, Daryaganj, New Delhi - 110002, India
www.woodheadpublishingindia.com

Published in North America by the American Society of Mechanical Engineers (ASME), Three Park Avenue, New York, NY 10016-5990, USA
www.asme.org

First published 2009, Woodhead Publishing Limited and Matthews Engineering Training Limited
© 2009, S. E. Hughes

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. Reasonable efforts have been made to publish reliable data and information, but the author and the publishers cannot assume responsibility for the validity of all materials. Neither the author nor the publishers, nor anyone else associated with this publication, shall be liable for any loss, damage or liability directly or indirectly caused or alleged to be caused by this book.

Neither this book nor any part may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming and recording, or by any information storage or retrieval system, without permission in writing from Woodhead Publishing Limited.

The consent of Woodhead Publishing Limited does not extend to copying for general distribution, for promotion, for creating new works or for resale. Specific permission must be obtained in writing from Woodhead Publishing Limited for such copying.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation, without intent to infringe.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data
A catalog record for this book is available from the Library of Congress.

Woodhead Publishing ISBN 978-1-84569-641-2 (book)
Woodhead Publishing ISBN 978-1-84569-767-9 (e-book)
ASME ISBN 978-0-7918-5950-6
ASME Order No. 859506
ASME Order No. 85950Q (e-book)

Typeset by Data Standards Ltd, Frome, Somerset, UK
Printed by Cromwell Press Limited, Trowbridge, Wiltshire, UK

Foreword by Series Editor

We are pleased to introduce this new book by Steve Hughes in our ongoing *Quick Guide* series. Intended for inspection engineers and technicians the *Quick Guide to Welding and Weld Inspection* will provide a useful reference for those involved in the plant inspection discipline.

Steve Hughes provides valuable lecturing input to our ASME Plant Inspector and API 510/570/653 Certified inspection training courses run by Matthews Engineering Training Ltd, bringing his technical knowledge underpinned by sound engineering practical experience. As a discipline, Plant Inspection incorporates many subjects, but welding and weld inspection forms a common thread running through the majority of them. For pressure systems in particular, weld joining design, welding techniques, NDE and defect acceptance are prime considerations in equipment repair and assessing the risk of existing equipment.

The documentation aspects of weld inspection often remain a bit of a mystery to plant inspectors (even experienced ones). In assessing the quality of weld repairs it becomes all too easy to assume that the necessary WPS/PQR weld documentation and welder qualifications are in place, without properly checking. Equally it is easy to be intimidated by the technical detail and terminology symbols and acronyms. You should find this book useful in helping you through such situations.

Despite the predominance given in technical publications to advanced technology-intensive NDE and inspection activities, it remains a fact that 90% of all pressure equipment inspections include only well-proven established inspection activities. In writing this book, Steve Hughes has addressed these real issues of most practical weld inspections, distilling a wide subject into a simplified and digestible format. You don't have to be a qualified welding engineer or metallurgist to read this book, and you may still benefit from it if you are.

Finally, we are always interested in hearing from people who are interested in writing a *Quick Guide* book (like this one) in an inspection-related subject. You don't have to be an experienced author or an acknowledged expert on the subject (all experts are self-appointed anyway – we know that), just have sufficient experience to know what you are talking about and be able to write it down in a way that other people will understand. We'll do the rest. Contact me on enquires@matthews-training.co.uk or through our website www.matthews-training.co.uk.

Cliff Matthews
Series Editor

Introduction

This book was written to provide a quick guide to welding inspection which is easy to read and understand. There are many books covering all aspects of welding (many of them go into great detail on the subject) but it is difficult to find books specifically covering weld inspection requirements. This book's subjects are purposely not covered in great detail because it is assumed that the reader is able to find detailed books on specific subject areas of particular interest. What this book will do is give you a basic understanding of the subject and so help you decide if you need to look further. In many cases the depth of knowledge required for any particular welding-related subject will be dependent on specific industry requirements. In all situations, however, the welding inspector's role is to ensure that welds have been produced and tested in accordance with the correct code specified procedures and that they are code compliant. Code compliance in this sense means that the weld meets all the requirements of the defect acceptance criteria specified within the code.

Inspectors considering training to achieve certified welding inspector status under certification schemes such as CSWIP (Certification Scheme for Welding Inspection Personnel) or PCN (Personal Certification Number) will find the book a useful pre-course learning aid giving coverage of the 'body of knowledge' they are expected to be familiar with. Non-welding personnel will find it a useful introduction to the world of welding inspection. Some people believe that a welding inspector must have previous welding experience, but this is not necessarily true as welding and welding inspection are two totally different subjects. Welding is naturally a mainly practical 'skill of hand' process and requires dexterity and good hand-to-eye coordination from the welder. The inspector does not require this practical skill but must be able to oversee the welding process, take accurate measurements, interpret the requirements of codes and standards, and ensure that completed welds are in compliance with the relevant code requirements. A good inspector is one who does not take shortcuts and ensures that procedures are properly followed.

S. E. Hughes
Author

A 10 Minute Guide

Here are some frequently asked questions about welding inspection subjects. You can answer them if you spend ten minutes reading the following information.

1. What is the difference between the main welding processes? (four minutes)

MMA has an arc struck between a consumable flux-coated electrode and the workpiece. The electrode melts and fills the joint with weld metal. The flux coating melts and produces gas to shield the arc. The melted flux also helps to remove impurities from the weld and forms a layer of slag on top of the weld. This slag must be removed between runs or it can cause slag inclusions within the weld. This is the most commonly used outdoor site welding process.

MIG/MAG has an arc struck between a reel-fed consumable solid wire electrode and the workpiece. It does not produce slag because the arc is shielded by an inert gas (active gas for MAG). It deposits weld metal quickly and can be used semi-automatically, mechanized or automated. This process can achieve fast weld metal deposition rates.

FCAW is similar to MIG/MAG but the consumable reel-fed electrode is hollow with a flux contained inside. The arc can be self-shielded using only the melting flux or a secondary gas shield can be used. Slag will be produced and must be removed between runs. Basically, this process gives a combination of the advantages (and disadvantages) of MMA and MIG/MAG.

TIG has an arc struck between a non-consumable tungsten electrode and the workpiece. Filler in wire or rod form is added separately. It does not produce slag because the arc is shielded by an inert gas. It is a very slow process but produces very high quality welds.

SAW has an arc struck between a reel-fed consumable solid wire electrode and the workpiece. The arc is formed underneath a layer of flux and is therefore not visible to the operator. This is a deeply penetrating process requiring good penetration control. It has very fast deposition rates but is generally restricted to welding in the flat or horizontal–vertical positions (otherwise the flux would fall out).

Spend four minutes reading through the welding processes, noting the differences and making sure you recognise the acronyms used for each process.

2. Why use welding symbols? (one minute)

Construction drawings need to pass on information clearly and within a limited space. Welding information is passed on by a method of conventions and symbols.

Spend one minute looking over the different symbols for butt welds and fillet welds.

3. What are codes and standards? (one minute)

Construction codes and application standards contain the rules that must be followed when providing a specific product or service. They contain information on design, manufacturing method, acceptable materials, workmanship, testing requirements and acceptable imperfection levels. They do not contain all the relevant data required for the design, manufacture, testing and inspection but will reference other standards and documents as required.

Spend one minute reading the definitions of codes and standards.

4. What is welding procedure qualification? (one minute)

Welding procedure qualification is carried out to prove that a welded joint meets the mechanical, metallurgical and physical properties required by a code or specification. It also enables repeatability by encouraging a systematic approach. The welding procedure qualification documents consist of a Welding Procedure Specification (WPS) and a Procedure Qualification Record (PQR) containing essential and non-essential variables.

Spend one minute reading about essential, non-essential and supplementary essential variables.

5. Why does a welder need qualification? (one minute)

Welders require qualification to prove that they have sufficient knowledge and skill to produce a weld in accordance with a welding procedure and achieve a result that meets the relevant specification. A welder qualification test form records the range of variables qualified from the essential variables.

Spend one minute looking at the variables contained in a WPQ form.

6. What are welding defects? (two minutes)

There is no such thing as a perfect weld because all welds contain imperfections of some sort. These imperfections need to be assessed to determine if they will have a detrimental effect on the welded joint. They are normally assessed against the acceptance criteria specified in the relevant code or standard but only an imperfection found to be outside the acceptance criteria limits will be classed as a defect and require action to be taken. The action to be taken could include:

- rejection of the component;
- removal of the defect and re-welding of the joint;
- a fitness-for-purpose analysis being carried out with a concession being granted allowing the defect to remain.

Spend two minutes determining what defects can occur in welds.

Chapter 1

Abbreviations, Terminology and Welding Symbols

Abbreviations

It is useful to have an awareness of abbreviations used within the world of welding and welding inspection. Some commonly used ones are:

API:	American Petroleum Institute
ASME:	American Society of Mechanical Engineers
BM:	base metal
BS:	British Standard
BSI:	British Standards Institute
CVI:	close visual inspection
DT:	destructive testing
EN:	European standard (Euro Norm)
GVI:	general visual inspection
HAZ:	heat affected zone
ISO:	International Standards Organisation
LCT:	lower critical temperature
MT/MPI:	magnetic testing/magnetic particle inspection
NDE/NDT:	non-destructive examination/non-destructive testing
PQR:	Procedure Qualification Record
PT/LPE:	penetrant testing/liquid penetrant examination
PWHT:	post-weld heat treatment
RT/RAD:	radiographic testing/radiography
SWI:	senior welding inspector
UCT:	upper critical temperature
UT:	ultrasonic testing
UTS:	ultimate tensile strength
VI:	visual inspection
VWI:	visual welding inspector

WI:	welding inspector
WM:	weld metal
WPS:	welding procedure specification

Common terms

Longitudinal direction:	along the length of the weldment (parallel to the weld)
Transverse direction:	along the width of the weldment (perpendicular to the weld)
Short transverse direction:	through the weldment thickness
Weldment:	the combined weld, HAZ and base metal
Fusion weld:	a weld produced by joining materials in a molten state
Yield point:	the point at which plastic deformation starts in a material

Joint terminology

Before welding takes place the parts to be joined must be prepared and arranged into the required form. The most common forms of joint are butt (or groove) joints, T joints and lap joints. Butt joint members are fitted edge to edge, T joints are fitted edge to surface and lap joints are fitted surface to surface (see Fig. 1.1).

Weld terminology

Types of weld used on the joints are butt (or groove) welds, fillet welds, edge welds, plug welds and spot welds. The type

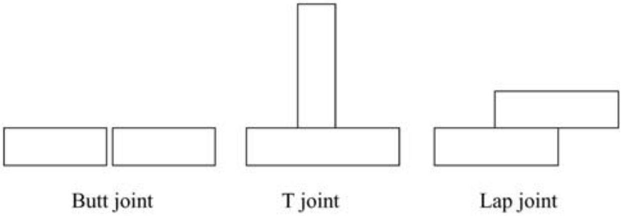

Figure 1.1 Common joint forms

of weld will be determined by the joint set-up but the most commonly used welds are butt or fillet welds.

Full penetration butt welds are normally the strongest type of weld with the strength contained within the throat of the weld, as indicated in Fig. 1.2(a). The throat is equal to the thinnest of the parent materials to be joined.

Fillet welds also contain their strength in the throat and the design throat size is normally (but not always) based on the leg size, as shown in Fig. 1.2(b). In visual inspections we can normally work out the design throat size of a mitre fillet with equal leg lengths by multiplying the leg length by 0.707. Conversely, the leg length can be calculated by multiplying the design throat by 1.414.

Butt weld joint preparation consists of preparing both edges and then arranging them together to permit the required depth of weld penetration to be achieved. Full penetration welds are the most common but many applications will only require partial penetration if the welded joint is either lightly loaded or is just a sealing weld.

The main purpose of the joint preparation is to permit the required level of fusion between the joint faces. The type of weld preparation applied will therefore be dependent upon the thickness of the material and the welding process to be used. Some typical butt weld joint preparations are shown in Fig. 1.3.

Single-sided preparations allow welding to take place from one side whereas double-sided preparations require welding to be done from both sides. A major disadvantage of having access to only one side is that the other side of the initial root run cannot be accessed to carry out removal of welding-induced defects. An example of this would be where a large gap was required to permit fusion throughout the whole cross-section of the weld but excessive penetration then occurred. Good control and formation of the root weld must therefore be maintained at the time of welding, and this can be assisted by using the root control measures shown in Fig. 1.4.

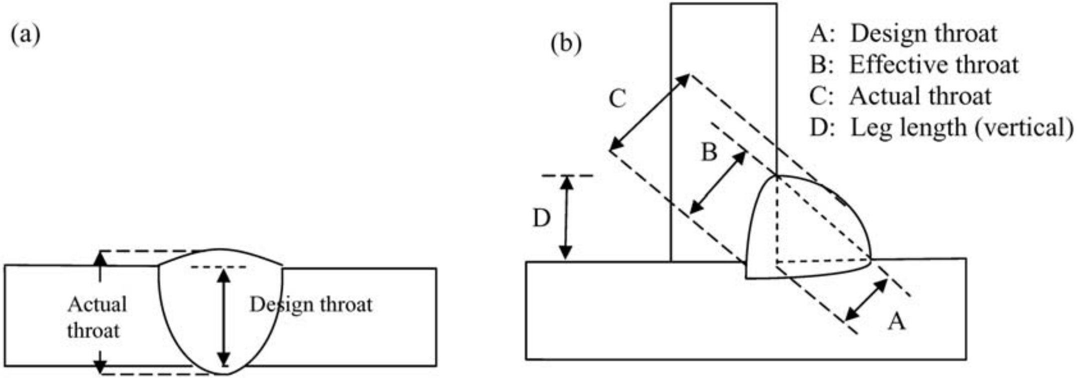
(a)
Actual
throat
Design throat
(b)
A: Design throat
B: Effective throat
C: Actual throat
D: Leg length (vertical)
C
B
D
A

Figure 1.2 Butt and fillet weld throat measurements

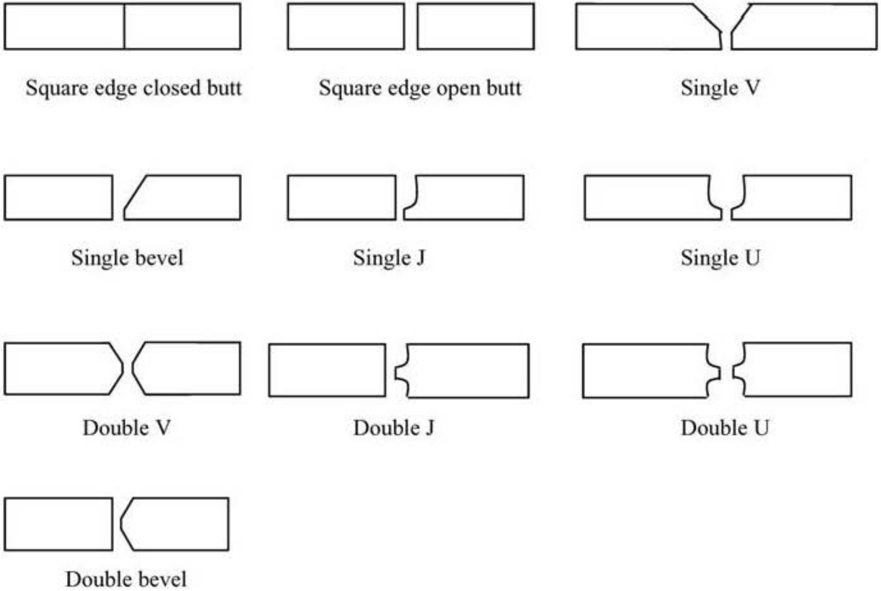
Square edge closed butt
Square edge open butt
Single V
Single bevel
Single J
Single U
Double V
Double J
Double U
Double bevel

Figure 1.3 Common butt joint types

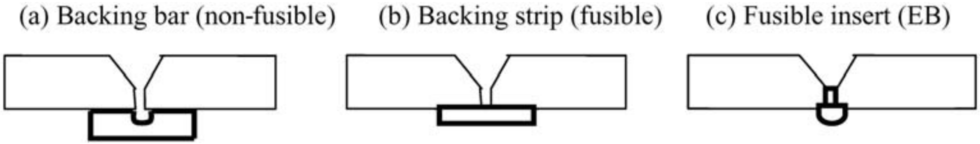

(a) Backing bar (non-fusible)
(b) Backing strip (fusible)
(c) Fusible insert (EB)

Figure 1.4 Penetration control methods

Woodhead Publishing Ltd – A Quick Guide to Welding and Weld Inspection
Data Standards Ltd, Frome, Somerset – 17/9/200901QG Welding chap1.3d Page 6 of 16

- Non-fusible backing bars. These are always removed and are normally made of copper (often water cooled) or ceramic and help form the root weld shape. It is important to ensure that the copper bar does not melt and contaminate the weld as this could cause weld metal cracking.
- Fusible backing strips. These are made of a material similar to the parent material and are tacked on to the parent material. They are fused into the root and are normally ground off, or occasionally left in place.
- Pre-placed filler such as EB inserts. These are used on pipe and welded using a TIG (tungsten inert gas) process. The EB stands for Electric Boat and is the name of the company that first supplied them. Use of EB inserts is a specialised procedure used in specialist applications such as the nuclear industry.

Fillet weld shapes are either mitre, concave or convex, as indicated in Figure 1.5.

- Mitre fillets are the most common and normally have equal leg lengths. They give a good combination of maximum design strength and toe blending for fatigue resistance.
- Concave fillets have a reduced throat measurement for their leg size, which gives them a reduced strength. The smoothly blended toe design gives them very good fatigue failure resistance.

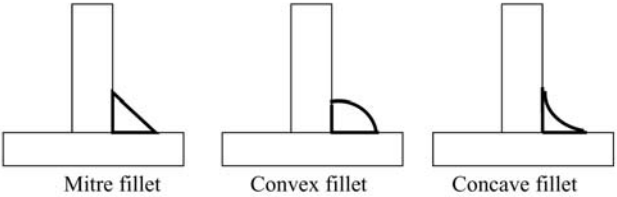

Figure 1.5 Types of fillet weld

- Convex fillet welds have sharp toes, which give very poor fatigue resistance but the excess weld metal they have on the throat gives maximum design strength (although it will increase the weight of the weld). The excess weld metal will also increase cost due to the extra consumables required and increased welding time. These welds are only used where strength is paramount and fatigue loading is not an issue.

Drawing rules and weld symbols

Weld symbols on drawings are a very efficient way to transfer fabrication information from the designer to the fabricator by showing the joint and welding information as a symbolic representation. This reduces the amount of information that would have to be put on the drawing in written form or as sketches. An inspector must have a good working knowledge of weld symbols as a large proportion of inspection time is spent verifying that the welder is complying with the approved fabrication drawing. Weld symbols themselves are similar between the major standards but there are some differences in how they are presented. It is important to understand the basic differences between the major standards and to be able to recognise any drawing standard being used. Reference should always be made to the applicable standard for specific symbolic information.

Common standards in use today are BS EN 22553 (which replaced BS 499) and AWS A2.4. Reference is still made to BS 499 because many old fabrication drawings will have been done to this standard. Most drawing standards follow a basic set of rules or conventions when formulating welding symbols. A weld symbol is composed of five main components common to most standards, consisting of an arrow line, a reference line, the welded joint symbol, dimensional information and finally any supplementary information.

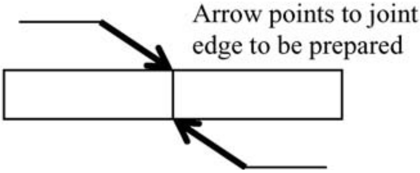

Figure 1.6 The arrow line

The arrow line

The arrow line is a single, straight, solid line (there is an exception to this in AWS A2.4 for single plate preparations where the arrow line is staggered) (Fig. 1.6). It touches the joint intersection and points to the plate edge that is to be prepared. For symmetrical joints it can point to either plate edge as they are both prepared in the same way. It must not be parallel to the bottom of the drawing and must always be finished with an arrow head.

The reference line

The reference line is a solid line that touches the arrow line (Fig. 1.7). It is preferably drawn parallel to the bottom of the drawing, but if this is not possible it is drawn perpendicular to the bottom. There must be an angle between the arrow line and reference line where they join. In BS EN 22553 a dashed line is also placed either above or beneath the solid reference line and relates specifically to the 'other side' of a joint.

The welded joint symbol

The joint symbol is used to represent the different joint categories and is generally similar in shape to the weld that it

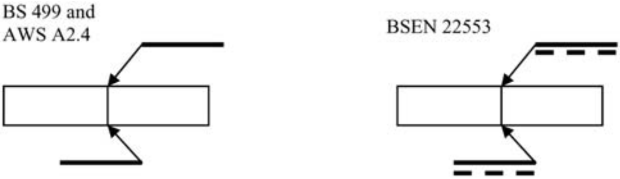

Figure 1.7 The reference line

represents. Depending on the standard used, it is normally positioned below or above the reference line to specify whether the weld is on the 'arrow side' or 'other side' respectively. In BS 499 and AWS A2.4 the symbol is placed below the reference line to represent the fact that the weld is made from the arrow side (i.e. the side that the arrow is touching). If the symbol is on top of the line it indicates that the weld will be made from the opposite side (i.e. the other side) to which the arrow is pointing (or touching). This often causes confusion and can lead to welds being put in the wrong position.

To avoid this confusion the 'other side' is represented by a separate dashed line in BSEN 22553 (see Fig. 1.8). It therefore does not matter if the symbol is above or below the solid line because it will always be on the arrow side except when placed on the dashed line. A symmetrical joint such as a double V, which is the exact same size on both sides, does not require a dashed line to represent the other side because both sides are the same.

Supplementary symbols may be used to complete a joint symbol with additional information such as a flat, concave or convex finish (Fig. 1.9). The absence of a supplementary symbol indicates that the weld is left in the 'as-welded' condition with no precise weld surface finish required.

Dimensions

Dimensions that relate to the joint cross-section, such as weld depth or fillet sizing, are positioned on the left of the symbol. If no weld depth is specified before the symbol on a butt weld it is assumed to be a full penetration weld. Fillet welds will always have a dimension relating to the leg length and/or throat dimension. Longitudinal dimensions such as weld lengths are positioned on the right of the symbols in most standards (see Fig. 1.10). If no longitudinal dimensions are specified after the symbol then the weld extends the full length of the joint.

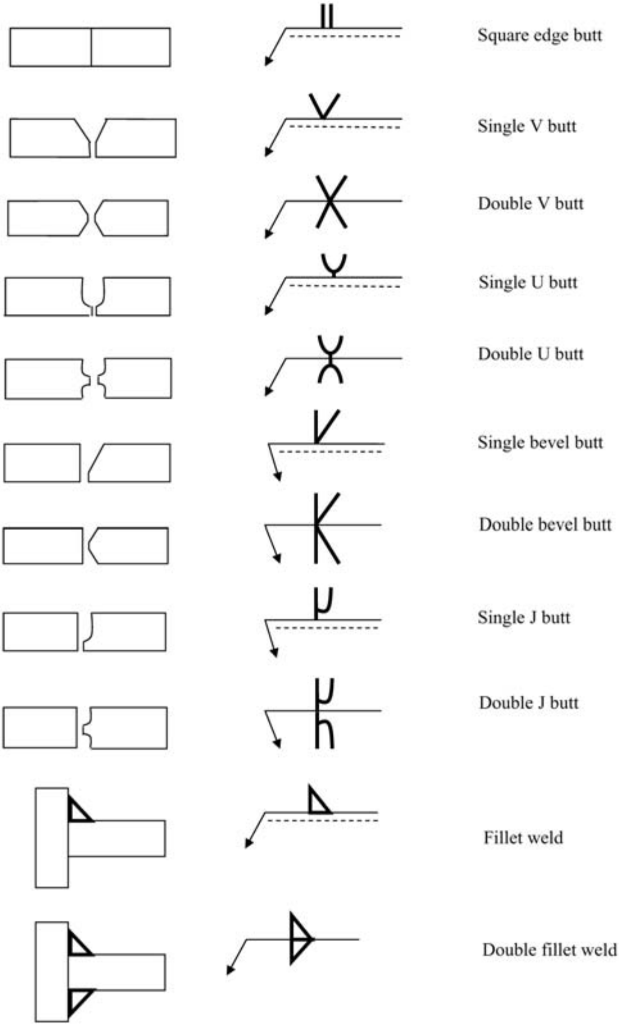
Square edge butt
Single V butt
Double V butt
Single U butt
Double U butt
Single bevel butt
Double bevel butt
Single J butt
Double J butt
Fillet weld
Double fillet weld

Figure 1.8 Typical weld drawing symbols

Supplementary Symbols (from BS EN 22553)		
	Shape of weld surface or weld	Symbol
(a)	Flat (usually finished flush)	—
(b)	Convex	⌒
(c)	Concave	‿
(d)	Toes shall be blended smoothly	
(e)	Permanent backing strip used	⌈M⌉
(f)	Removable backing strip used	⌈MR⌉

(a)

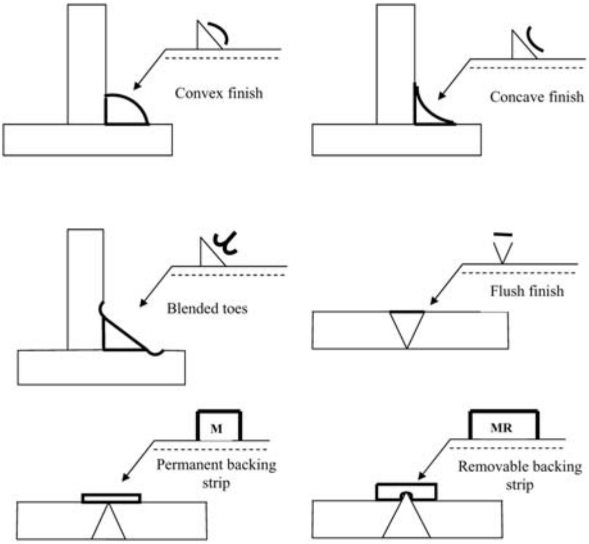
Convex finish
Concave finish
Blended toes
Flush finish
M
Permanent backing strip
MR
Removable backing strip

(b)

Figure 1.9 (a) Supplementary symbols; (b) examples of supplementary symbol use

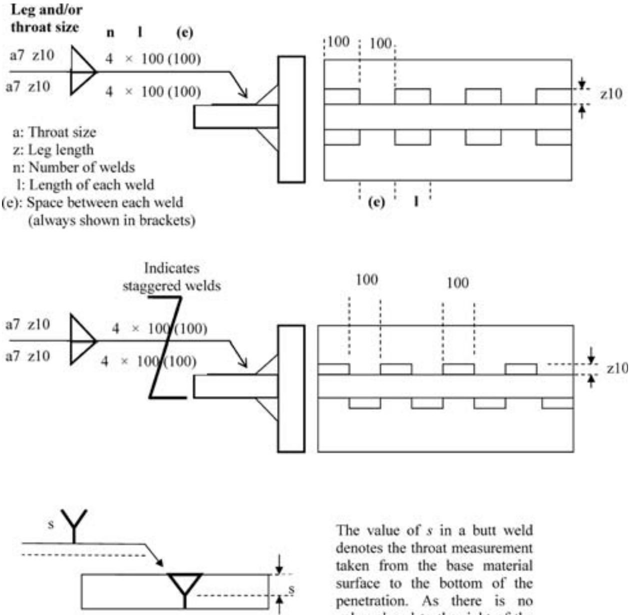
Leg and/or throat size
n
l
(e)
a7 z10
4 × 100 (100)
a7 z10
4 × 100 (100)
100
100
z10
a: Throat size
z: Leg length
n: Number of welds
l: Length of each weld
(e): Space between each weld (always shown in brackets)
(e)
l
Indicates staggered welds
a7 z10
4 × 100 (100)
a7 z10
4 × 100 (100)
100
100
z10
s
s

The value of *s* in a butt weld denotes the throat measurement taken from the base material surface to the bottom of the penetration. As there is no value placed to the right of the symbol the weld is continuous for its full length.

Figure 1.10 Weld dimensioning examples (BSEN 22553)

Supplementary information

Supplementary information specifying NDT, welding processes or special instructions can differ between standards and may be shown at the end of the reference line or adjacent to it (see Fig. 1.11). BSEN 22553 refers to this as complementary information. A site or field weld is indicated by a flag positioned at the joint between the arrow and reference lines. A circle at the same position is used to indicate that a peripheral weld is required all around a part.

Other common rules

When a symbol has a perpendicular line (single bevel, single J, fillet weld, etc.) it will always have the line positioned to the left. When the same symbol is shown on the top and bottom of the reference line (double V, double bevel, etc.) it will be a mirror image of itself (based on the reference line being the mirror).

BSEN 22553 specifics (Fig. 1.12)

- BSEN 22553 does not have a single reference line with the symbols placed above and/or below it to specify if the weld is on the 'arrow side' or 'other side', but adds a dashed line

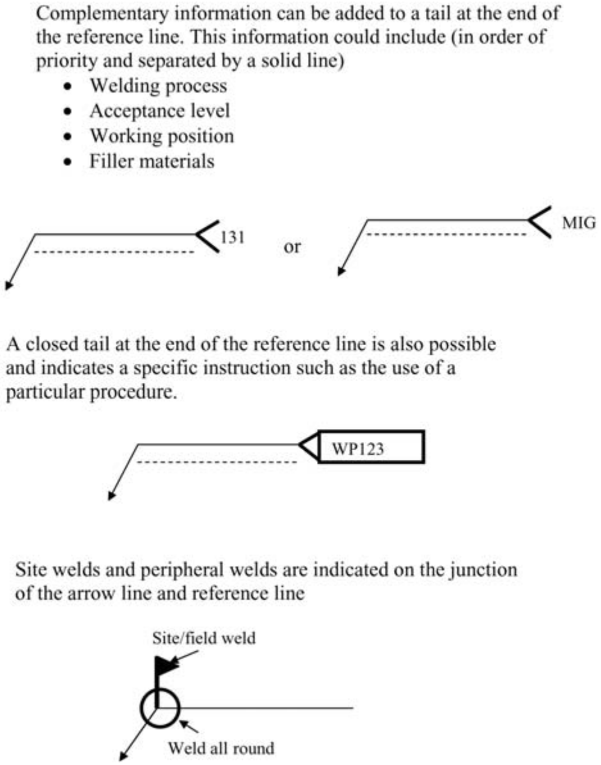

Complementary information can be added to a tail at the end of the reference line. This information could include (in order of priority and separated by a solid line)

- Welding process
- Acceptance level
- Working position
- Filler materials

A closed tail at the end of the reference line is also possible and indicates a specific instruction such as the use of a particular procedure.

Site welds and peripheral welds are indicated on the junction of the arrow line and reference line

Figure 1.11 Supplementary information

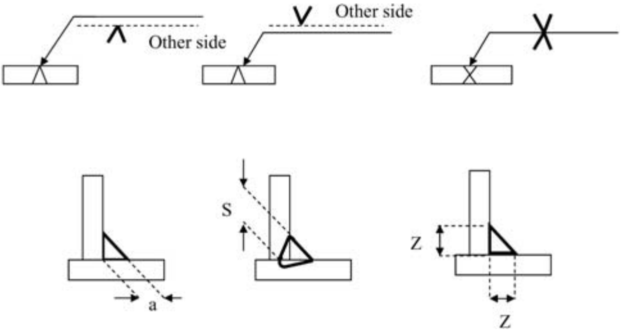

Figure 1.12 Specifics of BSEN 22553

to specify the ‘other side’. This makes good sense because engineering drawings represent what you cannot see as a dashed line and obviously you cannot see the ‘other side’ of a joint. It helps reduce the risk of drawings being read wrongly and welds being incorrectly placed on the wrong side of a joint.

- The dashed line specifying ‘other side’ can be omitted if a double-sided weld is symmetrical about the reference line.
- Fillets leg sizes must be prefaced with the letter *z*.
- Nominal fillet throats are preceded by the letter *a*.
- Effective throat thickness for deep penetration fillet welds and partial penetration butt welds is preceded by the letter *s*.

AWS A2.4 specifics (Fig. 1.13)

- AWS A2.4 may use more than one reference line from the arrow line to indicate the sequence of welding.
- Weld dimensions may be given as fractions or decimals and in metric or imperial units.
- Welding processes are indicated using standard AWS abbreviations.
- Single plate preparations are indicated by a directional

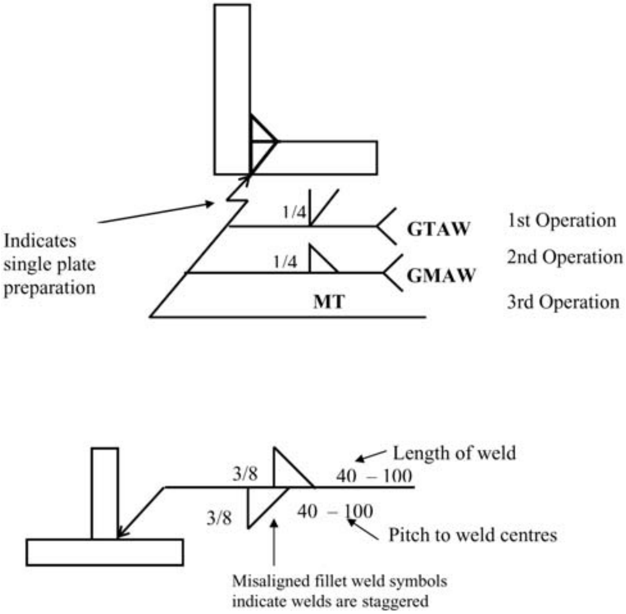
1/4
GTAW
1st Operation
Indicates
single plate
preparation
1/4
GMAW
2nd Operation
MT
3rd Operation
Length of weld
3/8
40 – 100
3/8
40 – 100
Pitch to weld centres
Misaligned fillet weld symbols
indicate welds are staggered

Figure 1.13 Specifics of AWS A2.4

change of arrow line but the arrow remains pointing to the plate edge requiring preparation.

- When plate preparation within a joint is obvious (i.e. a T joint) then the direction of the arrow line is optional.
- AWS A2.4 dimensions the pitch of intermittent fillet welds and plug welds to the centre of each weld. (The BS and BS EN codes dimension these to the start of each weld.)

Chapter 2

Duties of a Welding Inspector

A welding inspector has the responsibility to monitor all aspects of the welding process before, during and after welding to ensure that the finished weldment is fit for purpose. Fitness for purpose in this sense normally means that the finished weldment has been produced in accordance with a code or specification and complies with the stated requirements of that code or specification. The inspector must therefore be proficient in carrying out visual inspections and assessing his or her findings in accordance with the relevant code or specification acceptance criteria. The following checks are commonplace.

Before welding

Before welding is carried out the inspector checks that all of the welding variables specified on the weld procedure specification (WPS) can be achieved and that the correct equipment and documentation is available. The level of checking will vary depending on the code requirements but may involve the following checks.

Documentation

- Drawings, codes, specifications and standards are available and at the correct revision state.
- Correctly qualified welding procedures are in place covering the welding processes to be used in production.
- Welders are correctly qualified for the welding processes and the WPSs to be used in production.
- Correct NDE procedures and techniques are in place.
- NDE and heat treatment personnel are suitably qualified.
- Correct preheat and post-weld heat treatment procedures specifying temperatures, holding times and temperature measurement methods.

Equipment

- Welding machines are in date for calibration or validation (welding machines with meters are normally calibrated while machines without meters are validated).
- All ancillary equipment such as ovens, quivers, remote controls, torches, cables, etc., are in good condition.
- Any heat treatment equipment is in good condition and in date for calibration if required.

Materials

- Verify the correct materials are being used and are in an acceptable condition.
- Ensure the material certificates are the correct grade and the batch or heat numbers correspond to the actual material.
- Verify the correct consumables (covered electrodes, wires, fluxes or gases) are being used and have been correctly prepared.
- Verify any consumables certificates match consumables batch numbers.

Safety

- Ensure all safety precautions are being adhered to and that any work permits required are in place.
- Ensure all electrical equipment is in date for test.
- Ensure correct and sufficient ventilation is in place.
- Ensure the correct PPE is being used.

Weld joint preparation

- Ensure the correct weld joint preparation has been formed in accordance with the specified procedure using an approved method (flame cut, ground, machined, etc.).
- Ensure the joint is correctly sized (groove angles, groove radius, root face, etc.) and the fit-up is within tolerance (root gap, alignment, etc.) using correct restraints.
- Ensure any required pre-heat is applied using the correct

heating method and temperature measurement methods are approved for use.

- Ensure temperatures are in accordance with the WPS.
- Ensure any tack welds are correctly applied after any pre-heat requirement.

During welding

The inspector must monitor all aspects of the actual welding to check that the correct heat input is applied to the weld and that the weld is formed correctly in accordance with the WPS. Realistically, it is pointless making all the necessary preparations and checks before welding and then simply assuming that the welder will strictly adhere to the procedure. Many a weld has failed because a welder has taken a shortcut without appreciating the detrimental effect to the mechanical properties of a welded joint. One of the functions of the inspector is therefore to police all the parties involved in the formation and testing of the weld. Checks to be carried out include:

- Monitor and record the welding variables such as amperage, run-out lengths and voltage.
- Monitor and record environmental conditions (rain, snow, wind, etc.) that could affect the finished weld.
- Monitor and maintain pre-heat temperatures.
- Monitor the interpass temperatures and maintain them as required.
- Ensure interpass cleaning is carried out correctly.
- Inspect critical root runs and monitor hot pass times.

After welding

Once all welding is completed it is the responsibility of the inspector to ensure that the required quality inspections and tests are carried out by qualified personnel and meet the relevant code or specification acceptance requirements. The main checks are to:

- Ensure any required PWHT is correctly applied and

maintained by suitably qualified persons using suitable procedures and equipment. PWHT may be required before and after NDE when hydrogen cracking is a possibility.
- Ensure required NDE is carried out and review the results.
- Visually inspect and measure the completed weld and assess any imperfections in accordance with the relevant code acceptance criteria.
- Maintain weld records with weld identification.
- Ensure any required pressure testing of piping and pressure vessels is carried out after PWHT and results recorded.
- Take final dimensional checks after pressure testing.
- Collect and collate all the relevant documentation and pass the completed package to the relevant acceptance authority.

Repairs

When a defect is identified during the inspection, three main options exist, depending upon the severity of the defect and the component being welded. The component can be scrapped, it can be given a concession or it can be repaired.

- A **concession** is a situation where a defect is deemed not to have too detrimental an effect on the component and the client agrees in writing that the defect can remain in place. In some cases the concession conditions may mean a limitation is placed on the component service conditions and/or a repair is required at a more convenient time.
- A **repair** can in some cases require a different welding process, different consumables or some other change to the original procedure to be used. There may be limitations on the number of repairs or thermal cycles applied to the component depending on the material used. A repair procedure should be in place detailing what actions are required.

The simplest way to think about carrying out a repair is that the inspector monitors all aspects of the repair process before, during and after welding the component. This gives the best possible assurance that the weld is fit for its specified purpose.

Chapter 3

Analysis of a Fusion Weld

Components of a welded joint

A welded joint comprises the weld, the heat affected zone (HAZ) and the base metal. The components of the weld and HAZ are shown in Fig. 3.1. The main features are as follows.

Excess weld metal

This is excess to requirements and does not add to the strength of the weld. It is also known as the cap (a slang term), reinforcement (BS term) or crown (US term). Reinforcement is a poor term because it implies that adding reinforcement strengthens the weld, which is not the case. In practice, the addition of reinforcement (or excess weld metal) can effectively weaken the weld by increasing the stress concentration at the weld toes, hence reducing the fatigue life of the joint.

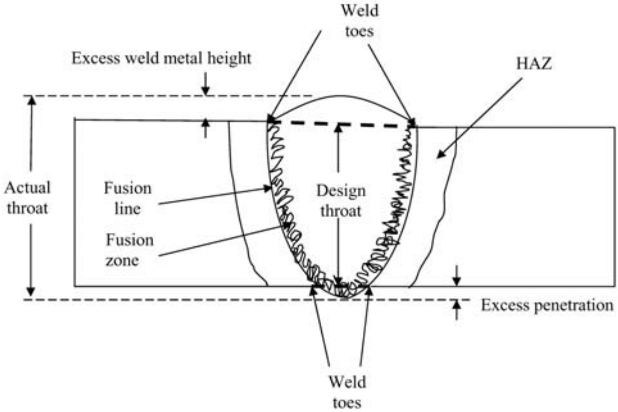
Weld toes
Excess weld metal height
HAZ
Actual throat
Fusion line
Fusion zone
Design throat
Excess penetration
Weld toes

Figure 3.1 Components of a welded joint

Weld toes

These are the points where the weld metal adjoins the base metal. There will normally be four weld toes on a full penetration butt weld (two on the face and two on the root) and two on a fillet weld. The greater the angle at the weld toe, the greater the risk of a fatigue failure if the component is subjected to cyclic stress.

Fusion line

This is where fusion takes place between the melted and unmelted material. It is commonly referred to as the fusion boundary or weld junction.

Fusion zone

The fusion zone is the region within the weld that contains the greatest dilution of filler metal with melted base metal. This region of highest dilution can contain defects owing to impurities or contaminants contained in the base metal being drawn into the weld. The centre of the weld will be the area of the lowest dilution and may consist of filler metal only.

HAZ

This is the region of the base metal that has not been melted but has been affected by the heat and had a change made to its grain structure. Cracks often occur in the HAZ after welding has taken place owing to it being hardened by the formation of a martensitic grain structure on cooling.

Excess penetration

Excess penetration is excess weld metal formed in the root of the weld. Some codes may specify a limit to the penetration due to the toe blending and/or the bore restriction caused in piping.

Design throat thickness

The throat is where the strength is contained within the weld. It is equal in size to the thinnest of the base materials being joined.

Actual throat thickness

This is the actual measurement made from the weld face to the root. The actual strength of the weld can be calculated using this actual throat measurement (but any measurement in excess of the design throat is ignored). This is more relevant for partial penetration welds or where root concavity is present.

What makes a good fusion weld?

There are four factors that must be satisfied to produce a good fusion weld. If any of these factors is not achieved then the result will be a weld that may not be fit for purpose. The four factors are as follows:

- Heat input. This is the heat that melts the parent material and filler (if required) to give the required fusion between the parts being joined. The heating could be from an arc, a laser, an electron beam or an oxyacetylene gas mixture. The most common process is arc welding.
- Protection from atmosphere. If exposed, a welding arc will pick up gases such as oxygen, nitrogen or hydrogen from the atmosphere and pass them into the weld pool. These gases can have a detrimental effect on the finished weld so the arc needs to be protected from them. Common methods of atmospheric protection used in arc welding processes include the use of:
  - a shielding gas for processes such as TIG, MIG, MAG and PAW;
  - a gaseous shroud, which is produced as the flux coating on the electrode melts in MMA or FCAW;
  - a flux blanket, which covers the arc in SAW;
  - a combination of a shielding gas and a gaseous shroud from melting flux in secondary shielded FCAW.
- Protection from external and internal contaminants (cleaning). The weld metal can pick up contaminants from the surface of the material so it is important that the material surface is free from scale, rust, paint, grease,

moisture and other possible weld contaminants. Internal contaminants such as sulphur or oxides can be present within the material and may also have to be removed or neutralised during welding. In effect, the process must be able to 'clean' the material and weld pool during welding to afford the correct protection. Protection from contaminants can be achieved by:

  - mechanical cleaning of the component (grinding, wire brushing, abrasion, etc.);
  - chemical cleaning of the component and filler wire/rods (acid, acetone, etc.);
  - use of a flux (containing deoxidisers);
  - use of correct polarity (d.c. +ve or a.c. when welding Al or Mg alloys to remove the surface oxide layer, called cathodic cleaning).

- Adequate mechanical properties. The finished welded joint must have adequate properties such as strength, toughness, hardness and ductility in the base material (including the HAZ) and weld metal. These properties are achieved (depending on the welding process) by:
  - using the correct base materials;
  - using the correct consumables (filler wire, electrodes, shielding gas, flux);
  - using correctly prepared consumables (correctly heat-treated electrodes, etc.);
  - using the correct pre-heat and/or post-weld heat treatments;
  - using the correct heat inputs (voltage, current and travel speeds).

All of the above factors can be achieved by adhering strictly to the requirements of an approved welding procedure specification, which contains the essential, supplementary essential (ASME IX only) and non-essential variables necessary to produce a sound weld.

In practice, when welders get lazy, take shortcuts and do not fully comply with a correctly tested and qualified welding procedure, problems can arise. A common example is when tack welds are done without applying the specified pre-heat. This leads to the tacks and HAZ around them being harder than they should be because they cool down too quickly (remember, the main reason for pre-heat is to retard the cooling rate and reduce temperature variations across the weldment). These hardened areas within a highly stressed region can then suffer cracking, with possible catastrophic consequences if they are not found.

Weld joint: preparation methods

The type of joint preparation and how it is made can have an effect on the final weldment properties. For example:

- Weld preps made using thermal cutting processes can be affected by the heat of the cutting process. This can lead to a possible loss of toughness or cracking in the HAZ, so approximately 3 mm of material is normally removed from the prep by mechanical means after thermal cutting. Typical thermal cutting processes (used to form straight-sided preps such as square edged or bevelled) include plasma, oxy-fuel gas or oxy-arc processes.
- If cutting is by an oxygen fuel gas process such as oxyacetylene then the weld preps can gain excess carbon from the process, leading to cracking in the weld (the carbon increases the weld metal hardness and therefore the risk of cracking). This is another reason to remove 3 mm from the prep by mechanical means after cutting.

  Remember that oxy-fuel gas cutting does not actually melt and blow away the molten material like plasma or oxy arc cutting does. What it actually does is to heat up the material to its ignition temperature and then introduces a stream of oxygen, turning the metal into instant rust, which is then blown away. This is why stainless steel (a rust-resistant material) cannot be cut by oxy-fuel gas unless special powders are added to the process.

- If mechanical cutting such as machining (normally required for weld joint preparations containing a radius such as a U or J prep) or shearing is used then consideration must be given to the possibility of having cutting fluids trapped in the cut edges. These can cause porosity or other weld defects unless they are properly cleaned out (the edge may look smooth but under a microscope it will be full of peaks and troughs).

Weld joint: shape

The shape of the joint (joint type) can affect not only the final mechanical properties of the weldment but also the level of distortion produced in it. The mechanical properties can be affected because a change of joint type may affect the overall heat input. This is due to the possible requirement for more (or less) weld metal to be welded into the joint. The actual level of final distortion is determined by factors such as the weld metal shrinkage direction (based on the angle of the groove) and whether the weld is formed from one or both sides.

Figure 3.2 shows the components of a typical single V joint and U joint, and some advantages of the single U over a single V in thicker materials.

Residual stress and distortion

Residual stresses are those stresses that remain in a component after a procedure such as bending or welding has been carried out. If a plate is bent then bending stresses will be introduced into the bend and the maximum tensile stress will be at the outer convex surface. This stress remains in the plate and can cause failure, especially if a defect or stress raiser is present in the stressed region (this of course is how bend tests are carried out).

Residual welding stress in a welded joint comes from the uneven expansion and contraction of the joint coupled with a restraint. Consider the following example.

An unrestrained strip of metal is heated up and expands. If

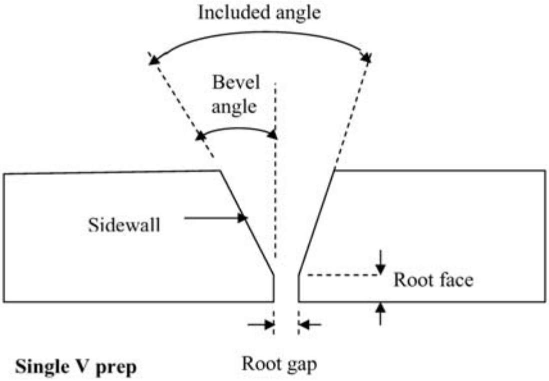
Included angle
Bevel angle
Sidewall
Root face
Root gap

Single V prep

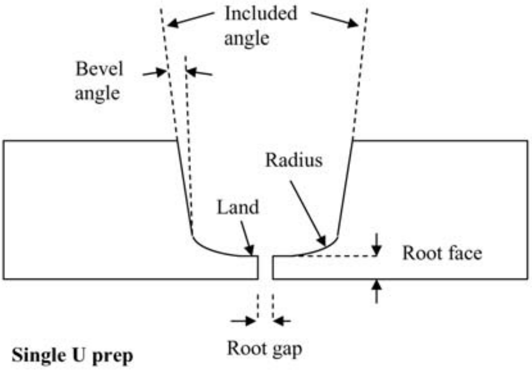
Included angle
Bevel angle
Radius
Land
Root face
Root gap

Single U prep

Some benefits of single U versus single V in thicker materials

- Less volume to fill
- Less filler required
- Lower heat input
- Less distortion

Figure 3.2 Single V and U preparations

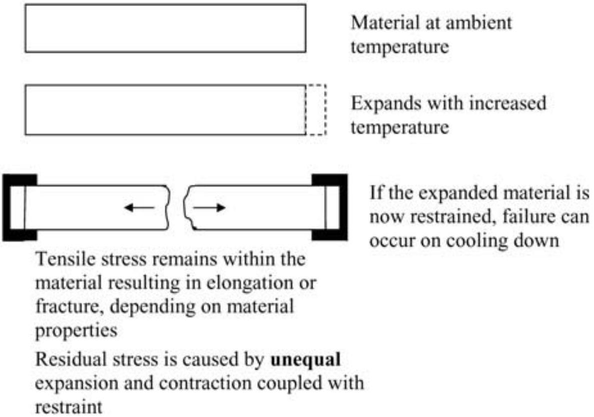
Material at ambient temperature
Expands with increased temperature
If the expanded material is now restrained, failure can occur on cooling down
Tensile stress remains within the material resulting in elongation or fracture, depending on material properties
Residual stress is caused by **unequal** expansion and contraction coupled with restraint

Figure 3.3 Unequal expansion and contraction

it is then allowed to cool it will shrink back to its original size and there will be no residual stress present. If, however, the expanded metal was to be restrained by jigs (see Fig. 3.3) to try and prevent it shrinking when it cools, then a tensile stress will be introduced into the material and can cause the material to deform plastically or fracture. This, in effect, is residual stress induced through uneven expansion and contraction coupled with restraint.

Now consider the mechanics of welding where a pool of molten weld metal cools down and contracts (shrinks). The base material acts as a restraint and tries to prevent the weld pool contracting, providing the ideal conditions required to leave residual stress in the completed weldment. It can be deduced from this that there will always be residual stresses present in welded joints. In some cases this may be high enough to approach the yield point.

Mechanical restraints such as jigs and clamps increase restraint stresses further by preventing the movement that would normally occur as the molten weld metal cools down and shrinks. This causes the joined components to distort. Allowing distortion to take place helps to reduce the overall stress in the component, but may make the final dimensional criteria unacceptable with the code requirements. Methods can be employed that can reduce both residual stress and distortion (see later).

Residual stress in a weld acts in many directions in a complex pattern because any weld will change volume and shape in all directions. There are three main stress directions caused through weld metal shrinkage to consider: the longitudinal, transverse and short transverse directions (see Fig. 3.4).

Distortion

If an unrestrained material is heated and cooled uniformly then there will be no distortion produced within it. If, however, the material is subjected to localised heating and cooling then distortion will occur due to the different rates of expansion and contraction experienced throughout the material, caused by temperature gradients. Welding does not heat and cool the material uniformly because the weld region will always be hotter than the surrounding region and the weld metal will therefore expand and contract at a much greater rate than the adjoining base material. The base material can be pre-heated to reduce the temperature gradient spreading from the weld outwards but obviously

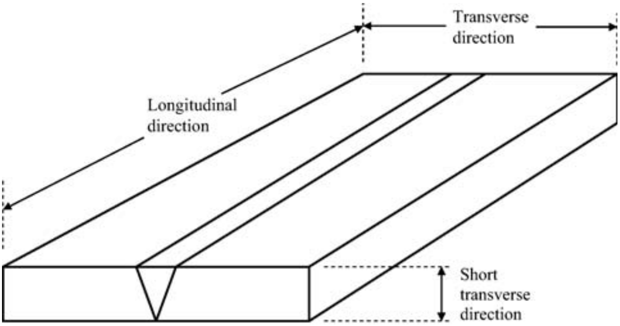

Figure 3.4 Main stress directions in a weld

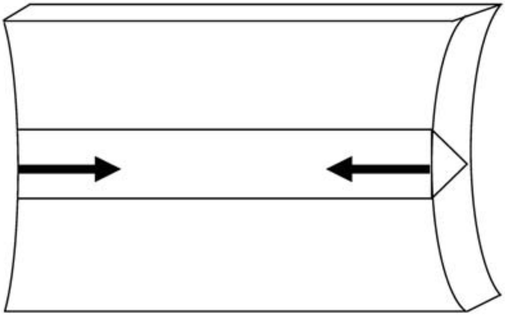

Longitudinal distortion

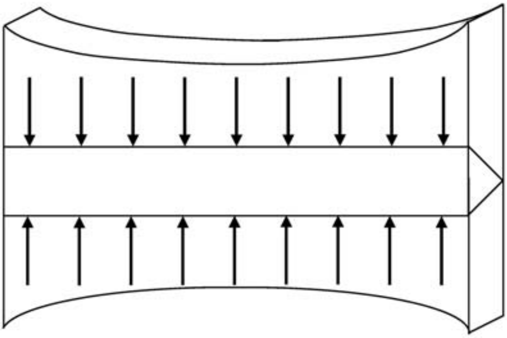

Transverse distortion

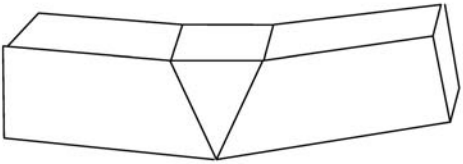

Angular distortion

Figure 3.5 Distortion caused by the effects of shrinkage

cannot be uniformly heated to the melting temperature required to weld. The net effect of this is that some distortion will always occur because the weld effectively acts as a form of localised heating.

The coefficient of thermal expansion of a material plays a large part in how much welding stress is introduced into a material and how much distortion can occur. The higher the coefficient of expansion, the higher the distortion level, which is why stainless steels suffer higher distortion levels than plain carbon steels.

Simplistically, the main distortions to consider are caused by the weld metal shrinkage in the longitudinal, transverse and short transverse directions. In reality, the shrinkage and distortion will follow a very complex pattern but a simplified exaggerated view of the effects is shown in Fig. 3.5.

The shape of a weld preparation will affect how much distortion is produced, because bevel angles will help to direct the distortion direction. The material thickness, amount of weld metal required and size of individual weld runs also affect stress and distortion levels. A single V butt will suffer considerable distortion, especially during the initial root runs when there is no restraint, because of the high level of weld metal shrinkage. Although further runs add to the distortion, the overall effect is not cumulative because the previous runs will help to restrain the joint from moving. A square-edged closed butt will actually suffer very little distortion. Figure 3.6 compares distortion levels based on the bevel angle of two single V butt welds and a square-edged closed butt weld.

Minimising stresses and distortion

There are various ways to counteract the effects of residual stress levels and distortion depending upon the type of welding process and joint type that is being used.

Offsetting (Fig. 3.7) is where plates are offset to a preset angle and welded without restraint. As the weld solidifies and shrinks it distorts the plates and pulls them into the correct

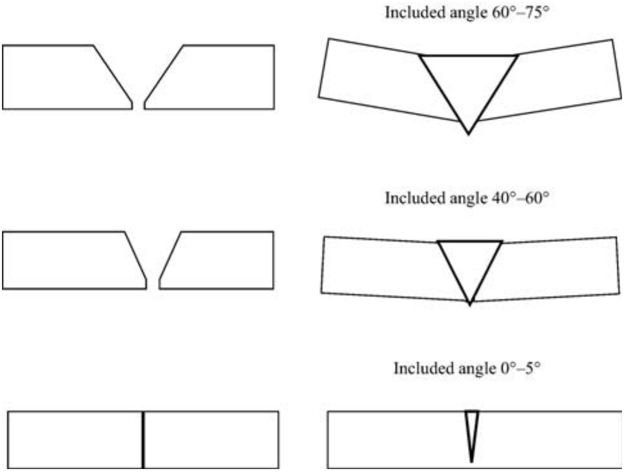
Included angle 60°–75°
Included angle 40°–60°
Included angle 0°–5°

Figure 3.6 Distortion and bevel angles

position. It can be used on fillet welded T-joints and butt welds but the amount of offset required by welding is normally only determinable by trial and error. It is a very cheap and simple way to control distortion if carried out

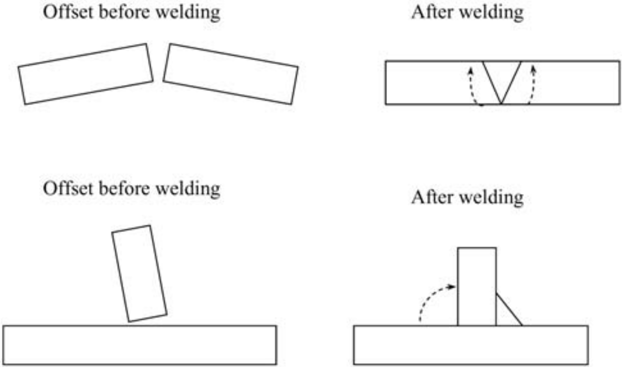
Offset before welding
After welding
Offset before welding
After welding

Figure 3.7 Offsetting

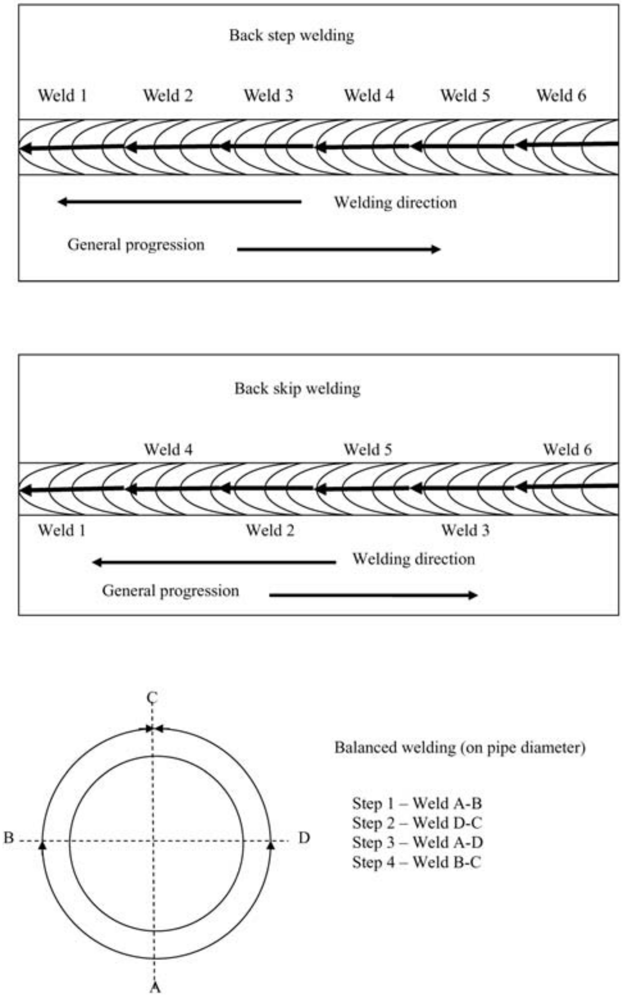
Back step welding
Weld 1
Weld 2
Weld 3
Weld 4
Weld 5
Weld 6
Welding direction
General progression
Back skip welding
Weld 4
Weld 5
Weld 6
Weld 1
Weld 2
Weld 3
Welding direction
General progression
C
B
D
A
Balanced welding (on pipe diameter)
Step 1 – Weld A-B
Step 2 – Weld D-C
Step 3 – Weld A-D
Step 4 – Weld B-C

Figure 3.8 Sequential welding examples

correctly. By allowing the plates to move and not restraining them the level of residual stress in the welded joint is reduced.

Sequential welding (Fig. 3.8) is a technique utilised to control the level of distortion by welding the joint in a specific way. There are various sequential welding techniques including balanced welding (welding about a neutral axis of the weld), back skip welding or back step welding (a short length is welded and the next weld starts a short distance behind the previous weld length).

Mechanical restraints including clamps and jigs allow accurate positioning of the component parts. They are normally left in position until the components are tack-welded together and then removed to allow the welder full access. Restraints will reduce the level of distortion by physically holding the components in position but will increase the stress levels introduced into the weld.

Flame straightening uses an oxyacetylene torch (but not a cutting torch) to give a localised intense heat. The distortion caused by this localised heating can be used in some cases to straighten or modify the shape of a component. Typical examples of this method would be straightening flanges or removing bulges from insert plates.

Heat treatment methods such as post-weld heat treatment stress relief can remove a high percentage of residual stresses.

Mechanical stress relieving methods include:

- using ultrasound to stress relieve fabrications;
- peening weld faces using pneumatic needle guns, to redistribute residual stresses by placing the weld face in compression.

Chapter 4

Materials and Their Weldability

There are many different definitions concerning the weldability of steel because it often means different things to different people. Simplistically it can be defined as the ability of a material to be welded and still retain its specified properties. This ability to be welded successfully depends on many factors including the type and composition of the material, the welding process used and the mechanical properties required. Poor weldability generally involves some type of cracking problem and this is dependent upon factors such as:

- residual stress level (from unequal expansion and contraction due to welding);
- restraint stress level (from local restraint such as clamps, jigs or fixtures);
- presence of a microstructure susceptible to cracking (the base material may have a susceptible microstructure or the HAZ and/or weld metal may form a microstructure susceptible to cracking owing to the welding).

Carbon equivalency

The susceptibility of a microstructure to cracking is heavily influenced by the amount of carbon and the type and amount of other alloying elements present in the steel. The carbon and other alloying elements can be put into a formula that determines the carbon equivalency value (Cev) of the material. This Cev is a measure of the hardenability of the steel. The higher the Cev, the more susceptible the material will be to cracking by brittle fracture.

Other factors affecting the likelihood of cracking include the base metal thickness and the combined joint thickness (i.e. a butt weld has two thicknesses whereas a fillet weld has three). The combined joint thickness is important because

Woodhead Publishing Ltd – A Quick Guide to Welding and Weld Inspection
Data Standards Ltd, Frome, Somerset – 17/9/200904QG Welding chap4.3d Page 36 of 48

$$Cev = C + \frac{Mn}{6} + \frac{Cr + Mo + V}{5} + \frac{Ni + Cu}{15}$$

This is not the full formula but contains as much as a welding inspector will need to use.

Figure 4.1 Simplified carbon equivalency formula

each material thickness acts as a heat sink and can cool the weld area more quickly, making it harder and therefore more susceptible to cracking.

Figure 4.1 shows the formula used to determine the Cev for a material. As a general guide the following Cev levels determine the weldability of steels:

- up to 0.4%: good weldability;
- 0.4 to 0.5%: limited weldability;
- above 0.5%: poor weldability.

Classification of steels

Low carbon steel: contains 0.01–0.3% carbon
Medium carbon steel: contains 0.3–0.6% carbon
High carbon steel: contains 0.6–1.4% carbon

Plain carbon steels contain only iron and carbon as main alloying elements. Traces of other elements such as Mn, Si, Al, S and P may also be present. It can be seen from the carbon diagram in Fig. 4.2 that an increase in carbon causes the ductility of steel to decrease while the tensile strength and hardness increase. Note also how the maximum tensile strength of plain carbon steel is achieved at 0.83% carbon content.

Alloy steels contain alloying elements such as Mn, Mo, Cr and Ni and are divided into two groups:

- Low alloy steels contain $<$ 7% total of other alloying elements.
- High alloy steels contain $>$ 7% total of other alloying elements.

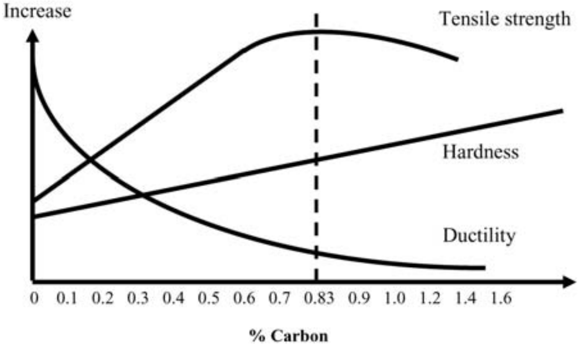

Figure 4.2 Carbon diagram

Alloying elements

The following are some basic properties of alloying elements added to steels:

Iron (Fe)	This is the basic constituent of steels.
Carbon (C)	Increases tensile strength and hardness but reduces ductility.
Manganese (Mn)	Improves toughness and strength when alloyed at levels of < 1.6% in steels. Can control solidification cracking in steels by neutralising the detrimental effects of sulphur.
Chromium (Cr)	Alloyed at levels > 12% to produce stainless steels. Gives corrosion resistance and promotes through-thickness hardenability. Hardenability is the ability of steel to harden at slower cooling rates when alloying elements are added to it. Do not confuse this term with hardness.
Molybdenum (Mo)	Gives high temperature creep resistance and strength in low alloy steels.
Nickel (Ni)	Improves strength, toughness, ductility

	and corrosion resistance of steels when alloyed at levels > 8%. It promotes austenite formation at temperatures below the lower critical temperature.
Silicon (Si)	Alloyed in small amounts as a deoxidiser in ferritic steels.
Aluminium (Al)	Used as a grain refiner in steels and a deoxidising agent in triple deoxidised steels.
Niobium (Nb) and Titanium (Ti)	Used to help carbide formation to stabilise stainless steel.

Titanium

Titanium is becoming more widely used in industry because of its excellent properties. These are:

- high strength to weight ratio (strong as steel but half the weight);
- excellent corrosion resistance;
- good mechanical properties at elevated temperatures.

Types of titanium in use are:

- Commercially pure (98–99.5% Ti). May be strengthened by small additions of O_2, N_2, C and Fe and is easily welded.
- Alpha alloys. Mainly single-phase alloys with up to 7% Al and a small amount of O_2, N_2 and C. These can be fusion welded in the annealed condition.
- Alpha-beta alloys. Two-phase alloys formed by the addition of up to 6% Al and varying amounts of beta-forming constituents such as V, Cr and Mo. Can be fusion welded in the annealed condition.
- Ni–Ti alloys that contain a large beta phase, stabilised by elements such as Cr, are not easily welded.

Commercially pure grades and variants of the 6% Al and 4% V alloy are widely used in industry.

Welding problems

There are many problems with welding titanium because it reacts readily with air, moisture, grease, dirt and refractory to form brittle compounds. It is therefore absolutely essential that the weld joint surfaces and filler wire are free of any contamination before and during the entire welding operation. Titanium cannot be welded to most other metals because embrittling metallic compounds that lead to weld cracking are formed.

Above 500 °C, titanium has a very high affinity for hydrogen, nitrogen and oxygen so an inert atmosphere protection must be maintained until the weld metal cools below 426 °C (800 °F). This normally means that a trailing gas shielding is required, i.e. a process utilising inert gases, and welding must be slow enough to enable the trailing gas shielding to be utilised.

- Improper welds might be less corrosion-resistant compared to the base metal.
- Reaction with gases and fluxes makes common welding processes such as gas welding, MMA, FCAW and SAW unsuitable.

Typical imperfections

Titanium welding imperfections are:

- **Porosity**, caused by gas bubbles trapped during solidification. The gas is hydrogen from moisture in the arc environment or from contamination on the filler and/or parent metal surface.
- **Contamination cracking**, caused by iron particles present on the material surface dissolving in the weld metal and reducing its corrosion resistance. They can also cause embrittlement at high iron concentrations. Iron particles present in the HAZ can melt and cause microcracking and corrosion.
- **Embrittlement**, caused by weld metal contamination from

gas absorption (due to poor shielding) or by dissolving contaminants such as dust (iron particles) on the surface.

On the positive side, solidification cracking and hydrogen cracking are not normally found in titanium or its alloys.

A thin layer of surface oxide generates an interference colour at the weld that can indicate whether the shielding was adequate or not. Typical colours are as follows:

- Silver or light straw is acceptable.
- Dark blue may be acceptable for certain service conditions.
- Light blue, grey, white or yellow powders are not acceptable.

Defect avoidance

Titanium is very expensive and many defects can be avoided if the following precautions are taken:

- Avoid steel fabrication operations near titanium components.
- Have dedicated tools used only for titanium.
- Scratch-brush the joint area immediately before welding.
- Do not handle the cleaned component with dirty gloves.
- Cover components to avoid airborne dust and iron particles settling on the surface.
- Maintain adequate shielding gas, trailing gas and purging gas levels.
- In short – keep it clean.

Duplex stainless steel

Duplex stainless steels have a two-phase structure consisting of almost equal parts of austenite and ferrite. They have become very popular because they are approximately twice as strong as common austenitic stainless steels but less expensive, owing to the lower levels of nickel they contain. The characteristic benefits of duplex stainless steels (strength, toughness, corrosion resistance and resistance to stress corrosion cracking) are achieved when there is at least 25% ferrite, with the balance being austenite.

The ferrite in a duplex weld metal is typically in the range of 25 to 60%. In some welding processes utilising a flux, the phase balance of the filler is biased to increased austenite to offset the loss of toughness associated with oxygen pickup from the flux. Thermal expansion and conductivity of a duplex stainless steel are between that of carbon steel and austenitic stainless steel. The operating temperature is normally kept below 300 °C to avoid a degradation mechanism called '475 °C embrittlement'.

Typical defects

- The problems most typical of duplex stainless steels are associated with the heat affected zone (HAZ) rather than the weld metal. The HAZ can suffer from:
  - loss of corrosion resistance;
  - loss of toughness;
  - post-weld cracking.
- The duplex structure is very sensitive to contaminants, particularly moisture.
- Detrimental reactions occur to the material properties if heat input times keep temperatures within the 705 to 980 °C range for too long.
- Rapidly quenched autogenous welds (welds without filler) such as arc strikes and repairs to arc strikes tend to have ferrite levels greater than 60%. These welds can have low toughness and reduced corrosion resistance.

Defect avoidance

- Allow rapid (but not extreme) cooling of the HAZ.
- Limit the temperature of the workpiece because it provides the most effective cooling of the HAZ.
- Limit the maximum interpass temperature to 150 °C (300 °F).
- When a large amount of welding is to be performed, plan

the welding to provide enough time for cooling between passes.

- Avoid PWHT stress relief because duplex steels are sensitive to even relatively short exposures to temperatures in the 300 to 1000 °C range.
- PWHT stress relief in the 300 to 700 °C range may cause precipitation of the alpha prime phase ('475 °C embrittlement'), causing a loss of toughness and corrosion resistance.
- Stress relief in the range of 700 to 1000 °C leads to rapid precipitation of intermetallic phases, resulting in loss of toughness and corrosion resistance.
- Heat treatment of duplex steel, for whatever reason, should be a full solution anneal followed by water quenching.

In summary, the best way to avoid problems with duplex steels is to avoid excessive ferrite levels and limit the total time at temperature in the HAZ.

Material properties

Materials are chosen for service use based on the properties that they possess.

Ductility is the ability of a material to be drawn or plastically deformed without fracture. It is therefore an indication of how 'soft' or malleable the material is. The ductility of steels varies depending on the types and levels of alloying elements present. An increase in carbon, for example, will increase the strength but decrease the ductility.

Hardness is the ability of a material to resist abrasion or penetration on its surface. The harder the material, the smaller the indentation left by an object such as a ball or diamond being impressed upon it. As a general rule there will be a higher risk of cracking as hardness increases.

Toughness is the ability of a material to resist impact (i.e. absorb the energy of an impact). The general rule is that a higher toughness will lessen the risk of cracking.

Heat treatment of steels

Metallic materials consist of a microstructure of small crystals called grains. The grain size and composition help determine the overall mechanical behaviour of the metal. Heat treatment provides an efficient way to manipulate the properties of the metal by controlling the formation of structures, changing the metal properties or controlling the rate of cooling within the microstructure. All heat treatments cycles contain three parts: the rate of climb to the hold temperature (including any hold points), the hold (or soak) time and the cooling rate (see Fig. 4.3).

Heat treatment methods include the use of open flames, electric resistance heating blankets, furnaces and autoclaves. Temperature measurements are taken using indicating crayons (tempil sticks), thermocouples, pyrometers or other methods depending on the level of accuracy required. An inspector should ensure that all heat treatments are carried out in accordance with a specified procedure, make a record of all parameters and ensure that all documents are retained within the quality files.

Annealing

Annealing is a heat treatment carried out to soften and reduce internal stresses on metals that have been work-hardened. The first stage of the annealing process is the

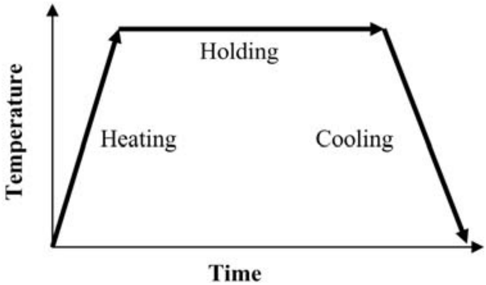

Figure 4.3 Heat treatment cycle

softening of the metal by removal of crystal defects and the internal stresses that they cause. The second stage is recrystallisation, where new grains replace those deformed by internal stresses. Further annealing after recrystallisation will lead to grain growth, which starts to coarsen the microstructure and may cause poor mechanical properties such as a loss of strength and toughness. In work-hardened non-ferrous metals, annealing is used to recrystallise work-hardened grains and the cooling rate is not always critical.

For steel, there are two basic kinds of annealing: full annealing and subcritical annealing.

Full annealing

This is commonly carried out on large castings. The steel is heated to around 50 °C above its upper critical temperature (UCT) and held in a furnace for sufficient time to allow the temperature to become uniform throughout the steel. It is then slowly cooled, causing grain growth. The tensile strength will not be particularly improved but toughness and ductility will increase. The upper critical temperature (UCT) of plain carbon steels ranges from 723 to 910 °C depending on carbon content, so the actual annealing temperature will be dependent upon the carbon content of the steel. Above the UCT the steel structure will be an austenite structure.

Subcritical annealing

Subcritical annealing methods are used to increase the machinability of high carbon steels or for softening worked-hardened mild steels to allow further cold work to be applied. The steel is heated to a temperature above which recrystallisation will take place but below the lower critical temperature (LCT) of 723 °C. This recrystallises the distorted ferrite grains so that the structure becomes softer again. The recrystallisation temperature and time held at temperature will be dependent on the carbon content of the steel.

Annealing generally puts a metal, or alloy, into its most ductile condition. In steels the resultant large grain structure means a reduction in toughness with low impact strength. The lower critical temperature (LCT) is the temperature below which the austenite forms into ferrite and cementite.

Normalising

When an annealed part is removed from the furnace and allowed to cool in air, it is called ‘normalising’. It is often used for hardenable steels to regain toughness after high heat input processes have formed large grain structures. The steel is heated to just above its upper critical temperature, held for a specified period and then allowed to cool in air. Small grains are formed, which give a harder and much tougher metal with normal tensile strength rather than the maximum ductility achieved by annealing.

Hardening (quenching)

Hardening of steel is achieved by heating the alloy to above its upper critical temperature until it is a fully austenitic structure and then cooling it rapidly with forced air, oil, water or brine. Upon being rapidly cooled, a portion of austenite (dependent on alloy composition) will transform into martensite. Martensite is very hard and strong but too brittle for most applications. It must therefore be subjected to a process called tempering, which will temper the martensite into a very strong and tough structure. Most applications require that quenched parts be tempered, to impart some toughness and further ductility, although some yield strength is lost.

Tempering

Tempering of steel is used to transform a hard and brittle martensitic structure into a tougher, more ductile structure. There is always a trade-off between ductility and brittleness, and the precise control of time and temperature during the tempering process is necessary to achieve a structure with the correct balance of these mechanical properties. The steel is normally tempered after thermal hardening by heating between 150 and 650 °C (300 and 1200 °F) dependent upon the material properties and the specific mechanical properties required. Tempering can continue up to the lower critical temperature of 723 °C, at which point most of the extra hardness produced by thermal hardening will have been removed, but the fine grain structure produced by the hardening process will remain. Quenched and tempered (QT) steels are normally tempered from between 550 and 650 °C giving them good toughness and strength

Quenching and tempering of 'precipitation hardening' alloys

Precipitation hardening metal alloys have their alloying elements trapped in solution during quenching, resulting in a soft material. Ageing a 'solutionised' metal will allow the alloying elements to diffuse through the microstructure and form intermetallic particles, which fall out of solution and increase the strength of the alloy. Alloys may age naturally at room temperature, or artificially at elevated temperatures. Some naturally ageing alloys can be prevented from age hardening until needed by storing at subzero temperatures.

Precipitation hardening alloys include 2000, 6000 and 7000 series aluminium alloys, some superalloys and some stainless steels. An age hardening alloy can be tempered after quenching by heating at temperatures below the solutionising temperature. During tempering, the alloying elements will diffuse through the alloy and react to form intermetallic compounds. These precipitate out and form small particles that strengthen the metal by impeding the movement of dislocations through the crystal structure of the alloy.

The mechanical properties of an alloy can be determined by careful control of the tempering time and temperature, affecting the size and amount of precipitates. Artificially aged alloys are tempered at elevated temperature, while naturally ageing alloys may be tempered at room temperature. Some superalloys may be subjected to several tempering operations where a different precipitate is formed during each operation. This results in a large number of different precipitates that are difficult to drive back into solution. This contributes to the high temperature strength of precipitation hardened superalloys.

Stress relief

The primary function of post-weld heat treatment (PWHT) stress relief is to relieve internal stresses in welded fabrications by permitting the steel to creep slightly at an elevated temperature. This elevated temperature lowers the material yield point and allows any high residual welding stresses to exceed the new yield stress, causing localised plastic deformation. When the material is cooled down, the yield point increases again but the residual stresses have now been reduced to a lower level.

Stress relief is also used to reduce the hardness of hardenable steels after welding to help prevent brittle fracture. The procedure is basically a form of subcritical annealing and most steels are stress relieved in the 550–700 °C temperature range depending upon their composition. The holding temperature, holding time and cooling rate are critical and must be sufficient to permit the required changes to take place throughout the whole material thickness. For this reason most codes that specify PWHT stress relief will normally specify the minimum holding time in minutes per millimetre thickness (min/mm) or hours per inch (h/in).

Chapter 5

Welding Processes

Welding processes usually utilise heat and/or pressure to form the welded joint. The heat can come from various sources such as an arc, flame, laser beam, electron beam, friction, etc. We will look at the most commonly used arc welding processes in this chapter.

Manual metal arc (MMA)/shielded metal arc welding (SMAW)

Process description

Figure 5.1 shows the equipment used in this process. Fusion is obtained from the heat of an arc formed between a consumable flux-coated electrode and the workpiece. The arc is protected from the atmosphere by a gaseous shroud produced from the melting flux while the weld metal is cleaned of contaminants by the flux, which forms a slag that floats to the top of the weld (Fig. 5.2). This slag must be removed after each weld run before the next pass is added to prevent slag inclusions within the completed weld.

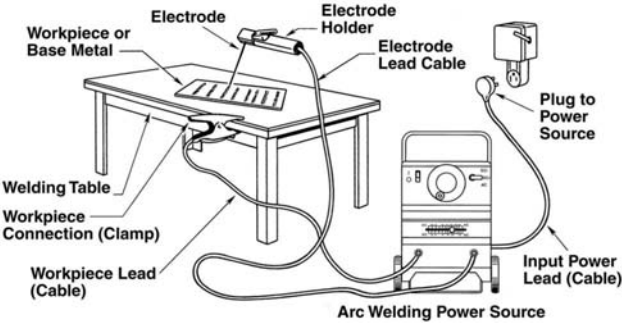
Electrode
Electrode Holder
Workpiece or Base Metal
Electrode Lead Cable
Plug to Power Source
Welding Table
Workpiece Connection (Clamp)
Workpiece Lead (Cable)
Input Power Lead (Cable)
Arc Welding Power Source

Figure 5.1 MMA welding equipment

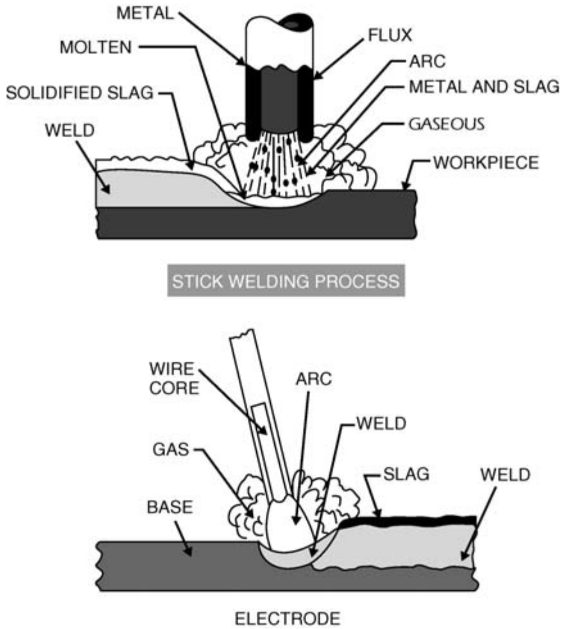
METAL
MOLTEN
SOLIDIFIED SLAG
WELD
FLUX
ARC
METAL AND SLAG
GASEOUS
WORKPIECE
STICK WELDING PROCESS
WIRE CORE
ARC
WELD
GAS
SLAG
WELD
BASE
ELECTRODE

Figure 5.2 MMA welding process

Polarities

Welding can be done by using an alternating current (a.c.) or direct current (d.c.). When direct current is used the welding electrode will be connected to either the positive or negative pole. This is referred to as direct current electrode positive (DCEP) or direct current electrode negative (DCEN). The type of current and polarity is determined by the electrode characteristics. The power source has what is termed a 'drooping' or constant current characteristic (Fig. 5.3). This means that a change of arc length (which controls voltage) will have only a small effect on the welding current as follows:

- As arc length increases, voltage increases and current

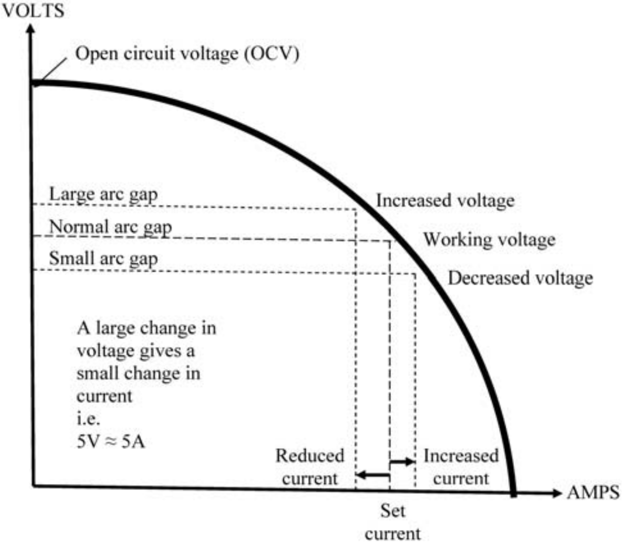

Figure 5.3 Constant current characteristic

decreases. The current decrease is very small and will have little effect on the burn-off rate of the electrode.

- As arc length decreases, voltage decreases and current increases. The current increase is very small and will have little effect on the burn-off rate of the electrode.

Consumables

The only consumables are flux-coated electrodes. These are three main types:

- Basic electrodes. Basic electrodes contain calcium compounds such as calcium carbonate and calcium fluoride in the flux coating. These compounds are helpful for all positional welding as they help produce a fast freezing slag covering on the weld metal. The shielding gas produced from the melting flux to protect the welding arc is mainly CO_2.

Basic electrodes contain low levels of H_2, and are therefore used where high quality welds with good mechanical properties are required (especially high strength welds in high restraint situations, which are susceptible to hydrogen-induced cold cracking (HICC). Basic electrodes require baking at temperatures above 150 °C, storing in an oven at temperatures up to 120 °C, and being used from heated quivers at about 70 °C to ensure they maintain the low H_2 levels required. A modern option is to use electrodes straight from vacuum packs, which give low H_2 levels as long as they are used in accordance with the manufacturer's instructions.

- Rutile electrodes. Rutile electrodes contain titanium oxide in the flux coating. As with basic electrodes the shielding gas produced is mainly CO_2. The rutile coating makes this electrode very welder-friendly due to its ease of use, low fume levels, low spatter levels and smooth weld beads. They are not baked or preheated before use but may be heated for a short period at temperatures up to 120 °C to ensure they are dry before use.
- Cellulosic electrodes. These are electrodes that contain cellulose (which is an organic material) in the flux coating. The shielding gas produced has high levels of H_2, producing a hotter burning arc than CO_2. This hotter arc gives deeper weld penetration and faster welding speeds and is commonly used in the stovepipe welding technique (which entails welding pipes in the vertical down direction).

  One problem is that high levels of H_2 are introduced into the weld from the shielding gas, leading to an increased risk of H_2 cracking. To reduce this risk the welding procedures specify that timed hot passes are required; i.e. applying further passes over the root pass while the weld is still hot. This acts as a 'H_2 soak' technique and allows the H_2 to remain as atomic H and dissipate out of the weldment.

Applications

Common uses of MMA are welding of pipelines, nozzles and nodes and medium to heavy fabrications. MMA is commonly used outdoors owing to the good gas shielding supplied from the melting flux.

Typical defects

Typical defects include slag inclusions, porosity, undercut, lack of fusion and profile defects.

Metal inert gas (MIG)/metal active gas (MAG)/gas metal arc welding (GMAW)

Process description

Figure 5.4 shows the equipment. An electric arc is struck between a continuously fed consumable solid electrode wire and the workpiece. The arc is protected by a shielding gas, which can be either inert or active, depending on the material being welded (Fig. 5.5). An inert gas such as argon or helium does not affect the weld pool properties but an active gas such as CO_2 does have an effect. MIG is known as a semi-automatic process because the welding wire is continually fed

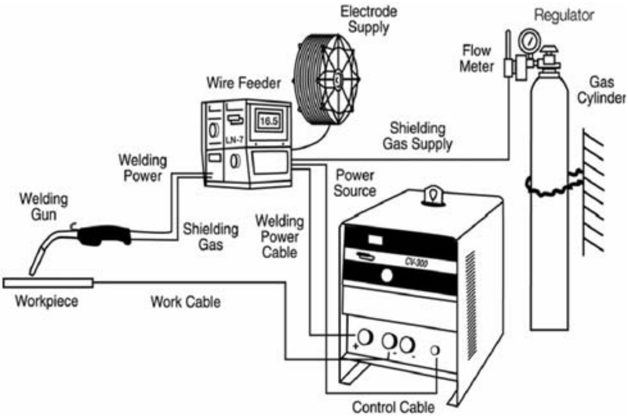
Electrode Supply
Regulator
Flow Meter
Wire Feeder
Gas Cylinder
16.5
LN-7
Shielding Gas Supply
Welding Power
Power Source
Welding Gun
Shielding Gas
Welding Power Cable
CV-300
Workpiece
Work Cable
+
-
Control Cable

Figure 5.4 MIG welding equipment

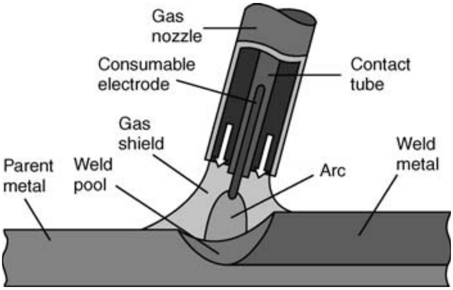

Figure 5.5 MIG welding process

from a reel by the machine but with the travel speed controlled by the welder.

Polarities

MIG/MAG almost always uses a d.c. power source with the polarity electrode positive. The power source has what is termed a ‘flat’ or constant voltage characteristic (Fig. 5.6). This means that a change of arc length (which controls voltage) will have a large effect on the welding current as follows:

- As arc length increases, voltage increases and current decreases. The current is controlled by wire feed speed and affects the burn-off rate of the wire so the wire will burn off slower and extend out back to its original length.
- As arc length decreases, voltage decreases and current increases. This causes the wire to burn off more quickly until it burns back to its original length.

This is referred to as the ‘self-adjusting’ arc (because the arc length is adjusted by the machine and not the welder).

Modes of metal transfer

The MIG/MAG process has varying modes of transferring the filler metal across the arc, dependent on what wire feed speed (current), voltage and shielding gas are being used. The

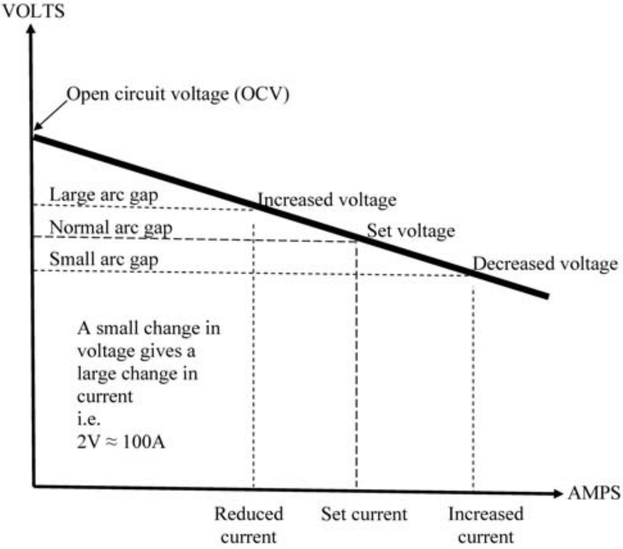

Figure 5.6 Constant voltage characteristic

main modes are short-circuiting transfer, globular transfer and spray transfer.

Short-circuiting (dip) transfer

When voltage and current are low the wire feed speed exceeds the burn-off rate of the wire. The wire ‘dips’ into the weld pool causing the arc to extinguish and short circuiting takes place. This short circuit increases the current in the wire and the end of the wire becomes molten. A magnetic effect takes place causing the wire to ‘neck’ and fall off into the weld pool as a molten droplet. The arc then re-establishes and the whole procedure starts again.

In this mode the welding current must be high enough to prevent the wire sticking and the voltage must be high enough to re-establish the arc. Because this mode of transfer has a low heat input it is best suited to the welding of thin materials and for all positional welding due to the small weld pool formed. The downside to this is that lack of fusion can occur in thick section materials.

Globular transfer

Globular transfer takes place between short circuiting and spray transfer modes at medium current and voltage levels. The molten droplets are larger than the wire diameter and some intermediate short circuiting can take place, leading to the arc being unstable and producing high spatter levels. This mode is rarely used except for some filling passes in the flat position.

Spray transfer

Spray transfer takes place with higher currents and voltages. As the current increases there is an increased flow of droplets across the arc and the diameter of the droplets become smaller. The transfer therefore takes place in the form of a fine spray, giving a high deposition rate coupled with deep penetration and a large weld pool. This can lead to difficulty using spray transfer with a thin sheet owing to the risk of burn-through. The large weld pool is also too difficult to control and maintain during all positional welding so it is mainly used with thick sections in the flat or horizontal–vertical positions only. Aluminium can be welded in all positions in the spray mode because the weld pool solidifies quickly, maintaining a smaller more manageable pool.

Pulse transfer

The ‘all positional’ thickness limitation of the spray transfer mode can be overcome by pulsing the arc to reduce the overall heat input to the work and allow the weld pool to shrink before it gets too large and collapses. This is achieved by regulating the current and voltage to operate in the spray mode for a set period of time, but then immediately reducing them to a level that just keeps the wire tip molten for an equivalent time. An example of this would be to operate on the spray mode for one second, giving deep penetration, but then reduce the amps/volts for one second to allow the weld pool to reduce in size before increasing back up to the spray mode, and so on. In this way the likelihood of getting a lack of fusion-type defects found with the short-circuiting mode is reduced.

Consumables

The only consumables used in a MIG/MAG process are solid wires between 0.6 and 2.4 mm and gases consisting of argon, helium, argon/helium mixtures, CO_2, Ar/CO_2 mixtures, Ar/O_2 mixtures or other proprietary mixtures. It is worth noting the following points in relation to gases:

- Pure CO_2 can be used with steels $\leqslant 0.4\%$ C and low alloy steels using triple deoxidised wire, but it is not usually used in the spray mode.
- Argon produces a better arc in the spray mode and is better with non-ferrous metals and alloys.
- Ar/O_2 (1 or 2%) mixtures are used for stainless steels.
- Helium is normally mixed with argon, oxygen or CO_2. The higher helium contents produce higher arc voltages and heat inputs and give deeper penetrating welds with higher welding speeds.
- Argon/CO_2 (5 to $<20\%$) mixtures are normally used to give a combination of good penetration, a stable arc, less spatter and a flatter weld profile. The lower 5% CO_2 is used in the spray mode and the higher 20% CO_2 is used in the short-circuiting mode. The higher CO_2 level is required to give better penetration in what is a low heat input transfer mode.

Applications

MIG/MAG is commonly used for the welding of structural steels, aluminium alloys and stainless steels. It combines good weld properties with fast deposition rates in light, medium and heavy fabrications.

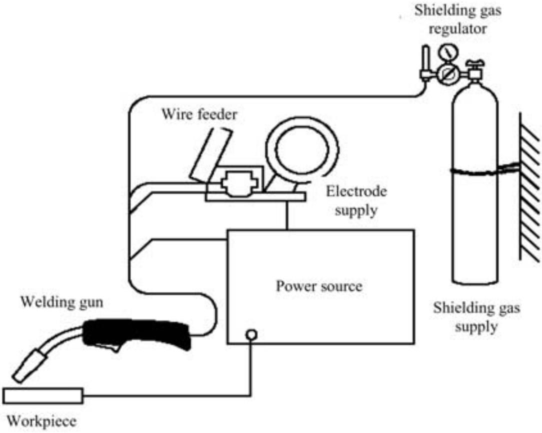
Shielding gas regulator
Wire feeder
Electrode supply
Power source
Welding gun
Shielding gas supply
Workpiece

Secondary gas shielded FCAW

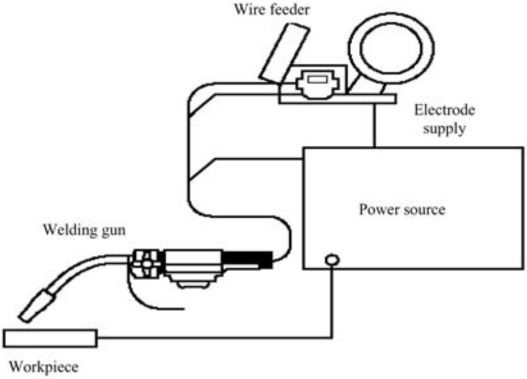
Wire feeder
Electrode supply
Power source
Welding gun
Workpiece

Self shielded FCAW

Figure 5.7 FCAW equipment

Typical defects

Porosity, lack of fusion defects (particularly in the short-circuiting mode), solidification cracking in the spray mode and crater pipes are typical defects.

Flux-cored arc welding (FCAW)

Process description

The equipment is similar to MIG/MAG but uses a flux-cored wire (Fig. 5.7). An arc is formed between a continuously fed tubular electrode wire containing a flux and the work. The arc is protected by a gaseous shroud formed by the flux melting. An external secondary shielding gas can also be supplied through the torch (Fig. 5.8). Using a flux-cored wire enables the addition of alloying elements and the production of a shielding gas more tolerant to outdoor use than MIG/MAG. This means that the benefits of the MMA process can be combined with the speed of the MIG/MAG process. The downside is that it requires slag removal between runs, a backing material for root runs and suitable equipment to remove the large volumes of fume produced from the self-shielding process.

Polarities

DCEP or DCEN is dependent on the wire being used. The power source has a 'flat' or constant voltage characteristic (Fig. 5.6).

Consumables

Cored wires may be self-shielded or gas shielded. Gases may be CO_2, Ar/CO_2 mixtures or Ar/O_2 mixtures. The Ar/O_2 mixture is often used to replace Ar/CO_2 to keep carbon levels at a minimum when welding stainless steels or high alloy materials.

Applications

This process is used in shipyards, structural applications, and other medium and heavy fabrications where positional welding would not be possible with solid wire welding.

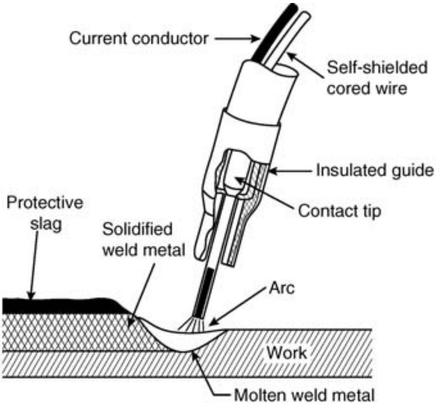
Current conductor
Self-shielded cored wire
Insulated guide
Contact tip
Protective slag
Solidified weld metal
Arc
Work
Molten weld metal

FCAW self shielded

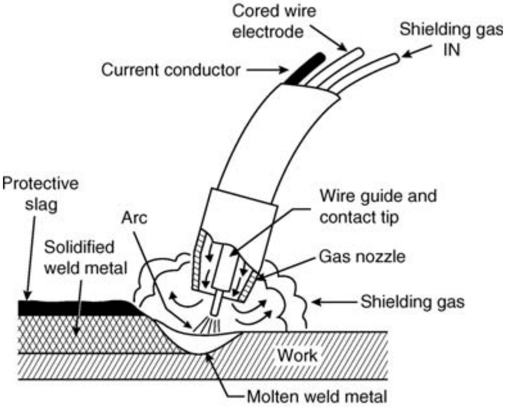
Cored wire electrode
Shielding gas IN
Current conductor
Protective slag
Arc
Wire guide and contact tip
Solidified weld metal
Gas nozzle
Shielding gas
Work
Molten weld metal

FCAW secondary gas shielded

Figure 5.8 FCAW process

Typical defects

Typical defects are slag inclusions, porosity, undercut, spatter, lack of fusion and profile defects.

Tungsten inert gas (TIG)/gas tungsten arc welding (GTAW)

Process description

Figure 5.9 shows the equipment. Fusion is obtained from the heat of an arc formed between a non-consumable tungsten electrode and the workpiece. The arc and weld pool are protected from the atmosphere by a gas shield supplied through the welding torch. Filler metal can be supplied separately into the weld pool (Fig. 5.10). If a joint is welded without the addition of filler it is called an autogenous weld.

Polarities

Steels are TIG welded using direct current with the welding electrode connected to the negative pole (direct current electrode negative, or DCEN). The reason for this is to keep most of the heat in the workpiece and less in the electrode and so prevent the electrode from overheating and possibly melting.

Aluminium alloys and magnesium alloys are welded using alternating current (a.c.). The positive cycle removes the high temperature oxide layer (known as cathodic cleaning) while the negative cycle helps to keep the electrode from melting.

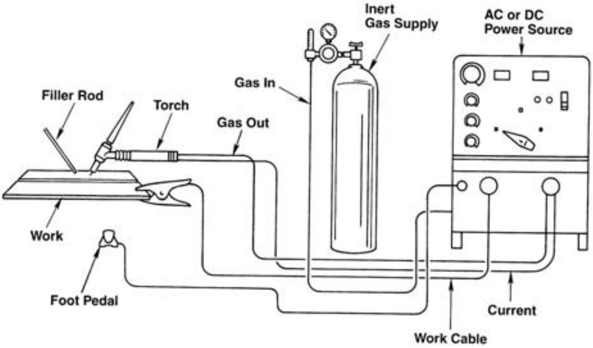

Figure 5.9 TIG welding equipment

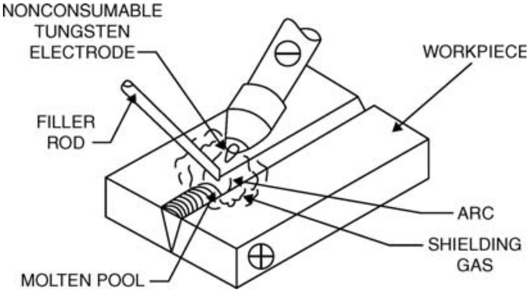

Figure 5.10 TIG welding process

The power source has a ‘drooping’ or constant current characteristic (see Fig. 5.3).

Consumables

The gas consumables used in the TIG process are mainly inert gases such as argon or helium. Active gas mixtures such as 95%Ar/5%H_2 may also be used in certain applications such as welding stainless steels or nickel alloys. Although this is actually tungsten active gas (TAG), it is generally still referred to with the generic TIG term (except by those trying to sound like they are offering something special).

Filler in wire or rod form is used in the majority of cases to fill the joint, but in some applications a fusible insert (often referred to as an EB after the Electric Boat Company who first supplied them) may be used. This is actually pre-placed filler used for the root run in pipe butt welds, but tends to be used in more specialised applications such as the nuclear industry.

Although the TIG electrode is non-consumable it is often classed as a process consumable owing to the fact that it becomes slowly ‘consumed’ over time as it is cleaned and shaped by grinding.

Applications

Because manual TIG is such a slow and expensive process it is generally not used on thick materials where high deposition rates are required. On thick materials it is commonly used for the initial root run(s) because the welder has such good control of the weld pool and can therefore achieve a better quality weld than with most other manual processes. Common uses are high quality welds in the aerospace industry, critical root welds in pipe and general light fabrications.

Typical defects

Typical defects are tungsten inclusions caused by touching the electrode into the weld pool during welding, porosity from the loss of gas shielding or surface contamination, oxidation from insufficient purging gas, root concavity from excess purging gas and crater pipes from breaking the arc too quickly.

Submerged arc welding (SAW)

Process description

Figure 5.11 shows the equipment used for SAW. An electric arc is struck between a continuously fed consumable solid electrode wire and the workpiece. The arc is formed and protected within a blanket of flux, which is partially consumed within the process. The flux is supplied from a hopper attached to the weld head and fed through a tube to form a continuous layer in front of the torch deep enough to contain the arc (Fig. 5.12). The weld metal is formed from a combination of the base metal, filler metal and flux constituents, and will therefore be affected by changes to currents or voltages. The weld metal can be cleaned of contaminants by flux additions and then protected by the slag that forms at the top of the weld. This slag must be removed after each weld run before the next pass is added to prevent slag inclusions within the completed weld. Any unused flux can be collected, mixed with new flux and re-used, provided it is not contaminated. Because a flux layer is used, SAW is normally restricted to the flat or horizontal–vertical positions, although some very specialised equipment

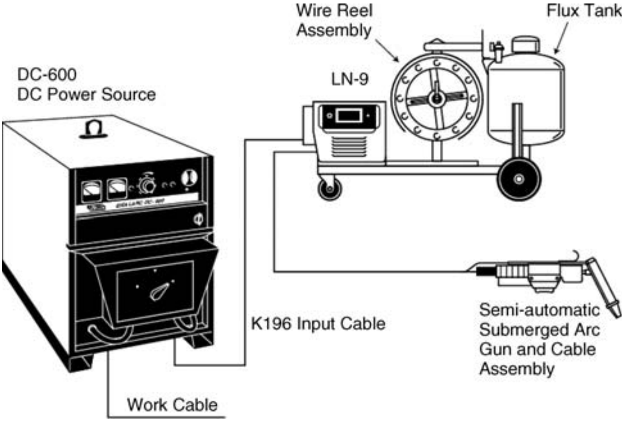
Wire Reel Assembly
Flux Tank
DC-600
DC Power Source
LN-9
K196 Input Cable
Semi-automatic Submerged Arc Gun and Cable Assembly
Work Cable

Figure 5.11 SAW equipment

does enable some other positions to be used. Root control is critical and some form of root backing is required.

Polarities

Polarities typically used are as follows.

- Alternating current is used for multihead systems using high currents to prevent magnetic arc-blow problems. Arc-blow is where the magnetic field surrounding the electric arc is affected by another magnetic field causing the arc to deflect. This is not a problem with a.c. because the magnetic field changes direction with each cycle.
- DCEP is used for welding applications to give deep penetration.
- DCEN is used for surfacing or cladding applications (DCEN gives shallow penetration).
- Multihead systems of two or more heads often use combinations of polarities to increase penetration and deposition rates without causing arc-blow. It is common for the leading head to be d.c. while the trailing head(s) are on a.c. The power source normally has a ‘flat’ or constant

voltage characteristic (Fig. 5.6) but can have a ‘drooping’ or constant current characteristic (see Fig. 5.3) in certain applications.

Consumables

Consumables consist of reel-mounted bare wire or flux-cored electrodes and granular fluxes.

Fluxes

The main types of SAW fluxes are agglomerated or fused.

Agglomerated fluxes are of light colour, globular and of irregular size. They are easily crushed and feel dry and smooth to the touch. Deoxidisers and additional alloying elements are added during manufacture. They are also hygroscopic

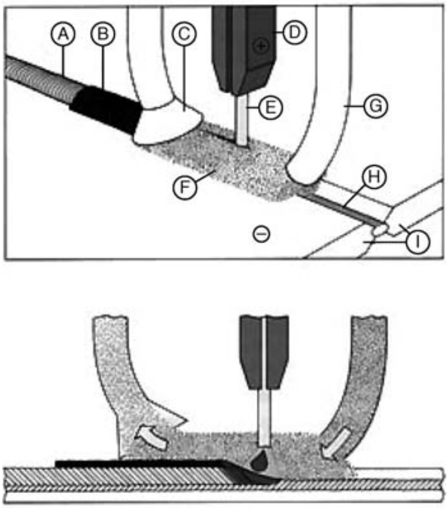

Schematic representation of submerged-arc welding

A Finished weld
B Slag
C Powder removal
D Electrode holder
E Filler wire

F Flux
G Flux supply
H Root bead
I Parent metal

Figure 5.12 SAW process

(absorb moisture) and therefore must be kept dry and preheated before use to remove the moisture. Agglomerated fluxes give better mechanical properties to the welded joint than fused fluxes and are used in low hydrogen applications.

Fused fluxes are formed by melting their constituents together at high temperature and then cooling them to form a glass-like flux. The granules are solid and feel like grains of sand when rolled between the fingers. They are moisture resistant, easy to use, give good weld profiles and have good slag detachability, but the weld metal properties are not as good as those using agglomerated fluxes. These are ideally used in general purpose applications.

Fluxes are also classified as basic or acidic, which refers to the ratio of basic oxides to acidic oxides that they contain. The higher the basicity then the greater the flux moisture absorption and related difficulty in removing it. Agglomerated fluxes will be basic whereas the fused fluxes will tend to be acidic. An increase in basicity will increase toughness but reduce arc stability, reduce weld profile and make slag removal more difficult. The flux with the highest basicity giving good arc stability, acceptable weld profile and slag removal should normally be chosen.

Applications

Because of the deep penetration and high deposition rates achievable with SAW it is commonly used in heavy fabrication environments such as shipbuilding and pressure vessel manufacture.

Typical defects

Typical defects associated with this process are shrinkage cavities formed where high weld depth/width ($> 2{:}3$) ratios are present, solidification cracking caused by high dilution levels with parent material or a high depth/width ratio, porosity due to damp fluxes or insufficient flux depth, hydrogen cracking from damp flux and lack of fusion from arc-blow or an incorrect technique.

Chapter 6

Non-destructive and Destructive Testing

Liquid penetrant testing (PT)

Liquid penetrant examination, often called dye penetrant or penetrant testing (PT), is used to find surface breaking defects only. It involves the use of a cleaner (degreaser), a liquid penetrant and a developer. The most common PT system on site involves the use of these three materials from cans and is referred to as the 'three can system'. Figure 6.1 shows the idea.

A typical colour contrast procedure involves preparing the surface to remove any spatter, slag or other imperfections that could retain the penetrant and mask relevant indications. The surface is then thoroughly cleaned, using the cleaner, to remove any surface oil or grease, which could prevent the red liquid penetrant being drawn into surface breaking cracks or indications. It is then dried using air or lint free cloths.

The penetrant is then applied by spray or brush and left for a dwell time as specified in the procedure. This dwell time must be long enough to enable the penetrant to be drawn into surface breaking defects by capillary action. Excess penetrant is then removed using cloths dampened with the cleaner. The cleaner must not be sprayed directly on to the component otherwise penetrant could be washed out of relevant indications.

Developer (a white chalk like substance) is then lightly sprayed on to the surface causing the red penetrant to be drawn out of any indications by reverse capillary action and the blotting effect of the developer. Any indication highlighted by the red penetrant against the white developer can then be assessed. In effect, PT is a form of enhanced visual

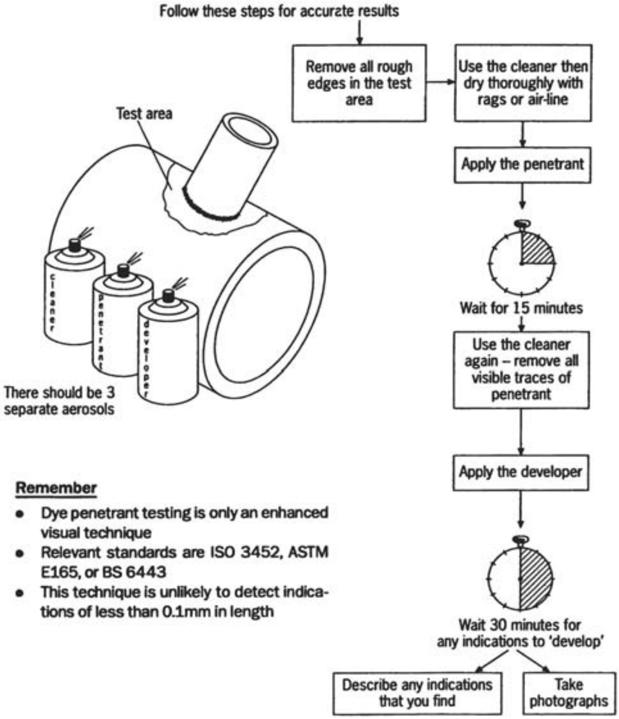
Follow these steps for accurzte results
Remove all rough edges in the test area
Use the cleaner then dry thoroughly with rags or air-line
Apply the penetrant
Wait for 15 minutes
Use the cleaner again – remove all visible traces of penetrant
Apply the developer
Wait 30 minutes for any indications to 'develop'
Describe any indications that you find
Take photographs
Test area
cleaner
penetrant
developer
There should be 3 separate aerosols

Remember

- Dye penetrant testing is only an enhanced visual technique
- Relevant standards are ISO 3452, ASTM E165, or BS 6443
- This technique is unlikely to detect indications of less than 0.1mm in length

Figure 6.1 Liquid penetrant testing

technique that is often used to confirm visual uncertainties. Viewing is done under good lighting conditions of 1000 lux (100 ft candles).

Typical colour contrast penetrants used are solvent removable, post-emulsifiable or water washable. Fluorescent penetrants can also be used in solvent removable, post-emulsifiable or water washable form but are viewed under ultraviolet (UV) light and tend to be more sensitive than the colour contrast methods. Viewing is done in a darkened room with a light intensity of 1000 W/cm^2 on the surface being viewed.

Advantages

- Can test most materials including non-magnetic ones.
- Cheap and simple to use.

Disadvantages

- Can only find surface breaking defects.
- Good surface preparation and cleaning is required.

Magnetic particle testing (MT)

Magnetic particle testing is used to find mainly surface breaking defects in ferromagnetic materials. Sometimes it is possible to find slightly subsurface defects when used with a permanent magnet or d.c. electromagnet. A magnetic flux (or field) is introduced into the material and any defects cutting across the magnetic flux can be detected when ink or powders containing ferromagnetic particles (iron filings) are applied to the material. What happens is that a flux leakage occurs at the defect, which effectively makes the defect a magnet in its own right. This ‘magnet’ attracts the ferromagnetic particles, which take the shape of the defect. Figure 6.2 shows the arrangement.

The magnetic flux can be introduced from:

- a permanent magnet;
- an electromagnet (either a.c. or d.c.);
- electric prods (either a.c. or d.c.) between which a current flows (and the current flow is surrounded by a magnetic field).

Ferromagnetic particles can be applied to the material as:

- a black ink (viewed against a pre-applied white contrast paint);
- fluorescent ink (viewed under UV light conditions);
- red or blue dry powders (used at higher temperatures).

The magnetic flux is applied in two mutually perpendicular directions. The reason for this is that if the flux direction runs parallel to the defect then magnetisation of the defect will not take place and the ferromagnetic particles will not be attracted to it. Testing in two directions ensures that the flux cuts any linear defects by at least 45 degrees. The magnetic field strength can be checked using a checker such as a ‘pie gauge’ or ‘burma castrol strip’.

1. Black magnetic ink and white contrast paint

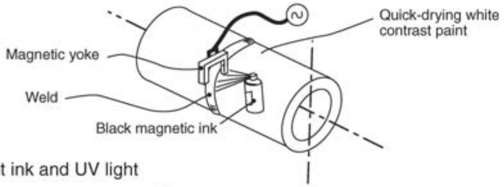
Quick-drying white contrast paint
Magnetic yoke
Weld
Black magnetic ink

2. Fluorescent ink and UV light

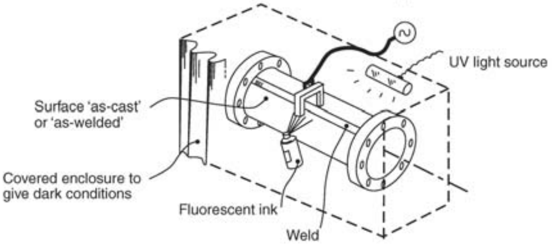
UV light source
Surface ‘as-cast’ or ‘as-welded’
Covered enclosure to give dark conditions
Fluorescent ink
Weld

3. Dry powder (red or blue)

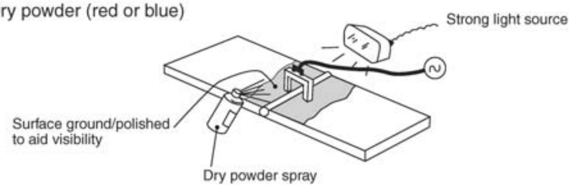
Strong light source
Surface ground/polished to aid visibility
Dry powder spray

Figure 6.2 Magnetic particle testing

Advantages

- Quick and simple to use.
- Can find slightly subsurface defects (under certain conditions).
- Portable when permanent magnets used.

Disadvantages

- Can only be used on ferromagnetic materials.
- Cannot be used on austenitic stainless steels (non-magnetic).
- Electric prods can cause arc strikes.
- The component may need demagnetising on completion of testing.

Ultrasonic testing (UT)

Ultrasonic testing is used to find internal defects within a weld or body of a component being tested. A probe emits a sound wave that is passed through the material. If this sound wave hits a defect then all or part of it gets rebounded back to a receiver in the probe and the size and position of the defect can be plotted on a graph by a skilled operator. Figure 6.3 shows the arrangement.

Angled probes send the wave into a weld at angles suitable for the weld preparation bevel angles used. A zero degree ‘compression’ probe is used first to check for any laminations in the parent material that could deflect angled waves and mask defects in the weld.

The surface of the component must be clean and smooth and a couplant applied to exclude air from between the probe and component. The couplant must be suitable for use on the material being tested and then be thoroughly cleaned off afterwards to prevent any risk of corrosion or degradation of the component in service. Wallpaper paste is often used as a couplant because when it dries it can be easily peeled off.

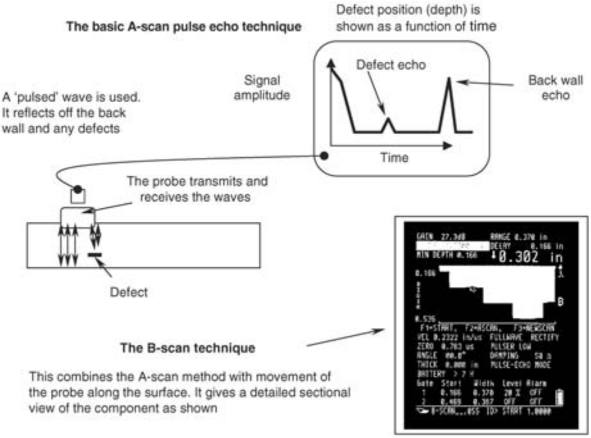
The basic A-scan pulse echo technique

Defect position (depth) is shown as a function of time
Signal amplitude
Defect echo
Back wall echo
Time
A 'pulsed' wave is used. It reflects off the back wall and any defects
The probe transmits and receives the waves
Defect

The B-scan technique

This combines the A-scan method with movement of the probe along the surface. It gives a detailed sectional view of the component as shown

```
GAIN  27.3dB          RANGE 0.370 in
                      DELAY     0.166 in
MIN DEPTH 0.166       ↓0.302 in
0.166
0.536
 F1=START,  F2=ASCAN,   F3=NEWSCAN
VEL 0.2322 in/us  FULLWAVE  RECTIFY
ZERO  0.703 us    PULSER LOW
ANGLE  00.0°      DAMPING    50 Ω
THICK  0.000 in   PULSE-ECHO MODE
BATTERY  > 7 H
Gate  Start   Width  Level Alarm
  1   0.166   0.370  20 %  OFF
  2   0.469   0.307  OFF   OFF
B-SCAN...055  ID> START 1.0000
```

The C-scan technique

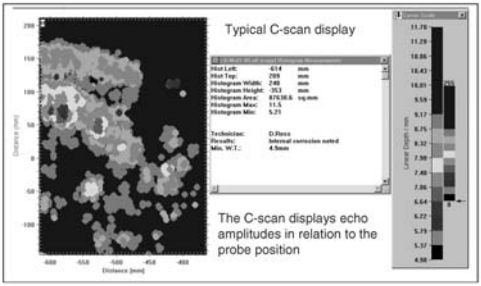
Typical C-scan display

```
Hist Left:            -614      mm
Hist Top:             209       mm
Histogram Width:      240       mm
Histogram Height:     -353      mm
Histogram Area:       87630.6   sq.mm
Histogram Max:        11.5
Histogram Min:        5.21

Technician:           D.Ross
Results:              Internal corrosion noted
Min. W.T.:            4.5mm
```

The C-scan displays echo amplitudes in relation to the probe position

Distance (mm)
200
150
100
50
0
-50
-100
-600
-550
-500
-450
-400
Distance (mm)
Linear Depth / mm
11.70
11.28
10.86
10.43
10.01
9.59
9.17
8.75
8.32
7.90
7.48
7.06
6.64
6.22
5.79
5.37
4.90

Figure 6.3 Ultrasonic testing

Advantages

- Can find linear type defects in most orientations.
- Is portable.
- Is safe and therefore does not require production to be stopped or the area cleared.

Disadvantages

- Surfaces must be smooth and free of all spatter and arc strikes.
- Very dependent on skill of the operator.
- Normally limited to materials above 6 mm thick.
- Basic technique does not give a permanent record.
- Not so good on coarse grained materials.

Radiographic testing (RT)

Radiographic testing is used to detect internal defects in welds. Although it can find planar (two-dimensional) defects such as cracks or lack of fusion it will not find them in all orientations. It will, however, more easily find volumetric defects such as porosity or slag inclusions or shape defects such as undercut or excess root penetration. It can also be used for profile surveys of pipework and components to check for loss of wall thickness caused by corrosion and/or erosion. Figure 6.4 shows the basic technique.

Gamma rays (from a radioactive isotope) or X-rays (from a machine) are passed through the material and strike a film causing it to darken. The film gets darker the more radiation that hits it so volumetric defects such as porosity that allow more radiation through the material will show as areas darker than the surrounding area. Conversely, areas such as excess penetration, where more radiation is absorbed by the material, will show as lighter than the surrounding area. Gamma radiography is commonly used on site because it is portable and does not require a power source. It is inherently dangerous, though, because the radioactive isotope (source) cannot be ‘turned off’ and is always radiating. Storage, transportation and use of the source must therefore be closely controlled to ensure the safety of the radiographers and workplace personnel.

The type of isotope used will depend on the thickness of material to be tested. The most common gamma isotope is iridium 192 but thinner materials may use other isotopes

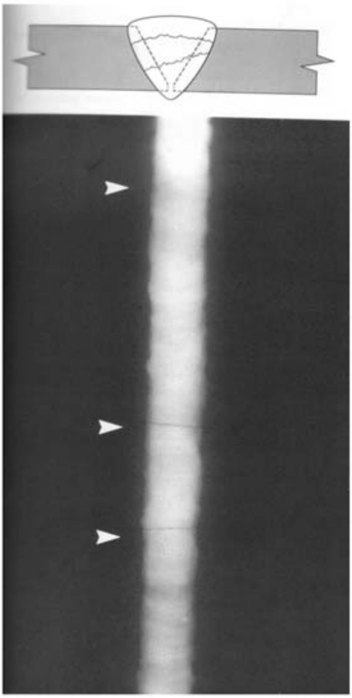

Figure 6.4 Radiographic testing

such as yturbium and (for thicker materials 40 mm +) cobalt 60. Cobalt 60 is incredibly dangerous and should only really be used when iridium 192 is impracticable. It is a good idea to finish having all the children you want before using cobalt 60 or, better still, get someone else to use it and keep well back. On a more serious note – treat it with great respect. The sources are contained in carriers made from depleted uranium and are only wound out for the required exposure time.

Quality of the radiographic film is measured using four main parameters:

1. **Density** is a measure of how much light passes through the film. The higher the density number, the darker the film. Normal acceptance levels are densities of between 1.8 minimum to 4.0 maximum for X-ray and 2.0 to 4.0 for gamma, but it varies between standards. The density measurement is normally taken along the area of interest (i.e. the weld length) using a piece of equipment called a densitometer.
2. **Sensitivity** is an indication of the smallest defect that can be seen on the image. A wire type or hole type image quality indicator (IQI) is used to determine the smallest defect visible on the image. The wire type IQI is the more common type used and consists of either six or seven wires depending on the standard used (see Fig. 6.5). The American ASTM standard specifies six wires but the European standard specifies seven. Sensitivity is expressed as a percentage derived from the thinnest wire visible divided by the material thickness. The acceptable percentage will vary depending on material thickness. To save calculations ASME V and EN standards give tables specifying the smallest wire that

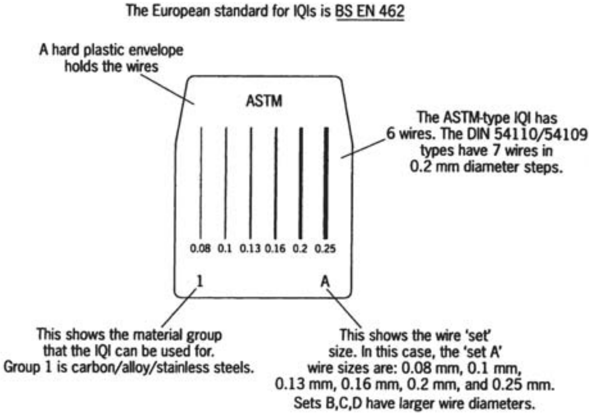
The European standard for IQIs is BS EN 462
A hard plastic envelope holds the wires
ASTM
The ASTM-type IQI has 6 wires. The DIN 54110/54109 types have 7 wires in 0.2 mm diameter steps.
0.08 0.1 0.13 0.16 0.2 0.25
1
A
This shows the material group that the IQI can be used for. Group 1 is carbon/alloy/stainless steels.
This shows the wire 'set' size. In this case, the 'set A' wire sizes are: 0.08 mm, 0.1 mm, 0.13 mm, 0.16 mm, 0.2 mm, and 0.25 mm. Sets B,C,D have larger wire diameters.

Figure 6.5 Image quality indicator (IQI)

must be visible to give the correct sensitivity value for different material thicknesses and different radiographic techniques.

3. **Geometric unsharpness (Ug)**, also known as penumbra, is a measure of the 'fuzziness' of the radiographic image. Geometric unsharpness (Ug) is calculated from the formula

   Ug = Fd/D
   where
   F = effective source size or focal spot
   d = object to film distance
   D = source to object distance

   Acceptance is based on the relevant code requirement. ASME V Section 2 gives recommended maximum limits for Ug ranging from 0.020 in (0.51 mm) for material thicknesses below 2 in to 0.070 in (1.78 mm) for materials greater than 4 in.

4. **Backscatter** is stray radiation that can expose the film. A lead letter 'B' is placed on the back of the film and if excessive backscatter is present a light image 'B' will be visible on the film, which should then be rejected. A darker image, on the other hand, is not a cause for rejection.

Advantages

- Gives a permanent record.
- Surface condition is not so critical.
- Lagging does not require removal.

Disadvantages

- Safety is an issue.
- May require production to be shut down.
- Personnel must be excluded from the vicinity.
- Will not find linear defects in all orientations.

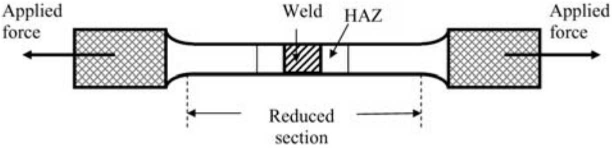

Figure 6.6 Tensile specimen: reduced section transverse test

Tension tests

The main objective of a tension test on a welded joint is to measure the yield and tensile strengths of the sample under test. A pulling force (load) is applied to the sample and the yield and ultimate tensile strength (UTS) are measured and recorded on the test form along with the material type, specimen type, specimen size and location of the fracture. Weldments are normally subjected to a reduced transverse tensile test (see Fig. 6.6).

Manufacturers of welding consumables carry out a longitudinal ‘all weld metal’ tensile test to measure the tensile strength, yield point and percentage elongation (*E*%) of the deposited weld metal. The sample is taken from the centre of the weld and consists of weld metal only.

Bend tests

Bend tests give an indication of weld quality and a rough indication of ductility by putting the weld and HAZ under tension. The weld and HAZ must be included within the bent portion and a limit given to the size of linear openings permitted on the surface under test. Bend tests are normally transverse tests taken across the weld and include the weld, HAZ and base material. Longitudinal tests are more unusual but can be used where dissimilar base materials with widely differing properties are welded or where the weld metal is of greatly differing properties to the base metal. Guided bend tests are those that have the sample bent into a guide (see Fig. 6.7).

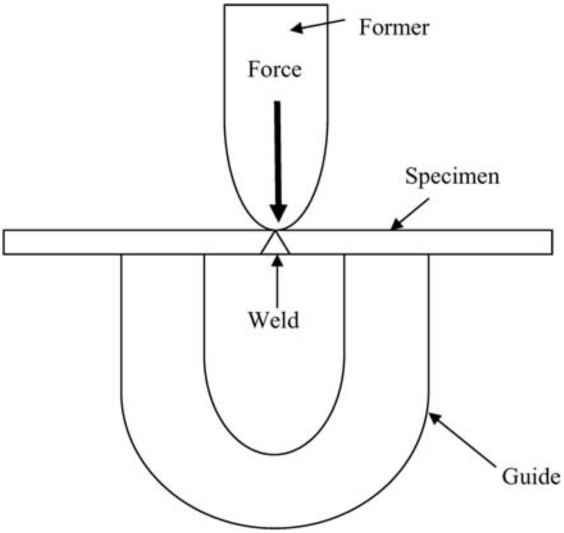

Figure 6.7 Guided bend test

Bend tests are classed as face, root or side bends (see Fig. 6.8) depending on which surface is under tension; a face bend will have the face under tension while a root bend will have the root under tension. A side bend test is used where the sample is too thick to form a face or root bend. With a side bend test the cross-section of the whole weld is put under tension and checked for internal defects such as lack of sidewall fusion or inter-run fusion. It is worth keeping in mind that the side bend is just a snapshot of the weld at one particular point within the weld length.

Charpy tests

Material impact toughness can be measured by various types of test such as the Charpy V-notch impact test, Izod test or K_{IC} test. The most commonly used test is the Charpy impact test (see Fig. 6.9), which gives an indication of the toughness of a material at a specified temperature. It is not a

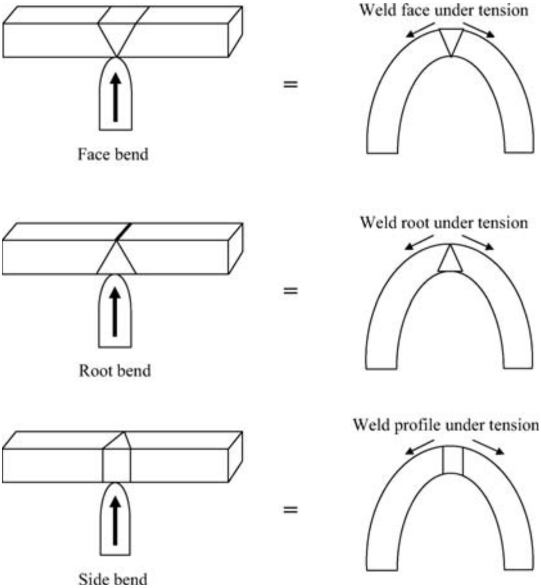

Figure 6.8 Types of bend test

particularly accurate test but can give a general indication of the ability of a material to resist brittle fracture at its minimum design material temperature. The test consists of holding a machined specimen, of a specified size (normally 55 mm × 10 mm × 10 mm) containing an accurately machined notch of specific dimension, at both ends as a simple beam. A pendulum impacts on the specimen and the start and finish heights of the pendulum are measured. The difference in height equates to the energy absorbed by the specimen before it fractures. This absorbed energy is usually measured in joules on a scale attached to the machine.

Three Charpy specimens are tested at each specified temperature and the final result taken as an average of the three. Tests can be done at various temperatures and recorded in a graph to form the Charpy curve (see

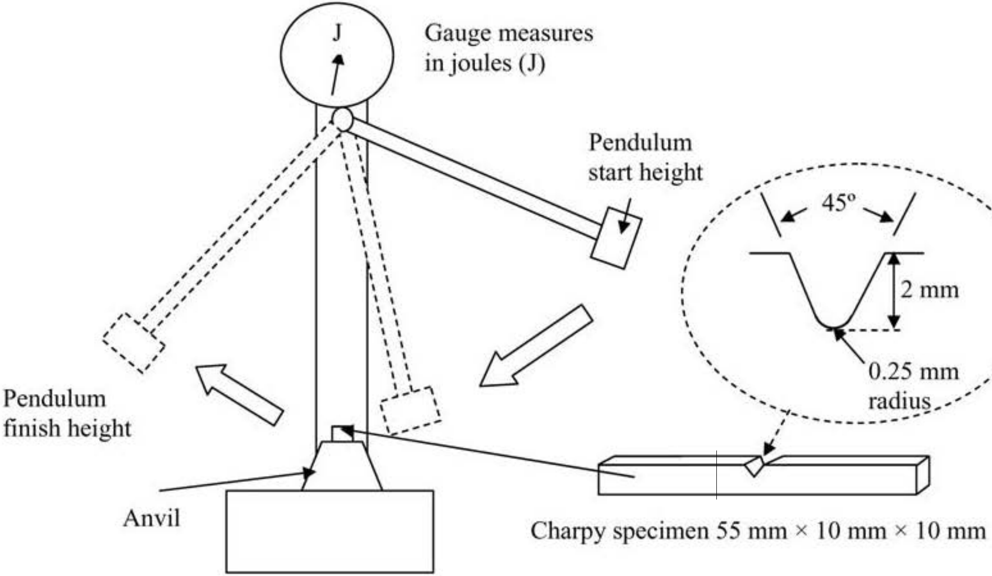
J
Gauge measures in joules (J)
Pendulum start height
45°
2 mm
0.25 mm radius
Pendulum finish height
Anvil
Charpy specimen 55 mm × 10 mm × 10 mm

Figure 6.9 Charpy V-notch impact test

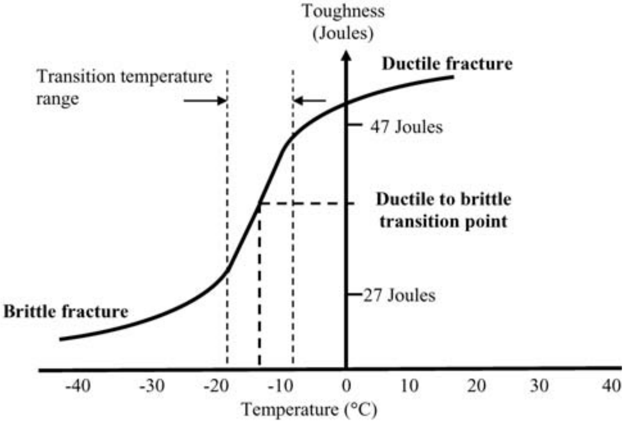

Figure 6.10 A typical Charpy curve

Fig. 6.10) and determine the ductile-to-brittle transition temperature. The ductile-to-brittle transition temperature is the temperature at which the test specimen will start to become more brittle than ductile. You would therefore not want to use the material at design temperatures below this as it would have an increased risk of failing in a brittle manner. Remember that this is not an accurate test reflecting the material behaviour under actual service conditions, so the results should therefore be used with caution.

A more accurate test to check a material's likelihood of failing in a brittle manner is the crack tip open displacement (CTOD) test, sometimes referred to as a K_{IC} test.

Hardness testing

The most common hardness tests are Vickers, Rockwell and Brinell. Hardness is defined as the ability of a material to resist indentation on its surface. Hardness tests consist of impressing a ball (Brinell or Rockwell) or diamond shape (Vickers or Rockwell) into the material under a specified loading and measuring the width of the indentation to give a relative hardness reading (Fig. 6.11). The smaller the width of the indentation the harder the material. Hardness testing normally encompasses the weld and HAZ and is usually done to confirm that PWHT has been carried out correctly. When hardness testing is done for weld procedure qualification purposes it is often done through the weldment thickness as hardness levels can vary considerably through the thickness.

The various types of hardness testing have their own units:

- Vickers test: HV (Vickers hardness);
- Brinell test: HB (Brinell hardness);
- Rockwell: HR (Rockwell hardness).

A Shore Schlerescope is a portable dynamic hardness test using equipment similar in size to a ballpoint pen. It drops a weight from a height on to the test surface and measures the height of the rebound. The higher the rebound the higher the hardness value, which can be read off in any selected unit. It may be used by the welding inspector to gauge hardness values on site, but the accuracy depends on the condition of

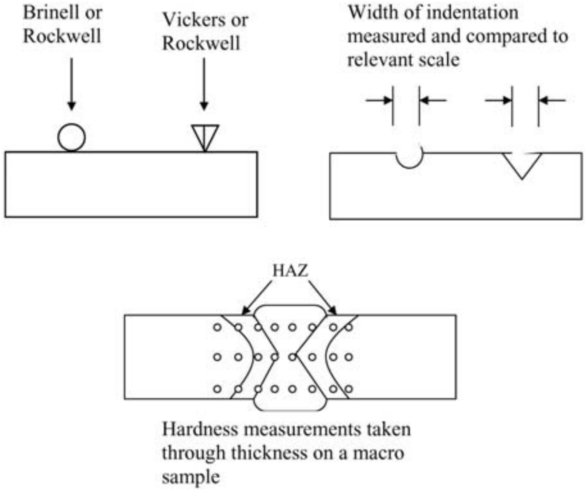
Brinell or Rockwell
Vickers or Rockwell
Width of indentation measured and compared to relevant scale
HAZ
Hardness measurements taken through thickness on a macro sample

Figure 6.11 Hardness tests

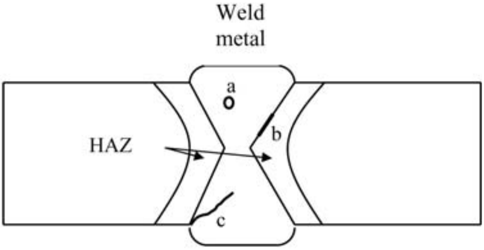

Double V butt macro sample showing
- a. A pore
- b. A lack of sidewall fusion
- c. A crack

Figure 6.12 Macro sample

the test surface and the support of the test piece during the test.

Macro samples

Macro tests (Fig. 6.12) are used to check the internal weld quality of a welded test coupon and so are used in welder performance qualification tests (welder approval tests). A section is removed from the test coupon and then the surface to be inspected is made smooth and polished by lapping to a finish of approximately 600 grit before etching it with ‘nital’ (a mixture of around 10% nitric acid in industrial alcohol). The mixture will depend on the material being tested.

The macro specimen is then inspected at a magnification of ×5–×10 for welder-induced imperfections such as porosity, slag inclusions, cracks, lack of fusion etc. It is common for the macro to be taken at a stop/start location (particularly in the root runs) as this is where defects are most likely to be found. Obviously any stop/starts in root or hot passes would have to be marked on the sample by an inspector before subsequent runs obscured them.

The sample should also be inspected when it is first removed to check for any imperfections which may then be

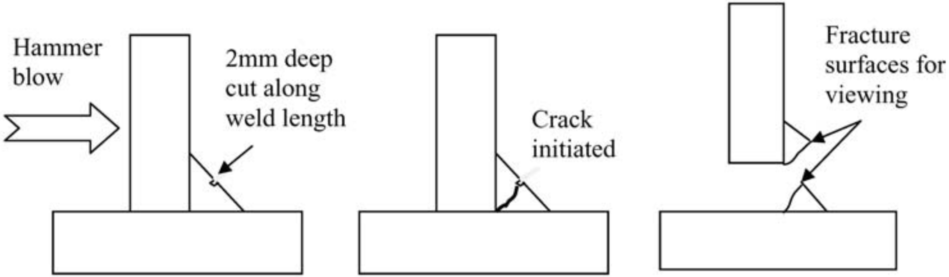
Hammer blow
2mm deep cut along weld length
Crack initiated
Fracture surfaces for viewing

Figure 6.13 Fillet weld fracture test

Woodhead Publishing Ltd – A Quick Guide to Welding and Weld Inspection
Data Standards Ltd, Frome, Somerset – 17/9/2009 06QG Welding chap6.3d Page 84 of 87

removed during subsequent preparation. It is common for macro samples to be varnished and mounted in plastic to preserve them and avoid direct handling of the prepared surface.

Micro samples

A micro sample is prepared to permit examination of the grain structure under much higher magnifications than that used for macros. Magnification is normally in excess of ×100 and can be up to ×2000+ using electron microscopes. The sample is prepared in a similar manner to a macro but the surface is finished with diamond paste to give a much more polished finish of around 1–3 μm before etching with a nital solution (see macro preparation). Micro samples are normally used for research purposes or investigations into the cause of defects or failures.

Fillet fracture test

Fillet fracture tests are used for welder approval testing using fillet welded test coupons. A typical test coupon will range from 150 mm (6 in) to 300 mm (12 in) long and will ideally contain a stop/start in the centre. The end 25 mm (1 in) will be removed either side and discarded (although it is a good idea to make one or both cut ends into macros). The central vertical section will be loaded in the direction shown in Fig. 6.13 until the specimen either fractures or bends flat. A common method of applying loading is to strike it sharply with a hammer in the loading direction. Fracture can be assisted by cutting a groove 2 mm deep along the centre of the weld to act as a fracture initiation point. The fracture surfaces can now be inspected for any internal defects such as porosity, slag inclusions or lack of fusion defects. The root can also be inspected to ensure root penetration along the full length has been achieved.

Butt weld fracture test (nick break)

Another welder approval test is the nick break test (Fig. 6.14), which is similar to the fillet fracture test but is used to

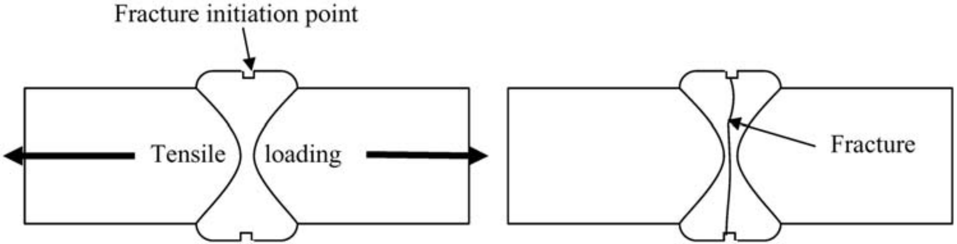
Fracture initiation point
Tensile loading
Fracture

Figure 6.14 Nick break test

assess a butt weld. The weld has a notch cut along the weld length and is then either put under a tensile loading until fracture occurs or is placed in a vice and fractured with a hammer blow. The fracture faces are then inspected for internal imperfections.

Chapter 7

Fracture Modes and Welding Defects

Imperfection categories

Imperfections can be broadly classified as those formed during fabrication and those formed during service. Fabrication imperfections include:

- Cracks
- Porosity and cavities
- Lack of fusion
- Solid inclusions
- Profile imperfections
- Wrong sizes

Service failure modes include:

- Brittle fractures (cracks)
- Fatigue fractures (cracks)
- Stress corrosion cracking (SCC)
- Re-heat cracking
- Creep failure

Cracks

Cracks can be classified by shape (longitudinal, transverse or branched) and position (HAZ, base metal, centreline, crater). An inspector will rarely classify a crack as a particular type (i.e. fatigue, HICC or SCC) because in most cases it will not be possible to determine the precise cause of a crack until an examination is carried out. The shape and position are facts, anything else is supposition.

Hydrogen-induced cold cracking (HICC)

HICC may occur in the HAZ of all hardenable steels (i.e. C, C/Mn) or in the weld metal of high strength low alloy (HSLA) steels that are microalloyed with small amounts of titanium, vanadium or niobium (typically <0.05%). The hydrogen breaks down at increased temperatures into atomic hydrogen (which has a small atomic size) and escapes to the atmosphere through the steel microstructure. When the temperature reduces to below around 300 °C the hydrogen starts reforming to the hydrogen element and will no longer be able to escape from the material. As the H_2 reforms it may build up an internal pressure stress within the material structure itself.

A ductile metal structure can absorb this H_2 without penalty as the metal is able to plastically deform. If, however, the metal is of a hardened (i.e. martensitic) structure less able to deform, then the stress level built up in the material may be enough to cause a fracture to occur. The four critical factors involved in a hydrogen crack are:

- A hydrogen content of > 15 ml/100 g weld metal. The hydrogen comes from moisture, paint, oil, grease, damp electrodes or fluxes, loss of shielding gases or cellulosic electrodes (H_2 is the shield gas).
- A stress level $>$ 50% yield. The stress comes from residual welding stress, restraint stress, etc.
- A hardness $>$ 350 Vickers. The hardness refers to a crack-sensitive microstructure and is related to the Cev of the steel and the formation of martensite, a hard structure caused by rapid cooling of steels.
- A temperature $<$ 300 °C. There is nothing that can be done to stop the temperature eventually falling below the critical level on completion of the welding and any required PWHT, so NDT needs to be carried out for up to 72 hours after welding to check for delayed cracking.

All of the four factors must be present at the same time for a crack to occur, so if any factor is reduced below its critical level a crack will be avoided. Each factor can be reduced as follows:

Reduce the hydrogen level by:

- use of correctly heat treated low H_2 electrodes;
- removal of moisture, oil, grease, paint, etc., from materials;
- use of a low H_2 welding process;
- carrying out a hydrogen soak.

Reduce the stress level by:

- removal of restraints;
- ensuring a good joint fit-up;
- carrying out PWHT stress relief.

Reduce the hardness by:

- preheating the joint;
- carrying out PWHT.

Carry out delayed NDT:

- The temperature will eventually fall below the critical level but a crack may not occur until days afterwards, so carry out delayed NDT.

Solidification cracking

Solidification cracking (also known as centreline cracking or hot cracking) is a fracture that occurs in the weld metal of ferritic steels with a high sulphur or phosphorus content or in joints with a large depth/width ratio.

It occurs in ferritic steels with a high sulphur content because during welding the sulphur joins with iron to form iron sulphide (FeS). This FeS has a lower melting point than steel and therefore remains as a liquid as the steel solidifies. The centre of the weld is the last place to cool so this liquid FeS is pushed to the centre of the weld and forms a liquid film on the grain boundaries, causing a lack of adhesion between the grains. The weld metal contraction due to cooling leaves a high tensile stress, which can pull the weld

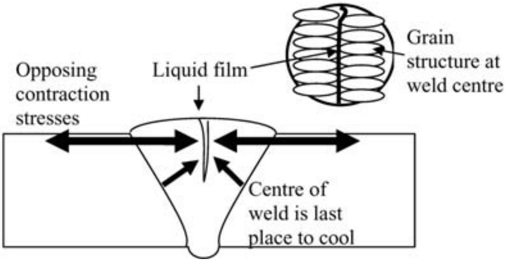

Figure 7.1 Solidification cracking

apart along the weakened centre of the weld bead (see Fig. 7.1).

Solidification cracking can be avoided by the following methods:

Adding manganese (to form manganese sulphide). The manganese combines with the liquid sulphur to form manganese sulphide (MnS), which has a melting point similar to steel. The manganese sulphide structure is spherical and forms between the solidifying weld metal grains without forming a liquid film. This maintains the cohesion between the grains and prevents a crack occurring.

Keep sulphur levels low:

- Specify low sulphur content material with a content <0.03% S.
- Keep surfaces clean of contaminants such as oils, paints and grease.
- Use welding processes that have a low dilution level of parent plate to weld metal (to reduce the levels of sulphur being introduced into the weld metal from the parent plate).
- Ensure temperature indicating crayons used close to welds do not contain sulphur.

Keep stress levels low:

- Reduce restraints to decrease restraint stress and residual stress levels.
- Keep heat inputs low.

Maintain low carbon levels. An increase to the carbon content can rapidly increase the manganese to sulphur ratio required to minimise cracking. Therefore use low carbon fillers and low carbon materials.

Solidification cracking occurs in austenitic stainless steels because the austenitic grain structure is intolerant to contaminants such as sulphur or phosphorus between the grains in the weld metal. A high level of these contaminants can cause the weld to crack.

This cracking can be avoided by making the weld metal about 5% ferrite because the ferrite structure is more accommodating to contaminants. Weld metal can be partially ferrite by the addition of a filler metal chosen using a Schaeffler diagram. It is worth noting that although austenitic stainless steel is non-magnetic the weld metal may be found to be slightly magnetic due to the slight ferrite content.

In welds with a large depth/width ratio (usually in excess of around 2:3) the centre of the weld is the last place to cool and therefore contains large columnar-shaped grains at this point. These grains impinge on each other as they form from both directions within the weld, leading to voids forming in the centre of the weld that cannot be filled with the molten filler being added from above. These voids weaken the weld metal, which is then subjected to severe contraction stresses caused by the shrinkage of the large weld metal volume, producing a centreline crack. Processes with deep penetration or large deposition rates such as SAW are particularly susceptible to this type of failure.

Reheat cracking

Reheat cracking occurs primarily in the HAZ of thick-section high strength low alloy (HSLA) steels, 300 series stainless steels and nickel-based alloys. During PWHT or elevated service temperatures intergranular cracking can occur due to stress relaxation in coarse grained regions under high restraint or residual stresses. The failure normally initiates at a stress concentration such as a notch or change in cross-section. Reheat cracking can be avoided by:

- adequate preheating to reduce the stress levels in the HAZ;
- using joint designs that require less restraint during welding in thick sections;
- removing stress concentrations caused by sharp changes in cross-section, such as sharp undercut, mechanical damage and poorly blended weld toes.

Lamellar tearing

Lamellar tearing occurs mainly in thick-section T-joints and closed corner joints in carbon and carbon manganese steels with a high sulphur content and/or high levels of restraint. It does not occur in cast or forged steels; only in rolled plate. It has the appearance of a 'steplike' crack (Fig. 7.2) and occurs in wrought (rolled) plates due to a combination of:

- contraction stresses from the cooling weld acting through the parent plate thickness plus

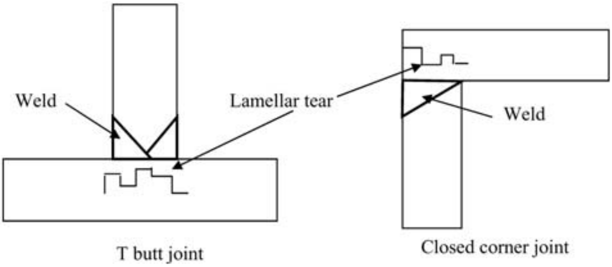

Figure 7.2 Lamellar tearing

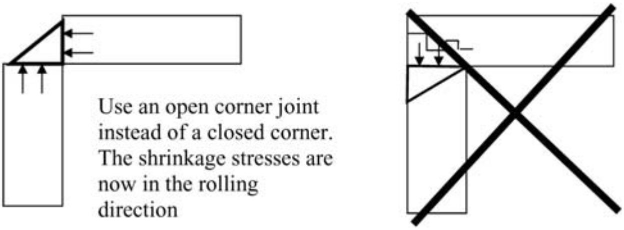
Use an open corner joint instead of a closed corner. The shrinkage stresses are now in the rolling direction

Figure 7.3 Change of joint design

- poor through-thickness ductility due to impurities in the steel (such as sulphides, sulphur, microinclusions and small laminations).

The main ways to avoid lamellar tearing occurring are to:

- *Change the joint design* (Fig. 7.3) to reduce the stress level across the rolling direction (through the thickness). Cracking does not occur when contractional stresses are in the rolling direction. You can relate this to a plank of wood where the grain is running down the length. If you apply a load in the grain direction then the wood is very strong. Apply a load across the grain, however, and the wood fractures more easily.
- *Use a buttering layer* around the joint. A buttering layer (Fig. 7.4) can be applied using a more ductile filler

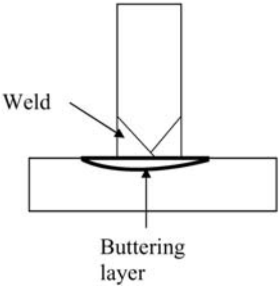
Weld
Buttering layer

Figure 7.4 Weld buttering layer

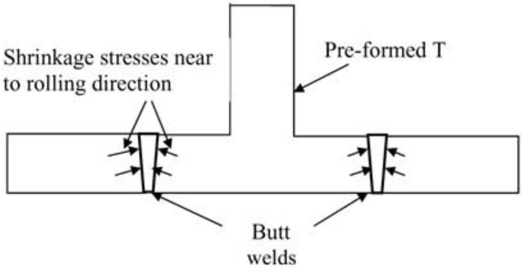

Figure 7.5 Use of pre-formed T joint

material. This allows plastic strain to be absorbed within the buttering layer and so reduce the contractional stress through the parent material thickness.

- *Reduce restraints* to reduce the restraint stress. This can be done by offsetting the parent material instead of using jigs or fixtures that reduce distortion but increase the restraint stress.
- *Use pre-formed T joints* (Fig. 7.5) which are then joined with butt welds. The butt welds can be designed to reduce the through-thickness stress by keeping most of the shrinkage stresses in the rolling direction.
- *Pre-heat* on the base material. Pre-heat will retard the cooling rate, reducing both the shrinkage stress and hardness of the base material. The reduction in hardness will decrease the risk of fracture in the material.
- Use material that has been *STRA tested*. STRA stands for short transverse reduction in area test. This is a tensile test for material that checks for ductility in the through-thickness (short transverse) direction. The greater the ductility, the less chance there is of lamellar tearing occurring. A measure of the ductility is the %*E* (% elongation) or %*A* (% reduction in area) properties of the base material.

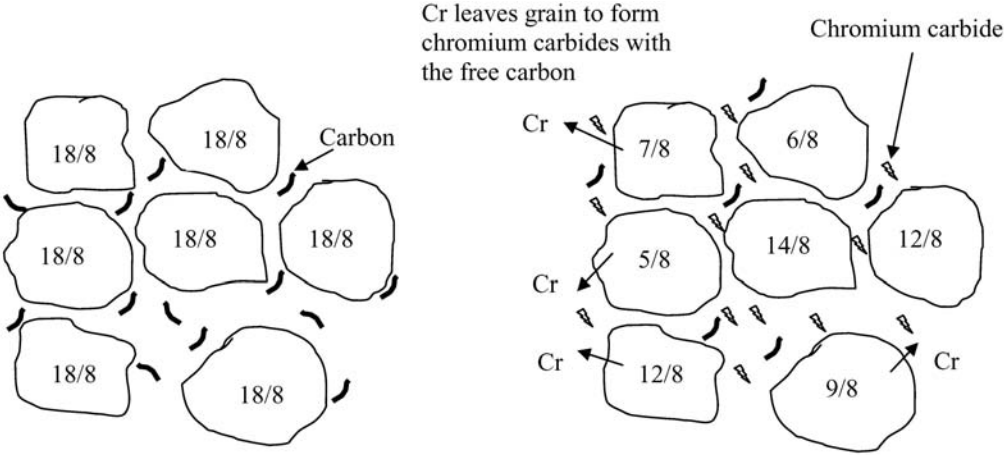
Cr leaves grain to form chromium carbides with the free carbon
Chromium carbide
Carbon
18/8
18/8
18/8
18/8
18/8
18/8
18/8
Cr
7/8
6/8
5/8
14/8
12/8
Cr
Cr
12/8
9/8
Cr

Typical 18/8 austenitic stainless steel structure – 18% Cr 8% Ni

Depleted austenitic stainless steel structure showing reduced Cr in grain

Figure 7.6 Grain depletion

Weld decay

Weld decay is a form of intergranular corrosion that occurs in the HAZ of unstabilised stainless steels. Within the temperature range of around 600—850 °C chromium comes out of solution (Fig. 7.6) to join with free carbon and form chromium carbides. The chromium was in the grain to help prevent corrosion so corrosion can now occur where it has been depleted. Once this chromium depletion occurs the depleted area is said to be *sensitised* (meaning it is susceptible to corrosion) and will corrode in the presence of an electrolyte. The critical region is usually in the HAZ parallel to the weld toes (Fig. 7.7) and once the area is sensitised, corrosion can lead to rapid failure. Weld decay can be avoided by:

- *Using low carbon grade stainless steels*. These steels have less carbon available to form chromium carbides. Reducing the carbon content also reduces the tensile strength, however, so 304L is used instead of 304 and 316L instead of 316. The L indicates low carbon grade stainless steel which may contain about 0.3% carbon rather than 0.8% carbon.
- *Using stabilised stainless steels* instead of unstabilised grades. To prevent the loss of strength associated with using low carbon grades of stainless steels, stabilised

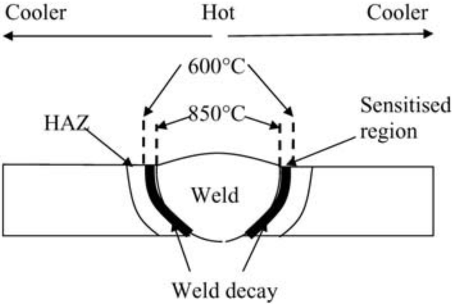

Figure 7.7 Weld decay

stainless steel grades such as 321 or 347 are used. These have stabilisers such as titanium (321) or niobium (347) added. The titanium and niobium are stronger carbide formers than chromium and form titanium carbide and niobium carbide thereby leaving the chromium in the grain.

- *Quench cooling*. Austenitic stainless steel is not hardened by quenching and is not generally susceptible to hydrogen cracking so cooling rapidly to reduce the time in the critical heat range can be done without detrimental effects.
- *Keeping heat inputs and interpass temperatures low*. These actions reduce the time that the material is held in the critical temperature range.
- *Solution heat treatment after welding*. This involves heating to around 1100 °C and quenching, which will dissolve the chromium carbides and restore the chromium to the grain. Do not carry out this procedure without consulting a metallurgist for advice because other problems can be induced in the material if you get the temperatures/timing wrong. Also, be aware that a major disadvantage of this method is the high level of distortion it causes.

Porosity

Porosity (Fig. 7.8) is the entrapment of gases (H_2, O_2, N_2, etc.) within the solidifying weld metal. Causes of porosity include:

- *Loss of gas shield*. Nitrogen and oxygen contamination result from poor gas shielding. As little as 1% air entrainment in the shielding gas will cause porosity. Draughts, leaks in the gas hose or incorrect gas flow rates are frequent causes of porosity.
- *Damp electrodes or fluxes*. Hydrogen can originate from moisture in insufficiently dried electrodes and fluxes.
- *Arc length too large*.
- *Damaged electrode flux*.
- *Moisture or contamination* on the parent material or

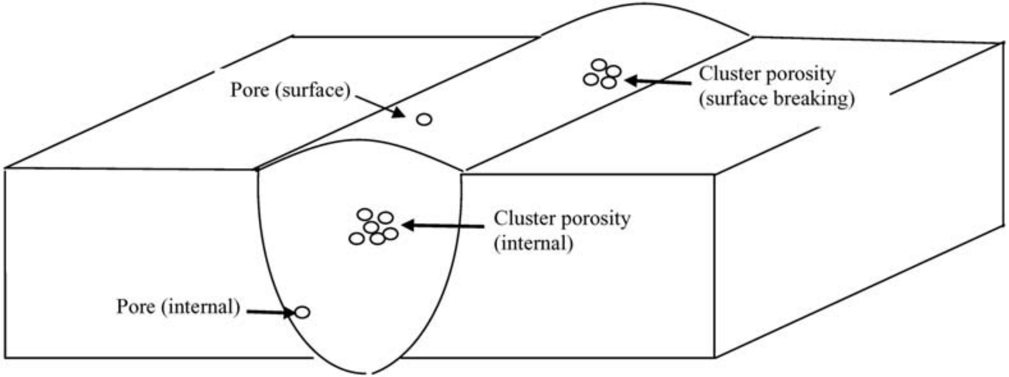
Pore (surface)
Cluster porosity
(surface breaking)
Cluster porosity
(internal)
Pore (internal)

Figure 7.8 Porosity

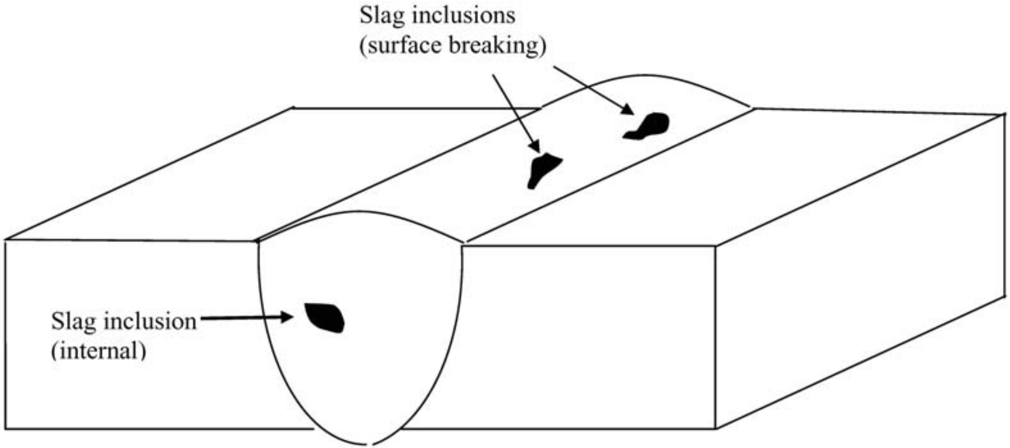
Slag inclusions
(surface breaking)
Slag inclusion
(internal)

Figure 7.9 Solid inclusions

Woodhead Publishing Ltd – A Quick Guide to Welding and Weld Inspection
Data Standards Ltd, Frome, Somerset – 17/9/2009 07QG Welding chap7.3d Page 100 of 107

consumables. Moisture, grease, paint or oil on the material surface or filler wire are common sources of hydrogen leading to porosity.

Solid inclusions

Solid inclusions (Fig. 7.9) can be metallic (i.e. tungsten, copper, etc.) or non-metallic (i.e. slag) and are formed within the weld metal. Causes of solid inclusions include:

- inadequate cleaning of slag originating from the welding flux;
- inadequate removal of silica inclusions in ferritic steels during MAG or TIG welding;
- touching the tungsten to the weld pool during TIG welding;
- the melting of the copper contact tube into the weld pool during MIG/MAG welding.

Lack of fusion

Lack of fusion (Fig. 7.10) is weld metal not correctly fused to the parent material or the previous weld bead. Causes of lack of fusion include:

- incorrect joint preparation (narrow root gap, large root face);
- incorrect welding parameters (current too low);
- poor welder technique (incorrect electrode tilt or slope angles);
- magnetic arc blow;

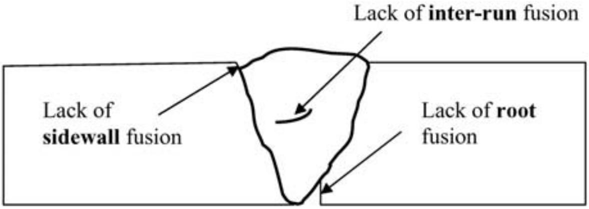

Figure 7.10 Lack of fusion

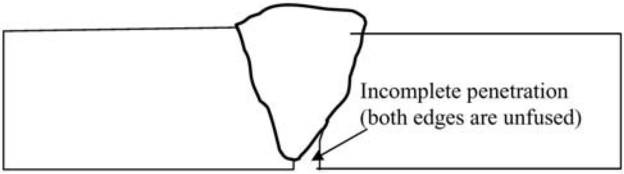

Figure 7.11 Incomplete root penetration

- poor surface cleaning.

Incomplete root penetration

Incomplete root penetration (Fig. 7.11) is where both edges of the root faces are not fused. Causes of incomplete penetration include:

- incorrect joint preparation (narrow root gap, large root face);
- incorrect welding parameters (current too low);
- poor welder technique (incorrect electrode tilt or slope angles);
- magnetic arc blow;
- poor surface cleaning.

Root concavity

Root concavity (Fig. 7.12) is a groove in the root of a butt weld, but with both edges correctly fused. It is sometimes referred to as 'suck back'. Causes of root concavity include:

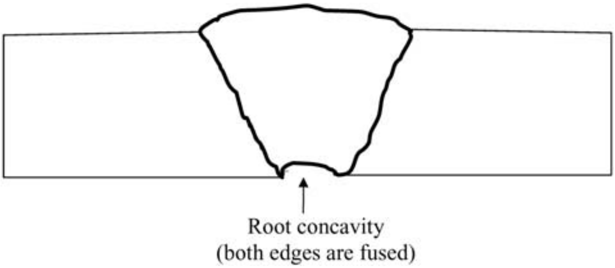

Figure 7.12 Root concavity

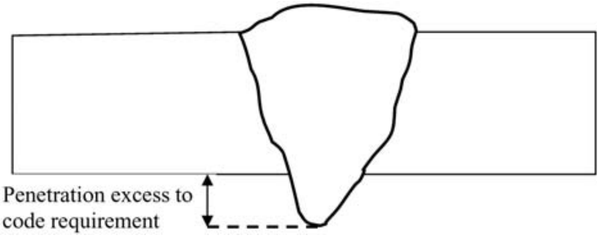

Figure 7.13 Excessive root penetration

- the root face or root gap too large;
- excessive purge pressure being applied when welding using the TIG process;
- excessive root bead grinding before the application of the second weld pass.

Excessive root penetration

A protruding penetration bead is classed as excess penetration because it is excess to requirements and does not contribute to the weld strength. If the level of root penetration is in excess of the design code acceptance criteria it is then classed as *excessive penetration* (Fig. 7.13). Do not confuse excess penetration (which may be acceptable to code requirements) with excessive penetration (which by definition is *not* acceptable to code requirements). Causes of excess/excessive penetration include:

- root faces too small;
- root gaps too large;
- excessive current leading to deeper than expected penetration;
- electrode travel speed too slow.

Overlap

Overlap (Fig. 7.14) is filler metal lying on the surface of the parent metal but not fused to it. Causes of overlap include:

- incorrect travel speed;

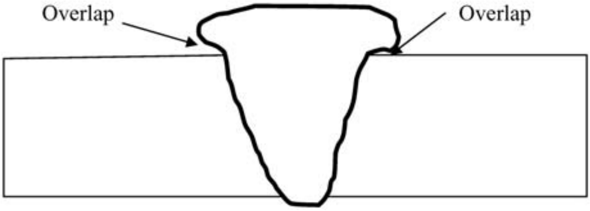

Figure 7.14 Overlap

- incorrect welding technique;
- current too low.

Underfill

Underfill (Fig. 7.15) is the term given to a joint that has not been completely filled to the parent metal surface but the edges of the joint have been fused. Causes of underfill include:

- too small an electrode being used;
- too few weld runs;
- poor welder technique.

Undercut

Undercut (Fig. 7.16) is an unfilled groove left at the toe of the weld following melting of the parent material. It can also be found in previously deposited weld metal. The problem with undercut is that it causes a reduction in material thickness and is a stress concentration from which failures such as

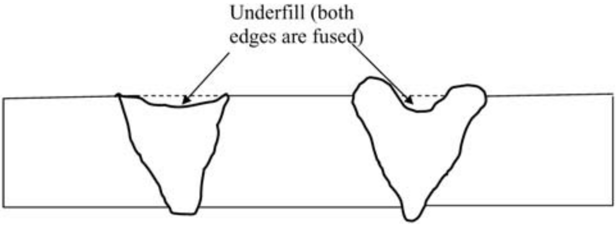

Figure 7.15 Underfill

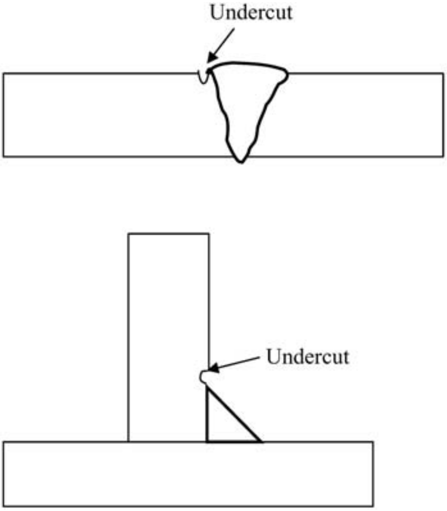

Figure 7.16 Undercut

fatigue fracture can propagate. Undercut is normally assessed by length, depth and profile (sharp or smooth) with acceptance specified by the design code. Causes of undercut include:

- excessive amps/volts;
- excessive travel speed;
- incorrect electrode angle;
- incorrect welding technique;
- electrode too large.

Crater pipe

A crater pipe results from shrinkage of the weld pool on solidification. It is usually caused by switching off a high welding current or breaking the arc resulting in the rapid solidification of a large weld pool. It can be effectively prevented when using the TIG process by progressively reducing the welding current (using a slope-down control) to reduce the weld pool size before breaking the arc and/or by adding filler to compensate for the weld pool shrinkage. Other processes may require run-off plates to be used at the end of weld runs or special techniques to be utilised to reduce the effects of breaking the arc too quickly.

Burn-through

Burn-through is a localised collapse of the weld pool during the root run indicated by excessive root penetration or an irregular cavity in the root bead. It is normally assessed on radiographs by looking at the density. Causes of burn-through include:

- excessive welding current;
- having a small or uneven root face;
- having a large or uneven root gap;
- using too slow a travel speed.

Root oxidation

Root oxidation (sometimes referred to as 'coking') is most common when welding stainless steels and leaves a blackened poorly fused root. It is caused by insufficient back purging gas.

Arc strike

An arc strike (or stray flash) is accidental arcing on to the parent material. This can lead to cracking on crack-sensitive materials due to the fast quenching of the arc strike, causing localised hardened regions. These hardened regions are susceptible to brittle fracture. They can also cause stress concentrations leading to in-service failures such as fatigue fractures. Arc strikes caused by poorly insulated cables or loose earth clamps may introduce copper or other dissimilar materials into the weldment, causing liquation cracking or other contamination problems. Arc strikes on susceptible materials require removal and PT or MT to ensure no cracking is present.

Spatter

Spatter is molten globules of consumable electrode that are ejected from the weld and quench quickly wherever they land on the weldment. They can therefore cause cracking on susceptible materials so they should be removed and then the area tested with PT or MT. Other problems caused by spatter include prevention of UT (because UT needs a smooth surface for the probes), unwanted retention of penetrant during PT and problems with paint retention. Causes of spatter include:

- excessive current;
- damp electrodes;
- surface contamination from oil, paint, moisture or grease;
- incorrect wire feed speed during MAG welding.

Magnetic arc blow

Magnetic arc blow is an uncontrolled deflection of the welding arc due to magnetism. This causes defects such as lack of root fusion or lack of sidewall fusion. Causes of magnetic arc blow include:

- deflection of the arc by the Earth's magnetic field (can occur in pipelines);
- poor position of the current return cable (the magnetic field surrounding the welding arc interacts with the current flow in the material to the current return cable and is sufficient to deflect the arc);
- residual magnetism in the material causing distortion of the magnetic field produced by the arc current.

Some methods of avoiding arc blow are:

- welding towards or away from the clamp;
- using a.c. instead of d.c.;
- demagnetising the steel before welding.

Chapter 8

Codes, Standards and Documentation

Codes and standards

Reference is often made in books (including this one) and quality documentation to 'compliance with the relevant specification, code or standard' but a common question is 'what is the difference (if any) between them?' Unfortunately there is no universally agreed definition of what specifically constitutes a code or standard but the following loose definitions are acceptable.

A **code** is generally:

- A set of rules that must be followed when providing a specific product or service.

A **standard** may:

- refer to standard procedures such as examinations and tests of materials and personnel;
- be a specification for a material or manufactured product and may either be written by companies for internal use or by national and international bodies for public use;
- be a document *referred to* by a code and contain optional or mandatory manufacturing, testing or measurement data. For example, a pressure vessel will be manufactured to a *code* using materials meeting the quality levels contained in a *standard*. In reality it is often not that simple because many standards are actually codes in all but name.

A *code of practice* is normally legally binding and contains all the rules required to design, build and test a specific product. Compliance with the ASME Boiler and Pressure Vessel (BPV) code, for example, is a legal requirement in most states of the USA and pressure vessels manufactured in accordance with Section VIII of the BPV code must comply with all aspects of the code before they can have the ASME stamp applied to them.

Conversely, compliance with the European unfired pressure vessel standard BS EN 13445 is not a legal requirement in the UK (unless it is a contractual requirement) but is the British equivalent to Section VIII of the BPV code (and is therefore realistically also a code).

The term code of practice (or code) is often used by ASME and API, or by individual companies or organisations for their own applications, but the term is not normally used by the British and European standards organisations. The term *application standard* is often used to describe a construction standard in the UK. The best way to look at this is to accept that the construction document may be called a code, an application standard, recommended practice or specification depending upon its country of origin and/or the organisation that released it and the inspector's job is to ensure the required compliance with it.

When all is said and done, a standard or code is a document of best practice that contains the lessons learned over time about the best method of manufacturing a product to meet an acceptable level of quality. Generally the higher the level of quality required then the more stringent the code/standard will be in terms of:

- design;
- the manufacturing method;
- acceptable materials;
- workmanship;
- testing requirements;
- acceptable imperfection levels.

Be aware that codes generally do not contain all the relevant data required for the design, manufacture, testing and inspection but will reference other standards and documents as required. Common construction code/application standards covering vessels, piping and pipelines are as follows:

- ASME VIII: *Rules for Construction of Pressure Vessels*
  One of the major construction codes used throughout the world is the ASME Boiler and Pressure Vessel (BPV) code. It is divided into 12 different sections, but when referring to a section within the ASME BPV code it is normal to just preface the section number with ASME. Therefore, Section V of the ASME BPV code is normally just referred to as ASME V, Section VIII as ASME VIII, Section IX as ASME IX and so on.
- BS EN 13445: *Unfired Pressure Vessels*
  This is the European harmonised standard for the construction of unfired pressure vessels. It superseded BS 5500 *Manufacture of Unfired Pressure Vessels*, which had such good technical information within it that the BSI reissued it as a Published Document (PD 5500) for reference purposes.
- ASME B31.1: *Design and construction rules for power piping*
- ASME B31.3: *Design and construction rules for process piping*
  These are the two most commonly used piping construction codes from the ASME B31 series and are used worldwide.
- BS EN 13480: *Metallic Industrial Piping*
- API 1104: *Welding of Pipelines and Related Facilities*
  This is a very common standard used throughout the world for pipeline construction.
- BS 4515-1:2009: *Specification for Welding of Steel Pipelines on Land and Offshore*
  This is the specification for carbon and carbon manganese steel pipelines.

The PED and UK PER

The European Pressure Equipment Directive (PED) is a directive that specifies the minimum essential safety requirements (ESRs) required for pressure equipment and any pressure equipment to be used within the EU must comply with these requirements by law. The PED is implemented in the UK by means of the UK Pressure Equipment Regulations (PERs).

Do not confuse the PED (a European directive) and PER (a UK regulation) with construction standards/codes. They are totally different entities. Directives and regulations are laws whereas construction codes and standards are normally taken to be 'arbiters of good practice' or guidelines. A manufacturer can choose to build a product following the requirements of a code or standard but the final product must also comply with the law (directive/regulation).

European harmonised standards

European harmonised standards are those that are considered to satisfy the relevant essential safety requirements (ESRs) specified in European product directives such as the Pressure Equipment Directive (PED). Harmonised standards contain an appendix Z, which defines which directives and ESRs the standard meets. The idea of building to a harmonised standard is that it gives a 'presumption of conformity' with any relevant European directive. Products demonstrate their compliance with relevant directives by having a CE mark affixed by the manufacturer. Not every European standard will necessarily be a harmonised standard but it is worth noting that when a European standard is released it replaces the relevant competing national standard from all the countries making up the EU.

If pressure equipment is manufactured to a non-harmonised standard it does not automatically have 'a presumption of conformity' with the PED and will therefore need to show compliance by other means. This usually entails utilising the services of a 'notified body' (NoBo) to prove conformance.

The notified body is an organisation that has been notified to the EU as having sufficient knowledge and experience to be able to ascertain compliance with the directive.

NDE standards

There are many standards dealing with all methods of NDE but it is worth being familiar with the following:

- EN 970: *Non-destructive Examination of Fusion Welds – Visual Examination*
  This European standard covers the visual examination of fusion welds in metallic materials. It is normally, but not always, performed on welds in the as-welded condition.
- EN 1206: *Non-destructive Examination of Welds – General Rules for Metallic Materials*
- ASME V: *Non-destructive Examination*
  This is actually Section V of the ASME BPV code and is also specified by other ASME and API codes for their NDE requirements.

European quality system requirements

- EN 719: *Welding Coordination – Tasks and Responsibilities*
  Welding is a special process that requires the coordination of welding operations in order to establish confidence in welding fabrication and reliable performance in service. The tasks and responsibilities of personnel need to be clearly defined. This standard identifies the quality-related responsibilities and tasks included in the coordination of welding-related activities.
- EN 729: *Quality Requirements for Welding – Fusion Welding of Metallic Materials*
  This standard contains guidelines to describe welding quality requirements suitable for application by manufacturers and is composed of the following parts:
    - ○ Part 1: Guidelines for selection and use
    - ○ Part 2: Comprehensive quality requirements
    - ○ Part 3: Standard quality requirements

  - Part 4: Elementary quality requirements

Welding procedure qualifications

Welding procedure qualification is carried out to prove that a welded joint meets the mechanical, metallurgical and physical properties required by a code or specification. It also enables repeatability by allowing a systematic approach to making the welded joint to be recorded and then used by any qualified welder or welding operator. The welding procedure qualification documents consist of a welding procedure specification (WPS) and a procedure qualification record (PQR). The PQR is also referred to as a welding procedure qualification record (WPQR) in the European procedure qualification standards. For the purposes of clarity this book will use the terms PQR and WPS.

Common procedure qualification standards are:

- ASME IX: *Welding and Brazing Qualifications*
  This specifically details the welding procedure qualification requirements for welding in accordance with the ASME BPV code but is also referenced by other American standard organisations including API. Although other standards will reference ASME IX they may specify additional essential variables or procedure qualification requirements.
- EN 288-2: *Specification and Approval of Welding Procedures for Metallic Materials – Part 2: Welding Procedure Specification for Arc Welding*
  EN 288-2 is the only remaining part from a series of EN 288 standards covering welding procedure qualification. The other parts have been superseded by EN ISO 15607 through to EN ISO 15614, which are standards detailing the specification and qualification of welding procedures for metallic materials with respect to:
  - general rules;
  - guidelines for a metallic materials grouping system;

- welding procedure specification for both arc and gas welding;
- qualification based on tested welding consumables;
- qualification based on previous welding experience;
- qualification by adoption of a standard welding procedure;
- qualification based on pre-production welding test;
- welding procedure test: arc and gas welding of steels and arc welding of nickel and nickel alloys.

Procedure qualification record (PQR)

The PQR is a record of the welding data and variables, which were used to weld a test coupon. It also contains the results of any NDE and/or mechanical tests carried out on the test coupon. Different qualification codes have different requirements but the PQR will contain, as a minimum, the essential variables used in the production of the test coupon. In many cases the non-essential variables will also be recorded, but this is not necessarily a mandatory requirement.

ASME IX is rather unusual in that it separates the toughness requirement of a welded joint from the other mechanical property requirements and assigns it a separate set of essential variables called *supplementary essential variables*, which are only required when impact testing for toughness is carried out and recorded on the PQR.

Figure 8.1 shows an example blank PQR form which contains the requirements recommended by ASME IX.

Note that the main procedure qualification codes (BSEN/ISO) used within Europe have different requirements to the American (ASME/API) codes and they do not fit well together. Procedures may have to be separately qualified to both codes if production welding is specified to comply with both European and American construction codes.

Welding procedure specification (WPS)

A WPS contains all the information required by the welder or welding operator to produce a production weld that meets

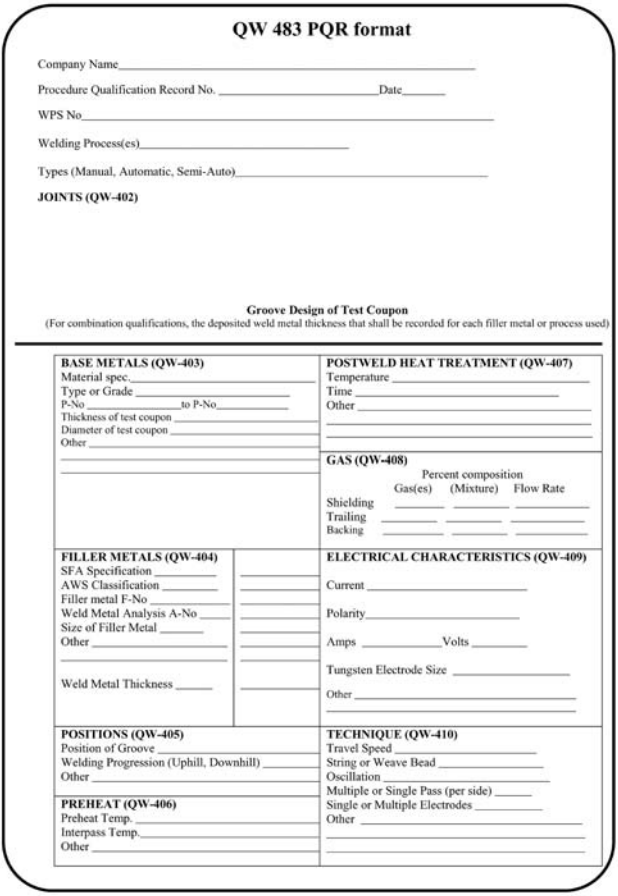

QW 483 PQR format

Company Name______________________________

Procedure Qualification Record No. ______________ Date______

WPS No______________________________

Welding Process(es)______________________

Types (Manual, Automatic, Semi-Auto)______________________

JOINTS (QW-402)

Groove Design of Test Coupon

(For combination qualifications, the deposited weld metal thickness that shall be recorded for each filler metal or process used)

BASE METALS (QW-403)

Material spec.______________

Type or Grade ______________

P-No __________ to P-No__________

Thickness of test coupon ______________

Diameter of test coupon ______________

Other ______________

FILLER METALS (QW-404)

SFA Specification ________ ________

AWS Classification ________ ________

Filler metal F-No ________ ________

Weld Metal Analysis A-No ________ ________

Size of Filler Metal ________ ________

Other ________ ________

Weld Metal Thickness ________ ________

POSITIONS (QW-405)

Position of Groove ______________

Welding Progression (Uphill, Downhill) ______________

Other ______________

PREHEAT (QW-406)

Preheat Temp. ______________

Interpass Temp.______________

Other ______________

POSTWELD HEAT TREATMENT (QW-407)

Temperature ______________

Time ______________

Other ______________

GAS (QW-408)

	Percent composition Gas(es)	(Mixture)	Flow Rate
Shielding			
Trailing			
Backing			

ELECTRICAL CHARACTERISTICS (QW-409)

Current ______________

Polarity______________

Amps __________ Volts ________

Tungsten Electrode Size ______________

Other ______________

TECHNIQUE (QW-410)

Travel Speed ______________

String or Weave Bead ______________

Oscillation ______________

Multiple or Single Pass (per side) ______

Single or Multiple Electrodes __________

Other ______________

Figure 8.1a PQR example (ASME IX)

the required mechanical and metallurgical properties. All the required essential and non-essential variables will be recorded in the WPS forming the instructions to the welder (and also to the inspector). The WPS will contain the

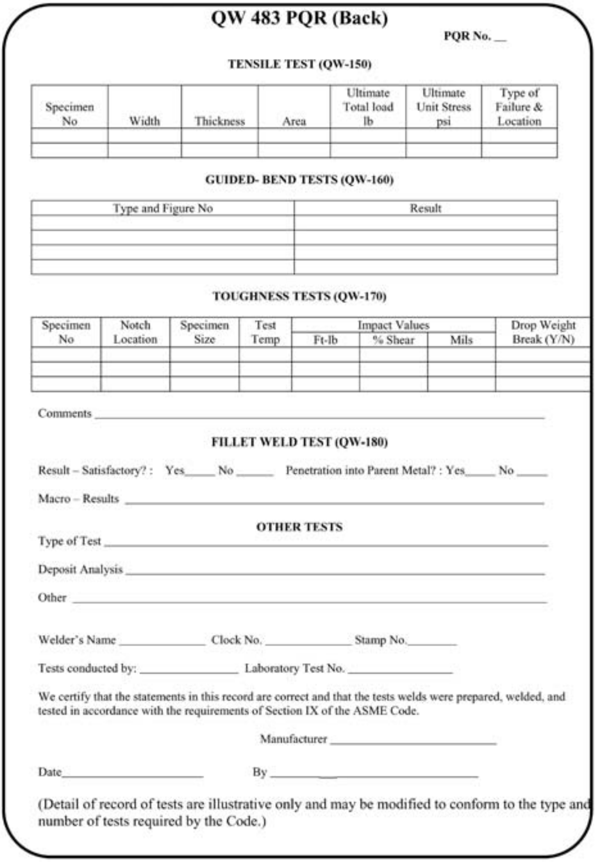

QW 483 PQR (Back)

PQR No. __

TENSILE TEST (QW-150)

Specimen No	Width	Thickness	Area	Ultimate Total load lb	Ultimate Unit Stress psi	Type of Failure & Location

GUIDED- BEND TESTS (QW-160)

Type and Figure No	Result

TOUGHNESS TESTS (QW-170)

Specimen No	Notch Location	Specimen Size	Test Temp	Impact Values			Drop Weight Break (Y/N)
				Ft-lb	% Shear	Mils	

Comments ____

FILLET WELD TEST (QW-180)

Result – Satisfactory? : Yes_____ No ______ Penetration into Parent Metal? : Yes_____ No _____

Macro – Results ____

OTHER TESTS

Type of Test ____

Deposit Analysis ____

Other ____

Welder's Name _____________ Clock No. _____________ Stamp No.________

Tests conducted by: _______________ Laboratory Test No. _______________

We certify that the statements in this record are correct and that the tests welds were prepared, welded, and tested in accordance with the requirements of Section IX of the ASME Code.

Manufacturer ________________________

Date______________________ By ______________________________

(Detail of record of tests are illustrative only and may be modified to conform to the type and number of tests required by the Code.)

Figure 8.1b PQR example (ASME IX)

qualified range of a variable based on the actual value, recorded in the PQR, which was used to weld the test coupon. The qualified range of a variable specified in the WPS is known as the 'range of approval' or 'extent of

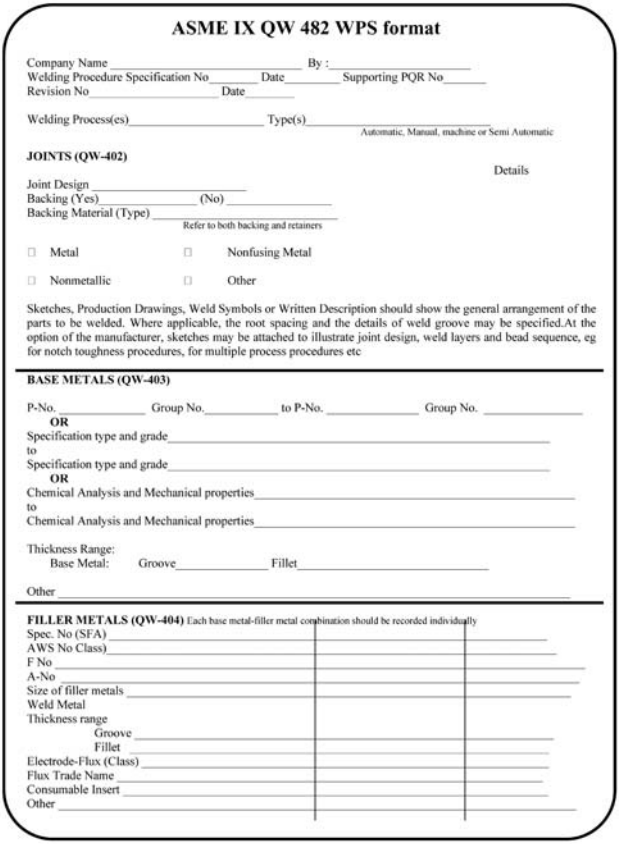

ASME IX QW 482 WPS format

Company Name ______________ By : ______________
Welding Procedure Specification No ________ Date ________ Supporting PQR No ________
Revision No ______________ Date ________

Welding Process(es) ______________ Type(s) ______________
Automatic, Manual, machine or Semi Automatic

JOINTS (QW-402)

Details

Joint Design ______________
Backing (Yes) ________ (No) ________
Backing Material (Type) ______________
Refer to both backing and retainers

☐ Metal ☐ Nonfusing Metal

☐ Nonmetallic ☐ Other

Sketches, Production Drawings, Weld Symbols or Written Description should show the general arrangement of the parts to be welded. Where applicable, the root spacing and the details of weld groove may be specified.At the option of the manufacturer, sketches may be attached to illustrate joint design, weld layers and bead sequence, eg for notch toughness procedures, for multiple process procedures etc

BASE METALS (QW-403)

P-No. ________ Group No. ________ to P-No. ________ Group No. ________
OR
Specification type and grade ______________
to
Specification type and grade ______________
OR
Chemical Analysis and Mechanical properties ______________
to
Chemical Analysis and Mechanical properties ______________

Thickness Range:
Base Metal: Groove ________ Fillet ________

Other ______________

FILLER METALS (QW-404) Each base metal-filler metal combination should be recorded individually
Spec. No (SFA) ______________
AWS No Class) ______________
F No ______________
A-No ______________
Size of filler metals ______________
Weld Metal
Thickness range
Groove ______________
Fillet ______________
Electrode-Flux (Class) ______________
Flux Trade Name ______________
Consumable Insert ______________
Other ______________

Figure 8.2a WPS example (ASME IX)

approval'. Figure 8.2 shows an example blank WPS form that contains the requirements recommended by ASME IX.

Welder qualifications

A welder is qualified by welding a test coupon in accordance with an approved welding procedure. The main purpose of

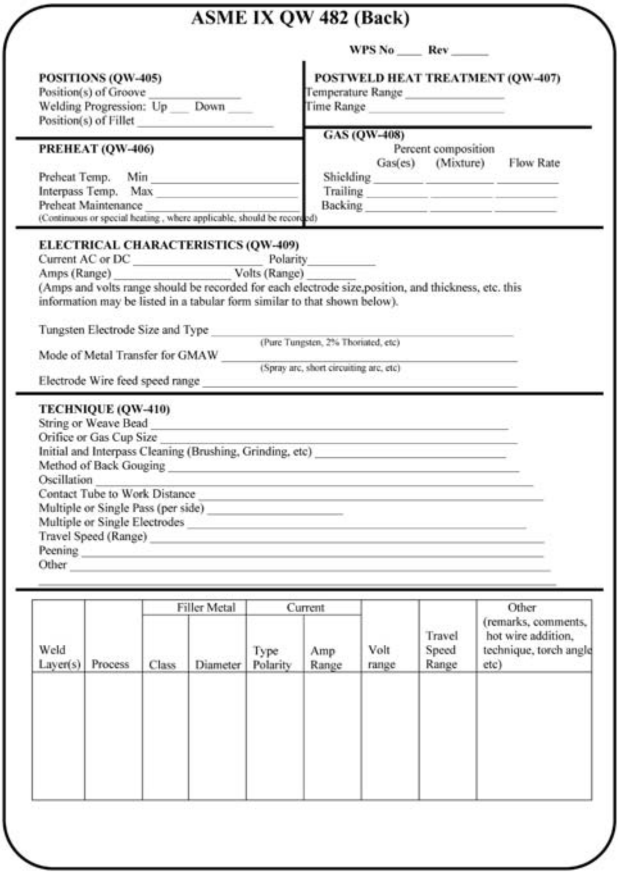

ASME IX QW 482 (Back)

WPS No ____ Rev ______

POSITIONS (QW-405)
Position(s) of Groove ______________
Welding Progression: Up ___ Down ____
Position(s) of Fillet ____________________

POSTWELD HEAT TREATMENT (QW-407)
Temperature Range _______________
Time Range ____________________

PREHEAT (QW-406)
Preheat Temp. Min ______________________
Interpass Temp. Max ______________________
Preheat Maintenance ______________________
(Continuous or special heating , where applicable, should be recorded)

GAS (QW-408)

	Gas(es)	Percent composition (Mixture)	Flow Rate
Shielding	________	__________	_________
Trailing	________	__________	_________
Backing	________	__________	_________

ELECTRICAL CHARACTERISTICS (QW-409)
Current AC or DC ____________________ Polarity__________
Amps (Range) __________________ Volts (Range) ________
(Amps and volts range should be recorded for each electrode size,position, and thickness, etc. this information may be listed in a tabular form similar to that shown below).

Tungsten Electrode Size and Type ____
(Pure Tungsten, 2% Thoriated, etc)
Mode of Metal Transfer for GMAW ____
(Spray arc, short circuiting arc, etc)
Electrode Wire feed speed range ____

TECHNIQUE (QW-410)
String or Weave Bead ____
Orifice or Gas Cup Size ____
Initial and Interpass Cleaning (Brushing, Grinding, etc) ______________________
Method of Back Gouging ____
Oscillation ____
Contact Tube to Work Distance ____
Multiple or Single Pass (per side) ____________________
Multiple or Single Electrodes ____
Travel Speed (Range) ____
Peening ____
Other ____

Weld Layer(s)	Process	Filler Metal Class	Filler Metal Diameter	Current Type Polarity	Current Amp Range	Volt range	Travel Speed Range	Other (remarks, comments, hot wire addition, technique, torch angle etc)

Figure 8.2b WPS example (ASME IX)

this qualification test is to prove that the welder has the knowledge, skill and dexterity necessary to produce a weld that meets the requirements of the relevant standard or specification.

Welder qualification testing is generally the responsibility of the manufacturer or contractor but some standards require the test to be witnessed by an independent third party examiner. One of the duties of a welding inspector is to oversee a welder qualification test being carried out and ensure that it is being done in compliance with the requirements of the relevant code or specification.

A common question asked is whether a welder who welds a test coupon to qualify a new welding procedure is automatically qualified to weld that procedure (even though the coupon has in effect been welded using an unqualified procedure). If not, it means that the welder would have to weld to the procedure once it was qualified to gain the qualification. The answer is very simple; it depends on what the standard says. European standards will normally permit this but others may not.

It is important to be aware of the differences between standards in how welders are qualified. The two main standards are ASME IX and BSEN 287. Many fabricators require their welders to be qualified to both if they are required to comply with the requirements of ASME and the requirements of European directives such as the Pressure Equipment Directive (PED). The ideal situation would be to qualify welders to both standards on the same form using the same test coupon, but unfortunately this is not possible because the qualified thickness ranges, pipe diameters and the non-destructive testing and destructive testing requirements are different. It is therefore better to qualify the welder separately to each standard and apply the relevant qualification in production.

Common welder qualification standards are:

- ASME IX: *Welding and Brazing Qualifications*
  Article III contains welder qualification requirements for welding done in accordance with ASME or API requirements. Requirements may also be referenced within other standards or specifications.
- EN 287-1:2004: *Approval Testing of Welders – Fusion*

Welding of Steels
This is the welder qualification standard used within Europe for fusion welding steels in accordance with European standards.

- EN ISO 9606: *Approval Testing of Welders – Fusion Welding*
  This is the welder qualification standard used within Europe for fusion welding of materials other than steels in accordance with European standards. EN ISO 9606 is divided into Parts 2 to 5 covering aluminium, copper, nickel, titanium, zirconium and their respective alloys. There is no Part 1 because agreement could not be reached for EN 287-1 to be replaced by it.

Figure 8.3 shows the suggested welder qualification form for ASME IX while Fig. 8.4 shows the welder qualification form for EN 287-1. An interesting point to note is that the ASME IX qualification lasts indefinitely as long as the welder is involved in the use of the specified process and the manufacturer signs his qualification history to this effect every six months. This is different to the EN287-1 qualification, which also requires signing every six months but normally only lasts for two years (although it can in some circumstances be extended for a further two years).

Essential variables

Essential variables are those that will affect the mechanical or metallurgical properties of a weldment. If they are changed outside their qualified range a new procedure has to be qualified. The essential variables will be determined by the code that is being used so you do not need to guess what they are. It is also worth bearing in mind that what is an essential variable in one code may very well not be an essential variable in another code so it is important to read the code requirements closely.

The essential variables are recorded in the PQR and it is from these values that the range (or extent) of approval will

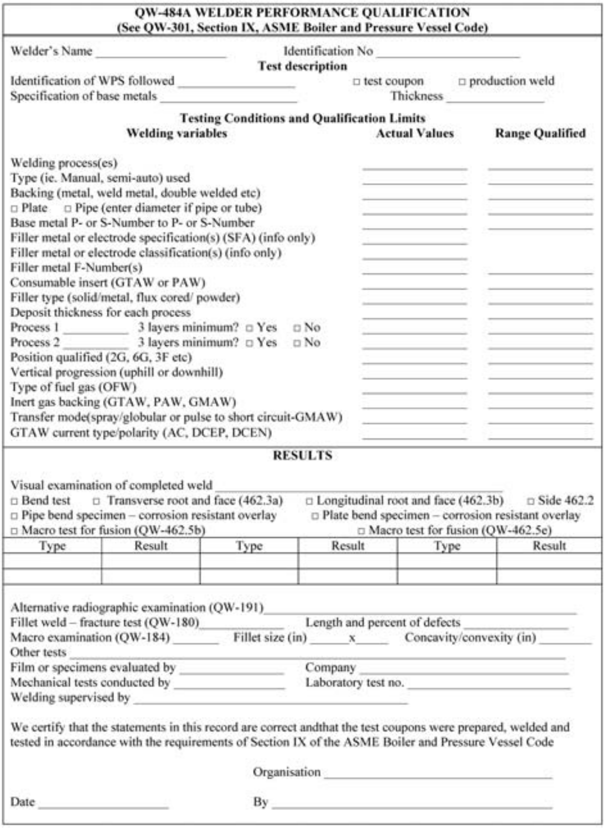

QW-484A WELDER PERFORMANCE QUALIFICATION
(See QW-301, Section IX, ASME Boiler and Pressure Vessel Code)

Welder's Name ____________________ Identification No ____________________

Test description

Identification of WPS followed __________________ ☐ test coupon ☐ production weld
Specification of base metals ____________________ Thickness ______________

Testing Conditions and Qualification Limits

Welding variables	**Actual Values**	**Range Qualified**
Welding process(es)	________	________
Type (ie. Manual, semi-auto) used	________	________
Backing (metal, weld metal, double welded etc)	________	________
☐ Plate ☐ Pipe (enter diameter if pipe or tube)	________	________
Base metal P- or S-Number to P- or S-Number	________	________
Filler metal or electrode specification(s) (SFA) (info only)	________	
Filler metal or electrode classification(s) (info only)	________	
Filler metal F-Number(s)	________	________
Consumable insert (GTAW or PAW)	________	________
Filler type (solid/metal, flux cored/ powder)	________	________
Deposit thickness for each process	________	________
Process 1 __________ 3 layers minimum? ☐ Yes ☐ No	________	________
Process 2 __________ 3 layers minimum? ☐ Yes ☐ No	________	________
Position qualified (2G, 6G, 3F etc)	________	________
Vertical progression (uphill or downhill)	________	________
Type of fuel gas (OFW)	________	________
Inert gas backing (GTAW, PAW, GMAW)	________	________
Transfer mode(spray/globular or pulse to short circuit-GMAW)	________	________
GTAW current type/polarity (AC, DCEP, DCEN)	________	________

RESULTS

Visual examination of completed weld ______________________________
☐ Bend test ☐ Transverse root and face (462.3a) ☐ Longitudinal root and face (462.3b) ☐ Side 462.2
☐ Pipe bend specimen – corrosion resistant overlay ☐ Plate bend specimen – corrosion resistant overlay
☐ Macro test for fusion (QW-462.5b) ☐ Macro test for fusion (QW-462.5e)

Type	Result	Type	Result	Type	Result

Alternative radiographic examination (QW-191)______________________________
Fillet weld – fracture test (QW-180)____________ Length and percent of defects ______________
Macro examination (QW-184) _______ Fillet size (in) ______x______ Concavity/convexity (in) ________
Other tests ____
Film or specimens evaluated by ________________ Company ____________________
Mechanical tests conducted by ________________ Laboratory test no. ____________________
Welding supervised by ______________________________

We certify that the statements in this record are correct andthat the test coupons were prepared, welded and tested in accordance with the requirements of Section IX of the ASME Boiler and Pressure Vessel Code

Organisation ____________________________

Date ________________ By ____________________________

Figure 8.3 WPQ form (ASME IX)

be determined and recorded on the WPS. Again, the ranges will be specified by the code.

Non-essential variables

Non-essential variables are those that do not affect the mechanical or metallurgical properties of a weldment. If changed outside their qualified range a new procedure will

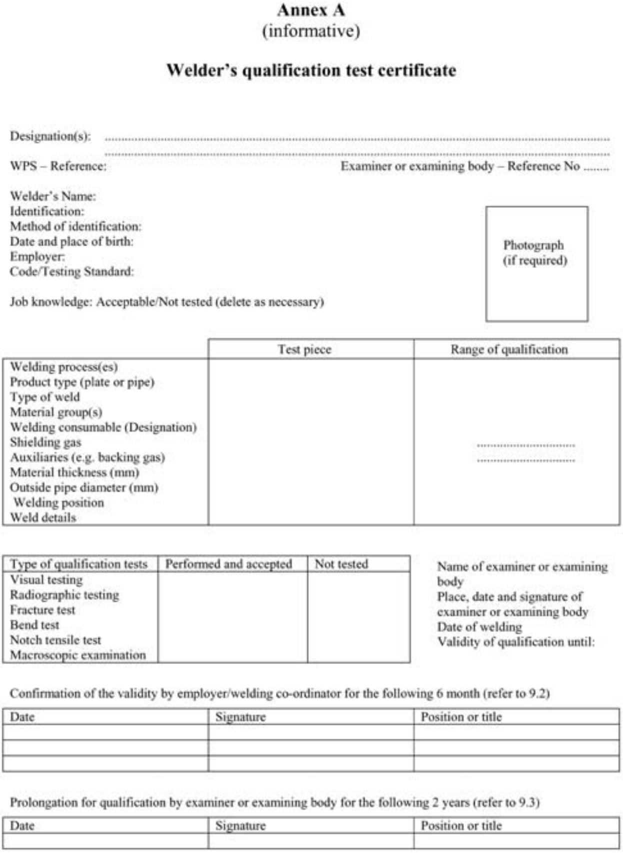

Annex A
(informative)

Welder's qualification test certificate

Designation(s): ....

WPS – Reference: Examiner or examining body – Reference No ....

Welder's Name:
Identification:
Method of identification:
Date and place of birth:
Employer:
Code/Testing Standard:

Photograph
(if required)

Job knowledge: Acceptable/Not tested (delete as necessary)

	Test piece	Range of qualification
Welding process(es)		
Product type (plate or pipe)		
Type of weld		
Material group(s)		
Welding consumable (Designation)		
Shielding gas		....
Auxiliaries (e.g. backing gas)		....
Material thickness (mm)		
Outside pipe diameter (mm)		
Welding position		
Weld details		

Type of qualification tests	Performed and accepted	Not tested
Visual testing		
Radiographic testing		
Fracture test		
Bend test		
Notch tensile test		
Macroscopic examination		

Name of examiner or examining body
Place, date and signature of examiner or examining body
Date of welding
Validity of qualification until:

Confirmation of the validity by employer/welding co-ordinator for the following 6 month (refer to 9.2)

Date	Signature	Position or title

Prolongation for qualification by examiner or examining body for the following 2 years (refer to 9.3)

Date	Signature	Position or title

Figure 8.4 WPQ form (EN 287-1)

not be required but the WPS has to be amended to reflect the new range before production welding takes place. Non-essential variables have to be addressed on the WPS but are not necessarily required on the PQR (again, the code will specify what is required).

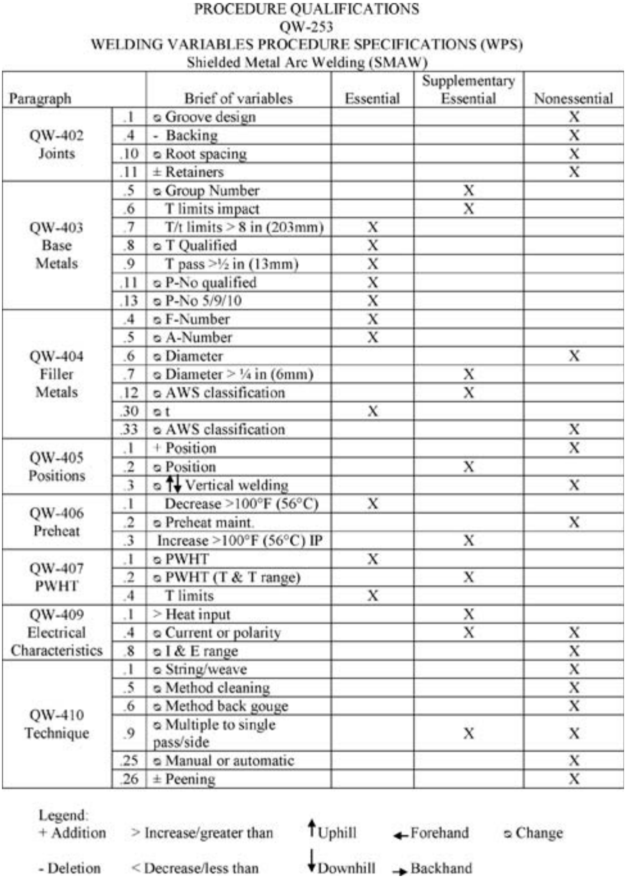

PROCEDURE QUALIFICATIONS
QW-253
WELDING VARIABLES PROCEDURE SPECIFICATIONS (WPS)
Shielded Metal Arc Welding (SMAW)

Paragraph		Brief of variables	Essential	Supplementary Essential	Nonessential
QW-402 Joints	.1	φ Groove design			X
	.4	- Backing			X
	.10	φ Root spacing			X
	.11	± Retainers			X
QW-403 Base Metals	.5	φ Group Number		X	
	.6	T limits impact		X	
	.7	T/t limits > 8 in (203mm)	X		
	.8	φ T Qualified	X		
	.9	T pass >½ in (13mm)	X		
	.11	φ P-No qualified	X		
	.13	φ P-No 5/9/10	X		
QW-404 Filler Metals	.4	φ F-Number	X		
	.5	φ A-Number	X		
	.6	φ Diameter			X
	.7	φ Diameter > ¼ in (6mm)		X	
	.12	φ AWS classification		X	
	.30	φ t	X		
	.33	φ AWS classification			X
QW-405 Positions	.1	+ Position			X
	.2	φ Position		X	
	.3	φ ↑↓ Vertical welding			X
QW-406 Preheat	.1	Decrease >100°F (56°C)	X		
	.2	φ Preheat maint.			X
	.3	Increase >100°F (56°C) IP		X	
QW-407 PWHT	.1	φ PWHT	X		
	.2	φ PWHT (T & T range)		X	
	.4	T limits	X		
QW-409 Electrical Characteristics	.1	> Heat input		X	
	.4	φ Current or polarity		X	X
	.8	φ I & E range			X
QW-410 Technique	.1	φ String/weave			X
	.5	φ Method cleaning			X
	.6	φ Method back gouge			X
	.9	φ Multiple to single pass/side		X	X
	.25	φ Manual or automatic			X
	.26	± Peening			X

Legend:
+ Addition   > Increase/greater than   ↑Uphill   ←Forehand   φ Change
- Deletion   < Decrease/less than   ↓Downhill   →Backhand

Figure 8.5 Brief of variables for SMAW

Supplementary essential variables

These are only found in ASME IX and become essential variables when toughness testing is required as part of the procedure approval. ASME IX contains tables called the 'Brief of Variables' for most of the common fusion welding processes and these tables specify whether a variable is essential, non-essential or supplementary. The 'Brief of Variables' table does not fully explain what the actual variable is but gives direction to the relevant code paragraph that does give a full explanation. Figure 8.5 shows the table for the SMAW welding process.

Chapter 9

Health and Safety

Health and Safety at Work Act

The Health and Safety at Work Act (HASAWA) is a piece of legislation that states that your employer has a duty under the law to ensure, so far as is reasonably practicable, your health, safety and welfare at work. In particular, your employer must:

- Assess the risks to your health and safety.
- Make arrangements for implementing the health and safety measures identified as being necessary by the assessment.
- If there are five or more employees, record the significant findings of the risk assessment and the arrangements for health and safety measures.
- Make sure that the workplace satisfies health, safety and welfare requirements, e.g. for ventilation, temperature, lighting, and sanitary, washing and rest facilities.
- Make sure that work equipment is suitable for its intended use, so far as health and safety is concerned, and that it is properly maintained and used.
- Prevent or adequately control exposure to substances that may damage your health.
- Take precautions against danger from flammable or explosive hazards, electrical equipment, noise and radiation.
- Avoid hazardous manual handling operations, and, where they cannot be avoided, reduce the risk of injury.
- Provide health surveillance as appropriate.
- Provide free any protective clothing or equipment, where risks are not adequately controlled by other means.

As an employee you have legal duties too. They include:

- Taking reasonable care for your own health and safety and that of others who may be affected by what you do or do not do.
- Co-operating with your employer on health and safety.
- Correctly using work items provided by your employer, including personal protective equipment, in accordance with training or instructions.
- Not interfering with or misusing anything provided for your health, safety or welfare.

A welding inspector is not a Health and Safety Officer but must be aware of the dangers associated with welding and cutting operations and comply with their legal duty under site procedures and/or national legislative requirements (i.e. HASAWA) to ensure the safety of themselves and other workers. Some dangers will be common to any workplace environment and some will be specific to the welding/cutting environment. The following sections contain some typical dangers associated with an environment where welding and cutting operations take place.

Electrical safety

When welding, electrical safety practices must be followed including identifying all electrical equipment and having it tested for electrical safety on a planned basis. Equipment must have the correct insulation, be in good condition and be suitably earthed. A low voltage 110 V supply should be used where appropriate for all power tools and the correct fittings and connections must be fitted to the cables and equipment.

A duty cycle is an important consideration and refers to the amount of current that can be safely carried by a conductor in a period of time. The time base is normally ten minutes and a 60% duty cycle means that the conductor can safely carry this current for six minutes in ten and then must rest and cool for four minutes. At a 100% duty cycle the equipment can carry the current continuously. Generally 60% and 100% duty cycles are given for welding equipment.

An example would be a welding machine that can operate at 350 amps at a 60% duty cycle but only 300 amps at a 100% duty cycle.

Do not confuse a duty cycle with a welding process's operating factor. They are often confused because both are given as a percentage. A *process operating factor* is multiplied by its deposition rate in economic calculations to calculate welding costs including process down time. The down time is the time when an arc is not struck and no welding is taking place.

Some typical process operating factors are:

- MMA: 30%;
- MIG/MAG: 60% (hence the possible confusion with a duty cycle);
- MIG/MAG mechanised/automated: 90%.

Remember that the operating factor percentage refers to the percentage of time that an arc is struck.

Welding/cutting process safety

The safety considerations required for cutting and welding operations are there to ensure:

- Suitable ventilation and extraction is in place and operating correctly.
- There are no combustible materials such as oily rags in the area.
- Confined spaces are gas and fume free with gas free certificates in place.
- PPE is worn at all times.
- Permits to work are in place and complied with.
- Oil and grease is kept away from oxygen gauges and fittings (it can ignite).
- Fittings with a high copper content are not used with acetylene gas systems because explosive copper acetylyde can be formed.

Personal protective equipment (PPE)

PPE is defined in the Personal Protective Equipment at Work Regulations (1992) as 'all equipment (including clothing affording protection against the weather) which is intended to be worn or held by a person at work and which protects him against one or more risks to his health or safety'. Hearing protection and respiratory protective equipment are not covered by these regulations because other regulations apply to them, but they do need to be compatible with any other PPE provided. The main requirement of the regulations is that PPE is to be supplied and used at work wherever there are risks to health and safety that cannot be adequately controlled in other ways and that it is:

- Properly assessed before use to ensure it is suitable.
- Maintained and stored properly.
- Provided with instructions on how to use it safely.
- Used correctly by employees.

Arc welding can cause injury in many ways so the correct PPE protection is essential and includes:

- Fire retardant overalls to protect skin from burns caused by the powerful visible and ultraviolet light emitted. They also protect from spatter (globules of burning metal) which is thrown out during welding.
- Gloves to protect hands from burns from hot metal.
- A welding mask with dark filter where the mask itself can afford direct protection from fumes and spatter. The correct grade of filter will protect the eyes from the visible and ultraviolet (UV) light emitted from the welding arc. This UV light can lead to a severe eye irritation called 'arc eye' (which feels like sand has been thrown into your eyes). Any welder who has experienced arc eye will tell you how painful it is and how little they would wish to repeat the experience.
- Boots with steel toecaps to protect from falling objects.

- Goggles/face mask for eye protection when grinding or chipping.
- Gaiters and spats can give protection from burning metal or spatter falling into boots and burning feet. Otherwise, ensure overalls are outside boots.
- A leather apron may be required depending on the process used.
- A welder's hat is especially useful when performing overhead welding.

Gases and fumes

Exposure to the dangerous fumes and gases produced by welding can lead to many health problems, ranging from minor respiratory problems to major respiratory problems (such as no longer breathing at all). Metallic fumes (fumes containing metal particles) and/or gases can come from electrodes, plating or base metals used during the welding cycle. Metallic fumes produced from welding cadmium and zinc are extremely toxic and can result in death, so the importance of removing coatings or plating before welding cannot be overemphasised. The use of the correct extraction or breathing system is essential to maintain a safe welding environment.

Dangerous gases that may be produced during the welding process include ozone, nitrous oxides and phosgene (a form of mustard gas caused by the breakdown of some degreasing agents in arc light). These gases are all extremely poisonous and overexposure can result in death. Other gases used in welding cause problems by displacing air or reducing the oxygen content and so cause death by asphyxiation or suffocation. An example is the use of inert argon gas in a confined space; the argon is heavier than air and will sink to the bottom of the space and reduce the oxygen content as it slowly displaces the air and fills the space from the bottom up. Helium, on the other hand, will fill a space from the top down.

Gases such as oxygen or acetylene can build up in a space and then ignite, causing a flash fire or explosion. Oxygen is particularly dangerous as it has no smell, but will cause a flash fire to occur if an ignition source is brought into an oxygen-rich atmosphere. In particular, if clothing is contaminated with oxygen, it will catch fire easily and burn very fiercely, resulting in severe injury. Even fire-retardant clothing will burn if contaminated with oxygen. Also, oxygen can cause explosions if used with incompatible materials. In particular, oxygen reacts explosively with oil and grease.

Most gases are stored under high pressure and gas cylinders should be stored and used in accordance with recognised practices. All fittings such as pressure regulators, flashback arrestors, flowmeters, gas hoses, etc., should be inspected at regular intervals and maintained in good condition.

COSHH and workplace exposure limits (WEL)

The Control of Substances Hazardous to Health (COSHH) Regulations require employers to prevent exposure to substances hazardous to health, if it is reasonably practicable to do so. If prevention is not reasonably practicable, the employer must adequately control exposure by putting in place measures including, in order of priority, one or more of the following:

- Use processes that minimise the amount of material used or produced, or equipment that totally encloses the process.
- Control exposure at source (e.g. local exhaust ventilation) and reduce the number of employees exposed to a minimum, the level and duration of their exposure and the quantity of hazardous substances used or produced in the workplace.
- Provide personal protective equipment (PPE), e.g. face masks, respirators, protective clothing, but only as a last resort and never as a replacement for other control measures that are required.

Examples of the effects of hazardous substances include:

- Skin irritation or dermatitis as a result of skin contact.
- Asthma as a result of developing allergy to substances used at work.
- Losing consciousness as a result of being overcome by toxic fumes.
- Cancer, which may appear long after the exposure to the chemical that caused it.

The Health and Safety Commission (HSC) in the UK has established workplace exposure limits (WEL) for a number of substances hazardous to health, including fumes or gases that may be produced through welding. These limits are intended to prevent excessive exposure to the substances by keeping exposure below a set limit. A WEL is the maximum concentration of an airborne substance, averaged over a reference period, to which employees may be exposed by inhalation. The WELs are listed in document EH40/2005: *Workplace Exposure Limits*. The reference periods used are eight hours for the long term exposure limit (LTEL) and 15 minutes for the short term exposure limit (STEL). The toxicity of a substance can be gauged by the value of the exposure limit, so applying good practice will mean actual exposures are controlled below the WEL. Advice on applying the principles can be found in the COSHH *Approved Code of Practice (ACOP)*. Be aware that the values given in Guidance Note EH40 can change annually. Table 9.1 contains examples of substances that may be present in welding under certain conditions.

Table 9.1 Workplace exposure limits (WELs)

Fume or gas	Long term exposure limit (LTEL) (8 hour reference period)	Short term exposure limit (STEL) (15 minute reference period)
Aluminium (respirable dust)	4 mg/m^3	Nil
Cadmium oxide fume	0.025 mg/m^3	0.05 mg/m^3
Iron oxide fume (as Fe)	5 mg/m^3	10 mg/m^3
Ozone	Nil	0.4 mg/m^3/0.2 ppm
Phosgene	0.08 mg/m^3/ 0.02 ppm	0.25 mg/m^3/ 0.06 ppm
Argon	No WEL value Air O_2 content to be controlled	

Lifting equipment/pressure system requirements

Many items of site welding equipment will require lifting equipment to place them into position for the site welding to take place. The correct lifting equipment and practices will therefore be required for all slinging and lifting operations. Lifting equipment is normally subject to regular inspection in accordance with a country's published regulations. In the UK the regulations governing lifting equipment and lifting operations is the *Lifting Operations and Lifting Equipment Regulations (LOLER)*, which lays down mandatory inspection periods for lifting equipment and accessories ranging from strops and shackles to cranes.

Fixed gas systems in Great Britain are subject to the requirements of the *Pressure System Safety Regulations (PSSR: 2000),* which deal with the in-service inspection of pressure systems containing a relevant fluid. Gases above 0.5 barg are included in the relevant fluid classifications. Another set of UK regulations that welding equipment will be subject to is the *Provision and Use of Working Equipment Regulations*

(PUWER), which covers tools and equipment used in the workplace.

Hand tools and grinding machines

Hand tools must be kept in a safe and serviceable condition and only used by persons trained to use them correctly. Portable cutting and grinding equipment should have the correct disc fitted for cutting or grinding suitable for the material in use. Fixed grinding machines should only have their wheels changed by a suitably qualified and approved person and must always be used in a safe and correct manner.

Other precautions before welding starts

- Ensure persons in the area are shielded from welding arcs (use screens).
- Give a warning to persons in close proximity before striking an arc.
- Do not arc weld close to degreasing baths as UV decomposes the vapour into a toxic gas.
- Ensure firefighting equipment is in good order and close to hand.
- Ensure fire sentries (if required) are briefed and in position.

Causes of accidents

Accidents do not just happen but are usually attributable to one of the following:

- Poor communication (by the spoken or written word).
- Unfamiliarity (with the equipment, workplace or working practices).
- Just doing something wrong (because you haven't checked if it's ok to do it).
- Complacency (you reckon it will be fine – but based on what?).

Do not leave things to chance. If in doubt – check it out.

Permit-to-work

Many companies operate a written permit system for hot work. The permit details the work to be carried out, how and when it is to be done, and the precautions to be taken. A written permit system is likely to result in a higher standard of care and supervision and must be adhered to.

Tanks and drums

Flammable liquids and vapours such as petrol, diesel, fuel oil, paints, solvents, glue, lacquer and cleaning agents are found in many places of work. If welding or thermal cutting is used on a tank or drum containing a flammable material the tank or drum can explode violently. People have been killed and seriously injured by such explosions. Tanks and drums that are ‘empty’ usually still have residues in the bottom, and in seams and crevices. Just a teaspoon of flammable liquid in a drum can be enough to cause an explosion when heated and turned into vapour.

You must never weld or thermally cut a drum or tank that has contained or may have contained flammable material unless you know it has been made safe. If it contains flammable material, it will need thorough cleaning or inerting (see the HSE guidance note CS15: *The Cleaning and Gas Freeing of Tanks Containing Flammable Residues*). It may be safer for a specialist company to carry out the work. If in doubt, ask.

Index

Alloying elements 38–43
Arc strike 106
AWS A2.4 15–16

Backscatter 76
Bend tests 77–78
BSEN22553 15
Burn-through 106
Butt weld fracture test (nick break) 85–87

Carbon diagram 38
Carbon equivalency 36–37
Charpy tests 78–81
Chromium depletion 97
Classification of steels 37–38
Codes and standards 108–124
  European harmonised standards 111–112
  European Pressure Equipment Directive (PED) 111
  European quality system requirements 112
  UK Pressure Equipment Regulations (UK PER) 111
Concession 20
COSHH and workplace exposure limits 130–131
Cracks and cracking 88–93
  hydrogen induced cold cracking (HICC) 88–90
  reheat cracking 93
  solidification cracking 90–93
Crater pipe 105–106
CTOD test 81

Density 75
Destructive testing 67–87
Dimensions 10
Distortion 30–32
Documentation 17, 108–124
  essential variables 120–121
  non-essential variables 121–122
  procedure qualification record (PQR) 114
  supplementary essential variables 123–124
  welder qualifications 117–120
  welding procedure qualifications 113–114
  welding procedure specification (WPS) 114–117
Duties of a welding inspector 17–21

Electrical safety 126–127
Essential variables 120–121

European harmonised standards 111–112
European Pressure Equipment Directive (PED) 111
European quality system requirements 112
Excess weld metal 22
Excessive root penetration 103

FCAW (flux-cored arc welding) xii, 59–60
Fillet fracture test 85
Fillet weld shapes 7–8
Flame straightening 35
Fracture modes 88–107

Geometric unsharpness 76
GMAW (gas metal arc welding) 53–59

Hardness testing 81–83
HAZ (heat affected zone) 23
Health and safety 125–134
  COSHH and workplace exposure limits 130–131
  electrical safety 126–127
  gases and fumes 129–130
  personal protective equipment 128–129
  welding/cutting process safety 127
  workplace exposure limits (WELs) 132
Heat treatment (of steels) 44–48
Hydrogen induced cold cracking (HICC) 88–90

Incomplete root penetration 102

K_{IC} test 81

Lack of fusion 101–102
Lamellar tearing 93–95
Liquid penetrant testing (PT) 67–69

Macro samples 83–85
MAG (metal active gas) welding 53–59
Magnetic arc blow 107
Magnetic particle testing (MT) 69–71
Materials 36–48
  material properties 43
Micro samples 85
MIG (metal inert gas) welding 53–59
  MIG/MAG xii
MMA (manual metal arc) welding xii, 49–53

Non-destructive testing 67–87
Non-essential variables 121–122
Notified body 111

Offsetting 32
Overlap 103–104

PED (European Pressure Equipment Directive) 111
Personal protective equipment (PPE) 128–129
Porosity 98–101
Precipitation hardening alloys 47–48
Procedure qualification record (PQR) 114

Radiographic testing (RT) 73–76
  backscatter 76
  density 75
  geometric unsharpness 76
  sensitivity 75
Reheat cracking 93
Repairs 20
Residual stress and distortion 27
Root concavity 102–103
Root oxidation 106

SAW (submerged arc welding) xiii, 63–66
  fluxes 65–66
Schaeffler diagram 92
Sensitivity 75
Sequential welding 35
Shrinkage 31
SMAW (shielded metal arc welding) 49–53
Solid inclusions 100–101
Solidification cracking 90–93
Spatter 107
Stress relief 48
Supplementary essential variables 123–124

Tension tests 77
Testing, destructive and non-destructive 67–87
  bend tests 77–78
  butt weld fracture test (nick break) 85–87
  Charpy tests 78–81
  CTOD test 81
  fillet fracture test 85
  hardness testing 81–83
  K_{IC} test 81
  penetrant testing 67–69
  macro samples 83–85
  magnetic particle testing 69–71
  micro samples 85
  radiographic testing 73–76
  tension tests 77
  ultrasonic testing 71–73
Throat thickness 24
TIG (tungsten inert gas) welding xii, 61–63

UK Pressure Equipment Regulations (UK PER) 111
Ultrasonic testing (UT) 71–73
Undercut 104–105
Underfill 104

Weld decay 97–98
Weld joint preparation 18
Weld terminology 2–8

Weld throat thickness 24
Welder qualifications 117–120
Welding defects xiv, 88–107
  arc strike 106
  burn-through 106
  crater pipe 105–106
  excessive root penetration 103
  incomplete root penetration 102
  lack of fusion 101–102
  lamellar tearing 93–95
  magnetic arc blow 107
  overlap 103–104
  porosity 98–101
  root concavity 102–103
  root oxidation 106
  solid inclusions 100–101
  spatter 107
  undercut 104–105
  underfill 104
Welding problems 40
Welding procedure qualifications xiii, 113–114
Welding procedure specification (WPS) 114–117
Welding processes 49–66
Welding symbols 8–16
Welding/cutting process safety 127
Workplace exposure limits (WELs) 132

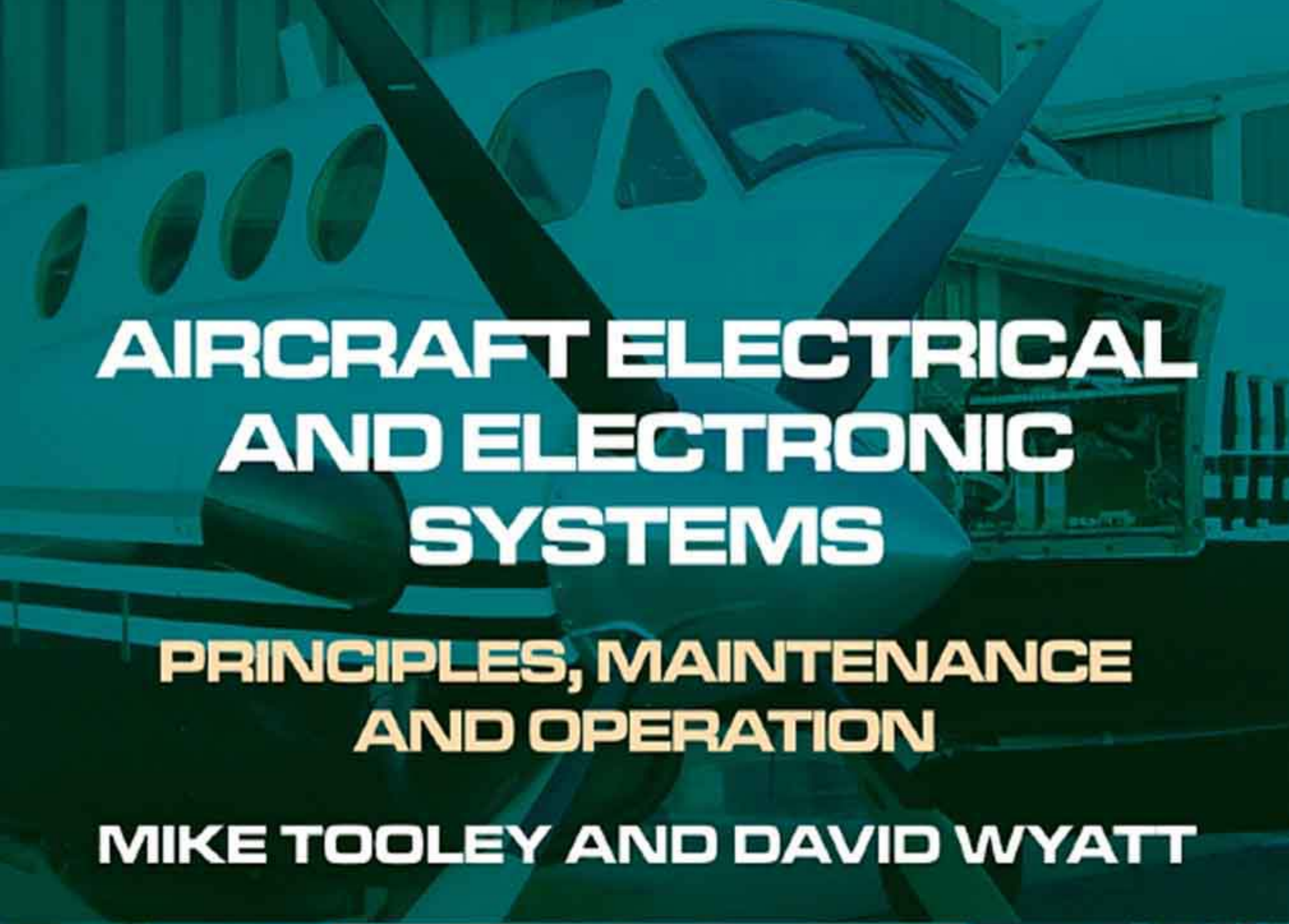

AIRCRAFT ELECTRICAL AND ELECTRONIC SYSTEMS

PRINCIPLES, MAINTENANCE AND OPERATION

MIKE TOOLEY AND DAVID WYATT

Aircraft Electrical and Electronic Systems

Principles, Operation and Maintenance

Aircraft Electrical and Electronic Systems

Principles, operation and maintenance

Mike Tooley and David Wyatt

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Butterworth-Heinemann is an imprint of Elsevier

Butterworth-Heinemann is an imprint of Elsevier
Linacre House, Jordan Hill, Oxford OX2 8DP, UK
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA

First edition 2009

Copyright © 2009, Mike Tooley and David Wyatt. Published by Elsevier Ltd.
All rights reserved

The right of Mike Tooley and David Wyatt to be identified as the authors of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44) (0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively you can submit your request online by visiting the Elsevier web site at http://elsevier.com/locate/permissions, and selecting *Obtaining permission to use Elsevier material*

Notice
No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein.

Disclaimer
This book draws on many sources. Some are facts, some hypotheses, some opinions. Most – including many of my own statements – are mixtures. Even 'facts' are unavoidably selective and can rarely be guaranteed. Despite careful checking, neither I nor my colleagues or publishers can accept responsibility for any errors, misinformation or unsuitable advice. This also applies to opinions – particularly on issues affecting health and safety. As any recommendations must balance complex, often opposing, factors, not everyone will reach the same conclusions. In this – as indeed in every issue this book touches on – every reader must make up her or his mind, for which they alone must be responsible.

I offer the best advice I am capable of, but every circumstance is different. Anyone who acts on this advice must make their own evaluation, and adapt it to their particular circumstances.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN: 978-0-7506-8695-2

For information on all Butterworth-Heinemann publications
visit our website at www.Elsevierdirect.com

Printed and bound in Great Britain

09 10 11 12 13 10 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER BOOK AID International Sabre Foundation

Contents

Preface x

Acknowledgements xvi

Chapter 1 Electrical fundamentals **1**
1.1 Electron theory 1
1.2 Electrostatics and capacitors 3
1.3 Direct current 6
1.4 Current, voltage and resistance 7
1.5 Power and energy 12
1.6 Electromagnetism and inductors 14
1.7 Alternating current and transformers 19
1.8 Safety 32
1.9 Multiple choice questions 33

Chapter 2 Electronic fundamentals **35**
2.1 Semiconductor theory 35
2.2 Diodes 38
2.3 Transistors 49
2.4 Integrated circuits 55
2.5 Multiple choice questions 56

Chapter 3 Digital fundamentals **59**
3.1 Logic gates 59
3.2 Combinational logic systems 60
3.3 Monostable devices 63
3.4 Bistable devices 65
3.5 Decoders 66
3.6 Encoders 69
3.7 Multiplexers 70
3.8 Bus systems 72
3.9 Computers 76
3.10 Multiple choice questions 78

Chapter 4 Generators and Motors **81**
4.1 Generator and motor principles 81
4.2 AC generators 89
4.3 Three-phase generation and distribution 93
4.4 AC motors 95

4.5 Practical aircraft generating systems 104
4.6 Multiple choice questions 109

Chapter 5 Batteries 111
5.1 Overview 111
5.2 Storage cells 112
5.3 Lead-acid batteries 113
5.4 Nickel-cadmium batteries 117
5.5 Lithium batteries 120
5.6 Nickel-metal hydride batteries 120
5.7 Battery locations 122
5.8 Battery venting 122
5.9 Battery connections 122
5.10 Multiple choice questions 125

Chapter 6 Power supplies 127
6.1 Regulators 127
6.2 External power 131
6.3 Inverters 131
6.4 Transformer rectifier units 133
6.5 Transformers 133
6.6 Auxiliary power unit (APU) 134
6.7 Emergency power 136
6.8 Multiple choice questions 136

Chapter 7 Wiring and circuit protection 137
7.1 Overview 137
7.2 Construction and materials 138
7.3 Specifications 139
7.4 Shielding/screening 140
7.5 Circuit protection 142
7.6 Multiple choice questions 146

Chapter 8 Distribution of power supplies 147
8.1 Single engine/general aviation 147
8.2 Twin engine general aviation aircraft 151
8.3 Larger aircraft systems 152
8.4 Split bus system 153
8.5 Parallel bus system 155
8.6 Split/parallel bus system 155
8.7 Standby and essential power 155
8.8 Battery charging 156
8.9 Control and protection 157
8.10 Load shedding 159
8.11 Multiple choice questions 163

Chapter 9 Controls and transducers 165
9.1 Switches 165
9.2 Relays and contactors 168
9.3 Variable resistors 170
9.4 Linear displacement transducers 171
9.5 Fluid pressure transducers 174
9.6 Temperature transducers 175
9.7 Strain transducers 179
9.8 Rotary position transducers 180
9.9 Multiple choice questions 182

Chapter 10 Engine systems 185
10.1 Starting and ignition 185
10.2 Indicating systems overview 195
10.3 Primary indicating systems 195
10.4 Secondary indicating systems 205
10.5 Electronic indicating systems 206
10.6 Multiple choice questions 210

Chapter 11 Fuel management 213
11.1 Storage overview 213
11.2 Fuel quantity measurement and indication 213
11.3 Fuel feed and distribution 218
11.4 Fuel transfer 220
11.5 Refuelling and defuelling 221
11.6 Fuel jettison 221
11.7 Fuel system venting 221
11.8 Fuel tank inerting 221
11.9 Multiple choice questions 223

Chapter 12 Lights 225
12.1 Lighting technologies 225
12.2 Flight compartment lights 226
12.3 Passenger cabin lights 230
12.4 Exterior lights 232
12.5 Multiple choice questions 238

Chapter 13 Cabin systems 241
13.1 Passenger address system 241
13.2 Galley equipment 243
13.3 In-flight entertainment (IFE) 243
13.4 Satellite communications 244
13.5 Multiplexing 247
13.6 Fibre optics 248
13.7 Air conditioning 249
13.8 Pressurization 252
13.9 Airstairs 254
13.10 Multiple choice questions 254

Chapter 14 Airframe control and indicating systems 257
14.1 Landing gear 257
14.2 Trailing edge flaps 260
14.3 Control surfaces 261
14.4 Electronic indicating systems 262
14.5 Multiple choice questions 264

Chapter 15 Warning and protection systems 265
15.1 Stall warning and protection 265
15.2 Airframe ice and rain protection 271
15.3 Windscreen ice and rain protection 275
15.4 Anti-skid 277
15.5 Configuration warnings 279
15.6 Aural warnings 281
15.7 Multiple choice questions 282

Chapter 16 Fire and overheat protection 283
16.1 Overview 283
16.2 Engine/APU fire detection 284
16.3 Cargo bay/baggage area 291
16.4 Fire extinguishing 294
16.5 Multiple choice questions 298

Chapter 17 Terrain awareness warning systems (TAWS) 299
17.1 System overview 299
17.2 System warnings and protection 301
17.3 External references 303
17.4 Ground proximity modes 306
17.5 Forward looking terrain awareness (FLTA) 313
17.6 Rotorcraft TAWS 316
17.7 Architecture and configurations 317
17.8 Future developments 318
17.9 Multiple choice questions 318

Chapter 18 Flight data and cockpit voice recorders 321
18.1 Flight data recorder history 321
18.2 Mandatory equipment requirements 323
18.3 Flight data recorder (FDR) specifications 326
18.4 Cockpit voice recorders 332
18.5 Health and usage monitoring system (HUMS) 334
18.6 Multiple choice questions 334

Chapter 19 Electrical and magnetic fields 337
19.1 Electromagnetic interference 337
19.2 EMI reduction 342
19.3 High-intensity/energy radiated fields 343

19.4 Lightning 344
19.5 Grounding and bonding 347
19.6 Electrostatic sensitive devices (ESSD) 348
19.7 Multiple choice questions 351

Chapter 20 Continuing airworthiness 353
20.1 Wire and cable installations 353
20.2 Bonding 359
20.3 Static charges 360
20.4 Earth returns 361
20.5 Aircraft manuals 362
20.6 Circuit testing 363
20.7 Automatic test equipment 366
20.8 On-board diagnostic equipment 366
20.9 Multiple choice questions 368

Appendix 1: Abbreviations and acronyms 369
Appendix 2: Revision papers 373
Appendix 3: Answers to multiple choice questions 379
Appendix 4: Electrical quantities, symbols and units 383
Appendix 5: Electrical formulae 385
Appendix 6: Decibels 387
Appendix 7: Wire and cable sizes 389
Appendix 8: ATA chapter/subsystem list 391
Appendix 9: Electrical and electronic symbols 395
Appendix 10: Wire numbering/coding 399

Index 401

Preface

Aircraft Electrical and Electronic Systems continues the series of textbooks written for aircraft engineering students. This book addresses the electrical contents of the EASA Part 66 Modules 11 and 13; it also provides reference material for the avionic and aircraft electrical units of various BTEC National and Higher National, City and Guilds, NVQ and Foundation Degree modules.

This book is designed to cover the essential knowledge base required by certifying mechanics, technicians and engineers engaged in engineering maintenance activities on commercial aircraft and in general aviation. In addition, this book should appeal to members of the armed forces and others attending training and educational establishments engaged in aircraft maintenance and related aeronautical engineering programmes. This book will also appeal to others within the aircraft industry who need an insight into electrical and electronic systems, e.g. pilots, engineering managers, etc.

The book provides an introduction to the fundamentals of electrical, electronic and digital theory that underpins the principles of systems covered in the remainder of the book. For the reader that already has background knowledge of the fundamentals, the subsequent chapters can be read as individual subjects. For the reader that requires a deeper understanding of related fundamentals, additional material can be found in related books in the series:

- *Aircraft Engineering Principles*
- *Aircraft Digital Electronic and Computer Systems*
- *Aircraft Communications and Navigation Systems*.

The books in this series have been designed for both independent and tutor-assisted studies. They are particularly useful to the 'self-starter' and to those wishing to update or upgrade their aircraft maintenance licence. The series also provides a useful source of reference for those taking ab initio training programmes in EASA Part 147 and FAR 147 approved organizations as well as those following related programmes in further and higher education institutions.

The title of this book, *Aircraft Electrical and Electronic Systems*, has been specifically chosen to differentiate between other avionic systems such as communications, navigation, flight guidance and instruments. The term *avionics* (aviation electronics) was first used in the late 1940s to identify electrical and electronic equipment such as radar, radio navigation and communications, although the term was not in general use until the late 1960s. During the 1970s, integrated computer-based systems were being developed, e.g. ground proximity warning systems; these used a number of existing aircraft sensors that monitored parameters such as barometric altitude, vertical speed and radio altitude.

The continued development and integration of electrical and electronic systems, together with the widespread use of integrated circuits, microprocessors, data communications and electronic displays, have given new meaning to the term avionics. Aircraft engineers will be exposed to in-service aircraft using older technology, together with the new aircraft entering service based on modern technology. Using trends from the last 40 years, there will be an ever-increasing dependence on avionic systems. The eventual outcome could be the **all-electric aircraft**, a concept where traditional mechanical linkages, hydraulics and pneumatics are totally replaced by electrical and electronic systems.

This book establishes a reference point for engineering students; it does not attempt to address all system types for all aircraft types. It is also important to note that this book does not attempt to provide the level of detail found in the aircraft publications, including the maintenance and wiring diagram manuals. Although there are many examples quoted in the book that are based on specific aircraft types, this is only done to illustrate a specific point.

Throughout the book, the principles and operation of systems are summarized by numerous 'key points'. The reader will be invited at regular intervals to assess knowledge via 'test your understanding' questions. Finally, the principles and operation of systems are put into the context of aircraft maintenance engineering by numerous 'key maintenance points'. Each chapter concludes with a number of multiple choice questions; the reader will then find revision papers in the appendices for additional assessment purposes. A summary of the book is as follows.

Chapter 1 sets the scene by providing an explanation of electricity in terms of the motion of electric charge and basic electrical quantities such as current, voltage, resistance and power. The chapter provides an introduction to electrostatics and capacitors and also to electromagnetism and inductors. Here the emphasis is on the key concepts and fundamental laws that underpin the operation of the electrical systems found in aircraft. The chapter provides a detailed introduction to alternating current and transformer principles, and concludes with an essential section on safety. This chapter will be particularly useful if you have not previously studied electrical principles. It has also been designed to help fill any gaps in your knowledge and bring you quickly up to speed.

Electronic fundamentals are introduced in Chapter 2. This chapter explains the principles, construction and basic application of a variety of common semiconductor devices including diodes, thyristors, transistors and integrated circuits. The chapter includes a detailed explanation of rectifier circuits, both half-wave and full-wave types, and the use of transistors as current amplifiers.

The advent of digital techniques and integrated circuits has revolutionized the scope and applications for avionic systems. Chapter 3 provides readers with an introduction to digital techniques. The function and operation of logic gates are established before moving on to explore the use of combinational and sequential logic in several typical aircraft applications. The chapter also provides an overview of coding systems and the logic systems that are used to represent numerical data. An introduction to aircraft data bus systems is provided together with a brief overview of the architecture and principal constituents of simple computer systems.

Generators and motors are widely used in modern aircraft. Chapter 4 explains the principles on which they operate as well as the theoretical and practical aspects of aircraft power generation and distribution. Three-phase systems and methods of connection are described in some detail.

If you are in any doubt as to whether or not you should work through Chapters 1 to 4, you can always turn to the multiple choice questions at the end of each chapter to assess your knowledge.

All electrical and electronic systems require a power source. Batteries are primary sources of electrical power found on most aircraft delivering direct current (DC). Chapter 5 reviews the battery types used on aircraft, typical applications and how they are installed and maintained. There are several types of battery used on aircraft, usually defined by the types of materials used in their construction; these include lead-acid and nickel-cadmium batteries. Other types of battery are being considered for primary power on aircraft; these include lithium and nickel-metal hydride.

In Chapter 6, we review the other sources of electrical power used on aircraft and their typical applications. Electrical power can be derived from a variety of sources; these are categorized as either primary or secondary sources. Batteries and generators are primary sources of electrical power; inverters and transformer rectifier units (TRU) are secondary sources of power. This power is either in the form of direct or alternating current depending on system requirements. Generators can either supply direct or alternating current (AC); the outputs of generators need to be regulated. Inverters are used to convert DC (usually from the battery) into alternating (AC). Transformer rectifier units (TRU) convert AC into DC; these are often used to charge batteries from AC generators. In some installations, transformers (as described in Chapter 1) are used to convert AC into AC, typically for stepping down from 115 to 26V AC. In addition to onboard equipment, most aircraft have the facility to be connected to an external power source during servicing or maintenance. An auxiliary power unit (APU) is normally used for starting the aircraft's main engines via the air distribution system. While the aircraft is on the ground, the APU can also provide electrical power. In the event of generator failure(s), continuous power can be provided by a ram air turbine (RAT).

The safe and economic operation of an aircraft is becoming ever more dependent on electrical and electronic systems. These systems are all interconnected with wires and cables; these take many forms. Chapter 7 describes the physical construction of wires and cables together with how they are protected from overload conditions before power is distributed to the various loads on the aircraft. Electrical wires and cables have to be treated as an integral part of the aircraft requiring careful installation; this is followed by direct ongoing inspection and maintenance requirements for continued airworthiness. Wire and cable installations cannot be considered (or treated) as 'fit and forget'. System reliability will be seriously affected by wiring that has not been correctly installed or maintained. We need to distribute the sources of electrical power safely and efficiently and control its use on the aircraft. Once installed, the wires and cables must be protected from overload conditions that could lead to overheating, causing the release of toxic fumes, possibly leading to fire. Legislation is being proposed to introduce a new term: **electrical wire interconnection system** (EWIS); this will acknowledge the fact that wiring is just one of many components installed on the aircraft. EWIS relates to any wire, wiring device, or combination of these, including termination devices, installed in the aircraft for transmitting electrical energy between two or more termination points.

Electrical power is supplied to the various loads in the aircraft via common points called busbars. In Chapter 8, we will focus on busbar configurations and how these are arranged for the protection and management of the various power supply sources available on the aircraft. The electrical power distribution system is based on one or more busbar(s); these provide pre-determined routes to circuits and components throughout the aircraft. The nature and complexity of the distribution system depend on the size and role of the aircraft, ranging from single-engine general aviation through to multi-engine passenger transport aircraft.

The word 'bus' (as used in electrical systems) is derived from the Latin word *omnibus* meaning 'for all'. The busbar can be supplied from one or more of the power sources previously described (generator, inverter, transformer rectifier unit or battery). Protection devices, whether fuses or circuit breakers, are connected in series with a specific system; they will remove the power from that system if an overload condition arises. There also needs to be a means of protecting the power source and feeder lines to the busbar, i.e. before the individual circuit protection devices.

There are many systems on an aircraft that need to be controlled and/or monitored, either manually by the crew, or automatically. Chapter 9 describes generic controls and transducer devices used on aircraft. A switch provides the simplest form of circuit control and monitoring. Switches can be operated manually by a person, activated by sensing movement, or controlled remotely. Many other aircraft parameters need to be measured; this is achieved by a variety of transducers; these are devices used to convert the desired parameter, e.g. pressure, temperature, displacement, etc. into electrical energy.

The aircraft engine is installed with many systems requiring electrical power. Chapter 10 describes engine starting, ignition and indicating system for both piston and gas turbine engines. The predominant electrical requirement (in terms of current consumption) is for the engine starting system. General aviation aircraft use electrical starter motors for both piston and gas turbine engines; larger transport aircraft use an air-start system (controlled electrically) derived from ground support equipment or by air cross-fed from another engine. Electrical starting systems on piston and gas turbine engines are very different. The trend towards the all-electric aircraft will see more aircraft types using electrical starting methods. The engine also requires electrical power for the ignition system; once again, the needs of piston and gas turbine engines are quite different. Although starting and ignition systems are described in this chapter as separate systems, they are both required on a co-ordinated basis, i.e. a means to rotate the engine and ignite the air/fuel mixture.

Electrical and electronic requirements for engines also include the variety of indicating systems required to operate and manage the engine. These indicating systems include (but are not limited to) the measurement and indication of: rotational speed, thrust, torque, temperature, fuel flow and oil pressure. Indications can be provided by individual indicators or by electronic displays.

The management of fuel is essential for the safe and economic operation of the aircraft. The scope of fuel management depends on the size and type of aircraft; fuel is delivered to the engines using a variety of methods. Chapter 11 provides an overview of fuel management on a range of aircraft types. The system typically comprises fuel quantity indication, distribution, refuelling, defuelling and fuel jettison. In the first instance, we need to measure the quantity of fuel on board. Various technologies and methods are used to measure fuel quantity: this depends mainly on the type and size of aircraft. Technologies range from sight gauges through to electronic sensors. On larger aircraft, fuel is fed to the engines by electrically driven pumps. On smaller aircraft, an engine-driven pump is used with electrical pumps used as back-up devices. Solenoid or motorized valves are used to isolate the fuel supply to engines under abnormal conditions. On larger aircrafts, the fuel can be transferred between tanks; this is controlled manually by the crew, or automatically by a fuel control computer.

Lighting is installed on aircraft for a number of reasons including: safety, operational needs, servicing and for the convenience of passengers. Chapter 12 reviews a number of lighting technologies and the type of equipment used in specific aircraft applications. The applications of aircraft lights can be broadly grouped into four areas: flight compartment (cockpit), passenger cabin, exterior and servicing (cargo and equipment bays). Lights are controlled by on/off switches, variable resistors or by automatic control circuits.

Passenger transport and business aircraft are fitted with a range of cabin electronic equipment for passenger safety, convenience and entertainment. Typical applications for this equipment includes lighting, audio and visual systems. Chapter 13 describes the many types of systems and equipment used for passenger safety, convenience and entertainment. Audio systems include the passenger address system used by the flight or cabin crew to give out safety announcements and other flight information. These announcements are made from hand-held microphones and are heard over loudspeakers in the cabin and passenger headsets. The same system can be used to play automatic sound tracks; this is often used for announcements in foreign languages, or to play background music during boarding and disembarkation. A range of galley equipment is installed on business and passenger aircraft. The nature of this equipment depends on the size and role of the aircraft. Air conditioning is provided in passenger aircraft for the comfort of passengers; pressurization is required for flying at high altitudes. Airstairs allow passengers, flight crew and ground personnel to board or depart the aircraft without the need for a mobile staircase or access to a terminal. All of these systems have electrical/electronic interfaces and control functions.

Chapter 14 reviews airframe systems such as landing gear control and indication, control surface position and indicating systems. Various sensors are needed for the monitoring and control of airframe systems. Broadly speaking, the sensors can be considered as detecting one of two states (or ‘conditions’), or a variable position. Two-state conditions include: landing gear (up or down) or cabin doors (open or closed). Variable positions include: control surfaces and flap position. Micro-switches or proximity sensors detect two-state positions; variable position devices are detected by a variety of devices including synchros and variable resistors.

Chapter 15 describes a variety of systems installed on aircraft to protect them from a variety of hazards including: stalling, ice, rain, unsafe take-off configuration and skidding. Stall protection systems provide the crew with a clear and distinctive warning before an unsafe condition is reached. Flying in ice and/or rain conditions poses a number of threats to the safe operation of the aircraft; ice formation can affect the aerodynamics and/or trim of the aircraft. The configuration warning system (also known as a take-off warning system) provides a warning if the pilot attempts to take off with specific controls not selected in the correct position, i.e. an unsafe configuration. The anti-skid system (also called an anti-lock braking system: ABS) is designed to prevent the main landing gear wheels from locking up during landing, particularly on wet or icy runway surfaces.

Fire on board an aircraft is a very serious hazard; all precautions must be taken to minimize the risk of a fire starting. In the event that a fire does occur, there must be adequate fire protection on the aircraft. Chapter 16 focuses on the equipment and systems used to detect fire and smoke together with the means of delivering the fire-extinguishing agent. The subject of fire protection theory is a branch of engineering in its own right. Basic fire protection theory is covered in this chapter to provide the reader with sufficient information to understand how this theory is applied through aircraft electrical systems.

During the 1970s, studies were carried out by accident investigators and regulatory authorities to examine one of the most significant causes of aircraft accidents of the time: controlled flight into terrain (CFIT). This can be defined as an accident where a serviceable aircraft, under the control of a qualified pilot, inadvertently flies into terrain, an obstacle or water. Chapter 17 describes the generic name given to this type of protection: terrain awareness warning system (TAWS). CFIT accidents usually occur during poor visual conditions, often influenced by other factors, e.g. flight crew distraction, malfunctioning equipment or air traffic control (ATC) miscommunication. With CFIT, the pilots are generally unaware of this situation until it is too late. Ground proximity warning system (GPWS) were developed in 1967 to alert pilots that their aircraft was in immediate danger of CFIT. This system was further developed into the enhanced ground proximity warning system (EGPWS) by adding a forward-looking terrain avoidance (FLTA) feature, made possible via global positioning system technology.

One of the fundamental electrical/electronic systems associated with the aircraft industry is the crash survivable flight data recorder (FDR). This is often referred to in the press as the 'black box', even though the item is actually bright orange! Flight data recorders used for accident investigation are mandatory items of equipment in commercial transport aircraft. Efforts to introduce crash-survivable flight data recorders can be traced back to the 1940s; the FDR has now been supplemented with the cockpit voice recorder (CVR). Chapter 18 reviews the range of FDR/CVR technologies that are employed for both accident investigation and trend monitoring. Recorders are used after an incident or accident as an integral part of the investigators' efforts to establish the cause(s). Data recorders can also be used to indicate trends in aircraft and engine performance. Algorithms are established for healthy and normal conditions during the aircraft's flight-testing programme and the early period of service. These algorithms include engine parameters such as engine exhaust temperature, oil pressure and shaft vibration for given speeds and altitudes. These parameters are then monitored during the aircraft's life; any deviations from the norm are analysed to determine if the engine requires inspection, maintenance or removal.

One of the consequences of operating electrical and electronic equipment is the possibility of disturbing, or interfering with, nearby items of electronic equipment. Chapter 19 looks at some of the implications of interference from electrical and magnetic fields. The term given to this type of disturbance is electromagnetic interference (EMI). Electrical or electronic products will both radiate and be susceptible to the effects of EMI. This is a paradox since many principles of electrical engineering are based on electromagnetic waves coupling with conductors to produce electrical energy and vice versa (generators and motors). Furthermore, systems are specifically designed to transmit and receive electromagnetic energy, i.e. radio equipment. In complex avionic systems, the consequences of EMI can cause serious, if not hard-to-find–problems. The ability of an item of equipment to operate alongside other items of equipment without causing EMI is electromagnetic compatibility (EMC).

Modern digital equipment operates at very high speed and relatively low power levels. In addition to EMI, high-intensity radiated fields (HIRF) are received from the external environment, e.g. from radio and radar transmitters, power lines and lightning. The high energy created by these radiated fields disrupts electronic components and systems in the aircraft. (This effect is also referred to as high-energy radiated fields – HERF.) The electromagnetic energy induces large currents to flow, causing direct damage to electronic components together with the secondary effects of EMI.

Advances in electronic technology bring many new features and benefits, e.g. faster processors, higher-density memory and highly efficient displays. These advances are primarily due to the reduction in the physical size of semiconductor junctions; this leads to higher-density components in given size of integrated circuit. One significant problem associated with certain types of semiconductor devices is that the smaller junctions are susceptible to damage from electrostatic voltages. This is a problem that can potentially affect a wide range of electronic equipment fitted in an aircraft. Effects range from weakening of semiconductor junctions through total failure of the equipment; both these effects can occur without any visible signs of damage to the naked eye! Electrostatic sensitive devices (ESSD) are electronic components that are prone to damage from stray electrical charge produced primarily from human operators. This problem is particularly prevalent with high-density memory devices and electronic displays. Weakening and damage to static-sensitive devices can result from mishandling and inappropriate methods of storage; the practical issues for handling ESSD are addressed in Chapter 19.

Many processes are required throughout the aircraft's operating life to ensure that it complies with the applicable airworthiness requirements and can be safely operated. The generic term for this range of processes is continuing airworthiness. Chapter 20 reviews some practical installation requirements, documentation and test equipment required by the avionics engineer to ensure the continued airworthiness of aircraft electrical and electronic systems. The term 'maintenance' is used for any combination of overhaul, repair, inspection, replacement, modification or defect rectification of an aircraft or component, with the exception of the pre-flight inspection. Particular emphasis is given to wire and cable installations since these cannot be considered (or treated) as 'fit and forget'. System reliability will be seriously affected by wiring that has not been correctly installed or maintained. Persons responsible for the release of an aircraft or a component after maintenance are the certifying staff. Maintenance of an aircraft and its associated systems requires a variety of test equipment and documentation; these are required by certifying staff to fulfil their obligations in ensuring continued airworthiness.

Supporting material for the book series (including interactive questions, media files, etc.) is available online at **www.66web.co.uk** or **www.key2study.com** and then follow the links for aircraft engineering.

Acknowledgements

The authors would like to express their thanks to the following persons for ideas, support and contributions to the book. We thank Lloyd Dingle, who had the original idea for the aircraft engineering series, and Alex Hollingsworth for commissioning this book. Thanks also to Lucy Potter and Holly Bathie, editorial assistants at Elsevier Science & Technology Books.

Thanks also to the following organizations for permission to reproduce their information:

- Advanced Technological Systems International Limited (new generation aircraft batteries)
- Aero Quality (battery maintenance details)
- Flight Display Systems (in-flight entertainment systems)
- Iridium Satellite LLC (satellite communication systems)
- Lees Avionics, Helicopter Services and Wycombe Air Centre (photo images taken at their premises)
- Specialist Electronic Services Ltd (flight data recorders).

Chapter 1 Electrical fundamentals

This chapter will provide you with an introduction to the essential electrical theory that underpins the rest of this book. It has been designed to help fill any gaps in your knowledge and bring you quickly up to speed. You will find this chapter particularly useful if you have not previously studied electrical principles. However, if you are in any doubt as to whether or not you should work through this chapter you can always turn to Section 1.9 on page 33 and see how you get on with the multiple choice questions at the end of this chapter.

1.1 Electron theory

All matter is made up of atoms or groups of atoms (**molecules**) bonded together in a particular way. In order to understand something about the nature of electrical charge we need to consider a simple model of the atom. This model, known as the Bohr model (see Fig. 1.1), shows a single atom consisting of a central nucleus with orbiting electrons.

Within the nucleus there are **protons** which are positively charged and **neutrons** which, as their name implies, are electrical neutral and *have no charge*. Orbiting the nucleus are electrons that have a negative charge, equal in magnitude (size) to the charge on the proton. These electrons are approximately two thousand times lighter than the protons and neutrons in the nucleus.

In a stable atom the number of protons and electrons are equal, so that overall, the atom is neutral and has no charge. However, if we rub two particular materials together, electrons may be transferred from one to another. This alters the stability of the atom, leaving it with a net positive or negative charge. When an atom within a material loses electrons it becomes positively charged and is known as a **positive ion**, when an atom gains an electron it has a surplus negative charge and so is referred to as a **negative ion**. These differences in charge can cause **electrostatic** effects. For example, combing your hair with a nylon comb may result in a difference in charge between your hair and the rest of your body, resulting in your hair standing on end when your hand or some other differently charged body is brought close to it.

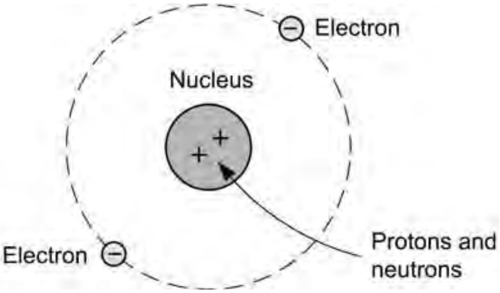

Figure 1.1 The Bohr model of the atom

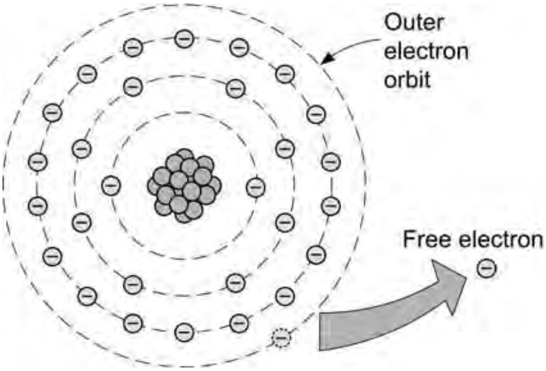

Figure 1.2 A material with a loosely bound electron in its outer shell

The number of electrons occupying a given orbit within an atom is predictable and is based on the position of the element within the periodic table. The electrons in all atoms sit in a particular orbit, or **shell**, dependent on their energy level. Each of these shells within the atom is filled by electrons from the nucleus outwards, as shown in Fig. 1.2). The first, innermost, of these shells can have up to two electrons, the second shell can have up to eight and the third up to 18.

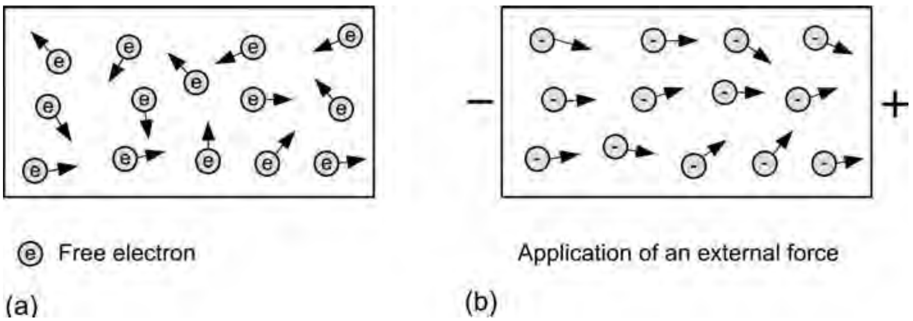

Figure 1.3 Free electrons and the application of an external force

All electrons and protons carry an electrostatic **charge** but its value is so small that a more convenient unit of charge is needed for practical use which we call the **coulomb**. One coulomb (C) is the total amount of the charge carried by 6.21×10^{18} electrons. Thus a single electron has a charge of a mere 1.61×10^{-19} C!

A material which has many free electrons available to act as charge carriers, and thus allows current to flow freely, is known as a **conductor**. Examples of good conductors include aluminium, copper, gold and iron. Figure 1.2 shows a material with one outer electron that can become easily detached from the parent atom. A small amount of external energy is required to overcome the attraction of the nucleus. Sources of such energy may include heat, light or electrostatic fields. The atom once detached from the atom is able to move freely around the structure of the material and is called a **free electron**. It is these free electrons that become the **charge carriers** within a material. Materials that have large numbers of free electrons make good **conductors** of electrical energy and heat.

In a material containing free electrons their direction of motion is random, as shown in Fig. 1.3(a), but if an external force is applied that causes the free electrons to move in a uniform manner (Fig. 1.3(b)) an electric **current** is said to flow.

Metals are the best conductors, since they have a very large number of free electrons available to act as charge carriers. Materials that do not conduct charge are called **insulators**; their electrons are tightly bound to the nuclei of their atoms. Examples of insulators include plastics, glass, rubber and ceramic materials.

The effects of electric current flow can be detected by the presence of one or more of the following effects: light, heat, magnetism, chemical, pressure and friction. For example, heat is produced when an electric current is passed through a resistive heating element. Light is produced when an electric current flows through the thin filament wire in the evacuated bulb of an electric lamp.

Key point

Electrons each carry a tiny amount of negative electrical charge.

Key point

Metals such as copper and silver are good conductors of electricity and they readily support the flow of electric current. Plastics, rubber and ceramic materials on the other hand are insulators and do not support the flow of electric current.

Test your understanding 1.1

1. Explain the following terms:
   (a) electron
   (b) ion
   (c) charge
   (d) conductor
   (e) insulator.
2. State, with reasons, whether an insulator or conductor is required in each of the following applications:
   (a) the body of a fuse
   (b) the outer protective sheath of a power cable
   (c) the fuselage covering of a transport aircraft
   (d) the radiating element of an antenna.

1.2 Electrostatics and capacitors

Electric charge is all around us. Indeed, many of the everyday items that we use in the home and at work rely for their operation on the existence of electric charge and the ability to make that charge do something useful. Electric charge is also present in the natural world and anyone who has experienced an electrical storm cannot fail to have been awed by its effects. In this section we begin by explaining what electric charge is and how it can be used to produce conduction in solids, liquids and gases.

We have already found that, if a conductor has a deficit of electrons, it will exhibit a net positive charge. If, on the other hand, it has a surplus of electrons, it will exhibit a net positive charge. An imbalance in charge can be produced by friction (removing or depositing electrons using materials such as silk and fur, respectively) or induction (by attracting or repelling electrons using a second body which is respectively positively or negatively charged).

If two bodies have charges with the same polarity (i.e. either both positively or both negatively charged) the two bodies will move apart, indicating that a force of repulsion exists between them. If, on the other hand, the charges on the two bodies are unlike (i.e. one positively charged and one negatively charged) the two bodies will move together, indicating that a force of attraction exists between them. From this we can conclude that like charges repel and unlike charges attract.

Static charges can be produced by friction. In this case, electrons and protons in an insulator are separated from each other by rubbing two materials together in order to produce opposite charges. These charges will remain separated for some time until they eventually leak away due to losses in the insulating **dielectric** material or in the air surrounding the materials. Note that more charge will be lost in a given time if the air is damp.

Static electricity is something that can cause particular problems in an aircraft and special measures are taken to ensure that excessive charges do not build up on the aircraft's structure. The aim is that of equalizing the potential of all points on the aircraft's external surfaces. The static charge that builds up during normal flight can be dissipated into the atmosphere surrounding the aircraft by means of small conductive rods connected to the aircraft's trailing surfaces. These are known as **static dischargers** or **static wicks** – see Fig. 1.4.

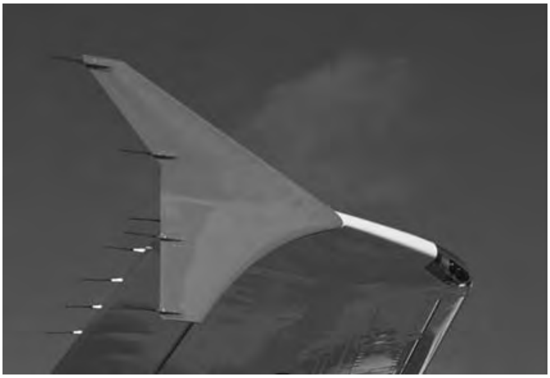

Figure 1.4 Static discharging devices

Key point

Charged bodies with the same polarity repel one another whilst charges with opposite polarity will attract one another.

Key point

A significant amount of charge can build up between conducting surfaces when they are insulated from one another. Where this might be a problem steps are taken to dissipate the charge instead of allowing it to accumulate uncontrolled.

Key maintenance point

Stray static charges can very easily damage static-sensitive devices such as semiconductors, memory devices and other integrated circuits. Damage can be prevented by adopting the appropriate electrostatic sensitive device (ESD) precautions (described in the aircraft maintenance manual) when handling such devices. Precautions usually involve using wrist straps and grounding leads as well as using static-dissipative packaging materials.

1.2.1 Electric fields

The force exerted on a charged particle is a manifestation of the existence of an electric field. The electric

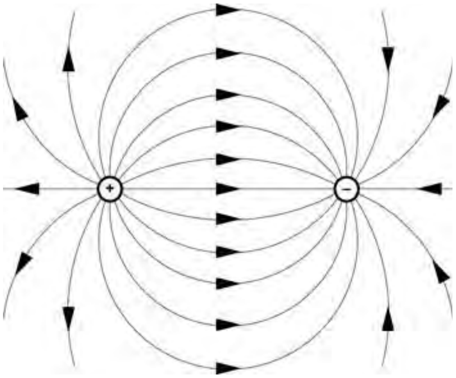

Figure 1.5 Electric field between isolated unlike charges

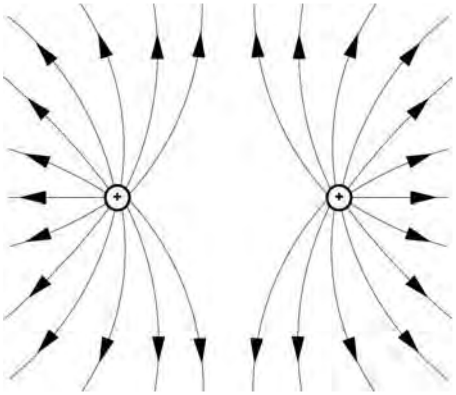

Figure 1.6 Electric field between isolated like charges

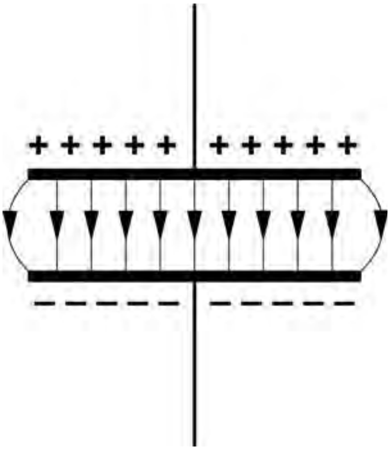

Figure 1.7 Electric field between the two charged parallel metal plates of a capacitor

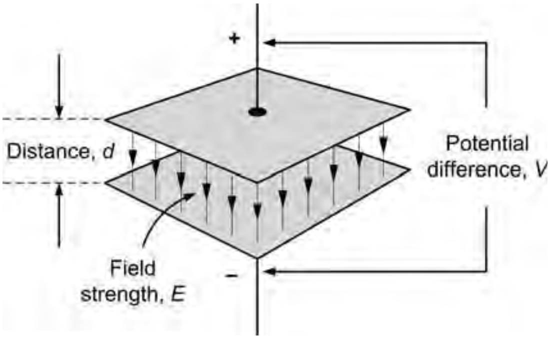

Figure 1.8 Electric field strength between two charged conducting surfaces

field defines the direction and magnitude of a force on a charged object. The field itself is invisible to the human eye but can be drawn by constructing lines which indicate the motion of a free positive charge within the field; the number of field lines in a particular region being used to indicate the relative strength of the field at the point in question.

Figures 1.5 and 1.6 show the electric fields between isolated unlike and like charges whilst Fig. 1.7 shows the field that exists between two charged parallel metal plates which forms a charge storage device known as a **capacitor**.

The strength of an electric field (E) is proportional to the applied **potential difference** and inversely proportional to the distance between the two conducting surfaces (see Fig. 1.8). The electric field strength is given by:

$$E = \frac{V}{d}$$

where E is the electric field strength (in V/m), V is the applied potential difference (in V) and d is the distance (in m).

The amount of charge that can be stored by a capacitor is given by the relationship:

$$Q = C \times V$$

where Q is the charge in coulomb, C is the capacitance in farads, F, and V is the voltage in volts, V. This relationship can be re-arranged to make C or V the subject as follows:

$$C = \frac{Q}{V} \text{ and } V = \frac{Q}{C}$$

Example 1.1

Two parallel conductors are separated by a distance of 25 mm. Determine the electric field strength if they are fed from a 600 V DC supply.

The electric field strength will be given by:

$$E = \frac{V}{d}$$

where $V = 600$ V and $d = 25$ mm $= 0.025$ m.
Thus:

$$E = \frac{600}{0.025} = 24{,}000\,\text{V/m} = 24\,\text{kV/m}$$

Example 1.2

The field strength between two parallel plates in a cathode ray tube is 18 kV/m. If the plates are separated by a distance of 21 mm determine the potential difference that exists between the plates.

The electric field strength will be given by:

$$E = \frac{V}{d}$$

Re-arranging this formula to make V the subject gives:

$$V = E \times d$$

Now $E = 16$ kV/m $= 18{,}000$ V/m and $d =$ 21 mm $=0.021$ m, thus:

$$V = 18{,}000 \times 0.021 = 378\,\text{V}$$

Example 1.3

A potential difference of 150 V appears across the plates of a 2 µF capacitor. What charge is present?

The charge can be calculated from:

$$Q = C \times V$$

where $C = 2$ µF and $V = 150$ V, thus:

$$Q = 2\,\mu\text{F} \times 150\,\text{V} = 300\,\mu\text{C}.$$

Example 1.4

A 68 µF capacitor is required to store a charge of 170 µC. What voltage should be applied to the capacitor?

The voltage can be calculated from:

$$V = \frac{Q}{C}$$

where $Q = 170$ µC and $C = 68$ µF; thus:

$$V = \frac{3.4\,\text{mC}}{6.8\,\mu\text{F}} = \frac{3{,}400\,\mu\text{C}}{6.8\,\mu\text{F}} = 500\,\text{V}$$

Key maintenance point

When replacing a capacitor it is essential to ensure that the replacement component is correctly rated in terms of type, value, working voltage and temperature. Capacitors are prone to failure if their maximum working voltage is exceeded and they should be de-rated when operated at a relatively high ambient temperature according to manufacturers' specifications. It is also essential to observe the correct polarity when replacing an electrolytic (polarized) component. This is usually clearly marked on the external casing.

Key maintenance point

When working with high-voltage capacitors it is essential to ensure that the capacitor is fully discharged before attempting to replace the component. In most cases, any accumulated charge will safely drain away within a few seconds after removal of power. However, this should not be relied upon and a safe discharge path through a high-value resistor (say 1 MΩ) fitted with appropriate probes will ensure that capacitor is safe to work on.

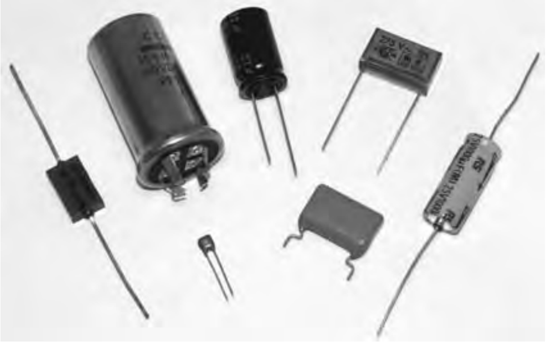

Figure 1.9 A selection of capacitors with values ranging from 12 pF to 1000 μF and working voltages ranging from 25 V to 450 V

Test your understanding 1.2

1. The two plates of a parallel plate capacitor are separated by a distance of 15 mm. If the potential difference between the plates is 300 V what will the electric field strength be?
2. The electric field between two conducting surfaces is 500 V/m. If the plates are separated by a distance of 2.5 mm, determine the potential difference between the plates.

1.3 Direct current

Direct current (DC) is current that flows in one direction only. DC circuits are found in every aircraft. An understanding of how and why these circuits work is an essential prerequisite to understanding more complex circuits. Because of their negative charge, electrons will flow from a point of negative potential to a point with more positive potential (recall that like charges attract and unlike charges repel). However, when we indicate the direction of current in a circuit we show it as moving from a point that has the greatest positive potential to a point that has the most negative potential. We call this **conventional current** and, although it may seem odd, you just need to remember that it flows in the *opposite* direction to that of the motion of electrons!

The most commonly used method of generating direct current is the electrochemical cell. A **cell** is a device that produces a charge when a chemical reaction takes place. When several cells are connected together they form a **battery**.

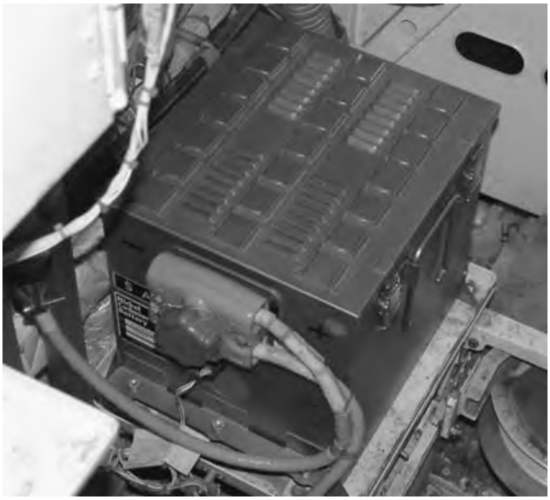

Figure 1.10 A typical aircraft battery

There are two types of cell: primary and secondary. **Primary cells** produce electrical energy at the expense of the chemicals from which they are made and once these chemicals are used up, no more electricity can be obtained from the cell. In **secondary cells**, the chemical action is reversible. This means that the chemical energy is converted into electrical energy when the cell is **discharged** whereas electrical energy is converted into chemical energy when the cell is being **charged**. You will find more information on aircraft batteries in Chapter 5.

Key point

Conventional current flows from positive to negative whilst electrons travel in the opposite direction, from negative to positive.

Key point

In a primary cell the conversion of chemical energy to electrical energy is irreversible and so these cells cannot be recharged. In secondary cells, the conversion of chemical energy to electrical energy is reversible. Thus these cells can be recharged and reused many times.

Key maintenance point

When removing and replacing batteries, it is essential to observe the guidance given in the aircraft maintenance manual (AMM) when removing, charging or replacing aircraft batteries. The AMM will describe the correct procedures for isolating the battery from the aircraft's electrical system prior to its physical removal.

Test your understanding 1.3

1. Explain the difference between a primary and a secondary cell.
2. Explain the difference between electron flow and conventional current.

1.4 Current, voltage and resistance

Current, I, is defined as the rate of flow of charge and its unit is the ampere, A. One ampere is equal to one coulomb C per second, or:

$$\text{One ampere of current } I = \frac{Q}{t}$$

Where t = time in seconds

So, for example: if a steady current of 3 A flows for two minutes, then the amount of charge transferred will be:

$$Q = I \times t = 3\,\text{A} \times 120\,\text{s} = 360 \text{ coulombs}$$

Example 1.5

A current of 45 mA flows from one point in a circuit to another. What charge is transferred between the two points in 10 minutes?

Here we will use $Q = It$ where $I = 45\,\text{mA} = 0.045\,\text{A}$ and $t = 10$ minutes $= 10 \times 60 = 600$ s.

Thus:

$$Q = 45\,\text{mA} \times 600\,\text{s} = 0.045\,\text{A} \times 600\,\text{s} = 27\,\text{C}$$

Example 1.6

A charge of 1.5 C is transferred to a capacitor in 30 seconds. What current is flowing in the capacitor?

Here we will use $I = Q/t$ where $Q = 1.5\,\text{C}$ and $t = 30$ s. Thus:

$$I = \frac{Q}{t} = \frac{1.5\,\text{C}}{30\,\text{s}} = 0.05\,\text{A} = 50\,\text{mA}$$

Key point

Current is the rate of flow of charge. Thus, if more charge moves in a given time, more current will be flowing. If no charge moves then no current is flowing.

1.4.1 Potential difference (voltage)

The force that creates the flow of current (or rate of flow of charge carriers) in a circuit is known as the **electromotive force** (or **e.m.f.**) and it is measured in volts (V). The **potential difference** (or **p.d.**) is the voltage difference, or voltage drop between two points.

One volt is the potential difference between two points if one Joule of energy is required to move one coulomb of charge between them. Hence:

$$V = \frac{W}{Q}$$

where W = energy and Q = charge, as before. Energy is defined later in Section 1.6.

Test your understanding 1.4

1. How much charge will be transferred when a current of 6 A flows for two minutes?
2. How long will it take for a charge of 0.2 C to be transferred using a current of 0.5 A?
3. If 0.4 J of energy is used to transfer 0.05 C of charge between two points what is the potential difference between the two points?

1.4.2 Resistance

All materials at normal temperatures oppose the movement of electric charge through them; this opposition to the flow of the charge carriers is known as the **resistance**, R, of the material. This resistance is due to collisions between the charge carriers (electrons) and the atoms of the material. The unit of resistance is the **ohm**, with symbol Ω.

Note that 1 V is the electromotive force (e.m.f.) required to move 6.21×10^{18} electrons (1 C) through a resistance of $1\,\Omega$ in 1 second. Hence:

$$V = \left(\frac{Q}{t}\right) \times R$$

where Q = charge, t = time, and R = resistance.

Re-arranging this equation to make R the subject gives:

$$R = \frac{V \times t}{Q}\ \Omega$$

Example 1.7

A 28 V DC aircraft supply delivers a charge of 5 C to a window heater every second. What is the resistance of the heater?

Here we will use $R = (V \times t)/Q$ where $V = 28\,\text{V}$, $Q = 5\,\text{C}$ and $t = 1\,\text{s}$. Thus:

$$R = \frac{V \times t}{Q} = \frac{28\,\text{V} \times 1\text{s}}{5\,\text{C}} = 5.6\,\Omega$$

Key point

Metals such as copper and silver are good conductors of electricity. Good conductors have low resistance whilst poor conductors have high resistance.

1.4.3 Ohm's law

The most basic DC circuit uses only two components; a cell (or battery) acting as a source of e.m.f., and a resistor (or *load*) through which a current is passing. These two components are connected together with wire conductors in order to form a completely closed circuit as shown in Fig. 1.11.

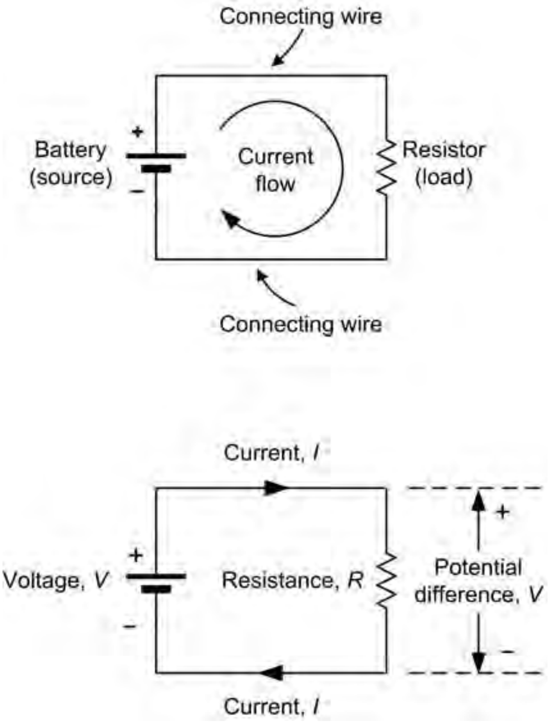

Figure 1.11 A simple DC circuit consisting of a battery (source) and resistor (load)

For any conductor, the current flowing is directly proportional to the e.m.f. applied. The current flowing will also be dependent on the physical dimensions (length and cross-sectional area) and material of which the conductor is composed. The amount of current that will flow in a conductor when a given e.m.f. is applied is inversely proportional to its resistance. Resistance, therefore, may be thought of as an 'opposition to current flow'; the higher the resistance the lower the current that will flow (assuming that the applied e.m.f. remains constant).

Provided that temperature does not vary, the ratio of p.d. across the ends of a conductor to the current flowing in the conductor is a constant. This relationship is known as Ohm's law and it leads to the relationship:

$$\frac{V}{I} = \text{a constant} = R$$

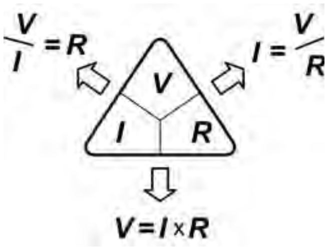

Figure 1.12 Relationship between *V*, *I* and *R*

where V is the potential difference (or voltage drop) in volts (V), I is the current in amps (A), and R is the resistance in ohms (Ω).

This important formula may be arranged to make V, I or R the subject, as follows:

$$V = I \times R \qquad I = \frac{V}{R} \qquad R = \frac{V}{I}$$

The triangle shown in Fig. 1.12 should help you remember these three important relationships. It is important to note that, when performing calculations of currents, voltages and resistances in practical circuits, it is seldom necessary to work with an accuracy of better than ±1% simply because component tolerances are invariably somewhat greater than this. Furthermore, in calculations involving Ohm's law, it is sometimes convenient to work in units of kΩ and mA (or MΩ and μA), in which case potential differences will be expressed directly in V.

Example 1.8

A current of 0.1 A flows in a 220 Ω resistor. What voltage drop (potential difference) will be developed across the resistor?

Here we must use $V = I \times R$ and ensure that we work in units of volts (V), amps (A), and ohms (Ω).

$$V = I \times R = 0.1\,\text{A} \times 22\,\Omega = 22\,\text{V}$$

Hence a p.d. of 22 V will be developed across the resistor.

Example 1.9

An 18 Ω resistor is connected to a 12 V battery. What current will flow in the resistor?

Figure 1.13 A selection of resistors with values ranging from 0.1 Ω to 10 MΩ and power ratings from 0.1 W to 15 W

Here we must use $I = V/R$ (where V = 12 V and R = 18 Ω):

$$I = \frac{V}{R} = \frac{24\,\text{V}}{10\,\Omega} = 2.4\,\text{A}$$

Hence a current of 2.4 A will flow in the resistor.

Example 1.10

A voltage drop of 15 V appears across a resistor in which a current of 1 mA flows. What is the value of the resistance?

Here we must use $R = V/I$ (where V = 15 V and I = 1 mA = 0.001 A)

$$R = \frac{V}{I} = \frac{15\,\text{V}}{0.001\,\text{A}} = 15{,}000\,\Omega = 15\,\text{k}\Omega$$

Note that it is sometimes more convenient to work in units of mA and V, which will produce an answer directly in kΩ, i.e.

$$R = \frac{V}{I} = \frac{15\,\text{V}}{1\,\text{mA}} = 15\,\text{k}\Omega$$

Test your understanding 1.5

1. An aircraft cable has a resistance of 0.02 Ω per foot. If a 20 foot length of this cable carries a current of 0.5 A what voltage will be dropped across the ends of the cable?

2. A relay has a coil resistance of 400 Ω. What current is required to operate the relay from a 24 V supply?
3. A current of 125 μA flows when an insulation tester delivers 500 V to a circuit. What is the resistance of the circuit?

Key maintenance point

When replacing a resistor it is essential to ensure that the replacement component is correctly rated in terms of type, value, power and temperature. Resistors are prone to failure if their maximum power is exceeded and they should be de-rated when operated at a relatively high ambient temperature according to manufacturers' specifications.

1.4.4 Kirchhoff's laws

Used on its own, Ohm's law is insufficient to determine the magnitude of the voltages and currents present in complex circuits. For these circuits we need to make use of two further laws; Kirchhoff's current law and Kirchhoff's voltage law.

Kirchhoff's current law states that the algebraic sum of the currents present at a junction (or *node*) in a circuit is zero – see Fig. 1.14. Kirchhoff's voltage law states that the algebraic sum of the potential drops present in a closed network (or *mesh*) is zero – see Fig. 1.15.

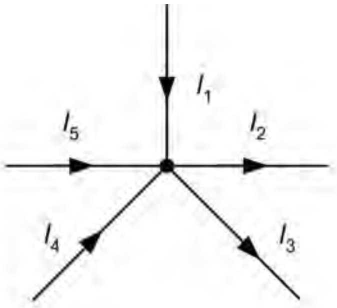

$$I_1 - I_2 - I_3 + I_4 + I_5 = 0$$

Convention:
Current flowing towards the junction is positive (+)
Current flowing away from the junction is negative (–)

Figure 1.14 Kirchhoff's current law

1.4.5 Series and parallel circuits

Ohm's law and Kirchhoff's laws can be combined to solve more complex series–parallel circuits. Before we show you how this is done, however, it's important to understand what we mean by 'series' and 'parallel' circuit!

Figure 1.16 shows three circuits, each containing three resistors, R_1, R_2 and R_3:

- In Fig. 1.16(a), the three resistors are connected one after another. We refer to this as a *series circuit*. In other words the resistors are said to be connected *in series*. It's important to note that, in this arrangement, *the same current flows through each resistor*.
- In Fig. 1.16(b), the three resistors are all connected across one another. We refer to this as a *parallel circuit*. In other words the resistors are said to be connected *in parallel*. It's important to note that, in this arrangement, *the same voltage appears across each resistor*.

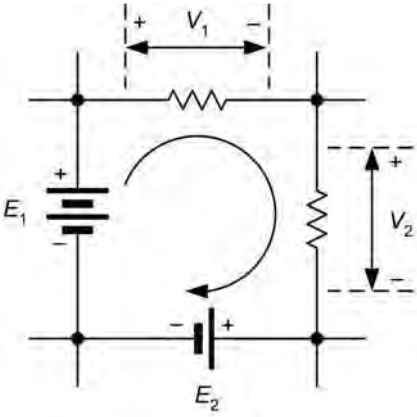

$$E_1 - V_1 - V_2 - E_2 = 0$$

Convention:
Move clockwise around the circuit starting with the positive terminal of the largest e.m.f.
Voltages acting in the same sense are positive (+)
Voltages acting in the opposite sense are negative (–)

Figure 1.15 Kirchhoff's voltage law

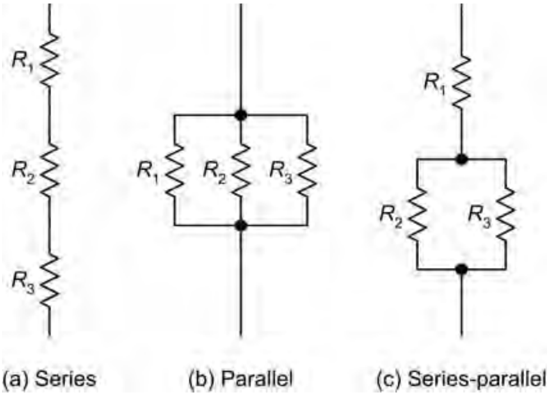

(a) Series (b) Parallel (c) Series-parallel

Figure 1.16 Series and parallel circuits

- In Fig. 1.16(c), we have shown a mixture of these two types of connection. Here we can say that R_1 is connected in series with the parallel combination of R_2 and R_3. In other words, R_2 and R_3 are connected *in parallel* and R_2 is connected *in series* with the parallel combination.

Example 1.11

Figure 1.17 shows a simple battery test circuit which is designed to draw a current of 2 A from a 24 V DC supply. The two test points, A and B, are designed for connecting a meter. Determine:

(a) the voltage that appears between terminals A and B (without the meter connected);
(b) the value of resistor, R.

We need to solve this problem in several small stages. Since we know that the circuit draws 2 A from the 24 V supply we know that this current must flow both through the 9 Ω resistor and through R (we hope that you have spotted that these two components are connected *in series*!).

We can determine the voltage drop across the 9 Ω resistor by applying Ohm's law:

$$V = I \times R = 2\,\text{A} \times 9\,\Omega = 18\,\text{V}$$

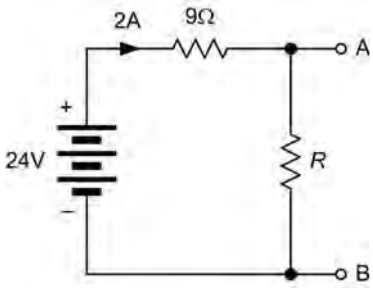

Figure 1.17 Battery test circuit – see Example 1.11

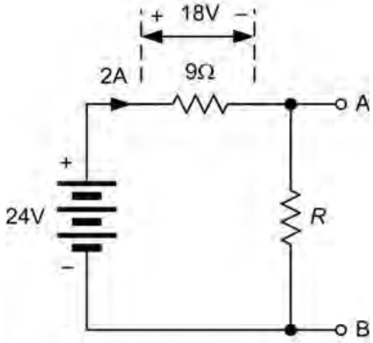

Figure 1.18 Using Ohm's law to find the voltage dropped across the 9 Ω resistor – see Example 1.11

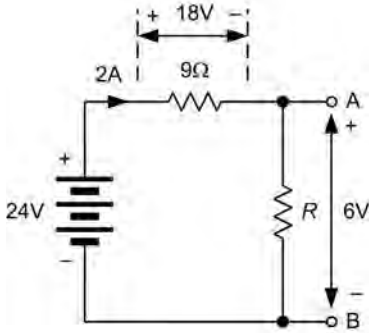

Figure 1.19 Using Kirchhoff's law to find the voltage that appears between terminals A and B – see Example 1.11

Next we can apply Kirchhoff's voltage law in order to determine the voltage drop, V, that appears across R (i.e. the potential drop between terminals A and B):

$$+24\,\text{V} - 18\,\text{V} - V = 0$$

From which:

$$V = +6\,\text{V}$$

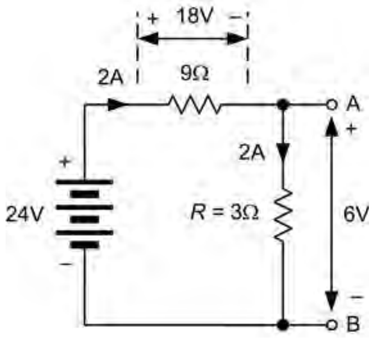

Figure 1.20 Using Ohm's law to find the value of R – see Example 1.11

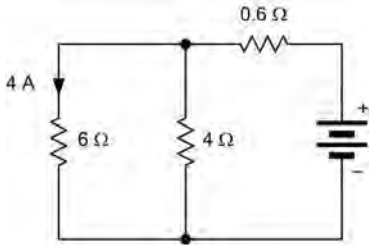

Figure 1.21 See Test your understanding 1.6

Finally, since we now know the voltage, V, and current, I, that flows in R, we can apply Ohm's law again in order to determine the value of R:

$$R = \frac{V}{I} = \frac{6\,\text{V}}{2\,\text{A}} = 3\,\Omega$$

Key point

Circuits with multiple branches can be solved using a combination of Kirchhoff's laws and Ohm's law.

Test your understanding 1.6

Determine the current and voltage present in each branch of the circuit shown in Fig. 1.21.

1.5 Power and energy

Power, P, is the rate at which energy is converted from one form to another and it is measured in *watts* (W). The larger the amount of power the greater the amount of energy that is converted in a given period of time.

$$1\ \text{watt} = 1\ \text{joule per second}$$

or

$$\text{Power, } P = \frac{\text{Energy, } W}{\text{time, } t}$$

thus:

$$P = \frac{W}{t}\ \text{W}$$

Like all other forms of energy, electrical energy is the capacity to do work. Energy can be converted from one form to another. An electric fire, for example, converts electrical energy into heat. A filament lamp converts electrical energy into light, and so on. Energy can only be transferred when a difference in energy levels exists.

The unit of energy is the *joule* (J). Then, from the definition of power,

$$1\ \text{joule} = 1\ \text{watt} \times 1\ \text{second}$$

hence:

$$\text{Energy, } W = (\text{power, } P) \times (\text{time, } t)$$
$$\text{with units of (watts} \times \text{seconds)}$$

thus

$$W = \text{P}t\ \text{J}$$

Thus joules are measured in **watt-seconds**. If the power was to be measured in kilowatts and the time in hours, then the unit of electrical energy would be the **kilowatt-hour** (kWh) (commonly knows as a **unit of electricity**). The electricity meter in your home records the amount of energy that you have used expressed in kilowatt-hours.

The power in an electrical circuit is equivalent to the product of voltage and current. Hence:

$$P = I \times V$$

where P is the power in watts (W), I is the current in amps (A), and V is the voltage in volts (V).

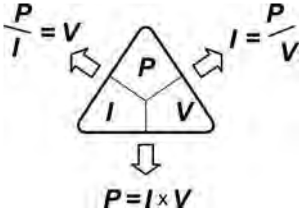

Figure 1.22 Relationship between *P*, *I* and *V*

The formula may be arranged to make *P*, *I* or *V* the subject, as follows:

$$P = I \times V \qquad I = \frac{P}{V} \qquad V = \frac{P}{I}$$

The triangle shown in Fig. 1.22 should help you remember these three important relationships. It is important to note that, when performing calculations of power, current and voltages in practical circuits it is seldom necessary to work with an accuracy of better than ±1% simply because component tolerances are invariably somewhat greater than this.

Finally, we can combine the Ohm's law relationship that we met earlier with the formulae for power to arrive at two further useful relationships:

$$P = I \times V = I \times (IR) = I^2R$$

$$P = I \times V = \left(\frac{V}{R}\right) \times V = \frac{V^2}{R}$$

Key point

Power is the rate of using energy and a power of one watt corresponds to energy being used at the rate of one joule per second.

Example 1.12

An auxiliary power unit (APU) provides an output of 1.5 kW for 20 minutes. How much energy has it supplied to the aircraft?

Here we will use $W = Pt$ where $P = 1.5\,\text{kW} = 1500\,\text{W}$ and $t = 20$ minutes $= 20 \times 60 = 1200$ s.

Thus:

$$W = 1500 \times 1200 = 1{,}800{,}000\,\text{J} = 1.8\,\text{MJ}$$

Example 1.13

The reservoir capacitor in a power supply is required to store 20 J of energy. How much power is required to store this energy in a time interval of 0.5 s?

Re-arranging $W = Pt$ to make P the subject gives:

$$P = \frac{W}{t}$$

We can now find P when $W = 20$ J and $t = 0.5$ s. Thus

$$P = \frac{W}{t} = \frac{20\,\text{J}}{0.5\,\text{s}} = 400\,\text{W}$$

Example 1.14

A main aircraft battery is used to start an engine. If the starter demands a current of 1000 A for 30 s and the battery voltage remains at 12 V during this period, determine the amount of electrical energy required to start the engine.

We need to solve this problem in two stages. First we need to find the power delivered to the starter from:

$$P = IV$$

where $I = 1000$ A and $V = 12$ V. Thus

$$P = 1000 \times 12 = 12{,}000\,\text{W} = 12\,\text{kW}$$

Next we need to find the energy from:

$$W = Pt$$

where $P = 12\,\text{kW}$ and $t = 30$ s. Thus

$$W = 12\,\text{kW} \times 30\,\text{s} = 360\,\text{kJ}$$

Example 1.15

A 24 V bench power unit is to be tested at its rated load current of 3 A. What value of load resistor is required and what should its minimum power rating be?

The value of load resistance required can be calculated using Ohm's law, as follows:

$$R = \frac{V}{I} = \frac{24\,\text{V}}{3\,\text{A}} = 8\,\Omega$$

The minimum power rating for the resistor will be given by:

$$P = IV = 3\,\text{A} \times 24\,\text{V} = 72\,\text{W}$$

Key point

Power is the rate at which energy is converted from one form to another. A power of one watt is equivalent to one joule of energy being converted every second.

Test your understanding 1.7

1. A window heater is rated at 150W. How much energy is required to operate the heater for one hour?
2. A resistor is rated at 11 Ω, 2W. What is the maximum current that should be allowed to flow in it?
3. An emergency locator transmitter is fitted with a lithium battery having a rated energy content of 18 kJ. How long can the unit be expected to operate if the transmitter consumes an input power of 2W?

1.6 Electromagnetism and inductors

Magnetism is an effect created by moving the elementary atomic particles in certain materials such as iron, nickel and cobalt. Iron has outstanding magnetic properties, and materials that behave magnetically, in a similar manner to iron, are known as ferromagnetic materials. These materials experience forces that act on them when placed near a magnet.

A magnetic field of flux is the region in which the forces created by the magnet have influence. This field surrounds a magnet in all directions, being strongest at the end extremities of the magnet, known as the poles. Magnetic fields are mapped by an arrangement of lines that give an indication of strength and direction of the flux as illustrated in Fig. 1.23.

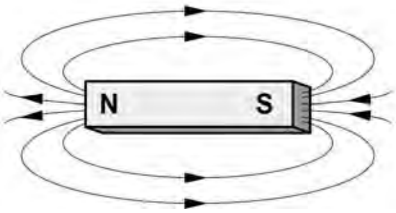

Figure 1.23 Field and flux directions for a bar magnet

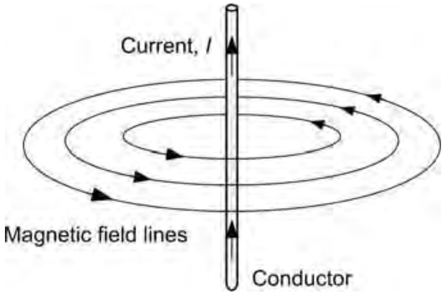

Figure 1.24 Field around a current-carrying conductor

Whenever an electric current flows in a conductor a magnetic field is set up around the conductor in the form of concentric circles, as shown in Fig. 1.24. The field is present along the whole length of the conductor and is strongest nearest to the conductor. Now like permanent magnets, this field also has direction. The direction of the magnetic field is dependent on the direction of the current passing through the conductor.

If we place a current-carrying conductor in a magnetic field, the conductor has a force exerted on it. Consider the arrangement shown in Fig. 1.25, in which a current-carrying conductor is placed between two magnetic poles. The direction of the current passing through it is into the page going away from us. Then by the right-hand screw rule, the direction of the magnetic field, created by the current in the conductor, is clockwise, as shown. We also know that the flux lines from the permanent magnet exit at a north

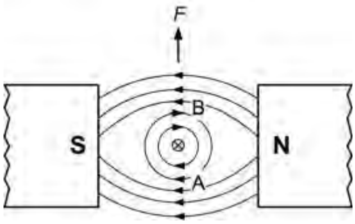

Figure 1.25 A current-carrying conductor in a magnetic field

pole and enter at a south pole; in other words, they travel from north to south, as indicated by the direction arrows. The net effect of the coming together of these two magnetic force fields is that at position A, they both travel in the same direction and reinforce one another. While at position B, they travel in the opposite direction and tend to cancel one another. So with a stronger force field at position A and a weaker force at position B the conductor is forced upwards out of the magnetic field. If the direction of the current was reversed, i.e. if it was to travel towards us out of the page, then the direction of the magnetic field in the current-carrying conductor would be reversed and therefore so would the direction of motion of the conductor.

Key point

A magnetic field of flux is the region in which the forces created by the magnet have influence. This field surrounds a magnet in all directions and is concentrated at the north and south poles of the magnet.

Key point

Whenever an electric current flows in a conductor a magnetic field is set up in the space surrounding the conductor. The field spreads out around the conductor in concentric circles with the greatest density of magnetic flux nearest to the conductor.

The magnitude of the force acting on the conductor depends on the current flowing in the conductor, the length of the conductor in the field, and the strength of the magnetic flux (expressed in terms of its **flux density**). The size of the force will be given by the expression:

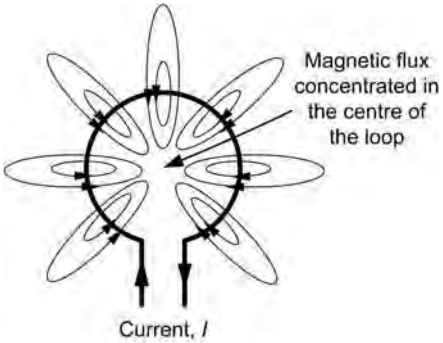

Figure 1.26 Magnetic field around a single turn loop

$$F = BIl$$

where F is the force in newtons (N), B is the flux density in tesla (T), I is the current (A) and l is the length (m).

Flux density is a term that merits a little more explanation. The total flux present in a magnetic field is a measure of the total magnetic intensity present in the field and it is measured in webers (Wb) and represented by the Greek symbol, Φ. The flux density, B, is simply the total flux, Φ, divided by the area over which the flux acts, A. Hence:

$$B = \frac{\Phi}{A}$$

where B is the flux density (T), Φ is the total flux present (Wb), and A is the area (m^2).

In order to increase the strength of the field, a conductor may be shaped into a loop (Fig. 1.26) or coiled to form a **solenoid** (Fig. 1.27).

Example 1.16

A flux density of 0.25 T is developed in free space over an area of 20 cm^2. Determine the total flux.

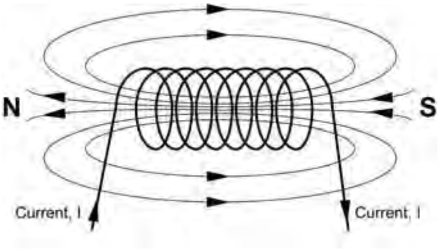

Figure 1.27 Magnetic field around a coil or solenoid

Re-arranging the formula $B = \frac{\Phi}{A}$ to make Φ the subject gives:

$$\Phi = BA$$

thus:

$$\Phi = 0.25\,\text{T} \times 20\,\text{cm}^2 = 0.25\,\text{T} \times 0.002\,\text{m}^2 = 0.0005\,\text{Wb or } 0.5\,\text{mWb}$$

Key point

If we place a current-carrying conductor in a magnetic field, the conductor has a force exerted on it. If the conductor is free to move this force will produce motion.

Key point

Flux density is found by dividing the total flux present by the area over which the flux acts.

1.6.1 Electromagnetic induction

The way in which electricity is generated in a conductor may be viewed as being the exact opposite to that which produces the motor force. In order to generate electricity we require movement in to get electricity out. In fact we need the same components to generate electricity as those needed for the electric motor, namely a closed conductor, a magnetic field and movement.

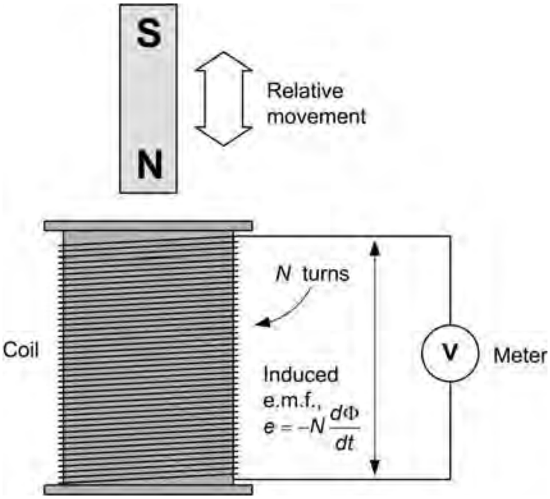

Figure 1.28 Demonstration of electromagnetic induction

Whenever relative motion occurs between a magnetic field and a conductor acting at right angles to the field, an e.m.f. is induced, or generated in the conductor. The manner in which this e.m.f. is generated is based on the principle of electromagnetic induction.

Consider Fig. 1.28, which shows relative movement between a magnet and a closed coil of wire. An e.m.f. will be induced in the coil whenever the magnet is moved in or out of the coil (or the magnet is held stationary and the coil moved). The magnitude of the induced e.m.f., e, depends on the number of turns, N, and the rate at which the flux changes in thecoil, $d\Phi/dt$. Note that this last expression is simply amathematical way of expressing the *rate of change of flux with respect to time*.

The e.m.f., e, is given by the relationship:

$$e = -N\frac{d\Phi}{dt}$$

where N is the number of turns and $d\Phi/dt$ is the rateof change of flux. The minus sign indicates that the polarity of the generated e.m.f. *opposes* the change.

Now the number of turns N is directly related to the length of the conductor, l, moving through a magnetic field with flux density, B. Also, the velocity with which

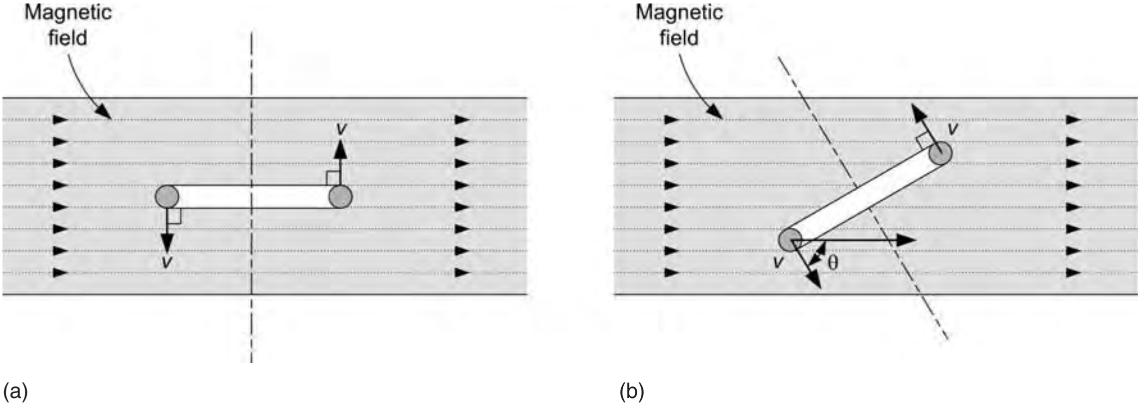

Figure 1.29 Cutting lines of flux and the e.m.f. generated: (a) cutting lines of flux at 90°, *e* = *Blv*; (b) cutting lines of flux at θ, *e* = *Blv* sin θ

the conductor moves through the field determines the rate at which the flux changes in the coil as it cuts the flux field. Thus the magnitude of the induced (generated) e.m.f., *e*, is proportional to the flux density, length of conductor and relative velocity between the field and the conductor.

The magnitude of the induced e.m.f. also depends on:

- the length of the conductor *l* in m
- the strength of the magnetic field, *B*, in tesla (T)
- the velocity of the conductor, *v*, in m/s.

Hence:

$$e \propto Blv$$

where *B* is the strength of the magnetic field (T), *l* is the length of the conductor in the field (m), and *v* is the velocity of the conductor (m/s).

Now you are probably wondering why the above relationship has the proportionality sign. In order to generator an e.m.f. the conductor must cut the lines of magnetic flux. If the conductor cuts the lines of flux at right angles (Fig. 1.29(a)) then the maximum e.m.f. is generated; cutting them at any other angle θ (Fig. 1.29(b)), reduces this value until $\theta = 0°$, at which point the lines of flux are not being cut at all and no e.m.f. is induced or generated in the conductor. So the magnitude of the induced e.m.f. is also dependent on sin θ . So we may write:

$$e = Blv \sin\theta$$

1.6.2 Faraday's and Lenz's laws

When a magnetic flux through a coil is made to vary, an e.m.f. is induced. The magnitude of this e.m.f. is proportional to the rate of change of magnetic flux. What this law is saying in effect is that relative movement between the magnetic flux and the conductor is essential to generate an e.m.f. The voltmeter shown in Fig. 1.29 indicates the induced (generated) e.m.f. and if the direction of motion changes the polarity of the induced e.m.f. in the conductor changes. Faraday's law also tells us that the magnitude of the induced e.m.f. is dependent on the relative velocity with which the conductor cuts the lines of magnetic flux.

Lenz's law states that the current induced in a conductor opposes the changing field that produces it. It is therefore important to remember that the induced current *always* acts in such a direction so as to oppose the change in flux. This is the reason for the minus sign in the formula that we met earlier:

$$e = -N\frac{d\Phi}{dt}$$

Key point

The induced e.m.f. tends to oppose any change of current and because of this we often refer to it as a *back e.m.f.*

Example 1.17

A closed conductor of length 15 cm cuts the magnetic flux field of 1.25 T with a velocity of 25 m/s. Determine the induced e.m.f. when:

(a) the angle between the conductor and field lines is 60°
(b) the angle between the conductor and field lines is 90°.

(a) The induced e.m.f. is found using $e = Blv \sin\theta$, hence:

$$\begin{aligned} e &= 1.25 \times 0.15 \times 25 \times \sin 60^\circ \\ &= 4.688 \times 0.866 = 4.06\,\text{V} \end{aligned}$$

(b) The maximum induced e.m.f. occurs when the lines of flux are cut at 90°. In this case $e = Blv \sin\theta = Blv$ (recall that $\sin 90^\circ = 1$), hence:

$$e = 1.25 \times 0.15 \times 25 = 4.69\,\text{V}$$

1.6.3 Self-inductance and mutual inductance

We have already shown how an induced e.m.f. (i.e. a **back e.m.f.**) is produced by a flux change in an inductor. The back e.m.f. is proportional to the rate of change of current (from Lenz's law), as illustrated in Fig. 1.30.

This effect is called *self-inductance* (or just *inductance*) which has the symbol L. Self-inductance is measured in henries (H) and is calculated from:

$$e = -L\frac{di}{dt}$$

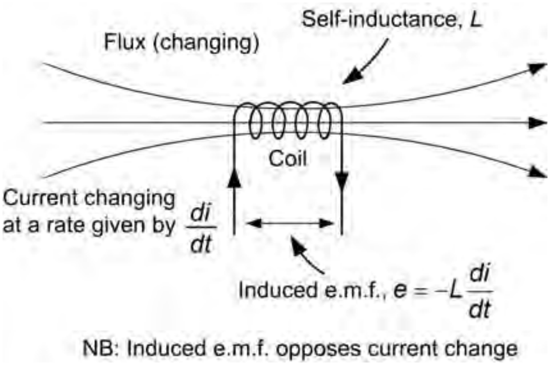
Self-inductance, L
Flux (changing)
Coil
Current changing at a rate given by $\frac{di}{dt}$
Induced e.m.f., $e = -L\frac{di}{dt}$
NB: Induced e.m.f. opposes current change

Figure 1.30 Self-inductance

where L is the self-inductance, di/dt is the rate ofchange of current and the minus sign indicates that the polarity of the generated e.m.f. *opposes* the change (you might like to compare this relationship with the one shown earlier for electromagnetic induction).

The unit of inductance is the henry (H) and a coil is said to have an inductance of 1 H if a voltage of 1 V is induced across it when a current changing at the rate of 1 A/s is flowing in it.

Example 1.18

A coil has a self-inductance of 15 mH and is subject to a current that changes at a rate of 450 A/s. What e.m.f. is produced?

Now $e = -L\,di/dt$ and hence:

$$e = -15 \times 10^{-3} \times 450 = -6.75\,\text{V}$$

Note the minus sign. This reminds us that a *back e.m.f.* of 6.75 V is induced.

Example 1.19

A current increases at a uniform rate from 2 A to 6 A in a time of 250 ms. If this current is applied to an inductor determine the value of inductance if a back e.m.f. of 15 V is produced across its terminals.

Now $e = -L\,di/dt$ and hence $L = -e\,di/dt$

Thus

$$\begin{aligned} L = -(-15)\times\frac{250 \times 10^{-3}}{(6-2)} &= 15 \times 62.5 \times 10^{-3} \\ &= 937.5 \times 10^{-3} \\ &= 0.94\,\text{H} \end{aligned}$$

Finally, when two inductors are placed close to one another, the flux generated when a changing current flows in the first inductor will cut through the other inductor (see Fig. 1.31). This changing flux will, in turn, induce a current in the second inductor. This effect is known as **mutual inductance** and it occurs whenever two inductors are inductively coupled. This is the principle of a very useful component, the **transformer**, which we shall meet later.

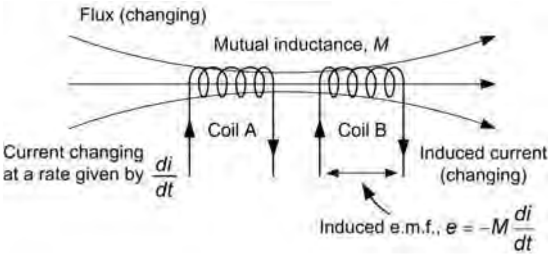
Flux (changing)
Mutual inductance, M
Coil A
Coil B
Current changing at a rate given by $\frac{di}{dt}$
Induced current (changing)
Induced e.m.f., $e = -M\frac{di}{dt}$

Figure 1.31 Mutual inductance

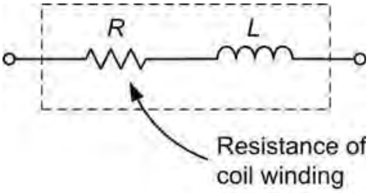

Figure 1.32 A real inductor has resistance as well as inductance

1.6.4 Inductors

Inductors provide us with a means of storing electrical energy in the form of a magnetic field. Typical applications include chokes, filters, and frequency selective circuits. The electrical characteristics of an inductor are determined by a number of factors including the material of the core (if any), the number of turns, and the physical dimensions of the coil.

In practice every coil comprises both inductance and resistance and the circuit of Fig. 1.32 shows these as two discrete components. In reality the inductance, *L*, and resistance, *R*, are both distributed throughout the component but it is convenient to treat the inductance and resistance as separate components in the analysis of the circuit.

Key point

An e.m.f. is produced when the magnetic flux passing through an inductor changes.

Figure 1.33 A selection of inductors with values ranging from 100 nH to 4 H and current ratings ranging from 0.1 A to 10 A

Key point

The current induced in a conductor always opposes the change that produces it.

Test your understanding 1.8

1. A 1.5 m length of wire moves perpendicular to a magnetic flux field of 0.75 T. Determine the e.m.f. that will be induced across the ends of the wire if it moves at 10 m/s.
2. An e.m.f. of 30 V is developed across the terminals of an inductor when the current flowing in it changes from zero to 10 A in half a second. What is the value of inductance?

1.7 Alternating current and transformers

Direct currents are currents which, even though their magnitude may vary, essentially flow only in one direction. In other words, direct currents are unidirectional. Alternating currents, on the other hand, are bi-directional and continuously reversing their direction of flow, as shown in Fig. 1.34.

A graph showing the variation of voltage or current present in a circuit is known as a **waveform**.

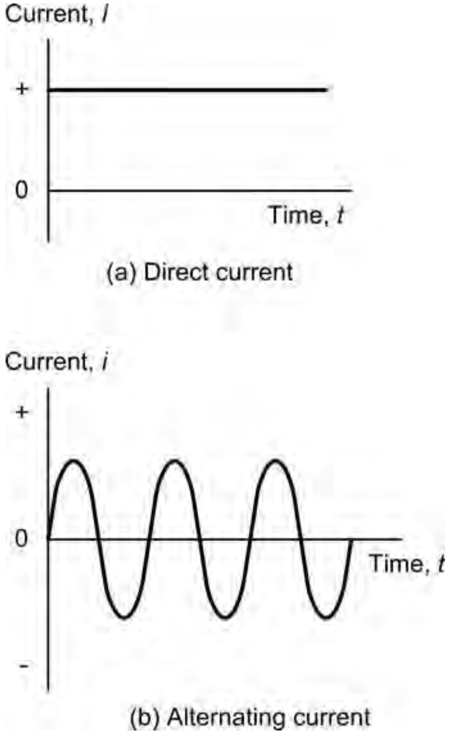

Figure 1.34 Comparison of direct and alternating current

There are many common types of waveform encountered in electrical circuits including sine (or sinusoidal), square, triangle, ramp or sawtooth (which may be either positive or negative), and pulse. Complex waveforms like speech or music usually comprise many components at different frequencies. Pulse waveforms found in digital circuits are often categorized as either repetitive or non-repetitive (the former comprises a pattern of pulses which regularly repeats whilst the latter comprises pulses which constitute a unique event). Several of the most common waveform types are shown in Fig. 1.35.

1.7.1 Frequency and periodic time

The **frequency** of a repetitive waveform is the number of cycles of the waveform that occur in unit time. Frequency is expressed in **hertz** (**Hz**). A frequency of 1 Hz is equivalent to one cycle per second. Hence, if an aircraft supply has a frequency of 400 Hz, 400 cycles of the supply will occur in every second (Fig. 1.36).

The **periodic time** (or **period**) of a waveform is the time taken for one complete cycle of the wave (see Fig. 1.37). The relationship between periodic time and frequency is thus:

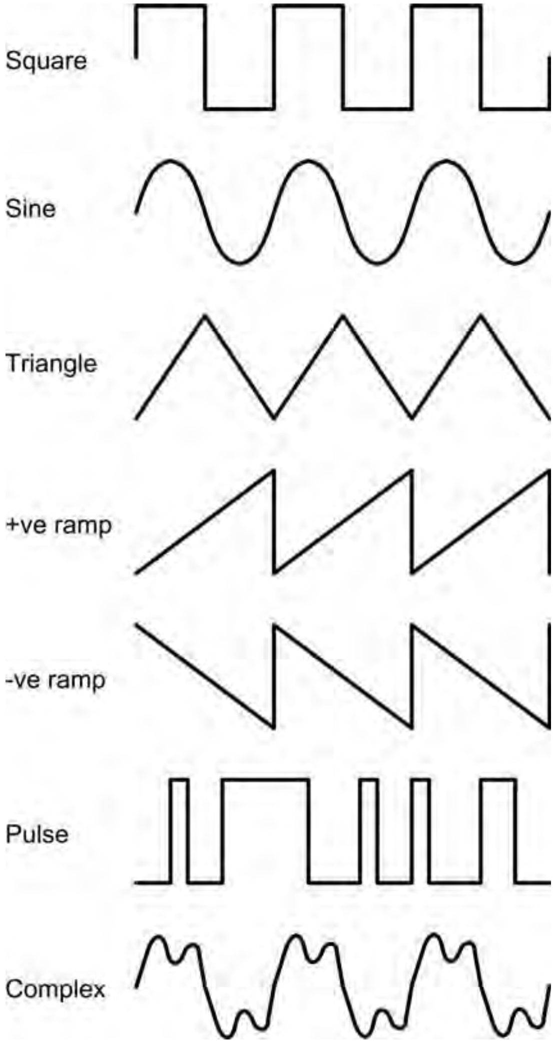

Figure 1.35 Various waveforms

$$t = \frac{1}{f} \text{ or } f = \frac{1}{t}$$

where t is the periodic time (in seconds) and f is the frequency (in Hz).

Example 1.20

An aircraft generator operates at a frequency of 400 Hz. What is the periodic time of the voltage generated?

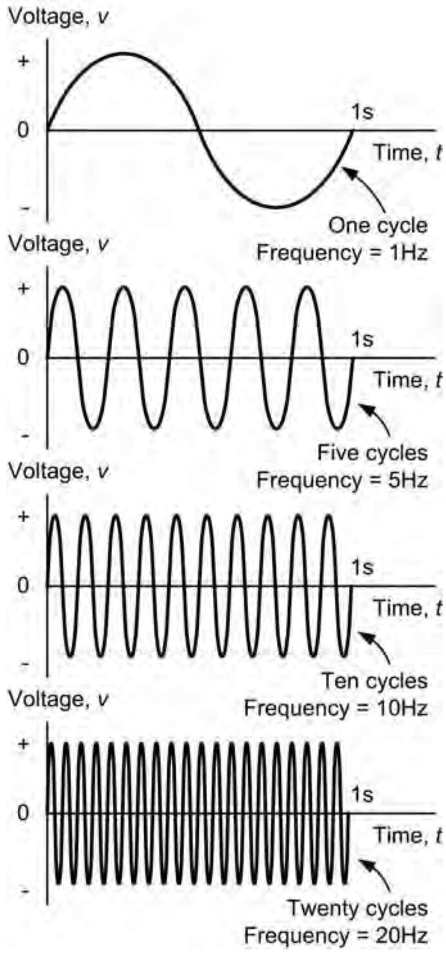

Figure 1.36 Waveforms with different frequencies

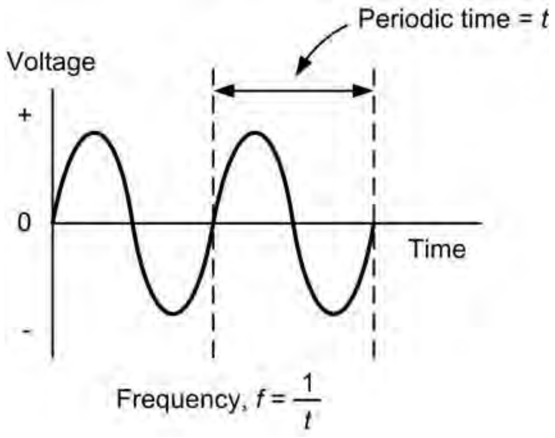

Figure 1.37 Periodic time

Now

$$t = \frac{1}{f} = \frac{1}{400\,\text{Hz}} = 0.0025\,\text{s} = 2.5\,\text{ms}$$

Hence the voltage will have a periodic time of 2.5 ms.

Example 1.21

A bench AC supply has a periodic time of 20 ms. What is the frequency of the supply?

Now

$$f = \frac{1}{t} = \frac{1}{20\,\text{ms}} = \frac{1}{0.02\,\text{s}} = 50\,\text{Hz}$$

Hence the supply has a frequency of 50 Hz.

1.7.2 Average, peak, peak–peak, and r.m.s. values

The **average value** of an alternating current which swings symmetrically above and below zero will obviously be zero when measured over a long period of time. Hence average values of currents and voltages are invariably taken over one complete half cycle (either positive or negative) rather than over on complete full cycle (which would result in an average value of zero).

The **peak value** (or **maximum value** or **amplitude**) of a waveform is a measure of the extent of its voltage or current excursion from the resting value (usually zero). The peak-to-peak value for a wave which is symmetrical about its resting value is twice its peak value.

The **root mean square** (**r.m.s.**) or **effective value** of an alternating voltage or current is the value which would produce the same heat energy in a resistor as a direct voltage or current of the same magnitude. Since the r.m.s. value of a waveform is very much dependent upon its shape, values are only meaningful when dealing with a waveform of known shape. Where the shape of a waveform is not specified, r.m.s. values are normally assumed to refer to sinusoidal conditions.

For a given waveform, a set of fixed relationships exist between average, peak, peak–peak, and r.m.s. values. The required multiplying factors for sinusoidal voltages and currents are summarized in the table shown below.

		Wanted quantity			
		average	*peak*	*peak–peak*	*r.m.s.*
Given quantity	average	1	1.57	3.14	1.11
	peak	0.636	1	2	0.707
	peak–peak	0.318	0.5	1	0.353
	r.m.s.	0.9	1.414	2.828	1

From the table we can conclude that, for example:

$$V_{\text{av}} = 0.636 \times V_{\text{pk}},\ V_{\text{pk}-\text{pk}} = 2 \times V_{\text{pk}},\ \text{and}\ V_{\text{r.m.s.}} = 0.707 \times V_{\text{pk}}$$

Similar relationships apply to the corresponding alternating currents, thus:

$$I_{\text{av}} = 0.636 \times I_{\text{pk}},\ I_{\text{pk}-\text{pk}} = 2 \times I_{\text{peak}},\ \text{and}\ I_{\text{r.m.s.}} = 0.707 \times I_{\text{pk}}$$

Example 1.22

A generator produces an r.m.s. sine wave output 110V. What is the peak value of the voltage?

Now

$$V_{\text{pk}} = 1.414 \times V_{\text{r.m.s.}} = 1.414 \times 110\,\text{V} = 155.5\,\text{V}$$

Hence the voltage has a peak value of 311V.

Example 1.23

A sinusoidal current of 40A peak–peak flows in a circuit. What is the r.m.s. value of the current?

Now

$$I_{\text{r.m.s}} = 0.353 \times I_{\text{pk}-\text{pk}} = 0.353 \times 40\,\text{A} = 14.12\,\text{A}$$

Hence the current has an r.m.s. value of 14.12 A.

Key point

The root mean square (r.m.s.) value of an alternating voltage will produce the same amount of heat in a resistor as a direct voltage of the same magnitude.

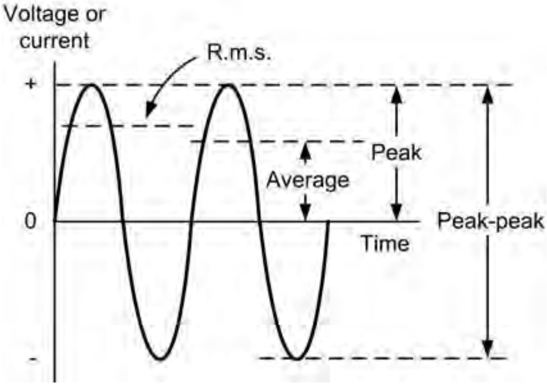

Figure 1.38 Average, r.m.s., peak and peak–peak values of a sine wave

1.7.3 Three-phase supplies

The most simple method of distributing an AC supply is a system that uses two wires. In fact, this is how AC is distributed in your home (the third wire present is simply an earth connection for any appliances that may require it for safety reasons). In many practical applications, including aircraft, it can be advantageous to use a **multiphase** supply rather than a **single-phase** supply (here the word *phase* simply refers to an AC voltage source that may not be rising and falling at the same time as other voltage sources that may be present).

The most common system uses three separate voltage sources (and three wires) and is known as **three-phase**. The voltages produced by the three sources are spaced equally in time such that the angle between them is 120° (or 360°/3). The waveforms for a three-phase supply are shown in Fig. 1.37 (note that each is a sine wave and all three sine waves have the same frequency and periodic time). We shall be returning to this topic in greater detail in Chapter 4 when we introduce three-phase power generation.

1.7.4 Reactance

In an AC circuit, **reactance**, like resistance, is simply the ratio of applied voltage to the current flowing. Thus:

$$I = \frac{V}{X}$$

where X is the reactance in ohms (Ω), V is the alternating potential difference in volts (V) and I is the alternating current in amps (A).

In the case of **capacitive reactance** (i.e. the reactance of a capacitor) we use the suffix, C, so that the reactance equation becomes:

$$X_C = \frac{V_C}{I_C}$$

Similarly, in the case of **inductive reactance** (i.e. the reactance of an inductor) we use the suffix, L, so that the reactance equation becomes:

$$X_L = \frac{V_L}{I_L}$$

The voltage and current in a circuit containing pure reactance (either capacitive or inductive) will be out of phase by 90°. In the case of a circuit containing pure capacitance the current will lead the voltage by 90° (alternatively we can say that the voltage lags the current by 90°). This relationship is illustrated by the waveforms shown in Fig. 1.40.

In the case of a circuit containing pure inductance the voltage will lead the current by 90° (alternatively we can also say that the current lags the voltage by 90°). This relationship is illustrated by the waveforms shown in Fig. 1.41.

Key point

A good way of remembering leading and lagging phase relationships is to recall the word *CIVIL*, as shown in Fig. 1.42. Note that, in the case of a circuit containing pure capacitance (*C*) the current (*I*) will lead the voltage (*V*) by 90° whilst in the case of a circuit containing pure inductance (*L*) the voltage (*V*) will lead the current (*I*) by 90°.

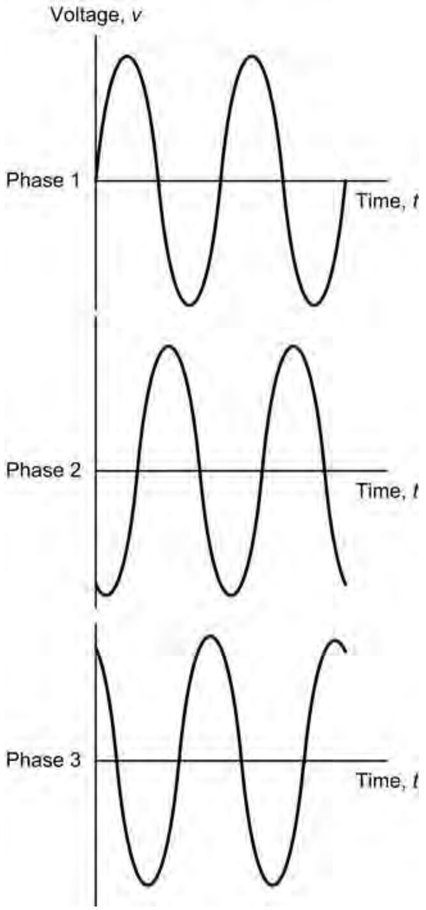

Figure 1.39 Waveforms for a three-phase AC supply

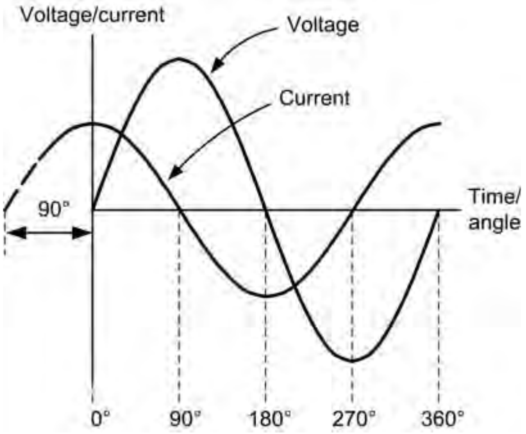

Figure 1.40 Voltage and current waveforms for a pure capacitor (the current leads the voltage by 90°)

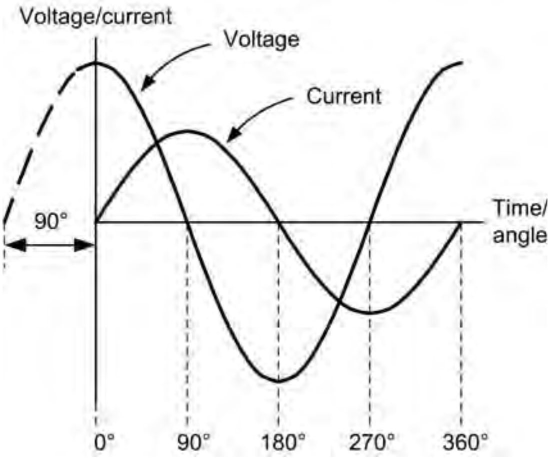

Figure 1.41 Voltage and current waveforms for a pure capacitor (the voltage leads the current by 90°)

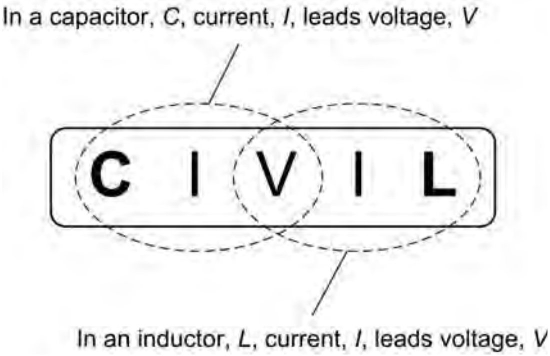
In a capacitor, C, current, I, leads voltage, V
C I V I L
In an inductor, L, current, I, leads voltage, V

Figure 1.42 Using CIVIL to determine phase relationships in circuits containing capacitance and inductance

The reactance of an inductor (**inductive reactance**) is directly proportional to the frequency of the applied alternating current and can be determined from the following formula:

$$X_L = 2\pi fL$$

where X_L is the reactance in Ω, f is the frequency in Hz, and L is the inductance in H.

Since inductive reactance is directly proportional to frequency ($X_L \propto f$), the graph of inductive reactance plotted against frequency takes the form of a straight line (see Fig. 1.43).

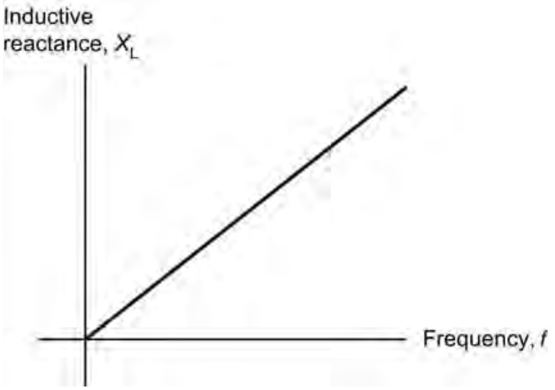

Figure 1.43 Variation of inductive reactance, X_L, with frequency, f

Example 1.24

Determine the reactance of a 0.1 H inductor at (a) 100 Hz and (b) 10 kHz.

(a) At 100 Hz, $X_L = 2\pi \times 100 \times 0.1 = 62.8\,\Omega$
(b) At 10 kHz, $X_L = 2\pi \times 10{,}000 \times 0.1 = 6280\,\Omega = 6.28\,\text{k}\Omega$.

The reactance of a capacitor (**capacitive reactance**) is inversely proportional to the frequency of the applied alternating current and can be determined from the following formula:

$$X_C = \frac{1}{2\pi fC}$$

where X_C is the reactance in Ω, f is the frequency in Hz, and C is the capacitance in F.

Since capacitive reactance is inversely proportional to frequency ($X_L \propto 1/f$), the graph of inductive reactance plotted against frequency takes the form of a rectangular hyperbola (see Fig. 1.44).

Example 1.25

Determine the reactance of a 1 μF capacitor at (a) 100 Hz and (b) 10 kHz.

(a) At 100 Hz,

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi \times 100 \times 1 \times 10^{-6}}$$
$$= \frac{0.159}{10^{-4}} = 1.59\,\text{k}\Omega$$

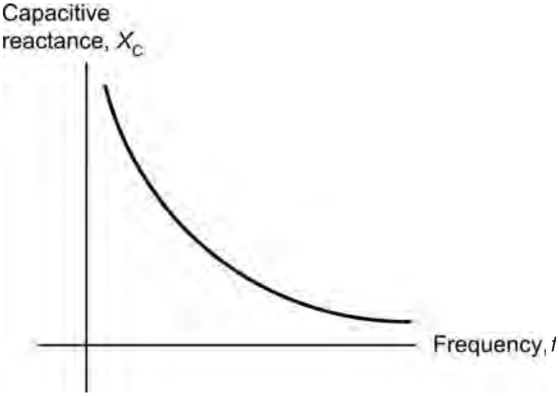

Figure 1.44 Variation of capacitive reactance, X_C, with frequency, f

(b) At 10 kHz,

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi \times 10 \times 10^3 \times 1 \times 10^{-6}}$$
$$= 0.159 \times 10^2 = 15.9\,\Omega$$

Key point

When alternating voltages are applied to capacitors or inductors the magnitude of the current flowing will depend upon the value of capacitance or inductance and on the frequency of the voltage. In effect, capacitors and inductors oppose the flow of current in much the same way as a resistor. The important difference being that the effective resistance (or *reactance*) of the component varies with frequency (unlike the case of a conventional resistor where the magnitude of the current does not change with frequency).

1.7.5 Impedance

Circuits that contain a mixture of both resistance and reactance (either capacitive reactance or inductive reactance or both) are said to exhibit **impedance**. Impedance, like resistance and reactance, is simply the ratio of applied voltage to the current flowing. Thus:

$$Z = \frac{V}{I}$$

where Z is the impedance in ohms (Ω), V is the alternating potential difference in volts (V) and I is the alternating current in amps (A).

Because the voltage and current in a pure reactance are at 90° to one another (we say that they are in **quadrature**) we can't simply add up the resistance and reactance present in a circuit in order to find its impedance. Instead, we can use the **impedance triangle** shown in Fig. 1.45. The impedance triangle takes into account the 90° phase angle and from it we can infer that the impedance of a series circuit (R in series with X) is given by:

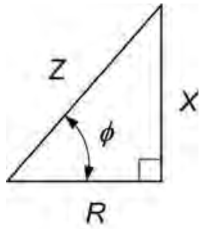

Figure 1.45 The impedance triangle

$$Z = \sqrt{R^2 + X^2}$$

where Z is the impedance (in Ω), X is the reactance, either capacitive or inductive (expressed in Ω), and R is the resistance (also in Ω).

We shall be explaining the significance of the **phase angle**, ϕ, later on. For now you simply need to be aware that ϕ is the angle between the impedance, Z, and the resistance, R. Later on we shall obtain some useful information from the fact that:

$$\sin\phi = \frac{opposite}{hypotenuse} = \frac{X}{Z} \text{ from which}$$
$$\phi = \arcsin\left(\frac{X}{Z}\right)$$

$$\cos\phi = \frac{adjacent}{hypotenuse} = \frac{R}{Z} \text{ from which}$$
$$\phi = \arccos\left(\frac{R}{Z}\right)$$

and

$$\tan\phi = \frac{opposite}{adjacent} = \frac{X}{R} \text{ from which}$$
$$\phi = \arctan\left(\frac{X}{R}\right)$$

Key point

Resistance and reactance combine together to make *impedance*. In other words, impedance is the *resultant* of combining resistance and reactance in the impedance triangle. Because of the *quadrature* relationship between voltage and current in a pure capacitor or inductor, the angle between resistance and reactance in the impedance triangle is always 90°.

Example 1.26

A resistor of 30 Ω is connected in series with a capacitive reactance of 40 Ω. Determine the impedance of the circuit and the current flowing when the circuit is connected to a 115 V supply.

First we must find the impedance of the C–R series circuit:

$$Z = \sqrt{R^2 + X^2} = \sqrt{30^2 + 40^2}$$
$$= \sqrt{2{,}500} = 50\,\Omega$$

The current taken from the supply can now be found:

$$I = \frac{V}{Z} = \frac{115}{50} = 2.3\,\text{A}$$

Example 1.27

A coil is connected to a 50 V AC supply at 400 Hz. If the current supplied to the coil is 200 mA and the coil has a resistance of 60 Ω, determine the value of inductance.

Like most practical forms of inductor, the coil in this example has both resistance *and* reactance (see Fig. 1.31). We can find the impedance of the coil from:

$$Z = \frac{V}{I} = \frac{50}{0.2} = 250\,\Omega$$

Since $Z = \sqrt{R^2 + X^2}$,

$$Z^2 = R^2 + X^2 \quad \text{and} \quad X^2 = Z^2 - R^2$$

from which:

$$X^2 = Z^2 - R^2 = 250^2 - 60^2$$
$$= 62{,}500 - 3{,}600$$
$$= 58{,}900$$

Thus

$$X = \sqrt{58{,}900} = 243\,\Omega$$

Now since $X_L = 2\pi fL$,

$$L = \frac{X_L}{2\pi f} = \frac{243}{6.28 \times 400}$$
$$= \frac{243}{2{,}512} = 0.097\,\text{H}$$

Hence

$$L = 97\,\text{mH}$$

1.7.6 Resonance

It is important to note that a special case occurs when $X_C = X_L$ in which case the two equal but opposite reactances effectively cancel each other out. The result of this is that the circuit behaves as if only resistance, R, is present (in other words, the impedance of the circuit, $Z = R$). In this condition the circuit is said to be **resonant**. The frequency at which resonance occurs is given by:

$$X_C = X_L$$

thus

$$\frac{1}{2\pi fC} = 2\pi fL$$

from which

$$f^2 = \frac{1}{4\pi^2 LC}$$

and thus

$$f = \frac{1}{2\pi\sqrt{LC}}$$

where f is the resonant frequency (in Hz), L is the inductance (in H) and C is the capacitance (in F).

Example 1.28

A series circuit comprises an inductor of 10 mH, a resistor of 50 Ω and a capacitor of 40 nF. Determine the frequency at which this circuit is resonant and the current that will flow in it when it is connected to a 20 V AC supply at the resonant frequency.

Using:

$$f = \frac{1}{2\pi\sqrt{LC}}$$

where $L = 10 \times 10^{-3}$ H and $C = 40 \times 10^{-9}$ F gives:

$$\begin{aligned} f &= \frac{1}{6.28\sqrt{10 \times 10^{-3} \times 40 \times 10^{-9}}} \\ &= \frac{0.159}{\sqrt{4 \times 10^{-10}}} = \frac{0.159}{2 \times 10^{-5}} \\ &= \frac{0.159}{2 \times 10^{-5}} = 7{,}950 = 7.95\,\text{kHz} \end{aligned}$$

At the resonant frequency the circuit will behave as a pure resistance (recall that the two reactances will be equal but opposite) and thus the supply current at resonance can be determined from:

$$I = \frac{V}{Z} = \frac{V}{R} = \frac{20}{50} = 0.4\,\text{A}$$

1.7.7 Power factor

The power factor in an AC circuit containing resistance and reactance is simply the ratio of true power to apparent power. Hence:

$$\text{Power factor} = \frac{\text{true power}}{\text{apparent power}}$$

The true power in an AC circuit is the power that is actually dissipated as heat in the resistive component. Thus:

$$\text{True power} = I^2R$$

where I is r.m.s. current and R is the resistance. True power is measured in watts (W).

The apparent power in an AC circuit is the power that is apparently consumed by the circuit and is the product of the supply current and supply voltage (which may not be in phase). Note that, unless the voltage and current are in phase (i.e. $\phi = 0°$), the apparent power *will not* be the same as the power which is actually dissipated as heat. Hence:

$$\text{Apparent power} = IV$$

where I is r.m.s. current and V is the supply voltage. To distinguish apparent power from true power, apparent power is measured in volt-amperes (VA).

Now since $V = IZ$ we can re-arrange the apparent power equation as follows:

$$\text{Apparent power} = IV = I \times IZ = I^2Z$$

Now returning to our original equation:

$$\begin{aligned} \text{Power factor} &= \frac{\text{true power}}{\text{apparent power}} \\ &= \frac{I^2R}{IV} = \frac{I^2R}{I \times IZ} = \frac{I^2R}{I^2Z} = \frac{R}{Z} \end{aligned}$$

From the *impedance triangle* shown earlier in Fig. 1.43, we can infer that:

$$\text{Power factor} = \frac{R}{Z} = \cos\phi$$

Example 1.29

An AC load has a power factor of 0.8. Determine the true power dissipated in the load if it consumes a current of 2 A at 110 V.

Now since:

$$\text{Power factor} = \cos\phi = \frac{\text{true power}}{\text{apparent power}}$$

True power = power factor × apparent power = power factor × VI. Thus:

$$\text{True power} = 0.8 \times 2 \times 110 = 176\,\text{W}$$

Example 1.30

A coil having an inductance of 150 mH and resistance of 250 Ω is connected to a 115 V 400 Hz AC supply. Determine:

(a) the power factor of the coil
(b) the current taken from the supply
(c) the power dissipated as heat in the coil.

(a) First we must find the reactance of the inductor, X_L, and the impedance, Z, of the coil at 400 Hz.

$$X_L = 2\pi \times 400 \times 150 \times 10^{-3} = 376\,\Omega$$

and

$$Z = \sqrt{R^2 + X_L{}^2} = \sqrt{250^2 + 376^2} = 452\,\Omega$$

We can now determine the power factor from:

$$\text{Power factor} = \frac{R}{Z} = \frac{250}{452} = 0.553$$

(b) The current taken from the supply can be determined from:

$$I = \frac{V}{Z} = \frac{115}{452} = 0.254\,\text{A}$$

(c) The power dissipated as heat can be found from:

$$\begin{aligned}\text{True power} &= \text{power factor} \times VI\\ &= 0.553 \times 115 \times 0.254\\ &= 16.15\,\text{W}\end{aligned}$$

Key point

In an AC circuit the power factor is the ratio of true power to apparent power. The power factor is also the cosine of the phase angle between the supply current and supply voltage.

Test your understanding 1.9

1. Determine the reactance of a 60 mH inductor at (a) 20 Hz and (b) 4 kHz.
2. Determine the reactance of a 220 nF capacitor at (a) 400 Hz and (b) 20 kHz.
3. A 0.5 μF capacitor is connected to a 110 V 400 Hz supply. Determine the current flowing in the capacitor.
4. A resistor of 120 Ω is connected in series with a capacitive reactance of 160 Ω. Determine the impedance of the circuit and the current flowing when the circuit is connected to a 200 V AC supply.
5. A capacitor or 2 μF is connected in series with a 100 Ω resistor across a 24 V 400 Hz AC supply. Determine the current that will be supplied to the circuit and the voltage that will appear across each component.
6. An 80 mH coil has a resistance of 10 Ω. Calculate the current flowing when the coil is connected to a 250 V 50 Hz supply.
7. Determine the phase angle and power factor for Question 6.
8. An AC load has a power factor of 0.6. If the current supplied to the load is 5 A and the supply voltage is 110 V determine the true power dissipated by the load.
9. An AC load comprises a 110 Ω resistor connected in parallel with a 20 μF capacitor. If the load is connected to a 220 V 50 Hz supply, determine the apparent power supplied to the load and its power factor.
10. A filter consists of a 2 μF capacitor connected in series with a 50 mH inductor. At what frequency will the filter be resonant?

1.7.8 Transformers

Transformers provide us with a means of stepping-up or stepping down an AC voltage. For a **step-up transformer**, the output (or **secondary**) voltage will be greater than the input (or **primary**) whilst for a **step-down transformer** the secondary voltage will be less than the primary voltage. Since the primary and secondary power must be the same (no increase in power is possible), an increase in secondary voltage can only be achieved at the expense of a corresponding reduction in secondary current, and vice versa (in fact, the secondary power will be very slightly less than the primary power due to losses within the transformer).

The principle of the transformer is illustrated in Fig. 1.46. The primary and secondary windings are wound on a common low-reluctance magnetic core consisting of a number of steel laminations. All of the alternating flux generated by the primary winding is therefore coupled into the secondary winding (very

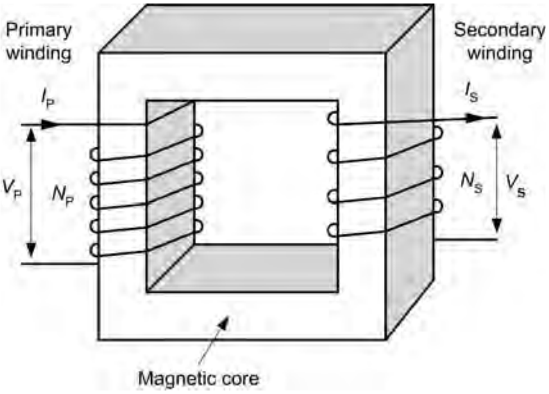

Figure 1.46 The principle of the transformer

little flux escapes due to leakage). A sinusoidal current flowing in the primary winding produces a sinusoidal flux within the transformer core.

At any instant the flux, Φ, in the transformer core is given by the equation:

$$\Phi = \Phi_{max} \sin(\omega t)$$

where Φ_{max} is the maximum value of flux (in Wb), f is the frequency of the applied current, and t is the time in seconds. You might like to compare this equation with the one that you met earlier for a sine wave voltage.

The r.m.s. value of the primary voltage (V_P) is given by:

$$V_P = 4.44 f N_P \Phi_{max}$$

Similarly, the r.m.s. value of the secondary voltage (V_P) is given by:

$$V_S = 4.44 f N_S \Phi_{max}$$

From these two relationships (and since the same magnetic flux appears in both the primary and secondary windings) we can infer that:

$$\frac{V_P}{V_S} = \frac{N_P}{N_S}$$

Furthermore, assuming that no power is lost in the transformer (i.e. as long as the primary and secondary powers are the same) we can conclude that:

$$\frac{I_P}{I_S} = \frac{N_S}{N_P}$$

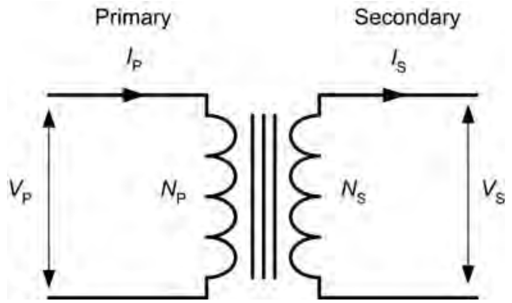

Figure 1.47 Transformer turns and voltages

The ratio of primary turns to secondary turns (N_P/N_S) is known as the **turns ratio**. Furthermore, since ratio of primary voltage to primary turns is the same as the ratio of secondary turns to secondary voltage, we can conclude that, for a particular transformer:

$$\text{Turns-per-volt (t.p.v.)} = \frac{V_P}{N_P} = \frac{V_S}{N_S}$$

The turns-per-volt rating can be quite useful when it comes to designing transformers with multiple secondary windings.

Example 1.31

A transformer has 2000 primary turns and 120 secondary turns. If the primary is connected to a 220 V AC mains supply, determine the secondary voltage.

Since $\frac{V_P}{V_S} = \frac{N_P}{N_S}$ we can conclude that:

$$V_S = \frac{V_P N_S}{N_P} = \frac{220 \times 120}{2000} = 13.2\,\text{V}$$

Example 1.32

A transformer has 1200 primary turns and is designed to operate with a 110 V AC supply. If the transformer is required to produce an output of 10 V, determine the number of secondary turns required.

Since $\frac{V_P}{V_S} = \frac{N_P}{N_S}$ we can conclude that:

$$N_S = \frac{N_P V_S}{V_P} = \frac{1200 \times 10}{110} = 109.1$$

Example 1.33

A transformer has a turns-per-volt rating of 1.2. How many turns are required to produce secondary outputs of (a) 50 V and (b) 350 V?

Here we will use N_S = turns-per-volt × V_S.

(a) In the case of a 50 V secondary winding:

$$N_S = 1.5 \times 50 = 75\,\text{turns}$$

(b) In the case of a 350 V secondary winding:

$$N_S = 1.5 \times 350 = 525\,\text{turns}$$

Example 1.34

A transformer has 1200 primary turns and 60 secondary turns. Assuming that the transformer is loss-free, determine the primary current when a load current of 20 A is taken from the secondary.

Since $\frac{I_S}{I_P} = \frac{N_P}{N_S}$ we can conclude that:

$$I_P = \frac{I_S N_S}{N_P} = \frac{20 \times 60}{1200} = 1\text{ A}$$

The output voltage produced at the secondary of a real transformer falls progressively as the load imposed on the transformer increases (i.e. as the secondary current increases from its no-load value). The **voltage regulation** of a transformer is a measure of its ability to keep the secondary output voltage constant over the full range of output load currents (i.e. from **no-load** to **full-load**) at the same power factor. This change, when divided by the no-load output voltage, is referred to as the **per-unit regulation** for the transformer.

Example 1.35

A transformer produces an output voltage of 110 V under no-load conditions and an output voltage of 101 V when a full load is applied. Determine the per-unit regulation.

The per-unit regulation can be determined from:

$$\text{Per-unit regulation} = \frac{V_{S(\text{no-load})} - V_{S(\text{full-load})}}{V_{S(\text{no-load})}}$$
$$= \frac{110 - 101}{110} = 0.081 \text{ (or 8.1\%)}$$

Most transformers operate with very high values of efficiency. Despite this, in high power applications the losses in a transformer cannot be completely neglected. Transformer losses can be divided into two types of loss:

- losses in the magnetic core (often referred to as **iron loss**)
- losses due to the resistance of the coil windings (often referred to as **copper loss**).

Iron loss can be further divided into **hysteresis loss** (energy lost in repeatedly cycling the magnetic flux in the core backwards and forwards) and **eddy current loss** (energy lost due to current circulating in the steel core).

Hysteresis loss can be reduced by using material for the magnetic core that is easily magnetized and has a very high permeability (see Fig. 1.48 – note that energy loss is proportional to the area inside the *B–H* curve). Eddy current loss can be reduced by laminating the core (using E and I laminations, for example) and also ensuring that a small gap is present. These laminations and gaps in the core help to ensure that there is no closed path for current to flow. Copper loss results from the resistance of the coil windings and it can be reduced by using wire of large diameter and low resistivity.

It is important to note that, since the flux within a transformer varies only slightly between the no-load and full-load conditions, iron loss is substantially constant regardless of the load actually imposed on a transformer. Copper loss, on the other hand, is zero when a transformer is under no-load conditions and rises to a maximum at full-load.

The efficiency of a transformer is given by:

$$\text{Efficiency} = \frac{\text{output power}}{\text{input power}} \times 100\%$$

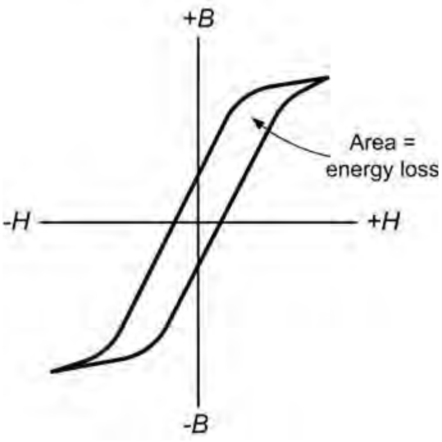

Figure 1.48 Hysteresis curves and energy loss

from which

$$\text{Efficiency} = \frac{\text{input power} - \text{losses}}{\text{input power}} \times 100\%$$

and

$$\text{Efficiency} = 1 - \frac{\text{losses}}{\text{input power}} \times 100\%$$

As we have said, the losses present are attributable to iron loss and copper loss but the copper loss appears in both the primary and the secondary windings. Hence:

$$\text{Efficiency} = 1 - \frac{\begin{array}{l}\text{iron loss} + \text{primary}\\ \text{copper loss} + \text{secondary}\\ \text{copper loss}\end{array}}{\text{input power}} \times 100\%$$

Example 1.36

A transformer rated at 500 VA has an iron loss of 3 W and a full-load copper loss (primary plus secondary) of 7 W. Calculate the efficiency of the transformer at 0.8 power factor.

The input power to the transformer will be given by the product of the apparent power (i.e. the transformer's *VA* rating) and the power factor. Hence:

$$\text{Input power} = 0.8 \times 5000 = 400\,\text{W}$$

Figure 1.49 A selection of transformers for operation at frequencies ranging from 50 Hz to 50 kHz

Now

$$\text{Efficiency} = 1 - \frac{(7+3)}{400} \times 100\% = 97.5\%$$

Key maintenance point

When replacing a transformer it is essential to ensure that the replacement component is correctly rated. The specifications for a transformer usually include the rated primary and secondary voltage and current, the power rating expressed in volt-amperes, VA (this is the maximum power that the transformer can deliver under a given set of conditions), the frequency range for the transformer (note that a transformer designed for operation at 400 Hz will not work at 50 Hz or 60 Hz), and the **per-unit regulation** of the transformer (this is the ability of the transformer to maintain its rated output when under load).

Test your understanding 1.10

1. A transformer has 480 primary turns and 120 secondary turns. If the primary is connected to a 110 V AC supply determine the secondary voltage.
2. A step-down transformer has a 220 V primary and a 24 V secondary. If the secondary winding has 60 turns, how many turns are there on the primary?

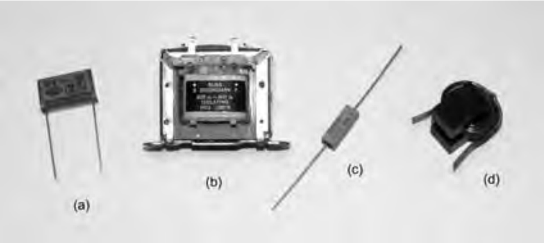

Figure 1.50 See Test your understanding 1.10

3. A transformer has 440 primary turns and 1800 secondary turns. If the secondary supplies a current of 250 mA, determine the primary current (assume that the transformer is loss-free).
4. A transformer produces an output voltage of 220 V under no-load conditions and an output voltage of 208 V when full load is applied. Determine the per-unit regulation.
5. A 1 kVA transformer has an iron loss of 15 W and a full-load copper loss (primary plus secondary) of 20 W. Determine the efficiency of the transformer at 0.9 power factor.
6. Identify the components shown in Fig. 1.50.

1.8 Safety

When working on aircraft electrical and electronic systems, personal safety (both yours and of those around you) should be paramount in everything that you do. Hazards can exist within many circuits – even those that, on the face of it, may appear to be totally safe. Inadvertent misconnection of a supply, incorrect earthing, reverse connection of components, and incorrect fitting can all result in serious hazards to personal safety as a consequence of fire, explosion or the generation of toxic fumes. In addition, there is a need to ensure that your work will not compromise the safety and integrity of the aircraft and not endanger the passengers and crew that will fly in it.

Potential hazards can be easily recognized and it is well worth making yourself familiar with them but perhaps the most important point to make is that electricity acts very quickly and you should always think carefully before working on circuits where mains or high voltages (i.e. those over 50 V or so) are present. Failure to observe this simple precaution can result in the very real risk of electric shock.

Voltages in many items of electronic equipment, including all items which derive their power from the aircraft's 400 Hz AC supply, are at a level which can cause sufficient current flow in the body to disrupt normal operation of the heart. The threshold will be even lower for anyone with a defective heart. Bodily contact with AC supplies and other high-voltage circuits can thus be lethal.

The most critical path for electric current within the body (i.e. the one that is most likely to stop the heart) is that which exists from one hand to the other. The hand-to-foot path is also dangerous but somewhat less dangerous than the hand-to-hand path. So, before you start to work on an item of electronic equipment, it is essential not only to switch off but to disconnect the equipment at the mains by removing the mains plug. If you have to make measurements or carry out adjustments on a piece of working (or 'live') equipment, a useful precaution is that of using one hand only to perform the adjustment or to make the measurement. Your 'spare' hand should be placed safely away from contact with anything metal (including the chassis of the equipment which may, or may not, be earthed).

The severity of electric shock depends upon several factors including the magnitude of the current, whether it is alternating or direct current, and its precise path through the body. The magnitude of the current depends upon the voltage which is applied and the resistance of the body. The electrical energy developed in the body will depend upon the time for which the current flows. The duration of contact is also crucial in determining the eventual physiological effects of the shock. As a rough guide, and assuming that the voltage applied is from the aircraft's 400 Hz AC supply, the following effects are typical:

Current	*Physiological effect*
less than 1 mA	Not usually noticeable
1 mA to 2 mA	Threshold of perception (a slight tingle may be felt)
2 mA to 4 mA	Mild shock (effects of current flow are felt)
4 mA to 10 mA	Serious shock (shock is felt as pain)
10 mA to 20 mA	Motor nerve paralysis may occur (unable to let go)
20 mA to 50 mA	Respiratory control inhibited (breathing may stop)
more than 50 mA	Ventricular fibrillation of heart muscle (heart failure)

Other hazards

Various other hazards, apart from electric shock, exist within the environment of an aircraft when work is being carried out on electrical and electronic systems. For example, accidental movement of the aircraft's spoilers can result in injury and/or damage to equipment. Whenever power sources are changed or when switches or circuit-breakers are opened that may cause movement of spoilers it is essential to ensure that the spoilers are deactivated or that all persons and equipment are removed from the vicinity.

Key maintenance point

The figures quoted in the table are provided as a guide – there have been cases of lethal shocks resulting from contact with much lower voltages and at relatively small values of current. The upshot of all this is simply that any potential in excess of 50V should be considered dangerous. Lesser potentials may, under unusual circumstances, also be dangerous. As such, it is wise to get into the habit of treating all electrical and electronic systems with great care.

Key maintenance point

It is essential to remove electrical power from an aircraft before removing or installing components in the power panels. Failure to observe this precaution can result in electric shock as well as damage to components and equipment.

1.9 Multiple choice questions

1. A material in which there are no free charge carries is known as:
   (a) a conductor
   (b) an insulator
   (c) a semiconductor.

2. Conventional current flow is:
   (a) always from negative to positive
   (b) in the same direction as electron movement
   (c) in the opposite direction to electron movement.

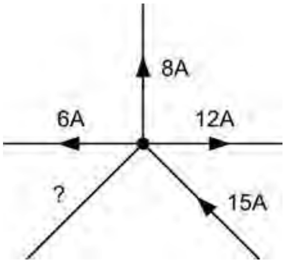

Figure 1.51 See Question 4

3. Which one of the following gives the symbol and abbreviated units for electric charge?
   (a) Symbol, *Q*; unit, C
   (b) Symbol, *C*; unit, F
   (c) Symbol, *C*; unit, V.

4. The unknown current shown in Fig. 1.51 will be:
   (a) 11 A flowing towards the junction
   (b) 17 A flowing away from the junction
   (c) 41 A flowing away from the junction.

5. The power factor in an AC circuit is the same as:
   (a) the sine of the phase angle
   (b) the cosine of the phase angle
   (c) the tangent of the phase angle.

6. An aircraft cabin has 110 passenger reading lamps each rated at 10W, 28V. What is the maximum load current imposed by these lamps?
   (a) 25.5 A
   (b) 39.3 A
   (c) 308 A.

7. Two isolated charges have dissimilar polarities. The force between them will be:
   (a) a force of attraction
   (b) a force of repulsion
   (c) zero.

8. The relationship between voltage, *V*, current, *I*, and resistance, *R*, for a resistor is:
   (a) $V = IR$
   (b) $V = R/I$
   (c) $V = IR^2$.

9. A potential difference of 7.5V appears across a 15 Ω resistor. Which one of the following gives the current flowing:
   (a) 0.25 A
   (b) 0.5 A
   (c) 2 A.

10. An aircraft supply has an RMS value of 115V. Which one of the following gives the approximate peak value of the supply voltage?
    (a) 67.5V
    (b) 115V
    (c) 163V.

11. An AC waveform has a period of 4 ms. Which one of the following gives its frequency?
    (a) 25Hz
    (b) 250Hz
    (c) 4kHz.

12. Which one of the following gives the angle between the successive phases of a three-phase supply?
    (a) 60°
    (b) 90°
    (c) 120°.

13. The peak value of current supplied to an aircraft TRU is 28 A.Which one of the following gives the approximate value of RMS current supplied?
    (a) 10 A
    (b) 14 A
    (c) 20 A.

14. A transformer has 2400 primary turns and 600 secondary turns. If the primary is supplied from a 220V AC supply which one of the following gives the resulting secondary voltage?
    (a) 55V
    (b) 110V
    (c) 880V.

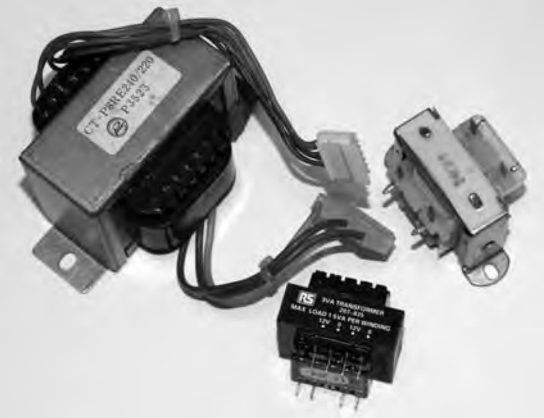

Figure 1.52 See Question 15

15. Identify the components shown in Fig. 1.52.
    (a) resistors
    (b) inductors
    (c) transformers.

Chapter 2

Electronic fundamentals

Electronics underpins all avionic systems, and semiconductor devices such as diodes, transistors and integrated circuits are the basic building blocks of these systems. This chapter provides you with an introduction to the theory and operation of semiconductors and their application in basic electronic circuits. As with Chapter 1, if you are in any doubt as to whether or not you should work through this chapter you can always turn to Section 2.5 on page 56 and see how you get on with the multiple choice questions at the end of the chapter.

2.1 Semiconductor theory

Materials that combine some of the electrical characteristics of conductors with those of insulators are known as **semiconductors**. Common types of semiconductor material are silicon, germanium, selenium and gallium. In the pure state, these materials may have relatively few free electrons to permit the flow of electric current. However, it is possible to add foreign atoms (called **impurity atoms**) to the semiconductor material in order to modify the properties of the semiconductor and allow it to conduct electricity.

You should recall that an atom contains both **negative charge carriers** (electrons) and **positive charge carriers** (protons). Electrons each carry a single unit of negative electric charge while protons each exhibit a single unit of positive charge. Since atoms normally contain an equal number of electrons and protons, the net charge present will be zero. For example, if an atom has eleven electrons, it will also contain eleven protons. The end result is that the negative charge of the electrons will be exactly balanced by the positive charge of the protons.

Electrons are in constant motion as they orbit around the nucleus of the atom. Electron orbits are organized into **shells**. The maximum number of electrons present in the first shell is two, in the second shell eight, and in the third, fourth and fifth shells it is 18, 32 and 50, respectively. In electronics, only the electron shell furthermost from the nucleus of an atom is important. It is important to note that the movement of electrons between atoms only involves those present in the outer **valence shell**.

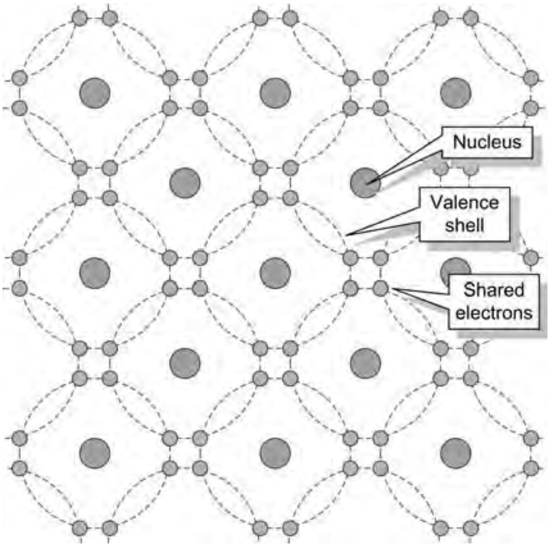

Figure 2.1 Regular lattice structure of a pure semiconductor material

If the valence shell contains the maximum number of electrons possible the electrons are rigidly bonded together and the material has the properties of an insulator. If, however, the valence shell does not have its full complement of electrons, the electrons can be easily detached from their orbital bonds, and the material has the properties associated with an electrical conductor.

In its pure state, silicon is an insulator because the covalent bonding rigidly holds all of the electrons leaving no free (easily loosened) electrons to conduct current. If, however, an atom of a different element (i.e. an **impurity**) is introduced that has five electrons in its valence shell, a surplus electron will be present (see Fig. 2.2). These free electrons become available for use as charge carriers and they can be made

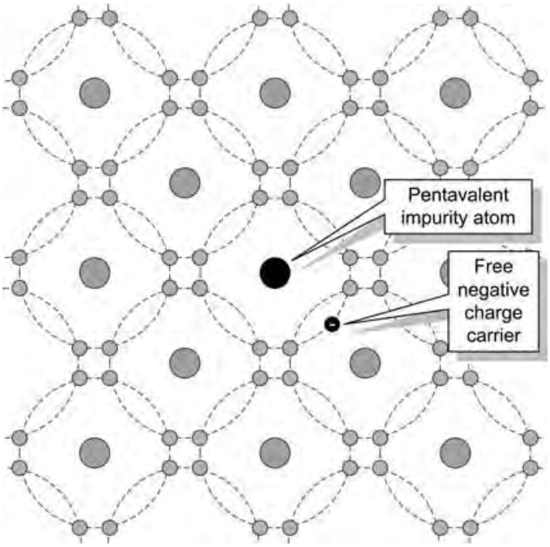

Figure 2.2 Effect of introducing a pentavalent impurity

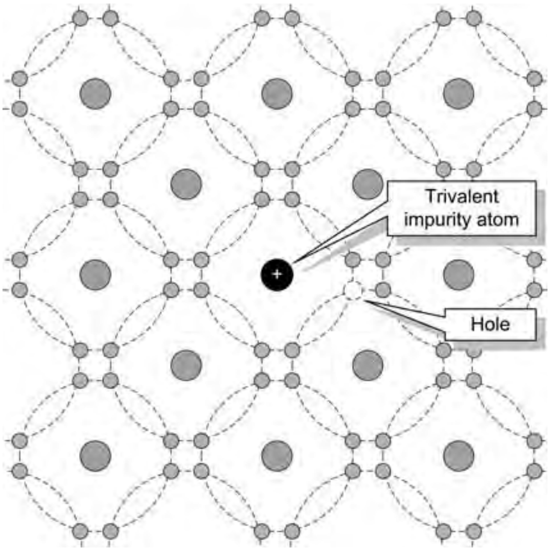

Figure 2.3 Effect of introducing a trivalent impurity

to move through the lattice by applying an external potential difference to the material.

Similarly, if the impurity element introduced into the pure silicon lattice has three electrons in its valence shell, the absence of the fourth electron needed for proper covalent bonding will produce a number of spaces into which electrons can fit (see Fig. 2.3). These gaps are referred to as **holes**. Once again, current will flow when an external potential difference is applied to the material.

Regardless of whether the impurity element produces surplus electrons or holes, the material will no longer behave as an insulator, neither will it have the properties that we normally associate with a metallic conductor. Instead, we call the material a semiconductor – the term simply indicates that the substance is no longer a good insulator nor a good conductor but has the properties of something between the two! Examples of semiconductors include **germanium** (**Ge**) and **silicon** (**Si**).

The process of introducing an atom of another (impurity) element into the lattice of an otherwise pure material is called **doping**. When the pure material is doped with an impurity with five electrons in its valence shell (i.e. a **pentavalent impurity**) it will become an N-type (i.e. negative type) material. If, however, the pure material is doped with an impurity having three electrons in its valence shell (i.e. a **trivalent impurity**) it will become P-type material (i.e. positive type). N-type semiconductor material contains an excess of negative charge carriers, and P-type material contains an excess of positive charge carriers.

2.1.1 Temperature effects

As explained in Chapter 1, all materials offer some resistance to current flow. In conductors the free electrons, rather than passing unobstructed through the material, collide with the relatively large and solid nuclei of the atoms. As the temperature increases, the nuclei vibrate more energetically, further obstructing the path of the free electrons, causing more frequent collisions. The result is that the resistance of a metal conductor increases with temperature.

Due to the nature of the bonding in insulators, there are no free electrons, except that when thermal energy increases as a result of a temperature increase, a few outer electrons manage to break free from their fixed positions and act as charge carriers. The result is that the resistance of an insulator decreases as temperature increases.

Semiconductors behave in a similar manner to insulators. At absolute zero (−273°C) both types of material act as perfect insulators. However, unlike the insulator, as temperature increases in a semiconductor large numbers of electrons break free to act as charge carriers. Therefore, as temperature

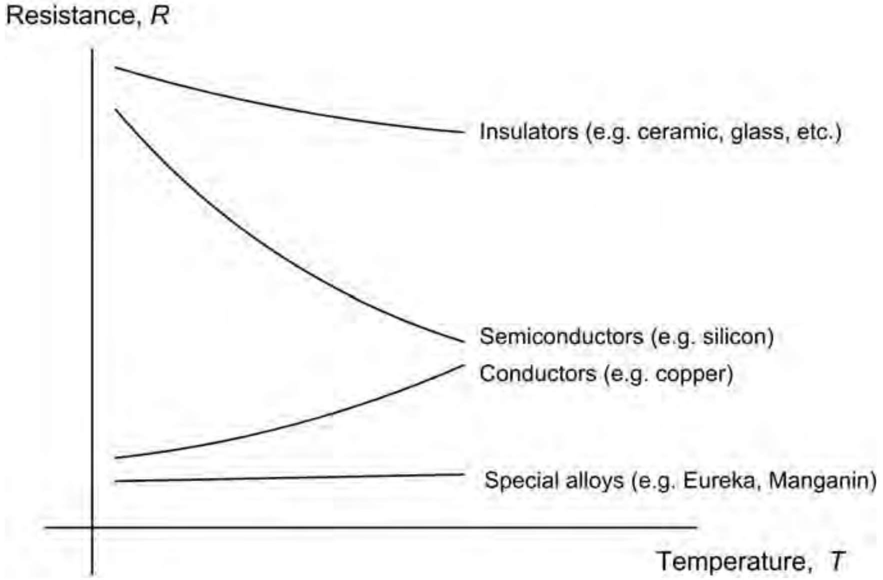

Figure 2.4 Variation of resistance with temperature for various materials

increases, the resistance of a semiconductor decreases rapidly.

By producing special alloys, such as Eureka and Manganin that combine the effects of insulators and conductors, it is possible to produce a material where the resistance remains constant with increase in temperature. Fig. 2.4 shows how the resistance of insulators, semiconductors and conductors change with temperature.

Unlike conventional resistors, the resistance of a **thermistor** is intended to change considerably with temperature. Thermistors are employed in a wide variety of temperature-sensing and temperature-compensating applications. Two basic types of thermistors are available: negative temperature coefficient (NTC) and positive temperature coefficient (PTC).

Typical NTC thermistors have resistances that vary from a few hundred (or thousand) ohms at 25°C to a few tens (or hundreds) of ohms at 100°C. PTC thermistors, on the other hand, usually have a resistance–temperature characteristic that remains substantially flat (usually at around 100 Ω) over the range 0°C to around 75°C. Above this, and at a critical threshold temperature (usually in the range 80°C to 120°C) their resistances rise rapidly up to, and beyond, 10 kΩ (see Figs 2.5 and 2.6).

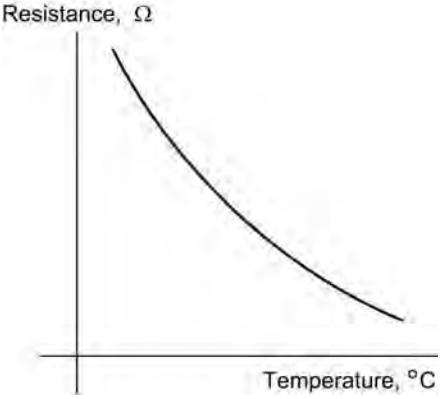

Figure 2.5 NTC thermistor characteristics

Key point

The properties of most semiconductor devices are profoundly affected by temperature. In order to maintain performance within specification, care must be taken to ensure that temperature variations are minimized. In some applications, additional thermal protection may be required to reduce the risk of **thermal runaway** (a condition in which uncontrolled temperature rise may result in failure of one or more semiconductor devices).

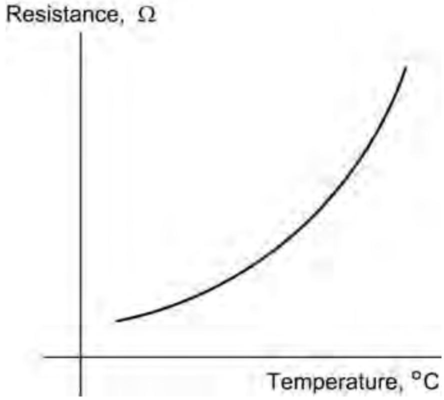

Figure 2.6 PTC thermistor characteristics

Key point

Thermistors provide us with a means of sensing temperature changes. The resistance of an NTC thermistor falls with temperature whilst that of a PTC thermistor increases with temperature.

Key maintenance point

Many semiconductor devices are may be fitted with dedicated cooling systems such as heat sinks, heat pipes and fans. These components provide cooling by means of a combination of thermal conduction and convection and they are crucial to the correct operation of the devices to which they are fitted. It is therefore essential to ensure that cooling systems are correctly fitted and are fully operational whenever maintenance is carried out on the semiconductor devices themselves.

Key maintenance point

Beryllium oxide (BeO_2) is used in the construction of some power semiconductors. This material is highly toxic and special care should be taken when handling and disposing of any semiconductors that use it. It is essential to refer to semiconductor manufacturers' recommendations concerning the fitting, handling and disposal of these devices.

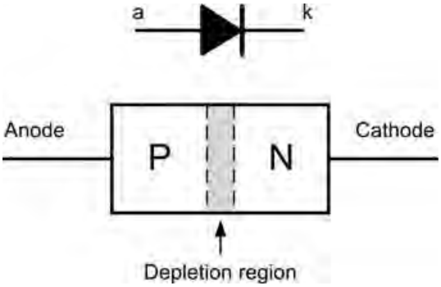

Figure 2.7 A P-N junction diode

Test your understanding 2.1

Explain why thermal runaway is undesirable and why cooling may be necessary in conjunction with semiconductor devices.

2.2 Diodes

When a junction is formed between N-type and P-type semiconductor materials, the resulting device is called a **diode**. This component offers an extremely low resistance to current flow in one direction and an extremely high resistance to current flow in the other. This characteristic allows diodes to be used in applications that require a circuit to behave differently according to the direction of current flowing in it. An ideal diode would pass an infinite current in one direction and no current at all in the other direction.

Connections are made to each side of the diode. The connection to the P-type material is referred to as the **anode** while that to the N-type material is called the **cathode**. With no externally applied potential, electrons from the N-type material will cross into the P-type region and fill some of the vacant holes. This action will result in the production of a region either side of the junction in which there are no free charge carriers. This zone is known as the **depletion region**.

If a positive voltage is applied to the anode (see Fig. 2.8), the free positive charge carriers in the P-type material will be repelled and they will move away from the positive potential towards the junction. Likewise, the negative potential applied to the cathode will cause the free negative charge carriers in the N-type material to move away from the negative potential towards the junction.

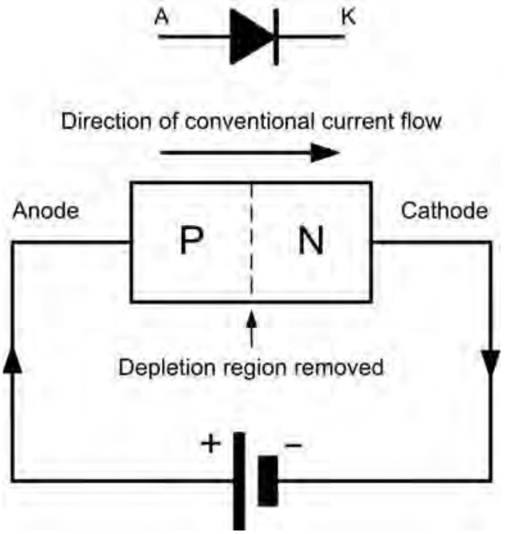

Figure 2.8 A forward-biased P-N junction diode

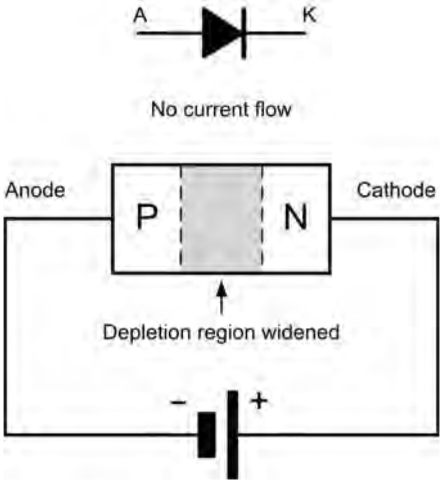

Figure 2.9 A reverse-biased P-N junction diode

When the positive and negative charge carriers arrive at the junction, they will attract one another and combine (recall that that unlike charges attract). As each negative and positive charge carrier combine at the junction, a new negative and positive charge carrier will be introduced to the semiconductor material from the voltage source. As these new charge carriers enter the semiconductor material, they will move towards the junction and combine. Thus, current flow is established and it will continue for as long as the voltage is applied. In this **forward-biased** condition, the diode freely passes current.

If a negative voltage is applied to the anode (see Fig. 2.9), the free positive charge carriers in the P-type material will be attracted and they will move away from the junction. Likewise, the positive potential applied to the cathode will cause the free negative charge carriers in the N-type material to move away from the junction. The combined effect is that the depletion region becomes wider. In this **reverse-biased** condition, the diode passes a negligible amount of current.

Key point

In the freely conducting forward-biased state, the diode acts rather like a closed switch. In the reverse-biased state, the diode acts like an open switch.

2.2.1 Diode characteristics

Typical *I*/*V* characteristics for germanium and silicon diodes are shown in Fig. 2.10. It should be noted from these characteristics that the approximate **forward conduction voltage** for a germanium diode is 0.2 V whilst that for a silicon diode is 0.6 V. This threshold voltage must be high enough to completely overcome the potential associated with the depletion region and force charge carriers to move across the junction.

Diodes are limited by the amount of forward current and reverse voltage they can withstand. This limit is based on the physical size and construction of the diode. In the case of a reverse biased diode, the P-type material is negatively biased relative to the N-type material. In this case, the negative potential at the P-type material attracts the positive carriers, drawing them away from the junction. This leaves the area depleted; virtually no charge carriers exist and therefore current flow is inhibited. If the reverse bias potential is increased above the **maximum reverse voltage** (**V_{RM}**) or **peak inverse voltage** (**PIV**) quoted by the manufacturer, the depletion region may suffer an irreversible breakdown. Typical values of maximum reverse voltage range from as low as 50 V to well over 500 V. Note that reverse breakdown voltage is usually very much higher than the forward threshold voltage. For example, a typical general-purpose diode may be specified as having a forward threshold voltage of 0.6 V and a reverse breakdown voltage of

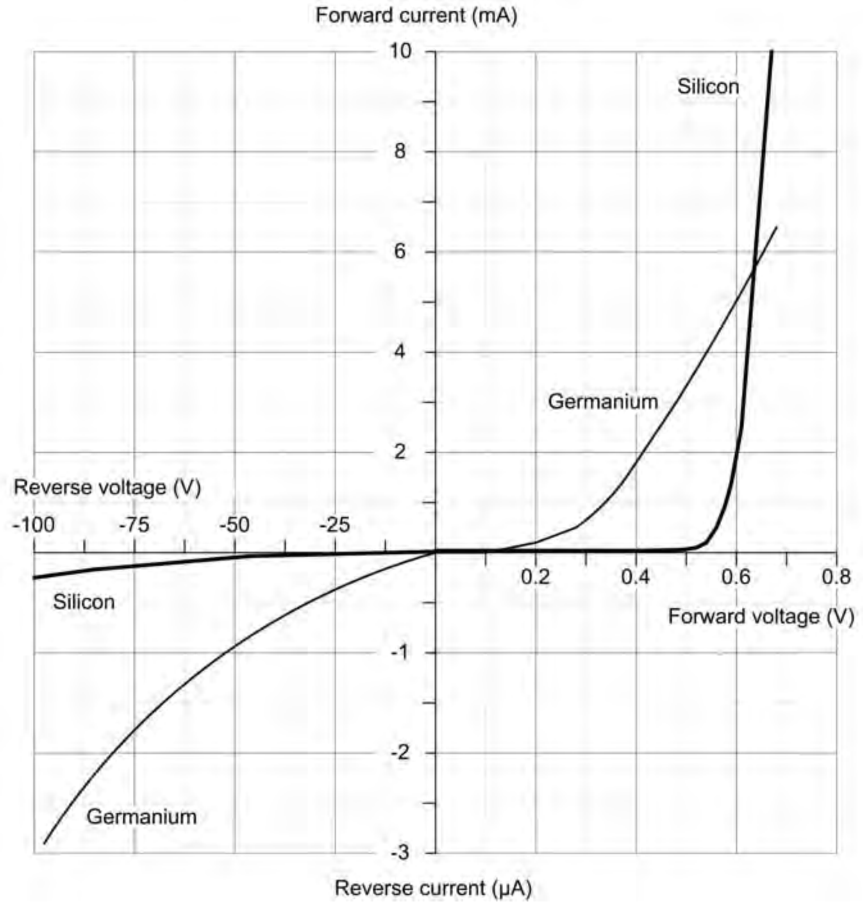

Figure 2.10 Typical *I*/*V* characteristics for germanium and silicon diodes

200V. If the latter is exceeded, the diode may suffer irreversible damage,

---

Key point

The forward voltage for a germanium diode is approximately 0.2V whilst that for a silicon diode is approximately 0.6V.

2.2.2 Zener diodes

Zener diodes are heavily doped silicon diodes that, unlike normal diodes, exhibit an abrupt reverse breakdown at relatively low voltages (typically less than

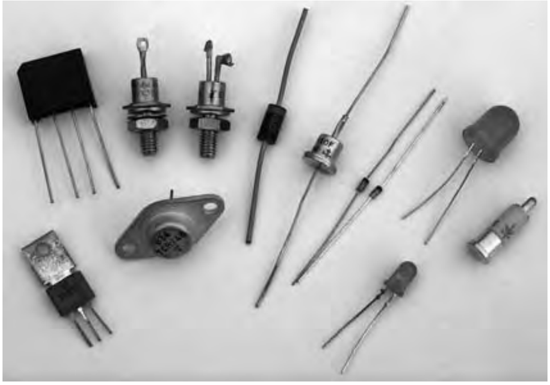

Figure 2.11 Various diodes (including rectifiers, signal diodes, Zener diodes, light-emitting diodes and silicon-controlled rectifiers)

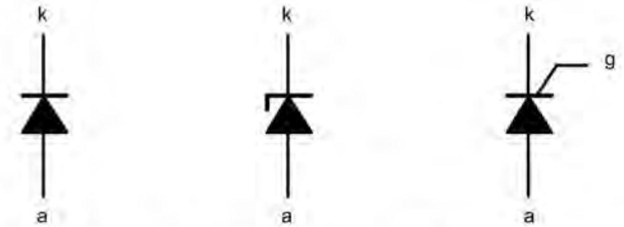

(a) Signal or rectifier diode (b) Zener diode (c) Silicon controlled rectifier (thyristor)

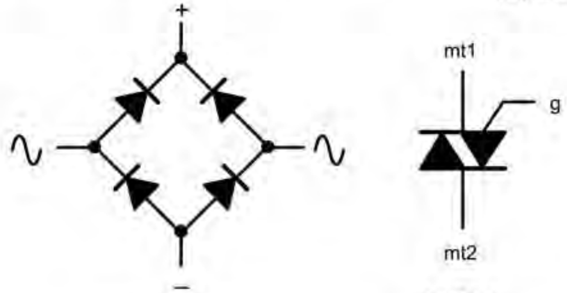

(d) Bridge rectifier (e) Triac

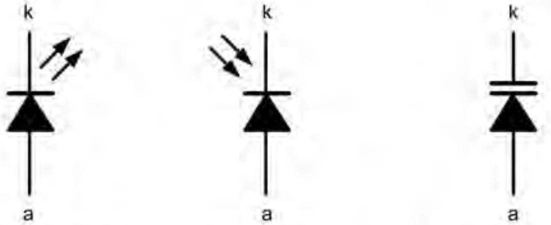

(f) Light emitting diode (g) Photodiode (h) Varactor diode

Figure 2.12 Symbols used for various common types of diode

6 V). A similar effect (avalanche) occurs in less heavily doped diodes. These avalanche diodes also exhibit a rapid breakdown with negligible current flowing below the avalanche voltage and a relatively large current flowing once the avalanche voltage has been reached. For avalanche diodes, this breakdown voltage usually occurs at voltages above 6 V. In practice, however, both types of diode are commonly referred to as Zener diodes. The symbol for a Zener diode was shown earlier in Fig. 2.12, whilst typical Zener diode characteristics are shown in Fig. 2.13.

Whereas reverse breakdown is a highly undesirable effect in circuits that use conventional diodes, it can be extremely useful in the case of Zener diodes where the breakdown voltage is precisely known. When a diode is undergoing reverse breakdown and provided its maximum ratings are not exceeded the voltage appearing across it will remain substantially constant (equal to the nominal Zener voltage) regardless of the current flowing. This property makes the Zener diode ideal for use as a **voltage regulator**.

Zener diodes are available in various families (according to their general characteristics, encapsulations and power ratings) with reverse breakdown (Zener) voltages in the range 2.4 V to 91 V.

A simple **voltage regulator** is shown in Fig. 2.14. The series resistor, R_S is included to limit the Zener current to a safe value when the load is disconnected. When a load (R_L) is connected, the Zener current will fall as current is diverted into the load resistance (it is usual to allow a minimum current of 2 mA to 5 mA in order to ensure that the diode regulates). The

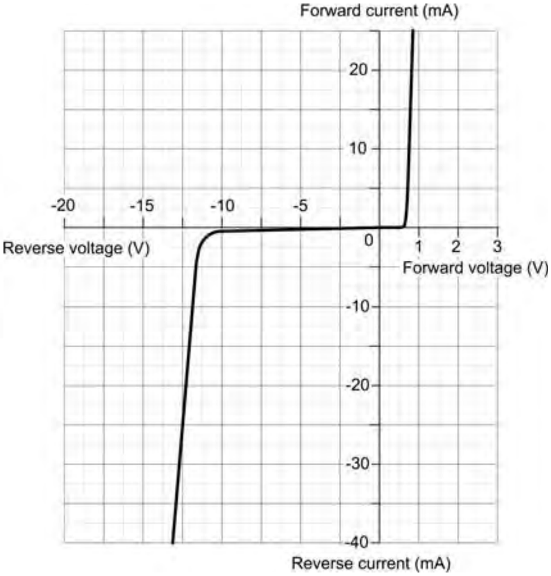

Figure 2.13 Typical Zener diode characteristic

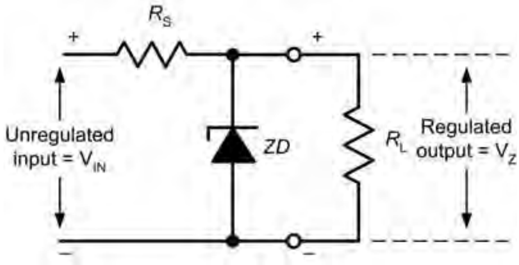

Figure 2.14 A simple Zener diode voltage regulator

output voltage will remain at the Zener voltage (V_Z) until regulation fails at the point at which the potential divider formed by R_S and R_L produces a lower output voltage that is less than V_Z. The ratio of R_S to R_L is thus important.

Key point

Zener diodes begin to conduct heavily when the applied voltage reaches a particular threshold value (known as the Zener voltage). Zener diodes can thus be used to maintain a constant voltage which is often used as a *voltage reference*.

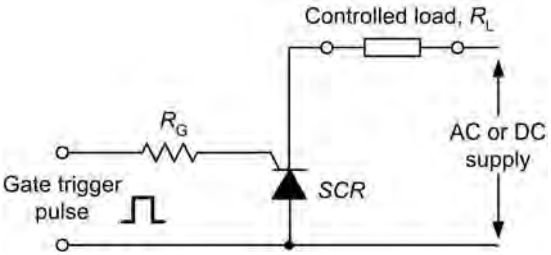

Figure 2.15 Triggering a silicon-controlled rectifier

2.2.3 Silicon-controlled rectifiers

Silicon-controlled rectifiers (or **thyristors**) are three-terminal devices which can be used for switching and AC power control. Silicon-controlled rectifiers can switch very rapidly from a conducting to a non-conducting state. In the off state, the silicon-controlled rectifier exhibits negligible leakage current, while in the on state the device exhibits very low resistance. This results in very little power loss within the silicon-controlled rectifier even when appreciable power levels are being controlled.

Once switched into the conducting state, the silicon-controlled rectifier will remain conducting (i.e. it is latched in the on state) until the forward current is removed from the device. In DC applications this necessitates the interruption (or disconnection) of the supply before the device can be reset into its non-conducting state. Where the device is used with an alternating supply, the device will automatically become reset whenever the main supply reverses. The device can then be triggered on the next half-cycle having correct polarity to permit conduction.

Like their conventional silicon diode counterparts, silicon-controlled rectifiers have anode and cathode connections; control is applied by means of a gate terminal. The symbol for a silicon-controlled rectifier was shown earlier in Fig. 2.12.

In normal use, a silicon-controlled rectifier is triggered into the conducting (on) state by means of the application of a current pulse to the gate terminal (see Fig. 2.15). The effective triggering of a silicon-controlled rectifier requires a gate trigger pulse having a fast rise time derived from a low-resistance source. Triggering can become erratic when insufficient gate current is available or when the gate current changes slowly.

Key point

Silicon-controlled rectifiers (SCR) are diodes that can be triggered into conduction by applying a small current to their gate input. SCRs are able to control large voltages and currents from a relatively small (low-current, low-voltage) signal.

2.2.4 Light-emitting diodes

Light-emitting diodes (LED) can be used as general-purpose indicators and, compared with conventional filament lamps, operate from significantly smaller voltages and currents. LEDs are also very much more reliable than filament lamps. Most LEDs will provide a reasonable level of light output when a forward current of between 5 mA and 20 mA is applied.

Light-emitting diodes are available in various formats with the round types being most popular. Round LEDs are commonly available in the 3 mm and 5 mm (0.2 inch) diameter plastic packages and also in a 5 mm × 2 mm rectangular format. The viewing angle for round LEDs tends to be in the region of 20° to 40°, whereas for rectangular types this is increased to around 100°. The symbol for an LED was shown earlier in Fig. 2.12.

Key point

Light-emitting diodes produce light when a small current is applied to them. They are generally smaller and more reliable than conventional filament lamps and can be used to form larger and more complex displays.

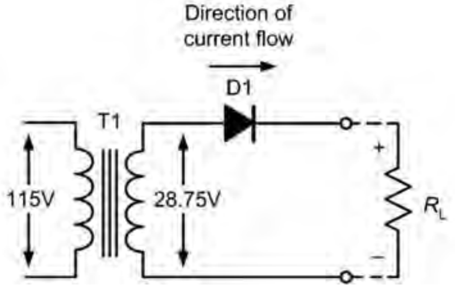

Figure 2.16 A simple half-wave rectifier circuit

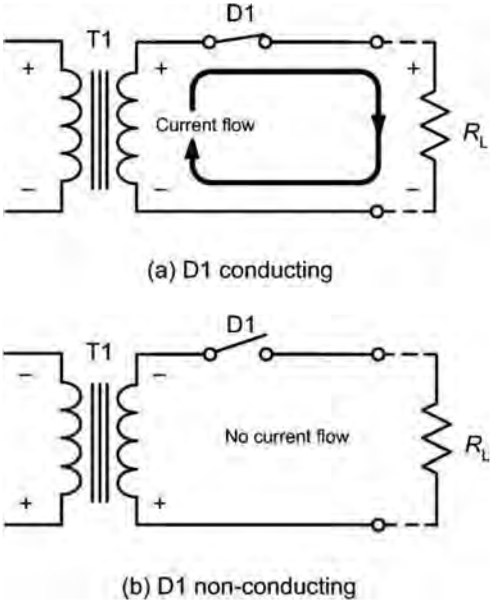

(a) D1 conducting

(b) D1 non-conducting

Figure 2.17 Switching action of the diode in the half-wave rectifier: (a) D1 forward biased, (b) D1 reverse biased

2.2.5 Rectifiers

Semiconductor diodes are commonly used to convert alternating current (AC) to direct current (DC), in which case they are referred to as **rectifiers**. The simplest form of rectifier circuit makes use of a single diode and, since it operates on only either positive or negative half-cycles of the supply, it is known as a half-wave rectifier.

Figure 2.16 shows a simple **half-wave rectifier** circuit. The AC supply at 115 V is applied to the primary of a step-down transformer (T1). The secondary of T1 steps down the 115 V 400 Hz supply to 28.75 V RMS (the turns ratio of T1 will thus be 115/28.75 or 4:1).

Diode D1 will only allow the current to flow in the direction shown (i.e. from anode to cathode). D1 will be forward biased during each positive half-cycle and will effectively behave like a closed switch, as shown in Fig. 2.17(a). When the circuit current tries to flow in the opposite direction, the voltage bias across the diode will be reversed, causing the diode to act like an open switch, as shown in Fig. 2.17(b).

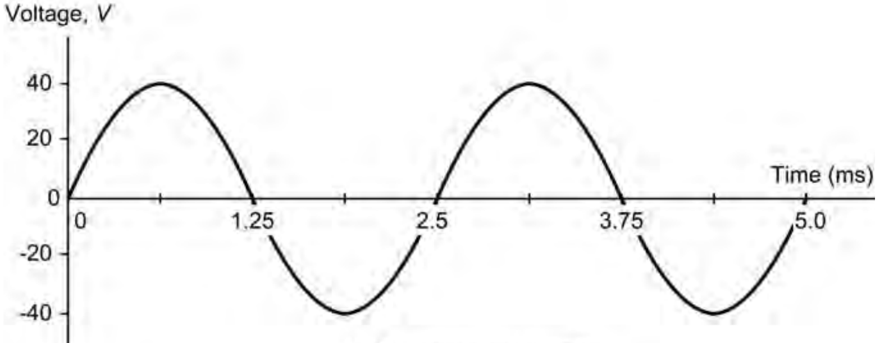

(a) T1 secondary voltage

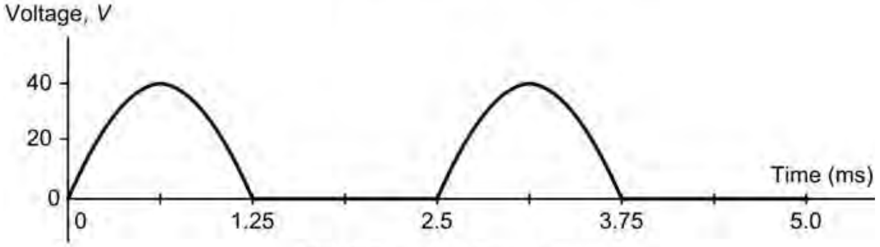

(b) Voltage developed across R_L

Figure 2.18 Waveforms of voltages in the simple half-wave power supply

The switching action of D1 results in a pulsating output voltage, which is developed across the load resistor (R_L). Since the supply is at 400 Hz, the pulses of voltage developed across R_L will also be at 400 Hz even if only half the AC cycle is present. During the positive half-cycle, the diode will drop the 0.6 V forward threshold voltage normally associated with silicon diodes. However, during the negative half-cycle the peak AC voltage will be dropped across D1 when it is reverse biased. This is an important consideration when selecting a diode for a particular application. Assuming that the secondary of T1 provides 28.75 V RMS, the peak voltage output from the transformer's secondary winding will be given by:

$$V_{pk} = 1.414 \times V_{RMS} = 1.414 \times 28.75 \text{ V} = 40.65 \text{ V}$$

The peak voltage applied to Dl will thus be a little over 40 V. The negative half-cycles are blocked by D1 and thus only the positive half-cycles appear across R_L. Note, however, that the actual peak voltage across R_L will be the 40.65 V positive peak being supplied from the secondary on T1, minus the 0.6 V forward threshold voltage dropped by D1. In other words, positive half-cycle pulses having a peak amplitude of almost exactly 40 V will appear across R_L.

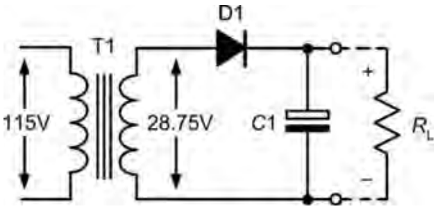

Figure 2.19 Adding a reservoir capacitor to a half-wave rectifier to make a simple DC power supply

Figure 2.19 shows a considerable improvement to the earlier simple rectifier. The capacitor, C1, has been added to ensure that the output voltage remains at, or near, the peak voltage even when the diode is not conducting. When the primary voltage is first applied to T1, the first positive half-cycle output from the secondary will charge C1 to the peak value seen across R_L. Hence C1 charges to 40 V at the peak of

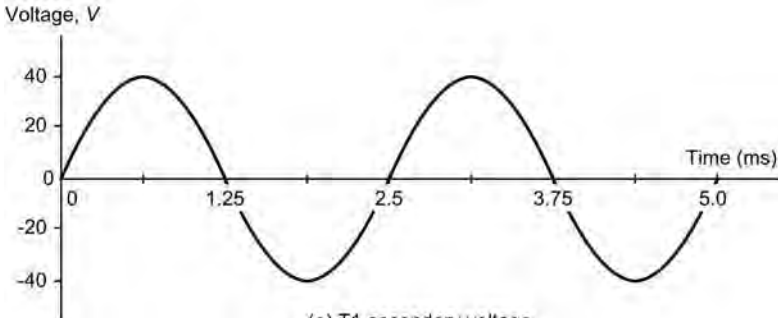

(a) T1 secondary voltage

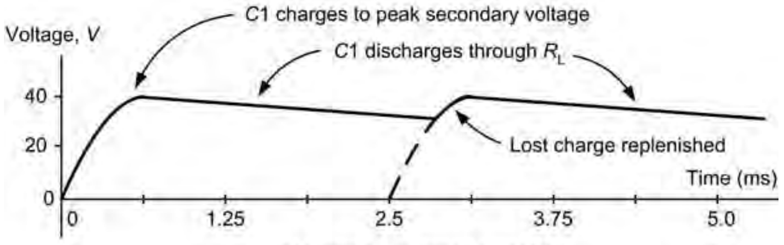

(b) Voltage developed across R_L

Figure 2.20 Effect of adding a reservoir capacitor on the output of the half-wave rectifier

the positive half-cycle. Because C1 and R_L are in parallel, the voltage across R_L will be the same as that developed across C1 (see Fig. 2.19).

The time required for C1 to charge to the maximum (peak) level is determined by the charging circuit time constant (the series resistance multiplied by the capacitance value). In this circuit, the series resistance comprises the secondary winding resistance together with the forward resistance of the diode and the (minimal) resistance of the wiring and connections. Hence C1 charges to 40 V at the peak of the positive half-cycle. Because C1 and R_L are in parallel, the voltage across R_L will be the same as that across C1.

The time required for C1 to discharge is, in contrast, very much greater. The discharge time constant is determined by the capacitance value and the load resistance, R_L. In practice, R_L is very much larger than the resistance of the secondary circuit and hence C1 takes an appreciable time to discharge. During this time, D1 will be reverse biased and will thus be held in its non-conducting state. As a consequence, the only discharge path for C1 is through R_L.

C1 is referred to as a **reservoir capacitor**. It stores charge during the positive half-cycles of secondary voltage and releases it during the negative half-cycles. The circuit shown earlier is thus able to maintain a reasonably constant output voltage across R_L. Even so, C1 will discharge by a small amount during the negative half-cycle periods from the transformer secondary. The figure below shows the secondary voltage waveform together with the voltage developed across R_L with and without C1 present. This gives rise to a small variation in the DC output voltage (known as **ripple**).

Since ripple is undesirable we must take additional precautions to reduce it. One obvious method of reducing the amplitude of the ripple is that of simply increasing the discharge time constant. This can be achieved either by increasing the value of C1 or by increasing the resistance value of R_L. In practice, however, the latter is not really an option because R_L is the effective resistance of the circuit being supplied and we don't usually have the ability to change it! Increasing the value of C1 is a more practical alternative and very large capacitor values (often in excess of 1000 μF) are typical.

The half-wave rectifier circuit is relatively inefficient as conduction takes place only on alternate half-cycles. A better rectifier arrangement would make use of both positive *and* negative half-cycles. These **full-wave rectifier** circuits offer a considerable improvement over their half-wave counterparts. They are not only more efficient but are significantly less demanding in terms of the reservoir and smoothing components. There are

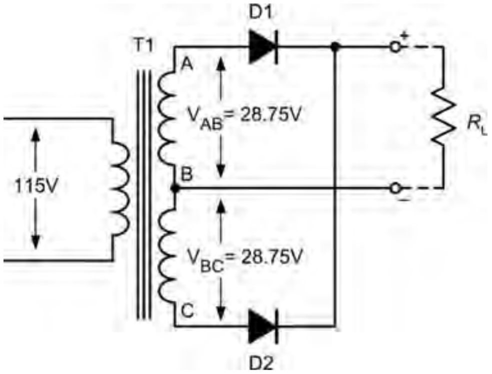

Figure 2.21 A simple bi-phase rectifier circuit

two basic forms of full-wave rectifier: the bi-phase type and the bridge rectifier type.

Figure 2.21 shows a simple **bi-phase rectifier** circuit. The AC supply at 115 V is applied to the primary of a step-down transformer (T1). This has two identical secondary windings, each providing 28.75 V RMS (the turns ratio of T1 will still be 115/28.75 or 4:1 for each secondary winding).

On positive half-cycles, point A will be positive with respect to point B. Similarly, point B will be positive with respect to point C. In this condition D1 will allow conduction (its anode will be positive with respect to its cathode) while D2 will not allow conduction (its anode will be negative with respect to its cathode). Thus D1 alone conducts on positive half-cycles.

On negative half-cycles, point C will be positive with respect to point B. Similarly, point B will be positive with respect to point A. In this condition D2 will allow conduction (its anode will be positive with respect to its cathode) while D1 will not allow conduction (its anode will be negative with respect to its cathode). Thus D2 alone conducts on negative half-cycles.

Figure 2.22 shows the bi-phase rectifier circuit with the diodes replaced by switches. In (a) D1 is shown conducting on a positive half-cycle whilst in (b) D2 is shown conducting on a negative half-cycle of the input. The result is that current is routed through the load *in the same direction* on successive half-cycles. Furthermore, this current is derived alternately from the two secondary windings.

As with the half-wave rectifier, the switching action of the two diodes results in a pulsating output voltage being developed across the load resistor (R_L).

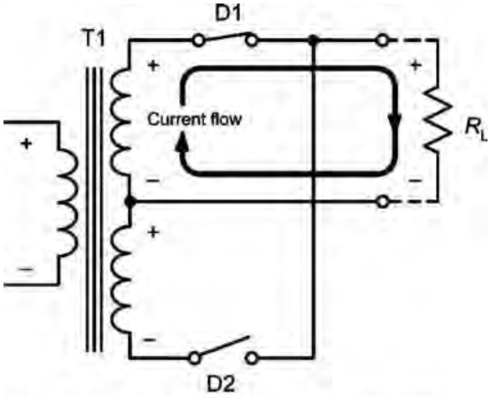

(a) D1 conducting and D2 non-conducting

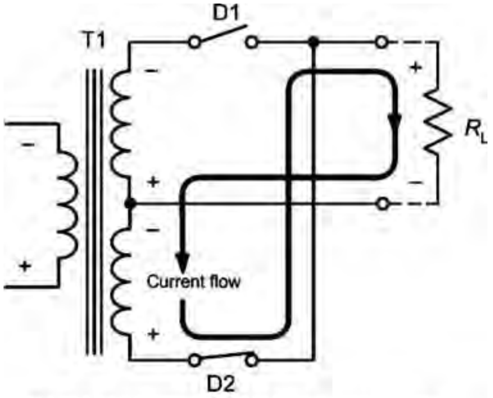

(b) D2 conducting and D1 non-conducting

Figure 2.22 Switching action of the diodes in the bi-phase rectifier: (a) D1 forward biased and D2 reverse biased, (b) D1 reverse biased and D2 forward biased

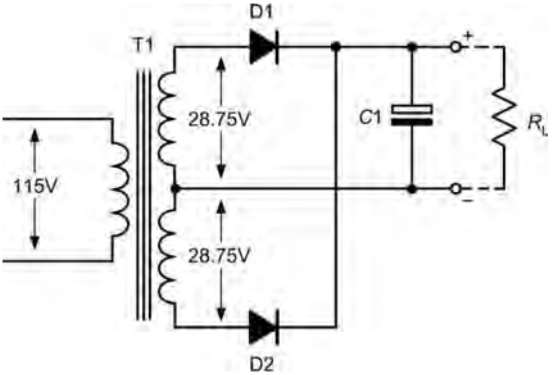

Figure 2.23 Bi-phase power supply with reservoir capacitor

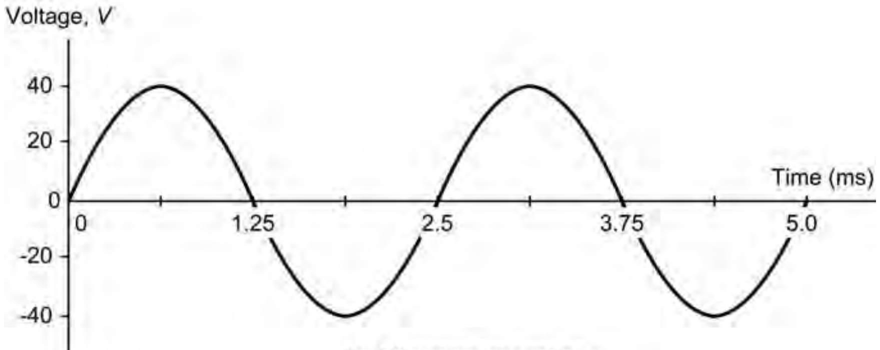

(a) T1 secondary voltage

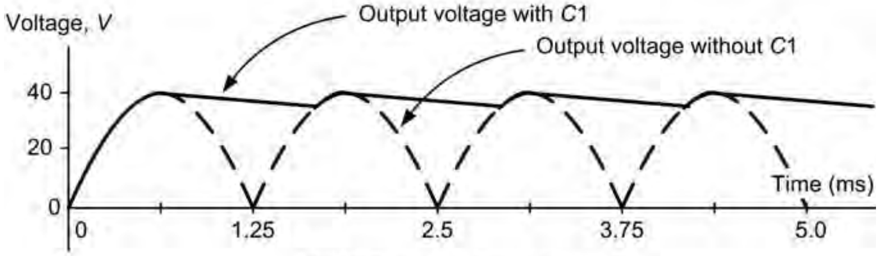

(b) Voltage developed across R_L

Figure 2.24 Waveforms in the bi-phase power supply with and without a reservoir capacitor

However, unlike the half-wave circuit the pulses of voltage developed across R_L will occur at a frequency of 800 Hz (not 400 Hz). This doubling of the ripple frequency allows us to use smaller values of reservoir and smoothing capacitor to obtain the same degree of ripple reduction (recall that the reactance of a capacitor is reduced as frequency increases). As before, the peak voltage produced by each of the secondary windings will be approximately 17 V and the peak voltage across R_L will be about 40 V (i.e. 40.65 V less the 0.6 V forward).

An alternative to the use of the bi-phase circuit is that of using a four-diode **bridge rectifier** in which opposite pairs of diode conduct on alternate half-cycles. This arrangement avoids the need to have two separate secondary windings. A full-wave bridge rectifier arrangement is shown in Fig. 2.25. The 115 V AC supply at 400 Hz is applied to the primary of a step-down transformer (T1). As before, the secondary winding provides 28.75 V RMS (approximately 40 V peak) and has a turns ratio of 4:1. On positive half-cycles, point A will be positive with respect to point B. In this condition Dl and D2 will allow conduction while D3 and D4 will not allow conduction. Conversely, on negative half-cycles, point B will be positive with respect to point A. In this condition D3 and D4 will allow conduction while D1 and D2 will not allow conduction.

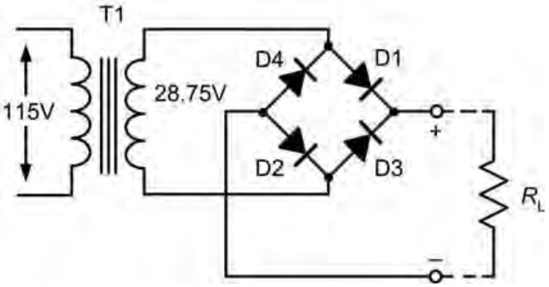

Figure 2.25 Full-wave power supply using a bridge rectifier

As with the bi-phase rectifier, the switching action of the two diodes results in a pulsating output voltage being developed across the load resistor (R_L).

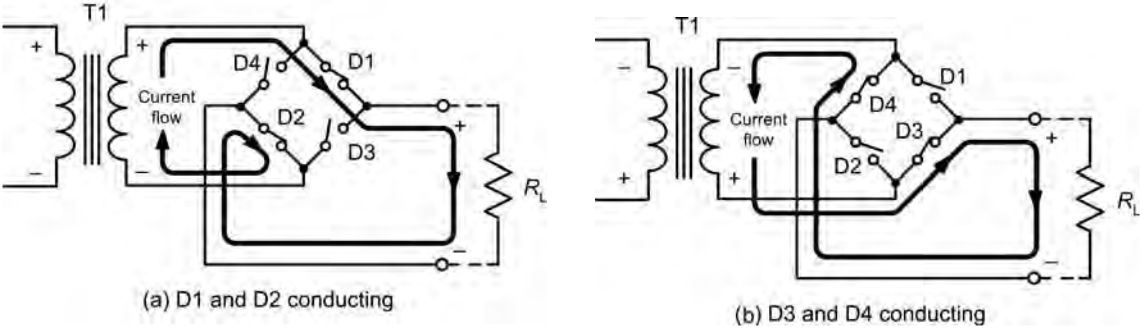

(a) D1 and D2 conducting

(b) D3 and D4 conducting

Figure 2.26 Switching action of the diodes in the full-wave bridge: (a) D1 and D2 forward biased whilst D3 and D4 are reverse biased, (b) D1 and D2 reverse biased whilst D3 and D4 are reverse biased

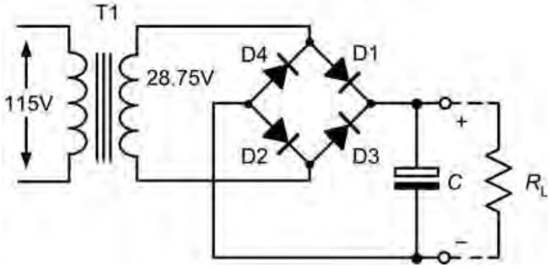

Figure 2.27 Full-wave bridge power supply with reservoir capacitor

Once again, the peak output voltage is approximately 40 V (i.e. 40.65 V less the 2 × 0.6 V forward threshold voltage of the two diodes).

Figure 2.27 shows how a reservoir capacitor (C1) can be added to the basic bridge rectifier circuit in order to ensure that the output voltage remains at, or near, the peak voltage even when the diodes are not conducting. This component operates in exactly the same way as for the bi-phase circuit, i.e. it charges to approximately 40 V at the peak of the positive half-cycle and holds the voltage at this level when the diodes are in their non-conducting states. These voltage waveforms are identical to those that we met earlier for the bi-phase rectifier.

Key point

The forward voltage for a germanium diode is approximately 0.2 V whilst that for a silicon diode is approximately 0.6 V.

Key point

The frequency of the ripple voltage present on the DC output of a half-wave power supply will be the same as the frequency of the AC input. The frequency of the ripple voltage present on the DC output of a full-wave power supply will be double that of the AC input.

Test your understanding 2.2

Explain the following terms when used in relation to a semiconductor diode:

(a) forward bias
(b) reverse bias
(c) forward conduction voltage.

Test your understanding 2.3

Match the following types of diode to each of the applications listed below:

Types of diode

(a) Silicon bridge rectifier
(b) Germanium signal diode
(c) Zener diode
(d) Red LED
(e) Thyristor.

Application

1. A voltage reference for use in a generator control unit

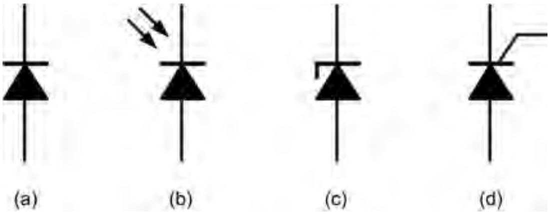

Figure 2.28 See Test your understanding 2.4

2. A power-on indicator for use in an emergency locator transmitter
3. An aircraft battery charger
4. An AC power controller
5. The detector circuit in an HF radio.

Test your understanding 2.4

Identify each of the diode symbols shown in Fig. 2.28.

2.3 Transistors

Typical applications for transistors in aircraft electrical and electronic systems are controlling generator field current, driving lights and warning displays, amplifying signals from sensors and transducers and for use as amplifying devices in cabin interphone and aircraft radio and navigation aids.

Conventional **bipolar junction transistors** (**BJT**) generally comprise **NPN** or **PNP** junctions of either silicon (Si) or germanium (Ge) material. The junctions are produced in a single slice of silicon by diffusing impurities through a photographically reduced mask. Silicon transistors are superior when compared with germanium transistors in the vast majority of applications (particularly at high temperatures) and thus germanium devices are very rarely encountered in modern electronic equipment.

The construction of typical NPN and PNP BJT are shown in Figs 2.29 and 2.30. In order to conduct the heat away from the junction (important in medium and high-power applications) the collector is often connected to the metal case of the transistor.

The symbols and simplified junction models for NPN and PNP transistors are shown in Fig. 2.31. It is important to note that the base region (P-type material in the case of an NPN transistor or N-type material in the case of a PNP transistor) is extremely narrow.

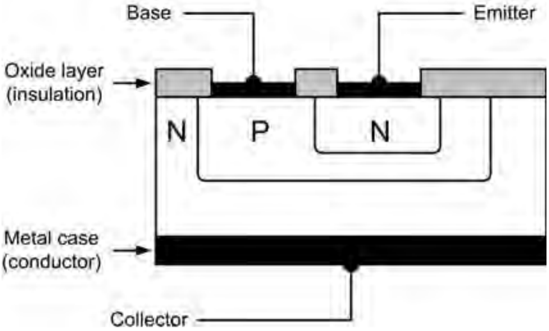

Figure 2.29 Construction of a typical NPN transistor

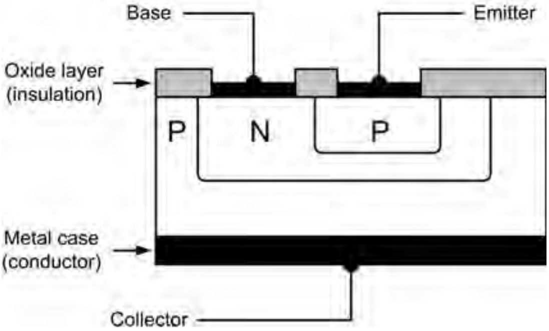

Figure 2.30 Construction of a typical PNP transistor

2.3.1 Bias and current flow

In normal operation (i.e. for operation as a linear amplifier) the base-emitter junction of a transistor is forward biased and the collector base junction is reverse biased. The base region is, however, made very narrow so that carriers are swept across it from emitter to collector so that only a relatively small current flows in the base. To put this into context, the current flowing in the emitter circuit is typically 100 times greater than that flowing in the base. The direction of conventional current flow is from emitter to collector in the case of a PNP transistor, and collector to emitter in the case of an NPN device, as shown in Fig. 2.32.

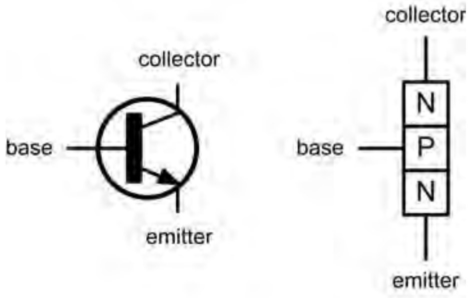

(a) NPN bipolar junction transistor (BJT)

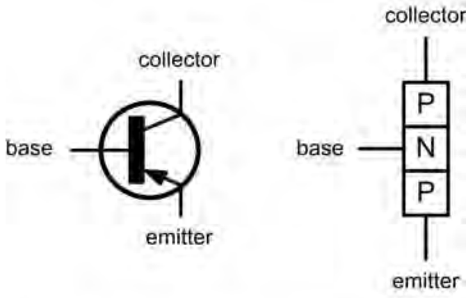

(b) PNP bipolar junction transistor (BJT)

Figure 2.31 NPN and PNP BJT symbols and simplified junction models

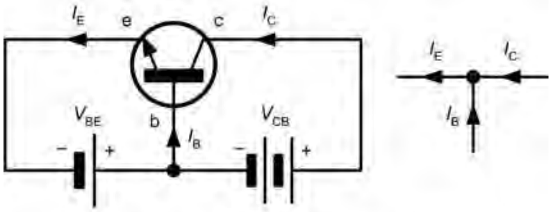

(a) NPN bipolar junction transistor (BJT)

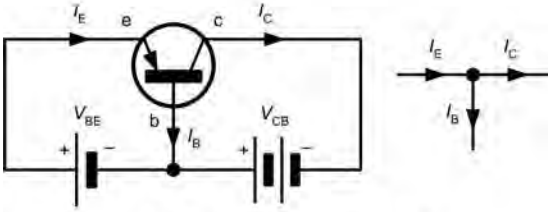

(b) PNP bipolar junction transistor (BJT)

Figure 2.32 Bias voltages and current flow in NPN and PNP bipolar junction transistors

The equation (based on Kirchhoff's current law) that relates current flow in the collector, base, and emitter of a transistor (see Fig. 2.32) is:

$$I_E = I_B + I_C$$

where I_E is the emitter current, I_B is the base current, and I_C is the collector current (all expressed in the same units)

Example 2.1

A transistor operates with a collector current of 2 A and an emitter current of 2.1 A. Determine the value of base current.

Now:

$$I_E = I_B + I_C$$

thus:

$$I_B = I_E - I_C$$

In this case, $I_C = 2\text{A}$ and $I_E = 2.1\text{A}$, thus:

$$I_B = 2.1\text{ A} - 2\text{ A} = 0.1\text{ A} = 100\text{ mA}$$

Test your understanding 2.5

A BJT operates with an emitter current of 1.25 A and a base current of 50 mA. What will the collector current be?

Key point

The emitter current of a transistor is the sum of its base and collector currents.

2.3.2 Transistor characteristics

The characteristics of a bipolar junction transistor are usually presented in the form of a set of graphs relating voltage and current present at the transistors terminals. Figure 2.33 shows a typical **input characteristic** (I_B plotted against V_{BE}) for an NPN bipolar junction transistor operating in **common-emitter mode**. In this mode, the input current is applied to the

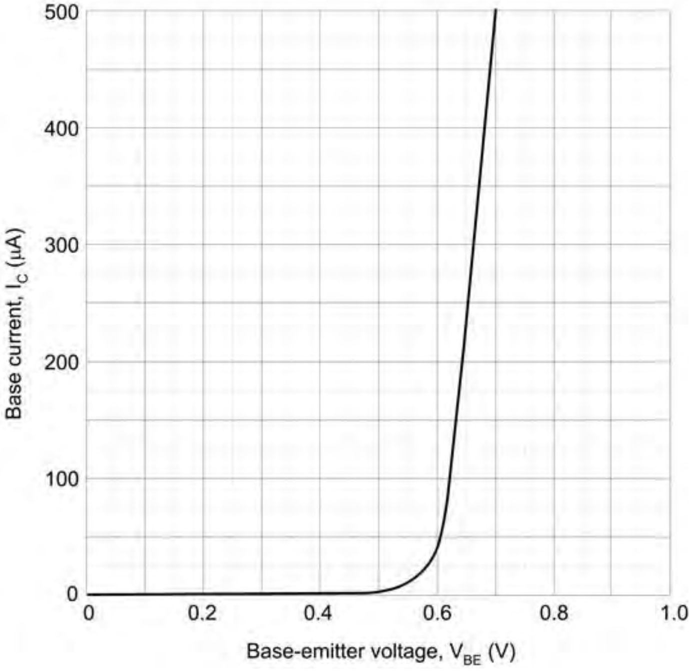

Figure 2.33 Input characteristic (I_B/V_{BE}) for an NPN bipolar junction transistor

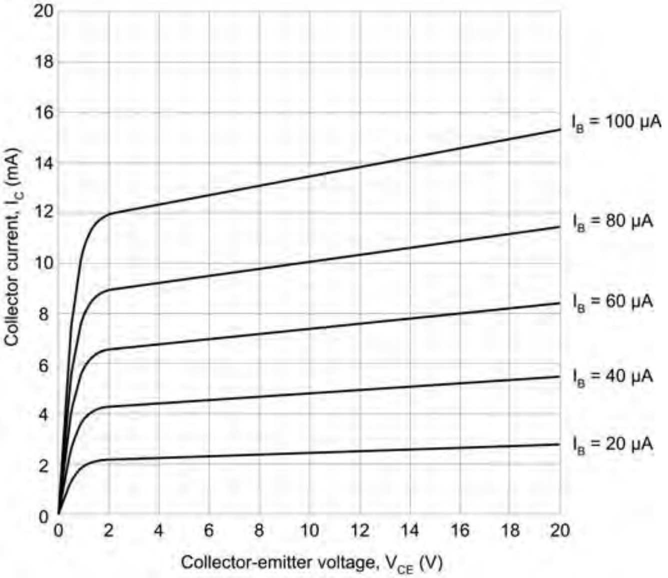

Figure 2.34 Output characteristic (I_C/V_{CE}) for an NPN bipolar junction transistor

base and the output current appears in the collector (the emitter is effectively common to both the input and output circuits).

The input characteristic shows that very little base current flows until the base emitter voltage V_{BE} exceeds 0.6 V. Thereafter, the base current increases rapidly (this characteristic bears a close resemblance to the forward part of the characteristic for a silicon diode).

Figure 2.34 shows a typical set of **output (collector) characteristics** (I_C plotted against V_{CE}) for an

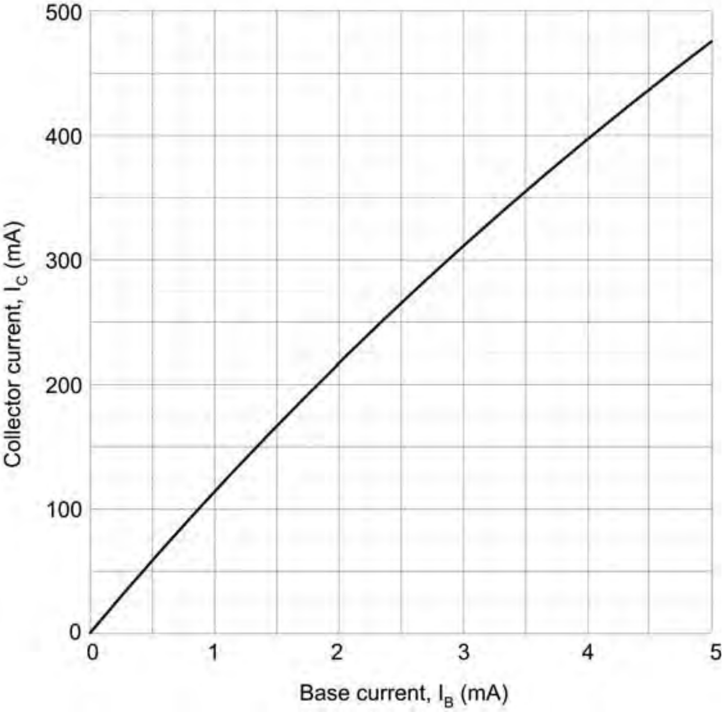

Figure 2.35 Transfer characteristic (I_C/I_B) for an NPN bipolar junction transistor

NPN bipolar transistor. Each curve corresponds to a different value of base current. Note the 'knee' in the characteristic below $V_{CE} = 2$V. Also note that the curves are quite flat. For this reason (i.e. since the collector current does not change very much as the collector-emitter voltage changes) we often refer to this as a *constant current characteristic*.

Figure 2.35 shows a typical **transfer characteristic** for an NPN bipolar junction transistor. Here I_C is plotted against I_B for a small-signal general purpose transistor. The slope of this curve (i.e. the ratio of I_C to I_B) is the **common emitter current gain** of the transistor. We shall explore this further in Section 2.2.6.

2.3.3 Transistor operating configurations

Three basic circuit arrangements are used for transistor **amplifiers** and these are based on the three circuit configurations that we met earlier (i.e. they depend upon which one of the three transistor connections is made common to both the input and the output). In the case of bipolar transistors, the configurations are known as **common emitter**, **common collector** (or emitter follower) and **common base**. Where field effect transistors are used, the corresponding configurations are common source, common drain (or source follower) and common gate.

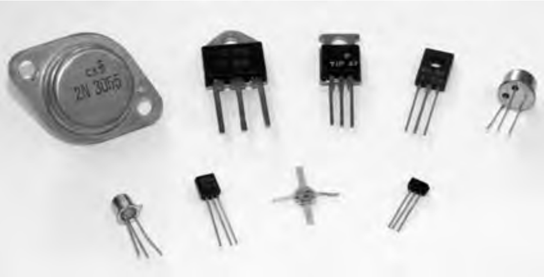

Figure 2.36 Various transistors (including low-frequency, high-frequency, high-voltage, small-signal and power types)

These basic circuit configurations depicted in Fig. 2.38 exhibit quite different performance characteristics, as shown in the table below.

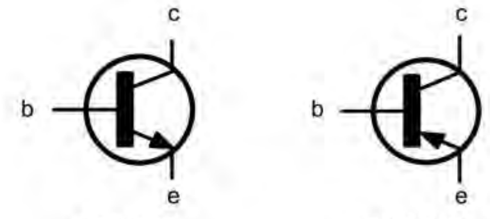

NPN BJT PNP BJT

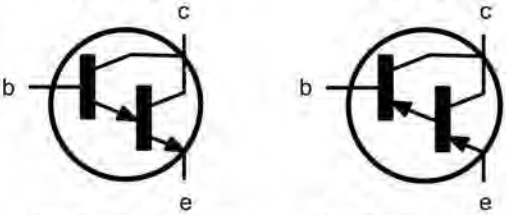

NPN Darlington BJT PNP Darlington BJT

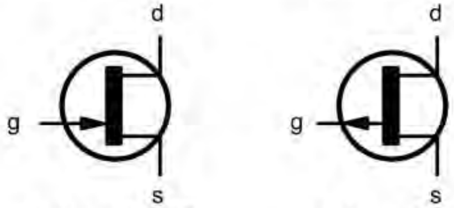

N-channel JFET P-channel JFET

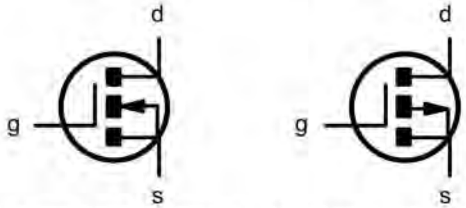

N-channel enhancement mode MOSFET P-channel enhancement mode MOSFET

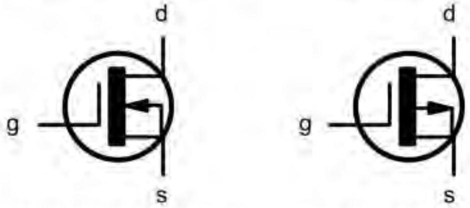

N-channel depletion mode MOSFET P-channel depletion mode MOSFET

Figure 2.37 Symbols used for various types of transistor

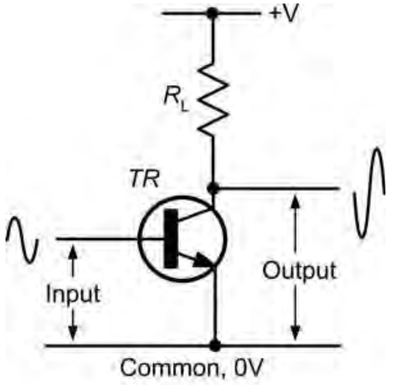
+V
R_L
TR
Input
Output
Common, 0V

(a) Common emitter

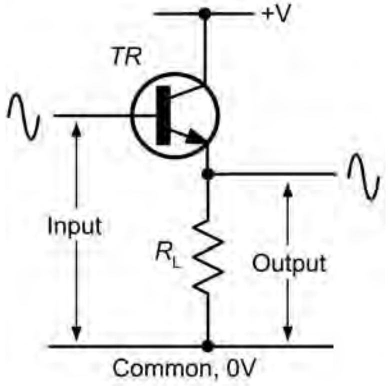
+V
TR
Input
R_L
Output
Common, 0V

(b) Common collector

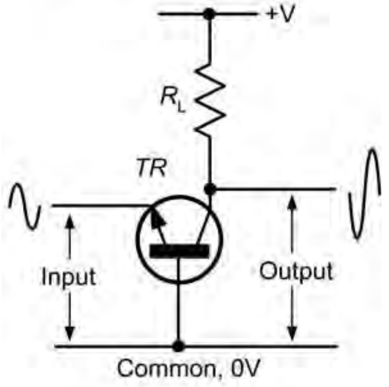
+V
R_L
TR
Input
Output
Common, 0V

(c) Common base

Figure 2.38 BJT circuit configurations

Parameter	*Common emitter*	*Common collector*	*Common base*
Voltage gain	medium/high (40)	unity (1)	high (200)
Current gain	high (200)	high (200)	unity (1)
Power gain	very high (8000)	high (200)	high (200)
Input resistance	medium (2.5 kΩ)	high (100 kΩ)	low (200 Ω)
Output resistance	medium/high (20 kΩ)	low (100 Ω)	high (100 kΩ)
Phase shift	180°	0°	0°

A requirement of most amplifiers is that the output signal should be a faithful copy of the input signal or be somewhat larger in amplitude. Other types of amplifier are 'non-linear', in which case their input and output waveforms will not necessarily be similar. In practice, the degree of linearity provided by an amplifier can be affected by a number of factors including the amount of bias applied and the amplitude of the input signal. It is also worth noting that a linear amplifier will become non-linear when the applied input signal exceeds a threshold value. Beyond this value the amplifier is said to be overdriven and the output will become increasingly distorted if the input signal is further increased.

The optimum value of bias for linear (**Class A**) amplifiers is that value which ensures that the active devices are operated at the mid-point of their characteristics. In practice, this means that a static value of collector current will flow even when there is no signal present. Furthermore, the collector current will flow throughout the complete cycle of an input signal (i.e. conduction will take place over an angle of 360°). At no stage should the transistor be **saturated** ($V_{CE} \approx 0V$ or $V_{DS} \approx 0V$) nor should it be **cut-off** ($V_{CE} \approx V_{CC}$ or $V_{DS} \approx V_{DD}$).

In order to ensure that a static value of collector current flows in a transistor, a small **bias current** must be applied to the base of the transistor. This current is usually derived from the same voltage rail that supplies the collector circuit via one or more resistors of appropriate value.

2.3.4 Current gain

In general terms, current gain is the ratio of output current to input current. When a transistor is operating in common emitter mode the input current appears at the base and the output current at the collector. Thus, for this mode of operation the current gain will simply be the ratio of collector current, I_C, to base current, I_B. We use the symbol h_{FE} to represent the static value of current gain when a transistor is connected in common emitter mode. Thus:

$$h_{FE} = \frac{I_C}{I_B}$$

Typical values of common emitter current gain vary from about 40 to 200.

Example 2.2

A current of 4 A is to be supplied to a generator field winding from a BJT. If the device is operated in common emitter mode with a base current of 20 mA, what is the minimum value of current gain required?

Using $h_{FE} = I_C/I_B$ where $I_C = 4\,\text{A}$ and $I_B = 20\,\text{mA}$ gives:

$$h_{FE} = \frac{I_C}{I_B} = \frac{4\,\text{A}}{80\,\text{mA}} = \frac{4\,\text{A}}{0.08\,\text{A}} = 50$$

Example 2.3

A small-signal BJT has a common emitter current gain of 125. If the transistor operates with a collector current of 50 mA, determine the value of base current.

Rearranging the formula $h_{FE} = I_C/I_B$ to make I_B the subject gives $I_B = I_C/h_{FE}$ from which:

$$I_B = \frac{50 \times 10^{-3}}{125} = 400\mu\text{A}$$

Test your understanding 2.6

A BJT has a common emitter current gain of 75. If a base current of 40 mA is applied to the device what will the collector current be?

Test your understanding 2.7

Estimate the common emitter current gain of the BJT whose transfer characteristic is shown in Fig. 2.35.

Key point

The common emitter current gain of a transistor is the ratio of collector current to base current and is typically in the range 40 to 200.

2.4 Integrated circuits

Considerable cost savings can be made by manufacturing all of the components required for a particular circuit function on one small slice of semiconductor material (usually silicon). The resulting **integrated circuit** may contain as few as 10 or more than 100,000 active devices (transistors and diodes). With the exception of a few specialized applications (such as amplification at high power levels) integrated circuits have largely rendered conventional circuits (i.e. those based on **discrete components**) obsolete.

Integrated circuits can be divided into two general classes, digital (logic) and linear (analogue).

A number of devices bridge the gap between the analogue and digital world. Such devices include **analogue to digital converters** (**ADC**), **digital to analogue converters** (**DAC**), and **timers**.

Digital integrated circuits have numerous applications quite apart from their obvious use in computing. Digital signals exist only in discrete steps or levels; intermediate states are disallowed. Conventional electronic logic is based on two binary states, commonly referred to as logic 0 (low) and logic 1 (high). A comparison between digital and analogue signals is shown in Fig. 2.39. We shall look at digital logic in much greater detail in Chapter 3.

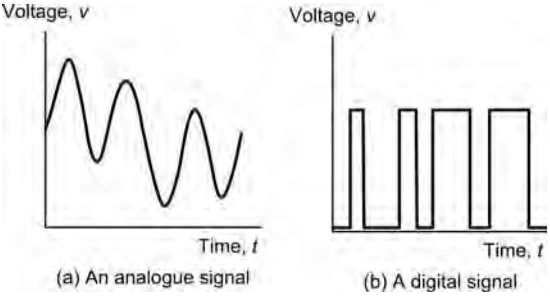

(a) An analogue signal (b) A digital signal

Figure 2.39 Digital and analogue signals

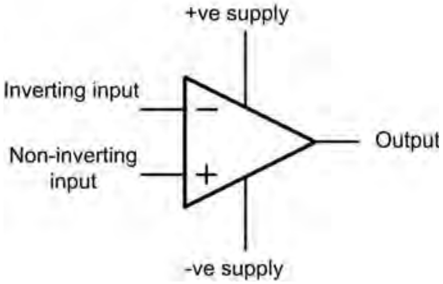

Figure 2.40 Symbol for an operational amplifier

Operational amplifiers are analogue integrated circuits designed for linear amplification that offer near-ideal characteristics (virtually infinite voltage gain and input resistance coupled with low output resistance and wide bandwidth).

Operational amplifiers can be thought of as universal ‘gain blocks’ to which external components are added in order to define their function within a circuit. By adding two resistors, we can produce an amplifier having a precisely defined gain. Alternatively, with three resistors and two capacitors we can realize a low-pass filter. From this you might begin to suspect that operational amplifiers are really easy to use. The good news is that they are!

The symbol for an operational amplifier is shown in Fig. 2.40. There are a few things to note about this. The device has two inputs and one output and no common connection. Furthermore, we often don’t show the supply connections – it is often clearer to leave them out of the circuit altogether!

In Fig. 2.40, one of the inputs is marked ‘−’ and the other is marked ‘+’. These polarity markings

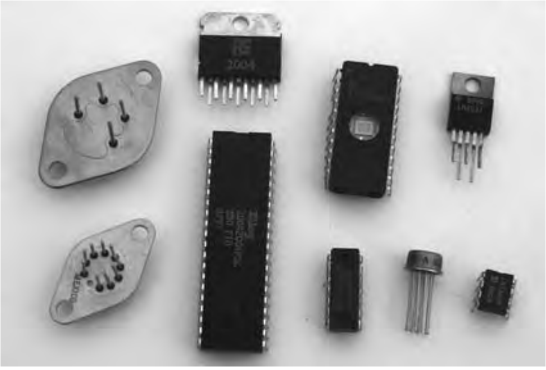

Figure 2.41 Various integrated circuits (including logic gates, operational amplifiers and memories)

have nothing to do with the supply connections – they indicate the overall phase shift between each input and the output. The '+' sign indicates zero phase shift whilst the '−' sign indicates 180° phase shift. Since 180° phase shift produces an inverted (i.e., turned upside down) waveform, the '−' input is often referred to as the 'inverting input'. Similarly, the '+' input is known as the 'non-inverting' input.

Most (but not all) operational amplifiers require a symmetrical power supply (of typically ±6V to ±15V). This allows the output voltage to swing both positive (above 0V) and negative (below 0V). Others types of operational amplifier operate from a single supply voltage of usually between 5V and 15V.

Key point

Integrated circuits contain large numbers of individual components fabricated on a single slice of silicon. Integrated circuits are often classified as either digital (logic) or linear (analogue).

Key point

Operational amplifiers are linear integrated circuits that can be used as versatile 'gain blocks' within a wide variety of linear circuits.

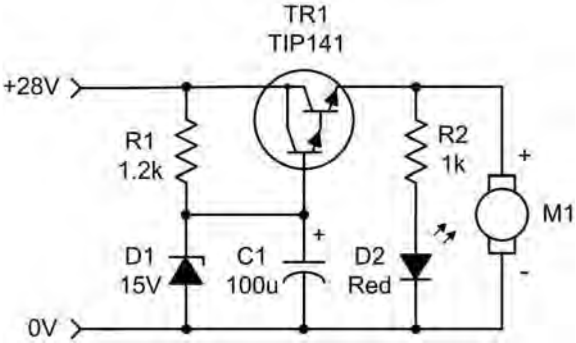

Figure 2.42 See Test your understanding 2.8

Test your understanding 2.8

The circuit of a motor controller is shown in Fig. 2.42.

1. What type of transistor is TR1?
2. What type of diode is D1?
3. What type of diode is D2?
4. Which resistor sets the brightness of D2?
5. What voltage appears across C1?
6. What voltage appears across R1?
7. What voltage is delivered to M1?
8. What operating configuration is used for TR1?
9. What would happen if R1 is open circuit?
10. What would happen if C1 is short circuit?

2.5 Multiple choice questions

1. The forward voltage drop of a conducting silicon diode is approximately:
   (a) 0.2V
   (b) 0.6V
   (c) 2V.

2. The region in a P-N junction diode where no free charge carriers exist is known as the:
   (a) collector
   (b) depletion region
   (c) enhancement region.

3. The device shown in Fig. 2.43 is:
   (a) a thyristor
   (b) a PNP transistor
   (c) an NPN transistor.

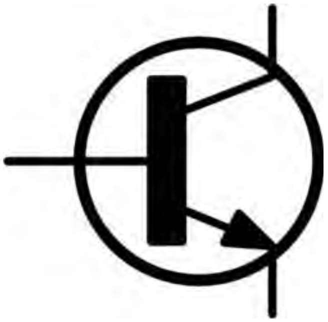

Figure 2.43 See Question 3

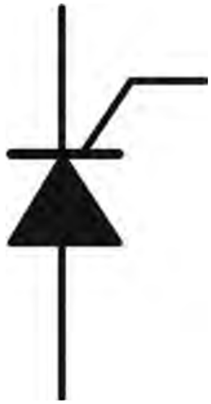

Figure 2.44 See Question 6

4. The typical common-emitter current gain for a bipolar transistor is:
   (a) less than 10
   (b) between 10 and 40
   (c) more than 40.
5. When a transistor is used in common-emitter mode:
   (a) the input is taken to the base and the output is taken from the collector
   (b) the input is taken to the collector and the output is taken from the emitter
   (c) the input is taken to the emitter and the output is taken from the collector.
6. The device shown in Fig. 2.44 is used for:
   (a) power control
   (b) acting as a voltage reference
   (c) producing a visual indication.

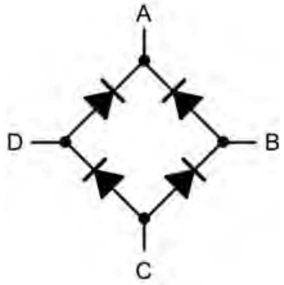

Figure 2.45 See Question 7

7. In the bridge rectifier arrangement shown in Fig. 2.45 the DC output is taken from:
   (a) A and B
   (b) A and C
   (c) B and D.
8. The connections on a thyristor (silicon-controlled rectifier) are labelled:
   (a) anode, cathode, gate
   (b) collector, base, emitter
   (c) collector, gate, emitter.
9. The anode of a silicon diode is connected to a $+5$ V DC supply and the cathode is connected to a $+4$ V DC supply. The diode is:
   (a) forward biased and not conducting
   (b) reverse biased and not conducting
   (c) forward biased and conducting.
10. A thyristor has which of the following?
    (a) high resistance when switched on
    (b) high resistance when switched off
    (c) a positive temperature coefficient
11. A practical operational amplifier has:
    (a) very high input impedance and very high voltage gain
    (b) very high input impedance and very low voltage gain
    (c) very low input impedance and very high voltage gain.
12. A typical application for a Zener diode is:
    (a) voltage stabilization
    (b) power rectification
    (c) signal detection.
13. A transistor operates with a base current of 45 mA and a collector current of 1.8 A. Which of

the following gives the current gain of the device when used in common emitter configuration?
(a) 40
(b) 90
(c) 225.

14. Which one of the following relationships is correct for a BJT?
(a) $I_B = I_C - I_E$
(b) $I_C = I_E - I_B$
(c) $I_E = I_C - I_B$

15. The frequency of the ripple voltage present on the output of a bridge rectifier fed from a 400 Hz AC supply will be:
(a) 200 Hz
(b) 400 Hz
(c) 800 Hz.

Chapter 3 Digital fundamentals

In this chapter we introduce the basic building blocks of the digital logic systems found in modern aircraft. We begin with a review of the different types of logic gates before moving on to explore their use in several typical aircraft applications. The chapter also provides a brief overview of coding systems and the logic arrangements that are used to generate and convert the codes that are used to represent numerical data. The chapter provides an introduction to aircraft bus systems and concludes with a brief overview of the architecture and principal constituents of simple computer systems.

3.1 Logic gates

Aircraft logic systems follow the same conventions and standards as those used in other electronic applications. In particular, the MIL/ANSI standard logic symbols are invariably used and the logic elements that they represent operate in exactly the same way as those used in non-aircraft applications. MIL/ANSI standard symbols for the most common logic gates are shown together with their truth tables in Fig. 3.1.

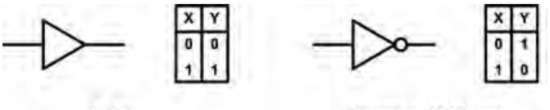

Buffer

Inverter or NOT gate

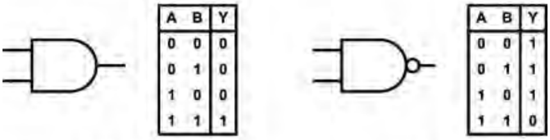

AND gate

NAND gate

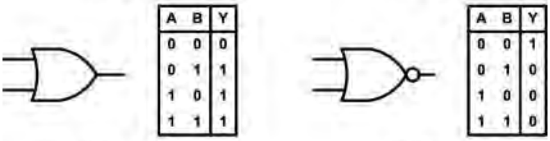

OR gate

NOR gate

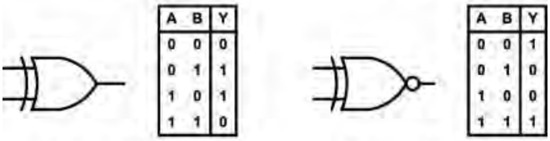

Exclusive-OR gate

Exclusive-NOR gate

Figure 3.1 MIL/ANSI symbols for standard logic gates together with truth tables

3.1.1 Buffers

Buffers do not affect the logical state of a digital signal (i.e. a logic 1 input results in a logic 1 output whereas a logic 0 input results in a logic 0 output). Buffers are normally used to provide extra current drive at the output but can also be used to regularize the logic levels present at an interface.

Inverters are used to complement the logical state (i.e. a logic 1 input results in a logic 0 output and vice versa). Inverters also provide extra current drive and, like buffers, are used in interfacing applications where they provide a means of regularizing logic levels present at the input or output of a digital system.

3.1.2 AND logic

AND gates will only produce a logic 1 output when all inputs are simultaneously at logic 1. Any other input combination results in a logic 0 output.

3.1.3 OR logic

OR gates will produce a logic 1 output whenever any one, or more, inputs are at logic 1. Putting this another way, an OR gate will only produce a logic 0 output whenever all of its inputs are simultaneously at logic 0.

3.1.4 NAND logic

NAND (i.e. NOT-AND) gates will only produce a logic 0 output when all inputs are simultaneously at logic 1. Any other input combination will produce a logic 1 output. A NAND gate, therefore, is nothing more than an AND gate with its output inverted. The circle shown at the output of the gate denotes this inversion.

3.1.5 NOR logic

NOR (i.e. NOT-OR) gates will only produce a logic 1 output when all inputs are simultaneously at logic 0. Any other input combination will produce a logic 0 output. A NOR gate, therefore, is simply an OR gate with its output inverted. A circle is again used to indicate inversion.

3.1.6 Exclusive-OR logic

Exclusive-OR gates will produce a logic 1 output whenever either one of the two inputs is at logic 1 and the other is at logic 0. Exclusive-OR gates produce a logic 0 output whenever both inputs have the same logical state (i.e. when both are at logic 0 or both are at logic 1).

3.1.7 Exclusive-NOR logic

Exclusive-NOR gates will produce a logic 0 output whenever either one of the two inputs is at logic 1 and the other is at logic 0. Exclusive-NOR gates produce a logic 1 output whenever both inputs have the same logical state (i.e. when both are at logic 0 or both are at logic 1).

3.1.8 Inverted inputs and outputs

The NAND and NOR gates that we have just met are said to have inverted outputs. In other words, they are respectively equivalent to AND and OR gates with their outputs passed through an inverter (or NOT gate) as shown in Fig. 3.2(a) and (b).

As well as inverted outputs, aircraft logic systems also tend to show logic gates in which one or more of the inputs is inverted. In Fig. 3.2(c) an AND gate is shown with one input inverted. This is equivalent to an inverter (NOT gate) connected to one input of the AND gate, as shown. In Fig. 3.2(d) an OR gate is shown with one input inverted. This is equivalent to an inverter (NOT gate) connected to one input of the OR gate, as shown.

Two further circuits with inverted inputs are shown in Fig. 3.2. In Fig. 3.2(e) both inputs of an AND gate are shown inverted. This arrangement is equivalent to the two-input NOR gate shown. In Fig. 3.2(f), both inputs of an OR gate are shown inverted. This arrangement is equivalent to the two-input NAND gate shown.

Key point

Logic circuits involve signals that can only exist in one of two mutually exclusive states. These two states are usually denoted by 1 and 0, 'on' or 'off', 'high' and 'low', 'closed' and 'open', etc.

3.2 Combinational logic systems

By using a standard range of logic levels (i.e. voltage levels used to represent the logic 1 and logic 0 states) logic circuits can be combined together in order to solve more complex logic functions. As an example, assume that a logic circuit is to be constructed that will produce a logic 1 output whenever two or more of its three inputs are at logic 1. This circuit is referred to as a majority vote circuit and its truth table is shown in Fig. 3.3. Figure 3.4 shows the logic circuitry required to satisfy the truth table.

3.2.1 Landing gear warning logic

Now let's look at a more practical example of the use of logic in the typical aircraft system shown in Fig. 3.5. The inputs to this logic system consist of five switches that detect whether or not the respective landing gear door is open. The output from the logic

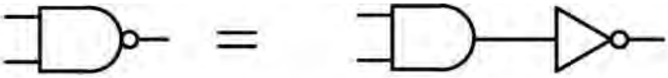

(a) AND gate with output inverted

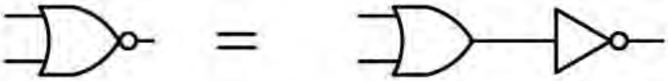

(b) OR gate with output inverted

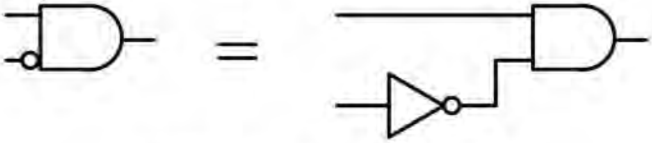

(c) AND gate with one input inverted

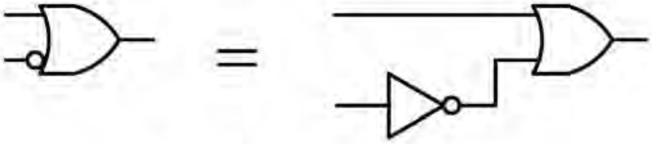

(d) OR gate with one input inverted

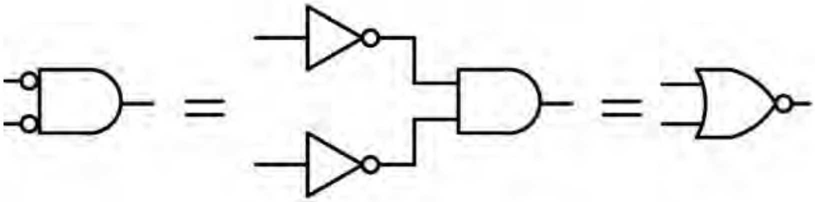

(e) AND gate with both inputs inverted

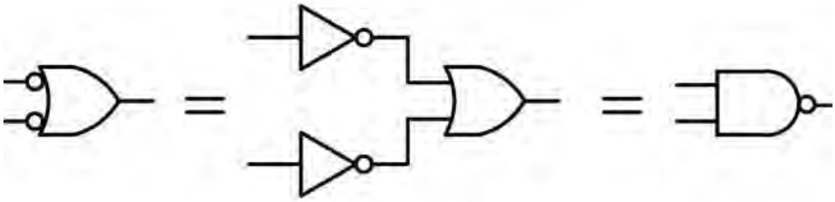

(f) OR gate with both inputs inverted

Figure 3.2 Logic gates with inverted outputs and inputs

system is used to drive six warning indicators. Four of these are located on the overhead display panel and show which door (or doors) are left open whilst an indicator located on the pilot's instrument panel provides a master landing gear door warning. A switch is also provided in order to enable or disable the five door warning indicators.

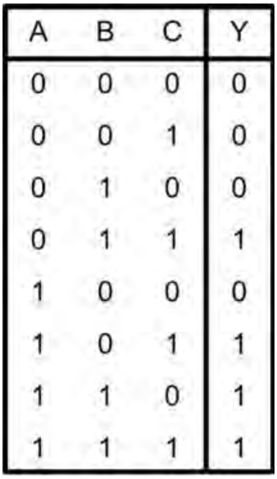

A	B	C	Y
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

Figure 3.3 The majority vote truth table

The landing gear warning logic primary module consists of the following integrated circuit devices:

A1	Regulated power supply for A5
A2	Regulated power supply for A7 and A11
A5	Ten inverting (NOT) gates
A7	Five-input NAND gate
A11	Six inverting (NOT) gates

Note that the power supply for A1 and A2 is derived from the essential services DC bus. This is a 28V DC bus which is maintained in the event of an

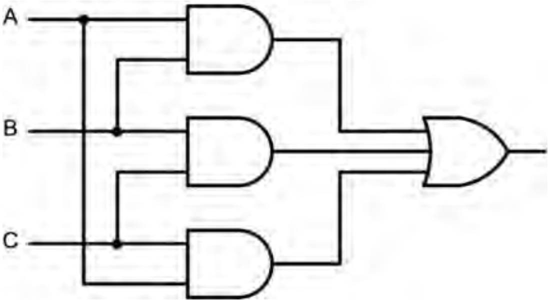

Figure 3.4 The majority vote logic

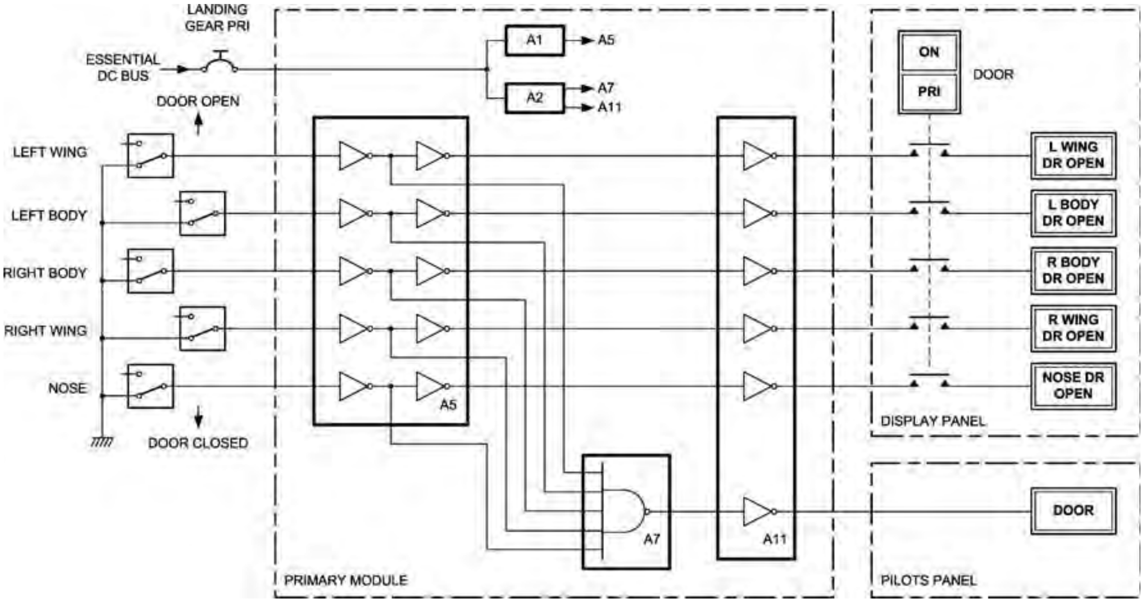
LANDING GEAR PRI
ESSENTIAL DC BUS
DOOR OPEN
LEFT WING
LEFT BODY
RIGHT BODY
RIGHT WING
NOSE
DOOR CLOSED
A1
A2
A5
A7
A11
A5
A7
A11
ON
PRI
DOOR
L WING DR OPEN
L BODY DR OPEN
R BODY DR OPEN
R WING DR OPEN
NOSE DR OPEN
DISPLAY PANEL
DOOR
PILOTS PANEL
PRIMARY MODULE

Figure 3.5 Landing gear warning logic

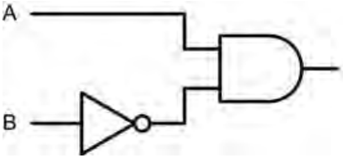

Figure 3.6 See Test your understanding 3.2

A	B	C	Y
0	0	0	0
0	0	1	0
0	1	0	1
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

Figure 3.7 See Test your understanding 3.2

aircraft power failure. Note also that the indicators are **active-low** devices (in other words, they require a logic 0 input in order to become illuminated).

Test your understanding 3.1

1. Show how a four-input AND gate can be made using three two-input AND gates.
2. Show how a four-input OR gate can be made using three two-input OR gates.

Test your understanding 3.2

1. Draw the truth table for the logic gate arrangement shown in Fig. 3.6.
2. Devise a logic gate arrangement that provides the output described by the truth table shown in Fig. 3.7.

3.3 Monostable devices

Monostable (or one-shot) devices provide us with a means of generating precise time delays. Such delays become important in many logic applications where logic states are not static but change with time.

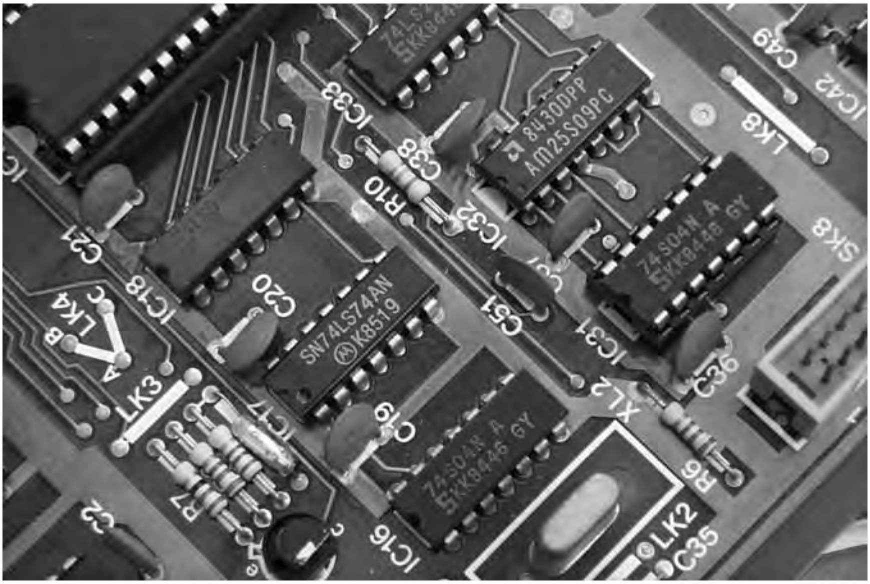
IC33
C38
R10
8430DPP
AM25S09PC
IC32
C51
74S04N A
5KK8446 GY
SN74LS74AN
K8519
C20
IC18
C21
LK4
LK3
IC17
C19
IC31
XL2
C36
R6
LK2
C35
IC16
R7
C32
SK8
LK8
C49
IC42

Figure 3.8 Logic circuit based on transistor-transistor logic (TTL) integrated circuits

The action of a monostable is quite simple – its output is initially logic 0 until a change of state occurs at its trigger input. The level change can be from 0 to 1 (positive edge trigger) or 1 to 0 (negative edge trigger). Immediately the trigger pulse arrives, the output of the monostable changes state to logic 1. The output then remains at logic 1 for a pre-determined period before reverting back to logic 0.

3.3.1 APU starter logic

An example of the use of a monostable is shown in the auxiliary power unit (APU) starter logic shown in Fig. 3.9. This arrangement has three inputs (APU START, APU SHUTDOWN, and APU RUNNING) and one output (APU STARTER MOTOR). The inputs are all active-high (in other words, a logic 1 is generated when the pilot operates the APU START switch, and so on). The output of the APU starter motor control logic goes to logic 1 in order to apply power to the starter motor via a large relay.

There are a few things to note about the logic arrangement shown in Fig. 3.9:

1. When the APU runs on its own we need to disengage the starter motor. In this condition the APU MOTOR signal needs to become inactive (i.e. it needs to revert to logic 0).
2. We need to avoid the situation that might occur if the APU does not start but the starter motor runs continuously (as this will drain the aircraft batteries). Instead, we should run the starter motor for a reasonable time (say, 60 seconds) before disengaging the starter motor. The 60 second timing is provided by means of a positive edge triggered monostable device. This device is triggered from the APU START signal.
3. Since the pilot is only required to momentarily press the APU START switch, we need to hold the condition until such time as the engine starts or times out (i.e. at the end of the 60 second period).

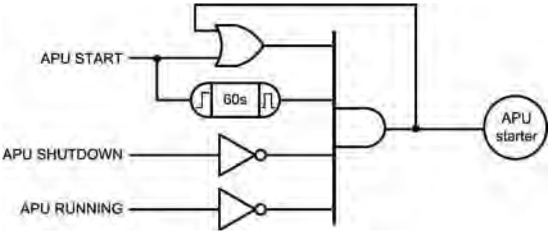
APU START
60s
APU SHUTDOWN
APU RUNNING
APU starter

Figure 3.9 APU starter logic

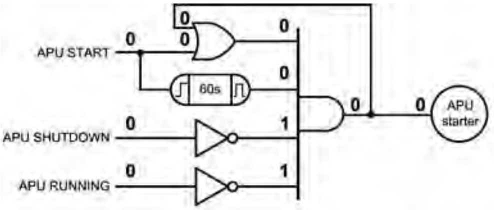
0
APU START
0
0
0
60s
0
APU SHUTDOWN
0
1
0
0
APU starter
APU RUNNING
0
1

(a) Normal flight; engine power generation; APU not running

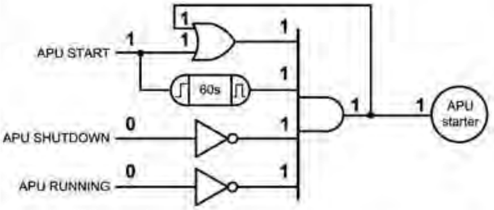
1
APU START
1
1
1
60s
1
APU SHUTDOWN
0
1
1
1
APU starter
APU RUNNING
0
1

(b) APU starter switch operated; APU starter motor begins to run

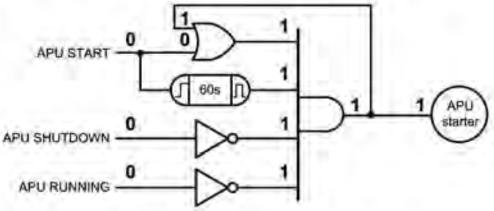
1
APU START
0
0
1
60s
1
APU SHUTDOWN
0
1
1
1
APU starter
APU RUNNING
0
1

(c) APU starter motor continues to run for up to 60s

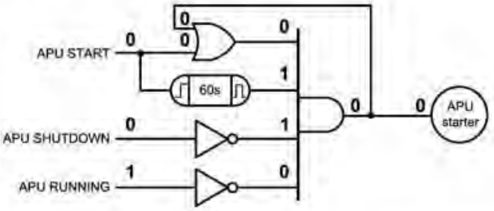
0
APU START
0
0
0
60s
1
APU SHUTDOWN
0
1
0
0
APU starter
APU RUNNING
1
0

(d) APU runs before 60s timeout; starter motor stops when APU runs

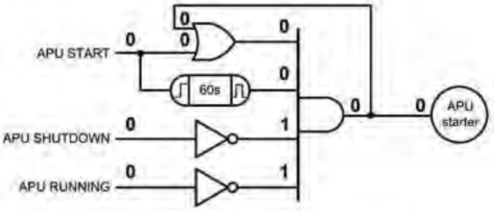
0
APU START
0
0
0
60s
0
APU SHUTDOWN
0
1
0
0
APU starter
APU RUNNING
0
1

(e) APU fails to run during 60s period; further APU START signal awaited

Figure 3.10 APU starter operation

We can achieve this by OR'ing the momentary APU START signal with the APU STARTER MOTOR signal.

4. We need to provide a signal that the pilot can use to shut down the APU (for example, when the aircraft's main engines are running or perhaps in the event of a fault condition).

In order to understand the operation of the APU starter motor logic system we can once again trace through the logic system using 1's and 0's to represent the logical condition at each point (just as we did for the landing gear door warning logic).

In Fig. 3.10(a) the APU is in normal flight and the APU is not running. In this condition the main engines are providing the aircraft's electrical power.

In Fig. 3.10(b) the pilot is operating the APU START switch. The monostable is triggered and output of the OR and AND gates both go to logic 1 in order to assert the APU STARTER MOTOR signal.

In Fig. 3.10(c) the APU START signal is removed but the output of the AND gate is held at logic 1 by feeding back its logical state via the OR gate. The monostable remains triggered and continues to produce a logic 1 output for its 60 second period.

In Fig. 3.10(d) the APU is now running and the APU RUNNING signal has gone to logic 1 in order to signal this conditions. This results in the output of the AND gate going to logic 0 and the APU STARTER MOTOR signal is no longer made active. The starter motor is therefore disengaged.

In Fig. 3.10(e) the APU has failed to run during the 60 second monostable period. In this timed out condition the output of the AND gate goes to logic 0 and the APU STARTER MOTOR signal becomes inactive. The system then waits for the pilot to operate the APU START button for a further attempt at starting!

Test your understanding 3.3

Part of an aircraft's generator control logic is shown in Fig. 3.11. Construct a truth table for the two outputs (GCB trip and GCR trip) and use this to explain the operation of the circuit. Under what conditions are the trip signals generated?

3.4 Bistable devices

A bistable device is a logic arrangement that is capable of 'remembering' a transient logic state such as a key press or a momentary overload condition. The output of a bistable circuit has two stables states (logic 0 or logic 1). Once **set** in one or other of these states, the output of a bistable will remain at a particular logic level for an indefinite period until **reset**. A bistable thus forms a simple form of memory as it remains in its latched state (either **set** or **reset**) until a signal is applied to it in order to change its state (or until the supply is disconnected).

The simplest form of bistable is the **R-S bistable**. This device has two inputs, SET and RESET, and complementary outputs, Q and /Q. A logic 1 applied to the SET input will cause the Q output to become (or remain at) logic 1 while a logic 1 applied to the RESET input will cause the Q output to become (or remain at) logic 0. In either case, the bistable will remain in its SET or RESET state until an input is applied in such a sense as to change the state.

Two simple forms of R-S bistable based on cross-coupled logic gates are shown in Fig. 3.12. Figure 3.12(a) is based on cross-coupled two-input NAND gates while Fig. 3.12(b) is based on cross-coupled two-input NOR gates.

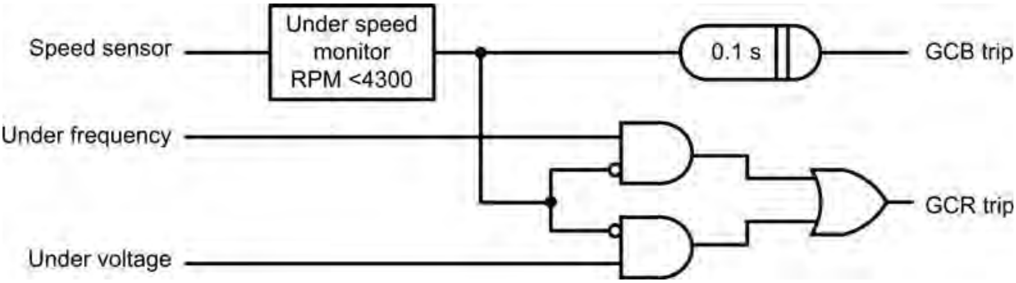

Figure 3.11 Generator control logic

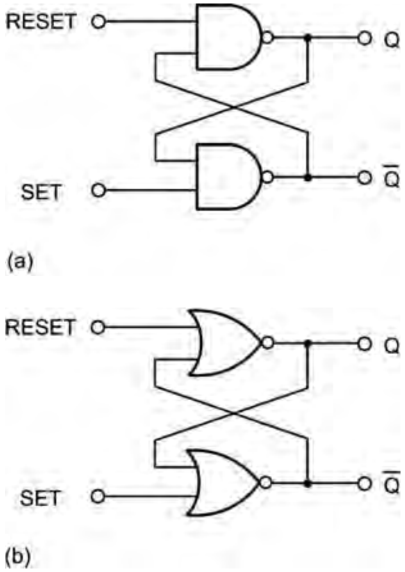

Figure 3.12 Simple R-S bistables based on: (a) NAND gates and (b) NOR gates

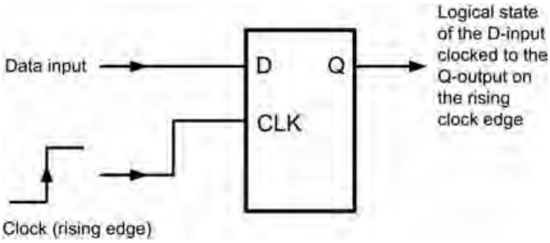
Data input
D
Q
CLK
Logical state of the D-input clocked to the Q-output on the rising clock edge
Clock (rising edge)

Figure 3.13 D-type bistable

Unfortunately, the simple cross-coupled logic gate bistable has a number of serious shortcomings (consider what would happen if a logic 1 was simultaneously present on both the SET and RESET inputs!) and practical forms of bistable make use of much improved purpose-designed logic circuits such as D-type and J-K bistables.

The **D-type bistable** has two inputs: D (standing variously for 'data' or 'delay') and CLOCK (CLK). The data input (logic 0 or logic 1) is clocked into the bistable such that the output state only changes when the clock changes state. Operation is thus said to be synchronous. Additional subsidiary inputs (which are invariably active low) are provided which can be used to directly set or reset the bistable. These are usually called PRESET (PR) and CLEAR (CLR). D-type bistables are used both as latches (a simple form of memory) and as binary dividers. The simple circuit arrangement in Fig. 3.13 together with the **timing diagram** shown in Fig. 3.14 illustrate the operation of D-type bistables.

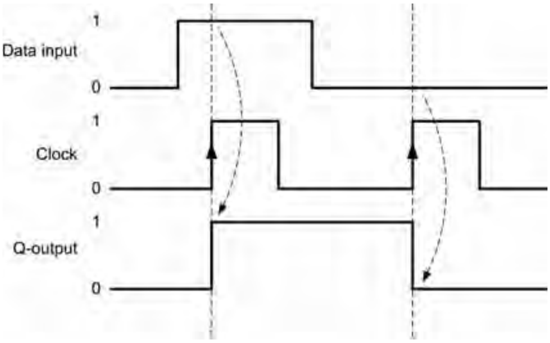

Figure 3.14 Timing diagram for the D-type bistable

J-K bistables (see Fig. 3.15) have two clocked inputs (J and K), two direct inputs (PRESET and CLEAR), a CLOCK (CK) input, and complementary outputs (Q and /Q). As with R-S bistables, the two outputs are complementary (i.e. when one is 0 the other is 1, and vice versa). Similarly, the PRESET and CLEAR inputs are invariably both active low (i.e. a 0 on the PRESET input will set the Q output to 1 whereas a 0 on the CLEAR input will set the Q output to 0). Figure 3.16 summarizes the input and corresponding output states of a J-K bistable for various input states. J-K bistables are the most sophisticated and flexible of the bistable types and they can be configured in various ways for use in binary dividers, shift registers, and latches.

3.5 Decoders

A variety of different coding schemes are used to represent numerical data in avionic systems. They include binary (or, more correctly, **natural binary**), **binary coded decimal** (**BCD**), **Gray code**, **octal** (base 8), and **hexadecimal** (base 16) as shown in Fig. 3.17.

BCD uses four digits to represent each numerical character. Thus decimal 11 is represented by 0001 0001 (or just 10001 omitting the three leading zeros). Gray code is an important code because only one digit changes at a time. This property assists with error correction. Notice also how the least significant three bits of each Gray coded number become reflected after a decimal count of seven. Because of this, Gray code is often referred to as a **reflected code**.

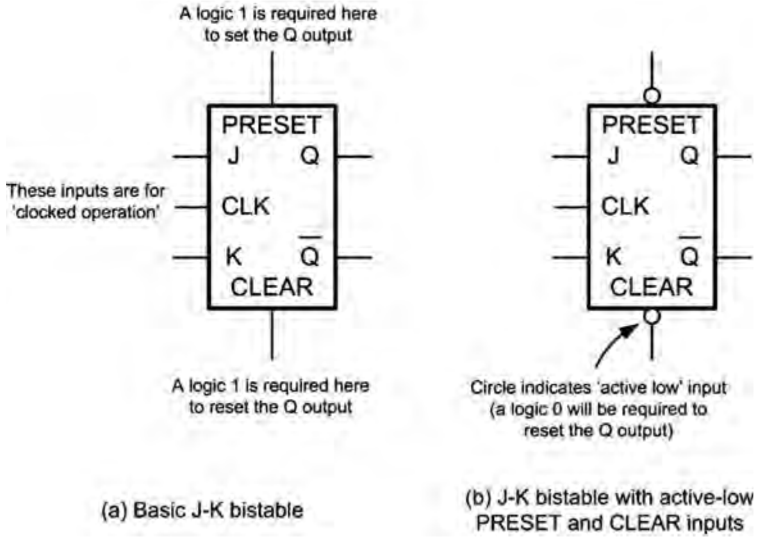
A logic 1 is required here to set the Q output
PRESET
J
Q
These inputs are for 'clocked operation'
CLK
K
$\overline{Q}$
CLEAR
A logic 1 is required here to reset the Q output
(a) Basic J-K bistable
Circle indicates 'active low' input (a logic 0 will be required to reset the Q output)
(b) J-K bistable with active-low PRESET and CLEAR inputs

Figure 3.15 J-K bistable symbols

(a) PRESET and CLEAR inputs

Inputs PRESET	Inputs CLEAR	Output Q	Comment
0	0	?	Indeterminate
0	1	0	Q output changes to 0 (i.e. Q is reset) regardless of the clock
1	0	1	Q output changes to 1 (i.e. Q is reset) on the next clock transition)
1	1	-	Enables clocked operation - refer to the next truth table

Note that the PRESET and CLEAR inputs are unaffected by the state of the clock

(b) Clocked operation using the J and K inputs

Inputs J	Inputs K	Output Q_{N+1}	Comment
0	0	Q_N	No change in state of the Q output on the next clock transition
0	1	0	Q output changes to 0 (i.e. Q is reset) on the next clock transition)
1	0	1	Q output changes to 1 (i.e. Q is reset) on the next clock transition)
1	1	Q_N	Q output changes to the opposite state on the next clock transition

Note that Q_N means 'Q in whatever state it was before' whilst Q_{N+1} means 'Q after the next clock transition'

Figure 3.16 Truth tables for the J-K bistable

Decoders are used to convert information from one number system to another, such as binary to octal or binary to decimal. A simple two to four line decoder is shown in Fig. 3.18. In this arrangement there are two inputs, A and B, and four outputs, Y_0, Y_1, Y_2 and Y_3.

The binary code appearing on A and B is decoded into one of the four possible states and corresponding output appears on the four output lines with Y_3 being the most significant. Because two to four and three to eight line decoders are frequently used as address decoders in computer systems (where memory and I/O devices are invariably enabled by a low rather than a high state), the outputs are active-low, as indicated by the circles on the logic diagrams.

The internal logic arrangement of a two to four line decoder is shown in Fig. 3.19. This arrangement uses two inverters and three two-input NAND gates. All four outputs are active-low; Y_0 will go low when A and B are both at logic 0, Y_1 will go low when A is at logic 0 and B is at logic 1, Y_2 will go low when A is

Dec.	Binary	BCD	Gray	Octal	Hex.
0	0000	0000 0000	0000	0	0
1	0001	0000 0001	0001	1	1
2	0010	0000 0010	0011	2	2
3	0011	0000 0011	0010	3	3
4	0100	0000 0100	0110	4	4
5	0101	0000 0101	0111	5	5
6	0110	0000 0110	0101	6	6
7	0111	0000 0111	0100	7	7
8	1000	0000 1000	1100	10	8
9	1001	0000 1001	1101	11	9
10	1010	0001 0000	1111	12	A
11	1011	0001 0001	1110	13	B
12	1100	0001 0010	1010	14	C
13	1101	0001 0011	1011	15	D
14	1110	0001 0100	1001	16	E
15	1111	0001 0101	1000	17	F

Figure 3.17 Various coding schemes for numerical data

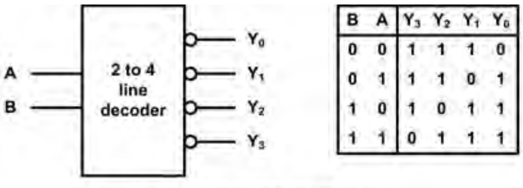

B	A	Y_3	Y_2	Y_1	Y_0
0	0	1	1	1	0
0	1	1	1	0	1
1	0	1	0	1	1
1	1	0	1	1	1

(a) Basic two to four line decoder

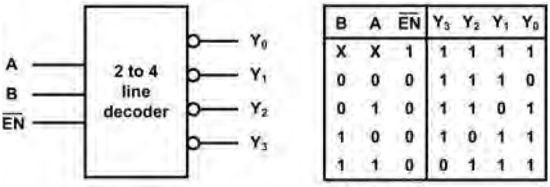

B	A	$\overline{EN}$	Y_3	Y_2	Y_1	Y_0
X	X	1	1	1	1	1
0	0	0	1	1	1	0
0	1	0	1	1	0	1
1	0	0	1	0	1	1
1	1	0	0	1	1	1

X = don't care

(b) Two to four line decoder with enable input

Figure 3.18 Two to four line decoders and their corresponding truth tables. The arrangement shown in (b) has a separate active-low enable input (when this input is high all of the outputs remain in the high state regardless of the A and B inputs)

at logic 1 and B is at logic 0, and Y_3 will go low when both A and B are at logic 1.

Test your understanding 3.4

1. Convert decimal 20 to (a) natural binary, (b) BCD, (c) octal, and (d) hexadecimal.
2. Convert natural binary 100001 to (a) decimal, (b) BCD, (c) octal, and (d) hexadecimal.

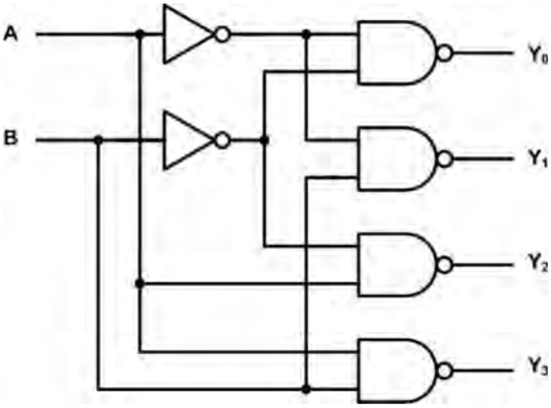

Figure 3.19 Internal logic of a simple two to four line decoder

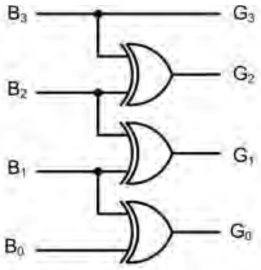

(a) Binary to Gray code conversion

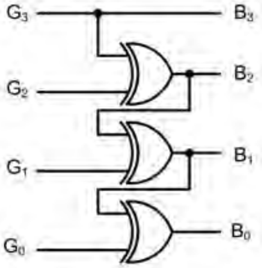

(b) Gray code to binary conversion

Figure 3.20 Natural binary to Gray code conversion

Test your understanding 3.5

Figure 3.20 shows simple logic arrangements that will perform natural binary to Gray code conversion and vice versa. Construct a truth table for each arrangement and use them to verify the operation of each circuit.

3.5.1 Gillham interface and Gillham code

Early altitude encoders were mechanical devices integrated into the barometric altimeter. The electrical output was a 10-bit parallel bus although more modern implementations use an expanded 11-bit bus. To ensure that only one bit of the bus changes on each transition (so that no transient false values are generated as the output changes between successive values) the altitude is represented in a form of Gray code, not natural binary code (see table below). The interface required to perform this operation is commonly referred to as a **Gillham interface** and, as a result, the altitude code that it produces (which operates in increments of 100’) is often referred to as **Gillham code**, see below:

Altitude (feet)	*Data line*									
	D4	*A1*	*A2*	*A4*	*B1*	*B2*	*B4*	*C1*	*C2*	*C4*
0	0	0	0	0	0	1	1	0	1	0
100	0	0	0	0	0	1	1	1	1	0
200	0	0	0	0	0	1	1	1	0	0
300	0	0	0	0	0	1	0	1	0	0
400	0	0	0	0	0	1	0	1	1	0
500	0	0	0	0	0	1	0	0	1	0
600	0	0	0	0	0	1	0	0	1	1
700	0	0	0	0	0	1	0	0	0	1
800	0	0	0	0	1	1	0	0	0	1
900	0	0	0	0	1	1	0	0	1	1
1000	0	0	0	0	1	1	0	0	1	0
1100	0	0	0	0	1	1	0	1	1	0
30500	0	1	0	0	0	0	0	0	1	0
30600	0	1	0	0	0	0	0	0	1	1
30700	0	1	0	0	0	0	0	0	0	1
30800	1	1	0	0	0	0	0	0	0	1
30900	1	1	0	0	0	0	0	0	1	1
31000	1	1	0	0	0	0	0	0	1	0
31100	1	1	0	0	0	0	0	1	1	0

Key maintenance point

Unfortunately, there has been at least one case when the failure of a single data line in a Gillham interface has generated erroneous altitude data which has caused an aircraft’s **Traffic Alert and Collision Avoidance System** (**TCAS**) to guide an aircraft into a **near miss** situation. Hence there is now a requirement to test and verify the operation of Gillham-based systems. Furthermore, the use of Gillham code interfaces has actively been discouraged in favour of more modern digital systems.

3.6 Encoders

Encoders provide the reverse function to that of a decoder. In other words, they accept a number of inputs and then generate a binary code corresponding to the state of those inputs. Typical applications for encoders include generating a binary code corresponding to the state of a keyboard/keypad or generating BCD from a decade (ten-position) rotary switch.

A particularly useful form of encoder is one that can determine the priority of its inputs. This device is known as a **priority encoder** and its inputs are arranged in priority order, from lowest to highest priority. If more

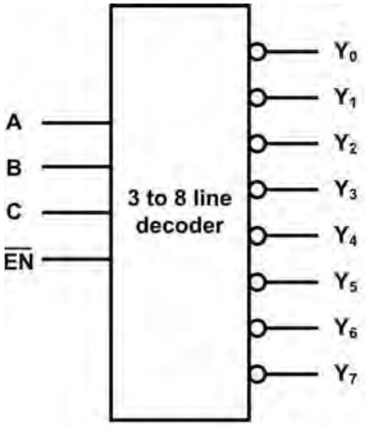

C	B	A	$\overline{EN}$	Y_7	Y_6	Y_5	Y_4	Y_3	Y_2	Y_1	Y_0
X	X	X	1	1	1	1	1	1	1	1	1
0	0	0	0	1	1	1	1	1	1	1	0
0	0	1	0	1	1	1	1	1	1	0	1
0	1	0	0	1	1	1	1	1	0	1	1
0	1	1	0	1	1	1	1	0	1	1	1
1	0	0	0	1	1	1	0	1	1	1	1
1	0	1	0	1	1	0	1	1	1	1	1
1	1	0	0	1	0	1	1	1	1	1	1
1	1	1	0	0	1	1	1	1	1	1	1

X = don't care

Figure 3.21 An eight to three line priority encoder and its corresponding truth table. Note that the enable (EN) input is active-low

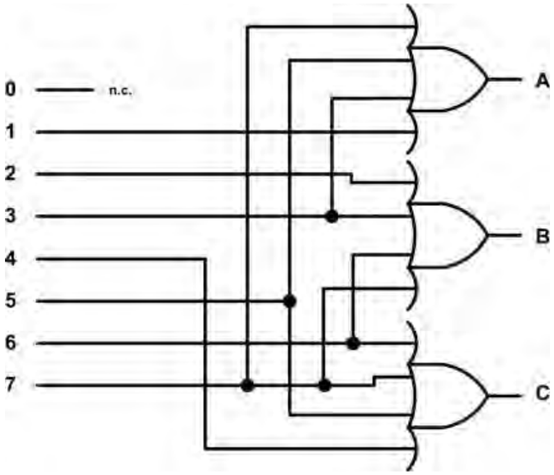

Figure 3.22 An octal to binary encoder

than one input becomes active, the input with the highest priority will be encoded and its binary coded value will appear on the outputs. The state of the other (lower priority) inputs will be ignored.

Figure 3.21 shows an eight to three line priority encoder together with its corresponding truth table. Note that this device has active-low inputs as well as active-low outputs.

Test your understanding 3.6

Figure 3.22 shows the circuit of an octal to binary encoder. Verify the operation of this circuit by constructing its truth table.

3.7 Multiplexers

Like encoders, multiplexers have several inputs. However, unlike encoders, they have only one output. Multiplexers provide a means of selecting data from one of several sources. Because of this, they are often referred to as **data selectors**. Switch equivalent circuits of some common types of multiplexer are shown in Fig. 3.23.

The single two-way multiplexer in Fig. 3.23(a) is equivalent to a simple SPDT (changeover) switch. The dual two-way multiplexer shown in Fig. 3.23(b) performs the same function but two independent circuits are controlled from the same select signal. A single four-way multiplexer is shown in Fig. 3.23(c). Note that two digital select inputs are required, A and B, in order to place the switch in its four different states.

Block schematic symbols, truth tables and simplified logic circuits for two to one and four to one multiplexers are shown in Figs 3.24 to 3.27.

Test your understanding 3.7

1. Verify the operation of the two to one multiplexer circuit shown in Fig. 3.25 by constructing its truth table.
2. Verify the operation of the four to one multiplexer circuit shown in Fig. 3.27 by constructing its truth table.

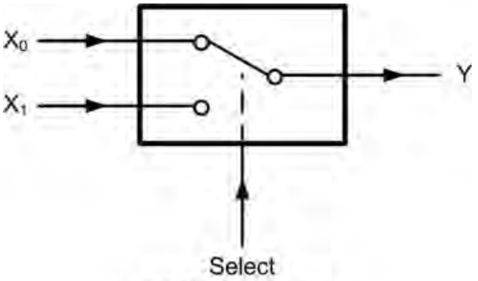
X_0
X_1
Y
Select

(a) Single two way

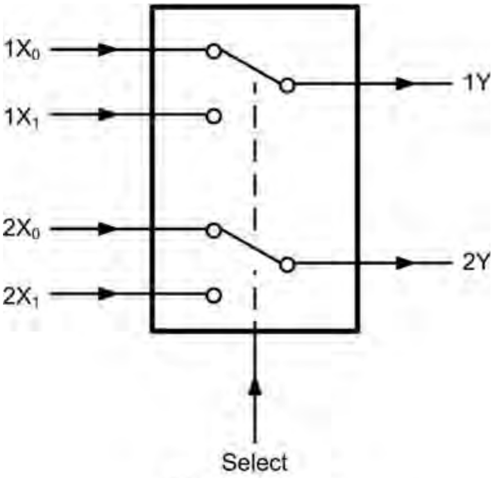
$1X_0$
$1X_1$
1Y
$2X_0$
$2X_1$
2Y
Select

(b) Dual two way

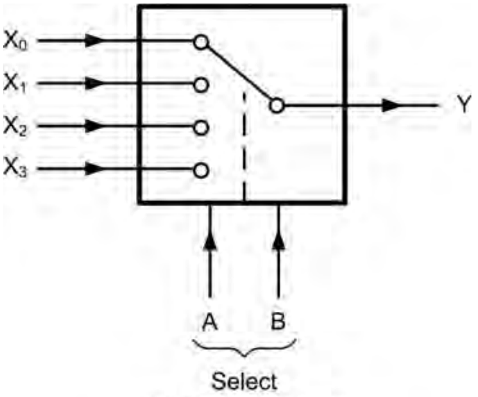
X_0
X_1
X_2
X_3
Y
A
B
Select

(c) Single four way

Figure 3.23 Switch equivalent circuits for some common types of multiplexer or data selector

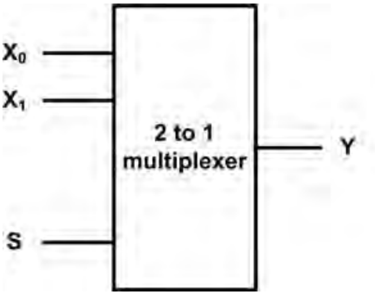
X_0
X_1
S
2 to 1 multiplexer
Y

S	Y
0	X_0
1	X_1

Figure 3.24 A basic two to one multiplexer arrangement with its corresponding truth table

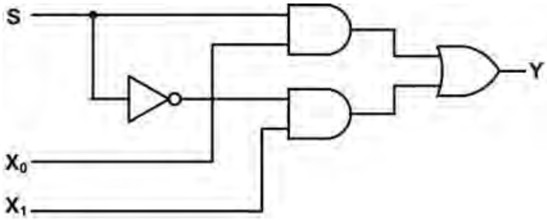

Figure 3.25 Logic circuit arrangement for the basic two to one multiplexer

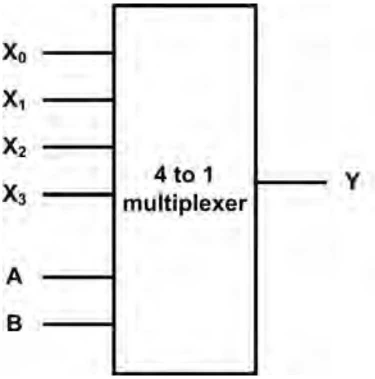
X_0
X_1
X_2
X_3
A
B
4 to 1 multiplexer
Y

B	A	Y
0	0	Y_0
0	1	Y_1
1	0	Y_2
1	1	Y_3

Figure 3.26 A four to one multiplexer. The logic state of the A and B inputs determines which of the four logic inputs (X_0 to X_3) appears at the output

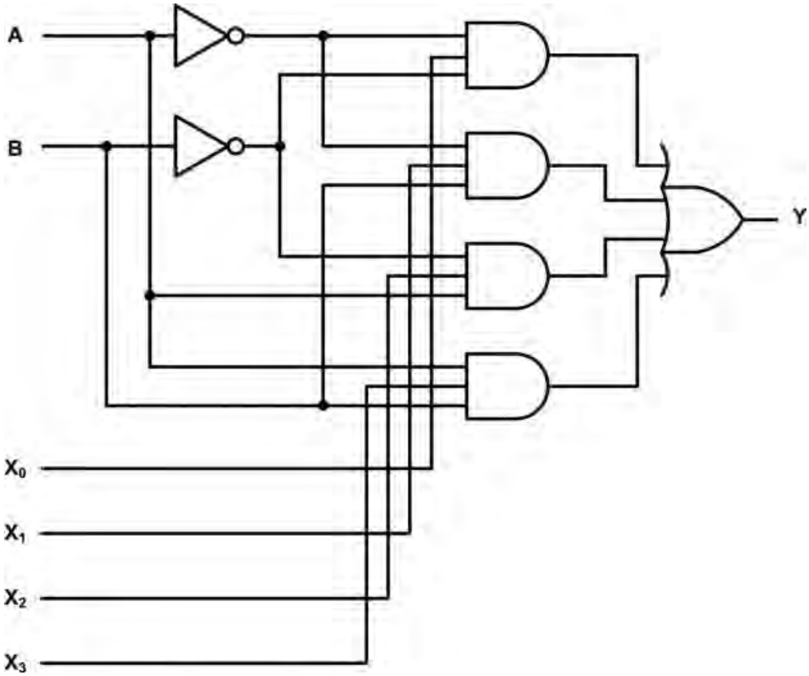

Figure 3.27 Logic gate arrangement for the four to one multiplexer

3.8 Bus systems

Aircraft data bus systems allow a wide variety of avionics equipment to communicate with one another and exchange data. Bus systems can be either **bidirectional** (one way) or **unidirectional** (two way), as shown in Fig. 3.28. They can also be serial (one bit of data transmitted at a time) or **parallel** (where often 8, 16 or 32 bits of data appear as a group on a number of data lines at the same time). Because of the constraints imposed by conductor length and weight, all practical aircraft bus systems are based on serial (rather than parallel) data transfer.

Bus systems provide an efficient means of exchanging data between the diverse avionic systems found in a modern aircraft (see Fig. 3.29). Individual **line replaceable units** (**LRU**), such as the engine data interface or flap/slat electronics units shown in Fig. 3.29, are each connected to the bus by means of a dedicated **bus coupler** and **serial interface module** (not shown in Fig. 3.29). Within the LRU, the dedicated digital logic and microprocessor systems that process data locally each make use of their own **local bus** system. These local bus systems invariably use parallel data transfer which is ideal for moving large amounts of data very quickly but only over short distances.

Key point

Modern aircraft use multiple redundant bus systems for exchanging data between the various avionic systems and sub-systems. These bus systems use serial data transfer because it minimizes the size and weight of aircraft cabling.

3.8.1 Serial bus principles

A simple system for serial data transfer between two line replaceable units, each of which comprises an avionic system in its own right, is shown in Fig. 3.30. Within the LRU data is transferred using an internal parallel data bus (either 8, 16, 32 or 64 bits wide). The link between the two LRUs is made using a simple serial cable (often with only two, four or six conductors). The required parallel-to-serial and serial-to-parallel data

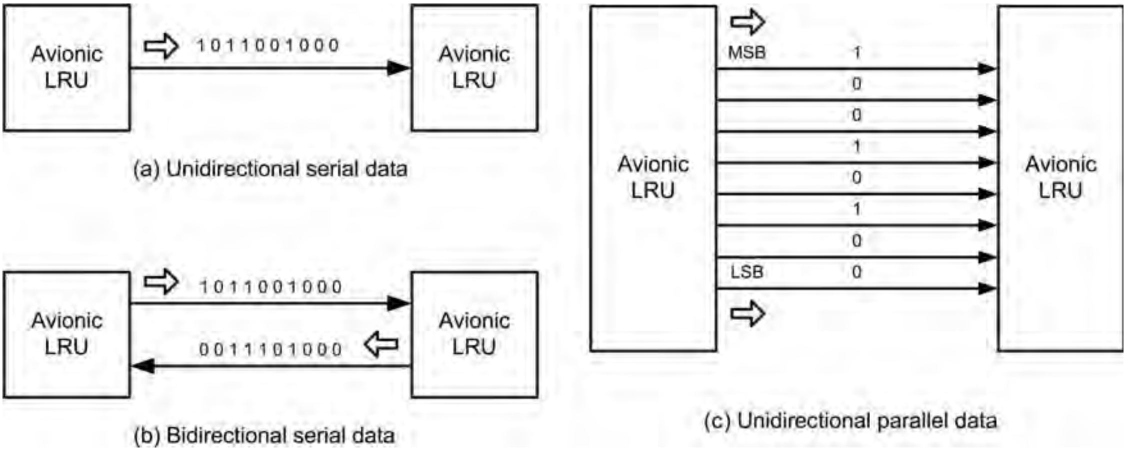
Avionic LRU
1011001000
Avionic LRU
(a) Unidirectional serial data
Avionic LRU
1011001000
0011101000
Avionic LRU
(b) Bidirectional serial data
Avionic LRU
MSB
1
0
0
1
0
1
0
LSB
0
Avionic LRU
(c) Unidirectional parallel data

Figure 3.28 Unidirectional and bidirectional serial and parallel data

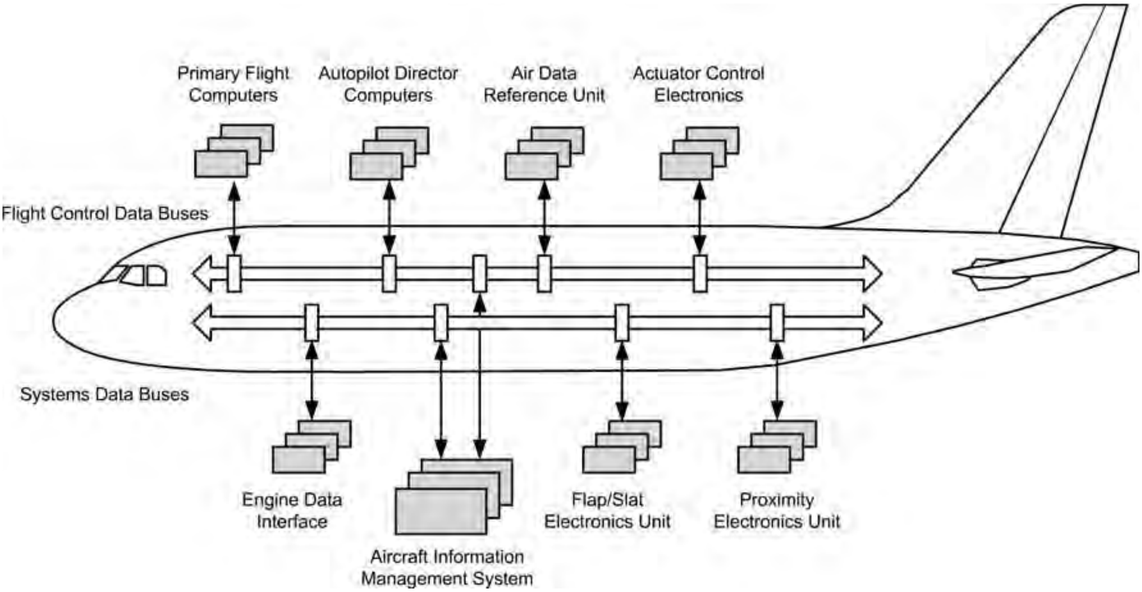
Primary Flight Computers
Autopilot Director Computers
Air Data Reference Unit
Actuator Control Electronics
Flight Control Data Buses
Systems Data Buses
Engine Data Interface
Aircraft Information Management System
Flap/Slat Electronics Unit
Proximity Electronics Unit

Figure 3.29 Bus systems implemented on a modern passenger aircraft

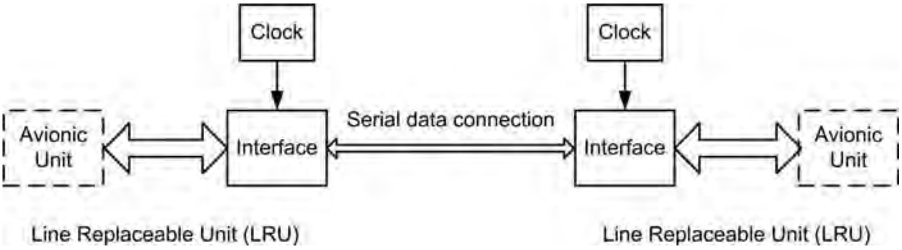
Clock
Clock
Avionic Unit
Interface
Serial data connection
Interface
Avionic Unit
Line Replaceable Unit (LRU)
Line Replaceable Unit (LRU)

Figure 3.30 A simple system for serial data transfer between two avionic systems

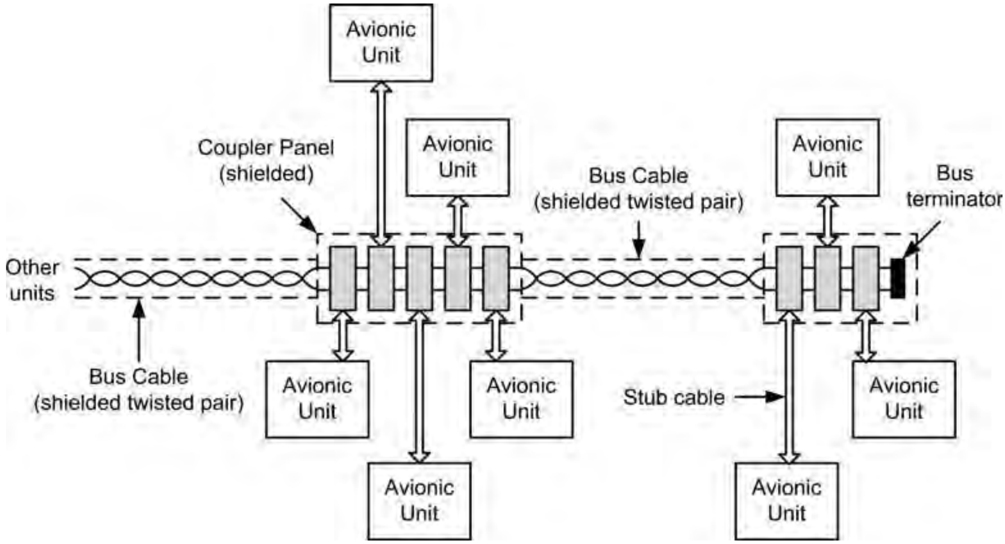

Figure 3.31 A practical aircraft data bus

conversion is carried out by a bus interface (often this is a single card or module within the LRU). The data to be transferred can be **synchronous** (using clock signals generated locally within each LRU) or it may be **asynchronous** (i.e. self-clocking).

The system shown in Fig. 3.30 has the obvious limitation that data can only be exchanged between two devices. In practice we need to share the data between many LRU/avionic units. This can be achieved by the bus system illustrated in Fig. 3.31. In this system, data is transferred using a **shielded twisted pair (STP) bus cable** with a number of **coupler panels** that are located at appropriate points in the aircraft (e.g. the flight deck, avionics bay, etc.). Each coupler panel allows a number of avionic units to be connected to the bus using a **stub cable**. In order to optimize the speed of data transfer and minimize problems associated with reflection and mismatch, the bus cable must be terminated at each end using a matched **bus terminator**.

Bus couplers are produced as either **voltage mode** or **current mode** units depending upon whether they use voltage or current sensing devices. Within each LRU/avionics unit, an interface is provided that performs the required serial-to-parallel or parallel-to-serial data conversion, as shown in Fig. 3.32.

3.8.2 ARINC 429

The ARINC 429 data bus has proved to be one of the most popular bus standards used in commercial aircraft. The ARINC 429 specification defines the electrical and data characteristics and protocols that are used.

ARINC 429 employs a unidirectional data bus standard known as Mark 33 Digital Information Transfer System (DITS). Messages are transmitted in packets of 32-bits at a bit rate of either 12.5 or 100 kilobits per second (referred to as low and high bit rate respectively). Because the bus is unidirectional, separate ports, couplers and cables will be required when an LRU wishes to be able to both transmit and receive data. Note that a large number of bus connections may be required on an aircraft that uses sophisticated avionic systems.

ARINC 429 has been installed on a wide variety of commercial transport aircraft including; Airbus A310/A320 and A330/A340; Boeing 727, 737, 747, 757, and 767; and McDonnell Douglas MD-11. More modern aircraft (e.g. Boeing 777 and Airbus A380) use significantly enhanced bus specifications (see p. xx) in order to reduce the weight and size of cabling and to facilitate higher data rates than are possible with ARINC 429. Despite these moves to faster, bidirectional bus standards, the ARINC 429 standard has proved to be highly reliable and so is likely to remain in service for many years to come.

ARINC 429 is a two wire **differential** bus which can connect a single transmitter or source to one or more receivers. The term ‘differential’ simply means that neither of the two twisted wires is grounded and

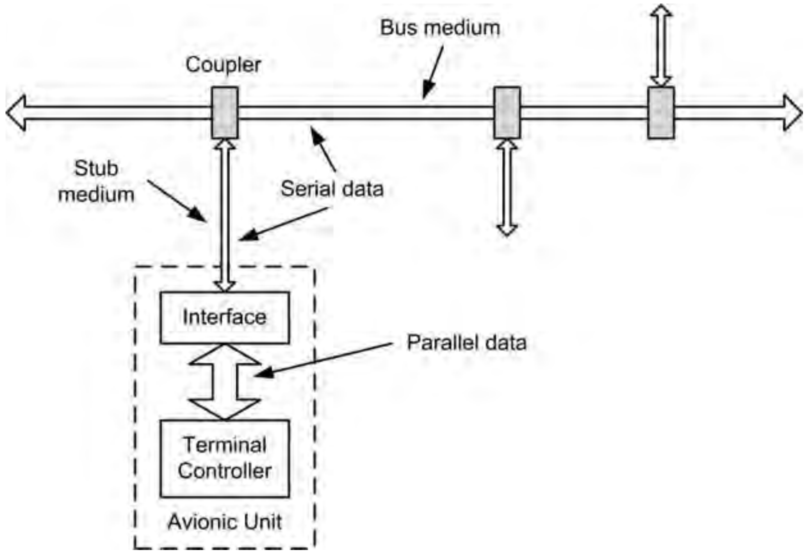

Figure 3.32 A basic bus interface

both convey signal voltages but of opposite polarity. This arrangement improves noise immunity which appears as a common mode signal induced on both of the conductors.

Two **bus speeds** are available, 12.5 kbps and 100 kbps. The data bus uses two signal wires to transmit 32 bit words. Transmission of sequential words is separated by at least four bit times of NULL (zero voltage). This eliminates the need for a separate clock signal and it makes the system **self-clocking**.

The ARINC 429 electrical characteristics are summarized below:

Voltage levels:	+5 V, 0 V, −5 V (each conductor with respect to ground) +10 V, 0 V, −10 V (conductor A with respect to conductor B)
Data encoding:	Bi-polar return to zero
Word size:	32 bits
Bit rate:	100 K bits per second (high), 12.5 K bits per second (low)

It is important to note that the received voltage on a serial bus depends on line length and the number of receivers connected to the bus. With ARINC 429, no more than 20 receivers should be connected to a single bus. Since each bus is unidirectional, a system needs to have its own transmit bus if it is required to respond or to send messages. Hence, to achieve bidirectional data transfer it is necessary to have two separate bus connections.

Since there can be only one transmitter on a twisted wire pair, ARINC 429 uses a very simple, point-to-point **protocol**. The transmitter is continuously sending 32-bit data words or is placed in the NULL state. Note that although there may only be one receiver on a particular bus cable the ARINC specification supports up to 20.

Other, faster and more sophisticated, bus systems are found on modern aircraft. They include ARINC 629 which supports a 20 Mbps data rate (20 times faster than ARINC 429), FDDI (Boeing's Fiber Distributed Data Interface), and 10 Mbps Ethernet.

Test your understanding 3.8

In relation to an aircraft bus system explain the meaning of each of the following terms:

(a) bus coupler
(b) bus terminator
(c) differential bus
(d) self-clocking
(e) NULL state.

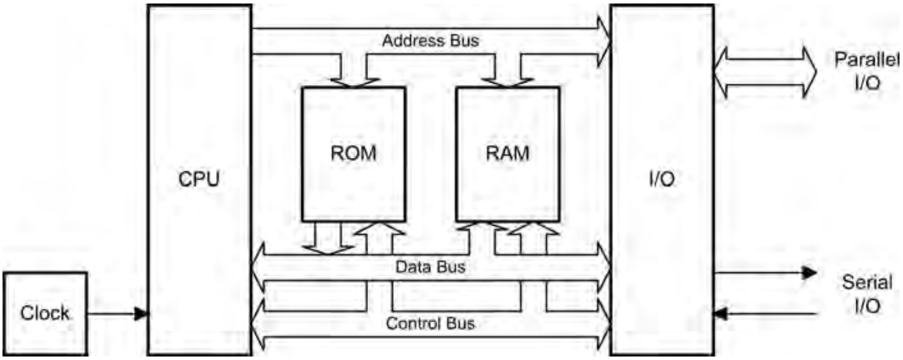

Figure 3.33 Basic components of a computer system

3.9 Computers

Modern aircraft use increasingly sophisticated avionic systems which involve the use of microprocessor-based computer systems. These systems combine hardware and software and are capable of processing large amounts of data in a very small time.

The basic components of a computer system are shown in Fig. 3.33. The main components are:

(a) a central processing unit (CPU)
(b) a memory, comprising both 'read/write' and 'read only' devices (commonly called RAM and ROM respectively)
(c) a means of providing input and output (I/O). For example, a keypad for input and a display for output.

In a microprocessor system the functions of the CPU are provided by a single very large scale integrated (VLSI) microprocessor chip. This chip is equivalent to many thousands of individual transistors.

Semiconductor devices are also used to provide the read/write and read-only memory. Strictly speaking, both types of memory permit 'random access' since any item of data can be retrieved with equal ease regardless of its actual location within the memory. Despite this, the term 'RAM' has become synonymous with semiconductor read/write memory.

The basic components of the system (CPU, RAM, ROM and I/O) are linked together using a multiple-wire connecting system know as a **bus** (see Fig. 3.33). Three different buses are present, these are:

(a) the **address bus** used to specify memory locations;
(b) the **data bus** on which data is transferred between devices; and
(c) the **control bus** which provides timing and control signals throughout the system.

The number of individual lines present within the address bus and data bus depends upon the particular microprocessor employed. Signals on all lines, no matter whether they are used for address, data, or control, can exist in only two basic states: logic 0 (**low**) or logic 1 (**high**). Data and addresses are represented by **binary numbers** (a sequence of 1 s and 0 s) that appear respectively on the data and address bus.

Some basic microprocessors designed for control and instrumentation applications have an 8-bit data bus and a 16-bit address bus. More sophisticated processors can operate with as many as 64 or 128 bits at a time.

The largest binary number that can appear on an 8-bit data bus corresponds to the condition when all eight lines are at logic 1. Therefore the largest value of data that can be present on the bus at any instant of time is equivalent to the binary number 11111111 (or 255). Similarly, most the highest address that can appear on a 16-bit address bus is 1111111111111111 (or 65,535). The full range of data values and addresses for a simple microprocessor of this type is thus:

Data	From	00000000
	To	11111111
Addresses	From	0000000000000000
	To	1111111111111111

Finally, a locally generated clock signal provides a time reference for controlling the transfer of synchronous data within the system. The clock signal usually

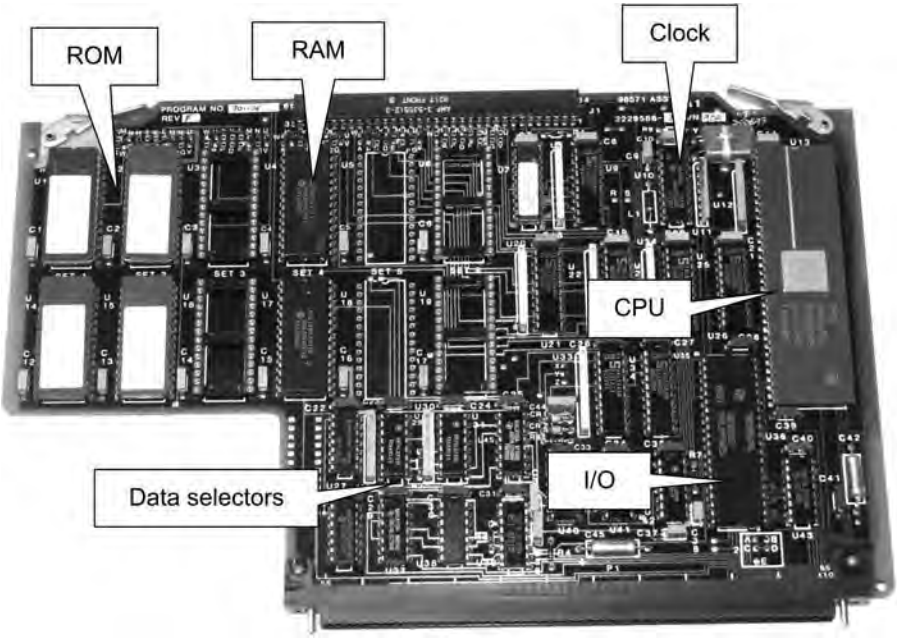

Figure 3.34 A typical aircraft computer system

consists of a high-frequency square wave pulse train derived from an accurate quartz crystal controlled oscillator.

Key point

A computer system consists of a central processing unit (CPU), a read-only memory (ROM), a read/write (random access) memory (RAM), and one or more input/output (I/O) devices. These elements are linked together using a local bus system that comprises an address bus, a data bus, and a control bus.

Key point

Depending on the size of their internal data bus, avionic computer systems usually manipulate data in groups of 8, 16 or 32 bits. When this data is obtained from an external serial bus this data must first be assembled into parallel form in order to facilitate internal processing.

3.9.1 Memories and data storage

The semiconductor read-only memory (ROM) within a microprocessor system provides storage for the program code as well as any permanent data that requires storage. All of this data is referred to as **non-volatile** because it remains intact when the power supply is disconnected.

The semiconductor RAM within a microprocessor system provides storage for the transient data and variables that are used by programs. Part of the RAM is also used by the microprocessor as a temporary store for data whilst carrying out its normal processing tasks.

It is important to note that any program or data stored in RAM will be lost when the power supply is switched off or disconnected. The only exception to this is low-power CMOS RAM that is kept alive by means of a small battery. This **battery-backed memory** is used to retain important data, such as the time and date.

When expressing the amount of storage provided by a memory device we usually use kilobytes (Kbyte). It is important to note that a kilobyte of memory is actually 1024 bytes (not 1000 bytes). The reason for choosing the Kbyte rather than the kbyte (1000 bytes) is that 1024 happens to be the nearest power of 2 (note that $2^{10} = 1024$).

The capacity of a semiconductor ROM is usually specified in terms of an address range and the number of bits stored at each address. For example, 2 K × 8 bits (capacity 2 Kbytes), 4 K × 8 bits (capacity 4 Kbytes), and so on. Note that it is not always necessary (or desirable) for the entire memory space of a computer to be populated by memory devices.

Key point

The term 'random access' simply refers to a memory device in which data may be retrieved from all locations with equal ease (i.e. access time is independent of actual memory address). This is important since our programs often involve moving sizeable blocks of data into and out of memory.

Key maintenance point

Stray static charges can very easily damage static-sensitive devices such as microprocessors and memory chips. Damage can be prevented by adopting the appropriate ESD procedures which usually involve using grounded wrist straps when handling chips and boards and using specially treated conductive packaging for transport and storage of component parts.

Test your understanding 3.9

In relation to a basic computer system explain the function of each of the following components:

(a) CPU
(b) clock
(c) RAM
(d) ROM
(e) I/O.

3.10 Multiple choice questions

1. The logic symbol shown in Fig. 3.35 is for:
   (a) a NAND gate
   (b) a NOR gate
   (c) an OR gate.

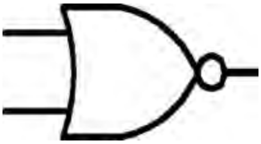

Figure 3.35 See Question 1

2. A logic 1 is present at the output of a two-input NOR gate. Which one of the following is true?
   (a) both of its inputs must be at logic 1
   (b) both of its inputs must be at logic 0
   (c) one or more of its inputs must be at logic 1.
3. The decimal number 10 is equivalent to the binary number:
   (a) 1000
   (b) 1010
   (c) 1100.
4. The BCD number 10001001 is equivalent to the decimal number:
   (a) 89
   (b) 141
   (c) 137.
5. The output of an R-S bistable is at logic 1. This means that the bistable is:
   (a) set
   (b) reset
   (c) in a high impedance state.
6. In a volatile memory:
   (a) the memory is lost as soon as power is removed
   (b) the memory is retained indefinitely
   (c) the memory needs to be refreshed constantly, even when power is on.
7. The main elements of a computer system are:
   (a) CPU, memory and I/O ports
   (b) ALU, registers and a ROM
   (c) RAM, ROM and I/O.
8. A NOR gate with both inputs inverted becomes a
   (a) NAND gate
   (b) AND gate
   (c) OR gate.
9. Another name for a data selector is:
   (a) a monostable
   (b) a multiplexer
   (c) an encoder.

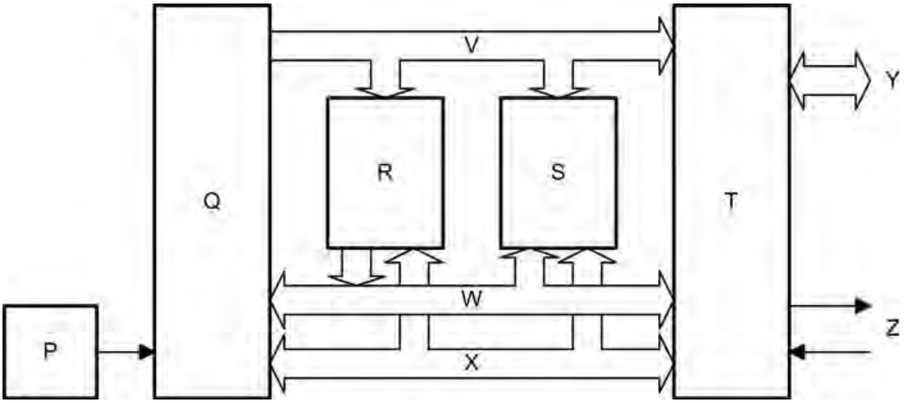

Figure 3.36 See Question 15

10. A group of bits transmitted one after another is:
    (a) a clock signal
    (b) parallel data
    (c) serial data.

11. A bus that supports the transfer of data in both directions is referred to as:
    (a) universal
    (b) bidirectional
    (c) asynchronous.

12. The main advantage of using a serial bus in an aircraft is:
    (a) there is no need for data conversion
    (b) it supports the highest possible data rates
    (c) reduction in the size and weight of cabling.

13. The maximum speed of an ARINC 429 bus is:
    (a) 12.5 Kbps
    (b) 100 Kbps
    (c) 10 Mbps.
14. What is the maximum number of receivers that may be connected to an ARINC 429 bus cable?
    (a) 1
    (b) 16
    (c) 20.
15. The feature marked Q in Fig. 3.36 is:
    (a) CPU
    (b) RAM
    (c) I/O.

This page intentionally left blank

Chapter 4

Generators and motors

Used in conjunction with an aircraft battery, engine-driven generators are a primary source of electrical power in an aircraft. Generators can supply either direct or alternating current (DC or AC), as appropriate to the needs of an individual aircraft type. Motors (which can also be either DC or AC types) are fitted in aircraft in order to satisfy a wide range of needs. Generators and motors share many similarities and this chapter provides an introduction to their operating principles by looking in more detail at the construction and operation of these indispensable aircraft electrical components. The chapter also includes a brief introduction to three-phase AC supplies, including the theoretical and practical aspects of their generation and distribution.

4.1 Generator and motor principles

In Chapter 1 we introduced the concept of electromagnetic induction. Put simply, this is the generation of an e.m.f. across the ends of a conductor when it passes through a change in magnetic flux. In a similar fashion, an e.m.f. will appear across the ends of a conductor if it remains stationary whilst the field moves. In either case, the action of cutting through the lines of magnetic flux results in a generated e.m.f. – see Fig. 4.1. The amount of e.m.f., e, *induced* in the conductor will be directly proportional to:

- the density of the magnetic flux, B, measured in tesla (T)
- the effective length of the conductor, l, within the magnetic flux
- the speed, v, at which the lines of flux cut through the conductor measured in metres per second (m/s)
- the sine of the angle, θ, between the conductor and the lines of flux.

The induced e.m.f. is given by the formula:

$$e = Blv \sin \theta$$

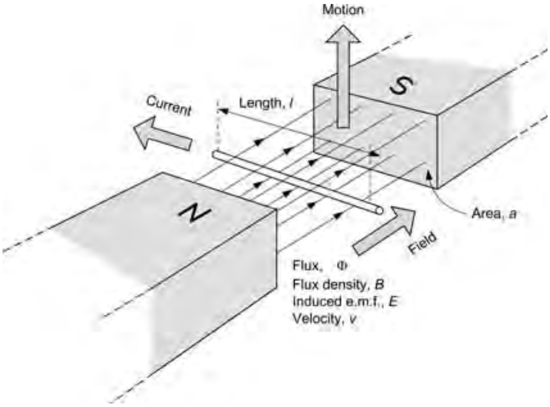

Figure 4.1 A conductor moving inside a magnetic field

Note that if the conductor moves at right angles to the field (as shown in Fig. 4.1) maximum e.m.f. will be induced. Conversely, if the conductor moves along the lines of flux (instead of at right angles) the induced e.m.f. will be zero.

Electricity and magnetism often work together to produce motion. In an electric **motor**, current flowing in a conductor placed inside a magnetic field produces motion. A **generator**, on the other hand, produces a voltage when a conductor is moved inside a magnetic field. These two effects are, as you might suspect, closely related to one another and they are vitally important in the context of aircraft electrical systems!

Key point

An e.m.f. will be induced across the ends of a conductor when there is relative motion between it and a magnetic field. The induced voltage will take its greatest value when moving at right angles to the magnetic field lines and its least value (i.e. zero) when moving along the direction of the field lines.

4.1.1 A simple AC generator

Being able to generate a voltage by moving a conductor through a magnetic field is extremely useful as it provides us with an easy way of generating electricity. Unfortunately, moving a wire at a constant linear velocity through a uniform magnetic field presents us with a practical problem simply because the mechanical power that can be derived from an aircraft engine is available in rotary (rather than linear) form!

The solution to this problem is that of using the rotary power available from the engine (via a suitable gearbox and transmission) to rotate a conductor shaped into the form of loop as shown in Fig. 4.2. The loop is made to rotate inside a permanent magnetic field with opposite poles (N and S) on either side of the loop.

There now remains the problem of making contact with the loop as it rotates inside the magnetic field but this can be overcome by means of a pair of carbon **brushes** and copper **slip-rings**. The brushes are spring loaded and held against the rotating slip-rings so that, at any time, there is a path for current to flow from the loop to the load to which it is connected.

The opposite sides of the loop consist of conductors that move through the field. At 0° (with the loop vertical as shown at A in Fig. 4.4) the opposite sides of the loop will be moving in the same direction as the lines of flux. At that instant, the angle, θ, at which the field is cut is 0° and since the sine of 0° is 0 the generated voltage (from $E = Blv \sin\theta$) will consequently also be zero.

If the loop has rotated to a position which is 90° (position B in Fig. 4.4) the two conductors will effectively be moving at right angles to the field. At that instant, the generated e.m.f. will take a maximum value (since the sine of 90° is 1).

At 180° from the starting position the generated e.m.f. will have fallen back to zero since, once again, the conductors are moving along the flux lines (but in the direction opposite to that at 0 °, as shown in C of Fig. 4.4).

At 270° the conductors will once again be moving in a direction which is perpendicular to the flux lines (but in the direction opposite to that at 90°). At this point (D of Fig. 4.4), a maximum generated e.m.f. will once again be produced. It is, however, important to note that the e.m.f. generated at this instant will be of opposite polarity to that which was generated at 90°.The reason for this is simply that the relative direction of motion (between the conductors and flux lines) has effectively been reversed.

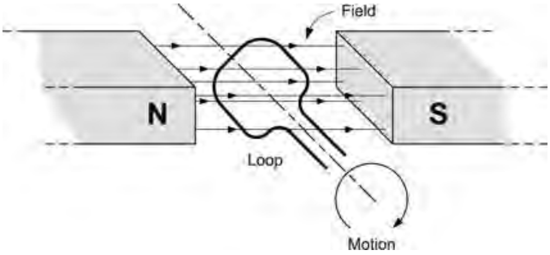

Figure 4.2 A loop rotating within a magnetic field

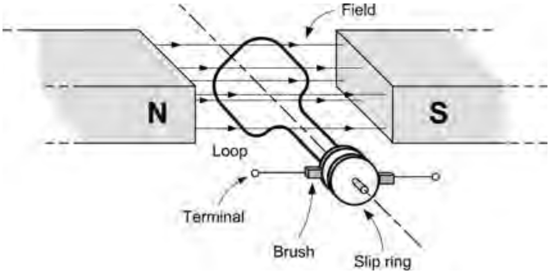

Figure 4.3 Brush arrangement

Since $E = Blv \sin\theta$, the e.m.f. generated by the arrangement shown in Fig. 4.4 will take a sinusoidal form, as shown in Fig. 4.5. Note that the maximum values of e.m.f. occur at 90° and 270° and that the generated voltage is zero at 0°, 180° and 360°.

In practice, the single loop shown in Fig. 4.2 would comprise a coil of wire wound on a suitable non-magnetic former. This coil of wire effectively increases the length of the conductor within the magnetic field and the generated e.m.f. will then be directly proportional to the number of turns on the coil.

Key point

In a simple AC generator a loop of wire rotates inside the magnetic field produced by two opposite magnetic poles. Contact is made to the loop as it rotates by means of slip-rings and brushes.

4.1.2 DC generators

When connected to a load, the simple generator shown in Fig. 4.3 produces a sinusoidal alternating current (AC) output. In many applications a steady direct current (DC) output may be preferred.

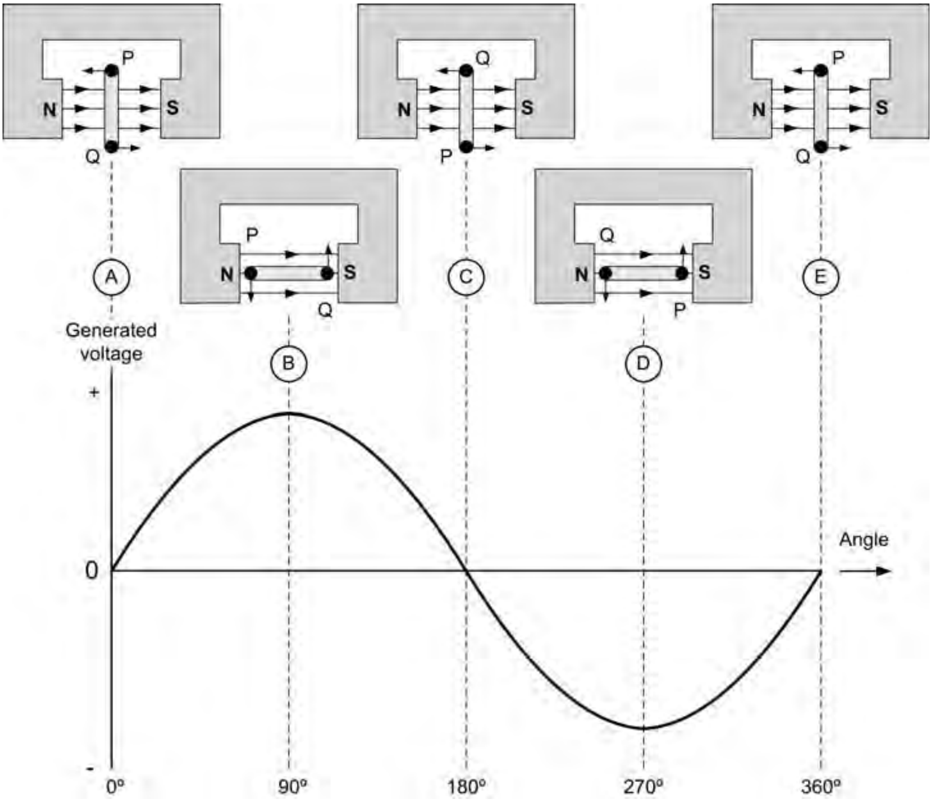

Figure 4.4 E.m.f. generated at various angles

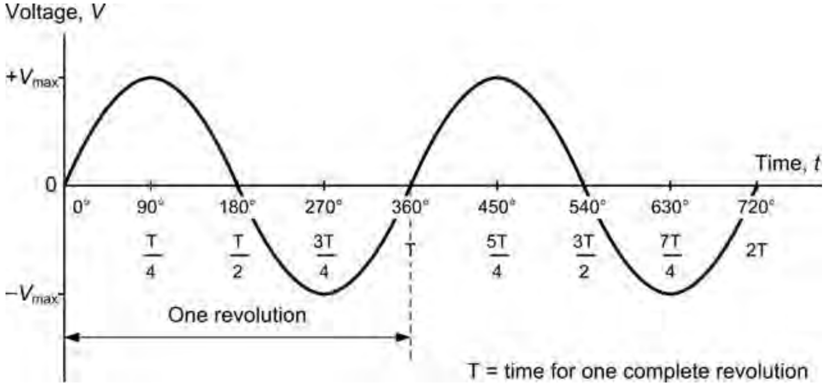

Figure 4.5 Sinusoidal voltage produced by the rotating loop

This can be achieved by modifying the arrangement shown in Fig. 4.3, replacing the brushes and slip-rings with a **commutator** arrangement, as shown in Fig. 4.6. The commutator arrangement functions as a rotating reversing switch which ensures that the e.m.f. generated by the loop is reversed after rotating through 180 °. The generated e.m.f. for this arrangement is shown in Fig. 4.7. It's worth comparing this waveform with that shown in Fig. 4.5—you should be able to spot the difference immediately!

The generated e.m.f. shown in Fig. 4.7, whilst **unipolar** (i.e. having only one polarity, either all positive or all negative), is clearly far from ideal since a DC power source should provide a constant voltage output rather than a series of pulses. One way of overcoming this problem is with the use of a second loop (or coil) at right angles to the first, as shown in Fig. 4.8. The commutator is then divided into four (rather than two) **segments** and the generated e.m.f. produced by this arrangement is shown in Fig. 4.9.

In real generators, a coil comprising a large number of turns of copper wire replaces the single-turn rotating loop. This arrangement effectively increases the total length of the conductor within the magnetic field and, as a result, also increases the generated output voltage. The output voltage also depends on the density of the magnetic flux through which the current-carrying conductor passes. The denser the field the greater the output voltage will be.

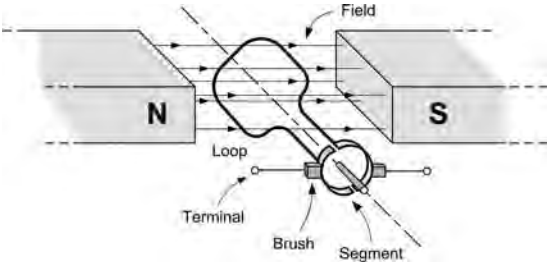

Figure 4.6 Commutator arrangement

Key point

A simple DC generator uses an arrangement similar to that used for an AC generator but with the slip-rings and brushes replaced by a commutator that reverses the current produced by the generator every 180°.

4.1.3 DC generators

A simple DC motor consists of a very similar arrangement to that of the DC generator that we met earlier.

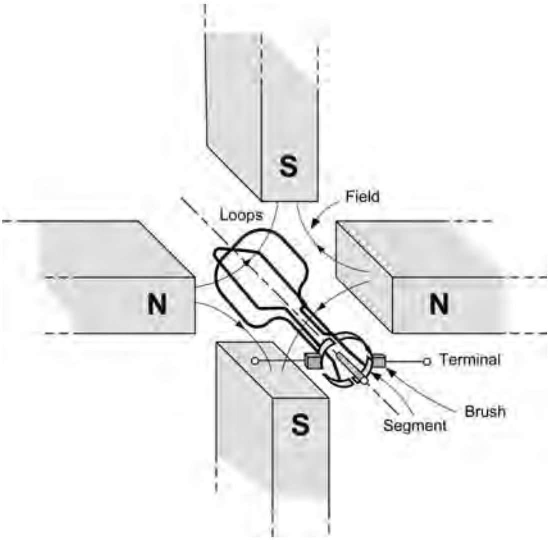

Figure 4.8 An improved DC generator

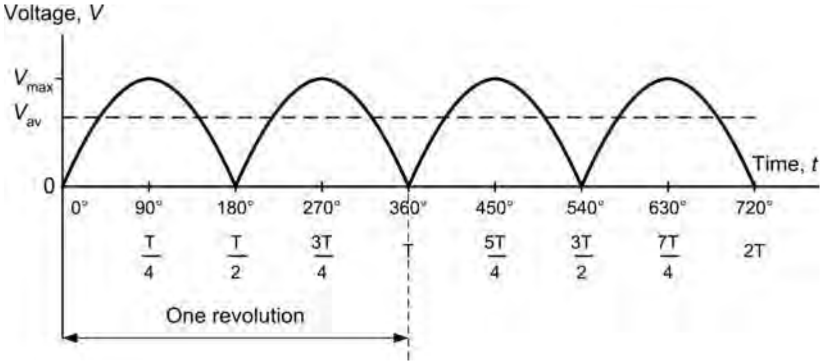

Figure 4.7 E.m.f. generated (compare with Fig. 4.5)

A loop of wire that is free to rotate is placed inside a permanent magnetic field (see Fig. 4.10). When a DC current is applied to the loop of wire, two equal and opposite forces are set up which act on the conductor in the directions indicated in Fig. 4.10.

The direction of the forces acting on each arm of the conductor can be established by again using the right-hand grip rule and Fleming's left-hand rule. Now because the conductors are equidistant from their pivot point and the forces acting on them are *equal and opposite*, then they form a **couple**. The **moment** of this couple is equal to the magnitude of a single force multiplied by the distance between them and this moment is known as **torque**, T. Now,

$$T = Fd$$

where T is the torque (in newton-metres, Nm), F is the force (N) and d is the distance (m).

We already know that the magnitude of the force F is given by $F = BIl$, therefore the torque produced by the current-carrying thus the torque expression can be written:

$$T = BIld$$

where T is the torque (Nm), B is the flux density (T), I is the current (A), l is the length of conductor in the magnetic field (m), and d is the distance (m).

The torque produces a turning moment such that the coil or loop rotates within the magnetic field. This rotation continues for as long as a current is applied. A more practical form of DC motor consists of a rectangular coil of wire (instead of a single turn loop of wire) mounted on a former and free to rotate about a shaft in a permanent magnetic field, as shown in Fig. 4.11.

In real motors, this rotating coil is know as the **armature** and consists of many hundreds of turns of conducting wire. This arrangement is needed in order to maximize the force imposed on the conductor by introducing the *longest possible* conductor into the

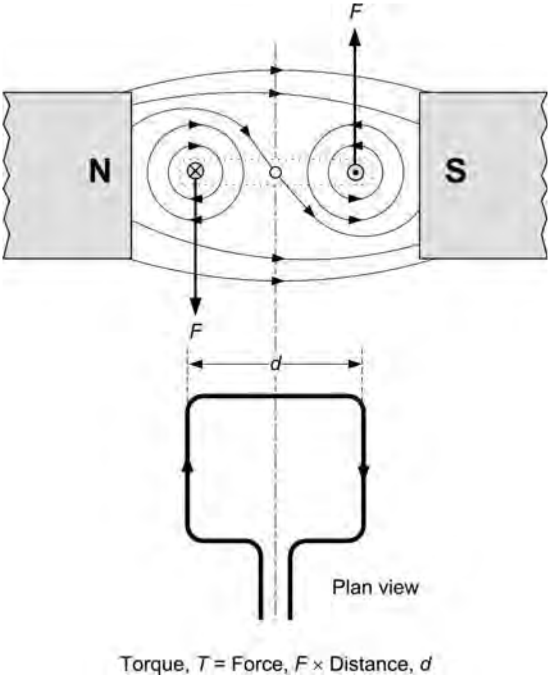

Figure 4.10 Torque on a current-carrying loop suspended within a permanent magnetic field

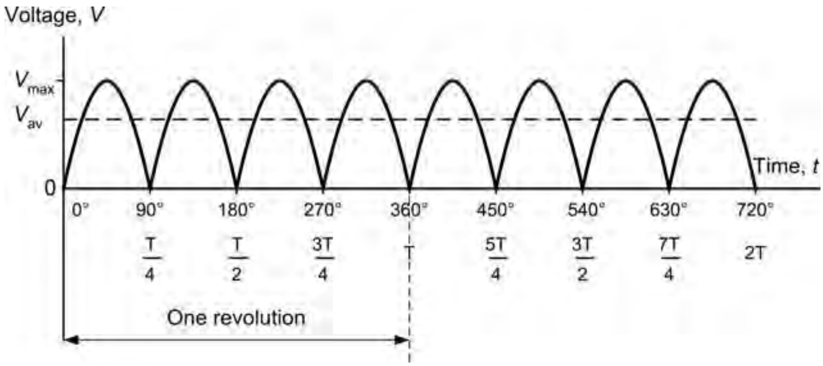

Figure 4.9 E.m.f. generated (compare with Fig. 4.7)

magnetic field. Also, from the relationship $F = BIl$ it can be seen that the force used to provide the torque in a motor is directly proportional to the size of the magnetic flux, B. Instead of using a permanent magnet to produce this flux, in a real motor, an electromagnet is used. Here an electromagnetic field is set up using the **solenoid** principle (Fig. 4.12). A long length of conductor is wound into a coil consisting of many turns and a current passed through it. This arrangement constitutes a **field winding** and each of the turns in the field winding assists each of the other turns in order to produce a strong magnetic field, as shown in Fig. 4.12.

As in the case of the DC generator, this field may be intensified by inserting a ferromagnetic core inside the coil. Once the current is applied to the conducting coil, the core is magnetized and all the time the current is on it acts in combination with the coil to produce a permanent magnet, having its own N–S poles.

Now returning to the simple motor illustrated in Fig. 4.11, we know that when current is supplied to the rotating armature (**rotor**) a torque is produced. In order to produce continuous rotary motion, this torque (turning moment) must always act in the same direction.

Therefore, the current in each of the armature conductors must be reversed as the conductor passes between the north and south magnetic field poles. The commutator acts like a rotating switch, reversing the current in each armature conductor at the appropriate time to achieve this continuous rotary motion. Without the presence of a commutator in a DC motor, only a half turn of movement is possible!

In Fig. 4.13(a) the rotation of the armature conductor is given by Fleming's left-hand rule (see Fig. 4.14). When the coil reaches a position mid-way way between the poles (Fig. 4.13(b)), no rotational torque is produced in the coil. At this stage the commutator reverses the current in the coil. Finally (Fig. 4.13(c)) with the current reversed, the motor torque now continues to rotate the coil in its original direction.

Key point

The torque produced by a DC motor is directly proportional to the product of the current flowing in the rotating armature winding.

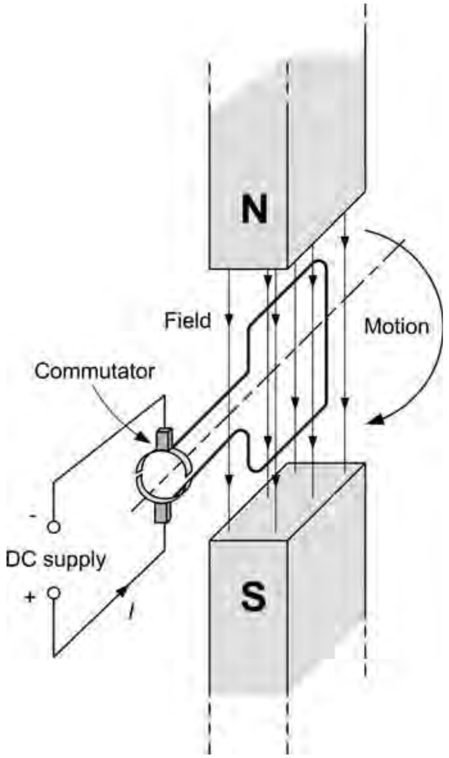

Figure 4.11 Simple electric motor with commutator

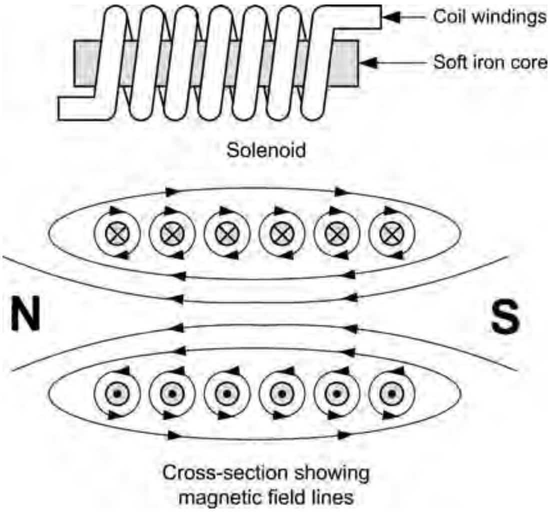

Figure 4.12 Magnetic field produced by a solenoid

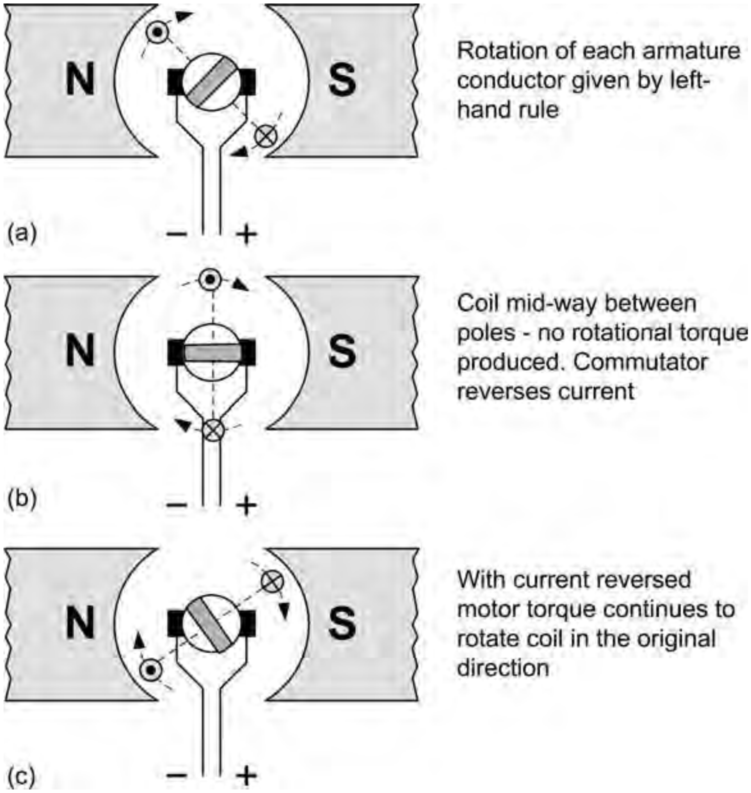
N
S
Rotation of each armature conductor given by left-hand rule
(a)
−
+
N
S
Coil mid-way between poles - no rotational torque produced. Commutator reverses current
(b)
−
+
N
S
With current reversed motor torque continues to rotate coil in the original direction
(c)
−
+

Figure 4.13 Action of the commutator

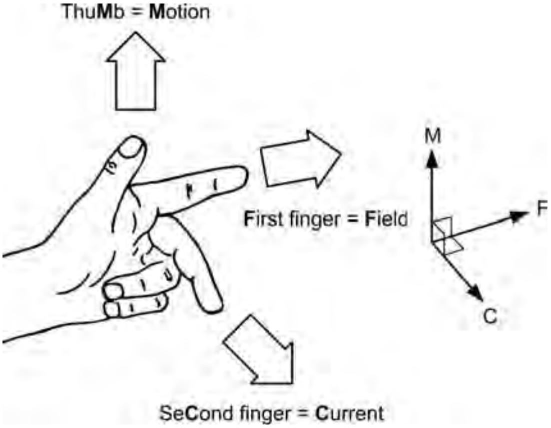
ThuMb = Motion
First finger = Field
SeCond finger = Current
M
F
C

Figure 4.14 Fleming's left-hand rule

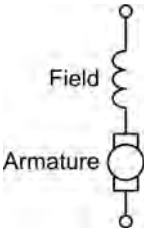
Field
Armature

Figure 4.15 Series-wound DC motor

4.1.4 Field connections

The field winding of a DC motor can be connected in various different ways according to the application envisaged for the motor in question. The following configurations are possible:

- series-wound
- shunt-wound
- compound-wound (where both series and shunt windings are present).

In the **series-wound** DC motor the field winding is connected in series with the armature and the full armature current flows through the field winding (see Fig. 4.15). This arrangement results in a DC motor

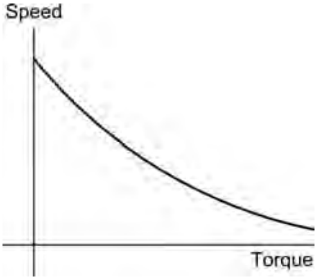

Figure 4.16 Typical torque and speed characteristic for a series-wound DC motor

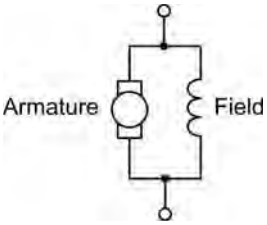

Figure 4.17 Shunt-wound DC motor

that produces a large starting torque at slow speeds. This type of motor is ideal for applications where a heavy load is applied from rest. The disadvantage of this type of motor is that on light loads the motor speed may become excessively high. For this reason this type of motor should not be used in situations where the load may be accidentally removed. A typical set of torque and speed characteristics for a series-wound DC motor is shown in Fig. 4.16.

In the **shunt-wound** DC motor the field winding is connected in parallel with the armature and thus the supply current is divided between the armature and the field winding (see Fig. 4.17). This arrangement results in a DC motor that runs at a reasonably constant speed over a wide variation of load but does not perform well when heavily loaded. A typical set of torque and speed characteristics for a shunt-wound DC motor is shown in Fig. 4.18.

The **compound-wound** DC motor has both series and shunt field windings (see Fig. 4.19) and is therefore able to combine some of the advantages of each type of motor. A typical set of torque and speed characteristics for a compound-wound DC motor is shown in Fig. 4.20.

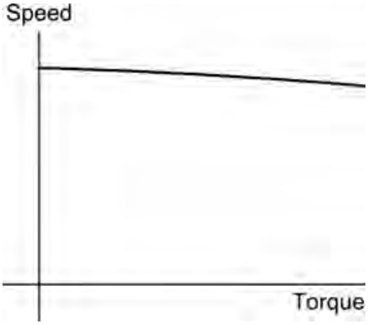

Figure 4.18 Typical torque and speed characteristic for a shunt-wound DC motor

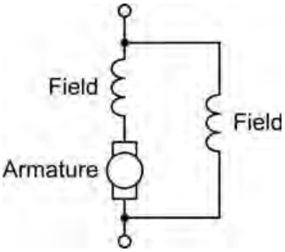

Figure 4.19 Compound-wound DC motor

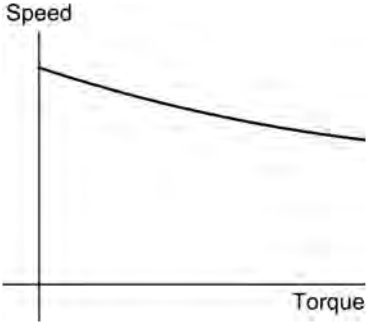

Figure 4.20 Typical torque and speed characteristic for a compound-wound DC motor

---

Key point

In order to avoid the need for a large permanent magnet in a DC machine (i.e. a motor or generator), a separate field winding is used. This winding is energized with direct current. In the case of a DC generator, this current can be derived from the output of the generator (in which case it is referred to as self-excited generator) or it can be energized from a separate DC supply.

Key point

There is no real difference between the construction of a DC generator and that of a DC motor. The only significant distinction is that in a generator rotational mechanical energy is being converted to electrical energy whilst in the case of a motor the converse applies.

Key point

The output of a generator can be regulated by controlling the intensity of its field: increasing the field current will produce a corresponding increase in output voltage whereas reducing the field current will produce a corresponding reduction in output.

4.1.5 Starter-generators

Some small aircraft starter generators eliminate the need for separate engine starter motors and DC generators. They usually have separate field windings (one for the starter motor and one for the generator) together with a common armature winding. When used for starting, the starter-generator is connected as a series-wound DC motor capable of producing a very high starting torque. However, when used as a generator the connections are changed so that the unit operates as shunt-wound generator producing reasonably constant current over a wide range of speed.

In the start condition, the low-resistance starter field and common armature windings of the starter-generator are connected in series across the DC supply via a set of contactors. This arrangement ensures that a torque is produced that is sufficient to start an aircraft's turbine engine.

When the engine reaches self-sustaining speed, the current is broken through the first set of contactors and a second set of contactors operate, removing the external DC power supply from the starter-generator and reconnecting the arrangement so that the generated the armature voltage is fed to the higher-resistance shunt field and the aircraft's main voltage regulator.

The advantage of this arrangement is not only that the starter-generator replaces two individual machines (i.e. a starter and a generator) with consequent savings in size and weight, but additionally that only a single mechanical drive is required between the engine and the starter-generator unit. The disadvantage of this arrangement is that the generator output is difficult to maintain at low engine r.p.m. and therefore starter-generators are mainly found on turbine-powered aircraft that maintain a relatively high engine r.p.m.

Test your understanding 4.1

1. In relation to a simple DC machine, explain the meaning of each of the following terms:
   (a) slip-rings
   (b) brushes
   (c) field winding.
2. Explain the advantages and disadvantages of series-wound DC motors compared with shunt-wound DC motors.

4.2 AC generators

AC generators, or **alternators**, are based on the principles that relate to the simple AC generator that we met earlier in Section 4.1.1. However, in a practical AC generator the magnetic field is rotated rather than the conductors from which the output is taken. Furthermore, the magnetic field is usually produced by a rotating electromagnet (the *rotor*) rather than a permanent magnet. There are a number of reasons for this including:

(a) the conductors are generally lighter in weight than the magnetic field system and are thus more easily rotated
(b) the conductors are more easily insulated if they are stationary
(c) the currents which are required to produce the rotating magnetic field are very much smaller than those which are produced by the conductors. Hence the slip-rings are smaller and more reliable.

Figure 4.22 shows the simplified construction of a **single-phase AC generator**. The *stator* consists of five coils of insulated heavy gauge wire located in

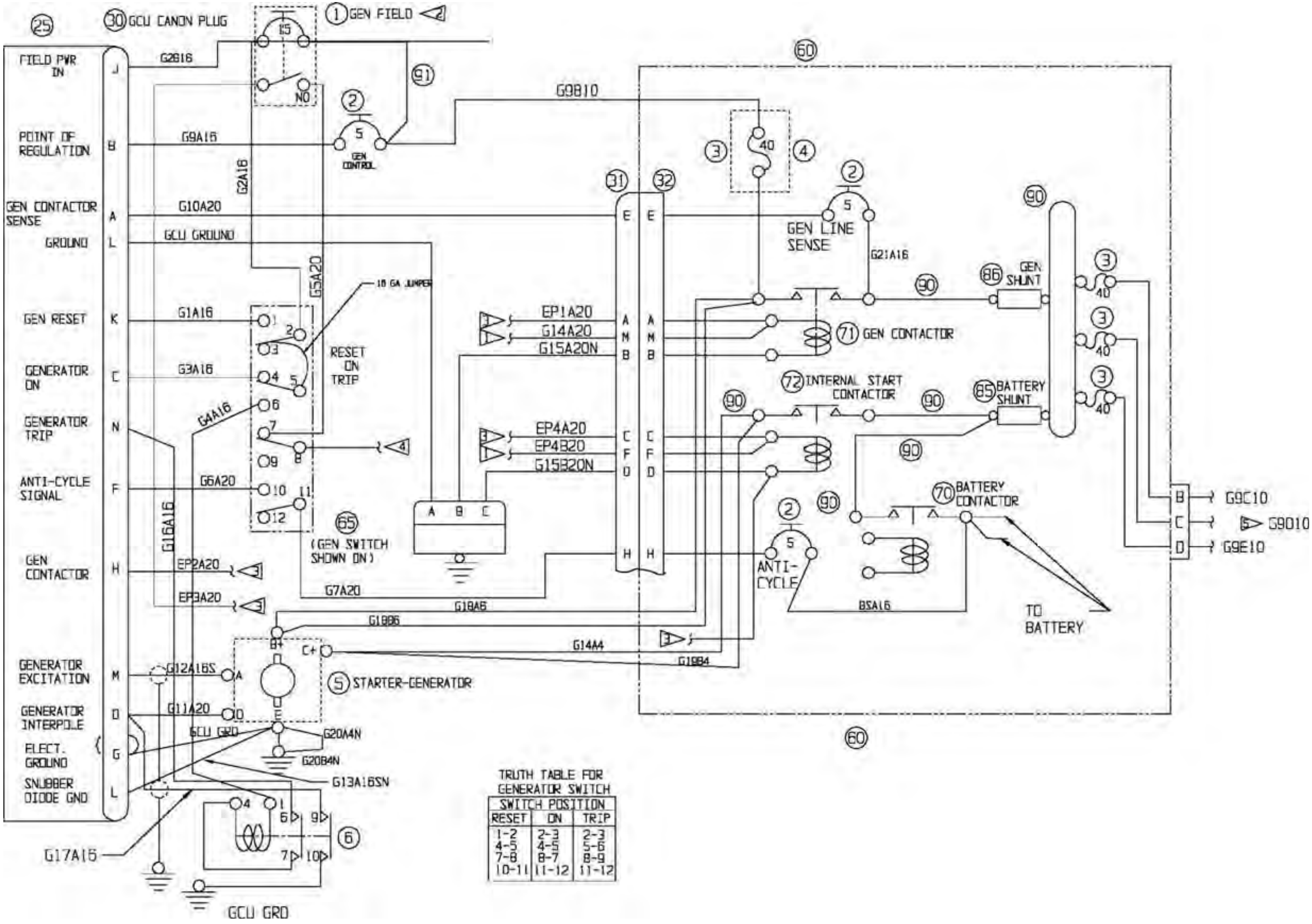
25
30 GCU CANON PLUG
1 GEN FIELD
2
15
NO
91
60
FIELD PWR IN
J
G2B16
POINT OF REGULATION
B
G9A16
5
GEN CONTROL
G9B10
G2A16
3
40
4
GEN CONTACTOR SENSE
A
G10A20
GROUND
L
GCU GROUND
31
32
E
GEN LINE SENSE
G21A16
90
G5A20
18 GA JUMPER
86 GEN SHUNT
GEN RESET
K
G1A16
RESET ON TRIP
EP1A20
G14A20
G15A20N
M
71 GEN CONTACTOR
GENERATOR ON
C
G3A16
G4A16
72 INTERNAL START CONTACTOR
85 BATTERY SHUNT
GENERATOR TRIP
N
EP4A20
EP4B20
G15B20N
F
D
ANTI-CYCLE SIGNAL
G6A20
65 (GEN SWITCH SHOWN ON)
70 BATTERY CONTACTOR
G9C10
G9D10
G9E10
GEN CONTACTOR
H
G16A16
EP2A20
EP3A20
G7A20
ANTI-CYCLE
B5A16
G18A6
G18B6
TO BATTERY
B+
C+
G14A4
G19B4
GENERATOR EXCITATION
M
G12A16S
5 STARTER-GENERATOR
GENERATOR INTERPOLE
G11A20
ELECT. GROUND
G
GCU GRD
G20A4N
G20B4N
G13A16SN
SNUBBER DIODE GND
6
G17A16
GCU GRD
TRUTH TABLE FOR GENERATOR SWITCH

SWITCH POSITION		
RESET	ON	TRIP
1-2	2-3	2-3
4-5	4-5	5-6
7-8	8-7	8-9
10-11	11-12	11-12

Figure 4.21 Starter-generator circuit showing changeover contactors

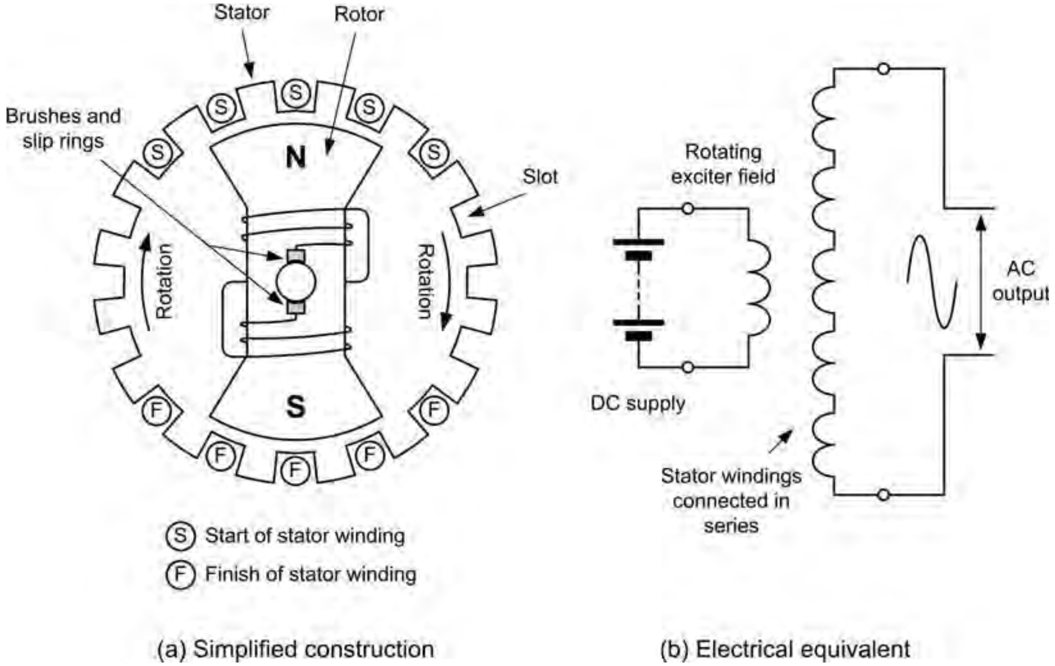
Stator
Rotor
Brushes and slip rings
N
S
Slot
Rotation
Rotation
S
F
Ⓢ Start of stator winding
Ⓕ Finish of stator winding
Rotating exciter field
DC supply
Stator windings connected in series
AC output

(a) Simplified construction (b) Electrical equivalent

Figure 4.22 Simplified construction of a single-phase AC generator

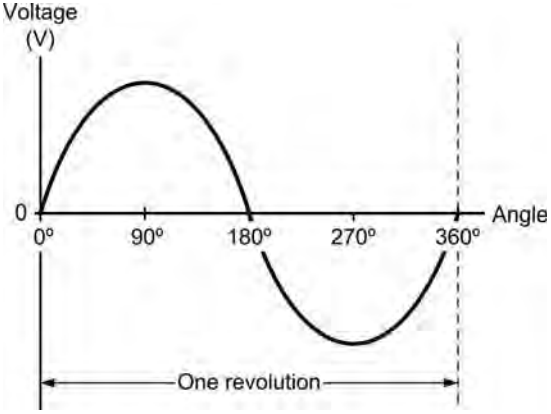

Figure 4.23 Output voltage produced by the single-phase AC generator shown in Fig. 4.22

slots in the high-permeability laminated core. These coils are connected in series to make a single stator winding from which the output voltage is derived.

The two-pole rotor comprises a field winding that is connected to a DC field supply via a set of slip-rings and brushes. As the rotor moves through one complete revolution the output voltage will complete one full cycle of a sine wave, as shown in Fig. 4.23.

By adding more pairs of poles to the arrangement shown in Fig. 4.22 it is possible to produce several cycles of output voltage for one single revolution of the rotor. The frequency of the output voltage produced by an AC generator is given by:

$$f = \frac{pN}{60}$$

where f is the frequency of the induced e.m.f. (in Hz), p is the number of pole pairs, and N is the rotational speed (in rev/min).

Example 4.1

An alternator is to produce an output at a frequency of 60 Hz. If it uses a 4-pole rotor, determine the shaft speed at which it must be driven.

Re-arranging $f = \frac{pN}{60}$ to make N the subject gives:

$$N = \frac{60f}{p}$$

A 4-pole machine has 2 pairs of poles, thus $p = 2$ and:

$$N = \frac{60 \times 60}{2} = 1{,}800 \text{ rev/min}$$

Key point

In a practical AC generator, the magnetic field excitation is produced by the moving rotor whilst the conductors from which the output is taken are stationary and form part of the stator.

4.2.1 Two-phase AC generators

By adding a second stator winding to the single-phase AC generator shown in Fig. 4.22, we can produce an alternator that produces two separate output voltages which will differ in phase by 90°. This arrangement is known as a two-phase AC generator.

When compared with a single-phase AC generator of similar size, a two-phase AC generator can produce more power. The reason for this is attributable to the fact that the two-phase AC generator will produce two positive and two negative pulses per cycle whereas the single-phase generator will only produce one positive and one negative pulse. Thus, over a period of time, a multi-phase supply will transmit a more evenly distributed power and this, in turn, results in a higher overall efficiency.

Key point

Three-phase AC generators are more efficient and produce more constant output than comparable single-phase AC generators.

4.2.2 Three-phase AC generators

The three-phase AC generator has three individual stator windings, as shown in Fig. 4.26. The output voltages produced by the three-phase AC generator are spaced by 120° as shown in Fig. 4.27. Each phase can be used independently to supply a different load or the generator outputs can be used with a three-phase distribution system like those described in the next section. In a practical three-phase system the three output voltages are identified by the colours red, yellow, and blue or by letters A, B, and C respectively.

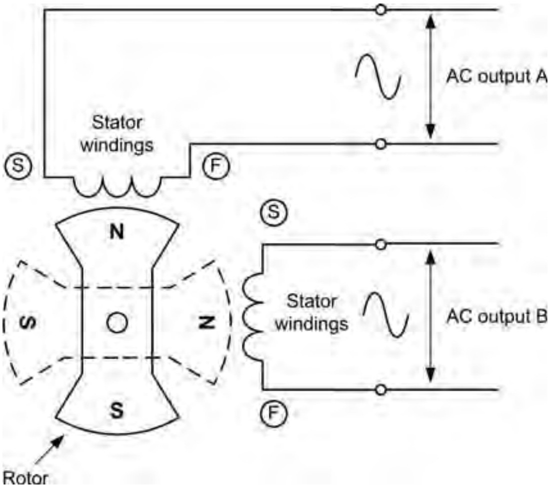

Figure 4.24 Simplified construction of a two-phase AC generator

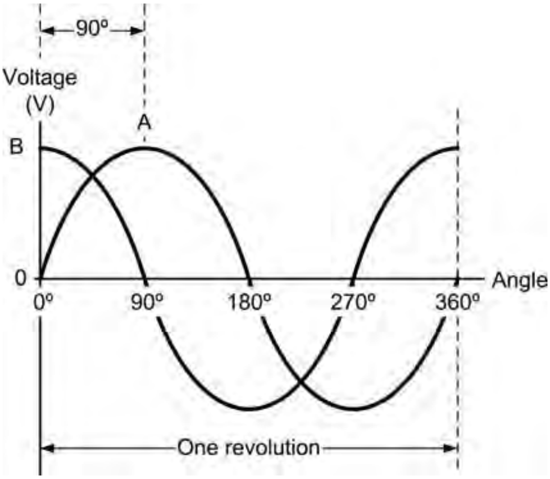

Figure 4.25 Output voltage produced by the two-phase AC generator shown in Fig. 4.24

Test your understanding 4.2

An alternator uses a 12-pole rotor and is to operate at a frequency of 400 Hz. At what speed must it be driven?

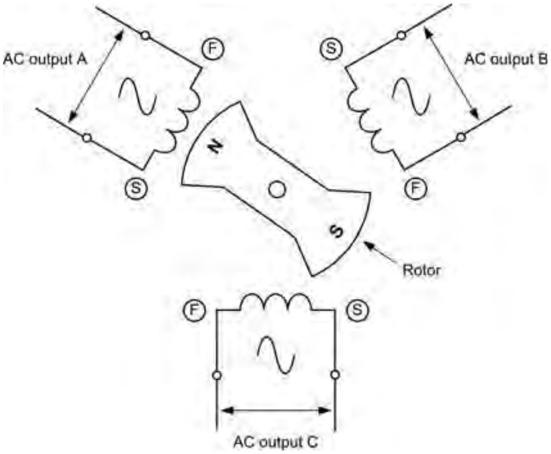

Figure 4.26 Simplified construction of a three-phase AC generator

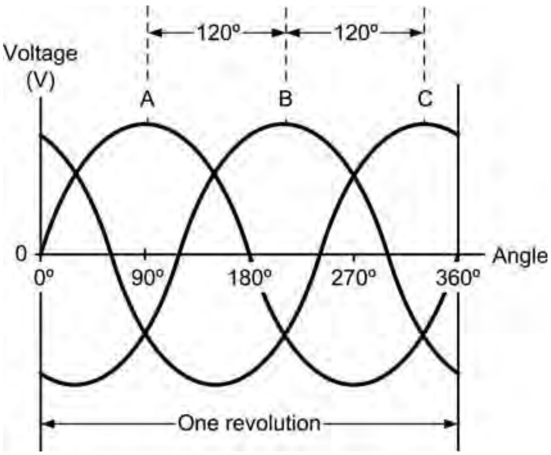

Figure 4.27 Output voltage produced by the three-phase AC generator shown in Fig. 4.26

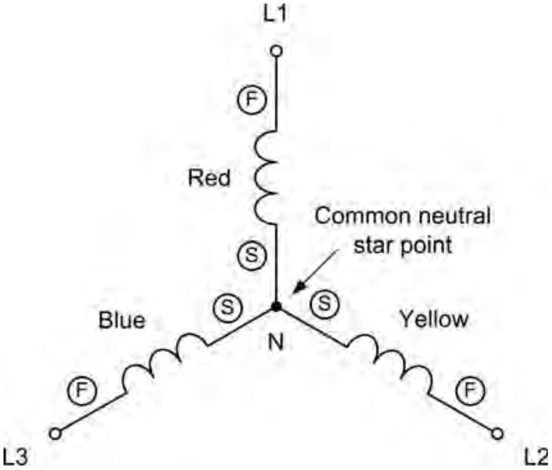

Figure 4.28 Star connection

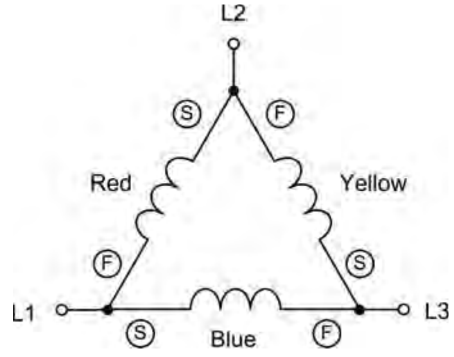

Figure 4.29 Delta connection

4.3 Three-phase generation and distribution

When three-phase supplies are distributed there are two basic methods of connection:

- star (as shown in Fig. 4.28)
- delta (as shown in Fig. 4.29)

A complete star-connected three-phase distribution system is shown in Fig. 4.30. This shows a three-phase AC generator connected a three-phase load. Ideally, the load will be **balanced** in which case all three load resistances (or impedances) will be identical.

The relationship between the line and phase voltages shown in Fig. 4.30 can be determined from the **phasor diagram** shown in Fig. 4.31. This diagram shows the relative directions of the three alternating phase voltages (V_P) and the voltages between the lines (V_L). From this diagram it is important to note that three line voltages are 120° apart and that the line voltages lead the phase voltages by 30°. In order to obtain the relationship between the line voltage, V_L, and the phase voltage, V_P, we need to resolve any one of the triangles, from which we find that:

$$V_L = 2(V_P \times \cos 30°)$$

Now

$$\cos 30° = \frac{\sqrt{3}}{2}$$

and hence:

$$V_L = 2\left(V_P \times \frac{\sqrt{3}}{2}\right)$$

from which:

$$V_L = \sqrt{3}\ V_P$$

Note also that the phase current is the same as the line current, hence:

$$I_P = I_L$$

An alternative, delta-connected three-phase distribution system is shown in Fig. 4.32. Once again this shows a three-phase AC generator connected a three-phase load. Here again, the load will ideally be balanced in which case all three load resistances (or impedances) will be identical.

In this arrangement the three line currents are 120° apart and that the line currents lag the phase currents by 30°. Using a similar phasor diagram to that which we used earlier, we can show that:

$$I_L = \sqrt{3}\ I_P$$

It should also be obvious that:

$$V_P = V_L$$

Example 4.2

In a star-connected three-phase system the phase voltage is 105 V. Determine the line voltage.

$$V_L = \sqrt{3}\ V_P = \sqrt{3} \times 115 = 199.2\,\text{V}$$

Example 4.3

In a delta-connected three-phase system the line current is 6 A. Determine the phase current.

$$I_L = \sqrt{3}\ I_P$$

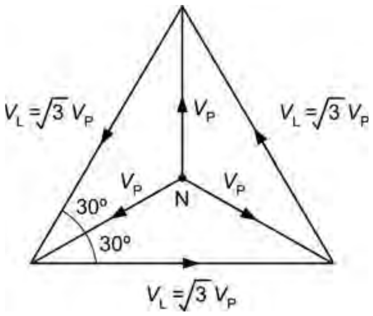

$V_L = \sqrt{3}\ V_P$
V_P
$V_L = \sqrt{3}\ V_P$
V_P
V_P
N
30°
30°
$V_L = \sqrt{3}\ V_P$

Figure 4.31 Phasor diagram for the three-phase star-connected system

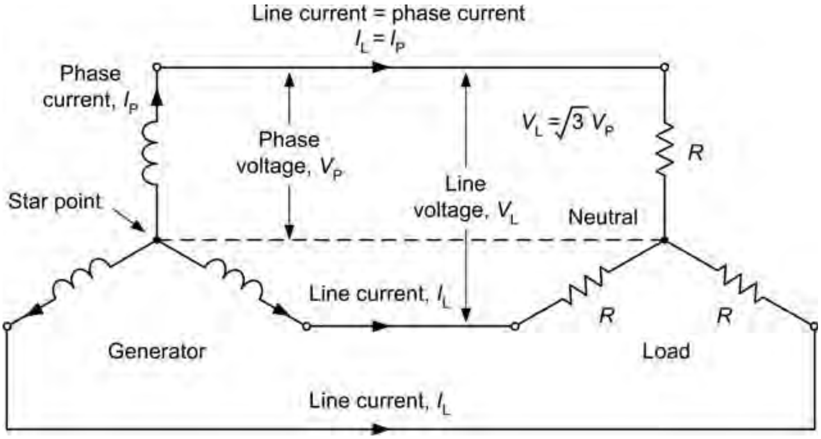

Line current = phase current
$I_L = I_P$
Phase current, I_P
Phase voltage, V_P
$V_L = \sqrt{3}\ V_P$
R
Line voltage, V_L
Star point
Neutral
Line current, I_L
R
R
Generator
Load
Line current, I_L

Figure 4.30 A complete star-connected three-phase distribution system

from which:

$$I_P = \frac{I_L}{\sqrt{3}} = \frac{6}{1.732} = 3.46\,\text{A}$$

4.3.1 Power in a three-phase system

In an unbalanced three-phase system the total power will be the sum of the individual phase powers. Hence:

$$P = P_1 + P_2 + P_3$$

or

$$P = V_1 I_1 \cos\phi_1 + V_2 I_2 \cos\phi_2 + V_3 I_3 \cos\phi_3$$

However, in the balanced condition the power is simply:

$$P = 3\,V_P I_P \cos\phi$$

where V_P and I_P are the phase voltage and phase current respectively and ϕ is the phase angle.

Using the relationships that we derived earlier, we can show that, for both the star and delta-connected systems the total power is given by:

$$P = \sqrt{3}\,V_L I_L \cos\phi$$

Example 4.4

In a three-phase system the line voltage is 110V and the line current is 12A. If the power factor is 0.8 determine the total power supplied.

Here it is important to remember (see Section 1.7.7) that:

$$\text{Power factor} = \cos\phi$$

and hence:

$$\begin{aligned} P &= \sqrt{3}\,V_L I_L \times \text{power factor} \\ &= \sqrt{3} \times 110 \times 12 \times 0.8 = 1829 = 1.829\text{ kW} \end{aligned}$$

Key point

The total power in a three-phase system is the sum of the power present in each of the three phases.

Test your understanding 4.3

1. The phase voltage in a star-connected AC system is 220V. What will the line voltage be?
2. The phase current in a delta-connected system is 12A.What will the line current be?
3. A three-phase system delivers power to a load consisting of three 12Ω resistors. If a current of 8A is supplied to each load, determine the total power supplied by the system.
4. In a three-phase system the line voltage is 105V and the line current is 8A. If the power factor is 0.75 determine the total power supplied.

4.4 AC motors

AC motors offer significant advantages over their DC counterparts. AC motors can, in most cases,

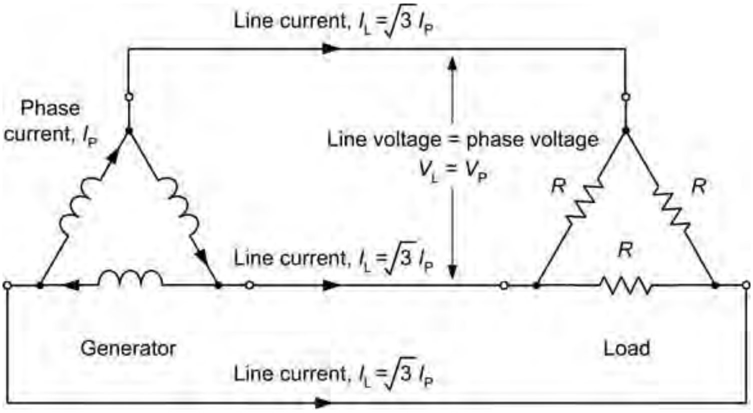
Line current, $I_L = \sqrt{3}\,I_P$
Phase current, I_P
Line voltage = phase voltage
$V_L = V_P$
R
R
R
Line current, $I_L = \sqrt{3}\,I_P$
Generator
Load
Line current, $I_L = \sqrt{3}\,I_P$

Figure 4.32 A complete delta-connected three-phase distribution system

duplicate the operation of DC motors and they are significantly more reliable. The main reason for this is that the commutator arrangements (i.e. brushes and slip-rings) fitted to DC motors are inherently troublesome. Because the speed of an AC motor is determined by the frequency of the AC supply that is applied it, AC motors are well suited to constant-speed applications.

AC motors are generally classified into two types:

- synchronous motors
- induction motors.

The synchronous motor is effectively an AC generator (i.e. an alternator) operated as a motor. In this machine, AC is applied to the stator and DC is applied to the rotor. The induction motor is different in that no source of AC or DC power is connected to the rotor. Of these two types of AC motor, the induction motor is by far the most commonly used.

Key point

The principle of all AC motors is based on the generation of a rotating magnetic field. It is this rotating field that causes the motor's rotor to turn.

4.4.1 Producing a rotating magnetic field

Before we go any further it's important to understand how a rotating magnetic field is produced. Take a look at Fig. 4.33 which shows a three-phase stator to which three-phase AC is applied. The windings are connected in delta configuration, as shown in Fig. 4.34. It is important to note that the two windings for each phase (diametrically opposite to one another) are wound in the *same* direction.

At any instant the magnetic field generated by one particular phase depends on the current through that phase. If the current is zero, the magnetic field is zero. If the current is a maximum, the magnetic field is a maximum. Since the currents in the three windings are 120° out of phase, the magnetic fields generated will also be 120° out of phase.

The three magnetic fields that exist at any instant will combine to produce one field that acts on the rotor. The magnetic fields inside the motor will combine to produce a moving magnetic field and, at the end of one complete cycle of the applied current, the magnetic field will have shifted through 360° (or one complete revolution).

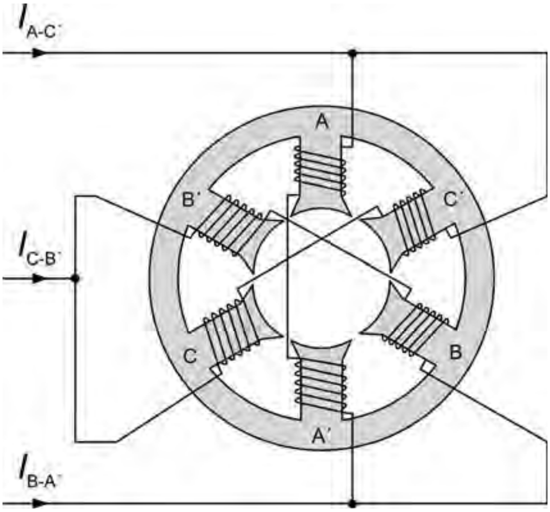

Figure 4.33 Arrangement of the field windings of a three-phase AC motor

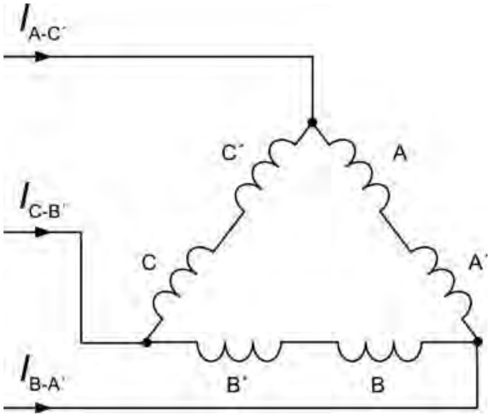

Figure 4.34 AC motor as a delta-connected load

Figure 4.35 shows the three current waveforms applied to the field system. These waveforms are 120° out of phase with each other. The waveforms can represent either the three alternating magnetic fields generated by the three phases, or the currents in the phases.

We can consider the direction of the magnetic field at regular intervals over a cycle of the applied current (i.e. every 60°). To make life simple we take the times at which one of the three current waveforms passes

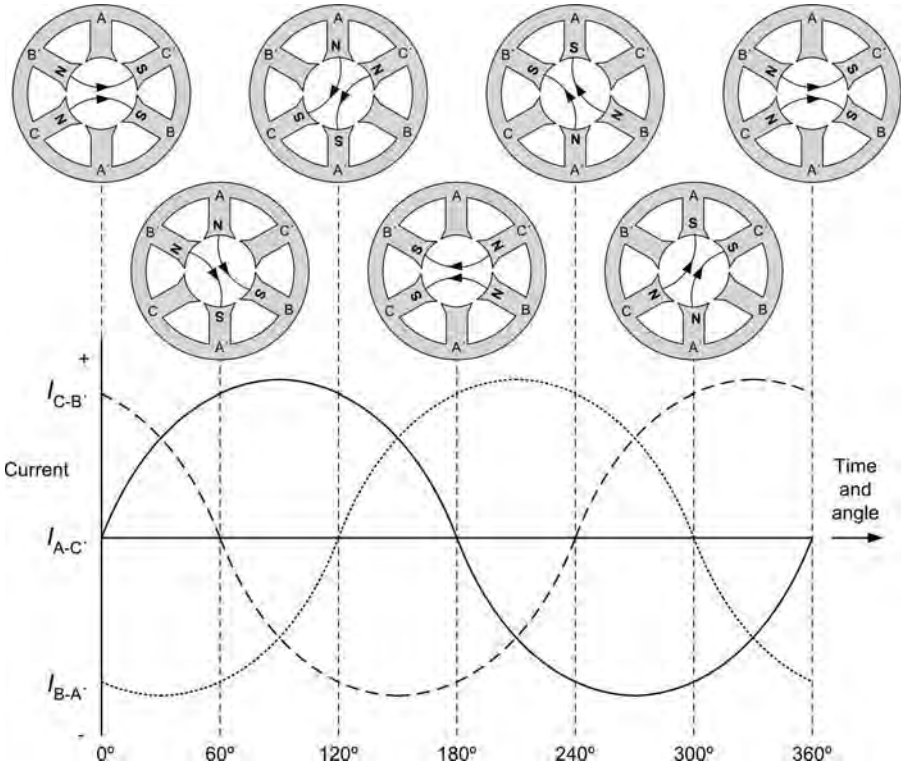

Figure 4.35 AC waveforms and magnetic field direction

through zero (i.e. the point at which there will be no current and therefore no field produced by one pair of field windings). For the purpose of this exercise we will use the current applied to A and C’ as our reference waveform (i.e. this will be the waveform that starts at 0° on our graph).

At 0°, waveform C–B′ is positive and waveform B–A′ is negative. This means that the current flows in opposite directions through phases B and C, and so establishes the magnetic polarity of phases B and C. The polarity is shown in Fig. 4.35. Note that B′ is a north pole and B is a south pole, and that C is a north pole and C′ is a south pole.

Since at 0° there is no current flowing through phase A, its magnetic field is zero. The magnetic fields leaving poles B′ and C will move towards the nearest south poles C′ and B. Since the magnetic fields of B and C are equal in amplitude, the resultant magnetic field will lie between the two fields, and will have the direction shown.

At the next point, 60° later, the current waveforms to phases A and B are equal and opposite, and waveform C is zero. The resultant magnetic field has rotated through 60°. At point 120°, waveform B is zero and the resultant magnetic field has rotated through another 60°. From successive points (corresponding to one cycle of AC), you will note that the resultant magnetic field rotates through one revolution for every cycle of applied current. Hence, by applying a three-phase alternating current to the three windings we have been able to produce a rotating magnetic field.

Key point

If three windings are placed round a stator frame, and three-phase AC is applied to the windings, the magnetic fields generated in each of the three windings will combine into a magnetic field that rotates. At any given instance, these fields combine together in order to produce a resultant field that which acts on the rotor. The rotor turns because the magnetic field rotates.

4.4.2 Synchronous motors

We have already shown how a rotating magnetic field is produced when a three-phase alternating current is applied to the field coils of a stator arrangement. If the rotor winding is energized with DC, it will act like a bar magnet and it will rotate in sympathy with the rotating field. The speed of rotation of the magnetic field depends on the frequency of the three-phase AC supply and, provided that the supply frequency remains constant, the rotor will turn at a constant speed. Furthermore, the speed of rotation will remain constant regardless of the load applied. For many applications this is a desirable characteristic; however, one of the disadvantages of a synchronous motor is that it cannot be started from a standstill by simply applying three-phase AC to the stator. The reason for this is that the instant AC is applied to the stator, a high-speed rotating field appears. This rotating field moves past the rotor poles so quickly that the rotor does not have a chance to get started. Instead, it is repelled first in one direction and then in the other.

Another way of putting this is simply that a synchronous motor (in its pure form) has no starting torque. Instead, it is usually started with the help of a small induction motor (or with windings equivalent to this incorporated in the synchronous motor). When the rotor has been brought near to synchronous speed by the starting device, the rotor is energized by connecting it to a DC voltage source. The rotor then falls into step with the rotating field. The requirement to have an external DC voltage source as well as the AC field excitation makes this type of motor somewhat unattractive!

Key point

The synchronous motor is so called because its rotor is synchronized with the rotating field set up by the stator. Its construction is essentially the same as that of a simple AC generator (alternator).

Key point

Synchronous motors are not self-starting and must be brought up to near synchronous speed before they can continue rotating by themselves. In effect, the rotor becomes ‘frozen’ by virtue of its inability to respond to the changing field!

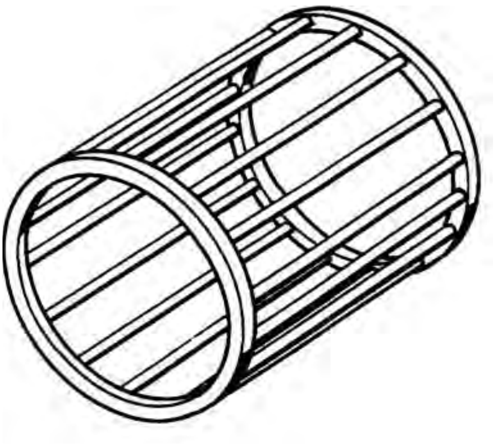

Figure 4.36 Squirrel cage rotor construction

4.4.3 Three-phase induction motors

The induction motor derives its name from the fact that AC currents are induced in the rotor circuit by the rotating magnetic field in the stator. The stator construction of the induction motor and of the synchronous motor are almost identical, but their rotors are completely different.

The induction motor rotor is a laminated cylinder with slots in its surface. The windings in these slots are one of two types. The most common uses so-called **squirrel cage** construction (see Fig. 4.36) that is made up of heavy copper bars connected together at either end by a metal ring made of copper or brass. No insulation is required between the core and the bars because of the very low voltages generated in the rotor bars. The air gap between the rotor and stator is kept very small so as to obtain maximum field strength.

The other type of winding contains coils placed in the rotor slots. The rotor is then called a **wound rotor**. Just as the rotor usually has more than one conductor, the stator usually has more than one pair of poles per coil, as shown in Fig. 4.37.

Key point

The induction motor is the most commonly used AC motor because of its simplicity, its robust construction and its relatively low cost. These advantages arise from the fact that the rotor of an induction motor is a self-contained component that is *not actually electrically connected to an external source of voltage.*

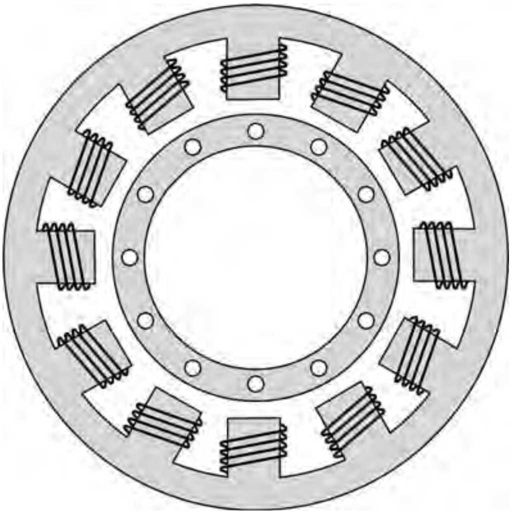

Figure 4.37 Typical stator construction

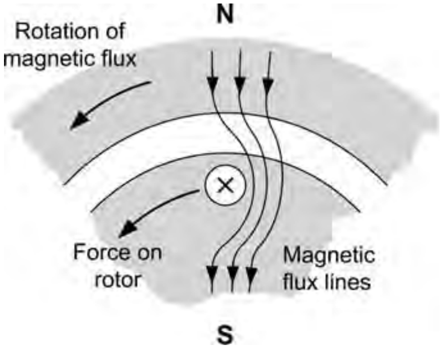

Figure 4.38 Force on the rotor of an induction motor

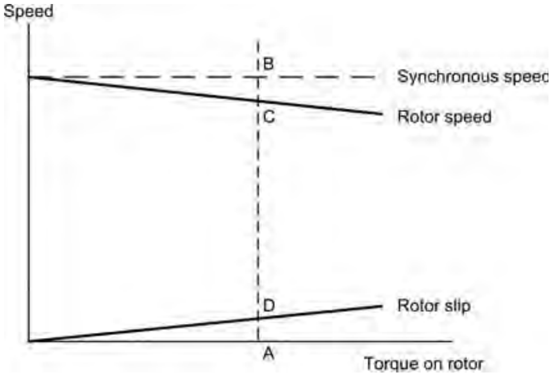

Figure 4.39 Relationship between torque and slip

Regardless of whether a squirrel cage or wound rotor is used, the basic principle of operation of an induction motor is the same. The rotating magnetic field generated in the stator induces an e.m.f. in the rotor. The current in the rotor circuit caused by this induced e.m.f. sets up a magnetic field. The two fields interact, and cause the rotor to turn. Fig. 4.38 shows how the rotor moves in the same direction as the rotating magnetic flux generated by the stator.

From **Lenz's law** (see Section 1.6.2) we know that an induced current opposes the changing field which induces it. In the case of an induction motor, the changing field is the rotating stator field and so the force exerted on the rotor (caused by the interaction between the rotor and the stator fields) attempts to cancel out the continuous motion of the stator field. Hence the rotor will move in the same direction as the stator field and will attempt to align with it. In practice, it gets as close to the moving stator field but never quite aligns perfectly with it!

Key point

The induction motor has the same stator as the synchronous motor. The rotor is different in that it does not require an external source of power. Current is induced in the rotor by the action of the rotating field cutting through the rotor conductors. This rotor current generates a magnetic field which interacts with the stator field, resulting in a torque being exerted on the rotor and causing it to rotate.

4.4.4 Slip, torque and speed

We have already said that the rotor of an induction motor is unable to turn in sympathy with the rotating field and, in practice, a small difference always exists. In fact, if the speeds were exactly the same, no relative motion would exist between the two, and so no e.m.f. would be induced in the rotor. For this reason the rotor operates at a lower speed than that of the rotating magnetic field. This phenomenon is known as **slip** and it becomes more significant as the rotor develops increased torque, as shown in Fig. 4.39.

From Fig. 4.39, for a torque of A the rotor speed will be represented by the distance AC whilst the slip will be represented by distance AD. Now:

$$AD = AB - AC = CB$$

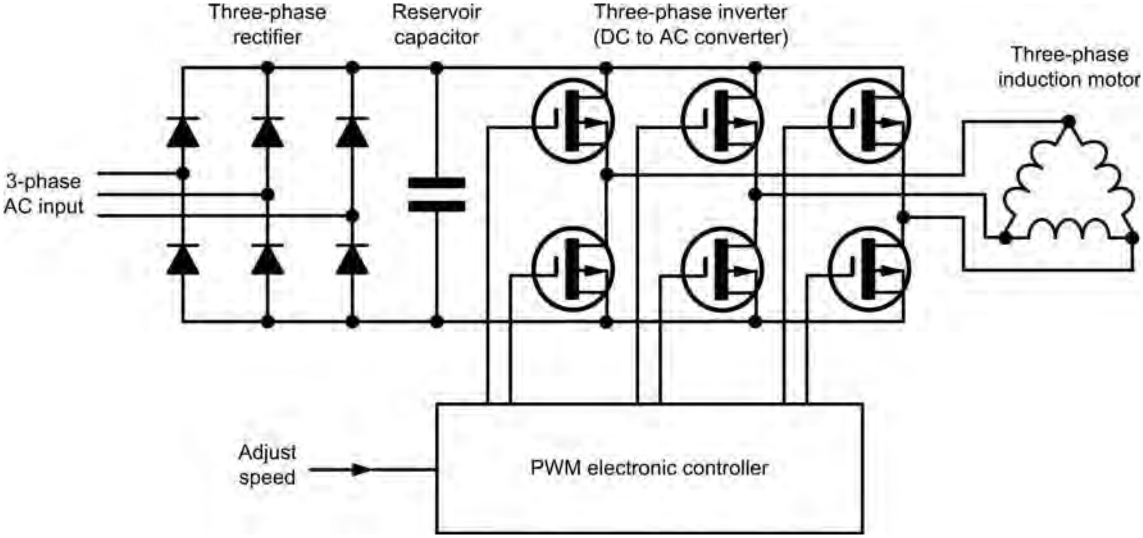
Three-phase rectifier
Reservoir capacitor
Three-phase inverter (DC to AC converter)
Three-phase induction motor
3-phase AC input
Adjust speed
PWM electronic controller

Figure 4.40 Using an inverter to produce a variable output speed from an AC induction motor

For values of torque within the working range of the motor (i.e. over the linear range of the graph shown in Fig. 4.39), the slip directly proportional to the torque and the **per-unit slip** is given by:

$$\text{Per-unit slip} = \frac{\text{slip}}{\text{synchronous speed}} = \frac{AD}{AB}$$

Now since AD = AB − BC,

$$\text{slip} = (\text{synchronous speed}) - (\text{rotor speed})$$

thus:

$$\text{Per-unit slip} = \frac{(\text{synchronous speed}) - (\text{rotor speed})}{\text{synchronous speed}}$$
$$= \frac{AB - BC}{AB}$$

The percentage slip is given by:

$$\text{Percentage slip} = \frac{(\text{synchronous speed}) - (\text{rotor speed})}{\text{synchronous speed}}$$
$$\times 100\% = \frac{AB - BC}{AB} \times 100\%$$

The actual value of slip tends to vary from about 6% for a small motor to around 2% for a large machine. Hence, for most purposes the induction motor can be considered to provide a constant speed (determined by the frequency of the current applied to its stator); however, one of its principal disadvantages is the fact that it is not easy to vary the speed of such a motor!

Note that, in general, it is not easy to control the speed of an AC motor unless a variable frequency AC supply is available. The speed of a motor with a wound rotor can be controlled by varying the current induced in the rotor but such an arrangement is not very practical as some means of making contact with the rotor windings is required. For this reason, DC motors are usually preferred in applications where the speed must be varied. However, where it is essential to be able to adjust the speed of an AC motor, the motor is invariably powered by an **inverter**. This consists of an electronic switching unit which produces a high-current three-phase **pulse-width-modulated** (**PWM**) output voltage from a DC supply, as shown in Fig. 4.40.

Key point

The rotor of an induction motor rotates at less than synchronous speed, in order that the rotating field can cut through the rotor conductors and induce a current flow in them. This percentage difference between the synchronous speed and the rotor speed is known as slip. Slip varies very little with normal load changes, and the induction motor is therefore considered to be a constant-speed motor.

Example 4.5

An induction motor has a synchronous speed of 3600 r.p.m. and its actual speed of rotation is measured as 3450 r.p.m. Determine (a) the per-unit slip and (b) the percentage slip.

(a) The per-unit slip is found from:

$$\text{Per-unit slip} = \frac{3{,}600 - 3{,}450}{3{,}600} = \frac{150}{3{,}600} = 0.042$$

(b) The percentage slip is given by:

$$\begin{aligned}\text{Percentage slip} &= \frac{3{,}600 - 3{,}450}{3{,}450} \times 100\% \\ &= \frac{150}{3{,}450} \times 100\% = 4.2\%\end{aligned}$$

Inside an induction motor, the speed of the rotating flux, N, is given by the relationship:

$$N = \frac{f}{p}$$

where N is the speed of the flux (in revolutions per second), f is the frequency of the applied AC (Hz) and p is the number of pole pairs.

Now the per-unit slip, s, is given by:

$$s = \frac{\text{AB} - \text{BC}}{\text{AB}} = \frac{N - N_r}{N}$$

where N is the speed of the flux (in revolutions per second) and N_r is the rotor speed.

Now:

$$sN = N - N_r$$

from which:

$$N_r = N - sN = N(1 - s)$$

and:

$$N_r = N(1 - s) = \frac{f}{p}(1 - s)$$

where N_r is the speed of the rotor (in revolutions per second), f is the frequency of the applied AC (Hz) and s is the per-unit slip.

Example 4.6

An induction motor has four poles and is operated from a 400 Hz AC supply. If the motor operates with a slip of 2.5% determine the speed of the output rotor.

Now:

$$\begin{aligned}N_r &= \frac{f}{p}(1 - s) = \frac{400}{2}(1 - 0.025) \\ &= 200 \times 0.975 = 195\end{aligned}$$

Thus the rotor has a speed of 195 revolutions per second (or 11,700 r.p.m.).

Example 4.7

An induction motor has four poles and is operated from a 60 Hz AC supply. If the rotor speed is 1700 r.p.m. determine the percentage slip.

Now:

$$N_r = \frac{f}{p}(1 - s)$$

from which:

$$\begin{aligned}s &= 1 - \frac{N_r p}{f} = 1 - \frac{\left(\frac{1{,}700}{60}\right) \times 2}{60} \\ &= 1 - \frac{56.7}{60} = 1 - 0.944 = 0.056\end{aligned}$$

Expressed as a percentage, this is 5.6%

4.4.5 Single- and two-phase induction motors

In the case of a two-phase induction motor, two windings are placed at right angles to each other. By exciting these windings with current which is 90° out of phase, a rotating magnetic field can be created. A single-phase induction motor, on the other hand, has only one phase. This type of motor is extensively used in applications which require small low-output motors. The advantage gained by using single-phase motors is that in small sizes they are less expensive to manufacture than other types. Also they eliminate the need for a three-phase supply. Single-phase motors are used in communication equipment, fans, portable power tools, etc. Since the field due to the single-phase AC voltage applied to the stator winding is pulsating, single-phase AC induction motors develop a pulsating torque. They are therefore less efficient than three-phase or two-phase motors, in which the torque is more uniform.

Single-phase induction motors have only one stator winding. This winding generates a field which can be said to alternate along the axis of the single winding, rather than to rotate. Series motors, on the other hand, resemble DC machines in that they have commutators and brushes.

When the rotor is stationary, the expanding and collapsing stator field induces currents in the rotor which generate a rotor field. The opposition of these fields exerts a force on the rotor, which tries to turn it 180° degrees from its position. However, this force is exerted through the centre of the rotor and the rotor will not turn unless a force is applied in order to assist it. Hence some means of **starting** is required for all single-phase induction motors.

Key point

Induction motors are available that are designed for three-phase, two-phase and single-phase operation. The three-phase stator is exactly the same as the three-phase stator of the synchronous motor. The two-phase stator generates a rotating field by having two windings positioned at right angles to each other. If the voltages applied to the two windings are 90° out of phase, a rotating field will be generated.

Key point

A synchronous motor uses a single- or three-phase stator to generate a rotating magnetic field, and an electromagnetic rotor that is supplied with DC. The rotor acts like a magnet and is attracted by the rotating stator field. This attraction will exert a torque on the rotor and cause it to rotate with the field.

Key point

A single-phase induction motor has only one stator winding; therefore the magnetic field generated does not rotate. A single-phase induction motor with only one winding cannot start rotating by itself.

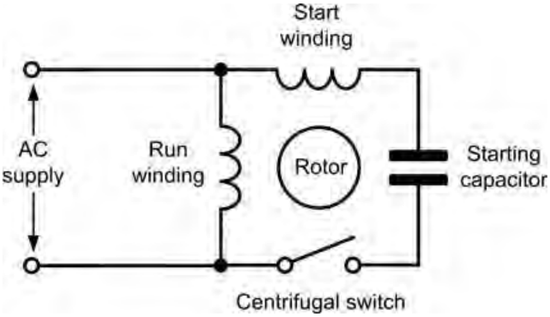

Figure 4.41 Capacitor starting arrangement

Once the rotor is started rotating, however, it will continue to rotate and come up to speed. A field is set up in the rotating rotor that is 90° out of phase with the stator field. These two fields together produce a rotating field that keeps the rotor in motion.

4.4.6 Capacitor starting

In an induction motor designed for capacitor starting, the stator consists of the main winding together with a starting winding which is connected in parallel with the main winding and spaced at right angles to it. A phase difference between the current in the two windings is obtained by connecting a capacitor in series with the auxiliary winding. A switch is included solely for the purposes of applying current to the auxiliary winding in order to start the rotor (see Fig. 4.41).

On starting, the switch is closed, placing the capacitor in series with the auxiliary winding. The capacitor is of such a value that the auxiliary winding is effectively a resistive–capacitive circuit in which the current leads the line voltage by approximately 45°. The main winding has enough inductance to cause the current to lag the line voltage by approximately 45°. The two field currents are therefore approximately 90° out of phase. Consequently the fields generated are also at an angle of 90°. The result is a revolving field that is sufficient to start the rotor turning.

After a brief period (when the motor is running at a speed which is close to its normal speed) the switch opens and breaks the current flowing in the auxiliary winding. At this point, the motor runs as an ordinary single-phase induction motor. However, since the two-phase induction motor is more efficient than a single-phase motor, it can be desirable to maintain the current in the auxiliary winding so that motor runs as a two-phase induction motor.

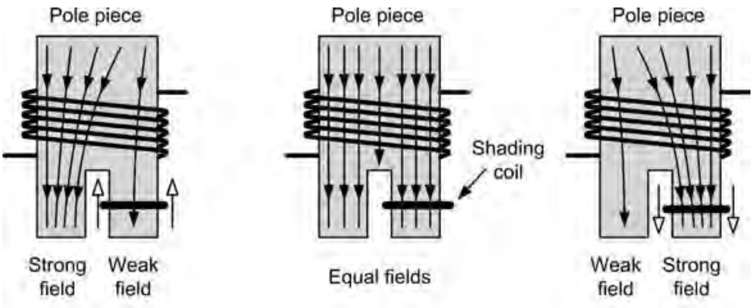

(a) Field increasing (b) Maximum field (c) Field decreasing

Figure 4.42 Action of a shaded pole

In some types of motor a more complicated arrangement is used with more than one capacitor switched into the auxiliary circuit. For example, a large value of capacitor could be used in order to ensure sufficient torque for starting a heavy load and then, once the motor has reached its operating speed, the capacitor value can be reduced in order to reduce the current in the auxiliary winding. A motor that employs such an arrangement, where two different capacitors are used (one for *starting* and one for *running*) is often referred to as **capacitor-start, capacitor-run** induction motor. Finally, note that, since phase shift can also be produced by an inductor, it is possible to use an inductor instead of a capacitor. Capacitors tend to be less expensive and more compact than comparable inductors and therefore are more frequently used.

Since the current and voltage in an inductor are also 90° out of phase, inductor starting is also possible. Once again, a starting winding is added to the stator. If this starting winding is placed in series with an inductor across the same supply as the running winding, the current in the starting winding will be out of phase with the current in the running winding. A rotating magnetic field will therefore be generated, and the rotor will rotate.

Key point

In order to make a single-phase motor self-starting, a starting winding is added to the stator. If this starting winding is placed in series with a capacitor across the same supply as the running winding, the current in the starting winding will be out of phase with the current in the running winding. A rotating magnetic field will therefore be generated, and the rotor will rotate. Once the rotor comes up to speed, the current in the auxiliary winding can he switched-out, and the motor will continue running as a single-phase motor.

4.4.7 Shaded pole motors

A different method of starting a single-phase induction motor is based on a **shaded-pole**. In this type of motor, a moving magnetic field is produced by constructing the stator in a particular way. The motor has projecting pole pieces just like DC machines; and part of the pole surface is surrounded by a copper strap or **shading coil**.

As the magnetic field in the core builds, the field flows effortlessly through the unshaded segment. This field is coupled into the shading coil which effectively constitutes a short-circuited loop. A large current momentarily flows in this loop and an opposing field is generated as a consequence. The result is simply that the unshaded segment initially experiences a larger magnetic field than does the shaded segment. At some later time, the fields in the two segments become equal. Later still, as the magnetic field in the unshaded segment declines, the field in the shaded segment strengthens. This is illustrated in Fig. 4.42.

Key point

In the shaded pole induction motor, a section of each pole face in the stator is shorted out by a metal strap. This has the effect of moving the magnetic field back and forth across the pole face. The moving magnetic field has the same effect as a rotating field, and the motor is self-starting when switched on.

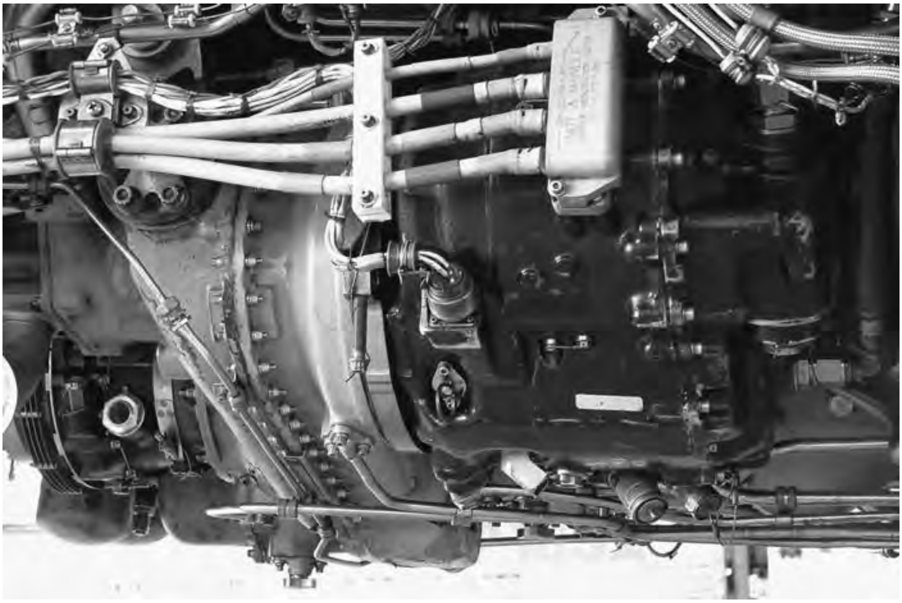

Figure 4.43 Main three-phase AC generator mounted in the engine compartment of a large passenger aircraft

Test your understanding 4.4

1. Explain the difference between synchronous AC motors and induction motors.
2. Explain the main disadvantage of the synchronous motor.
3. An induction motor has a synchronous speed of 7200 r.p.m. and its actual speed of rotation is measured as 7000 r.p.m. Determine (a) the per-unit slip and (b) the percentage slip.
4. An induction motor has four poles and is operated from a 400 Hz AC supply. If the motor operates with a slip of 1.8% determine the speed of the output rotor.
5. An induction motor has four poles and is operated from a 60 Hz AC supply. If the rotor speed is 1675 r.p.m. determine the percentage slip.
6. Explain why a single-phase induction motor requires a means of starting.

4.5 Practical aircraft generating systems

Generators are a primary source of power in an aircraft and can either produce direct or alternating current (DC or AC) as required. They are driven by a belt drive (in smaller aircraft), or engine/APU accessory gearbox in larger aircraft. Generators will have sufficient output to supply all specified loads and charge the battery(s). Most avionic equipment requires a regulated and stable power supply depending on its function, e.g. in the case of lighting, it would be inconvenient if the intensity of lighting varied with engine speed. Generator output is affected by internal heat and this has to be dissipated. Cooling methods can include natural radiation from the casing, however this is inadequate for high-output devices where ram-air is directed from a scoop and directed into the generator's brush-gear and commutator. In some installations, e.g. helicopters, a fan is installed to provide cooling when the aircraft is hovering. We know from basic theory that a generator's output will vary depending on the input shaft speed. A means of regulating the generator's output to the bus is required as is a means of overload protection.

4.5.1 DC generators

DC generators are less common on modern aircraft due to their low power-to-weight ratio, poor performance at low r.p.m. and high servicing costs. The latter is due to the need for inspection and servicing of brushes and commutators since they have irregular

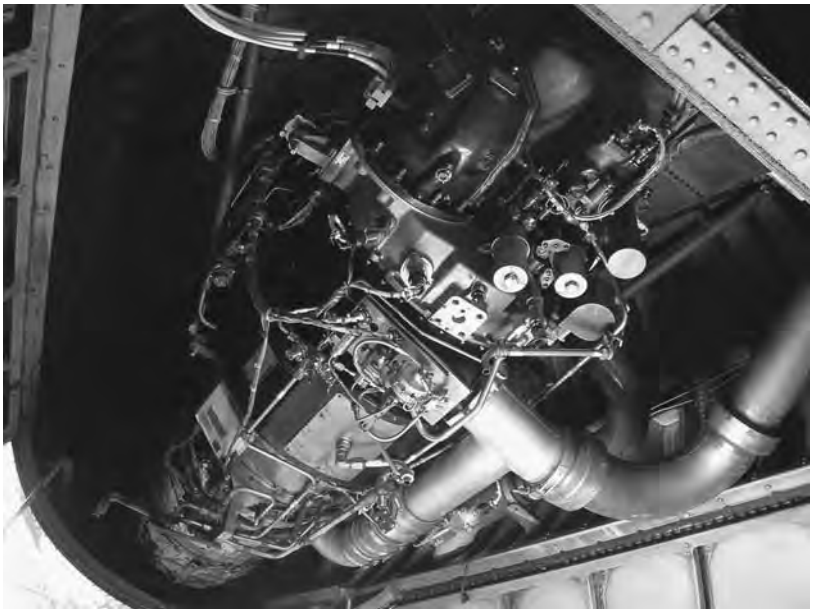

Figure 4.44 Tail-mounted auxiliary power unit (APU) with three-phase AC generator (top of picture)

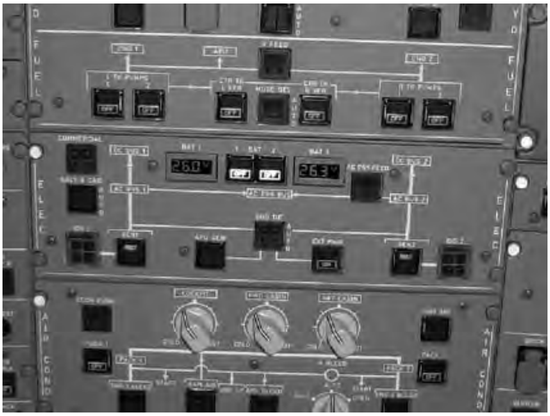

Figure 4.45 Overhead electrical panel fitted with main and APU generator indicators and displays. This panel provides the crew with an instant overview of the status of the electrical generating system and the AC and DC distribution buses)

surfaces/contact area and conduct the entire load current. Carbon brushes are porous and will absorb substances including moisture; this provides an amount of inherent lubrication. At altitude, the atmosphere is dryer and this leads to higher brush wear. Without any lubrication, arcing occurs and static charges build up; brush erosion is accelerated. Additives can be incorporated into the brushes that deposit a lubricating film on the commutator; this needs time to build up a sufficient protection; brushes need to be run in for several hours before the protective layer forms (this is often mistaken for contamination). The alternative is an in-built lubrication that is consumed as part of the natural brush wear, i.e. no film is deposited.

Key maintenance point

Automotive style alternators are normally installed on general aviation and light aircraft to overcome the shortfalls of DC generators. Larger aircraft use brushless AC generators.

4.5.2 Alternators

Automotive style alternators comprise a rotor, stator and rectifier pack. The rotor contains the field coil arranged in six sections around the shaft. Each section forms a pole piece that is supplied via slip-rings and brushes. The alternator has no residual magnetism, its field has to be excited by a DC supply (e.g. the battery). When energised, the rotor's pole pieces

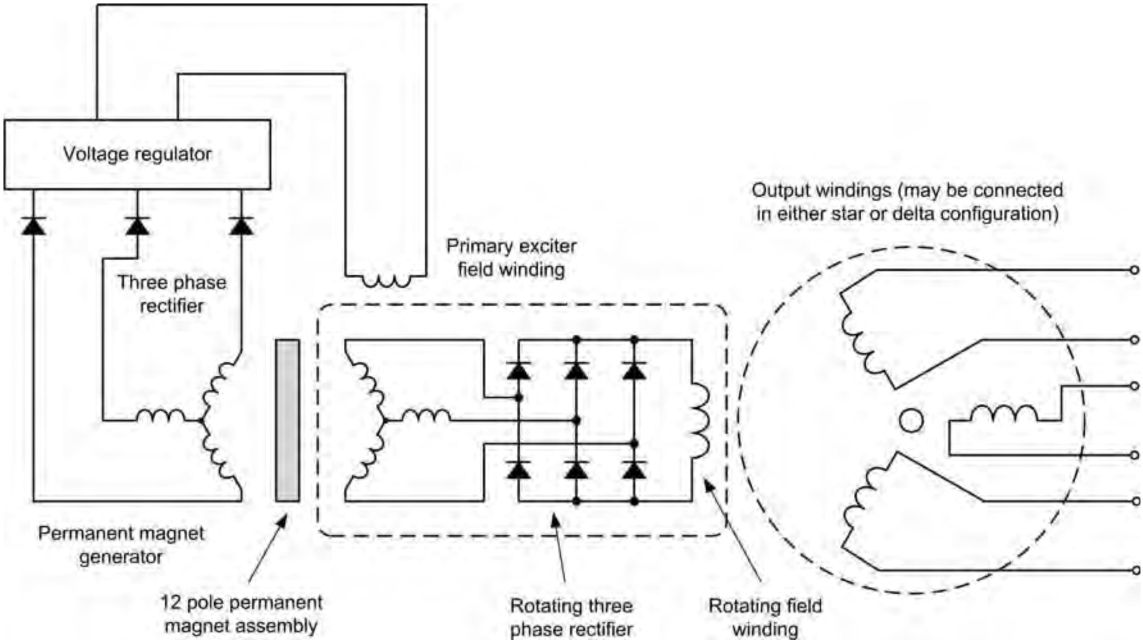
Voltage regulator
Three phase rectifier
Primary exciter field winding
Output windings (may be connected in either star or delta configuration)
Permanent magnet generator
12 pole permanent magnet assembly
Rotating three phase rectifier
Rotating field winding

Figure 4.46 A practical brushless AC generator arrangement

produce north and south poles. As these poles are rotated they induce currents in the stator windings; these are wound at 120° and this produces three-phase AC. The AC output is fed to a diode rectifier pack comprising six high-current diodes, see Fig. 4.46 which produces a DC output. This has to be regulated before connecting to the various aircraft systems. Voltage regulators used with alternators on general aviation aircraft can be electromechanical or electronic. There are two types of electromechanical regulators: sensing coil with contacts and carbon-pile. Modern solid-state electronic regulators are more reliable as they use no mechanical parts.

The alternators previously described rely on slip-rings and brushes, albeit with reduced current loading. Slip-rings and brushes require maintenance in the workshop thereby incurring an associated cost burden. The **brushless generator** is a more complex device but has significantly increased reliability coupled with reduced maintenance requirements. A schematic diagram for the brushless generator is shown in Fig. 4.46; the device can be divided into three main sections:

- permanent magnet generator
- rotating field
- three-phase output.

The AC generator uses a brushless arrangement based on a rotating rectifier and **permanent magnet generator** (**PMG**). The output of the PMG rectifier is fed to the voltage regulator which provides current for the primary exciter field winding. The primary exciter field induces current into a three-phase rotor winding. The output of this winding is fed to the shaft-mounted rectifier diodes which produce a pulsating DC output which is fed to the rotating field winding. It is important to note that the excitation system is an integral part of the rotor and that there is no direct electrical connection between the stator and rotor.

The output of the main three-phase generator is supplied via **current transformers** (one for each phase) that monitor the load current in each line. An additional current transformer can also be present in the neutral line to detect an **out-of-balance** condition (when the load is unbalanced an appreciable current will flow in the generator's neutral connection).

The generator output is fed to the various aircraft systems and a solid-state regulator. This rectifies the output and sends a regulated direct current to the stator exciter field of the PMG. The regulator maintains the output of the generator at 115V AC and is normally contained within a **generator control unit** (**GCU**); this is unit is further described in Chapter 7.

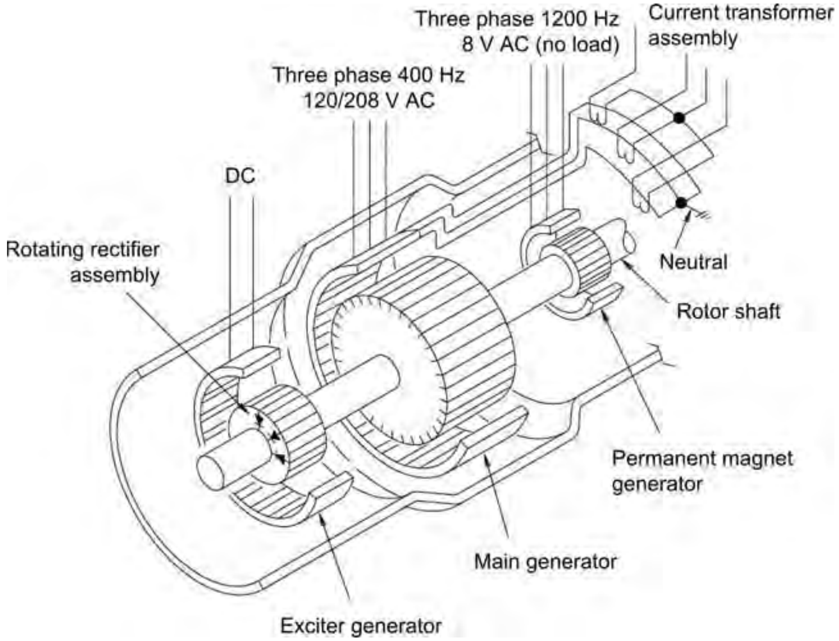

Figure 4.47 Layout of a typical brushless AC generator

Although the regulator controls the output voltage of the generator, its frequency will vary depending on the speed of shaft rotation. Variable frequency power supplies (sometimes called **frequency wild**) are acceptable for resistive loads, e.g. de-icing, but they are not suitable for many induction motor loads that need to run at constant speed, e.g. fuel pumps and gyroscopic instruments. Furthermore, certain loads are designed for optimum efficiency at the specified frequency of 400 Hz, e.g. cooling fans. Some larger multi-engine aircraft operate the generators in parallel; it is essential that each generator is operating at the same frequency. Constant frequency can be achieved in one of two ways: controlling the shaft speed by electromechanical methods using a **constant speed drive** (**CSD**) or by controlling the generator output frequency electronically (**variable speed constant frequency**: **VSCF**).

Key point

A three-phase AC generator can be made brushless by incorporating an integral excitation system in which the field current is derived from a rotor-mounted rectifier arrangement. In this type of generator the coupling is entirely magnetic and no brushes and slip-rings are required.

4.5.3 Constant speed drive/integrated drive generator

The CSD is an electromechanical device installed on each engine. The input shaft is connected to the engine gearbox; the output shaft is connected to the generator. The CSD is based on a variable ratio drive employing a series of hydraulic pumps and differential gears. CSDs can be disconnected from the engine via a clutch, either manually or automatically. Note that it is only possible to reconnect the clutch on the ground. Modern commercial aircraft employ a combined CSD and brushless AC generator, in one item – the integrated drive generator (IDG). Typical characteristics are a variable input speed of 4500/9000 r.p.m. and a constant output speed of 12,000 ± 150 r.p.m. The IDG on a large commercial aircraft is oil-cooled and produces a 115/200 V 400 Hz three-phase, 90 m kVA output.

4.5.4 Variable speed constant frequency (VSCF)

Both the constant speed drive and integrated drive generator are complex and very expensive electromechanical devices. Advances in semiconductor technology has facilitated the development of solid-state products that can convert variable frequencies into

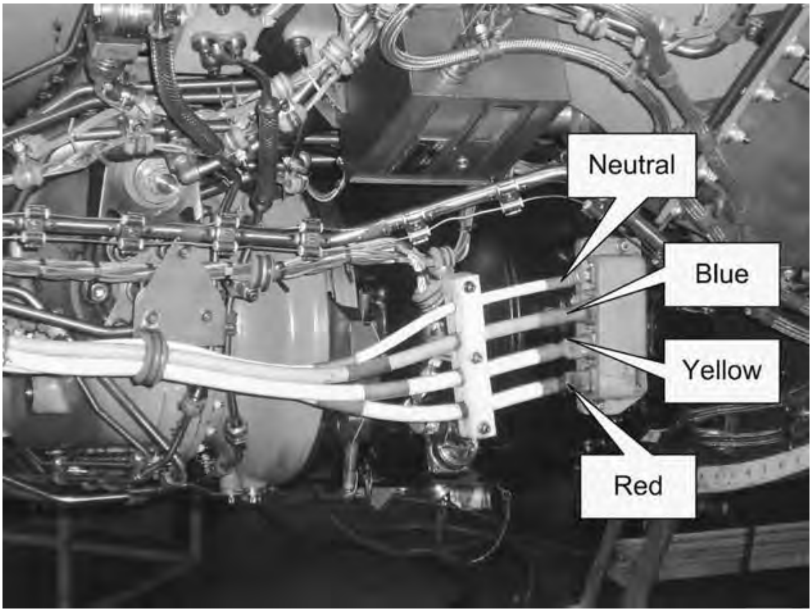
Neutral
Blue
Yellow
Red

Figure 4.48 Typical main generator showing three-phase output cabling

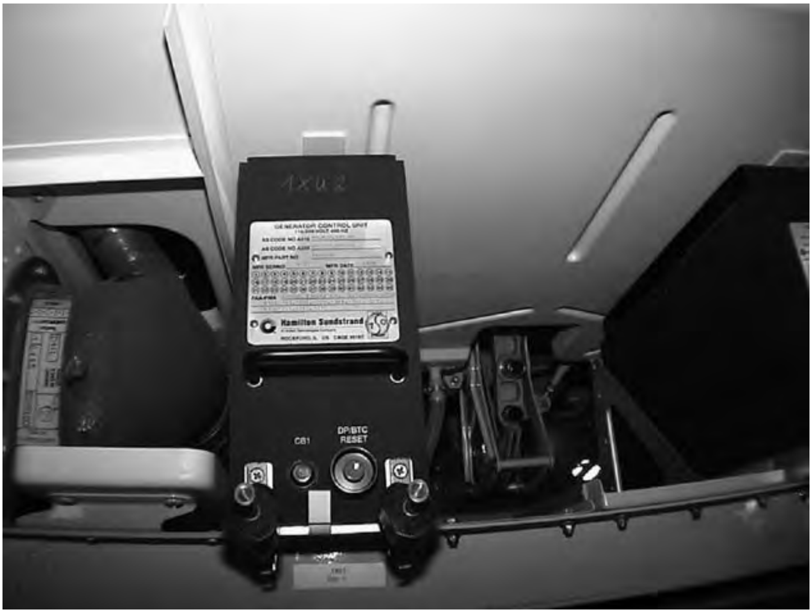
GENERATOR CONTROL UNIT
Hamilton Sundstrand
CB1
DP/BTC RESET

Figure 4.49 Typical generator control unit (GCU)

115/200 V AC, 400 Hz three-phase power supplies. Variable speed constant frequency (VSCF) systems comprise a generator and power converter.

A brushless AC generator is mounted onto the engine accessory gearbox as before; its output voltage and frequency varies in accordance with engine speed. The gearbox increases the generator speed by a ratio of 1:3, producing a variable output frequency between 1300 and 2500 r.p.m. The three-phase output of VSCF is full-wave rectified to produce a 270 V DC output. This direct current output is smoothed by large capacitors, filtered and fed into an inverter that produces a square wave output. The inverter converts the DC level into a three phase, pulse-width-modulated waveform. This is then converted in the sinusoidal output voltage. These outputs are then converted into AC through electronic circuits. The final output stage is monitored by a **current transformer** and **electromagnetic interference** filter (**CT**/**EMI**). Within the **generator control unit** (**GCU**) a **generator control relay** (**GCR**) energizes the field. This circuit can be interrupted by the pilot or automatically under fault conditions.

The VSCF **generator conversion control unit** (**GCCU**) can either be integrated with the generator as a single engine mounted device (weighing typically 65 kg) or it can be located in the airframe. The latter arrangement has the advantage of making the engine accessories smaller; this means a lower profile nacelle. In addition, the electronics can be located in a zone with reduced temperature and vibration.

VSCF systems are more reliable compared with constant speed drive and integrated drive generators since there are fewer moving parts. The VSCF system's moving parts consist of the generator's rotor and an oil pump used for cooling. The VSCF can be used for both primary and secondary power supplies; outputs of 110 kVA are achievable. Enabling technology for VSCF are the power transistors and diodes capable of handling currents in excess of 500 A. These diodes and transistors form the core of the rectifier and conversion circuits of the GCCU. The VSCF contains an oil pump mounted on the generator shaft that circulates oil through the system; this oil is passed through a heat exchanger. Oil temperatures and pressures are closely monitored; warnings are given to the crew in the event of malfunctions. Oil level can be checked during ground servicing through a sight glass.

4.6 Multiple choice questions

1. The slip-rings in an AC generator provide a means of:
   (a) connecting an external circuit to a rotating armature winding
   (b) supporting a rotating armature without the need for bearings
   (c) periodically reversing the current produced by an armature winding.

2. Decreasing the current in the field coil of a DC generator will:
   (a) decrease the output voltage
   (b) increase the output voltage
   (c) increase the output frequency.

3. The rotor of an AC induction motor consists of:
   (a) a laminated iron core inside a 'squirrel cage' made from copper or aluminium
   (b) a series of coil windings on a laminated iron core with connections via slip-rings
   (c) a single copper loop which rotates inside the field created by a permanent magnet.

4. The slip speed of an AC induction motor is the difference between:
   (a) the synchronous speed and the rotor speed
   (b) the frequency of the supply and the rotor speed
   (c) the maximum speed and the minimum speed.

5. When compared with three-phase induction motors, single-phase induction motors:
   (a) are not inherently 'self starting'
   (b) have more complicated stator windings
   (c) are significantly more efficient.

6. A three-phase induction motor has three pairs of poles and is operated from a 60 Hz supply. Which one of the following gives the motor's synchronous speed?
   (a) 1200 r.p.m.
   (b) 1800 r.p.m.
   (c) 3600 r.p.m.

7. In a star-connected three-phase system, the line voltage is found to be 200 V. Which one of the following gives the approximate value of phase voltage?
   (a) 67 V
   (b) 115 V
   (c) 346 V.

8. A single-phase AC generator has twelve poles and it runs at 600 r.p.m. Which one of the following gives the output frequency of the generator?
   (a) 50 Hz
   (b) 60 Hz
   (c) 120 Hz.

9. In a balanced star-connected three-phase system the line current is 2 A and the line voltage is

200 V. If the power factor is 0.75 which one of the following gives the total power in the load?
   (a) 300 W
   (b) 520 W
   (c) 900 W.
10. The commutator in a DC generator is used to:
   (a) provide a means of connecting an external field current supply
   (b) periodically reverse the connections to the rotating coil winding
   (c) disconnect the coil winding when the induced current reaches a maximum value.
11. The brushes fitted to a DC motor/generator should have:
   (a) low coefficient of friction and low contact resistance
   (b) high coefficient of friction and low contact resistance
   (c) low coefficient of friction and high content resistance.
12. Self-excited generators derive their field current from:
   (a) the current produced by the armature
   (b) a separate field current supply
   (c) an external power source.
13. In a shunt-wound generator:
   (a) none of the armature current flows through the field
   (b) some of the armature current flows through the field
   (c) all of the armature current flows through the field.
14. When combined with a CSD, a brushless three-phase AC generator is often referred to as:
   (a) a compound generator
   (b) a 'frequency wild' generator
   (c) an IDG.
15. An out-of-balance condition in an AC three-phase system can be detected by means of:
   (a) voltage sensors connected across each output line
   (b) a dedicated field coil monitoring circuit
   (c) a current transformer connected in the neutral line.

Chapter 5

Batteries

Aircraft electrical power can be derived from a variety of sources; these are categorized as either primary or secondary. Batteries are primary sources of electrical power found on most aircraft delivering direct current. (Secondary sources of power are described in Chapter 6.) There are several types of battery used on aircraft, defined by the types of materials used in their construction; these include lead-acid and nickel-cadmium batteries. The choice of battery type depends mainly on performance and cost. Other types of battery are being considered for primary power on aircraft; these include lithium and nickel-metal hydride. Electrical power delivered by batteries is used for a variety of applications, e.g. lights, radios, instruments, and motors. This chapter reviews the battery types used on aircraft, typical applications and how they are installed and maintained.

5.1 Overview

The main aircraft battery is a primary source of electrical power; its use can be controlled by the pilot or by automatic means. The main battery provides autonomous starting for the engine(s) or auxiliary power unit (APU) when external ground power is not available. Typical current requirement during APU starting is 1000A, albeit for a short period of time. Batteries also supply **essential loads** in the event of generator failure. It is an airworthiness requirement that the main battery(s) supplies essential services for a specified period of time. Other aircraft systems are supplied with their own dedicated batteries, e.g. aircraft emergency lights. Individual computers use their own battery sources to provide non-volatile memory. Battery type and maintenance requirements have to be understood by the aircraft engineer to ensure safe and reliable operation and availability.

The battery is constructed from a number of individual cells; generic cell features consist of two electrodes (the **anode** and **cathode**) and electrolyte contained within a casing. Cell materials vary depending on the type of battery performance required for a given cost. The simple primary cell (Fig. 5.1) causes an electron flow from the cathode (negative) through the external load to the anode (positive). The materials used refer to the two types of battery cell in widespread use on aircraft for the primary source of power: **lead-acid** or **nickel-cadmium**. These are maintained on the aircraft and treated as line-replaceable units; a full description of these two battery types is provided in this chapter. Cells used within other aircraft equipment or systems are typically made from **lithium** or **nickel-metal hydride** materials. These are not maintained as individual items on the aircraft, they are installed/removed as part of the equipment that they are fitted into; in this case only a brief description is provided.

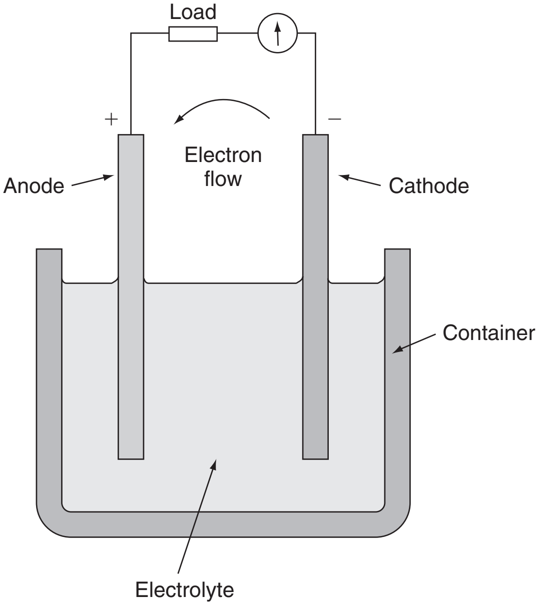

Figure 5.1 Electrical storage cell

Key point

The main aircraft battery is a primary source of electrical power; its use can be controlled by the pilot or by automatic means.

5.2 Storage cells

The basic function of any electrical cell is the conversion of chemical energy into electrical energy. The cells can be considered as a chemical means of storing electrical energy. Electrons are removed from the (**positive**) cathode and deposited on the (**negative**) anode. The electrolyte is the physical means of migration between the cathode/anode. The attraction of electrons between cathode/anode creates a potential difference across the cell; the cathode/anode are attached to external terminals for connection to the equipment or system. Material types used for the cathode/anode and electrolyte will determine the cell voltage.

Key maintenance point

Different battery types possess different characteristics both in terms of what they are used for, and in terms of how they should be maintained; always refer to maintenance manual instructions for servicing.

Cells are categorized as either primary (where they can only be used once) or secondary (where they can be recharged). In the primary cell, the chemical activity occurs only once, i.e. during discharge. By applying current through a secondary cell in the opposite direction to that of discharging, the chemical reaction is reversed and the cell can be used again. The cathode/anode are returned to their original charged form; the cell therefore becomes a chemical means of storing electrical energy.

Key maintenance point

It can be dangerous to attempt charging a primary cell. In the secondary cell, the chemical activity is reversible.

The energy storage **capacity** of a cell is determined by the amount of material available for chemical reaction. To maximize the storage capacity, the physical areas of the cathode and anode are made as large as possible, normally by constructing them as plates. Capacity is stated in ampere-hours; batteries are rated with low or high discharge rates, either 10 hours or 1 hour. The battery's capacity will gradually deteriorate over time depending on usage, in particular the charge and discharge rates. For aircraft maintenance purposes, we need to define the acceptable capacity of the main battery(s); this is the ratio of actual capacity and rated capacity, expressed as a percentage. Actual capacity must not fall below 80% of the rated capacity; therefore testing is required on a periodic basis. **Memory effect** is observed in some secondary cells that cause them to hold less charge; cells gradually lose their maximum capacity if they are repeatedly recharged before being fully discharged. The net result is the cell appears to retain less charge than specified.

All secondary cells have a finite life and will gradually lose their capacity over time due to secondary chemical reactions; this occurs whether the cell is used or not. They also have a finite number of charge and discharge cycles since they lose a very small amount of storage capacity during each cycle. Secondary cells can be damaged by repeated deep discharge or repeated over-charging.

Storage cells have **internal resistance**; this is usually very small but it has the effect of limiting the amount of current that the cell can supply and also reducing the amount of electromotive force (e.m.f.) available when connected to a load. Internal resistance varies significantly with the distance between plates. For this reason, the gap is made as small as practicably possible. This internal resistance is sometimes shown as a series resistor within the cell for design purposes, but it is normally omitted in circuit diagrams used in maintenance and wiring diagram manuals. Internal resistance is affected by temperature and this leads to practical issues for certain cell types.

Test your understanding 5.1

Describe the process whereby secondary cells gradually lose their maximum capacity (memory effect).

A number of cells are linked together in series to form a **battery**. The total battery terminal voltage is

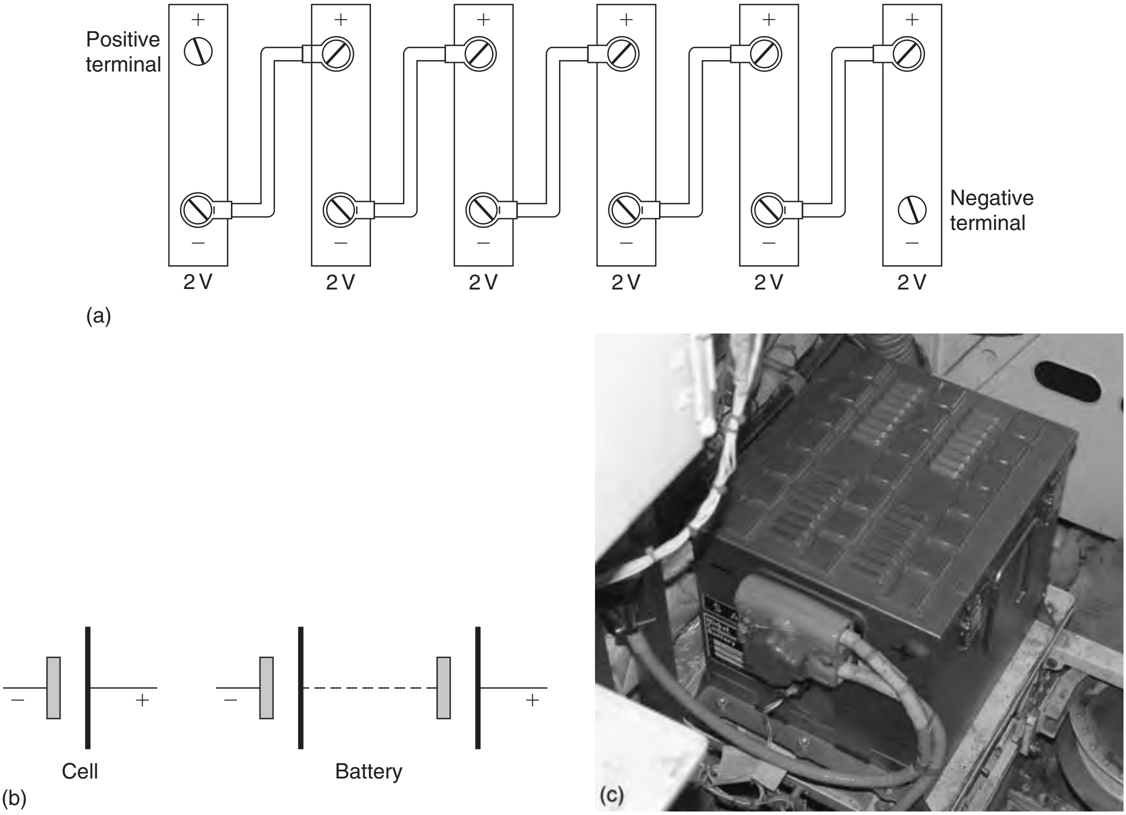

Figure 5.2 Cells and batteries: (a) connection of cells to form a battery; (b) symbols for cells and a battery; (c) typical battery casing

the sum of individual cell voltages, see Fig. 5.2(a). In this illustration, six cells are connected in series to for a 12 V battery. The circuit symbols for individual cells and a battery are shown in Fig. 5.2(b). All of the individual cells are contained within a battery case, see Fig. 5.2(c).

5.3 Lead-acid batteries

Developed in 1859, this is the oldest secondary cell technology in aircraft use today. Despite advances in alternative technologies, lead-acid batteries have retained market share (particularly in general aviation) due to the relatively low cost and mature technology. This type of battery has widespread applications on general aviation fixed and rotary wing aircraft due to the high current available for engine start and relatively low manufacturing cost (compared with nickel-cadmium batteries). The surface area of the plates, strength of the electrolyte and temperature determine the actual capacity of a lead-acid cell. There are two types of lead-acid battery used in aircraft: **flooded** (wet-cell) and **sealed**. The disadvantages of flooded batteries are that they require regular maintenance, they liberate gas during charging and the electrolyte can be spilt or leak. Spillage and/or leakage of the electrolyte requires immediate clean-up to avoid corrosion. These problems are overcome with sealed lead-acid batteries. Although lead-acid batteries remain popular with GA aircraft, this battery technology will eventually be phased out due to **environmental** issues.

5.3.1 Construction

Flooded cells are housed within an impact- and acid-resistant casing made from polystyrene-based materials. The casing retains the two terminals and includes a vent cap to prevent gas pressure build-up whilst not allowing the electrolyte to escape. A single battery cell contains a number of positive and negative **plate groups** constructed as illustrated in Fig. 5.3. The

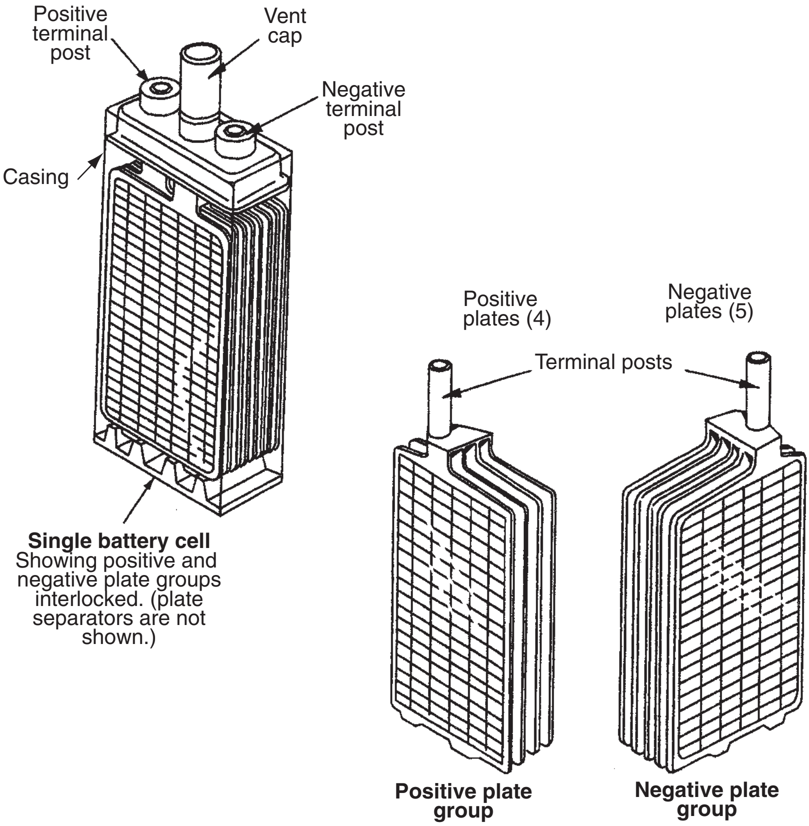
Positive terminal post
Vent cap
Negative terminal post
Casing
Single battery cell
Showing positive and negative plate groups interlocked. (plate separators are not shown.)
Positive plates (4)
Negative plates (5)
Terminal posts
Positive plate group
Negative plate group

Figure 5.3 Lead-acid cell construction

individual plates are separated by a porous material to prevent short circuit through physical contact; there is space below the plates to allow any material shed from the plates to accumulate without shorting the plates. Flooded cells can be accessed on an individual basis for checking the content and condition of the electrolyte.

Key point

Since each positive plate is always positioned between two negative plates, there is actually one more negative plate than the positives.

Each positive plate is a cast lead/antimony frame formed as a grid; this is impregnated with a paste of lead dioxide (PbO_2). The negative plate is a similar frame containing lead (Pb); this is sometimes referred to as '**spongy lead**'. In practice, a typical cell is constructed with several plates in order to get the required current output. Positive plates distort when chemical reactions take place on only one side; for this reason, there are always an even number of positive plates sandwiched between an odd number of negative plates. All positive plates are connected together as are all the negatives. The plates are interlaced and separated by a porous separator that allows free circulation of the electrolyte at the plate surfaces; the plates are all stacked within the cell container. The **electrolyte** is sulphuric acid diluted with distilled (pure) water (H_2SO_4).

Test your understanding 5.2

What are the advantages and disadvantages of flooded (wet-cell) batteries?

5.3.2 Charging/discharging

When fully charged, each cell has a potential difference of 2.5 V (falling to 2.2 V after a period of approximately one hour) at its terminals; when discharged,

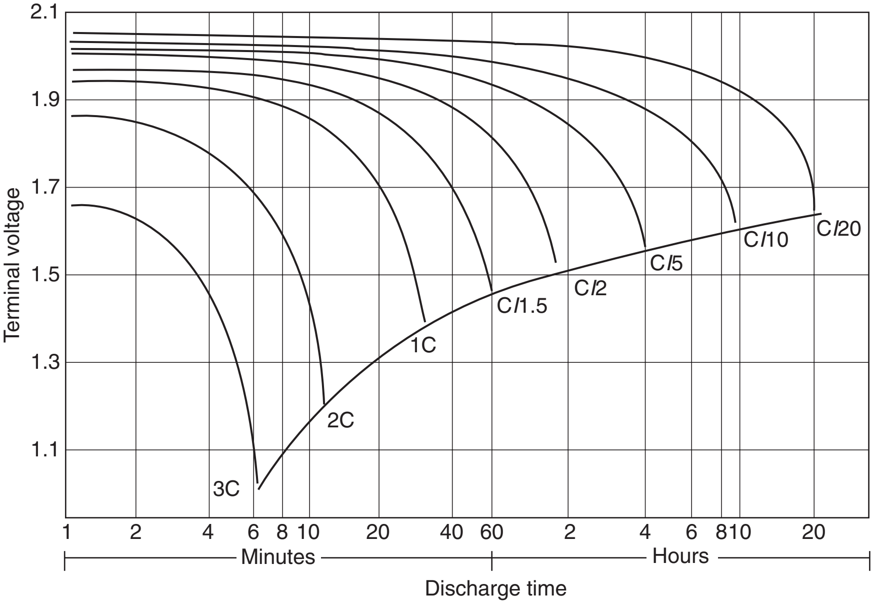

Figure 5.4 Lead-acid cell discharge characteristics

this potential difference is 1.8V. A six-cell battery would produce 13.2V fully charged, and 10.8V DC when discharged. A twelve-cell battery would produce 26.4V DC fully charged, and 21.6V DC when discharged. During normal use of lead-acid cells, the terminal voltage stays at around 2V for a long period of cell life, this is referred to as the cell's **nominal voltage**. When fully charged, the positive plate is lead dioxide (PbO_2) and the negative plate is lead (Pb).

Connecting an external load to the battery completes the electrical circuit, electrons are transferred from the negative plate and the battery starts to discharge. The chemical reaction that takes place during discharge changes each of the plates into lead sulphate ($PbSO_4$). Molecules of water are formed, thereby diluting the electrolyte. For a given battery capacity, a steady discharge rating forms part of the battery specification, e.g. a 20 hour rate produces a constant current for 20 hours until the cell is discharged. Figure 5.4 illustrates typical lead-acid battery characteristics at different discharge currents. The discharge current, in amps (A), is expressed as a fraction of the numerical value of C. For example, 0.1 C means C/10A, and discharging will take approximately 10 hours. If the battery capacity was 35 Ah, a discharge current of 3.5A can be expressed as 0.1 C (or C/10). This means that batteries of different sizes can be compared by a single set of graphs. Since a battery may be rated for different discharge times, its rated capacity will normally be an indication of current used. With a 20-hour discharge capacity, the chart shows that C/20 will discharge the battery at 1 A current in 20 hours.

The condition of each cell can be determined by the **specific gravity** (SG) of its electrolyte. When the battery is charged, the above process is reversed. The lead sulphate on the positive plate is returned to lead peroxide. The negative plate is returned to lead, and the electrolyte is restored to its original specific gravity; SG ranges will be from 1.25–1.3 (charged) down to 1.15–1.2 when discharged. Table 5.1 summarizes the chemical aspects of a charged and discharged lead-acid cell.

5.3.3 Maintenance

Flooded lead-acid batteries are susceptible to damage at low temperatures due to freezing of the electrolyte causing plate damage. The point at which the electrolyte freezes depends on its specific gravity; at a specific gravity of 1.15 (discharged) the freezing point is −15°C. To prevent freezing, the specific gravity should be maintained at higher levels; at a specific gravity of 1.275 (charged) the freezing point

Table 5.1 Chemical aspects of a charged and discharged lead-acid cell.

State	*Positive plate*	*Negative plate*	*Electrolyte*
Charged	Lead dioxide (PbO_2)	Lead (Pb)	Concentrated sulphuric acid (H_2SO_4)
Discharged	Lead sulphate ($PbSO_4$)	Lead sulphate ($PbSO_4$)	Weak sulphuric acid (H_2SO_4)

is −62ºC. Although this guards against freezing, the consequence of maintaining a battery in this condition is that it will gradually self-discharge.

Lead-acid batteries require a three-month **capacity check**, and have approximately 18–24 months' life. The condition of a fully charged lead-acid battery can be confirmed by three factors:

- The terminal voltage remains at its maximum level
- There is a free discharge of gas
- The SG is in the range 1.25–1.3.

The **specific gravity** of the electrolyte provides the definitive means of checking the charged condition of a lead-acid cell; this must be checked with a hydrometer on a periodic basis. (Specific gravity of a fluid is the relative density, or ratio of fluid's weight compared to pure water.) The electrolyte must always cover the plates; it can be topped up with distilled water. Differences of specific gravity readings between cells indicates that the battery is reaching the end of its useful life.

Key maintenance point

Specific gravity is temperature-dependent, and correction factors must be applied when taking measurements (refer to maintenance manual instructions).

Key maintenance point

When taking specific gravity readings from lead-acid cells, the acid sample should always be returned into the same cell.

Charging of lead-acid batteries should be from a constant voltage source. Excessive charging rates can lead to boiling of the electrolyte; fumes containing droplets of electrolyte can escape the battery. These fumes can become noxious unless the battery is properly ventilated. The voltage per cell during charging should not exceed 2.35 V.

Sulphation occurs when an excess of lead sulphate builds up on the plates. This happens with a fully charged battery over a period of several weeks when the battery self-discharges. To prevent this, the battery should be re-charged in accordance with the maintenance manual instructions. Sulphation can eventually occur on a permanent basis and the sulphate will not go back into solution when charged. Over time, the lead sulphate gradually occupies more space on the plates thereby reducing capacity. This can be removed by drawing a heavy charge current causing particles to be removed from the plates and subsequently accumulated at the bottom of the cell. Eventually the plates become uneven in cross-section and distorted, leading to cracks being formed. Particles will accumulate at the bottom of the cell and this can lead to shorting of the plates. Sulphating is accelerated by small (trickle) discharging/charging together with incorrect electrolyte strength and levels.

In the event of electrolyte **spillage/leaks**, (always refer to the aircraft maintenance manual for any specific requirements) the following generic actions should be taken:

1. Report the incident.
2. Mop-up the electrolyte with a damp rag or sponge.
3. Brush the affected area with a dilute solution of sodium bicarbonate.
4. Sponge the area with clean water; dry thoroughly.
5. Press a moist piece of blue litmus paper on the affected area; a change of colour to red indicates the presence of acid (repeat steps 3–4 until the acid is removed).
6. Leave for 24 hours, and then check for any evidence of corrosion.
7. Restore any protective finish to the aircraft structure.

5.3.4 Sealed batteries

Maintenance and servicing costs associated with flooded cells can be overcome with sealed lead-acid batteries. This technology was developed in the 1970s and has been in place since the 1980s and is known as **valve-regulated lead-acid** (VRLA); the sealed lead-acid (SLA) effectively provides maintenance-free lead-acid batteries.

Cell plates are made from lead calcium; the electrolyte is sulphuric acid diluted with distilled water. Plates are separated by an **absorbent glass mat** (AGM) that absorbs gasses liberated from the plates during charging. The lead plates are purer (99.99%) than flooded cell materials since they do not have to support their own weight. The electrolyte is absorbed between the plates and immobilized by a very fine fibreglass mat. This glass mat absorbs and immobilizes the acid while keeping the electrolyte in contact with the plates. This allows a fast reaction between and electrolyte and plate material during charge/discharge. There is no disintegration of the active materials leading to a short-circuit.

Key maintenance point

It is not possible to check the SG of the electrolyte; the battery can only be checked by measuring the terminal voltage.

The **internal resistance** of sealed lead-acid cells is lower than flooded cells, they can handle higher temperatures, and self-discharge more slowly. They are also more tolerant to the attitude of the aircraft. The product is inherently safer than the flooded cell due to reduced risk of spillage, leakage and gassing. Maintenance requirements are for a capacity check only. The overall capacity-to-weight ratio of sealed lead-acid batteries is superior to flooded lead-acid batteries. Since they are sealed, they can be shipped as non-hazardous material via ground or air.

5.4 Nickel-cadmium batteries

Nickel-cadmium battery technology became commercially available for aircraft applications in the 1950s. At that time the major sources of batteries for aircraft were either vented lead-acid or silver-zinc technology. The nickel-cadmium (Ni-Cd) battery (pronounced ‘nye-cad’) eventually became the preferred battery type for larger aircraft since it can withstand higher charge/discharge rates and has a longer life. Ni-Cd cells are able to maintain a relatively steady voltage during high discharge conditions. The disadvantages of nickel-cadmium batteries are that they are more expensive (than lead-acid batteries) and have a lower voltage output per cell (hence their physical volume is larger than a lead-acid battery).

5.4.1 Construction

Plates are formed from a nickel mesh on which a nickel powder is sintered. The sintering process (where powdered material is formed into a solid) is used to form the porous base-plates (called **plaques**). This process maximizes the available quantity of active material. The plaques are vacuum impregnated with nickel or cadmium salts, electrochemically deposited with the pores of the plaques. Nickel tabs are spot-welded onto the plates and formed into the terminals; these plates are then stacked and separated by a porous plastic in a similar fashion to the lead-acid battery. The electrolyte is potassium hydroxide (KOH) diluted in distilled water giving a specific gravity of between 1.24 and 1.3. Both the plates and electrolyte are sealed in a plastic container.

5.4.2 Charging

During charging, there is an exchange of ions between plates. Oxygen is removed from the negative plate, and transferred to the positive plate. This transfer takes place for as long as charging current exists, until all the oxygen is driven out of the negative plate (leaving metallic cadmium) and the positive plate becomes nickel oxide. The electrolyte acts as an ionized conductor and it does not react with the plates in any way. There is virtually no chemical change taking place in the electrolyte during charging or discharging, therefore its condition does not provide an indication of cell condition. Towards the end of charging, gassing occurs as a result of **electrolysis** and the water content of the electrolyte is reduced. Gas emitted by decomposition of water molecules is converted into hydrogen at the negative plate and oxygen at the positive plate. This gassing leads to the loss of some water; the amount of gas released is a function of electrolyte temperature and charging voltage. When fully charged, each cell has a potential difference of between 1.2 and 1.3 V across its terminals. This reduces to 1.1 V when discharged. An aircraft battery containing 19 cells at 1.3 V therefore produces a battery of 24.7 V. Charging voltage depends on the design and construction, but will be in the order of 1.4/1.5 V per cell.

5.4.3 Discharging

This is a reverse chemical activity of the charging process; the positive plate gradually loses oxygen and the negative plate gradually regains oxygen. No gassing takes place during a normal discharge; the electrolyte is absorbed into the plates and may not be visible over the plates. When fully charged, the volume of electrolyte is high; this is the only time that water should be added to a Ni-Cd battery. Ni-Cd battery electrolyte freezes at approximately –60ºC and is therefore less susceptible to freezing compared to lead-acid. The formation of white crystals of potassium carbonate indicates the possibility that overcharging has occurred.

Referring to Fig. 5.5, the nickel-cadmium cell voltage remains relatively constant at approximately 1.2V through to the end of discharge, at which point there is a steep voltage drop. The discharge characteristics of a cell are affected by the:

- discharge rate
- discharge time
- depth of discharge
- cell temperature
- charge rate and overcharge rate
- charge time, and rest period after charge
- previous cycling history.

Every nickel-cadmium cell (and hence a battery) has a specific:

- rated capacity
- discharge voltage
- effective resistance.

Individual cells are rated at a nominal 1.2V, and voltage for battery voltages are multiples of the individual cell nominal voltage. Five cells connected in series would therefore result in a 6 Vbattery. It can be seen from Fig. 5.6 that the discharge voltage will exceed 1.2V for some portion of the discharge period. Cell capacity is normally rated by stating a conservative estimate of the amount of capacity that can be discharged from a relatively new, fully charged cell. The cell rating in ampere-hours (or milliampere-hours) is therefore quoted by most manufacturers to a voltage of 0.9V at 5 hour discharge rate.

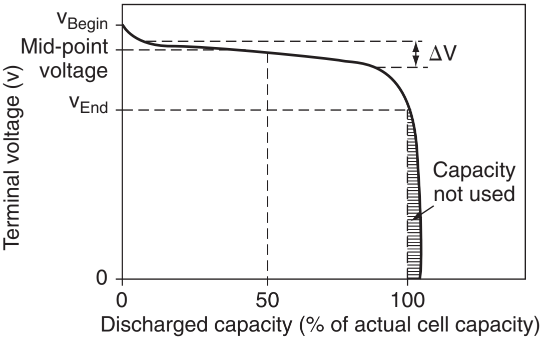

Figure 5.5 Nickel-cadmium cell discharge characteristics

Figure 5.6 shows that when rates of discharge are reduced, the available capacity becomes less dependent on the discharge rate. When rates of discharge rates increase, the available capacity decreases.

Charging of nickel-cadmium batteries needs specific methods since they can suffer from an effect called **thermal runaway**. This occurs at high temperatures and if the battery is connected to a constant charging voltage that can deliver high currents. Thermal runaway causes an increase in temperature and lower internal resistance, causing more current to flow into the battery. In extreme cases sufficient heat may be generated to destroy the battery. Dedicated battery charges (either on the aircraft or in the workshop) are designed to take this into account by regulating the charging current. Temperature sensors are installed in the batteries to detect if a runaway condition is occurring. Table 5.2 summarizes the chemical aspects of a charged and discharged nickel-cadmium cell.

5.4.4 Maintenance

Since there is virtually no chemical change taking place during nickel-cadmium cell charging or discharging, the condition of the electrolyte does not provide an indication of the battery's condition. Cell terminal voltage does not provide an indication of charge since it remains relatively constant. The only accurate and practical way to determine the condition of the nickel-cadmium battery is with a **measured discharge** in the workshop. The fully charged battery is tested after a two hour 'resting' period, after which the electrolyte is topped up using distilled or demineralized water. Note that since the electrolyte level depends on the state of charge, water should never be added to the battery on the aircraft. This could lead to the electrolyte overflowing when the battery discharges, leading to corrosion and self-discharging (both of which could lead to premature failure of the battery). Ni-Cd batteries emit gas near the end of the charging process and during overcharging. This is an explosive mixture and must

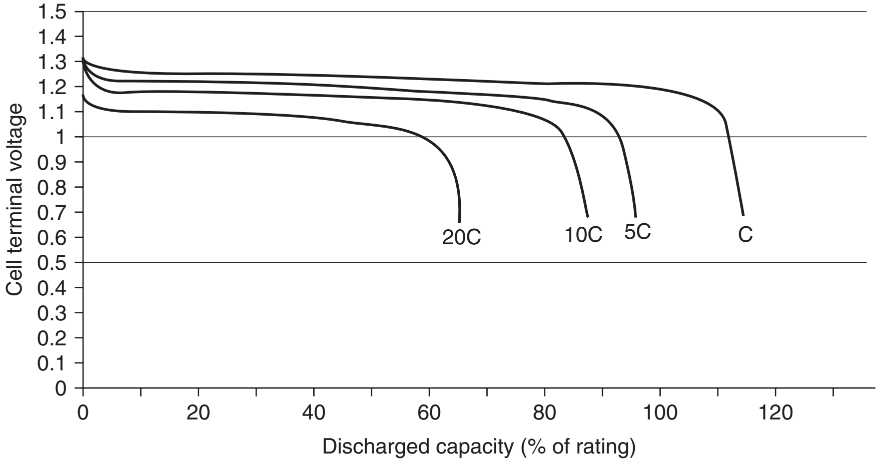

Figure 5.6 Nickel-cadmium cell discharge profiles and capacity

Table 5.2 Chemical aspects of a charged and discharged nickel-cadmium cell

State	*Positive plate*	*Negative plate*	*Electrolyte*
Charged	Nickel oxides (Ni_2O_2 Ni_3O_3)	Cadmium (Cd)	Potassium hydroxide (KOH)
Discharged	Nickel hydroxide $Ni(OH_2)$	Cadmium hydroxide $Cd(OH_2)$	Potassium hydroxide (KOH)

be prevented from accumulating; maintenance of the venting system is essential.

In the event of electrolyte **spillage/leaks** (always refer to the aircraft maintenance manual for specific details):

- report incident
- mop electrolyte with damp rag or sponge
- cover the area with a dilute solution of acetic acid, 5% solution of chromic acid, or 10% solution of boric acid
- press moist piece of red litmus paper on affected area; change of colour to blue indicates presence of alkaline
- leave for a minumum of 24 hours, check for corrosion
- restore protective finish.

In addition to providing primary power, Ni-Cd batteries are also used in aircraft for emergency equipment, e.g. lighting. This type of cell is sealed and the electrolyte cannot be topped up. Extreme care must be taken with how these batteries are charged.

Key maintenance point

Servicing equipment used for lead-acid batteries must not be used for nickel-cadmium batteries; sulphuric acid is detrimental to the operation of nickel-cadmium batteries.

Key maintenance point

If Ni-Cd batteries are replacing lead-acid batteries, always neutralize the battery compartment.

Key maintenance point

Electrolyte used in lead-acid and nickel cadmium batteries is actively corrosive.

Key maintenance point

Main batteries need to be kept upright in the aircraft to avoid spilling any of the electrolyte.

5.5 Lithium batteries

Lithium batteries include a family of over 20 different products with many types of anodes, cathodes and electrolytes. The type of materials selected depends on many factors, e.g. cost, capacity, temperature, life etc.; these are all driven by what the application requirements are.

Applications range from consumer products (accounting for the largest market requirement) through to specialist applications including communications and medical equipment. Aircraft are often equipped with systems requiring an autonomous source of energy, e.g. emergency locator beacons, life rafts and life jackets. Lithium (Li) is one of the alkali group of **reactive metals**; it is one of the lightest elements, giving it an immediate advantage for aircraft applications. It has a single valence electron with low combining power, therefore readily becoming a positive ion. The materials used in these cells are:

- electrolyte: lithium-ion
- cathode: cobalt
- anode: graphite.

Lithium-ion is a fast-growing and promising battery technology. This type of battery is often found in consumer products (mobile phones and laptop computers) because they have very high energy-to-weight ratios, no memory effect, and a slow discharge charge rate when not in use. They are being introduced for aircraft applications (e.g. in smoke detectors) on a cautious basis because they are significantly more susceptible to thermal runaway. Applications on aircraft now include engine start and emergency back-up power, the first such application of the devices in the business aviation sector. In the longer term, they are being developed for main battery applications. They offer several advantages compared to lead-acid and nickel-cadmium products, including:

- longer life
- less weight
- low maintenance
- reduced charging time.

Disadvantages are the higher product cost and the fact that the electrolyte is extremely flammable. They can lose up to 10% of their storage capacity every year from when they are manufactured, irrespective of usage. The rate at which the ageing process occurs is subject to temperature; higher temperatures results in faster ageing.

The lithium-ion main aircraft battery will not be a ‘drop-in’ replacement for main battery applications. Safety features are required within the aircraft as well as in the battery. These features include protection circuits and hardware to maintain voltage and current within safe limits. The nominal cell voltage is 3.6V, charging requires a constant voltage of 4.2V with associated current limiting. When the cell voltage reaches 4.2V, and the current drops to approximately 7% of the initial charging current, the cell is fully recharged. Figure 5.7 illustrates the typical discharge curve of a lithium-ion cell when discharged at the 0.2 C rate. Lithium-ion cells have a very flat discharge curve, and cell voltage cannot be used to determine the state of charge. The effective capacity of the lithium-ion cell is increased with low discharge rates and reduced if the cell is discharged at higher rates.

Software-based monitoring and alarms are needed for safe operation during charging. Specific design and maintenance considerations for these batteries in aircraft include:

- maintaining safe cell temperatures and pressures
- mitigating against explosion
- preventing the electrolyte escaping from the battery
- disconnecting the charging source in the event of over-temperature
- providing a low battery charge warning.

5.6 Nickel-metal hydride batteries

Nickel-metal hydride (Ni-MH) is a secondary battery technology, similar to the sealed nickel-cadmium product. Ni-MH batteries provide a constant voltage during discharge, excellent long-term storage and long cycle life (over 500 charge–discharge cycles). No maintenance is required on this type of battery; however, care must be taken in charging and discharging. The evolution of Ni-MH technology is being driven by the need for environmentally friendly materials and higher energy efficiency. The materials used in the Ni-MH battery technology are:

- anode: nickel and lanthanum
- cathode: nickel hydroxide
- electrolyte: potassium hydroxide.

The charging voltage is in the range of 1.4/1.6V per cell. A fully charged cell measures between 1.35 and 1.4V (unloaded), and supplies a nominal 1.2V

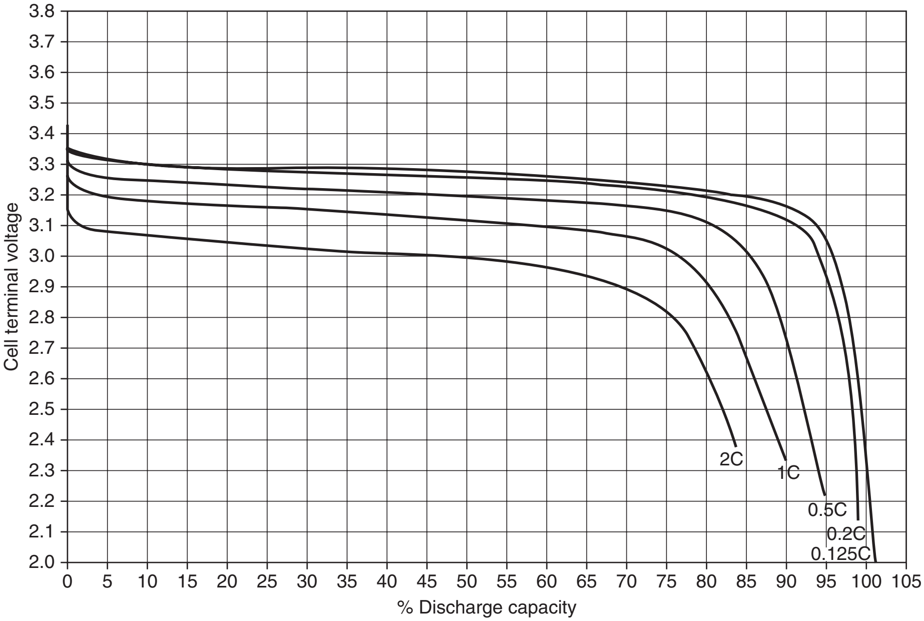

Figure 5.7 Lithium-ion cell discharge characteristics

per cell during use, reducing to approximately 1 volt per cell (further discharge may cause permanent damage). The Ni-MH cell requires a complex charging algorithm, and hence dedicated charger equipment.

Figure 5.8 illustrates the voltage profile of a metal-hydride cell, discharged at the 5-hour rate (0.2 C rate). This profile is affected by temperature and discharge rate; however under most conditions, the cell voltage retains a flat plateau that is ideal for electronics applications. As with nickel-cadmium cells, the nickel-metal hydride cell exhibits a sharp 'knee' at the end of the discharge where the voltage drops rapidly.

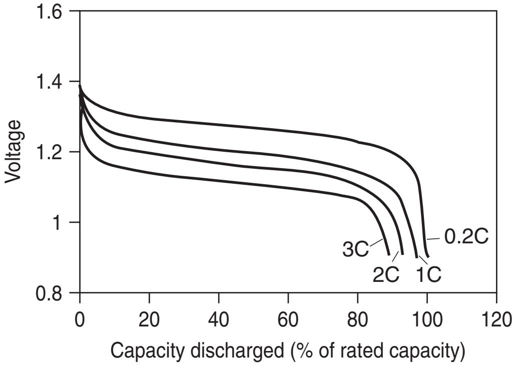

Figure 5.8 Metal-hydride cell discharge characteristics

Key maintenance point

A Ni-Cad charger should not be used as a substitute for a Ni-MH charger.

A new generation of nickel-metal hydride 12V batteries has been designed by Advanced Technological Systems International Limited (ATSI) as a direct replacement for the conventional sealed lead-acid battery typically used in gliders. It delivers more than twice the power of its lead acid counter part whilst having the same base footprint and lower weight. The integral advanced electronics guarantees that it will always deliver maximum output up to the point of total discharge. Unlike sealed lead-acid batteries, it does not suffer any loss of performance even after many deep discharge cycles, or storage in a discharged state, making it one of the most advanced batteries in the world today. The new battery type will be longer lasting than the equivalent sealed

Table 5.3 Metal hydride aircraft battery: technical specifications (courtesy of ATSI)

Dimensions	65 × 95 × 150 mm
Weight	1.9 kg
Nominal voltage	12 V
Capacity	10 Ah
Maximum discharge	5 A
Fuse	Self setting internal
Charge time	8 hours flat to full
Minimum charge cycles	600
Operating temperatures	−20/+60°C

lead-acid battery and requires a purpose-designed charging unit.

5.7 Battery locations

An aircraft is fitted with one or two main batteries depending on its size and role. The battery is located as close as possible to its point of distribution; this is to reduce IR losses through heavy-duty cables. In smaller general aviation (GA) aircraft, the battery can be located in the engine compartment, alternatively behind the luggage compartment in the rear fuselage, see Fig. 5.9(a). On some larger GA aircraft the battery is located in the leading edge of the wing, see Fig. 5.9(b). Other locations include the nose equipment bay on medium size helicopters (Fig. 5.9(c)) or attached to the **external airframe**, see Fig. 5.9(d). For larger aircraft, e.g. the Boeing 747, one battery is located in the **flight compartment**; the other is located in the **auxiliary power unit** (APU) bay at the rear of the aircraft. Batteries are installed in a dedicated box or compartment designed to retain it in position and provide ventilation. The battery compartment is usually fitted with a tray to collect any spilt electrolyte and protect the airframe. Tray material will be resistant to corrosion and non-absorbent. The structure around the battery compartment will be treated to reduce any damage from corrosion resulting from any spilt electrolyte or fumes given off during charging. Batteries must be secured to prevent them from becoming detached during aircraft manoeuvres; they are a **fire risk** if they become detached from their tray.

Key maintenance point

When installing batteries in the aircraft, extreme care must be taken not to directly connect (or 'short circuit') the terminals. This could lead to a high discharge of electrical energy causing personal harm and/or damage to the aircraft.

Key maintenance point

The battery must be secured without causing any deformation of the casing which could lead to plate buckling and internal shorting.

5.8 Battery venting

Main battery installations must be vented to allow gases to escape, and accommodate electrolyte spillage. Rubber or other non-corroding pipes are used as ventilation lines which direct the gases overboard, usually terminating at the fuselage skin. On pressurized aircraft the differential pressures between cabin and atmosphere are used to draw air through the venting system. Some installations contain traps to retain harmful gases and vapours. Figure 5.10 illustrates battery venting, acid traps and how pressurized cabin air is used to ventilate the battery.

Key maintenance points

- Avoid personal contact with battery electrolyte (fluid and fumes).
- Observe safety precautions for the protection of hands and eyes.
- Always use personal protective equipment (goggles, rubber gloves, aprons) when handling electrolyte to prevent serious burns.
- Seek first aid in the event of electrolyte contact.
- When mixing electrolyte, acid is always added to the water. (Adding water to acid is very dangerous.)

5.9 Battery connections

These depend on the type of battery and aircraft installation. On smaller aircraft the cable connections

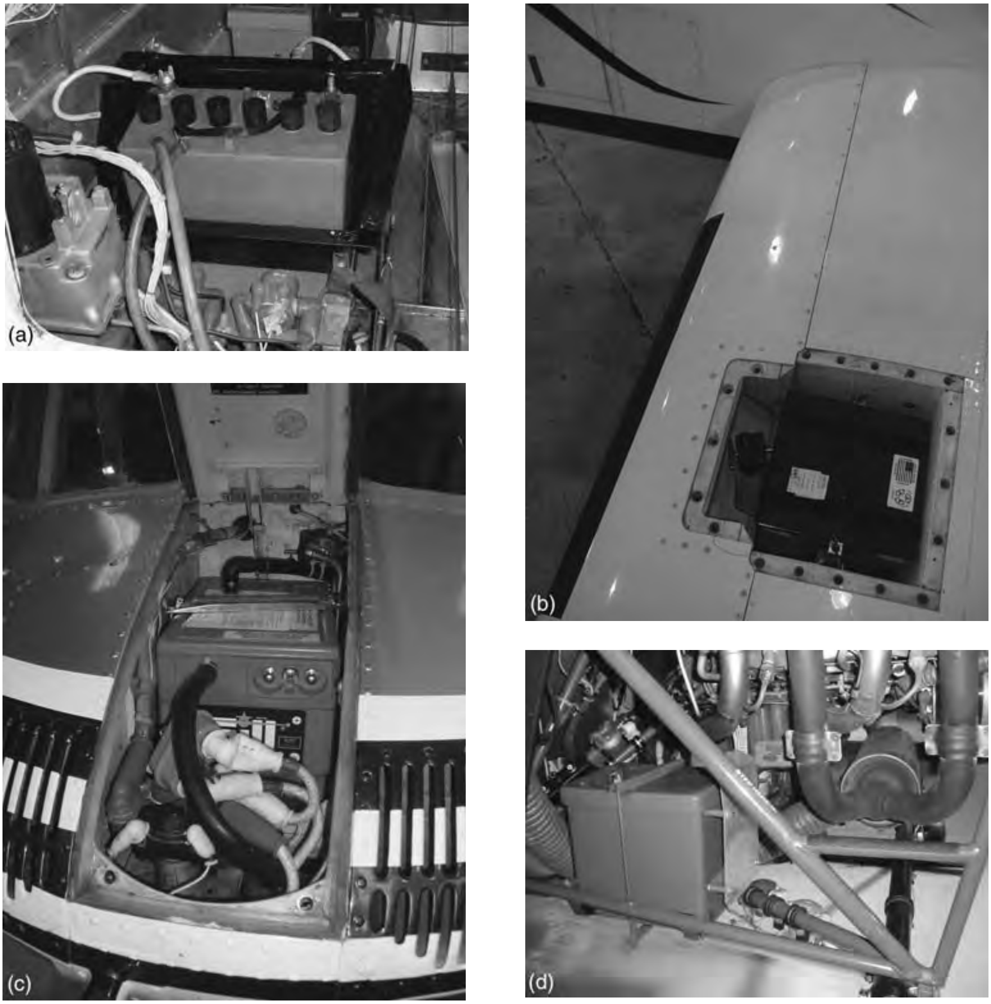

Figure 5.9 Typical battery locations: (a) battery compartment (GA aircraft); (b) wing leading edge (Beech King Air); (c) nose equipment bay (medium helicopter); (d) externally mounted (small helicopter)

simply fit over the terminal lugs and are secured with a nut, bolt and washers. On larger aircraft, the main batteries have **quick-release** connectors, see Fig. 5.11. These provide protection for the terminals and cable connections, the aircraft connector is a plastic housing with two shrouded spring-loaded terminals (for connecting the battery cables) and a hand-wheel with lead-screw. The battery connection is a plastic housing integrated into the casing; it contains two shrouded pins and a female lead screw. When the two halves are engaged, the lead screws are pulled together and eventually form a lock. This mechanism provides good contact pressure and a low resistance connection. The main battery(s) is connected into the aircraft distribution system; this is described in Chapter 7.

Key maintenance point

Batteries must not be exposed to temperatures or charging currents in excess of their specified values. This will result in the electrolyte boiling, rapid deterioration of the cell(s) eventually leading to battery failure.

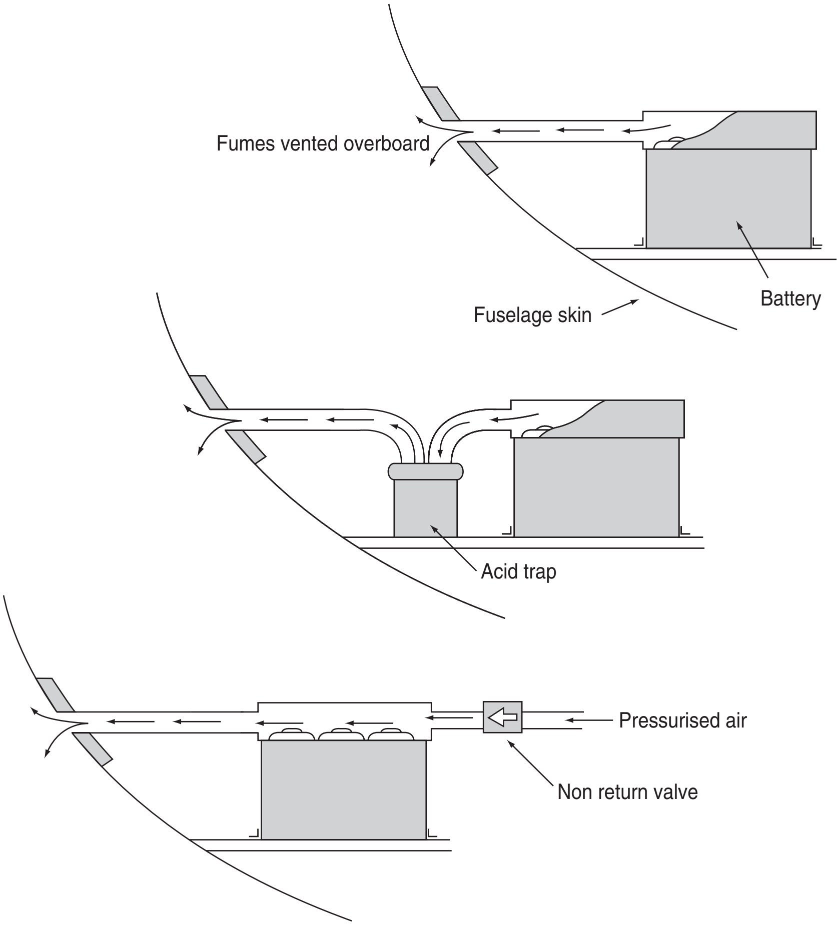
Fumes vented overboard
Battery
Fuselage skin
Acid trap
Pressurised air
Non return valve

Figure 5.10 Battery venting

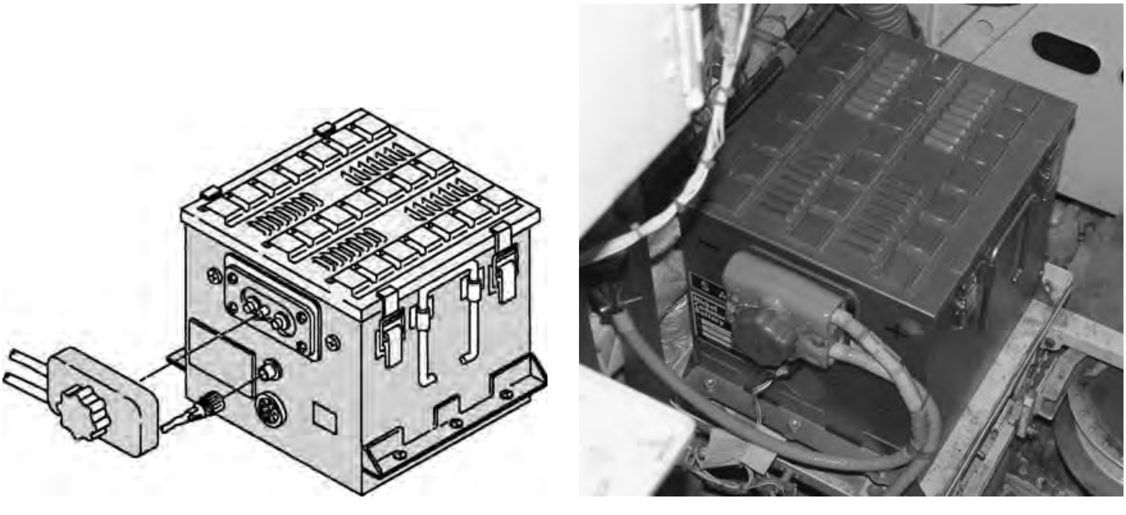

Figure 5.11 Battery connections

Key maintenance point

Removal of the aircraft battery can result in loss of power to any clocks that are electrically. It will usually be necessary to check and reset the clocks on the flight deck when battery power is eventually restored.

Key maintenance point

Some aircraft main batteries are heavy and may require a hoist for removal/installation into the aircraft.

5.10 Multiple choice questions

1. In a simple cell, electrons are:
   (a) removed from the (positive) cathode and deposited on the (negative) anode
   (b) removed from the (negative) cathode and deposited on the (positive) anode
   (c) removed from the (negative) anode and deposited on the (positive) cathode.

2. The energy storage capacity of a cell is determined by the:
   (a) terminal voltage
   (b) electrolyte specific gravity
   (c) amount of material available for chemical reaction.

3. When mixing electrolyte:
   (a) acid is always added to the water
   (b) water is always added to the acid
   (c) it is not important how water and acid are mixed.

4. Servicing equipment used for lead-acid batteries:
   (a) can also be used for nickel-cadmium batteries
   (b) must not be used for nickel-cadmium batteries
   (c) must be disposed of after use.

5. Battery capacity is measured in:
   (a) volts
   (b) amperes
   (c) ampere-hours.

6. Lead acid batteries are recharged by constant:
   (a) voltage
   (b) current
   (c) ampere-hours.

7. The only accurate and practical way to determine the condition of the nickel-cadmium battery is with a:
   (a) specific gravity check of the electrolyte
   (b) measured discharge in the workshop
   (c) check of the terminal voltage.

8. A cell that can only be charged once is called a:
   (a) secondary cell
   (b) metal-hydride cell
   (c) primary cell.

9. The only time that water should be added to a Ni-Cd battery is:
   (a) when fully charged, and the volume of electrolyte is high
   (b) when fully discharged, and the volume of electrolyte is high
   (c) when fully charged, and the volume of electrolyte is low.

10. The only accurate and practical way to determine the condition of the lead-acid battery is with a:
    (a) specific gravity check of the electrolyte
    (b) measured discharge in the workshop
    (c) check of the terminal voltage.

11. Referring to Fig. 5.12 (page 126), the reason for having an even number of positive plates in a lead-acid battery is because:
    (a) positive plates distort when chemical reactions take place on both sides
    (b) positive plates distort when chemical reactions take place on one side
    (c) negative plates distort when chemical reactions take place on both sides.

12. Referring to Fig. 5.13 (page 126), a 20 Ah battery, when discharging at 2A, will be fully discharged in approximately:
    (a) two hours
    (b) 15 minutes
    (c) ten hours.

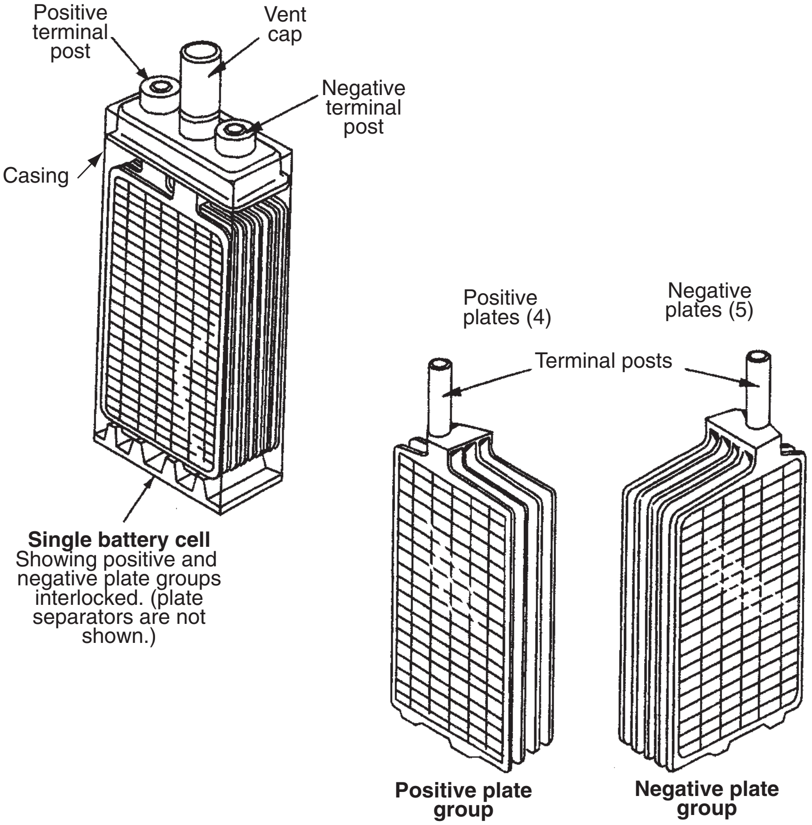
Positive terminal post
Vent cap
Negative terminal post
Casing
Single battery cell
Showing positive and negative plate groups interlocked. (plate separators are not shown.)
Positive plates (4)
Negative plates (5)
Terminal posts
Positive plate group
Negative plate group

Figure 5.12 See Question 11

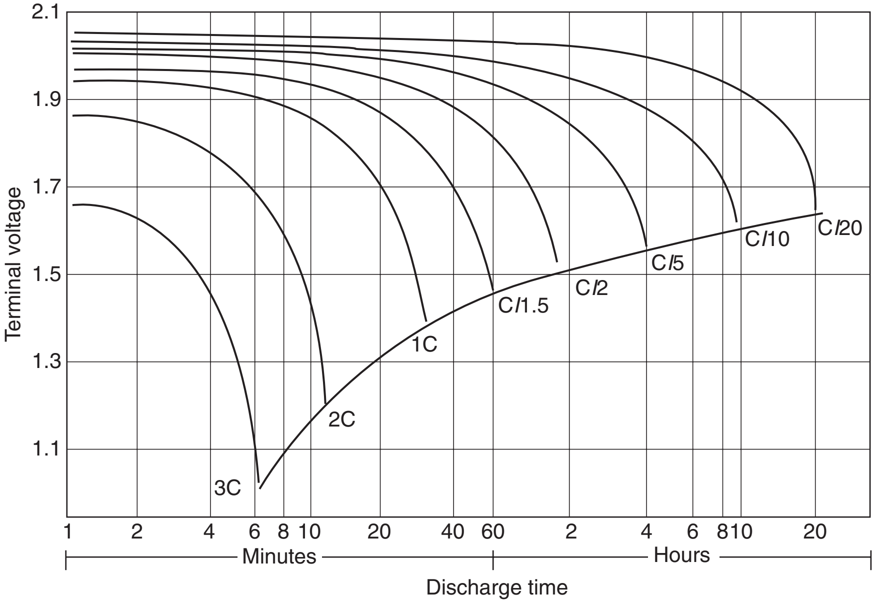

Figure 5.13 See Question 12

Chapter 6 Power supplies

Aircraft electrical power can be derived from a variety of sources; these are categorized as either primary or secondary sources. Batteries and generators are primary sources of electrical power; inverters and transformer rectifier units (TRU) are secondary sources of power. This power is either in the form of direct or alternating current depending on system requirements. In addition to onboard equipment, most aircraft have the facility to be connected to an external power source during servicing or maintenance. The basic power source found on most aircraft is the battery, delivering direct current (DC). Generators can supply either direct or alternating current; the outputs of generators need to be regulated. Alternating current generators are also referred to as alternators.

Inverters are used to convert DC (usually from the battery) into alternating current (AC). Transformer rectifier units (TRU) convert AC into DC; these are often used to charge batteries from AC generators. In some installations transformers (as described in Chapter 1) are used to convert AC into AC, typically for stepping down from 115 to 26V AC. An auxiliary power unit (APU) is normally used for starting the aircraft's main engines via the air distribution system. While the aircraft is on the ground, the APU can also provide electrical power. In the event of generator failure(s), continuous power can be provided by a ram air turbine (RAT). In this chapter, we review the various sources of electrical power used on aircraft and their typical applications.

6.1 Regulators

We know from basic theory that a generator's output will vary depending on the input shaft speed. A means of regulating the generator's output is therefore required.

6.1.1 Vibrating contact regulator

This device comprises voltage and current regulators as shown in Fig. 6.1. They are used on small general

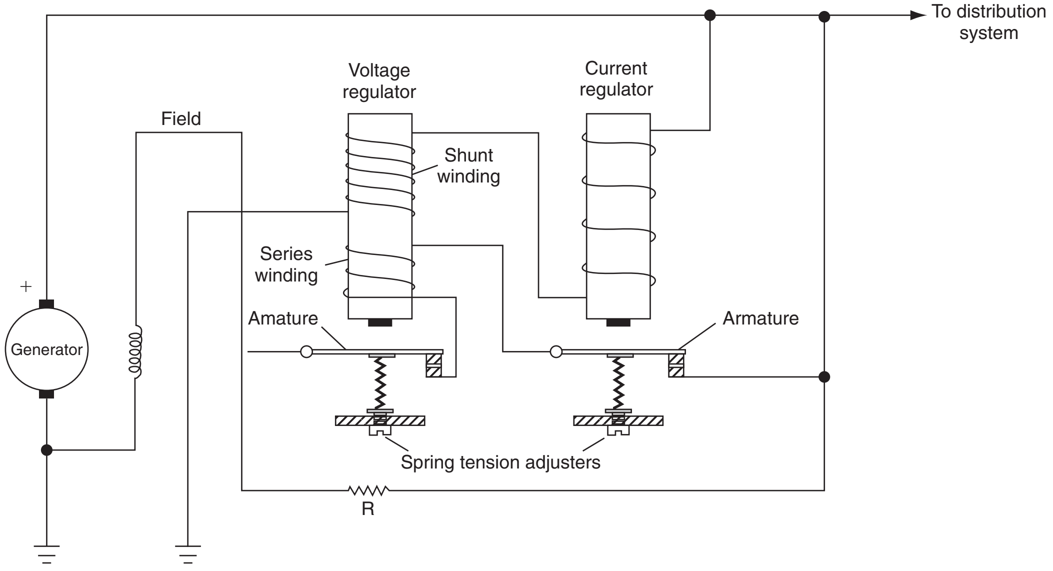

Figure 6.1 Vibrating contact regulator schematic

aviation (GA) aircraft that have relatively low generator power outputs. When the engine starts, the alternator output voltage builds up rapidly to the nominal aircraft level (either 14 or 28 V DC). Contacts of both regulators remain closed to allow current to flow into the field windings. When the generator output **voltage** increases beyond 14/28 V, the voltage coil contacts open and this introduces the resistor into the field windings, thereby reducing the field excitation current, and subsequently reduces the generator output. Once the output voltage drops to below 14/28 V, the contacts close (by a spring mechanism) and the resistor is bypassed, allowing full excitation current back into the field. The on/off cycle repeats between 50 and 200 times per second, or 50–200 Hz. This process regulates the generator output to a mean level, typically 14 ± 0.5 V (or 28 ± 1 volt).

Current regulation is achieved in a similar way, i.e. by controlling the field current. When loads are high, the voltage output may be insufficient to open the contacts. The result is that the output will continue to increase until the maximum rated current is reached. At this point, the current regulator contacts open and the resistor is connected into the field windings.

The accuracy of this type of regulation depends on the resistor value and spring tensions. In the event of high rotor speed and low electrical load on the generator, the output could exceed the specified system voltage despite the field being supplied via the resistor. In this event, the contact is pulled to ground, thereby reducing the output to below the regulated mean level. Although simple, this type of regulator has the disadvantages of contact wear; a typical vibrating contact regulator product is shown in Fig. 6.2.

Key maintenance point

The accuracy of the vibrating contact regulator depends on the resistor value and spring tension.

6.1.2 Carbon-pile regulator

Another type of electromechanical regulator is the carbon-pile device. This type of regulator is used in generator systems with outputs in excess of 50 A and provides smoother regulation compared with the vibrating contact regulator. Carbon-pile regulators consists of a variable resistance in series with the generator's shunt wound field coil. The variable resistance is achieved with a stack (or pile) of **carbon discs** (washers). These are retained by a ceramic rube that keeps the discs aligned. Figure 6.3 shows the main features of the regulator in cross-section. The surface of each disc is relatively rough; applying pressure to the discs creates more surface contact, thereby reducing the resistance of the pile. When pressure is reduced, the reverse process happens, and the resistance through the pile increases. Pressure is applied to the pile by a spring plate. This compression is opposed by the action of an electromagnet connected to the generator output; the strength of the electromagnet's flux varies in proportion with generator output voltage.

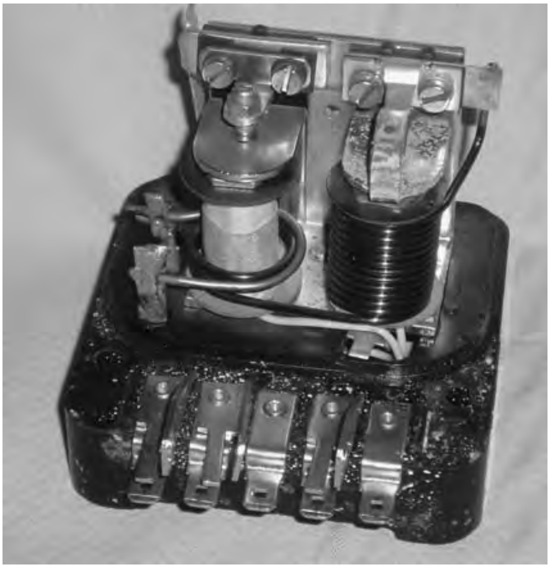

Figure 6.2 Vibrating contact regulator overview

Higher generator output increases the current in the electromagnet; this attracts the steel centre of the spring, which reduces compression on the pile, thereby increasing its resistance. Less field current reduces the generator output voltage; the current in the voltage coil reduces electromagnetic effect and the spring compresses the pile, reducing its resistance. The varying force applied by the electromagnet and spring thereby controls the pile's resistance to control field current and maintains a constant generator output voltage. The regulator is contained within a cylinder (typically three inches in diameter and six

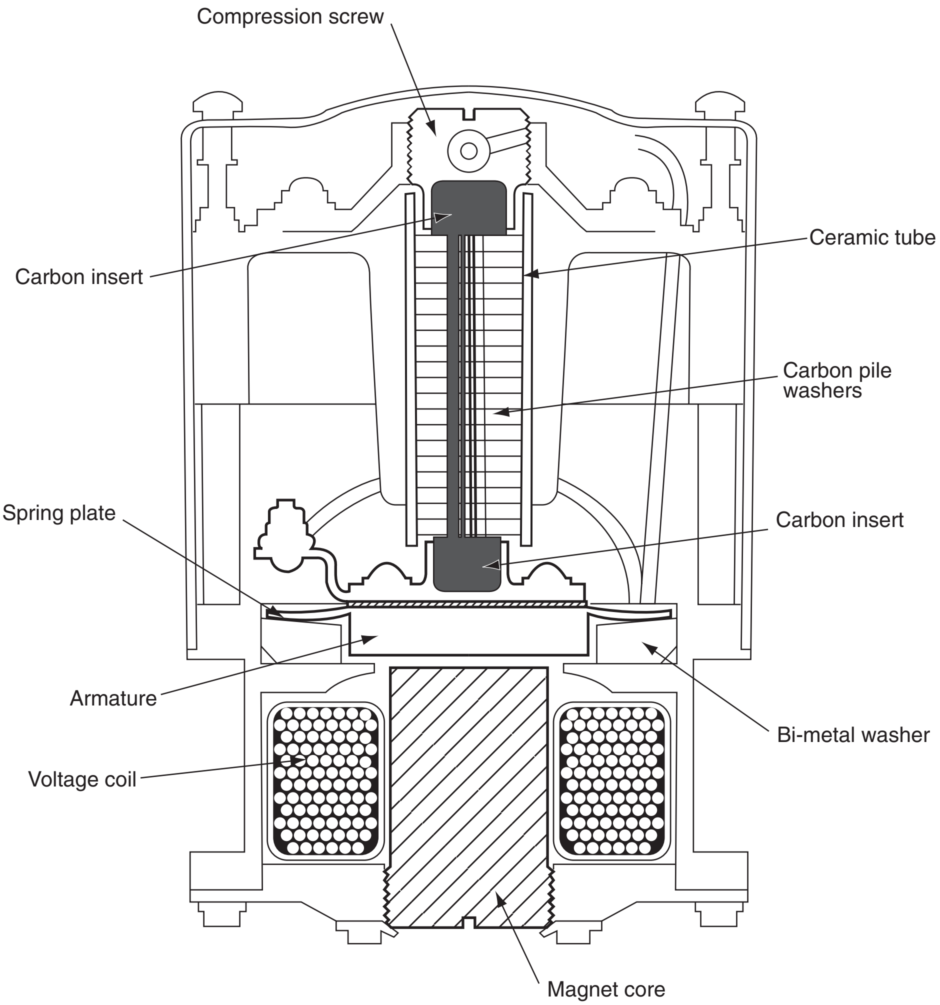
Compression screw
Carbon insert
Ceramic tube
Carbon pile washers
Spring plate
Carbon insert
Armature
Bi-metal washer
Voltage coil
Magnet core

Figure 6.3 Carbon-pile regulator – cross section

inches in length) with cooling fins. Functions of each component are as follows:

- **Compression screw**: the means of setting up compression on the pile and compensating for erosion of the pile during its life.
- **Spring plate and armature**: this compresses the pile to its minimum resistance position.
- **Voltage coil:** contains a large number of turns of copper wire and, with the core screw, forms an electromagnet when connected across the generator output.
- **Magnet core**: concentrates the coil flux; it is also used for voltage adjustment during servicing.
- **Bi-metallic washers**: providing temperature compensation.

Figure 6.4 shows the carbon-pile regulator connected into the generator's regulating circuit. The ballast resistor has a low-temperature coefficient and minimizes the effects of temperature on the voltage coil. The trimmer resistors (in series with the ballast resistor) allow the generator output voltage to be trimmed on the aircraft. The **boost** resistor is normally shorted out; if the switch is opened it allows a slight increase in generator output to meet short-term increases in loading. This is achieved by temporarily reducing the current through the voltage coil. The boost resistor can either be located in the regulator and/or at a remote location for easy access during maintenance.

Key point

Higher generator output increases the current in the carbon pile regulator, which reduces compression on the pile thereby increasing its resistance.

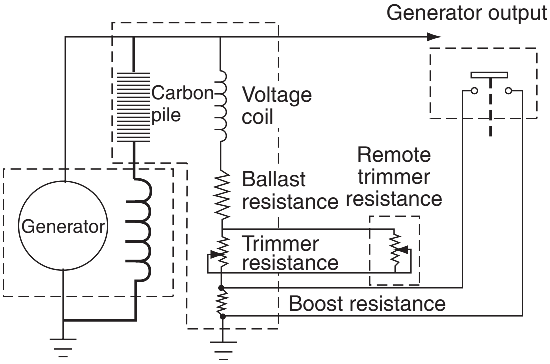
Generator output
Carbon pile
Voltage coil
Remote trimmer resistance
Ballast resistance
Generator
Trimmer resistance
Boost resistance

Figure 6.4 Carbon-pile regulator – schematic

Test your understanding 6.1

Explain the purpose of the carbon pile regulator voltage coil.

6.1.3 Electronic voltage regulator

There are many types and configurations of electronic voltage regulators. A representative type is illustrated in Fig. 6.5. The **alternator master switch** used in AC systems energizes the field relay and applies current to the base of TR_2 and the resistor network of R_1, R_2, R_{V1}. This network, together with the Zener diode (Z) is used to establish the nominal operating voltage. Current flows through the alternator's field coil via transistors TR_2 and TR_3, allowing the generator's output to increase. When the output reaches its specified value (14 or 28V DC depending on the installation) Zener diode Z conducts which turns on transistor TR_1, shorting out transistor TR_2 and TR_3. The generator voltage falls and Zener diode Z stops conducting, thereby turning of transistor TR_1. This turns transistors TR_2 and TR_3 back on, allowing the generator output to increase again. This operation is repeated

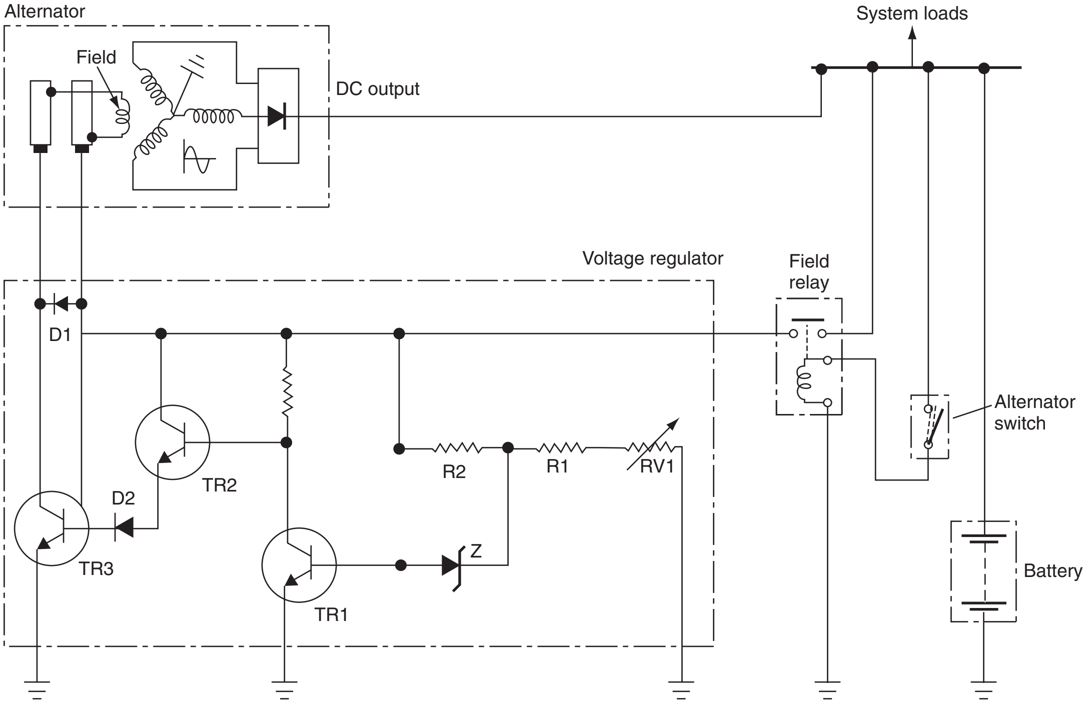
Alternator
Field
DC output
System loads
Voltage regulator
Field relay
D1
D2
TR2
TR3
TR1
R2
R1
RV1
Z
Alternator switch
Battery

Figure 6.5 Electronic voltage regulator

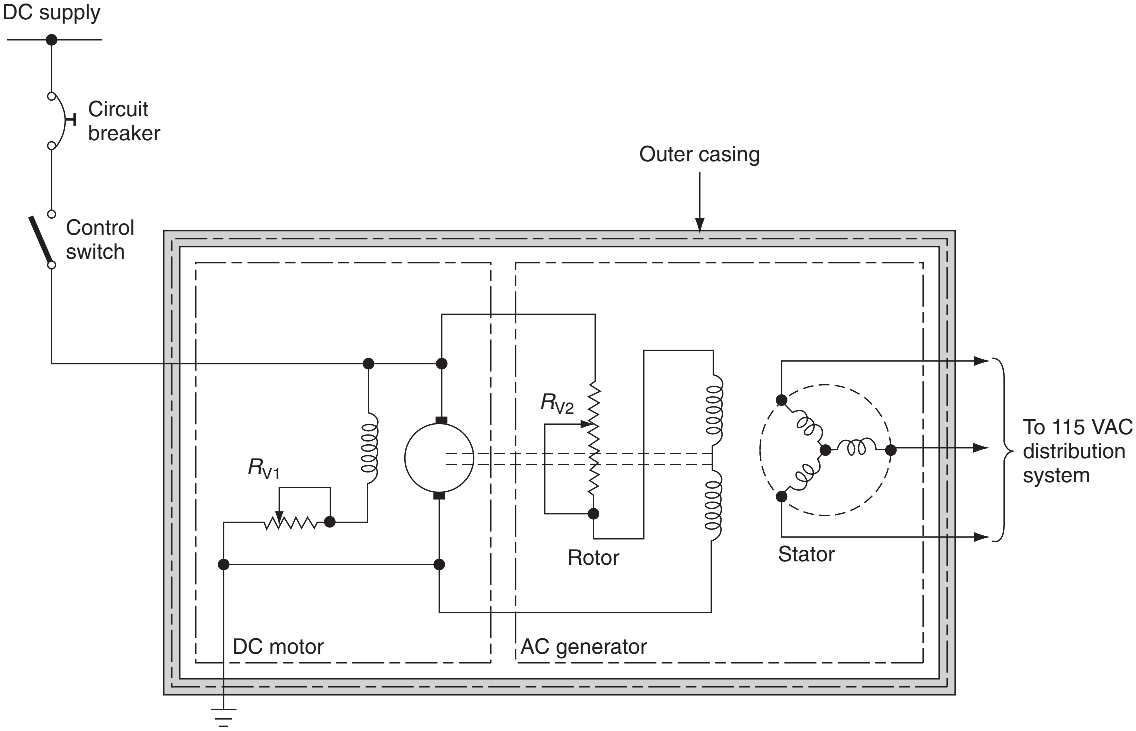

Figure 6.6 Rotary inverter schematic

many times per second as with the vibrating contact regulator; the difference being that electronic circuits have no moving parts and do not suffer from arcing across contacts. Diode D_1 provides protection against the back e.m.f. induced in the field each time TR_3 is switched. The trimming resistor R_{V1} can be used to adjust the nominal voltage output of the regulator.

6.2 External power

In addition to the onboard equipment that has been described, most aircraft have the facility to be connected to an external power source during servicing or maintenance. This allows systems to be operated without having to start the engines or use the battery. The external ground power can either be from a battery pack, a ground power unit (that has a diesel engine and generator) or from industrial power converters connected to the national grid. More details of external power and how it is used is given in Chapter 8.

6.2.1 Power conversion

Equipment used on aircraft to provide secondary power supplies include:

- inverters
- transformer rectifier units (TRU)
- transformers.

6.3 Inverters

Inverters are used to convert direct current into alternating current. The input is typically from the battery; the output can be a low voltage (26V AC) for use in instruments, or high voltage (115V AC single or three phase) for driving loads such as pumps. Older **rotary inverter** technology uses a DC motor to drive an AC generator, see Fig. 6.6. A typical rotary inverter has a four-pole compound DC motor driving a star-wound AC generator. The outputs can be single- or three-phase; 26V AC, or 115V AC. The desired output frequency of 400 Hz is determined by the DC input

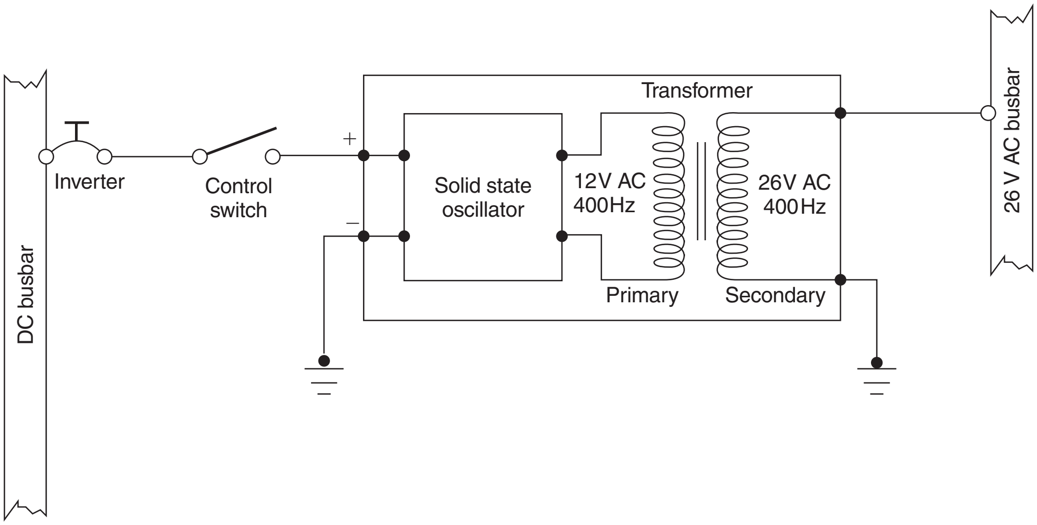

Figure 6.7 Static inverter schematic

voltage. Various regulation methods are employed, e.g. a trimming resistor (R_v) connected in series with the DC motor field sets the correct speed when connected to the 14 or 28V DC supply.

Key point

The desired output frequency of a rotary inverter is determined by the DC input voltage.

Modern aircraft equipment is based on the **static inverter**; it is solid state, i.e. it has no moving parts (see Fig. 6.7). The DC power supply is connected to an oscillator; this produces a low-voltage 400Hz output. This output is stepped up to the desired AC output voltage via a transformer.

The static inverter can either be used as the sole source of AC power or to supply specific equipment in the event that the main generator has failed. Alternatively they are used to provide power for passenger use, e.g. lap-top computers. The DC input voltage is applied to an oscillator that produces a sinusoidal output voltage. This output is connected to a transformer that provides the required output voltage. Frequency and voltage controls are usually integrated within the static inverter; it therefore has no external means of adjustment.

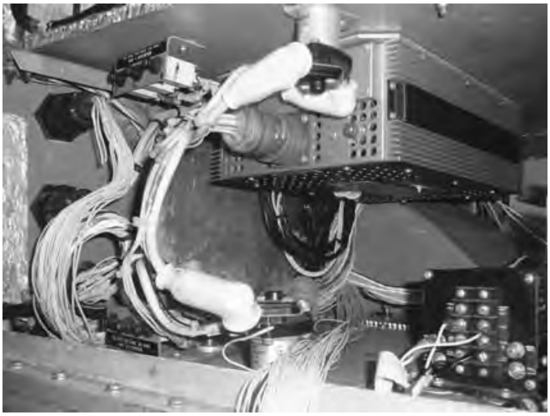

Figure 6.8 Static inverter installation

Key point

Inverters are used to convert direct current into alternating current.

A typical inverter used on a large commercial aircraft can produce 1kVA. Static inverters are located in an electrical equipment bay; a remote on/off switch in the flight compartment is used to isolate the inverter if required. Figure 6.8 shows an inverter installation in a general aviation aircraft.

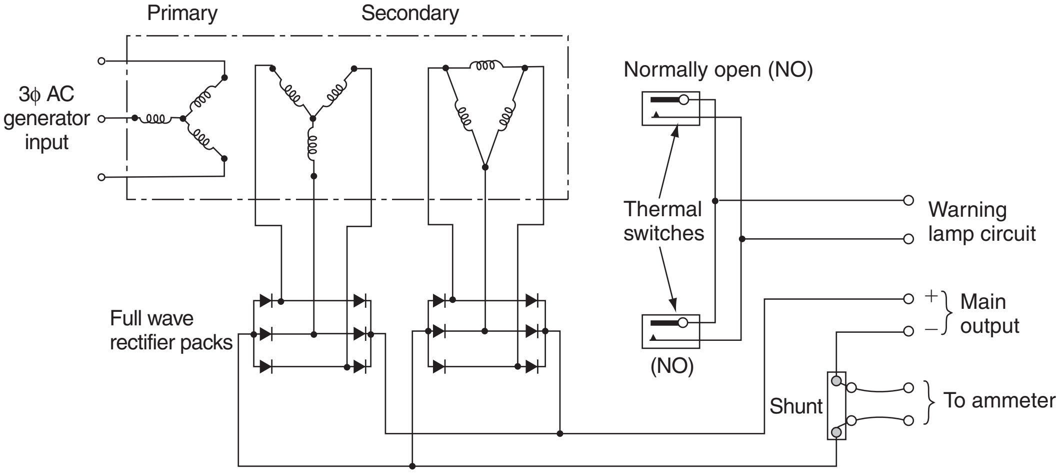

Figure 6.9 Transformer rectifier unit (TRU) schematic

This particular inverter has the following features:

Input:	28 V DC	39 A
Outputs:	115 V AC	6.5 A
	26 V AC	5.8 A
	400 Hz 750 VA continuous	
Power factor:	0.8 to 0.95	
Weight:	15.6 lb	
Dimensions:	270 × 220 × 100 mm	

6.4 Transformer rectifier units

Transformer rectifier units (TRU) convert AC into DC; these are often used to charge batteries from AC generators. A schematic diagram for a TRU is shown in Fig. 6.9. The three-phase 115/200 V 400 Hz input is connected to star-wound primary windings of a transformer. The dual secondary windings are wound in star and delta configuration. Outputs from each of the secondary windings are rectified and connected to the main output terminals. A series (**shunt**) resistor is used to derive the current output of the TRU. Overheat warnings are provided by locating thermal switches at key points within the TRU.

Test your understanding 6.2

Explain the applications of inverters and transformer rectifier units (TRUs).

6.5 Transformers

Transformers are devices that convert (or transfer) electrical energy from one circuit to another through inductively coupled electrical conductors. The transformer used as a power supply source can be considered as having an input (the primary conductors, or windings) and output (the secondary conductors, or windings). A changing current in the primary windings creates a changing magnetic field; this magnetic field induces a changing voltage in the secondary windings. By connecting a load in series with the secondary windings, current flows in the transformer. The output voltage of the transformer (secondary windings) is determined by the input voltage on the primary and ratio of turns on the primary and secondary windings. In practical applications, we convert high voltages into low voltages or vice versa; this conversion is termed step down or step up. (More transformer theory is given in Chapter 1.)

Circuits needing only small step-up/down ratios employ **auto-transformers**. These are formed from single winding, tapped in a specific way to form primary and secondary windings. Referring to Fig. 6.10(a), when an alternating voltage is applied to the primary (P_1–P_2) the magnetic field produces links with all turns on the windings and an EMF is induced in each turn. The output voltage is developed across the secondary turns (S_1–S_2) which can be connected for either step-up or step-down ratios. In practice, auto-transformers are smaller in size and weight than conventional transformers. Their disadvantage is that,

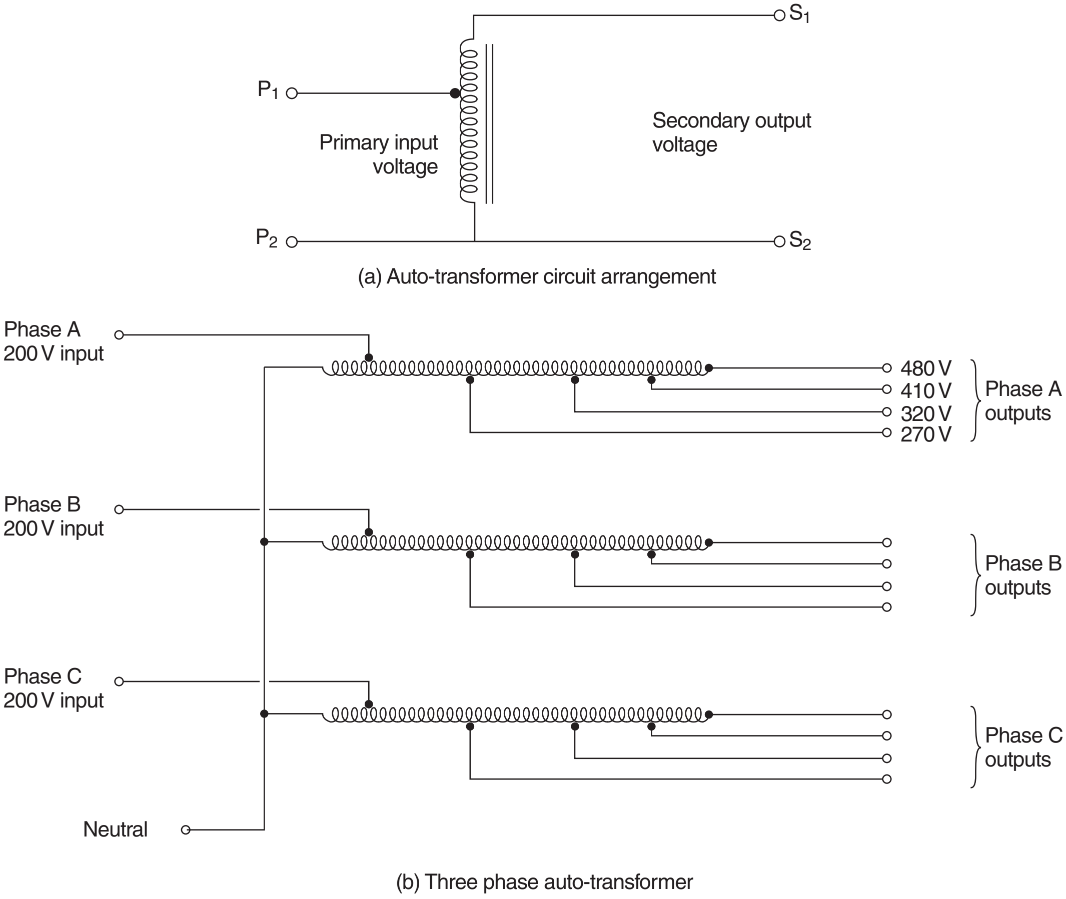
P1
P2
S1
S2
Primary input voltage
Secondary output voltage
(a) Auto-transformer circuit arrangement
Phase A 200 V input
Phase B 200 V input
Phase C 200 V input
Neutral
480 V
410 V
320 V
270 V
Phase A outputs
Phase B outputs
Phase C outputs
(b) Three phase auto-transformer

Figure 6.10 (a) Autotransformer principles; (b) three-phase autotransformer

since the primary and secondary windings are physically connected, a breakdown in insulation places the full primary e.m.f. onto the secondary winding.

The arrangement for a three-phase auto-transformer is shown in Fig. 6.10(b). This is a **star-connected** step-up configuration. Primary input voltage is the 200 V AC from the aircraft alternator; multiple outputs are derived from the secondary tappings: 270, 320, 410 and 480 V AC. Applications for this type of arrangement include **windscreen heating**.

Test your understanding 6.3

Explain the difference between conventional and auto-transformers.

6.6 Auxiliary power unit (APU)

An APU is a relatively small gas turbine engine, typically located in the tail cone of the aircraft. The APU is a two-stage **centrifugal compressor** with a single turbine. Bleed air is tapped from the compressor and connected into the aircraft's air distribution system. Once started (see Chapter 10) the APU runs at constant speed, i.e. there is no throttle control. The APU shuts down automatically in the event of malfunction.

APUs are used for starting the aircraft's main engines via the air distribution system. While the aircraft is on the ground, the APU can also provide:

- electrical power
- hydraulic pressure
- air conditioning.

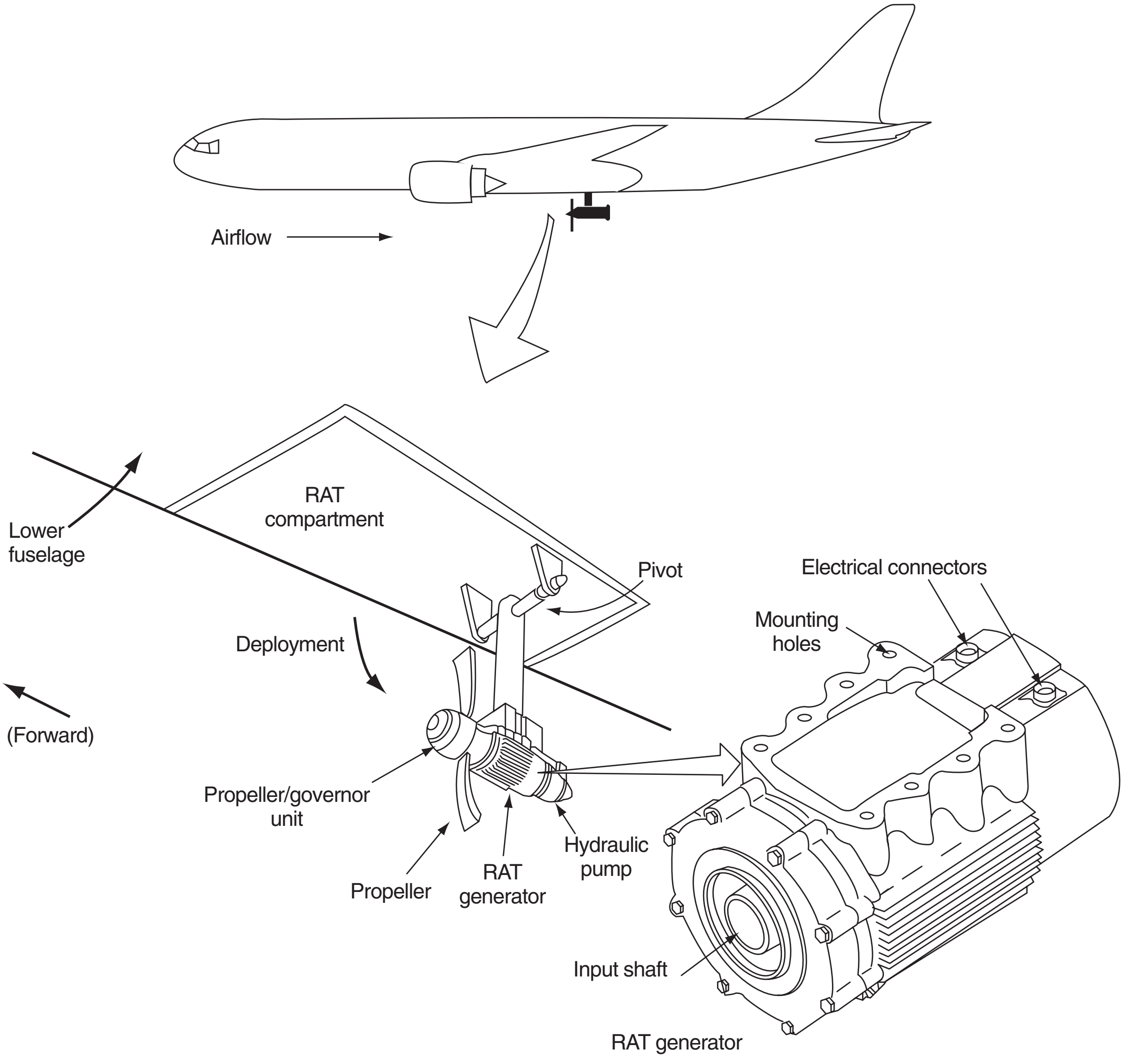
Airflow
Lower fuselage
RAT compartment
Pivot
Electrical connectors
Mounting holes
Deployment
(Forward)
Propeller/governor unit
Propeller
RAT generator
Hydraulic pump
Input shaft
RAT generator

Figure 6.11 Ram air turbine

The APU itself is started from the main aircraft battery. In some aircraft, the APU can also provide electrical power in the air in the event of main generator failure. The Boeing 787 aircraft has more electrical systems and less pneumatic systems than aircraft it is replacing. In this case the APU delivers only electrical power.

APUs fitted to extended-range twin-engine operations aircraft (ETOPS) are critical to the continued safe flight of the aircraft since they supply electrical power, hydraulic pressure and an air supply in the event of a failed main engine generator or engine. Some APUs on larger four-engined aircraft are not certified for use while the aircraft is in flight.

Key maintenance point

It is essential to remove electrical power from the relevant busbar (or in some cases the entire aircraft) before removing or installing electrical components. Failure to observe this precaution can result in electric shock as well as damage to components and equipment.

Key maintenance point

It is essential that electrical power is removed from an external power cable before connecting the cable to the aircraft. Failure to observe this precaution can result in electric shock as well as damage to components and equipment.

6.7 Emergency power

In the event of generator failure, continuous power can be provided by a ram air turbine (RAT). Also referred to as an air-driven generator, this is an emergency source of power that can be called upon when normal power sources are not available. The RAT is an air-driven device that is stowed in the wing or fuselage and deployed in the event that the aircraft loses normal power. When deployed, it derives energy from the airflow, see Fig. 6.11. RATs typically comprises a two-bladed fan, or propeller that drives the generator shaft via a governor unit and gearbox; the gear ratios increase the generator shaft speed.

The RAT can be deployed between aircraft speeds of 120 to 430 knots; some RATs feature variable pitch blades operated by a hydraulic motor to maintain the device at typical speeds of 4,800 r.p.m. Typical RAT generators produces an AC output of 7.5 kVA to a TRU. Heaters are installed in the RAT generator to prevent ice formation. RATs can weigh up to 400 lbs on very large transport aircraft, with blade diameters of between 40 and 60 inches depending on power requirements.

6.8 Multiple choice questions

1. Circuits needing only small step-up/down ratios employ:
   (a) auto-transformers
   (b) transformer rectifier units
   (c) inverters.

2. Inverters are used to convert what forms of electrical power?
   (a) AC into DC
   (b) AC into AC
   (c) DC into AC.

3. The accuracy of the vibrating contact regulator depends on the:
   (a) input voltage
   (b) generator output current
   (c) resistor value and spring tension.

4. The voltage coil of a carbon pile regulator contains a:
   (a) large number of copper wire turns connected across the generator output
   (b) low number of copper wire turns connected across the generator output
   (c) large number of copper wire turns connected in series with the generator output.

5. Transformer rectifier units (TRU) are often used to:
   (a) convert battery power into AC power
   (b) charge batteries from AC generators
   (c) connect batteries in series.

6. Stepping a power supply down from 115 to 26 V AC would normally be achieved by a:
   (a) transformer rectifier unit
   (b) inverter
   (c) transformer.

7. The desired output frequency of a rotary inverter is determined by the:
   (a) AC input voltage
   (b) input frequency
   (c) DC input voltage.

8. Inverters and transformer rectifier units (TRU) are sources of:
   (a) emergency power
   (b) secondary power
   (c) primary power.

9. Higher generator output increases the current in the carbon pile regulator, this:
   (a) reduces compression on the pile thereby increasing its resistance
   (b) increases compression on the pile thereby increasing its resistance
   (c) reduces compression on the pile thereby decreasing its resistance.

10. TRUs are used to convert what forms of electrical power:
   (a) AC into DC
   (b) AC into AC
   (c) DC into AC.

Chapter 7 Wiring and circuit protection

The safe and economic operation of an aircraft is becoming ever more dependent on electrical and electronic systems. These systems are all interconnected with wires and cables, and take many forms. Electrical wires and cables have to be treated as an integral part of the aircraft requiring careful installation; this is followed by direct ongoing inspection and maintenance requirements for continued airworthiness. Wire and cable installations cannot be considered (or treated) as 'fit and forget'. System reliability will be seriously affected by wiring that has not been correctly installed or maintained. Legislation is being proposed to introduce a new term: **electrical wire interconnection system** (EWIS); this will acknowledge the fact that wiring is just one of many components installed on the aircraft. EWIS relates to any wire, wiring device, or combination of these, including termination devices, installed in the aircraft for transmitting electrical energy between two or more termination points.

We need to distribute the sources of electrical power safely and efficiently and control its use on the aircraft. Once installed, the wires and cables must be protected from overload conditions that could lead to overheating, causing the release of toxic fumes, possibly leading to fire. This chapter describes the physical construction of wires and cables together with how they are protected from overload conditions before power is distributed to the various loads on the aircraft. (Practical installation considerations for wiring are addressed in Chapter 20.)

7.1 Overview

In the first instance, we need to make a distinction between **wires** and **cables**. Wires are formed from a single solid conductor or stranded conductors, contained within insulation and protective sheath materials. Cables can be defined as:

- two or more separate wires within the same insulation and protective sheath
- two or more wires twisted together
- any number of wires covered by a metallic braid, or sheath
- a single insulated conductor covered by a metallic outer conductor (co-axial cable).

The terms wires and cables are often interchanged. In this book, reference will be made to 'wiring' in the all-embracing generic sense; cables will be referred to in specific terms as and when required.

Key maintenance point

Wire and cable installations cannot be considered (or treated) as 'fit and forget'. System reliability will be seriously affected by wiring that has not been correctly installed or maintained.

7.1.1 Types of wire and cable

There is a vast range of wires and cable types used in an aircraft; these can be categorized with their specific uses and applications:

- airframe wires and cables
- equipment wires and cables
- ignition system cables
- thermocouple cables
- data bus cables
- radio-frequency (RF) cables.

This chapter addresses airframe wire and cable applications in detail. (Wiring inside of equipment is not specifically addressed.) The application of ignition and thermocouple cables are described in Chapter 10 (engine systems). Wires or cables designed for small signals, e.g. data bus cables carrying digital data, are screened to prevent their signals from being affected by electromagnet interference. Wires and cables that carry high power and/or high frequencies are also shielded to prevent them being the cause of electromagnetic interference. **Screening** and **shielding** to protect against electromagnetic interference (EMI) and electromagnetic compatibility (EMC) are described in this chapter; the subject of EMI/EMC protection is discussed in more detail in Chapter 19. Radio-frequency (RF) cables (or feeders) are specialized types of screened cables. (These are described in more detail in a related book in the series, *Aircraft Communications and Navigation Systems*.)

7.1.2 Operating environment

Wiring installed in aircraft has to operate in a harsher environment than that found in cars, buildings or industrial applications. In addition to conducting current (often at high voltages), aircraft wiring will be exposed to a variety of environmental and in-service conditions including contaminants, for example:

- hydraulic fluid
- fuel and/or oil
- temperature extremes
- abrasion
- vibration

General-purpose wiring used in the airframe comprises the majority of the installed material. There are also special considerations for wiring, e.g. when routed through fire zones. The materials used for the conductor and insulation must take these factors into account. Each application has to be planned from the initial design, taking into account the inspection and maintenance required for continued airworthiness of the wiring.

Test your understanding 7.1

Why are cables shielded?

7.2 Construction and materials

Aircraft wiring needs to be physically flexible to allow it to be installed, and then to withstand the vibration of the aircraft that will cause the wires to flex. **Multi-stranding** of the conductor increases the flexibility of the wire or cable, making it easier to install and withstand vibration of the aircraft. The insulating material has to be able to withstand the applied voltage; the sheath material needs to be able to withstand the specified contaminants. Conductors need to be able to carry the required current without overheating or burning; they must also have low insulation resistance to minimize voltage drops. From **Ohm's law**, we know that (at a given temperature) the voltage across a resistance is proportional to the current. For a given current (I) and resistance (R), the voltage drop is quoted in terms of ***IR* losses**. For these reasons, most aircraft conductors are constructed from copper or aluminium contained within man-made insulating material(s).

Aluminium conductors are sometimes used in aircraft; however, the majority of installations are copper. The choice of conductor material is a trade-off; the first consideration is the material's resistance over a given length (given the term **resistivity**, symbol ρ, measured in ohm-metres, abbreviated Ωm). Annealed copper at 20°C has a resistivity of $1.725 \times 10^{-8}\ \Omega$m; copper is more ductile than aluminium, and can be easily soldered. Aluminium has a resistivity value of $2.8 \times 10^{-8}\ \Omega$m, it is 60% lighter than copper but it is more expensive.

A major consideration for using aluminium is that it is **self-oxidizing**; this reduces manufacturing costs (no plating required) but extra precautions are necessary for terminating the conductors due to the increased termination resistance. The quantity and gauge of these strands depends on the current-carrying capacity (or rating) and degree of flexibility required. Individual strands of copper need to be coated to prevent oxidation. The choice of coating for the strands depends on the operating temperature of the wire. In general terms, three types of coating are used: tin, silver or nickel, giving temperature ratings of 135°C, 200°C and 260°C respectively.

To summarize, copper conductors are used extensively on aircraft due to the material's:

- low resistivity
- high ductility
- high tensile strength
- ease of soldering

Test your understanding 7.2

Explain why cables and wires are multi-stranded.

Conductors must be insulated to prevent short-circuits between adjacent circuits and the airframe. Power supplies of 12 or 28V do not pose a threat of electrical shock, but the wires must be insulated to prevent arcing, loss of system integrity and equipment failure. The combined effects of insulation damage and fluid contamination gives rise to **wet arc tracking**. This phenomenon can occur when insulating surfaces are contaminated with any material containing free ions; the surface then behaves as an electrically conductive medium (an **electrolyte**). Leakage currents are sufficiently high to vaporize the contamination; this drives away the electrolyte and results in the formation of localized dry areas. These areas now offer a higher resistance to the current flow. In turn, high voltages will develop across these areas and result in small surface discharges. Initially, these discharges will emit flashes of light at the insulation surface, and produce localized temperatures in the order of 1000ºC. These high temperatures cause degradation of the insulation material. The ability of aircraft wiring to resist wet arc tracking is highly dependent on the wire insulation material. The conductivity level of the electrolyte will influence the **failure mode** resulting from this wet arc tracking. Higher power supply voltages are potentially lethal and the insulation provides a level of protection against this. The insulation material and its thickness depend mainly on the operating temperature and system voltage; examples of insulating and sheath materials include:

- ethylene tetrafluoroethylene (ETFE): this is a fluorocarbon-based polymer (fluoropolymer) in the form of plastic material
- polytetrafluoroethylene (PTFE), a synthetic fluoropolymer
- fluorinated ethylene-propylene (FEP): this retains the properties of PTFE, but is easier to form
- polyvinylidene fluoride (PVF_2 or PVFD) has good abrasion and chemical resistance, and (like most fluoropolymers) is inherently flame-retardant.

Test your understanding 7.3

Explain the term wet arc tracking.

7.3 Specifications

These have become more complex over the years to address the higher performance needed from wires and cables. This has been driven largely from in-service experience with electrical fires and the drive to reduce weight as more and more avionics equipment is introduced onto the aircraft. Reduced weight for a given length and diameter of wire is achieved through reducing the wall thickness of the insulation. Typical specifications used for wires and cables are contained in the US military specification **MIL-W-M22759E**. This specification covers fluoropolymer-insulated single conductor electrical wires manufactured with copper or copper alloy conductors coated with either tin, silver or nickel. The fluoropolymer insulation of these wires can either be PTFE, PVF_2, FEP or ETFE. Wires manufactured to this specification are given a part number using the following format: M22759/x-xx-x.

- *M22759/x*-xx-x determines the specific wire type (Insulation, sheath and wire coating) from a table in the specification
- M22759/x-*xx*-x determines the wire size from a table in the specification
- M22759/x-xx-*x* determines the insulation colour from another specification (MIL-STD-681).

Examples of wires are shown in Fig. 7.1. The single walled construction in Fig. 7.1(a) has a composite insulator and protective sheath to reduce cost and weight. The twin-walled construction, Fig. 7.1(b), illustrates the conductor, insulator and separate protective sheath. Outer sheaths on either type of wire can crack over long periods of time; in the single-wall-type wire, moisture ingress can migrate into the wire through capillary action. This can lead to tracking inside the insulation, leading to overheating; twin walled wiring is more resistant to this effect.

Key maintenance point

The combined effects of insulation damage and fluid contamination gives rise to wet arc tracking.

7.3.1 Wire size

Wire sizes used on aircraft are defined in accordance with the **American Wire Gauge** or **Gage** (AWG). For a given AWG, the wire will have a specified diameter and hence a known conductance (the reciprocal of resistance). Cable size relates to the conductor's diameter; the overall wire or cable diameter is therefore larger due to the insulation. The largest wire size is 0000 AWG; the smallest is 40 AWG. The range of AWG wire/cable sizes is detailed in the Appendices. Selection of wire size depends on the specified current to be conducted. Larger diameter wires add

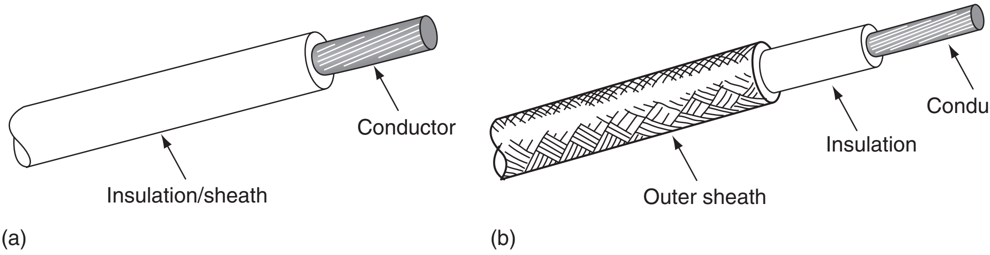

Figure 7.1 Typical aircraft wires and cables: (a) single-walled, (b) twin-walled

weight, but offer less voltage drop for a given length and lower heating effect due to I^2R losses.

Wire identification is printed on the outer surface of the sheath at intervals between 6 and 60 inches. This includes the wire part number and manufacturers' commercial and government entity (**CAGE**) designation. The printing is either green or white (depending on the actual colour of the wire).

Key point

Conductors must have low insulation resistance to minimize IR losses.

7.3.2 Performance requirements

Wires manufactured to MIL-W-M22759E have to demonstrate compliance with dimensions and construction, together with criteria to meet the following requirements:

- ease of removing the insulation
- ease of soldering
- dielectric testing
- flexibility
- elongation and tensile strength
- wicking
- high- and low-temperature testing
- flammability
- life cycle testing
- fluid immersion
- humidity
- smoke emission.

7.4 Shielding/screening

Shielded or screened wiring either prevents **radiation** from circuits switching high currents or protects **susceptible** circuits (see Fig. 7.2). The inner conductor carries the system current; the screen provides a low resistance path for coupling of electromagnetic fields. These fields are coupled into the shield and are dissipated to ground. (Further details on the theory of electromagnetic fields and shielding are given in Chapter 19.) Typical applications where shielding is used includes wiring installed near generators, ignition systems or contacts that are switching high currents. Shielded wires can be formed with single, twin, triple or quadruple cores. The classic example of screened wiring occurs with the Arinc 429 data bus which uses twisted screened pairs of cable to transmit digital data. This is being transmitted at either 12.5–14.5 (low speed) or 100 (high speed) k bits per second with a differential voltage of 10V between the pair of wires. Examples of screened wire and cable specifications are found in M27500-22TG1T-14.

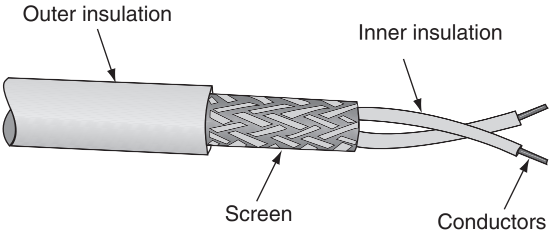

Figure 7.2 Screened cable

Key point

Wires and cables that carry high power, and/or high frequencies are shielded to prevent electromagnetic interference.

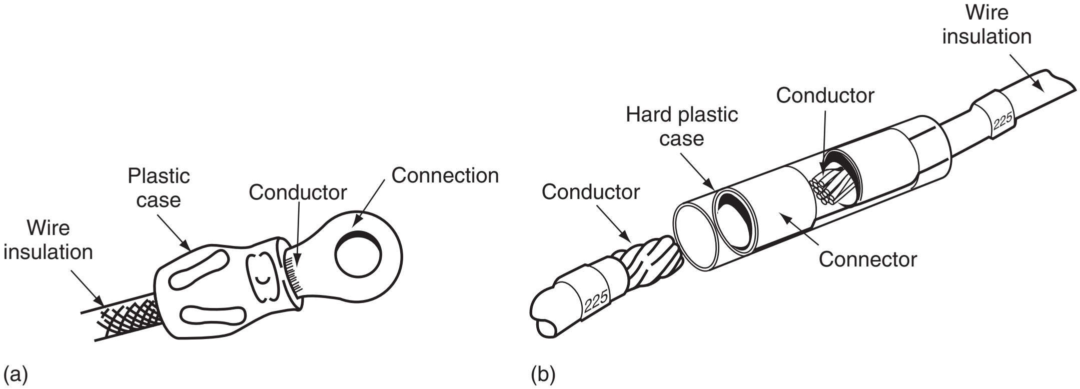

Figure 7.3 Wire termination and splicing: (a) ring tongue terminal, (b) in-line splice crimping

7.4.1 Crimps and splices

Individual wires and cables can be terminated or connected using crimps and splices. Care must be taken when stripping shielded cables; both the inner conductor and outer screen must be exposed in order to make the connection. The outer shield is formed into a pigtails and terminated with a **crimp**, or ring-tongue terminal, see Fig. 7.3(a). Alternatively, individual wires can be joined with an in-line **splice**, e.g. if a system is being modified with additional wiring, see Fig. 7.3(b). Crimps and splices are formed over the exposed conductor and insulating material. The entire crimp or splice termination is then protected mechanically with a hard plastic case.

Test your understanding 7.4

Give some examples of wire/cable performance specifications.

7.4.2 Coaxial cables

A specialized version of the shielded wire is the coaxial cable. The inner conductor is solid or stranded; it can be plain copper or plated. The outer conductor forms a shield and is a single wire **braid** made from fine strands of copper or steel. The inner and outer conductors are separated by a solid insulation, forming a **dielectric**. The outer sheath or jacket provides protection against fluid contaminants. Coaxial cables are normally used to guide radio-frequency (RF) energy between antennas and receivers or transmitters. The inner conductor is shielded by the braiding from electric and magnetic fields; the conductor's own field is contained within the same shield. The net result is that fields from the inner and outer conductors cancel each other out. In most practical RF applications, coaxial cable radiation and susceptibility are virtually eliminated.

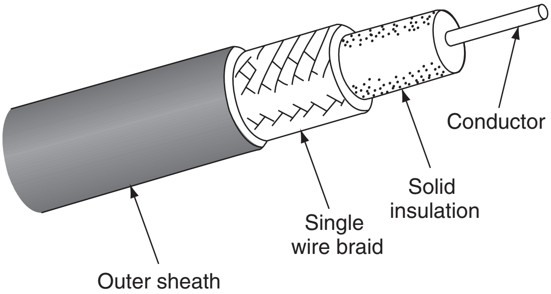

Figure 7.4 Coaxial cable overview

The typical construction of coaxial cable is shown in Fig. 7.4; the salient features are:

- **inner conductor:** silver plated copper
- **solid insulation:** (dielectric) foamed FEP
- **single wire braid:** (screen) tin plated copper
- **outer insulation:** (jacket) FEP.

(Further details on coaxial cables and antennas used in radio-frequency (RF) applications can be found in *Aircraft Communications and Navigation Systems*.) Other applications for coaxial cables include fuel quantity systems (see Chapter 11).

Coaxial cables are terminated with a **bayonet** or **thread-nut couplings** (BNCs and TNCs respectively). BNC has the advantage of a quick-release connection. TNC performs well under vibration conditions since

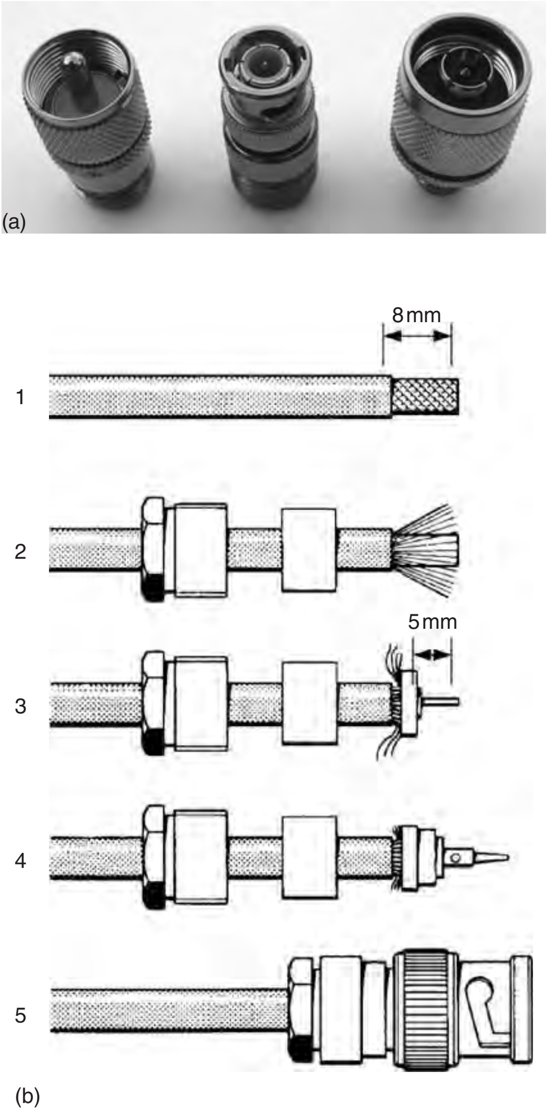

Figure 7.5 Coaxial cable connectors: (a) typical end fittings, (b) terminating the coax cable

they have reduced axial movement. Figure 7.5 gives an illustration of typical coaxial cable connectors and how they are fitted to the cable.

Test your understanding 7.5

What is the difference between shielded and coaxial cables?

7.5 Circuit protection

The current-carrying capacity of a wire or cable is determined by its length and cross sectional area; heat dissipation is determined by I^2R losses. When the circuit or system is designed, the wire size is selected to safely carry this current. Wires and cable are subjected to abrasion during the normal service life of the aircraft; this can lead to the conductor being exposed. This exposure could lead to a low resistance path between the conductor and the airframe and/or an adjacent conductor. Faulty equipment, low resistance paths or overloading from additional circuits will cause the current to increase and this might exceed the current-carrying limit of the conductor. Heat will build up in the wire leading to fumes, smoke and ultimately fire. It is vital that we protect against this whilst allowing for transients; the methods used in aircraft are selected from the following devices:

- fuse
- circuit-breaker
- limiting resistor.

7.5.1 Fuses

Fuses are links of wire that are connected in series with the circuit. Their current-carrying capacity is predetermined and they will heat up and melt when this is exceeded, thereby interrupting and isolating the circuit. Materials used for the **fusible link** include lead, tin-bismuth alloy, copper or silver alloys. Referring to Fig. 7.6(a), the fuse wire is contained with a glass or ceramic casing (or cartridge) to prevent any particles of hot metal escaping which could cause secondary damage. End-caps provide a connection for the fuse wire and make contact with the circuit wiring. Fuse holders consist of terminals and a panel clamp-nut.

Some fuse holders have an indication of the fuse condition, i.e. if the fuse has blown. The **indicating cap** is black with an integrated coloured light. When the fuse has blown, the cap illuminates; different colours indicate different power supply voltages. Heavy-duty fuses (typically protecting circuits with up to 50 A current) are constructed with a ceramic body and terminals, see Fig. 7.6(c). Fuses are either clipped into position on a terminal board, see Fig. 7.7, or screwed into a panel, see Fig. 7.8.

Fuses are relatively low cost items, but they can only be used once. In some applications, the fuse

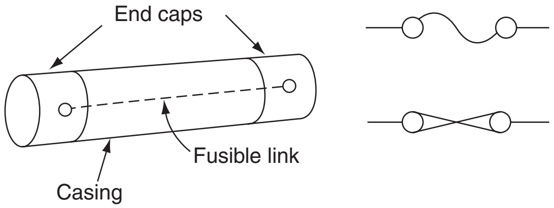
End caps
Fusible link
Casing

(a) Cartridge fuse & circuit symbols

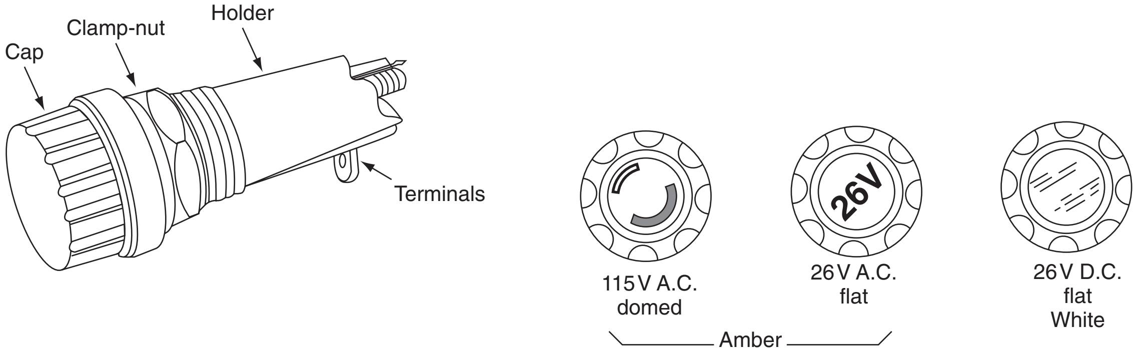
Cap
Clamp-nut
Holder
Terminals
26V
115 V A.C.
domed
26 V A.C.
flat
26 V D.C.
flat
White
Amber

(b) Fuse holders (left) and indicating caps (right)

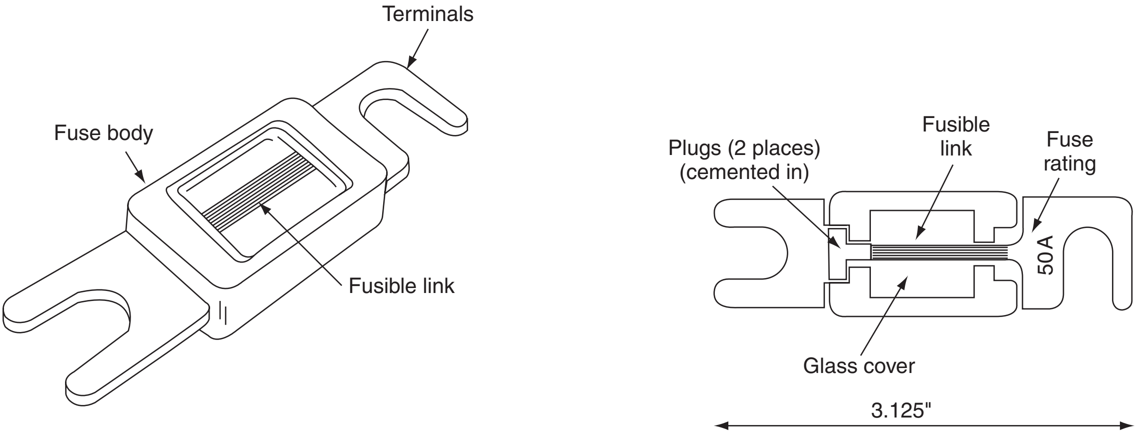
Terminals
Fuse body
Fusible link
Plugs (2 places)
(cemented in)
Fusible
link
Fuse
rating
50 A
Glass cover
3.125"

(c) Heavy duty fuse construction

Figure 7.6 Aircraft fuses

material and physical construction is designed to have a time delay; the so-called **slow-blow** fuse, or **current limiter**. This is made from a copper alloy that has a higher melting point than lead/tin. It has a single strip of material waisted into a narrow cross-section to provide the fusing point. Heavy-duty fuses are used at power distribution points. They have multi-strands of parallel elements and are rated up to 500 A. This type of fuse is fitted with a packing medium to contain the debris following rupture. Materials used include quartz, magnesium oxide, kieselguhr or calcium carbonate (chalk).

Fuses have a rating that determines the maximum current it can carry without melting. The fuse will also have a minimum fusing current that is affected by ageing; a process that occurs when the fuse is

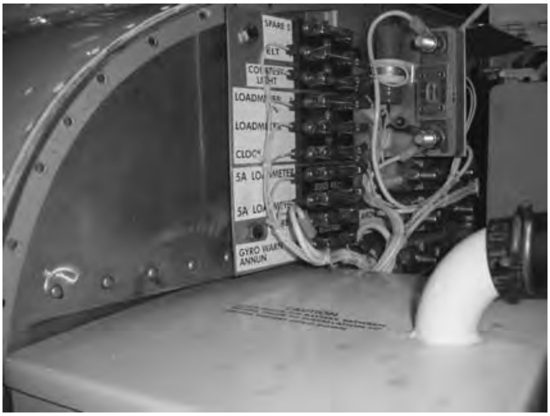

Figure 7.7 Fuse clipped directly into a terminal board

Figure 7.8 Fuses located in a panel

operated at the minimum fuse rating for prolonged periods of time. Ambient temperature affects the current rating and response time of a fuse. They must be located close to power source to minimize the length of unprotected wire; at the same time they have to be accessible for replacement. Spare fuses must be carried on the aircraft and be accessible to the flight crew. Typical requirements are to carry 50% of each rating as spares, e.g. if the aircraft is fitted with four 10 A and five 15 A fuses, then two 10 A and three 15 A fuses should be carried as spares.

Key maintenance points

1. Fuses can rupture, or 'blow', when exposed to vibration and thermal cycling as well as during overload conditions.
2. Determine the reason that caused the fuse to blow before replacing.
3. Never assume that the blown fuse is to be replaced with the same rating; always check the aircraft documents.

7.5.2 Circuit-breakers

Circuit-breakers are electromechanical devices that interrupt and isolate a circuit in the event of excessive current. Unlike fuses, circuit-breakers can be reset (assuming that the fault condition has cleared). There are two circuit-breaker principles: electromagnetic and thermal.

An **electromagnetic** circuit-breaker is essentially a relay with current flowing through a coil; the resulting magnetic field attracts an armature mechanism. The current is normally a proportion of the main load current; this increases in proportion to the main load current. The armature mechanism is linked to a pair of contacts that carry the main load current. These contacts are opened when the current through the coil exceeds a certain limit.

Thermal-type circuit-breakers consist of a bimetallic thermal element, switch contacts and mechanical latch. The internal schematic of a thermal circuit-breaker and its external features are illustrated in Fig. 7.9. The thermal element is a **bimetallic** spring that heats up as current passes through it; this eventually distorts and trips the mechanism when the rated level is exceeded. The mechanism is linked to the main switch contacts, when the circuit-breaker '**trips**' the contacts open, thereby disconnecting power from the circuit. When the contacts open, a **button** is pushed out of the circuit-breaker. This button is used to manually reset the contacts; a **white collar** just below the button provides visual indication that the circuit-breaker is closed or tripped.

Some circuit-breakers use a large collar grip to identify specific systems. They can be locked open if required, e.g. if the system is installed but not certified for operation. As with fuses, the circuit-breaker should be located as close as possible to source of power; they are often arranged on the panels in groups.

Circuit-breakers can also be used to conveniently isolate circuits, e.g. during maintenance. Certain circuit-breakers are fitted with removable collars so that they can be readily identified. The circuit-breaker current rating is engraved on the end of the button (see Fig. 7.10). Circuit-breakers can be single- or multi-pole devices (poles are defined as the number of links

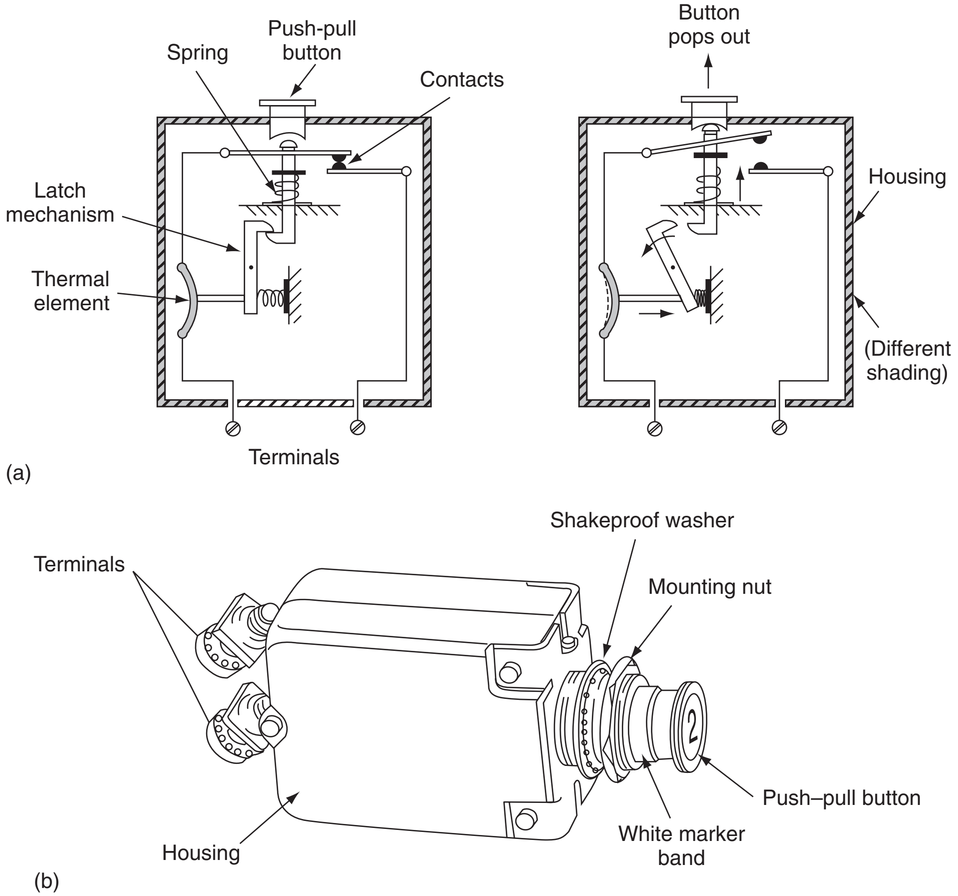
Spring
Push-pull button
Contacts
Button pops out
Latch mechanism
Thermal element
Housing
(Different shading)
Terminals
(a)
Shakeproof washer
Terminals
Mounting nut
2
Push–pull button
White marker band
Housing
(b)

Figure 7.9 Aircraft circuit-breakers (thermal type): (a) internal schematic (closed and tripped), (b) external features

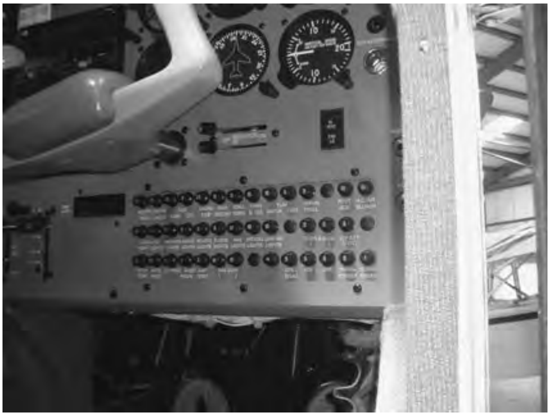

Figure 7.10 Circuit-breaker panel

that a switching device contains). Multi-pole devices are used in three-phase AC circuits.

Various configurations of circuit-breaker are installed on aircraft, including:

1. automatic reset
2. automatic trip/push to reset
3. switch types
4. trip free.

The circuit symbols for each of these types is illustrated in Fig. 7.11. **Trip-free** circuit-breaker contacts cannot be closed whilst a fault exists. This is the preferred type of circuit-breaker on aircraft, especially on new installations.

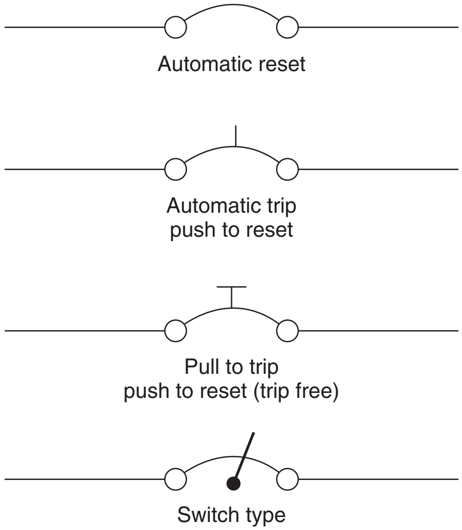

Figure 7.11 Circuit-breaker symbols/type

7.5.3 Limiting resistors

These are used to limit current surges, primarily in DC circuits, where the initial current surge is large. When these circuits are switched they create large current flows that can be harmful to other components and reduce the power supply voltage for a period of time (determined by the time constants of the circuit). Limiting resistors are connected in series with such circuits and then automatically shorted out once the circuit current has stabilized. Typical applications of limiting resistors are found in engine starting circuits and voltage regulators. Limiting resistors are also used in fire extinguishing systems (see also Chapter 16). Fire extinguishers are activated by applying direct current (DC) through current-limiting resistors to the associated squib; this ruptures a disc that allows extinguishing agent to be expelled under pressure. The limiting resistors prevent inadvertent operation of the squib. In electronics, limiting resistors are used to protect devices such as diodes.

7.6 Multiple choice questions

1. Wires and cables that carry high power and/or high frequencies are shielded to prevent:
   (a) the wire or cable being the cause of electromagnetic interference
   (b) the wire or cable being affected by electromagnetic interference
   (c) wet arc tracking.

2. Two or more separate wires within the same insulation and protective sheath is referred to as a:
   (a) screened wire
   (b) coaxial cable
   (c) cable.

3. Fuses have a rating that determines the:
   (a) maximum current it can carry without opening
   (b) minimum current it can carry without opening
   (c) time taken for the fuse to rupture.

4. Conductors must have low insulation resistance to minimize:
   (a) current surges
   (b) electromagnetic interference
   (c) IR losses.

5. The combined effects of insulation damage and fluid contamination gives rise to:
   (a) electromagnetic interference
   (b) wet arc tracking
   (c) current surges

6. For a given AWG, the wire will have a specified:
   (a) diameter and hence a known conductance
   (b) length
   (c) screening.

7. Coaxial cables are normally used for:
   (a) digital signals
   (b) motors and generators
   (c) radio-frequency (RF) signals.

8. Trip-free circuit-breaker contacts:
   (a) can always be closed whilst a fault exists
   (b) cannot be closed whilst a fault exists
   (c) are only used during maintenance.

9. Limiting resistors are connected:
   (a) momentarily in series
   (b) momentarily in parallel
   (c) permanently in series.

10. The white collar just below a circuit-breaker button provides visual indication of the:
   (a) circuit-breaker trip current
   (b) system being protected
   (c) circuit-breaker being closed or tripped.

Chapter 8 Distribution of power supplies

Electrical power is supplied to the various loads in the aircraft via common points called busbars. The electrical power distribution system is based on one or more busbar(s); these provide pre-determined routes to circuits and components throughout the aircraft. The nature and complexity of the distribution system depends on the size and role of the aircraft, ranging from single-engine general aviation through to multi-engine passenger transport aircraft. For any aircraft type, the distribution system will comprise the following items:

- busbar
- protection
- control
- wiring
- loads.

In this chapter, we will focus on busbar configurations and how these are arranged for the protection and management of the various power supply sources available on the aircraft.

The word 'bus' (as used in electrical systems) is derived from the Latin word *omnibus* meaning 'for all'. Busbars are often formed from thick strips of copper, hence the name 'bar'. These have holes at appropriative intervals for attaching one side of the protection device (circuit-breaker or fuse). Alternatively, they are made from heavy gauge wire. The busbar can be supplied from one or more of the power sources previously described (generator, inverter, transformer rectifier unit or battery). Protection devices, whether fuses or circuit-breakers, are connected in series with a specific system; they will remove the power from that system if an overload condition arises. There also needs to be a means of protecting the power source and feeder lines to the busbar, i.e. before the individual circuit protection devices. The various methods used for aircraft power supply distribution for a wide range of aircraft types are described in this chapter.

8.1 Single engine/general aviation aircraft

A simple busbar distribution system is fed by the battery and controlled by a **battery master switch**, see Fig. 8.1. Since the current supplied from the battery

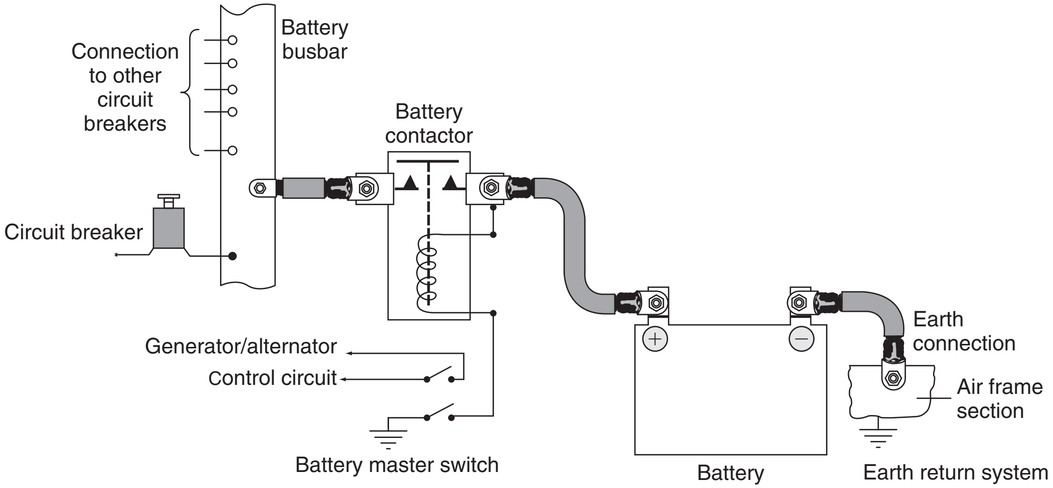

Figure 8.1 Battery busbar schematic

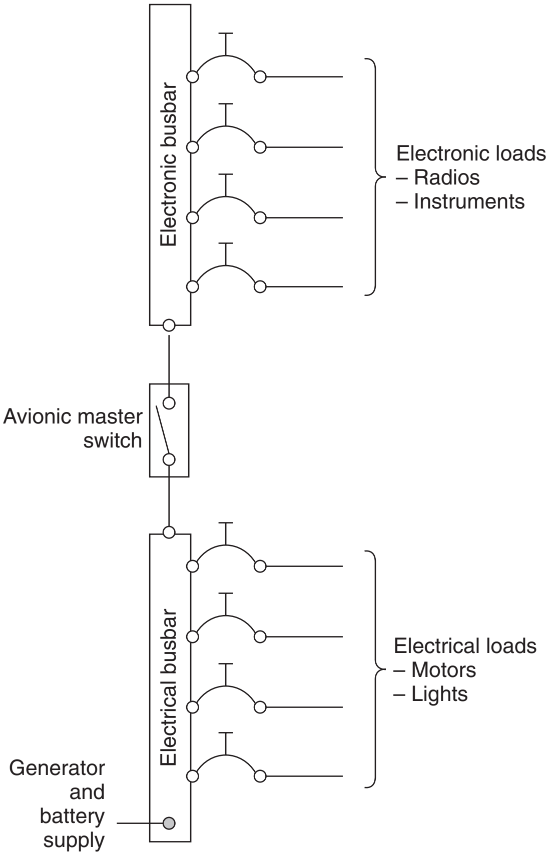

Figure 8.2 Split avionic bus schematic

to the bus will be very high, a heavy-duty relay or contactor is often used; the battery master switch activates the contactor when power is required on the bus. The same switch is also used for generator control.

For general aviation aircraft the battery is normally used for an engine start. The connection for this is taken before the bus to minimize voltage drop (IR) losses. A variation of the simple battery bus is the **split avionic bus**, see Fig. 8.2. This arrangement isolates electronic devices from heavy surges during engine start, or when using external power to start the engine. The arrangement for an AC generator system is shown schematically in Fig. 8.3. This illustrates the voltage regulator and battery master switch used for alternator control. It is normal practice for the battery charge/discharge current to be monitored via an ammeter.

For certification purposes, it is often a requirement that no single failure of any component should cause loss of power to the radios. One way of achieving this is supplying the bus from the main contactor and an emergency switched contact.

8.1.1 Reverse current relay

This circuit arrangement is needed on any DC generation system to prevent the battery from feeding excess current back through to the generator's armature. Without this protection, the battery would discharge when the generator output voltage is less than battery terminal voltage, e.g. during engine shutdown. An automatic method is needed to disconnect the generator from the battery under these conditions. Many circuits have been developed, some are standalone, some are integrated with other generator functions within the same unit e.g. voltage regulation. To illustrate the principles of the reverse current relay, an electromechanical device is described.

Referring to Fig. 8.4, voltage and current coils are wound onto the same soft-iron core. The voltage coil comprises many turns of fine gauge wire; this is connected in parallel with the generator output. The current coil is made of a few turns of thick gauge wire; it is connected in series with the generator output. The combined coils and core form an electromagnet.

A pair of high-current-rated switch contacts are held open by the force of a spring; the contacts can be closed by an armature that is controlled by the influence of the electromagnet's field.

When the generator output is higher than the terminal voltage of the battery, the contacts are held closed by the magnetic field created by the current in both the coils. (These are wound such that the magnetic fields assist each other.) If the engine is slowed down to the point where the generator output is less than the terminal voltage of the battery, the contacts are opened since current is now flowing through the current coil in the opposite direction (the electromagnetic field is weakened as a result). With the contacts open, the battery is effectively disconnected from the generator.

8.1.2 Current limiter

This device reduces the output from a DC generator under fault conditions, e.g. a short circuit on the busbar, to protect the generator from overheating and burning out its windings. The limiter reduces generator output if its maximum safe load is exceeded; if the fault condition is removed, the generator automatically becomes

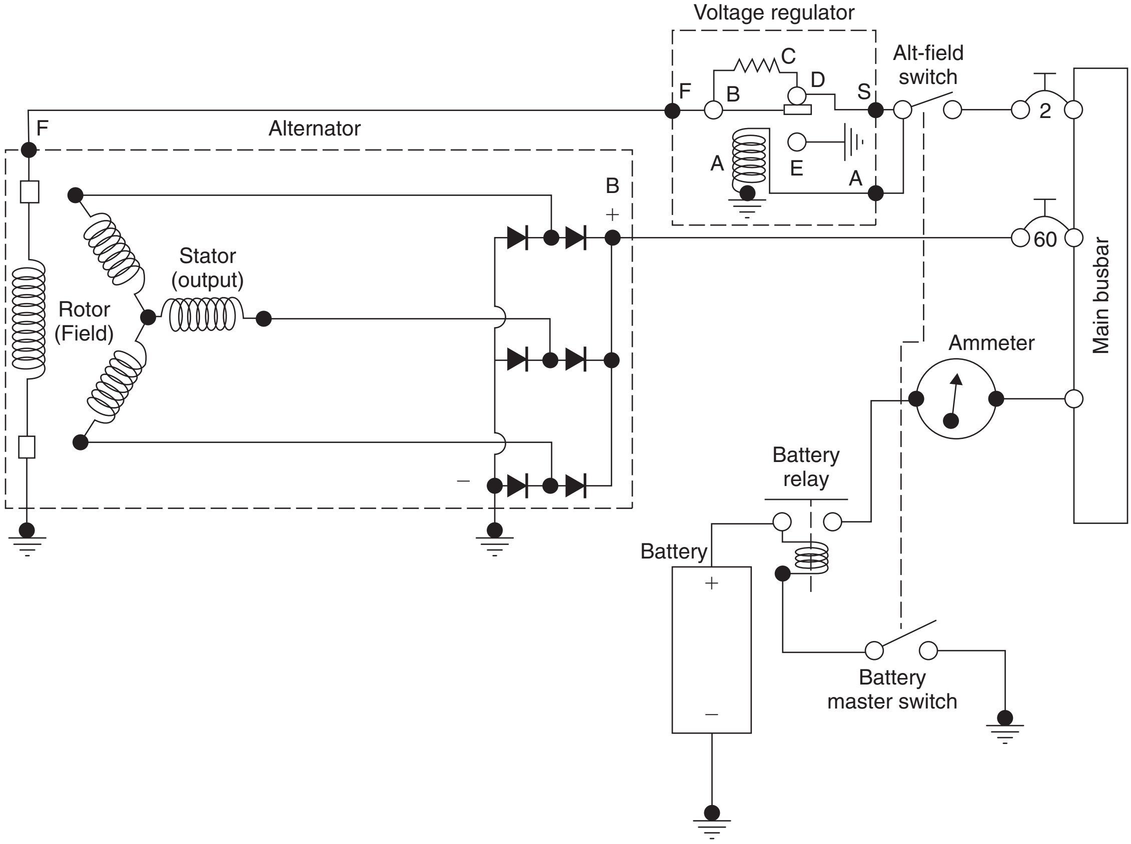
Voltage regulator
C
D
F
B
S
Alt-field switch
2
F
Alternator
A
E
A
B
+
Stator (output)
Rotor (Field)
60
Main busbar
Ammeter
−
Battery relay
Battery
+
−
Battery master switch

Figure 8.3 Alternator/busbar schematic

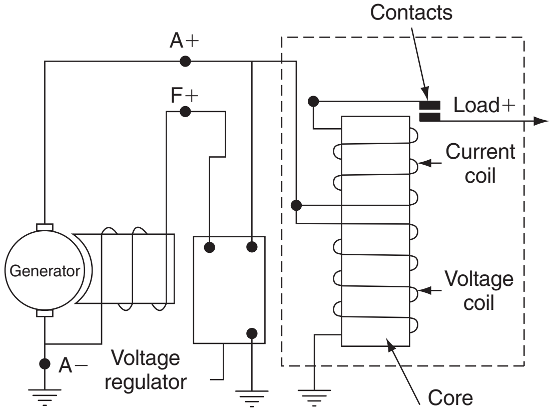
Contacts
A+
F+
Load+
Current coil
Generator
Voltage coil
A−
Voltage regulator
Core

Figure 8.4 Reverse current relay

available again. It operates on a similar principle to the electromechanical voltage regulator; as with the reverse current relay, it could be a stand-alone device or be built into a generator control unit. To illustrate the principles of the current limiter, an electromechanical device is described.

Referring to Fig. 8.5, the current coil is connected in series with the output of the generator. If the output exceeds a pre-determined limit, the coil's magnetic field opens a pair of switch contacts; this puts a resistor in series with the generator's field thereby reducing the generator's output. This decrease in generator output weakens the field of the current coil and the contacts close again (shorting out the resistor). If the fault condition has been removed, the generator remains connected to the bus. Should a permanent overload condition exist, the contacts open/close on a repetitive basis to keep the output within safe limits.

Test your understanding 8.1

Explain the different functions of a reverse current relay and current limiter.

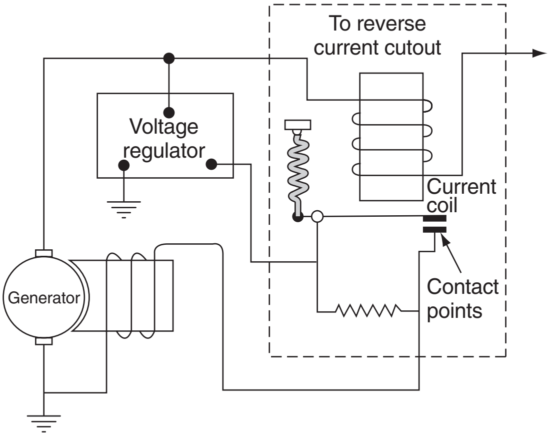
To reverse current cutout
Voltage regulator
Current coil
Contact points
Generator

Figure 8.5 Current limiter

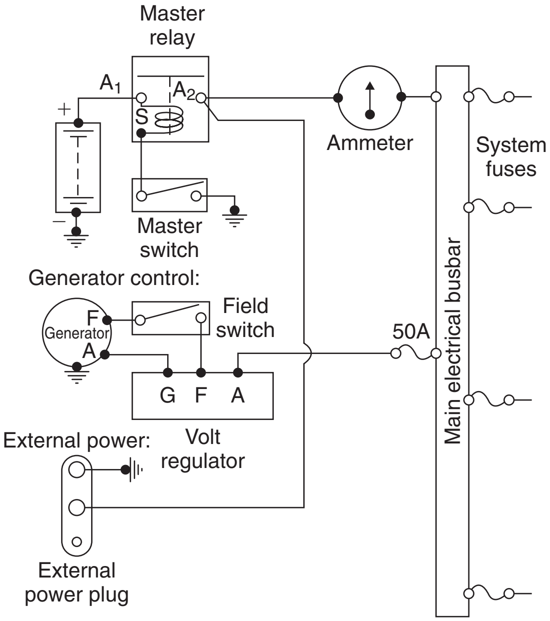
Master relay
A1
A2
S
+
−
Ammeter
System fuses
Master switch
Generator control:
F
Generator
A
Field switch
50A
Main electrical busbar
G F A
Volt regulator
External power:
External power plug

Figure 8.6 External power supply schematic

8.1.3 External power

The aircraft battery provides an autonomous means of starting the engine. Certain types of operation, e.g. cold weather and repeated starts, could lead to excessive demands, resulting in a battery that is not fully charged. An external power supply system schematic (as illustrated in Fig. 8.6) provides power to the aircraft, even when the battery is flat, or not installed.

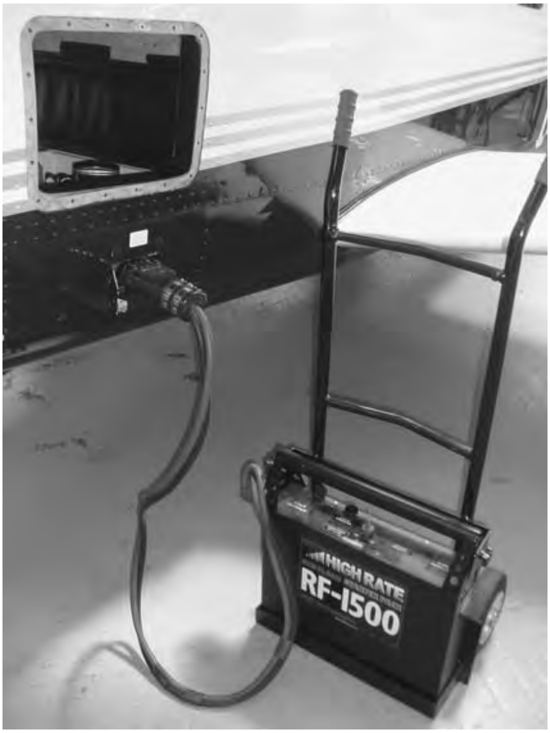
HIGH RATE
RF-1500

Figure 8.7 External ground power from a battery pack

External power can be from a ground power unit or simply from a battery pack as shown in Fig. 8.7. On larger aircraft installations, a connector with three sockets supplies external power. These sockets connect with three pins on the aircraft fuselage as shown in Fig. 8.8. Different-size pins are used on the connectors to prevent a reverse polarity voltage being applied.

Some aircraft installations have a **ground power relay**. Power can only be supplied into the aircraft via the main pins when the third (shorter) pin makes contact. The third pin is used to energize the ground power relay; this additional relay prevents arcing on the power connector as illustrated in Fig. 8.9. External AC power is applied in a similar way, except that the three phases have to be connected via individual circuits.

8.1.4 Battery charging: single engine

Battery charging on the ground is achieved by an external power supply (it would be unsafe to run the

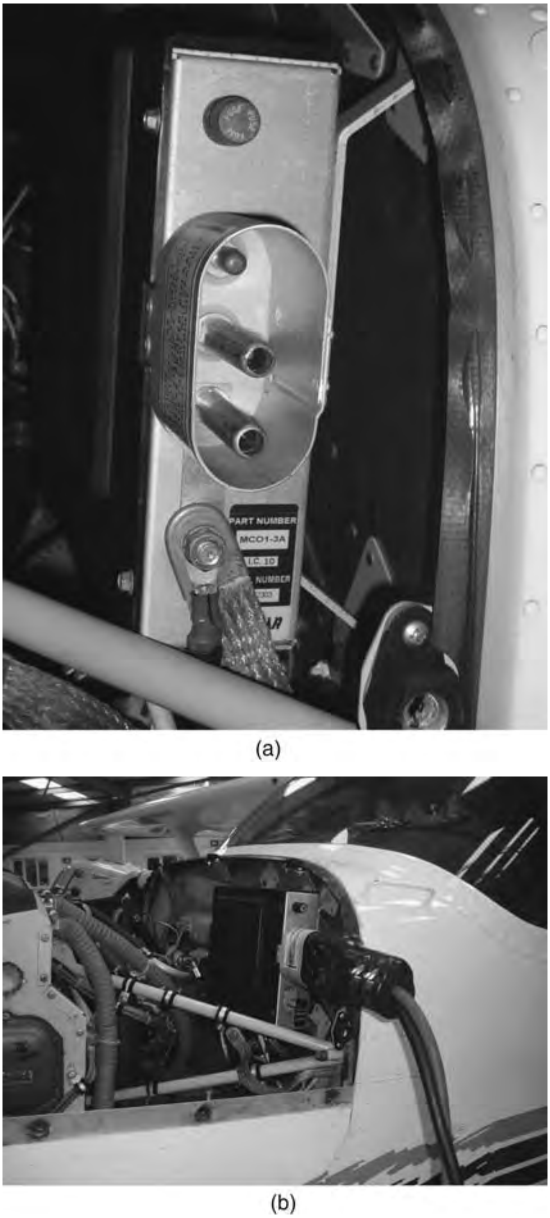

Figure 8.8 External ground power: (a) external ground power 3-pin connector, (b) external ground power connected

engine in the hangar). External power in the form of another battery, a diesel-powered generator or adapted mains supply provides a source of energy. With external power available and switched on, power is now available at the battery relay output terminal via the external power relay. Power is also supplied to the closing circuit (the resistor and diode reduces the charging voltage). When the battery master switch is selected on, this closes the battery relay thus providing full charging current. Protection against a short circuit battery is from a fuse.

Test your understanding 8.2

What kind of instrument would be used to measure battery charging/discharging current?

8.2 Twin engine general aviation aircraft

In the basic configuration, each engine drives its own generator, and the outputs are supplied to a common battery busbar. Engine speed variations mean that one generator could be supplying more output than the other generator and so an equalizing system, using carbon pile regulators, is employed to balance the outputs. This balancing circuit provides automatic adjustment so that each generator delivers an equal output.

Referring to Fig. 8.10, the left and right carbon pile regulators (items 3 and 4) each have equalizing coils (items 7 and 8) wound on the same core as the voltage regulator coils (items 5 and 6). Low value resistors (items 1 and 2, typically 0.01 Ω) are fitted in each generator ground connection; these develop voltage drops in proportion to the generator output currents.

Both equalizing coils are connected in series with the top end of the resistor; the system senses and adjusts the generator outputs to balance both generators at normal engine speeds. The generator outputs are connected through the reverse current relays (items 11 and 12) to the battery busbar. Equalizing circuits connect each resistor via equalizing coils and equalizing switches (items 13 and 14). When each generator has the same output, e.g. 50 A, the voltage dropped across each resistor is therefore $IR = 50 \times 0.01 = 0.5\,\text{V}$; there is no voltage difference and no current flows in the equalizing circuit, the circuit is now balanced. Over time, variations in output develop due to the mechanical differences in the regulators, e.g.

- carbon pile wear
- contact surface resistance
- mechanical variations in the armature.

For illustration purposes, assume that the left generator is regulating at an output less than the right generator. The voltage dropped across each of the resistors is

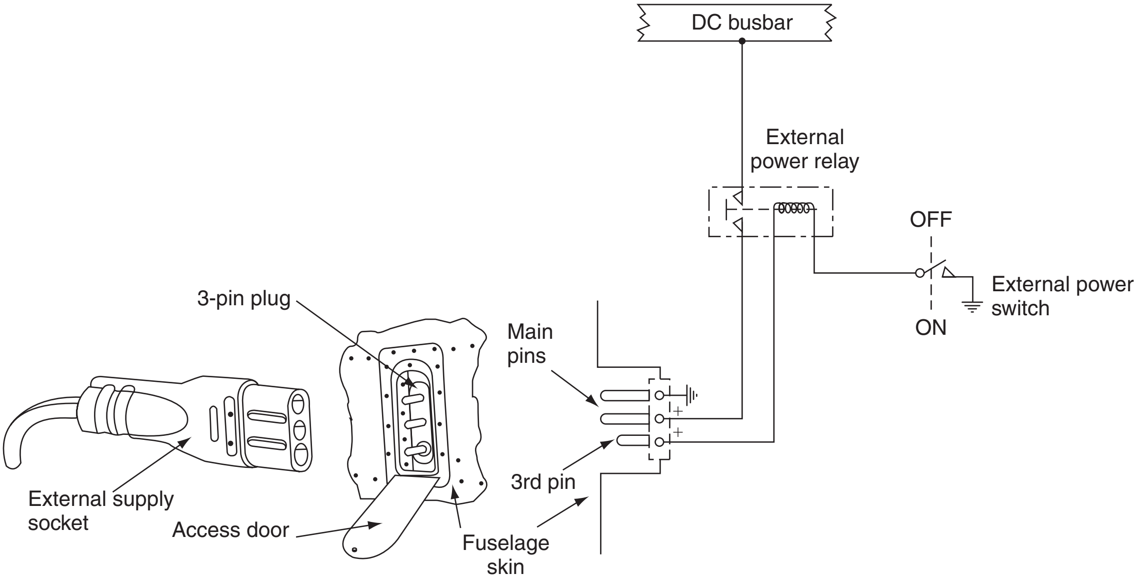

Figure 8.9 External ground power 3-pin connector schematic

now different since the left generator is supplying less current. To illustrate this effect, if the left generator is supplying 40A and the right generator is supplying 50A, the voltage drops across the respective resistors are:

$$V_{R1} = 40 \times 0.01 = 0.4 \text{ V}$$

$$V_{R2} = 50 \times 0.01 = 0.5 \text{ V}$$

This difference in voltage (0.1 volt) causes a current to flow from R2 through the right equalizing coil (8) into the left equalizing coil (7). This relaxes the right carbon pile regulator thereby decreasing the generator output. The left regulator coil compresses thereby increasing the left generator output. In practice, this circuit can only compensate for relatively small variations in output; the generators need to be trimmed on a periodic basis.

8.3 Larger aircraft systems

Larger (commuter, business and passenger) aircraft have many more electrical systems compared with general aviation aircraft; there is s requirement for a comprehensive approach to account for potential failures of generators, wiring, etc. The management of potential failures is addressed by categorizing the various loads and then disconnecting them in accordance with a predetermined sequence. The process of switching loads off the bus is called **load-shedding**; this can be achieved by automatic or manual control. These loads are connected onto specific busbars that fulfil a specific function. These can be categorized into a hierarchy as illustrated in Fig. 8.11. Connections between busbars are via heavy-duty contactors, or breakers. Aircraft types vary, however the following categories are typical for many installations.

Main bus: this is sometimes called the non-essential, generator, or load bus. It will include loads such as the galleys, in-flight entertainment (IFE) and main cabin lights. These loads can be disconnected and isolated in flight without affecting the safe operation of the aircraft.

Essential bus: this is sometimes called the vital or safety bus. It will include equipment and instruments required for the continued safe operation of the aircraft.

Battery bus: this is sometimes called the standby, or emergency bus. It supplies the equipment required for the safe landing of the aircraft, e.g. radios, fuel control, landing gear and fire protection.

There are three main types of distribution system architecture used on aircraft to fulfil the above:

- split bus system
- parallel system
- split parallel system.

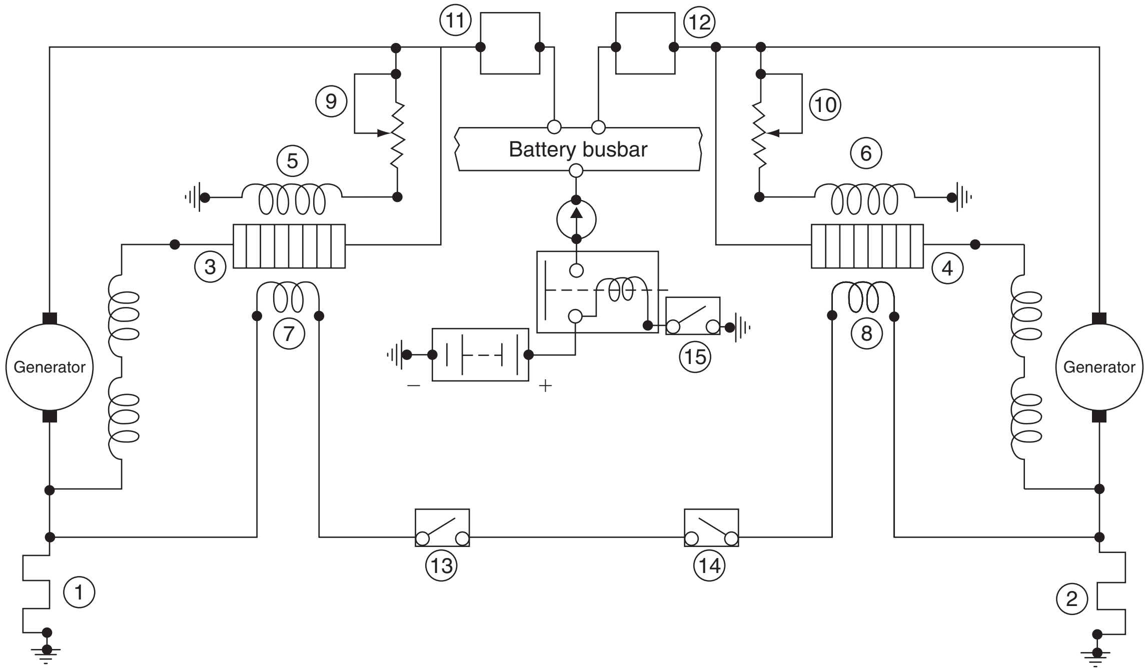

L	R	
①	②	Equalising resistors
③	④	Carbon pile regulators
⑤	⑥	Regulator voltage coils
⑦	⑧	Regulator equalising coils
⑨	⑩	Trimming resistors
⑪	⑫	Reverse current relays
⑬	⑭	Equalising switches
	⑮	Battery switch

Figure 8.10 Simple twin engine equalizing system schematic

8.4 Split bus system

This is a completely isolated twin generation system, sometimes called a **non-parallel** system used on twin-engine aircraft, see Fig. 8.12. **Primary power** is based on two main AC integrated drive generators (typically 40 kVA on each engine). An APU generator (40 kVA) is used as back-up in the event of a main integrated drive generator (IDG) failure. Note that the APU is normally a constant speed device in its own right; therefore an IDG is not required. The advantage of a split-bus system is that the generators do not need to be operating at exactly the same frequency and can be running out of phase with each other. **Secondary power** is derived from step-down transformers to provide 26 V AC; transformer rectifier units (TRU) provide 28 V DC for the DC busbars and battery charging.

Referring to Fig. 8.12, the right and left generators feed their own busbars to which specific loads are connected. Each generator bus is connected to a transfer bus via transfer relays. In the event of a generator failure, the remaining generator (engine or APU) supplies essential loads. Control of the system

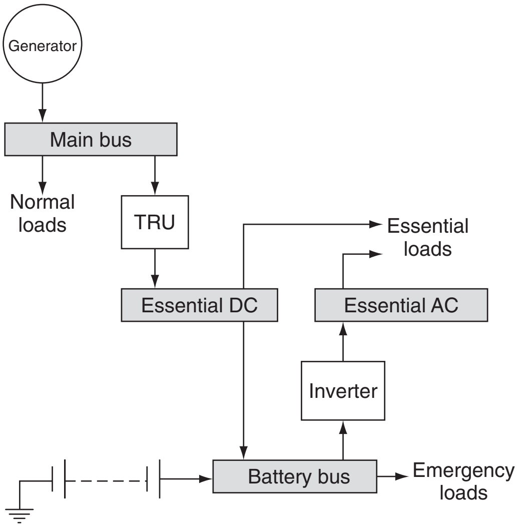
Generator
Main bus
Normal loads
TRU
Essential loads
Essential DC
Essential AC
Inverter
Battery bus
Emergency loads

Figure 8.11 Large aircraft electrical power hierarchy

is via a number of flight compartment switches, control breakers and relays arranged to connect and disconnect the generators and busbars. Typical control panel features for a split bus system are shown in Fig. 8.13; the features of this panel are:

Ammeters for the main generators to indicate load current
Ground power available (blue) when external power supply is connected
Ground power on/off switch to select ground power onto the aircraft
Transfer bus off (amber) when the transfer relay is de-energized (either normal or transfer)
Bus off (amber) both respective generator circuit-breakers (GCB) and bus tie breakers (BTB) open
Generator bus off (blue) if the respective GCB is open
APU generator bus off (blue) APU running at >95% RPM, no power from generator.

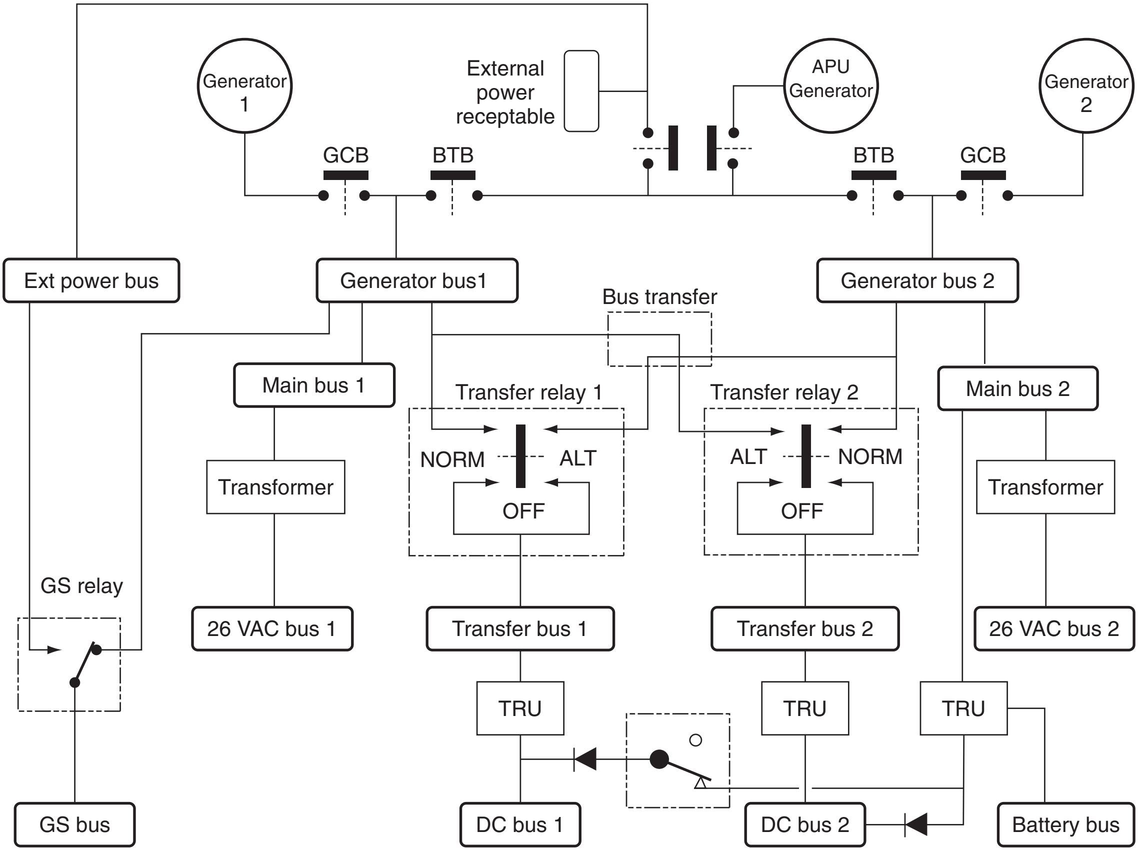
Generator 1
External power receptable
APU Generator
Generator 2
GCB
BTB
BTB
GCB
Ext power bus
Generator bus1
Generator bus 2
Bus transfer
Main bus 1
Main bus 2
Transfer relay 1
Transfer relay 2
NORM
ALT
OFF
ALT
NORM
OFF
Transformer
Transformer
GS relay
26 VAC bus 1
Transfer bus 1
Transfer bus 2
26 VAC bus 2
TRU
TRU
TRU
GS bus
DC bus 1
DC bus 2
Battery bus

Figure 8.12 Split bus system

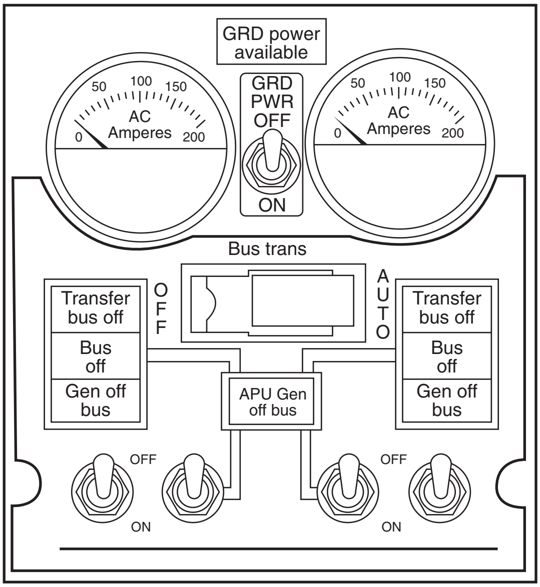

Figure 8.13 Typical electrical power control panel features

Key maintenance point

External AC power cannot be paralleled with the aircraft generators

---

8.5 Parallel bus system

The electrical distribution system on larger passenger aircraft (with three or four engines) are based on a parallel load distribution system, see Fig. 8.14. In this configuration, all generators are connected to their own AC load bus and a distribution bus; any generator can supply any load bus to provide equal load-sharing. All generator voltages, frequencies and phase relationships must be controlled to very close tolerances. Any attempt to connect generators in parallel before these conditions are met could result in loss of generator power due to large circulating currents. Referring to Fig. 8.14, when all four GCBs and BTBs are closed; all four generators are **synchronized** and connected to the tie (or synchronized) distribution busbar.

If one generator fails, its GCB is opened; this isolates the generator from its own load busbar. That busbar is now powered from the remaining generators. If this bus becomes overloaded, opening its GCB and BTB isolates it. With more than two generator failures, **load-shed**ding is introduced. External power can be made available by one or two power supply units (or carts). The APU can also be connected onto the distribution bus.

8.6 Split/parallel bus system

This is a flexible load distribution system for large passenger aircraft; it provides the advantages of the parallel system and maintains isolation when needed. Primary power supply features include: one IDG per engine, two APU generators and two external power connections. A split system breaker links left and right sides of distribution system. Any generator can supply any load busbar; any combination of generators can operate in parallel.

8.7 Standby and essential power

Essential 115 VAC power is provided as a single phase supply, see Fig. 8.15; this is selected from one phase of a main busbar. (On a four-engined aircraft, the normal source is the main AC bus number four.) If this normal source fails, a supply is maintained by selecting a different main bus from the remaining systems. The standby AC bus is normally powered from the essential AC bus; if this source fails, a static inverter (supplied from the main battery) is selected to provide standby AC power.

Some aircraft distinguish between **battery busbars** (that can be disconnected from the battery) and **hot battery busbars** that are connected directly to the battery, i.e. without any switching; this is illustrated in Fig. 8.15. This arrangement splits the battery bus with a direct connection (the ‘hot’ battery bus), and a switched battery bus controlled by the battery switch. Essential DC power is from a transformer rectifier unit (TRU) powered from the essential AC bus. Standby AC and DC bus power is normally from the respective essential supplies; when selected ON, standby AC and DC power is from an inverter and the battery respectively.

Test your understanding 8.3

Explain the difference between main, essential and battery busbars.

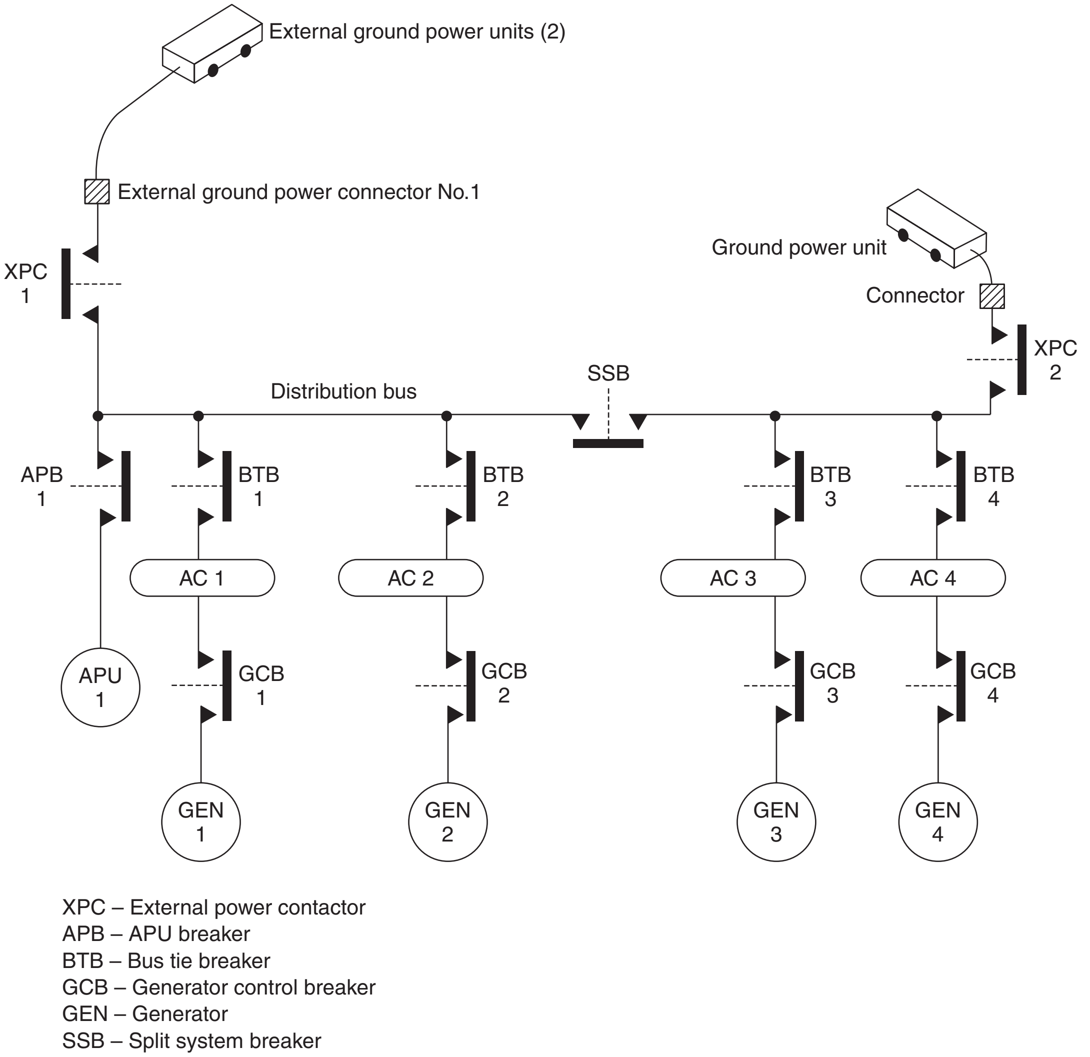
External ground power units (2)
External ground power connector No.1
XPC 1
Ground power unit
Connector
XPC 2
Distribution bus
SSB
APB 1
BTB 1
BTB 2
BTB 3
BTB 4
AC 1
AC 2
AC 3
AC 4
APU 1
GCB 1
GCB 2
GCB 3
GCB 4
GEN 1
GEN 2
GEN 3
GEN 4
XPC – External power contactor
APB – APU breaker
BTB – Bus tie breaker
GCB – Generator control breaker
GEN – Generator
SSB – Split system breaker

Figure 8.14 Parallel load distribution system

8.8 Battery charging

The battery charger on a large aircraft operates from 115-volt, three-phase, 400 Hz AC power supplied by the AC ground service bus. A completely discharged nickel-cadmium battery can be recharged in approximately 60–90 minutes. The battery charger operates in one of two modes depending on whether the aircraft is being supplied by external ground power or not.

With the aircraft on the ground, the charger is powered from the ground services bus and provides a constant current to the battery. When the battery's terminal voltage reaches a pre-defined level (adjusted for battery temperature), charging is automatically disconnected. When external power is not available, the battery is charged from a transformer rectifier unit (TRU) that provides 28 V DC to maintain battery charge, and supply loads on the battery bus. The battery's temperature sensor forms part of the charging system to prevent battery damage. If the battery temperature is outside a predefined range (typically –18°C to 60°C), the charging circuit is disconnected.

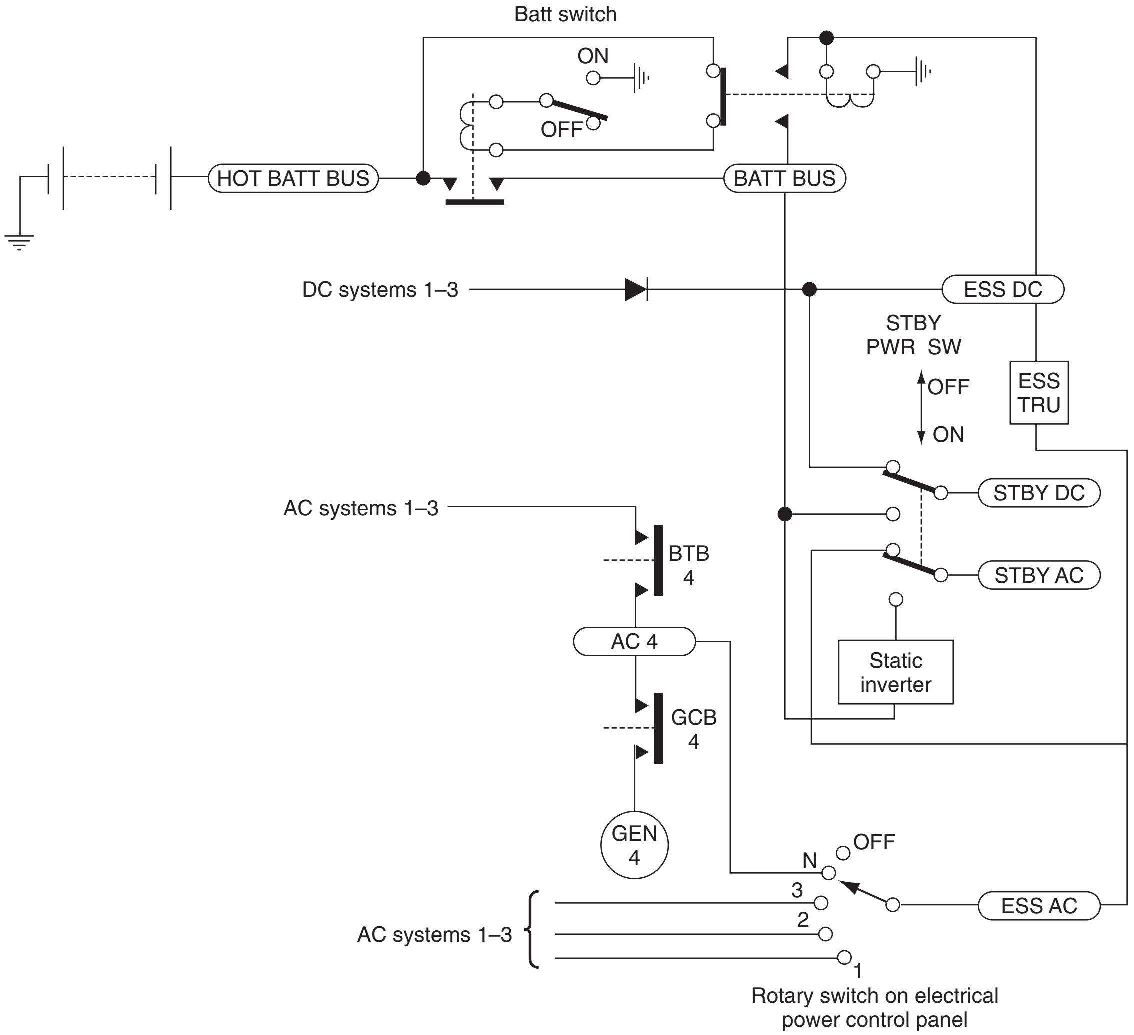
Batt switch
ON
OFF
HOT BATT BUS
BATT BUS
DC systems 1–3
ESS DC
STBY
PWR SW
OFF
ON
ESS
TRU
STBY DC
AC systems 1–3
BTB
4
STBY AC
AC 4
Static
inverter
GCB
4
GEN
4
OFF
N
3
2
1
AC systems 1–3
ESS AC
Rotary switch on electrical
power control panel

Figure 8.15 AC/DC standby and essential power

8.9 Control and protection

Various components are used for both control and protection of the power distribution system:

- current transformers
- differential current protection
- phase protection
- breakers/contactors.

8.9.1 Current transformers

These are used to sense current for control, protection and indication applications. The primary winding is the main heavy-duty AC feeder cable being monitored; the secondary winding is contained within a housing, see Fig. 8.16. The secondary windings are in the form of inductive pick-up coils. When current I_P flows in the feeder cable, the corresponding magnetic field induces current I_S into the secondary windings; this is the output signal that is used by a control, protection or indication device.

8.9.2 Differential current protection

This circuit detects short-circuits in AC generator feeder lines or busbars; it is a method of protecting the generator from overheating and burning out.

Assuming a three-phase AC generator is installed, each phase has its own protection circuit. For illustration purposes, the circuit for a single phase is described. Two control transformers (CT) are located at either end of the distribution system, see Fig. 8.17. CT_1 is located in the negative (earthed) connection of the generator's output. CT_2 is located at the output from the busbar performing a monitoring function in a **generator control unit** (GCU).

If a fault were to develop between the generator and busbar, a current I_F flows to ground. The net current received at the busbar is therefore the total generator output current I_T minus the fault current (I_F). The fault current flows back through the earth return system through CT_1 and back into the generator; the remaining current ($I_T - I_F$) flows through CT_2 and into the loads. Current transformer CT_1 therefore detects $(I_T - I_F) + I_F$ which is the total generator current. Current transformer CT_2 detects ($I_T - I_F$); the difference between control transformer outputs is therefore I_F. At a pre-determined differential current, the generator control relay (GCR) is automatically tripped by the GCU and this opens the generator field.

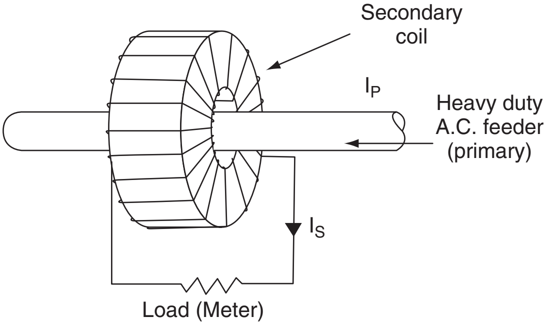

Figure 8.16 Current transformer

8.9.3 Phase protection (Merz Price circuit)

This circuit protects against faults between phases, or from individual phase to ground faults. Connections are shown for protection of a single phase in Fig. 8.18; a three-phase system would require the same circuit per phase. Two current transformers (CT) are located at each end of the feeder distribution line:

- CT_1 monitors the current output from the generator
- CT_2 monitors the current into the distribution system.

Secondary windings of each current transformer are connected via two relay coils; these windings are formed in the opposite direction. When current flows through the feeder, there is equal current in both coils; the induced EMF is balanced, so no current flows. If a fault develops in the feeder line, current CT_1 flows (but not CT_2), thereby creating an unbalanced condition.

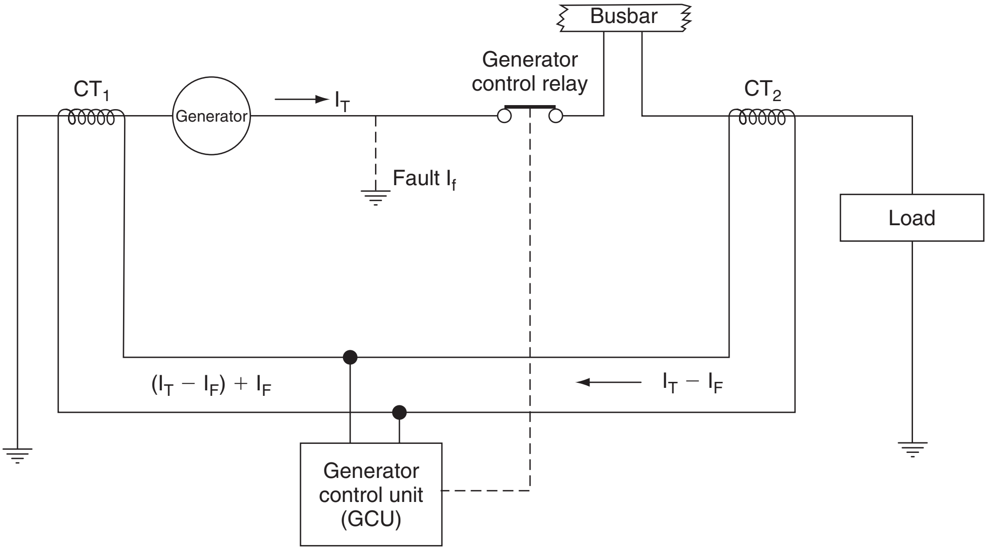

Figure 8.17 Differential current protection

Current flow in either of the coils opens the contacts and disconnects the feeder line at both ends.

8.9.4 Breakers/contactors

Breakers (sometimes referred to as contactors) are used in power generation systems for connecting feeder lines to busbars and for interconnecting various busbars. Unlike conventional circuit-breakers, these devices can be tripped on or off remotely. Referring to Fig. 8.19, they have several heavy-duty main contacts to switch power and a number of auxiliary contacts for the control of other circuits, e.g. warning lights, relays etc.

The breaker is closed by an external control switch via contacts A; the coil remains energized via contacts B to ground. With the coil energized, the main and auxiliary contacts are closed and the spring is compressed. Contacts A latch the breaker closed, assisted by the permanent magnet. When a trip signal is applied (either by a fault condition or manual selection) current flows to ground in the opposite direction. The spring assists the reversed electromagnetic field and this breaks the permanent magnet latch. A Zener diode suppresses arcing of coil current across the contacts. An electrical power breaker installation is shown in Fig. 8.20.

8.10 Load-shedding

Load-shedding can be defined as deciding on which systems to switch off and in what order this should be done to reduce the power consumption for the remaining power sources. (The worst-case scenario is when the aircraft is operating from the battery alone.) Non-essential loads are not required for the safe continuation of flight; these loads include the galleys, in-flight entertainment (IFE) system and main cabin lights as previously described. Essential loads are required for the safe continuation of flight. Emergency loads are required for the safe landing of the aircraft, e.g. radios, fuel control, landing gear and fire protection.

8.10.1 General aviation aircraft

In small general aviation (GA) aircraft, pilots need to be familiar with the electrical systems of their aircraft. Most light GA aircraft are only concerned about an alternator or generator failure. This type of aircraft

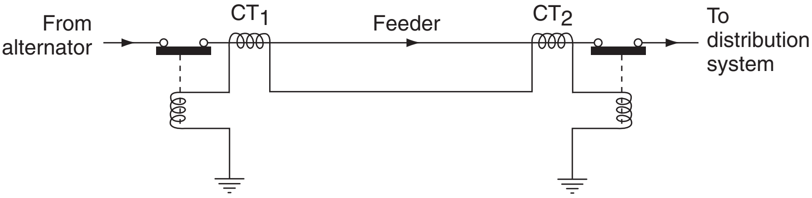

Figure 8.18 Phase protection (Merz Price circuit)

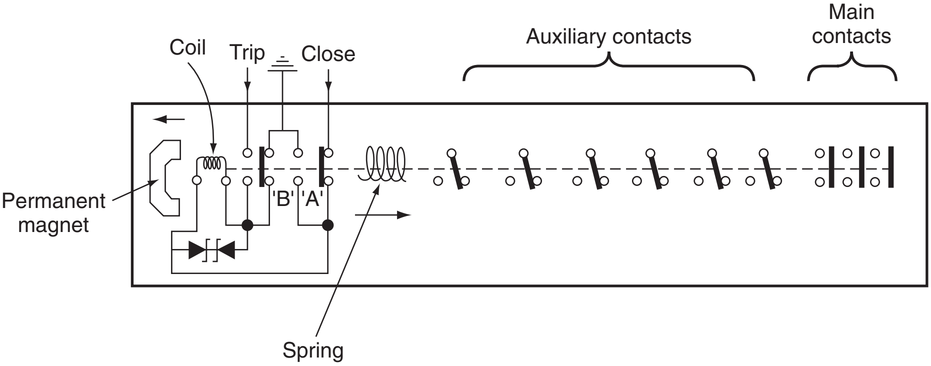

Figure 8.19 Electrical power breaker schematic

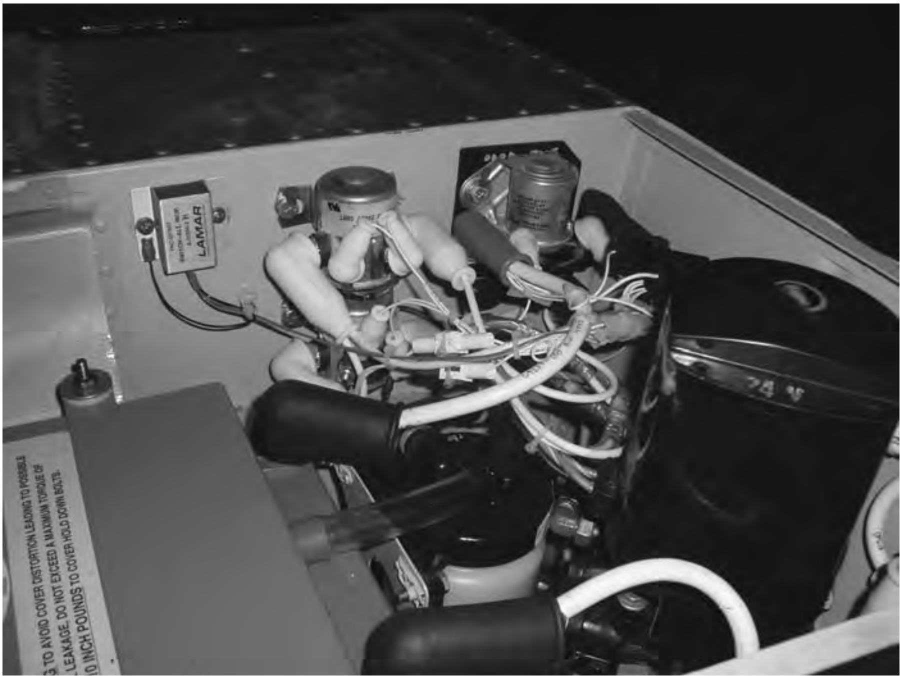
LAMAR
24 V
G TO AVOID COVER DISTORTION LEADING TO POSSIBLE
LEAKAGE. DO NOT EXCEED A MAXIMUM TORQUE OF
10 INCH POUNDS TO COVER HOLD DOWN BOLTS.

Figure 8.20 Electrical power breaker installation

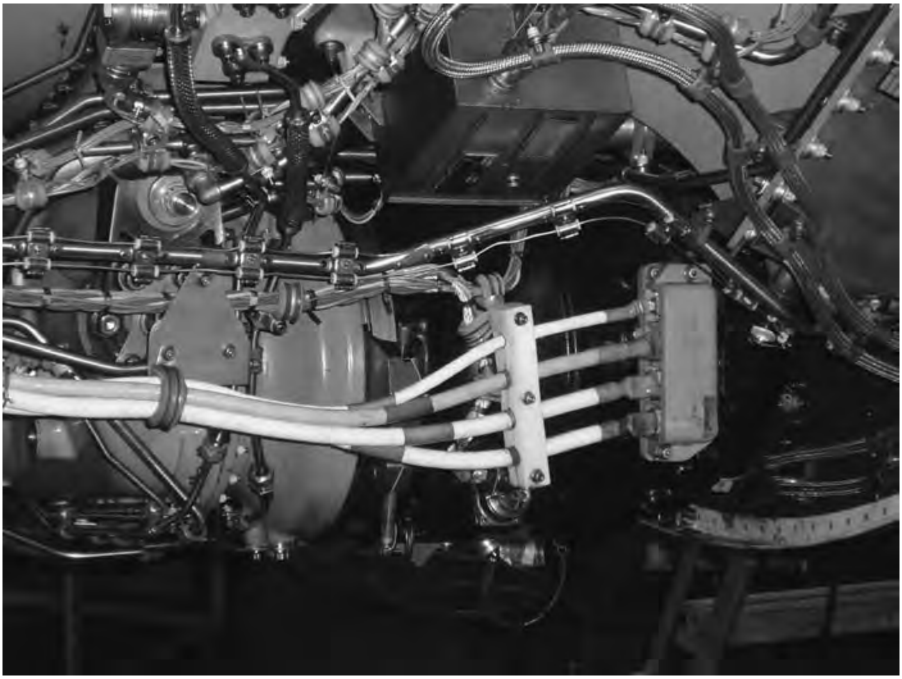

Figure 8.21 Generator three-phase output and neutral wires

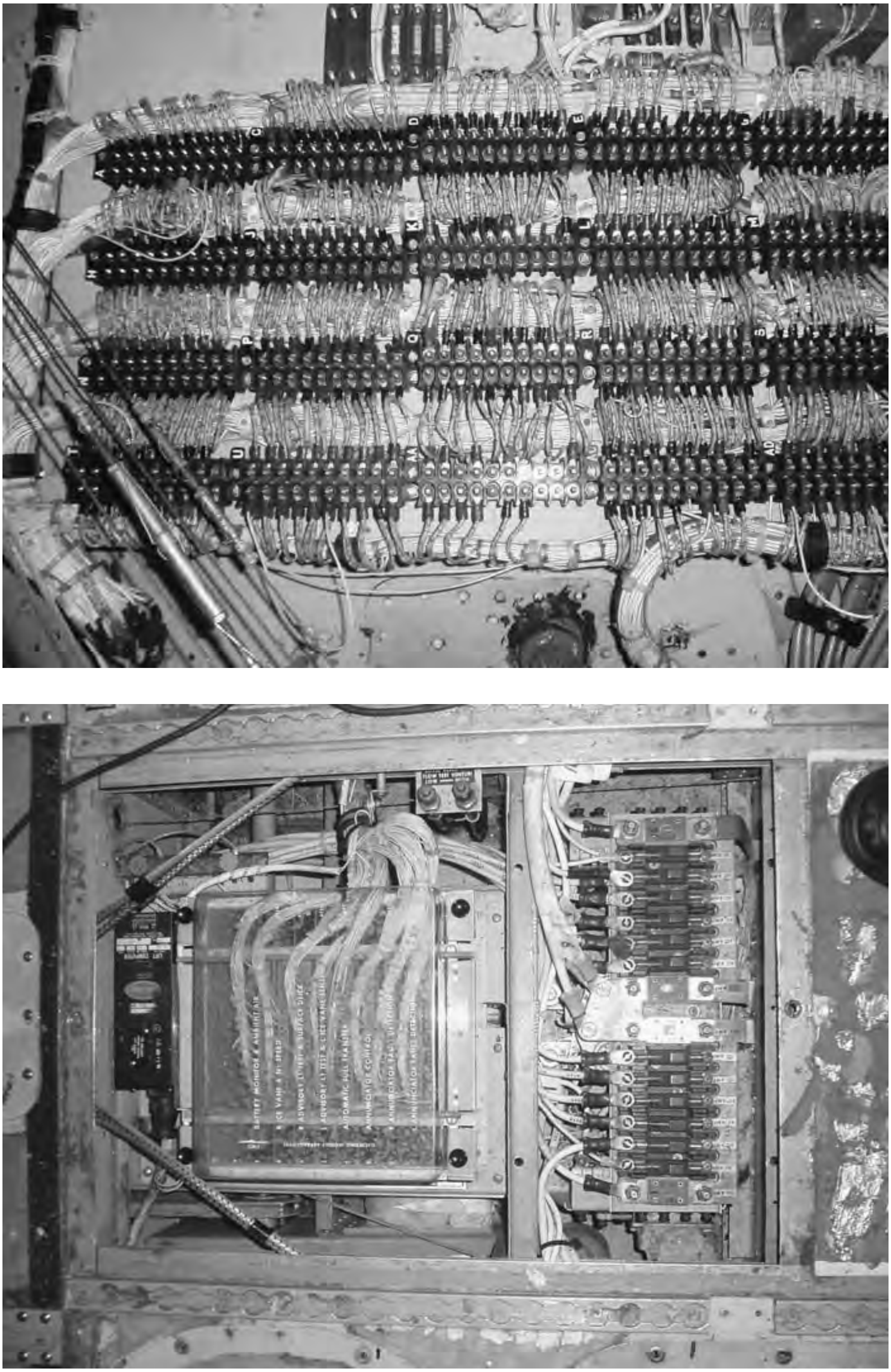

Figure 8.22 Typical power distribution

is normally fitted with an ammeter and/or a voltmeter. On some light GA aircraft an over-voltage light and ammeter can be used to determine the nature of electrical malfunctions. Some aircraft are also fitted with a light that comes on in the event of generator failure, where the battery alone is being used to supply the avionic bus. The aircraft battery has a finite capacity; the pilot has to make decisions on the use of the aircraft's power consumption in order to be able to make a safe landing.

8.10.2 Larger aircraft

More complex aircraft and/or twin-engined aircraft require that pilots have greater knowledge about the electrical systems. Automatic electrical load-shedding is provided on modern aircraft where a control unit detects the loss of electrical power, i.e. the loss of a single or both generators. Automatic load-shedding disconnects specified electrical equipment to maintain sufficient power and systems functionality for the safe continuation of flight.

Split bus or parallel systems need to have a centralized means of controlling power distribution. If a generator or bus fault develops, the relevant bus tie breakers (BTB) or generator circuit-breakers (GCB) must be configured to protect the remaining generators, whilst still delivering power to essential loads. Load-shedding is controlled automatically by a solid-state bus power control unit (BPCU). This unit receives inputs from the generator control unit (GCU), ground power control unit GPCU and BTBs, together with control transformers that sense real system current through the main power leads. The BPCU is programmed to detect abnormal conditions and open/close pre-determined BTBs.

Galley power represents a high percentage of the total electrical system's load since it comprises equipment such as beverage makers and ovens. These systems have a major influence on the electrical system's design.

8.10.3 Load-sharing of AC circuits

It is necessary to discriminate between real and reactive loads in AC load-sharing circuits. Current transformers in the feeder lines sense each load; these feeder lines are heavy-gauge wires connected to the output of the generator (Fig. 8.21). The feeder lines are connected through the various bus hierarchy as described earlier before finally being distributed to the individual loads via smaller-gauge wires and terminal blocks (Fig. 8.22).

Resistive loads consume **real power**; unbalances are corrected by adjusting the generator shaft input power (from the IDG or CSD). Inductive or capacitive loads consume **reactive power**; unbalances are corrected by adjusting the generator output via the field current control of the regulator. **Apparent power** consumed by entire circuit. Reactive power is the vector sum of inductive and capacitive currents and voltages, see Fig. 8.23.

- True power is measured in watts (W)
- Apparent power is measured in volt-amps (VA)
- Reactive power is measured in volt-amps-reactive (VAR).

Power factor (PF) is the ratio of true and apparent power (W/VA); this is a measure of circuit efficiency.

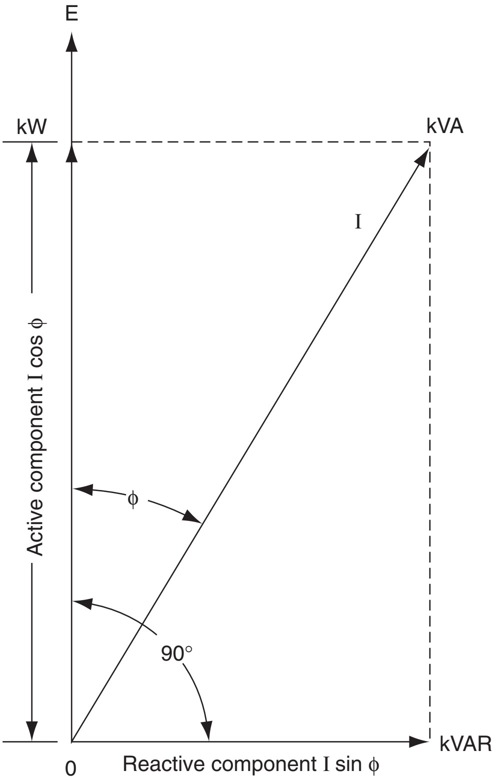

Figure 8.23 Components of currents due to phase difference

A power factor of one (unity) is for a purely resistive circuit; a power factor of zero is for a purely reactive circuit. Aircraft generators and inverters are rated in kVA, with a stated power factor over a given range.

8.11 Multiple choice questions

1. The reverse current relay is needed on any DC generation system to prevent the:
   (a) battery from feeding excess current back through to the generator's armature
   (b) generator from feeding excess current back through to the battery
   (c) battery from feeding excess current to the starter motor.

2. The current limiter device:
   (a) increases the output from a DC generator under fault conditions
   (b) reduces the output from a DC generator under fault conditions
   (c) reduces the output from a DC generator under normal conditions.

3. Battery charge/discharge current is monitored via which type of instrument?
   (a) voltmeter
   (b) contactor
   (c) ammeter.

4. Referring to Fig. 8.24 different size pins are used on external DC power connectors to:
   (a) prevent a reverse polarity voltage being applied
   (b) prevent excessive power being applied
   (c) prevent power being applied when the battery is discharged.

5. The main distribution bus is sometimes called the:
   (a) essential bus
   (b) emergency bus
   (c) non-essential bus.

6. The split bus system is sometimes called a:
   (a) non-parallel system
   (b) parallel system
   (c) standby and essential power system.

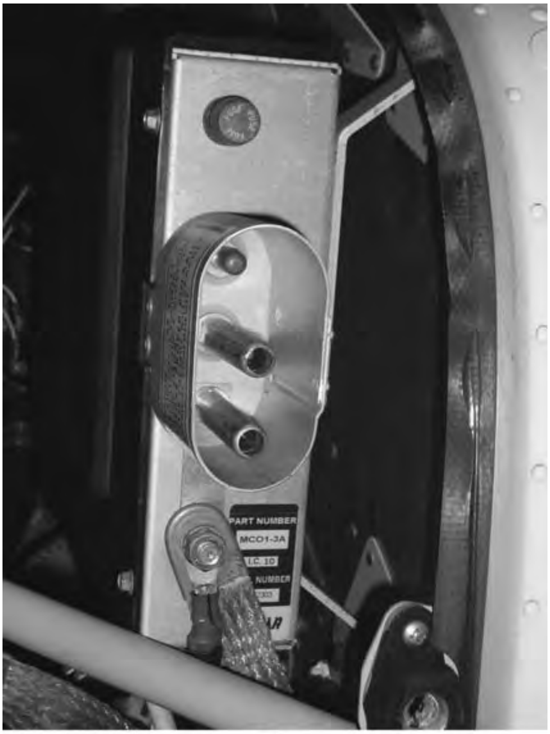

(a)

Figure 8.24 See Question 4

7. Inductive or capacitive loads consume:
   (a) real power
   (b) DC power
   (c) reactive power.

8. Essential DC power could be supplied from a:
   (a) transformer rectifier unit (TRU) powered from the essential DC bus
   (b) ttransformer rectifier unit (TRU) powered from the essential AC bus
   (c) inverter powered from the essential AC bus.

9. Referring to Fig. 8.25 apparent power is measured in:
   (a) volt-amps (VA)
   (b) watts (W)
   (c) volt-amps-reactive (VAR).

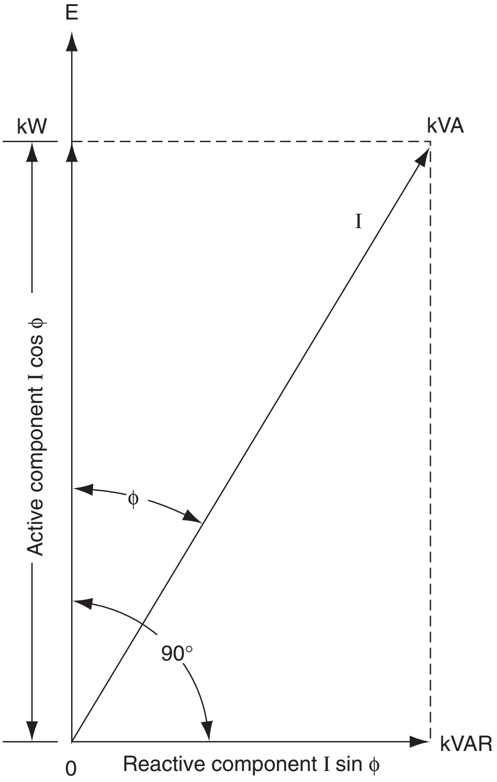
E
kW
kVA
I
Active component I cos ϕ
ϕ
90°
0
Reactive component I sin ϕ
kVAR

Figure 8.25 See Question 9

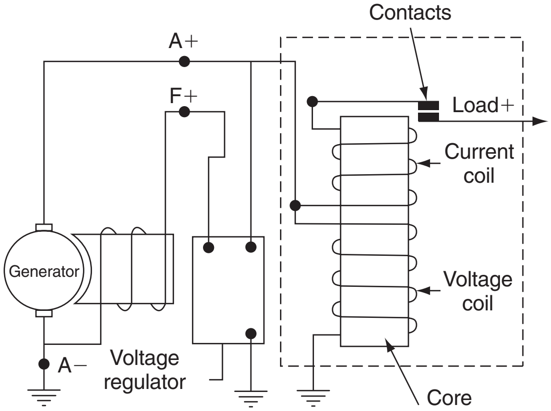
Contacts
A+
F+
Load+
Current coil
Generator
Voltage coil
A−
Voltage regulator
Core

Figure 8.26 See Question 11

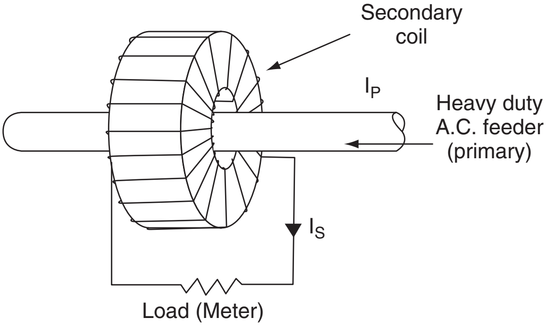
Secondary coil
I_P
Heavy duty A.C. feeder (primary)
I_S
Load (Meter)

Figure 8.27 See Question 12

10. The third pin on external DC power connectors is used to:
    (a) charge the battery
    (b) energize the reverse current relay
    (c) energize the ground power relay.
11. Referring to Fig. 8.26, this device is a
    (a) external power relay
    (b) current limiter
    (c) reverse current relay
12. Referring to Fig. 8.27, this is a schematic diagram for a
    (a) contractor
    (b) current transformer
    (c) auto-transformer

Chapter 9

Controls and transducers

There are many systems on aircraft that need to be controlled and/or monitored, either manually by the crew, or automatically. A switch provides the simplest form of control and monitoring. For example, the crew need to know if any doors are not closed as part of their pre-flight check; on larger aircraft the position of control surfaces is displayed in the cockpit. Using these two examples, doors can be either open or closed; control surfaces can move through an infinite number of positions (within their normal limits of travel). Many other aircraft parameters need to be measured, e.g. temperature and pressure. Measurements are made by a variety of **transducers**; these are devices used to convert the desired parameter, e.g. pressure, temperature, displacement etc. into electrical energy. This chapter describes generic controls and transducer devices used on aircraft.

9.1 Switches

The simplest form of switch is the on/off device used to isolate circuits. Other switch types are used to direct the current into pre-determined parts of a circuit. Switches are characterized by:

- number of poles
- number of switched positions
- type of switched contacts (permanent or momentary).

Switches are sometimes guarded or wire-locked with fuse wire to prevent inadvertent operation. Some switch designs have to be pulled out of a detent position before the position can be changed. They are designed to be operated in a number of ways, e.g. toggle, push/pull, rocker or rotary selectors. These switches are designed with multiple contacts; they can be arranged as permanent or momentary contacts.

The **toggle switch** is a very basic device; Fig. 9.1 illustrates its internal schematic and external features. Operating the lever/arm opens and closes switch contacts. Operating levers on toggle switches are sometimes ganged so that more than one circuit is operated.

The simplest switch has two contact surfaces that provide a link between circuits; these links are referred to as **poles**. Switch contacts can be normally open or closed and this is normally marked on the switch (NO/NC). The number of circuits that can be linked by a single switch operation is called the **throw**. The simplest form of switch would be single pole, single throw (SPST). Schematics of switch configurations commonly found in aircraft are illustrated in Fig. 9.2.

Some switches are designed with instinctive tactile features so that the risk of selecting the wrong system is minimized, e.g. the flap up/down switch-operating lever would be shaped in the form of an aerofoil; the undercarriage selection switch-operating lever would be shaped in the form of a wheel (see Chapter 14 for examples).

Push/pull-operated switches incorporate a spring to hold the contacts open or closed; the switch contacts are therefore push-to-make or push-to-break. Rocker switches (Fig. 9.2b) combine the action of toggle and push/pull devices. Rotary switches are formed by discs mounted onto a shaft; the contacts are opened and closed by the control knob.

9.1.1 Combined switch/light devices

Modern aircraft panels utilize a combined switch and light display; the display is engraved with a legend or caption indicating system status. These can be used in a variety of ways e.g. to show system on/off. The switch portion of the device can be momentary or continuous; small level signals are sent to a computer or high-impedance device. Internal backlighting is from two lights per legend for redundancy; these are projected via coloured filters. The two captions provide such information as press to test (P/TEST). Examples of combined switch/light devices are given in Fig. 9.3.

Test your understanding 9.1

Explain the switch terms ‘throw’ and ‘pole’.

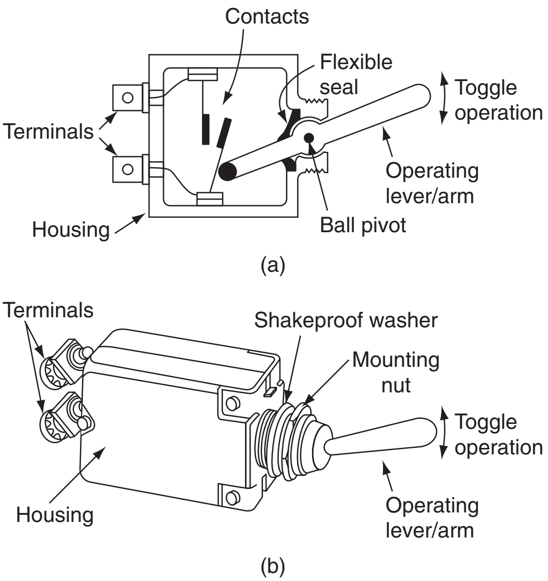
Contacts
Flexible seal
Toggle operation
Terminals
Operating lever/arm
Ball pivot
Housing
Terminals
Shakeproof washer
Mounting nut
Toggle operation
Housing
Operating lever/arm

(a)

(b)

Figure 9.1 Toggle switch: (a) internal schematic, (b) external features

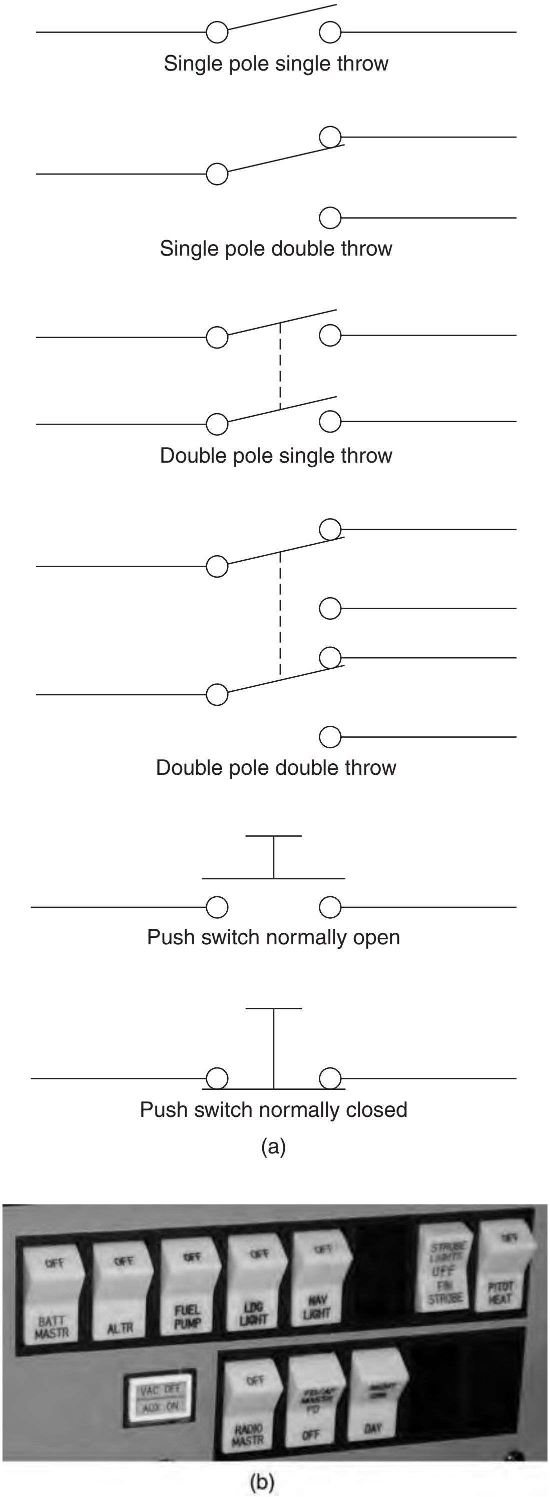
Single pole single throw
Single pole double throw
Double pole single throw
Double pole double throw
Push switch normally open
Push switch normally closed
OFF
BATT MASTR
ALTR
FUEL PUMP
LDG LIGHT
NAV LIGHT
STROBE LIGHTS
FIN STROBE
PITOT HEAT
VAC OFF
ALX ON
RADIO MASTR
FD
DAY

(a)

(b)

Figure 9.2 (a) Switch configurations, (b) rocker switches

9.1.2 Micro-switches

These are used to sense if a device has moved or has reached its limit of travel, e.g. flap drive or undercarriage mechanisms. Figure. 9.4 illustrates the internal schematic and electrical contacts of a typical micro-switch product. The contacts open and close with a very small movement of the plunger. The distance travelled by the armature between make/break is measured in thousandths of an inch, hence the name '**micro**'. A snap action is achieved with a contact mechanism that has a pre-tensioned spring.

Micro-switches are attached to the structure and the wiring is connected into a control circuit. An example of micro-switch application is to sense when the aircraft is on the ground; this is achieved by mounting a micro-switch on the **oleo leg**. When the aircraft is on the ground, the oleo leg compresses and the switch is operated. Micro-switches are used to sense the mechanical displacement of a variety of devices, including:

- control surfaces
- undercarriage
- pressure capsules
- bi-metallic temperature sensors
- mechanical timers.

Proximity switches perform the same function as micro-switches; they sense the presence of an object

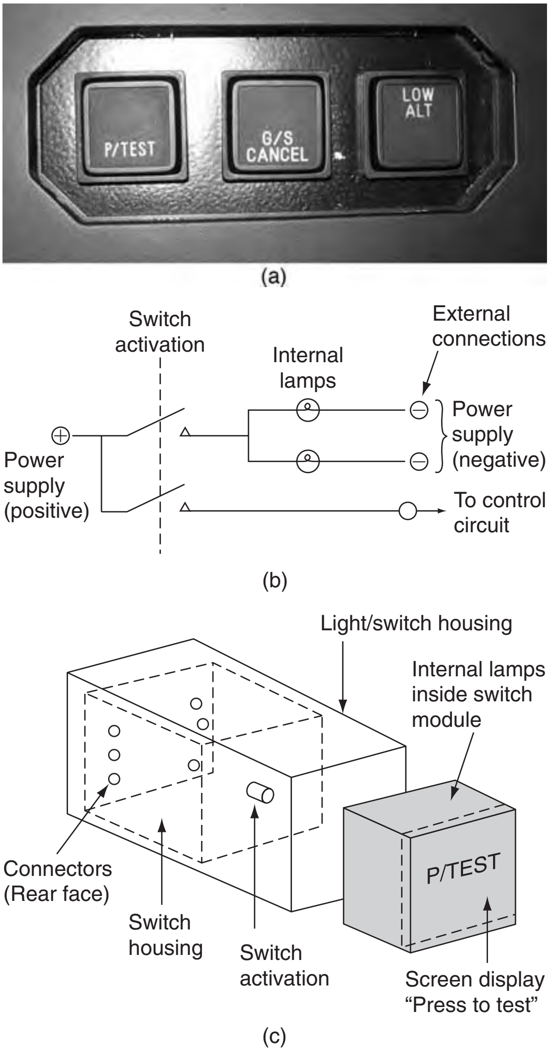
P/TEST
G/S CANCEL
LOW ALT
(a)
Switch activation
Internal lamps
External connections
Power supply (positive)
Power supply (negative)
To control circuit
(b)
Light/switch housing
Internal lamps inside switch module
Connectors (Rear face)
Switch housing
Switch activation
P/TEST
Screen display "Press to test"
(c)

Figure 9.3 Combined switch/light: (a) typical installation, (b) electrical schematic, (c) external features

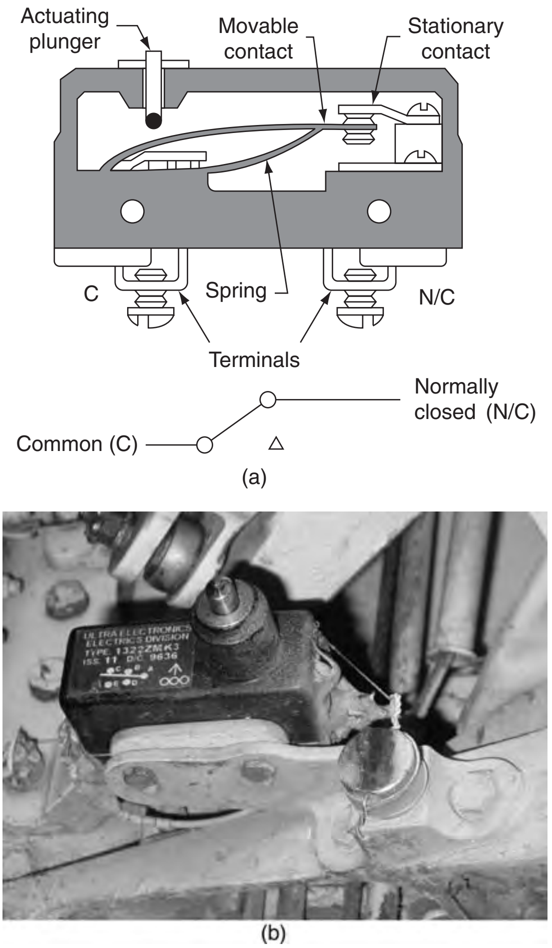
Actuating plunger
Movable contact
Stationary contact
C
Spring
N/C
Terminals
Normally closed (N/C)
Common (C)
(a)
ULTRA ELECTRONICS ELECTRICS DIVISION TYPE 1322ZMK3 ISS 11 D/C 9636
(b)

Figure 9.4 Micro-switch schematic: (a) internal schematic, (b) external features

by the interruption of a magnetic circuit. There are two types of proximity switch: reed and solid state.

The **reed switch** device comprises two hermetically sealed sections as illustrated in Fig. 9.5. One section (the actuator) contains a magnet; the other section (the sensor) contains a reed armature with rhodium-plated contacts. The usual arrangement is for the sensor unit to be fixed to the aircraft structure; the actuator is attached to the item being monitored, e.g. a door. When the gap between the actuator and the sensor reaches a pre-determined distance, the reed contacts close thereby completing the circuit. They open again when the actuator and sensor are moved apart.

The solid state proximity switch is based on an **inductance loop** and steel target as illustrated in Fig. 9.6. This inductance loop is the input stage of an electronic switch unit incorporated as part of the actuator. As the target moves closer to the coil, the inductance of the coil changes. An electronic circuit determines when the inductance has reached a pre-determined level. The obvious advantage of this type of switch is that there are no switch contacts, hence higher reliability.

9.1.3 Proximity switch electronic unit

Some aircraft are installed with a proximity switch electronic unit (PSEU). This unit receives the position of various items, e.g. flaps, gear, doors, etc and communicates this information to other systems including:

- take-off and landing configuration warnings
- landing gear position indicating and warning

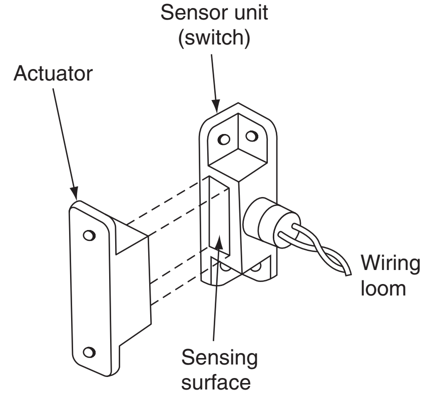

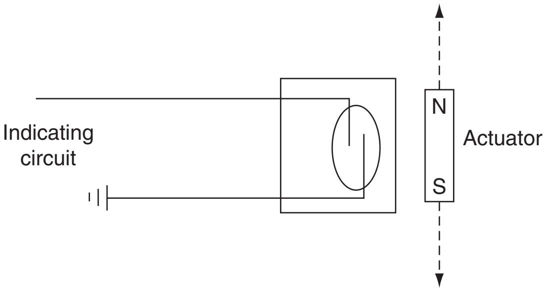

Figure 9.5 Proximity switch (reed type) schematic

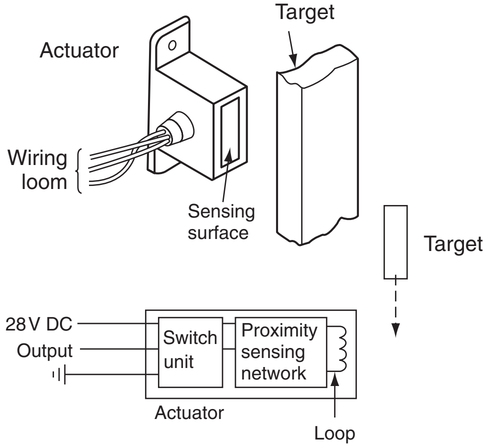

Figure 9.6 Proximity switch (solid state) schematic

- air/ground relays
- airstairs and door warnings.

The PSEU is integrated with the master caution system, and used to indicate if a problem exists that has to be corrected before flight. (The above systems are discussed in subsequent chapters.)

Key point

Typical applications for micro-switches and proximity switches include sensing the positions of flaps, gear, doors etc. On a large transport aircraft there can be up to 100 such devices.

Key maintenance point

The maintenance engineer will often need to adjust micro-switch or proximity switch positions to ensure that they are correctly positioned; this will compensate for any displacement of the switch.

Test your understanding 9.2

What is the difference between proximity and micro-switches?

9.2 Relays and contactors

These electromechanical devices interrupt or complete a circuit when activated from a remote source, see Fig. 9.7. **Changeover** relays consist of a coil, moving contact (armature) and external connections. When the coil has current flowing through it the electromagnetic effect pulls in the contact **armature**. The armature is pivoted and is held in position by the spring force; with no current flowing, the armature returns to its original position by the spring force.

Contactors operate in the same way; the difference between them is their physical construction and application. Relays are generally used for low current applications; contactors (also known as breakers) are used for switching higher currents, e.g. for connecting battery power to the aircraft. The features of a contactor include the main power contacts and auxiliary contacts used for indication and control of other devices, e.g. lights and relays in power distribution systems (contactors/breakers are described in this context in Chapter 8).

Reed relays are used in control circuit applications; they are generally found within components, e.g. mounted onto printed circuit boards. Figure. 9.8 illustrates the principles of a reed relay. **Slugged relays** have delayed operating times and are needed in specialized applications. The delay in opening/closing the contacts is achieved by a second coil wound around the main coil; the turns are arranged such that the build up of magnetic flux in the main coil is opposed by the build up of magnetic flux in the secondary coil.

9.2.1 Relay configurations

The simplest relay has two contact surfaces that provide a link between circuits; these links are referred to as **poles**. Relay contacts can be normally open or closed (NO/NC) and this is normally marked on the body of the relay. The number of circuits that can be linked by a single relay operation is called the **throw**. The simplest form of relay would be single pole, single throw (SPST). Relay configurations commonly found in aircraft are illustrated in Fig. 9.9. Typical relay installations on a general aviation aircraft are shown in Fig. 9.10.

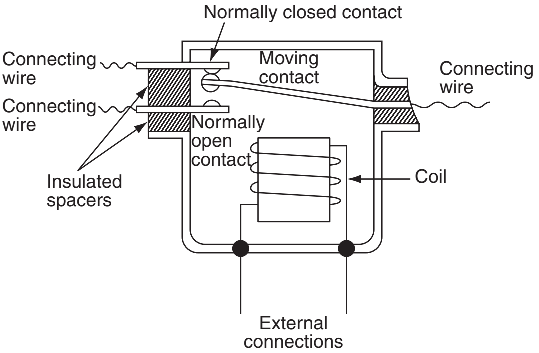
Normally closed contact
Connecting wire
Connecting wire
Moving contact
Connecting wire
Normally open contact
Coil
Insulated spacers
External connections

Figure 9.7 Changeover relay schematic

Key maintenance point

The pull-in voltage of a relay is the specified voltage need to attract the armature and close or open the switch contacts.

Key maintenance point

The drop-out voltage of a relay is the specified point at which the armature returns to its relaxed position when the magnetic force is overcome by the spring force.

9.2.2 Polarized relays

These are used in control circuits with very low voltages or currents; the relay is extremely sensitive and can respond to levels in the order of mA/mV. This low level is often not suitable for the conventional spring-loaded armature device since (at very low pull-in/drop-out voltages) the contacts would chatter, leading to spark erosion. Polarized relays use magnetic forces to attract and repel the armature instead of a spring force. The armature is a permanent magnet, pivoted between two pole faces formed by the frame of high

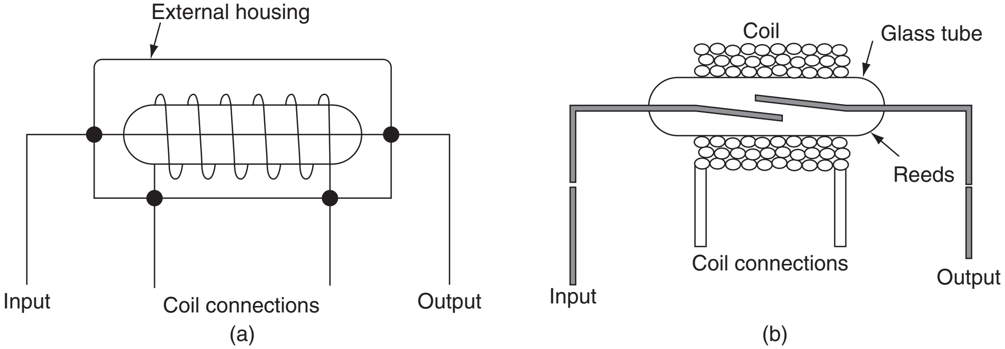
External housing
Input
Coil connections
Output
(a)
Coil
Glass tube
Reeds
Coil connections
Input
Output
(b)

Figure 9.8 Reed relay schematic (a) external schematic, (b) internal schematic

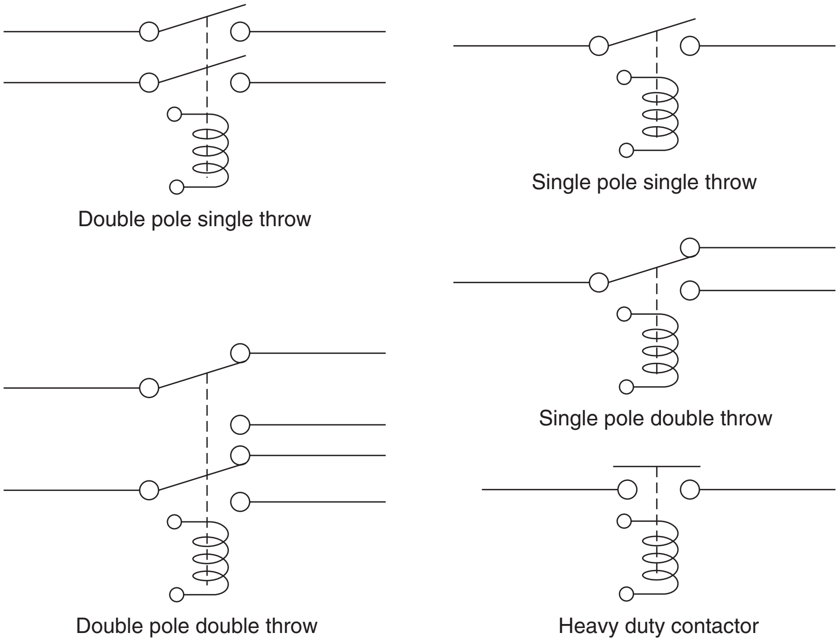

Double pole single throw

Single pole single throw

Single pole double throw

Double pole double throw

Heavy duty contactor

Figure 9.9 Relay configurations

Figure 9.10 Relay installation (GA aircraft)

permeability material. When current flows, the poles change and the frame becomes an electromagnet; this exceeds the force exerted by the permanent magnet and the armature changes position. In this position the N-S poles form a strong attractive force and the armature is retained in position. If the supply is interrupted the electromagnetic force is reduced to less than the permanent magnet and the armature returns to its original position.

Key maintenance point

Contacts in relays and contactors are made from silver alloy. The contacts need to have sufficient current flowing/switching to provide '**wetting**' of the contacts. In-service problems and failure associated with these contacts include:

- contact erosion
- welded contacts
- contact corrosion.

9.3 Variable resistors

Variable resistors are mounted on a linear slider or rotary shaft to provide a user-adjustable resistance; typical applications include the control of lighting, audio volume or generator regulator trimming. They are sometimes combined with micro-switches to provide an on/off control function. Variable resistors are produced as either:

- potentiometers
- preset resistors
- rheostats.

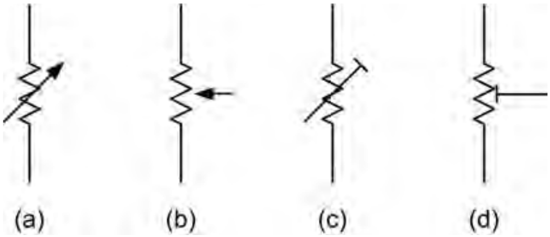

Figure 9.11 Variable resistors: potentiometers, preset or rheostats

Figure 9.11 provides some examples of symbols used for variable resistors. When the intention is for the pilot or maintenance engineer to adjust the circuit resistance for control purposes, e.g. audio volume or lighting intensity, the variable resistor device is used. If the circuit resistance is only intended to be adjusted in the workshop, a preset device is used.

A **potentiometer** (often called a 'pot' for short) is a type of variable resistor that is normally used as a **voltage divider**; this is a circuit used to supply a portion of the power supply voltage from a resistive contact. The potentiometer is typically a three-terminal resistor with a sliding centre contact (the wiper). If all three terminals are used, it can be used in the voltage divider application. If only two terminals are used (one end of the resistor and the wiper), it acts as a variable resistor.

Key maintenance point

Corrosion or wearing of the sliding contact can lead to audio volume or lighting intensity problems, especially if the centre contact (the wiper) is kept in one position for long periods.

Key point

A **rheostat** performs the same function as the potentiometer, but is physically much larger, being designed to handle much higher voltage and/or current. Rheostats are constructed with a resistive wire formed as a toroidal coil, with the centre contact/wiper moving over the surface of the windings.

9.4 Linear displacement transducers

Many aircraft parameters need to be measured, including: speed, altitude, temperature and pressure. Measurements are made by a variety of **transducers**; these are devices used to convert the desired parameter into an electrical signal that can be used in a control system and/or display.

9.4.1 Solenoids

The solenoid is a type of transducer that converts electrical energy into **linear displacement**. Typical applications include the actuation of pneumatic or hydraulic valves. They are electromechanical devices, consisting of an inductive coil wound around a steel or iron **armature**. The coil is formed such that the armature can be moved in and out of the solenoid's body. The armature is used to provide the mechanical force required to the item being moved.

9.4.2 LVDT

The **linear variable differential transformer** (LVDT) is a transducer used for measuring small linear displacements. The LVDT has three transformer windings located within a housing; the centre coil is the primary winding; the two outer coils are the secondary windings, see Fig. 9.12. A ferromagnetic core, connected to the item being measured, moves along the axis of the housing. The 26 VAC 400 Hz supply is connected to the primary windings, causing a voltage to be induced in each of the secondary windings proportional to its mutual inductance with the primary.

When the core is displaced from its neutral position, the mutual inductances change, causing the voltages induced in the secondary windings to change. The coils are wound such that the output voltage is the difference between the two secondary output voltages. When the core is in its neutral position (equidistant between the two secondaries) equal and opposite voltages are induced in the two secondary coils, so the output voltage is zero.

When the core is displaced, the voltage in one secondary coil increases as the other decreases; the resulting output voltage starts to increase in phase with the primary voltage. When the core moves in the opposite direction, the output voltage increases, but is out of phase with the primary. The output voltage is therefore proportional to displacement of the core; the

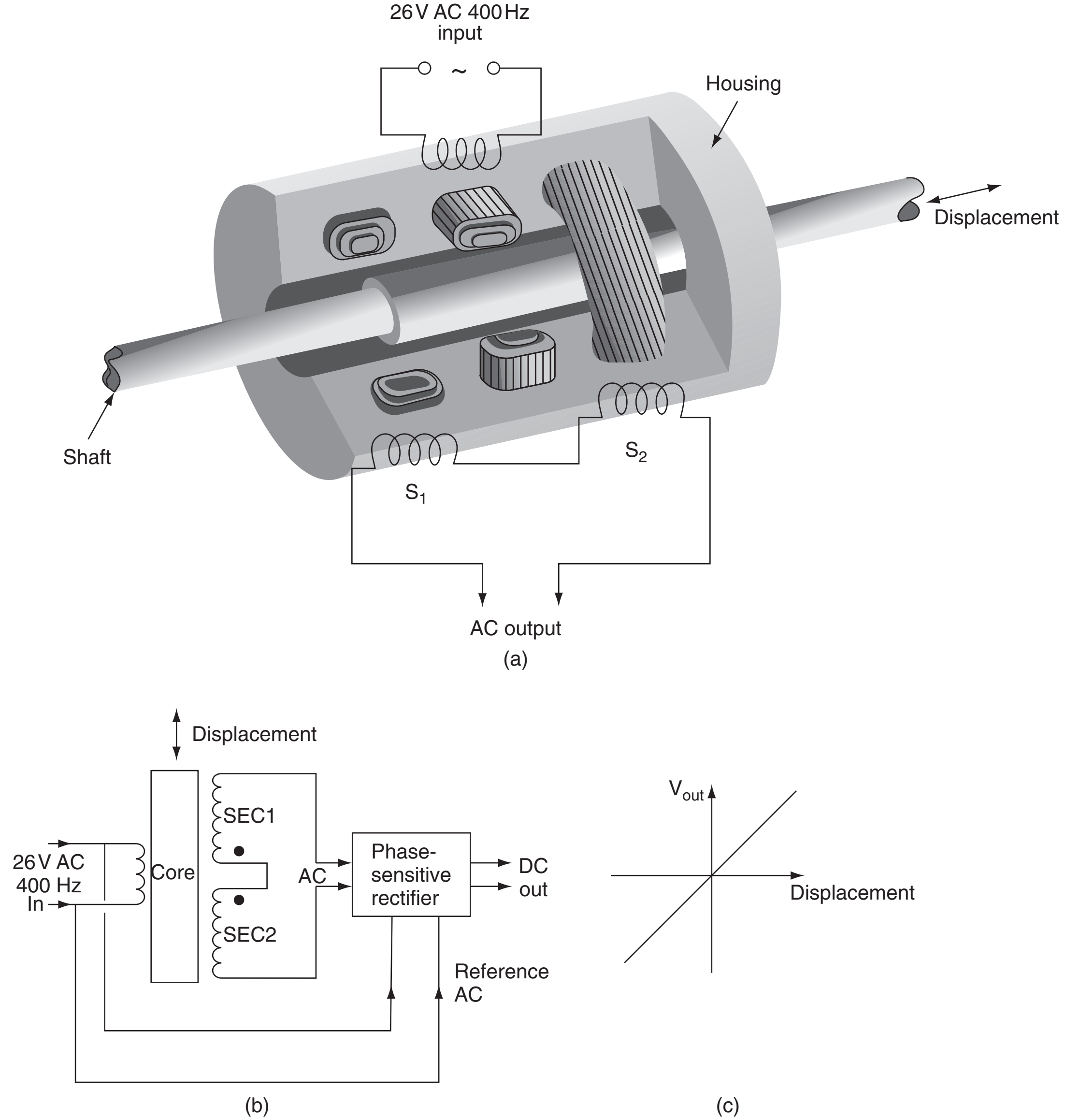
26 V AC 400 Hz input
Housing
Displacement
Shaft
S_2
S_1
AC output
(a)
Displacement
SEC1
26 V AC 400 Hz In
Core
AC
Phase-sensitive rectifier
DC out
SEC2
Reference AC
(b)
V_{out}
Displacement
(c)

Figure 9.12 Linear variable differential transformer (LVDT) (a) mechanical schematic (b) electrical schematic (c) output voltage vs displacement

output voltage phase is used to determine the direction of movement. This AC output is then fed into a phase sensitive rectifier; the phase difference is then indicated by the +/− polarity of the DC output.

9.4.3 EI sensor

Another device used for measuring very small movements in control systems is the EI sensor. This device gets its name from the shapes of the fixed E and pivoted I laminated cores, see Fig. 9.13(a). The centre limb of the E core is wound with the primary turns of a transformer; the outer limbs of the E core are wound with the secondary turns, see Fig. 9.13(b). These are wound in series, and in opposition to each other. The I bar is connected mechanically to the device being measured for displacement, e.g. a pressure capsule; it is pivoted and wound with a coil that is connected to a reference power supply.

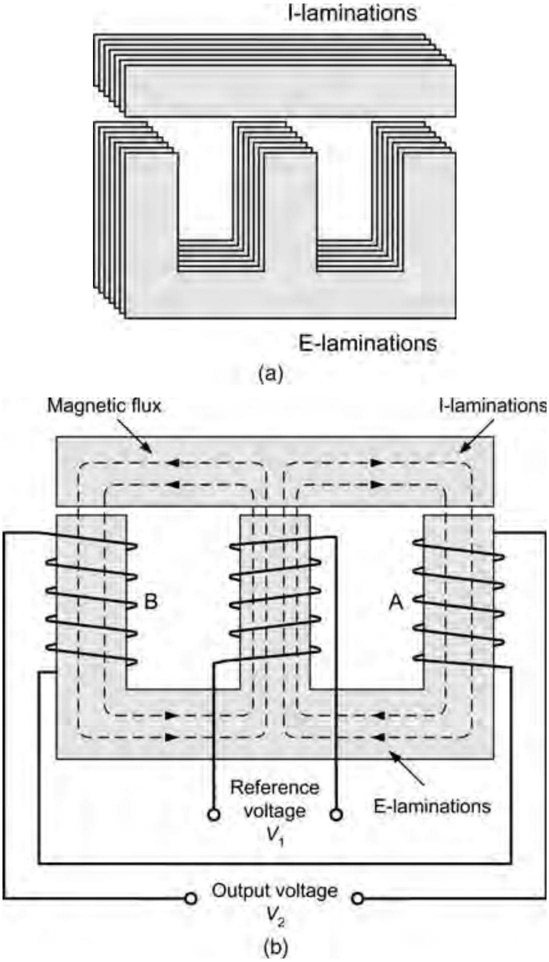
I-laminations
E-laminations
(a)
Magnetic flux
I-laminations
B
A
Reference voltage V_1
E-laminations
Output voltage V_2
(b)

Figure 9.13 EI sensor – general arrangement: (a) core arrangement, (b) core windings

With the I bar in a neutral position:

- the air gaps at each end are equal
- the magnetic flux is equal
- equal and opposite voltages are generated in the outer limbs
- the output from the sensor is zero.

Referring to the schematic in Fig. 9.14, when the I bar moves, the air gaps become unequal, causing the reluctance to change. Flux in one limb (with the smaller gap) increases, flux in other limb decreases. The induced voltages change; the output voltage is amplified for use in a control system. This measures the phase angle and magnitude of the EI sensor.

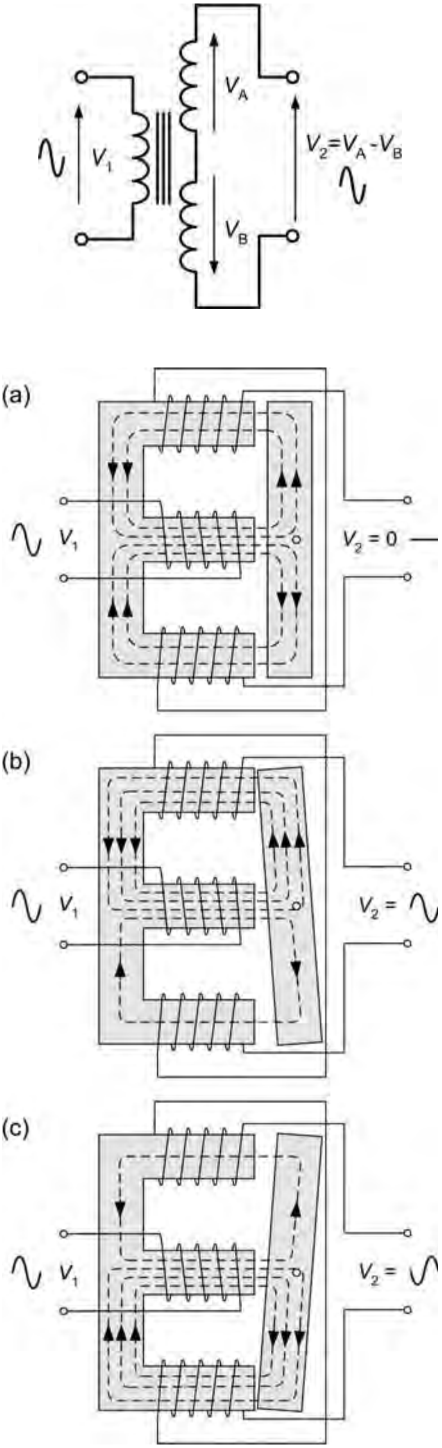
V_1
V_A
V_B
$V_2=V_A-V_B$
(a)
V_1
$V_2 = 0$
(b)
V_1
$V_2 =$
(c)
V_1
$V_2 =$

Figure 9.14 EI sensor operation

9.5 Fluid pressure transducers

There are two basic methods used to measure fluid pressure: the Bourdon tube or a capsule. Typical fluid pressures being measured on the aircraft include hydraulic oil and fuel.

9.5.1 Bourdon tube

The **Bourdon tube** was invented by Eugène Bourdon (1808–84), a French watchmaker and engineer. The pressure-sensing element is a tube with either a flat or elliptical section; it is formed as a spiral or curve, see Fig. 9.15. One end of the tube is sealed and connected to a pointer mechanism, the open end is connected to the fluid system via a pipe.

As the applied pressure from the fluid system increases, the tube will tend to straighten out, while a reduced pressure will cause the tube to return to its original shape. This movement is transferred via the gear mechanism to move a pointer. The pointer moves across a scale thereby providing a **direct reading** of pressure. Materials used for the tube are selected for the pressure range being measured; these include phosphor bronze (0–1000 psi) and beryllium copper (0–10,000 psi). The Bourdon tube principle can also be used to remotely measure pressure, see Fig. 9.16.

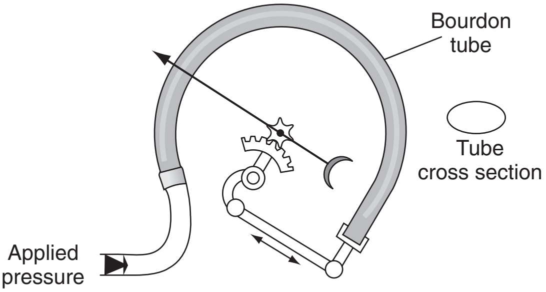

Figure 9.15 Bourdon tube principles

9.5.2 Pressure capsule

The other method used for measuring fluid pressure is with a **capsule**, see Fig. 9.17. As pressure is applied by the fluid, the capsule expands. This moves an iron bobbin within the envelope of two inductor coils.

These coils are part of an **AC ratiometer system** as illustrated in Fig. 9.18. The linear shaded pole motors are single-phase induction motors. An unbalance of currents in the inductor coil circuits is caused by displacement of the cores within the coil. Current then flows in one of two directions into the motor circuits. The rotors of the motor are the aluminium discs attached to a shaft. This arrangement produces a low starting torque to produce a responsive indicating system.

An alternative arrangement is to connect the iron bobbin to a micro-switch to provide indications of low/high pressure. The advantage of this capsule system is that pressure pipes do not need to be run all the way up to the cockpit indicators.

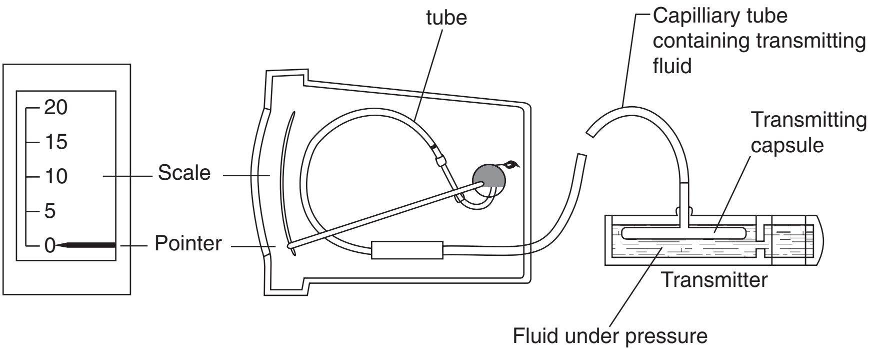

Figure 9.16 Bourdon tube/remote pressure sensing

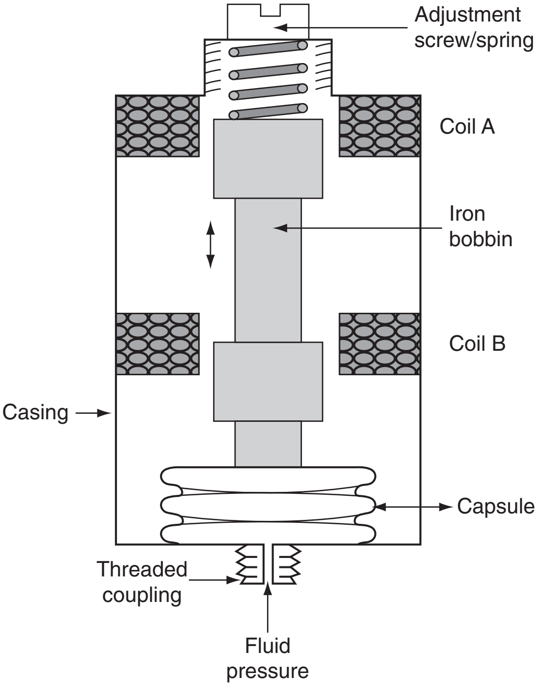
Adjustment screw/spring
Coil A
Iron bobbin
Coil B
Casing
Capsule
Threaded coupling
Fluid pressure

Figure 9.17 Pressure transducer

9.6 Temperature transducers

Aircraft use a variety of temperature transducers, or sensing devices, ranging from the simple bi-metallic strip or thermostat through to thermistors and thermocouples. The type of sensor used depends on the application and takes into account:

- temperature range
- accuracy requirements
- type of output required
- cost.

9.6.1 Bi-metallic strip

A bi-metallic strip converts temperature changes into mechanical displacement as an on/off measurement. The device consists of two strips of different metals with different coefficients of thermal expansion. The strips or elements are joined together by rivets, by brazing or by welding (see Fig. 9.19).

Differential expansion causes the element to bend one way when heated, and in the opposite direction when cooled below its nominal temperature. The metal

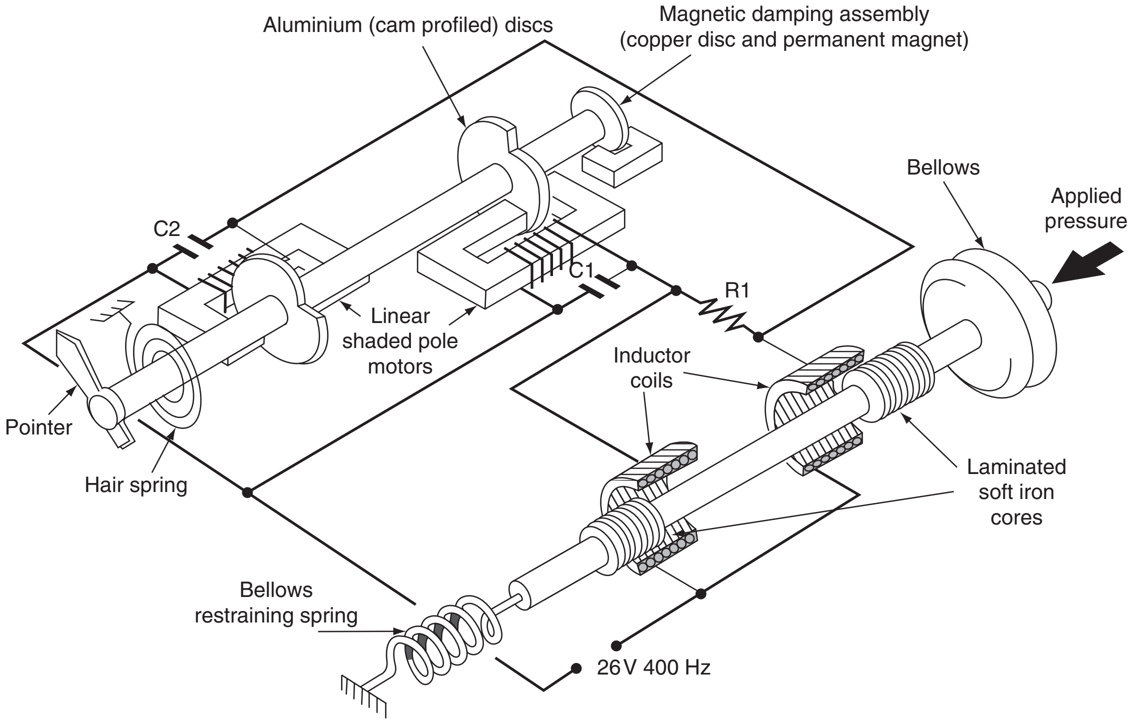
Aluminium (cam profiled) discs
Magnetic damping assembly (copper disc and permanent magnet)
Bellows
Applied pressure
C2
C1
R1
Linear shaded pole motors
Inductor coils
Pointer
Hair spring
Laminated soft iron cores
Bellows restraining spring
26 V 400 Hz

Figure 9.18 AC ratiometer

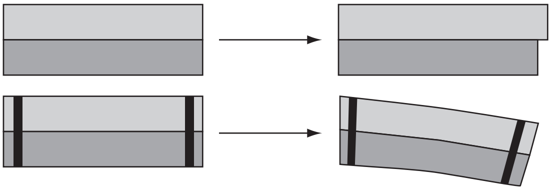
(a) (b)

Figure 9.19 Bi-metallic strip principles: (a) before heating, (b) after heating.

with the higher coefficient of expansion is on the outer side of the curve whilst the element is heated and on the inner side when cooled. The element can be formed as a switch contact, called a **thermostat**, to open or close a circuit when temperature limits are exceeded. Alternatively the element is formed in a spiral shape so that temperature changes cause a shaft to rotate; the typical application for this is an outside air temperature indicator, see Fig. 9.20 (a). The sensing portion of the indicator projects through windscreen (Fig. 9.20(b)) into the ambient air.

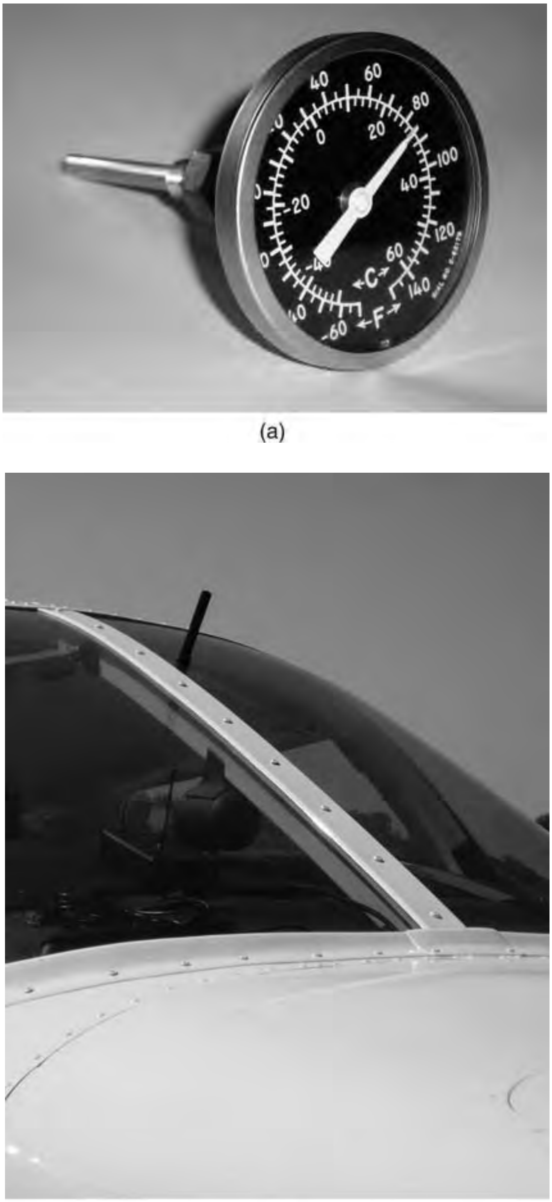

(a)

(b)

Figure 9.20 Bi-metallic strip application: (a) outside air temperature indicator, (b) outside air temperature (sensor projecting through windscreen)

9.6.2 Thermistors

A thermistor is a type of resistor (*therm*al res*istor*) used to measure temperature; the principle is based on the known relationship between resistance and temperature. The thermistor product used on aircraft is the resistance temperature device (RTD), also known as a resistance thermometer detector, or **temperature bulb.** Figure. 9.21 illustrates some features of an RTD. A length of very small diameter wire is wound on a bobbin to form resistance windings. The bobbin is contained within a stainless steel body to protect the windings against contaminants. Each end of the windings is soldered onto pins that are formed into an electrical connector; this is welded to the flange, resulting in a hermetically sealed unit.

The principles of this temperature-measuring device is based on measuring the change of resistance of a metal element and interpreting this as temperature. Metal elements have a **positive temperature coefficient**; certain metals, e.g. nickel and platinum, have a very stable and linear relationship between resistance and temperature, see Fig. 9.22. Nickel is less accurate than platinum, and is non-linear below 200°C, but has lower cost.

RTDs are typically used within a **DC ratiometer circuit**, see Fig. 9.23. A permanent magnet is attached to the pointer of the measuring instrument. When the RTD resistance is equal to the fixed resistor value, the current in each leg is equal and the pointer takes up the centre position. With a change in temperature, the RTD

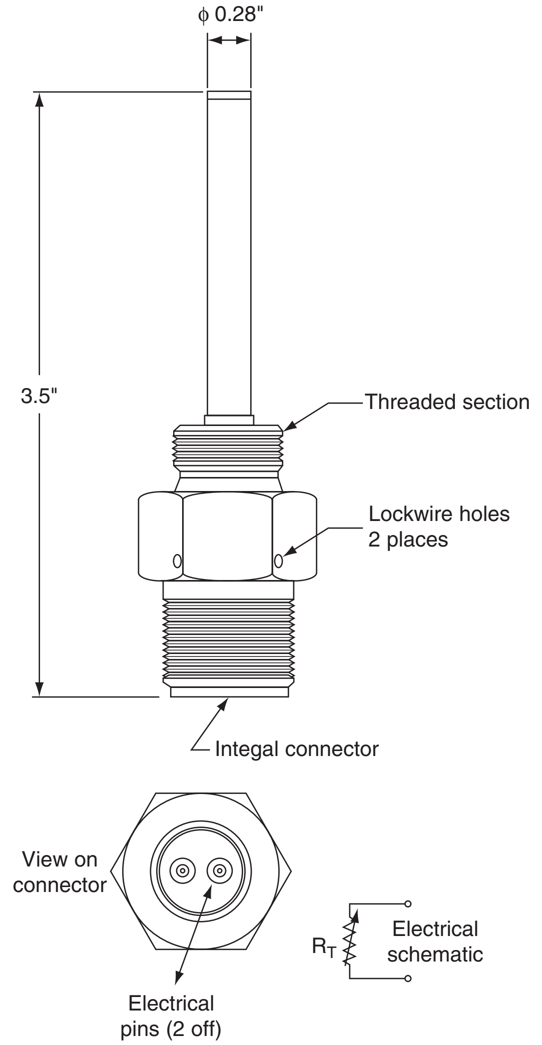
ϕ 0.28"
3.5"
Threaded section
Lockwire holes
2 places
Integal connector
View on
connector
Electrical
pins (2 off)
R_T
Electrical
schematic

Figure 9.21 Resistance temperature device (RTD) features

resistance value changes and the current through each leg is not equal, thereby causing the pointer to take up a new position. The position of the pointer is in proportion to the **ratio** of currents in each leg of the network.

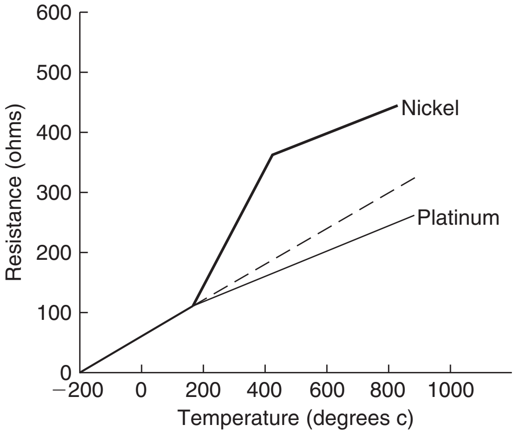

Figure 9.22 RTD temperature and resistance chart

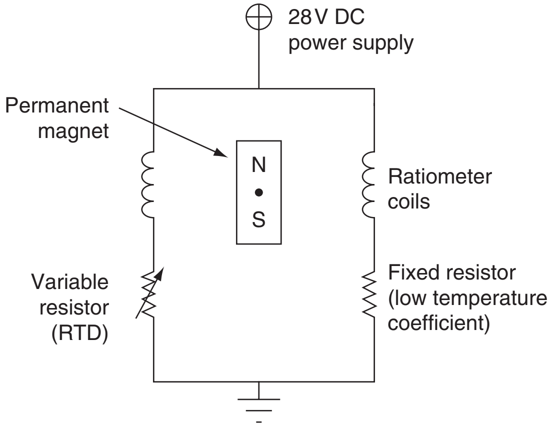
28 V DC
power supply
Permanent
magnet
N
S
Ratiometer
coils
Variable
resistor
(RTD)
Fixed resistor
(low temperature
coefficient)

Figure 9.23 DC ratiometer

Alternatively they are used in a Wheatstone bridge circuit, named after Sir Charles Wheatstone (1802–75), a British scientist. The circuit was actually invented by Samuel Hunter Christie (1784–1865), a British scientist and mathematician; Sir Charles Wheatstone further developed the circuit. It was used to measure an unknown value of resistance by balancing two legs of a bridge circuit, one leg of which includes the unknown value.

This circuit developed by Wheatstone is based on Ohm's and Kirchhoff's laws. It is typically used as an accurate method to overcome the effect of power supply variations. Referring to Fig. 9.24, resistors R_1, R_2, and R_3 have fixed values; the RTD is the variable resistor in the bridge network. Current I_T splits from the power supply splits into I_1 and I_2. Depending on temperature, if the RTD value equals R_1 then $I_1 = I_2$ and $V_A = V_B$ so I_3 is zero and the meter pointer has a zero deflection. A change in value of RTD (as a result of temperature) means that V_A is no longer equal to V_B. Current I_3 increases and the meter pointer is

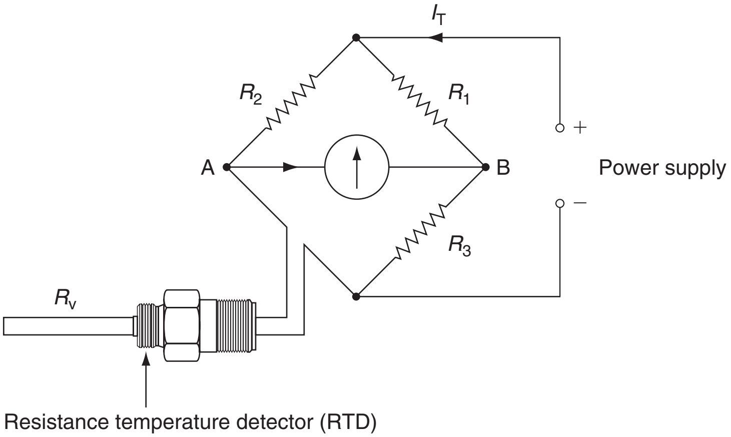
I_T
R_2
R_1
A
B
R_3
+
−
Power supply
R_v
Resistance temperature detector (RTD)

Figure 9.24 Wheatstone bridge

deflected, thereby indicating a change in temperature. For a given temperature, if the power supply voltage changes, then I_T will change in proportion, however I_1 and I_2 will be the same irrespective of the change in voltage. Thus, any variation in power supply does not affect the temperature reading. (R_1 and R_2 are used to prevent a short circuit if the RTD value is zero ohms.)

9.6.3 Thermocouples

The thermocouple principle is based on a **thermoelectric** effect; this is a generic expression used for temperature-dependent electrical properties of matter. Thermocouples use the potential difference that results from the difference in temperature between two junctions of dissimilar metals. This thermoelectric potential difference is called the **Seebeck effect**, named after the German physicist Thomas Seebeck (1770–1831). Two metal conductors made out of different materials are welded at each end to form **junctions**, see Fig. 9.25(a). The thermocouple junction is housed with a stainless steel tube to provide mechanical support; the wires are terminated as a junction in the thermocouple head.

When a temperature gradient between each junction is created, electrons will diffuse from the higher temperature junction to the colder junction. This creates an electric field within the conductors, which causes a current to flow, and an electromotive force (e.m.f.) to be generated. In a simple circuit, Fig. 9.25(b), the **hot junction** is exposed to an unknown temperature (T_2). The **cold junction** is contained with the indicator (measuring current); this is at a known temperature (T_1). The practical thermocouple circuit is shown in Fig. 9.25(c). The hot junction is located in the thermocouple head, or probe. Connecting cables back to the meter are either built into the probe as flying leads, or they are attached onto external connectors.

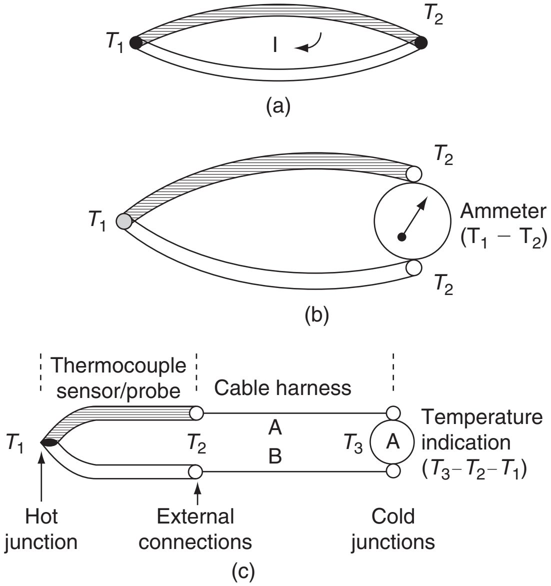
T_1
T_2
I
(a)
T_1
T_2
Ammeter
(T_1 − T_2)
T_2
(b)
Thermocouple sensor/probe
Cable harness
T_1
T_2
A
B
T_3
A
Temperature indication
(T_3-T_2-T_1)
Hot junction
External connections
Cold junctions
(c)

Figure 9.25 Thermocouple principles: (a) the Seebeck effect, (b) simple thermocouple circuit, (c) practical thermocouple circuit

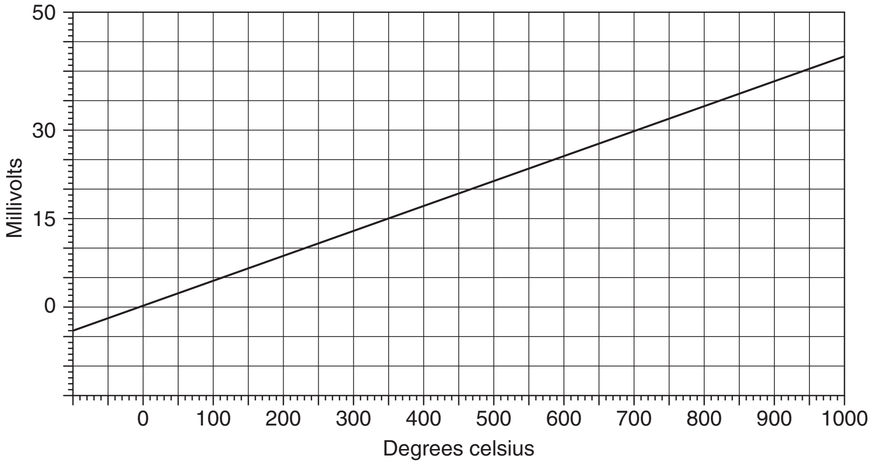

Figure 9.26 Thermocouple output (voltage/temperature)

Key maintenance point

It is important that the materials used in the harness are matched to the thermocouple type, otherwise additional junctions would be formed by dissimilar materials causing unwanted junctions and additional temperature gradients.

The relationship between the temperature gradient and EMF is linear; a typical thermocouple will generate approximately 4 mV per 100ºC of temperature difference, see Fig. 9.26.

There are many different combinations of materials used for the two thermocouple wires; the choice of materials depends mainly on the operational range of temperatures to be measured. Typical materials and upper temperatures are:

- nickel-chrome/nickel-aluminium alloys* (1000ºC)
- iron/constantan (400ºC)
- copper/constantan (300ºC)

Another term associated with thermoelectric principles is the Peltier effect, named after Jean Peltier (1785–1845). This is the opposite of the Seebeck effect; by applying a potential difference between two junctions, the temperature of the materials can be controlled.

Test your understanding 9.3

Explain the basic principles of bi-metallic switches, RTDs and thermocouples.

Key point

In an ideal thermocouple circuit, the cold junction should be maintained at constant temperature to ensure that the temperature difference is being accurately measured.

9.7 Strain transducers

Strain is defined as the deformation of a material caused by the action of **stress**. Defined as force per unit area, stress is a measure of the intensity of internal forces acting within a body. **Strain** is determined by a change that takes place between two material states, from the initial state to the final state. The physical displacement of two points in this material between these two states is used to express the numerical value of strain. A **strain gauge** is a device used to measure the deformation of a material when forces act upon it; the principles of a metallic foil strain gauge are illustrated in Fig. 9.27. Edward E. Simmons, an electrical engineer from the USA, developed the principle of the metallic foil type in 1938.

---

*These are often referred to as chromel and alumel; tradenames of the Hoskins Manufacturing Company. Further details of thermocouple applications are given in Chapter 10 (engine systems).

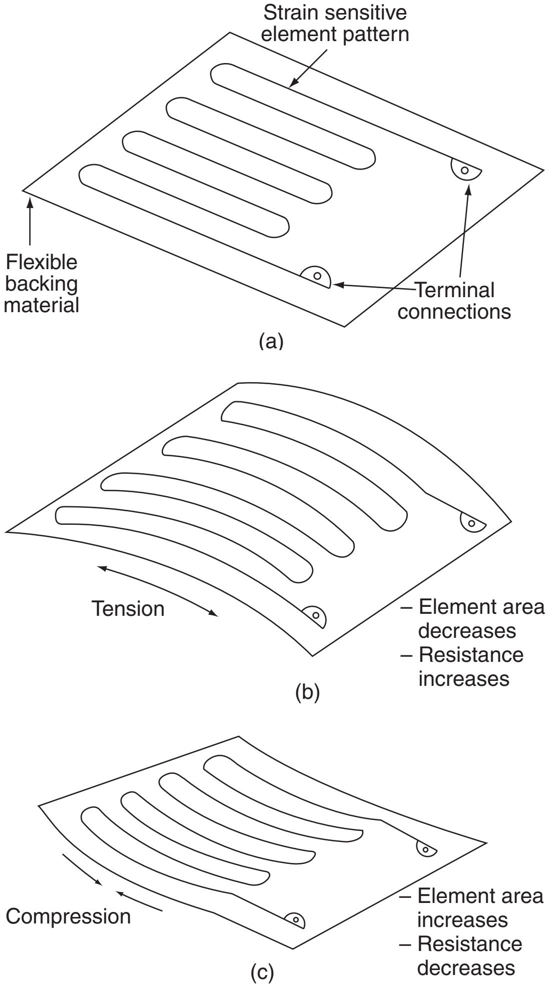

Figure 9.27 Strain gauge principles (a) no strain, (b) tension, (c) compression

The **piezo-resistive** strain gauge is based on the material's change of electrical resistance when subjected to mechanical stress. This effect was first discovered in 1856 by Lord Kelvin (1824–1907), a British mathematician, physicist and engineer. The principle was subsequently developed with silicon and germanium semiconductors during the 1950s.

Strain gauges have a number of applications on aircraft, e.g. the measurement of torque in an engine drive shaft. The metallic foil gauge is attached to the surface of the shaft, either by a suitable adhesive (metallic foil strain gauge) or it can be fabricated as an integral section of the tube surface. As the shaft is twisted by engine power, the strain gauge is deformed, causing its electrical resistance to change. Applications of this principle are discussed further in Chapter 10 (engine systems).

9.8 Rotary position transducers

There are many examples on an aircraft that require the measure of angular position; for example the position of shafts that drive control surfaces. By measuring the angle of the shaft mechanism that drives a control surface, and indication can be provided to the flight crew, or used in a computer. The type of power supply available (DC or AC) determines the main types of **synchro** systems used on aircraft.

9.8.1 DC synchro

The **DC synchro** consists of a transmitter and receiver as illustrated in Fig. 9.28. The transmitter is wire-wound circular resistance with three pick offs. The input shaft (of the device being monitored) is attached to the two contacts that slide across the resistance windings. These contacts are insulated from each other and connected to the power supply. The receiver contains a permanent magnet attached to a shaft; the magnet is positioned within three stator windings. DC synchros are known by various trade names including **Selsyn** and **Desyn**.

With the two contacts in the positions shown in Fig. 9.28, the power supply current enters the resistor at point A, current flows into the top coil (A) of the receiver and splits into the lower coils (B and C). This current sets up three magnetic fields around each of the three coils; the permanent magnet in the receiver takes up a position that aligns with the resulting magnetic field. A pointer attached to the receiver shaft moves to a position inside the indicator. When the input shaft of the transmitter is rotated, the contacts move to a new position. This alters the balance of currents through the resistance windings resulting in a change of currents in the receiver coils. The resulting magnetic flux rotates and the permanent magnet aligns to the new direction, moving the pointer to a different position.

Key maintenance point

The DC synchro suffers from contact wear on the resistance windings; this can lead to spurious operation.

9.8.2 AC torque synchro

The AC **torque synchro** also consists of a transmitter and receiver. These are similar in form; stator and rotor windings are illustrated in Fig. 9.29. Interconnections are shown in Fig. 9.30. The rotors consist of an iron core, single winding, slip-rings and brushes. Stators

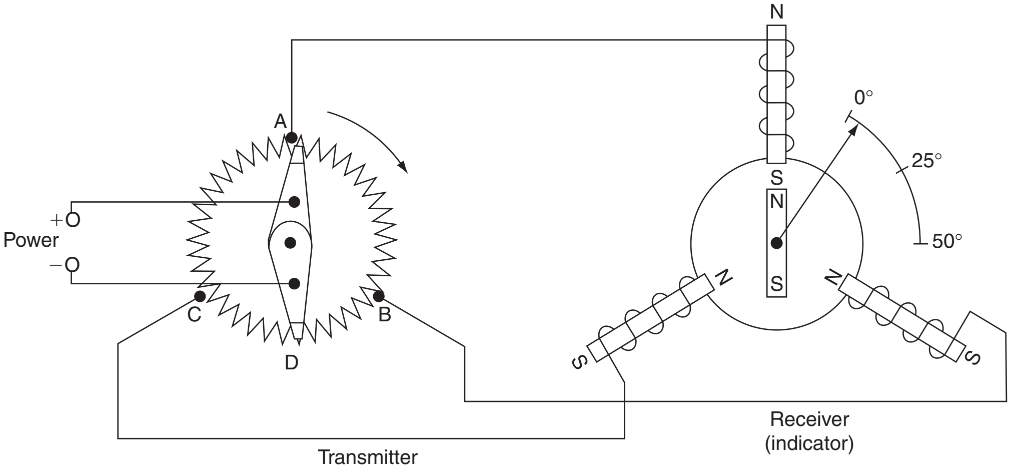

Figure 9.28 DC synchro (electrical schematic)

consist of three sets of windings each at 120 degrees to each other. The rotors of the transmitter and receiver are supplied from same power supply (normally 26V AC); this alternating supply flows through both the rotor coils via **slip-rings**. Current flowing through the transmitter's rotor creates an alternating magnetic field; this induces currents into each of its stator windings.

Key point

Torque synchros have the advantage of no contact wear on resistance windings.

Referring to Fig. 9.31, the current from each of the stators flows out of the transmitter and into the receiver where it creates magnetic fields in each of the stator windings. The receiver's rotor will be aligned in angular position with the transmitter's rotor. Fig. 9.31(a) shows the system in correspondence. When the transmitter rotor is moved, this unbalances the induced voltages in the receiver until the receiver's rotor is aligned with the resulting magnetic field.

The system is then out of correspondence as shown in Fig. 9.31(b). When the rotor of the receiver aligns itself with the field again, Fig. 9.31(c), the system is back in correspondence.

Key point

Torque synchros are sometimes referred to by the trade name of **Autosyn**.

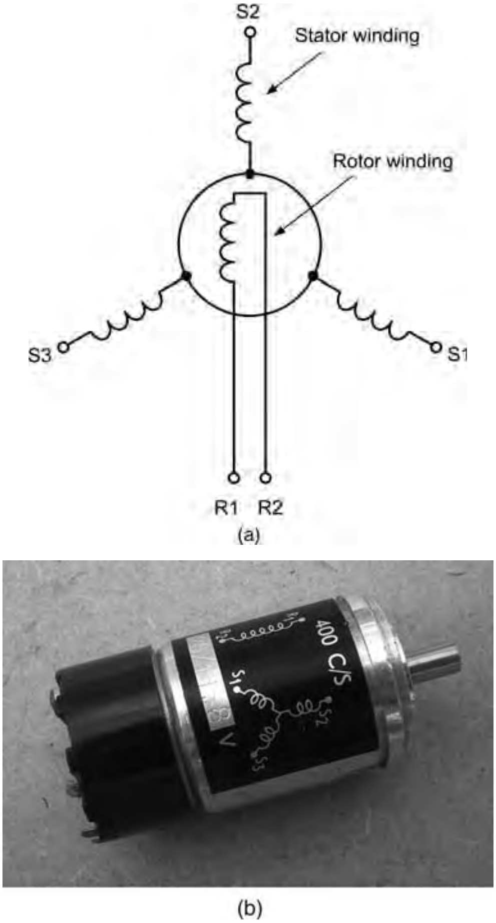

Figure 9.29 AC torque synchro system: (a) transmitter/receiver schematic, (b) transmitter/receiver hardware

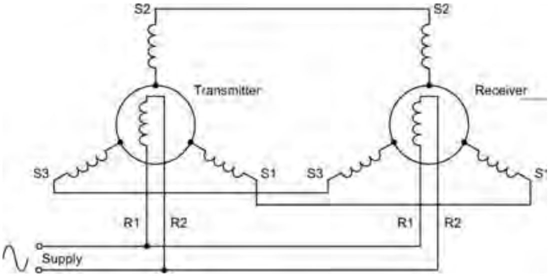
S2
S2
Transmitter
Receiver
S3
S1
S3
S1
R1
R2
R1
R2
Supply

Figure 9.30 AC torque synchro system schematic

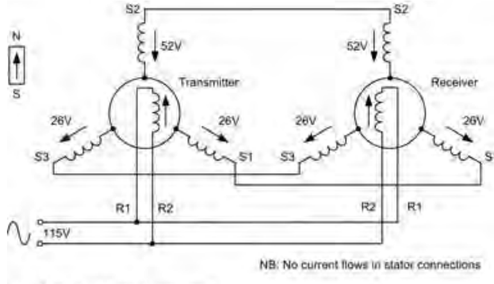
N
S
S2
S2
52V
52V
Transmitter
Receiver
26V
26V
26V
26V
S3
S1
S3
S1
R1
R2
R2
R1
115V
NB: No current flows in stator connections

(a) System in correspondence

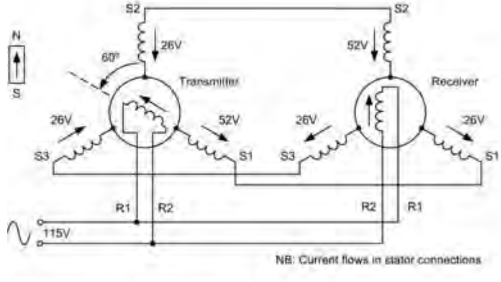
N
S
S2
S2
60°
26V
52V
Transmitter
Receiver
26V
52V
26V
26V
S3
S1
S3
S1
R1
R2
R2
R1
115V
NB: Current flows in stator connections

(b) System out of correspondence

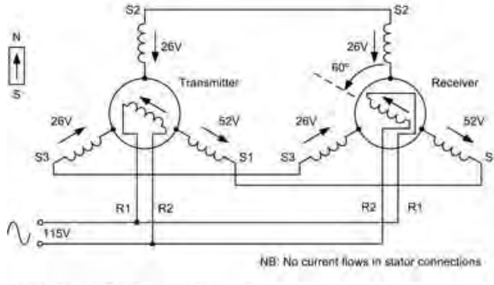
N
S
S2
S2
26V
26V
60°
Transmitter
Receiver
26V
52V
26V
52V
S3
S1
S3
S1
R1
R2
R2
R1
115V
NB: No current flows in stator connections

(c) System again in correspondence

Figure 9.31 AC torque synchro system operation

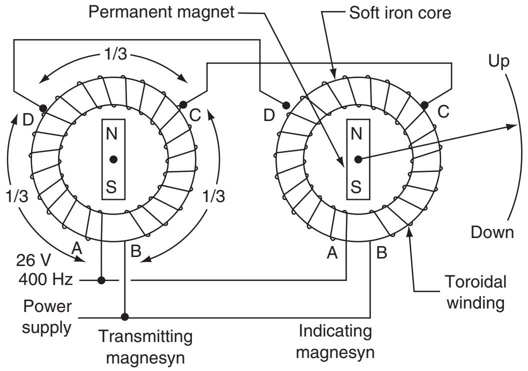
Permanent magnet
Soft iron core
Up
1/3
D
C
D
C
N
S
N
S
1/3
1/3
Down
26 V
400 Hz
A
B
A
B
Toroidal winding
Power supply
Transmitting magnesyn
Indicating magnesyn

Figure 9.32 AC Magnesyn synchro

9.8.3 Magnesyn

An alternative AC synchro is known by the trade name of **Magnesyn**. The transmitter and receiver are similar in construction; see Fig. 9.32. The rotors are permanent magnets, the stators are soft iron cores with toroidal windings. The transmitter rotor's permanent magnetic field is superimposed onto the stator field (generated from the power supply).

When the input shaft is rotated to a given position, the resulting magnetic field rotates and this causes the receiver's rotor shaft to take up the same position as the input shaft.

9.9 Multiple choice questions

1. The number of circuits that can be linked by a single switch operation is called the:
   (a) pole
   (b) throw
   (c) NO/NC contacts.
2. The bi-metallic temperature sensor consists of:
   (a) two strips of metal with different coefficients of thermal expansion
   (b) two strips of metal with the same coefficients of thermal expansion
   (c) two wires welded into a junction.
3. When a foil strain gauge is deformed, this causes:
   (a) an electromotive force (EMF) to be generated
   (b) its electrical resistance to change
   (c) different coefficients of thermal expansion.
4. Metal elements used in RTDs have a temperature coefficient that is:
   (a) negative, temperature increases cause an increase in resistance

(b) positive, temperature decreases cause an increase in resistance
(c) positive, temperature increases cause an increase in resistance.

5. The rotors of an torque synchro transmitter and receiver are supplied from:
   (a) the same power supply (normally 26V AC)
   (b) different power supplies (normally 26V AC)
   (c) the same power supply (normally 26V DC).

6. The thermocouple principle is based on the Seebeck effect; when heat is applied:
   (a) a change of resistance is measured
   (b) this causes the element to bend
   (c) an electromotive force (e.m.f.) is generated.

7. A rheostat performs the same function as a:
   (a) resistance temperature detector
   (b) potentiometer
   (c) Wheatstone bridge.

8. The solenoid is a type of transducer that converts:
   (a) electrical energy into linear motion
   (b) linear motion into electrical energy
   (c) electrical energy into thermal energy.

9. Proximity switches perform the same function as:
   (a) micro-switches
   (b) relays
   (c) toggle switches.

10. The linear variable differential transformer (LVDT) is used for measuring:
    (a) small rotary displacements
    (b) variable resistance
    (c) small linear displacements.

11. Referring to Fig. 9.33, these control devices are:
    (a) toggle switches
    (b) micro switches
    (c) rocker switches.

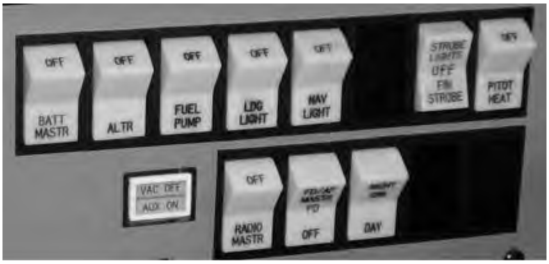
OFF
BATT MASTR
OFF
ALTR
OFF
FUEL PUMP
OFF
LDG LIGHT
OFF
NAV LIGHT
STROBE LIGHTS
OFF
FIN STROBE
OFF
PITOT HEAT
VAC OFF
AUX ON
OFF
RADIO MASTR
FD/AP MASTER
FD
OFF
DAY

Figure 9.33

12. Referring to Fig. 9.34, this device is a:
    (a) toggle switch
    (b) micro switch
    (c) changeover relay.

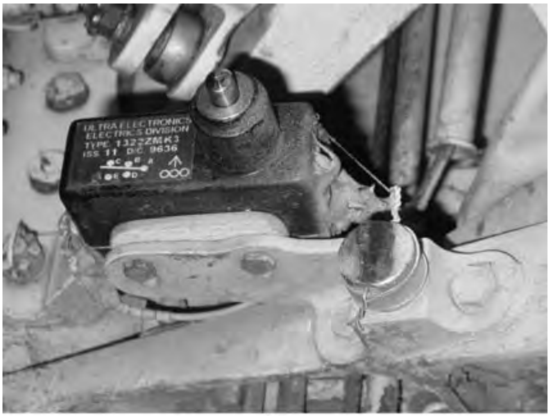
ULTRA ELECTRONICS
ELECTRICS DIVISION
TYPE 1322ZMK3
ISS 11 D/C 9636

Figure 9.34

13. Referring to Fig. 9.35, this circuit is called a:
    (a) wheatstone bridge
    (b) thermocouple bridge
    (c) ratiometer.

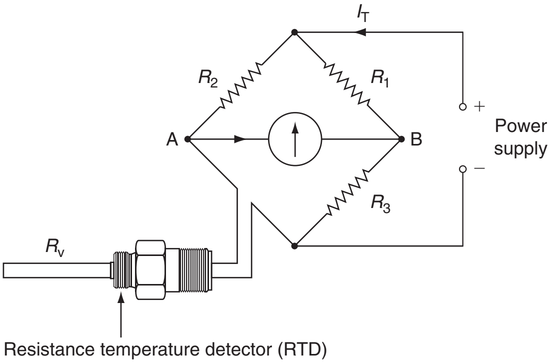
I_T
R_2
R_1
A
B
R_3
+
Power supply
−
R_V
Resistance temperature detector (RTD)

Figure 9.35

14. Referring to Fig. 9.36, this schematic is for a:
    (a) AC magnesyn synchro
    (b) AC torque synchro
    (c) DC synchro.

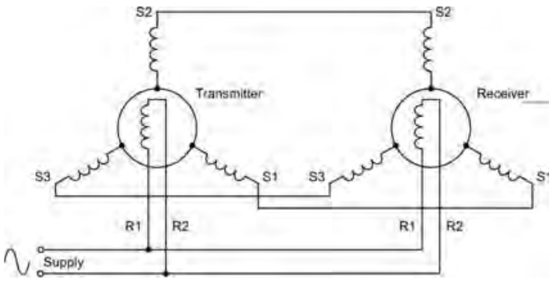
S2
S2
Transmitter
Receiver
S3
S1
S3
S1
R1
R2
R1
R2
Supply

Figure 9.36

15. Referring to Fig. 9.37, this chart shows the output of a:
    (a) RTD
    (b) thermocouple
    (c) strain gauge.

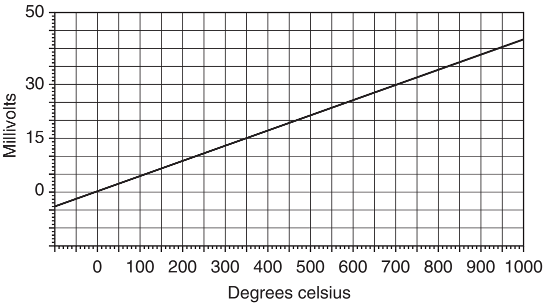

Figure 9.37

Chapter 10

Engine systems

The aircraft engine is installed with many systems requiring electrical power. The predominant requirement (in terms of current consumption) is for the starting system. General aviation aircraft use electrical starter motors for both piston and gas turbine engines; larger transport aircraft use an air-start system (controlled electrically) derived from ground support equipment or by air cross-fed from another engine. Electrical starting systems on piston and gas turbine engines are very different. The trend towards the all-electric aircraft will see more aircraft types using electrical starting methods. The engine also requires electrical power for the ignition system. Once again, the needs of piston and gas turbine engines are quite different. Although starting and ignition systems are described in this chapter as separate systems, they are both required on a co-ordinated basis, i.e. a means to rotate the engine and ignite the air/fuel mixture.

Electrical and electronic requirements for engines also include the variety of indicating systems required to operate and manage the engine. These indicating systems include (but are not limited to) the measurement and indication of: rotational speed, thrust, torque, temperature, fuel flow and oil pressure. Indications can be provided by individual indicators or by electronic displays. This chapter describes engine starting, ignition and indicating system for both piston and gas turbine engines.

10.1 Starting and ignition

10.1.1 Piston engines

The original way of starting piston (internal combustion) engines was by ‘swinging’ the propeller; this involves using the propeller as a lever to turn the engine shaft. The method in widespread use on most engines now is a starter motor powered by the aircraft battery. Basic electrical starting systems comprise a series or compound wound motor with engaging mechanism.

The simplest method of physically connecting the motor to the reciprocating engine is via a pinion on the motor that engages with a gear ring attached to the crankshaft; this mechanism disconnects after the engine has started. This pinion and gear ring provides a gear ratio in the order of 100:1 to turn the engine at sufficient speed to overcome compression and bearing friction.

Referring to Fig. 10.1, the battery master switch is selected on; this energizes the battery relay and power is applied to the busbar and starter relay. When the starter switch is closed, this energizes the starter relay and applies power to the starter motor. As soon as the engine fires and starts, the starter switch is released and power is removed from the motor; this opens the starter relay contacts.

10.1.2 Twin-engine (piston) starting system

The twin-piston engine aircraft is started in a similar way to the single engine. Referring to Fig. 10.2, power is applied from the battery to the battery relays. When the battery master switch is turned on, power is available at both the starter relay. Power is also made available at the busbar and the starter circuit breaker is fed to the dual starter switch. Each engine is started by its own set of contacts. When an engine has been started the switch is released and the spring-loaded contacts return it to the centre-off position.

10.1.2.1 Magneto ignition (high-tension type)

Ignition energy for piston engines is generated from a **magneto**; this provides pulses of electrical power via a **distributor** to spark plugs in each of the engine cylinders. The magneto operates on the principle of electromagnet induction (Fig. 10.3); it is a combined four-pole permanent magnet generator and autotransformer and can be used where there is no aircraft battery. The engine drives the input shaft of the magneto

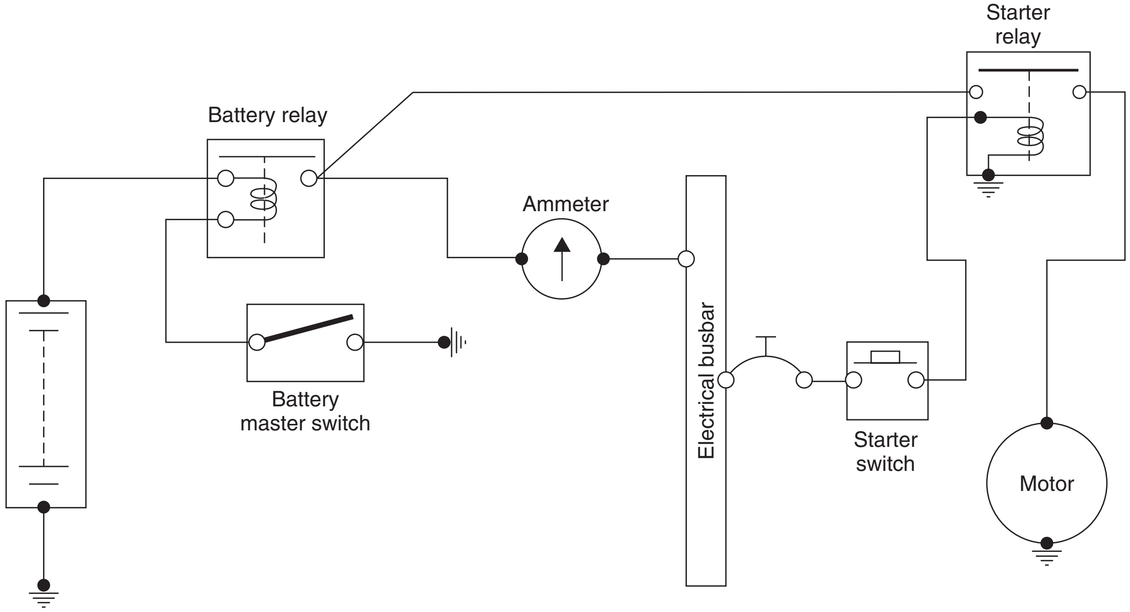

Figure 10.1 Electrical starting system

rotor via a gearbox; the relative movement of transformer windings and the poles of a permanent magnet can be arranged in one of three ways:

1. The transformer coils are on the shaft and the magnet is fixed to the housing (rotating armature type)
2. The permanent magnet is rotated by the shaft within stationary coils of the transformer (rotating magnet type)
3. A soft iron inductor is rotated between the permanent magnet and transformer windings (polar inductor type).

For every revolution of the shaft, a cam opens the contact breaker, interrupting the primary current; this causes the electromagnetic field in the primary coil winding to collapse. As the field collapses there is a voltage generated across the **primary coil**. A capacitor connected across the contacts discharges when the breaker contacts are closed and charges when they open. When the capacitor discharges, a high current flows through the primary coil, inducing high secondary voltages. The capacitor also prevents arcing across the breaker contacts, and determines the voltage across the primary coil thereby controlling the rate at which the electrical energy dissipates through the primary coil. The magneto's output is directed to the spark plugs via a distributor. The distributor shaft is connected via gears to the magneto shaft; this ensures that energy is applied to the spark plugs with the correct timing.

In the aircraft engine, each cylinder normally has two spark plugs, each driven from a separate magneto. This arrangement provides redundancy in the event of failure of one of the magnetos. Two sparks per cylinder also provides a more complete and efficient burn of the fuel/air mixture. The magneto's simplicity and self-contained design provides reliability as well as light weight. An on/off switch controls the system; in the off position, the primary winding is connected to ground, and this prevents current from being induced in the primary windings.

10.1.2.2 Magneto ignition (low-tension type)

A larger piston engine has more cylinders and is designed to operate at higher altitudes. Decreased atmospheric pressure means that the high-tension magneto system is prone to electrical insulation breakdown in the distribution cables. The low-tension magneto system is based on the **polar inductor** method. The output from the magneto is a low voltage; this is increased to a high voltage by secondary transformers located near the plugs. This reduces the length of high-tension cable and reduces the risk of insulation breakdown in the distribution cables.

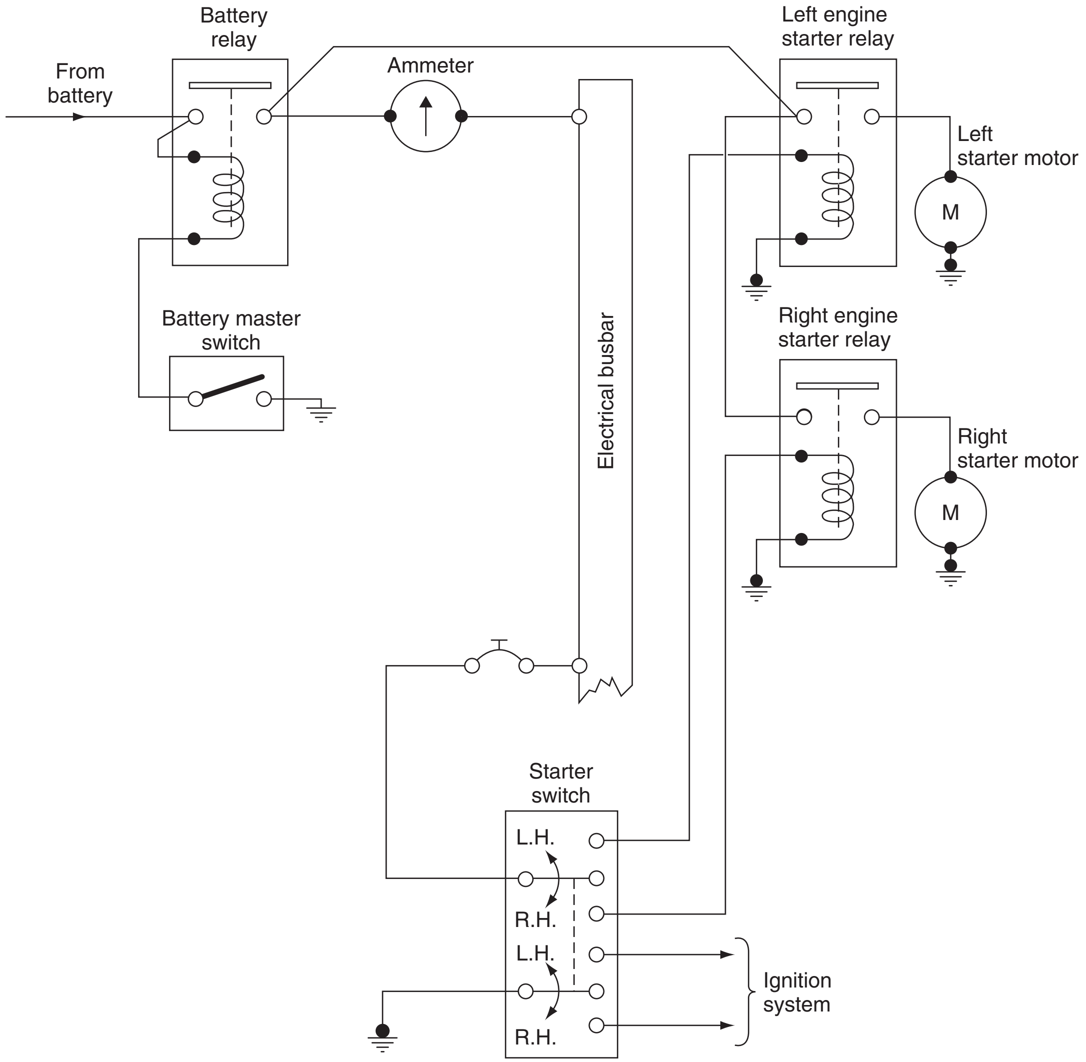
Battery relay
From battery
Ammeter
Left engine starter relay
Left starter motor
M
Battery master switch
Electrical busbar
Right engine starter relay
Right starter motor
M
Starter switch
L.H.
R.H.
L.H.
R.H.
Ignition system

Figure 10.2 Twin-piston engine aircraft starting

Brushes and commutators form the low-tension magneto distributor. **Ignition cables** carry the high energy from the magneto to the plugs. The single core of stranded conductors is insulated by a substantial thickness of material.

10.1.2.3 Spark plugs

These conduct the high-energy output from the magneto across an air gap, see Fig. 10.4. The fuel/air mixture across the gap is an insulator; as the voltage from the magneto increases, it alters the structure of the fuel/air mixture between the electrodes. Once the voltage exceeds the dielectric strength of the fuel/air mixture, it becomes ionized. The ionized fuel/air mixture then becomes a conductor and allows electrons to flow across the gap. When the electrons flow across the gap, this raises the local temperature to approximately 60,000 K. The electrical energy is discharged as heat and light across the air gap, thereby appearing as a spark with an audible 'clicking' sound. This energy ignites the fuel mixture in the cylinder.

The plug is fitted into the cylinder head's combustion chamber and is therefore exposed to high pressures and temperatures; plugs have to operate with

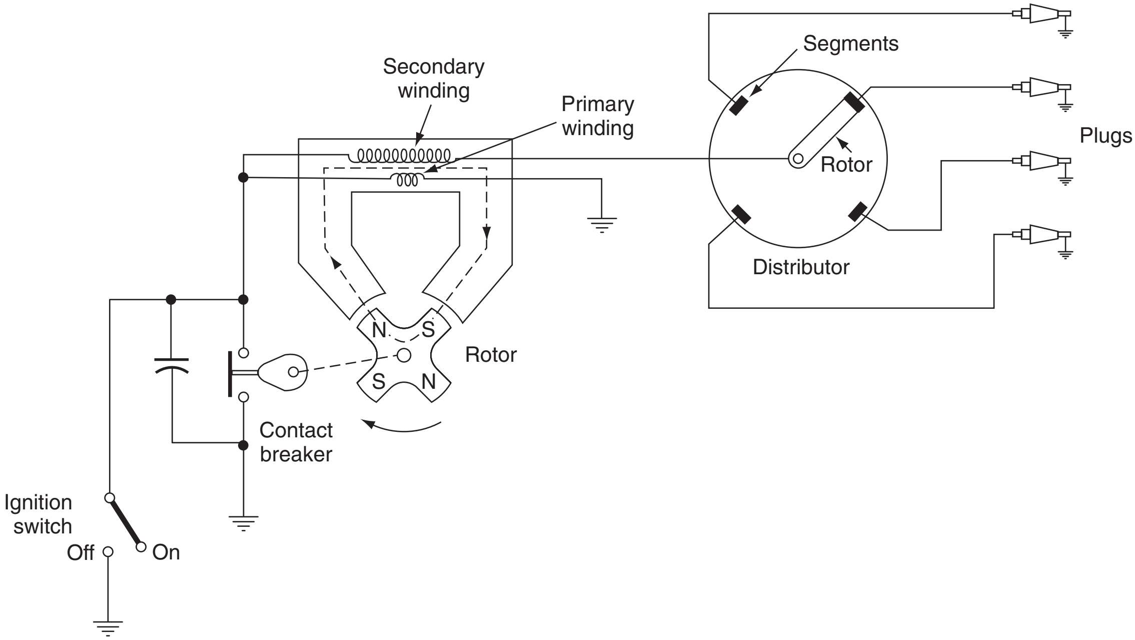
Secondary winding
Primary winding
Segments
Rotor
Plugs
Distributor
N
S
S
N
Rotor
Contact breaker
Ignition switch
Off
On

Figure 10.3 Magneto ignition principles

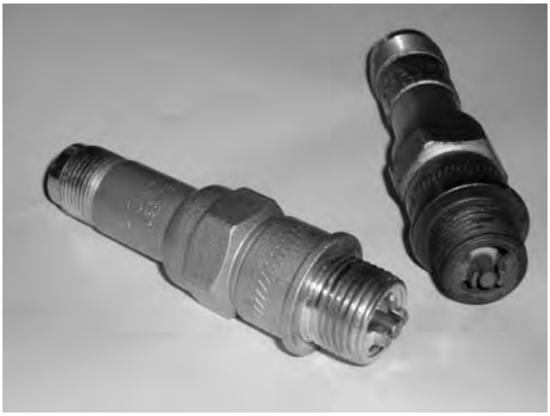

Figure 10.4 Spark plugs (new and used)

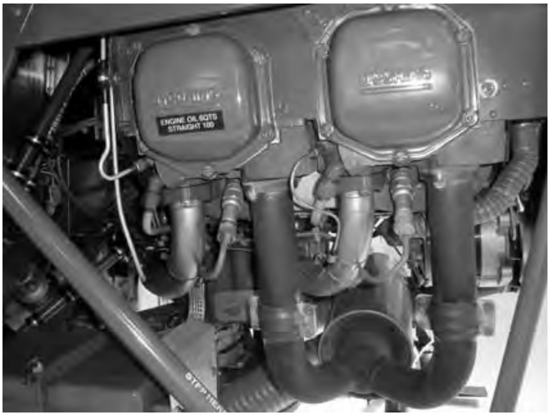
ENGINE OIL 6QTS
STRAIGHT 100

Figure 10.5 Spark plug installation

minimal deterioration over long periods of time in this harsh environment. The outer shell of the plug is made from high tensile steel with close tolerance threads to locate the plug in the cylinder head; a copper crush-washer completes the seal against high gas pressures. The outer shell is connected electrically to the cylinder head through its body. The centre electrode carries the high tension energy to the spark gap; this is formed from a material that is resistant to the repetitive arcing, typically nickel, platinum or iridium. An insulator separates the inner and outer sections of the plug, typical materials include mica, ceramic, aluminium oxide ceramic. Examples of spark plug installations, together with ignition cables, are shown in Fig. 10.5.

Key point

Two spark plugs per cylinder provides redundancy in the event of magneto failure and an efficient burn of the fuel/air mixture.

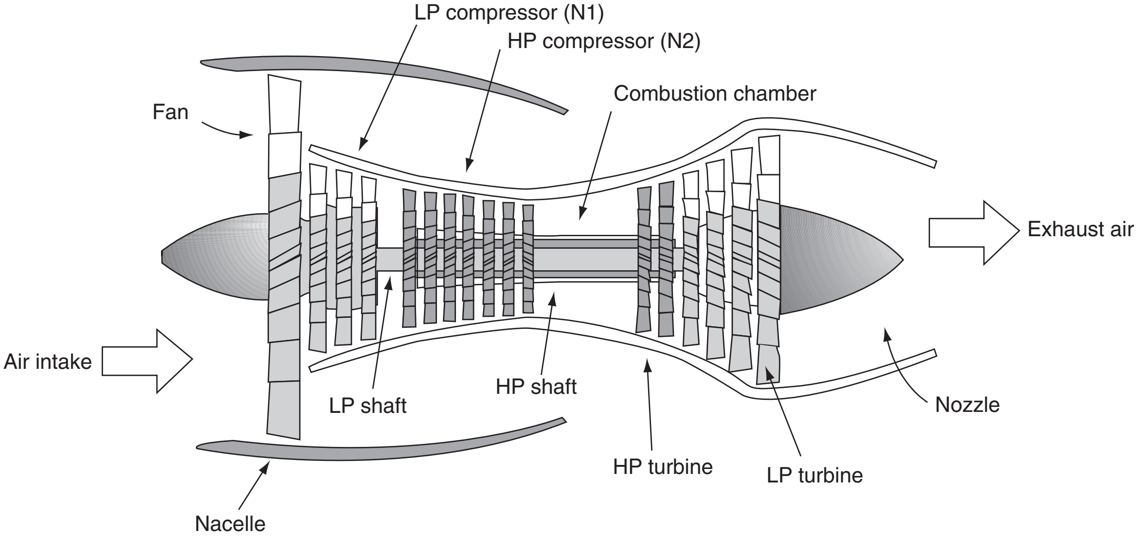

Figure 10.6 Gas turbine engine features

10.1.3 Turbine engine starting

Starting a turbine engine requires a higher-duty motor compared with piston engines; the motor has to overcome higher inertia and needs to achieve higher cranking speeds. The general features of a gas turbine engine are illustrated in Fig. 10.6. Air is compressed through a multi-stage system before entering the combustion chamber where it is mixed with fuel and ignited. The expanding exhaust gases are then directed through a turbine to produce thrust. The turbine also turns shafts to drive the compressor stages.

The starter motor has to overcome the inertia of the compressor; a large volume of air has to be drawn through engine and then accelerated and compressed until the engine's **self-sustaining speed** is reached. Self-sustaining speed is when sufficient energy is being developed by the engine to provide continuous operation without the starting device. When the turbine engine is driving a propeller (commonly called a **turboprop**) the starter has to overcome the additional inertia of the propeller (helicopter engines are normally connected to the rotor via a clutch).

A typical turbine engine starter circuit is illustrated in Fig. 10.7. Current from the busbar is connected to the coil of a starter relay via a push-button switch. This makes a circuit to a timer switch and starter motor via a limiting resistor. (The resistor limits excessive currents that would otherwise occur in overcoming the initial torque of the motor.) Once the motor speed has reached its nominal operating speed, its starting torque reduces and the timer switch operates contacts for the shorting relay. With the relay energized, current from the busbar is switched directly into the motor, i.e. bypassing the resistor. Ignition is switched on when self-sustaining speed is reached; power to the motor is removed.

Key point

The initial current through the turbine engine starter motor is in the order of 1000–1500 A, hence the need for a limiting resistor and timing circuit.

Direct-cranking electrical starting systems are similar to those used on piston engines. A typical 28 V DC four-pole motor produces 15–20 lb/ft of torque. The output shaft is clutched into a gear mechanism via the accessory gearbox. Typical duty cycles for motors in small- to medium-size engine requires speeds of 3800 r.p.m. for up to 90 seconds with current peaking at 1000 A. The motor must transmit sufficient torque to the engine's rotating assembly to provide smooth acceleration from rest up to the self-sustaining speed; at this point, the motor is disengaged.

Some aircraft are installed with a combined **starter-generator.** This involves a permanent coupling of the starter-generator shaft to the engine

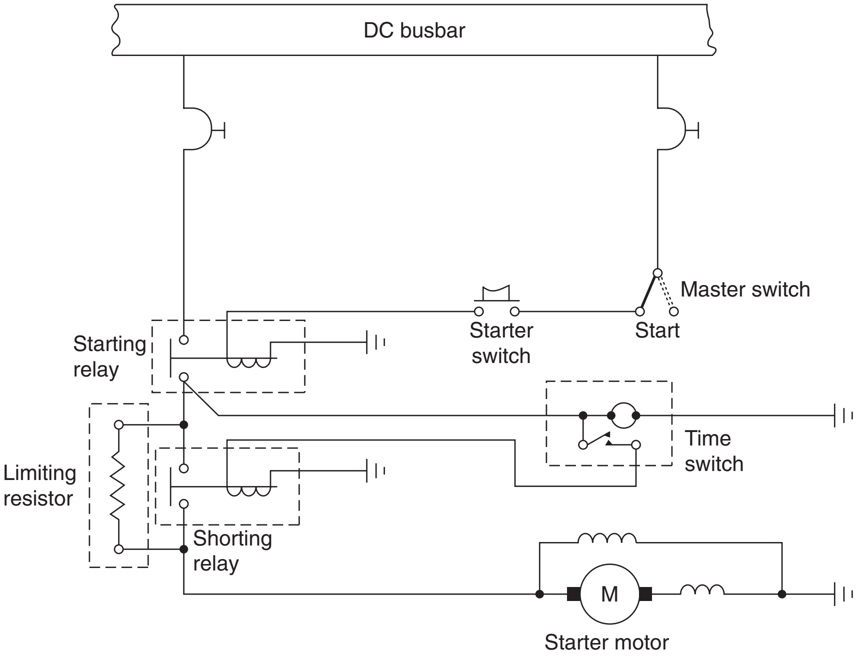

Figure 10.7 Turbine engine starter circuit

via a gearbox drive. The dual-purpose machine is compound-wound and the field is connected via a changeover relay. During engine start it acts like a conventional motor until the engine is up to speed. The changeover relay then automatically connects the field windings to a voltage regulator and it becomes a conventional generator. The starter-generator system has reduced weight and component parts compared with having a separate starter motor and generator, thereby reducing overall operating costs.

10.1.3.1 High-energy ignition unit (HEIU)

High-energy ignition is required for starting gas turbine engines; a dual system is normally installed for the main engines. The system comprises two HEIUs and two igniter plugs per engine. A typical HEIU installed on an engine is shown in Fig. 10.8. Turbine ignition systems are switched off after the engine has reached self-sustaining speed; the system is used as a precaution during certain flight conditions e.g. icing, rain or snow. **In-flight start** uses the wind-milling effect of the engine within the specified flight envelope of air speed and altitude. High voltages are required for the igniter plugs to accommodate the variations in atomized fuel over the range of atmospheric conditions. Electrical energy is stored in the HEIU and then dissipated across the igniter plug.

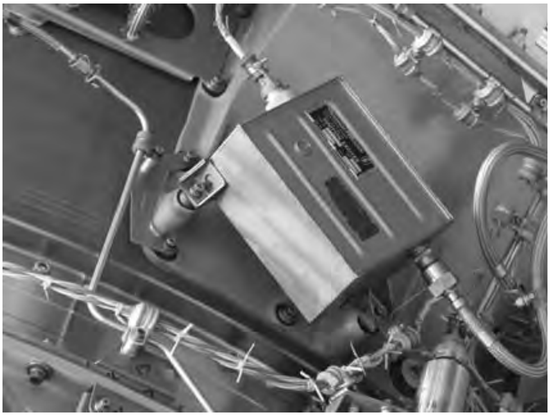

Figure 10.8 Ignition unit installation

Key maintenance point

The very high voltage output from an HEIU is potentially lethal. The HEIU can remain charged for several minutes; always refer to the maintenance manual for operating procedures.

10.1.3.2 Igniter plugs

Referring to Fig. 10.9, the end of the igniter plug is formed with the outer casing insulated from the centre electrode. The end of the igniter is coated with a semi-conductive ceramic material. The output from the HEIU heats the surface of the ceramic material and lowers its resistance. This creates a high-intensity flashover from the centre electrode to the outer casing. (Note this is a surface discharge, not a spark across an air gap.)

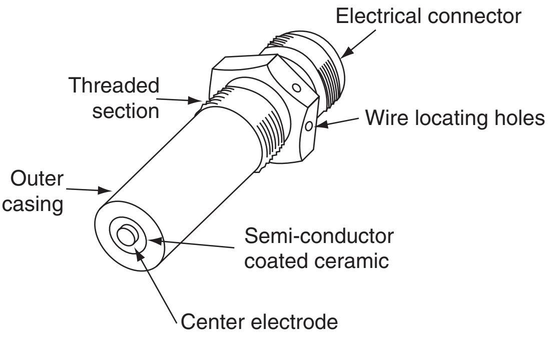
Electrical connector
Threaded section
Wire locating holes
Outer casing
Semi-conductor coated ceramic
Center electrode

Figure 10.9 Igniter plug features

10.1.4 Auxiliary power unit (APU) start and ignition

The APU on large transport aircraft is started electrically; air is then bled into the air distribution system for the main engines. The APU starter motor is engaged and disengaged automatically as part of the starting system. Referring to Fig. 10.10, a centrifugal switch connected to the APU shaft controls the start sequence via three sets of contacts that operate at 35%, 95% and 110% of maximum speed. An oil pressure switch ensures that the system cannot start until the lubrication oil pump builds up sufficient pressure.

With the system circuit breaker closed, and the start switch selected, the start and control relays are supplied via the closed 110% switch contacts; the starter motor is now connected to the electrical supply via the start relay. The starting sequence is confirmed via a green annunciator light on the control panel; the

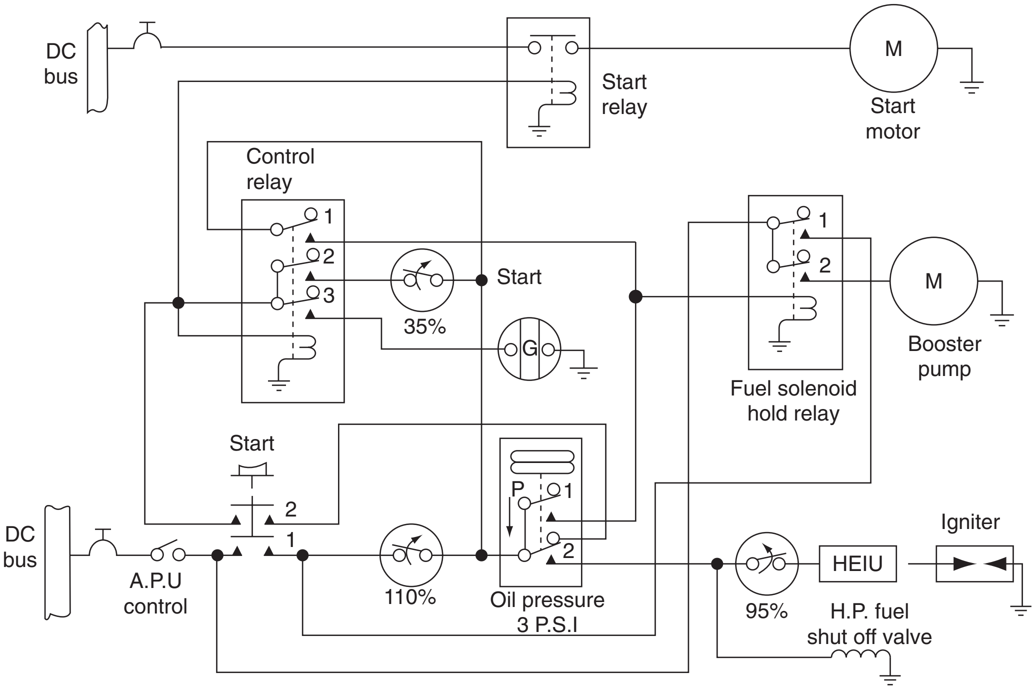
DC bus
Start relay
M
Start motor
Control relay
1
2
3
Start
35%
G
1
2
M
Booster pump
Fuel solenoid hold relay
Start
2
1
DC bus
A.P.U control
110%
P
1
2
Oil pressure 3 P.S.I
Igniter
95%
HEIU
H.P. fuel shut off valve

Figure 10.10 Auxiliary power unit (APU) start and ignition

light is connected to the power supply via contacts on the control relay. The fuel solenoid holding relay (FSHR) is operated via the control relay, and the 35% switch contacts; the low-pressure (LP) fuel pump motor is connected to the power supply via the FHSR. The start switch is now released; the starter circuits are maintained via the FHSR, 100/35% switch contacts and control relay. When the oil lube pump builds up sufficient pressure, this closes a switch that provides a retaining supply for the FSHR. The high-pressure fuel shut-off valve is then energized open (allowing fuel to be delivered under pressure into the APU) and the ignition system is supplied via the 95% contacts. The APU continues to accelerate; at 35% engine speed, the start and control relays open; the starter motor is disengaged, and the 'start' light is switched off. At 95% of maximum speed, the contacts open and the ignition is switched off. The APU start sequence is now completed, and the engine runs constantly at 100% r.p.m. (There is no throttle control on the APU.)

The APU can be shut down either manually (by selecting the APU to off) or automatically. Loss of oil pressure at any time will automatically shut the APU down. An overspeed condition (sensed by the 110% switch opening its contacts) will de-energize the fuel solenoid holding relay, close the high-pressure fuel shut-off valve and remove power from the low-pressure booster pump.

10.1.5 Main engine start

The main engines are normally started via air-driven motors; there are three sources of air for starting the main engines:

- APU
- ground air supply cart
- another engine.

A typical air distribution system is illustrated in Fig. 10.11.Valves, controlled either manually or automatically, are operated by motors. With the APU started and running at normal speed, a switch on the start control panel opens the APU bleed valve. Air is directed through the isolation valve and engine bleed valve to the engine start valve.

When using the ground air supply cart, an external connection is made and air is directed through the engine bleed valve to the start valve, or through the isolation valve and engine bleed valve to the start valve. When using another engine (that is already running), air is supplied from its bleed valve, through the isolation and bleed valves of the engine to be started and through to the start valve.

For illustration purposes, a twin-engined aircraft starting and ignition system is described; refer to Fig. 10.12. There is a combined start and ignition control panel located in the overhead panel; in the illustration, this is fitted with a rotary switch for each engine. The operation and functions of this switch are identical for each engine. The switch has to be pushed in before any selections can be made; this is to prevent accidental movement of the switch. Selecting ground (GRD) connects 28V DC to energize the start switch holding coil.

The circuit is completed through the cut-out contacts in the engine starter valve. The start switch is now held in the GRD position and the ground start sequence is initiated. The 28 V DC supply also energizes the start valve solenoid and this opens the valve, supplying air to drive a small turbine in the starter motor. The turbine connects through an accessory gearbox onto the engine's HP compressor shaft.

At approximately 16% of maximum rotational speed, the start lever is moved from the cut-off position to 'idle'. This applies 28V DC through a second pair of contacts of the start switch and ignition switch to supply the HEIU. Each igniter plug discharges at a high level, typically 20 joules of energy, at 60–90 discharges per minute. (This can be heard outside the engine as an audible 'clicking' sound.) At a predetermined cut-out speed, the centrifugal switch in the starter motor opens: the start switch is de-energized and returns (under spring force) to the off position. The 28V DC power supply is removed from the HEIU and the start valve motor drives to its closed position. The engine continues to accelerate to the **ground idle** speed; this is slightly above self-sustaining speed and occurs when the engine has stabilized. For a twin-shaft axial flow engine, ground idle is typically 60% of the high-pressure (HP) compressor speed.

Low-energy ignition (typically 4 joules of energy, at 30 discharges per minute) is used in certain phases of flight including take-off, turbulence and landing. Furthermore, if the aircraft is flying through clouds, rain or snow, continuous low-energy ignition is selected on the control panel. This closes a contact on the rotary switch and applies power to a second HEIU input.

In the event of an engine **flameout** during flight, the crew will attempt an in-flight start of the engine; this requires a modified procedure to that of the ground start. The engine will be wind-milling due to the forward speed of the aircraft. The starter valve

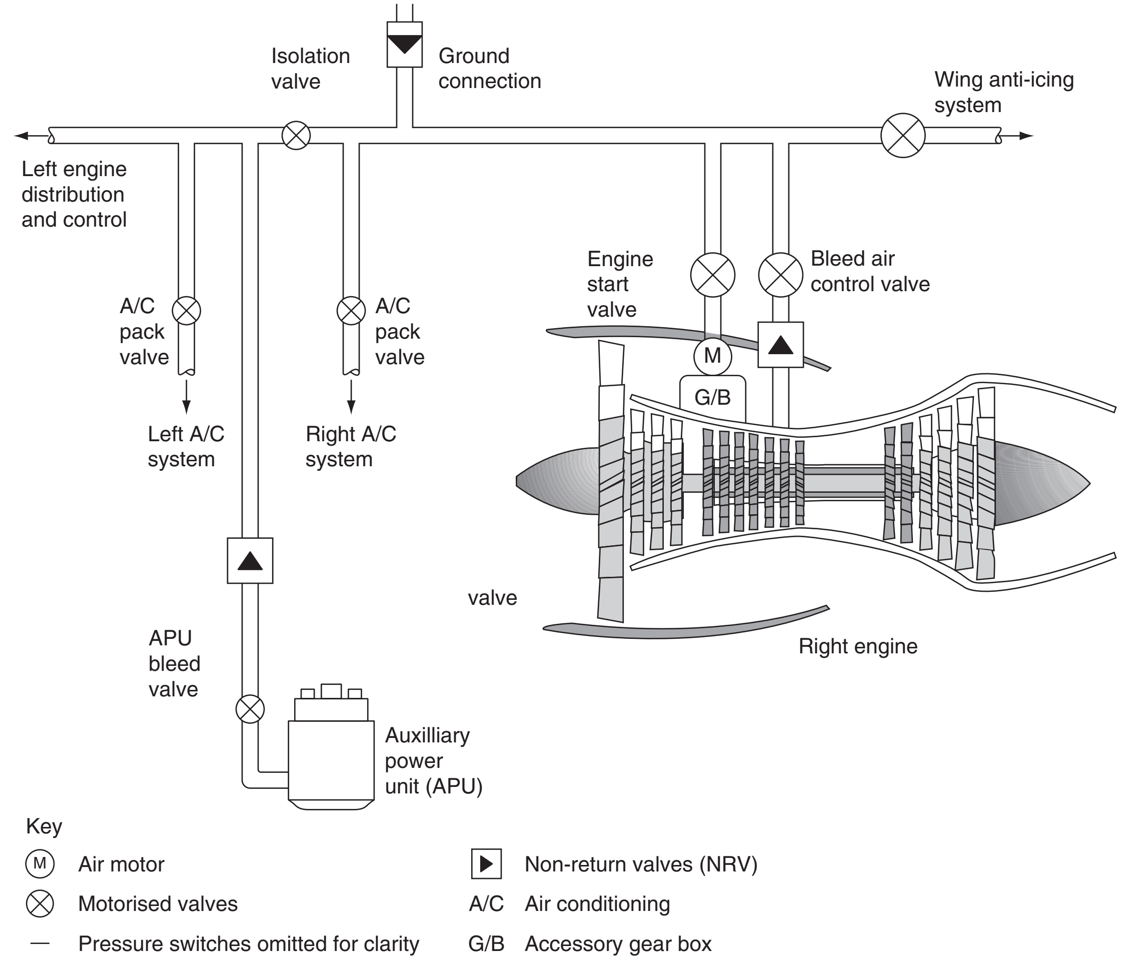
Isolation valve
Ground connection
Wing anti-icing system
Left engine distribution and control
Engine start valve
Bleed air control valve
A/C pack valve
A/C pack valve
M
G/B
Left A/C system
Right A/C system
valve
APU bleed valve
Right engine
Auxilliary power unit (APU)
Key
M Air motor
Motorised valves
— Pressure switches omitted for clarity
Non-return valves (NRV)
A/C Air conditioning
G/B Accessory gear box

Figure 10.11 Typical air distribution system

and motor are not selected as with the ground start. Low ignition (LOW IGN) and flight (FLT) are manually selected on the control panel until the engine reaches **flight idle** speed. In-flight restarts can only be attempted within certain airspeed and altitude limits.

Key maintenance point

The starting sequence for a gas turbine engine is to: (i) develop sufficient airflow to compress the air, (ii) turn on the ignition, and (iii) open the fuel valves. This sequence is critical since there must be sufficient airflow through the engine to support combustion before the fuel/air mixture is ignited.

Key maintenance point

Facing the aircraft into the wind augments gas turbine engine starting; this assists with engine acceleration, particularly for turbo-prop engines. The propellers are normally designed with a fine-blade angle for starting and ground running.

Gas turbine engines sometimes suffer from a starting problem that results in fuel entering the combustion chamber, but no ignition; this is sometimes referred to as a **wet start**. Engine indications would be the engine turning at the correct starter speed, with indications of fuel flow, but no increase in exhaust gas temperature

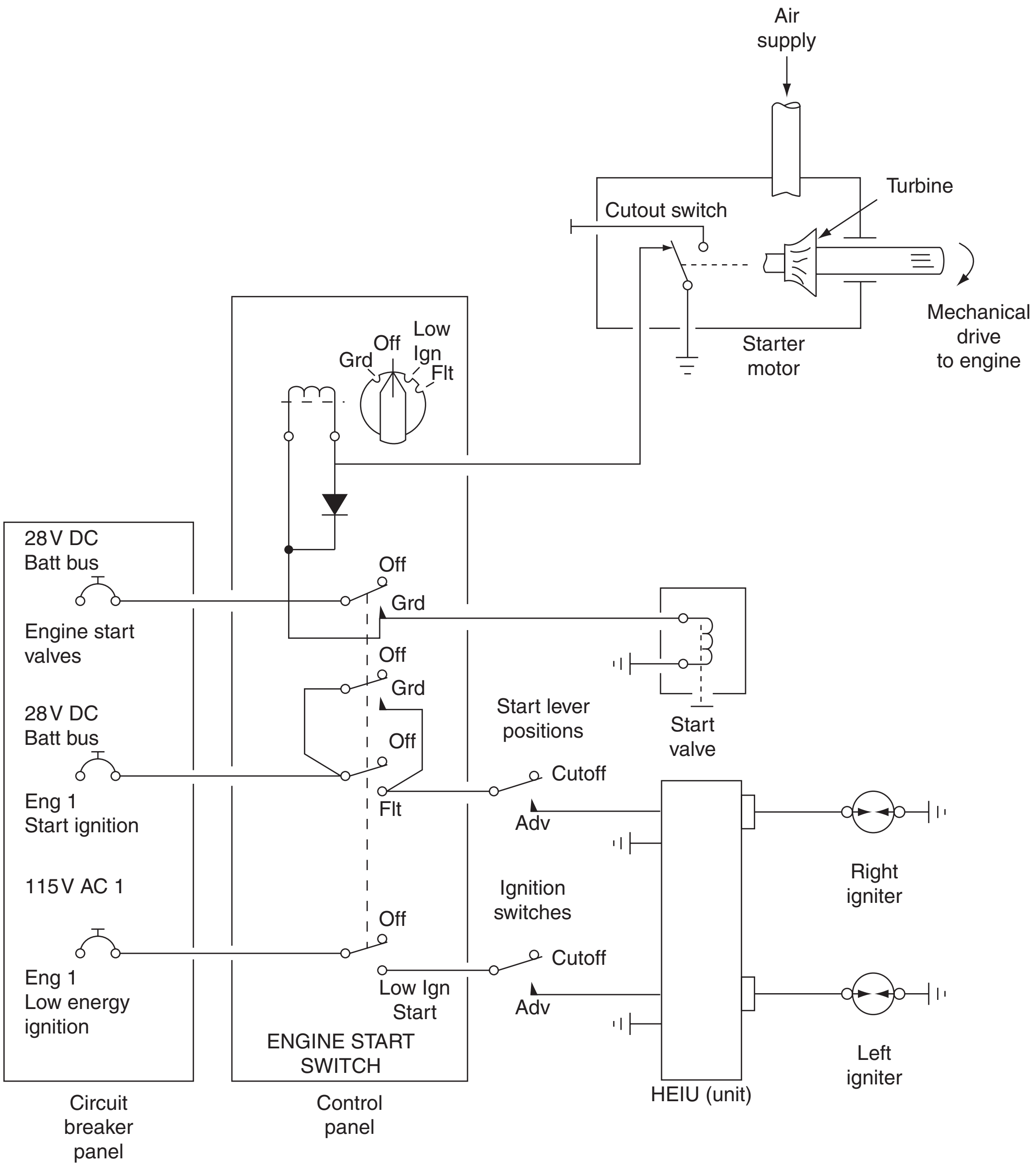
Air supply
Turbine
Cutout switch
Starter motor
Mechanical drive to engine
Grd
Off
Low Ign
Flt
28 V DC Batt bus
Engine start valves
28 V DC Batt bus
Eng 1 Start ignition
115 V AC 1
Eng 1 Low energy ignition
Off
Grd
Off
Grd
Off
Flt
Off
Low Ign Start
ENGINE START SWITCH
Start lever positions
Cutoff
Adv
Ignition switches
Cutoff
Adv
Start valve
HEIU (unit)
Right igniter
Left igniter
Circuit breaker panel
Control panel

Figure 10.12 Turbine starting and ignition system

(EGT). Observers outside the aircraft could see atomized fuel or vapour from the engine exhaust.

Test your understanding 10.1

Explain the term 'self sustaining speed'.

The cause of a wet start is most likely to be a defective HEIU and/or igniter plug. The net result is no ignition in the combustion chamber and the accumulation of fuel. If compressor outlet air gets hot enough, it could ignite the fuel, causing a rapid expansion of the fuel/air mix (effectively an explosion) that could lead damage of the turbine section and eject flames from the engine exhaust. The procedure is to shut off the fuel supply to the engine and continue turning over the engine with the starter motor to clear (or blow out) out the fuel. Some starter panels have a selectable **blow out** position to achieve this procedure.

Test your understanding 10.2

What is the difference between a flameout and a wet start?

10.2 Indicating systems overview

Engine indications can be broadly divided into primary and secondary systems. Some indication systems are unique to gas turbine, turboprop or piston engines, some are common to all types. Primary indicators include:

- speed
- temperature
- thrust
- fuel flow.

Secondary indicators include (but are not limited to):

- oil temperature
- oil quantity
- oil pressure
- vibration.

Measurements are made by a variety of **transducers**; these are devices used to convert the desired parameter, e.g. pressure, temperature, displacement etc. into electrical energy. The locations of engine instruments is normally between the two pilot's panels, see Fig. 10.13.

10.3 Primary indicating systems

10.3.1 Engine speed

This is a primary engine indication used on both piston and gas turbine engines. It is one of two methods used to indicate thrust on gas turbine engines (the other being EPR). In gas turbine engines, the usual practice is to display a **percentage** of maximum revolutions per minute (r.p.m.) rather than actual r.p.m. Typical gas turbine engine speeds are in the order of 8000–12,000 r.p.m. The mathematical notation for rotational speed is N; gas turbine engines can have up to three shafts (or spools), these are referred to as low, intermediate and high-pressure shafts (**LP, IP and HP**). Alternatively, they are referred to numerically; the individual shaft speeds are therefore N_1, N_2 **and** N_3. Engine speed is monitored by the crew at all times; particularly during start and take-off to make sure that engine limits are not exceeded. The two principal types of engine speed transducer are the tachometer and variable reluctance device.

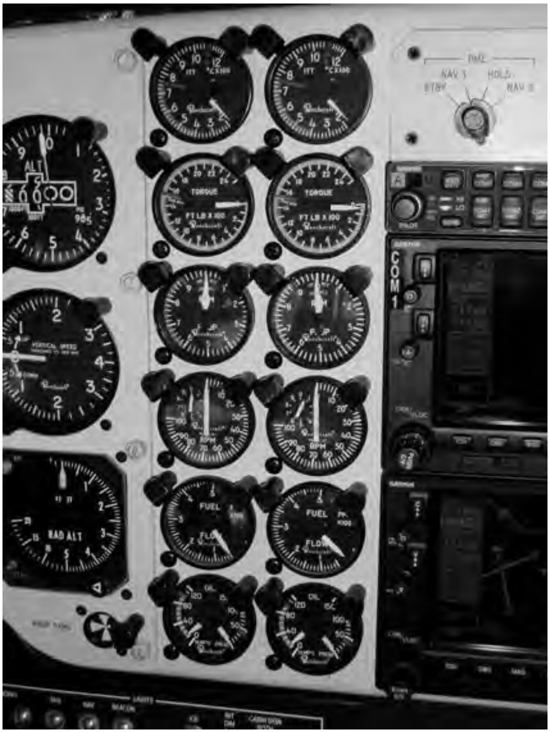

Figure 10.13 Typical engine instruments

10.3.1.1 Tachometer system

The **tachometer** indicating system is a small three-phase AC generator connected via a mechanical link to engine accessory gearbox. A tachometer system is found on most general aviation aircraft. Referring to fundamental principles, the tachometer's output increases with increased engine speed; the output is rectified and connected to a moving coil meter. The output from the generator is supplied to a three-phase AC synchronous motor in the indicator; see Fig. 10.14.

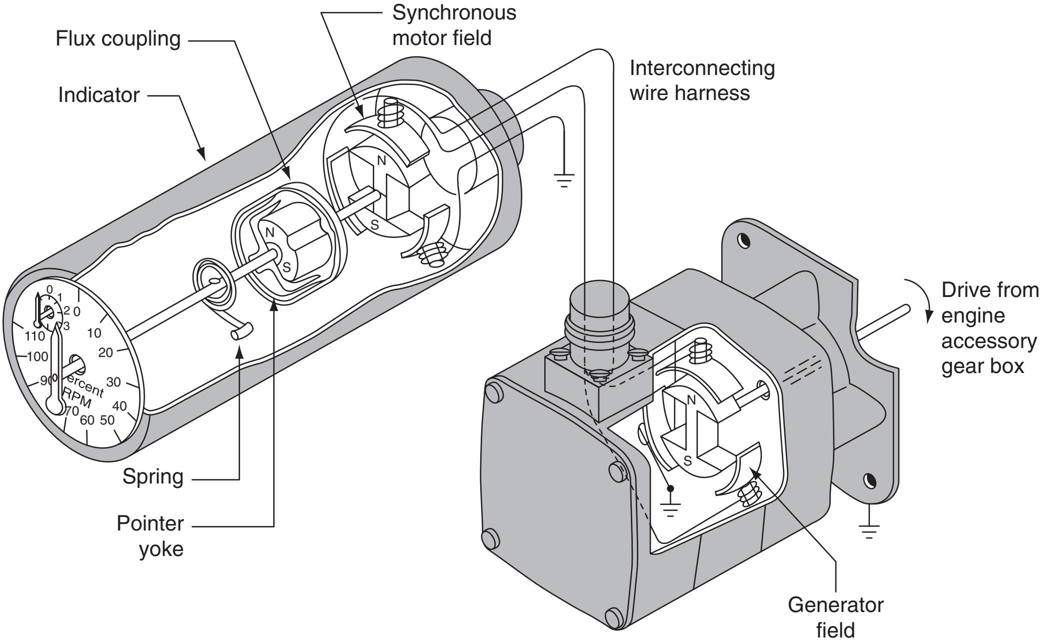
Synchronous motor field
Flux coupling
Indicator
Interconnecting wire harness
Drive from engine accessory gear box
Spring
Pointer yoke
Generator field
Percent RPM
0
1
2
3
10
20
30
40
50
60
70
90
100
110
N
S

Figure 10.14 Tachometer system schematic

The AC generator tachometer consists of a **permanent magnet** that is rotated inside stator windings wound in a three-phase star configuration. Three stator outputs are connected to the stator windings of the motor. As the engine turns, the permanent magnet induces currents in the stator windings. The three-phase output induces a rotating field in the motor stator windings.

Referring to Fig. 10.15, a permanent magnet is attached to the rotor shaft of the synchronous motor; this is connected to the pointer of the indicator. As the stator field rotates, the permanent magnet keeps itself aligned with the field. A second permanent magnet is attached to the indicator rotor; this is located within a **drag cup**, see Fig. 10.16. As this second magnet rotates, it induces **eddy currents** in the drag cup. These currents produce their own magnetic fields in opposition to the rotating magnet.

When the rotating magnet increases in speed, the drag (or torque) on the drag cup increases. A hairspring attached to the shaft opposes this torque; the net result is the pointer moves across the scale in proportion to the speed of the engine shaft.

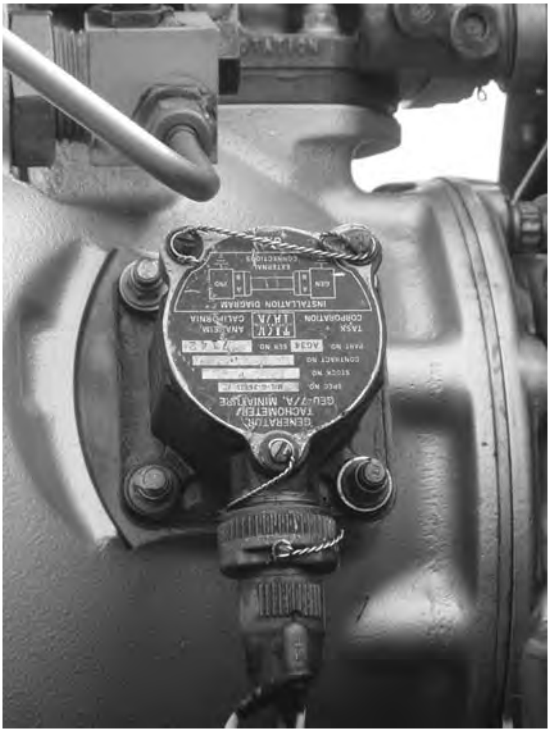

Figure 10.15 Tachometer system installation

10.3.1.2 Variable reluctance

The N1 **variable reluctance** speed transducer comprises a coil wound onto a permanent magnet core as shown in Fig. 10.17. This type of sensor is found on most gas turbine engines. When the blade tip passes the sensor, the magnetic field is disturbed. This induces a voltage into the coil; as the shaft speed increases the fan tips pass the sensor at an increase rate. The output from the coil is in the form of voltage ‘spikes’. These are counted by a processor and used to determine engine speed. Some engines are fitted with a low permeability material inserted into the blade tip to provide a distinctive pulse each time that a specific blade passes the sensor.

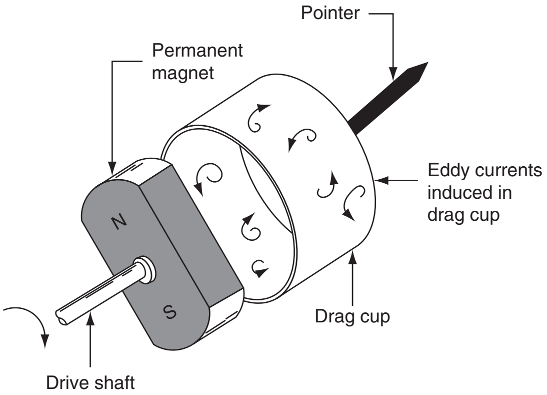
Pointer
Permanent magnet
Eddy currents induced in drag cup
N
S
Drag cup
Drive shaft

Figure 10.16 Drag cup features

The N2 speed transducer is located within the accessory gearbox, see Fig. 10.18. A gearbox-driven **target wheel** contains a permanent magnet inserted on the periphery of the wheel. The coil/core sensor field is disturbed each time that the target passes the sensor. The output is supplied into a processor that drives the indicator.

Test your understanding 10.3

Explain the terms N1, N2 and N3. How do these relate to the HP, IP and LP shafts?

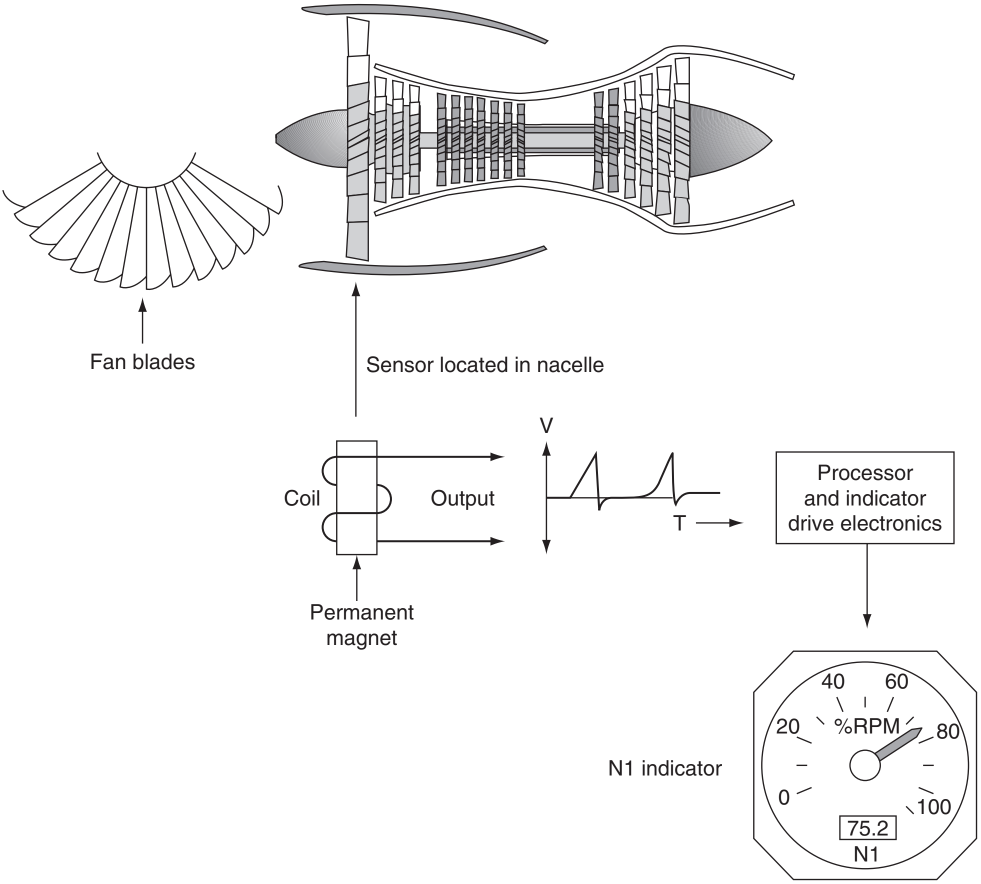
Fan blades
Sensor located in nacelle
Coil
Output
Permanent magnet
V
T
Processor and indicator drive electronics
N1 indicator
40
60
20
%RPM
80
0
100
75.2
N1

Figure 10.17 N1 engine speed indication

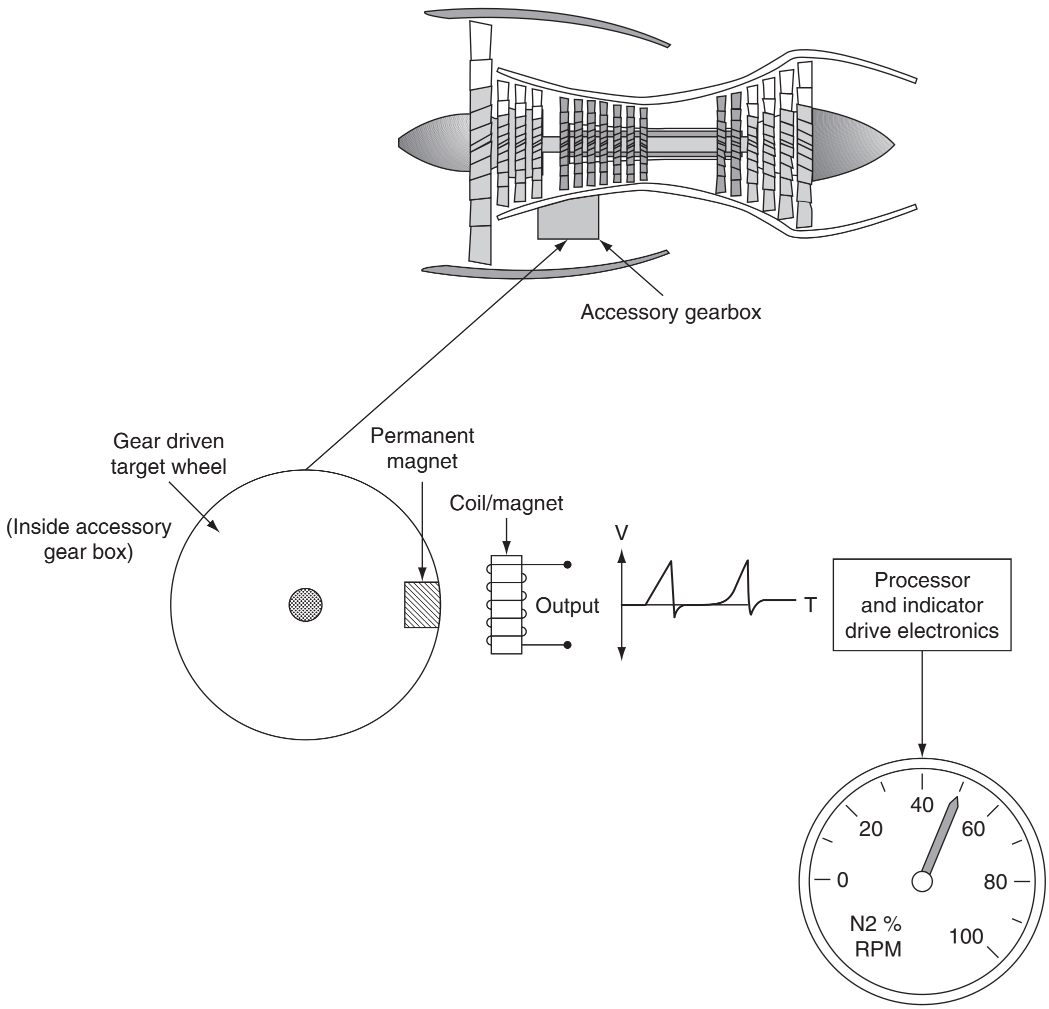
Accessory gearbox
Gear driven target wheel
(Inside accessory gear box)
Permanent magnet
Coil/magnet
Output
V
T
Processor and indicator drive electronics
0
20
40
60
80
100
N2 %
RPM

Figure 10.18 N2 engine speed indication

10.3.2 Engine temperature

Exhaust gas temperature is a primary gas turbine engine indication. Engine temperature is closely monitored at all times, particularly during start and take-off, to make sure that engine limits are not exceeded. It is sometimes referred to as:

- turbine inlet temperature (TIT)
- inter-turbine temperature (ITT)
- turbine outlet temperature (TOT)
- exhaust gas temperature (EGT)
- turbine gas temperature (TGT)
- jet pipe temperature (JPT).

The type of measurement depends on where the probe is located and according to individual engine manufacturer's terminology; Fig. 10.19 illustrates some of these locations. The turbine section runs at very high temperatures, typically 1000ºC; the transducer used in this application is the **thermocouple** (sometimes referred to as a 'temperature probe').

A typical engine temperature measurement system is illustrated in Fig. 10.20. On larger engines, it is possible to have a range of temperatures in the exhaust zone due to the turbulence of gases.

Some installations feature a thermocouple that has two or even three hot junctions within the same

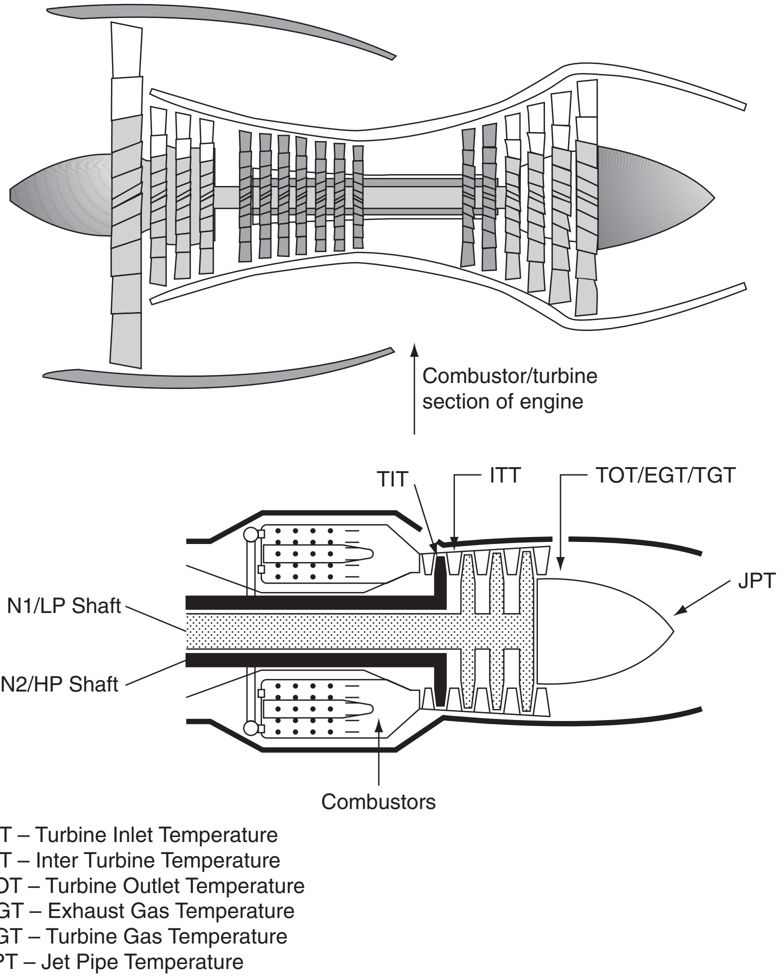
Combustor/turbine
section of engine
TIT
ITT
TOT/EGT/TGT
JPT
N1/LP Shaft
N2/HP Shaft
Combustors
TIT – Turbine Inlet Temperature
ITT – Inter Turbine Temperature
TOT – Turbine Outlet Temperature
EGT – Exhaust Gas Temperature
TGT – Turbine Gas Temperature
JPT – Jet Pipe Temperature

Figure 10.19 Engine temperature measurement

outer tube. This arrangement provides an average of temperature within the zone to provide an average temperature at different immersion distances in the engine, see Fig. 10.21.

Key point

Thermocouple junctions alone do not generate electromotive force (e.m.f.). The potential difference that develops at the heated junction is a function of both the hot and cold junction temperatures.

Small gas turbine engines are normally fitted with several thermocouples to provide an average temperature and some redundancy in case of failure. Larger engines can be fitted with up to 21 thermocouples; these are connected in parallel to provide an average reading of the gas temperature in the exhaust zone. The interconnecting cables between the thermocouple(s) and indicator have to be the same material throughout the system otherwise addition junctions will be formed, thereby generating unwanted voltages. A typical thermocouple cable installation is shown in Fig. 10.22.

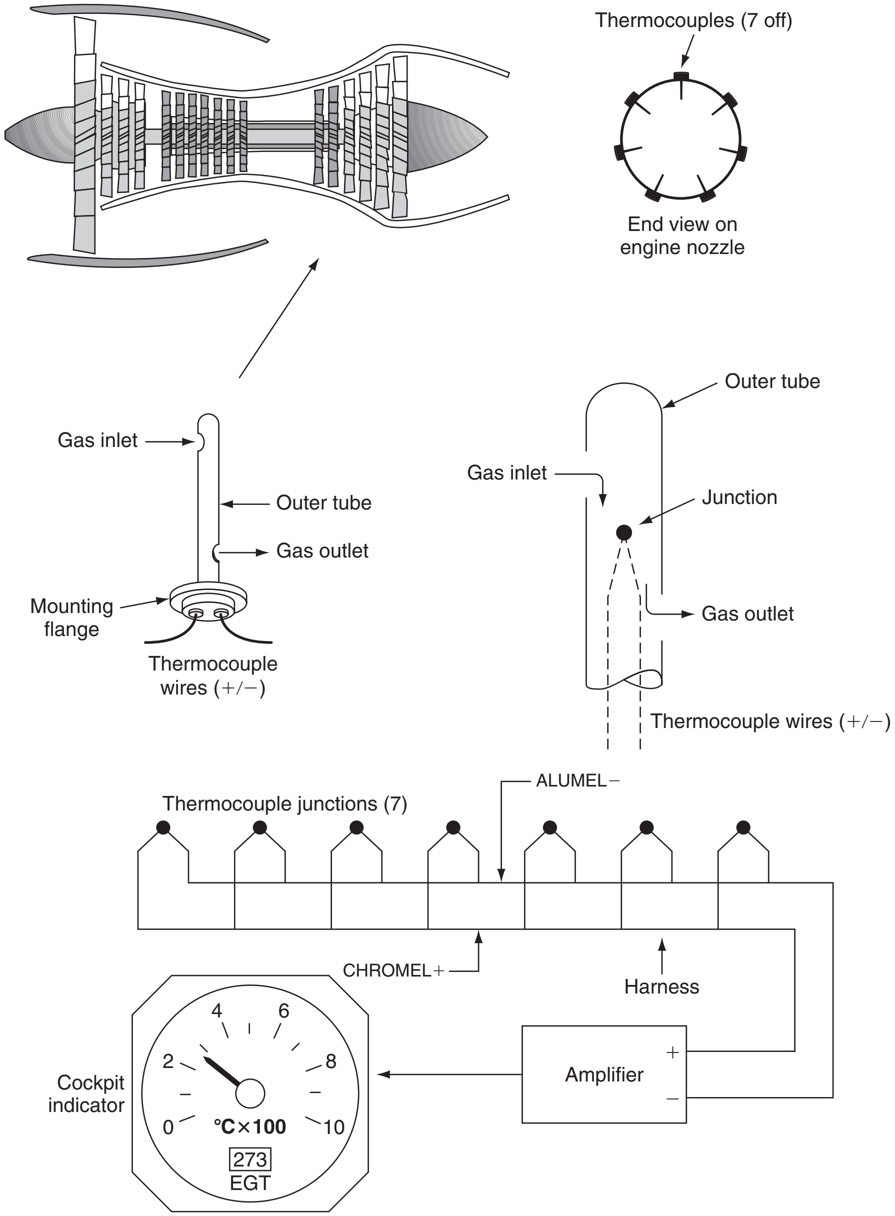
Thermocouples (7 off)
End view on engine nozzle
Gas inlet
Outer tube
Gas outlet
Mounting flange
Thermocouple wires (+/−)
Outer tube
Gas inlet
Junction
Gas outlet
Thermocouple wires (+/−)
ALUMEL−
Thermocouple junctions (7)
CHROMEL+
Harness
Cockpit indicator
0
2
4
6
8
10
°C×100
273
EGT
Amplifier
+
−

Figure 10.20 Engine temperature system

Key maintenance point

Thermocouple cables are colour-coded to reduce the likelihood of different materials being cross-connected, or mixed in the same installation (note that these codes vary in some countries; always refer to the maintenance manual):

- nickel-chrome (white)
- nickel-aluminium (green)
- iron (black)
- constantan (yellow)
- copper (red).

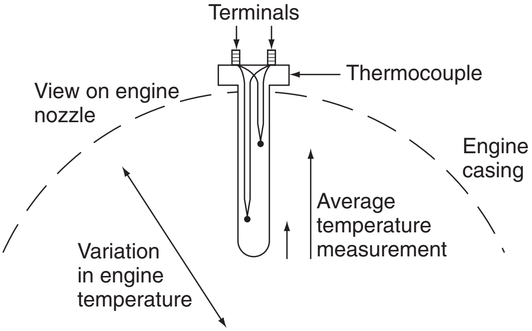
Terminals
Thermocouple
View on engine nozzle
Engine casing
Average temperature measurement
Variation in engine temperature

Figure 10.21 Average temperature measurement

Figure 10.22 Thermocouple/cable installation

On piston engines, the EGT indication has a different application; it is used to adjust the fuel–air mixture for economical running of the engine. The thermocouple is located in located in the exhaust system, reducing the amount of fuel into the cylinders (called **leaning** the mixture); this results in a more efficient combustion and the exhaust temperature increases. When this temperature reaches a peak, maximum efficiency has been achieved. Using one thermocouple per cylinder averages out individual differences and provides enhanced fuel efficiencies.

Piston engines also use thermocouples to monitor the **cylinder head temperature** (CHT). This indication is used to monitor the temperature of an air-cooled engine. The system can be arranged with either one CHT thermocouple per cylinder or just one CHT thermocouple per engine. In the latter case, the thermocouple is installed in the hottest running cylinder; on a horizontally opposed engine this would be the rear cylinder. The thermocouple can be formed to fit under the spark plug or directly into the head itself. The advantage of monitoring each cylinder is that trend indications can be derived.

Test your understanding 10.4

Why are thermocouple cables colour-coded?

10.3.3 Engine pressure ratio (EPR)

This is a primary gas turbine engine indication system and is one of two methods used to indicate **thrust** (the alternative method is described under engine speed). The principle of an EPR indication system is to measure the exhaust and inlet pressures and derive a ratio between them. EPR probes are located at the inlet and exhaust of the engine, see Fig. 10.23. The inlet pressure probe is a single device, located in the nacelle; there can be several exhaust probes connected via a manifold.

Pressures from the inlet and exhaust are sent via small diameter pipes to the EPR transducer; this comprises pressure sensors (capsules). These are coupled to a mechanism that determines the displacement of the capsules as ratio. Outputs from the ratio sensor are sent via a **linear variable differential transformer (LVDT)** to the indicator. The EPR signal from the transducer is transmitted as a voltage to the indicator. The typical range of EPR indications is between one and two.

Test your understanding 10.5

What EPR reading would be displayed when an engine is not running?

10.3.4 Fuel flow

This is a primary engine indication for which there are various types of transducer; these are located in the fuel delivery pipeline. A typical transducer type is based on the **metering vane** principle; see Fig. 10.24. The metering vane is attached to a shaft; this

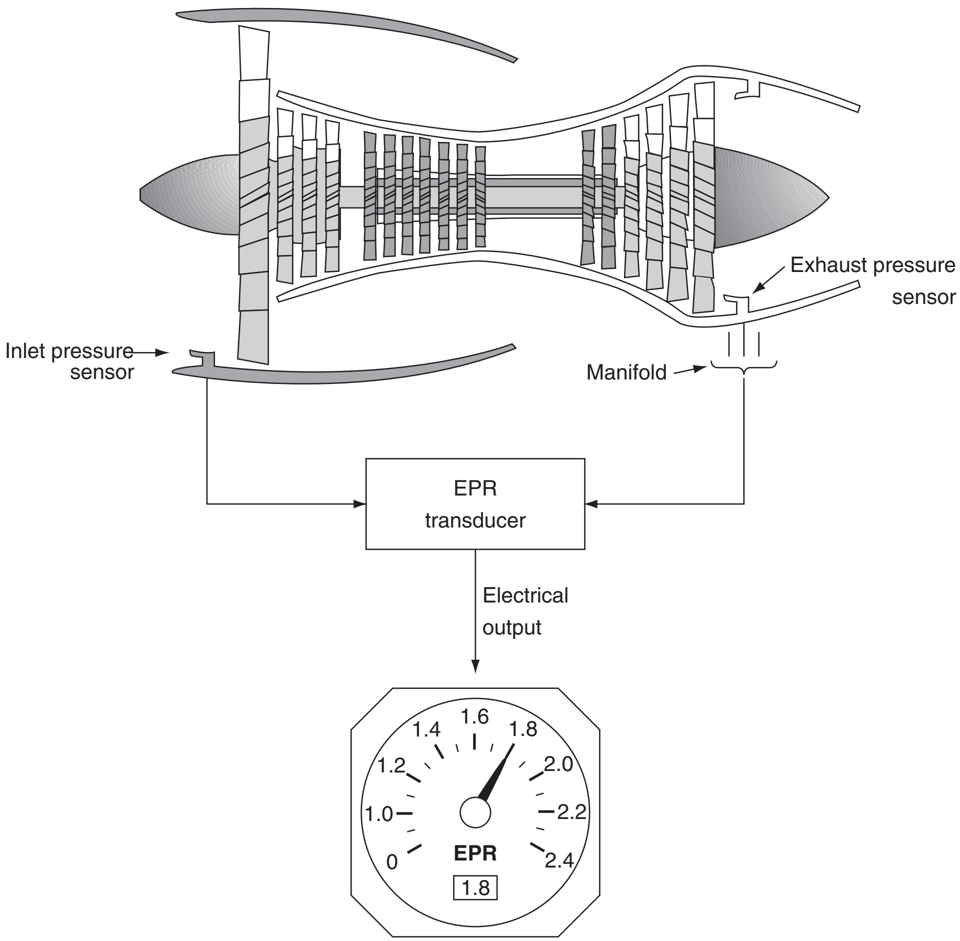

Figure 10.23 EPR schematic

rotates when the fuel supply passes through the body of the transducer. The circular chamber has sufficient clearance to allow fuel to pass through with minimal restriction.

Rotation of the vane is opposed by a spring; the angular position of the vane is measured and indicated by a synchro system, see Fig. 10.25. The metering vane system gives an indication of **volume flow**, e.g. gallons or litres of fuel flow per hour.

An alternative fuel flow transducer is the **motor-driven impeller** principle; see Fig. 10.26. An impeller is driven at a constant speed by an AC synchronous motor. The fuel that flows through the impeller is imparted with a rotational force; this causes the turbine blades to rotate against the force of a spring. The speed of turbine rotation is therefore proportional to fuel flow. Denser fuel imparts more energy into the turbine; the transducer is therefore measuring mass flow. The turbine is connected to the indicator via a synchro system; typical units of **mass flow** are given in pounds or kilograms of fuel per hour.

10.3.5 Torque

The power delivered by the engine to the propeller shaft can be derived from the relationship:

$$\text{Power} = \text{torque} \times \text{speed}$$

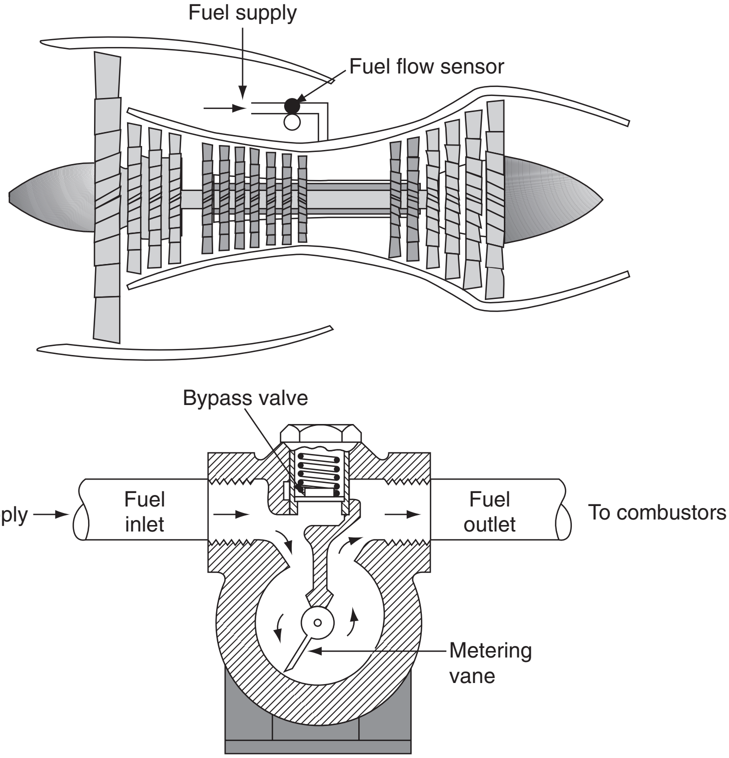
Fuel supply
Fuel flow sensor
Bypass valve
Fuel supply
Fuel inlet
Fuel outlet
To combustors
Metering vane

Figure 10.24 Fuel flow-metering vane principles

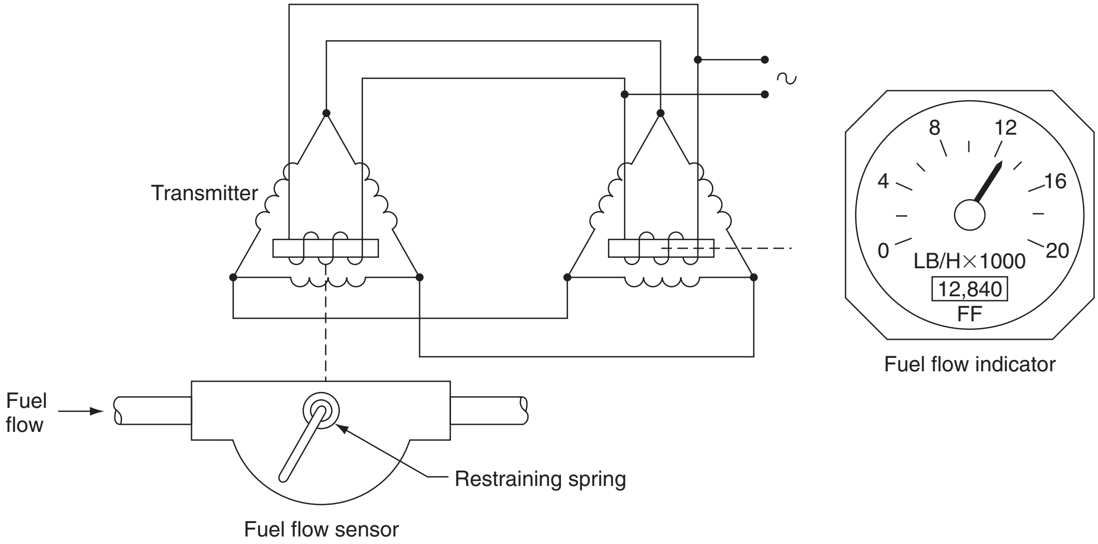
~
Transmitter
8
12
4
16
0
20
LB/H×1000
12,840
FF
Fuel flow indicator
Fuel flow
Restraining spring
Fuel flow sensor

Figure 10.25 Fuel flow indicating system

Power is derived from measuring both torque and speed; the indication is normally used for turboprops and helicopter rotors. Indicators are calibrated in a percentage of maximum torque or **shaft horsepower** (SHP). A typical torque indicator used on a twin-engined helicopter is shown in Fig. 10.27.

Indications are given for the output from both engines and main rotor (M/R) torque. This is the most effective way of indicating the power being produced by the engine.

There are several methods used to measure torque. The torque shaft is formed with toothed or **phonic wheels**, see Fig. 10.28. As the applied torque changes the phase difference between the signals obtained from the two speed sensors increases. The torque applied to the shaft results in a phase difference between the outputs of the two sensors, see Fig. 10.29.

Torque can also be measured by embedding **strain gauges** into the shaft to measure the deformation (strain) of the shaft; these can either be **metallic foil**

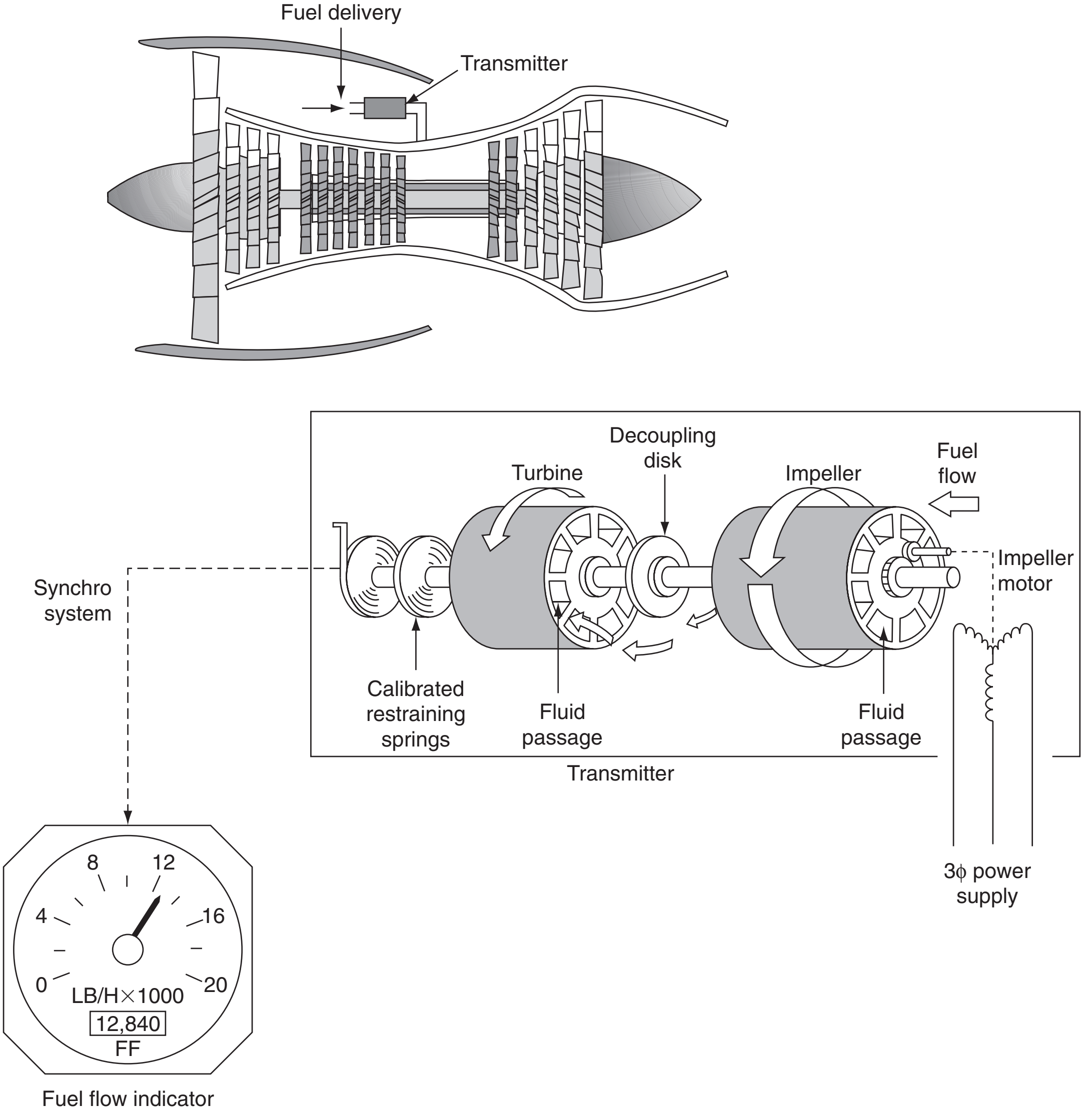
Fuel delivery
Transmitter
Synchro system
Decoupling disk
Turbine
Impeller
Fuel flow
Impeller motor
Calibrated restraining springs
Fluid passage
Fluid passage
Transmitter
3ϕ power supply
0
4
8
12
16
20
LB/H×1000
12,840
FF
Fuel flow indicator

Figure 10.26 Fuel flow – impeller principle

or semiconductor **piezo-resistors** (as described in Chapter 9).

Test your understanding 10.6

What is the difference between torque and thrust?

10.4 Secondary indicating systems

10.4.1 Oil/fuel temperature

There is a need for accurate measurement of fluid temperatures on and around the engine, e.g. fuel, engine lubrication (lube) oil and hydraulic oil. The typical temperature range of these fluids is between −40 and +150°C. When lube oil operates at high temperatures, its viscosity reduces and its lubrication

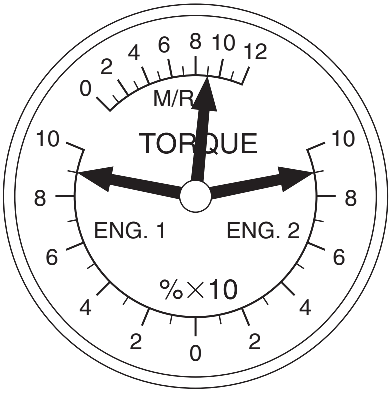
0
2
4
6
8
10
12
M/R
TORQUE
10
8
6
4
2
0
2
4
6
8
10
ENG. 1
ENG. 2
%×10

Figure 10.27 Typical helicopter torque indicator (twin engine)

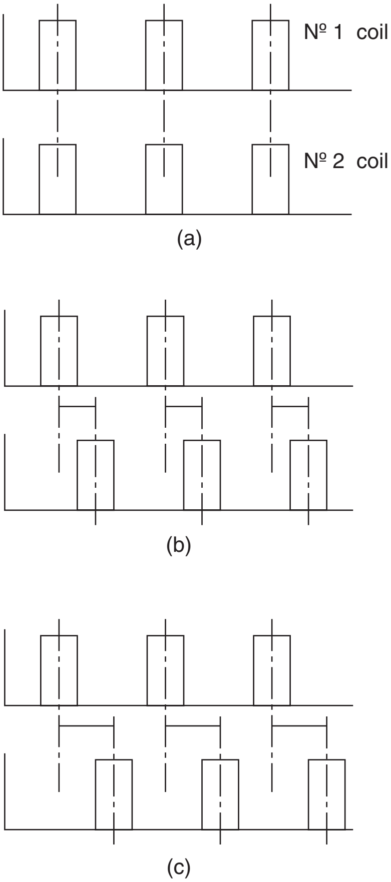
Nº 1 coil
Nº 2 coil
(a)
(b)
(c)

Figure 10.29 Shaft torque sensor phase differences (a) no load conditions, (b) engine lightly loaded, (c) engine heavily loaded.

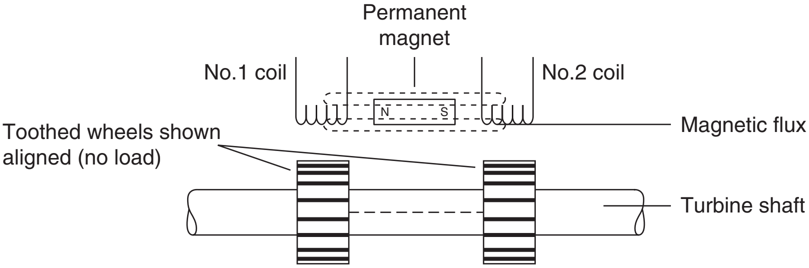
Permanent magnet
No.1 coil
No.2 coil
N
S
Magnetic flux
Toothed wheels shown aligned (no load)
Turbine shaft

Figure 10.28 Torque transducer principles

performance decreases. This leads to engine wear and eventual failure of bearings or other engine components. When lube oil operates at low temperatures, its **viscosity** increases, which has an effect on starting. When fuel is exposed to high temperatures, it can vaporize, causing fuel delivery problems and a potential risk of explosion; at low temperatures, ice formation can occur leading to blocked filters. (Fuel systems are further described in Chapter 11.)

The flight compartment indicator is a moving coil meter connected in series with the resistance winding of a resistance temperature device (RTD); this item is described in more detail in Chapter 9. As the temperature changes, the resistance changes and this causes the pointer to respond accordingly. The meter is acting as an ammeter, but since there is a linear relationship between temperature and resistance, and since we know (from Ohm's law) that the current in a circuit will vary in accordance with resistance, the moving coil meter can be calibrated as a temperature indicator.

This is all very well provided that the power supply voltage does not change. To illustrate this problem, assume a given temperature and hence a specific resistance of the RTD. The effect of a varying power supply voltage will be to vary the current through the RTD and meter; this would change the reading even though temperature has not changed; an undesirable situation. The method employed to overcome this is a circuit called the **Wheatstone bridge**, as described in Chapter 9. This circuit has a number of applications and is ideally suited for the RTD.

10.4.2 Vibration

Unbalanced shafts on a gas turbine engine can cause damage, particularly at high rotational speeds. These conditions can be anticipated by measuring the engine's vibration. Sensors used to detect vibration are based on **piezoelectric** crystals; a small electrical charge is created when the crystal is vibrated. Each engine has a sensor(s) located at strategic locations; the output is fed into a processor and displayed on an indicator, see Fig. 10.30.

The output from the processor can also be used to illuminate a warning light when pre-determined limits are exceeded. A test switch (incorporated into the light) is used to energize a relay (or equivalent electronic switch); this inserts a known frequency into the processing circuit to generate a warning. The vibration warning circuit also activates the master caution system.

10.4.3 Fluid pressure

Typical fluid pressures being measured on the engine include lubrication oil and fuel. There are two basic methods used to measure fluid pressure: the **Bourdon tube** or a capsule. The Bourdon tube comprises a tube formed into a curved or spiral shape. As the applied pressure from the fluid system increases, the tube will tend to straighten out, while a reduced pressure will cause the tube to return to its original shape. This movement is transferred via the gear mechanism to move a pointer. The pointer moves across a scale, thereby providing a **direct reading** of pressure. The other method used for measuring fluid pressure is with a **capsule**. As pressure is applied by the fluid, the capsule expands. This moves an iron bobbin within the envelope of two inductor coils. (The principles of Bourdon tubes, pressure capsules and measurement systems are described in Chapter 9.)

10.4.4 Propeller synchronization

This is a mechanism that automatically synchronizes all propellers of a multiengine aircraft to ensure that they rotate at the same speed. The main reason for synchronizing the propeller speed is for the comfort of crew and passengers; propellers that are turning at slightly different speeds create a 'beat' through **heterodyning**. This is the creation of new frequencies through mixing two different frequencies; one at the difference of the two mixed frequencies, and the other at their sum. Heterodyning creates beats that can become very tiresome to persons inside the aircraft over long periods of time. Some aircraft are installed with a visual indicator of propeller synchronization. Pilots can use this indicator to decide whether or not to engage automatic synchronization, or to assist with manually synchronization of propeller speeds by adjusting the throttles.

10.5 Electronic indicating systems

There are two formats of electronic indicating systems in widespread use on larger passenger aircraft: EICAS and ECAM. The **Engine Indication and Crew Alerting System** (EICAS) is a Boeing-developed system that provides all engine instrumentation and crew annunciations in an integrated format. The system used on Airbus aircraft is the **Electronic**

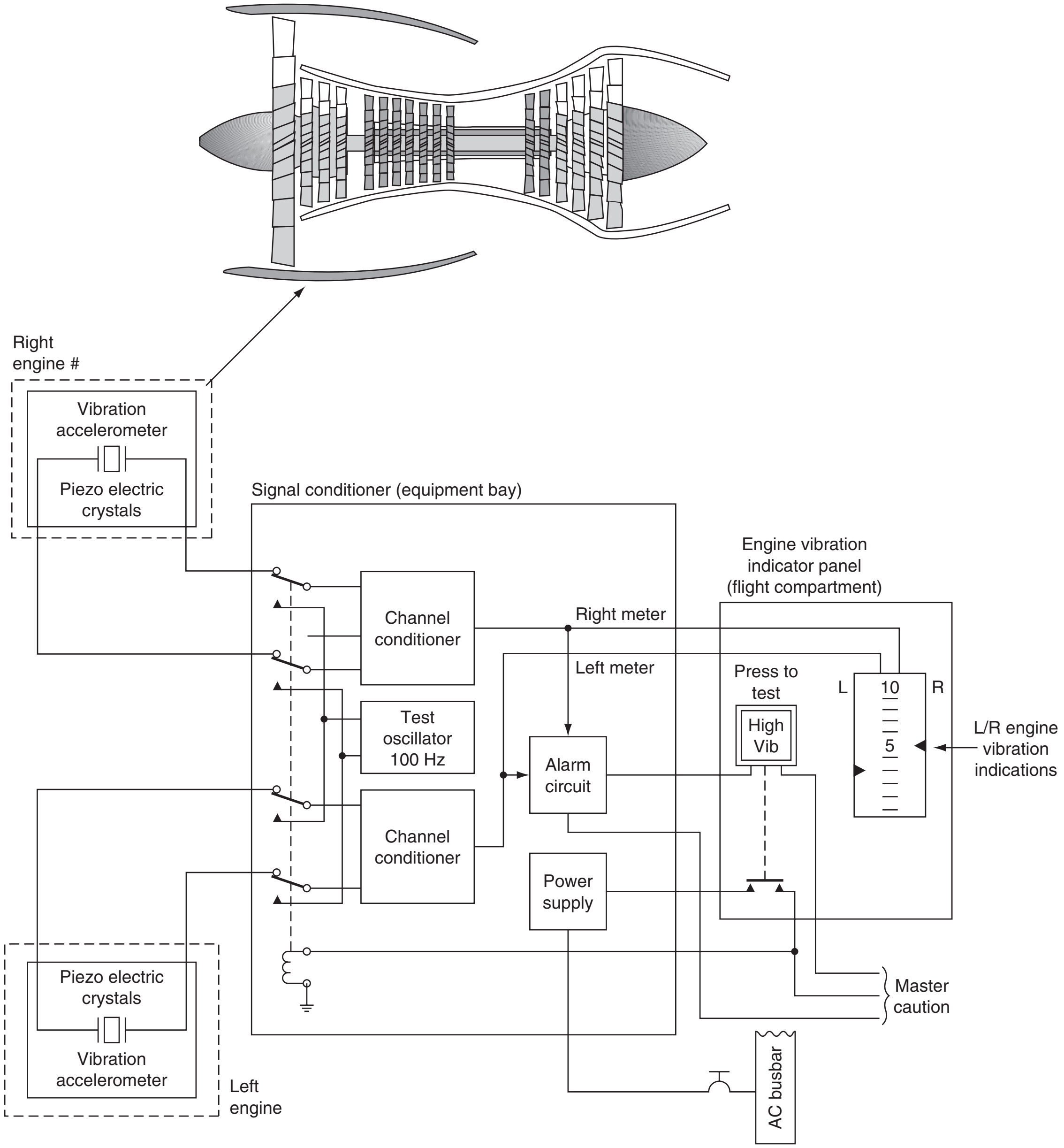
Right
engine #
Vibration
accelerometer
Piezo electric
crystals
Signal conditioner (equipment bay)
Channel
conditioner
Test
oscillator
100 Hz
Channel
conditioner
Right meter
Left meter
Alarm
circuit
Power
supply
Engine vibration
indicator panel
(flight compartment)
Press to
test
High
Vib
L
10
5
R
L/R engine
vibration
indications
Master
caution
Piezo electric
crystals
Vibration
accelerometer
Left
engine
AC busbar

Figure 10.30 Vibration sensor system

Centralized Aircraft Monitoring (ECAM) system. The two systems operate on different philosophies; however, their basic functions are to monitor aircraft systems and display relevant information to the pilots. Both systems produce warnings, cautions and advisory messages that need to be evaluated by the crew. In certain cases, the system provides the procedures required to address the problem. Each colour display unit uses either an active matrix liquid crystal display (AMLCD) or a cathode ray tube (CRT).

EICAS is arranged with two displays in the centre instrument panel (upper and lower), see Fig. 10.31. ECAM uses either a left and right display, see Fig. 10.32, or an upper and lower arrangement. The

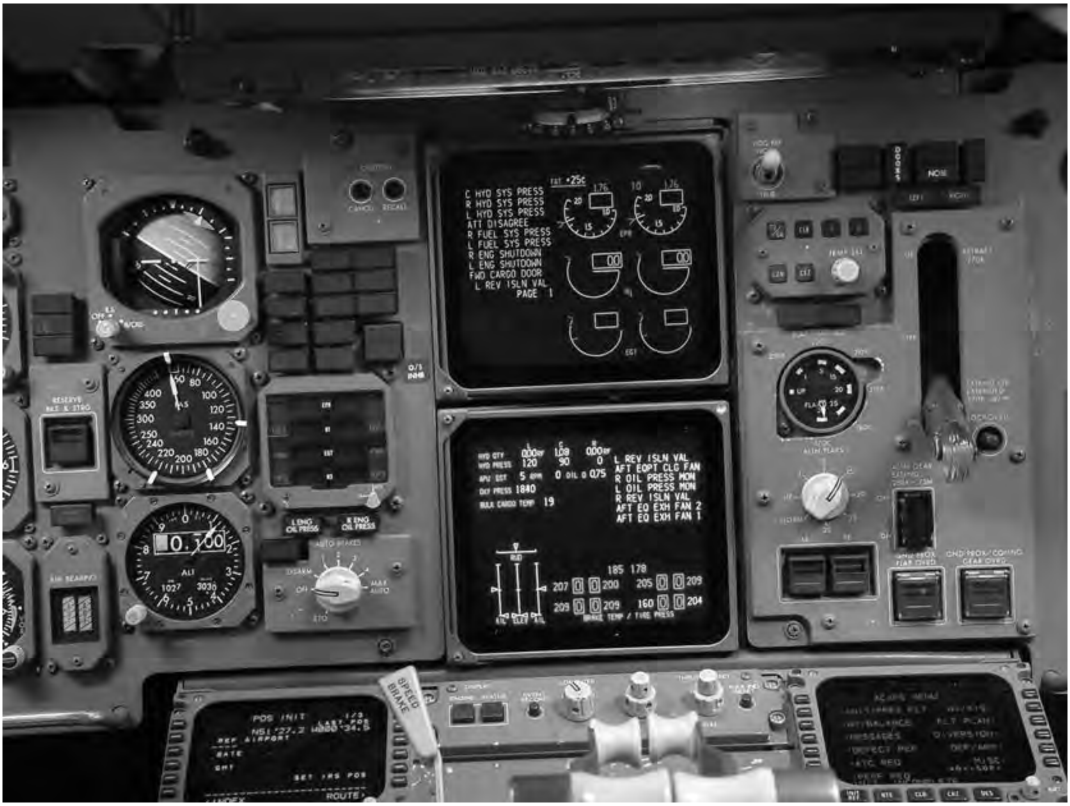
C HYD SYS PRESS
R HYD SYS PRESS
L HYD SYS PRESS
ATT DISAGREE
R FUEL SYS PRESS
L FUEL SYS PRESS
R ENG SHUTDOWN
L ENG SHUTDOWN
FWD CARGO DOOR
L REV ISLN VAL
PAGE 1
L REV ISLN VAL
AFT EQPT CLG FAN
R OIL PRESS MON
L OIL PRESS MON
R REV ISLN VAL
AFT EQ EXH FAN 2
AFT EQ EXH FAN 1
BRAKE TEMP / TIRE PRESS
SPEED BRAKE
POS INIT

Figure 10.31 EICAS

information displayed by EICAS/ECAM includes engine thrust, engine r.p.m., fuel flow, oil temperature and pressure. Graphical depiction of aircraft systems can also be displayed; this includes electrical, hydraulic, de-icing, environmental and control surface positions.

10.5.1 EICAS

System features include:

- primary engine parameters displayed on a full-time basis
- secondary engine parameters displayed as required
- automatic monitoring of systems
- fault codes, fault history, self test (lower display)
- two computers (includes interfaces, symbol generator, memory)
- select panel, control switches, annunciation, standby display.

This integrated system improves reliability through elimination of traditional engine gauges and simplifies the flight deck through fewer stand-alone indicators. EICAS also reduces crew workload by employing a graphical presentation that can be rapidly assimilated. A typical EICAS comprises two large high-resolution colour displays together with associated control panels, two or three EICAS data concentrator units and a lamp driver unit. The primary EICAS display presents primary engine indication instruments and relevant crew alerts. It has a fixed format providing engine data including:

- engine thrust (EPR)
- engine rotational speed (N_1)
- exhaust gas temperature (EGT).

If N1 or EGT limits are exceeded, their respective pointers and digital readouts change from white to yellow, and eventually to red. Exceedance information is stored in non-volatile memory (NVM) for access by maintenance engineers when troubleshooting the systems.

Messages are displayed as part of the Caution Alerting System (CAS); these are colour-coded to indicate their importance. Typical messages include

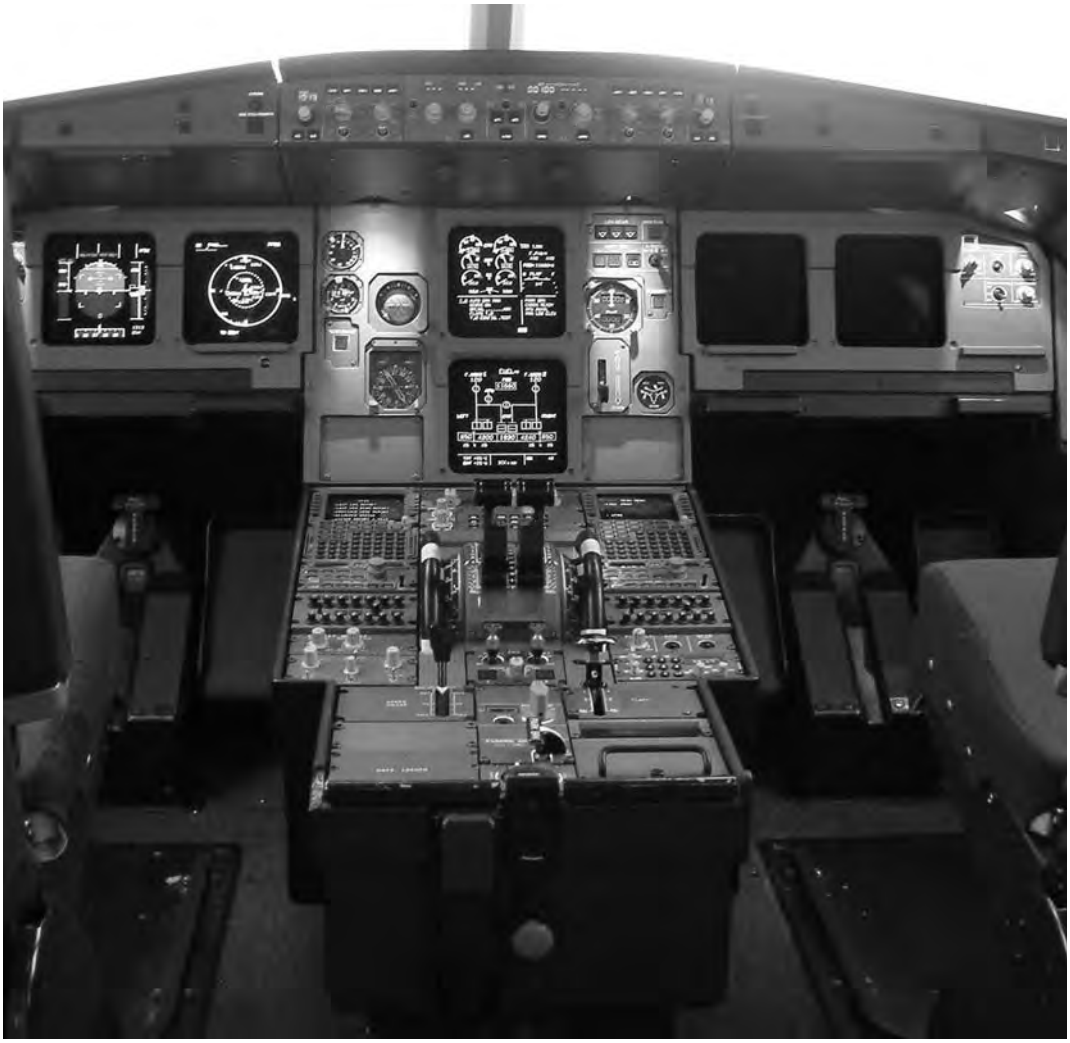

Figure 10.32 ECAM

low-pressure in the fuel or hydraulic oil systems, doors not closed etc. System warnings are automatically colour-coded and prioritized; they are accompanied by an audible warning depending on the message priority:

- **Warning** messages are red, accompanied by an audio alert (prompt action is required by the crew).
- **Caution** messages are yellow accompanied by an audio alert (timely action is required by the crew).
- **Advisory** messages are yellow, no audio alert (time available attention is required by the crew).

The secondary EICAS display indicates a wide variety of options to the crew and serves as a backup to the primary display. Pages are selectable using the EICAS control panel and includes many types of system information, for example:

- hydraulic systems pressure
- flying controls position
- cargo bay temperature
- brake temperatures
- tyre pressures.

The information on the aircraft system data buses is routed to both EICAS displays and both multifunction displays. A data concentrator unit (DCU) receives data in various formats from a variety of sensors, including the high- and low-speed ARINC 429 bus, from analogue and discrete inputs from the engines and other aircraft systems. In the event of a display system failure, the other system displays essential information in a compacted format. Primary engine indications are as before, secondary indications are digital readings only. In the event that both systems fail, a standby LED display is used for EPR, N1 and EGT.

10.5.2 Ecam

This is a flight-phase-related architecture; displays are automatically selected for:

- pre-flight
- take-off
- climb
- cruise
- descent
- approach.

To illustrate ECAM features, during the pre-flight phase, checklists are displayed on the left-hand screen; this includes information such as brake temperatures, APU status and advisory messages. During the pre-flight checks, the right-hand display provides a graphic representation of the aircraft, e.g. if doors are open or closed. The system automatically changes to the relevant pages needed for the next flight phase. System warnings are prioritized, ranging from level one to level three; the warning hierarchy is similar to EICAS.

Both EICAS and ECAM reduce flight deck clutter by integrating the many electro-mechanical instruments that previously monitored engine and aircraft systems. Reliability is increased, and the pilot workload is reduced. Further information on ECAM is given in Chapter 13.

10.6 Multiple choice questions

1. HEIUs can remain charged for several:
   (a) seconds
   (b) minutes
   (c) hours.

2. Ground idle speed occurs when the engine has:
   (a) stabilized (slightly above self-sustaining speed)
   (b) stabilized (slightly below self-sustaining speed)
   (c) just been started.

3. Power from an engine is derived from measuring:
   (a) torque and speed
   (b) temperature and speed
   (c) engine pressure ratio (EPR).

4. When lube oil operates at high temperatures, its viscosity:
   (a) reduces and its lubrication performance decreases
   (b) reduces and its lubrication performance increases
   (c) increases and its lubrication performance increases.

5. The starting sequence for a gas turbine engine is to:
   (a) turn on the ignition, develop sufficient airflow to compress the air, and then open the fuel valves
   (b) develop sufficient airflow to compress the air, open the fuel valves and then turn on the ignition
   (c) develop sufficient airflow to compress the air, turn on the ignition and then open the fuel valves.

6. EICAS warning messages are:
   (a) red, accompanied by an audio alert (prompt action is required by the crew)
   (b) yellow, accompanied by an audio alert (timely action is required by the crew)
   (c) yellow, no audio alert (time available attention is required by the crew).

7. Engine pressure ratio (EPR) is used to measure a gas turbine engine's:
   (a) torque
   (b) thrust
   (c) temperature.

8. Low-, intermediate- and high-pressure shafts are also referred to as:
   (a) N1, N2 and N3
   (b) N2, N3 and N1
   (c) N3, N1 and N2.

9. Typical units of fuel mass flow are given in:
   (a) pounds or kilograms per hour
   (b) gallons per hour
   (c) litres per hour.

10. A gas turbine engine's self-sustaining speed is when sufficient energy is being developed by the engine to provide continuous operation:
    (a) with the starting device still engaged
    (b) with the ignition system in operation
    (c) without the starting device and ignition.

11. Referring to Fig. 10.33, this transducer measures:
    (a) speed
    (b) torque
    (c) linear displacement

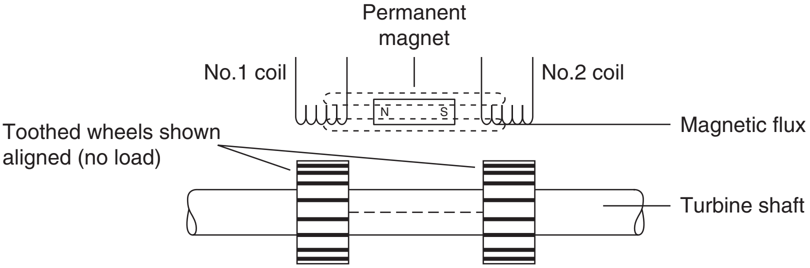
Permanent magnet
No.1 coil
No.2 coil
N
S
Magnetic flux
Toothed wheels shown aligned (no load)
Turbine shaft

Figure 10.33 See Question 11

12. In turboprop engines, power is measured from:
    (a) engine pressure ratio
    (b) torque
    (c) torque × speed.

This page intentionally left blank

Chapter 11 Fuel management

The management of fuel is essential for the safe and economic operation of aircraft. The scope of fuel management depends on the size and type of aircraft; fuel is delivered to the engines using a variety of methods. The system comprises fuel quantity indication, distribution, refuelling, defuelling and fuel jettison. On a typical passenger aircraft, the fuel is contained within the sealed wing box structure. The fuel tanks are divided into main tanks, reserve tanks and centre wing tanks. Fuel tanks on general aviation (GA) aircraft are rubberized bags (bladder tanks) contained within the structure of the aircraft; smaller GA aircraft use metal fuel tanks attached to the wings and/or fuselage.

In the first instance, we need to measure the quantity of fuel on board the aircraft. Various technologies and methods are used to measure fuel quantity: this depends mainly on the type and size of aircraft. Technologies range from sight gauges through to electronic sensors. On larger aircraft, fuel is fed to the engines by electrically driven pumps. On smaller aircraft, an engine-driven pump is used with electrical pumps used as back-up devices. Solenoid or motorized valves are used to isolate the fuel supply to engines under abnormal conditions. On larger aircraft, the fuel can be transferred between tanks; this is controlled manually by the crew, automatically by a fuel control computer. This chapter provides an overview of fuel management on a range of aircraft types.

11.1 Storage overview

Rigid tanks are usually found in smaller general aviation aircraft. They are installed within the fuselage and/or wings, and are designed to be removable for inspection, replacement, or repair. They do not form an integral part of the aircraft structure.

Bladder tanks are reinforced rubberized bags installed within specific areas of aircraft structure. The bladder can be inserted/removed via the fuel filler inlet or a dedicated access panel; the bladder is then secured to the airframe by clips inside the compartment.

Integral fuel tanks are located within the structure on larger aircraft; these are sealed to accommodate fuel storage. These tanks form part of the aircraft structure; they cannot be removed for service or inspection. Access panels are installed to allow internal inspection, repair, and servicing. On large passenger aircraft, there are four main tanks, two in each wing. The area between the forward and aft spars is divided into tank sections by solid ribs. The wing skin (upper and lower surfaces) completes the tank. Wing ribs act as baffles to prevent the fuel from ‘sloshing’ around the tank. The inter-spar area of the wing centre section is also used to store fuel. This centre wing tank is similar in construction to the main tanks.

11.2 Fuel quantity measurement and indication

Various technologies and methods are used to measure and display fuel quantity: this depends mainly on the type and size of aircraft. The fuel quantity methods described here could equally apply to other fluids, e.g. oil, hydraulic fluid or water. Some of the methods used are not actually part of an electrical system; however, they are described here to provide a complete review of the methods employed across a range of aircraft types. The methods used for measuring fuel quantity can be summarized as:

- sight glass
- float gauge
- resistance gauge
- under-wing measurement
- capacitance units.

11.2.1 Sight glass

The **sight glass** method illustrated in Fig. 11.1 is based on a simple glass or plastic tube located on the outside of the tank, and visible to the pilot. Fluid level in the tube is the same as the level in the tank; graduations on the tube provide an indication of tank

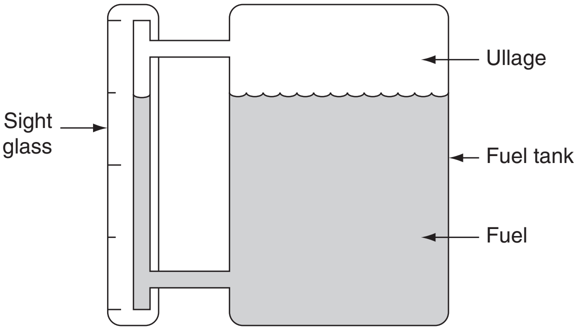

Figure 11.1 Fuel quantity sight glass

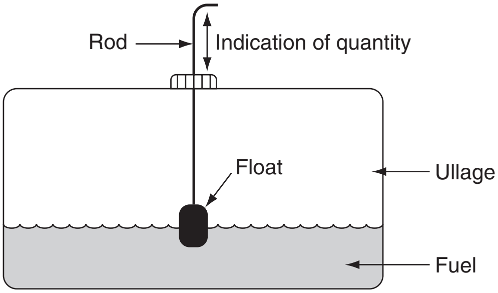

Figure 11.2 Fuel quantity float gauge

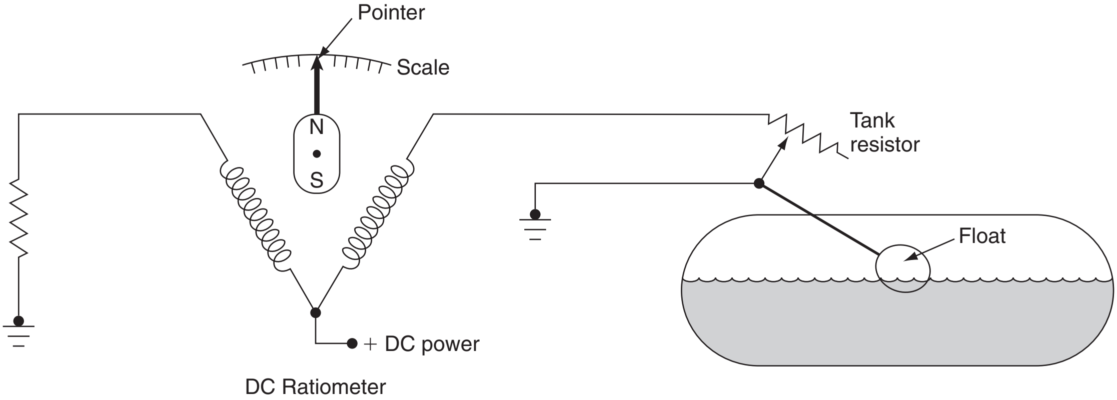

Figure 11.3 Fuel quantity float valve

contents. The advantage of this method is that it has no moving parts and is suitable for fuel or oil. This method is suitable for small aircraft where the pilot can see the sight glass; alternatively it is suited for ground serving applications.

Key point

The space above the fuel in a storage tank is called the ullage.

11.2.2 Float gauge

The **float gauge** uses a rod projecting through a hole in the tank cap; see Fig. 11.2. A float is attached to the base of the rod and this rises and falls with the fuel level. The pilot checks the amount of rod protruding through the cap and this provides a direct reading of fuel quantity. One disadvantage of this method is that it is not very stable during aircraft manoeuvres.

The majority of small general aviation (GA) aircraft use a float gauge system similar to that used in motor vehicles as shown in Fig. 11.3; this is based on a float connected to a variable resistor adjacent to the tank. The variable resistor is connected into a **DC ratiometer** circuit where two opposing magnetic fields are created in each of the coils. The pointer is formed with a permanent magnet and is aligned with the resulting magnetic field created by the coils; the pointer moves in accordance with the ratio of currents in the coil.

Key point

The DC ratiometer pointer position is not adversely affected by variations in power supply voltage.

11.2.3 Under-wing measurement

Under-wing measurement of fuel quantity is used during ground servicing only. The drip stick method uses a hollow tube pushed into the tank. During flight, the tube is stowed into a latched position, which is flush with the aircraft wing or fuselage surface. To take a reading, the tube is released from its stowed position and slowly withdrawn from the tank; when fuel starts to drip from the tank, a reading is taken from graduations on the tube.

Key maintenance point

Spilt fuel presents a fire hazard; take all necessary precautions to clean up any spillage and dispose of any rags soaked in fuel.

A safer alternative to the drip stick method utilizes a transparent plastic rod. This method uses the principle of **light refraction**, and is based on fuel and air having different refractive indexes. When the tip of the rod is moved above/below the surface of the fuel, the light intensity emerging from the viewing end of the rod changes. A reading is then taken from graduations on the rod.

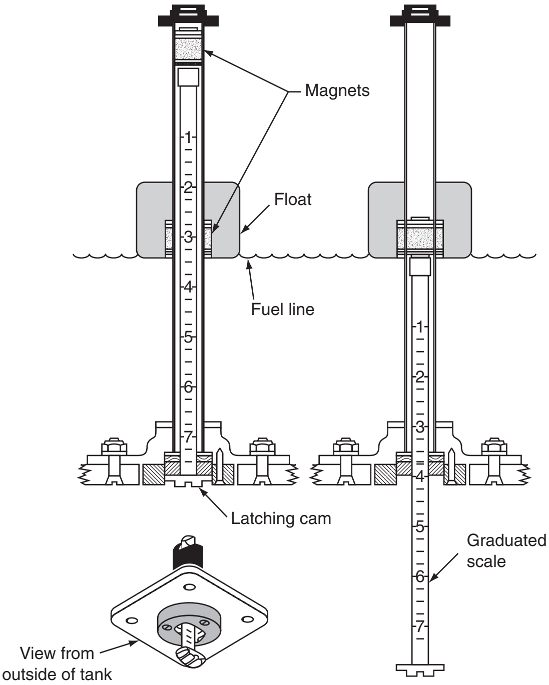

Figure 11.4 Fuel quantity float stick

Another version of the under-wing fuel gauge uses a **floatstick** that comprises a rod, float and magnets located inside the tank as shown in Fig. 11.4. The floatstick is stowed when not in use and released via a quarter-turn cam mechanism; it slides out of the tank until the two magnets align and is then retained in this position. The floatstick is moved in and out of the tank until the attraction of the magnets can be sensed. The fuel quantity reading is then taken from a reference point on the surface of the wing. When the reading has been taken, the rod is pushed back and locked into the stowed position.

Floatsticks can be used when the electronic fuel quantity system (see capacitive fuel quantity system) is unserviceable. All floatsticks are read for each tank and the measurements recorded, the quantity is then calculated from tables in the aircraft documentation. A typical medium-sized aircraft has six floatsticks in each wing tank and four in the centre tank.

Key point

Under-wing fuel quantity measurements are used during ground servicing only.

11.2.4 Capacitive fuel quantity system

The majority of turbine-powered aircraft fuel quantity-indicating systems (**FQIS**) use the **capacitive fuel quantity system**. On a typical twin-engine passenger aircraft the fuel tanks are contained within the aircraft structure as shown in Fig. 11.5. Fuel tank sensors are located throughout each tank and are monitored by an electronic control system. Accurate readings can be obtained by capacitive fuel quantity systems in large and irregular-shaped tanks.

11.2.4.1 Principle of operation

Fuel tank units are formed by concentric aluminium tubes; the inner and outer tubes are the capacitor plates, see Fig. 11.6. The primary advantages of this technology are no moving parts and fuel quantity is measured in **mass** rather than volume. (The mass of fuel determines the amount of energy available.) The **dielectric** is either fuel or air depending on the quantity of fuel in tank. Air has a dielectric of 1.0006 (practically unity); fuel has a dielectric of approximately

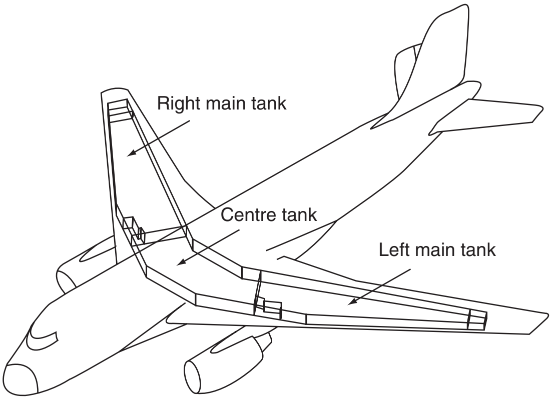
Right main tank
Centre tank
Left main tank

Figure 11.5 Integral fuel tanks

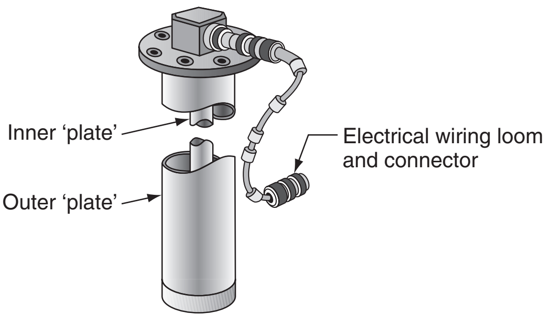
Inner ‘plate’
Electrical wiring loom and connector
Outer ‘plate’

Figure 11.6 Capacitance fuel quantity system – sensor

two; this provides a good relationship for the measurement of a variable quantity.

From basic fundamentals, we know that capacitance is proportional to:

- plate area
- air gap
- dielectric strength.

In the capacitive tank unit, the first two parameters are fixed; the capacitance varies in accordance with the dielectric, i.e. the amount of fuel in the tank as illustrated in Fig. 11.7. With a high quantity of fuel in the tank, the capacitance is high; capacitance varies in direct proportion to the amount of fuel in the tank.

The tank’s capacitance unit is connected into an **impedance bridge** circuit, see Fig. 11.8. Variation in capacitance (*C*) of the fuel tank unit causes a change in reactance (X_C); this is given by the formula:

$$X_C = \frac{1}{2}\pi FC$$

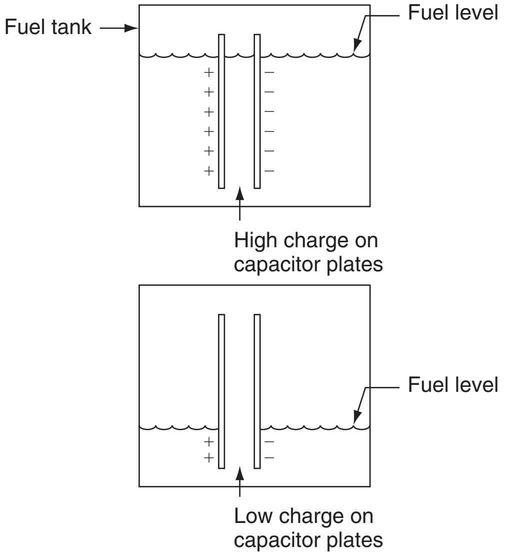
Fuel tank
Fuel level
High charge on capacitor plates
Fuel level
Low charge on capacitor plates

Figure 11.7 Capacitance fuel quantity system – principles

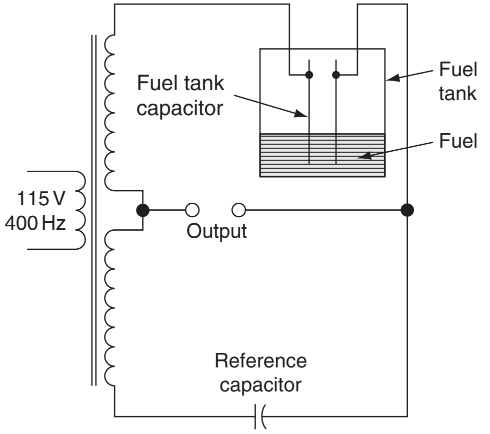
Fuel tank capacitor
Fuel tank
Fuel
115 V
400 Hz
Output
Reference capacitor

Figure 11.8 Capacitance fuel quantity system – impedance bridge circuit

where X_C is the reactance, Π is a constant, *F* is the frequency (400 Hz from the aircraft power supply) and *C* is the capacitance. The impedance (*Z*) of a capacitive network can be calculated from the formula:

$$Z = \sqrt{(R^2 + X_c^2)}$$

The current in a capacitive network can now be calculated from the formula:

$$I = \frac{V}{Z}$$

When the fuel is consumed during a flight, and the fuel level decreases, the capacitance decreases and the reactance increases; less current flows in the tank unit (compared with the reference capacitor). This unbalance causes a potential difference at points X–Y, proportional to fuel quantity; this signal is amplified and used to drive an indicator. Capacitance trimmers are used for calibration of the fuel indicators (or gauges).

Key point

Air and fuel have dielectrics of approximately unity and two respectively.

11.2.4.2 Density compensation

The volume of fuel in a tank varies with temperature; as the temperature changes, the mass of fuel remains the same, but the volume changes. The dielectric is therefore affected by fuel **density**; this density will change with temperature. Increased density is a result of reduced temperature that will cause increased capacitance. Changes in fuel density are measured by a **compensation unit**. This is an additional tank unit located in the bottom of the fuel tank, therefore it is always immersed in fuel. The compensating unit is connected into the impedance bridge such that changes in fuel density cause the bridge to become unbalanced and this compensates for the change in fuel level.

Key point

The volume of fuel in a tank varies with temperature; as the temperature changes, the mass of fuel remains the same, but the volume changes.

11.2.4.3 Multiple tanks

Multiple tank units are often employed in larger aircraft as illustrated in Fig. 11.9. On a typical medium-sized passenger aircraft there are twelve capacitive tank units in each main fuel tank, and nine in the centre fuel tank. The construction of all the tank units is the same, except for their length; this depends on the depth of the tank at that particular location. Each tank unit consists of two concentric aluminium tubes and a terminal block. The aluminium tubes are anodized and polyurethane coated to protect against corrosion. The air gap between the tubes is relatively wide to avoid electrical short-circuiting, caused by contamination in the fuel or fungus coating on the tubes.

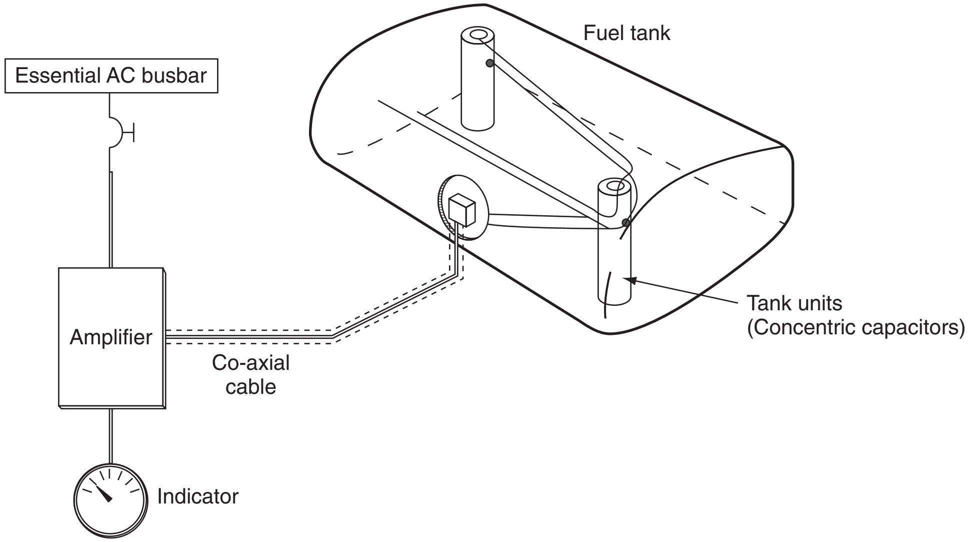

Figure 11.9 Capacitance fuel quantity system – multiple tank units

Key maintenance point

The fuel tank units are factory calibrated and cannot be adjusted on the aircraft.

A tank unit wiring harness attaches to the terminal block of the tank unit. The terminals on the tank wiring harness and the terminal posts have different dimensions to prevent cross-connection during installation. Tank unit end caps are insulated to ensure that the unit does not short to the ground; they also control stray capacitance that forms between the tank unit and ground plane (**capacitance fringing**). Two brackets made of nonferrous glass-filled nylon attach each tank unit to the fuel tank structure.

Intrinsic safety is a technique used for safe operation of electrical and electronic equipment in explosive atmospheres. It is essential that the available electrical and thermal energy in the fuel quantity system is always low enough that ignition of the hazardous atmosphere cannot occur.

Key maintenance point

When working with fuel systems, always use intrinsically safe lamps or torches.

11.3 Fuel feed and distribution

GA aircraft are normally fitted with an **engine-driven pump** (EDP), with electrical **boost pumps** fitted to prime the system during starting. The electrical boost pumps also provide fuel pressure should the EDP fail. A simple fuel pump system comprises an electrically driven boost pump motor controlled by an on/off switch; see Fig. 11.10.

This system is enhanced by a two-stage throttle control system, as illustrated in Fig. 11.11. When the boost pump selector switch is set at the 'low' setting, electrical power is switched through the resistor and the motor runs at a low speed. With the engine running, the selector is moved to the 'high' setting; this provides power through the normally closed (NC) contacts of the throttle micro-switch. When the throttle is set below one-third open, the resistor remains in series, and the motor continues to run at the low speed. When the throttle is advanced, the throttle micro switch changes over via the normally open (NO) contacts to bypass the resistor and apply full power to motor.

The typical fuel feed arrangement on a medium- to large-sized passenger aircraft comprises two booster pumps for each main tank; the motor is located on the tank bulkhead, with the pump located inside the tank. The **fuel distribution** system requires electrical power and is controlled by a panel in the flight compartment as illustrated in Fig. 11.12. Fuel shut-off valves are connected to the battery bus, and controlled by the engine start lever (see engine systems, Chapter 10) and fire handle (see fire protection, Chapter 16). The fuel system normally has the means of transferring fuel between tanks (see fuel transfer); this is controlled by a selector switch that operates a cross-feed valve. Each boost pump is driven by a 115 V AC three-phase motor selected on/off on the control panel via a relay.

The delivery output from each pump feeds into the system via a non-return valve (NRV). Under normal operating conditions, each pump feeds own engine via a motor driven **low-pressure (LP) cock**. If a centre tank is fitted as part of a three-tank instalation, this can feed either engine by a **fuel transfer** system as shown in Fig. 11.13. On some aircraft, tank pumps are located in a dry area of the wing root; on other aircraft the pumps are actually inside the fuel tank. In the latter case, the crew have to maintain certain minimum fuel levels. Control switches for all pumps, cocks and valves together with warning indications are located on the overhead panel or the flight engineer's station. LP cocks are automatically closed if the fire handle is activated (see fire protection, Chapter 16).

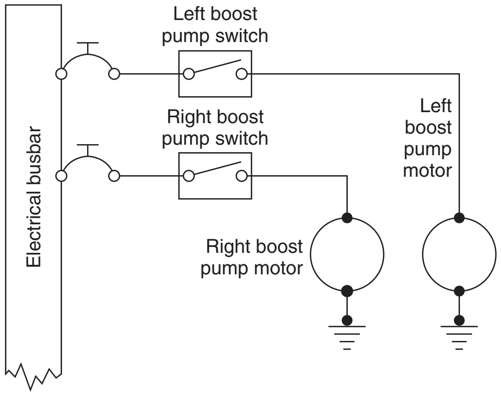

Figure 11.10 Simple fuel pump system

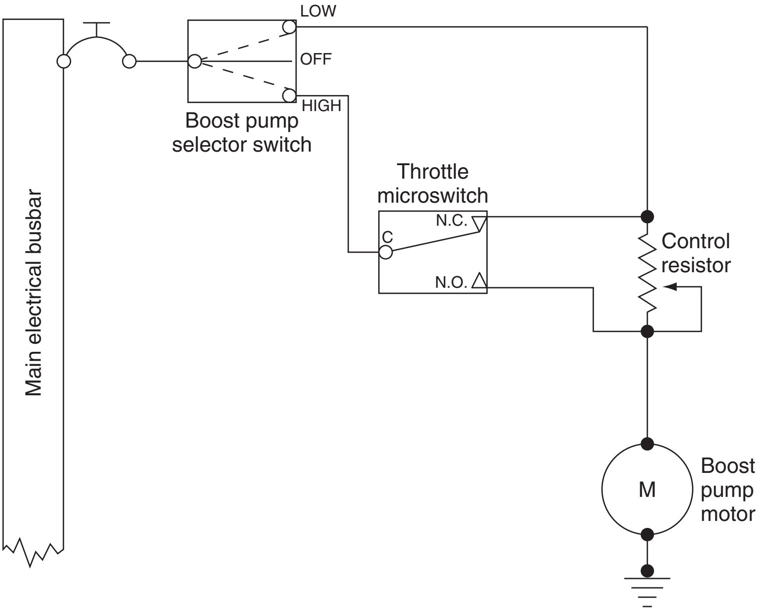
Main electrical busbar
LOW
OFF
HIGH
Boost pump selector switch
Throttle microswitch
C
N.C.
N.O.
Control resistor
M
Boost pump motor

Figure 11.11 Two-stage fuel pump system

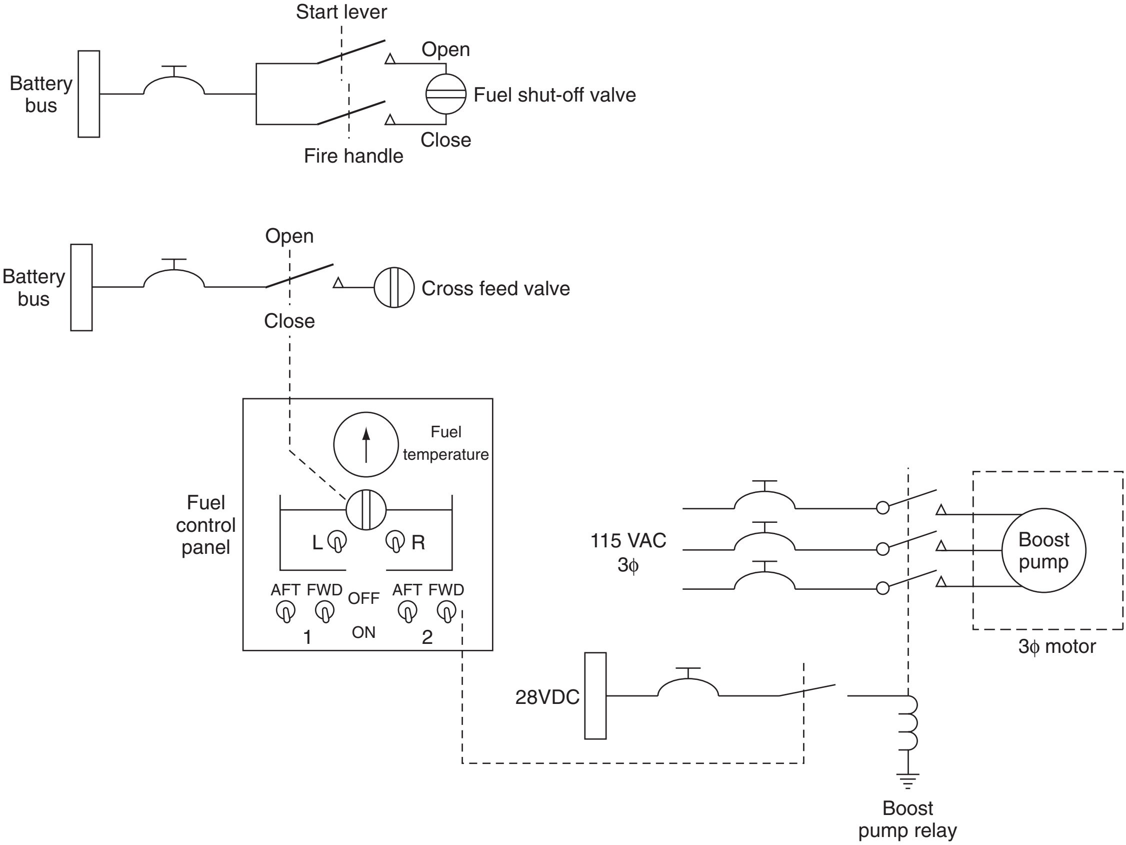
Battery bus
Start lever
Open
Fuel shut-off valve
Close
Fire handle
Battery bus
Open
Cross feed valve
Close
Fuel control panel
Fuel temperature
L
R
AFT FWD
OFF
AFT FWD
1
ON
2
115 VAC 3ϕ
Boost pump
3ϕ motor
28VDC
Boost pump relay

Figure 11.12 Fuel distribution system – electrical power

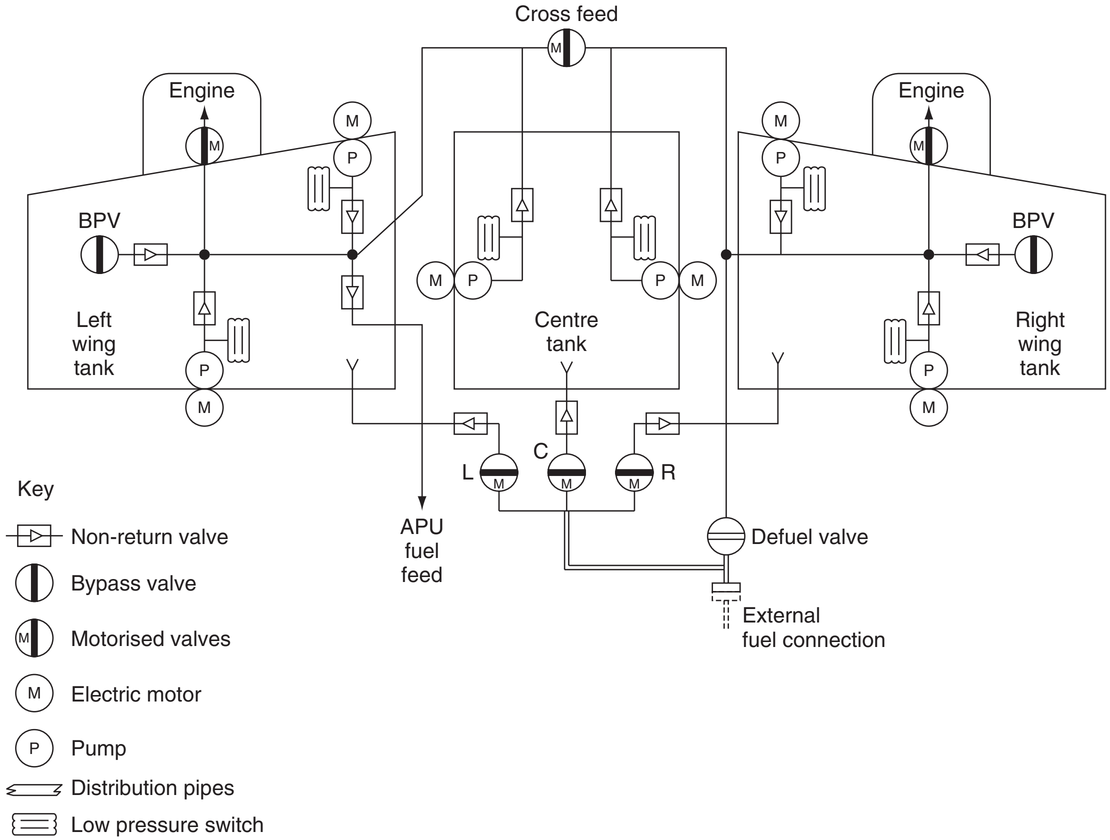
Cross feed
M
Engine
Engine
M
P
M
P
BPV
BPV
M
P
P
M
Left
wing
tank
Centre
tank
Right
wing
tank
P
M
P
M
C
L
R
M
M
M
Key
Non-return valve
Bypass valve
Motorised valves
Electric motor
Pump
Distribution pipes
Low pressure switch
APU
fuel
feed
Defuel valve
External
fuel connection

Figure 11.13 Fuel distribution – three-tank system

Key maintenance point

Leaving a fuel pump switched on with low fuel quantity is an explosion risk since the pump is likely to overheat.

Key maintenance point

If a pump is left running dry for extended periods, typically 10 minutes, it will have insufficient fuel for priming which will render it inoperative when the tank is subsequently refuelled.

11.4 Fuel transfer

This system is used to selectively transfer fuel between tanks; electrically driven fuel pumps and motorized valves are controlled either manually by the crew or by an automatic control system. A simple example of this is illustrated in Fig. 11.13, where fuel can be cross-fed between the left and right tanks. On larger aircraft, the complexity of this fuel transfer system increases. The system comprises a number of motorized valves. Engine valves are activated by the start lever or fire handles. The left, centre and right refuel/defuel valves are operated from an under-wing panel. Bypass valves (BPV) are operated if an electrical pump's fuel filter is blocked. (The BPV senses differential pressure across the filter.) Controls and indications are located on an overhead panel or flight engineer's station. An electrically operated cross-feed valve normally closed unless fuel is being transferred.

It is essential that **fuel temperature** is monitored, either manually or automatically. The maximum fuel temperature is typically +49°C; the minimum fuel temperature is –45ºC or freezing point +3ºC, whichever is higher; the typical freezing point of Jet

A1 fuel is −47ºC. Fuel temperature is measured by an RTD. If the fuel temperature is approaching the lower limits, some fuel could be transferred between tanks; alternatively the aircraft would have to descend into warmer air or accelerate to increase the kinetic heating.

11.5 Refuelling and defuelling

A refuelling control panel and pressure connections are normally located in one or both of the wing areas allowing the fuel to be supplied directly into the main fuel system. A **bonding lead** is always connected between the fuel bowser and aircraft to minimize the risk of static discharge. The fuel tank supply line is connected to the aircraft and pressure applied. Selective control of the system's motorized valves allow specific tanks to be filled as required. **Defuelling** is often required before maintenance, or if the aircraft is to be weighed. The fuel is transferred from the aircraft into a suitable container, typically a fuel bowser. This is achieved via a defuelling valve and applying suction through valves.

11.6 Fuel jettison

When an aircraft takes off fully loaded with passengers and fuel, and then needs to make an emergency landing, it will almost certainly be over its maximum landing weight. Fuel has to be disposed of to reduce the aircraft weight to prepare for the emergency landing. A large aircraft such as the Boeing 747 can be carrying over 100 tonnes of fuel; this is almost 50% of the aircraft's gross weight. One way of burning off fuel and reducing aircraft weight is to fly in a high drag configuration, e.g. 250 knots with the gear down (speed-brakes will further increase the drag). Aircraft can be certified for landings up to the maximum takeoff weight (MTOW) in an emergency; however, overweight landings would only be made if burning off fuel exposed the aircraft to additional hazards. Some aircraft are installed with a **fuel jettison**, or **fuel dumping** system, Fig. 11.14. This provides a means of pumping fuel overboard to rapidly decrease the aircraft's weight. Fuel can normally be jettisoned with landing gear and/or flaps extended. Two **jettison pumps** are installed in each main tank, fuel is pumped via a jettison manifold to nozzle valves located at each wing tip trailing edge.

Key maintenance point

Bonding of fuel system components is essential to ensure intrinsic safety.

11.7 Fuel tank venting

The venting system takes ram air from intakes on the underside of the wing for two specific purposes. When the aircraft is flying, ram air is used to pressurize the fuel to prevent vaporization at lower atmospheric pressures. The ram air also pressurizes the fuel tanks to ensure positive pressure on the inlet ports of each pump. The expansion space above the fuel, called **ullage**, changes with aircraft attitude. Float-operated vent valves are located at key points in the tank to allow fuel to escape into the vent system. Venting tanks (sometimes called **surge tanks**) in the wing tips collect this overspill from the main tanks; the fuel is then pumped back into one of the main tanks. Float-operated vent valves prevent inadvertent transfer of fuel between the tanks.

11.8 Fuel tank inerting

Aircraft have been destroyed due to explosions in empty centre wing fuel tanks. The most likely cause is the centre tank fuel pumps being left running in high ambient temperatures with low fuel quantity. The first centre wing tank (CWT) explosion on a large passenger aircraft occurred on a B707 in 1959; there have been other events in more recent times. In May 1990, shortly after pushback, the centre wing tank (CWT) exploded on a Philippine Airlines Boeing 737, killing eight people. The CWT had not been filled since March 9, 1990. Air-conditioning (A/C) packs had been running on the ground before pushback for approximately 30 to 45 minutes, ambient air was 35°C/95°F. On March 3, 2001, the centre wing tank explosion on Thai Airways International B737 was followed 18 minutes later by an explosion in the right wing tank. Residual fuel was in the CWT. Air-conditioning packs had been running continuously since the aircraft's previous flight, including some 40 minutes on the ground. Ambient air temperature was in excess of 30°C. On July 17, 1996, the CWT exploded on a TWA B747-100 shortly after takeoff from JFK International Airport with fatal consequences. The CWT contained a low amount of residual fuel. The A/C packs had been

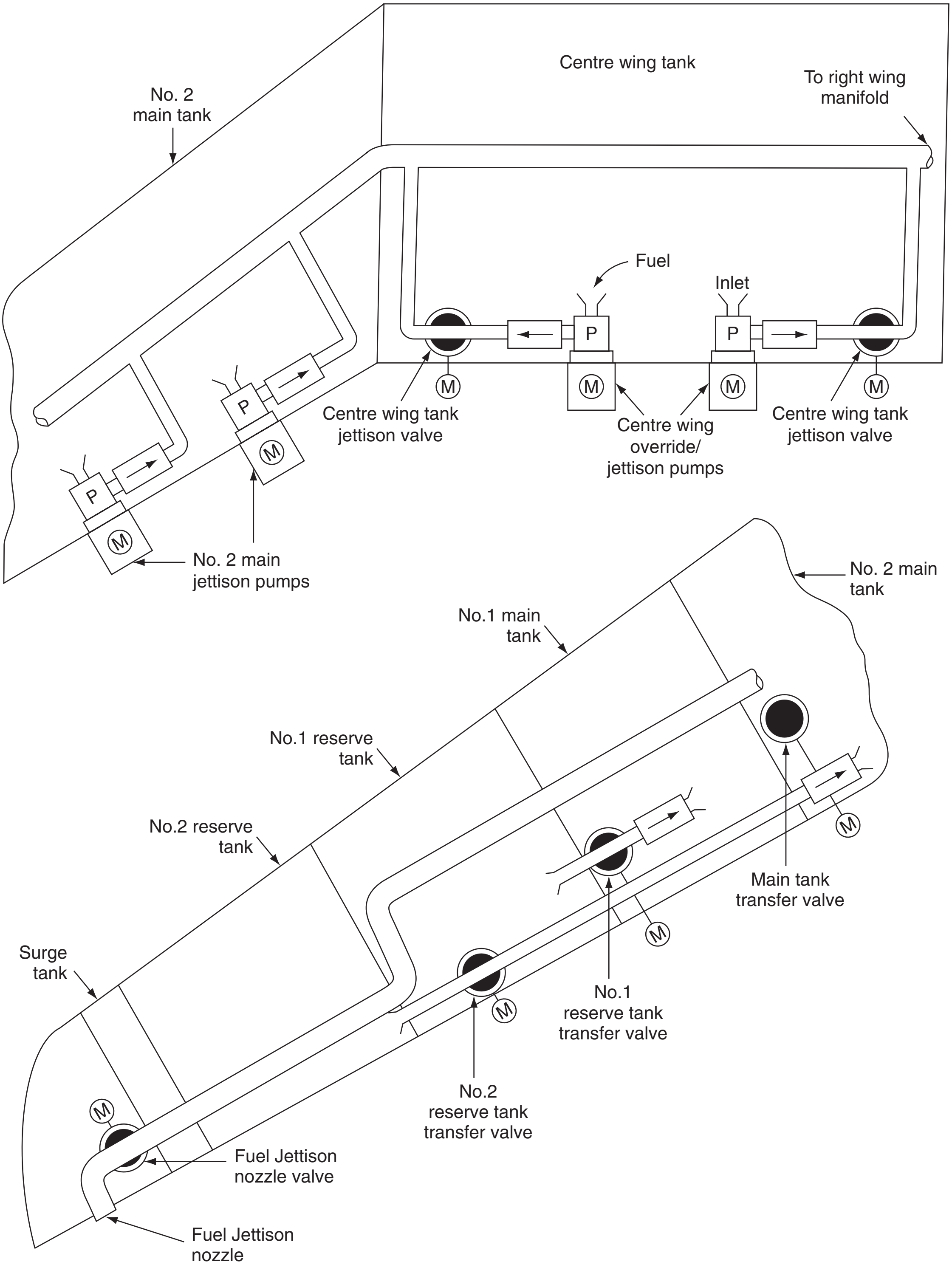
Centre wing tank
To right wing manifold
No. 2 main tank
Fuel
Inlet
P
P
P
P
M
M
M
M
M
M
Centre wing tank jettison valve
Centre wing override/ jettison pumps
Centre wing tank jettison valve
No. 2 main jettison pumps
No. 2 main tank
No.1 main tank
No.1 reserve tank
No.2 reserve tank
Surge tank
M
M
M
M
Main tank transfer valve
No.1 reserve tank transfer valve
No.2 reserve tank transfer valve
Fuel Jettison nozzle valve
Fuel Jettison nozzle

Figure 11.14 Fuel jettison system schematic

running on the ground for 2.5 hours before takeoff, ambient temperature was 28°C/82°F. These three fatal CWT fuel-tank explosions all had certain elements in common:

- all three aircraft had only a small amount of fuel in the wing tank
- the air-conditioning packs, located in non-vented bays directly under the CWT, had been running before the explosions
- the outside air temperatures were quite warm.

A small amount of fuel in the centre wing tank means that there is a large volume for fuel vapour to occupy. Air-conditioning packs generate heat, which contributes to fuel vaporization in the non-insulated tanks. The vaporization is increased with high outside air temperatures.

Empty fuel tanks retain some **unusable fuel** which, in high ambient temperature conditions, will evaporate and create an explosive mixture when combined with oxygen in the ullage. Regulatory mandates have introduced improvements to the design and maintenance of fuel tanks to reduce the chances of such explosions. These improvements include the enhanced design of the FQIS and fuel pumps, inspection regimes for wiring in the tanks and insulating fuel tanks that are in close proximity to sources of high temperature. The longer-term solution that will contribute more towards fuel tank safety is **inerting**; possible solutions include:

- ground-based inerting
- on-board ground inerting
- on-board inert gas generation system (OBIGGS)
- liquid nitrogen from a ground source.

For ground-based inerting, the fuel tanks would receive an amount of nitrogen-enriched air before pushback. Inerting would last for the taxi, takeoff and climb phases of flight, when the fuel vapours would be warmest. On-board ground inerting achieves the same objective; however, the inerting equipment is an aircraft system.

Although OBIGGS offers tremendous benefits, it is expensive. Nitrogen generating systems (NGS) have been developed which decrease the flammability risks of the tanks. The NGS is an onboard inert gas system; external (atmospheric) air is directed into an air separation module (ASM), this separates out the oxygen and nitrogen via a **molecular sieve**. After separation, the nitrogen-enriched air (NEA) is supplied into the centre wing tank and the oxygen-enriched air (OEA) is vented overboard. NEA decreases the oxygen content in the ullage to prevent combustion. Trials have determined that an oxygen level of 12% is sufficient to prevent ignition; this is achievable with one ASM on a medium-sized aircraft and up to six ASMs on larger wide-bodied aircraft.

In the fourth potential method, the aircraft would receive a supply of **liquid nitrogen** from a ground source. The liquid nitrogen would be stored on the aircraft in a vacuum-sealed insulated container, and would be fed to the fuel tanks under low pressure. Oxygen sensors in the fuel tanks provide feedback to the system's computer control function; the supply of liquid nitrogen is regulated to keep the fuel tanks inerted for critical phases of flight. The liquid nitrogen could also be used to supplement fire suppression in other areas of the aircraft. A computer would monitor temperature in the fuel tanks to determine if the ullage is within a flammable temperature range. This system would know the tank temperature at all times, together with the oxygen content of the tanks.

There remains an on-going industry debate regarding the effectiveness and cost of inerting. There needs to be a practical means of inerting fuel tanks for in-service and new-production civil transports.

11.9 Multiple choice questions

1. The volume of fuel in a tank varies with temperature; as the temperature changes:
    (a) the mass and volume of fuel remains the same
    (b) the mass of fuel changes but the volume remains the same
    (c) the mass of fuel remains the same, but the volume changes.

2. In the DC ratiometer fuel quantity circuit:
    (a) two opposing magnetic fields are created in each of the coils
    (b) three opposing magnetic fields are created in each of the coils
    (c) two complementing magnetic fields are created in each of the coils.

3. Air and fuel have dielectrics of approximately:
    (a) unity and zero respectively
    (b) unity and two respectively
    (c) two and unity respectively.

4. Fuel tank ullage is the:
    (a) expansion space above the fuel, and changes with aircraft attitude

   (b) unusable fuel, and changes with aircraft attitude
   (c) expansion space above the fuel, remaining constant with aircraft attitude.

5. When fuel level decreases, the capacitance of the fuel quantity sensor:
   (a) decreases and the reactance increases
   (b) increases and the reactance increases
   (c) increases and the reactance decreases.

6. Defuelling is often required:
   (a) following aircraft maintenance
   (b) before maintenance, or if the aircraft is to be weighed
   (c) after the aircraft has been weighed.

7. Under-wing fuel quantity measurements are used during:
   (a) level flight
   (b) all flight conditions
   (c) ground servicing only.

8. Intrinsic safety is a technique used for:
   (a) safe operation of electrical/electronic equipment in explosive atmospheres
   (b) ensuring fuel temperature does not become too low/high
   (c) reducing fuel quantity.

9. Fuel tank inerting is a system where:
   (a) nitrogen-enriched air (NEA) is vented overboard
   (b) nitrogen-enriched air (NEA) is supplied into the centre wing tank
   (c) oxygen-enriched air (OEA) is supplied into the centre wing tank.

10. Increased fuel density is a result of:
   (a) increased temperature and increased capacitance
   (b) reduced temperature and increased capacitance
   (c) reduced temperature and reduced capacitance.

Chapter 12 Lights

Lighting is installed on aircraft for a number of reasons including: safety, operational needs, servicing and for the convenience of passengers. The applications of aircraft lights can be broadly grouped into four areas: flight compartment (cockpit), passenger cabin, exterior and servicing (cargo and equipment bays). There are many types of lighting technologies used on aircraft. Lights are controlled by on/off switches, variable resistors or by automatic control circuits. This chapter will review each of these lighting technologies and the type of equipment used in specific aircraft applications.

12.1 Lighting technologies

Aircraft lighting is based on a number of technologies:

- incandescence
- light-emitting diode (LED)
- electro-luminescent
- fluorescence
- strobe.

Incandescence is the radiation of light from an electrical filament due to an increase in its temperature. The filament is a small length of wire, e.g. tungsten, which resists the flow of electrons when a voltage is applied, thereby heating the filament. Tungsten can be drawn into a very thin wire filament and has a very high melting point (3,659 K). The electron flow creates a voltage drop that heats the filament to a temperature where radiation is emitted in the visible spectrum.

Electro-luminescence is a combined optical and electrical phenomenon that causes visible light to be emitted. This can be achieved with electron flow through a semi-conductor material, or by a strong electric field applied across a phosphor material. Electro-luminescence is the effect of recombining electrons and holes in a **light-emitting diode (LED)** or phosphor material. The electrons are imparted with an external energy source and release their own energy as photons; this is radiated as light energy. In the LED semiconductor, electrons and holes are separated by a doping process to form the p-n junction. With the electro-luminescent phosphor display, electrons are imparted with energy by the impact of high-energy electrons that are accelerated by a strong electric field.

Fluorescent lamps are gas-discharge devices formed from a sealed tube of glass that is coated on the inside with phosphor; the glass tube contains mercury vapour mixed with an inert gas, e.g. argon or neon. The lamp uses a high voltage to energize the mercury vapour; this results in an **ionized gas** where the electrons are separated from the nucleus of their atoms creating **plasma**. The release of energy causes the phosphor coating to fluoresce, i.e. it produces visible light. Fluorescent lamps require a ballast resistor to regulate the flow of energy in the tube.

Strobe lights are formed from small diameter (typically 5 mm) sealed quartz or glass envelope/tube filled with xenon gas, see Fig. 12.1. Power from the aircraft bus is converted into a 400 V DC supply for the strobe. The tube is formed into the desired shape to suit the installation; Fig. 12.2 is a wing-tip anti-collision light, normally located behind a clear plastic protective cover. **Xenon** is an inert (or noble) gas, chemically very stable, and has widespread use used in light-emitting devices, e.g. aircraft anti-collision lights. The emission of light is initiated by ionizing the xenon gas mixture by applying a high voltage across the electrodes.

Figure 12.3 illustrates a typical strobe light circuit; this comprises a low-voltage power supply and a 400 V DC power pack that provides a dual power supply. Shielded cables are required between the power pack and strobes to minimize the effect of electromagnetic interference (EMI). A high-energy pulse of current is triggered through the ionized gas on a cyclic basis. The ionization decreases the electrical resistance of the gas such that a pulse in the order of thousands of amperes is conducted through the gas.

When this current pulse travels through the tube, it imparts energy into the electrons surrounding

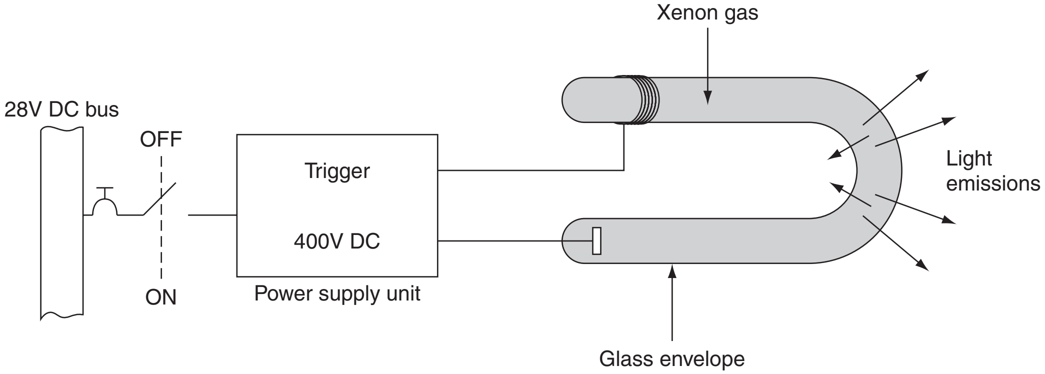

Figure 12.1 Strobe/xenon gas tube schematic

Figure 12.2 Strobe/xenon gas tube product

the xenon atoms causing them to move up to higher energy levels. The electrons immediately drop back into to lower energy levels, thereby producing photons. Strobes are a source of short-wavelength ultraviolet radiation together with intense emissions in the near infrared; the overall effect is a high-intensity white flash.

Key maintenance point

400V DC used in strobe circuits is dangerous: take all necessary health and safety precautions when working near the system.

Key maintenance point

Do not handle strobe tubes with bare hands; moisture causes local hot-spots that can lead to premature failure.

Test your understanding 12.1

Briefly explain the difference between the following lighting technologies:

- incandescent
- light-emitting diode (LED)
- electro-luminescent
- fluorescent
- strobe.

12.2 Flight compartment lights

Lighting is needed for the illumination of instruments, switches and panels. **Dome lights** located on the ceiling provide non-directional distribution of light in the compartment; it typically contains an incandescent lamp and is powered from the battery or ground services bus. **Flood lighting** in the flight compartment from incandescent lamps and/or fluorescent tubes provides a general illumination of instruments, panels, pedestals etc. Fluorescent tubes located beneath the glareshield provide overall illumination of the instrument panels. **Emergency lights** are installed in the flight compartment for escape purposes. The colour of flight compartment lights is normally white; this

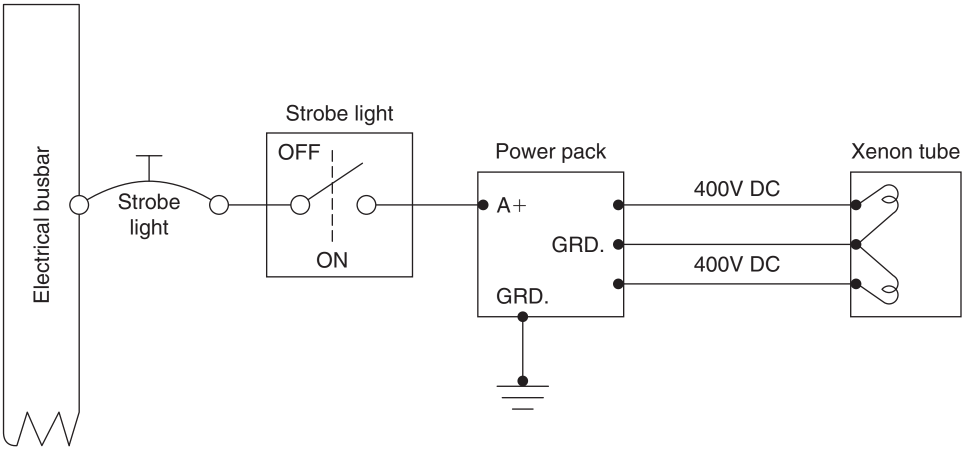
Electrical busbar
Strobe light
Strobe light
OFF
ON
Power pack
A+
GRD.
GRD.
400V DC
400V DC
Xenon tube

Figure 12.3 Strobe/xenon gas tube circuit

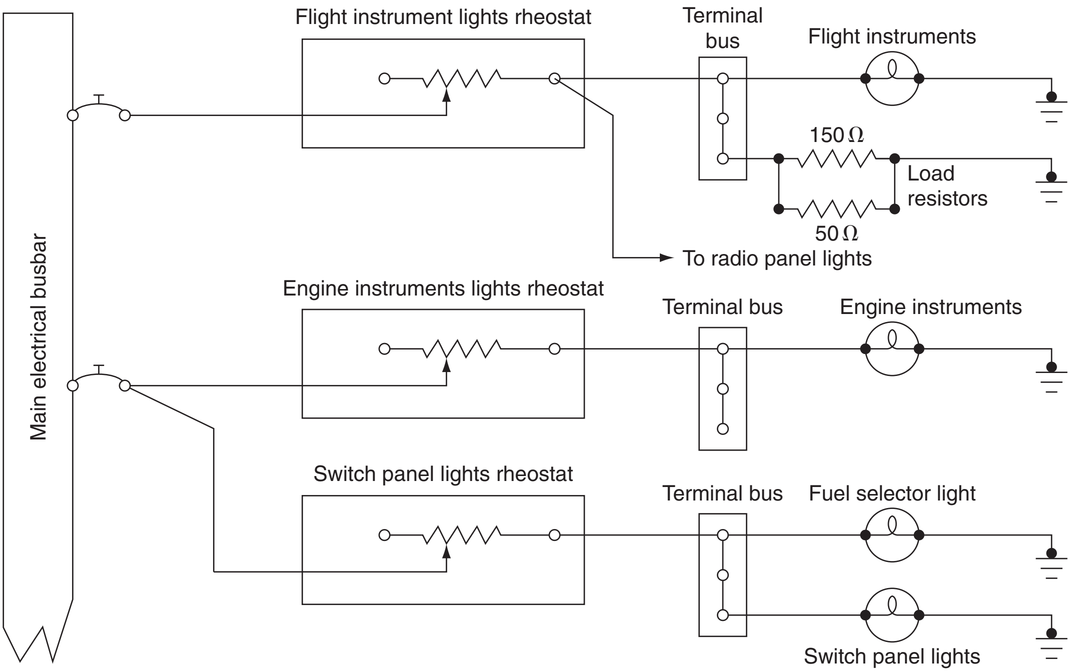
Main electrical busbar
Flight instrument lights rheostat
Terminal bus
Flight instruments
150 Ω
Load resistors
50 Ω
To radio panel lights
Engine instruments lights rheostat
Terminal bus
Engine instruments
Switch panel lights rheostat
Terminal bus
Fuel selector light
Switch panel lights

Figure 12.4 Rheostat light intensity control

reduces the power and heat, improves contrast on the instruments, and reduces eye fatigue.

12.2.1 Instruments

Internal instrument lighting is normally from incandescent lamps integrated within individual instruments; lighting must be shielded from causing any direct glare to the pilot and must be dimmable. External instrument lighting is provided by pillar (or bridge) lights positioned on the panels for individual instruments. The light intensity can be dimmed by a simple **rheostat** device as illustrated in Fig. 12.4; this

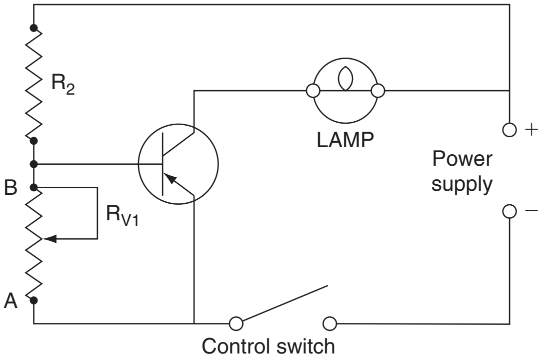

Figure 12.5 Transistor lighting control circuit

typical circuit is for flight instruments, engine instruments and switch panels.

The increased quantity of instruments requires electronic control in place of rheostat due to the higher loads. A **transistor circuit** provides electronic control as illustrated in Fig. 12.5; variable resistor R_{V1} varies the (relatively low) base current into the base of a PNP transistor; this controls the (relatively high) current through the collector and lamp. A typical transistor controlled lighting system is illustrated Fig. 12.6. The relatively low base currents in the respective transistors can now control a variety of lighting circuits:

- radio navigation systems
- compass
- fuel panels
- engine indications.

Instrument panels are often constructed from Perspex; the surface is painted and then engraved with the identification of switches and controls; the panel is illuminated from the edges. The light is dispersed through the panel, but is only seen through the engravings. Alternatively, electro-luminescent panels are used; these are AC powered and energy efficient. Referring to Fig. 12.7, the phosphorus layer is laminated between front and rear clear plastic layers. The phosphorus material glows when AC power is applied; the front of the panel is painted to match the colour of other panels. Engraved lettering or symbols remain clear and transmit light from the glowing phosphorus layer.

12.2.2 Master warning

An increasing number of systems are being designed into aircraft; this leads to more warning lights and larger panels with an increased possibility of a warning light being missed by the crew. This has led to centralized 'attention getters', or **master warning** and **caution light** panels, a typical arrangement is shown in Fig. 12.8. The lights are located within pilot's immediate view, typically on the glareshield. Master caution and warning systems were developed to ease pilot workload, particularly on aircraft designed for operation without a flight engineer. In addition to the attention getter, it also directs the pilot towards the problem area concerned. The system annunciators are usually arranged such that the cautions are in the same orientation as the main system panels, e.g. the overhead panel.

A simplified master warning circuit is illustrated in Fig. 12.9. In this example, switch contacts 1 and 2 represent the individual warning signals from two different systems. In the event of a failure or malfunction, closure of either switch causes its corresponding warning light to illuminate together with the master warning (MW) light. The master warning and caution light is cancelled by pressing the light caption, whilst the crew react and investigate the reason for the warning. Typical panels could have up to 50 individual warning lights, any one of which also illuminates the master warning light. The individual lights could be located on an overhead or side panel.

Key point

A diode is required in the warning light circuit to ensure that only the relevant system light is illuminated.

When the master warning or caution light is illuminated, the pilot cross-refers to a centralized group of warning lights on the relevant panel, each connected to the warning devices of specific systems. The individual systems are identified by the system name and are located within the pilot's scan. Warning lights can be tested by a separate **test** switch, or by a centralized master dim and test switch. The night/day switch is used to reduce the intensity of warning lights during low ambient lighting conditions. Referring to Fig. 12.10(a), warning and caution lights affecting system operation and aircraft safety are defined by specific colours:

- **warning**, red, an unsafe condition exists
- **caution**, amber, an abnormal condition exists, but it is not unsafe

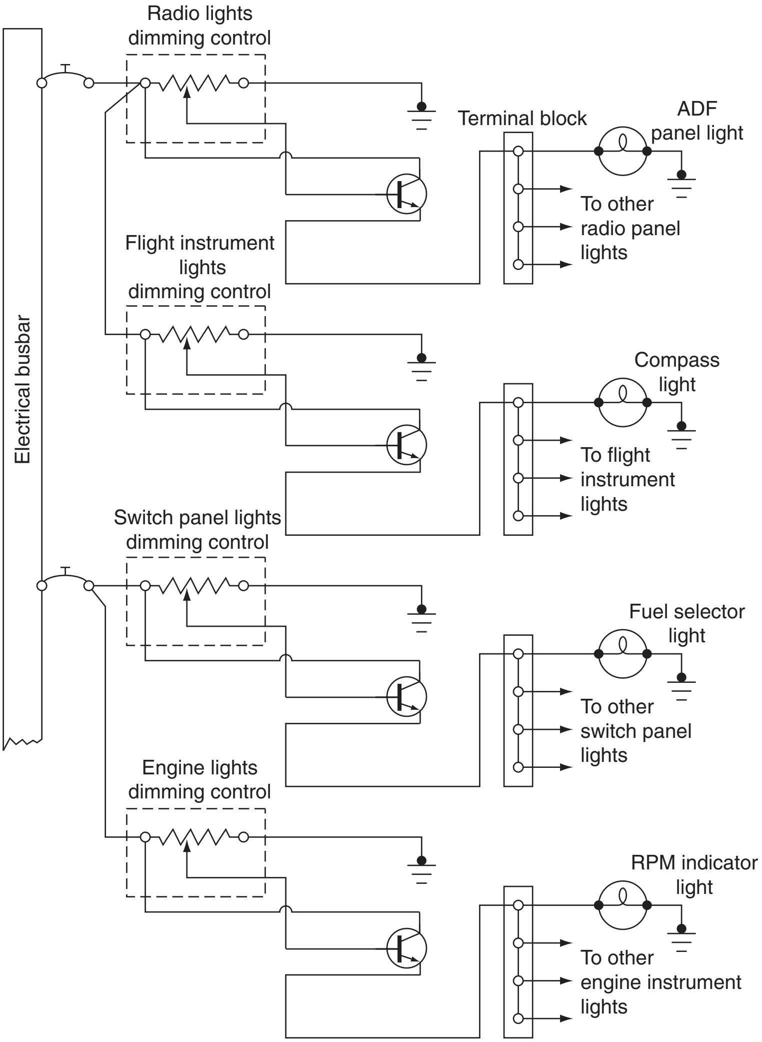
Radio lights
dimming control
Terminal block
ADF
panel light
To other
radio panel
lights
Flight instrument
lights
dimming control
Electrical busbar
Compass
light
To flight
instrument
lights
Switch panel lights
dimming control
Fuel selector
light
To other
switch panel
lights
Engine lights
dimming control
RPM indicator
light
To other
engine instrument
lights

Figure 12.6 Transistor lighting control system

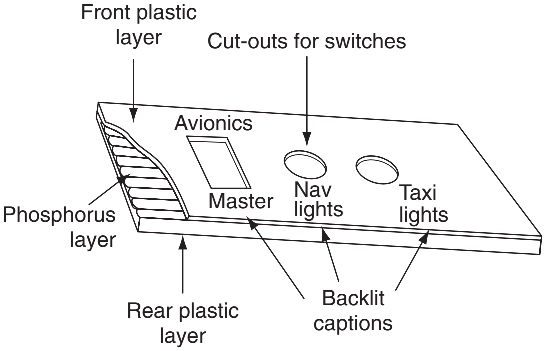
Front plastic
layer
Cut-outs for switches
Avionics
Master
Nav
lights
Taxi
lights
Phosphorus
layer
Rear plastic
layer
Backlit
captions

Figure 12.7 Phosphorus lighting panel

- **advisory**, green or blue, a safe condition exists, or for information e.g. gear down.

Some installations have a comprehensive master warning and caution light panel that occupies the entire upper instrument panel, see Fig. 12.10b.

12.2.3 Emerging technology

Digital technology is replacing traditional wire bundles from the flight compartment to the equipment bays. A digital processor on each warning and

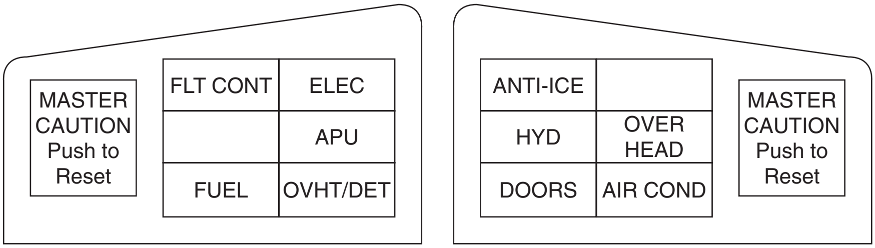

Figure 12.8 Master warning and caution lights

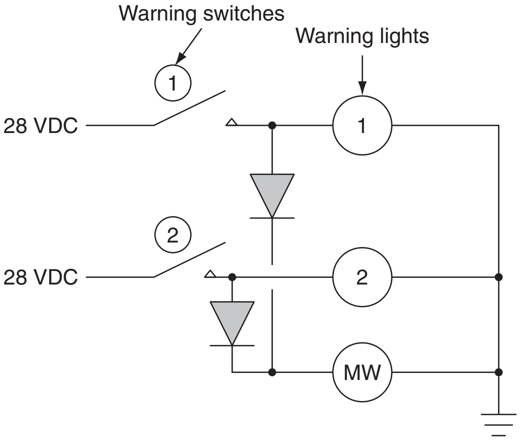

Figure 12.9 Simplified master warning circuit

caution module replaces lighted switches with the associated wire bundles. LEDs can now provide sufficient brightness to replace incandescent lighting with 30% power reduction compared with fluorescent lighting. The objectives are to accomplish weight, cost and reliability goals while giving flight crews what they are used to seeing in older flight compartments, including the tactile feel for push buttons and controls for lighting and dimming.

Key point

Lights used for certain warnings are not dimmable, e.g. fire and overheat. This is to ensure that the warnings are not missed under bright ambient lighting conditions.

12.3 Passenger cabin lights

Interior lighting installations for the passenger cabin vary depending on the size of aircraft; this ranges from a small quantity of roof-mounted incandescent lamps, through to integrated lighting concealed within the interior trim. These lights are controlled from the flight attendants' station. LED illumination is being specified on business and passenger aircraft that have pre-programmed settings for specific flight phases and time zones. The systems are automatically controlled to customize the mixing of colours and lighting levels; this is intended to help passengers combat the fatigue of long-distance travel.

Traditionally, each cabin light is controlled individually; this technology is being replaced by a central control unit connected to all the lights in the cabin. Those lights are linked via a data bus to their respective control units; each of the lights can be programmed for different scenarios.

Implemented through software, the lighting in the cabin is a mixed array of colours. LED technology offers higher reliability and reduced maintenance costs compared to incandescent and fluorescent lights. Cabin signs, e.g. 'return to seat' or 'no-smoking', are normally activated by the flight crew; on some aircraft the lights are armed by the crew and then activated automatically. Passenger reading lights are controlled from individual seat controls.

Additional entry floodlights are provided in the door areas. **Exit** lights are located adjacent to the emergency exits and are clearly visible, irrespective of whether the door is open or closed. **Floor path lighting** is used to in emergency situations to provide visual identification of escape routes along cabin aisle floor. These systems have sufficient energy to enable passengers to identify aisle boundaries. The system guides the passengers to designated **emergency exits**

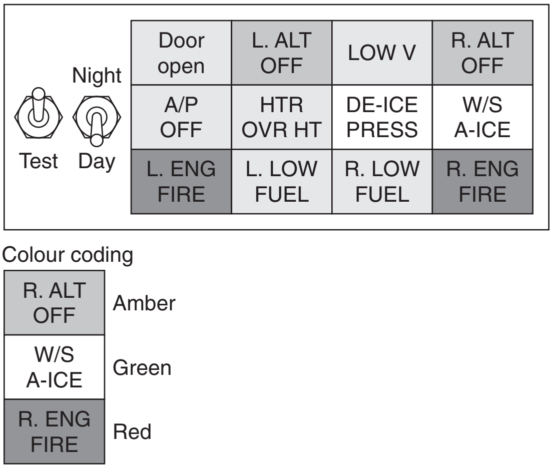
Night
Test
Day

Door open	L. ALT OFF	LOW V	R. ALT OFF
A/P OFF	HTR OVR HT	DE-ICE PRESS	W/S A-ICE
L. ENG FIRE	L. LOW FUEL	R. LOW FUEL	R. ENG FIRE

Colour coding

R. ALT OFF	Amber
W/S A-ICE	Green
R. ENG FIRE	Red

(a)

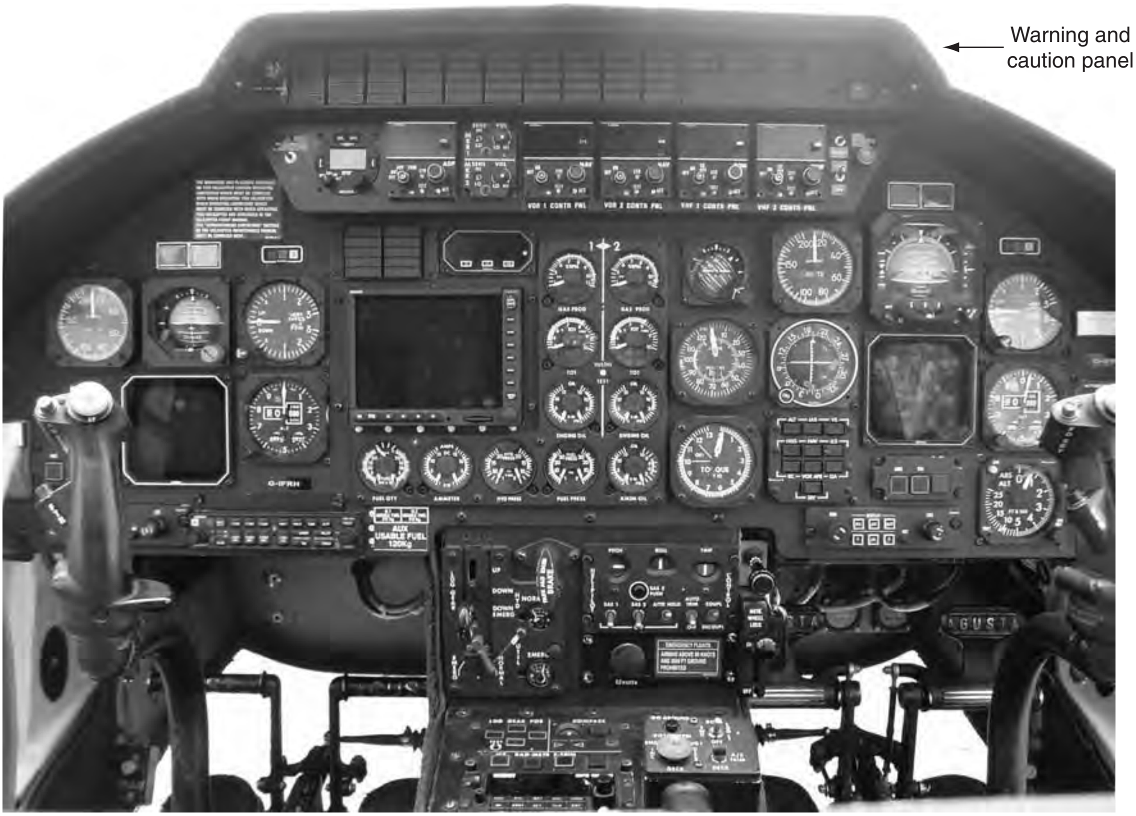
Warning and caution panel

(b)

Figure 12.10 Warning and caution panels: (a) typical arrangement, (b) helicopter installation

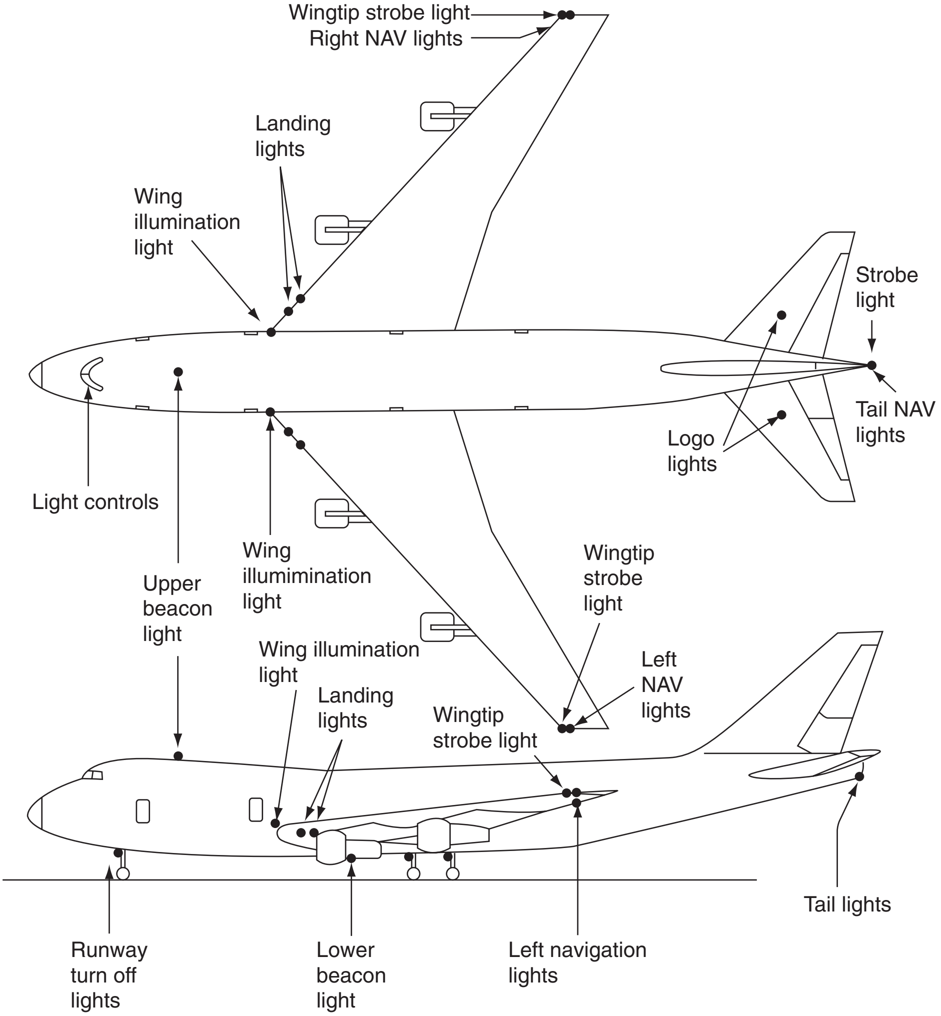
Wingtip strobe light
Right NAV lights
Landing lights
Wing illumination light
Strobe light
Tail NAV lights
Logo lights
Light controls
Upper beacon light
Wing illumimination light
Wingtip strobe light
Wing illumination light
Landing lights
Wingtip strobe light
Left NAV lights
Tail lights
Runway turn off lights
Lower beacon light
Left navigation lights

Figure 12.11 Exterior lights (large transport aircraft)

in accordance with the specific certification requirements of the aircraft.

12.4 Exterior lights

An overview of the exterior lighting arrangement on a large passenger aircraft is illustrated in Fig. 12.11. Exterior lighting is used for:

- logo illumination
- landing/taxiing
- wing illumination
- anti-collision/navigation.

12.4.1 Logo lights

Referring to Fig. 12.12, **logo lights** are used to illuminate the tail fin; this is primarily for promotional purposes, i.e. for the airline to highlight their logo during night operations at an airport. Apart from the advertising value at airports, they are often used for additional awareness in busy airspace. **Taxi lights** (or runway turn off lights) are sealed beam devices with 250 W filament lamps located on the nose, landing gear or wing roots. They are sometimes combined with the landing light and used when approaching or leaving the runway. Taxi lights improve visibility during ground operations; they are directed at higher

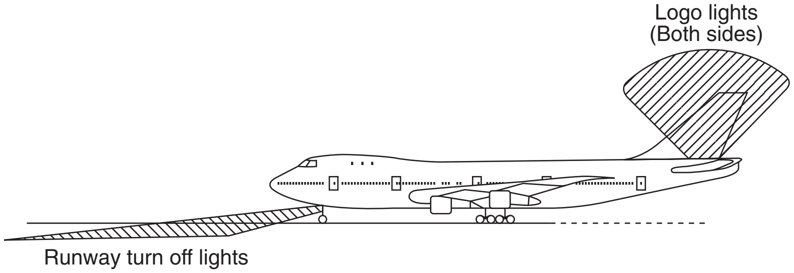

Figure 12.12 Exterior lights (turn off/logo)

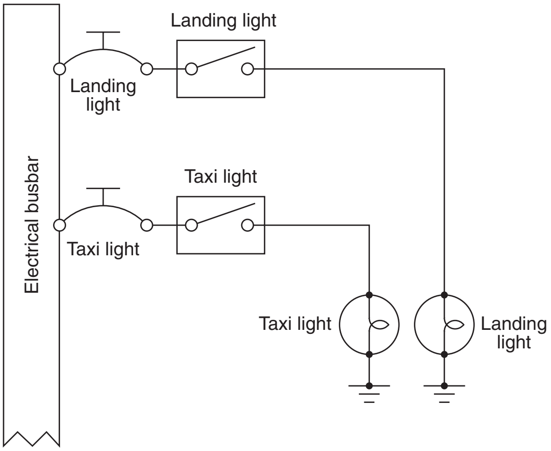

Figure 12.13 Fixed landing light circuit

angle than landing lights. Runway turnoff lights, in the wing roots, are normally only used at night on poorly lit runways. A typical fixed landing and taxi light circuit is shown in Fig. 12.13. On some aircraft types, taxi lights will switch off automatically with gear retraction. It is common practice to have taxi lights on whilst the aircraft is in motion as a warning to other aircraft and vehicles.

12.4.2 Landing lights

Landing lights are located on the wing tips, or on the front of the fuselage, usually at fixed angles to illuminate the runway. They are sealed beam devices with 600–1000 W filament lamps; a parabolic reflector concentrates light into a directional beam. The high current requirement is controlled via a relay. Some landing light installations have a retractable assembly located on the underside of the wing. This has a reversible motor and gear mechanism to drive the light out against the airflow; a typical circuit is illustrated in Fig. 12.14. The alternative location for a landing light is in the wing leading edge; this has a transparent cover to provide aerodynamic fairing. Inboard and outboard landing lights (Fig. 12.15) provide extended illumination of the landing area.

12.4.3 Wing illumination

Ice inspection lights (Fig. 12.15) are often installed to check ice formation on wing leading edges and engine intakes. Typical lights are the sealed beam type with filament lamps of 50–250 watts. They are recessed into the fuselage or engine nacelle with a preset direction that illuminates a section of the wing that can be viewed from the flight compartment. (More details on ice detection are given in Chapter 15.)

12.4.4 Service lights

Service lights are provided throughout the aircraft as illustrated in Fig. 12.16. These lights are powered from the aircraft ground servicing bus. Examples include:

- cargo bays
- wheel wells
- equipment bays
- fuelling panels.

Wheel wells lights are normally only used during the turnaround at night during the pre-flight inspection; they can also be used on to see the mechanical gear down-lock indications at night.

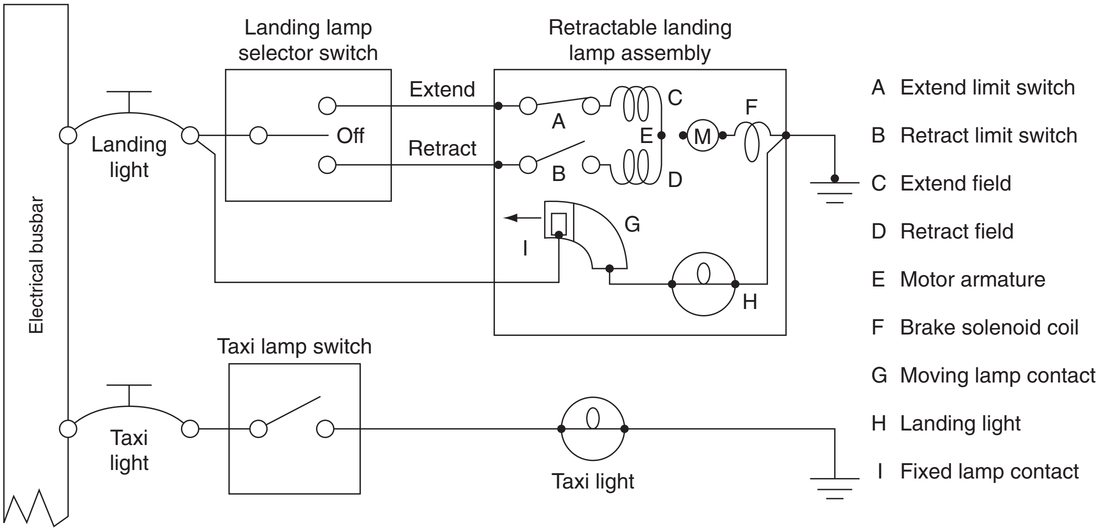
Landing lamp selector switch
Retractable landing lamp assembly
Extend
Off
Retract
Landing light
Electrical busbar
A
B
C
D
E
M
F
G
I
H
Taxi lamp switch
Taxi light
Taxi light
A Extend limit switch
B Retract limit switch
C Extend field
D Retract field
E Motor armature
F Brake solenoid coil
G Moving lamp contact
H Landing light
I Fixed lamp contact

Figure 12.14 Retractable landing/taxi light circuit

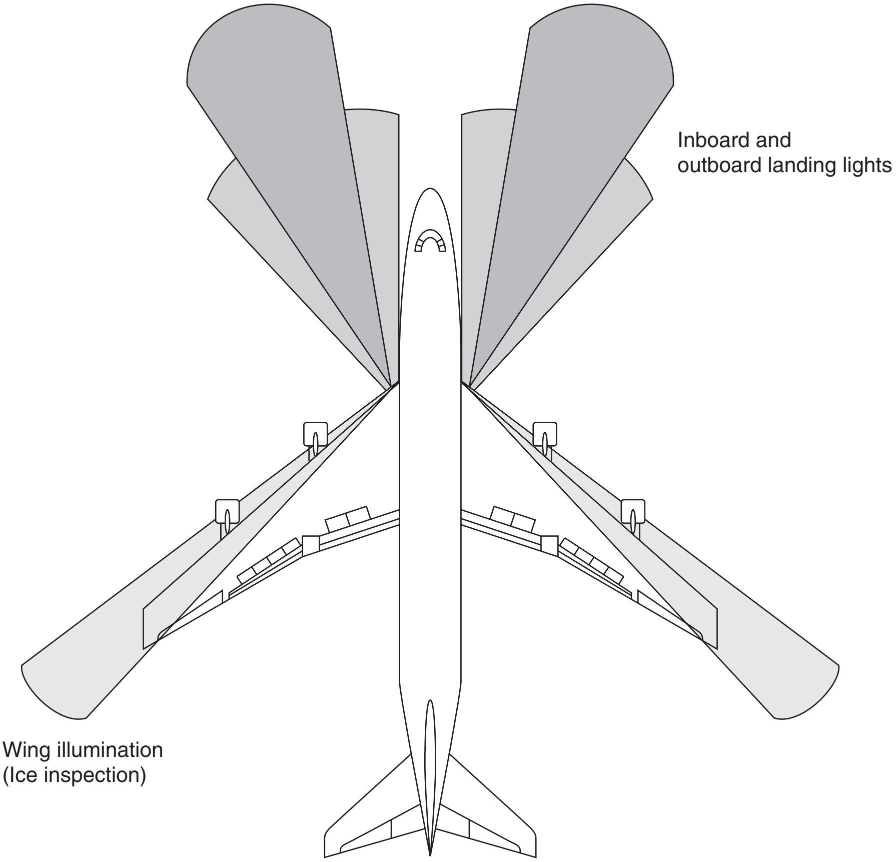
Inboard and outboard landing lights
Wing illumination (Ice inspection)

Figure 12.15 Exterior lights (landing/wing)

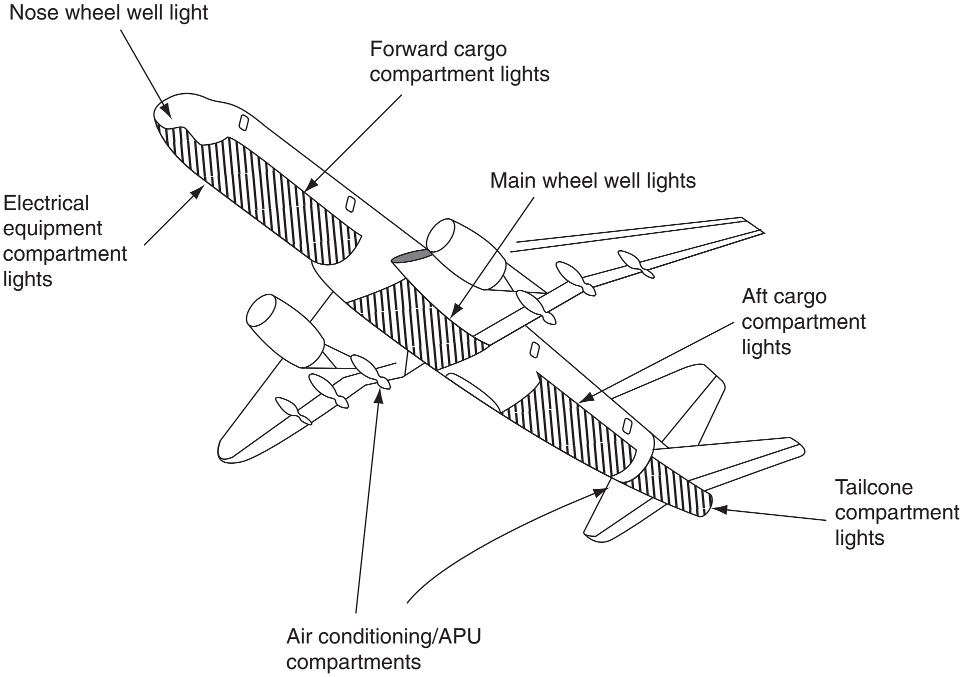
Nose wheel well light
Forward cargo compartment lights
Main wheel well lights
Electrical equipment compartment lights
Aft cargo compartment lights
Tailcone compartment lights
Air conditioning/APU compartments

Figure 12.16 Exterior lights (servicing)

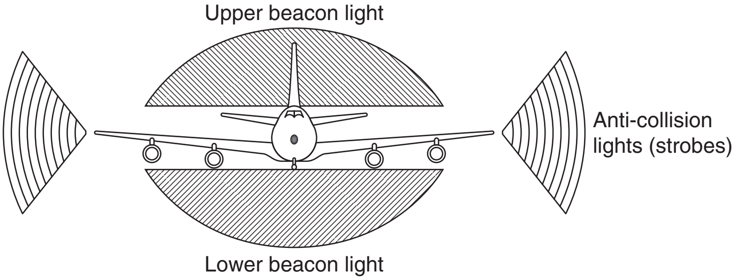
Upper beacon light
Anti-collision lights (strobes)
Lower beacon light

Figure 12.17 Exterior lights (beacons/strobes)

12.4.5 Navigation lights

The primary external lights required for navigation purposes are the **beacons** and **anti-collision** lights, see Fig. 12.17.

12.4.5.1 Navigation lights

The **navigation** (or position) lights are a legal requirement for night flying. Navigation lights are normally based on filament lamps, providing steady illumination. They are located at the extremes of the aircraft, see Fig. 12.18, and provide an indication of the aircraft's direction and manoeuvres. Navigation lights are based on regulations that define the colour, location and beam divergence such that the aircraft is visible from any viewing angle; these colours and divergence angles are:

- **green**, starboard wing, divergence of 110 degrees
- **red**, port wing, divergence of 110 degrees
- **clear** (white), tail, divergence of ±70 degrees either side of aircraft centreline (140 degrees total).

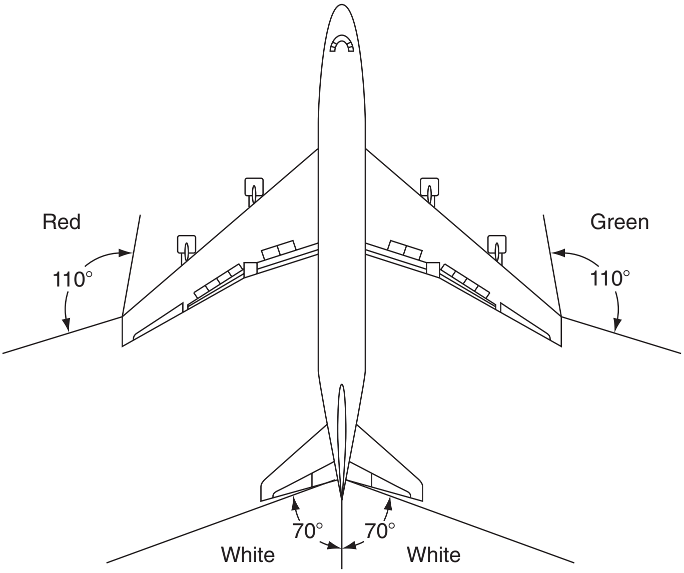
Red
110°
Green
110°
70°
70°
White
White

Figure 12.18 Navigation lights – angular coverage

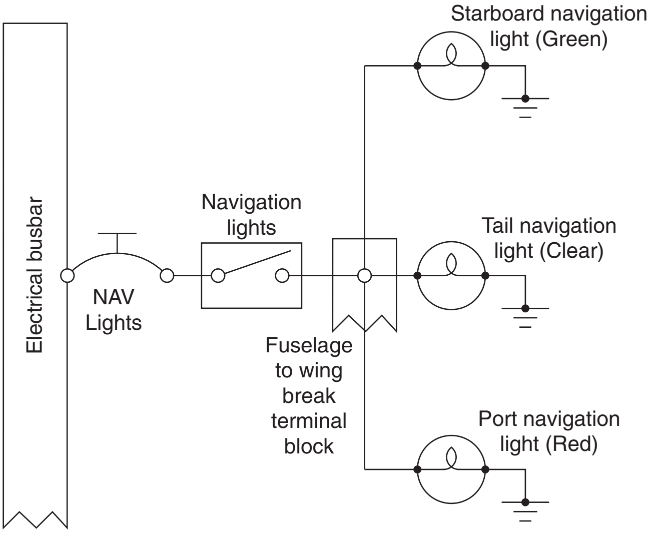
Starboard navigation light (Green)
Electrical busbar
Navigation lights
Tail navigation light (Clear)
NAV Lights
Fuselage to wing break terminal block
Port navigation light (Red)

Figure 12.19 Navigation light circuit

The traditional location of the white light is on the tail cone or fin tip; some aircraft have the rear facing light on the trailing edge of each wing tip. The wing lamps are 20W filament lamps, the tail lamp is 10W. Coloured filters produce the specific colours; these filters must not shrink, fade or become opaque. A typical navigation light circuit is shown in Fig. 12.19. Note that the lights are controlled by single switch and protection device.

Some aircraft are installed with LED position lights; these are formed with a bank of LEDs, see Fig. 12.20. Helicopters have varying navigation light installations due to their specific geometry, see Fig. 12.21.

12.4.5.2 Anti-collision lights

Anti-collision lights often supplement navigation lights; these can be provided either by a strobe

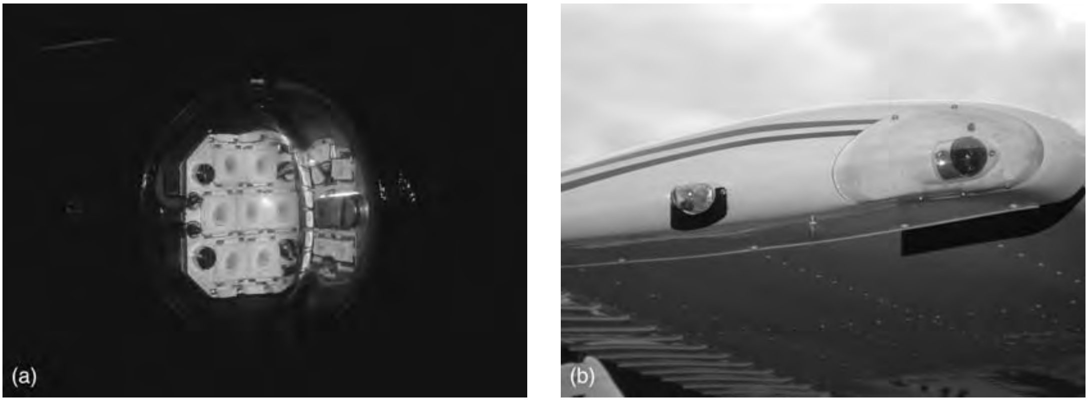

Figure 12.20 Position lights: (a) LED type; (b) wing tip location

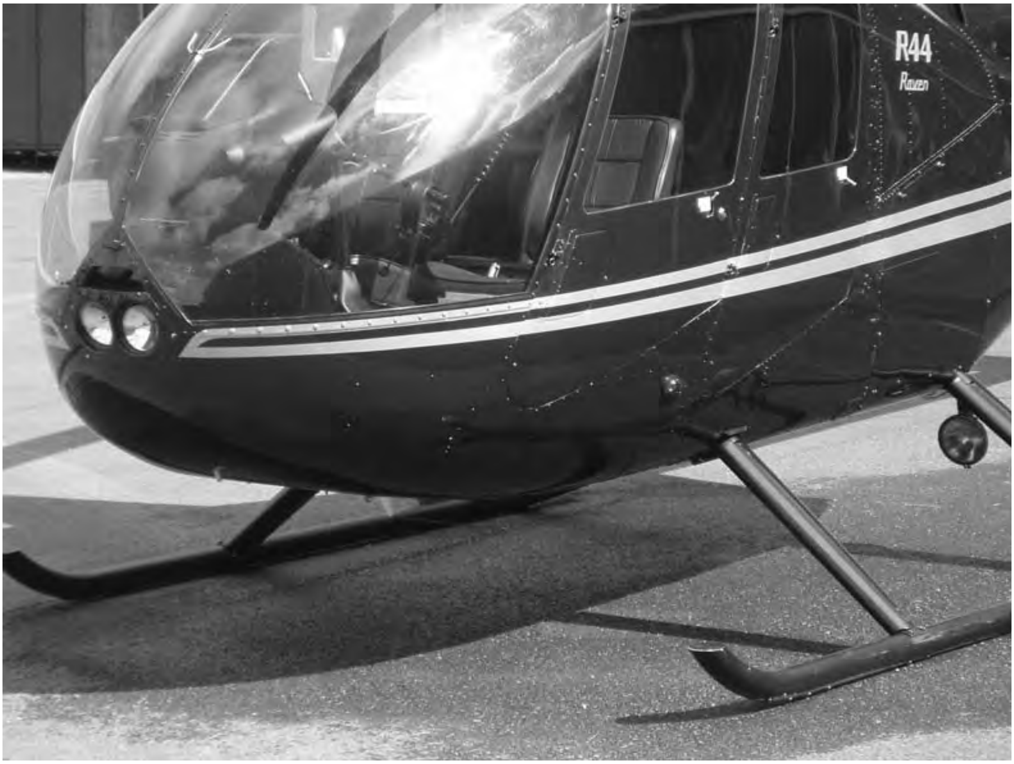

Figure 12.21 Helicopter lights: Twin landing lights on nose; Navigation light below door; Additional landing light on rear skid

light, rotating beacon or a combination of both, see Fig. 12.22. Anti-collision lights are also used as a warning that the engines are running or are about to be started. They are typically not switched off until it is considered safe for ground personnel to approach the aircraft. Strobe lights are typically located on the:

- vertical stabilizer
- wing tips

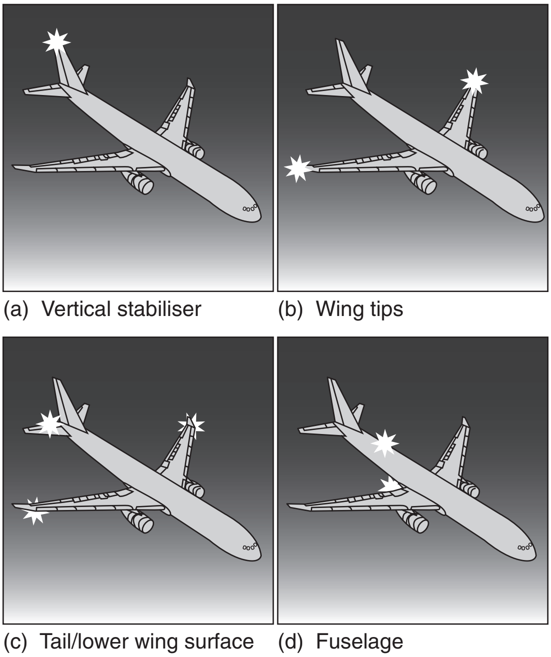
(a) Vertical stabiliser (b) Wing tips

(c) Tail/lower wing surface (d) Fuselage

Figure 12.22 Anti-collision strobe lights

- tail/lower wing surfaces
- fuselage.

These anti-collision lights are controlled by a single switch, with a single protection device. Anti-collision lights used in conjunction with the navigation lights enhance situational awareness for pilots in nearby aircraft, especially during night-time or in low-visibility conditions.

The **rotating beacon** comprises a filament lamp, reflector, motor and drive mechanism that gives the effect of a light through a red filter that flashes 40–50 times per second. A typical rotating beacon is illustrated in Fig. 12.23. They are located on tail fins and the upper and lower fuselage (or tail boom on a helicopter).

Strobe lights are wing-tip and tail-fin mounted to supplement navigation lights. The strobe light produces a high intensity white flash of 1 mS duration at approximately 70 flashes per minute through a white or red filter; these provide light that can be seen from several miles.

Many external lights are based on **sealed beams**. A sealed beam combines an incandescent filament lamp and reflector into a single assembly. The reflector concentrates the light from the lamp into a predetermined beam shape; the assembly is fitted with a clear glass front cover that is permanently sealed to the reflector and cannot be removed. The filament lamp is inserted through a hole in the rear of the reflector and retained by a locking mechanism.

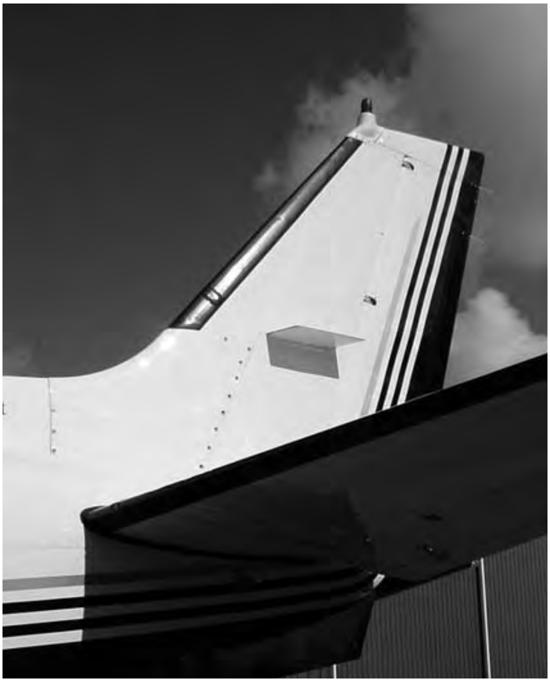
Figure 12.23 Rotating beacon (top of fin)

12.5 Multiple choice questions

1. A high-intensity white flash is produced from a:
   (a) strobe light
   (b) fluorescent tube
   (c) landing light.

2. A clear (white) light with a divergence of ±70 degrees either side of aircraft centreline is the:
   (a) landing light
   (b) rear position light
   (c) logo light.

3. Incandescence is the radiation of light from:
   (a) a gas-discharge device
   (b) an electrical filament due to an increase in its temperature
   (c) a combined optical and electrical phenomenon.

4. Anti-collision lights can be provided by:
   (a) rotating beacon or strobe lights
   (b) fluorescent tubes
   (c) retractable assembly.
5. Green or blue lights in the instrument panel indicate:
   (a) a safe condition exists
   (b) an unsafe condition exists
   (c) an abnormal condition exists.
6. Flood lighting in the flight compartment is normally from:
   (a) strobe lights
   (b) incandescent lamps and/or fluorescent tubes
   (c) position lights.
7. Sealed quartz or glass tubes filled with xenon gas are called:
   (a) luorescent tubes
   (b) LEDs
   (c) strobe lights.
8. The starboard wing tip navigation light has the following colour and divergence:
   (a) red and 110 degrees
   (b) green and 110 degrees
   (c) white and ±70 degrees.
9. Anti-collision lights are used in conjunction with the:
   (a) master warning lights
   (b) wing inspection lights
   (c) navigation lights.
10. Red warning lights in the instrument panel indicate the existence of:
    (a) unsafe conditions
    (b) abnormal conditions
    (c) safe conditions.
11. Referring to Fig. 12.24, this device is a:
    (a) strobe/xenon tube
    (b) fluorescent tube
    (c) LED.
12. Referring to Fig.12.25, the diode is required to:
    (a) provide dimming
    (b) ensure the correct light is illuminated
    (c) ensure both lights are illuminated.
13. Referring to Fig.12.26, this illustrates the angular coverage for:
    (a) wing illumination lights
    (b) rotating beacons
    (c) navigation (or position) lights.

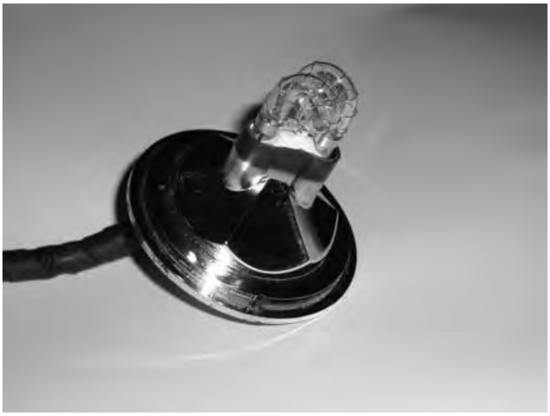

Figure 12.24 See Question 11

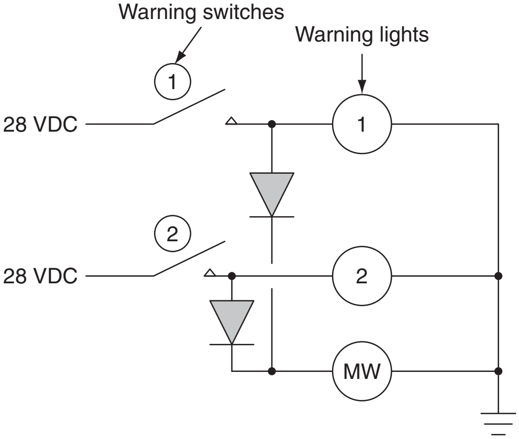

Figure 12.25 See Question 12

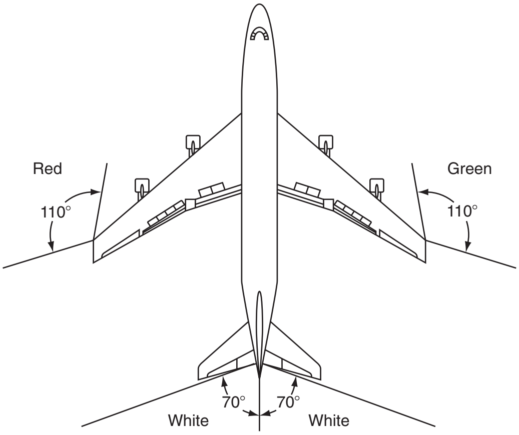

Figure 12.26 See Question 13

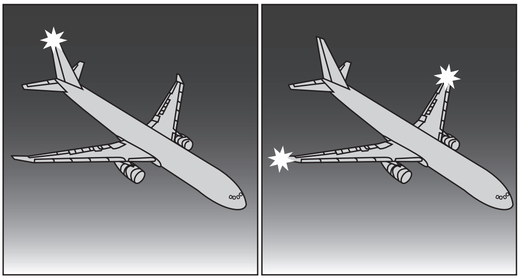

(a) Vertical stabiliser (b) Wing tips

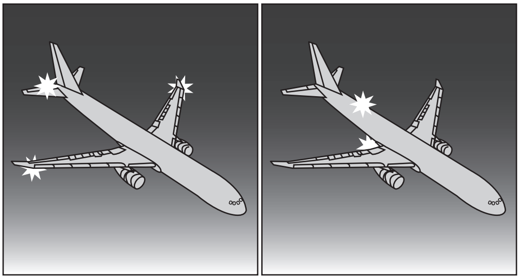

(c) Tail/lower wing surface (d) Fuselage

Figure 12.27 See Question 14

14. Referring to Fig. 12.27, these positions are for:
    (a) navigation lights
    (b) wing inspection lights
    (c) strobe lights.

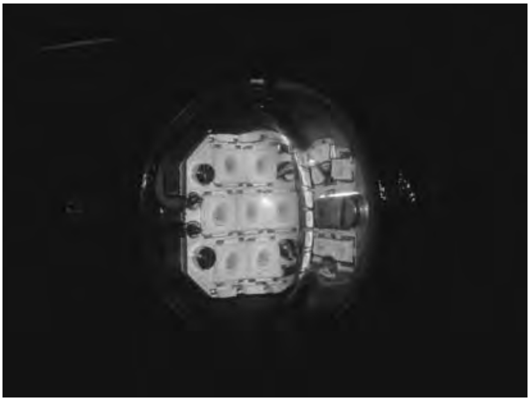

Figure 12.28 See Question 16

15. Master caution and warning lights are located on the:
    (a) lower instrument panel
    (b) upper instrument panel
    (c) overhead panel.

16. Referring to Fig.12.28, this light assembles comprises:
    (a) LEDs
    (b) strobes
    (c) incandescent elements.

Chapter 13

Cabin systems

Passenger transport and business aircraft are fitted with a range of cabin electronic equipment for passenger safety, convenience and entertainment. Typical applications for this equipment include lighting, audio and visual systems. Cabin lighting is used for the safety and comfort of passengers; this was described in Chapter 12. Audio systems include the passenger address system used by the flight or cabin crew to give out safety announcements and other flight information. These announcements are made from handheld microphones and are heard over loudspeakers in the cabin and passenger headsets. The same system can be used to play automatic sound tracks; this is often used for announcements in foreign languages, or to play background music during boarding and disembarkation. Audio systems are also available at the individual seat positions.

A range of galley equipment is installed on business and passenger aircraft. The nature of this equipment depends on the size and role of the aircraft. Air conditioning is provided in passenger aircraft for the comfort of passenger; pressurization is required for flying at high altitudes. Both these systems have electrical and electronic interfaces and control functions. Airstairs allow passengers, flight crew and ground personnel to board or depart the aircraft without the need for a mobile staircase or access to a terminal. This chapter describes the many types of systems and equipment used for passenger safety, convenience and entertainment.

13.1 Passenger address system

The passenger address (PA) system is primarily a safety system that provides passengers with voice announcements and chime signals from the flight or cabin crew; its secondary purpose is for the audio entertainment system. The crew make these voice announcements via the **interphone system** (see *Aircraft Communication and Navigation Systems*, a related book in the series). The PA amplifier provides a level of **sidetone** to the crew's handset or headsets during voice announcements. Sidetone is the technique of feeding back a small amount of sound from the mouthpiece and introducing this at low level back into the earpiece of the same handset, acting as feedback to confirm that the handset or headsets are functional.

The PA system audio outputs are transmitted through to speakers located in the cabin, see Fig. 13.1; typical locations include:

- passenger service units (PSUs)
- galleys
- washrooms
- cabin crew areas.

PA voice announcements are integrated with the passenger entertainment system so that safety announcements can be made over and above entertainment channels.

A typical PA system is controlled by a selector panel (Fig. 13.2) located at cabin crew stations; an amplifier makes chime sounds in response to discrete signals from the cabin interphone system; examples of these chimes include:

- a single low chime to indicate when the 'no smoking' and 'fasten seat belt' lights have been switched on
- a single high/low chime for calls between cabin crew stations
- three high/low chimes when the flight crew needs to gain the attention of the cabin crew
- a single high chime when a passenger needs to gain the attention of the cabin crew.

Test your understanding 13.1

What is the primary purpose of a cabin PA system?

PA systems are often integrated with video systems used for giving cabin safety information to

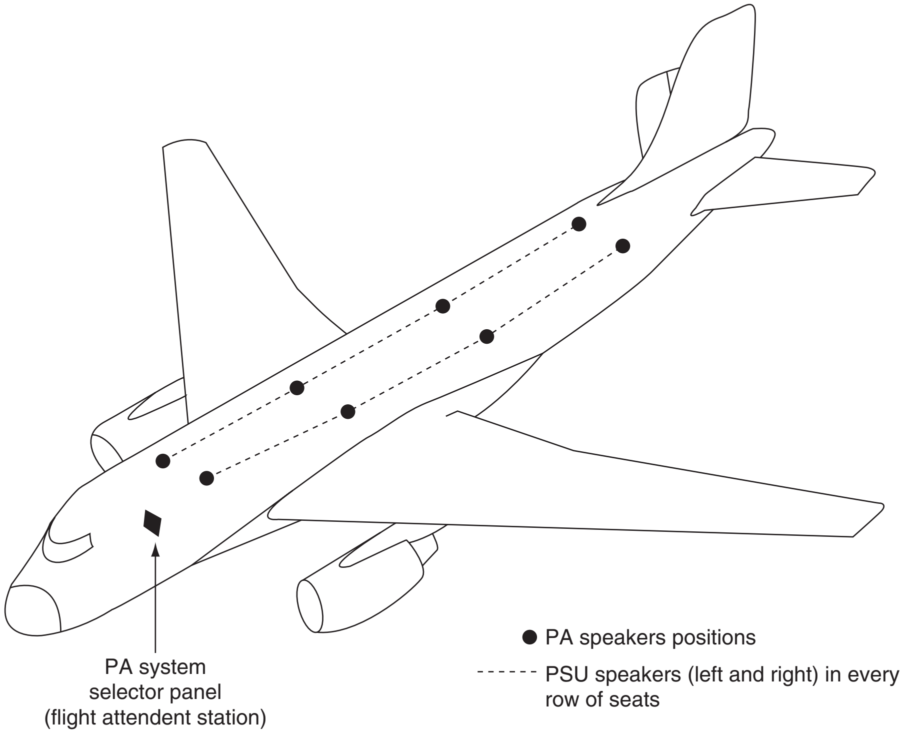
PA system
selector panel
(flight attendent station)
● PA speakers positions
------ PSU speakers (left and right) in every row of seats

Figure 13.1 Public address (PA) system overview

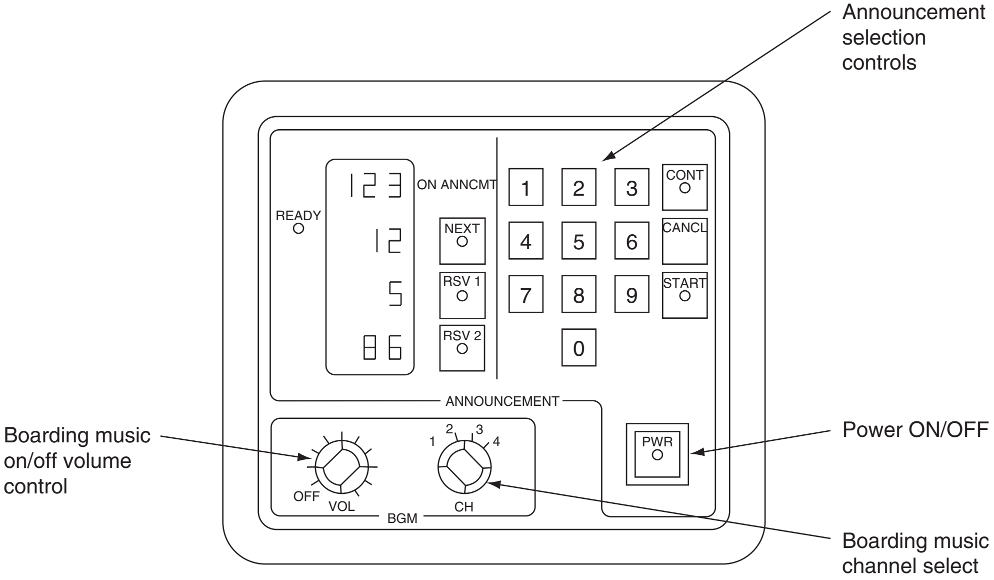
Announcement selection controls
READY
123
12
5
86
ON ANNCMT
NEXT
RSV 1
RSV 2
1
2
3
CONT
4
5
6
CANCL
7
8
9
START
0
ANNOUNCEMENT
Boarding music on/off volume control
OFF
VOL
1
2
3
4
CH
BGM
PWR
Power ON/OFF
Boarding music channel select

Figure 13.2 Public address (PA) selector panel

passengers. A complete **cabin management system** includes:

- crew announcements
- recorded announcements
- video system audio
- boarding music
- chimes.

13.2 Galley equipment

Galley equipment is a major consideration for the design of an aircraft's electrical power system since it can represent a high proportion of the total load requirements. The nature of the installed equipment depends on the size and role of aircraft, e.g. if it is operating short or long routes, the type of cabin classification (economy, business first class etc). In the 1930s, the Douglas Aircraft DC-3 was the first aircraft to have a purpose-built area galley area for the preparation of food and beverages. Typical galley equipment on a modern aircraft now includes:

- beverage makers
- ovens/cookers
- refrigeration units.

Typical power requirements for galley equipment on a typical narrow body aircraft are given in Table 13.1.

Power from the main distribution system is supplied to a dedicated galley bus which also contains its own protection circuits; galley power is the first non-essential load to be shed.

13.3 In-flight entertainment (IFE)

This was initially introduced onto aircraft to break the boredom of long flight times; IFE is now part of an airline's image and forms part of its competitive edge. There is now a vast range of cabin electronic equipment that is used for passenger safety, convenience and entertainment.

13.3.1 Overview

The first ever in-flight 'movie' was shown in 1925. These early systems were traditional projectors and screens. This was followed in the 1960s by centralized television screens placed at the front of each cabin. This concept was further developed into **video projectors** suspended on the aircraft ceiling and large screens on the bulkheads. Smaller business aircraft are installed with video screens fitted into the front and/or rear bulkheads; larger aircraft are installed with monitors integrated within the overhead panels or, more usually, into the back of the seat headrest.

Satellite-based communication is available with an individual transmitter/receiver and handset for business aircraft through to multiple access systems on larger passenger aircraft. The rapid development of broadband communications and satellite systems has also facilitated access to the Internet via laptop computers at an individual seat position. The use of mobile phones is not permitted by many regulatory authorities due to the potential interference with aircraft systems; technology is being developed to allow their use on aircraft.

Audio entertainment systems installed on aircraft range from a single source, e.g. a digital versatile disc (DVD) player on private aircraft through to multi-channel systems with multiple channels on larger aircraft. Passengers are able to select from a wide range of music channels, news bulletins, current affairs, documentaries and comedy channels etc. Audio channels are played through headphones that plug into an individual seat position. **Video entertainment** is traditionally provided via a large video screen at the

Table 13.1 Typical power requirements for galley equipment on a typical narrow body aircraft

Product	*QPA**	*Power supply*	*Rating*	*Protection*
Beverage maker	10	200 VAC three-phase 400 Hz	2750 W	7A
Ovens	5	200 VAC three-phase 400 Hz	1000 W	5A
Refrigeration units	5	200 VAC three-phase 400 Hz	1000 W	5A

*QPA quantity per aircraft

front of each cabin zone, together with smaller monitors suspended from the ceiling situated every few rows. Audio is supplied via the same headphones used for audio entertainment. In addition to entertainment, these screens are also used to broadcast safety information to passengers.

Personal video screens (PVS) for every passenger are becoming the norm, providing passengers with a selection of audio and visual channels. These screens are usually located in the seat backs or stowed in the armrests during take-off and landings. Some airlines also provide news and current affairs programmes, which are pre-recorded prior to the flight and made available to the aircraft before departure. PVS are operated via an in-flight management system which stores pre-recorded channels on a central server, and streams them to the individual seat positions during flight. Some airlines also provide video games to the individual seat positions. **Audio-video on demand** (AVOD) entertainment enables passengers to pause, rewind, fast-forward or stop a programme. AVOD also allows the passengers to choose between an assortment of audio-visual programmes stored in the aircraft computer system.

Passengers can obtain real-time flight information on a video channel at their seat position or on cabin video screens via **moving-map systems**. The typical display is a an electronic screen, typically using **liquid crystal display (LCD) technology** that illustrates the aircraft current position and direction of travel; the system can also display information such as aircraft altitude, airspeed, distance to destination, distance from origination and local time. Moving-map system information is derived from the aircraft's navigation systems. (LCD technology is covered in more detail in *Aircraft Digital Electronic and Computer Systems*, a related book in the series.)

Test your understanding 13.2

What does audio-video on demand (AVOD) entertainment provide for passengers?

13.3.2 Typical product specifications

Companies that offer a complete range of IFE products include Flight Display Systems in the USA. Products include DVD/CD/MP3 players that provide high-resolution images with high-quality sound; high-definition (HD) LCD screens from 5 to 42 inches in size; moving-map systems incorporate an internal database that interfaces with the aircraft's navigation system to display:

- custom waypoints
- takeoff, landing and owner interest sites
- standard GPS position coordinates
- corporate logo
- personalized cabin greeting
- expanded city database
- regional flight information.

Another popular addition to the range of IFE accessories is the very small high-resolution colour camera that is typically mounted in the flight compartment headliner. This provides an NTSC or PAL video output to the cabin entertainment system, enabling passengers to see an outside view from the pilot's perspective on the cabin monitors. Typical Flight Display Systems' products are shown in Fig. 13.3.

13.3.3 IFE system safety and regulation

One major consideration for designing and installing IFE systems is the system's safety. The primary consideration is the additional wiring required by the systems; insulation breakdown and arcing are potential sources of toxic fumes, overheating and fire. To address these potential issues, the IFE system is typically isolated from the aircraft's main electrical systems; many installations are fitted with a master switch in the flight compartment. Electrical equipment must not alter the safety or functionality of the aircraft as a result of a failure. IFE systems must be independent from the aircraft's main power supplies and digital systems. Protection of the aircraft power supplies and digital data links is required to mitigate against IFE failures and maintain the integrity of the aircraft's systems.

13.4 Satellite communications

Some aircraft are fitted with intranet-type **data communication** systems that provide full access to the Internet and email via satellite. A **passenger telephone system** is installed on some aircraft to allow telephone calls from passengers to the ground; these are used for outgoing calls on passenger aircraft, and two-way calls on private aircraft. Data communication is via the **Iridium** satellite communication system and allows passengers to connect to live **Internet**

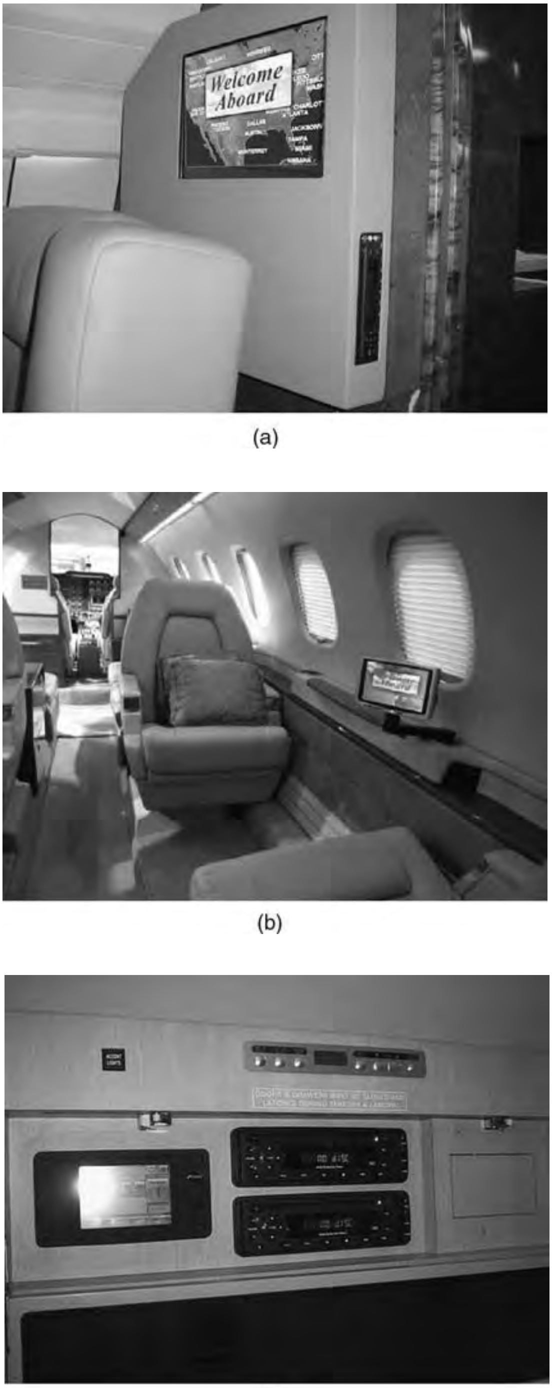

(a)

(b)

(c)

Figure 13.3 IFE products: (a) moving map display (bulkhead installation), (b) moving map display (seat position), (c) DVD/CD/MP3 player (courtesy of Flight Display Systems)

from the individual IFE units or their laptops. Next-generation systems IFE systems are being introduced to incorporate live data/TV reception and on-demand capabilities.

The Iridium system is a satellite-based, personal communications network providing global voice and data features. Iridium is a privately owned company based in the USA. The satellite communications system is illustrated in Fig. 13.4 and comprises three principal components:

- satellite network
- ground network (based on gateways)
- subscriber products (including phones and data modems).

The Iridium network allows voice and data messages to be routed anywhere in the world. The system operates between user and satellite in the L-band, 1616–1626.5 MHz. Voice and data messages are relayed from one satellite to another until they reach the satellite above the Iridium handset or terminal; the signal is then relayed back to a gateway. When an Iridium customer places a call from a handset or terminal, it connects to the nearest satellite, and is relayed among satellites around the globe to whatever satellite is above the appropriate gateway; this downlinks the call and transfers it to the global public voice network or Internet so that it reaches the recipient. Users can access the network via **aircraft earth stations** (AES) or Iridium subscriber units (ISU).

13.4.1 Satellite communication network

The **satellite communication network** comprises a constellation of 66 active satellites in a near-polar orbit at an altitude of 485 miles (780 km). The satellites fly in formation in six orbital planes, evenly spaced around the planet, each with 11 satellites equally spaced apart from each other in that orbital plane. A single satellite orbits the earth once every 100 minutes, travelling at a rate of 16,832 miles per hour; the time taken from horizon to horizon is approximately ten minutes. As a satellite moves out of view, the subscriber's call is seamlessly handed over to the next satellite coming into view.

Key point

The Iridium network allows voice and data messages to be routed anywhere in the world.

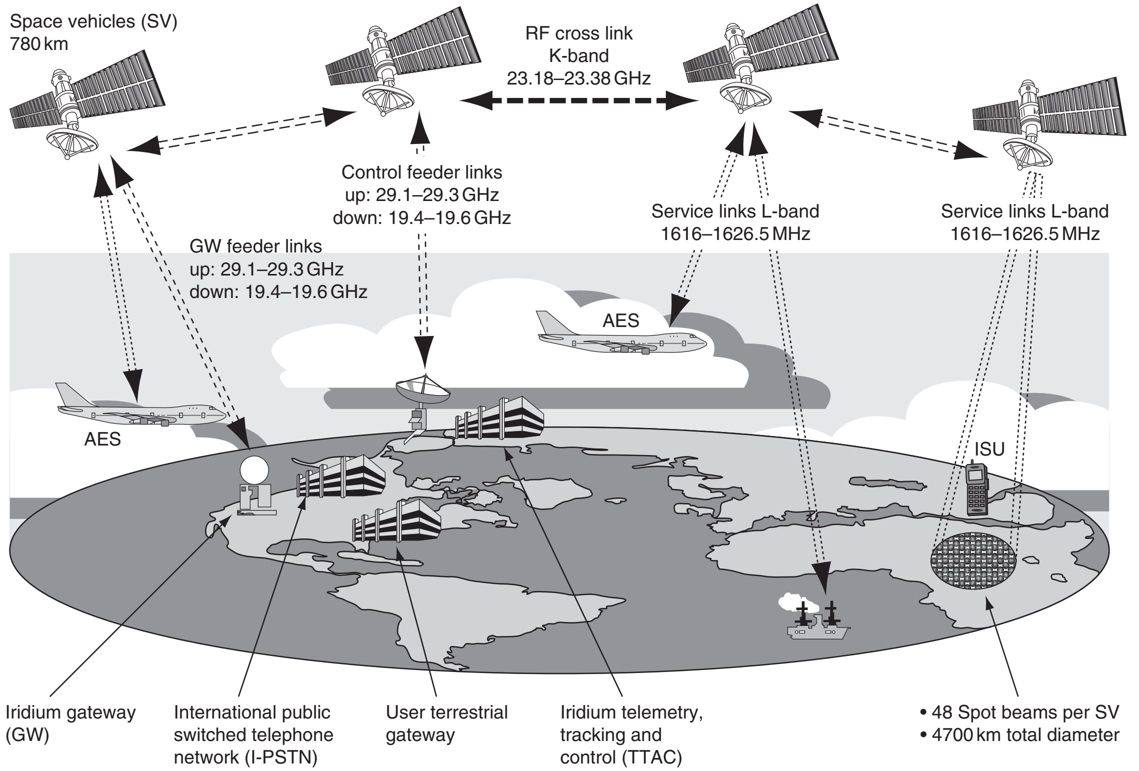
Space vehicles (SV)
780 km
RF cross link
K-band
23.18–23.38 GHz
Control feeder links
up: 29.1–29.3 GHz
down: 19.4–19.6 GHz
Service links L-band
1616–1626.5 MHz
Service links L-band
1616–1626.5 MHz
GW feeder links
up: 29.1–29.3 GHz
down: 19.4–19.6 GHz
AES
AES
ISU
Iridium gateway (GW)
International public switched telephone network (I-PSTN)
User terrestrial gateway
Iridium telemetry, tracking and control (TTAC)
• 48 Spot beams per SV
• 4700 km total diameter

Figure 13.4 Iridium satellite communication system

Since Iridium is a **low earth orbit** (LEO) satellite system, voice delays are minimal. Communication systems using geostationary earth orbits (GEOs) have satellites located 22,300 miles above the equator. As a result, latency can be quite high, causing the users to have to wait for each other to finish. GEOs are also largely ineffective in more northern or southern latitudes. The curvature of the earth disrupts message transmission when attempted at the edge of a GEO satellite's footprint. (Global positioning system (GPS) satellites used for navigation are in 20,200 km orbits.) A comparison of LEO, GPS and GEO orbits is illustrated in Fig. 13.5.

Each Iridium satellite is cross-linked to four other satellites – two satellites in the same orbital plane and two in an adjacent plane. These links create a dynamic network in space – calls are routed directly between satellites without reference to the ground, creating a highly secure and reliable connection. Cross-links make the Iridium system particularly impervious to natural disasters – such as hurricanes, tsunamis and earthquakes – that can damage ground-based wireless towers.

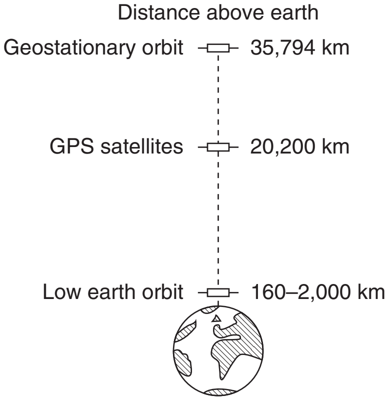
Distance above earth
Geostationary orbit
35,794 km
GPS satellites
20,200 km
Low earth orbit
160–2,000 km

Figure 13.5 Comparison of satellite orbits

Key point

Satellite communication systems use a low earth orbit to minimize voice delays.

13.4.2 Iridium ground network

The Iridium **ground network** comprises the system control segment and telephony gateways used to connect into the terrestrial telephone system. With centralized management of the Iridium network, the system control segment supplies global operational support and control services for the satellite constellation, delivers satellite tracking data to the gateways, and controls the termination of Iridium messaging services. The system control segment consists of three primary components:

- four telemetry tracking and command/control (TTAC) stations
- the operational support network
- the satellite network operation centre (SNOC).

Ku-band feeder links and cross-links throughout the satellite constellation supply the connections among the system control segment, the satellites and the gateways (ku-band is a section of the electromagnetic spectrum in the microwave range of frequencies between 12 and 18 GHz). **Telephony gateways** are the ground-based antennas and electronics that provide voice and data services, messaging, prepaid and postpaid billing services, as well as other customer services. The gateways are responsible for the support and management of mobile subscribers and the interconnection of the Iridium network to the terrestrial phone system. Gateways also provide management functions for their own network elements and links.

13.5 Multiplexing

Increasing the amount of IFE and communication system choices to each passenger position has the potential to rapidly increase the amount of wiring on the aircraft; this would make IFE very costly. One way of reducing the amount of wiring to a seat position is via **multiplexing**; this technology provides a means of selecting data from one of several sources. Equivalent circuits (depicted as switches for illustration purposes) of some common types of multiplexer are shown in Fig. 13.6.

The single two-way multiplexer (Fig. 13.6(a)) is equivalent to a simple single pole double throw (SPDT) changeover switch. The dual two-way multiplexer (Fig. 13.6(b)) performs the same function but two independent circuits are controlled from the same select signal. A single four-way multiplexer (Fig. 13.6c) requires two digital selector inputs (A and B) to place the switch in its four different states. These are coded in a simple truth table format with four possible outcomes, or selectable switch positions. (The subject of multiplexing is covered in more detail in *Aircraft Digital Electronic and Computer Systems*, a companion volume.)

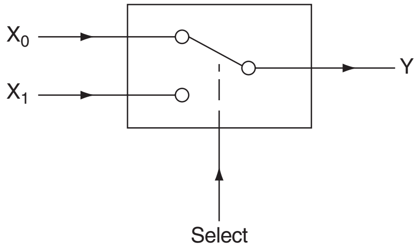

(a) Single two way

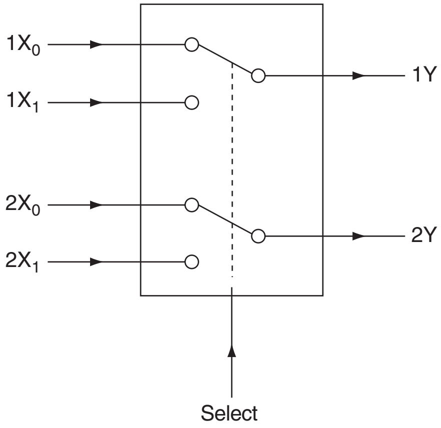

(b) Single two way

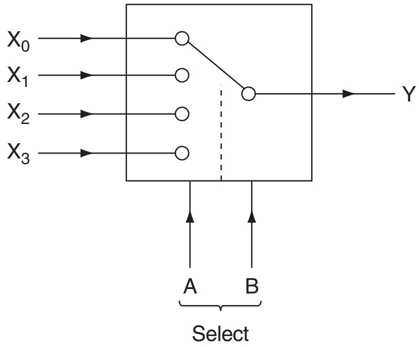

(c) Single four way

Figure 13.6 Multiplexer schematics

Test your understanding 13.3

What are the advantages of using multiplexing for IFE?

13.6 Fibre optics

Another technology adopted for IFE is based on **fibre optics**. This technology has been widely used as a transmission medium for ground-based data communications and in local area networks (LAN) for many years and they are now being introduced into passenger aircraft to satisfy the need for wideband networked avionic and cabin entertainment systems. Compared with copper wiring, optical fibres have a number of advantages:

- reduced weight, compact size
- immunity to electromagnet interference (EMI)
- exceptionally wide bandwidth
- significantly reduced noise and cross-talk
- relatively low values of attenuation within the medium
- high reliability coupled with long operational life
- electrical isolation and freedom from earth/ground loops.

These advantages mean that optical fibres are ideally suited as a replacement for conventional copper network cabling. The technology is relatively new in the civil aircraft industry and introduces new challenges for those involved with aircraft operation and maintenance. There are a number of disadvantages for optical fibres; these include:

- industry resistance to the introduction of new technology
- need for a high degree of precision when fitting cables and connectors
- concerns about the mechanical strength of fibres
- cable bends need to have a sufficiently large radius to minimize losses and damage.

The advantages are now outweighing the disadvantages and systems based on optical fibres are being installed on many aircraft types. A simple one-way (simplex) fibre optic data link is shown in Fig. 13.7. The optical transmitter consists of an infrared light-emitting diode (LED) or low-power semiconductor laser diode coupled directly to the optical fibre. The diode is supplied with pulses of current from a bus interface.

These pulses of current produce equivalent pulses of light that travel along the core of the fibre until they reach the optical receiver unit. The optical receiver unit consists of a photodiode or phototransistor that passes a relatively large current when illuminated and negligible current when not. The pulses of current at the transmitting end (logic inputs) are thus replicated at the receiving end (as logic outputs). **Cladding** is a layer of material with a lower refractive index than the core, and confines the optical signal inside the core.

13.6.1 Construction

The construction of a typical fibre optic cable is shown in Fig. 13.8; in this example, the cable consists of:

1. an outer jacket
2. aramid yarn (strength member)
3. separator tape (polyester)
4. two filler strands
5. five optical fibres.

Key maintenance point

Fibre optic cable bends need to have a sufficiently large radius to minimize signal losses and physical damage.

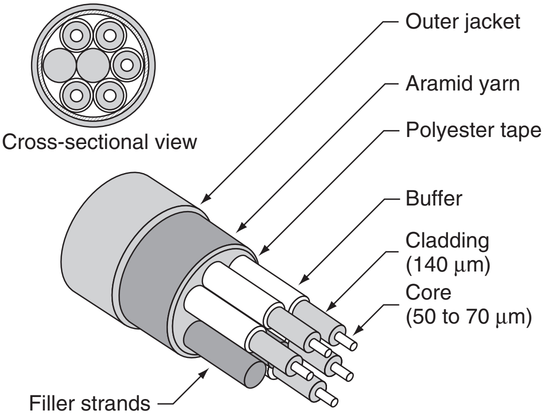

Figure 13.8 (Simplex) fibre optic data link

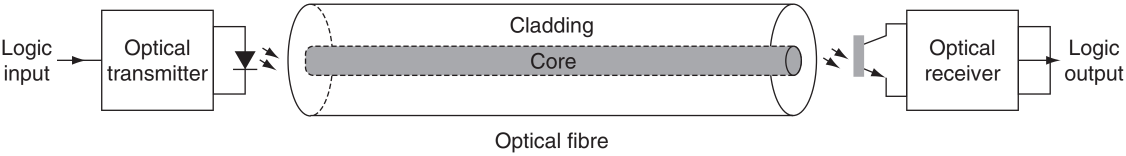

Figure 13.7 Fibre optic cable connector

A protective **buffer** covers each fibre and protects it during manufacture, increases mechanical strength and diameter in order to make handling and assembly easier. The buffers are coded in order to identify the fibres using colours (blue, red, green, yellow and white). The filler strands are made from polyester and are approximately 0.035 inches in diameter. A polyester separator tape covers the group of five fibres and two filler strands. This tape is manufactured from low-friction polyester and it serves to make the cable more flexible. A layer of woven aramid (or Kevlar) yarn provides added mechanical strength and protection for the cable assembly. The outer thermoplastic jacket (usually purple in colour) is fitted to prevent moisture ingress and also to provide insulation. A typical fibre optic cable connector arrangement is shown in Fig. 13.9.

13.6.2 Connectors

Each connector has alignment keys on the plug and matching alignment grooves on the receptacle. These are used to accurately align the connector optical components; the guide pins in the plug fit into cavities in the receptacle when the plug and receptacle connect. In order to ensure that the connector is not over-tightened (which may cause damage to the fibres) the pins of the plug are designed to provide a buffer stop against the bottom of the cavities in the receptacle. Plugs and receptacles have ceramic contacts that are designed to make physical contact when properly connected.

The light signal passes through holes in the end of the ceramic contacts when they are in direct physical contact with each other. The coupling nut on the plug barrel has a yellow band whilst the receptacle barrel has a red and a yellow band. A correct connection is made when the red band on the receptacle is at least 50 per cent covered by the coupling nut. This position indicates an effective connection in which the optical fibres in the plug are aligned end-to-end with the fibre in the receptacle. Three start threads on the plug and receptacle ensure a straight start when they join. The recessed receptacle components prevent damage from the plug if it strikes the receptacle at an angle. The plug and receptacle are automatically sealed in order to prevent the ingress of moisture, dust and other contamination. Colour-coded bands ensure that the plug and receptacle are fully mated. (The subject of optical fibres is covered in more detail in *Aircraft Digital Electronic and Computer Systems*, a companion volume.)

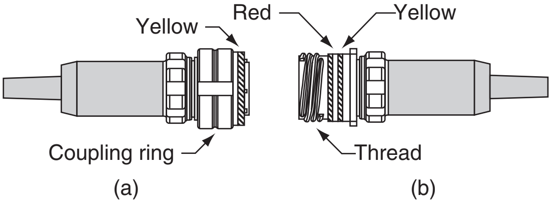

Figure 13.9 Fibre optic cable

Key point

The optical fibre receiver unit consists of a photodiode or phototransistor that passes a relatively large current when illuminated and negligible current when not illuminated.

13.7 Air conditioning

Air conditioning is provided in passenger aircraft for the comfort of passengers. The cabin on larger aircraft is divided into passenger compartment locations, or temperature control zones, see Fig. 13.10. Air conditioning, and the environmental control system (ECS), normally derive high-pressure air from the compressor stage of each turbine engine. The temperature and pressure of this bleed air varies, depending upon rotational speed of the engine. A pressure-regulating shutoff valve (PRSOV) restricts the flow as necessary to maintain the desired pressure for the ECS.

13.7.1 Environmental control system

To increase efficiency of the ECS, air is normally bled from two or three positions on the engine. Aircraft types vary, but the principles of air conditioning systems comprise five salient features:

- air supply
- heating
- cooling
- temperature control
- distribution.

The air conditioning system is based on an air cycle machine (ACM) cooling device as illustrated in

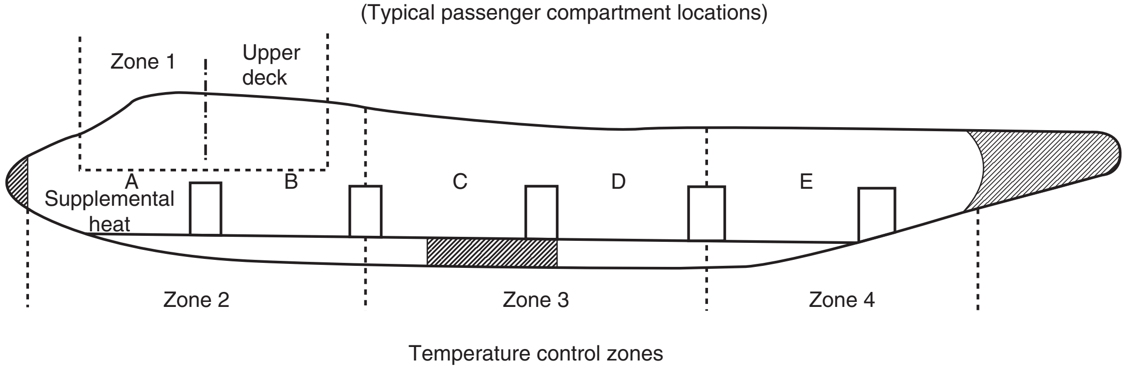
(Typical passenger compartment locations)
Zone 1
Upper deck
A
B
C
D
E
Supplemental heat
Zone 2
Zone 3
Zone 4
Temperature control zones

Figure 13.10 Air conditioning: passenger cabin zones

Fig. 13.11. The air conditioning packs (**A/C packs**) are located in varying places on aircraft, including:

- between the two wings in the lower fuselage
- in the rear fuselage tail section
- in the front of the aircraft beneath the flight deck.

Engine bleed air, with temperatures in the order of 150 and 200°C and a pressure of between 30–35 pounds per square inch (psi), is directed into a primary heat exchanger. External ram air (at ambient temperature and pressure) is the cooling medium for this air-to-air heat exchanger. The cooled bleed air then enters the centrifugal compressor of the air cycle machine (ACM). This compression heats the air (the maximum air temperature at this point is about 250°C) and it is directed into the secondary heat exchanger, which again uses ram air as the coolant. Pre-cooling through the primary heat exchanger increases the efficiency of the ACM by reducing the temperature of the air entering the compressor; less work is required to compress a given air mass (the energy required to compress a gas by a given ratio increases with higher temperature of the incoming air).

At this stage, the temperature of the heat exchanger air is higher than the outside air temperature; the temperature is sensed by an RTD and this is displayed on the air conditioning control panel and/or used as part of a control system input. The compressed, cooled air is then directed into the expansion turbine of the ACM; this extracts work from the air as it expands, cooling it to between −20°C and −30°C. (The ACM can cool the air to less than 0°C even when the aircraft is on the ground in a high ambient temperatures.) The work extracted by the turbine drives a shaft to turn the ACM's centrifugal compressor, together with a ram air inlet fan that draws in the external air during ground running. A motorized bypass valve controls the ratio of air being directed into the turbine. An electrically driven ram air fan within the system provides air flow across the heat exchangers when the aircraft is on the ground. To assist ram air recovery, some aircraft use modulating vanes on the ram air exhaust. Power for the air conditioning pack is obtained by reducing the pressure of the incoming bleed air relative to the cooled air output of the system.

Test your understanding 13.4

What type of sensor is used for regulation of cabin air temperature?

With the air now cooled, its water vapour condenses. To remove this, the moist air output of the expansion turbine is passed through a water separator; this uses centrifugal force to eject the water particles into a coalescing bag that absorbs the moisture. This condensate is sometimes fed back into the ram air entering the secondary heat exchanger to improve its performance. The cool dry air is now combined in a mixing chamber with a small amount of engine bleed air; the amount of trimming air mixed with cooled air is modulated to achieve the desired cabin air temperature before the air is ready for supply into the cabin. An RTD is installed to monitor the ACM outlet temperature; airflow into the cabin is monitored by a flow sensor.

The A/C pack outlet air for use in the cabin is mixed with filtered air from re-circulation fans, and

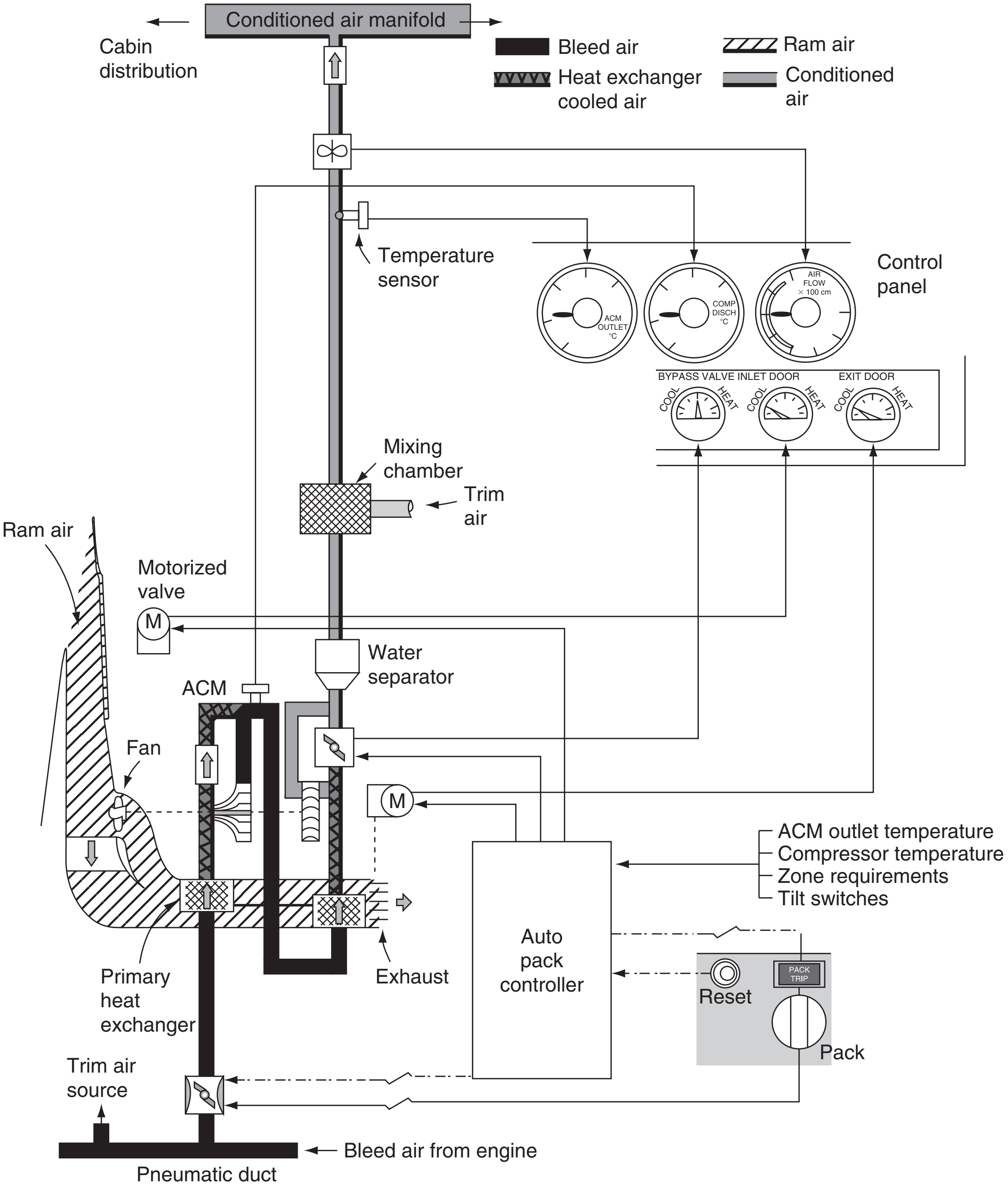
Conditioned air manifold
Cabin distribution
Bleed air
Heat exchanger cooled air
Ram air
Conditioned air
Temperature sensor
Control panel
ACM OUTLET °C
COMP DISCH °C
AIR FLOW × 100 cm
BYPASS VALVE INLET DOOR
EXIT DOOR
COOL
HEAT
COOL
HEAT
COOL
HEAT
Mixing chamber
Trim air
Ram air
Motorized valve
M
Water separator
ACM
Fan
M
ACM outlet temperature
Compressor temperature
Zone requirements
Tilt switches
Auto pack controller
Primary heat exchanger
Exhaust
Reset
PACK TRIP
Pack
Trim air source
Bleed air from engine
Pneumatic duct

Figure 13.11 Air conditioning system

then fed into the mixing manifold. On modern passenger aircraft, the airflow into the cabin is approximately 50% bleed air and 50% filtered air. Control of the air conditioning system on a Boeing 757 aircraft is from an overhead control panel, see Fig. 13.12.

Key point

To assist ram air recovery, some aircraft use modulating vanes on the ram air exhaust.

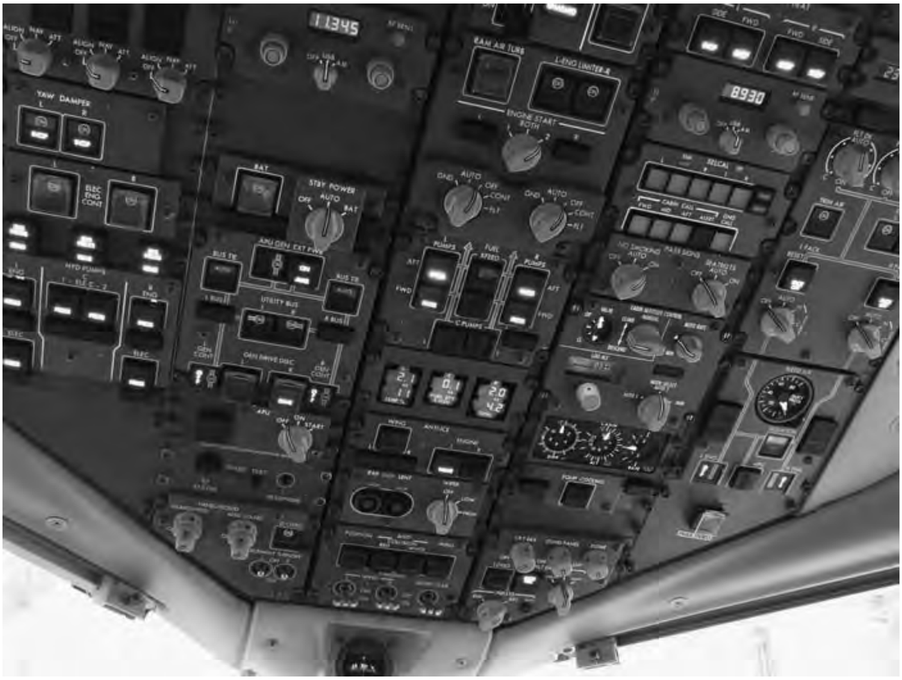

Figure 13.12 Air conditioning system control (right/centre of overhead panel)

13.7.2 Ventilation

A **re-circulation fan** is used to re-circulate filtered cabin air back into the cabin to reduce bleed air requirements. Up to 50% of the cabin air can re-circulated for passenger comfort. The fan will switch off if either A/C pack is in high flow, giving a net reduction in the ventilation rate. Optimum cooling and reduced bleed air demand (hence reduced fuel consumption) is achieved with a combination of air re-circulation and automatic operation of the A/C packs. The ventilation rate is increased on larger aircraft with additional re-circulation fans for comfort levels.

Some aircraft are installed with a **gasper fan**; this is an electric fan designed to increase pressure in the outlets above passenger seats. The gasper fan is typically used when there is a:

- low supply of air pressure
- high cold air demand
- high ambient temperature (on the ground on a hot day).

When the A/C packs are off, the gasper fan draws cabin air into the distribution ducts, into the main air manifold and mixing chamber where it is then blown (albeit not chilled) into the cabin.

13.7.3 Equipment cooling

Although not part of the cabin system for the comfort of passengers, some aircraft are installed with equipment cooling that takes air from the cabin system. Equipment cooling can be used for:

- flight compartment panels
- display units
- circuit breaker panels
- electrical and electronic (E & E) bay.

Replacing the warm air generated by equipment with cool air from the cabin is achieved with dedicated electric fans. On the ground, the air is directed overboard through the flow control or exhaust valves. During flight, the warm air generated by equipment is sometimes used for heating the cargo holds by exhausting air around their walls.

13.8 Pressurization

In addition to temperature control, aircraft flying at high altitudes need to be pressurized since the amount of oxygen in the atmosphere decreases with altitude. The effects of insufficient oxygen on people can begin

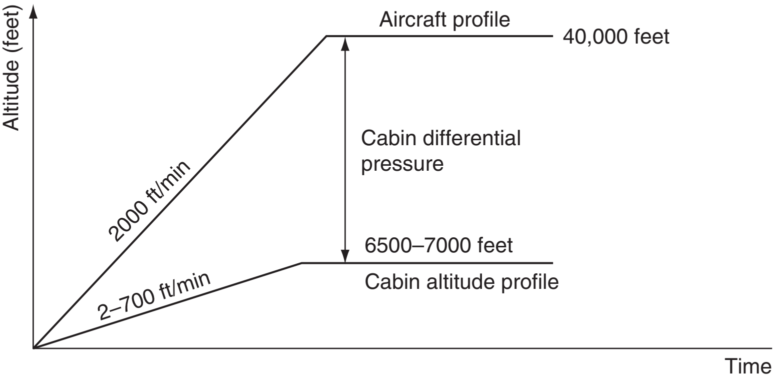

Figure 13.13 Cabin pressurization/altitude chart

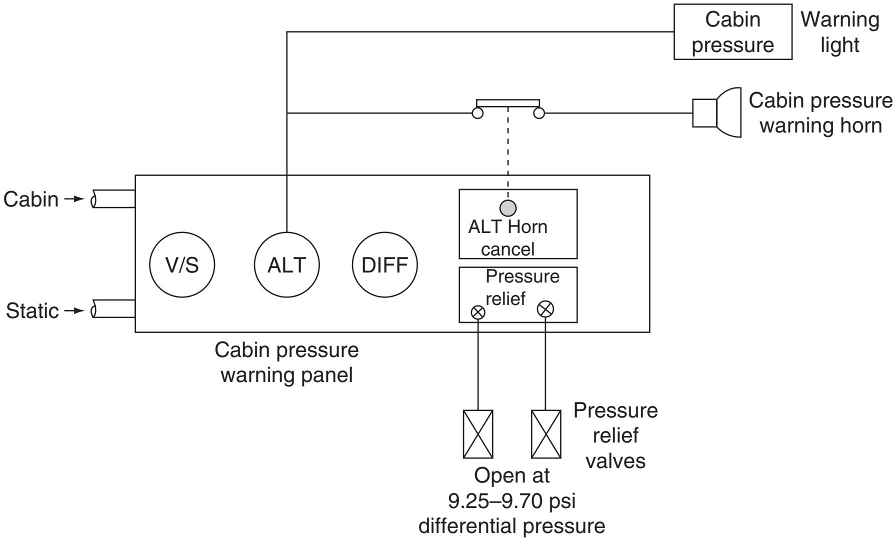
Cabin pressure
Warning light
Cabin pressure warning horn
Cabin
Static
V/S
ALT
DIFF
ALT Horn cancel
Pressure relief
Cabin pressure warning panel
Pressure relief valves
Open at 9.25–9.70 psi differential pressure

Figure 13.14 Cabin pressurization warning

as low as 1500 m (5000 ft) above sea level, although most passengers can tolerate altitudes of 2500 m (8000 ft) without ill effect. At this altitude, there is approximately 25% less oxygen compared to sea level. Flight operations above 3000 m (10,000 ft) generally require supplemental oxygen; this is achieved through pressurization on passenger aircraft.

The relationship between cabin pressure and atmospheric pressure is illustrated in Fig. 13.13. Cabin pressure is controlled automatically; this is managed by the crew via a control panel as illustrated in Fig. 13.12. Pressurization is monitored and warnings provided in the event of a malfunction. Fig. 13.14 illustrates a typical warning system schematic; this illuminates a warning light on the master caution panel and sounds a horn in the event of reduced cabin altitude.

Engine bleed air is supplied into the cabin and then allowed to pass out of the fuselage via **outflow valves** (OFV). Most aircraft have a single OFV located near the bottom aft end of the fuselage; some larger aircraft have two. By modulating the position of the outflow valve(s), the pressurization in the cabin can be maintained higher than atmospheric pressure. Modern commercial aircraft have a dual channel electronic controller for maintaining pressurization, with a manual back-up system. These systems maintain cabin air pressure at the equivalent to 2500 m (8000 ft) at high cruising altitudes.

A positive pressure relief valve opens in the event of excessive pressure in the cabin; this protects the aircraft structure from excessive loading. The maximum pressure differential between the cabin and atmospheric pressure is between 7.5 and 8 psi. This is the equivalent of approximately 10,000 feet; if the cabin were maintained at sea level pressure and then operated at cruise altitude, the differential would be the structural life of the aircraft would be reduced.

Key point

Bleeding air from the engine comes at the expense of overall aircraft efficiency. Some aircraft such as the Boeing 787 are now using **electrically driven compressors** to provide pressurization.

A pressurization failure above 10,000 ft results in an emergency situation; the cabin pressure might be 10 psi, while the atmospheric pressure is only 2 psi. The pilot must make an emergency descent and activate oxygen masks for everyone aboard. In most passenger aircraft, oxygen masks are automatically deployed if the cabin pressure drops below an equivalent cabin altitude of 14,000 feet. Gradual decompression is dangerous to all on board since it may not be detected. Rapid decompression causes the lungs to decompress faster than the cabin. Explosive decompression is when cabin pressure reduces faster than the lungs can decompress (less than 0.5 seconds). Rapid decompression of commercial aircraft is extremely rare, but the results are life-threatening. Furthermore, the air temperature will rapidly fall due to the expansion of the cabin air, potentially resulting in frostbite.

Key point

Aircraft cabin pressure is commonly referred to as a ‘cabin altitude’, typically 8000 feet or less, therefore a lower cabin altitude relates to a higher pressure.

Key point

The cargo compartment is normally pressurized to the same level as the cabin; the temperature in this compartment may be controllable depending on the nature of cargo being carried.

13.9 Airstairs

These are located in either forward or aft cabin locations, and can be raised or lowered while the aircraft is on the ground. Airstairs allow passengers, flight crew and ground personnel to board or depart the aircraft without the need for a mobile staircase or access to a terminal.

On forward door positions, the stair folds and stows directly into the floor, beneath the door. This type of airstair is found on many short-range aircraft to provide operational flexibility, although the stairs are quite heavy; many airlines remove this system to reduce aircraft weight. Ventral airstairs are fitted on most tail-engined airliners, and are incorporated as ramps which lower from the rear fuselage.

They can be operated from either internal or external control panels; both panels typically have normal and standby systems. Normal operation requires both AC and DC power, standby only requires DC power. External standby system power typically comes directly from the battery bus and so does not require the battery switch to be on. Airstairs should not be operated more than is necessary to prevent the motors from overloading; their typical duty cycle is three consecutive cycles of normal system operation within a 20 minute period.

13.10 Multiple choice questions

1. Multiplexing is a technique used for:
   (a) increasing the amount of IFE wiring to a seat position
   (b) reducing the amount of IFE wiring to a seat position
   (c) increased immunity to electromagnet interference (EMI).

2. Fibre optic cable bends need to have a sufficiently large radius to:
   (a) minimize losses and damage
   (b) maximize immunity to electromagnet interference (EMI)

(c) accurately align the connector optical components.

3. ACM outlet temperature is measured using:
   (a) a Bourdon tube
   (b) an RTD
   (c) a thermocouple.

4. The Iridium network allows voice and data messages to be routed:
   (a) anywhere in the world
   (b) between the flight crew and cabin crew
   (c) via a fibre optic network.

5. The passenger address (PA) system is primarily a safety system that provides passengers with:
   (a) in-flight entertainment
   (b) reduced amount of IFE wiring to a seat position
   (c) voice announcements and chime signals.

6. To assist ram air recovery, some aircraft use:
   (a) modulating vanes on the ram air exhaust
   (b) modulating vanes on the ram air inlet
   (c) modulating vanes on the ACM outlet.

7. Satellite communication systems use a low earth orbit to:
   (a) provide greater coverage
   (b) maintain a geostationary position
   (c) minimize voice delays.

8. Audio-video on demand (AVOD) entertainment enables passengers to:
   (a) pause, rewind, fast-forward or stop a programme
   (b) make phone calls via satellite communication
   (c) ignore PA system voice announcements and chime signals.

9. Supplemental oxygen is generally required flight operations above:
   (a) 3000 ft
   (b) 10,000 ft
   (c) sea level.

10. The optical fibre receiver unit consists of a photodiode or phototransistor that passes a:
    (a) relatively large current when illuminated and negligible current when not illuminated
    (b) relatively small current when illuminated and high current when not illuminated
    (c) relatively small current when illuminated and low current when not illuminated.

11. The gasper fan is used in cabin ventilation systems to:
    (a) increase air pressure in the air outlets
    (b) re-circulate filtered cabin air
    (c) assist with ram air recovery.

12. Cabin altitude is typically
    (a) cruising altitude
    (b) sea level
    (c) between 6–7000 feet.

13. The cargo compartment is normally pressorized to be:
    (a) equal to cabin pressure
    (b) above cabin pressure
    (c) below cabin pressure.

14. Referring to Fig. 13.15, cladding is used in fibre optic cables to:
    (a) protect the fibre
    (b) prevent moisture ingress
    (c) guide the light signal.

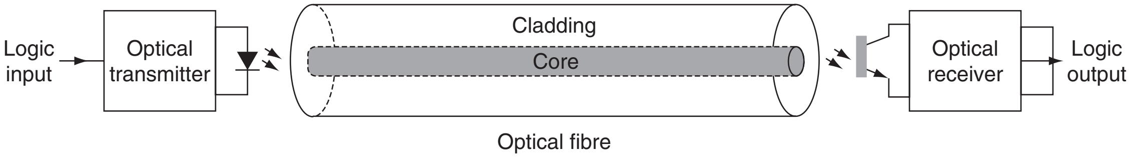

Figure 13.15 See Question 14

15. Referring to Fig. 13.16, the iridium system transfers messages around the world:
    (a) directly between satellites
    (b) via gateways on the earth's surface
    (c) directly between users.

16. The iridium system satellite orbits are:
    (a) random
    (b) equatorial
    (c) polar.

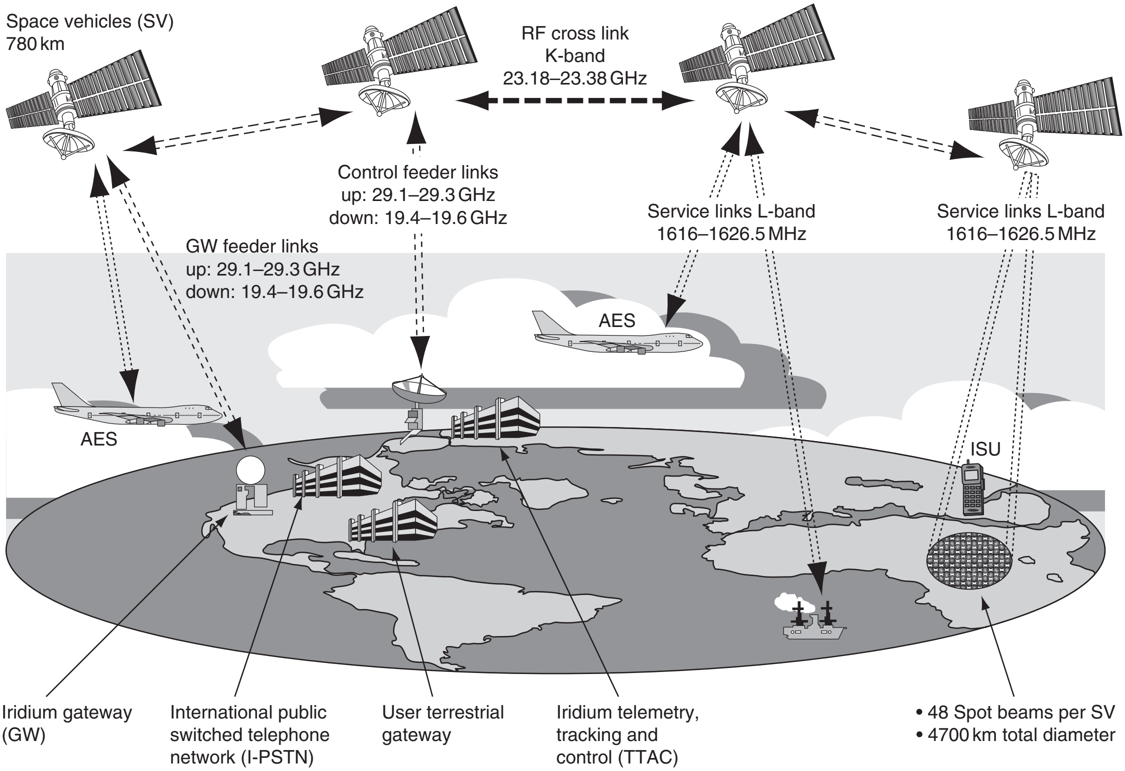
Space vehicles (SV)
780 km
RF cross link
K-band
23.18–23.38 GHz
Control feeder links
up: 29.1–29.3 GHz
down: 19.4–19.6 GHz
Service links L-band
1616–1626.5 MHz
Service links L-band
1616–1626.5 MHz
GW feeder links
up: 29.1–29.3 GHz
down: 19.4–19.6 GHz
AES
AES
ISU
Iridium gateway (GW)
International public switched telephone network (I-PSTN)
User terrestrial gateway
Iridium telemetry, tracking and control (TTAC)
• 48 Spot beams per SV
• 4700 km total diameter

Figure 13.16 See Question 15

Chapter 14

Airframe monitoring, control and indicating systems

Various sensors are needed for the monitoring and control of airframe systems. The indications and/or control circuit need to be informed of the position of a particular feature on the aircraft; this is achieved by a position sensor. Broadly speaking, the sensors can be considered as detecting one of two states (or 'conditions'), or a variable position. Two state conditions include: landing gear (up or down) or cabin doors (open or closed). Variable positions include: control surfaces and flap position. Two state position devices are detected by micro-switches and proximity sensors; variable position devices are detected by a variety of devices including synchros and variable resistors. This chapter reviews airframe systems such as landing gear control and indication, control surface position and indicating systems.

14.1 Landing gear

Three control and indicating system configurations are found on aircraft with retractable landing gear (or undercarriage):

- hydraulic control and operation, electrical indication of position
- electrical control and operation, electrical indication of position
- electro-hydraulic control and operation, electrical indication of position.

The landing gear comprises two or more wheels and shock absorbers, or **oleo legs**; these can be fixed in position or retractable. Large aircraft have hydraulic systems with electrical position sensors for gear up/down indications. General aviation (GA) aircraft normally have electrical control and operation, and electrical indication of position. Operation is via a series-wound split field electrical motor (Fig. 14.1) driving through to each undercarriage leg via a gearbox, torque tubes, cables and pulleys.

The 28V DC power supply is connected to the **reversible** motor via a safety switch and the position selector switch. This switch operates the landing gear relay via up-lock and down-lock micro-switches. When the landing gear starts to retract, the down-lock switch contacts close. When the landing gear is fully retracted the up-lock switch contacts open, and this removes power from the motor. When the landing gear is selected down, the up-lock contacts close and the gear can be extended.

The landing gear safety switch senses when the aircraft is on the ground to prevent the landing gear from being retracted. Some aircraft have a **micro-switch** (or **proximity-sensor**) attached to the oleo leg; when the aircraft is on the ground, the oleo leg is compressed from the weight of the aircraft; this closes the switch contacts. The switch (or sensor) is now being used to detect the **weight on wheels (WoW)**; this is sometimes referred to as a **squat** switch. When the aircraft takes off, the oleo leg extends when the aircraft weight is transferred to the wings and the WoW, or squat switch, opens its contacts. (Note that the switch can actually be configured as normally open or normally closed on the ground depending on the aircraft design.) WoW or squat switch contacts are also used by other systems that need to be set into air or ground modes, e.g. to prevent the pitot probes and ice detectors from being heated on the ground.

Key point

Two state conditions include landing gear (up or down) or cabin doors (open or closed).

Landing gear position indication is derived from simple micro-switches in two locations: gear up and gear down. Switches are operated by a cam or lever and this completes the circuit. The quantity of lights depends on the aircraft manufacturer and certification requirements of the aircraft type. The simple system shows if the nose and main gear are up or down with a single indication; it is more usual to

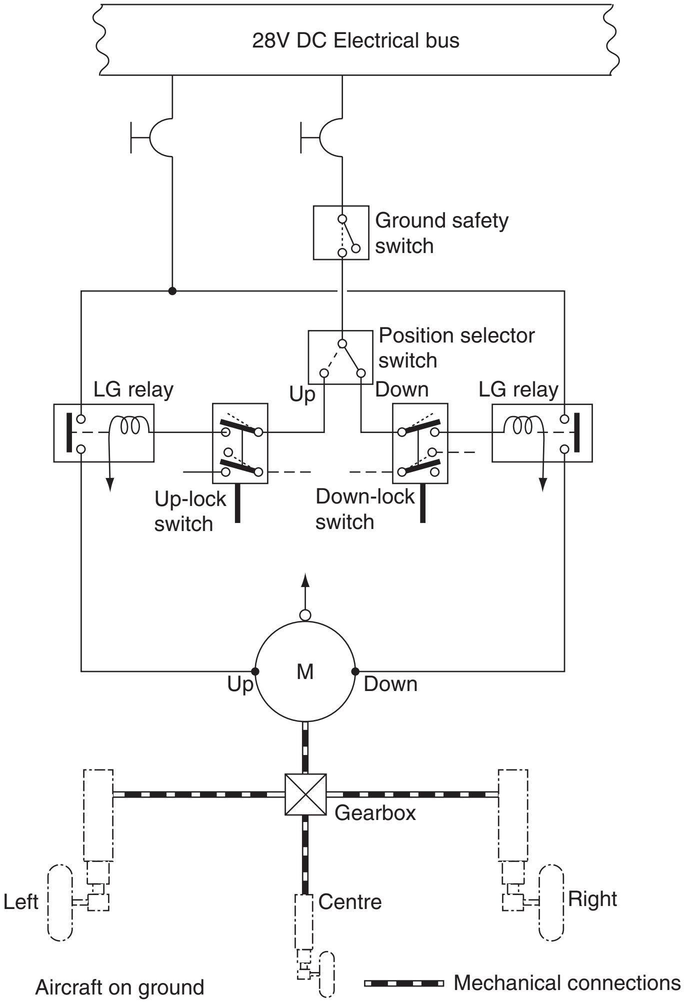
28V DC Electrical bus
Ground safety switch
Position selector switch
LG relay
Up
Down
LG relay
Up-lock switch
Down-lock switch
M
Up
Down
Gearbox
Left
Centre
Right
Aircraft on ground
Mechanical connections

Figure 14.1 Landing gear control system

have an indication of each gear leg position. Landing gear indications sometimes include an audible warning (klaxon or horn) when the throttles are retarded and the gear is not locked down. The system configuration in widespread use has the following indications:

- gear down and locked (three green lights)
- gear up and locked (all three lights out)
- gear in transit (three red lights).

Figure 14.2(a) illustrates a typical landing gear position indication system; note that this system includes additional green lights that can be selected in the event of lamp failure. The circuit is drawn with positions of switches for the aircraft on the ground.

With power applied, the three green lights are illuminated via the down-lock switch. When the gear is selected up, the down-lock switch opens, the green lights extinguish and the three red lights are turned on. When the gear is fully retracted and locked, the red lights are extinguished. When the gear is selected down, the up-lock switch closes, and three red lights come on. With the gear fully down and locked, the down-lock switch closes, the red lights are extinguished and the green lights are switched on. In the event that the throttles are closed (during an approach) and the gear is up, a warning horn will sound.

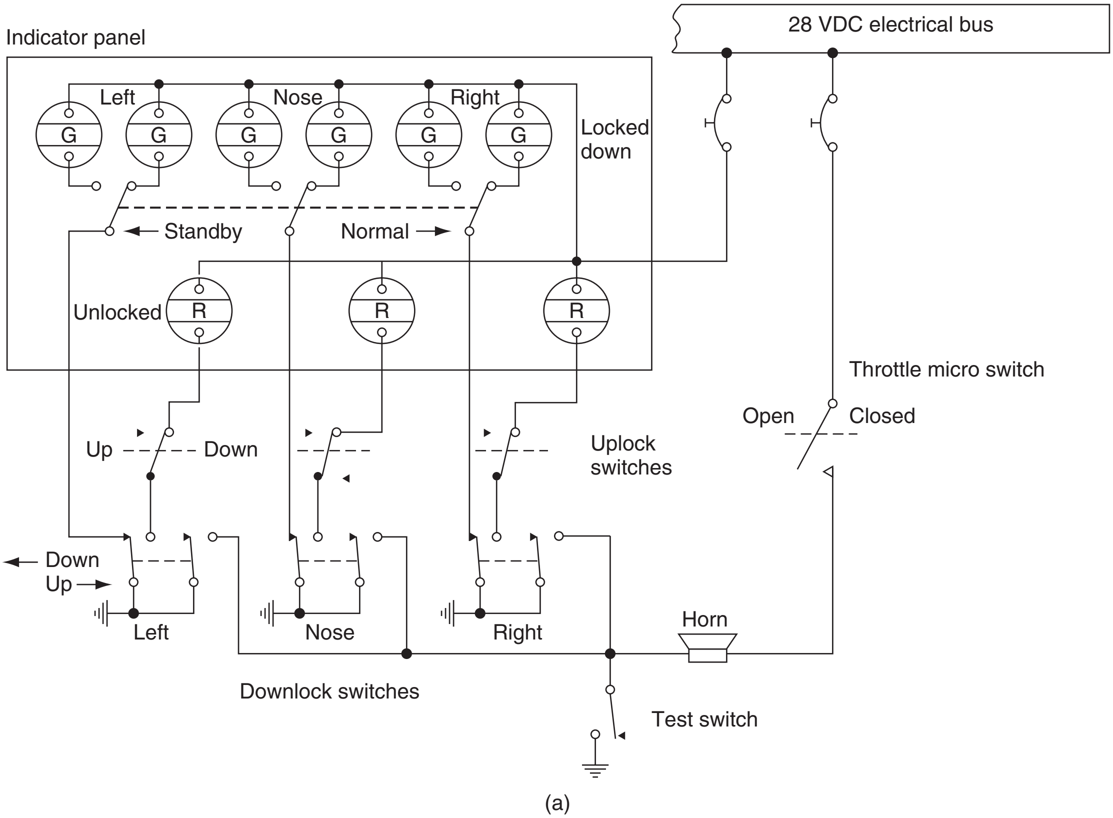
28 VDC electrical bus
Indicator panel
Left
Nose
Right
G
Locked down
Standby
Normal
Unlocked
R
Throttle micro switch
Open
Closed
Up
Down
Uplock switches
Down
Up
Left
Nose
Right
Horn
Downlock switches
Test switch
(a)

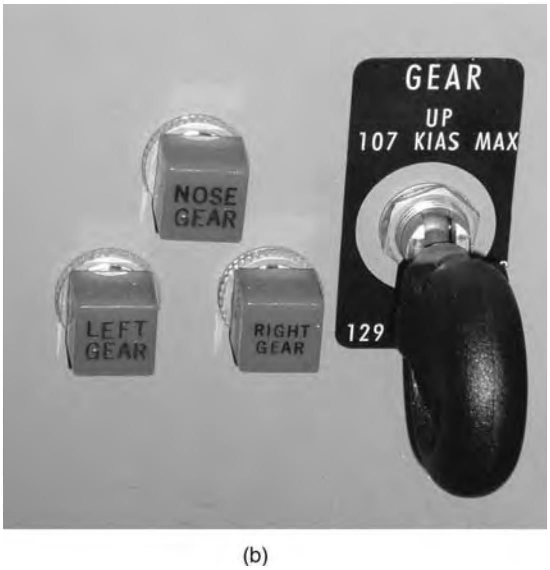
GEAR
UP
107 KIAS MAX
129
NOSE GEAR
LEFT GEAR
RIGHT GEAR
(b)

Figure 14.2 Landing gear position indication: (a) system schematic, (b) typical gear position selector and warning lights (general aviation)

Key point

Indications of landing gear fully down and locked are:

- red lights off
- green lights on.

Key point

When an electrically operated landing gear is fully retracted, the up-lock switch contacts open, thereby removing power from the motor

On smaller general aviation aircraft, the system is simplified by having three green lights for gear down and locked (Fig. 14.2(b)). In this example, the selector switch also has a placard stating the airspeeds for gear selection.

14.2 Trailing edge flaps

Three configurations are found on aircraft:

- hydraulic control and operation, electrical indication of position
- electrical control and operation, electrical indication of position
- electro-hydraulic control and operation, electrical indication of position.

The flap drive motor is a reversible DC motor that drives the flap control mechanism through a gearbox. GA aircraft have a simple three-position switch identified as OFF/UP/DOWN. The flaps on these aircraft are driven the full extent of travel until they operate up or down limit switches, see Fig. 14.3.

Selecting ‘flaps up’ applies power to the motor through the closed contacts of the up-limit switch; the motor operates and the flaps start to move. The down-limit switch changes over as shown by dotted lines, and the flaps travel to the fully retracted position; note that both the limit switches are closed during

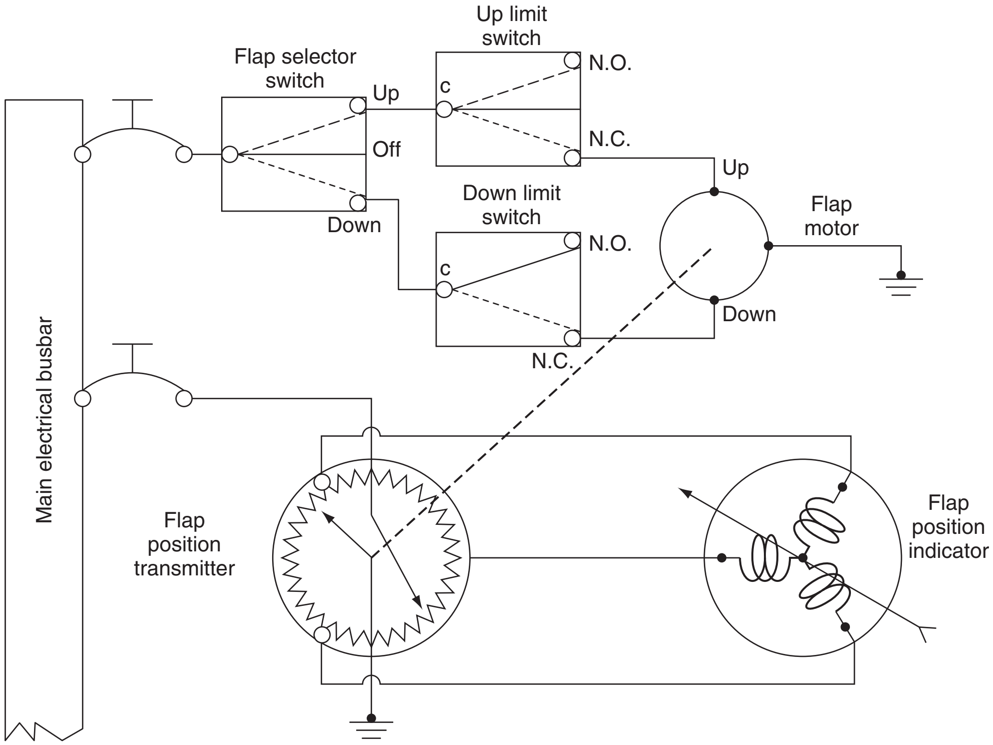
Up limit switch
Flap selector switch
Up
Off
Down
c
N.O.
N.C.
Down limit switch
c
N.O.
N.C.
Up
Flap motor
Down
Main electrical busbar
Flap position transmitter
Flap position indicator

Figure 14.3 Trailing edge flaps control and indication

travel, and this allows the pilot to stop/start the flaps. Selecting 'flaps down' applies power through the down limit switch and the motor runs in the opposite direction; the up-limit switch changes over to the position shown as solid lines in preparation for the next selection. The flaps continue to travel until they are fully extended and the down-limit switch is operated (solid line). This opens the supply to the motor. Selecting off at any time stops the motor at the given position.

The flap position indicator is an integral part of the system; a mechanical linkage is made between the flap control mechanism and position sensor. This varies the ratio of current in each of three coils in the indicator; flap position is displayed in degrees of flap movement. Flap position can be from a simple rheostat and ratiometer indication, or from a synchro system. Flaps on larger aircraft are normally driven by a hydraulic system; flap position is selected in specific settings referred to by simple up/down positions, or (on larger aircraft) by angular positions, see Fig. 14.4. The flap selection switch in this photograph is to the right of the pedestal and, in this example, are also referenced to the maximum airspeed for given flap settings.

Key point

An electrical flap drive system uses a reversible DC motor.

14.3 Control surfaces

Different aircraft types have a variety of control surface position indicators; the various flying control surfaces on a large passenger aircraft are illustrated in Fig. 14.5. These control surfaces typically include:

- leading edge slats
- spoilers
- trailing edge flaps
- rudder(s)
- elevators
- ailerons.

The position of each control surface can be displayed by an analogue indicator (Fig. 14.6) or form part of an electronic display. Each of the control

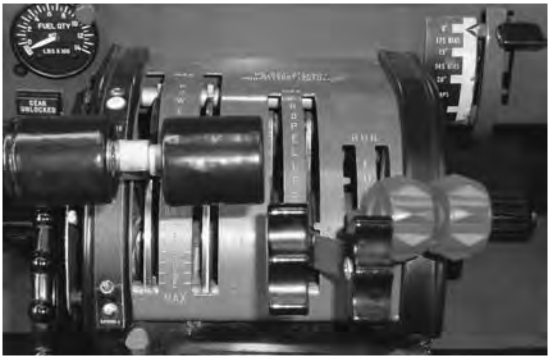

Figure 14.4 Flap position selection (top right of photo)

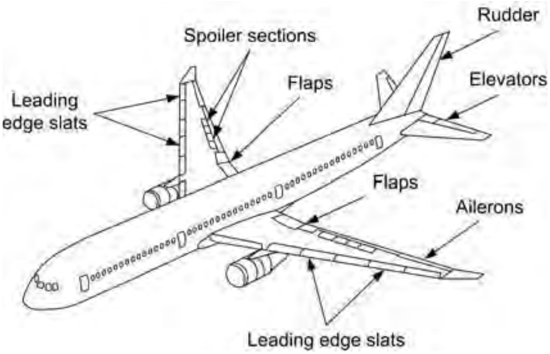

Figure 14.5 Flying control surfaces

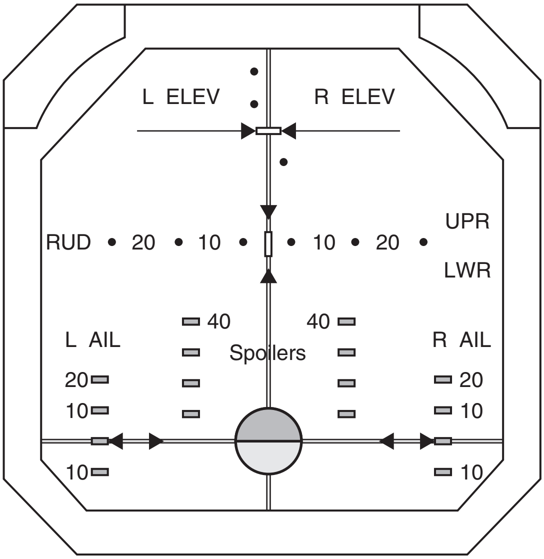

Figure 14.6 Control surface position indication

surfaces is fitted with a position sensor, typically a torque synchro.

Key point

Variable position features include control surface and flap position.

Key point

Two state position devices include micro-switches and proximity sensors.

14.4 Electronic indicating systems

The two systems used on the majority of passenger aircraft are EICAS and ECAM. The **Engine Indication and Crew Alerting System** (EICAS) is a Boeing developed system that provides all engine instrumentation and crew annunciations in an integrated format; this is covered in Chapter 10 (engine systems). The system used on Airbus aircraft is the **Electronic Centralized Aircraft Monitoring** (ECAM) system. The two systems operate on different philosophies; however their basic functions are to monitor aircraft systems and display relevant information to the pilots. Both systems produce warning, cautions and advisory messages that need to be evaluated by the crew; in certain cases, the system provides the procedures required to address the problem. Each colour display unit uses either an active matrix liquid crystal display (AMLCD) or a cathode ray tube (CRT), see Fig. 14.7.

ECAM provides the main features of EICAS but also displays corrective action to be taken by the crew as well as system limitations after the failures. Using a colour coded hierarchy, the crew can assimilate the information being presented and take the necessary corrective action. ECAM was first introduced in the Airbus A320 and provides the **paperless cockpit** since procedures for abnormal and emergency situations are presented on each of the displays.

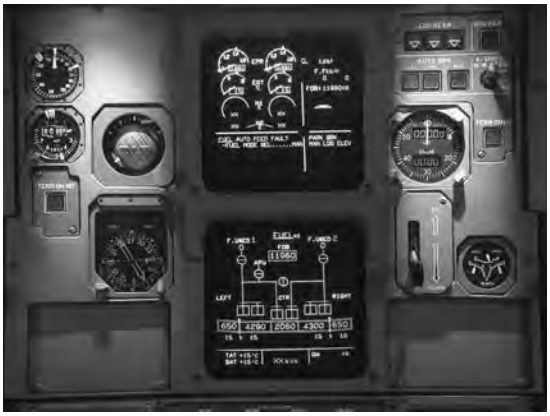

Figure 14.7 ECAM displays

ECAM comprises a series of integrated systems that display information to the crew in an efficient way. Referring to Fig. 14.8, aircraft sensors are categorized into key monitoring functions; these sensors transmit data into two system data acquisition concentrators (SDAC). The data is processed and supplied into flight warning computers (FWC). The FWCs are programmed to identify any inconsistencies in the data, and then output the data through three display management computers (DMC).

If system fault or event is detected, one of the FWCs generates the appropriate warning messages and aural alerts. Critical systems such as engine and fuel quantity are routed directly into the FWCs so that they can still be monitored in the event of both SDACs failing. (ECAM can tolerate the a failure of one SDAC and one FWC and still continue to operate.)

The typical synoptic for an electrical system is shown in Fig. 14.9. In this illustration, the power sources supplying specific busbars is shown graphically, together with information such as battery and transformer rectifier (TR) unit voltages and supply currents. Generator capacity is shown as a percentage of maximum, together with output voltage and frequency.

Aircraft system failures are prioritized as level 1, 2 or 3 failures for the upper and/or lower display. The warning and caution hierarchy is as follows.

Level 3 failures – red warnings: these are situations that require immediate crew action and indicate that the aircraft is in danger. Examples of level 3 failures include an engine fire or loss of cabin pressure. Level 3 system failures illuminate the red master warning light, a warning (red) ECAM

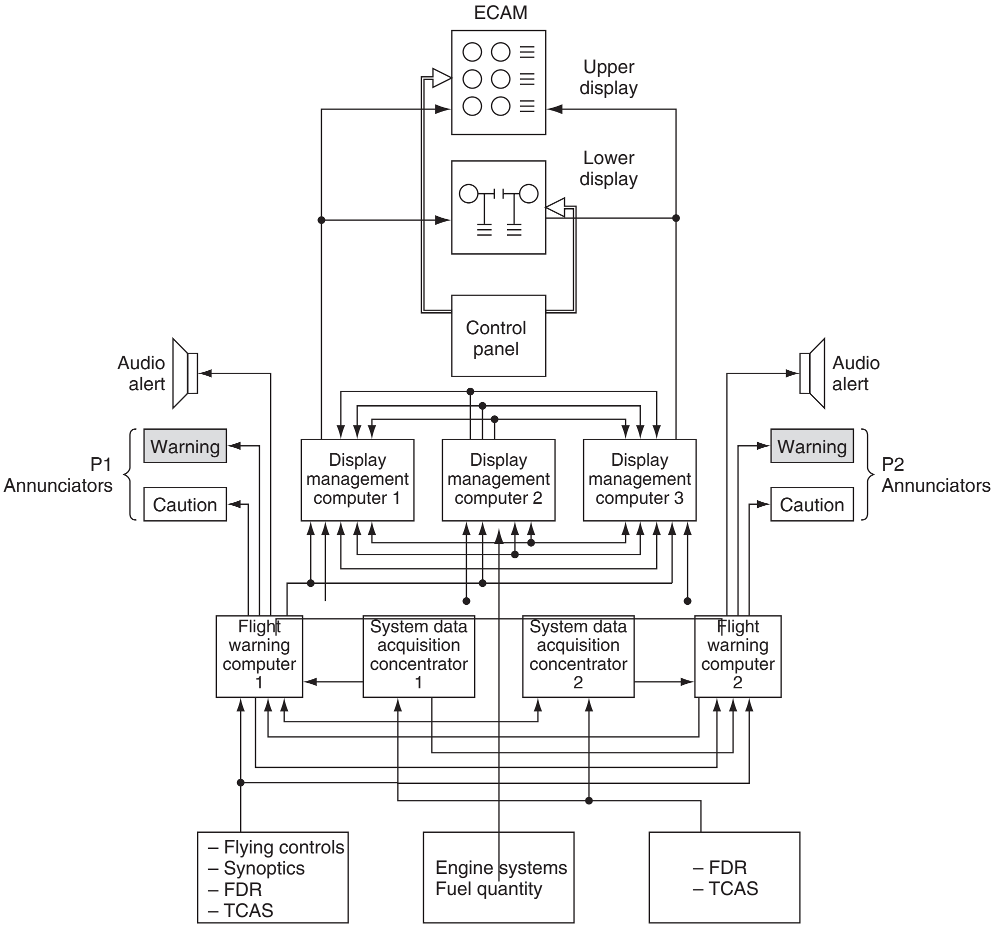
ECAM
Upper display
Lower display
Control panel
Audio alert
Audio alert
P1 Annunciators
Warning
Caution
Warning
Caution
P2 Annunciators
Display management computer 1
Display management computer 2
Display management computer 3
Flight warning computer 1
System data acquisition concentrator 1
System data acquisition concentrator 2
Flight warning computer 2
– Flying controls
– Synoptics
– FDR
– TCAS
Engine systems
Fuel quantity
– FDR
– TCAS

Figure 14.8 ECAM schematic

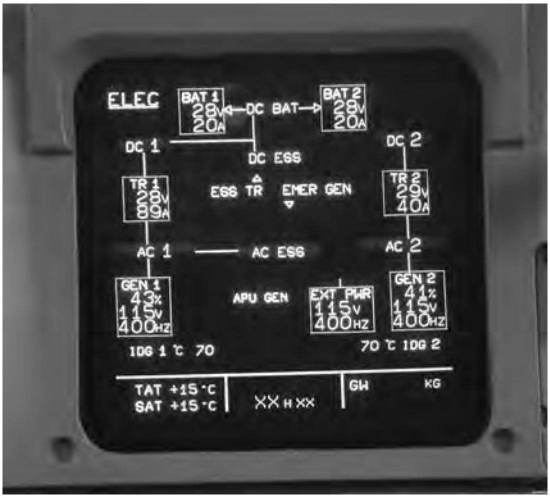
ELEC
BAT 1 28V 20A
DC BAT
BAT 2 28V 20A
DC 1
DC ESS
DC 2
TR 1 28V 89A
ESS TR
EMER GEN
TR 2 29V 40A
AC 1
AC ESS
AC 2
GEN 1 43% 115V 400HZ
APU GEN
EXT PWR 115V 400HZ
GEN 2 41% 115V 400HZ
IDG 1 °C 70
70 °C IDG 2
TAT +15 °C
SAT +15 °C
XX H XX
GW
KG

Figure 14.9 Electrical system synoptic

message and an aural warning; this can be a continuous repetitive chime, a specific sound or a synthetic voice. Pressing the master warning push button silences the aural warning.

Level 2 failures – amber cautions: these are situations that require crew attention but not immediate action. Examples of level 2 failures include bleed air failure or a fuel system fault. Level 2 failures have no immediate or direct impact on flight safety; cautions are displayed to the crew by an amber master caution light, an amber ECAM message and a single chime.

Level 1 failures – these are system failures and/or faults that could lead to a loss of system redundancy. Level 1 failures require monitoring but have no immediate impact on continued safe operation of the aircraft. Examples of level 1 failures include the loss of a fuel system temperature sensor. Level

1 failures are displayed to the crew by amber ECAM messages only (no aural warning).

14.5 Multiple choice questions

1. Indications of landing gear fully down and locked are:
   (a) red lights on, green lights off
   (b) red lights on, green lights on
   (c) red lights off, green lights on.
2. Two state conditions include:
   (a) flap position or doors (open or closed)
   (b) landing gear (up or down) or doors (open or closed)
   (c) landing gear (up or down) or control surface position
3. An electrical flap drive system uses a:
   (a) reversible DC motor
   (b) variable speed DC motor
   (c) unidirectional DC motor.
4. Level 3 ECAM failures are indicated by:
   (a) red warnings, requiring immediate crew action
   (b) amber cautions, requiring crew attention
   (c) red warnings, having no immediate impact on the aircraft.
5. Variable position features include:
   (a) doors (open or closed) and flap position
   (b) control surface and proximity sensor position
   (c) control surface and flap position.
6. When an electrically operated landing gear is fully retracted, the up-lock switch contacts:
   (a) open, thereby removing power from the motor
   (b) close, thereby removing power from the motor
   (c) open, thereby applying power to the motor.
7. Engine fire or loss of cabin pressure would be displayed on ECAM as:
   (a) level 3 failures
   (b) level 2 failures
   (c) level 1 failures.
8. Two state position devices include:
   (a) micro switches and variable resistors
   (b) synchros and proximity sensors
   (c) micro switches and proximity sensors.
9. Electrically driven flaps continue to travel until they are:
   (a) fully retracted and the down-limit switch is operated
   (b) fully extended and the down-limit switch is operated
   (c) fully extended and the up-limit switch is operated.
10. Variable position devices include:
    (a) synchros and variable resistors
    (b) micro switches and variable resistors
    (c) synchros and proximity sensors.

Chapter 15

Warning and protection systems

Various systems are installed on aircraft to protect them from a variety of hazards including: stalling, ice, rain, unsafe configuration during takeoff and skidding. Stall protection systems provide the crew with a clear and distinctive warning before the stall is reached. The primary sensor required for this protection system is the aircraft's angle of attack. Flying in ice and/or rain conditions poses a number of threats to the safe operation of the aircraft. Ice formation can affect the aerodynamics and/or trim of the aircraft. The anti-skid system (also called an anti-lock braking system: ABS) is designed to prevent the main landing gear wheels from locking up during landing, particularly on wet or icy runway surfaces. The configuration warning system (also known as a take-off warning system) provides a warning if the pilot attempts to take-off with specific controls not selected in the correct position, i.e. an unsafe configuration. This chapter describes a variety of protection systems found on a range of aircraft.

15.1 Stall warning and protection

Aircraft wings are in the form of an aerofoil designed to produce lift; the key features of an aerofoil together with specific terminology are shown in Fig. 15.1.

When the airflow passes over the wing without turbulence it is said to have a **streamline airflow**, see Fig. 15.2. The **angle of attack** (AoA) is the angle between the chord line of the wing and the vector representing the relative motion of the aircraft and the atmosphere. (The term angle of incidence is sometimes used instead of angle of attack.)

Basic aerodynamic principles state that increasing the angle of attack increases the lift produced by the wing; when the AoA reaches a certain angle, the airflow over the wing becomes turbulent and the lift is dramatically decreased. When this occurs, the aerofoil is in a **stall** condition as illustrated in Fig. 15.3. It can be seen that the streamlined airflow is breaking up and becoming turbulent.

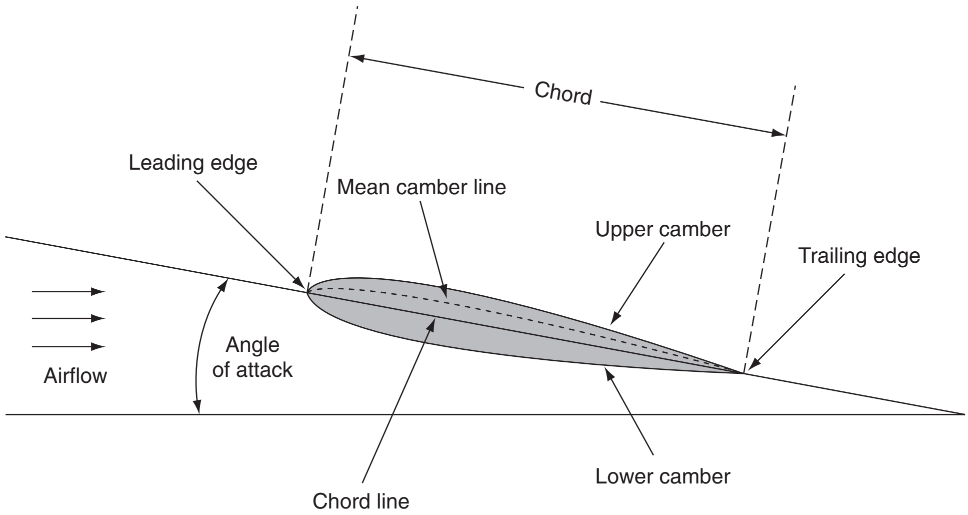

Figure 15.1 Aerofoil terminology

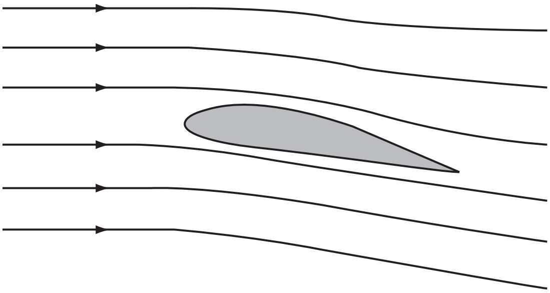

Figure 15.2 Streamline airflow

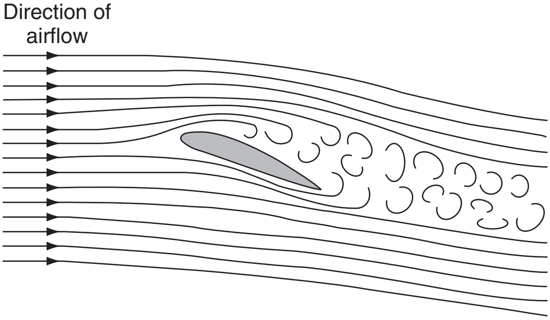

Figure 15.3 Aerofoil stall

It is important to note that an aerofoil stalls at a certain angle; the speed at which the aircraft is flying makes little difference to the stall angle. The relationship between speed and stalling can be related directly to the formula for lift:

$$\text{Lift} = \tfrac{1}{2}\rho V^2 S C_L$$

where ρ is the air density, V is the aircraft speed, S is the wing area, and C_L is the aerofoil's coefficient of lift.

Referring to Fig. 15.4, an aircraft can have a different angle of attack for a given pitch attitude. Aircraft pitch angle should not be confused with angle of attack. Relative airflow over the wing will change direction in flight in relation to the pitch angle of the aircraft.

The critical features of aerofoil performance are centred on the boundary layer as illustrated in Fig. 15.5. When the airflow divides between the upper and lower surfaces, the point of division is at the **stagnation point**. Airflow over the upper surface remains stable until, at higher AoA, it starts to detach; this occurs at the **transition point**, before giving rise to turbulence at the **separation point**. The turbulence behind the separation point increases with AoA.

The relationship between lift and stall is depicted in Fig. 15.6. The two axes on this chart are the lift coefficient (C_L) and angle of attack. It can be seen that lift

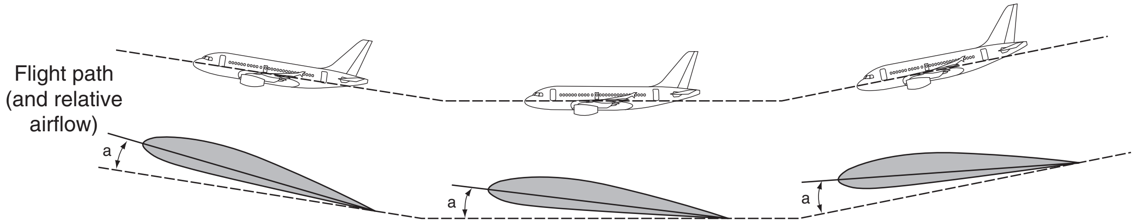

(a) Same AoA, built different pitching angles

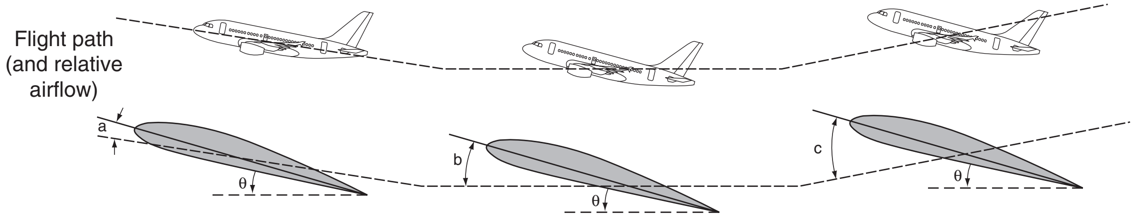

(b) Same pitching angle, but different angles of attack

Figure 15.4 Angle of attack and attitude

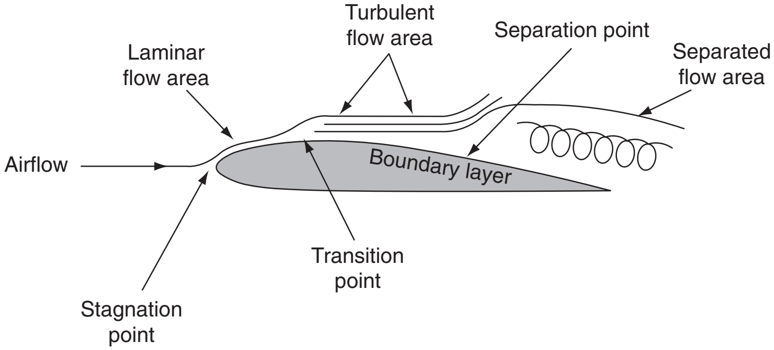

Figure 15.5 Boundary layer separation

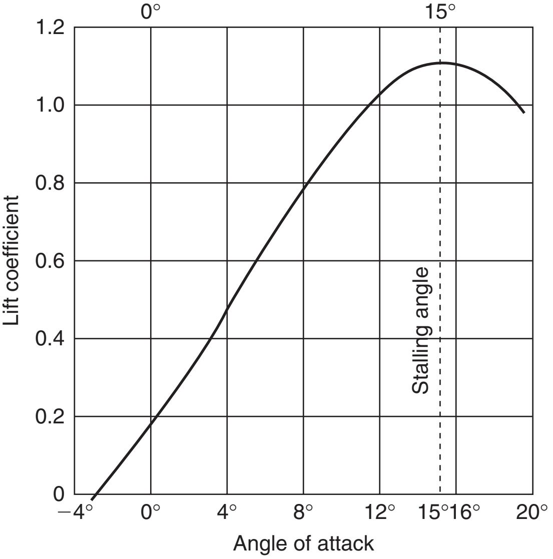

Figure 15.6 Angle of attack and lift

increases in proportion to the AoA over the ordinary angles of flight. At higher angles of attack, the stall condition is reached. Different aerofoil sections have different lift coefficients, and therefore different stall angles, but the principles remain the same.

Deploying lift augmentation devices, e.g. leading and/or trailing edge flaps, can change the wing area and also the wing camber, thereby changing the coefficient of lift. By reducing aircraft speed, the lift will decrease; for a given flap setting, the only other way of increasing lift is to increase the angle of attack. (Aerodynamic theory is covered in more detail in another book from the aircraft engineering series: *Aircraft Engineering Principles*.)

Key point

An aerofoil stalls at a certain angle; the speed at which the aircraft is flying makes little difference to the stall angle.

Warning systems are installed to provide the crew with clear and distinctive warning before the stall is reached. This can be a warning horn and/or warning light on smaller aircraft. On passenger aircraft, the stall warning is provided by a **stick shaker**. Warning systems require a basic sensor to detect this stall condition; typical sensors are:

- reed sensor
- vane sensor
- pressure sensor
- angle of attack sensor.

Key point

Stall warning systems provide the crew with a clear and distinctive warning before the stall is reached.

15.1.1 Reed sensor

Many GA aircraft are installed with a **reed sensor** (Fig. 15.7); this does not require any electrical power.

Airflow is directed from a scoop (Fig. 15.7, item 5) in the wing leading edge and into a reed (3) and horn assembly (2). In the pre-stall condition, the air pressure on the leading edge reduces (relative to cabin pressure)

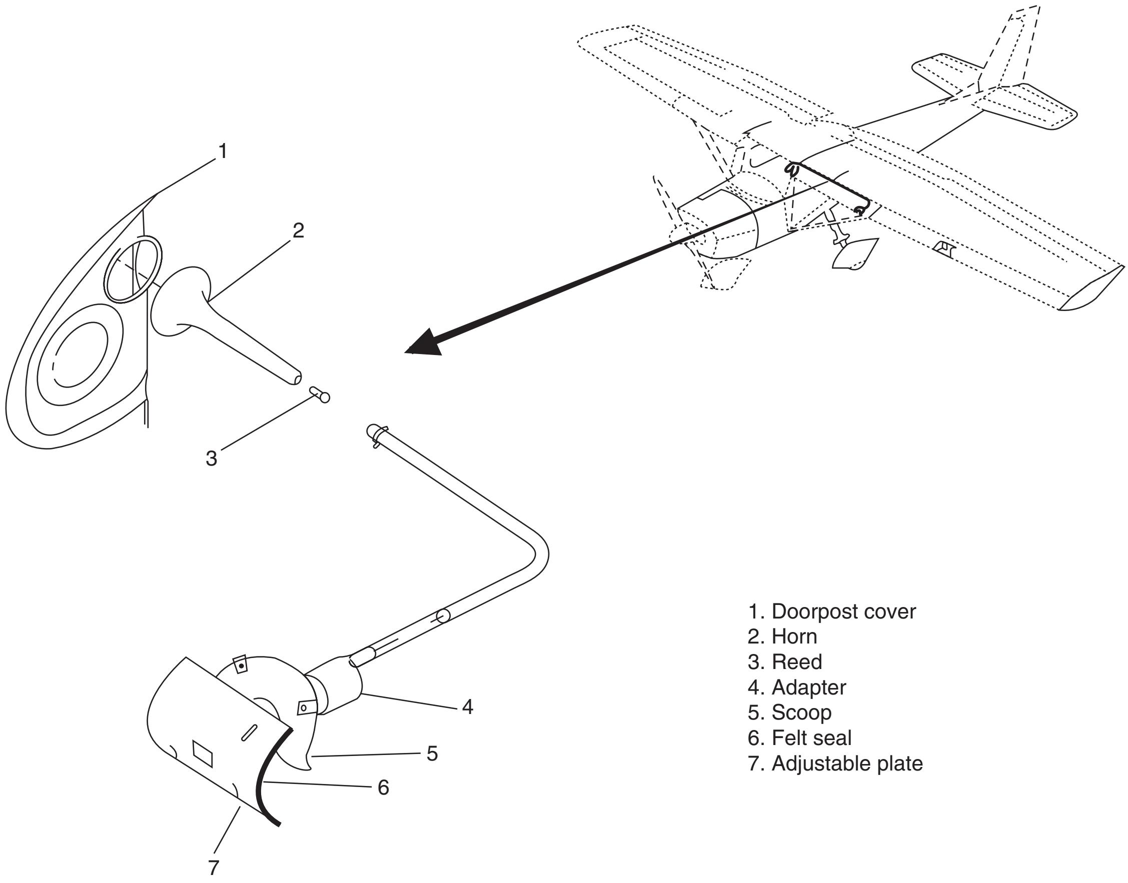

Figure 15.7 Stall warning: reed sensor

because the stagnation point has moved. This draws air through the reed, causing it to vibrate at an audible frequency (much like a musical instrument); this is amplified in an acoustic horn and the crew receive a tone. An adjustable plate (7) is used to alter the airflow intake such that the intake is aligned with stagnation point.

Key point

When the AoA reaches a certain angle, the airflow over the wing becomes turbulent and the lift is dramatically decreased

---

15.1.2 Vane sensor

An alternative device used on GA aircraft is the **vane sensor** located in the wing leading edge as shown in Fig. 15.8. As with the reed sensor, the vane is sensitive to the movement of the stagnation point. The vane is held in a forward position by an internal spring and connected to a micro-switch.

At normal angles of attack, the vane is held back by the airflow (against spring pressure) and the micro-switch is held in the open position, see Fig. 15.9. At higher angles of attack, airflow pressure on the switch is reduced and the vane eventually moves forward (by the spring force) thereby closing the micro-switch. This completes the warning circuit that activates a light and horn.

Key point

At normal angles of attack, the vane sensor is held back by the airflow against spring pressure.

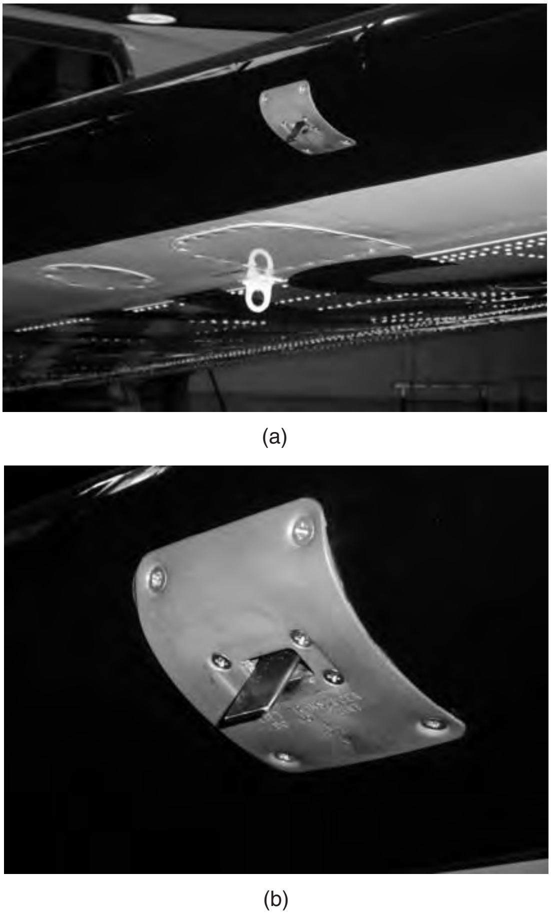

Figure 15.8 Stall warning: vane sensor installation

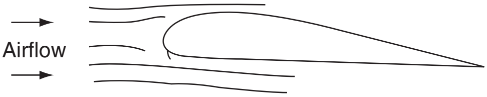

(a) Aerofoil at normal altitude
– vane held back by airflow

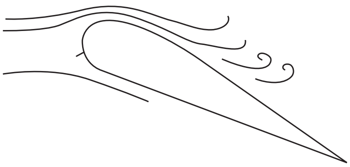

(b) Aerofoil in stall condition
– vane pushed forward by spirng

Figure 15.9 Stall warning: vane sensor principles

15.1.3 Pressure sensing sensor

Another angle of attack sensor technology is based on measuring the **pressure** at two points on the sensor housing, see Fig. 15.10. The conical housing rotates on its axis; slots A and B are connected to a pressure chamber that contains a pivoted vane. The sensor housing aligns with the angle of attack because slots A and B maintain equal pressures on the vane, and hence the conical housing.

If the angle of attack increases, slot B has increased pressure compared with slot B. The vane moves to re-align the housing with the airflow to equalize the pressures in both slots. Rotation of the housing is detected by a potentiometer; the centre contact picks off a signal voltage tapped from the resistance windings and this is used to measure the angle of attack.

Key point

When airflow passes over the wing without breaking up, it is said to have streamline airflow.

15.1.4 Sensor vane

The final type of stall warning sensor to be described is illustrated in Fig. 15.11. An angle of attack (AoA) **sensor vane** aligns itself with the prevailing airstream; this rotates a shaft inside the housing. The vane's shaft is connected to a synchro that provides an electrical output proportional to the angle of attack.

A **viscous damper** connected to the AoA vane stabilizes vane movements and reduces the effects of turbulence. The AoA sensor contains a heater that provides continuous de-icing/anti-icing, prevents condensation and reduces changes in damper fluid viscosity.

15.1.5 Stick shaker

On larger aircraft the stall warning system comprises an angle of attack vane and stick shaker. A motor is attached to one or both control columns; an out-of-balance weight is attached to the motor shaft so that it vibrates when the motor is running. The motor design is matched to the control column to provide a distinctive frequency of approximately 10–30 Hz together with physical movement of the controls.

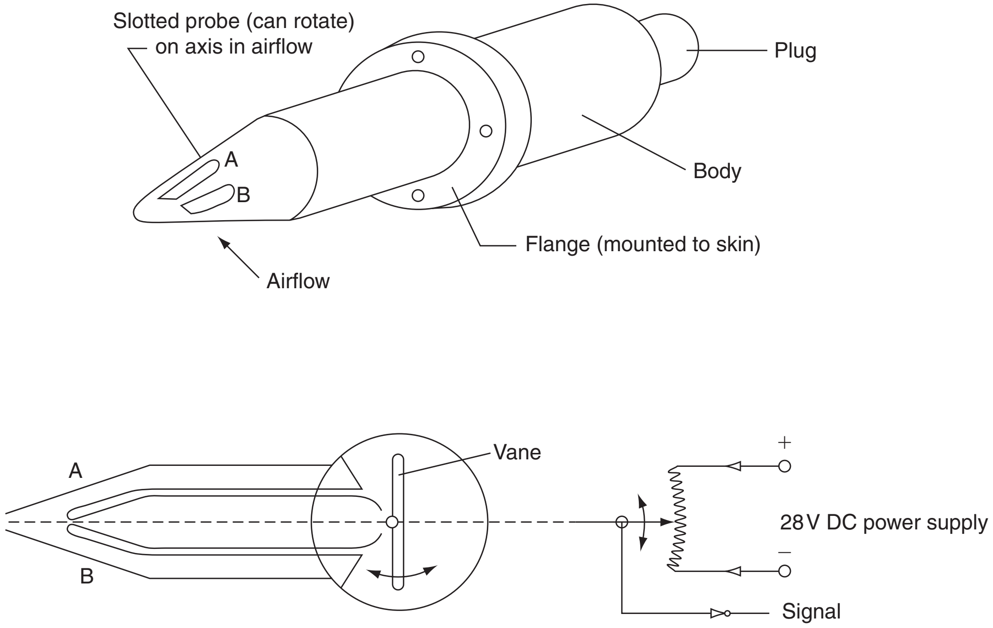
Slotted probe (can rotate) on axis in airflow
A
B
Plug
Body
Flange (mounted to skin)
Airflow
A
Vane
+
28 V DC power supply
–
B
Signal

Figure 15.10 Stall warning; pressure sensor

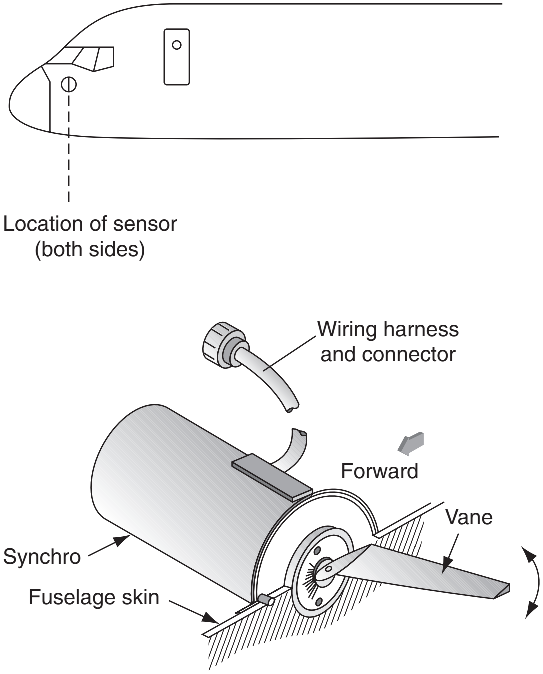
Location of sensor (both sides)
Wiring harness and connector
Forward
Vane
Synchro
Fuselage skin

Figure 15.11 Stall warning: AoA sensor

Key point

An angle of attack (AoA) sensor vane aligns itself with the prevailing airstream.

15.1.6 Stall identification system

Certain aircraft types, typically **tee-tailed aircraft**, are installed with stall identification systems; this is normally integrated with the stall warning system. The tee-tailed aircraft can be vulnerable to a deep-stall as illustrated in Fig. 15.12. In addition to the wings (or main-planes) stalling, the deep stall condition also makes the elevators less effective; this compounds the problem, since the elevators are the means of pitching the aircraft nose down to recover from the stall. Furthermore, if the aircraft has rear-mounted engines, the airflow into the engine is turbulent, possibly causing a loss of performance.

The stall identification system contains an actuator that pushes the control column forward with a sharp

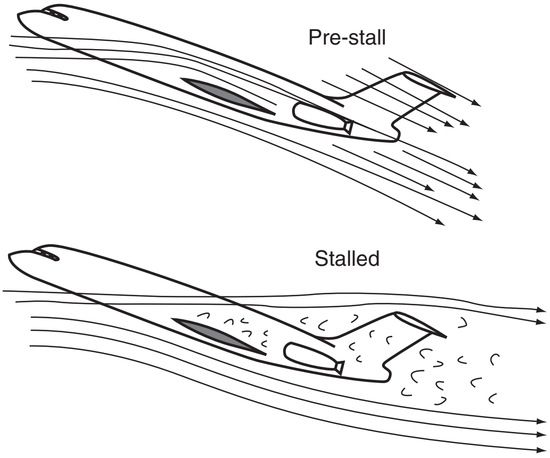

Figure 15.12 Deep stall

and positive action. The **stick push** is released once the aircraft is returned to an acceptable angle of attack. This system has to be highly reliable, both in terms of operating when need and having low probability of false operation.

Key point

The stall identification system contains an actuator that pushes the control column forward.

15.2 Airframe ice and rain protection

Flying in ice and/or rain conditions poses a number of threats to the safe operation of the aircraft. Ice formation can affect the aerodynamics and/or trim of the aircraft. Icing occurs when **super-cooled water** makes contact with the airframe and/or ice particles. Super-cooled water exists at temperatures below freezing point; this is because water needs nuclei to form ice crystals. The water freezes on the surface of the nuclei, and then grows in size by forming layers. Water freezes when energy is given up to its surroundings. Ice accumulation on an aircraft can causes secondary damage by breaking off and being ingested into the engine. Ice and rain also reduces visibility through the windscreen. External sensors and equipment can also be affected by the build-up of ice, e.g. pitot tubes that sense airspeed and angle of attack sensors, and these must not be allowed to freeze.

15.2.1 Ice detection

Various technologies are available to provide a warning to the crew of icing conditions. The simple method is to monitor the outside air temperature and atmospheric conditions. When flying at night, an ice inspection light (see Chapter 12) can be used to illuminate critical areas, e.g. wing leading edges. Automatic ice detectors are normally located at the front of the aircraft. Various technologies are employed; the function of the sensor is to detect **ice accretion** and provide a warning to the crew and/or turn on the ice protection systems.

One type of automatic ice detector consists of a motor-driven sensor that is located against a **knife-edge cutter**, see Fig. 15.13, at the front of the fuselage. As ice builds on the rotor, the gap closes with the knife-edge and torque is applied to the rotor. The body of the motor is held in position by springs, and as the torque increases, the motor starts to rotate in its mounting until a pre-determined point is reached and a micro-switch is closed, thereby operating a warning light.

An alternative technology uses an **ultrasonic** ice detector; see Fig. 15.14. The sensing probe is exposed to the airflow and is vibrated by an electromagnetic coil at a natural frequency of 40 kHz.The probe vibrates at a lower frequency when it accumulates ice due to its increased mass.

A logic unit within the detector housing (Fig. 15.15) determines when the probe vibration is less than 39,867 Hz; this occurs when a known mass of ice has been accumulated, and the cylinder is heated to melt the ice.

A timing schedule (Fig. 15.16) is programmed into the control unit and monitors the time taken for the frequency to change. The heater remains on until the probe vibration returns to its nominal 40 kHz value. Nominal heating time is six seconds; if the heater is on for more than 25 seconds, power is removed, and a fault condition is notified to the crew. The detector provides control functions for selection of both engine and wing ice protection.

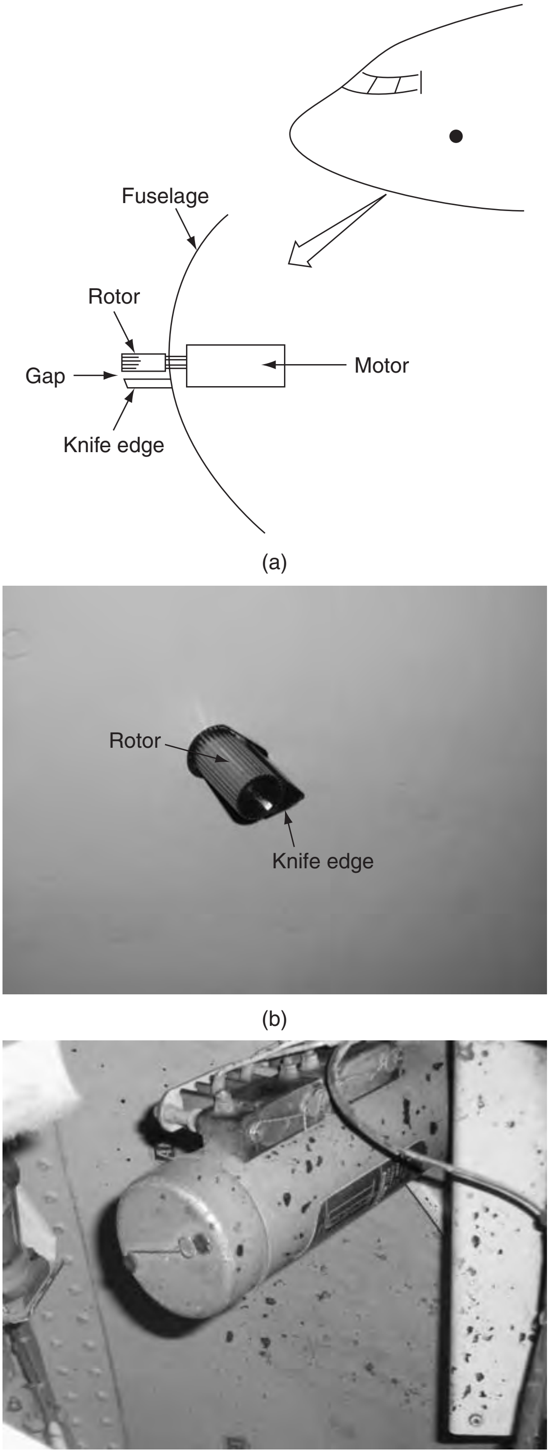

Figure 15.13 Ice detector: knife edge cutter. (a) Schematic, (b) sensor, (c) motor

---

Key point

When the ultrasonic ice detector probe accumulates ice, it vibrates at a lower frequency.

Fault conditions are reported to the on-board maintenance system. Weight on wheels logic prevents the detector from being heated on the ground. Press to test (PTT) is used to check the detector's logic functions and heater continuity.

15.2.2 Ice protection

Two strategies are used for ice protection: de-icing and anti-icing. De-icing allows ice to form and then be removed on a periodic basis. The build-up of ice will have been investigated during type testing of the aircraft and the build-up removed before it poses a hazard. Anti-icing is when ice is not permitted to form at all. Three primary methods are used for both de-icing and anti-icing:

- fluid
- pneumatic
- thermal.

All three are controlled and operated by electrical systems; an aircraft can be fitted with any one method, or a combination of all three. Specific areas to be protected from ice formation are as follows:

Airframe
  - leading edges
  - control surfaces
  - lift augmentation devices
  - windscreen (or windshield)

Propulsion
  - air intakes
  - propellers

External components
  - pitot tubes
  - temperature sensors
  - angle of attack sensors
  - water drains.

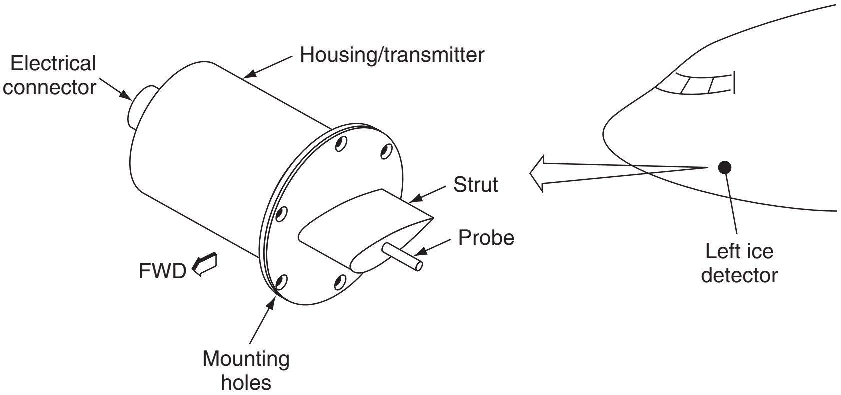
Electrical connector
Housing/transmitter
Strut
Probe
FWD
Mounting holes
Left ice detector

Figure 15.14 Ice detector: ultrasonic sensor

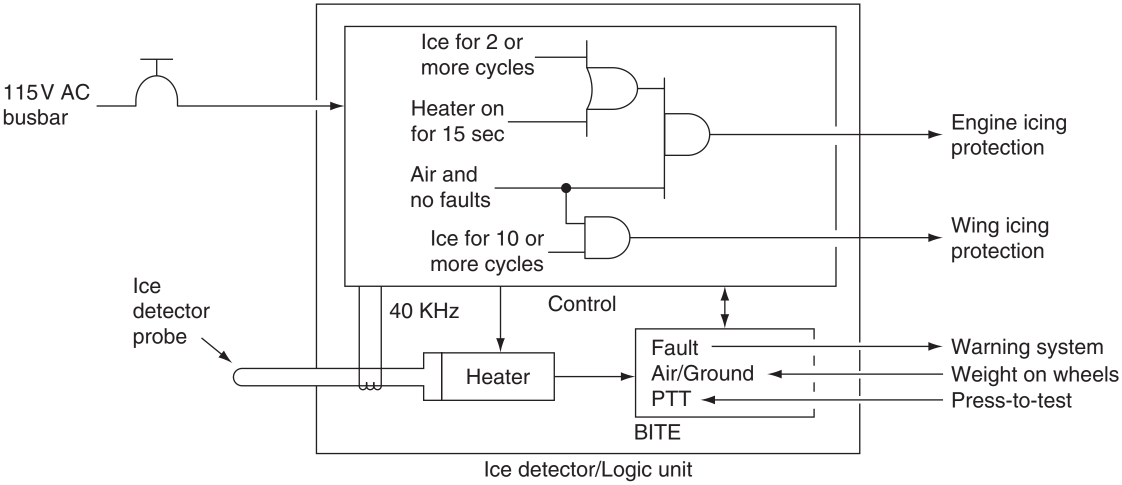
115 V AC busbar
Ice for 2 or more cycles
Heater on for 15 sec
Air and no faults
Ice for 10 or more cycles
Engine icing protection
Wing icing protection
Ice detector probe
40 KHz
Control
Heater
Fault
Air/Ground
PTT
BITE
Warning system
Weight on wheels
Press-to-test
Ice detector/Logic unit

Figure 15.15 Ice detector: ultrasonic system

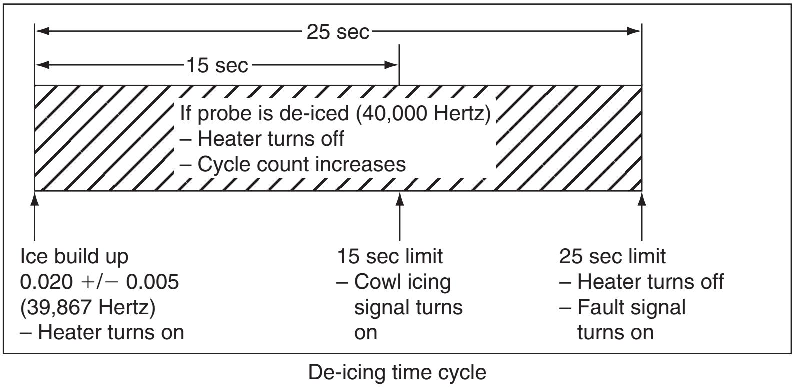
25 sec
15 sec
If probe is de-iced (40,000 Hertz)
– Heater turns off
– Cycle count increases
Ice build up
0.020 +/− 0.005
(39,867 Hertz)
– Heater turns on
15 sec limit
– Cowl icing signal turns on
25 sec limit
– Heater turns off
– Fault signal turns on
De-icing time cycle

Figure 15.16 Ice detector: ultrasonic system timer

15.2.2.1 De-icing fluids

Fluids are typically used on wings, vertical stabilizer, horizontal stabilizer, propellers and windscreens. On-board fluid protection is applied through pipes and/or small holes in the airframe. The fluid is transferred from a storage tank by electrical pumps and directed to the required zones by electrically operated valves.

The fluid can also be applied on the ground by specially equipped vehicles with booms to allow easy access to the entire aircraft. **De-icing fluids** are typically composed of ethylene glycol or propylene glycol, together with thickening agents, corrosion inhibitors, and coloured dye. The aircraft is sprayed with a fluid that melts any existing ice on the aircraft and also prevents ice formation prior to takeoff. The timing of when the aircraft is sprayed has to take into account the weather conditions (in particular ambient temperature, wind speed, precipitation, humidity) aircraft skin temperature, and the scheduled departure time.

Fluid performance is characterized by its **holdover time**: this is the period of time that an aircraft remains protected against ice prior to takeoff, and can be up to 80 minutes depending on conditions and fluid additives. The coloured dye is used so that treated areas can be readily identified.

Key point

De-icing allows ice to form and then be removed on a periodic basis.

15.2.2.2 Pneumatic ice protection

Pneumatic ice protection is used on the leading edge of wings and the vertical/horizontal stabilizers. Fabric reinforced rubber tubes (known as **boots**) are attached to the surfaces of leading edges of wings, tail planes and fins, see Fig. 15.17. These tubes are inflated and deflated with air on a cyclic basis to cause a slight deformation of the boot; this causes the ice to break off.

The air supply is from an electrically driven pump, or bled from the engine compressor via a regulator. Electrical solenoid valves direct the air into the boots either in sequence or simultaneously. The tubes are deflated and kept flush with the airframe when not in use; this is achieved by connecting the boots to a vacuum source. The system can be operated automatically or manually.

Key point

When not in use, pneumatic de-icing boots are deflated and kept flush with the airframe.

15.2.2.3 Thermal ice protection

Thermal ice protection can be achieved by one of two methods: hot air bled from the engine compressor, or electrical heaters. Hot air is directed to the required zone(s) by electrically operated valves controlled from the flight compartment or by a control unit. Temperature sensors are used as part of a control

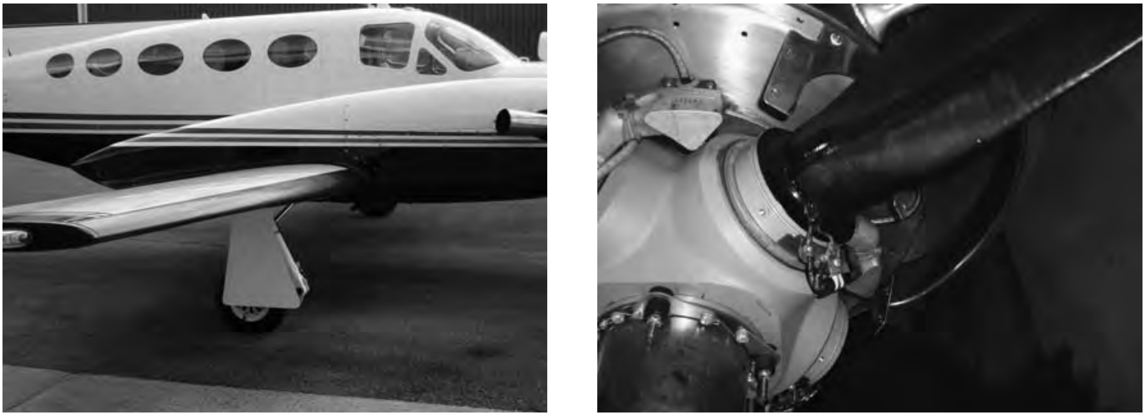

(a) (b)

Figure 15.17 Pneumatic ice protection: (a) wing leading edge, (b) electrical connections

and feedback system to provide overheat protection. Applications for **bleed air ice protection** include the leading edge of wings, vertical/horizontal stabilizers and engine air intakes.

Key point

Anti-icing is used where ice is not permitted to form at any time.

Electrical heating systems are used for both anti-ice and de-icing; heater mats are bonded to the:

- airframe
- engine intakes
- propellers
- rotor blades
- windscreens.

The heater mats are formed from fine-wire elements inside layers of insulation and protective materials. Typical heating elements are nickel, nickel-chrome or copper-nickel. Larger aircraft use 115V AC variable frequency power supply. Smaller aircraft use a 28V DC power supply.

Propeller de-icing is achieved by electrical heating elements that are bonded to the leading edge of the blades, see Fig. 15.18. Some aircraft have heating mats on the inboard and outboard sections of the propeller blade. The elements are connected to the 28V DC power supply via slip rings and brushes inside the hub. Typical current requirements for propeller heating are 15–20A.Power is applied from a timer control circuit that alternates the heating of outer and inner mats on each of the propeller blades over a 30 second cycle.

Engine intake leading edges are continuously heated when the system is turned on (anti-icing). This applies to gas turbine engines and turbo-props). The inner and outer surfaces are heated via a cyclic timer to provide de-icing. Engine intakes are installed with mats that are shaped to provide breaker strips; this ensures that the ice breaks off in pre-determined sections in a controlled sequence.

Test your understanding 15.1

What is the difference between anti-icing and de-icing?

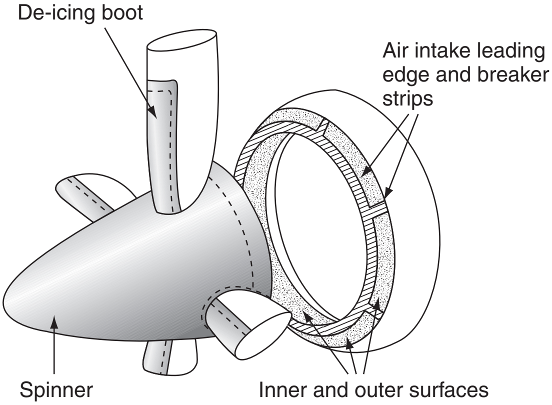

Figure 15.18 Propeller de-icing

Larger aircraft use engine bleed air for thermal anti-icing. Fig. 15.19 illustrates how ducts are located along the leading edge of the wing. Spray holes direct the air inside the leading edge before it is ejected overboard. Similar installations are used for tail plane and engine nacelles.

15.3 Windscreen ice and rain protection

Various methods of windscreen ice and rain protection are used on a range of aircraft types. One method uses a metallic film deposited onto the surface of the screen; this is connected to an electrical supply on either side of the panel. An alternative technology uses fine wire elements (Fig. 15.20) sandwiched within the laminated glass panel. Individual windscreen heaters use 4kW of power, to keep window temperature at approximately 30°C.

Automatic cycling of power regulates the amount of heat being absorbed by the windscreen. Power is isolated from the heaters in the event of fault conditions; temperature sensors detect overheat and current transformers detect electrical overload conditions. Temperature sensors are monitored by a control unit; these can be simple thermostats or thermistors.

Key point

Windscreens are kept clear of rain by wipers, rain repellent (sprayed on the windows), or by treating the glass with a hydrophobic coating.

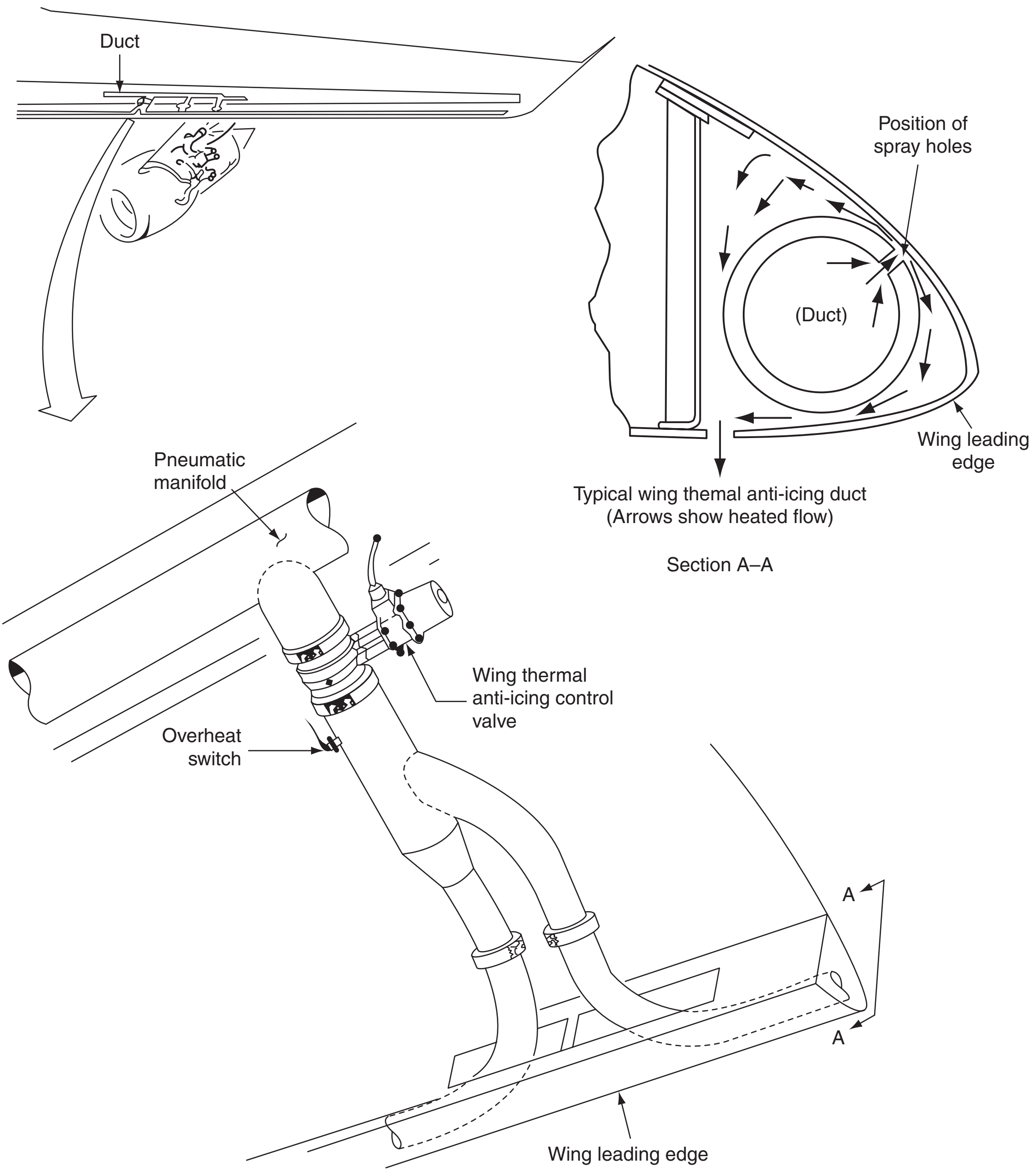
Duct
Position of spray holes
(Duct)
Wing leading edge
Typical wing themal anti-icing duct
(Arrows show heated flow)
Section A–A
Pneumatic manifold
Wing thermal anti-icing control valve
Overheat switch
A
A
Wing leading edge

Figure 15.19 Thermal anti-icing

15.3.1 Windscreen wiper

The windscreen wiper system is based on 28V DC variable speed motors; the rotary motion of the motor is changed by a gear mechanism in the converter to produce the sweeping motion of the wiper arm over the windscreen. A typical windscreen wiper circuit is shown in Fig. 15.21. The normal arrangement is to have one wiper assembly per screen to ensure that at least on pilot can keep a clear screen in the event of

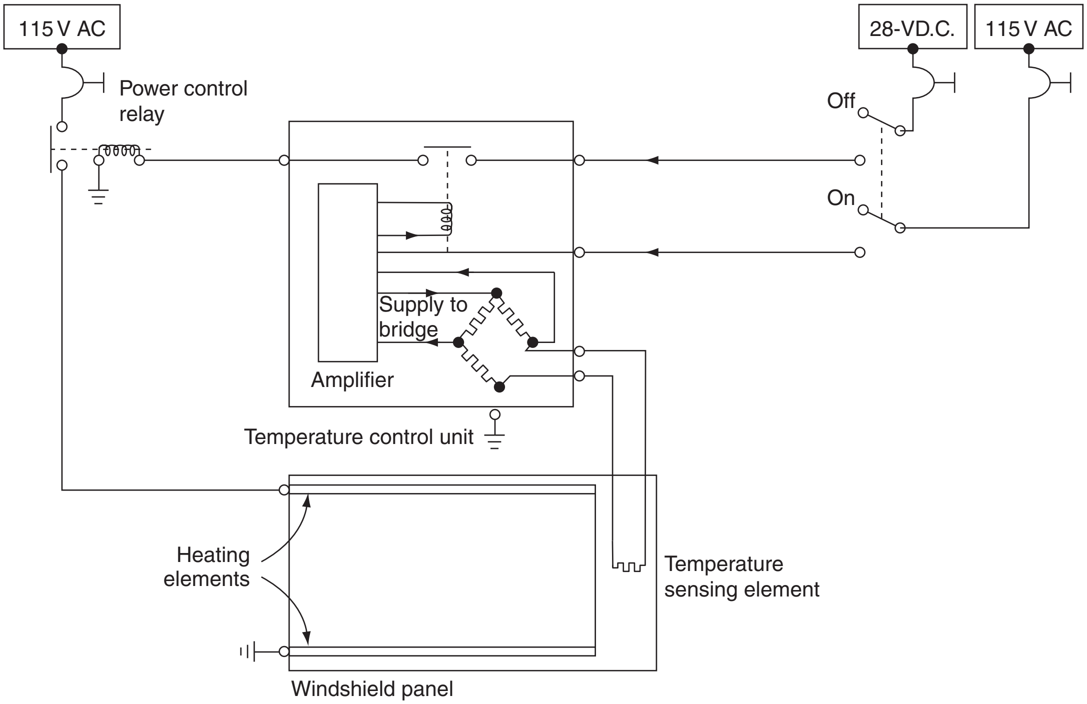

Figure 15.20 Windscreen de-icing circuit

failure. The motors are set by the control switch; typical wiper speeds are:

- **low**, 160 cycles per minute
- **high**, 250 cycles per minute.

A parking switch in the motor/converter sets the wiper blade to the park position when the wiper is selected off. In the off position, the park switch in each motor/converter closes; this causes the blades to position themselves at the bottom of each windscreen.

15.3.2 Rain repellent

This system is used to maintain a clear area on the windscreen during take-off and landing; a typical system is shown in Fig. 15.22. The rain repellent bottle is located inside the fuselage roof; this contains a pressure gauge, visual contents reservoir and a shut-off valve.

The control panel (normally combined with the wiper system) contains two switches to control the rain repellent system. Figure 15.23 illustrates a typical rain repellent system schematic. The repellent fluid is stored in a container; this is pressurized from an external air supply. Two electrically operated solenoid valves control the flow of repellent to the windscreens. Nozzles on the fuselage (forward of the windscreen) direct the repellent spray onto the windscreens.

The rain repellent switches are momentary make type; pushing a switch once activates the respective solenoid valve. A time delay circuit in the valve controls the amount of repellent being released by closing the valve after a short time period, typically ½ second. A pressure gauge indicates when the container needs replacing.

15.4 Anti-skid

The anti-skid system (also called an anti-lock braking system: ABS) is designed to prevent the main landing gear wheels from locking up during landing, particularly on wet or icy runway surfaces. Traditional method used to operate brakes is from hydraulic pressure controlled from the brake pedals. This pressure can be varied by the pilot by increasing or decreasing the amount of force being applied to the pedals. If too much pressure is applied by the pilot, the wheels will lock-up and the aircraft will skid on the runway.

The anti-skid system ensures optimum braking under all conditions by modifying the pressure being exerted by the pilot on the brake pedals.

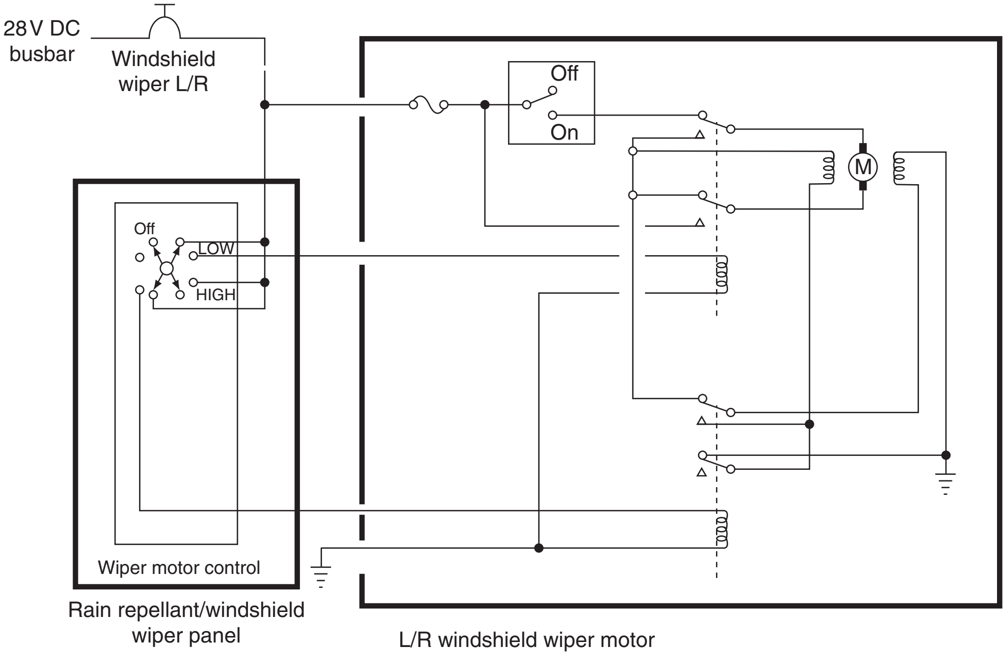
28V DC
busbar
Windshield
wiper L/R
Off
On
M
Off
LOW
HIGH
Wiper motor control
Rain repellant/windshield
wiper panel
L/R windshield wiper motor

Figure 15.21 Windscreen wiper circuit

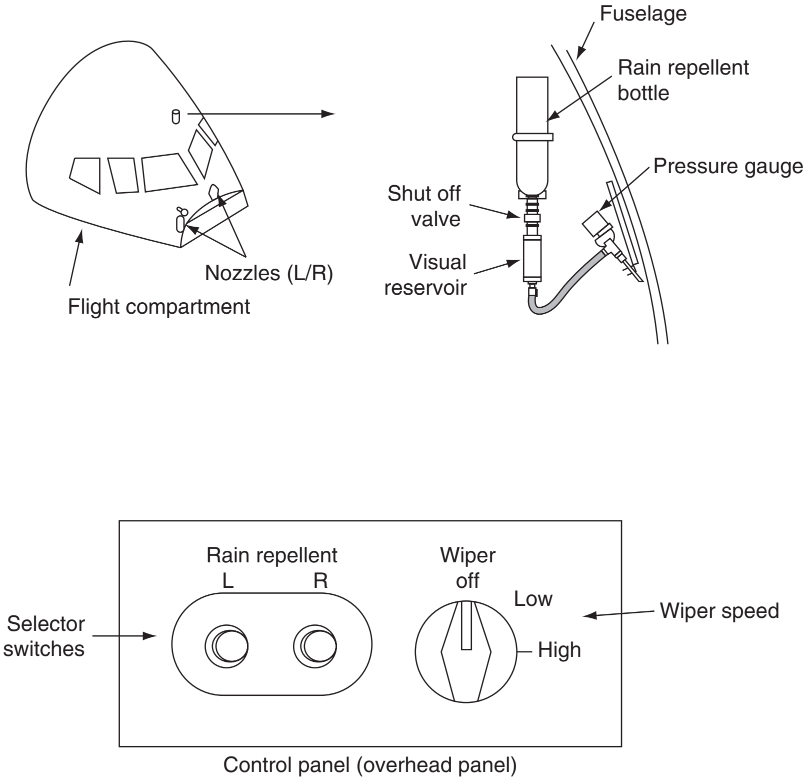
Fuselage
Rain repellent
bottle
Pressure gauge
Shut off
valve
Visual
reservoir
Nozzles (L/R)
Flight compartment
Rain repellent
L
R
Wiper
off
Low
High
Wiper speed
Selector
switches
Control panel (overhead panel)

Figure 15.22 Rain repellent system overview

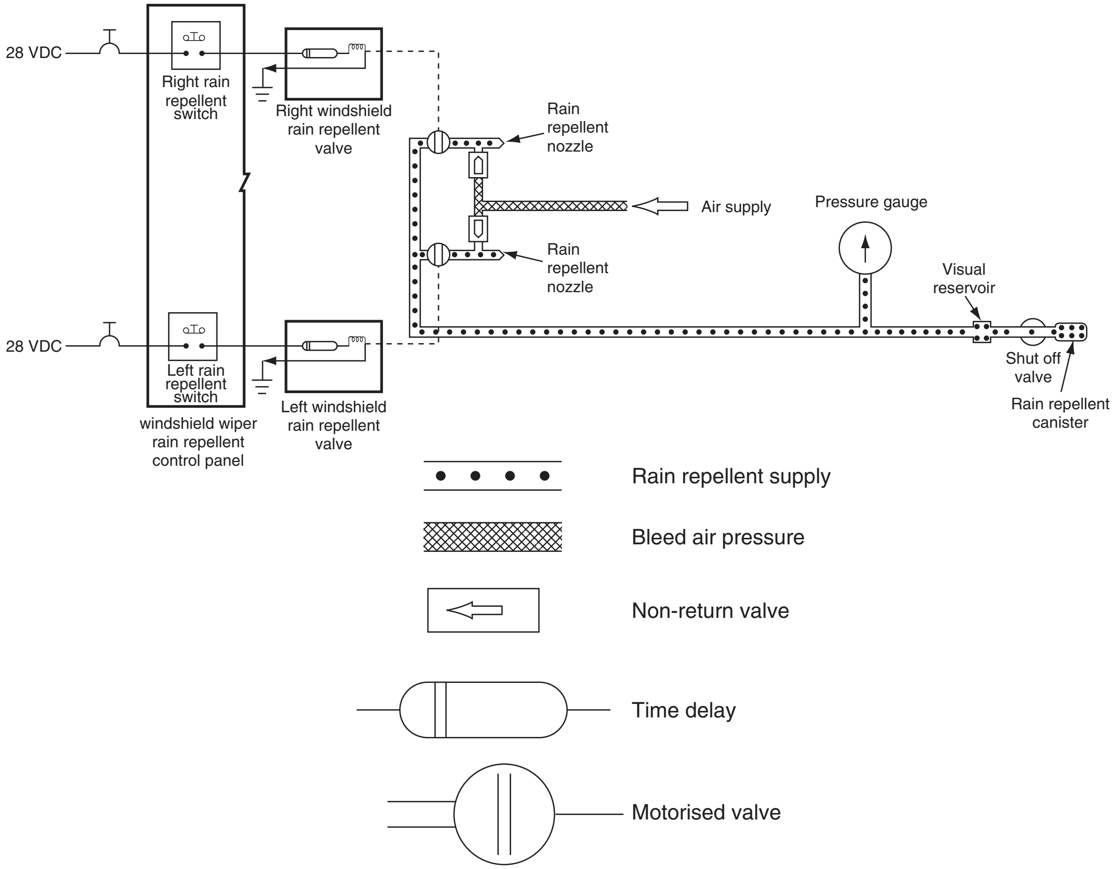
28 VDC
Right rain repellent switch
Right windshield rain repellent valve
Rain repellent nozzle
Air supply
Pressure gauge
Rain repellent nozzle
Visual reservoir
28 VDC
Left rain repellent switch
windshield wiper rain repellent control panel
Left windshield rain repellent valve
Shut off valve
Rain repellent canister
Rain repellent supply
Bleed air pressure
Non-return valve
Time delay
Motorised valve

Figure 15.23 Rain repellent system schematic

The applied pressure is reduced before the wheels lock up, and then reapplied to continue the braking action; this occurs as a repeated on/off cycle. Modifying the applied pressure is achieved by modulation; this can vary the applied pressure in both time and pressure.

The system comprises speed transducers (or sensors) on each wheel, a control unit and electro-hydraulic control valves, see Fig. 15.24. During a skid condition, the wheel(s) experience a rapid deceleration. Typical speed sensors are based on the tachometer; this is a small AC generator with the stator formed with a permanent magnet. The rotor coil is built into the axle and turns inside the stator field. Referring to fundamental principles, when the coil is rotated inside the field, currents are induced and a voltage is generated; the tachometer's output is proportional to wheel speed. The tachometer AC output is rectified in the control unit; the DC output is monitored for rate of change to determine if the wheel is accelerating or decelerating. A comparator circuit generates an error signal which is amplified and used to operate a control valve to modulate the applied brake pressure. Control laws are used to determine when the wheel is approaching a skid condition, this occurs when the wheel speed is decreasing at a given rate. Hydraulic pressure is decreased in the braking system and the wheel accelerates. Once the skid has been averted, hydraulic pressure can be reapplied.

Test your understanding 15.2

Explain how wheel speed is monitored to determine when an aircraft is approaching a skid condition.

15.5 Configuration warning

The system (also known as a **take-off warning** system) provides a warning if the pilot attempts to take-off with

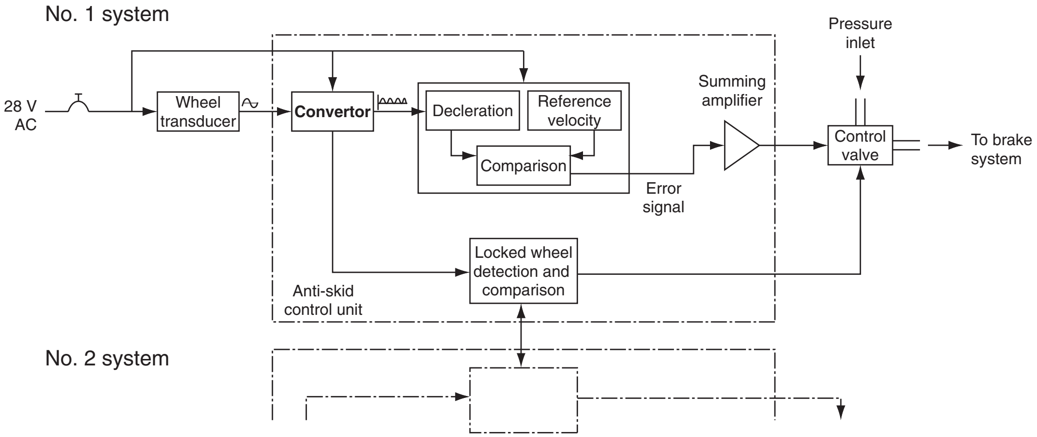
No. 1 system
28 V AC
Wheel transducer
Convertor
Decleration
Reference velocity
Comparison
Summing amplifier
Pressure inlet
Control valve
To brake system
Error signal
Locked wheel detection and comparison
Anti-skid control unit
No. 2 system

Figure 15.24 Anti-skid system

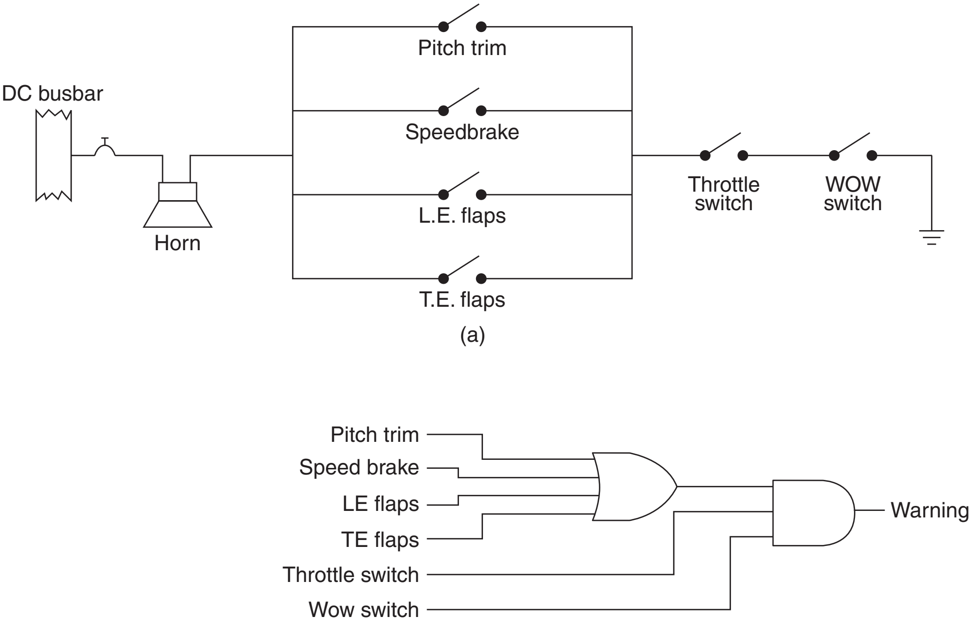
Pitch trim
DC busbar
Speedbrake
L.E. flaps
T.E. flaps
Throttle switch
WOW switch
Horn
(a)
Pitch trim
Speed brake
LE flaps
TE flaps
Throttle switch
Wow switch
Warning
(b)

Figure 15.25 Take-off warning system: (a) functional diagram, (b) logic diagram

specific controls not selected in the correct position, i.e. an unsafe configuration. A simple take-off warning system is illustrated in Fig. 15.25(a). The type of system fitted depends on the aircraft type and size; typical parameters being monitored include (but are not limited to):

- pitch trim
- speed brake
- leading edge (LE) flaps
- leading edge (LE) slats
- trailing edge (TE) flaps

The position of each of these controls is monitored together with the squat switch (weight on wheels) and throttle position. If an unsafe take-off configuration is detected, a warning horn is sounded. This system can be viewed in combination logic terms as illustrated in

Fig. 15.25(b). On larger aircraft, more parameters are monitored, and more logic functions are required for configuration warning (Fig. 15.26).

15.6 Aural warnings

Flight compartment aural warnings typically include the:

- fire bell
- take-off configuration warning
- cabin altitude warning
- landing gear configuration warning
- mach/airspeed overspeed
- stall warning
- TAWS
- TCAS.

Typical external aural warnings are also provided for a fire bell in the wheel well and ground call horn in the nose wheel-well for an equipment bay overheat. Certain warnings can be silenced whilst the condition exists. (TCAS is an automatic surveillance system that helps aircrews and ATC maintain safe separation of aircraft. It is covered in a related book in this series: *Aircraft Communications and Navigation Systems*.)

An aural warning system that combines many functions on has a **priority logic** based on alert levels to ensures that certain warnings, e.g. TAWS, are announced in a predetermined threat level. Examples of three priority levels are given in Table 15.1. Note that there could be 25 or more levels of aural warning system inputs.

Table 15.1 Aural warning system priority logic

Priority	*Description*
1	Windshear
2	Pull-up (sink rate)
3	Pull-up (terrain closure)

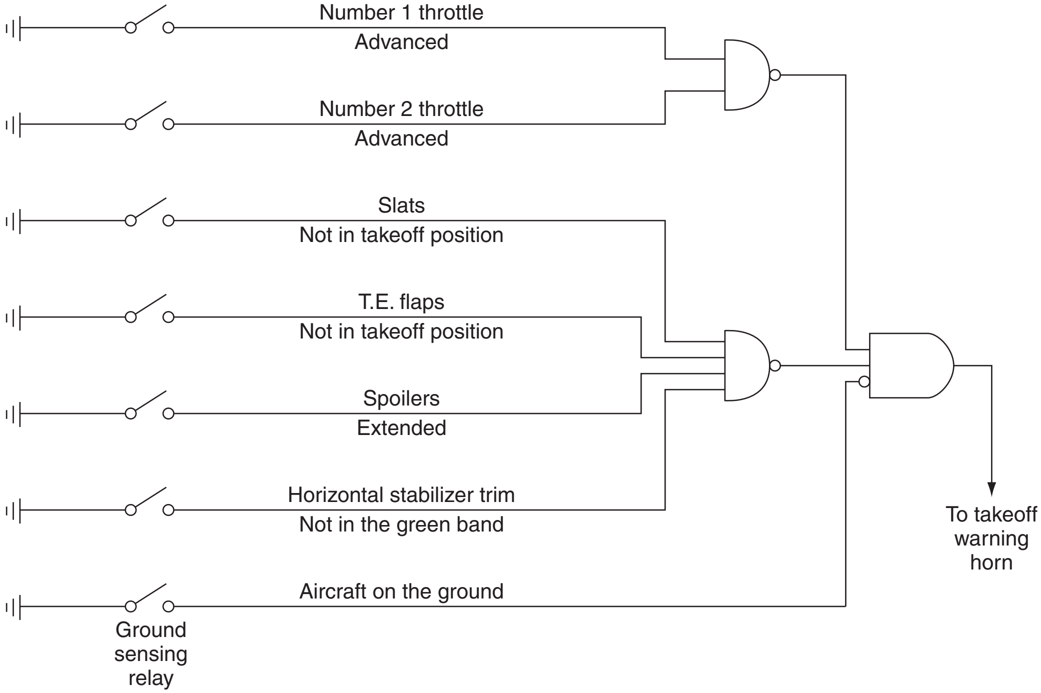
Number 1 throttle
Advanced
Number 2 throttle
Advanced
Slats
Not in takeoff position
T.E. flaps
Not in takeoff position
Spoilers
Extended
Horizontal stabilizer trim
Not in the green band
Aircraft on the ground
Ground sensing relay
To takeoff warning horn

Figure 15.26 Configuration warning system

15.7 Multiple choice questions

1. Stall warning systems provide the crew with a clear and distinctive warning:
   (a) before the stall is reached
   (b) after the stall is reached
   (c) at all angles of attack.

2. When an ultrasonic ice detector probe accumulates ice, it vibrates at a:
   (a) higher frequency
   (b) lower frequency
   (c) constant frequency.

3. The stall identification system contains an actuator that:
   (a) maintains the angle of attack
   (b) pulls the control column rearward
   (c) pushes the control column forward.

4. When not in use, pneumatic de-icing boots are:
   (a) inflated and kept flush with the airframe
   (b) deflated and kept flush with the airframe
   (c) cycled on a periodic basis.

5. When the AoA reaches a certain angle, the airflow over the wing:
   (a) becomes turbulent and the lift is dramatically decreased
   (b) becomes streamlined and the lift is dramatically decreased
   (c) becomes turbulent and the lift is dramatically increased.

6. When a wheel is approaching a skid condition, this is detected when the speed is:
   (a) decreasing at a given rate
   (b) increasing at a given rate
   (c) constant.

7. At normal angles of attack, the vane sensor is held:
   (a) back by spring pressure against the airflow
   (b) forward by the airflow against spring pressure
   (c) back by the airflow against spring pressure.

8. De-icing:
   (a) allows ice to form, and then be removed on a periodic basis
   (b) prevents ice from forming at any time
   (c) directs the repellent spray onto the windscreens.

9. An angle of attack (AoA) sensor vane aligns itself with the:
   (a) boundary layer
   (b) prevailing airstream
   (c) stagnation point.

10. When airflow passes over the wing without breaking up, it is said to have a:
    (a) boundary layer
    (b) streamline airflow
    (c) stalled airflow.

Chapter 16

Fire and overheat protection

Fire on board an aircraft is a very serious hazard; all precautions must be taken to minimize the risk of a fire starting. In the event that a fire does occur, there must be adequate fire protection on the aircraft. Fire protection includes the means of both detecting and extinguishing the fire. The subject of fire protection theory is a branch of engineering in its own right. Basic fire protection theory is covered in this chapter to provide the reader with sufficient information to understand how this theory is applied through aircraft electrical systems. The chapter focuses on the equipment and systems used to detect fire and smoke together with the means of delivering the fire-extinguishing agent.

16.1 Overview

The type of fire protection systems and equipment fitted to an aircraft can be sub-divided into specific areas of the aircraft:

- engines/APU
- cargo bay
- passenger cabin.

The fire protection technologies used on aircraft depend on these areas and specified fire risk. In addition to the fire risk in these areas, high temperatures resulting from engine **bleed air leaks** can also be hazardous. Bleed air is tapped from the compressor stage of a gas turbine engine, distributed in the pneumatic system for a number of purposes including thermal anti-icing. The consequences of hot air leaks are less severe than fire, but overheats can weaken structure and damage components. The pneumatic system takes air from the engines and distributes it throughout the aircraft; typical areas that are fitted with dedicated overheat detection systems are the wing leading edges, in the wheel wells and under the cabin floor.

Up until the 1990s, the two agents used extensively in the aircraft industry were:

- bromochlorodifluoromethane, also known by the trade name Halon 1211, or BCF
- bromotrifluoromethane, also known by the trade name Halon 1301, or BTM.

These are both inert gases and are either applied locally (typically in hand-held extinguishers) or as total flooding applications (typically in cargo bays). Both these agents are very effective at extinguishing fires, however they are in a group of **halogenated hydrocarbons** that have been shown to contribute to depletion of the earth's ozone layer. Although halon fire extinguishers are still being specified on aircraft, both gases come under the terms of the 1989 Montreal Protocol (and subsequent revisions), that prohibits the new production of these agents. The industry is supporting the supply of halon 1211 and 1301 through **recycling** of existing halon stocks for in-service and new production aircraft.

Alternative and replacement agents are being developed, and these are gradually being introduced throughout the fire protection industry. Examples of agents that do not deplete the ozone layer include:

- water
- dry powder
- vaporizing liquids (gaseous agents).

Combustion can be defined as a rapid and complex chemical reaction in which light and heat are evolved. All equipment in an aircraft is carefully designed and tested so that there is a low probability of starting and sustaining fire. Three factors are required to initiate and sustain combustion:

- fuel
- heat
- oxygen.

Fuels include solids, liquids and gases; each type of fuel requires a minimum temperature to be reached and maintained. Oxygen is normally available from the ambient air. The type of fire detector and extinguishing agent deployed is largely determined by which fuel is likely to combust. Detecting a fire

Table 16.1 Fire classifications and extinguishing strategies

Class of fire	*Fuel/heat source*	*Appropriate extinguishing strategies*
Class A	Solid/organic	Water, vaporizing gas
Class B	Flammable liquids	Carbon dioxide, dry powder, or vaporizing liquid
Class C	Flammable gasses	Dry powder, vaporizing liquids
Class D	Metals	Dry powder
Class E	Electrical equipment	Carbon dioxide, dry powder, or vaporizing liquid
Class F	Cooking fat/oil	Carbon dioxide, dry powder, or vaporizing liquid

can be achieved by one or more of the following strategies:

- thermal sensors
- optical sensors
- smoke detectors.

Extinguishing a fire can be achieved by one or more of the following strategies:

- limiting or eliminating the fuel
- limiting or eliminating the oxygen
- reducing the temperature of the fire
- interfering with the chemical reaction.

Fires can be categorized by the types of fuel that are combusting; this in turn determines the detection and extinguishing strategy. There are several classifications of fire used in the USA, Europe and Australasia. Typical fire classifications and extinguishing strategies are illustrated in Table16.1.

Primarily sources of risk/fuel in aircraft applications are:

- engines and APUs: class B
- cabin/flight compartment: class A
- cargo bay: class A.

Key point

Electrical fires are sometimes stated as Class E, or combined as Class A, B and D. The strategy adopted for this type of fire is to isolate electrical power and then deploy the agent.

Each of the fire extinguishing agents has different attributes. Water reduces the heat in a fire. It is often mixed with an additive, glycol, to enable it to be used in sub-zero temperatures. Vaporizing liquids reduce heat, but their significant contribution is to interfere with the chemical reaction of the fire. Carbon dioxide (CO_2) and dry powder both reduce the fuel's temperature and reduces the oxygen levels.

Key maintenance point

Always consult with the appropriate authority before deploying fire extinguishers.

16.2 Engine/APU fire detection

Engine/APU fire detectors are designed to very standards; they are needed when other systems, procedures and safety features have failed to prevent a hazardous situation developing.

In terms of fire protection, the engines and auxiliary power unit (APU) can be considered as having the same requirements. Fire detection within an aircraft engine can be achieved by thermal and/or optical methods. **Thermal** fire detectors sense heat energy that is radiated from the fire; they can be either unit detectors (sometimes referred to as point, or 'spot' detectors) or linear detectors. **Optical** fire detectors are based on sensing ultraviolet (UV) or infrared (IR) energy.

Whatever the sensor technology, the engine is fitted with sensors in positions that provide coverage of all areas where a fire could start; these positions are determined mainly from experience with an engine type. Ignition sources within the engine are electrical faults, and hot surfaces, e.g. exhaust ducts. Fuel and/or oil supply line failures in the engine will provide the means of sustaining the fire. There is a ready supply of oxygen in the ventilation air that is used to cool components.

When a fire is detected, the warning system activates an alarm bell and illuminates red warning lights in the flight compartment. The fire warning also illuminates a **master warning light** and another light (typically in the fire handle) to identify the affected engine(s). When the crew confirm that they have an engine fire, they operate the fire handle by first pulling and then twisting the handle. The action of pulling the fire handle activates micro-switches that shuts down the engine by:

- closing the fuel shutoff valves
- opening the field circuit of that engine's generator(s)
- closing the bleed air supply from the engine into the pneumatic system
- closing the hydraulic systems engine driven pump shutoff valve.

If the fire warning continues for a prescribed time after these actions, the fire handle is twisted; this closes micro-switches that activate the engine fire extinguisher.

16.2.1 Thermal fire detection

There are various techniques used to detect fire and overheat by including unit (or spot detectors), linear detectors and optical sensors.

16.2.1.1 Unit fire detectors

These are effectively thermostatic overheat switches based on the **bimetallic principle**. Figure 16.1 illustrates the external features and internal view and electrical schematic of a typical overheat switch. The two contacts are attached to spring bows; these are compressed during assembly and held apart by the outer barrel of the switch. As the detector is heated, it expands and the two contacts close, thus completing the warning circuit. An adjusting screw is found on the end-cap of some switches to allow for minor adjustments of operating temperature.

This type of detector can be used in engine fire detection systems; alternatively it can be used in the bleed air overheat detection system. The detector can only sense fire/overheat in a localized volume of the installation; it is more likely that multiple detectors are used to provide increased detection probability. A small gas turbine engine would typically be fitted with six of these detectors.

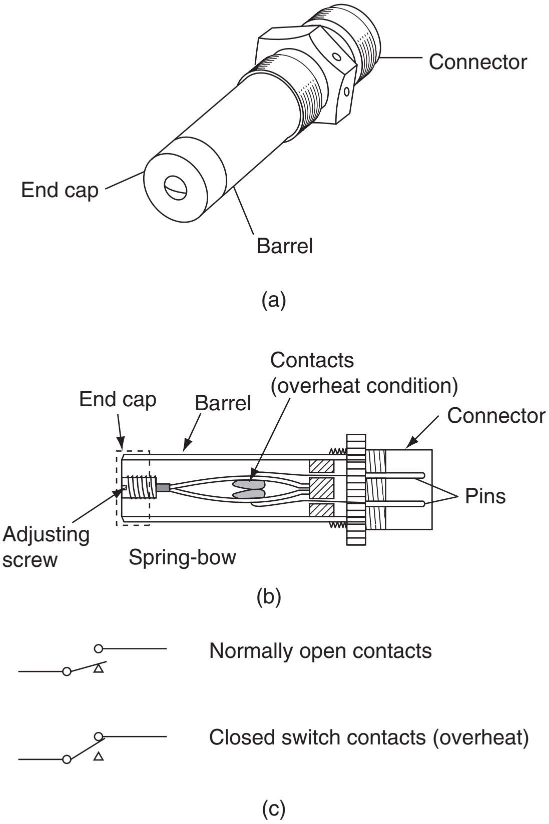

Figure 16.1 Overheat switch: (a) external features, (b) internal features, (c) electrical schematic

The wing leading of a large aircraft could be fitted with over 20 overheat switches. The switches are connected in parallel as shown in Fig. 16.2; this forms a **detection loop**. If any one of the switches closes due to a fire or overheat being detected, the alarm circuit is activated thereby illuminating a system warning light. A simple test circuit allows some of the circuit to be tested from the flight compartment. Some aircraft are fitted with a dual loop of detectors to provide a back-up system in the event of loop failure, e.g. open circuit wiring.

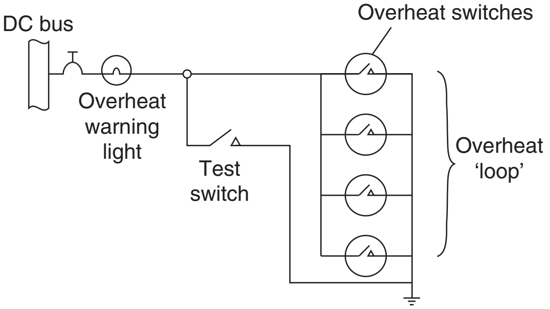

Figure 16.2 Overheat switches connected in parallel

Test your understanding 16.1

How would the crew know if more than one switch in a multiple detection system went into alarm?

Key maintenance point

The individual thermostat switch cannot be easily tested on the aircraft.

The actual alarm temperature of the unit fire detector is set by the amount of compression on the bow springs. During the aircraft design phase, the normal ambient temperature of the zone being protected is determined. The alarm temperature setting of the fire detector has to be carefully selected by the system designers. If it is too close to the nominal ambient temperature, there is a risk of false warnings. An operating temperature far higher than the zone's nominal ambient temperature will cause delays in the switch operating (if at all). Aircraft specifications and designs vary; however, a margin of approximately 50°C between ambient and operate temperature is typical. Detectors with different alarm temperatures are sometimes fitted in the same loop; this takes into account varying ambient temperatures in different zones. Typical settings are 230°C in the cooler forward sections of the engine nacelle and 370°C in the aft sections of the nacelle.

16.2.1.2 Linear fire detectors

Greater coverage of the engine nacelle can be achieved with linear fire detectors. Three types of

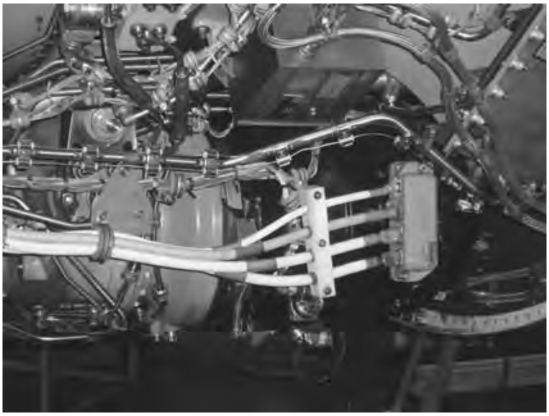

Figure 16.3 Linear sensor installation

linear fire detectors are used on aircraft; these are based on averaging or discrete detectors. The alarm temperature of **averaging** linear detectors depends on the length of element heated. **Discrete** linear detector systems are essentially independent of the length of element heated to provide an alarm; these eutectic salt. The response time for any fire detector is critical to the safety of the aircraft; the standard test for a linear fire detector is for it to respond to a 1100°C flame with five seconds. Actual response time will depend on many factors including the convection of heat energy; this is why it is so important to route the fire detector in the critical areas of the engine nacelle. Linear fire detectors are manufactured in lengths of up to three metres; some types can be joined in series to so that all these areas can be covered in convenient lengths. Installing the detector in manageable lengths facilitates installation and removal during maintenance.

Averaging linear detectors are based on either thermistor or pneumatic principles. Both detectors consist of a thin (typically 2 mm diameter) sensor that is routed around the engine. The sensor is attached to the engine using clips and bushes, see Fig. 16.3. The **thermistor** type fire detector is constructed in a coaxial form with a corrosion-resistant (typically stainless steel) outer sheath, nickel-chrome centre wire and temperature-sensitive separating, or insulating material, typically very fine granules of a silicon compound. Figure 16.4 illustrates the detector's physical features, a detector loop and the resistance temperature characteristics. This forms a complete electrical circuit, or **fire detector loop**.

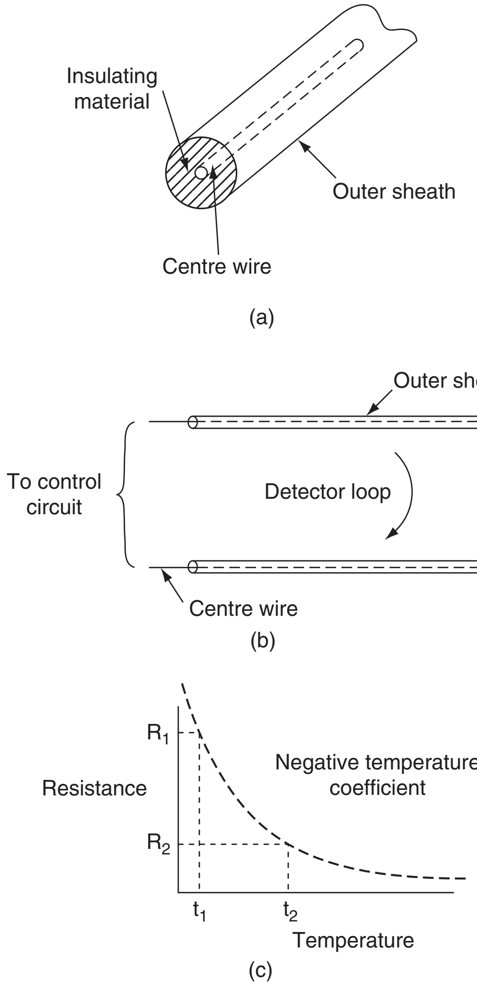
Insulating material
Outer sheath
Centre wire
(a)
Outer sheath
To control circuit
Detector loop
Centre wire
(b)
R_1
Resistance
Negative temperature coefficient
R_2
t_1
t_2
Temperature
(c)

Figure 16.4 Linear fire detector principles: (a) external features (b) detector loop, (c) resistance/temperature characteristics

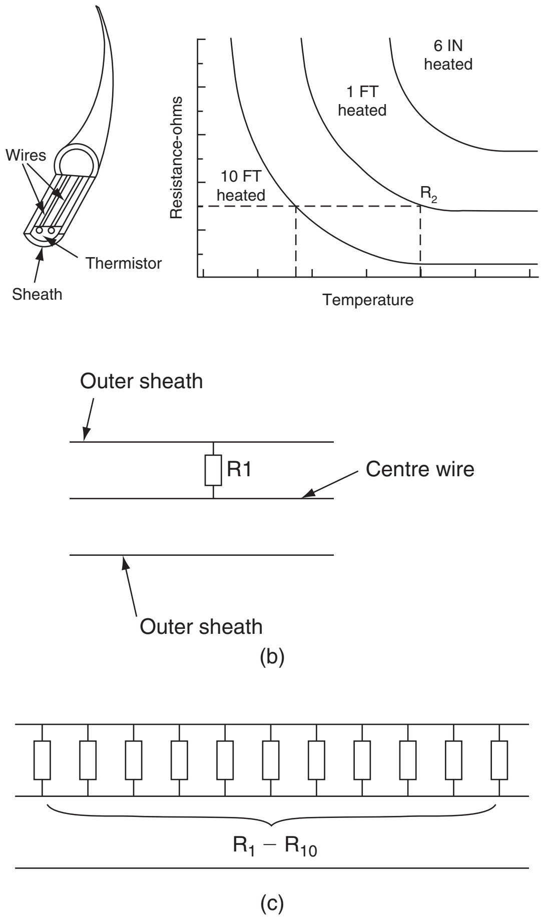
Wires
Thermistor
Sheath
Resistance-ohms
6 IN heated
1 FT heated
10 FT heated
R_2
Temperature
Outer sheath
R1
Centre wire
Outer sheath
(b)
$R_1 - R_{10}$
(c)

Figure 16.5 Thermistor characteristics: (a) effect of length being heated, (b) short length heated, (c) longer length heated

Referring to Fig. 16.4(c), when the sensor is at low ambient temperature (t_1), the resistance (R_1) between the outer sheath and centre wire is very high. The silicon insulating material has a negative temperature coefficient, i.e. when the sensor is heated, the resistance of the insulating material decreases.

When this temperature reaches the alarm point (t_2), the control unit senses a predetermined value, typically $1\,\text{k}\Omega$. A relay contact in the control unit closes and this activates the alarm circuit. When the fire has been extinguished, the sensor cools down, causing the resistance to increase; the relay opens and the alarm circuit is deactivated.

The thermistor fire detector is an **averaging** type, responsive to the length of element heated, see Fig. 16.5. It can be seen that if the alarm is tripped at an equivalent resistance indicated by R_2, a shorter length operates at a higher temperature. The thermistor fire detector can be thought of as an infinite number of resistors connecting the centre wire to the outer sheath.

If only one of these imaginary resistors is heated, i.e. a short length of element, R_1 has to reach a high temperature before its resistance drops to the alarm point. If a large number of these imaginary resistors is heated, i.e. a longer length of element, since the resistors R_1 through R_{10} are in parallel, the alarm point will be reached at a lower temperature.

There are certain fault conditions that can lead to false fire warnings; these can be caused by moisture ingress into the sensor and/or short circuit of the sensor wiring. Control units have been developed with the ability to discriminate between fire and fault conditions. Moisture ingress into the connectors is prevented by hermitically sealed terminal lugs.

Testing of the system is via a test switch located in the flight compartment. When the test circuit is activated, the following aspects are checked:

- continuity of centre element
- resistance of the insulating material
- alarm circuit.

Key maintenance point

It is very important that the sensing element and wiring does not come into contact with the engine structure.

Test your understanding 16.2

Why are averaging fire detectors dependent on the length of element heated to trigger an alarm?

Key maintenance point

Abrasion of the outer sheath will expose the insulating material leading to moisture ingress, decreased resistance and potential false warnings.

Key maintenance point

A short circuit on the sensor wires could be interpreted by the control unit as a fire condition.

Eutectic salt overheat detectors are constructed from a thin (typically 2 mm diameter) tube filled with a salt compound. They are particularly effective at detecting overheats when only a short length of element is heated. Referring to Fig. 16.6, the insulating material in this type of detector is a salt compound that melts at a specific temperature. All of the constituents crystallize simultaneously at this temperature and the insulation changes state into a molten liquid solution.

As with the thermistor system, a control unit monitors the resistance between the inner conductor and the outer sheath. When the salt melts, the resistance drops very rapidly and the control unit signals a warning. The resistance between the outer sheath and inner conductor decreases and the control unit senses this as a fire and/or overheat condition.

When the temperature is reduced, the salt compound solidifies and the resistance increases, thereby resetting the alarm output. This type of detector is used in both engine fire detection and bleed air overheat detection applications. The discrete detector is ideally suited for the detection of hot air leaks in:

- wing leading edges
- wheel wells
- under floor locations.

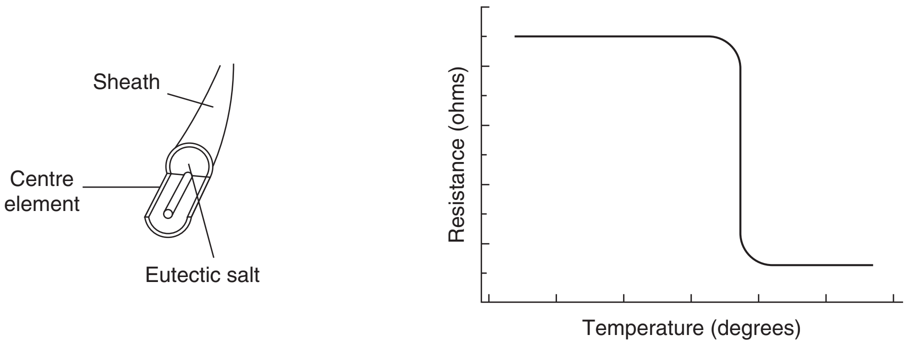

Figure 16.6 Eutectic salt overheat

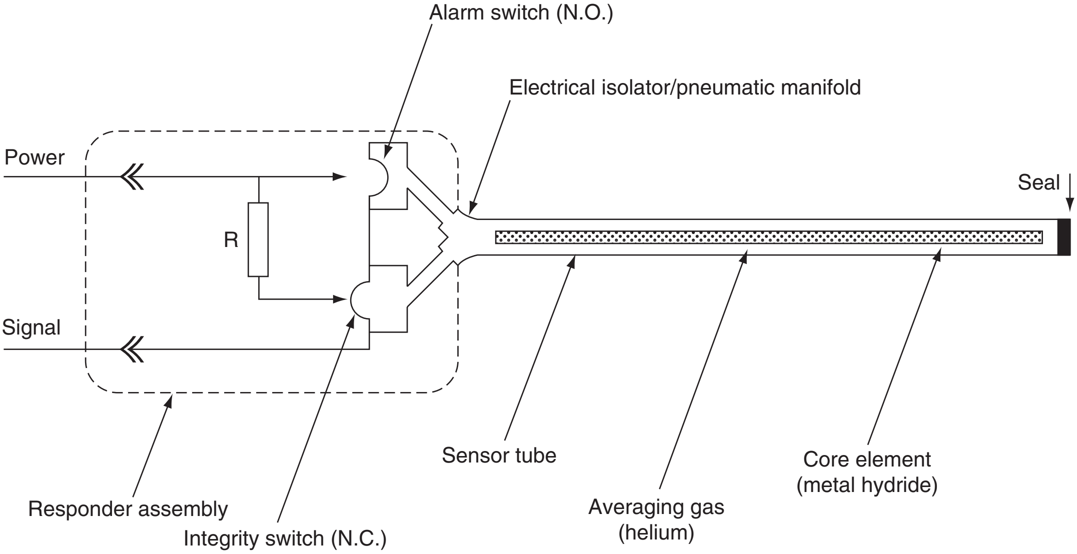

Figure 16.7 Pneumatic detector schematic

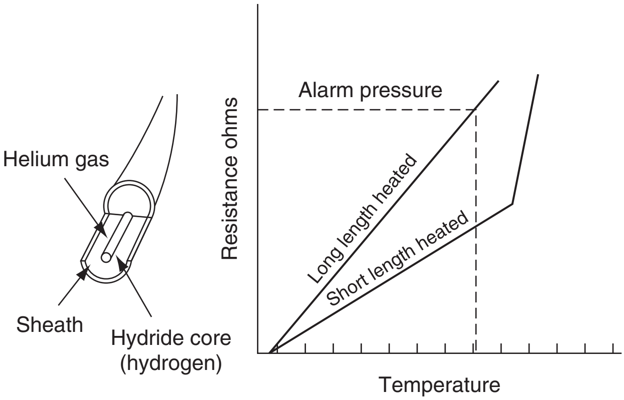

Figure 16.8 Pneumatic detector characteristics

Pneumatic fire detectors are constructed from a thin (typically 2 mm diameter) sensor tube filled with helium, and a gas-charged metal-hydride centre core. The tube is sealed at one end and attached to a pressure-sensing device (the **responder)** at the other end. The responder assembly is electrically isolated from the sensing tube and contains two pressure switches, a resistor and electrical connector, see Fig. 16.7.

The normally open (NO) **alarm switch** is closed when a fire or overheat causes the helium to expand. A normally closed (NC) **integrity switch** in the responder is held closed by the helium pressure; the internal resistor and switch contacts form the warning circuit. Measurement of the resistor value by an external circuit provides an indication of fire, overheat or loss of integrity. In the non-alarm condition, a monitoring circuit detects the known resistor value. When large sections of the sensor tube are heated, the increased pressure of the helium causes the alarm switch to close; this bypasses the resistor and the monitoring circuit signals an alarm condition. When the overheat condition is removed, the gas pressure decreases and the switch opens.

If only a small section of sensor tube is heated, there will be insufficient helium gas pressure to operate the alarm switch. To accommodate this localized heating, the centre core is formed with a **metallic hydride compound** which liberates hydrogen when rapidly heated; this now provides sufficient gas pressure to operate the alarm switch. When the sensor cools, the hydrogen is re-absorbed back into the centre core, pressure reduces and the alarm switch opens. Should a leak develop in the sensor or responder, the integrity switch opens. The control unit detects an open circuit condition from the change of resistance and this is signalled as a fault condition. The characteristics of a pneumatic detector are shown in Fig. 16.8. The entire assembly is hermitically sealed; the responder can be located in the engine fire zone as illustrated in Fig. 16.9.

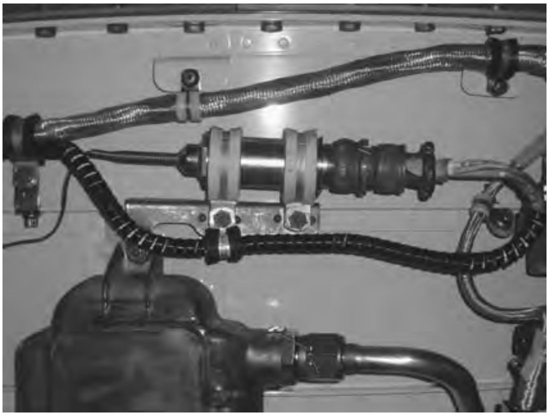

Figure 16.9 Responder located in the engine fire zone

Key point

Thermal fire and overheat detectors are often installed as **dual loops** to ensure that fire detection is available even if one system is inoperative.

Key point

Dual loop fire detector systems can also be configured to only go into alarm if both loops sense fire, thereby reducing the possibility of a false warning.

16.2.2 Optical fire detection

Optical sensors sense the light emitted from a flame in much the same way that a person recognizes fires. Our brains can distinguish between the light energy from a fire and the light energy from another source, e.g. a light bulb. The optical fire detector (or sensor) also has to be able to make this distinction. Some aircraft use optical fire detectors in place of thermal detector elements to simplify the installation. Depending on the size of engine, several optical sensors may be required. The sensors are easier to install and maintain than linear detector elements. Optical sensors offer the advantage of being able to monitor a specific volume of engine nacelle.

The spectral analysis of a hydrocarbon fuel fire reveals peaks of energy in the **infrared** (IR) and **ultraviolet** (UV) frequencies bands. Optical fire detectors are designed to detect radiation one or both of these frequency bands. The type of detection technology used depends on a number of factors including the:

- speed of response to a fire
- ambient temperature for where the sensor is located
- likely source of potential false alarms.

One of the characteristics of burning fuel in an aircraft engine is the emission of **infrared** (IR) radiation. A particular feature of a **hydrocarbon** fire is that it emits a high energy level at a nominal 4.4 micron band of the infrared radiation, referred to as the CO_2 spike, see Fig. 16.10(a). This is caused by the emission of energy from excited CO_2 molecules burning in the fuel.

Detecting a fire is one thing; being able to discriminate against other light sources is a vital part of the optical detector's performance. Extraneous light sources that are capable of causing a nuisance alarm emit very low levels of radiation in this range. These sources include:

- lightning
- arc welding
- x-rays
- sunlight
- hot surfaces.

An IR sensor designed for this specific frequency band provides a high level of reliable fire detection, while being relatively immune to nuisance alarm signals.

The detection cell incorporates a **pyro-electric cell** and an optical filter; the latter only transmits radiation within the wavelength band of 4.2 to 4.7 microns. This is packaged in a one-inch-diameter, three-inches-long cylindrical housing with an optical window at one end and electrical connector at the other. The **pyro-electric** detection cell responds to a fire by generating a signal when 4.4 micron radiation energy is detected. The optical fire detector has a **cone of vision** as illustrated in Fig. 16.10(b); 100% represents the maximum detection distance for a given fire. The sensitivity of the detector increases as the angle of incidence decreases.

As with any fire detector (whether thermal or optical) The **response time** of the detector depends on the:

- size of the fire
- rate of propagation
- type of fuel burning
- distance from the detector.

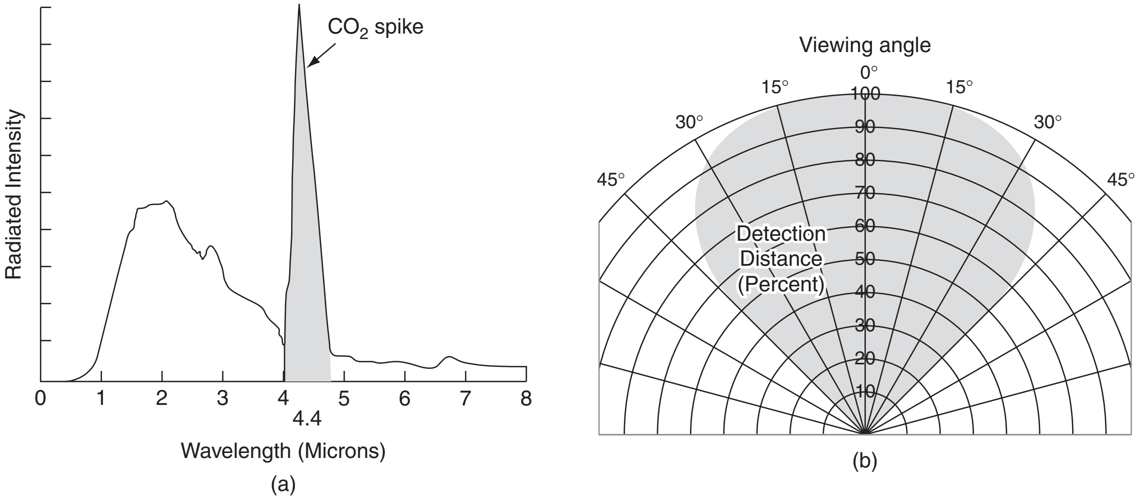

Figure 16.10 Optical fire detection: (a) hydrocarbon fire and 4.4 micron CO$_2$ 'spike', (b) optical fire detector viewing angle

The average **response** time of aircraft engine fire detectors is less than five seconds; this is in response to a standard 1100°C, six-inch-diameter flame.

A typical turbine engine fitted to a medium-sized general aviation (GA) aircraft is installed with three IR optical detectors in each engine nacelle, a control amplifier, warning lights and test switch. The detectors are positioned to receive direct and indirect radiation from flames, thus providing optimum coverage of the nacelle.

Key point

Increased temperature as a result of a fire is not a factor in the operation of an optical sensor.

Key point

The IR cell's output is an electrical signal proportional to the amount of IR energy falling onto it.

16.3 Cargo bay/baggage area

The most likely source of fire in the aircraft cargo bay or baggage compartment is from organic materials; the detection method in widespread use is the **smoke detector**. The primary design consideration for this application is ensuring that the smoke reaches the detector in a timely way. **Open area** smoke detectors rely on the transfer of particulate matter from the source of fire to the detector by **convection**. Smoke from burning materials in cargo/baggage areas often stratifies; in the early stages of a cargo/baggage fire there is insufficient heat energy to cause the smoke to rise through convection. Smoke will eventually accumulate in the roof area of the cargo/baggage area and so this is where the detectors are located. In larger cargo/baggage areas several detectors will be installed to speed up the detection time, which can be up to 60 seconds. In some aircraft, air samples are taken from numerous locations and supplied to a centralized **ducted** smoke detector.

16.3.1 Smoke detector principles

There are two basic principles used in smoke detectors: ionization and photoelectric.

16.3.1.1 Ionization smoke detectors

The **ionization** type monitors ionized products of combustion as they pass through a charged electrical field. This type of smoke detector can detect particles of smoke that are often too small to be visible. The

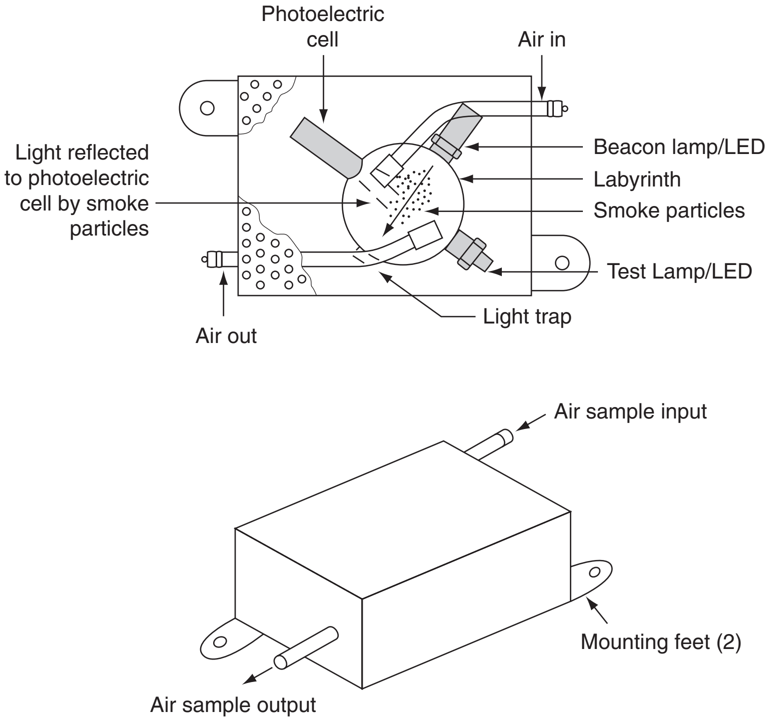

Figure 16.11 Optical smoke detector principles

detector uses a very small amount of **radioactive isotope** that is a source of **alpha radiation**. The radiation passes through an air-filled chamber between two electrodes; this radiation allows a low constant current to flow between the electrodes. Smoke entering the chamber absorbs the alpha particles, thereby reducing the ionization and interrupting the current; this interruption is used to trigger a warning circuit. Ionization smoke detectors can be manufactured at a lower cost compared with optical detectors; however they have several disadvantages:

- they will also go into an alarm if dust and/or fine water droplets enter the chamber
- the detector's sensitivity changes with pressure, i.e. altitude changes and also with age.

16.3.1.2 Photoelectric smoke detectors

Photoelectric smoke detectors measure light attenuation and/or reflection within a chamber. An air sample is passed through the detector via ducts, or pipes, see Fig. 16.11. This air sample passes trough a labyrinth chamber. A photoelectric cell is located at right angles to the light beam; when smoke particles enter the chamber, some of the light is reflected and scattered off the particles, causing energy to fall on the second cell.

Early types of photoelectric smoke detectors used a low-voltage incandescent light bulb projecting light across the chamber with a photosensitive cell on the other side of the chamber. Incandescent lamps do not have a very long life; this reduces the overall reliability of the detector. Modern aircraft smoke detectors now use a light-emitting diode (LED) as the light source. The smoke detector can be tested by illuminating an internal LED that simulates the deflected light.

Some aircraft use **open area** detectors where the smoke samples pass directly into the chamber. Open area cargo bay smoke detectors are arranged through the aircraft as shown in Fig. 16.12.

Main deck freighter aircraft typically use ducted smoke detectors installed as shown in Fig. 16.13. This configuration has a pair of smoke detectors per zone connected via orifices connected via manifolds. A centralized vacuum manifold draws air through all of the smoke detectors. Alarms are given on a zoned

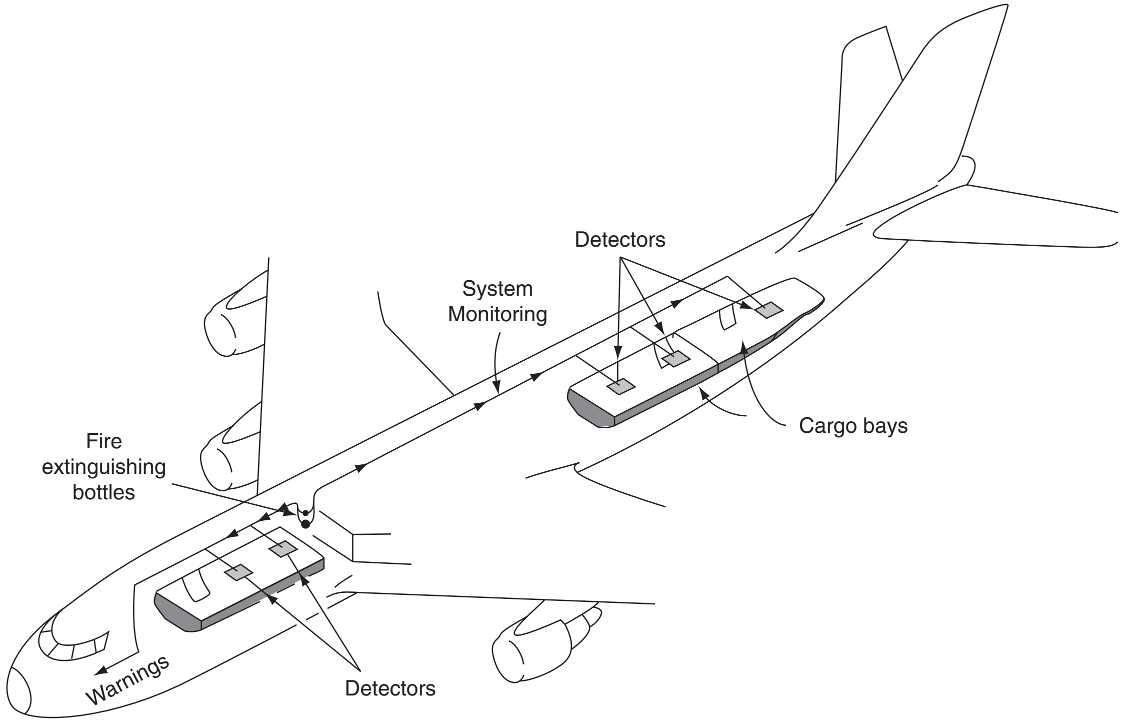

Figure 16.12 Cargo smoke detectors

basis when smoke enters through any of the sampling orifices.

A classification process approach is used to develop the fire protection strategy for cargo bay/baggage areas. The safety requirements are specified according to each class of compartments (A through E).

Class A cargo or baggage compartments are compartments where:

- the presence of a fire would be easily discovered by a crew member while at his/her station; and
- each part of the compartment is easily accessible in flight.

Class B cargo or baggage compartments are typically used on aircraft carrying a combination of cargo and passengers on the same deck. In these compartments:

- there is sufficient access during flight to enable crew members to reach any part of the compartment with the contents of a hand fire extinguisher
- when the access provisions are being used, no hazardous quantity of smoke, flames, or extinguishing agent will enter any compartment occupied by the crew or passengers
- there is a separate approved smoke detector or fire detector system to give warning at the pilot or flight engineer station.

Class C cargo or baggage compartments are typically compartments below the main deck of an aircraft, or on the same deck as passengers, but are not readily accessible. In these compartments:

- there is a separate approved smoke detector or fire detector system to give warning at the pilot or flight engineer station
- there is an approved built-in fire extinguishing or suppression system controllable from the cockpit
- there are means to exclude hazardous quantities of smoke, flames, or extinguishing agent from any compartment occupied by the crew or passengers
- there are means to control ventilation and drafts within the compartment so that the extinguishing agent used can control any fire that may start within the compartment.

Class D cargo or baggage compartments were typically compartments below the main deck of an aircraft, or on the same deck as passengers, not readily

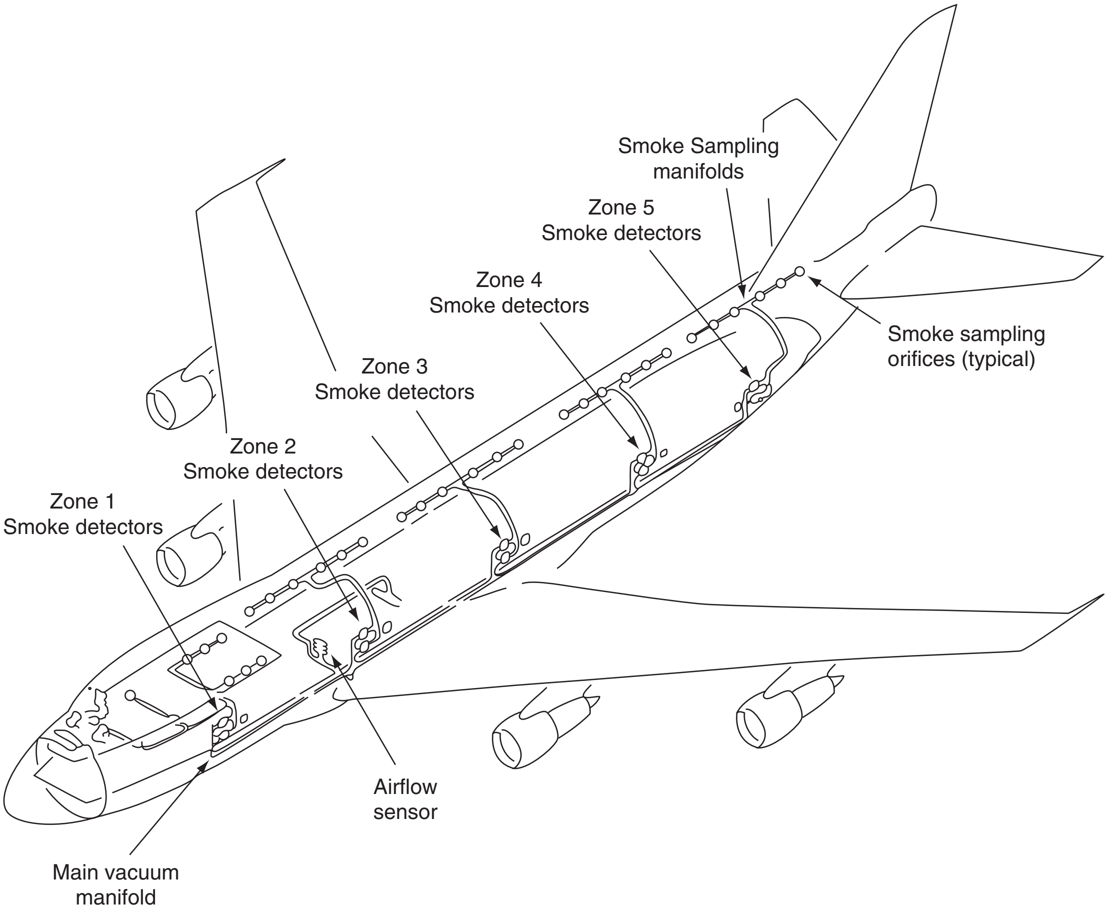
Smoke Sampling manifolds
Zone 5 Smoke detectors
Zone 4 Smoke detectors
Zone 3 Smoke detectors
Zone 2 Smoke detectors
Zone 1 Smoke detectors
Smoke sampling orifices (typical)
Airflow sensor
Main vacuum manifold

Figure 16.13 Main deck/freighter smoke detectors

accessible and were designed to use oxygen starvation to contain fire. This class of cargo or baggage compartments is being phased out due to incidents involving uncontained fires. They must now meet the same standards as Class C cargo or baggage compartments.

Class E cargo or baggage compartments are compartments on aircraft used only for the carriage of cargo and do not carry passengers. In these compartments:

- there is a separate approved smoke or fire detector system to give warning at the pilot or flight engineers station
- there are means to shut off the ventilating airflow to or within the compartment, and the controls for these means are accessible to the flight crew in the crew compartment
- there are means to exclude hazardous quantities of smoke, flames, or noxious gases from the flight crew compartment
- the required crew emergency exits are accessible under any cargo loading condition.

16.4 Fire extinguishing

16.4.1 Overview

Extinguishers are used in the engines, APUs, cargo bay, cabin and flight compartment. The type of fire extinguishing agent used on an aircraft, and the method of activating the extinguisher, depends on its application. The type of agent used depends primarily on the type of fire being addressed; other factors to be considered include toxicity and storage temperature. Extinguishers are constructed from a metal casing that contains the agent and a means of expelling the agent into the fire. Some agents, e.g. water and dry powder, are expelled from the container by high-pressure gas; some agents, e.g. carbon dioxide, are gaseous at room temperature, and can be stored as liquids under pressure. The container is fitted with a discharge head; this contains a valve that is opened to release the agent. The discharge valve can be opened manually, electrically or automatically by exposure to heat.

16.4.2 Cabin/flight compartment

Hand-held fire extinguishers are portable and operated by first removing a safety pin and then squeezing the handle. This instinctive squeezing action punctures a diaphragm in the discharge head and the agent is expelled under pressure through the aperture left by the ruptured diaphragm.

Waste bins located in the aircraft cabin washrooms are vulnerable from fire started by passengers disposing of cigarettes; this threat is largely eliminated by no-smoking flights; however, self-contained extinguishers are still used. The extinguisher is fitted with a pipe that is plugged by a fusible material; when the tip of the pipe is exposed to a predetermined temperature (typically 75°C), the plug melts and the extinguishing agent is released automatically into the fire.

16.4.3 Engine/APU and cargo bay

Engine/APU and cargo bay fire extinguishers are secured to the aircraft or engine structure outside of their respective fire zones; the agent is directed into the fire zone via a pipe. In some aircraft this pipe feeds into a manifold to distribute the agent. The containers (often referred to as **bottles**) are either cylindrical or spherical in shape; a typical engine fire 'bottle' installation is shown in Fig. 16.14.

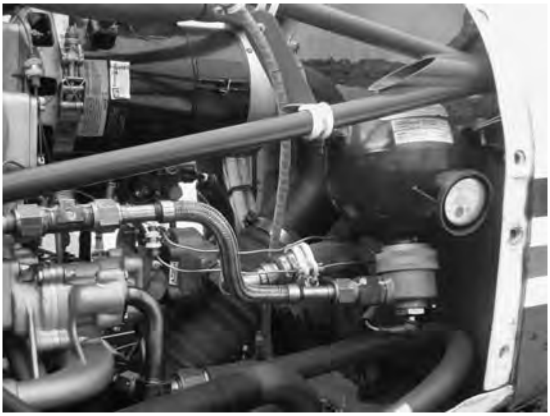

Figure 16.14 Engine fire extinguisher installation

This type of extinguisher is filled with a vaporizing liquid agent pressurized with dry nitrogen to 450 psi at 20°C. Figure 16.15 illustrates a two-outlet fire extinguisher. The discharge head contains a **cartridge** unit formed by small **pyrotechnic** device (sometimes called a **squib)** fitted with a heating coil.

When the fire switch is operated, current is passed through the heater coil. This causes the pyrotechnic charge to ignite, creating a rapid increase in pressure within the discharge head, thereby rupturing a **frangible diaphragm**. The agent is expelled under pressure through the aperture left by the ruptured diaphragm. When the agent has been expelled, a discharge plug moves to the end of the discharge head; this causes an indicator pin to become visible. On the aircraft fuselage, a polythene plug is forced into the position shown in Fig. 16.16; this provides additional confirmation that the extinguisher has been fired. If the fire extinguisher is exposed to high temperatures outside of its specified range, a pressure relief valve is opened and the contents discharged overboard.

The engine/APU fire extinguisher cartridge (or squib) is energized when the pilot operates the engine fire handle as previously described. Many aircraft are installed with extinguishers that have two or three outlets; a dedicated cartridge operates each outlet. These types of extinguishers are referred to as two-shot or three-shot bottles depending on the construction. The fire extinguisher outlets are connected to two separate engines; this provides the opportunity to discharge a second fire extinguisher into one engine in the event that the first discharge did not extinguish the fire, see Fig. 16.17. The so-called **two-shot fire extinguisher** is controlled from the fire handle as before; the first shot is discharged when the handle is turned in one direction, the second shot is discharged by twisting the handle in the opposite direction.

Key maintenance point

A discharge plug on the aircraft fuselage provides confirmation that the fire extinguisher has been fired.

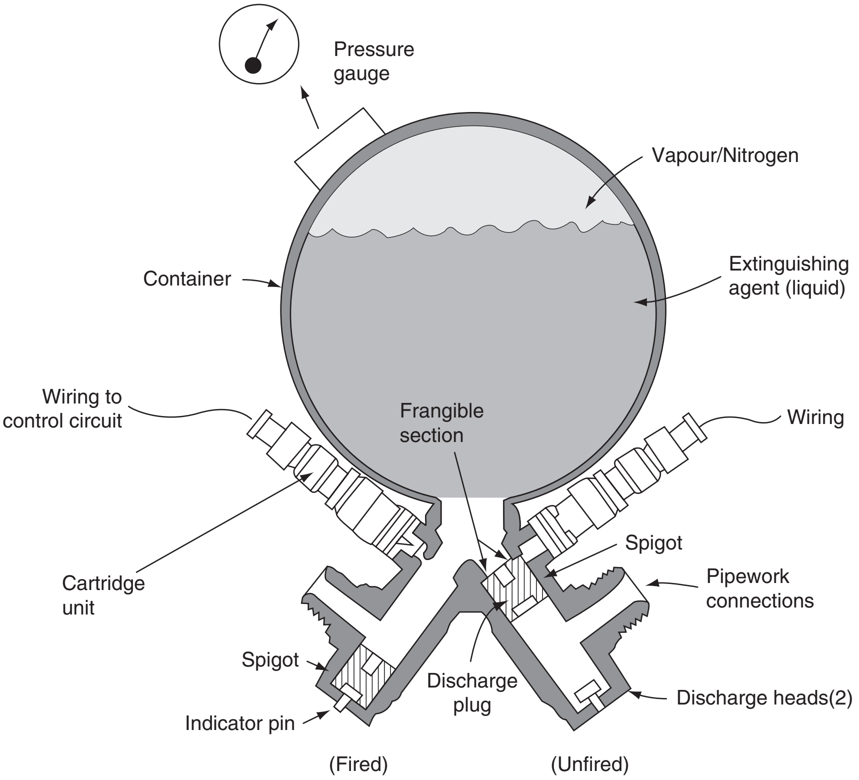
Pressure gauge
Vapour/Nitrogen
Container
Extinguishing agent (liquid)
Wiring to control circuit
Frangible section
Wiring
Spigot
Pipework connections
Cartridge unit
Spigot
Discharge plug
Discharge heads(2)
Indicator pin
(Fired)
(Unfired)

Figure 16.15 Engine fire extinguisher features

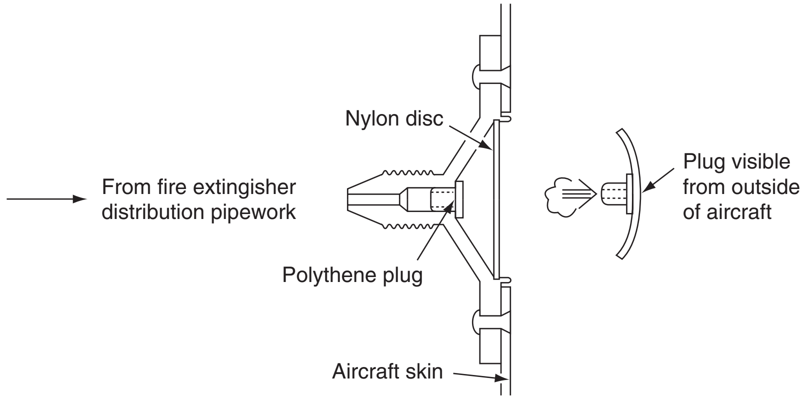
Nylon disc
From fire extingisher distribution pipework
Plug visible from outside of aircraft
Polythene plug
Aircraft skin

Figure 16.16 Fire extinguisher discharge plug

Cargo bay fire extinguishers are operated from a control panel selector. A typical cargo bay extinguishing system is illustrated in Fig. 16.18; this system features two fire extinguishers. The first bottle provides a high rate discharge to 'knock down' the fire; the second provides a low rate discharge to prevent re-ignition.

16.4.4 Fire extinguisher maintenance

Depending on the construction of the fire extinguisher, some of its contents could escape through the discharge head. Some extinguishers are **hermetically** sealed, some are constructed with gaskets. It is

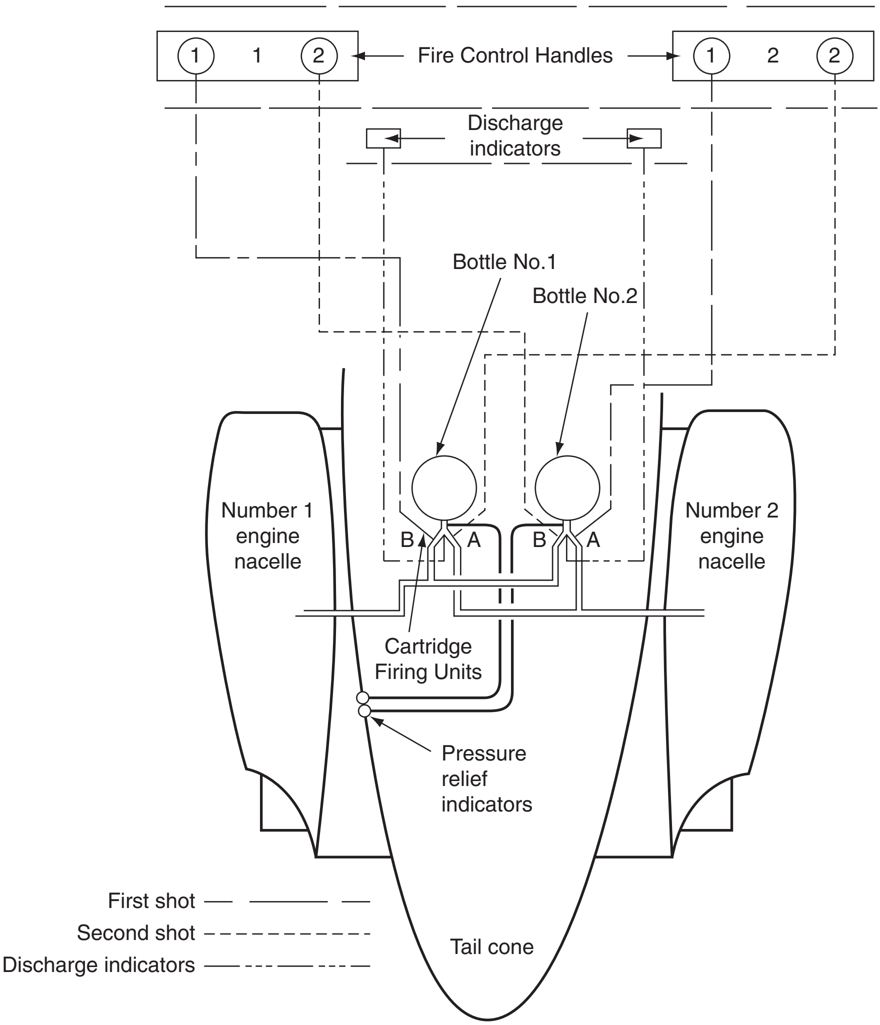
1
1
2
Fire Control Handles
1
2
2
Discharge indicators
Bottle No.1
Bottle No.2
Number 1 engine nacelle
B
A
B
A
Number 2 engine nacelle
Cartridge Firing Units
Pressure relief indicators
First shot
Second shot
Discharge indicators
Tail cone

Figure 16.17 Two-shot fire extinguisher system

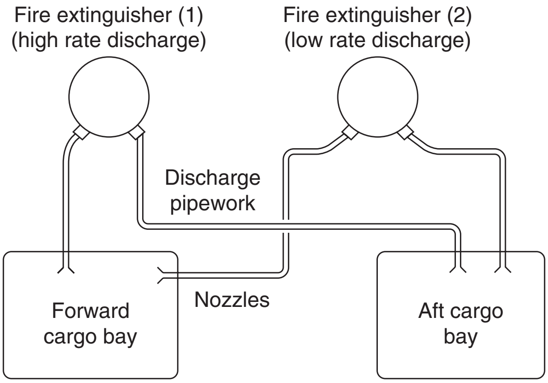
Fire extinguisher (1) (high rate discharge)
Fire extinguisher (2) (low rate discharge)
Discharge pipework
Nozzles
Forward cargo bay
Aft cargo bay

Figure 16.18 Cargo fire extinguisher system

essential that any loss of gas pressure and/or extinguishing agent be established. One way of ascertaining this is to weight the extinguisher on a periodic basis; the disadvantage of this method is the cost of removal/installation together with the logistic costs of transporting the item back to a workshop, together with the additional spares requirements. Most of these fire extinguishers are fitted with a means of indicating the **pressure** of the contents. This can either be a direct reading gauge, typically using a Bourdon tube, or a pressure switch that is linked via an indicating system to provide a warning to the crew that they have a extinguisher with low pressure; either from a leak or when the extinguisher has been discharged.

16.5 Multiple choice questions

1. The action of twisting a fire handle closes microswitches that:
   (a) activate the engine fire extinguisher
   (b) cancels the alarm
   (c) shuts off the fuel.

2. When eutectic salt melts, the resistance between the centre wire and outer sheath:
   (a) drops very rapidly and the control unit resets the warning
   (b) increases very rapidly and the control unit signals a warning
   (c) drops very rapidly and the control unit signals a warning.

3. Open area smoke detectors rely on the transfer of particulate matter from the source of fire to the detector by:
   (a) convection
   (b) radiation
   (c) conduction.

4. Dual-loop fire detector systems can be configured to only go into alarm if:
   (a) both loops sense fire, thereby reducing the probability of a false warning
   (b) either loop senses fire, thereby reducing the probability of a false warning
   (c) both loops sense fire, thereby increasing the probability of a false warning.

5. Multiple bimetallic overheat detectors are connected in parallel to provide:
   (a) variable alarm temperatures
   (b) increased detection probability
   (c) linear fire detection.

6. A discharge plug on the aircraft fuselage provides confirmation that the fire extinguisher:
   (a) is fully pressurized
   (b) squib is still intact
   (c) has been fired.

7. When a pneumatic fire detector is rapidly heated; hydrogen is liberated causing:
   (a) sufficient gas pressure to close the alarm switch
   (b) the integrity switch to open
   (c) the alarm switch to reset.

8. Photoelectric smoke detectors measure:
   (a) light attenuation and/or reflection within a chamber
   (b) radiated heat within a chamber
   (c) absorption of alpha particles within a chamber.

9. When a thermistor sensor is heated, the resistance of the insulating material:
   (a) decreases because of its negative temperature coefficient
   (b) decreases because of its positive temperature coefficient
   (c) increases because of its negative temperature coefficient.

10. To operate an engine fire extinguisher, the fire handle is:
    (a) twisted
    (b) twisted and then pulled
    (c) pulled and then twisted.

Chapter 17 Terrain awareness warning system (TAWS)

During the 1970s, studies were carried out by accident investigators and regulatory authorities to examine one of the most significant causes of aircraft accidents of the time: **controlled flight into terrain** (CFIT). This can be defined as an accident where a serviceable aircraft, under the control of a qualified pilot, inadvertently flies into terrain, an obstacle or water. CFIT accidents usually occur during poor visual conditions, often influenced by other factors, e.g. flight crew distraction, malfunctioning equipment or air traffic control (ATC) miscommunication. With CFIT, the pilots are generally unaware of this situation until it is too late. The outcome of these investigations was that many CFIT accidents could be avoided with a ground proximity warning system (GPWS). A system was developed in 1967 to alert pilots that their aircraft was in immediate danger of CFIT. This system was further developed into the enhanced ground proximity warning system (EGPWS) by adding a forward-looking terrain avoidance (FLTA) feature, made possible via global positioning system technology. This chapter describes the generic name given to this type of protection: terrain awareness warning system (TAWS).

17.1 System overview

The terrain awareness warning system comprises aircraft sensors, a computing function and warning outputs. The computer receives numerous inputs from on-board sensors and processes these to determine if a hazardous situation is developing. The TAWS computing function can either be from a self-contained unit; alternatively, it can integrated within another aircraft system.

Early ground proximity warning systems were based on aircraft sensors that detected:

- barometric altitude
- vertical speed
- radio altitude.

These early systems warned the crew of excessive descent rates or if they were flying towards rising terrain. TAWS can be considered as a **ground proximity warning system** (GPWS) combined with **forward-looking terrain avoidance** (FLTA), see Fig. 17.1. The combined systems provide a significant

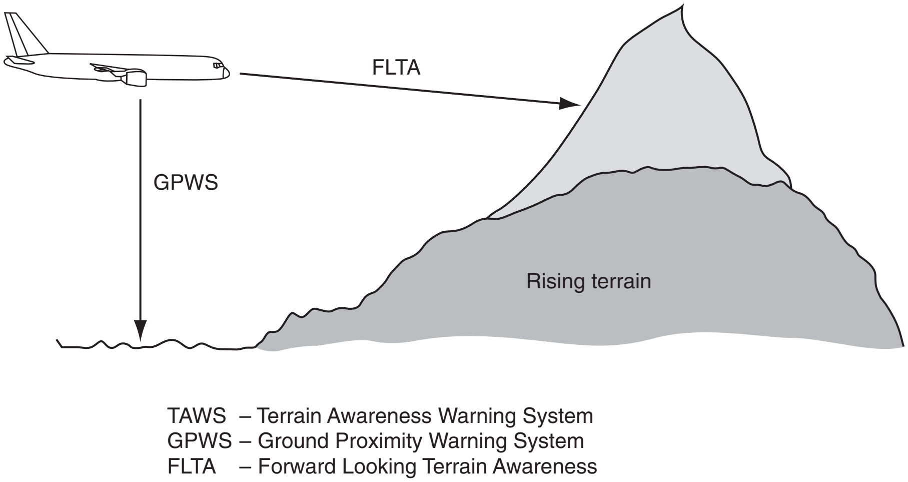
FLTA
GPWS
Rising terrain
TAWS – Terrain Awareness Warning System
GPWS – Ground Proximity Warning System
FLTA – Forward Looking Terrain Awareness

Figure 17.1 TAWS overview

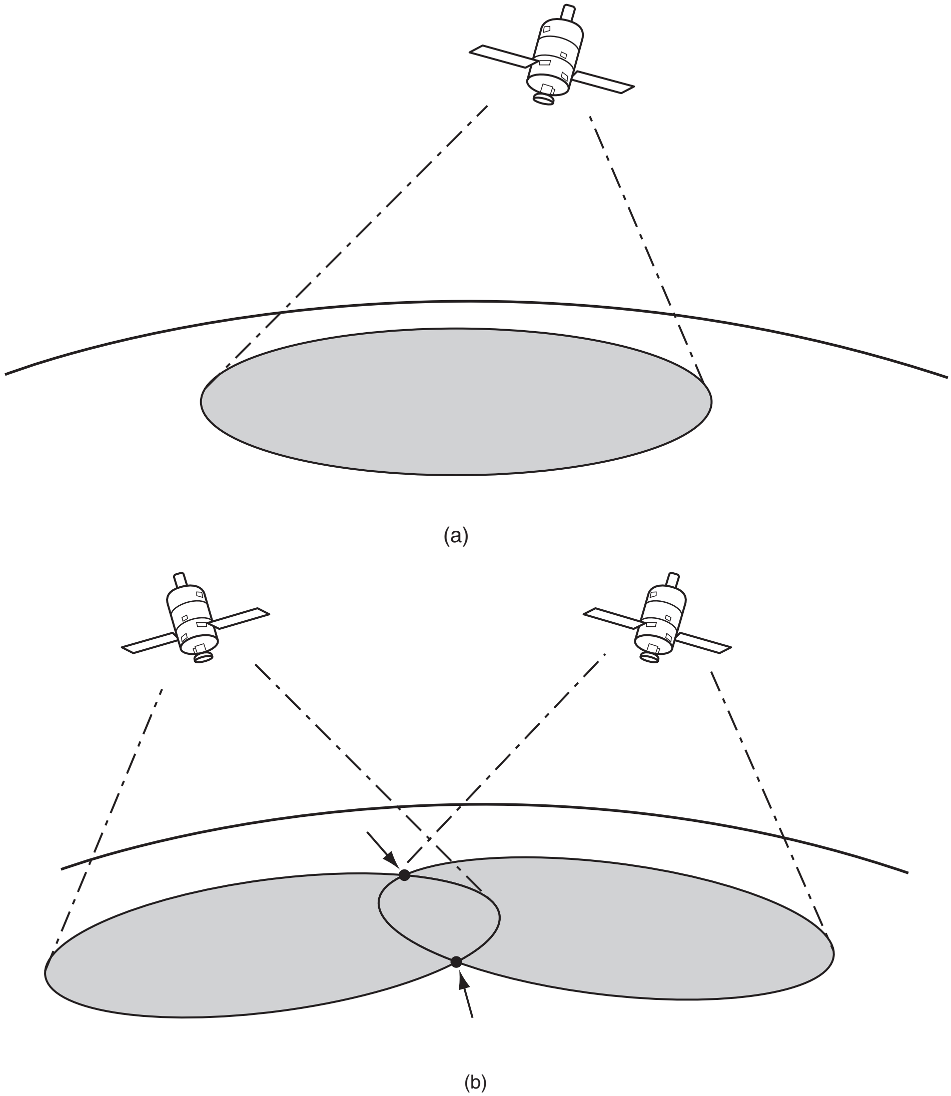

Figure 17.2 GPS principles: (a) Single satellite describes a circle on the earth's surface, (b) Two satellites define two unique positions

increase in safety margins by combining accurate position information (via the GPS) together with a terrain database.

GPS is an artificial constellation of navigation aids that was initiated in 1973 and referred to as Navstar (navigation system with timing and ranging). The global positioning system was developed for use by the US military; the first satellite was launched in 1978 and the full constellation was in place and operating by 1994. GPS is now widely available for use in many applications including aircraft navigation; the system calculates the aircraft position by triangulating the distances from a number of satellites, see Fig. 17.2.

Three satellites are required for the receiver to calculate a unique **lateral** position; a fourth satellite provides altitude, i.e. a **three-dimensional** position.

The availability of GPS technology thereby provides **forward-looking terrain avoidance** (FLTA)

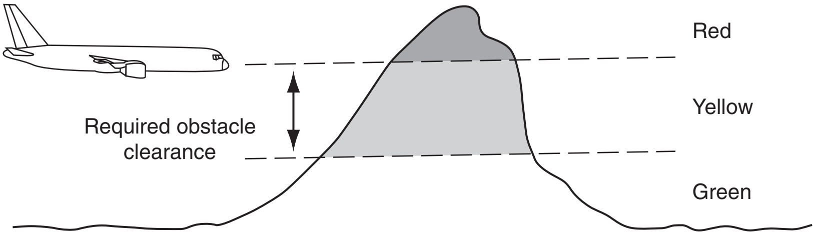

Figure 17.3 Obstacle clearance/colour coding

capability; it can look ahead of and below the aircraft flight path. This can be depicted on a display that is colour-coded (Fig. 17.3) to show the required obstacle clearance:

- **red** is used to indicate terrain above the aircraft's current altitude
- **yellow** is dependent on the flight phase, i.e. en route, terminal areas or approach
- **green** areas on the display are safe in terms of required terrain clearance.

The TAWS computer function creates a four-dimensional position comprising: latitude, longitude, altitude and time. It compares this position with the on-board database that contains details of terrain, obstacles and runways to determine any conflicts. (More information on GPS is provided in a related book in the series, *Aircraft Communications and Navigation Systems*.)

Key point

TAWS can be considered as a ground proximity warning system (GPWS) combined with forward-looking terrain avoidance (FLTA).

17.2 System warnings and protection

Cautions and warnings are provided to the crew depending on the level of CFIT threat; these include visual indications and audible announcements. TAWS equipment is designed to provide the flight crew with adequate time to take appropriate action. Equipment is classified either as Class A or B depending on the level of protection that the system provides. In general terms, passenger carrying aircraft require **TAWS-A**, smaller aircraft require **TAWS-B**; see Tables 17.1 and 17.2 for a summary of features.

Cautionary alerts are given by TAWS that require immediate crew awareness; subsequent corrective action will normally be necessary. Warning alerts are given for a terrain threat that requires immediate crew action. Aural alerts are synthesized voice messages that are produced through the audio panel into crew headsets and through the flight compartment speakers. Messages are presented on a suitable electronic display in text format, typically on a multi-function or other navigation display. Electronic navigation displays provide situational awareness, i.e. direction of flight and the location of nearby airports (given as four-letter codes). Local terrain features can be **overlaid** onto this presentation. Messages are either red or amber, depending on the threat level. Master caution and GPWS lights are illuminated depending on the threat level.

We have seen from the above that TAWS can be considered as a ground proximity warning system (GPWS) and forward-looking terrain avoidance (FLTA). Forward-looking terrain avoidance produces alerts from obstacles ahead of the aircraft using a terrain data base, GPS position and other reference data, e.g. aircraft speed etc. The ground proximity warning system produces alerts from the terrain below the aircraft and configuration inputs, e.g. gear and flap settings during approach and landing. FLTA features in Class A and B TAWS are identical and provide the same level of safety for the aircraft. The primary differences between Class A and B TAWS are in the GPWS features and terrain awareness display. Class A TAWS can be considered as being **autonomous**, i.e. all the aircraft reference inputs are from independent sources. Class B TAWS derives much of its references from within the TAWS computer (or TAWS section of

Table 17.1 TAWS: features, purpose and application (Class A or Class B)

Feature	*Purpose*	*A*	*B*
Forward-looking terrain avoidance (FLTA)	The system looks ahead of and below the aircraft's lateral and vertical flight path to provide a suitable alert if a potential threat exists. FLTA sub-functions are detailed below	✓	✓
FLTA reduced required terrain clearance alerts	Generated when the aircraft is currently above the terrain of the projected flight path of the aircraft and terrain clearance ahead of the aircraft is considered unsafe	✓	✓
FLTA imminent terrain impact alert	Generated when the aircraft is currently below the elevation of local terrain along the lateral flight path and when the projected vertical flight path is considered unsafe	✓	✓
FLTA high terrain impact alert	Generated when the terrain ahead of the aircraft is higher than the projected vertical path	✓	✓
FLTA flight path intent advisory alerts	Generated when the terrain ahead of the aircraft conflicts with the flight plan	✓	✓
Premature descent alert	The system compares the aircraft's current lateral and vertical position with the proximity of the nearest airport. Predicted flight path information and details from the database are used to determine if the aircraft is below the normal approach path for the nearest runway	✓	✓
Attention alerts	The system provides appropriate visual and audio alerts for both cautions and warnings	✓	✓
Terrain awareness display	The system provides terrain information to a suitable display system, e.g. a multi-function display (MFD)	✓	

Table 2 TAWS ground proximity warning Modes (Class A and Class B)

Mode	*Description*	*A*	*B*
1	Excessive descent rate	✓	✓
2	Excessive terrain closure rate	✓	
3	Negative climb rate or altitude loss after take-off	✓	✓
4	Flight into terrain when not in landing configuration	✓	
5	Excessive downward deviation from an ILS glide slope	✓	
6	Altitude callout at 500 feet	✓	✓
7	Wind shear	✓	

another computer). In summary, the following sensor inputs and displays are required by Class A TAWS, but not Class B:

- radar altitude
- air data
- gear position
- map display.

A terrain awareness display is only required for Class-A TAWS, although the display is found in many Class-B installations as an option. The graphical

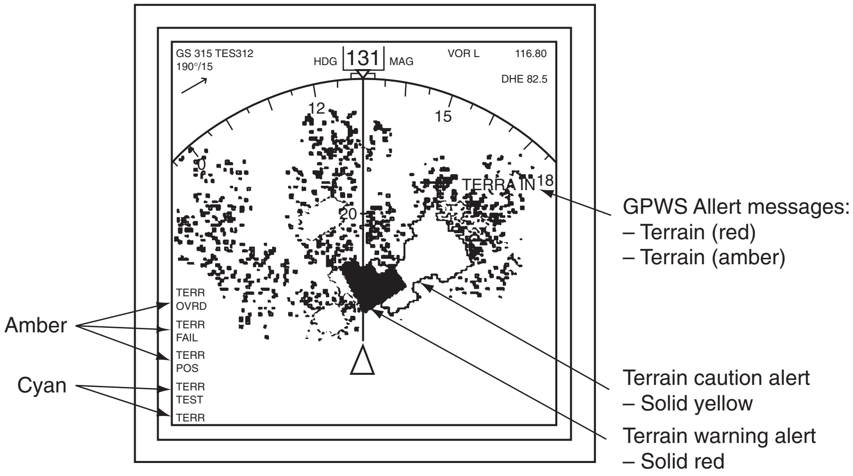
GS 315 TES312
190°/15
HDG 131 MAG
VOR L 116.80
DHE 82.5
12
15
0
18
20
TERRAIN
TERR OVRD
TERR FAIL
TERR POS
TERR TEST
TERR
Amber
Cyan
GPWS Allert messages:
– Terrain (red)
– Terrain (amber)
Terrain caution alert
– Solid yellow
Terrain warning alert
– Solid red

Figure 17.4 MFD/navigation display

display of terrain relative to the aircraft position and projected flight path is a major enhancement to **situational awareness**. Terrain can be displayed on general purpose multi-function displays (MFD) or on a navigation display (ND), see Fig. 17.4. Colour-coding and intensity are used to depict terrain relative to the aircraft.

Displays can be either a profile view or plan view. In addition to the terrain map, text messages are produced depending on system status and the terrain awareness threat level. When integrated with an electronic display, the surrounding terrain can be viewed relative to the aircraft position; this provides strategic terrain information up to 30 minutes in advance of a potential terrain conflict.

Test your understanding 17.1

Define the meaning of CFIT.

Key point

A terrain awareness display is only required for Class-A TAWS.

The two classes of TAWS are aimed at different types of aircraft; in broad terms Class A is intended for larger passenger aircraft, Class B is for general aviation aircraft. By deriving critical parameters such as synthetic radar altitude, the cost of Class B TAWS is substantially reduced in terms of both equipment and certification costs.

17.3 External references

The ground proximity Modes are derived from a number of signal inputs to the computer; each input is received from a variety of aircraft system outputs including:

- barometric air data
- radio altitude
- instrument landing system
- attitude and heading reference system.

The source of these inputs to the computer will vary depending on aircraft type.

Barometric air data includes altitude and vertical rate, or vertical speed. These two parameters are usually derived from an air data computer (ADC). This computer combines the functions of individual instruments, and can be used to determine data from the aircraft's pitot-static system:

- altitude
- vertical rate, or speed
- calibrated airspeed
- mach number.

Air data computers usually have an input of total air temperature to enable computation of static air temperature and true airspeed.

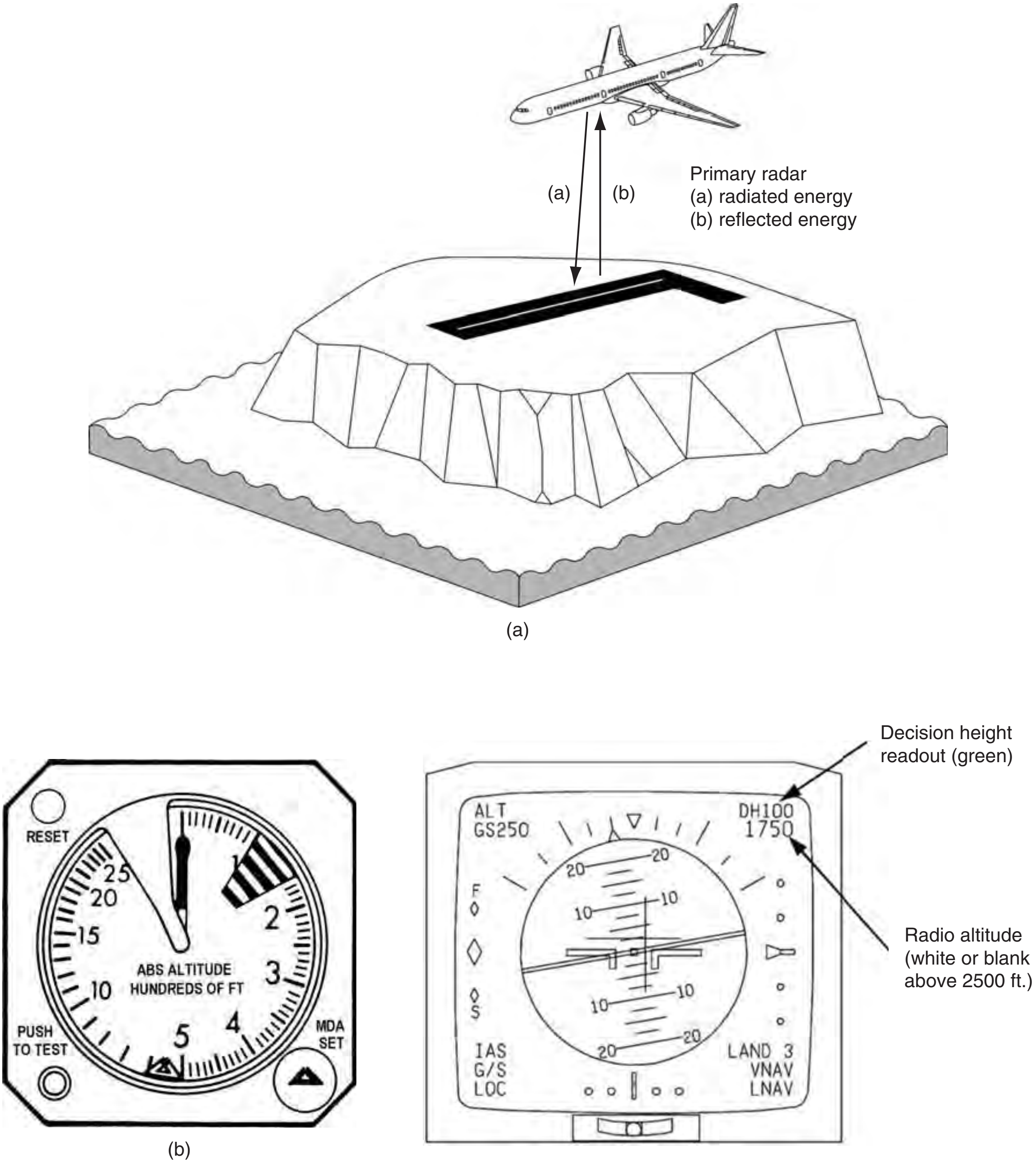
(a)
(b)
Primary radar
(a) radiated energy
(b) reflected energy
RESET
25
20
15
10
1
2
3
4
5
ABS ALTITUDE
HUNDREDS OF FT
PUSH
TO TEST
MDA
SET
Decision height
readout (green)
ALT
GS250
DH100
1750
20
10
F
S
IAS
G/S
LOC
LAND 3
VNAV
LNAV
Radio altitude
(white or blank
above 2500 ft.)

(a)

(b)

(c)

Figure 17.5 Low-range radio altimeter: (a) principle of operation, (b) analogue indication, (c) electronic indication

The **low-range radio altimeter** (LRRA) is a self-contained vertically directed primary radar system operating in the 4.2 to 4.4 GHz band. Airborne equipment comprises a combined transmitting and receiving antenna, LRRA transmitter/receiver and a flight deck indicator. Most aircraft are fitted with two independent systems. Radar energy is directed via a transmitting antenna to the ground (Fig. 17.5(a)); some of this energy is reflected back from the ground and is collected in the receiving antenna. This is displayed

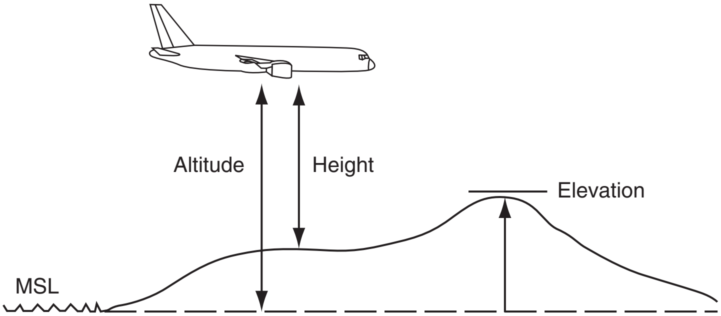
Altitude
Height
Elevation
MSL

Figure 17.6 Elevation, altitude and height

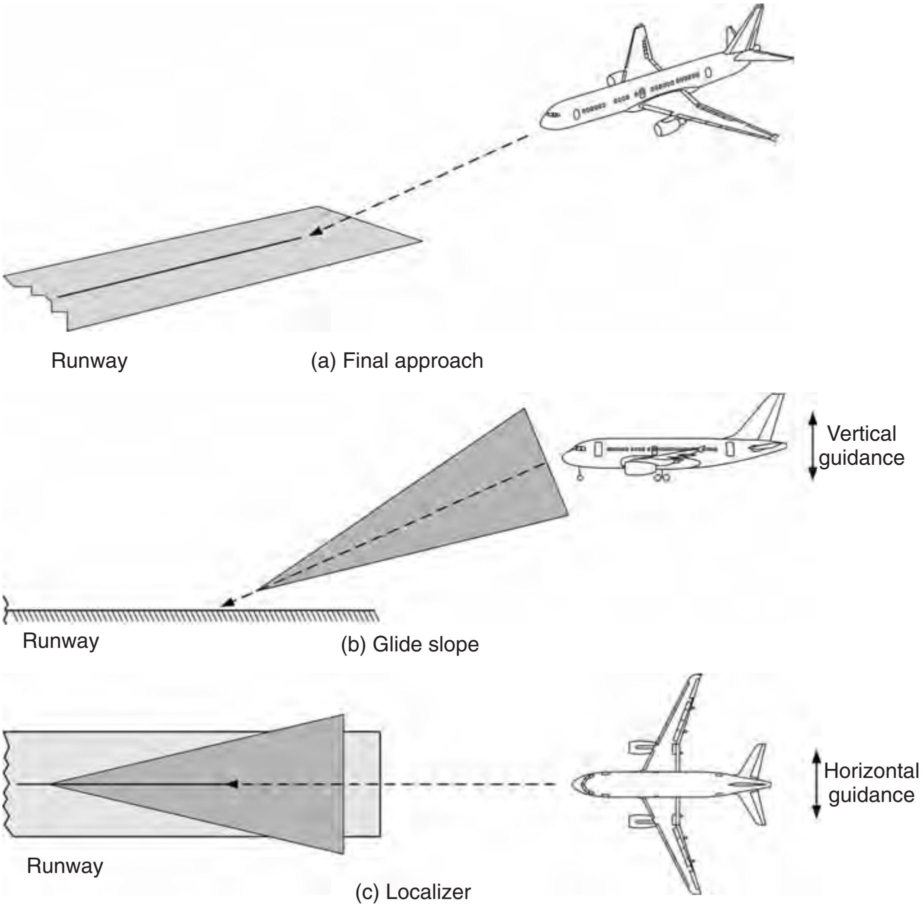
Runway
(a) Final approach
Vertical guidance
Runway
(b) Glide slope
Horizontal guidance
Runway
(c) Localizer

Figure 17.7 Instrument landing system

to the crew on an dedicated display (Fig. 17.5(b)) or an electronic instrument (Fig. 17.5(c)).

For aircraft operations, and in particular for TAWS, it is important to differentiate between elevation, altitude and height. Referring to Fig. 17.6, the basic reference point is mean sea level (MSL). Terrain that rises above MSL is measured as **elevation**. The aircraft's **altitude** is measured (normally in feet) above MSL, whereas the aircraft's **height** is measured above the terrain.

The **instrument landing system** is used during the final approach and is based on directional beams propagated from two transmitters at the airfield; see Fig. 17.7. One transmitter (the localizer) guides the

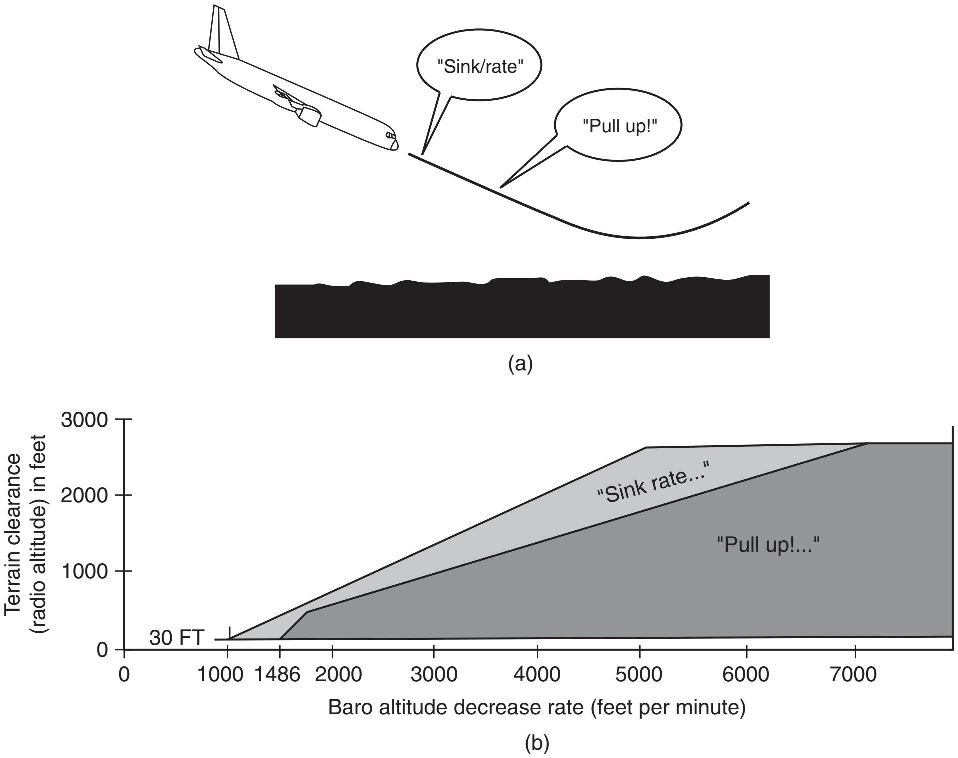

(a)

(b)

Figure 17.8 GPWS Mode 1: (a) excessive descent rate profile, (b) excessive descent rate envelope

aircraft in the horizontal plane. The second transmitter (the glide slope) provides guidance in the vertical plane and has a range of approximately 10 nm.

The **attitude** of an aircraft (pitch and roll) is sensed by gyroscopes. These sensors provide reference outputs that are processed to develop navigation and attitude data. Larger passenger aircraft derive this data from an inertial reference system (IRS). Developments in micro-electromechanical systems (MEMS) technology have led to silicon accelerometers that are more reliable and can be manufactured onto an integrated circuit. MEMS is the integration of mechanical elements, sensors and electronics on a common silicon substrate through micro-fabrication technology. This technology is being introduced onto general aviation aircraft for attitude and heading reference systems (AHRS). These systems are described in more detail in *Aircraft Navigation and Communication Systems*.

Key point

Warning alerts are given for a terrain threat that requires immediate crew action.

17.4 Ground proximity modes

17.4.1 Mode 1: excessive descent rate

Referring to Fig. 17.8, Mode 1 GPWS detects excessive sink (or descent) rate. The computer will provide a repeated aural alert of 'SINK RATE'. This is followed by a repeated aural warning of 'PULL UP' if the excessive descent rate continues, together with illumination of the **ground proximity warning light**. Amber and red text messages consistent with the aural messages are given on the navigation display.

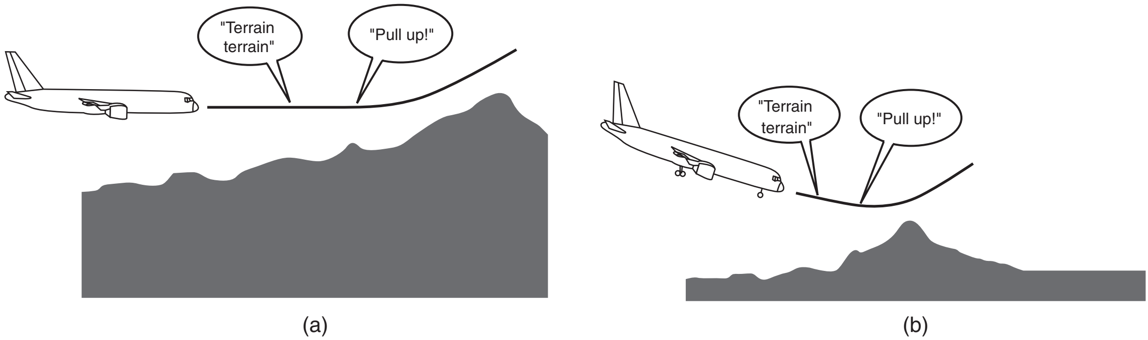

Figure 17.9 GPWS Mode 2 profiles: (a) excessive terrain closure rate, (b) excessive terrain closure rate, flaps and gear down

The computer derives excessive descent rate alert and warning from the following primary inputs:

- radio altitude
- vertical speed
- barometric altitude.

There are two envelopes within the TAWS software algorithms. The first envelope provides the 'SINK RATE' caution, if the aircraft flight path continues into the second envelope, the 'PULL UP' warning is produced. Mode 1 is inhibited below a minimum height, typically 30 feet; it is also independent of aircraft configuration, e.g. flaps and/or undercarriage positions.

17.4.2 Mode 2: excessive terrain closure rate

This mode provides protection in one of two ways (see Fig. 17.9): excessive terrain closure rate with flaps not in landing position (Mode 2A) and excessive terrain closure rate with flaps and gear in the landing position (Mode 2B).

Entering the Mode 2A envelope (see Fig. 17.10) will cause the repeated 'TERRAIN-TERRAIN' aural alert to sound. This is followed by the 'PULL UP' warning if the closure rate is too large. Airspeed schedules further refine the Mode 2A profile.

The height of the Mode 2B envelope floor varies dependent upon barometric descent rate and airspeed (Fig. 17.11). The cautionary voice message is accompanied with an amber text message consistent with the aural message. If the terrain closure rate penetrates the warning zone, an aural 'PULL UP' voice message is produced together with a red text message consistent with the aural message. Mode 2 is inhibited below a minimum terrain clearance, typically 30 feet.

17.4.3 Mode 3: negative climb rate or altitude loss after take off or go around

During an approach, in the landing configuration, Mode 3 arms when the aircraft descends below a specified altitude as illustrated in Fig. 17.12. The mode becomes active if either gear or flaps are retracted.

The cautionary alerts are 'DON'T SINK' or 'TOO LOW TERRAIN' (depending on what has been programmed into the computer during installation) together with an amber text message consistent with the aural message. Mode 3 does not have a warning zone; it is inhibited below a minimum terrain clearance and barometric altitude.

17.4.4 Mode 4: flight into terrain when not in landing configuration

This mode is programmed in one of three ways; see Fig. 17.13. Mode 4A is for unsafe terrain clearance with flaps in the landing position, but the gear is up; this will initially generate a 'TOO LOW TERRAIN'. This is followed by a 'TOO LOW GEAR' voice message with an amber text message consistent with the aural message.

Alternatively, if the flaps are up and the gear is down (Mode 4B), the voice message 'TOO LOW FLAPS' is produced. The third condition (Mode 4C)

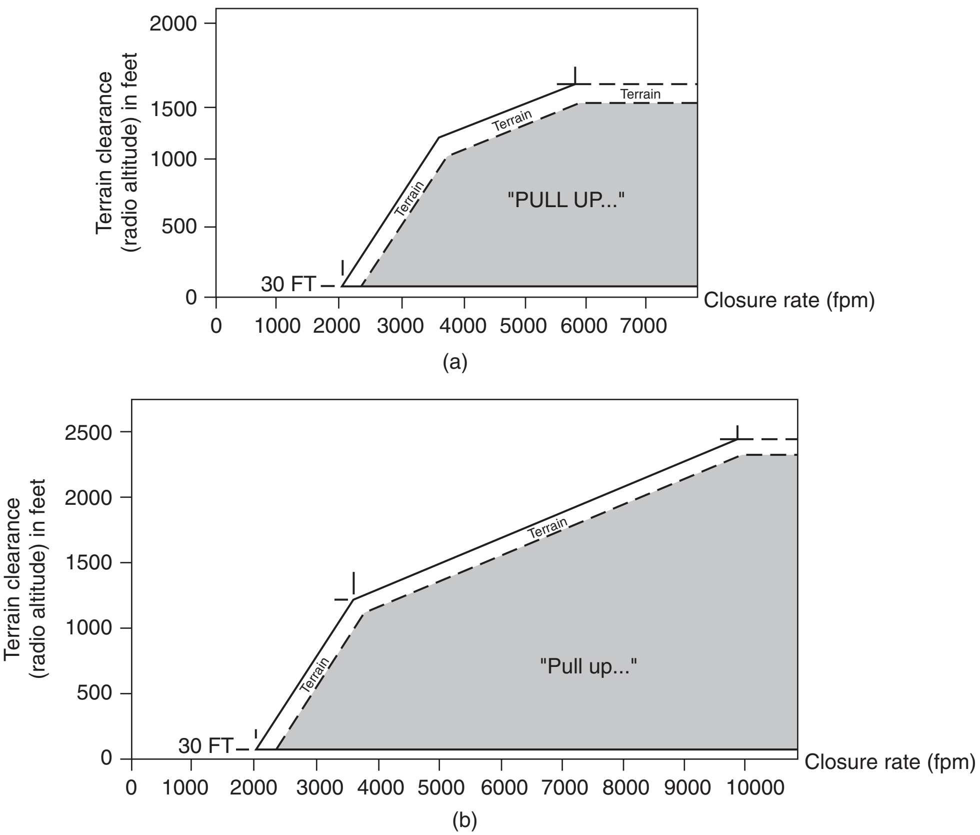

Figure 17.10 GPWS Mode 2A envelopes: (a) excessive terrain closure rate (airspeed < 220 knots), (b) excessive terrain closure rate (airspeed > 220 knots)

is caused by unsafe terrain clearance after take-off or go around; the will produce a 'TOO LOW TERRAIN' caution. The various envelopes associated with Mode 4 are illustrated in Fig. 17.14.

17.4.5 Mode 5: excessive downward deviation from an ILS glide slope

Mode 5 provides protection on the glide slope in both hard and soft alert areas (Fig. 17.15(a)). This mode becomes active when the number one ILS receiver has a valid signal, gear is down and radio altitude is between the limits shown in Fig. 17.15(b).

Descent below the glide slope (G/S) causes 'GLIDE SLOPE' voice message and an amber text message. Mode 5 can be inhibited by selecting the GPWS G/S push button switch when below 1000 ft RA. The mode automatically reactivates for a new envelope penetration.

17.4.6 Mode 6 Altitude callouts

When the aircraft descends to 500 feet above runway elevation, a voice message of 'Five hundred' is given; there are no text messages generated with this Mode. Within Mode 6, some manufactures offer predetermined altitude callouts, see Fig. 17.16.

This mode can also incorporate other optional features such as bank angle protection, see Fig. 17.17.

17.4.7 Mode 7 wind shear

Additional features on larger aircraft include a **wind shear** functionality within the TAWS computer. This is incorporated as **ground proximity Mode 7**. Wind shear can occur in both the vertical and horizontal directions; this is particularly hazardous to aircraft during take-off and landing. Specific weather conditions known a **microbursts** cause short-lived, rapid

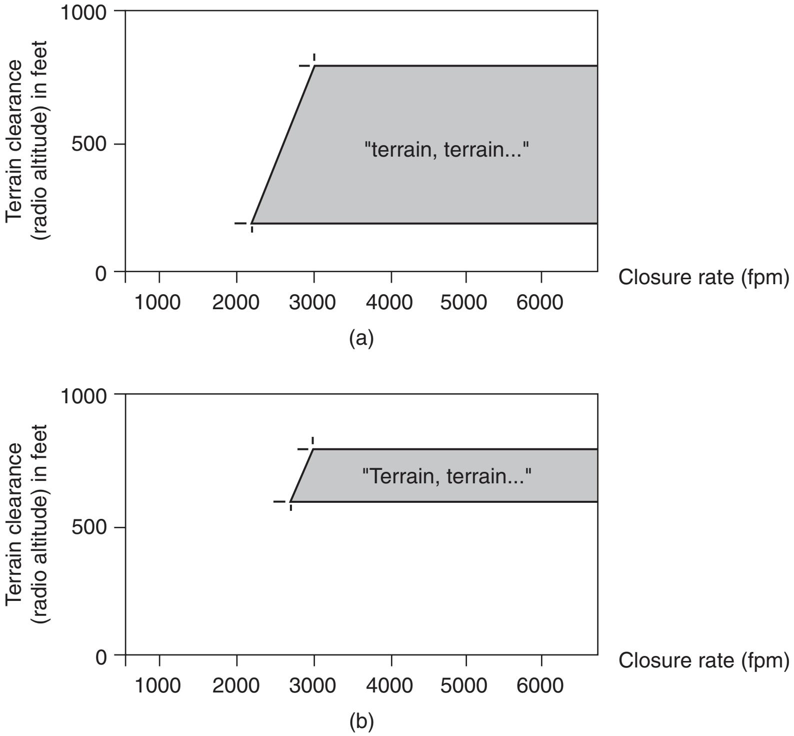

Figure 17.11 GPWS Mode 2B envelopes: (a) excessive terrain closure rate – flaps down (barometric rate < 400 f.p.m.), (b) excessive terrain closure rate envelope – flaps down (barometric rate > 1000 f.p.m.)

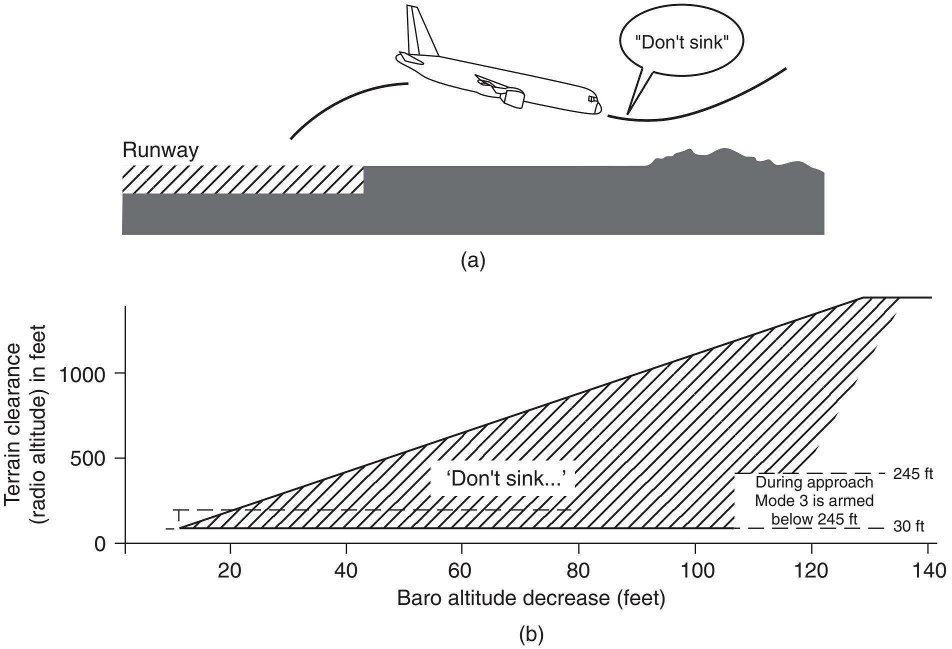
"Don't sink"
Runway
(a)
Terrain clearance (radio altitude) in feet
1000
500
0
'Don't sink...'
245 ft
During approach
Mode 3 is armed
below 245 ft
30 ft
20
40
60
80
100
120
140
Baro altitude decrease (feet)
(b)

Figure 17.12 GPWS Mode 3: (a) altitude loss after take-off profile, (b) altitude loss after take-off envelope

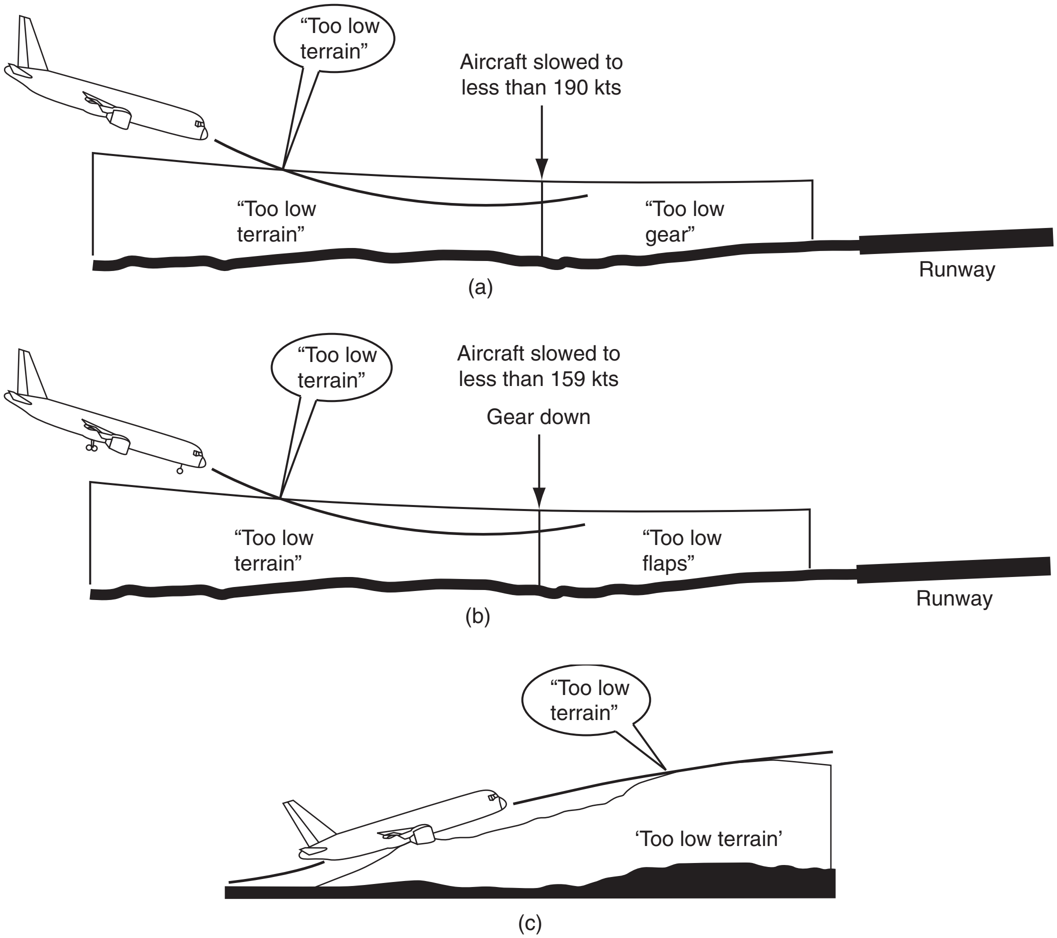
"Too low terrain"
Aircraft slowed to less than 190 kts
"Too low terrain"
"Too low gear"
Runway
(a)
"Too low terrain"
Aircraft slowed to less than 159 kts
Gear down
"Too low terrain"
"Too low flaps"
Runway
(b)
"Too low terrain"
'Too low terrain'
(c)

Figure 17.13 GPWS Mode 4 profiles: (a) flight into terrain when not in landing configuration – gear up, (b) flight into terrain when not in landing configuration – gear down/flaps up, (c) flight into terrain when not in landing configuration (take-off or go-around)

air movements from clouds towards the ground. When the air from the microburst reaches the ground it spreads in all directions, this has an effect on the aircraft depending on its relative position to the microburst.

17.4.7.1 Microburst

Referring to Fig. 17.18, when approaching the microburst, it creates an **increase** in headwind causing a temporary increase of airspeed and lift; if the pilot were unaware of the condition creating the increased airspeed; the normal reaction would be to reduce power.

When flying through the microburst, the headwind reduces and the aircraft is subjected to a **downdraught**. As the aircraft exits the microburst, the downdraught now becomes a tail wind, thereby **reducing** airspeed and lift. This complete sequence of events happens very quickly, and could lead to a sudden loss of airspeed and altitude. In the take-off and climb-out phase of flight, an aircraft is flying just above stall speed; wind shear is a severe threat. During approach and landing, engine thrust will be low; if a microburst is encountered, the crew will have to react very quickly to recognize and compensate for these conditions.

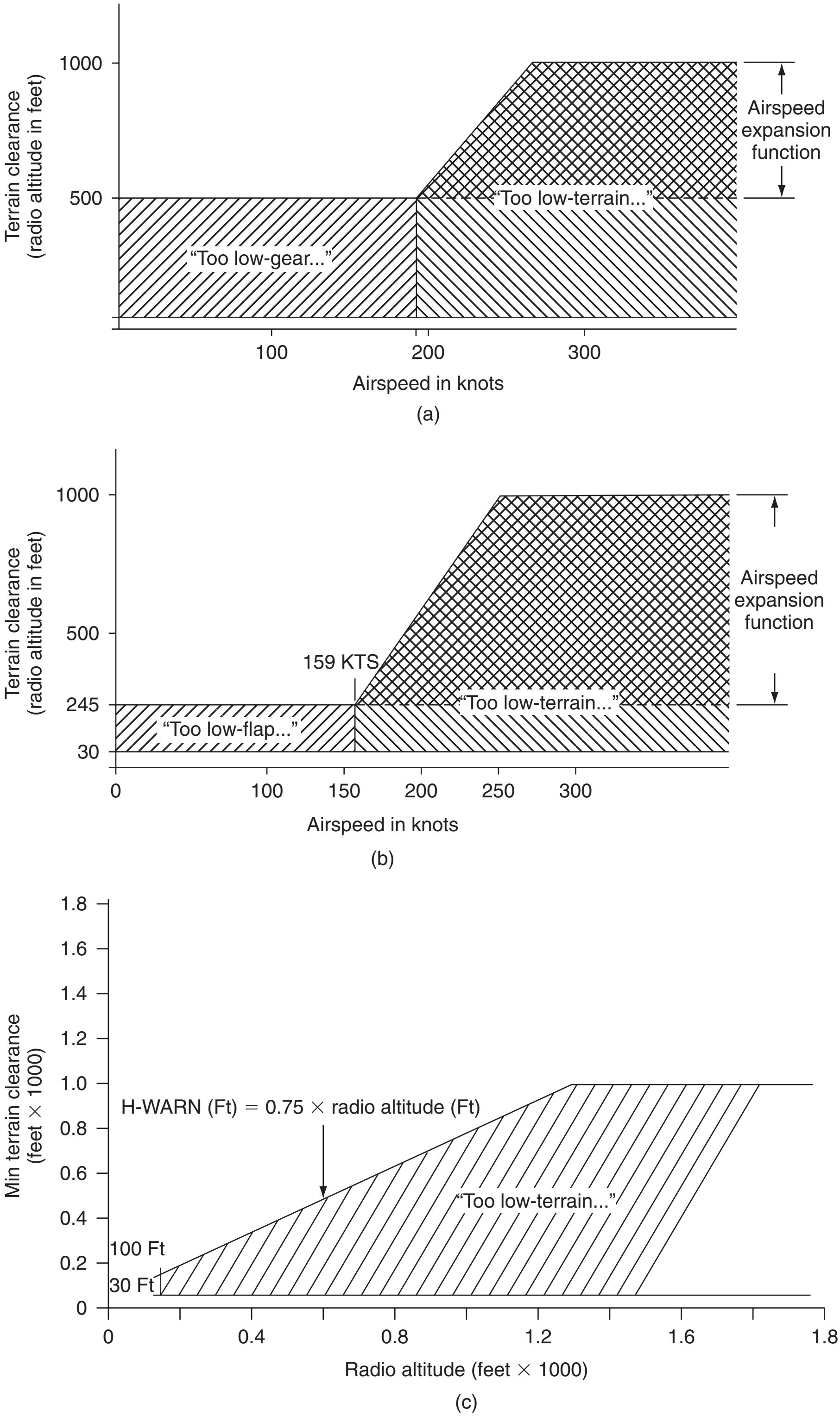

Figure 17.14 GPWS Mode 4 envelopes: (a) flight into terrain when not in landing configuration – gear up, (b) flight into terrain when not in landing configuration – gear down/flaps up, (c) flight into terrain when not in landing configuration (take-off or go-around)

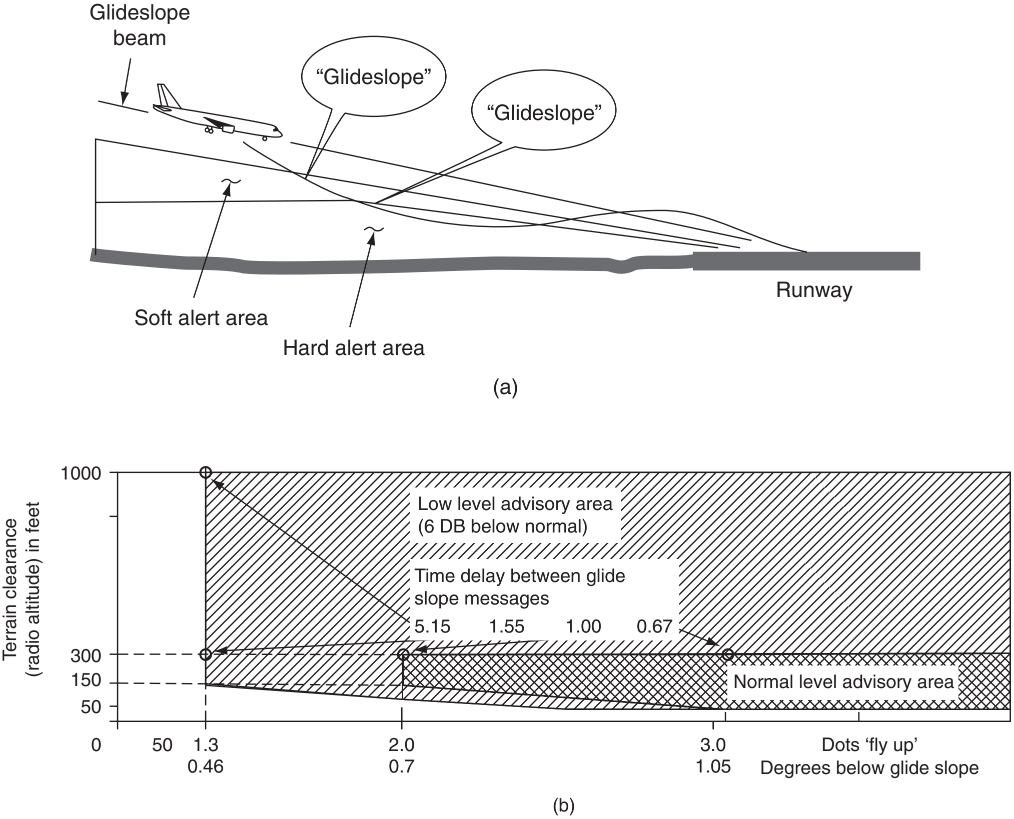
Glideslope beam
"Glideslope"
"Glideslope"
Runway
Soft alert area
Hard alert area
(a)
1000
300
150
50
0
Terrain clearance (radio altitude) in feet
Low level advisory area (6 DB below normal)
Time delay between glide slope messages
5.15 1.55 1.00 0.67
Normal level advisory area
50
1.3
0.46
2.0
0.7
3.0
1.05
Dots 'fly up'
Degrees below glide slope
(b)

Figure 17.15 GPWS Mode 5: (a) excessive downward deviation from an ILS glide slope profile, (b) excessive downward deviation from an ILS glide slope envelope

17.4.7.2 Doppler radar

Modern weather radar systems are able to detect the horizontal movement of droplets using **Doppler shift** techniques. Doppler is usually associated with self-contained navigation systems. (This subject is described further in *Aircraft Communications and Navigation Systems.*) The Doppler effect can be summarized here as: 'the frequency of a wave apparently changes as its source moves closer to, or farther away from an observer'. This feature allows wind shear created by microbursts to be detected. Referring to Fig. 17.19, the microwave energy pulses from the antenna are reflected by the water droplets as in the conventional weather radar system.

Using the Doppler shift principle, the frequency of energy pulse returned by droplets (B) moving towards the aircraft will be at a higher frequency than the transmitted frequency. The frequency of energy pulse returned by droplets moving away from the aircraft (A) will be at a lower frequency than the transmitted frequency. These Doppler shifts are used to determine the direction and velocity of the air movement resulting from a microburst.

17.4.7.3 Predictive wind shear

Visual and audible warnings of **predictive** wind shear conditions are provided to the crew. The visual warnings are given on the multi-function or navigation displays using a wind shear icon and message, together with warning lights on the glare shield. Audible warnings are provided as computer generated voice alerts over the cockpit speakers, typically 'WIND SHEAR WIND SHEAR'. The system automatically configures itself for the phase of flight; it is normally

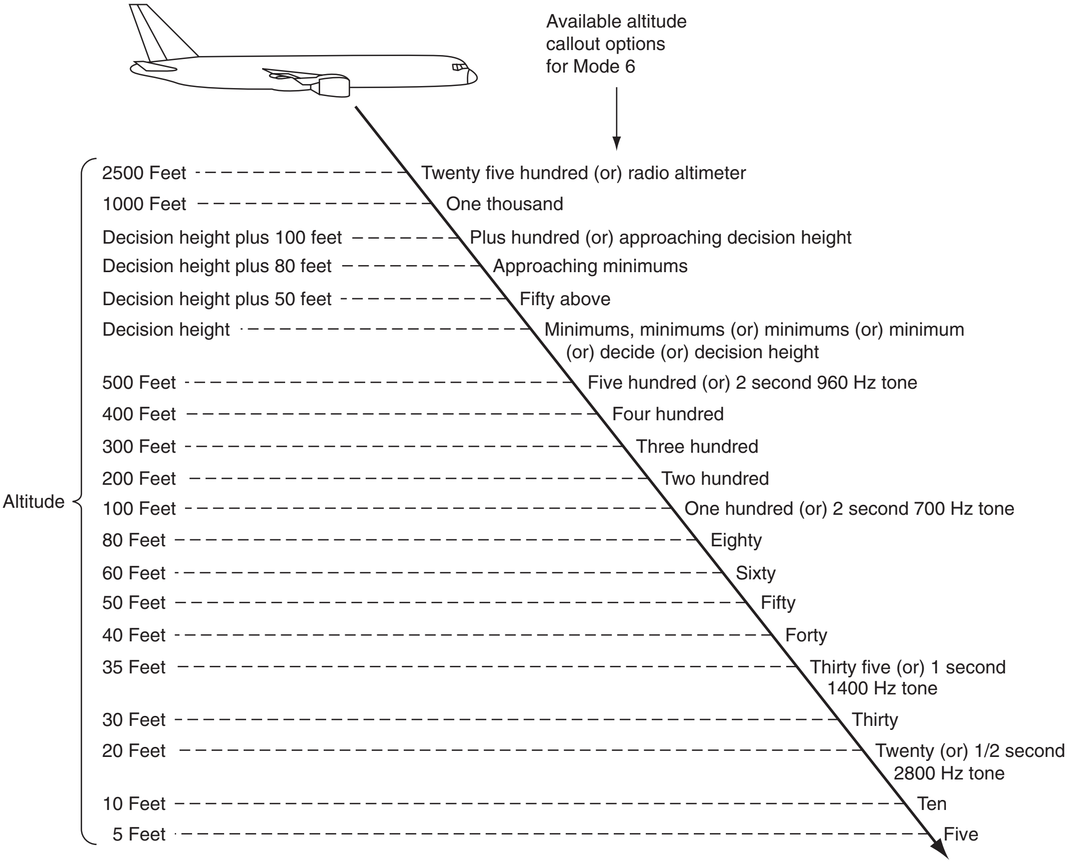
Available altitude
callout options
for Mode 6
Altitude
2500 Feet
Twenty five hundred (or) radio altimeter
1000 Feet
One thousand
Decision height plus 100 feet
Plus hundred (or) approaching decision height
Decision height plus 80 feet
Approaching minimums
Decision height plus 50 feet
Fifty above
Decision height
Minimums, minimums (or) minimums (or) minimum (or) decide (or) decision height
500 Feet
Five hundred (or) 2 second 960 Hz tone
400 Feet
Four hundred
300 Feet
Three hundred
200 Feet
Two hundred
100 Feet
One hundred (or) 2 second 700 Hz tone
80 Feet
Eighty
60 Feet
Sixty
50 Feet
Fifty
40 Feet
Forty
35 Feet
Thirty five (or) 1 second 1400 Hz tone
30 Feet
Thirty
20 Feet
Twenty (or) 1/2 second 2800 Hz tone
10 Feet
Ten
5 Feet
Five

Figure 17.16 GPWS Mode 6 altitude callouts

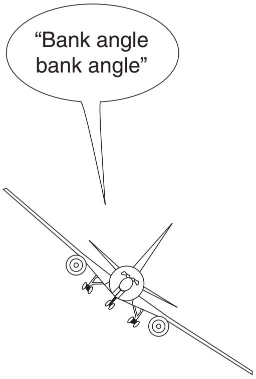
"Bank angle bank angle"

Figure 17.17 GPWS Mode 6 bank angle

inhibited below 50 feet radio altitude during take-off and landing. During an approach, Mode 7 is activated below 2500 feet radio altitude and has typical envelope incorporated as illustrated in Fig. 17.20.

Key point

When approaching a microburst, it creates an increase in headwind, causing a temporary increase of airspeed and lift.

17.5 Forward-looking terrain avoidance (FLTA)

This TAWS feature uses aircraft position, altitude and **flight path** information together with details of terrain, obstacles and known runways in the database. Referring to Fig. 17.21, the system provides an envelope of protection by comparing the flight path with the details contained in the database. There are two **look-ahead** distance features; these are functions of terrain clearance floor height and ground speed.

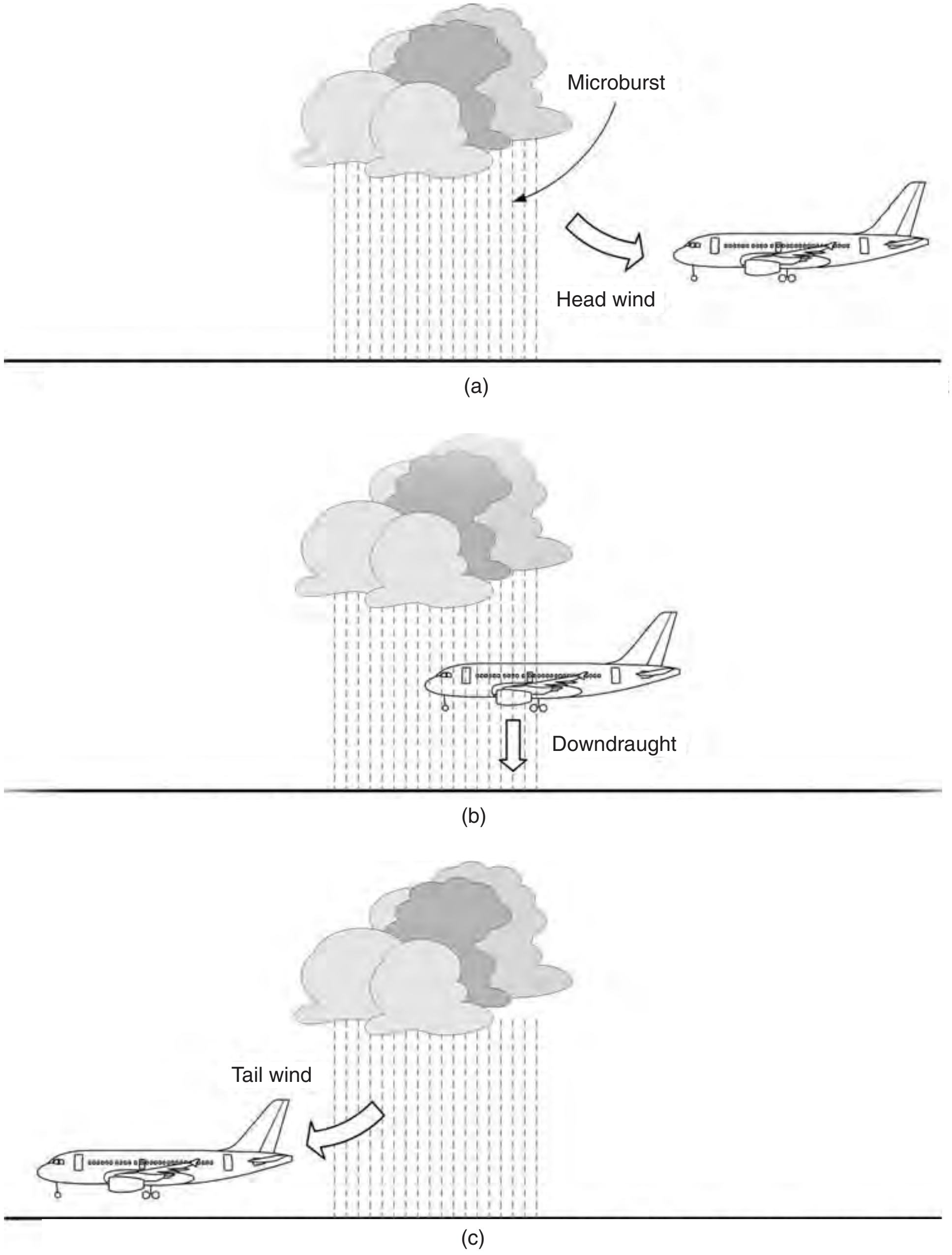
Microburst
Head wind
(a)
Downdraught
(b)
Tail wind
(c)

Figure 17.18 Microbursts

Referring to Fig. 17.21, alert areas are projected along the flight path in three levels:

- warning
- alert
- caution.

When a runway has been selected in the navigation system, a terrain clearance profile comes into effect as shown in Fig. 17.22. This profile take into account distance to the runway and height above ground level.

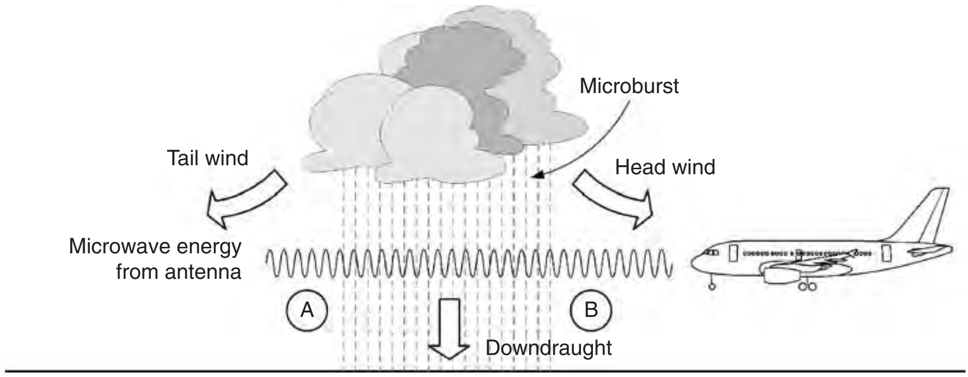
Microburst
Tail wind
Head wind
Microwave energy
from antenna
A
B
Downdraught

Doppler frequency shift between A and B indicates a wind shear condition

Figure 17.19 Doppler radar

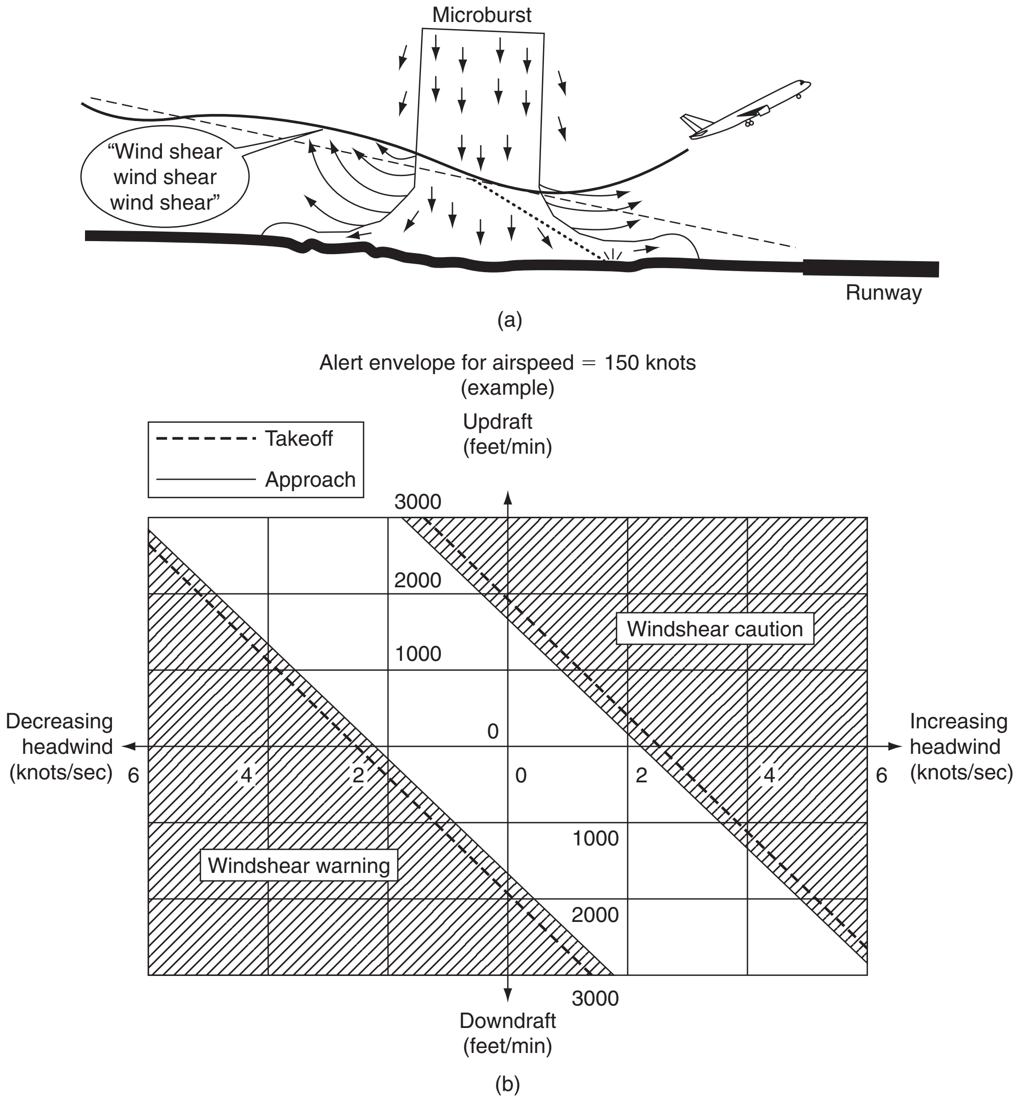
Microburst
"Wind shear
wind shear
wind shear"
Runway
Alert envelope for airspeed = 150 knots
(example)
Takeoff
Approach
Updraft
(feet/min)
3000
2000
1000
0
Windshear caution
Decreasing
headwind
(knots/sec)
6
4
2
0
2
4
6
Increasing
headwind
(knots/sec)
1000
2000
3000
Windshear warning
Downdraft
(feet/min)

(a)

(b)

Figure 17.20 GPWS Mode 7: (a) wind sheer detection profile, (b) wind sheer detection envelope

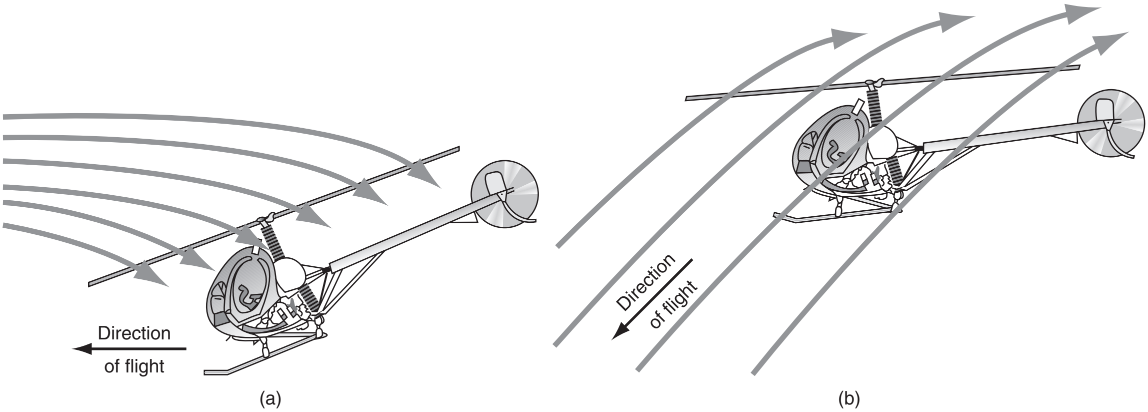
Direction
of flight
Direction
of flight
(a)
(b)

Figure 17.21 Rotorcraft autorotation: (a) normal powered flight, (b) autorotation

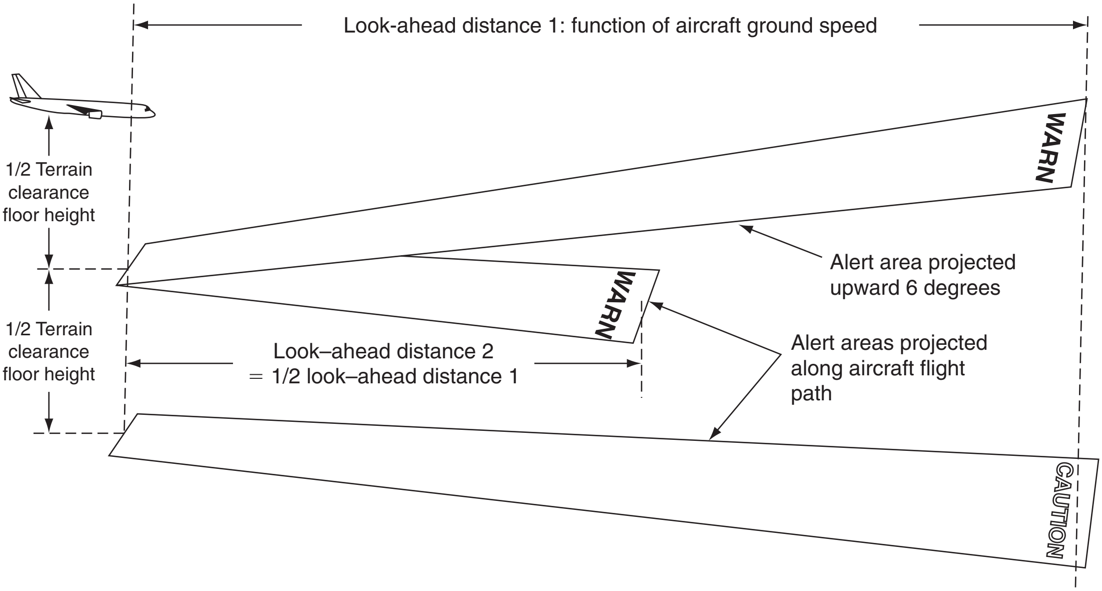
Look-ahead distance 1: function of aircraft ground speed
1/2 Terrain clearance floor height
1/2 Terrain clearance floor height
WARN
WARN
Alert area projected upward 6 degrees
Look–ahead distance 2 = 1/2 look–ahead distance 1
Alert areas projected along aircraft flight path
CAUTION

Figure 17.22 FLTA flight path

17.6 Rotorcraft TAWS

Rotorcraft, or helicopters, require additional TAWS considerations as result of their unique flying characteristics and operational roles. In addition to the functions already described, systems installed into rotorcraft need to include:

- excessive bank angle protection
- tail strike protection
- autorotation.

During **autorotation**, the relative upward airflow causes the main rotor blades to rotate at their normal speed; the blades are ‘gliding’ in their rotational plane, see Fig. 17.23. Autorotation facilitates a controlled descent through to an emergency landing in case of power plant failure; this flight regime has to be recognized by TAWS.

Many national aviation authorities including the FAA and CAA recommend the use of TAWS in helicopters performing specialized roles, e.g. emergency medical services and oil-pipe inspections these

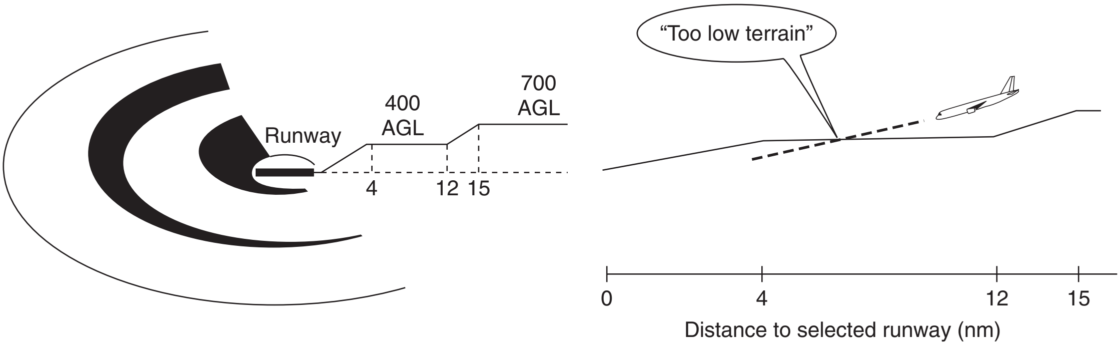
Runway
400
AGL
700
AGL
4
12
15
"Too low terrain"
0
4
12
15
Distance to selected runway (nm)

Figure 17.23 FLTA runway

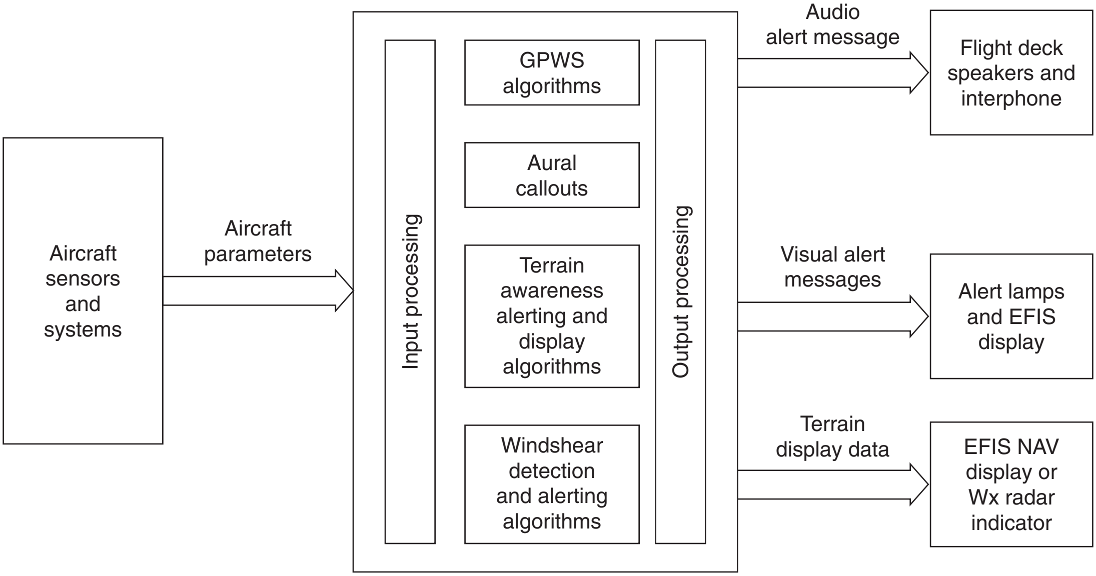
Aircraft sensors and systems
Aircraft parameters
Input processing
GPWS algorithms
Aural callouts
Terrain awareness alerting and display algorithms
Windshear detection and alerting algorithms
Output processing
Audio alert message
Flight deck speakers and interphone
Visual alert messages
Alert lamps and EFIS display
Terrain display data
EFIS NAV display or Wx radar indicator

Figure 17.24 TAWS architecture

often involve low-level operations at night or in low visibility.

17.7 Architecture and configurations

Typical Class A system architecture is illustrated in Fig. 17.24. Larger aircraft require TAWS Class A, smaller aircraft require TAWS Class B; this is reflected in the complexity (hence cost) of the hardware and certification requirements. The various TAWS configurations also take into account the type of interfaces that TAWS requires, e.g. whether the existing aircraft sensors are digital or analogue.

Smaller general aviation aircraft are adopting a simple terrain only feature. This is normally a subfeature of a GPS navigation system, referred to as **Class C TAWS**. Typical features include:

- terrain data base
- terrain elevations relative to the aircraft
- visual alerting
- reduced terrain clearance avoidance
- imminent terrain impact avoidance
- premature descent alert.

17.8 Future developments

TAWS is specified in transport category aircraft as a result of mandatory rulemaking. Two main categories of general aviation (GA) aircraft are not included in these rules, however: non-turbine, fixed-wing aircraft (particularly privately owned aircraft), and rotorcraft. Both of these categories fall under the voluntary Class C TAWS arrangement. Although there are no firm indications of general aviation's adoption of Class C TAWS equipment, the functionality is often included within existing equipment, and many aircraft owners would prefer some level of protection.

Some manufactures are developing an **assisted recovery** concept, where the aircraft's automatic flight control system automatically performs the pull-up manoeuvre following a TAWS alert. The system could also steer the aircraft away from prohibited airspace for security purposes.

Performance-based systems are being developed that utilize data from an active aircraft database that stores the aircraft's performance characteristics under varying conditions of engine thrust, aircraft weight, altitude, temperature and other inputs. When a TAWS 'PULLl UP' warning is given, the system determines if the aircraft can actually climb at the required rate to avoid impacting the terrain ahead. If the system predicts that this is not possible, the crew receive an 'AVOID TERRAIN' command, and the system indicates an alternate manoeuvre to fly the aircraft into safer areas.

To minimize nuisance warnings, the system limits its look-ahead terrain search profile to a narrow sector either side of the aircraft's projected track. Terrain advisory lines are indicated on a wider scan either side of the projected track to visually alert the crew of potential obstructions. When a terrain avoidance manoeuvre is initiated, the system switches to a wider scanning sector and anticipates the rate of turn while searching the terrain ahead on a from the original track. Terrain that the aircraft can safely climb over is then depicted with appropriate graphics on the navigation display.

The ground collision avoidance module (GCAM) is a predictive alerting technique derived from a terrain-following/terrain-avoidance system used by military aircraft. GCAM uses a collision prediction and alerting (CPA) function that combines terrain and airport databases with a Model of the aircraft's climb capability. The CPA function predicts potential terrain conflicts from the terrain environment and the aircraft's predicted flight path.

The former correlates precise aircraft position information from the GPS and/or other navigation sources with the digital database. The flight path prediction uses current aircraft flight path data to project the flight path more than 2 minutes ahead of the aircraft, and provides visual cautions up to 4 minutes ahead. Alerts are generated when the CPA system predicts the flight path could intersect the terrain, providing **terrain avoidance** rather than just a warning.

17.9 Multiple choice questions

1. A terrain awareness display is only required for:
   (a) Class A TAWS
   (b) Class-B TAWS
   (c) GPWS.

2. Mode 1 terrain awareness cautions are given for:
   (a) negative climb rate or altitude loss after take-off
   (b) excessive descent rate
   (c) altitude callout at 500 feet.

3. When approaching a microburst, it creates:
   (a) a decrease in headwind causing a temporary increase of airspeed and lift
   (b) an increase in headwind causing a temporary decrease of airspeed and lift
   (c) an increase in headwind causing a temporary increase of airspeed and lift.

4. Warning alerts are given for a terrain threat that requires:
   (a) immediate crew awareness
   (b) confirmation with air traffic control
   (c) immediate crew action.

5. Forward-looking terrain avoidance (FLTA) looks:
   (a) ahead of and below the aircraft's lateral and vertical flight path
   (b) ahead of and above the aircraft's lateral and vertical flight path
   (c) either side of and below the aircraft's lateral and vertical flight path.

6. Premature descent alert compares the aircraft's:
   (a) ground speed with the proximity of the nearest airport

(b) lateral and vertical position with the proximity of the nearest airport
(c) lateral and vertical position with the proximity of high terrain.

7. Red areas are used on TAWS displays to indicate terrain that is:
(a) above the aircraft's current altitude
(b) level with the aircraft's current altitude
(c) safe in terms of required terrain clearance.

8. The low range radio altimeter (LRRA) is a:
(a) self-contained vertically directed primary radar system
(b) self-contained horizontally directed primary radar system
(c) self-contained vertically directed secondary radar system.

9. Mode 7 (wind shear) is normally inhibited:
(a) above 50 feet radio altitude during take-off and landing
(b) below 50 feet radio altitude during take-off and landing
(c) during take-off and landing.

10. The TAWS computer function creates a four-dimensional situation comprising:
(a) latitude, longitude, heading and time
(b) latitude, speed, altitude and time
(c) latitude, longitude, altitude and time.

11. Referring to Fig.17.25, this illustrates a:
(a) FLTA profile
(b) GPWS Mode 2 profile
(c) GPWS Mode 1 profile.

12. Referring to Fig. 17.26, this illustrates a:
(a) GPWS Mode 6 caution
(b) GPWS Mode 3 caution
(c) GPWS Mode 1 warning.

13. Referring to Fig.17.27, this is the warning envelope for:
(a) GPWS Mode 5
(b) GPWS Mode 2
(c) GPWS Mode 1.

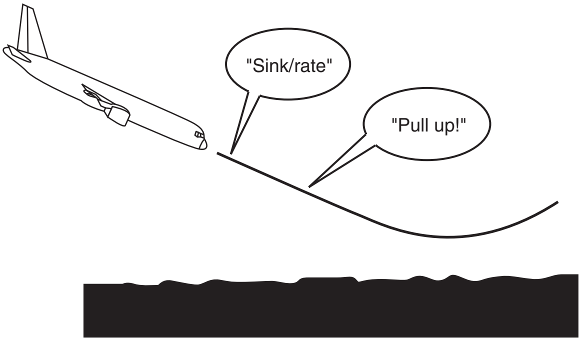

(a)

Figure 17.25 See Question 11

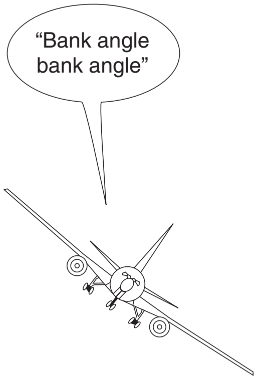

Figure 17.26 See Question 12

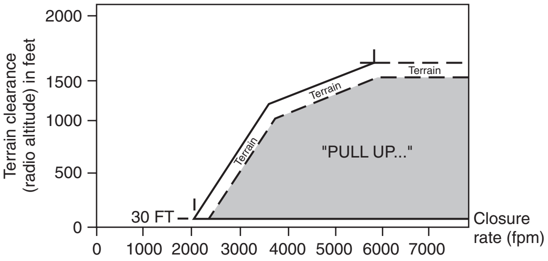

(a)

Figure 17.27a See Question 13

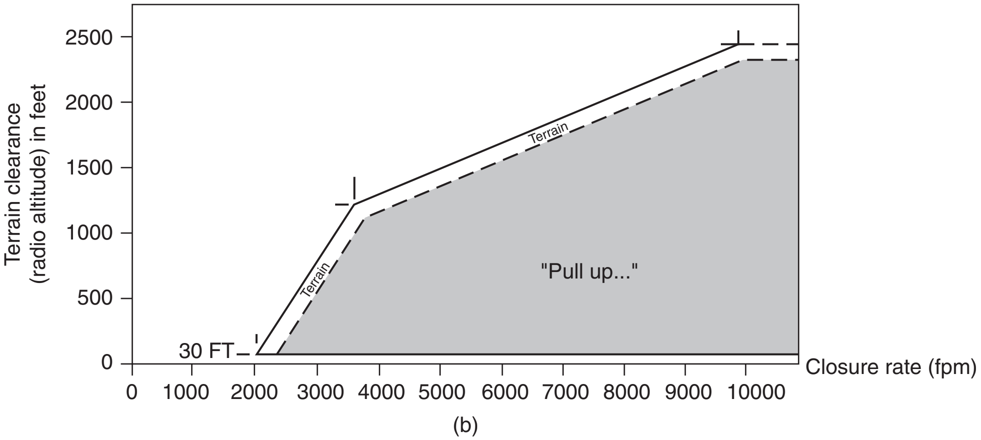

Figure 17.27b See Question 13

14. GPWS Mode 3 is defined as:
    (a) Excessive terrain closure rate
    (b) Negative climb rate, or altitude loss after take off or go around
    (c) excessive descent below the glide slope

15. Flight into terrain whilst not in the landing configuration is GPWS Mode:
    (a) one
    (b) three
    (c) four

Chapter 18 Flight data and cockpit voice recorders

One of the fundamental electrical/electronic systems associated with the aircraft industry is the crash-survivable flight data recorder (FDR). This is often referred to in the press as the 'black box', even though the item is painted bright orange! Flight data recorders used for accident investigation are mandatory items of equipment in commercial transport aircraft. Efforts to introduce crash-survivable flight data recorders can be traced back to the 1940s; the FDR has now been supplemented with the cockpit voice recorder (CVR). Recorders are used after an incident or accident as an integral part of the investigators' efforts to establish the cause(s).

Data recorders can also be used to indicate trends in aircraft and engine performance. Algorithms are established for healthy and normal conditions during the aircraft's flight-testing programme and the early period of service. These algorithms include engine parameters such as engine exhaust temperature, oil pressure and shaft vibration for given speeds and altitudes. These parameters are then monitored during the aircraft's life; any deviations from the norm are analysed to determine if the engine requires inspection, maintenance or removal. This chapter reviews the range of FDR/CVR technologies that are employed for both accident investigation and trend monitoring.

18.1 Flight data recorder history

A series of fatal accidents involving the De Havilland DH106 Comet in 1953 and 1954 led to the grounding of the entire Comet fleet pending an investigation. There were no survivors or witnesses to these accidents and an exhaustive investigation at the Royal Aircraft Establishment at Farnborough, UK, eventually concluded that metal fatigue was the primary cause of the accidents. One of the ideas that came out of these investigations was for the development of a **crash-survivable** device to record the flight crew's conversations and other aircraft data. Doctor David Warren of the Aeronautical Research Laboratories of Melbourne, Australia, was on one of the investigation committees; he concluded that such recorders would be of tremendous benefit in determining the likely cause(s) of an accident. The idea was that design improvements and/or operating procedures would be introduced to eliminate avoidable accidents from the same cause. The proposed recorder would have to survive both the high g forces and the potential fire resulting from an aircraft crash; the technology required for such a device did not become available until 1958.

Key point

Flight data recorders (FDR) are often referred to in the press as the 'black box', even though the item is painted bright orange.

Test your understanding 18.1

Describe how data recorders can also be used to indicate trends in aircraft and engine performance.

18.1.1 Scratch foil

The technology used in first-generation flight data recorders was based on a roll of **steel foil tape** embossed with five separate items of recorded data (parameters), see Fig. 18.1; these parameters were:

- heading
- altitude
- air speed
- vertical acceleration
- time.

The foil is made from a high nickel content steel; several hundred feet of foil, approximately 125 mm wide, is rolled onto a spool. When the recorder is switched on, the foil is transferred from a feed spool onto the take-up spool after passing under five recording needles (hardened or diamond-tipped styli) that move across the foil in response to their respective inputs. With the foil tape removed from the recorder, the various traces (Fig. 18.2) are compared with a transparent overlay with markings to measure the recorded parameters.

This early ‘scratched-foil’ technology gave the investigators more data than had been previously available; however, it was soon realized that five parameters did not provide sufficient data for meaningful accident investigation.

18.1.2 Magnetic recording

Second-generation recorders were introduced onto aircraft during the 1960s based on one of two recording media; **steel wire** or **magnetic tape**. When the recorder is switched on, the wire is transferred onto another drum after passing over record/replay heads. Steel wire recording technology was used by the telegraphy industry from the late 1890s; it was subsequently developed for other industrial applications and home entertainment until the introduction of magnetic tape recorders in the late 1940s. Even with very fine diameter of approximately 100 micrometers, steel wire is intrinsically more crash-survivable than magnetic tape, and this medium was used for recording **analogue data** from the 1950s through to the 1970s. Wire recorders have a high media density; this is made possible by the solid metal medium and relatively fast speed (typically 24 inches per second) over the record/replay heads. One hour of recorded data could be achieved with approximately 7000 feet of wire wound onto a three-inch-diameter drum.

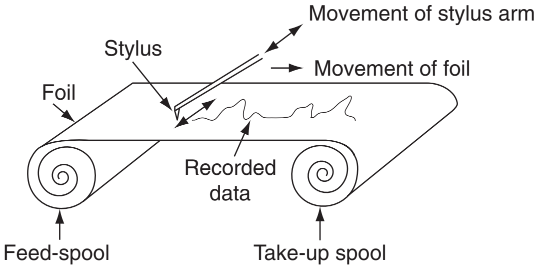

Figure 18.1 Scratched foil recorder principles

The technology required to protect magnetic wire or tape was eventually developed; the first application of magnetic tape recorders was to record the last thirty minutes of crew voice communications and noise within the flight compartment environment. The cockpit voice recorder (CVR) required substantial packaging to enable it to be crash-survivable. This magnetic tape technology (recording data in a **digital format**) was then applied to FDR applications. This third-generation FDR facilitated the recording of many more flight parameters from the airframe and engines while meeting higher crash and fire protection requirements. Inside the write head is a small coil; when voltages are applied to the coil, an electromagnetic field is created; particles in the wire or tape are aligned on a micro-level with the field as it passes over the head.

To analyse electromagnetic recorded tape, an electronic process is required such that digitally coded data can be converted back (decoded) into its analogue form required for presentation and/or display purposes. To read the data stored on the wire/tape, it is passed over a reading head; this is similar in physical construction to the writing head. As each section of wire or tape passes

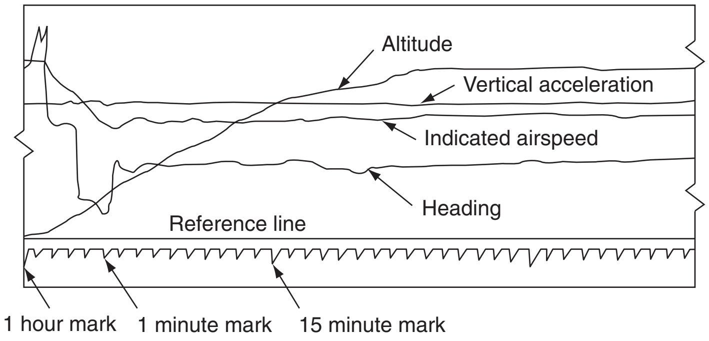

Figure 18.2 Scratched foil recorded data

over the head, the particles are aligned in specific directions, inducing small voltages into the coil. The polarity of the voltages indicates whether a binary one or zero is stored. The voltage signals are processed electronically to produce analogue signals for use in displays or charts.

18.1.3 Solid state data recorders

Solid-state digital flight data recorder (DFDR) technology became commercially viable in the 1990s. Data is stored in semiconductor memory via integrated circuits. Solid-state memory does not require any servicing, maintenance or overhaul. Furthermore, retrieval and interpretation of the data via personal computer-based software is a more efficient process.

Recorders are now also used for reasons other than accident investigation, e.g. to indicate trends in aircraft and engine performance. Algorithms are established for healthy conditions during the aircraft's flight-testing programme and the early period of service; these include engine parameters such as oil temperature, oil pressure and shaft vibration for given speeds and altitudes. These parameters are then monitored during the aircraft's life and any deviations from the norm are analysed to determine if the engine requires inspection, maintenance or removal. Low-cost, high-performance, rugged solutions are required for the processing and data recording. Equipment such as the S3DR product range developed by Specialist Electronics Services (SES) Ltd has many applications in the aerospace and defence industry. The S3DR product range developed and fielded by SES is designed with flexibility and expandability in mind.

The S3DR-F is a panel-mounted data recorder providing a recording solution where space and mass are at a premium. It is specified where access to the removable media is required from within the flight compartment. The unit features a high-speed recording function that is able to write data to a removable solid state memory module from multiple monitored interfaces at speeds up to 35 Mbits per second (MBPS). Memory modules are available in many sizes up to 16 GBytes. A Pentium® compatible processor, allowing application specific data gathering regimes to be accommodated, provides processing functions.

The Solid State Data Recorder-Expandable (S3DR-E) is a high-speed, high-capacity expandable data recorder with extensive data acquisition and communications flexibility. The recorder is built from the concept of multiple recording, data acquisition and communication modules that are assembled together to provide a system that meets the specific bespoke requirements of a customer application.

The recorder contains a master processor module which provides control and configuration of any other modules contained in the system. This master processor provides a high-performance general purpose processor that controls all other data acquisition, recording and communication modules. The processor module also provides multiple high-speed USB 2.0, Giga Ethernet and VGA interfaces.

The baseline S3DR-E recorder is built from a master processor combined with a recording module and an integrated power supply. Alternative recorder configurations can be built from any combination of up to seven data acquisition, communications and recording modules. Each recording module can be built from either a dual PCMCIA recorder with a dedicated I/O processor, or a high-capacity (100 byte), high-speed (up to 480 MBPS) dedicated recording module.

18.2 Mandatory equipment requirements

In the late 1960s, the UK Civil Aviation Authority (CAA) stipulated a need for additional parameters to be recorded; FDRs using scratched foil were developed to mark both sides of the foil. This allowed for the recording of additional parameters, e.g. pitch, roll and flap position. Although this provided additional information for the accident investigator, it increased the complexity of the recorder resulting in lower reliability. Furthermore, the additional information contained on the scratch-foil was becoming harder to read and required high skill levels to interpret the data.

Larger, more complex aircraft were introduced in the 1960/70s. These aircraft were flying faster, higher and farther than had been previously possible. They had reduced flight crew; the navigator and wireless operators were being replaced by computerized (albeit analogue) avionics bringing reduced operating costs. High-density passenger operations, reduced flight crew and longer flights were becoming the norm. This created new problems for the accident investigators; more data was required to identify the probable cause of an accident. Flight data acquisition units (FDAU) were introduced to provide an interface between aircraft systems and the FDR. The FDAU collects, or acquires, a variety of analogue signals from around the aircraft, converts them into a single digital data stream to the recorder. In the late 1970s and early 1980s the

digital flight data recorder (DFDR) could record 32 parameters for a 25 hour period. (This represented the round-trip time between cities in the USA, Europe and Asia.)

Key point

The FDAU collects, or acquires, a variety of analogue signals and converts them into a digital data stream to the recorder.

The specific parameters to be recorded are stipulated by the regulatory authorities. In Europe, commercial air transport (fixed wing and rotary wing) aircraft are regulated under JAR OPS 1 and 3 respectively. For flight data recorders, these regulations take into account when the aircraft was first issued with an individual Certificate of Airworthiness. Aircraft operating under JAR OPS 1 are required to install a **digital flight data recorder** (DFDR). The DFDR has to be able to retain the recorded data for a minimum of the last 25 hours of its operation (10 hours for aircraft below 5700 kg maximum take of mass). The flight data recorder must start to record data automatically prior to the aircraft being capable of moving under its own power. The recorder must stop automatically after the aircraft is incapable of moving under its own power. The flight data recorder must have a device to assist with locating that recorder if it is submerged in water.

The mandatory parameters required for an aircraft depend on the size of the aircraft and the prevailing regulatory rules applied to that aircraft. The mandatory parameters required for multi-engine turbine powered with maximum approved seating configuration for more than nine passengers and maximum take-off mass under 27,000 kg are shown in Table 18.1. (Note that all parameters must be referenced to a timescale or relative time count.)

For aircraft over 27,000 kg maximum take-off mass, additional mandatory requirements are required; these are listed in Table 18.2. Aircraft with an electronic flight instrument system (EFIS) are required to have the additional parameters listed in Table 18.3.

Electronic flight instrument system (EFIS) technology was introduced during the 1980s based on the cathode ray tube (CRT); the established technology is now liquid crystal displays (LCD). The attitude director indicator (ADI) and horizontal situation indicator (HSI) are included within a basic EFIS installation; other indications and displays are incorporated through various configurations.

Table 18.1 JAR OPS 1 mandatory FDR parameters

- Time or relative time count
- Pressure altitude
- Indicated airspeed
- Heading
- Normal acceleration
- Pitch attitude
- Roll attitude
- Manual radio transmission keying
- Propulsive thrust/power on each engine and thrust/power lever position (if applicable)
- Trailing edge flap or cockpit control selection
- Leading edge flap or cockpit control selection
- Thrust reverser status
- Ground spoiler position and/or speed brake selection
- Total or outside air temperature
- Autopilot and autothrottle mode and engagement status
- Longitudinal acceleration (body axis)
- Lateral acceleration[1]
- Angle of attack[2]

[1] Lateral acceleration on aircraft with 5700 kg maximum take-off mass.

[2] Angle of attack on aircraft with less than 5700 kg maximum take-off mass.

There are some exceptions to the above allowed under JAR OPS 1, that have to be agreed and authorized, e.g. if the required transducer is not available; or if the aircraft system or equipment producing the data needs to be modified; or if the signals are incompatible with the recording system. It is a fundamental requirement that the parameters being recorded are obtained from sources that enable accurate correlation with the information displayed to the flight crew.

Helicopters used for commercial transportation (with maximum certificated take-off mass over 3175 kg) also require a flight data recorder. This must be capable of recording data during at least the last eight hours of its operation; mandatory parameters are shown in Table 18.4. Helicopters over 7000 kg require the additional parameters shown in Table 18.5.

Table 18.2 JAR OPS 1 Mandatory FDR parameters (aircraft over 27,000 kg)

- Primary flight controls – control surface position and/or pilot input (pitch, roll, yaw)
- Pitch trim position
- Radio altitude
- Vertical beam deviation (ILS glide path or MLS elevation)[1]
- Horizontal beam deviation (coupled approaches: ILS localizer or MLS azimuth)
- Marker beacon passage
- Warnings (specified on the aircraft installation)
- Optional navigation receiver frequency selection
- Optional DME[2] distance
- Landing gear squat switch status or air/ground status
- Terrain awareness warning system (TAWS)
- Angle of attack
- Low pressure warning (hydraulic and pneumatic power)
- Groundspeed
- Landing gear or gear selector position

[1] Instrument landing and microwave landing systems (ILS and MLS) are used for precision approach and landing.

[2] Distance measurement equipment (DME) is a short/medium-range navigation system based on secondary radar principles. (These systems are described in a related book in the series: *Aircraft Communications and Navigation Systems*.)

Table 18.3 Additional parameters for aircraft with electronic instruments

- Selected barometric setting
- Selected altitude
- Selected speed
- Selected mach
- Selected vertical speed
- Selected heading
- Selected flight path
- Selected decision height
- EFIS display format
- Multi function/engine/alerts display format

Table 18.4 Mandatory FDR parameters for commercial transport helicopters (less than 7000 kg)

- Time or relative time count
- Pressure altitude
- Indicated airspeed
- Heading
- Normal acceleration
- Pitch attitude
- Roll attitude
- Manual radio transmission keying
- Power on each engine
- Engine power control position
- Main rotor speed
- Rotor brake (if installed)
- Primary flight controls
- Collective pitch
- Longitudinal cyclic pitch
- Lateral cyclic pitch
- Tail rotor pedal
- Controllable stabilator
- Hydraulic selection
- Warnings
- Outside air temperature
- Autopilot engagement status
- Stability augmentation system engagement

Table 18.5 Additional parameters for helicopters over 7000 kg

- Main gear box oil pressure
- Main gear box oil temperature
- Yaw rate or yaw acceleration
- Indicated sling load force (if installed)
- Longitudinal acceleration (body axis)
- Lateral acceleration
- Radio altitude
- Vertical beam deviation (ILS glide path or MLS elevation)
- Horizontal beam deviation (ILS localizer or MLS azimuth)
- Marker beacon passage
- Warnings
- Optional navigation receiver frequency selection
- Optional DME distance
- Optional navigation data
- Landing gear or gear selector position

Key point

The flight data recorder must stop automatically after the aircraft is incapable of moving under its own power.

Key point

Parameters being recorded are obtained from sources that are accurately correlated with the information displayed to the flight crew.

18.3 Flight data recorder (FDR) specifications

Initial FDR technology required that the unit be able to withstand 100 g of impact; the equipment was often installed alongside other equipment in the avionics bay. As more experience was gathered in the operational use of FDRs, the impact requirements were increased to 1000 g. It was also learned through experience that the optimum location for the recorder was at the rear of the aircraft. This rational is based on the statistical probability that, following initial impact, the rear structure of the aircraft will remain intact. When the front of the aircraft is crushed, it effectively reduces the shock impact that reaches the rear of the aircraft. Evidence from aircraft accidents shows that the rear fuselage and empennage are most likely to survive intact and therefore provide the optimum location for the recorder. It is imperative that, in the event of an accident, the data recorder be located as swiftly as possible. For this reason they are painted bright orange and are fitted with an **underwater location transmitter**, see Fig. 18.3.

Key point

The underwater location transmitter is sometimes called an underwater locator beacon (ULB).

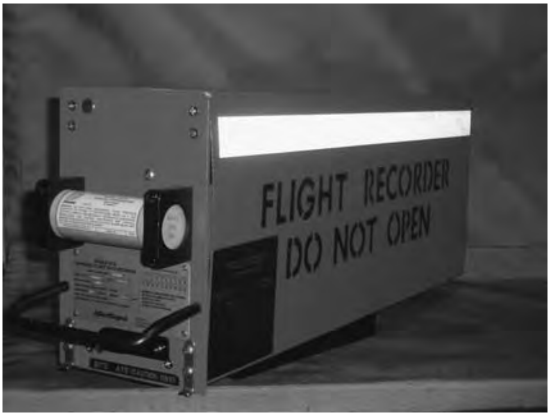

Figure 18.3 Underwater location transmitter

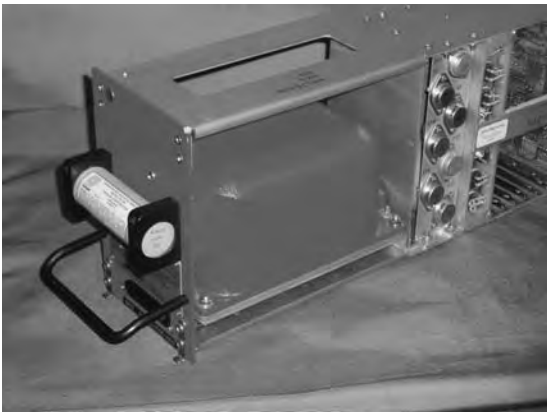

Figure 18.4 Crash-proof recorder

To ensure survivability of a flight data recorder, modern products are typically double-wrapped in strong corrosion-resistant stainless steel or titanium, with high-temperature insulation. The crash-survivable portion of the recorder (see Fig. 18.4) is designed and tested for a variety of severe conditions derived from aircraft crash data; the most significant requirements being:

Crash impact:	3400 g deceleration for 6.5 ms.
Crushing:	5000 pounds pressure applied against all six sides of the recorder
Piercing:	500 pounds mass with a hardened steel pin dropped from ten feet
Fire:	1100°C for 60 minutes
Water:	20,000 ft (submerged)

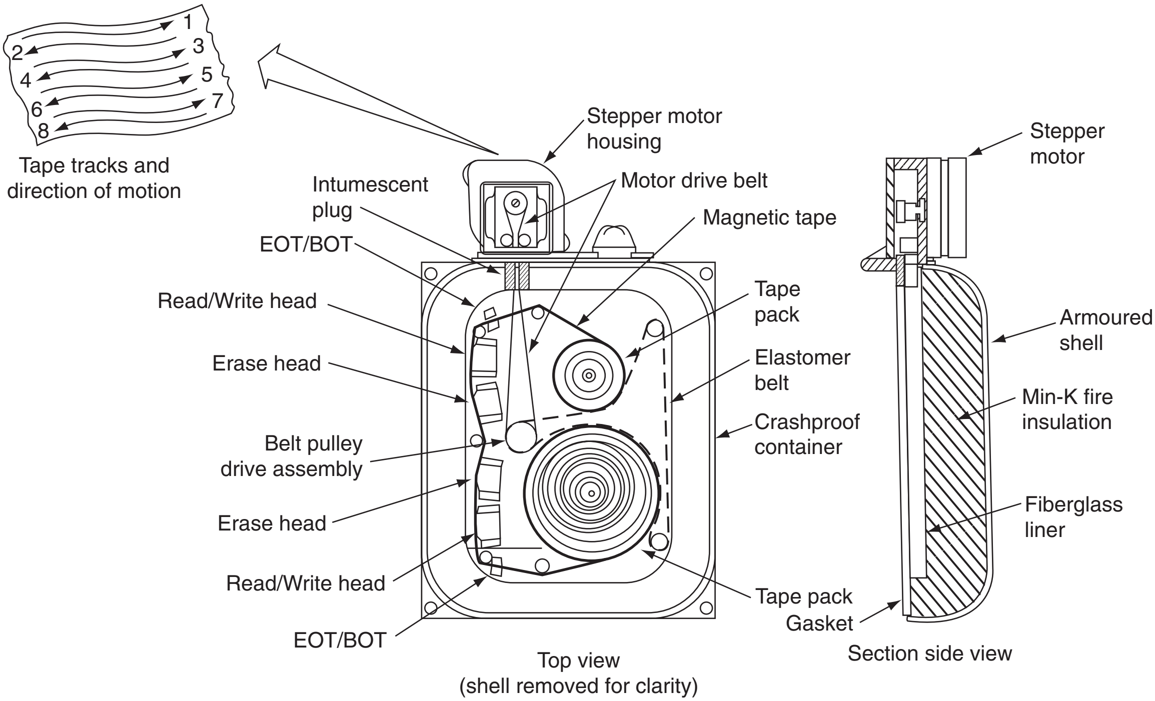
1
2
3
4
5
6
7
8
Tape tracks and direction of motion
Stepper motor housing
Intumescent plug
Motor drive belt
Magnetic tape
EOT/BOT
Read/Write head
Tape pack
Erase head
Elastomer belt
Belt pulley drive assembly
Crashproof container
Erase head
Read/Write head
Tape pack
EOT/BOT
Gasket
Top view (shell removed for clarity)
Stepper motor
Armoured shell
Min-K fire insulation
Fiberglass liner
Section side view

Figure 18.5 Crash-proof recorder tape drive

These are specific requirements for the **crash-survivable** section of the recorder. This part of the recorder fully encloses the tape and transport mechanism (Fig. 18.5); the drive motor is contained externally. Generic requirements for the entire product include resistance to a number of environmental conditions including:

- hydraulic fluids
- fuel de-icing fluids
- fire extinguishing agents etc.

Key point

The mandatory parameters required for an aircraft DFDR depend on the size of the aircraft and the prevailing regulatory rules applied to the operation of that aircraft.

18.3.1 Recording tape

Typical DFDR recording tape is 0.25 in. wide made of Kapton or Mylar; these are the trade names for self-lubricating polyamide films which can remain stable in a wide range of temperatures from −269°C to +400°C. Up to 500 feet of tape is arranged by eight bi-directional tracks with sequential switching. The tape is transferred from one reel to another by a motor-driven belt, and transported over the write/read heads. The motor is a reversible stepper motor, located outside of the crash-survivable enclosure rotating at a nominal rate of 700 steps per second. Odd-numbered tracks are used when the tape is travelling in the nominal forward direction; even-numbered tracks are used in the reverse direction, see Fig. 18.5. When an entire track has been used, the direction of travel is switched and the next track utilized. At the end of recording on track eight, the recording will be switched back to track one, and old data is thereby overwritten. The tape travels at speeds up to five inches per second on an incremental basis when data is being written, or five inches per second continuously when data is being read. Beginning of tape (BOT) and end of tape (EOT) sensors are used to reverse the direction of travel and switch to the next track.

Key point

The DFDR on large aircraft has to be able to retain the recorded data for a minimum of the last 25 hours of its operation.

Test your understanding 18.2

What are the five significant requirements that a crash-survivable recorder has to withstand?

Two four-channel write/read heads are used for recording and playback from the tape. One head is enabled when the tape moves in one direction, and writes the odd numbered tracks. When the tape direction is reversed, the other write/read is enabled and writes to the even numbered tracks. Data is written at a rate of 11.4 kilo bits per second (kbps). Before writing new data, the old data are removed on the engaged track by dedicated **erase heads**.

18.3.2 Data acquisition

Modern flight data recorders are based on digital technology to enable more parameters to be recorded. Digital flight data recorders (DFDR) installed on aircraft equipped with a distributed digital data bus (Fig. 18.6) are integrated as one of the many avionic systems.

There are many aircraft, however, that have limited or no digital equipment. Depending on the size and category of aircraft, e.g. if it is used for public transport, it still has to have a DFDR. The DFDR will be used for recording both mandatory and domestic parameters. The latter is used for the benefit of the operator; it might only have a 30 minute duration, sufficient for monitoring specific systems, e.g. to evaluate the performance of engines or automatic landings. This type of data is often recorded on a cassette termed a **quick access recorder** (QAR); the QAR is removed at the convenience of the operator and information is analysed via customized software packages.

Assuming that the aircraft has limited digital capability, parameters required for both mandatory and domestic recording are gathered in analogue form and converted into digital form via an **aircraft integrated data system** (AIDS). This system also allows the digital information to be retrieved and formatted for analogue purposes, e.g. for display on a monitor or printed on a chart; Fig. 18.7 illustrates a typical system.

The acquisition units sample the data to be recorded and converts them into digital form suitable for recording. The logic unit also inhibits the recording (REC) function until the aircraft is capable of moving under its own power; this switch is controlled by one of many possible options, e.g. when ground power is removed, when an alternator comes on line or when ground speed increases over a specified value. The TEST switch allows the system to be tested on the ground by overriding the REC logic.

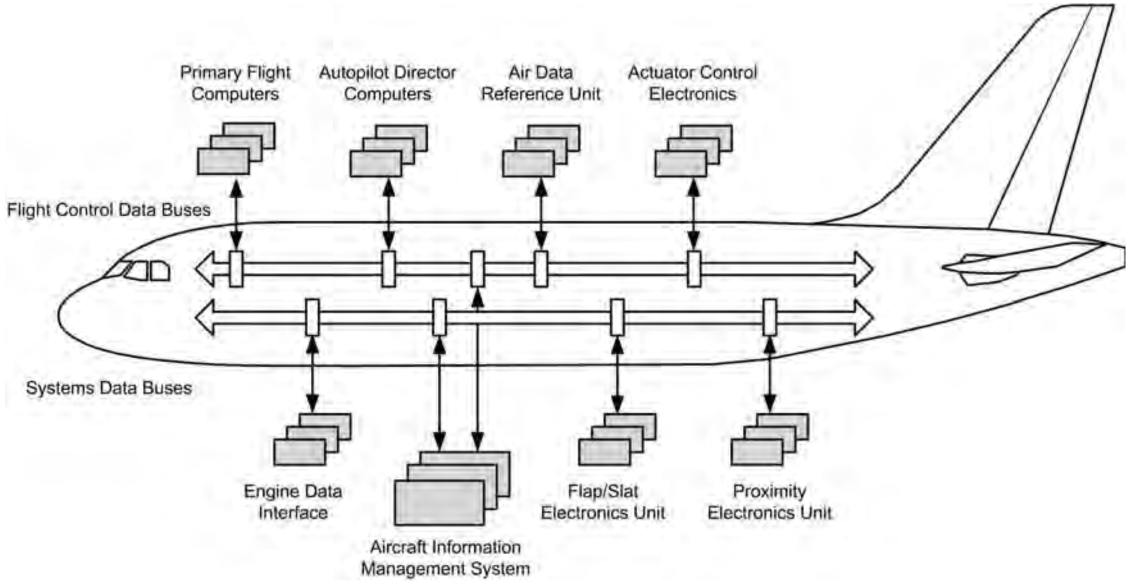
Primary Flight Computers
Autopilot Director Computers
Air Data Reference Unit
Actuator Control Electronics
Flight Control Data Buses
Systems Data Buses
Engine Data Interface
Aircraft Information Management System
Flap/Slat Electronics Unit
Proximity Electronics Unit

Figure 18.6 Digital data bus

Test your understanding 18.3

What is the purpose of a QAR?

The DFDR control display panel is used to enter the aircraft flight number (or some other identification, weight, date etc. Once the system is in operation, specific parameters being recorded can be displayed on the panel. The control panel often has a push-button momentary switch that is used to mark a particular event. (Pushing the event button puts a discrete signal on the recording to highlight the timing of the event.)

18.3.3 Digital data recording formats

Typical data output to the DFDR comprises a continuous stream of digital data formed into frames, see Fig. 18.8. Each frame is divided into four sub-frames, typically of one-second duration each; the sub-frame is formed by 64 12-bit words.

The first word in each sub-frame is a synchronizing word; the other three contain data. Each word contains 12 bits of digital information; the total number of bits in a sub-frame is therefore 768. Various formats are used to form the digital bits of logic one and logic zero, see Fig. 18.9. These formats include:

- non-return-to-zero (NRZ)
- bipolar-return-to-zero return-to-zero (RZ)
- Harvard bi-phase format.

Non-return-to-zero (NRZ) logic one is formed by a 5 V DC level; logic zero is indicated by 0 V DC. The two logic levels are therefore represented by one of two significant conditions, with no other neutral or rest condition. A clock waveform is required to distinguish between bits. Two wires are used to carry the signal, together with a third wire for a clock reference.

Bipolar-return-to-zero return-to-zero (RZ) describes a code in which the signal drops (returns) to zero between each pulse. This takes place even if a number of consecutive logic zeros or ones occur in the signal. Logic one is indicated by a 0.5 V DC level, and logic zero by a ±0.5 V DC level. The signal returns to 0 V DC in the second half of each bit. The signal is self-clocking; therefore separate clock pulses are not required alongside the signal.

In the **Harvard bi-phase** format, each bit changes state at its trailing edge; either from high to zero or

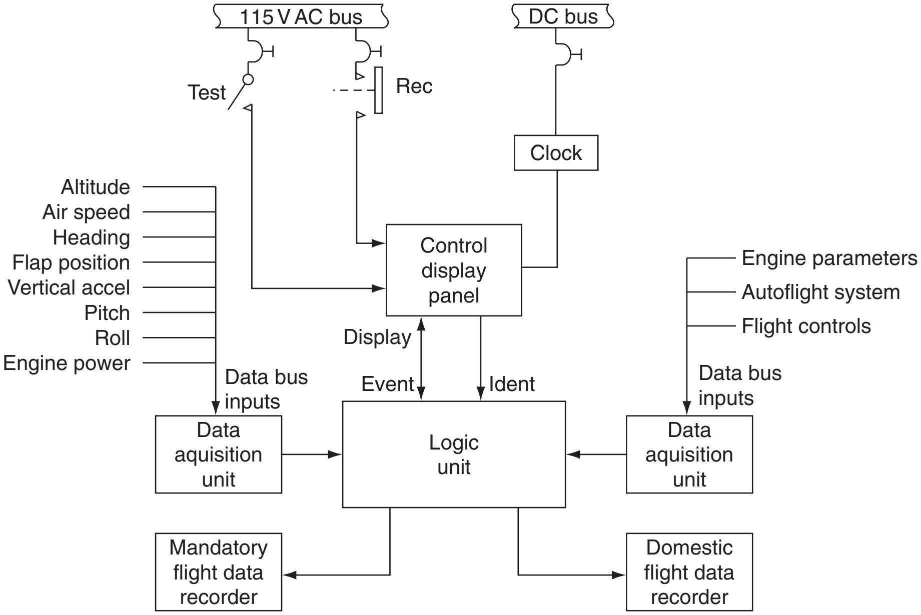
115 V AC bus
DC bus
Test
Rec
Clock
Altitude
Air speed
Heading
Flap position
Vertical accel
Pitch
Roll
Engine power
Control display panel
Engine parameters
Autoflight system
Flight controls
Display
Data bus inputs
Event
Ident
Data bus inputs
Data aquisition unit
Logic unit
Data aquisition unit
Mandatory flight data recorder
Domestic flight data recorder

Figure 18.7 AIDS overview

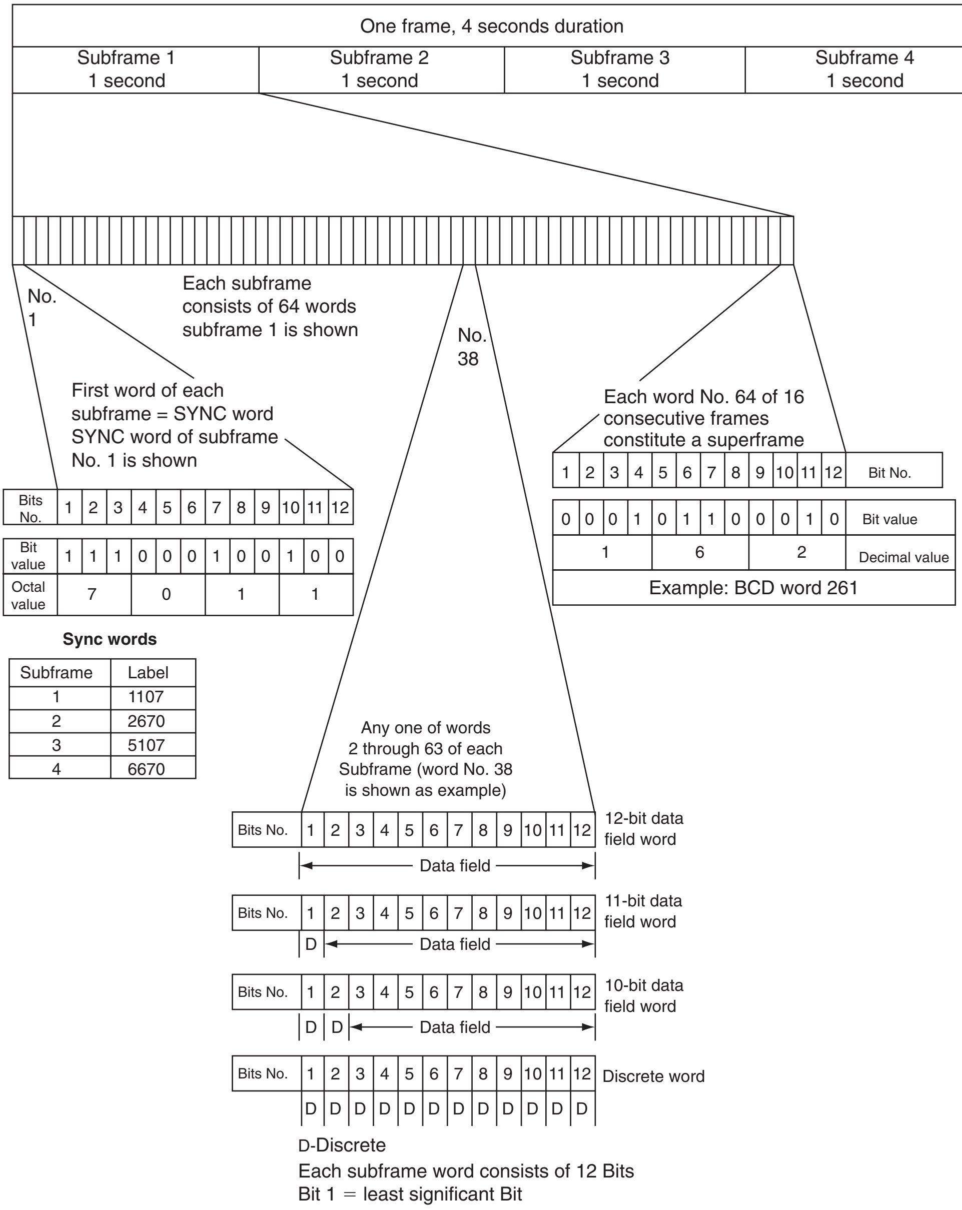
One frame, 4 seconds duration
Subframe 1 1 second
Subframe 2 1 second
Subframe 3 1 second
Subframe 4 1 second
No. 1
Each subframe consists of 64 words subframe 1 is shown
No. 38
First word of each subframe = SYNC word SYNC word of subframe No. 1 is shown
Each word No. 64 of 16 consecutive frames constitute a superframe
Bits No. 1 2 3 4 5 6 7 8 9 10 11 12
Bit value 1 1 1 0 0 0 1 0 0 1 0 0
Octal value 7 0 1 1
Sync words
Subframe | Label
1 | 1107
2 | 2670
3 | 5107
4 | 6670
Bit No. 1 2 3 4 5 6 7 8 9 10 11 12
Bit value 0 0 0 1 0 1 1 0 0 0 1 0
Decimal value 1 6 2
Example: BCD word 261
Any one of words 2 through 63 of each Subframe (word No. 38 is shown as example)
Bits No. 1 2 3 4 5 6 7 8 9 10 11 12
12-bit data field word
Data field
11-bit data field word
D Data field
10-bit data field word
D D Data field
Discrete word
D D D D D D D D D D D D
D-Discrete
Each subframe word consists of 12 Bits
Bit 1 = least significant Bit

Figure 18.8 Digital data formed into frames

zero to high independently of its value. A logic one is indicated by a mid-bit change of state; a logic zero is indicated by no mid-bit change of state.

The 64th word of 16 consecutive frames are combined into a super-frame. These words are formed in **binary coded decimal** (BCD) format; each word comprises four bits, used to represent the denary numbers zero to nine, see Table 18.6. This illustrates how binary numbers and then BCD represent the denary (or decimal) numbers 0–20.

The advantage of BCD is that it allows conversion to decimal digits for printing or display and faster decimal calculations. This is particularly useful where a numeric value is to be displayed, e.g. recording

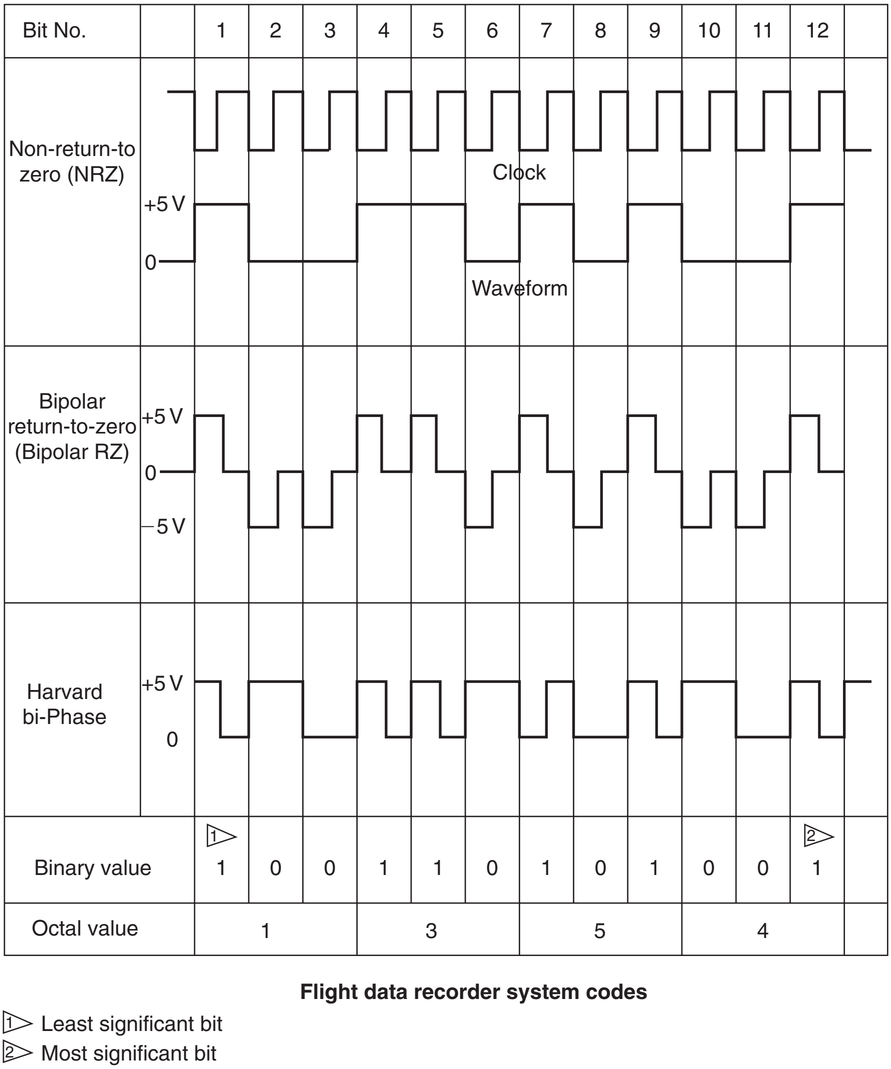
Bit No.
1
2
3
4
5
6
7
8
9
10
11
12
Non-return-to zero (NRZ)
Clock
+5 V
0
Waveform
Bipolar return-to-zero (Bipolar RZ)
+5 V
0
−5 V
Harvard bi-Phase
+5 V
0
Binary value
1 0 0 1 1 0 1 0 1 0 0 1
Octal value
1 3 5 4
Flight data recorder system codes
1 Least significant bit
2 Most significant bit

Figure 18.9 Digital bits of logic one and logic zero

systems consisting solely of digital logic. Manipulation of numerical data for the recording readout can be simplified by treating each digit as a separate single sub-circuit. This matches the physical reality of data displayed to the pilot and recorded data. The disadvantage of BCD is the complexity of circuits needed to implement mathematical operations; BCD is a relatively inefficient technique for encoding, i.e. it occupies more memory than a binary representation.

Test your understanding 18.4

How would the following denary numbers be represented in BCD: 27, 3, 1954?

The data transmitted by these super-frames includes frame count, error codes and voltage levels used in the calibration of analogue signals. The DFDR continuously verifies the data integrity from the tape; this

Table 18.6 Binary coded decimal (BCD)

Decimal	*Binary*	*Binary coded decimal*	
		A	*B*
0	00000	0000	0000
1	00001	0000	0001
2	00010	0000	0010
3	00011	0000	0011
4	00100	0000	0100
5	00101	0000	0101
6	00110	0000	0110
7	00111	0000	0111
8	01000	0000	1000
9	01001	0000	1001
10	01010	0001	0000
11	01011	0001	0001
12	01100	0001	0010
13	01101	0001	0011
14	01110	0001	0100
15	01111	0001	0101
16	10000	0001	0110
17	10001	0001	0111
18	10010	0001	1000
19	10011	0001	1001
20	10100	0010	0000

is accomplished by backing up the tape after a write operation and then reading the data.

A typical DFDR receives data from an acquisition unit in Harvard bi-phase code. This data is then processed and stored in solid-state **non-volatile memory** (NVM). The data is received at the interface card and is transferred to the power supply card. Arinc 573 data transmitters and receivers interface with a compression/storage acquisition processor; this processor stores the data in a buffer memory before sending it to a flash memory unit in the crash-survivable portion of the recorder (Fig. 18.10).

The compression/storage acquisition processor also monitors the system for faults via built in test routines. If any of these tests should fail, appropriate warning lights or LEDs are illuminated on the front of the unit.

18.4 Cockpit voice recorders

Commercial aircraft are required to carry a cockpit voice recorder (CVR). This unit captures and stores information derived from a number of the aircraft's audio channels that may later become invaluable in the event of a crash or serious incident. The CVR can provide valuable information that can later be analysed. The voice recorder preserves a continuing record of typically between 30 to 120 minutes of the most recent flight crew communications and conversations. The storage medium used with the CVR fitted to modern aircraft is usually based on one or more solid state memory devices whereas, on older aircraft, the CVR is usually based on a continuous loop of magnetic tape. The CVR storage unit must be recoverable in the event of an accident. This means that the storage media must be mounted in an enclosure that can withstand severe mechanical and thermal shock as well as the high pressure that exists when immersed in deep water.

The CVR control panel is usually fitted with a test switch, headphone jack and status light; the recorder itself is fitted with an underwater locator beacon (ULB) to facilitate undersea recovery. The ULB is a self-contained device (invariably attached to the front of the CVR) that emits an ultrasonic vibration (typically at 37.5 kHz) when the water-activated switch is immersed in water. A label on the ULB indicates the date by which the internal battery should be replaced. The specification for a typical ULB is shown in Table 18.7. An external view of a CVR showing its externally mounted ULB is shown in Fig. 18.11.

The audio input to the CVR is derived from the captain, first officer and observer position(s) and also from an open area microphone in the flight compartment which is usually mounted in the overhead panel and thus collects audio input from the entire flight-deck area.

In order to improve visibility and aid recovery, the external housing of the CVR is painted bright orange. The unit is thermally insulated and hermetically sealed to prevent the ingress of water. Because of the crucial nature of the data preserved by the flight, the unit should only be opened by authorized personnel following recovery from the aircraft.

Magnetic tape CVRs use a multi-track tape transport mechanism. This comprises a tape drive, four recording heads, a single (full-width) erase head, a monitor head and a bulk erase coil. The bias generator

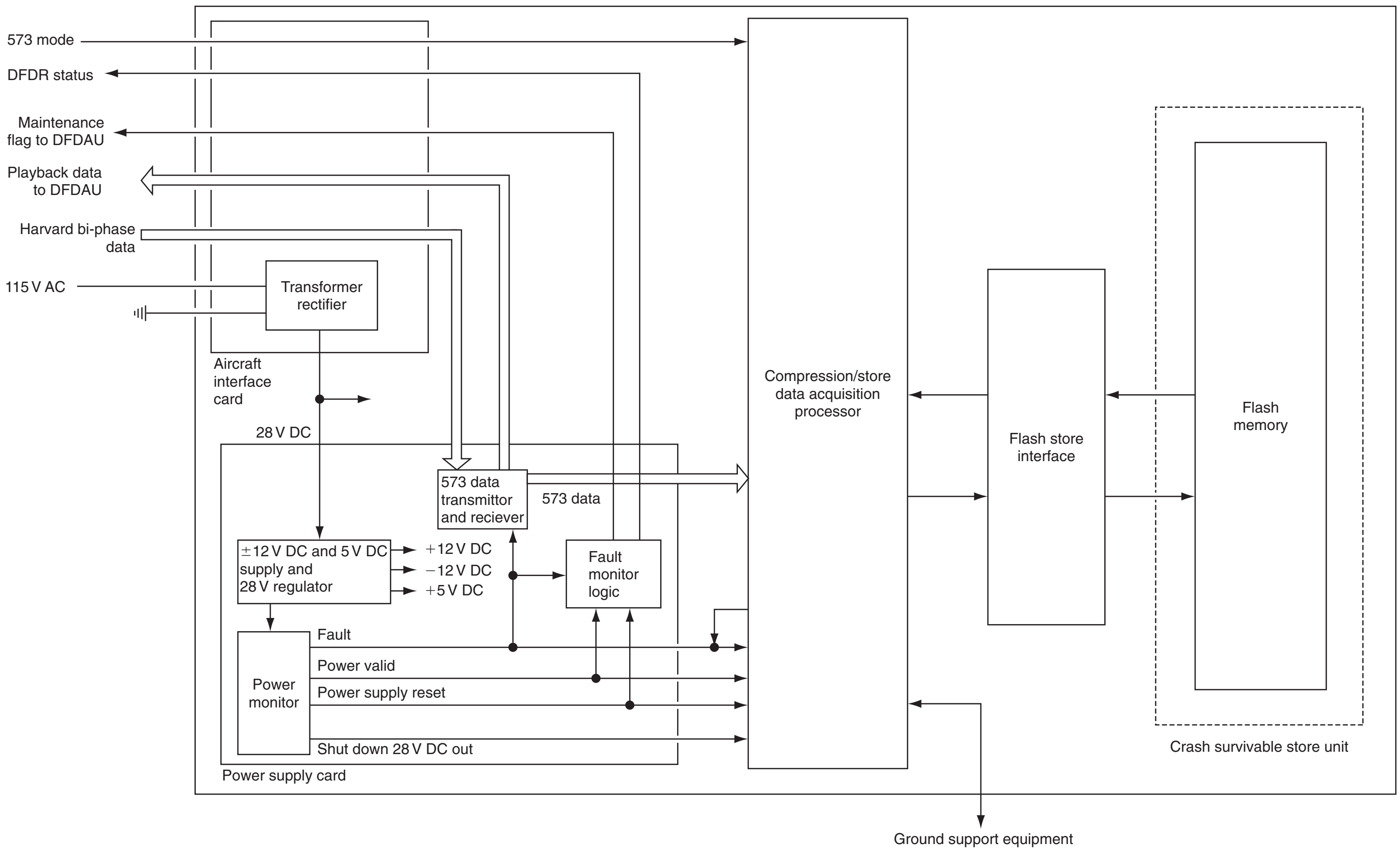
573 mode
DFDR status
Maintenance flag to DFDAU
Playback data to DFDAU
Harvard bi-phase data
115 V AC
Transformer rectifier
Aircraft interface card
28 V DC
573 data transmittor and reciever
573 data
±12 V DC and 5 V DC supply and 28 V regulator
+12 V DC
−12 V DC
+5 V DC
Fault monitor logic
Power monitor
Fault
Power valid
Power supply reset
Shut down 28 V DC out
Power supply card
Compression/store data acquisition processor
Flash store interface
Flash memory
Crash survivable store unit
Ground support equipment

Figure 18.10 Solid state recorder

Table 18.7 Typical ULB specification

Parameter	*Specification*
Operating frequency	37.5 kHz (±1 kHz)
Acoustic output	160 dB relative to 1 μPa at 1 m
Pulse repetition rate	0.9 pulses per second
Pulse duration	10 ms
Activation	Immersion in either salt water or fresh water
Power source	Internal lithium battery
Battery life	6 years standby (shelf-life)
Beacon operating life	30 days
Operating depth	20,000 ft (6096 m)
Housing material	Aluminium
Length	3.92 in. (9.95 cm)
Diameter	1.3 in. (3.3 cm)
Weight	6.7 oz (190 g)

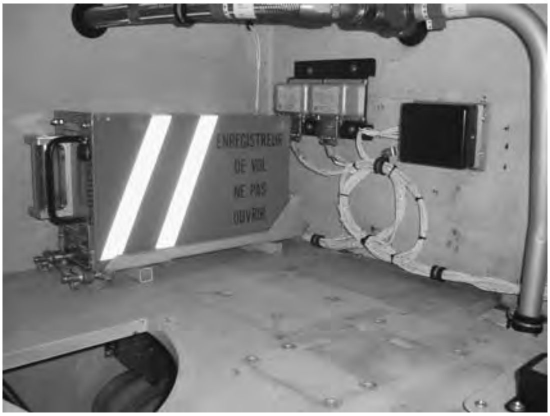

Figure 18.11 CVR/ULB installation

usually operates at around 65 kHz and an internal signal (at around 600 Hz) is provided for test purposes. **Bulk erase** can be performed by means of an erase switch; this is interlocked so that erasure can only be performed when the aircraft is on the ground and the parking brake is set. The erase current source is usually derived directly from the aircraft's 115 V AC 400 Hz supply. The magnetic tape (a continuous loop) is typically 308 feet in length and ¼ in. wide.

18.5 Health and usage monitoring system (HUMS)

HUMS was originally pioneered to monitor helicopters servicing the oil platforms in the North Sea. It monitors a variety of aircraft parameters, including aircraft vibration, engine and structural health. Ground support equipment and software are designed to decrease maintenance costs and increase operational readiness. The system includes on-board sensors and data processing to calculate specific maintenance recommendations. HUMS provides engine health and usage monitoring, exceedance monitoring, performance trends, operational usage monitoring and flight data monitoring and recording. Systems developed for helicopters now include rotor track and balance. All the required airborne data acquisition and processing, including crash-survivable cockpit voice recorders (CVR) and flight data recorders (FDR), are often combined into a single line replaceable unit.

18.6 Multiple choice questions

1. The flight data recorder must stop automatically:
   (a) after the aircraft is incapable of moving under its own power
   (b) before the aircraft is incapable of moving under its own power
   (c) after landing.

2. Parameters being recorded on the FDR are obtained from sources that are:
   (a) not displayed to the flight crew
   (b) independent of information displayed to the flight crew
   (c) accurately correlated with the information displayed to the flight crew.

3. Inputs form the audio system are recorded on the:
   (a) FDR
   (b) CVR
   (c) ULB.

4. The mandatory parameters required for an aircraft DFDR depend on the:
   (a) speed and weight of the aircraft
   (b) maximum weight of the aircraft
   (c) size of the aircraft and the prevailing regulatory rules applied to that aircraft.

5. The DFDR requires an electronic process such that:
   (a) digitally coded data can be converted back (decoded) into its analogue form
   (b) digitally coded data can be converted back (encoded) into its analogue form
   (c) analogue data can be converted back (decoded) into its digital form.

6. In binary coded decimal (BCD) format, each word comprises:
   (a) four bits, used to represent the denary numbers zero to nine
   (b) four bits, used to represent the binary numbers zero to nine
   (c) four bits, used to represent the denary numbers zero to ten.

7. The DFDR on large aircraft has to be able to retain the recorded data for a minimum of the last:
   (a) 30 minutes of its operation
   (b) 25 hours of its operation
   (c) 25 flights.

8. Lateral acceleration and radio altitude are typical parameters recorded on the:
   (a) FDR
   (b) CVR
   (c) ULB.

9. The FDAU collects, or acquires, a variety of:
   (a) analogue signals and converts them into a digital data stream for the recorder
   (b) analogue signals and outputs for the recorder
   (c) digital signals and converts them into an analogue data stream for the recorder.

10. The flight data recorder must start to record data automatically:
   (a) after the aircraft is capable of moving under its own power
   (b) after take-off
   (c) prior to the aircraft being capable of moving under its own power.

11. Referring to Fig. 18.12, the underwater locator beacon is activated when immersed in water:
   (a) after a time delay
   (b) immediately
   (c) at a specified depth.

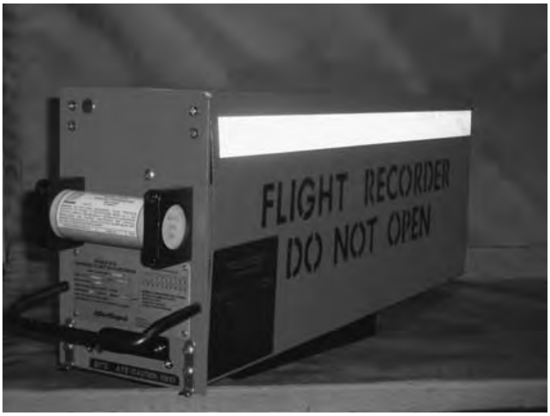

Figure 18.12 See Question 11

12. Referring to Fig. 8.13, data bits that change at the trailing edge are:
   (a) Harvard bi-phase
   (b) Non-return-to-zero
   (c) Return-to-zero.

13. Referring to Fig. 18.14, what is the purpose of EOT/BOT sensors?
   (a) convert digital data into a numerical display
   (b) convert analogue data in digital format
   (c) reverse the direction of tape travel and switch tracks.

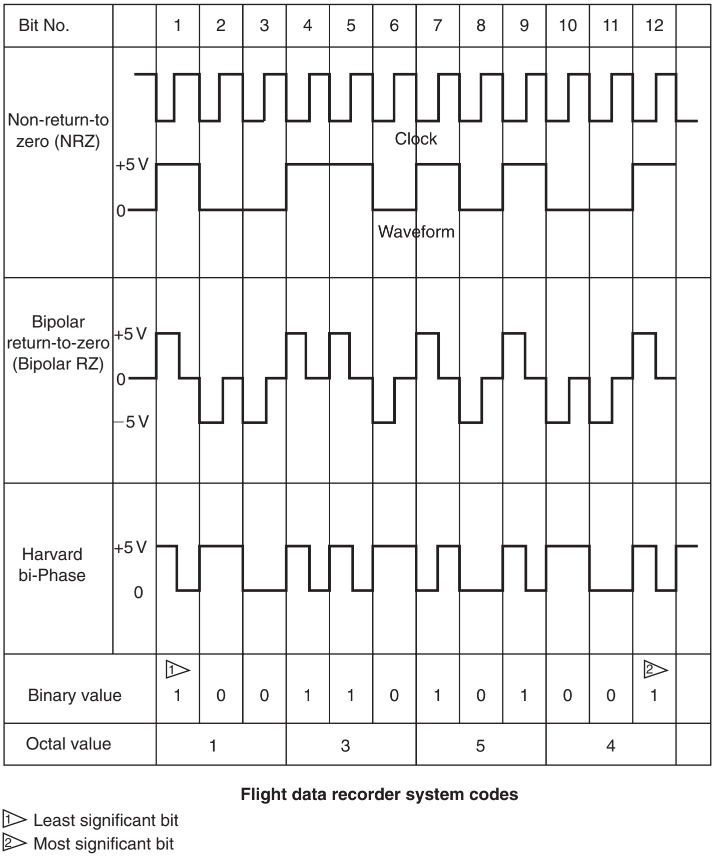
Bit No.
1
2
3
4
5
6
7
8
9
10
11
12
Non-return-to zero (NRZ)
Clock
+5 V
0
Waveform
Bipolar return-to-zero (Bipolar RZ)
+5 V
0
−5 V
Harvard bi-Phase
+5 V
0
Binary value
1
0
0
1
1
0
1
0
1
0
0
1
Octal value
1
3
5
4
Flight data recorder system codes
1 Least significant bit
2 Most significant bit

Figure 18.13 See Question 12

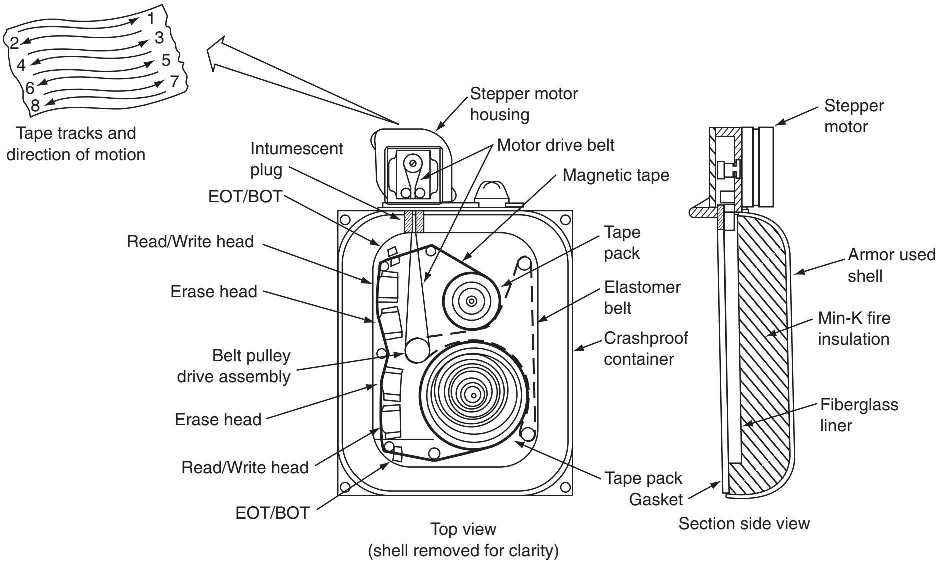
1
2
3
4
5
6
7
8
Tape tracks and direction of motion
Stepper motor housing
Motor drive belt
Intumescent plug
EOT/BOT
Magnetic tape
Read/Write head
Tape pack
Erase head
Elastomer belt
Belt pulley drive assembly
Crashproof container
Erase head
Read/Write head
Tape pack
Gasket
EOT/BOT
Top view
(shell removed for clarity)
Stepper motor
Armor used shell
Min-K fire insulation
Fiberglass liner
Section side view

Figure 18.14 See Question 13

Chapter 19

Electrical and magnetic fields

One of the consequences of operating electrical and electronic equipment is the possibility of disturbing, or interfering with, nearby items of electronic equipment. The term given to this type of disturbance is electromagnetic interference (EMI). Placing a portable radio receiver close to a computer and tuning through the radio's wavebands can illustrate this effect. The computer will radiate electromagnetic energy; this is received by the radio and heard as 'noise'. Radio equipment is designed to receive electromagnetic energy; however, the noise in this simple experiment is unwanted. Electrical or electronic products will both radiate and be susceptible to the effects of EMI. This is a paradox since many principles of electrical engineering are based on electromagnetic waves coupling with conductors to produce electrical energy and vice versa (generators and motors). Furthermore, systems are specifically designed to transmit and receive electromagnetic energy, i.e. radio equipment. The problem facing aircraft electrical and electronic systems is the unwanted noise; in the case of the computer/radio experiment, this unwanted noise is no more than a nuisance. In complex avionic systems, the consequences of EMI can be more serious. The ability of an item of equipment to operate alongside other items of equipment without causing EMI is electromagnetic compatibility (EMC).

Modern digital equipment operates at very high speed and relatively low power levels. In addition to EMI, high-intensity radiated fields (HIRF) are received from the external environment, e.g. from radio and radar transmitters, power lines and lightning. The high energy created by these radiated fields disrupts electronic components and systems in the aircraft. (This effect is also referred to as high-energy radiated fields – HERF.) The electromagnetic energy induces large currents to flow, causing direct damage to electronic components together with the secondary effects of EMI.

Advances in electronic technology bring many new features and benefits, e.g. faster processors, higher-density memory and highly efficient displays. These advances are primarily due to the reduction in the physical size of semiconductor junctions; this leads to higher-density components in given size of integrated circuit. One significant problem associated with the handling of semiconductor devices is that the smaller junctions are susceptible to damage from electrostatic voltages. This is a problem that can potentially affect a wide range of electronic equipment fitted in an aircraft. Effects range from weakening of semiconductor junctions through total failure of the equipment; both these effects can occur without any visible signs of damage to the naked eye! Electrostatic sensitive devices (ESSD) are electronic components that are prone to damage from stray electrical charge produced primarily from human operators. This problem is particularly prevalent with high-density memory devices and electronic displays. Weakening and damage to static sensitive devices can result from mishandling and inappropriate methods of storage.

This chapter looks at some of the practical aircraft maintenance aspects of EMI, EMC and HIRF together with background information and specific guidance on the correct handling and storage of ESSD.

19.1 Electromagnetic Interference

Electromagnetic interference (EMI) can be defined as the presence of unwanted voltages or currents that can adversely affect the performance of an electrical/electronic system. The effects of EMI include:

- errors in indications
- unwanted noise on audio signals
- random patterns on electronic displays
- repetitive 'buzzing' on intercom and cabin phone systems
- desensitizing of radio and radar receivers
- false indications in radar and navigation equipment
- nuisance triggering of alarms.

Consider the current-carrying conductor illustrated in Fig. 19.1. It can be seen that the field strength is

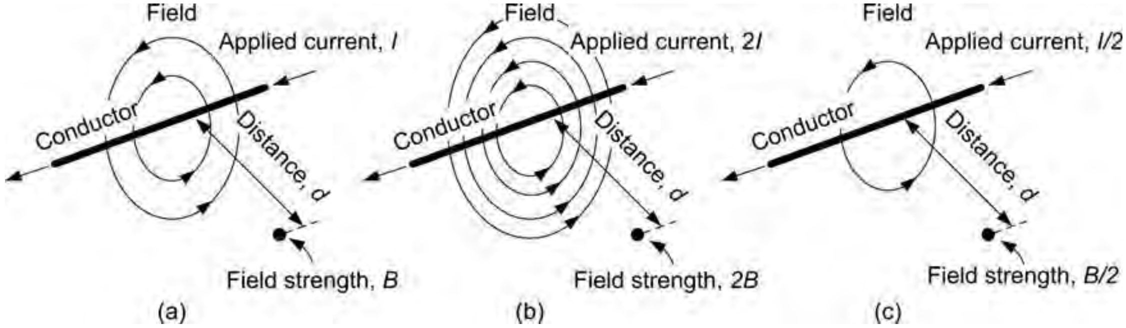
Field
Applied current, I
Conductor
Distance, d
Field strength, B
(a)
Field
Applied current, 2I
Conductor
Distance, d
Field strength, 2B
(b)
Field
Applied current, I/2
Conductor
Distance, d
Field strength, B/2
(c)

Figure 19.1 Current-carrying conductor

a measure of **flux density** at any given point. From first principles, the field strength *B* is proportional to the applied current and inversely proportional to the distance from the conductor. If an alternating current is applied to the conductor, this field will build up and collapse around the conductor. If a second conductor is now placed alongside the first conductor, the alternating field from the first conductor will induce currents in the second conductor. These induced currents will be superimposed onto any current flowing in the second conductor. The first conductor is **radiating** EMI; the second conductor is **susceptible** to EMI.

The total EMI effect depends primarily on the amount of current in the first conductor, rate of change of current (i.e. alternating, digital or switched currents) and the distance between the two conductors. Alternating currents, or digital signals being carried by a conductor, will set up alternating magnetic fields thus causing EMI as described. In summary, the amount of electromagnetic field radiated from a conductor depends on amount of current in the conductor and the rate of change of a magnetic field from the conductor.

Key maintenance point

EMI can also be created if a conductor carrying direct current is vibrated close to another conductor. The net result is relative movement of a magnetic field over a conductor with the consequences of induced currents. Wiring looms must be secured to prevent them vibrating excessively.

19.1.1 Shielding

For a given current, the distance between conductors and the provision of shielding (or screening) are the main considerations when trying to design and install wiring and equipment to minimize EMI. From Fig. 19.1 we can see that the distance between conductors has a direct effect on the amount of unwanted current induced into the second conductor. The addition of shielding on conductors limits the coupling of electromagnetic fields, see Fig. 19.2. The electromagnetic field created by current I_1 in the first conductor induces a current I_2 in the second conductor.

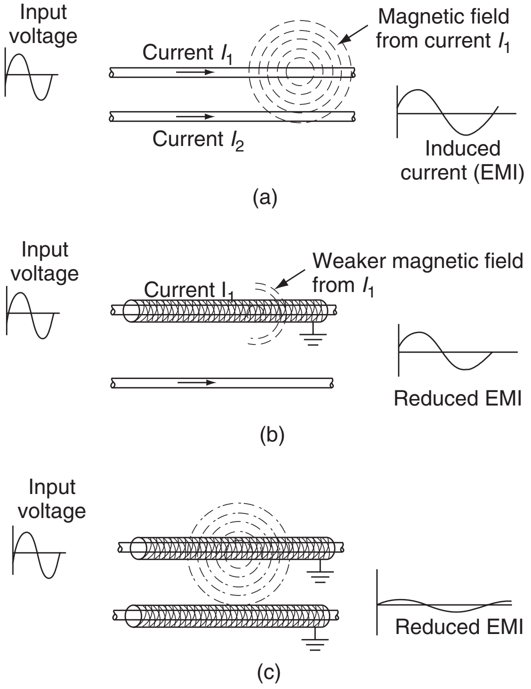
Input voltage
Current I_1
Magnetic field from current I_1
Current I_2
Induced current (EMI)
(a)
Input voltage
Current I_1
Weaker magnetic field from I_1
Reduced EMI
(b)
Input voltage
Reduced EMI
(c)

Figure 19.2 Shielding principles: (a) conductors unshielded, (b) one conductor shielded, (c) both conductors shielded

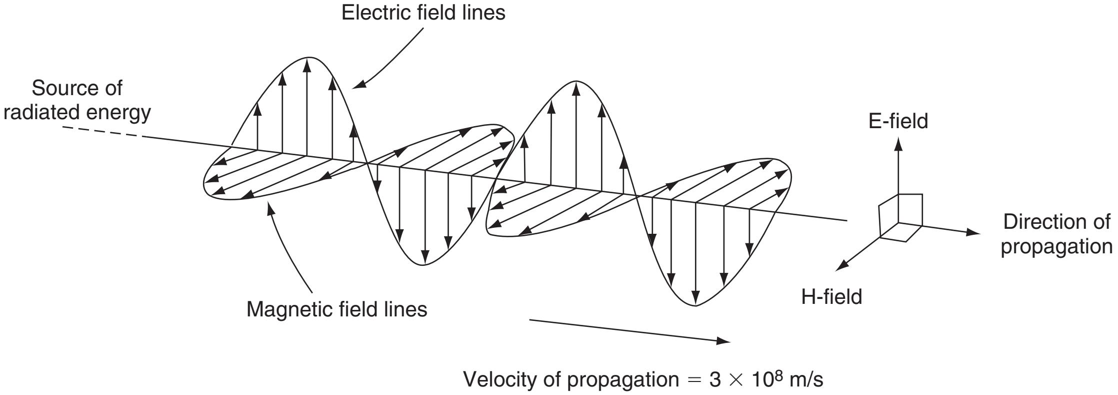

Figure 19.3 E and H lines mutually at right angles

With one conductor shielded, some of the electromagnetic field created by current I_1 induces current in the shielding; this current and taken away to a ground connection at one or both ends of the shielding. The shielding effectively absorbs some of the electromagnetic field created by I_1; the remaining (weaker) field induces a reduced current I_2 in the second conductor. If both conductors are shielded, most of the weaker field created by I_1 is absorbed into the shielding of the second conductor; this further reduces the current I_2 in the second conductor.

Shielding can reduce the coupling of electromagnetic fields between conductors. The amount of reduction depends upon the screening material used, its thickness, and the frequency of the applied current. Typical materials used for shielding are metallic mesh, or braiding formed on the outside of the conductor's protective sheath; this ensures that the wire or cable retains some degree of mechanical flexibility. The shielding impedes the radiation of signals from the first conductor, and also minimizes signals from being induced into the second conductor.

19.1.2 Electromagnet waves

We have been using the term electromagnet wave to develop the concept of EMI. The electromagnet wave actually comprises two components. As with light, electromagnet waves propagate outwards from a source of energy (transmitter) and comprise **electric** (*E*) and **magnetic** (*H*) fields at right angles to each other. These two components, the *E*-field and the *H*-field, are inseparable. The resulting wave travels away from the source with the *E* and *H* lines mutually at right angles to the direction of propagation, as shown in Fig. 19.3.

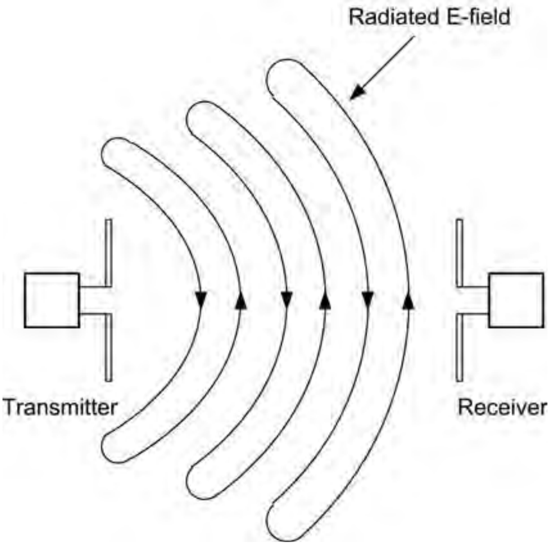

Figure 19.4 Radio transmitters and receivers

Electromagnet waves are said to be **polarized** in the plane of the electric (*E*) field. Thus, if the *E*-field is vertical, the signal is said to be vertically polarized whereas, if the *E*-field is horizontal, the signal is said to be horizontally polarized.

In the case of intentional propagation of electromagnet waves, i.e. radio transmitters and receivers (Fig. 19.4), the electric *E*-field lines are shown in the space between a transmitter and a receiver. The transmitter aerial (or antenna) is supplied with a high-frequency alternating current. This gives rise to an alternating electric field between the ends of the aerial/antenna and an alternating magnetic field around (and at right angles to) it. The direction of the *E*-field lines is reversed on each cycle of the signal as the wavefront moves outwards from the source.

The receiving aerial/antenna intercepts the moving field, and voltage and current is induced as a consequence. These voltages and currents are similar (but of smaller amplitude) to those produced by the transmitter. Note that in Fig. 19.4 (where the transmitter and receiver are close together) the radiated *E*-field is shown spreading out in a spherical pattern (this is known more correctly as the near field). The magnetic field (not shown) will be perpendicular to the *E*-field.

In practice, there will be some considerable distance between the transmitter and the receiver and so the wave that reaches the receiving aerial/antenna will have a plane wavefront. In this far field region, the angular field distribution is essentially independent of the distance from the transmitting antenna. (Radio wave theory is addressed in more detail in a related book in the series, *Aircraft Communications and Navigation Systems*.)

A simple wiring installation connects the shielding to ground, thereby 'soaking' away the unwanted currents, rather than inducing them into the second conductor. Although shielding provides some protection against EMI, it also adds cost and weight to the installation. Furthermore, the currents being carried away in the shielding radiate their own fields that can cause secondary EMI. This unwanted current is referred to as a **ground loop**; current flows in a conductor connecting two points that are intended to be at the same potential, e.g. ground potential. (The fact that current is flowing can only occur because they are actually at different potentials.) As with many engineering situations, solving one problem often comes at a cost together with the introduction of new problems!

Test your understanding 19.1

Explain the difference between EMI and EMC.

19.1.3 Twisted pair

Another technique used to minimize EMI in wiring is the **twisted pair**. This a form of wiring in which the two conductors are wound together to cancel out electromagnetic interference (EMI) from external sources, and minimize **cross-talk** between adjacent pairs of wires. Cross-talk is the consequence of EMI between one electrical circuit to another, i.e. when a signal transmitted in one circuit creates an undesired effect in another circuit. A twisted pair cable consists of two independently insulated wires twisted around one another.

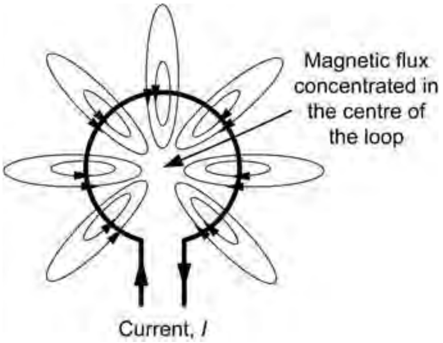

Figure 19.5 Magnetic field

Twisting the cables forms repetitive loops; each twist reverses the polarity of the loop, therefore magnetic fields can only couple into each loop and not the entire cable length. When current is supplied through a loop, a magnetic field is set up; it can be seen that the flux is concentrated in the centre of the loop as illustrated in Fig. 19.5. The twisted pair also provides a physical means of minimizing EMI in wires carrying digital signals; each wire is positioned alternatively next to the source of interference; the net effect is to cancel out any differential between the wires.

Twisting a pair of wires to form a cable is an extremely effective way of transmitting high-speed signals because:

- most of the electrical noise entering into and/or radiating from the cable can be eliminated
- cross-talk (signals 'leaking' between wires in the cable) is minimized.

In addition to the electrical energy flowing down each wire, energy can also be coupled between wires due to electrostatic and magnetic effects. For the electrostatic effects, the insulation between the two conductors is the dielectric of a capacitor. More surface area is created by longer cables; this leads to increased inter-wire capacitance. Higher frequency signals can lead to cross-talk between wires as a result of reduced capacitance.

Typical applications of signals transmitted down a twisted-pair of wires include Arinc 429 signals as

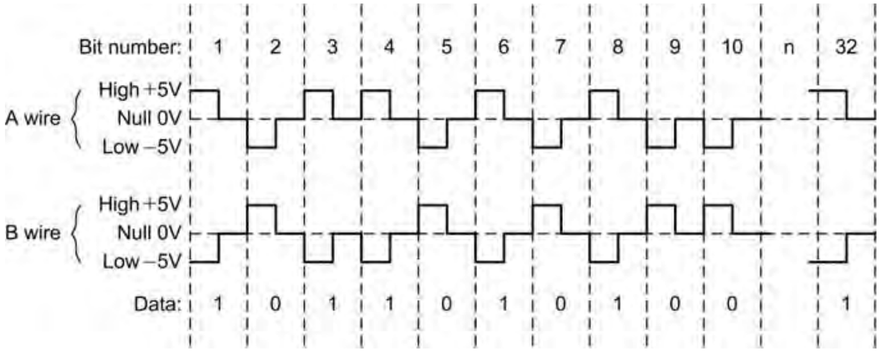
Bit number: 1 2 3 4 5 6 7 8 9 10 n 32
A wire
High +5V
Null 0V
Low −5V
B wire
High +5V
Null 0V
Low −5V
Data: 1 0 1 1 0 1 0 1 0 0 1

Figure 19.6 Digital signal

illustrated in Fig. 19.6. Any other system wires adjacent to this pair will be affected by cross-talk.

Two wires (A and B) each carry a ±5 volt digital signal arranged as bipolar return to zero (BPRZ). The extent of the cross-talk is equal to the sum of the digital signals; if this sum is zero (or nearly zero) then the affects of cross-talk are eliminated.

Consider three wires in a cable as illustrated in Fig. 19.7. Wires A and B are formed as a twisted pair carrying a digital signal. If the signal sent through wire A is ±10 volts with respect to wire B (a reference voltage of zero) then wire C wire picks up cross-talk noise. If the digital signals are arranged as per the second illustration, the opposite polarity signal of ±5 volts on wires A and B cancel each other out, the cross-talk effect wire C is eliminated.

Key maintenance point

Twisting wires together also provides a neater installation; the number of twists per given length should be specified in order to achieve the desired EMI protection.

Key maintenance point

In some installations, further EMI protection is given by the addition of shielding over the twisted pair of wires.

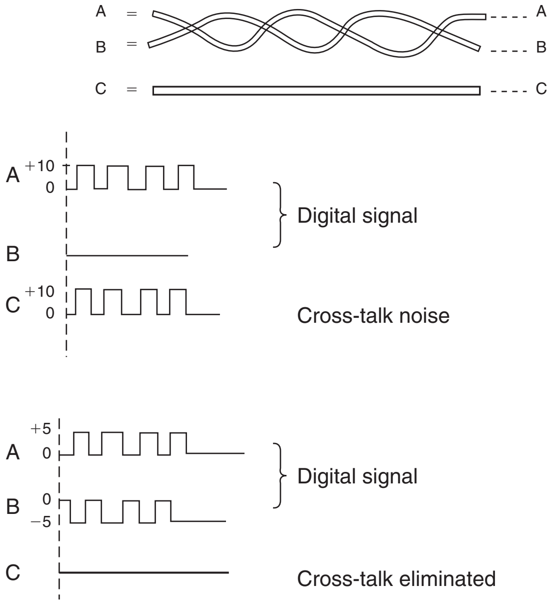
A =
B =
C =
A
B
C
A +10
0
B
C +10
0
Digital signal
Cross-talk noise
A +5
0
B 0
−5
C
Digital signal
Cross-talk eliminated

Figure 19.7 Twisted pairs and cross talk

19.1.4 Bandwidth

Bandwidth is the difference between the upper and lower cut-off frequencies of analogue amplifying circuits; the unit of **bandwidth** is hertz (Hz). The characteristics of bandwidth is illustrated in Fig. 19.8. This is a central concept in electronics, radio-frequency (RF) systems and signal processing. It also has to be considered in the context of EMI; high frequencies should be filtered out of a circuit wherever possible, without compromising its functionality.

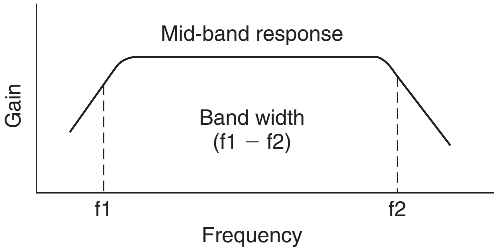

Figure 19.8 Bandwidth

In computers, digital bandwidth refers to the rate at which data is transmitted/received; this is measured in **bits per second (BPS)**. A digital communication network has a given bandwidth in terms of its overall channel capacity and throughput (consumption). Channel capacity (in BPS) is proportional to the analogue bandwidth in hertz (Hz); this is the maximum amount of error-free digital data that can be transmitted via a communication link with a specified bandwidth in the presence of EMI. The deeper theory associated with this subject is beyond the scope of this book; students wishing to research this subject in more detail should refer to the **Shannon–Hartley theorem**.

Test your understanding 19.2

Why do higher-frequency signals lead to increased cross-talk?

Test your understanding 19.3

How does twisting pairs of wiring reduce cross-talk?

19.1.5 Radiated EMI

There are many sources of EMI throughout the aircraft. Those sources known to **radiate** EMI include:

- fluorescent lights
- radio and radar transmitters
- power lines
- AC powered window heat controllers
- motors/generators
- switching and light dimming circuits
- microprocessors
- pulsed high-frequency circuits
- data bus cables (but not fibre optic cables)
- static discharge and lightning.

The energy generated by these sources is radiated as an electromagnetic field. From first principles, we know that the coupling only takes place when there is a relative movement of electromagnet field and conductor; digital circuits are (by definition) switching currents on/off with fast and short pulse rise/fall times. Unless adequate precautions are taken to eliminate the interference at source and/or to reduce the equipment's radiation of EMI, the energy can then become coupled into other circuits. In electromagnetic field radiation, energy is transmitted through electrically nonconductive paths, such as air, plastic materials, or fibreglass.

19.1.6 EMI susceptibility

There are many systems on the aircraft that which may be **susceptible** to electromagnetic interference. These include:

- radio and radar receivers
- microprocessors and other microelectronic systems
- electronic instruments
- control systems
- audio and in-flight entertainment systems (IFE).

Whether a system will have an adverse response to **electromagnetic interference** depends on the type and amount of emitted energy in conjunction with the susceptibility threshold of the receiving system. The threshold of susceptibility is the minimum interference signal level (conducted or radiated) that results in equipment performance that is indistinguishable from the normal operation. If the threshold is exceeded then the performance of the equipment will become degraded. Note that, when the susceptibility threshold level is greater than the levels of radiated emissions, electromagnetic interference problems do not exist. Systems to which this applies have **electromagnetic compatibility** (EMC). In other words, the systems will operate as intended and any EMI generated is at such a level that it does not affect normal operation.

19.2 EMI reduction

Planning for electromagnetic compatibility must be initiated in the design phase of a device or system.

If this is not satisfactorily addressed, interference problems may arise. The three factors necessary to produce an EMI problem are:

- source(s) of interference (sometimes called noise)
- a means of coupling (by conduction or radiation)
- susceptible components or circuits.

To reduce the effects of EMI, or **electrical noise**, at least one of these factors must be addressed. The following lists some techniques used for EMI reduction to tackle these three factors (note that some techniques address more than one factor).

1. Suppress the interference at source
    - Enclose the interference source(s) in a screened metal enclosure and then ensure that the enclosure is adequately grounded
    - Use transient suppression on relays, switches and contactors
    - Twist and/or shield bus wires and data bus connections
    - Use screened (i.e. coaxial) cables for audio and radio-frequency signals
    - Keep pulse rise times as slow and long as possible
    - Check that enclosures, racks and other supporting structures are grounded effectively.
2. Reduce noise coupling
    - Separate power leads from interconnecting signal wires
    - Twist and/or shield noisy wires and data bus connections
    - Use screened (i.e. coaxial) cables for audio and radio-frequency signals
    - Keep ground leads as short as possible
    - Pay close attention to potential ground loops
    - Filter noisy output leads
    - Physically relocate receivers and sensitive equipment away from interference sources.
3. Increase the susceptibility thresholds
    - Limit the bandwidth of circuits wherever possible
    - Limit the gain and sensitivity of circuits wherever possible
    - Ensure that enclosures are grounded and that internal screens are fitted
    - Fit components that are inherently less susceptible to the effect of stray radiated fields.

19.3 High-intensity/energy radiated fields

High-intensity/energy radiated fields (HIRF/HERF) are generated by certain radio-frequency (RF) sources that are **external** to the aircraft. These fields disrupt electronic components and systems within the aircraft via currents that are induced from these fields into the aircraft's wiring. (The terms HIRF and HERF mean the same thing; HIRF will be used for the remainder of this chapter.) Radio, radar and television transmitters (particularly highly directional broadcasts) have the capability of adversely affecting the operation of aircraft electrical and electronic systems. The HIRF environment has become a significant threat to aircraft using the electronic systems described in this book, together with the communications, navigation and flight guidance systems. These systems are potentially very susceptible to the HIRF environment. Accidents and incidents on aircraft with such systems have led to the need for a thorough understanding of, and increased protection from, high-intensity radiated fields.

Practical experience with the effects of HIRF often result in unexplained and unrelated disruption to aircraft systems e.g. simultaneous navigation errors and erroneous engine indications. These systems then start operating normally again without being attributed to any specific equipment failures. HIRF has been cited as the cause of misleading roll and pitch information on electronic displays and the total loss of engine power due to interference with electronic control systems. The need for protection of modern electrical and electronic systems from HIRF is required because of the:

- dependence on these systems used for the continued safe flight and landing of the aircraft
- increased use of composite materials (reducing the natural Faraday cage protection of metallic structures)
- Increased complexity of digital systems (faster operating speeds, higher-density integrated circuits)
- expanded frequency usage of microwave energy
- increased quantity and power of transmitters
- proliferation of RF transmitters.

19.3.1 HIRF environment

The **HIRF environment** is created by the transmission of high-power radio-frequency (RF) energy into free space. These transmissions can be from military systems, television, radio, radars and satellites communicating with ground-based equipment, ships or other aircraft. When an aircraft operates in a HIRF environment, this can have an adverse effect on the aircraft systems and equipment that could result in system failure, malfunction, or misleading information. The process whereby electromagnetic energy from an RF source is induced in a system by radiation is termed **coupling**. It is entirely possible that HIRF will affect individual components through to system level via individual wires (or wire bundles) that connect the items of equipment. In the event of the HIRF environment having an adverse effect on systems, it is essential that the aircraft has the capability for continued controlled flight and landing to a suitable location, albeit under emergency conditions.

19.3.2 HIRF characteristics

The analysis of HIRF is centred on the **frequency** of transmission and **field strength**. The practical considerations of RF transmissions are from approximately 10 kilohertz (10 kHz) through to 100 gigahertz (100 GHz). Field strength is defined as the magnitude of the electromagnetic energy propagating in free space expressed in volts per meter (V/m). Aircraft systems need to be tested and/or analysed across a range of frequencies and field strengths to determine their susceptibility characteristics. Certain systems (or individual items of equipment) are immune to HIRF and have the ability to continue to perform their intended function. This could occur as an inherent or system design feature of the equipment, e.g. if it is located behind a material (**reflection plane**) that reflects RF signals.

Alternatively, a decrease in electromagnetic field strength in transmission from one point to another can occur by **attenuation**, expressed in decibels (dB). Attenuation is the scalar ratio of the RF energy input magnitude and output magnitude. (Further details of attenuation are given in the appendices.). In the HIRF environment, the ratio of current induced in a wire bundle form the external HIRF field strength (as a function of frequency) is termed the **transfer function**. An item of equipment or system that is susceptible to the HIRF environment means that it is unable to perform its intended function; this will be defined by a susceptibility level, where the effects of interference from HIRF become apparent.

Figure 19.9 Electrical storm

19.4 Lightning

Lightning together with precipitation is associated with electrical activity within **cumulonimbus** clouds, see Fig. 19.9. Lightning results from the build up of huge amounts of static charge in the atmosphere. Precipitation may defined as the result of water vapour **condensing** in the atmosphere that subsequently falls to the earth's surface.

Electrical activity can originate at the top of thunderclouds or the outside edges of the precipitation area. If an aircraft is subjected to a **lightning strike**, or discharge, the structure and bonding are designed to dissipate this energy, see Fig. 19.10.

Key maintenance point

The probability of an aircraft experiencing a lightning strike on any given flight is one. High standards of design and maintenance are required to ensure that serious damage from a lightning strike is a rare occurrence.

19.4.1 Faraday cage

The metallic aircraft structure acts as a **Faraday cage (or Faraday shield)**, named after the British physicist Michael Faraday. This is an enclosure formed by a conducting material that shields the inside of the structure from electromagnetic effects. An external electrical field will cause the electrical charges

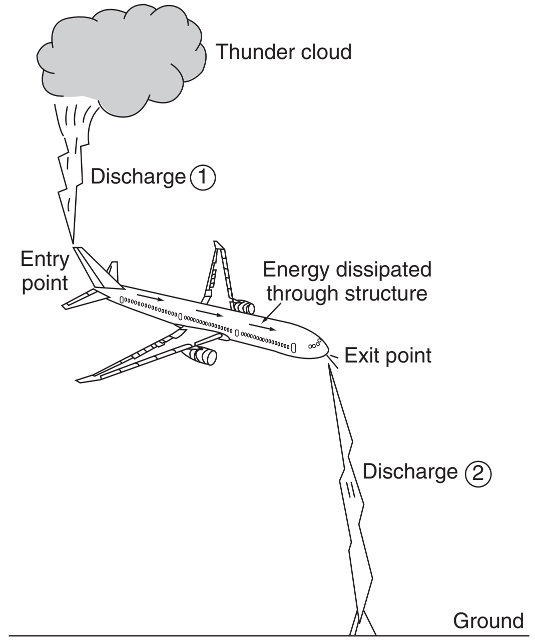

Figure 19.10 Lightning strike; dissipation of energy through the aircraft

within the structure to redistribute themselves so as to cancel the field's effects inside the fuselage. This effect is used very effectively in aircraft to protect electronic equipment from lightning strikes and other electrostatic discharges. Lightning normally enters the aircraft at an extremity, e.g. the nose cone or wing tip; currents flow through the conductive structure, and then exits the aircraft at another extremity. This can cause short-term interference with systems, but there should be no permanent damage. Some currents could enter the structure due to the high voltages, the energy will normally be conducted through bonding leads and back into the fuselage skin. The aircraft is categorized into specific areas when planning system and equipment locations, see Fig. 19.11:

- zone one is where lightning can be expected to enter and exit the aircraft
- zones two and three provide the conductive paths through the aircraft.

Key maintenance point

The majority of physical damage on the aircraft occurs at the exit point of the lightning strike.

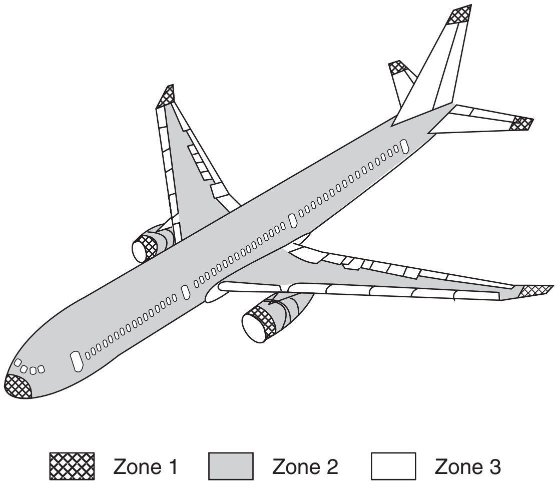

Figure 19.11 Lightning zones

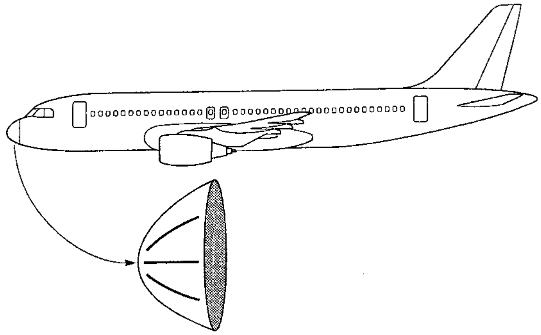

Figure 19.12 Radome protection

19.4.2 Aircraft construction

The metal structure dissipates electric currents generated from the external electromagnetic fields, thereby reducing or even eliminating electromagnetic interference. More aircraft are now being built using (non-conducting) **composite materials**, ranging from individual items of structure through to the complete exterior. In this case, the Faraday cage effect has to be designed into the composite structure as a mesh of conducting material. Additional design effort, testing and on-going maintenance are required as a consequence. One particular part of the aircraft often affected by lighting is the nose cone; this is made from a composite material and therefore has no inherent conductive paths. Lightning strips can be fitted to protect this area, see Fig. 19.12.

Key maintenance point

The aircraft is designed to provide conductive paths to dissipate the energy from lightning strikes. Over periods of time, certain parts of the aircraft will corrode and be subjected to vibration; it is essential that these conducive paths are maintained, e.g. by inspection of bonding leads.

Even when the lightning discharge is dissipated safely through the aircraft, structural damage can occur at the exit points of the discharge. Furthermore, the high circulating currents can causes EMI as previously discussed. Lightning discharges via the atmosphere from one cloud to another, or directly to the earth's surface. The atmosphere has a very high dielectric breakdown (depending on moisture content), typically three million volts per meter. The voltage generated by lightning depends on the length travelled by the lightning discharge; this will be approximately one gigavolt (10^9 volts) for a 300 m (1000 ft) lightning bolt. With a typical current of 100 kA, this gives rise to a power dissipation of 100 terawatts (100×10^{12} watts). The high energy radiated by a lightning discharge carries sufficient electromagnetic field strength to couple into the aircraft's wiring. This causes high currents of relatively short duration to be **injected** into electronic equipment. Adequate protection must be designed into the equipment to prevent disruption of signal processing and/or damage to components.

19.4.3 Certification of aircraft for HIRF and lightning protection

Although the sources of HIRF and lightning (HIRF/L) threats are very different, the strategies used in design, certification and maintenance are very similar, if not the same. The external HIRF and lightning (HIRF/L) environment has evolved over time and varies between aircraft types. Various specifications defining HIRF/L threat levels have emerged in response to the increasing understanding of the problem. In Europe, aircraft have been required to comply with HIRF requirements since early 1992. In-service aircraft are therefore specified with varying HIRF standards ranging from no requirement through to the current standards. HIRF threat levels can be broadly thought of in terms of field strength and frequency of the RF signal. Compliance with HIRF requirements typically involves a range of factors including:

- testing the entire aircraft
- equipment tests
- analogy/similarity with other installations.

Equipment and systems being certified for HIRF/L immunity and susceptibility are in fact achieved by any combination of the above. Comprehensive test programmes are required on new aircraft types; this is carried out on entire aircraft, with engines and systems operating. Various techniques are used for HIRF/L testing:

- **bulk current injection** (BCI) via a probe that injects RF signals directly into wire bundles when clamped around them.
- **direct drive testing** (DDT) involves connecting an RF signal source directly to the unit being tested.

Testing is an expensive and time-consuming activity, one that has to be planned and agreed with the certifying authorities well in advance of the aircraft going into service. An alternative strategy is to use existing HIRF compliance documentation and data from a similar system (or aircraft) to demonstrate HIRF compliance. Failure conditions resulting from HIRF/L are assessed against certification requirements and are given classifications, i.e. catastrophic, hazardous or major. The classifications for HIRF/L related system failures are assigned certification levels of:

A. **Catastrophic**: where system failure(s) would prevent the continued safe flight and landing of the aircraft
B. **Hazardous**: where system failure would significantly reduce the capability of the aircraft and/or the ability of the flight crew to respond to the failure(s)
C. **Major**: where system failure reduces the capability of the aircraft or the ability of the flight crew to respond to an adverse operating condition(s).

A safety assessment is prepared to address all potential adverse effects. The safety assessment could establish that some systems have different failure

Table 19.1 External HIRF environments

Environment	*Frequency range*	*Field strength (V/m) (peak)*	*Field strength (V/m) (average)*
I	10 kHz–40 GHz	3000	200
II	10 kHz–40 GHz	3000	230
III	10 kHz–40 GHz	7200	490

conditions in different phases of flight, e.g. a navigation system may have a catastrophic failure condition during the landing phase, but in cruise the classification could be reduced to a major failure condition. To cater for these different classifications on the same system, varying HIRF environments would be applied, see Table 19.1.

The field strength values shown in Table 19.1 are the maximum for any given frequency within the range. Field strength values are expressed in root-mean-square (RMS) units, measured during the peak of the modulation cycle.

19.4.4 Maintenance for HIRF/L protection

The maintenance engineer has a number of essential tasks to ensure that HIRF/L protection features and devices are serviceable for continued airworthiness of the aircraft. These tasks include, but are not limited to:

- bonding resistance measurements
- wire/cable resistance or impedance measurements
- inspection and disassembly of connectors to detect corrosion or termination failure.

19.4.5 Aircraft wiring and cabling

When many potential sources of EMI are present in a confined space, aircraft wiring and cabling has a crucial role to play in maintaining electromagnetic compatibility. Adequate wire separation should be maintained between noise source wiring and susceptible equipment. For example, radio-frequency (RF) navigation systems' wiring should be strategically routed in the aircraft to ensure a high level of EMC. Any changes to the routing of RF systems' wiring could have an adverse affect on EMC. Separation requirements for all wire categories must be maintained; wire lengths should be kept as short as possible to maintain coupling at a minimum. Where wire shielding is incorporated to protect against EMI caused by lightning, it is important that the shield grounds (pigtails) be kept to their designed length. An inch or two added to the length will result in degraded lightning protection. Equipment grounds must not be lengthened beyond design specification. A circuit ground with too much impedance may no longer be a true ground.

Key maintenance point

With the aid of the aircraft technical manuals, grounding and bonding integrity must be maintained. This includes proper preparation of the surfaces where electrical bonding is made.

19.5 Grounding and bonding

The aircraft structure provides an extremely important means of reducing EMI and also protecting the aircraft, its passengers, crew and systems, from the effects of lightning strikes and static discharge. Grounding and bonding are specific techniques that are used to achieve electrical integrity, see Fig. 19.13.

Grounding and bonding can also be instrumental in minimizing the effects of high-intensity radiated fields (HIRF) emanating from high-power radio transmitters and radar equipment. Grounding and bonding resistances of less than 0.001 ohms to 0.003 ohms are usually required.

Grounding can be defined as the technique of electrically connecting components to either a conductive structure or another return path for the purpose of completing a circuit. **Bonding** refers to the electrical connecting of two or more conducting objects that are not otherwise adequately connected. Bonding and grounding connections are made in an aircraft in order to accomplish the following:

- dissipate energy from a high-intensity radiated fields (HIRF) and lightning strikes
- dissipate static electricity
- limit the potential difference between equipment
- provide a low resistance path for earth return systems.

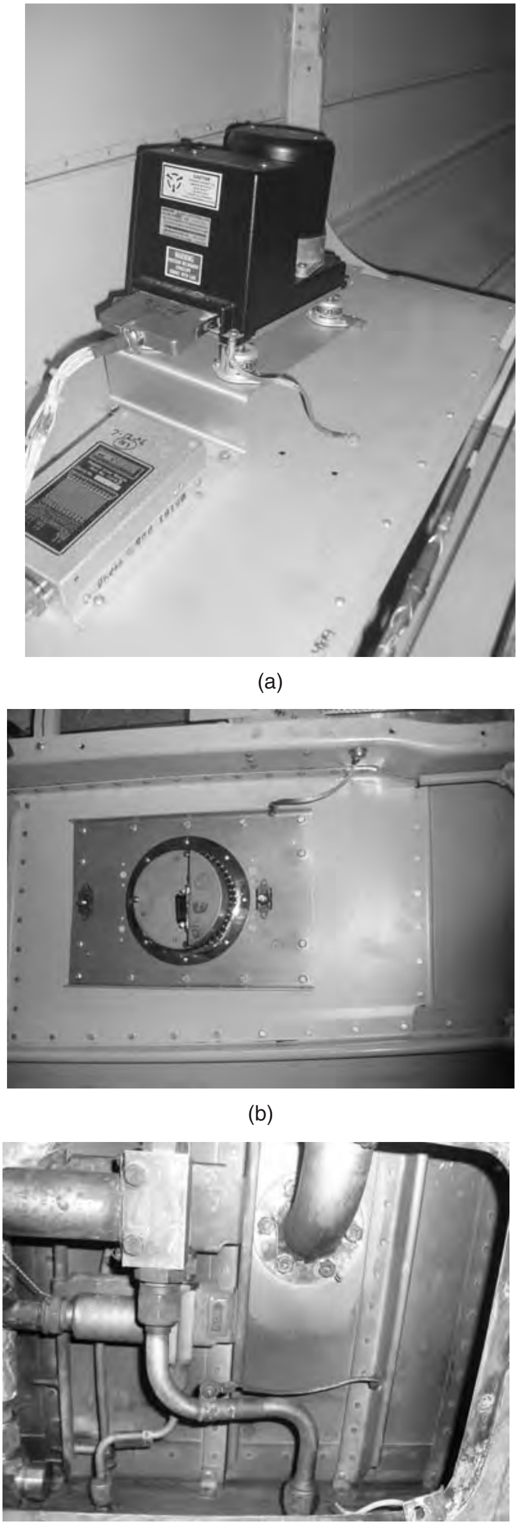

(a)

(b)

(c)

Figure 19.13 Examples of bonding

The following general procedures and precautions apply when making bonding or grounding connections:

- bond or ground parts to the primary aircraft structure where possible
- make bonding or grounding connections so that no part of the aircraft structure is weakened
- bond parts individually if feasible
- install bonding or grounding connections against smooth, clean surfaces
- install bonding or grounding connections so that vibration, expansion or contraction, or relative movement in normal service, will not break or loosen the connection
- check the integrity and effectiveness of a bonded or grounded connection using an approved bonding tester.

The main types of bonding are:

Equipment bonding. Low impedance paths to the aircraft structure are generally required for electronic equipment to provide radio-frequency return circuits and to facilitate reduction in EMI.

Metallic surface bonding. All conducting objects located on the exterior of the airframe should be electrically connected to the airframe through mechanical joints, conductive hinges, or bond straps, which are capable of conducting static charges and currents induced by lightning strikes.

Static bonds. All isolated conducting paths inside and outside the aircraft with an area greater than 3 in.^{2} and a linear dimension over 3 inches that are subjected to electrostatic charging should have a mechanically secure electrical connection to the aircraft structure of adequate conductivity to dissipate possible static charges.

19.6 Electrostatic sensitive devices (ESSD)

Advances in electronic technology bring many new features and benefits, e.g. faster processors, higher-density memory and highly efficient displays. These advances are primarily due to the reduction in the physical size of semiconductor junctions; this leads to higher-density components in given size of integrated circuit. One significant problem associated with the handling of semiconductor devices is that the smaller junctions are susceptible to damage from

electrostatic voltages. An example of static electricity that the reader might have encountered is the electric shock received when stepping out of a car. The synthetic materials used for clothing as well as the vehicle's interior are capable of producing large amounts of static charge which is only released when the driver or passenger sets foot on the ground!

19.6.1 Triboelectric effect

When two dissimilar non-conducting materials are rubbed together, the friction transfers electrical charge from one material to the other. This raises the electrical potential between the materials, and is known as the **triboelectric effect**. The build-up of charge, and subsequent attraction of materials, can be observed when clingfilm is separated from its roll. The polarity and strength of the electrical charges depend primarily on the materials, surface finish, ambient temperature and humidity. The triboelectric series classifies different materials according to how readily they create static electricity when rubbed with another material. The series is arranged on a **triboelectric scale** of increasingly positive and increasingly negative materials.

Materials that give up electrons and become positive when charged (thereby appearing as positive on the triboelectric scale) include glass, air and dry human skin. Materials that attract electrons become negatively charged (appearing as negative on the triboelectric scale) include:

- polyester
- polystyrene
- polyethylene
- poly(vinyl chloride) (PVC).

Certain materials do not readily attract or give up electrons when brought in contact or rubbed with other materials are neutral on the triboelectric scale; examples include cardboard, cotton and steel.

The largest amounts of static charge resulting from materials being rubbed together (or separated from each other in the case of cling film) are at the extreme ends of the triboelectric scale. For example, PVC rubbed against glass or polyester rubbed against dry human skin both produce an accumulation of charge. A common occurrence when working in a dry atmosphere is that people rapidly discharge current (sometimes producing a spark) when touching metal objects. This is because they have relatively dry skin (which can become highly positive in charge), and is accentuated when the clothes they wear are made of man-made material (such as polyester), which will acquire a negative charge. The effect is much less pronounced in a humid atmosphere where the stray charge can 'leak away' harmlessly into the atmosphere. (Note that moist skin tends to dissipate charge more readily.) People that build up static charges due to dry skin are advised to wear all-cotton clothes since cotton is neutral on the triboelectric scale.

Table 19.2 Representative values of electrostatic voltages generated in typical work situations

Situation	*20% relative humidity*	*80% relative humidity*
Walking over a wool/nylon carpet	35 kV	1.5 kV
Sliding a plastic box across a carpet	18 kV	1.2 kV
Removing parts from a polystyrene bag	15 kV	1 kV
Walking over vinyl flooring	11 kV	350 V
Removing shrink wrap packaging	10 kV	250 V
Working at a bench wearing overalls	8 kV	150 V

19.6.2 Working environment

The problem for the electronics industry is that high voltages can be accumulated in people and materials and then discharged into equipment. This discharge can weaken and/or damage electronic components. Representative values of electrostatic voltages generated in some typical working situations are shown in Table 19.2. (Note the significant difference in voltage generated at different values of relative humidity.)

Static voltage susceptibility varies for different types of semiconductor device or components, for example static discharge voltages as low as 20 V to 100 V can affect complex devices such as microprocessors. Single components such as silicon-controlled rectifiers are not affected until levels of between 4 kV to 15 kV are reached.

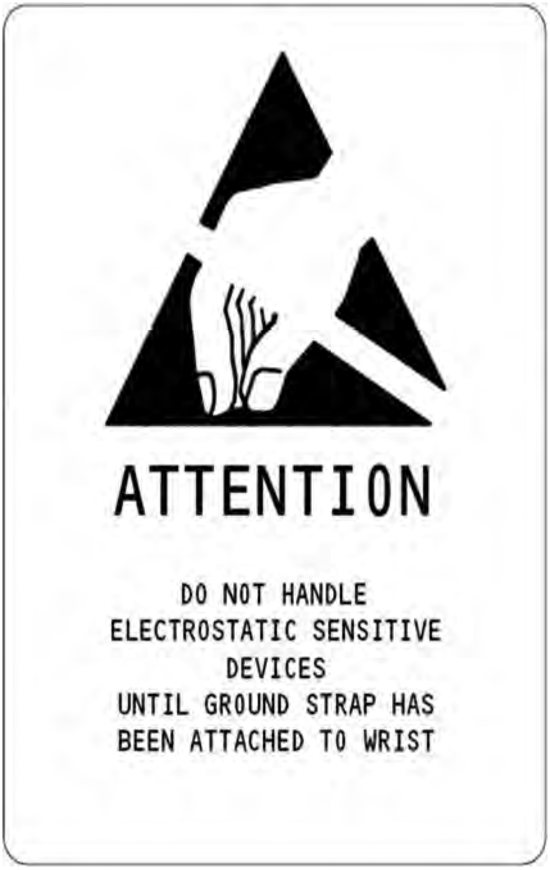

Figure 19.14 ESD warning notices

Electrostatic sensitive devices (including printed circuit boards, circuit modules, and plug-in devices) are invariably marked with warning notices. These are usually printed with black text on yellow backgrounds, as shown in Fig. 19.14.

Special precautions must be taken when handling removing ESSD. These precautions include use of the following:

- wrist/heel straps
- static dissipative floor and bench mats
- ground jacks
- grounded test equipment
- low-voltage soldering equipment and anti-static soldering stations (low-voltage soldering irons with grounded bits)
- anti-static insertion and removal for integrated circuits
- avoidance of nearby high-voltage sources (e.g. fluorescent light units)
- anti-static packaging (static-sensitive devices and printed circuit boards should be stored in their anti-static packaging until such time as they are required for use)
- Protective materials.

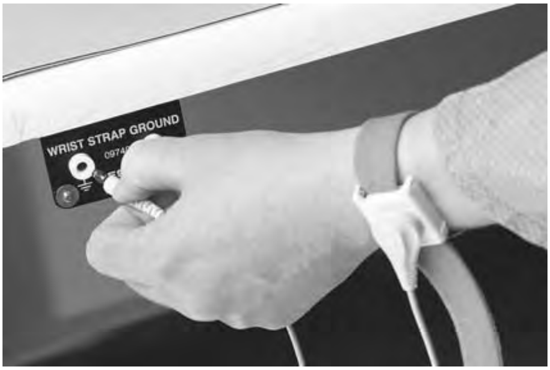

Figure 19.15 ESD Typical wrist strap grounding point

Wrist straps (and also heel straps) are conductive bands that are connected to an effective ground point by means of a short wire lead. The lead is usually fitted with an integral 1 MΩ resistor which minimizes currents arising between the person wearing the strap and the equipment being worked on. Wrist straps are usually stored at strategic points on the aircraft, or they may be carried by maintenance technicians.

Figure 19.15 shows a typical wrist strap being used for a bench operation. The bench itself must be grounded; this is achieved by connecting the conductive mat to an external earthing point outside of the building (Fig. 19.16).

There are three main classes of materials used for protecting static sensitive devices. These are:

- conductive materials, e.g. metal foils
- static dissipative materials (a cheaper form of conductive material)
- anti-static materials, e.g. cardboard and cotton.

Of these, conductive materials offer the greatest protection whilst anti-static materials offer the least protection. The working environment plays an important part in the safe handling of ESSD; extra vigilance is required when the relative humidity is low, e.g. in air conditioned workshops. (Further reading on the subject of EMI and ESSD can be found in a related book in the series, *Aircraft Digital and Electronic Computer Systems.*)

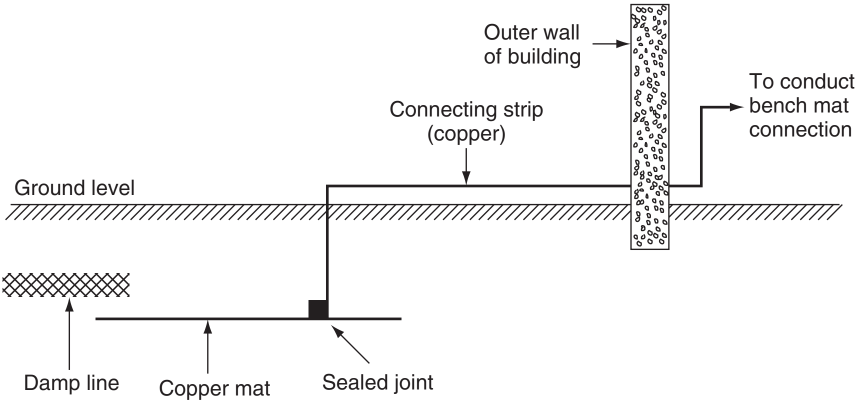

Figure 19.16 ESD work area protection

Key maintenance point

The effect of lower humidity on ESSD components in typical working environments will be to create higher voltages and pose more threat to the components.

19.7 Multiple choice questions

1. The ability of an item of equipment to operate alongside other items of equipment without causing EMI is called:
   (a) EMC
   (b) ESSD
   (c) HIRF.

2. Higher-frequency signals can lead to cross-talk between wires as a result of the:
   (a) dielectric effect of the wire insulation causing increased capacitance
   (b) dielectric effect of the wire insulation causing reduced capacitance
   (c) dielectric effect of the wire conductors causing reduced capacitance.

3. Radio and radar transmitters in the external environment are sources of:
   (a) EMC
   (b) ESSD
   (c) HIRF.

4. Of the materials used for protecting static sensitive devices:
   (a) conductive and anti-static materials offer the least protection
   (b) conductive materials offer the least protection whilst anti-static materials offer the most protection
   (c) conductive materials offer the greatest protection whilst anti-static materials offer the least protection.

5. Shielding of conductors:
   (a) reduces radiation and minimizes susceptibility
   (b) increases radiation and minimizes susceptibility
   (c) reduces radiation and increases susceptibility.

6. The presence of unwanted voltages or currents in systems is caused by:
   (a) EMC
   (b) ESSD
   (c) EMI.

7. When the lightning discharge is dissipated safely through the aircraft, structural damage is most likely to occur at the:
   (a) bonding between structures
   (b) exit points of the discharge
   (c) entry points of the discharge.

8. In the context of EMI, high-frequency signals should be:
   (a) filtered out of a circuit wherever possible
   (b) introduced into a circuit wherever possible
   (c) used only in unshielded cables.

9. The twisted pair is designed to:
   (a) maximize cross-talk between adjacent pairs of wires
   (b) protect static sensitive devices
   (c) minimize cross-talk between adjacent pairs of wires.
10. The effect of lower humidity on ESSD components in typical working environments will be to:
   (a) create lower voltages and pose more threat to the components
   (b) create higher voltages and pose more threat to the components
   (c) create higher voltages and pose less threat to the components.

Chapter 20 Continuing airworthiness

Many processes are required throughout the aircraft's operating life to ensure that it complies with the applicable airworthiness requirements and can be safely operated. The generic term for this range of processes is continuing airworthiness. The term 'maintenance' is used for any combination of overhaul, repair, inspection, replacement, modification or defect rectification of an aircraft or component, with the exception of the pre-flight inspection. Persons responsible for the release of an aircraft or a component after maintenance are the certifying staff. Maintenance of an aircraft and its associated systems requires a variety of test equipment and documentation; these are required by certifying staff to fulfil their obligations in ensuring continued airworthiness. Aircraft wiring cannot be considered as 'fit and forget'. Legislation is being proposed to introduce a new term: **electrical wire interconnection system** (EWIS); this will acknowledge the fact that wiring is just one of many components installed on the aircraft. EWIS relates to any wire, wiring device, or combination of these, including termination devices, installed in the aircraft for transmitting electrical energy between two or more termination points. This chapter reviews some practical installation requirements, documentation and test equipment required by the avionics engineer to ensure the continued airworthiness of aircraft electrical and electronic systems.

20.1 Wire and cable installation

The importance of aircraft wire and cable selection, installation and maintenance cannot be overstated. Modern aircraft (or upgraded older aircraft) are installed with more wiring, carrying more current than earlier generations of aircraft; much of this wiring carries digital signals. Wiring and cables must be treated as integral components of the aircraft; they are not to be treated as 'fit and forget'. Wires are formed from a single solid conductor or stranded conductors, contained within insulation and protective sheath materials. Cables can be defined as:

- two or more separate wires within the same insulation and protective sheath
- two or more wires twisted together
- any number of wires covered by a metallic braid, or sheath
- a single insulated conductor covered by a metallic outer conductor (co-axial cable).

The terms wires and cables are often interchanged. In this chapter, reference will be made to 'wiring' in the all-embracing generic sense; cables will be referred to in specific terms as and when required. Surveys and inspections of aircraft have revealed a number of issues and problems that require the close attention of system design and maintenance to ensure continuing airworthiness. Wire insulation can deteriorate over time (typically over ten years); exposed conductors create the environment for potential faults, spurious signals and **arcing**.

Wires are vulnerable to their installed environment, e.g. changes in temperature, exposure to moisture and vibration that can lead to **open circuits** and/or **chafing**. In certain areas of the aircraft, e.g. the leading/trailing edges of the wing and wheel wells, the physical environment is harsher than protected areas, e.g. the flight compartment or passenger cabin. The installation of wire and cable is very important; the following must be avoided:

- sharp bend radii
- unsupported wires
- routing high and low power circuits in the same bundle, or loom.

All the above can intensify the ageing and environmental effects on wiring. Certain older standards of wire insulation, notably single-walled aromatic polyamide, have a known vulnerability to high temperature that can lead to the insulation forming small cracks; this can lead to moisture ingress, and overheating. Under these conditions, the carbonized insulation becomes a conductor and the situation propagates leading to the possibility of fire. It is essential that older standards of wiring are inspected in accordance with maintenance schedule; the wiring should be replaced if there are any signs of deterioration. Care must be taken not to disturb or damage wiring during maintenance or inspection of nearby equipment.

Key maintenance point

Faulty wiring is often overlooked as the reason for systems becoming unserviceable. Line replaceable units (LRU) are often changed first, with the inevitable 'no fault found'.

20.1.1 Cable and wire looms

Grouping individual wires and cables into bundles forms a loom (or harness); these bundles are tied, or strapped together to form a secure assembly. Wire looms must be installed and maintained to ensure maximum integrity. The loom should be formed without twisting or overlapping the cables/wires. They can be tied together with waxed string, lacing cord or nylon straps. These ties are made at regular intervals, equally spaced along the loom; the loom is then secured to the airframe with clamps. The current-carrying capacity of wires reduces in looms since the inner wires are not able to radiate heat efficiently. To illustrate this point, a single 20-gauge wire in free circulating air is rated at 14A; this reduces depending on the number of wires in the loom:

- single wire, 14A
- three wires, 9A
- seven wires, 7A
- twelve wires, 5A.

Figure 20.1 provides some examples of installed wire looms. The loom will follow a path through the aircraft dictated by existing structure and obstacles; additional support and protection is provided where required. Any bends or branches in the loom must not be so sharp that the loom becomes kinked (never bend or form wires or cables with pliers or any sharp edge). Looms must be installed such that they are supported with clamps and protected from chaffing through contact with sharp edges, pipes and other wiring.

Cable clamps and grommets are used throughout the aircraft to provide this protection. General guidelines for the installation of wire looms are as follows:

- Install with a downward slope away from equipment to prevent moisture running into the equipment
- Do not install below fuel pipes
- Avoid areas of high temperature
- Install to minimize EMI
- Make provision for at least one remake of the wire
- Take care not to crush coaxial cables
- Areas of high vibration require additional loom support
- Breakout should not cross over the main loom.

Open looms are formed with bound wires in a bundle; these are supported by p-clips and protected by grommets. The use of open looms is dictated by temperature and length of loom, together with any EMI considerations. **Conduits** (Fig. 20.2) are used in specific areas, e.g. wing leading edges to protect wiring loom from rain and other fluids; these conduits are made from plastic or metal.

When looms can be passed through bulkheads, they must be prevented from chaffing; this can be achieved by clamping or potting, see Fig. 20.3. **Ducted looms** are formed in channels made from a suitable material, e.g. aluminium alloy or composite. They provide more support and are used to guide the looms through and around specific areas of the aircraft. When the loom passes through bulkheads, they are often sealed with a rubber bung, see Fig. 20.3.

Certain areas of the aircraft will experience high vibration (Fig. 20.4); wiring in these areas will be subjected to a harsher environment. These areas include:

- wheels wells
- empennage
- wing roots
- wing trailing edges
- wing leading edges
- engine pylons and nacelles.

The wiring in these areas will be exposed to severe wind and moisture problems (**SWAMP**); wiring specifications and inspection requirements must be adhered to.

Key maintenance point

Cracking and breakdown of the insulation material through exposure to moisture, the speed of breakdown depending on both temperature and stress; this phenomenon is known as **hydrolysis**.

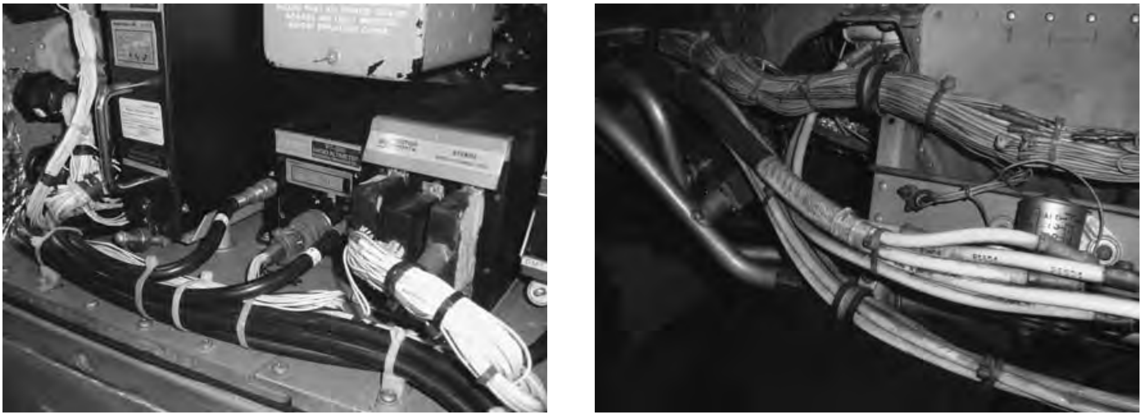

Figure 20.1 Examples of installed wire looms

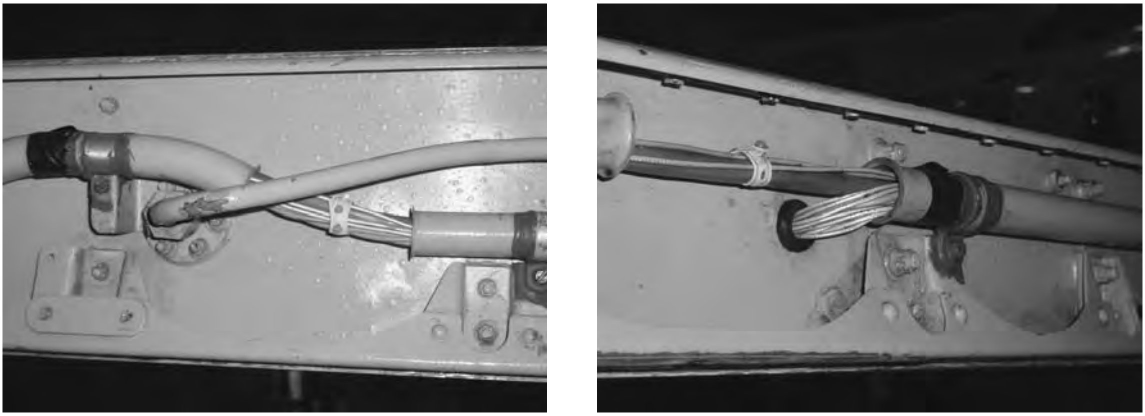

Figure 20.2 Examples of cable-loom conduit

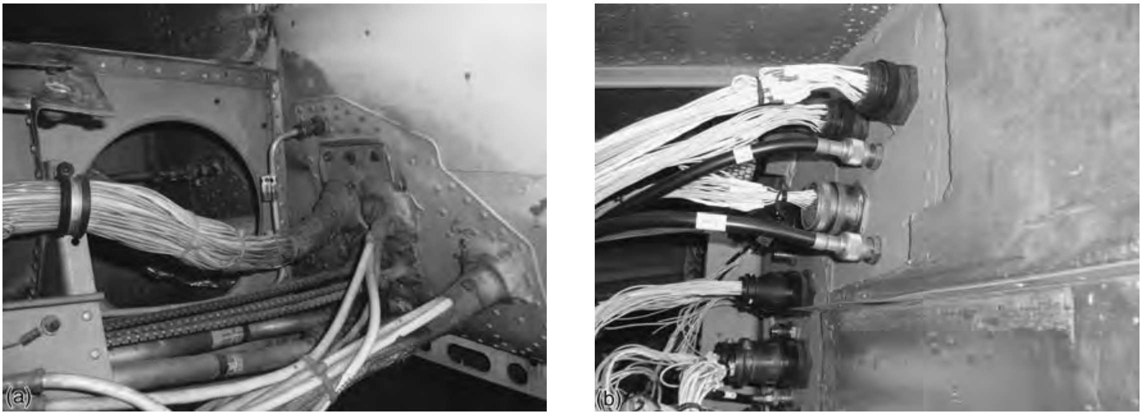

Figure 20.3 Bulkhead wire looms: (a) sealant/potting, (b) connectors

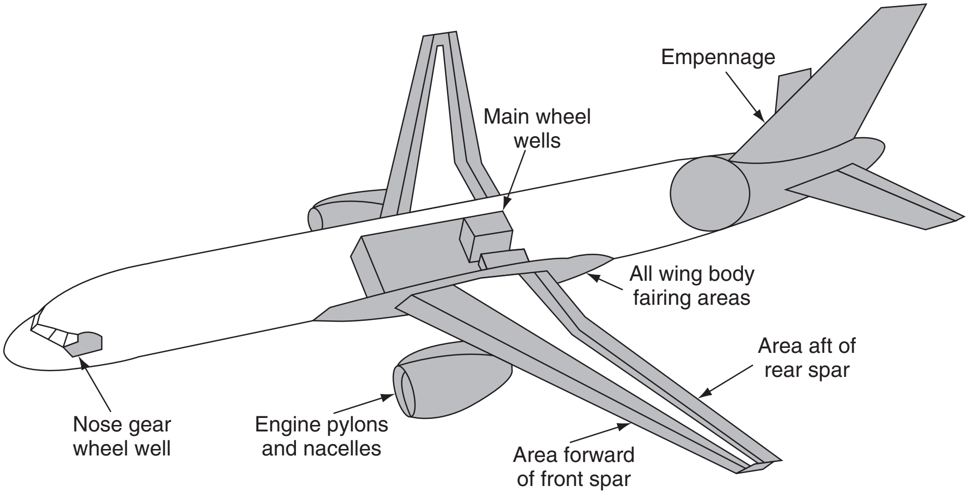
Empennage
Main wheel wells
All wing body fairing areas
Area aft of rear spar
Nose gear wheel well
Engine pylons and nacelles
Area forward of front spar

Figure 20.4 High vibration areas

Key maintenance point

Carbon arc tracking occurs when contaminating moisture (including aircraft fluids) creates a short circuit between an exposed conductor and the aircraft structure or an adjacent exposed conductor at a different potential; this phenomenon is known as **wet arc tracking**.

Key maintenance point

Carbon arc tracking occurs in dry conditions when one or more conductors are shorted as a result of abrasion from: the aircraft structure, wire to wire, or installation error. This phenomenon is known **dry arc tracking**.

20.1.2 Wire terminations and connections

It is highly unlikely that a single wire or cable will be routed from the power source directly to the load – it will invariably pass through bulkheads via connectors. Certain areas in the aircraft are not suitable for the routing of wires (e.g. zones exposed to high temperature or EMI). Wiring is invariably installed in sections and joined at intervals. Terminations and connection types used will depend on a number of factors driven by cost and continued airworthiness requirements. The size and configuration of the aircraft will determine where connectors are needed and located. Other considerations are when aircraft sections (wings, fuselage, etc.) are manufactured at various geographical locations and come together at a different place for final assembly. Finally, the need for inspection, removal and installation of equipment needs to be considered; quick-release connectors are used for most **line-replaceable units (LRU)**, see Fig. 20.5.

Wires can be joined by soldering, although this is normally only used within equipment. (Soldering aircraft wires reduces their flexibility and can lead to premature failure.) The majority of cable and wire terminations in the airframe are made by attaching crimp tags for use with terminal blocks or pin and sockets within connectors, see Fig. 20.6. There are two types of crimp: dimple and confined; the latter is the type normally used on aircraft.

The confined crimp is formed by compressing the crimp's shank onto the conductor. This results in the **cold flow** of metal between shank and conductor forming a homogeneous mass. The crimping operation also causes the crimp to form over the insulating and sheath materials. The crimping operation is performed with crimping pliers that contain two dies to form the crimp in a controlled and preset way. For larger-diameter cables and wires, power tools are required. Connections can be made to terminate a cable with a ring-tag, or join two cables together with a splice. These can be used for repairs or as permanent installations. They can also be used to form junctions, e.g. when modifying

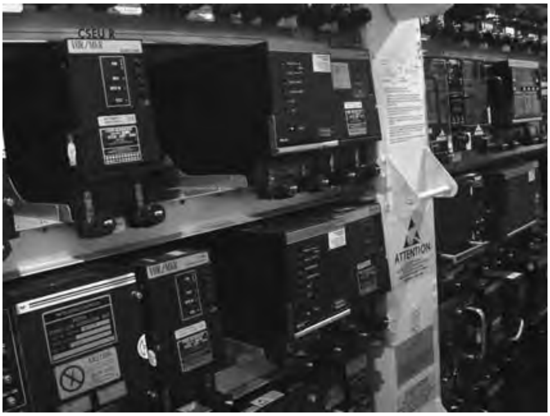

Figure 20.5 Quick-release connectors (LRU)

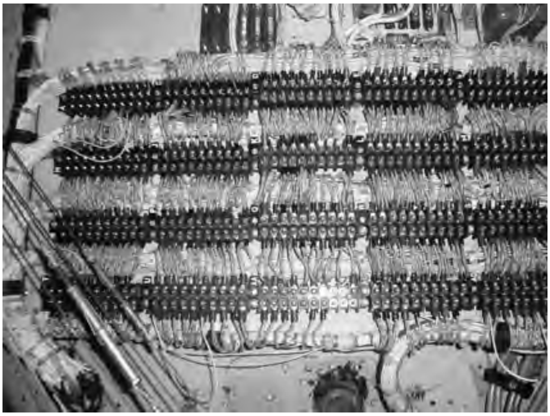

Figure 20.6 Terminal blocks with connections

an aircraft to 'tap' into a signal line. General rules for the use of splices will be given in the relevant aircraft documents; these will address the:

- spacing of splices in the same wire
- maximum number of splices in a given length of wire
- support of the splices (they should not be located in curved sections).

Both tags and splices have a plastic or nylon insulating sleeve covering the shank. This insulation is coloured red, blue or yellow depending on size; this colour coding relates to the specific crimping pliers required for the operation. To perform the crimping operation on a tag or splice, the wire must have its insulation/sheath stripped from the end. Stripping is carried out taking the following requirements into account:

- an approved stripping tool must be used with the correct size of cutting blades
- knives or side cutters must not be used since they can sever or damage the conductor(s)
- all insulating and sheath material must be removed to suit the crimp or splice.

If the above is not adhered to, problems will be introduced, e.g. reduced current-carrying capacity and reduced mechanical strength; this could lead to premature failure and/or overheating of the wire. Once stripped, the strands of a wire should be formed by hand so that they lie neatly together. They must be 100% inserted into the barrel of the tag or splice; if any strands are disturbed by the insertion, the wire must be withdrawn and the strands straightened. Excessive twisting will increase the diameter of the conductor, making it difficult to insert into the barrel. Once formed, the tag can be attached to terminal blocks with connections made to respective circuits. Some terminations benefit from a heat-shrink sleeve; this provides extra mechanical protection and support. Heat-shrink material is polythene based and reduces to a pre-determined diameter (but not length) when heated. The advantages of crimping are:

- good conductivity
- consistent operation
- good strength
- resistance to corrosion.

Key maintenance point

Always use approved crimping tools; crushing a termination does not constitute a good connection!

20.1.3 Connectors

Wires and cables also have to be routed through bulkheads or production breaks; this is normally achieved by using quick-release connectors. General aviation connectors (D-shaped) are the same type as used on personal computers, see Fig. 20.7.

Larger aircraft use circular quick-release connectors as shown in Fig. 20.8. Pins and sockets are crimped onto their respective ends of stripped wires; the connector is assembled usually with a supporting bracket.

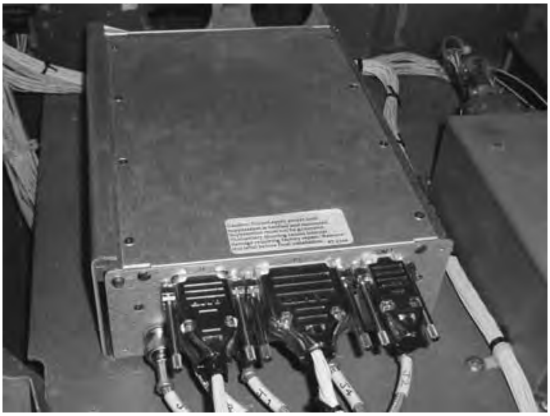

Figure 20.7 General aviation connectors (D-shaped)

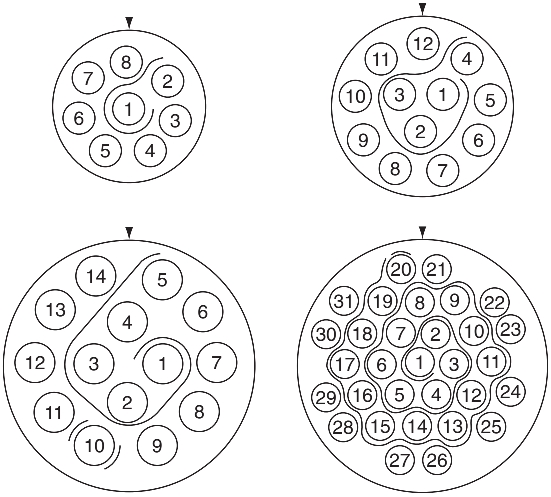

Figure 20.9 Pin/socket identification

Connectors are released by screws or by twisting the quick-release mechanism. Pins and sockets are removed and inserted using special tools; they are crimped onto the wires in a similar way to crimping of tags and splices.

Key maintenance point

If the connector is to be left undone for a period of time, always protect the pins/sockets with a cap or plastic bag.

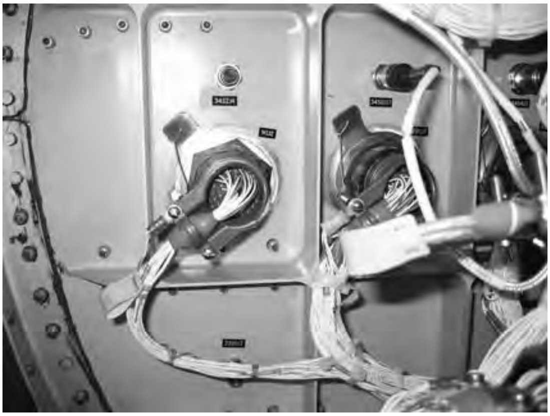

Figure 20.8 Circular quick-release connectors

The bodies and shells of quick-release connectors are made from light alloy, stainless steel or cadmium-plated alloys. Keyways ensure the correct alignment and prevent any twisting movement. The pins slide-fit into the respective socket; the pins are gripped by springs inside the socket to provide the electrical contact surface. A number identifies each pin/socket; see Fig. 20.9 for examples.

Key maintenance point

Pin/socket numbers of a connector are referenced to a physical point, or marking on the connector.

Key maintenance point

Connector socket spring-grips wear over time and can cause loss of continuity. Inserting anything but the correct pin can cause permanent damage.

Key maintenance point

When re-making connectors, always check the alignment of keyways of mating connectors before making the connection.

20.1.4 Aluminium wires or cables

Some applications use **aluminium** wires or cables to save weight. It is essential that the correct procedures, tools and materials are used when working with aluminium wires or cables. Aluminium oxidizes immediately when exposed to air, leading to high-resistance joints; compounds must be used to prevent this. Petrolatum (semi-solid mixture of hydrocarbons) and granular zinc compounds can either be applied during the crimping process, or be pre-filled within the crimp. When the crimp joint is made, the compound penetrates around each of the strands; the abrasive action of the compound scours the surfaces and removes the oxidation. When the crimping process is completed, the compound remains in the crimp to form a sealed termination. Aluminium wiring crimps joined into the main aircraft distribution system, it will be joined with relatively hard materials, e.g. copper or steel. Fixing hardware, e.g. nuts, bolts and washers must be made of materials such as cadmium-plated aluminium that will not cause electrolytic corrosion. It is essential that washers are used to protect the softer aluminium material. The surfaces must first be cleaned and treated with compound; the terminations are secured with nuts that must be torque-loaded.

Key maintenance point

Flexing of aluminium wires during installation together with aircraft vibration causes work-hardening leading to brittle joints.

20.2 Bonding

It is a mandatory requirement that aircraft structure and equipment are electrically bonded. Specific bonding and grounding connections are made in an aircraft to accomplish the following:

- dissipate energy from a high intensity radiated fields (HIRF) and lightning strikes
- dissipate static electricity
- limit the potential difference between equipment
- provide a low resistance path for earth return systems.

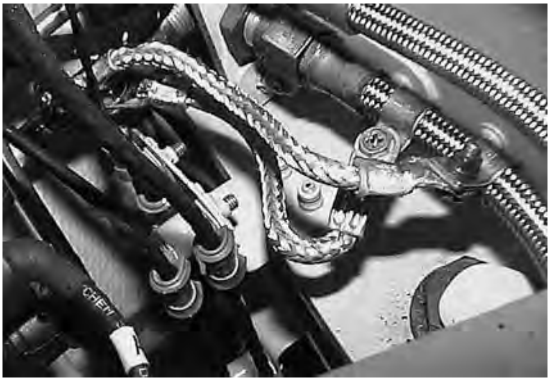

Figure 20.10 Bonding

Bonding connections are made between components and structure using purpose-made straps, see Fig. 20.10 (further examples are given in Chapter 19). Bonding is categorized as either primary or secondary; this is determined by the magnitude of current being conducted. **Primary** bonding is designed for carrying lightning discharges and to provide electrical return paths. **Secondary** bonding is used to dissipate static electricity and keep all structure at the same potential. Bonding straps (or leads) are pre-fabricated from braided copper or aluminium terminated with crimps.

20.2.1 Composite materials

There is an increasing use of **composite materials** being used in the construction of aircraft because of their good strength-to-weight ratio (compared with aluminium). Composite material has a high electrical resistance and is intrinsically unsuitable for bonding, earth returns and lightning strike dissipation. A ground plane has to be integrated into the airframe; this is normally achieved by bonding an aluminium wire mesh into the composite structure during manufacture. This mesh is accessed at key points around the aircraft to gain access to the ground plane.

Direct bonding (Fig. 20.11) is achieved by exposing the mesh (ground plane) and mounting the equipment directly onto the conductive path. **Indirect bonding** (Fig. 20.12) is achieved by exposing the mesh and installing a bonding wire and connector. The mesh must always be coated after making a connection since the aluminium will oxidize when exposed to air, leading to high resistance and unreliable joints.

Lightning protection in composite aircraft is achieved via aluminium wire integrated into the outer

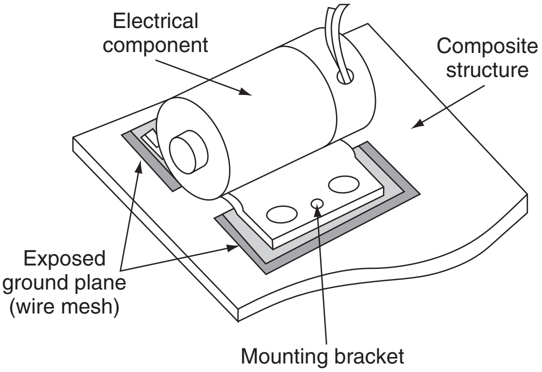

Figure 20.11 Direct bonding on composite structure

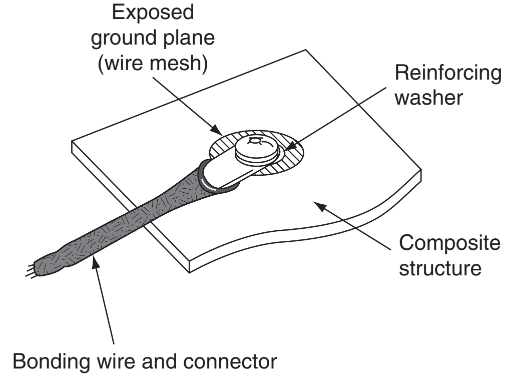

Figure 20.12 Indirect bonding on composite structure

layers of the composite construction. The lightning strike will enter and leave the aircraft at its extremities; the integrated wires are installed in anticipation of this and the energy dissipated through the aircraft along pre-determined routes to the exit point(s).

20.2.2 Maintenance requirements

Maintenance requirements for bonding include inspection of the two crimped tags to ensure that they are secure and not corroded. The braided conductor should be inspected for any visual signs of mechanical damage; it must not interfere with any moving items e.g. flying controls or actuators. The bonding lead must have sufficient slack to allow for aircraft flexing. All contacting surfaces must be bare metal; the crimp tags are then treated and sealed in accordance with the maintenance manual. Electrolytic corrosion can occur at the bonding lead connection; it is essential that the correct bonding lead materials and treatments are used.

Key maintenance point

If a bonding lead has to be changed, refer to the aircraft documentation to obtain the correct type and part number.

20.3 Static charges

Static charges build up on the airframe during flight and these must be discharged into the atmosphere to avoid interference with avionic equipment. Electric charge is dissipated from the surface of an aircraft during flight by an effect called **corona discharge**. Coronas can generate both audible and radio-frequency (RF) noise, thereby disrupting the avionic systems. (Power lines produce an audible sound because they are producing a corona discharge that is interacting with the surrounding air.) Static electricity is discharged from the aircraft to atmosphere through **static wicks** (Fig. 20.13), sometimes called **static dischargers**. These ensure that the discharge of static electricity is from pre-determined points.

Static dischargers are fabricated with a wick of conductive element that provides a continuous low-resistance discharge path between the aircraft and atmosphere. The corona discharge can cause a faint glow adjacent to the prominent sections of the aircraft, e.g. wing tips; this results from ionization of oxygen and the formation of ozone in the surrounding air. The electric field created by the flow of electrons into the conductive section of the wick contains enough energy to ionize the oxygen and nitrogen in the air. This can produce low-energy **plasma**, the corona discharge. Plasma is a fourth state of matter along with solids, liquids, and gases. It is similar in appearance to a gas or liquid, but the molecules are separated into atoms from which the electrons in the outer shell are released into the plasma.

Key maintenance point

Helicopters (or rotorcraft) build up additional static charges through their rotor blades.

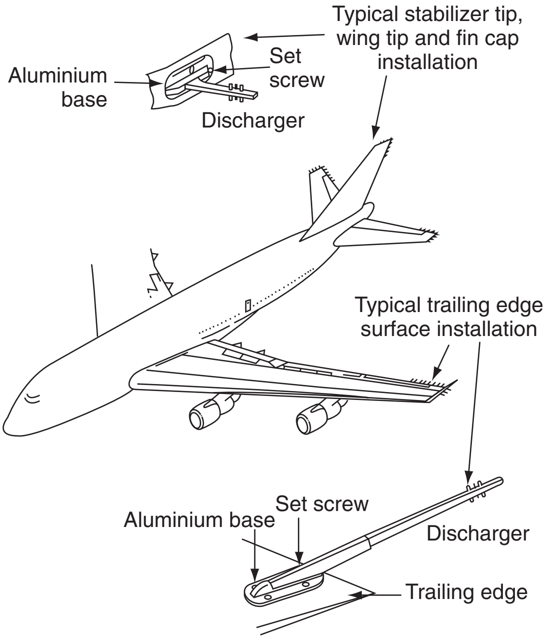
Typical stabilizer tip, wing tip and fin cap installation
Set screw
Aluminium base
Discharger
Typical trailing edge surface installation
Set screw
Aluminium base
Discharger
Trailing edge

Figure 20.13 Static discharger locations

Despite the precautions taken to dissipate static charge during flight, it is highly likely that some charge will remain in the aircraft after landing. This presents a personal shock hazard for crew, passengers and ground staff. Static electricity is dissipated through tyres when the aircraft lands; the tyres are impregnated with a conductive compound. The retained static electricity also presents a major hazard during refuelling since the aircraft and fuel bowser will almost certainly be at different potentials, thereby causing a spark to jump across the fuel connections. The aircraft and fuel bowser are therefore connected by a length of cable onto specified points before refuelling starts.

Key maintenance point

A loose bonding connection causes more problems than having no bonding at all.

20.4 Earth returns

Aircraft with metallic structure use the airframe as the means of completing the electrical circuit, thereby reducing the cost, weight and installation time of installing a return wire, see Fig. 20.14. In this example, the positive side of the battery power supply is connected to the navigation light via the circuit protection and control switch.

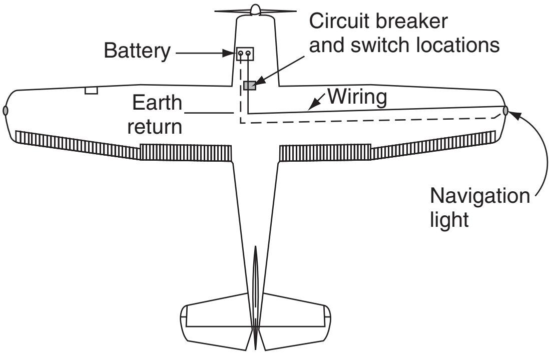

Figure 20.14 Airframe 'earth return' circuit

The earth (or ground) return is connected to the negative side of the power supply via the aircraft structure. In some installations, grounding of the load, e.g. the navigation light is via the body of the component's housing. In other installations, all the negative connections are collected at earth stations. (Wires terminating at an earth station are usually identified with a suffix letter N.)

The location of the earth stations depends on a number of factors including the:

- mechanical strength of the structure
- current through the connections
- corrosive effects (dissimilar materials)
- ease of accessing/making the connections.

It is essential that the ground (or earth points) are not mixed between different types of circuit e.g. AC and digital signals. Using a common return path can lead to corrupted signals as illustrated in Fig. 20.15(a). Separate return paths (Fig. 20.15(b)) should be considered on all new system designs; this separation must be maintained whenever specified in the aircraft manuals.

Taking these factors into account, installations will vary considerably; see Fig. 20.16 for some examples. In order to make good electrical contact, there must be minimal resistance between the conductor and structure. Earth stations incorporate an anti-corrosion tag or plate to prevent electrolytic reaction between the base of the assembly and aircraft structure. The stations are usually identified on the adjacent structure by numerals preceded by an asterisk.

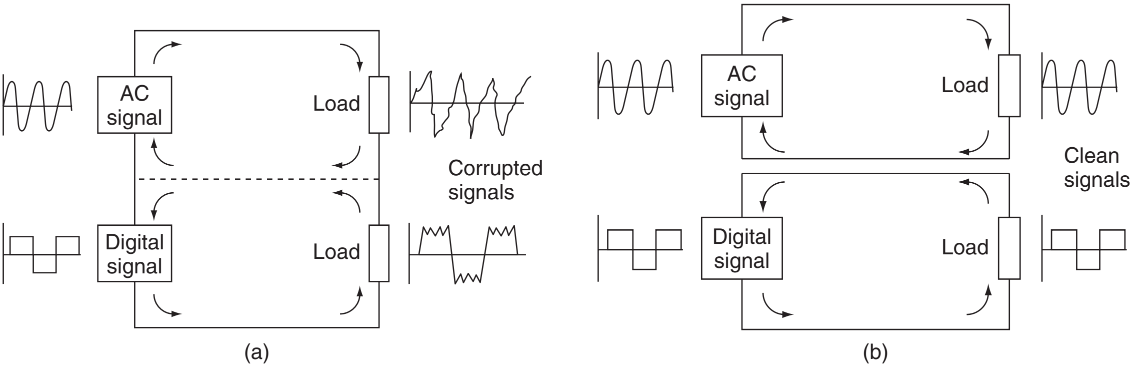

Figure 20.15 Earth/ground loops: (a) common return paths, (b) separate return paths

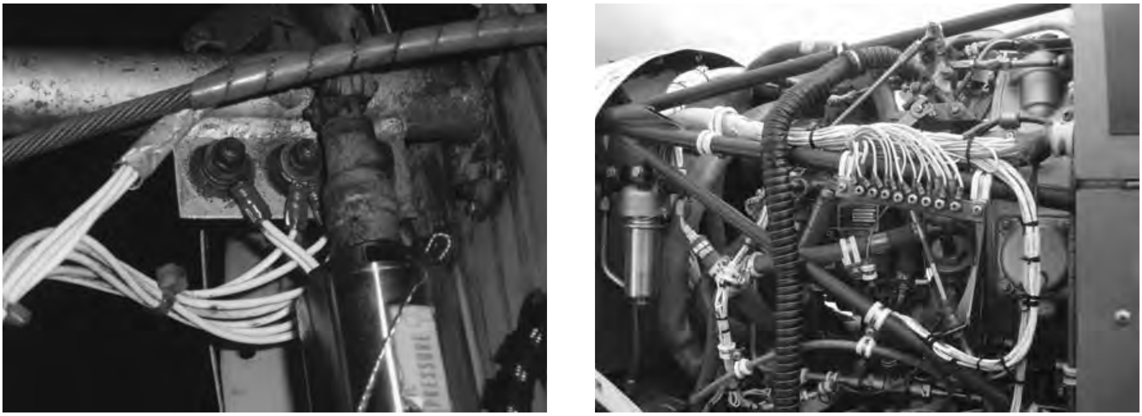

Figure 20.16 Earth station installations

Certain circuits must be isolated from each other, e.g. AC neutral and DC earth-returns must not be on the same termination; this could lead to current flow from the AC neutral through the DC system. Relay and lamp returns should not be on the same termination; if the relay earth connection has high resistance, currents could find a path through the low resistance of the filament lamp when cold.

For aircraft that have non-metallic **composite** structure, an alternative means of providing the return path must be made. This can be in the form of copper strips running the length of the fuselage or a wire mesh formed into the composite material. The principles of earth return remains the same as with bonding; the method of achieving it will vary.

20.5 Aircraft manuals

There is a wide range of technical manuals required by certifying staff, including, but not limited to:

- maintenance manuals
- illustrated parts catalogues
- wiring diagram manuals
- schematic diagrams.

These documents contain information required to explain how equipment operates, how it is maintained, and the correct part numbers that can be used when replacing components. For electrical and electronic systems, additional information is required to show how the systems are supplied with power, interconnections with other systems etc. For larger aircraft, a standardized manual format was developed by the Airline Transport Association (ATA) of America called **ATA Spec. 100**. This specification contains format and content guidelines for technical manuals written by aircraft manufacturers and suppliers. The format has widespread international acceptance by aircraft manufactures, airlines and equipment suppliers for the maintenance of their respective products. The specification is organized into **ATA chapters** that are specific to an aircraft system, e.g. ATA Chapter 24 contains all subjects within the electrical power systems, Chapter 33 contains details of the aircraft lights. (A complete list of ATA chapters is given in Appendix 8.) The ATA Spec. 100 is not widely adopted by the general aviation (GA) industry; the information required by certifying staff on GA aircraft is presented in a number of formats. In 2000, ATA Spec 100 (documentation) and ATA Spec 2100 (for inventory management) were incorporated into **ATA iSpec 2200**.

20.5.1 Maintenance manual

This is the primary source of information and data for certifying staff. When based on ATA Spec. 100, the maintenance manual (MM) is organized into specific sections for complete systems and/or individual components:

- detailed description (from p. 1)
- component locations (from p. 101)
- maintenance practices (from p. 201)
- servicing (from p. 301)
- removal and installation (from p. 401)
- adjustment /test (from p. 501)
- inspection/checking (from p. 601).

Not all of these sections apply to every system and/or individual component.

20.5.2 Wiring diagram manual

Aside from the aircraft maintenance manual, the primary manuals used by avionic engineers are the wiring diagram manuals (WDMs), schematic diagrams and/or circuit diagrams. Wires and cables should be identified to facilitate installation, trouble-shooting and potential modifications. There are various specifications that provide details of how wiring identification is implemented, including those adopted by the aircraft manufacturers. The simplest form of identifying cables is to mark the ends of the wires with the **source** and **destination** of the individual wire or cable. The wire/cable insulation is marked with indelible ink or laser printing. Wires are sometimes identified by the system type, wire number and gauge. Some manufacturers mark the cables at intervals along the cable. On more complex installations, where cables are bundled into looms, the entire loom may be given an identification code; this would be marked on a sleeve or band.

When created under ATA Spec. 100, the WDM provides details of equipment locations, wiring between equipment and connectors (including wire and cable identification). Schematic diagrams provide an overview of the system interconnections and equipment locations; individual wire numbers and connections are omitted for clarity. In some GA applications, basic electrical and electronic system information is contained within the maintenance manual. In smaller GA aircraft, much of the avionic equipment is installed as a customer option; in this case, the equipment installer provides the necessary details as supplements. This also applies in the case where new equipment is installed as part of an equipment upgrade. For the majority of cases, symbols used in wiring diagrams and schematics are standardized; however, some variations do exist among manufacturers. Symbols used for electrical and electronic equipment are provided in Appendix 9.

20.6 Circuit testing

The typical sequence in which electrical and electronic systems are checked and tested is as follows:

- **visual inspection**; this enables obvious faults to be identified and the appropriate action(s) taken
- **bonding test**; this checks the continuity of the earth return path back to the power supply(s)
- **continuity test**; this checks the interconnections between all circuits
- **insulation resistance (IR) test**; this checks for adequate insulation resistance between conductors and between conductors and the airframe (earth)
- **functional test**; there will be varying levels of test depending on complexity of circuit. These range from simple ‘self tests’ through to deeper tests that can only be carried out in the workshop.

Key maintenance point

For personal safety and to reduce the risk of damage to wiring and/or components, isolate power supplies when changing equipment or inspecting interconnecting wiring.

There are a number of techniques used to identify and locate faults in aircraft wiring and equipment, e.g. continuity and resistance measurements. There is a variety of equipment available to the avionics engineer to check and test aircraft electrical and electronic systems, including multimeters and oscilloscopes.

20.6.1 Multimeters

The most popular item of test equipment used for simple troubleshooting is the **multimeter**; these can be either analogue or digital instruments, see Fig. 20.17. Multimeters combine several functions into a single item of test equipment; the basic functions on simple multimeter include the measurement of current, voltage and resistance.

20.6.1.1 Analogue instruments

A popular analogue instrument used in the aircraft industry is the **Avometer**® (registered trade mark of Megger Group Limited). This instrument is often referred to simply as an **AVO**, deriving its name from the words amperes, volts, ohms. It has been in widespread use in the UK from the 1930s and can still be found in many hangars, workshops and repair stations. Referring to Fig. 20.18, features include the measurement of:

- alternating currents up to 10 A
- voltages up to 1000 V
- resistance from 0.1 Ω up to 200 kΩ.

The instrument is very accurate, typically ±1% of full-scale deflection (FSD) on DC ranges and ±2% on AC ranges. Maximum current consumption for

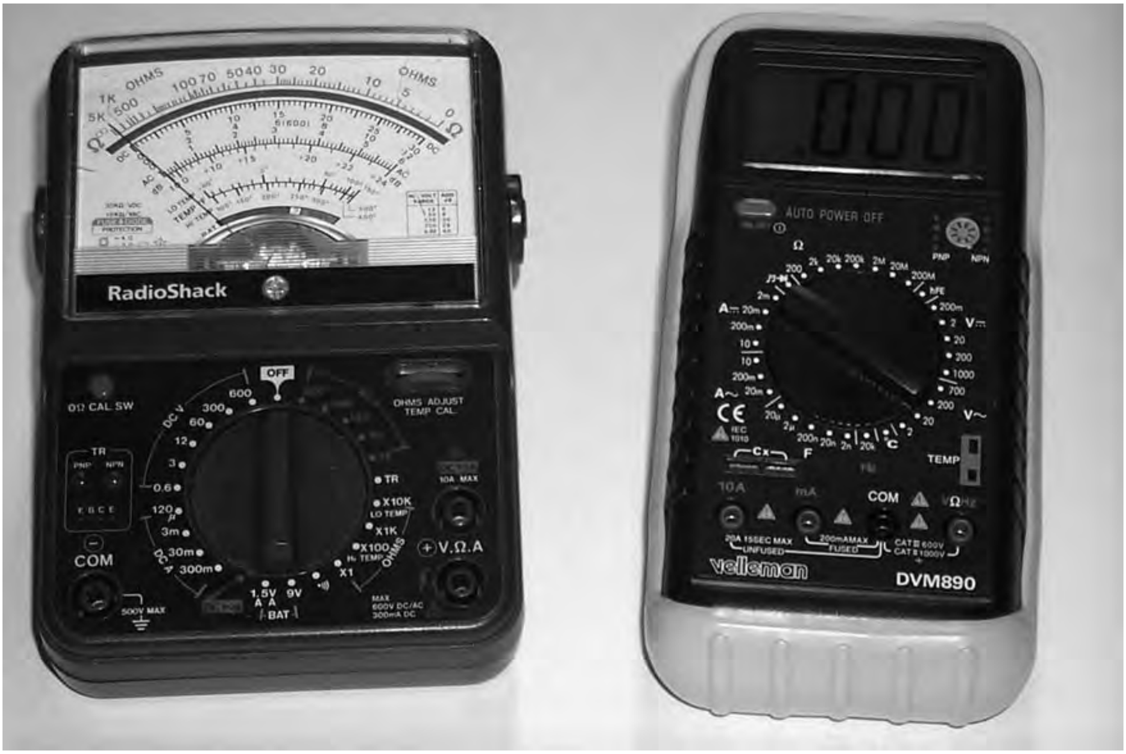

Figure 20.17 Multimeter; analogue or digital

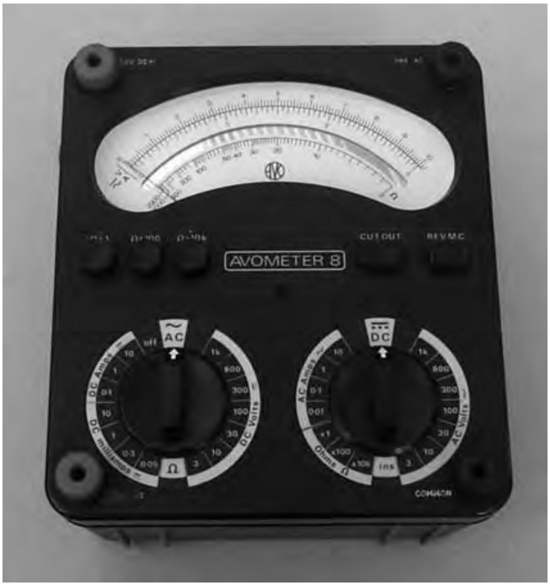

Figure 20.18 Avometer® (registered trademark of Megger Group Limited)

voltage measurement is 50 µA (corresponding to a sensitivity of 20,000 Ω per volt); thereby minimizing voltage measurement error. Two rotary switches are used to select the function and range to be measured; if the wrong combination of function or range is selected, an overload cut-out switch (similar to a circuit breaker) disconnects the test circuit. Some multimeters can measure the voltage drop across semiconductor junctions and measure additional circuit/component quantities such as:

- capacitance
- frequency
- temperature
- conductance
- inductance.

20.6.1.2 Digital multimeters

Digital multimeters are generally smaller than the analogue type of multimeter making them a useful handheld device for basic fault finding on the aircraft. Higher specification devices can measure to a very high degree of accuracy and will commonly be found in repair shops and calibration laboratories. Digital multimeters offer more features than basic analogue instruments; commonly available measurement features include:

- autoranging
- sample and hold
- graphical representation
- data acquisition
- personal computer interface.

Autoranging selects the appropriate range for the quantity under test so that meaningful digits are shown. For example, if a battery cell terminal voltage of 1.954 VDC was being measured, autoranging by a four-digit multimeter would automatically display this voltage instead of 0.019 (range set too high), or 0.999 (range set too low) via manual range selection. **Sample and hold** retains the most recent display for evaluation after the instrument is disconnected from the circuit being tested. **Graphical representation** of the quantity under test can be displayed in a number of ways, e.g. as a bar graph. This facilities observations of trends. Simple **data acquisition** features are used in some multimeters to record maximum and minimum readings over a given period of time, or to take a number of sample measurements at fixed time intervals. Higher-specification multimeters feature a **personal computer** interface, typically achieved by infrared (IR) links, or datalink connections, e.g. RS-232 or universal serial bus (USB). This interface allows the multimeter to upload measured data into the computer for storage and/or analysis.

20.6.2 Bonding meters

Aircraft bonding is tested with a bonding meter; the maintenance manual will define specific points on the airframe where measurements are made. The maximum resistance between specific points will be defined in the maintenance manual. The bonding meter consists of a low reading ohmmeter and two leads. One lead is 60 feet long with a single prodded end-piece; the other lead is 6 feet long with a double prodded end-piece. (Note that these leads must never be shortened.) The bond meter is first checked by:

- shorting all three prods together to check for zero-reading
- momentarily short the two prods of the 6 feet lead to check for full-scale deflection.

The maximum resistance for between extremities of fixed portions of a metallic aircraft and between bonded components/earth stations will be stated in the maintenance manual; the typical maximum value is 0.05 ohms for an aircraft with metal structure. The maximum value for composite aircraft will vary from type to type, but it will by necessity be very low.

20.6.3 Oscilloscopes

The **oscilloscope** (often referred to as a 'scope') is an item of electronic test equipment used to measure and view signal voltages as a two-dimensional graph (usually signal voltage on the vertical axis versus time on the horizontal axis). The oscilloscope will be found in most workshops, repair stations and calibration laboratories, see Fig. 20.19.

Original equipment was based on the cathode ray tube (CRT) display, making them heavy and bulky. Use of liquid crystal display (LCD) technology makes the oscilloscope useful as a portable device. (More information is provided on CRTs and LCDs in a related title in the series: *Aircraft Digital and Electronic Computer Systems.*) The portable oscilloscope has many uses during troubleshooting, e.g. checking for electrical noise and measurement of digital signals. PC-based oscilloscopes can be configured in an existing laptop via a specialized signal acquisition board and suitable hardware interfaces; the PC-based oscilloscope has a number of features:

- lower cost compared to a stand-alone oscilloscope
- efficient exporting of data into standard PC software, e.g. spreadsheets
- control of the instrument via custom programmes on the PC
- utilization of the PC's networking and disc storage functions
- larger-/higher-resolution colour displays
- colours can differentiate between waveforms
- portability.

Figure 20.19 Oscilloscope

20.7 Automatic test equipment

Automatic test equipment (ATE) is dedicated ground test equipment that provides a variety of different functional checks on line-replaceable units (LRUs) or printed circuit boards (PCBs). The equipment being tested is connected to a variety of external circuits that represent the aircraft interfaces; addition connections are often made for diagnostic purposes. ATE is able to gather and analyse a large amount of data very quickly, thus avoiding the need to make a very large number of manual measurements in order to assess the functional status of an item of equipment.

ATE usually incorporates computerized control with displays and printouts that indicate what further action (repair or adjustment) is necessary in order maintain the equipment. Equipment may then require further detailed tests and measurements following initial diagnosis. ATE tends to be dedicated to a particular type of avionic system; it is therefore expensive to develop, manufacture and maintain. Because of this, ATE tends to be only used by original equipment manufacturers (OEMs) and licensed repairers.

20.8 On-board diagnostic equipment

Systems have been developed to match the complexity of electrical and electronic systems to assist the avionics engineer in fault finding on aircraft systems. On-board diagnostics use a range of techniques that are built into and integrated with the aircraft systems.

20.8.1 Built in test equipment

As the name implies, **built-in test equipment (BITE)** is primarily a self-test feature built into aircraft electrical and electronic equipment as a means of:

- detecting and indicating specific equipment faults
- monitoring equipment performance
- detecting problems
- storing fault data
- isolate faulty sensors/components.

The origins of BITE started with simple on/off displays on the front of **line replaceable units** (LRU) to assist the avionics engineer with troubleshooting; this display (typically a light-emitting diode, LED) would indicate the **go/no-go status** for a particular unit, either as a result of system test or in-flight fault. If the LED indicates an LRU fault, the engineer either changes the LRU or check the interfaces with the unit. This simple technique is applied to individual LRUs, and will only indicate **real-time faults**. BITE techniques were developed alongside the increasing complexity of electrical and electronic equipment. Features such as fault storage provide an indication of faults over many flights; this information is often provided in coded form and needs to be interpreted by the engineer.

BITE is usually designed as a signal flow type test. If the signal flow is interrupted or deviates outside accepted levels, warning alerts indicate a fault has occurred. The functions or capabilities of BITE include the following:

- real-time monitoring of systems
- continuous display presentation
- sampled recorder readouts
- module and/or subassembly failure isolation
- verification of systems status
- go/no-go indications
- quantitative displays
- degraded operation status
- percentage of functional deterioration.

20.8.2 Centralized maintenance systems (CMS)

BITE technology has now been developed into centralized maintenance systems on modern aircraft. The electronic centralized aircraft monitoring system (ECAM) developed for Airbus aircraft oversees a variety of aircraft systems and also collects data on a continuous basis. While ECAM automatically warns of malfunctions, the flight crew can also manually select and monitor individual systems. Failure messages recorded by the flight crew can be followed up by maintenance personnel by using the system test facilities on the maintenance panel in the flight compartment. Printouts can be produced as permanent records for further analysis, see Fig. 20.20.

Dedicated CMS control display units (CDU) are installed on the centre pedestal. The CDU allows systems to be investigated via a menu selection; this provides a description of the **flight deck effect** (FDE) and likely reasons for the fault. Typical centralized maintenance systems have a dedicated **centralized maintenance computer** (CMC) to collect, analyse and store fault information. The CMC has up to 50 data bus inputs, 10 data bus outputs and can store up to 500 fault messages into non-volatile memory (NVM).

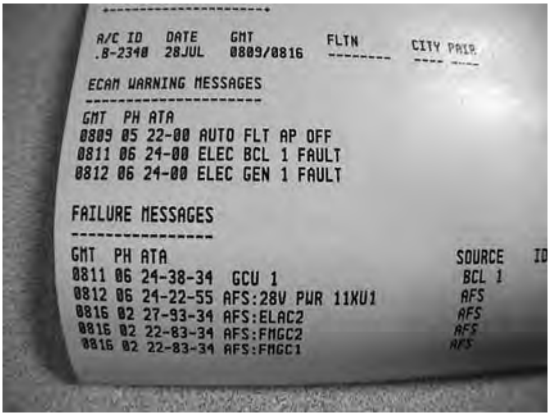

```
+----------------------+

A/C ID   DATE    GMT        FLTN       CITY PAIR
.B-2340  28JUL   0809/0816  --------   ---- ----

ECAM WARNING MESSAGES
---------------------
GMT  PH ATA
0809 05 22-00 AUTO FLT AP OFF
0811 06 24-00 ELEC BCL 1 FAULT
0812 06 24-00 ELEC GEN 1 FAULT

FAILURE MESSAGES
----------------
GMT  PH ATA                                  SOURCE  ID
0811 06 24-38-34  GCU 1                      BCL 1
0812 06 24-22-55 AFS:28V PWR 11XU1           AFS
0816 02 27-93-34 AFS:ELAC2                   AFS
0816 02 22-83-34 AFS:FMGC2                   AFS
0816 02 22-83-34 AFS:FMGC1                   AFS
```

Figure 20.20 ECAM printout

20.8.3 Aircraft communication addressing and reporting system

Some aircraft are installed with a system called ACARS (aircraft communication addressing and reporting system). This is a digital data link system transmitted in the VHF range (118 MHz to 136 MHz). ACARS provides a means by which aircraft operators can exchange maintenance and operational data directly without human intervention. This makes it possible for airline staff to communicate directly with the aircraft in their fleet in much the same way as it is possible to exchange data using a land-based digital network. ACARS uses an aircraft's unique identifier and the system has some features that are similar to those currently used for electronic mail. Typical ACARS messages are used to convey routine information such as:

- fuel data
- engine performance data
- aircraft fault data
- passenger loads
- departure reports
- arrival reports.

This information can be requested by the company and retrieved from the aircraft at periodic intervals or on demand. Prior to ACARS this type of information would have been transferred via VHF voice communications. (Additional information on ACARS is provided a related book in the series: *Aircraft Communications and Navigation Systems.*)

20.9 Multiple choice questions

1. ACARS is a digital data link system transmitted in the:
   (a) VHF range
   (b) LF range
   (c) UHF range.
2. Secondary bonding is designed for:
   (a) carrying lightning discharges through the aircraft
   (b) keeping all the structure at the same potential
   (c) discharging static electricity from the aircraft to atmosphere.
3. Some electrical installations use aluminium wires or cables to:
   (a) reduce EMI
   (b) save cost
   (c) save weight.
4. Visual inspection of wiring installations:
   (a) enables obvious faults to be identified and the appropriate action(s) taken
   (b) checks for adequate insulation resistance between conductors
   (c) checks the continuity of the earth return path back to the power supply(s).
5. In terms of bonding, composite material has a:
   (a) high resistance and is unsuitable for bonding
   (b) low resistance and is unsuitable for bonding
   (c) high resistance and is suitable for bonding.
6. For a given wire size, when installed in a loom it will be able to safely conduct:
   (a) more current compared to a wire in free circulating air
   (b) less current compared to a wire in free circulating air
   (c) the same current compared to a wire in free circulating air.
7. Bonding is categorized as primary or secondary determined by the:
   (a) use of composite or metal structure
   (b) locations of static wicks
   (c) magnitude of current being conducted.
8. The use of composite materials for aircraft structures results in:
   (a) less natural paths for bonding
   (b) more natural paths for bonding
   (c) higher probability of a lightning strike.
9. Bonding is made between components and structure using:
   (a) coaxial cable
   (b) purpose-made straps
   (c) general-purpose wiring.
10. Static electricity is discharged from the aircraft to atmosphere through:
    (a) composite structure
    (b) earth stations
    (c) static wicks.

Appendix 1 Abbreviations and acronyms

Abbreviations and acronyms are extensively used to describe electrical and electronic systems, as well as the parameters with which they operate. The abbreviations and acronyms used in this book, together with some commonly used terms, are listed below.

ABS	Automatic braking system
A/C	Air conditioning
AC	Alternating current
ACARS	Aircraft communication addressing and reporting system
ACAS	Airborne collision avoidance system
ACM	Aircraft cycle machine
ACMS	Aircraft condition monitoring system
ADC	Analogue to digital converter
ADC	Air data computer
ADF	Automatic direction finder
ADI	Attitude director indicator
ADIRS	Air data/inertial reference system
ADS	Air data system
AEEC	Airlines electronic engineering committee
AES	Aircraft earth stations
AFC	Automatic frequency control
AFCS	Auto flight control system (autopilot)
AFS	Automatic flight system (autopilot)
AGL	Above ground level
AGM	Absorbent glass mat
AHRS	Attitude/heading reference system
AHS	Attitude heading system
AI	Airbus industries
AIAA	American Institute of Aeronautics and Astronautics
AIDS	Aircraft integrated data system
AIMS	Airplane information management system
AIS	Aeronautical information system
AIV	Anti-icing valve
ALT	Altitude
AMLCD	Active matrix liquid crystal display
AMM	Aircraft maintenance manual
AoA	Angle of attack
AP	Autopilot
APP	Approach
APU	Auxiliary power unit
ARINC	Aeronautical radio incorporated
ASCB	Aircraft system common data bus
ASI	Air speed indicator
ASM	Air separation module
ASR	Aerodrome surveillance radar
ATA	Actual time of arrival or Airline Transport Association (ATA) of America
ATC	Air traffic control
ATE	Automatic test equipment
ATFM	Air traffic flow management
AVLAN	Avionics local area network
AVOD	Audio visual on demand
AWG	American wire gauge (US spelling gage)
BC	Bus controller
BCD	Binary coded decimal
BCF	Bromochlorodifluoromethane
BCI	Bulk current injection
BGA	Ball grid array
BIC	Backplane interface controller
BIOS	Basic input/output system
BIST	Built in self-test
BIT(E)	Built in test (equipment)
BIU	Bus interface unit
BJT	Bipolar junction transistor
BOT	Begining of tape (FDR)
BPCU	Bus power control unit
BPV	Bypass valve
BSU	Bypass switch unit
BTB	Bus tie breaker
BTM	Bromotrifluoromethane
CABLAN	Cabin local area network
CAGE	Commercial and government entity
CAI	Computer-aided instruction
CAN	Controller area network
CAS	Collision avoidance system
CAS	Crew alerting system
CAT	Clear-air turbulence
CDI	Course deviation indicator
CDU	Control display unit

CFDS	Central fault display system
CFIT	Controlled flight into terrain
CIDS	Cabin intercommunication data system
CMC	Central maintenance computer
CMOS	Complementary metal oxide semiconductor
CMS	Centralized maintenance system
CNS/ATM	Communication, navigation, surveillance and air traffic
CO_2	Carbon dioxide
CPU	Central processing unit
CRT	Cathode ray tube
CSD	Constant speed drive
CVR	Cockpit voice recorder
CWT	Centre wing tank

DA	Drift angle
DAC	Digital to analogue converter
DC	Direct current
DDR	Digital data recorder
DDT	Direct drive testing
DECU	Digital engine control unit
DEOS	Digital engine operating system
DFDAU	Digital flight data acquisition unit
DFDR	Digital flight data recorder
DFLD	Database field loadable data
DH	Decision height
DIL	Dual in-line
DIMM	Dual in-line memory module
DIP	Dual in-line plastic
DLS	DME-based landing system
DMA	Direct memory access
DMC	Display management computer
DME	Distance measuring equipment
DO	Design organization
DPM	Data position module
DPU	Display processor unit
DRAM	Dynamic random access memory
DSP	Display select panel
DVD	Digital versatile disc

EADI	Electronic attitude director indicator
EATMS	Enhanced air traffic management system
ECAM	Electronic centralized aircraft monitoring
ECM	Electronic countermeasures
ECS	Environmental control system
ECU	Electronic control unit
EDP	Engine-driven pump
EEC	Electronic engine control
EEPROM	Electrically erasable programmable read-only memory
EFIS	Electronic flight instrument system
EGPWS	Enhanced ground proximity warning system
EHSI	Electronic horizontal situation indicator
EICAS	Engine indication and crew alerting systems
EMC	Electromagnetic compatibility
EMF	Electromotive force
EMI	Electromagnetic interference
EOT	End of tape (FDR)
EPROM	Erasable programmable read-only memory
EROPS	Extended range operations
ESD	Electrostatic discharge
ESSD	Electrostatic sensitive device
ETA	Estimated time of arrival
ETFE	Ethylene tetrafluoroethylene
ETOPS	Extended range twin-engine operations
EWIS	Electrical wire interconnection system

FAC	Flight augmentation computer
FBW	Fly-by-wire
FCC	Flight control computer
FCGC	Flight control and guidance computer
FCU	Flight control unit
FDAU	Flight data acquisition unit
FDC	Fight director computer
FDD	Floppy disk drive
FDDI	Fibre distributed data interface
FDE	Flight deck effect
FDR	Flight data recorder
FDS	Flight director system
FEP	Fluorinated ethylene-propylene
FET	Field effect transistor
FG	Flight guidance
FGC	Flight guidance computer
FIFO	First-in first-out
FIR	Flight information region
FIS	Flight information system
FL	Flight level
FLS	Field loadable software
FLIR	Forward-looking infrared
FLTA	Forward-looking terrain avoidance
FMC	Flight management computer
FMS	Flight management system
FOG	Fibre optic gyros
FPM	Feet per minute
FSD	Full scale deflection
FQIS	Fuel quantity indicating system
FWC	Flight warning computer

GA	General aviation
GCB	Generator circuit breaker
GCU	Generator control unit
GEO	Geostationary earth orbit
GND	Ground
GNSS	Global navigation satellite system

GPCU	Ground power control unit
GPM	Ground position module
GPS	Global positioning system
GPWS	Ground proximity warning system
GS	Ground speed or glide slope
HDG	Heading
HEIU	High-energy ignition unit
HERF	High-energy radiated field
HFDS	Head-up flight display system
HGS	Head-up guidance system
HIRF	High intensity radiated field
HIRF/L	High intensity radiated field/lightning
HIRL	High intensity runway lights
HM	Health monitoring
HP	High pressure
HSI	Horizontal situation indicator
HUD	Head-up display
HUGS	Head-up guidance system
HUMS	Health and usage monitoring system
I/O	Input/output
IAPS	Integrated avionics processing system
IAS	Indicated air speed
IC	Integrated circuit
IDE	Integrated drive electronics
IDG	Integrated drive generator
IFE	In-flight entertainment
IFR	Instrument flight rules
IHUMS	Integrated health and usage monitoring system
ILS	Instrument landing system
IMA	Integrated modular avionics
INU	Inertial navigation unit
INS	Inertial navigation system
IOAPIC	Input/output advanced programmable input controller
IOM	Input/output module
IP	Internet protocol
IPC	Instructions per cycle
IPR	Intellectual property rights
IR	Infrared
IRS	Inertial reference system
ISA	International standard atmosphere
ISAS	Integrated situational awareness system
ISU	Iridium subscriber unit
LAAS	Local area augmentation system
LAN	Local area network
LCC	Leadless chip carrier
LCD	Liquid crystal display
LE	Leading edge
LED	Light emitting diode
LEO	Low earth orbit
LFC	Laminar flow control
LGC	Landing gear control
LIDAR	Light radar
LP	Low pressure
LRRA	Low range radio altimeter
LRU	Line replaceable unit
LSAP	Loadable aircraft software part
LSB	Least significant bit
LSD	Least significant digit
LSI	Large scale integration
LVDT	Linear variable differential transformer
MAU	Modular avionics unit
MCDU	Microprocessor controlled display units
MEL	Minimum equipment list
MEMS	Micro-electromechanical systems
MFD	Multi-function flight display
MFDS	Multi-function display system
MSL	Mean sea level
MLS	Microwave landing system
MLW	Maximum landing weight
MMI	Man machine interface
MMM	Mass memory module
MMS	Mission management system
MNPS	Minimum navigation performance specification
MOS	Metal oxide semiconductor
MOSFET	Metal oxide semiconductor field effect transistor
MPU	Multifunction processor unit
MRO	Maintenance/repair/overhaul
MSB	Most significant bit
MSD	Most significant digit
MSI	Medium scale integration
MTBF	Mean time between failure
MTBO	Mean time between overhaul
MTOW	Maximum takeoff weight
MWS	Master warning system
NC	Normally closed
ND	Navigation display
NEA	Nitrogen-enriched air
NGS	Nitrogen generating system
NO	Normally open
NRZ	Non-return to zero
NTSB	National transport safety board
NVM	Non-volatile memory
OBI	Omni bearing indicator
OBS	Omni bearing selector
OEA	Oxygen enriched air
OEM	Original equipment manufacturer
OFD	Optical fire detector
OFV	Out flow valve

OS	Operating system
PA	Public address
PAPI	Precision approach path indicator
PBGA	Plastic ball grid array
PCMCIA	Personal computer memory card international association
PC	Personal computer
PCB	Printed circuit board
PDL	Portable data loader
PFD	Primary flight display
PGA	Pin grid array
PLCC	Plastic leadless chip carrier
POST	Power-on self-test
PP	Pre-processor
PQFP	Plastic quad flat package
PROM	Programmable read-only memory
PRSOV	Pressure relief shut-off valve
PSEU	Proximity switch electronic unit
PSM	Power supply module
PSU	Passenger service unit
PTFE	Polytetrafluoroethylene
PVfD	Polyvinylidene fluoride
PVS	Personal video screen
QAR	Quick access recorder
QFP	Quad flat package
RAM	Random access memory
RAT	Ram air turbine
RF	Radio-frequency
RFI	Radio-frequency interference
ROM	Read-only memory
RMI	Radio magnetic indicator
RMS	Root mean square
RNAV	Area navigation
RTD	Resistance temperature device
RZ	Return to zero
SCR	Silicon controlled rectifier
SDAC	System data acquisition concentrator
SDD	System definition document
SDI	Source/destination identifier
SIL	Single in-line
SIP	Single in-line package
SLA	Sealed lead acid
SG	Signal generator
SNOC	Satellite network operation center
SRAM	Synchronous random access memory
SRD	System requirement document
SROM	Serial read only memory
SPDT	Single pole double throw
SSI	Small scale integration
SSM	Sign/status matrix
STP	Shielded twisted pair
SWAMP	Severe wind and moisture problems
TACAN	Tactical air navigation
TAS	True air speed
TAWS	Terrain awareness warning system
TBO	Time between overhaul
TCAS	Traffic alert and collision avoidance system
TE	Trailing edge
TRU	Transformer rectifier unit
TTAC	Telemetry tracking and command/control
TTL	Transistor transistor logic
ULB	Underwater locator beacon
ULSI	Ultra-large-scale integration
UMS	User modifiable software
USB	Universal serial bus
UTP	Unshielded twisted pair
UV	Ultraviolet
VAS	Virtual address space
VFR	Visual flight rules
VGA	Video graphics adapter
VLSI	Very large scale integration
VRLA	Valve regulated lead acid
VSCF	Variable speed constant frequency
WAAS	Wide area augmentation system
WDM	Wiring diagram manual
WORM	Write-once read-many
WoW	Weight on wheels
WX	Weather
WXP	Weather radar panel
WXR	Weather radar

Appendix 2 Revision papers

These papers are intended to provide you with practice in preparation for examinations. Each paper contains 20 questions; these should be attempted without the use of a calculator or notes. The time allocation for each paper is 25 minutes.

Paper 1

1. Which one of the following gives the symbol and abbreviated units for resistance?
   (a) symbol, R; unit, Ω
   (b) symbol, V; unit, V
   (c) symbol, R; unit, A.
2. When testing a diode with an ohmmeter, what resistance will be measured between the anode and the cathode?
   (a) high resistance one way
   (b) high resistance both ways
   (c) low resistance both ways.
3. A DC generator produces an output of 28 V at 20 A. The power supplied by the generator will be:
   (a) 14 W
   (b) 560 W
   (c) 1.4 kW.
4. A silicon transistor has a base-emitter voltage of 0 V. In this condition the transistor will be:
   (a) conducting heavily
   (b) conducting slightly
   (c) turned off.
5. Battery capacity is measured in:
   (a) volts
   (b) amperes
   (c) ampere-hours.
6. TRUs are used to convert what forms of electrical power:
   (a) AC into DC
   (b) AC into AC
   (c) DC into AC.
7. For a given AWG, a wire will have a specified:
   (a) diameter and hence a known conductance
   (b) length
   (c) screening.
8. The main distribution bus is sometimes called the:
   (a) essential bus
   (b) emergency bus
   (c) non-essential bus.
9. Metal elements used in RTDs have a temperature coefficient that is:
   (a) negative, temperature increases cause an increase in resistance
   (b) positive, temperature decreases cause an increase in resistance
   (c) positive, temperature increases cause an increase in resistance.
10. Engine pressure ratio (EPR) is used to measure a gas turbine engine's:
    (a) torque
    (b) thrust
    (c) temperature.
11. When fuel level decreases, the capacitance of the fuel quantity sensor:
    (a) decreases and the reactance increases
    (b) Increases and the reactance increases
    (c) Increases and the reactance decreases.
12. The starboard wing tip navigation light is has the following colour and divergence:
    (a) red and 110 degrees
    (b) green and 110 degrees
    (c) white and ±70 degrees.
13. Satellite communication systems use a low earth orbit to:
    (a) provide greater coverage
    (b) maintain a geostationary position
    (c) minimize voice delays.

14. Indications of landing gear fully down and locked are:
    (a) red lights on, green lights off
    (b) red lights on, green lights on
    (c) red lights off, green lights on.

15. Stall warning systems provide the crew with a clear and distinctive warning:
    (a) before the stall is reached
    (b) after the stall is reached
    (c) at all angles of attack.

16. The action of twisting a fire handle closes micro-switches that:
    (a) activate the engine fire extinguisher
    (b) cancels the alarm
    (c) shuts off the fuel.

17. Red areas are used on TAWS displays to indicate terrain that is:
    (a) above the aircraft's current altitude
    (b) level with the aircraft's current altitude
    (c) safe in terms of required terrain clearance.

18. The flight data recorder must start to record data automatically:
    (a) after the aircraft is capable of moving under its own power
    (b) after take off
    (c) prior to the aircraft being capable of moving under its own power.

19. Shielding of conductors:
    (a) reduces radiation and minimizes susceptibility
    (b) increases radiation and minimizes susceptibility
    (c) reduces radiation and increases susceptibility.

20. For a given wire size, when installed in a loom it will be able to safely conduct:
    (a) more current compared to a wire in free circulating air
    (b) less current compared to a wire in free circulating air
    (c) the same current compared to a wire in free circulating air.

Paper 2

1. The relationship between power, P, current, I, and resistance, R, is:
   (a) $P = IR$
   (b) $P = \frac{R}{I}$
   (c) $P = I^2R$.

2. In order to make a silicon-controlled rectifier (thyristor) conduct a:
   (a) small current pulse must be applied to its gate
   (b) high voltage must be applied to its anode
   (c) load must be connected between its anode and cathode.

3. A two-input NAND gate with both inputs inverted will become which of these:
   (a) NOR gate
   (b) AND gate
   (c) OR gate.

4. Decreasing the current in the field coil of a generator will:
   (a) decrease the output current
   (b) increase the output voltage
   (c) decrease the output frequency.

5. Static electricity is discharged from the aircraft to atmosphere through:
   (a) composite structure
   (b) earth stations
   (c) static wicks.

6. When the lightning discharge is dissipated safely through the aircraft, structural damage is most likely to occur at the:
   (a) bonding between structures
   (b) exit points of the discharge
   (c) entry points of the discharge.

7. The mandatory parameters required for an aircraft DFDR depend on the:
   (a) speed and weight of the aircraft
   (b) maximum weight of the aircraft
   (c) size of the aircraft and the prevailing regulatory rules applied to that aircraft.

8. Mode 1 terrain awareness cautions are given for:
   (a) negative climb rate or altitude loss after take-off
   (b) excessive descent rate
   (c) altitude callout at 500 feet.

9. Open area smoke detectors rely on the transfer of particulate matter from the source of fire to the detector by:
   (a) convection
   (b) radiation
   (c) conduction.

10. In the automatic braking system, when a wheel is approaching a skid condition, this is detected when the speed is:
   (a) decreasing at a given rate
   (b) increasing at a given rate
   (c) constant.

11. Engine fire or loss of cabin pressure would be displayed on ECAM as:
   (a) level 3 failures
   (b) level 2 failures
   (c) level 1 failures.

12. Audio-video on demand (AVOD) entertainment enables passengers to:
   (a) pause, rewind, fast-forward or stop a programme
   (b) make phone calls via satellite communication
   (c) ignore PA system voice announcements and chime signals.

13. Incandescence is the radiation of light from:
   (a) a gas-discharge device
   (b) an electrical filament due to an increase in its temperature
   (c) a combined optical and electrical phenomenon.

14. The volume of fuel in a tank varies with temperature; as the temperature changes:
   (a) the mass and volume of fuel remains the same
   (b) the mass of fuel changes but the volume remains the same
   (c) the mass of fuel remains the same, but the volume changes.

15. Ground idle speed occurs when the engine has:
   (a) stabilized (slightly above self-sustaining speed)
   (b) stabilized (slightly below self-sustaining speed)
   (c) Just been started.

16. The thermocouple principle is based on the Seebeck effect, when heat is applied:
   (a) a change of resistance is measured
   (b) this causes the element to bend
   (c) an electromotive force (EMF) is generated.

17. The reverse current relay is needed on any DC generation system to prevent the:
   (a) battery from feeding excess current back through to the generator's armature
   (b) generator from feeding excess current back through to the battery
   (c) battery from feeding excess current to the starter motor.

18. Two or more separate wires within the same insulation and protective sheath is referred to as a:
   (a) screened wire
   (b) coaxial cable
   (c) cable.

19. The desired output frequency of a rotary inverter is determined by the:
   (a) AC input voltage
   (b) input frequency
   (c) DC input voltage.

20. The energy storage capacity of a cell is determined by the:
   (a) terminal voltage
   (b) electrolyte specific gravity
   (c) amount of material available for chemical reaction.

Paper 3

1. A cabin reading lamp consumes 10 W from a 24 V DC supply. The current supplied will be:
   (a) 0.42 A
   (b) 0.65 A
   (c) 2.4 A.

2. The output frequency of an AC generator driven at a constant speed:
   (a) increases with the number of pole pairs
   (b) decreases with the number of pole pairs
   (c) is independent of the number of pole pairs.

3. A compound-wound generator has:
   (a) only a series field winding
   (b) only a shunt field winding
   (c) both a series and a shunt field winding.

4. The decimal equivalent of the natural binary code 10110 is:
   (a) 12
   (b) 22
   (c) 38.

5. Lead acid batteries are recharged by constant:
   (a) voltage
   (b) current
   (c) ampere-hours.

6. Inverters are used to convert what forms of electrical power?
   (a) DC into DC
   (b) AC into AC
   (c) DC into AC.

7. Trip-free circuit breaker contacts:
   (a) can always be closed whilst a fault exists
   (b) cannot be closed whilst a fault exists
   (c) are only used during maintenance.

8. Different size pins are used on external DC power connectors to:
   (a) prevent a reverse polarity voltage being applied
   (b) prevent excessive power being applied
   (c) prevent power being applied when the battery is discharged.

9. The rotors of an torque synchro transmitter and receiver are supplied from:
   (a) the same power supply (normally 26V AC)
   (b) Different power supplies (normally 26V AC)
   (c) the same power supply (normally 26V DC).

10. HEIUs can remain charged for several:
   (a) seconds
   (b) minutes
   (c) hours.

11. Air and fuel have dielectrics of approximately:
   (a) unity and zero respectively
   (b) unity and two respectively
   (c) two and unity respectively.

12. Green or indicator blue lights in the instrument panel are used inform the crew that:
   (a) a safe condition exists
   (b) an unsafe condition exists
   (c) an abnormal condition exists.

13. The passenger address (PA) system is primarily a safety system that provides passengers with:
   (a) in-flight entertainment
   (b) reduced amount of IFE wiring to a seat position
   (c) voice announcements and chime signals.

14. An electrical flap drive system uses a:
   (a) reversible DC motor
   (b) variable-speed DC motor
   (c) unidirectional DC motor.

15. The stall identification system contains an actuator that:
   (a) maintains the angle of attack
   (b) pulls the control column rearward
   (c) pushes the control column forward.

16. When a pneumatic fire detector is rapidly heated; hydrogen is liberated causing:
   (a) sufficient gas pressure to close the alarm switch
   (b) the integrity switch to open
   (c) the alarm switch to reset.

17. Forward-looking terrain avoidance (FLTA) looks:
   (a) ahead of, and below the aircraft's lateral and vertical flight path
   (b) ahead of, and above the aircraft's lateral and vertical flight path
   (c) either side of, and below the aircraft's lateral and vertical flight path.

18. The DFDR on large aircraft has to be able to retain the recorded data for a minimum of the last:
   (a) 30 minutes of its operation
   (b) 25 hours of its operation
   (c) 25 flights.

19. The ability of an item of equipment to operate alongside other items of equipment without causing EMI is called:
    (a) EMC
    (b) ESSD
    (c) HIRF.

20. Visual inspection of wiring installations:
    (a) enables obvious faults to be identified and the appropriate action(s) taken
    (b) checks for adequate insulation resistance between conductors
    (c) checks the continuity of the earth return path back to the power supply(s).

Paper 4

1. A generator delivers 250 W of power to a 50 Ω load. The current flowing in the load will be:
   (a) 0.2 A
   (b) 5 A
   (c) 10 A.

2. The power factor in an AC circuit is defined as the ratio of:
   (a) true power to apparent power
   (b) apparent power to true power
   (c) reactive power to true power.

3. In a PNP transistor, conventional current:
   (a) flows into the collector
   (b) flows out of the collector
   (c) there is no current flow in the collector.

4. Gillham code from an altitude encoder is based on:
   (a) natural binary code
   (b) a form of Gray code
   (c) self-correcting hexadecimal code.

5. In terms of bonding, composite material has a:
   (a) high resistance and is unsuitable for bonding
   (b) low resistance and is unsuitable for bonding
   (c) high resistance and is suitable for bonding.

6. Higher frequency signals can lead to cross-talk between wires as a result of the:
   (a) dielectric effect of the wire insulation causing increased capacitance
   (b) dielectric effect of the wire insulation causing reduced capacitance
   (c) dielectric effect of the wire conductors causing reduced capacitance.

7. Lateral acceleration and radio altitude are typical parameters recorded on the:
   (a) FDR
   (b) CVR
   (c) ULB.

8. Premature descent alert compares the aircraft's:
   (a) ground speed with the proximity of the nearest airport
   (b) lateral and vertical position with the proximity of the nearest airport
   (c) lateral and vertical position with the proximity of high terrain.

9. To operate an engine fire extinguisher, the fire handle is:
   (a) twisted
   (b) twisted and then pulled
   (c) pulled and then twisted.

10. When the AoA reaches a certain angle, the airflow over the wing:
    (a) becomes turbulent and the lift is dramatically decreased
    (b) becomes streamlined and the lift is dramatically decreased
    (c) becomes turbulent and the lift is dramatically increased.

11. When an electrically operated landing gear is fully retracted, the up-lock switch contacts:
    (a) open thereby removing power from the motor
    (b) close thereby removing power from the motor
    (c) open thereby applying power to the motor.

12. Fibre optic cable bends need to have a sufficiently large radius to:
    (a) minimize losses and damage
    (b) maximize immunity to electromagnet interference (EMI)
    (c) accurately align the connector optical components.

13. The starboard wing tip navigation light is has the following colour and divergence:
    (a) red and 110 degrees
    (b) green and 110 degrees
    (c) white and ±70 degrees.

14. Intrinsic safety is a technique used for:
    (a) safe operation of electrical/electronic equipment in explosive atmospheres
    (b) ensuring fuel temperature does not become too low/high
    (c) reducing fuel quantity.

15. The starting sequence for a gas turbine engine is to:
    (a) turn on the ignition, develop sufficient airflow to compress the air, and then open the fuel valves
    (b) develop sufficient airflow to compress the air, open the fuel valves and then turn on the ignition
    (c) develop sufficient airflow to compress the air, turn on the ignition and then open the fuel valves.

16. When a foil strain gauge is deformed, this causes:
    (a) an electromotive force (EMF) to be generated
    (b) its electrical resistance to change
    (c) different coefficients of thermal expansion.

17. The split bus system is sometimes called a:
    (a) non-parallel system
    (b) parallel system
    (c) standby and essential power system.

18. The white collar just below a circuit breaker button provides visual indication of the:
    (a) circuit breaker trip current
    (b) system being protected
    (c) circuit breaker being closed or tripped.

19. The voltage coil of a carbon pile regulator contains a:
    (a) large number of copper wire turns connected across the generator output
    (b) low number of copper wire turns connected across the generator output
    (c) large number of copper wire turns connected in series with the generator output.

20. The only accurate and practical way to determine the condition of the nickel-cadmium battery is with a:
    (a) specific gravity check of the electrolyte
    (b) measured discharge in the workshop
    (c) check of the terminal voltage.

Appendix 3 Answers to multiple choice questions

Chapter 1 (Electrical fundamentals)

1. b
2. c
3. a
4. a
5. b
6. b
7. b
8. a
9. b
10. c
11. b
12. c
13. c
14. a
15. c

Chapter 2 (Electronic fundamentals)

1. b
2. b
3. c
4. c
5. a
6. a
7. b
8. a
9. c
10. b
11. a
12. a
13. a
14. b
15. c

Chapter 3 (Digital fundamentals)

1. b
2. c
3. b
4. a
5. a
6. a
7. a
8. b
9. a
10. c
11. b
12. c
13. b
14. c
15. a

Chapter 4 (Electrical machines)

1. a
2. a
3. a
4. a
5. a
6. a
7. b
8. b
9. b
10. b
11. a
12. b
13. b
14. c
15. c

Chapter 5 (Batteries)

1. a
2. c
3. a
4. b
5. b
6. a
7. b
8. c
9. a
10. a
11. b
12. c

Chapter 6 (Power supplies)

1. a
2. c
3. c
4. a
5. b
6. c
7. c
8. b
9. a
10. a

Chapter 7 (Wiring and circuit protection)

1. a
2. c
3. a
4. c
5. b
6. a
7. c
8. b
9. a
10. c

Chapter 8 (Distribution of power supplies)

1. a
2. b
3. c
4. a
5. c
6. a
7. c
8. b
9. a
10. c
11. c
12. b

Chapter 9 (Controls and transducers)

1. b
2. a
3. b
4. c
5. a
6. c
7. b
8. a
9. a
10. c
11. a
12. b
13. a
14. b
15. b

Chapter 10 (Engine systems)

1. b
2. a
3. a
4. a
5. c
6. a
7. b
8. a
9. a
10. c
11. b
12. c

Chapter 11 (Fuel management)

1. c
2. a
3. b
4. a
5. a
6. b
7. c
8. a
9. b
10. b

Chapter 12 (Lights)

1. a
2. b
3. b
4. a
5. a
6. b
7. c
8. b
9. c
10. a
11. a
12. b
13. c
14. c
15. b
16. a

Chapter 13 (Cabin systems)

1. b
2. a
3. b
4. a
5. c
6. a
7. c
8. a
9. b
10. a
11. a
12. c
13. a
14. c
15. a
16. c

Chapter 14 (Airframe control and indicating systems)

1. c
2. b
3. a
4. a
5. c
6. a
7. a
8. c
9. b
10. a

Chapter 15 (Warning and protection systems)

1. a
2. b
3. c
4. b
5. a
6. a
7. c
8. a
9. b
10. b

Chapter 16 (Fire and overheat protection)

1. a
2. c
3. a
4. a
5. b
6. c
7. a
8. a
9. a
10. c

Chapter 17 (Terrain awareness warning systems)

1. a
2. b
3. c
4. c
5. a
6. b
7. a
8. a
9. b
10. c
11. c
12. a

Chapter 18 (Flight data and cockpit voice recording)

1. a
2. c
3. b
4. b
5. c
6. a
7. b
8. a
9. a
10. c
11. b
12. a
13. c

Chapter 19 (Electrical and magnetic fields)

1. a
2. b
3. c
4. c
5. a
6. c
7. b
8. a
9. c
10. b

Chapter 20 (Continuing airworthiness)

1. a
2. b
3. c
4. a
5. a
6. b
7. c
8. a
9. b
10. c

Revision paper 1

1. a
2. a
3. b
4. c
5. c
6. c
7. a
8. c
9. c
10. b
11. a
12. b
13. c
14. c
15. a
16. a
17. a
18. c
19. a
20. b

Revision paper 2

1. c
2. a
3. c
4. a
5. c
6. b
7. c
8. b
9. a
10. a
11. a
12. a
13. b
14. c
15. a
16. c
17. a
18. c
19. c
20. c

Revision paper 3

1. a
2. a
3. c
4. b
5. a
6. c
7. b
8. a
9. a
10. b
11. b
12. a
13. c
14. b
15. c
16. a
17. a
18. b
19. a
20. a

Revision paper 4

1. b
2. a
3. b
4. b
5. a
6. b
7. a
8. b
9. c
10. a
11. a
12. a
13. b
14. a
15. c
16. b
17. a
18. c
19. a
20. b

Appendix 4

Electrical quantities, symbols and units

Quantity	*Symbol*	*Unit*	*Abbreviated units*
Angle	ϕ	radian or degree	Rad or °
Capacitance	C	farad	F
Charge	Q	coulomb	C
Conductance	G	siemen	S
Current	I	ampere	A
Energy	W	joule	J
Flux	Φ	weber	Wb
Flux density	B	tesla	T
Frequency	f	hertz	Hz
Impedance	Z	ohm	Ω
Inductance	L	henry	H
Power	P	watt	W
Reactance	X	ohm	Ω
Resistance	R	ohm	Ω
Time	t	second	s
Voltage	V	volt	V

This page intentionally left blank

Appendix 5

Electrical formulae

Charge, current and voltage

$$Q = I \times t$$

Ohm's law

$$V = I \times R \text{ and } I = V/R \text{ and } R = V/I$$

Similarly if *resistance* is replaced by *reactance* or *impedance*:

$$V = I \times X \text{ and } I = V/X \text{ and } X = V/I$$
$$V = I \times Z \text{ and } I = V/Z \text{ and } Z = V/I$$

Power and energy

$$P = I \times V \text{ and } P = V^2/R \text{ and } P = I^2R$$
$$J = P \times t \text{ and since } P = I \times V \text{ so } J = IVt$$

Resistors in series

$$R_T = R_1 + R_2 + R_3$$

Resistors in parallel

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

but where there are *only two* resistors,

$$R_T = \frac{R_1 \times R_2}{R_1 + R_2}$$

Capacitance

$$C = \frac{\varepsilon A}{d}$$

where ε is the *permittivity* of the dielectric and $\varepsilon = \varepsilon_0\,\varepsilon_r$.

Capacitance, charge and voltage

$$Q = CV$$

Inductance

$$L = n^2\frac{\mu A}{l}$$

where μ is the *permeability* of the magnetic medium and $\mu = \mu_0\,\mu_r$.

Energy stored in a capacitor

$$W = \tfrac{1}{2}CV^2$$

Energy stored in an inductor

$$W = \tfrac{1}{2}LI^2$$

Inductors in series

$$L_T = L_1 + L_2 + L_3$$

Inductors in parallel

$$\frac{1}{L_T} = \frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3}$$

but where there are *only* two inductors

$$L_T = \frac{L_1 \times L_2}{L_1 + L_2}$$

Capacitors in series

$$\frac{1}{C_T} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}$$

but where there are *only* two capacitors

$$C_T = \frac{C_1 \times C_2}{C_1 + C_2}$$

Capacitors in parallel

$$C_T = C_1 + C_2 + C_3$$

Induced e.m.f. in an inductor

$$e = -L\frac{di}{dt}$$

where $\frac{di}{dt}$ is the rate of change of current with time.

Current in a capacitor

$$i = C\frac{dv}{dt}$$

where $\frac{dv}{dt}$ is the rate of change of voltage with time.

Sine wave voltage

$$v = V_{max}\sin(\omega t) \text{ or } v = V_{max}\sin(2\pi ft)$$

$$\text{because} \quad \omega = 2\pi f$$
$$f = 1/T$$

where T is the periodic time.

For a *sine wave*, to convert:

RMS to **peak** multiply by **1.414**
Peak to **RMS** multiply by **0.707**
Peak to **average** multiply by **0.636**
Peak to **peak-peak** multiply by **2**

Capacitive reactance

$$X_C = \frac{V_C}{I_C} = \frac{1}{2\pi fC}$$

Inductive reactance

$$X_L = \frac{V_L}{I_L} = 2\pi fL$$

Resistance and reactance in series

$$Z = \sqrt{(R^2 + X^2)} \text{ and } \phi = \arctan\left(\frac{X}{R}\right)$$

Resonance

$$X_L = X_C \text{ thus } \omega L = \frac{1}{\omega C} \text{ or } 2\pi f_o L = \frac{1}{2\pi f_o C}$$
$$f_o = \frac{1}{2\pi\sqrt{LC}}$$

Power factor

$$\begin{aligned}\text{Power factor} &= \text{true power/apparent power}\\ &= \text{watts/volt-amperes} = \text{W/VA}\\ \text{True power} &= V \times (I \times \cos\phi) = VI\cos\phi\\ \text{Power factor} &= \cos\phi = R/Z\\ \text{Reactive power} &= V \times (I \times \sin\phi) = VI\sin\phi\end{aligned}$$

Motors and generators

$$F = BIl$$
$$f = pn/60$$

where p is the number of pole *pairs* and n is the speed in r.p.m.

Three phase

Star connection

$$V_L = 1.732 \times V_P \text{ and } I_L = I_P$$

Note that $1.732 = \sqrt{3}$

Delta connection

$$V_L = V_P \text{ and } I_L = 1.732 \times I_P$$

Power in a three-phase load

$$P = 3 \times V_P I_P \cos\phi = 1.732 \times V_L I_L \cos\phi$$

Appendix 6

Decibels

Decibels (dB) are a convenient means of comparing power, voltage and current in electrical and electronic circuits. They are also commonly used for expressing gain (amplification) and loss (attenuation) in electronic circuits. They are used as a *relative* measure (i.e. comparing one voltage with another, one current with another, or one power with another). In conjunction with other units, decibels are sometimes also used as an *absolute* measure. Hence dB V are decibels relative to 1 V, dB m are decibels relative to 1 mW, etc. The decibel is one-tenth of a bel which, in turn, is defined as the logarithm (base 10) of the ratio of power, P, to reference power, P_R. Provided that the resistances/impedances are identical, voltage and current ratios may be similarly defined, i.e. one-twentieth of the logarithm (base 10) of the ratio of voltage, V, to the reference voltage, V_R, or one-twentieth of the logarithm (base 10) of the ratio of current, I, to the reference current, I_R.

Decibels (dB)	*Power ratio*	*Voltage ratio*	*Current ratio*
0	1	1	1
1	1.26	1.12	1.12
2	1.58	1.26	1.26
3	2	1.41	1.41
4	2.51	1.58	1.58
5	3.16	1.78	1.78
6	3.98	2	2
7	5.01	2.24	2.24
8	6.31	2.51	2.51
9	7.94	2.82	2.82
10	10	3.16	3.16
13	19.95	3.98	3.98
16	39.81	6.31	6.31
20	100	10	10
30	1,000	31.62	31.62
40	10,000	100	100
50	100,000	316.23	316.23
60	1,000,000	1,000	1,000
70	10,000,000	3,162.3	3,162.3

This page intentionally left blank

Appendix 7

Wire and cable sizes

The sizing of most, if not all, aircraft wiring installations is based on the **American wire gauge (AWG)**. The size of wire (or gauge) is a function of its diameter and cross-sectional area. The cross-sectional area of each gauge is an important factor for determining its current-carrying capacity. Increasing gauge numbers give decreasing wire diameters.

Measurement of wire gauge is based on a linear measurement of one thousandth of an inch (**one mil**). A **square mil** is a unit of area equal to the area of a square with sides of one thousandth of an inch. This unit of area is usually used in specifying the area of the cross-section of rectangular conductors, e.g. bus bars or terminal strips.

The **circular mil** is a unit of area, equal to the area of a circle with a diameter of one mil. Note that a circular mil is not the true cross sectional area of a wire (given by πr^2). One circular mil is equal to the cross sectional area of a 0.001 in. diameter wire. (Although the term 'mil' is used, there is no relationship to any metric units.)

Circular mils are a convenient unit for comparing wires and cables, without the need to reference π. Circular mils were introduced into the aircraft industry from the USA and should be used with caution since potential error will result from the lack of π in its definition! This is particularly relevant when working with the resistivity of metals; these figures are based on the true cross-sectional area.

When the wire is contained within a bundle (or loom) the current-carrying capacity reduces since heat dissipation is reduced. The voltage drop along a given length of wire will also be a limiting factor for the current-carrying capacity; tables are published that allow the circuit designer to calculate the wire gauge for a given power supply voltage, wire length and whether the current is intermittent or continuous.

Table A7.1 Copper wire, American Wire Gauge (AWG)

Wire size (AWG)	*Diameter (mils)*	*Cross-section (Circular mils)*	*Maximum current (A) in free air*
20	32.0	1,024	11
18	40.3	1,624	16
16	50.8	2,581	22
14	64.1	4,109	32
12	81.0	6,561	41
10	102.0	10,404	55
8	128.5	16,512	73
6	162.0	26,244	101
4	204.0	41,616	135
2	258.0	66,564	181

Key maintenance point

The maximum current (A) in Table A7.1 is for training purposes and guidance only; always consult the approved data.

This page intentionally left blank

Appendix 8 ATA chapter/subsystem list

For larger aircraft, a standardized manual format was developed by the Airline Transport Association (ATA) of America called ATA Spec. 100. This specification contains format and content guidelines for technical manuals written by aircraft manufacturers and suppliers. Note that in 2000, ATA Spec 100 (documentation) and ATA Spec 2100 (for inventory management) were incorporated into ATA iSpec 2200. Many aircraft documents that were written via the ATA Spec. 100 will be in service for many years to come.

The following gives a list of ATA chapters and subsections within the chapter.

Chapter 21 Air Conditioning

21-00-00 General
21-10-00 Compression
21-20-00 Distribution
21-30-00 Pressurization control
21-40-00 Heating
21-50-00 Cooling
21-60-00 Temperature control
21-70-00 Moisture/air contamination

Chapter 22 Autoflight

22-00-00 General
22-10-00 Autopilot
22-20-00 Speed-attitude correction
22-30-00 Autothrottle
22-40-00 System monitor
22-50-00 Aerodynamic load alleviating

Chapter 23 Communications

23-00-00 General
23-10-00 Speech communication
23-20-00 Data transmission, auto. calling
23-30-00 Passenger address and entertainment
23-40-00 Interphone
23-50-00 Audio integrating
23-60-00 Static discharging
23-70-00 Audio & video monitoring
23-80-00 Integrated automatic tuning

Chapter 24 Electrical Power

24-00-00 General
24-10-00 Generator drive
24-20-00 AC generation
24-30-00 DC generation
24-40-00 External power
24-50-00 AC electrical load distribution
24-60-00 DC electrical load distribution

Chapter 25 Equipment & Furnishings

25-00-00 General
25-10-00 Flight compartment
25-20-00 Passenger compartment
25-30-00 Buffet/galley
25-40-00 Lavatories
25-50-00 Cargo compartments
25-60-00 Emergency
25-70-00 Accessory compartments
25-80-00 Insulation

Chapter 26 Fire Protection

26-00-00 General
26-10-00 Detection
26-20-00 Extinguishing
26-30-00 Explosion suppression

Chapter 27 Flight Controls

27-00-00 General
27-10-00 Aileron & tab
27-20-00 Rudder & tab
27-30-00 Elevator & tab

27-40-00 Horizontal stabilizer
27-50-00 Flaps
27-60-00 Spoiler, drag devices, fairings
27-70-00 Gust lock & damper
27-80-00 Lift augmenting

Chapter 28 Fuel

28-00-00 General
28-10-00 Storage
28-20-00 Distribution
28-30-00 Dump
28-40-00 Indicating

Chapter 29 Hydraulic Power

29-00-00 General
29-10-00 Main
29-20-00 Auxiliary
29-30-00 Indicating

Chapter 30 Ice & Rain Protection

30-00-00 General
30-10-00 Airfoil
30-20-00 Air intakes
30-30-00 Pitot and static
30-40-00 Windows, windshields & doors
30-50-00 Antennas & radomes
30-60-00 Propellers & rotors
30-70-00 Water lines
30-80-00 Detection

Chapter 31 Indicating & Recording Systems

31-00-00 General
31-10-00 Instrument & control panels
31-20-00 Independent instruments
31-30-00 Recorders
31-40-00 Central computers
31-50-00 Central warning systems
31-60-00 Central display systems
31-70-00 Automatic data reporting systems

Chapter 32 Landing Gear

32-00-00 General
32-10-00 Main gear & doors
32-20-00 Nose gear & doors
32-21-01 Actuators
32-30-00 Extension & retraction
32-40-00 Wheels & brakes
32-50-00 Steering
32-60-00 Position and warning
32-70-00 Supplementary gear

Chapter 33 Lights

33-00-00 General
33-10-00 General compartment
33-20-00 Passenger compartments
33-30-00 Cargo and service compartments
33-40-00 Exterior
33-50-00 Emergency lighting

Chapter 34 Navigation

34-00-00 General
34-10-00 Flight environment data
34-20-00 Attitude & direction
34-30-00 Landing & taxiing aids
34-40-00 Independent position determining
34-50-00 Dependent position determining
34-60-00 Flight management computing
34-70-00 ATC Transponder

Chapter 35 Oxygen

35-00-00 General
35-10-00 Crew
35-20-00 Passenger
35-30-00 Portable

Chapter 38 Water/waste

38-00-00 General
38-10-00 Potable
38-20-00 Wash
38-30-00 Waste disposal
38-40-00 Air supply

Chapter 51 Structure

51-00-00 General

Chapter 52 Doors

52-00-00 General
52-10-00 Passenger/crew

52-20-00 Emergency exit
52-30-00 Cargo
52-40-00 Service
52-50-00 Fixed interior
52-60-00 Entrance stairs
52-70-00 Door warning
52-80-00 Landing gear

Chapter 53 Fuselage

53-00-00 General

Chapter 54 Nacelles/pylons

54-00-00 General
54-10-00 Nacelle section
54-50-00 Pylon section

Chapter 55 Horizontal & vertical stabilizers

55-00-00 General
55-10-00 Horizontal stabilizer or canard
55-20-00 Elevator
55-30-00 Vertical stabilizer
55-40-00 Rudder

Chapter 56 Windows

56-00-00 General
56-10-00 Flight compartment
56-20-00 Passenger compartment
56-30-00 Door
56-40-00 Inspection & observation

Chapter 57 Wings

57-00-00 General
57-10-00 Centre wing
57-20-00 Outer wing
57-30-00 Wing tip
57-40-00 Leading edge and leading edge
57-50-00 Trailing edge and trailing edge
57-60-00 Ailerons and elevons
57-70-00 Spoilers

Chapter 61 Propellers

61-00-00 General
61-10-00 Propeller assembly
61-20-00 Controlling
61-25-01 Governor, propeller
61-30-00 Braking
61-40-00 Indicating
61-50-00 Propulsor duct

Chapter 71 Power plant

71-00-00 General
71-10-00 Cowling
71-20-00 Mounts
71-30-00 Fireseals
71-40-00 Attach fittings
71-50-00 Electrical harness
71-60-00 Air intakes
71-70-00 Engine drains

Chapter 72 Engine – turbine

72-00-00 General

Chapter 73 Engine fuel and control

73-00-00 General
73-10-00 Distribution
73-15-00 Divider flow
73-20-00 Controlling
73-25-00 Unit fuel control
73-30-00 Indicating

Chapter 74 Ignition

74-00-00 General
74-10-00 Electrical power supply
74-15-01 Box, ignition exciter
74-20-00 Distribution

Chapter 75 Air

75-00-00 General
75-10-00 Engine anti-icing
75-20-00 Cooling
75-30-00 Compressor control
75-35-01 Valve HP & LP bleed
75-40-00 Indicating

Chapter 76 Engine controls

76-00-00 General
76-10-00 Power control
76-20-00 Emergency shutdown

Chapter 77 Engine indicating

77-00-00 General
77-10-00 Power
77-20-00 Temperature
77-30-00 Analysers
77-40-00 Integrated engine instrument systems

Chapter 78 Exhaust

78-00-00 General
78-10-00 Collector/nozzle
78-20-00 Noise suppressor
78-30-00 Thrust reverser
78-40-00 Supplementary air

Chapter 79 Oil

79-00-00 General
79-10-00 Storage
79-20-00 Distribution
79-30-00 Indicating

Chapter 80 Starting

80-00-00 General
80-10-00 Cranking

Appendix 9 Electrical and electronic symbols

Batteries

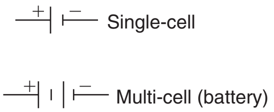
+
−
Single-cell
+
−
Multi-cell (battery)

Audible devices

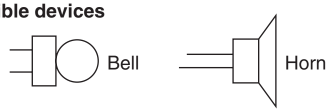

Busbar

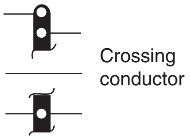
Crossing
conductor

Test jack

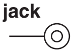

Slip ring

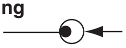

Connector test point

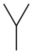

Single pin connector

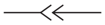

Resistors

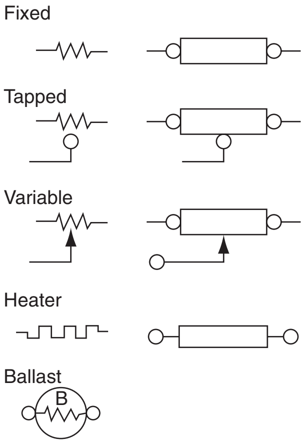
Fixed
Tapped
Variable
Heater
Ballast
B

Fuse

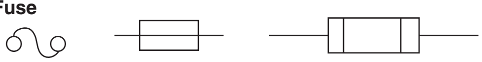

Gearbox

Grounds

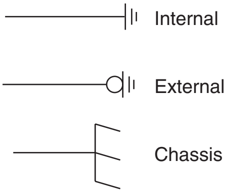
Internal
External
Chassis

Warning lights

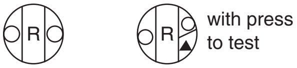
R
R
with press
to test

Meters

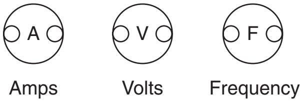
A
V
F
Amps
Volts
Frequency

Shunt

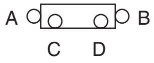

Complete connector

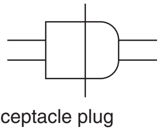
Receptacle plug

Wires

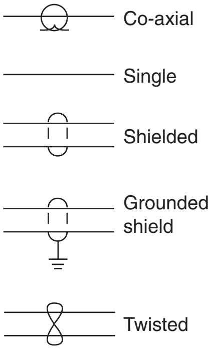
Co-axial
Single
Shielded
Grounded
shield
Twisted

Semiconductors

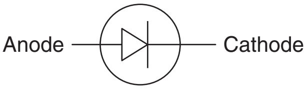

Diode

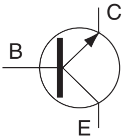

NPN transistor

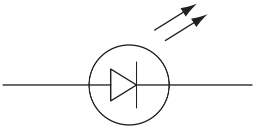

Light emitting diode

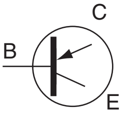

PNP transistor

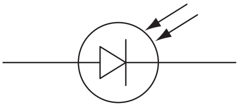

Photo sensitive diode

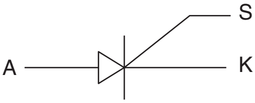

Thyristor

Transformers

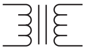

Basic

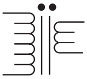

Step-down

• No phase shift

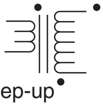

Step-up

• Phase shift 180°

Auto

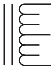

Fixed

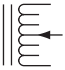

Variable

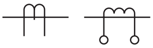

Current

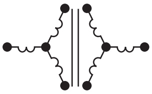

Wye-Wye

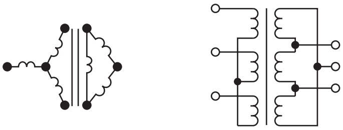

Wye-Delta

Thermal devices

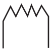

Sensing element

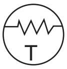

Thermal resistor

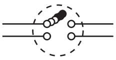

Thermal relay
with time delay

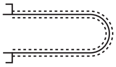

Continuous loop
detector

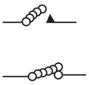

N.O.

N.C.

Thermal switch

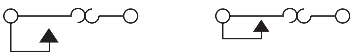

Contacts N.O.

Contacts N.C.

Thermal overlaod

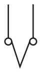

Thermocouple

Switches

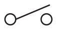
S.P.S.T.

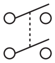
D.P.S.T.

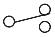
S.P.D.T.

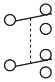
D.P.D.T.

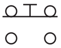
Push-pull

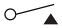
Off-Momentary ON

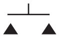
Push ON

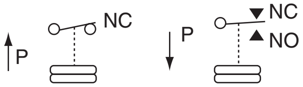

Pressure

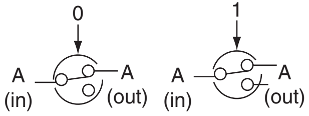

Solid-state

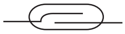
Reed

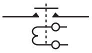
Push with
hold-in

Relays

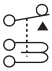
S.P.D.T.

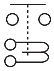
S.P.S.T.

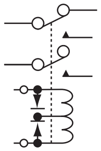
D.P.D.T. with
a.c. excitation

Antennas

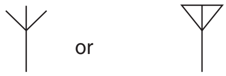

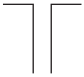
Dipole

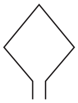
Loop

Circuit breakers

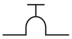
Trip free

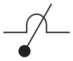
Switch type

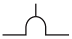
Automatic trip
push to reset

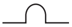
Automatic reset

Logic gates

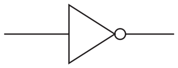
Inverter

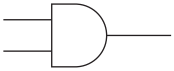
AND

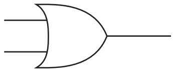
OR

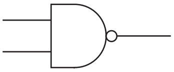
NAND

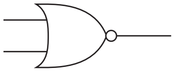
NOR

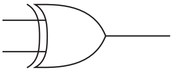
Exclusive OR

This page intentionally left blank

Appendix 10

Wire numbering/coding

Wires and cables should be identified to facilitate installation, trouble-shooting and potential modifications. There are various specifications that provide details of how wiring identification is implemented, including those adopted by the aircraft manufacturers. The simplest form of identifying cables is to mark the ends of the wires with the **source** and **destination** of the individual wire or cable. The wire/cable insulation is marked with indelible ink or laser printing. Wires are sometimes identified by the system type, wire number and gauge. Some manufacturers mark the cables at intervals along the cable. On more complex installations, where cables are bundled into looms, the entire loom may be given an identification code; this would be marked on a sleeve or band. The ATA specification is illustrated here to describe the principles of wire coding.

The code consists of a six-position combination of letters and numbers that is marked on the wire or cable, and is documented in the wiring diagram manuals. If the number cannot be marked on the wire or cable, e.g. due to its small diameter, it is printed onto sleeves that are fitted to the ends of the wire or cable. The following examples illustrate the principles of coding; the wire code would be marked thus:

2 P 4 B 20 N

Each of the six positions has a specific purpose:

$$2_1\ P_2\ 4_3\ B_4\ 20_5\ N_6$$

Position 1 indicates the system number, e.g. in a twin-engine aircraft; this would be the wiring associated with the number two engine.

Position 2 indicates the function of the circuit or system, e.g. P = electrical power. Other examples of circuit functions/system codes are given below.

Position 3 indicates the sequential number of wires in the same circuit; three wires used by the stator windings of synchro system would be numbered 1, 2 and 3.

Position 4 indicates the segment of wiring in a circuit, i.e. that portion of wire between two connections. This normally starts with A and builds up sequentially through the circuit.

Position 5 indicates the AWG size of the wire or cable; on certain cable types, e.g. coaxial and thermocouples, this number is not used, and is replaced by a (-).

Position 6 indicates whether the wire is being used as either: a connection to a ground/neutral point, an AC supply or a thermocouple. Codes used for position 6 are as follows:

N	ground/neutral
V	single-phase AC supply
A, B, C	three-phase AC supply
AL	alumel thermocouple
CH	chromel thermocouple
CU	aluminium thermocouple
CN	constantan thermocouple

Examples of circuit functions/system codes:

C	Flight control
D	De-icing/anti-icing
E	Engine instruments
F	Flight instruments
H	Heating/ventilation
J	Ignition
K	Engine control
L	Lighting
M	Miscellaneous
P	Power
Q	Fuel/oil
R	Radio navigation/communication
V	Inverter
W	Warning devices
X	AC power

This page intentionally left blank

Index

A/C pack, 250
AC generator, 82, 89, 91, 92, 104, 105, 106, 107
AC motor, 95, 96, 97
AC ratiometer, 174, 175
AC torque synchro, 180, 181, 182
ACARS, 367
ACM, 249, 250, 251
ADC, 55
AES, 245, 246
AGM, 117
AIDS, 328
AMLCD, 207, 262
AMM, 7
AND logic, 59
APU, 105, 111, 134, 135, 191
APU fire detection, 284
APU start, 191, 192
APU starter logic, 64
ARINC 429, 74, 75, 213
ARINC 631, 75
ATA chapter, 363, 391
ATA specification, 363
ATC, 299
ATE, 366
AVO, 342
AVOD, 244
AWG, 139, 389
Abbreviations, 369
Absorbent glass mat, 117
Accretion, 271
Acronymns, 369
Active matrix liquid crystal display, 262
Active-low, 63
Address bus, 76
Aerofoil, 265, 266
Air conditioning, 249, 250, 251
Air cycle machine, 249
Air distribution system, 193
Air traffic control, 299
Aircraft Maintenance Manual, 7
Aircraft battery, 111
Aircraft communication addressing and reporting system, 367
Aircraft earth station, 245
Aircraft integrated data system, 328
Aircraft manual, 362
Airline Transport Association, 391
Airstairs, 254
Alarm switch, 289
Alert, 262
Alpha radiation, 292
Alternating current, 19, 20
Alternator, 89, 105
Alternator master switch, 130
Alternator schematic, 149
Altitude, 305
Altitude callout, 306
Aluminium wire, 359
American wire gauge, 139, 389
Amplifier, 54
Amplitude, 21
Analogue multimeter, 364
Analogue signal, 55
Analogue to digital converter, 55
Angle of attack, 265, 266, 267
Anode, 38, 111, 112
Anti-collision light, 236, 238
Anti-icing, 275, 276
Anti-skid, 277, 280
AoA, 265, 266, 269
AoA sensor, 270
Apparent power, 27, 162
Armature, 85, 88, 168, 170
Assisted recovery, 318
Asynchronous, 74
Atom, 1
Attitude, 306
Audio entertainment system, 243
Audio-video on demand, 244
Aural warning, 281
Auto-transformer, 133, 134
Automatic test equipment, 366
Autoranging, 365
Autorotation, 316
Autosyn, 181
Auxiliary power unit, 105, 111, 134, 191
Average value, 21
Averaging linear detector, 286
Avionic bus, 148
Avometer, 364, 365

BCD, 66, 68, 69, 330, 331, 332
BCF, 283
BCI, 346
BITE, 366, 367
BJT, 49, 50
BJT circuit configurations, 53
BNC, 141
BPCU, 162
BPS, 342
BPV, 220
BTB, 154, 162
BTM, 283
Baggage compartment, 293
Balanced load, 93
Bandwidth, 341
Bank angle, 313
Bar magnet, 14
Battery, 6, 111, 113, 120
Battery bus, 152
Battery busbar, 155
Battery charging, 150, 156
Battery compartment, 123
Battery connection, 123, 124
Battery location, 122, 123
Battery maintenance, 118
Battery master switch, 147
Battery pack, 150
Battery venting, 122, 124
Battery-backed memory, 77
Beacon light, 236, 238
Bi-metallic strip, 175, 176
Bi-phase rectifier, 46
Bias, 49, 50, 54
Bidirectional bus, 72, 73
Bimetallic element, 144
Bimetallic principle, 285
Binary, 68, 76
Binary coded decimal, 66, 330, 332
Binary to Gray code conversion, 68
Bipolar junction transistor, 49
Bipolar return to zero format, 329
Bistable, 65, 66
Bits per second, 342
Bladder tank, 213
Bleed air, 279
Bleed air leak, 283
Blow out, 195
Bohr model, 1
Bonding, 347, 348, 359
Bonding meter, 365
Bonding test, 363
Boost pump, 218
Boost resistor, 131
Boots, 274
Boundary layer, 267

Bourdon tube, 174, 206
Breaker, 159
Bridge rectifier, 47, 48
Bromochlorodifluoromethane, 283
Bromotrifluoromethane, 283
Brushes, 82
Brushless generator, 106, 107
Buffer, 59, 248
Built-in test equipment, 366
Bulk current injection, 346
Bulk erase, 334
Bulkhead loom, 355
Bus, 152, 154
Bus cable, 74
Bus coupler, 72
Bus speed, 75
Bus system, 72, 73
Bus terminator, 74
Bus tie breaker, 154
Busbar schematic, 149
Bypass valve, 220

CAGE, 140
CAS, 208
CDU, 367
CFIT, 299, 301
CHT, 201
CIVIL, 23, 24
CMC, 367
CMOS RAM, 77
CMS, 367
CPA, 318
CPU, 76, 77
CRT, 207, 262, 324
CSD, 107
CVR, 322, 332, 334
CWT, 221, 222, 223
Cabin altitude, 254
Cabin light, 230
Cabin pressurization, 253
Cabin zone, 250
Cable, 137, 140, 353, 359, 389
Cable loom, 353, 354, 355
Cabling, 347
Camber, 265
Capacitance fringing, 218
Capacitance fuel quantity sensor, 216
Capacitive fuel measurement system, 215
Capacitive reactance, 23, 24, 25
Capacitor, 3, 4, 6, 23
Capacitor run, 103
Capacitor start, 103
Capacitor starting, 102
Capacity, 112
Capacity check, 116
Capsule, 174, 206
Carbon dioxide, 284
Carbon-pile regulator, 128, 129, 130
Cargo bay, 291, 293
Cargo bay fire extinguisher, 296, 297
Cargo compartment, 254
Cartridge fuse, 143
Cathode, 38, 111, 112
Cathode ray tube, 262
Caution alerting system, 208
Caution light, 228
Caution panel, 231
Cell, 6, 111, 112, 114
Cell capacity, 112
Central processing unit, 76
Centralized maintenance computer, 367
Centralized maintenance system, 367
Centre wing tank, 221, 222
Centrifugal compressor, 134
Changeover relay, 169
Charge, 2, 4, 6
Charge carrier, 2, 35
Charging, 114, 117, 150, 156
Circuit breaker, 144, 145
Circuit breaker panel, 145
Circuit breaker symbols, 146
Circuit protection, 142
Circuit testing, 363
Circuit trip, 144
Cladding, 248
Class A, 54
Class of fire, 284, 293
Clock, 76, 77
Coaxial cable, 141, 142
Coaxial connector, 142
Cockpit voice recorder, 321, 332
Coding, 399
Cold flow, 356
Cold junction, 180
Collector characteristic, 51
Collision prediction and alerting, 318
Combinational logic, 60
Combustion, 283
Commercial and government entity, 140
Common base, 52
Common base amplifier, 53, 54
Common collector, 52
Common collector amplifier, 53, 54
Common emitter, 52
Common emitter amplifier, 53, 54
Common emitter current gain, 52
Commutator, 83, 84, 86, 87
Compensation unit, 217
Complex wave, 20
Composite material, 359, 360
Compound-wound, 88
Compressor, 250, 254
Computer, 76, 77
Conductor, 2, 8
Conduit, 354, 355
Cone of vision, 290
Configuration warning, 279, 281
Connector, 355, 356, 358
Constant speed drive, 107
Contact armature, 168
Contactor, 159, 168
Continuity test, 363
Control bus, 76
Control display unit, 367
Control surfaces, 261
Controlled flight into terrain, 299
Conventional current, 6
Cooling, 252
Copper loss, 30
Core, 248
Corona discharge, 360
Coulomb, 2
Couple, 85
Coupler panel, 74
Crash-survivable, 321
Cross-talk, 248, 340, 341
Current, 2, 6, 7
Current gain, 52, 54
Current limiter, 143, 148, 150
Current mode, 74
Current protection, 156
Current regulation, 128
Current transformer, 106, 109, 156, 158
Cut-off, 54
Cylinder head temperature, 201

D-type bistable, 66
DAC, 55
DC, 6
DC circuit, 8
DC generator, 82, 84, 104
DC motor, 87
DC ratiometer, 177, 212
DC synchro, 180, 181
DCU, 209
DDT, 346
DFDR, 323, 324, 327, 328, 329
DITS, 74
DVD, 243, 244
Data bus, 74, 76, 328
Data communication, 244
Data concentrator unit, 209
Data recording formats, 329
Data selector, 70, 71, 77
De-icing fluid, 274
Decibels, 387
Decoder, 66, 68
Decompression, 254
Deep stall, 271
Defuelling, 221
Delta connection, 93, 94, 95
Depletion region, 38
Desyn, 180
Detection loop, 285
Dielectric, 3
Differential bus, 74
Differential current protection, 158

Differential protection, 156, 158
Digital data bus, 328
Digital flight data recorder, 323
Digital multimeter, 364, 365
Digital signal, 55, 341
Digital to analogue converter, 55
Digital versatile disk, 243
Diode, 38, 39, 40, 41
Diode characteristics, 39, 40
Diode symbols, 41
Direct bonding, 359, 360
Direct cranking, 189
Direct current, 6, 20
Direct drive testing, 346
Discharge, 6, 118, 360
Discharge characteristic, 115, 118, 119, 121
Discharge plug, 295
Discharger, 3
Discharging, 114, 118
Discrete component, 55
Discrete linear detector, 286
Distribution, 147, 161
Distributor, 185
Dome light, 226
Doping, 36
Doppler radar, 312, 315
Doppler shift, 312
Downdraught, 310, 315
Drag cup, 196
Ducted loom, 353, 355
Ducted smoke detector, 291

E-I sensor, 172, 173
E-field, 339
E.m.f., 7
ECAM, 206, 209, 210, 262, 263, 367
ECS, 249
EDP, 218
EFIS, 324
EGPWS, 299
EGT, 198, 208, 210
EICAS, 206, 207, 208, 209, 210, 262
EMC, 342
EMI, 109, 225, 248, 337, 338, 340, 342
EMI reduction, 342
EMI susceptibility, 342
EPR, 201, 208, 210
EPR transducer, 202
ESD, 3
ESD warning notices, 350
ESSD, 337, 348, 350
ETFE, 139
EWIS, 137, 353
Earth loop, 362
Earth return circuit, 361
Earth station, 362
Eddy current, 196
Eddy current loss, 30
Effective value, 21
Efficiency, 30
Eight to three line encoder, 70
Electric charge, 2, 4
Electric current, 2, 6, 7
Electric field, 3, 4, 339
Electric shock, 32
Electrical formulae, 385
Electrical noise, 343
Electrical power breaker, 159, 160
Electrical power panel, 155
Electrical quantities, 383
Electrical starting system, 186
Electrical symbols, 383, 395
Electrical system synoptic, 263
Electrical units, 383
Electrical wire interconnection system, 137, 353
Electroluminescence, 225
Electrolysis, 117
Electrolyte, 114, 115, 139
Electrolyte spillage, 119
Electromagnetic circuit breaker, 144
Electromagnetic compatibility, 337, 342
Electromagnetic induction, 16
Electromagnetic interference, 225, 248, 337, 342
Electromagnetic interference filter, 109
Electromagnetic wave, 339
Electromagnetism, 14
Electromotive force, 7
Electron, 1, 2
Electronic centralized aircraft monitoring, 206, 262
Electronic indicating systems, 206
Electronic symbols, 395
Electronic voltage regulator, 130
Electrostatic effect, 1
Electrostatic sensitive device, 3, 337, 348
Electrostatic voltage, 349
Electrostatics, 3
Elevation, 305
Emergency exit light, 230
Emergency light, 226
Emergency power, 136
Emitter follower, 52
Encoder, 69
Energy, 12
Engine electrical systems, 185
Engine fire detection, 284
Engine fire extinguisher, 295
Engine indicating and crew alerting system, 206, 262
Engine indicating system, 195
Engine instruments, 195
Engine pressure ratio, 201
Engine speed indication, 197
Engine start, 192
Engine temperature, 198, 199, 200
Engine torque indicator, 205
Engine-driven pump, 218
Enhanced ground proximity warning system, 299
Environment, 137
Environmental control system, 249
Equalizing system, 153
Equipment bonding, 348
Equipment cooling, 252
Erase head, 328
Essential bus, 152
Essential load, 111
Essential power, 155
Ethernet, 75
Ethylene tetrafluoroethylene, 139
Eutectic salt, 288
Exciter generator, 107
Exclusive-NOR, 60
Exclusive-OR, 60
Exhaust gas temperature, 198
Exit light, 230
Exterior light, 232, 234, 235
External power, 131, 150, 151, 152, 156
External reference, 303
Extinguisher, 294, 295

FDAU, 323, 324
FDDI, 75
FDE, 367
FDR, 321, 334
FDR parameters, 324, 325
FDR specifications, 326
FEP, 139
FLTA, 299, 300, 302, 313, 316
FLTA runway, 317
FQIS, 215
FSHR, 192
Failure mode, 139
Faraday cage, 344
Faraday shield, 344
Faraday's law, 17
Fibre distributed data interface, 75
Fibre optic, 248, 249
Fibre optic cable, 248
Fibre optic connector, 248, 249
Field, 88
Field connections, 87
Field strength, 344
Field winding, 86
Fire bottle, 295
Fire detection, 284, 286
Fire detector loop, 287
Fire extinguisher, 294, 295, 296
Fire protection, 283
Flameout, 193
Flaps, 260, 261
Fleming's rule, 86, 87
Flight data acquisition unit, 323
Flight data recorder, 321

Flight deck effect, 367
Flight idle speed, 193
Flight path, 313, 316
Float gauge, 212
Float stick, 215
Float valve, 212
Flood light, 226
Flooded cell, 113
Floor path lighting, 230
Fluid pressure measurement, 206
Fluid pressure transducer, 176
Fluorescent lamp, 225
Fluorinated ethylene-propylene, 139
Flux density, 15, 338
Foil recorder, 321
Formulae, 385
Forward conduction voltage, 39
Forward looking terrain avoidance, 299, 313
Forward-bias, 39
Four to one multiplexer, 71, 72
Frangible diaphragm, 295
Free electron, 2
Frequency, 20, 21
Fuel, 213
Fuel distribution, 218, 219, 220
Fuel dumping, 221
Fuel feed, 218
Fuel flow indication, 201, 203, 204
Fuel flow metering, 203
Fuel indication, 213
Fuel jettison, 221, 222
Fuel quantity indicating system, 215
Fuel quantity measurement, 213, 214, 215, 216, 217
Fuel solenoid holding relay, 192
Fuel tank, 217
Fuel tank venting, 221
Fuel temperature, 220
Fuel temperature indicating system, 205
Fuel transfer, 220
Full-load, 30
Full-wave rectifier, 45, 47, 48
Functional test, 363
Fuse, 142
Fuse construction, 143
Fuse holder, 143
Fuse location, 144
Fuse panel, 144
Fuse symbol, 143
Fusible link, 142

GCAM, 318
GCB, 154
GCCU, 109
GCR, 109, 158
GCU, 106, 108, 109, 158
GEO, 246
GPCU, 162
GPS, 246, 300
GPWS, 299, 301, 303
GPWS Mode 1, 306
GPWS Mode 2, 307, 309
GPWS Mode 3, 307, 309
GPWS Mode 4, 307, 310, 311
GPWS Mode 5, 308, 312
GPWS Mode 6, 308, 313
GPWS Mode 7, 308, 315
Galley equipment, 243
Gas turbine engine, 189
Gasper fan, 252
Gate, 59
Ge, 36
Generated e.m.f., 83, 84, 85
Generator, 81, 82, 84, 153, 154
Generator circuit breaker, 154
Generator control logic, 65
Generator control relay, 109
Generator control unit, 106, 108, 109, 158
Generator conversion control unit, 109
Generator output, 160
Geostationary earth orbit, 246
Germanium, 36, 40
Gillham code, 69
Gillham interface, 69
Glide slope, 305
Global positioning system, 246
Go/no-go status, 367
Gray code, 66, 68
Gray code to binary conversion, 68
Ground collision avoidance module, 318
Ground idle speed, 192
Ground loop, 340, 362
Ground power, 151, 152
Ground power available, 154
Ground power relay, 150
Ground proximity warning light, 306
Ground proximity warning system, 299
Grounding, 347, 348, 350

H-field, 339
HEIU, 190, 192, 195
HERF, 337, 343
HIRF, 337, 343, 344, 346
HIRF environment, 347
HIRF threat, 346
HUMS, 334
Half-wave rectifier, 43, 44, 45
Halogenated hydrocarbon, 283
Halon, 283
Hand-held fire extinguisher, 295
Harvard format, 329
Head wind, 310, 315
Health and usage monitoring system, 334
Height, 305
Hertz, 20
Hexadecimal, 66, 68
High state, 76
High-energy ignition unit, 190
High-energy radiated field, 337, 343
High-intensity radiated field, 337, 343
Holdover time, 274
Hole, 36
Hot battery bus, 155, 156
Hot junction, 178
Hydrocarbon fire, 290
Hydrolysis, 354
Hysteresis loss, 30

I/O, 76, 77
IDG, 107, 153
IFE, 243, 244, 245, 248
ILS, glide slope, 308
IR, 284, 290
IR lossses, 137
IR test, 363
IRS, 306
ISU, 245
Ice detection, 271
Ice detector, 272, 273
Ice inspection light, 233
Ice protection, 271, 272, 274, 275
Igniter, 191
Ignition cable, 187
Ignition system, 194
Ignition unit, 190
Impedance, 25
Impedance bridge, 216
Impedance triangle, 25
Impurity atom, 35, 36
Impurity element, 36
In-flight entertainment, 243
In-flight start, 190
In-line splice, 141
Incandescence, 225
Indicating cap, 142
Indirect bonding, 359, 360
Induced e.m.f., 16, 17, 81
Inductance, 28
Inductance loop, 167
Induction, 16
Induction motor, 96, 100, 101
Inductive reactance, 23, 24
Inductor, 14, 19
Inertial reference system, 306
Infrared, 284, 290
Input characteristic, 50, 51
Input/output, 76
Instrument landing system, 305
Instrument lighting, 227
Insulation resistance test, 363

Insulator, 2
Integral fuel tank, 213
Integrated circuit, 55, 56
Integrated drive generator, 107
Integrity switch, 289
Inter-turbine temperature, 198
Internal resistance, 112, 117
Internet, 244
Interphone system, 241
Intrinsic safety, 218
Inverted input, 60, 61
Inverted output, 60, 61
Inverter, 100, 131, 132
Inverting input, 55
Ion, 1
Ionization smoke detector, 291
Ionized gas, 225
Iridium, 244, 246, 247
Iridium subscriber unit, 245
Iron loss, 30
Isolated charge, 4
Isotope, 292

J-K bistable, 66, 67
JPT, 198
Jet pipe temperature, 198
Jettison pump, 219
Joule, 12
Junction, 177, 178
Junction diode, 38

KWh, 12
Kevlar, 249
Kilobyte, 77
Kilowatt, 12
Kirchhoff's laws, 10, 11
Knife-edge cutter, 271, 272
Ku-band, 247

L-band, 246
LAN, 248
LCD, 244, 324
LED, 43, 225, 230, 237, 248, 292
LEO, 246
LP cock, 218
LRRA, 304
LRU, 72, 73, 74, 356, 357, 366
LVDT, 170, 172, 201
Landing gear, 257, 258, 259
Landing gear warning logic, 60, 62
Landing light, 233, 234
Lead-acid battery, 113, 114
Lead-acid cell, 111
Leading edge, 265
Leak, 116, 119
Leaning, 201
Left-hand rule, 87
Lenz's law, 17, 99
Level 1 failure, 263
Level 2 failure, 263
Level 3 failure, 262
Lift, 267
Light, 234
Light control, 227
Light refraction, 215
Light-emitting diode, 43, 225
Lighting, 225, 230, 232
Lightning, 344
Lightning strike, 345
Lightning zone, 345
Lights, 225
Limiting resistor, 146
Line current, 94, 94, 95
Line replaceable unit, 72, 356, 366
Line voltage, 94, 95
Linear displacement transducer, 170
Linear fire detector, 286, 287
Linear variable differential transformer, 170, 172, 201
Liquid crystal display, 244
Lithium battery, 111, 120
Lithium-ion battery, 120
Load, 8, 30, 111
Load sharing, 162
Load shedding, 152, 155, 159
Local area network, 248
Local bus, 72
Localizer, 305
Logic, 59, 63
Logic 0, 76, 331
Logic 1, 76, 331
Logic gate, 59
Logic gate symbol, 59
Logo light, 232
Look ahead distance, 313
Loom, 353
Loop, 15, 340
Low earth orbit, 246
Low state, 76
Low-pressure cock, 218
Low-range altimeter, 304

MFD, 303
MIL/ANSI symbols, 59
MM, 363
MSL, 305
Magnesyn, 181, 182
Magnetic field, 14, 15, 16, 81, 86, 339, 340
Magnetic flux, 15
Magneto, 185
Magneto ignition, 185, 186, 188
Main bus, 152
Main engine start, 192
Main tank, 222
Maintenance manual, 363
Majority vote, 62
Manual, 362
Mark 33 digital information transfer system, 74
Mass flow, 202
Master switch, 147
Master warning circuit, 230
Master warning light, 228, 285
Maximum reverse voltage, 39
Maximum value, 21
Mean camber line, 265
Mean sea level, 305
Measured discharge, 118
Memory, 76
Memory effect, 112
Merz Price protection, 158, 159
Mesh, 10
Metallic hydride compound, 289
Metallic surface bonding, 348
Metering vane, 201
Microburst, 310, 314, 315
Microswitch, 166, 167, 257
Molecular sieve, 223
Monostable, 63
Montreal Protocol, 283
Motor, 81, 85, 86, 95
Motor field connections, 87
Motor-driven impeller, 202
Moving map system, 244
Multi-stranding, 137
Multimeter, 364
Multiphase supply, 22
Multiple fuel tanks, 217
Multiplexer, 70, 71
Multiplexing, 247
Mutual inductance, 28, 19

N1, 208, 210
N1 engine speed indication, 197
N2 engine speed indication, 198
NAND logic, 60
ND, 303
NEA, 223
NOR logic, 60
NPN, 49, 50
NRZ, 329
NTC, 37
NVM, 208, 332
Natural binary, 66, 68
Navigation display, 303
Navigation light, 235, 236, 238
Near miss, 69
Negative charge carrier, 35
Negative ion, 1
Negative ramp, 20
Neutron, 1
Ni-Cd, 117
Ni-MH, 121

Nickel-cadmium battery, 117
Nickel-metal hydride battery, 111, 121
Nitrogen enriched air, 223
No-load, 30
Node, 10
Nominal voltage, 115
Non return to zero format, 329
Non-inverting input, 55
Non-parallel system, 153
Non-volatile memory, 210, 332
Numbering, 399

OBIGGS, 223
OEA, 223
OFV, 253
OR logic, 60
Obstacle clearance, 231
Octal, 68
Octal to binary encoder, 70
Ohm, 8
Ohm's Law, 8, 137
Oil pressure, 109
Oil temperature, 109
Oil temperature indicating system, 205
Oleo leg, 166, 257
On-board diagnostic equipment, 366
On-board inert gas generation system, 223
One-shot, 63
Open area smoke detector, 291
Open loom, 353, 354
Operational amplifier, 55
Optical fire detector, 284, 290, 291
Oscilloscope, 366
Out-of-balance condition, 106
Outflow valve, 253
Output characteristic, 51
Overhead panel, 105
Overheat protection, 283
Overheat switch, 285, 286
Oxygen enriched air, 223

P-N junction, 38, 39
P.d., 7
P/TEST, 165
PA, 241, 242
PF, 162
PIV, 39
PMG, 106
PNP, 49, 50
PRSOV, 249
PSEU, 167, 168
PSOV, 249
PTC, 37, 38
PTFE, 139
PU, 241
PVC, 349
PVFD, 139
PVS, 244
PWM controller, 100
Panel, 155
Paperless cockpit, 262
Parallel bus, 72, 155
Parallel circuit, 10, 11
Parallel data, 73, 75
Parallel load distribution, 156
Parallel plate capacitor, 4
Parallel system, 152
Passenger address system, 241, 242
Passenger cabin light, 230
Passenger cabin zone, 250
Passenger service unit, 241
Passenger telephone system, 244
Peak inverse voltage, 39
Peak value, 21
Peak-peak value, 22
Peltier effect, 179
Pentavalent impurity, 36
Per-unit regulation, 30, 31
Per-unit slip, 100
Period, 20
Periodic time, 20, 21
Permanent magnet generator, 106
Personal video screen, 244
Phase angle, 25
Phase current, 94, 95
Phase protection, 158
Phase voltage, 95
Phasor diagram, 93, 94
Phonic wheel, 206
Phosphorous lighting, 229
Photoelectric smoke detector, 292
Piezo-resistive strain gauge, 179
Pin/socket identification, 358
Piston engine, 185
Plaque, 117
Plasma, 225, 360
Plate group, 113
Pneumatic fire detector, 289
Pneumatic ice protection, 274
Polar inductor, 186
Polarization, 339
Polarized relay, 169
Pole, 165, 169
Polytetrafluoroethylene, 139
Polyvinylidene fluoride, 139
Position light, 237
Positive charge carrier, 35
Positive ion, 1
Positive ramp, 20
Positive temperature coefficient, 177
Pot, 170
Potential difference, 7, 8
Potentiometer, 170
Potting, 355
Power, 12, 95
Power breaker, 159, 160
Power connector, 151
Power control, 156
Power conversion, 131
Power distribution, 147, 161
Power factor, 27, 95, 162
Power panel, 155
Power supply, 147
Pre-stall, 271
Predictive wind shear, 312
Premature descent alert, 302
Press to test, 165
Pressure capsule, 174
Pressure relief valve, 254
Pressure sensor, 269, 270
Pressure transducer, 175
Pressure-regulating shutoff valve, 249
Pressurization, 252
Primary, 28
Primary bonding, 359
Primary cell, 6
Primary coil, 186
Primary power, 153
Priority encoder, 69
Priority logic, 281
Projector, 243
Propeller de-icing, 275
Propeller synchronization, 208
Protection, 142, 156
Proton, 1
Proximity sensor, 257
Proximity switch, 166, 167, 168
Proximity switch electronic unit, 167
Pulse wave, 20
Pulse width modulation, 100
Pyro-electric cell, 290
Pyrotechnic device, 295

QAR, 328
Quadrature, 25
Quantities, 383
Quick access recorder, 328
Quick release connector, 123, 357, 358

R-S bistable, 65, 66
RAM, 76, 77
RAT, 127, 135, 136
RMS, 21, 22
ROM, 76, 77, 78
RTD, 176, 177, 178, 250
RTDE, 206
Radiated EMI, 338, 342
Radiated field, 339
Radiation, 140
Radio altimeter, 304
Radioactive isotope, 292
Radome protection, 345
Rain protection, 271, 275
Rain repellent, 277, 278, 279
Ram air turbine, 127, 135, 136
Ramp wave, 20

Random access memory, 76
Ratiometer, 174, 177
Reactance, 23, 24
Reactive material, 120
Reactive power, 162
Read-only memory, 76
Read-write memory, 76
Real power, 162
Real-time fault, 367
Receiver, 181
Recirculation fan, 252
Rectifier, 43
Reed relay, 169
Reed sensor, 268
Reed switch, 167, 168
Reflected code, 66
Refuelling, 221
Regulator, 127
Relay, 149, 148, 150, 168, 169, 170
Relay configurations, 170
Reserve tank, 222
Reservoir capacitor, 44, 45, 46, 47
Reset, 65
Resistance, 7, 8
Resistance temperature device, 176, 206
Resistivity, 137
Resistor, 9
Resistor symbols, 170
Resonance, 26
Resonant circuit, 26
Responder, 289
Response time, 290
Return path, 362
Return to zero format, 329
Reverse current relay, 148, 149
Reverse-bias, 39
Reversible motor, 257
Revision papers, 373
Rheostat, 170
Rheostat light control, 227
Rigid fuel tank, 211
Ring tongue terminal, 141
Ripple, 45
Rocker switch, 166
Root mean square, 21, 22
Rotary inverter, 131
Rotary position transducer, 179
Rotating beacon light, 236, 238
Rotating rectifier, 107
Rotor, 86, 98
Rotorcraft, 316

SDAC, 262
SG, 115
SNOC, 247
SPST, 165, 171
STP, 74
SV, 246
SWAMP, 354
Safety, 32
Sample and hold, 365
Satellite communication network, 245
Satellite communications, 244
Satellite network operation centre, 247
Saturation, 54
Scratch foil, 321
Scratch foil recorder, 322
Screened cable, 140
Screening, 137, 140
Sealant, 355
Sealed battery, 113, 116
Sealed beam, 236
Secondary, 28
Secondary bonding, 359
Secondary cell, 6
Secondary indicating system, 205
Secondary power, 153
Seebeck effect, 177
Segment, 84
Self-clocking, 75
Self-inductance, 28
Self-oxidizing, 137
Self-sustaining speed, 189
Selsyn, 180
Semiconductor, 35
Sensor vane, 269
Separation point, 266
Serial bus, 72
Serial data, 73, 75
Serial interface module, 72
Series circuit, 10, 11
Series wound motor, 87
Service light, 233
Set, 65
Severe wind and moisture problems, 354
Shaded pole, 103
Shading coil, 103
Shaft horsepower, 204
Shannon-Hartley theory, 342
Shell, 1, 35
Shielded twisted pair, 74
Shielding, 137, 140, 338
Shock hazard, 32
Shunt resistor, 133
Shunt-wound, 88
Si, 36
Sidetone, 241
Sight glass, 211, 212
Silicon, 36, 40
Silicon controlled rectifier, 42
Single-phase generator, 89, 91
Single-phase motor, 101
Single-phase supply, 22
Situational awareness, 303
Slats, 261
Slip, 99
Slip-rings, 82, 180
Slow-blow fuse, 143
Slugged relay, 169
Smoke detector, 291, 292, 293, 294
Solenoid, 15, 86, 170
Solid state recorder, 323, 333
Space vehicle, 246
Spark plug, 187, 188
Specific gravity, 115, 116
Spillage, 116, 119
Splice, 357
Split bus, 148, 152
Split bus system, 153
Split parallel system, 152
Split/parallel bus, 155
Spongy lead, 114
Square wave, 20
Squat switch, 257
Squib, 295
Squirrel cage, 98
Stagnation point, 266
Stall, 265, 266, 271
Stall identification system, 270
Stall protection, 265
Stall warning, 265, 267, 269, 270
Standby bus, 156
Standby power, 155
Star connection, 93, 94, 134
Starter-generator, 89, 90, 189
Starting, 102
Static bonds, 348
Static charge, 360
Static discharger, 3, 360, 361
Static inverter, 132
Static wick, 3, 360
Stator, 99
Step-down transformer, 28
Step-up transformer, 28
Stick push, 271
Stick shaker, 267, 269
Storage cell, 111, 112
Strain gauge, 179, 180, 204
Strain transducer, 179
Streamline flow, 266
Strobe light, 225, 226, 227, 235, 236, 238
Stub cable, 74
Sulphation, 116
Switch, 165, 166
Switch configurations, 166
Switch/light, 167
Symbols, 383, 395
Synchro receiver, 181
Synchro transmitter, 181
Synchronizing, 155
Synchronous, 74
Synchronous motor, 96, 98
Synoptic, 263
System data acquisition concentrator, 262

T-tailed aircraft, 270
T.p.v., 29

TAWS, 281, 299, 301, 302, 303, 316, 317, 318
TCAS, 69, 281
TGT, 198
TIT, 198
TNC, 141
TOT, 198
TRU, 127, 133, 153, 154
TTAC, 247
TTL, 63
Tachometer, 195, 196
Tail wind, 315
Take-off warning, 279, 280
Tape drive, 327
Target wheel, 197
Taxi light, 232, 234
Telemetry tracking and command/control, 247
Telephone gateway, 247
Telephone system, 244
Temperature, 36, 37
Temperature bulb, 176
Temperature coefficient, 37
Temperature transducer, 175
Terminal block, 357
Termination, 356
Terminator, 74
Terrain avoidance, 318
Terrain awareness display, 302
Terrain awareness warning system, 299
Test switch, 228
Testing, 363
Thermal anti-icing, 276
Thermal circuit breaker, 144
Thermal fire detection, 284, 285
Thermal ice protection, 274
Thermal runaway, 37, 118
Thermistor, 37, 38, 176, 286, 287
Thermocouple, 177, 178, 179, 198, 199, 200, 201
Thermoelectric effect, 177
Thermostat, 176
Three-phase cabling, 108
Three-phase generator, 92, 93
Three-phase power, 95
Three-phase supply, 22, 23
Throw, 165, 169
Thrust indication, 201
Thyristor, 42
Timer, 55
Toggle switch, 165, 166
Torque, 85, 88, 202
Torque indicator, 205
Torque transducer, 205
Torque-slip characteristic, 99
Torque-speed characteristic, 88
Traffic alert and collision avoidance system, 69
Trailing edge, 265
Trailing edge flaps, 260
Transducer, 165
Transfer bus, 154
Transfer characteristic, 52
Transfer relay, 154
Transformer, 19, 28, 29, 30, 31
Transformer rectifier unit, 127, 133
Transistor, 49, 52
Transistor characteristic, 50, 51
Transistor light control, 228, 229
Transistor symbols, 53
Transistor-transistor logic, 63
Transition point, 266
Transmitter, 181
Triangle wave, 20
Triboelectric effect, 349
Triboelectric scale, 349
Trickle charge, 116
Trip, 145
Trivalent impurity, 36
True power, 27
Truth table, 59, 62, 67
Turbine, 189
Turbine engine starter, 189, 190
Turbine gas temperature, 198
Turbine inlet temperature, 198
Turbine outlet temperature, 198
Turbine starting system, 194
Turboprop, 189
Turns-per-volt, 29
Twin-engine starting system, 185, 187
Twisted pair, 340, 341
Two to four line decoder, 68
Two to one multiplexer, 71
Two-phase motor, 101
Two-shot fire extinguisher, 295, 297

ULB, 326, 332
ULB specification, 334
UV, 284, 290
Ullage, 221, 223
Ultrasonic detector, 273
Ultrasonic ice detector, 271
Ultraviolet, 284, 290
Under-wing fuel measurement, 215
Underwater location beacon, 326, 332
Unidirectional bus, 72, 73
Unipolar voltage, 84
Unit fire detector, 285
Unit of electricity, 12
Units, 383
Unusable fuel, 223

VRLA, 116
VSCF, 107, 109
Valence shell, 35
Valve-regulated lead-acid battery, 116
Vane sensor, 268, 269
Variable reluctance, 197
Variable resistor, 170
Variable speed constant frequency, 107
Ventilation, 252
Venting, 122, 124, 221
Vibrating contact regulator, 127, 128
Vibration, 206, 356
Vibration sensor, 207
Video entertainment, 243
Video projector, 243
Viscosity, 206
Viscous damper, 269
Visual inspection, 363
Voltage, 7
Voltage divider, 170
Voltage mode, 74
Voltage regulation, 30
Voltage regulator, 41, 42, 127, 128, 130
Volume flow, 202

WDM, 363
Warning panel, 231
Waste bin fire extinguisher, 295
Watt, 12
Watt-second, 12
Waveform, 19, 20, 21
Weight on wheels, 257
Wet arc tracking, 139
Wet cell, 113
Wet start, 194
Wheatstone bridge, 177, 178, 206
Wind shear, 308, 312, 315
Windscreen de-icing, 277
Windscreen heater, 134
Windscreen ice protection, 275
Windscreen wiper, 276, 278
Wiper, 276, 278
Wire, 140, 353, 359, 389
Wire coding, 399
Wire construction, 137
Wire gauge, 139
Wire identification, 140
Wire loom, 354, 353
Wire materials, 137
Wire numbering, 399
Wire performance, 140
Wire recorder, 322
Wire size, 139
Wire specification, 139
Wire termination, 141, 356
Wiring, 137, 347
Wiring diagram manual, 363
Wiring loom, 353
WoW, 257
Wound rotor, 98
Wrist strap, 350

Xenon, 225, 226, 227

Zener diode, 40, 41, 42

All the Mathematics You Missed

[But Need to Know for Graduate School]

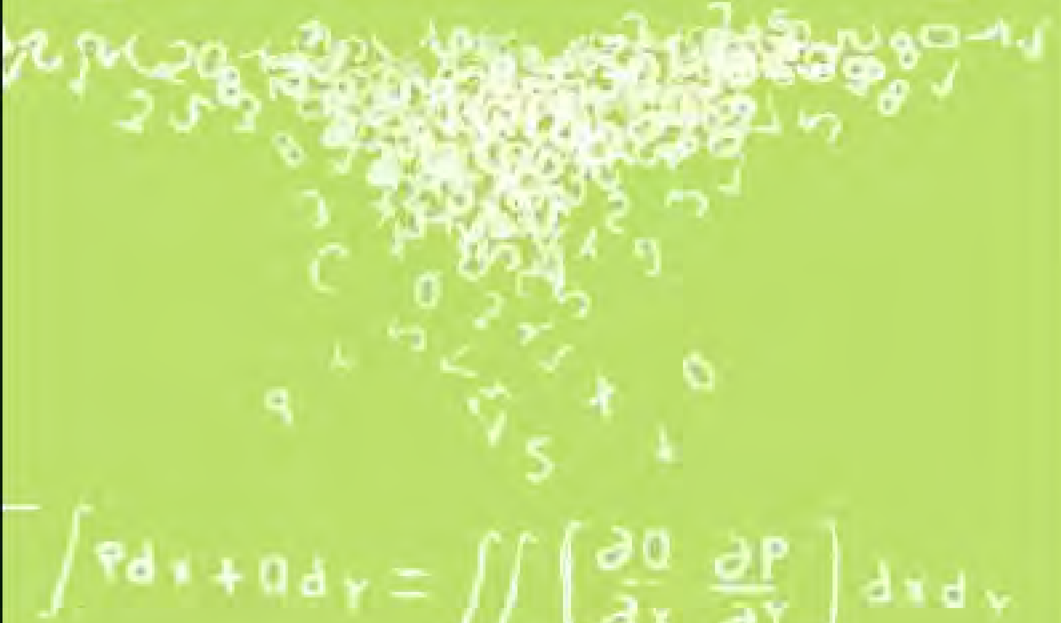
$\oint P\,dx + Q\,dy = \iint \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx\,dy$

THOMAS A. GARRITY

All the Mathematics You Missed

Beginning graduate students in mathematics and other quantitative subjects are expected to have a daunting breadth of mathematical knowledge, but few have such a background. This book will help students see the broad outline of mathematics and to fill in the gaps in their knowledge.

The author explains the basic points and a few key results of the most important undergraduate topics in mathematics, emphasizing the intuitions behind the subject. The topics include linear algebra, vector calculus, differential geometry, real analysis, point-set topology, differential equations, probability theory, complex analysis, abstract algebra, and more. An annotated bibliography offers a guide to further reading and more rigorous foundations.

This book will be an essential resource for advanced undergraduate and beginning graduate students in mathematics, the physical sciences, engineering, computer science, statistics, and economics, and for anyone else who needs to quickly learn some serious mathematics.

Thomas A. Garrity is Professor of Mathematics at Williams College in Williamstown, Massachusetts. He was an undergraduate at the University of Texas, Austin, and a graduate student at Brown University, receiving his Ph.D. in 1986. From 1986 to 1989, he was G.C. Evans Instructor at Rice University. In 1989, he moved to Williams College, where he has been ever since except in 1992–3, when he spent the year at the University of Washington, and 2000–1, when he spent the year at the University of Michigan, Ann Arbor.

All the Mathematics You Missed

But Need to Know for Graduate School

Thomas A. Garrity
Williams College

Figures by Lori Pedersen

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

© Thomas A. Garrity 2002

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2002

Printed in the United States of America

Typeface Palatino 10/12 pt.   *System* LATEX [AU]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data

Garrity, Thomas A., 1959 –
All the mathematics you missed : but need to know for graduate school / Thomas A. Garrity.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-79285-1 – ISBN 0-521-79707-1 (pb.)
1. Mathematics. I. Title.

QA37.3 .G37 2002
510–dc21   2001037644

ISBN 0 521 79285 1 hardback
ISBN 0 521 79707 1 paperback

Dedicated to the Memory

of

Robert Mizner

Contents

Preface xiii

On the Structure of Mathematics xix

Brief Summaries of Topics xxiii
- 0.1 Linear Algebra xxiii
- 0.2 Real Analysis xxiii
- 0.3 Differentiating Vector-Valued Functions xxiii
- 0.4 Point Set Topology xxiv
- 0.5 Classical Stokes' Theorems xxiv
- 0.6 Differential Forms and Stokes' Theorem xxiv
- 0.7 Curvature for Curves and Surfaces xxiv
- 0.8 Geometry xxv
- 0.9 Complex Analysis xxv
- 0.10 Countability and the Axiom of Choice xxvi
- 0.11 Algebra xxvi
- 0.12 Lebesgue Integration xxvi
- 0.13 Fourier Analysis xxvi
- 0.14 Differential Equations xxvii
- 0.15 Combinatorics and Probability Theory xxvii
- 0.16 Algorithms xxvii

1 Linear Algebra 1
- 1.1 Introduction 1
- 1.2 The Basic Vector Space $\mathbf{R}^n$ 2
- 1.3 Vector Spaces and Linear Transformations 4
- 1.4 Bases and Dimension 6
- 1.5 The Determinant 9
- 1.6 The Key Theorem of Linear Algebra 12
- 1.7 Similar Matrices 14
- 1.8 Eigenvalues and Eigenvectors 15

1.9 Dual Vector Spaces 20
1.10 Books 21
1.11 Exercises 21

2 ϵ and δ Real Analysis 23
2.1 Limits 23
2.2 Continuity 25
2.3 Differentiation 26
2.4 Integration 28
2.5 The Fundamental Theorem of Calculus 31
2.6 Pointwise Convergence of Functions 35
2.7 Uniform Convergence 36
2.8 The Weierstrass M-Test 38
2.9 Weierstrass' Example 40
2.10 Books 43
2.11 Exercises 44

3 Calculus for Vector-Valued Functions 47
3.1 Vector-Valued Functions 47
3.2 Limits and Continuity 49
3.3 Differentiation and Jacobians 50
3.4 The Inverse Function Theorem 53
3.5 Implicit Function Theorem 56
3.6 Books 60
3.7 Exercises 60

4 Point Set Topology 63
4.1 Basic Definitions 63
4.2 The Standard Topology on $\mathbf{R}^n$ 66
4.3 Metric Spaces 72
4.4 Bases for Topologies 73
4.5 Zariski Topology of Commutative Rings 75
4.6 Books 77
4.7 Exercises 78

5 Classical Stokes' Theorems 81
5.1 Preliminaries about Vector Calculus 82
  5.1.1 Vector Fields 82
  5.1.2 Manifolds and Boundaries 84
  5.1.3 Path Integrals 87
  5.1.4 Surface Integrals 91
  5.1.5 The Gradient 93
  5.1.6 The Divergence 93

5.1.7 The Curl 94
5.1.8 Orientability 94
5.2 The Divergence Theorem and Stokes' Theorem 95
5.3 Physical Interpretation of Divergence Thm. 97
5.4 A Physical Interpretation of Stokes' Theorem 98
5.5 Proof of the Divergence Theorem 99
5.6 Sketch of a Proof for Stokes' Theorem 104
5.7 Books 108
5.8 Exercises 108

6 Differential Forms and Stokes' Thm. 111
6.1 Volumes of Parallelepipeds 112
6.2 Diff. Forms and the Exterior Derivative 115
6.2.1 Elementary k-forms 115
6.2.2 The Vector Space of k-forms 118
6.2.3 Rules for Manipulating k-forms 119
6.2.4 Differential k-forms and the Exterior Derivative . . . 122
6.3 Differential Forms and Vector Fields 124
6.4 Manifolds 126
6.5 Tangent Spaces and Orientations 132
6.5.1 Tangent Spaces for Implicit and Parametric Manifolds 132
6.5.2 Tangent Spaces for Abstract Manifolds 133
6.5.3 Orientation of a Vector Space 135
6.5.4 Orientation of a Manifold and its Boundary 136
6.6 Integration on Manifolds 137
6.7 Stokes' Theorem 139
6.8 Books 142
6.9 Exercises 143

7 Curvature for Curves and Surfaces 145
7.1 Plane Curves 145
7.2 Space Curves 148
7.3 Surfaces 152
7.4 The Gauss-Bonnet Theorem 157
7.5 Books 158
7.6 Exercises 158

8 Geometry 161
8.1 Euclidean Geometry 162
8.2 Hyperbolic Geometry 163
8.3 Elliptic Geometry 166
8.4 Curvature 167

8.5 Books 168
8.6 Exercises 169

9 Complex Analysis **171**
9.1 Analyticity as a Limit 172
9.2 Cauchy-Riemann Equations 174
9.3 Integral Representations of Functions 179
9.4 Analytic Functions as Power Series 187
9.5 Conformal Maps 191
9.6 The Riemann Mapping Theorem 194
9.7 Several Complex Variables: Hartog's Theorem 196
9.8 Books 197
9.9 Exercises 198

10 Countability and the Axiom of Choice **201**
10.1 Countability 201
10.2 Naive Set Theory and Paradoxes 205
10.3 The Axiom of Choice 207
10.4 Non-measurable Sets 208
10.5 Gödel and Independence Proofs 210
10.6 Books 211
10.7 Exercises 211

11 Algebra **213**
11.1 Groups 213
11.2 Representation Theory 219
11.3 Rings 221
11.4 Fields and Galois Theory 223
11.5 Books 228
11.6 Exercises 229

12 Lebesgue Integration **231**
12.1 Lebesgue Measure 231
12.2 The Cantor Set 234
12.3 Lebesgue Integration 236
12.4 Convergence Theorems 239
12.5 Books 241
12.6 Exercises 241

13 Fourier Analysis **243**
13.1 Waves, Periodic Functions and Trigonometry 243
13.2 Fourier Series 244
13.3 Convergence Issues 250

13.4 Fourier Integrals and Transforms 252
13.5 Solving Differential Equations 256
13.6 Books 258
13.7 Exercises 258

14 Differential Equations 261
14.1 Basics 261
14.2 Ordinary Differential Equations 262
14.3 The Laplacian 266
    14.3.1 Mean Value Principle 266
    14.3.2 Separation of Variables 267
    14.3.3 Applications to Complex Analysis 270
14.4 The Heat Equation 270
14.5 The Wave Equation 273
    14.5.1 Derivation 273
    14.5.2 Change of Variables 277
14.6 Integrability Conditions 279
14.7 Lewy's Example 281
14.8 Books 282
14.9 Exercises 282

15 Combinatorics and Probability 285
15.1 Counting 285
15.2 Basic Probability Theory 287
15.3 Independence 290
15.4 Expected Values and Variance 291
15.5 Central Limit Theorem 294
15.6 Stirling's Approximation for $n!$ 300
15.7 Books 305
15.8 Exercises 305

16 Algorithms 307
16.1 Algorithms and Complexity 308
16.2 Graphs: Euler and Hamiltonian Circuits 308
16.3 Sorting and Trees 313
16.4 P=NP? 316
16.5 Numerical Analysis: Newton's Method 317
16.6 Books 324
16.7 Exercises 324

A Equivalence Relations 327

Preface

Math is Exciting. We are living in the greatest age of mathematics ever seen. In the 1930s, there were some people who feared that the rising abstractions of the early twentieth century would either lead to mathematicians working on sterile, silly intellectual exercises or to mathematics splitting into sharply distinct subdisciplines, similar to the way natural philosophy split into physics, chemistry, biology and geology. But the very opposite has happened. Since World War II, it has become increasingly clear that mathematics is one unified discipline. What were separate areas now feed off of each other. Learning and creating mathematics is indeed a worthwhile way to spend one's life.

Math is Hard. Unfortunately, people are just not that good at mathematics. While intensely enjoyable, it also requires hard work and self-discipline. I know of no serious mathematician who finds math easy. In fact, most, after a few beers, will confess as to how stupid and slow they are. This is one of the personal hurdles that a beginning graduate student must face, namely how to deal with the profundity of mathematics in stark comparison to our own shallow understandings of mathematics. This is in part why the attrition rate in graduate school is so high. At the best schools, with the most successful retention rates, usually only about half of the people who start eventually get their PhDs. Even schools that are in the top twenty have at times had eighty percent of their incoming graduate students not finish. This is in spite of the fact that most beginning graduate students are, in comparison to the general population, amazingly good at mathematics. Most have found that math is one area in which they could shine. Suddenly, in graduate school, they are surrounded by people who are just as good (and who seem even better). To make matters worse, mathematics is a meritocracy. The faculty will not go out of their way to make beginning students feel good (this is not the faculty's job; their job is to discover new mathematics). The fact is that there are easier (though, for a mathematician, less satisfying) ways to make a living. There is truth in the statement that you must be driven to become a mathematician.

Mathematics is exciting, though. The frustrations should more than be compensated for by the thrills of learning and eventually creating (or discovering) new mathematics. That is, after all, the main goal for attending graduate school, to become a research mathematician. As with all creative endeavors, there will be emotional highs and lows. Only jobs that are routine and boring will not have these peaks and valleys. Part of the difficulty of graduate school is learning how to deal with the low times.

Goal of Book. The goal of this book is to give people at least a rough idea of the many topics that beginning graduate students at the best graduate schools are assumed to know. Since there is unfortunately far more that is needed to be known for graduate school and for research than it is possible to learn in a mere four years of college, few beginning students know all of these topics, but hopefully all will know at least some. Different people will know different topics. This strongly suggests the advantage of working with others.

There is another goal. Many nonmathematicians suddenly find that they need to know some serious math. The prospect of struggling with a text will legitimately seem for them to be daunting. Each chapter of this book will provide for these folks a place where they can get a rough idea and outline of the topic they are interested in.

As for general hints for helping sort out some mathematical field, certainly one should always, when faced with a new definition, try to find a simple example and a simple non-example. A non-example, by the way, is an example that almost, but not quite, satisfies the definition. But beyond finding these examples, one should examine the reason why the basic definitions were given. This leads to a split into two streams of thought for how to do mathematics. One can start with reasonable, if not naive, definitions and then prove theorems about these definitions. Frequently the statements of the theorems are complicated, with many different cases and conditions, and the proofs are quite convoluted, full of special tricks.

The other, more mid-twentieth century approach, is to spend quite a bit of time on the basic definitions, with the goal of having the resulting theorems be clearly stated and having straightforward proofs. Under this philosophy, any time there is a trick in a proof, it means more work needs to be done on the definitions. It also means that the definitions themselves take work to understand, even at the level of figuring out why anyone would care. But now the theorems can be cleanly stated and proved.

In this approach the role of examples becomes key. Usually there are basic examples whose properties are already known. These examples will shape the abstract definitions and theorems. The definitions in fact are made in order for the resulting theorems to give, for the examples, the answers we expect. Only then can the theorems be applied to new examples and cases whose properties are unknown.

For example, the correct notion of a derivative and thus of the slope of a tangent line is somewhat complicated. But whatever definition is chosen, the slope of a horizontal line (and hence the derivative of a constant function) must be zero. If the definition of a derivative does not yield that a horizontal line has zero slope, it is the definition that must be viewed as wrong, not the intuition behind the example.

For another example, consider the definition of the curvature of a plane curve, which is in Chapter Seven. The formulas are somewhat ungainly. But whatever the definitions, they must yield that a straight line has zero curvature, that at every point of a circle the curvature is the same and that the curvature of a circle with small radius must be greater than the curvature of a circle with a larger radius (reflecting the fact that it is easier to balance on the earth than on a basketball). If a definition of curvature does not do this, we would reject the definitions, not the examples.

Thus it pays to know the key examples. When trying to undo the technical maze of a new subject, knowing these examples will not only help explain why the theorems and definitions are what they are but will even help in predicting what the theorems must be.

Of course this is vague and ignores the fact that first proofs are almost always ugly and full of tricks, with the true insight usually hidden. But in learning the basic material, look for the key idea, the key theorem and then see how these shape the definitions.

Caveats for Critics. This book is far from a rigorous treatment of any topic. There is a deliberate looseness in style and rigor. I am trying to get the point across and to write in the way that most mathematicians talk to each other. The level of rigor in this book would be totally inappropriate in a research paper.

Consider that there are three tasks for any intellectual discipline:

1. Coming up with new ideas.

2. Verifying new ideas.

3. Communicating new ideas.

How people come up with new ideas in mathematics (or in any other field) is overall a mystery. There are at best a few heuristics in mathematics, such as asking if something is unique or if it is canonical. It is in verifying new ideas that mathematicians are supreme. Our standard is that there must be a rigorous proof. Nothing else will do. This is why the mathematical literature is so trustworthy (not that mistakes don't creep in, but they are usually not major errors). In fact, I would go as far as to say that if any discipline has as its standard of verification rigorous proof, than that discipline must be a part of mathematics. Certainly the main goal for a math major in the first few years of college is to learn what a rigorous proof is.

Unfortunately, we do a poor job of communicating mathematics. Every year there are millions of people who take math courses. A large number of people who you meet on the street or on the airplane have taken college level mathematics. How many enjoyed it? How many saw no real point to it? While this book is not addressed to that random airplane person, it is addressed to beginning graduate students, people who already enjoy mathematics but who all too frequently get blown out of the mathematical water by mathematics presented in an unmotivated, but rigorous, manner. There is no problem with being nonrigorous, as long as you know and clearly label when you are being nonrigorous.

Comments on the Bibliography. There are many topics in this book. While I would love to be able to say that I thoroughly know the literature on each of these topics, that would be a lie. The bibliography has been cobbled together from recommendations from colleagues, from books that I have taught from and books that I have used. I am confident that there are excellent texts that I do not know about. If you have a favorite, please let me know at tgarrity@williams.edu.

While this book was being written, Paulo Ney De Souza and Jorge-Nuno Silva wrote *Berkeley Problems in Mathematics* [26], which is an excellent collection of problems that have appeared over the years on qualifying exams (usually taken in the first or second year of graduate school) in the math department at Berkeley. In many ways, their book is the complement of this one, as their work is the place to go to when you want to test your computational skills while this book concentrates on underlying intuitions. For example, say you want to learn about complex analysis. You should first read chapter nine of this book to get an overview of the basics about complex analysis. Then choose a good complex analysis book and work most of its exercises. Then use the problems in De Souza and Silva as a final test of your knowledge.

Finally, the book *Mathematics, Form and Function* by Mac Lane [82], is excellent. It provides an overview of much of mathematics. I am listing it here because there was no other place where it could be naturally referenced. Second and third year graduate students should seriously consider reading this book.

Acknowledgments

First, I would like to thank Lori Pedersen for a wonderful job of creating the illustrations and diagrams for this book.

Many people have given feedback and ideas over the years. Nero Budar, Chris French and Richard Haynes were student readers of one of the early versions of this manuscript. Ed Dunne gave much needed advice and help. In the spring semester of 2000 at Williams, Tegan Cheslack-Postava, Ben Cooper and Ken Dennison went over the book line-by-line. Others who have given ideas have included Bill Lenhart, Frank Morgan, Cesar Silva, Colin Adams, Ed Burger, David Barrett, Sergey Fomin, Peter Hinman, Smadar Karni, Dick Canary, Jacek Miekisz, David James and Eric Schippers. During the final rush to finish this book, Trevor Arnold, Yann Bernard, Bill Correll, Jr., Bart Kastermans, Christopher Kennedy, Elizabeth Klodginski, Alex Köronya, Scott Kravitz, Steve Root and Craig Westerland have provided amazing help. Marissa Barschdorff texed a very early version of this manuscript. The Williams College Department of Mathematics and Statistics has been a wonderful place to write the bulk of this book; I thank all of my Williams' colleagues. The last revisions were done while I have been on sabbatical at the University of Michigan, another great place to do mathematics. I would like to thank my editor at Cambridge, Lauren Cowles, and also Caitlin Doggart at Cambridge. Gary Knapp has throughout provided moral support and gave a close, detailed reading to an early version of the manuscript. My wife, Lori, has also given much needed encouragement and has spent many hours catching many of my mistakes. To all I owe thanks.

Finally, near the completion of this work, Bob Mizner passed away at an early age. It is in his memory that I dedicate this book (though no doubt he would have disagreed with most of my presentations and choices of topics; he definitely would have made fun of the lack of rigor).

On the Structure of Mathematics

If you look at articles in current journals, the range of topics seems immense. How could anyone even begin to make sense out of all of these topics? And indeed there is a glimmer of truth in this. People cannot effortlessly switch from one research field to another. But not all is chaos. There are at least two ways of placing some type of structure on all of mathematics.

Equivalence Problems

Mathematicians want to know when things are the same, or, when they are equivalent. What is meant by the same is what distinguishes one branch of mathematics from another. For example, a topologist will consider two geometric objects (technically, two topological spaces) to be the same if one can be twisted and bent, but not ripped, into the other. Thus for a topologist, we have

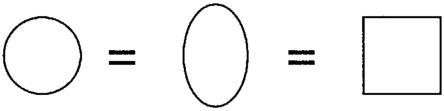

To a differential topologist, two geometric objects are the same if one can be smoothly bent and twisted into the other. By smooth we mean that no sharp edges can be introduced. Then

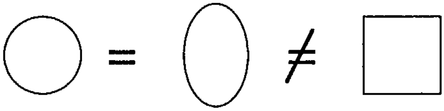

The four sharp corners of the square are what prevent it from being equivalent to the circle.

For a differential geometer, the notion of equivalence is even more restrictive. Here two objects are the same not only if one can be smoothly bent and twisted into the other but also if the curvatures agree. Thus for the differential geometer, the circle is no longer equivalent to the ellipse:

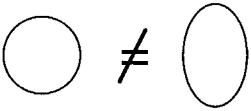

As a first pass to placing structure on mathematics, we can view an area of mathematics as consisting of certain *Objects*, coupled with the notion of *Equivalence* between these objects. We can explain equivalence by looking at the allowed *Maps*, or functions, between the objects. At the beginning of most chapters, we will list the Objects and the Maps between the objects that are key for that subject. The *Equivalence Problem* is of course the problem of determining when two objects are the same, using the allowable maps.

If the equivalence problem is easy to solve for some class of objects, then the corresponding branch of mathematics will no longer be active. If the equivalence problem is too hard to solve, with no known ways of attacking the problem, then the corresponding branch of mathematics will again not be active, though of course for opposite reasons. The hot areas of mathematics are precisely those for which there are rich partial but not complete answers to the equivalence problem. But what could we mean by a partial answer?

Here enters the notion of invariance. Start with an example. Certainly the circle, as a topological space, is different from two circles,

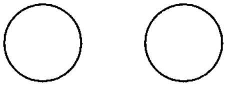

since a circle has only one connected component and two circles have two connected components. We map each topological space to a positive integer, namely the number of connected components of the topological space. Thus we have:

$$\text{Topological Spaces} \rightarrow \text{Positive Integers}.$$

The key is that the number of connected components for a space cannot change under the notion of topological equivalence (under bendings and twistings). We say that the number of connected components is an *invariant* of a topological space. Thus if the spaces map to different numbers, meaning that they have different numbers of connected components, then the two spaces cannot be topologically equivalent.

Of course, two spaces can have the same number of connected components and still be different. For example, both the circle and the sphere

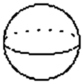

have only one connected component, but they are different. (These can be distinguished by looking at each space's dimension, which is another topological invariant.) The goal of topology is to find enough invariants to be able to always determine when two spaces are different or the same. This has not come close to being done. Much of algebraic topology maps each space not to invariant numbers but to other types of algebraic objects, such as groups and rings. Similar techniques show up throughout mathematics. This provides for tremendous interplay between different branches of mathematics.

The Study of Functions

The mantra that we should all chant each night before bed is:

$$\boxed{\text{Functions describe the World.}}$$

To a large extent what makes mathematics so useful to the world is that seemingly disparate real-world situations can be described by the same type of function. For example, think of how many different problems can be recast as finding the maximum or minimum of a function.

Different areas of mathematics study different types of functions. Calculus studies differentiable functions from the real numbers to the real numbers, algebra studies polynomials of degree one and two (in high school) and permutations (in college), linear algebra studies linear functions, or matrix multiplication.

Thus in learning a new area of mathematics, you should always "find the function" of interest. Hence at the beginning of most chapters we will state the type of function that will be studied.

Equivalence Problems in Physics

Physics is an experimental science. Hence any question in physics must eventually be answered by performing an experiment. But experiments come down to making observations, which usually are described by certain computable numbers, such as velocity, mass or charge. Thus the experiments in physics are described by numbers that are read off in the lab. More succinctly, physics is ultimately:

$$\boxed{\text{Numbers in Boxes}}$$

where the boxes are various pieces of lab machinery used to make measurements. But different boxes (different lab set-ups) can yield different numbers, even if the underlying physics is the same. This happens even at the trivial level of choice of units.

More deeply, suppose you are modeling the physical state of a system as the solution of a differential equation. To write down the differential equation, a coordinate system must be chosen. The allowed changes of coordinates are determined by the physics. For example, Newtonian physics can be distinguished from Special Relativity in that each has different allowable changes of coordinates.

Thus while physics is 'Numbers in Boxes', the true questions come down to when different numbers represent the same physics. But this is an equivalence problem; mathematics comes to the fore. (This explains in part the heavy need for advanced mathematics in physics.) Physicists want to find physics invariants. Usually, though, physicists call their invariants 'Conservation Laws'. For example, in classical physics the conservation of energy can be recast as the statement that the function that represents energy is an invariant function.

Brief Summaries of Topics

0.1 Linear Algebra

Linear algebra studies linear transformations and vector spaces, or in another language, matrix multiplication and the vector space $\mathbf{R}^n$. You should know how to translate between the language of abstract vector spaces and the language of matrices. In particular, given a basis for a vector space, you should know how to represent any linear transformation as a matrix. Further, given two matrices, you should know how to determine if these matrices actually represent the same linear transformation, but under different choices of bases. The key theorem of linear algebra is a statement that gives many equivalent descriptions for when a matrix is invertible. These equivalences should be known cold. You should also know why eigenvectors and eigenvalues occur naturally in linear algebra.

0.2 Real Analysis

The basic definitions of a limit, continuity, differentiation and integration should be known and understood in terms of ϵ's and δ's. Using this ϵ and δ language, you should be comfortable with the idea of uniform convergence of functions.

0.3 Differentiating Vector-Valued Functions

The goal of the Inverse Function Theorem is to show that a differentiable function $f : \mathbf{R}^n \to \mathbf{R}^n$ is locally invertible if and only if the determinant of its derivative (the Jacobian) is non-zero. You should be comfortable with what it means for a vector-valued function to be differentiable, why its derivative must be a linear map (and hence representable as a matrix, the Jacobian) and how to compute the Jacobian. Further, you should know the statement of the Implicit Function Theorem and see why is is closely related to the Inverse Function Theorem.

0.4 Point Set Topology

You should understand how to define a topology in terms of open sets and how to express the idea of continuous functions in terms of open sets. The standard topology on $\mathbf{R}^n$ must be well understood, at least to the level of the Heine-Borel Theorem. Finally, you should know what a metric space is and how a metric can be used to define open sets and hence a topology.

0.5 Classical Stokes' Theorems

You should know about the calculus of vector fields. In particular, you should know how to compute, and know the geometric interpretations behind, the curl and the divergence of a vector field, the gradient of a function and the path integral along a curve. Then you should know the classical extensions of the Fundamental Theorem of Calculus, namely the Divergence Theorem and Stokes' Theorem. You should especially understand why these are indeed generalizations of the Fundamental Theorem of Calculus.

0.6 Differential Forms and Stokes' Theorem

Manifolds are naturally occurring geometric objects. Differential k-forms are the tools for doing calculus on manifolds. You should know the various ways for defining a manifold, how to define and to think about differential k-forms, and how to take the exterior derivative of a k-form. You should also be able to translate from the language of k-forms and exterior derivatives to the language from Chapter Five on vector fields, gradients, curls and divergences. Finally, you should know the statement of Stokes' Theorem, understand why it is a sharp quantitative statement about the equality of the integral of a k-form on the boundary of a $(k+1)$-dimensional manifold with the integral of the exterior derivative of the k-form on the manifold, and how this Stokes' Theorem has as special cases the Divergence Theorem and the Stokes' Theorem from the previous chapter.

0.7 Curvature for Curves and Surfaces

Curvature, in all of its manifestations, attempts to measure the rate of change of the directions of tangent spaces of geometric objects. You should know how to compute the curvature of a plane curve, the curvature and the torsion of a space curve and the two principal curvatures, in terms of the Hessian, of a surface in space.

0.8 Geometry

Different geometries are built out of different axiomatic systems. Given a line l and a point p not on l, Euclidean geometry assumes that there is exactly one line containing p parallel to l, hyperbolic geometry assumes that there is more than one line containing p parallel to l, and elliptic geometries assume that there is no line parallel to l. You should know models for hyperbolic geometry, single elliptic geometry and double elliptic geometry. Finally, you should understand why the existence of such models implies that all of these geometries are mutually consistent.

0.9 Complex Analysis

The main point is to recognize and understand the many equivalent ways for describing when a function can be analytic. Here we are concerned with functions $f : U \to \mathbf{C}$, where U is an open set in the complex numbers $\mathbf{C}$. You should know that such a function $f(z)$ is said to be *analytic* if it satisfies any of the following equivalent conditions:
a) For all $z_0 \in U$,

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

exists.
b)The real and imaginary parts of the function f satisfy the *Cauchy-Riemann equations*:

$$\frac{\partial Ref}{\partial x} = \frac{\partial Imf}{\partial y}$$

and

$$\frac{\partial Ref}{\partial y} = -\frac{\partial Imf}{\partial x}.$$

c) If γ is any counterclockwise simple loop in $\mathbf{C}=\mathbf{R}^2$ and if z_0 is any complex number in the interior of γ, then

$$f(z_0) = \frac{1}{2\pi i} \int_\gamma \frac{f(z)}{z - z_0} \mathrm{d}z.$$

This is the *Cauchy Integral formula.*

d) For any complex number z_0, there is an open neighborhood in $\mathbf{C} = \mathbf{R}^2$ of z_0 on which

$$f(z) = \sum_{k=o}^{\infty} a_k (z - z_0)^k,$$

is a uniformly converging series.

Further, if $f : U \rightarrow \mathbf{C}$ is analytic and if $f'(z_0) \neq 0$, then at z_0, the function f is conformal (i.e., angle-preserving), viewed as a map from $\mathbf{R}^2$ to $\mathbf{R}^2$.

0.10 Countability and the Axiom of Choice

You should know what it means for a set to be countably infinite. In particular, you should know that the integers and rationals are countably infinite while the real numbers are uncountably infinite. The statement of the Axiom of Choice and the fact that it has many seemingly bizarre equivalences should also be known.

0.11 Algebra

Groups, the basic object of study in abstract algebra, are the algebraic interpretations of geometric symmetries. One should know the basics about groups (at least to the level of the Sylow Theorem, which is a key tool for understanding finite groups), rings and fields. You should also know Galois Theory, which provides the link between finite groups and the finding of the roots of a polynomial and hence shows the connections between high school and abstract algebra. Finally, you should know the basics behind representation theory, which is how one relates abstract groups to groups of matrices.

0.12 Lebesgue Integration

You should know the basic ideas behind Lebesgue measure and integration, at least to the level of the Lebesgue Dominating Convergence Theorem, and the concept of sets of measure zero.

0.13 Fourier Analysis

You should know how to find the Fourier series of a periodic function, the Fourier integral of a function, the Fourier transform, and how Fourier series relate to Hilbert spaces. Further, you should see how Fourier transforms can be used to simplify differential equations.

0.14 Differential Equations

Much of physics, economics, mathematics and other sciences comes down to trying to find solutions to differential equations. One should know that the goal in differential equations is to find an unknown function satisfying an equation involving derivatives. Subject to mild restrictions, there are always solutions to ordinary differential equations. This is most definitely not the case for partial differential equations, where even the existence of solutions is frequently unknown. You should also be familiar with the three traditional classes of partial differential equations: the heat equation, the wave equation and the Laplacian.

0.15 Combinatorics and Probability Theory

Both elementary combinatorics and basic probability theory reduce to problems in counting. You should know that

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

is the number of ways of choosing k elements from n elements. The relation of $\binom{n}{k}$ to the binomial theorem for polynomials is useful to have handy for many computations. Basic probability theory should be understood. In particular one should understand the terms: sample space, random variable (both its intuitions and its definition as a function), expected value and variance. One should definitely understand why counting arguments are critical for calculating probabilities of finite sample spaces. The link between probability and integral calculus can be seen in the various versions of the Central Limit Theorem, the ideas of which should be known.

0.16 Algorithms

You should understand what is meant by the complexity of an algorithm, at least to the level of understanding the question P=NP. Basic graph theory should be known; for example, you should see why a tree is a natural structure for understanding many algorithms. Numerical Analysis is the study of algorithms for approximating the answer to computations in mathematics. As an example, you should understand Newton's method for approximating the roots of a polynomial.

Chapter 1

Linear Algebra

Basic Object:	Vector Spaces
Basic Map:	Linear Transformations
Basic Goal:	Equivalences for the Invertibility of Matrices

1.1 Introduction

Though a bit of an exaggeration, it can be said that a mathematical problem can be solved only if it can be reduced to a calculation in linear algebra. And a calculation in linear algebra will reduce ultimately to the solving of a system of linear equations, which in turn comes down to the manipulation of matrices. Throughout this text and, more importantly, throughout mathematics, linear algebra is a key tool (or more accurately, a collection of intertwining tools) that is critical for doing calculations.

The power of linear algebra lies not only in our ability to manipulate matrices in order to solve systems of linear equations. The abstraction of these concrete objects to the ideas of vector spaces and linear transformations allows us to see the common conceptual links between many seemingly disparate subjects. (Of course, this is the advantage of any good abstraction.) For example, the study of solutions to linear differential equations has, in part, the same feel as trying to model the hood of a car with cubic polynomials, since both the space of solutions to a linear differential equation and the space of cubic polynomials that model a car hood form vector spaces.

The key theorem of linear algebra, discussed in section six, gives many equivalent ways of telling when a system of n linear equations in n unknowns has a solution. Each of the equivalent conditions is important. What is remarkable and what gives linear algebra its oomph is that they are all the same.

1.2 The Basic Vector Space $\mathbf{R}^n$

The quintessential vector space is $\mathbf{R}^n$, the set of all n-tuples of real numbers

$$\{(x_1, \ldots, x_n) : x_i \in \mathbf{R}\}.$$

As we will see in the next section, what makes this a vector space is that we can add together two n-tuples to get another n-tuple:

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$

and that we can multiply each n-tuple by a real number λ:

$$\lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n)$$

to get another n-tuple. Of course each n-tuple is usually called a vector and the real numbers λ are called scalars. When $n = 2$ and when $n = 3$ all of this reduces to the vectors in the plane and in space that most of us learned in high school.

The natural map from some $\mathbf{R}^n$ to an $\mathbf{R}^m$ is given by matrix multiplication. Write a vector $\mathbf{x} \in \mathbf{R}^n$ as a column vector:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Similarly, we can write a vector in $\mathbf{R}^m$ as a column vector with m entries. Let A be an $m \times n$ matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \ldots & \ldots & a_{mn} \end{pmatrix}.$$

Then $A\mathbf{x}$ is the m-tuple:

$$A\mathbf{x} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \ldots & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \ldots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n \end{pmatrix}.$$

For any two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbf{R}^n$ and any two scalars λ and μ, we have

$$A(\lambda\mathbf{x} + \mu\mathbf{y}) = \lambda A\mathbf{x} + \mu A\mathbf{y}.$$

In the next section we will use the linearity of matrix multiplication to motivate the definition for a linear transformation between vector spaces.

Now to relate all of this to the solving of a system of linear equations. Suppose we are given numbers $b_1, \ldots, b_m$ and numbers $a_{11}, \ldots, a_{mn}$. Our goal is to find n numbers $x_1, \ldots, x_n$ that solve the following system of linear equations:

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

Calculations in linear algebra will frequently reduce to solving a system of linear equations. When there are only a few equations, we can find the solutions by hand, but as the number of equations increases, the calculations quickly turn from enjoyable algebraic manipulations into nightmares of notation. These nightmarish complications arise not from any single theoretical difficulty but instead stem solely from trying to keep track of the many individual minor details. In other words, it is a problem in bookkeeping.

Write

$$\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \ldots & \ldots & a_{mn} \end{pmatrix}$$

and our unknowns as

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Then we can rewrite our system of linear equations in the more visually appealing form of

$$A\mathbf{x} = \mathbf{b}.$$

When $m > n$ (when there are more equations than unknowns), we expect there to be, in general, no solutions. For example, when $m = 3$ and $n = 2$, this corresponds geometrically to the fact that three lines in a plane will usually have no common point of intersection. When $m < n$ (when there are more unknowns than equations), we expect there to be, in general, many solutions. In the case when $m = 2$ and $n = 3$, this corresponds geometrically to the fact that two planes in space will usually intersect in an entire line. Much of the machinery of linear algebra deals with the remaining case when $m = n$.

Thus we want to find the $n \times 1$ column vector $\mathbf{x}$ that solves $A\mathbf{x} = \mathbf{b}$, where A is a given $n \times n$ matrix and $\mathbf{b}$ is a given $n \times 1$ column vector.

Suppose that the square matrix A has an inverse matrix A^{-1} (which means that A^{-1} is also $n \times n$ and more importantly that $A^{-1}A = I$, with I the identity matrix). Then our solution will be

$$\mathbf{x} = A^{-1}\mathbf{b}$$

since

$$A\mathbf{x} = A(A^{-1}\mathbf{b}) = I\mathbf{b} = \mathbf{b}.$$

Thus solving our system of linear equations comes down to understanding when the $n \times n$ matrix A has an inverse. (If an inverse matrix exists, then there are algorithms for its calculations.)

The key theorem of linear algebra, stated in section six, is in essence a list of many equivalences for when an $n \times n$ matrix has an inverse and is thus essential to understanding when a system of linear equations can be solved.

1.3 Vector Spaces and Linear Transformations

The abstract approach to studying systems of linear equations starts with the notion of a vector space.

Definition 1.3.1 *A set V is a* vector space *over the real numbers*[1] $\mathbf{R}$ *if there are maps:*

1. $\mathbf{R} \times V \to V$, *denoted by $a \cdot v$ or av for all real numbers a and elements v in V,*

2. $V \times V \to V$, *denoted by $v + w$ for all elements v and w in the vector space V,*

with the following properties:

a) There is an element 0, in V such that $0 + v = v$ for all $v \in V$.
b) For each $v \in V$, there is an element $(-v) \in V$ with $v + (-v) = 0$.
c) For all $v, w \in V$, $v + w = w + v$.
d) For all $a \in \mathbf{R}$ and for all $v, w \in V$, we have that $a(v + w) = av + aw$.
e) For all $a, b \in \mathbf{R}$ and all $v \in V$, $a(bv) = (a \cdot b)v$.
f) For all $a, b \in \mathbf{R}$ and all $v \in V$, $(a + b)v = av + bv$.
g) For all $v \in V$, $1 \cdot v = v$.

---

[1] The real numbers can be replaced by the complex numbers and in fact by any field (which will be defined in Chapter Eleven on algebra).

As a matter of notation, and to agree with common usage, the elements of a vector space are called *vectors* and the elements of **R** (or whatever field is being used) *scalars*. Note that the space $\mathbf{R}^n$ given in the last section certainly satisfies these conditions.

The natural map between vector spaces is that of a linear transformation.

Definition 1.3.2 *A* linear transformation $T : V \to W$ *is a function from a vector space V to a vector space W such that for any real numbers a_1 and a_2 and any vectors v_1 and v_2 in V, we have*

$$T(a_1v_1 + a_2v_2) = a_1T(v_1) + a_2T(v_2).$$

Matrix multiplication from an $\mathbf{R}^n$ to an $\mathbf{R}^m$ gives an example of a linear transformation.

Definition 1.3.3 *A subset U of a vector space V is a* subspace *of V if U is itself a vector space.*

In practice, it is usually easy to see if a subset of a vector space is in fact a subspace, by the following proposition, whose proof is left to the reader:

Proposition 1.3.1 *A subset U of a vector space V is a subspace of V if U is closed under addition and scalar multiplication.*

Given a linear transformation $T : V \to W$, there are naturally occurring subspaces of both V and W.

Definition 1.3.4 *If $T : V \to W$ is a linear transformation, then the* kernel *of T is:*

$$ker(T) = \{v \in V : T(v) = 0\}$$

and the image *of T is*

$$Im(T) = \{w \in W : \textit{there exists a } v \in V \textit{ with } T(v) = w\}.$$

The kernel is a subspace of V, since if v_1 and v_2 are two vectors in the kernel and if a and b are any two real numbers, then

$$\begin{aligned} T(av_1 + bv_2) &= aT(v_1) + bT(v_2) \\ &= a \cdot 0 + b \cdot 0 \\ &= 0. \end{aligned}$$

In a similar way we can show that the image of T is a subspace of W.

If the only vector spaces that ever occurred were column vectors in $\mathbf{R}^n$, then even this mild level of abstraction would be silly. This is not the case.

Here we look at only one example. Let $C^k[0,1]$ be the set of all real-valued functions with domain the unit interval $[0,1]$:

$$f:[0,1]\to \mathbf{R}$$

such that the kth derivative of f exists and is continuous. Since the sum of any two such functions and a multiple of any such function by a scalar will still be in $C^k[0,1]$, we have a vector space. Though we will officially define dimension next section, $C^k[0,1]$ will be infinite dimensional (and thus definitely not some $\mathbf{R}^n$). We can view the derivative as a linear transformation from $C^k[0,1]$ to those functions with one less derivative, $C^{k-1}[0,1]$:

$$\frac{\mathrm{d}}{\mathrm{d}x}:C^k[0,1]\to C^{k-1}[0,1].$$

The kernel of $\frac{\mathrm{d}}{\mathrm{d}x}$ consists of those functions with $\frac{\mathrm{d}f}{\mathrm{d}x}=0$, namely constant functions.

Now consider the differential equation

$$\frac{\mathrm{d}^2f}{\mathrm{d}x^2}+3\frac{\mathrm{d}f}{\mathrm{d}x}+2f=0.$$

Let T be the linear transformation:

$$T=\frac{\mathrm{d}^2}{\mathrm{d}x^2}+3\frac{\mathrm{d}}{\mathrm{d}x}+2I:C^2[0,1]\to C^0[0,1].$$

The problem of finding a solution $f(x)$ to the original differential equation can now be translated to finding an element of the kernel of T. This suggests the possibility (which indeed is true) that the language of linear algebra can be used to understand solutions to (linear) differential equations.

1.4 Bases, Dimension, and Linear Transformations as Matrices

Our next goal is to define the dimension of a vector space.

Definition 1.4.1 *A set of vectors* $(v_1,\ldots,v_n)$ *form a* basis *for the vector space* V *if given any vector* v *in* V*, there are unique scalars* $a_1,\ldots,a_n\in\mathbf{R}$ *with* $v=a_1v_1+\ldots+a_nv_n$.

Definition 1.4.2 *The* dimension *of a vector space* V*, denoted by* $dim(V)$*, is the number of elements in a basis.*

As it is far from obvious that the number of elements in a basis will always be the same, no matter which basis is chosen, in order to make the definition of the dimension of a vector space well-defined we need the following theorem (which we will not prove):

Theorem 1.4.1 *All bases of a vector space V have the same number of elements.*

For $\mathbf{R}^n$, the usual basis is

$$\{(1,0,...,0),(0,1,0,...,0),...,(0,...,0,1)\}.$$

Thus $\mathbf{R}^n$ is n dimensional. Of course if this were not true, the above definition of dimension would be wrong and we would need another. This is an example of the principle mentioned in the introduction. We have a good intuitive understanding of what dimension should mean for certain specific examples: a line needs to be one dimensional, a plane two dimensional and space three dimensional. We then come up with a sharp definition. If this definition gives the "correct" answer for our three already understood examples, we are somewhat confident that the definition has indeed captured what is meant by, in this case, dimension. Then we can apply the definition to examples where our intuitions fail.

Linked to the idea of a basis is:

Definition 1.4.3 *Vectors $(v_1,\ldots,v_n)$ in a vector space V are* linearly independent *if whenever*

$$a_1v_1+\cdots+a_nv_n=0,$$

it must be the case that the scalars $a_1,\ldots,a_n$ must all be zero.

Intuitively, a collection of vectors are linearly independent if they all point in different directions. A basis consists then in a collection of linearly independent vectors that span the vector space, where by span we mean:

Definition 1.4.4 *A set of vectors $(v_1,\ldots,v_n)$* span *the vector space V if given any vector v in V, there are scalars $a_1,\ldots,a_n \in \mathbf{R}$ with $v=a_1v_1+\cdots+a_nv_n$.*

Our goal now is to show how all linear transformations $T:V\to W$ between finite-dimensional spaces can be represented as matrix multiplication, provided we fix bases for the vector spaces V and W.

First fix a basis $\{v_1,...,v_n\}$ for V and a basis $\{w_1,...,w_m\}$ for W. Before looking at the linear transformation T, we need to show how each element of the n-dimensional space V can be represented as a column vector in $\mathbf{R}^n$ and how each element of the m-dimensional space W can be represented as a column vector of $\mathbf{R}^m$. Given any vector v in V, by the definition of basis, there are unique real numbers $a_1, ..., a_n$ with

$$v = a_1 v_1 + \cdots + a_n v_n.$$

We thus represent the vector v with the column vector:

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

Similarly, for any vector w in W, there are unique real numbers $b_1, ..., b_m$ with

$$w = b_1 w_1 + \cdots + b_m w_m.$$

Here we represent w as the column vector

$$\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

Note that we have established a correspondence between vectors in V and W and column vectors $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively. More technically, we can show that V is isomorphic to $\mathbf{R}^n$ (meaning that there is a one-one, onto linear transformation from V to $\mathbf{R}^n$) and that W is isomorphic to $\mathbf{R}^m$, though it must be emphasized that the actual correspondence only exists after a basis has been chosen (which means that while the isomorphism exists, it is not canonical; this is actually a big deal, as in practice it is unfortunately often the case that no basis is given to us).

We now want to represent a linear transformation $T : V \to W$ as an $m \times n$ matrix A. For each basis vector v_i in the vector space V, $T(v_i)$ will be a vector in W. Thus there will exist real numbers $a_{1i}, \ldots, a_{mi}$ such that

$$T(v_i) = a_{1i} w_1 + \cdots + a_{mi} w_m.$$

We want to see that the linear transformation T will correspond to the $m \times n$ matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \ldots & \ldots & a_{mn} \end{pmatrix}.$$

Given any vector v in V, with $v = a_1 v_1 + \cdots + a_n v_n$, we have

$$\begin{aligned} T(v) &= T(a_1 v_1 + \cdots + a_n v_n) \\ &= a_1 T(v_1) + \cdots + a_n T(v_n) \\ &= a_1 (a_{11} w_1 + \cdots + a_{m1} w_m) + \cdots \\ &\quad + a_n (a_{1n} w_1 + \cdots + a_{mn} w_m). \end{aligned}$$

But under the correspondences of the vector spaces with the various column spaces, this can be seen to correspond to the matrix multiplication of A times the column vector corresponding to the vector v:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \dots & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

Note that if $T : V \to V$ is a linear transformation from a vector space to itself, then the corresponding matrix will be $n \times n$, a square matrix.

Given different bases for the vector spaces V and W, the matrix associated to the linear transformation T will change. A natural problem is to determine when two matrices actually represent the same linear transformation, but under different bases. This will be the goal of section seven.

1.5 The Determinant

Our next task is to give a definition for the determinant of a matrix. In fact, we will give three alternative descriptions of the determinant. All three are equivalent; each has its own advantages.

Our first method is to define the determinant of a 1×1 matrix and then to define recursively the determinant of an $n \times n$ matrix.

Since 1×1 matrices are just numbers, the following should not at all be surprising:

Definition 1.5.1 *The* determinant *of a* 1×1 *matrix* (a) *is the real-valued function*

$$\det(a) = a.$$

This should not yet seem significant.

Before giving the definition of the determinant for a general $n \times n$ matrix, we need a little notation. For an $n \times n$ matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & \dots & a_{nn} \end{pmatrix},$$

denote by A_{ij} the $(n-1) \times (n-1)$ matrix obtained from A by deleting the ith row and the jth column. For example, if $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, then $A_{12} = (a_{21})$. Similarly if $A = \begin{pmatrix} 2 & 3 & 5 \\ 6 & 4 & 9 \\ 7 & 1 & 8 \end{pmatrix}$, then $A_{12} = \begin{pmatrix} 6 & 9 \\ 7 & 8 \end{pmatrix}$.

Since we have a definition for the determinant for 1 × 1 matrices, we will now assume by induction that we know the determinant of any $(n-1)\times(n-1)$ matrix and use this to find the determinant of an $n\times n$ matrix.

Definition 1.5.2 *Let A be an $n\times n$ matrix. Then the* determinant *of A is*

$$\det(A)=\sum_{k=1}^{n}(-1)^{k+1}a_{1k}\det(A_{1k}).$$

Thus for $A=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{pmatrix}$, we have

$$\det(A)=a_{11}\det(A_{11})-a_{12}\det(A_{12})=a_{11}a_{22}-a_{12}a_{21},$$

which is what most of us think of as the determinant. The determinant of our above 3 × 3 matrix is:

$$\det\begin{pmatrix} 2 & 3 & 5 \\ 6 & 4 & 9 \\ 7 & 1 & 8\end{pmatrix}=2\det\begin{pmatrix} 4 & 9 \\ 1 & 8\end{pmatrix}-3\det\begin{pmatrix} 6 & 9 \\ 7 & 8\end{pmatrix}+5\det\begin{pmatrix} 6 & 4 \\ 7 & 1\end{pmatrix}.$$

While this definition is indeed an efficient means to describe the determinant, it obscures most of the determinant's uses and intuitions.

The second way we can describe the determinant has built into it the key algebraic properties of the determinant. It highlights function-theoretic properties of the determinant.

Denote the $n\times n$ matrix A as $A=(A_1,\ldots,A_n)$, where A_i denotes the i^{th} column:

$$A_i=\begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni}\end{pmatrix}.$$

Definition 1.5.3 *The* determinant *of A is defined as the unique real-valued function*

$$\det : \mathit{Matrices}\rightarrow \mathbf{R}$$

satisfying:

a) $\det(A_1,\ldots,\lambda A_k,\ldots,A_n)=\lambda\det(A_1,\ldots,A_k)$.

b) $\det(A_1,\ldots,A_k+\lambda A_i,\ldots,A_n)=\det(A_1,\ldots,A_n)$ *for* $k\neq i$.

c) det*(Identity matrix)* = 1.

Thus, treating each column vector of a matrix as a vector in $\mathbf{R}^n$, the determinant can be viewed as a special type of function from $\mathbf{R}^n\times\ldots\times\mathbf{R}^n$ to the real numbers.

In order to be able to use this definition, we would have to prove that such a function on the space of matrices, satisfying conditions *a* through *c*, even exists and then that it is unique. Existence can be shown by checking that our first (inductive) definition for the determinant satisfies these conditions, though it is a painful calculation. The proof of uniqueness can be found in almost any linear algebra text.

The third definition for the determinant is the most geometric but is also the most vague. We must think of an $n \times n$ matrix A as a linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^n$. Then A will map the unit cube in $\mathbf{R}^n$ to some different object (a parallelepiped). The unit cube has, by definition, a volume of one.

Definition 1.5.4 *The* determinant *of the matrix A is the signed volume of the image of the unit cube.*

This is not well-defined, as the very method of defining the volume of the image has not been described. In fact, most would define the signed volume of the image to be the number given by the determinant using one of the two earlier definitions. But this can be all made rigorous, though at the price of losing much of the geometric insight.

Let's look at some examples: the matrix $A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ takes the unit square to

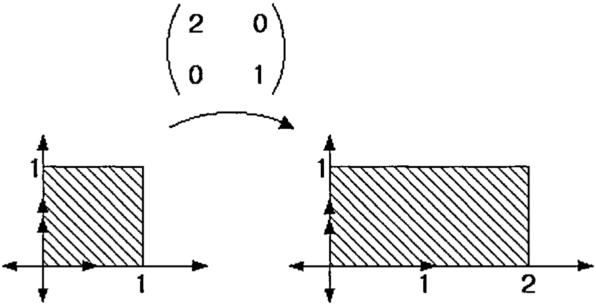

Since the area is doubled, we must have

$$\det(A) = 2.$$

Signed volume means that if the orientations of the edges of the unit cube are changed, then we must have a negative sign in front of the volume. For example, consider the matrix $A = \begin{pmatrix} -2 & 0 \\ 0 & 1 \end{pmatrix}$. Here the image is

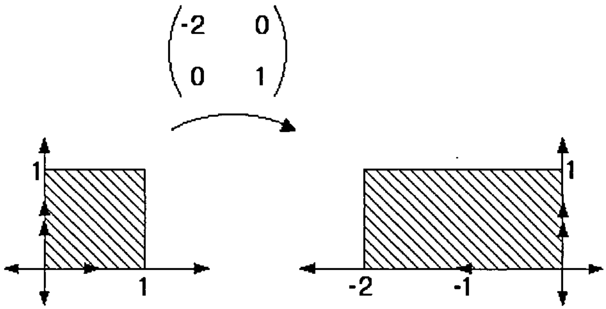

Note that the orientations of the sides are flipped. Since the area is still doubled, the definition will force

$$\det(A) = -2.$$

To rigorously define orientation is somewhat tricky (we do it in Chapter Six), but its meaning is straightforward.

The determinant has many algebraic properties. For example,

Lemma 1.5.1 : *If A and B are $n \times n$ matrices, then*

$$\det(AB) = \det(A)\det(B).$$

This can be proven by either a long calculation or by concentrating on the definition of the determinant as the change of volume of a unit cube.

1.6 The Key Theorem of Linear Algebra

Here is the the key theorem of linear algebra. (Note: we have yet to define eigenvalues and eigenvectors, but we will in section eight.)

Theorem 1.6.1 (Key Theorem) *Let A be an $n \times n$ matrix. Then the following are equivalent:*

1. *A is invertible.*
2. $\det(A) \neq 0$.
3. *ker*$(A) = 0$.
4. *If $\mathbf{b}$ is a column vector in $\mathbf{R}^n$, there is a unique column vector $\mathbf{x}$ in $\mathbf{R}^n$ satisfying $A\mathbf{x} = \mathbf{b}$.*

5. *The columns of A are linearly independent $n \times 1$ column vectors.*
6. *The rows of A are linearly independent $1 \times n$ row vectors.*
7. *The transpose A^t of A is invertible. (Here, if $A = (a_{ij})$, then $A^t = (a_{ji})$).*
8. *All of the eigenvalues of A are nonzero.*

We can restate this theorem in terms of linear transformations.

Theorem 1.6.2 (Key Theorem) *Let $T : V \to V$ be a linear transformation. Then the following are equivalent:*

1. *T is invertible.*
2. *$\det(T) \neq 0$, where the determinant is defined by a choice of basis on V.*
3. *$\ker(T) = 0$.*
4. *If b is a vector in V, there is a unique vector v in V satisfying $T(v) = b$.*
5. *For any basis $v_1, \ldots, v_n$ of V, the image vectors $T(v_1), \ldots, T(v_n)$ are linearly independent.*
6. *For any basis $v_1, \ldots, v_n$ of V, if S denotes the transpose linear transformation of T, then the image vectors $S(v_1), \ldots, S(v_n)$ are linearly independent.*
7. *The transpose of T is invertible. (Here the transpose is defined by a choice of basis on V).*
8. *All of the eigenvalues of T are nonzero.*

In order to make the correspondence between the two theorems clear, we must worry about the fact that we only have definitions of the determinant and the transpose for matrices, not for linear transformations. While we do not show it, both notions can be extended to linear transformations, provided a basis is chosen (in fact, provided we choose an inner product, which will be defined in Chapter Thirteen on Fourier series). But note that while the actual value $\det(T)$ will depend on a fixed basis, the condition that $\det(T) \neq 0$ does not. Similar statements hold for conditions (6) and (7). A proof is the goal of exercise 7, where you are asked to find any linear algebra book and then fill in the proof. It is unlikely that the linear algebra book will have this result as it is stated here. The act of translating is in fact part of the purpose of making this an exercise.

Each of the equivalences is important. Each can be studied on its own merits. It is remarkable that they are the same.

1.7 Similar Matrices

Recall that given a basis for an n dimensional vector space V, we can represent a linear transformation

$$T : V \to V$$

as an $n \times n$ matrix A. Unfortunately, if you choose a different basis for V, the matrix representing the linear transformation T will be quite different from the original matrix A. This section's goal is to find out a clean criterion for when two matrices actually represent the same linear transformation but under different choice of bases.

Definition 1.7.1 *Two $n \times n$ matrices A and B are* similar *if there is an invertible matrix C such that*

$$A = C^{-1}BC.$$

We want to see that two matrices are similar precisely when they represent the same linear transformation. Choose two bases for the vector space V, say $\{v_1, \ldots, v_n\}$ (the v basis) and $\{w_1, \ldots, w_n\}$ (the w basis). Let A be the matrix representing the linear transformation T for the v basis and let B be the matrix representing the linear transformation for the w basis. We want to construct the matrix C so that $A = C^{-1}BC$.

Recall that given the v basis, we can write each vector $z \in V$ as an $n \times 1$ column vector as follows: we know that there are unique scalars $a_1, \ldots, a_n$ with

$$z = a_1 v_1 + \cdots + a_n v_n.$$

We then write z, with respect to the v basis, as the column vector:

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

Similarly, there are unique scalars $b_1, \ldots, b_n$ so that

$$z = b_1 w_1 + \cdots + b_n w_n,$$

meaning that with respect to the w basis, the vector z is the column vector:

$$\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

The desired matrix C will be the matrix such that

$$C\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

If $C = (c_{ij})$, then the entries c_{ij} are precisely the numbers which yield:

$$w_i = c_{i1}v_1 + \ldots + c_{in}v_n.$$

Then, for A and B to represent the same linear transformation, we need the diagram:

$$\begin{tikzcd}
\mathbf{R}^n \arrow[r, "A"] \arrow[d, "C"'] & \mathbf{R}^n \arrow[d, "C"] \\
\mathbf{R}^n \arrow[r, "B"'] & \mathbf{R}^n
\end{tikzcd}$$

to commute, meaning that $CA = BC$ or

$$A = C^{-1}BC,$$

as desired.

Determining when two matrices are similar is a type of result that shows up throughout math and physics. Regularly you must choose some coordinate system (some basis) in order to write down anything at all, but the underlying math or physics that you are interested in is independent of the initial choice. The key question becomes: what is preserved when the coordinate system is changed? Similar matrices allow us to start to understand these questions.

1.8 Eigenvalues and Eigenvectors

In the last section we saw that two matrices represent the same linear transformation, under different choices of bases, precisely when they are similar. This does not tell us, though, how to choose a basis for a vector space so that a linear transformation has a particularly decent matrix representation. For example, the diagonal matrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

is similar to the matrix

$$B = \frac{1}{4}\begin{pmatrix} 1 & -4 & -5 \\ 1 & 8 & -1 \\ 5 & 4 & 15 \end{pmatrix},$$

but all recognize the simplicity of A as compared to B. (By the way, it is not obvious that A and B are similar; I started with A, chose a nonsingular matrix C and then used the software package Mathematica to compute $C^{-1}AC$ to get B. I did not just suddenly "see" that A and B are similar. No, I rigged it to be so.)

One of the purposes behind the following definitions for eigenvalues and eigenvectors is to give us tools for picking out good bases. There are, though, many other reasons to understand eigenvalues and eigenvectors.

Definition 1.8.1 *Let $T : V \rightarrow V$ be a linear transformation. Then a nonzero vector $v \in V$ will be an* eigenvector *of T with* eigenvalue λ, *a scalar, if*

$$T(v) = \lambda v.$$

For an $n \times n$ matrix A, a nonzero column vector $\mathbf{x} \in \mathbf{R}^n$ will be an eigenvector *with* eigenvalue λ, *a scalar, if*

$$A\mathbf{x} = \lambda \mathbf{x}.$$

Geometrically, a vector v is an eigenvector of the linear transformation T with eigenvalue λ if T stretches v by a factor of λ.

For example,

$$\begin{pmatrix} -2 & -2 \\ 6 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ -2 \end{pmatrix},$$

and thus 2 is an eigenvalue and $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$ an eigenvector for the linear transformation represented by the 2×2 matrix $\begin{pmatrix} -2 & -2 \\ 6 & 5 \end{pmatrix}$.

Luckily there is an easy way to describe the eigenvalues of a square matrix, which will allow us to see that the eigenvalues of a matrix are preserved under a similarity transformation.

Proposition 1.8.1 *A number λ will be an eigenvalue of a square matrix A if and only if λ is a root of the polynomial*

$$P(t) = \det(tI - A).$$

The polynomial $P(t) = \det(tI - A)$ is called the *characteristic polynomial* of the matrix A.

Proof: Suppose that λ is an eigenvalue of A, with eigenvector v. Then $Av = \lambda v$, or

$$\lambda v - Av = 0,$$

where the zero on the right hand side is the zero column vector. Then, putting in the identity matrix I, we have

$$0 = \lambda v - Av = (\lambda I - A)v.$$

Thus the matrix $\lambda I - A$ has a nontrivial kernel, v. By the key theorem of linear algebra, this happens precisely when

$$\det(\lambda I - A) = 0,$$

which means that λ is a root of the characteristic polynomial $P(t) = \det(tI - A)$. Since all of these directions can be reversed, we have our theorem. □

Theorem 1.8.1 *Let A and B be similar matrices. Then the characteristic polynomial of A is equal to the characteristic polynomial of B.*

Proof: For A and B to be similar, there must be an invertible matrix C with $A = C^{-1}BC$. Then

$$\begin{aligned} \det(tI - A) &= \det(tI - C^{-1}BC) \\ &= \det(tC^{-1}C - C^{-1}BC) \\ &= \det(C^{-1})\det(tI - B)\det(C) \\ &= \det(tI - B) \end{aligned}$$

using that $1 = \det(C^{-1}C) = \det(C^{-1})\det(C)$. □

Since the characteristic polynomials for similar matrices are the same, this means that the eigenvalues must be the same.

Corollary 1.8.1.1 *The eigenvalues for similar matrices are equal.*

Thus to see if two matrices are similar, one can compute to see if the eigenvalues are equal. If they are not, the matrices are not similar. Unfortunately in general, having equal eigenvalues does not force matrices to be similar. For example, the matrices

$$A = \begin{pmatrix} 1 & -7 \\ 0 & 2 \end{pmatrix}$$

and

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

both have eigenvalues 1 and 2, but they are not similar. (This can be shown by assuming that there is an invertible two-by-two matrix C with $C^{-1}AC = B$ and then showing that $\det(C) = 0$, contradicting C's invertibility.)

Since the characteristic polynomial $P(t)$ does not change under a similarity transformation, the coefficients of $P(t)$ will also not change under a similarity transformation. But since the coefficients of $P(t)$ will themselves be (complicated) polynomials of the entries of the matrix A, we now have certain special polynomials of the entries of A that are invariant under a similarity transformation. One of these coefficients we have already seen in another guise, namely the determinant of A, as the following theorem shows. This theorem will more importantly link the eigenvalues of A to the determinant of A.

Theorem 1.8.2 *Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues, counted with multiplicity, of a matrix A. Then*

$$\det(A) = \lambda_1 \cdots \lambda_n.$$

Before proving this theorem, we need to discuss the idea of counting eigenvalues "with multiplicity". The difficulty is that a polynomial can have a root that must be counted more than once (e.g., the polynomial $(x-2)^2$ has the single root 2 which we want to count twice). This can happen in particular to the characteristic polynomial. For example, consider the matrix

$$\begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

which has as its characteristic polynomial the cubic

$$(t-5)(t-5)(t-4).$$

For the above theorem, we would list the eigenvalues as 4, 5, and 5, hence counting the eigenvalue 5 twice.

Proof: Since the eigenvalues $\lambda_1, \dots, \lambda_n$ are the (complex) roots of the characteristic polynomial $\det(tI - A)$, we have

$$(t - \lambda_1) \cdots (t - \lambda_n) = \det(tI - A).$$

Setting $t = 0$, we have

$$(-1)^n \lambda_1 \cdots \lambda_n = \det(-A).$$

In the matrix $(-A)$, each column of A is multiplied by (-1). Using the second definition of a determinant, we can factor out each of these (-1)s, to get

$$(-1)^n \lambda_1 \cdots \lambda_n = (-1)^n \det(A)$$

and our result. □

Now finally to turn back to determining a "good" basis for representing a linear transformation. The measure of "goodness" is how close the matrix is to being a diagonal matrix. We will restrict ourselves to a special, but quite prevalent, class: symmetric matrices. By symmetric, we mean that if $A = (a_{ij})$, then we require that the entry at the ith row and jth column (a_{ij}) must equal to the entry at the jth row and the ith column (a_{ji}). Thus

$$\begin{pmatrix} 5 & 3 & 4 \\ 3 & 5 & 2 \\ 4 & 2 & 4 \end{pmatrix}$$

is symmetric but

$$\begin{pmatrix} 5 & 2 & 3 \\ 6 & 5 & 3 \\ 2 & 18 & 4 \end{pmatrix}$$

is not.

Theorem 1.8.3 *If A is a symmetric matrix, then there is a matrix B similar to A which is not only diagonal but with the entries along the diagonal being precisely the eigenvalues of A.*

Proof: The proof basically rests on showing that the eigenvectors for A form a basis in which A becomes our desired diagonal matrix. We will assume that the eigenvalues for A are distinct, as technical difficulties occur when there are eigenvalues with multiplicity.

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be the eigenvectors for the matrix A, with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Form the matrix

$$C = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n),$$

where the ith column of C is the column vector $\mathbf{v}_i$. We will show that the matrix $C^{-1}AC$ will satisfy our theorem. Thus we want to show that $C^{-1}AC$ equals the diagonal matrix

$$B = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Denote

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Then the above diagonal matrix B is the unique matrix with $B\mathbf{e}_i = \lambda_i \mathbf{e}_i$, for all i. Our choice for the matrix C now becomes clear as we observe that for all i, $C\mathbf{e}_i = \mathbf{v}_i$. Then we have

$$C^{-1}AC\mathbf{e}_i = C^{-1}A\mathbf{v}_i = C^{-1}(\lambda_i \mathbf{v}_i) = \lambda_i C^{-1}\mathbf{v}_i = \lambda_i \mathbf{e}_i,$$

giving us the theorem. □

This is of course not the end of the story. For nonsymmetric matrices, there are other canonical ways finding "good" similar matrices, such as the Jordan canonical form, the upper triangular form and rational canonical form.

1.9 Dual Vector Spaces

It pays to study functions. In fact, functions appear at times to be more basic than their domains. In the context of linear algebra, the natural class of functions is linear transformations, or linear maps from one vector space to another. Among all real vector spaces, there is one that seems simplest, namely the one-dimensional vector space of the real numbers **R**. This leads us to examine a special type of linear transformation on a vector space, those that map the vector space to the real numbers, the set of which we will call the *dual space*. Dual spaces regularly show up in mathematics.

Let V be a vector space. The *dual vector space*, or *dual space*, is:

$$\begin{aligned} V^* &= \{\text{linear maps from } V \text{ to the real numbers } \mathbf{R}\} \\ &= \{v^* : V \to \mathbf{R} \mid v^* \text{ is linear}\}. \end{aligned}$$

You can check that the dual space V^* is itself a vector space.

Let $T : V \to W$ be a linear transformation. Then we can define a natural linear transformation

$$T^* : W^* \to V^*$$

from the dual of W to the dual of V as follows. Let $w^* \in W^*$. Then given any vector w in the vector space W, we know that $w^*(w)$ will be a real number. We need to define T^* so that $T^*(w^*) \in V^*$. Thus given any vector $v \in V$, we need $T^*(w^*)(v)$ to be a real number. Simply define

$$T^*(w^*)(v) = w^*(T(v)).$$

By the way, note that the direction of the linear transformation $T : V \to W$ is indeed reversed to $T^* : W^* \to V^*$. Also by "natural", we do not mean that the map T^* is "obvious" but instead that it can be uniquely associated to the original linear transformation T.

Such a dual map shows up in many different contexts. For example, if X and Y are topological spaces with a continuous map $F : X \rightarrow Y$ and if $C(X)$ and $C(Y)$ denote the sets of continuous real-valued functions on X and Y, then here the dual map

$$F^* : C(Y) \rightarrow C(X)$$

is defined by $F^*(g)(x) = g(F(x))$, where g is a continuous map on Y.

Attempts to abstractly characterize all such dual maps were a major theme of mid-twentieth century mathematics and can be viewed as one of the beginnings of category theory.

1.10 Books

Mathematicians have been using linear algebra since they have been doing mathematics, but the styles, methods and the terminologies have shifted. For example, if you look in a college course catalogue in 1900 or probably even 1950, there will be no undergraduate course called linear algebra. Instead there were courses such as "Theory of Equations" or simply "Algebra". As seen in one of the more popular textbooks in the first part of the twentieth century, Maxime Bocher's *Introduction to Higher Algebra* [10], the concern was on concretely solving systems of linear equations. The results were written in an algorithmic style. Modern day computer programmers usually find this style of text far easier to understand than current math books. In the 1930s, a fundamental change in the way algebraic topics were taught occurred with the publication of Van der Waerden's *Modern Algebra* [113][114], which was based on lectures of Emmy Noether and Emil Artin. Here a more abstract approach was taken. The first true modern day linear algebra text was Halmos' *Finite-dimensional Vector Spaces* [52]. Here the emphasis is on the idea of a vector space from the very beginning. Today there are many beginning texts. Some start with systems of linear equations and then deal with vector spaces, others reverse the process. A long time favorite of many is Strang's *Linear Algebra and Its Applications* [109]. As a graduate student, you should volunteer to teach or TA linear algebra as soon as possible.

1.11 Exercises

1. Let $L : V \rightarrow W$ be a linear transformation between two vector spaces. Show that

$$\dim(ker(L)) + \dim(Im(L)) = \dim(V).$$

2. Consider the set of all polynomials in one variable with real coefficients of degree less than or equal to three.

a. Show that this set forms a vector space of dimension four.

b. Find a basis for this vector space.

c. Show that differentiating a polynomial is a linear transformation.

d. Given the basis chosen in part (b), write down the matrix representative of the derivative.

3. Let A and B be two $n \times n$ invertible matrices. Prove that

$$(AB)^{-1} = B^{-1}A^{-1}.$$

4. Let

$$A = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}$$

Find a matrix C so that $C^{-1}AC$ is a diagonal matrix.

5. Denote the vector space of all functions

$$f : \mathbf{R} \to \mathbf{R}$$

which are infinitely differentiable by $C^\infty(\mathbf{R})$. This space is called the space of smooth functions.

a. Show that $C^\infty(\mathbf{R})$ is infinite dimensional.

b. Show that differentiation is a linear transformation:

$$\frac{\mathrm{d}}{\mathrm{d}x} : C^\infty(\mathbf{R}) \to C^\infty(\mathbf{R}).$$

c. For a real number λ, find an eigenvector for $\frac{\mathrm{d}}{\mathrm{d}x}$ with eigenvalue λ.

6. Let V be a finite dimensional vector space. Show that the dual vector space V^* has the same dimension as V.

7. Find a linear algebra text. Use it to prove the key theorem of linear algebra. Note that this is a long exercise but is to be taken seriously.

Chapter 2

ϵ and δ Real Analysis

Basic Object:	The Real Numbers
Basic Maps:	Continuous and Differentiable Functions
Basic Goal:	The Fundamental Theorem of Calculus

While the basic intuitions behind differentiation and integration were known by the late 1600s, allowing for a wealth of physical and mathematical applications to develop during the 1700s, it was only in the 1800s that sharp, rigorous definitions were finally given. The key concept is that of a limit, from which follow the definitions for differentiation and integration and rigorous proofs of their basic properties. Far from a mere exercise in pedantry, this rigorization actually allowed mathematicians to discover new phenomena. For example, Karl Weierstrass discovered a function that was continuous everywhere but differentiable nowhere. In other words, there is a function with no breaks but with sharp edges at every point. Key to his proof is the need for limits to be applied to sequences of functions, leading to the idea of uniform convergence.

We will define limits and then use this definition to develop the ideas of continuity, differentiation and integration of functions. Then we will show how differentiation and integration are intimately connected in the Fundamental Theorem of Calculus. Finally we will finish with uniform convergence of functions and Weierstrass' example.

2.1 Limits

Definition 2.1.1 *A function* $f : \mathbf{R} \rightarrow \mathbf{R}$ *has a* limit L *at the point* a *if given any real number* $\epsilon > 0$ *there is a real number* $\delta > 0$ *such that for all* *real numbers x with*

$$0 < |x - a| < \delta,$$

we have

$$|f(x) - L| < \epsilon.$$

This is denoted by

$$\lim_{x \to a} f(x) = L.$$

Intuitively, the function $f(x)$ should have a limit L at a point a if, for numbers x near a, the value of the function $f(x)$ is close to the number L. In other words, to guarantee that $f(x)$ be close to L, we can require that x is close to a. Thus if we want $f(x)$ to be within an arbitrary $\epsilon > 0$ of the number L (i.e., if we want $|f(x) - L| < \epsilon$), we must be able to specify how close to a we must force x to be. Therefore, given a number $\epsilon > 0$ (no matter how small), we must be able to find a number $\delta > 0$ so that if x is within δ of a, we have that $f(x)$ is within an ϵ of L. This is precisely what the definition says, in symbols.

For example, if the above definition of a limit is to make sense, it must yield that

$$\lim_{x \to 2} x^2 = 4.$$

We will check this now. It must be emphasized that we would be foolish to show that x^2 approaches 4 as x approaches 2 by actually using the definition. We are again doing the common trick of using an example whose answer we already know to check the reasonableness of a new definition. Thus for any $\epsilon > 0$, we must find a $\delta > 0$ so that if $0 < |x - 2| < \delta$, we will have

$$|x^2 - 4| < \epsilon.$$

Set

$$\delta = \min\left(\frac{\epsilon}{5}, 1\right).$$

As often happens, the initial work in finding the correct expression for δ is hidden. Also, the '5' in the denominator will be seen not to be critical. Let $0 < |x - 2| < \delta$. We want $|x^2 - 4| < \epsilon$. Now

$$|x^2 - 4| = |x - 2| \cdot |x + 2|.$$

Since x is within δ of 2,

$$|x + 2| < (2 + \delta) + 2 = 4 + \delta \leq 5.$$

Thus

$$|x^2 - 4| = |x - 2| \cdot |x + 2| < 5 \cdot |x - 2| < 5 \cdot \frac{\epsilon}{5} = \epsilon.$$

We are done.

2.2 Continuity

Definition 2.2.1 *A function* $f : \mathbf{R} \to \mathbf{R}$ *is* continuous *at* a *if*

$$\lim_{x \to a} f(x) = f(a).$$

Of course, any intuition about continuous functions should capture the notion that a continuous function cannot have any breaks in its graph. In other words, you can graph a continuous function without having to lift your pencil from the page. (As with any sweeping intuition, this one will break down if pushed too hard.)

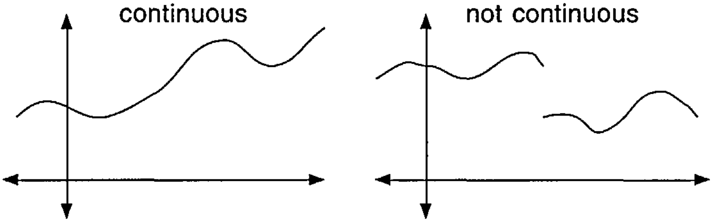

In ϵ and δ notation, the definition of continuity is:

Definition 2.2.2 *A function* $f : \mathbf{R} \to \mathbf{R}$ *is* continuous *at* a *if given any* $\epsilon > 0$, *there is some* $\delta > 0$ *such that for all* x *with* $0 < |x - a| < \delta$, *we have* $|f(x) - f(a)| < \epsilon$.

For an example, we will write down a function that is clearly not continuous at the origin 0, and use this function to check the reasonableness of the definition.

Let

$$f(x) = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x \leq 0 \end{cases}$$

Note that the graph of $f(x)$ has a break in it at the origin.

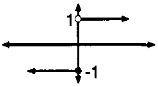

We want to capture this break by showing that

$$\lim_{x \to 0} f(x) \neq f(0).$$

Now $f(0) = -1$. Let $\epsilon = 1$ and let $\delta > 0$ be any positive number. Then for any x with $0 < x < \delta$, we have $f(x) = 1$. Then

$$|f(x) - f(0)| = |1 - (-1)| = 2 > 1 = \epsilon.$$

Thus for all positive $x < \delta$.

$$|f(x) - f(0)| > \epsilon.$$

Hence, for any $\delta > 0$, there are x with

$$|x - 0| < \delta$$

but

$$|f(x) - f(0)| > \epsilon.$$

This function is indeed not continuous.

2.3 Differentiation

Definition 2.3.1 *A function* $f : \mathbf{R} \to \mathbf{R}$ *is* differentiable *at* a *if*

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

exists. This limit is called the derivative *and is denoted by (among many other symbols)* $f'(a)$ *or* $\frac{\mathrm{d}f}{\mathrm{d}x}(a)$.

One of the key intuitive meanings of a derivative is that it should give the slope of the tangent line to the curve $y = f(x)$ at the point a. While logically the current definition of a tangent line must include the above definition of derivative, in pictures the tangent line is of course:

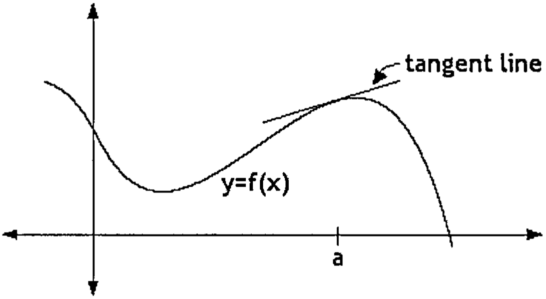

The idea behind the definition is that we can compute the slope of a line defined by any two points in the plane. In particular, for any $x \neq a$, the slope of the secant line through the points $(a, f(a))$ and $(x, f(x))$ will be

$$\frac{f(x) - f(a)}{x - a}.$$

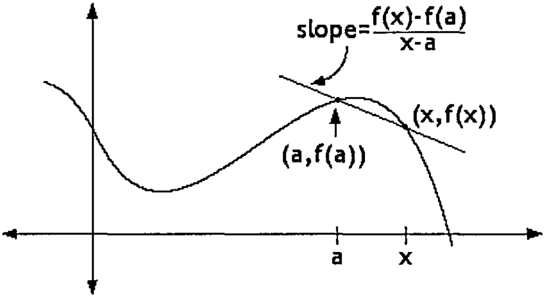

We now let x approach a. The corresponding secant lines will approach the tangent line. Thus the slopes of the secant lines must approach the slope of the tangent line.

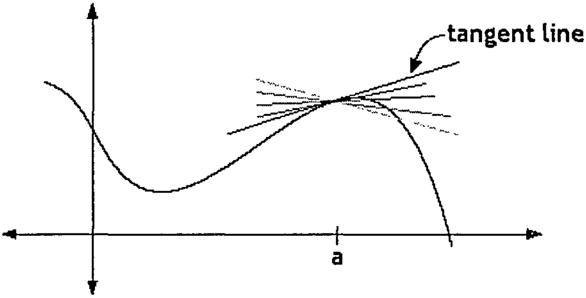

Hence the definition for the slope of the tangent line should be:

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}.$$

Part of the power of derivatives (and why they can be taught to high school seniors and first year college students) is that there is a whole calculational machinery to differentiation, allowing us to usually avoid the actual taking of a limit.

We now look at an example of a function that does not have a derivative at the origin, namely

$$f(x) = |x|.$$

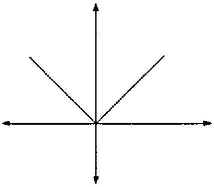

This function has a sharp point at the origin and thus no apparent tangent line there. We will show that the definition yields that $f(x) = |x|$ is indeed not differentiable at $x = 0$. Thus we want to show that

$$\lim_{x \to 0} \frac{f(x) - f(0)}{x - 0}$$

does not exist. Luckily

$$\frac{f(x) - f(0)}{x - 0} = \frac{|x|}{x} = \begin{cases} 1, & x > 0 \\ -1, & x < 0 \end{cases},$$

which we have already shown in the last section to not have a limit as x approaches 0.

2.4 Integration

Intuitively the integral of a positive function $f(x)$ with domain $a \le x \le b$ should be the area under the curve $y = f(x)$ above the x-axis.

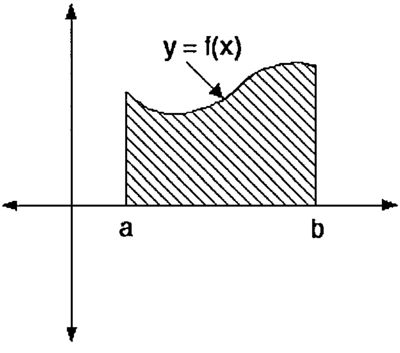

When the function $f(x)$ is not everywhere positive, then its integral should be the area under the positive part of the curve $y = f(x)$ minus the area above the negative part of $y = f(x)$.

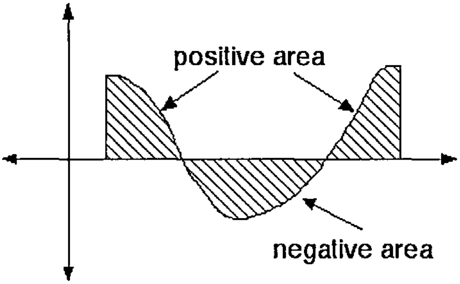

Of course this is hardly rigorous, as we do not yet even have a good definition for area.

The main idea is that the area of a rectangle with height a and width b is ab.

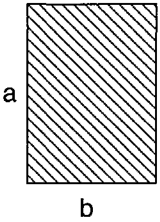

To find the area under a curve $y = f(x)$ we first find the area of various rectangles contained under the curve and then the area of various rectangles just outside the curve.

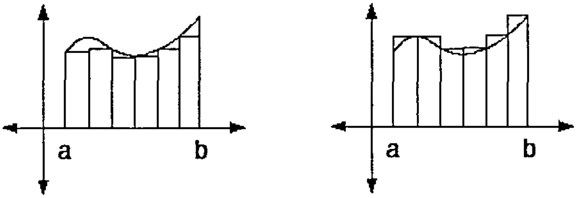

We then make the rectangles thinner and thinner, as in:

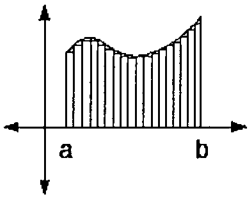

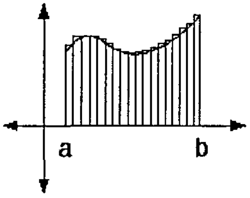

We take the limits, which should result in the area under the curve.

Now for the more technically correct definitions. We consider a real-valued function $f(x)$ with domain the closed interval $[a, b]$. We first want to divide, or partition, the interval $[a, b]$ into little segments that will be the widths of the approximating rectangles. For each positive integer n, let

$$\triangle t = \frac{b-a}{n}$$

and

$$\begin{aligned} a &= t_0, \\ t_1 &= t_0 + \triangle t, \\ t_2 &= t_1 + \triangle t, \\ &\vdots \\ t_n(=b) &= t_{n-1} + \triangle t. \end{aligned}$$

For example, on the interval $[0, 2]$ with $n = 4$, we have $\triangle t = \frac{2-0}{4} = \frac{1}{2}$ and

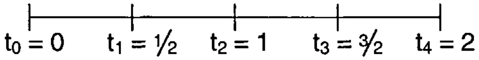

On each interval $[t_{k-1}, t_k]$, choose points l_k and u_k such that for all points t on $[t_{k-1}, t_k]$, we have

$$f(l_k) \leq f(t)$$

and

$$f(u_k) \geq f(t).$$

We make these choices in order to guarantee that the rectangle with base $[t_{k-1}, t_k]$ and height $f(l_k)$ is just under the curve $y = f(x)$ and that the rectangle with base $[t_{k-1}, t_k]$ and height $f(u_k)$ is just outside the curve $y = f(x)$.

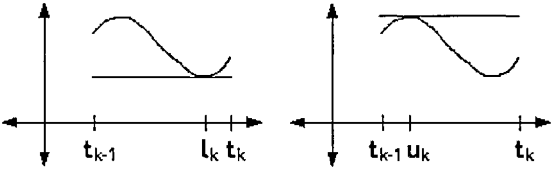

Definition 2.4.1 *Let $f(x)$ be a real-valued function defined on the closed interval $[a, b]$. For each positive integer n, let the* lower sum *of $f(x)$ be*

$$L(f, n) = \sum_{k=1}^{n} f(l_k)\Delta t$$

and the upper sum *be*

$$U(f, n) = \sum_{k=1}^{n} f(u_k)\Delta t.$$

Note that the lower sum $L(f, n)$ is the sum of the areas of the rectangles below our curve while the upper sum $U(f, n)$ is the sum of the areas of the rectangles sticking out above our curve.

Now we can define the integral.

Definition 2.4.2 *A real-valued function $f(x)$ with domain the closed interval $[a, b]$ is said to be* integrable *if the following two limits exist and are equal:*

$$\lim_{n\to\infty} L(f, n) = \lim_{n\to\infty} U(f, n).$$

If these limits are equal, we denote the limit by $\int_a^b f(x)\,\mathrm{d}x$ and call it the integral *of $f(x)$.*

While from pictures it does seem that the above definition will capture the notion of an area under a curve, almost any explicit attempt to actually calculate an integral will be quite difficult. The goal of the next section, the Fundamental Theorem of Calculus, is to see how the integral (an area-finding device) is linked to the derivative (a slope-finding device). This will actually allow us to compute many integrals.

2.5 The Fundamental Theorem of Calculus

Given a real-valued function $f(x)$ defined on the closed interval $[a, b]$ we can use the above definition of integral to define a new function, via setting:

$$F(x) = \int_a^x f(t)\,\mathrm{d}t.$$

We use the variable t inside the integral sign since the variable x is already being used as the independent variable for the function $F(x)$. Thus the value of $F(x)$ is the number that is the (signed) area under the curve $y = f(x)$ from the endpoint a to the value x.

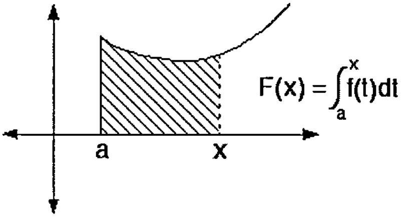

The amazing fact is that the derivative of this new function $F(x)$ will simply be the original function $f(x)$. This means that in order to find the integral of $f(x)$, you should, instead of fussing with upper and lower sums, simply try to find a function whose derivative is $f(x)$.

All of this is contained in:

Theorem 2.5.1 (Fundamental Theorem of Calculus) *Let $f(x)$ be a real-valued continuous function defined on the closed interval $[a, b]$ and define*

$$F(x) = \int_a^x f(t)\,\mathrm{d}t.$$

Then:

a) The function $F(x)$ is differentiable and

$$\frac{\mathrm{d}F(x)}{\mathrm{d}x} = \frac{\mathrm{d}\int_a^x f(t)\,\mathrm{d}t}{\mathrm{d}x} = f(x)$$

and

b) If $G(x)$ is a real-valued differentiable function defined on the closed interval $[a, b]$ whose derivative is:

$$\frac{\mathrm{d}G(x)}{\mathrm{d}x} = f(x),$$

then

$$\int_a^b f(x)\,\mathrm{d}x = G(b) - G(a).$$

First to sketch part a: We want to show that for all x in the interval $[a, b]$, the following limit exists and equals $f(x)$:

$$\lim_{h\to 0} \frac{F(x+h) - F(x)}{h} = f(x).$$

Note that we have mildly reformulated the definition of the derivative, from $\lim_{x \to x_0}(f(x) - f(x_0))/(x - x_0)$ to $\lim_{h \to 0}(f(x+h) - f(x))/h$. These are equivalent. Also, for simplicity, we will only show this for x in the open interval (a, b) and take the limit only for positive h. Consider

$$\begin{aligned} \frac{F(x+h) - F(x)}{h} &= \frac{\int_a^{x+h} f(t)\,\mathrm{d}t - \int_a^x f(t)\,\mathrm{d}t}{h} \\ &= \frac{\int_x^{x+h} f(t)\,\mathrm{d}t}{h}. \end{aligned}$$

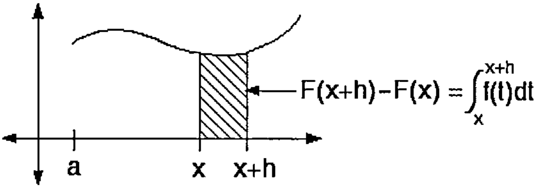

On the interval $[x, x+h]$, for each h define l_h and u_h so that for all points t on $[x, x+h]$, we have

$$f(l_h) \le f(t)$$

and

$$f(u_h) \ge f(t).$$

(Note that we are, in a somewhat hidden fashion, using that a continuous function on an interval like $[x, x+h]$ will have points such as l_h and u_h. In the chapter on point set topology, we will make this explicit, by seeing that on a compact set, such as $[x, x+h]$, a continuous function must achieve both its maximum and minimum.)

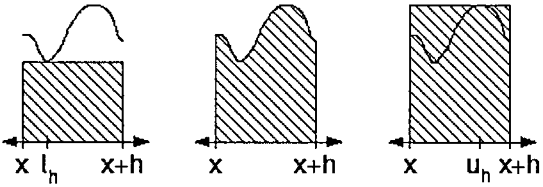

Then we have

$$f(l_h)h \le \int_x^{x+h} f(t)\,\mathrm{d}t \le f(u_h)h.$$

Dividing by $h > 0$ gives us:

$$f(l_h) \leq \frac{\int_x^{x+h} f(t)\,\mathrm{d}t}{h} \leq f(u_h).$$

Now both the l_h and the u_h approach the point x as h approaches zero. Since $f(x)$ is continuous, we have that

$$\lim_{h\to 0} f(l_h) = \lim_{h\to 0} f(u_h) = f(x)$$

and our result.

Turn to part b: Here we are given a function $G(x)$ whose derivative is:

$$\frac{\mathrm{d}G(x)}{\mathrm{d}x} = f(x).$$

Keep the notation of part a, namely that $F(x) = \int_a^x f(t)\,\mathrm{d}t$. Note that $F(a) = 0$ and

$$\int_a^b f(t)\,\mathrm{d}t = F(b) = F(b) - F(a).$$

By part a, we know that the derivative of $F(x)$ is the function $f(x)$. Thus the derivatives of $F(x)$ and $G(x)$ agree, meaning that

$$\frac{\mathrm{d}(F(x) - G(x))}{\mathrm{d}x} = f(x) - f(x) = 0.$$

But a function whose derivative is always zero must be a constant. (We have not shown this. It is quite reasonable, as the only way the slope of the tangent can always be zero is if the graph of the function is a horizontal line; the proof does take some work.) Thus there is a constant c such that

$$F(x) = G(x) + c.$$

Then

$$\int_a^b f(t)\,\mathrm{d}t = F(b) = F(b) - F(a)$$

$$= (G(b) + c) - (G(a) + c)$$

$$= G(b) - G(a)$$

as desired.

2.6 Pointwise Convergence of Functions

Definition 2.6.1 *Let* $f_n : [a,b] \to \mathbf{R}$ *be a sequence of functions*

$$f_1(x), f_2(x), f_3(x), \ldots$$

defined on an interval $[a,b] = \{x : a \leq x \leq b\}$. *This sequence* $\{f_n(x)\}$ *will* converge pointwise *to a function*

$$f(x) : [a,b] \to \mathbf{R}$$

if for all α *in* $[a,b]$,

$$\lim_{n \to \infty} f_n(\alpha) = f(\alpha).$$

In ϵ and δ notation, we would say that $\{f_n(x)\}$ *converges pointwise* to $f(x)$ if for all α in $[a,b]$ and given any $\epsilon > 0$, there is a positive integer N such that for all $n \geq N$, we have $|f(\alpha) - f_n(\alpha)| < \epsilon$.

Intuitively, a sequence of functions $f_n(x)$ will converge pointwise to a function $f(x)$ if, given any α, eventually (for huge n) the numbers $f_n(\alpha)$ become arbitrarily close to the number $f(\alpha)$. The importance of a good notion for convergence of functions stems from the frequent practice of only approximately solving a problem and then using the approximation to understand the true solution. Unfortunately, pointwise convergence is not as useful or as powerful as the next section's topic, uniform convergence, in that the pointwise limit of reasonable functions (e.g., continuous or integrable functions) does not guarantee the reasonableness of the limit, as we will see in the next example.

Here we show that the pointwise limit of continuous functions need not be continuous. For each positive integer n, set

$$f_n(x) = x^n$$

for all x on $[0,1]$.

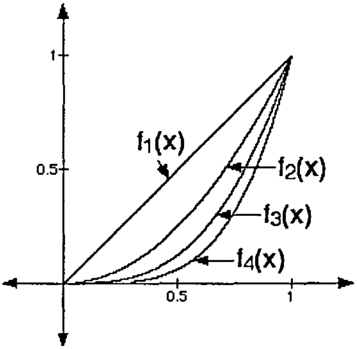

Set

$$f(x) = \begin{cases} 1, & x = 1 \\ 0, & 0 \le x < 1 \end{cases}$$

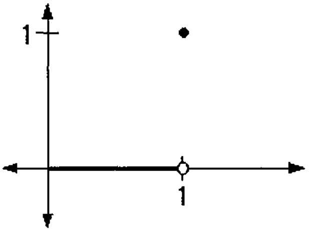

Clearly $f(x)$ is not continuous at the endpoint $x = 1$ while all of the functions $f_n(x) = x^n$ are continuous on the entire interval. But we will see that the sequence $\{f_n(x)\}$ does indeed converge pointwise to $f(x)$.

Fix α in $[0,1]$. If $\alpha = 1$, then $f_n(1) = 1^n = 1$ for all n. Then

$$\lim_{x \to \infty} f_n(1) = \lim_{n \to \infty} 1 = 1 = f(1).$$

Now let $0 \le \alpha < 1$. We will use (without proving) the fact that for any number α less than 1, the limit of α^n will approach 0 as n approaches ∞. In particular,

$$\begin{aligned} \lim_{n \to \infty} f_n(\alpha) &= \lim_{n \to \infty} \alpha^n \\ &= 0 \\ &= f(\alpha). \end{aligned}$$

Thus the pointwise limit of a sequence of continuous functions need not be continuous.

2.7 Uniform Convergence

Definition 2.7.1 *A sequence of functions $f_n : [a,b] \to \mathbf{R}$ will* converge uniformly *to a function $f : [a,b] \to \mathbf{R}$ if given any $\epsilon > 0$, there is a positive integer N such that for all $n \ge N$, we have*

$$|f(x) - f_n(x)| < \epsilon$$

for all points x.

The intuition is that if we put an ϵ-tube around the function $y = f(x)$, the functions $y = f_n(x)$ will eventually fit inside this band.

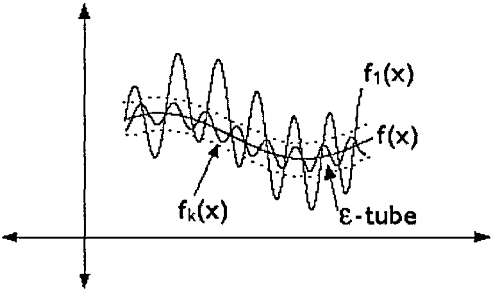

The key here is that the same ϵ and N will work for all x. This is not the case in the definition of pointwise convergence, where the choice of N depends on the number x.

Almost all of the desirable properties of the functions in the sequence will be inherited by the limit. The major exception is differentiability, but even here a partial result is true. As an example of how these arguments work, we will show

Theorem 2.7.1 *Let $f_n : [a,b] \rightarrow \mathbf{R}$ be a sequence of continuous functions converging uniformly to a function $f(x)$. Then $f(x)$ will be continuous.*

Proof: We need to show that for all α in $[a,b]$,

$$\lim_{x\rightarrow\alpha} f(x) = f(\alpha).$$

Thus, given any $\epsilon > 0$, we must find some $\delta > 0$ such that for $0 < |x-\alpha| < \delta$, we have

$$|f(x) - f(\alpha)| < \epsilon.$$

By uniform convergence, there is a positive integer N so that

$$|f(x) - f_N(x)| < \frac{\epsilon}{3}$$

for all x. (The reason for the $\frac{\epsilon}{3}$ will be seen in a moment.)

By assumption each function $f_N(x)$ is continuous at the point α. Thus there is a $\delta > 0$ such that for $0 < |x - \alpha| < \delta$, we have

$$|f_N(x) - f_N(\alpha)| < \frac{\epsilon}{3}.$$

Now to show that for $0 < |x - \alpha| < \delta$, we will have

$$|f(x) - f(\alpha)| < \epsilon.$$

We will use the trick of adding appropriate terms which sum to zero and then applying the triangle inequality ($|A+B| \leq |A| + |B|$). We have

$$\begin{aligned} |f(x) - f(\alpha)| &= |f(x) - f_N(x) + f_N(x) - f_N(\alpha) + f_N(\alpha) - f(\alpha)\ | \\ &\leq |f(x) - f_N(x)| + |f_N(x) - f_N(\alpha)| + |f_N(\alpha) - f(\alpha)| \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} \\ &= \epsilon, \end{aligned}$$

and we are done. □

We can now make sense out of series (infinite sums) of functions.

Definition 2.7.2 *Let $f_1(x), f_2(x), \ldots$ be a sequence of functions. The series of functions*

$$f_1(x) + f_2(x) + \ldots = \sum_{k=1}^{\infty} f_k(x)$$

converges uniformly *to a function $f(x)$ if the sequence of partial sums: $f_1(x), f_1(x) + f_2(x), f_1(x) + f_2(x) + f_3(x), \ldots$ converges uniformly to $f(x)$.*

In terms of ϵ and $\delta's$, the infinite series of functions $\sum_{k=1}^{\infty} f_k(x)$ converges uniformly to $f(x)$ if given any $\epsilon > 0$ there is a positive integer N such that for all $n \geq N$,

$$|f(x) - \sum_{k=1}^{n} f_k(x)| < \epsilon,$$

for all x.

We have

Theorem 2.7.2 *If each function $f_k(x)$ is continuous and if $\sum_{k=1}^{\infty} f_k(x)$ converges uniformly to $f(x)$, then $f(x)$ must be continuous.*

This follows from the fact that the finite sum of continuous functions is continuous and the previous theorem.

The writing of a function as a series of uniformly converging (simpler) functions is a powerful method of understanding and working with functions. It is the key idea behind the development of both Taylor series and Fourier series (which is the topic of Chapter Thirteen).

2.8 The Weierstrass M-Test

If we are interested in infinite series of functions $\sum_{k=1}^{\infty} f_k(x)$, then we must be interested in knowing when the series converges uniformly. Luckily the

Weierstrass M-test provides a straightforward means for determining uniform convergence. As we will see, the key is that this theorem reduces the question of uniform convergence of $\sum_{k=1}^{\infty} f_k(x)$ to a question of when an infinite series of numbers converges, for which beginning calculus provides many tools, such as the ratio test, the root test, the comparison test, the integral test, etc.

Theorem 2.8.1 *Let $\sum_{k=1}^{\infty} f_k(x)$ be a series of functions, with each function $f_k(x)$ defined on a subset A of the real numbers. Suppose $\sum_{k=1}^{\infty} M_k$ is a series of numbers such that:*

1. $0 \leq |f_k(x)| \leq M_k$, *for all* $x \in A$.
2. *The series* $\sum_{k=1}^{\infty} M_k$ *converges.*

Then $\sum_{k=1}^{\infty} f_k(x)$ converges uniformly and absolutely.

By *absolute convergence*, we mean that the series of absolute values $\sum_{k=1}^{\infty} |f_k(x)|$ also converges uniformly.
Proof: To show uniform convergence, we must show that, given any $\epsilon > 0$, there exists an integer N such that for all $n \geq N$, we have

$$|\sum_{k=n}^{\infty} f_k(x)| < \epsilon,$$

for all $x \in A$. Whether or not $\sum_{k=n}^{\infty} f_k(x)$ converges, we certainly have

$$|\sum_{k=n}^{\infty} f_k(x)| \leq \sum_{k=n}^{\infty} |f_k(x)|.$$

Since $\sum_{k=1}^{\infty} M_k$ converges, we know that we can find an N so that for all $n \geq N$, we have

$$\sum_{k=n}^{\infty} M_k < \epsilon.$$

Since $0 \leq |f_k(x)| \leq M_k$, for all $x \in A$, we have

$$|\sum_{k=n}^{\infty} f_k(x)| \leq \sum_{k=n}^{\infty} |f_k(x)| \leq \sum_{k=n}^{\infty} M_k < \epsilon,$$

and we are done. □

Let us look an easy example. Consider the series $\sum_{k=1}^{\infty} \frac{x^k}{k!}$, which from calculus we know to be the Taylor series for e^x. We will use the Weierstrass

M-test to show that this series converges uniformly on any interval $[-a, a]$. Here we have $f_k(x) = \frac{x^k}{k!}$. Set

$$M_k = \frac{a^k}{k!}.$$

Note that for all $x \in [-a, a]$, we have $0 < |x|^n/n! \leq a^n/n!$. Thus if we can show that the series$\sum_{k=1}^{\infty} M_k = \sum_{k=1}^{\infty} \frac{a^k}{k!}$ converges, we will have uniform convergence. By the ratio test, $\sum_{k=1}^{\infty} \frac{a^k}{k!}$ will converge if the limit of ratios

$$\lim_{k\to\infty} \frac{M_{k+1}}{M_k} = \lim_{k\to\infty} \frac{\left(\frac{a^{k+1}}{(k+1)!}\right)}{\left(\frac{a^k}{k!}\right)}$$

exists and is strictly less than one. But we have

$$\lim_{k\to\infty} \frac{\frac{a^{k+1}}{(k+1)!}}{\frac{a^k}{k!}} = \lim_{k\to\infty} \frac{a}{(k+1)} = 0.$$

Thus the Taylor series for e^x will converge uniformly on any closed interval.

2.9 Weierstrass' Example

Our goal is find a function that is continuous everywhere but differentiable nowhere. When Weierstrass first constructed such functions in the late 1800s, mathematicians were shocked and surprised. The conventional wisdom of the time was that no such function could exist. The moral of this example is that one has to be careful of geometric intuition.

We will follow closely the presentation given by Spivak in his *Calculus* [102] in Chapter 23. We need a bit of notation. Set $\{x\}$ = distance from x to the nearest integer. For example, $\{\frac{3}{4}\} = \frac{1}{4}$ and $\{1.3289\} = .3289$, etc. The graph of $\{x\}$ is:

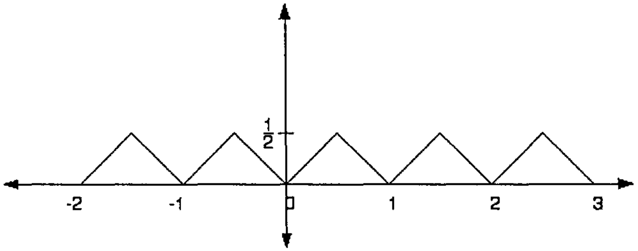

Define

$$f(x) = \sum_{k=1}^{\infty} \frac{1}{10^k}\{10^k x\}.$$

Our goal is:

Theorem 2.9.1 *The function $f(x)$ is continuous everywhere but differentiable nowhere.*

First for the intuition. For simplicity we restrict the domain to be the unit interval $(0, 1)$. For $k = 1$, we have the function $\frac{1}{10}$ $\{10x\}$, which has a graph:

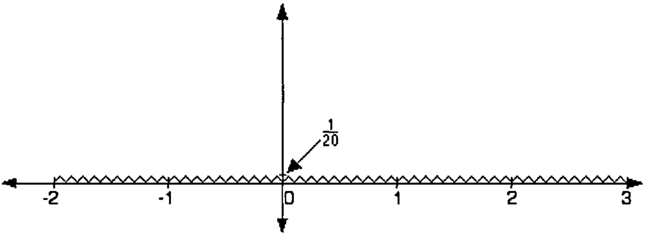

This function is continuous everywhere but not differentiable at the 19 points $.05, .1, .15, \ldots, .95$. Then $\{x\} + \frac{1}{10}\{10x\}$ has the graph:

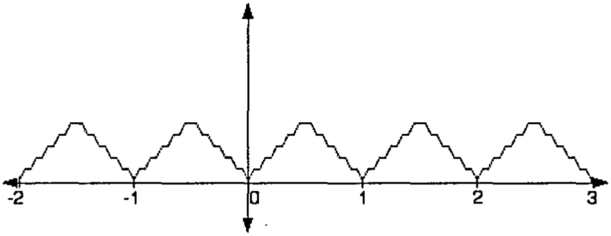

and is continuous everywhere but not differentiable at $.05, .1, .15, \ldots, .95$. For $k = 2$, the function $\frac{1}{100}$ $\{100x\}$ is continuous everywhere but is not differentiable at its 199 sharp points. Then the partial sum $\frac{1}{10}\{10x\} + \frac{1}{100}\{100x\}$ is continuous everywhere but not differentiable at the 199 sharp points. In a similar fashion, $\frac{1}{1000}$ $\{1000x\}$ is also continuous, but now loses differentiability at its 1999 sharp points. As we continue, at every sharp edge, we lose differentiability, but at no place is there a break in the graph. As we add all the terms in $\sum \frac{1}{10^k}\{10^k x\}$, we eventually lose differentiability at every point. The pictures are compelling, but of course we need a proof.

Proof: (We continue to follow Spivak)

The easy part is in showing that $f(x) = \sum_{k=1}^{\infty} \frac{1}{10^k}\{10^k x\}$ is continuous, as this will be a simple application of the Weierstrass M-test. We know that $\{x\} \leq \frac{1}{2}$ for all x. Thus we have, for all k, that

$$\frac{1}{10^k}\{10^k x\} \leq \frac{1}{2 \cdot 10^k}.$$

The series

$$\sum_{k=1}^{\infty} \frac{1}{2 \cdot 10^k} = \frac{1}{2}\sum_{k=1}^{\infty} \frac{1}{10^k}$$

is a geometric series and thus must converge (just use the ratio test). Then by the Weierstrass M-test, the series $f(x) = \sum_{k=1}^{\infty} \frac{1}{10^k}\{10^k x\}$ converges uniformly. Since each function $\frac{1}{10^k}\{10^k x\}$ is continuous, we have that $f(x)$ must be continuous.

It is much harder to show that $f(x)$ is not differentiable at every point; this will take some delicate work. Fix any x. We must show that

$$\lim_{h \to \infty} \frac{f(x+h) - f(x)}{h}$$

does not exist. We will find a sequence, h_m, of numbers that approach zero such that the sequence $\frac{f(x+h_m)-f(x)}{h_m}$ does not converge.

Write x in its decimal expansion:

$$x = a.a_1 a_2 \ldots,$$

where a is zero or one and each a_k is an integer between zero and nine. Set

$$h_m = \begin{cases} 10^{-m} & \text{if } a_m \neq 4 \text{ or if } a_m \neq 9 \\ -10^{-m} & \text{if } a_m = 4 \text{ or if } a_m = 9 \end{cases}$$

Then

$$x + h_m = \begin{cases} a.a_1 \ldots (a_m + 1)a_{m+1} \ldots & \text{if } a_m \neq 4 \text{ or if } a_m \neq 9 \\ a.a_1 \ldots (a_m - 1)a_{m+1} \ldots & \text{if } a_m = 4 \text{ or if } a_m = 9 \end{cases}$$

We will be looking at various $10^n(x + h_m)$. The 10^n factor just shifts where the decimal point lands. In particular, if $n > m$, then

$$10^n(x + h_m) = aa_1 \ldots (a_m \pm 1)a_{m+1} \ldots a_n.a_{n+1} \ldots,$$

in which case

$$\{10^n(x + h_m)\} = \{10^n x\}.$$

If $n \leq m$, then $10^n(x + h_m) = aa_1 \ldots a_n.a_{n+1} \ldots (a_m \pm 1)a_{m+1} \ldots$, in which case we have

$$\{10^n(x + h_m)\} = \begin{cases} 0.a_{n+1} \ldots (a_m + 1)a_{m+1} \ldots & \text{if } a_m \neq 4 \text{ or if } a_m \neq 9 \\ 0.a_{n+1} \ldots (a_m - 1)a_{m+1} \ldots & \text{if } a_m = 4 \text{ or if } a_m = 9 \end{cases}$$

We are interested in the limit of

$$\frac{f(x+h_m)-f(x)}{h_m}=\sum_{k=0}^{\infty}\frac{\frac{1}{10^k}\{10^k(x+h_m)\}-\frac{1}{10^k}\{10^k x\}}{h_m}.$$

Since $\{10^k(x+h_m)\}=\{10^k x\}$, for $k>m$, the above infinite series is actually the finite sum:

$$\sum_{k=0}^{m}\frac{\frac{1}{10^k}\{10^k(x+h_m)\}-\frac{1}{10^k}\{10^k x\}}{h_m}=\sum_{k=0}^{m}\pm 10^{m-k}(\{10^k(x+h_m)\}-\{10^k x\}).$$

We will show that each $\pm 10^{m-k}(\{10^k(x+h_m)\}-\{10^k x\})$ is a plus or minus one. Then the above finite sum is a sum of plus and minus ones and thus cannot be converging to a number, showing that the function is not differentiable.

There are two cases. Still following Spivak, we will only consider the case when $10^k x=.a_{k+1}\ldots<\frac{1}{2}$ (the case when $.a_{k+1}\ldots\geq\frac{1}{2}$ is left to the reader). Here is why we had to break our definition of the h_m into two separate cases. By our choice of h_m, $\{10^k(x+h_m)\}$ and $\{10^k x\}$ differ only in the $(m-k)$th term of the decimal expansion. Thus

$$\{10^k(x+h_m)\}-\{10^k x\}=\pm\frac{1}{10^{m-k}}.$$

Then $10^{m-k}(\{10^k(x+h_m)\}-\{10^k x\})$ will be, as predicted, a plus or minus one. □

2.10 Books

The development of ϵ and δ analysis was one of the main triumphs of 1800s mathematics; this means that undergraduates for most of the last hundred years have had to learn these techniques. There are many texts. The one that I learned from and one of my favorite math books of all times is Michael Spivak's *Calculus* [102]. Though called a calculus book, even Spivak admits, in the preface to the second and third editions, that a more apt title would be "An Introduction to Real Analysis". The exposition is wonderful and the problems are excellent.

Other texts for this level of real analysis include books by Bartle [6], Berberian [7], Bressoud [13], Lang [80], Protter and Morrey [94] and Rudin [96], among many others.

2.11 Exercises

1. Let $f(x)$ and $g(x)$ be differentiable functions. Using the definition of derivatives, show

a. $(f+g)' = f' + g'$.

b. $(fg)' = f'g + fg'$.

c. Assume that $f(x) = c$, where c is a constant. Show that the derivative of $f(x)$ is zero.

2. Let $f(x)$ and $g(x)$ be integrable functions.

a. Using the definition of integration, show that the sum $f(x) + g(x)$ is an integrable function.

b. Using the Fundamental Theorem of Calculus and problem 1.a, show that the sum $f(x) + g(x)$ is an integrable function.

3. The goal of this problem is to calculate $\int_0^1 x\,\mathrm{d}x$ three ways. The first two methods are not supposed to be challenging.

a. Look at the graph of the function $y = x$. Note what type of geometric object this is, and then get the area under the curve.

b. Find a function $f(x)$ such that $f'(x) = x$ and then use the Fundamental Theorem of Calculus to find $\int_0^1 x\,\mathrm{d}x$.

c. This has two parts. First show by induction that

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}.$$

Then use the definition of the integral to find $\int_0^1 x\,\mathrm{d}x$.

4. Let $f(x)$ be differentiable. Show that $f(x)$ must be continuous. (Note: intuitively this makes a lot of sense; after all, if the function f has breaks in its graph, it should not then have well-defined tangents. This problem is an exercise in the definitions.)

5. On the interval $[0, 1]$, define

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is not rational} \end{cases}$$

Show that $f(x)$ is not integrable. (Note: you will need to use the fact that any interval of any positive length must contain a rational number and an irrational number. In other words, both the rational and the irrational numbers are dense.)

6. This is a time-consuming problem but is very worthwhile. Find a calculus textbook. Go through its proof of the chain-rule, namely that

$$\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x)) \cdot g'(x).$$

7. Go again to the calculus book that you used in problem six. Find the chapter on infinite series. Go carefully through the proofs for the following tests for convergence: the integral test, the comparison test, the limit comparison test, the ratio test and the root test. Put all of these tests into the language of ϵ and δ real analysis.

Chapter 3

Calculus for Vector-Valued Functions

Basic Object:	$\mathbf{R}^n$
Basic Map:	Differentiable functions $f : \mathbf{R}^n \to \mathbf{R}^m$
Basic Goal:	Inverse Function Theorem

3.1 Vector-Valued Functions

A function $f : \mathbf{R}^n \to \mathbf{R}^m$ is called *vector-valued* since for any vector x in $\mathbf{R}^n$, the value (or image) of $f(x)$ is a vector in $\mathbf{R}^m$. If $(x_1, \ldots, x_n)$ is a coordinate system for $\mathbf{R}^n$, the function f can be described in terms of m real-valued functions by simply writing:

$$f(x_1, \ldots, x_n) = \begin{pmatrix} f_1(x_1, \cdots, x_n) \\ \vdots \\ \vdots \\ f_m(x_1 \ldots, x_n) \end{pmatrix}$$

Such functions occur everywhere. For example, let $f : \mathbf{R} \to \mathbf{R}^2$ be defined as

$$f(t) = \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}.$$

Here t is the coordinate for $\mathbf{R}$. Of course this is just the unit circle parametrized by its angle with the x-axis.

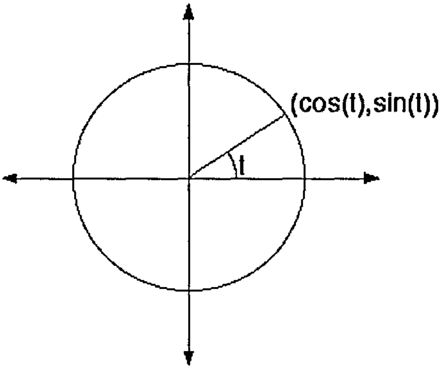

This can also be written as $x = \cos(t)$ and $y = \sin(t)$.

For another example, consider the function $f : \mathbf{R}^2 \to \mathbf{R}^3$ given by

$$f(x_1, x_2) = \begin{pmatrix} \cos x_1 \\ \sin x_1 \\ x_2 \end{pmatrix}.$$

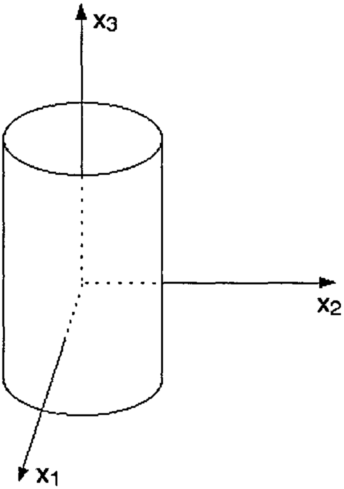

This function f maps the (x_1, x_2) plane to a cylinder in space.

Most examples are quite a bit more complicated, too complicated for pictures to even be drawn, much less used.

3.2 Limits and Continuity of Vector-Valued Functions

The key idea in defining limits for vector-valued functions is that the Pythagorean Theorem gives a natural way for measuring distance in $\mathbf{R}^n$.

Definition 3.2.1 *Let* $a = (a_1, \ldots, a_n)$ *and* $b = (b_1, \ldots, b_n)$ *be two points in* $\mathbf{R}^n$. *Then the* distance *between* a *and* b, *denoted by* $|a - b|$, *is*

$$|a - b| = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2}.$$

The length *of* a *is defined by*

$$|a| = \sqrt{a_1^2 + \cdots + a_n^2}.$$

Note that we are using the word "length" since we can think of the point a in $\mathbf{R}^n$ as a vector from the origin to the point.

Once we have a notion of distance, we can apply the standard tools from ϵ and δ style real analysis. For example, the reasonable definition of limit must be:

Definition 3.2.2 *The function* $f : \mathbf{R}^n \to \mathbf{R}^m$ *has* limit

$$L = (L_1, \ldots, L_m) \in \mathbf{R}^m$$

at the point $a = (a_1, \ldots, a_n) \in \mathbf{R}^n$ *if given any* $\epsilon > 0$, *there is some* $\delta > 0$ *such that for all* $x \in \mathbf{R}^n$, *if*

$$0 < | x - a | < \delta,$$

we have

$$|f(x) - L| < \epsilon.$$

We denote this limit by

$$\lim_{x \to a} f(x) = L$$

or by $f(x) \to L$ *as* $x \to a$.

Of course, continuity must now be defined by:

Definition 3.2.3 *The function* $f : \mathbf{R}^n \to \mathbf{R}^m$ *is* continuous *at a point* a *in* $\mathbf{R}^n$ *if* $\lim_{x \to a} f(x) = f(a)$.

Both the definitions of limit and continuity rely on the existence of a distance. Given different norms (distances) we will have corresponding definitions for limits and for continuity.

3.3 Differentiation and Jacobians

For single variable functions, the derivative is the slope of the tangent line (which is, recall, the best linear approximation to the graph of the original function) and can be used to find the equation for this tangent line. In a similar fashion, we want the derivative of a vector-valued function to be a tool that can be used to find the best linear approximation to the function.

We will first give the definition for the vector-valued derivative and then discuss the intuitions behind it. In particular we want this definition for vector-valued functions to agree with the earlier definition of a derivative for the case of single variable real-valued functions.

Definition 3.3.1 *A function* $f : \mathbf{R}^n \to \mathbf{R}^m$ *is* differentiable *at* $a \in \mathbf{R}^n$ *if there is an* $m \times n$ *matrix* $A : \mathbf{R}^n \to \mathbf{R}^m$ *such that*

$$\lim_{x \to a} \frac{|f(x) - f(a) - A \cdot (x - a)|}{|x - a|} = 0.$$

If such a limit exists, the matrix A *is denoted by* $Df(a)$ *and is called the* Jacobian

Note that $f(x), f(a)$ and $A \cdot (x - a)$ are all in $\mathbf{R}^m$ and hence

$$|f(x) - f(a) - A \cdot (x - a)|$$

is the length of a vector in $\mathbf{R}^m$. Likewise, $x - a$ is a vector in $\mathbf{R}^n$, forcing $|x - a|$ to be the length of a vector in $\mathbf{R}^n$. Further, usually there is an easy way to compute the matrix A, which we will see in a moment. Also, if the Jacobian matrix $Df(a)$ exists, one can show that it is unique, up to change of bases for $\mathbf{R}^n$ and $\mathbf{R}^m$.

We definitely want this definition to agree with the usual definition of derivative for a function $f : \mathbf{R} \to \mathbf{R}$. With $f : \mathbf{R} \to \mathbf{R}$, recall that the derivative $f'(a)$ was defined to be the limit

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}.$$

Unfortunately, for a vector-valued function $f : \mathbf{R}^n \to \mathbf{R}^m$ with n and m larger than one, this one-variable definition is nonsensical, since we cannot divide vectors. We can, however, algebraically manipulate the above one-variable limit until we have a statement that can be naturally generalized to functions $f : \mathbf{R}^n \to \mathbf{R}^m$ and which will agree with our definition.

Return to the one-variable case $f : \mathbf{R} \to \mathbf{R}$. Then

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

is true if and only if

$$0=\lim_{x\to a}\frac{f(x)-f(a)}{x-a}-f'(a),$$

which is equivalent to

$$0=\lim_{x\to a}\frac{f(x)-f(a)-f'(a)(x-a)}{x-a}$$

or

$$0=\lim_{x\to a}\frac{|f(x)-f(a)-f'(a)(x-a)|}{|x-a|}.$$

This last statement, at least formally, makes sense for functions $f:\mathbf{R}^n\to\mathbf{R}^m$, provided we replace $f'(a)$ (a number and hence a 1×1 matrix) by an $m\times n$ matrix, namely the Jacobian $Df(a)$.

As with the one-variable derivative, there is a (usually) straightforward method for computing the derivative without resorting to the actual taking of a limit, allowing us to actually calculate the Jacobian.

Theorem 3.3.1 *Let the function $f:\mathbf{R}^n\to\mathbf{R}^m$ be given by the m differentiable functions $f_1(x_1,\ldots,x_n),\ldots,f_m(x_1,\ldots,x_n)$, so that*

$$f(x_1,\ldots,x_n)=\begin{pmatrix} f_1(x_1,\ldots,x_n)\\ \vdots\\ \vdots\\ f_m(x_1,\ldots,x_n)\end{pmatrix}$$

Then f is differentiable and the Jacobian is

$$Df(x)=\begin{pmatrix}\frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n}\\ \vdots & & \vdots\\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n}\end{pmatrix}$$

The proof, found in most books on vector calculus, is a relatively straightforward calculation stemming from the definition of partial derivatives. But to understand it, we look at the following example. Consider our earlier example of the function $f:\mathbf{R}^2\to\mathbf{R}^3$ given by

$$f(x_1,x_2)=\begin{pmatrix}\cos x_1\\ \sin x_1\\ x_2\end{pmatrix},$$

which maps the (x_1, x_2) plane to a cylinder in space. Then the Jacobian, the derivative of this vector-valued function, will be

$$\begin{aligned} Df(x_1,x_2) &= \begin{pmatrix} \partial\cos(x_1)/\partial x_1 & \partial\cos(x_1)/\partial x_2 \\ \partial(\sin x_1)/\partial x_1 & \partial\sin(x_1)/\partial x_2 \\ \partial x_2/\partial x_1 & \partial x_2/\partial x_2 \end{pmatrix} \\ &= \begin{pmatrix} -sinx_1 & 0 \\ cosx_1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

One of the most difficult concepts and techniques in beginning calculus is the chain rule, which tells us how to differentiate the composition of two functions. For vector-valued forms, the chain rule can be easily stated (though we will not give the proof here). It should relate the derivative of the composition of functions with the derivatives of each component part and in fact has a quite clean flavor, namely:

Theorem 3.3.2 *Let* $f : \mathbf{R}^n \to \mathbf{R}^m$ *and* $g : \mathbf{R}^m \to \mathbf{R}^l$ *be differentiable functions. Then the composition function*

$$g \circ f : \mathbf{R}^n \to \mathbf{R}^l$$

is also differentiable with derivative given by: if $f(a) = b$, *then*

$$D(g \circ f)(a) = D(g)(b) \cdot D(f)(a).$$

Thus the chain rule says that to find the derivative of the composition $g \circ f$, one multiplies the Jacobian matrix for g times the Jacobian matrix for f.

One of the key intuitions behind the one-variable derivative is that $f'(a)$ is the slope of the tangent line to the curve $y = f(x)$ at the point $(a, f(a))$ in the plane $\mathbf{R}^2$. In fact, the tangent line through $(a, f(a))$ will have the equation

$$y = f(a) + f'(a)(x - a).$$

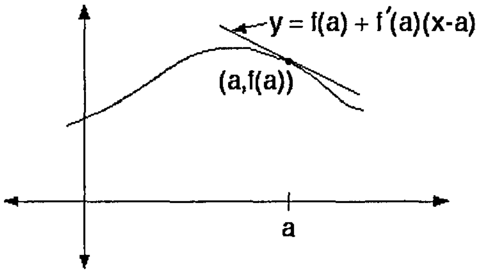

This line $y = f(a) + f'(a)(x - a)$ is the closest linear approximation to the function $y = f(x)$ at $x = a$.

Thus a reasonable criterion for the derivative of $f : \mathbf{R}^n \rightarrow \mathbf{R}^m$ should be that we can use this derivative to find a linear approximation to the geometric object $y = f(x)$, which lies in the space $\mathbf{R}^{n+m}$. But this is precisely what the definition

$$\lim_{x \rightarrow a} \frac{| f(x) - f(a) - Df(a)(x - a) |}{| x - a |} = 0$$

does. Namely, $f(x)$ is approximately equal to the linear function

$$f(a) + Df(a) \cdot (x - a).$$

Here $Df(a)$, as an $m \times n$ matrix, is a linear map from $\mathbf{R}^n \rightarrow \mathbf{R}^m$ and $f(a)$, as an element of $\mathbf{R}^m$, is a translation. Thus the vector $y = f(x)$ can be approximated by

$$y \approx f(a) + Df(a) \cdot (x - a).$$

3.4 The Inverse Function Theorem

Matrices are easy to understand, while vector-valued functions can be quite confusing. As seen in the last section, one of the points of having a derivative for vector-valued functions is that we can approximate the original function by a matrix, namely the Jacobian. The general question is now how good of an approximation do we have. What decent properties for matrices can be used to get corresponding decent properties for vector-valued functions?

This type of question could lead us to the heart of numerical analysis. We will limit ourselves to seeing that if the derivative matrix (the Jacobian) is invertible, then the original vector-valued function must also have an inverse, at least locally. This theorem, and its close relative the Implicit Function Theorem, are key technical tools that appear throughout mathematics.

Theorem 3.4.1 (Inverse Function Theorem) *For a vector-valued continuously differentiable function* $f : \mathbf{R}^n \rightarrow \mathbf{R}^m$, *assume that* $\det Df(a) \neq 0$, *at some point* a *in* $\mathbf{R}^n$. *Then there is an open neighborhood* U *of* a *in* $\mathbf{R}^n$ *and an open neighborhood* V *of* $f(a)$ *in* $\mathbf{R}^m$ *such that* $f : U \rightarrow V$ *is one to one, onto and has a differentiable inverse* $g : V \rightarrow U$ *(i.e.,* $g \circ f : U \rightarrow U$ *is the identity and* $f \circ g : V \rightarrow V$ *is the identity).*

Why should a function f have an inverse? Let us think of f as being approximated by the linear function

$$f(x) \approx f(a) + Df(a) \cdot (x - a).$$

From the key theorem of linear algebra, the matrix $Df(a)$ is invertible if and only if $\det Df(a) \neq 0$. Thus $f(x)$ should be invertible if $f(a)+Df(a)\cdot(x-a)$ is invertible, which should happen precisely when $\det Df(a) \neq 0$. In fact, consider

$$y = f(a) + Df(a) \cdot (x - a).$$

Here the vector y is written explicitly as a function of the variable vector x. But if the inverse to $Df(a)$ exists, then we can write x explicitly as a function of y, namely as:

$$x = a + Df(a)^{-1} \cdot (y - f(a)).$$

In particular, we should have, if the inverse function is denoted by f^{-1}, that its derivative is simply the inverse of the derivative of the original function f, namely

$$Df^{-1}(b) = Df(a)^{-1},$$

where $b = f(a)$. This follows from the chain rule and since the composition is $f^{-1} \circ f = I$.

For the case of $f : \mathbf{R} \to \mathbf{R}$, the idea behind the Inverse Function Theorem can be captured in pictures:

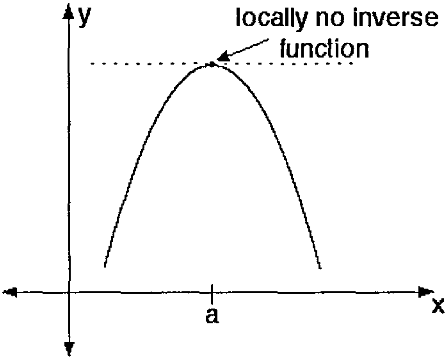

If the slope of the tangent line, $f'(a)$, is not zero, the tangent line will not be horizontal, and hence there will be an inverse.

In the statement of the theorem, we used the technical term "open set". There will be much more about this in the next chapter on topology. For now, think of an open set as a technical means allowing us to talk about all points near the points a and $f(a)$. More precisely, by an open neighborhood U of a point a in $\mathbf{R}^n$, we mean that, given any $a \in U$, there is a (small) positive ϵ such that

$$\{x : |x - a| < \epsilon\} \subset U.$$

In pictures, for example,

$$\{(x, y) \in \mathbf{R}^2 : |(x, y) - (0, 0)| = \sqrt{x^2 + y^2} \leq 1\}$$

is not open (it is in fact closed, meaning that its complement is open in the plane $\mathbf{R}^2$),

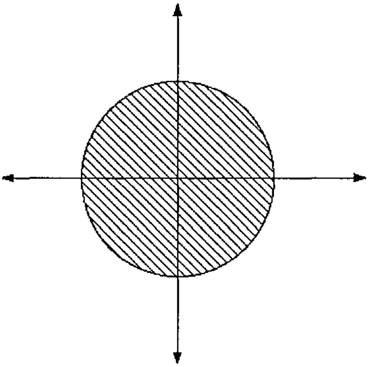

while the set

$$\{(x, y) \in \mathbf{R}^2 : |(x, y) - (0, 0)| < 1\}$$

is open.

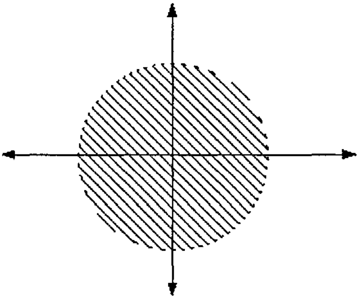

3.5 Implicit Function Theorem

Rarely can a curve in the plane be described as the graph of a one-variable function

$$y = f(x),$$

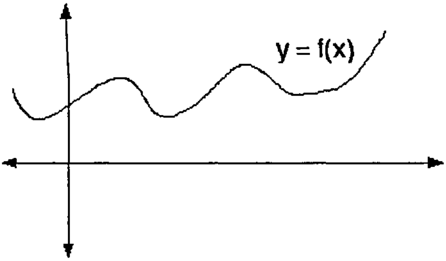

though much of our early mathematical experiences are with such functions. For example, it is impossible to write the circle

$$x^2 + y^2 = 1$$

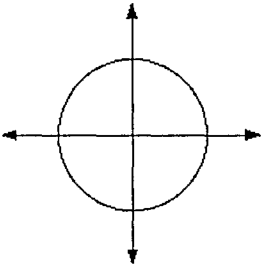

as the graph of a one-variable function, since for any value of x (besides -1 and 1) there are either no corresponding values of y on the circle or two corresponding values of y on the circle. This is unfortunate. Curves in the plane that can be cleanly written as $y = f(x)$ are simply easier to work with.

However, we can split the circle into its top and bottom halves.

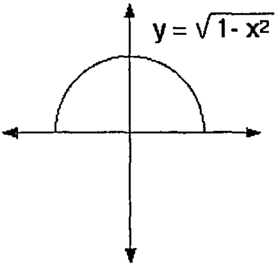

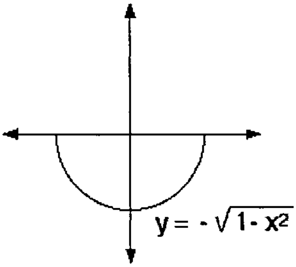

For each half, the variable y can be written as a function of x: for the top half, we have

$$y = \sqrt{1-x^2},$$

and for the bottom half,

$$y = -\sqrt{1-x^2}.$$

Only at the two points $(1,0)$ and $(-1,0)$ are there problems. The difficulty can be traced to the fact that at these two points (and only at these two points) the tangent lines of the circle are perpendicular to the x-axis.

This is the key. The tangent line of a circle is the best linear approximation to the circle. If the tangent line can be written as

$$y = mx + b,$$

then it should be no surprise that the circle can be written as $y = f(x)$, at least locally.

The goal of the Implicit Function Theorem is to find a computational tool that will allow us to determine when the zero locus of a bunch of functions in some $\mathbf{R}^N$ can locally be written as the graph of a function and thus in the form $y = f(x)$, where the x denote the independent variables and the y will denote the dependent variables. Buried (not too deeply) is the intuition that we want to know about the tangent space of the zero locus of functions.

The notation is a bit cumbersome. Label a coordinate system for $\mathbf{R}^{n+k}$ by

$$x_1, \ldots, x_n, y_1, \ldots, y_k$$

which we will frequently abbreviate as (x, y). Let

$$f_1(x_1, \ldots, x_n, y_1, \ldots, y_k), \ldots, f_k(x_1, \ldots, x_n, y_1, \ldots, y_k)$$

be k continuously differentiable functions, which will frequently be written as

$$f_1(x,y), \ldots, f_k(x,y).$$

Set

$$V = \{(x,y) \in \mathbf{R}^{n+m} : f_1(x,y) = 0, \ldots, f_k(x,y) = 0\}.$$

We want to determine when, given a point $(a,b) \in V$ (where $a \in \mathbf{R}^n$ and $b \in \mathbf{R}^k$), there are k functions

$$\rho_1(x_1, \ldots, x_n), \ldots, \rho_k(x_1, \ldots, x_n)$$

defined in a neighborhood of the point a on $\mathbf{R}^n$ such that V can be described, in a neighborhood of (a, b) on $\mathbf{R}^{n+k}$, as

$$\{(x, y) \in \mathbf{R}^{n+k} : y_1 = \rho_1(x_1, \ldots, x_n), \ldots, y_k = \rho_k(x_1, \ldots, x_n)\},$$

which of course is frequently written in the shorthand of

$$V = \{y_1 = \rho_1(x), \ldots, y_k = \rho_k(x)\},$$

or even more succinctly as

$$V = \{y = \rho(x)\}.$$

Thus we want to find k functions $\rho_1, \ldots, \rho_k$ such that for all $x \in \mathbf{R}^n$, we have

$$f_1(x, \rho_1(x)) = 0, \ldots, f_k(x, \rho_k(x)) = 0.$$

Thus we want to know when the k functions $f_1, \ldots, f_k$ can be used to define (implicitly, since it does take work to actually construct them) the k functions $\rho_1, \ldots, \rho_k$.

Theorem 3.5.1 (Implicit Function Theorem) *Let $f_1(x, y), \ldots, f_k(x, y)$ be k continuously differentiable functions on $\mathbf{R}^{n+k}$ and suppose that $p = (a, b) \in \mathbf{R}^{n+k}$ is a point for which*

$$f_1(a, b) = 0, \ldots, f_k(a, b) = 0.$$

Suppose that at the point p the $k \times k$ matrix

$$M = \begin{pmatrix} \frac{\partial f_1}{\partial y_1(p)} & \cdots & \frac{\partial f_1}{\partial y_k(p)} \\ \vdots & & \vdots \\ \frac{\partial f_k}{\partial y_1(p)} & \cdots & \frac{\partial f_k}{\partial y_k(p)} \end{pmatrix}$$

is invertible. Then in a neighborhood of a in $\mathbf{R}^n$ there are k unique, differentiable functions

$$\rho_1(x), \ldots, \rho_k(x)$$

such that

$$f_1(x, \rho_1(x)) = 0, \ldots, f_k(x, \rho_k(x)) = 0.$$

Return to the circle. Here the function is $f(x, y) = x^2 + y^2 - 1 = 0$. The matrix M in the theorem will be the 1×1 matrix:

$$\frac{\partial f}{\partial y_1} = 2y.$$

This matrix is not invertible (the number is zero) only where $y = 0$, namely at the two points $(1, 0)$ and $(-1, 0)$: only at these two points will there not be an implicitly defined function ρ.

Now to sketch the main ideas of the proof, whose outline we got from [103] In fact, this theorem is a fairly easy consequence of the Inverse Function Theorem. For ease of notation, write the k-tuple $(f_1(x, y), \dots, f_k(x, y))$ as $f(x, y)$. Define a new function $F : \mathbf{R}^{n+k} \to \mathbf{R}^{n+k}$ by

$$F(x, y) = (x, f(x, y)).$$

The Jacobian of this map is the $(n + k) \times (n + k)$ matrix

$$\begin{pmatrix} I & 0 \\ * & M \end{pmatrix}.$$

Here the I is the $n \times n$ identity matrix, M is the $k \times k$ matrix of partials as in the theorem, 0 is the $n \times k$ zero matrix and $*$ is some $k \times n$ matrix. Then the determinant of the Jacobian will be the determinant of the matrix M; hence the Jacobian is invertible if and only if the matrix M is invertible. By the Inverse Function Theorem, there will be a map $G : \mathbf{R}^{n+k} \to \mathbf{R}^{n+k}$ which will locally, in a neighborhood of the point (a, b), be the inverse of the map $F(x, y) = (x, f(x, y))$.

Let this inverse map $G : \mathbf{R}^{n+k} \to \mathbf{R}^{n+k}$ be described by the real-valued functions $G_1, \dots, G_{n+k}$ and thus as

$$G(x, y) = (G_1(x, y), \dots, G_{n+k}(x, y)).$$

By the nature of the map F, we see that for $1 \leq i \leq n$,

$$G_i(x, y) = x_i.$$

Relabel the last k functions that make up the map G by setting

$$\rho_i(x, y) = G_{i+n}(x, y).$$

Thus

$$G(x, y) = (x_1, \dots, x_n, \rho_1(x, y), \dots, \rho_k(x, y)).$$

We want to show that the functions $\rho_i(x, 0)$ are the functions the theorem requires.

We have yet looked at the set of points in $\mathbf{R}^{n+k}$ where the original k functions f_i are zero, namely the set that we earlier called V. The image of V under the map F will be contained in the set $(x, 0)$. Then the image $G(x, 0)$, at least locally around (a, b), will be V. Thus we must have

$$f_1(G(x, 0)) = 0, \dots, f_k(G(x, 0)) = 0.$$

But this just means that

$$f_1(x, \rho_1(x,0)) = 0, \ldots, f_k(x, \rho_k(x,0)) = 0,$$

which is exactly what we wanted to show.

Here we used the Inverse Function Theorem to prove the Implicit Function Theorem. It is certainly possible and no harder to prove the Implicit Function Theorem first and then use it to prove the Inverse Function Theorem.

3.6 Books

An excellent recent book on vector calculus (and for linear algebra and Stokes' Theorem) is by Hubbard and Hubbard [64]. Fleming [37] has been the standard reference for many years. Another, more abstract approach, is in Spivak's *Calculus on Manifolds* [103]. Information on vector calculus for three variable functions is in most calculus books. A good general exercise is to look in a calculus text and translate the given results into the language of this section.

3.7 Exercises

1. In the plane $\mathbf{R}^2$ there are two natural coordinate systems: polar coordinates (r, θ) with r the radius and θ the angle with the x-axis and Cartesian coordinates (x, y).

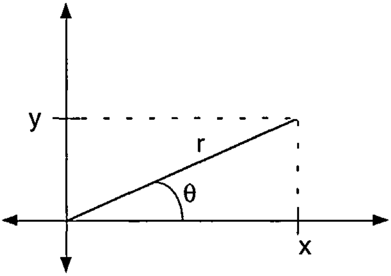

The functions that give the change of variables from polar to Cartesian coordinates are:

$$x = f(r, \theta) = rcos(\theta)$$

$$y = g(r, \theta) = rsin(\theta).$$

a. Compute the Jacobian of this change of coordinates.

b. At what points is the change of coordinates not well-defined (i.e., at what points is the change of coordinates not invertible)?

c. Give a geometric justification for your answer in part b.

2. There are two different ways of describing degree two monic polynomials in one variable: either by specifying the two roots or by specifying the coefficients. For example, we can describe the same polynomial by either stating that the roots are 1 and 2 or by writing it as $x^2 - 3x + 2$. The relation between the roots r_1 and r_2 and the coefficients a and b can be determined by noting that

$$(x - r_1)(x - r_2) = x^2 + ax + b.$$

Thus the space of all monic, degree two polynomials in one variable can be described by coordinates in the root space (r_1, r_2) or by coordinates in the coefficient space (a, b).

a. Write down the functions giving the change of coordinates from the root space to the coefficient space.

b. Compute the Jacobian of the coordinate change.

c. Find where this coordinate change is not invertible.

d. Give a geometric interpretation to your answer in part c.

3. Using the notation in the second question:

a. Via the quadratic equation, write down the functions giving the change of coordinates from the coordinate space to the root space.

b-d. Answer the same questions as in problem 2, but now for this new coordinate change.

4. Set $f(x, y) = x^2 - y^2$.

a. Graph the curve $f(x, y) = 0$.

b. Find the Jacobian of the function $f(x, y)$ at the point $(1, 1)$. Give a geometric interpretation of the Jacobian at this point.

c. Find the Jacobian of the function $f(x, y)$ at the point $(0, 0)$. Give a geometric interpretation for why the Jacobian is here the two-by-two zero matrix.

5. Set $f(x, y) = x^3 - y^2$.

a. Graph the curve $f(x, y) = 0$.

b. Find the Jacobian of the function $f(x, y)$ at the point $(1, 1)$. Give a geometric interpretation of the Jacobian at this point.

c. Find the Jacobian of the function $f(x, y)$ at the point $(0, 0)$. Give a geometric interpretation for why the Jacobian is here the two-by-two zero matrix.

Chapter 4

Point Set Topology

Basic Object:	Topological spaces
Basic Map:	Continuous functions

Historically, much of point set topology was developed to understand the correct definitions for such notions as continuity and dimension. By now, though, these definitions permeate mathematics, frequently in areas seemingly far removed from the traditional topological space $\mathbf{R}^n$. Unfortunately, it is not at first apparent that these more abstract definitions are at all useful; there needs to be an initial investment in learning the basic terms. In the first section, these basic definitions are given. In the next section, these definitions are applied to the topological space $\mathbf{R}^n$, where all is much more down to earth. Then we look at metric spaces. The last section applies these definitions to the Zariski topology of a commutative ring, which, while natural in algebraic geometry and algebraic number theory, is not at all similar to the topology of $\mathbf{R}^n$.

4.1 Basic Definitions

Much of point set topology consists in developing a convenient language to talk about when various points in a space are near to one another and about the notion of continuity. The key is that the same definitions can be applied to many disparate branches of math.

Definition 4.1.1 *Let X be a set of points. A collection of subsets $\mathbf{U} = \{U_\alpha\}$ forms a* topology *on X if*

1. *Any arbitrary union of the U_α is another set in the collection $\mathbf{U}$.*

2. *The intersection of any finite number of sets* U_α *in the collection* $\mathbf{U}$ *is another set in* $\mathbf{U}$.

3. *Both the empty set* ϕ *and the whole space* X *must be in* $\mathbf{U}$.

The $(X, \mathbf{U})$ *is called a* topological space.

The sets U_α in the collection $\mathbf{U}$ are called *open sets.* A set C is *closed* if its complement $X - C$ is open.

Definition 4.1.2 *Let* A *be a subset of a topological space* X. *Then the* induced topology *on* A *is described by letting the open sets on* A *be all sets of the form* $U \cap A$, *where* U *is an open set in* X.

A collection $\Sigma = \{U_\alpha\}$ of open sets is said to be an *open cover* of a subset A if A is contained in the union of the U_α.

Definition 4.1.3 *The subset* A *of a topological space* X *is* compact *if given any open cover of* A, *there is a finite subcover.*

In other words, if $\Sigma = \{U_\alpha\}$ is an open cover of A in X, then A being compact means that there are a finite number of the U_α, denoted let's say by $U_1, \ldots, U_n$, such that

$$A \subset (U_1 \cup U_2 \cup \ldots \cup U_n).$$

It should not be at all apparent why this definition would be useful, much less important. Part of its significance will be seen in the next section when we discuss the Heine-Borel Theorem.

Definition 4.1.4 *A topological space* X *is* Hausdorff *if given any two points* $x_1, x_2 \in X$, *there are two open sets* U_1, *and* U_2 *with* $x_1 \in U_1$ *and* $x_2 \in U_2$ *but with the intersection of* U_1 *and* U_2 *empty.*

Thus X is Hausdorff if points can be isolated (separated) from each other by disjoint open sets.

Definition 4.1.5 *A function* $f : X \to Y$ *is* continuous, *where* X *and* Y *are two topological spaces, if given any open set* U *in* Y, *then the inverse image* $f^{-1}(U)$ *in* X *must be open.*

Definition 4.1.6 *A topological space* X *is* connected *if it is not possible to find two open sets* U *and* V *in* X *with* $X = U \cup V$ *and* $U \cap V = \phi$.

Definition 4.1.7 *A topological space in X is* path connected *if given any two points a and b in X, there is a continuous map*

$$f : [0,1] \to X$$

with

$$f(0) = a \text{ and } f(1) = b.$$

Here of course

$$[0,1] = \{x \in \mathbf{R} : 0 \leq x \leq 1\}$$

is the unit interval. To make this last definition well-defined, we would need to put a topology on this interval $[0,1]$, but this is not hard and will in fact be done in the next section.

Though in the next section the standard topology on $\mathbf{R}^n$ will be developed, we will use this topology in order to construct a topological space that is connected but is not path connected. It must be emphasized that this is a pathology. In most cases, connected is equivalent to path connected.

Let

$$X = \{(0,t) : -1 \leq t \leq 1\} \cup \{y = sin(\frac{1}{x}) : x > 0\}.$$

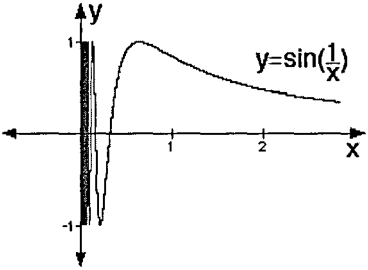

Put the induced topology on X from the standard topology on $\mathbf{R}^2$. Note that there is no path connecting the point $(0,0)$ to $(\frac{1}{\pi},0)$. In fact, no point on the segment $\{(0,t) ; -1 \leq t \leq 1\}$ can be connected by a path to any point on the curve $\{y = sin(\frac{1}{x}) : x > 0\}$. But on the other hand, the curve $\{y = sin(\frac{1}{x}) : x > 0\}$ gets arbitrarily close to the segment $\{(0,t) : -1 \leq t \leq 1\}$ and hence there is no way to separate the two parts by open sets.

Point set topology books would now give many further examples of various topological spaces which satisfy some but not all of the above conditions. Most have the feel, legitimately, of pathologies, creating in some the sense that all of these definitions are somewhat pedantic and not really essential. To counter this feel, in the last section of this chapter we will look at a nonstandard topology on commutative rings, the Zariski topology, which is definitely not a pathology. But first, in the next section, we must look at the standard topology on $\mathbf{R}^n$.

4.2 The Standard Topology on $\mathbf{R}^n$

Point set topology is definitely a product of the early twentieth century. However, long before that, people were using continuous functions and related ideas. Even in previous chapters, definitions were given for continuous functions, without the need to discuss open sets and topology. In this section we define the standard topology on $\mathbf{R}^n$ and show that the definition of continuity given in the last chapter in terms of limits agrees with the definition given in the last section in terms of inverse images of open sets. The important point is that the open set version can be used in contexts for which the limit notion makes no sense. Also, in practice the open set version is frequently no harder to use than the limit version.

Critical to the definition of the standard topology on $\mathbf{R}^n$ is that there is a natural notion of distance on $\mathbf{R}^n$. Recall that the distance between two points $a = (a_1, \dots, a_n)$ and $b = (b_1, \dots, b_n)$ in $\mathbf{R}^n$ is defined by

$$|a - b| = \sqrt{(a_1 - b_1)^2 + \ldots + (a_n - b_n)^2}.$$

With this, we can define a topology on $\mathbf{R}^n$ by specifying as the open sets the following:

Definition 4.2.1 *A set U in $\mathbf{R}^n$ will be* open *if given any $a \in \mathbf{R}^n$, there is a real number $\epsilon > 0$ such that*

$$\{x : |x - a| < \epsilon\}$$

is contained in U.

In $\mathbf{R}^1$, sets of the form $(a, b) = \{x : a < x < b\}$ are open, while sets of the form $[a, b] = \{x : a \le x \le b\}$ are closed. Sets like $[a, b) = \{x : a \le x < b\}$ are neither open nor closed. In $\mathbf{R}^2$, the set $\{(x, y) : x^2 + y^2 < 1\}$ is open.

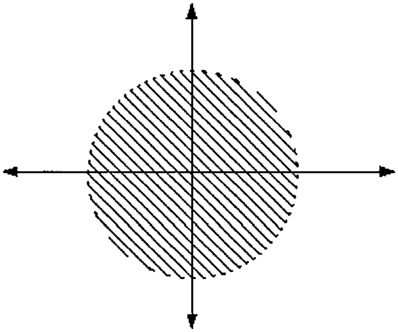

while $\{(x, y) : x^2 + y^2 \le 1\}$ is closed.

Proposition 4.2.1 *The above definition of an open set will define a topology on* $\mathbf{R}^n$.

(The proof is exercise 2 at the end of the chapter.) This is called the standard topology on R^n.

Proposition 4.2.2 *The standard topology on* $\mathbf{R}^n$ *is Hausdorff.*

This theorem is quite obvious geometrically:

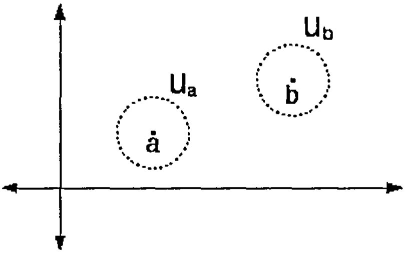

but we give a proof in order to test the definitions.

Proof: Let a and b be two distinct points in $\mathbf{R}^n$. Let $d = |a-b|$ be the distance from a to b. Set

$$U_a = \{x \in \mathbf{R}^n : |x-a| < \frac{d}{3}\}$$

and

$$U_b = \{x \in \mathbf{R}^n : |x-b| < \frac{d}{3}\}.$$

Both U_a and U_b are open sets with $a \in U_a$ and $b \in U_b$. Then $\mathbf{R}^n$ will be Hausdorff if

$$U_a \cap U_b = \phi.$$

Suppose that the intersection is not empty. Let $x \in U_a \cap U_b$. Then, by using the standard trick of adding terms that sum to zero and using the triangle inequality, we have

$$\begin{aligned} |a-b| &= |a-x+x-b| \\ &\leq |a-x| + |x-b| \\ &< \frac{d}{3} + \frac{d}{3} \\ &= \frac{2d}{3} \\ &< d. \end{aligned}$$

Since we cannot have $d = |a-b| < d$ and since the only assumption we made is that there is a point x in both U_a and U_b, we see that the intersection must indeed be empty. Hence the space $\mathbf{R}^n$ is Hausdorff. □

In Chapter Three, we defined a function $f : \mathbf{R}^n \to \mathbf{R}^m$ to be continuous if, for all $a \in \mathbf{R}^n$,

$$\lim_{x \to a} f(x) = f(a),$$

meaning that given any $\epsilon > 0$, there is some $\delta > 0$ such that if $|x - a| < \delta$, then

$$|f(x) - f(a)| < \epsilon.$$

This limit definition of continuity captures much of the intuitive idea that a function is continuous if it can be graphed without lifting the pen from the page. Certainly we want this previous definition of continuity to agree with our new definition that requires the inverse image of an open set to be open. Again, the justification for the inverse image version of continuity is that it can be extended to contexts where the limit version (much less the requirement of not lifting the pen from the page) makes no sense.

Proposition 4.2.3 *Let $f : \mathbf{R}^n \to \mathbf{R}^m$ be a function. For all $a \in \mathbf{R}^n$,*

$$\lim_{x \to a} f(x) = f(a)$$

if and only, if for any open set U in $\mathbf{R}^m$, the inverse image $f^{-1}(U)$ is open in $\mathbf{R}^n$.

Proof: First assume that the inverse image of every open set in $\mathbf{R}^m$ is open in $\mathbf{R}^n$. Let $a \in \mathbf{R}^n$. We must show that

$$\lim_{x \to a} f(x) = f(a).$$

Let $\epsilon > 0$. We must find some $\delta > 0$ so that if $|x - a| < \delta$, then

$$|f(x) - f(a)| < \epsilon.$$

Define

$$U = \{y \in \mathbf{R}^m : |y - f(a)| < \epsilon\}.$$

The set U is open in $\mathbf{R}^m$. By assumption the inverse image

$$\begin{aligned} f^{-1}(U) &= \{x \in \mathbf{R}^n : f(x) \in U\} \\ &= \{x \in \mathbf{R}^n : |f(x) - f(a)| < \epsilon\} \end{aligned}$$

is open in $\mathbf{R}^n$. Since $a \in f^{-1}(U)$, there is some real number $\delta > 0$ such that the set

$$\{x : |x - a| < \delta\}$$

is contained in $f^{-1}(U)$, by the definition of open set in $\mathbf{R}^n$. But then if $|x-a|<\delta$, we have $f(x) \in U$, or in other words,

$$|f(x)-f(a)|<\epsilon,$$

which is what we wanted to show. Hence the inverse image version of continuity implies the limit version.

Now assume that

$$\lim_{x \to a} f(x) = f(a).$$

Let U be any open set in $\mathbf{R}^m$. We need to show that the inverse $f^{-1}(U)$ is open in $\mathbf{R}^n$.

If $f^{-1}(U)$ is empty, we are done, since the empty set is always open. Now assume $f^{-1}(U)$ is not empty. Let $a \in f^{-1}(U)$. Then $f(a) \in U$. Since U is open, there is a real number $\epsilon > 0$ such that the set

$$\{y \in \mathbf{R}^m : |y - f(a)| < \epsilon\}$$

is contained in the set U. Since $\lim_{x \to a} f(x) = f(a)$, by the definition of limit, given this $\epsilon > 0$, there must be some $\delta > 0$ such that if $|x-a|<\delta$, then

$$|f(x)-f(a)|<\epsilon.$$

Therefore if $|x-a|<\delta$, then $f(x) \in U$. Thus the set

$$\{x : |x-a| < \delta\}$$

is contained in the set $f^{-1}(U)$, which means that $f^{-1}(U)$ is indeed an open set. Thus the two definitions of continuity agree. □

In the last section, a compact set was defined to be a set A on which every open cover $\Sigma = \{U_\alpha\}$ of A has a finite subcover. For the standard topology on $\mathbf{R}^n$, compactness is equivalent to the more intuitive idea that the set is compact if it is both closed and bounded. This equivalence is the goal of the Heine-Borel Theorem:

Theorem 4.2.1 (Heine-Borel) *A subset A of $\mathbf{R}^n$ is compact if and only if it is closed and bounded.*

We will first give a definition for boundedness, look at some examples and then sketch a proof of a special case of the theorem.

Definition 4.2.2 *A subset A is* bounded *in $\mathbf{R}^n$ if there is some fixed real number r such that for all $x \in A$,*

$$|x| < r$$

(i.e., A is contained in a ball of radius r).

For our first example, consider the open interval $(0,1)$ in $\mathbf{R}$, which is certainly bounded, but is not closed. We want to show that this interval is also not compact. Let

$$\begin{aligned} U_n &= \left(\frac{1}{n}, 1 - \frac{1}{n}\right) \\ &= \{x : \frac{1}{n} < x < 1 - \frac{1}{n}\} \end{aligned}$$

be a collection of open sets.

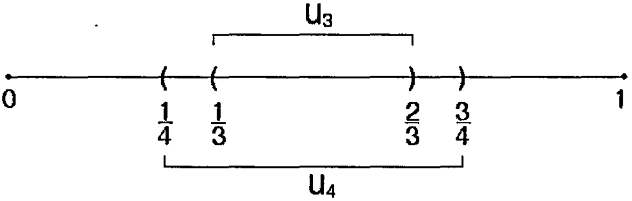

This collection will be an open cover of the interval, since every point in $(0,1)$ is in some U_n. (In fact, once a given point is in a set U_n, it will be in every future set U_{n+k}.) But note that no finite subcollection will cover the entire interval $(0,1)$. Thus $(0,1)$ cannot be compact.

The next example will be of a closed but not bounded interval. Again an explicit open cover will be given for which there is no finite subcover. The interval $[0,\infty) = \{x : 0 \leq x\}$ is closed but is most definitely not bounded. It also is not compact as can be seen with the following open cover:

$$U_n = (-1, n) = \{x : -1 < x < n\}.$$

The collection $\{U_n\}_{n=1}^{\infty}$, will cover $[0,\infty)$, but can contain no finite subcover.

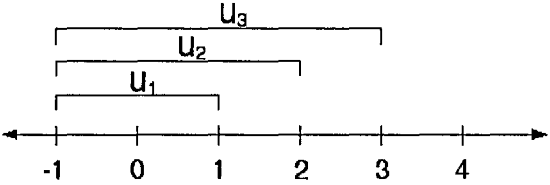

The proof of the Heine-Borel theorem revolves around reducing the whole argument to the special case of showing that a closed bounded interval on the real line is compact. (On how to reduce to this lemma, see the rigorous proof in Spivak [103], which is where we got the following argument.) This is the technical heart of the proof. The key idea actually pops up in a number of different contexts, which is why we give it here.

Lemma 4.2.1 *On the real line* $\mathbf{R}$*, a closed interval* $[a, b]$ *is compact.*

Proof: Let Σ be an open cover of $[a, b]$. We need to find a finite subcover. Define a new set

$$Y = \{x \in [a, b] : \text{there is a finite subcover in } \Sigma \text{ of the interval } [a, x]\}.$$

Our goal is to show that our interval's endpoint b is in this new set Y.

We will first show that Y is not empty, by showing that the initial point a is in Y. If $x = a$, then we are interested in the trivial interval $[a, a] = a$, a single point. Since Σ is an open cover, there is an open set $V \in \Sigma$ with $[a, a] \in V$. Thus for the admittedly silly interval $[a, a]$ there is a finite subcover, and thus a is in the set Y, meaning that, at the least, Y is not empty.

Set α to be the least upper bound of Y. This means that there are elements in Y arbitrarily close to α but that no element of Y is greater than α. (Though to show the existence of such a least upper bound involves the subtle and important property of completeness of the real number line, it is certainly quite reasonable intuitively that such an upper bound must exist for any bounded set of reals.) We first show that the point α is itself in the set Y and, second, that α is in fact the endpoint b, which will allow us to conclude that the interval is indeed compact.

Since $\alpha \in [a, b]$ and since Σ is an open cover, there is an open set U in Σ with $\alpha \in U$. Since U is open in $[a, b]$, there is a positive number ϵ with

$$\{x : |x - \alpha| < \epsilon\} \subset U.$$

Since α is the least upper bound of Y, there must be an $x \in Y$ that is arbitrarily close to but less than α. Thus we can find an $x \in Y \cap U$ with

$$\alpha - x < \epsilon,$$

Since $x \in Y$, there is a finite subcover $U_1, \ldots, U_N$ of the interval $[a, x]$. Then the finite collection $U_1, \ldots, U_N, U$ will cover $[a, \alpha]$. But this means, since each open set U_k and U are in Σ, that the interval $[a, \alpha]$ has a finite subcover and hence that the least upper bound α is in Y.

Now assume $\alpha < b$. We want to come up with a contradiction. We know that α is in the set Y. Hence there is a finite subcover $U_1, \ldots, U_n$ of the collection Σ which will cover the interval $[a, \alpha]$. Choose the open sets so that the point α is in the open set U_n. Since U_n is open, there is an $\epsilon > 0$ with

$$\{x : |x - \alpha| < \epsilon\} \subset U_n.$$

Since the endpoint b is strictly greater than the point α, we can actually find a point x that both is in the open set U_n and satisfies

$$\alpha < x < b.$$

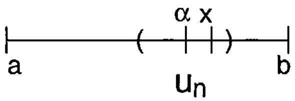

But then the finite subcover $U_1, \ldots, U_n$ will cover not only the interval $[a, \alpha]$ but also the larger interval $[a, x]$, forcing the point x to be in the set Y. This is impossible, since α is the largest possible element in Y. Since the only assumption that we made was that $\alpha < b$, we must have $\alpha = b$, as desired. □

There is yet another useful formulation for compactness in $\mathbf{R}^n$.

Theorem 4.2.2 *A subset A in $\mathbf{R}^n$ is compact if every infinite sequence (x_n) of points in A has a subsequence converging to a point in A. Thus, if (x_n) is a collection of points in A, there must be a point $p \in A$ and a subsequence x_{n_k} with $\lim_{k\to\infty} x_{n_k} = p$.*

The proof is one of the exercises at the end of the chapter.

Compactness is also critical for the following:

Theorem 4.2.3 *Let X be a compact topological space and let $f : X \to \mathbf{R}$ be a continuous function. Then there is a point $p \in X$ where f has a maximum.*

We give a general idea of the proof, with the details saved for the exercises. First, we need to show that the continuous image of a compact set is compact. Then $f(X)$ will be compact in $\mathbf{R}$ and hence must be closed and bounded. Thus there will be a least upper bound in $f(X)$, whose inverse image will contain the desired point p. A similar argument can be used to show that any continuous function $f(x)$ on a compact set must also have a minimum.

4.3 Metric Spaces

The natural notion of distance on the set $\mathbf{R}^n$ is the key to the existence of the standard topology. Luckily on many other sets similar notions of distance (called metrics) exist; any set that has a metric automatically has a topology.

Definition 4.3.1 *A* metric *on a set X is a function*

$$\rho : X \times X \to \mathbf{R}$$

such that for all points $x, y, z \in X$ we have:

1. $\rho(x,y) \geq 0$ *and* $\rho(x,y) = 0$ *if and only if* $x = y$.
2. $\rho(x,y) = \rho(y,x)$.
3. *(Triangle Inequality)*
$$\rho(x,z) \leq \rho(x,y) + \rho(y,z).$$

The set X with its metric ρ is called a metric space *and is denoted by (X, ρ).*

Fix a metric space (X, ρ).

Definition 4.3.2 *A set U in X is* open *if for all points $a \in U$, there is some real number $\epsilon > 0$ such that*

$$\{x : |x - a| < \epsilon\}$$

is contained in U.

Proposition 4.3.1 *The above definition for open set will define a Hausdorff topological space on the metric space (X, ρ).*

The proof is similar to the corresponding proof for the standard topology on $\mathbf{R}^n$. In fact, most of the topological facts about $\mathbf{R}^n$ can be quite easily translated into corresponding topological facts about any metric space. Unfortunately, as will be seen in section five, not all natural topological spaces come from a metric.

An example of a metric that is not just the standard one on $\mathbf{R}^n$ is given in Chapter Thirteen, when a metric and its associated topology is used to define Hilbert spaces.

4.4 Bases for Topologies

Warning: This section uses the notion of countability. A set is countable if there is a one-to-one onto mapping from the set to the natural numbers. More on this is in Chapter Ten. Note that the rational numbers are countable while the real numbers are uncountable.

In linear algebra, the word *basis* means a list of vectors in a vector space that generates uniquely the entire vector space. In a topology, a basis will be a collection of open sets that generate the entire topology. More precisely:

Definition 4.4.1 *Let X be a topological space. A collection of open sets forms a* basis *for the topology if every open set in X is the (possibly infinite) union of sets from the collection.*

For example, let (X, ρ) be a metric space. For each positive integer k and for each point $p \in X$, set

$$U(p, k) = \{x \in X : \rho(x, p) < \frac{1}{k}\}.$$

We can show that the collection of all possible $U(p, k)$ forms a basis for the topology of the metric space.

In practice, having a basis will allow us to reduce many topological calculations to calculating on sets in the basis. This will be more tractable if we can somehow limit the number of elements in a basis. This leads to

Definition 4.4.2 *A topological space is* second countable *if it has a basis with a countable number of elements.*

For example, $\mathbf{R}^n$, with the usual topology, is second countable. A countable basis can be constructed as follows. For each positive integer k and each $p \in \mathbf{Q}^n$ (which means that each coordinate of the point p is a rational number), define

$$U(p, k) = \{x \in \mathbf{R}^n : |x - p| < \frac{1}{k}\}.$$

There are a countable number of such sets $U(p, k)$ and they can be shown to form a basis.

Most reasonable topological spaces are second countable. Here is an example of a metric space that is not second countable. It should and does have the feel of being a pathology. Let X be any uncountable set (you can, for example, let X be the real numbers). Define a metric on X by setting $\rho(x, y) = 1$ if $x \neq y$ and $\rho(x, x) = 0$. It can be shown that this ρ defines a metric on X and thus defines a topology on X. This topology is weird, though. Each point x is itself an open set, since the open set $\{y \in X : \rho(x, y) < 1/2\} = x$. By using the fact that there are an uncountable number of points in X, we can show that this metric space is not second countable.

Of course, if we use the term "second countable", there must be a meaning to "first countable". A topological set is *first countable* if every point $x \in X$ has a countable neighborhood basis. For this to make sense, we need to know what a neighborhood basis is. A collection of open sets in X forms a *neighborhood basis* of some $x \in X$ if every open set containing x has in it an open set from the collection and if each open set in the collection contains the point x. We are just mentioning this definition for the sake of completeness. While we will later need the notion of second countable, we will not need in this book the idea of first countable.

4.5 Zariski Topology of Commutative Rings

Warning: This section requires a basic knowledge of commutative ring theory.

Though historically topology arose in the study of continuous functions on $\mathbf{R}^n$, a major reason why all mathematicians can speak the language of open, closed and compact sets is because there exists natural topologies on many diverse mathematical structures. This section looks at just one of these topologies. While this example (the Zariski topology for commutative rings) is important in algebraic geometry and algebraic number theory, there is no reason for the average mathematician to know it. It is given here simply to show how basic topological notions can be applied in a nonobvious way to an object besides $\mathbf{R}^n$. We will in fact see that the Zariski topology on the ring of polynomials is not Hausdorff and hence cannot come from a metric.

We want to associate a topological space to any commutative ring R. Our topological space will be defined on the set of all prime ideals in the ring R, a set that will be denoted by $Spec(R)$. Instead of first defining the open sets, we will start with what will be the closed sets. Let $\mathcal{P}$ be a prime ideal in R and hence a point in Spec R. Define *closed* sets to be

$$V_{\mathcal{P}} = \{\mathcal{Q} : \mathcal{Q} \text{ is a prime ideal in } \mathcal{R} \text{ containing } \mathcal{P}\}.$$

Then define Spec $R - V_{\mathcal{P}}$, where $\mathcal{P}$ is any prime ideal, to be an open set. The *Zariski topology* on Spec R is given by defining open sets to be the unions and finite intersections of all sets of the form Spec $R - V_{\mathcal{P}}$.

As will be seen in some of the examples, it is natural to call the points in Spec R corresponding to maximal ideals *geometric points.*

Assume that the ring R has no zero divisors, meaning that if $x \cdot y = 0$, then either x or y must be zero. Then the element 0 will generate a prime ideal, (0), contained in every other ideal. This ideal is called the *generic ideal* and is always a bit exceptional.

Now for some examples. For the first, let the ring R be the integers $\mathbf{Z}$. The only prime ideals in $\mathbf{Z}$ are of the form

$$(p) = \{kp : k \in \mathbf{Z}, p \text{ a prime number}\}$$

and the zero ideal (0). Then Spec $\mathbf{Z}$ is the set of all prime numbers:

2
3
5
7
11
13
17
19
23
29

and the zero ideal (0). The open sets in this topology are the complements of a finite number of these ideals.

For our second example, let the ring R be the field of complex numbers $\mathbf{C}$. The only two prime ideals are the zero ideal (0) and the whole field itself. Thus in some sense the space $\mathbf{C}$ is a single point.

A more interesting example occurs by setting $R = \mathbf{C}[x]$, the ring of one-variable polynomials with complex coefficients. We will see that as a point set this space can be identified with the real plane $\mathbf{R}^2$ (if we do not consider the generic ideal) but that the topology is far from the standard topology of $\mathbf{R}^2$. Key is that all one-variable polynomials can be factored into linear factors, by the Fundamental Theorem of Algebra; thus all prime ideals are multiples of linear polynomials. We denote the ideal of all of the multiples of a linear polynomial $x - c$ as:

$$(x-c) = \{f(x)(x-c) : f(x) \in \mathbf{C}[x], c \in \mathbf{C}\}.$$

Hence, to each complex number, $c = a + bi$ with $a, b \in \mathbf{R}$, there corresponds a prime ideal $(x - c)$ and thus Spec $\mathbf{C}[x]$ is another, more ring-theoretic description of the complex numbers. Geometrically, Spec $\mathbf{C}[x]$ is

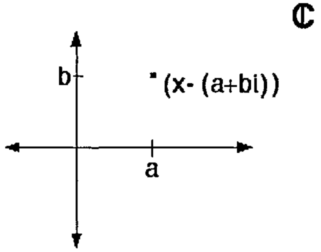

Note that while the zero ideal (0) is still a prime ideal in $\mathbf{C}[x]$, it does not correspond to any point in $\mathbf{C}$; instead, it is lurking in the background. The open sets in this topology are the complements of a finite number of the prime ideals. But each prime ideal corresponds to a complex number. Since the complex numbers $\mathbf{C}$ can be viewed as the real plane $\mathbf{R}^2$, we have that an open set is the complement of a finite number of points in the real plane. While these open sets are also open in the standard topology on $\mathbf{R}^2$, they are far larger than any open disc in the plane. No little ϵ-disc will be the complement of only a finite number of points and hence cannot be open in the Zariski topology. In fact, notice that the intersection of two of these Zariski open sets must intersect. This topology cannot be Hausdorff. Since all metric spaces are Hausdorff, this means that the Zariski topology cannot come from some metric.

Now let $R = \mathbf{C}[x, y]$ be the ring of two-variable polynomials with complex coefficients. Besides the zero ideal (0), there are two types of prime ideals: the maximal ideals, each of which is generated by polynomials of the form $x - c$ and $y - d$, where c and d are any two complex numbers and nonmaximal prime ideals, each of which is generated by an irreducible polynomial $f(x,y)$.

Note that the maximal ideals correspond to points in the complex plane $\mathbf{C} \times \mathbf{C}$, thus justifying the term 'geometric point'.

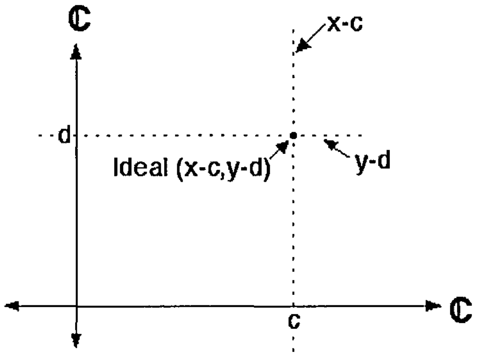

Since each copy of the complex numbers $\mathbf{C}$ is a real plane $\mathbf{R}^2$, $\mathbf{C} \times \mathbf{C}$ is $\mathbf{R}^2 \times \mathbf{R}^2 = \mathbf{R}^4$. In the Zariski topology, open sets are the complements of the zero loci of polynomials. For example, if $f(x,y)$ is an irreducible polynomial, then the set

$$U = \{(x,y) \in \mathbf{C}^2 : f(x,y) \neq 0\}$$

is open. While Zariski sets will still be open in the standard topology on $\mathbf{R}^4$, the converse is most spectacularly false. Similar to the Zariski topology on $\mathbf{C}[x]$, no ϵ-ball will be open in the Zariski topology on $\mathbf{C}[x,y]$. In fact, if U and V are two Zariski open sets that are non-empty, they must intersect. Thus this is also a non-Hausdorff space and hence cannot come from a metric space.

4.6 Books

Point set topology's days of glory were the early twentieth century, a time when some of the world's best mathematicians were concerned with the correct definitions for continuity, dimension and for a topological space. Most of these issues have long been settled. Today, point set topology is overwhelmingly a tool that all mathematicians need to know.

At the undergraduate level, it is not uncommon for a math department to use their point set topology class as a place to introduce students to proofs. Under the influence of E. H. Moore (of the University of Chicago) and of his student R.L. Moore (of the University of Texas, who advised an amazing number of Ph.D. students), many schools have taught topology under the Moore method. Using this approach, on the first day of class students are given a list of the definitions and theorems. On the second day people are asked who has proven Theorem One. If someone thinks they have a proof, they go to the board to present it to the class. Those who still want to think of a proof on their own leave the class for that part of the lecture. This is a powerful way to introduce students to proofs. On the other hand, not much material can be covered. At present, most people who teach using the Moore method modify it in various ways.

Of course, this approach comes close to being absurd for people who are already mathematically mature and just need to be able to use the results. The texts of the fifties and sixties were by Kelley [72] and Dugundji [30]. Overwhelmingly the most popular current book is Munkres' *Topology: A First Course* [88].

My own bias (a bias not shared by most) is that all the point set topology that most people need can be found in, for example, the chapter in Royden's *Real Analysis* [95] on topology.

4.7 Exercises

1. The goal of this problem is to show that a topology on a set X can also be defined in terms of a collection of closed sets, as opposed to a collection of open sets. Let X be a set of points and let $C = \{C_\alpha\}$ be a collection of subsets of X. Suppose that

- Any finite union of sets in the collection C must be another set in C.
- Any intersection of sets in C must be another set in C.
- The empty set ϕ and the whole space X must in the collection C.

Call the sets in C *closed* and call a set U *open* if its complement $X - U$ is closed. Show that this definition of open set will define a topology on the set X.

2. Prove Proposition 4.2.1.

3. Prove Theorem 4.2.2.

4. Prove Theorem 4.2.3.

5. Let V be the vector space of all functions

$$f : [0,1] \to \mathbf{R}$$

whose derivatives, including the one-sided derivatives at the endpoints, are continuous functions on the interval $[0,1]$. Define

$$|f|_\infty = \sup_{x\in[0,1]} |f(x)|$$

for any function $f \in V$. For each $f \in V$ and each $\epsilon > 0$, define

$$U_f(\epsilon) = \{g \in V : |f - g|_\infty < \epsilon\}.$$

a. Show that the set of all $U_f(\epsilon)$ is a basis for a topology on the set V.
b. Show that there can be no number M such that for all $f \in V$,

$$|\frac{\mathrm{d}f}{\mathrm{d}x}|_\infty < M|f|_\infty.$$

In the language of functional analysis, this means that the derivative, viewed as a linear map, is not *bounded* on the space V. One of the main places where serious issues involving point set topology occur is in functional analysis, which is the study of vector spaces of various types of functions. The study of such space is important in trying to solve differential equations.

Chapter 5

Classical Stokes' Theorems

Basic Objects:	Manifolds and boundaries
Basic Maps:	Vector-valued functions on manifolds
Basic Goal:	Function's average over a boundary = Derivative's average over interior

Stokes' Theorem, in all of its many manifestations, comes down to equating the average of a function on the boundary of some geometric object with the average of its derivative (in a suitable sense) on the interior of the object. Of course, a correct statement about averages must be put into the language of integrals. This theorem provides a deep link between topology (the part about boundaries) and analysis (integrals and derivatives). It is also critical for much of physics, as can be seen in both its historical development and in the fact that for most people their first introduction to Stokes' Theorem is in a course on electricity and magnetism.

The goal of Chapter Six is to prove Stokes' Theorem for abstract manifolds (which are, in some sense, the abstract method for dealing with geometric objects). As will be seen, to even state this theorem takes serious work in building up the necessary machinery. This chapter looks at some special cases of Stokes' Theorem, special cases that were known long before people realized that there is this one general underlying theorem. For example, we will see that the Fundamental Theorem of Calculus is a special case of Stokes' Theorem (though to prove Stokes' Theorem, you use the Fundamental Theorem of Calculus; thus logically Stokes' Theorem does not imply the Fundamental Theorem of Calculus). It was in the 1800s that most of these special cases of Stokes' Theorem were discovered, though, again, people did not know that each of these were special cases of one general result. These special cases are important and useful enough that they are now standard topics in most multivariable calculus courses and introductory classes in electricity and magnetism. They are Green's Theorem, the Divergence Theorem and Stokes' Theorem. (This Stokes' theorem is, though, a special case of the Stokes' Theorem of the next chapter.) This chapter develops the needed mathematics for these special cases. We will state and sketch proofs for the Divergence Theorem and Stokes' Theorem. Physical intuitions will be stressed.

There is a great deal of overlap between the next chapter and this one. Mathematicians need to know both the concrete special cases of Stokes' Theorem and the abstract version of Chapter Six.

5.1 Preliminaries about Vector Calculus

This is a long section setting up the basic definitions of vector calculus. We need to define vector fields, manifolds, path and surface integrals, divergence and curl. All of these notions are essential. Only then can we state the Divergence Theorem and Stokes' Theorem, which are the goals of this chapter.

5.1.1 Vector Fields

Definition 5.1.1 *A* vector field *on* $\mathbf{R}^n$ *is a vector-valued function*

$$\mathbf{F} : \mathbf{R}^n \to \mathbf{R}^m.$$

If $x_1, \ldots, x_n$ *are coordinates for* $\mathbf{R}^n$, *then the vector field* $\mathbf{F}$ *will be described by* m *real-valued functions* $f_k : \mathbf{R}^n \to \mathbf{R}$ *as follows:*

$$\mathbf{F}(x_1, \ldots, x_n) = \begin{pmatrix} f_1(x_1, \ldots, x_n) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{pmatrix}.$$

A vector field is *continuous* if each real-valued function f_k is continuous, *differentiable* if each real-valued f_k is differentiable, etc.

Intuitively, a vector field assigns to each point of $\mathbf{R}^n$ a vector. Any number of physical phenomenon can be captured in terms of vector fields. In fact, they are the natural language of fluid flow, electric fields, magnetic fields, gravitational fields, heat flow, traffic flow and much more.

For example, let $\mathbf{F} : \mathbf{R}^2 \to \mathbf{R}^2$ be given by

$$\mathbf{F}(x, y) = (3, 1).$$

Here $f_1(x, y) = 3$ and $f_2(x, y) = 1$. On $\mathbf{R}^2$ this vector field can be pictured by drawing in a few sample vectors.

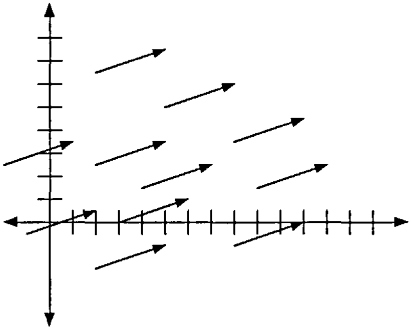

A physical example of this vector field would be wind blowing in the direction $(3, 1)$ with velocity

$$\text{length}(3,1) = \sqrt{9+1} = \sqrt{10}.$$

Now consider the vector field $\mathbf{F}(x, y) = (x, y)$. Then in pictures we have:

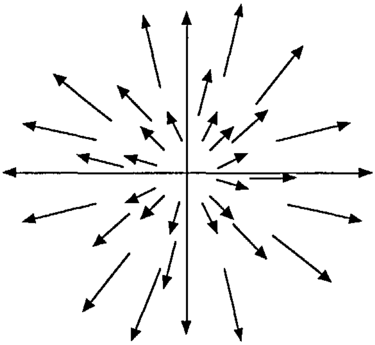

This could represent water flowing out from the origin $(0, 0)$.

For our final example, let $\mathbf{F}(x, y) = (-y, x)$. In pictures we have:

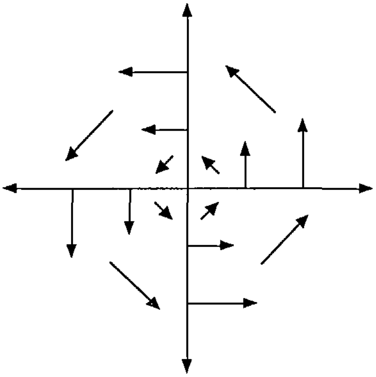

which might be some type of whirlpool.

5.1.2 Manifolds and Boundaries

Curves and surfaces appear all about us. Both are examples of manifolds, which are basically just certain naturally occurring geometric objects. The intuitive idea of a manifold is that, for a k-dimensional manifold, each point is in a neighborhood that looks like a ball in $\mathbf{R}^k$. In the next chapter we give three different ways for defining a manifold. In this chapter, we will define manifolds via *parametrizations.* The following definition is making rigorous the idea that locally, near any point, a k-dimensional manifold looks like a ball in $\mathbf{R}^k$.

Definition 5.1.2 *A* differentiable manifold M *of dimension* k *in* $\mathbf{R}^n$ *is a set of points in* $\mathbf{R}^n$ *such that for any point* $p \in M$*, there is a small open neighborhood* U *of* p*, a vector-valued differentiable function* $F : \mathbf{R}^k \to \mathbf{R}^n$ *and an open set* V *in* $\mathbf{R}^k$ *with*

a) $F(V) = U \cap M$

b) The Jacobian of F *has rank* k *at every point in* V*, where the* Jacobian *of* F *is the* $n \times k$ *matrix*

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_k} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_k} \end{pmatrix},$$

with $x_1, \ldots, x_k$ *a coordinate system for* $\mathbf{R}^k$*. The function* F *is called the (local)* parametrization *of the manifold.*

Recall that the rank of a matrix is k if the matrix has an invertible $k \times k$ minor. (A minor is a submatrix of a matrix.)

A circle is a one-dimensional manifold, with a parametrization

$$F : \mathbf{R}^1 \to \mathbf{R}^2$$

given by

$$F(t) = (\cos(t), \sin(t)).$$

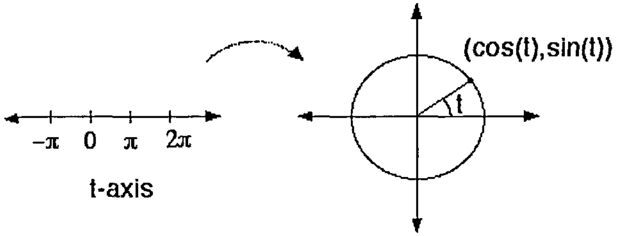

Geometrically the parameter t is the angle with the x-axis. Note that the Jacobian of F is $\binom{-\sin t}{\cos t}$. Since *sine* and *cosine* cannot simultaneously be zero, the Jacobian has rank 1.

A cone in three-space can be parametrized by

$$F(u,v) = (u, v, \sqrt{u^2 + v^2}).$$

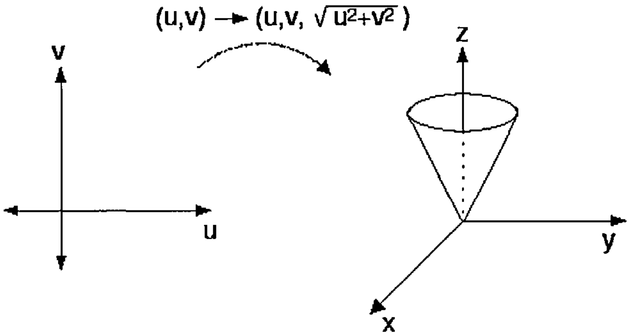

This will be a two dimensional manifold (a surface) except at the vertex $(0,0,0)$, for at this point the Jacobian fails to be well-defined, much less having rank two. Note that this agrees with the picture, where certainly the origin looks quite different than the other points.

Again, other definitions are given in Chapter Six.

Now to discuss what is the boundary of a manifold. This is needed since Stokes' Theorem and its many manifestations state that the average of a function on the boundary of a manifold will equal the average of its derivative on the interior.

Let M be a k-dimensional manifold in $\mathbf{R}^n$.

Definition 5.1.3 *The* closure *of M, denoted $\bar{M}$, is the set of all points x in $\mathbf{R}^n$ such that there is a sequence of points (x_n) in the manifold M with*

$$\lim_{n\to\infty} x_n = x.$$

The boundary *of M, denoted ∂M, is:*

$$\partial M = \bar{M} - M,$$

Given a manifold with boundary, we call the nonboundary part the *interior.*

All of this will become relatively straightforward with a few examples. Consider the map

$$r : [-1, 2] \to \mathbf{R}^2$$

where

$$r(t) = (t, t^2).$$

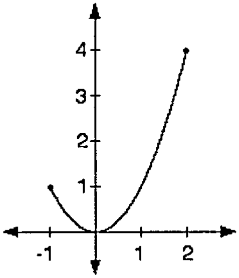

The image under r of the open interval $(-1, 2)$ is a one-manifold (since the Jacobian is the 2×1 matrix $(1, 2t)$, which always has rank one). The boundary consists of the two points $r(-1) = (-1, 1)$ and $r(2) = (2, 4)$.

Our next example is a two-manifold having a boundary consisting of a circle. Let

$$r : \{(x, y) \in \mathbf{R}^2 : x^2 + y^2 \leq 1\} \to \mathbf{R}^3$$

be defined by

$$r(x, y) = (x, y, x^2 + y^2).$$

The image of r is a bowl in space sitting over the unit disc in the plane:

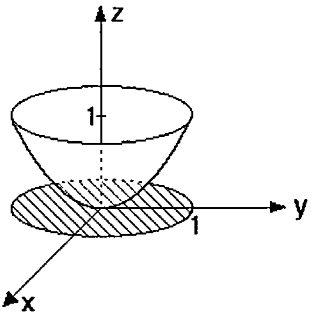

Now the image under r of the open disc $\{(x, y) \in \mathbf{R}^2 : x^2 + y^2 < 1\}$ is a two-manifold (since the Jacobian is

$$\begin{pmatrix} 1 & 0 & 2x \\ 0 & 1 & 2y \end{pmatrix},$$

which has rank two at all points). The boundary is the image of the boundary of the disc and hence the image of the circle $\{(x, y) \in \mathbf{R}^2 : x^2+y^2 = 1\}$. In this case, as can be seen by the picture, the boundary is itself a circle living on the plane $z = 1$ in space.

Another example is the unit circle in the plane. We saw that this is a one-manifold. There are no boundary points, though. On the other hand, the unit circle is itself the boundary of a two-manifold, namely the unit disc in the plane. In a similar fashion, the unit sphere in $\mathbf{R}^3$ is a two-manifold, with no boundary, that is itself the boundary of the unit ball, a three-manifold. (It is not chance that in these two cases that the boundary of the boundary is the empty set.)

We will frequently call a manifold with boundary simply a manifold. We will also usually be making the assumption that the boundary of an n-dimensional manifold will either be empty (in which case the manifold has no boundary) or is itself an $(n - 1)$-dimensional manifold.

5.1.3 Path Integrals

Now that we have a sharp definition for manifolds, we want to do calculus on them. We start with integrating vector fields along curves. This process is called a path integral or sometimes, misleadingly, a line integral.

A *curve* or *path* C in $\mathbf{R}^n$ is defined to be a one-manifold with boundary. Thus all curves are defined by maps $F : [a, b] \to \mathbf{R}^n$, given by

$$F(t) = \begin{pmatrix} f_1(t) \\ \vdots \\ f_n(t) \end{pmatrix}.$$

These maps are frequently written as

$$\begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix}.$$

We will require each component function $f_i : \mathbf{R} \to \mathbf{R}$ to be differentiable.

Definition 5.1.4 *Let $f(x_1, \ldots, x_n)$ be a real-valued function defined on $\mathbf{R}^n$. The* path integral *of the function f along the curve C is*

$$\begin{aligned} \int_c f \,\mathrm{d}s &= \int_c f(x_1, \ldots, x_n) \mathrm{d}s \\ &= \int_a^b f(x_1(t), \ldots, x_n(t)) \left(\sqrt{(\frac{\mathrm{d}x_1}{\mathrm{d}t})^2 + \cdots + (\frac{\mathrm{d}x_n}{\mathrm{d}t})^2} \right) \mathrm{d}t. \end{aligned}$$

Note that

$$\int_a^b f(x_1(t), \ldots, x_n(t)) \left(\sqrt{(\frac{\mathrm{d}x_1}{\mathrm{d}t})^2 + \cdots + (\frac{\mathrm{d}x_n}{\mathrm{d}t})^2} \right) \mathrm{d}t,$$

while looking quite messy, is an integral of the single variable t.

Theorem 5.1.1 *Let a curve C in $\mathbf{R}^n$ be described by two different parametrizations*

$$F : [a, b] \to \mathbf{R}^n$$

and

$$G : [c, d] \to \mathbf{R}^n,$$

with $F(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix}$ *and* $G(u) = \begin{pmatrix} y_1(u) \\ \vdots \\ y_n(u) \end{pmatrix}$.

The path integral $\int_C f \,\mathrm{d}s$ is independent of parametrization chosen, i.e.,

$$\int_a^b f(x_1(t), \ldots, x_n(t)) \sqrt{(\frac{\mathrm{d}x_1}{\mathrm{d}t})^2 + \cdots + (\frac{\mathrm{d}x_n}{\mathrm{d}t})^2} \,\mathrm{d}t$$

$$=$$

$$\int_c^d f(y_1(u), \ldots, y_n(u)) \sqrt{(\frac{\mathrm{d}y_1}{\mathrm{d}u})^2 + \cdots + (\frac{\mathrm{d}y_n}{\mathrm{d}u})^2} \,\mathrm{d}u.$$

While we will do an example in a moment, the proof uses critically and is an exercise in the chain rule. In fact, the path integral was defined with the awkward term

$$\mathrm{d}s = \sqrt{(\frac{\mathrm{d}x_1}{\mathrm{d}t})^2 + \cdots + (\frac{\mathrm{d}x_n}{\mathrm{d}t})^2}\,\mathrm{d}t$$

precisely in order to make the path integral independent of parametrization. This is why $\int_a^b f(x_1(t), \ldots, x_n(t))\,\mathrm{d}t$ is an incorrect definition for the path integral.

The symbol "ds" represents the infinitesimal arc length on the curve C in $\mathbf{R}^n$. In pictures, for $\mathbf{R}^2$, consider the following.

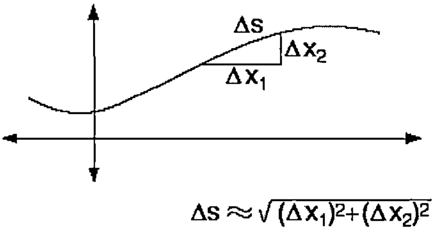

With Δs denoting the change in position along the curve C, we have by the Pythagorean Theorem

$$\begin{aligned} \Delta s &\approx \sqrt{(\Delta x_1)^2 + (\Delta x_2)^2} \\ &= \left(\sqrt{(\frac{\Delta x_1}{\Delta t})^2 + (\frac{\Delta x_2}{\Delta t})^2}\right)\Delta t. \end{aligned}$$

Then in the limit as $\Delta t \to 0$, we have, at least formally,

$$\mathrm{d}s = \left(\sqrt{(\frac{\mathrm{d}x_1}{\mathrm{d}t})^2 + (\frac{\mathrm{d}x_2}{\mathrm{d}t})^2}\right)\mathrm{d}t.$$

Thus the correct implementation of the Pythagorean Theorem will also force on us the term $\mathrm{d}s = \sqrt{(\frac{\mathrm{d}x_1}{\mathrm{d}t})^2 + \cdots + (\frac{\mathrm{d}x_n}{\mathrm{d}t})^2}\,\mathrm{d}t$ in the definition of the path integral.

Now for an example, in order to check our working knowledge of the definitions and also to see how the ds term is needed to make path integrals independent of parametrizations. Consider the straight line segment in the plane from $(0,0)$ to $(1,2)$. We will parametrize this line segment in two different ways, and then compute the path integral of the function

$$f(x,y) = x^2 + 3y$$

using each of the parametrizations.

First, define

$$F : [0,1] \to \mathbf{R}^2$$

by

$$F(t) = (t, 2t).$$

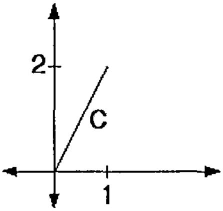

Thus we have $x(t) = t$ and $y(t) = 2t$. Denote this line segment by C. Then

$$\begin{aligned}
\int_C f(x,y)\mathrm{d}s &= \int_0^1 (x(t)^2 + 3y(t))\sqrt{(\frac{\mathrm{d}x}{\mathrm{d}t})^2 + (\frac{\mathrm{d}y}{\mathrm{d}t})^2}\,\mathrm{d}t \\
&= \int_0^1 (t^2 + 6t)\sqrt{5}\,\mathrm{d}t \\
&= \sqrt{5}(\frac{t^3}{3}\,|_0^1 + 3t^2\,|_0^1) \\
&= \sqrt{5}(\frac{1}{3} + 3) \\
&= \frac{10}{3}\sqrt{5}.
\end{aligned}$$

Now parametrize the segment C by:

$$G : [0,2] \to C$$

where

$$G(t) = (\frac{t}{2}, t).$$

Here we have $x(t) = \frac{t}{2}$ and $y(t) = t$. Then

$$\begin{aligned}
\int_C f(x,y)\mathrm{d}s &= \int_0^2 (x(t)^2 + 3y(t))\sqrt{(\frac{\mathrm{d}x}{\mathrm{d}t})^2 + (\frac{\mathrm{d}y}{\mathrm{d}t})^2}\,\mathrm{d}t \\
&= \int_0^2 (\frac{t^2}{4} + 3t)\sqrt{\frac{1}{4} + 1}\,\mathrm{d}t \\
&= \frac{\sqrt{5}}{2}(\frac{t^3}{12}\,|_0^2 + \frac{3t^2}{2}\,|_0^2) \\
&= \frac{\sqrt{5}}{2}(\frac{8}{12} + 6) \\
&= \frac{10}{3}\sqrt{5},
\end{aligned}$$

as desired.

5.1.4 Surface Integrals

Now to integrate along surfaces. A *surface* in $\mathbf{R}^3$ is a two-manifold with boundary. For the sake of simplicity, we will restrict our attention to those surfaces which are the image of a map

$$r : D \to \mathbf{R}^3,$$

given by

$$r(u,v) = (x(u,v), y(u,v), z(u,v)),$$

where x, y, z are coordinates for $\mathbf{R}^3$ and u, v are coordinates for $\mathbf{R}^2$. Here D is a *domain* in the plane, which means that there is an open set U in $\mathbf{R}^2$ whose closure is D. (If you think of U as an open disc and D as a closed disc, you usually will not go wrong.)

Definition 5.1.5 *Let $f(x,y,z)$ be a function on $\mathbf{R}^3$. Then the integral of $f(x,y,z)$ along the surface S is*

$$\int\int_S f(x,y,z)\mathrm{d}S = \int\int_D f(x(u,v), y(u,v), z(u,v)) \cdot \left|\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}\right| \mathrm{d}u\mathrm{d}v.$$

Here $|\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}|$ denotes the length of the cross product (which in a moment we will show to be the length of a certain normal vector) of the vectors $\frac{\partial r}{\partial u}$ and $\frac{\partial r}{\partial v}$,and is hence the determinant of

$$\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} = \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \partial x/\partial u & \partial y/\partial u & \partial z/\partial u \\ \partial x/\partial v & \partial y/\partial v & \partial z/\partial v \end{pmatrix}.$$

Thus the *infinitesimal area* dS is:

$$\text{length of } (\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}) \mathrm{d}u\mathrm{d}v =$$

$$\left| \left((\frac{\partial y}{\partial u}\frac{\partial z}{\partial v}) - (\frac{\partial z}{\partial u}\frac{\partial y}{\partial v}), (\frac{\partial x}{\partial v}\frac{\partial z}{\partial u}) - (\frac{\partial x}{\partial u}\frac{\partial z}{\partial v}), (\frac{\partial x}{\partial u}\frac{\partial y}{\partial v}) - (\frac{\partial x}{\partial v}\frac{\partial y}{\partial u}) \right) \right| \mathrm{d}u\mathrm{d}v.$$

In analogy with arc length, a surface integral is independent of parametrization:

Theorem 5.1.2 *The integral* $\int\int_S f(x,y,z)\ \mathrm{d}S$ *is independent of the parametri tion of the surface* S.

Again, the chain rule is a critical part of the proof.

Note that if this theorem were not true, we would define the surface integral (in particular the infinitesimal area) differently.

We now show how the vector field

$$\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}$$

is actually a normal to the surface. With the map $r : \mathbf{R}^2 \to \mathbf{R}^3$ given by $r(u,v) = (x(u,v), y(u,v), z(u,v))$, recall that the Jacobian of r is

$$\begin{pmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \\ \partial z/\partial u & \partial z/\partial v \end{pmatrix}.$$

But as we saw in Chapter Three, the Jacobian maps tangent vectors to tangent vectors. Thus the two vectors

$$(\frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}, \frac{\partial z}{\partial u})$$

and

$$(\frac{\partial x}{\partial v}, \frac{\partial y}{\partial v}, \frac{\partial z}{\partial v})$$

are both tangent vectors to the surface S. Hence their cross product must be a normal (perpendicular) vector $\mathbf{n}$. Thus we can interpret the surface integral as

$$\int\int_S f\ \mathrm{d}S = \int\int f \cdot |\mathbf{n}|\ \mathrm{d}u\mathrm{d}v$$

with dS =(length of the normal vector $\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}$) d$udv$.

5.1.5 The Gradient

The gradient of a function can be viewed as a method for differentiating functions.

Definition 5.1.6 *The* gradient *of a real-valued function* $f(x_1, \ldots, x_n)$ *is*

$$\nabla f = (\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}).$$

Thus

$$\nabla : (\text{Functions}) \rightarrow (\text{Vector fields}).$$

For example, if $f(x, y, z) = x^3 + 2xy + 3xz$, we have

$$\nabla(f) = (3x^2 + 2y + 3z, 2x, 3x).$$

It can be shown that if at all points on $M = (f(x_1, \ldots, x_n) = 0)$ where $\nabla f \neq 0$, the gradient ∇f is a normal vector to M.

5.1.6 The Divergence

The divergence of a vector field can be viewed as a reasonable way to differentiate a vector field. (In the next section we will see that the curl of a vector field is another way.) Let $\mathbf{F}(x, y, z) : \mathbf{R}^3 \rightarrow \mathbf{R}^3$ be a vector field given by three functions as follows:

$$\mathbf{F}(x, y, z) = (f_1(x, y, z), f_2(x, y, z), f_3(x, y, z)).$$

Definition 5.1.7 *The* divergence *of* $\mathbf{F}(x, y, z)$ *is*

$$div(\mathbf{F}) = \frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial y} + \frac{\partial f_3}{\partial z}.$$

Thus

$$\text{div} : (\text{Vector fields}) \rightarrow (\text{Functions}).$$

The Divergence Theorem will tell us that the divergence measures how much the vector field is spreading out at a point.

For example, let $\mathbf{F}(x, y, z) = (x, y^2, 0)$. Then

$$\text{div}(\mathbf{F}) = \frac{\partial x}{\partial x} + \frac{\partial (y^2)}{\partial y} + \frac{\partial (0)}{\partial z} = 1 + 2y.$$

If you sketch out this vector field, you do indeed see that the larger the y value, the more spread out the vector field becomes.

5.1.7 The Curl

The curl of a vector field is another way in which we can extend the idea of differentiation to vector fields. Stokes' Theorem will show us that the curl of a vector field measures how much the vector field is twirling or whirling or curling about. The actual definition is:

Definition 5.1.8 *The* curl *of a vector field* $\mathbf{F}(x, y, z)$ *is*

$$\begin{aligned} curl(\mathbf{F}) &= \det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ f_1 & f_2 & f_3 \end{pmatrix} \\ &= (\frac{\partial f_3}{\partial y} - \frac{\partial f_2}{\partial z}, -(\frac{\partial f_3}{\partial x} - \frac{\partial f_1}{\partial z}), \frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y}). \end{aligned}$$

Note that

$$\text{curl} : (\text{Vector fields}) \rightarrow (\text{Vector fields}).$$

Now to look at an example and see that the curl is indeed measuring some sort of twirling. Earlier we saw that the vector field $\mathbf{F}(x, y, z) = (-y, x, 0)$ looks like a whirlpool. Its curl is:

$$\text{curl}(\mathbf{F}) = \det \begin{pmatrix} i & j & k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ -y & x & 0 \end{pmatrix}.$$

$$= (0, 0, 2),$$

which reflects that the whirlpool action is in the xy-plane, perpendicular to the z-axis.

We will see in the statement of Stokes' Theorem that intuitively the length of the curl($\mathbf{F}$) indeed measures how much the vector field is twirling about while the vector curl($\mathbf{F}$) points in the direction normal to the twirling.

5.1.8 Orientability

We also require our manifolds to be orientable. For a surface, orientability means that we can choose a normal vector field on the surface that varies continuously and never vanishes. For a curve, orientability means that we can choose a unit tangent vector, at each point, that varies continuously.

The standard example of a nonorientable surface is the Möbius strip, obtained by putting a half twist in a strip of paper and then attaching the ends.

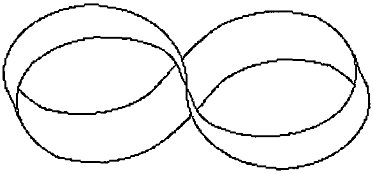

For an orientable manifold, there are always two choices of orientation, depending on which direction is chosen for the normal or the tangent. Further an oriented surface S with boundary curve ∂S will induce an orientation on ∂S, as will a 3-dimensional region induce an orientation on its boundary surface. If you happen to choose the wrong induced orientation for a boundary, the various versions of Stokes' Theorems will be off merely by a factor of (-1). Do not panic if you found the last few paragraphs vague. They were, deliberately so. To actually rigorously define orientation takes a little work. In first approaching the subject, it is best to concentrate on the basic examples and only then worry about the correct sign coming from the induced orientations. Rigorous definitions for orientability are given in the next chapter.

5.2 The Divergence Theorem and Stokes' Theorem

(For technical convenience, we will assume for the rest of this chapter that all functions, including those that make up vector fields, have as many derivatives as needed.)

The whole goal of this chapter is to emphasize that there must always be a deep link between the values of a function on the boundary of a manifold with the values of its derivative (suitably defined) on the interior of the manifold. This link is already present in

Theorem 5.2.1 (The Fundamental Theorem of Calculus) *Let*

$$f : [a, b] \to \mathbf{R}$$

be a a real-valued differentiable function on the interval $[a, b]$*. Then*

$$f(b) - f(a) = \int_a^b \frac{df}{dx} \mathrm{d}x.$$

Here the derivative $\frac{df}{dx}$ is integrated over the interval

$$[a, b] = \{x \in \mathbf{R} : a \leq x \leq b\},$$

which has as its boundary the points (a) and (b). The orientation on the boundary will be b and $-a$, or

$$\partial[a,b] = b - a.$$

Then the Fundamental Theorem of Calculus can be interpreted as stating that the value of $f(x)$ on the boundary is equal to the average (the integral) of the derivative over the interior.

One possible approach to generalizing the Fundamental Theorem is to replace the one-dimensional interval $[a,b]$ with something higher dimensional and replace the one variable function f with either a function of more than one variable or (less obviously) by a vector field. The correct generalizations will of course be determined by what can be proven.

In the divergence theorem, the interval becomes a three-dimensional manifold, whose boundary is a surface, and the function f becomes a vector field. The derivative of f will here be the divergence. More precisely:

Theorem 5.2.2 (The Divergence Theorem) *In $\mathbf{R}^3$, let M be a three-dimensional manifold with boundary ∂M a compact manifold of dimension two. Let $\mathbf{F}(x,y,z)$ denote a vector field on $\mathbf{R}^3$ and let $\mathbf{n}(x,y,z)$ denote a unit normal vector field to the boundary surface ∂M. Then*

$$\int\int_{\partial M} \mathbf{F}\cdot\mathbf{n}\, dS = \int\int\int_{M} (div\mathbf{F})\, dxdydz.$$

We will sketch a proof in section 5.5.

On the left hand side we have an integral of the vector field $\mathbf{F}$ over the boundary. On the right hand side we have an integral of the function div($\mathbf{F}$) (which involves derivatives of the vector field) over the interior.

In Stokes' Theorem, the interval becomes a surface, so that the boundary is a curve, and the function again becomes a vector field. The role of the derivative though will now be played by the curl of the vector field.

Theorem 5.2.3 (Stokes' Theorem) *Let M be a surface in $\mathbf{R}^3$ with compact boundary curve ∂M. Let $\mathbf{n}(x,y,z)$ be the unit normal vector field to M and let $\mathbf{T}(x,y,z)$ denote the induced unit tangent vector to the curve ∂M. If $\mathbf{F}(x,y,z)$ is any vector field, then*

$$\int_{\partial M} \mathbf{F}\cdot\mathbf{T}\, ds = \int\int_{M} curl(\mathbf{F})\cdot\mathbf{n}\, dS.$$

As with the Divergence Theorem, a sketch of the proof will be given later in this chapter.

Again, on the left hand side we have an integral involving a vector field $\mathbf{F}$ on the boundary while on the right hand side we have an integral on the interior involving the curl of $\mathbf{F}$ (which is in terms of the various derivatives of $\mathbf{F}$).

Although both the Divergence Theorem and Stokes' Theorem were proven independently, their similarity is more than a mere analogy; both are special cases, as is the Fundamental Theorem of Calculus, of one very general theorem, which is the goal of the next chapter. The proofs of each are also quite similar. There are in fact two basic methods for proving these types of theorems. The first is to reduce to the Fundamental Theorem of Calculus, $f(b) - f(a) = \int_a^b \frac{df}{dx} dx$. This method will be illustrated in our sketch of the Divergence Theorem.

The second method involves two steps. Step one is to show that given two regions R_1 and R_2 that share a common boundary, we have

$$\int_{\partial R_1} \text{function} + \int_{\partial R_2} \text{function} = \int_{\partial(R_1 \cup R_2)} \text{function}.$$

Step two is to show that the theorem is true on infinitesimally small regions. To prove the actual theorem by this approach, simply divide the original region into infinitely many infinitesimally small regions, apply step two and then step one. We take this approach in our sketch of Stokes' Theorem.

Again, all of these theorems are really the same. In fact, to most mathematicians, these theorems usually go by the single name "Stokes' Theorem".

5.3 A Physical Interpretation of the Divergence Theorem

The goal of this section is to give a physical meaning to the Divergence Theorem, which was, in part, historically how the theorem was discovered. We will see that the Divergence Theorem states that the flux of a vector field through a surface is precisely equal to the sum of the divergences of each point of the interior. Of course, we need to give some definitions to these terms.

Definition 5.3.1 *Let S be a surface in $\mathbf{R}^3$ with unit normal vector field $\mathbf{n}(x, y, z)$. Then the flux of a vector field $\mathbf{F}(x, y, z)$ through the surface S is*

$$\int\int_S \mathbf{F} \cdot \mathbf{n} \, \mathrm{d}S.$$

Intuitively we want the flux to measure how much of the vector field $\mathbf{F}$ pushes through the surface S.

Imagine a stream of water flowing along. The tangent vector of the direction of the water at each point defines a vector field $\mathbf{F}(x, y, z)$. Suppose the vector field $\mathbf{F}$ is:

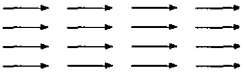

Place into the stream an infinitely thin sheet of rubber, let us say. We want the flux to measure how hard it is to hold this sheet in place against the flow of the water. Here are three possibilities:

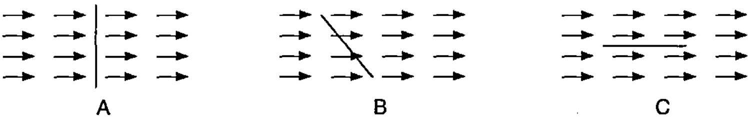

In case A, the water is hitting the rubber sheet head on, making it quite difficult to hold in place. In case C, no effort is needed to hold the sheet still, as the water just flows on by. The effort needed to keep the sheet still in case B is seen to be roughly halfway between effort needed in cases A and C. The key to somehow quantifying these differences of flux is to measure the angle between the vector field $\mathbf{F}$ of the stream and the normal vector field $\mathbf{n}$ to the membrane. Clearly, the dot product $\mathbf{F}\cdot\mathbf{n}$ works. Thus using that flux is defined by

$$\int\int_S \mathbf{F}\cdot\mathbf{n}\,\mathrm{d}S,$$

the flux through surface A is greater than the flux through surface B which in turn is greater than the flux through surface C, which has flux equal to 0.

The Divergence Theorem states that the flux of a vector field through a boundary surface is exactly equal to the sum (integral) of the divergence of the vector field in the interior. In some sense the divergence must be an infinitesimal measure of the flux of a vector field.

5.4 A Physical Interpretation of Stokes' Theorem

Here we discuss the notion of the circulation of a vector field with respect to a curve. We will give the definition, then discuss what it means.

Definition 5.4.1 *Let C be a smooth curve in $\mathbf{R}^3$ with unit tangent vector field $\mathbf{T}(x,y,z)$. The circulation of a vector field $\mathbf{F}(x,y,z)$ along the curve C is*

$$\int_C \mathbf{F}\cdot\mathbf{T}\,\mathrm{d}s.$$

Let **F** be a vector field representing a flowing stream of water, such as:

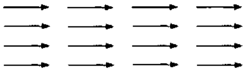

Put a thin wire (a curve C) into this stream with a small bead attached to it, with the bead free to move up and down the wire.

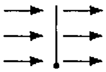

a

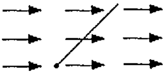

b

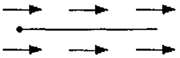

c

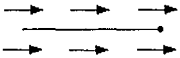

d

In case a, the water will not move the ball at all. In case b the ball will be pushed along the curve while in case c the water will move the ball the most quickly. In case d, not only will the ball not want to move along the curve C, effort is needed to even move the ball at all. These qualitative judgments are captured quantitatively in the above definition for circulation, since the dot product $\mathbf{F} \cdot \mathbf{T}$ measures at each point how much of the vector field **F** is pointing in the direction of the tangent **T** and hence how much of **F** is pointing in the direction of the curve.

In short, circulation measures how much of the vector field flows in the direction of the curve C. In physics, the vector field is frequently the force, in which case the circulation is a measurement of work.

Thus Stokes' Theorem is stating that the circulation of a vector field along a curve ∂M which bounds a surface M is precisely equal to the normal component of the vector field curl(**F**) in the interior. This is why the term 'curl' is used, as it measures the infinitesimal tendency of a vector field to have circulation, or in other words, it provides an infinitesimal measure of the "whirlpoolness" of the vector field.

5.5 Sketch of a Proof of the Divergence Theorem

This will only be a sketch, as we will be making a number of simplifying assumptions. First, assume that our three-dimensional manifold M (a solid) is *simple*, meaning that any line parallel to the $x-$axis, $y-$axis or $z-$axis can only intersect M in a connected line segment or a point. Thus

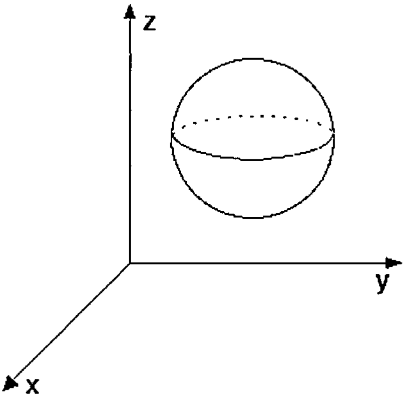

is simple while

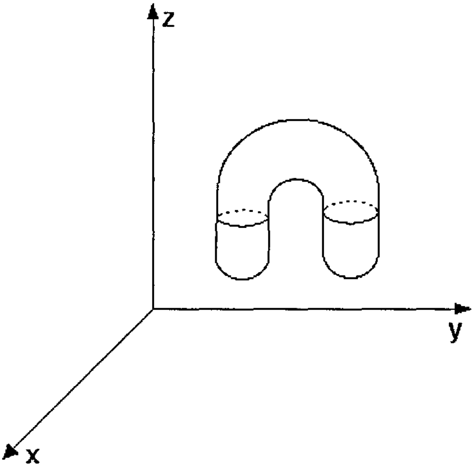

is not.

Denote the components of the vector field by

$$\begin{aligned}\mathbf{F}(z,y,z) &= (f_1(x,y,z), f_2(x,y,z), f_3(x,y,z)) \\ &= (f_1, f_2, f_3).\end{aligned}$$

On the boundary surface ∂M, denote the unit normal vector field by:

$$\begin{aligned}\mathbf{n}(x,y,z) &= (n_1(x,y,z), n_2(x,y,z), n_3(x,y,z)) \\ &= (n_1, n_2, n_3).\end{aligned}$$

We want to show that

$$\int\int_{\partial M} \mathbf{F}\cdot\mathbf{n}\,\mathrm{d}S = \int\int\int_M div(\mathbf{F})\mathrm{d}x\mathrm{d}y\mathrm{d}z.$$

In other words, we want

$$\int\int_{\partial M}(f_1n_1+f_2n_2+f_3n_3)\mathrm{d}S = \int\int\int_M\left(\frac{\partial f_1}{\partial x}+\frac{\partial f_2}{\partial y}+\frac{\partial f_3}{\partial z}\right)\mathrm{d}x\mathrm{d}y\mathrm{d}z.$$

If we can show

$$\begin{aligned}\int\int_{\partial M} f_1n_1\mathrm{d}S &= \int\int\int_M \frac{\partial f_1}{\partial x}\mathrm{d}x\mathrm{d}y\mathrm{d}z,\\ \int\int_{\partial M} f_2n_2\mathrm{d}S &= \int\int\int_M \frac{\partial f_2}{\partial y}\mathrm{d}x\mathrm{d}y\mathrm{d}z\\ \int\int_{\partial M} f_3n_3\mathrm{d}S &= \int\int\int_M \frac{\partial f_3}{\partial z}\mathrm{d}x\mathrm{d}y\mathrm{d}z\end{aligned}$$

we will be done.

We will just sketch the proof of the last equation

$$\int\int_{\partial M} f_3(x,y,z)n_3(x,y,z)\mathrm{d}S = \int\int\int_M \frac{\partial f_3}{\partial z}\mathrm{d}x\mathrm{d}y\mathrm{d}z,$$

since the other two equalities will hold for similar reasons.

The function $n_3(x,y,z)$ is the z-component of the normal vector field $\mathbf{n}(x,y,z)$. By the assumption that M is simple, we can split the boundary component ∂M into three connected pieces: $\{\partial M\}_{top}$, where $n_3 > 0$, $\{\partial M\}_{side}$, where $n_3 = 0$ and $\{\partial M\}_{bottom}$, where $n_3 < 0$.

For example, if ∂M is

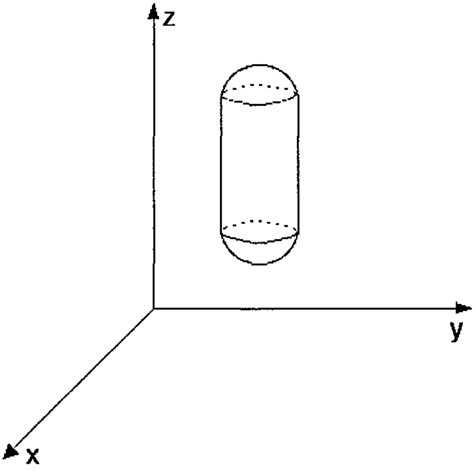

then

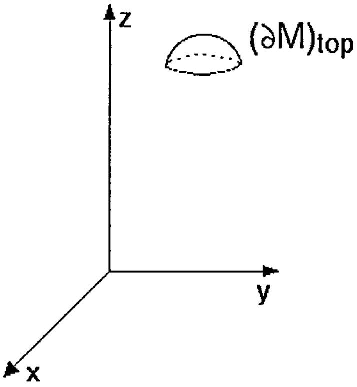
z
$(\partial M)_{top}$
y
x

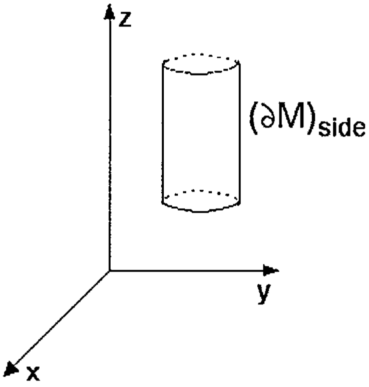
z
$(\partial M)_{side}$
y
x

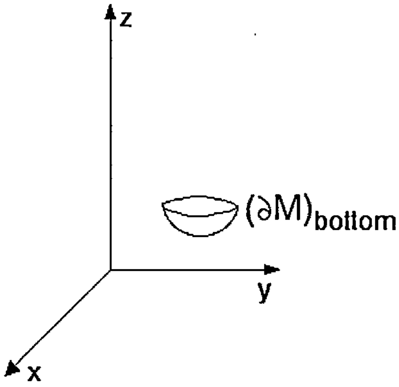
z
$(\partial M)_{bottom}$
y
x

Then we can split the boundary surface integral into three parts:

$$\int\int_{\partial M} f_3 n_3 \mathrm{d}S = \int\int_{\partial M_{top}} f_3 n_3 \mathrm{d}S + \int\int_{\partial M_{side}} f_3 n_3 \mathrm{d}S$$

$$+\int\int_{\partial M_{bottom}} f_3 n_3 \mathrm{d}S$$
$$= \int\int_{\partial M_{top}} f_3 n_3 \mathrm{d}S + \int\int_{\partial M_{bottom}} f_3 n_3 \mathrm{d}S,$$

since n_3, the normal component in the z direction, will be zero on ∂M_{side}.

Further, again by the assumption of simplicity, there is a region R in the xy-plane such that $\{\partial M\}_{top}$ is the image of a function

$$(x, y) \to (x, y, t(x, y))$$

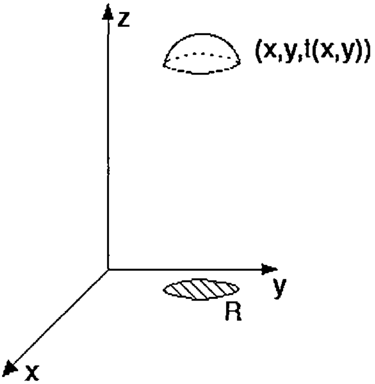

and $\{\partial M\}_{bottom}$ is the image of a function

$$(x, y) \to (x, y, b(x, y)).$$

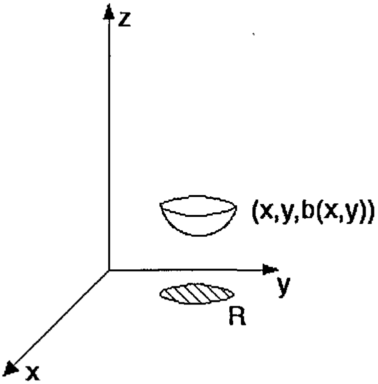

Then

$$\int\int_{\partial M} f_3 n_3 \mathrm{d}S = \int\int_{\partial M_{top}} f_3 n_3 \mathrm{d}S + \int\int_{\partial M_{bottom}} f_3 n_3 \mathrm{d}S$$

$$
\begin{aligned}
&= \int\int_R f_3(x,y,t(x,y))\mathrm{d}x\mathrm{d}y + \\
&\quad \int\int_R f_3(x,y,b(x,y))\mathrm{d}x\mathrm{d}y \\
&= \int\int_R (f_3(x,y,t(x,y)) - f_3(x,y,b(x,y)))\mathrm{d}x\mathrm{d}y,
\end{aligned}
$$

where the minus sign in front of the last term coming from the fact that the normal to ∂M_{bottom} points downward. But this is just

$$\int\int_R \int_{b(x,y)}^{t(x,y)} \frac{\partial f_3}{\partial z}\mathrm{d}x\mathrm{d}y\mathrm{d}z,$$

by the Fundamental Theorem of Calculus. This, in turn, is equal to

$$\int\int\int_M \frac{\partial f_3}{\partial z}\mathrm{d}x\mathrm{d}y\mathrm{d}z,$$

which is what we wanted to show.

To prove the full result, we would need to take any solid M and show that we can split M into simple parts and then that if the Divergence Theorem is true on each simple part, it is true on the original M. While not intuitively difficult, this is nontrivial to prove and involves some subtle questions of convergence.

5.6 Sketch of a Proof for Stokes' Theorem

Let M be a surface with boundary curve ∂M.

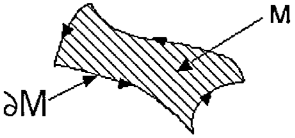

We break the proof of Stokes' Theorem into two steps. First, given two rectangles R_1 and R_2 that share a common side, we want

$$\int_{\partial R_1} \mathbf{F}\cdot\mathbf{T}\mathrm{d}s + \int_{\partial R_2} \mathbf{F}\cdot\mathbf{T}\mathrm{d}s = \int_{\partial R_1 \cup R_2} \mathbf{F}\cdot\mathbf{T}\mathrm{d}s,$$

where $\mathbf{T}$ is the unit tangent vector.

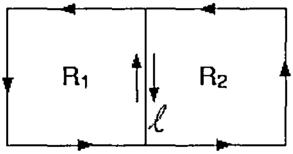

Second, we need to show that Stokes' Theorem is true on infinitesimally small rectangles.

The proof of the first is that for the common side ℓ of the two rectangles, the orientations are in opposite directions. This forces the value of the dot product $(\mathbf{F}\cdot\mathbf{T})$ along ℓ as a side of the rectangle R_1 to have opposite sign of the value of $(\mathbf{F}\cdot\mathbf{T})$ along ℓ as a side of the other rectangle R_2. Thus

$$\int_{\ell\subset\partial R_1}\mathbf{F}\cdot\mathbf{T}\,\mathrm{d}s=-\int_{\ell\subset\partial R_2}\mathbf{F}\cdot\mathbf{T}\,\mathrm{d}s.$$

Since the boundary of the union of the two rectangles $R_1\cup R_2$ does not contain the side ℓ, we have

$$\int_{\partial R_1}\mathbf{F}\cdot\mathbf{T}\,\mathrm{d}s+\int_{\partial R_2}\mathbf{F}\cdot\mathbf{T}\,\mathrm{d}s=\int_{\partial R_1\cup R_2}\mathbf{F}\cdot\mathbf{T}\,\mathrm{d}s.$$

Before proving that Stokes' Theorem is true on infinitesimally small rectangles, assume for a moment that we already know this to be true. Split the surface M into (infinitely many) small rectangles.

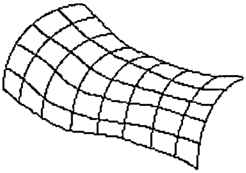

Then

$$\begin{aligned}\int\int_M curl(\mathbf{F})\cdot\mathbf{n}\,\mathrm{d}S &= \sum_{\text{small rectangles}}\int\int curl(\mathbf{F})\cdot\mathbf{n}\,\mathrm{d}S\\ &= \sum\int_{\partial(\text{each rectangle})}\mathbf{F}\cdot\mathbf{T}\,\mathrm{d}s,\end{aligned}$$

since we are assuming that Stokes' Theorem is true on infinitesimally small rectangles. But by the first step, the above sum will equal to the single integral over the boundary of the union of the small rectangles

$$\int_{\partial M}\mathbf{F}\cdot\mathbf{T}\,\mathrm{d}s,$$

which gives us Stokes' Theorem. Hence all we need to show is that Stokes' Theorem is true for infinitesimally small rectangles.

Before showing this, note that this argument is nonrigorous, as the whole sum is over infinitely many small rectangles, and thus subtle convergence questions would need to be solved. We pass over this in silence.

Now to sketch why Stokes' Theorem is true for infinitesimally small rectangles. This will also contain the justification for why the definition of the curl of a vector field is what it is.

By a change of coordinates, we can assume that our small rectangle R lies in the xy-plane with one vertex being the origin $(0,0)$.

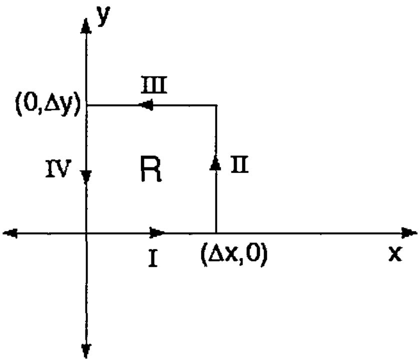

Its unit normal vector will be $\mathbf{n} = (0,0,1)$.

If the vector field is $\mathbf{F}(x,y,z) = (f_1, f_2, f_3)$, we have

$$curl(\mathbf{F}) \cdot \mathbf{n} = \frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y}.$$

We want to show that:

$$(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y})\mathrm{d}x\mathrm{d}y = \int_{\partial R} \mathbf{F} \cdot \mathbf{T}\, \mathrm{d}s,$$

where $\mathbf{T}$ is the unit tangent vector to the boundary rectangle ∂R and $\mathrm{d}x\ \mathrm{d}y$ is the infinitesimal area for the rectangle R.

Now to calculate $\int_{\partial R} \mathbf{F} \cdot \mathbf{T}\, \mathrm{d}s$.

The four sides of the rectangle ∂R have the following parametrizations.

Side	*Parametrization*	*Integral*
$I:$	$s(t) = (t\Delta x, 0), 0 \leq t \leq 1$	$\int_0^1 f_1(t\Delta x, 0)\Delta x\mathrm{d}t$
$II:$	$s(t) = (\Delta x, t\Delta y), 0 \leq t \leq 1$	$\int_0^1 f_2(\Delta x, t\Delta y)\Delta y\mathrm{d}t$
$III:$	$s(t) = (\Delta x - t\Delta x, \Delta y), 0 \leq t \leq 1$	$\int_0^1 -f_1(\Delta x - t\Delta x, \Delta y)\Delta x\mathrm{d}t$
$IV:$	$s(t) = (0, \Delta y - t\Delta y), 0 \leq t \leq 1$	$\int_0^1 -f_2(0, \Delta y - t\Delta y)\Delta y\mathrm{d}t$

It is always the case, for any function $f(t)$, that

$$\int_0^1 f(t)\mathrm{d}t = \int_0^1 f(1-t)\mathrm{d}t,$$

by changing the variable t to $1-t$. Thus the integrals for sides III and IV can be replaced by $\int_0^1 -f_1(t\Delta x, \Delta y)\Delta x \; \mathrm{d}t$ and $\int_0^1 -f_2(0, t\Delta y)\Delta y \; \mathrm{d}t$. Then

$$\begin{aligned}
\int_{\partial R} \mathbf{F}\cdot\mathbf{T}\,\mathrm{d}s &= \int_I \mathbf{F}\cdot\mathbf{T}\,\mathrm{d}s + \int_{II} \mathbf{F}\cdot\mathbf{T}\,\mathrm{d}s + \int_{III} \mathbf{F}\cdot\mathbf{T}\,\mathrm{d}s + \int_{IV} \mathbf{F}\cdot\mathbf{T}\,\mathrm{d}s \\
&= \int_0^1 (f_1(t\Delta x, 0)\Delta x + f_2(\Delta x, t\Delta y)\Delta y \\
&\quad -f_1(t\Delta x, \Delta y)\Delta x - f_2(0, t\Delta y)\Delta y)\mathrm{d}t \\
&= \int_0^1 (f_2(\Delta x, t\Delta y) - f_2(0, t\Delta y))\Delta y\mathrm{d}t \\
&\quad - \int_0^1 (f_1(t\Delta x, \Delta y) - f_1(t\Delta x, 0))\Delta x\mathrm{d}t \\
&= \int_0^1 \frac{f_2(\Delta x, t\Delta y) - f_2(0, t\Delta y)}{\Delta x} \\
&\quad - \frac{f_1(t\Delta x, \Delta y) - f_1(t\Delta x, y)}{\Delta y})\Delta x\Delta y\mathrm{d}t,
\end{aligned}$$

which converges to

$$\int_0^1 (\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y})\mathrm{d}x\mathrm{d}y\mathrm{d}t,$$

as $\Delta x, \Delta y \to 0$. But this last integral will be

$$(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y})\mathrm{d}x\mathrm{d}y$$

which is what we wanted.

Again, letting $\Delta x, \Delta y \to 0$ is a nonrigorous step. Also, the whole nonchalant way in which we changed coordinates to put our rectangle into the xy-plane would have to be justified in a rigorous proof.

5.7 Books

Most calculus books have sections near the end on the multivariable calculus covered in this chapter. A long time popular choice is Thomas and Finney's text [36]. Another good source is Stewart's *Calculus* [108].

Questions in physics, especially in electricity and magnetism, were the main historical motivation for the development of the mathematics in this chapter. There are physical "proofs" of the Divergence Theorem and Stokes' Theorem. Good sources are in Halliday and Resnick's text in physics [51] and in Feynmann's *Lectures in Physics* [35].

5.8 Exercises

1. Extend the proof of the Divergence Theorem, given in this chapter for simple regions, to the region:

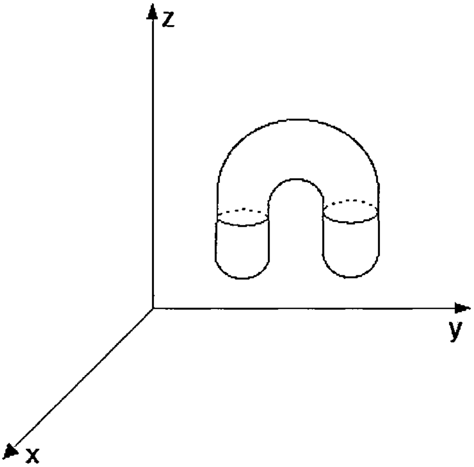

2. Let D be the disc of radius r, with boundary circle ∂D, given by the equations:

$$D = \{(x, y, 0) : x^2 + y^2 \leq r\}.$$

For the vector field

$$\mathbf{F}(x, y, z) = (x + y + z, 3x + 2y + 4z, 5x - 3y + z),$$

find the path integral $\int_{\partial D} \mathbf{F} \cdot \mathbf{T}\, ds$, where $\mathbf{T}$ is the unit tangent vector of the circle ∂D.

3. Consider the vector field

$$\mathbf{F}(x, y, z) = (x, 2y, 5z).$$

Find the surface integral $\int\int_{\partial M} \mathbf{F} \cdot \mathbf{n}\; \mathrm{d}S$, where the surface ∂M is the boundary of the ball

$$M = \{(x, y, z) : x^2 + y^2 + z^2 \leq r\}$$

of radius r centered at the origin and $\mathbf{n}$ is the unit normal vector.

4. Let S be the surface that is the image of the map

$$r : \mathbf{R}^2 \rightarrow \mathbf{R}^3$$

given by

$$r(u, v) = (x(u, v), y(u, v), z(u, v)).$$

Considering the image of the line $v =$ constant, justify to yourself that

$$(\frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}, \frac{\partial z}{\partial u})$$

is a tangent vector to S.

5. Green's Theorem is:

Theorem 5.8.1 (Green's Theorem) *Let σ be a simple loop in $\mathbf{C}$ and Ω its interior. If $P(x, y)$ and $Q(x, y)$ are two real-valued differentiable functions, then*

$$\int_{\sigma} P\ \mathrm{d}x + Q\ \mathrm{d}y = \int\int_{\Omega} (\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y})\ \mathrm{d}x\ \mathrm{d}y.$$

By putting the region Ω into the plane $z = 0$ and letting our vector field be $\langle P(x, y), Q(x, y), 0\rangle$, show that Green's Theorem follows from Stokes' Theorem.

Chapter 6

Differential Forms and Stokes' Theorem

Basic Objects:	Differential Forms and Manifolds
Basic Goal:	Stokes' Theorem

In the last chapter we saw various theorems, all of which related the values of a function on the boundary of a geometric object with the values of the function's derivative on the interior. The goal of this chapter is to show that there is a single theorem (Stokes' Theorem) underlying all of these results. Unfortunately, a lot of machinery is needed before we can even state this grand underlying theorem. Since we are talking about integrals and derivatives, we have to develop the techniques that will allow us to integrate on k-dimensional spaces. This will lead to differential forms, which are the objects on manifolds that can be integrated. The exterior derivative is the technique for differentiating these forms. Since integration is involved, we will have to talk about calculating volumes. This is done in section one. Section two defines differential forms. Section three links differential forms with the vector fields, gradients, curls and divergences from last chapter. Section four gives the definition of a manifold (actually, three different methods for defining manifolds are given). Section five concentrates on what it means for a manifold to be orientable. In section six, we define how to integrate a differential form along a manifold, allowing us finally in section seven to state and to sketch a proof of Stokes' Theorem.

6.1 Volumes of Parallelepipeds

In this chapter, we are ultimately interested in understanding integration on manifolds (which we have yet to define). This section, though, is pure linear algebra, but linear algebra that is crucial for the rest of the chapter.

The problem is the following: In $\mathbf{R}^n$, suppose we are given k vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$. These k vectors will define a parallelepiped in $\mathbf{R}^n$. The question is how to compute the volume of this parallelepiped. For example, consider the two vectors

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \text{ and } \mathbf{v}_2 = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}.$$

The parallelepiped that these two vectors span is a parallelogram in $\mathbf{R}^3$. We want a formula to calculate the area of this parallelogram. (Note: the true three dimensional volume of this flat parallelogram is zero, in the same way that the length of a point is zero and that the area of a line is zero; we are here trying to measure the two-dimensional "volume" of this parallelogram.)

We already know the answer in two special cases. For a single vector

$$\mathbf{v} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

in $\mathbf{R}^n$, the parallelepiped is the single vector $\mathbf{v}$. Here by "volume" we mean the length of this vector, which is, by the Pythagorean theorem,

$$\sqrt{a_1^2 + \cdots + a_n^2}.$$

The other case is when we are given n vectors in $\mathbf{R}^n$. Suppose the n vectors are

$$\mathbf{v}_1 = \begin{pmatrix} a_{11} \\ \vdots \\ a_{n1} \end{pmatrix}, \dots, \mathbf{v}_n = \begin{pmatrix} a_{1n} \\ \vdots \\ a_{nn} \end{pmatrix}.$$

Here we know that the volume of the resulting parallelepiped is

$$\left| \det \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ & \vdots & \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \right|,$$

following from one of the definitions of the determinant given in Chapter One. Our eventual formula will yield both of these results.

We will first give the formula and then discuss why it is reasonable. Write the k vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ as column vectors. Set

$$A = (\mathbf{v}_1, \ldots, \mathbf{v}_k),$$

a $k \times n$ matrix. We denote the transpose of A by A^T, the $n \times k$ matrix

$$A^T = \begin{pmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_k^T \end{pmatrix}$$

where each $\mathbf{v}_i^T$ is the writing of the vector $\mathbf{v}_i$ as a row vector. Then

Theorem 6.1.1 *The volume of the parallelepiped spanned by the vectors* $\mathbf{v}_1, \ldots, \mathbf{v}_k$ *is*

$$\sqrt{\det(A^T A)}.$$

Before sketching a proof, let us look at some examples. Consider the single vector

$$\mathbf{v} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

Here the matrix A is just $\mathbf{v}$ itself. Then

$$\begin{aligned}\sqrt{\det(A^T A)} &= \sqrt{\det(\mathbf{v}^T \mathbf{v})} \\ &= \sqrt{\det((a_1, \ldots, a_n) \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix})} \\ &= \sqrt{\det(a_1^2 + \cdots + a_n^2)} \\ &= \sqrt{a_1^2 + \cdots + a_n^2},\end{aligned}$$

the length of the vector $\mathbf{v}$.

Now consider the case of n vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Then the matrix A is $n \times n$. We will use that $\det(A) = \det(A^T)$. Then

$$\begin{aligned}\sqrt{\det(A^T A)} &= \sqrt{\det(A^T)\det(A)} \\ &= \sqrt{\det(A)^2} \\ &= |\det(A)|,\end{aligned}$$

as desired.

Now to see why in general $\sqrt{\det(A^TA)}$ must be the volume. We need a preliminary lemma that yields a more intrinsic, geometric approach to $\sqrt{\det(A^TA)}$.

Lemma 6.1.1 *For the matrix*

$$A = (\mathbf{v}_1, \ldots, \mathbf{v}_k),$$

we have that

$$A^TA = \begin{pmatrix} |\mathbf{v}_1|^2 & \mathbf{v}_1 \cdot \mathbf{v}_2 & \ldots & \mathbf{v}_1 \cdot \mathbf{v}_k \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{v}_k \cdot \mathbf{v}_1 & \mathbf{v}_k \cdot \mathbf{v}_2 & \ldots & |\mathbf{v}_k|^2 \end{pmatrix},$$

where $\mathbf{v}_i \cdot \mathbf{v}_j$ *denotes the dot product of vectors* $\mathbf{v}_i$ *and* $\mathbf{v}_j$ *and* $|\mathbf{v}_i| = \sqrt{\mathbf{v}_i \cdot \mathbf{v}_i}$ *denotes the length of the vector* $\mathbf{v}_i$.

The proof of this lemma is just looking at

$$A^TA = \begin{pmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_k^T \end{pmatrix} (\mathbf{v}_1, \ldots, \mathbf{v}_k).$$

Notice that if we apply any linear transformation on $\mathbf{R}^n$ that preserves angles and lengths (in other words, if we apply a rotation to $\mathbf{R}^n$), the numbers $|\mathbf{v}_i|$ and $\mathbf{v}_i \cdot \mathbf{v}_j$ do not change. (The set of all linear transformations of $\mathbf{R}^n$ that preserve angles and lengths form a group that is called the *orthogonal* group and denoted by $O(n)$.) This will allow us to reduce the problem to the finding of the volume of a parallelepiped in $\mathbf{R}^k$.
Sketch of Proof of Theorem: We know that

$$\sqrt{\det(A^TA)} = \sqrt{\det \begin{pmatrix} |\mathbf{v}_1|^2 & \mathbf{v}_1 \cdot \mathbf{v}_2 & \ldots & \mathbf{v}_1 \cdot \mathbf{v}_k \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{v}_k \cdot \mathbf{v}_1 & \mathbf{v}_k \cdot \mathbf{v}_2 & \ldots & |\mathbf{v}_k|^2 \end{pmatrix}}.$$

We will show that this must be the volume. Recall the standard basis for $\mathbf{R}^n$:

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

We can find a rotation of $\mathbf{R}^n$ that both preserves lengths and angles and more importantly, rotates our vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ so that they lie in the span of the first k standard vectors $\mathbf{e}_1, \dots, \mathbf{e}_k$. (To rigorously show this takes some work, but it is geometrically reasonable.) After this rotation, the last $n - k$ entries for each vector $\mathbf{v}_i$ are zero. Thus we can view our parallelepiped as being formed from k vectors in $\mathbf{R}^k$. But we already know how to compute this; it is

$$\sqrt{\det \begin{pmatrix} |\mathbf{v}_1|^2 & \mathbf{v}_1 \cdot \mathbf{v}_2 & \dots & \mathbf{v}_1 \cdot \mathbf{v}_k \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{v}_k \cdot \mathbf{v}_1 & \mathbf{v}_k \cdot \mathbf{v}_2 & \dots & |\mathbf{v}_k|^2 \end{pmatrix}}.$$

We are done. □

6.2 Differential Forms and the Exterior Derivative

This will be a long and, at times, technical section. We will initially define elementary k-forms on $\mathbf{R}^n$, for which there is still clear geometric meaning. We will then use these elementary k-forms to generate general k-forms. Finally, and for now no doubt the most unintuitive part, we will give the definition for the exterior derivative, a device that will map k-forms to $(k+1)$-forms and will eventually be seen to be a derivative-type operation. In the next section we will see that the gradient, the divergence and the curl of the last chapter can be interpreted in terms of the exterior derivative.

6.2.1 Elementary k-forms

We start with trying to understand elementary 2-forms in $\mathbf{R}^3$. Label the coordinate axis for $\mathbf{R}^3$ as x_1, x_2, x_3. There will be three elementary 2-forms, which will be denoted by $dx_1 \wedge dx_2$, $dx_1 \wedge dx_3$ and $dx_2 \wedge dx_3$. We must now determine what these symbols mean. (We will define 1-forms in a moment.)

In words, $dx_1 \wedge dx_2$ will measure the signed area of the projection onto the x_1x_2-plane of any parallelepiped in $\mathbf{R}^3$, $dx_1 \wedge dx_3$ will measure the signed area of the projection onto the x_1x_3-plane of any parallelepiped in $\mathbf{R}^3$ and $dx_2 \wedge dx_3$ will measure the signed area of the projection onto the x_2x_3-plane of any parallelepiped in $\mathbf{R}^3$.

By looking at an example, we will see how to actually compute with these 2-forms. Consider two vectors in $\mathbf{R}^3$, labelled

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \text{ and } \mathbf{v}_2 = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}.$$

These vectors span a parallelepiped P in $\mathbf{R}^3$. Consider the projection map $\pi : \mathbf{R}^3 \to \mathbf{R}^2$ of $\mathbf{R}^3$ to the x_1x_2 plane. Thus

$$\pi(x_1, x_2, x_3) = (x_1, x_2).$$

We define $\mathrm{d}x_1 \wedge \mathrm{d}x_2$ acting on the parallelepiped P to be the area of $\pi(P)$. Note that

$$\pi(\mathbf{v}_1) = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ and } \pi(\mathbf{v}_2) = \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

Then $\pi(P)$ is the parallelogram:

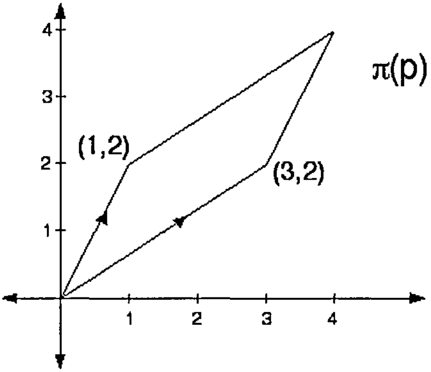

and the signed area is

$$\begin{aligned} \mathrm{d}x_1 \wedge \mathrm{d}x_2(P) &= \det(\pi(\mathbf{v}_1), \pi(\mathbf{v}_2)) \\ &= \det \begin{pmatrix} 1 & 3 \\ 2 & 2 \end{pmatrix} \\ &= -4 \end{aligned}$$

In general, given a 3×2 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix},$$

its two columns will define a parallelepiped. Then $\mathrm{d}x_1 \wedge \mathrm{d}x_2$ of this parallelepiped will be

$$\mathrm{d}x_1 \wedge \mathrm{d}x_2(A) = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

In the same way, $\mathrm{d}x_1 \wedge \mathrm{d}x_3$ will measure the area of the projection of a parallelepiped onto the x_1x_3-plane. Then

$$\mathrm{d}x_1 \wedge \mathrm{d}x_3(A) = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{pmatrix}.$$

Likewise, we need

$$\mathrm{d}x_2 \wedge \mathrm{d}x_3(A) = \det\begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}.$$

Before defining elementary k-forms in general, let us look at elementary 1-forms. In $\mathbf{R}^3$, there are three elementary 1-forms, which will be denoted by $\mathrm{d}x_1, \mathrm{d}x_2$ and $\mathrm{d}x_3$. Each will measure the one-dimensional volume (the length) of the projection of a one-dimensional parallelepiped in $\mathbf{R}^3$ to a coordinate axis. For example, with

$$\mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix},$$

its projection to the x_1-axis is just (1). Then we want to define

$$\mathrm{d}x_1(\mathbf{v}) = \mathrm{d}x_1\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = 1.$$

In general, for a vector

$$\begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix}$$

we have

$$\mathrm{d}x_1\begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} = a_{11},\ \mathrm{d}x_2\begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} = a_{21},\ \mathrm{d}x_3\begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} = a_{31}.$$

Now to define elementary k-forms on $\mathbf{R}^n$. Label the coordinates of $\mathbf{R}^n$ as $x_1, \ldots, x_n$. Choose an increasing subsequence of length k from $(1, 2, \ldots, n)$, which we will denote by

$$I = (i_1, \ldots, i_k)$$

with $1 \leq i_1 < \ldots < i_k \leq n$. Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1k} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \ldots & \ldots & a_{nk} \end{pmatrix}$$

be an $n \times k$ matrix. Its columns will span a k-dimensional parallelepiped P in $\mathbf{R}^n$. For convenience of exposition, let A_i be the ith row of A, i.e.,

$$A = \begin{pmatrix} A_1 \\ \vdots \\ A_n \end{pmatrix}.$$

We want the elementary k-form

$$\mathrm{d}x_I = \mathrm{d}x_{i_i} \wedge \cdots \wedge \mathrm{d}x_{i_k}$$

to act on the matrix A to give us the k-dimensional volume of the parallelepiped P projected onto the k-dimensional $x_{i_1}, \ldots, x_{i_k}$ space. This motivates the definition:

$$\mathrm{d}x_I(A) = \mathrm{d}x_{i_i} \wedge \cdots \wedge \mathrm{d}x_{i_k}(A) = \det \begin{pmatrix} A_{i_1} \\ \vdots \\ A_{i_k} \end{pmatrix}.$$

Elementary k-forms are precisely the devices that measure the volumes of k-dimensional parallelepipeds after projecting to coordinate k-spaces. The calculations come down to taking determinants of the original matrix with some of its rows deleted.

6.2.2 The Vector Space of k-forms

Recall back in Chapter One that we gave three different interpretations for the determinant of a matrix. The first was just how to compute it. The third was in terms of volumes of parallelepipeds, which is why determinants are showing up here. We now want to concentrate on the second interpretation, which in words was that the determinant is a multilinear map on the space of columns of a matrix. More precisely, if $M_{nk}(\mathbf{R})$ denotes the space of all $n \times k$ matrices with real entries, we had that the *determinant* of an $n \times n$ matrix A is defined as the unique real-valued function

$$\det : M_{nn}(\mathbf{R}) \to \mathbf{R}$$

satisfying:
a) $\det(A_1, ..., \lambda A_k, ..., A_n) = \lambda \det(A_1, ..., A_k)$.
b) $\det(A_1, ..., A_k + \lambda A_i, ..., A_n) = \det(A_1, ..., A_n)$ for $k \neq i$.
c) $\det(\text{Identity matrix}) = 1$.
A k-form will have a similar looking definition:

Definition 6.2.1 *A k-form ω is a real-valued function*

$$\omega : M_{nk}(\mathbf{R}) \to \mathbf{R}$$

satisfying:

$$\omega(A_1, ..., \lambda B + \mu C, ..., A_k) = \lambda\omega(A_1, ..., B, ..., A_k) + \mu\omega(A_1, ..., C, ..., A_k).$$

Thus ω is a multilinear real-valued function.

By the properties of determinants, we can see that each elementary k-form dx_I is in fact a k-form. (Of course this would have to be the case, or we wouldn't have called them elementary k-forms in the first place.) But in fact we have

Theorem 6.2.1 *The k-forms for a vector space $\mathbf{R}^n$ form a vector space of dimension $\binom{n}{k}$. The elementary k-forms are a basis for this vector space. This vector space is denoted by $\bigwedge^k(\mathbf{R}^n)$.*

We will not prove this theorem. It is not hard to prove that the k-forms are a vector space. It takes a bit more work to show that the elementary k-forms are a basis for $\bigwedge^k(\mathbf{R}^n)$.

Finally, note that 0-forms are just the real numbers themselves.

6.2.3 Rules for Manipulating k-forms

There is a whole machinery for manipulating k-forms. In particular, a k-form and an l-form can be combined to make a $(k+l)$-form. The method for doing this is not particularly easy to intuitively understand, but once you get the hang of it, it is a straightforward computational tool. We will look carefully at the $\mathbf{R}^2$ case, then describe the general rule for combining forms and finally see how this relates to the $\mathbf{R}^n$ case.

Let x_1 and x_2 be the coordinates for $\mathbf{R}^2$. Then dx_1 and dx_2 are the two elementary 1-forms and $dx_1 \wedge dx_2$ is the only elementary 2-form. But it looks, at least notationally, that the two 1-forms dx_1 and dx_2 somehow make up the 2-form $dx_1 \wedge dx_2$. We will see that this is indeed the case.

Let

$$\mathbf{v}_1 = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} \text{ and } \mathbf{v}_2 = \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}$$

be two vectors in $\mathbf{R}^2$. Then

$$dx_1(\mathbf{v}_1) = a_{11} \text{ and } dx_1(\mathbf{v}_2) = a_{12}$$

and

$$dx_2(\mathbf{v}_1) = a_{21} \text{ and } dx_2(\mathbf{v}_2) = a_{22}.$$

The 2-form $dx_1 \wedge dx_2$ acting on the 2×2 matrix $(\mathbf{v}_1, \mathbf{v}_2)$ is the area of the parallelogram spanned by the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ and is hence the determinant of the matrix $(\mathbf{v}_1, \mathbf{v}_2)$. Thus

$$dx_1 \wedge dx_2(\mathbf{v}_1, \mathbf{v}_2) = a_{11}a_{22} - a_{12}a_{21}.$$

But note that this equals

$$dx_1(\mathbf{v}_1)\, dx_2(\mathbf{v}_2) - dx_1(\mathbf{v}_2)\, dx_2(\mathbf{v}_1).$$

At some level we have related our 2-form $dx_1 \wedge dx_2$ with our 1-forms dx_1 and dx_2, but it is not clear what is going on. In particular, at first glance it would seem to make more sense to change the above minus sign to a plus sign, but then, unfortunately, nothing would work out correctly.

We need to recall a few facts about the permutation group on n elements, S_n. (There is more discussion about permutations in Chapter Eleven.) Each element of S_n permutes the ordering of the set $\{1, 2, \ldots, n\}$. In general, every element of S_n can be expressed as the composition of flips (or transpositions).

If we need an even number of flips to express an element, we say that the element has sign 0 while if we need an odd number of flips, then the sign is 1. (Note that in order for this to be well-defined, we need to show that if an element has sign 0 (1), then it can only be written as the composition of an even (odd) number of flips; this is indeed true, but we will not show it.)

Consider S_2. There are only two ways we can permute the set $\{1, 2\}$. We can either just leave $\{1, 2\}$ alone (the identity permutation), which has sign 0, or flip $\{1, 2\}$ to $\{2, 1\}$, which has sign 1. We will denote the flip that sends $\{1, 2\}$ to $\{2, 1\}$ by $(1, 2)$. There are six ways of permuting the three elements $\{1, 2, 3\}$ and thus six elements in S_3. Each can be written as the composition of flips. For example, the permutation that sends $\{1, 2, 3\}$ to $\{3, 1, 2\}$ (which means that the first element is sent to the second slot, the second to the third slot and the third to the first slot) is the composition of the flip $(1, 2)$ with the flip $(1, 3)$, since, starting with $\{1, 2, 3\}$ and applying the flip $(1, 2)$, we get $\{2, 1, 3\}$. Then applying the flip $(1, 3)$ (which just interchanges the first and third elements), we get $\{3, 1, 2\}$.

We will use the following notational convention. If σ denotes the flip $(1, 2)$, then we say that

$$\sigma(1) = 2 \text{ and } \sigma(2) = 1.$$

Similarly, if σ denotes the element $(1, 2)$ composed with $(1, 3)$ in S_3, then we write

$$\sigma(1) = 2, \sigma(2) = 3 \text{ and } \sigma(3) = 1,$$

since under this permutation one is sent to two, two is sent to three and three is sent to one.

Suppose we have a k-form and an l-form. Let $n = k+l$. We will consider a special subset of S_n, the (k, l) shuffles, which are all elements $\sigma \in S_n$ that have the property that

$$\sigma(1) < \sigma(2) < \cdots < \sigma(k)$$

and

$$\sigma(k+1) < \sigma(k+2) < \cdots < \sigma(k+l).$$

Thus the element σ that is the composition of $(1,2)$ with $(1,3)$ is a $(2,1)$ shuffle, since

$$\sigma(1) = 2 < 3 = \sigma(2).$$

Denote the set of all (k,l) shuffles by $S(k,l)$. One of the exercises at the end of the chapter is to justify why these are called shuffles.

We can finally formally define the wedge product.

Definition 6.2.2 *Let $A = (A_1, \ldots, A_{k+l})$ be an $N \times (k+l)$ matrix, for any N. (Here each A_i denotes a column vector.) Let τ be a k-form and ω be an l-form. Then define*

$$\tau \wedge \omega(A) = \sum_{\sigma \in S(k,l)} (-1)^{sign(\sigma)} \tau(A_{\sigma(1)}, \ldots, A_{\sigma(k)}) \omega(A_{\sigma(k+1)}, \ldots, A_{\sigma(k+l)}).$$

Using this definition allows us to see that the wedge in $\mathbf{R}^2$ of two elementary 1-forms does indeed give us an elementary 2-form. A long calculation will show that in $\mathbf{R}^3$, the wedge of three elementary 1-forms yields the elementary 3-form.

It can be shown by these definitions that two 1-forms will anti-commute, meaning that

$$dx \wedge dy = -dy \wedge dx.$$

In general, we have that if τ is a k-form and ω is an l-form, then

$$\tau \wedge \omega = (-1)^{kl} \omega \wedge \tau.$$

This can be proven by directly calculating from the above definition of wedge product (though this method of proof is not all that enlightening). Note that for k and l both being odd, this means that

$$\tau \wedge \omega = (-1)\omega \wedge \tau.$$

Then for k being odd, we must have that

$$\tau \wedge \tau = (-1)\tau \wedge \tau,$$

which can only occur if

$$\tau \wedge \tau = 0.$$

In particular, this means that it is always the case that

$$dx_i \wedge dx_i = 0$$

and, if $i \neq j$,

$$dx_i \wedge dx_j = -dx_j \wedge dx_i.$$

6.2.4 Differential k-forms and the Exterior Derivative

Here the level of abstraction will remain high. We are after a general notion of what can be integrated (which will be the differential k-forms) and a general notion of what a derivative can be (which will be the exterior derivative).

First to define differential k-forms. In $\mathbf{R}^n$, if we let $I = \{i_1, \ldots, i_k\}$ denote some subsequence of integers with

$$1 \leq i_1 < \ldots < i_k \leq n,$$

then we let

$$\mathrm{d}x_I = \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k}.$$

Then a *differential k-form* ω is:

$$\omega = \sum_{\text{all possible } I} f_I \, \mathrm{d}x_I,$$

where each $f_I = f_I(x_1, \ldots, x_n)$ is a differentiable function.

Thus

$$(x_1 + \sin(x_2))\mathrm{d}x_1 + x_1 x_2 \mathrm{d}x_2$$

is an example of a differential 1-form, while

$$e^{x_1+x_3} \, \mathrm{d}x_1 \wedge \mathrm{d}x_3 + x_2^3 \, \mathrm{d}x_2 \wedge \mathrm{d}x_3$$

is a differential 2-form.

Each differential k-form defines at each point of $\mathbf{R}^n$ a different k-form. For example, the differential 1-form $(x_1 + \sin(x_2))\mathrm{d}x_1 + x_1 x_2 \mathrm{d}x_2$ is the 1-form $3 \, \mathrm{d}x_1$ at the point $(3, 0)$ and is $5\mathrm{d}x_1 + 2\pi \mathrm{d}x_2$ at the point $(4, \frac{\pi}{2})$.

To define the exterior derivative, we first define the exterior derivative of a differential 0-form and then by induction define the exterior derivative for a general differential k-form. We will see that the exterior derivative is a map from k-forms to $(k+1)$-forms:

$$\mathrm{d} : k\text{-forms} \rightarrow (k+1)\text{-forms}.$$

A differential 0-form is just another name for a differentiable function. Given a 0-form $f(x_1, \ldots, x_n)$, its *exterior derivative*, denoted by $\mathrm{d}f$, is:

$$\mathrm{d}f = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \, \mathrm{d}x_i.$$

For example, if $f(x_1, x_2) = x_1 x_2 + x_2^3$, then

$$\mathrm{d}f = x_2 \mathrm{d}x_1 + (x_1 + 3x_2^2)\mathrm{d}x_2.$$

Note that the gradient of f is the similar looking $(x_2, x_1 + 3x_2^2)$. We will see in the next section that this is not chance.

Given a k-form $\omega = \sum_{\text{all possible } I} f_I \mathrm{d}x_I$, the *exterior derivative* $\mathrm{d}\omega$ is:

$$\mathrm{d}\omega = \sum_{\text{all possible } I} \mathrm{d}f_I \wedge \mathrm{d}x_I.$$

For example, in $\mathbf{R}^3$, let

$$\omega = f_1 \mathrm{d}x_1 + f_2 \mathrm{d}x_2 + f_3 \mathrm{d}x_3$$

be some 1-form. Then

$$\begin{aligned}
\mathrm{d}\omega &= \mathrm{d}f_1 \mathrm{d}x_1 + \mathrm{d}f_2 \mathrm{d}x_2 + \mathrm{d}f_3 \mathrm{d}x_3 \\
&= \left(\frac{\partial f_1}{\partial x_1}\mathrm{d}x_1 + \frac{\partial f_1}{\partial x_2}\mathrm{d}x_2 + \frac{\partial f_1}{\partial x_3}\mathrm{d}x_3\right) \wedge \mathrm{d}x_1 \\
&\quad + \left(\frac{\partial f_2}{\partial x_1}\mathrm{d}x_1 + \frac{\partial f_2}{\partial x_2}\mathrm{d}x_2 + \frac{\partial f_2}{\partial x_3}\mathrm{d}x_3\right) \wedge \mathrm{d}x_2 \\
&\quad + \left(\frac{\partial f_3}{\partial x_1}\mathrm{d}x_1 + \frac{\partial f_3}{\partial x_2}\mathrm{d}x_2 + \frac{\partial f_3}{\partial x_3}\mathrm{d}x_3\right) \wedge \mathrm{d}x_3 \\
&= \left(\frac{\partial f_3}{\partial x_1} - \frac{\partial f_1}{\partial x_3}\right)\mathrm{d}x_1 \wedge \mathrm{d}x_3 + \left(\frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2}\right)\mathrm{d}x_1 \wedge \mathrm{d}x_2 \\
&\quad + \left(\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3}\right)\mathrm{d}x_2 \wedge \mathrm{d}x_3.
\end{aligned}$$

Note that this looks similar to the curl of the vector field

$$(f_1, f_2, f_3).$$

Again, we will see that this similarity is not just chance.

Key to many calculations is:

Proposition 6.2.1 *For any differential k-form ω, we have*

$$\mathrm{d}(\mathrm{d}\omega) = 0.$$

The proof is one of the exercises at the end of the chapter, but you need to use that in $\mathbf{R}^n$ the order of differentiation does not matter, i.e.,

$$\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j} = \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i},$$

and that $\mathrm{d}x_i \wedge \mathrm{d}x_j = -\mathrm{d}x_j \wedge \mathrm{d}x_i$.

6.3 Differential Forms and Vector Fields

The overall goal for this chapter is to show that the classical Divergence Theorem, Green's Theorem and Stokes' Theorem are all special cases of one general theorem. This one general theorem will be stated in the language of differential forms. In order to see how it reduces to the theorems of last chapter, we need to relate differential forms with functions and vector fields. In $\mathbf{R}^3$, we will see that the exterior derivative, under suitable interpretation, will correspond to the gradient, the curl and the divergence.

Let x, y and z denote the standard coordinates for $\mathbf{R}^3$. Our first step is to define maps

$$\begin{aligned} T_0 : \text{0-forms} &\rightarrow \text{functions on } \mathbf{R}^3 \\ T_1 : \text{1-forms} &\rightarrow \text{vector fields on } \mathbf{R}^3 \\ T_2 : \text{2-forms} &\rightarrow \text{vector fields on } \mathbf{R}^3 \\ T_3 : \text{3-forms} &\rightarrow \text{functions on } \mathbf{R}^3. \end{aligned}$$

We will see that T_0, T_1 and T_3 have natural definitions. The definition for T_2 will take a bit of justification.

In the last section, we saw that differential 0-forms are just functions. Thus T_0 is just the identity map. From last section, we know that there are three elementary 1-forms: $\mathrm{d}x$, $\mathrm{d}y$ and $\mathrm{d}z$. Thus a general differential 1-form will be

$$\omega = f_1(x,y,z)\mathrm{d}x + f_2(x,y,z)\mathrm{d}y + f_3(x,y,z)\mathrm{d}z,$$

where f_1, f_2 and f_3 are three separate functions on $\mathbf{R}^3$. Then define

$$T_1(\omega) = (f_1, f_2, f_3).$$

The definition for T_3 is just as straightforward. We know that on $\mathbf{R}^3$ there is only a single elementary 3-form, namely $\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z$. Thus a general differential 3-form looks like:

$$\omega = f(x,y,z)\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z,$$

where f is a function on $\mathbf{R}^3$. Then we let

$$T_3(\omega) = f(x,y,z).$$

As we mentioned, the definition for T_2 is not as straightforward. There are three elementary 2-forms: $\mathrm{d}x \wedge \mathrm{d}y$, $\mathrm{d}x \wedge \mathrm{d}z$ and $\mathrm{d}y \wedge \mathrm{d}z$. A general differential 2-form looks like:

$$\omega = f_1(x,y,z)\mathrm{d}x \wedge \mathrm{d}y + f_2(x,y,z)\mathrm{d}x \wedge \mathrm{d}z + f_3(x,y,z)\mathrm{d}y \wedge \mathrm{d}z,$$

where, as expected, f_1, f_2 and f_3 are functions on $\mathbf{R}^3$. Define the map T_2 by:

$$T_2(\omega) = (f_3, -f_2, f_1).$$

One method for justifying this definition will be that it will allow us to prove the theorems needed to link the exterior derivative with the gradient, the curl and the divergence. A second method will be in terms of dual spaces, as we will see in a moment.

We want to show:

Theorem 6.3.1 *On* $\mathbf{R}^3$, *let* ω_k *denote a k-form. Then*

$$T_1(\mathrm{d}\omega_0) = grad(T_0(\omega_0)),$$

$$T_2(\mathrm{d}\omega_1) = curl(T_1(\omega_1)),$$

and

$$T_3(\mathrm{d}\omega_2) = div(T_2(\omega_2)).$$

Each is a calculation (and is an exercise at the end of this chapter). We needed to define T_2 as we did in order to make the above work; this is one of the ways that we can justify our definition for the map T_2.

There is another justification for why T_2 must be what it is. This approach is a bit more abstract, but ultimately more important, as it generalizes to higher dimensions. Consider $\mathbf{R}^n$ with coordinates $x_1, \ldots, x_n$. There is only a single elementary n-form, namely $\mathrm{d}x_1 \wedge \ldots \wedge \mathrm{d}x_n$. Thus the vector space $\bigwedge^n(\mathbf{R}^n)$ of n-forms on $\mathbf{R}^n$ is one-dimensional and can be identified to the real numbers $\mathbf{R}$. Label this map by

$$T : \bigwedge^n(\mathbf{R}^n) \to \mathbf{R}.$$

Thus $T(\alpha \mathrm{d}x_1 \wedge \ldots \wedge \mathrm{d}x_n) = \alpha$.

We now want to see that the dual vector space to $\bigwedge^k(\mathbf{R}^n)$ can be naturally identified with the vector space $\bigwedge^{n-k}(\mathbf{R}^n)$. Let ω_{n-k} be in $\bigwedge^{n-k}(\mathbf{R}^n)$. We first show how an $(n-k)$-form can be interpreted as a linear map on $\bigwedge^k(\mathbf{R}^n)$. If ω_k is any k-form, define

$$\omega_{n-k}(\omega_k) = T(\omega_{n-k} \wedge \omega_k).$$

It is a direct calculation that this is a linear map. From Chapter One we know that the dual vector space has the same dimension as the original vector space. By direct calculation, we also know that the dimensions for $\bigwedge^k(\mathbf{R}^n)$ and $\bigwedge^{n-k}(\mathbf{R}^n)$ are the same. Thus $\bigwedge^{n-k}(\mathbf{R}^n)$ is the dual space to $\bigwedge^k(\mathbf{R}^n)$.

Now consider the vector space $\bigwedge^1(\mathbf{R}^3)$, with its natural basis of $\mathrm{d}x, \mathrm{d}y$ and $\mathrm{d}z$. Its dual is then $\bigwedge^2(\mathbf{R}^3)$. As a dual vector space, an element of the natural basis is that which sends one of the basis vectors of $\bigwedge^1(\mathbf{R}^3)$ to one and the other basis vectors to zero. Thus the natural basis for $\bigwedge^2(\mathbf{R}^3)$, thought of as a dual vector space, is $dy \wedge dz$ (which corresponds to the 1-form dx, since $dy \wedge dz \wedge dx = 1 \cdot dx \wedge dy \wedge dz$), $-dx \wedge dz$ (which corresponds to dy) and $dx \wedge dy$ (which corresponds to dz). Then identifying dx with the row vector $(1,0,0)$, dy with $(0,1,0)$ and dz with $(0,0,1)$, we see that $dy \wedge dz$ should be identified with $(1,0,0)$, $dx \wedge dz$ with $(0,-1,0)$ and $dx \wedge dy$ with $(0,0,1)$. Then the 2-form

$$\omega = f_1 dx \wedge dy + f_2 dx \wedge dz + f_3 dy \wedge dz$$

should indeed be identified with $(f_3, -f_2, f_1)$, which is precisely how the map T_2 is defined.

6.4 Manifolds

While manifolds are to some extent some of the most natural occurring geometric objects, it takes work and care to create correct definitions. In essence, though, a k-dimensional manifold is any topological space that, in a neighborhood of any point, looks like a ball in $\mathbf{R}^k$. We will be at first concerned with manifolds that live in some ambient $\mathbf{R}^n$. For this type of manifold, we give two equivalent definitions: the parametric version and the implicit version. For each of these versions, we will carefully show that the unit circle S^1 in $\mathbf{R}^2$

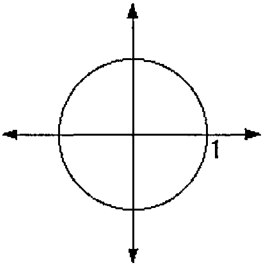

is a one-dimensional manifold. (Of course if we were just interested in circles we would not need all of these definitions; we are just using the circle to get a feel for the correctness of the definitions.) Then we will define an abstract manifold, a type of geometric object which need not be defined in terms of some ambient $\mathbf{R}^n$.

Consider again the circle S^1. Near any point $p \in S^1$ the circle looks like an interval (admittedly a bent interval). In a similar fashion, we want our definitions to yield that the unit sphere S^2 in $\mathbf{R}^3$ is a two-dimensional manifold, since near any point $p \in S^2$,

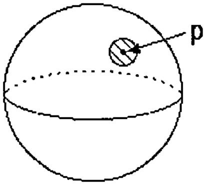

the sphere looks like a disc (though, again, more like a bent disc). We want to exclude from our definition of a manifold objects which contain points for which there is no well-defined notion of a tangent space, such as

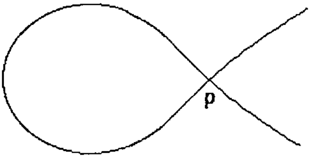

which has tangent difficulties at p, and the cone

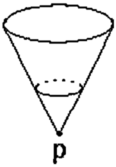

which has tangent difficulties at the vertex p. As a technical note, we will throughout this section let M denote a second countable Hausdorff topological space.

For $k \leq n$, a k-dimensional *parametrizing* map is any differentiable map

$$\phi : (\text{Ball in } \mathbf{R}^k) \to \mathbf{R}^n$$

such that the rank of the Jacobian at every point is exactly k. In local coordinates, if $u_1, \ldots, u_k$ are the coordinates for $\mathbf{R}^k$ and if ϕ is described by the n differentiable functions $\phi_1, \ldots, \phi_n$ (i.e., $\phi = (\phi_1, \ldots, \phi_n)$), we require that at all points there is a $k \times k$ minor of the $n \times k$ Jacobian matrix

$$D\phi = \begin{pmatrix} \frac{\partial \phi_1}{\partial u_1} & \cdots & \frac{\partial \phi_1}{\partial u_k} \\ \vdots & & \vdots \\ \frac{\partial \phi_n}{\partial u_1} & \cdots & \frac{\partial \phi_n}{\partial u_k} \end{pmatrix}$$

that is invertible.

Definition 6.4.1 (Parametrized Manifolds) *The Hausdorff topological space M in $\mathbf{R}^n$ is a k-*dimensional manifold *if for every point $p \in M$ in $\mathbf{R}^n$, there is an open set U in $\mathbf{R}^n$ containing the point p and a parametrizing map ϕ such that*

$$\phi(\text{Ball in } \mathbf{R}^k) = M \cap U.$$

Consider the circle S^1. At the point $p = (1, 0)$, a parametrizing map is:

$$\phi(u) = (\sqrt{1-u^2}, u),$$

while for the point $(0, 1)$, a parametrizing map could be:

$$\phi(u) = (u, \sqrt{1-u^2}).$$

Given the parametrization, we will see in section five that it is easy to find a basis for the tangent space of the manifold. More precisely the tangent space is spanned by the columns of the Jacobian $D\phi$. This is indeed one of the computational strengths of using parametrizations for defining a manifold.

Another approach is to define a manifold as the zero locus of a set of functions on $\mathbf{R}^n$. Here the normal vectors are practically given to us in the definition.

Definition 6.4.2 (Implicit Manifolds) *A set M in $\mathbf{R}^n$ is a k-*dimensional manifold *if, for any point $p \in M$ there is an open set U containing p and $(n-k)$ differentiable functions $\rho_1, \ldots, \rho_{n-k}$ such that*

1. $M \cap U = (\rho_1 = 0) \cap \cdots \cap (\rho_{n-k} = 0)$.
2. *At all points in $M \cap U$, the gradient vectors*

$$\nabla\rho_1, \ldots, \nabla\rho_{n-k}$$

are linearly independent.

It can be shown that the normal vectors are just the various $\nabla\rho_j$.

For an example, turn again to the circle S^1. The implicit method just notes that

$$S^1 = \{(x, y) : x^2 + y^2 - 1 = 0\}.$$

Here we have $\rho = x^2 + y^2 - 1$. Since

$$\nabla(x^2 + y^2 - 1) = (2x, 2y)$$

is never the zero vector, we are done.

The two definitions are equivalent, as discussed in the section on the implicit function theorem. But both of these definitions depend on our set M being in $\mathbf{R}^n$. Both critically use the properties of this ambient $\mathbf{R}^n$. There are situations where we still want to do calculus on a set of points which do not seem to live, in any natural way, in some $\mathbf{R}^n$. Historically this was first highlighted in Einstein's General Theory of Relativity, in which the universe itself was described as a 4-dimensional manifold that is neither $\mathbf{R}^4$ nor living in any natural way in a higher dimensional $\mathbf{R}^n$. By all accounts, Einstein was amazed that mathematicians had built up the whole needed machinery. Our goal here is to give the definition of an abstract manifold and then to show, once again, that S^1 is a manifold. Throughout this we will be using that we already know what it means for a function $f : \mathbf{R}^n \to \mathbf{R}^n$ to be differentiable.

Definition 6.4.3 (Manifolds) *A second countable Hausdorff topological space M is an n-*dimensional manifold *if there is an open cover (U_α) such that for each open set, U_α, we have a continuous map*

$$\phi_\alpha : \textit{Open ball in } \mathbf{R}^n \to U_\alpha$$

that is one-to-one and onto and such that the map

$$\phi_\alpha^{-1}\phi_\beta : \phi_\beta^{-1}(U_\alpha \cap U_\beta) \to \phi_\alpha^{-1}(U_\alpha \cap U_\beta)$$

is differentiable.

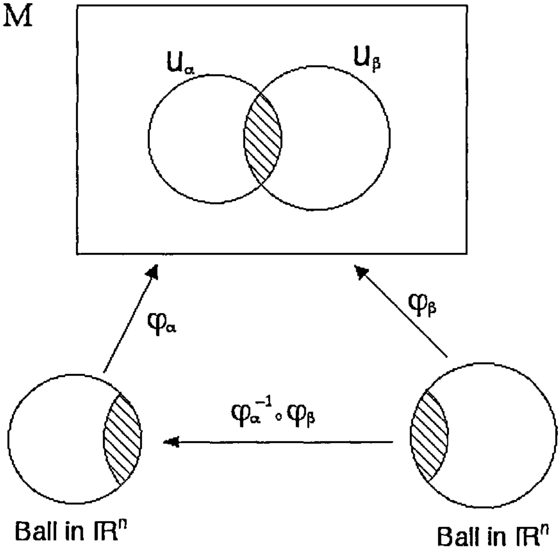

Note that $\phi_\beta^{-1}(U_\alpha \cap U_\beta)$ and $\phi_\alpha^{-1}(U_\alpha \cap U_\beta)$ are both open sets in $\mathbf{R}^n$ and thus we do know what it means for $\phi_\alpha^{-1}\phi_\beta$ to be differentiable, as discussed in Chapter Three. The idea is that we want to identify each open set U_α in M with its corresponding open ball in $\mathbf{R}^n$. In fact, if $x_1, \ldots, x_n$ are coordinates for $\mathbf{R}^n$, we can label every point p in U_α as the n-tuple given by $\phi_\alpha^{-1}(p)$. Usually people just say that we have chosen a coordinate system for U_α and identify it with the coordinates $x_1, \ldots, x_n$ for $\mathbf{R}^n$. It is this definition that motivates mathematicians to say that a manifold is anything that locally, around each point, looks like an open ball in $\mathbf{R}^n$.

Let us now show that S^1 satisfies this definition of a manifold. We will find an open cover of S^1 consisting of four open sets, for each of these write down the corresponding map ϕ_i and then see that $\phi_1^{-1}\phi_2$ is differentiable. (It is similar to show that the other $\phi_i^{-1}\phi_j$ are differentiable.)

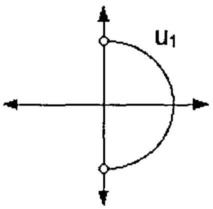

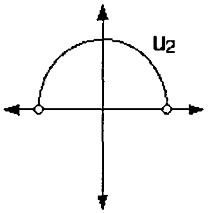

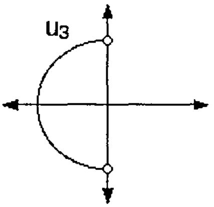

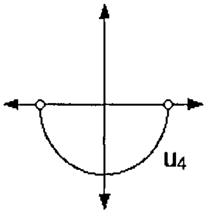

Set

$$U_1 = \{(x, y) \in S^1 : x > 0\}$$

and let

$$\phi_1 : (-1, 1) \to U_1$$

be defined by

$$\phi_1(u) = (\sqrt{1 - u^2}, u).$$

Here $(-1, 1)$ denotes the open interval $\{x : -1 < x < 1\}$. In a similar fashion, set

$$U_2 = \{(x, y) \in S^1 : y > 0\}$$

$$\begin{aligned} U_3 &= \{(x,y) \in S^1 : x < 0\} \\ U_4 &= \{(x,y) \in S^1 : y < 0\} \end{aligned}$$

and

$$\begin{aligned} \phi_2(u) &= (u, \sqrt{1-u^2}) \\ \phi_3(u) &= (-\sqrt{1-u^2}, u) \\ \phi_4(u) &= (u, -\sqrt{1-u^2}). \end{aligned}$$

Now to show on the appropriate domain that $\phi_1^{-1}\phi_2$ is differentiable. We have

$$\phi_1^{-1}\phi_2(u) = \phi_1^{-1}(u, \sqrt{1-u^2}) = \sqrt{1-u^2}$$

which is indeed differentiable for $-1 < u < 1$. (The other verifications are just as straightforward.)

We can now talk about what it means for a function to be differentiable on a manifold. Again, we will reduce the definition to a statement about the differentiability of a function from $\mathbf{R}^n$ to $\mathbf{R}$.

Definition 6.4.4 *A real-valued function f on a manifold M is* differentiable *if for an open cover (U_α) and maps ϕ_α :* Open ball in $\mathbf{R}^n \to U_\alpha$, *the composition function*

$$f \circ \phi_\alpha : \text{Open ball in } \mathbf{R}^n \to \mathbf{R}$$

is differentiable.

There is still one difficulty with our abstract definition of a manifold. The definition depends upon the existence of an open cover of M. Think of our open cover of the circle S^1. Certainly there are many other open covers that will also place a manifold structure on S^1, such as:

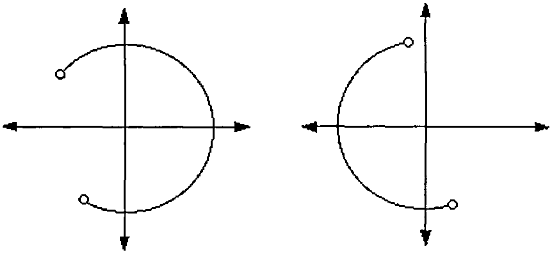

but still, it's the same circle. How can we identify these different ways of putting a manifold structure on the circle? We are led to the desire to find a natural notion of equivalence between manifolds (as we will see, we will denote this type of equivalence by saying that two manifolds are *diffeomorphic*). Before giving a definition, we need to define what it means to have a differentiable map between two manifolds. For notation, let M be an m-dimensional manifold with open cover (U_α) and corresponding maps ϕ_α and let N be an n-dimensional manifold with open cover (V_β) and corresponding maps η_β.

Definition 6.4.5 *Let $f : M \rightarrow N$ be a map from M to N. Let $p \in M$ with U_α an open set containing p. Set $q = f(p)$ and suppose that V_β is an open set containing q. Then f is differentiable at p if the map $\eta_\beta^{-1} \circ f \circ \phi_\alpha$ is differentiable in a neighborhood of the point $\phi_\alpha^{-1}(p)$ in $\mathbf{R}^m$. The map f is differentiable if it is differentiable at all points.*

We can now define our notion of equivalence.

Definition 6.4.6 *Two manifolds M and N are diffeomorphic if there exists a map $f : M \rightarrow N$ that is one-to-one, onto, differentiable and such that the inverse map, f^{-1}, is differentiable.*

Finally, by replacing the requirement that the various functions involved are differentiable by continuous functions, analytic functions, etc., we can define *continuous* manifolds, *analytic* manifolds, etc.

6.5 Tangent Spaces and Orientations

Before showing how to integrate differential k-forms along a k-dimensional manifold, we have to tackle the entirely messy issue of orientability. But before we can define orientability, we must define the tangent space to a manifold. If we use the implicit or parametric definition for a manifold, this will be straightforward. The definition for an abstract manifold is quite a bit more complicated (but as with most good abstractions, it is ultimately the right way to think about tangent vectors).

6.5.1 Tangent Spaces for Implicit and Parametric Manifolds

Let M be an implicitly defined manifold in $\mathbf{R}^n$ of dimension k. Then by definition, for each point $p \in M$ there is an open set U containing p and $(n-k)$ real-valued functions $\rho_1, \ldots, \rho_{n-k}$ defined on U such that

$$(\rho_1 = 0) \cap \ldots \cap (\rho_{n-k} = 0) = M \cap U$$

and, at every point $q \in M \cap U$, the vectors

$$\nabla \rho_1(q), \ldots, \nabla \rho_{n-k}(q)$$

are linearly independent. We have

Definition 6.5.1 *The* normal space $N_p(M)$ *to* M *at the point* p *is the vector space spanned by the vectors*

$$\nabla\rho_1(p), \ldots, \nabla\rho_{n-k}(p).$$

The tangent space $T_p(M)$ *to the manifold* M *at the point* p *consists of all vectors* $\mathbf{v}$ *in* $\mathbf{R}^n$ *that are perpendicular to each of the normal vectors.*

If $x_1, \ldots, x_n$ are the standard coordinates for $\mathbf{R}^n$, we have

Lemma 6.5.1 *A vector* $\mathbf{v} = (v_1, \ldots, v_n)$ *is in the tangent space* $T_p(M)$ *if for all* $i = 1, \ldots, n-k$ *we have*

$$0 = \mathbf{v} \cdot \nabla\rho_i(p) = \sum_{j=1}^{n} \frac{\partial\rho_i(p)}{\partial x_j} v_j.$$

The definition for the tangent space for parametrically defined manifolds is as straightforward. Here the Jacobian of the parametrizing map will be key. Let M be a manifold in $\mathbf{R}^n$, with the parametrizing map

$$\phi : (\text{Ball in } \mathbf{R}^k) \to \mathbf{R}^n$$

given by the n functions

$$\phi = (\phi_1, \ldots, \phi_n).$$

The Jacobian for ϕ is the $n \times k$ matrix

$$D\phi = \begin{pmatrix} \frac{\partial\phi_1}{\partial u_1} & \cdots & \frac{\partial\phi_1}{\partial u_k} \\ \vdots & & \vdots \\ \frac{\partial\phi_n}{\partial u_1} & \cdots & \frac{\partial\phi_n}{\partial u_k} \end{pmatrix}.$$

Definition 6.5.2 *The* tangent space $T_p(M)$ *for* M *at the point* p *is spanned by the columns of the matrix* $D\phi$.

The equivalence of these two approaches can, of course, be shown.

6.5.2 Tangent Spaces for Abstract Manifolds

Both implicitly and parametrically defined manifolds live in an ambient $\mathbf{R}^n$, which carries with it a natural vector space structure. In particular, there is a natural notion for vectors in $\mathbf{R}^n$ to be perpendicular. We used this ambient space to define tangent spaces. Unfortunately, no such ambient

$\mathbf{R}^n$ exists for an abstract manifold. What we do know is what it means for a real-valued function to be differentiable.

In calculus, we learn about differentiation as a tool to both find tangent lines and also to compute rates of change of functions. Here we concentrate on the derivative as a rate of change. Consider three-space, $\mathbf{R}^3$, with the three partial derivatives $\frac{\partial}{\partial x}, \frac{\partial}{\partial y}$ and $\frac{\partial}{\partial z}$. Each corresponds to a tangent direction for $\mathbf{R}^3$ but each also gives a method for measuring how fast a function $f(x, y, z)$ is changing, i.e.,

$$\frac{\partial f}{\partial x} = \text{how fast f is changing in the x-direction,}$$

$$\frac{\partial f}{\partial y} = \text{how fast f is changing in the y-direction}$$

and

$$\frac{\partial f}{\partial z} = \text{how fast f is changing in the z-direction.}$$

This is how we are going to define tangent vectors on an abstract manifold, as rates of change for functions. We will abstract out the algebraic properties of derivatives (namely that they are linear and satisfy Leibniz's rule).

But we have to look at differentiable functions on M a bit more closely. If we want to take the derivative of a function f at a point p, we want this to measure the rate of change of f at p. This should only involve the values of f near p. What values f achieves away from p should be irrelevant. This is the motivation behind the following equivalence relation. Let (f, U) denote an open set on M containing p and a differentiable function f defined on U. We will say that

$$(f, U) \sim (g, V)$$

if, on the open set $U \cap V$, we have $f = g$. This leads us to defining

$$C_p^\infty = \{(f, U)\}/\sim.$$

We will frequently abuse notation and denote an element of C_p^∞ by f. The space C_p^∞ is a vector space and captures the properties of functions close to the point p. (For mathematical culture sake, C_p^∞ is an example of a germ of a sheaf, in this case the sheaf of differentiable functions.)

Definition 6.5.3 *The* tangent space $T_p(M)$ *is the space of all linear maps*

$$v : C_p^\infty \to C_p^\infty$$

such that

$$v(fg) = fv(g) + gv(f).$$

To finish the story, we would need to show that this definition agrees with the other two, but this we leave as nontrivial exercises.

6.5.3 Orientation of a Vector Space

Our goal is to see that there are two possible orientations for any given vector space V. Our method is to set up an equivalence relation on the possible bases for V and see that there are only two equivalence classes, each of which we will call an orientation.

Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and $\mathbf{w}_1, \ldots, \mathbf{w}_n$ be two bases for V. Then there exists unique real numbers a_{ij}, with $i, j = 1, \ldots, n$ such that

$$\begin{aligned}
\mathbf{w}_1 &= a_{11}\mathbf{v}_1 + \cdots + a_{1n}\mathbf{v}_n \\
&\vdots \\
\mathbf{w}_n &= a_{n1}\mathbf{v}_1 + \cdots + a_{nn}\mathbf{v}_n.
\end{aligned}$$

Label the $n \times n$ matrix (a_{ij}) by A. Then we know that $\det(A) \neq 0$. We say that the bases $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and $\mathbf{w}_1, \ldots, \mathbf{w}_n$ have the same orientation if $\det(A) > 0$. If $\det(A) < 0$, then we say that they two bases have opposite orientation. It can be shown via matrix multiplication that

Lemma 6.5.2 *Having the same orientation is an equivalence relation on the set of bases for a vector space.*

The intuition is that two bases $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and $\mathbf{w}_1, \ldots, \mathbf{w}_n$ should have the same orientation if we can continuously move the basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ to $\mathbf{w}_1, \ldots, \mathbf{w}_n$ so that at each step we still have a basis. In pictures for $\mathbf{R}^2$, the bases $\{(1,0),(0,1)\}$ and $\{(1,1),(-1,1)\}$ have the same orientation but different from the basis $\{(-1,0),(0,1)\}$.

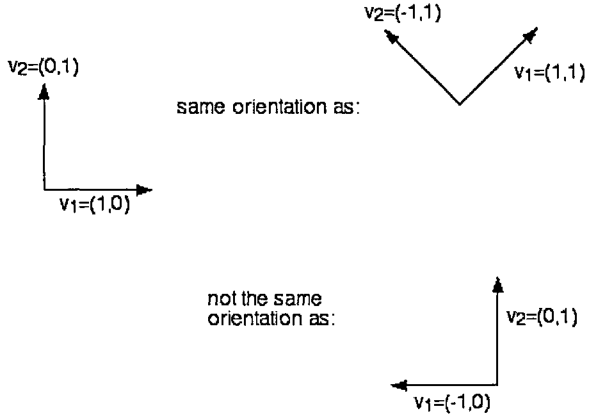

Choosing an orientation for a vector space means choosing one of the two possible orientations, i.e., choosing some basis.

6.5.4 Orientation of a Manifold and its Boundary

A manifold M has an orientation if we can choose a smoothly varying orientation for each tangent space $T_p(M)$. We ignore the technicalities of what 'smoothly varying' means, but the idea is that we can move our basis in a smooth manner from point to point on the manifold M.

Now let X^o be an open connected set in our oriented manifold M such that if X denotes the closure of X^o, then the boundary $\partial(X) = X - X^o$ is a smooth manifold of dimension one less than M. For example, if $M = \mathbf{R}^2$, an example of an X^o could be the open unit disc

$$D = \{(x, y) : x^2 + y^2 < 1\}.$$

Then the boundary of D is the unit circle

$$S^1 = \{(x, y) : x^2 + y^2 = 1\},$$

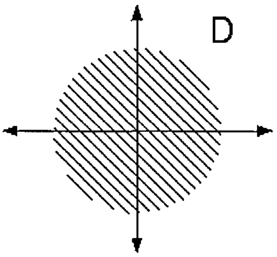

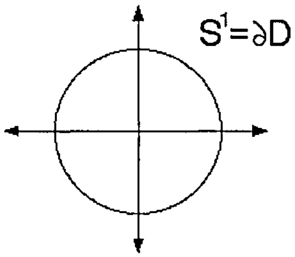

which is a one-dimensional manifold. The open set X^o inherits an orientation from the ambient manifold M. Our goal is to show that the boundary $\partial(X)$ has a canonical orientation. Let $p \in \partial(X)$. Since $\partial(X)$ has dimension one less than M, the normal space at p has dimension one. Choose a normal direction n that points out of X, not into X. The vector n, while normal to $\partial(X)$, is a tangent vector to M. Choose a basis $v_1, \ldots, v_{n-1}$ for $T_p(\partial(X))$ so that the basis $n, v_1, \ldots, v_{n-1}$ agrees with the orientation of M. It can be shown that all such chosen bases for $T_p(\partial(X))$ have the same orientation; thus the choice of the vectors $v_1, \ldots, v_{n-1}$ determines an orientation on the boundary manifold $\partial(X)$.

For example, let $M = \mathbf{R}^2$. At each point of $\mathbf{R}^2$, choose the basis $\{(1, 0), (0, 1)\}$.

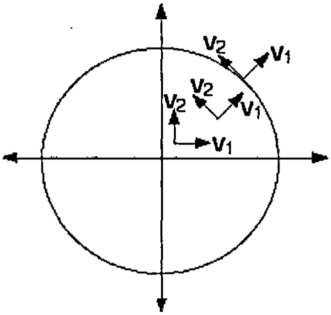

For the unit circle S^1, an outward pointing normal is always, at each point $p = (x, y)$, just the vector (x, y). Then the tangent vector $(-y, x)$ will give us a basis for $\mathbf{R}^2$ that has the same orientation as the given one. Thus we have a natural choice of orientation for the boundary manifold.

6.6 Integration on Manifolds

The goal of this section is to make sense out of the symbol

$$\int_M \omega,$$

where M will be a k-dimensional manifold and ω will be a differential k-form. Thus we want to (finally) show that differential k-forms are the things that will integrate along k-dimensional manifolds. The method will be to reduce all calculations to doing multiple integrals on $\mathbf{R}^k$, which we know how to do.

We will first look carefully at the case of 1-forms on $\mathbf{R}^2$. Our manifolds will be 1-dimensional and hence curves. Let C be a curve in the plane $\mathbf{R}^2$ that is parametrized by the map:

$$\sigma : [a, b] \to \mathbf{R}^2,$$

with

$$\sigma(u) = (x(u), y(u)).$$

If $f(x, y)$ is a continuous function defined on $\mathbf{R}^2$, then define the path integral, $\int_C f(x, y)\mathrm{d}x$, by the formula

$$\int_C f(x, y)\mathrm{d}x = \int_a^b f(x(u), y(u))\frac{\mathrm{d}x}{\mathrm{d}u}\mathrm{d}u.$$

Note that the second integral is just a one-variable integral over an interval on the real line. Likewise, the symbol $\int_C f(x,y)\mathrm{d}y$ is interpreted as

$$\int_C f(x,y)\mathrm{d}y = \int_a^b f(x(u),y(u))\frac{\mathrm{d}y}{\mathrm{d}u}\mathrm{d}u.$$

Using the chain rule, it can be checked that the numbers $\int_C f(x,y)\mathrm{d}x$ and $\int_C f(x,y)\mathrm{d}y$ are independent of the chosen parametrizations. Both of these are highly suggestive, as at least formally $f(x,y)\mathrm{d}x$ and $f(x,y)\mathrm{d}y$ look like differential 1-forms on the plane $\mathbf{R}^2$. Consider the Jacobian of the parametrizing map $\sigma(u)$, which is the 2×1 matrix

$$D\sigma = \begin{pmatrix} \mathrm{d}x/\mathrm{d}u \\ \mathrm{d}y/\mathrm{d}u \end{pmatrix}.$$

Letting $f(x,y)\mathrm{d}x$ and $f(x,y)\mathrm{d}y$ be differential 1-forms, we have by definition that at each point of $\sigma(u)$,

$$f(x,y)\mathrm{d}x(D\sigma) = f(x,y)\mathrm{d}x(\begin{pmatrix} \mathrm{d}x/\mathrm{d}u \\ \mathrm{d}y/\mathrm{d}u \end{pmatrix}) = f(x(u),y(u))\frac{\mathrm{d}x}{\mathrm{d}u}$$

and

$$f(x,y)\mathrm{d}y(D\sigma) = f(x,y)\mathrm{d}y(\begin{pmatrix} \mathrm{d}x/\mathrm{d}u \\ \mathrm{d}y/\mathrm{d}u \end{pmatrix}) = f(x(u),y(u))\frac{\mathrm{d}y}{\mathrm{d}u}.$$

Thus we could write the integrals $\int_C f(x,y)\mathrm{d}x$ and $\int_C f(x,y)\mathrm{d}y$ as

$$\int_C f(x,y)\mathrm{d}x = \int_a^b f(x,y)\mathrm{d}x(D\sigma)\mathrm{d}u$$

and

$$\int_C f(x,y)\mathrm{d}y = \int_a^b f(x,y)\mathrm{d}y(D\sigma)\mathrm{d}u.$$

This suggests how to define in general $\int_M \omega$. We will use that ω, as a k-form, will send any $n \times k$ matrix to a real number. We will parametrize our manifold M and take ω of the Jacobian of the parametrizing map.

Definition 6.6.1 *Let M be a k-dimensional oriented differentiable manifold in $\mathbf{R}^n$ such that there is a parametrizing one-to-one onto map*

$$\phi : B \to M$$

where B denotes the unit ball in $\mathbf{R}^k$. Suppose further that the parametrizing map agrees with the orientation of the manifold M. Let ω be a differential k-form on $\mathbf{R}^n$. Then

$$\int_M \omega = \int_B \omega(D\phi)\mathrm{d}u_1 \cdots \mathrm{d}u_k.$$

Via a chain rule calculation, we can show that $\int_M \omega$ is well-defined:

Lemma 6.6.1 *Given two orientation preserving parametrizations ϕ_1 and ϕ_1 of a k-dimensional manifold M, we have*

$$\int_B \omega(D\phi_1)\mathrm{d}u_1\cdots\mathrm{d}u_k = \int_B \omega(D\phi_2)\mathrm{d}u_1\cdots\mathrm{d}u_k.$$

Thus $\int_M \omega$ is independent of parametrization.

We now know what $\int_M \omega$ means for a manifold that is the image of a differentiable one-to-one onto map from a ball in $\mathbf{R}^k$. Not all manifolds can be written as the image of a single parametrizing map. For example, the unit sphere S^2 in $\mathbf{R}^3$ needs at least two such maps (basically to cover both the north and south poles). But we can (almost) cover reasonable oriented manifolds by a countable collection of non-overlapping parametrizations. More precisely, we can find a collection $\{U_\alpha\}$ of nonoverlapping open sets in M such that for each α there exists a parametrizing orientation preserving map

$$\phi_\alpha : B \to U_\alpha$$

and such that the space $M - \bigcup U_\alpha$ has dimension strictly smaller than k. Then for any differential k-form we set

$$\int_M \omega = \sum_\alpha \int_{U_\alpha} \omega.$$

Of course, this definition seems to depend on our choice of open sets, but we can show (though we choose not to) that:

Lemma 6.6.2 *The value of $\int_M \omega$ is independent of choice of set $\{U_\alpha\}$.*

While in practice the above summation could be infinite, in which case questions of convergence must arise, in practice this is rarely a problem.

6.7 Stokes' Theorem

We now come to the goal of this chapter:

Theorem 6.7.1 (Stokes' Theorem) *Let M be an oriented k-dimensional manifold in $\mathbf{R}^n$ with boundary ∂M, a smooth (k-1)-dimensional manifold with orientation induced from the orientation of M. Let ω be a differential (k-1)-form. Then*

$$\int_M \mathrm{d}\omega = \int_{\partial M} \omega.$$

This is a sharp quantitative version of the intuition:

Average of a function on boundary = Average of derivative on interior.

This single theorem includes as special cases the classical results of the Divergence Theorem, Green's Theorem and the vector-calculus Stokes' Theorem.

We will explicitly prove Stokes' Theorem only in the special case that M is a unit cube in $\mathbf{R}^k$ and when

$$\omega = f(x_1, \dots, x_k)\mathrm{d}x_2 \wedge \dots \wedge \mathrm{d}x_k.$$

After proving this special case, we will sketch the main ideas behind the proof for the general case.

Proof in unit cube case: Here

$$M = \{(x_1, \dots, x_k) : \text{ for each } i, 0 \leq x_i \leq 1\}.$$

The boundary ∂M of this cube consists of $2k$ unit cubes in $\mathbf{R}^{k-1}$. We will be concerned with the two boundary components

$$S_1 = \{(0, x_2, \dots, x_k) \in M\}$$

and

$$S_2 = \{(1, x_2, \dots, x_k) \in M\}.$$

For $\omega = f(x_1, \dots, x_k)\mathrm{d}x_2 \wedge \dots \wedge \mathrm{d}x_k$, we have

$$\begin{aligned} \mathrm{d}\omega &= \sum \frac{\partial f}{\partial x_i}\mathrm{d}x_i \wedge \mathrm{d}x_2 \wedge \dots \wedge \mathrm{d}x_k, \\ &= \frac{\partial f}{\partial x_1}\mathrm{d}x_1 \wedge dx_2 \wedge \dots \wedge \mathrm{d}x_k, \end{aligned}$$

since it is always the case that $\mathrm{d}x_j \wedge \mathrm{d}x_j = 0$.

Now to integrate $\mathrm{d}\omega$ along the unit cube M. We choose our orientation preserving parametrizing map to be the identity map. Then

$$\int_M \mathrm{d}\omega = \int_0^1 \cdots \int_0^1 \frac{\partial f}{\partial x_1}\mathrm{d}x_1 \cdots \mathrm{d}x_k.$$

By the Fundamental Theorem of Calculus we can do the first integral, to get

$$\begin{aligned} \int_M \mathrm{d}\omega &= \int_0^1 \cdots \int_0^1 f(1, x_2, \dots, x_k)\mathrm{d}x_2 \cdots \mathrm{d}x_k \\ &\quad - \int_0^1 \cdots \int_0^1 f(0, x_2, \dots, x_k)\mathrm{d}x_2 \cdots \mathrm{d}x_k. \end{aligned}$$

Now to look at the integral $\int_{\partial M} \omega$. Since $\omega = f(x_1, \dots, x_k) \mathrm{d}x_2 \wedge \dots \wedge \mathrm{d}x_k$, the only parts of the integral along the boundary that will not be zero will be along S_1 and S_2, both of which are unit cubes in $\mathbf{R}^{k-1}$, with coordinates given by $x_2, \dots, x_k$. They will have opposite orientations though. (This can be seen in the example for when M is a square in the plane; then S_1 is the bottom of the square and S_2 is the top of the square. Note how the orientations on S_1 and S_2 induced from the the orientation of the square are indeed opposite.)

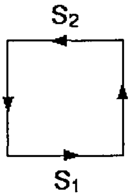

Then

$$\begin{aligned}\int_{\partial M} \omega &= \int_{C_1} \omega + \int_{C_2} \omega \\ &= \int_0^1 \cdots \int_0^1 -f(0, x_2, \dots, x_k) \mathrm{d}x_2 \cdots \mathrm{d}x_k \\ &\quad + \int_0^1 \cdots \int_0^1 f(1, x_2, \dots, x_k) \mathrm{d}x_2 \cdots \mathrm{d}x_k,\end{aligned}$$

which we have just shown to equal to $\int_M d\omega$, as desired. □

Now to sketch a false general proof for a manifold M in $\mathbf{R}^n$. We will use that the above argument for a unit cube can be used in a similar fashion for any cube. Also, any general differential $(k-1)$-form will look like:

$$\omega = \sum f_I \mathrm{d}x_I,$$

where each I is a (k-1)-tuple from $(1, \dots, n)$.

Divide M into many small cubes. Adjacent cubes' boundaries will have opposite orientation.

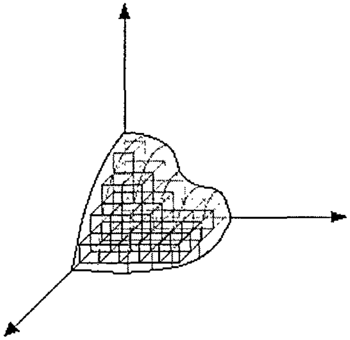

Then

$$\begin{aligned}\int_M d\omega &\approx \text{Sum over the cubes} \int_{\text{little cube}} d\omega \\ &= \text{Sum over the cubes} \int_{\partial(\text{little cube})} \omega \\ &\approx \int_{\partial(M)} \omega.\end{aligned}$$

The last approximation is from the fact that since the adjacent boundaries of the cubes have opposite orientations, they will cancel out. The only boundary parts that remain are those pushed out against the boundary of M itself. The final step would be to show that as we take more and more little cubes, we can replace the above approximations by equalities.

It must be noted that M cannot be split up into this union of cubes. Working around this difficultly is non-trivial.

6.8 Books

An excellent recent book is Hubbard and Hubbard's *Vector Calculus, Linear Algebra, and Differential Forms: A Unified Approach* [64], which contains a wealth of information, putting differential forms in the context of classical vector calculus and linear algebra. Spivak's *Calculus on Manifolds* [103] is for many people the best source. It is short and concise (in many ways the opposite of Spivak's leisurely presentation of ϵ and δ real analysis in [102]). Spivak emphasizes that the mathematical work should be done in getting the right definitions so that the theorems (Stokes' Theorem in particular) follow easily. Its briefness, though, makes it possibly not the best introduction. Fleming's *Functions of Several Variables* [37] is also a good introduction.

6.9 Exercises

1. Justify why it is reasonable for shuffles to indeed be called shuffles. (Think in terms of shuffling a deck of cards.)

2. In $\mathbf{R}^3$, let $\mathrm{d}x$, $\mathrm{d}y$ and $\mathrm{d}z$ denote the three elementary 1-forms. Using the definition of the wedge product, show that

$$(\mathrm{d}x \wedge \mathrm{d}y) \wedge \mathrm{d}z = \mathrm{d}x \wedge (\mathrm{d}y \wedge \mathrm{d}z)$$

and that these are equal to the elementary 3-form $\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z$.

3. Prove that for any differential k-form ω, we have

$$\mathrm{d}(\mathrm{d}\omega) = 0.$$

4. In $\mathbf{R}^n$, let $\mathrm{d}x$ and $\mathrm{d}y$ be one-forms. Show that

$$\mathrm{d}x \wedge \mathrm{d}y = -\mathrm{d}y \wedge \mathrm{d}x.$$

5. Prove Theorem 6.3.1.

6. Show that the map

$$\omega_{n-k}(\omega_k) = T(\omega_{n-k} \wedge \omega_k),$$

with $T : \bigwedge^n \mathbf{R^n} \to \mathbf{R}$ as defined in the chapter, provides a linear map from $\bigwedge^{n-k} \mathbf{R^n}$ to the dual space $\bigwedge^k \mathbf{R^n}^*$.

7. Prove that the unit sphere S^2 in $\mathbf{R}^3$ is a two-dimensional manifold, using each of the three definitions.

8. Consider the rectangle

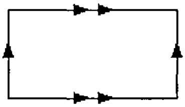

with opposite sides identified. Show first why this is a torus

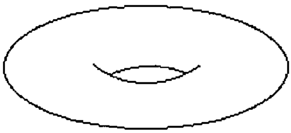

and then why it is a two-manifold.

9. The goal of this problem is to show that real projective space is a manifold. On $\mathbf{R}^{n+1} - 0$, define the equivalence relation

$$(x_0, x_1, \ldots, x_n) \sim (\lambda x_0, \lambda x_1, \ldots, \lambda x_n)$$

for any nonzero real number λ. Define real projective n-space by

$$\mathbf{P}^n = \mathbf{R}^{(n+1)} - (0)/\sim .$$

Thus, in projective three-space, we identify $(1,2,3)$ with $(2,4,6)$ and with $(-10,-20,-30)$ but not with $(2,3,1)$ or $(1,2,5)$. In $\mathbf{P}^n$, we denote the equivalence class containing $(x_0,\ldots,x_n)$ by the notation $(x_0 : \ldots : x_n)$. Thus the point in $\mathbf{P}^3$ corresponding to $(1,2,3)$ is denoted by $(1 : 2 : 3)$. Then in $\mathbf{P}^3$, we have $(1:2:3) = (2:4:6) \neq (1:2:5)$. Define

$$\phi_0 : \mathbf{R}^n \to \mathbf{P}^n$$

by

$$\phi_0(u_1,\ldots,u_n) = (1 : u_1 : \ldots : u_n),$$

define

$$\phi_1 : \mathbf{R}^n \to \mathbf{P}^n$$

by

$$\phi_1(u_1,\ldots,u_n) = (u_1 : 1 : u_2 : \ldots : u_n),$$

etc., all the way up to a defining a map ϕ_n. Show that these maps can be used to make $\mathbf{P}^n$ into an n-dimensional manifold.

10. Show that the Stokes' Theorem of this chapter has as special cases:

a. the Fundamental Theorem of Calculus. (Note that we need to use the Fundamental Theorem of Calculus to prove Stokes' Theorem; thus we cannot actually claim that the Fundamental Theorem of Calculus is a mere corollary to Stokes' Theorem.)

b. Green's Theorem.

c. the Divergence Theorem.

d. the Stokes' Theorem of Chapter Five.

Chapter 7

Curvature for Curves and Surfaces

Basic Objects:	Curves and surfaces in space
Basic Goal:	Calculating curvatures

Most of high school mathematics is concerned with straight lines and planes. There is of course far more to geometry than these flat objects. Classically differential geometry is concerned with how curves and surfaces bend and twist in space. The word "curvature" is used to denote the various measures of twisting that have been discovered.

Unfortunately, the calculations and formulas to compute the different types of curvature are quite involved and messy, but whatever curvature is, it should be the case that the curvature of a straight line and of a plane must be zero, that the curvature of a circle (and of a sphere) of radius r should be the same at every point and that the curvature of a small radius circle (or sphere) should be greater than the curvature of a larger radius circle (or sphere) (which captures the idea that it is easier to balance on the surface of the earth than on a bowling ball).

The first introduction to curvature-type ideas is usually in calculus. While the first derivative gives us tangent line (and thus linear) information, it is the second derivative that measures concavity, a curvature-type measurement. Thus we should expect to see second derivatives in curvature calculations.

7.1 Plane Curves

We will describe a plane curve via a parametrization:

$$r(t) = (x(t), y(t))$$

and thus as a map

$$r : \mathbf{R} \to \mathbf{R}^2.$$

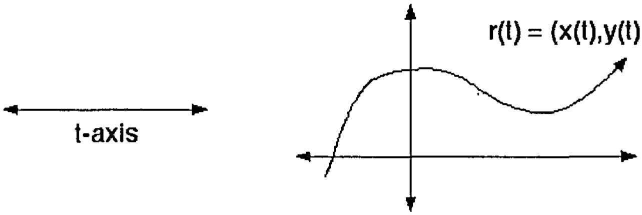

The variable t is called the *parameter* (and is frequently thought of as time). An actual plane curve can be parametrized in many different ways. For example,

$$r_1(t) = (\cos(t), \sin(t))$$

and

$$r_2(t) = (\cos(2t), \sin(2t))$$

both describe a unit circle. Any calculation of curvature should be independent of the choice of parametrization. There are a couple of reasonable ways to do this, all of which can be shown to be equivalent. We will take the approach of always fixing a canonical parametrization (the arc length parametrization). This is the parametrization $r : [a, b] \to \mathbf{R}$ such that the arc length of the curve is just $b - a$. Since the arc length is

$$\int_a^b \sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}s}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}s}\right)^2}\, \mathrm{d}s,$$

we need $\sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}s}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}s}\right)^2} = 1$. Thus for the arc length parametrization, the length of the tangent vector must always be one:

$$|\mathbf{T}(s)| = \left|\frac{\mathrm{d}r}{\mathrm{d}s}\right| = \left|\left(\frac{\mathrm{d}x}{\mathrm{d}s}, \frac{\mathrm{d}y}{\mathrm{d}s}\right)\right| = \sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}s}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}s}\right)^2} = 1.$$

Back to the question of curvature. Consider a straight line

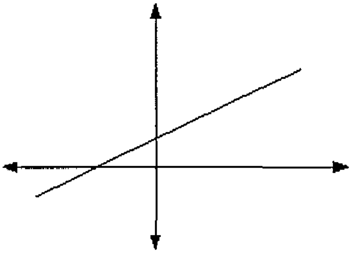

Note that each point of this line has the same tangent line.

Now consider a circle:

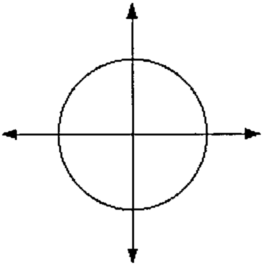

Here the tangent vectors' directions are constantly changing. This leads to the idea of trying to define curvature as a measure of the change in the direction of the tangent vectors. To measure a rate of change we need to use a derivative. This leads to:

Definition 7.1.1 *For a plane curve parametrized by arc length*

$$r(s) = (x(s), y(s)),$$

define the principal curvature κ *at a point on the curve to be the length of the derivative of the tangent vector with respect to the parameter* s*, i.e.,*

$$\kappa = \left|\frac{\mathrm{d}\mathbf{T}(s)}{\mathrm{d}s}\right|.$$

Consider the straight line $r(s) = (as + b, cs + d)$, where a, b, c and d are constants. The tangent vector is:

$$\mathbf{T}(s) = \frac{\mathrm{d}r}{\mathrm{d}s} = (a, c).$$

Then the curvature will be

$$\kappa = \left|\frac{\mathrm{d}\mathbf{T}(s)}{\mathrm{d}s}\right| = |(0,0)| = 0,$$

as desired.

Now consider a circle of radius a centered at the origin; an arc length parametrization is

$$r(s) = \left(a\cos\left(\frac{s}{a}\right), a\sin\left(\frac{s}{a}\right)\right),$$

giving us that the curvature is

$$\begin{aligned}
\kappa &= \left|\frac{\mathrm{d}\mathbf{T}(s)}{\mathrm{d}s}\right| \\
&= \left|\left(-\frac{1}{a}\cos\left(\frac{s}{a}\right), -\frac{1}{a}\sin\left(\frac{s}{a}\right)\right)\right| \\
&= \sqrt{\frac{1}{a^2}\cos^2\left(\frac{s}{a}\right) + \frac{1}{a^2}\sin^2\left(\frac{s}{a}\right)} \\
&= \frac{1}{a}.
\end{aligned}$$

Thus this definition of curvature does indeed agree with the intuitions about lines and circles that we initially desired.

7.2 Space Curves

Here the situation is more difficult; there is no single number that will capture curvature. Since we are interested in space curves, our parametrizations will have the form:

$$r(s) = (x(s), y(s), z(s)).$$

As in last section, we normalize by assuming that we have parametrized by arc length, i.e.,

$$\begin{aligned}
|\mathbf{T}(s)| &= \left|\frac{\mathrm{d}r}{\mathrm{d}s}\right| = \left|\left(\frac{\mathrm{d}x}{\mathrm{d}s}, \frac{\mathrm{d}y}{\mathrm{d}s}, \frac{\mathrm{d}z}{\mathrm{d}s}\right)\right| \\
&= \sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}s}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}s}\right)^2 + \left(\frac{\mathrm{d}z}{\mathrm{d}s}\right)^2} \\
&= 1.
\end{aligned}$$

Again we start with calculating the rate of change in the direction of the tangent vector.

Definition 7.2.1 *For a space curve parametrized by arc length*

$$r(s) = (x(s), y(s), z(s)),$$

define the principal curvature κ *at a point to be the length of the derivative of the tangent vector with respect to the parameter* s*, i.e.,*

$$\kappa = \left| \frac{\mathrm{d}\mathbf{T}(s)}{\mathrm{d}s} \right| .$$

The number κ is one of the numbers that captures curvature. Another is the torsion, but before giving its definition we need to do some preliminary work.

Set

$$\mathbf{N} = \frac{1}{\kappa}\frac{\mathrm{d}\mathbf{T}}{\mathrm{d}s}.$$

The vector $\mathbf{N}$ is called the *principal normal vector.* Note that it has length one. More importantly, as the following proposition shows, this vector is perpendicular to the tangent vector $\mathbf{T}(s)$.

Proposition 7.2.1

$$\mathbf{N} \cdot \mathbf{T} = 0$$

at all points on the space curve.

Proof: Since we are using the arc length parametrization, the length of the tangent vector is always one, which means

$$\mathbf{T} \cdot \mathbf{T} = 1.$$

Thus

$$\frac{\mathrm{d}}{\mathrm{d}s}(\mathbf{T} \cdot \mathbf{T}) = \frac{\mathrm{d}}{\mathrm{d}s}(1) = 0.$$

By the product rule we have

$$\frac{\mathrm{d}}{\mathrm{d}s}(\mathbf{T} \cdot \mathbf{T}) = \mathbf{T} \cdot \frac{\mathrm{d}\mathbf{T}}{\mathrm{d}s} + \frac{\mathrm{d}\mathbf{T}}{\mathrm{d}s} \cdot \mathbf{T} = 2\mathbf{T} \cdot \frac{\mathrm{d}\mathbf{T}}{\mathrm{d}s}.$$

Then

$$\mathbf{T} \cdot \frac{\mathrm{d}\mathbf{T}}{\mathrm{d}s} = 0.$$

Thus the vectors $\mathbf{T}$ and $\frac{\mathrm{d}\mathbf{T}}{\mathrm{d}s}$ are perpendicular. Since the principal normal vector $\mathbf{N}$ is a scalar multiple of the vector $\frac{\mathrm{d}\mathbf{T}}{\mathrm{d}s}$, we have our result. □

Set

$$\mathbf{B} = \mathbf{T} \times \mathbf{N},$$

a vector that is called the *binormal vector.* Since both **T** and **N** have length one, **B** must also be a unit vector. Thus at each point of the curve we have three mutually perpendicular unit vectors **T**, **N** and **B**. The torsion will be a number associated to the rate of change in the direction of the binormal **B**, but we need a proposition before the definition can be given.

Proposition 7.2.2 *The vector* $\frac{d\mathbf{B}}{ds}$ *is a scalar multiple of the principal normal vector* **N**.

Proof: We will show that $\frac{d\mathbf{B}}{ds}$ is perpendicular to both **T** and **B**, meaning that $\frac{d\mathbf{B}}{ds}$ must point in the same direction as **N**. First, since **B** has length one, by the same argument as in the previous proposition, just replacing all of the **T**s by **B**s, we get that $\frac{d\mathbf{B}}{ds} \cdot \mathbf{B} = 0$.

Now

$$\begin{aligned} \frac{d\mathbf{B}}{ds} &= \frac{d}{ds}(\mathbf{T} \times \mathbf{N}) \\ &= (\frac{d\mathbf{T}}{ds} \times \mathbf{N}) + (\mathbf{T} \times \frac{d\mathbf{N}}{ds}) \\ &= (\kappa\mathbf{N} \times \mathbf{N}) + (\mathbf{T} \times \frac{d\mathbf{N}}{ds}) \\ &= (\mathbf{T} \times \frac{d\mathbf{N}}{ds}). \end{aligned}$$

Thus $\frac{d\mathbf{B}}{ds}$ must be perpendicular to the vector **T**. □

Definition 7.2.2 *The* torsion *of a space curve is the number* τ *such that*

$$\frac{d\mathbf{B}}{ds} = -\tau\mathbf{N}.$$

We need now to have an intuitive understanding of what these two numbers mean. Basically, the torsion measures how much the space curve deviates from being a plane curve, while the principal curvature measures the curvature of the plane curve that the space curve wants to be. Consider the space curve

$$r(s) = (3\cos\left(\frac{s}{3}\right), 3\sin\left(\frac{s}{3}\right), 5),$$

which is a circle of radius three living in the plane $z = 5$. We will see that the torsion is zero. First, the tangent vector is

$$\mathbf{T}(s) = \frac{dr}{ds} = (-\sin\left(\frac{s}{3}\right), \cos\left(\frac{s}{3}\right), 0).$$

Then

$$\frac{d\mathbf{T}}{ds} = (-\frac{1}{3}\cos\left(\frac{s}{3}\right), -\frac{1}{3}\sin\left(\frac{s}{3}\right), 0),$$

which gives us that the principal curvature is $\frac{1}{3}$. The principal normal vector is

$$\mathbf{N} = \frac{1}{\kappa}\frac{\mathrm{d}\mathbf{T}}{\mathrm{d}s} = (-\cos\left(\frac{s}{3}\right), -\sin\left(\frac{s}{3}\right), 0).$$

Then the binormal is

$$\mathbf{B} = \mathbf{T} \times \mathbf{N} = (0, 0, 1),$$

and thus

$$\frac{\mathrm{d}\mathbf{B}}{\mathrm{d}s} = (0, 0, 0) = 0 \cdot \mathbf{N}.$$

The torsion is indeed zero, reflecting the fact that we are actually dealing with a plane curve disguised as a space curve.

Now consider the helix

$$r(t) = (\cos(t), \sin(t), t).$$

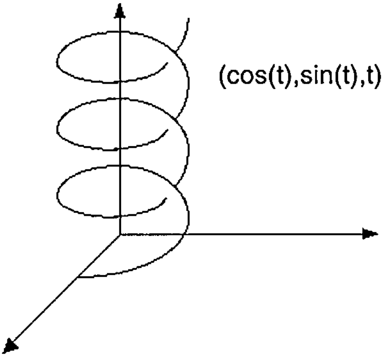

It should be the case that the principal curvature should be a positive constant, as the curve wants to be a circle. Similarly, the helix is constantly moving out of a plane, due to the t term in the z-coordinate. Hence the torsion should also be a nonzero constant. The tangent vector

$$\frac{\mathrm{d}r}{\mathrm{d}t} = (-\sin(t), \cos(t), 1)$$

does not have unique length. The arc length parametrization for this helix is simply

$$r(t) = (\cos\left(\frac{1}{\sqrt{2}}t\right), \sin\left(\frac{1}{\sqrt{2}}t\right), \frac{1}{\sqrt{2}}t).$$

Then the unit tangent vector is

$$\mathbf{T}(t) = (-\frac{1}{\sqrt{2}}\sin\left(\frac{1}{\sqrt{2}}t\right), \frac{1}{\sqrt{2}}\cos\left(\frac{1}{\sqrt{2}}t\right), \frac{1}{\sqrt{2}}).$$

The principal curvature κ is the length of the vector

$$\frac{d\mathbf{T}}{dt} = (-\frac{1}{2}\cos\left(\frac{1}{\sqrt{2}}t\right), -\frac{1}{2}\sin\left(\frac{1}{\sqrt{2}}t\right), 0).$$

Thus

$$\kappa = \frac{1}{2}.$$

Then the principal normal vector is

$$\mathbf{N}(t) = 2\frac{d\mathbf{T}}{dt} = (-\cos\left(\frac{1}{\sqrt{2}}t\right), -\sin\left(\frac{1}{\sqrt{2}}t\right), 0).$$

The binormal vector is

$$\mathbf{B} = \mathbf{T} \times \mathbf{N}$$

$$= (\frac{1}{\sqrt{2}}\sin\left(\frac{1}{\sqrt{2}}t\right), -\frac{1}{\sqrt{2}}\cos\left(\frac{1}{\sqrt{2}}t\right), \frac{1}{\sqrt{2}}).$$

The torsion τ is the length of the vector

$$\frac{d\mathbf{B}}{dt} = (\frac{1}{2}\cos\left(\frac{1}{\sqrt{2}}t\right), \frac{1}{2}\sin\left(\frac{1}{\sqrt{2}}t\right), 0),$$

and hence we have

$$\tau = \frac{1}{2}.$$

7.3 Surfaces

Measuring how tangent vectors vary worked well for understanding the curvature of space curves. A possible generalization to surfaces is to examine the variation of the tangent planes. Since the direction of a plane is determined by the direction of its normal vector, we will define curvature functions by measuring the rate of the change in the normal vector. For example, for a plane $ax + by + cz = d$, the normal at every point is the vector

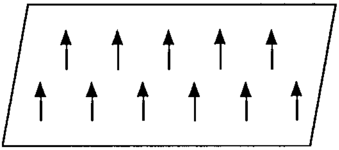

$$< a, b, c > .$$

The normal vector is a constant; there is no variation in its direction. Once we have the correct definitions in place, this should provide us with the intuitively plausible idea that since the normal is not varying, the curvature must be zero.

Denote a surface by

$$X = \{(x, y, z) : f(x, y, z) = 0\}.$$

Thus we are defining our surfaces implicitly, not parametrically. The normal vector at each point of the surface is the gradient of the defining function, i.e.,

$$\mathbf{n} = \nabla f = \langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \rangle.$$

Since we are interested in how the direction of the normal is changing and not in how the length of the normal is changing (since this length can be easily altered without varying the original surface at all), we normalize the defining function f by requiring that the normal $\mathbf{n}$ at every point has length one:

$$|\mathbf{n}| = 1.$$

We now have the following natural map:

Definition 7.3.1 *The* Gauss map *is the function*

$$\sigma : X \to S^2,$$

where S^2 *is the unit sphere in* $\mathbf{R}^2$, *defined by*

$$\sigma(p) = \mathbf{n}(p) = \nabla f = \langle \frac{\partial f}{\partial x}(p), \frac{\partial f}{\partial y}(p), \frac{\partial f}{\partial z}(p) \rangle.$$

As we move about on the surface X, the corresponding normal vector moves about on the sphere. To measure how this normal vector varies, we need to take the derivative of the vector-valued function σ and hence must look at the Jacobian of the Gauss map:

$$\mathrm{d}\sigma : TX \to TS^2,$$

where TX and TS^2 denote the respective tangent planes. If we choose orthonormal bases for both of the two dimensional vector spaces TX and TS^2, we can write $\mathrm{d}\sigma$ as a two-by-two matrix, a matrix important enough to carry its own name:

Definition 7.3.2 *The two-by-two matrix associated to the Jacobian of the Gauss map is the* Hessian.

While choosing different orthonormal bases for either TX and TS^2 will lead to a different Hessian matrix, it is the case that the eigenvalues, the trace and the determinant will remain constant (and are hence invariants of the Hessian). These invariants are what we concentrate on in studying curvature.

Definition 7.3.3 *For a surface X, the two eigenvalues of the Hessian are the* principal curvatures. *The determinant of the Hessian (equivalently the product of the principal curvatures) is the* Gaussian curvature *and the trace of the Hessian (equivalently the sum of the principal curvatures) is the* mean curvature.

We now want to see how to calculate these curvatures, in part in order to see if they agree with what our intuition demands. Luckily there is an easy algorithm that will do the trick. Start again with defining our surface X as $\{(x,y,z) : f(x,y,z) = 0\}$ such that the normal vector at each point has length one. Define the extended Hessian as

$$\tilde{H} = \begin{pmatrix} \partial^2 f/\partial x^2 & \partial^2 f/\partial x\partial y & \partial^2 f/\partial x\partial z \\ \partial^2 f/\partial x\partial y & \partial^2 f/\partial y^2 & \partial^2 f/\partial y\partial z \\ \partial^2 f/\partial x\partial z & \partial^2 f/\partial y\partial z & \partial^2 f/\partial z^2 \end{pmatrix}.$$

(Note that $\tilde{H}$ does not usually have a name.)

At a point p on X choose two orthonormal tangent vectors:

$$\begin{aligned} \mathbf{v}_1 &= a_1\frac{\partial}{\partial x} + b_1\frac{\partial}{\partial y} + c_1\frac{\partial}{\partial z} = (\,a_1 \quad b_1 \quad c_1\,) \\ \mathbf{v}_2 &= a_2\frac{\partial}{\partial x} + b_2\frac{\partial}{\partial y} + c_2\frac{\partial}{\partial z} = (\,a_2 \quad b_2 \quad c_2\,). \end{aligned}$$

Orthonormal means that we require

$$\mathbf{v}_i \cdot \mathbf{v}_j = (\,a_i \quad b_i \quad c_i\,)\begin{pmatrix} a_j \\ b_j \\ c_j \end{pmatrix} = \delta_{ij},$$

where δ_{ij} is zero for $i \neq j$ and is one for $i = j$. Set

$$h_{ij} = (\,a_i \quad b_i \quad c_i\,)\,\tilde{H}\begin{pmatrix} a_j \\ b_j \\ c_j \end{pmatrix}.$$

Then a technical argument, heavily relying on the chain rule, will yield

Proposition 7.3.1 *Coordinate systems can be chosen so that the Hessian matrix is the matrix H. Thus the principal curvatures for a surface X at a point p are the eigenvalues of the matrix*

$$H = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix}$$

and the Gaussian curvature is $\det(H)$ *and the mean curvature is trace*(H).

We can now compute some examples. Start with a plane X given by

$$(ax + by + cz - d = 0).$$

Since all of the second derivatives of the linear function $ax + by + cz - d$ are zero, the extended Hessian is the three-by-three zero matrix, which means that the Hessian is the two-by-two zero matrix, which in turn means that the principal curvatures, the Gaussian and the mean curvature are all zero, as desired.

Now suppose $X = \{(x, y, z) : \frac{1}{2r}(x^2 + y^2 + z^2 - r^2) = 0\}$, a sphere of radius r.

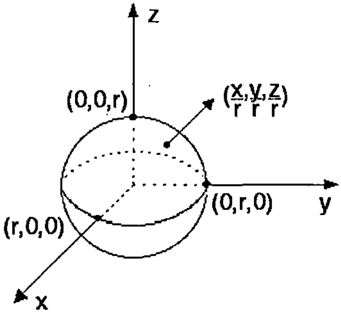

The normal is the unit vector

$$\left(\frac{x}{r} \quad \frac{y}{r} \quad \frac{z}{r} \right)$$

and the extended Hessian is

$$\tilde{H} = \begin{pmatrix} \frac{1}{r} & 0 & 0 \\ 0 & \frac{1}{r} & 0 \\ 0 & 0 & \frac{1}{r} \end{pmatrix} = \frac{1}{r} I.$$

Then given any two orthonormal vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, we have that

$$h_{ij} = \left(a_i \quad b_i \quad c_i \right) \tilde{H} \begin{pmatrix} a_j \\ b_j \\ c_j \end{pmatrix} = \frac{1}{r} \mathbf{v}_i \cdot \mathbf{v}_j,$$

and thus that the Hessian is the following diagonal matrix

$$H = \begin{pmatrix} \frac{1}{r} & 0 \\ 0 & \frac{1}{r} \end{pmatrix} = \frac{1}{r}I.$$

The two principal curvatures are both $\frac{1}{r}$ and are hence independent of which point is considered on the sphere, again agreeing with intuition.

For the final example, let X be a cylinder :

$$X = \{(x, y, z) : \frac{1}{2r}(x^2 + y^2 - r^2) = 0\}.$$

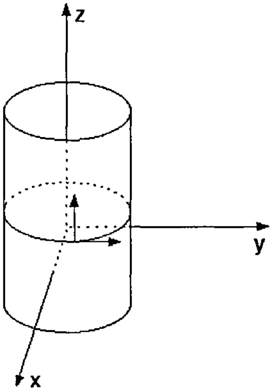

Since the intersection of this cylinder with any plane parallel to the xy plane is a circle of radius r, we should suspect that one of the principal curvatures should be the curvature of a circle, namely $\frac{1}{r}$. But also through each point on the cylinder there is a straight line parallel to the z-axis, suggesting that the other principal curvature should be zero. We can now check these guesses. The extended Hessian is

$$\tilde{H} = \begin{pmatrix} \frac{1}{r} & 0 & 0 \\ 0 & \frac{1}{r} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We can choose orthonormal tangent vectors at each point of the cylinder of the form

$$v_1 = \begin{pmatrix} a & b & 0 \end{pmatrix}$$

and

$$v_2 = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}.$$

Then the Hessian is the diagonal matrix

$$H = \begin{pmatrix} \frac{1}{r} & 0 \\ 0 & 0 \end{pmatrix},$$

meaning that one of the principal curvatures is indeed $\frac{1}{r}$ and the other is 0.

7.4 The Gauss-Bonnet Theorem

Curvature is not a topological invariant. A sphere and an ellipsoid are topologically equivalent (intuitively meaning that one can be continuously deformed into the other; technically meaning that there is a topological homeomorphism from one onto the other) but clearly the curvatures are different. But we can not alter curvature too much, or more accurately, if we make the appropriate curvature large near one point, it must be compensated for at other points. That is the essence of the Gauss-Bonnet Theorem, which we only state in this section.

We restrict our attention to compact orientable surfaces, which are topologically spheres, toruses, two-holed toruses, three-holed toruses, etc.

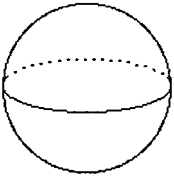

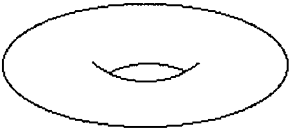

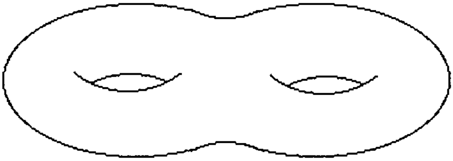

The number of holes (called the genus g) is known to be the only topological invariant, meaning that if two surfaces have the same genus, they are topologically equivalent.

Theorem 7.4.1 (Gauss-Bonnet) *For a surface X, we have*

$$\int_X \text{Gaussian curvature} = 2\pi(2-2g).$$

Thus while the Gaussian curvature is not a local topological invariant, its average value on the surface is such an invariant. Note that the left-hand side of the above equation involves analysis, while the right-hand side is topological. Equations of the form

Analysis information = Topological information

permeate modern mathematics, culminating in the Atiyah-Singer Index Formula from the mid 1960s (which has as a special case the Gauss-Bonnet Theorem). By now, it is assumed that if you have a local differential invariant, there should be a corresponding global topological invariant. The work lies in finding the correspondences.

7.5 Books

The range in texts is immense. In part this is because the differential geometry of curves and surfaces is rooted in the nineteenth century while higher dimensional differential geometry usually has quite a twentieth century feel to it. Three long time popular introductions are by do Carmo [29], Millman and Parker [85] and O'Neil [91]. A recent innovative text, emphasizing geometric intuitions is by Henderson [56]. Alfred Gray [48] has written a long book built around Mathematica, a major software package for mathematical computations. This would be a good source to see how to do actual calculations. Thorpe's text [111] is also interesting.

McLeary's *Geometry from a Differentiable Viewpoint* [84] has a lot of material in it, which is why it is also listed in the chapter on axiomatic geometry. Morgan [86] has written a short, readable account of Riemannian geometry. Then there are the classic texts. Spivak's five volumes [102] are impressive, with the first volume a solid introduction. The bible of the 1960s and 70s is *Foundations of Differential Geometry* by Kobayashi and Nomizu [74]; though fading in fashion, I would still recommend all budding differential geometers to struggle with its two volumes, but not as an introductory text.

7.6 Exercises

1. Let C be the plane curve given by $r(t) = (x(t), y(t))$. Show that the curvature at any point is

$$\kappa = \frac{x'y'' - y'x''}{((x')^2 + (y')^2)^{3/2}}.$$

(Note that the parametrization $r(t)$ is not necessarily the arc length parametrization.)

2. Let C be the plane curve given by $y = f(x)$. Show that a point $p = (x_0, y_0)$ is a point of inflection if and only if the curvature at p is zero. (Note that p is a point of inflection if $f''(x_0) = 0$.)

3. For the surface described by

$$z = x^2 + \frac{y^2}{4},$$

find the principal curvatures at each point. Sketch the surface. Does the sketch provide the same intuitions as the principal curvature calculations?

4. Consider the cone

$$z^2 = x^2 + y^2.$$

Find the image of the Gauss map. (Note that you need to make sure that the normal vector has length one.) What does this image have to say about the principal curvatures?

5. Let

$$A(t) = (a_1(t), a_2(t), a_3(t))$$

and

$$B(t) = (b_1(t), b_2(t), b_3(t))$$

be two 3-tuples of differentiable functions. Show that

$$\frac{d}{dt}(A(t) \cdot B(t)) = \frac{dA}{dt} \cdot B(t) + A(t) \cdot \frac{dB}{dt}.$$

Chapter 8

Geometry

Basic Objects:	Points and Lines in Planes
Basic Goal:	Axioms for Different Geometries

The axiomatic geometry of Euclid was the model for correct reasoning from at least as early as 300 BC to the mid 1800s. Here was a system of thought that started with basic definitions and axioms and then proceeded to prove theorem after theorem about geometry, all done without any empirical input. It was believed that Euclidean geometry correctly described the space that we live in. Pure thought seemingly told us about the physical world, which is a heady idea for mathematicians. But by the early 1800s, non-Euclidean geometries had been discovered, culminating in the early 1900s in the special and general theory of relativity, by which time it became clear that, since there are various types of geometry, the type of geometry that describes our universe is an empirical question. Pure thought can tell us the possibilities but does not appear able to pick out the correct one. (For a popular account of this development by a fine mathematician and mathematical gadfly, see Kline's *Mathematics and the Search for Knowledge* [73].)

Euclid started with basic definitions and attempted to give definitions for his terms. Today, this is viewed as a false start. An axiomatic system starts with a collection of undefined terms and a collection of relations (axioms) among these undefined terms. We can then prove theorems based on these axioms. An axiomatic system "works" if no contradictions occur. Hyperbolic and elliptic geometries were taken seriously when it was shown that any possible contradiction in them could be translated back into a contradiction in Euclidean geometry, which no one seriously believes contains a contradiction. This will be discussed in the appropriate sections of this chapter.

8.1 Euclidean Geometry

Euclid starts with twenty-three Definitions, five Postulates and five Common Notions. We will give a flavor of his language by giving a few examples of each (following Heath's translation of Euclid's *Elements* [32]; another excellent source is in Cederberg's *A Course in Modern Geometries* [17]).

For example, here is Euclid's definition of a line:

> *A* line *is breadthless length*

and for a surface:

> *A* surface *is that which has length and breadth only.*

While these definitions do agree with our intuitions of what these words should mean, to modern ears they sound vague.

His five Postulates would today be called axioms. They set up the basic assumptions for his geometry. For example, his fourth postulate states:

> *That all right angles are equal to one another.*

Finally, his five Common Notions are basic assumptions about equalities. For example, his third common notion is

> *If equals be subtracted from equals, the remainders are equal.*

All of these are straightforward, except for the infamous fifth postulate. This postulate has a different feel than the rest of Euclid's beginnings.

Fifth Postulate: *That, if a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.*

Certainly by looking at the picture

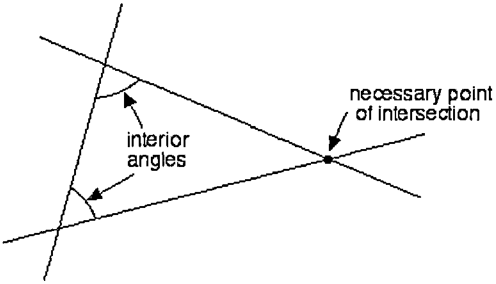

we see that this is a perfectly reasonable statement. We would be surprised if this were not true. What is troubling is that this is a basic assumption. Axioms should not be just reasonable but obvious. This is not obvious. It is also much more complicated than the other postulates, even in the superficial way that its statement requires a lot more words than the other postulates. In part, it is making an assumption about the infinite, as it states that if you extend lines further out, there will be an intersection point. A feeling of uneasiness was shared by mathematicians, starting with Euclid himself, who tried to use this postulate as little as possible.

One possible approach is to replace this postulate with another one that is more appealing, turning this troubling postulate into a theorem. There are a number of statements equivalent to the fifth postulate, but none that really do the trick. Probably the most popular is Playfair's Axiom:

Given a point off of a line, there is a unique line through the point parallel to the given line.

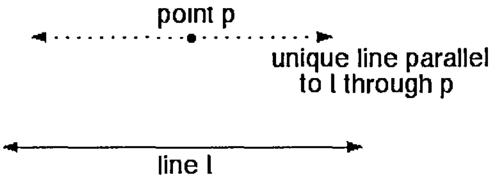

Certainly a reasonable statement. Still, it is quite bold to make this a basic assumption. It would be ideal if the fifth postulate could be shown to be a statement provable from the other axioms. The development of other geometries stemmed from the failed attempts in trying to prove the fifth postulate.

8.2 Hyperbolic Geometry

One method for showing that the fifth postulate must follow from the other axioms is to assume it is false and find a contradiction. Using Playfair's Axiom, there are two possibilities: either there are no lines through the point parallel to the given line or there are more than one line through the point parallel to the given line. These assumptions now go by the names:

Elliptic Axiom: Given a point off of a given line, there are no lines through the point parallel to the line.

This is actually just making the claim that there are no parallel lines, or that every two lines must intersect (which again seems absurd).

Hyperbolic Axiom: Given a point off of a given line, there is more than one line through the point parallel to the line.

What is meant by parallel must be clarified. Two lines are defined to be *parallel* if they do not intersect.

Geroloamo Saccheri (1667-1773) was the first to try to find a contradiction from the assumption that the fifth postulate is false. He quickly showed that if there is no such parallel line, then contradictions occurred. But when he assumed the Hyperbolic Axiom, no contradictions arose. Unfortunately for Saccheri, he thought that he had found such a contradiction and wrote a book, *Euclides ab Omni Naevo Vindicatus* (Euclid Vindicated from all Faults), that claimed to prove that Euclid was right.

Gauss (1777-1855) also thought about this problem and seems to have realized that by negating the fifth postulate, other geometries would arise. But he never mentioned this work to anybody and did not publish his results.

It was Lobatchevsky (1793-1856) and Janos Bolyai (1802-1860) who, independently, developed the first non-Euclidean geometry, now called hyperbolic geometry. Both showed, like Saccheri, that the Elliptic Axiom was not consistent with the other axioms of Euclid, and both showed, again like Saccheri, that the Hyperbolic Axiom did not appear to contradict the other axioms. Unlike Saccheri though, both confidently published their work and did not deign to find a fake contradiction.

Of course, just because you prove a lot of results and do not come up with a contradiction does not mean that a contradiction will not occur the next day. In other words, Bolyai and Lobatchevsky did not have a proof of consistency, a proof that no contradictions could ever occur. Felix Klein (1849-1925) is the main figure for finding models for different geometries that would allow for proofs of consistency, though the model we will look at was developed by Poincaré (1854-1912).

Thus the problem is how to show that a given collection of axioms forms a consistent theory, meaning that no contradiction can ever arise. The model approach will not show that hyperbolic geometry is consistent but instead show that it is as consistent as Euclidean geometry. The method is to model the straight lines of hyperbolic geometry as half circles in Euclidean geometry. Then each axiom of hyperbolic geometry will be a theorem of Euclidean geometry. The process can be reversed, so that each axiom of Euclidean geometry will become a theorem in hyperbolic geometry. Thus, if there is some hidden contradiction in hyperbolic geometry, there must also be a hidden contradiction in Euclidean geometry (a contradiction that no one believes to exist).

Now for the details of the model. Start with the upper half plane

$$H = \{(x, y) \in R^2 : y > 0\}.$$

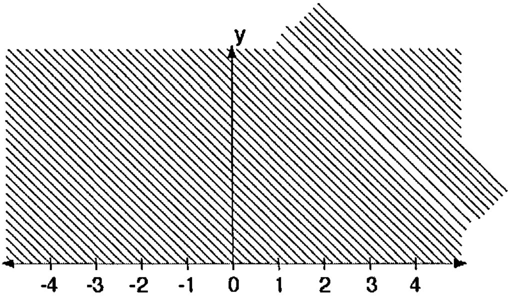

Our points will be simply the points in H. The key to our model of hyperbolic geometry is how we define straight lines. We say that a line is either a vertical line in H or a half-circle in H that intersects the x-axis perpendicularly.

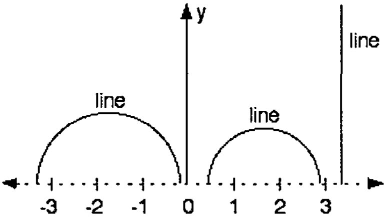

To see that this is indeed a model for hyperbolic geometry we would have to check each of the axioms. For example, we would need to check that between any two points there is a unique line (or in this case, show that for any two points in H, there is either a vertical line between them or a unique half-circle between them).

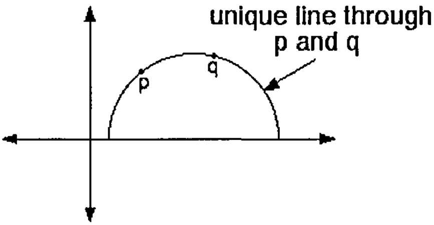

The main thing to see is that for this model the Hyperbolic Axiom is obviously true.

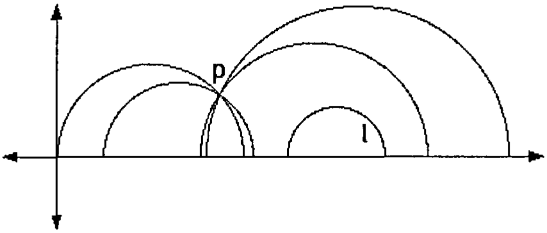

What this model allows us to do is to translate each axiom of hyperbolic geometry into a theorem in Euclidean geometry. Thus the axioms about lines in hyperbolic geometry become theorems about half-circles in Euclidean geometry. Therefore, hyperbolic geometry is as consistent as Euclidean geometry.

Further, this model shows that the fifth postulate can be assumed to be either true or false; this means that the fifth postulate is *independent* of the other axioms.

8.3 Elliptic Geometry

But what if we assume the Elliptic Axiom. Saccheri, Gauss, Bolyai and Lobatchevsky all showed that this new axiom was inconsistent with the other axioms. Could we, though, alter these other axioms to come up with another new geometry. Riemann (1826-1866) did precisely this, showing that there were two ways of altering the other axioms and thus that there were two new geometries, today called single elliptic geometry and double elliptic geometry (named by Klein). For both, Klein developed models and thus showed that both are as consistent as Euclidean geometry.

In Euclidean geometry, any two distinct points are on a unique line. Also in Euclidean geometry, a line must separate the plane, meaning that given any line l, there are at least two points off of l such that the line segment connecting the two points must intersect l.

For single elliptic geometry, we assume that *a line does not separate the plane*, in addition to the Elliptic Axiom. We keep the Euclidean assumption that any two points uniquely determine a line. For double elliptic geometry,, we need to assume that *two points can lie on more than one line*, but now keep the Euclidean assumption that a line will separate the plane. All of these sound absurd if you are thinking of straight lines as the straight lines from childhood. But under the models that Klein developed, they make sense, as we will now see.

For double elliptic geometry, our "plane" is the the unit sphere, the points are the points on the sphere and our "lines" will be the great circles on the spheres. (The *great circles* are just the circles on the sphere with greatest diameter.)

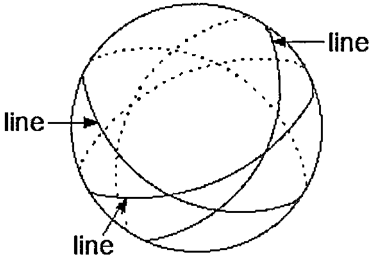

Note that any two lines will intersect (thus satisfying the Elliptic Axiom) and that while most pairs of points will uniquely define a line, points opposite to each other will lie on infinitely many lines. Thus statements about lines in double elliptic geometry will correspond to statements about great circles in Euclidean geometry.

For single elliptic geometry, the model is a touch more complicated. Our "plane" will now be the upper half-sphere, with points on the boundary circle identified with their antipodal points, i.e.,

$$\{(x,y,z) : x^2+y^2+z^2=1, z\geq 0\}/\{(x,y,0)\,\text{is identified with}\,(-x,-y,0)\}.$$

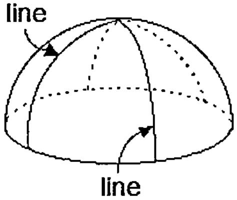

Thus the point on the boundary $(\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}},0)$ is identified with the point $(-\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}},0)$. Our "lines" will be the great half-circles on the half-sphere. Note that the Elliptic Axiom is satisfied. Further, note that no line will separate the plane, since antipodal points on the boundary are identified. Thus statements in single elliptic geometry will correspond to statements about great half-circles in Euclidean geometry.

8.4 Curvature

One of the most basic results in Euclidean geometry is that the sum of the angles of a triangle is 180 degrees, or in other words, the sum of two right angles.

Recall the proof. Given a triangle with vertices P, Q and R, by Playfair's Axiom there is a unique line through R parallel to the line spanned by P and Q. By results on alternating angles, we see that the the angles α, β and γ must sum to that of two right angles.

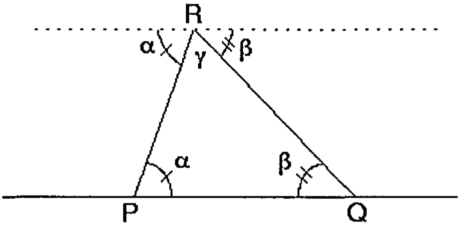

Note that we needed to use Playfair's axiom. Thus this result will not necessarily be true in non-Euclidean geometries. This seems reasonable if we look at the picture of a triangle in the hyperbolic upper half-plane and of a triangle on the sphere of double elliptic geometry.

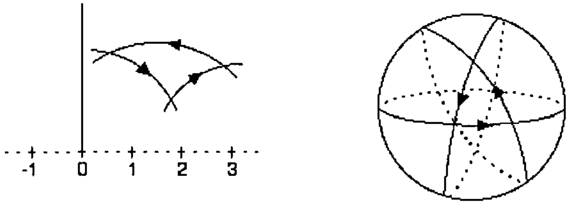

What happens is that in hyperbolic geometry the sums of the angles of a triangle are less than 180 degrees while, for elliptic geometries, the sum of the angles of a triangle will be greater than 180 degrees. It can be shown that the smaller that the area of the triangle is, the closer the sum of the triangle's angles will be to 180 degrees. This in turn is linked to the Gaussian curvature. It is the case (though it is not obvious) that methods of measuring distance (i.e., metrics) can be chosen so that the different types of geometry will have different Gaussian curvatures. More precisely, the Gaussian curvature of the Euclidean plane will be zero, of the hyperbolic plane will be -1 and of the elliptic planes will be 1. Thus differential geometry and curvature are linked to the axiomatics of different geometries.

8.5 Books

One of the best popular books in mathematics of all time is Hilbert and Cohn-Vossens' *Geometry and the Imagination* [58]. All serious students should study this book carefully. One of the 1900s best geometers (someone who actually researched in areas that nonmathematicians would recognize as geometry), Coxeter, wrote a great book, *Introduction to Geometry* [23]. More standard, straightforward texts on various types of geometry are by Gans [44], Cederberg [17] and Lang and Murrow [81] . Robin Hartshorne's *Geometry: Euclid and Beyond* [55] is an interesting recent book. Also, McLeary's *Geometry from a Differentiable Viewpoint* [84] is a place to see both non-Euclidean geometries and the beginnings of differential geometry.

8.6 Exercises

1. This problem gives another model for hyperbolic geometry. Our points will be the points in the open disc:

$$D = \{(x, y) : x^2 + y^2 < 1\}.$$

The lines will be the arcs of circles that intersect perpendicularly the boundary of D. Show that this model satisfies the Hyperbolic Axiom.

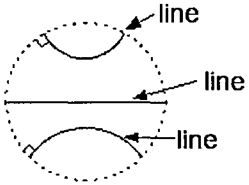

2. Show that the model in problem 1 and the upper half plane model are equivalent, if, in the upper half plane, we identify all points at infinity to a single point.
3. Give the analogue of Playfair's Axiom for planes in space.
4. Develop the idea of the upper half space so that if P is a "plane" and p is a point off of this plane, then there are infinitely many planes containing p that do not intersect the plane P.
5. Here is another model for single elliptic geometry. Start with the unit disc

$$D = \{(x, y) : x^2 + y^2 \leq 1\}.$$

Identify antipodal points on the boundary. Thus identify the point (a, b) with the point $(-a, -b)$, provided that $a^2 + b^2 = 1$. Our points will be the points of the disc, subject to this identification on the boundary.

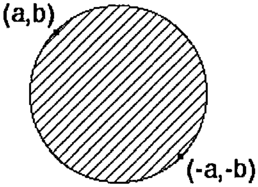

Lines will in this model be Euclidean lines, provided they start and end at antipodal points. Show that this model describes a single elliptic geometry.

6. Here is still another model for single elliptic geometry. Let our points be lines through the origin in space. Our lines in this geometry will be planes through the origin in space. (Note that two lines through the origin do indeed span a unique plane.) Show that this model describes a single elliptic geometry.

7. By looking at how a line through the origin in space intersects the top half of the unit sphere

$$\{(x,y,z) : x^2 + y^2 + z^2 = 1 \text{ and } z \geq 0\},$$

show that the model given in problem 6 is equivalent to the model for single elliptic geometry given in the text.

Chapter 9

Complex Analysis

Basic Object:	The complex numbers
Basic Map:	Analytic functions
Basic Goal:	Equivalences of analytic functions

Complex analysis in one variable studies a special type of function (called analytic or holomorphic) mapping complex numbers to themselves. There are a number of seemingly unrelated but equivalent ways for defining an analytic function. Each has its advantages; all should be known.

We will first define analyticity in terms of a limit (in direct analogy with the definition of a derivative for a real-valued function). We will then see that this limit definition can also be captured by the Cauchy-Riemann equations, an amazing set of partial differential equations. Analyticity will then be described in terms of relating the function with a particular path integral (the Cauchy Integral Formula). Even further, we will see that a function is analytic if and only if it can be locally written in terms of a convergent power series. We will then see that an analytic function, viewed as a map from $\mathbf{R}^2$ to $\mathbf{R}^2$, must preserve angles (which is what the term *conformal* means), provided that the function has a nonzero derivative. Thus our goal is:

Theorem 9.0.1 *Let $f : U \to \mathbf{C}$ be a function from an open set U of the complex numbers to the complex numbers. The function $f(z)$ is said to be* analytic *if it satisfies any of the following equivalent conditions:*

a) For all $z_0 \in U$,

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

exists. This limit is denoted by $f'(z_0)$ and is called the complex derivative.

b) The real and imaginary parts of the function f satisfy the Cauchy-Riemann equations*:*

$$\frac{\partial Re(f)}{\partial x} = \frac{\partial Im(f)}{\partial y}$$

and

$$\frac{\partial Re(f)}{\partial y} = -\frac{\partial Im(f)}{\partial x}.$$

c) Let σ be a counterclockwise simple loop in U such that every interior point of σ is also in U. If z_0 is any complex number in the interior of σ, then

$$f(z_0) = \frac{1}{2\pi i}\int_\sigma \frac{f(z)}{z - z_0}\mathrm{d}z.$$

d) For any complex number z_0, there is an open neighborhood in U of z_0 in which

$$f(z) = \sum_{n=0}^{\infty} a_n(z - z_0)^n,$$

a uniformly converging series.

Further, if f is analytic at a point z_0 and if $f'(z_0) \neq 0$, then at z_0, the function f is conformal (i.e., angle-preserving), viewed as a map from $\mathbf{R}^2$ to $\mathbf{R}^2$.

There is a basic distinction between real and complex analysis. Real analysis studies, in essence, differentiable functions; this is not a major restriction on functions at all. Complex analysis studies analytic functions; this is a major restriction on the type of functions studied, leading to the fact that analytic functions have many amazing and useful properties. Analytic functions appear throughout modern mathematics and physics, with applications ranging from the deepest properties of prime numbers to the subtlety of fluid flow. Know this subject well.

9.1 Analyticity as a Limit

For the rest of this chapter, let U denote an open set of the complex numbers $\mathbf{C}$.

Let $f : U \to \mathbf{C}$ be a function from our open set U of the complex numbers to the complex numbers.

Definition 9.1.1 *At a point $z_0 \in U$, the function $f(z)$ is* analytic *(or* holomorphic*) if*

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

exists. This limit is denoted by $f'(z_0)$ *and is called the* derivative.

Of course, this is equivalent to the limit

$$\lim_{h \to 0} \frac{f(z_0 + h) - f(z_0)}{h}$$

existing for $h \in \mathbf{C}$.

Note that this is exactly the definition for a function $f : \mathbf{R} \to \mathbf{R}$ to be differentiable if all **C**'s are replaced by **R**'s. Many basic properties of differentiable functions (such as the product rule, sum rule, quotient rule, and chain rule) will immediately apply. Hence, from this perspective, there does not appear to be anything particularly special about analytic functions. But the involved limits are not limits on the real line but limits in the real plane. This extra complexity creates profound distinctions between real differentiable functions and complex analytic ones, as we will see.

Our next task is to give an example of a nonholomorphic function. We need a little notation. The complex numbers **C** form a real two dimensional vector space. More concretely, each complex number z can be written as the sum of a real and imaginary part:

$$z = x + iy.$$

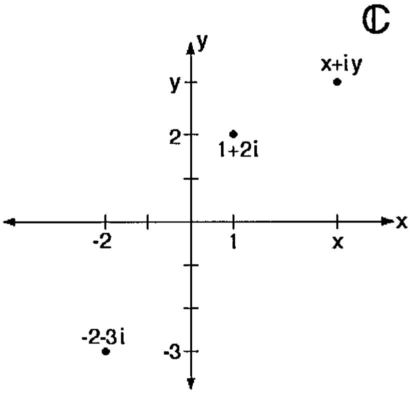

The *complex conjugate* of z is

$$\bar{z} = x - iy.$$

Note that the square of the length of the complex number z as a vector in $\mathbf{R}^2$ is

$$x^2 + y^2 = z\bar{z}.$$

Keeping in tune with this notion of length, the product $z\bar{z}$ is frequently denoted by:

$$z\bar{z} = |z|^2.$$

Fix the function

$$f(z) = \bar{z} = x - iy.$$

We will see that this function is not holomorphic. The key is that in the definition we look at the limit as $h \to 0$ but h must be allowed to be any complex number. Then we must allow h to approach 0 along any path in $\mathbf{C}$, or in other words, along any path in $\mathbf{R}^2$. We will take the limit along two different paths and see that we get two different limits, meaning that $\bar{z}$ is not holomorphic.

For convenience, let $z_0 = 0$. Let h be real valued. Then for this h we have

$$\lim_{h \to 0} \frac{f(h) - f(0)}{h - 0} = \lim_{h \to 0} \frac{h}{h} = 1.$$

Now let h be imaginary, which we label, with an abuse of notation, by hi, with h now real. Then the limit will be:

$$\lim_{hi \to 0} \frac{f(hi) - f(0)}{hi - 0} = \lim_{h \to 0} \frac{-hi}{hi} = -1.$$

Since the two limits are not equal, the function $\bar{z}$ cannot be a holomorphic function.

9.2 Cauchy-Riemann Equations

For a function $f : U \to \mathbf{C}$, we can split the image of f into its real and imaginary parts. Then, using that

$$z = x + iy = (x, y),$$

we can write $f(z) = u(z) + iv(z)$ as

$$f(x, y) = u(x, y) + iv(x, y).$$

For example, if $f(z) = z^2$, we have

$$\begin{aligned} f(z) &= z^2 \\ &= (x + iy)^2 \\ &= x^2 - y^2 + 2xyi. \end{aligned}$$

Then the real and imaginary parts of the function f will be:

$$\begin{aligned} u(x, y) &= x^2 - y^2 \\ v(x, y) &= 2xy. \end{aligned}$$

The goal of this section is to capture the analyticity of the function f by having the real-valued functions u and v satisfy a special system of partial differential equations.

Definition 9.2.1 *Real-valued functions* $u, v : U \to \mathbf{R}$ *satisfy the* Cauchy-Riemann *equations if*

$$\frac{\partial u(x,y)}{\partial x} = \frac{\partial v(x,y)}{\partial y}$$

and

$$\frac{\partial u(x,y)}{\partial y} = -\frac{\partial v(x,y)}{\partial x}.$$

Though not at all obvious, this is the most important system of partial differential equations in all of mathematics, due to its intimate connection with analyticity, described in the following theorem.

Theorem 9.2.1 *A complex-valued function* $f(x,y) = u(x,y) + iv(x,y)$ *is analytic at a point* $z_0 = x_0 + iy_0$ *if and only if the real-valued functions* $u(x,y)$ *and* $v(x,y)$ *satisfy the Cauchy-Riemann equations at* z_0.

We will show that analyticity implies the Cauchy-Riemann equations and then that the Cauchy-Riemann equations, coupled with the condition that the partial derivatives $\frac{\partial u}{\partial x}$, $\frac{\partial v}{\partial x}$, $\frac{\partial u}{\partial y}$ and $\frac{\partial v}{\partial y}$ are continuous, imply analyticity. This extra assumption requiring the continuity of the various partials is not needed, but without it the proof is quite a bit harder.

Proof: We first assume that at a point $z_0 = x_0 + iy_0$,

$$\lim_{h \to 0} \frac{f(z_0 + h) - f(z_0)}{h}$$

exists, with the limit denoted as usual by $f'(z_0)$. The key is that the number h is a complex number. Thus when we require the above limit to exist as h approaches zero, the limit must exist along any path in the plane for h approaching zero.

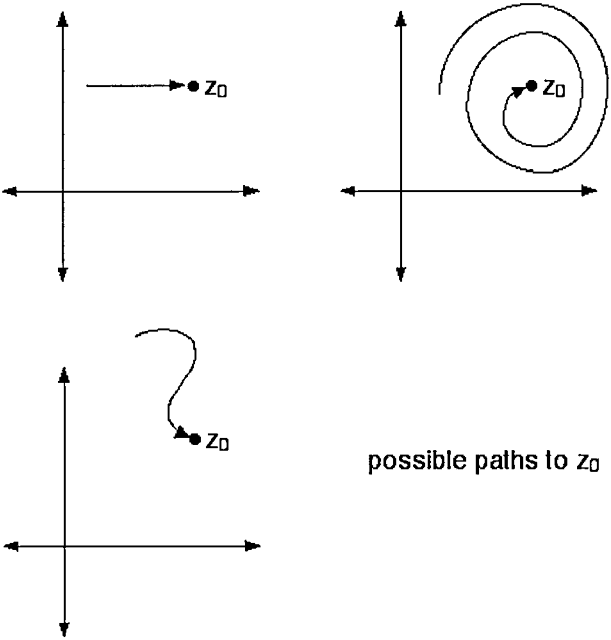

possible paths to z0

The Cauchy-Riemann equations will follow by choosing different paths for h.

First, assume that h is real. Then

$$f(z_0+h)=f(x_0+h,y)=u(x_0+h,y)+iv(x_0+h,y).$$

By the definition of analytic function,

$$\begin{aligned}
f'(z_0) &= \lim_{h\to 0}\frac{f(z_0+h)-f(z_0)}{h}\\
&= \lim_{h\to 0}\frac{u(x_0+h,y_0)+iv(x_0+h,y_0)-(u(x_0,y_0)+iv(x_0,y_0))}{h}\\
&= \lim_{h\to 0}\frac{u(x_0+h,y_0)-u(x_0,y_0)}{h}+i\lim_{h\to 0}\frac{v(x_0+h,y_0)-v(x_0,y_0)}{h}\\
&= \frac{\partial u}{\partial x}(x_0,y_0)+i\frac{\partial v}{\partial x}(x_0,y_0),
\end{aligned}$$

by the definition of partial derivatives.

Now assume that h is always purely imaginary. For ease of notation we denote h by hi, h now real. Then

$$f(z_0+hi)=f(x_0,y_0+h)=u(x_0,y_0+h)+iv(x_0,y_0+h).$$

We have, for the same complex number $f'(z_0)$ as before,

$$\begin{aligned} f'(z_0) &= \lim_{h\to 0} \frac{f(z_0+ih)-f(z_0)}{ih} \\ &= \lim_{h\to 0} \frac{u(x_0,y_0+h)+iv(x_0,y_0+h)-(u(x_0,y_0)+iv(x_0,y_0))}{ih} \\ &= \frac{1}{i}\lim_{h\to 0} \frac{u(x_0,y_0+h)-u(x_0,y_0)}{h} + \lim_{h\to 0}\frac{v(x_0,y_0+h)-v(x_0,y_0)}{h} \\ &= -i\frac{\partial u}{\partial y}(x_0,y_0)+\frac{\partial v}{\partial y}(x_0,y_0), \end{aligned}$$

by the definition of partial differentiation and since $\frac{1}{i}=-i$.

But these two limits are both equal to the same complex number $f'(z_0)$. Hence

$$\frac{\partial u}{\partial x}+i\frac{\partial v}{\partial x}=-i\frac{\partial u}{\partial y}+\frac{\partial v}{\partial y}.$$

Since $\frac{\partial u}{\partial x}, \frac{\partial v}{\partial x}, \frac{\partial u}{\partial y}$, and $\frac{\partial v}{\partial y}$ are all real-valued functions, we must have

$$\begin{aligned} \frac{\partial u}{\partial x} &= \frac{\partial v}{\partial y} \\ \frac{\partial u}{\partial y} &= -\frac{\partial v}{\partial x}, \end{aligned}$$

the Cauchy-Riemann equations.

Before we can prove that the Cauchy-Riemann equations (plus the extra assumption of continuity on the partial derivatives) imply that $f(z)$ is analytic, we need to describe how complex multiplication can be interpreted as a linear map from $\mathbf{R}^2$ to $\mathbf{R}^2$ (and hence as a 2×2 matrix).

Fix a complex number $a+bi$. Then for any other complex number $x+iy$, we have

$$(a+bi)(x+iy)=(ax-by)+i(ay+bx).$$

Representing $x+iy$ as a vector $\binom{x}{y}$ in $\mathbf{R}^2$, we see that multiplication by $a+bi$ corresponds to the matrix multiplication

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} ax-by \\ bx+ay \end{pmatrix}.$$

As can be seen, not all linear transformations $\binom{AB}{CD}:\mathbf{R}^2\to\mathbf{R}^2$ correspond to multiplication by a complex number. In fact, from the above we have

Lemma 9.2.1 *The matrix*

$$\begin{pmatrix} AB \\ CD \end{pmatrix}$$

corresponds to multiplication by a complex number $a+bi$ *if and only if* $A=D=a$ *and* $B=-C=-b$.

Now we can return to the other direction of the theorem. First write our function $f:\mathbf{C}\to\mathbf{C}$ as a map $f:\mathbf{R}^2\to\mathbf{R}^2$ by

$$f(x,y)=\begin{pmatrix}u(x,y)\\ v(x,y)\end{pmatrix}.$$

As described in Chapter Three, the Jacobian of f is the unique matrix

$$Df=\begin{pmatrix}\frac{\partial u}{\partial x}(x_0,y_0) & \frac{\partial u}{\partial y}(x_0,y_0)\\ \frac{\partial v}{\partial x}(x_0,y_0) & \frac{\partial v}{\partial y}(x_0,y_0)\end{pmatrix}$$

satisfying

$$\lim_{\frac{x}{y}\to\frac{x_0}{y_0}}\frac{\left|\begin{pmatrix}u(x,y)\\ v(x,y)\end{pmatrix}-\begin{pmatrix}u(x_0,y_0)\\ v(x_0,y_0)\end{pmatrix}-Df\cdot\begin{pmatrix}x-x_0\\ y-y_0\end{pmatrix}\right|}{|(x-x_0,y-y_0)|}=0.$$

But the Cauchy-Riemann equations, $\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y}$ and $\frac{\partial u}{\partial y}=-\frac{\partial v}{\partial x}$, tell us that this Jacobian represents multiplication by a complex number. Call this complex number $f'(z_0)$. Then, using that $z=x+iy$ and $z_0=x_0+iy_0$, we can rewrite the above limit as

$$\lim_{z\to z_0}\frac{|\,f(z)-f(z_0)-f'(z_0)(z-z_0)\,|}{|\,z-z_0\,|}=0.$$

This must also hold without the absolute value signs and hence

$$\begin{aligned}0 &= \lim_{z\to z_0}\frac{f(z)-f(z_0)-f'(z_0)(z-z_0)}{z-z_0}\\ &= \lim_{z\to z_0}\frac{f(z)-f(z_0)}{z-z_0}-f'(z_0).\end{aligned}$$

Thus

$$f'(z_0)=\lim_{z\to z_0}\frac{f(z)-f(z_0)}{z-z_0}$$

will always exist, meaning that the function $f:\mathbf{C}\to\mathbf{C}$ is analytic. □

9.3 Integral Representations of Functions

Analytic functions can also be defined in terms of path integrals about closed loops in **C**. This means that we will be writing analytic functions as integrals, which is what is meant by the term *integral representation.* We will see that for a closed loop σ,

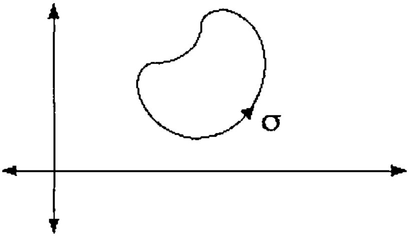

the values of an analytic function on interior points are determined from the values of the function on the boundary, which places strong restrictions on what analytic functions can be. The consequences of this integral representation of analytic functions range from the beginnings of homology theory to the calculation of difficult real-valued integrals (using residue theorems).

We first need some preliminaries on path integrals and Green's Theorem. Let σ be a path in our open set U. In other words, σ is the image of a differentiable map

$$\sigma : [0,1] \to U.$$

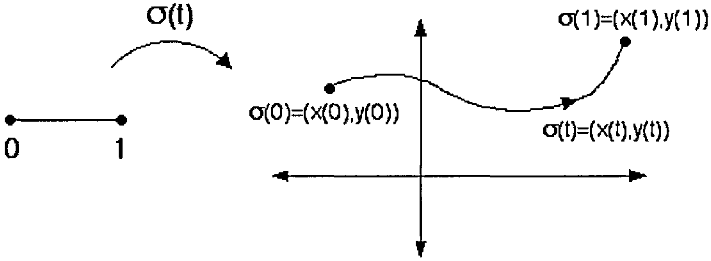

Writing $\sigma(t) = (x(t), y(t))$, with x denoting the real coordinate of **C** and y the imaginary coordinate, we have:

Definition 9.3.1 *If $P(x,y)$ and $Q(x,y)$ are real-valued functions defined on an open subset U of $\mathbf{R}^2 = \mathbf{C}$, then*

$$\int_\sigma P\mathrm{d}x + Q\mathrm{d}y = \int_0^1 P(x(t),y(t))\frac{\mathrm{d}x}{\mathrm{d}t}\mathrm{d}t + \int_0^1 Q(x(t),y(t))\frac{dy}{dt}dt.$$

If $f : U \to \mathbf{C}$ is a function written as

$$f(z) = f(x, y) = u(x, y) + iv(x, y) = u(z) + iv(z),$$

then

Definition 9.3.2 *The* path integral $\int_\sigma f(z)\mathrm{d}z$ *is defined by*

$$\begin{aligned}\int_\sigma f(z)\mathrm{d}z &= \int_\sigma \big(u(x,y) + iv(x,y)\big)(\mathrm{d}x + i\mathrm{d}y) \\ &= \int_\sigma \big(u(x,y) + iv(x,y)\big)\mathrm{d}x + \int_\sigma \big(iu(x,y) - v(x,y)\big)\mathrm{d}y.\end{aligned}$$

The goal of this section is to see that these path integrals have a number of special properties when the function f is analytic.

A path σ is a *closed loop* in U if there is a parametrization $\sigma : [0, 1] \to U$ with $\sigma(0) = \sigma(1)$.

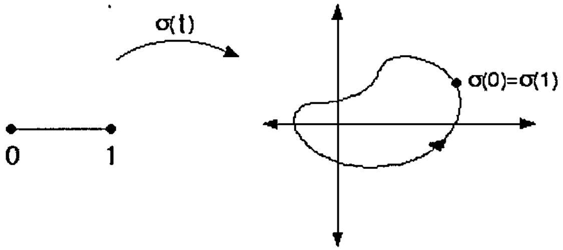

Note that we are using the same symbol for the actual path and for the parametrization function. The loop is *simple* if $\sigma(t) \neq \sigma(s)$, for all $s \neq t$, except for when t or s is zero or one.

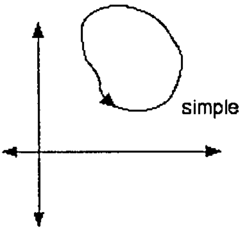

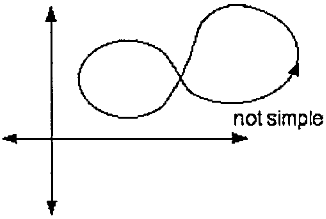

We will require all of our simple loops to be parametrized so that they are counterclockwise around their interior. For example, the unit circle is a counterclockwise simple loop, with parametrization

$$\sigma(t) = (\cos(2\pi t), \sin(2\pi t)).$$

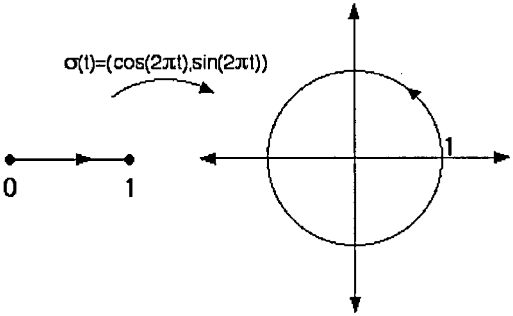

We will be interested in the path integrals of analytic functions around counterclockwise simple loops. Luckily, there are two key, easy examples that demonstrate the general results. Both of these examples will be integrals about the unit circle. Consider the function $f : \mathbf{C} \to \mathbf{C}$ defined by

$$f(z) = z = x + iy.$$

Then

$$\begin{aligned}
\int_\sigma f(z)\mathrm{d}z &= \int_\sigma z\mathrm{d}z \\
&= \int_\sigma (x+iy)(\mathrm{d}x + i\mathrm{d}y) \\
&= \int_\sigma (x+iy)\mathrm{d}x + \int_\sigma (xi - y)\mathrm{d}y \\
&= \int_0^1 (\cos(2\pi t) + i\sin(2\pi t))\frac{\mathrm{d}}{\mathrm{d}t}\cos(2\pi t)\mathrm{d}t \\
&\quad + \int_0^1 (i\cos(2\pi t) - \sin(2\pi t))\frac{\mathrm{d}}{\mathrm{d}t}\sin(2\pi t)\mathrm{d}t \\
&= 0,
\end{aligned}$$

when the integral is worked out.

On the other hand, consider the function $f(z) = \frac{1}{z}$. On the unit circle we have $|z|^2 = z\overline{z} = 1$ and hence $\frac{1}{z} = \overline{z}$. Then

$$\int_\sigma f(z)\mathrm{d}z = \int_\sigma \frac{\mathrm{d}z}{z} = \int_\sigma \overline{z}\mathrm{d}z = \int (\cos(2\pi t) - i\sin(2\pi t))(\mathrm{d}x + i\mathrm{d}y)$$

$$= 2\pi i,$$

when the calculation is performed. We will soon see that the reason that the path integral $\int_\sigma \frac{dz}{z}$ equals $2\pi i$ for the unit circle is that the function $\frac{1}{z}$ is not well-defined in the interior of the circle (namely at the origin). Otherwise the integral would be zero, as in the first example. Again, though not at all apparent, these are the two key examples.

The following theorems will show that the path integral of an analytic function about a closed loop will always be zero if the function is also analytic on the interior of the loop.

We will need, though, Green's Theorem:

Theorem 9.3.1 (Green's Theorem) *Let σ be a counterclockwise simple loop in $\mathbf{C}$ and Ω its interior. If $P(x,y)$ and $Q(x,y)$ are two real-valued differentiable functions, then*

$$\int_\sigma P\mathrm{d}x + Q\mathrm{d}y = \int\int_\Omega \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) \mathrm{d}x\mathrm{d}y.$$

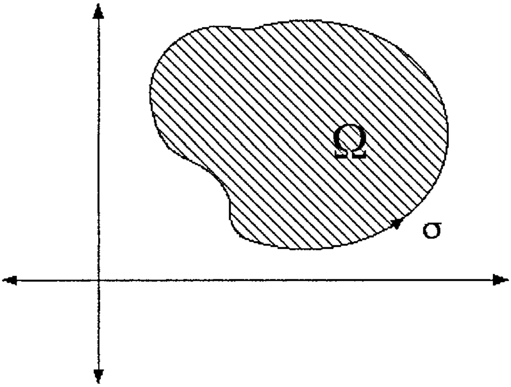

The proof is exercise 5 in Chapter Five.

Now on to Cauchy's Theorem:

Theorem 9.3.2 (Cauchy's Theorem) *Let σ be a counterclockwise simple loop in an open set U such that every point in the interior of σ is contained in U. If $f : U \to \mathbf{C}$ is an analytic function, then*

$$\int_\sigma f(z)\mathrm{d}z = 0.$$

Viewing the path integral $\int_\sigma f(z)\mathrm{d}z$ as some sort of average of the values of $f(z)$ along the loop σ, this theorem is stating the average value is zero for an analytic f. By the way, this theorem is spectacularly false for most functions, showing that those that are analytic are quite special.

Proof: (under the additional hypothesis, which can be removed with some work, that the complex derivative $f'(z)$ is continuous).

Write $f(z) = u(z) + iv(z)$, with $u(z)$ and $v(z)$ real-valued functions. Since $f(z)$ is analytic we know that the Cauchy-Riemann equations hold:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$$

and

$$-\frac{\partial u}{\partial y} = \frac{\partial v}{\partial x}.$$

Now

$$\begin{aligned} \int_\sigma f(z)\mathrm{d}z &= \int_\sigma (u+iv)(\mathrm{d}x + i\mathrm{d}y) \\ &= \int_\sigma (u\mathrm{d}x - v\mathrm{d}y) + i\int_\sigma (u\mathrm{d}y + v\mathrm{d}x) \\ &= \int\int_\Omega \left(-\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) \mathrm{d}x\mathrm{d}y + i\int\int_\Omega \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) \mathrm{d}x\mathrm{d}y, \end{aligned}$$

by Green's Theorem, where as before Ω denotes the interior of the closed loop σ. But this path integral must be zero by the Cauchy-Riemann equations. □

Note that while the actual proof of Cauchy's Theorem was short, it used two major earlier results, namely the equivalence of the Cauchy-Riemann equations with analyticity and Green's Theorem.

This theorem is at the heart of all integral-type properties for analytic functions. For example, this theorem leads (nontrivially) to the following, which we will not prove:

Theorem 9.3.3 *Let $f : U \to \mathbf{C}$ be analytic in an open set U and let σ and $\hat{\sigma}$ be two simple loops so that σ can be continuously deformed to $\hat{\sigma}$ in U (i.e., σ and $\hat{\sigma}$ are homotopic in U). Then*

$$\int_\sigma f(z)\mathrm{d}z = \int_{\hat{\sigma}} f(z)\mathrm{d}z.$$

Intuitively, two loops are homotopic in a region U if one can be continuously deformed into the other within U. Thus

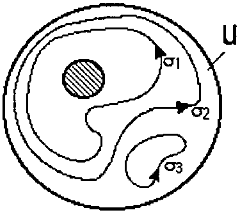

σ_1 and σ_2 are homotopic to each other in the region U but not to σ_3 in this region (though all three are homotopic to each other in **C**). The technical definition is:

Definition 9.3.3 *Two paths σ_1 and σ_2 are homotopic in a region U if there is a continuous map*

$$T : [0,1] \times [0,1] \to U$$

with

$$T(t,0) = \sigma_1(t)$$

and

$$T(t,1) = \sigma_2(t).$$

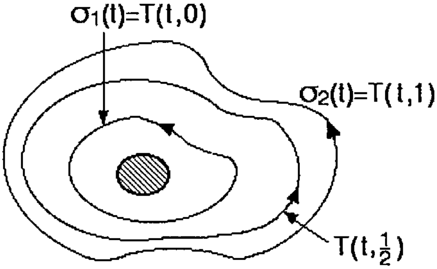

In the statement of Cauchy's Theorem, the requirement that all of the points in the interior of the closed loop σ be in the open set U can be restated as requiring that the loop σ is homotopic to a point in U.

We also need the notion of simply connected. A set U in **C** is *simply connected* if every closed loop in U is homotopic in U to a single point. Intuitively, U is simply connected if U contains the interior points of every closed loop in U. For example, the complex numbers **C** is simply connected, but **C**$-(0,0)$ is not simply connected, since **C**$-(0,0)$ does not contain the unit disc, even though it does contain the unit circle.

We will soon need the following slight generalization of Cauchy's Theorem:

Proposition 9.3.1 *Let U be a simply connected open set in* $\mathbf{C}$. *Let $f : U \to \mathbf{C}$ be analytic except possibly at a point z_0 but continuous everywhere. Let σ be any counterclockwise simple loop in U. Then*

$$\int_\sigma f(z)\mathrm{d}z = 0.$$

The proof is similar to that of Cauchy's Theorem; the extension is that we have to guarantee that all still works even if the point z_0 lies on the loop σ.

All of these lead to:

Theorem 9.3.4 (Cauchy Integral Formula) *Let $f : U \to \mathbf{C}$ be analytic on a simply connected open set U in* $\mathbf{C}$ *and let σ be a counterclockwise simple loop in U. Then for any point z_0 in the interior of σ, we have*

$$f(z_0) = \frac{1}{2\pi i}\int_\sigma \frac{f(z)}{z - z_0}\mathrm{d}z.$$

The meaning of this theorem is that the value of the analytic function f at any point in the interior of a region can be obtained by knowing the values of f on the boundary curve.

Proof: Define a new function $g(z)$ by setting

$$g(z) = \frac{f(z) - f(z_0)}{z - z_0},$$

when $z \neq z_0$ and setting

$$g(z) = f'(z_0)$$

when $z = z_0$.

Since $f(z)$ is analytic at z_0, by definition we have

$$f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0},$$

meaning that the new function $g(z)$ is continuous everywhere and analytic everywhere except for possibly at z_0.

Then by the last theorem we have $\int_\sigma g(z)\mathrm{d}z = 0$. Thus

$$0 = \int_\sigma \frac{f(z) - f(z_0)}{z - z_0}\mathrm{d}z = \int_\sigma \frac{f(z)}{z - z_0}\mathrm{d}z - \int_\sigma \frac{f(z_0)}{z - z_0}\mathrm{d}z.$$

Then

$$\begin{aligned}\int_\sigma \frac{f(z)}{z-z_0}\mathrm{d}z &= \int_\sigma \frac{f(z_0)}{z-z_0}\mathrm{d}z \\ &= f(z_0)\int_\sigma \frac{1}{z-z_0}\mathrm{d}z,\end{aligned}$$

since $f(z_0)$ is just a fixed complex number. But this path integral is just our desired $2\pi i f(z_0)$, by direct calculation, after deforming our simple loop σ to a circle centered at z_0. □

In fact, the converse is also true.

Theorem 9.3.5 *Let σ be a counterclockwise simple loop and $f : \sigma \to \mathbf{C}$ any continuous function on the loop σ. Extend the function f to the interior of the loop σ by setting*

$$f(z_0) = \frac{1}{2\pi i}\int_\sigma \frac{f(z)}{z-z_0}\mathrm{d}z$$

for points z_0 in the interior. Then $f(z)$ is analytic on the interior of σ. Further, f is infinitely differentiable with

$$f^k(z_0) = \frac{k!}{2\pi i}\int_\sigma \frac{f(z)}{(z-z_0)^{k+1}}\mathrm{d}z.$$

Though a general proof is in most books on complex analysis, we will only sketch why the derivative $f'(z_0)$ is capable of being written as the path integral

$$\frac{1}{2\pi i}\int_\sigma \frac{f(z)}{(z-z_0)^2}\mathrm{d}z.$$

For ease of notation, we write

$$f(z) = \frac{1}{2\pi i}\int_\sigma \frac{f(w)}{w-z}\mathrm{d}w.$$

Then

$$\begin{aligned}f'(z) &= \frac{\mathrm{d}}{\mathrm{d}z}f(z) \\ &= \frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{2\pi i}\int_\sigma \frac{f(w)}{w-z}\mathrm{d}w\right) \\ &= \frac{1}{2\pi i}\int_\sigma \frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{f(w)}{w-z}\right)\mathrm{d}w \\ &= \frac{1}{2\pi i}\int_\sigma \frac{f(w)}{(w-z)^2}\mathrm{d}w\end{aligned}$$

as desired.

Note that in this theorem we are not assuming that the original function $f : \sigma \to \mathbf{C}$ was analytic. In fact the theorem is saying that any continuous function on a simple loop can be used to define an analytic function on the interior. The reason that this can only be called a sketch of a proof was that we did not justify the pulling of the derivative $\frac{\mathrm{d}}{\mathrm{d}z}$ inside of the integral.

9.4 Analytic Functions as Power Series

Polynomials $a_n z^n + a_{n-1} z^{n-1} + \cdots + a_0$ are great functions to work with. In particular they are easy to differentiate and to integrate. Life would be easy if all we ever had to be concerned with were polynomials. But this is not the case. Even basic functions such as e^z, $\log(z)$ and the trig functions are just not polynomials. Luckily though, all of these functions are analytic, which we will see in this section means that they are almost polynomials, or more accurately, glorified polynomials, which go by the more common name as power series. In particular the goal of this section is to prove:

Theorem 9.4.1 *Let U be an open set in $\mathbf{C}$. A function $f : U \to \mathbf{C}$ is analytic at z_0 if and only if in a neighborhood of z_0, $f(z)$ is equal to a uniformly convergent power series, i.e.,*

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n.$$

Few functions are equal to uniformly convergent power series (these "glorified polynomials"). Thus we will be indeed showing that an analytic function can be described as such a glorified polynomial.

Note that if

$$\begin{aligned} f(z) &= \sum_{n=0}^{\infty} a_n (z - z_0)^n \\ &= a_0 + a_1(z - z_0) + a_2(z - z_0)^2 + \cdots, \end{aligned}$$

we have that

$$\begin{aligned} f(z_0) &= a_0, \\ f'(z_0) &= a_1, \\ f^{(2)}(z_0) &= 2a_2, \\ &\vdots \\ f^{(k)}(z_0) &= k! a_k. \end{aligned}$$

Thus, if $f(z) = \sum_{n=0}^{\infty} a_n(z - z_0)^n$, we have

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!}(z - z_0)^n,$$

the function's Taylor series. In other words, the above theorem is simply stating that an analytic function is equal to its Taylor series.

We first show that any uniformly convergent power series defines an analytic function by reviewing quickly some basic facts about power series and then sketching a proof.

Recall the definition of uniform convergence, given in Chapter Three.

Definition 9.4.1 *Let U be a subset of the complex numbers* $\mathbf{C}$. *A sequence of functions, $f_n : A \to \mathbf{C}$,* converges uniformly *to a function $f : U \to \mathbf{C}$ if given any $\epsilon > 0$, there is some positive integer N such that for all $n \geq N$,*

$$|f_n(z) - f(z)| < \epsilon,$$

for all points z in U.

In other words, we are guaranteed that eventually all the functions $f_n(z)$ will fall within any ϵ-tube about the limit function $f(z)$.

The importance of uniform convergence for us is the following theorem, which we will not prove here:

Theorem 9.4.2 *Let the sequence $\{f_n(z)\}$ of analytic functions converge uniformly on an open set U to a function $f : U \to \mathbf{C}$. Then the function $f(z)$ is also analytic and the sequence of derivatives $(f_n'(z))$ will converge pointwise to the derivative $f'(z)$ on the set U.*

Now that we have a definition for a sequence of functions to converge uniformly, we can make sense out of what it would mean for a series of functions to converge uniformly, via translating series statements into sequence statements using the partial sums of the series.

Definition 9.4.2 *A series $\sum_{n=0}^{\infty} a_n(z-z_0)^n$, for complex numbers a_n and z_0,* converges uniformly *in an open set U of the complex numbers* $\mathbf{C}$ *if the sequence of polynomials $\{\sum_{n=0}^{N} a_n(z - z_0)^n\}$ converges uniformly in U.*

By the above theorem and since polynomials are analytic, we can conclude that if

$$f(z) = \sum_{n=0}^{\infty} a_n(z - z_0)^n$$

is a uniformly convergent series, then the function $f(z)$ is analytic.

Now to sketch why any analytic function can be written as a uniformly convergent power series. The Cauchy Integral Formula from last section will be critical.

Start with a function f which is analytic about a point z_0. Choose a simple loop σ about z_0. By the Cauchy Integral Formula,

$$f(z) = \frac{1}{2\pi i} \int_{\sigma} \frac{f(w)}{w-z} dw,$$

for any z inside σ.

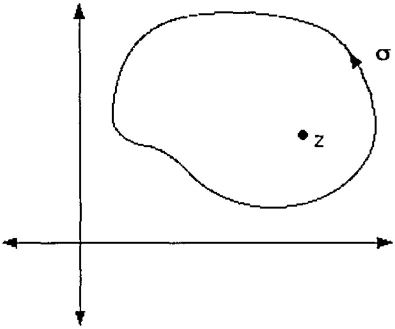

Knowing that the geometric series is

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r},$$

for $| r | <1$, we see that, for all w and z with $|z - z_0| < |w - z_0|$, we have

$$\begin{aligned} \frac{1}{w-z} &= \frac{1}{w-z_0} \cdot \frac{1}{1 - \frac{z-z_0}{w-z_0}} \\ &= \frac{1}{w-z_0} \cdot \sum_{n=0}^{\infty} \left(\frac{z-z_0}{w-z_0} \right)^n . \end{aligned}$$

Restrict the numbers w to lie on the loop σ. Then for those complex numbers z with $|z - z_0| < |w - z_0|$,

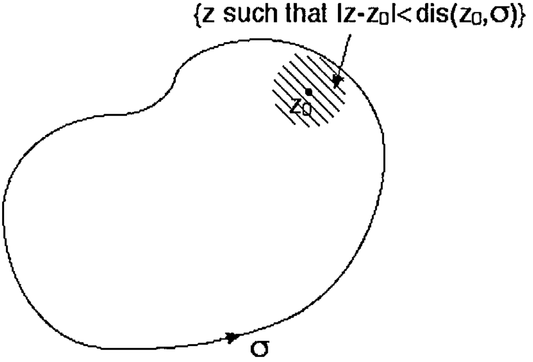

we have

$$
\begin{aligned}
f(z) &= \frac{1}{2\pi i}\int_\sigma \frac{f(w)}{w-z}\mathrm{d}w \\
&= \frac{1}{2\pi i}\int_\sigma \frac{f(w)}{w-z_0}\cdot\frac{1}{1-\frac{z-z_0}{w-z_0}}\mathrm{d}w \\
&= \frac{1}{2\pi i}\int_\sigma \frac{f(w)}{w-z_0}\sum_{n=0}^{\infty}\left(\frac{z-z_0}{w-z_0}\right)^n \mathrm{d}w \\
&= \frac{1}{2\pi i}\int_\sigma \sum_{n=0}^{\infty}\frac{f(w)}{w-z_0}\left(\frac{z-z_0}{w-z_0}\right)^n \mathrm{d}w \\
&= \frac{1}{2\pi i}\sum_{n=0}^{\infty}\int_\sigma \frac{f(w)}{w-z_0}\left(\frac{z-z_0}{w-z_0}\right)^n \mathrm{d}w \\
&= \frac{1}{2\pi i}\sum_{n=0}^{\infty}(z-z_0)^n\int_\sigma \frac{f(w)}{(w-z_0)^{n+1}}\mathrm{d}w \\
&= \sum_{n=0}^{\infty}\frac{f^{(n)}(z_0)}{n!}(z-z_0)^n,
\end{aligned}
$$

a convergent power series.

Of course the above is not quite rigorous, since we did not justify the switching of the integral with the sum. It follows, nontrivially, from the fact that the series $\sum_{n=0}^{\infty}\left(\frac{z-z_0}{w-z_0}\right)^n$ converges uniformly.

Note that we have also used the Cauchy Integral Formula, namely that

$$f^{(n)}(z_0) = \frac{n!}{2\pi i}\int_\sigma \frac{f(w)}{(w-z_0)^{n+1}}\mathrm{d}w.$$

9.5 Conformal Maps

We now want to show that analytic functions are also quite special when one looks at the geometry of maps from $\mathbf{R}^2$ to $\mathbf{R}^2$. After defining conformal maps (the technical name for those maps that preserve angles), we will show that an analytic function will be conformal at those points where its derivative is nonzero. This will be seen to follow almost immediately from the Cauchy-Riemann equations.

Before defining angle-preserving, we need a description for the angle between curves. Let

$$\sigma_1 : [-1,1] \to \mathbf{R}^2,$$

with $\sigma_1(t) = (x_1(t), y_1(t))$, and

$$\sigma_2 : [-1,1] \to \mathbf{R}^2,$$

with $\sigma_2(t) = (x_2(t), y_2(t))$, be two differentiable curves in the plane which intersect at

$$\sigma_1(0) = \sigma_2(0).$$

The angle between the two curves is defined to be the angle between the curves' tangent vectors.

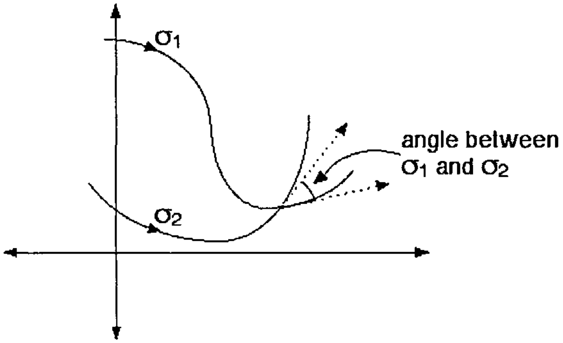

Thus we are interested in the dot product between the tangent vectors of the curves:

$$\begin{aligned}\frac{d\sigma_1}{dt}\cdot\frac{d\sigma_2}{dt} &= \left(\frac{dx_1}{dt},\frac{dy_1}{dt}\right)\cdot\left(\frac{dx_2}{dt},\frac{dy_2}{dt}\right)\\ &= \frac{dx_1}{dt}\frac{dx_2}{dt}+\frac{dy_1}{dt}\frac{dy_2}{dt}.\end{aligned}$$

Definition 9.5.1 *A function $f(x,y) = (u(x,y), v(x,y))$ will be* conformal *at a point (x_0, y_0) if the angle between any two curves intersecting at (x_0, y_0) is preserved, i.e., the angle between curves σ_1 and σ_2 is equal to the angle between the image curves $f(\sigma_1)$ and $f(\sigma_2)$.*

Thus

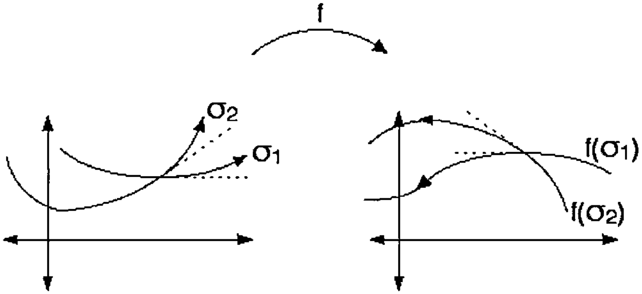

is conformal while

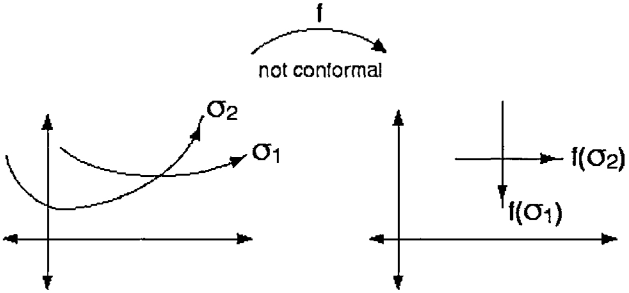

is not.

Theorem 9.5.1 *An analytic function $f(z)$ whose derivative at the point z_0 is not zero will be conformal at z_0.*

Proof : The tangent vectors are transformed under the map f by multiplying them by the two-by-two Jacobian matrix for f. Thus we want to show that multiplication by the Jacobian preserves angles. Writing f in its real and imaginary parts, with $z = x + iy$, as

$$f(z) = f(x,y) = u(x,y) + iv(x,y),$$

the Jacobian of f at the point $z_0 = (x_0, y_0)$ will be

$$Df(x_0,y_0) = \begin{pmatrix} \frac{\partial u}{\partial x}(x_0,y_0) & \frac{\partial u}{\partial y}(x_0,y_0) \\ \frac{\partial v}{\partial x}(x_0,y_0) & \frac{\partial v}{\partial y}(x_0,y_0) \end{pmatrix}.$$

But the function f is analytic at the point z_0 and hence the Cauchy-Riemann equations

$$\begin{aligned} \frac{\partial u}{\partial x}(x_0,y_0) &= \frac{\partial v}{\partial y}(x_0,y_0) \\ -\frac{\partial u}{\partial y}(x_0,y_0) &= \frac{\partial v}{\partial x}(x_0,y_0) \end{aligned}$$

hold, allowing us to write the Jacobian as

$$Df(x_0,y_0) = \begin{pmatrix} \frac{\partial u}{\partial x}(x_0,y_0) & \frac{\partial u}{\partial y}(x_0,y_0) \\ -\frac{\partial u}{\partial y}(x_0,y_0) & \frac{\partial u}{\partial x}(x_0,y_0) \end{pmatrix}.$$

Note that the columns of this matrix are orthogonal (i.e., their dot product is zero). This alone shows that the multiplication by the Jacobian will preserve angle. We can also show this by explicitly multiplying the Jacobian by the two tangent vectors $\frac{d\sigma_1}{dt}$ and $\frac{d\sigma_2}{dt}$ and then checking that the dot product between $\frac{d\sigma_1}{dt}$ and $\frac{d\sigma_2}{dt}$ is equal to the dot product of the image tangent vectors. □

This proof uses the Cauchy-Riemann equation approach to analyticity. A more geometric (and unfortunately a more vague) approach is to look carefully at the requirement for

$$\lim_{h\to 0} \frac{f(z_0+h)-f(z_0)}{h}$$

to exist, no matter what path is chosen for h to approach zero. This condition must place strong restrictions on how the function f alters angles.

This also suggests how to approach the converse. It can be shown (though we will not) that a conformal function f must satisfy either the limit for analyticity

$$\lim_{h\to 0} \frac{f(z_0+h)-f(z_0)}{h}$$

or that the limit holds for the conjugate function $\bar{f}$

$$\lim_{h\to 0} \frac{\bar{f}(z_0+h)-\bar{f}(z_0)}{h},$$

where the conjugate function of $f(z) = u(z) + iv(z)$ is

$$\bar{f}(z) = u(z) - iv(z).$$

9.6 The Riemann Mapping Theorem

Two domains D_1 and D_2 are said to be *conformally equivalent* if there is a one-to-one onto conformal map

$$f : D_1 \to D_2.$$

If such a function f exists, then its inverse function will also be conformal. Since conformal basically means that f is analytic, if two domains are conformally equivalent, then it is not possible to distinguish between them using the tools from complex analysis. Considering that analytic functions are special among functions, it is quite surprising that there are clean results for determining when two domains are conformally equivalent. The main result is:

Theorem 9.6.1 (Riemann Mapping Theorem) *Two simply connected domains, neither of which are equal to* $\mathbf{C}$, *are conformally equivalent.*

(Recall that a domain is *simply connected* if any closed loop in the domain is homotopic to a point in the domain, or intuitively, if every closed loop in the domain can be continuously shrunk to a point.) Frequently this result is stated as: for any simply connected domain D that is not equal to $\mathbf{C}$, there is a conformal one-to-one onto map from D to the unit disc. Thus the domain

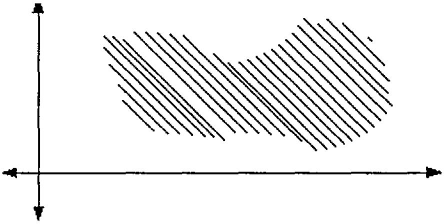

is conformally equivalent to

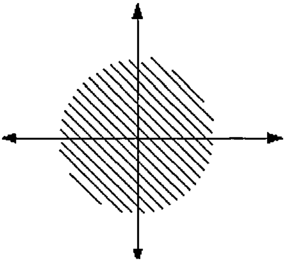

The Riemann Mapping Theorem, though, does not produce for us the desired function f. In practice, it is an art to find the conformal map. The standard approach is to first find conformal maps from each of the domains to the unit disc. Then, to conformally relate the two domains, we just compose various maps to the disc and inverses of maps to the disc.

For example, consider the right half plane

$$D = \{z \in \mathbf{C} : \mathrm{Re}(z) > 0\}.$$

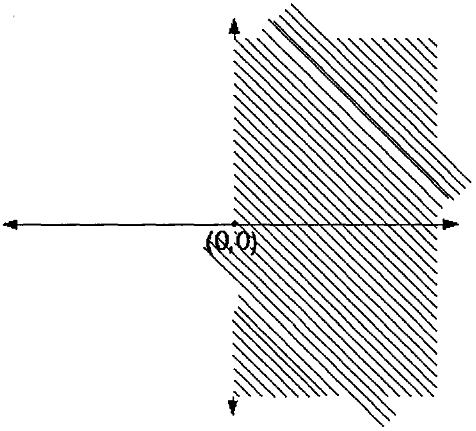

The function

$$f(z) = \frac{1 - z}{1 + z}$$

provides our conformal map from D to the unit disc. This can be checked by showing that the boundary of D, the y-axis, maps to the boundary of the unit disc. In this case, the inverse to f is f itself.

The Riemann Mapping Theorem is one reason why complex analysts spend so much time studying the function theory of the disc, as knowledge about the disc can be easily translated to knowledge about any simply connected domain.

In several complex variables theory, all is much more difficult, in large part because there is no higher dimensional analogue of the Riemann Mapping Theorem. There are many simply connected domains in $\mathbf{C}^n$ that are not conformally equivalent.

9.7 Several Complex Variables: Hartog's Theorem

Let $f(z_1, \dots, z_n)$ be a complex-valued function of n complex variables. We say that f is *holomorphic* (or *analytic*) in several variables if $f(z_1, \dots, z_n)$ is holomorphic in each variable z_i separately. Although many of the basic results for one variable analytic functions can be easily carried over to the several variable case, the subjects are profoundly different. These differences start with Hartog's Theorem, which is the subject of this section.

Consider the one-variable function $f(z) = \frac{1}{z}$. This function is holomorphic at all points except at the origin, where it is not even defined. It is thus easy to find a one-variable function that is holomorphic except for at one point. But what about the corresponding question for holomorphic functions of several variables? Is there a function $f(z_1, \dots, z_n)$ that is holomorphic everywhere except at an isolated point? Hartog's theorem is that no such function can exist.

Theorem 9.7.1 (Hartog's Theorem) *Let U be an open connected region in $\mathbf{C}^n$ and let V be a compact connected set contained in U. Then any function $f(z_1, \dots, z_n)$ that is holomorphic on $U - V$ can be extended to a holomorphic function that is defined on all of U.*

This certainly includes the case when V is an isolated point. Before sketching a proof for a special case of this theorem, consider the following question that is now quite natural, namely, is there a natural condition on open connected sets U so that there will exist holomorphic functions on U that cannot be extended to a larger open set. Such sets U are called *domains of holomorphy.* Hartog's Theorem says that regions like $U-$(isolated point) are not domains of holomorphy. In fact, a clean criterion does exist and involves geometric conditions on the boundary of the open set U (technically, the boundary must be pseudoconvex). Hartog's Theorem opens up a whole new world of phenomena for several complex variables.

One way of thinking about Hartog's Theorem is in considering the function $\frac{f(z_1,\dots,z_n)}{g(z_1,\dots,z_n)}$, where both f and g are holomorphic, as a possible counterexample. If we can find a holomorphic function g that has a zero at an isolated point or even on a compact set, then Hartog's Theorem will be false. Since Hartog's Theorem is indeed a theorem, an analytic function in more than one variable cannot have a zero at an isolated point. In fact, the study of the zero locus $g(z_1, \dots, z_n) = 0$ leads to much of algebraic and analytic geometry.

Now to sketch a proof of Hartog's Theorem, subject to simplifying assumptions that U is the polydisc

$$U = \{(z, w) : |z| < 1, |w| < 1\}$$

and that V is the isolated point $(0,0)$. We will also use the fact that if two functions that are holomorphic on an open connected region U are equal on an open subset of U, then they are equal on all of U. (The proof of this fact is similar to the corresponding result in one-variable complex analysis, which can be shown to follow from exercise three at the end of this chapter.)

Let $f(z,w)$ be a function that is holomorphic on $U - (0,0)$. We want to extend f to be a holomorphic function on all of U. Consider the sets $z = c$, where c is a constant with $|c| < 1$. Then the set

$$(z = c)\bigcap(U - (0,0))$$

is an open disc of radius one if $c \neq 0$ and an open disc punctured at the origin if $c = 0$. Define a new function by setting

$$F(z,w) = \frac{1}{2\pi i}\int_{|v|=\frac{1}{2}} \frac{f(z,v)}{v-w}dv.$$

This will be our desired extension. First, the function F is defined at all points of U, including the origin. Since the z variable is not varying in the integral, we have by Cauchy's Integral Formula that $F(z,w)$ is holomorphic in the w variable. Since the original function f is holomorphic with respect to the z variable, we have that F is holomorphic with respect to z; thus F is holomorphic on all of U. But again by Cauchy's Integral Formula, we have that $F = f$ when $z \neq 0$. Since the two holomorphic functions are equal on an open set of U, then we have equality on $U - (0,0)$.

The general proof of Hartog's Theorem is similar, namely to reduce the problem to slicing the region U into a bunch of discs and punctured discs and then using Cauchy's Integral Formula to create the new extension.

9.8 Books

Since complex analysis has many applications, there are many beginning textbooks, each emphasizing different aspects of the subject. An excellent introduction is in Marsden and Hoffman's *Basic Complex Analysis* [83]. Palka's *An Introduction to Complex Function Theory* [92] is also an excellent text. (I first learned complex analysis from Palka.) A recent beginning book is Greene and Krantz' *Function Theory of One Complex Variable* [49]. For a rapid fire introduction, Spiegels' *Complex Variables* [101] is outstanding, containing a wealth of concrete problems.

There are a number of graduate texts in complex analysis, which do start at the beginning but then build quickly. Ahlfors' book [1] has long been the standard. It reflects the mathematical era in which it was written (the 1960s) and thus approaches the subject from a decidedly abstract point of view. Conway's *Functions of One Complex Variable* [21] has long been the prime competitor to Ahlfors for the beginning graduate student market and is also quite good. The recent book by Berenstein and Gay [8] provides a modern framework for complex analysis. A good introduction to complex analysis in several variables is Krantz' *Function Theory in Several Variables* [77].

Complex analysis is probably the most beautiful subject in undergraduate mathematics. Neither Krantz' *Complex Analysis: The Geometric Viewpoint* [78] nor Davis' *The Schwarz Function and its Applications* [25] are textbooks but both show some of the fascinating implications contained in complex analysis and are good places to see how how analytic functions can be naturally linked to other parts of mathematics.

9.9 Exercises

1. Letting $z = x + iy$, show that the function

$$f(z) = f(x, y) = y^2$$

is not analytic. Show that it does not satisfy the Cauchy Integral Formula

$$f(z_0) = \frac{1}{2\pi i} \int_\sigma \frac{f(z)}{z - z_0} dz,$$

for the case when $z_0 = 0$ and when the closed loop σ is the circle of radius one centered at the origin.

2. Find a function $f(z)$ that is not analytic, besides the function given in problem one. If you think of $f(z)$ as a function of the two variables

$$f(x, y) = u(x, y) + iv(x, y),$$

almost any choice of functions u and v will work.

3. Let $f(z)$ and $g(z)$ be two analytic functions that are equal at all points on a closed loop σ. Show that for all points z in the interior of the closed loop we have the two functions equal. As a hint, start with the assumption that $g(z)$ is the zero function and thus that $f(z)$ is zero along the loop σ. Then show that $f(z)$ must also be the zero function inside the loop.

4. Find a one-to-one onto conformal map from the unit disc $\{(x, y) : x^2 + y^2 < 1\}$ to the first quadrant of the plane $\{(x, y) : x > 0 \text{ and } y > 0\}$.

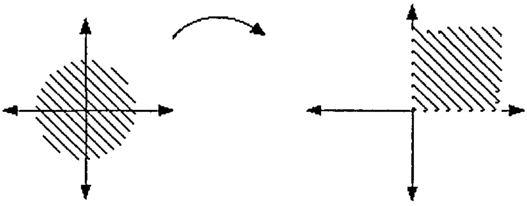

5. Let z_1, z_2 and z_3 be three distinct complex numbers. Show that we can find numbers a, b, c and d with $ad - bc = 1$ such that the map

$$T(z) = \frac{az+b}{cz+d}$$

maps z_1 to 0, z_2 to 1 and z_3 to 2. Show that the numbers a, b, c and d are uniquely determined, up to multiplication by -1.

6. Find $\int_{-\infty}^{\infty} \frac{\mathrm{d}x}{1+x^2}$ as follows:

a. Find

$$\int_{\gamma} \frac{\mathrm{d}z}{1+z^2},$$

where $\gamma = \gamma_1 + \gamma_2$ is the closed loop in the complex plane

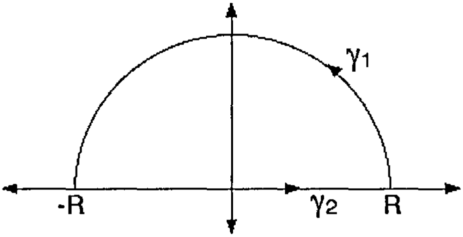

consisting of the path

$$\gamma_1 = \{Re^{\pi\theta} : 0 \leq \theta\pi\}$$

and

$$\gamma_2 = \{(x,0) \in \mathbf{R}^2 : -R \leq x \leq R\}.$$

b. Show that

$$\lim_{R\to\infty} \int_{\gamma_1} \frac{\mathrm{d}z}{1+z^2} = 0.$$

c. Conclude with the value for $\int_{-\infty}^{\infty} \frac{\mathrm{d}x}{1+x^2}$.

(This is a standard problem showing how to calculate hard real integrals easily. This is a hard problem if you have never used residues before; it should be straightforward if you have.)

7. The goal of this problem is to construct a conformal map from the unit sphere (minus the north pole) to the complex numbers. Consider the sphere $S^2 = \{(x, y, z) : x^2 + y^2 + z^2 = 1\}$.

a. Show that the map

$$\pi : S^2 - (0,0,1) \to \mathbf{C}$$

defined by

$$\pi(x,y,z) = \frac{x}{1-z} + i\frac{y}{1-z}$$

is one-to-one, onto and conformal.

b. We can consider the complex numbers **C** as sitting inside $\mathbf{R}^3$ by mapping $x + iy$ to the point $(x, y, 0)$. Show that the above map π can be interpreted as the map that sends a point (x, y, z) on $S^2 - (0, 0, 1)$ to the point on the plane $(z = 0)$ that is the intersection of the plane with the line through (x, y, z) and $(0, 0, 1)$.

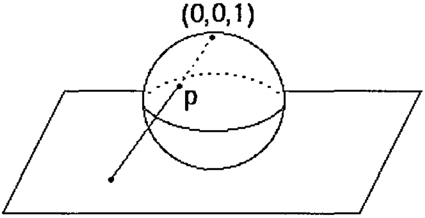

c. Justify why people regularly identify the unit sphere with $\mathbf{C} \cup \infty$.

Chapter 10

Countability and the Axiom of Choice

Basic goal: Comparing infinite sets

Both countability and the axiom of choice grapple with the elusive notions behind "infinity". While both the integers $\mathbf{Z}$ and the real numbers $\mathbf{R}$ are infinite sets, we will see that the infinity of the reals is strictly larger than the infinity of the integers. We will then turn to the Axiom of Choice, which, while straightforward and not an axiom at all for finite sets, is deep and independent from the other axioms of mathematics when applied to infinite collections of sets. Further, the Axiom of Choice implies a number of surprising and seemingly paradoxical results. For example, we will show that the Axiom of Choice forces the existence of sets of real numbers that cannot be measured.

10.1 Countability

The key is that there are different orders or magnitudes of infinity. The first step is to find the right definition for when two sets are of the same size.

Definition 10.1.1 *A set A is* finite of cardinality n *if there is a one-to-one onto function from the set $\{1, 2, 3, \cdots, n\}$ to A. The set A is* countably infinite *if there is a one-to-one onto function from the natural numbers $\mathbf{N} = \{1, 2, 3, \ldots,\}$ to A. A set that is either finite or countably infinite is said to be* countable. *A set A is* uncountably infinite *if it is not empty and not countable.*

For example, the set $\{a, b, c\}$ is finite with 3 elements. The more troubling and challenging examples appear in the infinite cases.

For example, the positive even numbers

$$2\mathrm{N} = \{2, 4, 6, 8, \cdots\},$$

while properly contained in the natural numbers **N**, are of the same size as **N** and hence are countably infinite. An explicit one-to-one onto map

$$f : \mathrm{N} \to 2\mathrm{N}$$

is $f(n) = 2 \cdot n$. Usually this one-to-one correspondence is shown via:

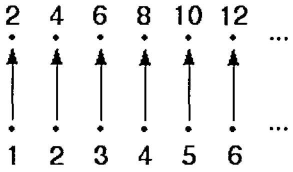

The set of whole numbers $\{0, 1, 2, 3, \ldots\}$ is also countably infinite, as seen by the one-to-one onto map

$$f : \mathrm{N} \to \{0, 1, 2, 3, \ldots\}$$

given by

$$f(n) = n - 1.$$

Here the picture is

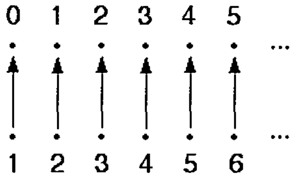

The integers **Z** are also countably infinite. The picture is

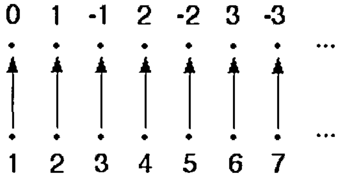

while an explicit one-to-one onto function

$$f : \mathbf{N} \to \mathbf{Z}$$

is, for even n,

$$f(n) = \frac{n}{2}$$

and, for odd n,

$$f(n) = -\frac{n-1}{2}.$$

It is typical for the picture to be more convincing than the actual function.

The rationals

$$\mathbf{Q} = \{\frac{p}{q} : p, q \in \mathbf{Z}, q \neq 0\}$$

are also countably infinite. The picture for showing that the positive rationals are countably infinite is as follows:

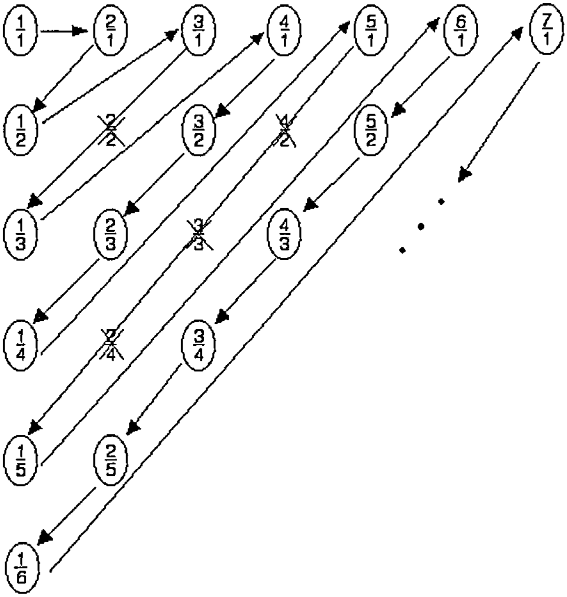

Every positive rational appears in the above array and will eventually be hit by a natural number.

In fact

Theorem 10.1.1 *Let A and B be two countably infinite sets. Then the Cartesian product $A \times B$ is also countably infinite.*

Proof: Since both A and B are in one-to-one correspondence with the natural numbers $\mathbf{N}$, all we need show is that the product $\mathbf{N}\times\mathbf{N}$ is countably infinite. For $\mathbf{N}\times\mathbf{N} = \{(n,m) : n,m \in \mathbf{N}\}$, the correct diagram is:

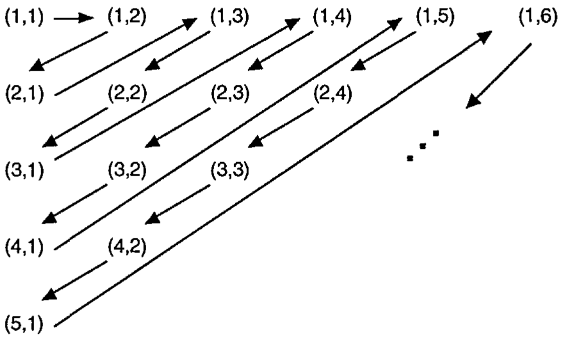

More algebraically, but less clearly, an explicit one-to-one onto map

$$f : \mathbf{N}\times\mathbf{N} \to \mathbf{N}$$

is

$$f(m,n) = \frac{(n+m-2)(n+m-1)}{2} + m. \ \square$$

Note that the fact that $\mathbf{N}\times\mathbf{N}$ is the same size as $\mathbf{N}$ is of course in marked contrast to the finite case. To make this painfully obvious, consider $A = \{a,b,c\}$, a set with three elements. Then $A \times A$ is the nine element set $\{(a,a),(a,b),(a,c),(b,a),(b,b),(b,c),(c,a),(c,b),(c,c)\}$.

There are infinite sets which, in some sense, are of size strictly larger than the natural numbers. Far from being esoteric, the basic example is the set of real numbers; the reals, while certainly not finite, are also not countably infinite.

We will give the famed Cantor diagonalization argument showing that the real numbers $[0,1] = \{x \in \mathbf{R} : 0 \le x \le 1\}$ cannot be countable.

Theorem 10.1.2 *The interval* $[0,1]$ *is not countable.*

Proof : The proof is by contradiction. We assume that there is a one-to-one onto map $f : \mathbf{N} \to [0,1]$ and then find a real number in $[0,1]$ that is not in the image, contradicting the assumption that f is onto. We will use that every real number in $[0,1]$ can be expressed as a decimal expansion

$$0.x_1x_2x_3x_4\ldots,$$

where each x_k is $0, 1, 2, 3, \ldots$ or 9. To make this expansion unique, we will always round up, except for the case $0.99999\ldots$ which we leave as is. Thus $0.32999\ldots$ will always be written as 0.3300.

Now let us take our assumed one-to-one correspondence $f : \mathbf{N} \to [0, 1]$ and start writing down its terms. Let

$$\begin{aligned} f(1) &= .a_1a_2a_3\cdots, \\ f(2) &= .b_1b_2b_3\cdots, \\ f(3) &= .c_1c_2c_3\cdots, \\ f(4) &= .d_1d_2d_3\cdots, \\ f(5) &= .e_1e_2e_3\cdots, \end{aligned}$$

and so forth. Note that the a_i, b_j, etc. are now fixed numbers between 0 and 9, given to us by the assumed one-to-one correspondence. They are not variables.

We will construct a new real number $.N_1N_2N_3N_4\ldots$ which will never appear in the above list, forcing a contradiction to the assumption that f is onto. Set

$$N_k = \begin{cases} 4, & \text{if the } k^{th}\text{entry of } f(k) \neq 4 \\ 5, & \text{if the } k^{th}\text{entry of } f(k) = 4 \end{cases}$$

(The choice of the numbers 4 and 5 are not important; any two integers between 0 and 9 would do just as well.)

Note that N_1 is 4 if $a_1 \neq 4$ and is 5 if $a_1 = 4$. Thus, no matter what,

$$.N_1N_2N_3\ldots \neq .a_1a_2\ldots = f(1).$$

Likewise N_2 is 4 if $b_2 \neq 4$ and is 5 if $b_2 = 4$ and hence

$$.N_1N_2N_3\ldots \neq .b_2b_2b_3\ldots = f(2).$$

This continues. Since our decimal expansions are unique, and since each N_k is defined so that it is not equal to the k^{th} term in $f(k)$, we must have that $.N_1N_2N_3\cdots$ is not equal to any $f(k)$, meaning that f cannot be onto. Thus there can never be an onto function from the natural numbers to the interval $[0, 1]$. Since the reals are certainly not finite, they must be uncountably infinite.

10.2 Naive Set Theory and Paradoxes

The question of what is a mathematical object was a deep source of debate in the last part of the eighteenth and first part of the nineteenth century. There has only been at best a partial resolution, caused in part by Gödel's work in logic and in part by exhaustion. Does a mathematical object exist only if an algorithm can be written that will explicitly construct the object or does it exist if the assumption of its existence leads to no contradictions, even if we can never find an example? The tension between constructive proofs versus existence proofs has in the last thirty years been eased with the development of complexity theory. The constructive camp was led by Kronecker (1823-1891), Brouwer (1881-1966) and Bishop (1928-1983). The existential camp, led by Hilbert (1862-1943), won the war, leading to most mathematicians' belief that all of mathematics can be built out of a correct set-theoretic foundation, usually believed to be an axiomatic system called Zermelo-Fraenkel plus the Axiom of Choice (for a list of those axioms, see Paul Cohen's *Set Theory and the Continuum Hypothesis* [20] Chapter II, Sections 1 and 2). This is in spite of the fact that few working mathematicians can actually write down these axioms, which certainly suggests that our confidence in our work does not stem from the axioms. More accurately, the axioms were chosen and developed to yield the results we already know to be true. In this section we informally discuss set theory and then give the famed Zermelo-Russell paradox, which shows that true care must be exercised in understanding sets.

The naive idea of a set is pretty good. Here a set is some collection of objects sharing some property. For example

$$\{n : n \text{ is an even number}\}$$

is a perfectly reasonable set. Basic operations are union, intersection and complement. We will see now how to build integers out of sets.

First for one subtlety. Given a set A, we can always form a new set, denoted by $\{A\}$, which consists of just one element, namely the set A. If A is the set of all even integers and thus containing an infinite number of elements, the set $\{A\}$ has only one element. Given a set A, we define the successor set A^+ as the union of the set A with the set $\{A\}$. Thus $x \in A^+$ if either $x \in A$ or $x = \{A\}$.

We start with the empty set $\emptyset$, the set that contains no elements. This set will correspond to the integer 0. Then we label the successor to the empty set by 1:

$$1 = \emptyset^+ = \{\emptyset\},$$

the successor to the successor of the empty set by 2:

$$2 = (\emptyset^+)^+ = \{\emptyset, \{\emptyset\}\},$$

and in general the successor to the set n by $n + 1$.

By thinking of the successor as adding by one, we can recover by recursion addition and thus in turn multiplication, subtraction and division.

Unfortunately, just naively proceeding along in this fashion will lead to paradoxes. We will construct here what appears to be a set but which cannot exist. First, note that sometimes a set can be a member of itself and sometimes not (at least if we are working in naive set theory; much of the mechanics of Zermelo-Fraenkel set theory is to prevent such nonchalant assumptions about sets). For example, the set of even numbers is not itself an even number and hence is not an element of itself. On the other hand, the set of all elements that are themselves sets with more than two elements is a member of itself. We can now define our paradoxical set. Set

$$\begin{aligned} X &= \{A : A \text{ is a set that does not contain itself}\} \\ &= \{A : A \notin A\}. \end{aligned}$$

Is the set X an element of itself? If $X \in X$, then by the definition of X, we must have $X \notin X$, which is absurd. But if $X \notin X$, then $X \in X$, which is also silly. There are problems with allowing X to be a set. This is the Zermelo-Russell paradox

Do not think this is just a trivial little problem. Russell (1872-1970) reports in his autobiography that when he first thought of this problem he was confident it could easily be resolved, probably that night after dinner. He spent the next year struggling with it and had to change his whole method of attack on the foundations of mathematics. (Russell, with Whitehead 1861-1947), did not use set theory but instead developed type theory; type theory is abstractly no better or worse than set theory, but mathematicians base their work on the language of set theory, probably by the historical accident of World War II, which led US mathematicians to be taught by German refugees, who knew set theory, as Zermelo (1871-1953 was German.)

Do not worry too much about the definitions of set theory. You should be nervous, though, if your sets refer to themselves, as this is precisely what led to the above difficulty.

10.3 The Axiom of Choice

The axioms in set theory were chosen and developed to yield the results we already know to be true. Still, we want these axioms to be immediately obvious. Overall, this is the case. Few of the actual axioms are controversial, save for the Axiom of Choice, which states:

Axiom 10.3.1 (Axiom of Choice) *Let* $\{X_\alpha\}$ *be a family of nonempty sets. Then there is a set* X *which contains, from each set* X_α*, exactly one element.*

For a finite collection of sets, this is obvious and not at all axiomatic (meaning that it can be proven from other axioms). For example, let $X_1 = \{a, b\}$ and $X_2 = \{c, d\}$. Then there is certainly a set X containing one element from X_1 and one element from X_2; for example, just let $X = \{a, c\}$.

The difficulties start to arise when applying the axiom to an infinite (possibly uncountably infinite) number of sets. The Axiom of Choice gives no method for finding the set X; it just mandates the existence of X. This leads to the observation that if the Axiom of Choice is needed to prove the existence of some object, then you will never be able to actually construct that object. In other words, there will be no method to actually construct the object; it will merely be known to exist.

Another difficulty lies not in the truth of the axiom of choice but in the need to assume it as an axiom. Axioms should be clear and obvious. No one would have any difficulty with its statement if it could be proven to follow from the other axioms.

In 1939, Kurt Gödel showed that the Axiom of Choice is consistent with the other axioms. This means that using the Axiom of Choice will lead to no contradictions that were not, in some sense, already present in the other axioms. But in the early 1960s, Paul Cohen [20] showed that the Axiom of Choice was independent of the other axioms, meaning that it cannot be derived from the other axioms and hence was truly an axiom. In particular, one can assume that the Axiom of Choice is false and still be confident that no contradictions will arise.

A third difficulty with the Axiom of Choice is that it is equivalent to any number of other statements, some of which are quite bizarre. To see some of the many equivalences to the Axiom of Choice, see Howard and Rubin's *Consequences of the Axiom of Choice* [62]. One of these equivalences is the subject of the next section.

10.4 Non-measurable Sets

Warning: This section will assume a working knowledge of Lebesgue measure on the real numbers. In particular, we will need that

- If a set A is measurable, its measure $m(A)$ is equal to its outer measure $m^*(A)$.
- If $A_1, A_2, \ldots$ are disjoint sets that are measurable, then the union is measurable, with

$$m(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} m(A_i).$$

This last condition corresponds to the idea that if we have two sets with lengths a and b, say, then the length of the two sets placed next to each other should be $a+b$. Also, this example closely follows the example of a nonmeasurable set in Royden's *Real Analysis* [95].

We will find a sequence of disjoint sets $A_1, A_2, \ldots$, all of which have the same outer measure and hence, if measurable, the same measure, whose union is the unit interval $[0,1]$. Since the Lebesgue measure of the unit interval is just its length, we will have

$$1 = \sum_{i=1}^{\infty} m(A_i).$$

If each A_i is measurable, since the measures are equal, this would mean that we can add a number to itself infinitely many times and have it sum to one. This is absurd. If a series converges, then the individual terms in the series must converge to zero. Certainly they cannot all be equal.

The point of this section is that to find these sets A_i, we will need to use the Axiom of Choice. This means that we are being fairly loose with the term "find", as these sets will in no sense actually be constructed. Instead, the Axiom of Choice will allow us to claim their existence, without actually finding them.

We say that x and y in $\in [0,1]$ are equivalent, denoted by $x \equiv y$, if $x-y$ is a rational number. It can be checked that this is an equivalence relation (see Appendix A for the basic properties of equivalence relations) and thus splits the unit interval into disjoint equivalency classes.

We now apply the Axiom of Choice to these disjoint sets. Let A be the set containing exactly one element from each of these equivalency classes. Thus the difference between any two elements of A cannot be a rational number. Note again, we do not have an explicit description of A. We have no way of knowing if a given real number is in A, but, by the Axiom of Choice, the set A does exist. In a moment we will see that A cannot be measurable.

We will now find a countable collection of disjoint sets, each with the same outer measure as the outer measure of the set A, whose union will be the unit interval. Now, since the rational numbers in $[0,1]$ are countable, we can list all rational numbers between zero and one as $r_0, r_1, r_2, \ldots$. For convenience, assume that $r_0 = 0$. For each rational number r_i, set

$$A_i = A + r_i \,(\mathrm{mod}\, 1).$$

Thus the elements of A_i are of the form

$$a + r_i - \text{greatest integer part of } (a + r_i).$$

In particular, $A = A_0$. It is also the case that for all i

$$m^*(A) = m^*(A_i),$$

which is not hard to show, but is mildly subtle since we are not just shifting the set A by the number r_i but are then modding out by one.

We now want to show that the A_i are disjoint and cover the unit interval. First, assume that there is a number x in the intersection of A_i and A_j. Then there are numbers a_i and a_j in the set A such that

$$x = a_i + r_i \text{ (mod 1)} = a_j + r_j \text{ (mod 1)}.$$

Then $a_i - a_j$ is a rational number, meaning that $a_i \equiv a_j$, which forces $i = j$. Thus if $i \neq j$, then

$$A_i \cap A_j = \emptyset.$$

Now let x be any element in the unit interval. It must be equivalent to some element a in A. Thus there is a rational number r_i in the unit interval with either

$$x = a + r_i \text{ or } a = x + r_i.$$

In either case we have $x \in A_i$. Thus the A_i are indeed a countable collection of disjoint sets that cover the unit interval. But then we have the length of the unit interval as an infinite series of the same number:

$$1 = \sum_{i=1}^{\infty} m(A_i) = \sum_{i=1}^{\infty} m(A),$$

which is impossible. Thus the set A cannot be measurable.

10.5 Gödel and Independence Proofs

In the debates about the nature of mathematical objects, all agreed that correct mathematics must be consistent (i.e., it should not be possible to both prove a statement and its converse). Eventually it was realized that most people were also implicitly assuming that mathematics was complete (meaning that any mathematical statement must ultimately be capable of being either proven or disproven). David Hilbert wanted to translate both of these goals into precise mathematical statements, each capable of rigorous proof. This attempt became known as Formalism. Unfortunately for Hilbert's school, K. Gödel (1906-1977) in 1931 destroyed any of these hopes. Gödel showed:

Any axiomatic system strong enough to include basic arithmetic must have statements in it that can be neither proven nor disproven, within the system.

Further, the example Gödel gave of a statement that could be neither proven nor disproven was that the given axiomatic system was itself consistent.

Thus in one fell swoop, Gödel showed that both consistency and completeness were beyond our grasp. Of course, no one seriously thinks that modern mathematics has within it a hidden contradiction. There are statements, though, that people care about that are not capable of being proven or disproven within Zermelo-Fraenkel set theory. The Axiom of Choice is an example of this. Such statements are said to be *independent* of the other axioms of mathematics. On the other hand, most open questions in mathematics are unlikely to be independent of Zermelo-Fraenkel set theory plus the Axiom of Choice. One exception is the question of P=NP (discussed in Chapter Sixteen), which many are now believing to be independent of the rest of mathematics.

10.6 Books

For many years the best source for getting an introduction to set theory has been Halmos' *Naive Set Theory* [53], which he wrote, in large part, to teach himself the subject. A more recent text is Moschovakis' *Notes on Set Theory* [87]. An introduction, not to set theory, but to logic is *Incompleteness Phenomenon* by Goldstern and Judah [46]. A slightly more advanced text, by a tremendous expositor, is Smullyan's *Gödel's Incompleteness Theorems* [100]. A concise, high level text is Cohen's *Set Theory and the Continuum Hypothesis* [20].

A long time popular introduction to Gödel's work has been Nagel and Newman's *Gödel's Proof* [89]. This is one of the inspirations for the amazing book of Hofstadter, *Gödel, Escher and Bach* [61]. Though not precisely a math book, it is full of ideas and should be read by everyone. Another impressive recent work is Hintikka's *Principles of Mathematics, Revisited* [60]. Here a new scheme for logic is presented. It also contains a summary of Hintikka's game-theoretic interpretation of Gödel's work.

10.7 Exercises

1. Show that the set

$$\{ax^2 + bx + c : a, b, c \in \mathbf{Q}\}$$

of all one variable polynomials of degree two with rational coefficients is countable.

2. Show that the set of all one variable polynomials with rational coefficients is countable.
3. Show that the set

$$\{a_0 + a_1 x + a_2 x^2 + \ldots : a_0, a_1, a_2, \ldots \in \mathbf{Q}\}$$

of all formal power series in one variable with rational coefficients is not countable.
4. Show that the set of all infinite sequences consisting of zeros and twos is uncountable. (This set will be used to show that the Cantor set, which will be defined in Chapter Twelve, is uncountable.)
5. In section two, the whole numbers were defined as sets. Addition by one was defined. Give a definition for addition by two and then a definition in general for whole numbers. Using this definition, show that $2 + 3 = 3 + 2$.
6.(Hard) A set S is partially ordered if there is an operation $<$ such that given any two elements x and y, we have $x < y$, $y < x$, $x = y$ or x and y have no relationship. The partial ordering is a total ordering if it must be the case that given any two elements x and y, it must be the case that $x < y$, $y < x$ or $x = y$. For example, if S is the real numbers, the standard interpretation of $<$ as less than places a total ordering on the reals. On the other hand, if S is the set of all subsets of some other set, then a partial ordering would exist if we let $<$ denote set containment. This is not a total ordering since given any two subsets, it is certainly not the case that one must be contained in the other. A partially ordered set is called a *poset.*

Let S be a poset. A chain in S is a subset of S on which the partial ordering becomes a total ordering. *Zorn's Lemma* states that if S is a poset such that every chain has an upper bound, then S contains a maximal element. Note that the upper bound to a chain need not be in the chain and that the maximal element need not be unique.

a. Show that the Axiom of Choice implies Zorn's Lemma.

b. Show that Zorn's Lemma implies the Axiom of Choice (this is quite a bit harder).

7. (Hard) The Hausdorff Maximal Principle states that every poset has a maximal chain, meaning a chain that is not strictly contained in any other chain. Show that the Hausdorff Maximal Principle is equivalent to the Axiom of Choice.
8. (Hard) Show that the Axiom of Choice (via the Hausdorff Maximal Principle) implies that every field is contained in an algebraically closed field. (For the definitions, see Chapter Eleven.)

Chapter 11

Algebra

Basic Objects:	Groups and rings
Basic Maps:	Group and ring homomorphisms

While current abstract algebra does indeed deserve the adjective *abstract*, it has both concrete historical roots and modern day applications. Central to undergraduate abstract algebra is the notion of a group, which is the algebraic interpretation of the geometric idea of symmetry. We can see something of the richness of groups in that there are three distinct areas that gave birth to the correct notion of an abstract group: attempts to find (more accurately, attempts to prove the inability to find) roots of polynomials, the study by chemists of the symmetries of crystals, and the application of symmetry principles to solve differential equations.

The inability to generalize the quadratic equation to polynomials of degree greater than or equal to five is at the heart of Galois Theory and involves the understanding of the symmetries of the roots of a polynomial. Symmetries of crystals involve properties of rotations in space. The use of group theory to understand the symmetries underlying a differential equation leads to Lie Theory. In all of these the idea and the applications of a group are critical.

11.1 Groups

This section presents the basic definitions and ideas of group theory.

Definition 11.1.1 *A nonempty set G that has a binary operation*

$$G \times G \to G,$$

denoted for all elements a and b in G by $a \cdot b$, is a group *if:*

i)There is an element $e \in G$ such that $e \cdot a = a \cdot e = a$, for all a in G. (The element e is of course called the identity.)

ii) For any $a \in G$, there is an element denoted by a^{-1} such that $aa^{-1} = a^{-1}a = e$. (Naturally enough, a^{-1} is called the inverse of a.)

iii) For all $a, b, c \in G$, we have $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ (i.e., we must have associativity).

Note that commutativity is not required.

Now for some examples. Let $\mathbf{GL(n,R)}$ denote the set of all $n \times n$ invertible matrices with real coefficients. Under matrix multiplication, we claim that $\mathbf{GL(n,R)}$ is a group. The identity element of course is simply the identity matrix

$$\begin{pmatrix} 1 & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & 1 \end{pmatrix}.$$

The inverse of an element will be its matrix inverse. The check that matrix multiplication is associative is a long calculation. The final thing to check is to see that if A and B are invertible $n \times n$ matrices, then their product, $A \cdot B$, must be invertible. From the key theorem of linear algebra, a matrix is invertible if and only if its determinant is nonzero. Using that $\det(A \cdot B) = \det(A)\det(B)$, we have

$$\det(A \cdot B) = \det(A) \cdot \det(B) \neq 0.$$

Thus $\mathbf{GL(n,R)}$ is a group.

Note that for almost any choice of two matrices

$$A \cdot B \neq B \cdot A.$$

The group is not commutative. Geometrically, we can interpret the elements of $\mathbf{GL(n,R)}$ as linear maps on $\mathbf{R}^n$. In particular, consider rotations in three-space. These do not commute (showing this is an exercise at the end of this chapter). Rotations can be represented as invertible 3×3 matrices and hence as elements in $\mathbf{GL(3,R)}$. If we want groups to be an algebraic method for capturing symmetry, then we will want rotations in space to form a group. Hence we cannot require groups to be commutative. (Note that rotations are associative, which is why we do require groups to be associative.)

The key examples of finite groups are the permutation groups. The permutation group, $\mathbf{S}_n$, is the set of all permutations on n distinct elements. The binary operation is composition while the identity element is the trivial permutation that permutes nothing.

To practice with the usual notation, let us look at the group of permutations on three elements:

$$S_3 = \{e, (12), (13), (23), (123), (132)\}.$$

Of course we need to explain the notation. Fix an ordered triple (a_1, a_2, a_3) of numbers. Here order matters. Thus (cow, horse, dog) is different from the triple (dog, horse, cow). Each element of S_3 will permute the ordering of the ordered triple. Specifically, the element (12) permutes (a_1, a_2, a_3) to (a_2, a_1, a_3):

$$(a_1, a_2, a_3) \overset{(12)}{\mapsto} (a_2, a_1, a_3).$$

For example, the element (12) will permute (cow, horse, dog) to the triple (horse, cow, dog). The other elements of the group S_3 act as follows: (13) permutes (a_1, a_2, a_3) to (a_3, a_2, a_1) :

$$(a_1, a_2, a_3) \overset{(13)}{\mapsto} (a_3, a_2, a_1),$$

(23) permutes (a_1, a_2, a_3) to (a_1, a_3, a_2):

$$(a_1, a_2, a_3) \overset{(23)}{\mapsto} (a_1, a_3, a_2),$$

(123) permutes (a_1, a_2, a_3) to (a_3, a_1, a_2):

$$(a_1, a_2, a_3) \overset{(123)}{\mapsto} (a_3, a_1, a_2),$$

(132) permutes (a_1, a_2, a_3) to (a_2, a_3, a_1):

$$(a_1, a_2, a_3) \overset{(132)}{\mapsto} (a_2, a_3, a_1),$$

and of course the identity element e leave the triple (a_1, a_2, a_3) alone:

$$(a_1, a_2, a_3) \overset{(e)}{\mapsto} (a_1, a_2, a_3).$$

By composition we can multiply the permutations together, to get the following multiplication table for S_3:

·	e	(12)	(13)	(23)	(123)	(132)
e	e	(12)	(13)	(23)	(123)	(132)
(12)	(12)	e	(123)	(132)	(13)	(23)
(13)	(13)	(132)	e	(123)	(23)	(12)
(23)	(23)	(123)	(132)	e	(12)	(13)
(123)	(123)	(23)	(12)	(13)	(132)	e
(132)	(132)	(13)	(23)	(12)	e	(123)

Note that S_3 is not commutative. In fact, S_3 is the smallest possible noncommutative group. In honor of one of the founders of group theory, Niels Abel, we have:

Definition 11.1.2 *A group that is commutative is* abelian.

The integers **Z** under addition form an abelian group. Most groups are not abelian.

We want to understand all groups. Of course, this is not actually doable. Hopefully we can at least build up groups from possibly simpler, more basic groups. To start this process, we make the following definition:

Definition 11.1.3 *A nonempty subset H of G is a* subgroup *if H is itself a group, using the binary operation of G.*

For example, let

$$H = \left\{ \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix} : \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in \mathbf{GL(2,R)}) \right\}.$$

Then H is a subgroup of the group $\mathbf{GL(3,R)}$ of invertible 3×3 matrices.

Definition 11.1.4 *Let G and $\hat{G}$ be two groups. Then a function*

$$\sigma : G \to \hat{G}$$

is a group homomorphism *if for all $g_1, g_2 \in G$,*

$$\sigma(g_1 \cdot g_2) = \sigma(g_1) \cdot \sigma(g_2).$$

For example, let $A \in \mathbf{GL(n,R)}$. Define $\sigma : \mathbf{GL(n,R)} \to \mathbf{GL(n,R)}$ by

$$\sigma(B) = A^{-1}BA.$$

Then for any two matrices $B, C \in \mathbf{GL(n,R)})$, we have

$$\begin{aligned} \sigma(BC) &= A^{-1}BCA \\ &= A^{-1}BAA^{-1}CA \\ &= \sigma(B) \cdot \sigma(C). \end{aligned}$$

There is a close relationship between group homomorphisms and a special class of subgroup. Before we can exhibit this, we need:

Definition 11.1.5 *Let H be a subgroup of G. The (left)* cosets *of G are all sets of the form*

$$gH = \{gh : h \in H\},$$

for $g \in G$.

This defines an equivalence class on G, with

$$g \sim \hat{g}$$

if the set gH is equal to the set $\hat{g}H$, i.e., if there is an $h \in H$ with $gh = \hat{g}$. In a natural way, the *right cosets* are the sets

$$Hg = \{hg : h \in H\},$$

which also define an equivalence relation on the group G.

Definition 11.1.6 *A subgroup H is* normal *if for all g in G, $gHg^{-1} = H$.*

Theorem 11.1.1 *Let H be a subgroup of G. The set of cosets gH, under the binary operation*

$$gH \cdot \hat{g}H = g\hat{g}H,$$

will form a group if and only if H is a normal subgroup. (This group is denoted by G/H and pronounced G mod H.)

Sketch of Proof: Most of the steps are routine. The main technical difficulty lies in showing that the binary operation

$$(gH) \cdot (\hat{g}H) = (g\hat{g}H)$$

is well defined. Hence we must show that the set $gH \cdot \hat{g}H$, which consists of the products of all elements of the set gH with all elements of the set $\hat{g}H$, is equal to the set $g\hat{g}H$. Since H is normal, we have

$$\hat{g}H(\hat{g})^{-1} = H.$$

Then as sets

$$\hat{g}H = H\hat{g}.$$

Thus

$$gH\hat{g}H = g\hat{g}H \cdot H = g\hat{g}H,$$

since $H \cdot H = H$, as H is a subgroup. The map is well defined.

The identity element of G/H is $e \cdot H$. The inverse to gH is $g^{-1}H$. Associativity follows from the associativity of the group G. □

Note that in writing $gH \cdot \hat{g}H = g\hat{g}H$, one must keep in mind that H is representing every element in H and thus that H is itself not a single element.

As an application of this new group G/H, we now define the cyclic groups $\mathbf{Z}/n\mathbf{Z}$. Here our initial group is the integers $\mathbf{Z}$ and our subgroup consists of all the multiples of some fixed integer n:

$$n\mathbf{Z} = \{nk : k \in \mathbf{Z}\}.$$

Since the integers form an abelian group, every subgroup, including $n\mathbf{Z}$, is normal and thus $\mathbf{Z}/n\mathbf{Z}$ will form a group, It is common to represent each coset in $\mathbf{Z}/n\mathbf{Z}$ by an integer between 0 and $n-1$:

$$\mathbf{Z}/n\mathbf{Z} = \{0, 1, 2, \ldots, n-1\}.$$

For example, if we let $n = 6$, we have $\mathbf{Z}/6\mathbf{Z} = \{0, 1, 2, 3, 4, 5\}$. The addition table is then

+	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

An enjoyable exercise is proving the following critical theorem relating normal subgroups and group homomorphisms.

Theorem 11.1.2 *Let $\sigma : G \to \hat{G}$ be a group homomorphism. If*

$$ker(\sigma) = \{g \in G : \sigma(g) = \hat{e}, \text{ the identity of } \hat{G}\},$$

then $ker(\sigma)$ is a normal subgroup of G. (This subgroup $ker(\sigma)$ is called the kernel *of the map σ.)*

The study of groups is to a large extent the study of normal subgroups. By the above, this is equivalent to the study of group homomorphisms and is an example of the mid-twentieth century tack of studying an object by studying its homomorphisms.

The key theorem in finite group theory, Sylow's Theorem, links the existence of subgroups from the knowledge of the number of elements in a group.

Definition 11.1.7 *The* order *of a group G, denoted by $\mid G \mid$, is equal to the number of elements in G.*

For example, $\mid S_3 \mid = 6$.

Theorem 11.1.3 (Sylow's Theorem) *Let G be a finite group.*

a) Let p be a prime number. Suppose that p^α divides $\mid G \mid$. Then G has a subgroup of order p^α.

b) If p^n divides $|\, G\, |$ but p^{n+1} does not, then for any two subgroups H and $\hat{H}$ of order p^n, there is an element $g \in G$ with $gHg^{-1} = \hat{H}$.

c) If p^n divides $|\, G\, |$ but p^{n+1} does not, then the number of subgroups of order p^n is $1 + kp$, for some k a positive integer.

Proofs can be found in Herstein's *Topics in Algebra* [57], Section 2.12.

The importance lies in that we gather quite a bit of information about a finite group from merely knowing how many elements it has.

11.2 Representation Theory

Certainly one of the basic examples of groups is that of invertible $n \times n$ matrices. Representation theory studies how any given abstract group can be realized as a group of matrices. Since $n \times n$ matrices, via matrix multiplication on column vectors, are linear transformations from a vector space to itself, we can rephrase representation theory as the study of how a group can be realized as a group of linear transformations.

If V is a vector space, let $\mathbf{GL(V)}$ denote the group of linear transformations from V to itself.

Definition 11.2.1 *A* representation *of a group G on a vector space V is a group homomorphism*

$$\rho : G \to \mathbf{GL(V)}.$$

We say that ρ is a representation of G.

For example, consider the group $\mathbf{S}_3$ of permutations on three elements. There is quite a natural representation of $\mathbf{S}_3$ on three space $\mathbf{R}^3$. Let

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \in \mathbf{R}^3.$$

If $\sigma \in \mathbf{S}_3$, then define the map ρ by:

$$\rho(\sigma)\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_{\sigma(1)} \\ a_{\sigma(2)} \\ a_{\sigma(3)} \end{pmatrix}.$$

For example, if $\sigma = (12)$, then

$$\rho(12)\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_2 \\ a_1 \\ a_3 \end{pmatrix}.$$

As a matrix, we have:

$$\rho(12) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

If $\sigma = (123)$, then since (123) permutes (a_1, a_2, a_3) to (a_3, a_1, a_2), we have

$$\rho(123) \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_3 \\ a_1 \\ a_2 \end{pmatrix}.$$

As a matrix,

$$\rho(123) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

The explicit matrices representing the other elements of $\mathbf{S}_3$ are left as an exercise at the end of the chapter.

The goal of representation theory is to find all possible representations for a given group. In order to even be able to start to make sense out of this question, we first see how to build new representations out of old.

Definition 11.2.2 *Let G be a group. Suppose we have representations of G:*

$$\rho_1 : G \rightarrow \mathbf{GL(V_1)}$$

and

$$\rho_2 : G \rightarrow \mathbf{GL(V_2)},$$

where V_1 and V_2 are possibly different vector spaces. Then the direct sum representation *of G on $V_1 \oplus V_2$, denoted by*

$$(\rho_1 \oplus \rho_2) : G \rightarrow \mathbf{GL(V_1)} \oplus \mathbf{GL(V_2)},$$

is defined for all $g \in G$ by:

$$(\rho_1 \oplus \rho_2)(g) = \rho_1(g) \oplus \rho_2(g).$$

Note that when we write out $\rho_1(g) \oplus \rho_2(g)$ as a matrix, it will be in block diagonal form.

If we want to classify representations, we should concentrate on finding those representations that are not direct sums of other representations. This leads to:

Definition 11.2.3 *A representation ρ of a group G on a nonzero vector space V is* irreducible *if there is no proper subspace W of V such that for all $g \in G$ and all $w \in W$,*

$$\rho(g)w \in W.$$

In particular if a representation is the direct sum of two other representations, it will certainly not be irreducible. Tremendous progress has been made in finding all irreducible representations for many specific groups.

Representation theory occurs throughout nature. Any time you have a change of coordinate systems, suddenly representations appear. In fact, most theoretical physicists will even define an elementary particle (such as an electron) as an irreducible representation of some group (a group that captures the intrinsic symmetries of the world). For more on this, see Sternberg's *Group Theory and Physics* [106], especially the last part of Chapter 3.9.

11.3 Rings

If groups are roughly viewed as sets for which there is an addition, then rings are sets for which there is both an addition and a multiplication.

Definition 11.3.1 *A nonempty set R is a* ring *if there are two binary operations, denoted by $\cdot$ and $+$, on R such that*

a) R with $+$ forms an abelian group. The identity is denoted by 0.

b) (Associativity) for all $a, b, c \in R$, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.

c) (Distributivity) for all $a, b, c \in R$,

$$a \cdot (b + c) = a \cdot b + a \cdot c$$

and

$$(a + b) \cdot c = a \cdot c + b \cdot c.$$

Note that rings are not required to be commutative for the $\cdot$ operation or, in other words, we do not require $a \cdot b = b \cdot a$.

If there exists an element $1 \in R$ with $1 \cdot a = a \cdot 1 = a$ for all $a \in R$, we say that R is a ring with unit element. Almost all rings that are ever encountered in life will have a unit element.

The integers $\mathbf{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$, with the usual addition and multiplication, form a ring. Polynomials in one variable x with complex coefficients, denoted by $\mathbf{C}[x]$, form a ring with the usual addition and multiplication of polynomials. In fact, polynomials in n variables $\{x_1, \ldots, x_n\}$ with complex coefficients, denoted by $\mathbf{C}[x_1, \ldots, x_n]$, will also form a ring in the natural way. By the way, the study of the ring theoretic properties of $\mathbf{C}[x_1, \dots, x_n]$ is at the heart of much of algebraic geometry. While polynomials with complex coefficients are the most common to study, it is of course the case that polynomials with integer coefficients ($\mathbf{Z}[x_1, \dots, x_n]$), polynomials with rational coefficients ($\mathbf{Q}[x_1, \dots, x_n]$) and polynomials with real coefficients ($\mathbf{R}[x_1, \dots, x_n]$) are also rings. In fact, if R is any ring, then the polynomials with coefficients in R form a ring, denoted by $R[x_1, \dots, x_n]$.

Definition 11.3.2 *A function $\sigma : R \to \hat{R}$ between rings R and $\hat{R}$ is a* ring homomorphism *if for all $a, b \in R$,*

$$\sigma(a+b) = \sigma(a) + \sigma(b)$$

and

$$\sigma(a \cdot b) = \sigma(a) \cdot \sigma(b).$$

Definition 11.3.3 *A subset I of a ring R is an* ideal *if I is a subgroup of R under $+$ and if, for any $a \in R$, $aI \subset I$ and $Ia \subset I$.*

The notion of an ideal in ring theory corresponds to the notion of a normal subgroup in group theory. This analogy is shown in the following theorems:

Theorem 11.3.1 *Let $\sigma : R \to \hat{R}$ be a ring homomorphism. Then the set*

$$ker(\sigma) = \{a \in R : \sigma(a) = 0\}$$

is an ideal in R. (This ideal $ker(\sigma)$ is called the kernel *of the map σ.)*

Sketch of Proof: We need to use that for all $x \in \hat{R}$,

$$x \cdot 0 = 0 \cdot x = 0,$$

which is an exercise at the end of the chapter. Let $b \in \ker(\sigma)$. Thus $\sigma(b) = 0$. Given any element $a \in R$, we want $a \cdot b \in \ker(\sigma)$ and $b \cdot a \in \ker(\sigma)$. We have

$$\begin{aligned} \sigma(a \cdot b) &= \sigma(a) \cdot \sigma(b) \\ &= \sigma(a) \cdot 0 \\ &= 0, \end{aligned}$$

implying that $a \cdot b \in \ker(\sigma)$.

By a similar argument, $b \cdot a \in \ker(\sigma)$, showing that $\ker(\sigma)$ is indeed an ideal. □

Theorem 11.3.2 *Let I be an ideal in R. The sets $\{a + I : a \in R\}$ form a ring, denoted R/I, under the operations $(a+I) + (b+I) = (a+b+I)$ and $(a+I) \cdot (b+I) = (a \cdot b + I)$.*

The proof is left as a (long) exercise at the end of the chapter.

The study of a ring comes down to studying its ideals, or equivalently, its homomorphisms. Again, it's a mid-twentieth century approach to translate the study of rings to the study of maps between rings.

11.4 Fields and Galois Theory

We are now ready to enter the heart of classical algebra. To a large extent, the whole point of high school algebra is to find roots of linear and quadratic polynomials. With more complicated, but in spirit, similar techniques, the roots for third and fourth degree polynomials can also be found. One of the main historical motivations for developing the machinery of group and ring theory was in showing that there can be no similar techniques for finding the roots of polynomials of fifth degree or higher. More specifically the roots of a fifth degree or higher polynomial cannot be obtained by a formula involving radicals of the coefficients of the polynomial. (For an historical account, see Edwards' *Galois Theory* [31].)

The key is to establish a correspondence between one variable polynomials and finite groups. This is the essence of Galois Theory, which explicitly connects the ability to express roots as radicals of coefficients (in analogue to the quadratic equation) with properties of the associated group.

Before describing this correspondence, we need to discuss fields and field extensions.

Definition 11.4.1 *A ring R is a* field *if*

1. *R has a multiplicative unit 1,*
2. *for all $a, b \in R$ we have $a \cdot b = b \cdot a$ and*
3. *for any $a \neq 0$ in R, there is an element denoted by a^{-1} with $a \cdot a^{-1} = 1$.*

For example, since the integers **Z** do not have multiplicative inverses, **Z** is not a field. The rationals **Q**, the reals **R** and the complexes **C** are fields. For the ring $\mathbf{C}[x]$ of one variable polynomials, there corresponds the field $\mathbf{C}(x) = \{\frac{P(x)}{Q(x)} : P(x), Q(x) \in \mathbf{C}[x], Q(x) \neq 0\}$.

Definition 11.4.2 *A field $\hat{k}$ is a* field extension *of a field k if k is contained in $\hat{k}$.*

For example, the complex numbers **C** is a field extension of the real numbers **R**.

Once we have the notion of a field, we can form the ring $k[x]$ of one variable polynomials with coefficients in k. Basic, but deep, is:

Theorem 11.4.1 *Let k be a field. Then there is a field extension $\hat{k}$ of k such that every polynomial in $k[x]$ has a root in $\hat{k}$.*

Such a field $\hat{k}$ is said to be *algebraically closed.* For a proof, see Garling's *A Course in Galois Theory* [45], Section 8.2. As a word of warning, the proof uses the Axiom of Choice.

Before showing how groups are related to finding roots of polynomials, recall that the root of a linear equation $ax + b = 0$ is simply $x = -\frac{b}{a}$. For second degree equations, the roots of $ax^2 + bx + c = 0$ are of course

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Already interesting things are happening. Note that even if the three coefficients a, b and c are real numbers, the roots will be complex if the discriminant $b^2 - 4ac < 0$. Furthermore, even if the coefficients are rational numbers, the roots need not be rational, as $\sqrt{b^2 - 4ac}$ need not be rational.

Both of these observations lead naturally to extension fields of the field of coefficients. We will restrict to the case when the coefficients of our (monic) polynomial are rational numbers.

Let

$$P(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_0,$$

with each $a_k \in \mathbf{Q}$. By the Fundamental Theorem of Algebra (which states that the algebraic closure of the real numbers is the complex numbers), there are complex numbers $\alpha_1, \ldots, \alpha_n$ with

$$P(x) = (x - \alpha)(x - \alpha_2) \cdots (x - \alpha_n).$$

Of course, the whole problem is that the fundamental theorem does not tell us what the roots are. We would like an analogue of the quadratic equation for any degree polynomial. As mentioned before, such analogues do exist for cubic and quartic polynomials, but the punchline of Galois Theory is that no such analogue exists for degree five or higher polynomials. The proof of such a statement involves far more than the tools of high school algebra.

Here is a rapid fire summary of Galois Theory. We will associate to each one variable polynomial with rational coefficients a unique finite dimensional vector space over the rational numbers that is also a field extension of the rational numbers contained in the complex numbers. Namely, if $\alpha_1, \ldots, \alpha_n$ are the roots of the polynomial $P(x)$, the smallest field in the complex numbers that contains both the rationals and the roots $\alpha_1, \ldots, \alpha_n$ is the desired vector space. We then look at all linear transformations from this vector space to itself, with the strong restriction that the linear transformation is also a field automorphism mapping each rational number to itself. This is such a strong restriction that there are only a finite number of such transformations, forming a finite group. Further, each such linear transformation will not only map each root of $P(x)$ to another root but is actually determined by how it maps the roots to each other. Thus the finite group of these special linear transformations are a subgroup of the permutation group on n letters. The final deep result lies in showing that these finite groups determine properties about the roots.

Now for some details. We assume that $P(x)$ is irreducible in $\mathbf{Q}[x]$, meaning that $P(x)$ is not the product of any polynomials in $\mathbf{Q}[x]$. Hence none of the roots α_i of $P(x)$ can be rational numbers.

Definition 11.4.3 *Let $\mathbf{Q}(\alpha_1, \ldots, \alpha_n)$ be the smallest subfield of $\mathbf{C}$ containing both $\mathbf{Q}$ and the roots $\alpha_1, \ldots, \alpha_n$.*

Definition 11.4.4 *Let E be a field extension of $\mathbf{Q}$ but contained in $\mathbf{C}$. We say E is a* splitting field *if there is a polynomial $P(x) \in \mathbf{Q}[x]$ such that $E = \mathbf{Q}(\alpha_1, \ldots, \alpha_n)$, where $\alpha_1, \ldots, \alpha_n$ are the roots in $\mathbf{C}$ of $P(x)$.*

A splitting field E over the rational numbers $\mathbf{Q}$ is in actual fact a vector space over $\mathbf{Q}$. For example, the splitting field $\mathbf{Q}(\sqrt{2})$ is a two-dimensional vector space, since any element can be written uniquely as $a + b\sqrt{2}$, with $a, b \in \mathbf{Q}$.

Definition 11.4.5 *Let E be an extension field of $\mathbf{Q}$. The* group of automorphisms *G of E over $\mathbf{Q}$ is the set of all field automorphisms $\sigma : E \to E$.*

By *field automorphism* we mean a ring homomorphism from the field E to itself that is one-to-one, onto, maps unit to unit and whose inverse is a ring homomorphism. Note that field automorphisms of an extension field have the property that each rational number is mapped to itself (this is an exercise at the end of the chapter).

Such field automorphisms can be interpreted as linear transformations of E to itself. But not all linear transformations are field automorphisms, as will be seen in a moment.

Of course, there is needed here, in a complete treatment, a lemma showing that this set of automorphisms actually forms a group.

Definition 11.4.6 *Given an extension field E over $\mathbf{Q}$ with group of automorphisms G, the* fixed field *of G is the set $\{e \in E : \sigma(e) = e, \text{ for all } \sigma \in G\}$.*

Note that we are restricting attention to those field automorphisms that contain $\mathbf{Q}$ in the fixed field. Further it can be shown that the fixed field is indeed a subfield of E.

Definition 11.4.7 *A field extension E of $\mathbf{Q}$ is*normal *if the fixed field of the group of automorphisms G of E over $\mathbf{Q}$ is exactly $\mathbf{Q}$.*

Let G be the group of automorphisms of $\mathbf{Q}(\alpha_1, \ldots, \alpha_n)$ over $\mathbf{Q}$, where $\mathbf{Q}(\alpha_1, \ldots, \alpha_n)$ is the splitting field of the polynomial

$$\begin{aligned} P(x) &= (x-\alpha_1)(x-\alpha_2)\ldots(x-\alpha_n) \\ &= x^n + a_{n-1}x^{n-1} + \ldots + a_0, \end{aligned}$$

with each $a_k \in \mathbf{Q}$. This group G is connected to the roots of the polynomial $P(x)$, as seen in:

Theorem 11.4.2 *The group of automorphisms G is a subgroup of the permutation group $\mathbf{S}_n$ on n elements. It is represented by permuting the roots of the polynomial $P(x)$.*

Sketch of Proof: We will show that for any automorphism σ in the group G, the image of every root α_i is another root of $P(x)$. Therefore the automorphisms will merely permute the n roots of $P(x)$. It will be critical that $\sigma(a) = a$ for all rational numbers a. Now

$$\begin{aligned} P(\sigma(\alpha_i)) &= (\sigma(\alpha_i))^n + a_{n-1}(\sigma(\alpha_i))^{n-1} + \cdots + a_0 \\ &= \sigma(\alpha_i)^n + \sigma(a_{n-1}(\alpha_i)^{n-1}) + \cdots + \sigma(a_0) \\ &= \sigma((\alpha_i)^n + a_{n-1}(\alpha_i)^{n-1} + \cdots + a_0) \\ &= \sigma(P(\alpha_i)) \\ &= \sigma(0) \\ &= 0. \end{aligned}$$

Thus $\sigma(\alpha_i)$ is another root. To finish the proof, which we will not do, we would need to show that an automorphism σ in G is completely determined by its action on the roots α. □

All of this culminates in:

Theorem 11.4.3 (Fundamental Theorem of Galois Theory) *Let $P(x)$ be an irreducible polynomial in $\mathbf{Q}[x]$ and let $E = \mathbf{Q}(\alpha_1, \ldots, \alpha_n)$ be its splitting field with G the automorphism group of E.*

i) Each field B containing $\mathbf{Q}$ and contained in E is the fixed field of a subgroup of G. Denote this subgroup by G_B.

ii) The field extension B of $\mathbf{Q}$ is normal if and only if the subgroup G_B is a normal subgroup of G.

iii) The rank of E as a vector space over B is the order of G_B. The rank of B as a vector space over $\mathbf{Q}$ is the order of the group G/G_B.

Unfortunately, in this brevity, none of the implications should be at all clear. It is not even apparent why this should be called the Fundamental Theorem of the subject. A brief hint or whisper of its importance is that it sets up a dictionary between field extensions B with $\mathbf{Q} \subset \mathbf{B} \subset \mathbf{E}$ and subgroups G_B of G. A see-saw type diagram would be

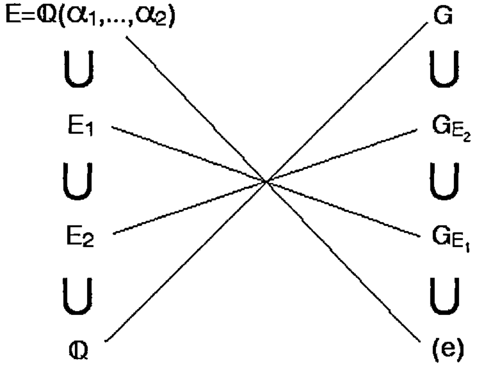

Here the lines connect subgroups with the corresponding fixed fields.

But what does this have to do with finding the roots of a polynomial. Our goal (which Galois Theory shows to be impossible) is to find an analogue of the quadratic equation. We need to make this more precise.

Definition 11.4.8 *A polynomial $P(x)$ is* solvable *if its splitting field $\mathbf{Q}(\alpha_1, \ldots, \alpha_n)$ lies in an extension field of $\mathbf{Q}$ obtained by adding radicals of integers.*

As an example, the field $\mathbf{Q}\left\{3\sqrt{2}, 5\sqrt{7}\right\}$ is obtained from $3\sqrt{2}$ and $5\sqrt{7}$, both of which are radicals. On the other hand, the field $\mathbf{Q}(\pi)$ is not obtained by adding radicals to $\mathbf{Q}$; this is a rewording of the deep fact that π is transcendental.

The quadratic equation $x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$ shows that each root of a second degree polynomial can be written in terms of a radical of its coefficients; hence every second degree polynomial is solvable. To show that no analogue of the quadratic equations exists for fifth degree or higher equations, all we need to show is that not all such polynomials are solvable. We want to describe this condition in terms of the polynomial's group of automorphisms.

Definition 11.4.9 *A finite group G is* solvable *if there is a nested sequence of subgroups $G_1, \ldots, G_n$ with $G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \ldots \supseteq G_n = (e)$, with each G_i normal in G_{i-1} and each G_{i-1}/G_i abelian.*

The link between writing roots as radicals and groups is contained in:

Theorem 11.4.4 *A polynomial $P(x)$ is solvable if and only if its associated group G of automorphisms of its splitting field is solvable.*

The impossibility of finding a clean formula for the roots of a high degree polynomial in terms of radicals of the coefficients now follows from showing that generically the group of automorphisms of an nth degree polynomial is the full permutation group $\mathbf{S}_n$ and

Theorem 11.4.5 *The permutation group on n elements, $\mathbf{S}_n$, is not solvable whenever n is greater than or equal to five.*

Of course, these are not obvious theorems. An excellent source for the proofs is Artins' *Galois Theory* [3].

Though there is no algebraic way of finding roots, there are many methods to approximate the roots. This leads to many of the basic techniques in numerical analysis.

11.5 Books

Algebra books went through quite a transformation starting in the 1930s. It was then that Van der Waerden wrote his algebra book *Modern Algebra* [113], which was based on lectures of Emmy Noether. The first undergraduate text mirroring these changes was *A Survey of Modern Algebra* [9], by Garrett Birkhoff and Saunders Mac Lane. The undergraduate text of the sixties and seventies was *Topics in Algebra* by Herstein [57]. Current popular choices are *A First Course in Abstract Algebra* by Fraleigh [41], and *Contemporary Abstract Algebra* by Gallian [43]. Serge Lang's *Algebra* [79] has been for a long time a standard graduate text, though it is not the place to start learning algebra. You will find, in your mathematical career, that you will read many texts by Lang. Jacobson's *Basic Algebra* [68], Artin's *Algebra* [4] and Hungerford's *Algebra* [65] are also good beginning graduate texts.

Galois Theory is definitely one of the most beautiful subjects in mathematics. Luckily there are a number of excellent undergraduate Galois Theory texts. One of the best (and cheapest) is Emil Artin's *Galois Theory* [3]. Other excellent texts are by Ian Stewart [107] and by Garling [45]. Edwards' *Galois Theory* [31] is an historical development. For beginning representation theory, I would recommend Hill's *Groups and Characters* [59] and Sternberg's *Group Theory and Physics* [106].

11.6 Exercises

1. Fix a corner of this book as the origin $(0,0,0)$ in space. Label one of the edges coming out of this corner as the x-axis, one as the y-axis and the last one as the z-axis. The goal of this exercise is to show that rotations do not commute. Let A denote the rotation of the book about the x-axis by ninety degrees and let B be the rotation about the y-axis by ninety degrees. Show with your book and by drawing pictures of your book that applying the rotation A and then rotation B is not the same as applying rotation B first and then rotation A.
2. Prove that the kernel of a group homomorphism is a normal subgroup.
3. Let R be a ring. Show that for all elements x in R,

$$x \cdot 0 = 0 \cdot x = 0,$$

even if the ring R is not commutative.
4. Let R be a ring and I an ideal in the ring. Show that R/I has a ring structure. (This is a long exercise, but it is an excellent way to nail down the basic definition of a ring.)
5. Show that the splitting field $\mathbf{Q}(\sqrt{2})$ over the rational numbers $\mathbf{C}$ is a two dimensional vector space over $\mathbf{C}$.
6. Start with the permutation group $\mathbf{S}_3$.
   a. Find all subgroups of $\mathbf{S}_3$.
   b. Show that the group $\mathbf{S}_3$ is solvable. (This allows us to conclude that for cubic polynomials there is an analogue of the quadratic equation.)
7. For each of the six elements of the group $\mathbf{S}_3$, find the corresponding matrices for the representation of $\mathbf{S}_3$ as described in section two of this chapter.
8. If H is a normal subgroup of a group G, show that there is a natural one-to-one correspondence between the left and the right cosets of H.
9. Let E be a field containing the rational numbers $\mathbf{Q}$. Let σ be a field automorphism of E. Note that this implies in particular that $\sigma(1) = 1$. Show that $\sigma(\frac{p}{q}) = \frac{p}{q}$ for all rational numbers $\frac{p}{q}$.
10. Let $T : G \to \tilde{G}$ be a group homomorphism. Show that $T(g^{-1}) = (T(g))^{-1}$ for all $g \in G$.
11. Let $T : G \to \tilde{G}$ be a group homomorphism. Show that the groups $G/\ker(T)$ and $Im(T)$ are isomorphic. Here $Im(T)$ denotes the image of the group G in the group $\tilde{G}$. This result is usually known as one of the Fundamental Homomorphism Theorems.

Chapter 12

Lebesgue Integration

Basic Object:	Measure Spaces
Basic Map:	Integrable Functions
Basic Goal:	Lebesgue Dominating Convergence Theorem

In calculus we learn about the Riemann integral of a function, which certainly works for many functions. Unfortunately, we must use the word 'many'. Lebesgue measure, and from this the Lebesgue integral, will allow us to define the right notion of integration. Not only will we be able to integrate far more functions with the Lebesgue integral but we will also understand when the integral of a limit of functions is equal to the limit of the integrals, i.e., when

$$\lim_{n\to\infty}\int f_n = \int \lim_{n\to\infty} f_n,$$

which is the Lebesgue Dominating Convergence Theorem. In some sense, the Lebesgue integral is the one that the gods intended us to use all along.

Our approach will be to develop the notion of Lebesgue measure for the real line $\mathbf{R}$, then use this to define the Lebesgue integral.

12.1 Lebesgue Measure

The goal of this section is to define the Lebesgue measure of a set E of real numbers. This intuitively means we want to define the length of E. For intervals

$$E = [a, b] = \{x \in \mathbf{R} : a \leq x \leq b\}$$

the length of E is simply:

$$\ell(E) = b - a.$$

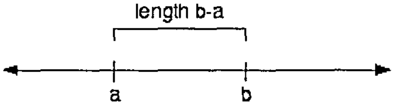

The question is to determine the length of sets that are not intervals, such as

$$E = \{x \in [0,1] : x \text{ is a rational number}\}.$$

We will heavily use that we already know the length of intervals. Let E be any subset of reals. A countable collection of intervals $\{I_n\}$, with each

$$I_n = [a_n, b_n],$$

covers the set E if

$$E \subset \bigcup I_n.$$

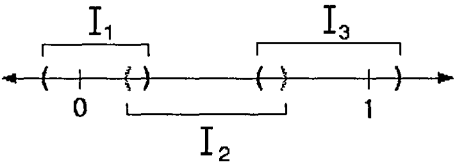

$E \subset I_1 \cup I_2 \cup I_3$

Whatever the length or measure of E is, it must be less than the sum of the lengths of the I_n.

Definition 12.1.1 *For any set E in* **R**, *the* outer measure *of E is*

$$m^*(E) = \inf\{\sum (b_n - a_n) : \textit{The collection of intervals } \{[a_n, b_n]\} \textit{ covers } E\}.$$

Definition 12.1.2 *A set E is* measurable *if for every set A,*

$$m^*(A) = m^*(A \cap E) + m^*(A - E).$$

The measure of a measurable set E, denoted by $m(E)$, is $m^*(E)$.

The reason for such a convoluted definition is that not all sets are measurable, though no one will ever construct a nonmeasurable set, since the existence of such a set requires the use of the Axiom of Choice, as we saw in Chapter Ten.

There is another method of defining a measurable set, via the notion of inner measure. Here we define the *inner measure* of a set E to be

$$m_*(E) = \sup(\sum(b_n - a_n) : E \supset \bigcup I_n \text{ and } I_n = [a_n, b_n] \text{ with } a_n \leq b_n\}.$$

Thus instead of covering the set E by a collection of open intervals, we fill up the inside of E with a collection of closed intervals.

If $m^*(E) < \infty$, then the set E can be shown to be measurable if and only if

$$m^*(E) = m_*(E).$$

In either case, we now have a way of measuring the length of almost all subsets of the real numbers.

As an example of how to use these definitions, we will show that the measure of the set of rational numbers (denoted here as E) between 0 and 1 is zero. We will assume that this set E is measurable and show its outer measure is zero. It will be critical that the rationals are countable. In fact, using this countability, list the rationals between zero and one as $a_1, a_2, a_3, \ldots$. Now choose an $\epsilon > 0$. Let I_1 be the interval

$$I_1 = [a_1 - \frac{\epsilon}{2}, a_1 + \frac{\epsilon}{2}].$$

Note that $\ell(I_1) = \epsilon$. Let

$$I_2 = [a_2 - \frac{\epsilon}{4}, a_2 + \frac{\epsilon}{4}].$$

Here $\ell(I_2) = \frac{\epsilon}{2}$. Let

$$I_3 = [a_3 - \frac{\epsilon}{8}, a_3 + \frac{\epsilon}{8}].$$

Here $\ell(I_3) = \frac{\epsilon}{4}$. In general let

$$I_k = [a_k - \frac{\epsilon}{2^k}, a_k + \frac{\epsilon}{2^k}].$$

Then $\ell(I_k) = \frac{\epsilon}{2^{k-1}}$.

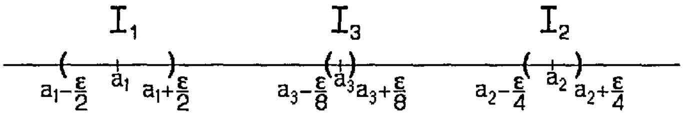
I_1
I_3
I_2
$a_1-\frac{\varepsilon}{2}$
a_1
$a_1+\frac{\varepsilon}{2}$
$a_3-\frac{\varepsilon}{8}$
a_3
$a_3+\frac{\varepsilon}{8}$
$a_2-\frac{\varepsilon}{4}$
a_2
$a_2+\frac{\varepsilon}{4}$

Certainly the rationals between zero and one are covered by this countable collection of open sets :

$$E \subset \bigcup I_n.$$

Then

$$\begin{aligned} m(E) &\leq \sum_{n=1}^{\infty} \ell(I_n) \\ &= \sum_{k=1}^{\infty} \frac{\epsilon}{2^{k-1}} \\ &= \epsilon \sum_{k=0}^{\infty} \frac{1}{2^k} \\ &= 2\epsilon. \end{aligned}$$

By letting ϵ approach zero, we see that $m(E) = 0$.

A similar argument can be used to show that the measure of any countable set is zero and in fact appears as an exercise at the end of this chapter.

12.2 The Cantor Set

While long a source of examples and counterexamples in real analysis, the Cantor set has recently been playing a significant role in dynamical systems. It is an uncountable, nowhere dense measure zero subset of the unit interval $[0, 1]$. By nowhere dense, we mean that the closure of the complement of the Cantor set will be the entire unit interval. We will first construct the Cantor set, then show that it is both uncountable and has measure zero.

For each positive integer k, we will construct a subset C_k of the unit interval and then define the Cantor set C to be

$$C = \bigcap_{k=1}^{\infty} C_k.$$

For $k = 1$, split the unit interval $[0, 1]$ into thirds and remove the open middle third, setting

$$\begin{aligned} C_1 &= [0,1] - (\frac{1}{3}, \frac{2}{3}) \\ &= [0, \frac{1}{3}] \bigcup [\frac{2}{3}, 1]. \end{aligned}$$

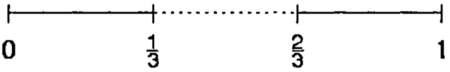

C_1

Take these two intervals and split them into thirds. Now remove each of their middle thirds to get

$$C_2 = [0, \frac{1}{9}] \bigcup [\frac{2}{9}, \frac{1}{3}] \bigcup [\frac{2}{3}, \frac{7}{9}] \bigcup [\frac{8}{9}, 1].$$

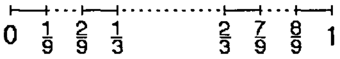

C_2

To get the next set C_3, split each of the four intervals of C_2 into three equal parts and remove the open middle thirds, to get eight closed intervals, each of length $\frac{1}{27}$. Continue this process for each k, so that each C_k consists of 2^k closed intervals, each of length $\frac{1}{3^k}$. Thus the length of each C_k will be

$$\text{length} = \frac{2^k}{3^k}.$$

The *Cantor set* C is the intersection of all of these C_k:

$$\text{Cantor set} = C = \bigcap_{k=1}^{\infty} C_k.$$

Part of the initial interest in the Cantor set was it was both uncountable and had measure zero. We will show first that the Cantor set has measure zero and then that it is uncountable. Since C is the intersection of all of the C_k, we get for each k that

$$m(C) < m(C_k) = \frac{2^k}{3^k}.$$

Since the fractions $\frac{2^k}{3^k}$ go to zero as k goes to infinity, we see that

$$m(C) = 0.$$

It takes a bit more work to show that the Cantor set is uncountable. The actual proof will come down to applying the trick of Cantor diagonalization, as discussed in Chapter Ten. The first step is to express any real number α in the unit interval $[0, 1]$ in its tri-adic expansion

$$\alpha = \sum_{k=1}^{\infty} \frac{n_k}{3^k},$$

where each n_k is zero, one or two. (This is the three-analog of the decimal expansion $\alpha = \sum_{k=1}^{\infty} \frac{n_k}{10^k}$, where here each $n_k = 0, 1, \ldots, 9$.) We can write the tri-adic expansion in base three notation, to get

$$\alpha = .n_1 n_2 n_3 \ldots$$

As with decimal expansion, the tri-adic expansion's coefficients n_k are unique, provided we always round up. Thus we will always say that

$$.102222\ldots = .11000\ldots$$

The Cantor set C has a particularly clean description in terms of the tri-adic or base three expansions. Namely

$$C = \{.n_1 n_2 n_3 \ldots \mid \text{each } n_k \text{ is either zero or two}\}.$$

Thus the effect of removing the middle thirds from all of the intervals corresponds to allowing no 1's among the coefficients. But then the Cantor set can be viewed as the set of infinite sequences of 0's and 2's, which was shown to be uncountable in the exercises of Chapter Ten.

12.3 Lebesgue Integration

One way to motivate integration is to try to find the area under curves. The Lebesgue integral will allow us to find the areas under some quite strange curves.

By definition the area of a unit square is one.

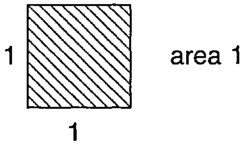

Hence the area of a rectangle with height b and base a will be ab.

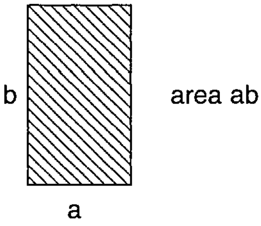

Let E be a measurable set on $\mathbf{R}$. Recall that the characteristic function of E, χ_E, is defined by

$$\chi_E(t) = \begin{cases} 1 & \text{if } t \in E \\ 0 & \text{if } t \in \mathbf{R} - E \end{cases}$$

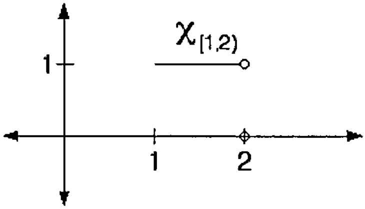

Since the height of χ_E is one, the area under the function (or curve) χ_E must be the length of E, or more precisely, $m(E)$. We denote this by $\int_E \chi_E$. Then the area under the function $a \cdot \chi_E$ must be $a \cdot m(E)$,

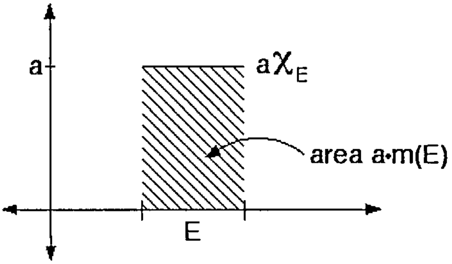

which we denote by $\int_E a\chi_E$.

Now let E and F be disjoint measurable sets. Then the area under the curve $a \cdot \chi_E + b \cdot \chi_F$ must be $a \cdot m(E) + b \cdot m(F)$,

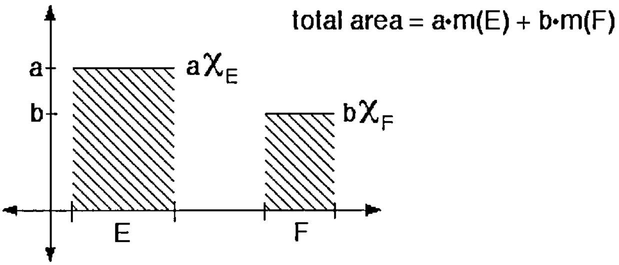

denoted by

$$\int_{E \cup F} a\chi_E + b\chi_F = a \cdot m(E) + b \cdot m(F).$$

For a countable collection of disjoint measurable sets A_i, the function

$$\sum a_i \chi_{Ai}$$

is called a *step function.* Let E be a measurable set. Let

$$\sum a_i \chi_{Ai}$$

be a step function. Then define

$$\int_E (\sum a_i \chi_{Ai}) = \sum a_i m(A_i \cap E).$$

We are about ready to define $\int_E f$.

Definition 12.3.1 *A function $f : E \to \mathbf{R} \cup (\infty) \cup (-\infty)$ is* measurable *if its domain E is measurable and if, for any fixed $\alpha \in \mathbf{R} \cup (\infty) \cup (-\infty)$,*

$$\{x \in E : f(x) = \alpha\}$$

is measurable.

Definition 12.3.2 *Let f be a measurable function on E. Then the* Lebesgue integral *of f on E is*

$$\int_E f = \inf\{\int_E \sum a_i \chi_{A_i} : \textit{for all } x \in E, \sum a_i \chi_{A_i}(x) \geq f(x)\}.$$

In pictures:

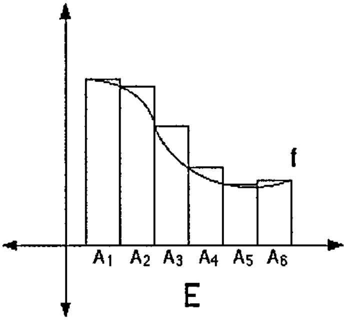

Thus we use that we know the integral for single step functions and then approximate the desired integral by summing the integrals of these step functions.

Every function that is integrable in beginning calculus is Lebesgue integrable. The converse is false, with the canonical counterexample given by the function $f : [0,1] \to [0,1]$ which is one at every rational and zero at every irrational. The Lebesgue integral is

$$\int_{[0,1]} f = 0,$$

which is one of the exercises at the end of the chapter, but this function has no Riemann integral, which is an exercise in Chapter Two.

12.4 Convergence Theorems

Not only does the Lebesgue integral allow us to integrate more functions than the calculus class (Riemann) integral, it also provides the right conditions to judge when we can conclude that

$$\int \lim_{k\to\infty} f_k = \lim_{k\to\infty} \int f_k.$$

In fact, if such a result were not true, we would have chosen another definition for the integral.

The typical theorem is of the form:

Theorem 12.4.1 (Lebesgue Dominating Convergence Thm.) *Let $g(x)$ be a Lebesgue integrable function on a measurable set E and let $\{f_n(x)\}$ be a sequence of Lebesgue integrable functions on E with $|f_k(x)| \leq g(x)$ for all x in E and such that there is a pointwise limit of the $f_k(x)$, i.e., there is a function $f(x)$ with*

$$f(x) = \lim_{k\to\infty} f_k(x).$$

Then

$$\int_E \lim_{k\to\infty} f_k(x) = \lim_{k\to\infty} \int_E f_k(x).$$

For a proof, see Royden's *Real Analysis* [95], Chapter 4, in section 4. We will just give a sketch here. Recall that if $f_k(x)$ converges uniformly to $f(x)$, then we know from ϵ and δ real analysis that

$$\lim_{k\to\infty} \int f_k(x) = \int f(x).$$

(i.e., the sequence of functions $f_k(x)$ converges uniformly to $f(x)$ if given any $\epsilon > 0$, there exists a positive integer N with

$$|f(x) - f_k(x)| < \epsilon,$$

for all x and all $k \geq N$. More quaintly, if we put an $\epsilon-$*tube* around $y = f(x)$, eventually the $y = f_k(x)$ will fall inside this tube.) The idea in the proof is that the $f_k(x)$ will indeed converge uniformly to $f(x)$, but only away from a subset of E of arbitrarily small measure. More precisely, the proposition we need is:

Proposition 12.4.1 *Let $\{f_n(x)\}$ be a sequence of measurable functions on a Lebesgue measurable set E, with $m(E) < \infty$. Suppose that $\{f_n(x)\}$ converges pointwise to a function $f(x)$. Then given $\epsilon > 0$ and $\delta > 0$, there is a positive integer N and a measurable set $A \subset E$ with $| f_k(x) - f(x) | < \epsilon$ for all $x \in E - A$ and $k \geq N$ and $m(A) < \delta$.*

The basic idea of the proof of the original theorem is now that

$$\begin{aligned} \int_E \lim_{n\to\infty} f_n &= \int_{E-A} \lim_{n\to\infty} f_n + \int_A \lim_{n\to\infty} f_n \\ &= \lim_{n\to\infty} \int_{E-A} f_n + max \mid g(x) \mid \cdot m(A). \end{aligned}$$

Since we can choose our set A to have arbitrarily small measure, we can let $m(A) \to 0$, which gives us our result.

The proposition can be seen to be true from the following. (After Royden's proof in Chapter 3, Section 6.) Set

$$G_n = \{x \in E : \mid f_n(x) - f(x) \mid \geq \epsilon\}.$$

Set

$$E_N = \bigcup_{n=N}^{\infty} G_n = \{x \in E : \mid f_n(x) - f(x) \mid \geq \epsilon, n \geq N\}.$$

Then $E_{N+1} \subset E_N$. Since we have $f_k(x)$ converging pointwise to $f(x)$, we must have $\cap E_n$, which can be thought of as the limit of the sets E_n, be empty. For measure to have any natural meaning, it should be true that $\lim_{N\to\infty} m(E_N) = 0$. Thus given $\delta > 0$, we can find an E_N with $m(E_N) < \delta$.

This is just an example of what can be accomplished with Lebesgue integration. Historically, the development of the Lebesgue integral in the early part of the twentieth century led quickly to many major advances. For example, until the 1920s, probability theory had no rigorous foundations. With the Lebesgue integral, and thus a correct way of measuring, the foundations were quickly laid.

12.5 Books

One of the first texts on measure theory was by Halmos [54]. This is still an excellent book. The book that I learned measure theory from was Royden's [95] and has been a standard since the 1960s. Rudin's book [96] is another excellent text. Frank Jones, one of the best teachers of mathematics in the country, has recently written a fine text [70]. Folland's recent text [40] is also quite good.

12.6 Exercises

1. Let E be any countable set of real numbers. Show that $m(E) = 0$.
2. Let $f(x)$ and $g(x)$ be two Lebesgue integrable functions, both with domain the set E. Suppose that the set

$$A = \{x \in E : f(x) \neq g(x)\}$$

has measure zero. What can be said about $\int_E f(x)$ and $\int_E g(x)$?
3. Let $f(x) = x$ for all real numbers x between zero and one and let $f(x)$ be zero everywhere else. We know from calculus that

$$\int_0^1 f(x)dx = \frac{1}{2}.$$

Show that this function $f(x)$ is Lebesgue integrable and that its Lebesgue integral is still $\frac{1}{2}$.
4. On the interval $[0, 1]$, define

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is not rational} \end{cases}.$$

Show that $f(x)$ is Lebesgue integrable, with

$$\int_0^1 f(x)dx = 0.$$

Chapter 13

Fourier Analysis

Basic Object:	Real-valued functions with a fixed period
Basic Maps:	Fourier transforms
Basic Goal:	Finding bases for vector spaces of periodic functions

13.1 Waves, Periodic Functions and Trigonometry

Waves occur throughout nature, from water pounding a beach to sound echoing off the walls at a club to the evolution of an electron's state in quantum mechanics. For these reasons, at the least, the mathematics of waves is important. In actual fact, the mathematical tools developed for waves, namely Fourier series (or harmonic analysis), touch on a tremendous number of different fields of mathematics. We will concentrate on only a small sliver and look at the basic definitions, how Hilbert spaces enter the scene, what a Fourier transform looks like and finally how Fourier transforms can be used to help solve differential equations.

Of course, a wave should look like:

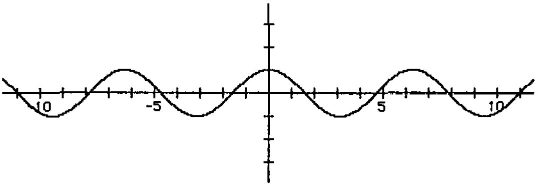

or

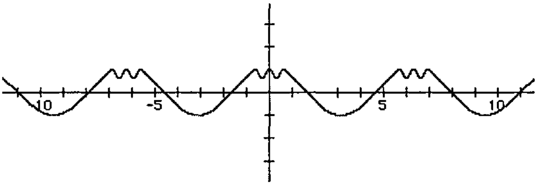

Both of these curves are described by periodic functions.

Definition 13.1.1 *A function* $f : \mathbf{R} \to \mathbf{R}$ *is* periodic *with* period L *if for all* x, $f(x+L) = f(x)$.

In other words, every L units the function must start to repeat itself. The quintessential periodic functions are the trigonometric functions $\cos(x)$ and $\sin(x)$, each with period 2π. Of course, functions like $\cos(\frac{2\pi x}{L})$ and $\sin(\frac{2\pi x}{L})$ are also periodic, both with period L.

Frequently people will say that a function $f(x)$ has period L if not only do we have that $f(x+L) = f(x)$, but also that there is no smaller number than L for which $f(x)$ is periodic. According to this convention, $\cos(x)$ will have period 2π but not period 4π, despite the fact that, for all x, $\cos(x+4\pi) = \cos(x)$. We will not follow this convention.

The central result in beginning Fourier series is that almost every periodic function is the, possibly infinite, sum of these trigonometric functions. Thus, at some level, the various functions $\cos(\frac{2\pi x}{L})$ and $\sin(\frac{2\pi x}{L})$ are not merely examples of periodic functions; they generate all periodic functions.

13.2 Fourier Series

Now to see how we can write a periodic function as an (infinite) sum of these cosines and sines. First suppose that we have a function $f : [-\pi, \pi] \to \mathbf{R}$ that has already been written as a series of sines and cosines, namely as

$$a_0 + \sum_{n=1}^{\infty} (a_n \cos(nx) + b_n \sin(nx)).$$

We want to see how we can naively compute the various coefficients a_k and b_k, ignoring all questions of convergence for these infinite series (convergence issues are faced in the next section). For any given k, consider

$$\int_{-\pi}^{\pi} f(x)\cos(kx)\mathrm{d}x \quad = \quad \int_{-\pi}^{\pi} (a_0 + \sum_{n=1}^{\infty} (a_n cos(nx) + b_n sin(nx)))\cos(kx)\mathrm{d}x$$

$$\begin{aligned} &= \int_{-\pi}^{\pi} a_0 \cos(kx)\mathrm{d}x \\ &\quad + \sum_{n=1}^{\infty} \int_{-\pi}^{\pi} \cos(nx)\cos(kx)\mathrm{d}x \\ &\quad + \sum_{n=1}^{\infty} \int_{-\pi}^{\pi} \sin(nx)\cos(kx)\mathrm{d}x. \end{aligned}$$

By direct calculation we have

$$\begin{aligned} \int_{-\pi}^{\pi} \cos(kx)\mathrm{d}x &= \begin{cases} 2\pi & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases} \\ \int_{-\pi}^{\pi} \cos(nx)\cos(kx)\mathrm{d}x &= \begin{cases} \pi & \text{if } k = n \\ 0 & \text{if } k \neq n \end{cases} \\ \int_{-\pi}^{\pi} \sin(nx)\cos(kx)\mathrm{d}x &= 0. \end{aligned}$$

Then we would expect

$$\int_{-\pi}^{\pi} f(x)\cos(kx)\mathrm{d}x = \begin{cases} 2\pi a_0 & \text{if } k = 0 \\ \pi a_n & \text{if } k \neq n \end{cases} .$$

By a similar calculation, using, though, the integrals $\int_{-\pi}^{\pi} f(x)\sin(nx)\mathrm{d}x$, we can get similar formulas for the b_n. This suggests how we could try to write any random periodic function as the infinite sum of sines and cosines:

Definition 13.2.1 *The* Fourier series *for a function* $f : [-\pi, \pi] \to \mathbf{R}$ *is*

$$a_0 + \sum_{n=1}^{\infty} (a_n \cos(nx) + b_n \sin(nx))$$

where

$$a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\mathrm{d}x$$

and

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\cos(nx)\mathrm{d}x$$

and

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\sin(nx)\mathrm{d}x.$$

The coefficients a_i and b_j are called the amplitudes, *or* Fourier coefficients *for the Fourier series.*

Of course, such a definition can only be applied to those functions for which the above integrals exist. The punchline, as we will see, is that most functions are actually equal to their Fourier series.

There are other ways of writing the Fourier series for a function. For example, using that $e^{ix} = \cos x + i \sin x$, for real numbers x, the Fourier series can also be expressed by

$$\sum_{n=-\infty}^{\infty} C_n e^{inx},$$

where

$$C_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{inx} \mathrm{d}x.$$

The C_n are also called the *amplitudes* or *Fourier coefficients.* In fact, for the rest of this section, but not for the rest of the chapter, we will write our Fourier series as $\sum_{n=-\infty}^{\infty} C_n e^{inx}$

The hope (which can be almost achieved) is that the function $f(x)$ and its Fourier series will be equal. For this, we must first put a slight restriction on the type of function we allow.

Theorem 13.2.1 *Let $f : [-\pi, \pi] \to \mathbf{R}$ be a square-integrable function. (i.e.,*

$$\int_{-\pi}^{\pi} |f(x)|^2 \mathrm{d}x < \infty.)$$

Then at almost all points,

$$f(x) = \sum_{n=-\infty}^{\infty} C_n e^{inx},$$

its Fourier series.

Note that this theorem contains within it the fact that the Fourier series of a square-integrable function will converge. Further, the above integral is the Lebesgue integral. Recall that almost everywhere means at all points except possibly for points in a set of measure zero. As seen in exercise 2 in Chapter Twelve, two functions that are equal almost everywhere will have equal integrals. Thus, morally, a square-integrable function is equal to its Fourier series.

What the Fourier series does is associate to a function an infinite sequence of numbers, the amplitudes. It explicitly gives how a function is the (infinite) sum of complex waves e^{inx}. Thus there is a map $\mathfrak{F}$ from square-integrable functions to infinite sequences of complex numbers,

$$\mathfrak{F}: \begin{array}{l}\text{Vector Space of square-} \\ \text{integrable functions}\end{array} \rightarrow \begin{array}{l}\text{Certain vector space of in-} \\ \text{finite sequences of com-} \\ \text{plex numbers}\end{array}$$

or

$$\mathfrak{F}: \begin{array}{l}\text{Vector Space of square-} \\ \text{integrable functions}\end{array} \rightarrow \begin{array}{l}\text{Vector space of infinite se-} \\ \text{quences of amplitudes}\end{array}$$

which, by the above theorem, is one-to-one, modulo equivalence of functions almost everywhere.

We now translate these statements into the language of Hilbert spaces, an extremely important class of vector space. Before giving the definition of a Hilbert space, a few definitions must be made.

Definition 13.2.2 *An* inner product $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbf{C}$ *on a complex vector space* V *is a map such that*

1. $\langle av_1 + bv_2, v_3 \rangle = a\langle v_1, v_3 \rangle + b\langle v_2, v_3 \rangle$ *for all complex numbers* $a, b \in \mathbf{C}$ *and for all vectors* $v_1, v_2, v_3, \in V$.

2. $\langle v, w \rangle = \overline{\langle w, v \rangle}$ *for all* $v, w \in V$.

3. $\langle v, v \rangle \geq 0$ *for all* $v \in V$ *and* $\langle v, v \rangle = 0$ *only if* $v = 0$.

Note that since $\langle v, v \rangle = \overline{\langle v, v \rangle}$, we must have, for all vectors v, that $\langle v, v \rangle$ is a real number. Hence the third requirement that $\langle v, v \rangle \geq 0$ makes sense.

To some extent, this is the complex vector space analogue of the dot product on $\mathbf{R}^n$. In fact, the basic example of an inner product on $\mathbf{C}^n$ is the following: let

$$\begin{aligned} v &= (v_1, \ldots, v_n) \\ w &= (w_1, \ldots, w_n) \end{aligned}$$

be two vectors in $\mathbf{C}^n$. Define

$$\langle v, w \rangle = \sum_{k=1}^{n} v_k \overline{w}_k.$$

It can be checked that this is an inner product on $\mathbf{C}^n$.

Definition 13.2.3 *Given an inner product* $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbf{C}$, *the* induced norm *on* V *is given by:*

$$|v| = \langle v, v \rangle^{1/2}.$$

In an inner product space, two vectors are *orthogonal* if their inner product is zero (which is what happens for the dot product in $\mathbf{R}^n$). Further, we can interpret the norm of a vector as a measure of the distance from the vector to the origin of the vector space. But then, with a notion of distance, we have a metric and hence a topology on V, as seen in Chapter Four, by setting

$$\rho(v,w) = |v-w|.$$

Definition 13.2.4 *A metric space (X,ρ) is* complete *if every Cauchy sequence converges, meaning that for any sequence $\{v_i\}$ in X with $\rho(v_i,v_j) \to 0$ as $i,j \to \infty$, there is an element v in X with $v_i \to v$ (i.e., $\rho(v,v_i) \to 0$ as $i \to \infty$).*

Definition 13.2.5 *A* Hilbert space *is an inner product space which is complete with respect to the topology defined by the inner product.*

There is the following natural Hilbert space.

Proposition 13.2.1 *The set of Lebesgue square-integrable functions*

$$\mathbf{L}^2[-\pi,\pi] = \{\mathbf{f} : [-\pi,\pi] \to \mathbf{C} \mid \int_{-\pi}^{\pi} |\mathbf{f}|^2 < \infty\}$$

is a Hilbert space, with inner product

$$\langle f,g \rangle = \int_{-\pi}^{\pi} f(x) \cdot \overline{g(x)} \mathrm{d}x.$$

This vector space is denoted by $L^2[-\pi,\pi]$.

We need to allow Lebesgue integrable functions in the above definition in order for the space to be complete.

In general, there is, for each real number $p \geq 1$ and any interval $[a,b]$, the vector space:

$$\mathbf{L}^p[a,b] = \{f : [a,b] \to \mathbf{R} \mid \int_a^b |f(x)|^p \mathrm{d}x < \infty\}.$$

The study of these vector spaces is the start of *Banach Space* theory.

Another standard example of a Hilbert space is the space of square-integrable sequences, denoted by l^2:

Proposition 13.2.2 *The set of sequences of complex numbers*

$$l^2 = \{(a_0, a_1, \ldots) \mid \sum_{j=0}^{\infty} |a_j|^2 < \infty\}$$

is a Hilbert space with inner product

$$\langle (a_0, a_1, \ldots), (b_0, b_1, \ldots) \rangle = \sum_{j=0}^{\infty} a_j \overline{b_j}.$$

We can now restate the fact that square-integrable functions are equal to their Fourier series, almost everywhere, into the language of Hilbert spaces.

Theorem 13.2.2 *For the Hilbert space* $\mathbf{L}^2[-\pi, \pi]$, *the functions*

$$\frac{1}{\sqrt{2\pi}} e^{inx}$$

are an orthonormal (Schauder) basis, meaning that each has length one, that they are pairwise orthogonal and that each element of $\mathbf{L}^2[-\pi, \pi]$ *is the unique infinite linear combination of the basis elements.*

Note that we had to use the technical term of *Schauder* basis. These are not quite the bases defined in Chapter One. There we needed each element in the vector space to be a unique finite linear combination of basis elements. While such do exist for Hilbert spaces, they do not seem to be of much use (the proof of their existence actually stems from the Axiom of Choice). The more natural bases are the above, for which we still require uniqueness of the coefficients but now allow infinite sums.

While the proof that the functions $\frac{1}{\sqrt{2\pi}} e^{inx}$ are orthonormal is simply an integral calculation, the proof that they form a basis is much harder and is in fact a restatement that a square-integrable function is equal to its Fourier series, namely:

Theorem 13.2.3 *For any function* $f(x)$ *in the Hilbert space* $\mathbf{L}^2[-\pi, \pi]$, *we have*

$$f(x) = \sum_{n=-\infty}^{\infty} \langle f(x), \frac{1}{\sqrt{2\pi}} e^{inx} \rangle \frac{1}{\sqrt{2\pi}} e^{inx},$$

almost everywhere.

Hence, the coefficients of a function's Fourier series are simply the inner product of $f(x)$ with each basis vector, exactly as with the dot product for vectors in $\mathbf{R}^3$ with respect to the standard basis $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. Further, we can view the association of a function with its Fourier coefficients (with its amplitudes) as a linear transformation

$$\mathbf{L}^2[-\pi,\pi] \to l^2.$$

Naturally enough, these formulas and theorems have versions for functions with period $2L$, when the Fourier series will be:

Definition 13.2.6 *A function $f : [-L, L] \to \mathbf{R}$ has Fourier series*

$$\sum_{n=-\infty}^{\infty} C_n e^{i\frac{n\pi x}{L}},$$

where

$$C_n = \frac{1}{2L}\int_{-L}^{L} f(x)e^{-\frac{in\pi x}{L}}\,\mathrm{d}x.$$

We have ignored, so far, a major subtlety, namely that a Fourier series is an infinite series. The next section deals with these issues.

13.3 Convergence Issues

Already during the 1700s mathematicians were trying to see if a given function was equal to its Fourier series, though in actual fact the theoretical tools needed to talk about such questions were not yet available, leading to some nonsensical statements. By the end of the 1800s, building on work of Dirichlet, Riemann and Gibbs, much more was known.

This section will state some of these convergence theorems. The proofs are hard. For notation, let our function be $f(x)$ and denote its Fourier series by

$$a_0 + \sum_{n=1}^{\infty}(a_n\cos(nx) + b_n\sin(nx)).$$

We want to know what this series converges to pointwise and to know when the convergence is uniform.

Theorem 13.3.1 *Let f(x) be continuous and periodic with period 2π. Then*

$$\lim_{N\to\infty}\int_{-\pi}^{\pi}\Big(f(x) - [a_0 + \sum_{n=1}^{N}(a_n\cos(nx) + b_n\sin(nx))]\Big)\mathrm{d}x = 0.$$

Thus for continuous functions, the area under the curve

$$y = \text{partial sum of the Fourier series}$$

will approach the area under the curve $y = f(x)$. We say that the Fourier series *converges in the mean* to the function $f(x)$.

This is telling us little about what the Fourier series converges to at any given fixed point x. Now assume that $f(x)$ is piecewise smooth on the closed interval $[-\pi, \pi]$, meaning that $f(x)$ is piecewise continuous, has a derivative at all but a finite number of points and that the derivative is piecewise continuous. For such functions, we define the one sided limits

$$f(x+) = \lim_{h \to 0 \text{ and } h>0} f(x+h)$$

and

$$f(x-) = \lim_{h \to 0 \text{ and } h>0} f(x-h).$$

Theorem 13.3.2 *If $f(x)$ is piecewise smooth on $[-\pi, \pi]$, then for all points x, the Fourier series converges pointwise to the function*

$$\frac{f(x+) + f(x-)}{2}.$$

At points where $f(x)$ is continuous, the one sided limits are of course each equal to $f(x)$. Thus for a continuous, piecewise smooth function, the Fourier series will converge pointwise to the function.

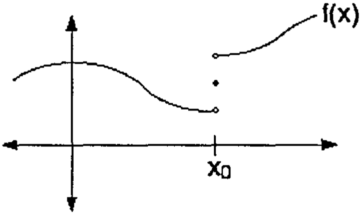

But when f is not continuous, even if it is piecewise smooth, the above pointwise convergence is far from uniform. Here the *Gibbs' phenomenon* becomes relevant. Denote the partial sum of the Fourier series by

$$S_N(x) = \frac{a_0}{2} + \sum_{n=1}^{N} (a_n \cos(nx) + b_n \sin(nx))$$

and suppose that f has a point of discontinuity at x_0. While the partial sums $S_N(x)$ do converge to $\frac{f(x+)+f(x-)}{2}$, the rate of convergence at different x is wildly different. In fact, the better the convergence is at the point of discontinuity x_0, the worse it is near x_0. In pictures, what happens is:

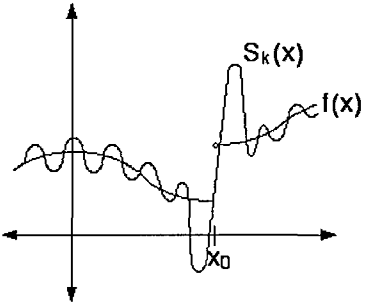

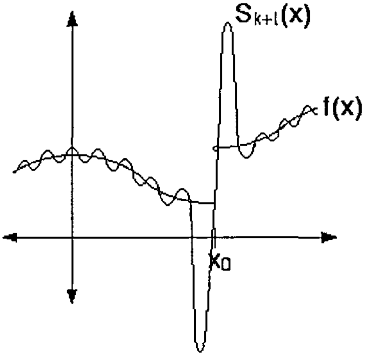

Note how the partial sums soar away from the function $f(x)$, destroying any hope of uniform convergence.

Luckily this does not happen if the function is continuous and piecewise smooth.

Theorem 13.3.3 *Let $f(x)$ be continuous and piecewise smooth on $[-\pi, \pi]$, with $f(-\pi) = f(\pi)$. Then the Fourier series will converge uniformly to $f(x)$.*

Thus for reasonably decent functions, we can safely substitute their Fourier series and still do basic calculus.

For proofs of these results, see Harry F. Davis' *Fourier Series and Orthogonal Functions* [24], chapter 3.

13.4 Fourier Integrals and Transforms

Most functions $f : \mathbf{R} \to \mathbf{R}$ will of course not be periodic, no matter what period L is chosen. But all functions, in some sense, are infinitely periodic. The Fourier integral is the result when we let the period L approach infinity (having as a consequence that $\frac{n\pi x}{L}$ approaches zero). The summation sign in the Fourier series becomes an integral. The result is:

Definition 13.4.1 *Let $f : \mathbf{R} \to \mathbf{R}$ be a function. Its* Fourier integral *is*

$$\int_0^{\infty} (a(t)\cos(tx) + b(t)\sin(tx))\mathrm{d}t,$$

where

$$a(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} f(x)\cos(tx)\mathrm{d}x$$

and

$$b(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} f(x)\sin(tx)\mathrm{d}x.$$

The Fourier integral can be rewritten as

$$\int_{-\infty}^{\infty} C(t)e^{itx}\mathrm{d}t,$$

where

$$C(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)e^{itx}\mathrm{d}x.$$

There are other forms, all equivalent up to constants.

The main theorem is:

Theorem 13.4.1 *Let $f : \mathbf{R} \to \mathbf{R}$ be integrable (i.e., $\int_{-\infty}^{\infty} |f(x)|\mathrm{d}x < \infty$). Then, off of a set of measure zero, the function $f(x)$ is equal to its Fourier integral.*

As with Fourier series, this integral is the Lebesgue integral. Further, again recall that by the term 'a set of measure zero', we mean a set of Lebesgue measure zero and that throughout analysis, sets of measure zero are routinely ignored.

As we will see, a large part of the usefulness of Fourier integrals lies in the existence of the Fourier transform.

Definition 13.4.2 *The* Fourier transform *of an integrable function $f(x)$ is:*

$$\Im(f(x))(t) = \int_{-\infty}^{\infty} f(x)e^{-itx}\mathrm{d}x.$$

The idea is that the Fourier transform can be viewed as corresponding to the coefficients a_n and b_n of a Fourier series and hence to the amplitude of the wave. By a calculation, we see that

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Im(f(x))(t)e^{itx}\mathrm{d}t,$$

provided we place suitable restrictions on the function $f(x)$. Thus indeed the Fourier transform is the continuous analogue of the amplitudes for Fourier series, in that we are writing the original function $f(x)$ as a sum (an integral) of the complex waves e^{itx} with coefficients given by the transform. (Also, the constant $\frac{1}{2\pi}$ is not fixed in stone; what is required is that the product of the constants in front of the integral in the Fourier transform (here it is 1) and the above integral be equal to $\frac{1}{2\pi}$.)

As we will see in the next section, in applications you frequently know the Fourier transform before you know the original function.

But for now we can view the Fourier transform as a one-to-one map

$$\Im : \text{Vector Space of Functions} \to \text{Different Vector Space of Functions.}$$

Thinking of the Fourier transform as an amplitude, we can rewrite this as:

$$\Im : \text{Position Space } \to \text{Amplitude Space.}$$

Following directly from the linearity of the Lebesgue integral, this map is linear.

Much of the power of Fourier transforms is that there is a dictionary between the algebraic and analytic properties of the functions in one of these vector spaces with those of the other vector space.

Proposition 13.4.1 *Let $f(x,t)$ be an integrable function with $f(x,t) \to 0$ as $x \to \pm\infty$. Let $\Im(f(x))(u)$ denote the Fourier transform with respect to the variable x. Then*

i) $\Im\{\frac{\partial f}{\partial x}\}(u) = iu\Im(f(x))(u)$.

ii) $\Im\{\frac{\partial^2 f}{\partial x^2}\}(u) = -u^2\Im(f(x))(u)$.

iii) $\Im\{\frac{\partial f(x,t)}{\partial t}\}(u) = \frac{\partial}{\partial t}\{\Im(f(x,t))\}(u)$.

We will show (i), where the key tool is simply integration by parts and sketch the proof of (iii).

By the definition of the Fourier transform, we have

$$\Im\{\frac{\partial f}{\partial x}\}(u) = \int_{-\infty}^{\infty} \frac{\partial f}{\partial x} e^{-iux}\mathrm{d}x,$$

which, by integration by parts, is

$$e^{-iux}f(x,t)\mid_{-\infty}^{\infty} +iu\int_{-\infty}^{\infty} f(x,t)e^{-iux}\mathrm{d}x = iu\int_{-\infty}^{\infty} f(x,t)e^{iux}\mathrm{d}x,$$

since $f(x,t) \to 0$ as $x \to \pm\infty$, and hence equals

$$iu\Im(f).$$

For (iii), we have

$$\Im\{\frac{\partial f(x,t)}{\partial t}\}(u) = \int_{-\infty}^{\infty} \frac{\partial f(x,t)}{\partial t}e^{-iux}\mathrm{d}x.$$

Since this integral is with respect to x and since the partial derivative is with respect to t, this is equal to:

$$\frac{\partial}{\partial t}\int_{-\infty}^{\infty} f(x,t)e^{-iux}\mathrm{d}x.$$

But this is just:

$$\frac{\partial}{\partial t}\{\Im(f(x,t))\}(u),$$

and thus (iii) has been shown. □

In the next section we will use this proposition to reduce the solving of a partial differential equation to the solving of an ordinary differential equation (which can almost always be solved). We need one more preliminary definition.

Definition 13.4.3 *The convolution of two functions $f(x)$ and $g(x)$ is*

$$(f*g)(x) = \int_{-\infty}^{\infty} f(u)g(x-u)\mathrm{d}u.$$

By a direct calculation, the Fourier transform of a convolution is the product of the Fourier transforms of each function, i.e.,

$$\Im(f*g) = \Im(f)\cdot\Im(g).$$

Thus the Fourier transform translates a convolution in the original vector space into a product in the image vector space. This will be important when trying to solve partial differential equations, in that at some stage we will have the product of two Fourier transforms, which we can now recognize as the Fourier transform of a single function, the convolution.

13.5 Solving Differential Equations

The idea is that the Fourier transform will translate a differential equation into a simpler one (one that is, vaguely, more algebraic). We will apply this technique to solving the partial differential equation that describes the flow of heat. Here the Fourier transform will change the partial differential equation into an ordinary differential equation, which can be solved. Once we know the Fourier transform, we can almost always recover the original function.

In the next chapter, we will derive the heat equation, but for now we will take as a given that the flow of heat through an infinitely thin, long bar is described by

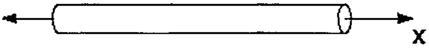

$$\frac{\partial h}{\partial t} = c\frac{\partial^2 h}{\partial x^2},$$

where $h(x,t)$ denotes the temperature at time t and position x and where c is a given constant. We start with an initial temperature distribution $f(x)$. Thus we want to find a function $h(x,t)$ that satisfies

$$\frac{\partial h}{\partial t} = c\frac{\partial^2 h}{\partial x^2},$$

given the initial condition,

$$h(x,0) = f(x).$$

Further, assume that as $x \to \pm\infty$, we know that $f(x) \to 0$. This just means basically that the bar will initially have zero temperature for large values of x. For physical reasons we assume that whatever is the eventual solution $h(x,t)$, we have that $h(x,t) \to 0$ as $x \to \pm\infty$.

Take the Fourier transform with respect to the variable x of the partial differential equation

$$\frac{\partial h}{\partial t} = k \cdot \frac{\partial^2 h}{\partial x^2},$$

to get

$$\Im(\frac{\partial h(x,t)}{\partial t})(u) = \Im(k \cdot \frac{\partial^2 h(x,t)}{\partial x^2})(u),$$

yielding

$$\frac{\partial}{\partial t}\Im(h(x,t))(u) = -ku^2\Im(h(x,t))(u).$$

Now $\Im(h(x,t))(u)$ is a function of the variables u and t. The x is a mere symbol, a ghost reminding us of the original PDE.

Treat the variable u as a constant, which is of course what we are doing when we take the partial derivative with respect to t. Then we can write the above equation in the form of an ODE:

$$\frac{\mathrm{d}}{\mathrm{d}t}\Im(h(x,t))(u) = -ku^2\Im(h(x,t))(u).$$

The solution to this ODE, as will be discussed in the next section but which can also be seen directly by (unpleasant) inspection, is:

$$\Im(h(x,t))(u) = C(u)e^{-ku^2t},$$

where $C(u)$ is a function of the variable u alone and hence, as far as the variable t is concerned, is a constant. We will first find this $C(u)$ by using the initial temperature $f(x)$. We know that $h(x,0) = f(x)$. Then for $t = 0$,

$$\Im(h(x,0))(u) = \Im(f(x))(u).$$

When $t = 0$, the function $C(u)e^{-ku^2t}$ is just $C(u)$ alone. Thus when $t = 0$, we have

$$\Im(f(x))(u) = C(u).$$

Since $f(x)$ is assumed to be known, we can actually compute its Fourier transform and thus we can compute $C(u)$. Thus

$$\Im(h(x,t))(u) = \Im(f(x))(u) \cdot e^{-ku^2t}.$$

Assume for a moment that we know a function $g(x,t)$ such that its Fourier transform with respect to x is:

$$\Im(g(x,t))(u) = e^{-ku^2t}.$$

If such a function $g(x,t)$ exists, then

$$\Im(h(x,t))(u) = \Im(f(x))(u) \cdot \Im(g(x,t))(u).$$

But a product of two Fourier transforms can be written as the Fourier transform of a convolution. Thus

$$\Im(h(x,t))(u) = \Im(f(x) * g(x,t)).$$

Since we can recover that original function from its Fourier transform, this means that the solution to the heat equation is

$$h(x,t) = f(x) * g(x,t).$$

Thus we can solve the heat equation if we can find this function $g(x,t)$ whose Fourier transform is e^{-ku^2t}. Luckily we are not the first people to attempt this approach. Over the years many such calculations have been done and tables have been prepared, listing such functions. (To do it oneself, one needs to define the notion of the inverse Fourier transform and then to take the inverse Fourier transform of the function e^{-ku^2t}; while no harder than the Fourier transform, we will not do it.) However it is done, we can figure out that

$$\Im\left(\frac{1}{\sqrt{4\pi kt}}e^{\frac{-x^2}{4kt}}\right) = e^{-ku^2t}.$$

Thus the solution of the heat equation will be:

$$h(x,t) = f(x) * \frac{1}{\sqrt{4\pi kt}}e^{\frac{-x^2}{4kt}}.$$

13.6 Books

Since Fourier analysis has applications ranging from CAT scans to questions about the distribution of the prime numbers, it is not surprising that there are books on Fourier series aimed at wildly different audiences and levels of mathematical maturity. Barbara Hubbard's *The World According to Wavelets* [63] is excellent. The first half is a gripping nontechnical description of Fourier series. The second half deals with the rigorous mathematics. Wavelets, by the way, are a recent innovation in Fourier series that have had profound practical applications. A solid, traditional introduction is given by Davis in his *Fourier Series and Orthogonal Functions* [24]. A slightly more advanced text is Folland's *Fourier Analysis and its Applications* [38]. A brief, interesting book is Seeley's *An Introduction to Fourier Series and Integrals* [98]. An old fashioned but readable book is Jackson's *Fourier Series and Orthogonal Polynomials* [67]. For the hardcore student, the classic inspiration in the subject since the 1930s has been Zygmund's *Trigonometric Series* [116].

13.7 Exercises

1. On the vector space

$$\mathbf{L}^2[-\pi,\pi] = \{f : [-\pi,\pi] \to \mathbf{C} \mid \int_{-\pi}^{\pi} |f|^2 < \infty\},$$

show that

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x) \cdot \overline{g(x)} \mathrm{d}x$$

is indeed an inner product, as claimed in this chapter.

2. Using Fourier transforms, reduce the solution of the wave equation

$$\frac{\partial^2 y}{\partial t^2} = k\frac{\partial^2 y}{\partial x^2},$$

with k a constant, to solving an ordinary (no partial derivatives involved) differential equation.

3. Consider the functions

$$f_n(x) = \begin{cases} 2n & \text{if } \frac{-1}{n} < x < \frac{1}{n} \\ 0 & \text{otherwise} \end{cases} .$$

Compute the Fourier transforms of each of the functions $f_n(x)$. Graph each of the functions f_n and each of the Fourier transforms. Compare the graphs and draw conclusions.

Chapter 14

Differential Equations

Basic Object:	Differential Equations
Basic Goal:	Finding Solutions to Differential Equations

14.1 Basics

A differential equation is simply an equation, or a set of equations, whose unknowns are functions which must satisfy (or solve) an equation involving both the function and its derivatives. Thus

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 3y$$

is a differential equation whose unknown is the function $y(x)$. Likewise,

$$\frac{\partial^2 y}{\partial x^2} - \frac{\partial^2 y}{\partial x \partial t} + \frac{\partial y}{\partial x} = x^3 + 3yt$$

is a differential equation with the unknown being the function of two variables $y(x,t)$. Differential equations fall into two broad classes: ordinary and partial. Ordinary differential equations (ODEs) are those for which the unknown functions are functions of only one independent variable. Thus $\frac{\mathrm{d}y}{\mathrm{d}x} = 3y$ and

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + \frac{\mathrm{d}y}{\mathrm{d}x} + \sin(x)y = 0$$

are both ordinary differential equations. As will be seen in the next section, these almost always have, in principle, solutions.

Partial differential equations (PDEs) have unknowns that are functions of more than one variable, such as

$$\frac{\partial^2 y}{\partial x^2} - \frac{\partial^2 y}{\partial t^2} = 0$$

and

$$\frac{\partial^2 y}{\partial x^2} + \left(\frac{\partial y}{\partial t}\right)^3 = \cos(xt).$$

Here the unknown is the function of two variables $y(x, t)$. For PDEs, everything is much murkier as far as solutions go. We will discuss the method of separation of variables and the method of clever change of variables (if this can be even called a method). A third method, discussed in Chapter Thirteen, is to use Fourier transforms.

There is another broad split in differential equations: linear and nonlinear. A differential equation is *homogeneous linear* if given two solutions f_1 and f_2 and any two numbers λ_1 and λ_2, then the function

$$\lambda_1 f_1 + \lambda_2 f_2$$

is another solution. Thus the solutions will form a vector space. For example, $\frac{\partial^2 y}{\partial x^2} - \frac{\partial^2 y}{\partial t^2} = 0$ is homogeneous linear. The differential equation is *linear* if by subtracting off from the differential equation a function of the independent variables alone changes it into a homogeneous linear differential equation. The equation $\frac{\partial^2 y}{\partial x^2} - \frac{\partial^2 y}{\partial t^2} = x$ is linear, since if we subtract off the function x we have a homogeneous linear equation. The important fact about linear differential equations is that their solution spaces form linear subspaces of vector spaces, allowing linear algebraic ideas to be applied. Naturally enough a *nonlinear* differential equation is one which is not linear.

In practice, one expects to have differential equations arise whenever one quantity varies with respect to another. Certainly the basic laws of physics are written in terms of differential equations. After all, Newton's second law:

$$\text{Force} = (\text{mass}) \cdot (\text{acceleration})$$

is the differential equation

$$\text{Force} = (\text{mass}) \cdot \left(\frac{\mathrm{d}^2(\textit{position})}{\mathrm{d}x^2}\right).$$

14.2 Ordinary Differential Equations

In solving an ordinary differential equation, one must basically undo a derivative. Hence solving an ordinary differential equation is basically the same as performing an integral. In fact, the same types of problems occur in ODEs and in integration theory.

Most reasonable functions (such as continuous functions) can be integrated. But to actually recognize the integral of a function as some other, well-known function (such as a polynomial, trig function, inverse trig function, exponential or log) is usually not possible. Likewise with ODEs, while almost all have solutions, only a handful can be solved cleanly and explicitly. Hence the standard sophomore-level engineering-type ODE course must inherently have the feel of a bag of tricks applied to special equations.[1]

In this section we are concerned with the fact that ODEs have solutions and that, subject to natural initial conditions, the solutions will be unique. We first see how the solution to a single ODE can be reduced to solving a system of first order ODEs, which are equations with unknown functions $y_1(x), \ldots, y_n(x)$ satisfying

$$\begin{aligned}\frac{\mathrm{d}y_1}{\mathrm{d}x} &= f_1(x, y_1, \ldots, y_n)\\ &\vdots\\ \frac{\mathrm{d}y_n}{\mathrm{d}x} &= f_n(x, y_1, \ldots, y_n)\end{aligned}$$

Start with a differential equation of the form:

$$a_n(x)\frac{\mathrm{d}^n y}{\mathrm{d}x^n} + \ldots + a_1(x)\frac{\mathrm{d}y}{\mathrm{d}x} + a_0(x)y(x) + b(x) = 0.$$

We introduce new variables:

$$\begin{aligned}y_0(x) &= y(x)\\ y_1(x) &= \frac{\mathrm{d}y_0}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}x}\\ y_2(x) &= \frac{\mathrm{d}y_1}{\mathrm{d}x} = \frac{\mathrm{d}^2 y_0}{\mathrm{d}x^2} = \frac{\mathrm{d}^2 y}{\mathrm{d}x^2}\\ y_{n-1}(x) &= \frac{\mathrm{d}y_{n-2}}{\mathrm{d}x} = \ldots = \frac{\mathrm{d}^{n-1}y_0}{\mathrm{d}x^{n-1}} = \frac{\mathrm{d}^{n-1}y}{\mathrm{d}x^{n-1}}.\end{aligned}$$

Then a solution $y(x)$ to the original ODE will give rise to a solution of the following system of first order ODEs:

$$\begin{aligned}\frac{\mathrm{d}y_0}{\mathrm{d}x} &= y_1\\ \frac{\mathrm{d}y_1}{\mathrm{d}x} &= y_2\\ &\vdots\end{aligned}$$

---

[1] There are reasons and patterns structuring the bag of tricks. These involve a careful study of the underlying symmetries of the equations. For more, see Peter Olver's *Applications of Lie Groups to Differential Equations* [90].

$$\frac{\mathrm{d}y_{n-1}}{\mathrm{d}x} = -\frac{1}{a_n(x)}(a_{n-1}(x)y_{n-1} + a_{n-2}(x)y_{n-2} + \ldots + a_0(x)y_0 + b(x)).$$

If we can solve all such systems of first order ODEs, we can then solve all ODEs. Hence the existence and uniqueness theorems for ODEs can be couched in the language of systems of first order ODEs.

First to define the special class of functions we are interested in.

Definition 14.2.1 *A function $f(x, y_1, \ldots, y_n)$ defined on a region T in $\mathbf{R}^{n+1}$ is* Lipschitz *if it is continuous and if there is a constant N such that for every $(x, y_1, \ldots, y_n)$ and $(\hat{x}, \hat{y_1}, \ldots, \hat{y_n})$ in T, we have*

$$|f(x, y_1, \ldots, y_n) - f(\hat{x}, \hat{y_1}, \ldots, \hat{y_n})| \leq N \cdot (|y_1 - \hat{y_1}| + \ldots + |y_n - \hat{y_n}|).$$

It is not a major restriction on a function to require it to be Lipschitz. For example, any function with continuous first partial derivatives on an open set will be Lipschitz on any connected compact subset.

Theorem 14.2.1 *A system of first order ordinary differential equations*

$$\begin{aligned} \frac{\mathrm{d}y_1}{\mathrm{d}x} &= f_1(x, y_1, \ldots, y_n) \\ &\vdots \\ \frac{\mathrm{d}y_n}{\mathrm{d}x} &= f_n(x, y_1, \ldots, y_n), \end{aligned}$$

with each function $f_1, \ldots, f_n$ being Lipschitz in a region T, will have, for each real number x_0, an interval $(x_0 - \epsilon, x_0 + \epsilon)$ on which there are solutions $y_1(x), \ldots, y_n(x)$. Further, given numbers $a_1, \ldots, a_n$, with $(x_0, a_1, \ldots, a_n)$ in the region T, the solutions satisfying the initial conditions

$$\begin{aligned} y_1(x_0) &= a_1 \\ &\vdots \\ y_n(x_0) &= a_n \end{aligned}$$

are unique.

Consider a system of two first order ODEs:

$$\frac{\mathrm{d}y_1}{\mathrm{d}x} = f_1(x, y_1, y_2)$$

$$\frac{\mathrm{d}y_2}{\mathrm{d}x} = f_2(x, y_1, y_2).$$

Then a solution $(y_1(x), y_2(x))$ will be a curve in the plane $\mathbf{R}^2$. The theorem states that there is exactly one solution curve passing through any given point (a_1, a_2). In some sense the reason why ODEs are easier to solve than PDEs is that we are trying to find solution curves for ODEs (a one-dimensional type problem) while for PDEs the solution sets will have higher dimensions and hence far more complicated geometries.

We will set up the *Picard Iteration* for finding solutions and then briefly describe why this iteration actually works in solving the differential equations.

For this iterative process, functions $y_{1_k}(x), \ldots, y_{n_k}(x)$ will be constructed that will approach the true solutions $y_1(x), \ldots, y_n(x)$. Start with setting

$$y_{i_0}(x) = a_i$$

for each i. Then, at the k^{th} step, define

$$y_{1_k}(x) = a_1 + \int_{x_0}^{x} f_1(t, y_{1_{k-1}}(t), \ldots, y_{n_{k-1}}(t))\mathrm{d}t$$

$$\vdots$$

$$y_{n_k}(x) = a_n + \int_{x_0}^{x} f_n(t, y_{1_{k-1}}(t), \ldots, y_{n_{k-1}}(t))\mathrm{d}t.$$

The crucial part of the theorem is that each of these converges to a solution. The method is to look at the sequence, for each i,

$$y_{i_0}(x) + \sum_{k=1}^{\infty}(y_{i_k}(x) - y_{i_{k-1}}(x)),$$

which has as its N^{th} partial sum the function $y_{i_N}(x)$. To show that this sequence converges comes down to showing that

$$|y_{i_k}(x) - y_{i_{k-1}}(x)|$$

approaches zero quickly enough. But this absolute value is equal to

$$\left|\int_{x_0}^{x} [f_i(t, y_{1_{k-1}}(t), \ldots, y_{n_{k-1}}(t)) - f_i(t, y_{1_{k-2}}(t), \ldots, y_{n_{k-2}}(t))]\mathrm{d}t\right|$$

$$\leq \int_{x_0}^{x} |f_i(t, y_{1_{k-1}}(t), \ldots, y_{n_{k-1}}(t)) - f_i(t, y_{1_{k-2}}(t), \ldots, y_{n_{k-2}}(t))|\mathrm{d}t.$$

The last integral's size can be controlled by applying the Lipschitz conditions and showing that it approaches zero.

14.3 The Laplacian

14.3.1 Mean Value Principle

In $\mathbf{R}^n$, the *Laplacian* of a function $u(x) = u(x_1, \ldots, x_n)$ is

$$\triangle u = \frac{\partial^2 u}{\partial {x_1}^2} + \ldots + \frac{\partial^2 u}{\partial {x_n}^2}.$$

One can check that the PDE

$$\triangle u = 0$$

is homogeneous and linear and thus that the solutions form a vector space. These solutions are important enough to justify their own name.

Definition 14.3.1 *A function $u(x) = u(x_1, \ldots, x_n)$ is* harmonic *if $u(x)$ is a solution to the Laplacian:*

$$\triangle u = 0.$$

Much of the importance of the Laplacian is that its solutions, harmonic functions, satisfy the Mean Value Principle, which is our next topic. For any point $a \in \mathbf{R}^n$, let

$$S_a(r) = \{x \in \mathbf{R}^n : |x - a| = r\},$$

be the sphere of radius r centered at a.

Theorem 14.3.1 (Mean Value Principle) *If $u(x) = u(x_1, \ldots, x_n)$ is harmonic, then at any point $a \in \mathbf{R}^n$,*

$$u(a) = \frac{1}{\text{area of } S_a(r)} \int_{S_a(r)} u(x).$$

Thus $u(a)$ is equal to the average value of $u(x)$ on any sphere centered at a. For a proof of the case when n is two, see almost any text on complex analysis. For the general case, see G. Folland's *Introduction to Partial Differential Equations* [39], section 2.A.

Frequently, in practice, people want to find harmonic functions on regions subject to given boundary conditions. This is called:

The Dirichlet Problem: Let R be a region in $\mathbf{R}^n$ with boundary ∂R. Suppose that g is a function defined on this boundary. The *Dirichlet Problem* is to find a function f on R satisfying

$$\triangle f = 0$$

on R and

$$f = g$$

on ∂R.

One way this type of PDE arises naturally in classical physics is as a potential. It is also the PDE used to study a steady-state solution of the heat equation. We will see in the next section that heat flow satisfies the PDE:

$$\frac{\partial^2 u}{\partial x_1^2} + \ldots + \frac{\partial^2 u}{\partial x_n^2} = c \cdot \frac{\partial u}{\partial t},$$

where $u(x_1, \ldots, x_n, t)$ denotes the temperature at time t at place $(x_1, \ldots, x_n)$. By a steady-state solution, we mean a solution that does not change over time, hence a solution with

$$\frac{\partial u}{\partial t} = 0.$$

Thus a steady state solution will satisfy

$$\triangle u = \frac{\partial^2 u}{\partial x_1^2} + \ldots + \frac{\partial^2 u}{\partial x_n^2} = 0,$$

and hence is a harmonic function.

14.3.2 Separation of Variables

There are a number of ways of finding harmonic functions and of solving the Dirichlet Problem, at least when the involved regions are reasonable. Here we discuss the method of separation of variables, a method that can also frequently be used to solve the heat equation and the wave equation. By the way, this technique does not always work.

We will look at a specific example and try to find the solution function $u(x, y)$ to

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0,$$

on the unit square, with boundary conditions

$$u(x, y) = \begin{cases} h(x) & \text{if } y = 1 \\ 0 & \text{if } x = 0,\ x = 1 \text{ or } y = 0 \end{cases}$$

where $h(x)$ is some initially specified function defined on the top side of the square.

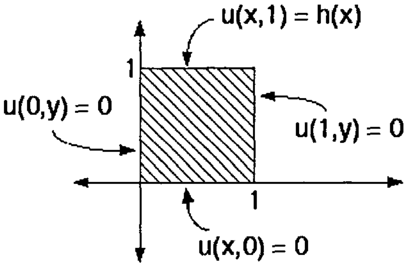

The key assumption will be that the solution will be of the form

$$u(x,y) = f(x) \cdot g(y),$$

where

$$f(0) = 0,\ g(0) = 0,\ f(1) = 0,\ f(x) \cdot g(1) = h(x).$$

This is wild. Few two-variable functions can be written as the product of two functions, each a function of one-variable alone. The only possible justification is if we can actually find such a solution, which is precisely what we will do. (To finish the story, which we will not do, we would need to prove that this solution is unique.) If $u(x,y) = f(x) \cdot g(y)$ and if $\Delta u = 0$, then we need

$$\frac{\mathrm{d}^2 f}{\mathrm{d}x^2} g(y) + f(x) \frac{\mathrm{d}^2 g}{\mathrm{d}y^2} = 0.$$

Thus we would need

$$\frac{\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}}{f(x)} = -\frac{\frac{\mathrm{d}^2 g}{\mathrm{d}y^2}}{g(y)}.$$

Each side depends on totally different variables, hence each must equal to a constant. Using the boundary conditions $f(0) = f(1) = 0$, one can show that this constant must be negative. We denote it by $-c^2$. Thus we need

$$\frac{\mathrm{d}^2 f}{\mathrm{d}x^2} = -c^2 f(x)$$

and

$$\frac{\mathrm{d}^2 g}{\mathrm{d}y^2} = c^2 g(y),$$

both second order ODEs, which have solutions

$$f(x) = \lambda_1 \cos(cx) + \lambda_2 \sin(cx)$$

and

$$g(y) = \mu_1 e^{cy} + \mu_2 e^{-cy}.$$

We now apply the boundary conditions. We have that $f(0) = 0$, which implies that

$$\lambda_1 = 0.$$

Also $g(0) = 0$ forces

$$\mu_1 = -\mu_2$$

and $f(1) = 0$ means that

$$\lambda_2 \sin(cx) = 0.$$

This condition means that the constant c must be of the form

$$c = k\pi, \text{ with } k = 0, 1, 2, \ldots.$$

Hence the solution must have the form

$$u(x, y) = f(x) \cdot g(y) = C_k \sin(k\pi x)(e^{k\pi y} - e^{-k\pi y}),$$

with C_k some constant.

But we also want $u(x, 1) = h(x)$. Here we need to use that the Laplacian is linear and thus that solutions can be added. By adding our various solutions for particular $c = k\pi$, we set

$$u(x, y) = \sum C_k (e^{k\pi y} - e^{-k\pi y}) \sin(k\pi x).$$

All that is left is to find the constants C_k. Since we require $u(x, 1) = h(x)$, we must have

$$h(x) = \sum C_k (e^{k\pi} - e^{-k\pi}) sin(k\pi x).$$

But this is a series of sines. By the Fourier analysis developed in the last chapter, we know that

$$C_k (e^{k\pi} - e^{-k\pi}) = 2 \int_0^1 h(x) \sin(k\pi x) \mathrm{d}x = \frac{2h(x)(1 - \cos k\pi)}{k\pi}.$$

Thus the solution is

$$u(x, y) = \frac{2h(x)}{\pi} \sum_{k=1}^{\infty} \frac{1 - \cos k\pi}{k(e^{k\pi} - e^{-k\pi})} \sin(k\pi x)(e^{k\pi y} - e^{-k\pi y}).$$

While not pleasant looking, it is an exact solution.

14.3.3 Applications to Complex Analysis

We will now quickly look at an application of harmonic functions. The goal of Chapter Nine was the study of complex analytic functions $f : U \to \mathbf{C}$, where U is an open set in the complex numbers. One method of describing such $f = u + iv$ was that the real and imaginary parts of f had to satisfy the Cauchy-Riemann equations:

$$\frac{\partial u(x,y)}{\partial x} = \frac{\partial v(x,y)}{\partial y}$$

and

$$\frac{\partial u(x,y)}{\partial y} = -\frac{\partial v(x,y)}{\partial x}.$$

Both real-valued functions u and v are harmonic. The harmonicity of u (and in a similar fashion that of v) can be seen, using the Cauchy-Riemann equations, via:

$$\begin{aligned}\triangle u &= \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \\ &= \frac{\partial}{\partial x}\frac{\partial v}{\partial y} + \frac{\partial}{\partial y}\frac{-\partial v}{\partial x} \\ &= 0.\end{aligned}$$

One approach to complex analysis is to push hard on the harmonicity of the real-valued functions u and v.

14.4 The Heat Equation

We will first describe the partial differential equation that is called the Heat Equation and then give a physics-type heuristic argument as to why this particular PDE should model heat flow. In a region in $\mathbf{R}^3$ with the usual coordinates x, y, z, let

$$u(x,y,z,t) = \text{temperature at time } t \text{ at } (x,y,z).$$

Definition 14.4.1 *The* heat equation *is:*

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = c\frac{\partial u}{\partial t},$$

where c is a constant.

Frequently one starts with an initial specified temperature distribution, such as

$$u(x, y, z, 0) = f(x, y, z),$$

with $f(x, y, z)$ some known, given function.

Surprisingly, the heat equation shows up throughout mathematics and the sciences, in many contexts for which no notion of heat or temperature is apparent. The common theme is that heat is a type of diffusion process and that the heat equation is the PDE that will capture any diffusion process. Also, there are a number of techniques for solving the heat equation. In fact, using Fourier Analysis, we solved it in the one-dimensional case in Chapter Thirteen. The method of separation of variables, used in last section to solve the Laplacian, can also be used.

Now to see why the above PDE deserves the name 'heat equation'. As seen in the last section,

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}$$

is the Laplacian. In non-rectilinear coordinates, the Laplacian will have different looking forms, but the heat equation will always be:

$$\Delta u = c\frac{\partial u}{\partial t}.$$

For simplicity, we restrict ourselves to the one-dimensional case. Consider an infinitely long rod, which we denote by the x-axis.

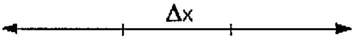

x-axis

Though the basic definitions of heat and temperature are and were fraught with difficulties, we will assume that there is a notion of temperature and that heat is measured via the change in temperature. Let $u(x, t)$ denote the temperature at position x at time t. We now denote the change in a variable by Δu, Δx, Δt, etc. Note that here Δ is not denoting the Laplacian of these variables.

There are three important constants associated to our rod, all coming from the real world: the density ρ, the thermal conductivity k and the specific heat σ. The density arises in that the mass m of the rod over a distance Δx will be the product $\rho \cdot \Delta x$. The specific heat is the number σ that, if a length Δx of the rod has its temperature u raised to $u + \Delta u$, then its heat will change by $\sigma \cdot (mass) \cdot \Delta u$. Note that this last number is the same as $\sigma \cdot \rho \cdot \Delta x \cdot \Delta u$. Here we are using the notion that heat is a measure of the change in temperature. Finally, the thermal conductivity k is the constant that yields

$$k \cdot \frac{\Delta u}{\Delta x}|_x$$

as the amount of heat that can flow through the rod at a fixed point x. Via physical experiments, these constants can be shown to exist.

We want to see how much heat flows in and out of the interval $[x, x+\Delta x]$. By calculating this heat flow by two different methods, and then letting $\Delta x \to 0$, the heat equation will appear. First, if the temperature changes by Δu, the heat will change by

$$\sigma \cdot \rho \cdot \Delta x \cdot \Delta u.$$

Second, at the point $x + \Delta x$, the amount of heat flowing out will be, over time Δt,

$$k \cdot \frac{\Delta u}{\Delta x}|_{x+\Delta x}\Delta t.$$

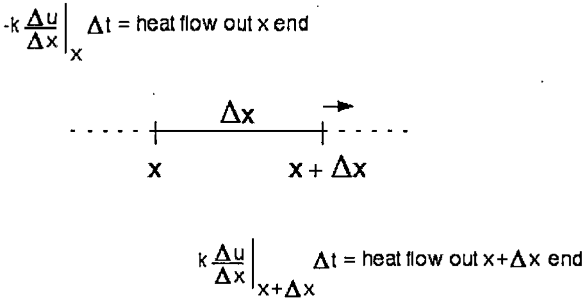
-k Δu/Δx |x Δt = heat flow out x end
Δx
x
x + Δx
k Δu/Δx |x+Δx Δt = heat flow out x+Δx end

At the point x, the amount of heat flowing out will be, over time Δt,

$$-k \cdot \frac{\Delta u}{\Delta x}|_x\Delta t.$$

Then the heat change over the interval Δx will also be

$$(k\frac{\Delta u}{\Delta x}|_{x+\Delta x} - k\frac{\Delta u}{\Delta x}|_x)\Delta t.$$

Thus

$$k \cdot (\frac{\Delta u}{\Delta x}|_{x+\Delta x} - \frac{\Delta u}{\Delta x}|_x)\Delta t = \sigma\rho\Delta x\Delta u.$$

Then

$$\frac{\left(\frac{\triangle u}{\triangle x}|_{x+\triangle x} - \frac{\triangle u}{\triangle x}|_{x}\right)}{\triangle x} = \frac{\sigma\rho}{k}\frac{\triangle u}{\triangle t}.$$

Letting $\triangle x$ and $\triangle t$ approach 0, we get by the definition of partial differentiation the heat equation

$$\frac{\partial^2 u}{\partial x^2} = \frac{\sigma\rho}{k} \cdot \frac{\partial u}{\partial t}.$$

In fact, we see that the constant c is

$$c = \frac{\sigma\rho}{k}.$$

Again, there are at least two other methods for solving the heat equation. We can, for example, use Fourier transforms, which is what we used to solve it in Chapter Thirteen. We can also use the method of separation of variables, discussed in the previous section.

14.5 The Wave Equation

14.5.1 Derivation

As its name suggests, this partial differential equation was originally derived to describe the motion of waves. As with the heat equation, its basic form appears in many apparently non-wave-like areas. We will state the wave equation and then give a quick heuristic description of why the wave equation should describe waves.

A transverse wave in the $x-y$ plane travelling in the x-direction should look like:

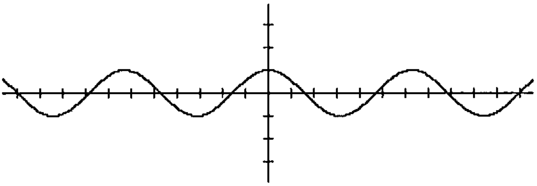

The solution function is denoted by $y(x, t)$, which is just the y coordinate of the wave at place x at time t. The wave equation in two independent variables is

$$\frac{\partial^2 y}{\partial x^2} - c\frac{\partial^2 y}{\partial t^2} = 0,$$

where c is a positive number. Usually we start with some type of knowledge of the initial position of the wave. This will of course mean that we are given an initial function $f(x)$ such that

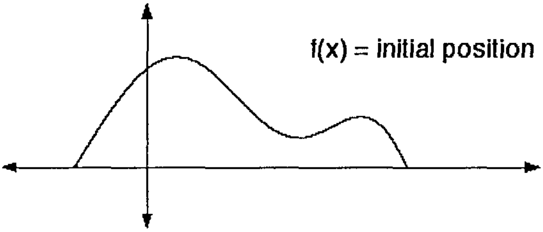

$$y(x,0) = f(x).$$

In general, the wave equation in n variables $x_1, \ldots, x_n$ with initial condition $f(x_1, \ldots, x_n)$ is

$$\frac{\partial^2 y}{\partial {x_1}^2} + \ldots + \frac{\partial^2 y}{\partial {x_n}^2} - c \cdot \frac{\partial^2 y}{\partial t^2} = 0$$

with initial condition

$$y(x_1, \ldots, x_n, 0) = f(x_1, \ldots, x_n).$$

In nonrectilinear coordinates, the wave equation will be:

$$\Delta y(x_1, \ldots, x_n, t) - c \cdot \frac{\partial^2 y}{\partial t^2} = 0.$$

Now to see the heuristics behind why this partial differential equation is even called the wave equation. Of course we need to make some physical assumptions. Assume that the wave is a string moving in an 'elastic' medium, meaning that subject to any displacement, there is a restoring force, something trying to move the string back to where it was. We further assume that the initial disturbance is small. We will use that

$$\text{Force} = (\text{mass}) \cdot (\text{acceleration}).$$

We let our string have density ρ and assume that there is a tension T in the string (this tension will be what we call the restoring force) which will act tangentially on the string. Finally, we assume that the string can only move vertically.

Consider the wave

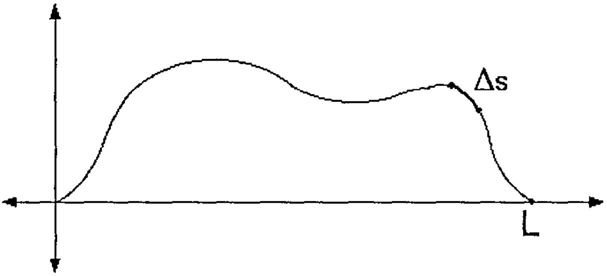

Let s denote the arc length of the curve. We want to calculate the restoring force acting on the segment Δs of the curve in two different ways and then let $\Delta s \to 0$. Since the density is ρ, the mass of the segment Δs will be the product $(\rho \cdot \Delta s)$. The acceleration is the second derivative. Since we are assuming that the curve can only move vertically (in the y-direction), the acceleration will be $\frac{\partial^2 y}{\partial t^2}$. Thus the force will be

$$(\rho \cdot \Delta s) \cdot \frac{\partial^2 y}{\partial t^2}.$$

By the assumption that the displacement is small, we can approximate the arc length Δs by the change in the x-direction alone.

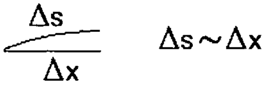

Hence we assume that the restoring force is

$$(\rho \Delta x) \cdot \frac{\partial^2 y}{\partial t^2}.$$

Now to calculate the restoring force in a completely different way. At each point in the picture

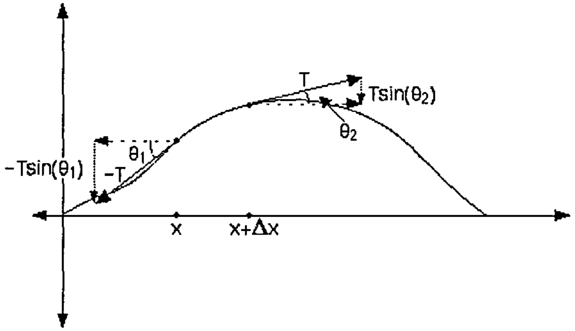

the tension T gives rise to an acceleration tangent to the curve. We want the y component. At the point $x + \Delta x$, the restoring force will be

$$T sin\theta_2.$$

At the point x, the restoring force will be

$$-T sin\theta_1.$$

Since both angles θ_1 and θ_2 are small, we can use the following approximation

$$\begin{aligned} sin\theta_1 &\sim tan\theta_1 = \frac{\partial y}{\partial x}|_x. \\ sin\theta_2 &\sim tan\theta_2 = \frac{\partial y}{\partial x}|_{x+\Delta x}. \end{aligned}$$

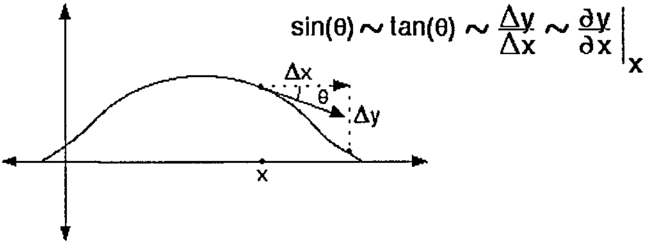

Then we can set the restoring force to be

$$T(\frac{\partial y}{\partial x}|_{x+\Delta x} - \frac{\partial y}{\partial x}|_x).$$

As we have now calculated the restoring force in two different ways, we can set the two formulas equal:

$$T(\frac{\partial y}{\partial x}|_{x+\Delta x} - \frac{\partial y}{\partial x}|_x) = \rho \Delta x \cdot \frac{\partial^2 y}{\partial t^2}.$$

or

$$\frac{\frac{\partial y}{\partial x}|_{x+\Delta x} - \frac{\partial y}{\partial x}|_x}{\Delta x} = \frac{\rho}{T}\frac{\partial^2 y}{\partial t^2}.$$

Letting $\Delta x \to 0$, we get

$$\frac{\partial^2 y}{\partial x^2} = \frac{\rho}{T}\frac{\partial^2 y}{\partial t^2},$$

the wave equation.

Now to see what solutions look like. We assume that $y(0) = 0$ and $y(L) = 0$, for some constant L. Thus we restrict our attention to waves which have fixed endpoints.

An exercise at the end of the chapter will ask you to solve the wave equation using the method of separation of variables and via Fourier transforms. Your answer will in fact be:

$$y(x,t) = \sum_{n=1}^{\infty} k_n sin\left(\frac{n\pi x}{L}\right) cos\left(\frac{n\pi t}{L}\right)$$

where

$$k_n = \frac{2}{L}\int_0^L f(x) sin\left(\frac{n\pi x}{L}\right) \mathrm{d}x.$$

14.5.2 Change of Variables

Sometimes a clever change of variables will reduce the original PDE to a more manageable one. We will see this in the following solution of the wave equation. Take an infinitely long piece of string. Suppose we pluck the string in the middle and then let go.

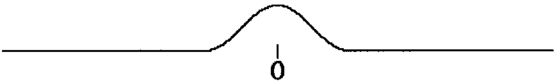

After a short time, we should get:

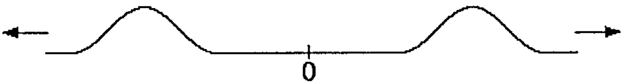

with seemingly two waves moving in opposite directions but at the same speed. With much thought and cleverness, one might eventually try to change coordinate systems in an attempt to capture these two waves.

Thus suppose we want to solve

$$\frac{\partial^2 y}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 y}{\partial t^2} = 0,$$

subject to the initial conditions

$$y(x,0) = g(x) \text{ and } \frac{\partial y}{\partial t}(x,0) = h(x)$$

for given functions $g(x)$ and $h(x)$. Note that we have relabelled the constant in the wave equation to be $\frac{1}{c^2}$. This is done solely for notational convenience, as we will in a moment.

Now to make the change of variables. Set

$$u = x + ct \text{ and } v = x - ct.$$

Using the chain rule, this coordinate change transforms the original wave equation into:

$$\frac{\partial^2 y}{\partial u \partial v} = 0.$$

We can solve this PDE by two straightforward integrations. First integrate with respect to the variable u to get

$$\frac{\partial y}{\partial v} = a(v),$$

where $a(v)$ is an unknown function of the variable v alone. This new function $a(v)$ is the 'constant of integration', constant with respect to the u variable. Now integrate this with respect to v to get

$$y(u,v) = A(v) + B(u),$$

where $A(v)$ is the integral of $a(v)$ and $B(u)$ is the term representing the 'constant of integration' with respect to v. Thus the solution $y(u,v)$ is the sum of two, for now unknown, functions, each a function of one variable alone. Plugging back into our original coordinates means that the solution will have the form:

$$y(u,v) = A(x - ct) + B(x + ct).$$

We use our initial conditions to determine the functions $A(x - ct)$ and $B(x + ct)$. We have

$$g(x) = y(x,0) = A(x) + B(x)$$

and

$$h(x) = \frac{\partial y}{\partial t}(x,0) = -cA'(x) + cB'(x).$$

For this last equation, integrate with respect to the one variable x, to get that

$$\int_0^x h(s)\mathrm{d}s + C = -cA(x) + cB(x).$$

Since we are assuming that the functions $g(x)$ and $h(x)$ are known, we can now solve for $A(x)$ and $B(x)$, to get:

$$A(x) = \frac{1}{2}g(x) - \frac{1}{2c}\int_0^x h(s)\mathrm{d}s - \frac{C}{2c}$$

and

$$B(x) = \frac{1}{2}g(x) + \frac{1}{2c}\int_0^x h(s)\mathrm{d}s + \frac{C}{2c}.$$

Then the solution is:

$$\begin{aligned} y(x,t) &= A(x-ct) + B(x+ct) \\ &= \frac{g(x-ct)+g(x+ct)}{2} + \frac{1}{2c}\int_{x-ct}^{x+ct} h(s)ds, \end{aligned}$$

This is called the *d'Alembert formula.* Note that if the initial velocity $h(x) = 0$, then the solution is simply

$$y(x,t) = \frac{g(x-ct)+g(x+ct)}{2},$$

which is two waves travelling in opposite directions, each looking like the initial position. (Though this is a standard way to solve the wave equation, I took the basic approach from Davis' *Fourier Series and Orthogonal Functions* [24].)

This method leaves the question of how to find a good change of coordinates unanswered. This is an art, not a science.

14.6 The Failure of Solutions: Integrability Conditions

There are no known general methods for determining when a system of partial differential equations has a solution. Frequently, though, there are necessary conditions (usually called 'integrability conditions') for there to be a solution.

We will look at the easiest case. When will there be a two-variable function $f(x,y)$, defined on the plane $\mathbf{R}^2$, satisfying:

$$\frac{\partial f}{\partial x} = g_1(x,y)$$

and

$$\frac{\partial f}{\partial y} = g_2(x,y),$$

where both g_1 and g_2 are differentiable functions? In this standard result from multivariable calculus, there are clean necessary and sufficient conditions for the solution function f to exist:

Theorem 14.6.1 *There is a solution f to the above system of partial differential equations if and only if*

$$\frac{\partial g_1}{\partial y} = \frac{\partial g_2}{\partial x}.$$

In this case, the *integrability* condition is $\frac{\partial g_1}{\partial y} = \frac{\partial g_2}{\partial x}$. As we will see, this is the easy part of the theorem; it is also the model for integrability conditions in general.

Proof: First assume that we have our solution f satisfying $\frac{\partial f}{\partial x} = g_1(x,y)$ and $\frac{\partial f}{\partial y} = g_2(x,y)$. Then

$$\frac{\partial g_1}{\partial y} = \frac{\partial}{\partial y}\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}\frac{\partial f}{\partial y} = \frac{\partial g_2}{\partial x}.$$

Thus the integrability condition is just a consequence that the order for taking partial derivatives does not matter.

The other direction takes more work. As a word of warning, Green's Theorem will be critical. We must find a function $f(x,y)$ satisfying the given system of PDEs. Given any point (x,y) in the plane, let γ be any smooth path from the origin $(0,0)$ to (x,y). Define

$$f(x,y) = \int_\gamma g_1(x,y)\mathrm{d}x + g_2(x,y)\mathrm{d}y.$$

We first show that the function $f(x,y)$ is well-defined, meaning that its value is independent of which path γ is chosen. This will then allow us to show that $\frac{\partial f}{\partial x} = g_1(x,y)$ and $\frac{\partial f}{\partial y} = g_2(x,y)$. Let τ be another smooth path from $(0,0)$ to (x,y).

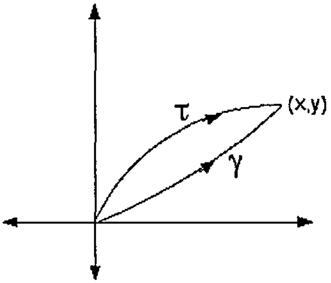

We want to show that

$$\int_\gamma g_1(x,y)\mathrm{d}x + g_2(x,y)\mathrm{d}y = \int_\tau g_1(x,y)\mathrm{d}x + g_2(x,y)\mathrm{d}y.$$

We can consider $\gamma - \tau$ as a closed loop at the origin, enclosing a region R. (Note: it might be the case that $\gamma - \tau$ encloses several regions, but then just apply the following to each of these regions.) By Green's Theorem we have

$$\begin{aligned}\int_\gamma g_1\mathrm{d}x + g_2\mathrm{d}y - \int_\tau g_1\mathrm{d}x + g_2\mathrm{d}y &= \int_{\gamma-\tau} g_1\mathrm{d}x + g_2\mathrm{d}y \\ &= \int_R (\frac{\partial g_2}{\partial x} - \frac{\partial g_1}{\partial y})\mathrm{d}x\mathrm{d}y \\ &= 0\end{aligned}$$

by the assumption that $\frac{\partial g_1}{\partial y} = \frac{\partial g_2}{\partial x}$. Thus the function $f(x,y)$ is well-defined.

Now to show that this function f satisfies $\frac{\partial f}{\partial x} = g_1(x,y)$ and $\frac{\partial f}{\partial y} = g_2(x,y)$. We will just show the first, as the second is similar. The key is that we will reduce the problem to the Fundamental Theorem of Calculus. Fix a point (x_0, y_0). Consider any path γ from $(0,0)$ to (x_0, y_0) and the extension $\gamma' = \gamma + \tau$, where τ is the horizontal line from (x_0, y_0) to (x, y_0).

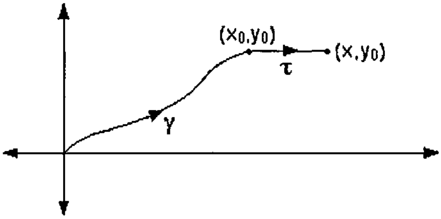

Then

$$\begin{aligned}\frac{\partial f}{\partial x} &= \lim_{x\to x_0} \frac{f(x,y_0) - f(x_0,y_0)}{x - x_0} \\ &= \lim_{x\to x_0} \frac{\int_{x_0}^{x} g_1(t,y_0)\mathrm{d}t}{x - x_0},\end{aligned}$$

since there is no variation in the y-direction, forcing the g_2 part of the path integral to drop out. This last limit, by the Fundamental Theorem of Calculus, is equal to g_1, as desired. □

14.7 Lewy's Example

Once you place any natural integrability conditions on a system of partial differential equations, you can then ask if there will always be a solution.

In practice, often such general statements about the existence of solutions can be made. For example, in the middle of the twentieth century it was shown that given any complex numbers $a_1, \ldots, a_n$ and any smooth function $g(x_1, \ldots, x_n)$, there always exists a smooth solution $f(x_1, \ldots, x_n)$ satisfying

$$a_1 \frac{\partial f}{\partial x_1} + \ldots + a_n \frac{\partial f}{\partial x_n} = g.$$

Based in part on these types of results, it was the belief that all reasonable PDEs would have solutions. Then, in 1957, Hans Lewy showed the amazing result that the linear PDE

$$\frac{\partial f}{\partial x} + i \frac{\partial f}{\partial y} - (x + iy) \frac{\partial f}{\partial z} = g(x, y, z)$$

will have a solution f only if g is real-analytic. Note that while this PDE does not have constants as coefficients, the coefficients are about as reasonable as you could want. Lewy's proof, while not hard (see Folland's book on PDEs [39]), did not give any real indication as to why there is no solution. In the early 1970s, Nirenberg showed that the Lewy PDE did not have a solution due to that there existed a three-dimensional CR structure (a certain type of manifold) that could not be embedded into a complex space, thus linking a geometric condition to the question of existence of this PDE. This is a common tack, namely to concentrate on PDEs whose solutions have some type of geometric meaning. Then, in trying to find the solution, use the geometry as a guide.

14.8 Books

Since beginning differential equations is a standard sophomore level course, there are many beginning text books. Boyce and Diprima's book [12] has long been a standard. Simmon's book [99] is also good. Another approach to learning basic ODEs is to volunteer to TA or teach such a class (though I would recommend that you teach linear algebra and vector calculus first). Moving into the realm of PDEs the level of text becomes much harder and more abstract. I have learned a lot from Folland's book [39]. Fritz John's book [69] has long been a standard. I have heard that Evans' recent book [33] is also excellent.

14.9 Exercises

1. The most basic differential equation is probably

$$\frac{dy}{dx} = y,$$

subject to the boundary condition $y(0) = 1$. The solution is of course the exponential function $y(x) = e^x$. Use Picard iteration to show that this is indeed the solution to $\frac{dy}{dx} = y$. (Of course you get an answer as a power series and then need to recognize that the power series is e^x. The author realizes that if you know the power series for the exponential function you also know that it is its own derivative. The goal of this problem is see explicitly how Picard iteration works on the simplest possible differential equation.)

2. Let $f(x)$ be a one variable function, with domain the interval $[0, 1]$, whose first derivative is continuous. Show that f is Lipschitz.
3. Show that $f(x) = e^x$ is not Lipschitz on the real numbers.
4. Solve the wave equation

$$\frac{\partial^2 y}{\partial x^2} - c\frac{\partial^2 y}{\partial t^2} = 0$$

subject to the boundary conditions $y(0, t) = 0$ and $y(L, t) = 0$ and the initial condition $y(x, 0) = f(x)$ for some function $f(x)$.

a. Use the method of separation of variables as described in the section on the Laplacian.

b. Now find the solutions using Fourier transforms.

Chapter 15

Combinatorics and Probability Theory

Basic Goals:	Cleverly Counting Large Finite Sets Central Limit Theorem

Beginning probability theory is basically the study of how to count large finite sets, or in other words, an application of combinatorics. Thus the first section of this chapter deals with basic combinatorics. The next three sections deal with the basics of probability theory. Unfortunately, counting will only take us so far in probability. If we want to see what happens as we, for example, play a game over and over again, methods of calculus become important. We concentrate on the Central Limit Theorem, which is where the famed Gauss-Bell curve appears. The proof of the Central Limit Theorem is full of clever estimates and algebraic tricks. We include this proof not only due to the importance of the Central Limit Theorem but also to show people that these types of estimates and tricks are sometimes needed in mathematics.

15.1 Counting

There are many ways to count. The most naive method, the one we learn as children, is simply to explicitly count the elements in a set, and this method is indeed the best one for small sets. Unfortunately, many sets are just too large for anyone to merely count the elements. Certainly in large part the fascination in card games such as poker and bridge is that while there are only a finite number of possible hands, the actual number is far too large for anyone to deal with directly, forcing the players to develop strategies and various heuristical devices. Combinatorics is the study of how to cleverly count. Be warned that the subject can quickly get quite difficult and is becoming increasingly important in mathematics.

We will look at the simplest of combinatorial formulas, ones that have been known for centuries. Start with n balls. Label each ball with a number $1, 2, \ldots, n$ and then put the balls into an urn. Pull one out, record its number and then put the ball back in. Again, pull out a ball and record its number and put it back into the urn. Keep this up until k balls have been pulled out and put back into the urn. We want to know how many different k-tuples of numbers are possible.

To pull out two balls from a three-ball urn (here $n = 3$ and $k = 2$), we can just list the possibilities:

$$(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3).$$

But if we pull out seventy-six balls from a ninety-nine ball urn (here $n = 99$ and $k = 76$), it would be ridiculous to make this list.

Nevertheless, we can find the correct number. There are n possibilities for the first number, n possibilities for the second, n for the third, etc. Thus all told there must be n^k possible ways to choose k-tuples of n numbers. This is a formula that works no matter how many balls we have or how many times we choose a ball.

For the next counting problem, return to the urn. Pull out a ball, record its number and keep it out. Now pull out another ball, record its number and keep it out. Continue pulling out balls and not replacing them. Now we want to find out how many k-tuples of n numbers there are without replacement. There are n possibilities for the first number, only $(n-1)$ possibilities for the second, $(n-2)$ for the third, etc. Thus the number of ways of choosing from n balls k times without replacement is:

$$n(n-1)(n-2)\cdots(n-k+1).$$

For our next counting problem, we want to find out how many ways there are for pulling out k balls from an urn with n balls, but now not only not replacing the balls but also not caring about the order of the balls. Thus pulling out the balls $(1,2,3)$ will be viewed as equivalent to pulling out the balls $(2,1,3)$. Suppose we have already pulled out k of the balls. We want to see how many ways there are of mixing up these k balls. But this should be the same as how many ways are there of choosing from k balls k times, which is

$$k(k-1)(k-2)\cdots 2\cdot 1 = k!.$$

Since $n(n-1)(n-2)\cdots(n-k+1)$ is the number of ways of choosing from n balls k times with order mattering and with each ordering capable of being mixed up $k!$ ways, we have

$$\frac{n(n-1)\dots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!},$$

which is the number of ways of choosing k balls from n balls without replacement and with order not mattering. This number comes up so often it has its own symbol

$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

pronounced 'n choose k'. It is frequently called the *binomial coefficient*, due to its appearance in the Binomial Theorem:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

The idea is that $(a+b)^n = (a+b)(a+b)\dots(a+b)$. To calculate how many different terms of the form $a^k b^{n-k}$ we can get, we note that this is the same as counting how many ways we can choose k things from n things without replacement and with ordering not mattering.

15.2 Basic Probability Theory

We want to set up the basic definitions of elementary probability theory. These definitions are required to yield the results we all know, such as that there is a fifty-fifty chance of flipping a coin and getting heads, or that there is a one in four chance of drawing a heart from a standard deck of 52 cards. Of course, as always, the reason for worrying about the basic definitions is not just to understand the obvious odds of getting heads but that the correct basic definition will allow us to compute the probabilities of events that are quite complicated.

We start with the notion of a sample space ω, which technically is just another name for a set. Intuitively, a sample space ω is the set whose elements are what can happen, or more precisely, the possible outcomes of an event. For example, if we flip a coin twice, ω will be a set with the four elements

$$\{(heads, heads), (heads, tails), (tails, heads), (tails, tails)\}.$$

Definition 15.2.1 *Let ω be a sample space and A a subset of ω. Then the* probability *of A, denoted by $P(A)$, is the number of elements in A divided by the number of elements in the sample space ω. Thus*

$$P(A) = \frac{|A|}{|\omega|},$$

where $|A|$ denotes the number of elements in the set A.

For example, if

$$\omega = \{(heads, heads), (heads, tails), (tails, heads), (tails, tails)\},$$

and if $A = \{(heads, heads)\}$, then the probability of flipping a coin twice and getting two heads will be

$$P(A) = \frac{|A|}{|\omega|} = \frac{1}{4},$$

which agrees with common sense.

In this framework, many of the basic rules of probability reduce to rules of set theory. For example, via sets, we see that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Frequently, a subset A of a sample space ω is called an *event*.

There are times when it is too much trouble to actually translate a real-world probability problem into a question of size of sets. For example, suppose we are flipping an unfair coin, where there is a 3/4 chance of getting a head and a 1/4 chance of getting tails. We could model this by taking our sample set to be

$$\omega = \{heads_1, heads_2, heads_3, tails\},$$

where we are using subscripts to keep track of the different ways of getting heads, but this feels unnatural. A more natural sample space would be

$$\omega = \{heads, tails\},$$

and to somehow account for the fact that it is far more likely to get heads than tails. This leads to another definition of a probability space:

Definition 15.2.2 *A* probability space *is a set ω, called the* sample space, *and a function*

$$P : \omega \to [0, 1]$$

such that

$$\sum_{a \in \omega} P(a) = 1.$$

We say that the probability of getting an 'a' is the value of $P(a)$.

If on a sample space ω it is equally likely to get any single element of ω, i.e., for all $a \in \omega$ we have

$$P(a) = \frac{1}{|\omega|},$$

then our 'size of set' definition for probability will agree with this second definition. For the model of flipping an unfair coin, this definition will give us that the sample set is:

$$\omega = \{heads, tails\},$$

but that $P(heads) = 3/4$ and $P(tails) = 1/4$.

We now turn to the notion of a random variable.

Definition 15.2.3 *A* random variable $\mathbf{X}$ *on a sample space* ω *is a real-valued function on* ω*:*

$$\mathbf{X} : \omega \to \mathbf{R}.$$

For example, we now create a simplistic gambling game which requires two flips of a coin. Once again let the sample space be

$$\omega = \{(heads, heads), (heads, tails), (tails, heads), (tails, tails)\}.$$

Suppose that, if the first toss of a coin is heads, you win ten dollars. If it is tails, you lose five dollars. On the second toss, heads will pay fifteen dollars and tails will cost you twelve dollars. To capture these stakes (for an admittedly boring game), we define the random variable

$$\mathbf{X} : \omega \to \mathbf{R}$$

by

$$\mathbf{X}(heads, heads) = 10 + 15 = 25$$

$$\mathbf{X}(heads, tails) = 10 - 12 = -2$$

$$\mathbf{X}(tails, heads) = -5 + 15 = 10$$

$$\mathbf{X}(tails, tails) = -5 - 12 = -17.$$

15.3 Independence

Toss a pair of dice, one blue and one red. The number on the blue die should have nothing to do with the the number on the red die. The events are in some sense independent, or disjoint. We want to take this intuition of independence and give it a sharp definition.

Before giving a definition for independence, we need to talk about conditional probability. Start with a sample space ω. We want to understand the probability for an event A to occur, given that we already know some other event B has occurred. For example, roll a single die. Let ω be the six possible outcomes on this die. Let A be the event that a 4 shows up. Certainly we have

$$P(A) = \frac{|A|}{|\omega|} = \frac{1}{6}.$$

But suppose someone tells us, before we look at the rolled die, that they know for sure that on the die there is an even number. Then the probability that a 4 will occur should be quite different. The set $B = \{2,4,6\}$ is the event that an even number occurs. Then the probability that a 4 shows up should now be $1/3$, as there are only three elements in B. Note that

$$\frac{1}{3} = \frac{|A \cap B|}{|B|} = \frac{\frac{|A \cap B|}{|\omega|}}{\frac{|B|}{|\omega|}} = \frac{P(A \cap B)}{P(B)}.$$

This motivates the definition:

Definition 15.3.1 *The* conditional probability *that A occurs given that B has occurred is:*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

What should it mean for an event A to be independent from an event B? At the least, it should mean that knowing about the likelihood of event B occurring should have no bearing on the likelihood that A occurs, i.e., knowing about B should not effect A. Thus if A and B are independent, we should have

$$P(A|B) = P(A).$$

Using that $P(A|B) = \frac{P(A \cap B)}{P(B)}$, this means that a reasonable definition for independence is:

Definition 15.3.2 *Two events A and B are* independent *if*

$$P(A \cap B) = P(A) \cdot P(B).$$

15.4 Expected Values and Variance

In a game, how much should you be expected to win in the long run? This quantity is the expected value. Further, how likely is it that you might lose big time, even if the expected value tells you that you will usually come out ahead? This type of information is contained in the variance and in its square root, the standard deviation. We start with some definitions.

Definition 15.4.1 *The* expected value *of a random variable* $\mathbf{X}$ *on a sample space* ω *is:*

$$E(\mathbf{X}) = \sum_{a\in\omega} \mathbf{X}(a)\cdot P(a).$$

For example, recall the simplistic game defined at the end of section two, where we flip a coin twice and our random variable represents our winnings: $\mathbf{X}(heads, heads) = 10+15 = 25, \mathbf{X}(heads, tails) = 10-12 = -2, \mathbf{X}(tails, heads) = -5+15 = 10$, and $\mathbf{X}(tails, tails) = -5-12 = -17$. The expected value is simply:

$$\begin{aligned} E(\mathbf{X}) &= 25\left(\frac{1}{4}\right) + (-2)\left(\frac{1}{4}\right) + 10\left(\frac{1}{4}\right) + (-17)\left(\frac{1}{4}\right) \\ &= 4. \end{aligned}$$

Intuitively, this means that on average you will win four dollars each time you play the game. Of course, luck might be against you and you could lose quite a bit.

The expected value can be viewed as a function from the set of all random variables to the real numbers. As a function, the expected value is linear.

Theorem 15.4.1 *On a probability space, the expected value is linear, meaning that for all random variables* $\mathbf{X}$ *and* $\mathbf{Y}$ *and all real numbers* λ *and* μ, *we have*

$$E(\lambda\mathbf{X}+\mu\mathbf{Y}) = \lambda E(\mathbf{X}) + \mu E(\mathbf{Y}).$$

Proof: This is a straightforward calculation from the definition of expected value. We have

$$\begin{aligned} E(\lambda\mathbf{X}+\mu\mathbf{Y}) &= \sum_{a\in\omega}(\lambda\mathbf{X}+\mu\mathbf{Y})(a)\cdot P(a) \\ &= \sum_{a\in\omega}(\lambda\mathbf{X}(a)+\mu\mathbf{Y}(a))\cdot P(a) \\ &= \sum_{a\in\omega}\lambda\mathbf{X}(a)\cdot P(a) + \sum_{a\in\omega}\mu\mathbf{Y}(a)\cdot P(a) \end{aligned}$$

$$\begin{aligned} &= \lambda \sum_{a\in\omega} \mathbf{X}(a)\cdot P(a) + \mu \sum_{a\in\omega} \mathbf{Y}(a)\cdot P(a) \\ &= \lambda E(\mathbf{X}) + \mu E(\mathbf{Y}). \ \square \end{aligned}$$

The expected value will only tell a part of the story, though. Consider two classes, each with ten students. On a test, in one of the classes five people got 100s and five got 50s, while in the other everyone got a 75. In both classes the average was a 75 but the performances were quite different. Expected value is like the average, but it does not tell us how far from the average you are likely to be. For example, in the first class you are guaranteed to be 25 points from the average while in the second class you are guaranteed to be exactly at the average. There is a measure of how likely it is that you are far from the expected value:

Definition 15.4.2 *The* variance *of a random variable* **X** *on a sample space* ω *is*

$$V(\mathbf{X}) = E[\mathbf{X} - E(\mathbf{X})]^2.$$

The idea is we set up a new random variable,

$$[\mathbf{X} - E(\mathbf{X})]^2.$$

Note that the expected value $E(\mathbf{X})$ is just a number. The farther **X** is from its expected value $E(\mathbf{X})$, the larger is $[\mathbf{X} - E(\mathbf{X})]^2$. Thus it is a measure of how far we can be expected to be from the average. We square $\mathbf{X} - E(\mathbf{X})$ in order to make everything non-negative.

We can think of the variance V as a map from random variables to the real numbers. While not quite linear, it is close, as we will now see. First, though, we want to show that the formula for variance can be rewritten.

Lemma 15.4.1 *For a random variable* **X** *on a probability space, we have*

$$V(\mathbf{X}) = E(\mathbf{X}^2) - [E(\mathbf{X})]^2$$

Proof: This is a direct calculation. We are interested in the new random variable

$$[\mathbf{X} - E(\mathbf{X})]^2.$$

Now

$$[\mathbf{X} - E(\mathbf{X})]^2 = \mathbf{X}^2 - 2\mathbf{X}E(\mathbf{X}) + [E(\mathbf{X})]^2.$$

Since $E(\mathbf{X})$ is just a number and since the expected value, as a map from random variables to the reals, is linear, we have

$$\begin{aligned} V(\mathbf{X}) &= E[\mathbf{X} - E(\mathbf{X})]^2 \\ &= E[\mathbf{X}^2 - 2\mathbf{X}E(\mathbf{X}) + [E(\mathbf{X})]^2] \\ &= E(\mathbf{X}^2) - 2E(\mathbf{X})E(\mathbf{X}) + [E(\mathbf{X})]^2 \\ &= E(\mathbf{X}^2) - [E(\mathbf{X})]^2, \end{aligned}$$

as desired. □

This will allow us to show that the variance is almost linear.

Theorem 15.4.2 *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be any two random variables that are independent on a probability space and let* λ *be any real number. Then*

$$V(\lambda \mathbf{X}) = \lambda^2 V(\mathbf{X})$$

and

$$V(\mathbf{X}+\mathbf{Y}) = V(\mathbf{X}) + V(\mathbf{Y}).$$

It is the λ^2 term that prevents the variance from being linear.

Proof: Since the expected value is a linear function, we know that $E(\lambda\mathbf{X}) = \lambda E(\mathbf{X})$. Then

$$\begin{aligned} V(\lambda \mathbf{X}) &= E[(\lambda\mathbf{X})^2] - [E(\lambda\mathbf{X})]^2 \\ &= \lambda^2 E(\mathbf{X}^2) - [\lambda E(\mathbf{X})]^2 \\ &= \lambda^2 [E(\mathbf{X}^2) - [E(\mathbf{X})]^2] \\ &= \lambda^2 V(\mathbf{X}). \end{aligned}$$

For the second formula, we will need to use that the independence of $\mathbf{X}$ and $\mathbf{Y}$ means that

$$E(\mathbf{X}\mathbf{Y}) = E(\mathbf{X})E(\mathbf{Y}).$$

By the above lemma's description of variance, we have

$$\begin{aligned} V(\mathbf{X}+\mathbf{Y}) &= E[(\mathbf{X}+\mathbf{Y})^2] - [E(\mathbf{X}+\mathbf{Y})]^2 \\ &= E[\mathbf{X}^2 + 2\mathbf{X}\mathbf{Y} + \mathbf{Y}^2] - [E(\mathbf{X}) + E(\mathbf{Y})]^2 \\ &= E[\mathbf{X}^2] + 2E[\mathbf{X}\mathbf{Y}] + E[\mathbf{Y}^2] \\ &\quad -[E(\mathbf{X})]^2 - 2E(\mathbf{X})E(\mathbf{Y}) - [E(\mathbf{Y})]^2 \\ &= (E[\mathbf{X}^2] - [E(\mathbf{X})]^2) + (2E[\mathbf{X}\mathbf{Y}] \\ &\quad -2E(\mathbf{X})E(\mathbf{Y})) + (E[\mathbf{Y}^2] - [E(\mathbf{Y})]^2) \\ &= V(\mathbf{X}) + V(\mathbf{Y}), \end{aligned}$$

as desired. □

A number related to the variance is its square root, the *standard deviation*:

$$\text{standard deviation}(\mathbf{X}) = \sigma(\mathbf{X}) = \sqrt{V(\mathbf{X})}.$$

15.5 Central Limit Theorem

In the last section we defined the basic notions of probability in terms of counting. Unfortunately, combinatorics can only take us so far. Think about flipping a coin. After many flips, we expect that the total number of heads should be quite close to one half of the total number of flips. In trying to capture this notion of flipping a coin over and over again, we need to introduce the following:

Definition 15.5.1 *Repeated independent trials are called* Bernoulli trials *if there are only two possible outcomes for each trial and if their probabilities remain the same throughout the trials.*

Let A be one of the outcomes and suppose the probability of A is $P(A) = p$. Then the probability of A not occurring is $1 - p$, which we will denote by q. Let the sample space be

$$\omega = \{A, \text{not } A\}.$$

We have

$$P(A) = p, P(\text{not } A) = q.$$

We now want to see what happens when we take many repeated trials. The following theorem is key:

Theorem 15.5.1 (Central Limit Theorem) *Consider a sample space $\omega = \{A, not\ A\}$ with $P(A) = p$ and $P(not\ A) = 1 - p = q$. Given n independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$, each taking*

$$\mathbf{X}_i(A) = 1,\ \mathbf{X}_i(not\,A) = 0,$$

set

$$S_n = \sum_{i=1}^{n} \mathbf{X}_i$$

and

$$S_n^* = \frac{S_n - E(S_n)}{\sqrt{V(S_n)}}.$$

Then for any real numbers a and b,

$$\lim_{n\to\infty} P\{a \leq S_n^* \leq b\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{\frac{-x^2}{2}} dx.$$

What this is saying is that if we perform a huge number of repeated Bernoulli trials, then the values of $\mathbf{S}_n$ will be distributed as:

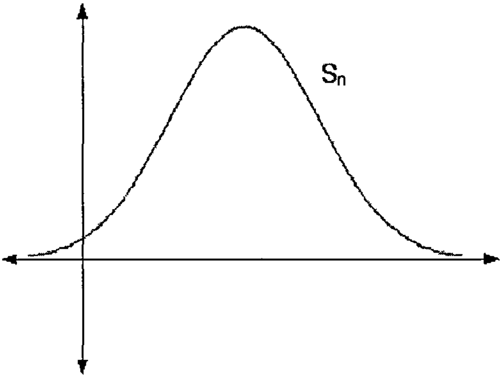

But we have even more. Namely, by normalizing S_n to the new random variable S_n^* (which, as we will see in a moment, has mean zero and variance one), we always get the same distribution, no matter what the real world situation we start with is, just as long as the real world problem can be modelled as a Bernoulli trial. By the way, the distribution for any Bernoulli trial is simply the graph of the function $\lim_{n\to\infty} S_n$. We call S_n^* the *normal distribution.* Its graph is the *Gauss-Bell curve.*

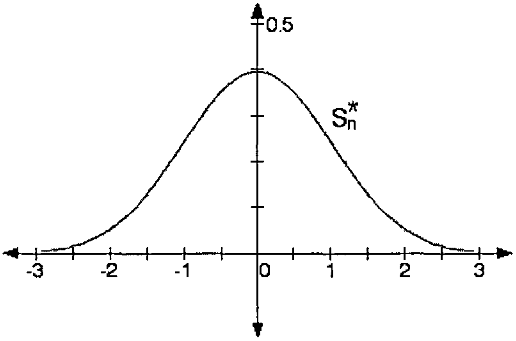

Before sketching a proof of the Central Limit Theorem (whose general outline is from [18]), let us look at the random variables S_n and S_n^*.

Lemma 15.5.1 *The expected value of S_n is np and its variance is npq. The expected value of S_n^* is 0 and its variance is 1.*

Proof of Lemma: We know that for all k,

$$E(\mathbf{X}_k) = \mathbf{X}_k(A)P(A) + \mathbf{X}_k(\text{not } A)P(\text{not } A) = 1 \cdot p + 0 \cdot q = p.$$

Then by the linearity of the expected value function

$$\begin{aligned} E(S_n) &= E(\mathbf{X}_1 + \ldots + \mathbf{X}_n) \\ &= E(\mathbf{X}_1) + \ldots + E(\mathbf{X}_n) \\ &= np. \end{aligned}$$

As for the variance, we know that for any k,

$$\begin{aligned} V(\mathbf{X}_k) &= E(\mathbf{X}_k^2) - [E(\mathbf{X}_k)]^2 \\ &= \mathbf{X}_k^2(A)P(A) + \mathbf{X}_k^2(\text{not } A)P(\text{not } A) - p^2 \\ &= 1^2 \cdot p + 0^2 \cdot q - p^2 \\ &= p - p^2 \\ &= p(1-p) \\ &= pq. \end{aligned}$$

Then we have

$$\begin{aligned} V(S_n) &= V(\mathbf{X}_1 + \ldots + \mathbf{X}_n) \\ &= V(\mathbf{X}_1) + \ldots + V(\mathbf{X}_n) \\ &= npq. \end{aligned}$$

Now

$$\begin{aligned} E(S_n^*) &= E\left(\frac{S_n - E(S_n)}{\sqrt{V(S_n)}}\right) \\ &= \frac{1}{\sqrt{V(S_n)}} E(S_n - E(S_n)) \\ &= \frac{1}{\sqrt{V(S_n)}} (E(S_n) - E(E(S_n))), \end{aligned}$$

which, since $E(S_n)$ is just a number, is zero.

Now for the variance. First, note that for any random variable that happens to be a constant function, the variance must be zero. In particular, since the expected value of a random variable is a number, we must have that the variance of an expected value is zero:

$$V(E(\mathbf{X})) = 0.$$

Using this, we have that

$$V(S_n^*) = V\left(\frac{S_n - E(S_n)}{\sqrt{V(S_n)}}\right)$$

$$\begin{aligned} &= (\frac{1}{\sqrt{V(S_n)}})^2 V(S_n - E(S_n)) \\ &= \frac{1}{V(S_n)}(V(S_n) - V(E(S_n))) \\ &= 1, \end{aligned}$$

as desired. □

Before discussing the proof of the Central Limit Theorem, let us look at the formula

$$\lim_{n\to\infty} P(a \leq S_n^* \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{\frac{-x^2}{2}} \mathrm{d}x.$$

It happens to be the case that for any particular choice of a and b, it is impossible to explicitly calculate the integral $\frac{1}{\sqrt{2\pi}} \int_a^b e^{\frac{-x^2}{2}} dx$; instead people must numerically approximate the answers, which of course can easily be done with standard software packages like Maple or Mathematica. Surprisingly enough, $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} dx$ can be shown to be exactly one. We first show why this must be the case if the Central Limit Theorem is true and then we will explicitly prove that this integral is one.

For any sequence of events and for any n, S_n^* must be some number. Thus for all n,

$$P(-\infty \leq S_n^* \leq \infty) = 1,$$

and thus its limit as n goes to infinity must be one, meaning that our integral is one. Thus if $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} dx$ is not one, the Central Limit Theorem would not be true. Thus we need to prove that this integral is one. In fact, the proof that this integral is one is interesting in its own right.

Theorem 15.5.2

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} dx = 1.$$

Proof: Surprisingly, we look at the square of the integral:

$$(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} \mathrm{d}x)^2 = (\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} \mathrm{d}x)(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} \mathrm{d}x).$$

Since the symbol x just denotes what variable we are integrating over, we can change the x in the second integral to a y without changing the equality:

$$(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} \mathrm{d}x)^2 = (\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} \mathrm{d}x)(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-y^2}{2}} \mathrm{d}y).$$

Since the x and the y have nothing to do with each other, we can combine these two single integrals into one double integral:

$$\begin{aligned}(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\frac{-x^2}{2}}\mathrm{d}x)^2 &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} e^{\frac{-y^2}{2}}\mathrm{d}x\mathrm{d}y \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\frac{-(x^2+y^2)}{2}}\mathrm{d}x\mathrm{d}y,\end{aligned}$$

which is now a double integral over the real plane. The next trick is to switch over to polar coordinates, to reduce our integrals to doable ones. Recall that we have $\mathrm{d}x\mathrm{d}y = r\mathrm{d}r\mathrm{d}\theta$ and $x^2 + y^2 = r^2$

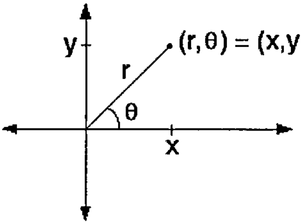

in polar coordinates. Then we have

$$\begin{aligned}(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\frac{-x^2}{2}}\mathrm{d}x)^2 &= \frac{1}{2\pi}\int_{0}^{2\pi}\int_{0}^{\infty} -e^{-\frac{r^2}{2}} r\mathrm{d}r\mathrm{d}\theta \\ &= \frac{1}{2\pi}\int_{0}^{2\pi} e^{-\frac{r^2}{2}} \mid_0^{\infty} \mathrm{d}\theta \\ &= \frac{1}{2\pi}\int_{0}^{2\pi} \mathrm{d}\theta \\ &= 1,\end{aligned}$$

as desired. □

Proof of Central Limit Theorem: (Again, we got this argument from [18].) At a critical stage of this proof, there will be a summation of terms of the form

$$\binom{n}{k} p^k q^{n-k}$$

which we will replace by

$$\frac{1}{\sqrt{2\pi npq}} e^{-\frac{x_k^2}{2}},$$

where the x_k will be defined in a moment. We will see that the justification for this replacement is a corollary of Stirling's formula for $n!$, next section's topic.

We are interested in $P(a \leq S_n^* \leq b)$. But, at least initially, the random variable S_n is a bit easier to work with. We want to link S_n with S_n^*. Suppose that we know that $S_n = k$, which means that after n trials, there have been exactly k occurrences of A (and thus $n-k$ occurrences of not A). Let x_k denote the corresponding value for S_n^*. Then

$$x_k = \frac{k - E(S_n)}{\sqrt{V(S_n)}}.$$

Since $E(S_n) = np$ and $V(S_n) = npq$, we have

$$x_k = \frac{k - np}{\sqrt{npq}},$$

and thus

$$k = np + \sqrt{npq}x_k.$$

Then

$$P(a \leq S_n^* \leq b) = \sum_{\{a \leq x_k \leq b\}} P(S_n = k).$$

First we need to show that

$$P(S_n = k) = \binom{n}{k} p^k q^{n-k}.$$

Now $S_n = k$ means that after n trials there are exactly k A's. Since $P(A) = p$ and $P(\text{not } A) = q$, we have that the probability of any particular pattern of k A's is $p^k q^{n-k}$ (for example, if the first k trials yield A's and the last $n-k$ trials yield not A's). But among n trials, there are $\binom{n}{k}$ different ways for there to be k A's. Thus $P(S_n = k) = \binom{n}{k} p^k q^{n-k}$.

Then we have

$$P(a \leq S_n^* \leq b) = \sum_{\{a \leq x_k \leq b\}} \binom{n}{k} p^k q^{n-k}.$$

We now replace $\binom{n}{k} p^k q^{n-k}$ with $\frac{1}{\sqrt{2\pi npq}} e^{-\frac{x_k^2}{2}}$ (which, again, will be justified in the next section), giving us

$$\begin{aligned} P(a \leq S_n^* \leq b) &= \sum_{\{a \leq x_k \leq b\}} \frac{1}{\sqrt{2\pi npq}} e^{-\frac{x_k^2}{2}} \\ &= \sum_{\{a \leq x_k \leq b\}} \frac{1}{\sqrt{2\pi}\sqrt{npq}} e^{-\frac{x_k^2}{2}}. \end{aligned}$$

Note that

$$x_{k+1} - x_k = \frac{k+1-np}{\sqrt{npq}} - \frac{k-np}{\sqrt{npq}} = \frac{1}{\sqrt{npq}}.$$

Thus

$$P(a \leq S_n^* \leq b) = \sum_{\{a \leq x_k \leq b\}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_k^2}{2}} (x_{k+1} - x_k).$$

As we let n approach infinity, the interval $[a, b]$ is split into a finer and finer partition by the x_k. The above sum is a Riemann sum and can thus be replaced, as n approaches infinity, by our desired integral:

$$\lim_{n\to\infty} P(a \leq S_n^* \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{\frac{-x^2}{2}} dx. \ \square$$

15.6 Stirling's Approximation for $n!$

Stirling's formula tells us that for large n we can replace $n!$ by $\sqrt{2\pi n} n^n e^{-n}$. We need this approximation to complete the proof of the Central Limit Theorem. (We are still following [18].)

First, given two functions $f(n)$ and $g(n)$, we say that

$$f(n) \sim g(n)$$

if there exists a nonzero constant c such that

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = c.$$

Thus the functions $f(n)$ and $g(n)$ grow at the same rate as n goes to infinity. For example

$$n^3 \sim 5n^3 - 2n + 3.$$

Theorem 15.6.1 (Stirling's Formula)

$$n! \sim \sqrt{2\pi n} n^n e^{-n}$$

Proof: This will take some work and some algebraic manipulations.

First note that

$$\sqrt{2\pi n} n^n e^{-n} = \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}.$$

We will show here that

$$\lim_{n\to\infty} \frac{n!}{n^{n+\frac{1}{2}} e^{-n}} = k,$$

for some constant k. To show that $k = \sqrt{2\pi}$, we use the following convoluted argument. Assume that we have already shown that $n! \sim kn^{n+\frac{1}{2}}e^{-n}$. Use this approximation in our replacement of $\binom{n}{k}p^k q^{n-k}$ in the following corollary and, more importantly, in the proof in the last section of the Central Limit Theorem. If we follow the steps in that proof, we will end up with

$$\lim_{n\to\infty} P(a \leq S_n^* \leq b) = \frac{1}{k}\int_a^b e^{\frac{-x^2}{2}}\,dx.$$

Since for each n, we must have S_n^* equal to some number, we know that $P(-\infty \leq S_n^* \leq \infty) = 1$ and thus $\lim_{n\to\infty} P(-\infty \leq S_n^* \leq \infty) = 1$. Then we must have

$$\frac{1}{k}\int_{-\infty}^{\infty} e^{\frac{-x^2}{2}}\,dx = 1.$$

But in the last section we calculated that $\int_{-\infty}^{\infty} e^{\frac{-x^2}{2}}\,dx = \sqrt{2\pi}$. From this calculation, we see that k must be $\sqrt{2\pi}$.

Now for the meat of the argument, showing that such a k exists. This will take some work and involve various computational tricks. Our goal is to show that there is a nonzero constant k such that

$$\lim_{n\to\infty} \frac{n!}{n^{n+\frac{1}{2}}e^{-n}} = k.$$

Since we have no clue for now as to what k is, save that it is positive, call it e^c, with c some other constant (we will be taking logarithms in a moment, so using e^c will make the notation a bit easier). Now,

$$\lim_{n\to\infty} \frac{n!}{n^{n+\frac{1}{2}}e^{-n}} = e^c$$

exactly when

$$\lim_{n\to\infty} \log\left(\frac{n!}{n^{n+\frac{1}{2}}e^{-n}}\right) = c.$$

Using that logarithms change multiplications and divisions into sums and differences, this is the same as

$$\lim_{n\to\infty} (\log(n!) - (n+\frac{1}{2})\log(n) + n) = c.$$

For notational convenience, set

$$d_n = \log(n!) - (n+\frac{1}{2})\log(n) + n.$$

We want to show that d_n converges to some number c as n goes to ∞. Here we use a trick. Consider the sequence

$$\sum_{i=1}^{n}(d_i - d_{i+1}) = (d_1 - d_2) + (d_2 - d_3) + \ldots (d_n - d_{n+1}) = d_1 - d_{n+1}.$$

We will show that the infinite series $\sum_{i=1}^{\infty}(d_i - d_{i+1})$ converges, which means that the partial sums $\sum_{i=1}^{n}(d_i - d_{i+1}) = d_1 - d_{n+1}$ converge. But this will mean that d_{n+1} will converge, which is our goal.

We will show that $\sum_{i=1}^{\infty}(d_i - d_{i+1})$ converges by the comparison test. Specifically, we will show that

$$|d_n - d_{n+1}| \leq \frac{2n+1}{2n^3} - \frac{1}{4n^2}.$$

Since both $\sum_{i=1}^{\infty} \frac{2n+1}{2n^3}$ and $\sum_{i=1}^{\infty} \frac{1}{4n^2}$ converge, our series will converge.

This will be a long calculation. We will need to use that, for any x with $|x| < \frac{2}{3}$,

$$\log(1+x) = x - \frac{x^2}{2} + \theta(x)$$

where $\theta(x)$ is a function such that for all $|x| < \frac{2}{3}$,

$$|\theta(x)| < |x|^3.$$

This follows from the Taylor series expansion of $\log(1+x)$. The requirement that $|x| < \frac{2}{3}$ is not critical; all we must do is make sure that our $|x|$ are sufficiently less than one.

Now,

$$\begin{aligned}
|d_n - d_{n+1}| &= [\log(n!) - (n+\frac{1}{2})\log(n) + n] - \\
&\quad [\log((n+1)!) - (n+1+\frac{1}{2})\log(n+1) + n + 1] \\
&= [\log(n) + \ldots + \log(1) - (n+\frac{1}{2})\log(n) + n] \\
&\quad -[\log(n+1) + \cdots + \log(1) \\
&\quad -(n+1+\frac{1}{2})\log(n+1) + n + 1] \\
&= -(n+\frac{1}{2})\log(n) + (n+\frac{1}{2})\log(n+1) - 1 \\
&= (n+\frac{1}{2})\log\left(\frac{n+1}{n}\right) - 1 \\
&= (n+\frac{1}{2})\log(1+\frac{1}{n}) - 1
\end{aligned}$$

$$
\begin{aligned}
&= (n+\frac{1}{2})(\frac{1}{n}-\frac{1}{2n^2}+\theta(\frac{1}{n}))-1\\
&= (n+\frac{1}{2})\theta(\frac{1}{n})-\frac{1}{4n^2}\\
&\le \frac{(n+\frac{1}{2})}{n^3}-\frac{1}{4n^2},
\end{aligned}
$$

which gives us our result. □

While Stirling's formula is important in its own right, we needed to use its following corollary in the proof of the Central Limit Theorem:

Corollary 15.6.1.1 *Let A be a constant. Then for $x_k \le A$, we have*

$$\binom{n}{k}p^k q^{n-k} \sim \frac{1}{\sqrt{2\pi npq}}e^{-\frac{x_k^2}{2}}.$$

Here the notation is the same as that used in the last section. In particular, if $S_n = k$, we set $S_n^* = x_k$. Then we have

$$k = np + \sqrt{npq}x_k,$$

and subtracting both sides of this equation from n, we have

$$n-k = n-np-\sqrt{npq}x_k = nq-\sqrt{npq}x_k.$$

If, as in the corollary, $x_k \le A$, then we must have

$$k \sim np$$

and

$$n-k \sim nq.$$

In the following proof, at a critical stage we will be replacing k by np and $n-k$ by nq.

Proof of Corollary: By definition

$$
\begin{aligned}
\binom{n}{k}p^k q^{n-k} &= \frac{n!}{k!(n-k)!}p^k q^{n-k}\\
&\sim \frac{(\frac{n}{e})^n\sqrt{2\pi n}}{(\frac{k}{e})^k\sqrt{2\pi k}(\frac{n-k}{e})^{n-k}\sqrt{2\pi(n-k)}}p^k q^{n-k},\\
&\quad \text{using Stirling's formula, which in turn yields}\\
&= \sqrt{\frac{n}{2\pi k(n-k)}}\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}\\
&\sim \sqrt{\frac{n}{2\pi(np)(nq)}}\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k},
\end{aligned}
$$

using here that $k \sim np$ and $n - k \sim nq$. This in turn equals

$$\sqrt{\frac{1}{2\pi npq}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}.$$

If we can show that

$$\left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} \sim e^{-\frac{x_k^2}{2}},$$

we will be done. Using that we can replace $\log(1+x)$ by $x - \frac{x^2}{2}$, for small x, we will show that

$$\log\left(\left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}\right) \sim -\frac{x_k^2}{2}.$$

Now

$$\begin{aligned}
\log\left(\left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}\right) &= k\log\left(\frac{np}{k}\right) + (n-k)\log\left(\frac{nq}{n-k}\right) \\
&= k\log\left(1 - \frac{\sqrt{npq}x_k}{k}\right) \\
&\quad +(n-k)\log\left(1 + \frac{\sqrt{npq}x_k}{n-k}\right),
\end{aligned}$$

using that the equality $k = np + \sqrt{npq}x_k$ implies

$$\frac{np}{k} = \frac{k - \sqrt{npq}x_k}{k} = 1 - \frac{\sqrt{npq}x_k}{k}$$

and a similar argument for the $(n-k)$. But then we can replace the log terms in the above to get

$$\begin{aligned}
&\sim k\left(-\frac{\sqrt{npq}x_k}{k} - \frac{npqx_k^2}{2k^2}\right) + (n-k)\left(\frac{\sqrt{npq}x_k}{n-k} - \frac{npqx_k^2}{2(n-k)^2}\right) \\
&= -\frac{npqx_k^2}{2k} - \frac{npqx_k^2}{2(n-k)} \\
&= -\frac{npqx_k^2}{2}\left(\frac{1}{k} + \frac{1}{n-k}\right) \\
&= -\frac{npqx_k^2}{2}\left(\frac{n}{k(n-k)}\right) \\
&= -\frac{x_k^2}{2}\left(\frac{np}{k}\right)\left(\frac{nq}{n-k}\right) \\
&\sim -\frac{x_k^2}{2},
\end{aligned}$$

since earlier we showed that $np \sim k$ and $nq \sim n-k$. □

The proof of Stirling's formula and of its corollary were full of clever manipulations. Part of the reason that these steps are shown here is to let people see that despite the abstract machinery of modern mathematics, there is still a need for cleverness at computations.

15.7 Books

From informed sources, Brualdi's book [14] is a good introduction to combinatorics. An excellent, but hard, text is by van Lint and Wilson [115]. Cameron's text [16] is also good. Polya, Tarjan and Woods' book [93] is fascinating. To get a feel of how current combinatorics is used, Graham, Knuth and Patashnik's [47] book is great. Stanley's text [105] is a standard text for beginning graduate students in combinatorics.

For probability theory, it is hard to imagine a better text than Feller [34]. This book is full of intuitions and wonderful, nontrivial examples. Grimmett and Stirzaker [50] is also a good place to begin. Another good source is Chung's book [18], which is where, as mentioned, I got the flow of the above argument for the Central Limit Theorem. More advanced work in probability theory is measure theoretic.

15.8 Exercises

1. The goal of this exercise is to see how to apply the definitions for probability to playing cards.

a. Given a standard deck of fifty-two cards, how many five card hands are possible (here order does not matter).

b. How many of these five card hands contain a pair? (This means that not only must there be a pair in the hand, but there cannot be a three-of-a-kind, two pair, etc.)

c. What is the probability of being dealt a hand with a pair?

2. The goal of this exercise is to see how the formulas for $\binom{n}{k}$ are linked to Pascal's triangle.

a. Prove by induction that

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}.$$

b. Prove this formula by counting how to choose k objects from n objects (order not mattering) in two different ways.

c. Prove that the binomial coefficients $\binom{n}{k}$ can be determined from Pascal's triangle, whose first five rows are:

$$\begin{array}{ccccccccc} &&&&1&&&& \\ &&&1&&1&&& \\ &&1&&2&&1&& \\ &1&&3&&3&&1& \\ 1&&4&&6&&4&&1 \end{array}$$

d. Give a combinatorial proof of the identity

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n.$$

4. Find a formula for determining how many monomials of degree k can be made out of n variables. (Thus for the two variables x,y, the number of monomials of degree two is three, since we can simply count the list

$$(x^2, xy, y^2).)$$

5. The pigeonhole principle states:
If (n+1) objects are placed into n different boxes, at least one box must have at least two objects in it.

Let $a_1, \ldots, a_{n+1}$ be integers. Show that there is at least one pair of these integers such that $a_i - a_j$ is divisible by the integer n.

6. The goal of this problem is to prove the Inclusion-Exclusion Principle, the statement of which is part c.

a. Let A and B be any two sets. Show that

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

b. Let A_1, A_2 and A_3 be any three sets. Show that

$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3|.$$

c. Let $A_1, \ldots, A_n$ be any n sets. Show that

$$|A_1 \cup \ldots \cup A_n| = \Sigma|A_i| - \Sigma|A_i \cap A_j| + \ldots + (-1)^{n+1}|A_1 \cap \ldots \cap A_n|.$$

7. Show that

$$\binom{2n}{n} \sim (\pi n)^{-1/2} 2^{2n}.$$

Chapter 16

Algorithms

Basic Object:	Graphs and Trees
Basic Goal:	Computing the Efficiency of Algorithms

The end of the 1800s and the beginning of the 1900s saw intense debate about the meaning of existence for mathematical objects. To some, a mathematical object could only have meaning if there was a method to compute it. For others, any definition that did not lead to a contradiction would be good enough to guarantee existence (and this is the path that mathematicians have overwhelmingly chosen to take). Think back to the section on the Axiom of Choice in Chapter Ten. Here objects were claimed to exist which were impossible to actually construct. In many ways these debates had quieted down by the 1930s, in part due to Gödel's work, but also in part due to the nature of the algorithms that were eventually being produced. By the late 1800s, the objects that were being supposedly constructed by algorithms were so cumbersome and time-consuming, that no human could ever compute them by hand. To most people, the pragmatic difference between an existence argument versus a computation that would take a human the life of the universe was too small to care about, especially if the existence proof had a clean feel.

All of this changed with the advent of computers. Suddenly, calculations that would take many lifetimes by hand could be easily completed in millionths of a second on a personal computer. Standard software packages like Mathematica and Maple can outcompute the wildest dreams of a mathematician from just a short time ago. Computers, though, seem to have problems with existence proofs. The need for constructive arguments returned with force, but now came a real concern with the efficiency of the construction, or the complexity of the algorithm. The idea that certain constructions have an intrinsic complexity has increasingly become basic in most branches of mathematics.

16.1 Algorithms and Complexity

An accurate, specific definition for an algorithm is non-trivial and not very enlightening. As stated in the beginning of Cormen, Leiserson and Rivest's book *Introduction to Algorithms* [22],

> *Informally, an* **algorithm** *is any well-defined computational procedure that takes some value, or set of values, as* **input** *and produces some values, or set of values, as* **output**. *An algorithm is thus a sequence of computational steps that transform the input into the output.*

Much of what has been discussed in this book can be recast into the language of algorithms. Certainly, much of the first chapter on linear algebra, such as the definition of the determinant and Gaussian elimination, is fundamentally algorithmic in nature.

We are concerned with the efficiency of an algorithm. Here we need to be concerned with asymptotic bounds on the growth of functions.

Definition 16.1.1 *Let $f(x)$ and $g(x)$ be two one-variable real-valued functions. We say that $f(x)$ is in* O(g(x)) *if there exists a positive constant C and a positive number N so that for all $x > N$, we have $|f(x)| \leq C|g(x)|$.*

This is informally known as *big O* notation.

Typically we do not use the symbol "x" for our variable but "n". Then the class of functions in $O(n)$ will be those that grow at most linearly, those in $O(n^2)$ grow at most quadratically, etc. Thus the polynomial $3n^4+7n-19$ is in $O(n^4)$.

For an algorithm there is the *input size*, n, which is how much information needs to be initially given, and the *running time*, which is how long the algorithm takes as a function of the input size. An algorithm is *linear* if the running time $r(n)$ is in $O(n)$, *polynomial* if the running time $r(n)$ is in $O(n^k)$ for some integer k, etc.

There are further concerns, such as the space size of an algorithm, which is how much space the algorithm requires in order to run as a function of the input size.

16.2 Graphs: Euler and Hamiltonian Circuits

An analysis of most current algorithms frequently comes down to studying graphs. This section will define graphs and then discuss graphs that have Euler circuits and Hamiltonian circuits. We will see that while these two have similar looking definitions, their algorithmic properties are quite different.

Intuitively a graph looks like:

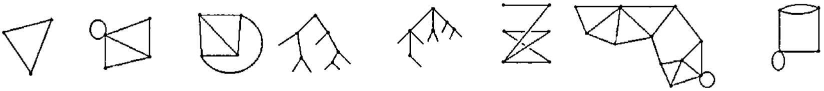

The key is that a graph consists of vertices and edges between vertices. All that matters is which vertices are linked by edges. Thus we will want these two graphs, which have different pictures in the plane, to be viewed as equivalent.

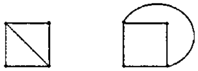

Definition 16.2.1 *A graph G consists of a set $V(g)$, called* vertices, *and a set $E(G)$, called* edges, *and a function*

$$\sigma : E(G) \rightarrow \{\{u,v\} : u,v \in V(G)\}.$$

We say that elements v_i and v_j in $V(G)$ are connected by an edge e if $\sigma(e) = \{v_i, v_j\}$.

Note that $\{v_i, v_j\}$ denotes the set consisting of the two vertices v_i and v_j.

For the graph G:

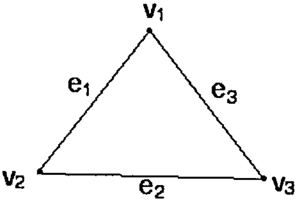

we have

$$V(G) = \{v_1, v_2, v_3\}$$
$$E(G) = \{e_1, e_2, e_3\}$$

and

$$\sigma(e_1) = \{v_1, v_2\}, \sigma(e_2) = \{v_2, v_3\}, \sigma(e_3) = \{v_1, v_3\}.$$

Associated to a graph is its *adjacency* matrix $A(G)$. If there are n vertices, this will be the following $n \times n$ matrix. List the vertices:

$$V(G) = \{v_1, v_2, ..., v_n\}.$$

For the (i, j)-entry of the matrix, put in a k if there are k edges between v_i and v_j and a 0 otherwise. Thus the adjacency matrix for:

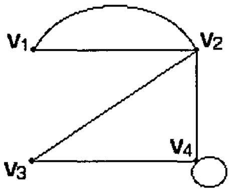

will be the 4 × 4 matrix:

$$A(G) = \begin{pmatrix} 0 & 2 & 0 & 0 \\ 2 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}.$$

The '1' in the $(4,4)$ entry reflects that there is an edge from v_4 to itself and the '2' in the $(1,2)$ and $(2,1)$ entries reflects that there are two edges from v_1 to v_2.

A *path* in a graph G is a sequence of edges that link up with each other. A *circuit* is a path that starts and ends at the same vertex. For example, in the graph:

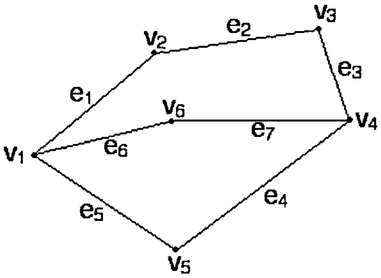

the path e_6e_7 starts at vertex v_1 and ends at v_4 while $e_1e_2e_3e_4e_5$ is a circuit starting and ending at v_1.

We can now start to talk about Euler circuits. We will follow the traditional approach and look first of the Königsberg bridge problem. The town of Königsberg had the following arrangement:

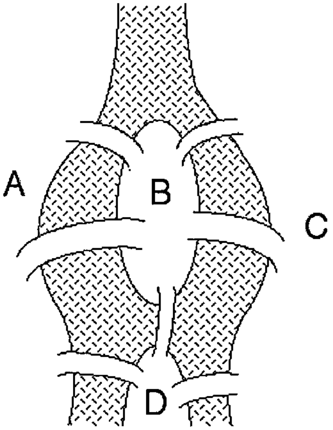

Here A, B, C and D denote land.

The story goes that in the 1700s, the people of Königsberg would try to see if they could cross every bridge exactly once so that at the end they returned to their starting spot. Euler translated this game into a graph theory question. To each connected piece of land he assigned a vertex and to each bridge between pieces of land he assigned an edge. Thus Königsberg became the graph

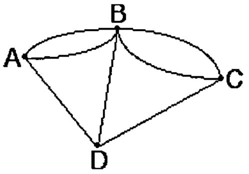

Then the game will be solved if in this graph there is a circuit that contains each edge exactly once. Such circuits have a special name, in honor of Euler:

Definition 16.2.2 *An* Euler circuit *on a graph is a circuit that contains each edge exactly once.*

To solve the Königsberg bridge problem, Euler came up with a clean criterion for when any graph will have an Euler circuit.

Theorem 16.2.1 *A graph has an Euler circuit if and only if each vertex has an even number of edges coming into it.*

Thus in Königsberg, since vertex A is on three edges (and in this case every other vertex also has an odd number of edges), no one can cross each bridge just once.

The fact that each vertex must be on an even number of edges is not that hard to see. Suppose we have an Euler circuit. Imagine deleting each edge as we transverse the graph. Each time we enter, then leave, a vertex, two edges are deleted, reducing the number of edges containing that vertex by two. By the end, there are no edges left, meaning that the original number of edges at each vertex had to be even.

The reverse direction is a bit more complicated but is more important. The best method (which we will not do) is to actually construct an algorithm that produces an Euler circuit. For us, the important point is that there is a clean, easy criterion for determining when an Euler circuit exists.

Let us now make a seemingly minor change in the definition for an Euler circuit. Instead of finding a circuit that contains each edge only once, now let us try to find one that contains each vertex only once. These circuits are called:

Definition 16.2.3 *A graph has a* Hamiltonian circuit *if there is a circuit that contains each vertex exactly once.*

For example, for the graph:

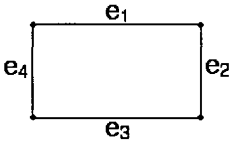

the circuit $e_1e_2e_3e_4$ is Hamiltonian, while for the graph:

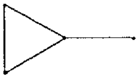

there is no Hamiltonian circuit. In this last graph, one can simply list all possible circuits and then just check if one of them is Hamiltonian. This algorithm of just listing all possible circuits will work for any graph, as there can only be a finite number of circuits, but this listing unfortunately takes $O(n!)$ time, where n is the number of edges. For any graph with a fair number of edges, this approach is prohibitively time-consuming. But this is fairly close to the best known method for determining if a Hamiltonian circuit exists. As we will see in section four, the problem of finding a Hamiltonian circuit seems to be intrinsically difficult and important.

16.3 Sorting and Trees

Suppose you are given a set of real numbers. Frequently you want to order the set from smallest number to largest. Similarly, suppose a stack of exams is sitting on your desk. You might want to put the exams into alphabetical order. Both of these problems are sorting problems. A sorting algorithm will take a collection of elements for which an ordering can exist and actually produce the ordering. This section will discuss how this is related to a special class of graphs called *trees* and that the lower bound for any sorting algorithm is $O(n\log(n))$.

Technically a *tree* is any graph that is connected (meaning that there is a path from any vertex to any other vertex) and contains within it no circuits. Thus

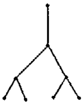

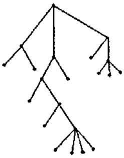

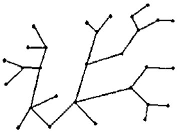

are trees while

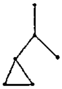

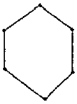

are not. Those vertices contained on exactly one edge are called *leaves.* These are in some sense the vertices where the tree stops. We will be concerned with *binary* trees, which are constructed as follows. Start with a vertex called the *root.* Let two edges come out from the root. From each of the two new vertices at the end of the two edges, either let two new edges stem out or stop. Continue this process a finite number of steps. Such a tree looks like:

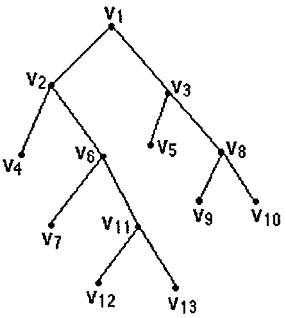

where v_1 is the root and v_4, v_5, v_7, v_9, v_{10}, v_{12} and v_{13} are the leaves. We will draw our binary trees top down, with the root at the top and the leaves at the bottom. At each vertex, the two edges that stem down are called the *left* edge and *right* edge, respectively. The two vertices at the ends of these edges are called the *left* child and the *right* child, respectively. The *height* of a tree is the number of edges in the longest path from the root to a leaf. Thus the height of

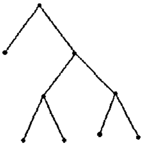

is three while the height of

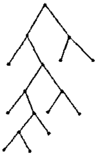

is six.

We now want to see why sorting is linked to binary trees. We are given a collection of elements $\{a_1, \ldots, a_n\}$. We will assume that all we can do is compare the size of any two elements. Thus given, say, elements a_i and a_j, we can determine if $a_i < a_j$ or if $a_j < a_i$. Any such sorting algorithm can only, at each stage, take two a_i and a_j and, based on which is larger, tell us what to do at the next stage. Now to show that any such algorithm can be represented as a tree. The root will correspond to the first pair to be compared in the algorithm. Say this first pair is a_i and a_j. There are two possibilities for the order of a_i and a_j. If $a_i < a_j$, go down the left edge and if $a_j < a_i$, go down the right edge. An algorithm will tell us at this stage which pair of elements to now compare. Label the new vertices by these pairs. Continue this process until there is nothing left to compare. Thus we will have a tree, with each vertex labeled by a pair of elements in our set and each leaf corresponding to an ordering of the set.

For example, take a three element set $\{a_1, a_2, a_3\}$. Consider the following simple algorithm (if anything this easy deserves to be called an algorithm):

Compare a_1 and a_2. If $a_1 < a_2$, compare a_2 and a_3. If $a_2 < a_3$, then the ordering is $a_1 < a_2 < a_3$. If $a_3 < a_2$, compare a_1 and a_3. If $a_1 < a_3$, then the ordering is $a_1 < a_3 < a_2$. If we had $a_3 < a_1$, then the ordering is $a_3 < a_1 < a_2$. Now we go back to the case when $a_2 < a_1$. Then we next compare a_1 and a_3. If $a_1 < a_3$, the ordering is $a_2 < a_1 < a_3$. If we have $a_3 < a_1$, we compare a_2 and a_3. If $a_2 < a_3$, then the ordering is $a_2 < a_3 < a_1$. If $a_3 < a_2$, then the ordering is $a_3 < a_2 < a_1$ and we are done. Even for this simple example, the steps, presented in this manner, are confusing. But when this method is represented as a tree it becomes clear:

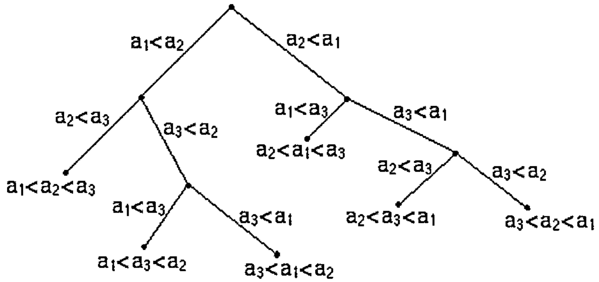

We now want to show that for a binary tree there is an intrinsic lower bound on its height, which means that there is an intrinsic lower bound on the time needed to sort.

Theorem 16.3.1 *A binary tree of height n has at most 2^n leaves.*

Proof: By induction. Suppose the height is zero. This means that the tree is a single vertex and thus has $2^0 = 1$ leaf, which of course in this case is also the root and is easy to sort.

Now suppose that we know the theorem is true for any tree of height $n-1$. Look at a tree of height n. Thus there is at least one path from the root to a leaf with length n. Remove all leaves, and their attaching edges, that are of length n from the root. We have a new tree of height $n-1$. The induction hypothesis kicks in, so we know that for this new tree there are at most 2^{n-1} leaves. Let two edges stem out from each of these 2^{n-1} leaves, forming still another new tree which has height n and which contains our original tree. But we are adding two new vertices for each of the 2^{n-1} leaves of the tree of height $n-1$. Thus this final new tree has at most $2 \cdot 2^{n-1} = 2^n$ leaves. Since each leaf of our original tree is a leaf of this tree, we have our result. □

This allows us to finally see that any algorithm that sorts n objects must be in at least $O(n \log(n))$.

Theorem 16.3.2 *Any sorting algorithm based on pairwise comparisons must be in at least $O(n \log(n))$.*

Proof: Given a set of n elements, there are $n!$ different ways they can be initially ordered. For any sorting algorithm, for the corresponding tree there must be a way, starting with the root, to get to one of these $n!$ different initial orderings. Thus the tree must have at least $n!$ leaves. Thus from the previous theorem, the tree must have height at least h, where

$$2^h \geq n!.$$

Thus we must have

$$h \geq \log_2(n!).$$

Any sorting algorithm must take at least h steps and hence must be in at least $O(\log_2(n!))$. Now we have, for any number K, $\log(K) = \log(2)\log_2(K)$, where of course, log is here the natural log, $\log_e$. Further, by Stirling's formula, we have for large n that

$$n! \sim \sqrt{2\pi n}\, n^n e^{-n}.$$

Then

$$\log(n!) \sim \log(\sqrt{2\pi n}) + n\log(n) - n\log(e),$$

which gives us that

$$\begin{aligned} O(\log(n!)) &= O(\log(\sqrt{2\pi n}) + n\log(n) - n\log(e)) \\ &= O(n\log(n)), \end{aligned}$$

since $n\log(n)$ dominates the other terms. Thus the complexity of any sorting algorithm is in at least $O(\log_2(n!)$, which equals $O(n\log(n))$, as desired. □

To show that sorting is actually equal to $O(n\log(n))$, we would need to find an algorithm that runs in $O(n\log(n))$. Heapsort, merge and other algorithms for sorting do exist that are in $O(n\log(n))$.

16.4 P=NP?

The goal of this section is to discuss what is possibly the most important open problem in mathematics: "P=NP?". This problem focuses on trying to determine the difference between the finding of a solution for a problem and the checking of a candidate solution for the problem. The fact that it remains open (and that it could well be independent of the other axioms of mathematics) shows that mathematicians do not yet understand the full meaning of mathematical existence versus construction.

A problem is in *polynomial* time if, given input size n, there is an algorithm that is in $O(n^k)$, for some positive integer k. A problem is in NP if, given input size n, a candidate solution can be checked for accuracy in polynomial time. The N in the NP is somewhat of a joke; NP stands for "not polynomial".

Think of a jigsaw puzzle. While it can be quite time consuming to put a jigsaw puzzle together, it is easy and quick to tell if someone has finished such a puzzle. For a more mathematical example, try to invert an $n \times n$ matrix A. While doable, it is not particularly easy to actually construct A^{-1}. But if someone hands us a matrix B and claims that it is the inverse, all we have to do to check is to multiply out AB and see if we get the identity I. For another example, start with a graph G. It is difficult to determine if G contains a Hamiltonian circuit. But if someone hands us a candidate circuit, it is easy to check whether or not the circuit goes through every vertex exactly once. Certainly it appears that the problem of finding a solution should be intrinsically more difficult than the problem of checking the accuracy of a solution.

Amazingly enough, people do not know if the class of NP problems is larger than the class of polynomial time problems (which are denoted as P problems). "P=NP" is the question:

Is the class of problems in P equal to the class of problems in NP?

This has been open for many years. While initially the smart money was on P≠NP, today the belief is increasingly that statement 'P=NP' is independent of the other axioms of mathematics. Few believe that P=NP.

Even more intriguing is the existence of NP *complete* problems. Such a problem is not only in NP but also must be a yes/no question and, most importantly, every other NP problem must be capable of being translated into this problem in polynomial time. Thus if there is a polynomial time solution to this NP yes/no problem, there will be a polynomial time solution of every NP problem.

Every area of math seems to have its own NP complete problems. For example, the question of whether or not a graph contains a Hamiltonian circuit is a quintessential NP complete problem and, since it can be explained with little high level math, is a popular choice in expository works.

16.5 Numerical Analysis: Newton's Method

Since the discovery of calculus, there has been work on finding answers to math questions that people can actually use. Frequently this comes down to only finding approximate solutions. Numerical Analysis is the field that tries to find approximate solutions to exact problems. How good of an approximation is good enough and how quickly the approximation can be found are the basic questions for a numerical analyst. While the roots of this subject are centuries old, the rise of computers has revolutionized the field. An algorithm that is unreasonable to perform by hand can often be easily solved for a standard computer. Since numerical analysis is ultimately concerned with the efficiency of algorithms, I have put this section in this chapter. It must be noted that in the current math world, numerical analysts and people in complexity theory are not viewed as being in the same subdiscipline. This is not to imply that they don't talk to each other; more that complexity theory has evolved from computer science and numerical analysis has always been a part of mathematics.

There are certain touchstone problems in numerical analysis, problems that are returned to again and again. Certainly efficient algorithms for computations in linear algebra are always important. Another, which we will be concerned with here, is the problem of finding zeros of functions. Many problems in math can be recast into finding a zero of a function. We will first look at Newton's method for approximating a zero of a real valued differentiable function $f : \mathbf{R} \to \mathbf{R}$, and then quickly see how the ideas behind this method can be used, at times, to approximate the zeros of other types of functions.

Let $f : \mathbf{R} \to \mathbf{R}$ be a differentiable function. We will first outline the geometry behind Newton's method. Suppose we know its graph (which of course in real life we will rarely know; otherwise the problem of approximating zeros would be easy) to be:

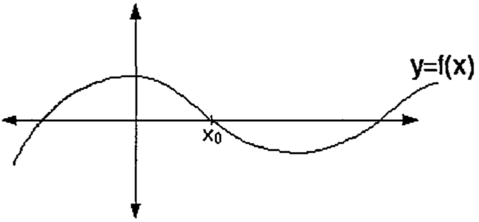

We thus want to approximate the point x_0. Choose any point x_1. Draw the tangent line to the curve $y = f(x)$ at the point $(x_1, f(x_1))$ and label its intersection with the x-axis by $(x_2, 0)$.

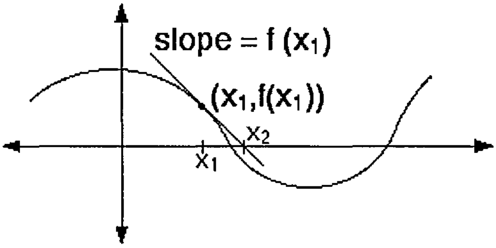

Then we have

$$f'(x_1) = \frac{0 - f(x_1)}{x_2 - x_1},$$

which, solving for x_2, yields

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

In the picture, it looks like our newly constructed x_2 is closer to our desired x_0 than is x_1. Let us try the same thing but replacing the x_1's with x_2. We label x_3 as the x-coordinate of the point of intersection of the tangent line of $y = f(x)$ through the point $(x_2, f(x_2))$ and get:

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)}.$$

Again, it at least looks like x_3 is getting closer to x_0. Newton's method is to continue this process, namely to set

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}.$$

For this to work, we need $x_k \to x_0$. There are difficulties. Consider the picture:

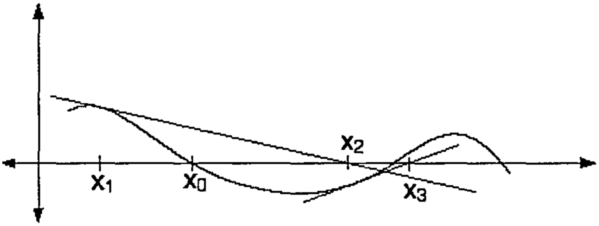

With this choice of initial x_1, the x_k will certainly not approach the zero x_0, though they do appear to approach a different zero. The problem of course is that this choice of x_1 is near a local maximum, which means that the derivative $f'(x_1)$ is very small, forcing $x_2 = x_1 - f(x_1)/f'(x_1)$ to be far from x_0.

We will now make this technically correct. Here we will see many ideas from calculus playing a critical role in proving that Newton's method will, subject to specific conditions, always produce an approximation to the true zero. We will look at functions $f : [a, b] \to [a, b]$ which have continuous second derivatives, i.e., functions in the vector space $C^2[a, b]$. As an aside, we will be using throughout the Mean Value Theorem, which states that for any function $f \in C^2[a,b]$, there exists a number c with $a \leq c \leq b$ such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Our goal is:

Theorem 16.5.1 *Let $f \in C^2[a,b]$. Suppose there exists a point $x_0 \in [a,b]$ with $f(x_0) = 0$ but $f'(x_0) \neq 0$. Then there exists a $\delta > 0$ such that, given any point x_1 in $[x_0 - \delta, x_0 + \delta]$, if for all k we define*

$$x_k = x_{k-1} - \frac{f(x_{k-1})}{f'(x_{k-1})},$$

we have that $x_k \to x_0$.

This theorem states that Newton's method will produce an approximation of the zero provided our initial choice x_1 is close enough to the zero.
Proof: We will alter the problem from finding a zero of a function f to the finding of a fixed point of a function g. Set

$$g(x) = x - \frac{f(x)}{f'(x)}.$$

Note that $f(x_0) = 0$ if and only if $g(x_0) = x_0$. We will show that Newton's method will produce an approximation to a fixed point of g.

We first need to see how to choose our $\delta > 0$. By taking derivatives and doing a bit of algebra, we have

$$g'(x) = \frac{f(x)f''(x)}{(f'(x))^2}.$$

Since the second derivative of f is still a continuous function, we have that $g'(x)$ is a continuous function. Further, since $f(x_0) = 0$, we have that $g'(x_0) = 0$. By continuity, given any positive number α, there exists a $\delta > 0$ such that for all $x \in [x_0 - \delta, x_0 + \delta]$, we have

$$|g'(x)| < \alpha.$$

We choose α to be strictly less than one (the reason for this restriction will be clear in a moment).

We will reduce the problem to proving the following three lemmas:

Lemma 16.5.1 *Let $g : [a,b] \to [a,b]$ be any continuous function. Then there is a fixed point in $[a,b]$.*

Lemma 16.5.2 *Let $g : [a,b] \to [a,b]$ be any differentiable function such that for all $x \in [a,b]$ we have*

$$|g'(x)| < \alpha < 1$$

for some constant α. Then there is a unique fixed point in the interval $[a,b]$.

Lemma 16.5.3 *Let $g : [a,b] \to [a,b]$ be any differentiable function such that for all $x \in [a,b]$ we have*

$$|g'(x)| < \alpha < 1$$

for some constant α. Then given any $x_1 \in [a,b]$, if we set

$$x_{k+1} = g(x_k),$$

then the x_k will approach the fixed point of g.

Assume briefly that all three lemmas are true. Note by our choice of δ, we have the function $g(x) = x - \frac{f(x)}{f'(x)}$ satisfying each of the conditions in the above lemma. Further we know that the zero x_0 of the function $f(x)$ is the fixed point of $g(x)$. Then we know that iterating any point in $[x_0 - \delta, x_0 + \delta]$ by g, we will approach x_0. But writing out this iteration is precisely Newton's method.

Now to prove the lemmas.

Proof of first lemma: This will be a simple application of the Intermediate Value Theorem. If $g(a) = a$ or if $g(b) = b$, then a or b is our fixed point and we are done. Suppose neither holds. Since the range of g is contained in the interval $[a,b]$, this means that

$$a < g(a) \text{ and } b > g(b).$$

Set

$$h(x) = x - g(x).$$

This new function is continuous and has the property that

$$h(a) = a - g(a) < 0$$

and

$$h(b) = b - g(b) > 0.$$

By the Intermediate Value Theorem, there must be a $c \in [a,b]$ with

$$h(c) = c - g(c) = 0$$

giving us our fixed point. □

Proof of second lemma: We will now use the Mean Value Theorem. Suppose there are two distinct fixed points, c_1 and c_2. Label these points so that $c_1 < c_2$. By the Mean Value Theorem, there is some number c with $c_1 \leq c \leq c_2$ such that

$$\frac{g(c_2) - g(c_1)}{c_2 - c_1} = g'(c).$$

Since $g(c_1) = c_1$ and $g(c_2) = c_2$, we have

$$g'(c) = \frac{c_2 - c_1}{c_2 - c_1} = 1.$$

Here is our contradiction, as we assumed that at all points that the absolute value of the derivative was strictly less than one. There cannot be two fixed points. □

Proof of third lemma: This will be another application of the Mean Value Theorem. By the second lemma, we know that g has a unique fixed point. Call this fixed point x_0. We will regularly replace x_0 by $g(x_0)$.

Our goal is to show that $|x_k - x_0| \to 0$. We will show that for all k

$$|x_k - x_0| \leq \alpha |x_{k-1} - x_0|.$$

Then by shifting subscripts we will have

$$|x_{k-1} - x_0| \leq \alpha |x_{k-2} - x_0|,$$

which will mean that

$$|x_k - x_0| \leq \alpha |x_{k-1} - x_0| \leq \alpha^2 |x_{k-2} - x_0| \leq \ldots \leq \alpha^k |x_1 - x_0|.$$

Since α is strictly less than one, we will have $|x_k - x_0| \to 0$.

Now

$$|x_k - x_0| = |g(x_{k-1}) - g(x_0)|.$$

By the Mean Value Theorem, there is some point c between x_0 and x_{k-1} with

$$\frac{g(x_{k-1}) - g(x_0)}{x_{k-1} - x_0} = g'(c)$$

which is equivalent to

$$g(x_{k-1}) - g(x_0) = g'(c)(x_{k-1} - x_0).$$

Then

$$|g(x_{k-1}) - g(x_0)| = |g'(c)||x_{k-1} - x_0|.$$

Now we just have to observe that by assumption $|g'(c)| \leq \alpha$, and we are done. □

All this theorem is telling us is that if we start with an initial point close enough to the zero of a function, Newton's method will indeed converge to the zero. It does not tell us how to make our initial choice and does not tell us the speed of the convergence.

Now let us see how to try to use Newton's method in other contexts. Suppose we have a map $L : V \to W$ from one vector space to another. How can we approximate a zero of this map? Let us assume that there is some notion of a derivative for the map L, which we will denote by DL. Then just formally following the Newton's method, we might, starting with any element $v_1 \in V$, recursively define

$$v_{k+1} = v_k - DL(v_k)^{-1} L(v_k)$$

and hope that the v_k will approach the zero of the map. This could be at least an outline of a general approach. The difficulties are in understanding DL and in particular in dealing with when DL has some type of inverse.

For example, consider a function $F : \mathbf{R}^2 \to \mathbf{R}^2$, given in local coordinates by

$$F(x,y) = (f_1(x,y), f_2(x,y)).$$

The derivative of F should be the two-by-two Jacobian matrix

$$DF = \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix}.$$

Starting with any $(x_1, y_1) \in \mathbf{R}^2$, we set

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ y_k \end{pmatrix} - DF^{-1}(x_k, y_k) \cdot \begin{pmatrix} f_1(x_k, y_k) \\ f_2(x_k, y_k) \end{pmatrix}.$$

Newton's method will work if the (x_k, y_k) approach a zero of F. By placing appropriate restrictions on the zero of F, such as requiring that $\det(DF(x_0, y_0)) \neq 0$, we can find an analogous proof to the one-dimensional case. In fact, it generalizes to any finite dimension.

More difficult problems occur for infinite dimensional spaces V and W. These naturally show up in the study of differential equations. People still try to follow a Newton-type method, but now the difficulty of dealing with the right notion for DL becomes a major stumbling block. This is why in trying to solve differential equations you are led to the study of infinite dimensional linear maps and are concerned with the behavior of the eigenvalues, since you want to control and understand what happens when the eigenvalues are, or are close to, zero, for this is the key to controlling the inverse of DL. The study of such eigenvalue questions falls under the rubric of Spectral Theorems, which is why the Spectral Theorem is a major part of beginning Functional Analysis and a major tool in PDE theory.

16.6 Books

The basic text for algorithms is *Introduction to Algorithms* by Cormen, Leiserson and Rivest [22]. Another source is *Data Structures and Algorithms* by Aho, Hopcroft, and Ullman [2].

Numerical Analysis has a long history. Further, many people, with widely varying mathematical backgrounds, need to learn some numerical analysis. Thus there are many beginning texts (though it must be stated that my knowledge of these texts is limited).

Atkinson's *Introduction to Numerical Analysis* [5] comes highly recommended. Another basic text that has long been the main reference for people studying for the numerical methods part of the actuarial exams is *Numerical Methods* by Burden and Faires [15]. Trefethon and Bau's text [112] is a good source for numerical methods for linear algebra. For numerical methods for differential equations, good sources are the books by Iserles [66] and Strikwerda [110]. Finally, for links with optimization theory, there is Ciarlet's *Introduction to Numerical Linear Algebra and Optimization* [19].

16.7 Exercises

1. Show that there are infinitely many nonisomorphic graphs, each having exactly k vertices.

2. How many nonisomorphic graphs with exactly three vertices and four edges are there?

3. Assume that the time for multiplying and adding two numbers together is exactly one.

a. Find an algorithm that runs in time (n-1) that adds n numbers together.

b. Find an algorithm that computes the dot product of two vectors in $\mathbf{R}^2$ in time (2n-1).

c. Assume that we can work in parallel, meaning that we allow algorithms that can compute items that do not depend on each other simultaneously. Show that we can add n numbers together in time $\log_2(n-1)$.

d. Find an algorithm that computes the dot product of two vectors in $\mathbf{R}^2$ in parallel in time $\log_2(n)$.

4. Let A be the adjacency matrix for a graph G.

a. Show that there is a nonzero (i, j) entry of the matrix A^2 if and only if there is a path containing two edges from vertex i to vertex j.

b. Generalize part (a) to linking entries in the matrix A^k to the existence of paths between various vertices having exactly k edges.

c. Find an algorithm that determines whether or not a given graph is connected.

5. Use Newton's method, with a calculator, to approximate $\sqrt{2}$ by approximating a root of the polynomial $x^2 - 2$.

6. Let $f : \mathbf{R}^n \to \mathbf{R}^n$ be any differentiable function from $\mathbf{R}^n$ to itself. Let x_0 be a point in $\mathbf{R}^n$ with $f(x_0) = 0$ but with $\det(Df(x_0)) \neq 0$, where Df denotes the Jacobian of the function f. Find a function $g : \mathbf{R}^n \to \mathbf{R}^n$ that has the point x_0 as a fixed point.

Appendix A

Equivalence Relations

Throughout this text we have used equivalence relations. Here we collect some of the basic facts about equivalence relations. In essence, an equivalence relation is a generalization of equality.

Definition A.0.1 (Equivalence Relation) *An equivalence relation on a set X is any relation '$x \sim y$' for $x, y \in X$ such that*

1. *(Reflexivity) For any $x \in X$, we have $x \sim x$.*
2. *(Symmetry) For all $x, y \in X$, if $x \sim y$ then $y \sim x$.*
3. *(Transitivity) For all $x, y, z \in X$, if $x \sim y$ and $y \sim z$, then $x \sim z$.*

The basic example is that of equality. Another example would be when $X = \mathbf{R}$ and we say that $x \sim y$ if $x - y$ is an integer. On the other hand, the relation $x \sim y$ if $x \leq y$ is not an equivalence relation, as it is not symmetric.

We can also define equivalence relations in term of subsets of the ordered pairs $X \times X$ as follows:

Definition A.0.2 (Equivalence Relation) *An equivalence relation on a set X is a subset $R \subset X \times X$ such that*

1. *(Reflexivity) For any $x \in X$, we have $(x, x) \in R$.*
2. *(Symmetry) For all $x, y \in X$, if $(x, y) \in R$ then $(y, x) \in R$.*
3. *(Transitivity) For all $x, y, z \in X$, if $(x, y) \in R$ and $(y, z) \in R$, the $(x, z) \in R$.*

The link between the two definitions is of course that $x \sim y$ means the same as $(x, y) \in R$.

An equivalence relation will split the set X into disjoint subsets, the equivalence classes.

Definition A.0.3 (Equivalence Classes) *An* equivalence class C *is a subset of* X *such that if* $x, y \in C$*, then* $x \sim y$ *and if* $x \in C$ *and if* $x \sim y$*, then* $y \in C$.

The various equivalence classes are disjoint, a fact that follows from transitivity.

Exercises: 1. Let G be a group and H a subgroup. Define, for $x, y \in G$, $x \sim y$, whenever $xy^{-1} \in H$. Show that this forms an equivalence relation on the group G.

2. For any two sets A and B, define $A \sim B$ if there is a one-to-one, onto map from A to B. Show that this is an equivalence relation.

3. Let (v_1, v_2, v_3) and (w_1, w_2, w_3) be two collections of three vectors in $\mathbf{R}^3$. Define $(v_1, v_2, v_3) \sim (w_1, w_2, w_3)$ if there is an element $A \in \mathbf{GL(n,R)}$ such that $Av_1 = w_1$, $Av_2 = w_2$ and $Av_3 = w_3$. Show that this is an equivalence relation.

4. On the real numbers, say that $x \sim y$ if $x - y$ is a rational number. Show that this forms an equivalence relation on the real numbers. (This equivalence was used in Chapter Ten, in the proof that there exists non-measurable sets.)

Bibliography

[1] Ahlfors, Lars V., *Complex Analysis: An Introduction to the Theory of Analytic Functions of One Complex Variable*, Third edition, International Series in Pure and Applied Mathematics, McGraw-Hill Book Co., New York, 1978. xi+331 pp.

[2] Aho, A., Hopcroft, J. and Ullman, J., *The Design and Analysis of Computer Algorithms*, Addison-Wesley, Reading, NY, 1974.

[3] Artin, E., *Galois Theory*, (edited and with a supplemental chapter by Arthur N. Milgram), Reprint of the 1944 second edition, Dover Publications, Inc., Mineola, NY, 1998. iv+82 pp.

[4] Artin, M., *Algebra*, Prentice Hall, 1995. 672 pp.

[5] Atkinson, Kendall E., *An Introduction to Numerical Analysis*, Second edition, John Wiley and Sons, Inc., New York, 1989. xvi+693 pp.

[6] Bartle, Robert G., *The Elements of Real Analysis*, Second edition, John Wiley and Sons, New York-London-Sydney, 1976. xv+480 pp.

[7] Berberian, Sterling K., *A First Course in Real Analysis*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1998. xii+237 pp.

[8] Berenstein, Carlos A.and Gay, Roger, *Complex Variables: An Introduction*, Graduate Texts in Mathematics, 125. Springer-Verlag, New York, 1997. xii+650 pp.

[9] Birkhoff, G. and Mac Lane, S, *A Survey of Modern Algebra*, Akp Classics, A K Peters Ltd, 1997. 512 pp.

[10] Bocher, M., *Introduction to Higher Algebra*, MacMillan, New York, 1907.

[11] Bollobas, B., *Graph Theory: An Introductory Course*, Graduate Texts in Mathematics, 63, Springer-Verlag, New York-Berlin, 1979. x+180 pp.

[12] Boyce, W.F. and Diprima, R. C., *Elementary Differential Equations and Boundary Value Problems*, Sixth Edition, John Wiley and Sons, 1996, 768 pp.

[13] Bressoud, David M., *A Radical Approach to Real Analysis* Classroom Resource Materials Series, 2, Mathematical Association of America, Washington, DC, 1994. xii+324 pp.

[14] Brualdi, Richard A., *Introductory Combinatorics*, Second edition, North-Holland Publishing Co., New York, 1992. xiv+618 pp.

[15] Burden, R. and Faires, J., *Numerical Methods*, Seventh edition, Brooks/Cole Publishing Co., Pacific Grove, CA, 2001. 810 pp

[16] Cameron, Peter J., *Combinatorics: Topics, Techniques, Algorithms*, Cambridge University Press, Cambridge, 1995. x+355 pp.

[17] Cederberg, Judith N, *A Course in Modern Geometries*, Second edition, Undergraduate Texts in Mathematics, Springer-Verlag, New York-Berlin, 2001. xix+439 pp.

[18] Chung, Kai Lai, *Elementary Probability Theory with Stochastic Processes*, Second printing of the second edition, Undergraduate Texts in Mathematics. Springer-Verlag New York, New York-Heidelberg, 1975. x+325 pp.

[19] Ciarlet, Phillippe,, *Introduction to Numerical Linear Algebra and Optimisation* Cambridge Texts in Applied Mathematics, Vol. 2, Cambridge University Press, 1989, 452 pp.

[20] Cohen, Paul J., *Set Theory and the Continuum Hypothesis*, W. A. Benjamin, Inc., New York-Amsterdam 1966 vi+154 pp.

[21] Conway, John B., *Functions of One Complex Variable* Second edition. Graduate Texts in Mathematics, 11, Springer-Verlag, New York-Berlin, 1995. xiii+317 pp.

[22] Cormen, Thomas H.; Leiserson, Charles E.; Rivest, Ronald L., *Introduction to Algorithms*, The MIT Electrical Engineering and Computer Science Series, MIT Press, Cambridge, MA; McGraw-Hill Book Co., New York, 1990. xx+1028 pp.

[23] Coxeter, H. S. M., *Introduction to Geometry*, Second edition, Reprint of the 1969 edition, Wiley Classics Library. John Wiley and Sons, Inc., New York, 1989. xxii+469 pp.

[24] Davis, Harry, *Fourier Series and Orthogonal Functions*, Dover, 1989, 403 pp.

[25] Davis, Philip J., *The Schwarz function and its applications*, The Carus Mathematical Monographs, No. 17, The Mathematical Association of America, Buffalo, N. Y., 1974. 241 pp.

[26] De Souza, P. and Silva, J., *Berkeley Problems in Mathematics*, Springer-Verlag, New York, 1998, 457 pp.

[27] do Carmo, Manfredo P., *Differential Forms and Applications*, Universitext, Springer-Verlag, Berlin, 1994. x+118 pp.

[28] do Carmo, Manfredo P., *Riemannian Geometry*, Translated from the second Portuguese edition by Francis Flaherty, Mathematics: Theory & Applications, Birkhauser Boston, Inc., Boston, MA, 1994. xiv+300 pp.

[29] do Carmo, Manfredo P., *Differential Geometry of Curves and Surfaces*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1976. viii+503 pp.

[30] Dugundji, James, *Topology*, Allyn and Bacon, Inc., Boston, Mass. 1966 xvi+447 pp.

[31] Edwards, H., *Galois Theory*, Graduate Texts in Mathematics, 101, Springer, 1984.

[32] Euclid, *The Thirteen Books of Euclid's Elements*, translated from the text of Heiberg. Vol. I: Introduction and Books I, II., Vol. II: Books III–IX, Vol. III: Books X–XIII and Appendix, Translated with introduction and commentary by Thomas L. Heath, Second edition, Dover Publications, Inc., New York, 1956. xi+432 pp.; i+436 pp.; i+546 pp.

[33] Evans, Lawrence C., *Partial Differential Equations*, Graduate Studies in Mathematics, 19, American Mathematical Society, Providence, RI, 1998. xviii+662 pp.

[34] Feller, William, *An Introduction to Probability Theory and its Applications*, Vol. I, Third edition, John Wiley and Sons, Inc., New York-London-Sydney 1968 xviii+509 pp.

[35] Feynmann, R., Leighton, R. and Sands, M., *Feynmann's Lectures in Physics*, Vol. I, II and III, Addison-Wesley Pub Co, 1988.

[36] Finney, R. and Thomas, G., *Calculuc and Analytic Geometry*, Ninth edition, Addison-Wesley Pub Co., 1996.

[37] Fleming, Wendell, *Functions of Several Variables*, Second edition, Undergraduate Texts in Mathematics. Springer-Verlag, New York-Heidelberg, 1987. xi+411 pp.

[38] Folland, Gerald B., *Fourier Analysis and Its Applications*, The Wadsworth and Brooks/Cole Mathematics Series, Wadsworth and Brooks/Cole Advanced Books and Software, Pacific Grove, CA, 1992. 444 pp.

[39] Folland, Gerald B., *Introduction to Partial Differential Equations*, Second edition, Princeton University Press, Princeton, NJ, 1995. 352 pp.

[40] Folland, Gerald B., *Real analysis: Modern Techniques and their Application*, Second edition. Pure and Applied Mathematics. A Wiley-Interscience Publication, John Wiley and Sons, Inc., New York, 1999. xvi+386 pp.

[41] Fraleigh, John B., *A First Course in Abstract Algebra*, Sixth edition Addison-Wesley Pub Co. 1998, 576 pp.

[42] Fulton, W. and Harris, J., *Representation Theory: A First Course*, Graduate Texts in Mathematics, 129, Springer-Verlag, New York, 1991.

[43] Gallian, J., *Contemporary Abstract Algebra*, Fouth edition, Houghton Mifflin College, 1998. 583 pp.

[44] Gans, David, *An Introduction to Non-Euclidean Geometry*, Academic Press, New York-London, 1973. xii+274 pp.

[45] Garling, D. J. H., *A Course in Galois Theory*, Cambridge University Press, Cambridge-New York, 1987. 176 pp.

[46] Goldstern, Martin and Judah, Haim, *The Incompleteness Phenomenon*, A K Peters, Ltd., Natick, MA, 1998. 264 pp.

[47] Graham, Ronald L., Knuth, Donald E.and Patashnik, Oren, *Concrete mathematics: A Foundation for Computer Science*, Second edition, Addison-Wesley Publishing Company, Reading, MA, 1994. xiv+657 pp.

[48] Gray, Alfred, *Modern Differential Geometry of Curves and Surfaces with Mathematica*, Second edition, CRC Press, Boca Raton, FL, 1997. 1088 pp.

[49] Greene, Robert E.and Krantz, Steven G., *Function Theory of One Complex Variable*, Pure and Applied Mathematics. A Wiley-Interscience Publication, John Wiley and Sons, Inc., New York, 1997. xiv+496 pp.

[50] Grimmett, G. R.and Stirzaker, D. R., *Probability and Random Processes*, Second edition, The Clarendon Press, Oxford University Press, New York, 1992. 600 pp.

[51] Halliday, D., Resnick, R and Walker, J., *Fundamentals of Physics*, Fifth edition, John Wiley and Sons, 5th edition, 1996, 1142 pp/

[52] Halmos, Paul R., *Finite-Dimensional Vector Spaces*, Reprinting of the 1958 second edition, Undergraduate Texts in Mathematics, Springer-Verlag, New York-Heidelberg, 1993. viii+200 pp.

[53] Halmos, Paul R. *Naive Set Theory*, Reprinting of the 1960 edition, Undergraduate Texts in Mathematics, Springer-Verlag, New York-Heidelberg, 1974. vii+104 pp.

[54] Halmos, Paul R., *Measure Theory*, Graduate Texts in Mathematics, 18, Springer-Verlag, New York, 1976, 305 pp.

[55] Hartshorne, Robin, *Geometry: Euclid and Beyond*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 2000. xii+526 pp.

[56] David Henderson, *Differential Geometry: A Geometric Introduction*, Prentice Hall, 1998. 250 pp.

[57] Herstein, I., *Topics in Algebra*, Second edition, John Wiley & Sons, 1975.

[58] Hilbert, D. and Cohn-Vossen, S., *Geometry and the Imagination*, AMS Chelsea, 1999. 357 pp.

[59] Hill, Victor E., IV, *Groups and Characters*, Chapman and Hall/CRC, Boca Raton, FL, 1999.256 pp.

[60] Hintikka, Jaakko, *The Principles of Mathematics Revisited*, With an appendix by Gabriel Sandu, Cambridge University Press, Cambridge, 1998. 302 pp.

[61] Hofstadter, Douglas R., *Gödel, Escher, Bach: An Eternal Golden Braid*, Basic Books, Inc., Publishers, New York, 1979. 777 pp.

[62] Howard, Paul and Rubin, Jean, *Consequences of the Axiom of Choice*, Mathematical Surveys and Monographs, 59, American Mathematical Society, Providence, RI, 1998. viii+432 pp.

[63] Hubbard, Barbara Burke, *The World According to Wavelets: The Story of a Mathematical Technique in the Making*, Second edition, A K Peters, Ltd., Wellesley, MA, 1998. 286 pp.

[64] Hubbard, J. and Hubbard, B., *Vector Calculus, Linear Algebra, and Differential Forms: A Unified Approach*, Prentice Hall, 1999. 687 pp.

[65] Hungerford, T., *Algebra*, Eighth edition, Graduate Texts in Mathematics, 73, Springer, 1997. 502 pp.

[66] Iserles, Arieh, *A First Course in the Numerical Analysis of Differential Equations*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, 1996. 396 pp.

[67] Jackson, Dunham, *Fourier Series and Orthogonal Polynomials*, Carus Monograph Series, no. 6, Mathematical Association of America, Oberlin, Ohio, 1941. xii+234 pp.

[68] Jacobson, N., Basic Algebra, Vol. I and II, Second edition, W.H. Freeman, 1985.

[69] John, Fritz, *Partial Differential Equations*, Reprint of the fourth edition. Applied Mathematical Sciences, 1, Springer-Verlag, New York, 1991. x+249 pp.

[70] Jones, Frank, *Lebesgue Integration on Euclidean Space*, Revised edition, Jones and Bartlett Publishers, Boston, MA, 2001. 608 pp.

[71] Kac, Mark, *Statistical Independence in Probability, Analysis and Number Theory*, The Carus Mathematical Monographs, No. 12, Mathematical Association of America, New York 1969 xiv+93 pp.

[72] Kelley, John L., *General Topology*, Graduate Texts in Mathematics, 27. Springer-Verlag, New York-Berlin, 1975. xiv+298 pp.

[73] Kline, Morris, *Mathematics and the Search for Knowledge*, Oxford University Press, New York, 1972. 1256 pp.

[74] Kobayashi, Shoshichi and Nomizu, Katsumi, *Foundations of Differential Geometry*, Vol. I, Wiley Classics Library. A Wiley-Interscience Publication, John Wiley and Sons, Inc., New York, 1996. xii+329.

[75] Kobayashi, Shoshichi and Nomizu, Katsumi, *Foundations of Differential Geometry*, Vol. II, Wiley Classics Library, A Wiley-Interscience Publication, John Wiley and Sons, Inc., New York, 1996. xvi+468 pp

[76] Kolmogorov, A. N.and Fomin, S. V., *Introductory Real Analysis*, Translated from the second Russian edition and edited by Richard A. Silverman, Dover Publications, Inc., New York, 1975. xii+403 pp.

[77] Krantz, Steven G., *Function Theory of Several Complex Variables*, Second edition, AMS Chelsea, 2001. 564 pp.

[78] Krantz, Steven G., *Complex Analysis: The Geometric Viewpoint*, Carus Mathematical Monographs, 23, Mathematical Association of America, Washington, DC, 1990. 210 pp.

[79] Lang, Serge, *Algebra*, Third edition, Addison-Wesley, 1993 , 904 pp.

[80] Lang, Serge, *Undergraduate Analysis*, Second edition, Undergraduate Texts in Mathematics, Springer-Verlag, 1997, 642 pp.

[81] Lang, Serge and Murrow, Gene, *Geometry*, Second edition, Springer-Verlag, 2000, 394 pp.

[82] Mac Lane, Saunders, *Mathematics, Form and Function*, Springer-Verlag, New York-Berlin, 1986. xi+476 pp.

[83] Marsden, Jerrold E. and Hoffman, Michael J., *Basic Complex Analysis*, Third edition, W. H. Freeman and Company, New York, 1999. 600 pp.

[84] McCleary, John, *Geometry from a Differentiable Viewpoint*, Cambridge University Press, Cambridge, 1995. 320 pp.

[85] Millman, Richard and Parker, George D., *Elements of Differential Geometry* , Prentice-Hall Inc., Englewood Cliffs, N. J., 1977. xiv+265 pp.

[86] Morgan, Frank, *Riemannian Geometry: A Beginner's Guide*, Second edition, A K Peters, Ltd., Wellesley, MA, 1998. 160 pp.

[87] Moschovakis, Yiannis N., *Notes on Set Theory*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1994. xiv+272 pp.

[88] Munkres, James R., *Topology: A First Course*, Second edition, Prentice-Hall, Inc., Englewood Cliffs, N.J., 2000. 537 pp.

[89] Nagel, Ernest and Newman, James R., *Gödel's Proof*, New York University Press, New York 1960 ix+118 pp.

[90] Olver, P., *Applications of Lie Groups to Diferential Equations*, Second edition, Graduate Texts in Mathematics, 107, Springer-Verlag, New York, 1993.

[91] O'Neill, Barrett, *Elementary Differential Geometry*, Second edition, Academic Press, New York-London 1997. 448 pp.

[92] Palka, Bruce P., *An Introduction to Complex Function Theory*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1991. xviii+559 pp.

[93] Polya, George, Tarjan, Robert E.and Woods, Donald R., *Notes on Introductory Combinatorics*, Progress in Computer Science, 4, Birkhauser Boston, Inc., Boston, Mass., 1990. v+192 pp

[94] Protter, Murray H. and Morrey, Charles B., Jr. *A First Course in Real Analysis*, Second edition. Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1991. xviii+534 pp.

[95] Royden, H. L., *Real Analysis*, Third edition, Prentice-Hall, 1988. 434 pp.

[96] Rudin, Walter. *Principles of Mathematical Analysis*, Third edition, International Series in Pure and Applied Mathematics, McGraw-Hill Book Co., New York-Auckland-Dsseldorf, 1976. x+342 pp.

[97] Rudin, Walter, *Real and complex analysis*, Third edition, McGraw-Hill Book Co., New York, 1986. xiv+416 pp.

[98] Seeley, Robert T., *An Introduction to Fourier Series and Integrals*, W. A. Benjamin, Inc., New York-Amsterdam 1966 x+104 pp.

[99] Simmons, George, *Differential Equations With Applications and Historical Notes*, McGraw-Hill Higher Education, 1991, 640 pp.

[100] Smullyan, Raymond M., *Gödel's Incompleteness Theorems*, Oxford Logic Guides, 19, The Clarendon Press, Oxford University Press, New York, 1992. xvi+139 pp.

[101] Spiegel, M., *Schaum's Outline of Complex Variables*, McGraw-Hill, 1983.

[102] Spivak, M., *Calculus*, Third edition, Publish or Perish, 1994. 670 pp.

[103] Spivak, Michael, *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus*, Westview Press, 1971. 160 pp.

[104] Spivak, Michael, *A Comprehensive Introduction to Differential Geometry*, Vol. I - V. Third edition, Publish or Perish, Inc., 1979.

[105] Stanley, Richard P., *Enumerative Combinatorics*, Vol. 1, With a foreword by Gian-Carlo Rota, Cambridge Studies in Advanced Mathematics, 49. Cambridge University Press, Cambridge, 1997. 337 pp.

[106] Sternberg, S., *Group Theory and Physics*, Cambridge University Press, Cambridge, 1995. 443 pp.

[107] Stewart, Ian, *Galois theory*, Second edition, Chapman and Hall, Ltd., London, 1990. xxx+202 pp.

[108] Stewart, J., *Calculus*, Brooks/Cole Pub Co, third edition, 1995, 1015 pp.

[109] Strang, G., *Linear Algebra and its Applications*, Third edition, Harcourt College, 1988. 505 pp.

[110] Strikwerda, John C.. *Finite Difference Schemes and Partial Differential Equations*, The Wadsworth and Brooks/Cole Mathematics Series, Wadsworth and Brooks/Cole Advanced Books and Software, Pacific Grove, CA, 1989. xii+386 pp.

[111] Thorpe, John A., *Elementary Topics in Differential Geometry*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1994. xiv+267 pp.

[112] Trefethen, Lloyd and Bau, David, III, *Numerical Linear Algebra*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997. xii+361 pp.

[113] van der Waerden, B. L., *Algebra*, Vol 1, Based in part on lectures by E. Artin and E. Noether, Translated from the seventh German edition by Fred Blum and John R. Schulenberger, Springer-Verlag, New York, 1991. xiv+265 pp.

[114] van der Waerden, B. L., *Algebra*, Vol 2, Based in part on lectures by E. Artin and E. Noether, Translated from the fifth German edition by John R. Schulenberger, Springer-Verlag, New York, 1991. xii+284 pp.

[115] van Lint, J. H. and Wilson, R. M., *A Course in Combinatorics*, Second edition, Cambridge University Press, Cambridge, 2001. 550 pp.

[116] Zygmund, A., *Trigonometric Series*, Vol. I, II, Reprinting of the 1968 version of the second edition with Volumes I and II bound together, Cambridge University Press, Cambridge-New York-Melbourne, 1988. 768 pp.

Index

$C^k[0,1]$, 6
$O(n)$, 114
$\mathbf{L}^2$ Spaces, 248
GL(n,R), 214
$\mathbf{R}^n$, 2
    Standard Topology, 66
        Heine-Borel, 69
$\mathbf{S}_n$, 214

Abelian Group, 216
Absolute Convergence, 39
Adams, C., xvii
Adjacency Matrix, 309
Ahlfors, L., 197
Aho, A., 324
Algebraic Closure, 224
Algorithms, 308
    Big O Notation, 308
    Input, 308
    NP, 317
    Output, 308
    P=NP?, 316
    Parallel, 324
    Polynomial Time, 316
    Sorting, 314
Amplitudes, 245
Analytic Functions
    and Harmonic Functions, 270
    as Limit, 171, 172
    Cauchy Integral Formula, 185
    Cauchy's Theorem, 182
    Cauchy-Riemann Equations, 172, 175
    Complex Derivative, 173
    Conformal Maps, 172, 191
        Definition, 191
        Riemann Mapping Theorem, 194
    Hartog's Theorem, 196
    Integral Representations, 172, 179
    Nonholomorphic Function, 173
    Power Series, 172, 187, 188
Angle between Curves, 191
Arc Length
    Parametrization, 146, 148
Arnold, T., xvii
Artin, E., 21, 228
Artin, M., 228
Atkinson, K., 324
Axiom of Choice, 201, 206
    Algebraically Closed, 224
    Hausdorff Maximal Principle, 212
    Non-measurable Sets, 208
    Statement, 207
    Zorn's Lemma, 212

Barnard, Y., xvii
Barrett, D., xvii
Barschdorff, M., xvii
Bartle, R., 43
Basis for Topology, 74
    First Countable, 74
    Neighborhood Basis, 74
    Second Countable, 74
Basis of a Vector Space, 6
    Schauder Basis, 249
Bau, D., 324
Berberian, S., 43
Berenstein, C., 198
Bernoulli Trials, 294
Big O Notation, 308
Binomial Coefficients, 287
Birkoff, G., 228
Bishop, E., 206
Bocher, M., 21
Bolyai, J., 164
Boundary, 86, 136
    Induced Orientation, 136
Bounded Sets, 69
Boyce, W., 282
Bressoud, D., 43
Brouwer, L., 206
Brualdi, R., 305
Budar, N., xvii
Burden, R., 324

Burger, E., xvii

Cameron, P., 305
Canary, D., xvii
Cantor Diagonalization, 204
Cantor Set, 212, 234
  Measure Zero, 235
  Uncountable, 235
Cardinality, 201
Cauchy Integral Formula, 185
Cauchy's Theorem, 182
Cauchy-Riemann Equations, 172, 175
Cederberg, J., 162, 169
Central Limit Theorem, 294
Chain Rule, 52
Characteristic Functions, 237
Characteristic Polynomial, 16
  Similar Matrices, 17
Cheslack-Postava, T., xvii
Chung, K., 305
Ciarlet, P., 324
Circle
  as Manifold, 126, 128, 130
  Curvature, 145
    Calculation, 148
Circuit
  Euler, 310
  Hamiltonian, 312
  of a Graph, 310
Closed Sets, 64
Cohen, P., 206, 208, 211
Cohn-Vossen, S., 168
Combinatorial Formulas
  $\binom{n}{k}$, 287
  Binomial Coefficients, 287
  Examples, 286
Compact Sets, 64
Complexity Theory, *see* Algorithms
Conditional Probability, 290
Conformal, 172, 191
  Definition, 191
  Riemann Mapping Theorem, 194
Conjugate, 173
Connected Spaces, 64
Continuous Functions
  Single-Variable
    Definition, 25
    Example, 25
  Vector-Valued
    Definition, 49
  Via open sets, 64
Convolution, 255
Conway, J., 198
Cooper, B., xvii
Cormen, T., 308, 324
Correll, B., xvii
Cosets, 216
Countability, 201
  Cantor Diagonalization, 204
  Cardinality n, 201
  Countably Infinite, 201
    Measure Zero, 233
  Integers, 202
  Polynomials, 211
  Products of Sets, 203
  Rational Numbers, 203
  Uncountability of Real Numbers, 204
  Uncountably Infinite, 201
    Cantor Set, 235
Cowles, L., xvii
Coxeter, H.S.M., 169
Crystals and Symmetries, 213
Curl, 94
  as Exterior Derivative, 124
Curvature
  Circle, 145
    Calculation, 148
  Gauss-Bonnet, 157
  Geometry, 167
    Angles of Triangles, 168
    Gaussian Curvature, 168
  Plane, 145
  Plane Curves
    Definition, 147
    Varying Tangents, 147
  Space Curves
    Binormal Vector, 150
    Helix, 151
    Principal Curvature, 149
    Principal Normal Vector, 149
    Torsion, 150
  Sphere, 145
  Straight Line, 145
  Surfaces, 152
    Gauss Map, 153
    Gaussian Curvature, 154
    Hessian, 153
    Mean Curvature, 154
    Principal Curvatures, 154
Curve, 87
  Angle between Curves, 191
  Arc Length Parametrization, 146, 148
  Plane Curves and Curvature, 147
Cyclic Group, 217

Cylinder
    Curvature, 156

d'Alembert Formula and the Wave Equation, 279
Davis, H., 258, 279
Davis, P., 198
De Souza, P., xvi
Dennison, K., xvii
Determinant, 9
    Multilinear Function Definition, 10
    Product of Eigenvalues, 18
    Recursive Definition, 9
    Signed Volume Definition, 11
    Volume of Parallelepiped, 113
Differentiable Functions
    Single-Variable
        Definition, 26
        Example, 28
    Vector-Valued
        Chain Rule, 52
        Computations, 51
        Definition, 50
        Implicit Function Theorem, 58
        Inverse Function Theorem, 53
Differential Equations, 261
    Fourier Transform Methods, 256
    Heat Equation, 256
        Definition, 270
        Derivation, 271
        Steady-State Solutions, 267
    Integrability Conditions, 279
    Laplacian
        Dirichlet Problem, 266
        Harmonic, 266
        Mean Value Principle, 266
    Lewy's Example, 281
    Linear, 262
        Homogeneous, 262
    Nonlinear, 262
    Ordinary
        Definition, 261
        Existence and Uniquenss of Solutions, 264
        Lipschitz, 264
        Picard Iteration, 265
    Partial
        Definition, 261
    Separation of Variables, 267
    Symmetries, Groups and Lie Theory, 213
    Wave Equation
        Change of Variables, 277
        d'Alembert Formula, 279
        Definition, 274
        Derivation, 274
Differential Forms, 111, 115
    k-forms, 118
        Manipulating, 119
        Multilinear Maps, 118
        Vector Space of, 118
        Wedge Product, 121
        Wedging, 119
    0-forms and functions, 124
    1-forms and Vector Fields, 124
    2-forms and Vector Fields, 124
    3-forms and functions, 124
    and Vector Fields, 124
    Definition, 122
    Elementary k-forms, 115, 118
        and Parallelepiped, 118
    Elementary 1-forms, 117
    Elementary 2-forms, 115
    Exterior Derivative
        0-forms, 122
        Definition, 123
    Integrating k-forms, 138
Dimension
    Vector Space, 6
Diprima, R., 282
Dirichlet Problem, 266
Divergence, 93
    as Exterior Derivative, 124
Divergence Theorem, 82, 96
    Physical Interpretation, 97
    Proof, 99
do Carmo, M., 158
Doggart, C., xvii
Double Elliptic Geometry, 166
    Model, 166
Dual Space, 20
    and k-forms, 125
    Dual Map, 20
Dugundji, J., 78
Dunne, E., xvii

Edges of a Graph, 309
Edwards, H., 223, 228
Eigenvalue, 16
Eigenvector, 16
Elliptic Geometries, 166
    Axiom, 163
    Double, 166
        Model, 166
    Single, 166
        Model, 167

Empty Set, 206
Equivalence Classes, 328
Equivalence Problems, xix
    Physics, xxii
Equivalence Relations, 327
Euclid, 161, 164
Euclidean Geometry, 161, 162
    Fifth Postulate, 162
    Line, 162
    Playfair's Axiom, 163
    Surface, 162
Euler Circuit, 310
Evans, L., 282
Expected Value, 291
    Linearity, 291
Exterior Derivative, 111, 115
    and Curl, 124
    and Divergence, 124
    and the Gradient, 124
    Definition, 123
        0-forms, 122

Faires, J., 324
Feller, W., 305
Feynmann, R., 108
Fields, 223
    Algebraically Closed, 224
    Field Extensions, 223
        Fixed Fields, 225
        Group of Automorphisms, 225
        Normal, 226
        Splitting Fields, 225
Fifth Postulate, 162
Finney, R., 108
Fleming, W., 60, 142
Folland, G., 241, 258, 266, 282
Fomin, S., xvii
Fourier Analysis
    Fourier Integral, 253
    Fourier Series
        Amplitudes, 245
        Convergence in the Mean, 251
        Definition, 245, 246
        Fourier Coefficients, 245
        Gibb's Phenomenon, 251
        Uniform Convergence, 252
    Fourier Transform
        Solving Differential Equations, 256
        Convolution, 255
        Definition, 253
        Properties, 254
    Hilbert Spaces, *see* Hilbert Spaces
Fraleigh, J., 228
French, C., xvii
Functions
    Analytic, *see* Analytic Functions
    as 0-forms, 124
    as 3-forms, 124
    Characteristic, 237
    Continuity via open sets, 64
    Gradient, 93
    Harmonic, 266
    Holomorphic, *see* Analytic Functions
    Integral Representations of, 179
    Lipschitz, 264
    Measurable, 238
    on Manifolds, 131
    Periodic, 244
    Random Variables, 289
    Single-Variable
        Continuous, 25
        Differentiable, 26
        Limit existing, 23
        Pointwise Convergence, 35
        Uniform Convergence, 36
        Weierstrass Example, 41
    Step, 238
    Vector-Valued
        Chain Rule, 52
        Continuous, 49
        Definition, 47
        Differentiable, 50
        Example, 48
        Implicit Function Theorem, 58
        Inverse Function Theorem, 53
        Limit existing, 49
Fundamental Theorem of Calculus, 31, 81, 95, 224
    Statement, 32

Gödel's Theorem, 210
Gödel, K., 208, 210, 307
Gallian, J., 228
Galois Theory, 223
    Field Extensions, 223
    Fixed Fields, 225
    Fundamental Theorem, 226
    Group of Automorphisms, 225
    Normal, 226
    Quadratic Equation, 224
    Splitting Fields, 225
Gans, D., 169
Garling, D., 224, 228
Gauss Map, 153, 159

    Hessian, 153
Gauss, C., 164
Gauss-Bell Curve, 285, 295
Gauss-Bonnet Theorem, 157
Gaussian Curvature, 154
    Geometries, 168
Gay, R., 198
Genus, 157
Geometry
    Curvature, 167
        Angles of Triangle, 168
        Gaussian, 168
    Elliptic, 166
        Axiom, 163
        Double, 166
        Model for Double, 166
        Model for Single, 167
        Single, 166
    Euclidean, 161, 162
        Fifth Postulate, 162
        Line, 162
        Playfair's Axiom, 163
        Surface, 162
    Hyperbolic, 163
        Axiom, 164
        Unit Disc, 169
        Upper Half Plane, 164
Gibb's Phenomenon, 251
Goldstern, M., 211
Gradient, 93
    as Exterior Derivative, 124
    as Normal Vector, 153
Graham, R., 305
Graph
    Adjacency Matrix, 309
    Circuit, 310
    Connected, 325
    Definition, 309
    Edges, 309
    Euler Circuit, 310
    Examples, 309
    Hamiltonian Circuit, 312
    Königsberg Bridge Problem, 310
    Path, 310
    Trees
        Binary, 314
        Definition, 313
        Leaves, 313
        Root, 313
        Sorting, 314
    Vertices, 309
Gray, A., 158
Green's Theorem, 82, 179, 182
Greene, R., 197
Grimmett, G., 305
Group, 213
    Abelian, 216
    Cosets, 216
    Cyclic, 217
    Definition, 213
    Homomorphism, 216
    Identity, 214
    Inverse, 214
    Kernel, 218
    Order, 218
    Orthogonal, 114
    Permutation, 120, 214
        Flips, 120
        Notation, 215
        Shuffles, 121, 143
        Sign of, 120
        Transpositions, 120
    Quotient Group, 217
    Representation Theory, *see* Representation Theory
    Rotations, 214
    Solvable, 227
    Subgroup, 216
        Normal, 217
    Sylow's Theorem, 218

Halliday, D., 108
Halmos, P., 21, 211, 241
Hamiltonian Circuit, 312
Harmonic Analysis, *see* Fourier Analysis
Harmonic Functions, 266
    and Analytic Functions, 270
Hartog's Theorem, 196
Hartshorne, R., 169
Hausdorff, 64, 127
Hausdorff Maximal Principle, 212
Haynes, R., xvii
Heat Equation, *see* Differential Equations
Heine-Borel Theorem, 69
Helix, Curvature of, 151
Henderson, D., 158
Herstein, I., 219, 228
Hessian, 153
    Eigenvalues as Principal Curvatures, 154
    Gaussian Curvature, 154
    Mean Curvature, 154
Hilbert Spaces
    Complete, 248

  Definition, 248
  Inner Product, 247
  Orthogonal Vectors, 248
  Schauder Basis, 249
  Square-integrable Functions, 248
Hilbert, D., 168, 206
Hill, V., 228
Hinman, P., xvii
Hintikka, J., 211
Hoffman, M., 197
Hofstadter, R., 211
Holomorphic, *see* Analytic Functions
Homotopy, 184
  Simply Connected, 194
Hopcroft, J., 324
Howard, P., 208
Hubbard, B., 60, 142, 258
Hubbard, J., 60, 142
Hungerford, T., 228
Hyperbolic Geometry, 163
  Axiom, 164
  Unit Disc, 169
  Upper Half Plane, 164

Ideal in Ring, 222
Implicit Function Theorem, 56
Independence in Probability Theory, 290
Independence Results, 210
  Completeness, 210
  Consistency, 210
Inner Measure, 233
Integers
  as Rings, 221
  Size of, 201, 202
Integrability Conditions, 279
Integral Representations
  Analytic Functions, 179
Integrals
  and Differential Forms, 138
  Area, 29
  Path or line, 88
    Infinitesimal Arc Length, 89
  Riemann, 28
    Definition, 31
    Lower Sums, 31
    Upper Sums, 31
  Surface, 91
Intermediate Value Theorem, 321
Inverse Function Theorem, 53
Iserles, A., 324

Jackson, D., 258
Jacobian
  Chain Rule, 52
  Computation, 51
  Definition, 50
  Implicit Function Theorem, 58
  Inverse Function Theorem, 53
  Parametrized Manifolds, 84, 127
Jacobson, N., 228
James, D., xvii
John, F., 282
Jones, F., 241
Judah, H., 211

Königsberg Bridge Problem, 310
Köronya, A., xvii
Karni, S., xvii
Kastermans, B., xvii
Kelley, J., 78
Kennedy, C., xvii
Kernel, 5
  Group, 218
Klein, F., 164, 166
Kline, M., 161
Klodginski, E., xvii
Knapp, F. G., xvii
Knuth, D., 305
Kobayashi, S., 158
Krantz, S., 197, 198
Kravitz, S., xvii
Kronecker, L, 206

Lang, S., 43, 169, 228
Laplacian, *see* Differential Equations
Lebesgue Dominating Convergence Theorem, 231, 239
Lebesgue Integration, *see* Measure Theory
Lebesgue Measure, *see* Measure Theory
Leiserson, C., 308, 324
Length of Complex Number, 174
Lenhart, W., xvii
Lewy's Example, 281
Lewy, H., 282
Limit
  Single-Variable
    Definition, 23
    Example, 24
  Vector-Valued
    Definition, 49
Line Integral, 88
  Infinitesimal Arc Length, 89
Linear Transformations, 1
  as Matrices, 7
  Definition, 5

    - Eigenvalue, 16
    - Eigenvector, 16
    - Image, 5
    - Kernel, 5
- Lipschitz Functions, 264
- Lobatchevsky, N., 164
- Loop
    - Closed, 180
        - Counterclockwise, 180
        - Homotopic, 184
        - Simple, 180
- Lower Sums, 31

- Möbius Strip, 94
- Mac Lane, S., xvi, 228
- Manifolds, 111, 126
    - Abstract, 129
        - Tangent Spaces, 133
    - Boundary, 136
        - Induced Orientation, 136
    - Circle, 126, 128, 130
    - Curve, 87
    - Differentiable Functions, 131
    - Equivalence, 132
    - Implicit, 128
        - Gradients, 128
        - Normal Space, 133
        - Tangent Space, 133
    - Orientability, 94, 135
        - Induced on Boundary, 136
    - Parametrized, 84, 128
        - Boundary, 86
        - Jacobian, 84, 127
        - Tangent Space, 133
    - Projective Space, 143
    - Simple, 99
    - Sphere, 127, 143
    - Surface, 91
        - Surface Integral, 91
    - Torus, 143
    - Transition Functions, 129
- Marsden, J., 197
- Matrices, 1
    - **GL(n,R)**, 214
    - Adjacency, 309
    - as Linear Transformations, 7
    - Characteristic Polynomial, 16
    - Determinant
        - Product of Eigenvalues, 18
    - Eigenvalue, 16
    - Eigenvector, 16
    - Multiplication by Complex Numbers, 177
    - Similar, 14
    - Symmetric, 19
    - Transpose, 13
- McLeary, J., 158, 169
- Mean Curvature, 154
- Mean Value Principle for Laplacian, 266
- Mean Value Theorem, 322
- Measurable Functions, 238
- Measure Theory
    - Inner Measure, 233
    - Lebesgue Integral
        - Definition, 238
        - Lebesgue Dominating Convergence Theorem, 239
    - Measurable Functions, 238
    - Measurable Sets, 232
    - Non-measurable sets, 208
    - Outer Measure, 232
- Metric Spaces, 73
    - Complete, 248
- Miekisz, J., xvii
- Millman, R., 158
- Mizner, R., xvii
- Moore, E., 77
- Moore, R., 77
- Morgan, F., xvii, 158
- Morrey, C., 43
- Moschovakis, Y., 211
- Multilinear
    - k-forms, 118
        - Wedge Product, 121
    - Determinant, 10
- Munkres, J., 78
- Murrow, G., 169

- Nagel, E., 211
- Newman, J., 211
- Newton's Method, 317
- Nirenberg, L., 282
- Noether, E., 21, 228
- Nomizu, K., 158
- Non-measurable Sets, 208
- Normal Distribution, 295
- NP Algorithms, 317
- Numerical Analysis
    - Newton's Method, 317

- O'Neil, B., 158
- Olver, P., 263
- Open Sets, 64
- Ordinary Differential Equations, *see* Differential Equations
- Orientability, 94, 136

    as an Equivalence Relation, 135
    Induced on Boundary, 136
    Möbius Strip, 94
    Vector Spaces, 135
Orthogonal Group, 114
Outer Measure, 232

P=NP?, 316
Palka, B., 197
Parallelepiped, 112
    and k-forms, 118
    Volume, 113
Parametrizing Map for Manifolds, 127
Parker, G., 158
Partial Differential Equations, *see* Differential Equations
Partially Ordered Set, 212
Pascal's Triangle, 305
Patashnik, O., 305
Path Connected Spaces, 65
Path Integral, 88, 179
    Infinitesimal Arc Length, 89
Pedersen, L., xvii
Periodic Functions, 244
Picard Iteration, 265
Pigeonhole Principle, 306
Playfair's Axiom, 163
Poincaré, H., 164
Pointwise Convergence of Functions
    Definition, 35
    Example, 35
Polar Coordinates, 298
Polya, G., 305
Polynomial
    Ring Structure, 221
    Roots
        Approximating, 318
        Galois Theory, 213
        Multiplicity, 18
Polynomial Time, 316
Poset, 212
Power Series, *see* Analytic Functions
Probability Theory, 287
    Bernoulli Trials, 294
    Central Limit Theorem
        Statement, 294
    Conditional Probability, 290
    Definition
        Probability, 288
    Expected Value, 291
        Linearity, 291
    Gauss-Bell Curve, 285, 295
    Independence, 290
    Normal Distribution, 295
    Probability
        Definition, 287
    Random Variables, 289
    Sample Space, 287, 288
    Standard Deviation, 293
    Variance, 292
Projective Space
    as Manifold, 143
Protter, M., 43

Quadratic Equation, 224
Quotient Group, 217

Random Variables, 289
Rational Numbers
    Measure Zero, 233
    Size of, 203
Real Numbers
    Size of, 201
    Uncountability, 204
Reflexivity, 327
Representation Theory
    Direct Sum Representation, 220
    Irreducible Representation, 221
    Representation
        Definition, 219
Resnick, R., 108
Riemann Integral, 28
    Definition, 31
Riemann Mapping Theorem, 194
Ring, 213
    Definition, 221
    Example
        Integers, 221
        Polynomials, 221
    Ideal, 222
    Quotient Ring, 222
    Zariski Topology, 75
        $Spec(R)$, 75
Rivest, R., 308, 324
Root, S., xvii
Royden, H., 239, 241
Rubin, J., 208
Rudin, W., 43, 241
Russell, B., 207
Russell-Zermelo Paradox, 207

Sacceri, G., 164
Sample Space, 287, 288
Scalars, 2, 5
Schauder Basis, 249
Schippers, E., xvii

Seeley, R., 258
Series of Functions
    Convergence, 38
Shuffles, 143
    Permutation Group, 121
Silva, C., xvii
Silva, J., xvi
Similar Matrices, 14
    Characteristic Polynomial, 17
Simmon, G., 282
Single Elliptic Geometry, 166
    Model, 167
Smullyan, R., 211
Solvable Group, 227
Sorting Algorithms, 314
Space Curves, *see* Curvature
Spectral Theorem, 323
Sphere
    as Manifold, 127, 143
    Curvature, 145
        Calculation, 155
Spiegel, M., 197
Spivak, M., 40, 41, 43, 60, 142, 158
Standard Deviation, 293
Stanley, R., 305
Step Functions, 238
Sternberg, S., 221, 228
Stewart, I., 228
Stewart, J., 108
Stirling's Formula, 300
Stirzaker, D., 305
Stokes' Theorem, 82, 111
    Classical, 96
    General Form, 139
    Physical Interpretation, 98
    Proof, 104
    Special Cases
        Classical Stokes' Theorem, 96
        Divergence Theorem, 82, 96
        Fundamental Theorem of Calculus, 95
        Green's Theorem, 82, 179, 182
Strang, G., 21
Strikwerda, J., 324
Successor Set, 206
Surface, 91
    Curvature, 152
        Gauss Map, 153
        Gauss-Bonnet, 157
        Gaussian Curvature, 154
        Hessian, 153
        Mean Curvature, 154
        Principal Curvatures, 154
    Genus, 157
    Normal Vectors, 153
Surface Integral, 91
Sylow's Theorem, 218
Symmetric Matrices, 19
Symmetry, *see* Group, 327

Tangent Spaces, 133
Tarjan, R., 305
Thomas, G., 108
Thorpe, J., 158
Topology
    Basis, 74
        First Countable, 74
        Neighborhood Basis, 74
        Second Countable, 74, 127
    Bounded Sets, 69
    Closed Sets, 64
    Compact Sets, 64
    Connected, 64
    Continuous Functions, 64
    Definition, 63
    Gauss Bonnet, 157
    Hausdorff, 64, 127
    Induced Topology, 64
    Metric Spaces, 73
        Complete, 248
    Open Sets, 64
    Path Connected, 65
    Standard Topology on $\mathbf{R}^n$, 66
        Heine-Borel, 69
    Zariski, 75
Torsion
    of Space Curves, 150
Torus
    as Manifold, 143
Transitivity, 327
Transpose, 13
Trees, *see* Graphs
Trefethon, L., 324
Type Theory, 207

Ullman, J., 324
Uniform Convergence of Functions
    Continuity, 37
    Definition, 36
    Fourier Series, 252
    Series of Functions, 38
Upper Half Plane, 164

Van der Waerden, B., 21, 228
van Lint, J., 305
Variance, 292

Vector Fields
    and Differential Forms, 124
    as 1-forms, 124
    as 2-forms, 124
    Continuous, 82
    Curl, 94
    Definition, 82
    Differentiable, 82
    Divergence, 93
Vector Space, 1
    $\mathbf{R}^n$, 2
    Basis, 6
    Definition, 4
    Dimension, 6
    Dual Space, 20
        Dual Map, 20
    Hilbert Space, *see* Hilbert Spaces
    Image, 5
    Inner Product, 247
        Orthogonality, 248
    Kernel, 5
    Key Theorem, 12
    Linear Independence, 7
    Span, 7
    Subspace, 5
Vector-Valued Functions, *see* Functions
Vertices of a Graph, 309
Volume of Parallelepiped, 113
Wave Equation, *see* Differential Equations
Waves
    Examples, 243
Wedge Product, 121
Weierstrass M-test, 38
    Application
        Weierstrass Example, 42
    Example, 39
    Statement, 39
Weierstrass, K., 23, 40
Westerland, C., xvii
Whitehead, A., 207
Wilson, R., 305
Woods, D., 305

Zariski Topology, 75
    $Spec(R)$, 75
Zermelo, E., 207
Zermelo-Fraenkel Axioms, 206
Zermelo-Russell Paradox, 207
Zorn's Lemma, 212
Zygmund, A., 258

(almost) everything you wanted to know about...

Soldering

Contributed by **Thijs Beckers, Luc Lemmens,** and **Jan Visser** (Elektor Labs)

Solder Station
ST - 50A
POWER
0
150
200
250
300
350
400
450
°C
TOOLCRAFT
ESD
50W
TOOLCRAFT

Q *What's a good soldering iron for electronics?*

A From 8-watt pencil to 150-watt "poker", it depends on your goal. If you occasionally have a bit of soldering to do and do not expect to embark on large jobs, it's sufficient to use a cheap mains operated soldering iron with a power from 25 to about 40 watts. At these lower power levels, the solder tip is usually small enough for electronics; at power levels above 40 watts roughly, the solder tip is usually unwieldy.
But if you have more extensive soldering to do and especially if you are planning on doing SMD work, a temperature-controlled 'iron' rated at about 100 watts is recommended. Such stations usually come with a diversity of solder tips, which together with the temperature control feature make the station suitable for large and small solder joints. Nowadays cheap soldering stations are available too but keep in mind that for no-name equipment the availability of spare tips can be a problem. Although cheap products aren't necessarily bad, if you are after a product you can enjoy using for years, you should not cut back on it.

Q *What is the best temperature for a soldering iron?*

A That depends on a number of things:

- lead-free solder tin (higher melting point) or leaded (lower melting point);
- tip thickness / heating capacity / heating element construction / solder station quality;
- soldering to a small pad or a large ground plane;
- personal soldering habits — for "slow" soldering (i.e. long tip-to-component contact time) set a relatively low temperature; for "fast" soldering, set a high temperature. Too hot soldering will cause flux evaporation or component overheating.

Q *How can I prevent the solder tin from appearing dull?*

A Both the temperature and the duration of the heating process must be correct. Make sure that the solder joint (i.e. the metal surfaces to be joined) is properly preheated to (almost) the melting point of the tin before adding it. If not there is a fair chance of the flux evaporating before it can do its job. BTW RoHS compliant solder tin always becomes dull.

Q *Which solder tin is best?*

A That is often a personal preference. At Elektor Labs we use tin/lead solder wire with resin core (flux) for our prototypes — the flux cleans the surfaces where the solder joint is to be made, and ensures that the tin flows beautifully. Note that lead/tin solder wire is still permitted for use in prototypes, but not in the production of equipment!
For larger work like through-hole (TH) parts we use solder wire with a diameter of 0.8 mm to 1 mm; for SMDs, 0.5 mm works much better.

Q *When do you use solder paste?*

A Mainly when using an SMD oven or hot-air (gun/pencil) soldering. But sometimes it's also useful to use a dispenser or pointed object to apply solder paste in advance to a spot that is not easily accessible.

Q *Why do you use flux?*

A Flux breaks the surface tension of molten tin and causes it to flow neatly. It also has a slight etching and cleaning effect on metal surfaces, which ensures proper adhesion of the tin.
There are many types of flux available with different chemical compositions. A specific type usually indicated as "no-clean" does not require after treatment, while others need to be removed after soldering because they are (slightly) conductive, or their etching effect continues and in time causes poor connections.

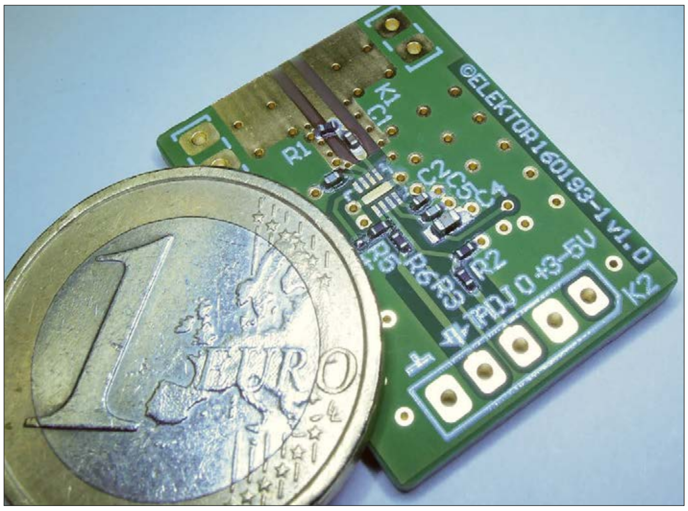

Q *How do I keep my soldering iron clean?*

A First and foremost, a clean solder tip is an absolute requirement for good soldering work. A contaminated or corroded tip results in poor heat transfer to the solder joint. Therefore, wipe the tip frequently (best after each connection) on a damp sponge or dry metal scouring pad.

- Do not heat up the soldering iron unnecessarily and switch it off immediately after use. A hot surface oxidises faster. Also, flux remnants carbonise, forming small black dots in the solder tin.
- On a damp (not: soaking wet) sponge you can wipe the solder tip clean. Do it with a quick swipe, otherwise the sponge burns and cools the tip down possibly to the extent of solidifying tin which precludes proper cleaning.
- A metal scouring pad. Contrary to what the name suggests, the intention is to lightly wipe the tip clean on the pad — real sanding would irreparably damage or even remove the metal coating from the tip.
- A product called 'tip activator' can aid in removing stubborn spots of oxidation and contaminations from the tip. The heated tip is dipped into the activator and lightly moved up and down, after which the contaminants on a metal abrasive sponge can be cleaned away. Trim the tip richly with fresh, fluxed solder tin, wipe clean again on the sponge and finally pre-tin lightly again. Instead of (fairly expensive) activator, flux can also be used in this process, but usually several cleaning cycles are required to achieve the desired result.

Note: officially, cleaning agents should not be necessary if the tip is regularly wiped clean, but it does no harm to properly inspect the tip before and after each soldering session. Damage to the coating is unavoidable however and will prompt tip replacement sooner or later.

Q *What and how to desolder?*

A Use a regular solder iron, possibly with a suction pump (hand operated or electric) and/or a desoldering iron (if necessary with extra flux) to remove as much solder tin as possible.

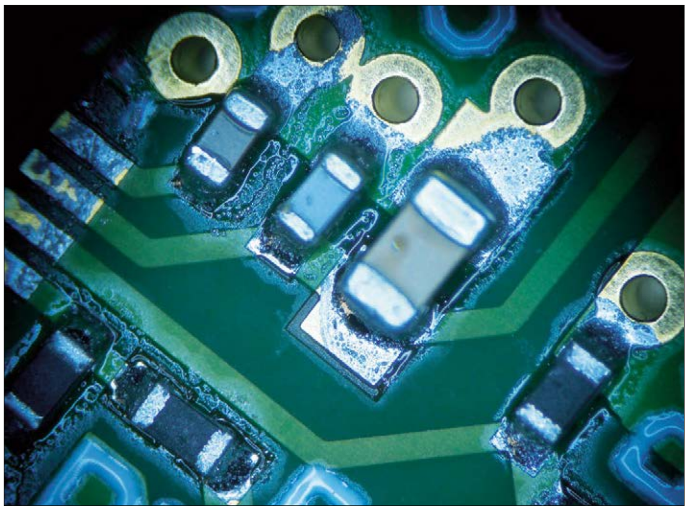

A hot-air pencil or gun is not only useful for soldering or removing SMD components, it can also be very useful for desoldering TH parts. It can be used to melt several solder joints simultaneously, for example three legs of a transistor are heated simultaneously to remove the component in one go. However, be careful with nearby plastic parts like connector housings and wiring, which may melt and burn before you can say Jack Robinson.
When it's not terribly important for the component to survive the desoldering operation, it is often handy to cut off the legs and then desolder them one by one.
After removing a component, always remove as much solder as possible with the suction pen or desoldering braid. The latter will leave resin residues on the board surface, which require cleaning with IPA or a special PCB cleaner before installing new components.

Q *What do I use to solder SMA parts a.k.a. SMDs?*

A Larger SMDs can still be handled with a small soldering iron, preferably with a thin tip and thin, resin-cored solder wire. But the smaller the components become, the more difficult the job, see the images in this article. For example, resistors and capacitors of size 0805 or smaller tend to stick to the solder tip instead of remaining on the PCB. Likewise, the smaller the pin spacing of ICs, the greater the risk of solder bridges forming short circuits. That too can be solved, but there will be a point at which working with solder paste in combination with a hot-air pencil or even an SMD oven is the only option. Moreover, there are components (especially ICs) whose connections can no longer be reached with a soldering iron. The purchase of a hot-air station should not be a dramatic financial loss — under 100 euros will buy you an acceptable station. An oven is not absolutely necessary for most parts, but it is a must-have for larger circuit boards or small-volume productions.

Q *How do I solder a component with a BGA housing?*

A A component with ball grid array (BGA) connections can only be soldered with hot air, meaning in an oven or using a hot-air station. The most important things are the correct amount of solder tin (i.e. the diameter of the balls on the component), the absolute cleanliness of the PCB pads (remove all tin with desoldering braid) and a sufficiently high temperature to warrant for all connections. Here now we assume a new delivered part with the right size of solder balls on the connections, a clean board and the use of a high-end hot-air station. Apply a reasonable amount of flux to the BGA terminals on the PCB. Position the IC as far to the very best of your ability, observing the marking(s) for pin 1. The temperature and airflow settings can be determined experimentally and will be different for each station and each nozzle. Take into account the temperature of any plastic headers or other 'sensitive' parts in the direct vicinity: these may be able to withstand 300 degrees C but should not be challenged. In case of doubt, apply protective Kapton tape (refer to the datasheet). When using airflow, avoid blowing other components off the board. The exact skill is to get sufficient heat energy conveyed with temperature and airflow combined. Heat the IC and the PCB. A circular movement is the best way to achieve this. Also heat an extra area of about 1-2 cm around the IC in order to heat the PCB sufficiently. Initially the flux will go in all directions, but a part will remain under the IC and will ensure a good flow of the tin. When all the tin has melted, the IC — if correctly positioned — will align itself in exactly the right place. Pay close attention during the heating: the alignment happens quite suddenly and at the same time it's a sign that all the tin has melted and that the IC is in place. When the hot air gun is now moved back and forth quickly, you should be able to see the IC 'swaying' a little bit on the molten grid balls. Then it's perfect. Limit this to 2-3 seconds and then remove the pencil/gun. Allow the IC to cool down gently. Afterwards, the flux can be cleaned with IPA (or in an ultrasonic bath). Presto! To be sure, measuring whether a short circuit has occurred between supply voltage and ground is not a bad thing.

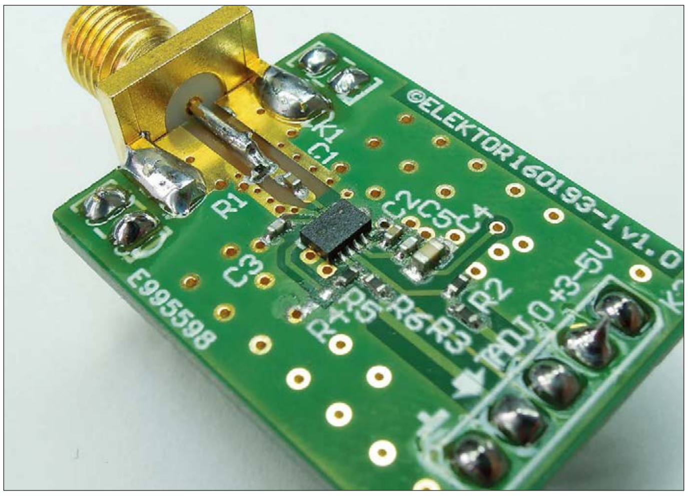

Q *Do I need a fume extractor?*

A Fumes from modern solder wire alone are not known to be toxic or hazardous in small amounts but they can cause irritation and trigger certain allergies. Burnt plastics though are very undesirable. These risks, and your housemates, should be enough reason to ventilate your e-workspace, or even better, buy a fume extractor. Now available from the Elektor Store is a fume extractor for table top use (SKU 18450 on www.elektor com). It extracts solder fumes either horizontally through a rectangular opening in the case, or through a flexible hose and mouthpiece assembly you can mount optionally.

(160697)

Sanding and filing

Note: the following does not refer to soldering irons used in electronics. Only if your solder tip is made of copper you may clean it with a file or a piece of sandpaper. Otherwise, **never**. The surface of a solder tip typically applied in electronics is covered with a layer of silver, which is removed or at least damaged by sanding or abrasive action. This will only make the soldering worse instead of easier. Also, any scratches you make with sandpaper quickly fill with small debris deteriorating the soldering process.

Bush Skills (/overland-tech-travel/?category=Bush+Skills), Equipment (/overland-tech-travel/?category=Equipment), Repair (/overland-tech-travel/?category=Repair)

Battery welding done right (/overland-tech-travel/2016/5/2/battery-welding-done-right)

Jonathan Hanson (/overland-tech-travel/?author=502d020ae4b0b0975770cfbe) · May 2, 2016 (/overland-tech-travel/2016/5/2/battery-welding-done-right)

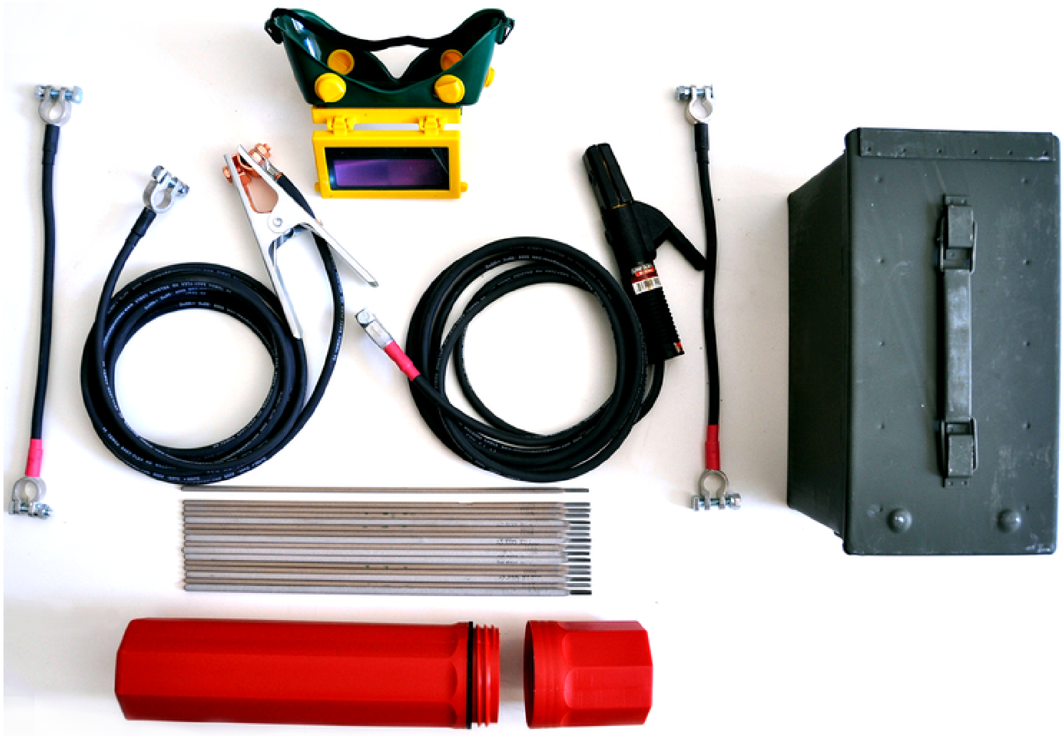

The magic of battery welding has been known to field mechanics for years now, and one of the most popular classes at the Overland Expo is the hands-on introduction to the skill, taught by experts such as welding wizard Tim Scully. Briefly, by combining several standard automotive

Search

Index of Past Posts ⌄

RETURN TO L
(HTTP://WWW.EXPLORING
TECH-TRA

SUBSCRIBE

Overland News is our monthly email report on all things overland—including Overland Expo and our trip and training announcements. **CLICK HERE** (http://www.overlandexpo.com/overlandnews/) **to subscribe** (opens new window, on the Overland Expo website).

Overland Tech & Travel RSS (http://www.exploringoverland.com/tech-travel/?format=rss)

May 17-19 '19
Flagstaff
AZ USA
Nov 9-11 '18
Asheville
NC USA

batteries in series—that is, positive terminal to negative terminal, in a chain—you create in effect one large 24V (with two batteries) or 36V (three batteries) cell, and this produces enough power to weld a lot of things that can be prone to breaking on vehicles used in rugged conditions: shock and spring mounts, motor mounts, ancillary brackets, roof racks—the list is endless, and full of items that can bring a trip to a sudden halt.

Until now, most impromptu battery welding has been jury rigged with standard jumper cables. While this works, it is far from ideal. Jumper-cable wire is cheap stuff designed for a few second's starting duty at a time; its coarse strands are inefficient at conducting the power produced by series-connected batteries. And the toothed clamps, although adequate for attaching to battery terminals, are poor for gripping slender welding rods. In addition, you need a way to connect the batteries to each other, which either requires another set (or two) of jumper cables, or yet more jury rigging with your existing battery leads.

All that just ended, thanks to the Trail Weld kit, developed by Tim Scully himself. Tim evaluated the compromises that go into the normal battery-welding setup, and fixed them all.

- The cables are now fine-strand, four-gauge Temco welding wire, with 360-degree crimps on all fittings. A 12-foot length on positive and negative leads allows you to put a safe distance between the batteries and sparks. The flexible cable makes controlling the clamp and rod much easier.
- The positive lead ends in a proper welding-rod clamp, and the negative lead ends in a dedicated ground clamp.
- Two short leads of Temco wire make quick work of connecting batteries, and reduce voltage loss.
- All battery connections are high-quality terminal clamps, greatly enhancing conductivity and thus efficiency.
- All connections are color-coded with heat-shrink wrap.
- A selection of correctly sized welding rods is included in a plastic protective case.
- A pair of self-powered auto-darkening goggles is also included.
- Available containers range from a simple bag to a .50-caliber ammo can to a Pelican Case.

The complete system is so efficient that Tim reports two batteries are sufficient to weld material that requires three if using jumper cables. As he told me, "If you are using three batteries you'd better be welding at least quarter-inch-thick stock." Since an increasing number of overland vehicles these days are equipped with two batteries, that means you can be completely self-sufficient for field-welding repairs.

For welding amateurs such as myself, the auto-darkening goggles make all the difference when welding with rod, as there's no pre-positioning the rod at the correct gap and then fumbling with a standard goggle or, worse, a jury-rigged square of welding glass taped to a cardboard face shield, as I've always carried. (Of course you'll still want face, arm, and hand protection.)

For more information, visit the Trail Weld site, here (http://www.trailweld.com). At the upcoming Overland Expo WEST, Trail Weld kits will be on display during the welding classes, and available for purchase at the 7P booth. Highly recommended.

❤ 4 Likes   Share

COMMENTS (3)   Newest First   Subscribe via e-mail

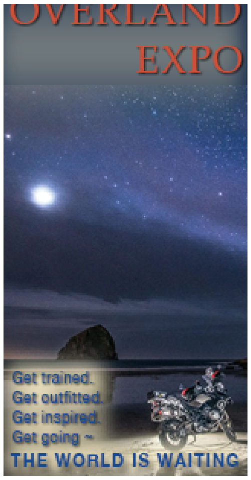
OVERLAND
EXPO
Get trained.
Get outfitted.
Get inspired.
Get going ~
THE WORLD IS WAITING
(http://www.overlandexpo.com)

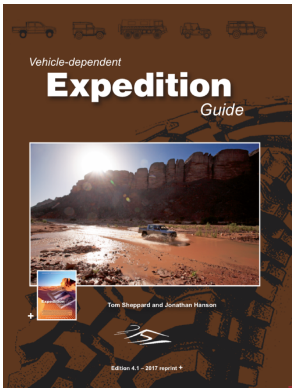

(http://www.exploringoverland.com,

BOOKS

OVERLAND
EXPO
(http://www.overlandexpo.com/gea
shop)

OVERLAND DECALS, HATS, T-SHIRTS & MORE . . .

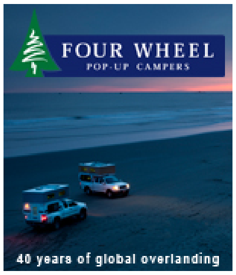

(http://www.fourwheelcampers.c

Trump Ends Another Obama Era Program

If you owe less than $679,650 on your home, use Congress's free mortgage relief program. The program itself is totally free and doesn't add any cost to your refi. You'll be shocked when you see how much you can save.

TAP YOUR AGE:

18-25
26-35
36-45
46-55
56-65
65+

RECALCULATE YOUR HOUSE PAYMENT

©2019 NMLS ID 167283; 3306 nmlsconsumeraccess.org

LowerMyBills.com

advertisement

Batteries + Jumpercables = DC Welder

By TimAnderson (/member/TimAnderson/) in Workshop (/workshop/) > Tools (/workshop/tools/) 256,045 403 101 Featured

All Rights Reserved Download Favorite

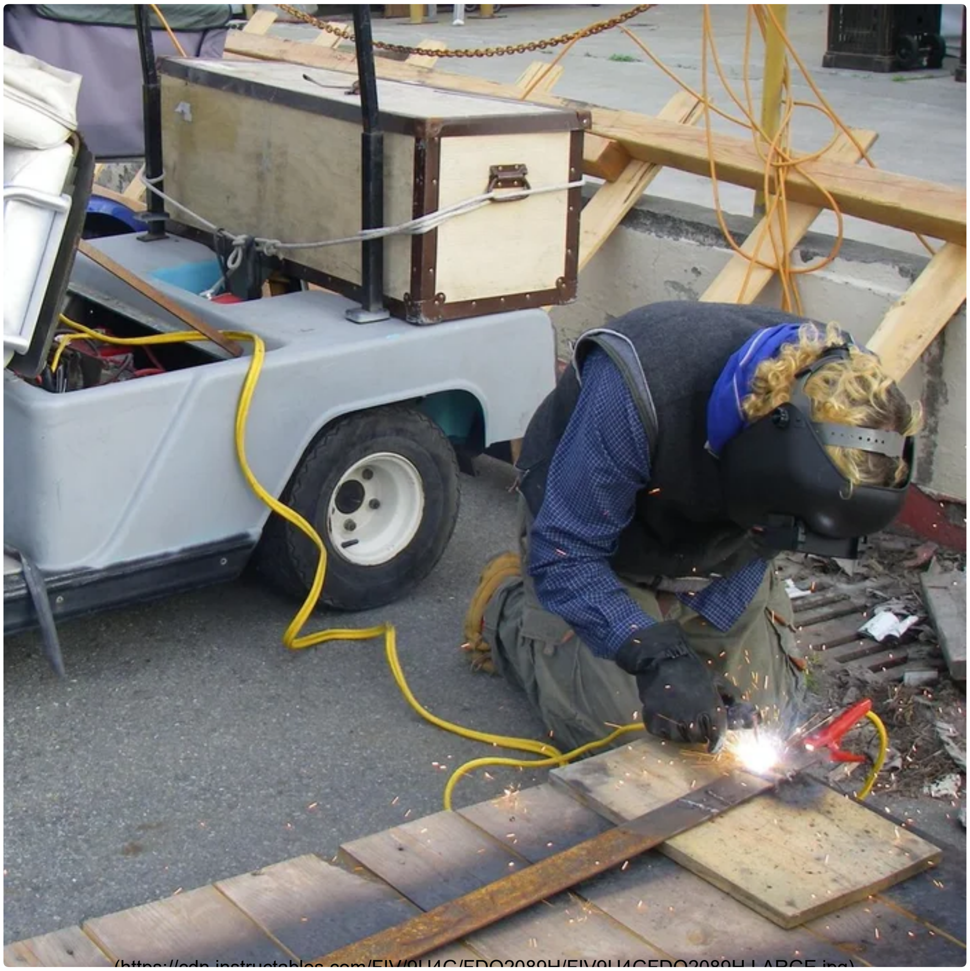

(https://cdn.instructables.com/FIV9U4C/FDO2089H/FIV9U4CFDO2089H.LARGE.jpg)

(/member/TimAnderson/)

By **TimAnderson (/member/TimAnderson/)**

Follow

More by the author:

About: Tim Anderson is the author of the "Heirloom Technology" column in Make Magazine. He is co-founder of www.zcorp.com, manufacturers of "3D Printer" output devices. His detailed drawings of traditional Pacific... More About

TimAnderson » (/member/TimAnderson/)

Here's the simplest welder you can make.
It's just a pair of jumper cables and a welding rod.
Oh yeah. And some batteries to power it.
I happened to have some nice juicy batteries in the form of a Golf Cart.

Here's me welding a trailer hitch with 1/8" 6011 rods. Welding current peaks at around 140-150 amps with these rods and around 120 amps with 3/32 6013 rods according to a Fluke 1010 clamp probe.

Greatly inspired by South Africans battery-welding their landrovers. (http://www.lcool.org/technical/80_series/bat_weld/battery_welding.html)
More battery welders (http://www.fieldlines.com/story/2006/5/8/6559/50626)
Also check out this great homemade AC stick welder (https://www.instructables.com/id/Build-a-Microwave-Transformer-Homemade-Welder/).

Oops! Almost forgot a warning. DON''T ELECTOCUTE, BURN, OR BLIND YOURSELF. DON'T BREATHE THE FUMES OR DETONATE YOUR BATTERIES.
That said, read the specs for your battery. Big ones are designed to produce enough current for welding. Don't let the nay-sayers keep you from trying it. Check your batteries and don't overheat them. Lead-acid batteries can go into a thermal-runaway self-discharge mode.

Action photos by Andrea (https://www.instructables.com/member/seedlingproject/), video by Pete Lynne.

Add Tip Ask Question Comment Download

Merrell
Reach Any Peak in These Hiking Boots
Learn more
advertisement

Step 1: Cables and Rod

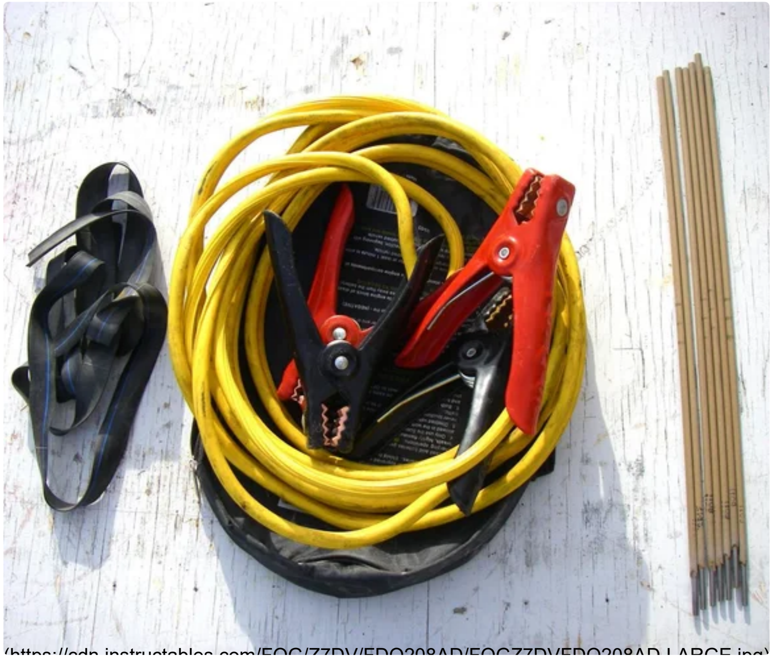

(https://cdn.instructables.com/FOG/Z7DV/FDO208AD/FOGZ7DVFDO208AD.LARGE.jpg)

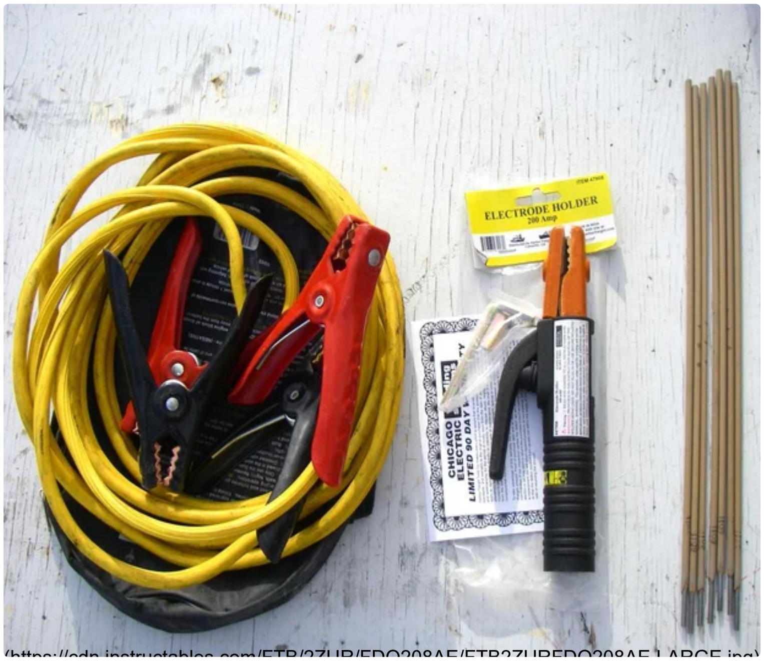

(https://cdn.instructables.com/FTB/2ZUB/FDO208AE/FTB2ZUBFDO208AE.LARGE.jpg)

Here's the stuff.
The jumper cables were $12 from Harbor Freight.
The strip of bicycle innertube is optional, to wrap the end of an alligator clip and make it grip harder.
If you're feeling fancy you can replace one of your alligator clips with an electrode holder.
This nice one was made in India and cost $7 from Harbor Freight.
The welding rods I tried are 3/32" 6013 and 1/8" 6011, 5lbs for $6 or so. Available in Africa, S.America, anywhere.

You'll also want to wear gloves,
and a welding helmet, $16 from Northern Tool (http://www.northerntool.com/webapp/wcs/stores/servlet/product_6970_200305009_200305009)

Add Tip Ask Question Comment Download

Step 2: Alligator Clip Orthodontics

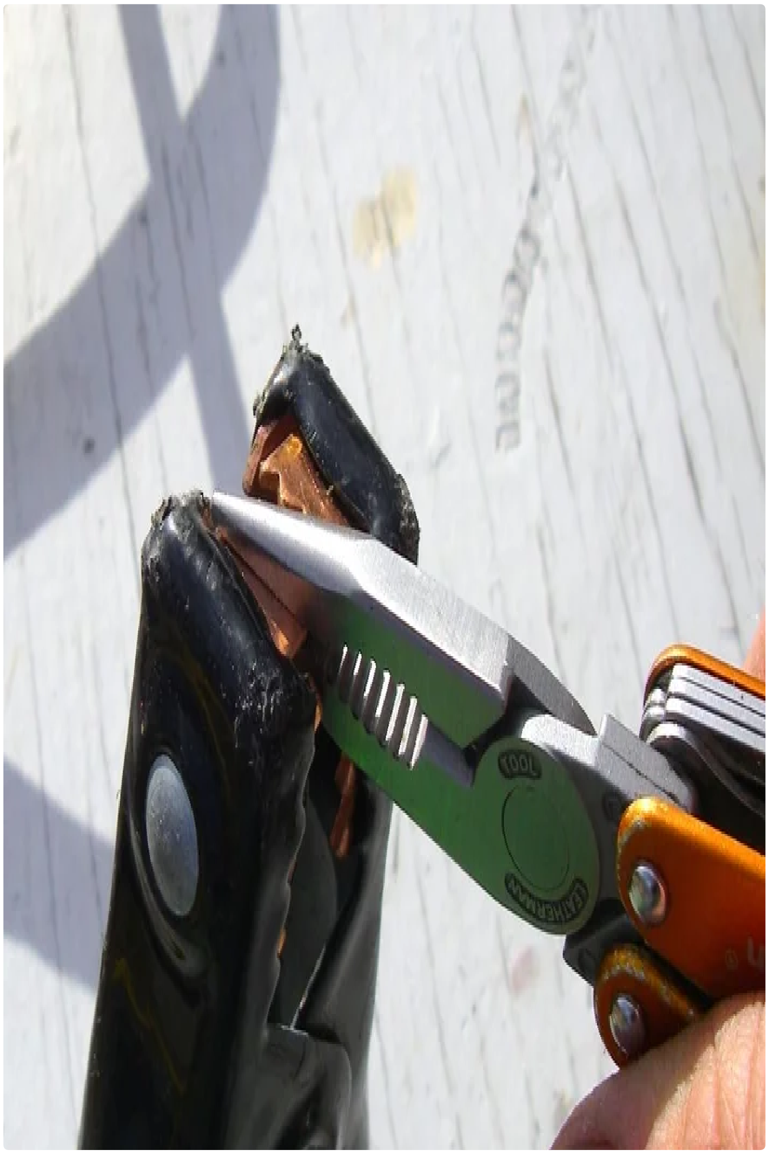

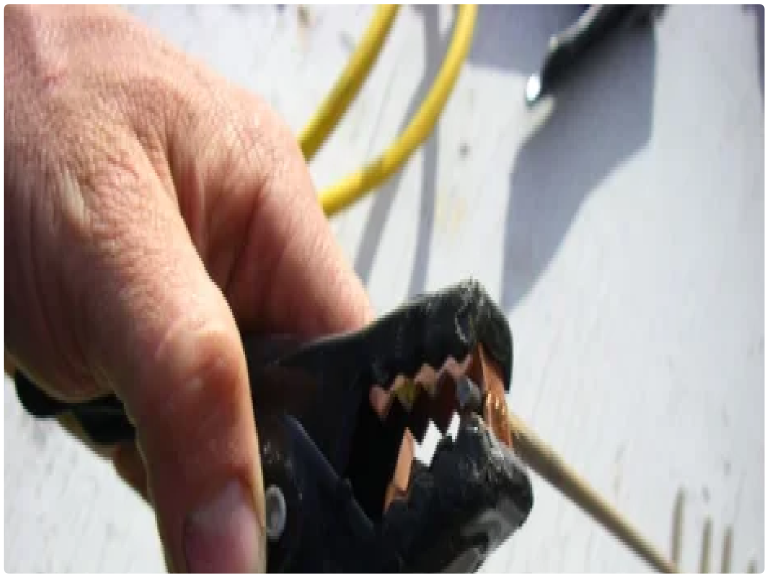

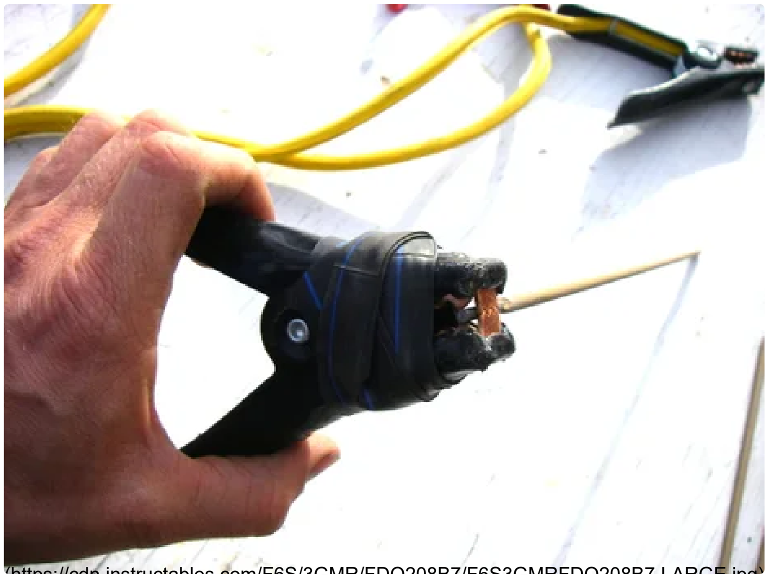

(https://cdn.instructables.com/F6S/3CMB/FDO208B7/F6S3CMBFDO208B7.LARGE.jpg)

Bend the alligator clip's incisors forward so the canines can grip the rod properly as seen in the 2nd photo.
Strengthen the alligator's jaw muscles with a wrapping of bicycle innertube strip.

Add Tip   Ask Question   Comment   Download

Step 3: Batteries!

(https://cdn.instructables.com/FZX/O6RC/FDO208BD/FZXO6RCFDO208BD.LARGE.jpg)

My golf cart has 6 of these 6 volt T-105 size batteries that I bought used from THE GUY WHO MADE THEM!!!

They're in series making 36 volts total. I can clip in anywhere in the chain to get 36,30,24, etc. volts.
The batteries have about 220 Amp-hours capacity, so that means I could weld at 150 amps for more than an hour (consult books printed on paper for the details) without recharging. Plenty good since I'll never have a project with enough metal laid out ahead of time to spend that much time a-welding in a day.
If your batteries aren't as nice as this, don't worry, you can do plenty of welding with three old car batteries in series.

I got these from the guy who made them?
That's right, a family run battery manufacturer right in Sausalito in a little workshop.
It's really amazing, like an old time blacksmith's shop, but high-tech. They can make any size battery.
Alrick, the proprietor, grew up making batteries and he really knows his stuff.
He sells batteries to all the big boats there, and he knows exactly how long every brand of battery lasts in every set of conditions.
Don't bother him about your electric car though, he's been plagued about that sort of thing enough already.

Arc battery company
90 gate 5 rd. sausalito ca 94965
415-332-3272

Add Tip Ask Question Comment Download

Step 4: Welding Tests

(https://cdn.instructables.com/FOB/LLBR/FDQ208CF/FOBLLBRFDQ208CF.LARGE.jpg)

I welded on 1/4" thick cold-rolled mild steel plate.
I used DC reverse-polarity at 36,30, and 24 volts.
I used 1/8" 6011 rod and 3/32" 6013 rod

Add Tip Ask Question Comment Download

Step 5: The Welds

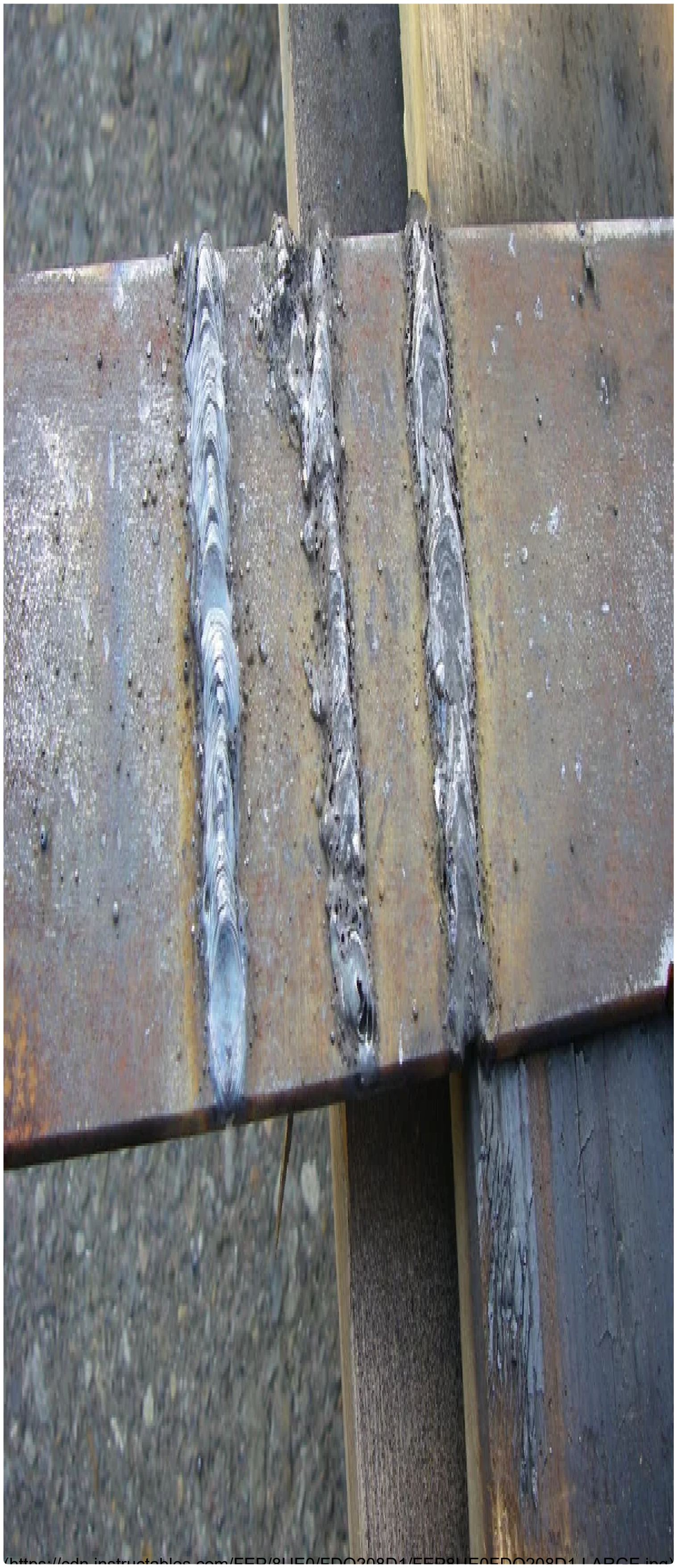

(https://cdn.instructables.com/FFB/8UE0/FDQ208D1/FFB8UE0FDQ208D1.LARGE.jpg)

(https://cdn.instructables.com/F7W/0W43/FDQ208D3/F7W0W43FDQ208D3.LARGE.jpg)

(https://cdn.instructables.com/F1W/RJMO/FDQ208D2/F1WRJMOFDQ208D2.LARGE.jpg)

A short arc but very hot. It melted deep and fast into the 1/4" plate.
I couldn't have stuck the rod to the plate if I wanted to.

These are the welds after chipping and brushing the slag off.

I did the right hand weld first with 1/8" 6011 rod at 36 Volts.
I've had the rods for a while and they're a bit damp and splattered a lot.
It's a deep good weld as you can see in the next photo of the backside of the plate, where it's melted through.

The mess at the top of the middle weld is trying a 6011 at 24 volts. Not enough juice, a lot of splattering. The full length bead next to it is a 6011 at 30 volts.

The left hand weld, the best looking one, is a 3/32" thick 6013 rod at 30 volts.

This golf cart is a good welder and I will definitely use it to weld projects in the future.

Add Tip Ask Question Comment Download

Step 6: Arc Stabilizer and Stinger

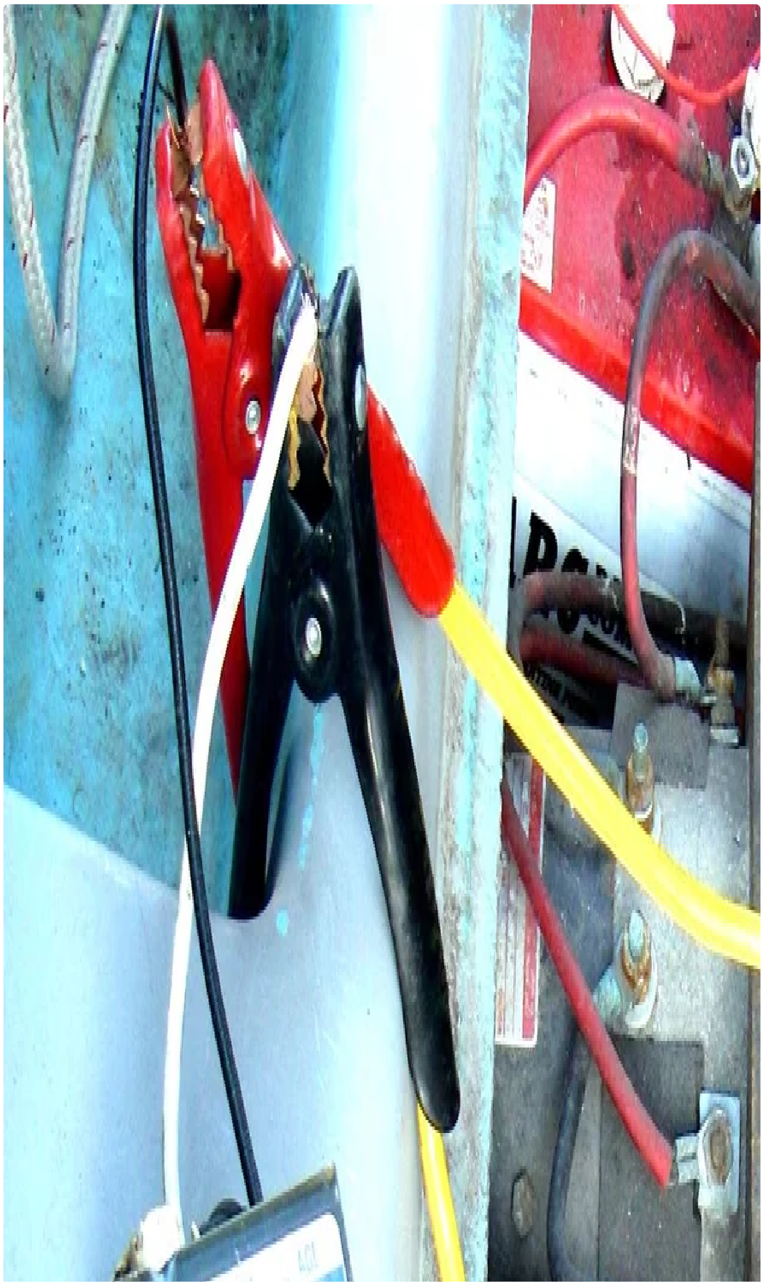

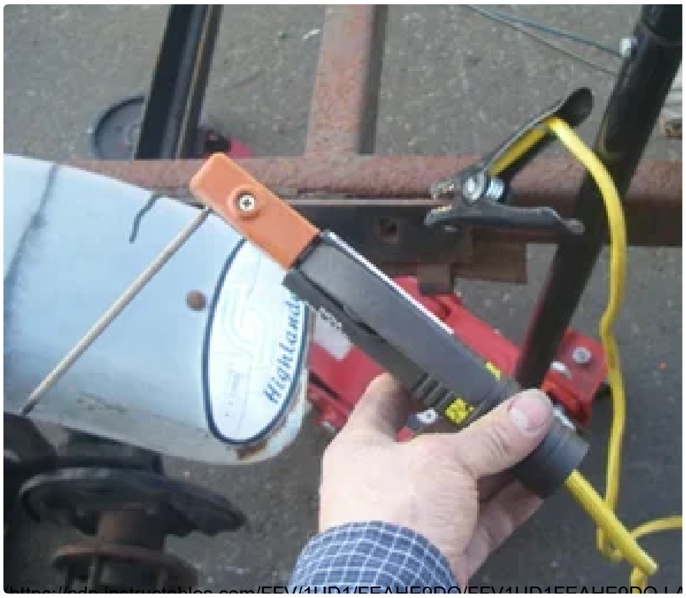

(https://cdn.instructables.com/FEV/1UD1/FEAHE9DQ/FEV1UD1FEAHE9DQ.LA

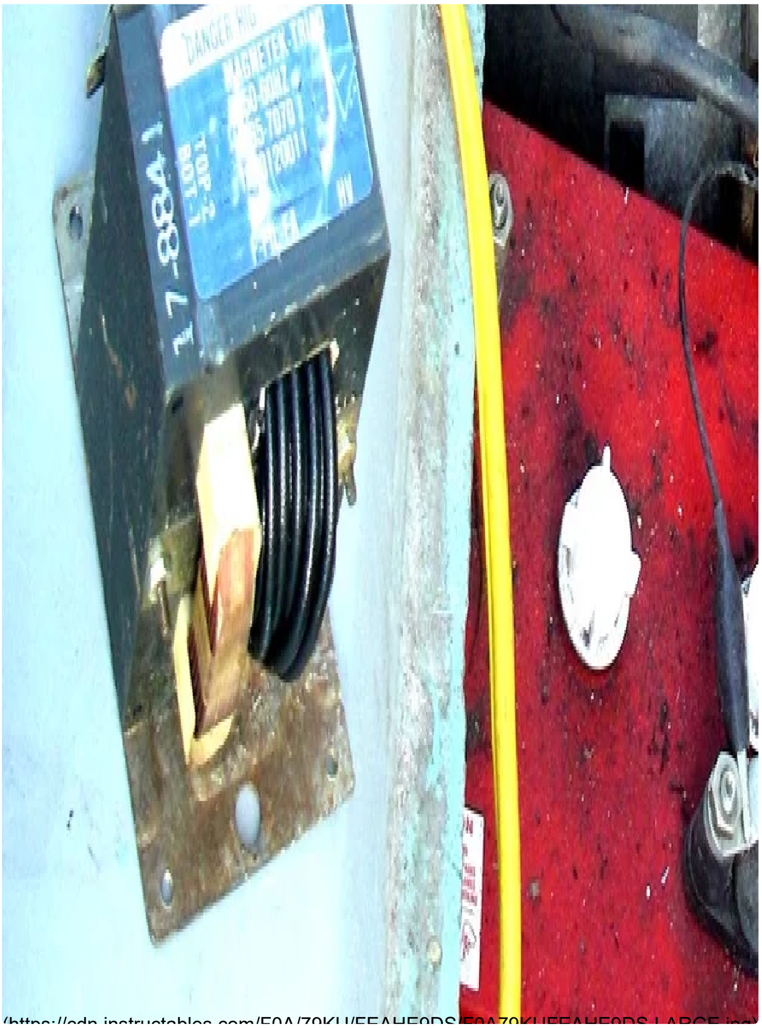

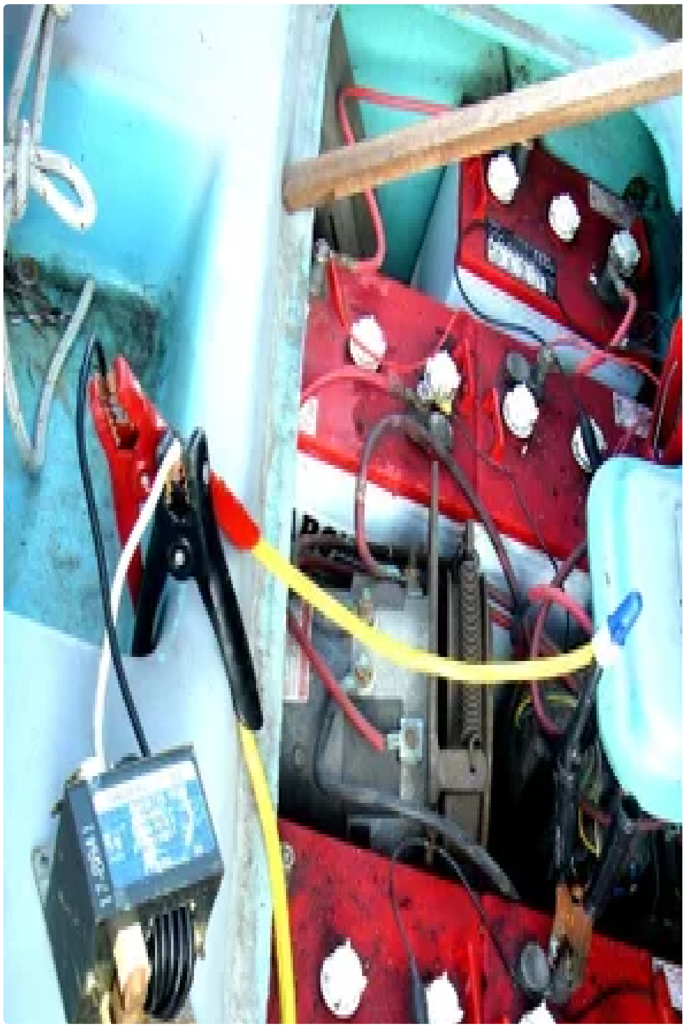

(https://cdn.instructables.com/F0A/79KU/FFAHE9DS/F0A79KUFFAHE9DS.LARGE.jpg)

Thanks to readers who recommended putting an inductor in series, called an "arc stabilizer".
Zan (https://www.instructables.com/member/ZanA/) and I made one from a transformer we scavenged from a dead microwave oven. We used this method (https://www.instructables.com/id/SSOXF48FAKC0MQT/) to chop out the secondary and wind 20 turns of 10-guage wire, which is all that would fit.
We left the primary intact, it doesn't affect the function here. An inductor gives the current "momentum" to blast through bits of flux and rust that get in the way. The arc sputters a lot less, welding is easier, and the welds are better.

I got a great Indian-made stinger (electrode holder) for $6 from harborfreight (http://www.harborfreight.com/cpi/ctaf/displayitem.taf?Itemnumber=47908) and replaced one of the jumper cable clamps with that.

Then I welded some repairs on this very badly rusted boat trailer. Too pitted to grind clean very well. I used 1/8" 6011 rod at 36 volts and it was fabulous. I melted good deep welds into all that rusty metal. Very satisfying.

We've got a self-propelled welder!

Add Tip Ask Question Comment Download

Step 7: Cutting Steel With a Stick Welder

(https://cdn.instructables.com/FO7/7QJT/FFEA6YBS/FO77QJTFFEA6YB

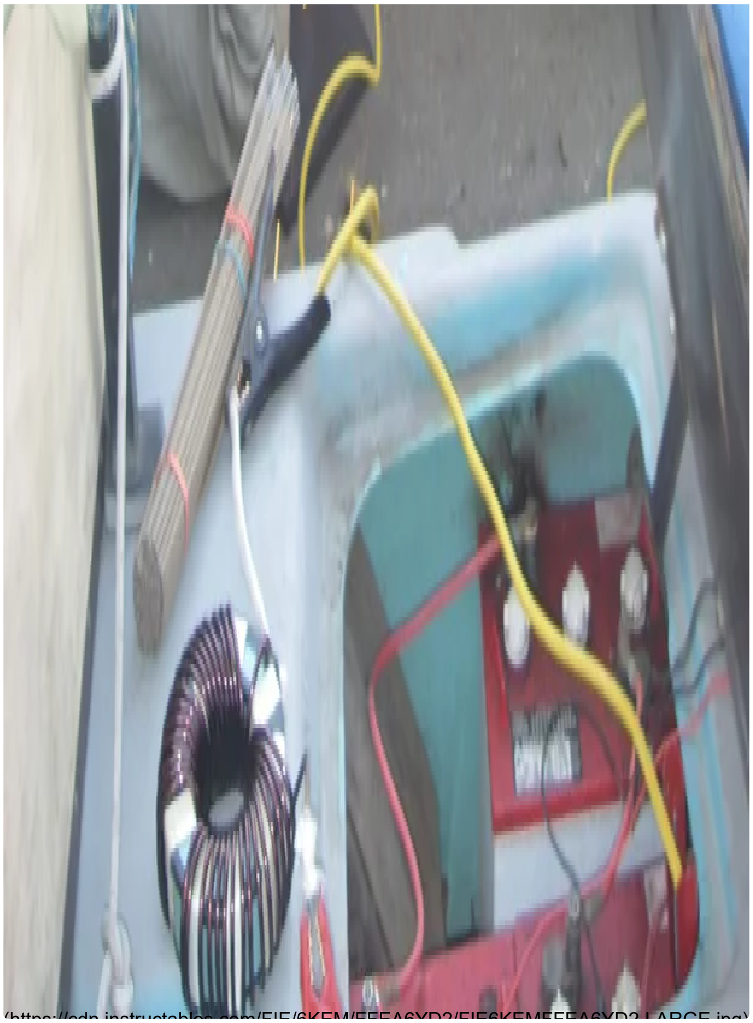

(https://cdn.instructables.com/FIF/6KFM/FEEA6YD2/FIF6KFMFEEA6YD2.LARGE.jpg)

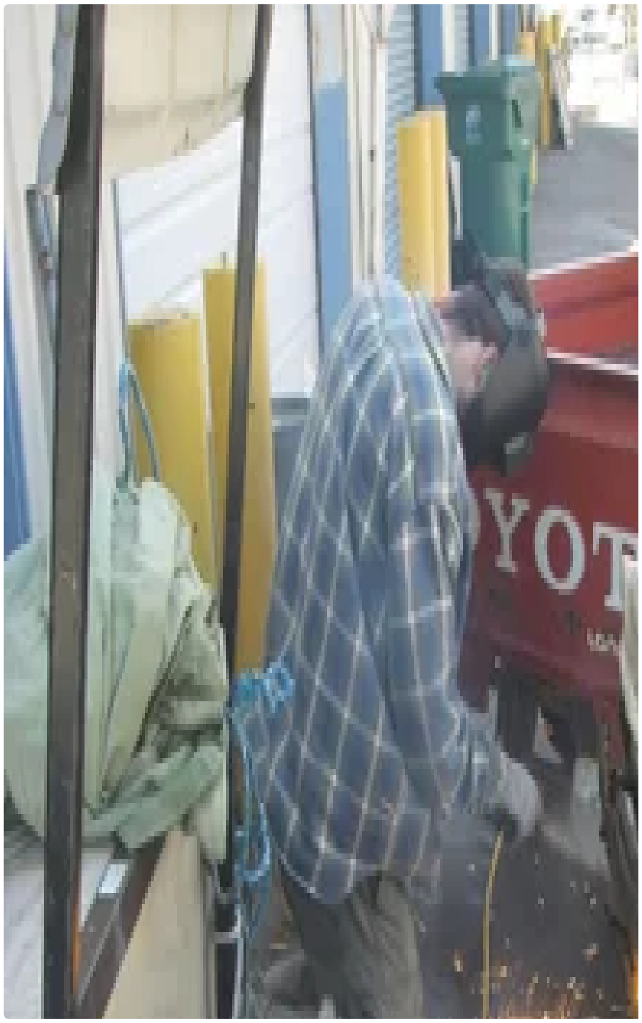

The bumper hitch on my 1978 Toyota truck was rusted rotten. Ray (https://www.instructables.com/member/Tangerinepiggy/) and I used the golfcart welder with 1/8" 6011 rods to weld some beefy reinforcements on it.

The big toroidal series inductor seen here came from [www.accrc.org/ ACCRC] Maker Day. I should put a few dozen more turns on it to really smooth out the arc. We did some strong agricultural-style welding that isn't going to break or appear in textbooks.

Here's Ray (https://www.instructables.com/member/Tangerinepiggy/) cutting holes for the trailer safety chain to hook on.
We dipped the rods in water and then blasted and pushed the steel out of the way. It works well and makes a surprisingly clean hole while using less rod than you'd think.
Thanks to a reader who suggested dipping the rods in water- I can't find your comment now, what other nifty tricks do you have up your sleeve?

I also tried an improvised air-arc torch using compressed air. It didn't work at all. The air blast blew the arc away, chilled and froze the molten metal instead of burning it. I guess that's why air-arc cutting torches use 500 amps or more

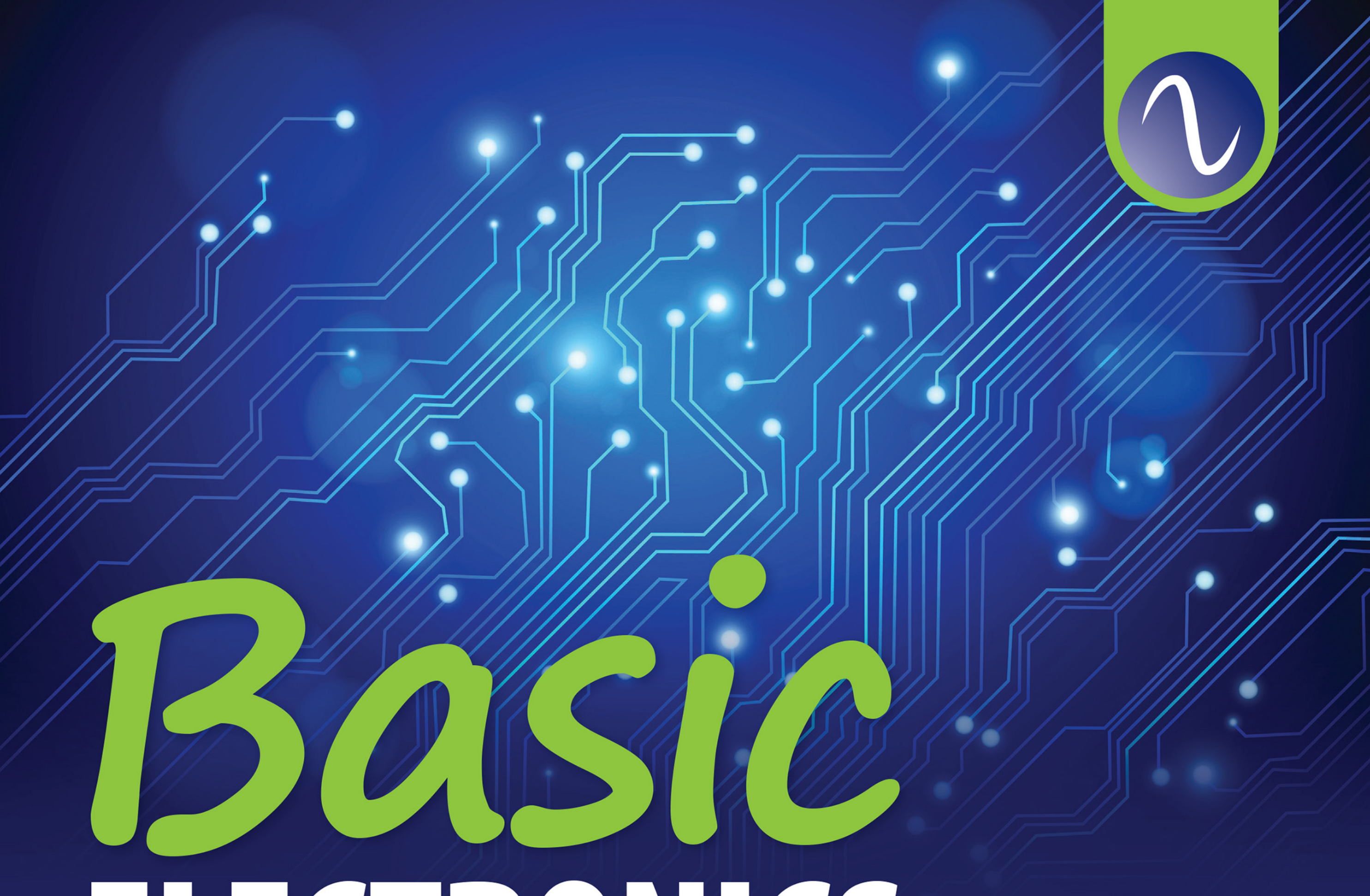

Basic ELECTRONICS TUTORIALS

For Beginners and Beyond

by **Wayne Storr**

FREE EDITION

BASIC ELECTRONICS TUTORIALS

By Wayne Storr of Electronics-Tutorials.ws

Terms of Use

The information contained within this Basic Electronics Tutorials guide is provided "as-is" and free of charge for general information purposes only on a range of Electronic related topics. All the information and material being published and presented within this particular e-book, including text, graphics and images is the copyright or similar such rights of Wayne Storr representing the website: *www.electronics-tutorials.ws*, unless otherwise expressly stated.

The purpose of this e-book is as a free study reference guide for the education of its readers who wish to learn Electronics. While every effort and reasonable care has been taken with respect to the accuracy of the information given herein, the author makes no representations or warranties of any kind, expressed or implied, about the completeness, accuracy, omission of errors, reliability, or suitability with respect to the information or related graphics contained within this e-book for any purpose. Any reliance you place on such information is therefore strictly at your own risk.

The information presented within this free e-book is intended for general information and personal use only and is not intended to address your particular problem or requirement. We can not and do not offer any specific technical advice, troubleshooting assistance or solutions to your individual circumstances. However you are encouraged to consult other technical resources or seek professional advice for any additional information or clarification you may need.

Caution

A word of caution. When working with Electrical or Electronics components and circuits, all Electrical Safety procedures must be adhered to and no responsibility is assumed for any injury and/or damage to persons or property as a result of using this guide. It is essential that careful supervision is given to children when working with mains powered electrical and electronic circuits and soldering irons. Thank you and take care.

For more information visit our website at: www.electronics-tutorials.ws

Copyright © 2013 Wayne Storr
www.electronics-tutorials.ws
All rights reserved

Basic Electronics Tutorials

CONTENTS

1.0 Electrical Theory 3
  1.1 The Structure of Atoms 3
  1.2 The Unit of Charge 3
  1.3 Electric Current 4
  1.4 Potential Difference 4
  1.5 AC Waveforms 6
  1.6 Half-wave Rectified Waveforms 7
  1.7 Full-wave Rectified Waveforms 7
  1.8 Triangular Waveforms 8
  1.9 Sawtoothed Waveforms 8
  1.10 Trigger Pulse 8
2.0 Resistances and their Circuits 9
  2.1 Resistance 9
  2.2 Unit of Resistance 9
  2.3 Factors Affecting Its Resistance 10
  2.4 Ohms Law 11
  2.5 Resistor Power 12
  2.6 Resistor Colour Codes 12
  2.7 Connecting Resistors Together 14
  2.8 Resistors in Series 14
  2.9 Voltage Divider Networks 14
  2.10 Resistors in Parallel 15
  2.11 Two Resistors in Parallel 15
3.0 Capacitors and their Circuits 15
  3.1 Capacitance 15
  3.2 Unit of Capacitance 16
  3.3 Capacitor Types 17
  3.4 Connecting Capacitors Together 17
  3.5 Capacitors in Parallel 17
  3.6 Capacitors in Series 17
4.0 Inductors and their Circuits 18
  4.1 Inductance 18
  4.2 Connecting Inductors Together 19

4.3 Inductors in Series....20
4.4 Inductors in Parallel....20
4.5 Two Inductors in Parallel....20
5.0 Impedance and Reactance 21
5.1 Impedance 21
5.2 AC Resistance 21
5.3 Capacitive Reactance 22
5.4 AC Capacitance....22
5.5 Inductive Reactance 22
5.6 AC Inductance 23
6.0 RC Filters 23
6.1 Low Pass RC Filter....24
6.2 High Pass RC Filter 24
6.3 Band Pass RC Filter....25
6.4 Band Stop RC Filter 25
7.0 Semiconductors 26
7.1 The Semiconductor Diode 27
7.2 Applications Of Diodes 28
7.3 The Zener Diode....29
7.4 Zener Voltage Regulator 30
7.5 The Bipolar Junction Transistor 30
7.6 The Field Effect Transistor....32
7.6 The MOSFET....33
8.0 Operational Amplifiers 34
8.1 Inverting Amplifier....35
8.2 Non-inverting Amplifier....35
8.3 Voltage Follower 35
8.4 Differential Amplifier (Subtractor) 36
8.5 Voltage Summing Amplifier 36
8.6 Integrator Amplifier 36
8.7 Differentiator Amplifier 37

Basic Electronics Tutorials

1.0 ELECTRICAL THEORY

1.1 THE STRUCTURE OF ATOMS

Atoms are not solid but composed of three fundamental particles: ***electrons, protons,*** and ***neutrons*** arranged in various combinations. The electron is defined as being negatively (-) charged and revolves around the nucleus of the atom in various concentric paths called ***orbits***. The proton is defined as being positively (+) charged. Neutrons are defined as being uncharged or neutral. Protons and neutrons are tightly bound together within the atoms nucleus and are not free to orbit.

In an atom the number of negative electrons and positive protons are the same making it electrically neutral. The number of protons present within the atoms nucleus specifies its atomic number. The corresponding numbers of electrons are arranged in different elliptical orbits, called ***shells***, around the nucleus. Electrons in different orbits can rotate around the nucleus in all directions, thus producing a three-dimensional atom.

The electrons in the nearest orbit having a great force of attraction while the electrons in the farthest orbit having the least force of attraction. Electrons in the farthest orbit which are loosely held to the nucleus are called ***valence electrons*** and therefore rotate around the valence shell. Copper for instance, has one valence electron.

An Atoms Structure

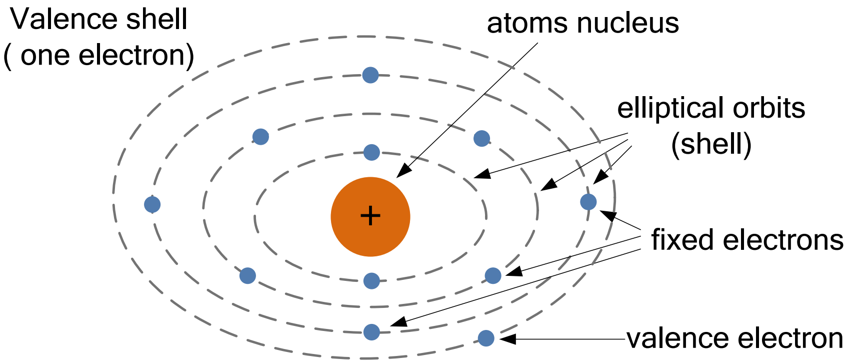

The loosely held electrons in the outer shell often break free due to an input of energy such as heat allowing them to move randomly around through the space in between the various orbits of the other atoms. Such loose electrons are called "free electrons". An atom that loses an electron in this way is left positively charged since it now has an excess of protons. If more electrons attach themselves to the valence shell, then there are more electrons than protons and the atom becomes negatively charged.

1.2 THE UNIT OF CHARGE

The unit of charge is the ***Coulomb***, **C**. The symbol of electric charge is **Q**. The charge of one electron is given as: 1.6×10^{-19}, so one Coulomb of charge is equal to $1/ 1.6 \times 10^{-19}$ or 6.25×10^{18} electrons.

The more free electrons the valence shell has the better the conduction of current by the atom

Thus:

$$\boxed{Q = 1 \text{ Coulomb} = \text{charge of } 6.25 \times 10^{18} \text{ Electrons}}$$

1.3 ELECTRIC CURRENT

Electric current is the flow of electric charge in the form of free electrons. Current is measured by the number of free electrons passing a particular point within a circuit per second. Therefore the flow of charge per unit second defines the amount of electric current. When the charge moves at the rate of 6.25 x 10^{18} electrons flowing past a given point per second, the value of the current is one ampere. This is the same as saying one coulomb of charge per second.

The ***SI*** (*International System of Units*) unit of current is the ***ampere*** with letter symbol **A**. A constant current has symbol **I**, while a time-varying current has a symbol **i** for intensity. Mathematically we can define the relationship between charge (Q) and electric current (I) as:

$$I(\text{amperes}) = \frac{Q(\text{coulombs})}{t(\text{seconds})}$$

Where:

I = Average Current flowing

Q = Total Charge passing a fixed point

t = Time taken to pass point

Electric current has a direction associated with it. Conventional current flow is in the direction of positive charge movement from positive to negative. Electron flow is in the opposite direction from negative to positive. The arrow in a circuit specifies the direction of positive current flow.

In solid metals only negatively charged free electrons move to produce a current flow, the positive protons can not move. But in a liquid or a gas, both the positive protons and negative electrons move to produce a current flow.

> An electric current of one ampere flows in a circuit when a charge of one coulomb passes a given point in one second

Since electric circuits consist almost entirely of solid metal conductors such as copper wire, only negatively charged electrons produce a current flow. Current is also a measure of how intense or concentrated the electron flow is.

1.4 POTENTIAL DIFFERENCE

When two positive charges or two negative charges are brought near to each other they repel while a positive and negative charge are attracted to each other. Then a charged particle has the ability to do work. The ability of a charged particle to do work is called an electric potential. Thus two dissimilar charges have a *difference of potential* and the unit of potential difference (pd) is called the volt.

The ***volt*** unit of potential difference, named after Alessandro Volta (1745–1827), involves work which is the measure of the amount of work required to move an electric charge, which in turn involves force and distance. The SI unit of work is the *joule* with unit symbol **J**, the SI unit of force is the *newton* with unit symbol **N**, and the SI unit for distance is the meter with unit symbol **m**.

W (joules) = N (newtons) x m (meters)

Energy is the capacity to do work. P*otential energy,* which is the energy a body has because of its physical position.

Basic Electronics Tutorials

In electronics, potential difference is commonly referred to as voltage, with the symbol **V**. Sometimes the symbol **U** or **E** for emf (electromotive force) is used, but the standard symbol **V** represents any potential difference. This applies either to the voltage generated by a source such as a battery or solar cell, or to the voltage dropped across a passive component such as a resistor.

The voltage difference (also called potential difference) between two points is the work in joules required to move one coulomb of charge from one point to the other. The SI unit of voltage is the volt and is given as:

$$V(\text{volts}) = \frac{W(\text{joules})}{Q(\text{coulombs})}$$

Where:

V = is the Voltage in Volts

W = is the Work done

Q = Total Charge passing a fixed point

A constant voltage is called a DC voltage and a voltage that varies sinusoidally with time is called an AC voltage. A voltage source such as a battery or solar cell provides a constant DC voltage, for example 12 VDC. A voltage source such as an alternator or generator provides an alternating AC voltage, so for example 240 VAC.

If point a is positive with respect to point b, moving a positive charge around a closed circuit from a to b (or a negative charge from b to a) requires work. The difference between the two points is the *voltage polarity.* This voltage polarity is indicated by a positive sign (+) at point a and a negative sign (-) at point b.

> The potential difference between two points is one volt if one joule of work is done displacing one coulomb of charge

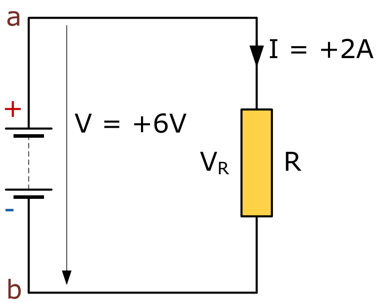

In the circuit the arrow indicates the direction of current flow. If the arrow points in the same direction of the positive charge carriers (conventional current flow) the numeric vale of the current receives a positive sign, +2A. If current flow is opposite (electron flow), the numerical value receives a minus sign, -2A. The sign indicates the direction of current flow with the arrow (+) or in reverse (-).

Then **Conventional Current Flow** gives the flow of electrical current from positive to negative and **Electron Current Flow** around a circuit from the negative terminal to the positive. The item which provides a path for the electrons to flow is called a conductor.

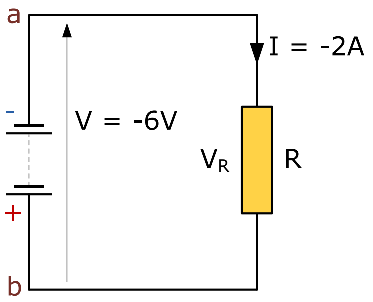

The polarity of the voltage is also indicated by the direction of an arrow. If the arrow points from more positive to more negative potential, the numerical value of the voltage has a positive sign, +6V. If it points from a more negative to a more positive potential, then the numerical value receives a minus sign, -6V.

The battery symbol is often used to denote a DC voltage source, but it may not always be a battery. Usually the positive (+) and negative (-) signs are not shown because, by convention, the long end line denotes the positive terminal (the Anode) and the short end line the negative terminal (the Cathode). Thus, it is not necessary to put + and - signs on the diagram.

For more information visit our website at: www.electronics-tutorials.ws

1.5 AC WAVEFORMS

Alternating current, (AC) is current that changes direction (usually many times per second) cyclically, passing first in one direction, then in the other through a circuit. Such alternating currents are produced by generators and other such voltage sources whose polarities alternate between a positive direction and a negative direction rather than being fixed in a constant direction as with DC sources. By convention, alternating currents are called *AC currents* and alternating voltages are called *AC voltages.* The most common AC source is the commercial AC power system that supplies energy to your home.

The variation of an AC voltage or an AC current over time is called a *waveform*. Since these waveforms vary with time, AC supplies are designated by lowercase letters $v(t)$ for voltage, and $i(t)$ for current instead of uppercase letters **V** and **I** for DC values. Note that the subscript *(t)* represents time.

There are many different types and shapes of waveforms but the most fundamental is the sine wave (also called sinusoid). The sine wave or sinusoidal AC waveform is the voltage and current waveform shape at the wall socket outlets in your home.

The waveform starts at zero, increases to a positive maximum (called the peak), and then decreases to zero, changes polarity, increases to a negative maximum, then returns again back to zero. One complete variation between the same points on the waveform is referred to as a ***cycle*.** Since the waveform repeats itself at regular intervals over time, it is called a ***periodic waveform***.

Sinusoidal waveforms are produced by rotating a coil of wire at a constant angular velocity within a fixed magnetic field

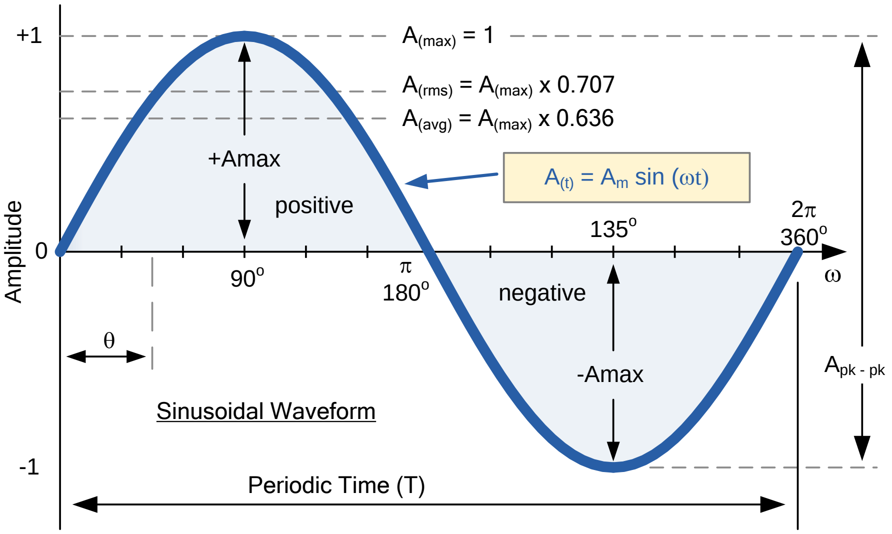

$$f = \frac{1}{T}\ \text{(Hz)}$$

$$\omega = 2\pi f\ \text{(rads/s)}$$

$$A_{(RMS)} = \frac{A_{max}}{\sqrt{2}} = 0.7071V_{max}$$

$$A_{(AVG)} = \frac{2A_{max}}{\pi} = 0.637V_{max}$$

$$\text{Form Factor} = \frac{\pi}{2\sqrt{2}} = 1.11$$

$$\text{Crest Factor} = \sqrt{2} = 1.414$$

The calculation of Average, R.M.S, Form Factor and Crest Factor can be use with any type of periodic waveform including Triangular, Square, Sawtoothed or any other irregular or complex voltage or current waveform shape.

For a pure sinusoidal waveform the effective or R.M.S. value will always be equal to 1/√2 x Amax which is equal to 0.707 x Amax and this relationship also holds true for RMS values of current. The RMS value for a sinusoidal waveform is always greater than its Average value.

Each repetition of a set of positive and negative instantaneous values of the alternating waveform is called a ***cycle***. The sine wave function is periodic in time. This means that the instantaneous value at time *t* will be exactly the same at a later time. The time taken by the alternating waveform to complete one full cycle is known as its **time period** (also called wavelength in radio), denoted by T seconds.

An Alternating Current (AC) waveform is defined as one which changes periodically both in magnitude and direction

The number of cycles per second of a waveform is defined as its **frequency**. Then we can say that frequency is the reciprocal of the time period as $f = 1/T$, and has the unit of inverse seconds, s^{-1}

In the SI system, the unit of frequency is the **hertz** (Hz) and by definition:

1 Hertz (Hz) = 1 cycle per second

Angular frequency (**ω**) is the waveforms frequency expressed in electrical radians per second. As one cycle of an alternating waveform corresponds to 2π radians, the angular frequency can therefore be expressed as: 2π cycles/sec. Angular frequency has the units of radians/second. The relationship between frequency, *f* and angular frequency, ω is:

$$\omega = 2\pi f \text{ rad/sec.} \quad \text{or} \quad \omega = \frac{2\pi}{T} \text{ radians/sec.}$$

The advantage of using alternating voltages and currents for electronic power supplies is that they can be raised and lowered with the help of a device called a transformer. In DC circuits, raising and lowering voltages is not so easy because transformers cannot be used with direct current.

1.6 HALF-WAVE RECTIFIED WAVEFORMS

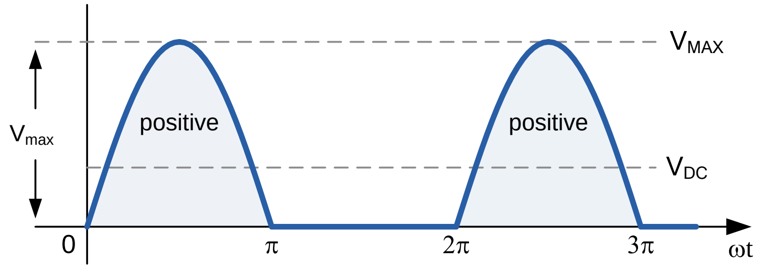

$$V_{DC} = \frac{V_{max}}{\pi} = 0.318V_{max}$$

$$\text{Ripple Factor} = 1.21$$

$$\text{Form Factor} = \frac{\pi}{\sqrt{2}} = 2.22$$

1.7 FULL-WAVE RECTIFIED WAVEFORMS

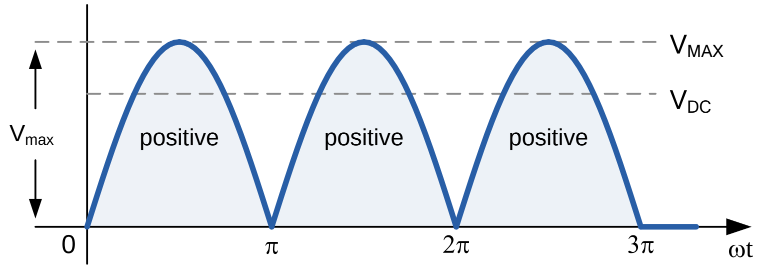

$$V_{DC} = \frac{2V_{max}}{\pi} = 0.636V_{max}$$

$$\text{Ripple Factor} = 0.482$$

$$\text{Form Factor} = \frac{\pi}{2\sqrt{2}} = 1.11$$

1.8 TRIANGULAR WAVEFORMS

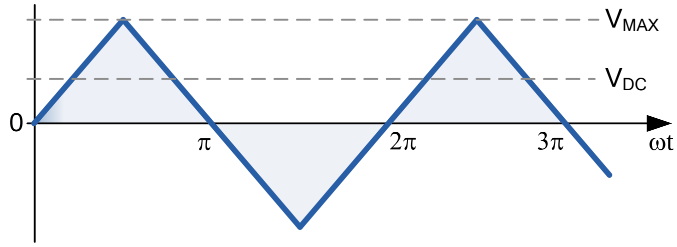

$$V_{DC} = \frac{V_{max}}{\sqrt{3}} = 0.577V_{max}$$

$$\text{Crest Factor} = \sqrt{3}$$

$$\text{Form Factor} = \frac{2}{\sqrt{3}} = 1.155$$

1.9 SAWTOOTHED WAVEFORMS

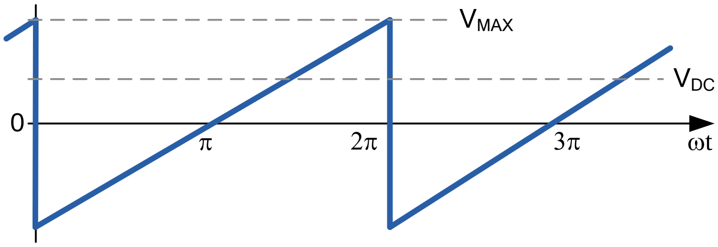

$$V_{DC} = \frac{V_{max}}{\sqrt{3}} = 0.577V_{max}$$

$$\text{Crest Factor} = \sqrt{3}$$

$$\text{Form Factor} = \frac{2}{\sqrt{3}} = 1.155$$

1.10 TRIGGER PULSE

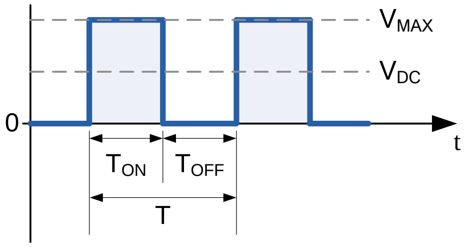

$$V_{DC} = V_{MAX} \times \frac{T_{ON}}{T}$$

$$V_{RMS} = V_{MAX} \times \sqrt{\frac{T_{ON}}{T}}$$

$$f = \frac{1}{T}\ (\text{Hz}) \quad T = T_{ON} + T_{OFF}$$

$$\text{Duty Cycle} = \frac{T}{T_{ON}}$$

Not all alternating voltage waveforms are sine waves. There are also square waves, asymmetrical triangle, rectangular and complex waveforms. Complex waveforms generally consist of base fundamental waveform plus various harmonics superimposed on top. The exact appearance of a complex waveform will depend on the frequencies, magnitudes, and phase relationships of the voltage waves superimposed upon the fundamental wave.

Note that the terms ***wave*** and ***waveform*** do not refer to the same thing as a wave is a varying voltage or current, but a waveform is a graphical representation of such a varying voltage or current.

For more information visit our website at: www.electronics-tutorials.ws

2.0 RESISTANCES AND THEIR CIRCUITS

2.1 RESISTANCE

The current flowing in an electrical circuit not only depends upon the voltage pushing this current around but also on the resistance of the wires, connections and components that make up the circuit. **Resistance**, (**R**) of a circuit is its ability to resist or prevent the flow of current (electron flow) through itself making it necessary to apply a greater voltage to the electrical circuit to cause the current to flow again. Resistance opposes current flow.

The amount of resistance a circuit element has determines whether the element is a "good conductor" with low resistance, or a "bad conductor" (insulator) with high resistance or somewhere in between.

Low resistance, for example one ohm or less implies that the circuit is a good conductor made from materials with lots of free electrons in its valence shell. Examples of good conductors are generally metals such as copper, aluminium, gold, silver or non-metals such as carbon, mercury and some acids and salts.

High resistance, one mega-ohm or more implies the circuit is a bad conductor of electricity made from insulating materials with no free electrons, or tightly grouped electrons in its valence shell. Examples of insulators include glass, porcelain, rubber, pvc (polyvinyl chloride) plastics, mineral oils and dry wood or paper, etc.

Resistance is the opposition to current flowing around an electrical circuit. The unit of resistance is the Ohm

Copper Cable

Insulator
Conductor

2.2 UNIT OF RESISTANCE

The SI unit of resistance is the ***Ohm*** with Greek symbol **Ω** (Omega). A conductor is said to have a resistance of one ohm when one volt causes one ampere of current to flow through it. Note that Resistance cannot be negative in value only positive and in AC circuits, AC resistance equals DC resistance, $Z_R = R$.

Prefixes used for Ohms

Prefix for Ohms	Description	Abbreviation	Value
Mega-Ohm	One Million	MΩ	10^{6} Ω
Kilo-Ohm	One Thousand	kΩ	10^{3} Ω
Ohm	One	Ω	10^{0} Ω
Milli-Ohm	One Thousandth	mΩ	10^{-3} Ω
Micro-Ohm	One Millionth	μΩ	10^{-6} Ω

The most common prefixes used are: **Kilo-ohms** (kΩ = 10^{3}Ω) and **Mega-ohms** (MΩ = 10^{6}Ω).

2.3 Factors Affecting Its Resistance

1. Length of Material: The resistance of a material is directly proportional to its length. The longer the material the more resistance it has.
2. Cross-sectional Area: The resistance of a material is indirectly proportional to its width. The wider or thicker the material is the less resistance it has allowing more free electrons to flow.
3. Type of Material: The type of material affects the amount of free electrons able to flow through it. A material which is a conductor has less resistance while a material which is an insulator has more resistance.
4. Temperature: The temperature of the material affects its resistance. Some materials such as thermocouples and thermistors are design to change their resistance with temperature.

Then the resistance of any material which has a uniform cross-sectional area, **A** and length, **ℓ** can be represented in mathematical form as:

$$R = \rho \frac{\ell}{A}$$

Where: ρ is known as the ***resistivity*** of the material in ohm-meters

The circuit element used to model this current resisting behaviour of a material is called a ***resistor***. The resistor is the simplest passive element used in Electrical and Electronic circuits that is they contain no source of power or amplification but only attenuate or reduce the voltage or current signal passing through them. The circuit symbols used to show a resistor in schematic diagrams are given below, where **R** stands for the resistance of the resistor, in this case 100Ω's.

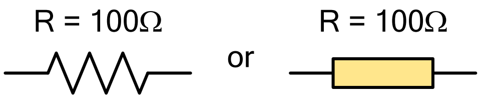

The symbol used in schematic and electrical drawings for a Resistor can either be a "zig-zag" type line or a rectangular box.

A resistor can either be fixed or variable. Most resistors are of the fixed type, meaning their resistance remains constant.

Resistor Symbols

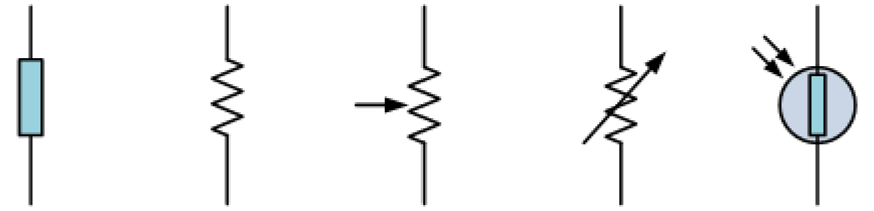

Fixed Value Resistor (IEC Symbol) | Fixed Value Resistor (IEEE Symbol) | Variable Resistor (Potentiometer) | Variable Resistor (Rheostat) | (LDR) Light Dependant Resistor

The two most common types of fixed resistors are wirewound and carbon composition Carbon composition resistors are used when large resistance is needed while wirewound resistors with their metal finned body are used for very high wattage applications. Variable resistors, called potentiometers or rheostats can be either linear or logarithmic types having an adjustable resistance value from zero ohms to their maximum resistance.

2.4 OHMS LAW

The relationship between Voltage, Current and Resistance in any DC electrical circuit was firstly discovered by the German physicist **Georg Ohm**. **Georg Ohm** found that, at a constant temperature, the electrical current flowing through a fixed linear resistance is directly proportional to the voltage applied across it, and also inversely proportional to the resistance.

This relationship between the Voltage, Current and Resistance forms the bases of **Ohms Law** and is shown below.

$$\text{Current (I)} = \frac{\text{Voltage (V)}}{\text{Resistance (R)}} \text{ in Amperes (A)}$$

Ohms Law states that a voltage, V across a resistor, R is directly proportional to the current, I flowing through the same resistor

By knowing any two values of the Voltage, Current or Resistance quantities we can use **Ohms Law** to find the third missing value. **Ohms Law** is used extensively in electronics formulas and calculations so it is "very important to understand and accurately remember these formulas".

To find the Voltage, (**V**)

$$[\ V = I \times R\ ] \quad V\ (\text{volts}) = I\ (\text{amps}) \times R\ (\Omega)$$

To find the Current, (**I**)

$$[\ I = V \div R\ ] \quad I\ (\text{amps}) = V\ (\text{volts}) \div R\ (\Omega)$$

To find the Resistance, (**R**)

$$[\ R = V \div I\ ] \quad R\ (\Omega) = V\ (\text{volts}) \div I\ (\text{amps})$$

A resistor (or resistance) which obeys the above rules of Ohm’s law is known as a *linear* resistor. Linear resistors have a constant resistance for all values of positive or negative voltages and currents. This linear relationship gives a current-voltage (*I-V*) characteristic of a straight line. A *nonlinear* resistor does not obey Ohm’s law and will therefore have a non-straight I-V curve.

For more information visit our website at: www.electronics-tutorials.ws

2.5 RESISTOR POWER

The unit of electrical ***power*** is the ***Watt*** (**W**). One watt of power is equal to the work done in one second by one volt of potential difference in moving one coulomb of charge around a circuit. As an ampere is equal to one coulomb per second, electrical power therefore equals the product of volts times amperes, **P = VxI**

A resistor can be used at any combination of voltage (within reason) and current so long as its "Dissipating Power Rating" is not exceeded. The ***power rating*** of a resistor, also called its “wattage” rating, is an indication how much heat a resistor or resistive element can safely dissipate or convert into heat before becoming damaged. If more heat is generated by the resistor than can be dissipated, the resistor will overheat and become damaged. Resistor power rating is specified in watts.

$$\text{Power, (P)} = V \times I = I^2 \times R = \frac{V^2}{R} \text{ Watts}$$

Where: ***V*** is the voltage across the resistor, ***R*** in ohms, producing the current, ***I*** in amperes, for power, ***P*** in watts.

When calculating the power in resistors or resistances, the main equation to use whenever there is current flowing in the resistance is **I^2R**.

The physical size of a resistor is no indication of its resistance as a small resistor can have a very low or a very high resistance value. A resistors physical size, however, does give some indication of its power rating. As the dissipated resistor power rating is linked to their physical size, a 1/4 (0.250)W resistor is physically smaller than a 1W resistor, and resistors that are of the same ohmic value are also available in different power or wattage ratings. Carbon and metal film resistors, for example, are commonly made in wattage ratings of 1/8 (0.125)W, 1/4 (0.250)W, 1/2 (0.5)W, 1W, and 2 Watts.

Generally speaking the larger their physical size the higher its wattage rating. However, it is always better to select a particular size resistor that is capable of dissipating two or more times the calculated power. When resistors with higher wattage ratings are required, wirewound resistors fitted to metal heatsinks are generally used to dissipate the excessive heat.

Whenever current flows through a resistor heat is produced which represents electrical power in Watts

When selecting or replacing a resistor for a circuit, first determine the required resistance value using R = V/I, then calculate the amount of power that will be dissipated by the resistor using any one of the power formulas above. When selecting the appropriate resistor for a circuit, always try to select a resistor with a higher wattage rating than the actual calculated power dissipation for safety reasons as resistors that conduct lots of current can become very hot.

2.6 RESISTOR COLOUR CODES

Small resistors use coloured painted bands which represent a number and multiplier to indicate both their resistive value and their tolerance with the physical size of the resistor indicating its wattage rating. These coloured painted bands produce a system of identification generally known as a **Resistors Colour Code**.

4 Coloured Bands

Resistor Colour Code Table

Colour	Digit	Multiplier	Tolerance
Black	0	1	
Brown	1	10	± 1%
Red	2	100	± 2%
Orange	3	1,000	
Yellow	4	10,000	
Green	5	100,000	± 0.5%
Blue	6	1,000,000	± 0.25%
Violet	7	10,000,000	± 0.1%
Grey	8		
White	9		
Gold		0.1	± 5%
Silver		0.01	± 10%
None			± 20%

There are two resistor colour coded systems, the four-band system and the five-band system, which use coloured bands or rings that completely encircle its body to identify the resistors value. These coloured bands are usually printed towards one end of the resistors body to indicate the first digit with the colours being read from left to right.

In the four-band system, the first band closest to the edge represents the first digit of the resistance value, the second band is the second digit, the third band is the decimal multiplier, which tells us how many zeros to add after the first two digits and the fourth band is the tolerance giving *Digit, Digit, Multiplier, Tolerance*. Resistors having standard tolerance value of 20% usually do not have a fourth band.

The five-band system displays the coloured bands the same as for the four-band, except for an additional third coloured band to represent a third significant digit giving, *Digit, Digit, Digit, Multiplier, Tolerance.* The five-band system is used for high precision resistors with low tolerance.

Preferred Values of E6, E12 and E24 Series Resistors

E6	1.0				1.5				2.2			
E12	1.0		1.2		1.5		1.8		2.2		2.7	
E24	1.0	1.1	1.2	1.3	1.5	1.6	1.8	2.0	2.2	2.4	2.7	3.0

E6	3,3				4,7				6,8			
E12	3,3		3,9		4,7		5,6		6,8		8,2	
E24	3,3	3.6	3,9	4.3	4.7	5.1	5.6	6.2	6.8	7.5	8.2	9.1

2.7 CONNECTING RESISTORS TOGETHER

Individual resistors can be connected together in series connections, a parallel connections or combinations of both series and parallel together to produce more complex resistor networks. These resistive networks have an equivalent resistance which is a combination of the individual resistors. Then complicated resistor networks can be replaced by a single equivalent resistor, $R_{(EQUIVALENT)}$. It makes no matter what the combination or complexity of the resistor network is, all resistors obey the same basic rules defined by *Ohm's Law* above.

2.8 RESISTORS IN SERIES

Resistors are said to be connected in "**Series**", when they are daisy chained together in a single line. Since all the current flowing through the first resistor has no other way to go it must also pass through the second resistor and the third and so on. Resistors in series have a **Common Current** flowing through them as the current that flows through one resistor must also flow through the others as it can only take one path. Then: $\mathbf{I_{R1} = I_{R2} = I_{R3} = I_{R4}}$ etc.

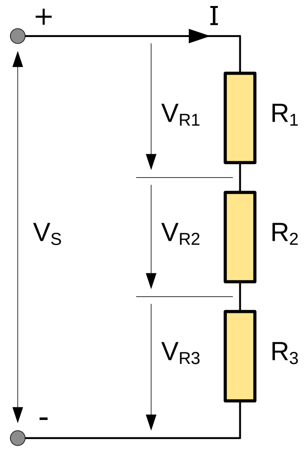

$$R_T = R_1 + R_2 + R_3 + \ldots$$

$$V_S = V_{R1} + V_{R2} + V_{R3} + \ldots$$

$$I = \frac{V_S}{R_T} = \frac{V_{R1}}{R_1} = \frac{V_{R2}}{R_2} = \frac{V_{R3}}{R_3}$$

2.9 VOLTAGE DIVIDER NETWORKS

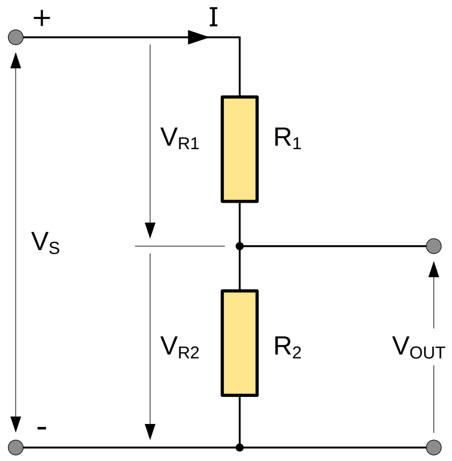

$$R_T = R_1 + R_2$$

$$V_S = V_{R1} + V_{R2}$$

$$V_{OUT} = V_S \frac{R_2}{R_T} = V_S \frac{R_2}{R_1 + R_2}$$

2.10 RESISTORS IN PARALLEL

Resistors are said to be connected together in "**Parallel**" when both of their terminals are respectively connected to each terminal of the other resistor or resistors. Unlike the previous series circuit, in a parallel resistor network the current can take more than one path.

Since there are multiple paths for the supply current to flow through, the current is not the same at all points in a parallel circuit. However, the voltage drop across all of the resistors in a parallel resistive network is the same. Then, **Resistors in Parallel** have a **Common Voltage** across them and this is true for all parallel connected elements.

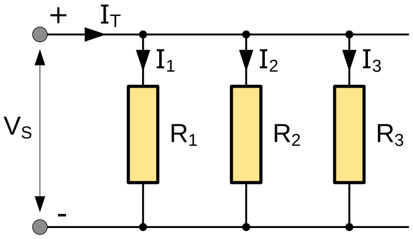

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \dots$$

$$I_T = I_1 + I_2 + I_3 + \dots$$

$$V_S = I_{R1} = I_{R2} = I_{R3} = \dots$$

This method of calculation can be used for calculating any number of individual resistances connected together within a single parallel network. If however, there are only two individual resistors in parallel then a much simpler and quicker formula can be used to find the total resistance value, and this is given as:

2.11 TWO RESISTORS IN PARALLEL

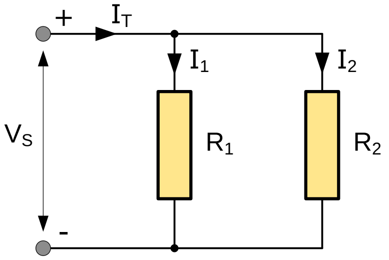

$$R_T = \frac{R_1 \times R_2}{R_1 + R_2}$$

3.0 CAPACITORS AND THEIR CIRCUITS

3.1 CAPACITANCE

Capacitors, **C** are also simple passive devices. The capacitor is a component which has the ability or "capacity" to store energy in the form of an electrical charge producing a potential difference across its plates. Capacitors consists of two or more parallel conductive metal or foil plates which are not connected or touching each other, but are electrically separated either by air or by some form of insulating material such as paper, mica, ceramic or plastic and which is commonly called the capacitors **Dielectric**.

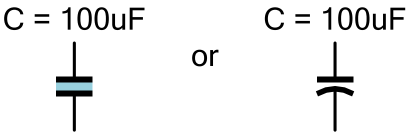

The symbol used in schematic and electrical drawings for a Capacitor can either be a two parallel lines or a straight and curved line.

A capacitors ability to become charged by a voltage and then hold that charge indefinitely allows capacitors to be used in electrical and electronic circuits in a variety of ways, from smoothing out fluctuations in voltage power supply levels to timing and filter circuits when used in conjunction with a resistor.

Capacitor Symbols

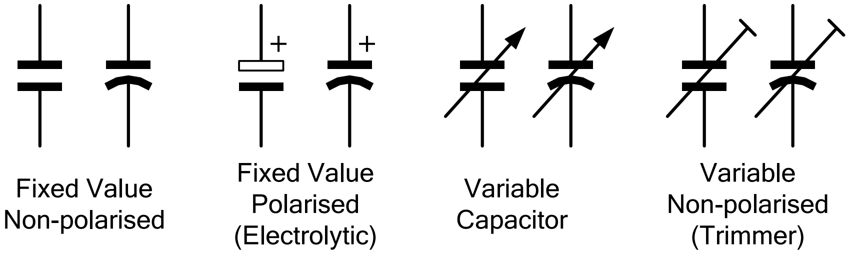
+
+
Fixed Value Non-polarised
Fixed Value Polarised (Electrolytic)
Variable Capacitor
Variable Non-polarised (Trimmer)

When the plates of a capacitor are connected across a DC supply voltage, it takes some time for the charge (in the form of electrons) on the plates to reach their full intensity. When a sufficient amount of charge, **Q** (measured in units of coulombs) have been transferred from the source voltage to the capacitors plates, the voltage across the plates, Vc will be equal to the source voltage, Vs and the flow of electrons will cease.

The voltage developed across the capacitors plates is not instantaneous but builds up slowly at a rate that depends on the capacitance value of the plates, the greater the capacitance, the slower the rate of change of voltage in the plates.

The material used to separates the plates of a capacitor from each other is called the dielectric

The capacitance, **C** value of a capacitor is an expression of the ratio between the amount of charge flowing and the rate of voltage change across the capacitors plates. A capacitance of one ***farad*****,** **F**, represents a charging current of one ampere when there is a voltage, V increase or decrease at a rate of of one volt per second. Then one coulomb of charge exists when a capacitance of one farad is subjected to one volt of potential difference and for a parallel plate capacitor the ratio of **Q ÷ V** is a constant called the capacitance, **C** as shown.

$$Q = VC, \quad C = Q \div V, \quad V = Q \div C$$

Where: **V** in volts, **Q** in coulombs and **C** in farads. Note that when **C** is given in microfarads, (μF) and **V** in volts, charge **Q** will be in micro-coulombs (μC).

3.2 UNIT OF CAPACITANCE

The SI unit of capacitance is the **Farad** (abbreviated to F) named after the British physicist Michael Faraday and is defined as a capacitor has the capacitance of **One Farad** when a charge of **One Coulomb** is stored on the plates by a voltage of **One volt**. Capacitance, C is always positive and has no negative units. However, the Farad is a very large unit of measurement to use on its own so sub-multiples of the Farad are generally used such as micro-farads, nano-farads and pico-farads, for example.

Prefixes used for Ohms

Microfarad **(μF)** 1μF = 1/1,000,000 = 0.000001 = 10^{-6} F

Nanofarad **(nF)** 1nF = 1/1,000,000,000 = 0.000000001 = 10^{-9} F

Picofarad **(pF)** 1pF = 1/1,000,000,000,000 = 0.000000000001 = 10^{-12} F

3.3 CAPACITOR TYPES

The names that are used to describe the different types of capacitors are the names of the dielectric materials used in its construction, because the performance of a capacitor is usually dependant upon the type of material that is used for its dielectric. Also like resistors, there are also variable types of capacitors which allow us to vary their capacitance value for use in radio or "frequency tuning" type circuits.

The various types of capacitors include, disc and tubular ceramics made from aluminium oxide or titanium oxide, silvered mica, metallised film made using strips of waxed or oiled paper and aluminium foil, or with plastic dielectrics such as polyethylene, mylar, polypropylene, polycarbonate, and polyester, and finally large electrolytic capacitors in the form of Aluminum Electrolytic Capacitors and Tantalum Electrolytic Capacitors either polarised or non-polarised.

Variable capacitors change value due to the variation in the overlapping area of the plates, or by varying the spacing between parallel plates. Air dielectric is used for the larger capacitance values. Trimmers and smaller variable types use very thin mica or plastic sheets as the dielectric between the plates.

3.4 CONNECTING CAPACITORS TOGETHER

Like resistors, capacitors can be connected in series, parallel, and series-parallel combinations. Placing capacitors in series effectively increases the thickness of the dielectric, decreases the total capacitance. The total capacitance of capacitors in series is calculated like the total resistance of parallel resistors.

Connecting capacitors together in parallel effectively increases the area of the plates making the total capacitance equal to the sum of the individual capacitances like the total resistance of series resistors. Capacitors in parallel all charge to the same voltage.

3.5 CAPACITORS IN PARALLEL

Capacitors are said to be connected together "in parallel" when both of their terminals are respectively connected to each terminal of the other capacitor or capacitors. The voltage, Vs connected across all the capacitors that are connected in parallel is the same. Then, **Capacitors in Parallel** have a common voltage supply across them.

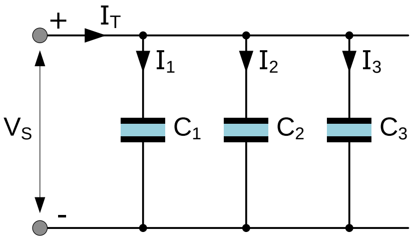

$$C_T = C_1 + C_2 + C_3 + \dots$$

$$Q_T = Q_1 + Q_2 + Q_3 + \dots$$

$$I_T = I_{C1} + I_{C2} + I_{C3} + \dots$$

$$V_C = V_S = \frac{Q_T}{C_T} = \frac{Q_1}{C_1} = \frac{Q_2}{C_2} = \frac{Q_3}{C_3} = \dots$$

3.6 CAPACITORS IN SERIES

Capacitors are said to be connected together "in series" when they are effectively "daisy chained" together in a single line. The charging current (I_C) flowing through the capacitors is **THE SAME** for all capacitors as it only has one path to follow. Then, **Capacitors in Series** all have the same current so each capacitor stores the same amount of charge regardless of its capacitance. Capacitors connected together in series all have the same amount of charge.

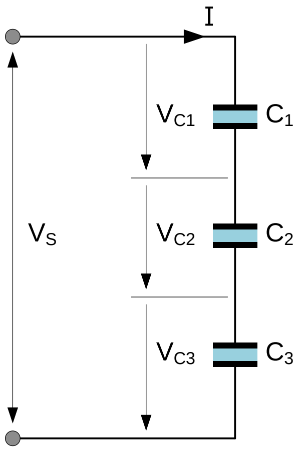

$$\frac{1}{C_T} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots$$

$$V_S = V_{C1} + V_{C2} + V_{C3} + \ldots$$

$$Q = V_S C_T = V_1 C_1 = V_2 C_2 = V_3 C_3 = \ldots$$

$$I_T = I_{C1} = I_{C2} = I_{C3} = \ldots$$

4.0 INDUCTORS AND THEIR CIRCUITS

4.1 INDUCTANCE

When an electrical current flows through a length of wire, a magnetic field is built up around the wire conductor. The direction of this magnetic field can be thought in terms of a wood screw being screwed into the conductor in the direction of the flow of current, with the head of the wood screw being rotated in the direction of the lines of force.

If we now take this length of wire and form it into a coil of N turns, the magnetic flux surrounding the coil is increased many times over for a given coil of wire compared with the flux produced by a single straight length. Also, if the current which is flowing through the coils conductor is increased in magnitude, the magnetic flux produced around the coil will also increase in value.

However, as the strength of the magnetic flux increases, it induces a secondary voltage within the coil called a back emf (electro-motive force). Then for a coil of wire, a self-induced voltage is developed *across* the coil due to the change in current flowing *through* the coil. The polarity of this self-induced voltage produces a secondary current in the coil that generates another magnetic flux which opposes any changes to the original flux.

An Inductor is a coil of wire which opposes the flow of current through itself in the form of a magnetic field

In other words, the instant the main current begins to increase (or decrease) in value, there will be an opposing effect trying to limit this change. But because the coil of wire is extremely long, the current through the coil cannot change instantaneously it takes a while for the current to change due mainly to the resistance of the wire and the self-induced effects of the wire coil.

The ability of a coil to oppose any change in current is a result of the ***self-inductance***, ***L*** of the coil. This self-inductance, simply called inductance, value of an inductor is measured in Henries, (H). The Henry is a large unit, so the milli-henry (mH) and the micro-henry (μH) are more commonly used instead. Then the greater the inductance value of the coil, the slower is the rate of change of current for a given source voltage.

Then ***Inductance*** is the characteristic of an electrical conductor that opposes a change in current flow. An ***inductor*** is a device that stores energy within itself in the form of a magnetic field.

The ***Inductor***, also called a choke, is another passive type electrical component which is just a coil of wire that is designed to take advantage of this relationship by inducing a magnetic field in itself or in the core as a result of the current passing through the coil. This results in a much stronger magnetic field than one that would be produced by a simple coil of wire. Inductors can also be fixed or variable.

Inductors are mainly designed to introduce specific amounts of inductance into a circuit. They are formed with wire tightly wrapped around a solid central core which can be either a straight cylindrical rod or a continuous loop or ring to concentrate their magnetic flux. The inductance of a coil varies directly with the magnetic properties of the central core. Ferrite and powdered iron materials are mainly used for the core to increase the inductance by increasing the flux linking the coil.

Then the inductance of a coil which has a uniform cross-sectional area, **A** length, **ℓ** and coil turns, **N** can be represented in mathematical form as:

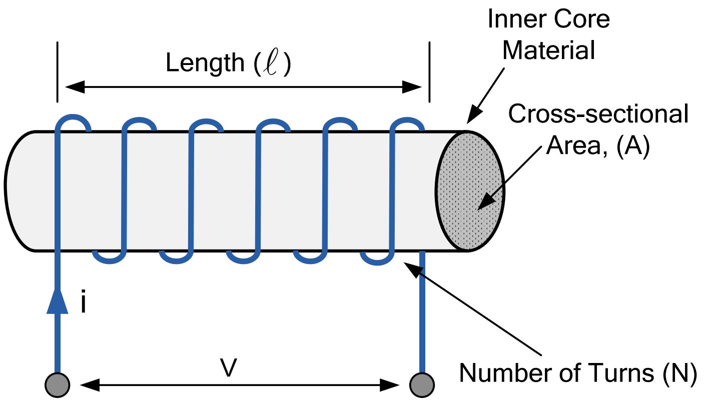

$$L = \frac{N^2 \mu A}{\ell}$$

Where: **N** is the number of coil turns, **μ** is the permeability of the central core, **A** is the area in m^2 and **ℓ** is the mean length of the core in meters.

Inductor Symbols

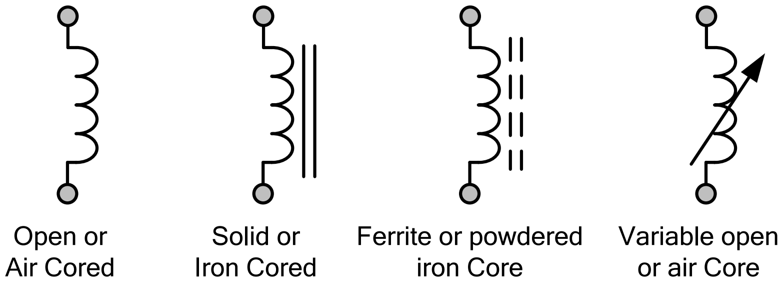

4.2 CONNECTING INDUCTORS TOGETHER

Inductors, like resistors and capacitors, can be connected together in series or parallel combinations. Increasing levels of inductance can be obtained by connecting the inductors in series, while decreasing levels can be obtained by connecting inductors in parallel. However, there are certain rules for connecting inductors in series or parallel and these are based on the fact that no mutual inductance or magnetic coupling exists between the individual inductors.

For more information visit our website at: www.electronics-tutorials.ws

4.3 INDUCTORS IN SERIES

Inductors are said to be connected in "Series" when they are daisy chained together in a straight line, end to end. In the *Resistors in Series* tutorial we saw that the different values of the resistances connected together in series just "add" together and this is also true of inductance. Inductors in series are simply "added together" because the number of coil turns is effectively increased, with the total circuit inductance L_T being equal to the sum of all the individual inductances added together.

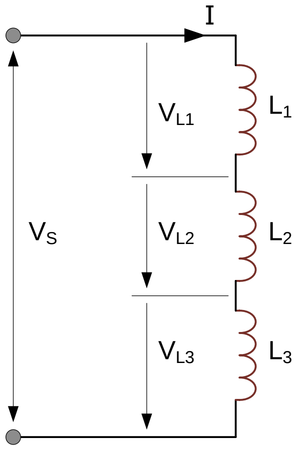

$$L_T = L_1 + L_2 + L_3 + \ldots$$
$$I = I_{L1} = I_{L2} = I_{L3} + \ldots$$
$$V_S = V_{L1} + V_{L2} + V_{L3}$$

4.4 INDUCTORS IN PARALLEL

Inductors are said to be connected together in "**Parallel**" when both of their terminals are respectively connected to each terminal of the other inductor or inductors. The voltage drop across all of the inductors in parallel will be the same. Then, **Inductors in Parallel** have a **Common Voltage** across them and in our example below the voltage across the inductors is given as:

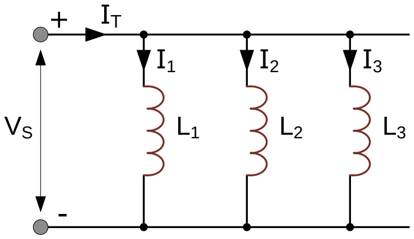

$$\frac{1}{L_T} = \frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} + \ldots$$
$$I_T = I_{L1} + I_{L2} + I_{L3} + \ldots$$
$$V_S = V_{L1} = V_{L2} = V_{L3} = \ldots$$

4.5 TWO INDUCTORS IN PARALLEL

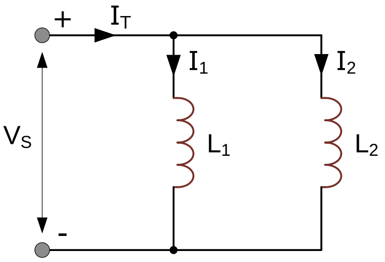

$$L_T = \frac{L_1 \times L_2}{L_1 + L_2}$$

5.0 IMPEDANCE AND REACTANCE

5.1 IMPEDANCE

Any ideal basic circuit element such as a resistor can be described mathematically in terms of its voltage and current, and we saw previously that the voltage across a pure ohmic resistor is linearly proportional to the current flowing through it as defined by *Ohm's Law* . If the voltage across a resistor varies sinusoidally with respect to time, as it does in an AC circuit, the current flowing through the resistor will also vary.

In an AC resistance, the current and voltage are both "in-phase" as there is no phase difference between them. The current flowing through the resistance is directly proportional to the voltage across it with their peak values related by Ohm's law.

Impedance is the opposition to sinusoidal current flow in an AC circuit being the phasor sum of both reactance and resistance

In an AC circuit, impedance represents the opposition to the flow of AC or sinusoidal current. A circuit consisting of reactance (inductive or capacitive) and a resistance will have an equivalent AC resistance known as Impedance, Z. Impedance is the phasor sum of the circuit's reactance, X and the resistance, R. Note that although impedance represents the ratio of two phasors, it is not a phasor itself, because it does not correspond to a sinusoidal varying quantity.

Impedance, which is given the letter Z, in a pure ohmic resistance is a complex number consisting only of a real part being the actual AC resistance value, (R) and a zero imaginary part, (j0). Because of this Ohm's Law can be used in circuits containing an AC resistance to calculate these voltages and currents. As a pure resistor has no reactance, resistance is, for all practical purposes, unaffected by the frequency of the applied sinusoidal voltage or current.

5.2 AC RESISTANCE

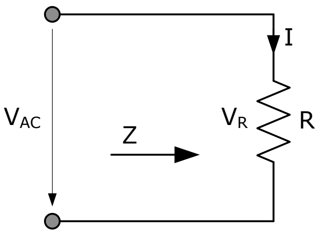

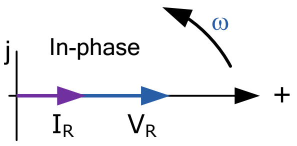

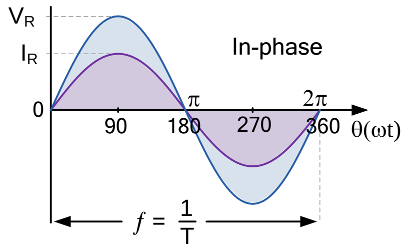

$$Z = R$$

$$Z = \frac{V_R}{I}$$

$$Z = R + j0$$

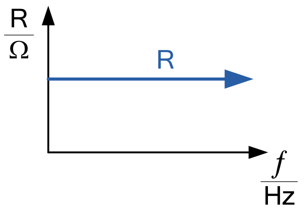

Many AC resistive circuits such as heating elements and lamps consist of a pure ohmic resistance only and have negligible values of inductance or capacitance adding to its impedance. In such circuits we can use both Ohms Law and Kirchoff's laws as well as simple circuit rules for calculating the voltage, current, impedance and power as we would in DC circuit analysis. When working with such rules it is usual to use rms values only.

5.3 CAPACITIVE REACTANCE

When an alternating sinusoidal voltage is applied to the plates of a capacitor, the capacitor is first charged in one direction and then again in the opposite direction, changing polarity at the same rate as the AC supply voltage. This instantaneous change in voltage across the capacitor is opposed by the fact that it takes a certain amount of time to deposit (or release) this charge onto the plates and is given by **V = Q/C**.

Capacitors oppose these changes in sinusoidal voltage with the flow of electrons through the capacitor being directly proportional to the rate of voltage change across its plates as the capacitor charges and discharges. Unlike a resistor were the opposition to current flow is its actual resistance, the opposition to current flow in a capacitor is called **Reactance**.

Like resistance, reactance is measured in Ohm's but is given the symbol "**X**" to distinguish it from a purely resistive ohmic **R** value and as the component in question is a capacitor, the reactance of a capacitor is called ***Capacitive Reactance***, (X_C) which is also measured in Ohms. In an AC capacitance circuit, this capacitive reactance value, X_C is equal to $1/(2\pi f C)$ or $1/(j\omega C)$.

5.4 AC CAPACITANCE

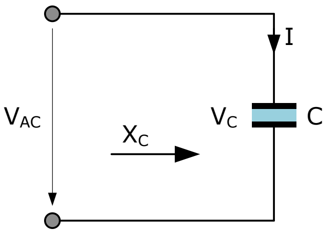

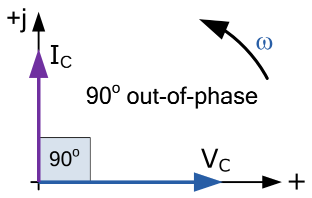

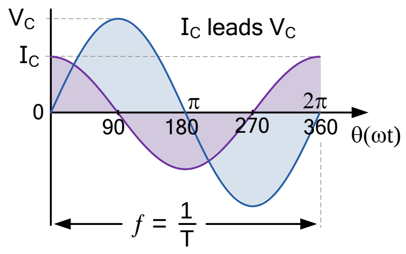

$$X_C = \frac{1}{2\pi f C} = \frac{1}{j\omega C}$$

$$f = 0\text{Hz},\ X_C = \infty\Omega,\ I_C = 0\text{ Amps}$$

$$f = \infty,\ X_C = 0\Omega,\ I_C = \text{Max Amps}$$

$$Z = 0 + jX_C$$

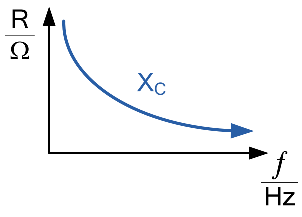

In a pure **AC Capacitance** circuit, the voltage and current are both "out-of-phase" with the current leading the voltage by 90° (or $\pi/2$ rads). So for a pure capacitor, V_C "lags" I_C by 90°, or we can say that I_C "leads" V_C by 90°.

5.5 INDUCTIVE REACTANCE

When an alternating or AC voltage is applied across an inductor the flow of current through it behaves very differently to that of an applied DC voltage. The effect of a sinusoidal supply produces a phase difference between the voltage and the current waveforms. In an AC circuit, the opposition to current flow through an inductors coil windings not only depends upon the inductance of the coil but also the frequency of the AC waveform.

The opposition to current flowing through a coil in an AC circuit is determined by the AC resistance, more commonly known as **Impedance** (**Z**), of the circuit. As the component we are interested in is an inductor, the reactance of an inductor is therefore called "Inductive Reactance". In other words, an inductors electrical resistance when used in an AC circuit is called ***Inductive Reactance***.

Inductive Reactance which is given the symbol $\mathbf{X_L}$, and is the property in an AC circuit which opposes the change in the current. In an AC inductive circuit, this capacitive reactance value, $\mathbf{X_L}$ is equal to $2\pi f L$ or jωC L.

5.6 AC INDUCTANCE

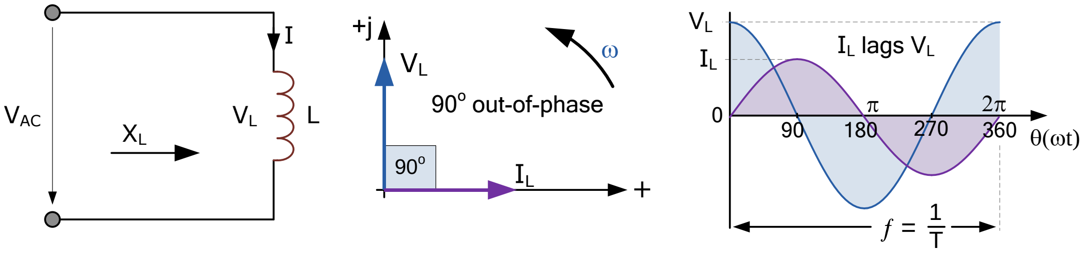
I
V_{AC}
X_L
V_L
L
+j
V_L
ω
90° out-of-phase
90°
I_L
+
V_L
I_L
0
I_L lags V_L
π
2π
90
180
270
360
θ(ωt)
$f = \frac{1}{T}$

$$X_L = 2\pi f L = j\omega L$$
$$f = 0\text{Hz},\ X_L = 0\Omega,\ I_L = \text{Max Amps}$$
$$f = \infty,\ X_L = \text{High }\Omega,\ I_L = 0\,\text{Amps}$$
$$Z = 0 + jX_L$$

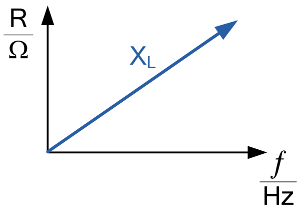

In a pure **AC Inductive** circuit, the voltage and current are both "out-of-phase" with the current lagging the applied voltage by 90° (or π/2 rads). So for a purely inductive AC circuit, V_L "leads" I_L by 90°, or we can say that I_L "lags" V_L by 90°.

6.0 RC FILTERS

RC filters are designed to modify, reshape or reject all unwanted frequencies of an electrical signal and accept or pass only those signals wanted by the circuit's designer. In other words they "filter-out" unwanted signals. Filters are named according to the frequency range of signals that they allow to pass through them, while blocking or "attenuating" the rest. The point at which a filter passes or blocks a signal is called the corner frequency, (fc).

Ideal Filter Response

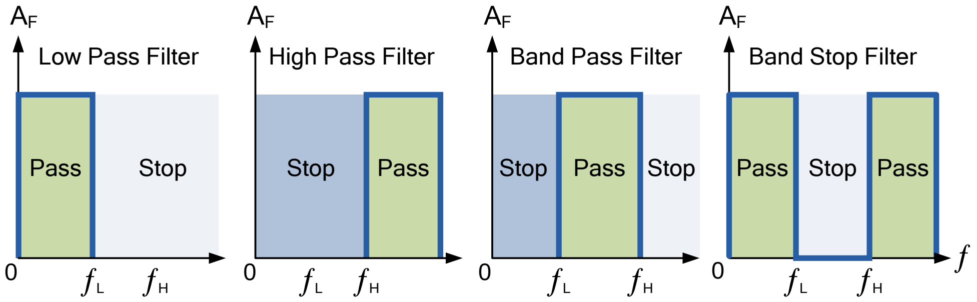
A_F
Low Pass Filter
Pass
Stop
0
f_L
f_H
A_F
High Pass Filter
Stop
Pass
0
f_L
f_H
A_F
Band Pass Filter
Stop
Pass
Stop
0
f_L
f_H
A_F
Band Stop Filter
Pass
Stop
Pass
0
f_L
f_H
f

The pass band allows signals to freely pass unattenuated, while the stop band blocks all signals.

6.1 LOW PASS RC FILTER

A low pass filter allows all low frequency signals to pass from DC (0Hz) up to its corner frequency, (*f*c) and blocks all high frequency signals above this point. A first-order RC low pass filter is constructed as follows:

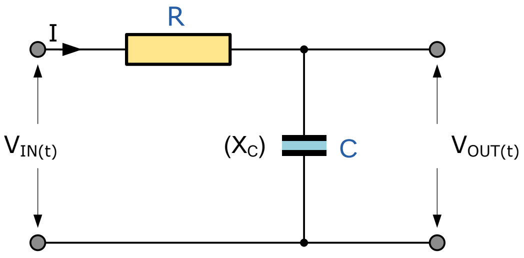

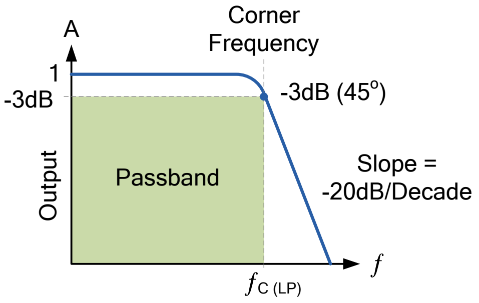

$$A = \frac{V_{OUT}}{V_{IN}} = \frac{X_C}{Z} = \frac{X_C}{\sqrt{R^2 + X_C^2}}$$

$$A = \frac{1}{\sqrt{1 + (\omega RC)^2}}$$

$$f_C = \frac{1}{2\pi RC} \quad \text{or} \quad \omega_C = \frac{1}{RC}$$

$$A_f = -3\text{dB} = \frac{1}{\sqrt{2}} = 0.7071V_{IN}$$

$$\text{Roll-off} = -20\text{dB/Decade}$$
$$\text{or} \quad -6\text{dB/Octave}$$

6.2 HIGH PASS RC FILTER

A high pass filter blocks all low frequency signals from DC up to its corner frequency, (*f*c) and allows all high frequency signals to pass above this point. A first-order RC high pass filter is constructed as follows:

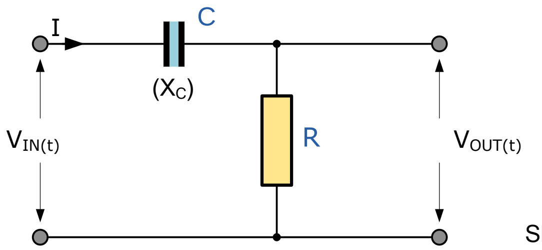

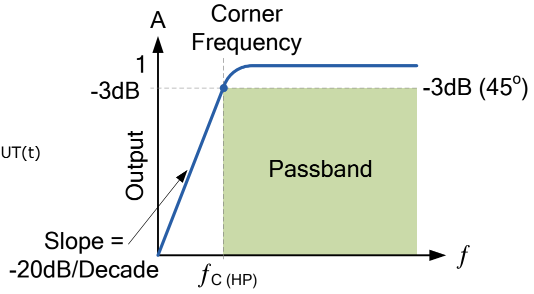

$$A = \frac{V_{OUT}}{V_{IN}} = \frac{R}{Z} = \frac{R}{\sqrt{R^2 + X_C^2}}$$

$$A = \frac{R}{\sqrt{R^2 + \left(\frac{1}{\omega C}\right)^2}}$$

$$f_C = \frac{1}{2\pi RC} \quad \text{or} \quad \omega_C = \frac{1}{RC}$$

$$A_f = -3\text{dB} = \frac{1}{\sqrt{2}} = 0.7071V_{IN}$$

$$\text{Roll-off} = -20\text{dB/Decade}$$
$$\text{or} \quad -6\text{dB/Octave}$$

6.3 BAND PASS RC FILTER

A *band pass* filter blocks all low frequencies below it lower corner frequency and all high frequency above its upper corner frequency only allowing frequency signals in between to pass. Then a band pass filter has two corner or cut-off frequencies. A second-order RC band pass filter is constructed as follows:

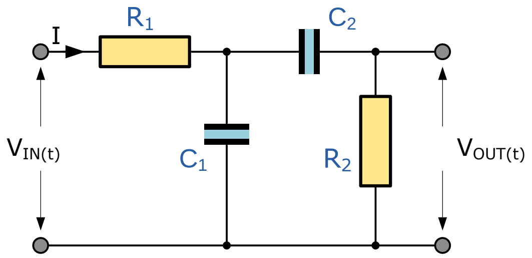

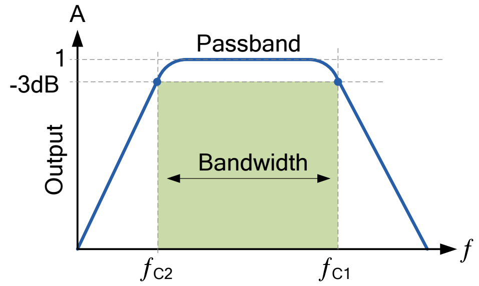

$$R_1C_1 = \text{Low Pass}$$

$$R_2C_2 = \text{High Pass}$$

$$f_{C1} = \frac{1}{2\pi R_1C_1}$$

$$f_{C2} = \frac{1}{2\pi R_2C_2}$$

$$\text{if } R_1 = R_2 \text{ and } C_1 = C_2$$

$$f_C = \frac{1}{2\pi RC} \quad \text{or} \quad \omega_C = \frac{1}{RC}$$

$$f_{\text{PASSBAND}} = \sqrt{f_{C1} - f_{C2}} = \text{Bandwidth}$$

$$f_{\text{CENTER}} = \sqrt{f_{C1} \times f_{C2}}$$

$$Q = \frac{f_{\text{CENTER}}}{f_{C1} - f_{C2}}$$

6.4 BAND STOP RC FILTER

The *band stop* or *notch* filter is another filter design with two corner or cut-off frequencies. The band stop filter blocks (rejects) or severely attenuates a certain band of frequencies between its two corner frequencies while allowing all those outside of this stop-band to pass. A second-order RC band stop filter is constructed as follows:

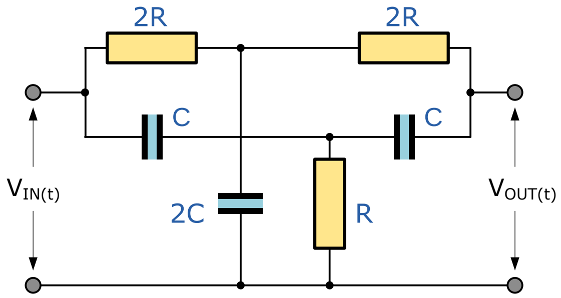

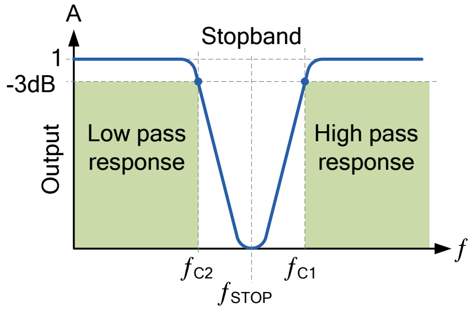

The frequency at which maximum stop attenuation occurs is given as:

$$\boxed{f_{\text{STOP}} = \frac{1}{2\pi RC}}$$

7.0 SEMICONDUCTORS

Semiconductors materials such as Silicon (Si), Germanium (Ge) and Gallium Arsenide (GaAs), have electrical properties somewhere between those of a "conductor" and an "insulator". They neither are not good conductors nor are they good insulators (hence their name "semi"-conductors). They have very few "fee electrons" in their valence shell because their atoms are closely grouped together in a tight crystalline pattern called a "crystal lattice". However, their ability to conduct electricity can be greatly improved by adding certain "impurities" to this crystalline structure thereby, producing more free electrons than holes or vice versa.

By controlling the amount of impurities added to the semiconductor material it is possible to control its conductivity. These impurities are called donors or acceptors depending on whether they produce electrons or holes respectively. This process of adding impurity atoms to semiconductor atoms (the order of 1 impurity atom per 10 million (or more) atoms of the semiconductor) is called **Doping.**

Silicon Atom Structure

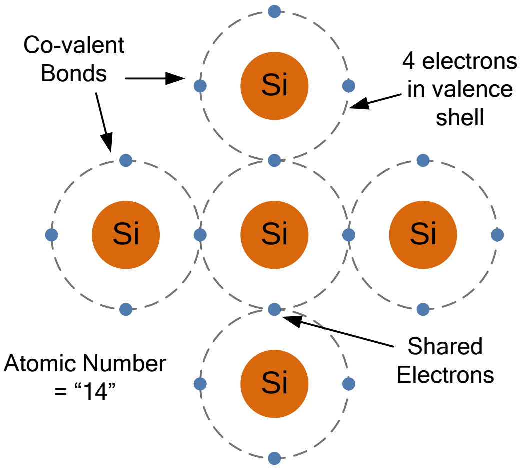

In order for a silicon crystal to conduct electricity, we need to introduce an impurity atom that has five outer electrons in its outermost valence shell to share with its neighbouring atoms. The most common type of "*pentavalent"* (5-electron) impurities used to dope silicon are **Antimony** (symbol Sb) or **Phosphorus** (symbol P), because they have 51 electrons arranged in five shells around their nucleus with the outermost orbital having five electrons.

The resulting semiconductor basics material has an excess of current-carrying electrons, each with a negative charge, and is therefore referred to as an "**N-type**" material. Then N-type Semiconductors are materials which have pentavalent impurity atoms (Donors) added and conduct by "electron" movement. In these types of materials the donors are positively charged and there are a large number of free electrons.

If we go the other way, and introduce a "*trivalent*" (3-electron) impurity into the crystalline structure, such as **Boron** (symbol B) or **Indium** (symbol In), which have only three valence electrons available in their outermost orbital, the fourth closed bond cannot be formed. Therefore, a complete connection is not possible, giving the semiconductor material an abundance of positively charged carriers known as "holes" in the structure of the crystal where electrons are effectively missing.

The doping of Boron atoms causes conduction to consist mainly of positive charge carriers resulting in what is called a "**P-type**" material. Then P-type Semiconductors are a material which have trivalent impurity atoms (Acceptors) added and conducts by the movement of "holes". In these types of materials the acceptors are negatively charged and there are a large number of holes for free electrons to fill.

Semiconductor materials made from silicon are used to produce diodes, mosfet's, bipolar transistors and all types of integrated circuits

So by using different doping agents to a base material of either Silicon (S) or Germanium (Ge), it is possible to produce different types of basic semiconductor materials, either N-type or P-type for use in electronic semiconductor components, microprocessor and solar cell applications.

For more information visit our website at: www.electronics-tutorials.ws

The most basic semiconductor device is the diode formed from the fusing together of both N-type and P-type semiconductor materials. The semiconductor diode is a device that allows current to pass through it in only one direction. This characteristic of a diode has many useful applications in electronics such as rectification of AC voltages and currents to DC.

7.1 THE SEMICONDUCTOR DIODE

A diode is one of the simplest semiconductor devices, which has the characteristic of passing current in one direction only. However, unlike a resistor, a diode does not behave linearly with respect to the applied voltage as the diode has an exponential I-V relationship and therefore we can not described its operation by simply using an equation such as Ohm's law.

Diode I-V Characteristics

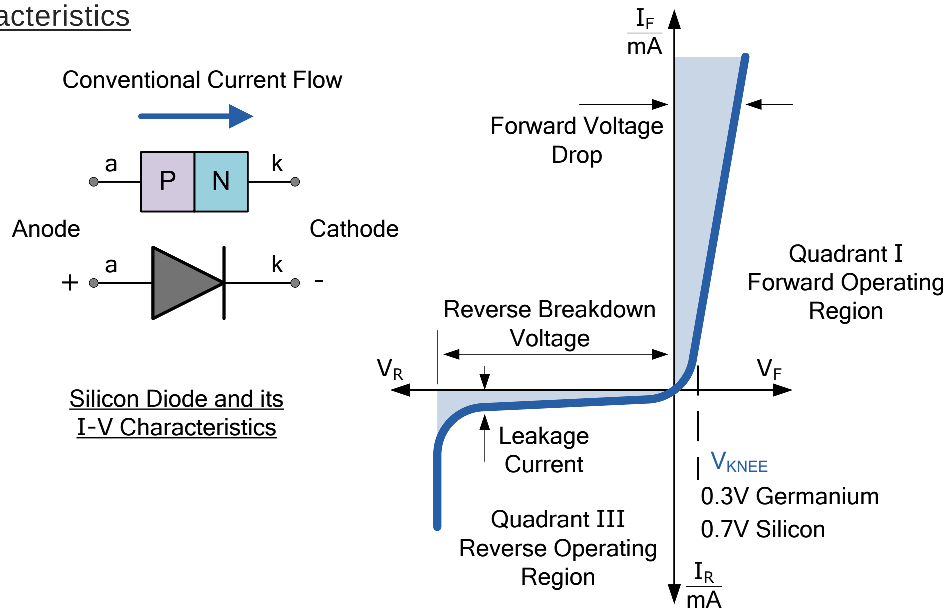
Conventional Current Flow
a
P
N
k
Anode
Cathode
+
a
k
-
I_F
mA
Forward Voltage Drop
Quadrant I
Forward Operating Region
Reverse Breakdown Voltage
V_R
V_F
Leakage Current
V_KNEE
0.3V Germanium
0.7V Silicon
Quadrant III
Reverse Operating Region
I_R
mA

Silicon Diode and its I-V Characteristics

When the diode's anode terminal is more positive than its cathode terminal by at least 0.7 volts for a silicon device, the diode is classed as being "forward biased" allowing current to flow through the device. However, when the anode is made more negative than the cathode, the diode is classed as being "reversed biased" and blocks the flow of current up to its reverse breakdown voltage at which point the diode looses control.

Note that the arrow points in the direction of conventional current flow and the diodes two connections are known as the *Anode*, (A) and *Cathode*, (K). The cathode (negative end) is often marked with a band for identification.

Semiconductor diodes are formed using either Silicon or Germanium semiconductor materials. Diodes are classed as either small signal diodes for use in a variety of small voltage applications or classed as power diodes for use in rectifying and mains powered circuits. Germanium diodes, unlike silicon diodes, only require a forward-biasing voltage of 0.2 volts or greater for conduction to occur.

Diodes being one way devices cannot be connected together in series randomly. Only circuit, "A" will conduct current.

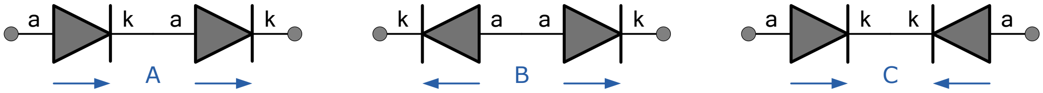
a
k
a
k
A
k
a
a
k
B
a
k
k
a
C

7.2 APPLICATIONS OF DIODES

A very common application of diodes is half-wave rectification, where either the positive or negative half of the input voltage is blocked.

Half-wave Rectification

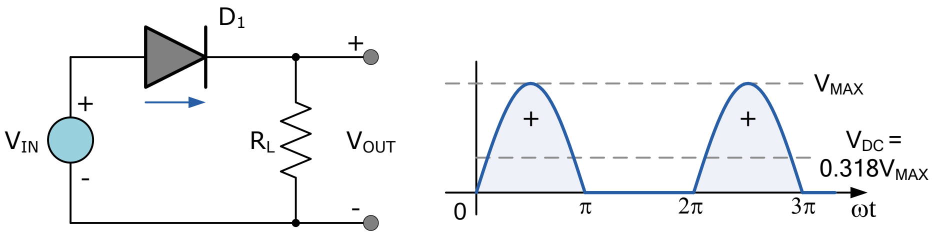
D1
+
-
VIN
RL
+
VOUT
-
VMAX
+
+
VDC =
0.318VMAX
0
π
2π
3π
ωt

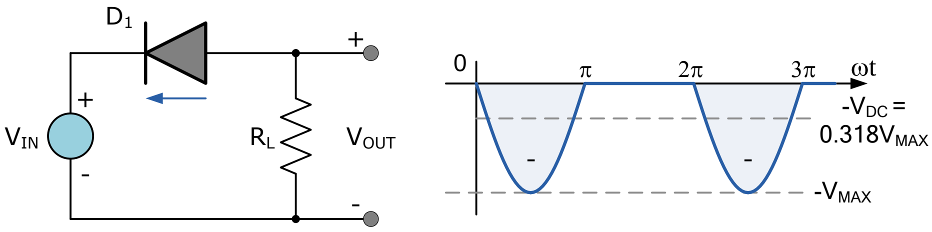
D1
+
-
VIN
RL
+
VOUT
-
0
π
2π
3π
ωt
-VDC =
0.318VMAX
-
-
-VMAX

Full-wave Rectification

A full-wave rectifier passes both the negative and positive half cycles of the input, while inverting the negative half of the input. A full-wave rectifier circuit reduces the ripple by a factor of two. Also the output DC ripple is twice the input supply frequency.

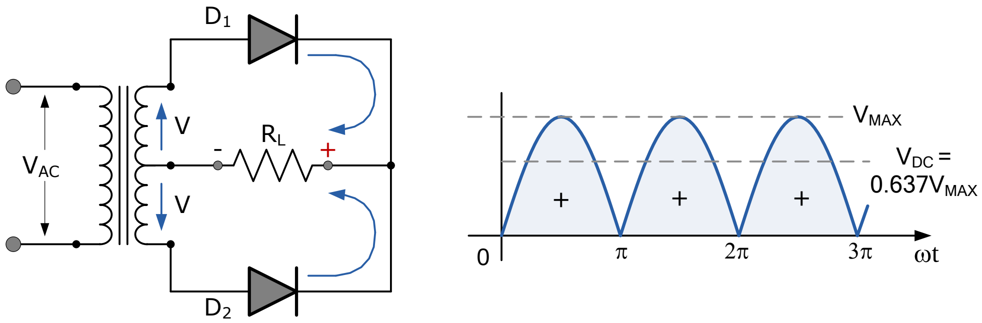
D1
VAC
V
V
RL
-
+
D2
VMAX
VDC =
0.637VMAX
+
+
+
0
π
2π
3π
ωt

Bridge Rectifier

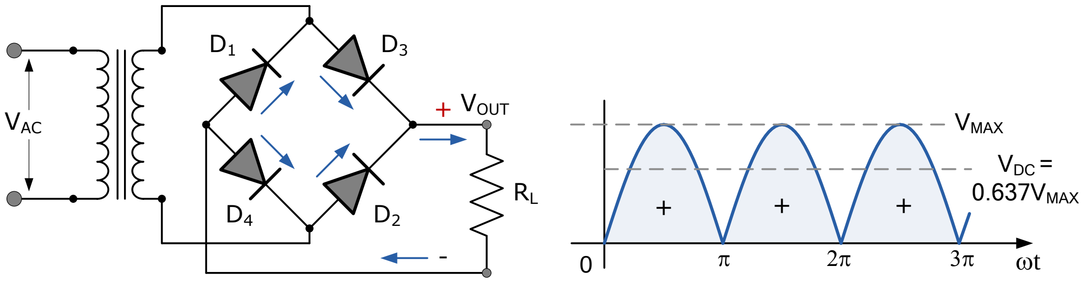
VAC
D1
D3
D4
D2
+ VOUT
RL
-
VMAX
VDC =
0.637VMAX
+
+
+
0
π
2π
3π
ωt

7.3 THE ZENER DIODE

As stated previously, a semiconductor junction diode blocks the flow of current in the reverse direction. However, if we increase the reverse voltage across the diode sufficiently high enough, the diode's PN-junction will breakdown and the diode will become damaged allowing current to flow. The reverse voltage level at which this breakdown occurs is called the *breakdown voltage,* or *peak reverse voltage.*

A zener diode is a special type of diode that conducts normally in the forward-biased mode but is designed to operate in the reverse-biased mode so that at a certain breakdown voltage point, the reverse voltage causes the diode to conduct in a controlled way allowing a reverse current to flow through a series limiting resistor, (Rz). This breakdown voltage is called the **zener voltage**, (**Vz**).

> The zener diode is a special type of diode designed to operate in the reverse bias direction and is used to stabilize or regulate a voltage

The breakdown voltage point of a zener diode, Vz is determined by the resistivity of the diodes junction which is controlled by the doping technique used during its manufacture with zener breakdown voltages ranging from 2.7V to about 200V. The voltage across the zener diode, Vz, remains reasonably constant over a wide range of reverse currents passing through the zener diode.

Zener Diode I-V Characteristics

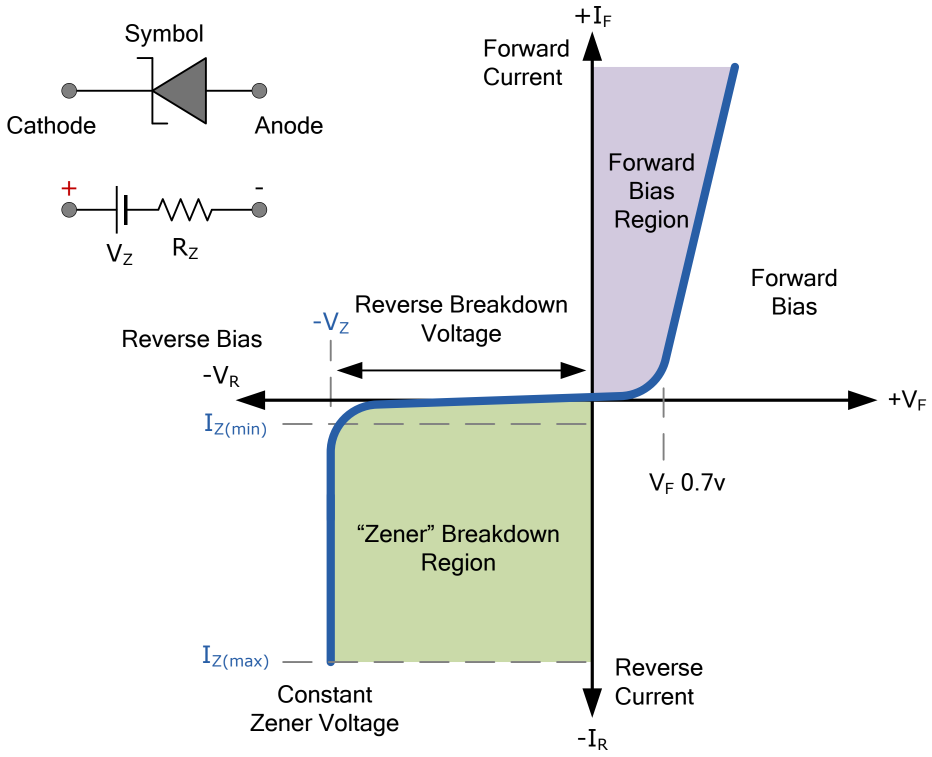
Symbol
Cathode
Anode
+
-
V_Z
R_Z
$+I_F$
Forward Current
Forward Bias Region
Forward Bias
Reverse Breakdown Voltage
$-V_Z$
Reverse Bias
$-V_R$
$+V_F$
$I_{Z(min)}$
V_F 0.7v
"Zener" Breakdown Region
$I_{Z(max)}$
Constant Zener Voltage
Reverse Current
$-I_R$

The zener diode has a region in its reverse bias characteristics of almost a constant negative voltage regardless of the value of the current flowing through the diode and remains nearly constant even with large changes in load current. This ability to control itself can be used to great effect to regulate or stabilise a voltage source against supply or load variations.

7.4 ZENER VOLTAGE REGULATOR

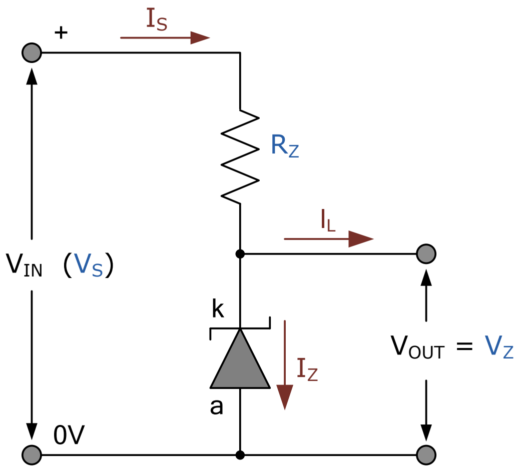

$$I_S = I_L + I_Z$$

$$R_Z = \frac{V_S - V_Z}{I_L + I_Z}$$

$$I_{Z(MAX)} = \frac{V_S - V_Z}{R_Z}$$

$$P_Z = V_Z \times I_Z$$

7.5 THE BIPOLAR JUNCTION TRANSISTOR

The transistor is another electronic device made from different semiconductor materials that can act as either an insulator or a conductor by the application of a small change in current on one lead to produce a large change in voltage, current and power through the other two leads.

In electronic circuits, transistors have two basic functions: "switching" (digital electronics) or "amplification" (analogue electronics). Transistors are made by adding an additional semiconductor layer to a PN-junction diode. In fact bipolar transistors can be thought of as two diodes back-to-back.

> Transistors are electronic devices that can amplify voltages, currents and power by the application of a small signal voltage on one lead

The **Bipolar Transistor** basic construction consists of two PN-junctions producing three connecting terminals with each terminal being given a name to identify it from the other two. These three terminals are known and labelled as the ***Emitter*** (**E**), the ***Base*** (**B**) and the ***Collector*** (**C**) respectively.

Bipolar transistors are current regulating devices that control the amount of current flowing through them in proportion to the amount of biasing voltage applied to their base terminal acting like a current-controlled switch. The principle of operation of the two transistor types PNP and NPN, is exactly the same the only difference being in their biasing and the polarity of the power supply for each type.

Bipolar transistors have the ability to operate within three different regions:

1. Active Region - the transistor operates as an amplifier and Ic = β.Ib

2. Saturation Region - the transistor is "Fully-ON" operating as a switch and Ic = I(saturation)

3. Cut-off Region - the transistor is "Fully-OFF" operating as a switch and Ic = 0

Bipolar Transistor Configurations

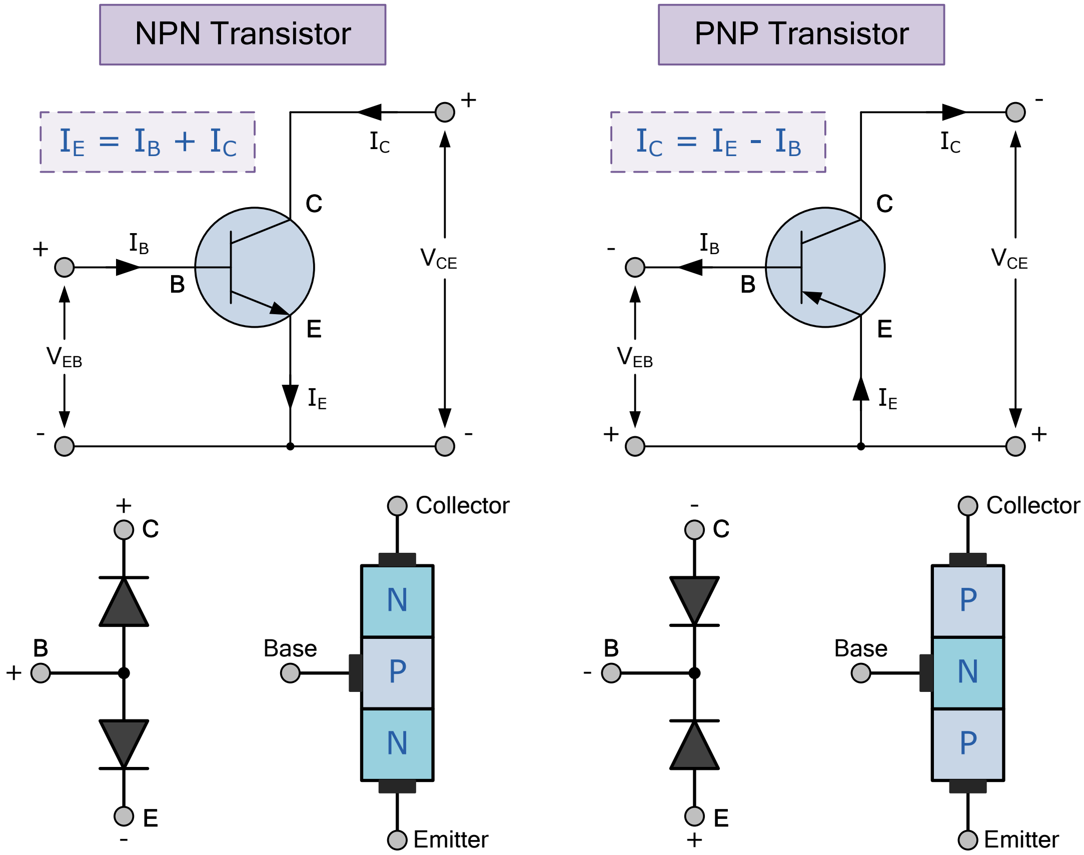
NPN Transistor
PNP Transistor
$I_E = I_B + I_C$
$I_C = I_E - I_B$
I_C
I_B
I_E
V_{CE}
V_{EB}
C
B
E
+
-
Collector
Base
Emitter
N
P
N
P
N
P

The construction and circuit symbols for both the PNP and NPN bipolar transistor are given above with the arrow in the circuit symbol always showing the direction of "conventional current flow" between the base terminal and its emitter terminal. The direction of the arrow always points from the positive P-type region to the negative N-type region for both transistor types, exactly the same as for the standard diode symbol.

Transistor Amplifier

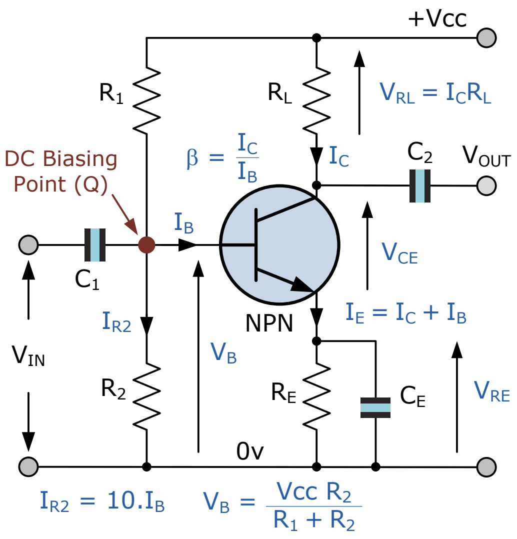
+Vcc
R_1
R_L
$V_{RL} = I_C R_L$
DC Biasing Point (Q)
$\beta = \frac{I_C}{I_B}$
I_C
C_2
V_{OUT}
I_B
C_1
V_{CE}
I_{R2}
NPN
$I_E = I_C + I_B$
V_{IN}
V_B
R_2
R_E
C_E
V_{RE}
0v
$I_{R2} = 10.I_B$
$V_B = \frac{Vcc\ R_2}{R_1 + R_2}$

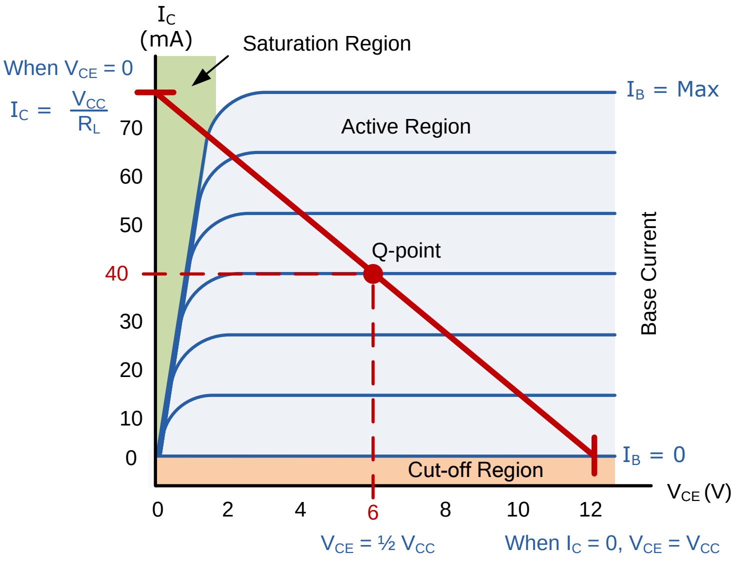

7.6 THE FIELD EFFECT TRANSISTOR

The ***Field Effect Transistor***, or **FET** is another three terminal semiconductor device used for switch and amplifier application as it has very similar characteristics to those of the previous bipolar junction transistor. The field effect transistor is a unipolar device since the flow of current through it is carried out by the majority carriers only. However, unlike the previous bipolar transistor which is a current operated device, the field effect transistor is a "voltage" operated device.

We saw previously that there are two basic types of Bipolar Transistor construction, NPN and PNP, which basically describes the physical arrangement of the P-type and N-type semiconductor materials from which they are made. This is also true of field effect transistors as there are also two basic classifications of FET, called the N-channel FET and the P-channel FET.

The **Field Effect Transistors** basic construction consists of a semiconductor "channel" which corresponds in function to the Collector and the Emitter respectively of the bipolar transistor. This resistive channel is the main current carrying path through the FET and may be made of either a P-type or an N-type semiconductor material.

Like the bipolar transistor, the FET is also a three terminal device with each terminal being given a name to identify it from the other two. These three terminals are known as the ***Drain*** (**D**), the ***Gate*** (**G**) and the ***Source*** (**S**).

Field Effect Transistor Configurations

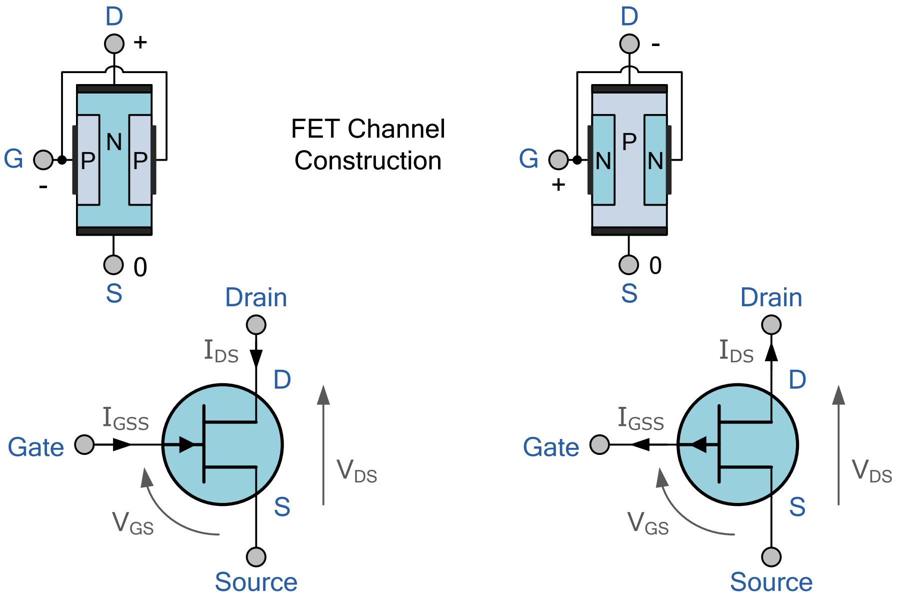

N-channel FET

P-channel FET

Comparison of terminals between the Field Effect Transistor and the Bipolar Junction Transistor.

Bipolar Transistor		Field Effect Transistor
Emitter - (E)	>>	Source - (S)
Base - (B)	>>	Gate - (G)
Collector - (C)	>>	Drain - (D)

Field Effect Transistor Amplifier

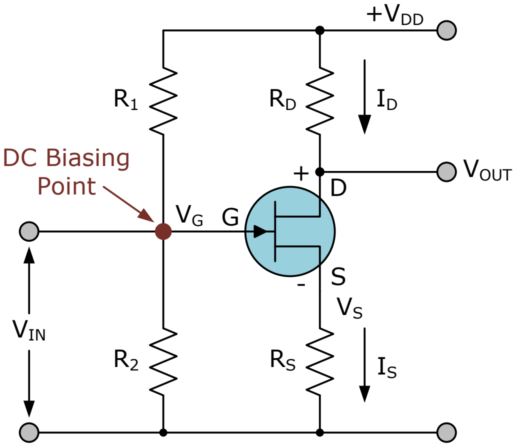

$$V_S = I_D R_S = \frac{V_{DD}}{4}$$

$$V_S = V_G - V_{GS}$$

$$V_G = V_{DD}\left(\frac{R_2}{R_1 + R_2}\right)$$

$$I_D = \frac{V_S}{R_S} = \frac{V_G - V_{GS}}{R_S}$$

The control of the Drain current by a negative Gate potential makes the **Junction Field Effect Transistor** useful as a switch and it is essential that the Gate voltage is never positive for an N-channel JFET as the channel current will flow to the Gate and not the Drain resulting in damage to the JFET. The principals of operation for a P-channel JFET are the same as for the N-channel JFET, except that the polarity of the voltages need to be reversed.

7.6 THE MOSFET

As well as the Junction Field Effect Transistor (JFET), there is another type of Field Effect Transistor available whose Gate input is electrically insulated from the main current carrying channel and is therefore called an **Insulated Gate Field Effect Transistor** or **IGFET**. The most common type of insulated gate FET which is used in many different types of electronic circuits is called the **Metal Oxide Semiconductor Field Effect Transistor** or **MOSFET** for short.

The **IGFET** or **MOSFET** is a voltage controlled field effect transistor that differs from a JFET in that it has a "Metal Oxide" Gate electrode which is electrically insulated from the main semiconductor N-channel or P-channel by a thin layer of insulating material. This insulated metal gate electrode can be thought of as one plate of a capacitor. The isolation of the controlling Gate makes the input resistance of the **MOSFET** extremely high in the Mega-ohms region.

MOSFET's are voltage controlled devices which have insulated gates resulting in an input impedance many times higher than a JFET.

As the Gate terminal is isolated from the main current carrying channel "NO current flows into the gate" and just like the JFET, the **MOSFET** also acts like a voltage controlled resistor were the current flowing through the main channel between the Drain and Source is proportional to the input voltage.

Like the previous JFET, MOSFETs are three terminal devices with a *Gate*, *Drain* and *Source* and both P-channel (PMOS) and N-channel (NMOS) MOSFETs are available. The main difference this time is that MOSFETs are available in two basic forms:

1. Depletion Type (Normally ON) - the transistor requires the Gate-Source voltage, ($\mathbf{V_{GS}}$) to switch the device "OFF". The depletion mode MOSFET is equivalent to a "Normally Closed" switch.

2. Enhancement Type (Normally OFF) - the transistor requires a Gate-Source voltage, ($\mathbf{V_{GS}}$) to switch the device "ON". The enhancement mode MOSFET is equivalent to a "Normally Open" switch.

MOSFET Configurations

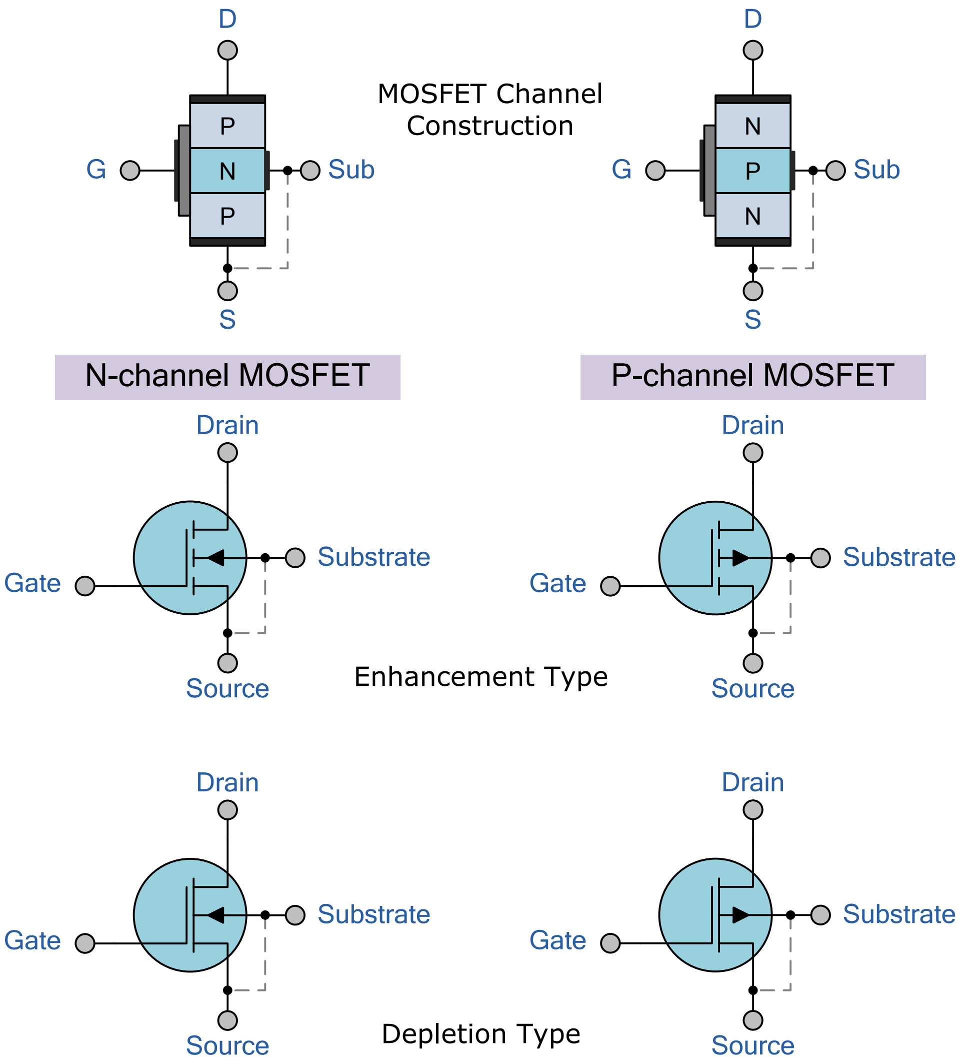
D
MOSFET Channel Construction
D
G
P
N
P
Sub
G
N
P
N
Sub
S
S
N-channel MOSFET
P-channel MOSFET
Drain
Drain
Gate
Substrate
Gate
Substrate
Source
Enhancement Type
Source
Drain
Drain
Gate
Substrate
Gate
Substrate
Source
Depletion Type
Source

One of the main disadvantages of MOSFET devices is their very high input impedance making them extremely sensitive to electrostatic discharge (ESD) on the gate terminal. The metal oxide insulating layer around the gate terminal is extremely thin and can be easily punctured by static so be careful when handling MOSFET devices.

8.0 OPERATIONAL AMPLIFIERS

Operational Amplifiers, or "Op-amps" as they are more commonly called, are linear devices that have all the properties required for nearly ideal DC amplification and are therefore used extensively in signal conditioning, filtering or to perform mathematical operations (hence their name "operational") such as add, subtract, integration and differentiation.

An ideal **Operational Amplifier** is basically a three-terminal device which consists of two high impedance inputs, one is called the **Inverting Input**, marked with a negative or "minus" sign, (-) and the other one is called the **Non-inverting Input**, marked with a positive or "plus" sign (+). A voltage level applied to only one of these inputs will be amplified if the second input is grounded or held at some fixed voltage level.

For more information visit our website at: www.electronics-tutorials.ws

Operational amplifiers can have infinite open-loop voltage gain allowing them to operate at their maximum amplification. However this very high open-loop gain causes the op-amp to become unstable so the gain is reduced to a more practical level by using negative feedback.

Negative feedback in electronics refers to taking a small part of the output signal from the amplifiers output and feeding it back either to aid or to oppose the input signal. Negative feedback means that the returning signal has a phase that opposes the input signal and can significantly improve the performance of an operational amplifier.

Operational amplifiers can be connected into two basic configurations, **Inverting** and **Non-inverting**.

8.1 INVERTING AMPLIFIER

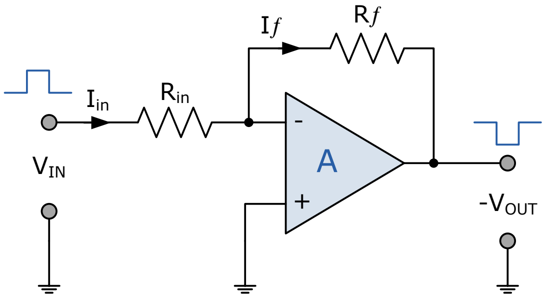

$$A_V = \frac{V_{OUT}}{V_{IN}} = -\frac{Rf}{R_{IN}}$$

$$V_{OUT} = -V_{IN}\left(\frac{Rf}{R_{IN}}\right)$$

$$Z_{IN} = R_{IN}$$

8.2 NON-INVERTING AMPLIFIER

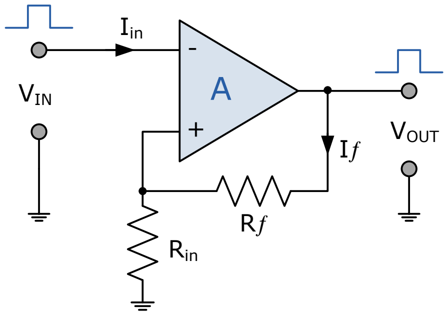

$$A_V = \frac{V_{OUT}}{V_{IN}} = 1 + \frac{Rf}{R_{IN}}$$

$$V_{OUT} = V_{IN}(A_V)$$

$$Z_{IN} = \infty$$

8.3 VOLTAGE FOLLOWER

By connecting the output directly back to the negative input terminal, 100% feedback is achieved resulting in a **Voltage Follower** (buffer) circuit with a constant gain of 1 (Unity).

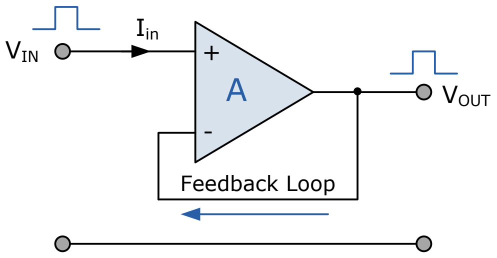

$$V_{IN} = +V, \quad V_{OUT} = -V$$

$$\therefore V_{OUT} = V_{IN}$$

$$A_V = \frac{V_{OUT}}{V_{IN}} = 1$$

$$Z_{IN} = \infty$$

Op amps are used in a wide variety of applications and by adding more input resistors to either the inverting or non-inverting inputs, op-amp **Voltage Adders**, **Voltage Subtractors, Integrators and Differentiators** can be made.

8.4 DIFFERENTIAL AMPLIFIER (SUBTRACTOR)

The **Differential Amplifier** produces an output that is proportional to the difference between the 2 input voltages.

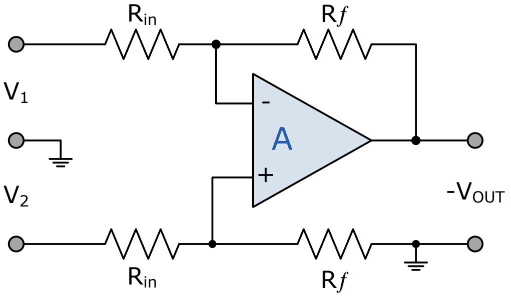

$$V_{OUT} = -\frac{Rf}{R_{IN}}\left(V_2 - V_1\right)$$

$$\therefore V_{OUT} = A\left(V_2 - V_1\right)$$

8.5 VOLTAGE SUMMING AMPLIFIER

Voltage Adders or **Summing Amplifiers** are made by adding more input resistors.

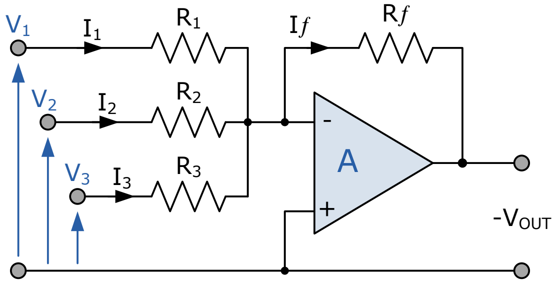

$$If = I_1 + I_2 + I_3 + \ldots$$

$$V_{OUT} = -\left(\frac{Rf}{R_1}V_1 + \frac{Rf}{R_2}V_2 + \frac{Rf}{R_3}V_3\right)$$

If $R_1 = R_2 = R_3$

$$V_{OUT} = -\frac{Rf}{R_{IN}}\left(V_1 + V_2 + V_3 + \ldots\right)$$

8.6 INTEGRATOR AMPLIFIER

The **Integrator Amplifier** produces an output that is the mathematical operation of integration.

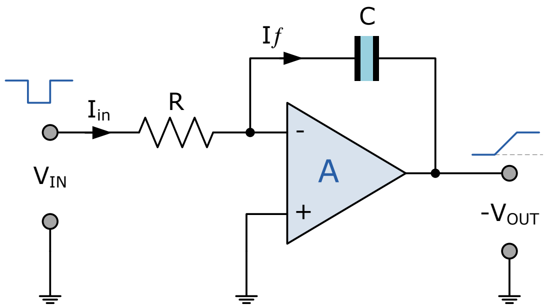

$$V_{OUT} = \frac{1}{2\pi f RC}V_{IN} \quad \text{and} \quad \tau = RC$$

with a sinusoidal input

$$V_{OUT} = \frac{X_C}{R} = \frac{1}{2\pi f RC} \text{ or } \frac{1}{\omega RC}$$

$$\theta = 90^{o}$$

8.7 DIFFERENTIATOR AMPLIFIER

The **Differentiator Amplifier** produces an output that is the mathematical operation of differentiation.

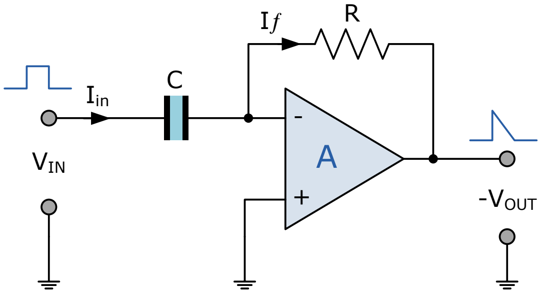

$$V_{OUT} = -\frac{dV_{IN}}{dt}RC \quad \text{and} \quad \tau = RC$$

with a sinusoidal input

$$V_{OUT} = \frac{R}{X_C} = \omega RC$$

$$\theta = -90^{o}$$

The end, for now.

I hope you found this free Electronics reference guide useful and if so please feel free to pass it on to all your Electronics friends and colleagues.

Thank you.
Wayne Storr
http://www.electronics-tutorials.ws

For more information visit our website at: www.electronics-tutorials.ws

Third Edition

Beginner's Guide to Reading Schematics

+12 V
TP3
C3
R3
TP2
Output
C2
Q1
R4
Input
TP1
R2
+12 V
C1
R1

Stan Gibilisco

TAB
Make Great Stuff!

Beginner's Guide to Reading Schematics

Third Edition

Stan Gibilisco

New York Chicago San Francisco Athens London
Madrid Mexico City Milan New Delhi
Singapore Sydney Toronto

Copyright © 2014, 1991, 1983 by McGraw-Hill Education. All rights reserved. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

ISBN: 978-0-07-182779-9

MHID: 0-07-182779-X

The material in this eBook also appears in the print version of this title: ISBN: 978-0-07-182778-2, MHID: 0-07-182778-1.

E-book conversion by codeMantra
Version 1.0

All trademarks are trademarks of their respective owners. Rather than put a trademark symbol after every occurrence of a trademarked name, we use names in an editorial fashion only, and to the benefit of the trademark owner, with no intention of infringement of the trademark. Where such designations appear in this book, they have been printed with initial caps.

McGraw-Hill Education eBooks are available at special quantity discounts to use as premiums and sales promotions or for use in corporate training programs. To contact a representative, please visit the Contact Us page at www.mhprofessional.com.

McGraw-Hill Education, the McGraw-Hill Education logo, TAB, and related trade dress are trademarks or registered trademarks of McGraw-Hill Education and/or its affiliates in the United States and other countries and may not be used without written permission. All other trademarks are the property of their respective owners. McGraw-Hill Education is not associated with any product or vendor mentioned in this book.

Information contained in this work has been obtained by McGraw-Hill Education from sources believed to be reliable. However, neither McGraw-Hill Education nor its authors guarantee the accuracy or completeness of any information published herein, and neither McGraw-Hill Education nor its authors shall be responsible for any errors, omissions, or damages arising out of use of this information. This work is published with the understanding that McGraw-Hill Education and its authors are supplying information but are not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought.

TERMS OF USE

This is a copyrighted work and McGraw-Hill Education and its licensors reserve all rights in and to the work. Use of this work is subject to these terms. Except as permitted under the Copyright Act of 1976 and the right to store and retrieve one copy of the work, you may not decompile, disassemble, reverse engineer, reproduce, modify, create derivative works based upon, transmit, distribute, disseminate, sell, publish or sublicense the work or any part of it without McGraw-Hill Education's prior consent. You may use the work for your own noncommercial and personal use; any other use of the work is strictly prohibited. Your right to use the work may be terminated if you fail to comply with these terms.

THE WORK IS PROVIDED "AS IS." McGRAW-HILL EDUCATION AND ITS LICENSORS MAKE NO GUARANTEES OR WARRANTIES AS TO THE ACCURACY, ADEQUACY OR COMPLETENESS OF OR RESULTS TO BE OBTAINED FROM USING THE WORK, INCLUDING ANY INFORMATION THAT CAN BE ACCESSED THROUGH THE WORK VIA HYPERLINK OR OTHERWISE, AND EXPRESSLY DISCLAIM ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. McGraw-Hill Education and its licensors do not warrant or guarantee that the functions contained in the work will meet your requirements or that its operation will be uninterrupted or error free. Neither McGraw-Hill Education nor its licensors shall be liable to you or anyone else for any inaccuracy, error or omission, regardless of cause, in the work or for any damages resulting therefrom. McGraw-Hill Education has no responsibility for the content of any information accessed through the work. Under no circumstances shall McGraw-Hill Education and/or its licensors be liable for any indirect, incidental, special, punitive, consequential or similar damages that result from the use of or inability to use the work, even if any of them has been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause arises in contract, tort or otherwise.

About the author

Stan Gibilisco has authored or coauthored more than 50 books on physics, electronics, mathematics, and computing. He began his career as an electronics engineer, wireless broadcast technician, and magazine editor. The American Library Association named Stan's *McGraw-Hill Encyclopedia of Personal Computing* (1995) as one of the "Best References of 1996." In addition to authoring several books in McGraw-Hill's *DeMYSTiFieD* series of home study guides, he has written three books in McGraw-Hill's *Know-It-All* series for students who enjoy mathematics. Stan's *Teach Yourself Electricity and Electronics* has become a classic in the field.

This page intentionally left blank

Contents

Introduction ix

1 **What's the scheme? *1***

Block diagrams 2
Schematic diagrams 2
Schematic symbology 4
Schematic interconnections 6
Visual language 8

2 **Block diagrams *13***

A simple example 13
Functional drawings 13
Current and signal paths 16
Flowcharts 18
Process paths 22
Summary 24

3 **Component symbols *25***

Resistors 25
Capacitors 30
Inductors and transformers 34
Switches 39
Conductors and cables 43
Diodes and transistors 46

Electron tubes 49
Cells and batteries 53
Logic gates 54
Summary 56

4 Simple circuits *57*
Getting started 57
Component labeling 66
Troubleshooting with schematics 71
A more complex circuit 75
Schematic/block combinations 78
Summary 81

5 Complex circuits *83*
Identifying the building blocks 83
Page breaks 91
Some more circuits 94
Getting comfortable with large schematics 106
Summary 111

6 Let's learn by doing *113*
Your breadboard 113
Wire wrapping 117
Kirchhoff's current law 119
Kirchhoff's voltage law 123
A resistive voltage divider 125
A diode-based voltage reducer 132
Mismatched lamps in series 137
Summary and conclusion 144

A Schematic symbols *145*

B Resistor color codes *163*

Suggested additional reading 167

Index 169

Introduction

Have you "caught the electronics bug," only to grow sick with apprehension as you encountered diagrams with arcane symbols the moment you decided to build, troubleshoot, or repair something? If so, you're holding the cure in your hands right now.

A little knowledge of electronics symbology can eliminate a lot of fear and confusion. Don't give up electronics just because you've come across some spooky graphics. That would be like giving up a sport because you fear the pain of training. That's where the coach comes in! A good coach streamlines your training in any sport and helps you get past the pain. Schematic diagrams, well drawn and clearly portrayed, can help you to design, build, maintain, troubleshoot, and repair electronic equipment.

Schematic diagrams are like road maps of electronic highway networks. These drawings can help you find your way through the elements of simple circuits, complex devices, and massive systems. Once you've learned what the symbols and notations stand for, reading a schematic will come as easily to you as planning a trip with the aid of a road atlas.

This book contains all the information that you'll need in order to begin exploring electronic circuits. It can help you build a solid background for a career in electronics, whether you decide to go into design, maintenance, or repair. This book explains the rationale behind schematic diagrams, how to use or interpret each symbol, and how the symbols interconnect to reflect real-world circuits. You'll even get a chance to diagram and perform some experiments!

Once you've completed this book, you'll have plenty of information and confidence, so that you can continue your quest to enter whatever field of electronics suits your fancy, whether it's something as humble as shortwave or amateur radio, or something as cutting-edge and exotic as bioelectronics, space communications, or mechatronics.

I welcome your suggestions for future editions. Please visit my website at www.sciencewriter.net. You can e-mail me from there.

Above all, have fun!

Stan Gibilisco

1

What's the scheme?

You'll encounter three types of diagrams in electricity and electronics: block, schematic, and pictorial. Each type of diagram serves its own special purpose.

1. A block diagram gives you an overview of how the discrete circuits within a device or system interact. Each circuit is represented with a "block" (a rectangle or other shape, depending on the application). Interconnecting lines, sometimes with arrows on one or both ends, reveal the relationships between the circuits.
2. A schematic diagram (often simply called a schematic) includes every component that a circuit contains, with each component having its own special symbol. This book is devoted mostly to schematics.
3. A pictorial diagram, sometimes called a layout diagram, shows the actual physical arrangement of the circuit elements on the circuit board or chassis, so that you can quickly find and identify components to test or replace.

When you troubleshoot an unfamiliar electronic circuit, you'll usually start with the block diagram to find where the trouble originates. Then you'll refer to the schematic diagram (or part of it) to find the faulty component in relation to other components in the circuit. A pictorial diagram can then tell you where the faulty component physically resides, so that you can test it and, if necessary, replace it.

Block diagrams

Block diagrams work well in conjunction with schematics to aid circuit comprehension and to streamline troubleshooting procedures. Each block represents all of the schematic symbols related to that part of the circuit. In addition, each block has a label that describes or names the circuit it represents. However, the block does nothing to explain the actual makeup of the circuit it represents. The blocks play a functional role only; they describe the circuit's purpose without depicting its actual components. Once you've gained a basic understanding of the circuit functions by looking at the block diagram, you can consult the schematic for more details.

To understand how you might use block diagrams, consider the following two examples.

First, suppose that you want to design an electronic device to perform a specific task. You can simplify matters by beginning with a block diagram that shows all of the circuits needed to complete the project. From that point, you can transform each block into a schematic diagram. Eventually, you'll end up with a complete schematic that replaces all of the blocks.

Alternatively, you can go at the task the other way around. Imagine that you have a complicated schematic, and you want to use it to troubleshoot a device. Because the schematic shows every single component, you might find it difficult to determine which part of the device has the problem. A block diagram can provide a clear understanding of how each part operates in conjunction with the others. Once you've found the troublesome area with the help of the block diagram, you can return to the schematic for more details.

Schematic diagrams

A schematic diagram acts, in effect, as a map of an electronic circuit, showing all of the individual components and how they interconnect with one another. According to one popular dictionary, the term *schematic* means "of or relating to a scheme; diagrammatic." Therefore, you can call any drawing that depicts a scheme—electronic, electrical, physiological, or whatever—a schematic diagram.

One of the most common schematic diagrams finds a place in almost every car or truck in the United States. It's a road map, of course, and it portrays a specific sort of scheme. The scheme might involve the paths of travel within a small town, within a state or province, or across multiple states or provinces. Like a schematic diagram of an electronic circuit, the road map shows all the components relevant to the scheme it addresses. Motorists make up their own schemes, which often comprise small portions of the total scheme included in the road map. Likewise, an electronic schematic shows all of the relevant components, and it allows a technician to extrapolate the components and interconnections when testing, troubleshooting, and repairing a small circuit, a large device, or a gigantic system.

Imagine that you want to travel in your automobile from point A to point B. Your road map lists all of the towns and cities that lie between these two points. By comparison, a schematic diagram lists all the components between a similar point A and point B in an electronic circuit. Nevertheless, both of these schematics indicate much more than mere points. You need to know more than which towns or cities lie between two fixed points to get an idea of the overall scheme of things. Indeed, you could easily write down the names of various towns or landmarks, in which case you would not have to resort to a road map at all. From an electronics standpoint, you could do the same thing by compiling a list of the components in a certain circuit, such as:

- 120-ohm resistor
- 1000-ohm resistor
- PNP transistor
- 0.47-microfarad capacitor
- 2 feet of hookup wire
- 1.5-volt battery
- Switch

This list tells us nothing about the circuit in a practical, put-together sense. We know all of the components that we would need if we wanted to build the thing, but we don't know what it would be if we did! In fact, these components might go together in two or three different ways to make two or three different circuits with different characteristics.

A schematic drawing must indicate not only all components necessary to make a specific scheme, but also how these components interrelate to one another. The road map connects various towns, cities, and other trip components with lines that represent streets and highways. A line that indicates a secondary road differs from a line that represents a four-lane highway. With practice, you can learn to tell immediately which types of lines indicate which types of roads. Likewise, an electronic schematic drawing uses a plain, straight line to indicate a standard conductor; other types of lines represent cables, logical pathways, shielding components, and wireless links. In all cases, when you draw the interconnecting lines, you draw them in order to indicate relationships between the connected components.

Schematic symbology

A schematic diagram reveals the scheme of a system by means of *symbology*. On a map, the lines that indicate roadways constitute symbols. But of course, a single black line that portrays Route 522 in no way resembles the actual appearance of this highway as we drive on it! We need know only the fact that the line symbolizes Route 522. We can make up the other details in our minds. If people always had to see pictorial drawings of highways on paper road maps, those maps would have to be thousands of times larger than those folded-up things we keep in our vehicle glove compartments, and they would be impossible for anybody to read.

Tip

Since the previous edition of this book was published in 1991, portable computers and the Internet have evolved so that, today, you actually can find and access road maps that show pictorial drawings of some roads and highways! You can look at photographs taken from satellites, aircraft, and sometimes even vehicles that have driven along specific routes. Check out "Google Maps," for example. These maps aren't on paper; instead, they reside in cyberspace. You need a computer or tablet device to use such "supermaps," but they do exist, and they're getting better every day.

On a decent road map, you'll find a key to the symbols used. The key shows each symbol and explains in plain language what each one means. If a small airplane drawn on the map indicates an airport and you know this fact, then each time you see the airplane symbol, you'll know that an airport exists at that particular site, as shown on the map. Symbology depicts a physical object (such as an airport outside a large city) in the form of another physical object (such as an airplane image on a piece of paper).

A good road map contains many different symbols. Each symbol is human engineered to appear logical to the human mind. For instance, when you see a miniature airplane on a road map, you'll reasonably suppose that this area has something to do with airplanes, so a detailed explanation should not be necessary. If, on the other hand, the map maker used a beer bottle to represent an airport, anyone who failed to read the key would probably think of a saloon or liquor store, not an airport! Because a map needs many different symbols, a good map maker will always take pains to make sure that the symbols make logical sense.

Pure logic will take us only up to a certain point in devising schemes to represent complicated things, especially when we get into the realm of electronic circuits and systems. For example, a circle forms the basis for a transistor symbol, a light-emitting-diode (LED) symbol, a vacuum tube symbol, and an electrical outlet symbol. Additional symbols inside the circle tell us which type of component it actually represents. A transistor is an active device, capable of producing an output signal of higher amplitude than the input signal. We can say the same thing about a vacuum tube, but not about an LED or an electrical outlet.

A circle with electrode symbols inside has been used for many years to represent a vacuum tube. Transistors were developed as active devices to take the places of vacuum tubes, so the schematic symbol for the transistor also started with a circle. Electrode symbols were inserted into this circle as before, but a transistor's elements differ from a tube's elements, so the transistor symbol has different markings inside the circle than the tube symbol does. The logic revolves around the circle symbol. Transistors accomplish many of the same functions in electronic circuits as vacuum tubes do (or did), so symbolically they are somewhat similar.

Inconsistencies arise in schematic symbology, and that's a bugaboo that makes electronics-related diagrams more sophisticated than road maps. A circle can make up a part of an electrical symbol for a device that doesn't resemble a tube or transistor at all. An LED, for example, can be portrayed as a circle with a diode symbol inside and a couple of arrows outside. An LED is not a transistor or tube, and the electrode symbol at the center clearly reveals this difference. An electrical outlet can serve as another example. It's absolutely nothing like a tube, transistor, or LED! Yet the basis for the symbol is a circle, just like the circle for a tube or transistor or LED. You'll learn more about specific schematic symbols in Chap. 3.

Schematic interconnections

To further explore how schematic diagrams are used, let's consider a single component, a PNP transistor. This device has three electrode elements, and although many different varieties of PNP transistors exist, we draw all their symbols in exactly the same way. We might find a PNP transistor in any one of thousands of different circuits! A good schematic will tell us how the transistor fits into the circuit, what other components work in conjunction with it, and which other circuit elements depend on it for proper operation. A transistor can act as a switch, an amplifier, an oscillator, or an impedance-matching device. A single, specific transistor can serve any one of these purposes. Therefore, if a transistor functions in one circuit as an amplifier, you can't say that the component will work as an amplifier only, and nothing else. You could pull this particular transistor out of the amplifier circuit and put it into another device to serve as the "heart" of an oscillator.

Tip

By knowing the type of component alone, you can't tell what role it plays in a circuit until you have a good schematic diagram showing all the components in the circuit, and how they all interconnect. Rarely can you get all this information in easy-to-read form by examining the physical hardware. You need a road map—a schematic diagram—to show you all the connections that the engineers and technicians made when they designed and built the circuit.

Suppose that you plan to drive your car from Baltimore, Maryland to Los Angeles, California. Even if you've made the trip several times in the past, you probably don't recall all of the routes that you'll need to take and all of the towns and cities that you'll pass along the way. A road map will give you an overall picture of the entire trip. Because all of the trip data exists in a form that you can scan at a glance, the road map plays a critical role in allowing you to see the entire trip rather than each and every segment, one at a time. A schematic diagram does the same thing for a "trip" through an electronic circuit.

Continuing with the road map and the coast-to-coast trip as an example, imagine that you have memorized the entire route from Baltimore to Los Angeles. Assume also that one of the prime highways on the way is under construction, forcing you to take an alternate route. Without a road map, you'll have no idea as to what detours exist, which alternate route is the best one to take, and which detour constitutes a path that will keep you on course as much as possible and eventually return you to the original travel route with a minimum of delay and inconvenience.

An electronic circuit has many electrical highways and byways. Occasionally, some of these routes break down, making it necessary to seek out the problem and correct it. Even if you can visualize the circuit in your head as it appears in physical existence, you'll find it impossible to keep in your "mind's eye" all the different routes that exist, one or more of which could prove defective. When I speak here of visualizing the circuit, I don't mean the schematic equivalent of the circuit, but the actual components and interconnections, known as the *hard wiring*.

A schematic diagram gives you an overall picture of a circuit and shows you how the various routes and components interact with other routes and components. When you can see how the overall circuit depends on each individual circuit leg and component, you can diagnose and repair the problem. Without such a view, you'll have to "shoot in the dark" if you want to get the circuit working again, and you'll just as likely introduce new trouble as get rid of the original problem!

Fear not!

Look at the schematic of Fig. 1-1. If you've had little or no experience with these types of diagrams, you might wonder how you'll ever manage to interpret it and follow the flow of electrical currents through the circuit that it represents. Fear not! By the time you finish this book, assuming that you already know some basic electricity and electronics principles, you'll wonder how you ever could have let a diagram like this intimidate you. By the way, you'll see this diagram again in Chap. 5.

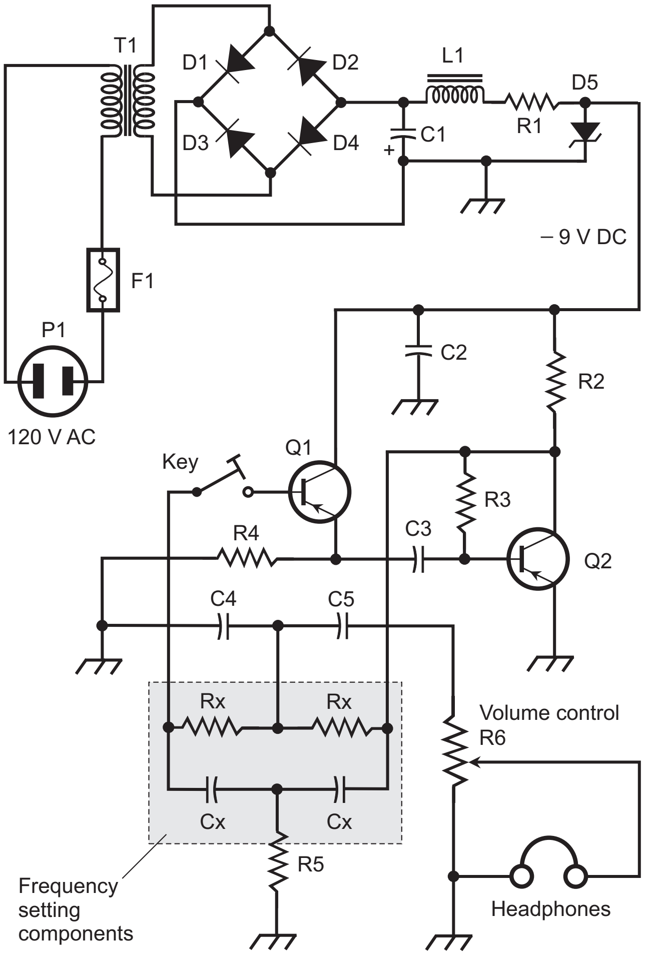
T1
D1
D2
L1
D5
D3
D4
C1
+
R1
– 9 V DC
F1
P1
120 V AC
C2
R2
Q1
Key
R3
R4
C3
Q2
C4
C5
Rx
Rx
Volume control
R6
Cx
Cx
R5
Frequency
setting
components
Headphones

FIG. 1-1. *An example of a fairly complicated schematic diagram. By the time you finish this book, you'll find it simple!*

Visual language

It can prove difficult to explain schematic diagramming in detail to people who have just begun their study of electronics. It helps to think of this form of symbology as a language, that is to say, a system of symbols that helps us to communicate ideas. The English language (or whatever your native language happens to be) is a scheme with a system of symbology with which you're familiar, of course, because you're reading this book!

Every word spoken in English or any other verbal language is a complex symbol made from simpler elements called *characters*. Let's take the word "stop," for example. Without a reference key, this sound means nothing. A newborn infant hears noise coming out of your mouth, that's all! However, through learning the symbology from shortly after birth, this word begins to mean something because the child, who has begun learning to speak and understand, can compare "stop" to other words, and also to actions. We can even say that the word "stop" is a sort of symbology within symbology. The communicator's intent, when using the word "stop," can also be expressed by the phrase "Do not proceed further." This phrase also constitutes symbology, expressing a mental image of a desired action.

If we could all communicate by mental telepathy, then we wouldn't need language or the symbols that it comprises. Thinking happens a lot faster than speaking or writing or reading can go; and brain processes are the same from human to human, regardless of what language any particular individual employs when speaking, writing, or reading. A newborn baby speaks and understands no language whatsoever. However, whether that baby was born in the United States, South Africa, Asia, or wherever, thought processes take place.

The baby knows when it is hungry, in pain, frightened, or happy. It needs no language to comprehend these states. But the baby does have to communicate right from the start. For this reason, all newborns communicate in the same language (crying and laughing, mostly). As newborns comprehend more and more of their environment through improved sensory equipment (eyes, ears, nose, fingers), they collect more and more data. At this point the various languages come into play, with different societies using different verbal symbols to express mental processes. The human brain still carries on the same nonlinguistic thought processes as before, however, because thinking in terms of symbols would take far too much time and "brain storage."

The brain helps a human to transpose complex thoughts into a language, and vice versa, just as a computer translates programming languages into electronic impulses and vice versa. Imagine that a child is about to step in front of a speeding automobile. If the brain had to handle millions of data elements symbolically, we humans would spend all of our lives waiting for our brains to deliver the correct processed information, and that child would probably get killed before we could even begin to react. Rather, the brain scans all the data received by the sensory organs in real time and then sums it up into a single symbol for communication. A good audible (and hopefully loud) symbol in the above-mentioned case is "Stop!" You, seeing a child about to go out into heavy traffic, might shout that word and produce in the child's brain the appropriate sequence of processes.

Not all languages involve the spoken word. Have you heard of sign language, whereby the arms and hands are used to communicate ideas? If you've done any amateur (or "ham") radio communication, especially if you got your "ham" license back in the time when I got mine (the 1960s), you know about the Morse code as a set of communications symbols. In most instances, an entire communicating language of visual symbols is not as efficient for us humans as one composed of words and visual symbols combined. Using the symbol "stop" again, we can utter this word in many different ways. The word in itself means something, but the way we say it (our "tone of voice") augments the meaning. We can't do all that with the printed or displayed characters S, T, O, and P all by themselves in plain text.

We humans have arrived at some universal methods of modifying visual symbols. To many of us, the color red denotes something that demands immediate attention. Often, however, this color is used in conjunction with the visual symbol for a spoken word. Think of a "stop" sign, for example. It's red, right? Or think of a "yield" sign. It's yellow, representing something that demands attention, but in a less forceful way than the color red does.

Tip

Schematic diagrams rarely include color. Look in the back of a technical manual for an amateur radio transceiver, for example. Does the schematic have color? I'll bet you that it doesn't. (A few good magazines, however, do put color into their schematics.) A technical manual's schematic might not even have grayscale shading. Schematics resemble printed text or Morse code in this respect; we must convey a lot of information with a limited set of symbols, and we're constrained even as to the way in which we can portray and read those symbols.

Schematics don't lend themselves to any form of oral (audible) symbology either. When you see the symbol for, say, a field-effect transistor (FET) in a schematic diagram, you don't hear the paper or computer say, "Field-effect transistor, for heaven's sake, not bipolar transistor!" You have to make sure that you read the symbol correctly. If you want to build the circuit and you mistakenly put a bipolar transistor where an FET should go—maybe because you didn't look carefully enough at the schematic—you have no right to expect that the final device or system will work. Something might even burn out, so that when you recognize your error and replace the FET with a bipolar transistor, you'll have to troubleshoot the whole circuit before you can use it!

Our senses along with our central processor, the brain, render us less than proficient at mentally conceiving all of the workings of electronic circuits by dealing with them directly. Therefore, we have to accept data a small step at a time, compiling it in hardcopy form (through symbology) and providing hardcopy readout. We can liken this method to the "connect-the-dots" drawings in children's workbooks. Individually, the dots mean nothing, but once they are arranged in logical form and connected by lines, we get an overall picture. The dots' relationships to each other and the order in which they are connected tell us everything that we need to know.

The remaining chapters in this book start with the symbols for individual electronic components, then move on to simple circuits, and finally show you a few rather complicated circuits. Schematic symbols and diagrams are designed for human beings, so human logic constitutes a prime factor in determining which symbols mean what. In that respect, the creation and reading of schematic diagrams resembles mathematics, and in particular, good old-fashioned plane geometry!

Tip

Schematic diagrams are *encoded* representations of circuits, while pictorials show us the physical objects, often proportioned according to their relative size, and sometimes rendered so as to look three-dimensional by means of shading and perspective. Schematic diagrams depict circuit components as symbols only, without regard to their real-world size or shape, and in two dimensions (a flat piece of paper or computer screen), completely lacking depth or perspective.

This page intentionally left blank

2

Block diagrams

A block diagram portrays the general construction of an electronic device or system. A block diagram can also provide a simplified version of a circuit by separating the main parts and showing you how they are interconnected.

A simple example

Figure 2-1 is a block diagram of a device that converts alternating current (AC) to direct current (DC). The terminal at the left accepts the AC input. In sequence going from left to right, the electricity passes through the transformer, the rectifier, and the filter before arriving at the output as DC. In this case, the lines that connect the blocks do not have arrows because readers will naturally assume that the flow goes from left to right. The input terminal resides at the left-hand end of the diagram, and the output lies on the extreme right. In more complicated block diagrams, the interconnecting lines may include arrows to show which block affects which, or to indicate the general direction of signal flow when it might not otherwise be clear.

Functional drawings

Engineers and technicians employ block diagrams in various ways. Commonly, block diagrams indicate the interconnections between

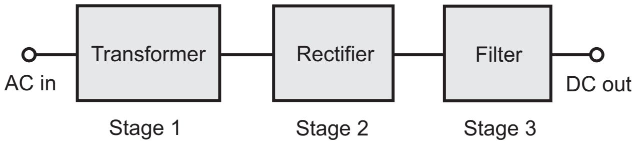

FIG. 2-1. *Block diagram of an AC-to-DC converter. The electricity flows from left to right.*

small circuits in a larger device, or between diverse devices in a large system. When drawn as shown in Fig. 2-1, block diagrams can also be called *functional diagrams* because they reveal the basic functioning of the electronic circuit. The functional diagram offers a simplistic explanation of how the device operates; it can lead to more detailed information provided by a schematic diagram.

Someone who wants to draw a schematic diagram for a complex electronic circuit designed from scratch can start with a block diagram. This diagram will show all of the circuit sections (stages) needed to arrive at a functioning device, but none of the internal details of those stages. Then the designer will develop schematic diagrams of circuits that can fill each block and serve the appropriate function in the overall system. The first block in the diagram will then be replaced by the schematic diagram of the circuit it represents. The engineer or technician will move through the blocks according to functional order, creating schematic diagrams that can be used to build each stage in the system. As soon as the final block has been filled in with a schematic for the applicable stage, the comprehensive schematic is complete, and a total (but so far only theoretical) system design is portrayed in detail.

Another way of using block diagrams starts with a finished schematic diagram. Imagine that the schematic is complicated, and that the equipment whose circuit it represents does not work properly. Although schematic diagrams can describe the functioning of an electronic circuit, they are not as clear and basic as a functional block diagram for that purpose. In the absence of a preexisting block diagram, a technician would have to start with the schematic, laboriously identify each stage in the system, and then draw the entire system diagram in block form. When finished, the block diagram would reveal how each stage interacts with the others. Using this method, one or more stages could be identified as a possible trouble area. Then the technician would refer to the original schematic and conduct tests in specific areas, based on his or her knowledge of how each stage works at the component level.

Tip

In practice, you'll often encounter block diagrams. If presented without accompanying schematics, a block diagram describes the basic functional operation of an electronic device or system. The block diagram can prove most useful when you don't need to know the functions of individual components.

We can describe the operation of a specific type of wireless transmitter, say an *amplitude-modulated* (AM) voice transmitter, such as the type found in Citizens Band (CB) radios, by means of a block diagram. This diagram will apply to most other AM voice radio transmitters. Of course, no two transmitters built by different manufacturers are exactly alike, but all of them contain the same basic circuit sections as far as functionality goes. One type of oscillator might work differently from another type, but they all do the same thing: generate a radio-frequency (RF) signal! When we need to know, or portray, individual differences between circuits that do essentially the same things, then we need schematic diagrams.

The block diagram in Fig. 2-2 illustrates the various parts of a strobe light circuit. Let's go through the diagram block by block to understand how it works. The input signal enters at the left; it's utility AC,

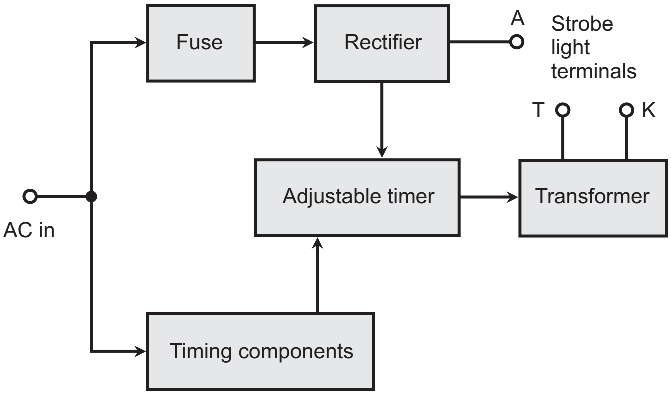

FIG. 2-2. *Block diagram of a circuit designed to provide power to a strobe light. Arrows show the direction of elecricity flow.*

such as we get from a standard wall outlet. In the United States and some other countries, this AC has a nominal voltage of 117 volts (V) and a frequency of 60 hertz (Hz), where "hertz" means "cycles per second." (In some countries, the voltage is about 234 V, and in some countries, you'll find a frequency of 50 Hz rather than 60 Hz.) The input AC goes to a fuse, and also to a combination of components that provide timing. The top path, where the fuse is located, leads to a diode-type rectifier, and the rectifier output passes directly to one terminal of the three-terminal strobe lamp. The rectifier also outputs to an adjuster that provides a variable flash rate for the lamp. The output from that adjuster goes to a transformer, which provides the remaining two outputs required to operate the lamp.

Current and signal paths

Figure 2-3 shows a power supply that produces several different voltage outputs. As you go through this diagram from left (the input) to the bottom and the right (the outputs), note that the circuit is powered with 120 volts AC (120 VAC), quite close to the nominal 117 VAC commonly found at utility outlets in the United States. The input

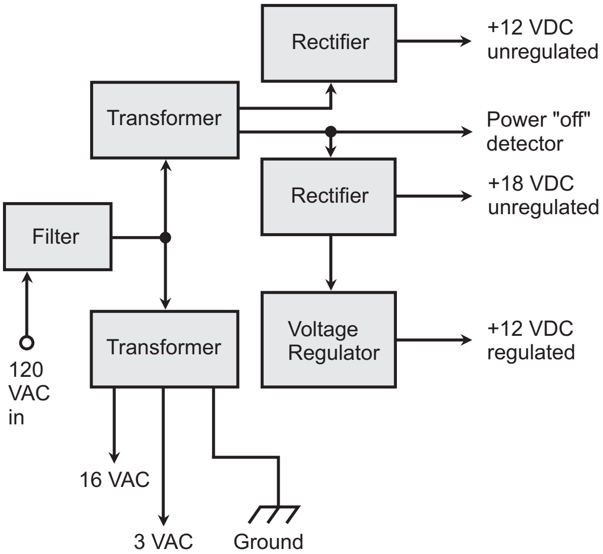
Rectifier
+12 VDC
unregulated
Transformer
Power "off"
detector
Rectifier
+18 VDC
unregulated
Filter
Transformer
Voltage
Regulator
+12 VDC
regulated
120
VAC
in
16 VAC
3 VAC
Ground

FIG. 2-3. *Block diagram of a power supply with several different outputs.*

AC goes through a filter and then splits into two paths. Part of the AC goes to the "lower" transformer that provides 16 VAC and 3 VAC output, along with a ground connection.

From the filter, the input voltage gets fed to another transformer that derives the voltages to be converted to DC electricity. One output of the transformer goes to a rectifier that provides 12 volts DC (12 VDC) without any voltage regulation. The other transformer output goes to a separate rectifier that provides 18 VDC, also unregulated. This transformer output also serves as a diagnostic detector for a power "off" condition. That line is further tapped to join with the output of the voltage regulator to provide 12 VDC with voltage regulation.

Tip

Block diagrams are comparatively easy to draw, comprising squares or rectangles along with interconnecting lines (sometimes with arrows). More sophisticated block diagrams also include triangles to represent circuit blocks built around specialized amplifiers constructed within *integrated circuits* (ICs), also known as *chips*.

Figure 2-4 is a block diagram of an AM radio transmitter. The microphone preamplifier stage goes to the input of the audio amplifier stage

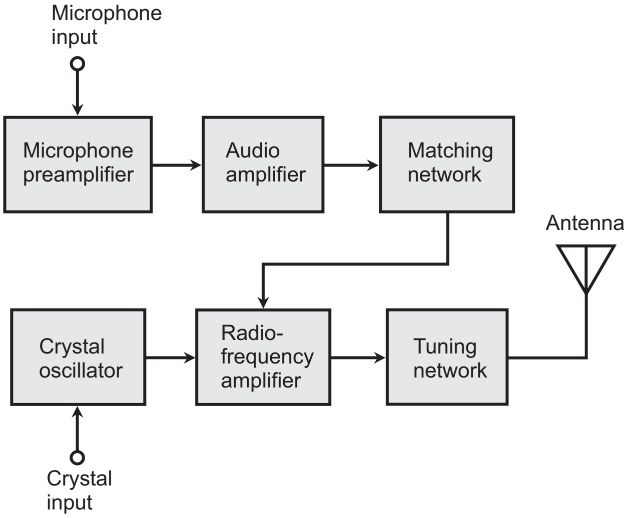
Microphone input
Microphone preamplifier
Audio amplifier
Matching network
Antenna
Crystal oscillator
Radio-frequency amplifier
Tuning network
Crystal input

FIG. 2-4. *Block diagram of an amplitude-modulated (AM) radio transmitter.*

(note the direction of the arrow). The output of the audio amplifier goes to the matching network, which in turn goes to the RF amplifier section. The crystal oscillator is also connected to the RF amplifier section, whose output leads into the RF tuning network. Only one connection exists between the audio section of the circuit and the RF section: the one between the matching network and the RF amplifier. This block diagram, with its arrows, tells us not only how the components of the system connect to one another, but also the sequence of events or direction of signal flow.

Flowcharts

Block diagrams can describe the functioning of electronic circuits, but in the world of computers, another form of diagramming is sometimes used to portray the functioning of a program. This system is called *flowcharting*. A flowchart resembles a block diagram, except that the symbology applies to the sections of a computer program, an intangible thing (as opposed to an electronic circuit, a tangible thing). A flowchart provides a graphic representation of the logical paths that a computer will take as it executes a particular program. Flowcharts are often prepared in conjunction with specifications, and are modified as the requirements change to fit within the constraints of the computer system.

For complex problems, a formal written specification might be necessary to ensure that everyone involved understands and agrees on what the problem is, and on what the results of the program should be. To illustrate this concept, let's suppose that a teacher wants to write a computer program that will determine a student's final grade for a course by calculating an average from grades the student has received over a certain period of time. The teacher will supply the grades to the program as input. Only the average grade is needed as an output. Now, we can make an orderly list of what the program has to do:

- Input the individual grades.
- Add the grade values together to find their sum.
- Divide the sum by the number of grades to find the average grade.
- Print out the average grade.

We can prepare a flowchart of the program, as shown in Fig. 2-5. As we can see, the flowchart graphically presents the structure of the program, revealing the relationship between the steps and paths. When the flow of control is complicated by many different paths that result from many decisions, a good flowchart can help the programmer sort things out. The flowchart can serve as a thinking-out tool to understand the problem and to aid in program design. The flowchart symbols have English narrative descriptions rather than programming language statements because we want to describe *what* happens, not

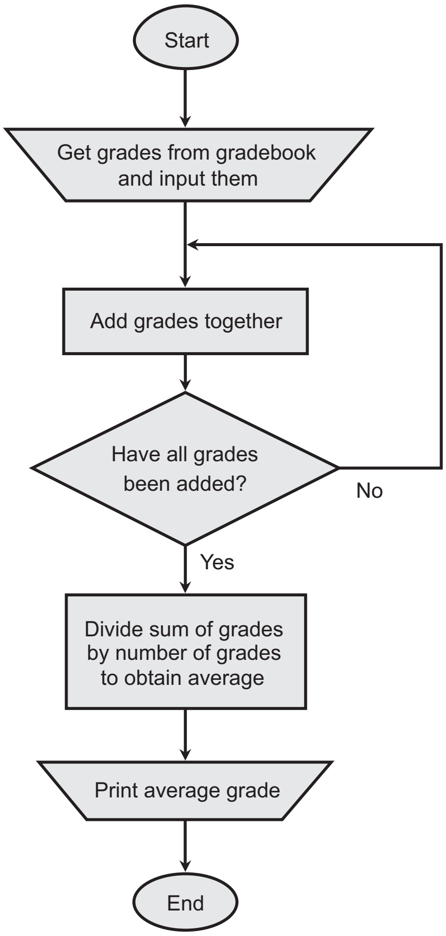
Start
Get grades from gradebook and input them
Add grades together
Have all grades been added?
No
Yes
Divide sum of grades by number of grades to obtain average
Print average grade
End

FIG. 2-5. *Example of a program flowchart.*

how it happens. At a later stage, if formal flowcharts are required for documentation, the flowchart can contain statements in a programming language. These flowcharts might prove helpful to another person who at some future time wants to understand the program.

It takes a lot of time to conceive and draw up a formal flowchart, and modifying a flowchart to incorporate changes, once a program has been written and its flowchart composed, can prove difficult. Because of these limitations, some programmers will shy away from the use of a flowchart, but for others, it can provide valuable assistance in understanding a program. In order to promote uniformity in flowcharts, standard symbols have been adopted, the most common of which are shown and defined in Fig. 2-6.

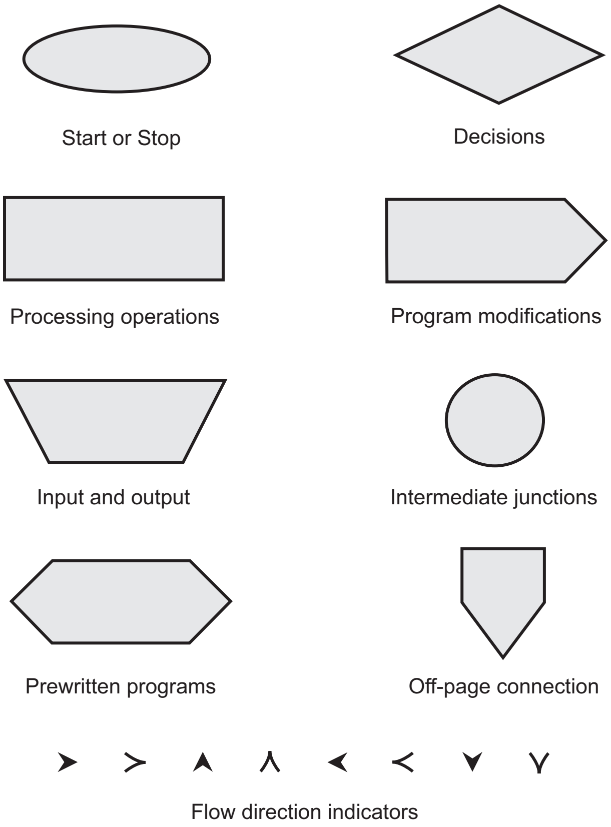
Start or Stop
Decisions
Processing operations
Program modifications
Input and output
Intermediate junctions
Prewritten programs
Off-page connection
Flow direction indicators

FIG. 2-6. *Common symbols for flowcharts intended to represent computer programs.*

Follow the flow

The normal direction of processes in a flowchart runs from top to bottom and from left to right, the same way as people read books in most of the world. Arrowheads on flow lines indicate direction. The arrows can be omitted if, but only if, the direction of flow is obvious without them.

Figure 2-7 is a flowchart for a program that duplicates punched cards, and at the same time prints the data on each card. Keep in mind that this particular "beast" is of historical interest only! (Were you born long enough ago to remember punch cards for inputting programs to computers? I recall using them, all the way back in the 1970s, when I attended the University of Minnesota. I guess that little factoid dates me, doesn't it?)

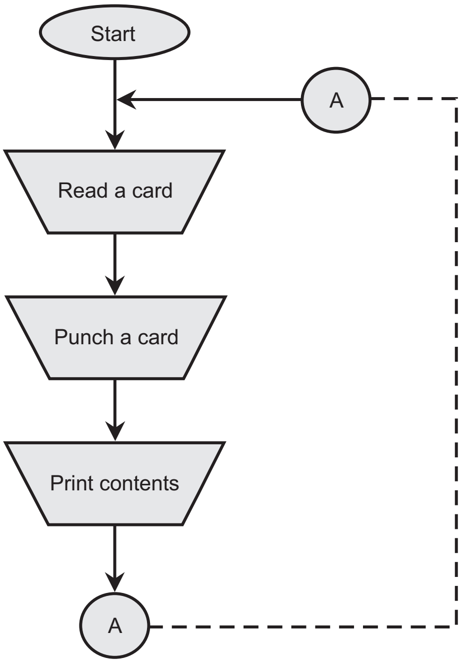

FIG. 2-7. *Flowchart for a program for duplicating punched cards. The circles labeled A represent the inflow and outflow points in the feedback loop shown by the dashed line.*

Let's trace the flow of the program through Fig. 2-7. The program begins at the "Start" oval at the top and proceeds in the direction of the arrows. In the first box below "Start," the program reads a card. Then the program punches the card's contents (data) as holes in a blank piece of heavy paper and sends the data to a printer. The program then goes back along the dashed line to the top and reads the next card. The circles marked A represent inflow and outflow points. In this case, they're superfluous, but in complicated flowcharts, they can be useful when it would create a mess to include all the applicable dashed lines. The program repeats itself as long as it has cards to read and punch.

In a sophisticated flowchart, we might see several different symbols of the sort shown in Fig. 2-6, and maybe even all of them. Oval boxes show start or stop points. Arithmetic operations go in rectangular boxes. Input and output instructions go in upside-down trapezoids. If we want to show a program that someone wrote earlier within the context of a larger flowchart, we don't necessarily have to draw the flowchart for the inside program. Rather, we might represent the entire program as a flattened hexagon. If a box indicates a decision, we use a diamond shape. A five-sided box portrays a part of the program that changes itself. A small circle identifies a processing junction point. Such a point in the program can go to several places. A small five-sided box, which has the shape of the home plate on a baseball field, shows where one page of a flowchart connects to the next, if the entire flowchart has more than one page. The intermediate junction and off-page connection points are labeled with numbers and letters to let readers know that all like symbols with the same character inside are meant to be connected together. Arrows indicate the direction of the flow.

Process paths

Returning to the flowchart for duplicating punched cards (Fig. 2-7), suppose that you want to change the card-punching program so that the computer skips blank cards and duplicates only those cards with some holes in them. Because the computer must make a decision about each card, you'll need to include a decision block in the flowchart. Figure 2-8 shows the result.

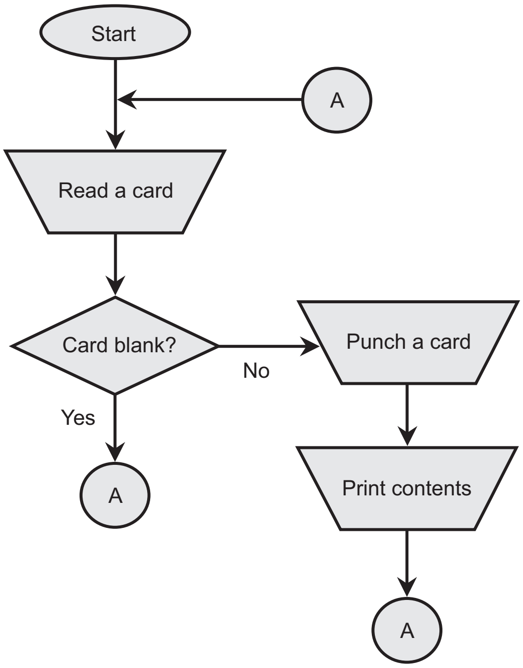

FIG. 2-8. *Example of a flowchart that includes a decision block (the diamond). The circles labeled A all represent a single junction point through which data flows in the directions shown by the arrows.*

Follow the flow

Except for the decision block, Fig. 2-8 shows the same process as Fig. 2-7 does. The program begins in the "Start" oval at the top and then goes to the block marked "Read a card." From there, the program moves on to the decision block labeled "Card blank?" If the answer is "Yes," the program proceeds to the connection circle marked "A" and back to the top to read the next card. If the answer is "No" (the card has holes in it), the program instructs the *hardware* (the physical components of the computer) to punch a duplicate card and print its contents. Then the program goes to another circle marked "A" and back to the starting point.

Tip

Figure 2-8 is a simple flowchart, showing a process that uses only input and output devices and that does no calculations. Most programs and flowcharts involve more complicated processes.

The field of microcomputers uses many different types of diagrams that deal mostly with *software* (the operating systems and programs) rather than hardware (the physical components). From a purely electronic standpoint, functional diagrams abound and are usually more numerous than the schematic diagrams in the computer world. From an understanding standpoint, block diagrams can serve to display machine functions in general, but hardware maintenance and repair procedures require well-defined schematic drawings. Computers take advantage of the latest state-of-the-art developments in electronic components and are relatively simple from this standpoint, especially when you consider all they can do. However, from a pure electronics standpoint and as far as schematic diagrams are concerned, computers are highly complex; it would take many pages of schematics to represent even the most rudimentary computer.

Summary

Block diagramming can help you understand the general functioning of electronic circuits. Block diagrams are easy to draw, usually requiring only a marking instrument, some paper, and a straightedge (or a vector graphics computer program and a little bit of training on it). Schematic diagrams, in contrast, need more tools and can, in some cases, take many hours to render in a form that people can easily read and interpret.

3

Component symbols

On a road map, symbols illustrate towns, cities, secondary roads, primary roads, airports, railroad tracks, and geographical landmarks. The same rule applies to schematic drawings; symbols indicate conductors, resistors, capacitors, solid-state components, and other electronic parts. Every time a new component comes out, a new schematic symbol is derived for it. Often, a new type of component is a modification of one that already exists, so the new schematic symbol ends up as a modification of the symbol for the preexisting component.

Tip

In this chapter, you'll find most of the schematic component symbols commonly used in electricity and electronics. Appendix A contains a more complete listing in alphabetic, tabular form. You'll find it in the back of this book.

Resistors

Resistors are among the most simple electronic components. As the term implies, they resist the flow of electrical current. The value or "size" of a resistor is measured in units called *ohms*; typical real-world resistors are rated from about one ohm up to millions of ohms. Less commonly, you'll encounter resistors with values of less than an ohm, or hundreds of millions (or even billions) of ohms.

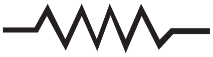

FIG. 3-1. *Standard schematic symbol for a fixed-value resistor.*

Regardless of the ohmic value, all fixed-value resistors are schematically indicated by the symbol shown in Fig. 3-1. This is the most universally accepted symbol for a resistor. The two horizontal lines indicate the leads or conductors that exit from both ends of the physical component. (Sometimes the resistor contacts are not wire leads but more substantial metal terminals.) Figure 3-2 shows a "transparent" functional drawing of a *carbon-composition* fixed resistor with leads on both ends. Figure 3-3 shows pictorial drawings of two other types of resistors. Any resistor of the sort shown pictorially in Fig. 3-2 or Fig. 3-3 is indicated schematically by the symbol in Fig. 3-1.

A variable resistor has the ability to change ohmic value by means of a slide or rotary tap that can be moved along the resistive element. The variable resistor is usually set to one value, and it remains at this point until manually changed. The electronic circuit, therefore, "sees" the component as a fixed resistor. However, when a variable resistor is required for the proper functioning of a specific circuit, it is necessary to indicate to any person who might build it from a schematic drawing that the resistor is actually a variable type. Figure 3-4 shows the schematic symbol for a variable resistor with two leads. Other types of variable resistors exist, and they have three leads (two end leads and a tap). Figure 3-5 shows two examples of schematic symbols for a three-terminal variable resistor, known as a *potentiometer* or a *rheostat* depending on the method of construction. Notice that

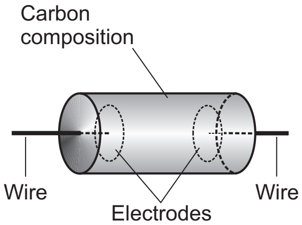

FIG. 3-2. *Pictorial illustration showing the anatomy of a carbon-composition resistor.*

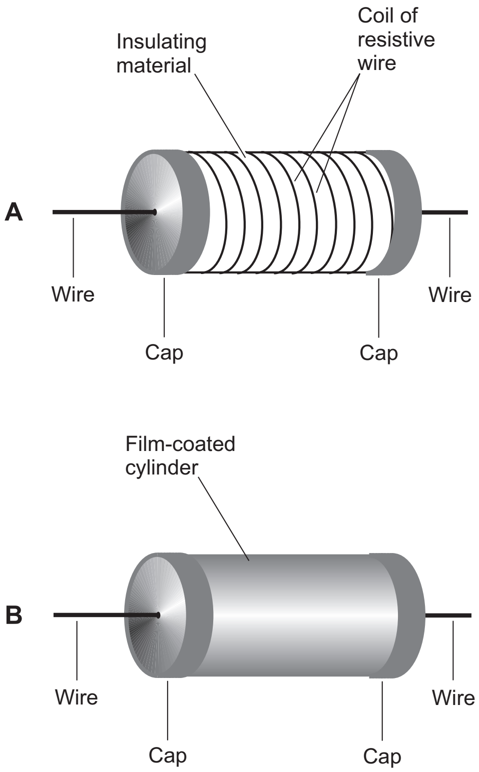
Insulating material
Coil of resistive wire
A
Wire
Wire
Cap
Cap
Film-coated cylinder
B
Wire
Wire
Cap
Cap

FIG. 3-3. *Pictorial illustrations showing the anatomy of a wirewound resistor (at A) and a film type resistor (at B).*

both examples use the standard resistor configuration, and indicate that it's a variable type by means of an arrow symbol pointing to the zig-zag part.

Did you know?

Rheostats are in effect the same as potentiometers, but mechanically they differ. A rheostat contains a wirewound resistance element, while a potentiometer is normally of the carbon-composition type. Therefore, a rheostat's value varies in small increments or steps, while a potentiometer's value can be adjusted over a continuous range.

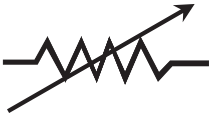

FIG. 3-4. *Schematic symbol for a two-terminal variable resistor.*

Tip

In schematic drawings, an arrow often indicates variable properties of a component, but not always! Transistors, diodes, and some other solid-state devices have arrows in their schematic symbols. These arrows don't have anything to do with variable or adjustable properties. Arrows can also sometimes indicate the direction of current or signal flow in complex circuits.

Figure 3-6 is a pictorial drawing of a variable resistor of the wire-wound type, manufactured so that the resistance wire is exposed. A sliding metallic collar, which goes around the body of the resistor, can be adjusted to intercept different points along the coil of resistance wire. The collar is attached by a flexible conductor to one of the two end leads. The collar, therefore, shorts out more or less of the coil turns, depending on where it rests along the length of the coil. As the collar moves toward the opposite resistor lead, the ohmic value of the component decreases.

Figure 3-7 shows a functional drawing of a rotary potentiometer (at A), along with the schematic symbol (at B). The symbol looks like

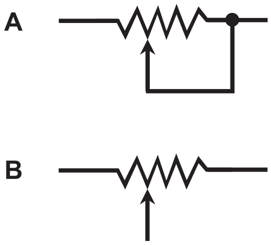

FIG. 3-5. *Alternate symbols for variable resistors, also known as potentiometers or rheostats (depending on the physical construction method). The device at A connects one end to the tap; and the device at B uses a three-terminal arrangement.*

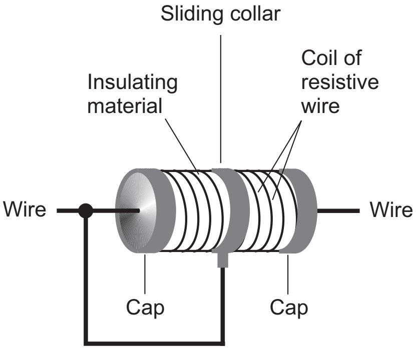
Sliding collar
Insulating material
Coil of resistive wire
Wire
Wire
Cap
Cap

FIG. 3-6. *Pictorial drawing of a wirewound variable resistor.*

the variable resistor equivalent, but has three discrete contact points. Using the potentiometer control, the portion of the circuit that comes off the arrow lead can be varied in resistance to two circuit points, each connected to the two remaining control leads. Figure 3-8 shows a pictorial drawing of a typical potentiometer.

The variable resistor shown pictorially in Fig. 3-6 can be changed into a rheostat by severing the connection between the collar and

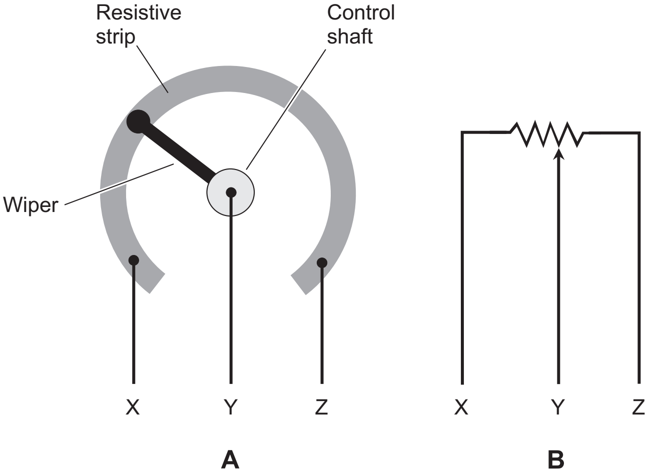
Resistive strip
Control shaft
Wiper
X
Y
Z
X
Y
Z
A
B

FIG. 3-7. *Simplified functional drawing of a rotary potentiometer (A) and its schematic symbol with corresponding connections (B).*

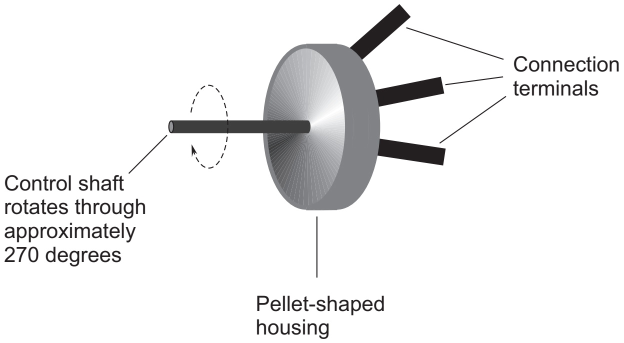

FIG. 3-8. *Pictorial drawing of a full-size potentiometer, suitable for mounting on the front panel of an electronic device such as a radio receiver.*

the end. Now, the collar can be used as the third or variable contact. Likewise, a rheostat or potentiometer can be turned into a two-lead variable resistor by shorting out the variable contact point with the lead on either end.

The schematic symbol for a resistor, all by itself, tells us nothing about the ohmic value, or anything else about the component such as its power rating or physical construction, either. Various specifications for the component can be written alongside the resistor symbol, but these details might also appear in a separate components table and referenced by an alphabetic/numeric designation printed next to the schematic symbol (such as R1, R2, R3, and so on).

Tip

You can usually determine the ohmic value of a fixed resistor by looking at the colored bands or zones on it. Appendix B lists the resistor color codes that specify the ohmic values of fixed resistors.

Capacitors

Capacitors are electronic components that have the ability to block direct current (DC), while passing alternating current (AC). They also store electrical energy. The basic unit of capacitance is the *farad* (symbolized F). The farad is a huge electrical quantity, and most

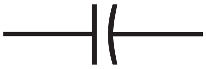

FIG. 3-9. *Standard symbol for a fixed capacitor. The curved line represents the plate (or set of plates) electrically closer to ground.*

real-world components are, therefore, rated in tiny fractions of a farad—*microfarads* or *picofarads*. A microfarad (symbolized μF) equals a millionth of a farad (0.000001 F), and a picofarad (symbolized pF) equals a millionth of a microfarad (0.000001 μF) or a trillionth of a farad (0.000000000001 F).

Figure 3-9 shows the common schematic symbol for a fixed capacitor. On occasion, you might see alternative symbols, such as those in Fig. 3-10. Many different types of capacitors exist. Some are *nonpolarized* devices, meaning that you can connect them in either direction and it doesn't make any difference. Others are *polarized*, having a positive and a negative terminal, and you must take care to connect them so that any DC voltage that happens to appear across them has the correct polarity. Most types of capacitors contain only two leads, although every now and then, you'll come across one with three or more leads.

The basic capacitor symbol consists of a vertical line followed by a space and then a parenthesis-like symbol. Horizontal lines connect to the centers of the vertical line and the parenthesis to indicate the component leads. The parenthesis side of a capacitor indicates the lead that should go to electrical ground, or to the circuit point more nearly connected to electrical ground. Unless the symbol includes a polarity sign, it indicates a nonpolarized capacitor, which might be made from metal plates surrounding ceramic, mica, glass, paper, or other solid nonconducting material (and, in some cases, air or a vacuum). The material designation indicates the insulation, technically known as a *dielectric*, that separates the two major parts of the component.

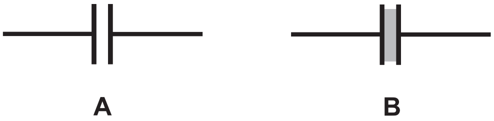

FIG. 3-10. *Alternate symbols for fixed capacitors. At A, air dielectric; at B, solid dielectric.*

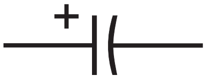

FIG. 3-11. *Schematic symbol for a polarized capacitor. The side with the plus sign (+) should carry a positive voltage relative to the other side.*

Physically, a typical fixed-value capacitor comprises two tiny sheets of conductive material close to each other but kept electrically separated by the dielectric layer.

Figure 3-11 shows the schematic symbol for a polarized or electrolytic capacitor. Notice that this symbol is the same as the one for the nonpolarized component, but a plus (+) sign has been added to one side. This sign indicates that the positive terminal of the component goes to the external circuitry. Occasionally, a negative (–) symbol will also appear on the opposite side. When you see the plus sign, you know that the component is polarized, and therefore, that you must connect it to the remainder of the circuit in observance of the proper polarity. That means the positive capacitor electrode must go to the more positive DC voltage point in the circuit, and the other electrode must go to the more negative DC voltage point in the circuit.

Tip

Polarized capacitors have external markings that tell you the polarity. Some have a plus sign, and some have a minus sign, and a few have both. Often, you'll need a magnifying glass to resolve the symbols, so beware: You should never connect a polarized capacitor the wrong way around!

All the capacitors that we've seen so far have a fixed design. In other words, the components specified have no provision for changing the capacitance value, which is determined at the time of manufacture. Some capacitors, however, do have the ability to change value. These components are generally called *variable capacitors*, although some specialized types are known as *trimmer capacitors* or *padder capacitors*.

Figure 3-12 shows the most common symbol for a variable capacitor. An arrowed line reveals the variable property; it runs diagonally through a fixed capacitor symbol. Figure 3-13 shows two alternative

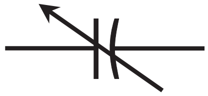

FIG. 3-12. *Standard symbol for a variable capacitor. The curved line represents the rotor, and the straight line represents the stator.*

ways of indicating this same component. Most of the time, the symbol shown in Fig. 3-12 will indicate a variable capacitance, regardless of the physical construction details.

An *air variable capacitor* (one with an air dielectric) can tune many types of radio-frequency (RF) equipment including antenna matching networks, transmitter output circuits, and old-fashioned radios. A typical air variable has many interlaced plates, with the plates connected together alternately to form two distinct contact points. The set of plates that you can rotate is called the *rotor*; the set of plates that remains stationary is called the *stator*. All variable capacitors are nonpolarized components, meaning that the external DC voltage you connect to them can go either way and it doesn't make any difference.

Tip

In most air variables, the rotor should go to electrical ground. The rotor connects physically to the shaft that you turn. By grounding that shaft along with the rotor, you minimize *external capacitance* effects so that if you touch the shaft, the addition of your body into the system doesn't upset the performance of the circuit. In addition, you're protected against the risk of electric shock when the shaft that you touch goes directly to ground!

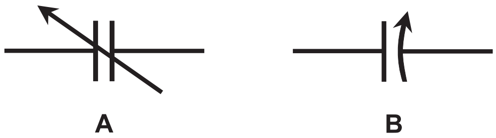

FIG. 3-13. *Alternate symbols for variable capacitors. At A, the stator is not distinguished from the rotor; at B, the rotor appears as a curved line with an arrow.*

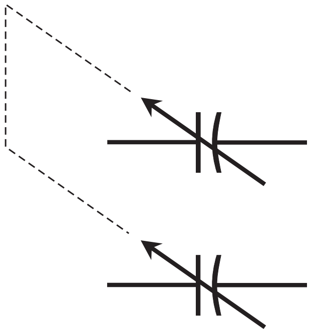

FIG. 3-14. *Schematic symbol for two variable capacitors ganged together.*

Sometimes, two separate variable capacitors are connected together or *ganged* in a circuit. In a ganged arrangement, two or more units are used to control two or more electronic circuits, but both components are varied simultaneously by tying the rotors of the two units together. Figure 3-14 shows the schematic symbol for two variable capacitors ganged together. The minimum and maximum capacitance values of the two components might be the same, but they don't have to be the same. They will, however, always track together. In a ganged system, when one of the capacitors increases in value, the others all increase as well.

As is the case with most electronic components, the schematic symbol for the capacitor serves only to identify it and to show whether it is fixed or variable, and if fixed, whether or not it is polarized. The component value might be written alongside the schematic symbol, or the component might be given a letter and number designation (for example, C1, C2, C3, and so on) for reference to a components list or table that goes along with the diagram.

Inductors and transformers

A basic *inductor* comprises a length of wire that is coiled up in order to introduce *inductance* into a circuit. Inductance is the property that opposes change in existing current; it acts in practice only while current increases or decreases. Coils or inductors can range in physical size from microscopic to gigantic, depending upon the inductance

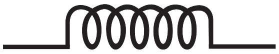

FIG. 3-15. *Standard symbol for an air-wound (or air-core) inductor.*

value of the component, and on the amount of current that it can handle.

The basic unit of inductance is the *henry* (symbolized H), a large electrical quantity. Most practical inductors are rated in *millihenrys* (symbolized mH), where 1 mH = 0.001 H, or in *microhenrys* (μH), where 1 μH = 0.001 mH = 0.000001 H. Occasionally, you'll see an inductor whose value is specified in *nanohenry*s (nH), where 1 nH = 0.001 μH = 0.000000001 H.

Figure 3-15 shows the basic schematic symbol for an *air-core inductor*. The two leads are designated by straight lines that merge into the coiled part. An air-core coil has nothing inside the windings that can affect the inductance. Some air-core coils are wound from stiff wire and support themselves mechanically, and their cores do, in fact, comprise nothing but air. In most cases, however, a nonconductive and noninductive form made out of plastic, mica, or ceramic material serves as a support for the coil turns, keeping them in place and enhancing the physical ruggedness of the component.

Did you know?

In some old radio receivers, you'll find air-core inductors wound around small waxed cardboard cylinders resembling short lengths of drinking straw. Some hobbyists even use waxed wooden dowels to support "air-core" coils!

Figure 3-16 shows the schematic symbol for a tapped air-core inductor; in this case, the coil has two tap points along its length.

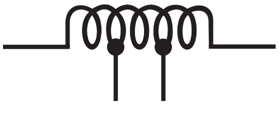

FIG. 3-16. *Schematic symbol for an air-core inductor with two taps.*

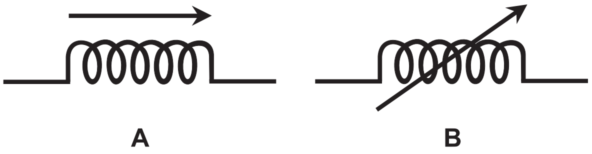

FIG. 3-17. *Schematic symbols for a variable air-core inductor. At A, arrow above coil symbol; at B, arrow passing through coil symbol.*

Whereas the fixed coil had only two leads, a tapped coil has three or more. When a coil is tapped, separate conductors are attached to one or more of the turns for intermediate connection. Maximum inductance is obtained from connecting the end leads to the external circuitry. A tapped arrangement allows for the selection of an input or output point that offers lower inductance than the full coil does.

As an alternative to taps, a coil might have a sliding contact that can be advanced along the entire length of the windings. This sliding contact allows adjustment of the inductance value, rather than having a select fixed point with the tapping arrangement. A variable coil can be indicated by either of the symbols shown in Fig. 3-17. The arrow indicates that the component can be adjusted from a maximum inductance value to a minimum inductance value.

Figure 3-18 shows symbols for a fixed air-core coil (at A), a tapped air-core coil (at B), and an adjustable air-core coil (at C).

An inductor meant for low-frequency applications can consist of a coiled wire wound around a solid or laminated (layered) iron core. Here, the iron, which constitutes a *ferromagnetic material*,

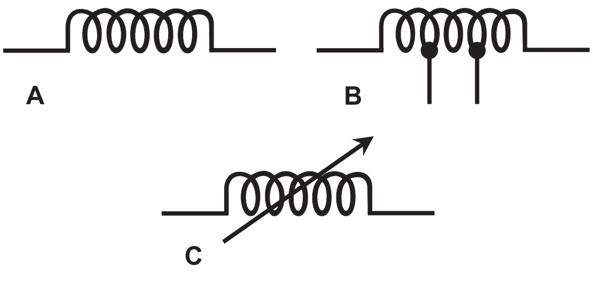

FIG. 3-18. *Schematic symbols for fixed (A), tapped (B), and adjustable (C) air-core inductors.*

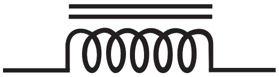

FIG. 3-19. *Schematic symbol for an inductor with a solid or laminated iron core.*

replaces the previous empty or air core. For example, a 60-Hz *choke*, intended for use in power-supply filters, will usually contain a single coil wound around a circular iron form. The ferromagnetic material greatly increases the *magnetic flux density* inside the coil windings, thereby increasing the inductance by a factor of many hundreds, or even thousands, of times compared with the inductance of an air-core coil having the same physical dimensions.

Figure 3-19 shows the schematic symbol for an iron-core inductor. Notice that it is the basic fixed coil discussed earlier, along with two close-spaced straight lines that run for its entire length. Sometimes the iron-core inductor is drawn as shown in Fig. 3-20, with the straight lines inside the coil turns in the symbol. (This is not the approved method of indicating an iron-core inductor, but you'll still see it now and then.) Some iron-core inductors contain taps for sampling different inductance values, and some might even be adjustable. The equivalent schematic symbols for these types of components appear in Fig. 3-21.

At higher frequencies, solid-iron and laminated-iron cores aren't efficient enough to function in inductors. Engineers would say that they have too much *loss*. At frequencies above a few kilohertz (kHz), a special core is needed if you want to increase the inductance over what you can get with *nonferromagnetic* core materials, such as air, plastic, ceramic, or wood. The most common substance for this purpose consists of iron material that has been shattered into myriad tiny fragments, each of which has a layer of insulation applied to it. After the fragmentation and insulation process has been completed, the particles are compressed to form a physically solid sample called

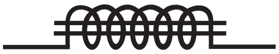

FIG. 3-20. *Alternate symbol for an inductor with a solid or laminated iron core.*

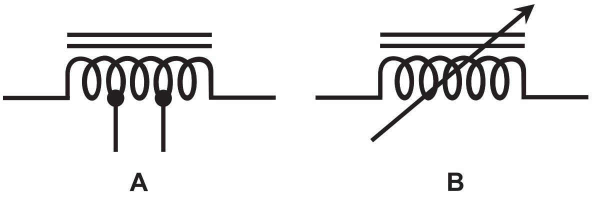

FIG. 3-21. *Symbols for a tapped coil (A) and an adjustable coil (B) with solid- or laminated-iron cores.*

a *powdered-iron core*. Figure 3-22 shows schematic symbols for powdered-iron-core inductors.

Tip

The symbols for powdered-iron-core inductors are nearly identical to those for solid- or laminated-iron-core inductors, except that the straight lines are broken up instead of solid. These types of components, like all other types of inductors, can be tapped or continuously variable.

A *transformer* is made up of multiple inductors with the coil turns interspersed or wound around different parts of a single core. Figure 3-23 shows the symbol for a basic air-core transformer. It looks like

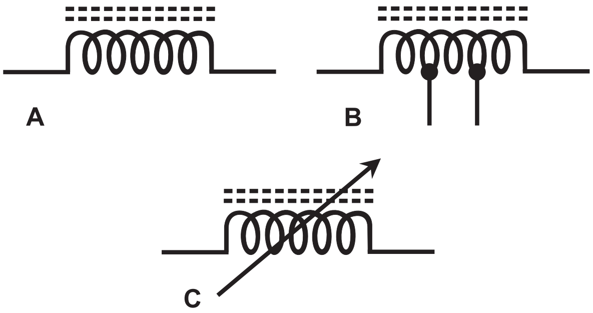

FIG. 3-22. *Schematic symbols for fixed (A), tapped (B), and adjustable (C) inductors with powdered-iron cores.*

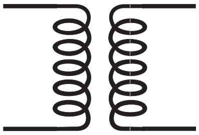

FIG. 3-23. *Schematic symbol for a transformer with an air core.*

two air-core coils drawn back-to-back. A transformer has the ability to transfer AC energy from one circuit to another at the same frequency. Because transformers are made by combining inductors, the schematic symbols are similar. Figure 3-24 shows some transformers that contain iron cores. The ones at A and B have solid or laminated cores; the ones at C and D have powdered cores.

Switches

A *switch* is a device, mechanical or electrical, that completes or breaks the path of current. Additionally, a switch can be used to allow

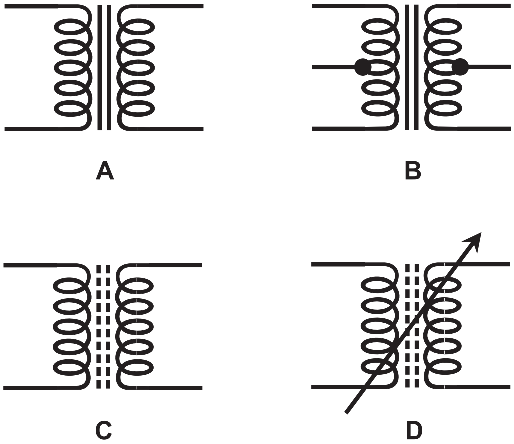

FIG. 3-24. *At A, a transformer with a solid- or laminated-iron core. At B, a transformer with a solid- or laminated-iron core and tapped windings. At C, a transformer with a powdered-iron core. At D, an adjustable transformer with a powdered-iron core.*

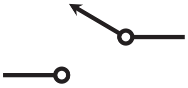

FIG. 3-25. *Schematic symbol for an SPST switch.*

current to pass through different circuit elements. Figure 3-25 shows the schematic symbol for a single-pole/single-throw (SPST) switch. This component can make or break a contact at only one point in a circuit; it's a two-position device (on-off or make-break).

Figure 3-26 shows a different type of switch, designated as a single-pole/double-throw (SPDT) component. Symbolically, the *pole* coincides with the point of contact at the base of the arrowed line. A *throw* is the contact point to which the arrow can point. The SPDT switch contains one pole contact and two throw positions; the input to the pole can be switched to either the upper or lower circuit point.

Some switches contain two or more poles. Figure 3-27A shows the symbol for a double-pole/single-throw (DPST) switch, while Fig. 3-27B shows the symbol for a double-pole/double-throw (DPDT) switch. Some switches have even more elements. The one shown in Fig. 3-28 has five poles, each of which can be switched to two separate positions. Engineers and technicians might call it a five-pole/double-throw (5PDT) arrangement.

This last designation can actually be covered under the heading of *multicontact switches*. This category takes in most switches that have more than two poles or two throw positions. For example, a rotary switch has a single pole and several throw positions; Fig. 3-29 shows an example. The arrow still indicates the pole contact. In this case the switch has 10 throw positions. Technically, then, it's a single-pole/10-throw (SP10T) device!

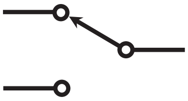

FIG. 3-26. *Schematic symbol for an SPDT switch.*

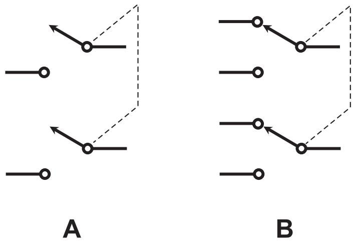

FIG. 3-27. *At A, symbol for a DPST switch. At B, symbol for a DPDT switch.*

Occasionally, you'll encounter sets of rotary switches ganged together, much like two or more variable capacitors can be made to rotate in sync with one another. Figure 3-30 shows the schematic symbol for an arrangement that uses two rotary switches. The dashed line tells us that the two switches are ganged. The two arrowed lines, which indicate the throw positions, go around "in sync" with each other. So, for example, when the left-hand switch (or pole number 1) rests at throw number 3 (as is the case here), the right-hand switch (or pole number 2) also rests at throw number 3.

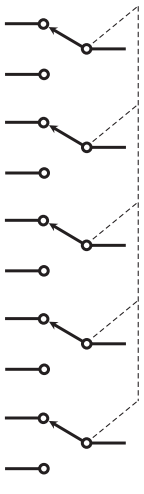

FIG. 3-28. *Schematic symbol for a five-pole double-throw (5PDT) switch.*

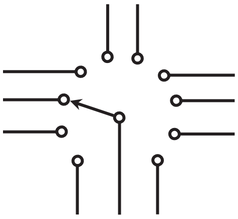

FIG. 3-29. *Schematic symbol for a rotary (or wafer) switch. This one has a single pole and 10 throws (SP10T).*

In each case, every switch contact point (pole or throw) is represented by a tiny circle. The variable element or pole is indicated by an arrow. The symbols shown here are all standard. You'll seldom see any significant variations.

Did you know?

Some amateur radio operators use a special switch called a *Morse code key*. This old-fashioned device, also called a *hand key* or a *straight key*, makes or breaks a circuit for the purpose of sending Morse code manually. It's an SPST switch with a lever and a spring, causing the device to return to the off position when the operator lets go of the lever. Figure 3-31 shows its schematic symbol.

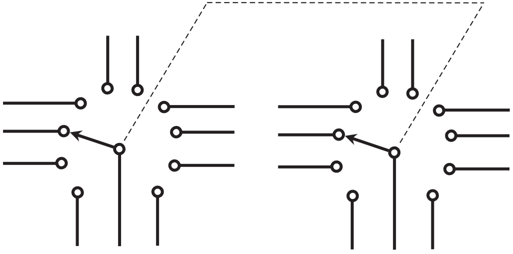

FIG. 3-30. *Schematic symbol for two rotary switches ganged together. This one has two poles and 10 throws (2P10T).*

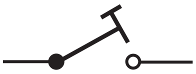

FIG. 3-31. *Schematic symbol for a Morse code key (hand key or straight key).*

Conductors and cables

Throughout this discussion, a straight line has always indicated a conductor, but most circuits contain a large number of conductors. When you draw a diagram of a complicated circuit or system, you'll often find it necessary to have lines cross over each other, whether the represented wires actually make contact in the physical system or not.

Figure 3-32 shows two conductors that must cross each other in a diagram, but that are not connected to each other in the physical circuit (at least not at the point where they cross in the schematic). This diagram geometry does not imply that when you build the circuit, the conductors must physically cross over each other at that exact place. It simply means that in order to make the schematic drawing, you have to draw one conductor across another to reach various circuit points without introducing a whole lot of confusion and clutter, or resorting to three dimensions to make your drawing.

Aha!

A real-world circuit exists in three-dimensional (3D) space, but when you want to diagram it, you must do it on a two-dimensional (2D) surface. To carry off that feat, you must learn a few tricks to make sure that your readers see things right!

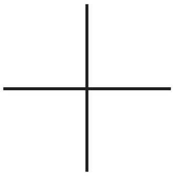

FIG. 3-32. *Schematic symbol for conductors that cross paths but are not electrically connected.*

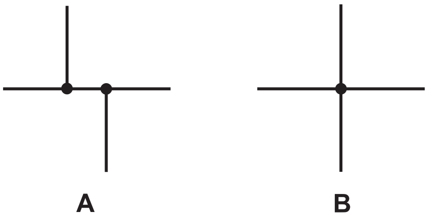

FIG. 3-33. *At A, preferred symbol for conductors that intersect and are electrically connected to each other. At B, alternate symbol for the same situation.*

Figure 3-33 shows two ways of portraying a point where two wires cross and they are electrically connected at that point. In the drawing at A, one of the conductors is "broken in two" so that it appears to contact the other one at two different points. This geometry makes it clear that the two conductors (the "divided" vertical one and the "solid" horizontal one) connect to each other electrically. Black dots indicate electrical connection. In the drawing at B, the two conductors cross (at right angles in this example), and a single black dot is drawn at the junction. This dot tells us that the conductors connect at this point. The method shown at B might look better at first glance, but the neatness comes along with a problem: Some readers might overlook the black dot and think that the two conductors are *not* meant to connect. The method at A makes that potential misinterpretation impossible.

Just as a reader might miss a black dot at a crossing point, as in Fig. 3-33B, another reader might see Fig. 3-32 and imagine a black dot when it isn't there! Then the reader will think the two wires connect when in fact they do not. This problem rarely occurs in well-engineered schematics where the draftsperson makes sure to use big black dots and good quality printing presses. However, in some older schematics you will see nonconnecting, crossed wires shown as in Fig. 3-34. One of the wires has a half loop that makes it look like it jumps over the other wire to avoid contact. That trick (which should never have gone out of style, in my opinion) gets rid of any doubt as to whether the wires electrically connect at the crossover point or not.

A *cable* consists of two or more conductors inside a single insulating jacket. In many cases, unshielded cables are not specifically

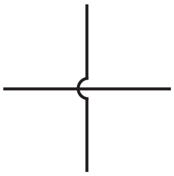

FIG. 3-34. *Archaic (but clear) representation of conductors that cross paths but are not electrically connected to each other.*

indicated in a schematic drawing, but appear as two or more lines that run parallel to indicate multiple conductors. Shielded cables require additional symbology along with the conductors. Figure 3-35 shows examples of shielded wire, often used to indicate the use of *coaxial cable* in an electronic circuit. Coaxial cable contains a single wire called the *center conductor* surrounded by a cylindrical, conduit-like conductive *shield*. An insulating layer, called the *dielectric*, keeps the two conductive elements isolated from each other. In most coaxial cables, the dielectric material consists of solid or foamed polyethylene.

Tip

Figure 3-36 shows a symbol for coaxial cable when the shield connects to a *chassis ground*, such as the metal plate on which an electronic circuit is constructed. The chassis ground might lead to an earth ground, but that's not always the case. In a truck, for example, no earth ground exists, so the chassis of the trucker's CB radio would go to the vehicle frame.

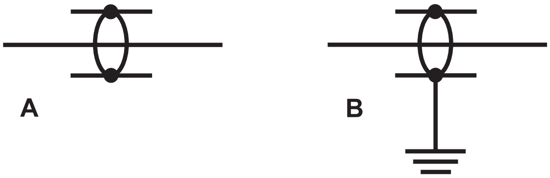

FIG. 3-35. *At A, symbol for a coaxial cable with an ungrounded shield. At B, symbol for a coaxial cable with an earth-grounded shield.*

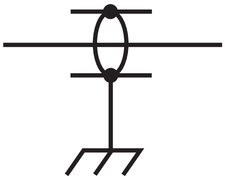

FIG. 3-36. *Symbol for a coaxial cable with a chassis-grounded shield.*

In some cables, a single shield surrounds two or more conductors. Figure 3-37 shows the schematic symbol for a two-conductor shielded cable. This symbol is identical to the one for coaxial cable, except that an extra inner conductor exists. If more than two inner conductors exist, then the number of straight, parallel lines going through the elliptical part of the symbol should equal the number of conductors. For example, if the cable in Fig. 3-37 contained five conductors, then five horizontal lines would run through the elliptical part of the symbol.

Diodes and transistors

Figure 3-38 shows the basic symbol for a *semiconductor diode*. In this symbol, an arrow and a vertical line indicate parts of the diode, and the horizontal lines to the left and right indicate the leads. The symbol in Fig. 3-38 portrays a *rectifier diode*. The arrowed part of the symbol corresponds to the diode's *anode*, and the short, straight line at the arrow's tip corresponds to the *cathode*. Under normal operating conditions, a rectifier diode conducts when the electrons move

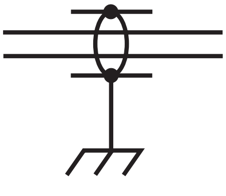

FIG. 3-37. *Symbol for a shielded two-conductor cable, in this case with a chassis ground for the shield.*

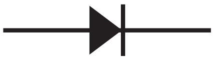

FIG. 3-38. *Symbol for a general-purpose semiconductor diode or rectifier.*

against the arrow, i.e., when the anode has a positive voltage with respect to the cathode.

Figure 3-39 shows the symbols for some specialized diode types. At A, we see a *varactor diode*, which can act as a variable capacitor when we apply an adjustable DC voltage to it. At B, we see a *Zener diode*, which can serve as a voltage regulator in a power supply. At C, we see a *Gunn diode*, which can act as an oscillator or amplifier at microwave radio frequencies.

A *silicon-controlled rectifier* (SCR) is, in effect, a semiconductor diode with an extra element and corresponding terminal. Its schematic symbol appears in Fig. 3-40. In the SCR representation, a circle often (but not always) surrounds the diode symbol, and the control element, called the *gate*, shows up as a diagonal line that runs outward from the tip of the arrow. In all cases, the lead that goes to the base of the arrow is the *anode* of the device, and the one connected to the short straight line at the arrow's tip is the *cathode*.

Figure 3-41 shows the schematic symbols for *bipolar transistors*. The PNP type is shown at A, followed by the NPN variety at B. The only distinction between the two is the direction of the arrow. In the PNP device, the arrow points into the straight line for the base electrode. In the NPN device, the arrow points outward from the base. Occasionally, the circle that surrounds the base, emitter, and

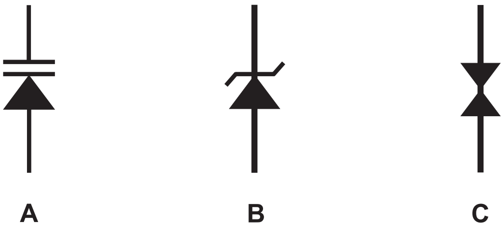

FIG. 3-39. *Symbols for a varactor diode (A), a Zener diode (B), and a Gunn diode (C).*

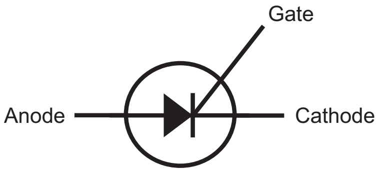

FIG. 3-40. *Symbol for a silicon-controlled rectifier (SCR).*

collector leads is omitted from the bipolar transistor symbol. Besides the bipolar variety, many other types of transistors exist. Figure 3-42 shows the symbols for four of these devices, as follows:

- At A, we see an *N-channel junction field-effect transistor* (JFET).
- At B, we see a *P-channel JFET.*
- At C, we see an *N-channel metal-oxide-semiconductor field-effect transistor* (MOSFET).
- At D, we see a *P-channel MOSFET.*

Tip

Transistors can be made from various types of semiconductor materials and metal-oxide compounds, but the schematic symbol, all by itself, tells us nothing about the elemental semiconductor material used in manufacture. The symbol merely indicates component functionality.

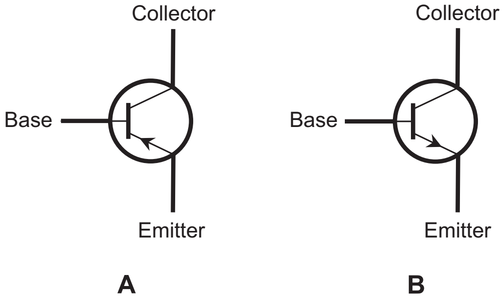

FIG. 3-41. *Symbols for a PNP bipolar transistor (A) and an NPN bipolar transistor (B).*

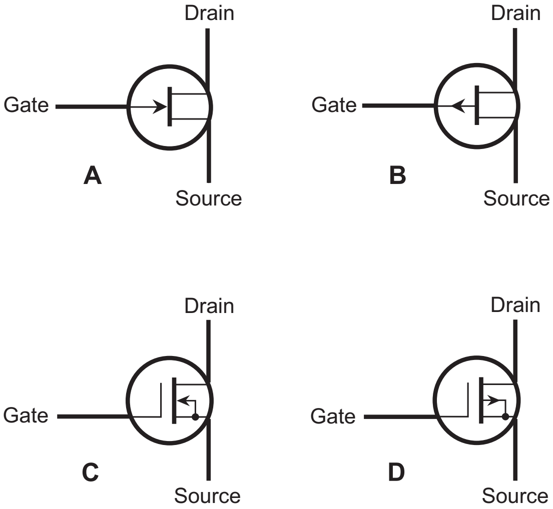

FIG. 3-42. *At A, symbol for an N-channel JFET. At B, symbol for a P-channel JFET. At C, symbol for an N-channel MOSFET. At D, symbol for a P-channel MOSFET.*

Electron tubes

Although vacuum tubes aren't used in electronics nearly as often as they were a few decades ago, many designs still exist that do employ them. When you want to create the symbol for a vacuum tube, you should start by drawing a fairly large circle, and then you should add the necessary symbols inside the circle to symbolize the type of tube involved. Figure 3-43 shows the schematic symbols for the various types of tube elements commonly used in schematic drawings.

Figure 3-44 shows the schematic symbol for a *diode vacuum tube*. This two-element device contains an *anode* (also called a *plate*) and a *cathode*. Just as with the semiconductor diode, the anode is normally positive with respect to the cathode when the device conducts current. The cathode emits electrons that travel through the vacuum to the anode. A hot-wire *filament*, something like a miniature low-wattage light bulb, heats the cathode to help drive electrons from it. In Fig. 3-44, the filament has been omitted for simplicity, a common practice in all vacuum tube symbology when the filament and cathode are physically separate, an arrangement known as an *indirectly heated cathode*.

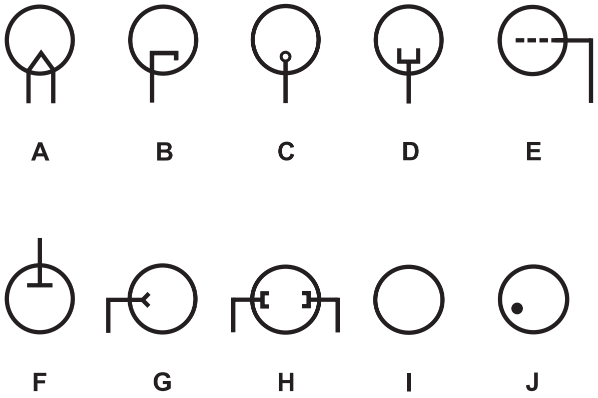

FIG. 3-43. *Symbols for tube elements and characteristics. A: Filament or directly heated cathode. B: Indirectly heated cathode. C: Cold cathode. D: Photocathode. E: Grid. F: Anode (plate). G: Deflection plate. H: Beam-forming plates. I: Envelope for vacuum tube. J: Envelope for gas-filled tube.*

Tip

All tube elements are surrounded by a circle, which represents the tube envelope. Occasionally, the circle is omitted from some tube symbols in schematic drawings, but that's not standard practice.

Figure 3-45 shows two versions of a *triode* vacuum tube, which consists of the same elements as the diode previously discussed, with the addition of a dashed line to indicate the grid. But there's another difference, too, in drawing A. Can you see it? Look closely

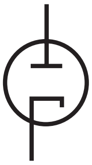

FIG. 3-44. *Schematic symbol for a diode vacuum tube with an indirectly heated cathode. Although a filament exists, it is often omitted to reduce clutter in symbols for tubes with indirectly heated cathodes.*

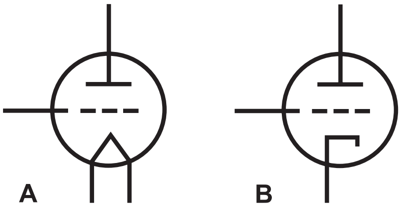

FIG. 3-45. *Symbols for a triode tube with a directly heated cathode (A) and an indirectly heated cathode (B).*

at the cathode. The tube at A has a *directly heated cathode*, in which the filament and the cathode are the very same physical object! We apply the negative cathode voltage directly to the filament wire; no separate cathode exists at all. In Fig. 3-45B, we see the symbol for a triode tube with an indirectly heated cathode. In this symbol, the filament is inside the cathode, which comprises a metal cylinder running along the central vertical axis of the tube.

Tetrode vacuum tubes have two grids. To represent one of them, we need an additional dashed line, as shown in the drawings of Fig. 3-46. In the tetrode, the upper grid, closer to the anode, is called the *screen*. Figure 3-47 shows symbols for the so-called *pentode* tube, which has three grids and a total of five elements. In the pentode, the second grid (going from the bottom up) is the screen, and the third grid (just underneath the plate) is called the *suppressor*. In both Figs. 3-46 and 3-47, the left-hand symbol (at A) portrays a device with a directly heated cathode, while the right-hand drawing (at B) shows a device with an indirectly heated cathode.

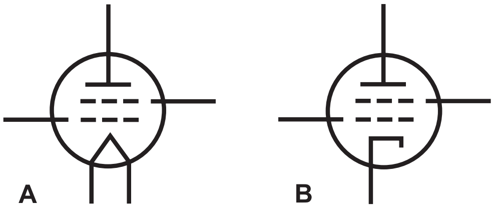

FIG. 3-46. *Symbols for a tetrode tube with a directly heated cathode (A) and an indirectly heated cathode (B).*

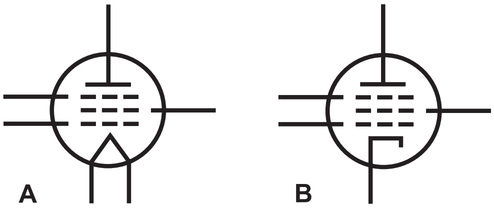

FIG. 3-47. *Symbols for a pentode tube with a directly heated cathode (A) and an indirectly heated cathode (B).*

Follow the flow

In all the vacuum tube symbols shown here, electrons normally flow from the bottom up. They come off the cathode, travel through the grid or grids (if any), and end up at the plate. Once in awhile you'll see a vacuum tube symbol lying on its side. In that sort of situation, you can simply remember that the electrons go from the cathode to the plate under normal operating conditions.

Some vacuum tubes consist of two separate, independent sets of electrodes housed in a single envelope. These components are called *dual tubes*. If the two sets of electrodes are identical, the entire component is called a *dual diode*, *dual triode*, *dual tetrode*, or *dual pentode*. Figure 3-48 shows the schematic symbol for a dual triode vacuum tube with indirectly heated cathodes.

In some older radio and television receivers, tubes with four or five grids were sometimes used. These tubes had six and seven elements, respectively, and were called *hexodes* and *heptodes*. These esoteric devices were used mainly for *mixing*, a process in which two RF signals having different frequencies are combined to get new signals at the sum and difference frequencies. The schematic symbol for a hexode is shown in Fig. 3-49A; the symbol for a heptode is shown

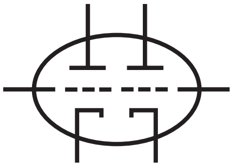

FIG. 3-48. *Schematic symbol for a dual triode tube.*

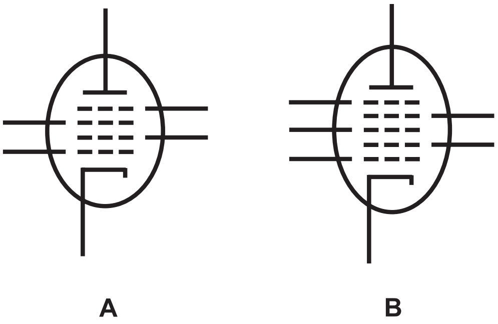

FIG. 3-49. *At A, symbol for a hexode tube. At B, symbol for a heptode tube, also known as a pentagrid converter.*

in Fig. 3-49B. Some engineers called the heptode tube a *pentagrid converter*. Both of these symbols show devices with indirectly heated cathodes.

Tip

You won't encounter hexodes and heptodes in modern electronics, but if you like to work with antique radios, you should get familiar with them. But take this warning: You'll probably have a difficult time finding a replacement component, should one of these relics go "soft" on you!

Cells and batteries

A cell or battery is often used as a power source for electronic circuits. Figure 3-50 shows the schematic symbol for a single *electrochemical cell*, such as the sort that you'll find in a flashlight. A single-cell component such as this usually has an output of approximately 1.5 V DC. *Electrochemical batteries* with higher voltage outputs comprise multiple cells connected in series (negative-to-positive in a chain or string);

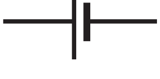

FIG. 3-50. *Schematic symbol for a single electrochemical cell.*

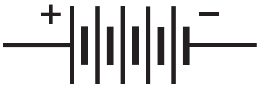

FIG. 3-51. *Schematic symbol for a self-contained multicell electrochemical battery.*

the schematic representation for a multicell battery takes this design into account, as shown in Fig. 3-51.

The multicell battery symbol is simply a number of single-cell symbols placed end-to-end without any intervening lines. If a circuit calls for the use of three individual, discrete single-cell batteries in a series connection, you might draw three cell symbols in series with wire conductor symbols between them (Fig. 3-52). Alternatively, if multiple individual cells are set in a "battery holder" designed for direct series connection, you can use a battery symbol to portray the whole bunch.

Standard practice calls for polarity signs to go with the symbols for cells or batteries. Unfortunately, some draftspeople neglect this detail. Then when you see the schematic, you'll have to infer the polarity by scrutinizing the rest of the circuit.

Logic gates

All digital electronic devices employ switches that perform specific logical operations. These switches, called *logic gates*, can have anywhere from one to several inputs and (usually) a single output. Logic devices have two states, represented by the digits 0 and 1. The 0 digit is normally called "low" and the 1 digit is called "high."

- A *logical inverter*, also called a *NOT gate*, has one input and one output. It reverses, or inverts, the state of the input. If the input equals 1, then the output equals 0. If the input equals 0, then the output equals 1.
- An *OR gate* can have two or more inputs (although it usually has only two). If both, or all, of the inputs equal 0, then the

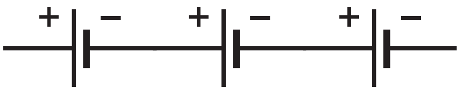

FIG. 3-52. *Symbol for three single electrochemical cells connected in series to form a battery.*

output equals 0. If any of the inputs equals 1, then the output equals 1. Mathematical logicians would tell us that such a gate performs an *inclusive-OR operation* because it "includes" the case where both variables are high.

- An *AND gate* can have two or more inputs (although it usually has only two). If both, or all, of the inputs equal 1, then the output equals 1. If any of the inputs equals 0, then the output equals 0.
- An OR gate can be followed by a NOT gate. This combination gives us a *NOT-OR gate*, more often called a *NOR gate*. If both, or all, of the inputs equal 0, then the output equals 1. If any of the inputs equals 1, then the output equals 0.
- An AND gate can be followed by a NOT gate. This combination gives us a *NOT-AND gate,* more often called a *NAND gate*. If both, or all, of the inputs equal 1, then the output equals 0. If any of the inputs equals 0, then the output equals 1.
- An e*xclusive OR gate*, also called an *XOR gate*, has two inputs and one output. If the two inputs have the same state (either both 1 or both 0), then the output equals 0. If the two inputs have different states, then the output equals 1. Mathematicians use the term *exclusive-OR operation* because it doesn't "include" the case where both variables are high.

Figure 3-53 illustrates the schematic symbols that engineers and technicians use to represent these gates in circuit diagrams.

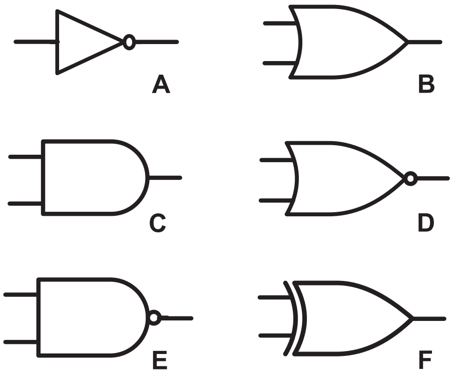

FIG. 3-53. *Symbols for a logical inverter or NOT gate (A), an OR gate (B), an AND gate (C), a NOR gate (D), a NAND gate (E), and an XOR gate (F).*

Summary

You'll encounter lots of symbols in electronics other than the common ones shown in this chapter. Appendix A is a comprehensive table of schematic symbols. In addition to the ones already discussed, you will see symbols for jacks and plugs, piezoelectric crystals, lamps, microphones, meters, antennas, and many other electronic components.

It might, at first thought, seem like a massive chore to memorize all of these symbols, but their usage and correct identification will come to you with practice and with time. The best way to begin the learning process is to read simple schematics and refer to Appendix A in this book whenever a symbol crops up that you can't identify. Within a few hours you'll be able to move on to more complex schematics, again looking up the unknown symbols. After a few weekends of practice, you should be thoroughly familiar with most electronic symbols used in schematic representations, so that when you see one in a diagram, you'll recognize it without having to think about it.

Schematic symbols are the fundamental elements of a communication scheme, like the symbols in mathematical expressions or architectural blueprints. Most schematic symbols in electronics are based on the structure of the components or devices they represent. Schematic symbols often appear in groups, each of which bears some relationship to the others. For example, you'll encounter many different types of transistors, but they're all represented in a similar fashion. Minor symbol changes portray variations in internal structure, but all can be easily identified as some type of transistor. The same rule applies to the symbols for diodes, resistors, capacitors, inductors, transformers, meters, lamps, and most other electronic components.

4

Simple circuits

When it comes to teaching people how to read and draw schematic diagrams, two schools of thought prevail. One school feels that you should learn to read diagrams before you learn to draw anything. The other school feels that you should learn to read diagrams as part of the process of drawing them. Both sides of this issue have their merits, so let's take advantage of them both! Certainly, you must start out by learning the basic component symbols. Then you should try to read as many schematic drawings as you can find. When this business starts to grow boring, you can draw your own simple diagrams. If you alternately read and draw schematics as the moods strike, you'll get comfortable with them before long! I recommend that you devote half of your study time to reading and the other half to drawing.

Getting started

This chapter deals with reading and drawing diagrams of some simple electronic circuits shown in pictorial and schematic form. Using this method, you can actually see the physical layout for a device, and then see how the schematic representation derives from it. Some commercial schematics are produced in this manner. However, in most instances, a circuit is designed schematically first, and then built and tested from the schematic. If a circuit is experimental, some bugs will exist in the prototype, necessitating various component deletions, substitutions, or modifications. When these changes are made to the test circuit, the results are noted, and the schematic is changed accordingly. In the end, the finished and corrected schematic is a product of design theory, actual testing, and modification.

Figure 4-1 shows a simple circuit that everyone has used at one time or another. Basically, it's a flashlight with the external case and the on/off switch removed. The device consists of a single electrochemical cell and an electric light bulb. This pictorial representation also shows the conductors, which attach to the light bulb and the battery. The conductors provide a current path between the battery and the light bulb.

Follow the flow

In the circuit of Fig. 4-1, electrons travel from the negative terminal of the cell through the bulb element and back to the positive terminal of the cell, "leapfrogging" from atom to atom in the metal wire and the bulb filament. That's how most electricians and engineers look at this situation. But some physicists will tell you that the current actually goes from the positive cell pole to the negative cell pole. That's called *theoretical current* or *conventional current*.

In order to draw a schematic diagram of the flashlight drawn in Fig. 4-1, you need to know three schematic symbols. These represent the electrochemical cell, the conductors, and the bulb, as shown in Fig. 4-2. Once you know the symbols, you can assemble them in a

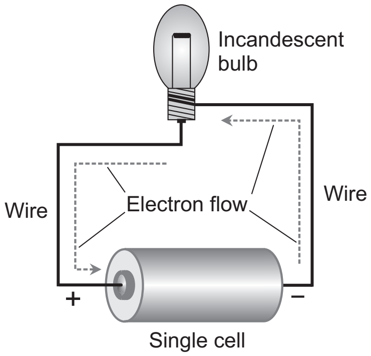

FIG. 4-1. *Pictorial drawing of a flashlight circuit using a single electrochemical cell (or single-cell battery), some wire, and an incandescent bulb.*

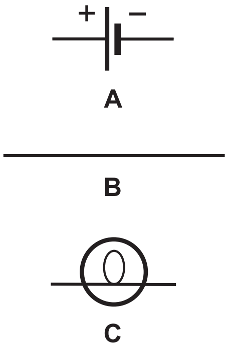

FIG. 4-2. *Schematic symbols for an electrochemical cell (A), an electrical conductor such as wire (B), and an incandescent bulb (C).*

logical manner based on the appearance of the circuit in the pictorial drawing.

Start by drawing the cell symbol. You can think of the cell as the heart of the circuit because it supplies all of the power for the device; it "pumps the electrons" through everything! Next comes the symbol for the light bulb, which you can draw at any point near the cell. Using this example, you should try to make the schematic symbols fall in line with the way the pictorial diagram appears. This layout places the light bulb above the cell.

Now that you've drawn the two major symbols, you can use the conductor symbols (plain, straight, solid, black lines) to hook them together. Notice that the pictorial drawing shows two conductors. Therefore, the schematic diagram also has two conductors. Figure 4-3 shows the completed schematic drawing, which is the symbolic equivalent of Fig. 4-1.

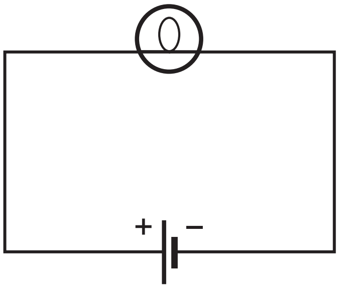

FIG. 4-3. *Schematic diagram of the single-cell flashlight.*

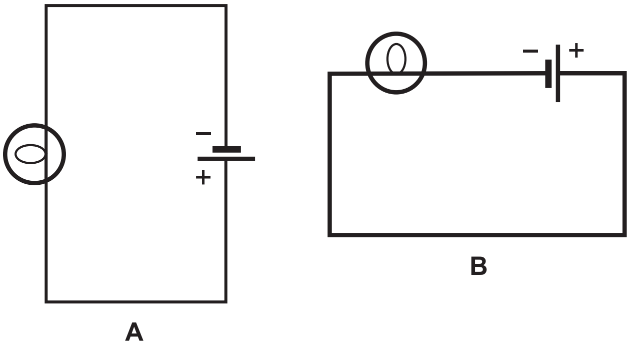

FIG. 4-4. *Alternative arrangements for the flashlight schematic. At A, cell and bulb "vertical;" at B, cell and bulb both on top.*

Figure 4-3 is by no means the only way that you can represent this simple circuit in schematic form. But any schematic representation will require the use of the same three basic symbols: cell, bulb, and conductors. The only changes that can occur involve the positioning of the component symbols on the page. Figure 4-4 shows two different alternatives for portraying the same circuit. All three of these diagrams (the one in Fig. 4-3 and the two in Fig. 4-4) are electrically equivalent, but they look somewhat different as a result of the relative positions of the components on the page.

Let's change this circuit a little bit, in order to gain proficiency in reading and writing schematics. Figure 4-5 shows the same basic flashlight circuit, but an additional cell and a switch have been added. This

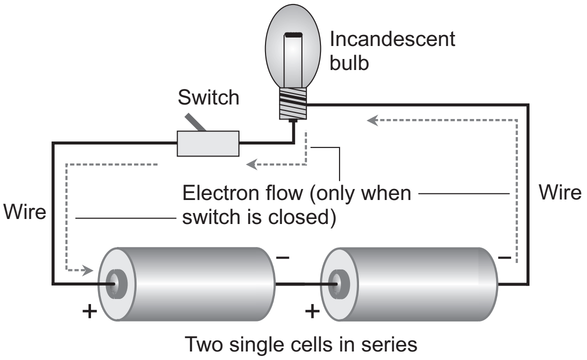
Incandescent bulb
Switch
Electron flow (only when switch is closed)
Wire
Wire
–
–
+
+
Two single cells in series

FIG. 4-5. *Pictorial drawing of a flashlight circuit using two single cells in series, some wire, a switch, and an incandescent bulb.*

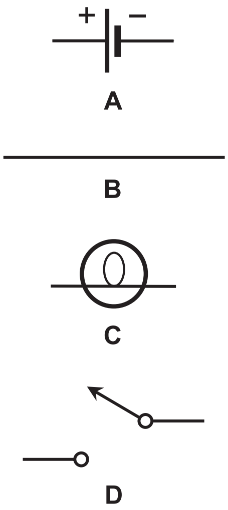

FIG. 4-6. *Symbols for components in the two-cell switched flashlight: Cell (A), wire (B), bulb (C), and switch (D).*

configuration is quite common for flashlights sold in the United States. By examining this pictorial drawing, you can see that any schematic representation will need symbols for the cells, the conductors, the light bulb, and the switch. Figure 4-6 shows the symbols that you'll need to produce an accurate and complete schematic drawing of this circuit. Again, you should draw the symbols in the same basic order as the components are wired in the circuit. Figure 4-7 shows the resulting schematic. Note that the two cell symbols are drawn separately, connected in series, with polarity markings provided for each one. In the series connection, the positive terminal of one cell goes to the negative

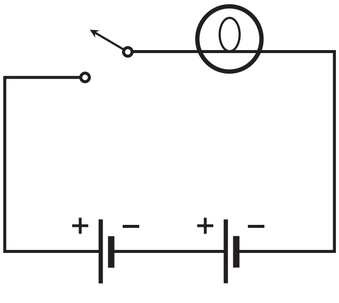

FIG. 4-7. *Schematic diagram of the two-cell switched flashlight.*

terminal of the other. The same two conductors are used from the cell terminals, but you need a third one to connect the switch to the light bulb, and you might also need a fourth one to connect the two cells together to form a battery (unless the cells rest directly against each other, a common state of affairs inside commercially manufactured flashlights). Figure 4-7 shows the switch in the off position.

Eureka!

Now you know what a common two-cell flashlight looks like when represented with schematic symbology. The next time that you switch one of those things on, you can imagine the switch symbol in Fig. 4-7 moving from the off (or open) position to the on (or closed) position.

Figure 4-8 is a pictorial representation of a device called a *field-strength meter*. Wireless communications engineers sometimes use this type of meter to see whether or not an RF electromagnetic (EM) field exists at a given location. You'll find this little circuit quite handy if you enjoy amateur radio, or if you need to locate the source of something that's causing RF interference. The circuit consists of an antenna, an RF diode, a microammeter (a sensitive current meter graduated in millionths of an ampere), and a coil. In order to draw this circuit

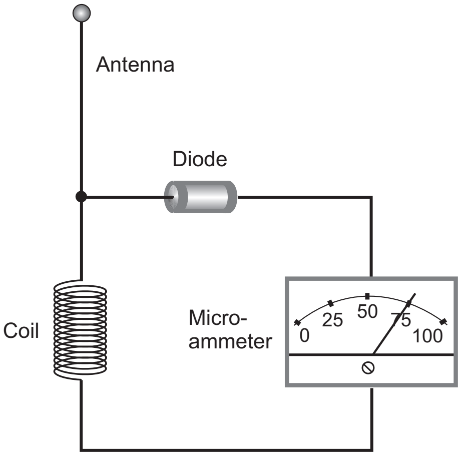

FIG. 4-8. *Pictorial representation of a field-strength meter circuit.*

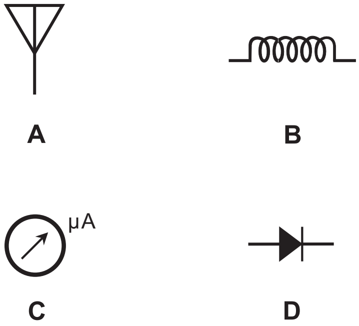

FIG. 4-9. *Schematic symbols for the components in the field-strength meter: Antenna (A), coil (B), microammeter (C), and diode (D).*

schematically, you need to know the symbols for an antenna, a coil, a microammeter, and a diode, all of which appear in Fig. 4-9. Using the same method as before, you can draw the schematic by connecting the symbols in the same geometric sequence as the components they represent appear in the circuit.

Figure 4-10 is a schematic of the field-strength meter shown pictorially in Fig. 4-8. This drawing involves nothing more than substitution of the schematic symbols for the pictorial symbols. As before, the parts need not be physically placed in the same positions as the schematic diagram suggests, but they must be interconnected precisely as indicated in the schematic. When you build a circuit from a schematic diagram that you trust, you should double-check and triple-check your actual component interconnections to make sure that they agree with the schematic. If you try to build the circuit shown in Fig. 4-10

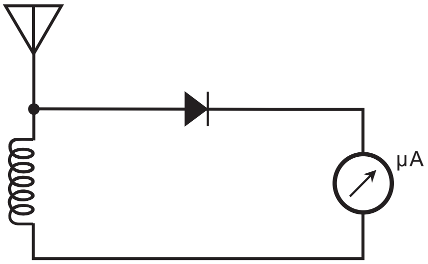

FIG. 4-10. *Schematic diagram of the field-strength meter.*

and make a mistake in the wiring connections, you cannot expect it to work. In more sophisticated devices and systems, wiring mistakes can cause component damage and, once in awhile, give rise to dangerous situations!

Follow the flow

In the circuit of Figs. 4-8 and 4-10, an EM field induces an RF current in the antenna and coil. That current is high-frequency AC. The diode rectifies the AC wave by "chopping off" either the positive half or the negative half of every cycle (depending on the diode's polarity) to produce pulsating DC like the output of a simple rectifier. The microammeter registers this current. As the strength of the EM field increases, the current increases, and the meter reading goes up.

Previously, we compared schematic drawings to road maps. A road map is supposed to indicate exactly what a motorist will experience in practice. The schematic drawing does the same thing. The highway lines that interconnect towns on a map correspond to the conductor lines that interconnect components in a schematic. A road map shows the route that a motorist takes while driving from town to town on the earth's surface, whereas a schematic diagram shows the route that a current or signal takes through the components in an electronic circuit.

Now let's look at something that's a little more complicated. Figure 4-11 is a schematic diagram of a *power supply* that produces pure, battery-like DC from utility AC. As you read this diagram from left to right, you'll see that a power plug goes to the transformer primary winding (the one to the left of the pair of vertical lines) through a fuse. At the top of the transformer secondary winding (the one on the right), a rectifier diode is connected in series. Following the diode, an electrolytic capacitor (note the polarity sign) is connected between the output of the rectifier and the bottom of the transformer secondary. A fixed resistor is connected in parallel with the capacitor. The DC output appears at the extreme right.

The physical size and weight of a real-world power supply, which you can build on the basis of Fig. 4-11, will depend on the voltage and current that you need to get from it. Because DC power

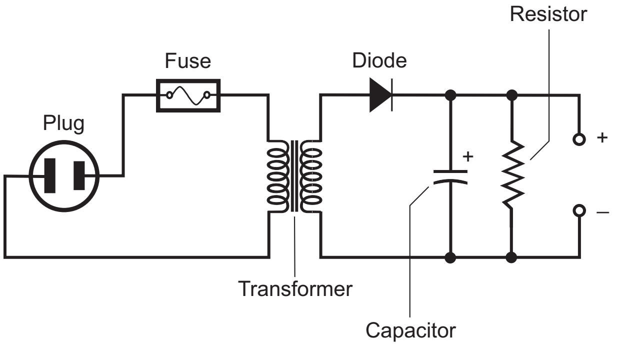

FIG. 4-11. *Schematic diagram of a simple DC power supply.*

supplies have polarized outputs, positive and negative signs indicate the output voltage polarity. Any power supply that uses a single diode, capacitor, and resistor will have this same basic configuration. Whether the output is 5 V at 1 A or 5000 V at 50 A, the schematic drawing will look the same. Figure 4-11 says nothing about how many volts or amperes the transformer, diode, capacitor, and resistor are meant to handle. You could add special features, such as a voltage regulator, overcurrent protector, voltmeter, or ammeter to your circuit and insert the symbols at the proper points in the schematic; but all half-wave DC power supplies are built around "cores" whose diagrams look like Fig. 4-11.

Follow the flow

In the circuit of Fig. 4-11, utility AC appears at the plug on the left. The AC travels through the fuse and flows in the transformer primary. In the secondary, AC also flows, but the voltage across the transformer secondary might be higher or lower than the voltage across the primary (depending on the transformer specifications). The diode allows current to flow only one way; in this case the electrons can go only from right to left (against the arrow). As a result, pulsating DC comes out of the diode. The capacitor gets rid of the pulsations, called *ripple*, on the DC output from the diode. The resistor discharges, or *bleeds*, the capacitor when you unplug the whole device from the utility outlet.

Component labeling

Figure 4-12 is a schematic of the same circuit as the one shown in Fig. 4-11, but in this case, each component has an alphabetic/numeric designation. These designators all refer to a components list at the bottom. Now you can see that this power supply uses a transformer with a primary winding rated at 125 V and a secondary winding that yields 12 V. The circuit has a diode rated at 50 peak inverse volts (PIV) and a forward current of 1 A; a 100-microfarad, 50-V capacitor; and a 10,000-ohm, 1-W carbon resistor. The fuse is rated at 0.5 A and 125 V.

The letters that identify each component are more or less standard. Notice that each letter is followed by the number 1. The designation T1, for instance, indicates that the component is a transformer (T) and that it's the first such component referenced. If this circuit had two transformers, then one of them would bear the label T1 and the other one would bear the label T2. The numbers reference the position or order on the components list; they serve no other purpose. The diode carries the reference designator D1, with D serving as the standard abbreviation for most diodes. Standardization is not universal, though! In some instances, the diode might bear the label SR1, where the letters SR stand for *silicon rectifier*. Some Zener diodes are labeled as ZD1, ZD2, and so on. This labeling makes little difference as long as you write the component designations next to the corre-

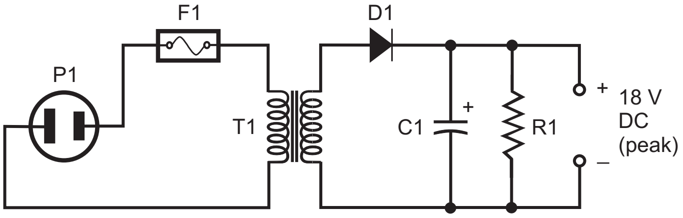

C1 - 100 microfarad electrolytic 50 V DC
D1 - 50 peak inverse volts at 1 ampere
F1 - 0.5 ampere at 125 volts
P1 - Male line plug
R1 - 10,000 ohm 1 watt carbon
T1 - 125 volt primary, 12 volt secondary, 1 ampere

FIG. 4-12. *Schematic diagram of the power supply with component designators and specifications.*

sponding symbols. If you replaced the designation D1 with SR1, your readers would still know that the abbreviation went with the symbol for the diode, as long as you made sure to put the abbreviation close enough to the symbol.

In the situation of Fig. 4-12, you don't have to include a number next to each component designation because only one of each component is used to make up the entire schematic! You could simply write P for the plug, F for the fuse, T for the transformer, D for the diode, C for the capacitor, and R for the resistor. Or, if you had confidence that your readers knew all the symbols, you could leave out designators altogether! Nevertheless, standard diagramming practice requires that you always include a letter and a number, even if only one of a certain component type exists in the whole circuit.

In complicated electronic systems, several hundred components of the same type (resistors, for example) might exist, many of which come from the same family. For instance, if you see the designation R101, then you know that the system contains at least 101 resistors. If you want to know the type and value of resistor R101, you will have to look up R101 in the components list to find its specifications.

Tip

You can use the schematic of Fig. 4-12 to build a power supply with a peak output of about 18 V DC. But before each component was referenced, the schematic had no practical use. Figure 4-11 illustrates a generic half-wave power supply, but that schematic doesn't give you any practical information other than the relative arrangement of the components.

Table 4-1 shows the standard letter designations for most types of electronic components that you'll encounter in schematic diagrams. Some of these designations can vary in real-world documentation, depending upon the idiosyncrasies of the person making the drawing or designing the circuit. You should find it easy to memorize (yes, that's right, memorize!) the information in Table 4-1 because most of the designations merely comprise the first letters of the component names. If the component has a complex name, such as *silicon-controlled rectifier*, the first letters from each of the three words is used, so you get SCR1. A resistor is designated by R, a capacitor by

Table 4-1
Letter designations for components that commonly appear in schematic diagrams

Abbreviation	Full component name
ANT	Antenna
B	Battery
C	Capacitor
CB	Circuit board
D	Diode
EP	Earphone
F	Fuse
GND	Ground
I	Incandescent lamp
IC	Integrated circuit
J	Receptacle, jack, or terminal strip
K	Relay
L	Inductor
LED	Light-emitting diode
M	Meter
NE	Neon lamp
P	Plug
PC	Photocell
PH	Earphone
Q	Transistor
R	Resistor
RFC	Radio-frequency choke
RY	Relay
S	Switch or telegraph key
SCR	Silicon-controlled rectifier
SPK, SPKR	Speaker
SR	Silicon rectifier
T	Transformer
TP	Terminal or test point
U	Integrated circuit
V	Vacuum tube
Y	Quartz crystal
Z	Circuit assembly

C, a fuse by F, and so on. Conflicts do arise, of course. If you want to designate a relay, you need to use some letter other than R because R indicates a resistor! The same thing happens if you want to label a crystal; you can't use C because that letter refers to a capacitor. Look through Table 4-1 from time to time as you read and draw schematic diagrams, and eventually you'll absorb all the information in there.

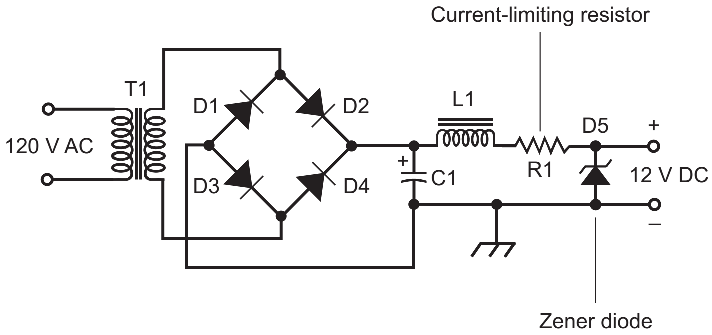

FIG. 4-13. *A power supply that uses full-wave bridge rectification (four rectifier diodes) and Zener-diode voltage regulation.*

The circuit of Fig. 4-13 has a full-wave bridge rectifier along with a better ripple filter than the simple capacitor used in the previous power supply. The inductor, L1, is a filter choke, which, along with capacitor C1, does an excellent job of "smoothing" out the DC so it resembles what comes from a 12-V battery (pure DC with no ripple).

Figure 4-14 shows a *voltage-doubler power supply*. The two capacitors, C1 and C2, charge up from the full transformer secondary output after the current goes through diodes D1 and D2. Because the two capacitors are connected in series, they act like two batteries in series, giving you twice the voltage. But there's a catch! A voltage doubler power supply works well only at low current levels. If you try to draw too much current from one of these power supplies, you'll "draw down" the capacitors and the voltage will decrease.

In Figs. 4-13 and 4-14, the letter designations are the same for each component type, but the numbers advance, one by one, up to the total number of units. So, for example, in Fig. 4-13 you see diodes

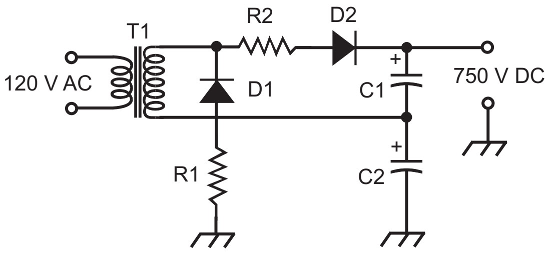

FIG. 4-14. *A voltage-doubler power supply.*

D1 through D5 because the circuit contains five diodes. (The Zener diode to the right of R1 has the letter D just like the rectifier diodes have, but you can tell it's a Zener diode because of the bent line in the symbol.) All the other components have only one of each type. In Fig. 4-14, you see two diodes, two capacitors, and two resistors, so the numbers for D, C, and R go up to 2. The transformer is all alone, so you see only the number 1 following the letter T.

Tip

Even though multiple components might all have the same value (820 ohms, for example, or 50 microfarads), they must nevertheless get separate *numerical* designations when two or more of them exist in a single circuit.

Schematics don't reveal every physical detail of a device, the way a photograph or detailed pictorial would do. Schematics depict schemes, that's all! The schematic diagram for a device allows engineers and technicians to make the correct electrical connections when putting it together, and to locate the various components when testing, adjusting, debugging, or troubleshooting it. If you find all this talk overly philosophical, maybe a real-world example will clear things up. Remember that solid lines in schematic drawings represent conductors. However, a conductor doesn't have to be a length of wire. It might be part of a component lead, or perhaps a *foil run* on a printed circuit board (the latter-day equivalent of a connecting wire). Whether or not a separate length of wire is needed to interconnect two components will depend on how close together those components are in the physical layout.

Examine the simple schematic of Fig. 4-15. The circuit contains three resistors, all of which go together in a parallel arrangement.

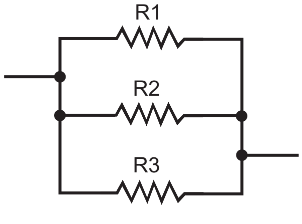

FIG. 4-15. *A simple circuit comprising three resistors in parallel.*

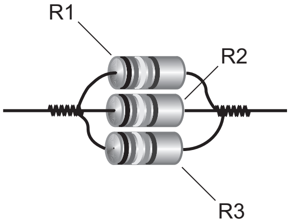

FIG. 4-16. *Pictorial diagram of three resistors in parallel with leads intertwined.*

Taking the schematic literally, a conductor connects the left-hand side of R1 to the left-hand side of R2. Another conductor goes between the left-hand side of R2 and the left-hand side of R3. Two other conductors connect the right-hand sides of the components. In practice, the connections might be made with wires attached to the resistor leads, but if the components are close enough together, the leads themselves can form the interconnections. Then Fig. 4-15 will represent the physical arrangement shown in Fig. 4-16.

Naturally, if you want to follow good engineering principles, you'll want to make all of your electronic circuits as compact (and dependable) as possible by using a minimum amount of point-to-point wiring and trying to make the component leads serve for interconnection purposes whenever you can. Of course, in the above example, if the three resistors had to go in different parts of the circuit separated by some physical distance, then you would need to use interconnecting conductors between them. However, as you design the physical layout of a circuit, you should try to minimize the overall length (that is, the total length) of all the interconnecting wires or foil runs combined.

Troubleshooting with schematics

Engineers and technicians use schematic diagrams to create electronic devices, but these diagrams can also prove invaluable for troubleshooting equipment when problems develop. Knowing how to read schematic diagrams, however, is not enough. You also need to know what tasks the various components actually perform, as well as how the diverse circuits work together in a complete system. No matter how proficient you might get at electronics troubleshooting, seemingly simple repair jobs can explode into major headaches without complete, accurate, and clear schematic representations of the hardware.

Remember!

Schematic diagrams clarify circuits. They present the circuit elements in a logical and easy-to-understand manner. They tell you very little, if anything, about the component layouts in actual devices.

When you build a circuit from a schematic drawing, the physical object rarely bears much physical resemblance to the schematic. It's impractical to build a complex electronic circuit by placing the components in the exact same geometrical relationship as they appear in the schematic. The diagrams purposely spread out the components on the page for easy reading. Schematic diagrams are two-dimensional, whereas real-world electronic components are three-dimensional. You need only to look inside of a major electronic device, such as a television set or computer, to realize the complexities that you'd face in troubleshooting a complex system without the help of a schematic diagram.

If you know a fair amount about electronic components and how they operate in various circuits, then you can use a schematic diagram to get a good idea (without any equipment testing) where a particular problem might occur. Then, by testing various circuit parameters at these critical points and comparing your findings with what the schematic diagram indicates should be present, you can make a quick assessment of the trouble. For example, if a schematic diagram shows a direct connection between two components in a circuit, and a check with an ohmmeter reveals a high resistance between the two, then you can assume that a conductor is broken or a contact has been shaken loose. If a schematic diagram shows only a capacitor between two components (with no other circuit routes around it) and a reading with your ohmmeter shows zero ohms or only a couple of ohms, you can assume that the capacitor has shorted out and you'll have to replace it.

Beginners to electronics troubleshooting and diagram reading sometimes assume that a professional can instantly isolate a problem to the component level by looking at the schematic. This idealized state of affairs might prevail for a few simple circuits, but in complex designs the situation grows a lot more involved. Often, the schematic diagram allows a technician to make educated guesses as to where or what the trouble might be, but an exhaustive diagnosis will nearly always require testing. A particular malfunction in an electronic device will not necessarily have a single, easy-to-identify cause. Often there are many possible causes, and the technician must whittle the situation down to a single cause by following a process of elimination.

Suppose that a circuit will not activate, and no voltage can be detected through testing at any contact point indicated by the schematic. Chances are good that no current is passing through the circuit at all. However, you don't know from this observation exactly what has caused the failure. Has one of the components in the power supply become defective? Has the line cord been accidentally pulled from the wall outlet? Has a conductor broken between the output of the power supply and the input to the electronic device? Has the fuse blown?

In a scenario of this sort, you will almost certainly want to consult the schematic diagram as you go through all of the standard test procedures. You might wish to find the contact point that serves as the power supply output, indicated on the schematic. If you test the voltage at this point and it appears normal, then you can assume that the problem lies somewhere further along in the circuit. The schematic diagram and the test instrument readings allow you to methodically search out and isolate the problem by starting at a point in the circuit where operation is normal and proceeding forward until you get to the point where the circuit shows some abnormality.

Continuing with the same example, if no output comes from the power supply, you know that you must search backward toward the trouble point. You will continue testing until you reach a point of normal operation and then proceed from there. Using a schematic diagram, you'll follow your progress and thereby narrow the problem area down to something between two points (the point farthest back from the output at which the problem exists, and the point furthest forward from the input where things test normal). Chances are good that this narrowing process will isolate the trouble to a single component or circuit connection.

Tip

You could follow all of the foregoing steps without a schematic diagram, although it would take you a lot longer to do it, and it would increase the risk of your making a mistake. As you become more experienced in the art of electronics troubleshooting, the information contained in schematic drawings becomes increasingly valuable.

Go back and look again at the flashlight circuit of Fig. 4-7. Although the schematic diagram does not say so, the two batteries in series should yield a DC potential of 3 V because a typical flashlight cell provides 1.5 V, and DC voltages add up in series connections. Some schematic diagrams provide voltage test points and maximum/minimum readings that you should expect, but this simple example doesn't.

Suppose that the flashlight has stopped working, and you decide to test the circuit with a *volt-ohm-milliammeter* (VOM), also called a *multimeter*, with the help of Fig. 4-7. First of all, you can measure the individual voltages across the cells. With the meter's positive probe placed at the positive cell terminal and the negative probe at the negative terminal, you should get a reading of 1.5 volts across each cell. If both read zero, then you know that both cells have lost all their electrical charge. If one cell reads normal and the other one reads zero, then in theory you should only have to replace the one that reads zero. (In practice, it's a good idea to replace entire sets of cells all at once, even if some of them still test okay). If both cells read normal, then you can test the voltage across the bulb. Here, you should expect a reading of 3 V under normal operation with the switch closed. If you do indeed observe 3 V here, then you can diagnose the problem by looking at the schematic. The bulb must have burned out! The schematic shows you that current must go through the light bulb if the bulb can conduct, so it must light up. If voltage is available at the base of the light bulb, then current will flow through the element unless it has opened up. But of course, if the bulb filament has broken apart, no current can flow through the bulb, so it won't light up. In fact, with a burned-out bulb, no current will flow anywhere at all in the circuit.

On the other hand, let's say that you get a normal reading at the batteries, but no reading whatsoever at the light bulb. Obviously, a break must exist in the circuit between these two circuit points. Three conductors are involved here: one between the negative terminal of the battery and one side of the bulb, another between the positive battery terminal and the switch, and another between the switch and the other side of the bulb. Obviously, one of the conductors has broken (or a contact has been lost where the conductor attaches to the battery), or maybe the switch is defective. While you keep an eye on the schematic, you can test for a defective switch by placing the negative meter probe on the negative battery terminal and the positive probe on the input to the switch. If you see a normal voltage reading, then the switch must be defective. If you still get no voltage reading, then one of the conductors has come loose or broken.

Admittedly, the scenario just described presents only a basic example of troubleshooting using a schematic diagram—almost as simple as things can get! But imagine that the flashlight circuit is highly complex, one you know nothing about. Then the schematic diagram becomes an invaluable aid and a necessary adjunct to the standard test procedures with the VOM. This same basic test procedure will be used over and over again when testing highly complex electronic circuits of a similar nature. In most instances, no matter how complicated the circuit design looks, it's actually a combination of many simple circuits. But if you have to do a comprehensive troubleshooting operation, you might have to test each and every one of those circuits individually.

A more complex circuit

Figure 4-17 shows a diagram for a somewhat more complicated, real-world electronic circuit presented in a form intended to assist a troubleshooting technician. The circuit has a single NPN bipolar transistor along with some resistors and capacitors. Note that *test points* (abbreviated TP) exist at three different locations: TP1 at the emitter of the transistor, TP2 at the base of the transistor, and TP3 at the collector of the transistor. If you need to troubleshoot this circuit (which happens to be a low-power amplifier of the sort you might find in a vintage radio receiver or hi-fi set), you'll connect your VOM between chassis ground and each one of these three points, in turn. You'll carefully note the meter readings and compare them with known normal values.

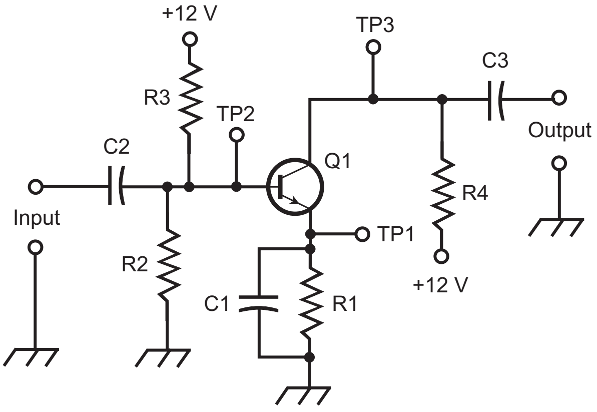

FIG. 4-17. *Schematic diagram of an amplifier circuit that includes component designators and three test point (TP) locations.*

Follow the flow

The circuit of Fig. 4-17 receives a weak input AC signal (such as the output of an ultrasonic pickup) and boosts it so that it can drive a device that consumes significant power (such as a switching circuit). The general signal flow goes from left to right. The original AC signal enters at the input terminals, passes through capacitor C2, and reaches the base (left-hand electrode) of transistor Q1. The base acts as an adjustable "current valve" that causes large current fluctuations through Q1 as the electrons flow from ground through R1 to the emitter (the bottom electrode), then onward to the collector (top electrode), and out through R4 to the positive power supply terminal. Capacitors C2 and C3 allow the AC signals to pass while blocking DC from the power supply so that the DC won't upset the operation of external circuits. Capacitor C1 keeps the transistor's emitter at a constant DC voltage while allowing the input signal to enter unimpeded. The resistors R2 and R3 have values carefully chosen to place precisely the right DC voltage, called *bias*, on the base of Q1, ensuring that the transistor will work as well as it possibly can in this application.

In many electronic circuits, actual voltages can deviate from design values by up to 20 percent; if this information is important, you'll usually find the error range at the bottom of the schematic drawing or in the accompanying literature. If the readings obtained are within this known error range (called the *component tolerance*), then you can tentatively assume that this part of the circuit is working properly. However, if the readings obtained are zero or well outside of the tolerance range, then you have pretty good reason to suspect a problem with the associated circuit portion, or possibly with other circuits that feed it.

Many schematic drawings that accompany electronic equipment, especially "projects" that you build from a kit containing individual components, include information that aids not only in troubleshooting, but also in the initial testing and alignment procedures that you must follow as soon as you've completed the physical assembly process. As a further aid, the literature might include pictorial diagrams that show you where each part belongs on the circuit board or chassis. That way, you can follow the circuit not only according to its electrical details, but along the physical pathways as they actually look.

According to standard schematic drawing practice, every component should bear a unique alphabetic/numeric label to designate it, as you see in Fig. 4-17. However, a few alternative labeling forms are also acceptable. Figure 4-18 shows the same circuit as the one in Fig. 4-17, but the parts list has been eliminated and the diagram contains no alphabetic/numeric designations. Instead, the compo-

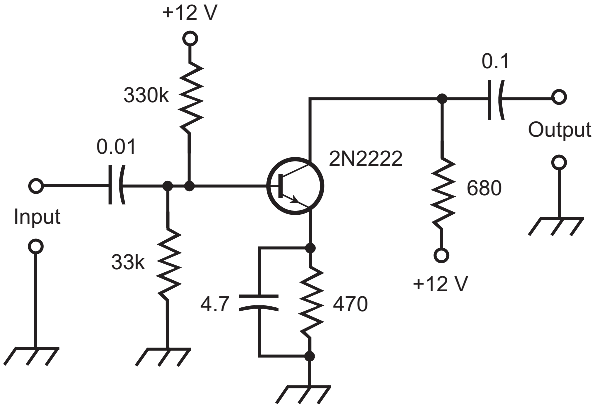

FIG. 4-18. *Schematic diagram of the amplifier circuit from Fig. 4-17, but with component values instead of sequential designators. All capacitances are in microfarads (μF). All resistances are in ohms (Ω); k = 1000.*

nents are identified only by their schematic symbols along with value designations or industry standard part designations. Using the example shown, we know that the transistor is a 2N2222 type, and that the resistors have values of 470, 33k (33,000), 330k (330,000), and 680 ohms. The input capacitor has a value of 0.01 microfarad (μF) and the output capacitor has a value of 0.1 μF. The emitter capacitor, which goes across the 470-ohm resistor, has a value of 4.7 μF.

Tip

In situations like the one in Fig. 4-18, you'll usually see a statement at the bottom of a schematic diagram that includes information about the units for the value designations. Such a statement might read "All capacitors are rated in microfarads (μF). All resistances are given in ohms (Ω), where k = 1000 and M = 1,000,000."

Schematic/block combinations

Once in awhile, you'll encounter a "hybrid" drawing that consists of a block diagram and a schematic diagram combined. Figure 4-19 shows an example of this approach, which works well when you want to highlight and explain a particular circuit, and to clarify its relationship to other circuits in a system. The detailed schematic portion of Fig. 4-19 shows a buffer amplifier of the sort you'll find in a radio transmitter. An oscillator (which precedes the buffer) and an amplifier (which follows the buffer) are represented as blocks with labels inside.

This diagram serves two purposes. First, as you read the schematic portion, you can study the actual component makeup of the buffer circuit. Second, you get a good idea as to the buffer's place in the overall system relative to the other circuits. The block/schematic representation in Fig. 4-19 clearly shows you that the buffer receives its input from a crystal-controlled oscillator, and also that the buffer sends its output to an amplifier. Another schematic diagram and block diagram combination might describe another portion of this same device, which you might recognize as a simple radio transmitter.

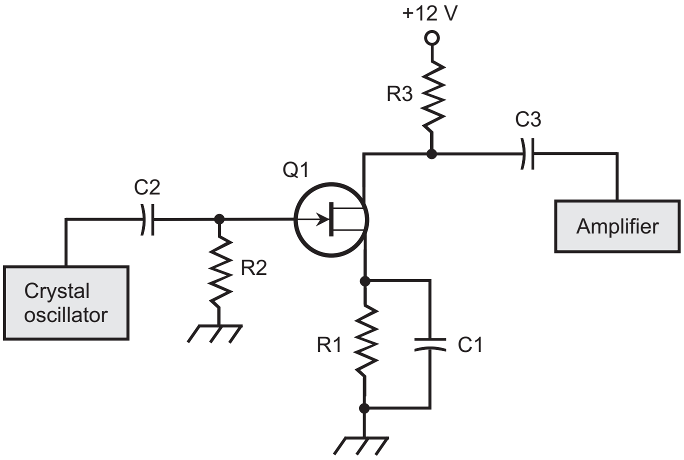
+12 V
R3
C3
Q1
C2
Amplifier
R2
Crystal oscillator
R1
C1

FIG. 4-19. *A combination block/schematic diagram of an oscillator, buffer, and amplifier system, showing the schematic details of the buffer portion.*

Figure 4-20 is a schematic/block "hybrid" diagram in which the oscillator is portrayed in schematic detail, but the buffer and the amplifier are represented only as blocks. This figure tells you that the oscillator output goes to the buffer, which then sends the signal along to the amplifier. The only new information obtained here is contained in

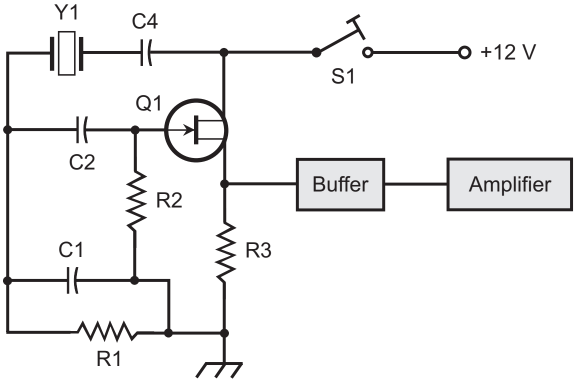
Y1
C4
S1
+12 V
Q1
C2
Buffer
Amplifier
R2
C1
R3
R1

FIG. 4-20. *A block/schematic diagram of the system from Fig. 4-19, showing the schematic details of the oscillator portion.*

the schematic representation of the oscillator. If you look at Figs. 4-19 and 4-20 together, however, you can picture in your mind a diagram in which all of the system's detail is included up to the point where the signal enters the amplifier. The oscillator portion of Fig. 4-20 also tells you that the oscillator is designed to generate Morse code signals because it contains a telegraph key!

Now let's "shift gears" and recall the block diagram of the AC-to-DC converter that you saw all the way back in Fig. 2-1. Figure 4-21 shows the schematic representation of that device, along with a duplicate of Fig. 2-1 to save you the trouble of having to flip pages. As you compare the two diagrams, note that in the schematic (at A), all of the actual components are shown, rather than merely portraying the stages as labeled blocks (at B). Figure 4-22 is a more comprehensive diagram that shows how the block diagram of this same device relates to the schematic diagram, while revealing all of the details originally present in both diagrams.

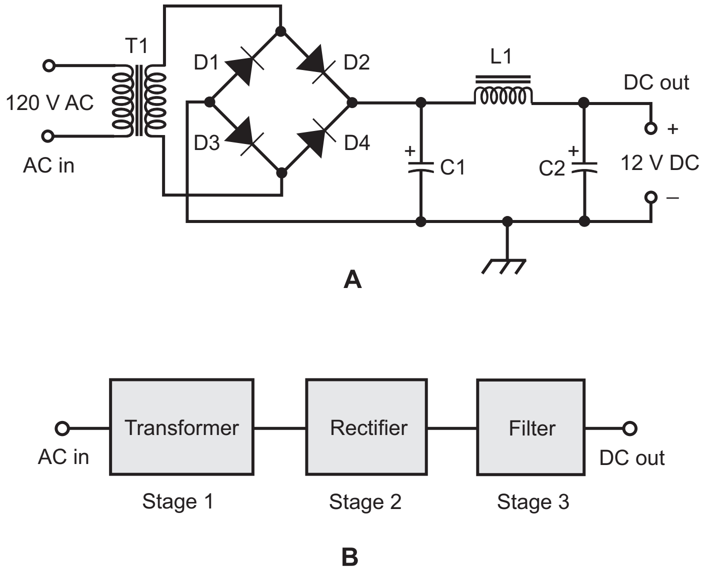

FIG. 4-21. *Schematic diagram of an AC-to-DC converter (at A), such as the one portrayed in the block diagram all the way back in Fig. 2-1 (reproduced here at B).*

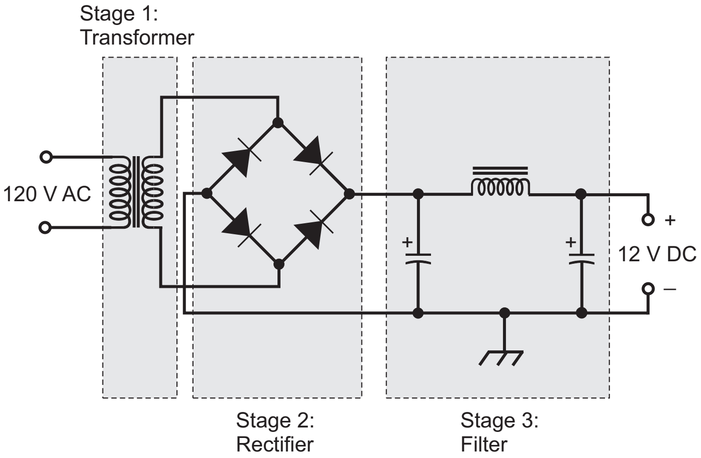

FIG. 4-22. *Here's how the block diagram of Fig. 2-1 relates to the schematic diagram of Fig. 4-21. (Component designators have been omitted to avoid clutter.)*

Follow the flow

To briefly explain the AC-to-DC converter circuit, more commonly called a power supply, the input electricity enters at the extreme left-hand end of all versions of the diagram (Figs. 4-21 and 4-22). The AC goes to the transformer to set up the proper ratio for conversion to DC. The four diodes in the diamond configuration constitute the rectifier, which converts the AC to pulsating DC, making use of both halves of the cycle. The final stage, the ripple filter, acts to smooth out the pulsations in the DC after the conversion. The "smoothed-out" or "pure" DC then goes along to the output terminals at the right-hand end, where it appears as a voltage just like the electricity that you'd find at the terminals of a 12-V battery.

Summary

This chapter has provided a brief sampling of simple electronic circuits, showing how you can read diagrams to learn how circuits are put together, or draw diagrams on the basis of known circuit details.

After a bit more training, you should be able to view a simple schematic and easily visualize what the finished circuit will look like, maybe even getting an idea of how you might go about arranging the components on a circuit board (even though the schematics or block diagrams don't explicitly tell you what to do in that respect).

You know that a roadway, as you drive along it in your car, does not look like the line that represents it on a map. You can visualize a secondary road or even a superhighway and do so whenever you see their symbols (although there's nothing like traveling along a road in both directions to learn all about it). The same principle holds true when reading schematic diagrams. When you see a bunch of schematic symbols in a diagram, you can visualize the component interconnections and perhaps make a few mental notes on the physical aspects of the circuit construction (although there's nothing like looking at, and using, a finished electronic device). Schematic symbology will become a new language for you—one in which you will eventually learn to think, just as you can think in terms of words, mathematical equations, musical scores, architectural blueprints, Morse code, or any other language.

5

Complex circuits

As you learn to read and draw schematic diagrams, you might get discouraged by the complexities that arise. You might think, "Oh sure, anyone can learn to read schematics of simple circuits, like those that have a transistor or two, a few capacitors, and a few resistors. But it must be impossible to decipher more complicated schematics without decades of experience." That's not true! You'll have to put some effort into the learning process, but you can always break a large, complicated system down into smaller, simpler circuits.

Identifying the building blocks

Even a circuit whose diagram looks overwhelming at first glance comprises multiple building blocks interconnected in an orderly, logical way. This principle never fails! Each building block represents a simple circuit. A device that contains six diodes, 10 inductors, 15 transistors, and dozens of resistors and capacitors might break down into 15 simple circuits, each containing a single transistor, a few resistors and capacitors, and a diode and inductor or two. If you look at the entire system schematic all at once, it's like trying to eat a jumbo hamburger in a single swallow. With the burger, and with the diagram as well, you face a far easier job if you assimilate the thing in little bites or pieces.

Tip

Nowadays, circuits of great complexity are etched onto semiconductor wafers and housed in *integrated circuits* (ICs), also called *chips*, that come in packages that actually look like building blocks! As you examine a modern circuit board, you might imagine that the whole system was designed and put together from a block diagram, not a schematic! When you see the complete schematic, it will contain rectangles representing the chips, along with a great many lines connecting them. In most real-world circuits, those lines represent foil runs on a printed circuit board.

Figure 5-1 shows a "crystal radio" receiver built with an antenna, a tapped air-core inductor, a variable capacitor, an RF diode, and a fixed capacitor. The term "crystal" comes from the original construction of RF diodes as they emerged in the early 1900s. In order to get a one-way current gate to act as a signal *detector* (or *demodulator*), radio experimenters placed a piece of fine wire called a *cat's whisker* into contact with a piece of crystalline lead sulfide called *galena*. Today, semiconductor diodes do the same thing as those old "crystals" did, even though modern RF diodes don't look anything like crystals.

You can't call the circuit of Fig. 5-1 "complicated" by any stretch of your imagination, but it performs some sophisticated tricks nevertheless! Aside from the antenna (an external component such as a wire running out your window to a tree) and a sensitive earphone or headphone that you can connect at the output terminals in order to

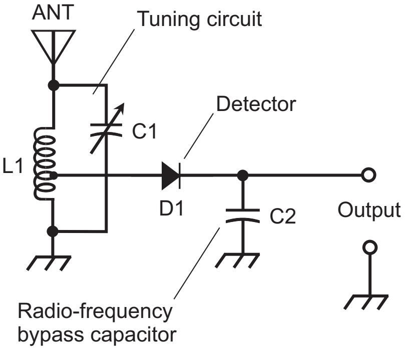

FIG. 5-1. *Schematic diagram of a "crystal radio" tuned circuit and detector stages. It produces an extremely weak audio signal.*

hear radio stations (however faintly), this circuit contains only four components: the coil, the diode, and two capacitors.

You can connect an amplifier to the output of the "crystal radio" in order to boost the audio volume to the point where it can drive a headset to a comfortable listening level. (That's the reason why Fig. 5-1 doesn't include a headset at the output.) To accomplish that feat, you'll need another circuit, along with a battery or other source of DC power to make the output signal strong enough.

Follow the flow

In the "crystal radio," RF energy from the antenna *resonates* (reverberates) in L1 and C1. The *resonant frequency* depends on the size of the coil and also on the setting of C1. That frequency determines which station you hear, if any. Diode D1 converts the signal to pulsating DC that contains both the RF and the audio from the original transmitter. Capacitor C2 shorts the RF part of the signal to ground, leaving only audio-frequency (AF) energy at the output. The position of the coil tap must be adjusted experimentally for maximum output.

Figure 5-2 shows another fairly simple schematic diagram: an audio amplifier that makes use of a single NPN bipolar transistor. In addition, this circuit has four resistors and three capacitors for a total of eight components. It also needs a source of DC power, such as a battery, which provides 12 V. This circuit can accept low-level audio signals (the output of a "crystal radio," for example) at the input terminals and boost the power to a level strong enough to make a nice loud sound come out of a headset. It might take an experienced engineer two or three minutes to draw the diagram, and a couple of hours to build and test the complete circuit, "tweaking" component values to get the best possible performance.

Because the circuit of Fig. 5-2 takes an extremely weak signal and boosts it to a reasonable (but not particularly powerful) level, it's sometimes called a *preamplifier*. If you want to drive a set of speakers so that the students in a classroom, or the audience in an auditorium, can hear the output from the "crystal radio," you'll need more amplification. That extra audio boost can be provided by one or more additional circuits called *power amplifiers* connected to the output of the preamplifier.

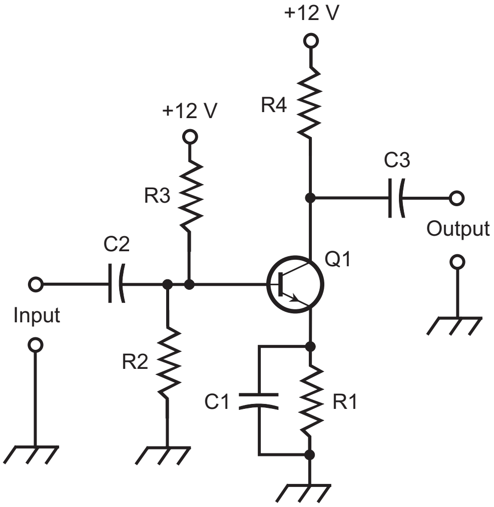

FIG. 5-2. *An audio preamplifier circuit that can be used with the "crystal radio" to produce a signal strong enough to drive a headset (but not a loudspeaker).*

Follow the flow

In the circuit of Fig. 5-2, audio AC passes from the input through C2 to the base of Q1. The capacitor C2 keeps the power supply's DC from affecting the behavior of the previous circuit (the "crystal radio" set itself). The tiny current variations at the base of Q1 cause larger current variations through the transistor. The amplified audio signal passes through C3 to the output. The resistors govern the current flowing through Q1, and must be chosen by experiment for optimum amplification. Capacitor C1 keeps the emitter at audio signal ground while allowing some DC voltage to exist there.

Figure 5-3 is a diagram of a circuit that looks, at first glance, more complicated than Fig. 5-1 or Fig. 5-2. But is it really? If you look at Fig. 5-3 for a minute or two, you'll notice that it's nothing more than the composite of the "crystal radio" and the audio preamplifier that we've just seen. The components are renumbered generally going from left to right, the direction of signal flow through the system. (You never want to duplicate a component designator in any schematic.) In Fig. 5-3, the connection between the original "crystal radio" and the preamplifier corresponds to the short, horizontal connecting line that goes from the dot above C2 to the left-hand side of C3.

Now that you can envision the two building blocks that make up the circuit of Fig. 5-3, the whole diagram looks less complicated, doesn't it? You can follow the signal flow through the system by thinking of the signal going through the "crystal radio" and then through the audio preamplifier. Of course, the process goes so fast that it would seem instantaneous if you could actually see the RF and audio impulses. The whole process, from the signal arriving at the antenna to the audio appearing at the output, takes place in a minuscule fraction of a second. Currents in electrical conductors, in general, travel at roughly 10 percent of the speed of light in free space. That's 30,000 kilometers, or 18,600 miles, per second!

The circuit of Fig. 5-3 has a low audio output power level, although it's a lot louder than the feeble audio signal that comes from the "crystal radio" all by itself, which gets its power only from the signal current in the antenna! Nevertheless, even the amplified audio at the output of the circuit of Fig. 5-3 isn't enough to provide a comfortable listening volume in a loudspeaker. In order to boost the audio power level some more, you'll need an *audio power amplifier*. Figure 5-4 shows a two-transistor circuit that will perform this task. It's called a

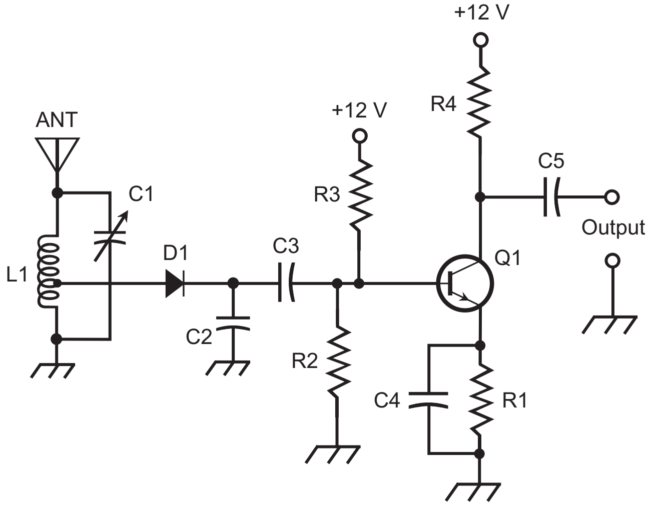

FIG. 5-3. *Combination of "crystal radio" and audio preamplifier circuits. Some of the component designators in the preamplifier stage are updated from Fig. 5-2.*

push-pull circuit. One of the transistors amplifies the "positive" half of the AC audio wave, and the other transistor amplifies the "negative" half. You might say that Q1 does the "pushing" and Q2 does the "pulling," so that when you combine their outputs, you get an amplified version of the complete audio input wave.

The circuit of Fig. 5-4 will accept the output of a low-level audio amplifier and boost it. However, the output from the original "crystal radio" would not be enough to drive the power amplifier. When you hear the term *power amplifier*, you should remember that the circuit does exactly what its name implies. It takes an input signal with a certain amount of *power* and produces an output signal that has *more power*. If the original input signal contains little or no power, then the circuit of Fig. 5-4 won't get enough *drive* (input power) to produce any output signal. In order to work properly, a power amplifier needs a signal that has a modest amount of power to begin with; otherwise it will simply "suck up" the feeble input signal and produce nothing from it other than a tiny amount of heat in the components. The circuit of Fig. 5-3 (the audio preamplifier) provides enough output power to adequately drive a push-pull audio amplifier. The circuit of Fig. 5-1 (the "crystal radio" alone) doesn't.

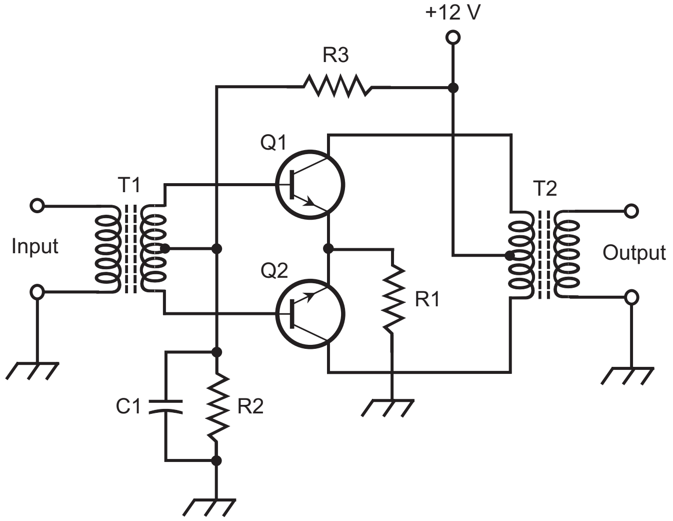

FIG. 5-4. *Audio power amplifier circuit, suitable for driving a speaker or set of speakers.*

Follow the flow

In the circuit of Fig. 5-4, the input signal goes through T1 and appears across its secondary. The details of the ensuing process get rather complicated, involving wave phase and transistor characteristics—technical matters that go beyond the scope of this book. Basically, Q1 handles half of the signal cycle and Q2 handles the other half, as the audio input AC flows alternately up and down through the secondary winding of T1. Transformer T2 combines the two amplified cycle halves back into a complete AC audio wave that's stronger than the wave that came in through T1, thanks to the transistors.

Now if you combine the circuits from Figs. 5-1, 5-2, and 5-4 in *cascade* (one after the other), you get a complete amplitude-modulation (AM) radio receiver that will produce decent sound from a loudspeaker! Figure 5-5 shows the entire three-transistor AM radio receiver in a single schematic diagram. Again, some of the component designators are changed from previous diagrams, so that they increase generally as you go from the original input at the antenna to the final output at the speaker.

So there!

If you were to have seen Fig. 5-5 at the start of this chapter, you might have experienced some major frustration. But now that you can see how the building blocks go together, you know that you don't have to "choke down the whole burger in one gulp." So it's pretty simple after all, isn't it?

The basic process used to make all electronic circuits breaks down into a sequence of combinations. First, the individual components (resistors, capacitors, diodes, and so on) combine to form simple circuits. Then, simple circuits combine to make more complex circuits. After that, complex circuits combine to form complete devices. Several different devices can combine to create a large system. An amateur radio station is a good example of such a system. It might consist of a *transceiver* (transmitter/receiver in a single box), an *antenna tuner*, a computer, an *interface* unit that goes between the computer and the transceiver, and a *speech processor* that goes between the microphone

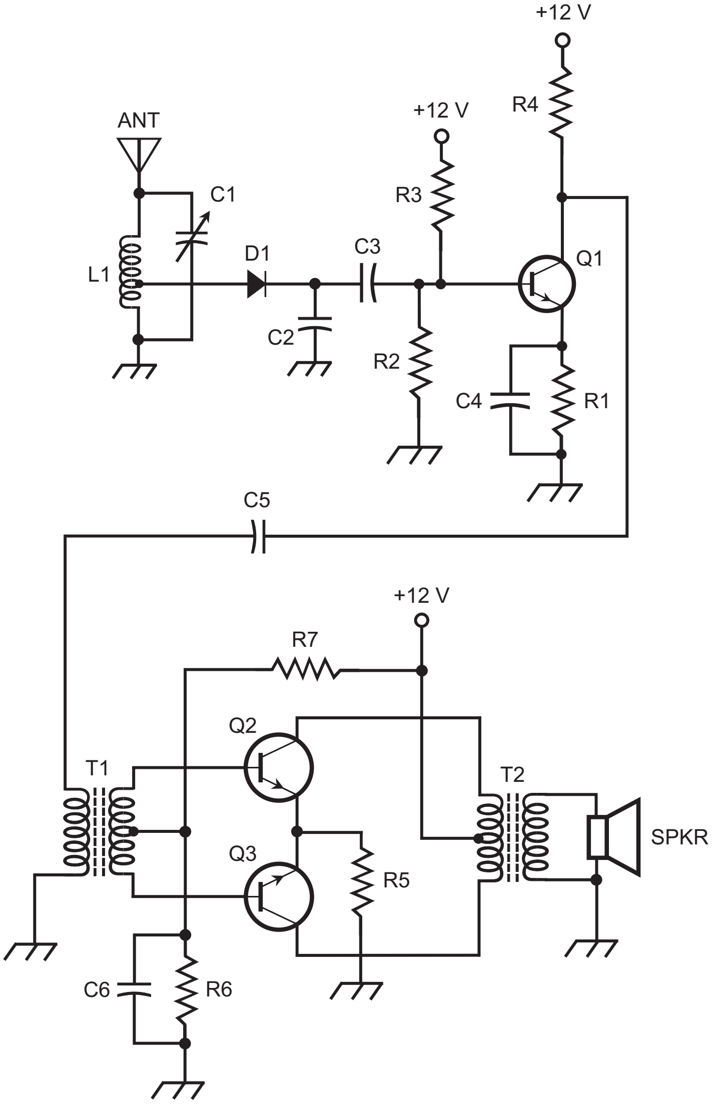
+12 V
R4
ANT
+12 V
C1
R3
D1
C3
Q1
L1
C2
R2
C4
R1
C5
+12 V
R7
Q2
T1
T2
SPKR
Q3
R5
C6
R6

FIG. 5-5. *Complete radio receiver circuit. Some of the component designators in the audio power amplifier stage are updated from Fig. 5-4. The speaker is also shown.*

and the transceiver. Each of these individual devices contains numerous complex circuits, which in turn, comprise multiple simple circuits, which ultimately break down to the individual resistors, capacitors, diodes, transistors, and other components.

Page breaks

Figure 5-5 is a "respectably complicated" diagram. It requires us to draw the system in a "two-story" format with the detector and preamplifier on top, and the audio power amplifier on the bottom. A long, tortuous line, broken in the middle by C5, represents the connection between the preamplifier output and the power amplifier input. There's nothing technically wrong with this diagram, but some people might rather see it all on one level. In order to put Fig. 5-5 all on one level, we'd either have to make it horribly small, or else draw it sideways on the page. But there's another option. We can produce it on a foldout page (the sort of thing that they do in those upscale print magazines when they want to show you something spectacular).

We have yet another alternative, though! We can split the diagram up among multiple pages. That approach isn't necessary here, but when we get to truly complicated systems, such as amateur radio transceivers, television sets, or complete computers, it's an option that engineers often use. Figure 5-6 shows how we can take advantage of this technique with the diagram of our radio receiver. Figure 5-6A puts the detector and the audio preamplifier right-side-up on a single page along with an output designator that appears as an X inside a wedge-like arrow that points off the page toward the right. Figure 5-6B shows the audio power amplifier with an input designator comprising an X inside a wedge-like arrow that points off the page toward the left.

Tip

In Figs. 5-6A and 5-6B, the wedge-like arrows represent points meant to connect directly to each other. They take the place of the long, tortuous line in Fig. 5-5. In this case, we need only one set of arrows of this type. Some circuits need two or more, so we might label them as X, Y, and Z, for example. Then the two X's would connect, the two Y's would connect, and the two Z's would connect.

Let's follow the signal through Fig. 5-6A. A radio wave causes current to flow in the antenna, and also through the inductor L1. Capacitor C1 causes the inductor/capacitor combination (called an *LC circuit*, where L stands for inductance and C stands for capacitance) to resonate at the frequency of the radio signal that we want to hear. Diode D1 detects the RF signal, splitting the AF and the RF portions apart. Capacitor C3 passes the AF part of that signal along to the base of transistor Q1. Capacitor C2 shunts the RF portion of the diode's output to ground because the circuit doesn't need the RF energy anymore, and its presence would only cause trouble. Transistor Q1 acts as an amplifier for the extremely weak AF signal at its base. Resistors R1, R2, R3, and R4 ensure that Q1 gets optimum DC voltage (called *bias*), so that it will produce the greatest possible amount of amplification. Capacitor C4 keeps the emitter at AF signal ground, while allowing some DC voltage to exist there. The AF output signal, along with some DC from the power supply (+12 V), goes off the page through the rightward-pointing arrow marked X.

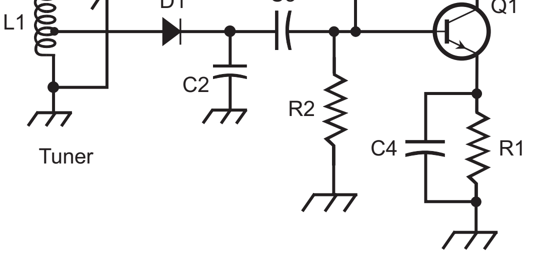

FIG. 5-6A. *Tuner, detector, and audio preamplifier stages in the radio receiver. The wedge X represents an extension to illustration B on the next page.*

Now let's look at Fig. 5-6B and follow the signal after it comes in from the previous diagram. The AF energy, along with some DC, appears at the leftward-pointing arrow marked X. Capacitor C5 blocks the DC so only the AF wave gets to potentiometer R8, which serves as a volume control. The full AF voltage appears across the entire resistance of R8, represented by its left-hand and right-hand terminals. The slider "picks off" various AF voltages that can range from zero (all the way to the right-hand end of the zig-zag, at ground) to the full audio voltage (all the way to the left-hand end of the zig-zag, at C5). This AF voltage goes to the primary of transformer T1. From there, the signal flows in precisely the same way as described in "Follow the flow" for Fig. 5-4. The only difference between this situation and that one lies in the numbering of the component designators. We continue going on upward in numbers from the previous part of the system. We also add a speaker to the output of T2, rather than simply labeling it "Output."

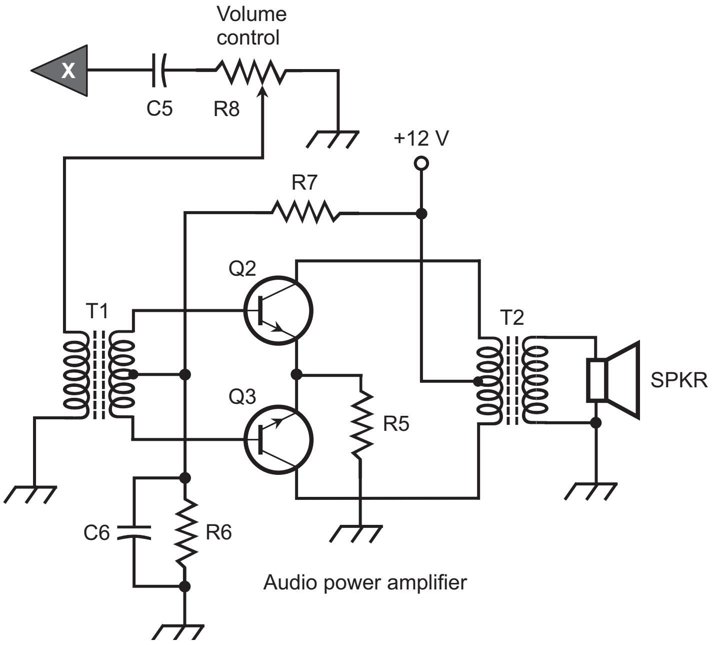

FIG. 5-6B. *Audio power amplifier and speaker in the radio receiver. The wedge X represents an extension from illustration A on the previous page.*

Tip

While looking at Figs. 5-6A and 5-6B, have you noticed that we've added labels to each stage to describe its function? You should do that whenever you draw diagrams of multistage systems such as this one, as long as you have enough room on the page. In Fig. 5-6A we see the labels "Detector" and "Audio preamplifier;" in Fig. 5-6B we see the label "Audio power amplifier."

Some more circuits

Figure 5-7 shows an antenna matching circuit known as an *L network*. Here, the letter L refers to the general layout of the components in the diagram, not the property of inductance. (Actually, to make the coil and capacitor in Fig. 5-7 actually take the shape of an L in the layout, you'll have to rotate the page 90 degrees clockwise and then hold it up to a mirror! But you get the general idea, right?)

Figure 5-8 shows another type of antenna matching network, which consists of the circuit from Fig. 5-7 with an extra capacitor added at the input end. Engineers sometimes call this type of circuit a *pi network* because its components, in the schematic layout, resemble the shape of the upper case Greek letter pi (Π). Both the L network and the pi network are commonly used in antenna tuning and matching circuits for radio transmitters at frequencies ranging up to around 150 MHz.

Figure 5-9 shows a circuit that's a little more complicated than the ones in Figs. 5-7 and 5-8, but in a certain sense, contains both of them

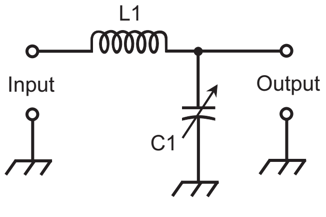

FIG. 5-7. *An L network comprising an inductor and a variable capacitor.*

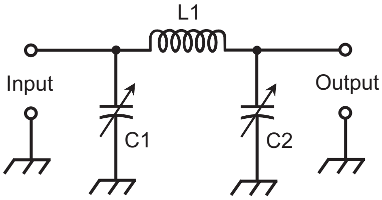

FIG. 5-8. *A pi network comprising an inductor and two variable capacitors.*

put together. When we follow a pi network with an L network, we get a so-called *pi-L network*. The advantage of a circuit like the one in Fig. 5-9, compared to those in the previous two diagrams, lies in its ability to make radio transmitters work with antennas that would otherwise not accept power very well. Those two extra components can go a long way!

Tip

The circuits of Figs. 5-7, 5-8, and 5-9 can all be made more versatile by using variable inductors rather than fixed inductors. One type of variable inductor has become popular among radio amateurs. It's called a *roller inductor*. Just for fun, try an Internet search on "roller inductor" by entering the term in the phrase box of your favorite search engine. You should find some images of these components. They allow for precise adjustment of inductance, and some of them have calibrated crankshafts so that you can easily reset them to any previous position.

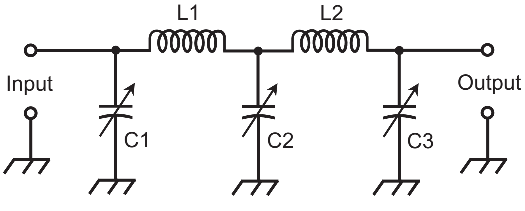

FIG. 5-9. *A pi-L network comprising two inductors and three variable capacitors.*

If you're old enough, you'll remember the days when you had to learn the *International Morse code* (often simply called "the code") to get an amateur radio operator's license. Those days have passed into history, but some radio amateurs still enjoy conversing in this mode. In order to do that, of course, you need to learn the code. To that end, you can build a *code practice oscillator,* such as the one diagrammed in Fig. 5-10. Basically, this circuit is an audio oscillator that you can switch on and off with a Morse code *straight key* or *telegraph key*, labeled "Key" in the figure. (Because a telegraph key technically constitutes an SPST switch, you could label it S1.)

When you first examine Fig. 5-10, you might wonder why an audio oscillator circuit has to be so complicated. Can't you just build a simple audio amplifier, like the preamplifier or power amplifier discussed earlier, and feed some of the output back to the input? Well, yes, you can in fact do that; but if you want a decent audio tone to

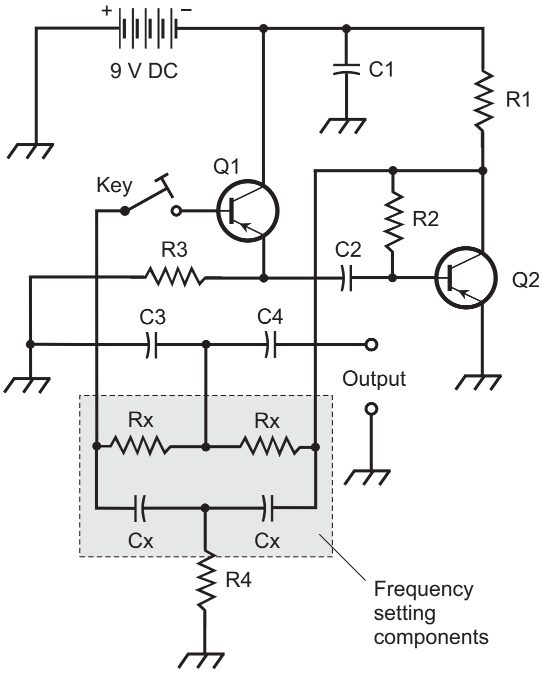

FIG. 5-10. *An audio code-practice oscillator using two PNP bipolar transistors. The values of Rx and Cx determine the frequency.*

come out of your code practice oscillator, you'll get a lot better results with a circuit like the one in Fig. 5-10. This particular circuit is called a *twin-T oscillator* because of the T-shaped configurations including the resistors marked Rx and the capacitors marked Cx. The twin-T oscillator produces an AF tone that's pleasing to the ears, and that also occurs at a predictable and stable *pitch* (frequency).

Follow the flow

The AF signal in the circuit of Fig. 5-10 goes around and around; that's how oscillation happens. We can start pretty much anywhere and follow the signal back to the same point. If we begin at the key, the signal goes into the base of Q1, where it's amplified and undergoes *phase inversion* (the wave is turned upside down). The output of Q1 comes from the emitter of Q1, rather than from the collector; that's a matter of engineering choice because it provides a stable and reliable circuit. The signal then goes into the base of Q2, where it's amplified and inverted again, so it comes out of Q2 in *phase coincidence* (reinforcing rather than opposing) with respect to the signal at the key. The signal emerges from the collector of Q2 and gets routed down to the twin-T network comprising resistors Rx and capacitors Cx, whose values determine the frequency of oscillation. From there, the signal goes back to the key, and gets ready for another round trip through the whole circuit! We take the output through C4, from the point between the two resistors Rx.

The circuit of Fig. 5-10 uses a 9-V battery as its power source. Note that the transistors are of the PNP type, so the collectors get a negative voltage, while the positive battery terminal goes straight to ground. This circuit, therefore, constitutes a *positive-ground system.* We might build a power supply designed to produce -9 V DC from the AC utility mains (about 120 V AC at 60 Hz in the United States), rather than relying on a battery. If we decide to do that, we'll need to make sure that we design the power supply so that it produces a negative voltage, not a positive voltage, with respect to ground.

Figure 5-11 shows a power supply that will do a good job of providing a constant, pure –9V DC. It's just about the same circuit as the one you saw back in Chap. 4 at Fig. 4-13. The basic component

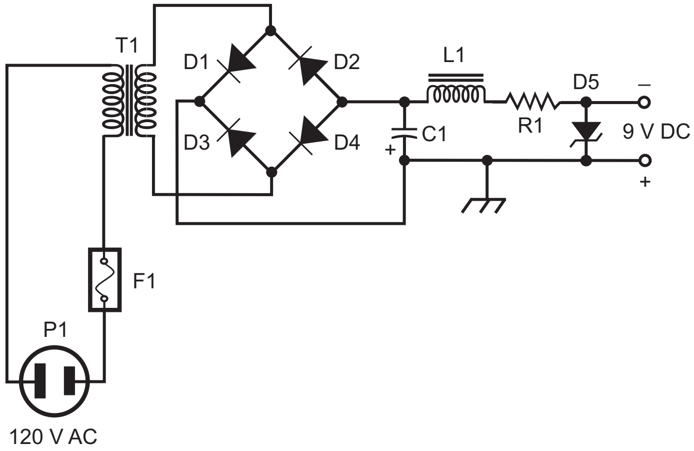

FIG. 5-11. *A regulated −9 V DC power supply that can be used with the code-practice oscillator. Note the positive ground in this supply, designed for use with PNP transistor circuits.*

layout is the same, except that all the diodes go in the opposite direction in Fig. 5-11 as compared to Fig. 4-13 (including the Zener diode), and the electrolytic capacitor polarity is reversed. And of course, the voltage of the power supply shown in Fig. 5-11 is lower than the voltage of the supply shown in Fig. 4-13.

Figure 5-12 shows a complete code practice oscillator system that will operate from the AC utility mains. It includes the power supply from Fig. 5-11 along with the oscillator from Fig. 5-10, all in one big schematic. As we did with the radio receiver circuit diagram of Fig. 5-5, we have connected the power supply output to the oscillator in Fig. 5-12 with a single line, although not as long as the one in Fig. 5-5. We've also included a volume control and a pair of headphones at the output of the twin-T oscillator circuit.

We can apply the same general layout tactics to Fig. 5-12 as we used with Fig. 5-5. We might draw the diagram with the page tilted sideways (landscape orientation instead of portrait orientation), putting the power supply on the left and the oscillator on the right. Or we might use a foldout page. However, an even better alternative (in my opinion) is to split the diagram up into multiple pages so we can draw everything in an uncluttered, right-side-up fashion.

Let's take this business a step further. We can boost the output of the twin-T oscillator so that it will drive a loudspeaker, letting us

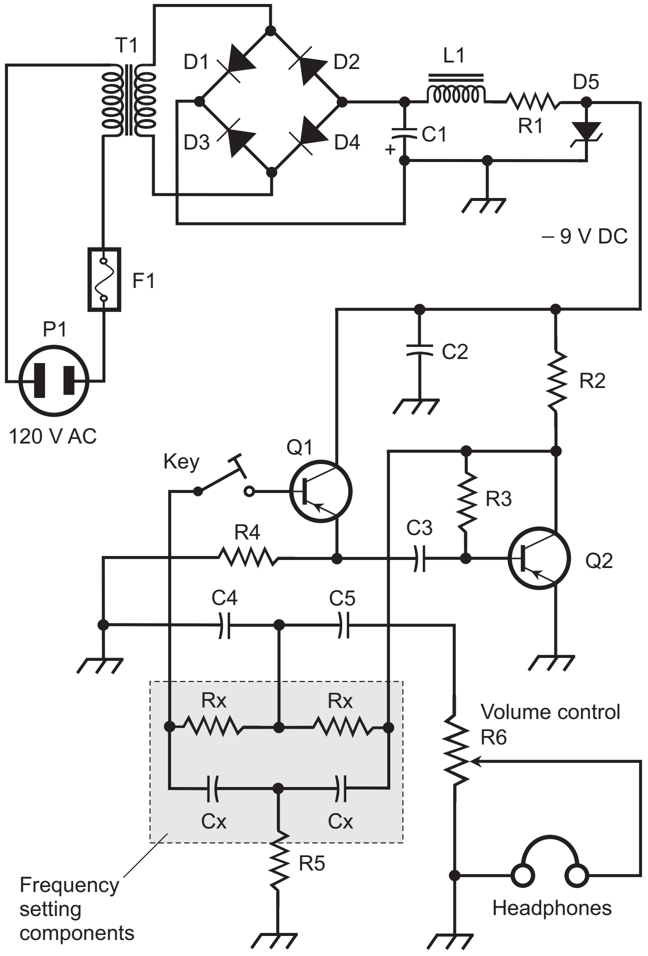
T1
D1
D2
D3
D4
L1
R1
D5
C1
+
– 9 V DC
F1
P1
120 V AC
C2
R2
Key
Q1
R3
R4
C3
Q2
C4
C5
Rx
Rx
Cx
Cx
Volume control
R6
R5
Frequency
setting
components
Headphones

FIG. 5-12. *Combination of the regulated power supply and code-practice oscillator. Note the addition of the volume control and headphones.*

send code to a classroom full of eager Morse-code students! (Finding enough people interested in learning the code to fill up a whole classroom would constitute an entirely different challenge.) A push-pull audio amplifier, just like the one we used for the radio receiver (Fig. 5-4) except with PNP rather than NPN transistors, will work here. Now we have a system with three essential circuits: a power supply, an oscillator, and an amplifier. Figures 5-13A, 5-13B, and 5-13C show how we can draw the complete schematic diagram of this system, letting it spread across three different pages.

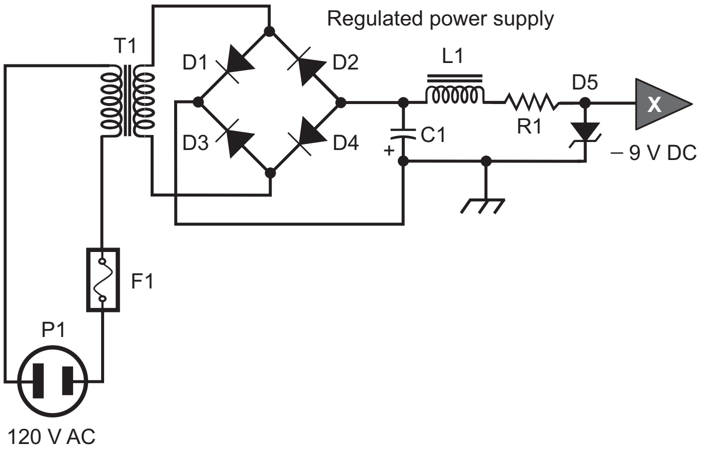
Regulated power supply
T1
D1
D2
D3
D4
L1
D5
R1
C1
+
X
– 9 V DC
F1
P1
120 V AC

FIG. 5-13A. *Regulated power supply for a classroom code-practice system. The wedge X represents an extension to illustrations B and C on the next two pages.*

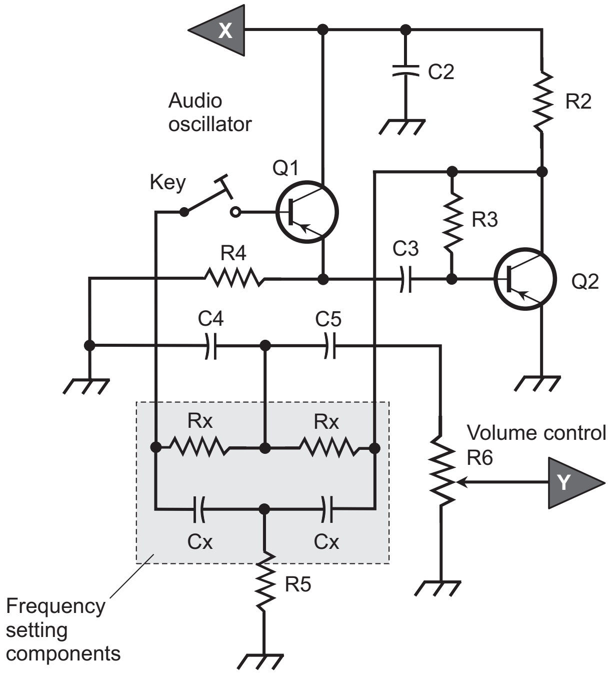
X
C2
R2
Audio
oscillator
Q1
Key
R3
R4
C3
Q2
C4
C5
Rx
Rx
Volume control
R6
Y
Cx
Cx
R5
Frequency
setting
components

FIG. 5-13B. *Twin-T audio oscillator for a classroom code-practice system. The wedge X represents an extension from illustration A on the previous page. The wedge Y represents an extension to illustration C on the next page.*

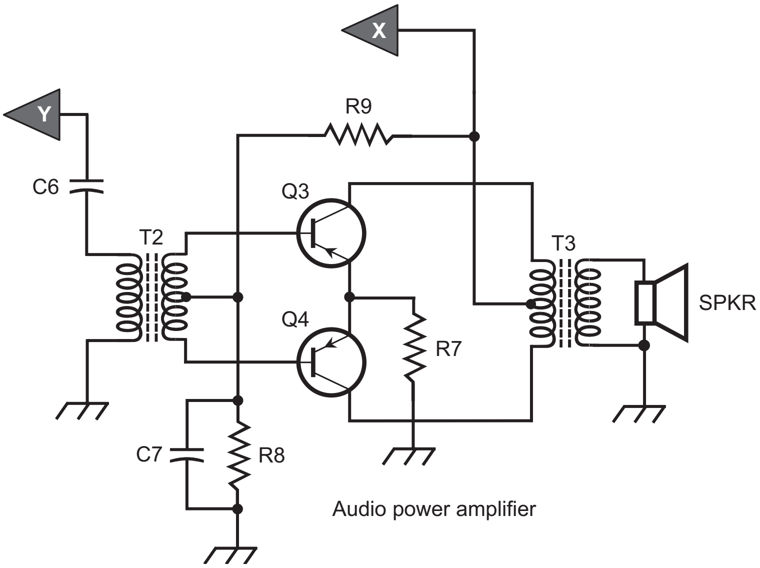
X
Y
R9
C6
Q3
T2
T3
SPKR
Q4
R7
C7
R8
Audio power amplifier

FIG. 5-13C. *Audio power amplifier for a classroom code-practice system. Note the PNP transistors, consistent with the negative power-supply voltage (positive-ground system). The wedge X represents an extension from illustration A on the page before last. The wedge Y represents an extension from illustration B on the previous page.*

Heads up!

As you examine Figs. 5-13A, B, and C, take careful note of the designators inside the arrows. The right-pointing arrow X in Fig. 5-13A goes to two places on subsequent pages: the left-pointing arrow X in Fig. 5-13B and also the left-pointing arrow X in Fig. 5-13C. The arrows marked Y apply only to Figs. 5-13B and C.

Figure 5-14 shows a simple *inductance-capacitance* **(LC) circuit**. Note its resemblance to the L network of Fig. 5-7. In this case, the inductor and capacitor have been reversed from their relative positions in the earlier case. In addition, this circuit has a different purpose than the other one does. The circuit in Fig. 5-7 works mainly for the purpose of tuning an antenna system, or matching the output of a transmitter to the characteristics of a particular antenna. The circuit in Fig. 5-14 is designed to let signals pass through (or not) depending on their frequency. In this case, we have a so-called *highpass filter* because it lets signals go through more easily as the frequency increases. The exact frequency at which the transition from high attenuation (or lots of signal loss) starts to change over to low attenuation (little or no signal loss) depends on the values of the capacitor and inductor.

Figure 5-15 shows a more complicated LC circuit that consists of two filters, one after the other. The first filter, made up of capacitor C1 and inductor L1, is the same highpass filter design as the one in Fig. 5-14. The second LC combination, made up of inductor L2 and capacitor C2, forms a *lowpass filter*. It works in the opposite manner from a highpass filter, letting signals through more easily as the frequency goes down. When we follow a highpass filter with a lowpass filter, and if we choose the *cutoff frequencies* (or transition points)

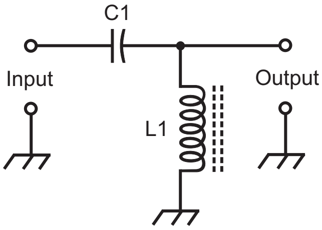

FIG. 5-14. *A simple frequency-sensitive filter circuit.*

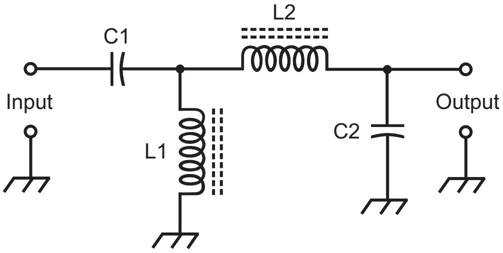

FIG. 5-15. *A complex frequency-sensitive filter comprising two simple, but different, filters connected in cascade.*

so that the highpass filter's cutoff frequency lies below the lowpass filter's cutoff frequency, we can get a *bandpass filter* in which a signal can get through easily only when its frequency lies between the two cutoffs.

Tip

You'll often find repetition of circuit configurations, one after another, in electronic system design. Sometimes you'll see several identical or similar circuits connected together. Sometimes their components will all have the same values, and sometimes they won't. Sometimes the circuits will be connected in series (end to end, like the links in a chain); sometimes they'll be connected in parallel (across each other, like the rungs in a ladder).

If you know how one circuit in a repetitive system operates, then you know, by extension, how all of the circuits work. A problem that occurs in one circuit might also occur in any of the others, and you can use the schematic diagram to track down the problem. For example, if, through testing, you know that an oscillator has changed frequency because of a defective resistor in the base circuit, then if another oscillator of the same configuration develops the same malfunction, you can consult the schematic, locate the base resistor, and conduct some tests to see if it, too, has gone bad. Without the schematic, you'd have a hard time locating the proper resistor. This single miscreant component that costs a few pennies, could, all by itself, bring down a sophisticated system!

Breaking a massive system down into complex circuits makes it easier to determine which circuit might be creating the problem for the entire system. This complex circuit is then broken down into simple circuits, and a determination is made as to which simple circuit might be at fault. The simple circuit is further divided into separate components, which are examined individually for a possible fault. Through this systematic method of elimination (all done with the help of schematics), the equipment can be repaired with a lot less trouble than you'd experience if you didn't have the schematics. The process of narrowing-down a problem zone from the system to a complex circuit, then to a simple circuit, and finally to an individual component, is called *troubleshooting to the component level.*

You're hired!

If you can troubleshoot large electronic systems to the component level, you'll find yourself in high demand. All you'll have to do is show up on time when someone calls you for service, do the job, and leave. When word gets around about your competence and reliability, you'll have a job for the rest of your life.

Many failures in complex electronic systems arise from a problem with a single component. Sometimes this failure will cause other components to go bad as well, but the repair must start with the first failure. Occasionally, two simultaneously defective components will be discovered, but this sort of coincidence is a rarity. Certainly, you must become familiar with the equipment you intend to repair, but after you have studied and understood the system's normal state of operation, you can use a schematic drawing to identify the general area of any future faults. By following this procedure, you can identify the most suspect individual component(s). Then all you have to do is find the suspect component(s) in the physical system, get into the equipment with the appropriate test instrument, and check the suspect component(s) one by one.

Tip

Even if you have good test equipment, you'll find it difficult to quickly identify defective sections (and especially individual components) without a schematic diagram because you probably won't know where on the circuit board or chassis to look!

Getting comfortable with large schematics

No matter how much you love electronics, you can't expect to sit down as a rank beginner and read complicated schematics straight away. You've got to climb a learning curve. First of all, you must make certain that you know all of the schematic symbols that you expect to deal with. Complex schematics can serve as a great learning tool because they contain lots of symbols, some of which you probably won't know. Although you can't sit down with a complex schematic at the start and understand everything that happens in the circuit that it represents, you can use such diagrams to help you learn the symbols.

Once you feel comfortable with the individual symbols, put away the complex schematics and start looking over diagrams of basic, simple, widely used circuits. You'll find lots of them in magazines geared to electronics enthusiasts. (You'll find some simple projects and related diagrams in *Electricity Experiments You Can Do at Home*, published by McGraw-Hill, as well!) Don't study only one type of schematic, such as those that portray only amplifiers. Check into oscillators, power supplies, solid-state switches, RF circuits, audio circuits, and anything else you can find. You'll discover similarities between different types of circuits, sometimes with no major differences other than a few changes in the component values. When you can identify an amplifier or oscillator or detector circuit by looking at its schematic, then you'll know you're making progress.

Once you can comfortably identify simple electronic circuits from schematic diagrams, you should move on to more complex drawings—but don't go overboard here! If you try to move too quickly, you might grow frustrated and give up altogether. Your next step will include devices that combine a few of the simple circuits you have previously studied. Sometimes additional components are added to match the output of one circuit to the input of another. Select books and publications that offer both theoretical and practical discussions of the circuit that the schematic depicts. Even better, you can build some simple circuits in a workshop at home.

You'll likely get a surprise when you see how your first "homebrewed" electronic circuit looks when compared with the schematic drawing. Your study will continue from this point by examining the functional circuit and noting the relationship of the physical components to those in the schematic drawing. You can further your electronics knowledge by experimenting with these circuits (substituting different components, for example). You might find a way to improve the circuit operation. All improvements should be duly noted, and a new schematic can be drawn up indicating your changes. You might want to simply pencil in the changes on the schematic diagram you were building from.

When you feel comfortable building electronic circuits from simple schematic diagrams, you might want to combine two or more circuits to make a single, more sophisticated device. Take two schematic diagrams from a projects book and combine them on paper. You will have to draw your own schematic to serve as your plan for the building procedure. You might know enough by this time to design and build some additional circuits that can *interface* the two (connect the output of the first circuit to the input of the second one so they both work at their best). When you combine electronics circuit-building with the task of learning to read and create schematics, you can improve your electronics knowledge faster than you can by simply looking at and drawing theoretical diagrams.

Tip

Admittedly, you might start to grow bored as you pore over schematic diagrams for hours on end. But when you can refer to a portion of a schematic drawing and then wire the components in place, much of the boredom will go away. If you're lucky, or if you happen to be a budding *technophile* (technology lover), the stuff will get downright fascinating!

Before you know it, you'll have obtained a solid basic knowledge of schematic diagrams and circuit-building. The circuits that you once imagined as complicated will soon seem familiar and elementary. Nevertheless, you should exercise some caution here. You might feel the temptation to stay with the sorts of circuits that you know best, and not venture into new territory. Don't let that little bit of laziness get the better of you! As soon as you reach one stage of comfort, move on to diagrams that are more difficult and make you feel a little bit uncomfortable again. If you don't "push yourself" this way, you'll find yourself stuck at one stage of development for a long time. Try building more and more complicated devices. Admittedly, this practice can grow expensive if you overdo it; so if you can't build everything in sight, keep on reading schematics and deciphering the various and diverse circuit components anyway.

You'll never cease to be amazed at what you know and what you don't know. For instance, many people new to electronics feel that a commercial AM radio transmitter must constitute a highly complex system. Most electronics novices are astonished to learn that the AM transmitter is technically less complex than an old-fashioned transistorized pocket receiver that you might use to intercept the broadcasts. A commercial radio transmitter is a rather simple system, even though it might have the size and mass of a full-sized refrigerator. The transmitter size is directly related to the component size, which in turn is directly related to the amount of power that the system consumes. The power-supply transformer for a commercial transmitter, all by itself, might weigh as much as the aforementioned refrigerator! This and other components make the commercial broadcast transmitter necessarily large. However, the power-supply transformer for a small amateur radio transmitter will likely weigh less than a kilogram. Schematically, both transformers look the same in schematic diagrams. Circuit complexity bears no relation to the physical size of a component. Complexity depends on the number of components as well as on the number of circuits that the system has.

The massive pieces of equipment are rarely the most complicated ones, both electronically and schematically. The tiny units that can be held in the palm of your hand often take the prize for schematic complexity when you break them down to the component level. A tablet computer is an excellent example. The integrated circuits (ICs or chips) inside a device like that contain millions of individual diodes, transistors, capacitors, and resistors. For this reason, the beginner to electronics and schematics should not shy away from any particular circuit, device, or equipment just because its size suggests complexity. You might be wrong, but even if you're right, every schematic diagram will contain portions that you can comprehend.

After you have passed the intermediate stage of learning schematics, then you can tackle complex circuits, devices, and systems. You can break them down into multiple-circuit stages or devices, and ultimately into simple circuits. Try to obtain schematics of a complex nature that also include a thorough explanation of how the circuits work.

Recall the block diagram of Fig. 2-2, the strobe light circuit that we saw in Chap. 2. Compare it to Fig. 5-16, a schematic that shows all of

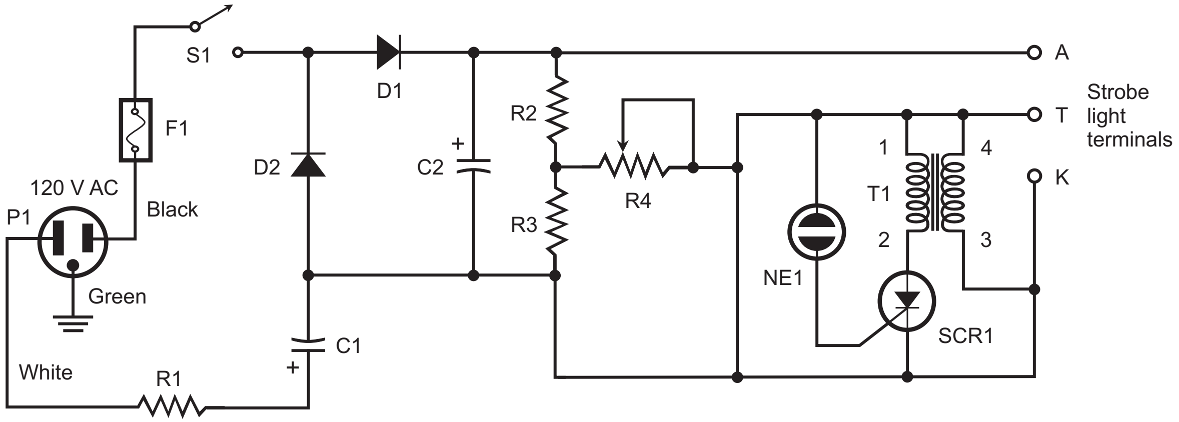
S1
D1
A
Strobe
light
terminals
T
K
F1
R2
R4
D2
C2
+
120 V AC
Black
P1
R3
1
4
T1
2
3
NE1
Green
SCR1
C1
+
White
R1

FIG. 5-16. *Complete strobe light circuit originally shown in the block diagram of Fig. 2-2 (Chap. 2). In order to fit it on the page, the whole diagram is rotated 90 degrees.*

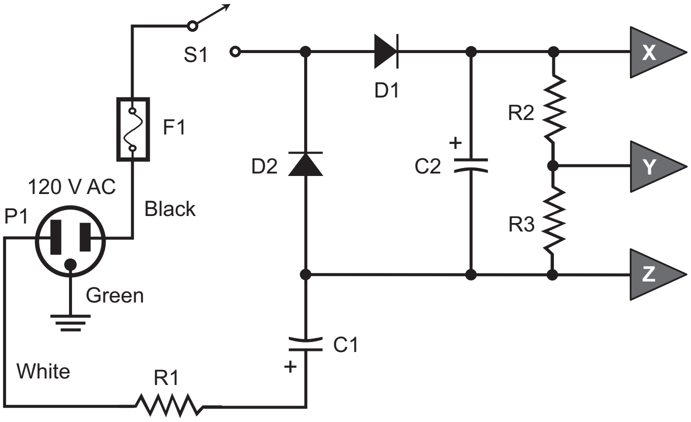

FIG. 5-17A. *Plug, fuse, and rectifier portions of the strobe light circuit. Wedges X, Y, and Z represent extensions to illustration B on the next page.*

the individual components in the circuit. The whole diagram is tilted on its side, allowing it to fit on the page neatly. The circuit is powered with 120 V AC, which comes in at the left side of the schematic (after you've rotated the page). The three terminals of the 120 V AC line take three separate paths along color-coded wires. One wire goes to the fuse, another wire goes to the timing components, and the third wire, coming from the ground prong of the plug, goes to electrical ground.

Following the top signal path, current passes through the fuse F1, assuming that the switch S1 is closed. Current then passes through the rectifier diodes D1 and D2 in one direction only. Conventional current travels with the arrows, and electron current goes against the arrows. From the diodes, one path goes all the way across the top wire to the A terminal of the strobe light, and the other path goes to the adjustable timing components comprising resistors R1 through R4, capacitors C1 and C2, and the neon lamp NE1. The adjustable timing components determine the frequency at which the light operates (how long it stays off between extremely brief flashes). The silicon-controlled rectifier SCR1 performs the switching operation for the strobe lamp. Resistors R2 and R3, besides assisting with the timing operation, provide a junction where the bottom portion of the circuit interacts with the top portion, sending the proper signal to the transformer T1 and finally to the strobe lamp terminals T and K.

Figures 5-17A and B show the same circuit as Fig. 5-16 does, except that the diagram is split between two pages so that it can all

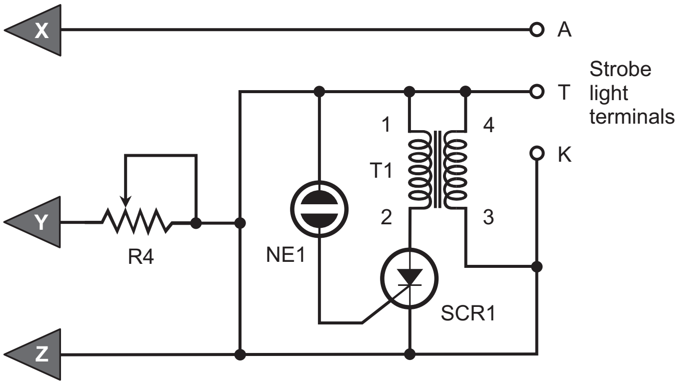

FIG. 5-17B. *Timing and transformer portions of the strobe light circuit. Wedges X, Y, and Z represent extensions from illustration A on the previous page.*

be presented right-side-up instead of sideways. The first part (Fig. 5-17A) shows the power supply and part of the timing circuitry. The second part (Fig. 5-17B) shows the frequency-adjusting potentiometer R4 along with the rest of the timing circuitry, the switching device, and the transformer that provides the strobe light with the voltage it needs to operate properly.

Summary

Reading and drawing schematic diagrams involves breaking down complex circuits into simple ones. The secret is to look at the system's parts and how they relate to each other, rather than to regard the whole thing as a huge monolithic entity. As you systematically study a complex schematic drawing, the relationships between the circuits will get more and more apparent. Once in awhile, you'll suddenly see all of a system's "secrets" revealed at once: An "Aha" moment!

Tip

When you use schematics for electronics troubleshooting, you won't always have to understand the function of every single system element. In many cases, you'll only have to worry about the circuits or components that represent potential trouble spots.

Learning to read and write schematic diagrams is a lot like learning to receive and send the old Morse Code. Morse Code is a language of audible symbols, just like schematic drawing is a language of printed symbols. Once you learn either language, you can use it to communicate; Morse code communicates words and sentences, while schematics communicate principles and concepts.

Using Morse Code as a further example, a long sequence of dots and dashes will mean nothing unless you can break the data down into words. As your proficiency increases, you'll stop hearing mere dots and dashes (or, as some people say, dits and dahs) and hear letters of the alphabet instead. As you keep practicing with the code, you'll start to hear entire words. Eventually, if you keep at it long enough (and especially if you get fond of the code for its own sake, communicating with it for hours on end, as I have done over the years as an amateur radio operator), you'll hear whole phrases and sentences. It will have become a complete language for you.

Reading and writing schematic diagrams follows a similar pattern of development. At first you'll see individual component symbols. Later, you'll begin to see simple circuits hidden within complex circuits. Then you'll be able to identify and decipher complex circuits. Finally, you'll begin to envision entire systems. This knowledge will not come quickly, but your proficiency will improve every time you practice, if you keep pushing yourself (gently, of course) into new knowledge zones.

6

Let's learn by doing

Earlier in this book, I mentioned a volume called *Electricity Experiments You Can Do at Home*. In this chapter, I've adapted a few of the experiments from that book along with layouts and schematics. You should find these activities entertaining and informative in their own right. As you go, you'll "automatically" bolster your proficiency in reading and interpreting schematics. Let's start with some setup details, including a parts list and the construction of a simple circuit-testing board. Then, if you want, you can do the experiments here. If you enjoy them, then you can buy *Electricity Experiments You Can Do at Home* and do a whole lot more!

Tip

You don't have to buy all the parts, build the breadboard, and do the experiments described here, if you aren't in the mood. I think you'll have more fun and learn things better that way, but if you don't want to spend money and time on "nuts and bolts stuff," you can learn a fair amount by simply following along and conducting the experiments in your imagination.

Your breadboard

Every experimenter needs a good workbench. Mine consists of a piece of plywood, weighted down over the keyboard of an old upright piano, and hung from the cellar ceiling by brass-plated chains. Yours doesn't have to be that exotic, and you can put it anywhere as long as it won't shake or collapse. The surface should be made of a nonconducting material such as wood, protected by a plastic mat or a small piece of close-cropped carpet (a doormat works great). A desk lamp, preferably the "high-intensity" type with an adjustable arm, completes the arrangement.

Before you begin any of the tasks described in this chapter, buy a good pair of safety glasses at your local hardware store. Wear the glasses at all times as you play around with the hardware. Get into the habit of wearing safety glasses whenever you work on electrical or electronic circuits, whether you think you need them or not. You never know when a little piece of wire will go flying when you snip it off with diagonal cutters, or a spark will fly right at one of your eyeballs!

Tip

Table 6-1 lists the items you'll need for the experiments in this chapter. You can find many of these components at Radio Shack retail stores, or you can order them through the Radio Shack website. A few of them are available at hardware stores and department stores.

For the experiments described in this chapter, you'll need a prototype-testing circuit board called a *breadboard*. I patronized a local lumber yard to get the wood for my breadboard. I found a length of "12-inch by 3/4-inch" pine in their scrap heap. The actual width of a "12-inch" board is about 10.8 inches or 27.4 centimeters, and the actual thickness is about 0.6 inch or 15 millimeters. They didn't charge me anything for the wood itself, but they demanded a couple of dollars to make a clean cut so I could have a fine rectangular piece of pine measuring 12.5 inches (31.8 centimeters) long.

Using a ruler, divide the breadboard lengthwise at 1-inch (25.4-millimeter) intervals, centered so as to get 11 evenly spaced marks. Do the same going sideways to obtain nine marks at 1-inch (25.4-millimeter) intervals. Using a ball-point or roller-point pen, draw lines parallel to the edges of the board to obtain a grid pattern. Label the grid lines from A to K and 1 to 9, as shown in Fig. 6-1. That'll give you 99 intersection points, each of which you can designate with a letter-number pair, such as D-3 or G-8.

Table 6-1
Components list for simple electricity experiments. You can find these items at retail stores throughout the United States. Abbreviations: AWG = American Wire Gauge, A = amperes, V = volts, W = watts, and PIV = peak inverse volts.

Quantity	Store type or Radio Shack part number	Description
1	Lumber yard	Pine board, approx. 10.8 × 12.5 × 0.6 inches
1	Hardware store	Pair of safety glasses
1	Hardware store	Small hammer
12	Hardware store	Flat-head wood screws, 6 × 3/4
100	Hardware store	Polished steel finishing nails, 1-1/4 inches long
1	Department store	12-inch plastic or wooden ruler
1	Hardware store	Small tube of contact cement
1	Hardware store or Radio Shack	Digital multimeter, GB Instruments GDT-11 or equivalent
1	Hardware store	Diagonal wire cutter/stripper
1	Hardware store	Small needle-nose pliers
1	Hardware store	Roll of AWG No. 24 solid uninsulated (bare) copper wire
1	278-1156	Package of insulated test/jumper leads
4	Hardware store	Alkaline AA cells rated at 1.5 V
1	270-391A	Holder for four size AA cells in series
1	271-1111	Package of five resistors rated at 220 ohms and 1/2 W
1	271-1113	Package of five resistors rated at 330 ohms and 1/2 W
1	271-1115	Package of five resistors rated at 470 ohms and 1/2 W
1	271-1117	Package of five resistors rated at 680 ohms and 1/2 W
1	271-1118	Package of five resistors rated at 1000 (1 k) ohms and 1/2 W
1	271-1120	Package of five resistors rated at 1500 (1.5 k) ohms and 1/2 W
1	271-1122	Package of five resistors rated at 3300 (3.3 k) ohms and 1/2 W
1	276-1104	Package of two rectifier diodes rated at 1 A and 600 PIV
2	272-357	Miniature screw-base lamp holder
1	272-1130	Package of two screw-base miniature lamps rated at 6.3 V
1	272-1133	Package of two screw-base miniature lamps rated at 7.5 V

Once you've marked the grid lines, gather together a bunch of 1.25-inch (31.8-millimeter) polished-steel finishing nails. Place the board on a solid surface that can't suffer any damage from scratching or scraping. A concrete or asphalt driveway will serve this purpose. Pound a nail into each of the grid intersection points shown by the black dots in Fig. 6-1 (53 nails in all). Make certain that the nails are made of polished steel, preferably with "tiny heads." The nails must not have any coating of paint, plastic, or other electrically insulating material. Each nail should go into the board just far enough so that you can't wiggle it around. I pounded every nail down to a depth of approximately 0.3 inch (8 millimeters), a distance amounting to halfway down through the board.

Using 6 × 32 flat-head wood screws, secure the two miniature lamp holders to the board at the locations shown in Fig. 6-1. Using short lengths of thin, solid, bare copper wire, connect the terminals of one lamp holder to breadboard nails A-2 and D-1. Connect the terminals of the other lamp holder to nails D-2 and G-1. Wrap the wire tightly at least twice, but preferably four times, around each nail. Snip off any

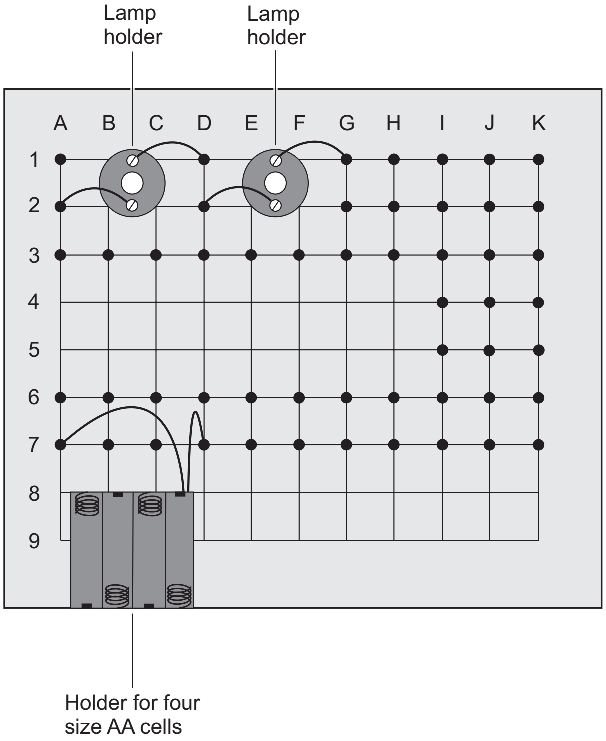

FIG. 6-1. *Layout of the breadboard for simple electricity experiments. I used a "12-inch" pine board (actually 10.8 inches wide) with a thickness of "3/4 inch" (actually about 0.6 inch), cut to a length of 12.5 inches. Solid dots show the positions of the nails. Grid squares measure 1 inch by 1 inch.*

excess wire that remains. Glue the four-cell AA battery holder to the breadboard with contact cement. Allow the cement to harden. That process will need a few hours, so you can take a break for awhile!

When the contact cement has solidified, strip 1 inch of the insulation from the ends of the cell-holder leads and connect the leads to the nails, as shown in Fig. 6-1. Remember that the red lead goes to the positive battery terminal, and the black lead goes to the negative terminal. Use the same wire-wrapping technique that you used for the lamp-holder wires. Place four brand new AA alkaline cells in the holders with the negative terminals against the springs. Now you have a 6-V battery, and the breadboard awaits your exploits.

Wire wrapping

The following breadboard-based experiments employ a construction method called *wire wrapping*. Each nail forms a terminal to which you can attach several component leads or wires. To make a connection, wrap an uninsulated wire or lead around a nail in a tight coil. Make at least two, but preferably four or five, complete wire turns, as shown in Fig. 6-2.

When you wrap the end of a length of wire, cut off the excess wire after wrapping. For small components, such as resistors and diodes,

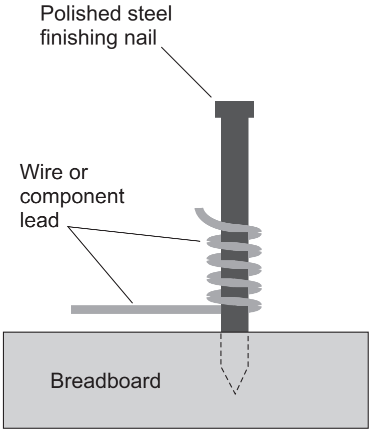

FIG. 6-2. *Wire-wrapping technique. Wind the wire or component lead at least twice, but preferably four or five times, around the nail. Extra wire should be snipped off, if necessary, using a diagonal cutter.*

wrap the leads around the nails as many times as is necessary to use up the entire lead length. That way, you won't have to cut down the component leads. You'll be able to easily unwrap and reuse the components for later experiments. Needle-nose pliers can help you to wrap wires or leads that you can't wrap with your fingers alone.

When you want to make multiple connections to a single nail, you can wrap one wire or lead over the other, but you shouldn't have to do that unless you've run out of nail space. Each nail should protrude approximately 1 inch above the board surface, so you won't be cramped for wrapping space.

Tip

Again, you must make absolutely sure that the nails are made of polished steel *without any coating*. They should be new and clean, so they'll function as efficient electrical terminals.

When you perform the experiments that follow, the exact arrangement of parts on the breadboard is up to you. I've provided schematic and pictorial layout diagrams to show you how the components are interconnected. I recommend that you follow my layout suggestions, only because that way, you can focus on how the actual appearance of the circuit compares with the schematic diagram, even if you haven't bothered to build the breadboard and work with the hardware directly.

Small components such as resistors should always go between adjacent nails, so that you can wrap each lead securely around each nail. *Jumper wires* (also known as *clip leads*) should be secured to the nails so that the "jaws" can't easily work their way loose. It's best to clamp jumpers to nails sideways, so that the wires come off horizontally. If you try to put one of these so-called *alligator clips* down on a nail vertically, there's a good chance that it will pop off in the middle of a mission-critical operation!

Caution!

Please let me repeat: Wear safety glasses at all times as you perform these experiments, whether you think you need to or not.

Kirchhoff's current law

In this experiment, you'll construct a network that demonstrates one of the most important principles in DC electricity. You'll need five resistors: two rated at 330 ohms, one rated at 1000 ohms (1 k), and two rated at 1500 ohms (1.5 k). You'll also need four AA cells.

Mount the resistors on the breadboard by wire-wrapping the leads around the terminal nails, as shown in the layout diagram, Fig. 6-3. Test each resistor with your multimeter (set to work as an ohmmeter) to verify their ohmic values before you install them. Use a 5-inch length of bare copper wire to interconnect the three terminals I-1, J-1, and K-1. Do the same thing with I-3, J-3, and K-3, and also with I-5, J-5, and K-5.

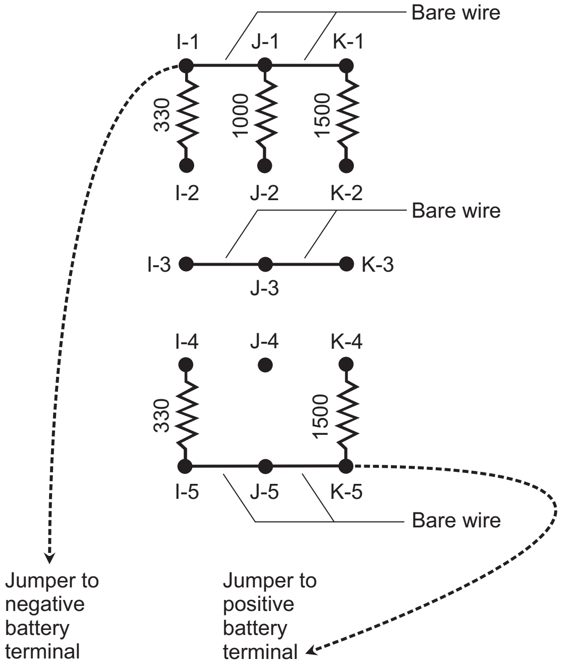
Bare wire
I-1
J-1
K-1
330
1000
1500
I-2
J-2
K-2
Bare wire
I-3
K-3
J-3
I-4
J-4
K-4
330
1500
I-5
J-5
K-5
Bare wire
Jumper to negative battery terminal
Jumper to positive battery terminal

FIG. 6-3. *Arrangement of resistors on breadboard for demonstration of Kirchhoff's current law. All resistance values are in ohms. Solid dots indicate breadboard terminals. Solid lines show interconnections with bare copper wire. Dashed lines indicate jumpers.*

Tip

Engineers symbolize current *as a variable* by writing an uppercase italic letter *I*. Amperes *as units* are abbreviated as an uppercase nonitalic A. Voltage *as a variable* can be symbolized by an uppercase italic *E* or an uppercase italic *V*. Volts *as units* are abbreviated as an uppercase nonitalic V. Resistance *as a variable* is symbolized with the uppercase italic *R*. Ohms *as units* can be written out in full, although some texts use the uppercase nonitalic Greek letter omega (Ω).

Gustav Robert Kirchhoff (1824–1887) did research and formulated theories in a time when no one knew much about electrical current. He used common sense to deduce fundamental properties of DC circuits. Kirchhoff reasoned that the current going into any *branch point* in a circuit must always equal the current going out of that point. Figure 6-4 shows a generic example of this principle, known as *Kirchhoff's first law*. We can also call it *Kirchhoff's current law* or the *principle of conservation of current*.

Mathematically, the sum of the currents entering a branch point always equals the sum of the currents leaving that same branch point. In the example of Fig. 6-4, two branches enter the point and three branches leave it, so

$$I_1 + I_2 = I_3 + I_4 + I_5$$

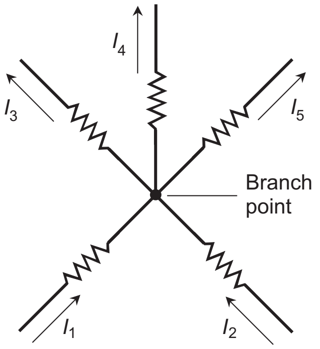

FIG. 6-4. *According to Kirchhoff's current law, the sum of the currents flowing into any branch point is equal to the sum of the currents flowing out of that branch point. In this example,* $I_1 + I_2 = I_3 + I_4 + I_5$.

Kirchhoff's current law holds true no matter how many branches come into or go out of a particular point.

Connect your four-cell battery to the resistive network and measure the currents in each branch. Every test point should be metered individually, while all the other test points are shorted with jumpers. Figure 6-5 is a schematic diagram that shows the actual values of the resistors in my network (yours will be slightly different, of course), along with the value I got when I measured I_1, the current through the smaller of the two input resistors.

As you measure each of the four other current values in turn, make certain that the meter polarity always agrees with the battery polarity. The black meter probe should go to the more negative point, and the red meter probe should go to the more positive point. That way, you'll avoid getting negative current readings that might throw off your calculations. When I tested my four-cell battery to determine its

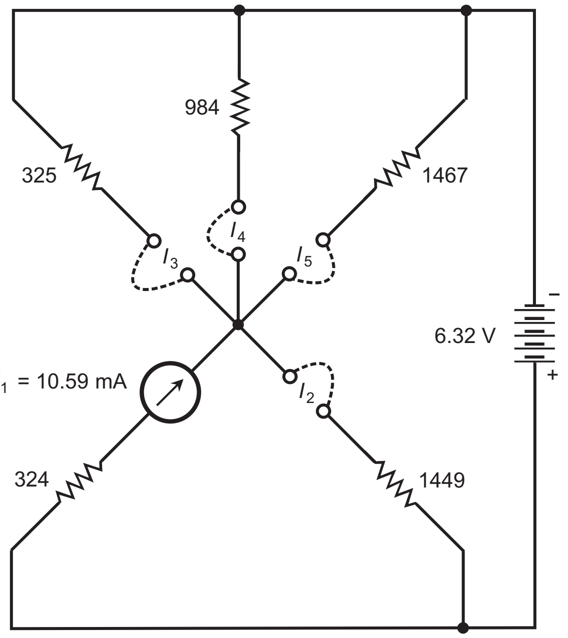

FIG. 6-5. *Network for verifying Kirchhoff's current law. All resistance values are in ohms. The battery voltage, the current* I_1, *and the resistances are the values I measured. Dashed lines show interconnections with jumpers.*

voltage, I got 6.32 V. When I measured I_1 through I_5, I got the following results, accurate to the nearest hundredth of a milliampere (mA):

$$I_1 = 10.59 \text{ mA}$$

$$I_2 = 2.40 \text{ mA}$$

$$I_3 = 8.35 \text{ mA}$$

$$I_4 = 2.79 \text{ mA}$$

$$I_5 = 1.88 \text{ mA}$$

It's "mission critical" that all test points *not* undergoing current measurement be shorted out with jumpers. Otherwise, your network will be incomplete and your current measurements will come out wrong. After you've finished making measurements, remove all the jumpers to conserve battery energy.

Now you can input your numbers to the Kirchhoff formulas and see how close the sum of the input currents comes to the sum of the output currents. Here are my results for the sum of the currents entering the branch point:

$$\begin{aligned} I_1 + I_2 &= 10.59 + 2.40 \\ &= 12.99 \text{ mA} \end{aligned}$$

When I added the currents leaving the branch point, I got

$$\begin{aligned} I_3 + I_4 + I_5 &= 8.35 + 2.79 + 1.88 \\ &= 13.02 \text{ mA} \end{aligned}$$

Tip

When you do experiments of this sort, you should expect a slight discrepancy. That principle explains the 0.03 mA current difference in the branch points in my test. An error of three hundredths of a milliampere at 13 mA amounts to well under 1 percent, which is acceptable.

Now compare!

Compare the layout diagram (Fig. 6-3) with the schematic diagram (Fig. 6-5) without the meter or jumpers.

Kirchhoff's voltage law

In this experiment, you'll construct a network that demonstrates another important DC circuit rule. You'll need four resistors: one rated at 220 ohms, one rated at 330 ohms, one rated at 470 ohms, and one rated at 680 ohms. You'll also need four AA cells.

According to *Kirchhoff's second law*, the sum of the voltages across the individual components in a series DC circuit, taking polarity into account, always equals zero. We can also call this rule *Kirchhoff's voltage law* or the principle of *conservation of voltage*.

Consider the generic series DC circuit shown in Fig. 6-6. According to Kirchhoff's voltage law, the battery voltage E must equal the sum of the potential differences (voltages) across the resistors, although the polarity will be reversed. Mathematically, we can state this fact as

$$E + E_1 + E_2 + E_3 + E_4 = 0$$

If we measure the voltages across the individual resistors and the battery, one at a time, with a DC voltmeter and *disregard the polarity*, we should find that

$$E = E_1 + E_2 + E_3 + E_4$$

Check each of the four resistors with your ohmmeter to verify their actual values. Mount the resistors in the upper right-hand corner of your breadboard by wire-wrapping the leads around nails, as shown in Fig. 6-7. Connect the battery to the network as shown, and measure the voltage across each resistor. Figure 6-8 illustrates the actual values of the resistors in my network (yours will be slightly different),

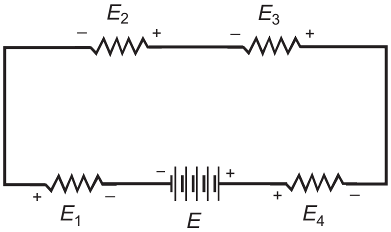

FIG. 6-6. *According to Kirchhoff's voltage law, the sum of the voltages across the resistances in a series DC circuit is equal and opposite to the battery voltage. If we disregard polarity, then in this example, we'll observe that* $E = E_1 + E_2 + E_3 + E_4$.

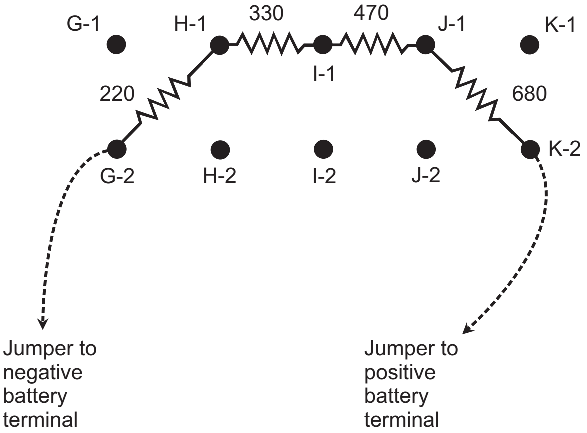
G-1
H-1
330
470
J-1
K-1
I-1
220
680
K-2
G-2
H-2
I-2
J-2
Jumper to negative battery terminal
Jumper to positive battery terminal

FIG. 6-7. *Suggested arrangement of resistors on breadboard for demonstration of Kirchhoff's voltage law. All resistance values are in ohms. Solid dots indicate terminals. Dashed lines indicate jumpers.*

along with the voltage I got for E_2. I measured $E = 6.30$ V across the battery when it was connected to the resistors.

As you measure each voltage E_1 through E_4, the black meter probe should go to the more negative voltage point, and the red probe should go to the more positive point. That way, you'll avoid getting negative

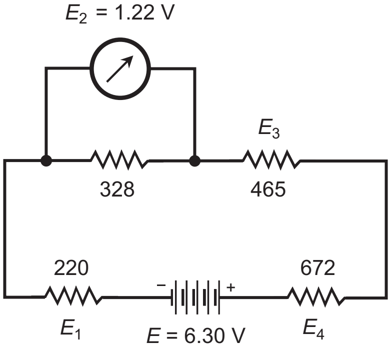
E_2 = 1.22 V
E_3
328
465
220
672
E_1
E = 6.30 V
E_4
−
+

FIG. 6-8. *Network for verifying Kirchhoff's voltage law. All resistance values are in ohms. The battery voltage* E, *the voltage across the second resistor, and the resistances are the values I measured.*

readings that might throw off your calculations. When I measured the voltages across the individual resistors, I got

$$E_1 = 0.82 \text{ V}$$

$$E_2 = 1.22 \text{ V}$$

$$E_3 = 1.75 \text{ V}$$

$$E_4 = 2.52 \text{ V}$$

When you've finished making your measurements, remove one of the jumpers to take the load off the battery.

After you've double-checked and written down your voltage measurements, input the numbers to the modified Kirchhoff formula

$$E = E_1 + E_2 + E_3 + E_4$$

and see how closely it works out. For the left-hand side of this equation, I measured

$$E = 6.30 \text{ V}$$

In the right-hand side of the foregoing equation, I added my numbers to get

$$\begin{aligned} E_1 + E_2 + E_3 + E_4 &= 0.82 + 1.22 + 1.75 + 2.52 \\ &= 6.31 \text{ V} \end{aligned}$$

That's an error of only 0.01 V out of a net potential difference of 6.30 V, amounting to less than two-tenths of one percent error.

Now compare!

Compare the layout diagram (Fig. 6-7) with the schematic diagram (Fig. 6-8), and "follow the flow" through both versions.

A resistive voltage divider

You can use the components from the previous experiment to obtain several different voltages from a single battery. Keep the resistors on the breadboard in the same arrangement as you had them in the experiment for Kirchhoff's voltage law.

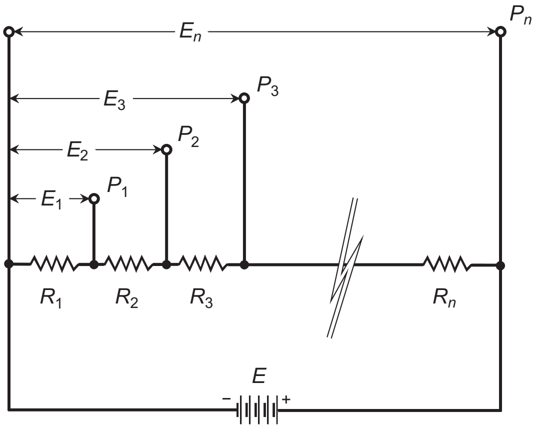

FIG. 6-9. *A voltage divider takes advantage of the potential differences across individual resistors connected in series with a DC power source. Note the use of italics and subscripts for the resistor designators!*

When you connect two or more resistors in series with a DC power source, those resistors produce specific voltage ratios. You can tailor these ratios using specific resistances that "fix" the intermediate voltages. This type of circuit works best when the resistance values are fairly small. Figure 6-9 illustrates the principle of a *resistive voltage divider*. The individual resistances are R_1, R_2, R_3, …, and R_n. The total resistance R is the sum

$$R = R_1 + R_2 + R_3 + \ldots + R_n$$

If we call the power supply voltage E, then Ohm's law tells us that the current I at any point in the circuit must be

$$I = E/R$$

as long as we express I in amperes, E in volts, and R in ohms. At the points P_1, P_2, P_3, …, and P_n, the voltages relative to the negative battery terminal are E_1, E_2, E_3, …, and E_n, respectively. The last (and highest) voltage E_n is the same as the battery voltage E. The voltages at the various points increase according to the sum total of the resistances up to each point, in proportion to the total resistance, multiplied by the supply voltage. In theory, then, we should find that the following equations hold true:

$$
\begin{aligned}
E_1 &= ER_1/R \\
E_2 &= E\,(R_1 + R_2)/R \\
E_3 &= E\,(R_1 + R_2 + R_3)/R \\
&\downarrow \\
E_n &= E\,(R_1 + R_2 + R_3 + \ldots + R_n)/R \\
&= ER/R \\
&= E
\end{aligned}
$$

Tip

Note the use of italics and subscripts for the resistor designators here! Instead of R1, R2, R3, and so on, we now write R_1, R_2, R_3, and so on. This alternative notation hasn't appeared previously in this book, but lots of engineers use it in schematics. Don't get surprised when you encounter it. This notation is commonly used for capacitors, inductors, diodes, and other components as well.

During this experiment, I measured $E = 6.30$ V across the battery as it worked under load, as shown in Fig. 6-10A. This diagram shows the *rated* values of the resistors. Your actual values will differ slightly from the rated values. In my case, the actual values were

$R_1 = 220$ ohms

$R_2 = 328$ ohms

$R_3 = 465$ ohms

$R_4 = 672$ ohms

Set your meter to measure current in milliamperes (mA). Connect the battery to the resistive network through the meter, as shown in Fig. 6-10B, and measure the current. In theory, I expected the milliammeter to indicate a value equal to the battery voltage divided by the sum of the actual resistances, or

$$
\begin{aligned}
I &= E/R \\
&= 6.30/(220 + 328 + 465 + 672) \\
&= 6.30/1685 \\
&= 0.00374 \text{ A} \\
&= 3.74 \text{ mA}
\end{aligned}
$$

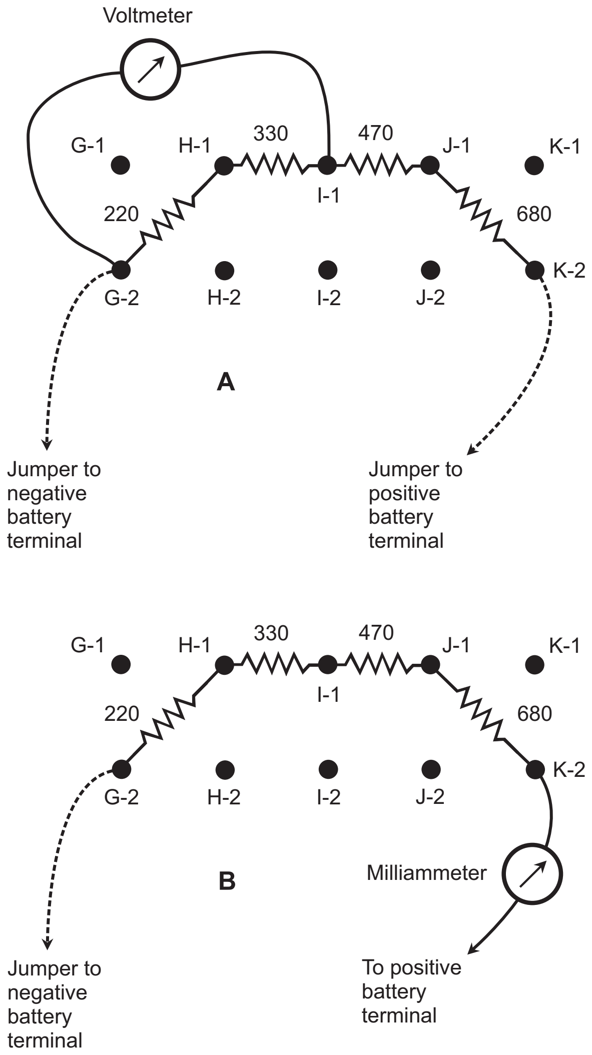
Voltmeter
G-1
H-1
330
470
J-1
K-1
I-1
220
680
G-2
H-2
I-2
J-2
K-2
A
Jumper to negative battery terminal
Jumper to positive battery terminal
G-1
H-1
330
470
J-1
K-1
I-1
220
680
G-2
H-2
I-2
J-2
K-2
B
Milliammeter
Jumper to negative battery terminal
To positive battery terminal

FIG. 6-10. *At A, arrangement for measuring voltages in a resistive divider. Here, the voltmeter is connected to measure the voltage* E_2 *across the first and second resistors. All resistance values are in ohms. Solid dots indicate terminals. Dashed lines indicate jumpers. At B, arrangement for measurement of current through the network.*

When I measured the current, I got 3.73 mA, a value comfortably within the limits of acceptable error.

Now measure the intermediate voltages E_1 through E_4 with your meter set for a moderate DC voltage range. The black meter probe should go directly to the negative battery terminal and stay there. The red meter probe should go to each positive voltage point in turn. First measure the voltage E_1 that appears across R_1 only. Then measure, in order, the following voltages, as illustrated in the schematic of Fig. 6-11:

- The potential difference E_2 across $R_1 + R_2$
- The potential difference E_3 across $R_1 + R_2 + R_3$
- The potential difference E_4 across $R_1 + R_2 + R_3 + R_4$

Figure 6-11 illustrates the arrangement for measuring E_2 to provide a specific example. When I measured each voltage in turn, my meter displayed these results:

$$E_1 = 0.82 \text{ V}$$

$$E_2 = 2.04 \text{ V}$$

$$E_3 = 3.79 \text{ V}$$

$$E_4 = 6.30 \text{ V}$$

After you've finished measuring all the voltages, remove one of the jumpers to conserve battery energy.

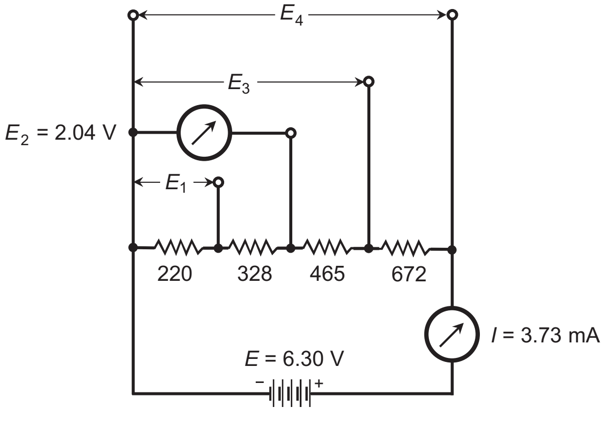

FIG. 6-11. *Network for testing the operation of a resistive voltage divider. All resistance values are in ohms. The battery voltage* E, *the voltage* E_2 *across the first and second resistors, and the resistance values represent my measurements.*

Now connect the meter across the combination $R_1 + R_2$. Run a couple of jumper wires to a *load resistor* located elsewhere on the breadboard, as shown in the layout diagram of Fig. 6-12. This arrangement will make the voltage source E_2 force current through the load resistor, which we'll call R_L. Try every resistor in your repertoire in the place of R_L. If you obtained all the resistors in the parts list, you'll have seven tests to do, using resistors rated at values ranging from 220 to 3300 ohms.

Alternately connect and disconnect one of the jumper wires between the voltage divider and R_L, so that you can observe the effect of the load on E_2. As you can see, the external load affects the behavior of the voltage divider. As R_L decreases, so does E_2. The effect becomes

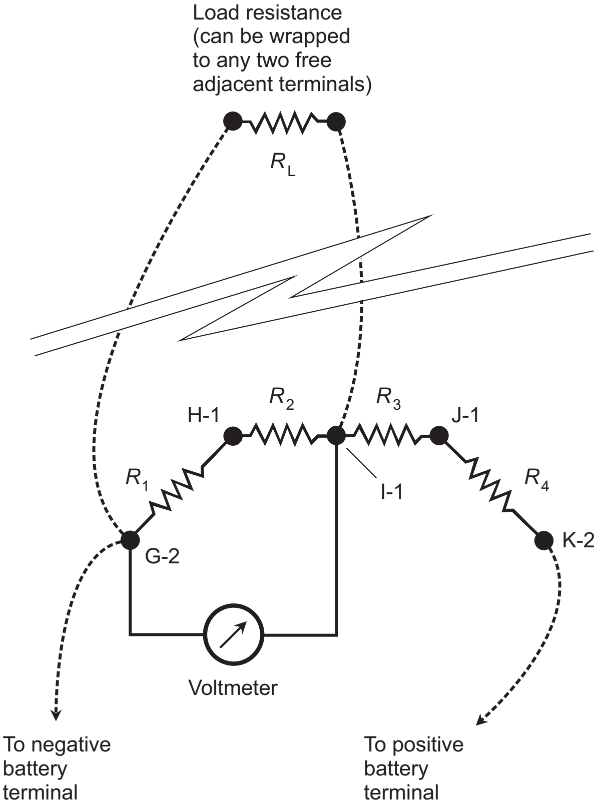
Load resistance
(can be wrapped
to any two free
adjacent terminals)
R_L
R_2
R_3
H-1
J-1
R_1
I-1
R_4
G-2
K-2
Voltmeter
To negative
battery
terminal
To positive
battery
terminal

FIG. 6-12. *Circuit for testing a resistive voltage divider under load. Dashed lines indicate jumpers. This diagram shows the arrangement for measuring variations in* E_2 *as the load resistance* R_L *is alternately connected and disconnected from the series combination of* R_1 *and* R_2.

Table 6-2
Here are the voltages that I measured across various loads in a resistive voltage divider constructed according to Fig. 6-12. My network resistor values were R_1 = 220 ohms, R_2 = 328 ohms, R_3 = 465 ohms, and R_4 = 672 ohms. The load resistance values (left column) are the actual measured values for components rated at 3300, 1500, 1000, 680, 470, 330, and 220 ohms, respectively as you read down.

Load resistance (ohms)	Output voltage (volts)
3250	1.83
1470	1.63
983	1.49
671	1.32
466	1.14
326	0.96
220	0.76

dramatic when R_L becomes small, representing a "heavy load." Table 6-2 shows the results I got. Plot your results as points on a coordinate grid with R_L on the horizontal axis and E_2 on the vertical axis, and then "connect the dots" to get a *characteristic curve* showing the voltage as a function of the load resistance. Figure 6-13 is the graph I made.

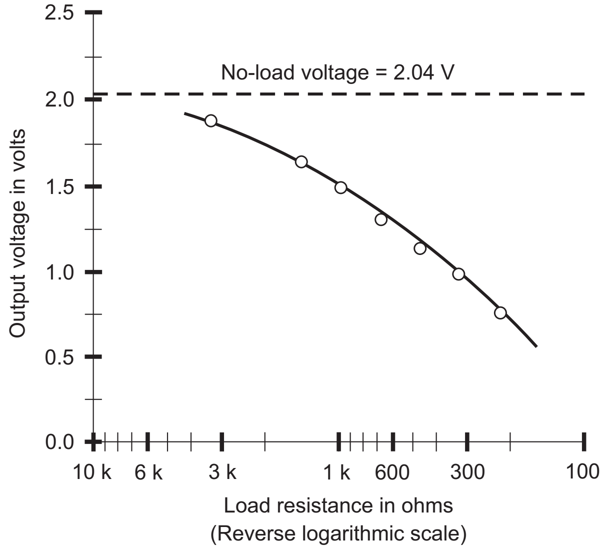

FIG. 6-13. *My results for output voltage versus load resistance in the voltage divider. The dashed line shows the open-circuit (no-load) voltage across the combination of* R_1 *and* R_2 *in series. Open circles show the measured voltages under various loads. The solid curve reveals how the circuit behaves as the load resistance goes down.*

I used a *reverse logarithmic scale* to portray R_L, so that the values would be reasonably spread out. This graph scale allows for a graph that provides a clear picture of what happens as the *conductance* of the load *increases*.

Here's a thought!

What do you think will happen to the voltage across the load if you use two, three, four, or five 220-ohm resistors in parallel, getting R_L values of about 110, 73, 55, and 44 ohms respectively? Try the arrangements one by one, and see what happens! Your original package from Radio Shack has five 220-ohm resistors in it, right?

Tip

The results of this experiment suggest that when engineers build voltage dividers, they had better know what sort of external load the circuit will have to deal with. If the load resistance fluctuates greatly, especially if it sometimes gets low, a resistive voltage divider won't work very well.

A diode-based voltage reducer

Rectifier diodes can reduce the output voltage of a low-voltage DC battery or power supply, providing a better (or at least more predictable) way to obtain specific desired voltages than the resistive divider can do. For this experiment, you'll need two diodes. The ones I obtained were rated at 1 A and 600 *peak inverse volts* (PIV), available at Radio Shack stores as part number 276-1104. You'll also need at least one of each of the resistors listed in Table 6-1, along with some jumpers.

Figure 6-14 shows the schematic symbol for a simple rectifier diode, which is manufactured by joining a piece of *P-type* semiconductor material to a piece of *N-type* material. The N-type semiconductor, represented by the short, straight line, forms the diode's *cathode*. The P-type semiconductor, represented by the arrow, composes the *anode*. Under most conditions, electrons can travel easily from the cathode to the anode (in the direction opposite the arrow), but

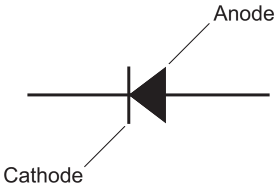
Anode
Cathode

FIG. 6-14. *Schematic symbol for a semiconductor diode. The short line represents the cathode. The arrow represents the anode.*

not from the anode to the cathode (in the direction of the arrow). Conventional current, which always goes from positive to negative, moves in the same direction as the arrow points.

If you connect a battery and a resistor in series with a diode, current will flow if the negative terminal of the battery faces the cathode and the positive terminal faces the anode, as shown in Fig. 6-15A. This condition is called *forward bias.* No current will flow if the bat-

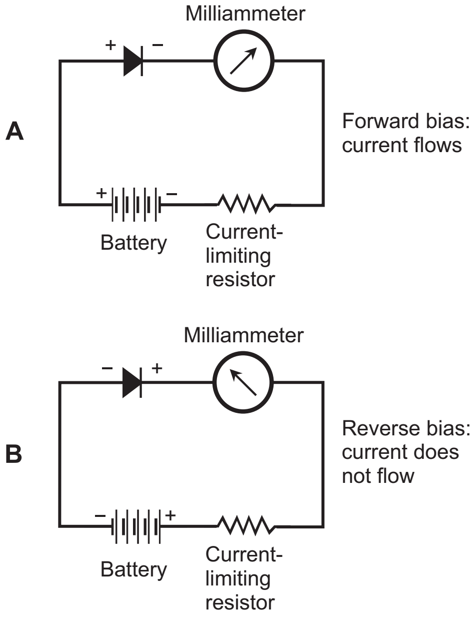
Milliammeter
+
−
A
Forward bias: current flows
+
−
Battery
Current-limiting resistor
Milliammeter
−
+
B
Reverse bias: current does not flow
−
+
Battery
Current-limiting resistor

FIG. 6-15. *Series connection of a battery, a resistor, a current meter, and a diode. At A, forward bias causes current to flow if the voltage equals or exceeds the forward breakover threshold. At B, reverse bias results in no current through the diode, unless the voltage gets very high.*

tery is reversed, as shown in Fig. 6-15B (except when the voltage is very high). This condition is called *reverse bias*. The resistor prevents destruction of the diode by excessive current under conditions of forward bias.

It takes a certain minimum voltage to cause current to flow through a forward-biased semiconductor diode. Engineers call this "threshold" the *forward breaker voltage*. In most diodes, it's a fraction of a volt, but it varies somewhat depending on how much current the diode is forced to carry. If the forward-bias voltage across the diode's *P-N junction* is not at least as great as the forward breaker voltage, then the diode will not conduct. When a diode is forward-biased and connected in series with a battery, the voltage goes down to an extent approximately equal to the forward breakover voltage. Unlike the voltage reduction that takes place with resistors, the diode's voltage-dropping capability does not change much when the external load resistance goes up or down.

Although current won't normally flow through a diode that's reverse-biased, exceptions do occur. If the reverse voltage gets high enough (usually far greater than the forward breakover value), a diode will conduct current because of the so-called *avalanche effect*. Zener diodes, which are used to regulate voltages in power supplies, work according to this principle.

Did you know?

When you connect two or more identical rectifier diodes in series with their polarities in agreement, the forward-breakover voltages add up (give or take a little), so you can get stable and predictable voltage drops that come quite close to whole-number multiples of the forward breakover voltage of a single diode. This technique works as long as you connect a load in the circuit, so that the diodes are forced to carry some current.

You can set up a voltage reducer with two diodes in series and their polarities in agreement, as shown in Fig. 6-16, so that current flows through the load resistor R_L if the forward bias is large enough. Figure 6-17 is a pictorial layout diagram showing an arrangement for mounting the components on your breadboard. Set your meter to indicate DC voltage in a moderate range, such as 0 to 20 V. Connect the meter across the load resistance, paying attention to the polarity so you get positive voltage readings. As you did in the previous

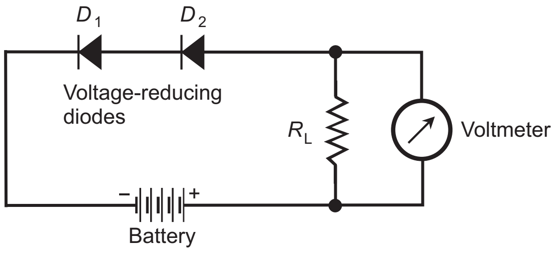
D_1
D_2
Voltage-reducing diodes
R_L
Voltmeter
−
+
Battery

FIG. 6-16. *Schematic diagram showing the method of voltage measurement across a load resistance* R_L *in a two-diode voltage reducer.*

experiment, try every resistor you have for R_L. Measure the voltage across R_L in each case. You'll have seven tests to do, with resistances ranging from 220 to 3300 ohms.

The load resistance R_L affects the behavior of a diode-based voltage reducer, but in a different way than it affects the behavior of a resistive voltage divider. When you do these tests, you'll see that as the load resistance R_L decreases, the potential difference across it goes down, but only a little bit. The voltage across the load tends to drop *more and more slowly* as R_L decreases. Contrast this behavior

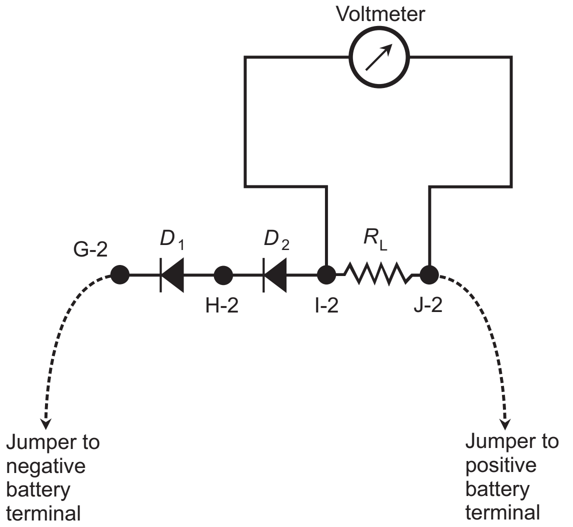
Voltmeter
G-2
D_1
D_2
R_L
H-2
I-2
J-2
Jumper to negative battery terminal
Jumper to positive battery terminal

FIG. 6-17. *Suggested breadboard layout for measurements of the voltages across the load resistance in a two-diode voltage reducer. Solid dots show breadboard terminals. Dashed lines indicate jumpers. Pay attention to the diode polarity! The cathodes should go toward the negative battery terminal.*

with that of the resistive divider, in which the voltage drops off *more and more rapidly* as the load resistance goes down. Table 6-3 shows the results I got when I measured the voltages across various load resistances with this two-diode arrangement.

Plot your results as points on a coordinate grid with the load resistance on the horizontal axis and the voltage across the load resistor on the vertical axis, and then approximate the curve as you did in the previous experiment. When I did this little exercise, I got the graph shown in Fig. 6-18. As before, I used a reverse logarithmic scale to portray R_L. Compare this graph with Fig. 6-13 from the previous experiment.

Here's a thought!

Repeat this experiment with only one diode. Then, if you're willing to make another trip to Radio Shack, get another package of diodes and try the experiment with three or four of them in series. You might also obtain some more resistors, covering a range of values of, say, 100 ohms to 100,000 ohms, and test the circuit using those resistors in the place of R_L.

Caution!

Don't use a resistor of less than about 75 ohms as the load here. In this arrangement, a 1/2-watt resistor of less than 75 ohms will let too much current flow, risking destruction of the resistor (and possibly the diodes as well, if you make the resistance really low).

Table 6-3
Here are the output voltages that I obtained with various loads connected to a diode-based voltage reducer. The circuit consisted of two diodes rated at 1 A and 600 PIV, forward-biased and placed in series with a 6.30-V battery.

Load resistance (ohms)	Output voltage (volts)
3250	5.08
1470	4.99
983	4.96
672	4.91
466	4.88
326	4.84
220	4.79

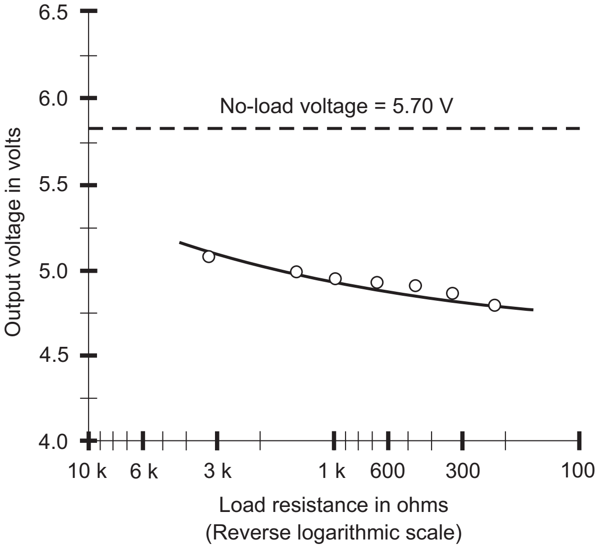

FIG. 6-18. *Output voltage versus load resistance for the diode-based voltage reducer. These are my results. The dashed line shows the open-circuit (no-load) voltage. Open circles show measured voltages under various loads. The solid curve approximates the circuit's characteristic function.*

Mismatched lamps in series

When two dissimilar incandescent lamps operate in series, they receive different voltages and consume different amounts of *volt-ampere* (VA) power, as this experiment demonstrates. (Remember from your basic electricity courses that in a DC circuit, power in watts equals voltage in volts times current in amperes, hence the term *volt-ampere* for simple DC power.) You'll need a 6.3-V lamp, a 7.5-V lamp, four AA cells, and some jumpers.

Your breadboard has two screw-base lamp holders. Position the board so that one holder lies on the left, and the other one lies to its right. Install a 6.3-V lamp in the left-hand socket, and a 7.5-V lamp in the right-hand socket. Connect a short length of bare wire securely between terminals D1 and D2, so that the top terminal of the left-hand lamp holder goes to the bottom terminal of the right-hand lamp holder. Then connect jumpers between the free lamp socket terminals and the battery terminals, so that the lamps are connected in series with each other. Both lamps should glow at partial brilliance.

Call the lamp that's closer to the negative battery terminal "lamp N." That will be the one on the left. Call the lamp that's closer to the positive battery terminal "lamp P." That'll be the one on the right. Unscrew lamp N. At the instant the contact fails, lamp P will go dark. Screw lamp N back in, and then unscrew lamp P. Lamp N will go out. This phenomenon is typical of the behavior of a series circuit. If any component opens up, all the others lose power. A break at any point in a series circuit prevents current from flowing anywhere in that circuit.

Now short out lamp N with a jumper. Lamp N will go dark because it no longer has any voltage across it; lamp P will attain nearly full brilliance. Then disconnect the jumper from lamp N, and move the jumper so that it shorts out lamp P instead. Lamp P will go dark while lamp N glows at full brilliance. Again, this behavior is typical of series circuits. If any component shorts out, all of the others receive more power than they did before.

Set your meter to read DC volts. Connect the jumpers so that the lamps are in series and are both glowing at partial brilliance. Measure the voltage E_1 across lamp N, as shown in the schematic of Fig. 6-19 and in the equivalent layout pictorial of Fig. 6-20. Then measure the voltage E_2 across lamp P, as shown in the schematic of Fig. 6-21 and the equivalent layout pictorial of Fig. 6-22. When I performed these tests, I got the following voltages:

$$E_1 = 2.20 \text{ V}$$

$$E_2 = 3.64 \text{ V}$$

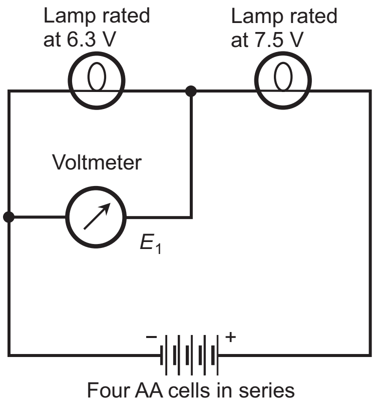

FIG. 6-19. *Measurement of voltage* E$_1$ *across the more negative of two dissimilar lamps in series (called N, rated at 6.3 V).*

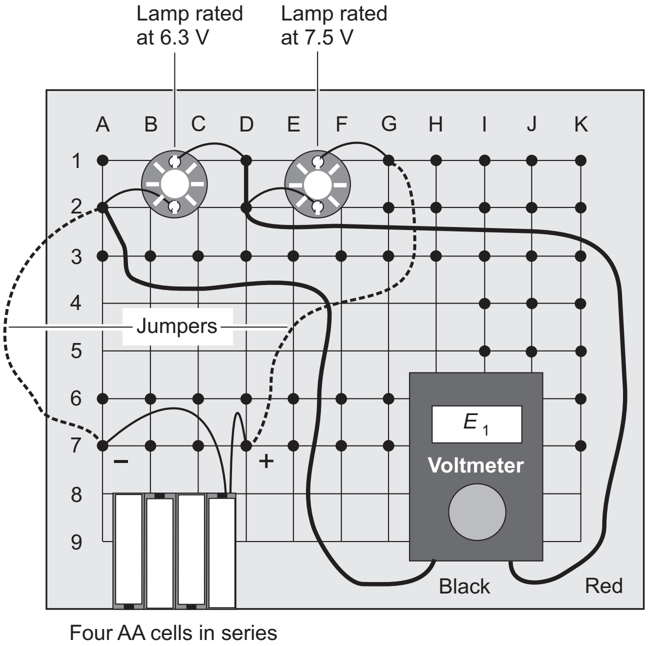
Lamp rated at 6.3 V
Lamp rated at 7.5 V
A
B
C
D
E
F
G
H
I
J
K
1
2
3
4
5
6
7
8
9
Jumpers
–
+
E 1
Voltmeter
Black
Red
Four AA cells in series

FIG. 6-20. *Breadboard layout rendition of the schematic shown in Fig. 6-19.*

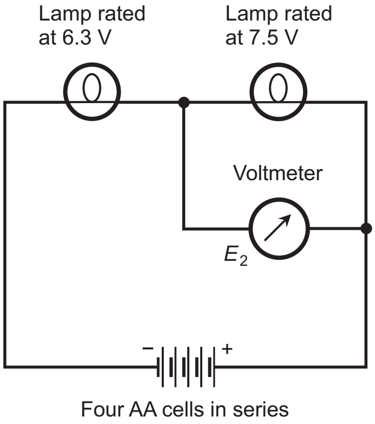
Lamp rated at 6.3 V
Lamp rated at 7.5 V
Voltmeter
E_2
–
+
Four AA cells in series

FIG. 6-21. *Measurement of voltage* E_2 *across the more positive lamp (called P, rated at 7.5 V).*

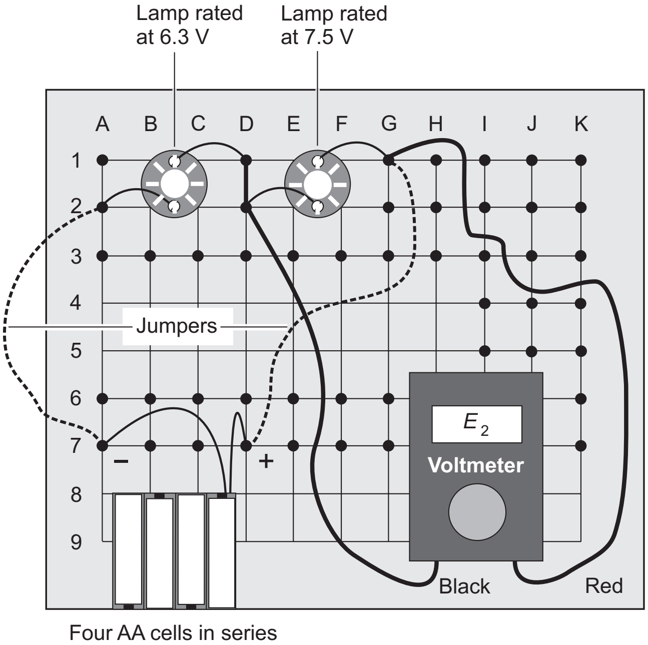
Lamp rated at 6.3 V
Lamp rated at 7.5 V
A
B
C
D
E
F
G
H
I
J
K
1
2
3
4
5
6
7
8
9
Jumpers
–
+
E_2
Voltmeter
Black
Red
Four AA cells in series

FIG. 6-22. *Breadboard layout rendition of the schematic shown in Fig. 6-21.*

Tip

If you can't get the exact same Radio Shack lamps as specified in the parts list for these experiments (Table 6-1), you can use lamps from another source, such as a hardware store or hobby store. This experiment will work as long as the lamps are rated for slightly different voltages or slightly different power levels, and both are rated between 6 V and 12 V.

Now determine the voltage E across the series combination of lamps, as shown in the schematic at Fig. 6-23 and the layout in Fig. 6-24. In theory, the reading should be

$$E = E_1 + E_2$$

When I put my numerical results into this formula, I predicted that I would see

$$\begin{aligned} E &= 2.20 + 3.64 \\ &= 5.84\ \text{V} \end{aligned}$$

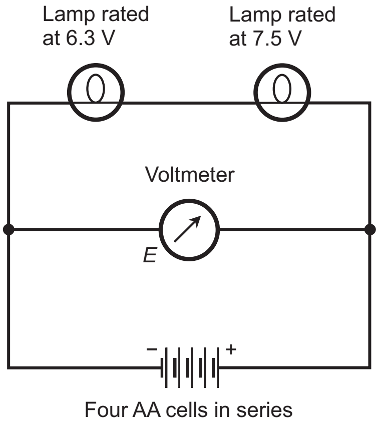
Lamp rated at 6.3 V
Lamp rated at 7.5 V
Voltmeter
E
−
+
Four AA cells in series

FIG. 6-23. *Measurement of voltage* E *across the combination of two dissimilar lamps in series.*

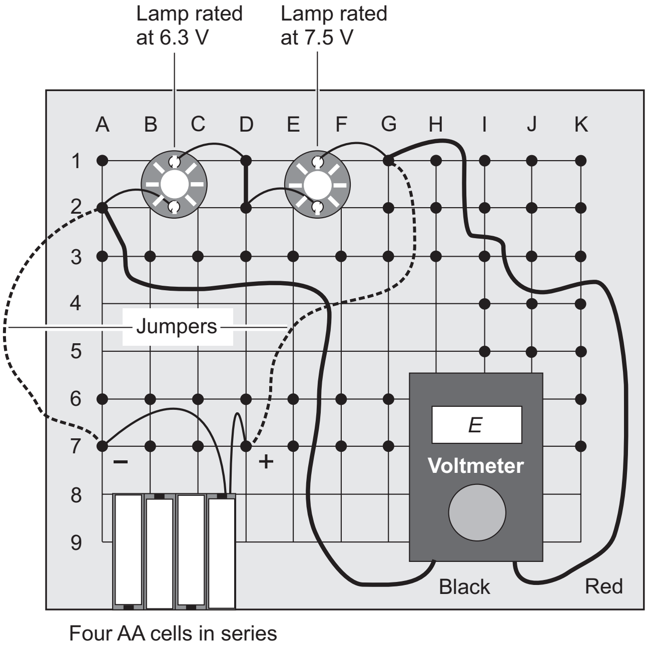
Lamp rated at 6.3 V
Lamp rated at 7.5 V
A
B
C
D
E
F
G
H
I
J
K
1
2
3
4
5
6
7
8
9
Jumpers
E
Voltmeter
−
+
Black
Red
Four AA cells in series

FIG. 6-24. *Breadboard layout rendition of the schematic shown in Fig. 6-23.*

My meter reading was $E = 5.85$ V, a value significantly lower than the 6.30 V I started out with under no-load conditions. Evidently, these bulbs "tax" the four-AA-cell battery rather heavily. I also entertained the notion that because I'd been using the battery for a lot of other experiments besides the ones described here, my AA cells might have grown a little weak.

Set your meter for DC milliamperes, and connect it, as shown in the schematic diagram of Fig. 6-25 and in the equivalent layout pictorial of Fig. 6-26. That connection will allow you to measure the current drawn by the series combination of lamps. I obtained a reading of 139 mA (or 0.139 A).

Now that you know the voltage across each lamp and the current going through the whole circuit, you can determine the VA power figures for the lamps individually and together. Let $P_{\mathrm{VA}1}$ represent the VA power consumed by lamp N, and use the formula

$$P_{\mathrm{VA}1} = E_1 I$$

When I plugged in my results, I got

$$\begin{aligned} \mathrm{P}_{\mathrm{VA}1} &= 2.20 \times 0.139 \\ &= 0.306 \text{ VA} \end{aligned}$$

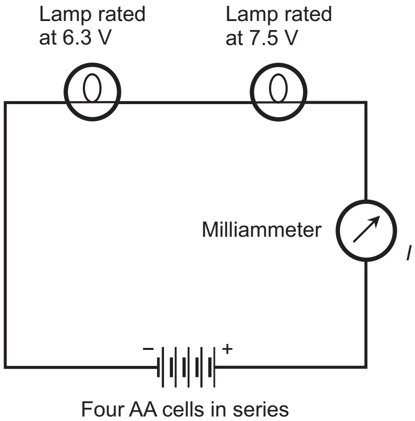

FIG. 6-25. *Measurement of current* I *drawn by the combination of two dissimilar lamps in series.*

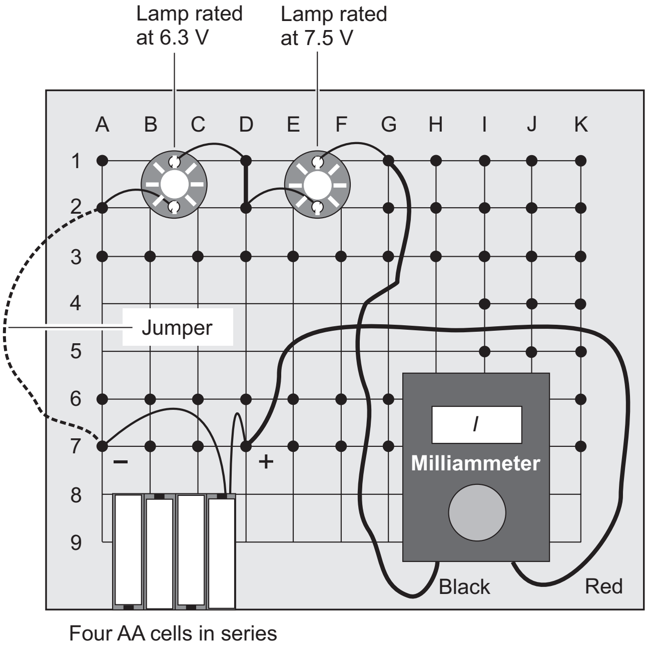
Lamp rated at 6.3 V
Lamp rated at 7.5 V
A B C D E F G H I J K
1
2
3
4
5
6
7
8
9
Jumper
−
+
I
Milliammeter
Black
Red
Four AA cells in series

FIG. 6-26. *Breadboard layout rendition of the schematic shown in Fig. 6-25.*

Letting P_{VA2} represent the VA power consumed by lamp P, the formula is

$$P_{VA2} = E_2 I$$

My result turned out as

$$\begin{aligned} P_{VA2} &= 3.64 \times 0.139 \\ &= 0.506 \text{ VA} \end{aligned}$$

Now let's say that P_{VA} represents the VA power consumed by both lamps operating together. In that case, theory predicts that

$$P_{VA} = E I$$

Plugging in my experimental results gave me

$$\begin{aligned} P_{VA} &= 5.85 \times 0.139 \\ &= 0.813 \text{ VA} \end{aligned}$$

In theory, the VA power consumed by the lamp combination should also work out as the sum of the two VA power quantities taken individually:

$$P_{\mathrm{VA}} = P_{\mathrm{VA1}} + P_{\mathrm{VA2}}$$

Adding my results of $P_{\mathrm{VA1}} = 0.306$ and $P_{\mathrm{VA2}} = 0.506$, I got

$$\begin{aligned} P_{\mathrm{VA}} &= 0.306 + 0.506 \\ &= 0.812 \text{ VA} \end{aligned}$$

The error between my two results was a small fraction of one percent, a state of affairs that made me joyful indeed.

Summary and conclusion

When you want to design, build, debug, and troubleshoot electronic equipment, you'll do best if you have a good schematic (or set of schematics) to work with. Pictorial diagrams, including layouts, can help you find your way too. But when it comes down to doing the work, you'll never find any substitute for playing around with real-world hardware. If you have both schematic and pictorial diagrams available, your job will be as easy as it can get.

If you did all the experiments in this chapter, you probably came up with results that differ slightly from mine. If you had to make major parts substitutions, for example with the lamps in the last experiment, then some of your results doubtless came out a lot different than mine did. Whether you did the experiments or merely followed along as a spectator, you got a chance to see how schematics, layout diagrams, literal pictorials, graphs, and tables can work together! All of these tools belong in an engineer's knowledge base.

Again, I'd like to make a "plug" for *Electricity Experiments You Can Do at Home*. You'll get some good hands-on lab experience, combined with some theory and some rather strange phenomena, from that book. If you want a more theoretical and exhaustive presentation of electricity and electronics, with plenty of schematics and just enough mathematics to keep a true geek from getting bored, I recommend the latest edition of *Teach Yourself Electricity and Electronics*. Both books are published by McGraw-Hill, and you can find them at all major book retailers.

A

Schematic symbols

ammeter	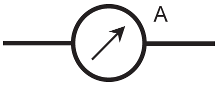
amplifier, general	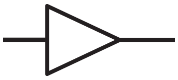
amplifier, inverting	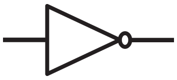
amplifier, operational	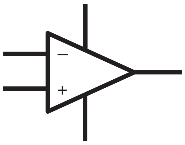
AND gate	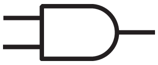
antenna, balanced	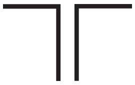
antenna, general	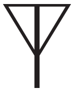

antenna, loop	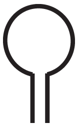
antenna, loop, multiturn	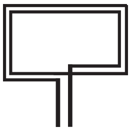
battery, electrochemical	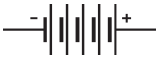
capacitor, feedthrough	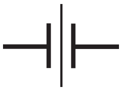
capacitor, fixed	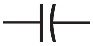
capacitor, variable	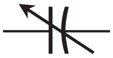
capacitor, variable, split-rotor	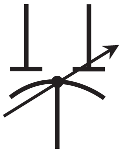
capacitor, variable, split-stator	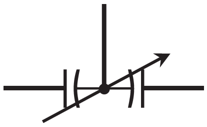
cathode, electron-tube, cold	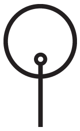
cathode, electron-tube, directly heated	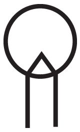

cathode, electron-tube, indirectly heated	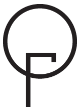
cavity resonator	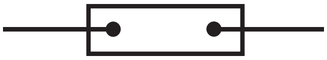
cell, electrochemical	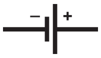
circuit breaker	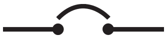
coaxial cable	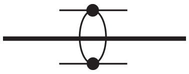
crystal, piezoelectric	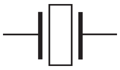
delay line	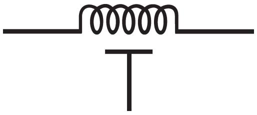
diac	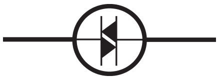
diode, field-effect	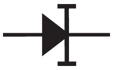
diode, general	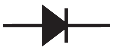
diode, Gunn	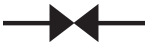
diode, light-emitting	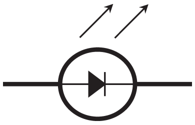
diode, photosensitive	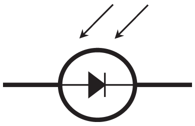

diode, PIN	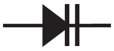
diode, Schottky	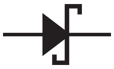
diode, tunnel	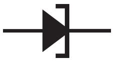
diode, varactor	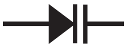
diode, Zener	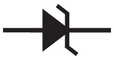
directional coupler	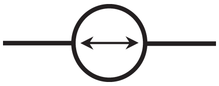
directional wattmeter	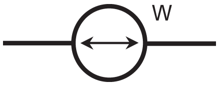
exclusive-OR gate	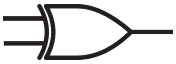
female contact, general	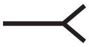
Ferrite bead	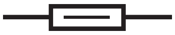
filament, electron-tube	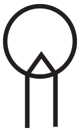
fuse	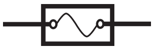
galvanometer	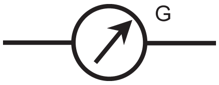
grid, electron-tube	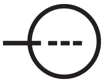

ground, chassis	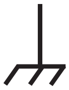
ground, earth	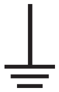
handset	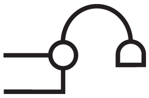
headset, double	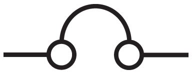
headset, single	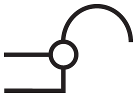
headset, stereo	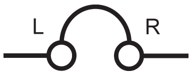
inductor, air core	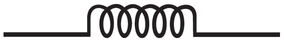
inductor, air core, bifilar	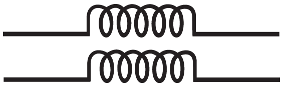
inductor, air core, tapped	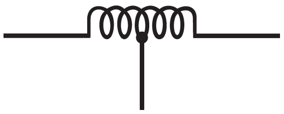
inductor, air core, variable	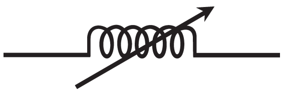
inductor, iron core	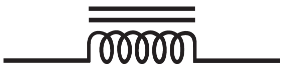
inductor, iron core, bifilar	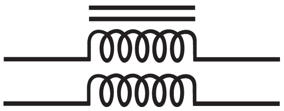

inductor, iron core, tapped

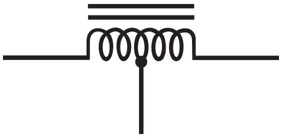

inductor, iron core, variable

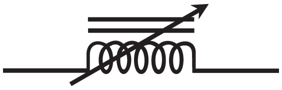

inductor, powdered-iron core

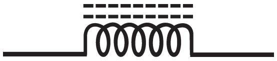

inductor, powdered-iron core, bifilar

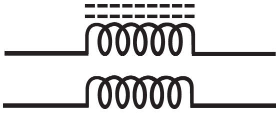

inductor, powdered-iron core, tapped

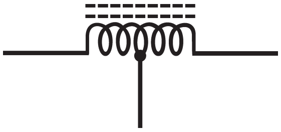

inductor, powdered-iron core, variable

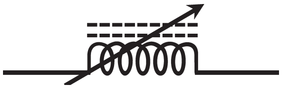

integrated circuit, general

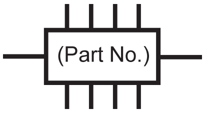

jack, coaxial or phono

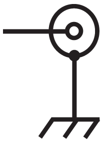

jack, phone, 2-conductor

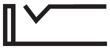

jack, phone, 3-conductor

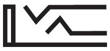

key, telegraph

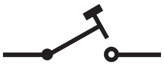

lamp, incandescent	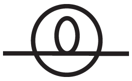
lamp, neon	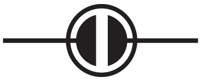
male contact, general	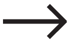
meter, general	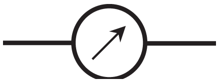
microammeter	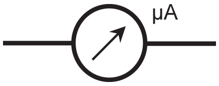
microphone	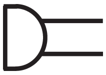
microphone, directional	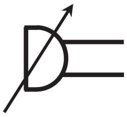
milliammeter	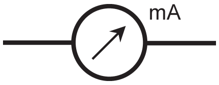
NAND gate	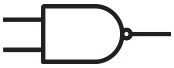
negative voltage connection	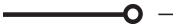
NOR gate	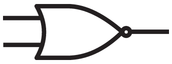
NOT gate	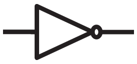
optoisolator	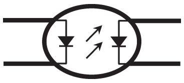

OR gate	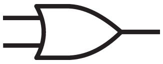
outlet, 2-wire, nonpolarized	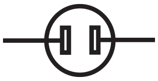
outlet, 2-wire, polarized	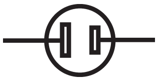
outlet, 3-wire	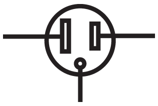
outlet, 234-volt	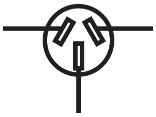
plate, electron-tube	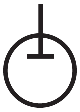
plug, 2-wire, nonpolarized	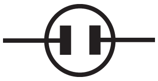
plug, 2-wire, polarized	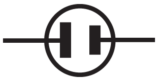
plug, 3-wire	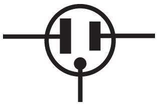
plug, 234-volt	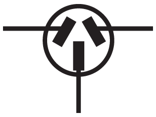
plug, coaxial or phono	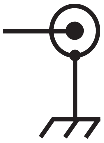

plug, phone, 2-conductor	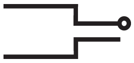
plug, phone, 3-conductor	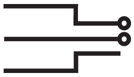
positive voltage connection	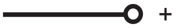
potentiometer	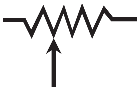
probe, radio-frequency	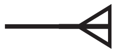 or 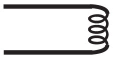
rectifier, gas-filled	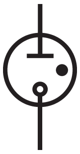
rectifier, high-vacuum	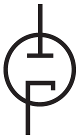
rectifier, semiconductor	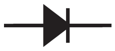
rectifier, silicon-controlled	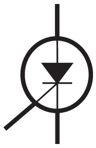

relay, double-pole,
double-throw

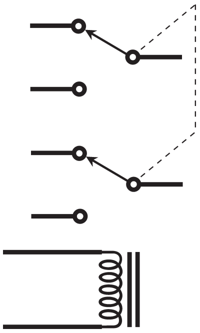

relay, double-pole,
single-throw

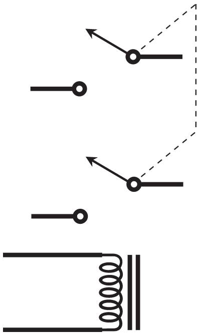

relay, single-pole,
double-throw

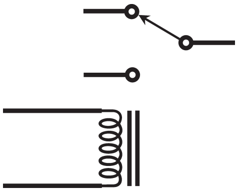

relay, single-pole,
single-throw

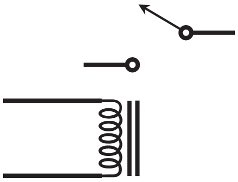

resistor, fixed

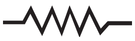

resistor, preset

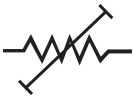

resistor, tapped	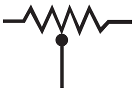
resonator	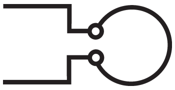
rheostat	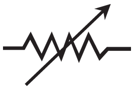
saturable reactor	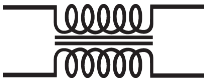
signal generator	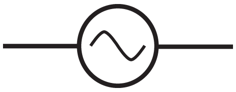
solar battery	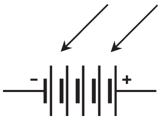
solar cell	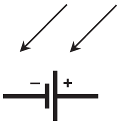
source, constant-current	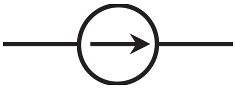
source, constant-voltage	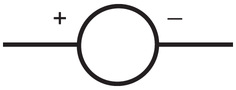
speaker	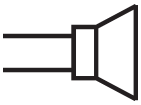

switch, double-pole, double-throw

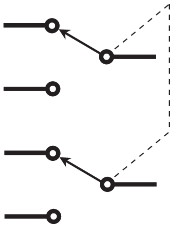

switch, double-pole, rotary

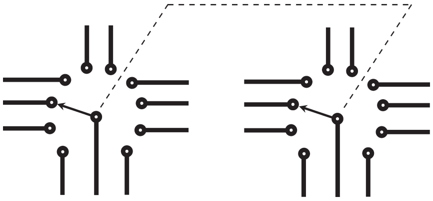

switch, double-pole, single-throw

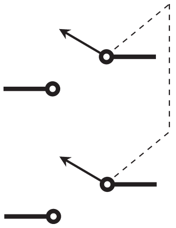

switch, momentary-contact

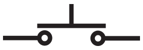

switch, silicon-controlled

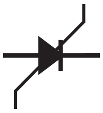

switch, single-pole, double-throw

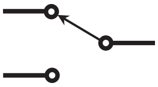

switch, single-pole, rotary

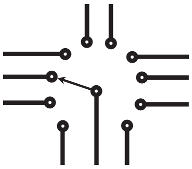

switch, single-pole, single-throw

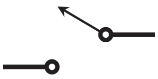

terminals, general, balanced	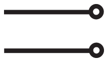
terminals, general, unbalanced	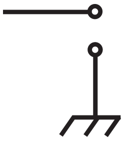
test point	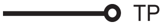
thermocouple	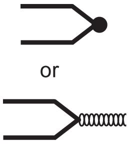
transformer, air core	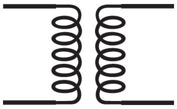
transformer, air core, step-down	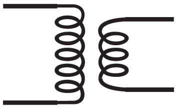
transformer, air core, step-up	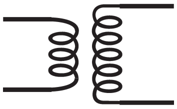
transformer, air core, tapped primary	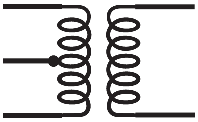
transformer, air core, tapped secondary	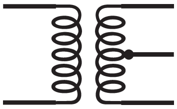
transformer, iron core	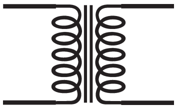

transformer, iron core, step-down

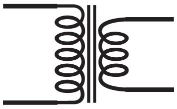

transformer, iron core, step-up

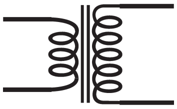

transformer, iron core, tapped primary

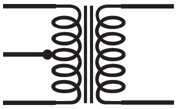

transformer, iron core, tapped secondary

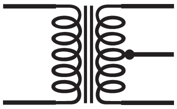

transformer, powdered-iron core

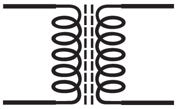

transformer, powdered-iron core, step-down

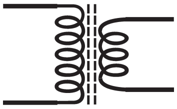

transformer, powdered-iron core, step-up

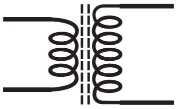

transformer, powdered-iron core, tapped primary

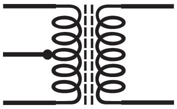

transformer, powdered-iron core, tapped secondary

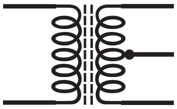

transistor, bipolar, NPN

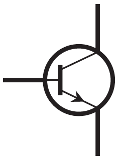

transistor, bipolar, PNP

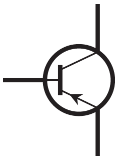

transistor, field-effect,
N-channel

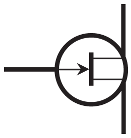

transistor, field-effect,
P-channel

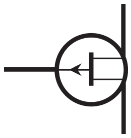

transistor, MOS field-effect,
N-channel

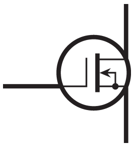

transistor, MOS field-effect,
P-channel

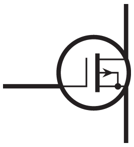

transistor, photosensitive,
NPN

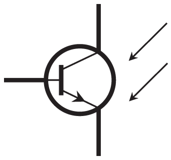

transistor, photosensitive,
PNP

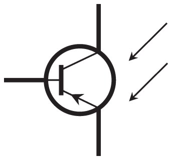

transistor, photosensitive,
field-effect, N-channel

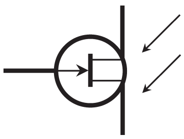

transistor, photosensitive, field-effect, P-channel

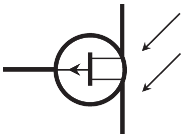

transistor, unijunction

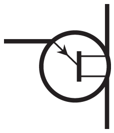

triac

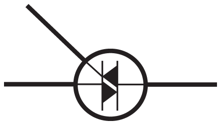

tube, diode

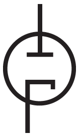

tube, heptode

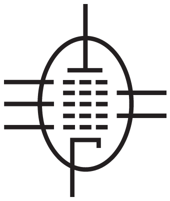

tube, hexode

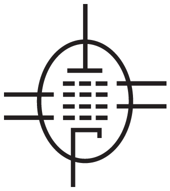

tube, pentode

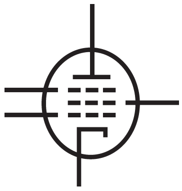

tube, photosensitive

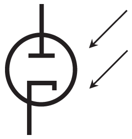

tube, tetrode

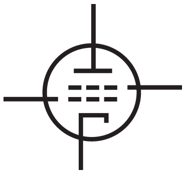

tube, triode

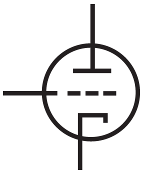

unspecified unit or component

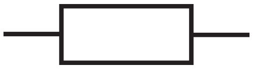

voltmeter

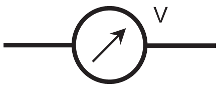

wattmeter

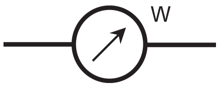

waveguide, circular

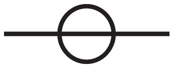

waveguide, flexible

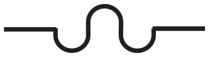

waveguide, rectangular

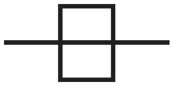

waveguide, twisted

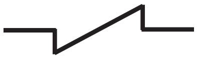

wires, crossing. connected

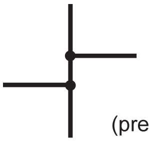

(preferred)

or

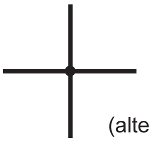

(alternative)

wires, crossing, not connected

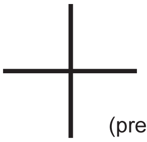

(preferred)

or

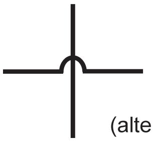

(alternative)

B

Resistor color codes

Some resistors have *color bands* that indicate their values and tolerances. You'll see three, four, or five bands around carbon-composition resistors and film resistors. Other resistors have enough physical bulk to allow for printed numbers that tell you the values and tolerances directly.

On resistors with *axial leads* (wires that come straight out of both ends), the first, second, third, fourth, and fifth bands are arranged as shown in Fig. B-1. On resistors with *radial leads* (wires that come off the ends at right angles to the axis of the component body), the colored regions are arranged as shown in Fig. B-2. The first two regions represent single digits 0 through 9, and the third region represents a multiplier of 10 to some power. (For the moment, don't worry about

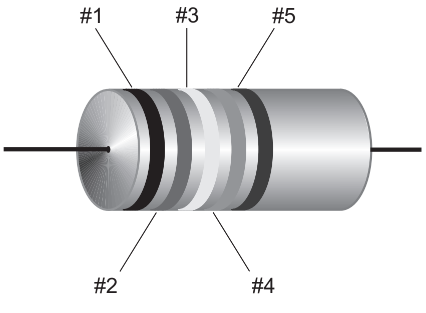

FIG. B-1. *Locations of color-code bands on a resistor with axial leads.*

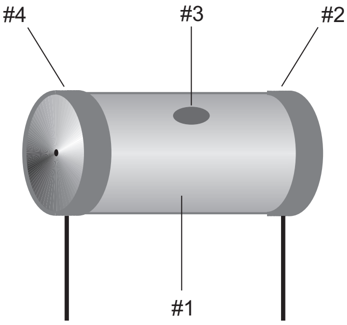

FIG. B-2. *Locations of color code designators on a resistor with radial leads.*

the fourth and fifth regions.) Table B-1 indicates the numerals corresponding to various colors.

Suppose that you find a resistor with three bands: yellow, violet, and red, in that order. You can read as follows, from left to right, referring to the table:

- Yellow = 4
- Violet = 7
- Red = ×100

You conclude that the rated resistance equals 4700 ohms, or 4.7 k.

Table B-1
The color code for the first three bands that appear on fixed resistors. See text for discussion of the fourth and fifth bands

Color of band	Numeral (first and second bands)	Multiplier (third band)
Black	0	1
Brown	1	10
Red	2	100
Orange	3	1000 (1 k)
Yellow	4	10^4 (10 k)
Green	5	10^5 (100 k)
Blue	6	10^6 (1 M)
Violet	7	10^7 (10 M)
Gray	8	10^8 (100 M)
White	9	10^9 (1000 M or 1 G)

As another example, suppose you find a resistor with bands of blue, gray, and orange. You refer to Table B-1 and determine that:

- Blue = 6
- Gray = 8
- Orange = ×1000

This sequence tells you that the resistor is rated at 68,000 ohms, or 68 k.

If a resistor has a fourth colored band on its surface (#4 as shown in Fig. B-1 or B-2), then that mark tells you the tolerance. A silver band indicates ±10%. A gold band indicates ±5%. If no fourth band exists, then the tolerance is ±20%.

The fifth band, if any, indicates the maximum percentage by which you should expect the resistance to change after the first 1000 hours of use. A brown band indicates a maximum change of ±1% of the rated value. A red band indicates ±0.1%. An orange band indicates ±0.01%. A yellow band indicates ±0.001%. If the resistor lacks a fifth band, it tells you that the resistor might deviate by more than ±1% of the rated value after the first 1000 hours of use.

A competent engineer or technician always tests a resistor with an ohmmeter before installing it in a circuit. If the component turns out defective or mislabeled, you can prevent potential future headaches by following this simple precaution. It takes only a few seconds to check a resistor's ohmic value. If you skip that simple step, build a circuit, and then discover that it won't work because of some miscreant resistor, you might have to spend hours tracking it down!

This page intentionally left blank

Suggested additional reading

Frenzel, Louis E., Jr., *Electronics Explained*. Burlington, MA: Newnes/Elsevier, 2010.

Geier, Michael, *How to Diagnose and Fix Everything Electronic*. New York: McGraw-Hill, 2011.

Gerrish, Howard, *Electricity and Electronics*. Tinley Park, IL: Goodheart-Wilcox Co., 2008.

Gibilisco, Stan, *Electricity Demystified*, 2nd ed. New York: McGraw-Hill, 2012.

Gibilisco, Stan, *Electricity Experiments You Can Do at Home*. New York: McGraw-Hill, 2010.

Gibilisco, Stan, *Electronics Demystified*, 2nd ed. New York: McGraw-Hill, 2011.

Gibilisco, Stan, *Teach Yourself Electricity and Electronics*, 5th ed. New York: McGraw-Hill, 2011.

Gussow, Milton, *Schaum's Outline of Basic Electricity*, 2nd ed. New York: McGraw-Hill, 2009.

Horn, Delton, *Basic Electronics Theory with Experiments and Projects*, 4th ed. New York: McGraw-Hill, 1994.

Horn, Delton, *How to Test Almost Everything Electronic*, 3rd ed. New York: McGraw-Hill, 1993.

Kybett, Harry, *All New Electronics Self-Teaching Guide*, 3rd ed. Hoboken, NJ: Wiley Publishing, 2008.

Miller, Rex, and Miller, Mark, *Electronics the Easy Way*, 4th ed. Hauppauge, NY: Barron's Educational Series, 2002.

Mims, Forrest M., *Getting Started in Electronics.* Niles, IL: Master Publishing, 2003.

Morrison, Ralph, *Electricity: A Self-Teaching Guide*, 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc., 2003.

Shamieh, Cathleen, and McComb, Gordon, *Electronics for Dummies*, 2nd ed. Hoboken, NJ: Wiley Publishing, 2009.

Slone, G. Randy, *TAB Electronics Guide to Understanding Electricity and Electronics*, 2nd ed. New York: McGraw-Hill, 2000.

Index

A
AC-to-DC converter, 13–14, 80–81
air-core inductor, 35–36
air-core transformer, 39
alligator clip, 118
AM radio receiver, 89–94
AM voice transmitter, 17–18
amateur radio station, 89
amplifier, 76–79, 85–94, 102
amplifier drive, 89
AND gate, 55
antenna, 63
antenna tuner, 89
arrows in flowcharts, 21
audio power amplifier, 85–89, 93, 102
avalanche effect, 134

B
battery, electrochemical, 54
bipolar transistor, 47–48
block diagram:
    definition of, 1–2
    examples and uses, 13–24
block- to schematic-diagram conversion, 14
breadboard, 113–117
buffer, 79
bulb, incandescent, 59, 61

C
capacitor:
    fixed, 31–32
    padder, 32
    polarized, 32
    trimmer, 32
    variable, 32–34
cat's whisker, 84
cell, electrochemical, 53–54, 59, 61
characters, 9
chassis ground, 35–36
chip, 17, 84
choke, 37
clip lead, 118
coaxial cable, 45–46
code practice oscillator, 96–103
coil, 63
component labeling, 66–71
component tolerance, 77
components, letter designations for, 68
components for experiments, 115

conductors:
  - crossing, connected, 44
  - crossing, not connected, 43, 45
  - single, 59, 61

conservation of current, 120
conservation of voltage, 123
conventional current, 58
"crystal radio" receiver, 84–85
current, conservation of, 120
current, conventional or theoretical, 58
current law, Kirchhoff's, 119–122
cutoff frequencies, 103–104

D

decision block, 22–23
demodulator, 84
detector, 84
diode:
  - semiconductor, 46–47, 63
  - vacuum-tube, 49–50

directly heated cathode, 50
DPDT switch, 41
DPST switch, 41
drive, for amplifier, 88
dual triode vacuum tube, 52

E

earth ground, 45
electrochemical battery, 54
electrochemical cell, 53–54, 59, 61
electron tube (*see* vacuum tube)
exclusive-OR gate, 55
experiments, components for, 115
external capacitance effects, 32

F

farad, 30–31
ferromagnetic material, 37
field-effect transistor, 48–49
field-strength meter, 62–64
filament in vacuum tube, 49–50
first law, Kirchhoff's, 120
fixed capacitor, 31–32
fixed inductor, 35–38
flashlight:
  - single-cell, 58–60
  - two-cell, 60–62

flowchart, 18–24
flux density, magnetic, 37
forward bias, 133–134
forward breakover voltage, 134
full-wave bridge rectification, 69
functional diagram, 13–16

G

galena, 84
ganged capacitors, 34
Gunn diode, 46

H

hand key, 42–43
hard wiring, 7
hardware, 23
henry, 35
heptode vacuum tube, 52–53
hexode vacuum tube, 52–53
highpass filter, 103–104

I

incandescent bulb, 59, 61
indirectly heated cathode, 49–50

inductance-capacitance circuit, 92, 103–104
inductor:
  air-core, 35–36
  fixed, 35–38
  laminated-iron-core, 37
  powdered-iron-core, 38
  solid-iron-core, 37
  tapped, 35–36, 38
  variable, 36, 38
input and output programming symbol, 20
integrated circuit, 17, 84
interconnections, schematic, 6–8
intermediate junction program symbol, 20
International Morse code, 96
iron-core inductor, 37–37
iron-core transformer, 39

J–K
jumper wire, 118

key for sending Morse code, 42–43
Kirchhoff, Gustav Robert, 120
Kirchhoff's current law, 119–122
Kirchhoff's first law, 120
Kirchhoff's second law, 123
Kirchhoff's voltage law, 123–125

L
L network, 94
labeling of components, 66–71
laminated-iron-core inductor, 37
laminated-iron-core transformer, 39
lamps, mismatched, 137–144
language:
  sign, 10
  visual, 8–11
LC circuit, 92, 103–104
letter designations for components, 68
load resistor, 130
logic gates, 54–55
logical inverter, 54–55
lowpass filter, 103–104

M
magnetic flux density, 37
microammeter, 63
microfarad, 31
microhenry, 35
millihenry, 35
mismatched lamps, 137–144
Morse code key, 42–43
multicontact switch, 40–42
multimeter, 74

N
N-type semiconductor, 132
NAND gate, 55
nanohenry, 35
nonferromagnetic material, 37
nonpolarized device, 31
NOR gate, 55
NOT gate, 54–55

O
off-page connection flowchart symbol, 20
ohm, unit of resistance, 25
OR gate, 55
oscillator, 79, 96–103

P
P-N junction, 134

P-type semiconductor, 132
padder capacitor, 32
page breaks, 91–94
parallel resistors, 70–71
peak inverse volts, 132
pentagrid converter, 53
pentode vacuum tube, 51–52
phase coincidence, 97
phase inversion, 97
pi network, 94–95
pi-L network, 95
picofarad, 31
pictorial diagram, definition of, 1
pictorial versus schematic diagrams, 11
polarized capacitor, 32
polarized device, 31
pole, switch, 40
positive-ground system, 97
potentiometer, 26–30
powdered-iron-core inductor, 38
powdered-iron-core transformer, 39
power amplifier, audio, 85–89
power supply, 16–17, 64–65, 80–81, 97–100
preamplifier, 85–87, 91–94
prewritten program symbol, 20
process paths, 22–24
processing operation programming symbol, 20
program decision symbol, 20
program flowchart, 19–24
program modification symbol, 20
punched-card program flowchart, 21–22
push-pull circuit, 87–89

R

radio receiver, 89–94
rectifier diode, 46–47
resistor:
    carbon composition, 26
    film type, 27
    fixed-value, 26–27
    load, 130
    variable, 28–30
    wirewound, 27, 29
resistors in parallel, 70–71
resonant frequency, 85
reverse bias, 133–134
rheostat, 26–29
ripple, 65
roller inductor, 95
rotary switch, 42
rotor of variable capacitor, 33

S

schematic-and-block combinations, 78–81
schematic diagram, definition of, 1–4
schematic interconnections, 6–8
schematic symbology, purpose of, 4–6
schematic- to block-diagram conversion, 14–15
schematic versus pictorial diagrams, 11
second law, Kirchhoff's, 123
sign language, 10
silicon-controlled rectifier, 47
single-cell flashlight, 58–60
software, 24

solid-iron-core inductor, 37
SPDT switch, 40
speech processor, 89–90
SPST switch, 40
start programming symbol, 20
stator of variable capacitor, 33
stop programming symbol, 20
straight key, 42–43, 96
strobe light circuit, 15–16, 108–111
switch:
    DPDT, 40–41
    DPST, 41
    multicontact, 40–42
    rotary, 42
    SPDT, 40
    SPST, 40, 61

T
tapped inductor, 35–36, 38
telegraph key, 96
test points, 75–76
tetrode vacuum tube, 51
theoretical current, 58
throw, switch, 40
tolerance, component, 77
transformers, all types, 39
transistor:
    bipolar, 47–48
    field-effect 48
trimmer capacitor, 32
triode vacuum tube, 50–51
troubleshooting, 71–75, 103
tube (*see* vacuum tube)
tuner, in radio receiver, 92–94
twin-T oscillator, 96–103
two-cell flashlight, 60–62

U–Z
vacuum tube:
    diode, 49–50
    dual triode, 52
    heptode, 52–53
    hexode, 52–53
    pentode, 51–52
    tetrode, 51
    triode, 50–51
varactor diode, 46
variable capacitor, 32–34
variable inductor, 36, 38
visual language, 8–11
voice transmitter, 17–18
volt-ohm-milliammeter, 74
voltage, conservation of, 123
voltage divider, 125–132
voltage doubler, 69–70
voltage law, Kirchhoff's, 123–125
voltage reducer, 132–137
voltage regulation, 69

wire wrapping, 117–118

XOR gate, 55

Zener diode, 46, 134

TECHNOLOGY IN ACTION™

Building Your Own Electronics Lab

A Guide to Setting Up Your Own Gadget Workshop

Dale Wheat

For your convenience Apress has placed some of the front matter material after the index. Please use the Bookmarks and Contents at a Glance links to access them.

friendsof™

Apress®

Contents at a Glance

- About the Author xi
- About the Technical Reviewer xii
- Acknowledgments xiii
- Introduction xiv
- Chapter 1: Planning Your Electronics Workshop 1
- Chapter 2: Building Your Tool Chest 19
- Chapter 3: Components 57
- Chapter 4: A Portable Mini-Lab 87
- Chapter 5: The Cozy Corner Lab 105
- Chapter 6: The Small Group Lab and Classroom 135
- Appendix A: Getting Started with Tool Building 145
- Index 161

Introduction

We're surrounded by electronics these days. Did you ever wonder how it all worked? Do you know just enough about electricity to be dangerous? Would you like to learn more? Then this book is for you.

There are many excellent textbooks on how electricity works. This is not one of them. This book helps you get started learning about electricity and electronics by showing you how to set up your own electronics laboratory. You might have guessed that from the title.

Along the way, several simple exercises are presented to show you some basic concepts about how electricity works in the real world—not just in a lecture hall. By the time you get done reading this short book, you'll have a good idea of where to begin and what kinds of tools and components you might need, and you'll have some excellent advice about how to keep it all organized.

Once you've got your own space set up to perform your electronic experiments, the sky is the limit.

Here's what you have to look forward to in this book, one chapter at a time.

Chapter 1: Planning Your Electronics Workshop

Your very own electronics laboratory is a great place to build, repair, invent, and learn more about electronics. We seem to be surrounded by electronics more and more every day. We can easily become dependent upon them, without even understanding how they work. Having your own electronics lab can help you gain some control over your electronic minions. Here you can learn about their inner mysteries, including how to repair them when possible or, better yet, improve them with your own custom modifications. You can also turn your own ideas into reality by building electronic circuits from scratch.

What will you need to plan your lab? What if you already have a basic understanding of electronics and the beginnings of a workshop at your disposal? This chapter helps get you going in the right direction.

Chapter 2: Building Your Tool Chest

You're going to need some tools in your lab, as well as the skills to use them effectively. This chapter will get you started.

If you're starting from scratch, don't worry. There's not a whole lot you absolutely *must* have to get started. Some basic hand tools and a place to keep them organized is all you will need. You don't even have to spend a lot of money at first. As you progress in your hobby, you will most likely want to add to your tool chest and upgrade some of those tools. Electronics can be the perfect hobby because it can take up (1) all your spare time and (2) all your discretionary income. What more could you ask for?

Chapter 3: Components

You're going to need some components to play with, as well as the knowledge to use them effectively. This chapter will introduce you to some of the bits and pieces that make up modern electric and electronic circuits. You'll also learn a little bit about how to identify components from their appearance and markings, when available.

Once you've got an idea about what these parts do in a circuit, you'll learn a little more about how to measure their electrical properties and put them to use. You'll also be shown what *not* to do, in some select examples.

Chapter 4: A Portable Mini-Lab

Having a portable (or at least (*trans*portable) electronics lab comes in handy in several circumstances. Maybe you don't have a place (yet) for a permanent home for all that equipment. Maybe you need to bounce at a moment's notice, heading off on electronic adventures at the drop of a hat. Or maybe you like to keep everything where you can find it in a hurry, without having to rummage through shelves and boxes, looking for just the right tool or component.

Whatever your motivation for wanting a portable lab at your disposal, this chapter should help you get started.

Chapter 5: The Cozy Corner Lab

You don't need an entire garage or extra office for a functional laboratory for your electronic endeavors. It's certainly nice if these areas are available to you, but you might be surprised at how much you can do in just a small space, if you set it up properly and maintain it with determination.

Chapter 6: The Small Group Lab and Classroom

The first part of this book deals with how to get along with electrons and make them do your bidding. In this chapter, you'll explore a completely different topic: other people, and how to get along with them.

You might not ever figure out how to get *them* to do your bidding, but at least you can keep them from stealing your tools. Maybe.

Appendix A: Getting Started with Tool Building

Once you've become comfortable in your lab and had time to play with a few circuits, you might start to notice that some of the "tools" you've been using are just simple electronic circuits themselves. Good examples are power supplies and meters. These are great tool-building projects because you can see useful results early.

This appendix takes a look at building a couple of simple electronic tools that might be of use in your lab. Hopefully you will develop a better understanding of how these tools work. Ideally, you will progress from the basic question of "Does it work?" to the more involved questions of "How well does it work and what can I do to improve it?"

CHAPTER 1

Planning Your Electronics Workshop

Are you interested in electronics? Would you like to set up your own “electro lab” to conduct amazing experiments, build crazy gizmos, and repair or modify your existing electronics? Then this book is for you!

Your very own electronics laboratory is a great place to build, repair, invent, and learn more about electronics. We seem to be surrounded by electronics more and more every day. We can easily become dependent upon them, without even understanding how they work. Having your own electronics lab can help you gain some control over your electronic minions. Here you can learn about their inner mysteries, including how to repair them when possible or, better yet, improve them with your own custom modifications. You can also turn your own ideas into reality by building electronic circuits from scratch.

What will you need to plan your lab? What if you already have a basic understanding of electronics and the beginnings of a workshop at your disposal? This chapter helps get you going in the right direction.

What to Expect

This book can’t teach you everything about electronics, and it doesn’t try. That would take several lifetimes. What it can do is introduce you to the tools and the skills you will need to set up your own electronics lab. This will include a very basic introduction to electricity, a little bit of theory, some safety tips, and a whole lot of example projects. It’s a great starting place. Where it will lead you is mostly determined by you and what you’re wanting to do with electronics.

If you’re already familiar with electricity and electronic concepts, there is still plenty of fun and interesting stuff waiting for you to explore. The field of electronics itself is constantly expanding. From the time this book was written to the time that you’ve finished reading this sentence, many advances will have occurred in both our understanding of electronics as well as the development of new applications and electronic devices. This is one of the many things that makes the study of electronics so interesting and exciting.

Every lab is different. Your lab will reflect not only your immediate electronics goals but also your personality. Feel free to pick and choose from all the ideas presented in this book and add in some of your own. The repair shop and the design studio are necessarily going to be arranged differently. There really is no right or wrong way to go.

Even the best laid plans, so the saying goes, end up somewhere unexpected. When planning your electronics lab for the first time, or for the tenth time, keep in mind that your interests, resources, and reasons for wanting to work on electronics are going to change over time. Please feel free to reinvent yourself and your lab as conditions permit. You might also be forced to reconsider your priorities when other factors in your life exert themselves. That being said, stay tuned for "A Cautionary Tale" at the end of this chapter.

Some configurations are more appropriate for certain endeavors, and nothing replaces the right tool for the job. This book should at least give you an idea of what is involved and get you pointed down the right road.

It Starts with a Plan

Most of the decisions you'll make when planning your laboratory will depend on what *you're* interested in doing there. Having a good plan at the beginning is like having a clear map when taking a long journey. It helps you from getting lost and also is handy for figuring out when you've arrived.

■ **Note** If you fail to plan, you plan to fail.

Remember, at this point in the game, you don't have to make any final decisions about what you want to be able to do in your lab. This book is all about illustrating possibilities and giving you the information and advice necessary to branch out into the areas that hold the most interest for *you*.

A Broad Outline

Here's what you need to get started: a little time, a small amount of space, some basic tools, and a few components. Those are the boring parts. The fun stuff includes your ideas, goals, and inspiration. Combine all that with some fundamental information about how electricity and electronics work, and you're in business.

A Little Time

Like any good hobby, setting up your electronics lab and conducting experiments there is going to take up some of your time. For some, this is the *main reason* for getting started in electronics and building a lab: a place to escape, unwind, and tinker with ideas, prototypes, and complex systems. Maybe it's your job. Maybe it's your passion. Perhaps it's a bit of both.

You're going to need time to plan and build your lab. This book will spell out several specific arrangements for various-sized labs, but they are just examples and not hard-and-fast rules. More time will be required for obtaining the proper tools and materials that you will use there. Time will be spent not only *doing* things in your lab, but also in just *thinking* about doing things in your lab. For example, you might build a clever little gizmo that ought to do a simple task, only to find out that it refuses to work. You might need to spend a little bit of time thinking of ways of testing it to find out where the problem lies. This is sometimes called *troubleshooting* or *debugging*.

This is the basis of the scientific method. To help understand how something works, you first form a theory that would explain its behavior. Next you prepare one or more experiments that will demonstrate the accuracy of your theory. The experiments are conducted and their results are analyzed. If the results are in agreement with your theory's predictions, then your theory is more than likely correct. If not, you may need to modify or extend your theory to accommodate the behavior observed. All of these things take time, which helps explain why we don't have jet packs or time machines yet.

A Little Space

You don't have to start out with a cavernous underground bunker for your electronics lab. If you already have one, that's great. You might have an extra classroom available, some unused warehouse space, or an empty bench in a garage.

Then again, you might only have a corner of the dining-room table, and then only between mealtimes. That's plenty of room when you're just getting started! Once you learn more about electronics, and especially when you learn more about what it is that *you* want to do with electronics, then you can start looking for a more permanent home for your toys, tools, and spare parts.

The main points that will be stressed throughout this book when planning and working in your space are safety, good lighting, and organization, in that order.

- Your lab needs to be relatively safe for you, and *very* safe for your visitors and neighbors. The "Go Away!" sign on the door does not guarantee that you will be free from curious visitors, pesky and otherwise. People are naturally attracted to the creative and mysterious activities going on in your lab.
- Be prepared to buy, build, or borrow more lighting, because you're going to need it. Proper lighting improves the quality of your doodlings and tinkering enormously. You might think you already have enough light, and unless you were a photographer in a previous life, you don't. (Photographers understand that you can never have too much light.)
- An organized lab is a productive lab. Don't waste time looking for parts or tools when it would have only taken mere moments to put them in their proper places in the first place. A place for everything, and everything in its place. You get to decide where all these tools and bits get to live, so spend some time and do it wisely.

These three tenets will be repeated throughout the book, so you might as well get used to hearing about them.

Basic Tools

You really don't need a lot of tools to get started. There is a tendency to want to be a bit overprepared before tackling any new project, but it really isn't necessary when it comes to planning and building a new electronics lab. This is also true when sorting out a lab that's gone badly out of control.

You're going to be working with wire quite often. Claude Debussy has been quoted as saying that music is the space between the notes. It's the connections between the various electronic components that make the magic happen in an electronic circuit, in the same way that it's the relationships between people that make their lives so interesting.

The connections between all the various imaginable electronic components almost always start out as bits of wire. These connections can then be optimized into copper traces etched on a printed circuit board or even as metallization links in an integrated circuit. You're going to be working with a lot of wires in this hobby, and that usually boils down to two basic hand tools: *wire cutters* and *wire strippers*.

A wire cutter does what it says and says what it does: it cuts wire. A wire stripper is a more specialized tool that helps remove the outer layer of insulation from a wire. Ever more specialized versions of both tools abound. Some tools aim to perform both tasks at once, while never perfecting the art of either.

Resist the temptation to use your teeth for cutting and stripping wire. If you're caught in a situation without proper tools, use your fingernails. They grow back faster than teeth.

■ **Tip** Use the right tool for the job. Knowing which tool is the right tool is half the battle.

If you're performing detailed work on smaller circuits, you might benefit from having some small needle-nosed pliers, a small vise, clip, or helping-hands jig, and possibly a magnifying glass.

A popular and handy skill in the electronic arena is *soldering*. Soldering is the process of making electrical and mechanical connections. This is done by heating the items to be joined together and adding solder. The solder melts and fills in all the gaps between the items being joined. Once cooled, the solder forms a solid, conductive link.

■ **Note** Soldering is similar to welding, with the exception that welding actually melts the parts being joined.

A suitably sized soldering iron and the right kind of solder are indispensible for many projects, including repair, prototyping, and production. Picking up the skill to make quality connections is obtained in the same way as all worthwhile skills: with lots of practice.

Besides these basic fabrication tools, you will also want some *test and measurement equipment*. These are devices used to measure various aspects of electrical circuit behavior. These range from the very simple, such as voltage meters, current meters, and continuity testers, to the very exotic (and sometimes expensive), which include oscilloscopes and various logic, network, and signal analyzers. See the Appendix for information about building your own simple test equipment for your lab.

This is just a quick summary of the kinds of tools you might want in your electronics lab. Chapter 2 goes into much more detail about the tools you will probably need or want in your lab, as well as the skills needed to operate them safely and effectively.

Fundamental Components

As mentioned in the previous section, you're more than likely going to be wiring up various contraptions and rewiring others. This is probably going to involve a lot of wire.

For small voltages and currents, small-gauge insulated copper wire is often used. It's easy to handle, cut, and splice. This can be done with mechanical connectors such as wire-nuts or screw terminals, as well as with soldering.

Back in the day when computers were all connected using fat cables, lots of tiny wires were bundled up into cables and snaked all over a typical desktop computer installation. With the increasing popularity of wireless devices, those fat, juicy cables are lying around, discarded, unwanted, and just waiting to be harvested for their valuable conductors. You can also buy various kinds of wire in all colors and sizes from most hardware stores and from online suppliers.

Now consider the types of projects you'll be working on:

- If you're most interested in building your own lighting, you're going to want some lightbulbs, sockets, and LEDs, or whatever lighting technology appeals most to you. Don't forget that you're also going to need switches, knobs, dimmers, housings, plugs, and *more wire*.

- If you like to work with things that move, you're going to need motors, relays, solenoids, servos, and fans. You'll probably be needing some gears, pulleys, chains, belts, shafts, and couplings, along with connectors, switches, controllers, position sensors, and *more wire*.
- If you're into making lots and lots of sound, you'll be needing speakers, headphones, microphones, pickups, transducers, transformers, amplifiers, connectors, jacks, plugs, and, you guessed it, *more wire*.
- Are computers and digital devices more to your liking? Be thinking about getting yourself a wide variety of chips, microcontrollers, software, cables, sockets, connectors, and still *more wire*.

Now you've got an idea of some of the different components you might want to have on hand, depending on what kind of inspiration happens to strike. You're going to need a place to keep all these different parts sorted. This will play a big part in the overall planning phase of your lab. Maybe your old fishing tackle box has enough room for your basic hand tools and enough secret compartments to hold an interesting variety of components. On the other end of the spectrum, maybe you need to start thinking about some industrial-grade shelving units or pallet racks. Now is the time to think about all these factors and start making the right kind of plans to help make your lab a successful one.

Again, this is just a quick summary of some of the types of components you might want to work on in your lab. Chapter 3 is going to take you much farther into understanding many of the more common electrical and electronic components that you are likely to encounter in your lab.

Ideas

Admit it: you've already got some ideas for cool projects that you'd like to be working on in your own electronics lab. That's great! You're well on your way to making those dreams a reality.

The best possible thing you could do with your ideas is to write them down somewhere. It can be as simple as a sketch on a piece of scrap paper or as formal as a complete set of project documentation. The important thing is that you actually get the idea across using either words or pictures in a form that you will be able to revisit in the future. This will be of immense benefit to you, as well as anyone else that might be interested in your ideas.

It's also a great idea to keep a log as you work on your projects. That's the perfect place to jot down those ideas as they come popping out of your brain. Be as detailed as you can be. You will thank yourself later.

Even if you don't already have a backlog of ideas jostling for your attention in the lab, it's still a good idea to keep a log and explore the things that interest you. Leonardo da Vinci kept detailed notebooks on almost every possible subject. Anything that interested him got written down, sketched, or diagrammed in his notes. Even today, people are still getting inspiration from his doodles (see Figure 1-1).

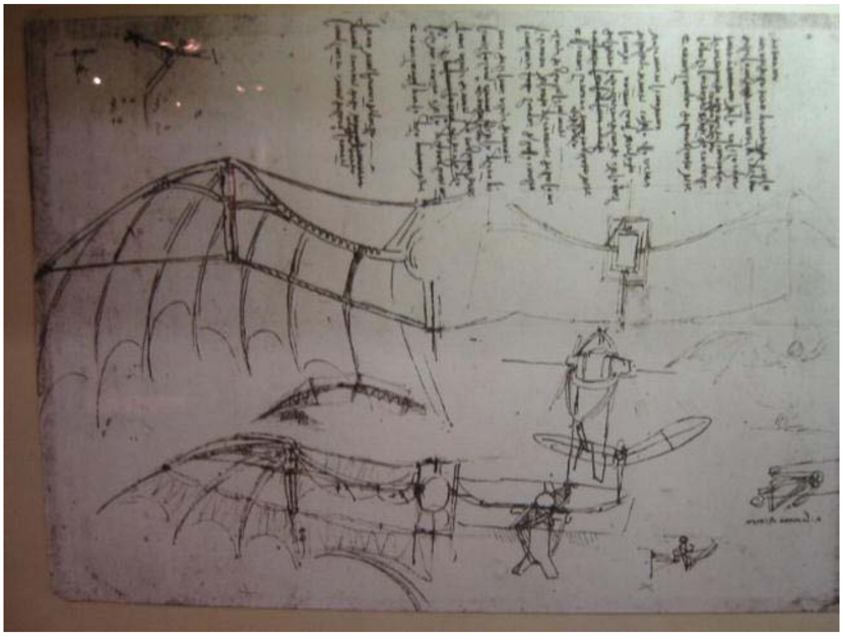

Figure 1-1. *Sketches and notes for a "flying machine" by Leonardo da Vinci. Photograph by Arnaud 25 (public domain), via Wikimedia Commons.*

Writing down your ideas also helps you pick which projects you want to work on first. It just isn't possible to work on all of them at the same time! Narrowing down the list helps get you focused and lets you concentrate all of your talents and energies on the most important projects.

When you're just beginning, the "most important projects" are the ones that provide the most positive reinforcement in the shortest amount of time. Plan a few simple, straightforward projects with a high probability of success at first. You can become more ambitious and take greater risks when you've built up a good reserve of self-confidence and techniques. Remember, the only experiments that fail are the ones from which you don't learn anything. Sometimes learning how *not* to do something is just as important as learning how to do it correctly.

Goals

The best way to turn great ideas into spectacular projects is to have a plan. Set some achievable goals for yourself. When you're first starting, don't be afraid to take those baby steps. You have to crawl before you can teleport.

Project planning is its own art form, and a dark and mysterious art it can be sometimes. It really helps to have clearly defined goals and expectations early in a project. Otherwise, how will you know when you're finished?

Any large, complex problem or project can be broken down into smaller, more manageable subprojects. Unfortunately, this can be carried too far. You don't need to reinvent the wheel every time you start a new project. There's a saying that suggests, "Don't build anything you can buy." The counterpoint to this wisdom is the equally ancient saying, "Don't buy what you can build." You'll have to be the judge as to which one applies best to you and your situation.

Do you have lots of time and very little money? This is good news! This helps simplify the decision-making process enormously. Design and build everything yourself, from scratch. This can include your own tools as well as custom-made components. The trick to this approach is not get lost in all the inevitable details that will arise.

No time to spare, but there's a bit of a budget available? This is likewise good news. With few exceptions, there's someone out there that's willing to part with just what you need for a reasonable price. The trick here is to find them and make that connection.

Inspiration!

Remind yourself what your inspiration is. What initially prompted you to start thinking about building your own electronics lab? Why did you want to get into electronics as a hobby or a career in the first place? Do you even remember?

It really helps to have a firm grasp on what your underlying motives are (or were) when faced with the challenges that regularly present themselves along the way. While it's true that some things just need doing, does it really have to be *you* that does them? Could you farm out some of the work and still feel a sense of satisfaction in the completion of the project? Or do you feel the need to take the do-it-yourself mind frame to its utmost limit?

Whatever your inspiration is, or wherever it came from, do your best to keep it in your sights while pursuing your dreams and goals. It makes it that much more satisfying when you eventually attain them.

How Electricity Works

There are two ways to go about learning how electricity actually works. The first involves just a whole bunch of *fascinating* theoretical information about atoms, electrons, fairies, and dragons. The second way is to just assume that it does indeed work, and spend some time actually putting it to work.

You don't need to know everything about electricity to have fun with it. A passing understanding is sufficient to get you started. As your technical requirements grow more elaborate, you will find yourself going back to revisit some of the basic concepts, until your comprehension is sufficient to carry you forward.

The Theoretical Approach

The prevailing theories about how electricity works center around the orbits of negatively charged subatomic particles called *electrons*. Electrons are quite small, even when compared to individual atoms. Electrons normally whiz around the *nucleus* of the atom, which is usually made up of other subatomic particles called *protons* and *neutrons*. The exception is hydrogen, whose most common isotope has only a single proton and no neutrons. Protons have a positive electrical charge and neutrons have no electrical charge.

When an atom has a balance of protons and electrons, it is said to be electrically neutral. If an atom has more electrons than protons, it has a negative electrical charge. Conversely, if there are more protons than electrons, the atom has a positive charge. An atom that has either a positive or negative electrical charge is called an *ion*.

While flying round and round the nucleus, electrons tend to cluster into well-defined orbits, or *shells*. The innermost shell of an atom can hold as many as 2 electrons. Once this orbit is "full," up to 8 more electrons can be accommodated in the next larger shell. The next shell holds up to 18 electrons, and the one after that can contain up to 32 electrons. After that, the shells get smaller again, stepping back down symmetrically from 18 to 8 to 2.

Each unique combination of electrons and protons represents one of the chemically distinct *elements*. Each element has its own particular chemical characteristics, and cannot be further subdivided by conventional chemical or mechanical means. Toward the end of the *periodic table*, which arranges all the known elements into meaningful rows and columns, you'll start to find some exceptions to the shell symmetry. These heavier elements, however, don't usually turn up in most electronics labs, so we'll not be spending a lot of time there.

Individual atoms can combine into more complex structures called *molecules*. One popular molecule is water, which is made up of two hydrogen atoms and one oxygen atom. Each of the hydrogen atoms has a single electron. The oxygen atom has eight electrons. The first two electrons in the oxygen atom occupy the first shell, leaving six electrons in the next outer shell. That leaves two "open" places in the second shell, which is where the two electrons from the two hydrogen atoms come into play. The electrons are shared by the three atoms, linking them all together and forming a *covalent bond*. Water has vastly different physical properties than either hydrogen or oxygen.

It's also possible to bump an electron out of orbit around a nucleus. When this happens, the atom becomes an ion and is electrically positive. This makes it especially attractive to any loose electrons that happen to be in the vicinity. This is because electrons are negatively charged, and in the case of electrical charges, opposites attract. This movement of electrons is how an electric charge flows through a substance.

Some materials are very good at slinging those electrons around, and as such are called good *conductors* of electrical energy. Examples of good conductors are copper and silver. Some materials resist the flow of electrical charge, and can be used as *insulators*.

An important concept to understand about the flow of electricity is that it is based on the movements of tiny, *negatively charged particles*. This is referred to as the flow of *electron current*, but this term is rarely used. For some reason, the more popular, or conventional, idea of electric current flow is from the more positive point to the more negative point in a circuit. This is referred to as *conventional current* and is the standard used throughout this book.

The fairies and the dragons come into the picture when you start to look any deeper into how the building blocks of matter are actually (or allegedly) constituted. The previous quick summary of how electricity works covers less than 1 percent of the physics needed to understand even how basic atomic principles operate. Even then there are various schools of thought as to how, *exactly*, all that tiny stuff works together to produce rainbows and cousins and moon rocks. A proton, once thought to be an indivisible *elementary particle*, is now thought to be built out of a "down" quark and two "up" quarks. There are also "strange" and "charmed" quarks flitting about, along with an entire pantheon of mythological particles. Go figure.

Now all of that *fascinating* information only scratched the surface of how *electricity* works. Electricity does a lot of work for us. We use it for lighting, heating, motion; just all kinds of things. *Electronics* is where we start to get into some really bizarre physics involving semiconductors, transistors, computer chips, and other miracles of the modern age. Explaining even the basic ideas about how *electronics* works is way beyond the scope of this little book. It's certainly possible to learn and understand it, but for the purposes of setting up a lab and building some example projects, we can just assume that it does indeed work. We can start from there and have some fun with it.

If your curious mind just can't stand *not knowing* how electronics works, the interwebs are your friends, and so are the helpful folks at your local library.

A Practical Approach

An easier way to think about electricity is to skip over *why* it flows and concentrate on *how* it flows. That's what we'll be doing in the remainder of this book, for the most part.

■ **Note** Electricity flows in a circle.

Any continuous flow of electrical current travels in a circle. There are exceptions to this broad generalization, but most of the example circuits encountered in this book will have a definite circular arrangement, if you know how to look for it.

Here are some technical terms that will be used throughout this book. You don't have to memorize any of them right now (there will *not* be a test), but you should at least take a glance at them so you know where to come back when you want to know more.

> *Circuit*: A circuit is anything that forms a continuous loop or circle. The term *circuit* is formed from the same Latin word that means *circle* or *wheel.*
>
> *Amps*: The *intensity* of an electrical current is measured in units called *amperes*, named after André-Marie Ampère. This is usually shortened to *amp* in popular usage. The symbol for amperage in a circuit is *I* (for intensity), while the unit is abbreviated *A*. For example, you could say that the intensity of the electrical current (I) in a circuit was measured to be 3.75A (amperes). Electrical current is measured with an *ammeter*.
>
> *Volts*: The voltage in a circuit is the measure of the difference in electrical *potential*, or how much work it *could* do. The unit of voltage is the volt (V), named after Alessandro Volta. This is sometimes referred to as pressure or tension. For example, a "high-tension wire" means a high-voltage line, not necessarily one that has been pulled really, really tight. Voltage is, not surprisingly, measured with a *voltmeter*.
>
> *Watts*: Electrical power is measured in watts (W), named after James Watt. The easy way to calculate watts, or the amount of energy that is converted into work (either useful or not) is to multiply the voltage across the circuit by the current flowing through the circuit. For example, if a 12V battery is connected to a lightbulb that draws 2A, the bulb converts the electrical energy into both light and heat (mostly heat) to the tune of 24W (12V × 2A = 24W).
>
> *Ohms*: The resistance to the flow of electrical current is measured in *ohms*, named after Georg Ohm. A *resistor* is a simple electronic component that exhibits a fixed resistance to the flow of electrical current. The symbol for resistance is *R* and the unit of measure is the capital Greek letter omega, Ω.

Safety

Make no mistake: electricity can kill you. It bears repeating, as well as emphasis.

■ **Warning** Electricity can kill you.

Working safely with electricity is like working safely with a sharp knife, for example. A good knife is a useful tool. Life would be cumbersome and crude (not to mention short) without good tools. However, when using *any* tools it's important to understand their proper use and operation. The same goes for electricity.

The point here is not to be overdramatic about the dangers of electricity. Billions of people are safely and peacefully coexisting with electricity, largely due to wise safety rules and well-designed electrical appliances. However, there are three major ways that you can be harmed when working around electricity, and it is important that you be aware of them:

- Your body uses tiny amounts of electricity to send nerve impulses to and from your brain. Sending relatively giant shocks of electricity through your body is a good way to stop your heart from beating or your lungs from breathing. You need to keep doing both of those things, so be careful.
- Since your body is not a very good conductor of electricity, it converts the electrical current into heat, which can produce very damaging burns, both on the inside and the outside of your body.
- Additionally, many electrically related injuries are sustained while trying to *get away* (i.e., your body's involuntary reflex when exposed to dangerous voltages).

Yet working safely with electricity is easy. Observe the following rules and you'll be relatively safe, both in your lab and outside of it:

- Always assume a circuit is "live." Electrical devices can retain lethal voltages even when completely disconnected from power.
- Don't work with high voltage (more than 36 volts) unless you're specifically trained. Reading this book does *not* mean you're "specifically trained," even if you read it *twice*.
- Don't short-circuit batteries (i.e., deliberately connect the positive and negative terminals together). Even very small batteries can hold dangerous amounts of power and can produce enough heat to burn you. Even worse, it can start a fire, which is yet another unpleasant way to die.
- Be careful when using metal (conductive) tools around battery terminals so that you don't accidentally short the terminals together or to other circuitry.
- Don't assume other people are aware of these dangers.

Lighten Up!

Now it's time to have a little fun with electricity, don't you think? While it's important to keep safety in mind, it's just as important to be curious and willing to try new things. Don't let the (important) safety rules immobilize your creativity. Use these rules, like any tool, to leverage your enjoyment of this fun and exciting hobby.

Let's start with a simple circuit that is almost guaranteed to lighten things up. Let's take apart a flashlight (or *torch* for you metric types) and see how it works.

First, find a typical household flashlight. The flashlight in Figure 1-2 was purchased for $3, and it included two new D-cell batteries. Put the batteries in the flashlight and make sure it lights up when you turn on the switch. If the bulb is burned out or the batteries are dead, then this experiment is not going to be much fun at all.

DORCY
Frost Brite
ClearView Design
TrueSpot Reflector
Modèle transparent
Réflecteur TrueSpot
DORCY
HEAVY DUTY

Figure 1-2. *A typical household flashlight will be the ~~victim~~ subject of our first electrical experiment.*

Once we've established that the flashlight works as intended, we can begin the experiment.

DISSECTING A FLASHLIGHT

We don't need most of the pieces of the flashlight body. Let's take it apart and have a look inside.

1. Open the flashlight be unscrewing the end with the lightbulb in it.
2. Remove the batteries. Set them aside for the moment. We'll use them in a bit.
3. Remove the plastic retainer that holds the bulb in the reflector.
4. Take out the bulb.
5. We'll need to replace the wiring that was contained within the flashlight's body. Find a short piece of wire 8" to 12" long. A specialized jumper wire with alligator clips on the end was used in the photos for this exercise, but almost any kind of wire will do. See Figure 1-3.

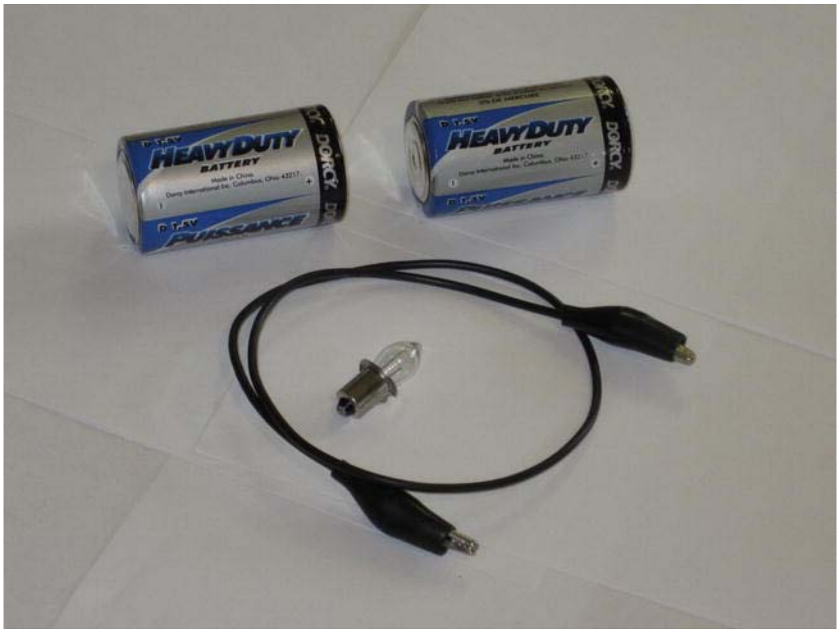

Figure 1-3. *All the pieces you'll need to light up the bulb.*

6. If you're using plain, insulated wire, you'll need to remove a small bit of it from both ends of the wire to expose the *conductor*, or metal wire part inside.
7. This step can be tricky. Try to hold both batteries in one hand, along with the bulb. The batteries need to be pointed in the same direction. If this is too hard to do, try it with just one battery.

8. The positive end of the battery has a small cap on it. Touch the bottom point of the bulb to the top of the battery.
9. Now, using your other hand, connect the wire from the bottom of the battery to the metal base of the bulb.
10. The lightbulb should start to glow. See Figure 1-4.

■ **Caution** If you have a super-bright bulb, it can get hot. Don't burn yourself!

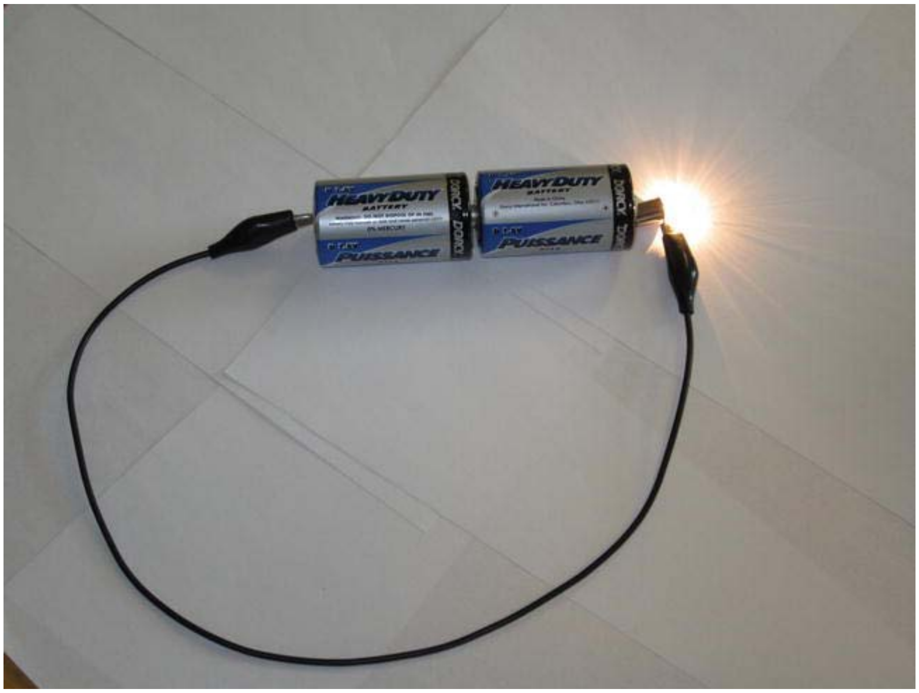

Figure 1-4. *The completed circuit, showing the glowing bulb. Electricity is flowing!*

Several important things are going on in this very simple circuit. Electrical current is flowing, producing useful work (in this case, light). Can you see the "circle"? When you *break* the circuit, the bulb stops glowing. This is exactly how the switch in the flashlight body works. Peek inside the flashlight and see if you can tell how it works. By connecting and disconnecting the wire, you can make the bulb flash. If you only use one battery, the bulb turns on, but is not as bright. Why do you think that happens?

Environmental Impact

You affect your environment every day, and your environment affects you, as well. As you learn more about electronics and all the interesting and wonderful things it can teach you, be mindful of what kind of trail you're leaving behind for others to clean up.

A good way to get a feel for your environmental impact is to take a look at your garbage. It doesn't sound that interesting or fun, but it *will* teach you a lot about yourself and your habits. What are you throwing away? Could you have found either another use for it, or perhaps another home? Is it going to be a nuisance to someone else, or even worse, a hazard?

A good example is the use of "disposable" batteries. We used batteries (technically *cells*, but that distinction will be spelled out in more detail in Chapter 3) in our very first experiment. We buy new batteries and plug them into our devices, and when they run out of power we take them out and throw them away. Where do they go? How do they then interact with the environment? They don't just "go away."

Disposable batteries are very convenient. They are generally inexpensive, compact, self-contained, and reliable. This *convenience*, however, comes with a price all its own. Consider using *rechargeable* batteries whenever possible in your lab. They cost a little more than their disposable counterparts, but can easily and cheaply be refilled with electrical goodness over and over again.

You should also be aware of any hazardous materials that you use or store in your lab. For example, lead is a very common material used in many electrical devices, even though its use is being reduced worldwide. It is found in older-technology rechargeable batteries, solder, and printed circuit boards.

Some of the chemicals used to etch printed circuit boards can be dangerous if not used, stored, and disposed of properly. Take the time to understand what, exactly, you're working with when you're conducting your experiments in your laboratory.

Just like the captain of a seagoing vessel, you're in charge of what happens in *your* lab. With this authority come important responsibilities that rest squarely upon your shoulders. Please take these responsibilities seriously.

Budgeting

How much is all this stuff going to cost? A lot of it depends on how creative you want to be in stocking your lab with tools and parts.

If your budget is small, then start small. If your budget is nonexistent, then you'll just have to start even smaller. You should generally approach acquisitions for your lab under one of these three headings:

- Buying new
- Finding used
- Harvesting or recycling

Buying New

Buying everything new is obviously the most expensive approach, at least from the *money* side of it. What a lot of people fail to take into account is that their *time* is also a very valuable commodity. Try to balance the one with the other, as befits your particular situation.

Except for the folks in the most *dreadful* of hurries, it always makes sense to shop around for the tools, equipment, and components that you are going to need in your lab. You'll find that a lot of "professional-grade" equipment carries an enormous price premium compared with "consumer-" or "hobby-grade" alternatives. This is generally to be blamed on the fact that there is a correspondingly enormous amount of money to be made in the field of electronics.

New components are almost always more desirable to use, especially when you're looking at a production environment and you need to make sure everything produced is of the highest possible quality and reliability. New, unused components are also more appropriate when conducting complex experiments, as this helps to reduce, if not eliminate, many variables in the testing and analysis phases. You don't have to worry that someone (probably you) might have "borrowed" some parts for a quick fix or some other temporary usage and then surreptitiously replaced them back in inventory, having possibly compromised them in some unseen way.

Finding Used

A prudent way to fill up a new lab, especially when first starting out, is to look for bargains on used tools and components. This is especially true of the fixtures, shelves, cabinets, benches, and other furniture that belongs in a lab.

Tools wear out. Tools are replaced. This is a fact of life when you're in the business of working with tools. The good news is that "one man's trash is another man's treasure." If we consider the possibility of women participating, we get three more potential combinations, all of which can end up in a win-win situation.

Many used tools can be retrofitted and placed back into service in your lab. On the other hand, there are a lot of tools that are simply used up and can not feasibly be brought back into useful service. Either their repair is cost prohibitive or they are no longer supported by their original manufacturers.

Another factor working in favor of the patient and observant collector of discarded tools is the inevitable march of progress. What was shiny and top-of-the-line a year ago might be intolerable or insufficient next year, from the standpoint of the original purchaser. Many bargain opportunities arise from this simple understanding.

This can sometimes be applied to the purchase of components that have never been used, but have been sitting somewhere waiting for their chance. A commonly used phrase to describe these parts, which are often sold at a discount, is *new old stock (NOS)*. Sometimes these parts are sold this way because of overstock or because a manufacturing run was cancelled or cut short.

Harvesting or Recycling

There's an absurdly large amount of perfectly good equipment, tooling, fixtures, and components sitting out there, looking for a new home. A lot of companies, as well as individuals, either outgrow their present fittings or turn toward other interests and endeavors. This can be an excellent opportunity to acquire the furnishings you need to populate your lab.

On the other hand, please don't go crazy. You don't have to pick up every discarded appliance or chair you see sitting by the roadside. Why pay rent for something you can easily and cheaply obtain when you actually need it? Are you really wanting to pay the heating and air-conditioning bill for that box full of odd-sized capacitors that are probably past their prime? Don't fall into the trap of the *false economy* when it comes to bargain-hunting for your lab. The most important ingredient in your lab is *you*. Leave a little room to move around a bit.

Another factor to keep in mind is *shelf life*. Not all parts (or tools) last forever, especially when not being used. Batteries are the worst. You can prolong the shelf life of nonrechargeable batteries by storing them in your refrigerator, as this slows down the electrochemical processes within them that produce the electricity. Even solar panels will wear out with either constant use or constant disuse, although this can take years to happen.

Some Example Workshops

Here's a little peek into some actual electronic laboratories, each with its own personality and history. Each one grew from something small into something, well, not so small.

A Fresh Start

Harold Timmis is an electronic engineer, software developer, and author. His curiosity about how things worked started as a small child, which led him from taking apart everything in his parents' house to studying engineering at the Florida Institute of Technology.

Harold keeps a lab in his home where he can work on gadgets in his spare time. Harold recently moved, and used the opportunity to "clean house" and set up everything in his lab exactly the way he wanted things.

He took this photo of his new lab after working on a couple of projects (see Figure 1-5). Note that everything is within easy reach when sitting at the desk. It's still quite tidy, as you can see.

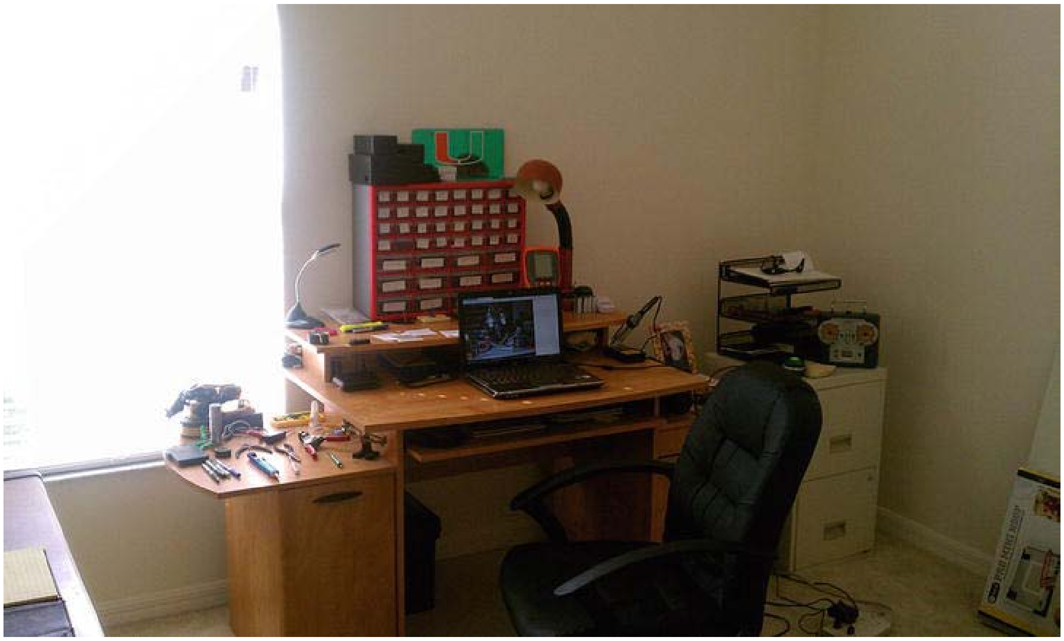

Figure 1-5. *The reorganized home lab of Harold Timmis (photo by Harold Timmis)*

The Robot Clubhouse

The Dallas Personal Robotics Group (`http://dprg.org`) is the world's oldest personal robotics group known to man. From 2002 to 2009, Mike Dodson of Modern Assemblies donated the use of a portion of one of his warehouses to DPRG. This gave DPRG a permanent home for meetings, contests, and social gatherings.

Robot builders enjoyed a spacious and comfortable area for working on robot projects, among other pursuits. Over the years, members contributed time, equipment, and lots of hard work to improving the space. The photo in Figure 1-6 shows one end of the available space, after an extensive remodeling. Several workstations along the wall are in the process of being set up. That good-looking fellow seated at the table is the author.

When Mike retired in 2009, DPRG began a search for another home.

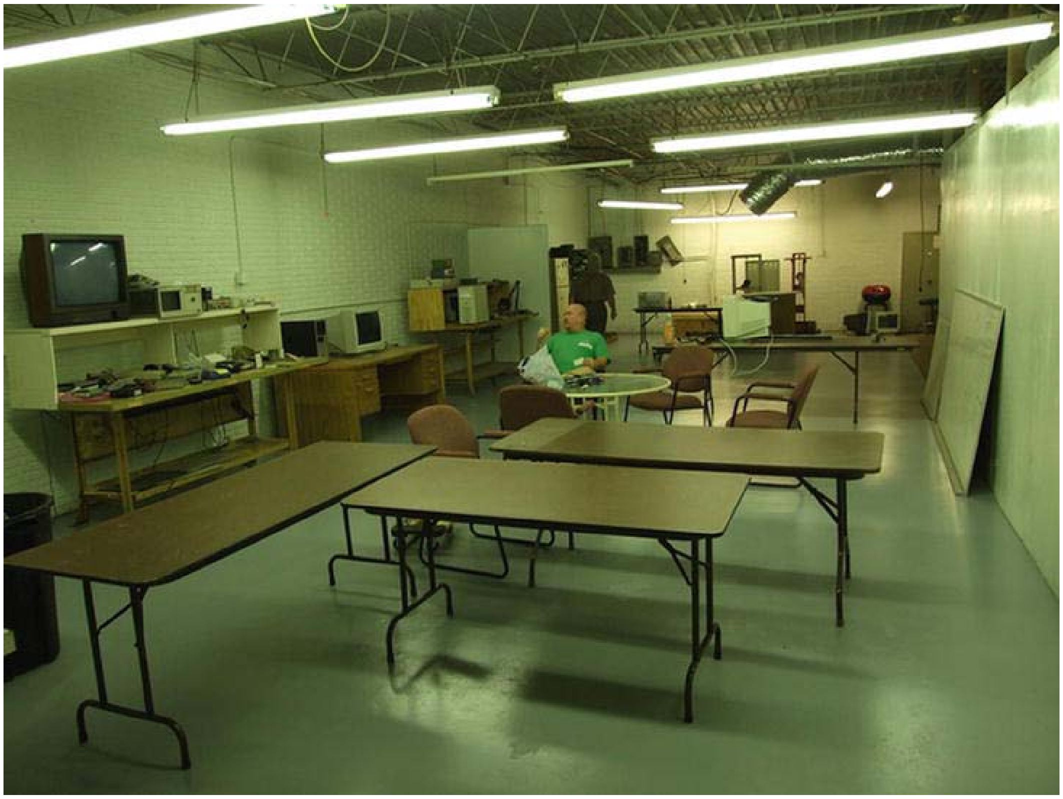

Figure 1-6. *DPRG's world headquarters from 2002 to 2009 (photo by R. Steve Rainwater)*

A Cautionary Tale

It's relatively easy to go from a well-designed and effective lab to something, let's say, less desirable. This doesn't happen overnight, but it can happen. Let this be a warning to you about how even the best of intentions can still produce embarrassing and unproductive confusion (see Figure 1-7).

Unfinished projects and leftover parts and tools compete for the limited space available in this photo. Not only is a lot of time wasted in searching for both components and tools, but the limited space restricts the scale of projects that can be entertained.

Don't let this happen to you.

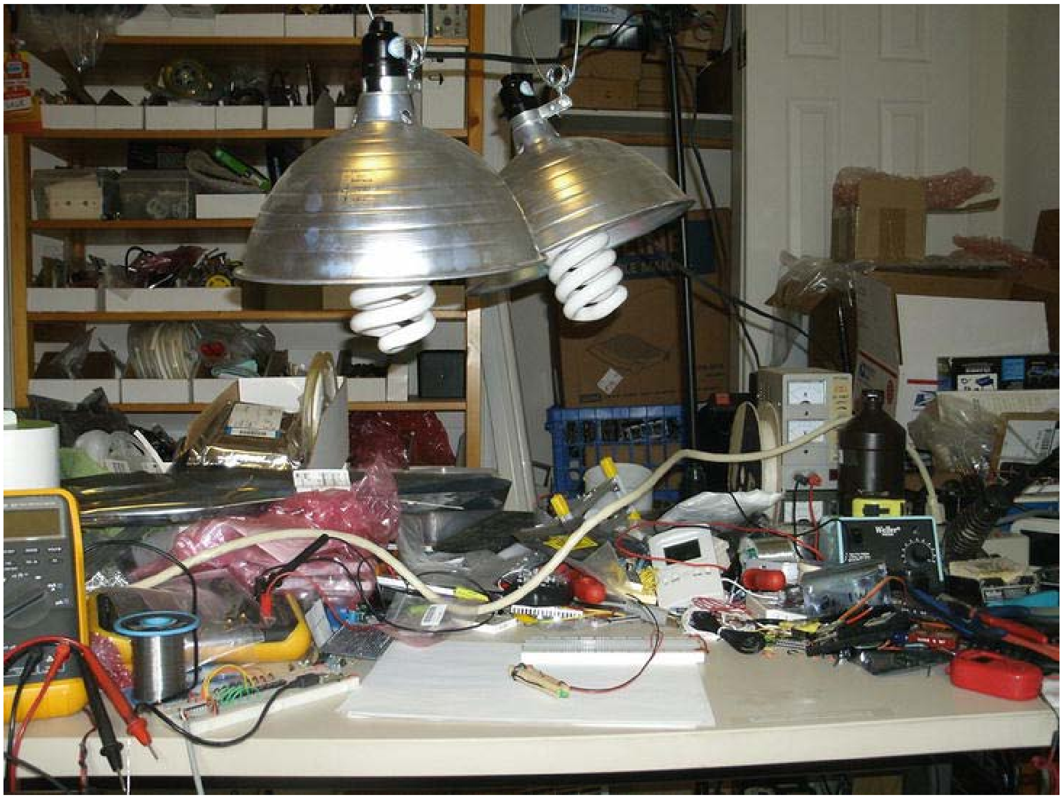

Figure 1-7. *It's not safe. It's not pretty. It's not even properly "eccentric." You have been warned.*

Summary

So now you know a little something about electricity and electronics. Hopefully this chapter has covered your very basic questions about how all this magic stuff actually works.

Now you're ready to start poking around some circuits. You're going to need the right tools for the job. Chapter 2 will get you started. You'll learn about some basic hand tools that are very useful in the lab, as well as some fairly high-tech machines that you might require in the future.

CHAPTER 2

Building Your Tool Chest

You're going to need some tools in your lab, as well as the skills to use them effectively. This chapter will get you started. If you're starting from scratch, don't worry. There's not a whole lot you absolutely *must* have to get started. Some basic hand tools and a place to keep them organized is all you will need. You don't even have to spend a lot of money at first. As you progress in your hobby, you will most likely want to add to your tool chest and upgrade some of those tools. Electronics can be the perfect hobby because it can take up (1) all your spare time and (2) all of your discretionary income. What more could you ask?

The very basic hand tools, such as screwdrivers and tweezers, are mostly self-evident in their usage. Try not to poke yourself in the eye. Where some helpful advice is appropriate, it will be given.

Here's an excellent example of a simple safety rule that will absolutely improve and extend your life: when using any kind of cutting tool, always direct the cutting motion away from yourself. For example, you might receive a package in the mail, and get all excited about opening it. You then discover it's sealed up with indestructible packing tape and can't simply be ripped open with your bare hands. Use a small knife or box cutter and make a single cut at a time, starting from the point closest to you on the package and cutting outward. Using this method, should the knife slip for any reason, the blade will travel away from you. This allows your blood and other bits to stay inside you, where they can do the most good. Never cut toward yourself.

If you've already got some tools (or even a *lot* of tools), it never hurts to review their proper usage. Also, if you're suffering from an *overabundance* of tools, to the point where you can no longer keep them organized, this chapter might help you to prioritize them, at least as far as permanent residency in your lab is concerned.

■ **Tip** Every tool in your shop must *earn* its place there. It must *deserve* to be there. *No exceptions*.

The Most Important Tool Ever

The most important tool you'll ever use is your mind. Every other tool you will ever work with depends on the correct and proper functioning of your mind. If you're careful with it, you won't misplace it, neglect it, damage it, or use it for the wrong job. Your mind is a wonderful asset when cared for properly, and a terrible liability when ill-treated.

Your mind is in many ways like a muscle. The more you use it, the stronger it gets. Learning new things does not "push out" the old things you've already learned. Your brain is capable of forging new neural pathways when challenged, and there's plenty of extra capacity in that old noggin of yours to handle whatever may arise.

■ **Tip** Try to learn something new every day.

Take care of that brain of yours. Just like you wouldn't use a dull drill bit or a rusty saw, don't try to work on electrical or electronic projects when your mind is not in the right place. Ideally, you should never try to get any work done in your lab when you are sleepy, distracted, angry, upset, medicated, or overly tired. Deal with whatever problems you're facing in a responsible manner and then get back to work in the lab.

- Are you sleepy? Take a nap or head off to bed for the night. Really bad mistakes are made when the brain does not get enough sleep. Sleep deprivation is notorious for being obvious (and obnoxious!) to everyone else except you. "Really bad mistakes" made while working with electrical or electronic projects can produce very hazardous conditions for you and everyone around you, such as fire, damaged equipment, injuries, and even death.
- Are you angry or upset? Then your mind is not going to be adequately focused on what you need to do in the lab. "Fools! I'll destroy them all!" are famous last words *for a reason.* Give yourself some time to calm down from whatever has angered or upset you. Also, it's OK for you to give yourself permission to admit that you are angry or upset about something in the first place. Ignoring it probably won't make it go away, and probably won't make your feelings about it go away, either. Walk away, count to ten, or whatever works for you.
- Feeling distracted? Then now is *not* the time to be conducting electronic experiments in the lab. You need to be *present* in the lab, both bodily and mentally. There are just too many ways to make simple blunders that have far-reaching and sometimes overly expensive consequences.
- Have your cognitive functions been compromised? This can happen from taking certain medicines or even recreational intoxicants. If the little bottle says, "Don't operate machinery or drive an automobile," then you probably shouldn't be making important decisions or working with dangerous circuits in your lab, either. Wait until you're clear-headed enough to proceed with your plan for world domination, adding blinking eyes to a plush doll, or what have you.

The worst thing about cognitive impairment is that the first thing that gets impaired is the ability to judge the level of your cognitive impairment!

Care and Feeding of Your Brain

Here are a few tips to keep your brain (and the rest of you) in good working order:

- Stay hydrated. Drink more water than you think you need. This might result in more trips to the bathroom, but this is much more desirable than the alternative. A good flow of water through your systems helps your body feed and repair itself, as well as aiding in the elimination of toxins. Drink more water.
- Avoid excess consumption of coffee, tea, sodas, and especially "energy" drinks. These contain ingredients that stress both your body and your mental capacity. They do not contain energy, but trick your body into going into overdrive by triggering a stress reaction.
- Take frequent breaks. Get up and walk around, if possible. Raise your arms over your head and hold them there as you count to ten. This redirects blood flow and helps restore a balance within your peripherals.

All Those Other Bits, Too

Just as your mind is the most important tool you'll ever use, the rest of your bodily bits and personal pieces are vital and important in your work in the lab. For the most part, there are no spare parts laying around for things like your eyes, your ears, your hands, or your skin.

Should you wear eye protection in the lab? Yes, you should. Are you always going to do so? Probably not. It's true, it can be inconvenient or perhaps a tad uncomfortable. Consider the cost, however. Even the tiniest scratch on your eyeball can take weeks to heal and leave you susceptible to infection or vision loss. It's just not worth it to be lazy. Ask the next one-eyed person you see. If you're a bit uncomfortable asking one-eyed people, "Hey, what's with the one eye?" then perhaps you can just try to imagine them advocating or encouraging you along the same path they took, whether intentionally or accidentally. They won't do it.

Invest in quality eye protection that is comfortable and properly sized for you. Provide eye protection for visitors to your lab as well, if they will be present when soldering, welding, machine tooling, or wire clipping is being performed. Then invest in the habit of wearing your eye protection whenever it is appropriate.

■ **Note** Don't pretend that your prescription glasses with polycarbonate lenses are "safety glasses." They aren't. They do not protect either your eyes or your face from impact or objects entering from the side, as proper safety glasses should. If you usually wear glasses, use safety glasses that comfortably accommodate your glasses when worn.

Will you be working with really loud noises? These could be from powerful audio amplifiers ("it goes to 11") or power tools. Hearing loss from exposure to excessively loud noises is cumulative. This means it slowly (or suddenly) chips away at your hearing, bit by bit. Why do you think old people turn up the TV so loud?

Hearing protection is both affordable and comfortable. For temporary protection, you can use disposable earplugs. If you need protection all day, consider earmuffs designed for hearing protection. Don't let "Hey, y'all, check this out" be the last thing you ever hear.

It's hard to work with hand tools with no hands. It's possible; just really hard. Take care of your hands. There are a hundred ways to burn, tear, cut, or otherwise mangle those mitten-fittings of yours. Have a variety of work gloves on hand to handle the different needs of your lab. The choice will be ultimately determined by the particular hazards present in your lab. Are you working with chemicals as well as electronics? How about sharp edges or glassware? Normally, you won't need to wear gloves for soldering small electronics, but if you start working with really large-scale projects, such as sculptures or plumbing, you might.

Make sure you are still "handy" when wearing the proper gloves for the job. Make sure the gloves fit you properly and allow you the adequate dexterity you need.

■ **Tip** Be safe. Don't be lazy.

Basic Hand Tools

Assuming that you're going to start small and build toward bigger things, you should also start small with some small hand tools. These tools will make it much easier to work on small parts and wires.

Good hand tools increase your effective strength and improve the accuracy and precision of your motions. They allow you to focus a great deal of your strength on exactly the part that needs it, without wasting a lot of effort. Good hand tools also insulate you from sharp, hot, or otherwise unpleasant objects when you work on them.

Take care of your hand tools. Put them up when you're done with them. Make sure they are clean and ready to be used again the next time you need them. Few things are more frustrating than attempting to start a quick repair job and finding out that you need to spend twice as much time preparing your tools for use.

Have some sort of methodical plan or reasoning involved with the storage of your tools. They need a safe and comfortable home, just like you do. You need to be able to find them quickly when you want them. Try out different organizational strategies until you find one that works for you and your lab. This is an easy task when you have just a few tools to organize. Once you've accumulated a large collection of hand tools, this is no longer optional. You *must* know what tools are available and where they are located.

Frequently used tools should be within easy reach of your primary workstation. If your tools need any sort of adjustment or setup before use, make sure you've got all the proper bits handy. For example, a drill press will often have a chuck key used to tighten and loosen the chuck that holds a drill bit in place. That key needs to be in the immediate vicinity of the drill press, as well as easy to locate. Don't spend a lot of time needlessly hunting for tools. Decide for yourself on some safe and practical policies for where you keep your tools.

■ **Tip** Don't waste time guessing where your tools are located. *Decide* where they are located.

Just like everything else, hand tools have a limited useful lifetime. While some tools can easily outlast you, it's up to you to keep track of exactly how much utility you can continue to expect from your tools. This doesn't mean that you have to trash your hand tools once they develop some tiny flaw. If it's beyond your capabilities to repair the tool, then consider replacing it and reassigning it to the second string or backup tool chest. Having a spare tool, even a slightly imperfect one, on hand when you need it can make the difference between success and failure in a project.

Bear in mind, however, that some tools are simply too dangerous to keep around once they've outlived their useful life span. A dull cutting tool that cannot be reasonably resharpened is a good example of a tool that has lost its honored place among your other tools. Thank it for its service, wipe away that little tear, and then figure out a good way to recycle any of its bits that are still functional.

Wire Cutters and Wire Strippers

As you might have picked up from Chapter 1, you can spend a lot of time working with wire in this hobby. Understanding a little bit about how wire is made and how it functions in a circuit will aid you in your experiments.

How Wire is Made—The Very Short Version

Most wire is made of copper or a copper alloy. Some wire used for electrical wiring in houses is made of aluminum. The wire itself starts out as a fairly thick strand and is stretched and pulled—a process called *drawing*—until the desired diameter is obtained. This strand is then wrapped up on spools or bobbins.

- Some wire is made of a single strand of copper (or aluminum). This is called *solid-core* or *single-strand* wire. It is a single strand of wire, usually circular in cross-section.
- Multiple, thinner strands of wire are often bundled or twisted together to make larger-diameter wire. Stranded wire is much more flexible than solid-core wire, but is more expensive to produce because of the extra manufacturing steps required. Because stranded wire has a larger surface area per unit of length, it is more susceptible to environmental hazards such as corrosion.

Most electrical wire is covered with an insulating jacket, usually made of plastic or rubber. This sheathing must be removed to make electrical connections. There is sometimes more than one layer of insulation on a wire, depending on what its final application is destined to be.

More information about wire as a *component* will be presented in Chapter 3. For now, the emphasis is on the characteristics of wire that help determine the best tools to use when working with wire.

Selecting the Correct Tool

Both copper and the aluminum alloys used for electrical wiring are fairly soft metals and can be easily cut or formed with simple hand tools. Wire *cutters* are specialized tools made to make a clean cut through the wire's strand or stands. Wire *strippers* are another kind of specialized tool used to remove the insulation from a wire.

Figure 2-1 shows two rows of commonly available wire cutters.

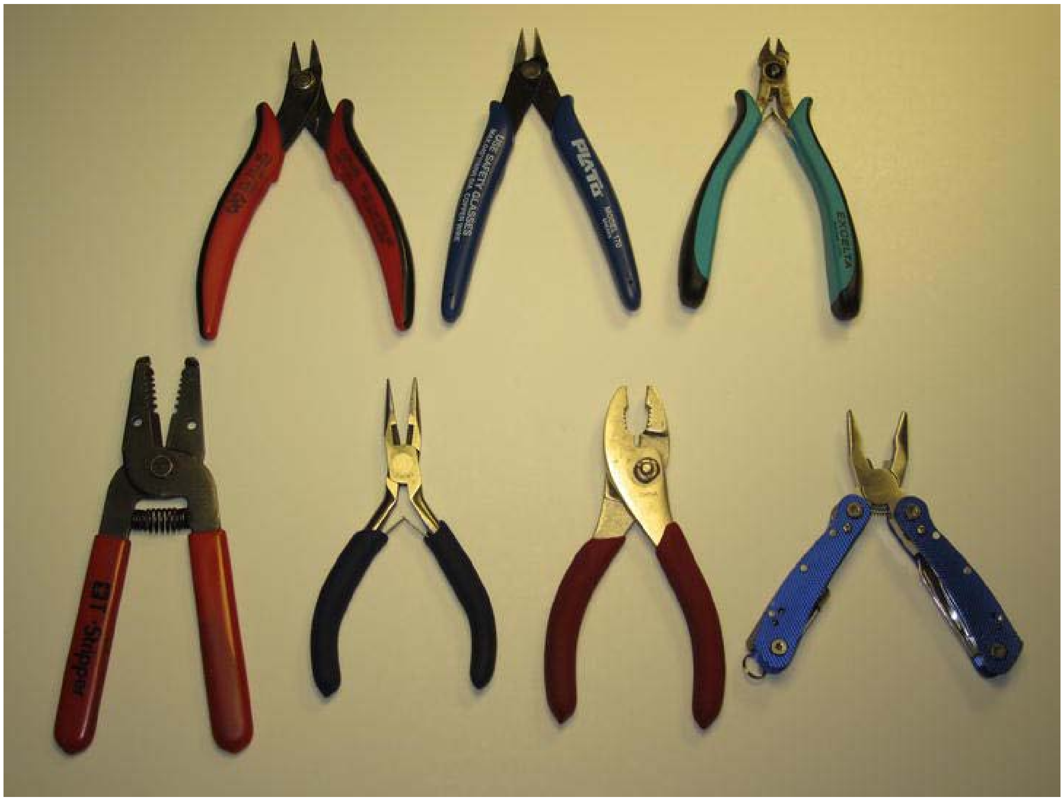

Figure 2-1. *Commonly available wire-cutting tools. The top row shows three styles of diagonal cutters, which are the preferred tools for cutting small-gauge copper wire. The bottom row shows some tools that offer wire cutting as a secondary facility, and should only be used when proper wire cutters are not available.*

The top row of cutting implements in Figure 2-1 are various forms of what are known as diagonal cutters, side cutters, or flush cutters. These cutters have beveled cutting edges that meet in the middle of the wire being cut. One side of the cutters is flat, or *flush*, allowing wires to be cut with some precision. These are the preferred cutters for small-gauge copper wire, as well as for clipping excess component leads from a printed circuit board (PCB). See Figure 2-2.

Compare the cutting action of the side cutters to the other, less optimized wire cutters in the second row of cutters in Figure 2-1. These cutters either use a shearing action to cut the wire or include a flush-cutting area as a secondary feature.

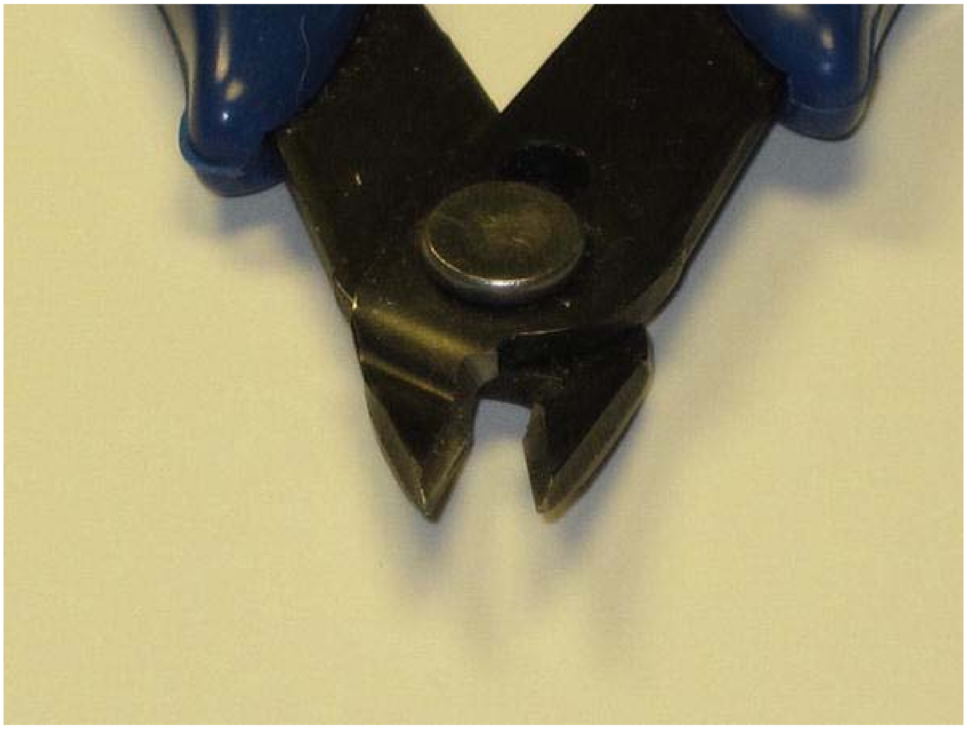

Figure 2-2. *A close-up look at the business end of a pair of side cutters. When the two cutting edges meet in the middle, the bottom side is flat, or flush, allowing the user to get quite close to where a wire or component lead emerges from a solder joint.*

Use caution when clipping component leads or small wires with any wire cutters. The cut lead can easily fly away at sufficient velocity to poke you right in the eye or other sensitive area. If possible, hold *both* ends of a wire being cut or direct the cut leads *away* from your face.

■ **Caution** Use eye protection when clipping leads. Don't say you weren't warned about this hazard.

Wire Strippers

Wire strippers are tools that allow you to easily strip the nonconductive insulation from wires. You can try chewing it off with your remaining teeth, but this is not recommended.

Wire strippers usually consist of a pair of hinged blades with some sort of machined notch. The notch allows the blades to cut the insulation around the wire without cutting or nicking the underlying conductor. A good example is shown in Figure 2-3.

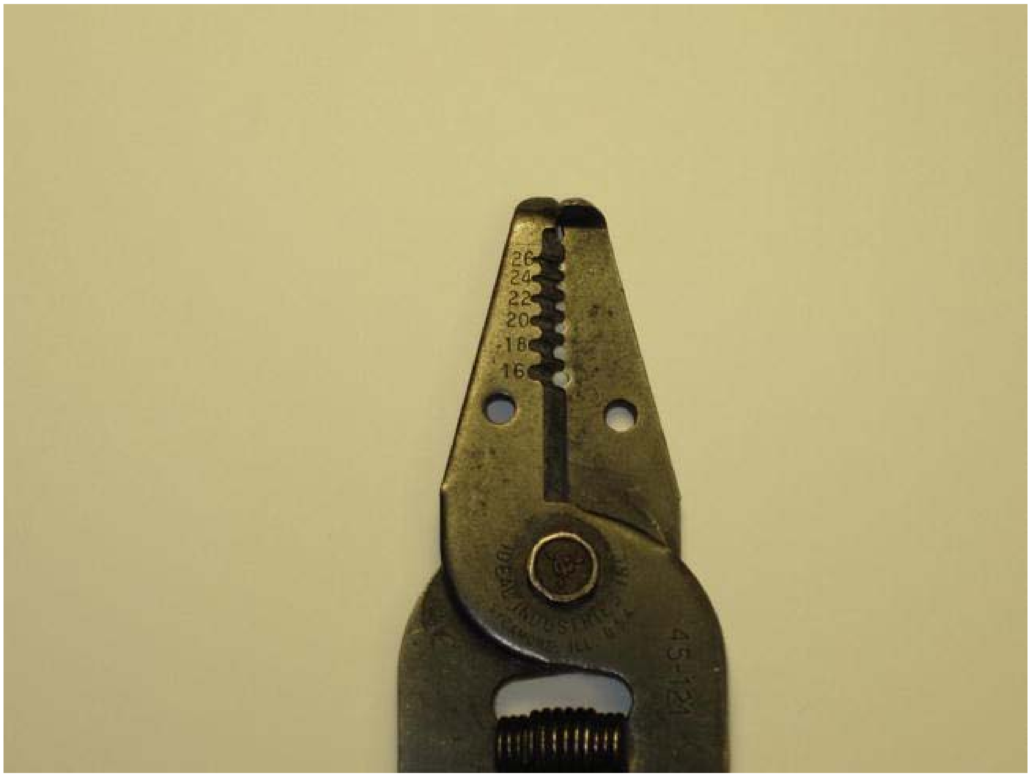

Figure 2-3. *The popular T-Stripper combination wire cutter and stripper tool from Ideal Industries has been around for over half a century. Note the wire gauge sizes stamped into the face of the cutter blade. New models feature laser-engraved text.*

The wire strippers in Figure 2-3 can easily strip the insulation from wires with a conductor diameter, or *gauge*, between 16 (0.0508" or 1.291 mm) and 26 (0.0159" or 0.405 mm). The higher-numbered wire gauges represent *smaller*-diameter wires. Once upon a time, the gauge number represented the number of *drawing* operations that were performed on a wire when being manufactured. The more times it was drawn through the processing equipment, the thinner the wire became. Today, the wire-manufacturing industry in the United States and Canada has adopted the American Wire Gauge standard (AWG). See Table 2-1.

Table 2-1. *The AWG. A more complete listing is available at http://en.wikipedia.org/wiki/American_wire_gauge, from which this partial list was obtained.*

AWG	Diameter Area			Copper Resistance	
	Inch mm		mm^2	Ω/km mΩ/m	Ω/1000 ft mΩ/ft
0000 (4/0) 0.46	00	11.684 107		0.1608	0.04901
000 (3/0)	0.4096	10.404	85.0 0.20	28	0.06180
00 (2/0)	0.3648	9.266	67.4 0.25	57	0.07793
0 (1/0)	0.3249	8.252	53.5 0.32	24	0.09827
1 0.28	93	7.348	42.4 0.40	66	0.1239
2 0.25	76	6.544	33.6 0.51	27	0.1563
3 0.22	94	5.827	26.7 0.64	65	0.1970
4 0.20	43	5.189	21.2 0.81	52	0.2485
5 0.18	19	4.621	16.8 1.02	8	0.3133
6 0.16	20	4.115	13.3 1.29	6	0.3951
7 0.14	43	3.665	10.5 1.63	4	0.4982
8 0.12	85	3.264	8.37 2.06	1	0.6282
10 0.10	19	2.588	5.26 3.27	7	0.9989
12 0.08	08	2.053	3.31 5.21	1	1.588
14 0.06	41	1.628	2.08 8.28	6	2.525
16 0.05	08	1.291	1.31 13.1	7	4.016
18 0.04	03	1.024	0.823 20.9	5	6.385
20 0.03	20	0.812	0.518 33.3	1	10.15

22 0.02	53	0.644	0.326 52.9	6	16.14
24 0.02	01	0.511	0.205 84.2	2	25.67
26 0.01	59	0.405	0.129	133.9	40.81
28	0.0126 0.32	1	0.0810 212.	9	64.90
30	0.0100 0.25	5	0.0509 338.	6	103.2

We'll return to this table again in Chapter 3 when the electrical properties of wire are discussed. For now, note that as the wire gauge number goes *up*, the wire's physical diameter goes *down*. Similarly, as a conductor's physical size goes down, its ability to carry current also goes down, and its electrical *resistance* goes up.

■ **Note** Other standards exist throughout the world. Steel-wire manufacturers use a different scale, as their interest is more on the tensile strength of the wire, not its electrical properties.

Note that the AWG is for single-strand copper wire. Equivalent gauges in stranded wire are physically a bit larger, although they have the same effective cross-sectional area and can carry the same amount of current.

When using wire strippers, keep in mind that your goal is to cut and remove only the nonconductive insulation from the wire, leaving the internal conductor intact. Even a small nick can result in a weak point in the wire, both electrically and mechanically.

Practice Makes Perfect

The best way to attain proficiency with a new tool is with practice—lots and lots of practice. This is true for working with your hands as well as with that big brain of yours. Ideally, any task worth doing will involve at least a little bit of both manual dexterity and focused brain power.

Cutting and stripping wires is an excellent example. For your first practice assignment, spend some time turning long, beautiful lengths of wire into little bits of wire with stripped ends.

CUTTING AND STRIPPING WIRE

For this exercise, you’re going to need the following:

- Wire cutters
- Wire strippers
- Much wire
- A comfortable, well-lit location

Cut the Wire into Segments

Don’t watch TV or be otherwise distracted. Pay attention to what you are doing.

Begin by cutting the wire into various-sized segments from 2" to 6" in length. Accuracy in length is not critical at this point. You need to get a good feel for your wire cutters and how much force is needed to cut the wire, without cutting yourself or others.

Collect the cut wire segments into a bowl or small container for the next step.

Strip the Wire Segments

Using your wire strippers, remove about 1/4" (6 mm) of insulation from both ends of each wire. Try not to cut or nick the wire core. The insulation that you remove can either be discarded or collected for use in some sort of art or craft project.

Once you’ve completed the cutting and stripping portion of this exercise, sort the wires according to their length, or their color if you used different colors of wire. Compare the quality and consistency of your work by looking at the first few wires and the last ones you did. Do you see any improvements in your work?

If you spend about an hour working on your wire-cutting and wire-stripping technique, you will perhaps master a skill that will prove useful for the rest of your life, as well as prepare a big pile of wires that you can use in many of the sample projects illustrated in this book.

Pliers and Tweezers

Every day, the size of the “typical” electronic component gets smaller and smaller. This miniaturization process has brought about a world where you can carry a phenomenally powerful computer in your pocket, and sometimes even make phone calls with it. The individual transistors, resistors, capacitors, and other components inside such modern marvels are literally microscopic in size. You need a fairly powerful microscope to be able to see them at all.

In your typical home electronics lab, and even in many professional labs, you will not be dealing with these tiny things, except as large clumps all tied together inside a chip or other framework. Still, the typical components you will use are often a bit too small for all but the tiniest of fingers. That’s where pliers and tweezers come into play. They are an extension of your own hands and fingers, increasing your precision and allowing you to perform delicate tasks with tiny parts.

There are several types of pliers and tweezers that you may use on a daily basis in your lab. See Figure 2-4.

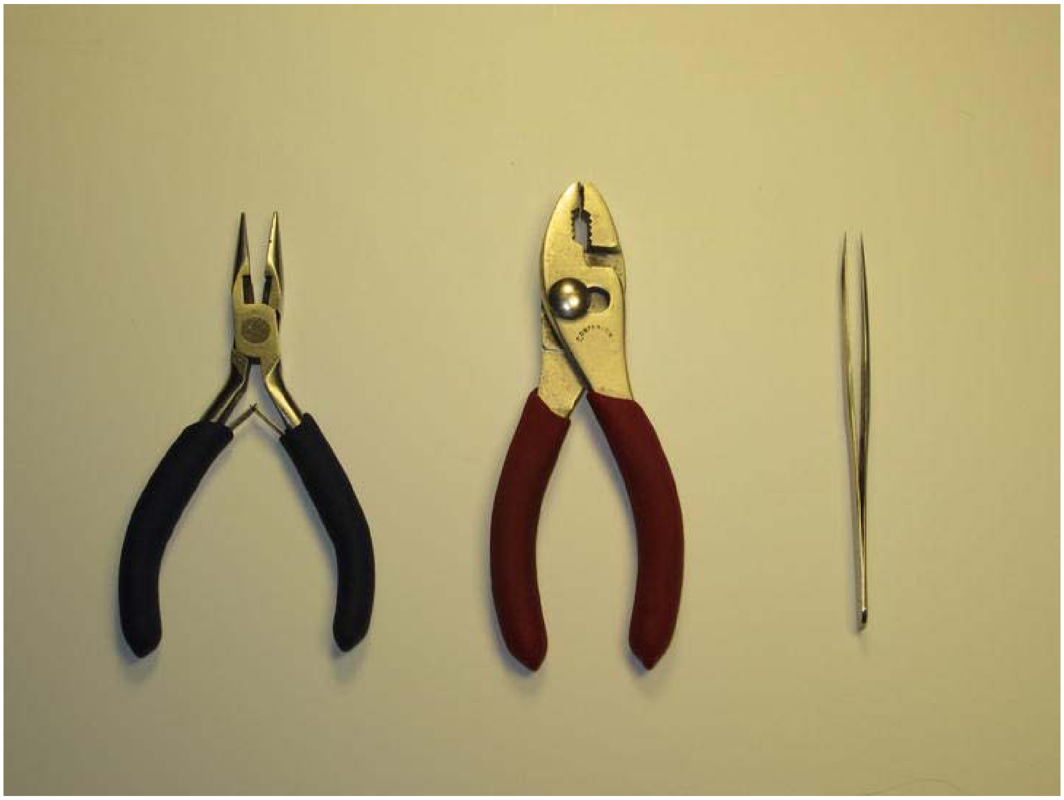

Figure 2-4. *Pliers and tweezers aid in working with small parts. On the left is a pair of needle-nosed pliers with cushioned grips and leaf springs. In the center is a pair of slip-joint pliers that can be adjusted to accommodate two different ranges of gripping action. On the right is a pair of needle-sharp tweezers, which can be used to pick up and place even the tiniest of electronic components—if your hand is steady enough!*

Screwdrivers, Nut Drivers, Wrenches, and Ratchets

A lot of electrical connections can be made using only a screwdriver. Figure 2-5 shows an electrical connection using a *barrier strip*. A specialized connector is crimped onto the end of the wire and held in place with a screw. A variety of crimp-on connectors are available and should be installed using the appropriate tool, as indicated by the connector's manufacturer. Just smashing the connector onto the wire with some pliers is not recommended.

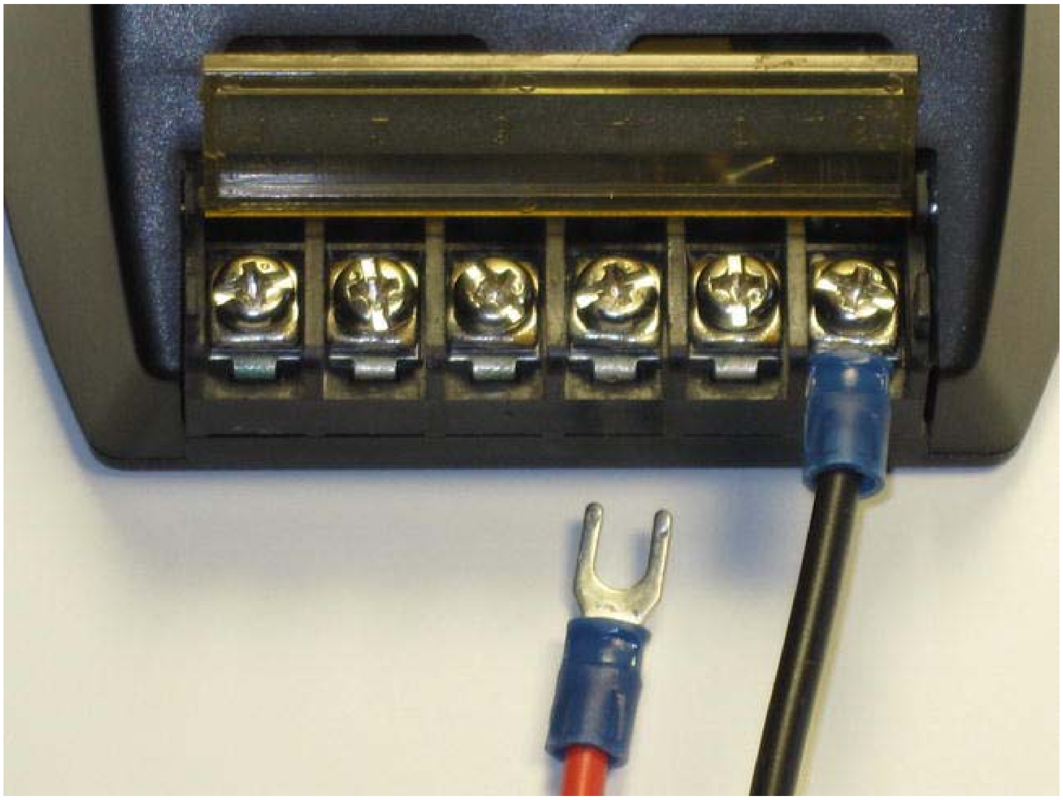

***Figure 2-5.** A barrier strip on a low-cost solar-charge controller. No soldering required! The specialized connector is held in place with a screw. The connector is affixed to the end of the wire by deformation using a specially designed crimping tool. You can see the indentations in the insulating collar around the connector.*

Another type of connector that is tightened with a screw is called a *terminal block*. Unlike a barrier strip, no specialized connector is required on the wire. You can simply insert a stripped wire into the terminal block and tighten the connector using a screwdriver. Try to strip the wire only enough to allow the bare conductor to enter the terminal block, without leaving any noninsulated wire exposed that could accidentally short out against another wire. See Figure 2-6.

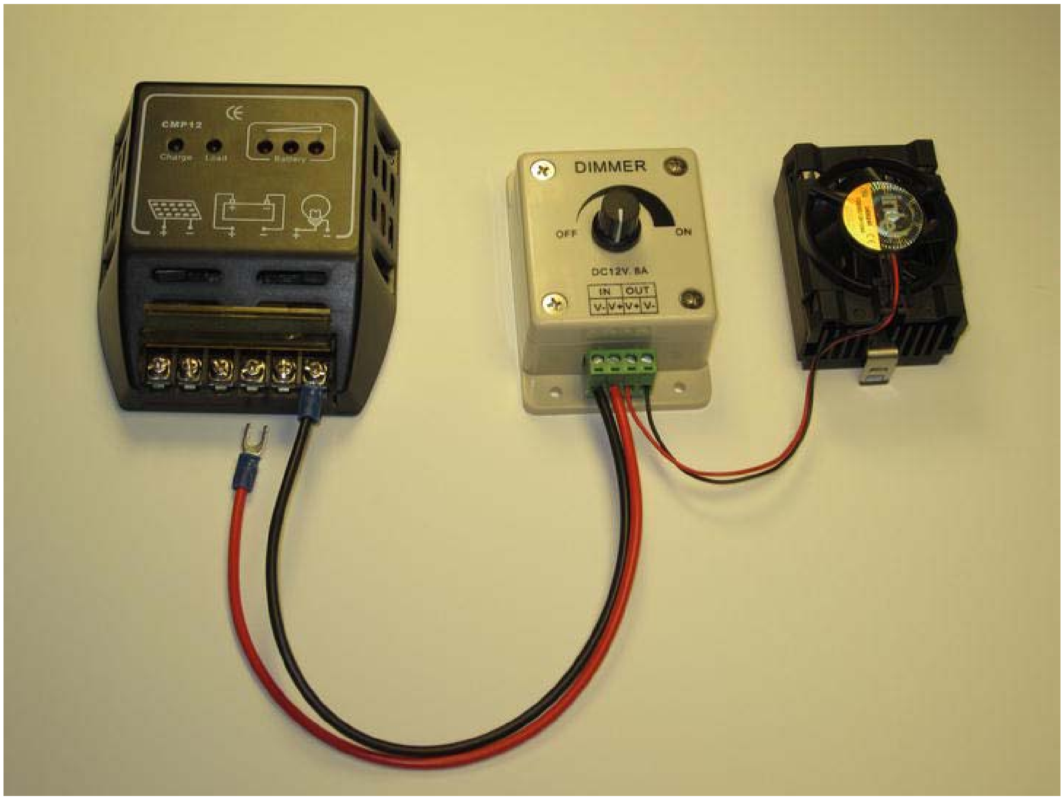
CMP12
Charge
Load
Battery
DIMMER
OFF
ON
DC12V. 8A
IN
OUT
V-
V+
V+
V-

Figure 2-6. *A terminal block uses a screw for tightening the connection with a bare wire.*

Besides being useful for making electrical connections using barrier strips and terminal blocks, screwdrivers are also quite handy in their original function, which is to tighten and loosen mechanical fastening devices such as screws and bolts. See Figure 2-7.

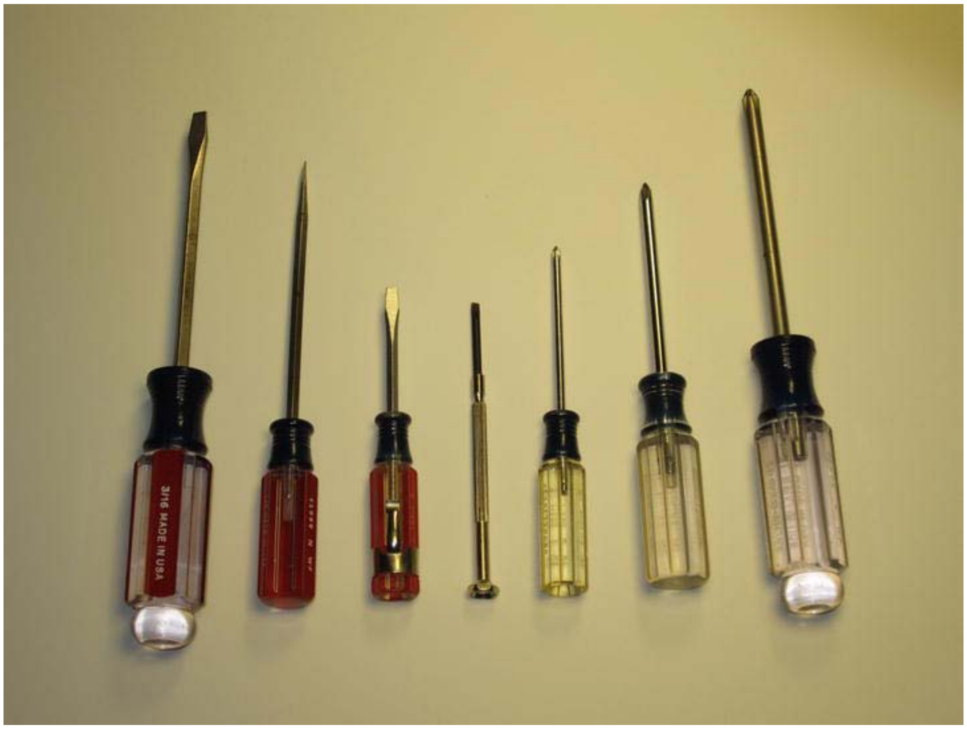

Figure 2-7. *A variety of screwdrivers. The four on the left are standard, or straight-bladed, screwdrivers. The three on the right are Phillips-style screwdrivers. Other specialty bits are also available.*

You should only use screwdrivers to drive screws and bolts. Don't try to use them as levers, pry bars, chisels, punches, or other tools. When you need a chisel, use a chisel. Keep your screwdrivers clean and dry. Put them back where they belong when you're done with them. You get to decide where they belong.

Screws, Bolts, and Nuts

There are a *very* large number of different kinds of screws and bolts. Some screws, such as wood screws, are made to be driven directly into wood and create their own threads in the material as they are installed. Other screws, largely referred to as machine screws, mate with a premachined or tapped hole, or a machine nut.

Nuts are often shaped with four or six sides that allow them to be held in place while the screw is turned. Alternately, the nuts themselves are turned while the screw is held in place. Some screws and bolts have a similarly shaped head that allows them to be turned with a wrench or nut driver instead of a traditional screwdriver.

Please resist the urge to overtighten screws and bolts. The mechanical advantage gained by the inclined plane of the screw thread multiplies the force used to turn the screw into a terrific amount of pressure on the screw itself. You can easily damage either the screw or the threads in the material to be fastened by overtightening a screw.

More Specialized Hand Tools

A good set of wire tools, pliers, and screwdrivers will go a long way when working with electronics. However, sometimes you need a more specialized tool for more specialized work.

Here are some of the more common tools that are a little more focused on a particular task than they are for general-purpose usage.

Wire Crimpers

As mentioned previously in this chapter, you can make solid electrical and mechanical connections using barrier strips and specially terminated wires. The optimized connectors used for this purpose require optimized tools for best performance.

Figure 2-8 depicts some examples of specialized crimp tools for electrical connections.

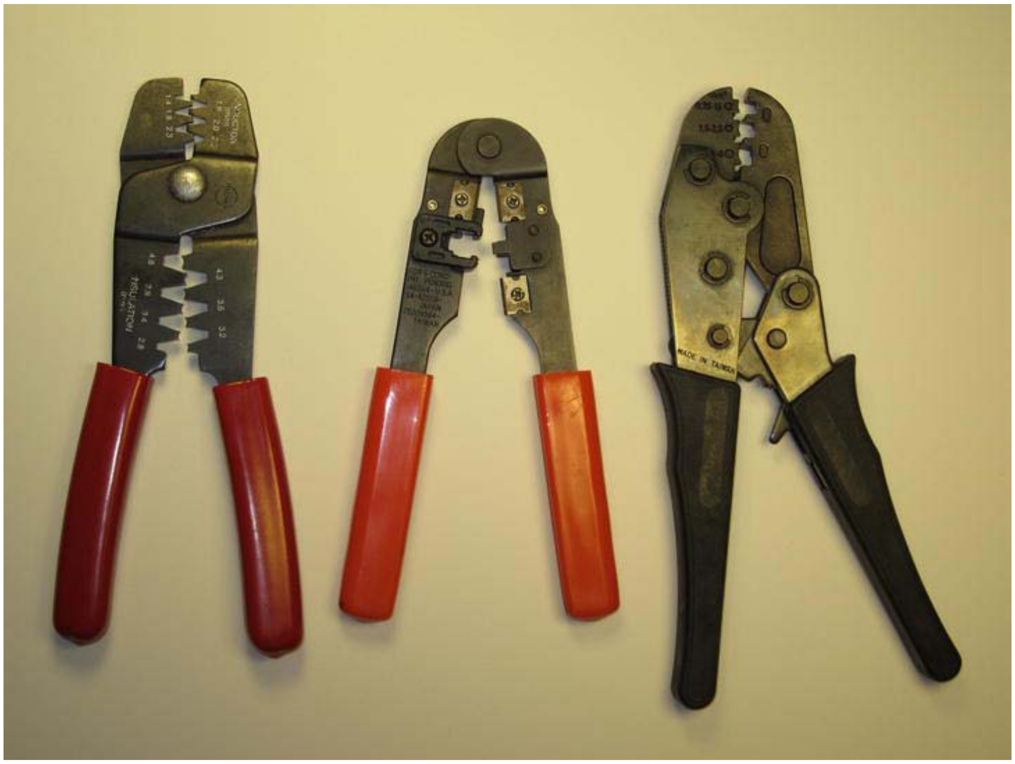

Figure 2-8. *Some common crimping tools for affixing electrical connectors to wire*

Referring again to Figure 2-8, the crimpers on the left are for crimping tiny pins to the ends of tiny wires that are then inserted into plastic connector bodies. The center crimp tool is for crimping modular jacks onto telephone cables. It also includes a wire-cutter blade (bottom) and a pair of controlled-depth stripper blades for removing only the outside layer of insulation from silver-satin flat telephone cord, without disturbing the individually insulated wires contained within. The crimp tool on the right features a ratcheting action that increases the crimping force applied to the connectors. This is the tool used to crimp the connectors show previously in Figure 2-5.

Computers

You might not think about computers as being "power tools" in your toolbox. As the size of computers, laptops, tablets, and smart phones decreases, and their computing power, networking ability, and storage capacities increase, they become much better candidates for use in the lab.

Traditional "desktop" personal computers (PCs), laptops, netbooks, and tablets can perform multiple roles in your laboratory setting. If you are lucky enough to have Internet access on your computer, it can become a conduit for all types of useful electronics information, in addition to its more accepted role as a source of funny cat videos. See Figure 2-9.

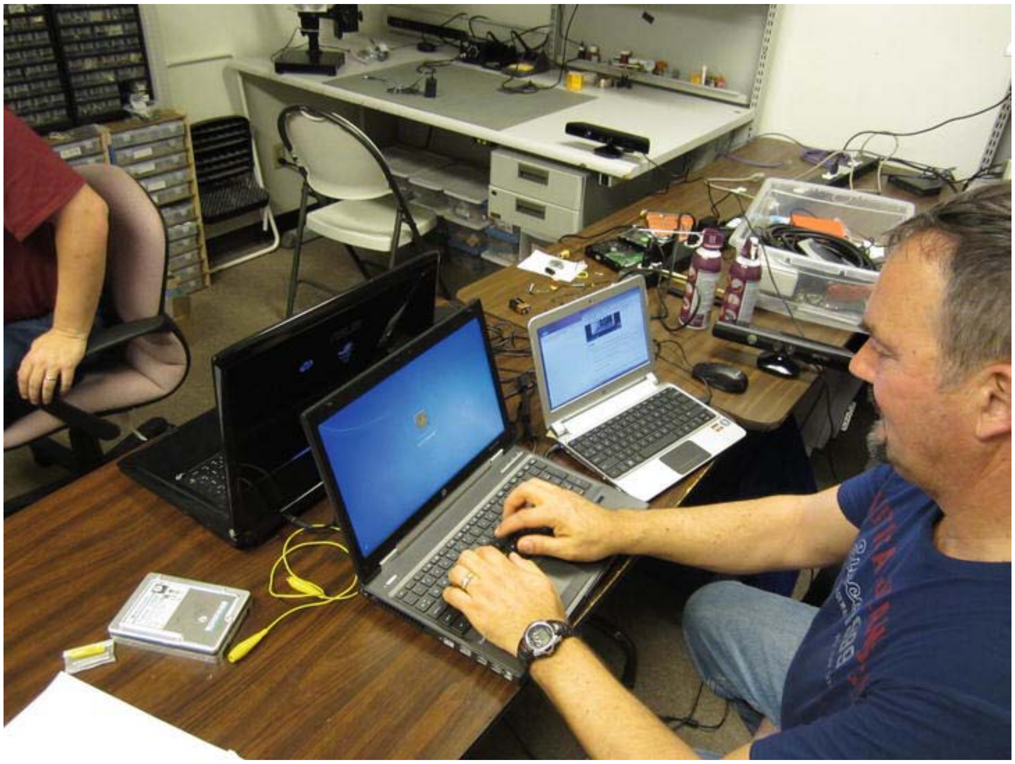

Figure 2-9. *Computers being used in the lab. Paul Bouchier of DPRG uses two laptops at once when testing new robot software with the Microsoft Kinect (just to the right of his mouse).*

Here are just a few of the many possible areas of practical knowledge, learning opportunities, and bits of interesting information that you can access using your computer:

> *Data sheets*: Almost all electronics manufacturers provide online access to a library of technical data sheets and application notes for their products. These should be your first choice for information about how and when to use their products. These sources also contain important product *errata*, which contains corrections and updates for products that have been made since their first publication.
>
> *Online tutorials*: The Web abounds with detailed and thought-provoking tutorials and background information on every topic imaginable, including, not surprisingly, mountains of information on computers and electronics.
>
> *User forums*: People tend to cluster around others with similar interests. Once upon a time when people where geographically limited and unable to communicate or interact with folks around the world, this severely restricted a lot of the more specialized areas of knowledge. Today, you can find user groups centered on every conceivable human endeavor (and some inconceivable ones, too) that are filled with helpful, knowledgeable members that are more than happy to answer questions and share experiences with you. Do try to be courteous, respectful, and fair when asking for help with your projects, and always try to repay in kind by helping those with similar questions that you've had to master.
>
> *Sharing/collaboration web sites*: As the cost and complexity of publishing your work on the Internet becomes more manageable, it becomes faster and easier to let others know about your progress, as well as your problems. You can easily set up your own web page for little or no cash, or join many of the available user groups that allow their members to post, browse, and otherwise benefit from the works of other similarly minded individuals.
>
> *Design tools*: You can find many free or reasonably priced electronic design and simulation tools online. Some manufacturers also provide free tools specifically for optimizing designs for their products.
>
> *Online suppliers*: Where you once had to walk (or drive) down to the store and see what was available in your area as far as electronic components and tools went, now you can browse through the catalogs of suppliers from all over the world, from the comfort of your own lab. Not only does this provide a much larger field from which to choose, it also levels the playing field, so to speak, forcing worldwide competition for your hard-earned cash. Bonus: Some manufacturers offer a limited numbers of free samples. Yip!

This list only scratches the surface of the kinds of information you can access using a computer. You can also use a computer as part of your electronic projects, and several examples of doing this are included later in the book.

Tools for Soldering

You don't have to learn how to solder to work with electronics, but it helps. It really does. It opens a whole new world of possible projects, repairs, and custom modifications to existing equipment. Just remember, as always, to keep your work area clean and safe. See Figure 2-10.

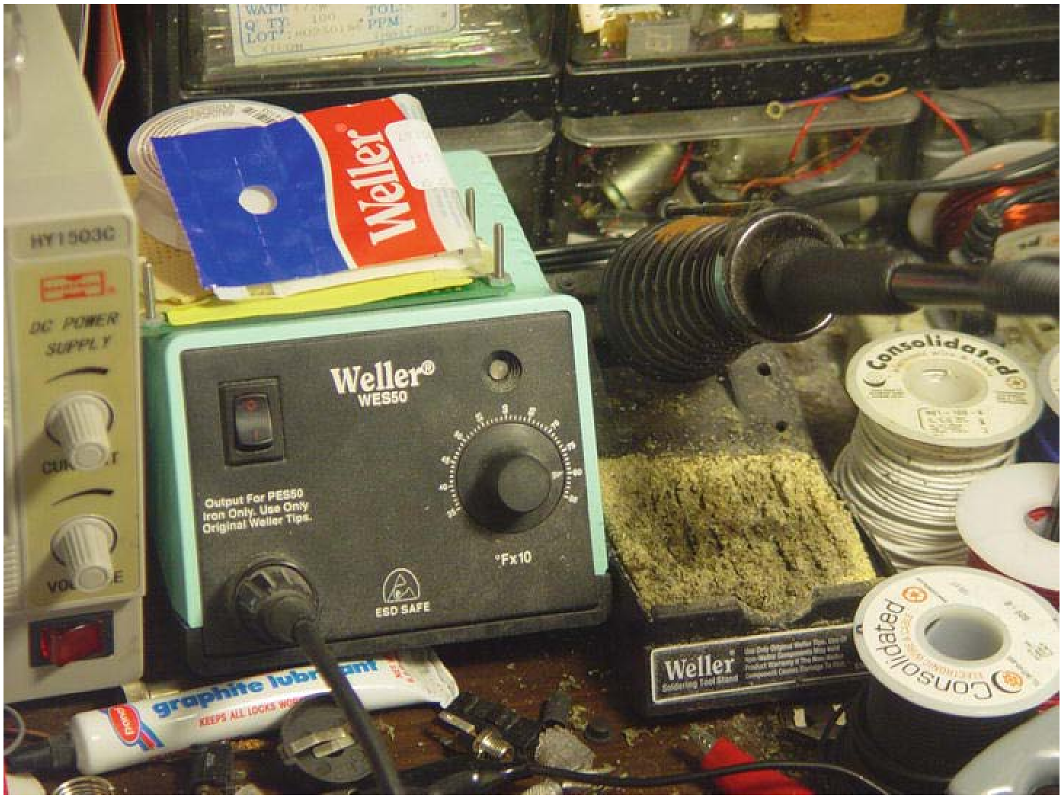
Weller
Weller®
WES50
Output For PES50 Iron Only. Use Only Original Weller Tips.
°Fx10
ESD SAFE
Weller
Soldering Tool Stand
HY1503C
DC POWER SUPPLY
graphite lubricant
KEEPS ALL LOCKS WORKING
Consolidated
Consolidated

Figure 2-10. *A Weller controlled-temperature soldering station. What a mess!*

Like any worthwhile skill, soldering is best mastered with lots and lots of practice. The best way to learn this useful skill is to have someone show you how it's done. It's difficult, but not impossible, to learn proper soldering techniques from a book. That being said, have a look at Figure 2-11.

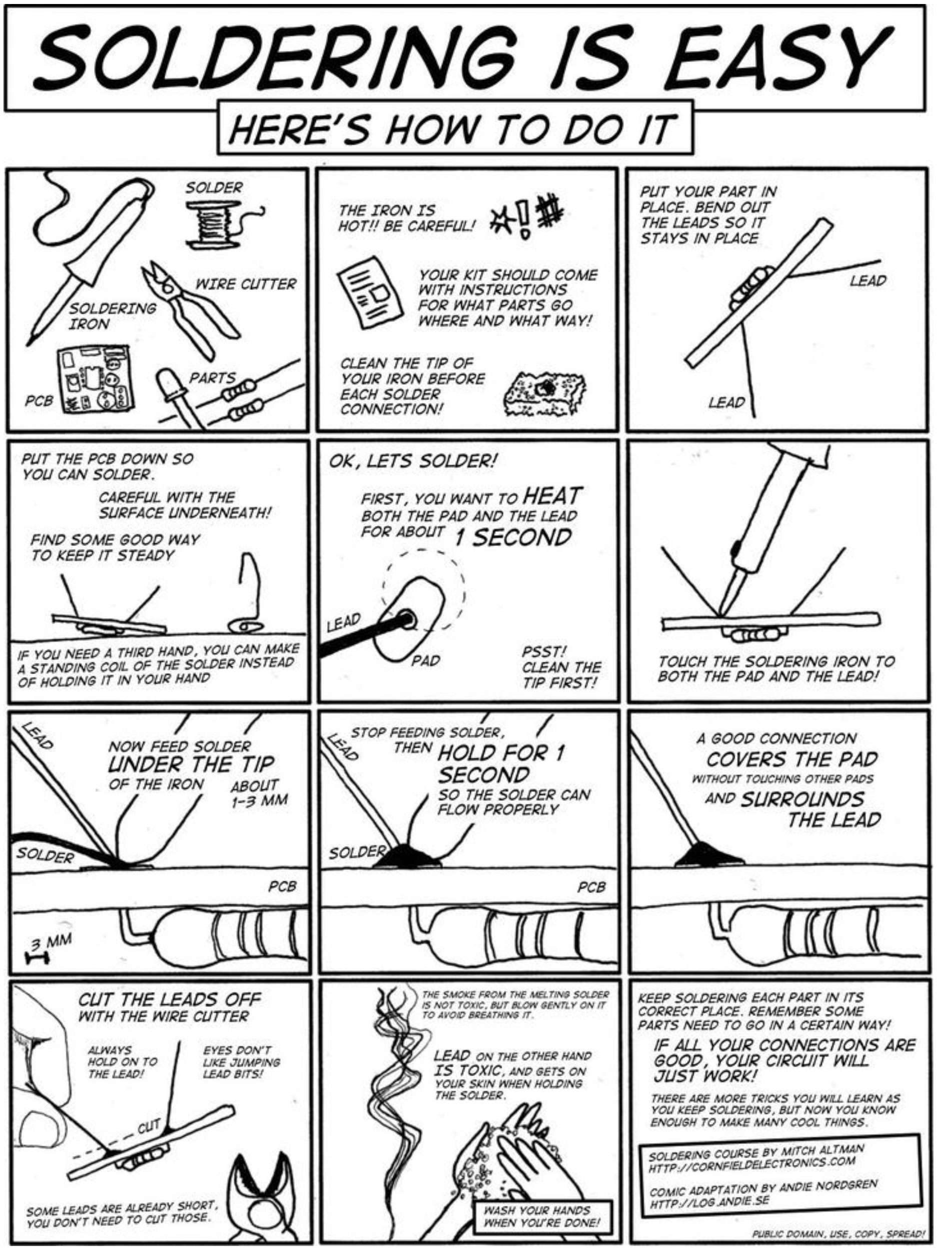
SOLDERING IS EASY
HERE'S HOW TO DO IT
SOLDER
WIRE CUTTER
SOLDERING IRON
PCB
PARTS
THE IRON IS HOT!! BE CAREFUL! ☆!#
YOUR KIT SHOULD COME WITH INSTRUCTIONS FOR WHAT PARTS GO WHERE AND WHAT WAY!
CLEAN THE TIP OF YOUR IRON BEFORE EACH SOLDER CONNECTION!
PUT YOUR PART IN PLACE. BEND OUT THE LEADS SO IT STAYS IN PLACE
LEAD
LEAD
PUT THE PCB DOWN SO YOU CAN SOLDER.
CAREFUL WITH THE SURFACE UNDERNEATH!
FIND SOME GOOD WAY TO KEEP IT STEADY
IF YOU NEED A THIRD HAND, YOU CAN MAKE A STANDING COIL OF THE SOLDER INSTEAD OF HOLDING IT IN YOUR HAND
OK, LETS SOLDER!
FIRST, YOU WANT TO HEAT BOTH THE PAD AND THE LEAD FOR ABOUT 1 SECOND
LEAD
PAD
PSST! CLEAN THE TIP FIRST!
TOUCH THE SOLDERING IRON TO BOTH THE PAD AND THE LEAD!
LEAD
NOW FEED SOLDER UNDER THE TIP OF THE IRON
ABOUT 1-3 MM
SOLDER
PCB
3 MM
LEAD
STOP FEEDING SOLDER, THEN HOLD FOR 1 SECOND SO THE SOLDER CAN FLOW PROPERLY
SOLDER
PCB
A GOOD CONNECTION COVERS THE PAD WITHOUT TOUCHING OTHER PADS AND SURROUNDS THE LEAD
CUT THE LEADS OFF WITH THE WIRE CUTTER
ALWAYS HOLD ON TO THE LEAD!
EYES DON'T LIKE JUMPING LEAD BITS!
CUT
SOME LEADS ARE ALREADY SHORT, YOU DON'T NEED TO CUT THOSE.
THE SMOKE FROM THE MELTING SOLDER IS NOT TOXIC, BUT BLOW GENTLY ON IT TO AVOID BREATHING IT.
LEAD ON THE OTHER HAND IS TOXIC, AND GETS ON YOUR SKIN WHEN HOLDING THE SOLDER.
WASH YOUR HANDS WHEN YOU'RE DONE!
KEEP SOLDERING EACH PART IN ITS CORRECT PLACE. REMEMBER SOME PARTS NEED TO GO IN A CERTAIN WAY!
IF ALL YOUR CONNECTIONS ARE GOOD, YOUR CIRCUIT WILL JUST WORK!
THERE ARE MORE TRICKS YOU WILL LEARN AS YOU KEEP SOLDERING, BUT NOW YOU KNOW ENOUGH TO MAKE MANY COOL THINGS.
SOLDERING COURSE BY MITCH ALTMAN HTTP://CORNFIELDELECTRONICS.COM
COMIC ADAPTATION BY ANDIE NORDGREN HTTP://LOG.ANDIE.SE
PUBLIC DOMAIN, USE, COPY, SPREAD!

Figure 2-11. *Instructions for soldering. This delightful illustration by Andie Nordgren captures the basics of soldering as taught by soldering guru Mitch Altman (http://log.andie.se/post/397677855/soldering-is-easy).*

To solder, you will need a soldering iron. You will also need some solder and probably some wire cutters. That's just the tip of the iceberg, however. Like any good tool, your soldering station will soon accumulate many *accessories* to enhance your soldering prowess.

To begin your soldering adventure, you will need a *small* soldering iron, some solder, and a damp sponge. The big, pistol-grip style of soldering irons are for much larger projects, such as plumbing. Ideally you want something in the 15W to 50W range. See Figure 2-12.

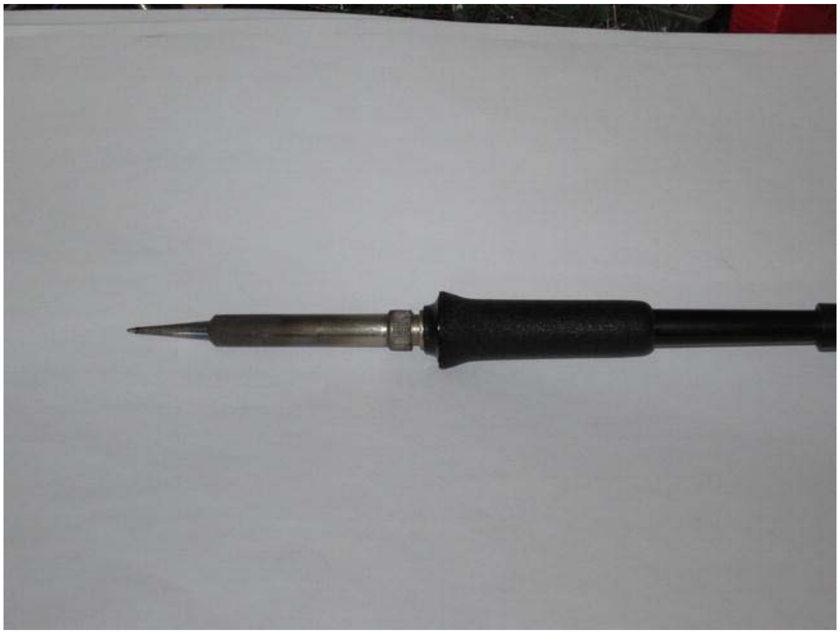

Figure 2-12. *The pencil-grip style of soldering irons is better suited for working on small electronics projects.*

The soldering iron in Figure 2-12 has a replaceable, conical tip. These tips are generally interchangeable for a particular model of soldering iron, but unfortunately not interchangeable between different models or manufacturers. Different tip shapes are also available, depending on what types of things you're wanting to solder.

The tip is usually made of copper, which is an excellent conductor of heat. Since copper really likes to combine with other metals at high temperatures (which is exactly what happens when soldering), most tips are coated in a thin layer of steel to give them much longer useful lifetimes. Tips still wear out, however, and should be replaced.

In this particular soldering iron, the next part of the iron is the removable barrel that holds the tip in place. This barrel is where the heating element is housed, so remember that it is just as hot as the tip! Try not to touch it. You will probably not want to make that mistake more than once.

The next section is the padded and insulated handle of the soldering iron, which is made of dense plastic foam. A quality soldering iron will be comfortable in your hand and won't produce excessive fatigue, even after hours of use. Lower-quality and lower-priced models will more rapidly take their toll on your hands.

Not shown in Figure 2-12 is the flexible power cord running back to the soldering iron's base. This cord supplies the electrical power that is converted to heat in the barrel, as well as some wires for a temperature sensor. This allows the unit to control the temperature of the tip with great accuracy. Note the temperature dial in Figure 2-10. Lower-cost units omit the temperature control features and run at a constant power output.

It will take some amount of time for the soldering iron tip to reach a working temperature. Likewise, it will also take some time to cool off once you've turned off the soldering iron.

■ **Caution** Soldering irons are very hot in use and can cause painful burns. Always treat them with respect!

The solder you use will be determined by what kind of soldering you want to do. For most small, electronics-related soldering projects, a small-gauge tin-lead or lead-free solder can be used. Bear in mind that lead is a poisonous material that slowly accumulates in your system, should you come into contact with it, as your body has no good way of eliminating it. Wash your hands before and after soldering, just to be safe.

Soldering works best when everything being soldered is sparklingly clean. To help facilitate this cleanliness, chemical compounds known as soldering rosin are used. These compounds chemically clean the surfaces being soldered when exposed to high levels of heat. Some types of solder have rosin incorporated into them. You can also obtain rosin as a paste into which you can dip wires and component leads before soldering. Other types can be applied with a dispensing pen.

■ **Caution** Don't breathe the smoke emitted during the soldering process. If possible, use a fan or fume extractor to redirect the soldering fumes away from you.

Once your soldering iron is up to its proper operating temperature, clean the tip by wiping it gently on a damp sponge. Your tips will last longer if you resist the urge to clean them before replacing the soldering iron in its holder. Even though it seems counterintuitive to "put it up dirty," this will more than double the life of your soldering iron tips. Just be sure to clean the tip *before* you solder anything.

Power Tools

Everybody loves power tools, don't they? Sometimes it's enough to just *imagine* that you might need one to justify a trip to the store and subsequent purchase. While this may be human nature (or perhaps only the nature of one particular gender), it doesn't fully justify having one of each in the lab, *just in case*.

Drills and Drill Bits

You will, however, find a few particular tools to be especially useful in your lab. A good hand drill or drill press is one of those tools. There seems to be no end of things electronical that need a few more holes drilled into them. Electronic enclosures and control panels need good, clean holes drilled into them. If you make your own PCBs at home, you will find that they need *lots* of holes drilled into them.

Like any power tool, an electric drill needs and deserves your respect. You'd be surprised how much unintentional damage you can cause with such a simple device. Observe these simple safety rules when operating a power drill:

- Keep your work area clean and well illuminated.
- Remove any watches or jewelry you might be wearing.
- If you have long hair, keep it pulled back and tied out of the way.
- Always clamp down any work to be drilled on a drill press.
- Always remove drill bits from the drill chuck after use.
- Discard dull or rusted drill bits that cannot be resharpened.
- Always wear safety glasses when operating power tools.

Hand drills come in all shapes, sizes, and price ranges. Some are powered by hand cranks and some are electric. Rechargeable batteries make for very portable drills that you can carry almost anywhere. See Figure 2-13. The only drawback is their limited charge time. That's why you should always have at least one spare battery charged up and ready to go. Electric drills with wires attached don't have this problem. Just don't trip over the cord!

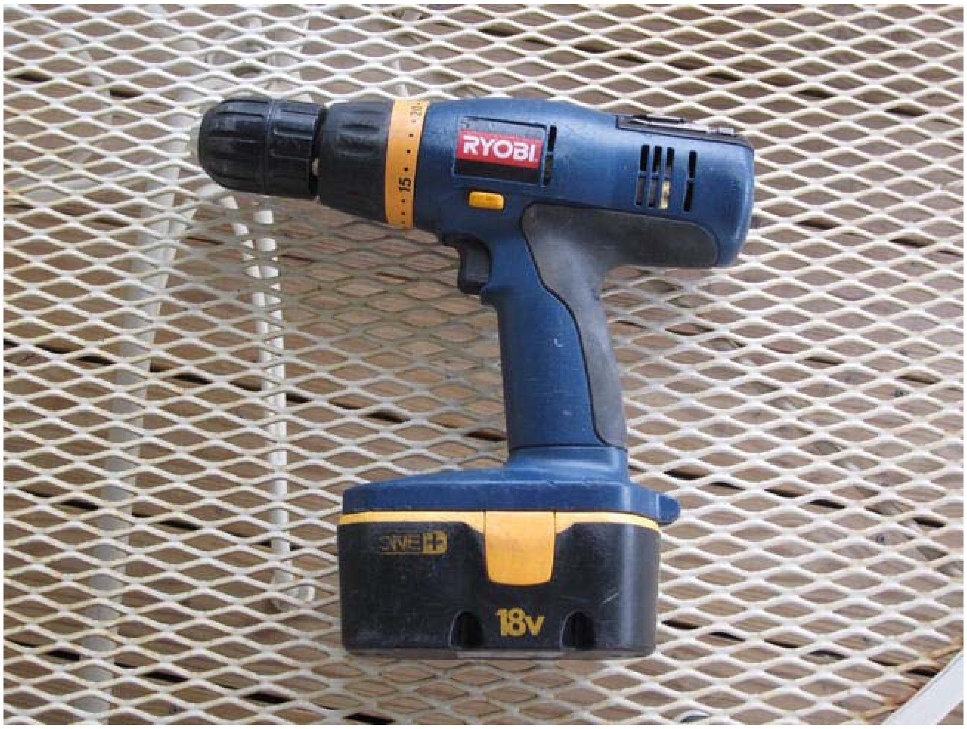

Figure 2-13. *A rechargeable hand drill. This model is both variable speed and reversible, which are two very desirable features in a hand drill.*

■ **Note** Power drills are also very handy when used with screwdriver bits. They can save you time and many blisters!

Drill bits likewise come in all sizes and shapes. You can buy sets with a variety of sizes, and these will usually come with a handy storage container that conveniently labels all the sizes for you. This only works, however, if you take the time to return the drill bits to their proper location once you've finished working with them.

A drill press has a large frame that facilitates drilling a precisely aligned hole, over and over again. You can make spindle-speed adjustments to drill presses (even small ones, such as the one pictured in Figure 2-14) either through an electronic speed control or by changing up the belts and pulleys that drive the spindle. There are even small frames for mounting handheld drills or rotary tools, such as the Dremel brand of tools. You can often, however, find actual, dedicated drill presses on sale for about the same price as the daintier Dremel tools.

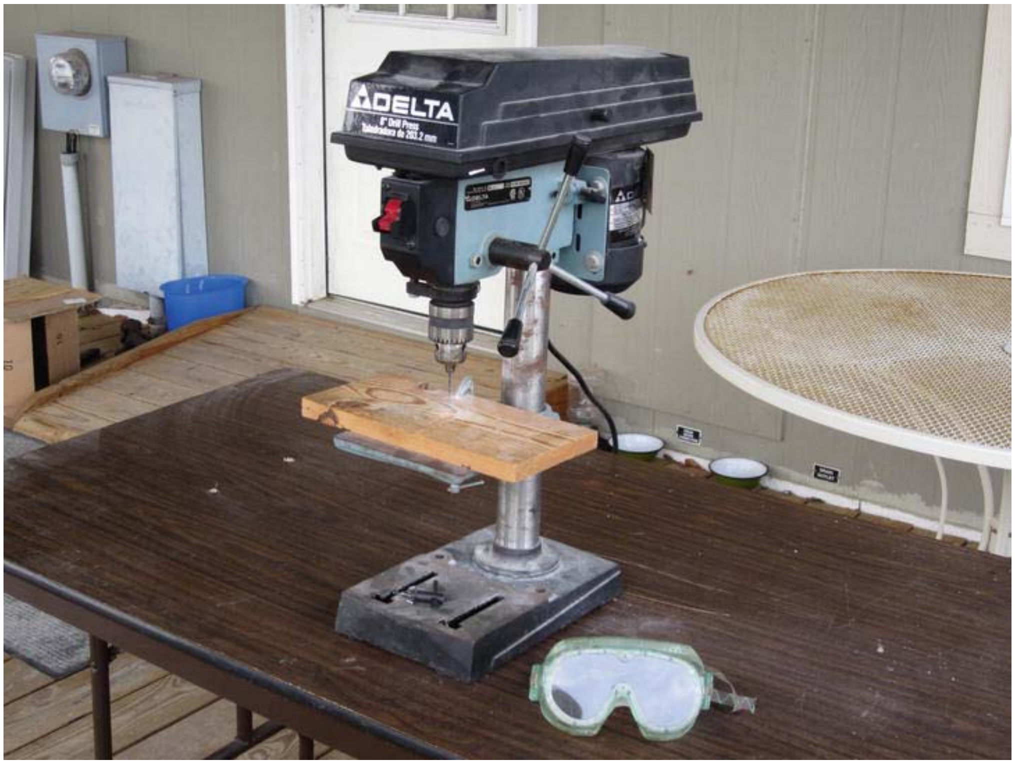

Figure 2-14. *A small tabletop drill press*

■ **Caution** When working with drills and presses, always keep your work area clear. Always wear safety glasses when using power tools. Try to keep the chuck key handy. Don't leave drill bits mounted in the chuck when not in use. Many chuck keys have built-in springs that prevent them from remaining in the chuck. Those that don't risk being accidentally left in the chuck when the drill its turned on, which can either fling the chuck across the workshop or make a big mess of your favorite hands.

Laser Cutters and 3D Printers

While not everyone can afford some of the more expensive and esoteric tools that are available, some of these tools are getting more popular than others; excellent examples are laser cutters and 3D printers. See Figures 2-15 and 2-16.

Laser cutters and engravers use a powerful carbon-dioxide laser to cut through wood and non-chlorine-based plastics. The laser beam is focused onto the work surface using mirrors that are mounted in an XY gantry, allowing the cutting area to be controlled by a computer. Typical installations require venting to the outside because of the various aromas produced in the process, as well as a water-cooling system for the laser tube itself.

There are also some newer laser cutters being built by hobbyists from scratch.

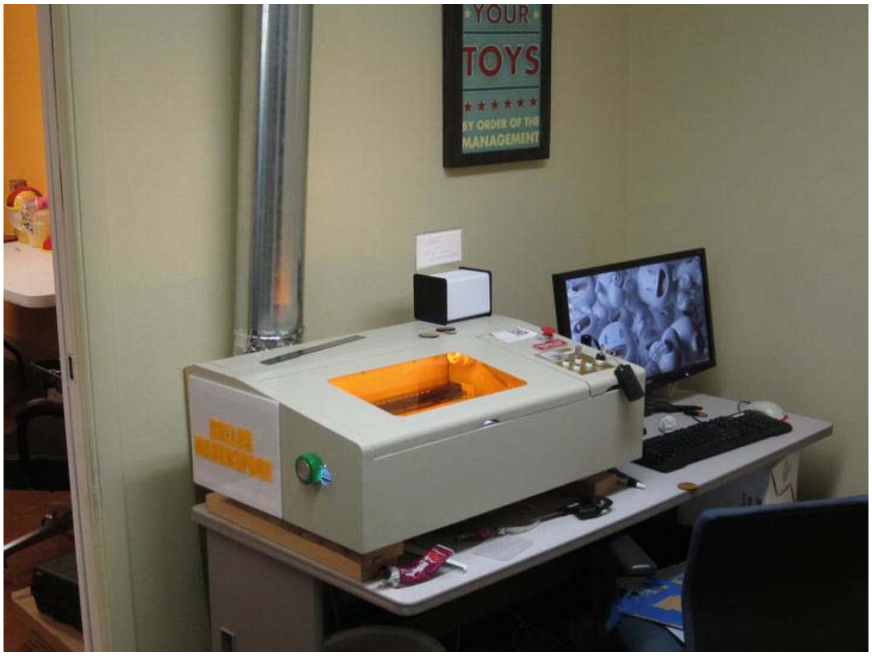

Figure 2-15. *A laser cutter. A powerful 40W carbon-dioxide laser is used to cut through wood and non-chlorine-based plastic, all under computer control. It can also be used as an engraver. This model is manufactured by Full Spectrum Engineering.*

3D printers seem to be all the rage these days. We're all pretty accustomed to the idea of printing on paper with a computer, although it's still a pretty recent innovation, as far as the history of printing goes. The idea of printing a solid object, on the other hand, is considered to be something of a game-changer.

The simplest 3D printers use a plastic extrusion and a motorized gantry to operate as a kind of a robotic glue gun. A plastic filament is fed into a heated fixture that is moved around a print bed, leaving a trail of cooling, gooey plastic. Either the extruder or the print bed is moved using motors under computer control to build up a layer at a time of the desired object.

Higher-technology 3D printers use this and other techniques to produce tangible, solid objects. Some use layers of fine powder that is glued together with a binder dispensed from a print head, similar to an inkjet printer. As each layer is printed, the object is lowered and another layer is deposited. This can even be done with sugar, as in the case of the CandyFab 4000, designed by the clever folks at Evil Mad Scientist Laboratories (`www.evilmadscientist.com/article.php/candyfab`). A stream of heated air is used to fuse a layer of sugar into a single solid structure, and an object is built up of successive layers.

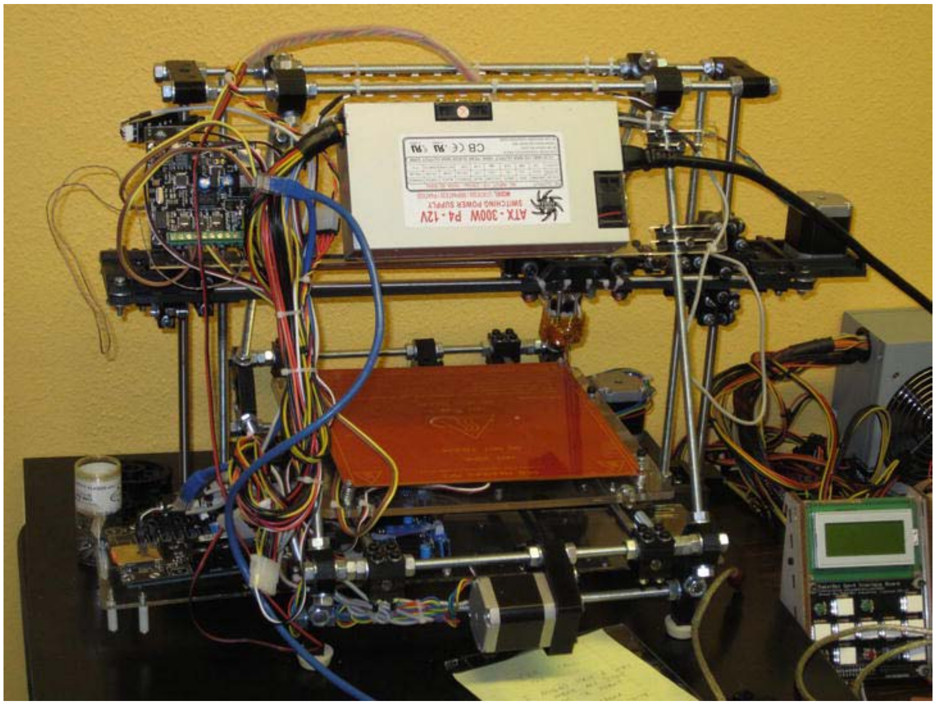

Figure 2-16. *A 3D printer at Dallas Makerspace. This machine uses a heated extruder to build 3D objects out of plastic.*

Basic Test and Measurement Equipment

When it is behaving itself, electricity is generally invisible. To "see" what it is doing, or not doing, you will need some specialized tools. These fall into the two general categories of *test* and *measurement* equipment.

The line between testing and measuring is sometimes blurry. One possible distinction might be that testing verifies proper operation within acceptable parameters, while measurement actually assigns values to those parameters. An example of testing is, Does that flashlight work? The answer is yes or no, arrived at by the simple expedient of turning it on and observing the results. An example of measurement is, How much current does that flashlight draw? The answer might be 0.78A, and this would be determined using the appropriate measurement device, which in this case could be an ammeter.

Inspection Equipment

Sometimes, however, measuring things isn't hard at all, especially when the things being measured are actual, tangible items that you can hold in your hand. What you will need are some basic tools, primarily to be able to *see* what it is that you are measuring, and secondarily to take accurate measurements of them, at least in regard to their physical dimension. This comes in very handy when trying to identify components during repair or equipment upgrades, when the original technical documentation is not at hand.

Even the youngest and finest of eyeballs will need help with the teeny-tiny components that are available nowadays. You should try to find at least a small magnifying glass to examine fine details. If your budget allows, you might want to consider a microscope of modest power to look at your handiwork. Careful shopping will result in finding some older microscopes at bargain prices. Expect to pay a premium price for the shiny, new stuff. See Figure 2-17.

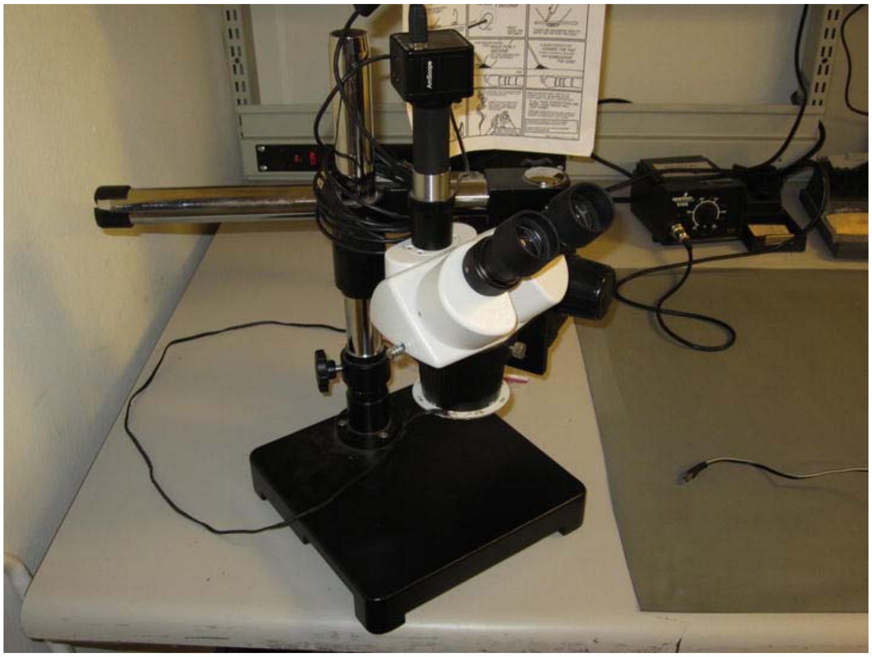

Figure 2-17. *A very nice stereo microscope. The power cord running to the white ring (partially hidden) supplies power to a ring of white LEDs that help illuminate the field of view. The base is very heavy to help stabilize the microscope.*

Even a humble magnifying lens can be very helpful in the lab, especially if it has a built-in light. See Figure 2-18.

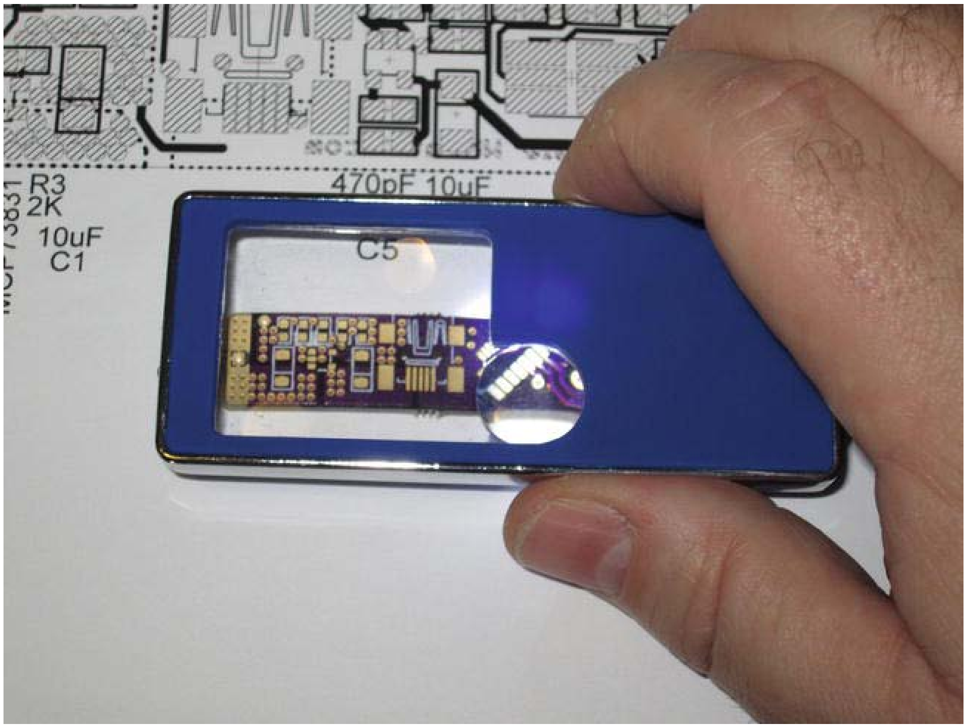

Figure 2-18. *A small, handheld magnifier with a built-in LED light is very handy for examining small features and details, such as are found on many PCBs.*

Physical Measurement Devices

Once you know what something looks like, at extreme close-ups, can you describe in absolute terms how big it is?

Take a look at Figure 2-19. Here we have, from the top to the bottom, a traditional tape measure, marked in both English and metric units (fractional inches and decimal centimeters), a digital caliper that can render measurements in either system at the touch of a button, and an architect's scale. While it may not be evident in the top-down view, the architect's scale is triangular in cross-section, resulting in not two but *six* different measurement options. This is mostly due to architects liking to draw things to scale, instead of full-sized.

The digital caliper is interesting in that it also provides a computer interface so that its readings can be pulled in by a computing device of some sort. These devices used to be quite expensive, but mass production has brought their prices down significantly.

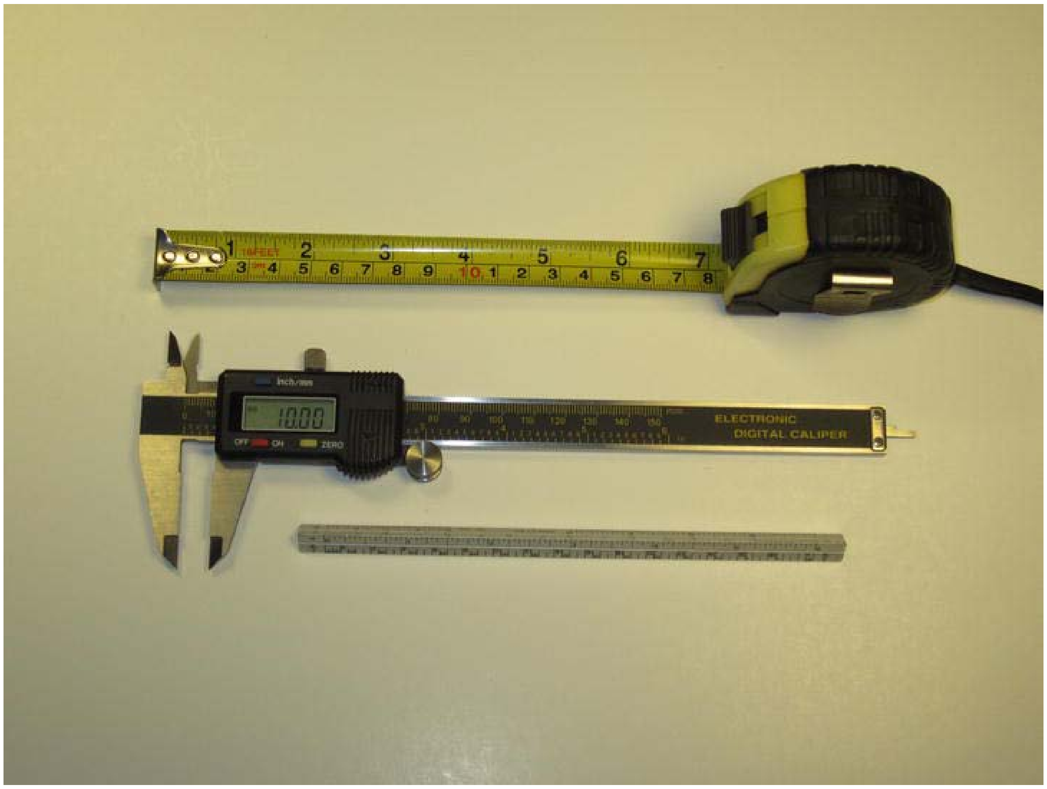

Figure 2-19. *From top to bottom, a tape measure, a set of digital calipers, and an architect's scale.*

Simple Electronic Measurement Tools

Now that you've got the tools to examine what these things look like and how big or small they are, let's move on to the more mysterious art of measuring their electrical properties.

Voltmeters

Voltage is a measure of electromotive force, and the standard unit of measure is the *volt*, named after Alessandro Volta, an Italian physicist. Voltage is sometimes compared to the pressure of water flowing through a garden hose, while amperage is compared to the amount of water flowing through the hose at any given time. It's not particularly useful to push this water analogy too far, but it's helpful in the beginning.

Voltage can be accurately measured with, you guessed it, a voltmeter. See Figure 2-20.

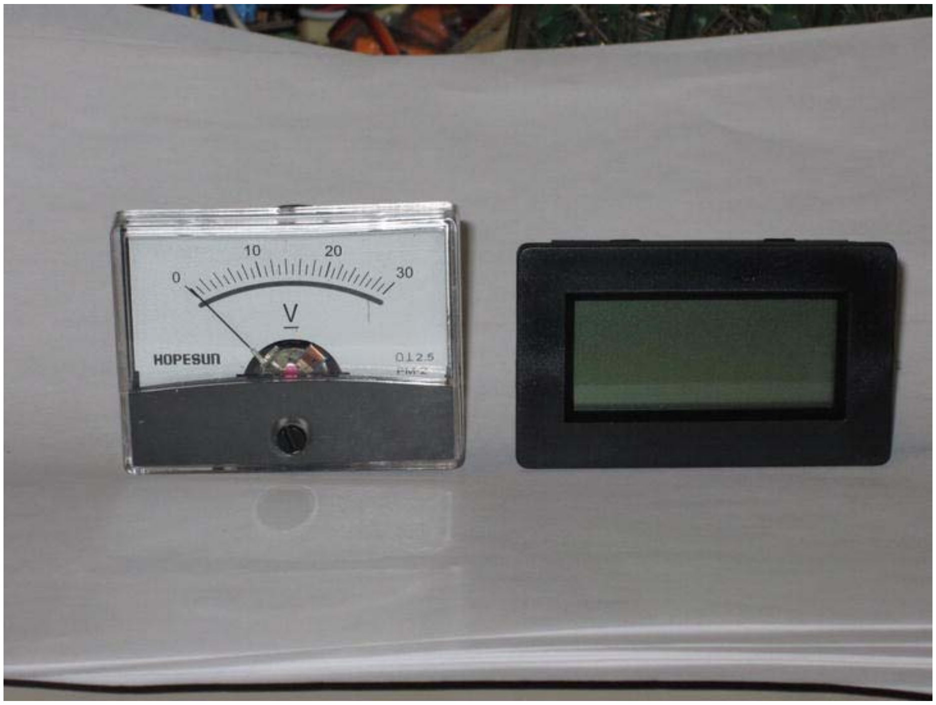

Figure 2-20. *A panel-mount analog voltmeter (left) and a digital version (right)*

A voltmeter measures the voltage, or *potential difference*, in a circuit across two points. Sometimes one of the points is a fixed reference point, such as the electrical *ground* mentioned in many circuits. This is often just a convenience for the purpose of measuring things, as the ground in an electrical circuit may or may not be connected to the ground under your feet.

A voltmeter will have two terminals, labeled positive (+) and negative (–). Looking at the analog voltmeter in Figure 2-20 (the one on the left), we see a needle that can swing from 0 on the left to 30 on the right. The prominent capital *V* on the dial indicates that the unit of measure is the volt. This indicates that you can use this meter to directly read from 0 to 30 volts, just by placing its terminals across a voltage. The meter itself is electromechanical in design, and creates a magnetic field that works against a spring and a fixed magnet. This action causes the needle to move to a point generally corresponding to the labeled voltage on the dial. These needles are not especially accurate, but can be corrected using the screwdriver adjustment at the bottom of the meter.

An analog panel meter such as this one does not require a separate power supply for proper operation. It draws all the power it needs from the signal being measured. This makes it unsuitable for measuring low-power circuits, where the current draw from the meter would unduly influence the reading.

The panel meter on the right is a digital circuit that measures the voltage and displays it on the built-in liquid crystal display (LCD). It requires a separate power supply for normal operation.

Ammeters

An *ammeter* looks and operates just like a voltmeter, except that it measures electrical current flow instead of voltage. The flow of electrical current is measured in units called *amperes* (A). The basic mechanism is almost identical to a voltmeter. To convert a current flow into a voltage, an electrical device known as a *shunt* is used. The shunt is simply a fixed-value resistor that produces a voltage across its terminals in relation to the amount of current that is flowing through it and its inherent resistance to the flow of current.

This relationship between voltage, current, and resistance is elegantly expressed in *Ohm's Law*. If you know two of the measurements in a circuit, you can calculate the third, based on their fixed and proportional relationship to each other. You will find this law to be exceedingly helpful to you when designing, repairing, and upgrading electronics.

To avoid using fractions, which seem to terrify so many people for no good reason, you can use Ohm's Law thusly:

- Voltage (volts) = current (amps) × resistance (ohms)
- Current (amps) = voltage (volts) ÷ resistance (ohms)
- Resistance (ohms) = voltage (volts) ÷ current (amps)

That's all there is to it, if you're just working with volts, amps, and ohms. Things can get weird, however, when you start to work with alternating current (AC), which is the kind of electricity you find in wall outlets and audio circuits.

Ohmmeters

An excellent application of Ohm's Law is found in building a resistance meter, also called an ohmmeter, because it measures resistance. Resistance is measured in units called ohms, often abbreviated with the capital Greek letter omega (Ω).

There are two ways to do measure resistance. One is to pass a known voltage though the unknown resistance and measure the current that is flowing in the circuit. The other is to pass a fixed current though an unknown resistance and measure the voltage across it. This second method is the one most commonly used.

You're most likely going to use a multimeter to measure resistance, as this is one of the more common functions of these handy devices. Mutlimeters are covered in the very next section.

Advanced Test and Measurement Equipment

It would seem that for simple circuits, having just a voltmeter, an ammeter, and an ohmmeter would be all you need. For the most part, this is correct. However, have you considered what would happen if you applied 100V to the 30V meter shown in Figure 2-20? What if you simply reversed the polarity of the signal by accident? Something bad for the meter, that's for sure. It might not actually catch on fire *right away*, but it would certainly damage the delicate, spring-loaded mechanism, possibly rendering it useless as a reliable measuring instrument.

The same can be said for an unprotected ammeter or ohmmeter. Luckily, the nice people that make good test equipment stay up late at night, thinking of the stupid things we might do with their products, and protect them accordingly. Let's look at some meters that measure multiple electrical characteristics, which strangely enough are called *multimeters*.

Multimeters

Multimeters will almost always allow you to measure various ranges of DC (direct current) volts, AC volts, current, and resistance. More advanced meters also measure things like frequency and duty cycle or even temperature. Some meters will even directly measure capacitors, inductors, and transistors using specialized connectors.

As with most tools, you generally get what you pay for with multimeters. The meter on the left in Figure 2-21 was purchased for under $3, including shipping from China. It arrived DOA (dead on arrival) and was not deemed important enough to return. It will make an interesting clock or fake lie detector someday. The unit on the right was purchased new for nearly $400 and has provided over a decade of useful service. The unit in the center is modeled after the unit on the right, it would seem, but can be bought new for between $10 and $20. The unit in the photo was bought wholesale for $4, because sometimes you get lucky. It's often handy to have a second meter in the lab—for example, when you need to measure the voltage *and* the current flowing in a circuit simultaneously, which you just can't do with a single meter.

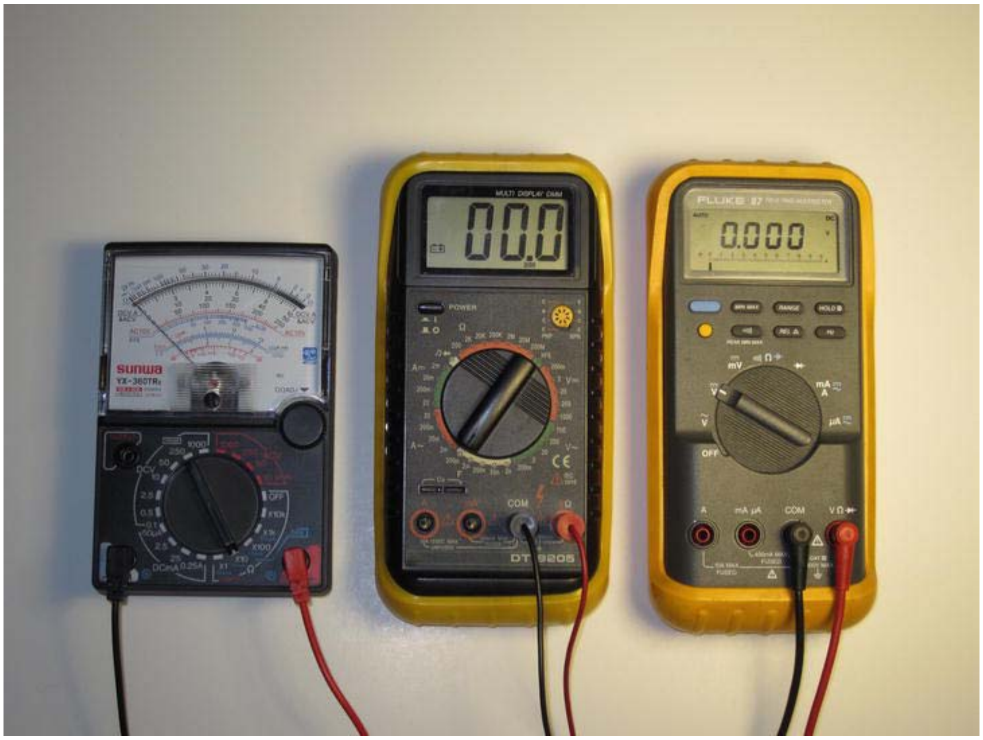

Figure 2-21. *Examples of some multimeters. The meter on the left features an analog display that is easy to read and acommodates several readout scales simultaneously. The remaining two meters have digital readouts and offer more accuracy. All the meters use a large rotary knob for selecting the proper function. The higher-cost unit on the right features auto-ranging, intellignetly deciding what range to use. The right two units also have high-impact boots to help protect them in the event of a fall.*

Your present budget will help determine what level of multimeter you might want to obtain. If you're wanting to sink less cash into a multimeter right now, consider looking for used equipment in good shape. As with any good tool, you can spend as much as you like on these. The sky really is the limit here.

Most all-in-one meters protect their inputs using fuses. You will have to disassemble your meter to replace these fuses, so think twice and measure once (to paraphrase the old carpenter's saying, "Measure twice and cut once").

■ **Tip** Be sure to set up your meter for the proper function and range *before* taking measurements.

Oscilloscopes

Once you've gone beyond measuring simple DC and AC voltages and move on to more complex signals and waveforms, you'll have to lay aside your trusty meter and fire up your favorite oscilloscope. This instrument is more complex than your traditional multimeter, and offers a different perspective on what's going on in your circuit.

While many modern oscilloscopes have many modes of operation, their basic function is to draw a picture of your signal on a small screen. In the olden days, this was done with a moving dot that was swept across a phosphorescent screen, leaving a residual, fading line of light. If this pattern was repeated over time, a visual representation of the signal could be seen on the screen. See Figure 2-22.

Modern oscilloscopes use a computer monitor or built-in screen display to draw the signal. This saves space, weight, and power requirements.

When purchasing an oscilloscope, the features you pay most for are the number of available channels and the bandwidth. The least expensive oscilloscopes will have a single channel. Having multiple channels allows you to "see" multiple signals simultaneously, which is often helpful in understanding what is going on in a circuit.

Bandwidth refers to the range of frequencies that can accurately be measured by the oscilloscope. Lower bandwidth scopes require much simpler electronics, and are thus cheaper to produce.

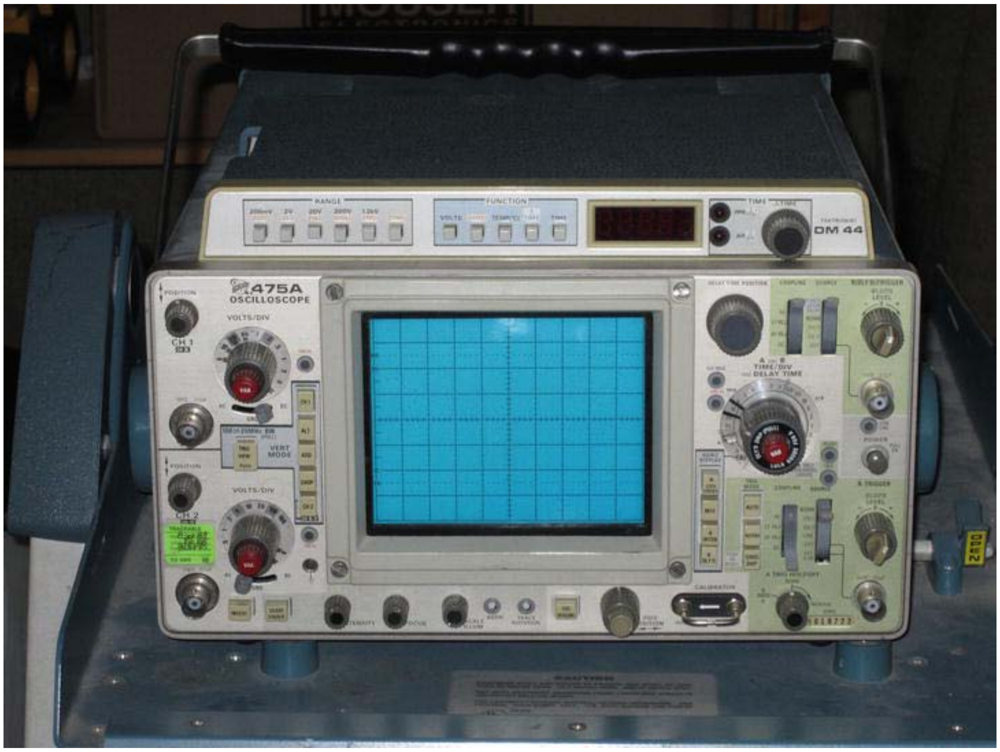

Figure 2-22. *A Tektronix 475A oscilloscope. This piece of vintage test equipment is considered by some to be the best analog oscilloscope ever made. It also has a digital multimeter built in along the top edge.*

The Golden Age

Once upon a time, not so long ago, a quality oscilloscope weighed 40 pounds and had a folding handle to make it, if not portable, at least *transportable*. It also had its own line of furniture to place it on, so that you could wheel it around the lab as you needed. Oh, and it also cost about three months' salary, and that's if you had a *good* job.

The good news? That same scope still works today. It can be repaired or calibrated to the exact same standards as when it was brand new, *plus* you can buy it on eBay for a few hundred dollars, sometimes less. That's because when these devices were designed back in the 1970s, they were made to be the best in the world, and cost was not a factor.

Modern Oscilloscopes

Today you can buy a quality oscilloscope for a few hundred to a few thousand dollars. These scopes not only display your signal on a large, colorful screen, but they can also analyze the signal and tell you all kinds of interesting things about it. For an example of such an oscilloscope, see Figure 2-23.

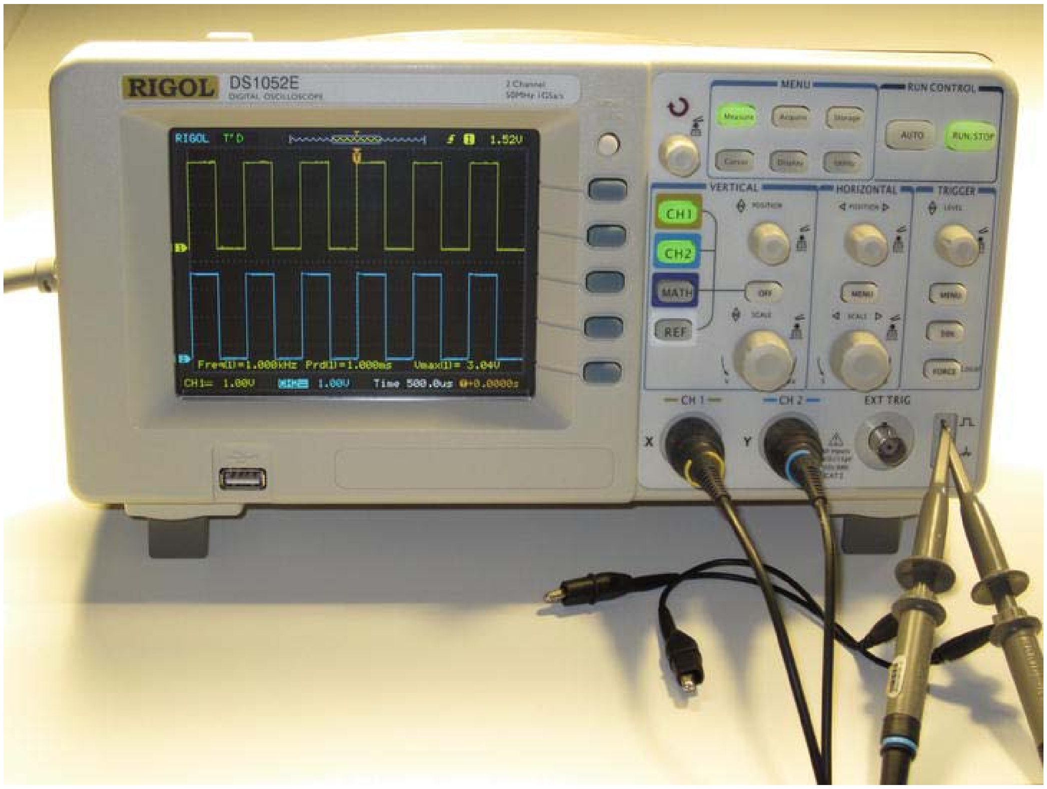
RIGOL DS1052E
DIGITAL OSCILLOSCOPE
2 Channel
50MHz 1GSa/s
MENU
RUN CONTROL
Measure
Acquire
Storage
Cursor
Display
Utility
AUTO
RUN/STOP
VERTICAL
HORIZONTAL
TRIGGER
CH1
CH2
MATH
REF
OFF
MENU
50%
FORCE
CH 1
CH 2
EXT TRIG
X
Y

Figure 2-23. *A modern digital storage oscilloscope. The RIGOL DS1052E is an affordable scope that will handle most troubleshooting and design challenges that you might encounter in your lab.*

This new breed of digital storage oscilloscope (DSO) can even connect to your PC to be remotely operated and allow you to capture screen shots that can then be pasted into books about how to build your own electronics lab. It's true. See Figure 2-24.

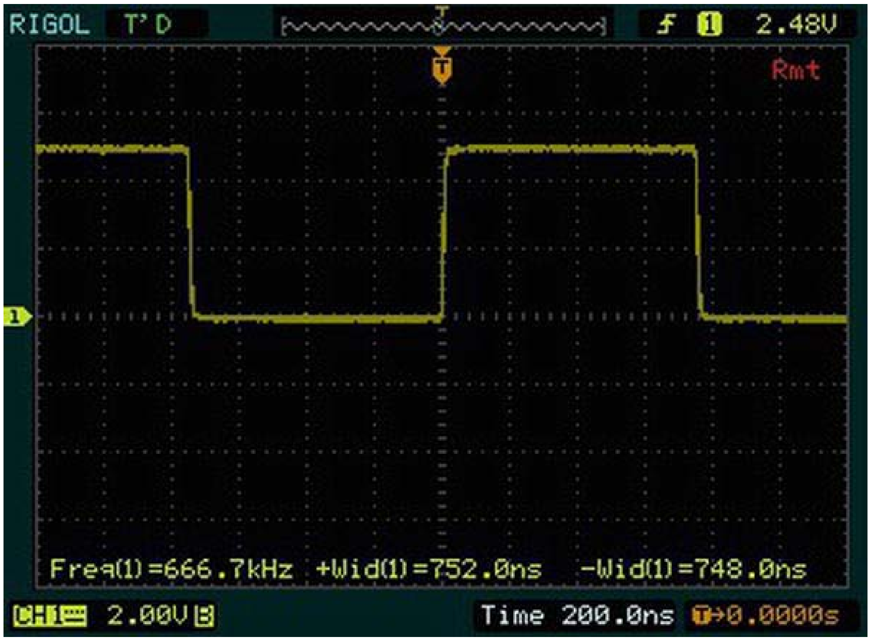

Figure 2-24. *A screen capture made with the RIGOL DS1052E oscilloscope.*

Tiny Oscilloscopes

The march of electronic miniaturization never stops, and so it is with oscilloscopes. If both your needs and your budget are modest enough, perhaps you can get by with something like the pocket-sized wonder shown in Figure 2-25.

While little scopes like this lack the accuracy and bandwidth of their larger relatives, they are definitely handy to have when you need to look at a signal. If you can keep your expectations in check when it comes to the user interface, you can get a lot of utility from such a small package.

Additionally, one of the benefits of a device such as this is that the hardware and internal software (yes, there's a little computer in there, somewhere) is all open source. This means that you have access to all the schematics and source code. And what does that mean? It means that if you have the skills and the right equipment, you can redesign the entire apparatus to fit your own requirements, assuming that you understand the basic limitations of the electronics. The sky is the limit. You are empowered.

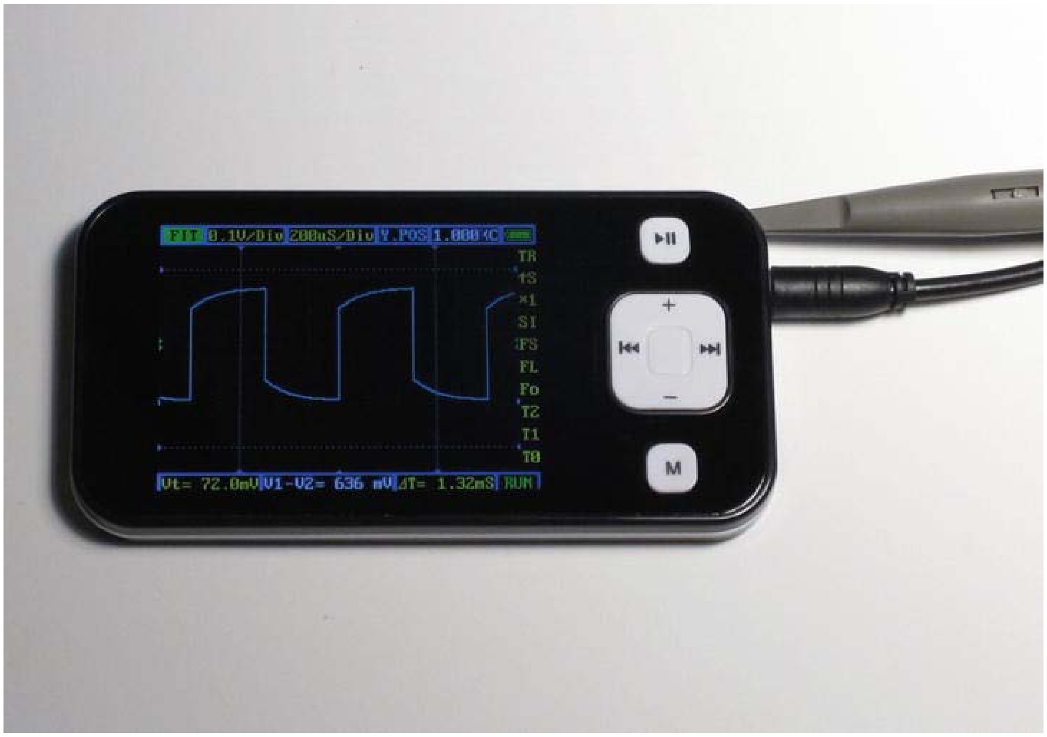
FIT 0.1V/Div 200uS/Div Y.POS 1.000 C
TR
↑S
×1
SI
FS
FL
Fo
T2
T1
T0
Vt= 72.0mV V1-V2= 636 mV ΔT= 1.32mS RUN
+
−
M

Figure 2-25. *A handheld oscilloscope, which can also fit in your pocket. This is the DSO201, which has been discontinued by the manufacturer, but can still be snapped up for a song from various online distributors, as well as eBay. This unit was purchased new for $60 and features a single channel with a bandwidth of 1 MHz.*

Power Supplies

One of the most important pieces of test equipment in your lab is a reliable power supply. Building a good bench power supply is an excellent first or second electronics project (see Chapter 5 for more on adding power supply to your lab and Appendix A for more on building your own). This is because the basic design is very simple and easy to build, yet allows you to add any number of additional features later.

Once you get tired of throwing away spent 9V batteries and grow weary of waiting for the red light on the battery charger to go out, it's time to invest (either time or money or both) in a decent bench power supply.

It doesn't have to be fancy. In fact, the simpler the better at first. What you're aiming for primarily at this point is reliability. You want to know that when you turn on your 5V supply, it's putting out 5 volts, not 2 or 11, no matter what you do to it. If it's overloaded, it needs to deal with it and not be catching on fire. If it's underloaded, or not loaded at all, it still needs to maintain a nice, clean output.

Want a cheap or free power supply? Do you have any dead or discarded electrical gizmos that came with a charger? These make excellent starter supplies for the new lab. Do you have any extra USB chargers? These are even better, because they supply up to 2.5W (and sometimes more) of *regulated* 5V. We'll use some of these in experiments in Chapter 3.

Organizing Your Tools

So now you have an idea about some of the tools you want (or already have) for your lab. Where are they all going to live? This is a big question, and deserves a big answer.

One way to look at this book is that the entire latter half is dedicated to answering this question. There are, obviously, many possible arrangements that will work for most people. But we're talking about *your* lab, with *your* tools, arranged and ordered to meet, specifically, *your* needs.

As you progress through Chapter 3 (the very next chapter in this book), you'll get a chance to play with both your tools and some interesting electronic components, all in the interest of getting you familiar with some basic concepts. These concepts and ideas are not new, but they may be new to you. The good news is that time invested today and in the very near future should pay off big dividends to you later in life. Who knows what you'll devise in that electronics lab of yours?

Summary

Starting with your brain and working outward, this chapter has covered some of the tools and devices you might use in your electronics lab. These include basic hand tools, computers, tools for soldering, some common power tools you might find handy, as well as some fancy-schmancy tools you might only find in a hackerspace, school, or electronics clubs, such as laser cutters and 3D printers.

You also learned about some of the basic test and measurement tools you'll be needing when working on electronic circuits in your lab. These include magnifying glasses, measuring tapes, and meters (oh my!), as well as oscilloscopes and power supplies.

Additionally, *lots* of advice on safety was presented. These topics were learned the hard way. Do try to learn from the mistakes of others, won't you?

Just a hint was given about how you might want to organize your tools within your lab. Take care of your tools and they will most certainly take care of you. The details about possible organizational scenarios is left for the second half of the book, where you will learn about setting up labs big and small, as befits your needs, budget, and interests.

Chapter 3 ties in much that you have learned here in Chapter 2 by introducing you to not only some of the basic electronics components and their uses and characteristics, but also how to use a good number of the tools described in this chapter. Hopefully this chapter has whetted your appetite and readied you to dive into some circuits and see what kind of mischief you can wreak on the world.

CHAPTER 3

Components

You're going to need some components to play with, as well as the knowledge to use them effectively. This chapter will introduce you to some of the bits and pieces that make up modern electric and electronic circuits. It will also teach you a little bit about how to identify components from their appearance and markings, when available.

Once you've got an idea about what these parts do in a circuit, you'll learn a little more about how to measure their electrical properties and put them to use. You'll also be shown what *not* to do, in some select examples.

There are an enormous number of both manufacturers and vendors of electronic components these days. Trying to list even a few of them here in print doesn't really make much sense, as a good percentage of that information will be incorrect or obsolete as soon as the ink dries on the page.

Having a variety of basic components readily available is convenient, until your collection grows beyond your ability to house or organize it. This happens far too often, in reality. Most common electronic parts can be obtained within a day or two from local sources, unless you live in a really secluded part of the planet. Also, most of the more common parts will typically cost only a few pennies apiece, so there aren't a lot of reasons to stockpile random parts. Keep a few of each on hand for experimentation, but hold off on pointless hoarding . . . if you can.

Conductors and Interconnects

In any electrical or electronic circuit, there are just as many connections between components as there are components. These include wires, cables, connectors, plugs, sockets, and all those other pieces that help the little electrons zip around the circuit.

First, let's take a look at wires and interconnects in a very abstract way. We can think of wires as connections between other components, without having to get into a lot of detail about the wire itself. We just want a way to indicate that there is an electrical connection among all the bits, and using an electrical schematic diagram is an easy and effective way to do this.

Electrical Schematics

A *schematic* diagram is any kind of stylized drawing that represents a simplified view of something. Usually some sorts of symbols are used to represent the parts that make up the whole, as well as the connections or relationships that exist between them and possibly the outside world. You can have schematics for plumbing, hiking trails, or tribbles, a well as electrical and electronic circuits. Let's talk about electronic schematics for a little bit.

The electrical symbol for a wire is almost always just a plain, solid line. If there's something really special about the wire that needs to be used, it is usually indicated in a note somewhere on the schematic. This can be used to indicate what color the wire needs to be, what size is appropriate, whether there are any special requirements for the insulation, and so forth. Most of the time it's just a plain, solid line. For an example of just how simple it can be to draw a wire in a schematic, see Figure 3-1.

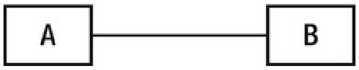

Figure 3-1. *A very simple electrical schematic diagram, showing how to get from point A to point B with a single wire. Drawing wires in schematics is easy!*

If a wire starts in just one place and ends in just one place, everything stays nice and easy. If multiple wires need to converge on a single point, or one signal wire needs to split up to go to more than one destination, it gets just a little more complicated, but not too much. See Figure 3-2.

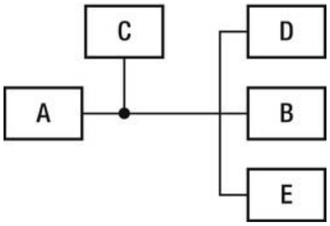

Figure 3-2. *Wires that are connected to each other usually have a dot at their intersection. If there is no dot, there is no connection between overlapping wires.*

In the figure, points A, B, and C are connected together. Points D and E are connected together, but are not connected to points A, B or C, even though the wires look like they are crossing. In a professional schematic, a careful draftsperson would route the wire from point D to point E on the right side of the components to avoid any possible confusion.

Some folks like to draw a little bump, or bridge, when two unconnected wires cross in a schematic, just to make sure that everyone knows that the wires aren't connected. This gets to be a bit tedious once you've got more than two or three components in your schematic, and it can look a bit unsightly. Feel free, however, to draw your schematics however you see fit. Just remember that others might not share your particular drafting sensibilities.

Wire and Cables

Chapter 2 took a quick look at wire and how it is made to help you pick out the best wire-cutting and wire-stripping tools. Now it is time to consider wire as an electrical component within your circuits. Knowing what a wire can and cannot do will help you pick the right kind of wire to use. Yes, wire has limits! This is important both for building new circuits from scratch as well as repairing, replacing, or upgrading wiring in existing circuits.

Wire can be used for all sorts of artsy and craftsy purposes. Copper wire is the most common wire used in electronics. Sometimes the copper conductor is *tinned* with another conductive metal with a low melting point to make it easier to solder. When you cut this wire, you can see the copper on the inside.

Pure copper wire is getting very expensive these days. Unscrupulous people try to pass off copper alloy with inferior electrical properties as the real deal. A quick way to test is with a magnet. Copper is not attracted to magnets.

As mentioned in Chapter 2, wire comes in two major categories: solid core and stranded. Solid core wire is just a single piece of copper wire, while stranded wire is made up of many smaller strands of copper wire, usually twisted together. Stranded wire is generally more flexible than solid core wire.

Copper wire is measured using a gauge system. Different systems exist around the world. A common system used in the United States and Canada is the American Wire Gauge (AWG) standard. It uses a series of single numbers to denote the relative size of the wire, with larger numbers representing smaller-diameter wires. These numbers go all the way down to zero (0 gauge), but they don't stop there. Larger wires than 0 gauge are called 00 gauge, double-ought, or sometimes 2/0. There are also three- and four-digit versions, but after that, those giant wires are specified with their actual cross-sectional diameter.

Using Wire: The Very Short Version

Fatter wires carry more current with less heat being generated due to resistance. The wire's insulation is rated for a maximum voltage and temperature. If this temperature is exceeded, the insulation will soften, melt, or catch on fire. Avoid this by using the right sized wire for the job. Exceeding the maximum voltage can cause more problems if the current arcs through the insulation. This can cause a short circuit and may even cause a fire. It is also likely to give off noxious fumes. Choose wisely!

For low voltages (under 36V) and small current loads (under 1A), you can use whatever you want. It doesn't matter. Pick something pretty. A thing of beauty is a joy forever.

For higher-voltage applications, make sure the wire's insulation is rated *far in excess* of what you anticipate. The rating should be visibly printed on the insulation itself. See Figure 3-3.

Figure 3-3. *The wire's insulation only works up to a point. That point is usually printed on the insulation itself, but you may need to put on your good reading glasses to see it. This wire states, "LL33911 CSA TEW 105°C 600V FTT 14AWG," which translates to the manufacturer's part number, Canadian Standards Association certified thermoplastic equipment wire, maximum temperature of 105°C, 600V insulation dialectric strength, 14 gauge.*

For higher-current applications, you really need to use the highest-gauge wire that is necessary to carry the required current. Hot wires cause fires.

Using Wire: The Longer Version

If you will spend some time learning about the properties of wire, you will be able to make *much* better decisions regarding its use and application. Assuming that the overall shape of the cross section of the wire is round, the gauge of the wire determines how much resistance to the flow of electrical current the wire will exhibit over a given distance. This resistance, in turn, helps determine the amount of electrical current that the wire can safely carry.

When electric current meets resistance, the electrical power is converted into some other form of energy. Usually this energy is given off as heat, or thermal energy. It can also produce magnetic fields, sound, light, and radio waves.

A good example of this is demonstrated by the heater coils on an electric stovetop. When the stove's switch is turned on, current flows through the coil. Most of the current is converted directly into heat. Electric heaters are very efficient this way. A small amount of the electrical power is converted into light. You can see this as a red glow coming from the heater coil when it is hot. The coil also creates a magnetic field.

If you know the gauge of a wire, you can look up the wire's characteristic resistance. These numbers will be close but perhaps not exact, as there will always be small variations in wire diameter, conductor purity, and other factors. Generally, however, the numbers are pretty close.

Let's take another look at the AWG table (see Table 3-1), and look at a few simple examples to help understand what these wires can do for us. (A more complete listing is available at `http://en.wikipedia.org/wiki/American_wire_gauge`, from which this partial list was obtained.)

Table 3-1. *The AWG*

AWG	Diameter Area			Copper Resistance	
	Inch mm		mm^2	Ω/km mΩ/m	Ω/1000 ft mΩ/ft
0000 (4/0) 0.46	00 11.6	84	107	0.1608	0.04901
000 (3/0)	0.4096	10.404	85.0 0.20	28	0.06180
00 (2/0)	0.3648	9.266	67.4 0.25	57	0.07793
0 (1/0)	0.3249	8.252	53.5 0.32	24	0.09827
1 0.28	93	7.348	42.4 0.40	66	0.1239
2 0.25	76	6.544	33.6 0.51	27	0.1563
3 0.22	94	5.827	26.7 0.64	65	0.1970
4 0.20	43	5.189	21.2 0.81	52	0.2485
5 0.18	19	4.621	16.8 1.02	8	0.3133
6 0.16	20	4.115	13.3 1.29	6	0.3951
7 0.14	43	3.665	10.5 1.63	4	0.4982

8 0.12	85	3.264	8.37 2.	06 1	0.6282
10 0.10	19	2.588	5.26 3.	27 7	0.9989
12 0.08	08	2.053	3.31 5.	21 1	1.588
14 0.06	41	1.628	2.08 8.	28 6	2.525
16 0.05	08	1.291	1.31 13	.1 7	4.016
18 0.04	03	1.024	0.823 20	.9 5	6.385
20 0.03	20	0.812	0.518 33	.3 1	10.15
22 0.02	53	0.644	0.326 52	.9 6	16.14
24 0.02	01	0.511	0.205 84	.2 2	25.67
26 0.01	59	0.405	0.129	133.9	40.81
28	0.0126 0	.32 1 0	.08 10	212.9	64.90
30	0.0100 0	.25 5 0	.05 09	338.6	103.2

Conductivity

The first thing you need to know is that copper wire really, really wants to conduct electricity. It wants to conduct electricity more than it wants to *resist* the flow of electrical current. Even so, copper is not a perfect conductor, and it has a certain amount of inherent resistance. This resistance causes the energy carried by the electrical current to be sidetracked, or converted into heat and other forms of energy. The higher the resistance in the wire, the more heat will be produced.

Normally this heat is a very small thing and is easily dissipated into the surrounding air. You usually can't even tell by touching that a wire is a fraction of a degree warmer than its surroundings—but this is only true when the amount of energy is relatively small, like in our rewired flashlight example. When you start to push a *lot* of power through a wire, the heat buildup can outpace the cooling-off provided by the surrounding air, and the temperature of the wire increases.

Again, this is no big deal as long as that temperature stays within a comfortable range. But what happens when the temperature gets really high?

First, the insulation will get soft, melt, or even catch on fire. After the insulation, if any, has burned off, the wire will glow red hot and eventually melt. This is known as *catastrophic wire failure*, which can result in what electrical engineers refer to as an unauthorized thermal event, which we call a fire. Yikes!

But look around you. You're probably sitting right next to some wires right now, very few of which (hopefully) are on fire, even though they are carrying variously large loads. How do they keep from burning up? The answer is in the proper selection of the right *size* of wire for the job. Bigger wires have less resistance to the flow of electricity, so they are able to carry more power with less heat being produced.

Now, you don't want to have to wire up all of your electronic circuits using jumper cables from your car's trunk, as that gets to be cumbersome after a while. Let's look at some good guidelines for what size wire to use for your projects.

In Chapter 2 you learned about (or were reminded of) Ohm's Law, which illustrates the relationships between voltage, current, and resistance in any electrical circuit. In summary, $V = I \times R$, where V is volts, I is current in amps, and R is resistance in ohms. It's not just a good idea, it's the law!

So how much electricity will we lose going through a foot or so of 16-gauge wire? If we know that one foot of 16-gauge wire has a resistance of 4mΩ (and we know this from looking it up in the chart), and we build a circuit that pushes 1A of current through this wire, we can figure out a couple of interesting things using Ohm's Law and a calculator.

Applying Ohm's Law

If you know two of the three variables used in Ohm's Law, you can calculate the third variable. Since we know R and I, we can calculate V. Since $V = I \times R$, we can substitute our values of 1A and 4mΩ (0.004Ω) into the equation, and get 0.004V, or 4 millivolts.

That's not very many volts at all! In this case, that's a good thing. The voltage we just calculated is the *voltage drop* across the length of the wire. Those millivolts never make it from one end of the wire to the other—they're lost. Poor millivolts, they never had a chance.

OK, let's not let their heroic sacrifice be in vain. If we know volts, and we know amps, we can also calculate power, which is the rate at which energy is converted to work. The unit of measure for electrical power is *watts*, and is also the same as Joules per second (J/s), but that's another story. We usually use the capital letter *W* to abbreviate wattage, but we use the capital letter *P* to represent the *value* in equations.

Calculating power is simple if you know the volts and the amps. Just multiply them together: $P = V \times I$. Easy.

So in our example so far, we have 4mV, 1A, and 4mΩ. How much power is being dissipated (wasted) in this circuit? $P = 0.004V \times 1A = 0.004W$, or 4 milliwatts, which is just four-thousandths of a watt. That's just about enough power to light up a dim LED, and not much else. So you're not going to be starting any fires with these numbers. You most likely would not even be able to measure the rise in temperature that is occurring in the wire (oh yes, it's occurring, but it's very, very small).

Let's put 10A through a foot of 30-gauge wire. On second thought, let's not, and just do the math instead. The voltage drop will be $V = I \times R$, or $10A \times 0.1032\Omega = 1.032V$. The power dissipated would be $P = V \times I$, or $1.032V \times 10A = 10.32W$. That wire would be a crispy critter in no time. Ten watts doesn't sound like much when you're talking about a home stereo or a lightbulb or a hair dryer, but that skinny little wire has almost no surface area with which to dissipate the generated heat. This means that the temperature goes up, quickly. Bye-bye, little piece of wire.

As a side note, you just learned how most current-limiting fuses operate. The conductor within the fuse literally *fuses*, or evaporates, when the internally generated heat produced from an "overcurrent" situation rises beyond its melting point. Poof, the circuit is opened, and property and lives are saved.

■ **Warning** Never circumvent the protection afforded by a fuse.

Ideally what you want is the smallest wire that will carry the required amount of current with an acceptable rise in temperature. Remember that any rise in temperature is in addition to the *ambient* temperature. If you're sitting in a room that is "room temperature," which is assumed to be 25°C or 77°F (the same thing, if you want to check it), and you're expecting a 50°C rise in temperature due to waste heat, your wire is going to be 75°C, which is far too hot to touch. Remember that 100°C is the boiling point of (pure) water (at sea level). Hot!

One more thing that complicates the thermal calculations is that copper, as a conductor, has a *positive thermal coefficient*, meaning that the characteristic resistance of a piece of wire (which we can look up in the AWG table) goes *up* by a known amount as the temperature increases. You need to consider this when you build circuits that are *supposed* to get hot, such as electric heaters. In most other cases, it's best to keep the heat to a minimum, as it represents wasted power, plain and simple.

Testing Your Wires

Can something as simple as a wire go bad? Yes, it most certainly can. How can you tell? Usually, it's because something doesn't work the way you think it should, or *worse*, it works sometimes and not others. A great deal of time, the problem is right at the end of the wire, where it is trying to make a connection with something else. However, it's possible for a wire to develop a problem right in the middle. This can be caused by a tiny nick in the insulation that lets in nasty elements such as rain and dirt, allowing the inner conductor to corrode. This type of problem takes time to make itself apparent, but it can certainly happen, especially in equipment that is exposed to the outside world for any length of time.

A quick way to test a wire is with a piece of test equipment called a *continuity tester*. This simple device usually is battery powered and has two probes. When you connect the probes together, the tester makes a noise, lights up a light, or maybe does both. Make sure your continuity tester's batteries are fresh and that the probes, if they are removable, are firmly seated in their sockets. Even test equipment needs testing from time to time.

■ **Note** A continuity tester is one of the easiest pieces of test equipment to cobble together yourself. See Appendix A for more information about building your own tools.

As mentioned previously, the problem is almost always at the very end of the wire. This is a good thing to know, so you can start looking in the most likely places first, saving valuable troubleshooting time.

A much better way to test a wire is with an ohmmeter. Using an ohmmeter, you can determine the *exact* amount of resistance in a wire, which under ideal circumstances should be very low. One ohm or less is typical for most normal-sized wires that you are likely to encounter. Super-long wires can have higher resistance, and that's where you can consult the handy AWG table to find out what the resistance is *supposed* to be.

When testing wires, especially wires with attached connectors, it's a good idea to wiggle the wire about vigorously, as this will reveal some intermittent connections for you.

Storing Your Wire

If you're buying new wire, it will probably come on a reel. It's often handy to keep common sizes and colors on your bench, if you have room. A paper towel–holder makes an excellent wire dispenser rack, although you can also spend more time or money buying a prebuilt one or designing and fabricating your own.

A custom red-and-black wire dispenser is shown in Figure 3-4. You might use a fixture like this if many pairs of identical-length wires are needed for a project. Not having to chase the wire spools all over the bench really saves a lot of time. This structure was built using a MicroRax miniature slotted aluminum extrusion (*http://www.microrax.com*), which is perhaps a bit of overkill, but was a lot of fun to plan and build. A coat hanger, artfully bent, would have also done the trick.

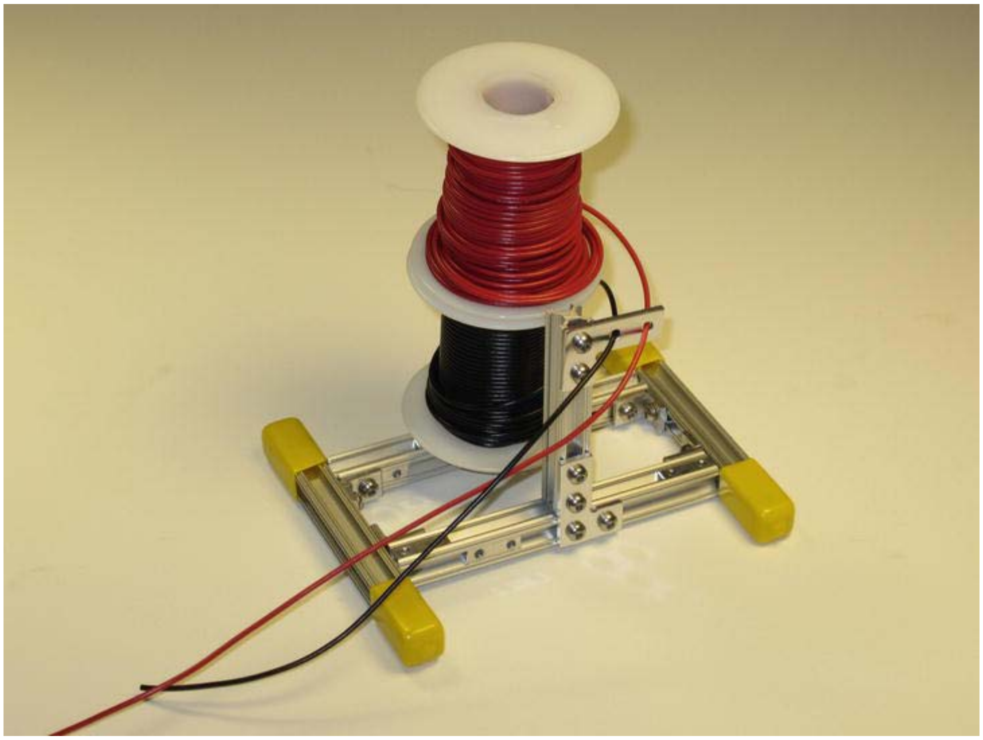

Figure 3-4. *A custom wire-dispensing frame made from modular aluminum extrusion. This kind of fixture makes it easier to cut both a red and a black wire to almost exactly the same length, which happens a lot when you use red and black wire for positive and negative power connections, respectively.*

Connectors

Connectors make life easier for the casual wiring enthusiast. The hardcore types solder everything together, but that is not always practical or even the right thing to do in every circumstance.

Keep in mind that every connector, just like every wire, has some sort of electrical limits. These limits must be respected.

Clip Leads

Also called *alligator clips* or *test leads*, you can buy ready-made wires or build your own from individual connectors. The flexible rubber or plastic boot extends the insulation of the wire. A spring-loaded jaw, often with serrated teeth, is used to clamp down on the wire, lead, or electrical terminal desired (see Figure 3-5).

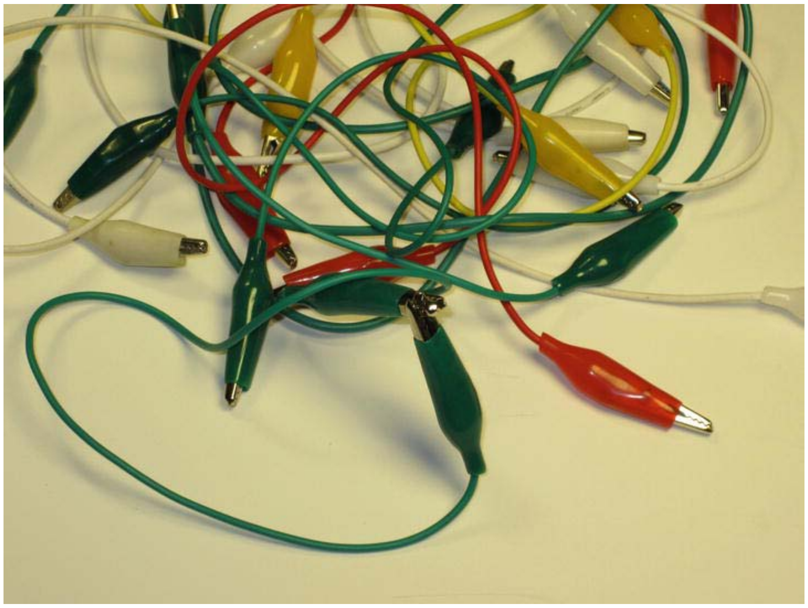

Figure 3-5. *Clip leads come in all sizes, gauges, colors, and terminal types. They are handy for making quick electrical connections.*

Clip leads are quite handy to have around your bench, if you've got the room for them. They tend to get tangled up, so some people hang them from a specially slotted rack, making them easier to pluck at a moment's notice.

Keep in mind that clip leads are generally for *temporary* electrical connections, such as when you're testing something or trying out a prototype circuit. They don't make very good long-term solutions to your electrical connection needs for a variety of reasons.

The first problem is that even when the spring-loaded jaws make a solid, low-impedance connection when first attached, this electrical connection will deteriorate over time as oxidation inevitably forms on the contacts.

Another problem is that many mass-produced clip leads are inexpertly smashed onto the ends of wires, without taking advantage of the alligator clips' built-in features, such as strain reliefs and soldering points. This leads to additional resistance in your circuit at both ends of the clip lead, where you would probably never think to look. The next time you buy some brand-new clip leads, hook them up to your ohmmeter and see just how conductive they really are. A good practice is to carefully remove the insulating boot and make a proper electrical connection yourself, so that you've got one less thing to worry about.

Solderless Breadboards

A very versatile prototyping platform is the solderless breadboard. Once upon a time, people would tack together simple electrical circuits using wires, small nails, and a hammer, utilizing a wooden plank or small board as the substrate. The kitchen breadboard was often just right in size and weight and would mysteriously disappear into the workshop. You can still cobble together simple circuits using this technique.

The solderless breadboard is a miracle of modern manufacturing techniques. A body made of precision-molded plastic has an array of small openings built in, and within each opening is a row of small spring clips that are just the right size to grasp and hold a small-gauge (22–26 AWG) wire. Adjacent points within the array are electrically connected together, which lets you easily connect small components into a circuit using your bare hands, with no soldering required. About all you would normally need would be some solid-core wire (preferably), some wire cutters, and some wire strippers. You can even buy precut jumper wires specifically made for use with solderless breadboards. They're terrifically handy! See Figure 3-6.

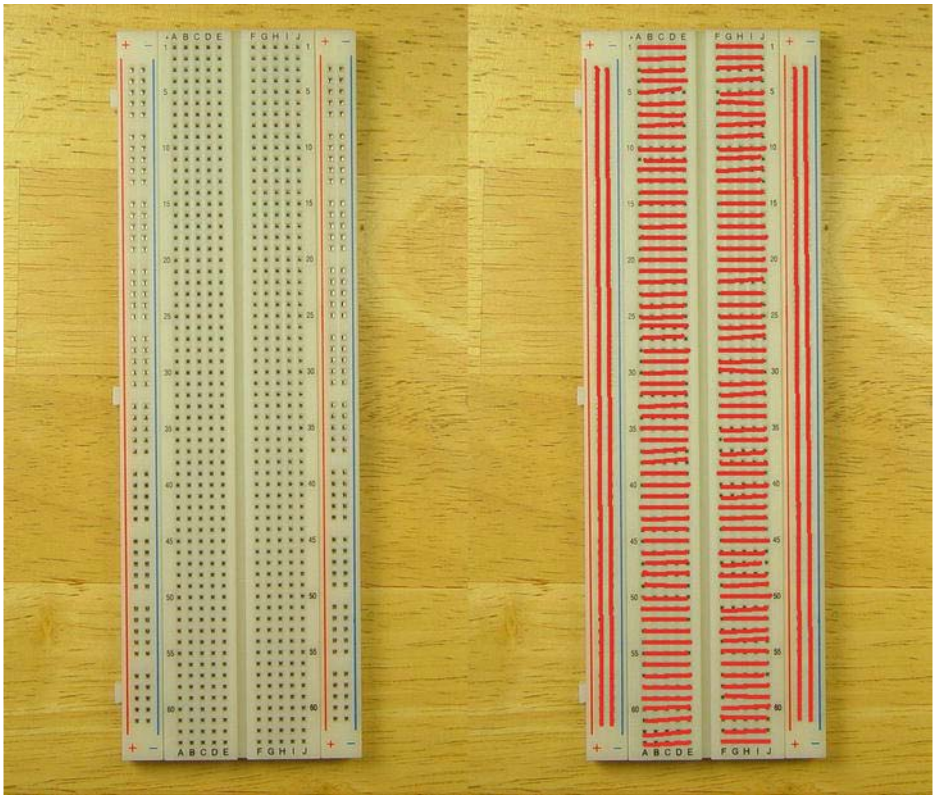

Figure 3-6. *On the left is a solderless breadboard with 830 tie points. Each hole has a spring clip inside that can hold a small wire or component lead. On the right, the underlying electrical connections between the tie points are illustrated.*

The spacing between adjacent tie points is exactly 0.1", which corresponds to many common “footprints” for electronic components. You can easily build simple circuits in just a few minutes. You can even build fairly complex circuits using integrated circuits and other components. Figure 3-7 shows a complete Arduino-compatible microcomputer that you can build with your bare hands and then program using your PC and the free Arduino software.

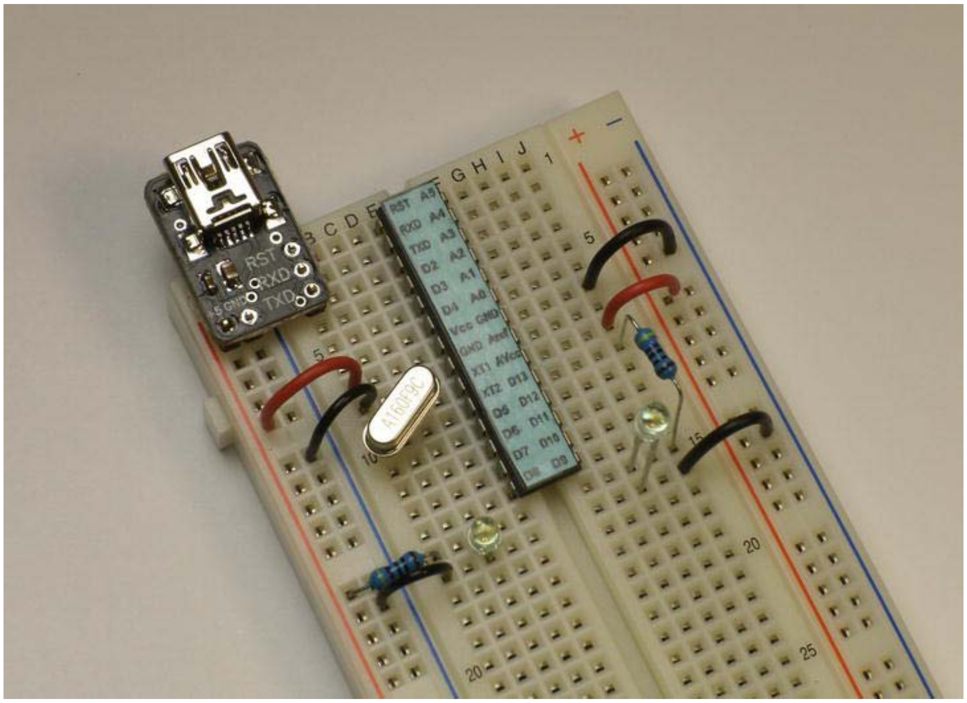

Figure 3-7. *A breadboard Arduino, which is compatible with the popular Arduino Duemilanove microcontroller, can be built using just a handful of components on a solderless breadboard, leaving plenty of room for more experimentation. The black, rectangular component in the upper-left corner is a miniature USB adapter that supplies power, and provides communication and control signals for the microcontroller.*

You’ll find that projects tend to expand to fill the available area (and budget). See Figure 3-8.

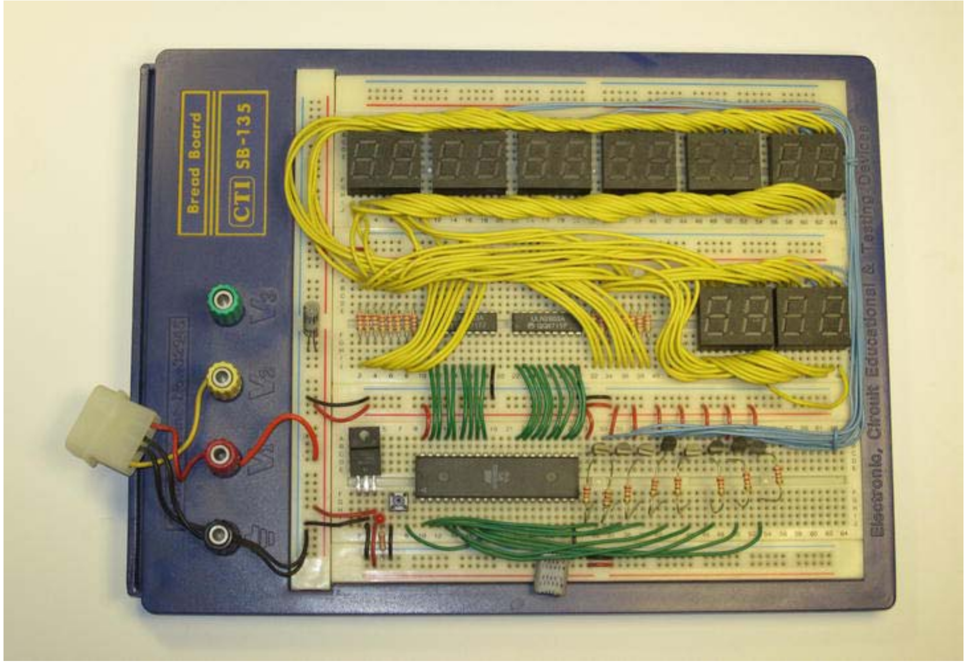

***Figure 3-8.** A prototype LED display with computer control, built on a solderless breadboard. That's a lot of wires!*

Passive Components

Most electronic components fall into one of two different categories: *passive* and *active*. There are a few different ideas floating around as to what, exactly, is the difference between passive and active components. For the purposes of this book, we will use the simple definition that an *active* component requires a power source to work properly, while a *passive* component does not.

Resistors

A good example of a passive component is a *resistor*. You've already learned a lot about resistance in an electronic circuit. Resistance is measured in ohms, and it converts electricity into heat. When you're working with wires, cables, and other conductors, resistance is usually a *bad thing*. You want the least amount of resistance possible. However, there are times when you both want and need resistance in a circuit.

One example of when you would want resistance in a circuit is when you want to light up an LED. Unlike an incandescent lightbulb, an LED has no built-in current-limiting ability. Without a fixed amount of external resistance added to the circuit, an LED would self-destruct in a very short time, because it would overheat from excessive current flowing through it. A resistor would limit the amount of current flowing and control the brightness of the LED.

The exception to this dire prediction is when the power source that is lighting up the LED has insufficient gumption to blow up anything, such as with a small coin-cell battery. Anything bigger than that normally has no trouble convincing the LED to release its magic smoke, without which it can no longer function.

A typical resistor has a fixed resistance and a maximum power rating. The resistance is determined by the material from which the resistor is made. The value of a resistor, as you already know, is measured in ohms, and can range from just a fraction of an ohm (often expressed as milliohms) up to millions of ohms.

Through-Hole and Surface-Mount Resistors

Resistors come in all shapes and sizes. Generally speaking, the larger the surface area of the resistor, the more power it can safely dissipate. A typical resistor that you might work with on a solderless breadboard or on an older PCB is shown in Figure 3-9. These are called *through-hole* components because their extended leads are meant to be formed and threaded through holes that are drilled into PCBs. Also in Figure 3-9, for comparison, are some teensy-weensy *surface-mount* resistors. Surface-mount resistors have no leads and are meant to be soldered directly to pads laid out on the surface of the PCB.

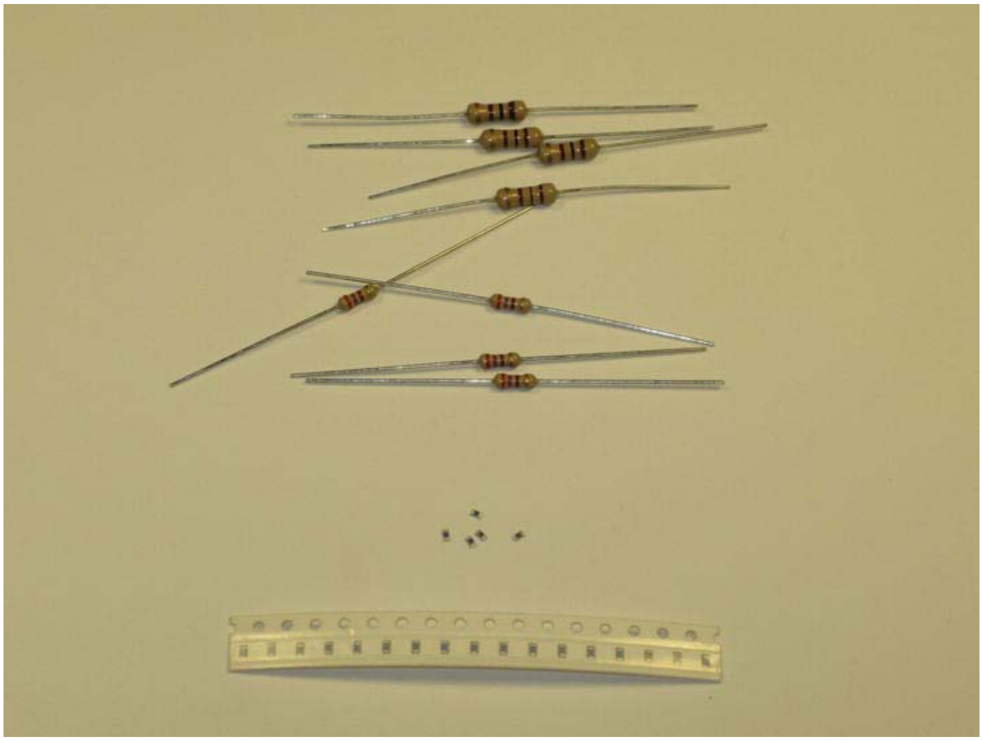

Figure 3-9. *Some through-hole resistors (top) and surface mount resistors (bottom). The surface mount resistors are so small that they are shipped packaged in tape, which is usually wound on reels for automated placement in PCB assembly.*

You can easily form the leads of through-hole resistors by hand or by using a *lead-bending jig*, a small tool with accurately spaced notches. Traditionally, these resistors have their leads formed so as to be threaded through holes in a PCB, and then soldered to the PCB. The extra lead length on the bottom of the PCB is then clipped off. See Figure 3-10.

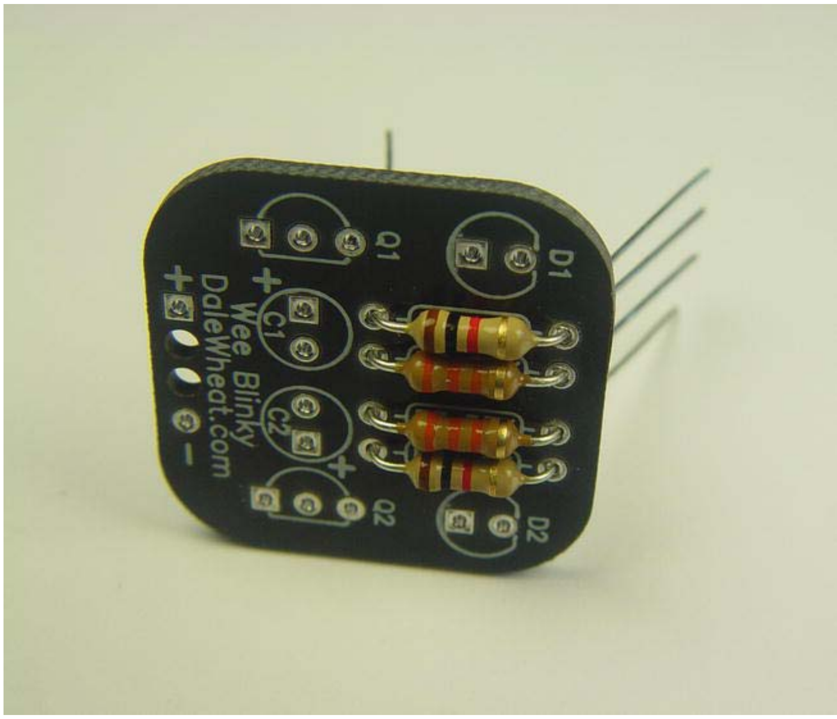

Figure 3-10. *Resistor leads are bent at right angles and inserted into holes in the PCB to be soldered.*

Schematic Symbols and Reference Designators

The schematic symbol for a fixed-value resistor is a zigzag line. See Figure 3-11. Every component in your circuit should have a name, which is also called a *reference designator*. Typically, most reference designators will have a short letter abbreviation to describe what kind of part it is (*R* is for *resistor*) and a sequential number assigned as more parts are added to the schematic. R1 would be the first resistor, R2 would be the second resistor, and so on.

If there is only one resistor in your circuit, then you don't have to number them, do you? You can just called it *R*, or *Galactic Emperor R the Mighty*, or whatever strikes your fancy. Most people stick to simple numbers.

Your numbering scheme depends a lot on the complexity of your schematic. If you only have a few resistors in your circuit, you can start at one and go up as needed. If you have a complex circuit with many subcircuits, you might allocate R101, R102, R103, and so on, to the first subcircuit (e.g., a power supply), and then assigned R201, R202, and so on, to a different section, such as a display. It's entirely up to you. Use a system that helps *you* keep track of things.

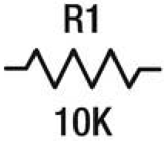

Figure 3-11. *The schematic symbol for a fixed-value resistor is a zigzag. The reference designator for this resistor is R1 and its value is 10,000 ohms, or 10KΩ. A simple rectangle symbol is also used to represent resistors, although many components are available in this physical form factor. Use the symbol you like best.*

Standard Resistor Values and Tolerances

Remember how wires have their own peculiar numbering scheme to describe their basic characteristics (i.e., the AWG, which uses *smaller* numbers to represent *larger* wires)? Well, resistors have their own peculiar numbering system as well to describe their electrical properties, or *value*. Even though the *value* of a resistor is measured in ohms and succinctly describes its characteristic resistance, not every possible value is readily available.

Some smart cookies at the Electronics Industries Alliance (EIA; see `http://www.ecaus.org/eia/site/index.html#`) have come up with some standard values that fill in most of the commonly used values for resistors and other components. Their scale is not linear but logarithmic, meaning that it takes larger steps as the values go up. This makes a little more sense once you understand how component value *tolerances* work.

You can buy, for example, a 1Ω resistor and expect it to be pretty close to being 1Ω at room temperature, but it's not going to be *exactly* 1Ω. It's going to be pretty close. How close? Well, that mostly depends on how much money you're willing to spend on this single resistor. If the answer is "not that much," then "pretty close" is going to mean within 5 percent of the stated value. That means your 1Ω resistor could actually measure as little as 0.95Ω or as much as 1.05Ω and still be "within tolerance." Five-percent resistors cost only a few cents even in small quantities.

Given that there's a little wiggle room for each resistor value, people who can tolerate their resistor values being off by as much as 5 percent don't need to be ordering 50 resistors of 0.99Ω and another 50 of 1.01Ω when they could just as easily (and certainly less expensively) order 100 resistors of 1Ω with a 5 percent tolerance.

This overlap in the values of the resistors allows us to skip from 1.0Ω to 1.1Ω without having to keep track of all the little resistors in between. This is because a 1.1Ω resistor could measure as little as 1.05Ω or as much as 1.15Ω and still be in tolerance.

With a standard resistor tolerance of 5 percent, we can use just 24 values to cover the range between 1Ω and 10Ω. The same multiples can then be used to cover the range between 10Ω and 100Ω, and so on. Aren't logarithms fun? See Table 3-2.

Table 3-2. *The E24 Standard Values for 5 Percent Resistor Value Tolerances (Higher and Lower Decades Are Also Available)*

1.0 10		100	1.0K	10K	100K	1.0M
1.1 11		110	1.1K	11K	110K	1.1M
1.2 12		120	1.2K	12K	120K	1.2M
1.3 13		130	1.3K	13K	130K	1.3M
1.5 15		150	1.5K	15K	150K	1.5M

1.6 16		160	1.6K	16K	160K	1.6M
1.8 18		180	1.8K	18K	180K	1.8M
2.0 20		200	2.0K	20K	200K	2.0M
2.2 22		220	2.2K	22K	220K	2.2M
2.4 24		240	2.4K	24K	240K	2.4M
2.7 27		270	2.7K	27K	270K	2.7M
3.0 30		300	3.0K	30K	300K	3.0M
3.3 33		330	3.3K	33K	330K	3.3M
3.6 36		360	3.6K	36K	360K	3.6M
3.9 39		390	3.9K	39K	390K	3.9M
4.3 43		430	4.3K	43K	430K	4.3M
4.7 47		470	4.7K	47K	470K	4.7M
5.1 51		510	5.1K	51K	510K	5.1M
5.6 56		560	5.6K	56K	560K	5.6M
6.2 62		620	6.2K	62K	620K	6.2M
6.8 68		680	6.8K	68K	680K	6.8M
7.5 75		750	7.5K	75K	750K	7.5M
8.2 82		820	8.2K	82K	820K	8.2M
9.1 91		910	9.1K	91K	910K	9.1M

Color Codes

Even the old-fashioned through-hole parts are a bit too small to legibly label with a numeric value and tolerance factor. A system of color codes was developed to designate resistor values, using colors to represent both single digits, a power-of-ten multiple, and the resistor's tolerance. These colors are painted on the resistors in bands. A little memorization and careful practice will have you reading resistor color codes in no time.

For 5 percent tolerance resistors, only four bands are needed to completely specify the resistor's value and tolerance. The first two bands represent the first two significant digits of the value. The next value is the multiplier. The final value is a special case to tell the resistor's tolerance.

The first two bands, as well as the multiplier value, use the color codes shown in Table 3-3 to represent the digits 0 through 9. (Unfortunately, these colors were chosen before the subtleties of colorblindness were widely understood.)

Table 3-3. *The Resistor Color Codes for the Significant Digits and Multiplier Digit of the Resistor's Value*

Digit	Color
0	Black
1	Brown
2	Red
3	Orange
4	Yellow
5	Green
6	Blue
7	Violet
8	Gray
9	White

The last band indicates the tolerance of the resistor using a different set of colors, as shown in Table 3-4. The most common are 5 percent (gold) and 1 percent (brown). There is supposed to be a slightly larger gap between the digits and the tolerance bands, but this is sometimes hard to see.

Table 3-4. *Resistor Color Codes for the Final Color Band*

Tolerance	Color
0.05%	Orange
0.25%	Blue
0.5%	Green
1%	Brown
2%	Red
5%	Gold
10%	Silver
20%	None

The multiplier digit can be thought of as either the number of 0s after the specified digits or as the exponent to which the number 10 is raised, as in scientific notation. Some examples make this easy to understand and remember.

A resistor marked with brown, black, red, and gold stripes has a resistance of 1KΩ with a tolerance of 5 percent. An easy way to decipher this code is to do a simple substitution for the first two digits (brown is 1 and black is 0, so we have 10 so far). Next decipher the multiplier band and write down that many 0s after the first two digits. Red is 2, so write down two 0s. That gives us 1000 so far. The final band is gold, and that means a 5 percent tolerance; 1,000 ohms is usually abbreviated as 1.0KΩ or simply 1KΩ.

■ **Note** *K* stands for *kilo*, the Greek prefix for 1,000. Likewise, *M* stands for *mega*, the prefix for 1 million.

This method works for most values from 10Ω to 10MΩ. A problem occurs, however, when we need to specify a value less than 10, as a negative multiplier value must be used to get an exponent of less than 0. For the value 1 and those less than 10Ω, a gold band is used for the multiplier. For values less than 1Ω, a silver band is used.

To save even more precious space on our schematics, another convention is to replace the decimal point with the multiplier. For example, 4.7KΩ would be written as just 4K7. 1.8MΩ would become 1M8 and 1.5Ω would be 1E5. 10KΩ would still be 10K using this method. If it is clearly understood that the unit of measure is the ohm, then even the Ω can be omitted.

Resistor Power Ratings

Any resistor has a maximum power-handling capability. This rating is usually expressed in watts or fractions of a watt. The larger through-hole resistors shown previously in Figure 3-9, for example, have a power rating of 1/2W (or 0.5W, if you prefer). This means that if you attempt to dissipate more than half a watt of power through the resistor, it is most likely going to quickly overheat and fail.

Thermal failure in a resistor starts with a thin curl of smoke, and then results in the discoloration of the resistor itself and sometimes actual flames. In any case, it gets dangerously hot, so do try to avoid this situation.

If you know the voltage across the resistor, as well as the current, you can easily calculate the power being dissipated using the simple formula P = V × I. Remember that P is the power value in watts, V is the voltage in volts, and I is the current in amps. One volt at one amp is one watt, which is way more than enough to smoke our little resistors. Be careful! Any time you calculate the correct value for a resistor in ohms, be sure to also calculate the required power-handling capability.

Variable Resistors

So far we've only looked at *fixed*-value resistors. They usually exhibit the same resistance day in and day out. That's what we love about them. But what if you need to vary the resistance in a circuit? We have resistors for that, too! A good example is a volume control on an amplifier. Turn it to the left (down), and the volume goes down. Turn it to the right (up), and the volume goes up. How does that work?

One kind of variable resistor is called a *potentiometer*. Potentiometers are sometimes abbreviated as *pots*. Pots usually have three terminals instead of two. Two of the terminals connect to the fixed resistor that is inside the potentiometer. Instead of being all sealed up like the fixed resistors you've been studying, the internal, resistive conductor is exposed along a track or groove. The third terminal, which is often physically in the center, between the two other terminals, is called the *wiper*, and can be moved from one end of the fixed resistor to the other end, making electrical connection to the inner part of the fixed resistor as it goes. As the wiper moves along the length of the resistive element, it taps into the resistor at a different point, varying the resistance. This is how a potentiometer can be used as a variable resistor.

Figure 3-12 shows a typical three-terminal potentiometer with a round body. The shaft can be turned to adjust the resistance. There are many variations on this basic design.

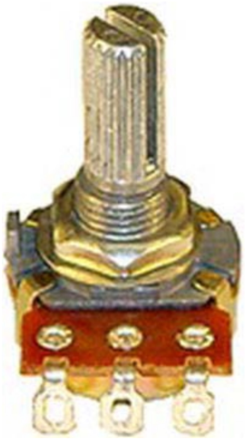

Figure 3-12. *A typical potentiometer, or variable resistor. Image by BG Micro (http://bgmicro.com).*

The schematic symbol for a potentiometer looks suspiciously like a resistor with an extra terminal, which happens to be exactly what it is. The extra terminal is the wiper, shown as the upward-pointing arrow. See Figure 3-13.

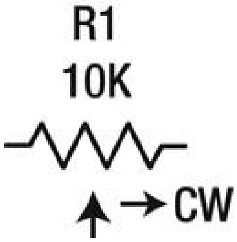

Figure 3-13. *The schematic symbol for a potentiometer is a fixed resistor with a wiper terminal. The "CW" arrow indicates which way the wiper goes when the shaft is rotated clockwise.*

Potentiometers generally come with two distinctive *response curves*, or *tapers*:

- A *linear* curve means that every equal step of rotation about the axis changes the resistance by the same amount. For example, if you were to center the shaft in the middle of its rotation, the resistance from the wiper contact to either of the end contacts would be nearly exactly half of the total resistance between the two end contacts. This is quite useful for when you need to adjust a voltage up or down in a relatively straight line. This is sometimes indicated by adding a capital *B* to the end of the resistance value (e.g., 10KΩB).
- In many audio circuits, however, it has been discovered that the human ear is anything *but* linear. A special response curve that more accurately describes this sensitivity is called an *audio* curve or taper. An *A* is used as a suffix to indicate that an audio taper is used for a potentiometer.

Note that linear and audio tapers are *not* interchangeable. If you find some random potentiometers in a parts bin, you can get a quick idea of what you've got by centering the shafts and seeing if the resistance is "symmetrical" or not, using an ohmmeter. Equal resistance to either end terminal indicates a linear taper, and anything else is mostly probably an audio taper.

Just like any other resistor, a potentiometer has a maximum power rating as well as a tolerance. These are usually either printed on the body of the potentiometer, if there is room, or hidden away in a product data sheet somewhere.

Specialty Resistors

There are many different kinds of resistors made for many special kinds of applications. High-power resistors are built out of sturdier materials that can take the thermal stress required to dissipate large amounts of heat. See Figure 13-14.

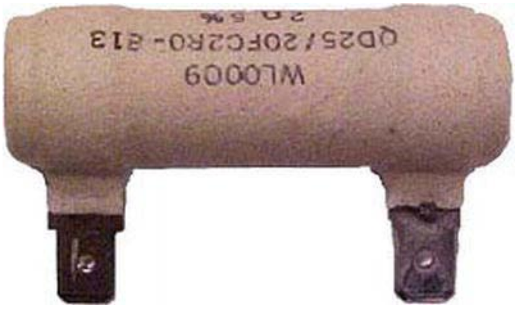

Figure 13-14. *A power resistor can handle much more heat dissipation than the smaller though-hole or surface-mount varieties. They are often so big that they need special mounting hardware. Image by BG Micro (http://bgmicro.com).*

Another popular option is the *multiturn* potentiometer. These potentiometers have an internal gear system that allows their shafts or wipers to be moved slowly across their internal resistive tracks by multiple turns of the control element. Increasing the number of turns allows the user much finer control of the resistance adjustment. It also takes the user longer to get from one point to another along the scale.

Capacitors

The basic definition of a *capacitor* is two conductors separated by an insulator. By itself, it doesn't sound like that would be a very interesting configuration at all. It turns out, however, that there's a lot more to it.

When a voltage is applied across a capacitor, no steady current flows through it. Remember, there is an *insulator* in the middle of that capacitor. Any insulator, just like the insulation on a piece of wire, *does not conduct electricity*.

What happens is that an *electric field* is created between the two conductors, or *plates*, as they are called when they are part of a capacitor. Now we get back to the fairies and dragons and molecules and so forth, way down deep at the heart of the problem. This electric field is a form of energy. Once it is built up by the difference in voltage, or *potential*, between the two conductors, it has nowhere to go, so there it stays . . . at least until something else in the circuit changes.

For this reason, capacitors are sometimes used as *reservoirs* of electrical energy in a circuit, holding a bit of spark here and there as needed. Just as it takes electrical energy to build up this electric field, a certain amount of electrical energy will be released when the field collapses. This will happen when the voltage between the two conductors drops.

Using capacitors as tiny electric batteries in a circuit is only one of many applications for this versatile part. While a capacitor will not conduct a *steady* flow of electric current between its terminals, it will often *appear* to conduct *alternating current* (as opposed to *direct current*), which is why you see them so often in audio circuits. As a charge builds up on one plate within the capacitor, it then attracts electrons from the other side, pulling them in from the other side. Even though they don't "jump the gap" of the central insulator, it looks like a tug-of-war is going on from both sides.

This behavior can also be used to send the higher-frequency components of a signal one way in a circuit and the lower-frequency components of a signal in another direction. This is one way that analog filters are built, and it has all kinds of useful and interesting applications.

Capacitors come in all shapes and sizes. They are made of many different materials, depending on what particular characteristics are needed. See Figure 3-15.

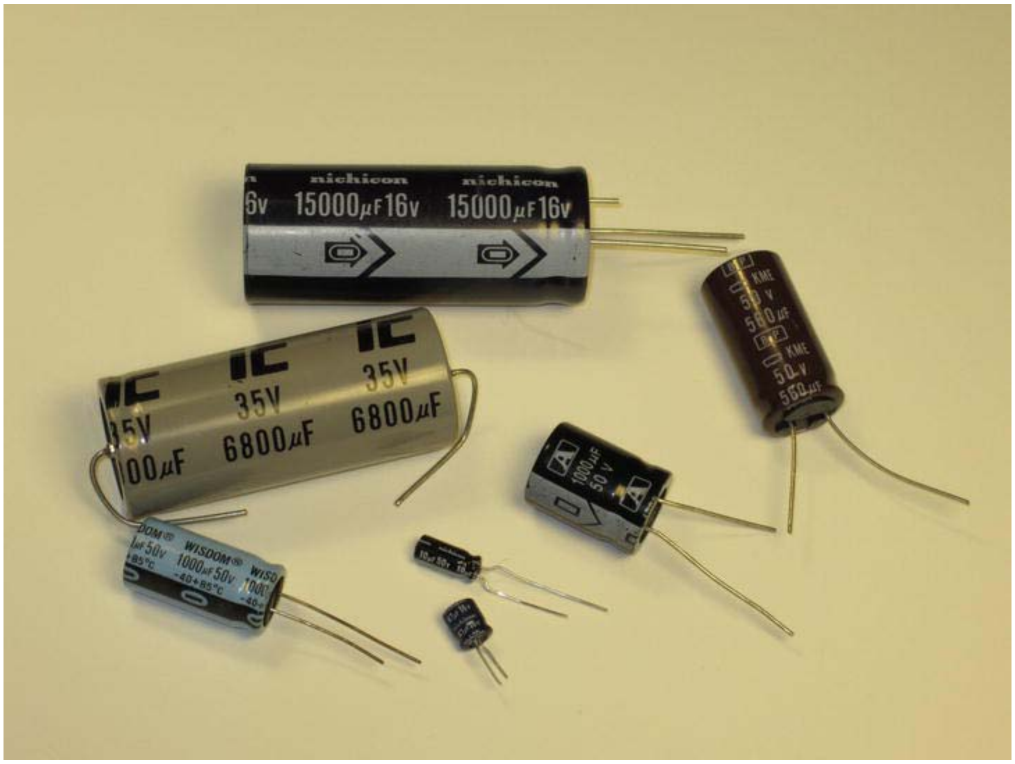
nichicon
15000μF 16v
nichicon
15000μF 16v
35V
6800μF
35V
6800μF
KME
50 V
560μF
1000μF
50 V
WISDOM
1000μF 50v
-40+85°C

Figure 3-15. *Some electrolytic capacitors of various sizes and capacities. Both much larger and much smaller capacitors exist and are very common in electronics.*

Schematic Symbols

While there are a large number of special-purpose capacitors that are custom made for certain applications, a very large number of capacitors fall into two broad categories: polarized and nonpolarized. The schematic symbols for both types of capacitors display the fundamental construction of two conducting plates separated by an insulator. See Figure 3-16.

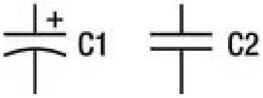

Figure 3-16. *The schematic symbols for polarized capacitors (left) and nonpolarized capacitors (right)*

C1 is a *polarized* capacitor, meaning one terminal is more positive than the other. In practice, this is indicated on the part itself in a variety of ways. One convention is that the *anode*, or more positive lead, is physically a little bit longer than the *cathode*, or negative lead. There is generally some sort of mark made on the capacitor body itself. Usually this is a stripe on the *negative* side of the capacitor.

What's really odd and potentially confusing is that one type of capacitor, specifically that made primarily of tantalum, has a stripe that indicates the *positive* lead of the polarized capacitor. Know your capacitors!

A polarized capacitor that is installed backward will experience terrible physical and thermal stresses and can sometimes explode, sending bits of itself all over the place.

C2 in Figure 3-16 illustrates a *nonpolarized* capacitor. It doesn't matter which way it is installed in a circuit. It is electrically symmetrical and works just as well in either orientation.

Inductors

An inductor is almost always made out of coils of wire. Sometimes inductors are quite large, as with transformers, while other times they are as small as a speck. An inductor would appear, electrically, as a simple short-circuit if it weren't configured to take advantage of certain principles of *electromagnetism*, or the tendency for the flow of electric current in a conductor to form a *magnetic field* around the conductor. The shape and orientation of the loops of wire that compose the inductor determine its electrical characteristics.

In many ways, an inductor is the opposite or complement of the capacitor. While a capacitor does not conduct direct current from one terminal to another, an inductor will happily do so, at least for a time. While a capacitor will resist a change in voltage across its terminal as it builds up its electric field, an inductor will resist a change in current flow across its terminal once it has established a magnetic field.

Because of their complementary natures, the two components can be used together to form a *tuned circuit*—one that resonates at a particular frequency, or allows the passage of only a certain range of frequencies, with the capacitor shuffling the high-frequency portions and blocking the low-frequency portions, while the inductor does the opposite, blocking the high-frequency components and passing the low-frequency ones. This is how many older radios and televisions worked before the advent of low-cost solid-state electronics.

Once an inductor has built up a magnetic field, any drop in the current flow across its terminals will result in the collapse of the field and the generation of a very-high-voltage spike. This is the basic electrical circuit that was used in automobile engines to create the high-voltage spark needed to ignite the fuel-and-air mixture in the combustion chamber.

Inductors can also be used to efficiently pass alternating current via a magnetic field from one inductor to another. This is how the primary windings of a power transformer *induce* a current in the neighboring secondary coil, which, depending on the ratio of windings and a few other factors, determines if the transformer *steps up* the voltage or *steps down* the voltage. It can even keep the voltage the same, for when electrical isolation is needed between two different systems.

Most electric motors employ some sort of inductor that alternately attracts and repels another electromagnet or permanent magnet. The most basic direct-current (DC) motors require a fairly complex mechanical arrangement to perform the necessary switch-swap to alternate the fields. This process is called *commutation*. In other types of electric motors, such as brushless motors and stepper motors, this commutation process is done electronically, externally to the motor itself, and requires more complex circuitry to control the speed and direction of the motors. See Figure 3-17.

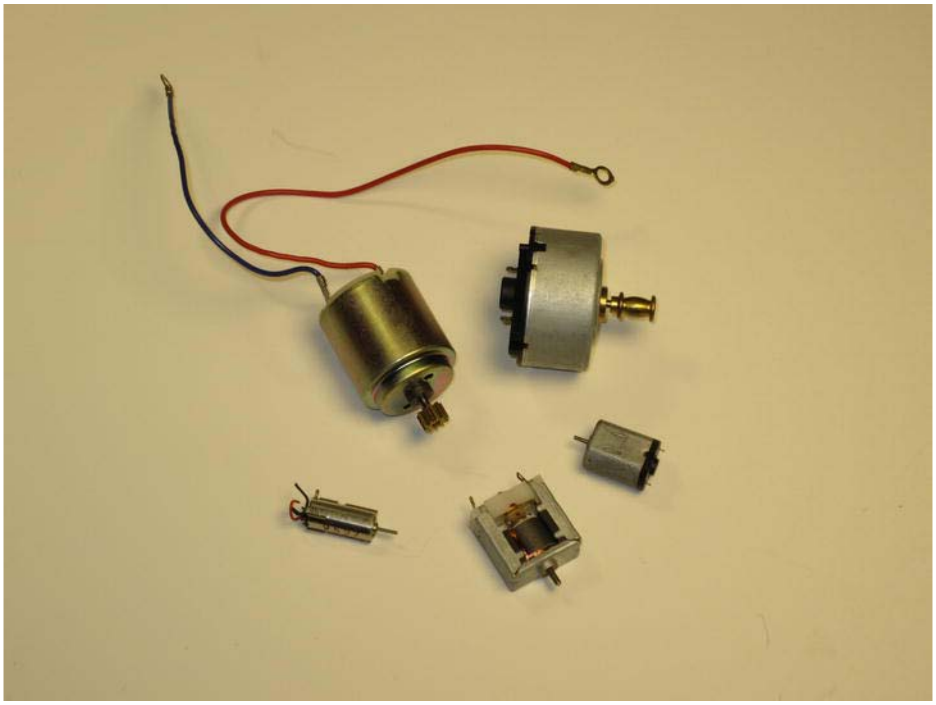

Figure 3-17. *Some simple DC motors. When direct current is applied to their terminals, they spin. The magic happens with an electrical and mechanical arrangement inside that performs the necessary commutation of the magnetic fields.*

Switches

The most basic way to control the flow of electric current in a circuit it to literally *break* the circuit. This means that instead of forming a conductive loop from power source to components and back again, a gap is introduced somewhere in the loop.

This gap can come anywhere in the loop. It can be right next to the battery, for example, or somewhere off in the middle—as long as the circle is broken, no current will flow.

If you perform the lightbulb experiment in Chapter 1, you can see how this works. When everything is lined up properly and the final connection is made, the lightbulb lights up. When any of the connections are broken, the light goes out. It's really as simple as that.

Most of the electrical switches that you operate every day are really mechanical in nature. Their main function is to physically move two conductors either together or apart.

In the lightbulb experiment, you could use the wire as a switch, by physically touching either the battery terminal or the lightbulb (depending on how you were holding everything together). With a little practice, you could probably send Morse code, or at least a variation on "one if by land, two if by sea."

There are many kinds of switches for many kinds of switching tasks. The most basic are *momentary action* switches. These are usually push buttons, and only conduct electricity when pressed. Alternately, they can be built so that they conduct electricity all the time, *except* when pressed:

- The first option is called *normally open*. When a circuit is open, it is not conducting electricity. When you *close* the circuit, by completing the loop, electricity can flow. A normally open switch would be welcome in a doorbell circuit. When no one is pressing the button, no bells are ringing. It would be annoying the other way around.
- A *normally closed* switch acts like a solid piece of wire (being a conductor of electricity) until it is activated, when it ceases to provide a path for the electricity to follow. This type of switch can be used as a *kill switch* on a gasoline-powered engine, such as a lawn mower, for example. By interrupting the flow of electricity generated by the magneto and powering the spark plug, you interrupt the internal combustion cycle, which will eventually cause the engine to slow down and then stop.

This type of switch is called a *single-throw* switch. It's either on or it's off—conducting electricity or not conducting electricity.

■ **Note** For almost every possible configuration of switches, there is a corresponding *opposite* configuration that is almost as popular.

Another useful configuration for a switch is to switch from one circuit to another circuit, and not just on and off. This is referred to as a *double-throw* circuit. A switch like this normally has three terminals. A *common* terminal is usually, but not always, in the center. At any rate, the common terminal is electrically connected to the moving part of the switch that is going to make contact with one or the other two remaining contacts. The *normally closed* contact is the one that will be making contact with the common terminal when the switch is not activated, or at rest. The *normally open* contact is simply the other terminal. This assumes that the switch has some sort of bias or mechanical spring action that returns it to a certain orientation when no one is activating it.

Some double-throw switches can remain in either state indefinitely. A typical light switch used in indoor lighting is a good example. It's just as happy to stay on as off.

Switches can have more than one circuit built into them. These switches are called *double pole*, *triple pole*, and so forth. They're like two or more single-pole switches physically taped together, so that when one is switched, they all are switched. Usually these poles are electrically isolated from each other.

We'll play with switches again in Chapter 5, where you'll learn some interesting ways to wire them up.

Active Components

While it could truthfully be said that *all* electronic components require some electricity to do anything, the distinction made in this book between *passive* and *active* components is that *active* components need a little more power to be really interesting.

First we'll look at sources of this magical electricity, and then we'll move on to some of the fascinating and complex components that will happily use it all up for you.

Power Supplies

You're going to need some sort of power source for any electric or electronic circuit you run across. If it doesn't use electricity in some structured form, it's not really electronics, now is it?

Batteries

You can use regular batteries to power a lot of your circuits, but they eventually wear out, and then you have no real recourse other than to throw them away or recycle them, if possible. If you're going to be using batteries, at least get some good ones. Many of the bargain brands are terribly inferior in their manufacture and are only providing you with a false economy. Spend a little more and get higher-quality parts, and you will be rewarded with longer life and more reliable operation.

A great way to work with batteries is to use the appropriate battery connectors or battery holders. Sometimes you can remove these from obsolete electronic devices that have outlived their useful lifetime. Always observe the proper orientation and polarity of batteries. Even smaller batteries contain an annoying amount of power, especially when it all decides to come out at once, such as when you accidentally plug something in backward or short something out. This can result in damaged parts, burns, or even fires. Be careful with batteries.

Having a fuse holder wired inline with your battery-powered supply is an excellent and cheap way to prevent such "unauthorized thermal events." Keep several spare fuses on hand as well. Fuses cost little and save much.

Rechargeable batteries are becoming more common, and they keep coming down in price while improving in longevity and power capacity, and decreasing in size and weight. Any rechargeable battery technology beats any kind of primary, or nonrechargeable, battery in the long run.

■ **Caution** The same warnings about battery polarity and fuses apply to rechargeable batteries as well.

Since there's no easy way to "turn off" a battery, it would be a good idea to include a power switch in your supply, if possible. Just make sure that whatever switch you select can stand up to the maximum current you will want to be extracting from your power supply. You should also make your switch large and easy to see, so that you can find it quickly when you need it. An obvious power-on indicator is also a great plus for a homemade power supply. This can be as simple as a green LED (with appropriate current-limiting resistor—see below) or as complex as a voltage meter indicating how much juice is left.

Transformers

Another great power supply that you probably already own is a power adapter or transformer from another electronic appliance. You should use caution, of course, with anything that plugs into the wall for power, especially if you're planning on monkeying around with the other end of it.

Almost all commercially produced power adapters have a slew of certifications plastered on them. Somewhere in there should be the basic data concerning what type of power input it expects, what it is going to try to produce, and what the maximum limits are on the output. Make sure that whatever it is that you're going to plug into your wall is specifically designed to plug into *your* wall, as electrical power standards vary from one location to another.

Make sure that any repurposed power supply you use does not get overly warm when in use. This is an indication that it is presently or has been overloaded in the past. Such a device is a fire-starter waiting for its magic moment.

You need to know how much voltage is coming out of your power supply. Even if it is clearly printed on the device itself, if you have access to a voltmeter, you need to verify the output voltage before connecting it to anything. Many lower-cost power adapters are *unregulated* and will produce an output voltage that will vary wildly depending on the load applied to it. A *regulated* voltage output, on the other hand, will provide consistent voltage levels on the output, even with varying loads.

USB adapters and chargers are becoming quite common and can provide *regulated* +5V at up to 1/2A and sometimes more. These are great devices to use for small electronic experiments. They are generally small, lightweight, and reliable.

A dedicated bench power supply is a very handy piece of test and measurement gear to have available in your lab. A good supply will have a variable voltage and a readout to show you what's going on (or coming out). A step up and you get to set the maximum current that it will supply (to keep things from getting out of control too fast). Additional desirable features include multiple outputs and remote control from your PC.

No matter what kind of power supply you use, use it responsibly. Don't leave equipment powered up if you're not using it. Don't attempt to bypass safety measures that are in place for *your* protection.

Diodes

A diode is a semiconductor device that will conduct current in one direction only. That's the simple story. Of course, the reality is much more complex. A typical diode is composed of two pieces of a semiconductor material, usually silicon, but sometimes germanium or some other exotic substance. The diode action occurs in the *junction* between the two types of semiconductors, which have been specially manufactured to contain the exact right balance of impurities required. This is generally not the kind of thing you can build in your lab, unless you're Jeri Ellsworth.

These two bits of semiconductor, called the *N region* and the *P region* (for negative and positive, respectively), are sealed up in either a glass or plastic enclosure. A standard diode has two terminals—an *anode* and a *cathode*—which represent the positive and negative leads, again respectively. Conventional current (i.e., current flowing from the more positive voltage to the more negative voltage) can flow through the diode from the anode to the cathode, but not the other way around. It's magic! Well, technically it's semiconductor physics, but either answer will serve for our purposes.

Diodes come in all sizes. Generally speaking, the more power a diode is expected to handle, the larger it is going to be. *Small-signal* diodes are the same size or smaller than typical resistors. Power diodes, often called *rectifiers*, depending on their application, can be so large they have to be mounted with large bolts.

The two most important electrical characteristics for the happy care and feeding of diodes are their maximum reverse voltage and forward current capability. You don't want to exceed either one.

Since a diode is, at best, only ever going to have a cryptic part number written on it, you'll need to consult the manufacturer's data sheet to find out all the necessary information. Many semiconductor producers offer "commodity" parts that conform to standard part numbering schemes. That way, you know that a 2N4148 diode from one manufacturer is going to be pretty close to a 2N4148 diode from any other manufacturer, but you should always check to make sure.

Diodes typically have a *forward voltage* that must be exceeded before current will flow. This voltage is essentially lost and is dissipated as waste heat from the device. Silicon diodes will have an average forward voltage around 0.7V and up. Germanium diodes will have a smaller forward voltage, in the area of 0.3V. Schottky diodes use a metal-to-semiconductor interface region and achieve substantially lower voltage drops, making them much more suitable for higher-power applications.

LEDs

One special kind of diode is the popular *light-emitting diode*. You see these little guys everywhere these days. In fact, some folks go around with black tape, just trying to cover them all up so they can get their darkness back at night.

The reason that they are everywhere is that they are cheap, reliable, and easy to use. They don't take up much room, don't take up much power, and don't generate that much heat. The exceptions to these admirable qualities, of course, are in the newer generation of LED lighting products that emit tremendous amounts of light and produce tremendous amounts of heat. Every month it seems that a new record is set for efficiency. Let's hope that trend continues.

Being essentially diodes, LEDs have both an anode and a cathode. LEDs only conduct current in a single direction. You'll need to make sure you limit the amount of current flowing through your LEDs, because the LEDs themselves sure aren't going to do it.

LEDs designed to be indicator lights can be powered with as little as 1mA. High-efficiency LEDs can be used for illumination purposes and still only draw 20–30mA each. These very modest power requirements make LEDs a favorite for microcontroller projects, as these output power levels are easily handled by most popular chips.

The forward voltage drop across LEDs is typically much higher than for a rectifier or signal diode. In fact, the voltage goes up as the emitted color of the LED goes from the infrared and red end of the spectrum (1.2–2V) up to the blue and ultraviolet end of the spectrum (3–3.6V).

Transistors

A true modern miracle, the semiconductor transistor ushered in the age of solid-state electronics. More than a million individual transistors can be used to build modern computer processors, all on a single chip.

The first types of transistors to be built and used were called *bipolar junction transistors (BJTs)*. This type of transistor is similar in construction to the diode, except that there are two semiconductor junction areas inside.

This arrangement of magic beans allows a small amount of electrical current to be applied to one of the three terminals of the transistor, called the *base*, which in turn allows a much greater amount of electrical current to flow between the remaining two terminals, known as the *collector* and *emitter*.

Two complementary types of BJTs are made, depending on the way that the positive and negative semiconductors are sandwiched together. NPN transistors will conduct conventional current from the collector to the emitter when current flows from the base to the emitter. The opposite variety, called the PNP transistor, does precisely the opposite, conducting conventional current from the emitter to the collector when the base is negatively biased.

Transistors can be used to make amplifiers, switches, current limiters, voltage regulators, oscillators, and any number of other useful circuits. They are generally very low cost, lightweight, and reliable.

As with any other electrical components, you need to be aware of the electrical limitations of your transistors. Even when a transistor is fully saturated and conducting very efficiently, it's still dissipating a certain amount of its energy as heat. Too much heat and say goodbye to your transistor.

A newer type of transistor is becoming quite popular in power circuit applications, and uses a different type of semiconductor property to accomplish its mission. The *MOSFET (metal-oxide–semiconductor field-effect transistor)* has three terminals, just like the BJT. Each terminal performs a similar job, but does it in a different way. The *gate* of a MOSFET is electrically isolated from the conduction *channel* between the *drain* and *source* terminals. However, when a proper voltage is applied to the gate, it creates an electrical field that influences the conductivity of the channel between the drain and the source.

The important differences between BJTs and MOSFETs lie in the type of signal needed to drive each of the transistors and how much power is wasted by being converted to heat. A BJT is a current-controlled device, where more current into (or out of) the base results in more current flow through the device. The MOSFET is a voltage-controlled device, where more voltage applied to the gate increases the current flowing through the channel.

Additionally, MOSFETs typically have much lower on-state resistance than BJTs, meaning they can more efficiently switch higher currents. Finally, MOSFETs have a *positive temperature coefficient*, meaning that they conduct less power as they heat up—which is a good thing. BJTs are subject to *thermal runaway*—conducting more current as they heat up, which in turn generates more heat, and so on, until something snaps, crackles, or pops.

Integrated Circuits

As parts get more specialized, they become less fun. There are a hundred ways to wire up some transistors, resistors, and capacitors, but there's generally only one way to hook up a thermocouple interface chip.

The good news is that it has become much easier for companies to design and build vastly complex chips that can do things for us that, honestly, most people cannot comprehend. And that's OK. We need airplanes that can, mostly, fly themselves. If we had to depend on humans to get everything done, you can imagine what kind of world we'd live in. Humans are notoriously sleepy, hungry, distracted, and forgetful. A voltage comparator chip does nothing but compare two voltages, all day, every day, perfectly, all the time.

Integrated circuits come in an unbelievably wide range of functions, packages, and capabilities. The first integrated circuits were amplifiers. Integrated circuits allowed manufacturers to make chips that have dozens or hundreds of transistors and other parts *all at once*. Just plop that little silicon chip into a plastic package with legs on it, and you're done.

What kind of chips do you need in your lab? Well, that's going to depend heavily on what kind of circuits you're wanting to design, build, or repair.

Many older chips had a standard 0.1" pin spacing that made them easy to plug into a solderless breadboard for prototyping purposes. Most newer chips have been considerably shrunk, because phones and tablet computers are all the rage these days. It's still possible to work with these newer chips, but it almost always requires mounting them to some sort of breakout board or expander that brings out their tiny leads to a more reasonable, and reachable, size.

Acquiring Components

Where do components come from? The same place as everything else—the store—just like food, just like money.

The electronic component manufacturing industry spans the globe. If someone's not putting something electronic together, they're taking something electronic apart.

Once upon a time it took some hunting to find sources for electronic parts. Swap meets and "ham fests" were great places to find components, assemblies, and entire appliances just ready to be hacked, studied, improved, or rebuilt. Now you have to spend time looking for places to *get away* from all the electronic hustle and bustle.

One of the great things about today's Internet is that it takes almost no effort to find both electronic parts and the information needed to use them. The real trick these days is finding the best deal.

If you're strictly a hobbyist, looking to spend some quality time alone or with your kids, learning about electronics and building fun projects, then it just about doesn't matter if you get the absolute, rock-bottom price on every component and tool. Even if you're a budding entrepreneur, you're still in a much better position today to leverage those brilliant ideas floating around in that noggin of yours than you would have been even ten years ago, all because of the ubiquitous nature of electronics and technology saturating our lives. Not only can you get the right tools and the best parts shipped right to your door for a song, you can easily network with knowledgeable folk from all over the globe, contributing ideas and brainstorming across time zones and borders.

Buying New Components

It would be more than a full-time job just keeping up with all of the new tools, parts, software, and hardware that are available today. Even if you could, it would take many lifetimes to make and grow the connections and relationships you need to successfully compete in the global economy.

That being said, a good place to start is in your own backyard. If at all possible, you should deal locally for the things you need. When what you want or need cannot be obtained locally, expand outward until you find it.

While it's certainly nice to be able to find all your needs met close to home, it's also good to know that the world is shrinking every day. The sort of deals and relationships that couldn't possibly have existed in your parents' time are happening right now.

If you are in a position to spend some money on quality tools and components, then by all means do so. You will not regret it. Buying questionable items is like buying your groceries out of some guy's van in a parking lot. Maybe he'll be there next week and maybe he won't.

Even the small fry can get treated like an honored guest when dealing with the right people. The trick is finding the right people. The best way to find them is to ask your friends and others with similar interest where they trade. Ask them how they feel about their relationships with their vendors. Are they just a number? Or are they a partner?

Recycling Used Components

It's really a bad idea to waste anything. In a more perfect world, there'd be no "other" category at the recyclers. Everything would have an exit strategy, especially the complex and fragile toys we're so addicted to these days.

Until that day, try to look at your trash with new eyes. Is it conceivable that someone else might be able to use this equipment, or is it truly past its useful life? Or are you just being lazy?

You can do your part, and you should believe that others are doing their part as well. Take the time to find them, find out what they can offer you, and find out what you have that they might be able to use. You won't know until you ask.

Be careful when harvesting components with which you are unfamiliar. Novice recyclers should give yesterday's cathode ray tube (CRT) TVs and monitors a wide berth. They can retain *lethal* voltages for considerable periods of time after being retired from service. Just leave them on the side of the road. The same goes for microwave ovens, or any equipment that has refrigerant in it. The stuff they use nowadays is quite toxic.

There is a growing surplus of both components and "last year's models" available to the student of electronics these days. Sometimes it's worth your time and money just to take something apart to see how it works. Then sometimes you find some useful parts inside that you can use in projects of your own.

Component Data Sheets

Every component ever manufactured has a data sheet. This data sheet contains everything you ever wanted or needed to know about this particular component. It's probably available right now on the manufacturer's web site, along with a dozen application notes. If not, then it's probably available on a university's computer system in a project folder.

This isn't always the case, sadly. Past a certain point, it doesn't make sense to allocate resources to information that no one wants or needs anymore. That point varies with each company, with some diligently archiving their information and others tossing it out the window at the end of the quarter.

When possible, collect and maintain your own private set of data sheets. When you acquire a new piece of equipment, get everything you can find in electronic form and stash it away somewhere where you can find it when you need it. If you're experimenting with some new components, squirrel away those data sheets, just in case you strike gold.

Have you built something cool? Write up your own data sheet for it! Pretend that you're still going to be interested in this project next year. Better yet, imagine that *someone else* could take an interest in your work. You'll be pleasantly surprised at how useful you'll find this information in the future, after you've slept a bit.

Summary

This chapter has covered a lot of different types of electronic components, but it has only scratched the surface of what is presently available.

It's perfectly possible to spend a lifetime pleasantly tinkering with electronics and never have a clue how any of the underlying principles work. Sometimes all you need is a *practical* understanding of what the different components do in a circuit. Hopefully this chapter has given you a rough outline of what is involved. Ideally, your interest has been sparked enough to pursue more detailed study elsewhere. Good luck and happy discovering!

Now that you're armed with some basic tools and some ideas about some of the available components, it's time to get busy building your own lab, one experiment at a time.

The remainder of this book will take you on a journey of trial, error, and experimentation that will help you build the best possible lab for your needs, both for now and in the future.

CHAPTER 4

A Portable Mini-Lab

Having a portable (or at least (*trans*portable) electronics lab comes in handy in several circumstances. Maybe you don't have a place (yet) for a permanent home for all that equipment. Maybe you need to bounce at a moment's notice, heading off on electronic adventures at the drop of a hat. Or maybe you like to keep everything where you can find it in a hurry, without having to rummage through shelves and boxes, looking for just the right tool or component.

Whatever your motivation for wanting a portable lab at your disposal, this chapter should help you get started.

Have Lab, Will Travel

First, a note or two is in order concerning *travelling* with your electronics lab. Nothing makes security screeners at an airport or customs inspectors at border crossings more nervous than a bunch of mysterious devices, especially devices that are tangled up in a rat's nest of wires.

Do what you can to keep your collection of tools and components tidy and in working order. You will most likely be asked to at least turn on any questionable items and demonstrate that they are not hollowed-out shells for transporting questionable items. It also doesn't hurt to clearly label equipment as to its function and ownership, which makes sense in any case, not just travel scenarios.

The Compact Executive Model

"A full mongoose is a slow mongoose." This is an important lesson learned from Rudyard Kipling's *Rikki-Tikki-Tavi*. Rikki, a mongoose, has to stay fast on his feet to be able to combat Nag the cobra and hope to win. Therefore, Rikki learns to only take a few bites at a time when eating and never gorge himself.

Although hopefully not in a life-or-death situation like Rikki's, you can learn a thing or two from this simple philosophy. Only take what you absolutely need when travelling, and never take something you can obtain at your destination. These things will only slow you down. Don't be a slow mongoose.

■ **Tip** Travel light.

A slim and lightweight briefcase or attaché can hold many of the tools and components you might need for your mobile electronic experimentations. The main advantage is that you will only need one hand to carry it (leaving one hand free to open doors, bribe guards, or fend off ninja—assuming you have two hands to begin with). See Figure 4-1.

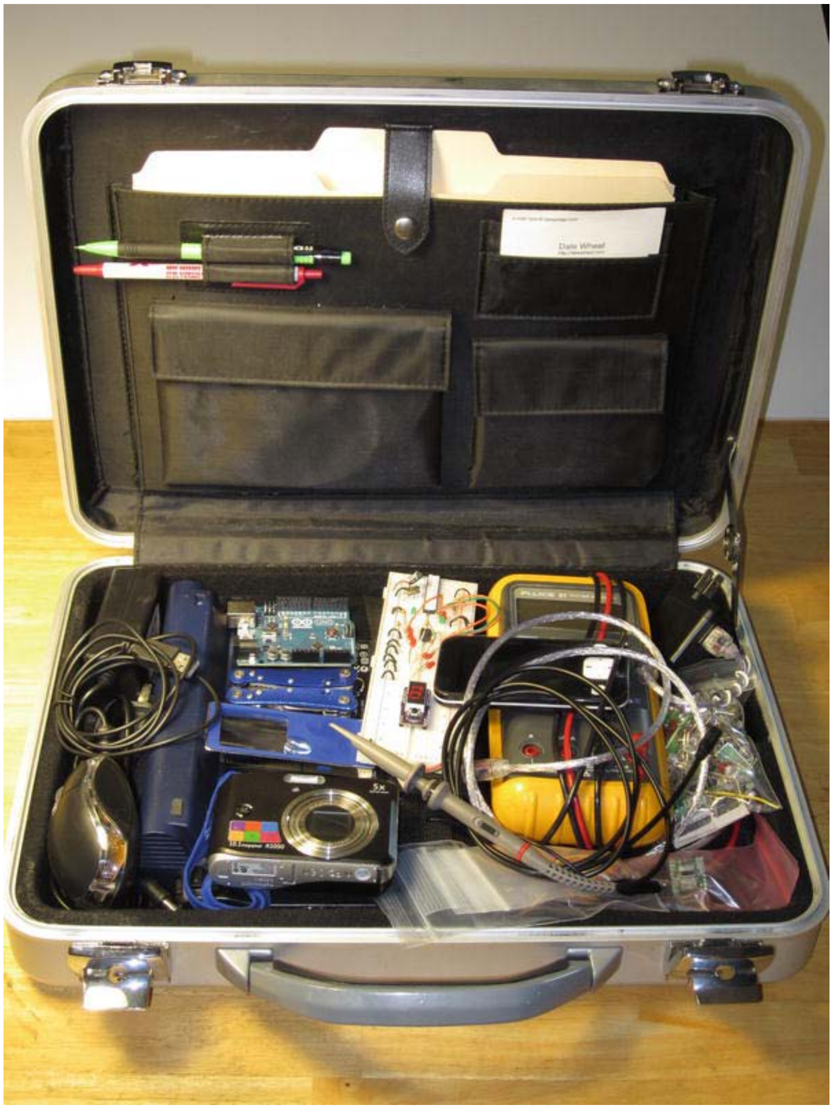

Figure 4-1. *The compact executive mobile electronics laboratory. There's plenty of room for lots of tiny things, but not much else. Also handy for transporting big stacks of cash, should the situation arise.*

Let's take a look at what all you can take with you without a lot of heavy lifting. Starting on the left and moving to the right, we have

- Netbook PC with AC charger and USB mouse (it might be hard to see in Figure 4-1 because it's upside down and covered by other stuff)
- Arduino Uno programmable microcontroller (or *physical computing platform*)
- Folding multitool
- Lighted magnifier
- Digital camera
- Solderless breadboard with lots of precut jumper wires
- Digital multimeter and probes
- Compact single-channel oscilloscope with probe
- AC-powered USB charger—can also be used as a stand-alone regulated 5V power supply
- USB cable
- A variety of LEDs in all shapes and sizes
- Additional passive components
- A variety of batteries, both rechargeable and otherwise, and various battery holders (not shown)
- Static-sensitive integrated circuits and modules stored in antistatic bags
- Notebooks, writing implements, blank media, and documentation
- Your contact information in case of loss

One of the interesting features of this lab is that it is completely self-contained, power-wise. While it's nice to have access to AC power, it's not strictly required, at least for a while. The Arduino and the attached breadboard are powered by the battery in the netbook, via the USB cable. The multimeter has its own 9V battery, while the mini-oscilloscope can be recharged from a spare USB connector. With an extended-life battery, you can expect four to five hours of on-location electronic experimentation.

If your netbook or PC has Wi-Fi or other connectivity to the Internet, you can browse tutorials, data sheets, user forums, and even this book, unless you've already downloaded these items to the internal hard drive.

Have a look at this particular lab deployed and ready for action in Figure 4-2.

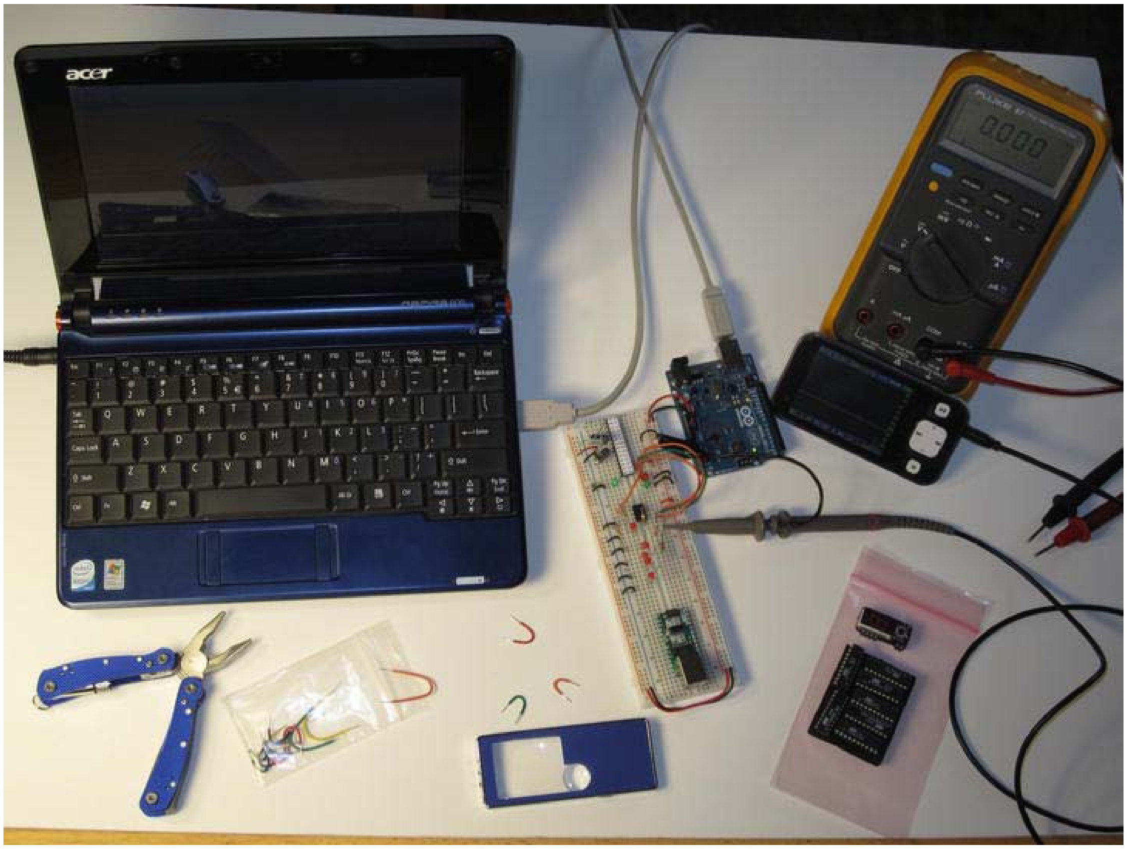

Figure 4-2. *The compact executive lab is deployed on a desktop. You have access to lots of electronic building and debugging capabilities in a small area. Using a solderless breadboard helps reduce the number of fabrication tools that you need to carry with you on your adventures.*

The Arduino was included for its small size, popularity, and ease of use. There's a whole world waiting for you inside that little board, waiting to be explored, but that's a story for another day.

You can use the USB cable to extract a regulated +5V supply from your PC for your low-voltage-experimenting pleasure. You can also use an inexpensive USB charger to do the same thing, if you don't feel like hauling a whole PC around with you. Just don't try to draw more than half an amp of current, which is the "official" USB limit for supplying power to USB devices. Technically, *unenumerated* devices (i.e., just a USB cable with nothing intelligent enough to identify itself to your computer) can only draw up to 100 mA (0.1A) of current. A USB charger, on the other hand, is not so picky.

No Soldering Required

One notable omission from this particular configuration is any sort of soldering tool. You could certainly add a small soldering iron, some solder, and a few more hand tools. You'll either need to boot out some of the bulkier items, such as the PC or the full-sized multimeter, or get a slightly bigger carrying case.

The Ever-Shrinking Computing Device

One option that will be gaining momentum in the very near future is the use of tablet computers or smart phones instead of the physically larger laptops or netbooks. They already have a lot of the computing capabilities of reasonably powered portable computers. Also, you can look forward to innovative hardware and software interfaces that will make them more useful in the lab as programmable test and measurement devices. While name-brand products still command a premium price, many no-name or OEM (original equipment manufacturer) products are being sold for a fraction of the cost.

Electrical Experiments You Can Perform

Here are some simple electronic experiments that you can perform using only the tools and components found in the compact executive mobile electronics lab. Start out with something simple and incrementally build on your success. Don't try the most complex problems first. What you want is to build up both your knowledge *and* your confidence at the same time.

Lighten Up, 20th Century Style

Here's an easy experiment that will teach you several interesting things about electronics, believe it or not. You can light up a typical LED using only a lithium coin cell. It needs to be a lithium coin cell that produces at least three volts (marked "3V"), not one of the smaller batteries that only supplies 1.5V, such as a watch or hearing-aid battery, as these will not produce enough voltage to light up a typical LED.

Slide the coin cell between the leads of your LED. There are four possible outcomes, three of which are "nothing interesting happens." Let's start with the least likely possibility, in which the LED lights up. Sometimes you just get lucky.

First Possible Outcome: The LED Lights Up When Attached to the Coin Cell

Figure 4-3 shows the results of this outcome. Notice that the longer lead of the sLED (the positive, or anode, lead) is touching the side of the coin cell marked with a +. No current-limiting resistor is required in this circuit because the battery's internal resistance is sufficient to keep from blowing up the LED. It just happens to work out that way, this time. Most other times, you will need to add a current-limiting resistor to the circuit.

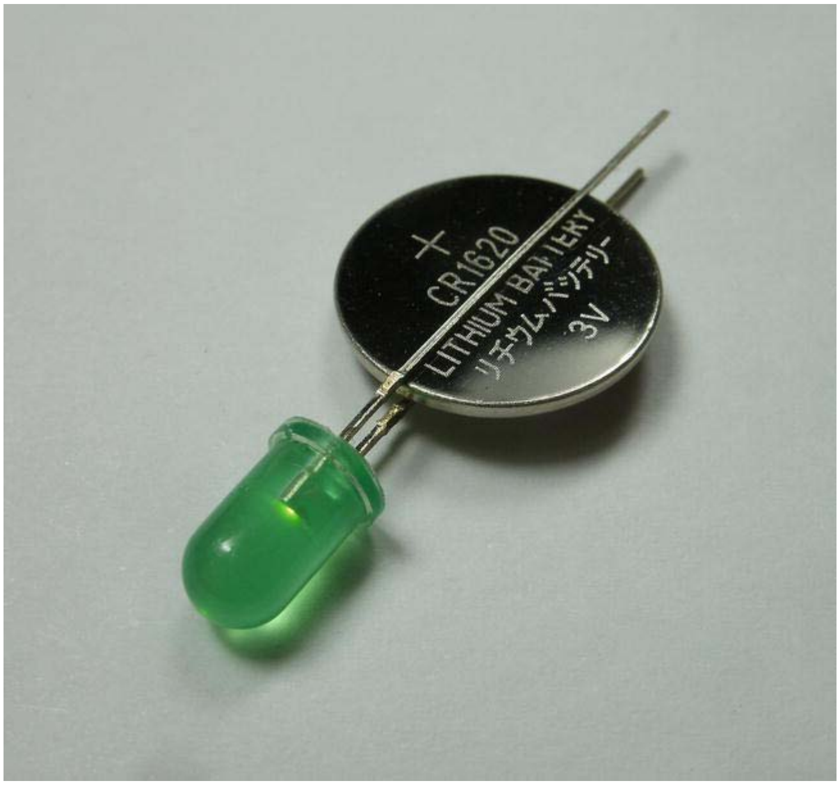

Figure 4-3. *A typical LED lights up when attached to a lithium coin cell. No extra wires needed!*

The other three scenarios? Let's look at them one at a time.

Second Possible Outcome: Nothing Interesting Happens

Why? Because the LED has been installed backward. Alternatively, the coin cell has been installed backward. Remember from Chapter 2 that LEDs are light-emitting diodes, and diodes only carry current in one direction. Therefore, when connected backward, they block the flow of electric current, preventing anything interesting from happening in the rest of the circuit.

Typical lithium coin cells have one broad, flat side that usually has something printed or etched on it, almost always with a + to indicate the positive polarity of the cell. The other side is often textured or rippled in some way to help make better electrical contact when installed in a device. This other side is the negative terminal of the battery.

The LED, being a diode, is *polarized.* It will only work when installed in the proper orientation inside the circuit. Most LEDs will offer one or more ways of easily identifying which lead is which (see Figure 4-4).

- Brand-new LEDs will have extended leads about an inch or so long. One lead will be slightly longer than the other one. The longer lead is the *anode*, or positive lead. The shorter lead is the *cathode*, or negative lead. This information is lost when the leads of the LED are trimmed—for example, after being installed in a PCB.
- Another industry standard for identifying the polarity of an LED is to look at the small ridge or shoulder at the base of the LED body. There will be a flat portion molded into the otherwise circular outline. This flat spot marks the cathode, or negative terminal of the LED.

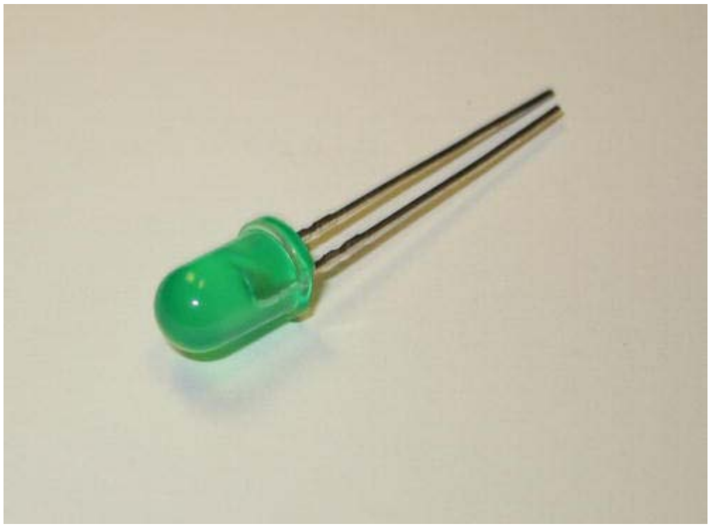

Figure 4-4. *Typical LEDs in the 3mm, 5mm, 8mm, and 10mm sizes almost always have two polarity indicators. The first is that the longer lead is the positive lead, or anode (top). The second is that the flat side on the base indicates the negative lead, or cathode (bottom).*

On really tiny LEDs, there may be an even tinier mark printed or etched onto the LED to indicate the cathode. You will sometimes need a magnifying glass just to see it!

The quick and easy fix for this problem is to swap the direction of either the LED or the battery. If this still produces no good results, we have to consider some more possibilities.

Third Possible Outcome: Nothing Interesting Happens

This time, after establishing that the LED has been properly oriented with respect to the polarity of the battery, we must entertain the possibility that the battery has been previously depleted. Coin cells have a limited amount of energy stored within them, and this can be easily drained in normal use over time.

A quick way to measure a battery's electrical goodness is to use a voltmeter. See Figure 4-5.

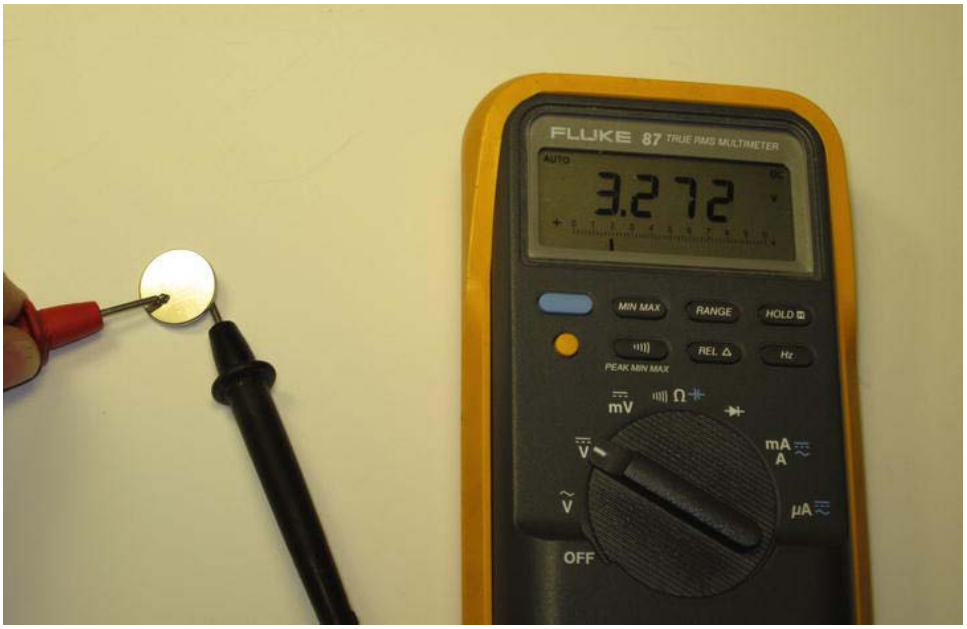

Figure 4-5. *Measuring a coin cell's electrical goodness using a digital multimeter. This one is chock-full of goodness!*

If you have a multimeter, be sure to set it to read DC voltage, in the appropriate voltage range (unless you have a fancy "autoranging" meter). Touch one of the meter's probes to one side of the battery and the other probe to the other side of the battery, and see how much voltage is measured. Anything below 3V should be considered depleted and of no further use, as far as electronic-type projects are concerned. The batteries are shiny and round, so they may have a second career in some sort of art project, but otherwise they should be disposed of in a responsible manner.

Also, perhaps because of their very shininess and roundness, they like to be eaten by small children. Don't let this happen! Take whatever steps are necessary to prevent small children from handling or even touching your electronic tools and components. Use extra care when your lab is mobilized.

■ **Caution** No unsupervised children, no matter how adorable, are allowed in the lab!

Fourth Possible Outcome: Nothing Interesting Happens Because the LED Is Defective

This can happen when an LED is subjected to too much current and the tiny, tiny wires within the plastic body overheat and fail. It can also occur when the LED has been exposed to the elements and moisture has penetrated the casing, allowing the internal connections to oxidize. This scenario, while within the realm of possibility, is certainly the least likely of the four, unless you are habitually cruel to your LEDs. It's also easy to test, if you have a known-good lithium coin cell handy.

Where's the Resistor?

You might be wondering where the current-limiting resistor can be found in this circuit. It's in there, all right, but it's hidden within the battery, so to speak. Coin cells are designed to provide tiny amounts of current over a long period of time. You're not going to arc-weld or start your car with one. As such, they have a much higher *internal resistance* than most any other kind of battery. We use this "feature" to our advantage when testing LEDs using just a coin cell, as the maximum amount of current that will normally flow out of the battery is below that necessary to destroy the LED.

A Slightly More Permanent Circuit

You can build a more permanent circuit for the coin cell and LED using a solderless breadboard. If you recall from the "Solderless Breadboards" section of Chapter 3 (see Figure 3-6), a solderless breadboard is an array of spring-loaded tie points, which serve to both physically hold electronic components and provide electrical connections between certain points. Let's build an LED circuit on a solderless breadboard using an LED, a coin cell, a resistor, a jumper wire, and a custom battery holder (see Figure 4-6). Later on, in Figure 4-7, you can see details of the custom battery holder.

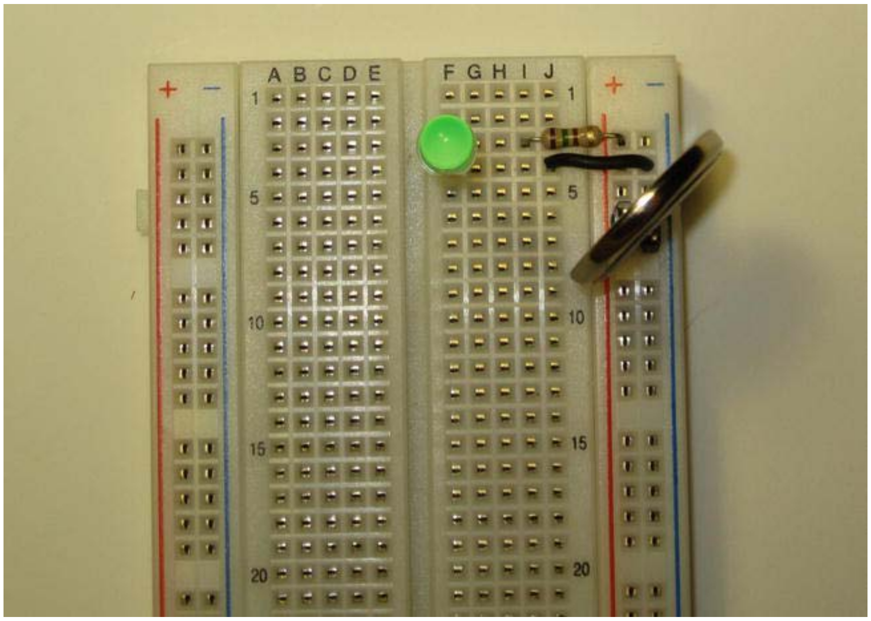

Figure 4-6. *A slightly more permanent LED circuit, built on a solderless breadboard. The tie points on the breadboard are labeled left to right with the letters A–E and F–J for the columns. The rows are numbered starting at 1 at the top. The LED's anode, for example, is installed in tie point F3. The power rails on either side are simply marked with + for positive and – for negative.*

Referring again to Figure 4-6, you can see the overall layout of the LEDcircuit. The power rails along both sides of the breadboard are just longer versions of the electrical connections that connect each of the groups of five tie points together. In this example, the right-hand power rails are used, but the left-hand side is not.

The schematic diagram of this circuit shows how simple it is, at least in principle (see Figure 4-7). Can you see the circle?

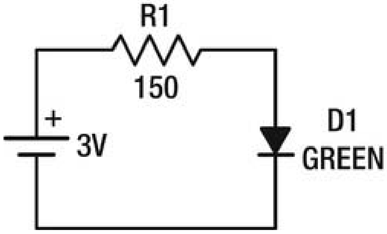

Figure 4-7. *The schematic diagram of the LED circuit as built on the solderless breadboard. Schematics are intentionally simpler than reality (most of the time) in order to help convey the very basic idea of how things are to be wired and the relationships between the components, without bogging you down in all the inevitable details that are involved.*

The Battery

The symbol for the battery actually looks like the battery in this case. The same symbol is used for most single-cell batteries, even when they are of completely different shapes or proportions. Multicell batteries are usually represented with repeated stacks of long and short lines.

The battery symbol's long line represents the positive terminal of the battery, and the short line represents the negative terminal. Again, it's just a coincidence that the wide side of the lithium coin cell is the positive terminal. The addition of the + sign in the schematic is redundant, but you'd be surprised how many people accidentally reverse this symbol. Better safe than sorry when it comes to electricity!

The Resistor

The battery's positive terminal is connected via a wire (drawn as a simple, solid line) to one side of the resistor. The schematic symbol for the resistor is a zigzag line.

The resistor is labeled "R1" (resistor 1) in this schematic, even though it's the only resistor and could just as well have been labeled "RESISTOR #1," or just plain "resistor." The numeric value underneath the resistor is the component's primary electrical characteristic, at least as far as we're concerned when looking at the schematic. In this case, it indicates a resistor whose *value* is 150Ω. In this simple example, the value of the resistor is quite flexible. A larger value, such as 1KΩ (1,000 ohms) or more, would also work, except that the LED would not shine as brightly. This is because *more* resistance results in *less* current flow, and it's the current flow that makes the LED shine. More current, more shiny.

We haven't specified any of the other important characteristics of the resistor, such as the tolerance or the power-handling capacity. Because this is such a low-power circuit, almost anything will do. When you start to work with really big LEDs that require correspondingly really big power supplies, this becomes a much more critical issue. For now, we start simply.

The LED

The other side of the resistor is connected, again, with a simple, solid line, to the anode of the LED, whose *reference designator* in this schematic is D1 (for diode 1). You could have called it "LED," and future generations would probably have understood your intentions, which should be your overriding goal. Its characteristic "value" in this case is "green." Of course, any color will do, as long as you have the eyes to see it.

The LED's schematic diagram effectively communicates its polarized nature. The arrow indicates the flow of *conventional current*. The flat bar indicates the cathode, or negative terminal of the LED. This corresponds to the flat side of the LED body in this example. The cathode is connected via a wire back to the negative terminal of the battery, completing the circuit.

Can you see the circle yet? Hint: It's disguised as a rectangle. The important point is that it forms a complete loop.

Building the Actual Circuit

Comparing the schematic to the photograph of the completed circuit might not immediately click in your head. Take a look at Figure 4-8 for a slightly different angle, where you can see some more details, such as how the battery is being mounted.

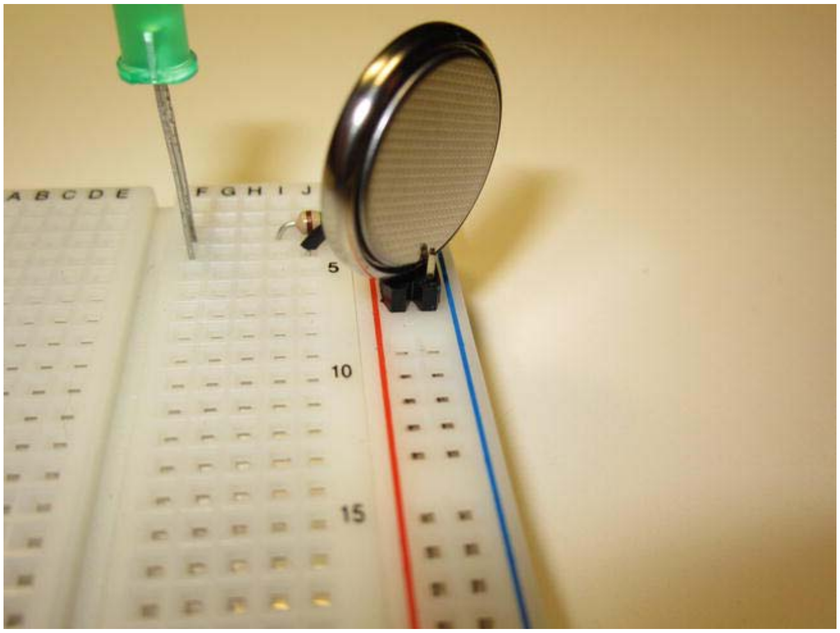

Figure 4-8. *Another view of the LED circuit as built on a solderless breadboard. The battery holder is a modified pin block that has had a couple of its pins removed, allowing the coin cell to fit in on the diagonal. A thinner coin cell could have fit without your having to remove the corner pins. A proper coin cell holder is inexpensive and would provide a much more robust solution to the problem.*

Looking at Figure 4-8, you can see how the positive side of the battery is making physical and electrical contact with the gold-plated pin of the pin block. The remainder of the pin emerges underneath the black plastic molding to make contact with the spring-loaded tie point within the body of the solderless breadboard, which in this case is the negative power rail. The positive terminal, in a similar fashion, makes contact with the positive power rail.

By attaching the battery directly to the right-side power rails, you ensure that your circuits are only a short jumper away from power, no matter where along the breadboard you decide to build them. Handy!

Note that the left-side and right-side power rails are not connected at this time. If you would like them to be, you can easily attach them using jumper wires. See Figure 4-9.

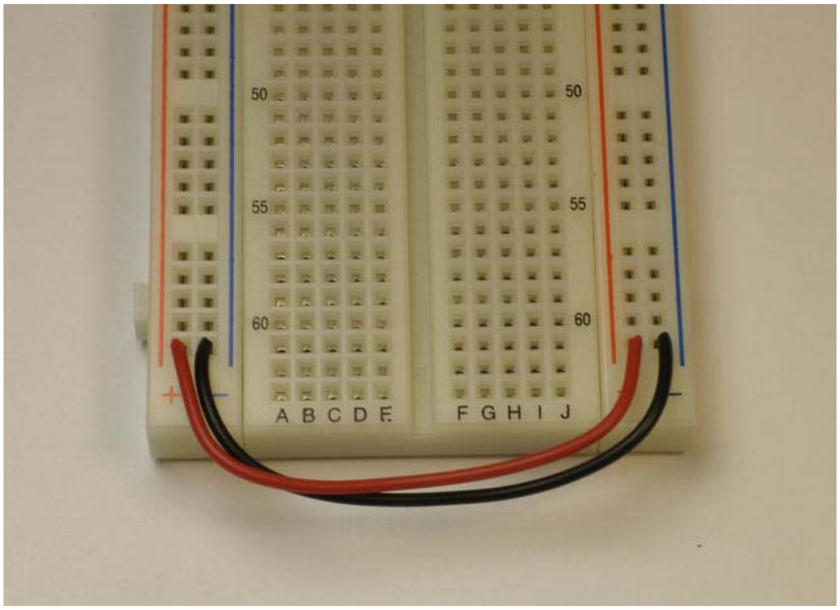

Figure 4-9. *Attaching the left- and right-side power rails using jumpers. By default, the power rails are not connected. There are some situations where you need them to not be connected, such as when using multiple supply voltages within a single circuit. These things happen.*

Now, referring *all the way back* to Figure 4-6 (sorry about that), you should be able to see how the positive terminal of the battery is indeed connected to one side of the resistor. The other side of the resistor, then, makes an electrical connection with the anode of the LED (the non-flat side, when seen from above). The cathode is connected to the negative power rail using a short jumper wire. The negative power rail then completes the circuit back to the battery. The circle is complete.

Note that the resistor could have been placed either before the LED in the circuit (as shown in the photographs) or after the LED. The only important points are that it is in the same circuit, or circle, and that the current must flow through *both* the LED and the resistor. It really doesn’t matter which one comes first, especially in this simple circuit. It will only make a difference when you start to do fancy stuff, like adding switches and such to complicate things.

If you’re using a different kind of solderless breadboard (and there are a few distinct varieties), you may or may not have power rails available. If so, fret not. You can still make point-to-point connections using more jumper wires. Now you’re starting to understand why you were assigned to cut and strip so many jumper wires back in Chapter 2! They should start to be coming in handy just about now.

Multiple Uses for the LED

Now we've moved from the 19th century incandescent bulb to the 20th century LED. What kind of lighting devices will we encounter in the centuries ahead?

One of the nice features of LEDs is that they don't typically require a great deal of power to operate. This is certainly true of indicator-style LEDs, whose only purpose is to be seen as either on or off. We can leave this LED installed right where it is and use it as a handy power-on indicator for future projects. You might be surprised how many heads get scratched in puzzlement, wondering why a newly wired prototype isn't working, only to find out after exhaustive investigation that the power wasn't turned on! Then again, you might not be surprised by this.

Now let's take this perfectly working circuit and break it—on purpose. If we interrupt the circuit, for example, by removing any of the components or their connections, the circuit stops working. To that end, remove the short jumper wire and replace it with two longer jumper wires. Plug in each of the new, longer jumper wires where the two ends of the short jumper were installed. See Figure 4-10.

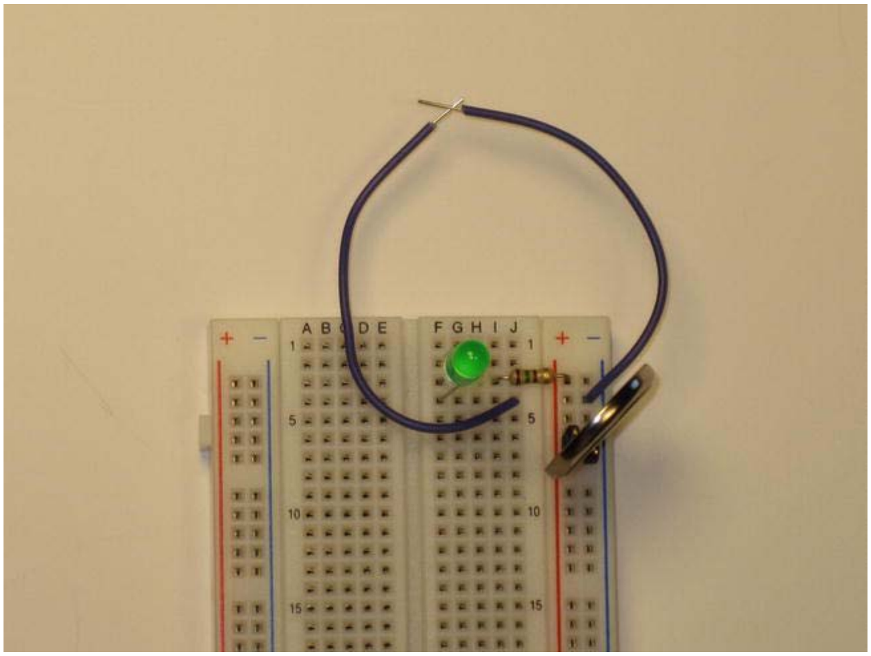

Figure 4-10. *By intentionally breaking the circuit, we stop the flow of electricity through the components, extinguishing the LED. By reconnecting the longer jumpers (or just touching them together), we reestablish the circuit, and the LED should light back up again. We can use this mechanism to test whether there is continuity in another circuit.*

Now, by touching the two probe wires together, you recomplete the circuit, and the LED should glow in happiness. This is exactly how a switch works. You can use this basic idea to build a continuity tester that lights up the LED when the current flows in the circuit. This is very handy for testing cables and wiring in a more complex assembly.

More electronic experiments, including how to both calculate and measure what's happening in a circuit, can be found in Chapter 5.

The Special Project Portable Laboratory

When you move past the simple joys of lighting LEDs and wiring up tutorial circuits, where do you go and what do you build? How about a robot?

Doug Paradis is an award-winning writer and robot builder. When Doug wants to work on one of his robots, he packs up a small kit of tools and heads over the local robot club, which in Doug's case is the Dallas Personal Robotics Group (DPRG). They meet every Tuesday at the Dallas Makerspace for an informal evening of robot building and socializing. See Figure 4-11.

Figure 4-11. *Doug Paradis of DPRG, arriving at the Dallas Makerspace for an evening of robot building. Doug uses the club's laser cutter and other specialized tools to give his robots that winning edge. He brings his own set of hand-picked tools along with him.*

Figure 4-12 shows what Doug brings with him to work on his robots.

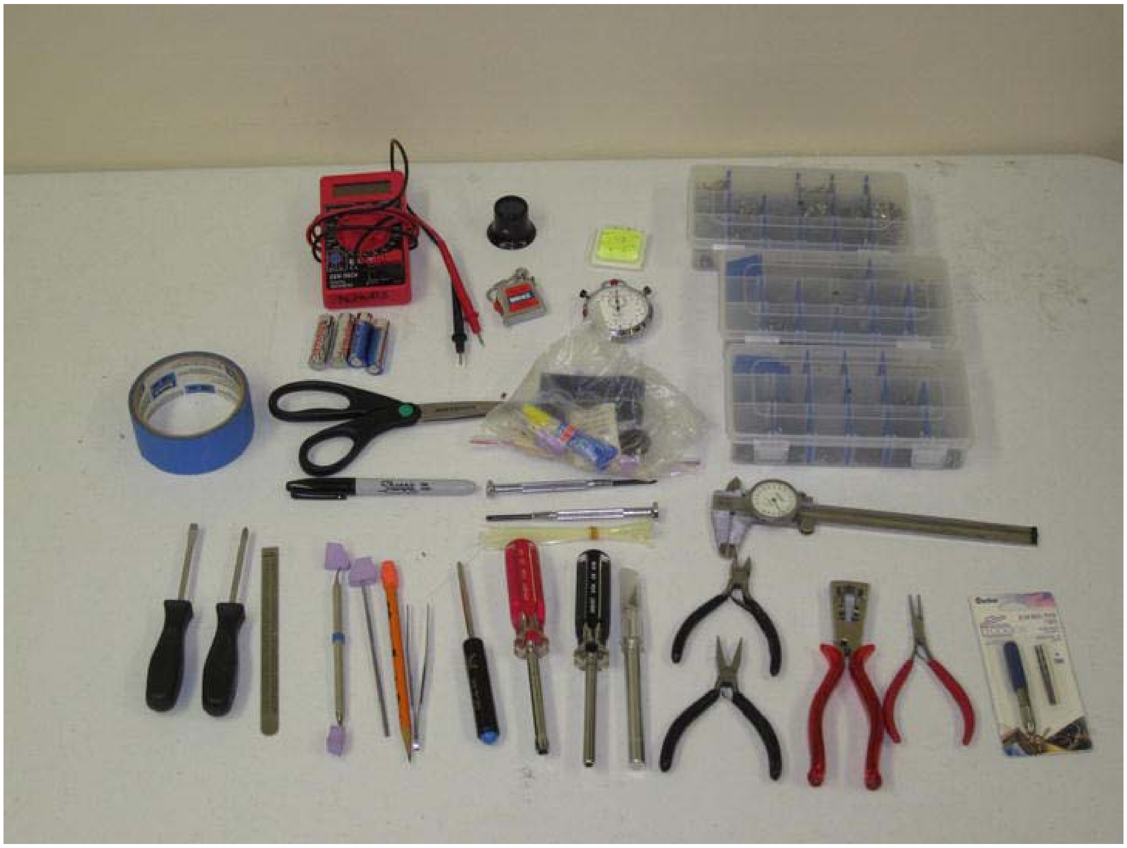

Figure 4-12. *The contents of Doug's portable workshop. A variety of hand tools, precison measurement devices, and fasteners (including tape, tie wraps, and super glue) are easily stored in a small tool bag for easy transport. The comparmentalized storage containers at the upper right contain a large assortment of screws, washers, and nuts.*

Doug has learned the lesson of travelling light. The electronic portions of his robotic creations evolve slowly over time, but the mechanical aspects get a lot of tweaking along the way.

Not everyone has perfected the art of light travelling. See Figure 4-13.

Figure 4-13. *A sort-of-portable robotics lab, with lots of tools and spare parts. One way of looking at this bounteous splendor is "three trips to the car."*

This caravan of tools, parts, and miscellaneous supplies forms the entourage of Kyle the Robot. Kyle is very much a work-in-progress. The basic framework is made of hand-tooled aluminum. The large wheels use O-rings for tires and are driven by 24V gear motors. The 24V supply is made up of two 12V, 5Ah rechargeable lead-acid batteries. Ah stands for amp hours, and represents the capacity of the battery. In this case, it can deliver an amp of current for five hours, or five amps for one hour. The custom PCB holds the electronics for the drive motors and will be controlled by a separate microcontroller that has yet to be added. See Figure 4-14.

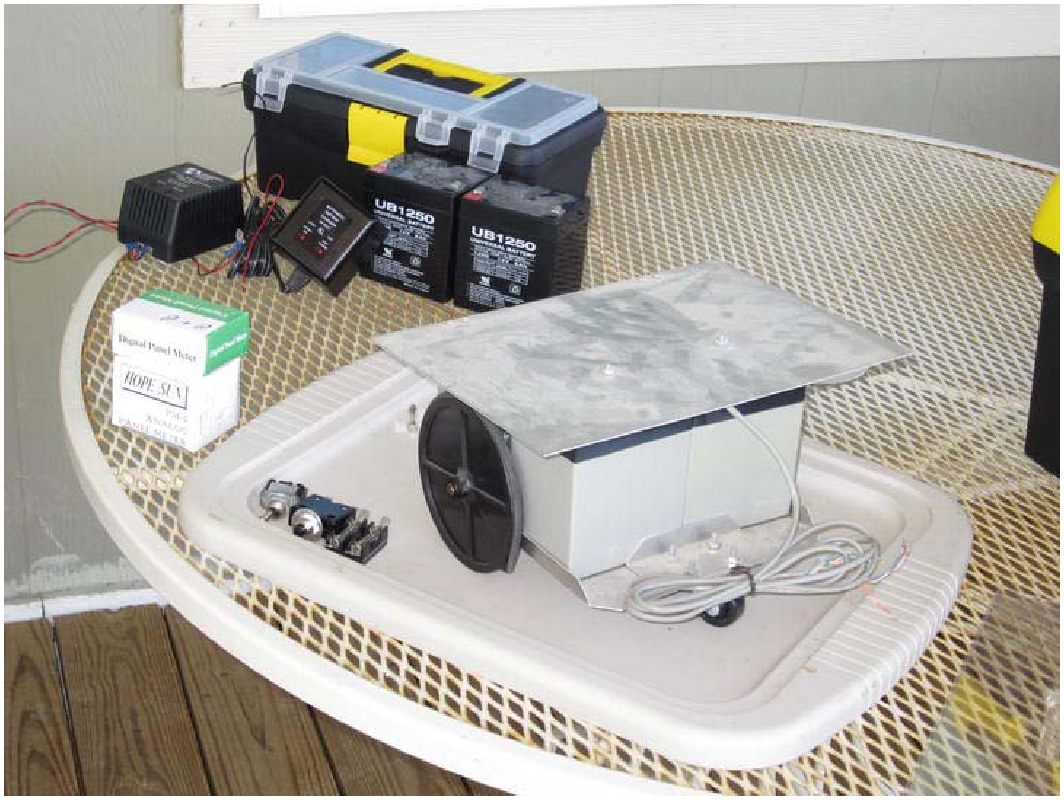

Figure 4-14. Kyle the Robot is a work-in-progress who requires a wide range of specialized tools and components. Some new components are being fitted, including dedicated battery voltmeters, fuses, a circuit breaker, and a master power switch. Once the sharp corners are filed off, Kyle will be able to roll around the house, looking for chores to do.

Summary

As we come to the end of this chapter, you should have some good ideas bouncing around in your head about how you want to pack your mobile electronics lab, what tools and components you *absolutely* must have with you, and how you want them organized. The best way to find out for sure what works for you is to make a few field trips, paying careful attention to what was actually used compared to what you anticipated needing. This will help you whittle down your cargo to the bare essentials, which is what you want. Remember, travel light. Excess baggage is a purely optional curse.

In the next chapter, we'll look at some suggestions for organizing and working in a more permanent setting. This could be a domain as small as the corner of a desk or as large as a warehouse. You will soon discover, if you haven't already, that this hobby, like many other projects, will quickly expand to fill the space available. Try to keep in mind that *you* are the one that ultimately makes the decisions. Effort spent planning ahead of time will help keep the sprawl to a minimum, while helping you stay focused on what you wanted to be able to do in the first place.

CHAPTER 5

The Cozy Corner Lab

Chapter 4 focused on the needs of the travelling electronic adept. Here in Chapter 5 you can come home to a lab that is always ready and waiting for some electrical fun and doodling. Nice!

You don't need an entire garage or extra office for a functional laboratory for your electronic endeavors. It's certainly nice if these areas are available to you, but you might be surprised at how much you can do in just a small space, if you set it up properly and maintain it with determination.

The first part of this chapter will deal with the task of rehabilitating a small, dedicated work area that has succumbed to years of poor organizational habits. These things happen. The second half of the chapter explores the possibilities of starting afresh with a nice, clean space—either something brand new or just newly cleaned. Some more projects help break in your new work area, as well as familiarize you with more interesting information about some of the basic electronic components reviewed in Chapter 3.

The Rehabilitation of a Small but Useful Area

Let's perform a little magic, in the form of a complete makeover for a small workstation. Have a look at Figure 5-1. Don't let this happen to you!

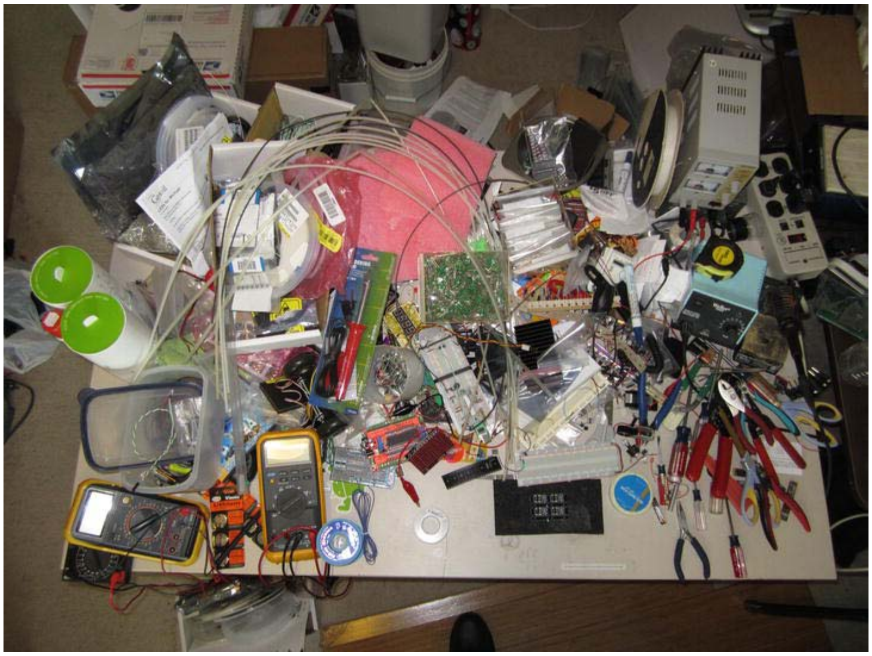

Figure 5-1. *A workstation that has been in continuous use for several years. Not much planning went into the original organization of this work area. Such a space can often accumulate flotsam and jetsam that might have at one time seemed appropriate, but now just competes for precious space. The effective work area is down to only a few dozen square inches! This limits the scope of projects that can be comfortably undertaken. It also causes unnecessary delays in rooting around for tools and parts. It's not pretty and it's not safe.*

Where to begin? Like any other project, this big project is made up of many smaller projects, none of which are especially difficult, if approached in the right order and with the right attitude.

If you're starting from scratch with your electronics lab and don't want to wade through this mess, jump on over to the "Adding a Power Source" section of this chapter.

Science to the Rescue!

Let's use the science of *taxonomy* to help break down this task into smaller, more manageable jobs. Taxonomy comes from the Greek words for "arrangement" (*taxis*) and "method" (*nomia*). That sound like *exactly* what is needed here! An "arrangement method" would assist in building up a plan for attacking this project.

■ **Tip** Method plus rigor equals success! Have a plan, pursue it with determination, and you will succeed.

Let's break down the overall category of "big mess on table" into two major *taxa* (singular: *taxon*), or categories. Ideally, what we will find here are *tools* and *components*. There will be other bits and pieces that don't easily fit into one category or the other (or will maybe fit into both), but it gives us a place to start.

Organization

Divide and *conquer*! These are the two words that will help us through the ordeal of rehabilitating this workspace. We'll get to do a lot of dividing here, and end up as conquerors. The spoils will be a useful (and tidy!) work area that will encourage you to begin new electronic adventures, instead of scaring small children.

The Great Divide

It would be possible to reorganize this workspace *in situ* (i.e., in place), but it would require a great deal of inner peace and lots of patience. Lacking these, a complete reboot is in order. Let's clear off the entire table and start afresh. This will provide the excellent opportunity to arrange things in exactly the best possible order.

First, get two big containers. These can be cardboard boxes or plastic storage bins—it really doesn't matter. We'll just be using them as temporary staging areas as we make like an embryo (and divide and divide again). See Figure 5-2.

Figure 5-2. *Two large cardboard boxes are enlisted to help with the initial organization of the contents of the tabletop. One will be for tools and the other one for components. Further subdivisions will be made later.*

Now, the first thing you will want to do is remove any items that really don't belong in the lab at all, such as coffee cups, personal effects, and any obvious trash. Everything that is about to go into one of the two boxes should eventually find its way back to the work area, so now is the time to omit the things just don't belong here.

Is it a tool? Is it a component? How do we tell? Ask five people and you will get five different answers. For the purposes of this little endeavor, a *component* is something that will become a part of a project, while a *tool* is something that is used to modify or manipulate a component. Another way of looking at it is that components get used up quickly, while tools get used up slowly. If you build enough projects, you will be lucky enough to see exceptions to these guidelines.

Here are some examples. Wire is a component. Wire cutters are tools. That was easy. A flashlight is a "visual inspection aid," so it goes under the category of tools. Batteries for the flashlight are consumable, so they are components. There's no real need to be ultra-precise at this point. Use your best judgment and make up policies that make sense to *you*.

It might help, from a physical logistics standpoint, to use smaller subcontainers to hold the really small bits as you're lumping them into the two broad categories. Folding cardboard bin boxes are handy both on the tabletop as well as on shelves, if you have them. The cardboard bin boxes shown in Figure 5-3 are from Uline (`http://uline.com`), part numbers S-16268 (4" wide) and S-16269 (6" wide).

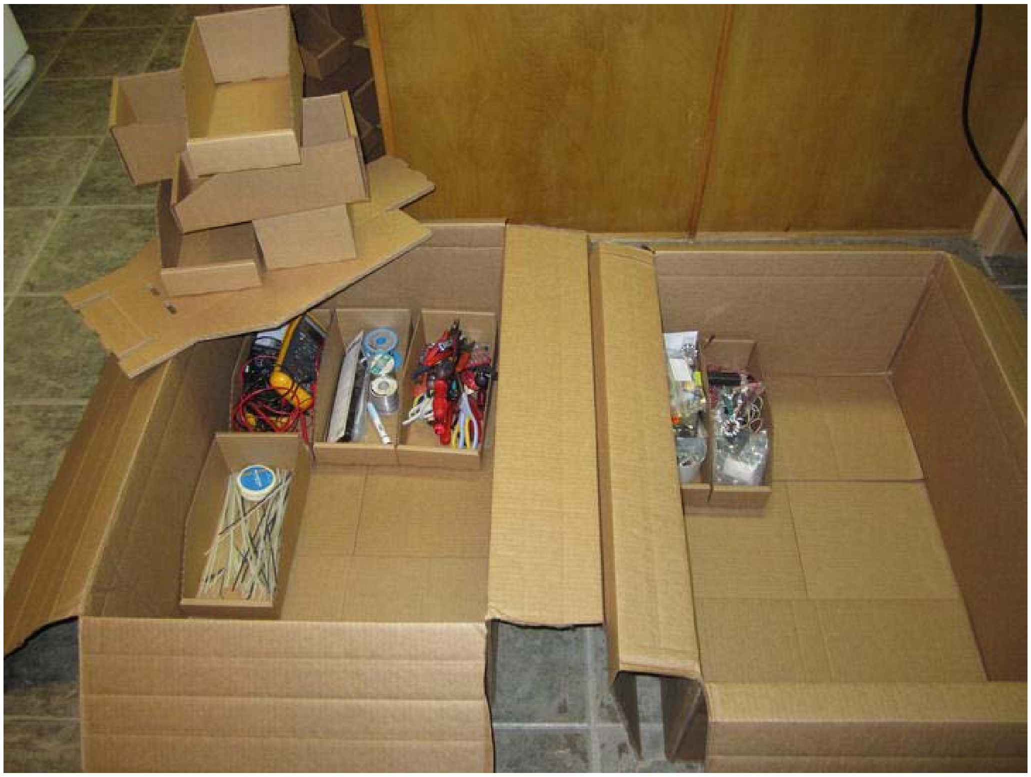

Figure 5-3. *Folding cardboard bin boxes help contain some of the smaller bits and pieces during the Big Sort. You can use whatever happens to be available. Some preliminary subsubdivision is going on here. For the most part, the tools are on the left and the components are on the right.*

After everything is removed from the table and the cardboard boxes are nearly full, the table itself should get a good washing. Behold the *tabula rasa* (blank slate) depicted in Figure 5-4.

Figure 5-4. *The tabletop has been cleared of tools, components, and various bits of debris. A good washing with soap and water completes the cleansing. A burned spot (the result of a poorly supervised thermal test) is covered with a small sticker (see inset).*

Adding a Power Source

Here's where a bit of planning will save you a lot of headaches in the future. What you never seem to have enough of in any laboratory, especially electronics labs, are lots of conveniently located power outlets. Let's take care of that issue right away, before building anything else onto the tabletop. See Figure 5-5.

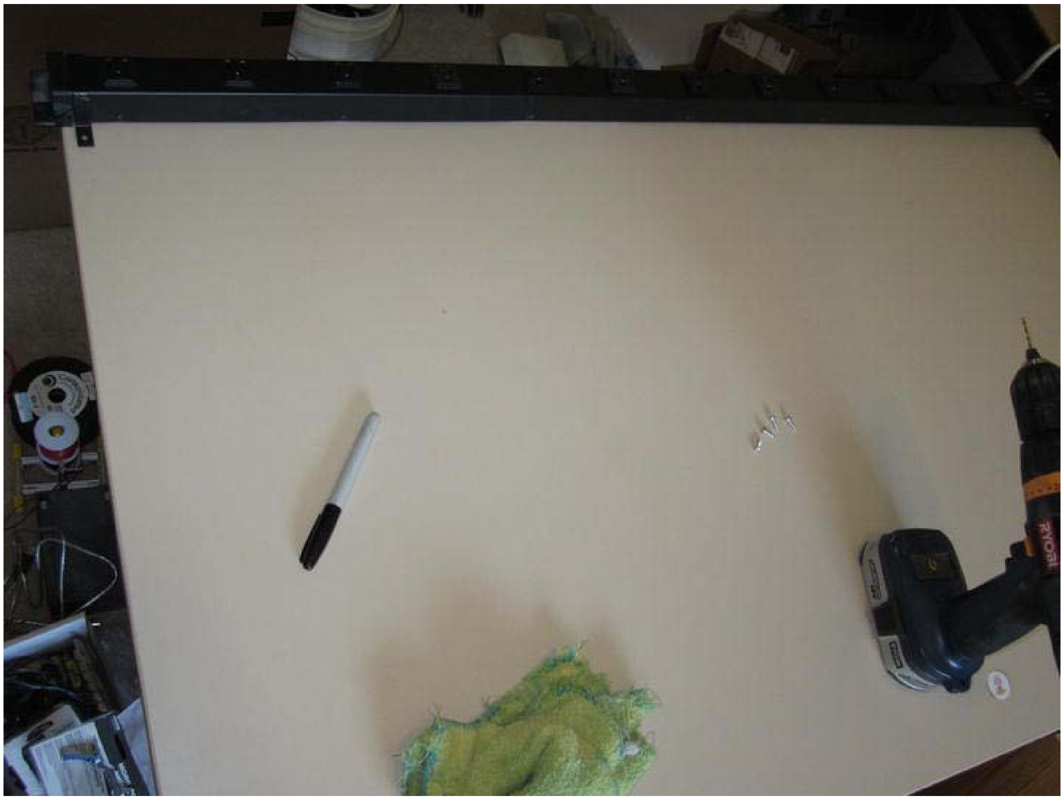

Figure 5-5. *A 12-outlet power strip is attached to the back edge of the tabletop. This will provide a safe and convenient source of power for all those gadgets you'll eventually want on your bench. Tools needed for the job include a drill motor to drill out some small pilot holes for the brackets, a small drill bit, and a felt-tipped marker. Before drilling the pilot holes, align the brackets where you want them to go and use the marker to put a dot where the pilot hole should be drilled. Make sure the power strip is not plugged in when you're drilling right next to it!*

Having lots of extra power outlets is a nice bonus, but is not absolutely required. What you want to avoid, however, is excessive daisy-chaining of power strips and extension cords. This is a fire hazard. Remember from Chapter 3 that every conductor has a certain amount of resistance. This resistance turns electricity into heat. Enough heat makes things catch on fire.

If you do have the luxury of being able to mount a permanent power strip (which, technically, remains a temporary power tap as far as the manufacturer is concerned), then be sure to install it securely using the hardware provided. See Figure 5-6.

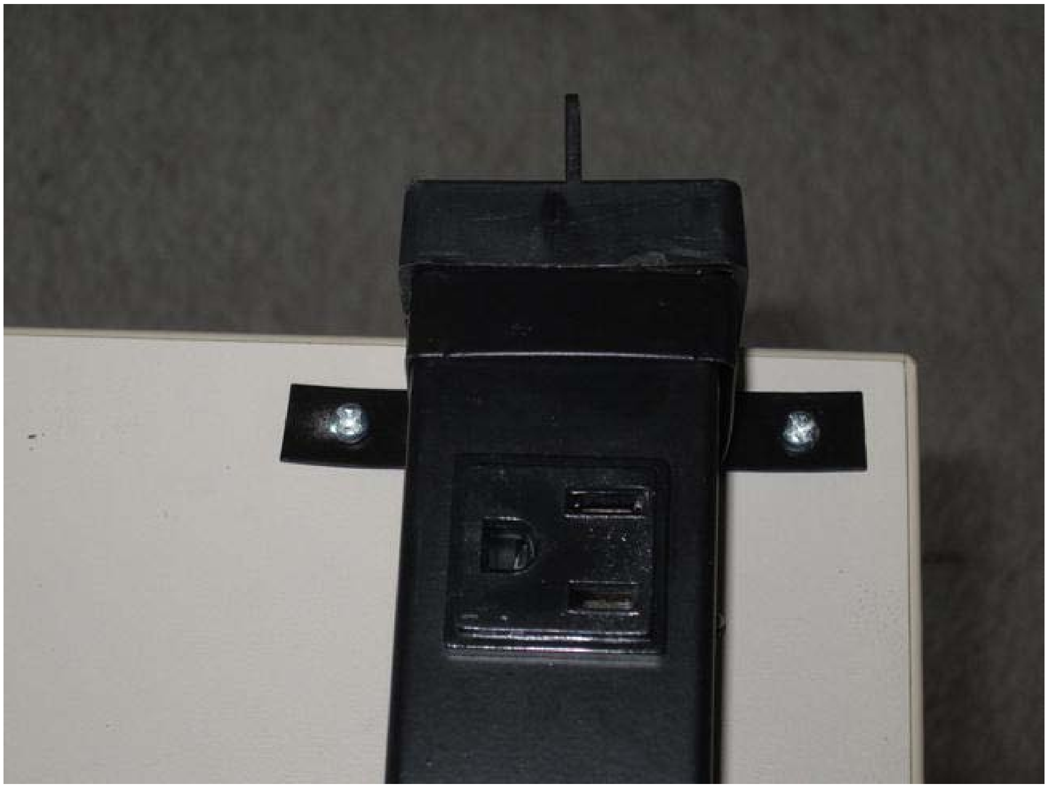

Figure 5-6. *The power strip is fastened securely to the table using the provided hardware. You do not want your power strip wiggling around, or worse yet, falling off the table, dragging all of your power tools with it!*

■ **Tip** Test every outlet of your power strip before permanently mounting it to your workbench.

A Clean, Well-Lighted Space

You can never have enough light, it seems. What passes for normal room illumination might be sufficient for such adventurous undertakings as finding your way to the door or looking for the remote control for the television. For anything more demanding, you're going to have to add some more lighting, and probably lots of it.

At a minimum, you are going to want some *task lighting*, which helps illuminate the immediate area of your focus. In this case, where a small table is being used as a workbench, an articulated fluorescent lamp with a built-in magnifier will do a great job, at least until more lighting can be strung up. See Figure 5-7.

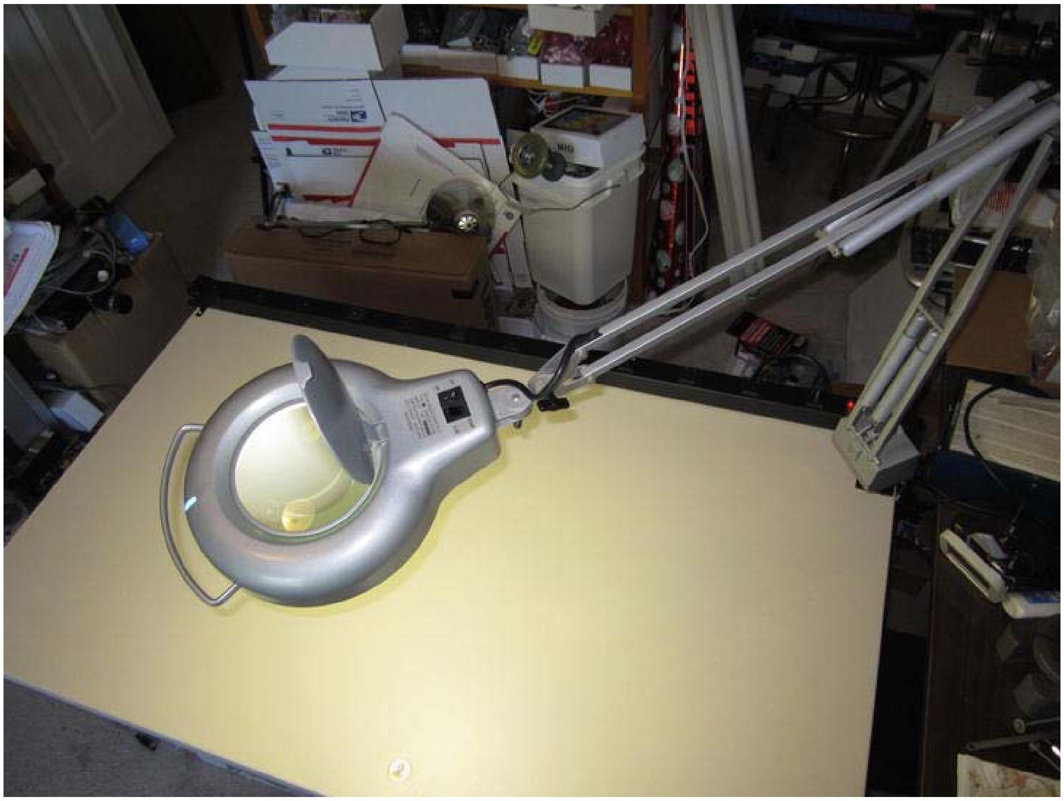

Figure 5-7. *You will need some task lighting for your bench. This desk lamp also includes a handy magnifier for really close work.*

Tools

Now we're ready to bring some of the tools back in to the lab. Here's where you get to decide which tools get a *permanent* home on the bench and which ones get put up and stored when not in use.

The permanently deployed tools on your workbench should be the ones that you find that you are using the most. The ones that get put up after every use are the ones that either only get used infrequently or pose some sort of hazard, such as sharp edges or hot surfaces. Again, you get to decide. It's your lab.

Tools for Soldering

An activity that is practiced frequently in many electronics labs is soldering. If you haven't learned how to solder yet, you are strongly encouraged to do so. Here's a handy setup for soldering, sitting in one corner of the newly refurbished workbench. See Figure 5-8.

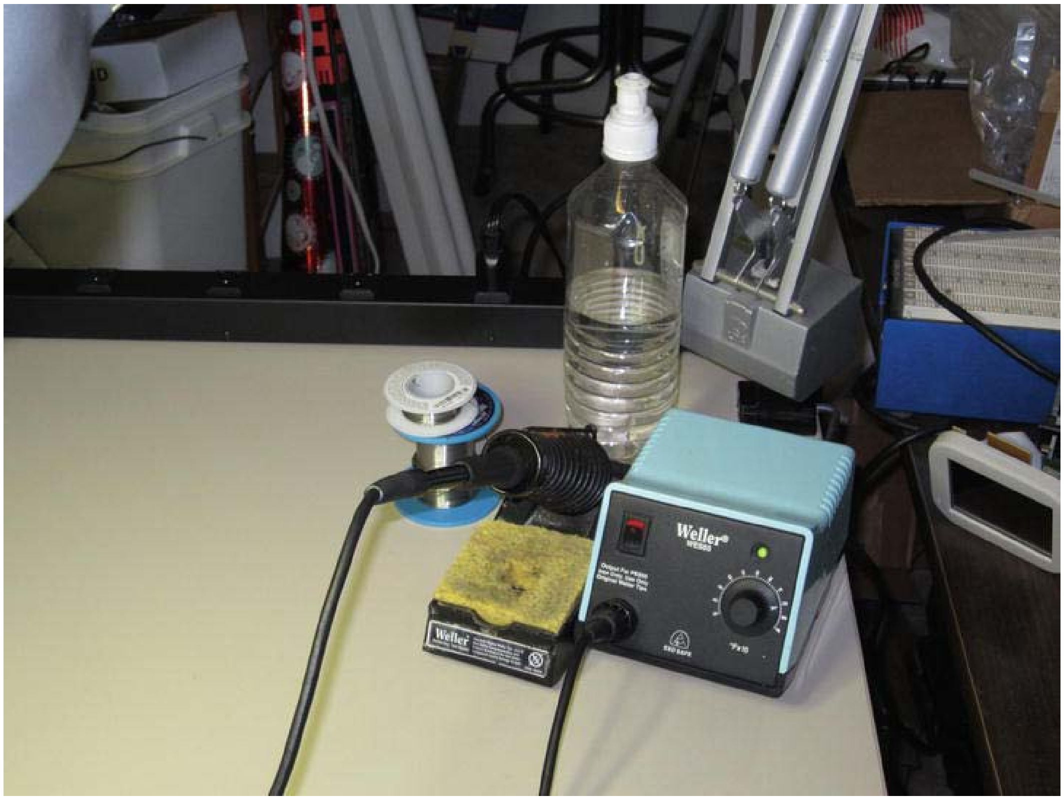

Figure 5-8. *A temperature-controlled soldering station with a stand for the soldering iron, a water bottle to wet the tip-wiping sponge, and some different kinds of solder. Some people prefer to have the soldering station nearer to the front edge of the table. Use whatever configuation makes the most sense in your lab.*

■ **Caution** Soldering irons are hot and can cause serious burns. Do not place a soldering iron where you might accidentally knock it off the work table or brush against it.

Hand Tools and Some Test Equipment

No matter what you work on at the bench, you will most likely discover the need for some common hand tools. When working with electronics, especially, you will often need wire cutters, wire strippers, and small pliers. These tools have been laid out conveniently (for a right-handed person) at the lower-right-hand side of the bench, while the less-often-used hand tools can reside in a box.

A regulated, adjustable bench power supply is also very handy when working with electronics of all kinds. A good, reliable multimeter can also claim a permanent home on the bench, as these will get a lot of use. See Figure 5-9.

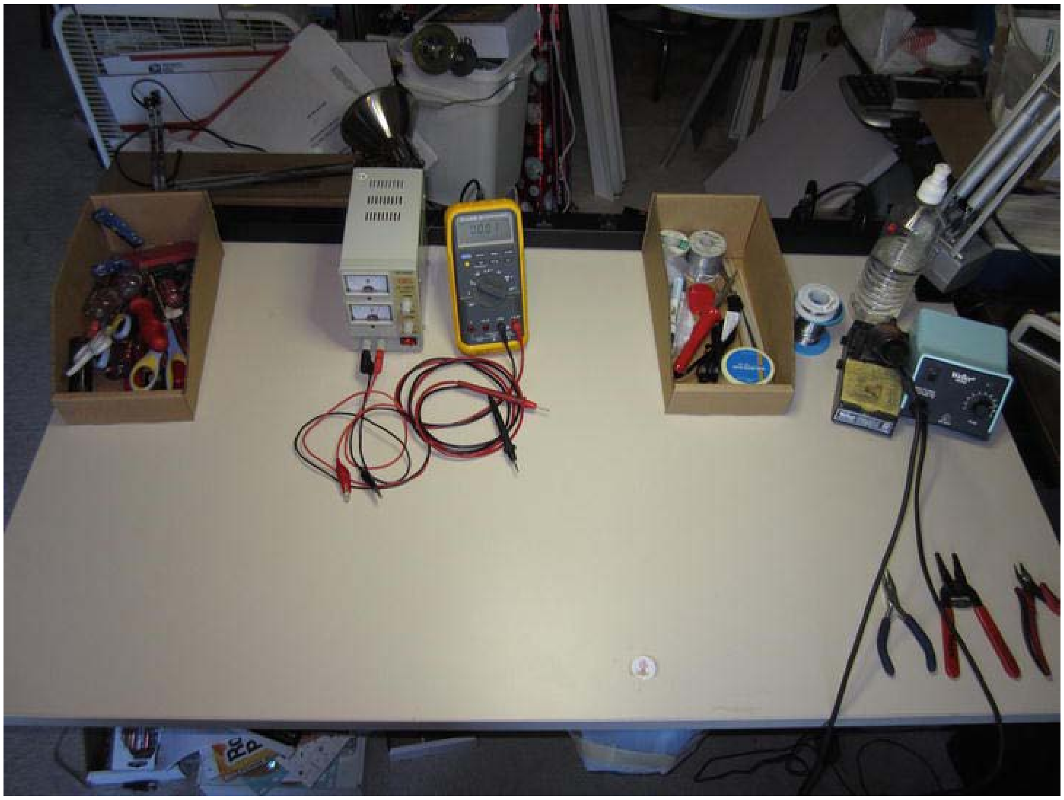

Figure 5-9. *A clean bench with just enough tools to have some fun. Extra soldering supplies, including more kinds of solder , extra soldering iron tips, soldering flux, and an emergency backup soldering iron, sit in a box adjacent to the soldering station. Hand tools and basic test-and-measurement equipment complete the set of "most-useful tools." There's still room for a few more, when you need them.*

Under the Desk

If your workbench has the luxury of drawers, by all means use them! Try to keep your under-desk environment as free of clutter as possible. It's not really a very convenient place to store unused tools or components. You're going to drop the occasional item from time to time, so having to dig out a bunch of junk just to find it is just going to be more frustrating for you. Keep it clean!

You *should* keep a medium-sized trash can under or near your work area, as you will inevitably generate some waste products in this hobby. By minimizing the amount of food-related garbage deposited in your laboratory trash receptacle, you can postpone the invasion of little critters looking for a free meal. Also, bits of wire and old components don't tend to stink after a few days, but a half-eaten sandwich might.

■ **Tip** Copper easily recycles at a premium price. Collect your copper wire snippings and cash them in someday.

A nice addition to the under-the-table accessories you might consider is some low-wattage strip lighting. LEDs are especially good for this. You will appreciate this more once you conduct your first search-and-rescue mission for that teensy-tiny component that dropped off the edge of your table and into the carpet.

Components

Most of our tools are now ready to go. The most frequently used tools are permanently set up, ready for action. The less frequently used tools are carefully stored where we can get to them when we need them, but are not in the way the rest of the time. This is a big improvement over the situation you saw back in Figure 5-1.

Now we need some components with which to work, using our scientifically arrayed tools. Have a look at Figure 5-10 to see one example of how to keep your parts organized yet within easy reach.

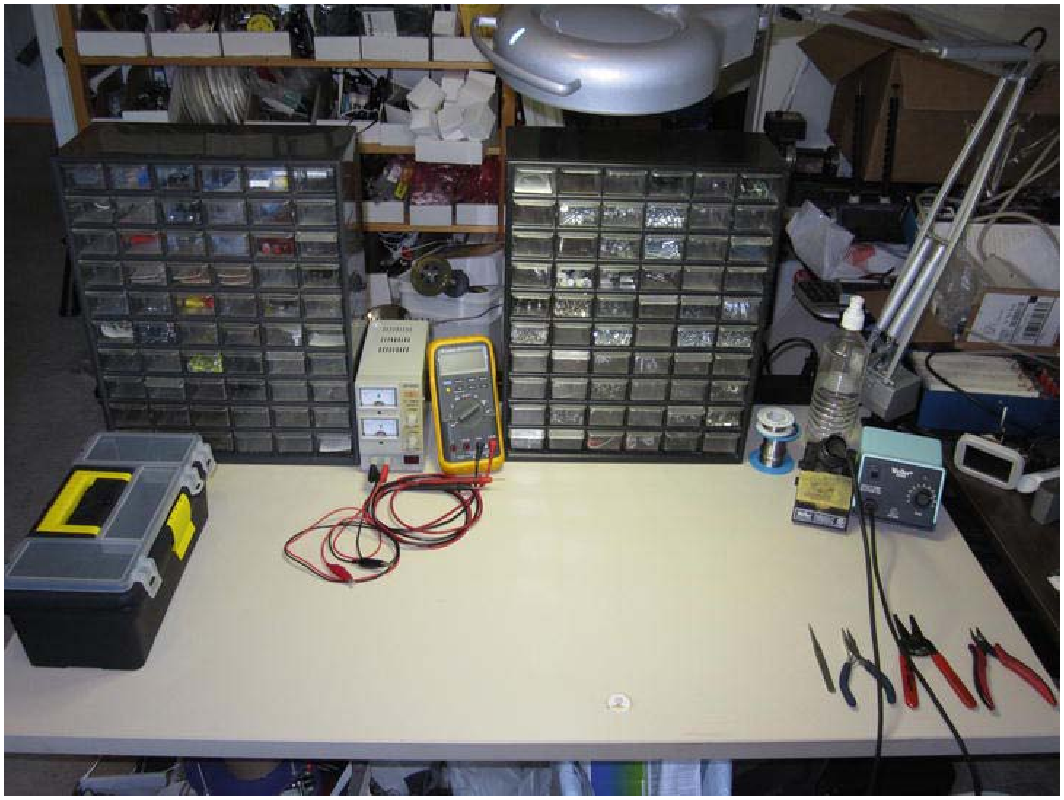

Figure 5-10. *Parts bins are used to organize components. This style of container is available in many sizes with different numbers and sizes of drawers. Even more parts are stored on a shelf behind the work area. The most commonly used hand tools are laid out on the right. The other hand tools are collected in a small toolbox on the left.*

Where to Go from Here

This work space could be further augmented with shelving, both above the work surface and underneath. More lighting wouldn't hurt, either. However, this is plenty to get started. Let's build some electronics projects on this squeaky-clean table and see if we can dirty it up a bit.

Projects

Let's look at a few smaller circuits you can easily build on your tidy, new work area using a solderless breadboard. Expanding on the LED-and-coin-cell experiments in Chapter 4, we'll add a few more components and hopefully start to see how they all interact.

b

There are any number of different types of solderless breadboards available today. A range of them is shown in Figure 5-11.

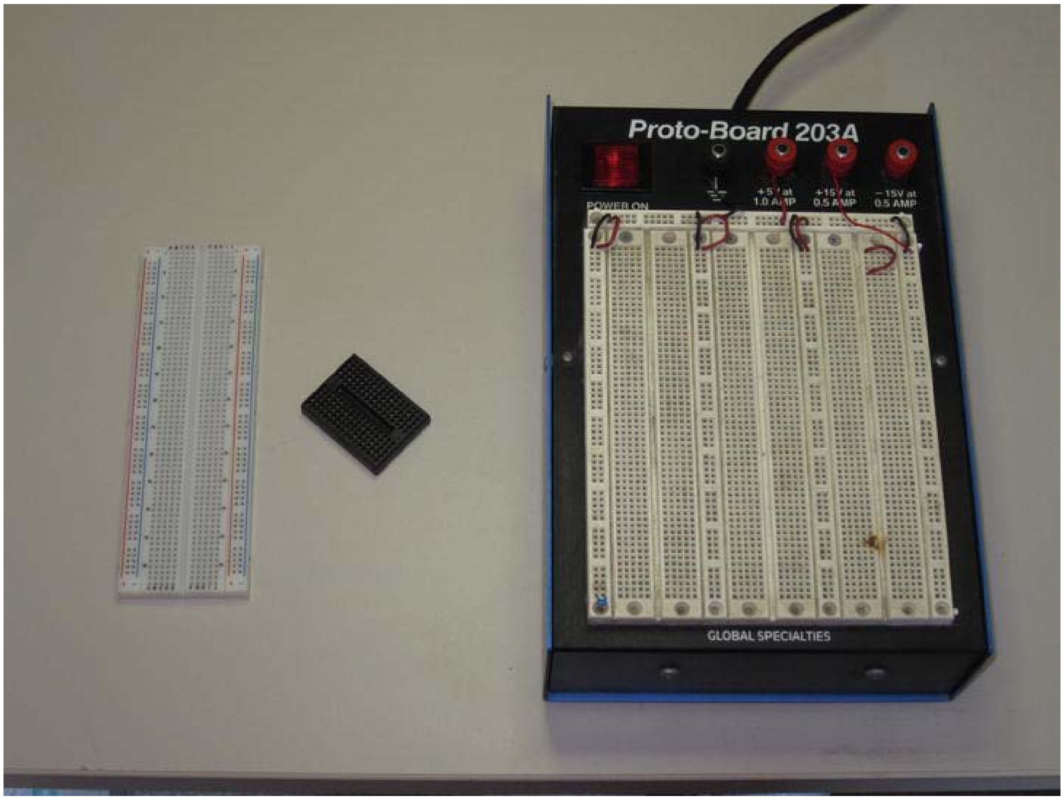

Figure 5-11. *Different sizes of solderless breadboards, all with the same lead pitch and internal connections. The small black unit in the center is perfect for little circuits with only a few small components. The white unit on the left has more tie points and offers power rails along each side. The deluxe, powered unit on the right adds even more space, as well as multiple regulated power supplies for all your tinkering requirements.*

Parallel and Series Circuits

The circuits presented so far have all illustrated the "loop" characteristic of electrical circuits. Electricity generally flows in a circle. It's possible to have more than one path through a circuit, however. What happens then?

Like most good questions, the answer is, "It depends." Let's hook up *two* LEDs at once and try to figure out exactly what's going on. Figure 5-12 shows how it's put together. See Figure 5-13 for the schematic diagram.

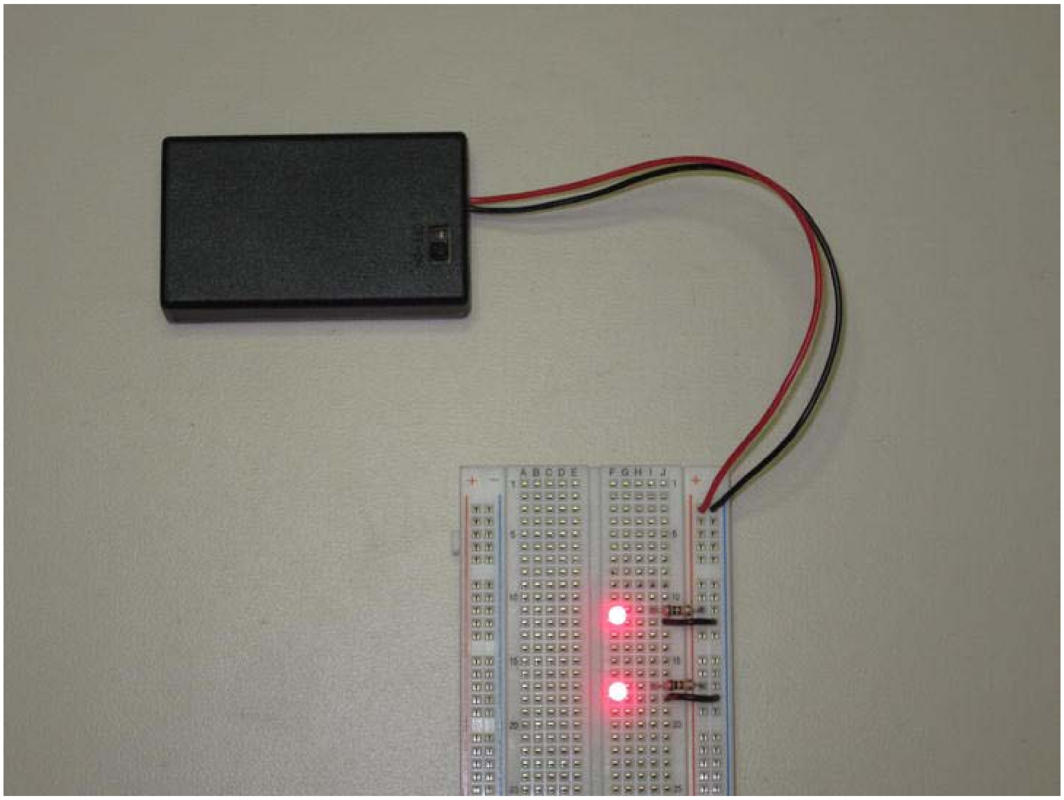

Figure 5-12. *Two LEDs are illuminated at the same time. This simple circuit demonstrates both parallel and series electrical circuits. A 3×AAA battery holder with an integrated power switch is used to provide power to the circuit.*

To build this circuit as shown, you will need the following components:

- A solderless breadboard with power rails
- A 3×AAA battery holder (which holds three AAA cells) and three fresh AAA cells—rechargeable cells are better
- Two LEDs (any visible color will do
- Two 100Ω 1/4W resistors
- Two short jumper wires

Using the letters printed on the breadboard to find the columns across the top and the numbers to find the rows, we can describe each of the tie points using a single letter and a number, such as A-1 or B-2. The following exercise shows how to put it all together.

ASSEMBLY INSTRUCTIONS

1. Install the top LED with its anode (longer lead) in tie point F-11 and the cathode in tie point F-12.
2. Insert one of the resistors from the red power rail on the right at row 11 across to tie point I-11.
3. Connect the right-hand blue power rail at row 12 to tie point J-12 using a short jumper wire.
4. Install another LED just below the top LED, with its anode in tie point F-17 and the cathode in F-18.
5. Install the other resistor from the right-hand power rail at row 17 across to tie point I-17.
6. Use another short jumper wire to connect the blue power rail on the right at row 18 to tie point J-18.
7. Install three fresh AAA cells into the battery holder.
8. Insert the red wire (positive) coming from the battery holder to the red power rail on the right at row 3 (the top).
9. Insert the black wire (negative) from the battery holder to the blue power rail on the right at row 3.
10. If your battery holder has a power switch, turn it on now.

The circuit is now complete. Please verify that both LEDs are illuminated, as shown in Figure 5-12.

Troubleshooting

If both LEDs are not shining at this point, you're going to have to do some troubleshooting.

- The first thing to do is double-check your wiring. Does your breadboard look like the one in Figure 5-12?
- Next, make sure you're using good batteries. If you have a voltmeter, you can test the batteries by measuring their voltage. New alkaline or carbon zinc cells should read 1.5V or more. Rechargeable nickel-cadmium (NiCd) or nickel–metal hydride (NiMH) batteries should read around 1.2V each when completely charged.
- Are the cells properly installed in the holder? Are they in the proper orientation and firmly seated?

- Is the power switch (if available) not in the On position? This happens more frequently than anyone would like to admit. Wiggle the switch back and forth and see if it's only making intermittent connection.
- Check the orientation of the LEDs. They will only work if installed in the correct orientation. It only takes a second to unplug them, swap them around, and try them the other way. At these low voltages, it won't hurt them to be installed backward.
- Are your resistors the right value? The 100Ω value is not critical. Anything from 100Ω to 1KΩ will do nicely. Don't use smaller values, as this will allow too much current to flow through the circuit, perhaps damaging the LEDs. How much current? We'll figure that out in just a bit. Larger-value resistors will decrease the amount of light emitted by the LEDs, down to the point where you can't see them anymore.
- Are you sure your LEDs emit visible light? Millions of infrared LEDs are manufactured every month for use in remote controls and night-vision gear. This light is not visible with the human eye, although some digital cameras can see it. Just because it looks like an LED doesn't mean it *is* an LED.

A Different View of the Same Thing

Your version of the circuit in Figure 5-12 might look a little different, and that's just fine. The important point is that both LEDs are lit up at the same time when you apply power.

Let's look at the traditional schematic for this circuit to get an idea of what's going on here. See Figure 5-13.

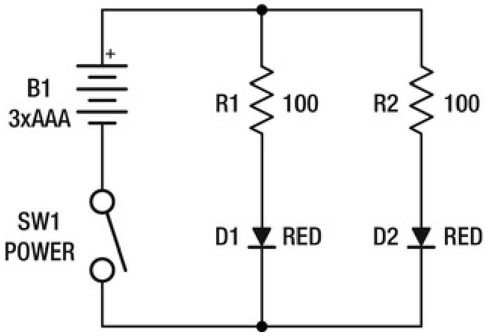

Figure 5-13. *The traditional schematic diagram of the two-LED circuit. This diagram should convey the very basic information needed to build the circuit, as well as understand the important aspects of how the circuit is supposed to work, without going into much detail at all about the actual implementation or assembly of the components.*

By now you should be able to look at the schematic and see the loop formed by the circuit. The only new twist is that there are *two* paths along which the electricity can flow, either through R1 and D1 (the top LED on the breadboard) or through R2 and D2 (the bottom LED, or the initials of a certain adorable robot from the movies).

The three AAA cells in the battery holder are shown as B1 in the diagram. The power switch, labeled SW1, is shown as a separate component to remind us that the circuit can easily be broken, halting the flow of electricity, which is something we want to happen from time to time. The connections between the two resistors and the two LEDs should be somewhat obvious. I'll discuss the details for the parallel and series connections momentarily.

Remember that an electronic schematic diagram is an abstract representation, boiled down to only the essential items involved and the relationships among them. As such, it doesn't need to resemble the final product, but it could. See Figure 5-14.

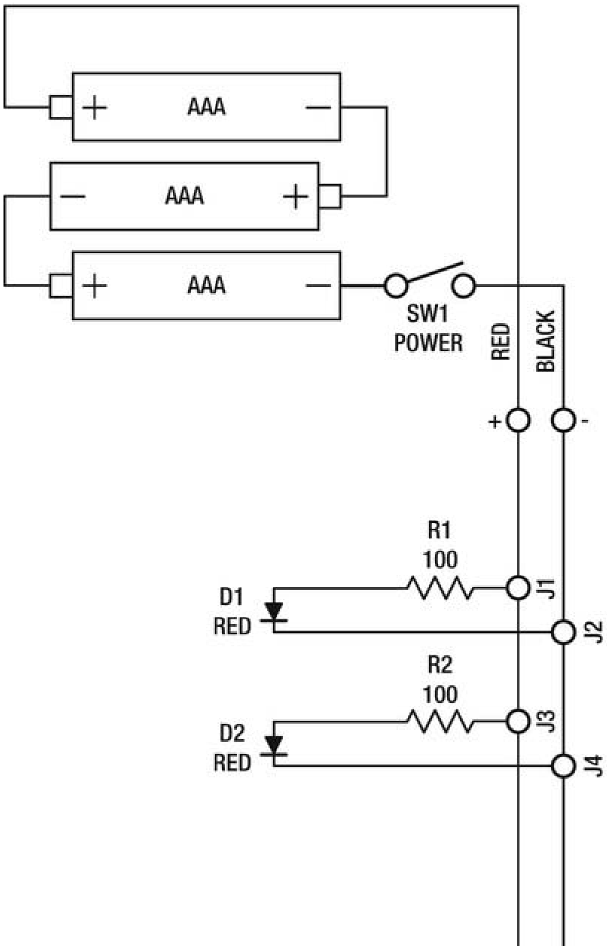

Figure 5-14. *Another schematic diagram of the two-LED circuit, showing the approximate physical layout of the components on the solderless breadboard. This style of schematic is often used when describing how to actually build a circuit, showing the constraints of the actual, physical components and the connections between them. It conveys more specific information than the previous example in Figure 5-13.*

Batteries in Series and Parallel

Let's look at the batteries first. Technically, as mentioned in Chapter 1, there is one *battery* in this circuit, and it happens to be composed of three *cells*. The term *battery* is very often used interchangeably with *cells*, and the distinction is minor. It's better to avoid confusion than it is to be exactly, precisely *correct*—usually.

We see three cells sitting side by side in both the schematic in Figure 5-14 and the actual battery holder shown in Figure 5-12. The cells are connected *in series*, meaning that one connects to another that connects to another, and so forth. They are like beads on a single wire. You should be able to trace the wiring (the solid lines) that connects the individual cells in the schematic.

Note the polarity of each of the cells within the battery. The positive terminal of the bottom-most cell is wired to the negative terminal of the center cell. The positive terminal of the center cell is connected to the negative terminal of the topmost cell.

The negative terminal of the bottom-most cell is connected to one side of an *SPST (single-pole, single-throw)* switch. The other side of the switch emerges from the battery holder's body as the black, or negative, lead from the battery pack. The positive terminal of the topmost cell connects directly to the red, or positive, lead coming out of the battery pack.

The switch could have just as easily been installed in the positive lead, and it would have worked exactly the same. The battery holder used in the photograph happens to switch the negative lead, as was revealed by examining the internal wiring using a multimeter as a continuity tester.

When the switch is *open* (i.e., in the Off position), the circle is broken and no electricity is going to flow. When the switch is *closed*, the circuit is complete and power is made available to the remainder of the LED circuit, assuming that it is wired up correctly.

The amount of *voltage* being supplied by the battery holder depends on the electrical characteristics of the individual cells. If alkaline batteries are used, each cell will provide 1.5V, assuming that the cells are relatively fresh. Rechargeable NiCd or NiMH cells will provide around 1.2V when fully charged.

When batteries like these are connected in series, the total voltage available is the sum of each of the cells. Just add all the voltages together. For example, using fresh alkaline batteries, the total voltage would be 3 × 1.5V = 4.5V. Rechargeable cells would provide 3 × 1.2V = 3.6V when fully charged.

If the cells were wired *in parallel*, with all of the positive terminals being connected together and all of the negative terminals wired together, the total output voltage would only be 1.5V for alkaline cells. The total output *current*, however, would be triple the capacity of each individual cell. See Figure 5-15 for a comparison of series, parallel, and combination circuits using batteries.

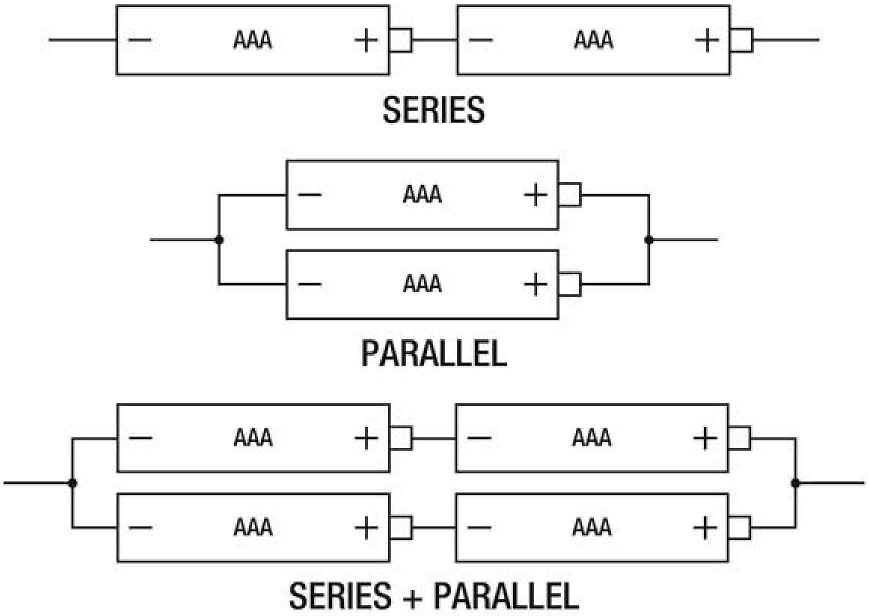

Figure 5-15. *Different arrangements of the wiring of the batteries produce different results. The top two cells are arranged in series. Their respective voltages are added together. The center circuit consists of two cells arranged in parallel. This configuration doubles the current that can be delivered. The bottom drawing shows a combination of both series and parallel circuits, doubling both the voltage and the current.*

When connected in series, the total current output is the same as the current capacity of each individual cell.

In summary, if you need more voltage, stack the cells end to end. If you need more current, stack them side by side. Even more complex arrangements are possible, using both parallel and series circuits to achieve the desired power output characteristics. This practice is very common in hobby model aircraft and battery-powered vehicles. A designation such as 3S2P means that there are two parallel circuits (2P) of three cells in series (3S).

Looking back at the circuit in Figures 5-13 and 5-14, we see that the “battery” used in this circuit consists of three cells in series. This means that if alkaline batteries of 1.5V each are used, the total voltage available to the circuit is 4.5V. If typical rechargeable cells (NiCd or NiMH) are used, we can expect 3.6V to flow through the circuit.

Switches in Series

The power switch, labeled SW1 in the schematic, is *in series* with the rest of the circuit. As mentioned in the previous section, when this switch is open, no current flows through the circuit at all. The switch is intentionally placed at this location in this circuit to act as a bottleneck. When the power switch is turned off, no power flows through the circuit. That seems like a good idea, considering the desired operational goal.

The schematic symbol for SW1 shows a pretty good likeness to the internal wiring of any simple SPST switch. There are two terminals, and a conductor that can be moved to either electrically connect them or electrically isolate them. As drawn, the switch is in the open position.

If we added another identical switch in series with the existing switch, what would happen? Take a look at Figure 5-16.

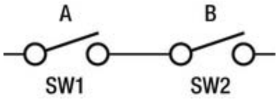

Figure 5-16. *Two switches in series. Both switches must be closed for electricity to flow through this part of the circuit. If either switch is open, the circuit is broken and no current will flow.*

Both switch A (SW1) and switch B (SW2) must be moved to the closed position for any electricity to flow through the circuit. If either switch is opened, then no current will flow. This is sometimes referred to as an *AND* circuit. Both A *and* B must be on for the circuit to operate.

A circuit just like this is often used in security systems. Switches are attached to doors and windows. All of the switches are wired in series, forming a large loop. The switches are all configured to be closed when the doors and windows are closed, and to open when doors or windows are opened.

If all the doors and windows are closed, then the loop of switches will allow the flow of electricity. Although it sounds exactly *backward*, when electricity is flowing in the circuit, this keeps the security alarm *turned off*. When any one of the switches is opened, the circuit is broken. The alarm then sounds.

Most real security systems are more complex than this, of course. The main advantage of this particular configuration is that it is easy to install and requires the least amount of wiring. The chief disadvantage is that when the alarm is triggered, you have no idea exactly *which switch* caused it.

Switches in Parallel

Just like with batteries, switches behave differently when configured in a different manner. Let's connect two switches *in parallel* this time and see what happens. See Figure 5-17.

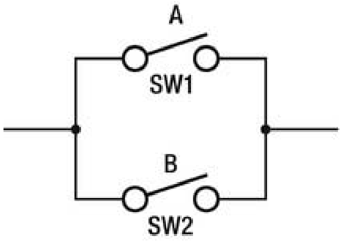

Figure 5-17. *Two SPST switches in parallel. If either one of the switches is closed, then power can flow through this part of the circuit. Both switches must be opened to stop the flow of current.*

The circuit in Figure 5-17 will conduct electricity if switch A (SW1) *or* switch B (SW2) are closed. To stop the flow of current, you have to open both switches. This can be used as a simple *OR* circuit. If switch A *or* switch B is closed, then power can flow through the circuit.

Looking back at the schematics in Figures 5-13 and 5-14, we see that SW1 is in series with the rest of the circuit. This makes it a very effective way to turn power on and off from a single location. Handy!

The Three-Way Switch

You might think that the circuit shown in Figure 5-17 could be used to turn lights on and off from two different locations. You would be partly correct. You could turn lights *on* from either switch location, since flipping either switch to On will cause current to flow and lights (or other appliances) to operate.

Unfortunately, you would not be able to turn *off* the light from either switch if the other switch were on. So this configuration is good for turning lights on but not off. A slightly more complex arrangement is required to do both.

We will need two switches, as before, but they will need to have a little more functionality to accomplish what we've set out to do. Instead of a switch that's either open or closed, we need a switch that alternates between two different routes, depending on how it is positioned. This type of switch is called a *double-throw* switch, as it connects a *common* terminal to either one of two different terminals. It can be *thrown* in two different directions. This should be much easier to understand when looking at a schematic symbol. See Figure 5-18.

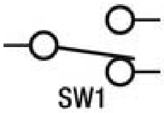

Figure 5-18. *A single-pole, double-throw (SPDT) switch. The common terminal on the left side is connected to either the top or bottom terminal on the right. Variations of this type of switch have a neutral position in the middle. Push-button (i.e., momentary contact) versions of this switch will label the two right-side terminals as either "normally open" or "normally closed," depending on the physical configuration of the switch.*

Now comes the clever bit. We'll connect a pair of these switches together by tying the two non-common terminals on one switch to their counterparts on the other switch. This results in four possible combinations of switch positions (both down, one up, the other up, and both up). Two possibilities turn the light on, and the other two turn the light off. The important point is that the "on" or "off" state of the entire circuit can be changed from either location, independently of the state of the other switch. Problem solved! Again, this should be easier to see when presented as a diagram. See Figure 5-19.

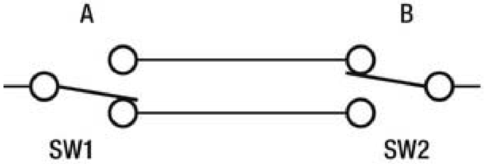

Figure 5-19. *A three-way switch circuit built using two SPDT switches. Power is either carried by the upper wire or the lower wire, as long as both switches are in the same position. If not, then no power flows across this part of the circuit. As shown, the three-way switch is effectively off. Flip either switch to turn it on.*

If you ask a real electrician how to wire a three-way switch, you'll get several answers, or more likely, you'll get, "It depends." That's because *it really does* depend on the actual wiring situation. Sometimes the power is routed through one or the other of the two switches, and sometimes the power is routed through the load. Each of these scenarios requires a different routing of wires and connections.

You are strongly advised to consult a real electrician concerning wiring in your house, where your family, pets, and all your stuff live. Less-than-perfect wiring causes fires or electrocution, which are both very unpleasant ways to die.

■ **Caution** This book does *not* teach you how to wire electricity in your home. Consult a professional.

Resistors in Series

Just like batteries and switches, resistors perform different functions in a circuit depending on their configuration. Let's look at using resistors *in series* first, and then consider using them in parallel later. Both combinations have their uses, as you will see.

Let's look at just a small part of the schematic that we've been working with so far in this chapter. See Figure 5-20.

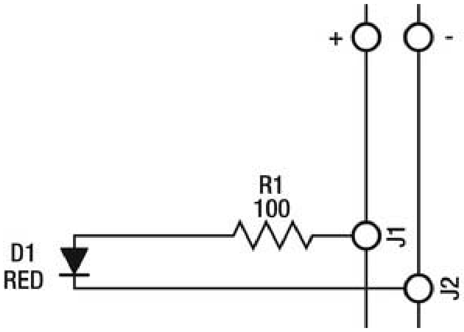

Figure 5-20. *One part of the circuit showing an LED, a resistor, and the power rails from our solderless breadboard. The resistor and LED are connected in series. The battery and power switch are not shown.*

LED Curiosities

D1 is an LED. It is a polarized component and will only work when installed in the proper orientation. The schematic indicates a red LED should be used, but feel free to use whatever color LED you like or happen to have available. The triangle shape of the schematic symbol helps remind us of the direction of *conventional current* through the device. The actual flow of electrons is the other way, but that's only because electrons are negatively charged particles and do everything backward. Apparently this was not well understood when electricity was first being tamed for our use.

LEDs are fascinating devices, emitting light in a wide range of colors, coming in an astounding array of shapes and sizes, and generally requiring very little power to operate. They are almost as easy to use as incandescent bulbs and are certainly more power efficient.

Being semiconductors, however (which leads to their being semimagical in their inner workings), LEDs generally only operate correctly within tightly bounded parameters. For example, a typical LED will only need between 2 and 3V to shine quite brightly. Incandescent lightbulbs, on the other hand, can operate from less than 1V to many hundreds of volts, depending on their construction and intended purpose.

While LEDs have a minimum *forward voltage* requirement that must be met for them to work at all, the brightness of an LED is determined solely by the amount of current flowing through it.

LEDs Have No Self-Control

In addition, LEDs have no internal, inherent mechanism to limit the amount of current flowing through them. Incandescent bulbs, when cold, have a low internal resistance, allowing much current to flow through them when connected to the right power supply. Remember that less resistance means more current can flow, and vice versa. This resistance almost instantly goes way up as the filament of the bulb glows white hot. The filament itself then becomes the limiting factor in how much current flows through the bulb, usually stabilizing at the proper power level.

LEDs have no such sense of self-control. They will conduct as much current though themselves as you are willing to provide, up to and beyond the point where the internal heat generated by the small amount of resistance they do possess causes them to literally burn up. In general, it's usually the tiny, almost microscopic internal bond wires that fizzle out, not the actual LED chip itself. However, dead is dead, and LEDs are generally considered to be "unrepairable" subassemblies. (Believe it or not, that's the correct technical term, even though it's not the correct English word—*irreparable* is the correct term when speaking of non-electrical things).

Care and Feeding of LEDs

The simplest possible solution to this ~~problem~~ opportunity for excellence is the use of a *current-limiting resistor* in series with the LED. Recall that a resistor will resist the flow of current in a circuit. The resistor used in this circuit, R1, is indicated to exhibit a resistance of 100Ω.

Let's say that D1, the red LED in our circuit, requires 2V in normal operation. This is pretty common for red LEDs. Generally speaking, the forward voltage requirement for LEDs goes up as you get farther up the color spectrum. Blue LEDs, for example, often require 3.0V to 3.6V to operate.

If we're using rechargeable cells in our prototype (and we really should be, you know), we can expect 3.6V to be made available to the resistor-plus-LED loop within our circuit. If 2V is needed for the LED, where are the rest of those volts going?

The answer is that they are being used to heat up the current-limiting resistor R1. Resistors turn current into heat. That's what they do. They do it, however, in a very precise and predictable way, which makes it easy for us to calculate how much current, voltage, and heat will be churning around in our circuits.

■ **Caution** Third-grade arithmetic ahead: Addition, subtraction, multiplication, and (shudder) division. You have been warned.

If we started with 3.6V, and two of those volts are being used by the LED, then our good friend Subtraction tells us that we've got 1.6V flowing through R1. For a thorough explanation of why this happens the way it does, feel free to look up Kirchhoff's Voltage Law. It basically says that all the voltages in a circuit end up being equal. The voltage across the resistor and LED will equal the battery voltage. The voltages within the LED-in-series-with-resistor circuit add up, just like battery voltages add up in series.

If we know the voltage across R1 (1.6V) and the resistance of R1 (100Ω), then we can use Ohm's Law to calculate the current flowing through R1. Reviewing what you learned in Chapter 3 about the predictable and inviolable relationships between voltage, resistance, and current, you can see that I = E / R (i.e., current equals voltage divided by resistance), where I is the current in amps, E is the voltage in volts. and R is the resistance in ohms.

It's not that hard! It sounds much worse than it actually is. The answer is 0.016A, or 16mA (1.6V/100Ω). That's the teeny-tiny amount of current needed to light up a typical red LED. Most small LEDs can handle up to 25mA safely, and some can handle much more than that, especially if they are intended for illumination purposes.

One quick side trip to multiplication land and we'll know everything we need to know about R1 in this circuit. While the schematic specified the resistance of R1 as 100Ω (the primary characteristic or component *value*), nothing was said about the power-handling capability of this component.

Power is calculated as voltage multiplied by current. That's all there is to it. We know the voltage: 1.6V. We know the current: 16mA. We multiply them together to get 0.0256W (watts), or almost 26mW. Remember that you have to enter the current as 0.016 because the formula calls for the number of *amps*, not milliamps.

Not quite 26mA is not quite much of anything. You will not be able to tell that the resistor is shedding all those extra volts as heat with your fingertip. It might not even register with a sensitive thermometer. It's just a really tiny amount of power. A typical small resistor (a 1/4W resistor) will handle 250mW of power, which is almost ten times what is needed in this application.

■ **Tip** It's always a good idea to calculate the power requirements of your components, just so you're not *too* surprised when the magic smoke leaves the circuit, rendering it crispy and nonfunctional.

OK, we're back from the frightening land of multiplication. It wasn't *that* bad was it? The remainder of understanding resistors in series depends only upon the science of addition. Counting on your fingers will be good enough.

We can change the brightness of the LED by changing the amount of current flowing through it. We can change the amount of current flowing through the LED by either raising the supply voltage or by changing the resistance of R1. Since we can easily substitute any other value of resistor for the 100Ω resistor with which we began, this is easy to do. Just don't go *too low* in resistor value, as this will allow *too much* current to flow through the LED, possibly damaging it permanently.

Adding resistors in series adds their resistances together. For example, if you placed two 100Ω resistors end to end, the total resistance across them would be 200Ω. That's really all there is to it: 100Ω + 100Ω = 200Ω.

If you perform this experiment with the LED circuit we built on the breadboard, you can see that the LED will not shine as brightly as it did before. This is because the resistance has doubled, which means, all other factors remaining the same, that the current through the LED has been cut in half. Half the current results in half the brightness, since LEDs are controlled by the amount of current flowing through them.

What's interesting about this arrangement is that the power-handling capacity of the two resistors is *also* double what it was when only one resistor was in the circuit. This is because each resistor is now dissipating half the power when compared with the previous scenario. If two 1/4W resistors are used in series, they can together safely dissipate up to 1/2W of power, although pushing any component to 100 percent of its capacity is just asking for trouble, especially when heat is concerned.

By connecting resistors end to end, you add their resistances together, as well as their power-handling capabilities. By keeping a small variety of basic values on hand (e.g., 1Ω, 10Ω, 100Ω, 1KΩ, 10KΩ, etc.), you can easily stack up the right combination needed for almost any electronic application. This is the method used in the very handy device known as a *decade resistor* or *decade box*, which has dials for each digit in the desired resistance, allowing you to dial up any value of resistor needed.

Extra credit: If you've followed along this far, consider researching voltage dividers built using only resistors.

Resistors in Parallel

What happens when resistors are connected side by side instead of end to end? That's where things get interesting, at least from a mathematical point of view.

An easy way to look at parallel resistors is to think about the amount of current flowing through each one, and then add them all up. See Figure 5-21.

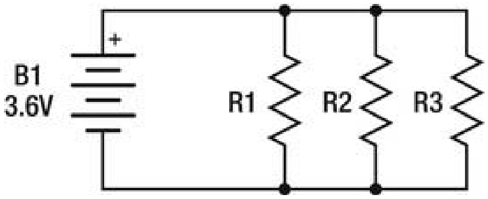

Figure 5-21. *Parallel resistors each conduct a portion of the total current flow through the circuit. Each resistor allows a certain amount of current to flow. The sum of their currents can be calculated if their individual resistances are known.*

Let's assume that R1, R2, and R3 have resistances of 100Ω, 200Ω, and 300Ω, respectively. The total resistance of the circuit across the battery terminals is *not* 600Ω, as would be the case if they were in series instead of parallel. The total resistance is less than 100Ω, which is the least of the resistors. How can this be?

First, let's think in terms of how much current is flowing in each of the three parallel pathways within this circuit. If B1 is putting out 3.6V, and R1 is 100Ω, then we know to use Ohm's Law to calculate the current, with I = E / R. In this case, I = 3.6V / 100Ω = 0.036A, or 36mA. *In addition*, the path through R2 is drawing half that (because R2 is double the value of R1—get it?), or 18 mA. *In addition*, the final pathway is drawing one-third as much current as the first one, or 12 mA. Just to verify: I = 3.6 / 300Ω = 0.012A, or 12mA. Everything checks out.

Now add all those currents together: 36mA + 18mA + 12mA = 66mA. That's the total current, using this method. Again, since we know both the voltage and the current, we can find the *effective* or *equivalent resistance* of all three resistors in parallel, as if they were one single resistor. Rearranging the formula for Ohm's Law, we find that R = E / I. In our case, R = 3.6 / 0.066A = ~54.55Ω. See? Smaller than 100Ω, somehow.

Thinking about the individual *currents* flowing through each resistor is the same as adding the current-supplying capacity of battery cells together when arrayed in parallel. Just add them together. However, we had to calculate each current separately and then add them together. There is a more direct way, depending upon how you look at things, but beware: it involves not only *fractions*, but *fractions upon fractions.* Oh, the horror.

■ **Note** Some people don't dig math. Weird, but true.

Here is the equation for calculating the total resistance (R_T) of three resistors in parallel (R_1, R_2, R_3):

$$R_T = \frac{1}{\left(\frac{1}{R_1}\right) + \left(\frac{1}{R_2}\right) + \left(\frac{1}{R_3}\right)}$$

The good thing about this equation is that you don't need to know what the supply voltage is to get the correct answer. All you need to know are the values of the resistors. If you need to use more resistors, just add more terms to the equation.

Calculating the value of parallel resistors occurs quite frequently when designing electronics. Some special cases are simple enough to do in your head. For example, if two identical-value resistors are placed in parallel, their combined resistance is half the value of each individual resistor. That is, if you placed two 100Ω ohm resistors in parallel, their effective resistance would be 50Ω. Similarly, if you have three identical resistors, the net resistance is one-third the value of each constituent resistor. Make sense? These are special cases of the general equation just shown.

This happens so much that an alternate notation method has been developed: $R_T = R_1|R_2|R_3$. This is not often used for hobby electronics, however.

The power-handling ratings of resistors are added together, so if you have two 10W resistors in parallel, the net power-handling capacity is now 20W. At least that part was easy.

Capacitors in Series and Parallel

Capacitors are completely different from resistors. When calculating their values in series and in parallel, you have to use *exactly the opposite* methods. Isn't electronics interesting?

The values of capacitors *in parallel* are added together. This makes them much more like batteries in parallel than resistors. However, you still have to use the fractions-upon-fractions method to calculate the value of capacitors in series.

Again, some shortcuts are available. Two identical capacitors in series have *half* the capacitance of each individual capacitor, but *double* the working voltage, which is the maximum voltage they were designed to withstand before failing.

Inductors in Series and Parallel

It gets more interesting, believe it or not. Inductors act like resistors as far as adding their values together in series and taking the reciprocal of the sum of the reciprocals of their values in parallel, but *only* when the magnetic fields of all of the inductors involved do not interfere with each other. If the magnetic fields generated when current flows though the inductors overlap, then you have to factor in their *mutual inductance*, which starts to get complicated.

Extra credit: Figure out how all that works and let me know.

LEDs in Series and Parallel

You learned (or were reminded of) a few of the basic characteristics of LEDs in the previous section, although it was really to help illustrate the operation of resistors in series and parallel. A couple of major oversimplifications were introduced to help concentrate on the resistor. We'll address a couple of these issues here.

Lighting up a single LED using a battery and a fixed current-limiting resistor is fun, but it's pretty basic stuff. Lighting up a *lot* of LEDs is where it gets to be a *lot* of fun. To do so, you need to learn a little more about the weirdness that is *semiconductor*.

LEDs are diodes (*LED* stands for *light-emitting diode*, after all). Diodes are semiconductors and are made by elves in a hollow tree. Semiconductors do not have nice linear voltage-to-current curves.

For the purposes of simplicity, in the previous section, it was suggested that LEDs have a certain minimum voltage that they need to operate. This is called their *forward voltage*, and is sometimes referred to as their *voltage drop* within a circuit. So far, this is all true. You can find an LED's forward voltage specification in the data sheet published by the manufacturer.

What you'll find, when you look, is that every manufacturer only publishes a *range* of voltages, such as 1.8V–2.2V, giving a minimum and a maximum voltage. This is what they are contractually obligated to deliver in their product. Anything outside of this range is considered *defective*, or at least out of tolerance. You might get a good deal on "rejects" that fall outside these parameters but still light up quite nicely. Such is the provenance of many of the penny LEDs available from overseas. Yes, they exist. And yes, you get what you pay for.

There are two reasons why LED manufacturers are being so deliberately vague in their specifications. The first is that process variations are going to produce a certain randomness to these values. One batch might be identical to the next batch, but then the third batch might be all over the place. Such things happen.

The second reason is where the real weirdness begins. Forward voltage is not a fixed value in any given device, but instead increases as the current increases. Again, this increase is *nonlinear* and cannot be calculated with a simple formula.

The best way to demonstrate this to yourself is to connect an LED to a power source via a variable resistor, such as a potentiometer. For the protection of your LED, add a fixed resistor in series with the potentiometer so that you establish a maximum current level, even when the potentiometer is turned all the way up (i.e., has almost zero resistance).

Now apply power and adjust the potentiometer until the LED just barely lights up. Now measure the voltage across the terminals of the LED, using a voltmeter. It should be in the neighborhood of 2V–3V. For extra credit, measure the current flowing *through* the LED, using an ammeter. This measurement will most likely be less than 1mA.

Compare these readings to what you find when the LED is at full brightness. The forward voltage across the LED goes up as the current goes up. To map the nonlinearity of this function, you would need to record many readings of both voltage and current, and then plot them in a graph.

Just as you can wire resistors, capacitors, and inductors in series and in parallel, you can do the same with LEDs. As with resistors, the forward voltage of the LEDs are added together when arranged in a series circuit. This raises the minimum voltage required to light up a string of LEDs in series. This means that you can't light up ten LEDs in a single string using a 9V battery. You have to first overcome the minimum forward voltage just to get any current at all flowing. Then you have to up the current to get the desired brightness, which simultaneously raises the total forward voltage through the circuit. You still need a current-limiting resistor in this circuit—but only one is needed. You don't have to have one for each LED.

The amount of current flowing through the string of LEDs is determined by the supply voltage, the cumulative forward voltage of all the LEDs, and the resistance of the current-limiting resistor. The same amount of current is flowing in every part of the circuit. If 25mA is being drawn from the power supply to light up your single string of LEDs, that means that each and every LED has 25mA of juice flowing through it—more than enough to light them up quite nicely.

If you have a 12V power supply (or a car battery), you could string up to five or maybe six red LEDs in series, with a small current-limiting resistor. That's because each red LED will typically have a forward voltage between 1.8V and 2.2V. Other colors higher up the rainbow might need more voltage per LED, meaning fewer can be lit at once using a single string of LEDs in series.

But don't fret. You can light up as many LEDs as you want by placing multiple strings of LEDs in parallel with each other. Each additional string of LEDs adds to the total current draw, so you will eventually find a limit along that dimension as well. For an example of a small LED array using both parallel and series circuits, see Figure 5-22. The schematic is shown in Figure 5-23.

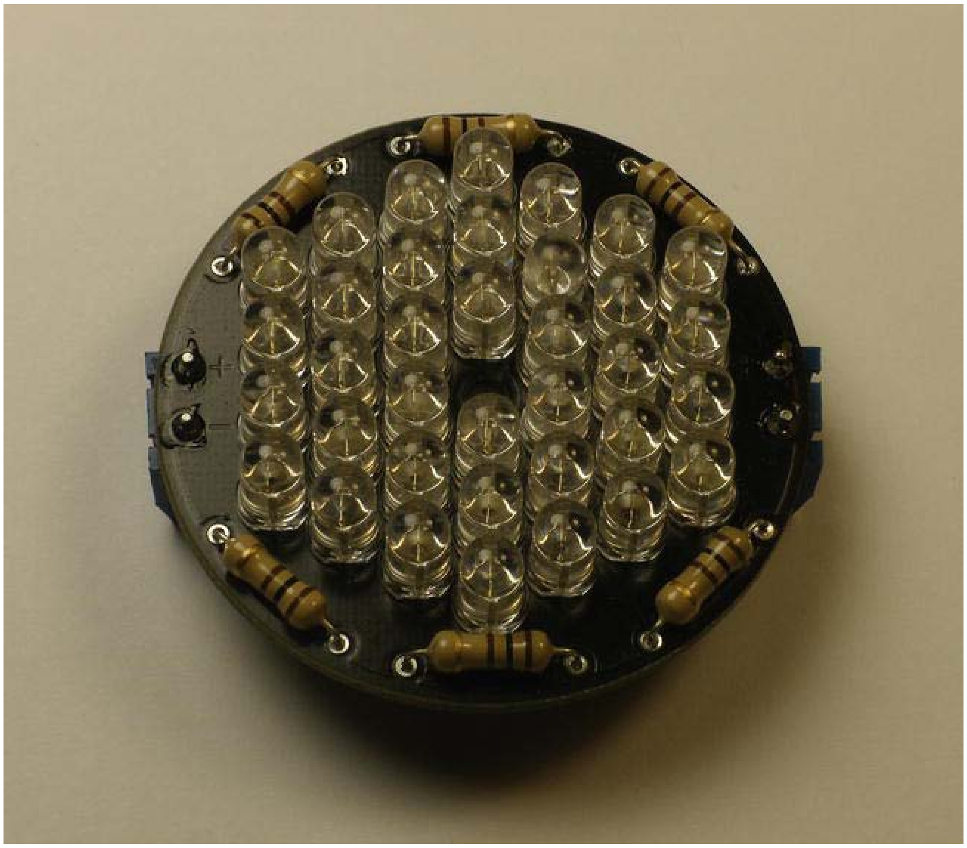

Figure 5-22. *An array of six parallel strings of six infrared LEDs in series. These infrared spotlights are used for covert surveillance and security applications.*

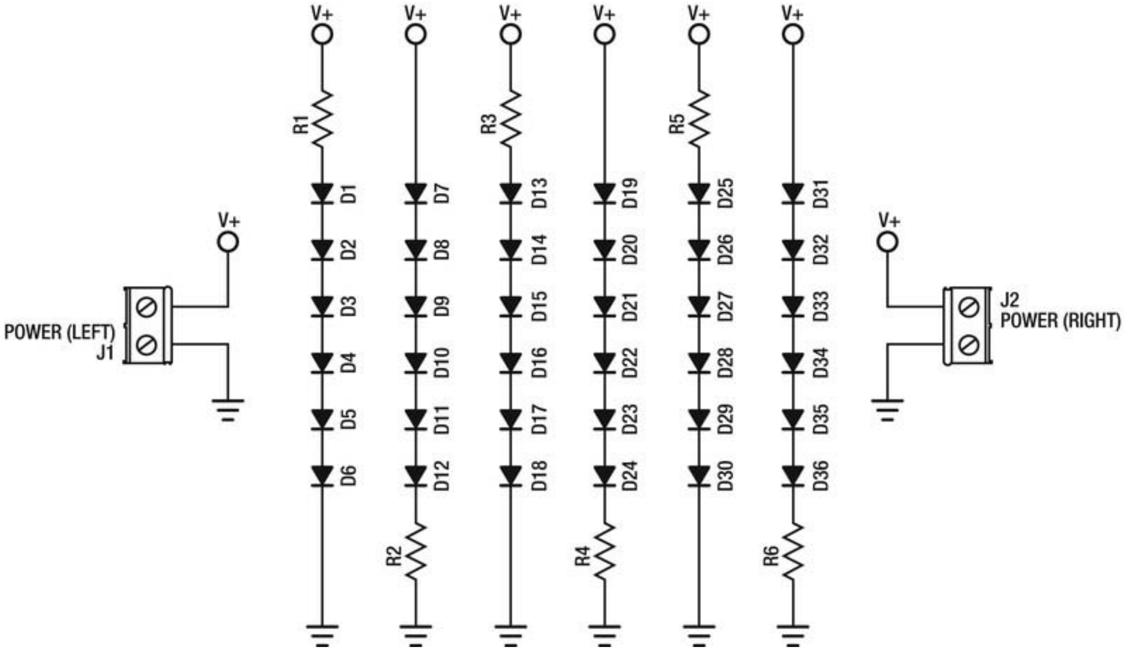

Figure 5-23. *The schematic of the LED array, showing both parallel and series circuits. The circuit offers two power connectors to make it easy to wire multiple LED arrays using the same power supply.*

Summary

Even if your workbench doesn't look as bad as the one presented at the beginning of this chapter (or even if it looks *worse*), hopefully you've learned a thing or two about the importance of they saying, "A place for everything and everything in its place." Since *you* get to decide where everything goes, you can either make it easy or hard on yourself as the years go by.

The workspace in this chapter extended to a total of 10 square feet. If this is what you have to work with, you can certainly set up a comfortable, safe, and productive area for your electronic projects. If you have even more space available (adding a shelf either above or below the desk is a great way to do this), then even better.

Hopefully some of the more detailed descriptions of how electronic components can be wired up have given you some ideas for projects of your own. The very best possible way to learn more about electronics is to think up some simple experiments and try to build them in your lab. Have a goal, take lots of notes, and have some fun—even your failures can be used as stepping stones to more ambitious and creative projects in the future.

Have a *lot* more room? Then perhaps you should be thinking about setting up a group lab or classroom. Chapter 6 will give you some helpful advice about what you are going to need to get started. See you there!

CHAPTER 6

The Small Group Lab and Classroom

The first part of this book covered what you need to know to build your own personal electronics lab. Now it's time to share the love with your fellow electronics enthusiasts. Where to begin?

Like most important questions, the answer remains, "It depends." What resources are available to you and your group? How many folks will likely want to be working together, and for how long? How well do you really know these people? And these are just a few of the questions you're going to encounter along the way.

Take the time to understand exactly who "you" (plural; in Texas, *y'all*) are and what your requirements and goals could be. By doing so, you map out the future of your successful facility. Likewise, *not* planning ahead or making reasonable assumptions about your needs is simply an invitation to problems, problems, and more problems. Who wants *more* problems?

■ **Tip** If you fail to plan, you plan to fail.

Know Yourself

The ancient Greek maxim "know thyself" has as many definitions as it has authors, it seems. Plato indicated in several of his dialogues that Socrates would use this phrase to admonish his students to learn more about themselves and their own characters and capabilities before attempting to figure out the rest of the universe. Wise advice then and now.

As far as knowing yourself in relation to setting up a shared electronics lab, you should first think about what role *you* are going to play in this endeavor. Yes, *you.* Here are some of the major ones, and you might end up being all of them at some point.

Facilitator or Coordinator

In the role of the facilitator, you plan on providing an environment for others to use for their projects and explorations in electronics. You may or may not partake in these activities yourself, but have been tasked with (or volunteered for) the organization and coordination of the lab.

It's critically important to understand the needs of your intended ~~lab monkeys~~ users. See the "Know Your Group" section later in the chapter for more information on this topic.

As the lab facilitator, lab workers will come to you with questions, requests, comments, and complaints. Be prepared to address their needs in a responsive and professional manner—even if they're just your cousin.

Your responsibilities will include

- Communicating the scope of the lab (i.e., deciding what *is* and what *is not* acceptable or expected)
- Publishing any important information or updates regarding location, available times, features, rules, and possibly ongoing projects
- Crafting policies involving safety, proper tool usage and training, and participant conduct
- Maintaining supplies of commonly used materials
- Inspecting and repairing tools as the need arises

To balance out your responsibilities, you should be granted adequate authority in the following areas:

- Characterizing the scope of the lab, based on present and anticipated future needs
- Enforcing the policies of the lab
- Budgetary discretion in acquiring, provisioning, and maintaining lab assets
- Granting or revoking access to the lab based on established parameters

If you're going to be the Keeper of the Lab, then you should have both a personal interest in keeping it in good shape as well as the wherewithal to decide how things should be done.

Good luck finding the right balance there.

Teacher or Presenter

Let's assume that your lab has already been set up (or can be set up in a jiffy), and your main goal is to present information or teach classes there. Again, it should be apparent that a good knowledge of your intended audience will play a vital part in your success. See the "Know Your Group" section for more information.

Once you understand the needs of your class and the goals of your presentation, the next step is to leverage the existing facilities to your best advantage. Will you need specialized presentation equipment or materials? If so, see what is already available, and then make a shopping list of everything else you will need. Examples include lighting, a projector, specialized power distribution, and test equipment.

Are you making an interactive presentation? If so, what materials will you need to provide for your students to be able to fully participate in your presentation?

What will be your optimum teacher-to-student ratio? If you have a simple message that's easily communicated and reinforced, a single speaker can effectively address a rather large crowd. Once you start adding layers of complexity to the subject matter, you're going to have to be prepared to backtrack and explain things in more detail from time to time. If you're leading a group through a technical exercise, it's often profitable to have assistants available to help just those individuals that need a little more help, without having to stop the class in its tracks. These assistants need to be up to speed on the subject matter as well as have good communication and troubleshooting skills.

Remember that students look to the teacher for leadership, not only as a subject-matter expert. Take the lead, set a good example, and share your enthusiasm for your topic with your students.

A successful teacher leads a class on a journey of discovery. Make sure you don't lose too many along the way.

Participant

Your goal as primarily a group lab participant is to make effective use of the facilities at hand, making the most of your time and energy. Having access to a well-stocked lab with a good array of tools can be a decided advantage for both your immediate project goals, and it can be an excellent opportunity to accelerate your knowledge of electronics and other mysterious arts.

Imagining yourself as an active participant is the best perspective when laying out the flow of your lab. Set up workstations that have almost everything you need, leaving room for the complex or specialized stuff elsewhere. There are more suggestions about specific workstation configurations in the "Workstations" section of this chapter.

While the active-participant model helps you squeeze out the maximum utilization of your lab, be sure to consider the sort-of-working-but-also-sort-of-goofing-off model. Not everyone wants to be working full-tilt all of the time. Leave room for a certain amount of idleness, contemplation, and whimsy. You might be surprised at the results.

Observer

You want to be in on the action, but not necessarily working on anything in particular. That's great! Don't underestimate the importance of the purely social aspects of technical gatherings.

Some people just like to hang out with other like-minded individuals. This is a good thing—when resources and circumstances allow it. Watching creative, talented, and driven folks working on projects with passion is very often the best kind of inspiration. Let your kids see you having fun and making cool stuff in your lab, and the next thing you know, they are going to want to be there whenever you are. Pretty soon, they are going to want to start working on projects of their own—and that is its own special reward.

Know Your Group

Now you know a little bit about who you are and what your expectations of your lab may be. Now it's time to get to know the group of people who will most benefit from access to the lab.

Identify the needs and requirements of your group by first identifying *precisely* who your intended audience is going to be. Knowing this one item, up front, before any big plans are laid, is the best way to ensure success down the road.

There are several broad categories for which you might need to prepare. The more you know about your particular group, the better. Let's start fairly close to home and expand outward.

Your Family

Once you've established your own electronics lab at home, it's great to be able to share some of the fun with those that are closest to you. This might be your children, parents, siblings, or extended family. A well-planned and well-maintained electronics lab is going to generate some interest, or at least some curiosity. Here's how you can carve out some room for them, while still being able to get things done yourself.

It's very natural to want to share your hobbies and interest with others, and a safe and well-stocked electronics lab is a great place to do so. This is an especially ripe opportunity for parents to pass on their knowledge, skills, values, and work ethic to the next generation.

The family lab can be a logical extension of the family home. Everything is shared, including responsibilities and privileges. In more agrarian societies, this is already well understood at an early age, where chores around the farm and in the fields and a respect for the tools being used help enhance a sense of group belonging and worth. Everyone has a stake in the prosperity of the family endeavor.

Do you want your kids to appreciate what you give them? Then give them chores before you give them presents. Let them clean up the lab, put away tools, inventory parts, and take out the trash before spoiling them with all the fun stuff you get to do in the lab. This will help foster a sense of ownership in the long-term success of the lab, whether you're there to supervise or not.

It bears repeating (from Chapter 4) that no unsupervised children, no matter how adorable, should be allowed in the lab. Would you allow unsupervised cooking in the kitchen? It's the same basic scenario. When everything goes well, everyone has a good time. It only takes a split-second distraction to go from a good time to a trip to the emergency room in either case. Have rules and enforce them, even if it's your elders that you are protecting by doing so.

Sometimes having a resource like a shared electronic laboratory is the first opportunity people have to assert some sort of control over their relationships. Use the opportunity to make friends with your relatives, while you still can.

Your Friends

Unlike your family, as the saying goes, you get to *pick* your friends. Now you have the opportunity to provide a cool place to hang out as well as build all manner of awesome gizmos. It's only natural to want to share some of the fun with your like-minded friends.

The basic framework of rules and accountability apply to the friendly lab as well as any other group endeavor. Let your friends know up front what your expectations of them are in regard to use of the lab. Can they use your tools? Can they store artwork-in-progress at your place? When is it cool to come over to tinker, and when is it not? These kinds of questions should be dealt with early in the game; otherwise, these issues can flare up and become trouble spots. There's no substitute for good communication.

Once you extend your laboratory's hospitality to those outside your immediate family, it's a good idea to clearly designate what is your personal property and what belongs to the lab and all its participants. This can be as simple as putting name tags on all your tools and marking shared resources appropriately.

This is where *trust* really factors into the success or failure of a group lab. Just like in math, you want to eliminate as many variables as possible to ensure that a good time is had by all, and to ensure the long-term safety and viability of the facility itself. Which of your friends do you *trust* in your lab? Think about these things at length before opening the doors to just anybody.

Coworkers and Fellow Students

So you've been tasked with organizing or supervising a lab for a school class or university department. You may or may not have any choice in who will be granted access, or even when they can avail themselves to the facilities.

The balance of responsibility and authority is crucial in this scenario, just like in most other situations in life. Hopefully you will be in a position to enforce whatever rules are already in place. Ideally, you will have some say in crafting new rules as they become necessary, or adjust existing rules when they need it. In either case, it's just not fair for someone to expect you to do a good job, yet not provide you the tools to get it done.

Positions like this become available from time to time in most schools. Sometimes there's an opportunity for a small paycheck, and other times all you are going to get are points with the administration. Hopefully, at least you will also enjoy lab privileges. That's got to be worth something, yes?

If you are helping to run an existing lab, there are probably others you can consult who have had this particular honor in the past. Try to find out who these folks might be and see if you can get any helpful advice from them about your lab and its care and feeding.

If you're starting from scratch, you will enjoy the double-edged sword of setting things up the way you want, as well as setting things up the way everyone else wants it. You *get* to establish whatever level of functionality you want, with the caveat that you then must maintain that level. It's a freedom, like most freedoms, that comes with a price.

Special-Interest Groups

Are you a member of club that has a special interest in electronics? Several popular interests, such as ham radio operators, remote-control model flyers or drivers, robot builders, and computer users, have a love-hate relationship with electronics. Having a dedicated electronics lab available to the membership is a great way to build up the love as well as work out the hate.

Given a highly specialized crowd, such as the members of a university or trade school, you might even find people organized into an electronics club, especially these days when almost every aspect of our lives is touched by this technology in one way or another.

The specialized nature of your group helps focus what areas of electronics are of most practical interest, at least from a lab point of view. This cascades into easier choices when it comes to stocking the lab with workstations, tools, and components.

The Dallas Personal Robotics Group (DPRG; see `http://dprg.org`), for example, had the use of an 1,800-square-foot warehouse for over six years, rent free, due to the generosity of one of the members, Mike Dodson of Modern Assemblies. Having access to an excellent industrial setting allowed the club members to dream of and then realize much more ambitious robot projects than could be attempted at home. After Mike's retirement, the DPRG went into a nomadic phase and finally spun off a dedicated workshop and meeting area, now called the Dallas Makerspace.

The General Public

A lot of people see the popularity of do-it-yourself (DIY) and *maker-movement* projects and envision themselves as entrepreneurs, seeing a need and being able to meet it. Unfortunately, most of these people are wrong.

Building a lab dedicated to public use sounds like a great and noble aim. The realities of such an undertaking, including the financial, logistical, and business related, are not to be underestimated. While it seems to work well in certain areas and at certain times, it's almost always the result of the right combination of people coming together at the right time.

Hackerspaces (a terribly unfortunately chosen label) are springing up all over the world. The term *hacker*, while understood within some technical communities to mean a person of respectable skills, has a decidedly *unrespectable* taint to it when viewed from the outside. To the vast majority of the English-speaking world, a hacker is a rogue computer jockey, a criminal intent on breaking and entering into computer systems for nefarious purposes.

The term *maker*, on the other hand, leaves a more positive first impression. The Dallas Makerspace members deliberately made this decision when naming themselves, while they were still just a subcommittee within the board of directors of the nomadic DPRG, looking for a new home. It was important to them to maintain the very positive image that the DPRG had established and built over its long history, while still looking to the future.

If you're going to set up a lab for use by the general public, whether you decide to call it a hackerspace, a makerspace, or something entirely different, you get to make almost *no* assumptions about the good intentions of your membership. It's a sad state of affairs, to be sure, but it seems to be the prevailing wisdom of the day.

Some popular and successful organizations—such as Noisebridge (`www.noisebridge.net`), in San Francisco, California, which was incorporated in the state as a "non-profit educational corporation for public benefit"—operate almost entirely on a "consensus basis," sometimes known as "anarchic chaos." Noisebridge's overriding motto is "Be excellent to each other," very often shortened to just "Be excellent." Its very real success and sense of community among members is based on what seems to be a very local tolerance that is not to be found in many parts of the globe. If you ever get a chance, do stop by and prepare to be impressed. It's an amazing place.

A good parallel would be to visit your local public library. In the areas that still maintain libraries (shockingly, they are on the decline), you will find a finely tuned machine that meshes well with the surrounding locality. Are there bars on the windows? If so, you're going to need some bars on your windows. Are there metal detectors at the entrance? Oh, dear. Do you see any uniformed security forces present? These are not good signs.

Don't be completely discouraged, however. Sometimes a community in decline needs a little boost of energy and an injection of esprit de corps. This can often be accomplished by introducing some new alternatives for people's spare time. An ancient Muslim saying reads, "Trust in Allah, but tie your camel."

A Shared Resource

Do you have a dedicated space available? If so, you are way ahead of the game. Finding a good location is half the battle. If your space is truly dedicated to your pursuits and activities (i.e., a private home, club, or makerspace), you have the advantage of being able to leave tools, components, and projects in place, returning to them (and finding them still there!) when the opportunity arises. "Truly dedicated" in this sense can be equated with having your own key to the door and reasonably free access. While a 24-hour facility sounds like a great idea, there are a lot of logistical problems associated, such as security, utility costs, scheduling, and parking, just to name a few.

If your space is not-so-dedicated, such as a school lab or friend's garage, you can still make it a home away from home if you are able to arrange for secure storage of all your shiny toys. This could take the form of individual lockers or a storage closet.

No matter what the logistical arrangements, the main challenge to providing a versatile and effective setting for electronic and other kinds of tinkers is basically a social one. Once it's more than just you, it can start to get complicated.

Luckily, there are some time-tested methods for making these types of things run smoothly.

Be a Hospitable Host

If you are taking the lead in arranging a shared work space, please remember that you are responsible for extending the appropriate hospitality to your present and future guests. This single factor will greatly determine the vibe, feeling, or ambiance of the facility, even more than having the latest and greatest tools and gizmos on hand.

This task may or may not fall entirely on your shoulders. Many hands make light work. If you are able to delegate some of the duties of a hospitable host to other members of your group, consider doing so.

Think of your task in the same light as planning a successful party, even if it's one that you anticipate will last for days on end. These basic ideas are as old as civilized living. Think about fabulous parties or memorable events that you have been privileged to attend in the past, and then plan on taking it up a notch.

Try to anticipate the human needs of your attendees as well as their technological expectations. People have what seems like an endless list of requirements, some of which you can address, some not.

Make your location known. Make this information easy to find to those interested, and easy to update when the venue changes. A web site is trivial to set up these days, and even a post to a popular forum will tend to be accessible via Internet search engines, in time. There's a new social media outlet popping up every third Tuesday, so avail yourself to what makes sense for your audience. Word of mouth only goes so far these days.

Once you've gone to the trouble of publishing your location, go the extra mile and put out some sort of actual sign or indicator that your arriving guests have found the right spot. If you have an identifiable logo or symbol associated with your group, be sure to proudly wave that flag. People don't like feeling lost, or unsure whether they're entering a welcoming space or about to trespass on private property.

Let everyone know when they are expected as well as when they are *not* expected. Post your regular meeting times, if you have them. Alternatively, make known the appropriate hours of operation. It's disappointing to be turned away at the door just because it's 1:00 a.m. on a school night.

When possible, welcome your guests. This makes a big impression on both newcomers as well as the regulars. Being part of a group is supposed to *enhance* a feeling of inclusiveness and belonging. We are social animals at heart, and even a simple hello is often the prelude to an interesting and productive time together.

Have rules. Make these rules known. A sign on the front door is not out of place, as long as it's not too negative or off-putting. Rules should *prevent* problems from occurring, and not be enacted as a spontaneous reaction when problems inevitably occur.

Ensure your guests' comfort when attending. Is there enough room for the largest (and the smallest) participant to safely enter and exit the facilities? Is there adequate lighting in the common areas, if they exist? Is there a place for everyone to sit? Who has dominion over the thermostat?

Do you plan on making refreshments available? Will this duty rotate among the membership? Keep track of such things, if only to keep from overextending yourself in all the myriad details of hosting a successful meeting. While massive banqueting and consumption of yummies make for a festive and enjoyable association, these activities are not always aligned with maintaining an effective laboratory. Separate the feeding trough from the work areas. Stay focused on what you're trying to accomplish here.

Will everyone have access to restrooms, and whose job is it to keep these areas tidy?

And the most important question of all: Who takes out the trash?

Be a Courteous Guest

Just as it's important to play the role of the hospitable host, it is likewise critical to be a courteous guest when visiting a facility, whether you are a paid-up member in good standing or an invited guest. Again, these guidelines will vary according to the intents and purposes of your lab, but the basics will always remain the same.

If possible, and when appropriate, let your host know if you are planning on attending. This won't apply in a causal, drop-in type of laboratory. Even so, by making your intentions of attending known ahead of time, you might influence (one way or the other) the attendance of other potential participants. Nothing succeeds like success. A well-attended party has a better potential for awesomeness than a sparsely peopled event.

Arrive on time or within the established time frames set forth by your host. Similarly, unless you have been granted special access, plan on leaving at a reasonable time as well. Benjamin Franklin wrote in his *Poor Richard's Almanack* (1736) that "Fish & Visitors stink in 3 days." Don't be a stinker.

Try to contribute something. Don't show up empty-handed, expecting to be fed, entertained, and enlightened.

Workstations

Now it's time to talk about your workstations. Even if it's just you in the lab, having a separate workstation for each type of activity is a great idea. For example, you might have one workstation set up for programming a computer, another for soldering electronics, and another for working with power tools, such as a drill press or a saw.

Workstations dedicated to a particular task should have everything within easy reach, including tooling, supplies, power, and lighting. Don't waste time by forcing people to wander around the lab looking for the right drill bit or magnifying glass. Keep everything close so you can find what you need, when you need it.

Having a place for everything ahead of time is the best way to follow the wise adage, "A place for everything and everything in its place." You get to decide where everything lives. It doesn't hurt to go ahead and label everything conspicuously so that others can find their way around the lab.

If you have several people wanting to work on the same kinds of projects at the same time, you're going to need multiple, nearly identical workstations. The exact number depends on your available space and the number of people you expect to want to participate simultaneously, as well as your budget.

Most indoor-oriented work can often be performed at a work table. Exceptions include giant-scale mechatronics, pyrotechnics, and underwater robotics. Some of their constituent bits and pieces, however, can be crafted on an adequately prepared work table.

Make sure each work area has plenty of power outlets available. You're almost guaranteed to run short at some point. Take care not to overload your circuits, however. If you anticipate needing a lot more power than is presently provisioned, you're going to have to consider the real possibility of having your facility expanded by qualified, *licensed* professionals. These requirements will vary from one location to the next. Check with your local building codes to be sure.

As repeated often in previous chapters, you can never have *too much* lighting available, especially for the more detail-oriented chores. Typical indoor room lighting is just about adequate for finding your way out of the room, and that's about all. You'll need area lighting for each workstation, and possible additional task lighting, depending on your particular requirements.

Providing network access to each workstation can be challenging. Wireless networking is a good way to accommodate a relatively small number of users at once. More than a handful of computers trying to access the Internet at the same time can bring many wireless routers to a slow crawl. Dedicated high-speed Ethernet connectivity is more expensive and takes more effort to deploy, but offers much greater bandwidth to your lab-folk.

If you're especially cramped (or *dimensionally constrained*) in your lab, you'll need to get inventive. For example, a table placed in the center of the room can be approached from multiple sides. There's no reason all the workbenches have to be pushed up against a wall, allowing access from only a single side. Get creative!

Safety Planning

While it's tempting to fill up all the available space in a lab with workstations, especially when the budget permits it, do try to leave a bit of common area here and there. You always want to be able to walk *away* from a workstation with the same (or sometimes even more) rapidity you approached it with. Fire is an excellent example of such a motivator.

■ **Tip** Keep a properly selected and maintained fire extinguisher at each workstation, if possible.

Post emergency procedures prominently, where everyone, including visitors, can see. Have an evacuation plan, complete with map and "You are here" indicator posted at each workstation. Hopefully you will never need it.

Presentations

Are you excited about your latest project? Have you just learned a great new skill? Setting up your lab for group presentations is an excellent way to communicate these ideas to your friends, family, and coworkers.

Make your ~~victims~~ guests comfortable. Have adequate seating for everyone, with plenty of room between the rows (if you have more than one row of chairs). Ensure that everyone has a good view of what you intend to present.

Will the attendees be expected to take notes? Then perhaps you should set up tables where they can comfortably jot down some of the more spectacular aspects of your presentation. If you will be providing a handout or other documentation packages, make sure you have enough for the anticipated demand. Will you be entertaining questions or comments either during the presentation or afterward? Then make sure that not only can your audience see *you*, but that *you* can see *them.*

If you're going to be using some sort of projector for your presentation, by all means set aside some time to get to know the equipment ahead of time. Nothing ever seems to work just exactly as you imagined that it should. This seems to be an invariant law of presentation physics. No one is really sure why.

Your audience's time is just as valuable as your own. In fact, if there are only two people attending, then their collective time is worth approximately *twice* as much as yours. Think about it. Have all your ducks in a row, so to speak. Wasting time figuring out which cable goes where is not a good use of your time together.

Summary

Wow! You can really go crazy setting up a shared electronics lab, can't you?

Before you even flip on the light in your shiny new group lab, you should already have an insightful idea of who is going to be using it and why. You should also have some honest expectations of what your own expectations are going to be, and how you would like to interact with both the lab itself and its population.

The details, of course, are always going to be specific to your particular situation. Be on guard, however, for the future. It's going to sneak up on you, someday. Not too long ago, for example, it took a fairly hefty desk to keep a modern personal computer up off the floor. That kind of computing power now fits in a pocket. Think of what better uses you could dream up for that space these days!

That kind of forward thinking is exactly what turns the dreams of today into the reality of tomorrow. It's forward-thinking people like you, dear reader, that help map out those realities. Take the opportunity you have right now to spend some quality time daydreaming about how things *could* be. Then take those steps necessary to make it happen. Your family, your society, and your world expect great things from you. You can do it!

APPENDIX A

Getting Started with Tool Building

Once you've become comfortable in your lab and have had time to play with a few circuits, you might start to notice that some of the "tools" you've been using are just simple electronic circuits themselves. Good examples are power supplies and meters. These are great tool-building projects because you can see useful results early.

Let's take a look at building a couple of simple electronic tools that might be of use in your lab. Hopefully you will develop a better understanding of how these tools work. Ideally, you will progress from the basic question of "Does it work?" to the more involved questions of "How well does it work and what can I do to improve it?"

Please bear in mind that these exercises, while hopefully useful, are to illustrate what needs to go into the building of practical, reliable tools. At every stage you should be thinking about what design choices you have and which choices will enhance the usefulness of your tool-building process.

A Very Simple Power Supply

Almost all electronic projects need electricity to work. Having a good, dependable power supply at your disposal on your bench will get you started faster on new projects. You won't have to worry about dead batteries and flaky power adapters. See Figure A-1.

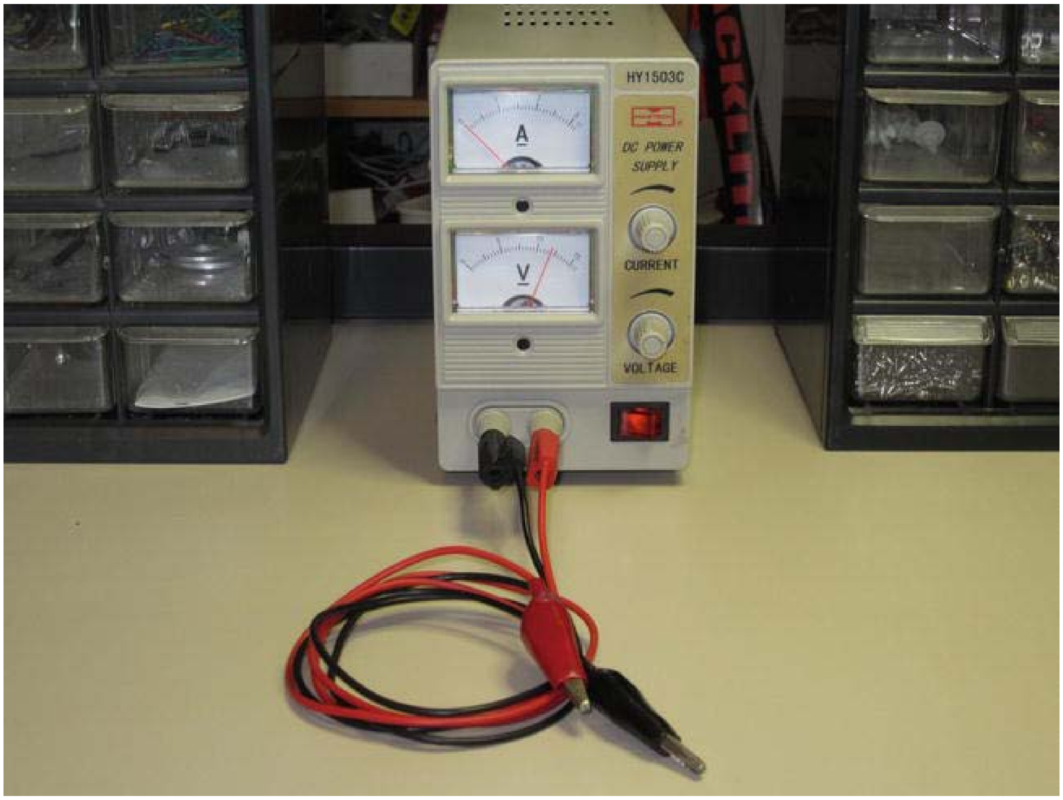

Figure A-1. *A professional bench power supply. This device allows both the voltage and the maximum current to be adjusted. It provides two dedicated meters to show how much power is flowing: one for voltage and the other for current. It also provides a modular set of plugs (banana plugs) for easily connecting and disconnecting power leads.*

A bench power supply such as the one shown in Figure A-1 has many features and includes extra circuitry to make it safe and easy to use. It will also set you back a good chunk of change (i.e., cost money). We're going to build something a bit simpler to start. Feel free to add more bells and whistles as your needs arise.

■ **Caution** Anything plugged into the wall has the potential to shock, burn, or kill you. Do not defeat their safety features!

An Easy 2.5 Watts

You probably have a spare USB charger lying around somewhere. If you don't already, you more than likely will at some point in the near future, as these are becoming more common with smaller electronic devices. If not, you can obtain one relatively cheaply if you do some smart shopping. The USB charger shown in Figure A-2 was obtained from overseas via eBay for less than $1. You will also need a USB cable about 3 or 4 feet long. Again, don't spend a lot of money on a new cable. If you don't already have one floating around in a drawer somewhere, you probably know someone that does.

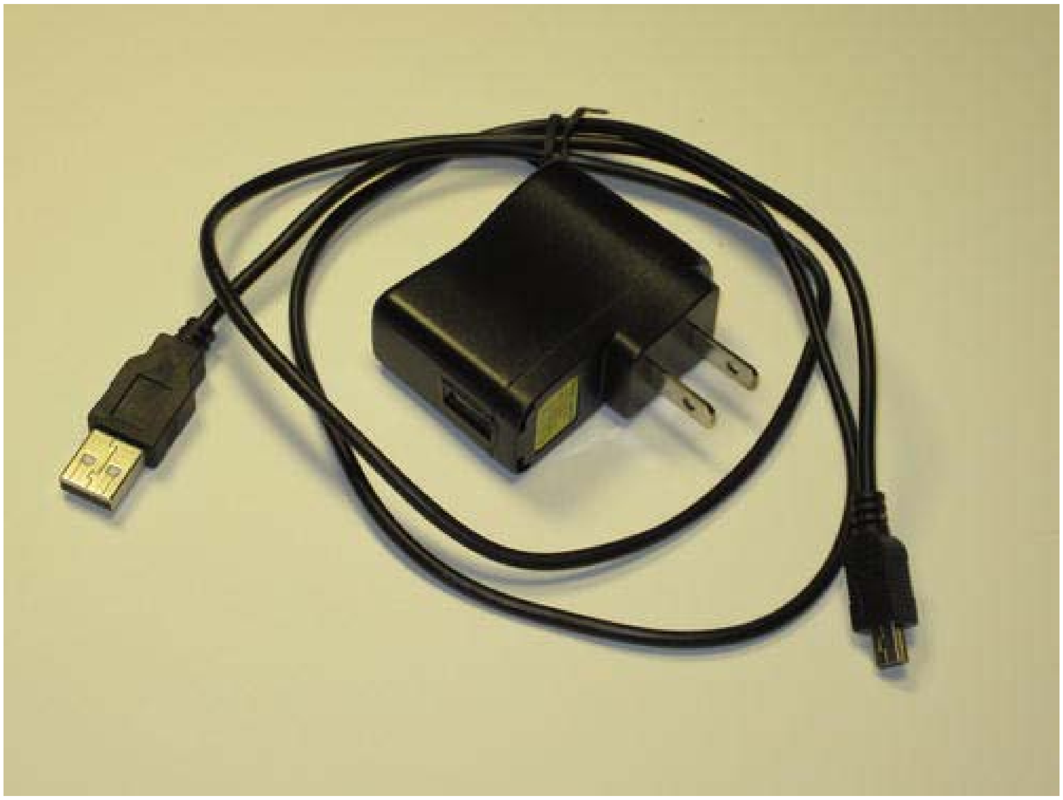

Figure A-2. *A USB charger that plugs into the wall and provides a regulated +5V supply. This plug is intended for US wiring, but will operate from 110AC to 240VAC at 50 or 60Hz, which pretty much covers the whole world, if you have the right adapter plug.*

It doesn't matter what type of USB cable you find, because we're going to cut off the smaller end and use the bare wires inside the cable. It just has to have the USB A–style plug shown on the left side of Figure A-2. The plug will fit into the socket of the charger. You do not need to open or modify the USB charger at all, and you are seriously advised not to do so for any reason.

Let's take a closer look at the label on the charger to find out exactly what it's supposed to be doing for us. See Figure A-3.

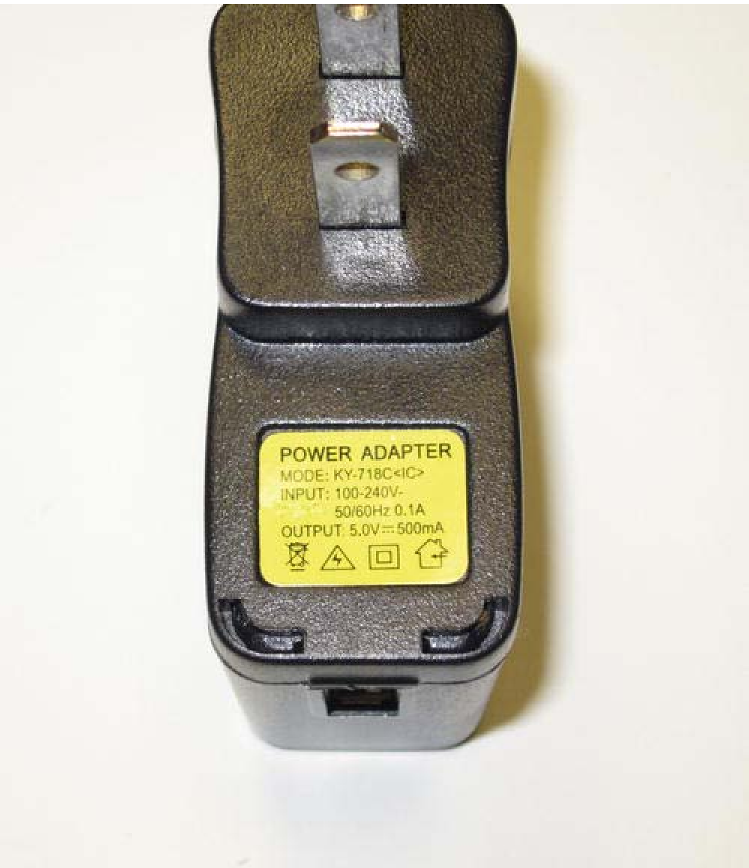

Figure A-3. *The label of the USB charger tells us that it accepts anything from 110VAC to 240VAC at either 50Hz or 60Hz. This wide range of input power allows the same internal electronic design to be used worldwide. Only the prongs that plug into the wall vary from location to location. The label also tells us that the USB charger will output +5V at up to 500mA. That's 2.5W!*

The "OUTPUT: 5.0V ⎓ 500mA" marking tells us what we need to know about the flavor of the power coming out of the adapter. The "5.0V" is the measure of the voltage. This is a very useful voltage level for small projects. Many digital chips require a 5V supply, although 3.3V is also becoming quite popular.

The ⎓ symbol represents *direct current (DC)* voltage. The symbol for *AC (alternating current)* voltage is the tilde, or wavy line symbol ~. The "500mA" is the maximum current that can be supplied by this charger (500 milliamps, or 0.5A), although you are cautioned not to expect a low-cost item such as this one to produce 100 percent of its rated power 100 percent of the time.

While this USB charger will most likely work for our intended purposes, it's best to keep those expectations in check. Note that "Model:" is misspelled ("Mode:"); and also note the absence of the magic letters *USB* or the USB logo. Also, there is a complete lack of any safety certifications present on this humble device, as befits its ~~cheap~~ economically optimized provenance. Compare this with the veritable constellation of safety certifications on the name-brand laptop charger shown in Figure A-4.

Figure A-4. *A name-brand laptop charger with several international safety certifications. Every one of these certifications costs money to obtain, and shows that the device was demonstrated to be safe for use in a typical home or business application.*

Next we adapt the USB cable to our needs. In the words of the Red Queen, “Off with his head!”

■ **Caution** Make sure the USB charger is not plugged into the wall or a power strip when making modifications to the cable.

First, plug the USB cable into the USB charger. Now cut off the *other* end of the cable. See Figure A-5.

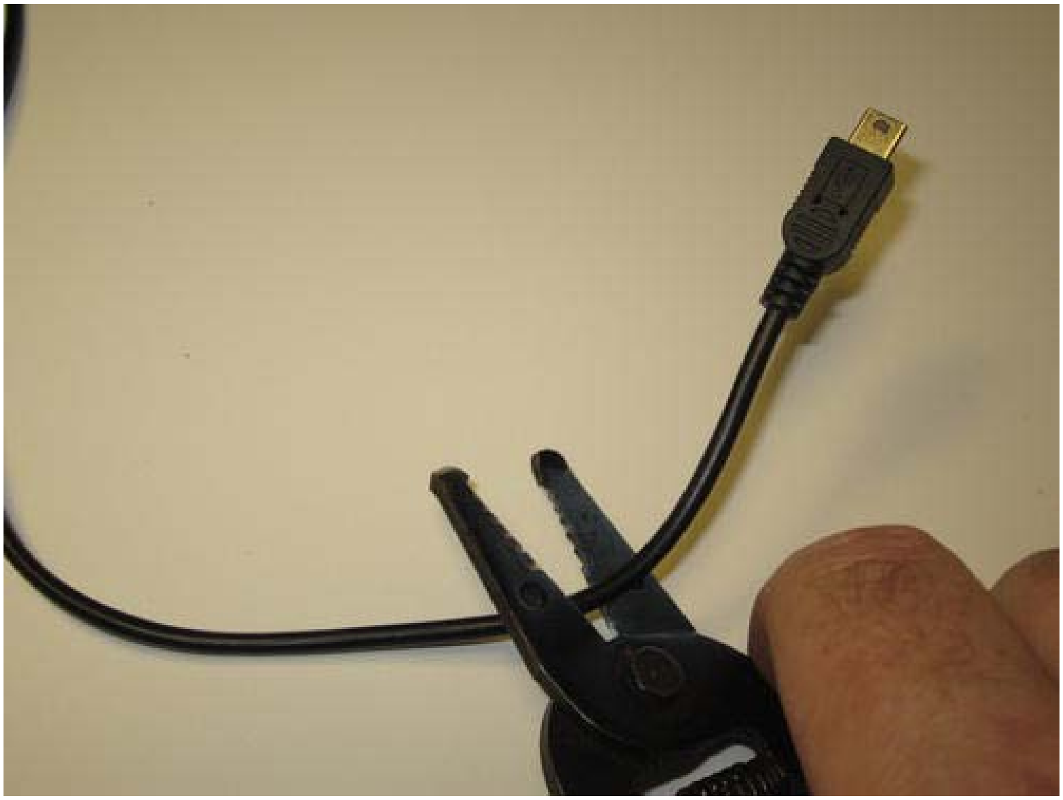

Figure A-5. *Cut off the device end of the USB cable. If you cut off the wrong end, you're not going to be able to plug it back into the charger, at least not without a considerable amount of effort. To ensure you cut the correct end, plug the host end of the cable into the USB charger before snipping, and then cut off the other end.*

Now carefully strip off about an inch of the cable's insulating jacket. Do your best to not nick the insulation on the wires inside. Within the jacket you should find four wires. Two of these wires carry the +5V and ground from the power supply. The other two wires carry the data on the USB connection. Since we're only interested in the power, we can ignore the other two wires for now. See Figure A-6.

s

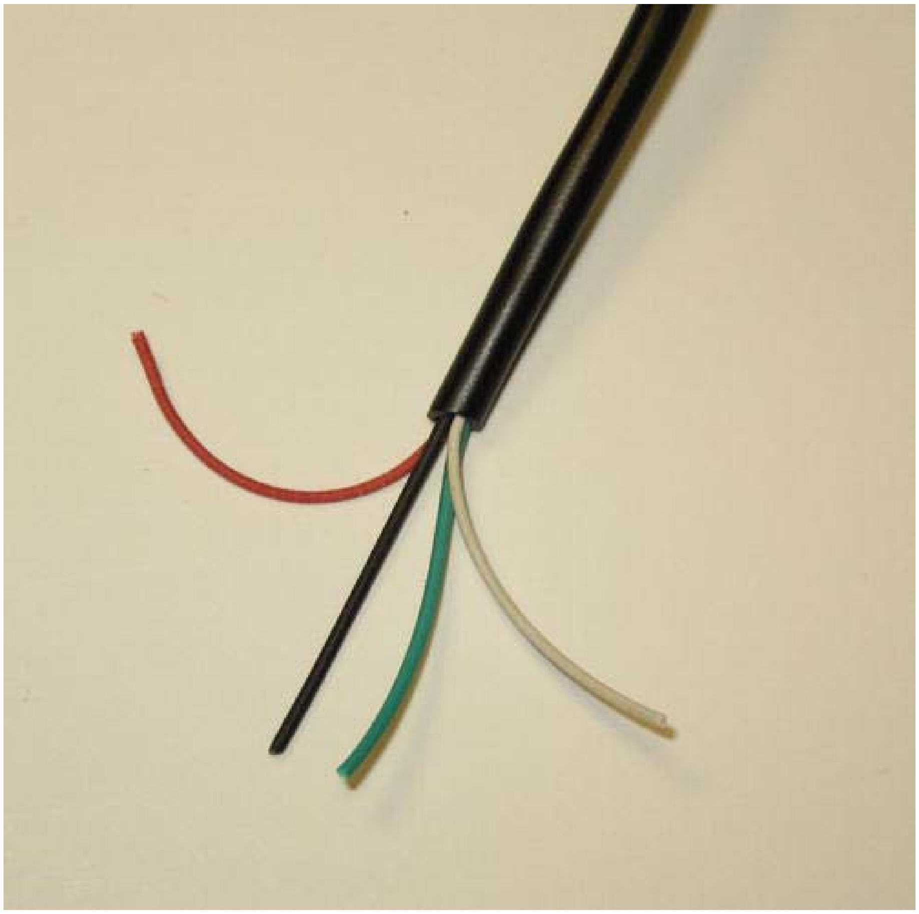

Figure A-6. *The four internal wires within the USB cable. From left to right they are red (VBUS or +5V), black (ground), green (USB D+), and white (USB D–). Use the red and black wires to pull a regulated +5V from the charger.*

The USB standard specifies every little detail about how to build and mark USB cables, even down to the colors of the internal wires in the cable that you're not ever supposed to see! That's dedication.

■ **Note** Apple violates the USB standard by not having a raised USB symbol molded into its cables.

However, sometimes cable manufacturers fudge a bit and use whatever four-conductor cable they happen to have lying around. A very common substitute cable uses red, blue, yellow, and brown wires internally. See Table A-1 for a list of which wires go where.

Table A-1. *Standard USB Wire Colors and Their Functions, Along with Popular Alternate Colors*

Function	**Pin Number**	**Standard USB Color**	**Alternate Color**
VBUS (+5V)	1	Red	Red
D– 2		White	Blue
D+ 3		Green	Yellow
GND 4		Black	Brown

Now is the time to verify the wiring before we get much further. Even if your cable has the official wire colors within the insulating jacket, you are strongly advised to double-check both the voltage and the polarity coming out of the charger before connecting it to your tender, sensitive circuits. Minimize your assumptions about the compliance of others. Testing proves testing works!

Strip about 1/4 inch of the insulation from the ends of the red and black wires. Get out your meter and set it up to read voltage at or above the 5V range. If your fancy meter has autoranging, then just set it to read DC voltage. Plug the USB cable into the charger, if it's not already installed. Make sure the red and black wires are not touching, and then plug the charger into the wall or a convenient power strip.

Now measure the voltage between the red and black wires. Hopefully you should see a reading at or near 5.0V. The USB standard specifies a tolerance of ±10 percent, so anything between 4.5V and 5.5V is considered OK. See Figure A-7.

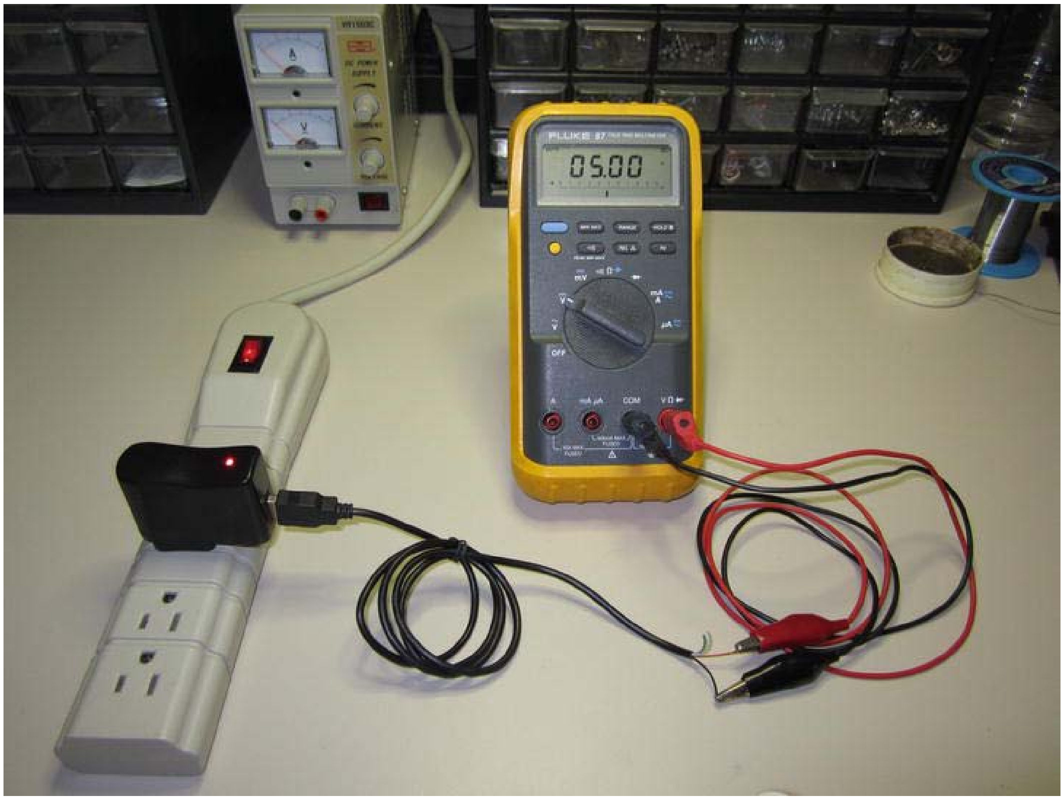

Figure A-7. *Measure the voltage coming from the USB charger before connecting it to any of your circuits. This one reads 5V on the dot! This is a much safer method than trial by fire (i.e., plugging it into your circuit and looking for a fire).*

Now that you've confirmed that the red and black wires do indeed carry the regulated +5V from the USB charger, you can clip off the unused green and white wires. At this point you can either use the bare, stripped wires of the USB cable or solder on a more robust connector. Since this power supply is destined to be used with a solderless breadboard, a small, two-pin connector with the appropriate spacing will be attached. See Figure A-8.

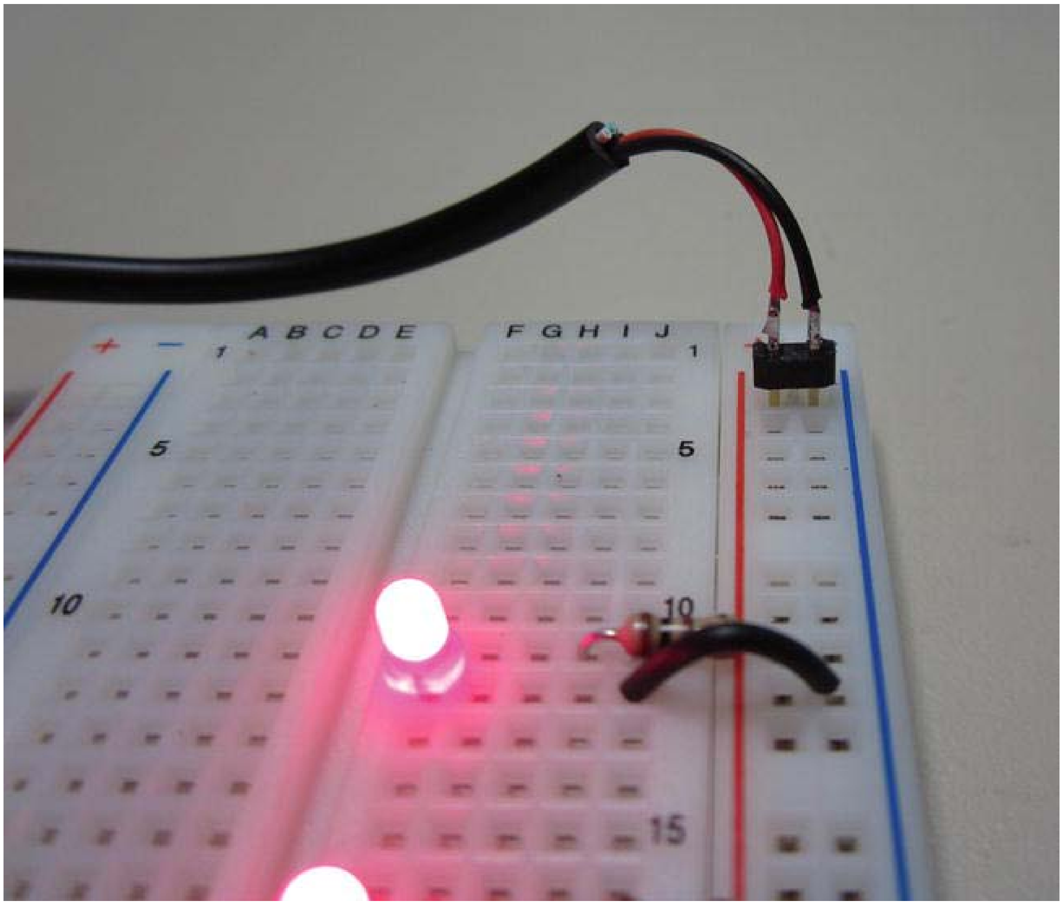

Figure A-8. *A gold-plated pin connector is soldered to the modified USB cable. The red and black insulation is left visible to be a polarity indicator when installing the cable to the solderless breadboard. Another happy LED!*

Adding Features

This power supply works just fine for most projects. If your requirements are, “Translate AC power from the wall into regulated DC power I can use for small projects, without spending a lot of money,” then this contraption fits the bill precisely.

You can see for yourself in Figure A-8 that it does indeed work as advertised. Looking back at Figure A-7, you can see that the charger itself has a power-on indicator LED built into the housing. What it doesn’t have is a power control or a switch. The power to the circuit is controlled by the switch on the power strip. This works just fine as long as you don’t mind turning on and off everything else that happens to be plugged into the same power strip.

Various Indicators

Not all USB chargers have a power indicator, however. Some have both a power indicator LED as well as a *charge status indicator*, which can tell you at a glance if your USB device is fully charged or still charging. A charge status indicator is usually implemented by measuring the amount of current being drawn from the supply. If more than a preset current is flowing, the power supply assumes a battery is still being charged. When that current drops below a certain level, it assumes that the battery has been completely (or maybe just mostly) charged. Some chargers use a red or orange LED to indicate charging, and a green or blue LED to show that the battery is finished charging. Again, this type of charge status display is provided to give the user a quick status update with a single glance. You don't need to consult the product manual to interpret the status code, if you can remember that red means "charging" and green means "charged."

Let's rearrange some of the parts on the solderless breadboard to make room for some new features. See Figure A-9.

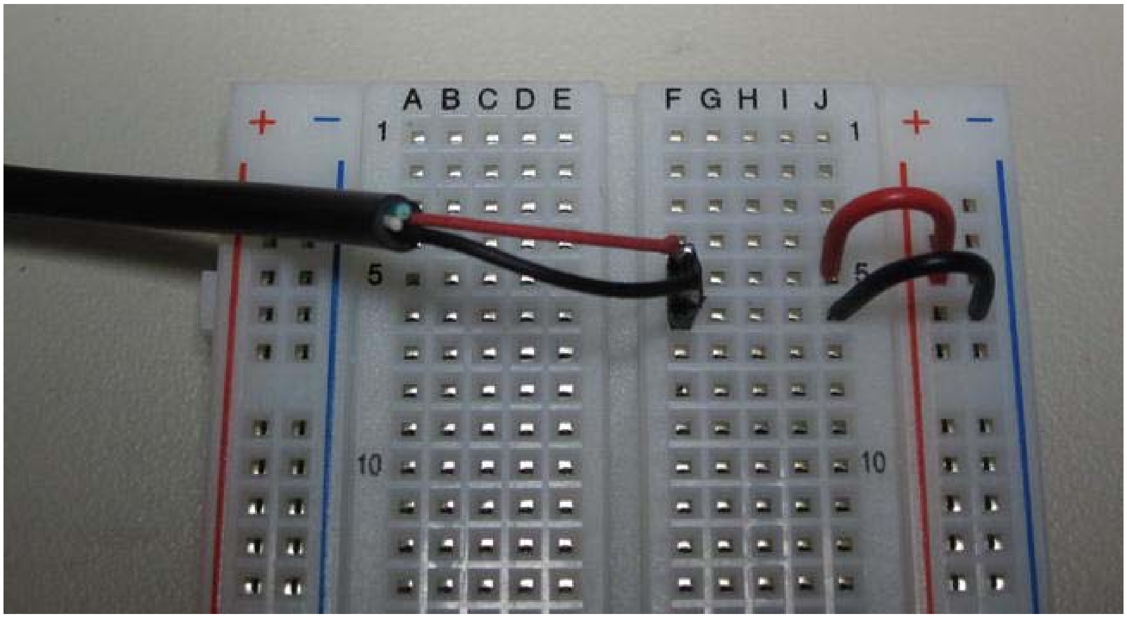

Figure A-9. *The regulated power coming from the USB charger is no longer directly connected to the breadboard's power rails. Instead, it is routed to some internal tie points and then connected to the power rails via two short jumpers.*

Power-On Indicator

First, we add a local power-on indicator by reinstalling the LED and current-limiting resistor used in the previous experiments. While this duplicates the function of any LEDs that might be present on the USB charger itself, it brings the indicator closer to the area of interest, which in this case is the circuitry onboard the solderless breadboard.

The anode of the LED goes into tie point F-3, with the cathode being installed at tie point F-4. The current-limiting resistor spans the gap between tie point I-3 and the right-hand positive (red) power rail. A short jumper connects tie point J-4 and the right-hand negative power rail. See Figure A-10.

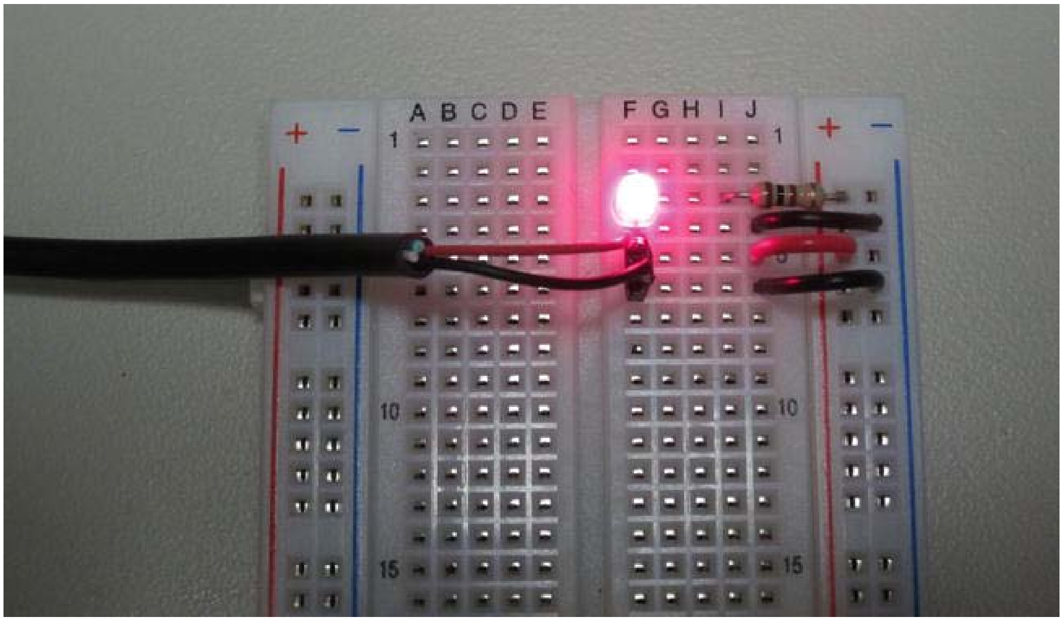

Figure A-10. *Add an LED and a current-limiting resistor across the power rails of the solderless breadboard to give an at-a-glance power indicator, right where the action is taking place.*

Now we have a good power-on indicator, right where we need it. It may or may not be redundant, depending on the features of your USB charger. This type of indicator is digital in the sense that it tells us that the circuit is either on or off, without giving us a lot of precise information about anything in between. For a lot of applications, however, this is more than enough information.

Power Supply Voltage

What if you *did* either want or need more precise information about the power supply? One method would be to permanently connect your multimeter to the circuit so that you can see how many volts are being delivered. That method assumes that you're not going to need your multimeter for anything else, ever.

A practical variation of that method, however, it to provide convenient *test points* in your circuit so that you can quickly and easily take a voltage reading when the need arises. Be sure to position the test points so that they are easy to access. Having a nice, big, readable label wouldn't hurt, either. Also, try to arrange things so that it's not likely that contact with a metal object, such as a tool or your wristwatch, will accidentally short out the power supply by touching both test points at the same time. Even simple "features" get complicated quickly, it seems!

If you want an accurate (and permanent) voltage reading on your circuit, consider adding a voltmeter. Small panel meters are relatively inexpensive and reasonably accurate. A DC voltmeter has two leads, positive and negative, that are attached across the voltage to be measured. See Figure A-11.

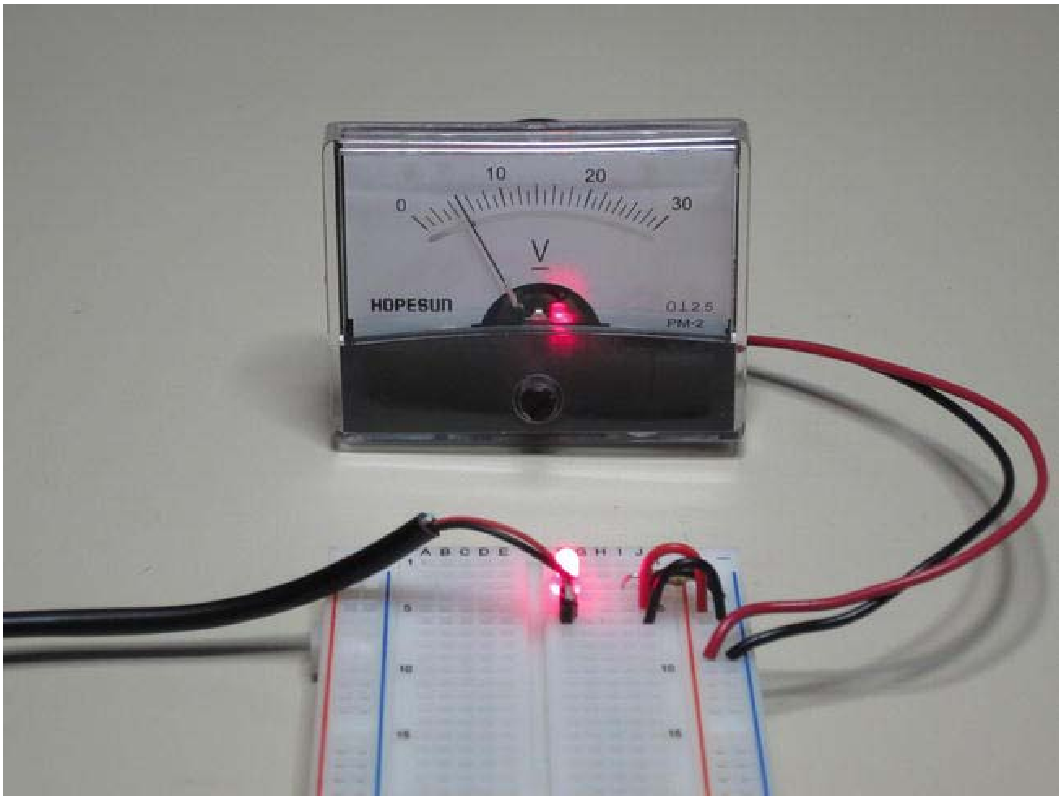

Figure A-11. *A DC voltmeter is attached to the circuit to accurately measure the supply voltage. The V on the meter's legend stands for voltage. The underline beneath the V indicates that it is designed to measure DC voltages, as opposed to AC voltages.*

The panel meter shown in Figure A-11 has a range of 0–30VDC. That's a bit more than we need, since our little power supply is only supposed to produce 5.0VDC ±10 percent. Since the meter's range does not accurately match the expected range of supply voltages, we lose a bit of precision when attempting to read the display. All the action is crowded together on the left side of the meter.

It might seem that a 0–5VDC meter would be preferable, but that would not allow for any overvoltage conditions at all. Remember, the USB standard allows for up to 5.5VDC to be delivered over the cable. A 0–6VDC meter, if available, would be more suitable. The meter introduced in Figure A-11, however, would be perfect when used with the laptop supply shown in Figure A-4, because its output voltage is 19.0V. That leaves a bit of room for overvoltage conditions and gives a wide enough swing of the needle to produce fine detail.

Power Supply Current

Measuring the *voltage* at any point in a circuit is easy. Just put a voltmeter between two points and voilà! In our previous example, we measured the VBUS power supply voltage with reference to GND, the ground potential of the circuit. Many voltage measurements are made with respect to ground.

To measure the current flowing through a circuit, we have two options. The first option is the most direct route. Install an ammeter (a meter that measures *amps* of current) inline with the circuit you want to measure. This requires that you *break* the circuit to make room for the meter. Once installed, the current flowing in the circuit actually flows through the ammeter as well.

The second option involves using a bit of trickery and some math. Instead of installing an ammeter directly inline with the current flow to be measured, we install a resistor of known resistance and measure the *voltage* that develops across the resistor.

Since we know that the values of volts, ohms, and amps are always related in any circuit, we can use Ohm's Law to determine how much current is flowing through a circuit by measuring the voltage across a known resistance.

To make things very easy, let's use a 1Ω resistor in series with the power supply, inserted in the circuit between where the positive lead of the power supply arrives at the circuit and where it is connected to the rest of the circuitry.

Unlike our previous use of Ohm's Law to calculate the missing value based on two known values, we only use *one* known value (the value of the resistor, called a *shunt resistor* when used in this manner). This one known value (resistance) then establishes the relationship between the two other values, voltage and current. As the current increases in the circuit, so does the voltage across the resistor. The two values will mirror each other—as one increases or decreases, the other will follow.

Since we used a 1Ω resistor in the circuit, this sets the relationship between current through the circuit and voltage across the resistor at a 1:1 ratio. For every ampere of current through the circuit, for example, 1V will be measured across the resistor.

Our power supply is only supposed to be able to deliver 0.5A of current, maximum, so we'll be dealing with voltages of less than 0.5V, unless something terribly unexpected happens.

Again, we could permanently attach a multimeter across the leads of a 1Ω resistor inserted into the circuit, or we could attach a voltmeter that is sensitive enough to give accurately readable information in the range we need. The 0–30VDC meter used before is much too large in range to do us any good here. We might see a tiny wiggle in the needle, but that's all.

Instead, we'll use a different kind of meter entirely. A digital voltmeter contains its own dedicated measurement circuitry and a numeric display to indicate the voltage reading. It also requires its own separate power supply. See Figure A-12.

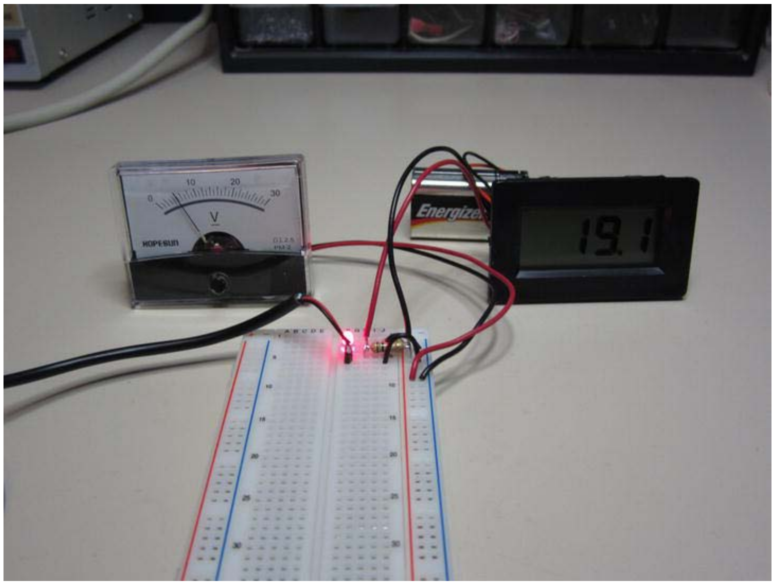

Figure A-12. *A digital voltage meter is used to read the voltage across the shunt resistor installed in the circuit. Because the shunt resistor has a 1.0Ω value, Ohm's Law dictates that the reading in millivolts on the digital meter correspond exactly to the number of milliamps flowing through the circuit. Math is useful, sometimes.*

The jumper wire connecting the incoming power supply directly to the power rail has been replaced with a 1Ω resistor. Leads from both ends of the resistor have been connected to the input terminals of the digital voltage meter on the right. The digital meter, unlike the analog meter on the left, requires an independent power supply to operate, in this case a 9V battery.

So now we've added some features to our simple power supply, including a power-on indicator, a supply voltage reading, and a supply current reading.

These are very useful features to have on any power supply. However, at this point in the prototyping stage, we've just barely covered what the added circuitry is supposed to do. Packaging all this up into an easy-to-use and reliable device takes as much if not more thought and effort.

First, you have to rebuild this entire circuit in a more permanent fashion. Next, you will have to wedge all that clever gadgetry into some sort of enclosure to protect the delicate bits while still permitting access to the inputs, outputs, and controls, if any. Then you get to label everything so that you can remember how to use it in the future.

Summary

So now you just might be developing a certain respect for how much work goes into building even the simplest tools—tools that are useful, accurate, reliable, and an asset to your toolbox. Remember, every tool on your workbench has to earn its place there. This applies to tools you design and build yourself, as well.

Good luck with your tool-building projects. Sometimes it's just as much fun to work on your tools as it is to work *with* your tools.

Index

A

Active electrical and electronic components
  batteries, 81
  diodes, 82–83
  integrated circuits, 84
  LEDs, 83
  power supplies, 81
  transformers, 82
  transistors, 83–84
Active-participant model, 137
American wire gauge (AWG), 26–28, 60–61
Ammeters, 49
Anarchic chaos, 140
AND circuit, 124
Arduino-compatible microcomputer, 67
AWG. *See* American wire gauge (AWG)

B

Barrier strip, 30–31
Bipolar junction transistors (BJTs), 83, 84
Bolts, 33
Budgeting, 14–15

C

Capacitors
  definition, 76
  electrolytic capacitors, 77
  schematic symbols, 78
  in series and parallel, 130
Charge status indicator, 155
Coin cell
  electrical goodness measurement, 94
  lithium
    installed backward, 92
    LED lights up, 91–92
Components, electric and electronic circuits
  acquiring components, 84
    new components purchase, 85
    recycle, used components, 85–86
  active components, 81
    batteries, 81
    diodes, 82–83
    integrated circuits, 84
    LEDs, 83
    power supplies, 81
    transformers, 82
    transistors, 83–84
  conductors and interconnects
    connectors, 64–68
    description, 57
    electrical schematics, 57–58
    wire and cables, 58–64
  data sheets, 86
  passive components
    capacitors, 76–78
    inductors, 78–79
    resistors, 68–76
    switches, 79–80
Computers, 35–36
Conductors and interconnects
  connectors
    clip leads, 65
    description, 64
    solderless breadboards, 66–68
  description, 57
  electrical schematics, 57–58
  wire and cables
    conductivity, 61–62
    description, 58–59
    longer version, 60–61
    Ohm's law, 62–63
    short version, 59–60
    storing, 63–64
    testing, 63
Connectors
  clip leads, 65
  description, 64
  solderless breadboards

Connectors, solderless breadboards (*cont.*)
        Arduino-compatible microcomputer, 67
        description, 66
        prototype LED display, 68
        tie points, 66
Cozy corner lab
    capacitors in series and parallel, 130
    inductors in series and parallel, 130
    LEDs, in series and parallel, 131–133
    parallel and series circuits
        assembly instructions, 119
        batteries in, 122–123
        components, 118
        in different view, 120–121
        schematic diagram, 118
        troubleshooting, 119–120
    rehabilitation, work station, 105–106
        components, 115–116
        divide, 107–109
        lighting, 112
        power source, 110–111
        taxonomy, 106–107
        tools, 112–114
        under-desk environment, 115
    resistors
        in parallel, 129
        in series, 126
    switches
        in parallel, 124
        in series, 123
        three-way, 125–126

D

Dallas Personal Robotics Group (DPRG), 17, 101–102, 139
DC voltmeter, 156–157
Digital caliper, 46, 47
Digital storage oscilloscope (DSO), 53–54
Digital voltage meter, 159
Diodes, 82–83
Double-throw switches, 80
DPRG. *See* Dallas Personal Robotics Group(DPRG)
Drawing, 23
Drills, 40–43
DSO. *See* Digital storage oscilloscope (DSO)

E, F, G

Electrical experiment
    LED
        defective, 95
        installed backward, 92–93
        lights up when attached to lithium coin cell, 91–92
        on solderless breadboard, 95–98
    parallel and series circuits, 117–118
Electrolytic capacitors, 77
Electronic measurement tools
    ammeters, 49
    ohmmeters, 49
    voltmeters, 47–48
Electronics lab
    and classroom (*see* Lab and classroom)
    budgeting, 14–15
    cautionary tale, 17
    electricity,working principles
        practical approach, 9
        theoretical approach, 7–8
    environmental impact, 14
    flashlight, 11–13
    fundamental components, 4–5
    goals, 6
    ideas, 5–6
    of Harold Timmis, 16
    planning, 2
    portable (*see* Portable mini-lab)
    Robot Clubhouse, 17
    safety, 10
    space for, 3
    tools, 3–4 (*see also* Tools)
Environmental impact, 14

H

Hackerspaces, 139
Hand tools, 114
    computers, 35–36
    pliers and tweezers, 29–30
    screwdriver, 30–33
    screws, bolts and nuts, 33
    wire crimpers, 34–35
    wire cutters, 23–25, 29
    wire strippers, 25–29
Harold Timmis' lab, 16

I, J, K

Indicators, 155
Inductors
  DC motors, 79
  description, 78
  usage, 79
Infrared LEDs, 132
Inspection equipment, 45–46
Insulators, 8
Integrated circuits, 84

L

Lab and classroom
  group
    coworkers and fellow students, 138
    family, 137
    friends, 138
    general public, 139–140
    special-interest groups, 139
  presentations, 143
  role
    facilitator/coordinator, 135
    observer, 137
    participant, 137
    teacher/presenter, 136
  shared resource
    courteous guest, 141
    hospitable host, 140–141
  workstations
    description, 142
    safety planning, 142
Laboratory set up. *See* Electronics lab
Laptop charger, 149
Laser cutters, 43
Light-emitting diodes (LEDs), 83
  care and feeding, 127–128
  circuit on solderless breadboard, 95–98
  curiosities, 126–127
  defective, 95
  installed backward, 92–93
  in series and parallel, 131–33
  lights up when attached to lithium coin cell, 91–92
  self-control, lack of, 127
  uses, 100

M

Magnifying lens, 46
Metal-oxide–semiconductor field-effect transistor (MOSFET), 84
Modern digital storage oscilloscope, 52–53
Mobile electronics lab. *See* Portable mini-lab
MOSFET. *See* Metal-oxide–semiconductor field-effect transistor (MOSFET)
Multimeters, 50–51, 94

N

Noisebridge organization, 140
Nuts, 33

O

Ohmmeters, 49
Ohm’s law
  description, 49
  application, 49, 62–63
OR circuit, 124
Oscilloscopes
  functions of, 51
  modern, 52–54
  tiny, 54–55

P, Q

Parallel and series circuits
  assembly instructions, 119
  batteries in, 122–123
  in different view, 120–121
  schematic diagram, 118
  troubleshooting, 119–120
Passive electrical and electronic components
  capacitors, 76–78
  inductors, 78–79
  resistors, 68–76
  switches, 79–80
Pencil-grip style, soldering irons, 39
Permanent circuit, 95–97
Physical measurement devices, 46–47
Pliers and tweezers, 29–30
Portable mini-lab
  advantages, 88
  compact executive model

Portable mini-lab, compact executive model (*cont.*)
        electrical experiments (*see* Electrical experiments)
        lab deployed on desktop, 89–90
    components, 89
    DPRG, 101–102
    features, 89
    robotics lab, 103
    travel, 87
Potentiometer, 74–75
Power-on indicator, 155
Power rail attachment, 99
Power supplies, 55–56, 81–82, 145–146
Power tools
    3D printers, 44
    drills, 40–43
    laser cutters, 43
Professional bench power supply, 146

R

Resistors
    color codes, 72–73
    description, 68
    location, 95
    in parallel, 129
    in permanent circuit, 97
    power ratings, 73
    schematic symbols and reference designators, 70–71
    in series, 126
    specialty resistors, 75–76
    standard values and tolerances, 71–72
    through-hole and surface-mount resistors, 69–70
    variable resistors, 74–75
RIGOL DS1052E oscilloscope, 54
Robot Clubhouse, 17

S

Shunt resistor, 158
Single-pole double-throw (SPDT) switch, 125
*Si*ngle-throw switches, 80
Soldering
    definition, 4
    instructions for, 37–38
    tools for, 113
Solderless breadboards, 66
    Arduino-compatible microcomputer, 67
    description, 66
    prototype LED display, 68
    tie points, 66
    types of, 117
SPDT switch. *See* Single-pole double-throw (SPDT) switch
Stereo microscope, 45
Surface-mount resistors, 69–70
Switches
    description, 79
    double-throw switch, 80
    in parallel, 124
    in series, 123
    single-throw switch, 80
    three-way, 125–126

T

Tablet computers, 91
Tabletop drill press, 42
Tektronix 475A oscilloscope, 52
Temperature-controlled soldering station, 113
Terminal block, 31–32
Test and measurement equipment
    multimeters, 50–51
    oscilloscopes
        functions of, 51
        modern, 52–54
        Tektronix 475A, 52
        tiny, 54–55
    power supplies, 55
3D printers, 44
Three-way switch, 125–126
Through-hole resistors, 69–70
Tools
    earmuffs, 22
    electronic measurement
        ammeters, 49
        ohmeters, 49
        voltmeters, 47–48
    eye protection, 21
    for soldering, 37–40
    hand
        computers, 35–36
        pliers and tweezers, 29–30
        protection, 22
        screwdriver, 30–33
        screws, bolts and nuts, 33
        wire crimpers, 34–35

        wire cutters, 23–25, 29
        wire strippers, 25–29
    hearing protection, 22
    indicators, 155
    inspection equipment, 45–46
    mind, 19–21
    physical measurement devices, 46–47
    power
        3D printers, 44
        drills, 40–43
        laser cutters, 43
    power-on indicator, 155
    power supply, 145–46
    USB charger
        description, 147–148
        voltage measurement, 149–153
Transformers, 82
Transistors
    BJTs, 83
    description, 83
    MOSFET, 84
Troubleshooting, 2

U

Unrepairable subassemblies, 127
USB charger
    cable
        colors and functions, 152
        internal wires, 151
    description, 147–148
    regulated power from, 155
    voltage measurement, 149–153

V

Variable resistors, 74–75
Voltmeters, 47–48

W, X, Y, Z

Weller controlled-temperature soldering station, 37
Wire and cables
    conductivity, 61–62
    description, 58–59
    longer version, 60–61
    Ohm’s law, 62–63
    short version, 59–60
    storing, 63–64
    testing, 63
Wire crimpers, 34–35
Wire cutters, 23–25, 29
Wire strippers, 25–29
Wood screws, 33
Workshop. *See* Electronics lab

Building Your Own Electronics Lab

■ ■ ■

Dale Wheat

Apress®

Building Your Own Electronics Lab

Copyright © 2012 by Dale Wheat

All rights reserved. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage or retrieval system, without the prior written permission of the copyright owner and the publisher.

ISBN-13 (pbk): 978-1-4302-4386-1

ISBN-13 (electronic): 978-1-4302-4387-8

Trademarked names, logos, and images may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, logo, or image we use the names, logos, and images only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark. The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

President and Publisher: Paul Manning
Lead Editor: Matt Moodie
Technical Reviewer: Cliff Wooton
Editorial Board: Steve Anglin, Mark Beckner, Ewan Buckingham, Gary Cornell, Morgan Ertel, Jonathan Gennick, Jonathan Hassell, Robert Hutchinson, Michelle Lowman, James Markham, Matthew Moodie, Jeff Olson, Jeffrey Pepper, Douglas Pundick, Ben Renow-Clarke, Dominic Shakeshaft, Gwenan Spearing, Matt Wade, Tom Welsh
Coordinating Editor: Brigid Duffy
Copy Editors: Damon Larson
Compositor: Bytheway Publishing Services
Indexer: SPi Global
Cover Designer: Anna Ishchenko

Distributed to the book trade worldwide by Springer Science+Business Media, LLC., 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail `orders-ny@springer-sbm.com`, or visit `www.springeronline.com`.

For information on translations, please e-mail `rights@apress.com`, or visit `www.apress.com`.

Apress and friends of ED books may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Special Bulk Sales–eBook Licensing web page at `http://www.apress.com/bulk-sales`.

The information in this book is distributed on an “as is” basis, without warranty. Although every precaution has been taken in the preparation of this work, neither the author(s) nor Apress shall have any liability to any person or entity with respect to any loss or damage caused or alleged to be caused directly or indirectly by the information contained in this work.

Any source code or other supplementary materials referenced by the author in this text is available to readers at `www.apress.com`. For detailed information about how to locate your book’s source code, go to `http://www.apress.com/source-code/`.

To my father, John H. Wheat, who patiently taught me about electricity and electronics over 40 years ago. To this day, I remain fascinated by it. My life has been enriched in so many ways by the many things that I learned from my father.

Contents

- **About the Author** ... **xi**
- **About the Technical Reviewer** ... **xii**
- **Acknowledgments** ... **xiii**
- **Introduction** ... **xiv**
- **Chapter 1: Planning Your Electronics Workshop** ... **1**
  - What to Expect ... 1
  - It Starts with a Plan ... 2
  - A Broad Outline ... 2
    - A Little Time ... 2
    - A Little Space ... 3
    - Basic Tools ... 3
    - Fundamental Components ... 4
    - Ideas ... 5
    - Goals ... 6
    - Inspiration! ... 7
  - How Electricity Works ... 7
    - The Theoretical Approach ... 7
    - A Practical Approach ... 9
  - Safety ... 10
  - Lighten Up! ... 11
  - Environmental Impact ... 14

Budgeting 14
Buying New 14
Finding Used. 15
Harvesting or Recycling. 15
Some Example Workshops 16
A Fresh Start.. 16
The Robot Clubhouse. 17
A Cautionary Tale. 17
Summary 18
■ **Chapter 2: Building Your Tool Chest 19**
The Most Important Tool Ever...19
Care and Feeding of Your Brain. 21
All Those Other Bits, Too 21
Basic Hand Tools 22
Wire Cutters and Wire Strippers. 23
Pliers and Tweezers 29
Screwdrivers, Nut Drivers, Wrenches, and Ratchets. 30
Screws, Bolts, and Nuts 33
More Specialized Hand Tools 34
Wire Crimpers.. 34
Computers 35
Tools for Soldering. 37
Power Tools 40
Drills and Drill Bits.. 40
Laser Cutters and 3D Printers 43
Basic Test and Measurement Equipment...44

Physical Measurement Devices 46
Simple Electronic Measurement Tools 47
Advanced Test and Measurement Equipment 49
Multimeters 50
Oscilloscopes 51
Power Supplies 55
Organizing Your Tools 56
Summary 56
Chapter 3: Components 57
Conductors and Interconnects 57
Electrical Schematics 57
Wire and Cables 58
Connectors 64
Passive Components 68
Resistors 68
Capacitors 76
Inductors 78
Switches 79
Active Components 81
Power Supplies 81
Diodes 82
LEDs 83
Transistors 83
Integrated Circuits 84
Acquiring Components 84
Buying New Components 85
Recycling Used Components 85

Component Data Sheets 86

Summary 86

■ **Chapter 4: A Portable Mini-Lab........ 87**

Have Lab, Will Travel 87

The Compact Executive Model........ 87

No Soldering Required........ 90

The Ever-Shrinking Computing Device........ 91

Electrical Experiments You Can Perform 91

Lighten Up, 20th Century Style........ 91

A Slightly More Permanent Circuit........ 95

Multiple Uses for the LED 100

The Special Project Portable Laboratory........ 101

Summary 104

■ **Chapter 5: The Cozy Corner Lab........ 105**

The Rehabilitation of a Small but Useful Area 105

Science to the Rescue!........ 106

Organization 107

The Great Divide 107

Adding a Power Source 110

A Clean, Well-Lighted Space 111

Tools........ 112

Under the Desk........ 115

Components........ 115

Where to Go from Here........ 116

Projects........ 116

Parallel and Series Circuits 117

Batteries in Series and Parallel 122

Switches in Series 123
Switches in Parallel 124
The Three-Way Switch 125
Resistors in Series 126
Resistors in Parallel 129
Capacitors in Series and Parallel 130
Inductors in Series and Parallel 130
LEDs in Series and Parallel 131
Summary 133
■ **Chapter 6: The Small Group Lab and Classroom 135**
Know Yourself 135
Facilitator or Coordinator 135
Teacher or Presenter 136
Participant 137
Observer 137
Know Your Group 137
Your Family 137
Your Friends 138
Coworkers and Fellow Students 138
Special-Interest Groups 139
The General Public 139
A Shared Resource 140
Be a Hospitable Host 140
Be a Courteous Guest 141
Workstations 142
Safety Planning 142
Presentations 143
Summary 143

■ **Appendix A: Getting Started with Tool Building........145**
A Very Simple Power Supply........145
An Easy 2.5 Watts........147
Adding Features........154
Various Indicators........155
Power-On Indicator........155
Power Supply Voltage........156
Power Supply Current........158
Summary........159
■ **Index........161**

About the Author

■ **Dale Wheat** is a full-time freelance writer, specializing in electronics and embedded systems. He has written several articles for technical and hobbyist magazines such as *Circuit Cellar*, O'Reilly's *MAKE* magazine, and *Elektor*. He teaches classes on electronics, microcontrollers, and soldering skills. He designs and sells DIY electronics kits from his web site, `http://dalewheat.com`. Before becoming a full-time writer, Dale consulted as a computer programmer and systems analyst for several companies, including IBM, MCI, and GTE (now Verizon). Dale is a two-term past president of the Dallas Personal Robotics Group, the world's oldest personal robotics club. He is a member of the National Honor Society, Phi Theta Kappa, and intends to continue his education as long as they keep the doors open. He lives with his wife, Anne, near Dallas, Texas.

About the Technical Reviewer

■ Cliff Wootton is a former Interactive TV systems architect at BBC News. The "News Loops" service developed there was nominated for a BAFTA and won a Royal Television Society Award for Technical Innovation. An invited speaker on pre-processing for video compression at the Apple WWDC conference. Taught post graduate MA students about real-world computing, multimedia, video compression, metadata and researching the deployment of next generation interactive TV systems based on open standards.

Currently working on R&D projects investigating new Interactive TV technologies, involved with MPEG standards working groups, writing more books on the topic and speaking at conferences when not lecturing on Multimedia at the University of the Arts in London.

Acknowledgments

This book was Michelle Lowman's idea. I just wrote down some of the words.

Corbin Collins and Brigid Duffy, coordinating editors at Apress, helped keep me on track with this project, despite all my best efforts. Cliff Wootton and Alan Smith provided excellent insight and suggestions to help the chapters along. The entire Apress editorial and production teams make what could be a hard job into a thing of joy.

My wife, Anne, continues to buy into the fantasy that I am a "professional writer," giving me the time and space I need to do this. Her patience, love, support, and trust make all the difference in the world.

My good neighbor and friend Jerry Johnson's fervent prayers and constant encouragement and enthusiasm really helped pull this project through the difficult times that were encountered. While I was complaining about "narrative pacing and clarity," he underwent brain surgery (twice) and chemotherapy for cancer, always with a smiling laugh and a twinkle in his eye.

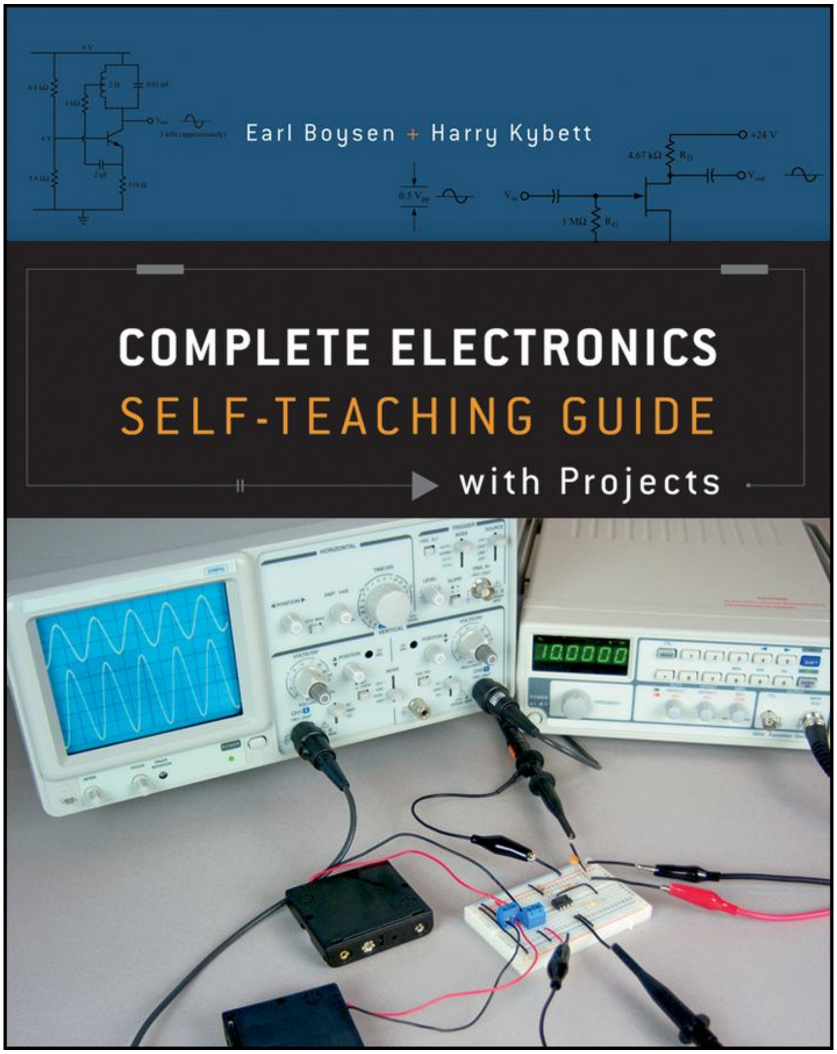
+24 V
4.67 kΩ
R_D
V_out
0.5 V_PP
V_in
1 MΩ
R_G
10.0000

Earl Boysen + Harry Kybett

COMPLETE ELECTRONICS SELF-TEACHING GUIDE

with Projects

Table of Contents

Chapter 1: DC Review and Pre-Test
Current Flow
Ohm's Law
Resistors in Series
Resistors in Parallel
Power
Small Currents
The Graph of Resistance
The Voltage Divider
The Current Divider
Switches
Capacitors in a DC Circuit
Summary
DC Pre-Test
Chapter 2: The Diode
Understanding Diodes
Diode Breakdown
The Zener Diode
Summary
Self-Test
Chapter 3: Introduction to the Transistor
Understanding Transistors
The Junction Field Effect Transistor (JFET)
Summary
Self-Test
Chapter 4: The Transistor Switch
Turning the Transistor On
Turning Off the Transistor
Why Transistors Are Used as Switches
The Three-Transistor Switch

Alternative Base Switching
Switching the JFET
Summary
Self-Test
Chapter 5: AC Pre-Test and Review
The Generator
Resistors in AC Circuits
Capacitors in AC Circuits
The Inductor in an AC Circuit
Resonance
Summary
Self-Test
Chapter 6: Filters
Capacitors in AC Circuits
Capacitors and Resistors in Series
Phase Shift of an RC Circuit
Resistor and Capacitor in Parallel
Inductors in AC Circuits
Phase Shift for an RL Circuit
Summary
Self-Test
Chapter 7: Resonant Circuits
The Capacitor and Inductor in Series
The Output Curve
Introduction to Oscillators
Summary
Self-Test
Chapter 8: Transistor Amplifiers
Working with Transistor Amplifiers
A Stable Amplifier
Biasing

The Emitter Follower
Analyzing an Amplifier
The JFET as an Amplifier
The Operational Amplifier
Summary
Self-Test
Chapter 9: Oscillators
Understanding Oscillators
Feedback
The Colpitts Oscillator
The Hartley Oscillator
The Armstrong Oscillator
Practical Oscillator Design
Simple Oscillator Design Procedure
Oscillator Troubleshooting Checklist
Summary and Applications
Self-Test
Chapter 10: The Transformer
Transformer Basics
Transformers in Communications Circuits
Summary and Applications
Self-Test
Answers to Self-Test
Chapter 11: Power Supply Circuits
Diodes in AC Circuits Produce Pulsating DC
Level DC (Smoothing Pulsating DC)
Summary
Self-Test
Chapter 12: Conclusion and Final Self-Test
Conclusion
Final Self-Test

Appendix A: Glossary
Appendix B: List of Symbols and Abbreviations
Appendix C: Powers of Ten and Engineering Prefixes
Appendix D: Standard Composition Resistor Values
Appendix E: Supplemental Resources
Web Sites
Books
Magazines
Suppliers
Appendix F: Equation Reference
Appendix G: Schematic Symbols Used in This Book
Introduction
What This Book Teaches
How This Book Is Organized
Conventions Used in This Book
How to Use This Book

Chapter 1

DC Review and Pre-Test

Electronics cannot be studied without first understanding the basics of electricity. This chapter is a review and pre-test on those aspects of direct current (DC) that apply to electronics. By no means does it cover the whole DC theory, but merely those topics that are essential to simple electronics.

This chapter reviews the following:

- Current flow
- Potential or voltage difference
- Ohm's law
- Resistors in series and parallel
- Power
- Small currents
- Resistance graphs
- Kirchhoff's Voltage Law
- Kirchhoff's Current Law
- Voltage and current dividers
- Switches
- Capacitor charging and discharging
- Capacitors in series and parallel

Current Flow

1 Electrical and electronic devices work because of an electric current.

Question

What is an electric current?

Answer

An *electric current* is a flow of electric charge. The electric charge usually consists of negatively charged electrons. However, in semiconductors, there are also positive charge carriers called *holes* .

2 There are several methods that can be used to generate an electric current.

Question

Write at least three ways an electron flow (or current) can be generated.

Answer

The following is a list of the most common ways to generate current:

- *Magnetically* —This includes the induction of electrons in a wire rotating within a magnetic field. An example of this would be generators turned by water, wind, or steam, or the fan belt in a car.
- *Chemically* —This involves the electrochemical generation of electrons by reactions between chemicals and electrodes (as in batteries).
- *Photovoltaic generation of electrons* —This occurs when light strikes semiconductor crystals (as in solar cells).

Less common methods to generate an electric current include the following:

- *Thermal generation* —This uses temperature differences between thermocouple junctions. Thermal generation is used in generators on spacecrafts that are fueled by radioactive material.

Electrochemical reaction —This occurs between hydrogen, oxygen, and electrodes (fuel cells).

Piezoelectrical —This involves mechanical deformation of piezoelectric substances. For example, piezoelectric material in the heels of shoes power LEDs that light up when you walk.

3 Most of the simple examples in this book contain a battery as the voltage source. As such, the source provides a potential difference to a circuit that enables a current to flow. An *electric current* is a flow of electric charge. In the case of a battery, electrons are the electric charge, and they flow from the terminal that has an excess number of electrons to the terminal that has a deficiency of electrons. This flow takes place in any complete circuit that is connected to battery terminals. It is this difference in the charge that creates the potential difference in the battery. The electrons try to balance the difference.

Because electrons have a negative charge, they actually flow from the negative terminal and return to the positive terminal. This direction of flow is called *electron flow* . Most books, however, use current flow, which is in the opposite direction. It is referred to as *conventional current flow* , or simply current flow. In this book, the term conventional current flow is used in all circuits.

Later in this book, you see that many semiconductor devices have a symbol that contains an arrowhead pointing in the direction of conventional current flow.

Questions

A. Draw arrows to show the current flow in Figure 1.1 . The symbol for the battery shows its polarity.

Figure 1.1

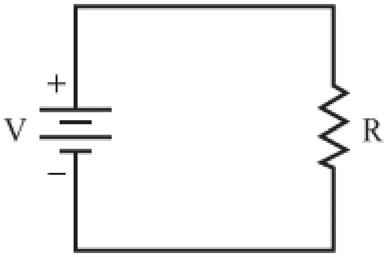

B. What indicates that a potential difference is present? __________

C. What does the potential difference cause? __________

D. What will happen if the battery is reversed? __________

Answers

A. See Figure 1.2 .

Figure 1.2

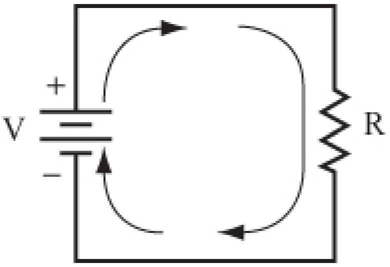

B. The battery symbol indicates that a difference of potential (also called *voltage*) is being supplied to the circuit.

C. Voltage causes current to flow if there is a complete circuit present, as shown in Figure 1.1 .

D. The current flows in the opposite direction.

Ohm's Law

4 Ohm's law states the fundamental relationship between voltage, current, and resistance.

Question

What is the algebraic formula for Ohm's law?

_____

Answer

$$V = I \times R$$

This is the most basic equation in electricity, and you should know it well. Some electronics books state Ohm's law as E = IR. E and V are both symbols for voltage. This book uses V to indicate voltage. When V is used after a number in equations and circuit diagrams, it represents volts, the unit of measurement of voltage. Also, in this formula, resistance is the opposition to current flow. Larger resistance results in smaller current for any given voltage.

5 Use Ohm's law to find the answers in this problem.

Questions

What is the voltage for each combination of resistance and current values?

A. R = 20 ohms, I = 0.5 amperes

V = _____

B. R = 560 ohms, I = 0.02 amperes

V = _____

C. R = 1,000 ohms, I = 0.01 amperes

V = _____

D. R = 20 ohms I = 1.5 amperes

V = _____

Answers

A. 10 volts

B. 11.2 volts

C. 10 volts

D. 30 volts

6 You can rearrange Ohm's law to calculate current values.

Questions

What is the current for each combination of voltage and resistance values?

A. V = 1 volt, R = 2 ohms

I = _____

B. V = 2 volts, R = 10 ohms

I = _____

C. V = 10 volts, R = 3 ohms

I = _____

D. V = 120 volts, R = 100 ohms

I = _____

Answers

A. 0.5 amperes

B. 0.2 amperes

C. 3.3 amperes

D. 1.2 amperes

7 You can rearrange Ohm's law to calculate resistance values.

Questions

What is the resistance for each combination of voltage and current values?

A. V = 1 volt, I = 1 ampere

R = _____

B. V = 2 volts, I = 0.5 ampere

R = _____

C. V = 10 volts, I = 3 amperes

R = _____

D. V = 50 volts, I = 20 amperes

R = _____

Answers

A. 1 ohm

B. 4 ohms

C. 3.3 ohms

D. 2.5 ohms

8 Work through these examples. In each case, two factors are given and you must find the third.

Questions

What are the missing values?

A. 12 volts and 10 ohms. Find the current. __________

B. 24 volts and 8 amperes. Find the resistance. __________

C. 5 amperes and 75 ohms. Find the voltage. _____

Answers

A. 1.2 amperes

B. 3 ohms

C. 375 volts

Inside the Resistor

Resistors are used to control the current that flows through a portion of a circuit. You can use Ohm's law to select the value of a resistor that gives you the correct current in a circuit. For a given voltage, the current flowing through a circuit increases when using smaller resistor values and decreases when using larger resistor values.

This resistor value works something like pipes that run water through a plumbing system. For example, to deliver the large water flow required by your water heater, you might use a 1-inch diameter pipe. To connect a bathroom sink to the water supply requires much smaller water flow and, therefore, works with a 1/2-inch pipe. In the same way, smaller resistor values (meaning less resistance) increase current flow, whereas larger resistor values (meaning more resistance) decrease the flow.

Tolerance refers to how precise a stated resistor value is. When you buy *fixed resistors* (in contrast to *variable resistors* that are used in some of the projects in this book), they have a particular resistance value. Their tolerance tells you how close to that value their resistance will be. For example, a 1,000-ohm resistor with ± 5 percent tolerance could have a value of anywhere from 950 ohms to 1,050 ohms. A 1,000-ohm resistor with ± 1 percent tolerance (referred to as a *precision resistor*) could have a value ranging anywhere from 990 ohms to 1,010 ohms. Although you are assured that the resistance of a precision resistor will be close to its stated value, the resistor with ± 1 percent tolerance costs more to manufacture and, therefore, costs you more than twice as much as a resistor with ± 5 percent.

Most electronic circuits are designed to work with resistors with ± 5 percent tolerance. The most commonly used type of resistor with ± 5 percent tolerance is called a *carbon film resistor* . This term refers to the manufacturing process in which a carbon film is deposited on an insulator. The thickness and width of the carbon film determines the resistance (the thicker the carbon film, the lower the resistance). Carbon film resistors work well in all the projects in this book.

On the other hand, precision resistors contain a metal film deposited on an insulator. The thickness and width of the metal film determines the resistance. These resistors are called *metal film resistors* and are used in circuits for precision devices such as test instruments.

Resistors are marked with four or five color bands to show the value and tolerance of the resistor, as illustrated in the following figure. The four-band color code is used for most resistors. As shown in the figure, by adding a fifth band, you get a five-band color code. Five-band color codes are used to provide more precise values in precision resistors.

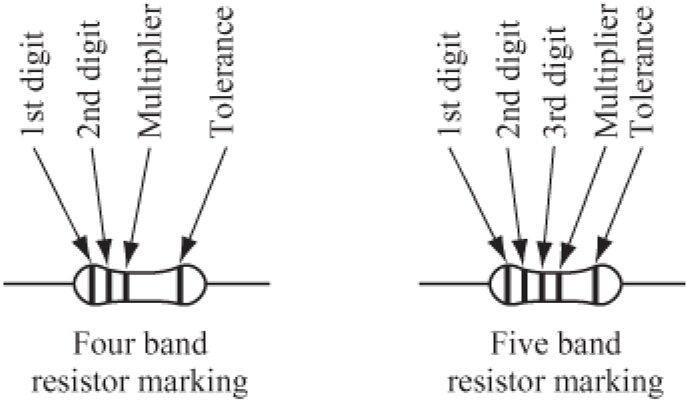
1st digit
2nd digit
Multiplier
Tolerance
Four band
resistor marking
1st digit
2nd digit
3rd digit
Multiplier
Tolerance
Five band
resistor marking

The following table shows the value of each color used in the bands:

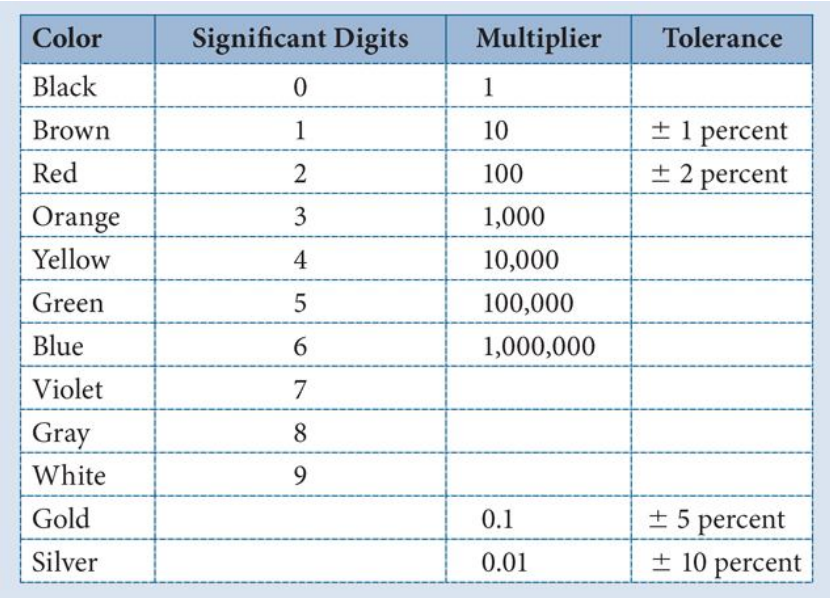

Color	Significant Digits	Multiplier	Tolerance
Black	0	1	
Brown	1	10	± 1 percent
Red	2	100	± 2 percent
Orange	3	1,000	
Yellow	4	10,000	
Green	5	100,000	
Blue	6	1,000,000	
Violet	7		
Gray	8		
White	9		
Gold		0.1	± 5 percent
Silver		0.01	± 10 percent

By studying this table, you can see how this code works. For example, if a resistor is marked with orange, blue, brown, and gold bands, its nominal resistance value is 360 ohms with a tolerance of ± 5 percent. If a resistor is marked with red, orange, violet, black, and brown, its nominal resistance value is 237 ohms with a tolerance of ± 1 percent.

Resistors in Series

9 You can connect resistors in series. Figure 1.3 shows two resistors in series.

Figure 1.3

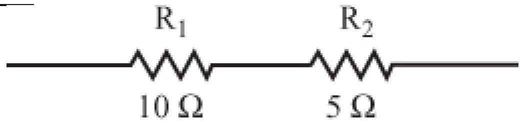

Question

What is their total resistance? _____

Answers

$$R_T = R_1 + R_2 = 10\text{ ohms} + 5\text{ ohms} = 15\text{ ohms}$$

The total resistance is often called the *equivalent series resistance* and is denoted as R eq .

Resistors in Parallel

10 You can connect resistors in parallel, as shown in Figure 1.4 .

Figure 1.4

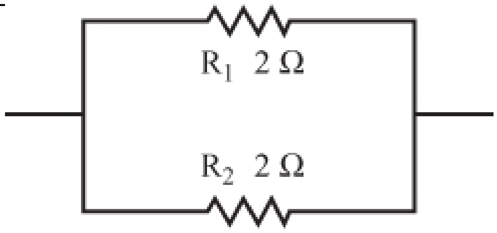

Question

What is the total resistance here? _____

Answers

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} = \frac{1}{2} + \frac{1}{2} = 1;\text{ thus } R_T = 1\text{ ohm}$$

R T is often called the *equivalent parallel resistance* .

11 The simple formula from problem 10 can be extended to include as many resistors as wanted.

Question

What is the formula for three resistors in parallel? _____

Answers

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

You often see this formula in the following form:

$$R_T = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}}$$

12 In the following exercises, two resistors are connected in parallel.

Questions

What is the total or equivalent resistance?

A. R_1 = 1 ohm, R_2 = 1 ohm

R_T = _____

B. R_1 = 1,000 ohms, R_2 = 500 ohms

R_T = _____

C. R_1 = 3,600 ohms, R_2 = 1,800 ohms

R_T = _____

Answers

A. 0.5 ohms

B. 333 ohms

C. 1,200 ohms

R_T is always smaller than the smallest of the resistors in parallel.

Power

13 When current flows through a resistor, it dissipates power, usually in the form of heat.

Power is expressed in terms of watts.

Question

What is the formula for power? _____

Answers

There are three formulas for calculating power:

$$P = VI \text{ or } P = I^2R \text{ or } P = \frac{V^2}{R}$$

14 The first formula shown in problem 13 allows power to be calculated when only the voltage and current are known.

Questions

What is the power dissipated by a resistor for the following voltage and current values?

A. V = 10 volts, I = 3 amperes

P = _____

B. V = 100 volts, I = 5 amperes

P = _____

C. V = 120 volts, I = 10 amperes

P = _____

Answers

A. 30 watts.

B. 500 watts, or 0.5 kW. (The abbreviation kW indicates kilowatts.)

C. 1,200 watts, or 1.2 kW.

15 The second formula shown in problem 13 allows power to be calculated when only the current and resistance are known.

Questions

What is the power dissipated by a resistor given the following resistance and current values?

A. R = 20 ohm, I = 0.5 ampere

P = _____

B. R = 560 ohms, I = 0.02 ampere

P = _____

C. V = 1 volt, R = 2 ohms

P = _____

D. V = 2 volt, R = 10 ohms

P = _____

Answers

A. 5 watts

B. 0.224 watts

C. 0.5 watts

D. 0.4 watts

16 Resistors used in electronics generally are manufactured in standard values with regard to resistance and power rating. Appendix D shows a table of standard resistance values for 0.25- and 0.05-watt resistors. Quite often, when a certain resistance value is needed in a circuit, you must choose the closest standard value. This is the case in several examples in this book.

You must also choose a resistor with the power rating in mind. Never place a resistor in a circuit that requires that resistor to dissipate more power than its rating specifies.

Questions

If standard power ratings for carbon film resistors are 1/8, 1/4, 1/2, 1, and 2 watts, what power ratings should be selected for the resistors that were used for the calculations in problem 15?

A. For 5 watts _____

B. For 0.224 watts _____

C. For 0.5 watts _____

D. For 0.4 watts _____

Answers

A. 5 watt (or greater)

B. 1/4 watt (or greater)

C. 1/2 watt (or greater)

D. 1/2 watt (or greater)

Most electronics circuits use low-power carbon film resistors. For higher-power levels (such as the 5-watt requirement in question A), other types of resistors are available.

Small Currents

17 Although currents much larger than 1 ampere are used in heavy industrial equipment, in most electronic circuits, only fractions of an ampere are required.

Questions

A. What is the meaning of the term *milliampere* ? __________

B. What does the term *microampere* mean?

____

Answers

A. A milliampere is one-thousandth of an ampere (that is, 1/1000 or 0.001 amperes). It is abbreviated mA.

B. A microampere is one-millionth of an

ampere (that is, 1/1,000,000 or 0.000001 amperes). It is abbreviated μA.

18 In electronics, the values of resistance normally encountered are quite high. Often, thousands of ohms and occasionally even millions of ohms are used.

Questions

A. What does kΩ mean when it refers to a resistor? __________

B. What does MΩ mean when it refers to a resistor? __________

Answers

A. Kilohm (k = kilo, Ω = ohm). The resistance value is thousands of ohms. Thus, 1 kΩ = 1,000 ohms, 2 kΩ = 2,000 ohms, and 5.6 kΩ = 5,600 ohms.

B. Megohm (M = mega, Ω = ohm). The resistance value is millions of ohms. Thus, 1 MΩ = 1,000,000 ohms, and 2.2 MΩ = 2,200,000 ohms.

19 The following exercise is typical of many performed in transistor circuits. In this example, 6 volts is applied across a resistor, and 5 mA of current is required to flow through the resistor.

Questions

What value of resistance must be used and what power will it dissipate?

R = _____ P = _____

Answers

$$R = \frac{V}{I} = \frac{6\text{ volts}}{5\text{ mA}} = \frac{6}{0.005} = 1200\text{ ohms} = 1.2\,\text{k}\Omega$$

$$P = V \times I = 6 \times 0.005 = 0.030\text{ watts} = 30\,\text{mW}$$

20 Now, try these two simple examples.

Questions

What is the missing value?

A. 50 volts and 10 mA. Find the resistance. __________

B. 1 volt and 1 MΩ. Find the current. __________

Answers

A. 5 kΩ

B. 1 μA

The Graph of Resistance

21 The voltage drop across a resistor and the current flowing through it can be plotted on a simple graph. This graph is called a *V-I curve* .

Consider a simple circuit in which a battery is connected across a 1 kΩ resistor.

Questions

A. Find the current flowing if a 10-volt battery is used. __________

B. Find the current when a 1-volt battery is used. __________

C. Now find the current when a 20-volt battery is used. __________

Answers

A. 10 mA

B. 1 mA

C. 20 mA

22 Plot the points of battery voltage and current flow from problem 21 on the graph shown in Figure 1.5, and connect them together.

Figure 1.5

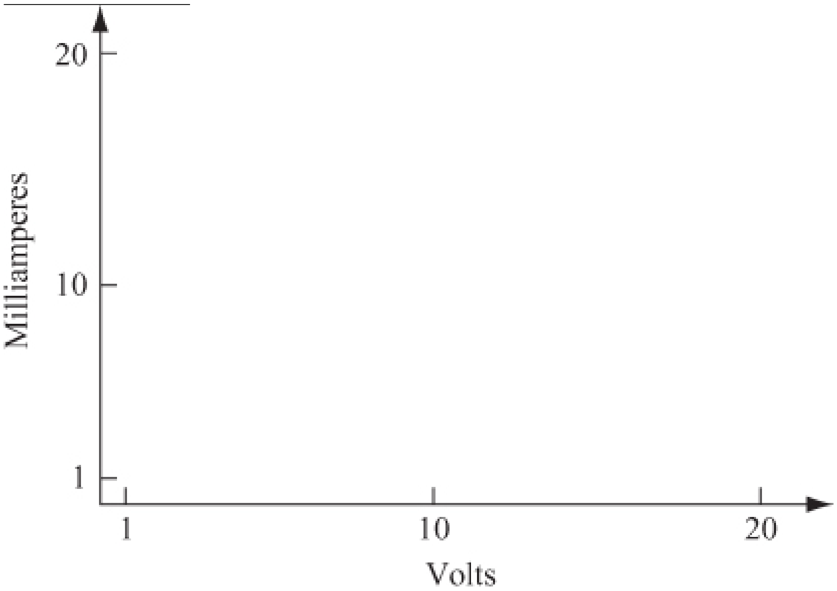

Question

What would the slope of this line be equal to?

_____

Answers

You should have drawn a straight line, as in the graph shown in Figure 1.6.

Figure 1.6

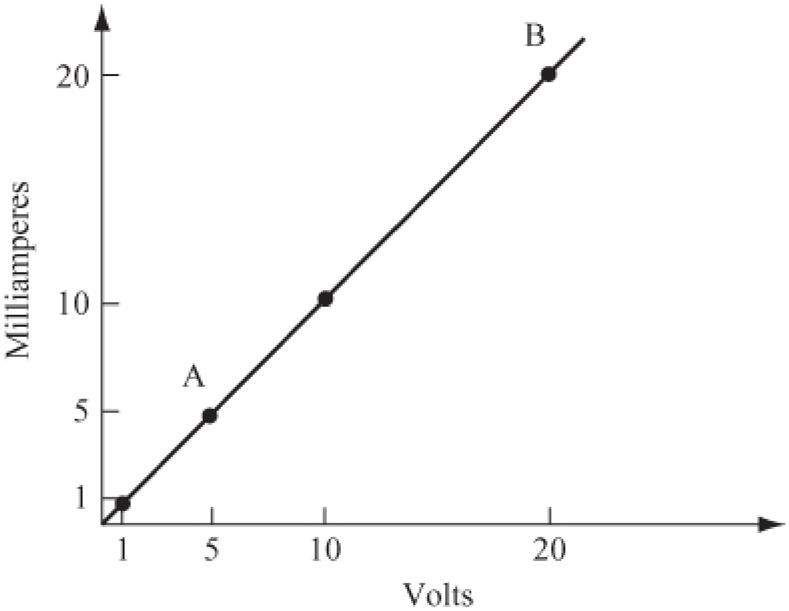

Sometimes you need to calculate the slope of the line on a graph. To do this, pick two points and call them A and B.

For point A, let V = 5 volts and I = 5 mA

For point B, let V = 20 volts and I = 20 mA

The slope can be calculated with the following formula:

$$\text{Slope} = \frac{V_B - V_A}{I_B - I_A} = \frac{20\text{ volts} - 5\text{ volts}}{20\text{ mA} - 5\text{ mA}} = \frac{15\text{ volts}}{15\text{ mA}} = \frac{15\text{ volts}}{0.015\text{ ampere}} = 1\text{ k}\,\Omega$$

In other words, the slope of the line is equal to the resistance.

Later, you learn about V-I curves for other components. They have several uses, and often they are not straight lines.

The Voltage Divider

23 The circuit shown in Figure 1.7 is called a *voltage divider*. It is the basis for many important theoretical and practical ideas you encounter throughout the entire field of electronics.

Figure 1.7

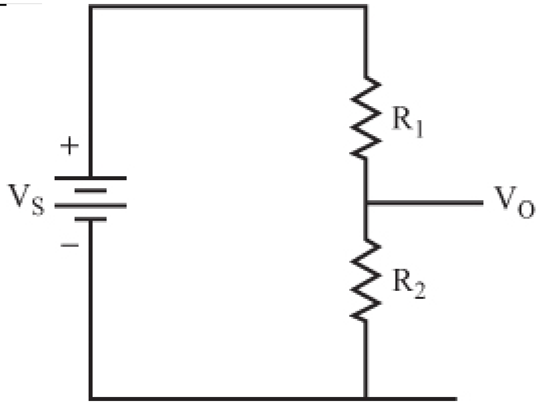

The object of this circuit is to create an output voltage (V_0) that you can control based upon the two resistors and the input voltage. V_0 is also the *voltage drop* across R_2.

Question

What is the formula for V_0? _____

Answers

$$V_o = V_S \times \frac{R_2}{R_1 + R_2}$$

$R_1 + R_2 = R_T$, the total resistance of the circuit.

24 A simple example can demonstrate the use of this formula.

Question

For the circuit shown in Figure 1.8, what is V

0 ? _____

Figure 1.8

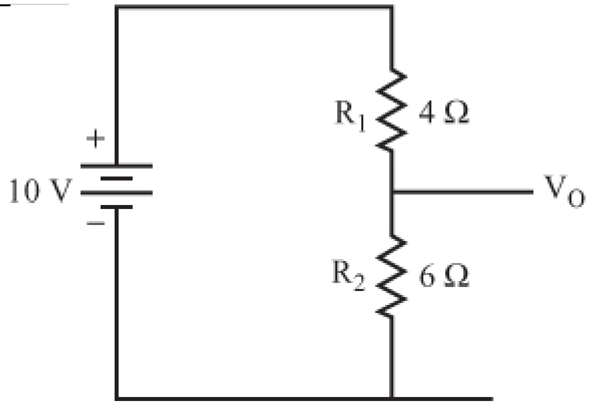

Answers

$$V_O = V_S \times \frac{R_2}{R_1+R_2}$$

$$= 10 \times \frac{6}{4+6}$$

$$= 10 \times \frac{6}{10}$$

$$= 6 \text{ volts}$$

25 Now, try these problems.

Questions

What is the output voltage for each combination of supply voltage and resistance?

A. V_S = 1 volt, R_1 = 1 ohm, R_2 = 1 ohm

V_0 = _____

B. V_S = 6 volts, R_1 = 4 ohms, R_2 = 2 ohms

V_0 = _____

C. V_S = 10 volts, R_1 = 3.3. kΩ, R_2 = 5.6 kΩ

V_0 = _____

D. V_S = 28 volts, R_1 = 22 kΩ, R_2 = 6.2 kΩ

V_0 = _____

Answers

A. 0.5 volts

B. 2 volts

C. 6.3 volts

D. 6.16 volts

26 The output voltage from the voltage divider is always less than the applied voltage. Voltage dividers are often used to apply specific voltages to different components in a circuit. Use the voltage divider equation to answer the following questions.

Questions

A. What is the voltage drop across the 22 k Ω resistor for question D of problem 25? __________

B. What total voltage do you get if you add this voltage drop to the voltage drop across the 6.2 k Ω resistor? __________

Answers

A. 21.84 volts

B. The sum is 28 volts.

The voltages across the two resistors add up to the supply voltage. This is an example of *Kirchhoff's Voltage Law (KVL)* , which simply means that the voltage supplied to a circuit must equal the sum of the voltage drops in the circuit. In this book, KVL is often used without actual reference to the law.

Also the voltage drop across a resistor is proportional to the resistor's value. Therefore, if one resistor has a greater value than another in a series circuit, the voltage drop across the higher-value resistor is greater.

Using Breadboards

A convenient way to create a prototype of an electronic circuit to verify that it works is to build it on a *breadboard* . You can use breadboards to build the circuits used in the projects later in this book. As shown in the following figure, a breadboard is a sheet of plastic with several contact holes. You use these holes to connect electronic components in a circuit. After you verify that a circuit works with this method, you can then create a permanent circuit using soldered connections.

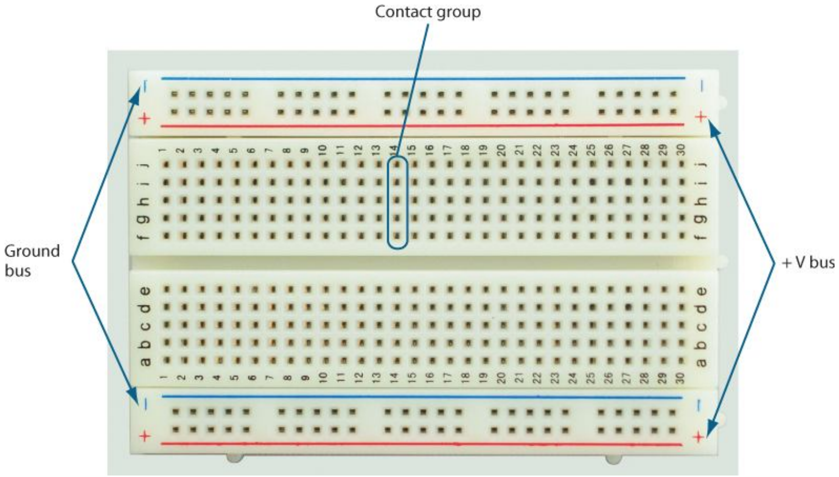
Contact group
Ground bus
+V bus
a b c d e
f g h i j
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

Breadboards contain metal strips arranged in a pattern under the contact holes, which are used to connect groups of contacts together. Each group of five contact holes in a vertical line (such as the group circled in the figure) is connected by these metal strips. Any components plugged into one of these five contact holes are, therefore, electrically connected.

Each row of contact holes marked by a "+" or "–" are connected by these metal strips. The rows marked "+" are connected to the positive terminal of the battery or power supply and are referred to as the *+V bus*. The rows marked "–" are connected to the negative terminal of the battery or power supply and are referred to as the *ground bus*. The 1V buses and ground buses running along the top and bottom of the breadboard make it easy to connect any component in a circuit with a short piece of wire called a *jumper wire*. Jumper wires are typically made of 22-gauge solid wire with approximately 1/4 inch of insulation stripped off each end.

The following figure shows a voltage divider circuit assembled on a breadboard. One end of R1 is inserted into a group of contact holes that is also connected by a jumper wire to the 1V bus. The other end of R1 is inserted into the same group of contact holes that contains one end of R2. The other end of R2 is inserted into a group of contact holes that is also connected by a jumper wire to the ground bus. In this example, a 1.5 kΩ resistor was used for R1, and a 5.1 kΩ resistor was used for R2.

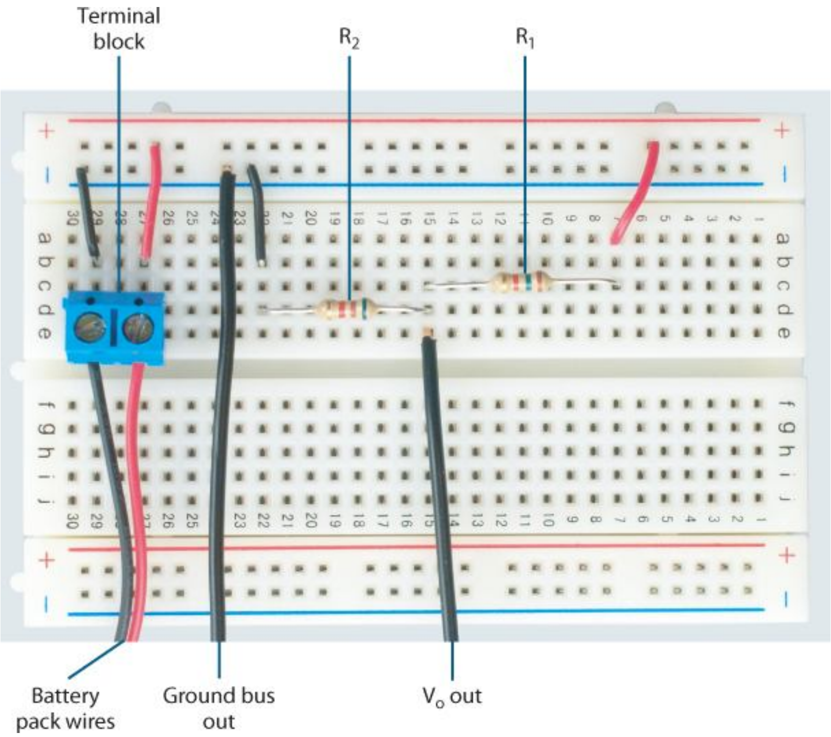

A *terminal block* is used to connect the battery pack to the breadboard because the wires supplied with battery packs (which are stranded wire) can't be inserted directly into breadboard contact holes. The red wire from a battery pack is attached to the side of the terminal block that is inserted into a group of contact holes, which is also connected by a jumper wire to the 1V bus. The black wire from a battery pack is attached to the side of the terminal block that is inserted into a group of contact holes, which is also connected by a jumper wire to the ground bus.

To connect the output voltage, Vo, to a multimeter or a downstream circuit, two additional connections are needed. One end of a jumper wire is inserted in the same group of contact holes that contain both R 1 and R 2 to supply Vo. One end of another jumper wire is inserted in a contact hole in the ground bus to provide an electrical contact to the negative side of the battery. When connecting test equipment to the breadboard, you should use a 20-gauge jumper wire because sometimes the 22-gauge wire is pulled out of the board when attaching test probes.

The Current Divider

27 In the circuit shown in Figure 1.9 , the current splits or divides between the two resistors that are connected in parallel.

Figure 1.9

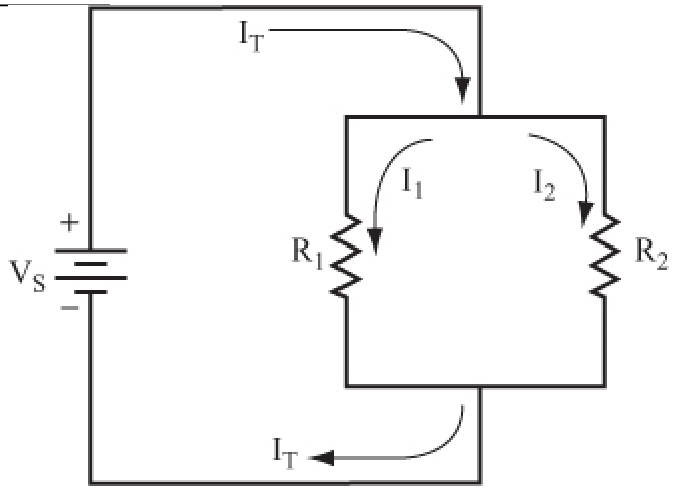

I_T splits into the individual currents I_1 and I_2, and then these recombine to form I_T.

Questions

Which of the following relationships are valid for this circuit?

A. $V_S = R_1 I_1$

B. $V_S = R_2 I_2$

C. $R_1 I_1 = R_2 I_2$

D. $I_1/I_2 = R_2/R_1$

Answers

All of them are valid.

28 When solving current divider problems, follow these steps:

1. Set up the ratio of the resistors and currents:

$I_1/I_2 = R_2/R_1$

2. Rearrange the ratio to give I_2 in terms of I_1:

$$I_2 = I_1 \times \frac{R_1}{R_2}$$

3. From the fact that $I_T = I_1 + I_2$, express I_T in terms of I_1 only.

4. Now, find I_1.

5. Now, find the remaining current (I_2).

Question

The values of two resistors in parallel and the total current flowing through the circuit are shown in Figure 1.10. What is the current through each individual resistor?

Figure 1.10

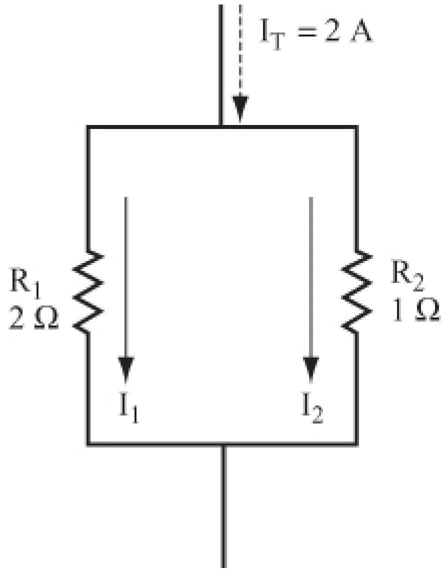

Answers

Work through the steps as shown here:

1. $I_1/I_2 = R_2/R_1 = 1/2$
2. $I_2 = 2I_1$
3. $I_T = I_1 + I_2 = I_1 + 2I_1 = 3I_1$
4. $I_1 = I_T/3 = 2/3$ ampere
5. $I_2 = 2I_1 = 4/3$ amperes

29 Now, try these problems. In each case, the total current and the two resistors are given. Find I_1 and I_2.

Questions

A. $I_T = 30$ mA, $R_1 = 12$ kΩ, $R_2 = 6$ kΩ

____

B. $I_T = 133$ mA, $R_1 = 1$ kΩ, $R_2 = 3$ kΩ

____

C. What current do you get if you add I_1 and I_2? __________

Answers

A. I_1 = 10 mA, I_2 = 20 mA
B. I_1 = 100 mA, I_2 = 33 mA
C. They add back together to give you the total current supplied to the parallel circuit.

Question C is actually a demonstration of *Kirchhoff's Current Law (KCL)* . Simply stated, this law says that the total current entering a junction in a circuit must equal the sum of the currents leaving that junction. This law is also used on numerous occasions in later chapters. KVL and KCL together form the basis for many techniques and methods of analysis that are used in the application of circuit analysis.

Also, the current through a resistor is inversely proportional to the resistor's value. Therefore, if one resistor is larger than another in a parallel circuit, the current flowing through the higher value resistor is the smaller of the two. Check your results for this problem to verify this.

30 You can also use the following equation to calculate the current flowing through a resistor in a two-branch parallel circuit:

$$I_1 = \frac{(I_T)(R_2)}{(R_1+R_2)}$$

Question

Write the equation for the current I_2 . _____
Check the answers for the previous problem using these equations.

Answer

$$I_2 = \frac{(I_T)(R_2)}{(R_1 + R_2)}$$

The current through one branch of a two-branch circuit is equal to the total current times the resistance of the opposite branch, divided by the sum of the resistances of both branches. This is an easy formula to remember.

Using the Multimeter

A *multimeter* is a must-have testing device for anyone's electronics toolkit. A multimeter is aptly named because it can be used to measure multiple parameters. Using a multimeter, you can measure current, voltage, and resistance by setting the rotary switch on the multimeter to the parameter you want to measure, and connecting each mulitmeter probe to a wire in a circuit. The following figure shows a multimeter connected to a voltage divider circuit to measure voltage. Following are the details of how you take each of these measurements.

Voltage

To measure the voltage in the circuit shown in the figure, at the connection between R_1 and R_2, use jumper wire to connect the red probe of a multimeter to the row of contact holes containing leads from both R_1 and R_2. Use another jumper wire to connect the black probe of the multimeter to the ground bus. Set the rotary switch on the multimeter to measure voltage, and it returns the results.

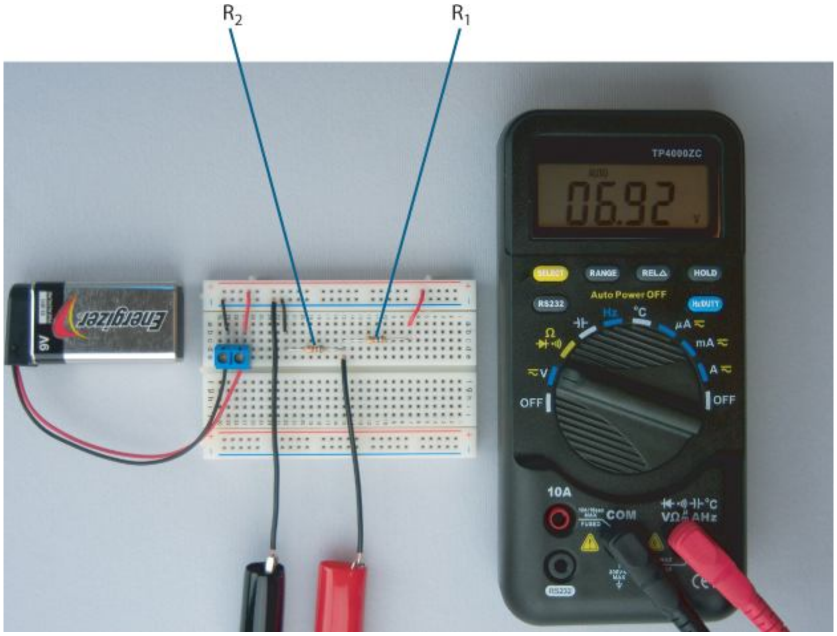
R_2
R_1
TP4000ZC
06.92
Energizer
9V
SELECT
RANGE
REL△
HOLD
RS232
Auto Power OFF
Hz/DUTY
OFF
OFF
10A
COM

Note The circuit used in a multimeter to measure voltage places a large-value resistor in parallel with R 2 so that the test itself does not cause any measurable drop in the current passing through the circuit.

Tip Whenever you perform tests on a circuit, attach alligator clips or test clips with plastic covers to the ends of the probes. This aids the probes in grabbing the jumper wires with little chance that they'll cause a short.

Current

The following figure shows how you connect a multimeter to a voltage divider circuit to measure current. Connect a multimeter in series with components in the circuit, and set the rotary switch to the appropriate ampere range, depending upon the magnitude of the expected current. To connect the multimeter in series with R1 and R2, use a jumper wire to connect the +V bus to the red lead of a multimeter, and another jumper wire to connect the black lead of the multimeter to R1. These connections force the current flowing through the circuit to flow through the multimeter.

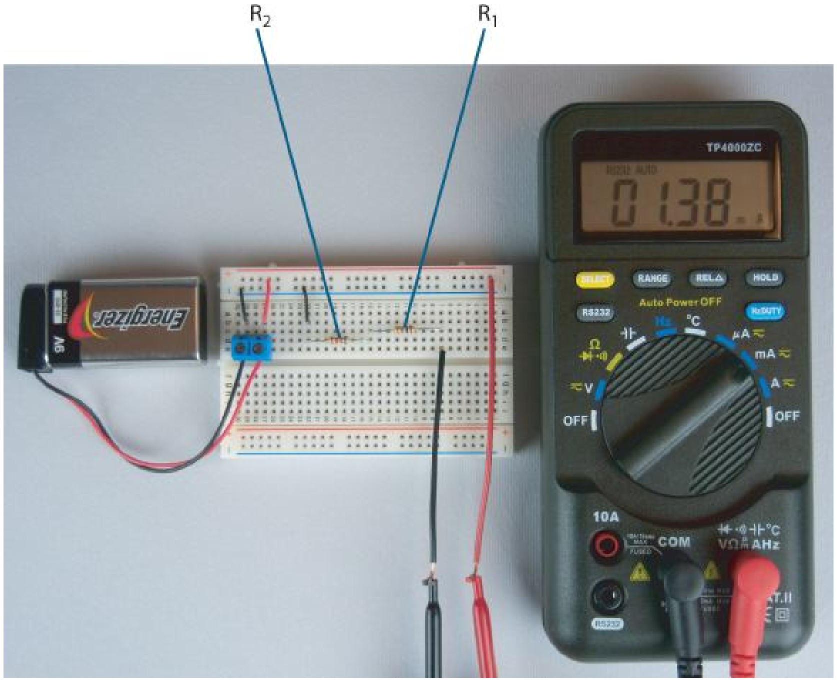
R2
R1
TP4000ZC
01.38
SELECT
RANGE
REL△
HOLD
RS232
Auto Power OFF
Hz/DUTY
Ω
Hz
°C
μA
mA
A
V
OFF
OFF
10A
COM
VΩ
AHz
9V
Energizer

Note The circuit used in a multimeter to measure current passes the current through a low-value resistor so that the test itself does not cause any measurable drop in the current.

Resistance

You typically use the resistance setting on a multimeter to check the resistance of individual components. For example, in measuring the resistance of R 2 before assembling the circuit shown in the previous figure, the result was 5.0 kΩ, slightly off the nominal 5.1 kΩ stated value.

You can also use a multimeter to measure the resistance of a component in a circuit. A multimeter measures resistance by applying a small current through the components being tested, and measuring the voltage drop. Therefore, to prevent false readings, you should disconnect the battery pack or power supply from the circuit before using the multimeter.

Switches

31 A *mechanical switch* is a device that completes or breaks a circuit. The most familiar use is that of applying power to turn a device on or off. A switch can also permit a signal to pass from one place to another, prevent its passage, or route a signal to one of several places.

In this book, you work with two types of switches. The first is the simple on-off switch, also called a *single pole single throw* switch.

The second is the *single pole double throw* switch. Figure 1.11 shows the circuit symbols for each.

Figure 1.11

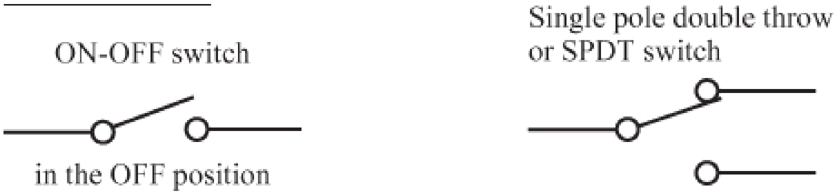
ON-OFF switch
in the OFF position
Single pole double throw
or SPDT switch

Keep in mind the following two important facts about switches:

A closed (or ON) switch has the total circuit current flowing through it. There is no voltage drop across its terminals.

An open (or OFF) switch has no current flowing through it. The full circuit voltage appears between its terminals.

The circuit shown in Figure 1.12 includes a closed switch.

Figure 1.12

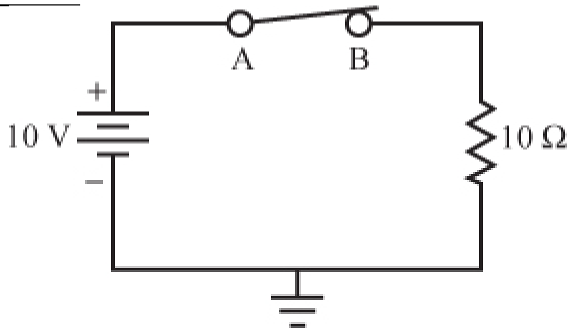

Questions

A. What is the current flowing through the switch? __________

B. What is the voltage at point A and point B with respect to ground? __________

C. What is the voltage drop across the switch? __________

Answers

A. $$\frac{10\text{ volts}}{10\text{ ohms}}=1\text{ ampere}$$

B. V A = V B = 10 volts

C. 0 V (There is no voltage drop because both terminals are at the same voltage.)

32 The circuit shown in Figure 1.13 includes an open switch.

Figure 1.13

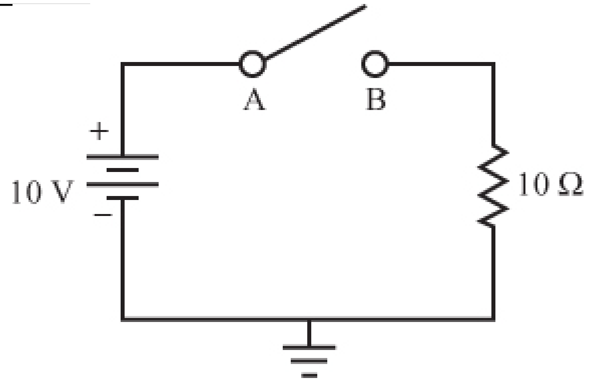

Questions

A. What is the voltage at point A and point B? __________

B. How much current is flowing through the switch? __________

C. What is the voltage drop across the switch? __________

Answers

A. V A = 10 volts; V B = 0 volts.

B. No current is flowing because the switch is open.

C. 10 volts. If the switch is open, point A is the same voltage as the positive battery terminal, and point B is the same voltage as the negative battery terminal.

33 The circuit shown in Figure 1.14 includes a single pole double throw switch. The position of the switch determines whether lamp A or lamp B is lit.

Figure 1.14

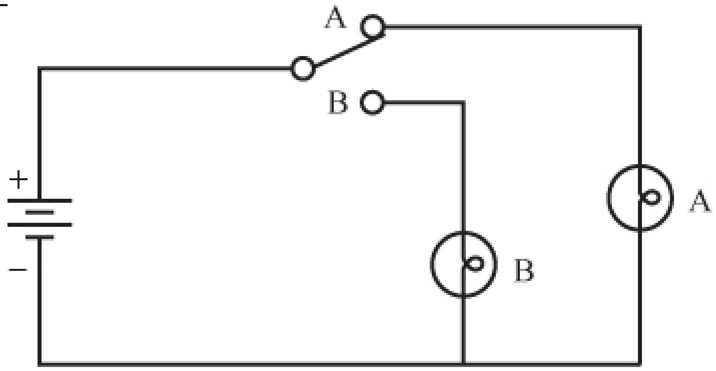

Questions

A. In the position shown, which lamp is lit? __________

B. Can both lamps be lit simultaneously? __________

Answers

A. Lamp A.

B. No, one or the other must be off.

Capacitors in a DC Circuit

34 Capacitors are used extensively in electronics. They are used in both alternating

current (AC) and DC circuits. Their main use in DC electronics is to become charged, hold the charge, and, at a specific time, release the charge.

The capacitor shown in Figure 1.15 charges when the switch is closed.

Figure 1.15

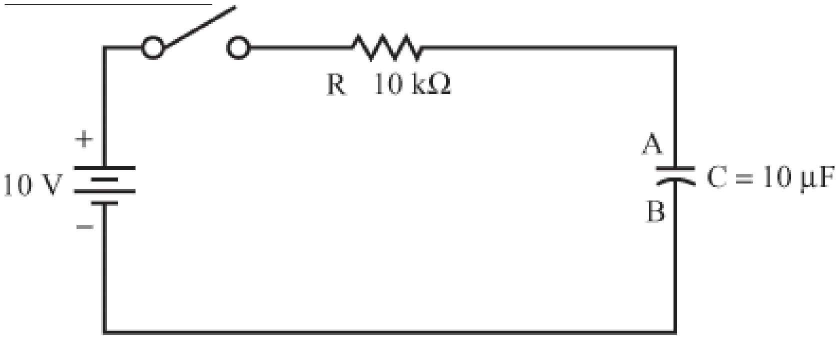

Question

To what final voltage will the capacitor charge? _____

Answers

It will charge up to 10 volts. It will charge up to the voltage that would appear across an open circuit located at the same place where the capacitor is located.

35 How long does it take to reach this voltage? This is an important question with many practical applications. To find the answer you must know the *time constant* (τ) (Greek letter tau) of the circuit.

Questions

A. What is the formula for the time constant of this type of circuit? __________

B. What is the time constant for the circuit shown in Figure 1.15 ? __________

C. How long does it take the capacitor to reach 10 volts? __________

D. To what voltage level does it charge in one time constant? __________

Answers

A. τ = R × C.

B. τ = 10 kΩ × 10 μF = 10,000 Ω × 0.00001 F = 0.1 seconds. (Convert resistance values to ohms and capacitance values to farads for this calculation.)

C. Approximately 5 time constants, or about 0.5 seconds.

D. 63 percent of the final voltage, or about 6.3 volts.

36 The capacitor does not begin charging until the switch is closed. When a capacitor is uncharged or discharged, it has the same voltage on both plates.

Questions

A. What is the voltage on plate A and plate B of the capacitor in Figure 1.15 before the switch is closed? __________

B. When the switch is closed, what happens to the voltage on plate A? __________

C. What happens to the voltage on plate B? __________

D. What is the voltage on plate A after one time constant? __________

Answers

A. Both will be at 0 volts if the capacitor is totally discharged.

B. It will rise toward 10 volts.

C. It will stay at 0 volts.

D. About 6.3 volts.

37 The *capacitor charging graph* in Figure 1.16 shows how many time constants a voltage must be applied to a capacitor before it reaches a given percentage of the applied voltage.

Figure 1.16

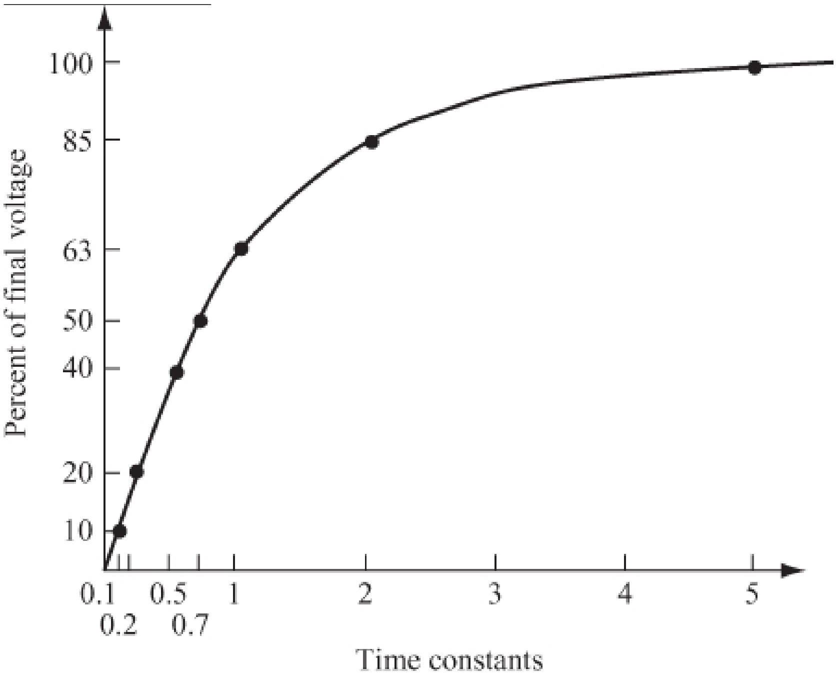

Questions

A. What is this type of curve called? __________

B. What is it used for? __________

Answers

A. It is called an *exponential* curve.

B. It is used to calculate how far a capacitor has charged in a given time.

38 In the following examples, a resistor and a capacitor are in series. Calculate the time constant, how long it takes the capacitor to fully charge, and the voltage level after one time constant if a 10-volt battery is used.

Questions

A. R = 1 kΩ, C = 1,000 μF __________

B. R = 330 kΩ, C = 0.05 μF __________

Answers

A. τ = 1 second; charge time = 5 seconds; V C = 6.3 volts.

B. τ = 16.5 ms; charge time = 82.5 ms; V C = 6.3 volts. (The abbreviation “ms” indicates milliseconds.)

39 The circuit shown in Figure 1.17 uses a double pole switch to create a discharge path for the capacitor.

Figure 1.17

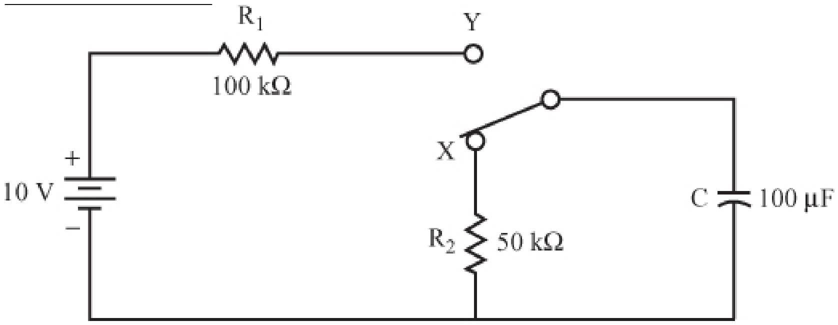

Questions

A. With the switch in position X, what is the voltage on each plate of the capacitor?

____

B. When the switch is moved to position Y, the capacitor begins to charge. What is its charging time constant? __________

C. How long does it take to fully charge the capacitor?

Answers

A. 0 volts

B. $\tau = R \times C = (100\ \mathrm{k}\Omega)\ (100\ \mu\mathrm{F}) = 10$ secs

C. Approximately 50 seconds

40 Suppose that the switch shown in Figure 1.17 is moved back to position X after the capacitor is fully charged.

Questions

A. What is the discharge time constant of the capacitor? __________

B. How long does it take to fully discharge the capacitor? __________

Answers

A. $\tau = R \times C = (50\ \mathrm{k}\Omega)\ (100\ \mu\mathrm{F}) = 5$ seconds (discharging through the 50 kΩ resistor)

B. Approximately 25 seconds

The circuit powering a camera flash is an example of a capacitor's capability to store a charge and then discharge upon demand. While you wait for the flash unit to charge, the camera uses its battery to charge a capacitor. When the capacitor is charged, it holds that charge until you click the Shutter button, causing the capacitor to discharge, which powers the flash.

Inside the Capacitor

Capacitors store an electrical charge on conductive plates that are separated by an insulating material, as shown in the following figure. One of the most common types of capacitor is a *ceramic capacitor* , which has values ranging from a few μF up to approximately 47 μF. The name for a ceramic capacitor comes from the use of a ceramic material to provide insulation between the metal plates.

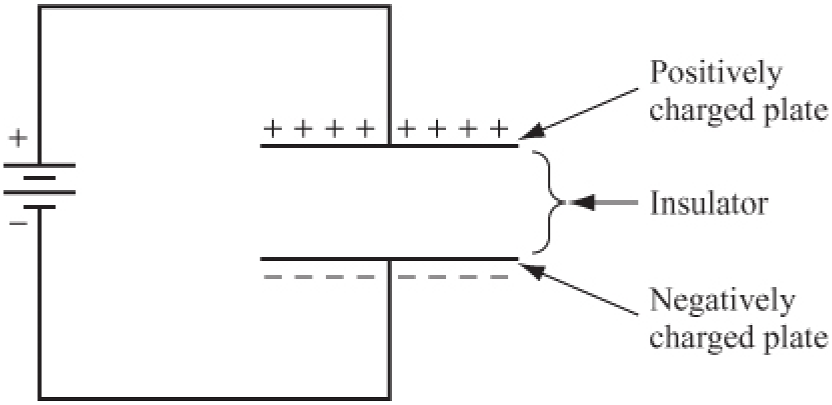

Another common type of capacitor is an electrolytic capacitor, available with capacitance values ranging from 0.1 μF to several thousand μF. The name electrolytic comes from the use of an electrolytic fluid, which, because it is conductive, acts as one of the “plates,” whereas the other plate is made of metal. The insulating material is an oxide on the surface of the metal.

Unlike ceramic capacitors, many electrolytic capacitors are polarized, which means that you must insert the lead marked with a “+” in the circuit closest to the positive voltage source. The symbol for a capacitor indicates the direction in which you insert polarized capacitors in a circuit. The curved side of the capacitor symbol indicates the negative side of the capacitor, whereas the straight side of the symbol indicates the positive side of the capacitor. You can see this orientation later in this chapter in Figure 1.22 .

Units of capacitance are stated in pF (picofarad), μF (microfarad), and F (farad). One μF equals 1,000,000 pF and one F equals 1,000,000 μF**.** Many capacitors are marked with their capacitance value, such as 220 pF. However, you'll often find capacitors that use a different numerical code, such as 224. The first two numbers in this code are the first and second significant digits of the capacitance value. The third number is the multiplier, and the units are pF. Therefore, a capacitor marked with 221 has a value of 220 pF, whereas a capacitor with a marking of 224 has a value of 220,000 pF. (You can simplify this to 0.22 μF**.**)

41 Capacitors can be connected in parallel, as shown in Figure 1.18 .

Figure 1.18

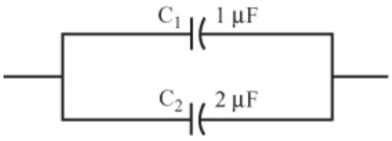

(1)

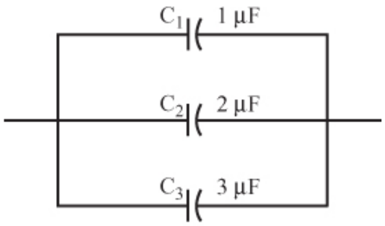

(2)

Questions

A. What is the formula for the total capacitance? __________

B. What is the total capacitance in circuit 1? __________

C. What is the total capacitance in circuit 2? __________

Answers

A. $C_T = C_1 + C_2 + C_3 + \ldots + C_N$

B. $C_T = 1 + 2 = 3$ μF

C. $C_T = 1 + 2 + 3 = 6$ μF

In other words, the total capacitance is found by simple addition of the capacitor values.

42 Capacitors can be placed in series, as shown in Figure 1.19.

Figure 1.19

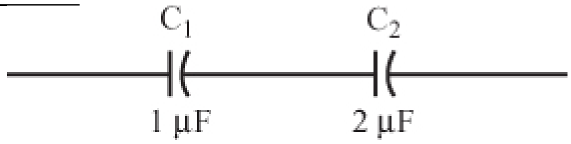

Questions

A. What is the formula for the total capacitance? __________

B. In Figure 1.19, what is the total capacitance? __________

Answers

A.

$$\frac{1}{C_T} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots + \frac{1}{C_N}$$

B.

$$\frac{1}{C_T} = \frac{1}{1} + \frac{1}{2} = 1 + \frac{1}{2} = \frac{3}{2}; \text{thus } C_T = \frac{2}{3}$$

43 In each of these examples, the capacitors are placed in series. Find the total capacitance.

Questions

A. $C_1 = 10$ μF, $C_2 = 5$ μF __________

B. $C_1 = 220$ μF, $C_2 = 330$ μF, $C_3 = 470$ μF __________

C. $C_1 = 0.33$ μF, $C_2 = 0.47$ μF, $C_3 = 0.68$ μF __________

Answers

A. 3.3 μF

B. 103.06 μF

C. 0.15 μF

Summary

The few simple principles reviewed in this chapter are those you need to begin the study of electronics. Following is a summary of these principles:

The basic electrical circuit consists of a source (voltage), a load (resistance), and a path (conductor or wire).

The voltage represents a charge difference.

If the circuit is a complete circuit, then electrons flow, which is called current flow.

The resistance offers opposition to current flow.

The relationship between V, I, and R is given by Ohm's law:

$$V = I \times R$$

Resistance could be a combination of resistors in series, in which case you add the values of the individual resistors together to get the total resistance.

$$R_T = R_1 + R_2 + \cdots + R_N$$

Resistance can be a combination of resistors in parallel, in which case you find the total by using the following formula:

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots + \frac{1}{R_N} \quad \text{or} \quad R_T = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots + \frac{1}{R_N}}$$

You can find the power delivered by a source by using the following formula:

$$P = VI$$

You can find the power dissipated by a resistance by using the following formula:

$$P = I^2 R \quad \text{or} \quad P = \frac{V^2}{R}$$

If you know the total applied voltage, V S , you can find the voltage across one resistor in a series string of resistors by using the following voltage divider formula:

$$V_1 = \frac{V_S R_1}{R_T}$$

You can find the current through one resistor in a two resistor parallel circuit with the total current known by using the current divider formula:

$$I_1 = \frac{I_T R_2}{(R_1 + R_2)}$$

Kirchhoff's Voltage Law (KVL) relates the voltage drops in a series circuit to the total applied voltage.

$$V_S = V_1 + V_2 + \cdots + V_N$$

Kirchhoff's Current Law (KCL) relates the currents at a junction in a circuit by saying that the sum of the input currents equals the sum of the output currents. For a simple parallel circuit, this becomes the following, where I T is the input current:

$$I_T = I_1 + I_2 + \cdots + I_N$$

A switch in a circuit is the control device that directs the flow of current or, in many cases, allows that current to flow.

Capacitors are used to store electric charge in a circuit. They also allow current or voltage to change at a controlled pace. The circuit time constant is found by using the following formula:

$$\tau = R \times C$$

At one time constant in an RC circuit, the values for current and voltage have reached 63 percent of their final values. At five time constants, they have reached their final values.

Capacitors in parallel are added to find the total capacitance.

$$C_T = C_1 + C_2 + \cdots + C_N$$

Capacitors in series are treated the same as resistors in parallel to find a total capacitance.

$$\frac{1}{C_T} = \frac{1}{C_1} + \frac{1}{C_2} + \cdots + \frac{1}{C_N} \quad \text{or} \quad C_T = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots + \frac{1}{C_N}}$$

DC Pre-Test

The following problems and questions test your understanding of the basic principles presented in this chapter. You need a separate sheet of paper for your calculations. Compare your answers with the answers provided following the test. You can work many of the problems in more than one way.

Questions 1–5 use the circuit shown in Figure 1.20. Find the unknown values indicated using the values given.

Figure 1.20

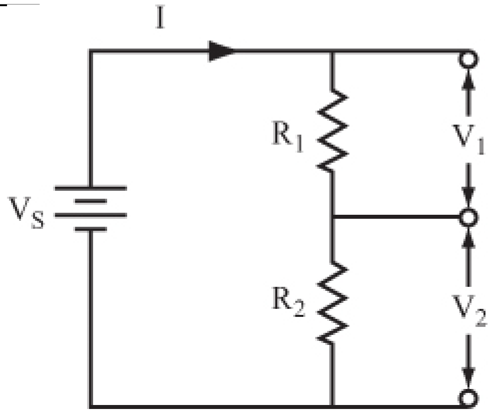

1. R_1 = 12 ohms, R_2 = 36 ohms, V_S = 24 volts

RT = _____ , I = _____

2. R1 = 1 kΩ, R2 = 3 kΩ, I = 5 mA

V1 = _____ , V2 = _____ , VS = _____

3. R1 = 12 kΩ, R2 = 8 kΩ, VS = 24 volts

V1 = _____ , V2 = _____

4. VS = 36 V, I = 250 mA, V1 = 6 volts

R2 = _____

5. Now, go back to problem 1. Find the power dissipated by each resistor and the total power delivered by the source.

P1 = _____ , P2 = _____ , PT = _____

Questions 6–8 use the circuit shown in Figure 1.21. Again, find the unknowns using the given values.

Figure 1.21

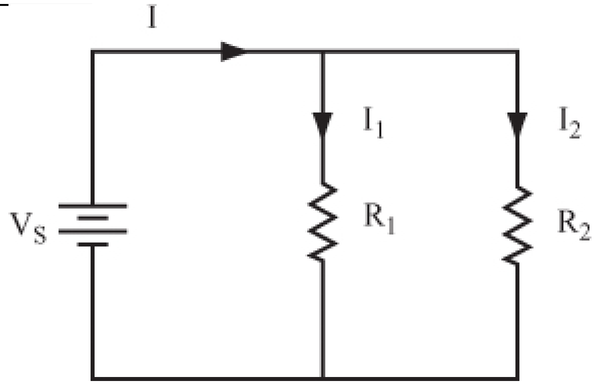

6. R1 = 6 kΩ, R2 = 12 kΩ, VS = 20 volts

RT = _____ , I = _____

7. I = 2 A, R1 = 10 ohms, R2 = 30 ohms

I1 = _____ , I2 = _____

8. VS = 12 volts, I = 300 mA, R1 = 50 ohms

R2 = _____ , P1 = _____

9. What is the maximum current that a 220-ohm resistor can safely have if its power rating is 1/4 watt?

I MAX = _____

10. In a series RC circuit the resistance is 1 kΩ, the applied voltage is 3 volts, and the time constant should be 60 μsec.

A. What is the required value of C?

C = _____

B. What is the voltage across the capacitor 60 μsec after the switch is closed?

V C = _____

C. At what time will the capacitor be fully charged?

T = _____

11. In the circuit shown in Figure 1.22 , when the switch is at position 1, the time constant should be 4.8 ms.

Figure 1.22

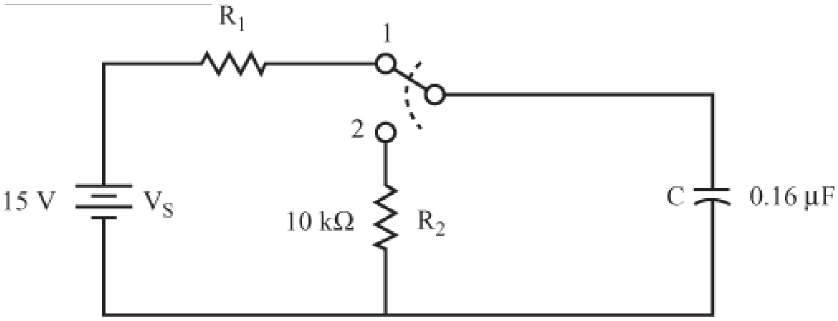

A. What should be the value of resistor R 1 ?

R 1 = _____

B. What will be the voltage on the capacitor when it is fully charged, and how long will it take to reach this voltage?

V_C = _____, T = _____

C. After the capacitor is fully charged, the switch is thrown to position 2. What is the discharge time constant, and how long will it take to completely discharge the capacitor?

τ = _____ , T = _____

12. Three capacitors are available with the following values:

C_1 = 8 μF; C_2 = 4 μF; C_3 = 12 μF.

A. What is C_T if all three are connected in parallel?

C_T = _____

B. What is C_T if they are connected in series?

C_T = _____

C. What is C_T if C_1 is in series with the parallel combination of C_2 and C_3?

C_T = _____

Answers to DC Pre-Test

If your answers do not agree with those provided here, review the problems indicated in parentheses before you go to Chapter 2, “The Diode.” If you still feel uncertain about these concepts, go to a website such as www.BuildingGadgets.com and work through DC tutorials listed there.

It is assumed that Ohm's law is well known, so problem 4 will not be referenced.

1. R_T = 48 ohms, I = 0.5 ampere (problem 9)

2. V_1 = 5 volts, V_2 = 15 volts, V_S = 20 volts (problems 23 and 26)
3. V_1 = 14.4 volts, V_2 = 9.6 volts (problems 23 and 26)
4. R_2 = 120 ohms (problems 9 and 23)
5. P_1 = 3 watts, P_2 = 9 watts, P_T = 12 watts (problems 9 and 13)
6. R_T = 4 kΩ, I = 5 mA (problem 10)
7. I_1 = 1.5 amperes, I_2 = 0.5 ampere (problems 28 and 29)
8. R_2 = 200 ohms, P_1 = 2.88 watts (problems 10 and 13)
9. I_{MAX} = 33.7 mA (problems 13, 15, and 16)
10A. C = 0.06 μF (problems 34 and 35)
10B. V_C = 1.9 volts (problem 35)
10C. T = 300 μsec (problems 34–38)
11A. R_1 = 30 kΩ (problems 33, 39, and 40)
11B. V_C = 15 V, T = 24 ms (problem 35)
11C. τ = 1.6 ms, T = 8.0 ms (problems 39–40)
12A. C_T = 24 μF (problems 41 and 42)
12B. C_T = 2.18 μF (problem 42)
12C. C_T = 5.33 μF (problems 42–43)

Chapter 2

The Diode

The main characteristic of a diode is that it conducts electricity in one direction only. Historically, the first vacuum tube was a diode; it was also known as a *rectifier* . The modern *diode* is a semiconductor device. It is used in all applications where the older vacuum tube diode was used, but it has the advantages of being much smaller, easier to use, and less expensive.

A *semiconductor* is a crystalline material that, depending on the conditions, can act as a conductor (allowing the flow of electric current) or an insulator (preventing the flow of electric current). Techniques have been developed to customize the electrical properties of adjacent regions of semiconductor crystals, which allow the manufacture of small diodes, as well as transistors and integrated circuits.

When you complete this chapter, you can do the following:

Specify the uses of diodes in DC circuits.

Determine from a circuit diagram whether a diode is forward- or reverse-biased.

Recognize the characteristic V-I curve for a diode.

Specify the knee voltage for a silicon or a germanium diode.

Calculate current and power dissipation in a diode.

Define diode breakdown.

Differentiate between zeners and other diodes.

Determine when a diode can be considered “perfect.”

Understanding Diodes

1 Silicon and germanium are semiconductor materials used in the manufacture of diodes, transistors, and integrated circuits. Semiconductor material is refined to an extreme level of purity, and then minute, controlled amounts of a specific impurity are added (a process called *doping*). Based on which impurity is added to a region of a semiconductor crystal, that region is said to be *N type* or *P type.* In addition to electrons (which are negative charge carriers used to conduct charge in a conventional conductor), semiconductors contain positive charge carriers called *holes.* The impurities added to an N type region increases the number of electrons capable of conducting charge, whereas the impurities added to a P type region increase the number of holes capable of conducting charge.

When a semiconductor chip contains an N doped region adjacent to a P doped region, a *diode junction* (often called a *PN junction*) is formed. Diode junctions can also be made with either silicon or germanium. However, silicon and germanium are never mixed when making PN junctions.

Question

Which diagrams in Figure 2.1 show diode junctions?_____

Figure 2.1

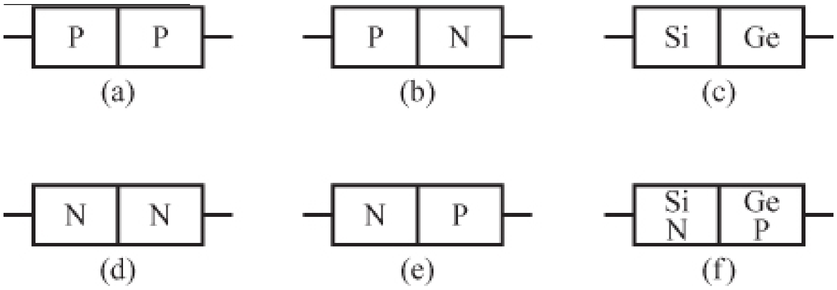

Answer

Diagrams (b) and (e) only

2 In a diode, the P material is called the *anode.* The N material is called the *cathode.*

Question

Identify which part of the diode shown in Figure 2.2 is P material and which part is N material._____

Figure 2.2

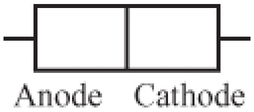

Answer

The anode is P material; the cathode is N material.

3 Diodes are useful because electric current can flow through a PN junction *in one direction only* . Figure 2.3 shows the direction

in which the current flows.

Figure 2.3

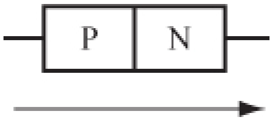

Figure 2.4 shows the circuit symbol for a diode. The arrowhead points in the direction of current flow. Although the anode and cathode are indicated here, they are not usually indicated in circuit diagrams.

Figure 2.4

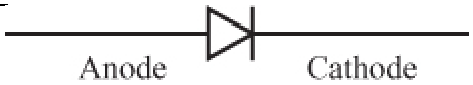

Question

In a diode, does current flow from anode to cathode, or cathode to anode?_____

Answer

Current flows from anode to cathode.

4 In the circuit shown in Figure 2.5 , an arrow shows the direction of current flow.

Figure 2.5

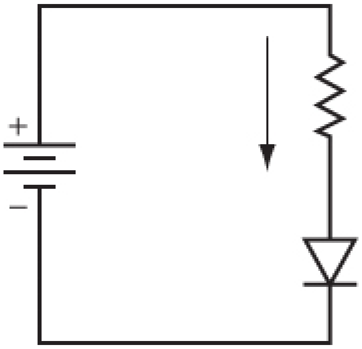

Questions

A. Is the diode connected correctly to permit current to flow?_____

B. Notice the way the battery and the diode connect. Is the anode at a higher or lower voltage than the cathode?_____

Answers

A. Yes.

B. The anode connects to the positive battery terminal, and the cathode connects to the negative battery terminal. Therefore, the anode is at a higher voltage than the cathode.

5 When the diode is connected so that the current flows, it is *forward-biased.* In a forward-biased diode, the anode connects to a higher voltage than the cathode, and current flows. Examine the way the diode is connected to the battery shown in Figure 2.6 .

Figure 2.6

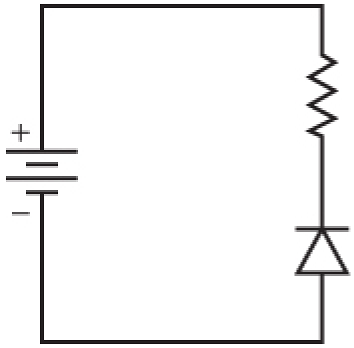

Question

Is the diode forward-biased? Give the reasons for your answer._____

Answer

No, it is not forward-biased. The cathode is connected to the positive battery terminal, and the anode is connected to the negative battery terminal. Therefore, the cathode is at a higher voltage than the anode.

6 When the cathode is connected to a higher voltage level than the anode, the diode cannot conduct. In this case, the diode is *reverse-biased* .

Question

Draw a reverse-biased diode in the circuit shown in Figure 2.7 ._____

Figure 2.7

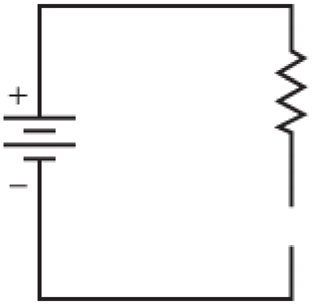

Answer

Your drawing should look something like Figure 2.8 .

Figure 2.8

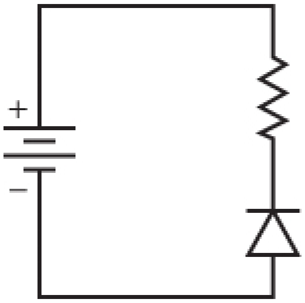

7 In many circuits, the diode is often

considered to be a *perfect* diode to simplify calculations. A perfect diode has zero voltage drop in the forward direction and conducts no current in the reverse direction.

Question

From your knowledge of basic electricity, what other component has zero voltage drop across its terminals in one condition and conducts no current in an alternative condition?_____

Answer

The mechanical switch. When closed, it has no voltage drop across its terminals, and when open, it conducts no current.

8 A forward-biased perfect diode can thus be compared to a closed switch. It has no voltage drop across its terminals, and current flows through it.

A reverse-biased perfect diode can be compared to an open switch. No current flows through it, and the voltage difference between its terminals equals the supply voltage.

Question

Which of the switches shown in Figure 2.9 performs like a forward-biased perfect diode?_____

Figure 2.9

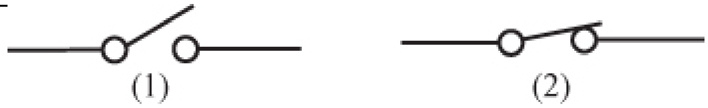

Answers

Switch (2) represents a closed switch and, like a forward-biased perfect diode, allows current to flow through it. There is no voltage drop across its terminals.

Project 2.1: The Diode

If you have access to electronic equipment, you may want to perform the simple project described in the next few problems. If this is the first time you have tried such a project, get help from an instructor or someone who is familiar with electronic projects.

When building electronic circuits, eventually you'll make a mistake (as all of us do), and sometimes those mistakes cause circuits to fry. If you smell hot electronic components, disconnect the battery from the circuit, and then check the circuit to determine what connections you should change.

When fixing a circuit, follow some simple safety rules. Do not try to rearrange connections with the battery connected because you may short leads together.

Also, don't touch bare wires with live electricity. Even with batteries, you have a chance of being burned or seriously injured. If your skin is wet, it forms an electrical connection with lower resistance, allowing more current to flow, potentially injuring you.

If you do not have access to equipment, do not skip this project. Read through the project, and try to picture or imagine the results. This is sometimes called “dry-labbing” the experiment. You can learn a lot from reading about this project, even though it is always better to actually perform the project. This advice also applies to the other projects in many of the following chapters.

Objective

The objective of this project is to plot the V-I curve (also called a *characteristic curve*) of a diode, which shows how current flow through the diode varies with the applied voltage. As shown in Figure 2.10 , the I-V curve for a diode demonstrates that if low voltage is applied to a diode, current does not flow. However, when the applied voltage exceeds a certain value, the current flow increases quickly.

Figure 2.10

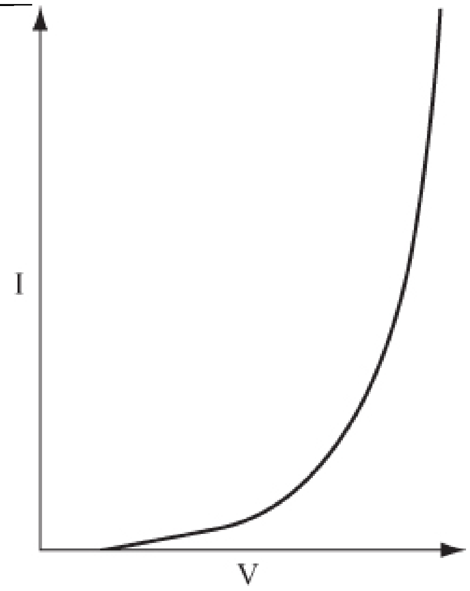

General Instructions

While the circuit is set up, measure the current for each voltage value. As you perform the project, observe how much more rapidly the current increases for higher applied voltages.

Parts List

- One 9 V battery
- One snap battery connector
- One multimeter set to measure current (mA)
- One multimeter set to measure DC voltage
- One 330-ohm, 0.5-watt resistor
- One 1N4001 diode
- One breadboard
- One 1 MΩ potentiometer
- One terminal block

Step-by-Step Instructions

Set up the circuit as shown in Figure 2.11 . The circled “A” designates a multimeter set to measure current, and the circled “V” designates a multimeter set to measure DC voltage. If you have some experience in building circuits, this schematic (along with the previous parts list) should provide all the information you need to build the circuit. If you need a bit more help in building the circuit, look at the photos of the completed circuit in the “Expected Results” section.

Figure 2.11

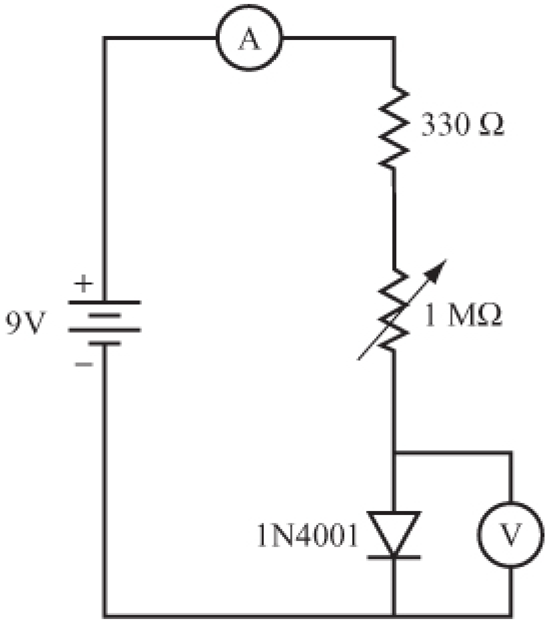

Carefully check your circuit against Figure 2.11 , especially the direction of the battery and the diode. The diode has a band at one end. Connect the lead at the end of the diode with the band to the ground bus on the breadboard.

After you check your circuit, follow these steps, and record your measurements in the blank table following the steps:

1. Set the potentiometer to its highest value. This sets the voltage applied to the diode to its lowest possible value.
2. Measure and record the voltage applied to the diode.
3. Measure and record the current.
4. Adjust the potentiometer slightly to give a higher voltage.

5. Measure and record the new values of voltage and current.

6. Repeat steps 4 and 5 until the lowest resistance of the potentiometer is reached, taking as many readings as possible. This results in the highest voltage and current readings for this circuit. At this point, the potentiometer resistance is zero ohms, and the current is limited to approximately 27 mA by the 330-ohm resistor. This resistor is included in the circuit to avoid overheating the components when the potentiometer is set to zero ohms. If your circuit allows currents significantly above this level as you adjust the potentiometer, something is wrong. You should disconnect the battery and examine the circuit to see if it were connected incorrectly. If V gets large—above 3 or 4 volts—and I remains small, then the diode is backward. Reverse it and start again.

V (volts)	I (mA)

7. Graph the points recorded in the table using the blank graph shown in Figure 2.12. Your curve should look like the one shown in Figure 2.10.

Figure 2.12

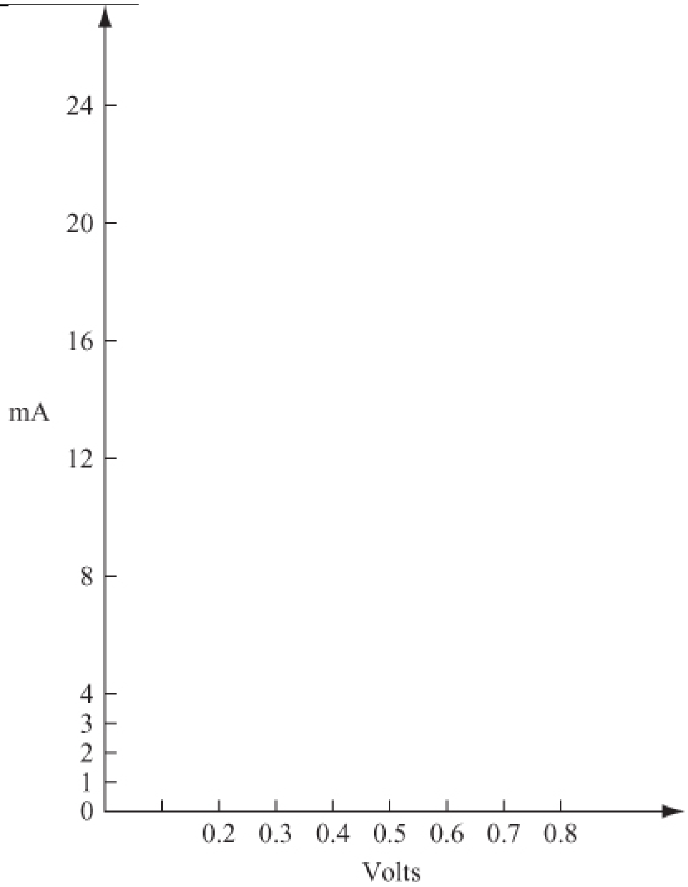

Expected Results

Figure 2.13 shows the breadboarded circuit for this project.

Figure 2.13

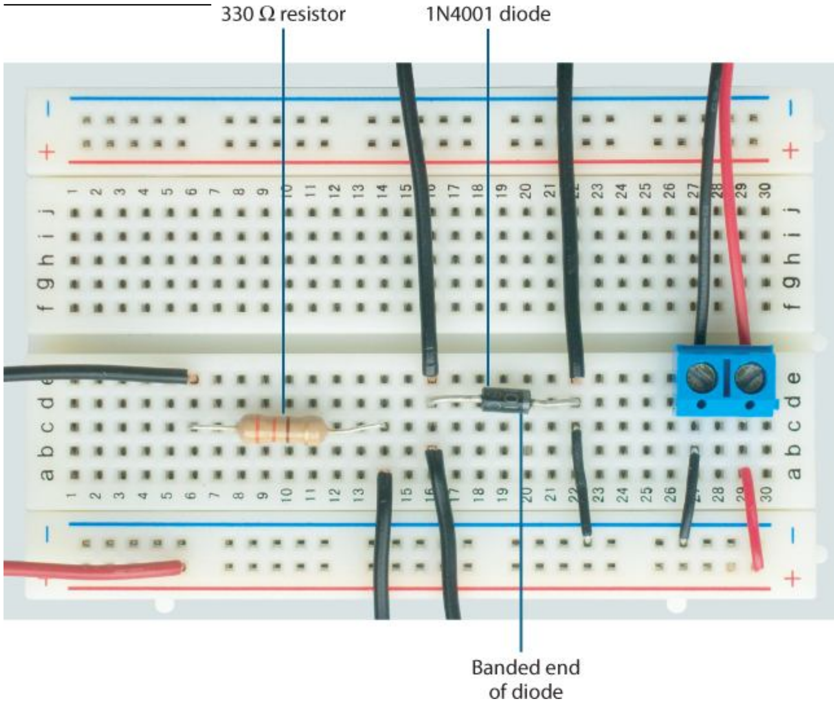
330 Ω resistor
1N4001 diode
Banded end of diode

Figure 2.14 shows the test setup for this project.

Figure 2.14

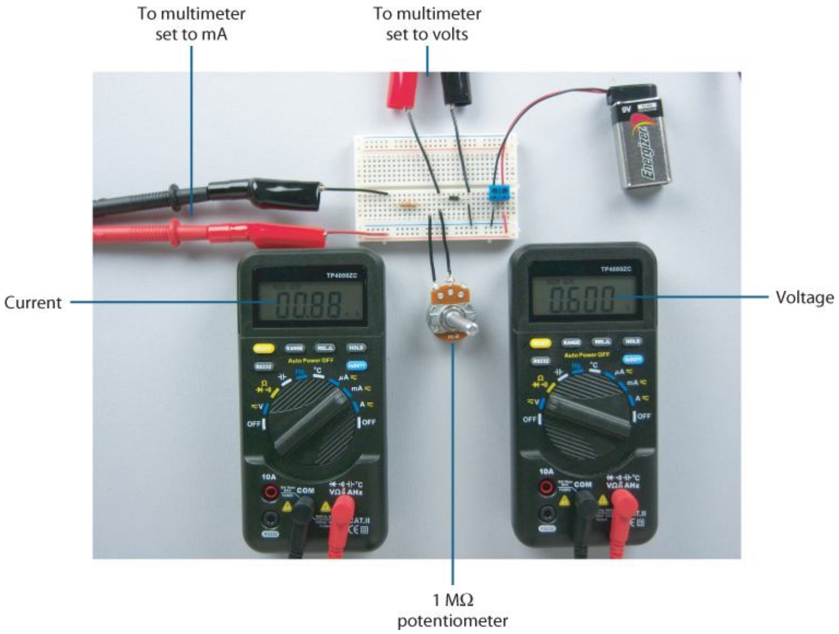
To multimeter set to mA
To multimeter set to volts
Current
Voltage
1 MΩ potentiometer

Compare your measurements with the ones shown in the following table:

V (volts)	I (mA)
0.44	0.00
0.46	0.01
0.50	0.06
0.52	0.11
0.55	0.23
0.58	0.49
0.60	0.92
0.63	1.74
0.68	4.86
0.72	15.1
0.73	20.9
0.74	25.2

Further reductions in the resistance below the 330 Ω included in the circuit causes little increase in the voltage but produces large increases in the current.

Figure 2.15 shows the V-I curve generated using the measurements shown in the preceding table.

Figure 2.15

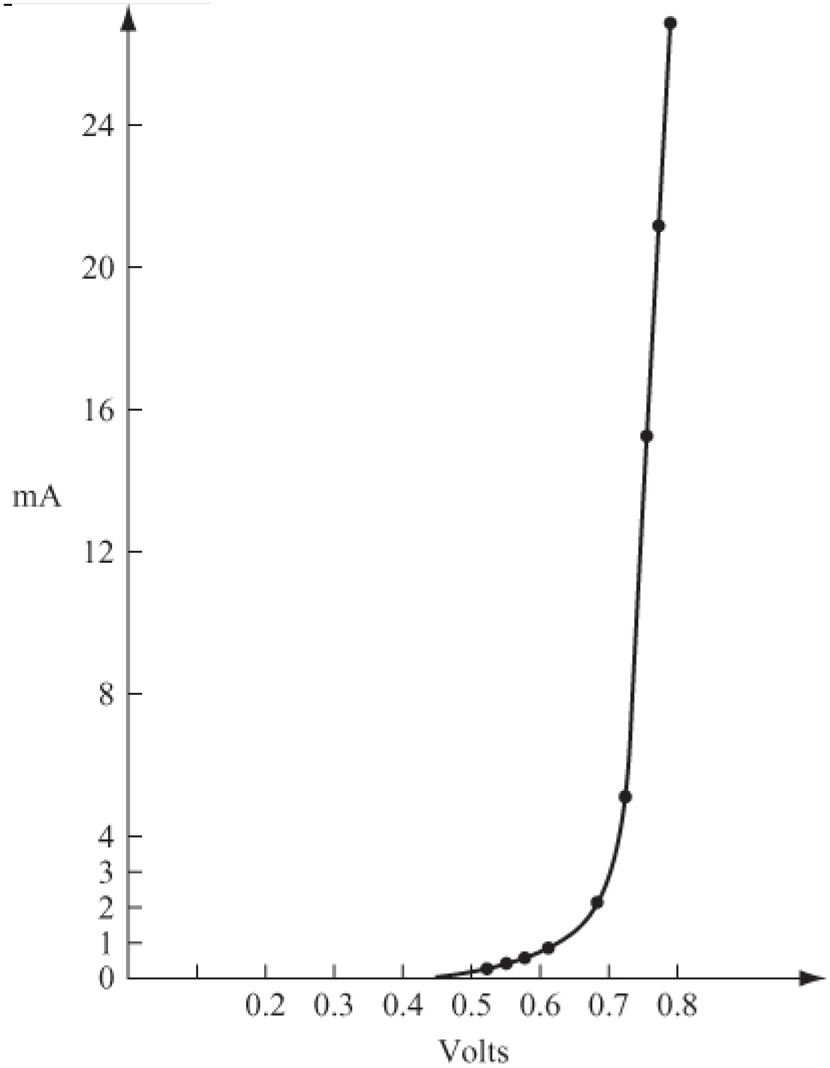

The V-I curve (or diode characteristic curve) is repeated in Figure 2.16 with three important regions marked on it.

Figure 2.16

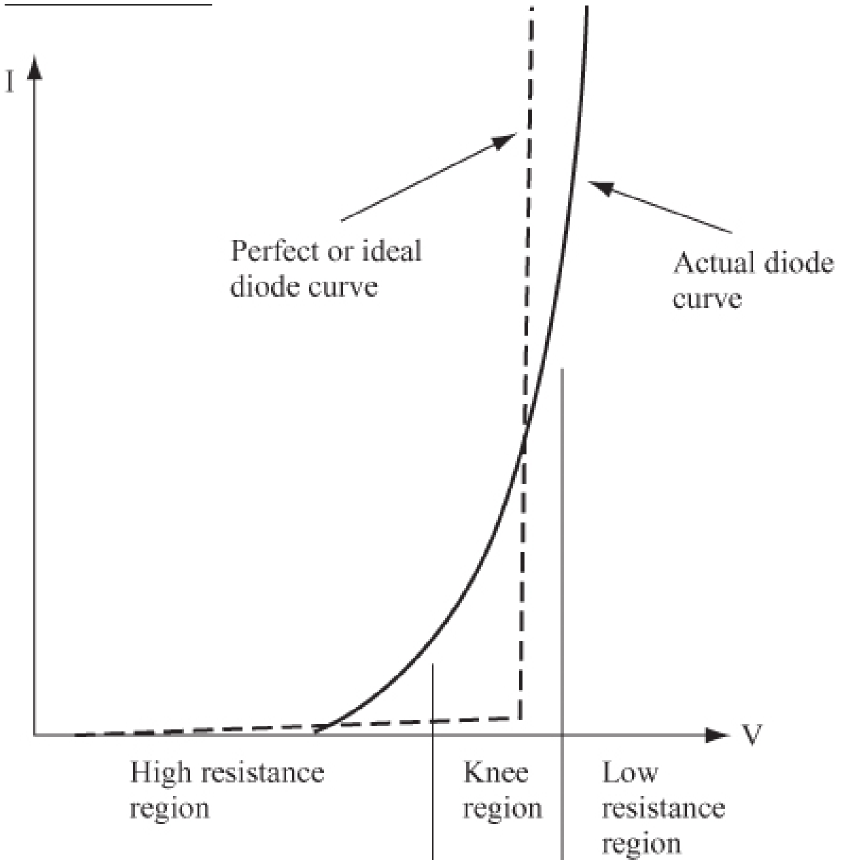
I
V
Perfect or ideal diode curve
Actual diode curve
High resistance region
Knee region
Low resistance region

The most important region is the *knee region*. This is not a sharply defined changeover point, but it occupies a narrow range of the curve where the diode resistance changes from high to low.

The ideal curve is shown for comparison.

For the diode used in this project, the knee voltage is about 0.7 volt, which is typical for a silicon diode. This means (and your data should verify this) that at voltage levels below 0.7 volt, the diode has such a high resistance that it limits the current flow to a low value. This characteristic knee voltage is sometimes referred to as a *threshold voltage*. If you use a germanium diode, the knee voltage is approximately 0.3 volt.

9 The knee voltage is also a *limiting voltage*. That is, it is the highest voltage that can be obtained across the diode in the forward direction.

Questions

A. Which has the higher limiting voltage, germanium or silicon?_____

B. What happens to the diode resistance at the limiting or knee voltage?_____

Answers

A. Silicon, with a limiting voltage of 0.7 volt, is higher than germanium, which has a limiting voltage of only 0.3 volt.

B. It changes from high to low.

Note You use these knee voltages in many later chapters as the voltage drop across the PN junction when it is forward-biased.

10 Refer back to the diagram of resistance regions shown in Figure 2.16.

Question

What happens to the current when the voltage becomes limited at the knee?_____

Answer

It increases rapidly.

11 For any given diode, the knee voltage is not exactly 0.7 volt or 0.3 volt. Rather, it varies slightly. But when using diodes in practice (that is, imperfect diodes), you can make two assumptions:

The voltage drop across the diode is either 0.7 volt or 0.3 volt.

You can prevent excessive current from flowing through the diode by using the appropriate resistor in series with the diode.

Questions

A. Why are imperfect diodes specified here?_____

B. Would you use a high or low resistance to prevent excessive current?_____

Answers

A. All diodes are imperfect, and the 0.3 or 0.7 voltage values are only approximate. In fact, in some later problems, it is assumed that the voltage drop across the diode, when it is conducting, is 0 volts. This assumes, then, that as soon as you apply any voltage above 0, current flows in an ideal resistor. (That is, the knee voltage on the V-I curve for an ideal diode is 0 volts.)

B. Generally, use a high resistance. However, you can calculate the actual resistance value given the applied voltage and the maximum current the diode can withstand.

12 Calculate the current through the diode in the circuit shown in Figure 2.17 using the steps in the following questions.

Figure 2.17

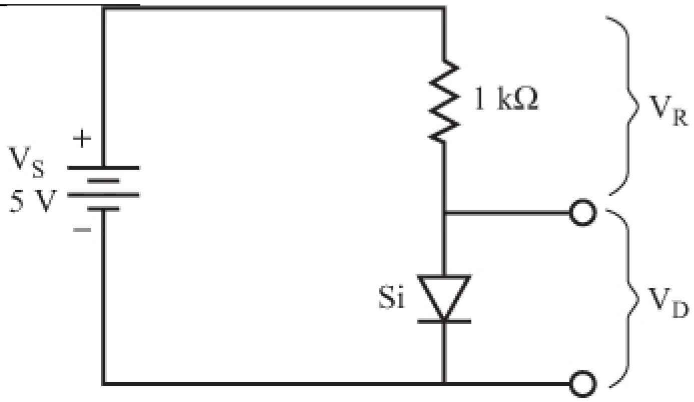

Questions

A. The voltage drop across the diode is known. It is 0.7 volt for silicon and 0.3 volt for germanium. (“Si” near the diode means it is silicon.) Write down the diode voltage drop. V_D = _____

B. Find the voltage drop across the resistor. This can be calculated using $V_R = V_S - V_D$. This is taken from KVL, which was discussed in Chapter 1, “DC Review and Pre-Test.” V_R = _____

C. Calculate the current through the resistor. Use $I = V_R/R$.

I = _____

D. Finally, determine the current through the diode. I =

Answers

You should have written these values:

A. 0.7 volt

B. $V_R = V_S - V_D$ = 5 volts – 0.7 volt = 4.3 volts

C.

$$I = \frac{V_R}{R} = \frac{4.3\,\text{volts}}{1\,\text{k}\Omega} = 4.3\,\text{mA}$$

D. 4.3 mA (In a series circuit, the same current runs through each component.)

13 In practice, when the battery voltage is 10 volt or above, the voltage drop across the diode is often considered to be 0 volt, instead of 0.7 volt.

The assumption here is that the diode is a perfect diode, and the knee voltage is at 0 volts, rather than at a threshold value that must be exceeded. As discussed later, this assumption is often used in many electronic designs.

Questions

A. Calculate the current through the silicon diode, as shown in Figure 2.18 .

Figure 2.18

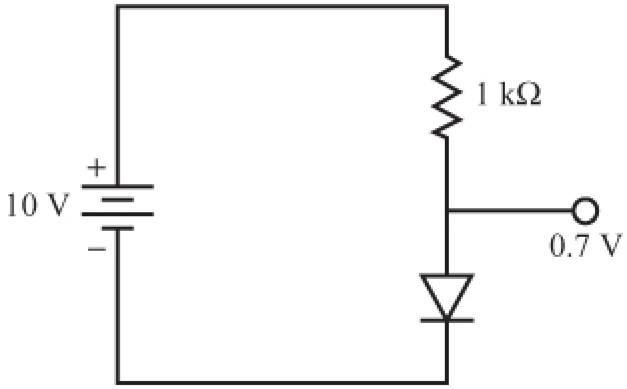

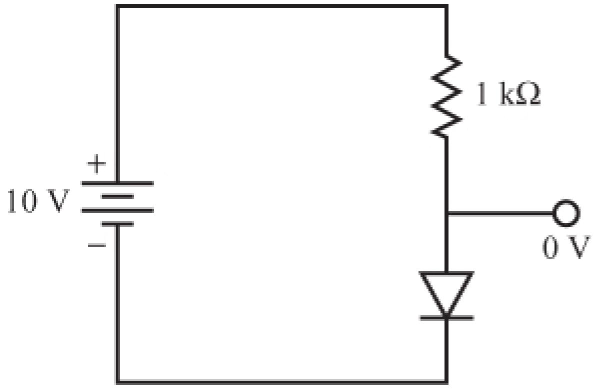

VD = _____

VR = VS – VD = _____

$$I = \frac{V_R}{R} = \_\_\_\_\_$$

ID = _____

B. Calculate the current through the perfect diode, as shown in Figure 2.18 .

VD = _____

VR = VS – VD = _____

$$I = \frac{V_R}{R} = \_\_\_\_\_$$

I_D = _____

Answers

A. 0.7 volt; 9.3 volt; 9.3 mA; 9.3 mA

B. 0 volt; 10 volt; 10 mA; 10 mA

14 The difference in the values of the two currents found in problem 13 is less than 10 percent of the total current. That is, 0.7 mA is less than 10 percent of 10 mA. Many electronic components have a plus or minus 5 percent tolerance in their nominal values. This means that a 1 kΩ resistor can be anywhere from 950 ohms to 1,050 ohms, meaning that the value of current through a resistor can vary plus or minus 5 percent.

Because of a slight variance in component values, calculations are often simplified if the simplification does not change values by more than 10 percent. Therefore, a diode is often assumed to be perfect when the supply voltage is 10 volts or more.

Questions

A. Examine the circuit in Figure 2.19. Is it safe to assume that the diode is perfect?

_____

Figure 2.19

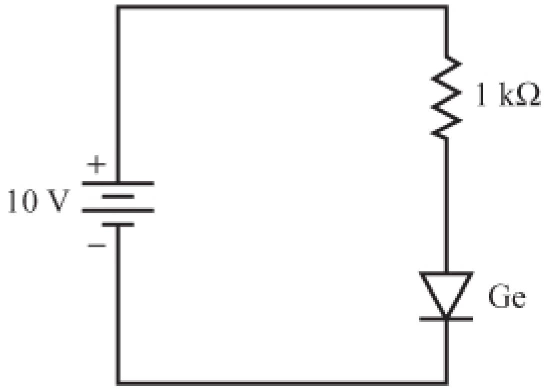

B. Calculate the current through the diode.

_____

Answers

A. Yes, it can be considered a perfect diode.

B. I = 10 mA

15 When a current flows through a diode, it causes heating and power dissipation, just as with a resistor. The power formula for resistors is P = V × I. This same formula can be applied to diodes to find the power dissipation.

To calculate the power dissipation in a diode, you must first calculate the current as shown previously. The voltage drop in this formula is assumed to be 0.7 volt for a silicon diode, even if you considered it to be 0 volts when calculating the current.

For example, a silicon diode has 100 mA flowing through it. Determine how much power the diode dissipates.

$$P = (0.7\,\text{volt})(100\,\text{mA}) = 70\,\text{mW}$$

Question

Assume a current of 2 amperes is flowing through a silicon diode. How much power is being dissipated? _____

Answer

$$P = (0.7\,\text{volts})(2\,\text{amperes}) = 1.4\,\text{watts}$$

16 Diodes are made to dissipate a certain amount of power, and this is quoted as a maximum power rating in the manufacturer's specifications of the diode.

Assume a silicon diode has a maximum power rating of 2 watts. How much current can it safely pass?

$$P = V \times I$$

$$I = \frac{P}{V}$$

$$= \frac{2\,\text{watts}}{0.7\,\text{volt}}$$

$$2.86\,\text{amperes (rounded off to two decimal places)}$$

Provided the current in the circuit does not exceed this, the diode cannot overheat and burn out.

Question

Suppose the maximum power rating of a germanium diode is 3 watts. What is its highest safe current? _____

Answer

$$I = \frac{3\,\text{watts}}{0.3\,\text{volt}} = 10\,\text{amperes}$$

17 Answer the following questions for another example.

Questions

A. Could a 3-watt silicon diode carry the current calculated for the germanium diode for problem 16? _____

B. What would be its safe current? _____

Answers

A. No, 10 amperes would cause a power dissipation of 7 watts, which would burn up the diode.

B.

$$I = \frac{3}{0.7} = 4.3\,\text{amperes}$$

Any current less than this would be safe.

18 The next several examples concentrate on finding the current through the diode. Look at the circuit shown in Figure 2.20 .

Figure 2.20

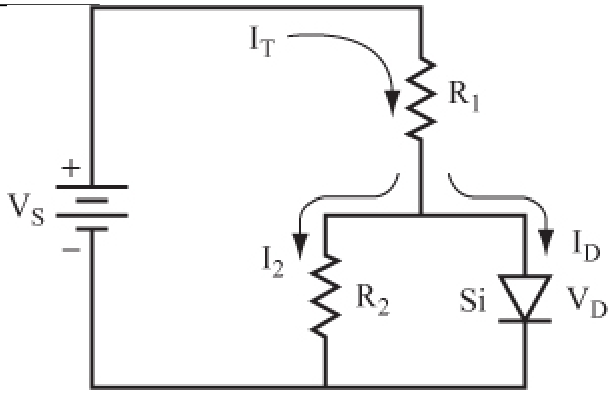

The total current from the battery flows through R_1, and then splits into I_2 and I_D. I_2 flows through R_2, and I_D flows through the diode.

Questions

A. What is the relationship between I_T, I_2, and I_D? _____

B. What is the value of V_D? _____

Answers

A. Remember KCL, $I_T = I_2 + I_D$ _____

B. $V_D = 0.7$ volt _____

19 To find I_D, you need to go through the following steps because there is no way to find I_D directly:

1. Find I_2. This is done using $V_D = R_2 \times I_2$.
2. Find V_R. For this, use $V_R = V_S - V_D$ (KVL again).
3. Find I_T (the current through R_1). Use $V_R = I_T \times R_1$.
4. Find I_D. This is found by using $I_T = I_2 + I_D$ (KCL again).

To find I_D in the circuit shown in Figure 2.21, go through these steps, and then check your answers.

Figure 2.21

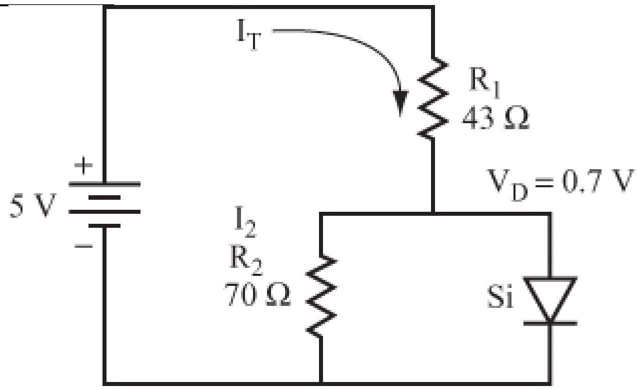

Questions

A. I_2 = _____

B. V_R = _____

C. I_T = _____

D. I_D = _____

Answers

A.

$$I_2 = \frac{V_D}{R_2} = \frac{0.7\,\text{volt}}{R_2} = \frac{0.7\,\text{volt}}{70\,\text{ohms}} = 0.01\,\text{ampere} = 10\,\text{mA}$$

B. $V_R = V_S - V_D$ = 5 volts – 0.7 volt = 4.3 volts

C.

$$I_T = \frac{V_R}{R_1} = \frac{4.3\ \text{volts}}{43\,\text{ohms}} = 0.1\,\text{ampere} = 100\,\text{mA}$$

D. $I_D = I_T - I_2$ = 100 mA – 10 mA = 90 mA

20 For this problem, refer to your answers in problem 19.

Question

What is the power dissipation of the diode in problem 19? _____

Answer

$P = V_D \times I_D$ = (0.7 volt)(90 mA) = 63 milliwatts

21 To find the current in the diode for the circuit shown in Figure 2.22, answer the following questions in order.

Figure 2.22

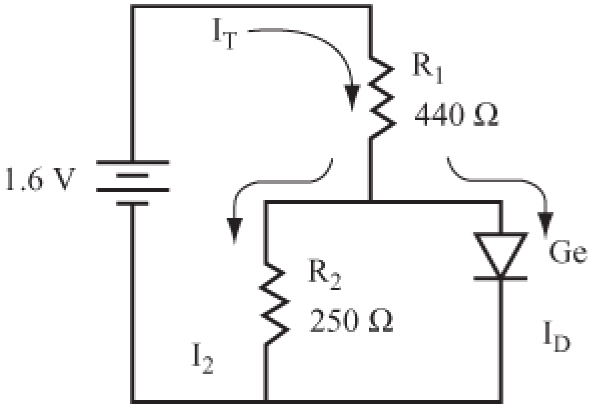

Questions

A. I_2 = _____

B. V_R = _____

C. I_T = _____

D. I_D = _____

Answers

A. $$I_2 = \frac{0.3}{250} = 1.2\,\text{mA}$$

B. $V_R = V_S - V_D = 1.6 - 0.3 = 1.3$ volts

C. $$I_T = \frac{V_R}{R_1} = \frac{1.3}{440} = 3\,\text{mA}$$

D. $I_D = I_T - I_2 = 1.8$ mA

If you want to take a break soon, this is a good stopping point.

Diode Breakdown

22 Earlier, you read that if the circuit in Project 2.1 was not working correctly, then the diode may be in backward. If you place the diode in the circuit backward—as shown on the right in Figure 2.23 —then almost no

current flows. In fact, the current flow is so small, it can be said that no current flows. The V-I curve for a reversed diode looks like the one shown on the left in Figure 2.23.

Figure 2.23

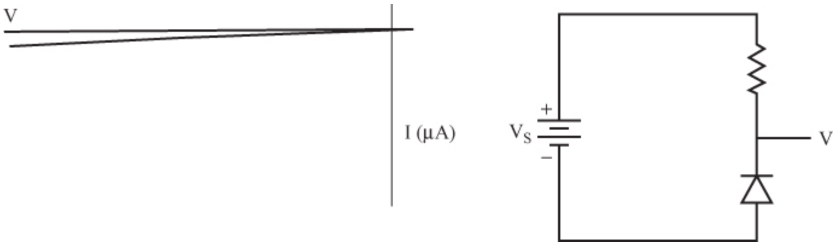

The V-I curve for a perfect diode would show zero current for all voltage values. But for a real diode, a voltage is reached where the diode “breaks down” and the diode allows a large current to flow. The V-I curve for the diode breakdown would then look like the one in Figure 2.24.

Figure 2.24

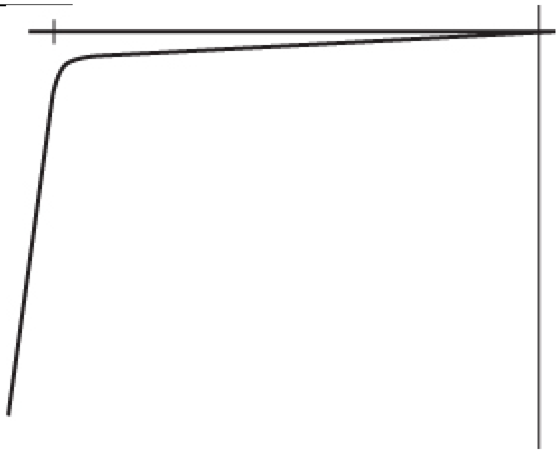

If this condition continues, the diode will burn out. You can avoid burning out the diode, even though it is at the breakdown voltage, by limiting the current with a resistor.

Question

The diode in the circuit shown in Figure 2.25 is known to break down at 100 volts, and it can safely pass 1 ampere without overheating. Find the resistance in this circuit that would limit the current to 1 ampere.

Figure 2.25

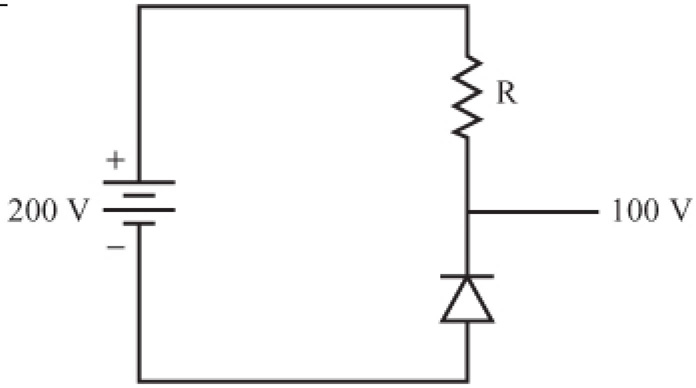

Answer

$$V_R = V_S - V_D = 200\text{ volts} - 100\text{ volts} = 100\text{ volts}$$

Because 1 ampere of current is flowing, then

$$R = \frac{V_R}{I} = \frac{100\text{ volts}}{1\text{ ampere}} = 100\text{ ohms}$$

23 All diodes break down when connected in the reverse direction if excess voltage is applied to them. The breakdown voltage (which is a function of how the diode is made) varies from one type of diode to another. This voltage is quoted in the manufacturer's data sheet.

Breakdown is not a catastrophic process and does not destroy the diode. If the excessive supply voltage is removed, the diode can recover and operate normally. You can use it safely many more times, provided the current is limited to prevent the diode from burning out.

A diode always breaks down at the same voltage, no matter how many times it is used.

The breakdown voltage is often called the *peak inverse voltage* (*PIV*) or the *peak reverse voltage* (*PRV*). Following are the PIVs of some common diodes:

Diode	PIV
1N4001	50 volts
1N4002	100 volts
1N4003	200 volts
1N4004	400 volts
1N4005	600 volts
1N4006	800 volts

Questions

A. Which can permanently destroy a diode, excessive current or excessive voltage? _____

B. Which is more harmful to a diode, breakdown or burnout? _____

Answers

A. Excessive current. Excessive voltage cannot harm the diode if the current is limited.

B. Burnout. Breakdown is not necessarily harmful, especially if the current is limited.

Inside the Diode

At the junction of the N and P type regions of a diode, electrons from the N region are trapped by holes in the P region, forming a depletion region as illustrated in the following figure.

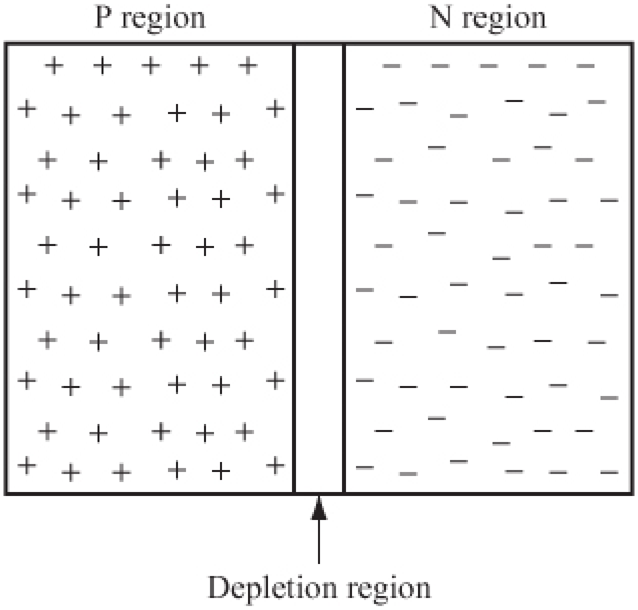

When the electrons are trapped by holes, they can no longer move, which is what produces this depletion region that contains no mobile electrons. The static electrons give this region the properties of an electrical insulator.

You apply a forward-biased voltage to a diode by applying negative voltage to the N region and positive voltage to the P region. Electrons in the N region are repelled by the negative voltage, pushing more electrons into the depletion region. However, the electrons in the N region are also repelled by the electrons already in the depletion region. (Remember that like charges repel each other.) When the forward-biased voltage is sufficiently high (0.7 volts for silicon diodes, and 0.3 volts for germanium diodes), the depletion region is eliminated. As the voltage is raised further, the negative voltage, in repelling electrons in the N region, pushes them into the P region, where they are attracted by the positive voltage.

This combination of repulsive and attractive forces allows current to flow, as illustrated in the following figure.

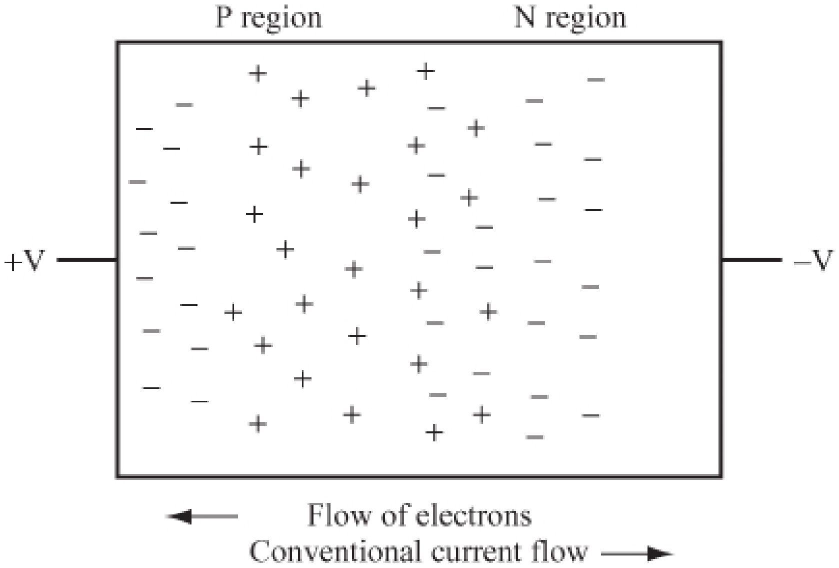

Note that an interesting aspect of diodes is that while the N and P regions of the diode have mobile charges, they do not have a net charge. The mobile charges (electrons in N regions and holes in P regions) are simply donated by the impurity atoms used to dope the semiconductor material. The impurity atoms have more electrons in the N region and less in the P region than are needed to bond to the adjacent silicon or germanium atoms in the crystalline structure. However, because the atoms have the same number of positive and negative charges, the net charge is neutral.

In a similar way, metal is a conductor because it has more electrons than are needed to bond together the atoms in its crystalline structure. These excess electrons are free to move when a voltage is applied, either in a metal or an N type semiconductor, which makes the material electrically conductive. Electrons moving through the P region of the diode jump between the holes, which physicists model as the holes moving. In a depletion region, where electrons have been trapped by holes, there is a net negative charge.

You apply a reverse-biased voltage to a diode by applying positive voltage to the N type region and negative voltage to the P type region. In this scenario, electrons are attracted to the positive voltage, which pulls them away from the depletion region. No current can flow unless the voltage exceeds the reverse breakdown voltage.

Diodes are marked with one band indicating the cathode (the N region) end of the diode. The band on the diode corresponds to the bar on the diode symbol, as shown in the following figure. Use this marking to orient diodes in a circuit.

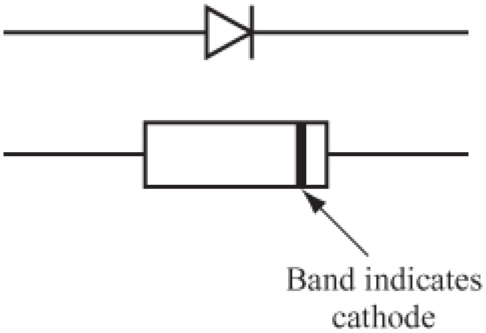

The part number is marked on diodes. However, the part numbers don't tell you much about the diode. For example, the 1N4001 is a silicon diode that can handle a peak reverse voltage of 50 volts, whereas the 1N270 is a germanium diode that can handle a peak reverse voltage of 80 volts. Your best bet is to refer to the manufacturer's data sheet for the diode peak reverse voltage and other characteristics. You can easily look up data sheets on the Internet. Also, you can find links to the data sheets for the components used in this book on the website at www.buildinggadgets.com/index_datasheets.htm.

The Zener Diode

24 Diodes can be manufactured so that breakdown occurs at lower and more precise voltages than those just discussed. These types of diodes are called *zener diodes*, so named because they exhibit the "Zener effect"—a particular form of voltage breakdown. At the *zener voltage*, a small current flows through the zener diode. This current must be maintained to keep the diode at the *zener point*. In most cases, a few milliamperes are all that is required. Figure 2.26 shows the zener diode symbol and a simple circuit.

Figure 2.26

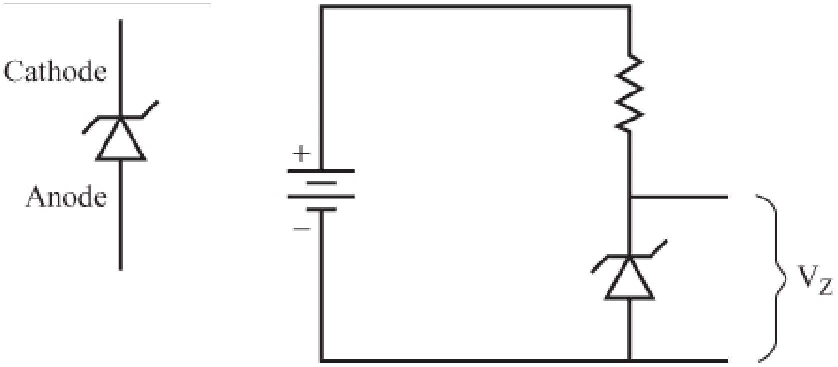

In this circuit, the battery determines the applied voltage. The zener diode determines the voltage drop (labeled V_Z) across it. The resistor determines the current flow. Zeners are used to maintain a constant voltage at some point in a circuit.

Question

Why are zeners used for this purpose, rather than ordinary diodes? _____

Answer

Because zeners have a precise breakdown voltage.

25 Examine an application in which a constant voltage is wanted—for example, a lamp driven by a DC generator. In this example, when the generator turns at full speed, it puts out 50 volts. When it runs more slowly, the voltage can drop to 35 volts. You want to illuminate a 20-volt lamp with this generator. Assume that the lamp draws 1.5 amperes. Figure 2.27 shows the circuit.

Figure 2.27

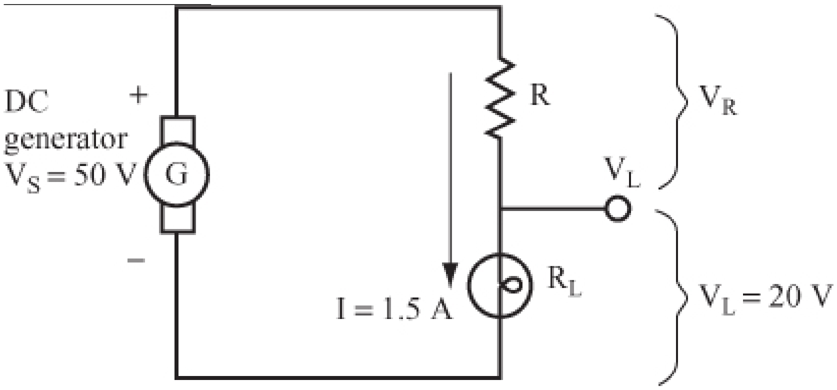

You need to determine a suitable value for the resistance. Follow these steps to find a suitable resistance value:

1. Find R_L, the lamp resistance. Use the following formula:

$$R_L = \frac{V_L}{I}$$

2. Find V_R. Use $V_S = V_R + V_L$.

3. Find R. Use the following formula:

$$R = \frac{V_R}{I}$$

Questions

Work through these steps, and write your answers here.

A. R_L = _____

B. V_R = _____

C. R = _____

Answers

A. $$R_L = \frac{20 \text{ volts}}{1.5 \text{ amperes}} = 13.33 \text{ ohms}$$

B. $$V_R = 50 \text{ volts} - 20 \text{ volts} = 30 \text{ volts}$$

C. $$R = \frac{50 \text{ volts} - 20 \text{ volts}}{1.5 \text{ amperes}} = \frac{30 \text{ volts}}{1.5 \text{ amperes}} = 20 \text{ ohms}$$

26 Assume now that the 20-ohm resistor calculated in problem 25 is in place, and the voltage output of the generator drops to 35 volts, as shown in Figure 2.28. This is similar to what happens when a battery gets old. Its voltage level decays and it will no longer have sufficient voltage to produce the proper current. This results in the lamp glowing less brightly, or perhaps not at all. Note, however, that the resistance of the lamp stays the same.

Figure 2.28

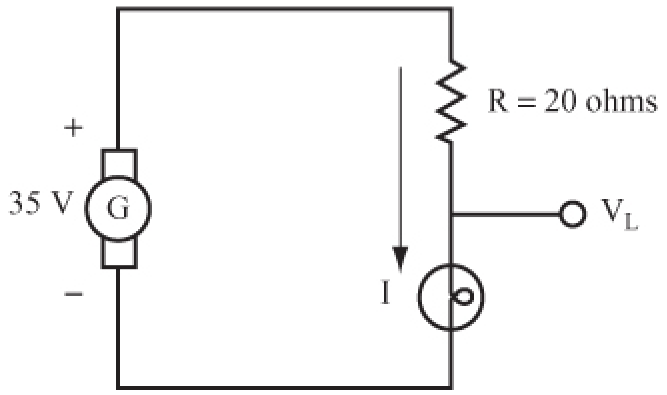

Questions

A. Find the total current flowing. Use the following formula:

$$I_T = \frac{V_S}{(R + R_L)}$$

$$I_T =$$

B. Find the voltage drop across the lamp. Use $V_L = I_T \times R_L$.

$V_L =$ _____

C. Have the voltage and current increased or decreased? _____

Answers

A.

$$I_T = \frac{35 \text{ volts}}{20\,\Omega + 13.3\,\Omega} = \frac{35 \text{ volts}}{33.3\,\Omega} = 1.05 \text{ amperes}$$

B.

$$V_L = 1.05 \text{ amperes} \times 13.3\Omega = 14 \text{ volts}$$

C. Both have reduced in value.

27 In many applications, a lowering of voltage across the lamp (or some other

component) may be unacceptable. You can prevent this by using a zener diode, as shown in the circuit in Figure 2.29 .

Figure 2.29

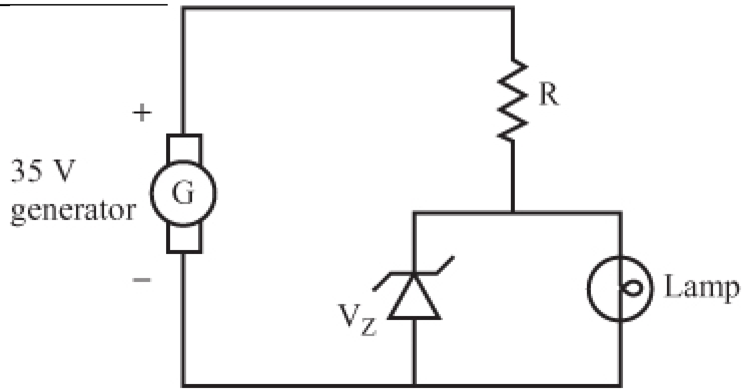

If you choose a 20-volt zener (that is, one that has a 20-volt drop across it), then the lamp always has 20 volts across it, no matter what the output voltage is from the generator (provided, of course, that the output from the generator is always above 20 volts).

Questions

Say that the voltage across the lamp is constant, and the generator output drops.

A. What happens to the current through the lamp? _____

B. What happens to the current through the zener? _____

Answers

A. The current stays constant because the voltage across the lamp stays constant.

B. The current decreases because the total current decreases.

28 To make this circuit work and keep 20 volts across the lamp at all times, you must find a suitable value of R. This value should allow sufficient total current to flow to provide 1.5 amperes required by the lamp, and the small amount required to keep the diode at its zener voltage. To do this, you start at the "worst case" condition. ("Worst case" design is a common practice in electronics. It is used to ensure that equipment can work under the most adverse conditions.) The worst case here would occur when the generator puts out only 35 volts. Figure 2.30 shows the paths that current would take in this circuit.

Figure 2.30

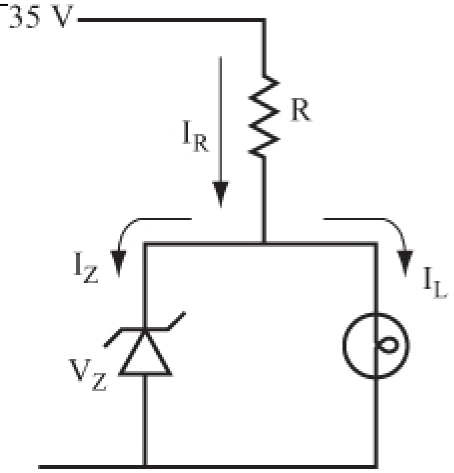

Find the value of R that enables 1.5 amperes to flow through the lamp. How much current can flow through the zener diode? You can choose any current you like, provided it is above a few milliamperes, and provided it does not cause the zener diode to burn out. In this example, assume that the zener current I_Z is 0.5 amperes.

Questions

A. What is the total current through R?

I_R = _____

B. Calculate the value of R.

R = _____

Answers

A.

$$I_R = I_L + I_Z = 1.5 \text{ amperes} + 0.5 \text{ amperes} = 2 \text{ amperes}$$

B.

$$R = \left(\frac{V_S - V_Z}{I_R}\right) = \frac{(35 \text{ volts} - 20 \text{ volts})}{2 \text{ amperes}} = 7.5 \text{ ohms}$$

A different choice of I_Z here would produce another value of R.

29 Now, take a look at what happens when the generator supplies 50 volts, as shown in Figure 2.31 .

Figure 2.31

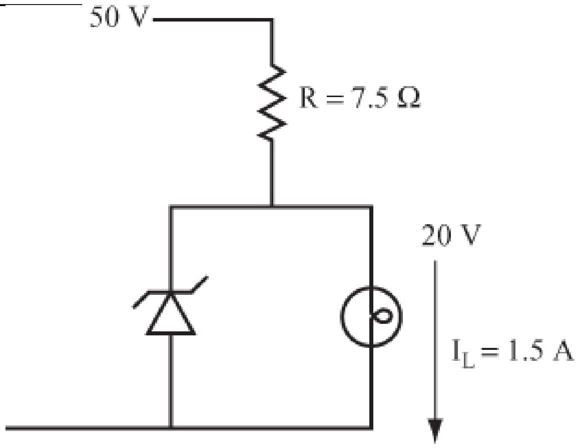

Because the lamp still has 20 volts across it, it can still draw only 1.5 amperes. But the total current and the zener current change.

Questions

A. Find the total current through R.

I_R = _____

B. Find the zener current.

I_Z = _____

Answers

A.
$$I_R = \frac{(V_S - V_Z)}{R} = \frac{(50 - 20)}{7.5} = 4\,\text{amperes}$$

B. $I_Z = I_R - I_L = 4 - 1.5 = 2.5$ amperes

30 Although the lamp voltage and current remain the same, the total current and the zener current both changed.

Questions

A. What has happened to I_T (I_R)? _____

B. What has happened to I_Z? _____

Answers

A. I_T has increased by 2 amperes.

B. I_Z has increased by 2 amperes.

The increase in I_T flows through the zener diode and not through the lamp.

31 The power dissipated by the zener diode changes as the generator voltage changes.

Questions

A. Find the power dissipated when the generator voltage is 35 volts. _____

B. Now, find the power when the generator is at 50 volts. _____

Answers

A. $P_Z = V \times I$ = (20 volts) (0.5 ampere) = 10 watts

B. $P_Z = V \times I$ = (20 volts) (2.5 ampere) = 50 watts

If you use a zener diode with a power rating of 50 watts or more, it does not burn out.

32 Use Figure 2.32 to answer the following question.

Figure 2.32

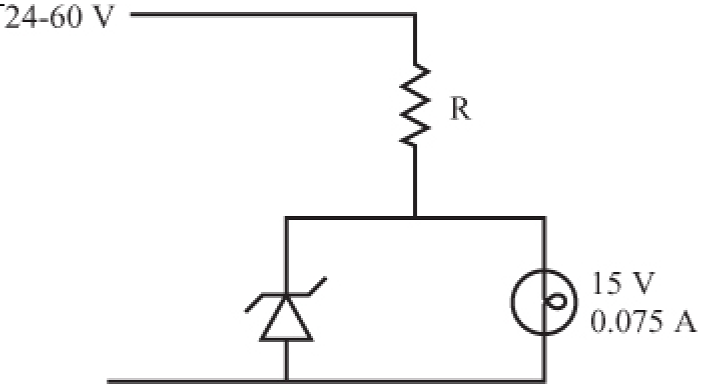

Question

For the circuit shown in Figure 2.32, what power rating should the zener diode have? The current and voltage ratings of the lamp are given. _____

Answer

At 24 volts, assuming a zener current of 0.5 ampere:

$$R = \frac{9}{0.575} = 15.7\,\text{ohms}$$

At 60 volts:

$$I_R = \frac{45}{15.7} = 2.87 \text{ amperes; therefore } I_Z \quad 2.8 \text{ amperes}$$

P_Z = (15 volts)(2.8 amperes) = 42 watts

Project 2.2: The Zener Diode Voltage Regulator

Objective

The objective of this project is to measure the voltage applied to the lamp, and the current through the lamp for different supply voltages, demonstrating the use of a zener diode to provide a steady voltage and current to a lamp when the supply voltage changes.

General Instructions

While the circuit is set up, measure the lamp current, zener diode current, and supply voltage as the voltage from the 9-volt battery drops.

Parts List

- One 9-volt battery
- One battery snap connector
- Two multimeters set to measure current (mA)
- One multimeter set to measure DC voltage
- One 56-ohm, 0.5-watt resistor
- One 1N4735A zener diode
- One breadboard
- One lamp rated for approximately 25 mA at 6 volts. (Part # 272-1140 from Radio Shack is a good fit for this project.)
- Two terminal blocks

Step-by-Step Instructions

Set up the circuit shown on Figure 2.33. The circled "A" designates a multimeter set to measure current, and the circled "V" designates a multimeter set to measure DC voltage. If you have some experience in building circuits, this schematic (along with the previous parts list) should provide all the information you need to build the circuit. If you need a bit more help in building the circuit, look at the photos of the completed circuit in the "Expected Results" section.

Figure 2.33

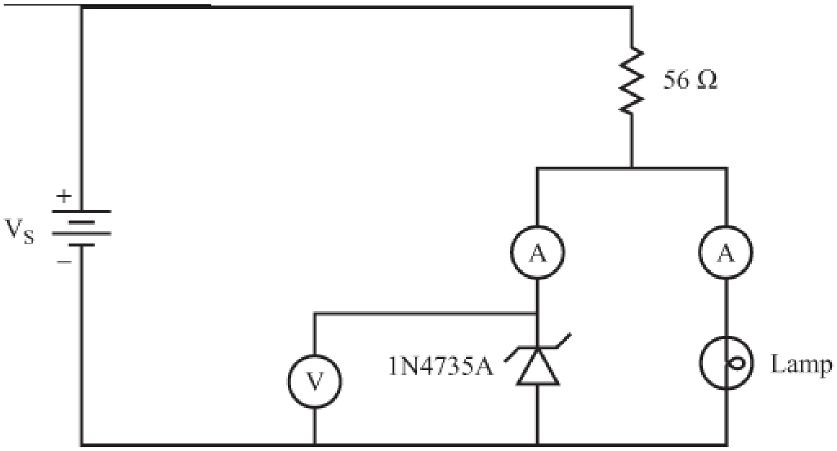

Carefully check your circuit against the diagram, especially the direction of the battery and the diode. The diode has a band at one end. Connect the lead at the end of the diode without the band to the ground bus on the breadboard.

After you check your circuit, follow these steps, and record your measurements in the blank table following the steps:

1. Measure and record the supply voltage.
2. Measure and record the lamp current.
3. Measure and record the zener current.
4. Wait 30 minutes.
5. Measure and record the new values of voltage and current.
6. Repeat steps 4 and 5 four times.

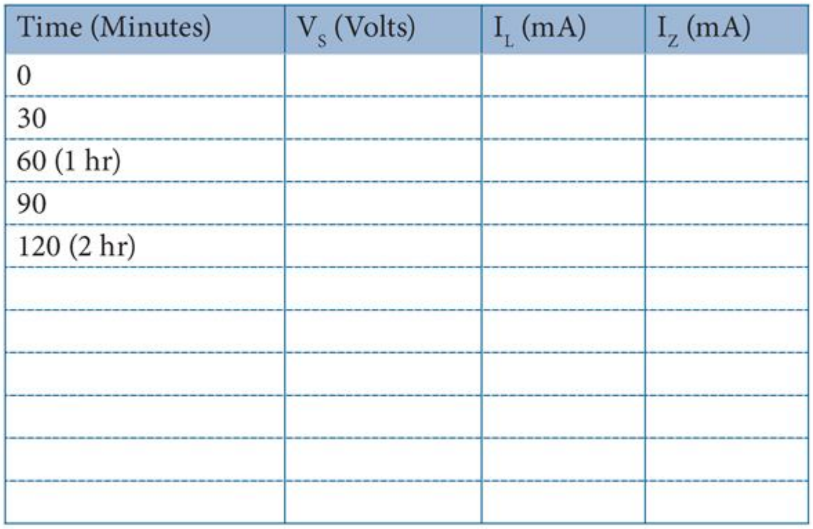

Time (Minutes)	V_S (Volts)	I_L (mA)	I_Z (mA)
0			
30			
60 (1 hr)			
90			
120 (2 hr)			

Expected Results

Figure 2.34 shows the breadboarded circuit for this project.

Figure 2.34

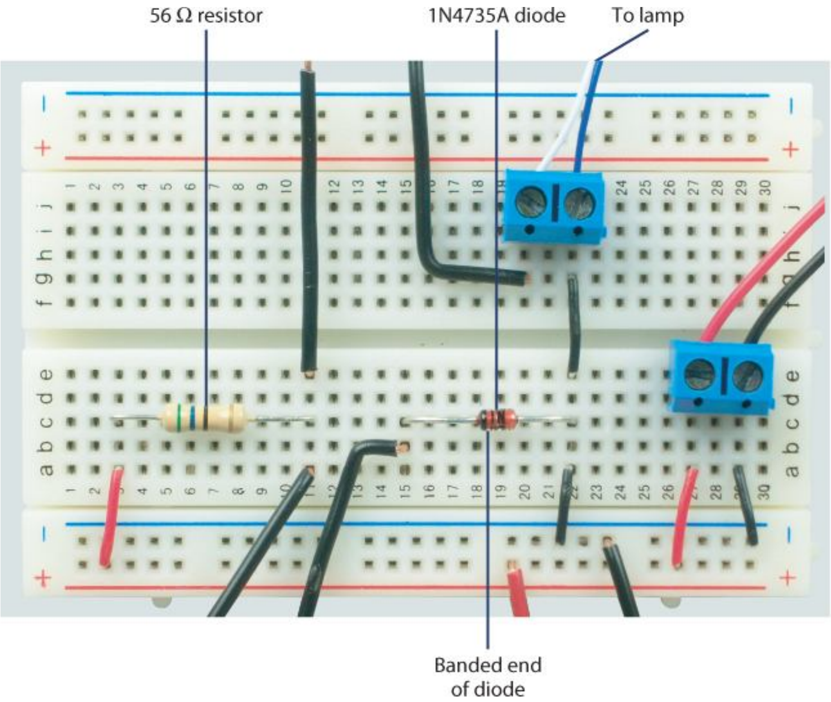
56 Ω resistor
1N4735A diode
To lamp
Banded end of diode

Figure 2.35 shows the test setup for this project.

Figure 2.35

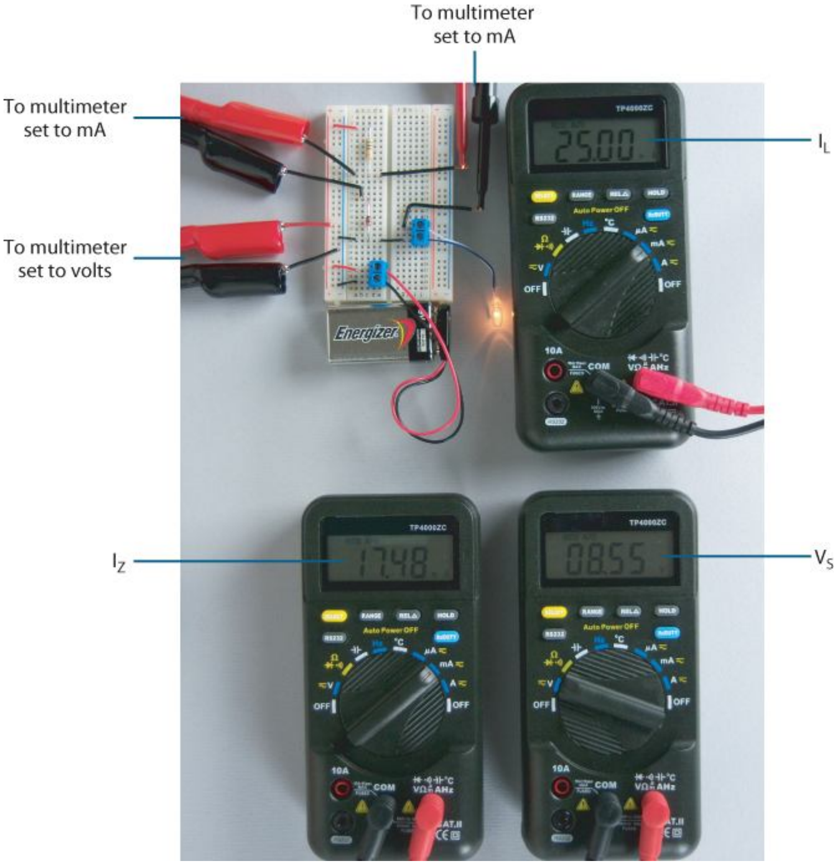
To multimeter set to mA
To multimeter set to mA
To multimeter set to volts
I_L
25.00
I_Z
17.48
V_S
08.55

Compare your measurements with the ones shown in the following table. You should see a similar trend in the measured values, but not exactly the same values.

Time (Minutes)	V_S (Volts)	I_L (mA)	I_Z (mA)
0	9.06	25.1	25.6
30	8.45	25.0	15.6
60 (1 hr)	8.21	25.0	11.8
90	8.03	24.9	9.0

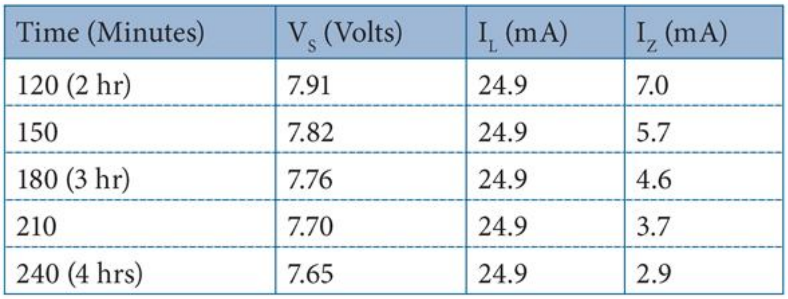

Time (Minutes)	V_S (Volts)	I_L (mA)	I_Z (mA)
120 (2 hr)	7.91	24.9	7.0
150	7.82	24.9	5.7
180 (3 hr)	7.76	24.9	4.6
210	7.70	24.9	3.7
240 (4 hrs)	7.65	24.9	2.9

As you can see in this data, even though the supply voltage dropped by approximately 15 percent, the lamp current stayed roughly constant, showing less than a 1 percent drop.

Summary

Semiconductor diodes are used extensively in modern electronic circuits. Following are the main advantages of semiconductor diodes:

- They are small.
- They are rugged and reliable if properly used. You must remember that excessive reverse voltage or excessive forward current could damage or destroy the diode.
- Diodes are easy to use because there are only two connections to make.
- They are inexpensive.
- They can be used in all types of electronic circuits, from simple DC controls to radio and TV circuits.
- They can be made to handle a wide range of voltage and power requirements.

Specialized diodes (which have not been covered here) can perform particular functions, which no other components can handle.

Finally, as you see in Chapter 3, "Introduction to the Transistor," diodes are an integral part of transistors.

All the many uses of semiconductor diodes are based on the fact they conduct in *one direction only*. Diodes are often used for the following:

- Protecting circuit components from voltage spikes
- Converting AC to DC
- Protecting sensitive components from high-voltage spikes
- Building high-speed switches
- Rectifying radio frequency signals

Self-Test

The following questions test your understanding of this chapter. Use a separate sheet of paper for your diagrams or calculations. Compare your answers with the answers that follow the test.

1. Draw the circuit symbol for a diode, labeling each terminal. _____
2. What semiconductor materials are used in the manufacture of diodes? _____
3. Draw a circuit with a battery, a resistor, and a forward-biased diode. _____

4. What is the current through a reverse-biased perfect diode? _____

5. Draw a typical V-I curve of a forward-biased diode. Show the knee voltage. _____

6. What is the knee voltage for silicon? _____ Germanium? _____

7. In the circuit shown in Figure 2.36, V_S = 10 volts and R = 100 ohms. Find the current through the diode, assuming a perfect diode. _____

Figure 2.36

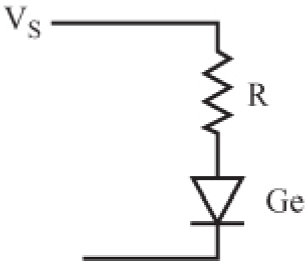

8. Calculate question 7 using these values: V_S = 3 volts and R = 1 kΩ. _____

9. In the circuit shown in Figure 2.37, find the current through the diode.

V_S = 10 volts
R_1 = 10 kΩ
R_2 = 1 kΩ

_____

Figure 2.37

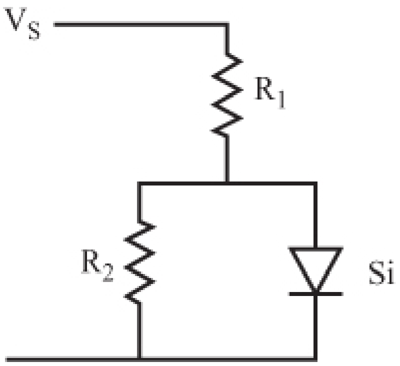

10. In the circuit shown in Figure 2.38, find the current through the zener diode.

$V_S = 20$ volts
$V_Z = 10$ volts
$R_1 = 1$ kΩ
$R_2 = 2$ kΩ

_____

Figure 2.38

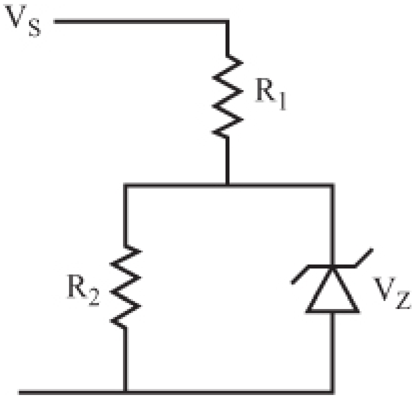

11. If the supply voltage for question 10 increases to 45 volts, what is the current in the zener diode? _____

12. What is the maximum power dissipated for the diode in questions 10 and 11? _____

Answers to Self-Test

If your answers do not agree with those given here, review the problems indicated in parentheses before you go to Chapter 3, "Introduction to the Transistor."

1. See Figure 2.39.

Figure 2.39

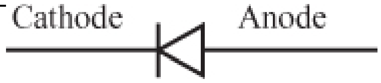

(problem 3)

2. Germanium and silicon. (problem 1)

3. See Figure 2.40 .

Figure 2.40

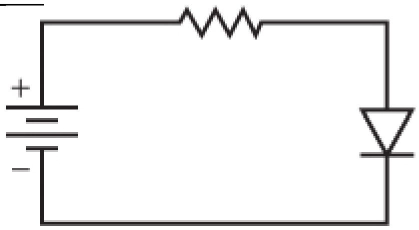

(problem 4)

4. There is zero current flowing through the diode. (problem 6)

5. See Figure 2.41 .

Figure 2.41

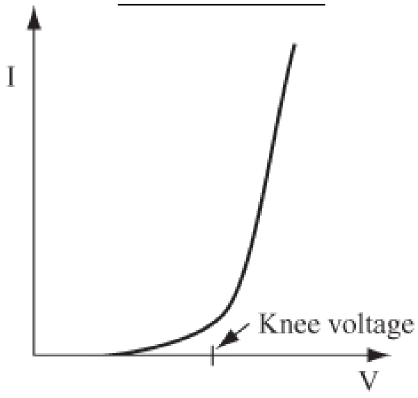

(Project 2.1)

6. Si = 0.7 volt; Ge = 0.3 volt (These are approximate.) (Project 2.1)
7. I_D = 100 mA. (problem 12)
8. As V_S = 3 volt, do not ignore the voltage drop across the diode. Thus, I_D = 2.7 mA. (problem 12)
9. Ignore V_D in this case. Thus, I_D = 0.3 mA. If V_D is not ignored, I_D = 0.23 mA. (problem 19)
10. I_Z = 5 mA. (problem 29)
11. I_Z = 30 mA. (problem 29)
12. The maximum power will be dissipated when I_Z is at its peak value of 30 mA. Therefore, $P_{Z(MAX)}$ = 0.30 watt. (problem 31)

Chapter 3

Introduction to the Transistor

The transistor is undoubtedly the most important modern electronic component because it has enabled great and profound changes in electronics and in your daily lives since its discovery in 1948.

This chapter introduces the transistor as an electronic component that acts similarly to a simple mechanical switch, and it is actually used as a switch in many modern electronic devices. A transistor can be made to conduct or not conduct an electric current—exactly what a mechanical switch does.

Most transistors used in electronic circuits are *bipolar junction transistors* (*BJTs*), commonly referred to as *bipolar transistors* *junction field effect transistors* (*JFETs*) or *metal oxide silicon field effect transistors* (*MOSFETs*). This chapter (along with Chapter 4, “The Transistor Switch,” and Chapter 8, “Transistor Amplifiers”) illustrates how BJTs and JFETs function and how they are used in electronic circuits. Because JFETs and MOSFETs function in similar fashion, MOSFETs are not covered here.

Projects in this chapter can help you to build a simple one-transistor circuit. You can easily set up this circuit on a home workbench. You should take the time to obtain the few components required, and actually build and operate the circuit.

In Chapter 4, you continue to study transistor circuits and the operation of the transistor as a switch. In Chapter 8, you learn how a transistor can be made to operate as an amplifier. In this mode, the transistor produces an output that is a magnified version of an input signal, which is useful because many electronic signals require amplification. These chapters taken together present an easy introduction to how transistors function, and how they are used in basic electronic circuits.

When you complete this chapter, you can do the following:

Describe the basic construction of a BJT.
Describe the basic construction of a JFET.
Specify the relationship between base and collector current in a BJT.
Specify the relationship between gate voltage and drain current in a JFET.
Calculate the current gain for a BJT.
Compare the transistor to a simple mechanical switch.

Understanding Transistors

1 The diagrams in Figure 3.1 show some common transistor cases (also called *packages*). The cases protect the semiconductor chip on which the transistor is built and provide leads that can be used to connect it to other components. For each transistor, the diagrams show the lead designations and how to identify them based on the package design.

Transistors can be soldered directly into a circuit, inserted into sockets, or inserted into breadboards. When soldering, you must take great care because transistors can be destroyed if overheated. A heat sink clipped to the transistor leads between the solder joint and the transistor case can reduce the possibility of overheating. If you use a socket, you can avoid exposing the transistor to heat by soldering the connections to the socket before inserting the transistor.

Figure 3.1

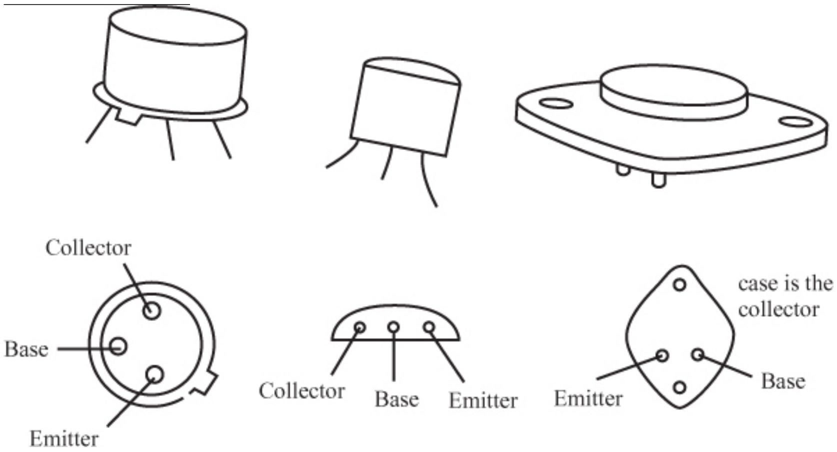

Questions

A. How many leads are there on most transistors? _____

B. Where there are only two leads, what takes the place of the third lead? _____

C. What are the three leads or connections called? _____

D. Why should you take care when soldering transistors into a circuit? _____

Answers

A. Three.

B. The case can be used instead, as indicated in the diagram on the right side of Figure 3.1 . (This type of case is used for power transistors.)

C. Emitter, base, and collector.

D. Excessive heat can damage a transistor.

2 You can think of a bipolar junction transistor as functioning like two diodes, connected back-to-back, as illustrated in Figure 3.2 .

Figure 3.2

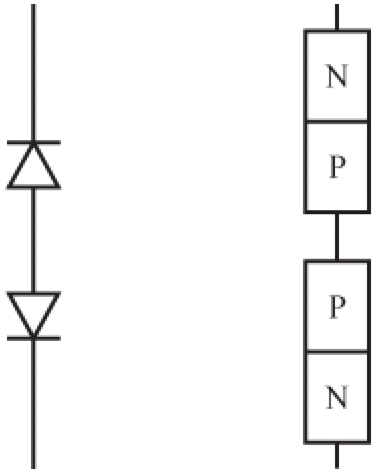

However, in the construction process, one important modification is made. Instead of two separate P regions, as shown in Figure 3.2 , only one thin region is used, as shown in Figure 3.3 .

Figure 3.3

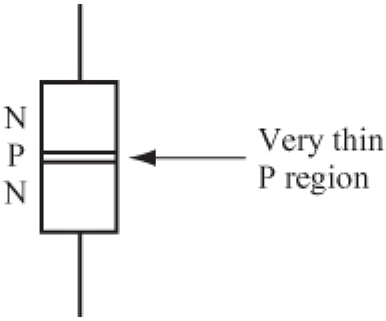

Question

Which has the thicker P region, the transistor shown in Figure 3.3 or two diodes connected back-to-back? _____

Answer

Two diodes. The transistor has a thin P region.

3 Because two separate diodes wired back-to-back share two thick P regions, they cannot behave like a transistor.

Question

Why don't two diodes connected back-to-back function like a transistor? _____

Answer

The transistor has one thin P region, whereas the diodes share two thick P regions.

4 The three terminals of a transistor (the base, the emitter, and the collector) connect, as shown in Figure 3.4 .

Figure 3.4

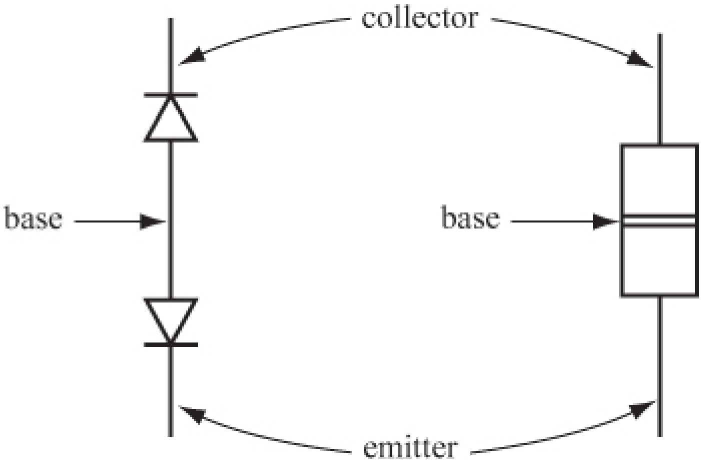
collector
base
base
emitter

When talking about a transistor as two diodes, you refer to the diodes as the *base-emitter diode* and the *base-collector diode* .

Figure 3.5 shows the symbol used in circuit diagrams for the transistor, with the two diodes and the junctions shown for comparison. Because of the way the semiconductor materials are arranged, this is known as an *NPN transistor* .

Figure 3.5

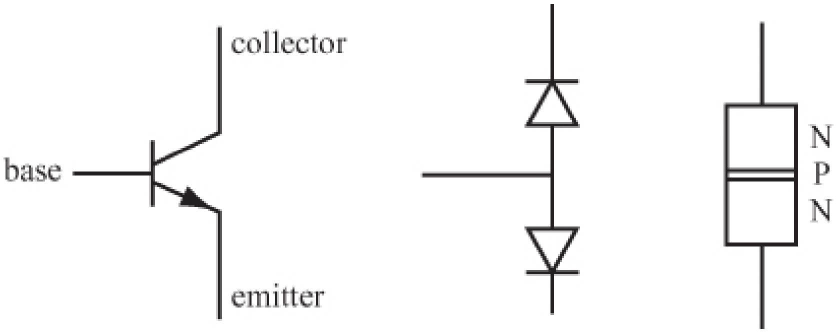
collector
base
emitter
N
P
N

Question

Which transistor terminal includes an

arrowhead? _____

Answer

The emitter

5 It is also possible to make transistors with a PNP configuration, as shown in Figure 3.6 .

Figure 3.6

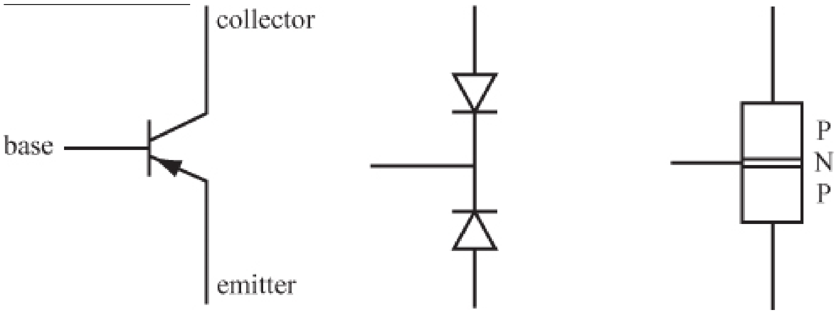

Both NPN and PNP type transistors can be made from either silicon or germanium.

Questions

A. Draw a circuit symbol for both an NPN and a PNP transistor. (Use a separate sheet of paper for your drawings.)

B. Which of the transistors represented by these symbols might be silicon? _____

C. Are silicon and germanium ever combined in a transistor? _____

Answers

A. See Figure 3.7 .

B. Either or both could be silicon. (Either or both could also be germanium.)

C. Silicon and germanium are *not* mixed in any commercially available transistors. However, researchers are attempting to develop ultra-fast transistors that contain both silicon and germanium.

Figure 3.7

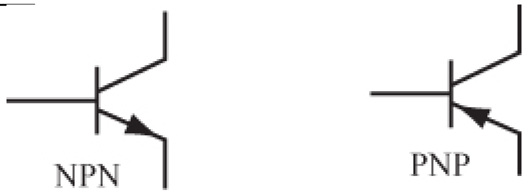

6 Take a look at the simple examples using NPN transistors in this and the next few problems.

If a battery is connected to an NPN transistor, as shown in Figure 3.8, then a current will flow in the direction shown.

Figure 3.8

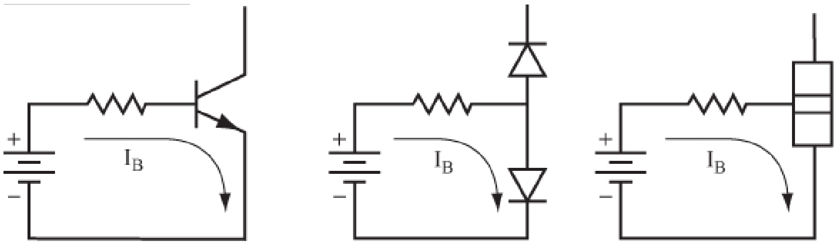

The current flowing through the base-emitter diode is called *base current* and is represented by the symbol I_B.

Question

Would base current flow if the battery were reversed? Give a reason for your answer.

_____

Answer

Base current would not flow because the diode would be back-biased.

7 In the circuit shown in Figure 3.9, you can calculate the base current using the techniques covered in Chapter 2, “The Diode.”

Figure 3.9

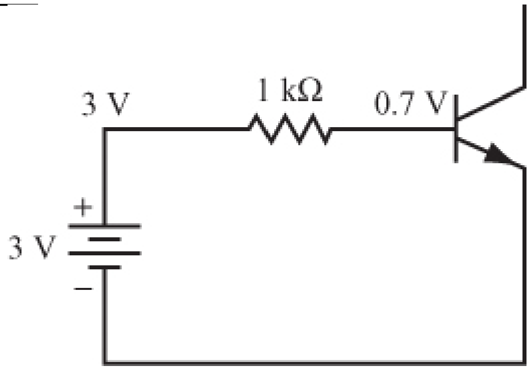

Question

Find the base current in the circuit shown in Figure 3.9 . (*Hint:* Do not ignore the 0.7-volt drop across the base-emitter diode.)

I B = _____

Answer

Your calculations should look something like this:

$$I_B = \frac{(V_S \quad 0.7\,\text{volt})}{R} = \frac{(3 \quad 0.7)}{1\text{k}\Omega} = \frac{2.3\,\text{volts}}{1\text{k}\Omega} = 2.3\,\text{mA}$$

8 For the circuit shown in Figure 3.10 , because the 10 volts supplied by the battery is much greater than the 0.7-volt diode drop, you can consider the base-emitter diode to be a perfect diode and thus assume the voltage drop is 0 volts.

Figure 3.10

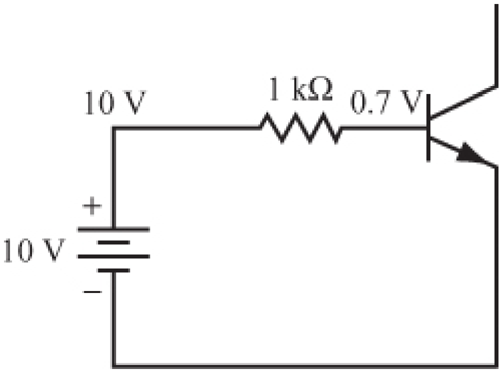
10 V
1 kΩ
0.7 V
+
10 V
−

Question

Calculate the base current.

I_B = _____

Answer

$$I_B = \frac{(10 \quad 0)}{1\text{k}\Omega} = \frac{10}{1\text{k}\Omega} = 10\,\text{mA}$$

9 Look at the circuit shown in Figure 3.11 .

Figure 3.11

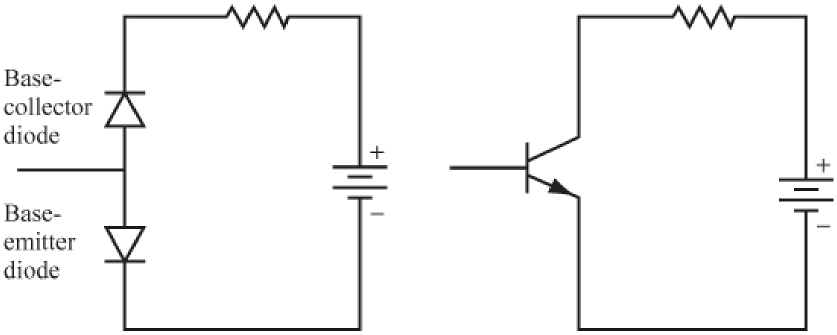
Base-collector diode
Base-emitter diode
+
−
+
−

Question

Will current flow in this circuit? Why or why not? _____

Answer

It cannot flow because the base-collector diode is reverse-biased.

10 Examine the circuit shown in Figure

3.12 . Batteries are connected to both the base and collector portions of the circuit.

Figure 3.12

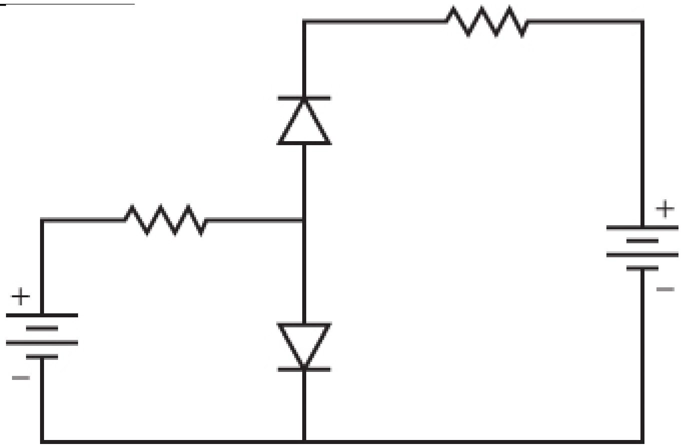

When you connect batteries to both the base and the collector portions of the circuit, currents flowing through the circuit demonstrate a key characteristic of the transistor. This characteristic is sometimes called *transistor action* —if base current flows in a transistor, collector current will also flow.

Examine the current paths shown in Figure 3.13 .

Figure 3.13

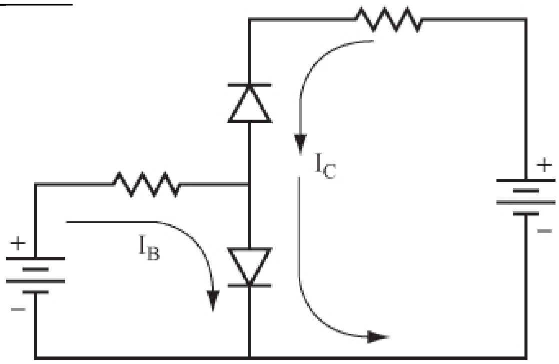

Questions

A. What current flows through the base-collector diode? _____

B. What current flows through the base-emitter diode? _____

C. Which of these currents causes the other to flow? _____

Answers

A. I_C (the collector current).

B. I_B and I_C. Both of them flow through the base-emitter diode.

C. Base current causes collector current to flow.

No current flows along the path shown by the dotted line in Figure 3.14 from the collector to the base.

Figure 3.14

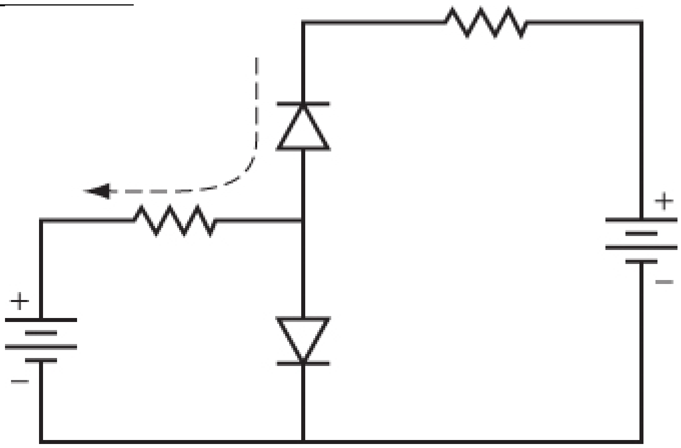

11 Up to now, you have studied the NPN bipolar transistor. PNP bipolar transistors work in the same way as NPN bipolar transistors and can also be used in these circuits.

There is, however, one important circuit difference, which is illustrated in Figure 3.15 . The PNP transistor is made with the diodes oriented in the reverse direction from the NPN transistor.

Figure 3.15

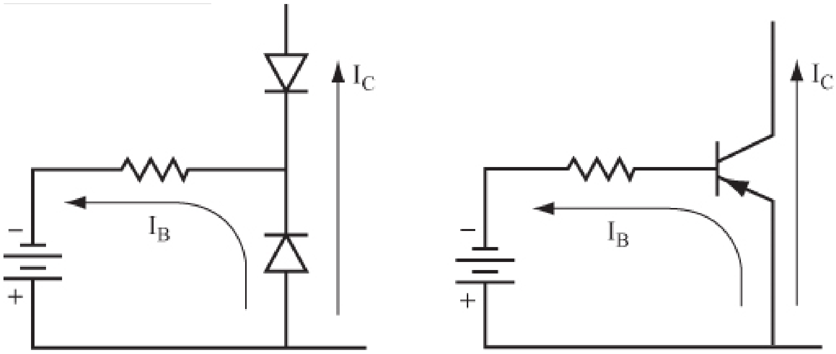

Questions

Compare Figure 3.15 with Figure 3.13 . How are the circuits different relative to the following?

A. Battery connections: _____

B. Current flow: _____

Answers

A. The battery is reversed in polarity.

B. The currents flow in the opposite direction.

12 Figure 3.16 shows the battery connections necessary to produce both base current and collector current in a circuit that uses a PNP transistor.

Figure 3.16

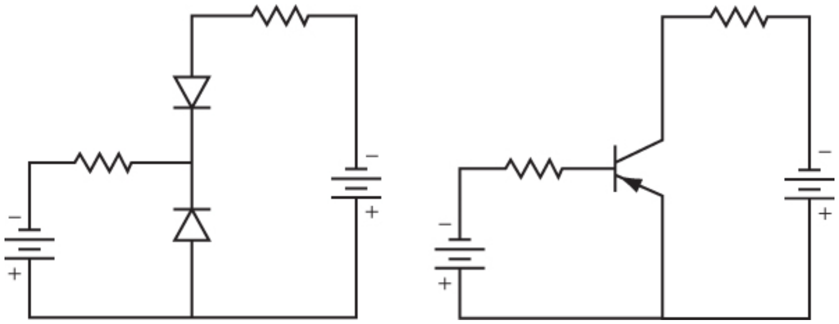

Question

In which direction do these currents circulate—clockwise or counterclockwise?

—————

Answer

Base current flows counterclockwise.

Collector current flows clockwise.

As stated earlier, NPN and PNP bipolar transistors work in much the same way: Base current causes collector current to flow in both. The only significant difference in using a PNP versus an NPN bipolar transistor is that the polarity of the supply voltage (for both the base and collector sections of the circuit) is reversed. To avoid confusion, bipolar transistors used throughout the rest of this book are NPNs.

13 Consider the circuit shown in Figure 3.17. It uses only one battery to supply voltage to both the base and the collector portions of the circuit. The path of the base current is shown in the diagram.

Figure 3.17

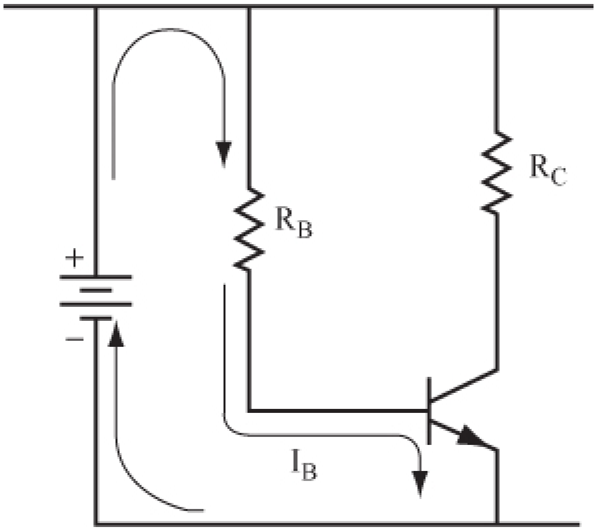

Questions

A. Name the components through which the base current flows. _____

B. Into which terminal of the transistor does the base current flow? _____

C. Out of which transistor terminal does the base current flow? _____

D. Through which terminals of the transistor does base current not flow? _____

Answers

A. The battery, the resistor R B , and the transistor

B. Base

C. Emitter

D. Collector

14 Take a moment to recall the key physical characteristic of the transistor.

Question

When base current flows in the circuit shown in Figure 3.17 , what other current can flow, and which components will it flow through?

_____

Answer

Collector current will flow through the resistor R_C and the transistor.

15 In Figure 3.18 the arrows indicate the path of the collector current through the circuit.

Figure 3.18

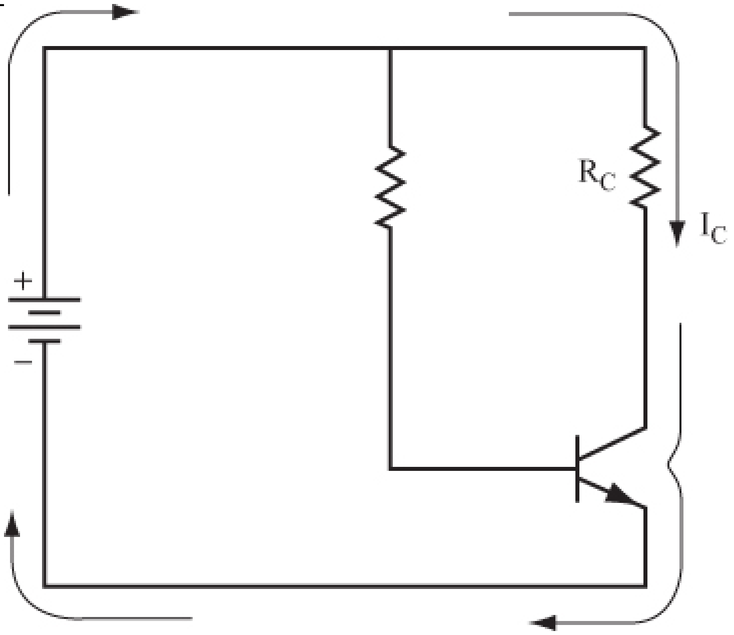

Questions

A. List the components through which the collector current flows. _____

B. What causes the collector current to flow?

_____

Answers

A. The resistor R_C, the transistor, and the battery.

B. Base current. (Collector current doesn't flow unless base current is flowing.)

16 It is a property of the transistor that the *ratio of collector current to base current is constant*. The collector current is always much larger than the base current. The ratio of the two currents is called the *current gain* of the transistor, and is represented by the symbol β, or beta. Typical values of β range from 10 to 300.

Questions

A. What is the ratio of collector current to base current called? _____

B. What is the symbol used to represent this? _____

C. Which is larger—base or collector current? _____

D. Look back at the circuit in problem 13. Will current be greater in R_B or in R_C? _____

Answers

A. Current gain.

B. β.

C. Collector current is larger.

D. The current is greater in R_C because it is the collector current.

Note The β introduced here is referred to in manufacturers' specification sheets as h_{FE}. Technically, it is referred to as the static or DC β. For the purposes of this chapter, it is called β. Discussions on transistor parameters in general, which are well covered in many textbooks, will not be covered here.

17 The mathematical formula for current gain is as follows:

$$\beta = \frac{I_C}{I_B}$$

In this equation, the following is true:

I_B = base current

I_C = collector current

The equation for β can be rearranged to give $I_C = \beta I_B$. From this, you can see that if no base current flows, no collector current flows. Also, if more base current flows, more collector current flows. This is why it's said that the base current controls the collector current.

Question

Suppose the base current is 1 mA and the collector current is 150 mA. What is the current gain of the transistor? _____

Answer

150

Inside the Bipolar Transistor

Now take a closer look inside a bipolar transistor. In a bipolar transistor, with no voltages applied, two depletion regions exist. As shown in the following figure, one *depletion region* exists at the emitter and base junction, and one at the collector and base junction. There are no free electrons or holes in these depletion regions, which prevents any current from flowing.

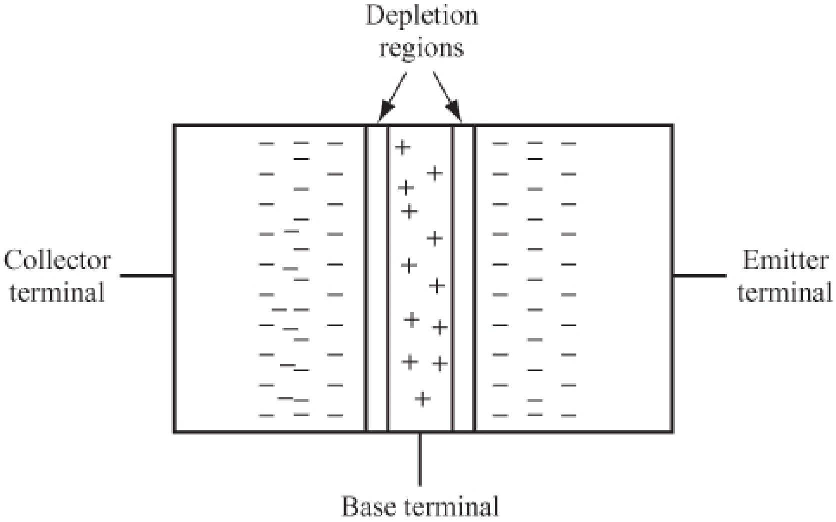

The following figure shows a bipolar transistor with a positive voltage applied to the base and a negative voltage applied to the emitter (forward-biasing the base-emitter diode), as well as a positive voltage applied to the collector. The forward bias on the base-emitter diode allows electrons to flow from the emitter into the base. A small fraction of these electrons would be captured by the holes in the base region and then flow out of the base terminal and through the base resistor. The base of a transistor is thin, which allows most of the electrons from the emitter to flow through the base and into the collector. The positive voltage on the collector terminal attracts these electrons, which flow out of the collector terminal and through the collector resistor.

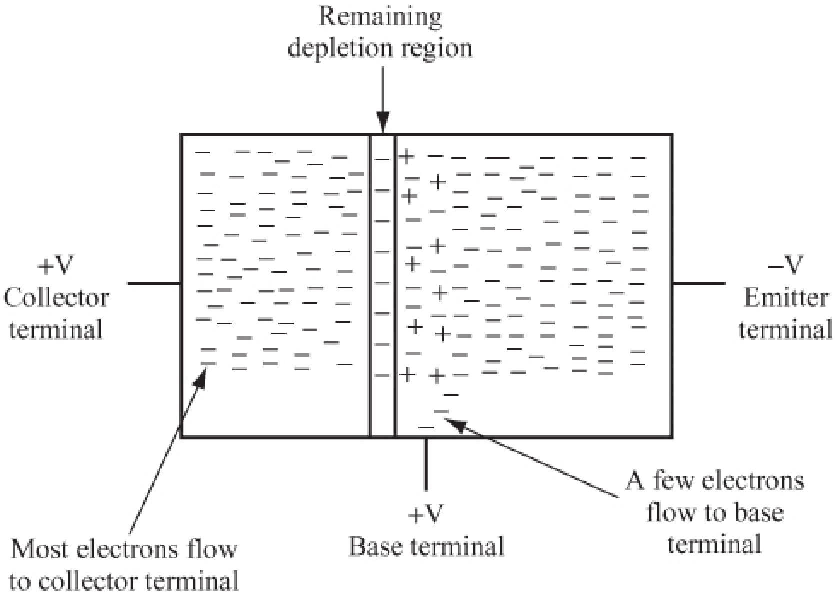
Remaining depletion region
+V
Collector terminal
–V
Emitter terminal
+V
Base terminal
A few electrons flow to base terminal
Most electrons flow to collector terminal

For a transistor with β = 100, only one electron flows out to the base terminal for every 100 electrons that flow to the collector terminal. β is controlled by two factors: the thickness of the base region and the relative concentration of the impurities providing holes in the P region to the concentration of the impurities providing electrons in the N regions. A thinner base region plus the lower relative concentration of holes allow more electrons to pass through the base without being captured, resulting in a higher **β.** (Remember, the direction that conventional electrical current flows in is opposite to the direction in which electrons flow.)

Before you connect any bipolar transistor to other components in a circuit, you must identify the emitter, base, and collector leads (referred to as the transistor's *pinout*) and determine whether the transistor is NPN or PNP.

Transistors are marked with a part number, such as 2N3904, 2N3906, BC337, and PN2222. However, the part numbers don't tell you much about the transistor. For these transistors, the 2N3904, BC337, and PN2222 are NPN, whereas the 2N3906 is a PNP, which is not obvious from the part number.

Also, the transistor pinout is not identified on the part number. For example, one NPN transistor, the BC337, uses the opposite leads for the emitter and collector than the 2N3904 transistor, as shown in the following figure.

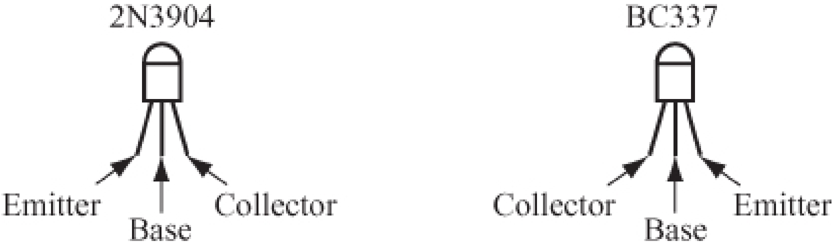

Your best bet is to refer to the manufacturer's data sheet for the transistor pinout and other characteristics. You can easily look up data sheets on the Internet. Also, you can find links to the data sheets for the transistors used in projects in this book on the website at www.buildinggadgets.com/index_datasheets.htm

.

18 *Current gain* is a physical property of transistors. You can find its value in the manufacturers' published data sheets, or you can determine it by experimenting.

In general, β is a different number from one transistor part number to the next, but transistors with the same part number have β values within a narrow range of each other.

One of the most frequently performed calculations in transistor work is to determine the values of either collector or base current, when β and the other current are known.

For example, suppose a transistor has 500 mA of collector current flowing, and you know it has a β value of 100. To find the base current, use the following formula:

$$\beta = \frac{I_C}{I_B}$$

$$I_B = \frac{I_C}{\beta} = \frac{500\,\text{mA}}{100} = 5\,\text{mA}$$

Questions

Calculate the following values:

A. I_C = 2 ampere, = 20. Find I_B. _____

B. I_B = 1 mA, β = 100. Find I_C. _____

C. I_B = 10 μA, β = 250. Find I_C. _____

D. I_B = 0.1 mA, I_C = 7.5 mA. Find β. _____

Answers

A. 0.1 ampere, or 100 mA

B. 100 mA

C. 2500 μA, or 2.5 mA

D. 75

19 This problem serves as a summary of the first part of this chapter. You should be able to answer all these questions. Use a separate sheet of paper for your drawing and calculations.

Questions

A. Draw a transistor circuit utilizing an NPN transistor, a base resistor, a collector resistor, and one battery to supply both base and collector currents. Show the paths of I_B and I_C.

B. Which current controls the other? _____

C. Which is the larger current, I_B or I_C? _____

D. $I_B = 6$ μA, $\beta = 250$. Find I_C. _____

E. $I_C = 300$ mA, $\beta = 50$. Find I_B. _____

Answers

A. See Figure 3.17 and Figure 3.18.

B. I_B (base current) controls I_C (collector current).

C. I_C

D. 1.5 mA

E. 6 mA

Project 3.1: The Transistor

Objective

The objective of this project is to find β of a particular transistor by setting several values of base current and measuring the corresponding values of collector current. Next, you divide the values of collector current by the values of the base current to determine β. The value of β will be almost the same for all the measured values of current. This demonstrates that β is a constant for a transistor.

General Instructions

While the circuit is set up, measure the collector voltage for each current value. This demonstrates (experimentally) some points that are covered in future problems. As you perform the project, observe how the collector voltage V_C drops as the collector current increases.

Parts List

- One 9 V battery (or a lab power supply)
- One multimeter set to μA
- One multimeter set to mA
- One multimeter set to measure DC voltage
- One 10 kΩ resistor
- One 510 ohm resistor
- One 2N3904 transistor
- One breadboard
- One 1 MΩ potentiometer

Step-by-Step Instructions

Set up the circuit shown in Figure 3.19 on a breadboard. If you have some experience in building circuits, this schematic (along with the previous parts list) should provide all the information you need to build the circuit. If you need a bit more help building the circuit, look at the photos of the completed circuit in the “Expected Results” section.

Figure 3.19

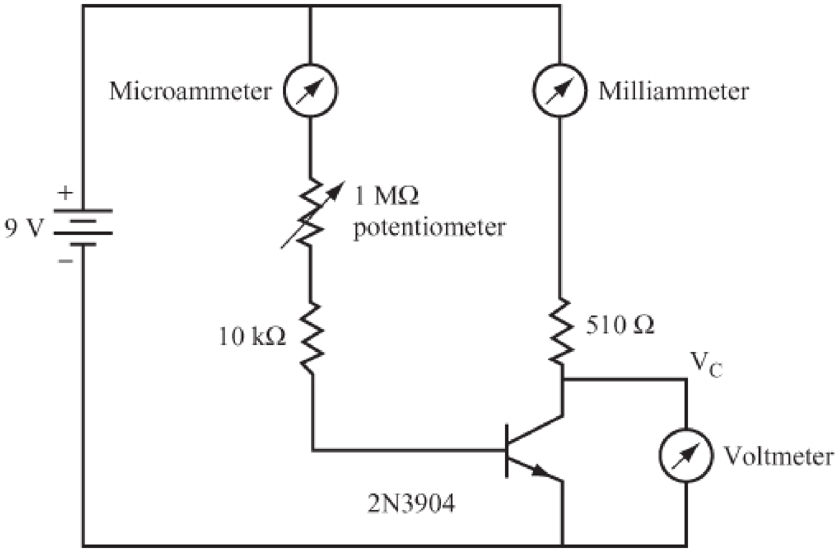
Microammeter
Milliammeter
+
9 V
−
1 MΩ
potentiometer
10 kΩ
510 Ω
V_C
Voltmeter
2N3904

Follow these steps, recording your measurements in the blank table following the steps.

1. Set the potentiometer to its highest value; this sets I_B to its lowest possible value.
2. Measure and record I_B.
3. Measure and record I_C.
4. Measure and record V_C. This voltage is sometimes referred to as the *collector-emitter voltage* (V_{CE}), because it is taken across the collector-emitter leads if the emitter is connected to ground or the negative of the power supply.

5. Adjust the potentiometer to give the next targeted value of I_B. You do not need to hit these values exactly. For example, for a target of 20 μA, a measured value of 20.4 μA is fine.
6. Measure and record the new values for I_B, I_C, and V_C.
7. Adjust the potentiometer to give the next targeted value of I_B.
8. Measure and record the new values for I_B, I_C, and V_C again.
9. Repeat steps 7 and 8 for each of the targeted values of I_B.
10. For each value of I_B and its corresponding value of I_C, calculate the value of β (β = I_C/I_B). The values will vary slightly but will be close to an average. Did you get a consistent β?

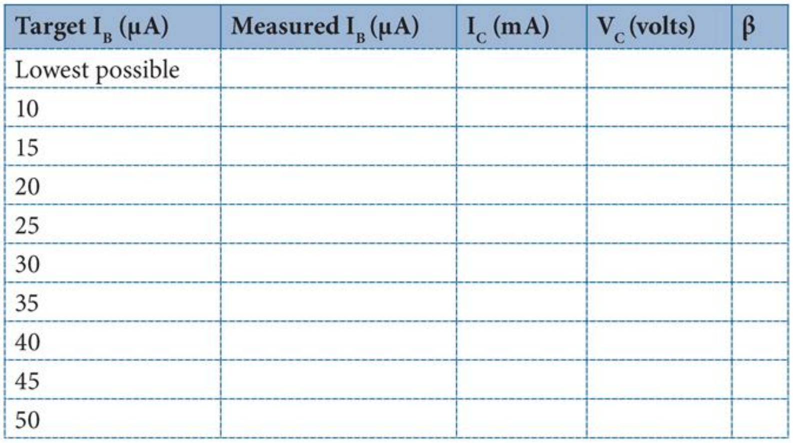

Target I_B (μA)	Measured I_B (μA)	I_C (mA)	V_C (volts)	β
Lowest possible				
10				
15				
20				
25				
30				
35				
40				
45				
50				

Save this circuit. You use it later in this chapter in Project 3.2.

Expected Results

Figure 3.20 shows the breadboarded circuit for this project.

Figure 3.20

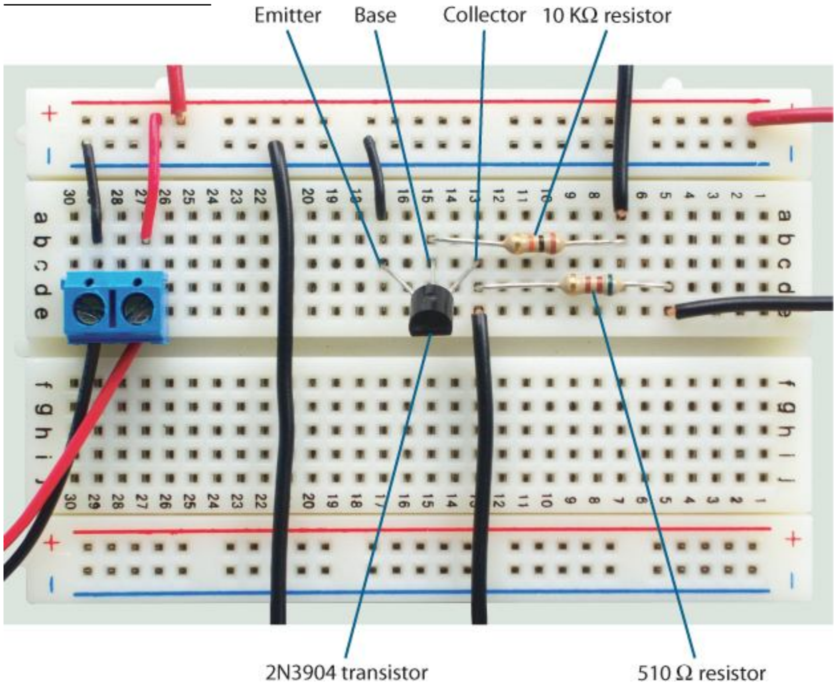
Emitter
Base
Collector
10 KΩ resistor
2N3904 transistor
510 Ω resistor

Figure 3.21 shows the test setup for this project.

Compare your measurements with the ones shown in the following table.

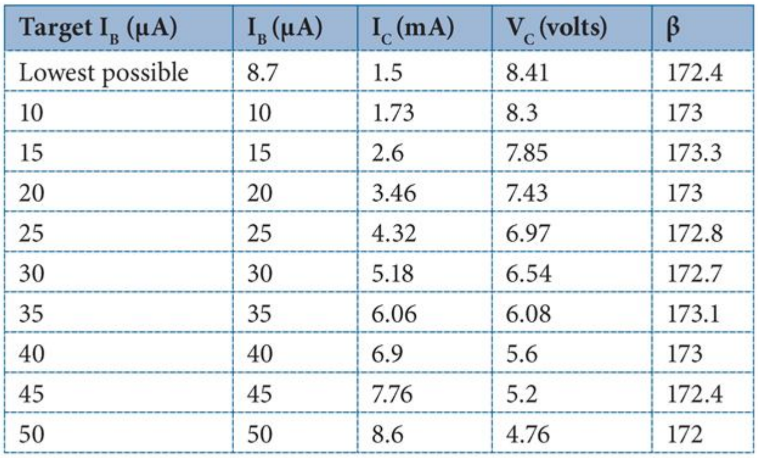

Target I_B (µA)	I_B (µA)	I_C (mA)	V_C (volts)	β
Lowest possible	8.7	1.5	8.41	172.4
10	10	1.73	8.3	173
15	15	2.6	7.85	173.3
20	20	3.46	7.43	173
25	25	4.32	6.97	172.8
30	30	5.18	6.54	172.7
35	35	6.06	6.08	173.1
40	40	6.9	5.6	173
45	45	7.76	5.2	172.4
50	50	8.6	4.76	172

Figure 3.21

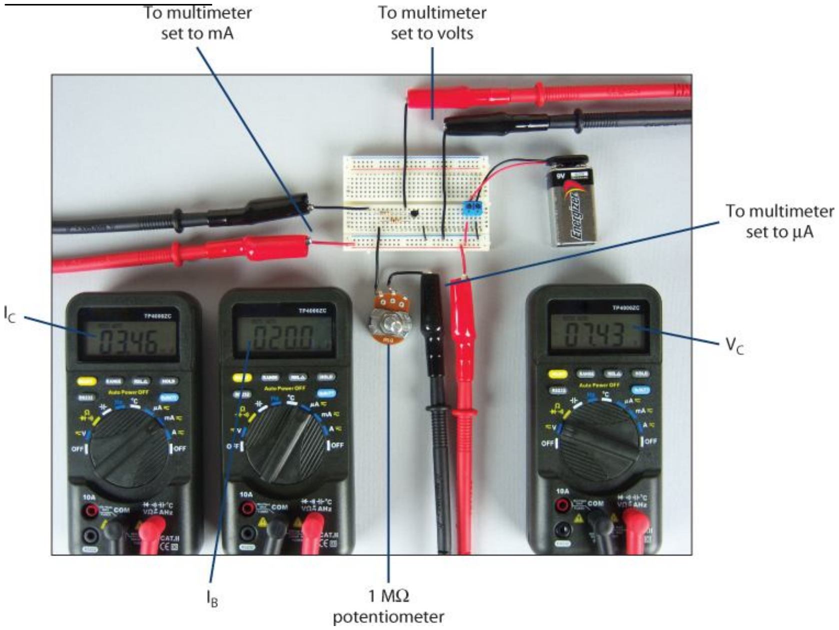
To multimeter set to mA
To multimeter set to volts
To multimeter set to µA
I_C
V_C
03.46
020.0
07.43
I_B
1 MΩ potentiometer

Don't worry if your results give a different value of β. The manufacturing process that produces transistors can allow variation of the base thickness and doping levels, which causes variation of β in the finished transistors.

20 In Project 3.1, you measured the voltage level at the collector (V_C) and recorded your measurements. Now, examine how to determine the voltage at the collector, when it's not possible to measure the voltage level.

Use the values shown in the circuit in Figure 3.22 to complete the following steps:

1. Determine I_C.
2. Determine the voltage drop across the collector resistor R_C. Call this V_R.
3. Subtract V_R from the supply voltage to calculate the collector voltage.

Figure 3.22

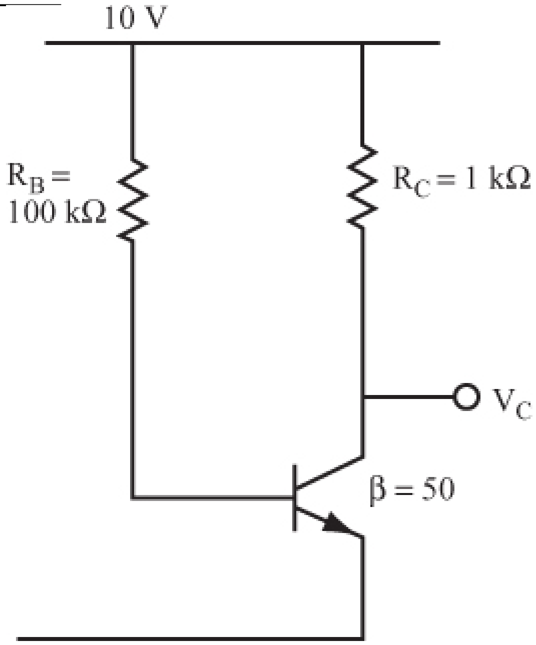

Here is the first step:

1. To find IC, you must first find IB.

$$I_B = \frac{10\,\text{volts}}{100\,\text{k}\Omega} = 0.1\,\text{mA}$$

$$I_C = \beta \quad I_B = 50 \quad 0.1\,\text{mA} = 5\,\text{mA}$$

Now, perform the next two steps.

Questions

A. VR = _____

B. VC = _____

Answers

A. To find VR:

VR = RC × IC = 1 kΩ × 5 mA = 5 volts

B. To find VC:

VC = VS – VR = 10 volts – 5 volts = 5 volts

21 Determine parameters for the circuit shown in Figure 3.22 using the value of β = 75.

Questions

Calculate the following:

A. IC = _____

B. VR = _____

C. VC = _____

Answers

A.

$$I_B = \frac{10\,\text{volts}}{100\,\text{k}\Omega} = 0.1\,\text{mA}$$

$$I_C = 75 \quad 0.1\,\text{mA} = 7.5\,\text{mA}$$

B. VR = 1 kΩ × 7.5 mA = 7.5 volts

C. VC = 10 volts – 7.5 volts = 2.5 volts

22 Determine parameters for the same circuit, using the values of RB = 250 kΩ and

$\beta = 75$.

Questions

Calculate the following:

A. I_C = _____

B. V_R = _____

C. V_C = _____

Answers

A.
$$I_B = \frac{10\,\text{volts}}{250\,\text{k}\Omega} = \frac{1}{25}\,\text{mA}$$
$$I_C = 75 \quad \frac{1}{25}\,\text{mA} = 3\,\text{mA}$$

B. V_R = 1 kΩ × 3 mA = 3 volts

C. V_C = 10 volts−3 volts = 7 volts

23 From the preceding problems, you can see that you can set V_C to any value by choosing a transistor with an appropriate value of β or by choosing the correct value of R_B.

Now, consider the example shown in Figure 3.23. The objective is to find V_C. Use the steps outlined in problem 20.

Figure 3.23

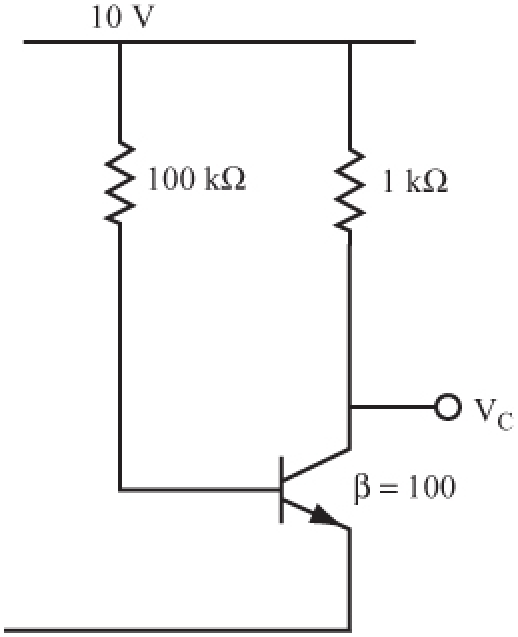

Questions

Calculate the following:

A. I_B = _____

I_C = _____

B. V_R = _____

C. V_C = _____

Answers

Your results should be as follows:

A.

$$I_B = \frac{10 \text{ volts}}{100\,\text{k}\Omega} = 0.1\,\text{mA}$$

$$I_C = 100 \quad 0.1\text{ mA} = 10\text{ mA}$$

B. V_R = 1 kΩ × 10 mA = 10 volts

C. V_C = 10 volts −10 volts = 0 volts.

Here the base current is sufficient to produce a collector voltage of 0 volts and the maximum collector current possible, given the stated values of the collector resistor and supply voltage. This condition is called *saturation*.

24 Look at the two circuits shown in Figure 3.24 and compare their voltages at the point labeled V_C.

Figure 3.24

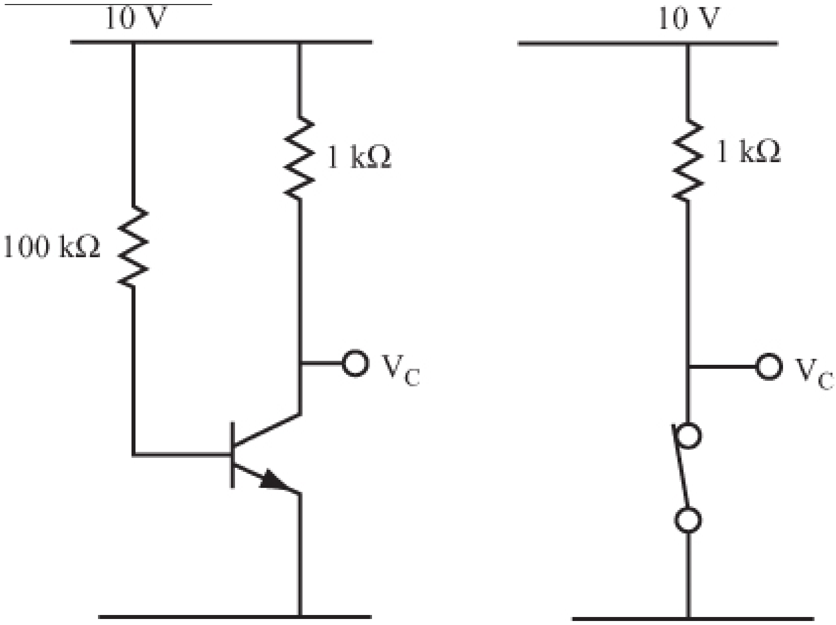

Consider a transistor that has sufficient base current and collector current to set its collector voltage to 0 volts. Obviously, this can be compared to a closed mechanical switch. Just as the switch is said to be ON, the transistor is also said to be "turned on" (or just ON).

Questions

A. What can you compare a turned on transistor to? _____

B. What is the collector voltage of an ON transistor? _____

Answers

A. A closed mechanical switch

B. 0 volts

Project 3.2: The Saturated Transistor

Objective

Normally, for a transistor, $I_C = \beta \times I_B$. However, this relationship does not hold when a transistor is saturated. The objective of this project is to determine the relationship of the collector current to the base current when a transistor is saturated.

General Instructions

Using the same breadboarded circuit you built in Project 3.1, set the base current to several values, starting at 90 μA and increasing the base current. Record measurements of the collector current and collector voltage at each value of the base current.

Step-by-Step Instructions

Follow these steps and record your measurements in the blank table following the steps.

1. Set up the circuit you built in Project 3.1.
2. Adjust the potentiometer to set I_B at approximately 90 μA.
3. Measure and record I_B.
4. Measure and record I_C.
5. Measure and record V_C.

6. Adjust the potentiometer slightly to lower its resistance, which sets a larger value of I_B.

7. Measure and record the new values for I_B, I_C, and V_C.

8. Repeat steps 6 and 7 until you reach the lower limit of the potentiometer, which is also the highest value of I_B.

I_B (μA)	I_C (mA)	V_C (volts)

Expected Results

Figure 3.25 shows the test setup for this project with the potentiometer set at its lower limit, providing the highest value of I_B. The test setup is the same as that used in Project 3.1; however, the value of I_B is considerably higher than in Project 3.1.

Compare your measurements to the ones shown in the following table.

I_B (μA)	I_C (mA)	V_C (volts)
91	14.5	1.53
101	15.9	0.843
126	16.9	0.329
150	17.0	0.256
203	17.1	0.211
264	17.1	0.189
389	17.2	0.163
503	17.2	0.149
614	17.2	0.139
780	17.2	0.127
806	17.2	0.126

Your data will probably have slightly different values but should indicate that I_C stays constant for values of V_C of 0.2 and below, whereas I_B continues to rise. In this region, the transistor is fully ON (saturated) and I_C can't increase further. This agrees with the data sheet published by Fairchild Semiconductor for the 2N3904 transistor, which indicates that the transistor saturates at V_C = 0.2 volts.

Figure 3.25

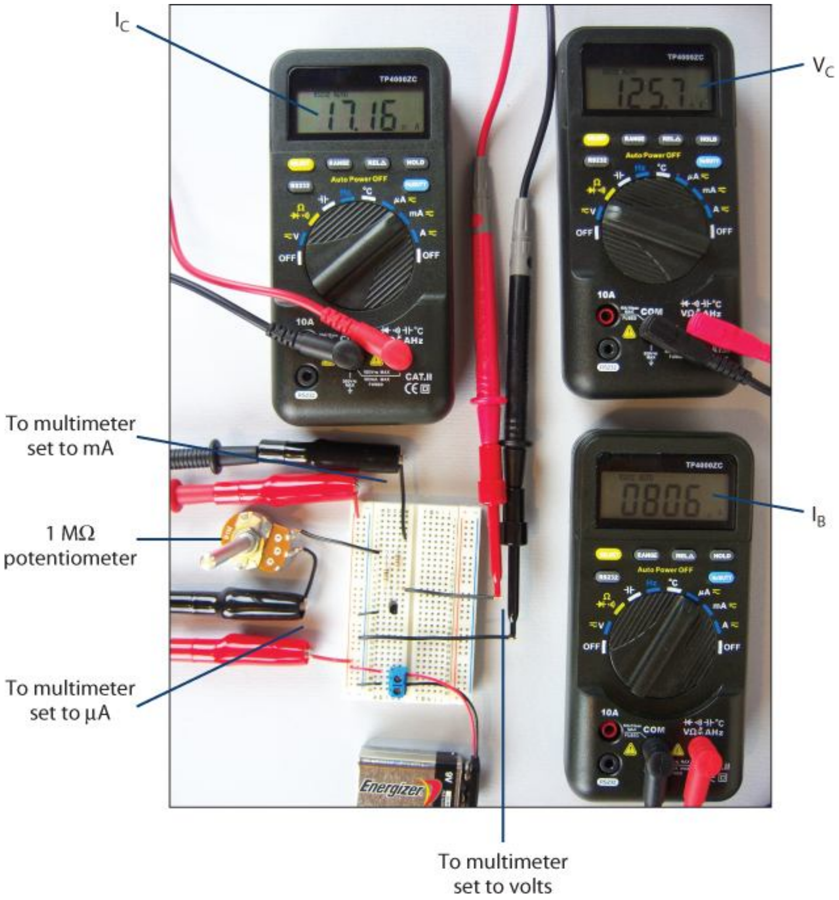
I_C
V_C
17.16
125.7
TP4000ZC
To multimeter set to mA
1 MΩ potentiometer
0806
I_B
To multimeter set to μA
Energizer
To multimeter set to volts

25 Now, compare the circuits shown in Figure 3.26 .

Figure 3.26

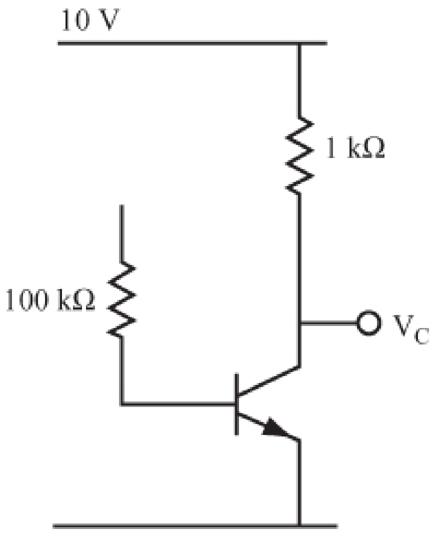

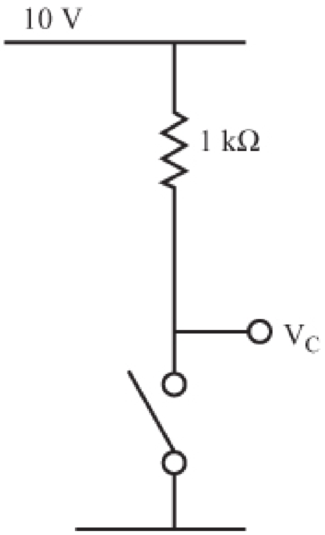

Because the base circuit is broken (that is, it is not complete), there is no base current flowing.

Questions

A. How much collector current is flowing? _____

B. What is the collector voltage? _____

C. What is the voltage at the point V C in the mechanical switch circuit? _____

Answers

A. None.

B. Because there is no current flowing through the 1 kΩ resistor, there is no voltage drop across it, so the collector is at 10 volts.

C. 10 volts because there is no current flowing through the 1 kΩ resistor.

26 From problem 25, it is obvious that a transistor with no collector current is similar to an open mechanical switch. For this reason, a transistor with no collector current and its collector voltage at the supply voltage level is said to be "turned off" (or just OFF).

Question

What are the two main characteristics of an OFF transistor? _____

Answer

It has no collector current, and the collector voltage is equal to the supply voltage.

27 Now, calculate the following parameters for the circuit in Figure 3.27, and compare the results to the examples in problems 25 and 26. Again, the objective here is to find V_C.

Figure 3.27

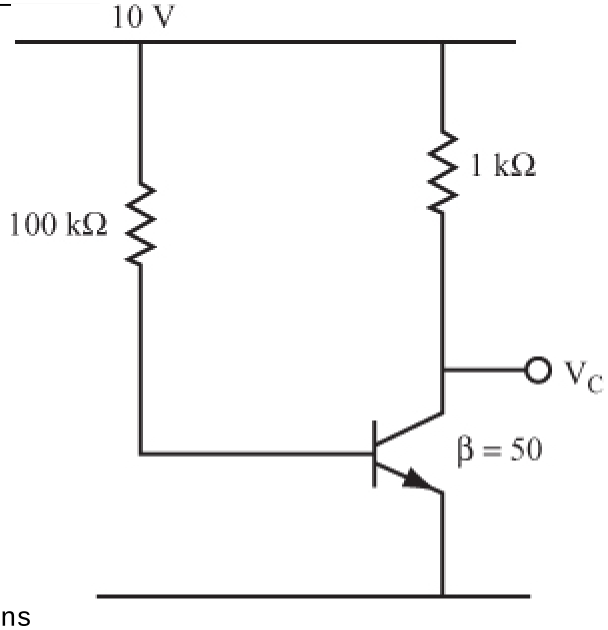

Questions

A. I_B = _____

I_C = _____

B. V_R = _____

C. V_C = _____

Answers

A.
$$I_B = \frac{10\,\text{volts}}{100\,\text{k}\Omega} = 0.1\text{mA}$$
$$I_C = 50 \quad 0.1\text{mA} = 5\text{mA}$$

B. V_R = 1 kΩ × 5 mA = 5 volts

C. V_C = 10 volts−5 volts = 5 volts

Note The output voltage in this problem is half of the supply voltage. This condition is important in AC electronics and is covered in Chapter 8.

The Junction Field Effect Transistor (JFET)

28 Up to now, the only transistor described has been the BJT. Another common transistor type is the JFET. Like the BJT, the JFET is used in many switching and amplification applications. The JFET is preferred when a high input impedance circuit is needed. The BJT has a relatively low input impedance as compared to the JFET. Like the BJT, the JFET is a three-terminal device. The terminals are called the *source* , *drain* , and *gate* . They are similar in function to the emitter, collector, and base, respectively.

Questions

A. How many terminals does a JFET have,

and what are these terminals called? _____

B. Which terminal has a function similar to the base of a BJT? _____

Answers

A. Three, called the source, drain, and gate.

B. The gate has a control function similar to that of the base of a BJT.

29 The basic design of a JFET consists of one type of semiconductor material with a channel made of the opposite type of semiconductor material running through it. If the channel is N material, it is called an N-channel JFET; if it is P material, it is called a P-channel.

Figure 3.28 shows the basic layout of N and P materials, along with their circuit symbols. Voltage on the gate controls the current flow through the drain and source by controlling the effective width of the channel, allowing more or less current to flow. Thus, the voltage on the gate acts to control the drain current, just as the voltage on the base of a BJT acts to control the collector current.

Questions

A. Which JFET would use electrons as the primary charge carrier for the drain current? _____

B. What effect does changing the voltage on the gate have on the operation of the JFET? _____

Figure 3.28

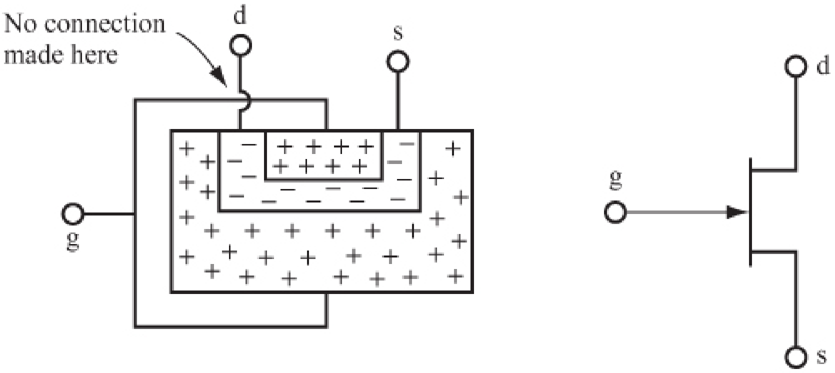
No connection made here
d
s
g
d
g
s

N-channel JFET

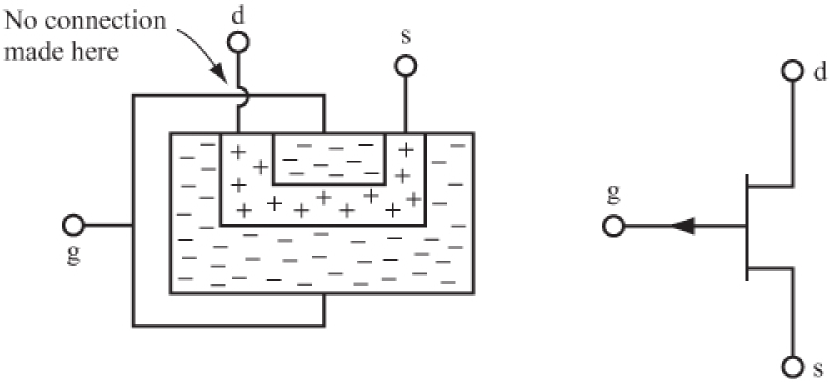
No connection made here
d
s
g
d
g
s

P-channel JFET

Answers

A. N-channel because N material uses electrons as the majority carrier.

B. It changes the current in the drain. The channel width is controlled electrically by the gate potential.

30 To operate the N-channel JFET, apply a positive voltage to the drain with respect to the source. This allows a current to flow through the channel. If the gate is at 0 volts, the drain current is at its largest value for safe operation, and the JFET is in the ON condition.

When a negative voltage is applied to the gate, the drain current is reduced. As the gate voltage becomes more negative, the current lessens until cutoff, which occurs when the JFET is in the OFF condition.

Figure 3.29 shows a typical biasing circuit for the N-channel JFET. For a P-channel JFET, you must reverse the polarity of the bias supplies.

Figure 3.29

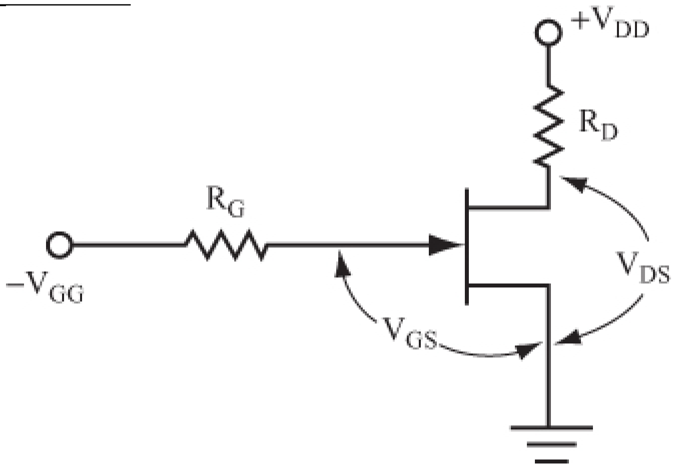

Question

How does the ON-OFF operation of a JFET compare to that of a BJT? _____

Answer

The JFET is ON when there are 0 volts on the gate, whereas you turn the BJT ON by applying a voltage to the base. You turn the

JFET OFF by applying a voltage to the gate, and the BJT is OFF when there are 0 volts on the base. The JFET is a “normally ON” device, but the BJT is considered a “normally OFF” device. Therefore, you can use the JFET (like the BJT) as a switching device.

31 When the gate to source voltage is at 0 volts ($V_{GS} = 0$) for the JFET (refer to Figure 3.29), the drain current is at its maximum (or saturation) value. This means that the N-channel resistance is at its lowest possible value, in the range of 5 to 200 ohms. If R_D is significantly greater than this, the N-channel resistance, r_{DS} , is assumed to be negligible.

Questions

A. What switch condition would this represent, and what is the drain to source voltage (V_{DS})? _____

B. As the gate becomes more negative with respect to the source, the resistance of the N-channel increases until the cutoff point is reached. At this point, the resistance of the channel is assumed to be infinite. What condition does this represent, and what is the drain to source voltage? _____

C. What does the JFET act like when it is operated between the two extremes of current saturation and current cutoff? _____

Answers

A. Closed switch, $V_{DS} = 0$ volts, or low value

B. Open switch, $V_{DS} = V_{DD}$

C. A variable resistance

Inside the JFET

Now take a closer look at the inside of an N-channel JFET. With 0 volts applied to the gate, the channel is at its widest, and the maximum amount of current can flow between the drain and the source. If you apply a negative voltage to the gate, electrons in the channel are repelled from the negative voltage, forming depletion regions on each side of the channel, which narrows the channel, as shown in the following figure.

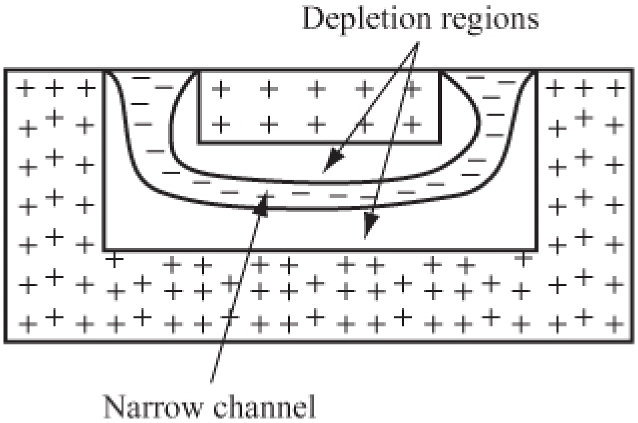

Further increasing the negative voltage on the gate repels additional electrons, increasing the width of the depletion region and decreasing the width of the channel. The narrower the channel, the higher its resistance. When you apply high enough negative voltage, the depletion regions completely block the channel, as shown in the following figure, cutting off the flow of current between the drain and source. This voltage is called the *cutoff voltage* .

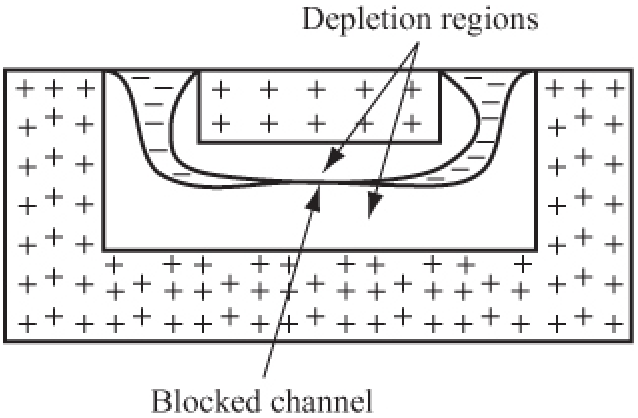

Before you connect any JFET to other components in a circuit, you must identify the drain, gate, and source leads (referred to as the JFET's pinout) and determine whether the component is an N-channel or a P-channel JFET.

Transistors are marked with a part number. For example, 2N3819, 2N5951, and 2N5460 are all part numbers of JFETs. However, the part numbers don't tell you much about the JFET. For these three transistors, the 2N3819 and 2N5951 are N channel JFETs, whereas the 2N5460 transistor is a P channel JFET. This is not obvious from the part numbers.

Also, the JFET pinout is not identified on the part number. For example, one N-channel JFET, the 2N3819, uses different leads for the gate, drain, and source than the 2N5951

N-channel JFET, as shown in the following figure.

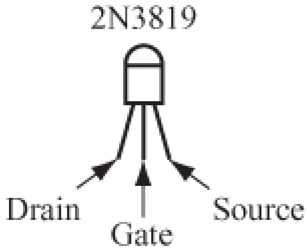

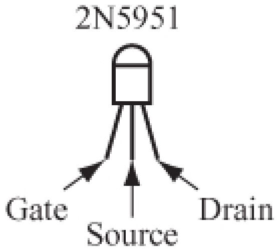

Refer to the manufacturer's data sheet for the transistor pinout and other characteristics. You can easily look up data sheets on the Internet. Also, you can find links to the data sheets for the JFETs used in this book on the website at www.buildinggadgets.com/index_datasheets.htm.

Summary

At this point, it's useful to compare the properties of a mechanical switch with the properties of both types of transistors, as summarized in the following table.

	Switch	BJT	JFET
OFF (or open)			
	No current.	No collector current.	No drain current.
	Full voltage across terminals.	Full supply voltage between collector and emitter.	Full supply voltage between drain and source.
ON (or closed)			
	Full current.	Full circuit current.	Full circuit

current.

No voltage across terminals. Collector to emitter voltage is 0 volts. Drain to source voltage is 0 volts.

The terms ON and OFF are used in digital electronics to describe the two transistor conditions you just encountered. Their similarity to a mechanical switch is useful in many electronic circuits.

In Chapter 4 you learn about the transistor switch in more detail. This is the first step toward an understanding of digital electronics. In Chapter 8 you examine the operation of the transistor when it is biased at a point falling between the two conditions, ON and OFF. In this mode, the transistor can be viewed as a variable resistance and used as an amplifier.

Self-Test

The following questions test your understanding of the concepts presented in this chapter. Use a separate sheet of paper for your drawings or calculations. Compare your answers with the answers provided following the test.

1. Draw the symbols for an NPN and a PNP bipolar transistor. Label the terminals of each.

2. Draw the paths taken by the base and collector currents, as shown in Figure 3.30 .

Figure 3.30

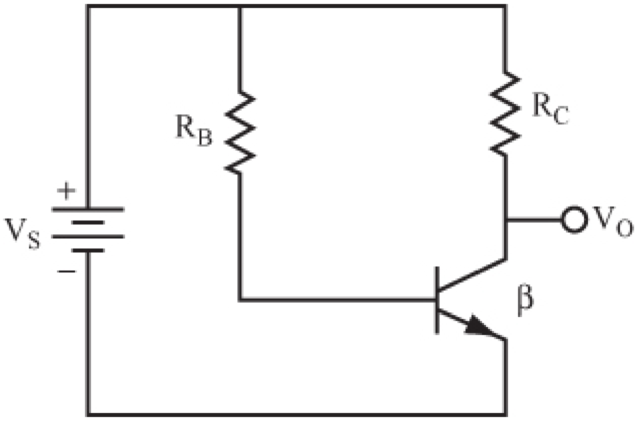

3. What causes the collector current to flow? _____

4. What is meant by the term *current gain* ? What symbol is used for this? What is its algebraic formula? _____

Use the circuit in Figure 3.30 to answer questions 5 through 10.

5. Assume that the transistor is made of silicon. Set R_B = 27 kΩ and V_S = 3 volts. Find I_B. _____

6. If R_B = 220 kΩ and V_S = 10 volts. Find I_B. _____

7. Find V_O when R_B = 100 kΩ, V_S = 10 volts, R_C = 1 kΩ, and β = 50. _____

8. Find V_O when R_B = 200 kΩ, V_S = 10 volts, R_C = 1 kΩ, and β = 50. _____

9. Now use these values to find V_O: R_B = 47 kΩ, V_S = 10 volts, R_C = 500 ohms, and β = 65. _____

10. Use these values to find V_O: R_B = 68 kΩ, V_S = 10 volts, R_C = 820 ohms, and β

= 75. _____

11. Draw the symbols for the two types of JFETs and identify the terminals. _____

12. What controls the flow of current in both a JFET and a BJT? _____

13. In the JFET common source circuit shown in Figure 3.31, add the correct polarities of the power supplies, and draw the current path taken by the drain current.

Figure 3.31

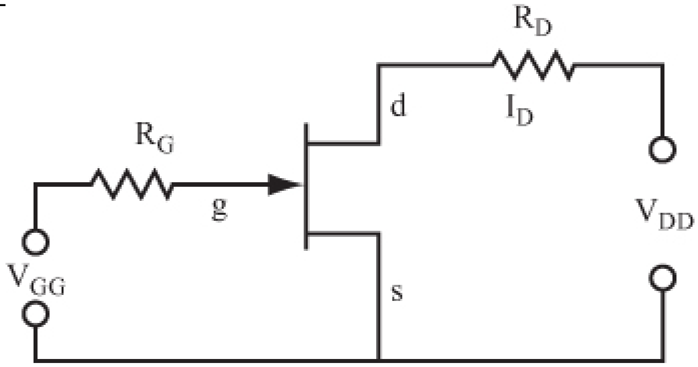

14. When a base current is required to turn a BJT ON, why is there no gate current for the JFET in the ON state. _____

15. Answer the following questions for the circuit shown in Figure 3.32.

Figure 3.32

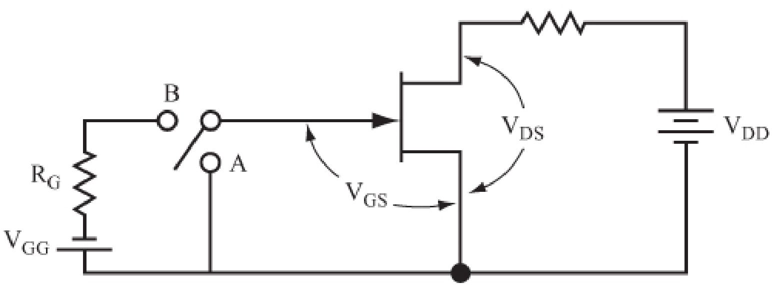

A. If the switch is at position A, what will the drain current be, and why? _____

B. If the switch is at position B, and the gate supply voltage is of sufficient value to cause cutoff, what will the drain current be, and why? _____

C. What is the voltage from the drain to the source for the two switch positions? _____

Answers to Self-Test

If your answers do not agree with those that follow, review the problems indicated in parentheses before you go to Chapter 4.

1. See Figure 3.33.

Figure 3.33

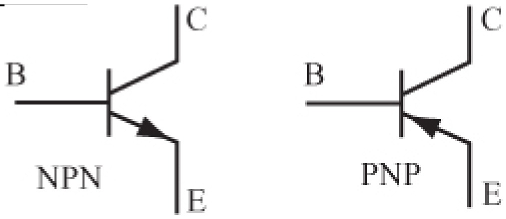

(problems 4 and 5)

2. See Figure 3.34.

Figure 3.34

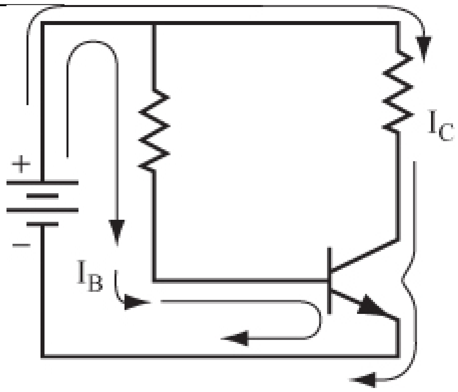

(problems 13 and 15)

3. Base current. (problem 15)

4. Current gain is the ratio of collector current to base current. It is represented by the symbol $\beta.\beta = I_C/I_B$. (problems 16 and 17)

5.

$$I_B = \frac{(V_S \quad 0.7\,\text{volt})}{R_B} = \frac{(3\text{ volts} \quad 0.7\text{ volt})}{27\,\text{k}\Omega} = \frac{2.3\text{ volts}}{27\,\text{k}\Omega} = 85..\text{A}$$

(problem 7)

6. $$I_B = \frac{10\text{ volts}}{220\,\text{k}\Omega} = 45.50..\text{A}$$ (problem 7)

7. 5 volts (problems 20–23)

8. 7.5 volts (problems 20–23)

9. 3.1 volts (problems 20–23)

10. 1 volt (problems 20–23)

11. See Figure 3.35 .

Figure 3.35

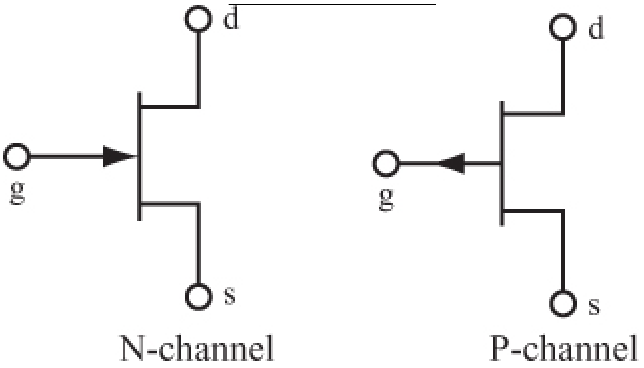

(problem 29)

12. The voltage on the gate controls the flow of drain current, which is similar to the base voltage controlling the collector current in a BJT. (problem 29)

13. See Figure 3.36 .

Figure 3.36

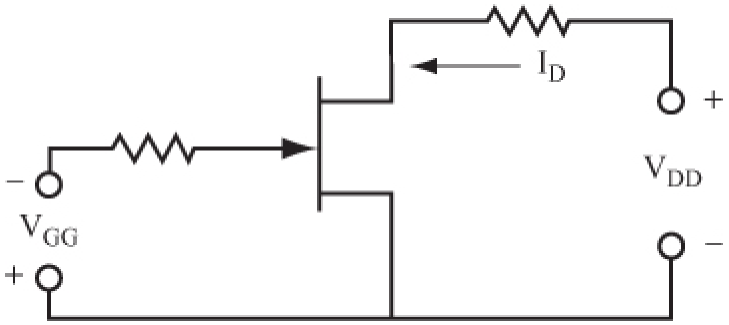

(problem 30)

14. The JFET is a high-impedance device and does not draw current from the gate circuit. The BJT is a relatively low-impedance device and does, therefore, require some base current to operate. (problem 28)

15A. The drain current will be at its maximum value. In this case, it equals V DD /R D because you can ignore the drop across the JFET. The gate to source voltage is 0 volts, which reduces the channel resistance to a small value close to 0 ohms. (problem 31)

15B. The drain current now goes to 0 ampere because the channel resistance is at infinity (very large), which does not allow electrons to flow through the channel.

15C. At position A, V DS is approximately 0 volts. At position B, V DS = V DD .

Chapter 4

The Transistor Switch

Transistors are everywhere. You can't avoid them as you move through your daily tasks. For example, almost all industrial controls, and even your MP3 player, stereo, and television may use transistors as switches.

In Chapter 3, “Introduction to the Transistor,” you saw how a transistor can be turned ON and OFF, similar to a mechanical switch. Computers work with Boolean algebra, which uses only two logic states—TRUE and FALSE. These two states are easily represented electronically by a transistor that is ON or OFF. Therefore, the transistor switch is used extensively in computers. In fact, the logic portions of microprocessors (the brains of computers) consist entirely of transistor switches.

This chapter introduces the transistor's simple and widespread application—switching, with emphasis on the bipolar junction transistor (BJT).

When you complete this chapter, you will be able to do the following:

- Calculate the base resistance, which turns a transistor ON and OFF.
- Explain how one transistor turns another ON and OFF.
- Calculate various currents and resistances in simple transistor switching circuits.

Calculate various resistances and currents in switching circuits, which contain two transistors.

Compare the switching action of a junction field effect transistor (JFET) to a BJT.

Turning the Transistor On

1 Start by examining how to turn a transistor ON by using the simple circuit shown in Figure 4.1 . In Chapter 3, R B was given, and you had to find the value of collector current and voltages. Now, do the reverse. Start with the current through R C, and find the value of R B that turns the transistor ON and permits the collector current to flow.

Figure 4.1

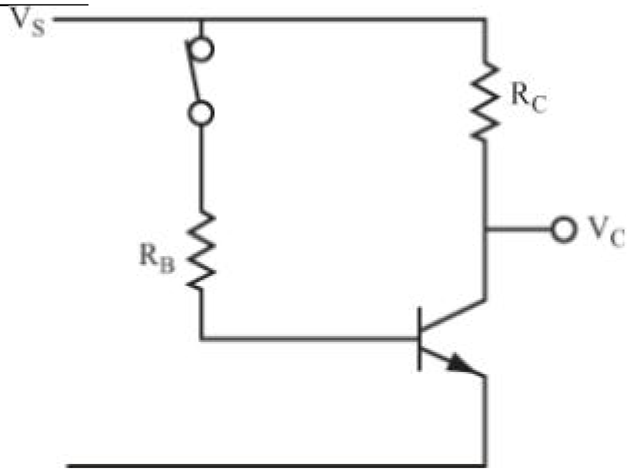

Question

What current values do you need to know to find R B ? _____

Answer

The base and collector currents

2 In this problem circuit, a lamp can be substituted for the collector resistor. In this case, R_C (the resistance of the lamp) is referred to as the *load*, and I_C (the current through the lamp) is called the *load current*.

Questions

A. Is load current equivalent to base or collector current? _____

B. What is the path taken by the collector current discussed in problem 1? Draw this path on the circuit.

Answers

A. Collector current

B. See Figure 4.2. In this figure, note that the resistor symbol has been replaced by the symbol for an incandescent lamp.

Figure 4.2

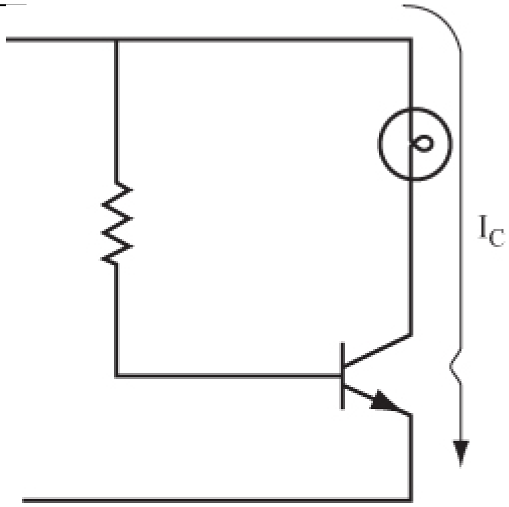

3 For the transistor switch to perform effectively as a CLOSED switch, its collector voltage must be at the same voltage as its emitter voltage. In this condition, the transistor is said to be turned ON.

Questions

A. What is the collector voltage when the transistor is turned ON? _____

B. What other component does an ON transistor resemble? _____

Answers

A. The same as the emitter voltage, which, in this circuit, is 0 volts

B. A closed mechanical switch

Note In actual practice, there is a small voltage drop across the transistor from the collector to the emitter. This is actually a saturation voltage and is the smallest voltage drop that can occur across a transistor when it is ON as “hard” as possible. The discussions in this chapter consider this voltage drop to be a negligible value; therefore, the collector voltage is said to be 0 volts. For a quality switching transistor, this is a safe assumption.

4 The circuit in Figure 4.3 shows a lamp with a resistance of 240 ohms in place of R C .

Figure 4.3

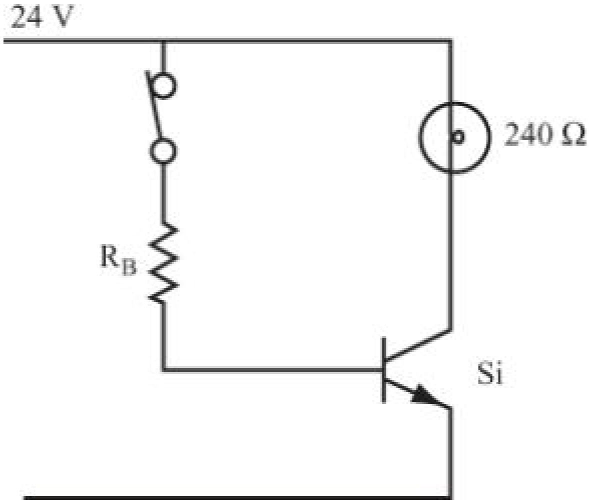

This figure shows the supply voltage and the collector resistance. Given these two values, using Ohm's law, you can calculate the load current (also called the *collector current*) as follows:

$$I_L = I_C = \frac{V_S}{R_C} = \frac{24 \text{ volts}}{240 \text{ ohms}} = 100\,\text{mA}$$

Thus, 100 mA of collector current must flow through the transistor to fully illuminate the lamp. As you learned in Chapter 3 collector current does not flow unless base current flows.

Questions

A. Why do you need base current? _____

B. How can you make base current flow? _____

Answers

A. To enable collector current to flow so that the lamp lights up

B. By closing the mechanical switch in the base circuit

5 You can calculate the amount of base current flowing. Assume that β = 100.

Question

A. What is the value of the base current I B ? _____

Answer

$$I_B = \frac{I_C}{\beta} = \frac{100\,\text{mA}}{100} = 1\,\text{mA}$$

6 The base current flows in the direction shown in Figure 4.4 . Base current flows through the base-emitter junction of the transistor as it does in a forward-biased diode.

Figure 4.4

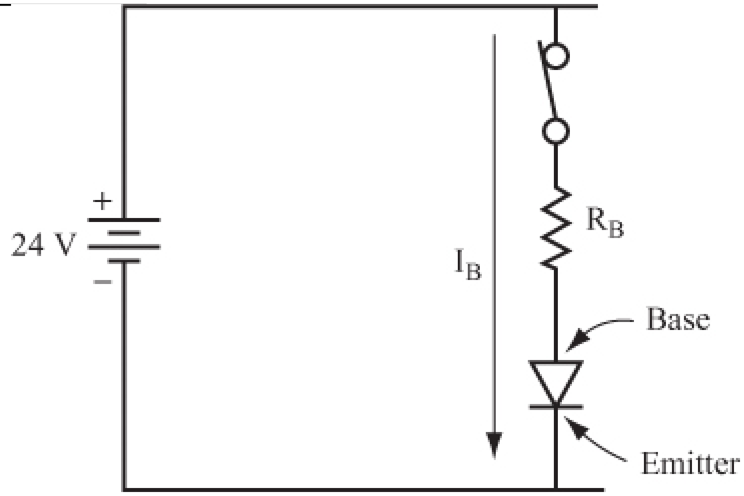

Questions

A. What is the voltage drop across the base-emitter diode? _____

B. What is the voltage drop across R_B? _____

Answers

A. 0.7 volt because it is a silicon transistor

B. 24 volts if the 0.7 is ignored; 23.3 volts if it is not

7 The next step is to calculate R_B. The current flowing through R_B is the base current I_B, and you determined the voltage across it in problem 6.

Question

1. Calculate R_B. _____

Answer

$$R_B = \frac{23.3 \text{ volts}}{1 \text{ mA}} = 23{,}300 \text{ ohms}$$

Figure 4.5 shows the final circuit, including the calculated current and resistance values.

Figure 4.5

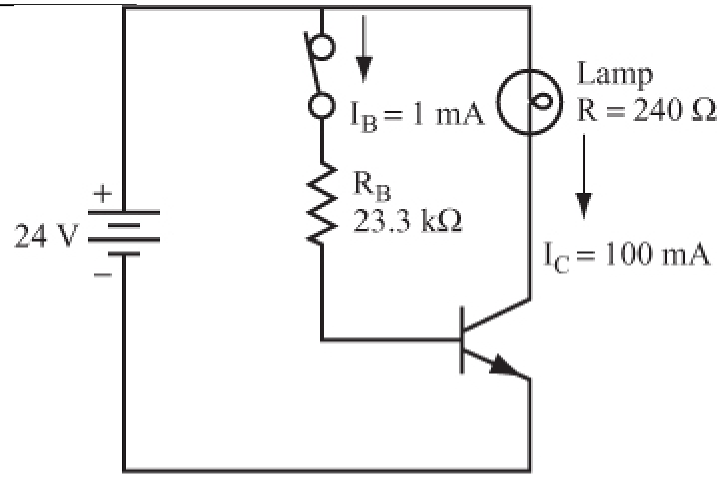

8 Use the following steps to calculate the values of I_B and R_B needed to turn a transistor ON:

1. Determine the required collector current.
2. Determine the value of β.
3. Calculate the required value of I_B from the results of steps 1 and 2.
4. Calculate the required value of R_B.
5. Draw the final circuit.

Now, assume that $V_S = 28$ volts, that you are using a lamp requiring 50 mA of current, and that $\beta = 75$.

Questions

A. Calculate I_B. _____

B. Determine R_B. _____

Answers

A. The collector current and β were given. Thus:

$$I_B = \frac{I_C}{\beta} = \frac{50\,\text{mA}}{75} = 0.667\,\text{mA}$$

B. $$R_B = \frac{28\ \text{volts}}{0.667\,\text{mA}} = 42\text{k}\Omega$$

This calculation ignores V_{BE}.

9 Now, assume that $V_S = 9$ volts, that you are using a lamp requiring 20 mA of current, and that $\beta = 75$.

Question

Calculate R_B. _____

Answer

$$R_B = 31.1\text{k}\Omega$$

In this calculation, V_{BE} is included.

10 In practice, if the supply voltage is much larger than the 0.7-volt drop across the base-emitter junction, you can simplify your calculations by ignoring the 0.7-volt drop, and assume that all the supply voltage appears across the base resistor R_B. (Most resistors are only accurate to within +/- 5 percent of their stated value anyway.) If the supply voltage is less than 10 volts, however, you shouldn't ignore the 0.7-volt drop across the base-emitter junction.

Questions

Calculate R_B for the following problems, ignoring the voltage drop across the base-emitter junction, if appropriate.

A. A 10-volt lamp that draws 10 mA. β = 100. _____

B. A 5-volt lamp that draws 100 mA. β = 50. _____

Answers

A.

$$I_B = \frac{10\,\text{mA}}{100} = 0.1\,\text{mA}$$

$$R_B = \frac{10\ \text{volts}}{0.1\,\text{mA}} = 100\,\text{k}\Omega$$

B.

$$I_B = \frac{100\,\text{mA}}{50} = 2\,\text{mA}$$

$$R_B = \frac{(5\ \text{volts} - 0.7\ \text{volts})}{2\,\text{mA}} = \frac{4.3\ \text{volts}}{2\,\text{mA}} = 2.15\,\text{k}\Omega$$

Turning Off the Transistor

11 Up to now, you have concentrated on turning the transistor ON, thus making it act like a closed mechanical switch. Now you focus on turning it OFF, thus making it act like an open mechanical switch. If the transistor is OFF, no current flows through the load (that is, no collector current flows).

Questions

A. When a switch is open, are the two terminals at different voltages or at the same voltage? _____

B. When a switch is open, does current flow? _____

C. For a transistor to turn OFF and act like an open switch, how much base current is needed? _____

Answers

A. At different voltages, the supply voltage and ground voltage.

B. No.

C. The transistor is OFF when there is no base current.

12 You can be sure that there is no base current in the circuit shown in Figure 4.6 by opening the mechanical switch.

Figure 4.6

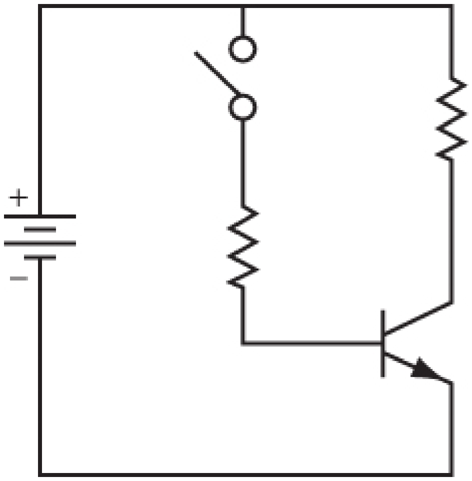

To ensure that the transistor remains OFF when the base is not connected to the supply voltage, you add a resistor (labeled R 2 in Figure 4.7) to the circuit. The base of the transistor connects to ground or 0 volts through this resistor. Therefore, no base current can possibly flow.

Figure 4.7

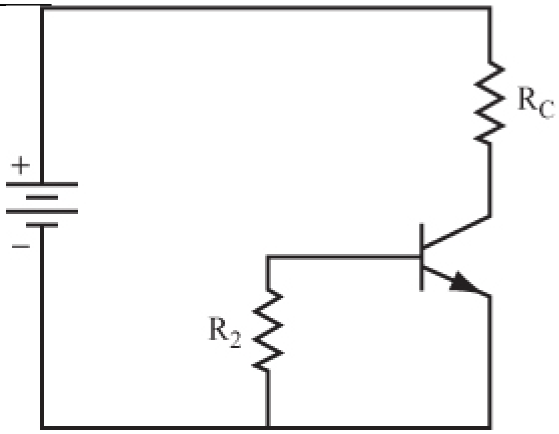

Questions

A. Why doesn't current flow from the supply voltage to the base-emitter junction? _____

B. How much current flows from collector to base? _____

C. Why doesn't current flow from collector to base through R 2 ground? _____

D. Why is the transistor base at 0 volts when R 2 is installed? _____

Answers

A. There is no current path from the supply voltage through the base-emitter junction. Thus, there is no base current flowing.

B. None at all.

C. The internal construction of the transistor prevents this, because the collector-to-base junction is basically a reverse-biased diode.

D. Because there is no current through R 2 , there is no voltage drop across R 2 and, therefore, the transistor base is at ground (0 volts).

13 Because no current is flowing through R 2 , you can use a wide range of resistance values. In practice, the values you find for R 2 are between 1 kΩ and 1 MΩ.

Question

Which of the following resistor values would you use to keep a transistor turned off? 1 ohm, 2 kΩ, 10 kΩ, 20 kΩ, 50 kΩ, 100 kΩ, 250 kΩ, and 500 kΩ. _____

Answer

They would all be suitable except the 1 ohm because the rest are all above 1 kΩ and below 1 MΩ.

14 Figure 4.8 shows a circuit using both R_1 and R_2. Note that the circuit includes a two-position switch that you can use to turn the transistor ON or OFF.

Figure 4.8

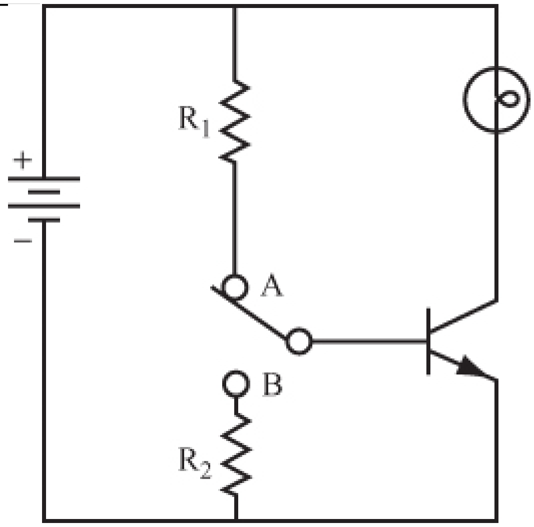

Questions

A. As shown in Figure 4.8, is the transistor ON or OFF? _____

B. Which position, A or B, can cause the collector current to be 0 amperes? _____

Answers

A. ON—the base-to-emitter diode is forward-biased. Therefore, base current can flow.

B. Position B—the base is tied to ground. Therefore, no base current can flow, and the transistor is OFF.

Why Transistors Are Used as Switches

15 You can use the transistor as a switch (as you saw in the previous problems) to perform simple operations such as turning a lamp current on and off. Although often used between a mechanical switch and a lamp, there are other uses for the transistor.

Following are a few other examples that demonstrate the advantages of using a transistor in a circuit as a switch:

Example 1—Suppose you must put a lamp in a dangerous environment, such as a radioactive chamber. Obviously, the switch to operate the lamp must be placed somewhere safe. You can simply use a switch outside the chamber to turn the transistor switch ON or OFF.

Example 2—If a switch controls equipment that requires large amounts of current, then that current must flow through the wires that run between the switch and the lamp. Because the transistor switch can be turned ON or OFF using low voltages and currents, you can connect a mechanical switch to the transistor switch using small, low-voltage wire and, thereby, control the larger current flow. If the mechanical switch is any distance from the equipment you're controlling, using low-voltage wire can save you time and money.

Example 3—A major problem with switching high current in wires is that the current induces interference in adjacent wires. This can be disastrous in communications equipment such as radio transceivers. To avoid this, you can use a transistor to control the larger current from a remote location, reducing the current needed at the switch located in the radio transceiver.

Example 4—In mobile devices (such as a radio-controlled airplane), using transistor switches minimizes the power, weight, and bulk required.

Example 5—When you use a sensor to activate devices, the sensor provides a low current to the transistor, which then acts as a switch controlling the larger current needed to power the equipment. An everyday example is a sensor that detects a light beam across a doorway. When the beam is blocked by a person or object passing through, the sensor stops generating a current, switching a transistor OFF, which activates a buzzer.

Question

What features mentioned in these examples make using transistors as switches desirable?

_____

Answer

The switching action of a transistor can be directly controlled by an electrical signal, as well as by a mechanical switch in the base circuit. This provides a lot of flexibility for the design and allows for simple electrical control. Other factors include safety, reduction of interference, remote switching control, and lower design costs.

16 The following examples of transistor switching demonstrate some other reasons for using transistors:

Example 1 —You can control the ON and OFF times of a transistor accurately, whereas mechanical devices are not accurate. This is important in applications such as photography, where it is necessary to expose a film or illuminate an object for a precise period of time. In these types of uses, transistors are much more accurate and controllable than any other device.

Example 2 —A transistor can be switched ON and OFF millions of times a second and will last for many years. In fact, transistors are one of the longest lasting and most reliable components known, whereas mechanical switches usually fail after a few thousand operations.

Example 3 —The signals generated by most industrial control devices are digital. These control signals can be simply a high or low voltage, which is ideally suited to turning transistor switches ON or OFF.

Example 4 —Modern manufacturing techniques enable the miniaturization of transistors to such a great extent that many of them (even hundreds of millions) can be fabricated into a single silicon chip. Silicon chips on which transistors (and other electronic components) have been fabricated are called *integrated circuits* (ICs). ICs are little, flat, black plastic components built into almost every mass-produced electronic device and are the reason that electronic devices continue to get smaller and lighter.

Question

What other features, besides the ones mentioned in the previous problem, are demonstrated in the examples given here? _____

Answer

Transistors can be accurately controlled, have high-speed operation, are reliable, have a long life, are small, have low power consumption, can be manufactured in large numbers at low cost, and are extremely small.

17 At this point, consider the idea of using one transistor to turn another one ON and OFF, and of using the second transistor to operate a lamp or other load. (This idea is explored in the next section of this chapter.)

If you must switch many high-current loads, then you can use one switch that controls several transistors simultaneously.

Questions

A. With the extra switches added, is the current that flows through the main switch more or less than the current that flows through the load? _____

B. What effect do you think the extra transistor has on the following?

1. Safety _____
2. Convenience to the operator _____
3. Efficiency and smoothness of operation _____

Answers

A. Less current flows through the main switch than through the load.

B.

1. It increases safety and allows the operator to stay isolated from dangerous situations.
2. Switches can be placed conveniently close together on a panel, or in the best place for an operator, rather than the switch position dictating the operator position.
3. One switch can start many things, such as in a master lighting panel in a television studio or theater.

18 This problem reviews your understanding of the concepts presented in problems 15, 16, and 17.

Question

Indicate which of the following are good reasons for using a transistor as a switch:

A. To switch equipment in a dangerous or inaccessible area on and off

B. To switch low currents or voltages

C. To lessen the electrical noise that might be introduced into communication and other circuits

D. To increase the number of control switches

E. To use a faster, more reliable device than a mechanical switch _____

Answer

A, C, and E.

Project 4.1: The Transistor Switch

Objective

The objective of this project is to demonstrate how light can switch a transistor ON or OFF to control a device.

General Instructions

This project uses two breadboarded circuits. The circuit shown on the left side of Figure 4.9 is used to generate infrared light. Another circuit, shown on the right side of Figure 4.9 , switches on a buzzer when the infrared light is blocked by an object.

Figure 4.9

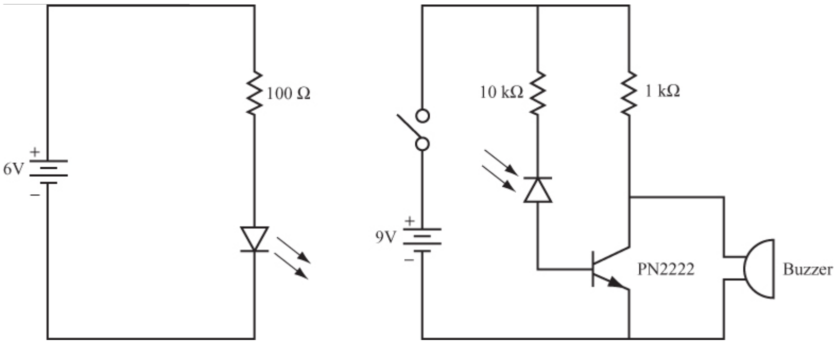

The infrared light in this project is generated by a light-emitting diode (LED). In an LED, a current runs through a PN junction that generates light. This same process occurs with all diodes. Infrared LEDs are simply diodes with a transparent case that enables the infrared light to show through. LEDs also have a PN junction made with semiconductor material that produces a large amount of infrared light. Figure 4.10 shows a typical LED and its schematic symbol, the symbol for a diode with arrows pointing outward, indicating that light is generated.

Figure 4.10

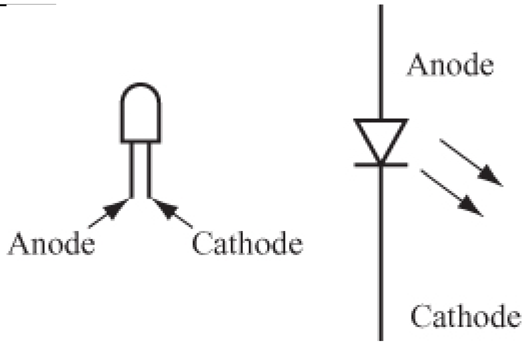

In this project, a photodiode detects the infrared light. When light strikes a PN junction in a photodiode (or any diode), a current is generated. Infrared photodiodes also have a transparent case and junction material that produces a large current when it absorbs infrared light. Figure 4.11 shows a typical photodiode and its schematic symbol consisting of the symbol for a diode with arrows pointing inward, indicating that light is absorbed.

Figure 4.11

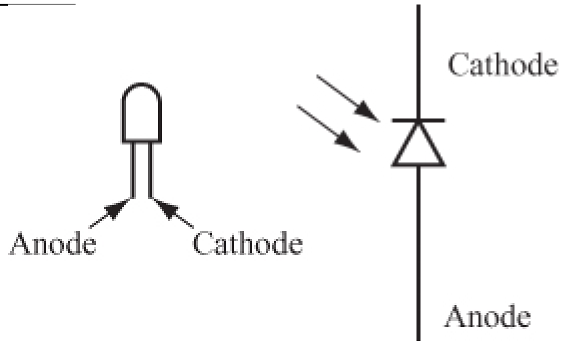

When the circuits are set up, the buzzer sounds whenever the infrared light is blocked from the photodiode.

Parts List

- One 9-volt battery.
- One 6-volt battery pack (4 AA batteries).
- Two battery snap connectors.
- One 100-ohm, 0.5-watt resistor.
- One 1 kΩ, 0.25-watt resistor.
- One 10 kΩ, 0.25-watt resistor.
- Two breadboards.
- Two terminal blocks.
- One piezoelectric buzzer with a minimum operating voltage of 3 volts DC. Using a buzzer with pins (such as part # SE9-2202AS by Shogyo International) enables you to insert the buzzer directly into the breadboard. If you use a buzzer with wire leads (such as part # PK-27N26WQ by Mallory), you need another terminal block.

One infrared LED.

One infrared photodiode.

One PN2222 transistor. Figure 4.12 shows the pinout diagram for the PN2222.

Figure 4.12

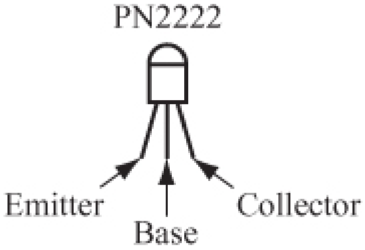

Step-by-Step Instructions

Set up the circuits shown in Figure 4.9 . If you have some experience in building circuits, this schematic (along with the previous parts list) should provide all the information you need to build the circuit. If you need a bit more help building the circuit, look at the photos of the completed circuit in the "Expected Results" section.

Carefully check your circuit against the diagram, especially the connection of the long and short leads to the LED and photodiode. The LED is connected so that it is forward-biased, whereas the photodiode is connected so that it is reverse-biased, as indicated by the direction of the schematic symbols in the circuit diagrams.

1. Align the rounded top of the LED toward the rounded top of the photodiode with the circuit boards a few feet apart from each other. (If you use a typical LED and

photodiode, you must bend their leads to align them.) Note that the rounded top of both the LED and photodiode shown in Figures 4.10 and 4.11 contain a lens to emit or absorb light. Some LEDs and photodiodes have lenses on the side, instead of on the top. If it isn't obvious where the lens is in your components, check the manufacturer's data sheet.

2. Turn on the power switch. When the power switch is on, the buzzer should sound whenever the photodiode does not sense infrared light.
3. Bring the circuits close enough together so that the buzzer shuts off.
4. Block the infrared light; the buzzer should turn on.

Expected Results

Figure 4.13 shows the breadboarded buzzer circuit for this project.

Figure 4.13

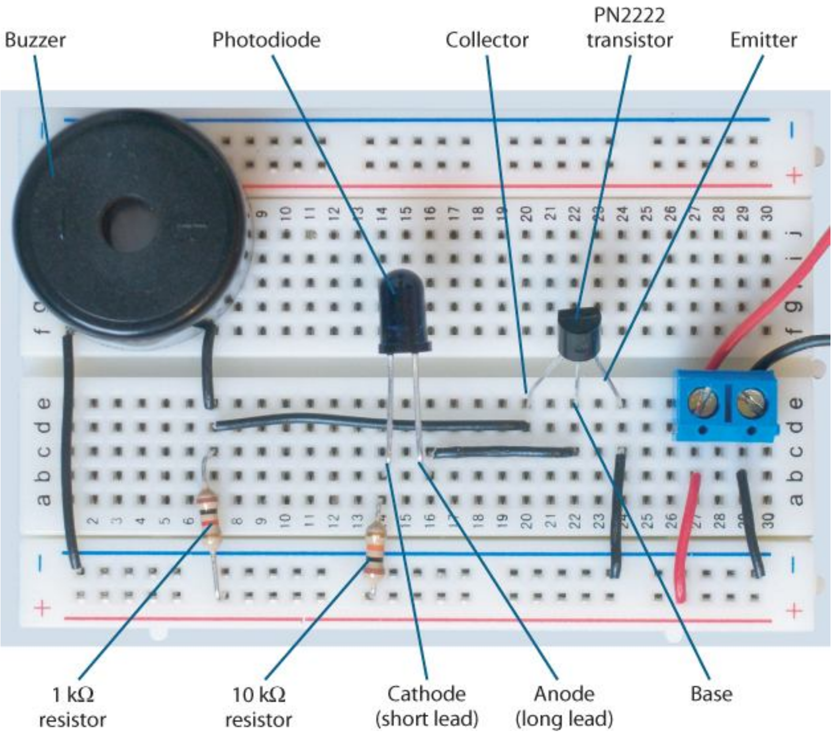
Buzzer
Photodiode
Collector
PN2222 transistor
Emitter
1 kΩ resistor
10 kΩ resistor
Cathode (short lead)
Anode (long lead)
Base

Figure 4.14 shows the breadboarded LED circuit for this project.

Figure 4.14

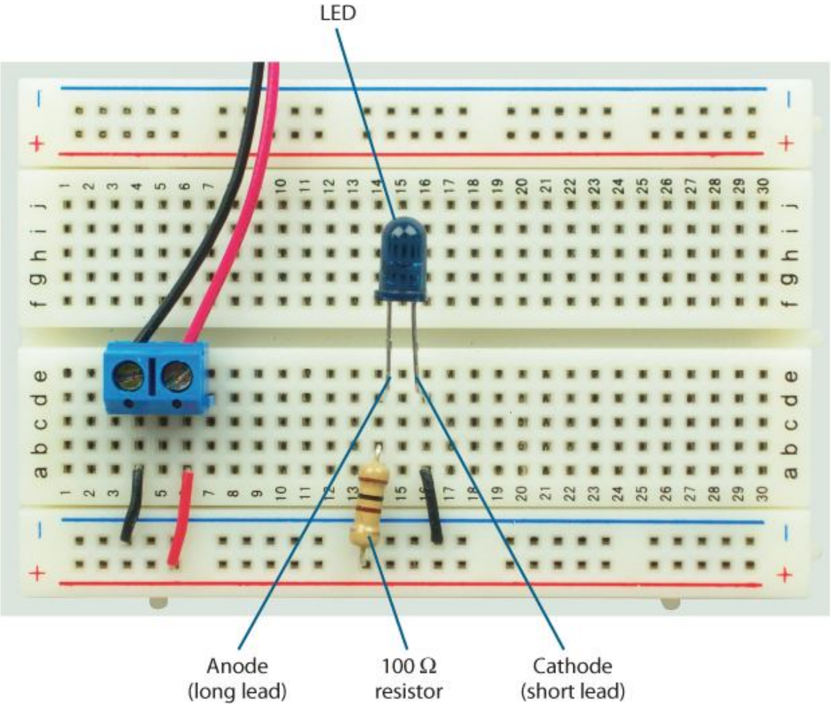
LED
Anode
(long lead)
100 Ω
resistor
Cathode
(short lead)

Figure 4.15 shows the test setup for this project with the rounded top of the LED and photodiode aligned toward each other.

Figure 4.15

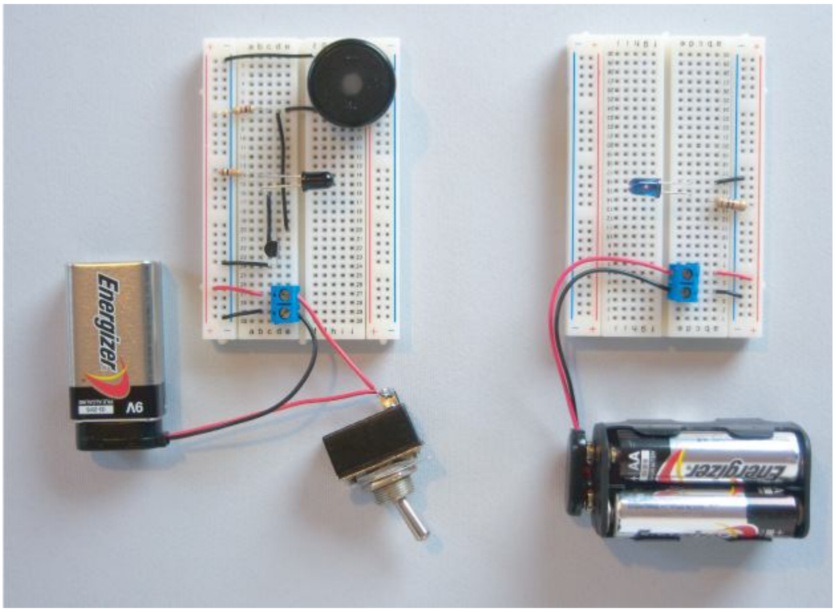

The photodiode is connected to the base of a transistor. Therefore, current generated by the photodiode turns the transistor ON. When the transistor is ON, V_C is about 0 volts, turning off the buzzer. When the infrared light is blocked, the photodiode stops generating current, which turns OFF the transistor, increasing V_C, which turns on the buzzer.

These circuits work with the LED and photodiode about 7 inches apart. With more complicated photo detectors that have circuitry to amplify the detected signal, this technique can work over several feet. One common application of this technique is a buzzer that sounds when a shopper enters a store, blocking the light, setting off a sound, and alerting the sales staff.

19 Many types of electronic circuits contain multiple switching transistors. In this type of circuit, one transistor is used to switch others ON and OFF. To illustrate how this works, again consider the lamp as the load and the mechanical switch as the actuating element. Figure 4.16 shows a circuit that uses two transistors to turn a lamp on or off.

Figure 4.16

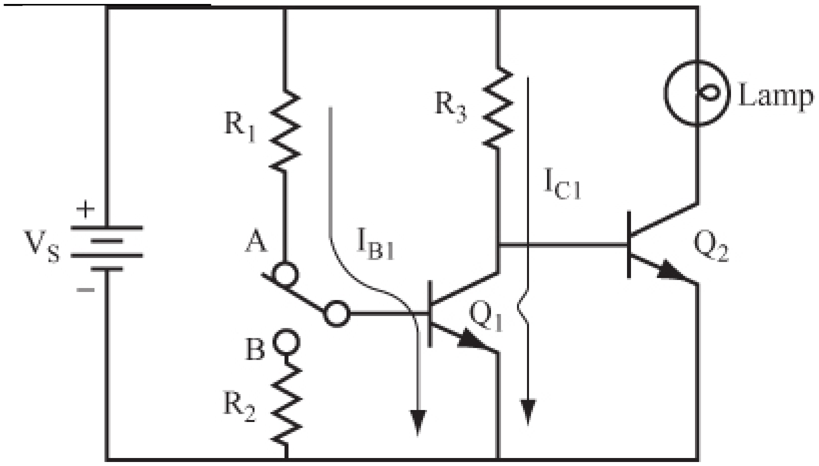

When the switch is in position A, the base-emitter junction of Q 1 is forward-biased. Therefore, base current (I B1) flows through R 1 and through the base-emitter diode of Q 1 , turning the transistor ON. This causes the collector current (I C1) to flow through Q 1 to ground, and the collector voltage drops to 0 volts, just as if Q 1 were a closed switch. Because the base of Q 2 is connected to the collector of Q 1 , the voltage on the base of Q 2 also drops to 0 volts. This ensures that

Q 2 is turned OFF and the lamp remains unlit.

Now, flip the switch to position B, as shown in Figure 4.17 . The base of Q 1 is tied to ground, or 0 volts, turning Q 1 OFF. Therefore, no collector current can flow through Q 1 . A positive voltage is applied to the base of Q 2, and the emitter-base junction of Q 2 is forward-biased. This enables current to flow through R 3 and the emitter-base junction of Q 2 , which turns Q 2 ON, allowing collector current (I C2) to flow, and the lamp is illuminated.

Figure 4.17

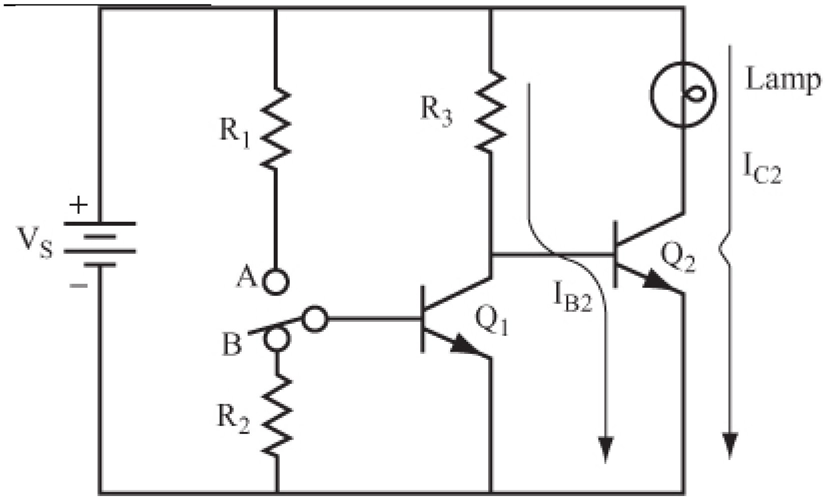

Now that you have read the descriptions of how the circuit works, answer the following questions. First assume that the switch is in position A, as shown in Figure 4.16 .

Questions

A. What effect does I B1 have on transistor Q 1 ? _____

B. What effect does turning Q 1 ON have on the following?

1. Collector current I C1 _____
2. Collector voltage V C1 _____

C. What effect does the change to V C1 covered in the previous question have on the following?

1. The base voltage of Q 2 _____
2. Transistor Q 2 (that is, is it ON or OFF) _____

D. Where does the current through R 3 go? _____

E. In this circuit is the lamp on or off? _____

Answers

A. I B1 , along with a portion of V S (0.7 volts if the transistor is silicon), turns Q 1 ON.

B. (1) I C1 flows; (2) V C1 drops to 0 volts.

C. (1) base of Q 2 drops to 0 volts; (2) Q 2 is OFF.

D. I C1 flows through Q 1 to ground.

E. Off.

20 Now, assume that the switch is in the B position, as shown in Figure 4.17 , and answer these questions.

Questions

A. How much base current I B1 flows into Q 1 ? _____

B. Is Q 1 ON or OFF? _____

C. What current flows through R 3 ? _____

D. Is Q 2 ON or OFF? _____

E. Is the lamp on or off? _____

Answers

A. None

B. OFF

C. I B2

D. ON

E. On

21 Refer to the circuit in Figures 4.16 and 4.17 . Now, answer these questions assuming the supply voltage is 10 volts.

Questions

A. Is the current through R 3 ever divided between Q 1 and Q 2 ? Explain. _____

B. What is the collector voltage of Q 2 with the switch in each position? _____

C. What is the collector voltage of Q 1 with the switch in each position? _____

Answers

A. No. If Q 1 is ON, all the current flows through it to ground as collector current. If Q 1 is OFF, all the current flows through the base of Q 2 as base current.

B. In position A, 10 volts because it is OFF.

In position B, 0 volts because it is ON.

C. In position A, 0 volts because it is ON.

In position B, the collector voltage of Q 1 equals the voltage drop across the forward-biased base-emitter junction of Q 2 , because the base of Q 2 is in parallel with the collector of Q 1 . The voltage drop across the forward-biased base-emitter junction does not rise to 10 volts, but can rise only to 0.7 volt if Q_2 is made of silicon.

22 Now, calculate the values of R_1, R_2, and R_3 for this circuit. The process is similar to the one you used before, but you must expand it to deal with the second transistor. This is similar to the steps you used in problem 8. Follow these steps to calculate R_1, R_2, and R_3:

1. Determine the load current I_{C2}.
2. Determine β for Q_2. Call this β_2.
3. Calculate I_{B2} for Q_2. Use $I_{B2} = I_{C2}/\beta_2$.
4. Calculate R_3 to provide this base current. Use $R_3 = V_S/I_{B2}$.
5. R_3 is also the load for Q_1 when Q_1 is ON. Therefore, the collector current for Q_1 (I_{C1}) has the same value as the base current for Q_2, as calculated in step 3.
6. Determine β_1, the β for Q_1.
7. Calculate the base current for Q_1. Use $I_{B1} = I_{C1}/\beta_1$.
8. Find R_1. Use $R_1 = V_s/I_{B1}$.
9. Choose R_2. For convenience, let $R_2 = R_1$.

Continue to work with the same circuit shown in Figure 4.18. Use the following values:

Figure 4.18

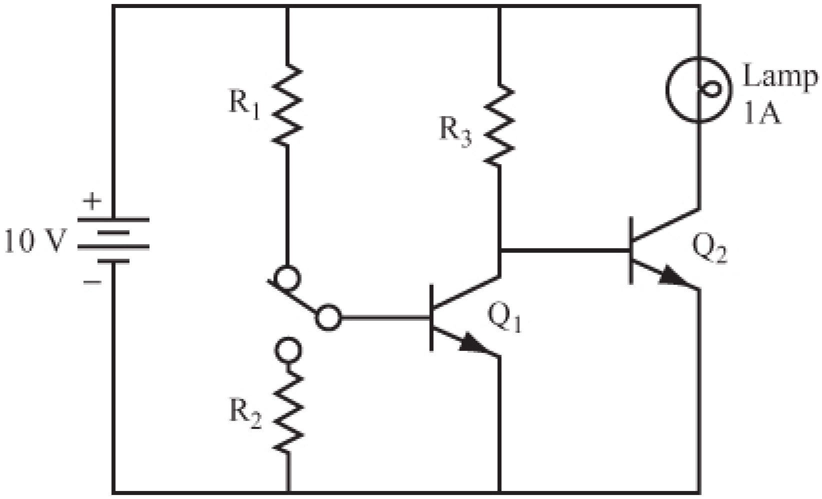

A 10-volt lamp that draws 1 ampere; therefore $V_S = 10$ volts, $I_{C2} = 1$ A.

$\beta_2 = 20$, $\beta_1 = 100$

Ignore any voltage drops across the transistors.

Questions

Calculate the following:

A. Find I_{B2} as in step 3.

I_{B2} = _____

B. Find R_3 as in step 4.

R_3 = _____

C. Calculate the load current for Q_1 when it is ON, as shown in Step 5.

I_{C1} = _____

D. Find the base current for Q_1.

I_{B1} = _____

E. Find R_1 as in step 8.

R_1 = _____

F. Choose a suitable value for R_2. R_2 = _____

Answers

The following answers correspond to the steps.

A.

1. I_{C2} is given as 1 ampere.

2. $\beta_2 = 20$ (given). This is a typical value for a transistor that would handle 1 ampere.

3. $$I_{B2} = \frac{1\text{ ampere}}{20} = 50\,\text{mA}$$

B.

4. $$R_3 = \frac{10\text{ volts}}{50\text{ mA}} = 200\Omega$$

Note that the 0.7 volt base-emitter drop has been ignored.

C.

5. $I_{C1} = I_{B2} = 50$ mA

D.

6. $\beta_1 = 100$

7. $$I_{B1} = \frac{50\,\text{mA}}{100} = 0.5\,\text{mA}$$

E.

8. $$R_1 = \frac{10\text{ volts}}{0.5\text{ mA}} = 20\text{k}\Omega$$

Again, the 0.7-volt drop is ignored.

F.

9. For convenience, choose a value for R_2 that is the same as R_1, or 20 kΩ. This reduces the number of different components in the circuit. The fewer different components you have in a circuit, the less components

you must keep in your parts bin. You could, of course, choose any value between 1 kΩ and 1 MΩ.

23 Following the same procedure, and using the same circuit shown in Figure 4.18, work through this example. Assume that you are using a 28-volt lamp that draws 560 mA, and that $\beta_2 = 10$ and $\beta_1 = 100$.

Questions

Calculate the following:

A. I_{B2} = _____
B. R_3 = _____
C. I_{C1} = _____
D. I_{B1} = _____
E. R_1 = _____
F. R_2 = _____

Answers

A. 56 mA
B. 500 ohms
C. 56 mA
D. 0.56 mA
E. 50 kΩ
F. 50 kΩ by choice

The Three-Transistor Switch

24 The circuit shown in Figure 4.19 uses three transistors to switch a load on and off. In this circuit, Q_1 is used to turn Q_2 ON and OFF, and Q_2 is used to turn Q_3 ON and OFF. The calculations are similar to those you performed in the last few problems, but a few additional steps are required to deal with the third transistor. Use the circuit diagram in Figure 4.19 to determine the answers to the following questions.

Figure 4.19

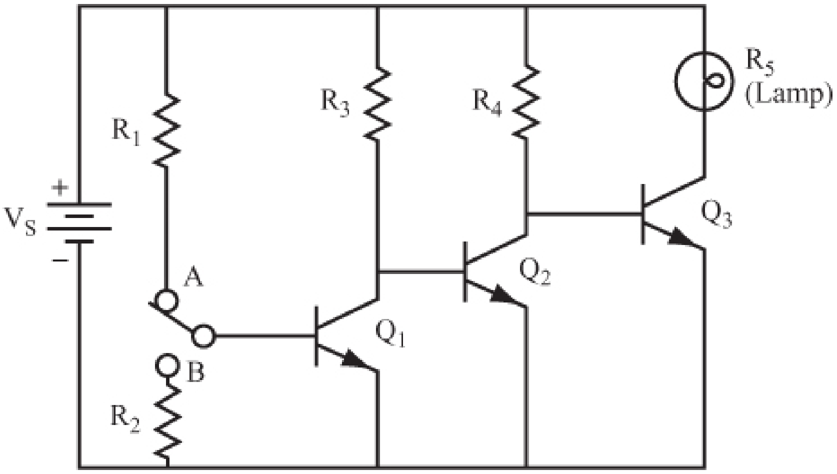

Questions

Assume that the switch is in position A.

A. Is Q 1 ON or OFF? _____

B. Is Q 2 ON or OFF? _____

C. Where is current through R 4 flowing? _____

D. Is Q 3 ON or OFF? _____

Answers

A. ON

B. OFF

C. Into the base of Q 3

D. ON

25 Now use the same circuit as in problem 24.

Questions

Assume that the switch is in position B.

A. Is Q 1 ON or OFF? _____

B. Is Q 2 ON or OFF? _____

C. Where is the current through R 4 flowing? _____

D. Is Q 3 ON or OFF? _____

E. Which switch position turns on the lamp? _____

F. How do the ON/OFF positions for the switch in the three-transistor switch differ from the ON/OFF positions for the switch in the two-transistor switch circuit? _____

Answers

A. OFF.

B. ON.

C. Through Q 2 to ground.

D. OFF.

E. Position A.

F. The positions are opposite. Therefore, if a circuit controls lamps with two transistors and another circuit controls lamps with three transistors, flipping the switch that controls both circuits would change which lamps (or which other loads) are on.

26 Work through this example using the same equations you used for the two-transistor switch in problem 22. The steps are similar but with a few added steps, as shown here:

1. Find the load current. This is often given.

2. Determine the current gain of Q 3 . This is

β_3 and usually it is a given value.

3. Calculate I_{B3}. Use $I_{B3} = I_{C3}/\beta_3$.
4. Calculate R_4. Use $R_4 = V_S/I_{B3}$.
5. Assume $I_{C2} = I_{B3}$.
6. Find β_2. Again this is a given value.
7. Calculate I_{B2}. Use $I_{B2} = I_{C2}/\beta_2$.
8. Calculate R_3. Use $R_3 = V_S/I_{B2}$.
9. Assume $I_{C1} = I_{B2}$.
10. Find β_1.
11. Calculate I_{B1}. Use $I_{B1} = I_{C1}/\beta_1$.
12. Calculate R_1. Use $R_1 = V_s/I_{B1}$.
13. Choose R_2.

For this example, use a 10-volt lamp that draws 10 amperes. Assume that the βs of the transistors are given in the manufacturer's data sheets as $\beta_1 = 100$, $\beta_2 = 50$, and $\beta_3 = 20$. Now, work through the steps, checking the answers for each step as you complete it.

Questions

Calculate the following:

A. I_{B3} = _____

B. R_4 = _____

C. I_{B2} = _____

D. R_3 = _____

E. I_{B1} = _____

F. R_1 = _____

G. R_2 = _____

Answers

The answers here correspond to the steps.

A.

1. The load current is given as 10 amperes.
2. β_3 is given as 20.
3.

$$I_{B3} = \frac{I_{C3}}{\beta_3} = \frac{10\ \text{amperes}}{20} = 0.5\ \text{ampere} = 500\,\text{mA}$$

B.

4. $$R_4 = \frac{10\,\text{volts}}{500\,\text{mA}} = 20\,\text{ohms}$$

C.

5. $I_{C2} = I_{B3} = 500\,\text{mA}$
6. β_2 is given as 50.
7. $$I_{B2} = \frac{I_{C2}}{\beta_2} = \frac{500\,\text{mA}}{50} = 10\,\text{mA}$$

D.

8. $$R_3 = \frac{10\ \text{volts}}{10\,\text{mA}} = 1\text{k}\Omega$$

E.

9. $I_{C1} = I_{B2} = 10\,\text{mA}$
10. β_1 is given as 100.
11. $$I_{B1} = \frac{I_{C1}}{\beta_1} = \frac{10\,\text{mA}}{100} = 0.1\,\text{mA}$$

F.

12. $$R_1 = \frac{10\ \text{volts}}{0.1\,\text{mA}} = 100\text{k}\Omega$$

G.

13. R_2 can be chosen to be 100 kΩ also.

27 Determine the values in the same circuit for a 75-volt lamp that draws 6 amperes. Assume that $\beta_3 = 30$, $\beta_2 = 100$, and $\beta_1 = 120$.

Questions

Calculate the following values using the steps in problem 26:

A. I_{B3} = _____
B. R_4 = _____
C. I_{B2} = _____
D. R_3 = _____
E. I_{B1} = _____
F. R_1 = _____
G. R_2 = _____

Answers

A. 200 mA
B. 375 Ω
C. 2 mA
D. 37.5 kΩ
E. 16.7 μA
F. 4.5 MΩ
G. Choose R_2 = 1 MΩ

Alternative Base Switching

28 In the examples of transistor switching, the actual switching was performed using a small mechanical switch placed in the base circuit of the first transistor. This switch has three terminals and switches from position A to position B. (This is a single-pole, double-throw switch.) This switch does not have a definite ON or OFF position, as does a simple ON-OFF switch.

Question

Why couldn't a simple ON-OFF switch with only two terminals have been used with these examples? _____

Answer

An ON-OFF switch is either open or closed, and cannot switch between position A and position B, as shown earlier in Figure 4.19 .

29 If you connect R 1 , R 2 , and a switch together, as shown in Figure 4.20 , you can use a simple ON-OFF switch with only two terminals. (This is a single-pole, single-throw switch.)

Figure 4.20

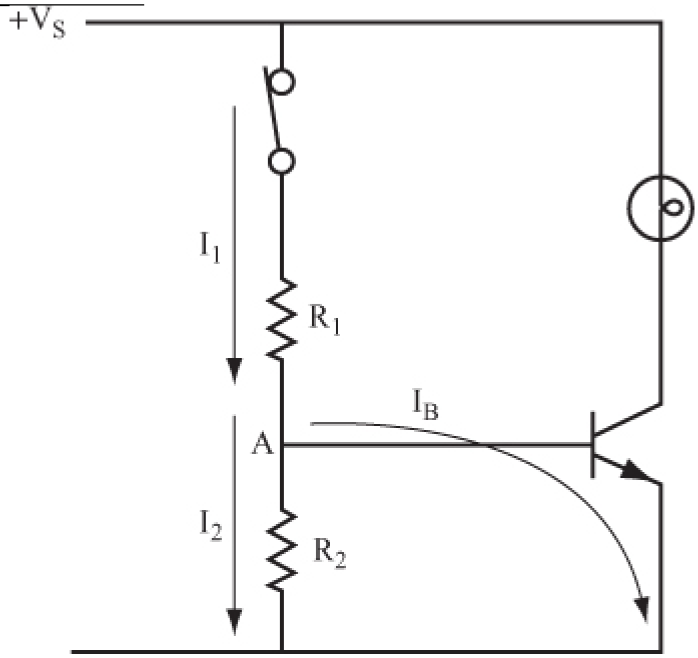

Questions

A. When the switch is open, is Q 1 ON or OFF? _____

B. When the switch is closed, is the lamp ON or Off? _____

Answers

A. OFF

B. ON

30 When the switch is closed, current flows through R_1. However, at point A in Figure 4.20, the current divides into two paths. One path is the base current I_B, and the other is marked I_2.

Question

How could you calculate the total current I_1? _____

Answer

$$I_1 = I_B + I_2$$

31 The problem now is to choose the values of both R_1 and R_2 so that when the current divides, there is sufficient base current to turn Q_1 ON.

Question

Consider this simple example. Assume the load is a 10-volt lamp that needs 100 mA of current and $\beta = 100$. Calculate the base current required.

1. $I_B =$ _____

Answer

$$I_B = \frac{100\,\text{mA}}{100} = 1\,\text{mA}$$

32 After the current I_1 flows through R_1, it must divide, and 1 mA of it becomes I_B. The remainder of the current is I_2. The difficulty at this point is that there is no unique value for either I_1 or I_2. In other words, you could assign them almost any value. The only restriction is that both must permit 1 mA of current to flow into the base of Q_1.

You must make an arbitrary choice for these two values. Based on practical experience, it is common to set I_2 to be 10 times greater than I_B. This split makes the circuits work reliably and keeps the calculations easy:

$$I_2 = 10I_B$$
$$I_1 = 11I_B$$

Question

In problem 31 you determined that I_B = 1 mA. What is the value of I_2? _____

Answer

I_2 = 10 mA

33 Now you can calculate the value of R_2. The voltage across R_2 is the same as the voltage drop across the base-emitter junction of Q_1. Assume that the circuit uses a silicon transistor, so this voltage is 0.7 volt.

Questions

A. What is the value of R_2? _____

B. What is the value R_1? _____

Answers

A. $$R_2 = \frac{0.7\ \text{volt}}{10\ \text{mA}} = 70\ \text{ohms}$$

B.

$$R_2 = \frac{(10\text{ volts} - 0.7\text{ volt})}{11\text{mA}} = \frac{9.3\text{ volts}}{11\text{mA}} = 800\text{ohms (approximately)}$$

You can ignore the 0.7 volt in this case, which would give $R_1 = 910$ ohms.

34 The resistor values you calculated in problem 33 ensure that the transistor turns ON and that the 100 mA current (I_C) you need to illuminate the lamp flows through the lamp and the transistor. Figure 4.21 shows the labeled circuit.

Figure 4.21

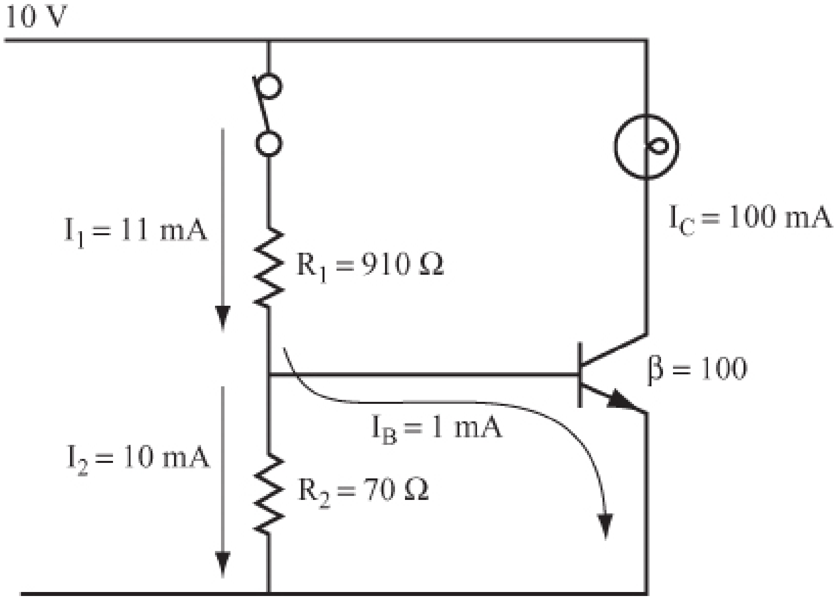

Questions

For each of the following lamps, perform the same calculations you used in the last few problems to find the values of R_1 and R_2.

A. A 28-volt lamp that draws 56 mA. β = 100 _____

B. A 12-volt lamp that draws 140 mA. β =

Answers

A.

$$I_B = \frac{56\,\text{mA}}{100} = 0.56\,\text{mA}$$

$$I_2 = 5.6\,\text{mA}$$

$$R_2 = \frac{0.7\,\text{volt}}{5.6\,\text{mA}} = 125\,\text{ohms}$$

$$R_1 = \frac{28\,\text{volts}}{6.16\,\text{mA}} = 4.5\text{k}\Omega$$

B.

$$R_2 = 25\,\text{ohms}$$

$$R_1 = 400\,\text{ohms}$$

35 The arbitrary decision to make the value of I_2 10 times the value of I_B is obviously subject to considerable discussion, doubt, and disagreement. *Transistors are not exact devices; they are not carbon copies of each other.*

In general, any transistor of the same type has a different β from any other because of the variance in tolerances found in component manufacturing. This leads to a degree of inexactness in designing and analyzing transistor circuits. The truth is that if you follow exact mathematical procedures, it can complicate your life. In practice, a few “rules of thumb” have been developed to help you make the necessary assumptions. These rules lead to simple equations that provide workable values for components that you can use in designing circuits.

The choice of $I_2 = 10I_B$ is one such rule of thumb. Is it the only choice that works? Of course not. Almost any value of I_2 that is at least 5 times larger than I_B can work. Choosing 10 times the value is a good option for three reasons:

It is a good practical choice. It always works.

It makes the arithmetic easy.

It's not overly complicated and doesn't involve unnecessary calculations.

Question

In the example from problem 32, $I_B = 1$ mA and $I_2 = 10$ mA. Which of the following values can also work efficiently for I_2 ?

A. 5 mA

B. 8 mA

C. 175 mA

D. 6.738 mA

E. 1 mA

Answers

Choices A, B, and D. Value C is too high to be a sensible choice, and E is too low.

36 Before you continue with this chapter, answer the following review questions.

Questions

A. Which switches faster, the transistor or the mechanical switch? _____

B. Which can be more accurately controlled? _____

C. Which is the easiest to operate remotely? _____

D. Which is the most reliable? _____

E. Which has the longest life? _____

Answers

A. The transistor is much faster.

B. The transistor.

C. The transistor.

D. The transistor.

E. Because transistors have no moving parts, they have a much longer operating lifetime than a mechanical switch. A mechanical switch will fail after several thousand operations, whereas transistors can be operated several million times a second and can last for years.

Switching the JFET

37 The use of the junction field effect transistor (JFET) as a switch is discussed in the next few problems. You may want to review problems 28 through 31 in Chapter 3 where this book introduced the JFET.

The JFET is considered a “normally on” device, which means that with 0 volts applied to the input terminal (called the *gate*), it is ON, and current can flow through the transistor. When you apply a voltage to the gate, the device conducts less current because the resistance of the drain to the source channel increases. At some point, as the voltage increases, the value of the resistance in the channel becomes so high that the device "cuts off" the flow of current.

Questions

A. What are the three terminals for a JFET called, and which one controls the operation of the device? _____

B. What turns the JFET ON and OFF? _____

Answers

A. Drain, source, and gate, with the gate acting as the control.

B. When the gate voltage is zero (at the same potential as the source), the JFET is ON. When the gate to source voltage difference is high, the JFET is OFF.

Project 4.2: The JFET

Objective

The objective of this project is to determine the drain current that flows when a JFET is fully ON, and the gate voltage needed to fully shut the JFET OFF, using the circuit shown in Figure 4.22 .

Figure 4.22

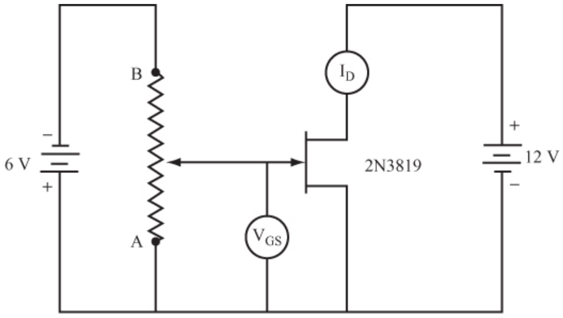

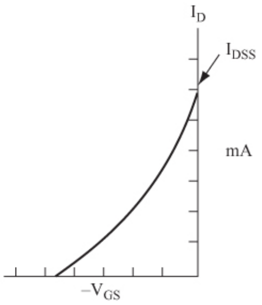

General Instructions

After the circuit is set up, change the gate voltage (V_{GS}) by adjusting the potentiometer. Measure the drain current (I_D) for each V_{GS} value. As you work through the project, observe how the drain current drops toward zero as you increase V_{GS}. When the JFET is OFF, I_D is at zero; when the JFET is fully ON, I_D is at its maximum (called I_{DSS}).

Parts List

You need the following equipment and supplies:

- One 6-volt battery pack (4 AA batteries)
- One 12-volt battery pack (8 AA batteries)
- One multimeter set to mA
- One multimeter set to measure DC voltage
- One 10 kΩ potentiometer
- One breadboard
- Two terminal blocks
- One 2N3819 JFET (Figure 4.23 shows the pinout for the 2N3819.)

Figure 4.23

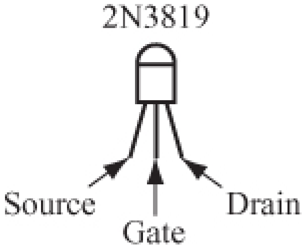

Step-by-Step Instructions

Set up the circuit shown in Figure 4.22 . If you have some experience in building circuits, this schematic (along with the previous parts list) should provide all the information you need to build the circuit. If you need a bit more help building the circuit, look at the photos of the completed circuit in the "Expected Results" section.

Carefully check your circuit against the diagram, especially the orientation of the JFET to ensure that the drain, gate, and source leads are connected correctly. One unusual aspect of this circuit you may want to check is that the +V bus of the 6-volt battery pack should be connected to the ground bus of the 12-volt battery pack.

After you check your circuit, follow these steps, and record your measurements in the blank table following the steps:

1. Adjust the potentiometer to set V_{GS} at 0 volts. (Your multimeter may indicate a few tenths of a millivolt; that's close enough.)
2. Measure and record V_{GS} and I_D .
3. Adjust the potentiometer slightly to give a higher value of V_{GS} .

4. Measure and record the new values of V_{GS} and I_D.

5. Repeat steps 3 and 4 until I_D drops to 0 mA.

V_{GS} (Volts)	I_D (mA)

6. Graph the points recorded in the table, using the blank graph in Figure 4.24. Draw a curve through the points. Your curve should look like the one in Figure 4.22.

Figure 4.24

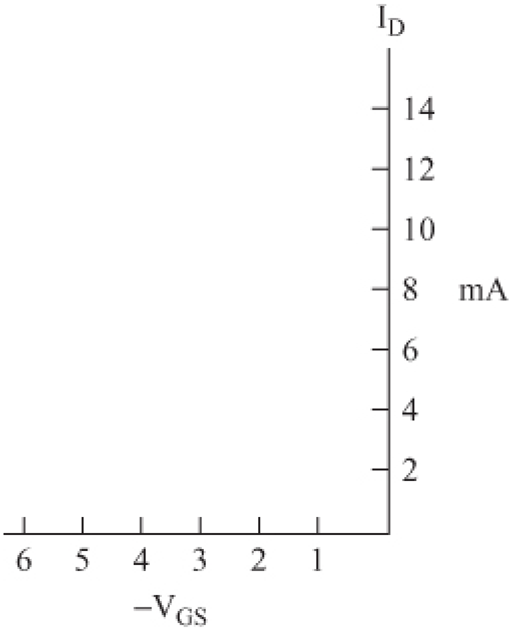

Expected Results

Figure 4.25 shows the breadboarded circuit for this project.

Figure 4.25

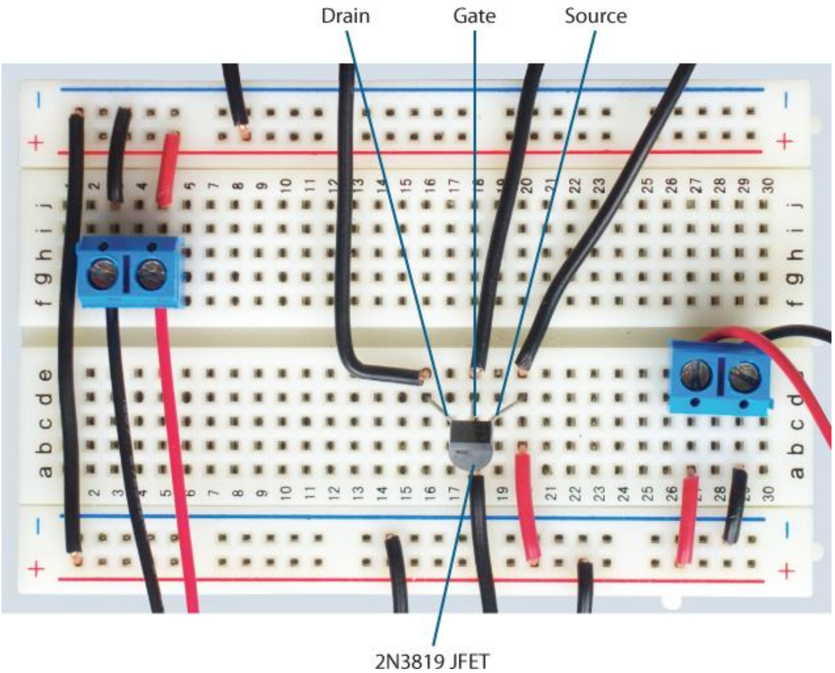
Drain
Gate
Source
2N3819 JFET

Figure 4.26 shows the test setup for this project.

Figure 4.26

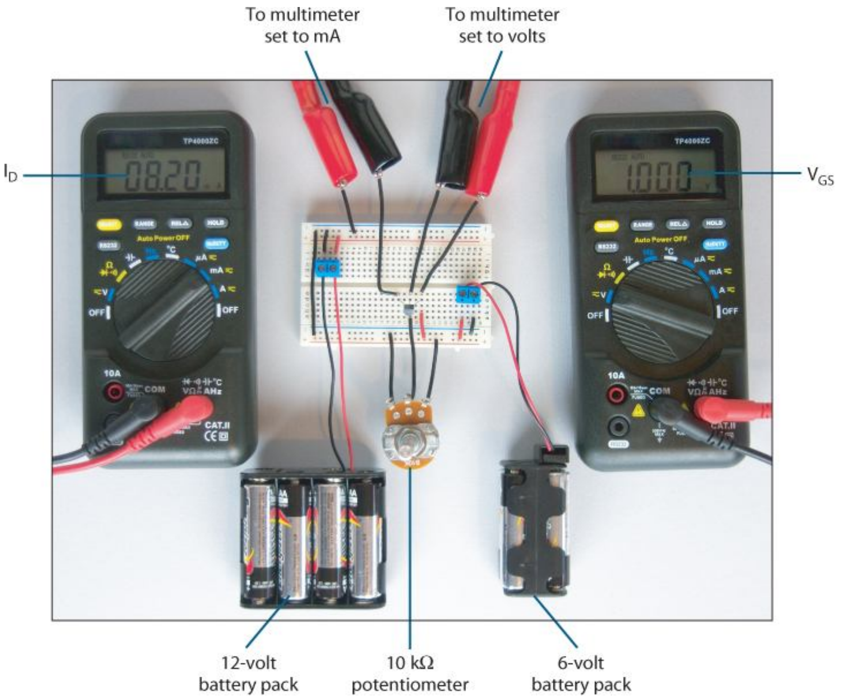
To multimeter set to mA
To multimeter set to volts
I_D
V_{GS}
12-volt battery pack
10 kΩ potentiometer
6-volt battery pack

Compare your measurements with the ones shown in the following table. You should see a similar trend in the measured values, not exactly the same values.

V_{GS} (Volts)	I_D (mA)
0	12.7
0.4	10.7
0.6	9.8
0.8	8.9
1.0	8.1
1.3	6.8
1.5	6.0
1.8	4.9
2.0	4.1
2.3	3.1

2.5	2.5
2.7	1.9
3.0	1.1
3.3	0.5
3.5	0.2
3.7	0.1
4.0	0

Figure 4.27 is the V-I curve generated using the measurements shown in the preceding table. This graph is called the transfer curve for the JFET.

Figure 4.27

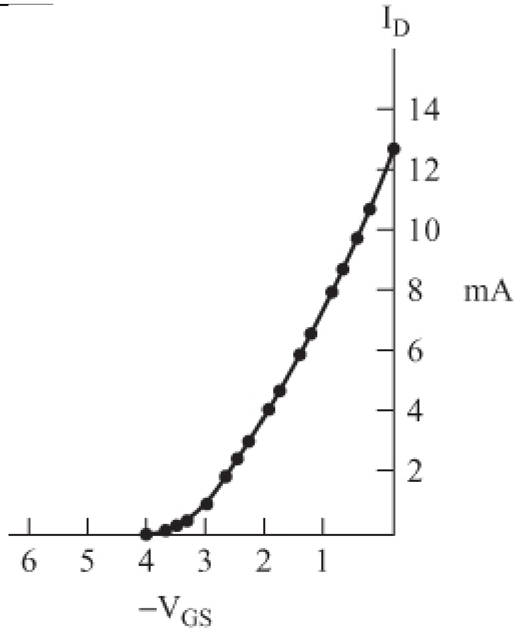

With the potentiometer set to 0 ohms (point A in Figure 4.22), the voltage from the gate to the source is zero (V GS = 0). The current that flows between the drain and source terminals of the JFET at this time is at its maximum value and is called the *saturation current* (I_{DSS}).

Note One property of the saturation current is that when V_{GS} is set at zero, and the transistor is fully ON, the current doesn't drop as long as the value of V_{DS} is above a few volts. If you have an adjustable power supply, you can determine the value of V_{DS} at which I_D starts to drop by starting with the power supply set at 12 volts. Watch the value of I_D as you lower the power supply voltage until you see I_D start to decrease.

38 Refer to the transfer curve shown in Figure 4.28 to answer the following questions.

Figure 4.28

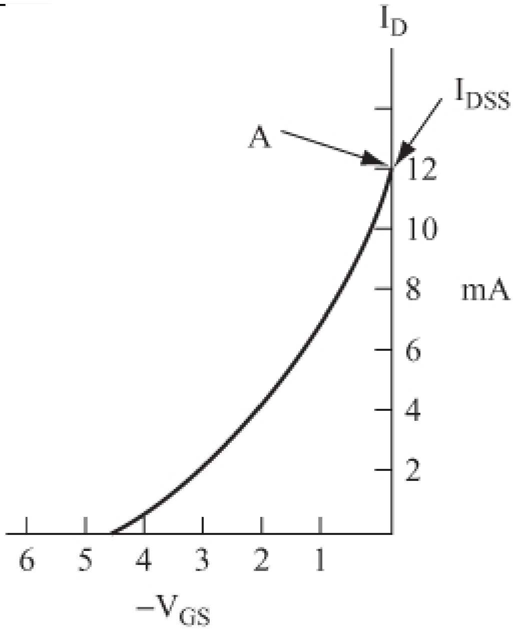

Questions

Using the transfer curve shown in Figure 4.28

, answer the following:

A. With $V_{GS} = 0$, what is the value of the drain current? _____

B. Why is this value called the drain saturation current? _____

C. What is the gate to source cutoff voltage for the curve shown? _____

D. Why is this called a cutoff voltage? _____

Answers

A. 12 mA on the graph.

B. The word “saturation” is used to indicate that the current is at its maximum.

C. Approximately −4.2 V on the graph.

D. It is termed a “cutoff voltage” because at this value, the drain current goes to 0 ampere.

39 Now, look at the circuit shown in Figure 4.29 . Assume that the JFET has the transfer characteristic shown by the curve in problem 38.

Figure 4.29

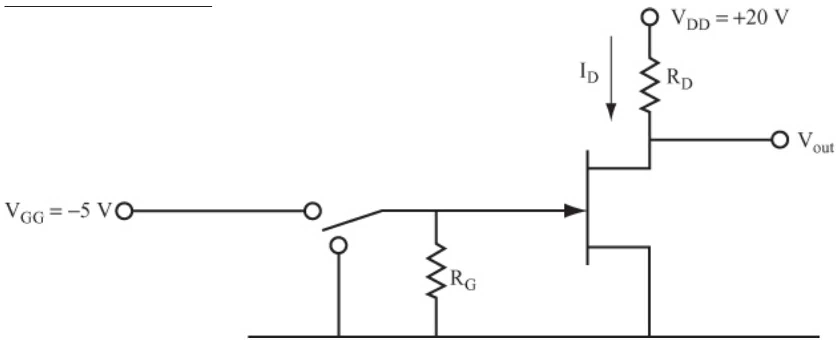

When the gate is connected to ground, the drain current will be at 12 mA. Assuming that the drain to source resistance is negligible, you can calculate the required value for R_D using the following formula:

$$R_D = \frac{V_{DD}}{I_{DSS}}$$

If you know the drain to source voltage, then you can include it in the calculation.

$$R_D = \frac{(V_{DD} - V_{DS})}{I_{DSS}}$$

Question

What should the value of R_D be for the I_{DSS} shown at point A in the curve? _____

Answer

$$R_D = \frac{20\,\text{volts}}{12\,\text{mA}} = 1.67\,\text{k}\Omega$$

40 For the JFET circuit shown in Figure 4.29 , assume that V_{DS} = 1 volt when the I_D is at saturation.

Questions

A. What is the required value of R_D ? _____

B. What is the effective drain to source resistance (r_{DS}) in this situation? _____

Answers

A.

$$R_D = \frac{(20\,\text{volts} \quad 1\,\text{volt})}{12\,\text{mA}} = 1583\,\text{ohms}$$

B.

$$r_{DS} = \frac{V_{DS}}{I_{DSS}} = \frac{1\,\text{volt}}{12\,\text{mA}} = 83\,\text{ohms}$$

Note You can see from this calculation that R_D is 19 times greater than r_{DS} . Thus, ignoring V_{DS} and assuming that $r_{DS} = 0$ does not greatly affect the value of R_D. The 1.67 kΩ value is only about 5 percent higher than the 1583 ohm value for R_D.

41 Now, turn the JFET OFF. From the curve shown in Figure 4.28, you can see that a cutoff value of −4.2 volts is required. Use a gate to source value of −5 volts to ensure that the JFET is in the “hard OFF” state. The purpose of resistor R_G is to ensure that the gate is connected to ground while you flip the switch between terminals, changing the gate voltage from one level to the other. Use a large value of 1 MΩ here to avoid drawing any appreciable current from the gate supply.

Question

When the gate is at the −5 V potential, what is the drain current and the resultant output voltage? _____

Answer

$I_D = 0$ ampere and $V_{out} = V_{DS} = 20$ volts, which is V_{DD}

Summary

In this chapter, you learned about the transistor switch and how to calculate the resistor values required to use it in a circuit.

- You worked with a lamp as the load example because this provides an easy visual demonstration of the switching action.

All the circuits shown in this chapter work when you build them on a breadboard, and the voltage and current measurements are close to those shown in the text.

You have not yet learned all there is to transistor switching. For example, you haven't found out how much current a transistor can conduct before it burns out, what maximum voltage a transistor can sustain, or how fast a transistor can switch ON and OFF. You can learn these things from the data sheet for each transistor model, so these things are not covered here.

When you use the JFET as a switch, it does not switch as fast as a BJT, but it does have certain advantages relating to its large input resistance. The JFET does not draw any current from the control circuit to operate. Conversely, a BJT will draw current from the control circuit because of its lower input resistance.

Self-Test

These questions test your understanding of the concepts introduced in this chapter. Use a separate sheet of paper for your diagrams or calculations. Compare your answers with the answers provided.

For the first three questions, use the circuit shown in Figure 4.30. The objective is to find the value of R_B that turns the transistor ON. As you may know, resistors are manufactured with “standard values.” After you have calculated an exact value, choose the nearest standard resistor value from Appendix D, “Standard Resistor Values.”

Figure 4.30

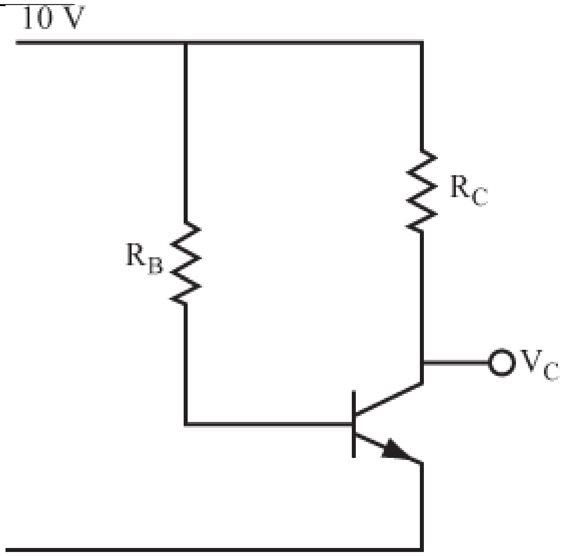

1. R_C = 1 kΩ, β = 100
R_B = _____
2. R_C = 4.7 kΩ, β = 50 R_B = _____
3. R_C = 22 kΩ, β = 75
R_B = _____

For questions 4–6, use the circuit shown in Figure 4.31 . Find the values of R_3 , R_2 , and R_1 that ensure that Q_2 is ON or OFF when the switch is in the corresponding position. Calculate the resistors in the order given. After you find the exact values, again choose the nearest standard resistor values.

Figure 4.31

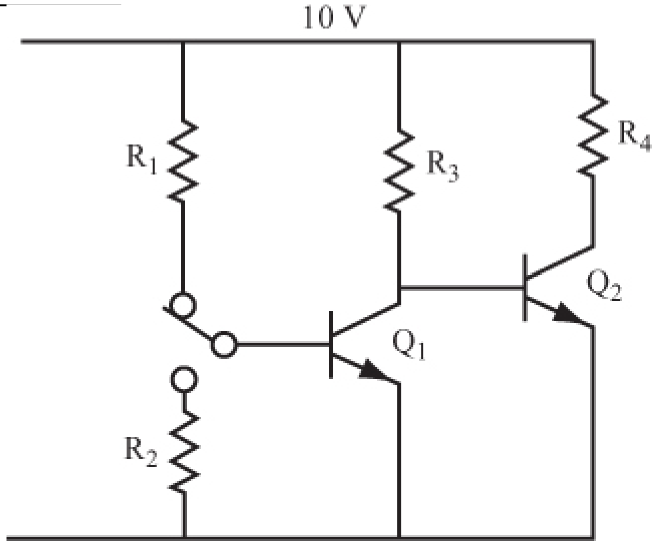

Note Rounding off throughout a problem, or rounding off the final answer, could produce slightly different results.

4. R_4 = 100 ohms, β_1 = 100, β_2 = 20.

R_3 = _____

R_1 = _____

R_2 = _____

5. R_4 = 10 ohms, β_1 = 50, β_2 = 20.

R_3 = _____

R_1 = _____

R_2 = _____

6. R_4 = 250 ohms, β_1 = 75, β_2 = 75.

R_3 = _____

R_1 = _____

R_2 = _____

For questions 7–9, find the values of the resistors in the circuit shown in Figure 4.32

that ensure that Q_3 will be ON or OFF when the switch is in the corresponding position. Then, select the nearest standard resistor values.

Figure 4.32

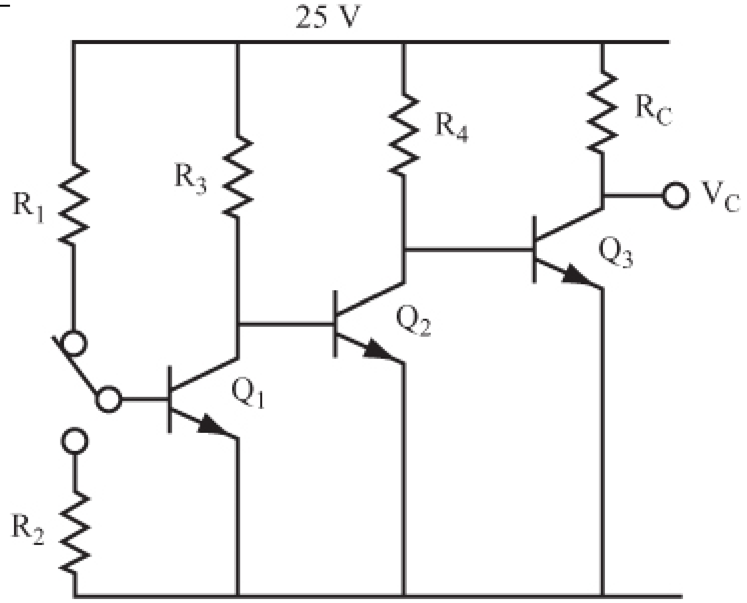

7. R_C = 10 ohms, β_3 = 20, β_2 = 50, β_1 = 100.

R_4 = _____
R_2 = _____
R_3 = _____
R_1 = _____

8. R_C = 28 ohms, β_3 = 10, β_2 = 75, β_1 = 75.

R_4 = _____
R_2 = _____
R_3 = _____
R_1 = _____

9. R_C = 1 ohm, β_3 = 10, β_2 = 50, β_1 = 75.

R_4 = _____

R_2 = _____

R_3 = _____

R_1 = _____

Questions 10–12 use the circuit shown in Figure 4.33. Find values for R_1 and R_2 that ensure that the transistor turns ON when the switch is closed and OFF when the switch is open.

Figure 4.33

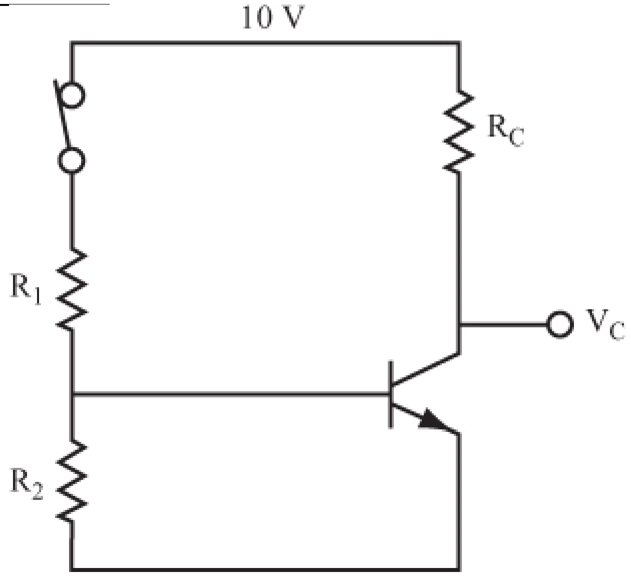

10. R_C = 1 kΩ, β = 100.

R_1 = _____

R_2 = _____

11. R_C = 22 kΩ, β = 75.

R_1 = _____

R_2 = _____

12. R_C = 100 Ω, β = 30.

R_1 = _____

R_2 = _____

13. An N-channel JFET has a transfer curve with the following characteristics. When V_{GS} = 0 volt, the saturation current (I_{DSS}) is 10.5 mA, and the cutoff voltage is −3.8 volts. With a drain supply of 20 volts, design a biasing circuit that switches the JFET from the ON state to the OFF state.

Answers to Self-Test

The exercises in this Self-Test show calculations that are typical of those found in practice, and the odd results you sometimes get are quite common. Thus, choosing a nearest standard value of resistor is a common practice. If your answers do not agree with those given here, review the problems indicated in parentheses before you go on to Chapter 5.

1. 100 kΩ (problem 8)
2. 235 kΩ. Choose 240 kΩ as a standard value. (problem 8)
3. 1.65 MΩ. Choose 1.6 MΩ as a standard value. (problem 8)
4. R_3 = 2 kΩ; R_1 = 200 kΩ; R_2 = 200 kΩ. Use these values. (problem 22)
5. R_3 = 200 ohms; R_1 = 10 kΩ; R_2 = 10 kΩ. Use these values. (problem 22)
6. R_3 = 18.8 kΩ. Choose 18 kΩ as a

standard value. (problem 22)

$R_1 = 1.41$ MΩ. Choose 1.5 MΩ as a standard value.

Select 1 MΩ for R_2.

7. $R_4 = 200$ ohms; $R_3 = 10$ kΩ; $R_2 = 1$ MΩ; $R_1 = 1$ MΩ. Use these values. (problem 26)

8. $R_4 = 280$ ohms. Choose 270 ohms as a standard value. (problem 26)

$R_3 = 21$ kΩ. Choose 22 kΩ as a standard value.

$R_2 = 1.56$ MΩ. Choose 1.5 or 1.6 MΩ as a standard value.

$R_1 = 1.56$ MΩ. Choose 1.5 or 1.6 MΩ as a standard value.

9. $R_4 = 10$ ohms. Choose 10 ohms as a standard value. (problem 26)

$R_3 = 500$ ohms. Choose 510 ohms as a standard value.

$R_2 = 37.5$ kΩ. Choose 39 kΩ as a standard value.

$R_1 = 37.5$ kΩ. Choose 39 kΩ as a standard value.

10. $R_2 = 700$ ohms. Choose 680 or 720 ohms as a standard value. (problems 31–33)

$R_1 = 8.45$ kΩ. Choose 8.2 kΩ as a standard value.

If 0.7 is ignored, then $R_1 = 9.1$ kΩ.

11. $R_2 = 11.7$ kΩ. Choose 12 KΩ as a standard value. (problems 31–33)

R_1 = 141 kΩ. Choose 140 or 150 kΩ as a standard value.

12. R_2 = 21 ohms. Choose 22 ohms as a standard value. (problems 31–33)

R_1 = 273 ohms. Choose 270 ohms as a standard value.

13. Use the circuit shown in Figure 4.29. Set the gate supply at a value slightly more negative than −3.8 volts. A value of −4 V would work. Make resistor R_G = 1 MΩ. Set R_D at a value of (20 volts)/(10.5 mA), which calculates a resistance of 1.9 kΩ. You can wire a standard resistor of 1 kΩ in series with a standard resistor of 910 ohms to obtain a resistance of 1.91 kΩ. (problems 39 and 41)

Chapter 5

AC Pre-Test and Review

You need to have some basic knowledge of alternating current (AC) to study electronics. To understand AC, you must understand sine waves.

A *sine wave* is simply a shape, like waves in the ocean. Sine waves in electronics are used to represent voltage or current moving up and down in magnitude. In AC electronics, some signals or power sources (such as the house current provided at a wall plug) are represented by sine waves. The sine wave shows how the voltage moves from 0 volts to its peak voltage and back down through 0, its negative peak voltage, at 60 cycles per second, or 60 Hertz (Hz).

The sound from a musical instrument also consists of sine waves. When you combine sounds (such as all the instruments in an orchestra), you get complex combinations of many sine waves at various frequencies.

The study of AC starts with the properties of simple sine waves and continues with an examination of how electronic circuits can generate or change sine waves.

This chapter discusses the following:

Generators

Sine waves

Peak-to-peak and root mean square voltages

Resistors in AC circuits
Capacitive and inductive reactance
Resonance

The Generator

1 In electronic circuits powered by direct current (DC), the voltage source is usually a battery or solar cell, which produces a constant voltage and a constant current through a conductor.

In electronic circuits or devices powered by alternating current (AC), the voltage source is usually a *generator* , which produces a regular output waveform, such as a sine wave.

Question

Draw one cycle of a sine wave.

Answer

Figure 5.1 shows one cycle of a sine wave.

Figure 5.1

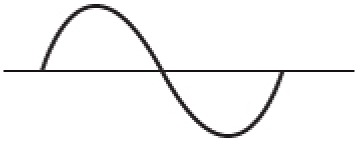

2 A number of electronic instruments are used in the laboratory to produce sine waves. For purposes of this discussion, the term *generator* means a sine wave source. These generators enable you to adjust the voltage and frequency by turning a dial or pushing a button. These instruments are called by various names, generally based on the method of producing the sine wave, or the application as a test instrument. The most popular generator at present is called a *function generator*. It provides a choice of functions or waveforms, including a square wave and a triangle wave. These waveforms are useful in testing certain electronic circuits.

The symbol shown in Figure 5.2 represents a generator. Note that a sine wave shown within a circle designates an AC sine wave source.

Figure 5.2

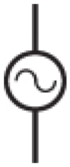

Questions

A. What is the most popular instrument used in the lab to produce waveforms? _____

B. What does the term AC mean? _____

C. What does the sine wave inside a generator symbol indicate? _____

Answers

A. Function generator.

B. Alternating current, as opposed to direct current.

C. The generator is a sine wave source.

3 Figure 5.3 shows some key parameters of sine waves. The two axes are voltage and time.

Figure 5.3

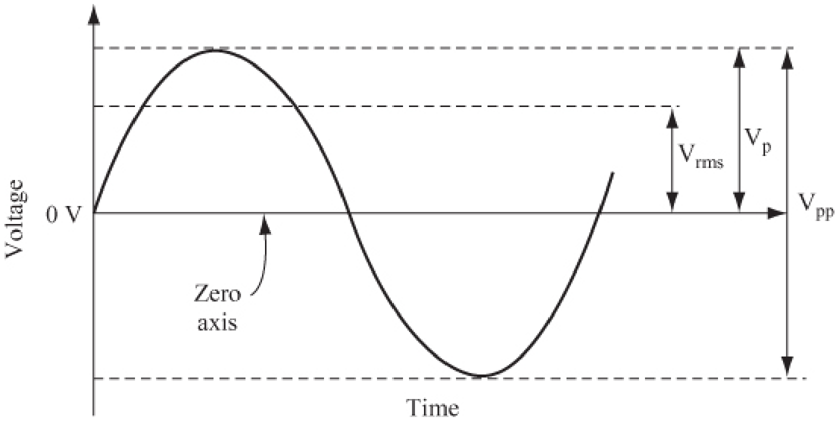

The *zero axis* is the reference point from which all voltage measurements are made.

Questions

A. What is the purpose of the zero axis? _____

B. What is the usual point for making time measurements? _____

Answers

A. It is the reference point from which all voltage measurements are made.

B. Time measurements can be made from any point in the sine wave, but usually they are made from a point at which the sine wave crosses the zero axis.

4 The three most important voltage or amplitude measurements are the *peak (p)* , *peak-to-peak (pp)* , and the *root mean square (rms)* voltages.

The following equations show the relationship between p, pp, and rms voltages for sine waves. The relationships between p, pp, and rms voltages differ for other waveforms (such as square waves).

$$V_p a = \sqrt{2} \quad V_{rms}$$

$$V_{pp} = 2V_p = 2 \quad \sqrt{2} \quad V_{rms}$$

$$V_{rms} = \frac{1}{\sqrt{2}} \quad V_p = \frac{1}{\sqrt{2}} \quad \frac{V_{pp}}{2}$$

Note the following:

$$\sqrt{2} = 1.414$$

$$\frac{1}{\sqrt{2}} = 0.707$$

Question

If the pp voltage of a sine wave is 10 volts, find the rms voltage. _____

Answer

$$V_{rms} = \frac{1}{\sqrt{2}} \times \frac{V_{pp}}{2} = 0.707 \times \frac{10}{2} = 3.535\,V$$

5 Calculate the following for a sine wave.

Question

If the rms voltage is 2 volts, find the pp voltage. _____

Answer

$$V_{pp} = 2 \times \sqrt{2} \times V_{rms} = 2 \times 1.414 \times 2 = 5.656\,V$$

6 Calculate the following for a sine wave.

Questions

A. V pp = 220 volts. Find V rms . _____

B. V rms = 120 volts. Find V pp . _____

Answers

A. 77.77 volts

B. 340 volts (This is the common house current supply voltage; 340 V pp = 120 V rms .)

7 There is a primary time measurement for sine waves. The duration of the complete sine wave is shown in Figure 5.4 and referred to as a *cycle* . All other time measurements are fractions or multiples of a cycle.

Figure 5.4

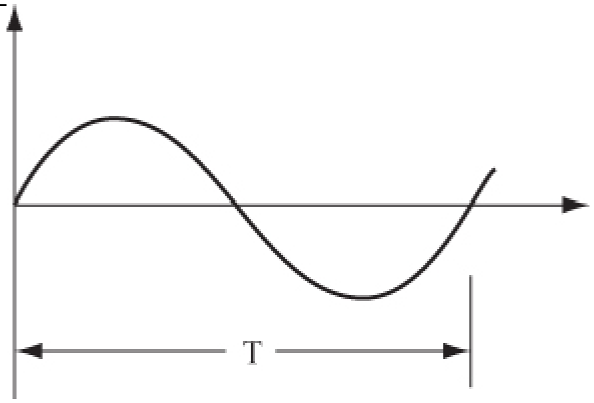

Questions

A. What is one complete sine wave called? _____

B. What do you call the time it takes to complete one sine wave? _____

C. How is the frequency of a sine wave related to this time? _____

D. What is the unit for frequency? _____

E. If the period of a sine wave is 0.5 ms, what is its frequency? What is the frequency of a sine wave with a period of 40 μsec? _____

F. If the frequency of a sine wave is 60 Hz, what is its period? What is the period of sine waves with frequencies of 12.5 kHz and 1 MHz? _____

Answers

A. Cycle

B. The period, T

C. f = 1/T

D. Hertz (Hz) is the standard unit for frequency. One Hertz equals one cycle per second.

E. 2 kHz, 25 kHz

F. 16.7 ms, 80 μsec, 1 μsec

8 Choose all answers that apply.

Questions

Which of the following could represent electrical AC signals?

A. Simple sine wave

B. Mixture of many sine waves, of different frequencies and amplitudes

C. Straight line

Answer

A and B

Resistors in AC Circuits

9 Alternating current is passed through components, just as direct current is. Resistors interact with alternating current just as they do with direct current.

Question

Suppose an AC signal of 10 V pp is

connected across a 10-ohm resistor. What is the current through the resistor? _____

Answers

Use Ohm's law:

$$I = \frac{V}{R} = \frac{10\,V_{pp}}{10\,\text{ohms}} = 1\,A_{pp}$$

Because the voltage is given in pp, the current is a pp current.

10 An AC signal of 10 V rms is connected across a 20-ohm resistor.

Question

Find the current. _____

Answer

$$I = \frac{10\,V_{rms}}{20\,\text{ohms}} = 0.5\,A_{rms}$$

Because the voltage was given in rms, the current is in rms.

11 You apply an AC signal of 10 V pp to the voltage divider circuit, as shown in Figure 5.5 .

Figure 5.5

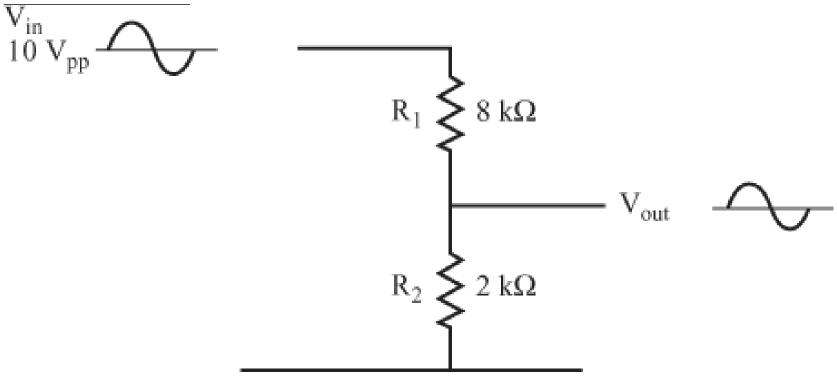

Question

Find V out . _____

Answer

$$V_{out} = V_{in} \times \frac{R_2}{(R_1 + R_2)} = 10 \times \frac{2k}{(8k\ \ + 2k\ \ )} = 10 \times \frac{2}{10} = 2V_{pp}$$

Capacitors in AC Circuits

12 A capacitor opposes the flow of an AC current.

Questions

A. What is this opposition to the current flow called? _____

B. What is this similar to in DC circuits? _____

Answers

A. Reactance

B. Resistance

13 Just as with resistance, you determine reactance by using an equation.

Questions

A. What is the equation for reactance? _____

B. What does each symbol in the equation stand for? _____

C. How does the reactance of a capacitor change as the frequency of a signal increases? _____

Answers

A.
$$X_C = \frac{1}{2\ \ fC}$$

B. X C = the reactance of the capacitor in ohms.

f = the frequency of the signal in hertz.

C = the value of the capacitor in farads.

C. The reactance of a capacitor decreases as the frequency of the signal increases.

14 Assume the capacitance is 1 μF and the frequency is 1 kHz.

Question

Find the capacitor's reactance. (*Note* : 1/(2π) = 0.159, approximately.) _____

Answers

$$X_C = \frac{1}{2\ \ fC}$$

f = 1 kHz = 10 3 Hz

C = 1 μF = 10 −6 F

Thus,

$$X_C = \frac{0.159}{10^3 \times 10^{\ \ 6}} = 160\,\text{ohms}$$

15 Now, perform these simple calculations. In each case, find X C1 (the capacitor's reactance at 1 kHz) and X C2 (the capacitor's reactance at the frequency specified in the question).

Questions

Find X C1 and X C2 :

A. C = 0.1 μF, f = 100 Hz. _____

B. C = 100 μF, f = 2 kHz. _____

Answers

A. At 1 kHz, X C1 = 1600 ohms; at 100 Hz, X C2 = 16,000 ohms

B. At 1 kHz, X C1 = 1.6 ohms; at 2 kHz, X C2 = 0.8 ohms

A circuit containing a capacitor in series with a resistor (as shown in Figure 5.6) functions as a voltage divider.

Figure 5.6

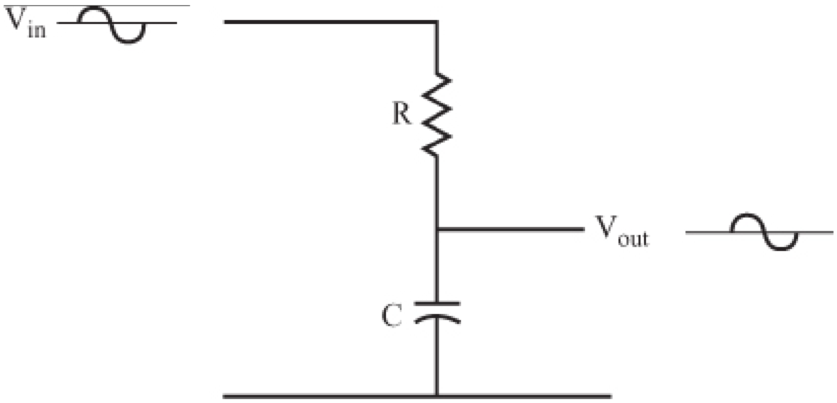

Although this voltage divider provides a reduced output voltage, just like a voltage divider using two resistors, there's a complication. If you view the output and input voltage waveforms on an oscilloscope, you see that one is shifted away from the other. The two waveforms are said to be “out of phase.” *Phase* is an important concept in understanding how certain electronic circuits work. In Chapter 6, “Filters,” you learn about phase relationships for some AC circuits. You also encounter this again when you study amplifiers.

Using the Oscilloscope

You use an oscilloscope to measure AC signals generated by a circuit, or to measure the effect that a circuit has on AC signals. The key parameters you measure with an oscilloscope are *frequency* and *peak-to-peak voltage*. An oscilloscope can also be used to show the shape of a signal's waveform so that you can ensure that the circuit works properly. When using an oscilloscope to compare a circuit's input signal to its output signal, you can determine the phase shift, as well as the change in V_{pp}.

The following figure shows an oscilloscope whose probe connects to the output of an oscillator circuit to measure the frequency of the signal generated by the oscillator. (Oscillator circuits are discussed in Chapter 9, “Oscillators.”) This example uses an analog oscilloscope, but you can also use a digital oscilloscope, which automates many of the measurements.

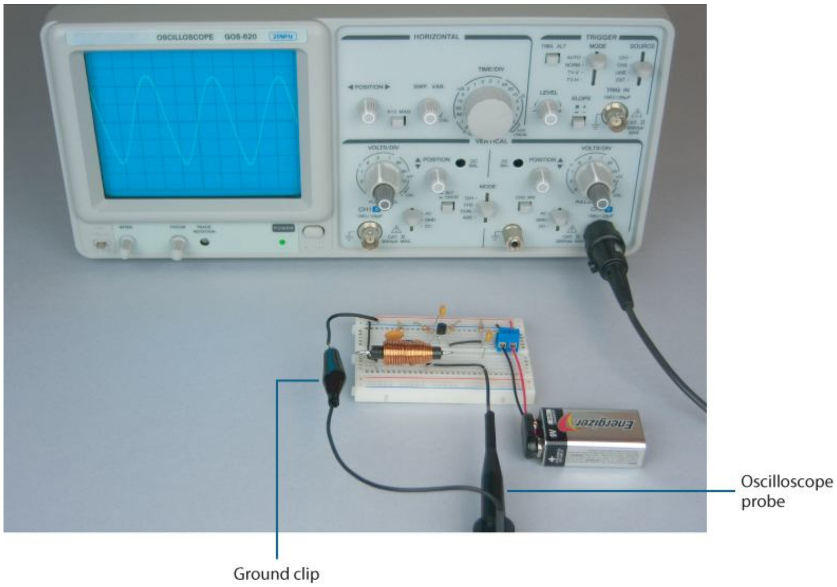

This oscilloscope has two channels, which provide the capability to measure two waveforms at once. Only channel 2 was used for the measurement in the preceding figure. The oscilloscope probe was clipped to a jumper wire connecting to

V out for the circuit, and the oscilloscope ground clip was clipped to a jumper wire connecting to the ground bus.

The following figure shows the oscilloscope control panel. You use the VOLTS/ DIVcontrol to adjust the vertical scale and the TIME/DIVcontrol to adjust the horizontal scale. Set the vertical position knob and the horizontal position knob to adjust the position of the waveform against the grid to make it easier to measure.

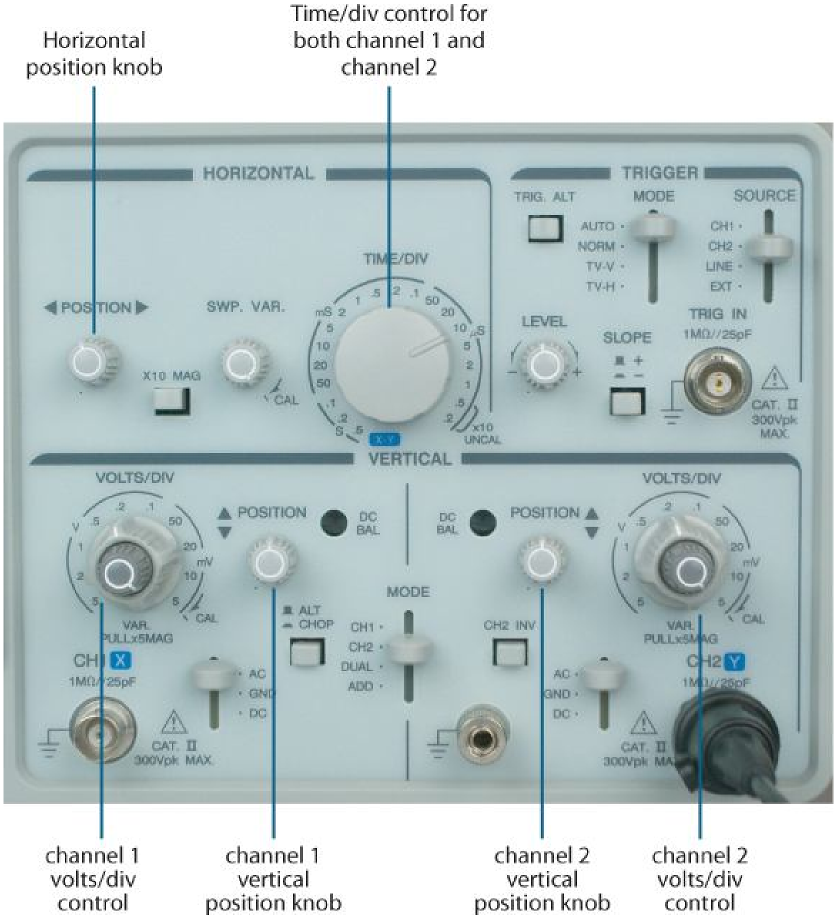
Horizontal position knob
Time/div control for both channel 1 and channel 2
HORIZONTAL
TRIGGER
TRIG. ALT
MODE
SOURCE
AUTO
NORM
TV-V
TV-H
CH1
CH2
LINE
EXT
POSITION
SWP. VAR.
TIME/DIV
LEVEL
SLOPE
TRIG IN
X10 MAG
CAL
X-Y
x10 UNCAL
VERTICAL
VOLTS/DIV
POSITION
DC BAL
MODE
ALT
CHOP
CH1
CH2
DUAL
ADD
CH2 INV
AC
GND
DC
VAR. PULLx5MAG
CH1 X
CH2 Y
1MΩ//25pF
CAT. II 300Vpk MAX.
channel 1 volts/div control
channel 1 vertical position knob
channel 2 vertical position knob
channel 2 volts/div control

You can determine the period of this waveform by counting the number of horizontal divisions the waveform takes to complete one cycle, and then multiplying the number of divisions by the TIME/DIVsetting.

In the following figure, the period of the sine wave generated by this oscillator circuit is approximately 3.3 divisions wide.

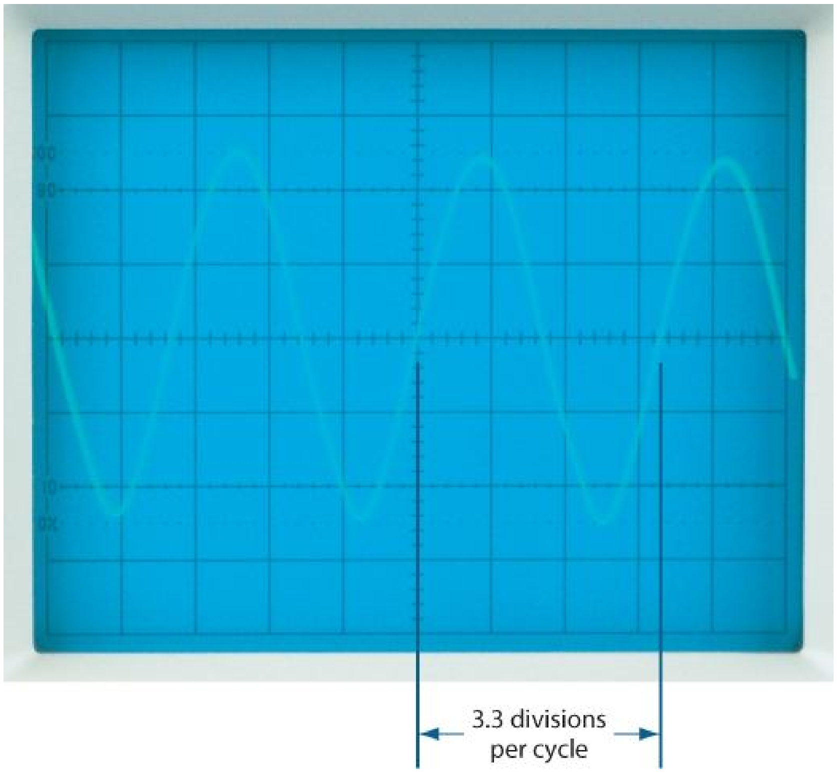

Because the TIME/DIVknob is set at 10 μs, the period of this sine wave is 33 μs. The frequency of this sine wave is therefore calculated as follows:

$$f = \frac{1}{T} = \frac{1}{33\ \ s} = \frac{1}{.000033\,\mathrm{sec}} = 30303\ \mathrm{Hz} = 30.3\ \mathrm{kHz}$$

You can also measure the effect of a circuit on a signal of a particular frequency. Supply the signal from a function generator to the input of the circuit. Attach the oscilloscope probe for channel 2 to the input of the circuit. Attach the oscilloscope probe for channel 1 to the output of the circuit. The following figure shows a function generator and oscilloscope attached to the voltage divider circuit shown in Figure 5.6 .

In this example, the red lead from the function generator was clipped to a jumper wire connected to the resistor in the voltage divider circuit, and the black lead was clipped to a jumper wire connected to the ground bus. The oscilloscope probe for channel 2 is clipped to a jumper wire connected to the resistor, and the ground clip is clipped to a jumper wire attached to the ground bus. The oscilloscope probe for channel 1 is clipped to a jumper wire connected to the voltage divider circuit V out , and the ground clip is clipped to a jumper wire connected to the ground bus.

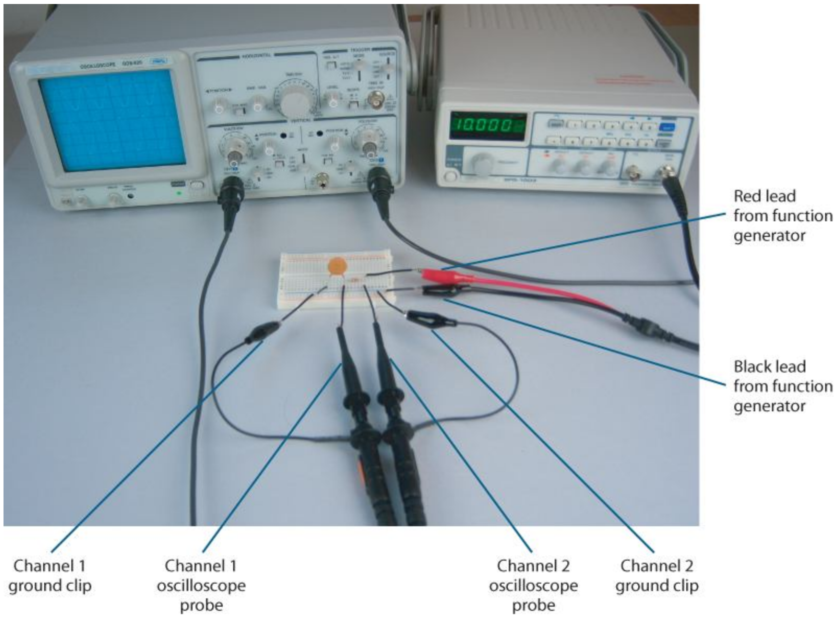
Red lead from function generator
Black lead from function generator
Channel 1 ground clip
Channel 1 oscilloscope probe
Channel 2 oscilloscope probe
Channel 2 ground clip

The function generator supplies an input signal at a frequency of 10 kHz and an amplitude of 10 V_{pp}. The input signal is represented by the upper sine wave on the oscilloscope. Many function generators (such as the one shown here) have an amplitude adjustment knob without a readout. You set the input signal amplitude to 10 V_{pp} with the amplitude knob on the function generator while monitoring the amplitude on the oscilloscope.

The output signal is represented by the lower sine wave in the following figure. Adjust the VOLT/DIVcontrols and vertical position controls for channels 1 and 2 to fit both sine waves on the screen, as shown here.

You can measure V_{pp} for each sine wave by multiplying the number of vertical divisions between peaks by the setting on the VOLT/DIVknobs. For the input sine wave in this example, this measurement is two divisions at 5 VOLTS/DIV, for a total of 10 volts. For the output sine wave, this measurement is 3 divisions at 2 VOLTS/DIV, for a total of 6 volts. This indicates that the circuit has decreased the input signal from 10 V_{pp} to 6 V_{pp}.

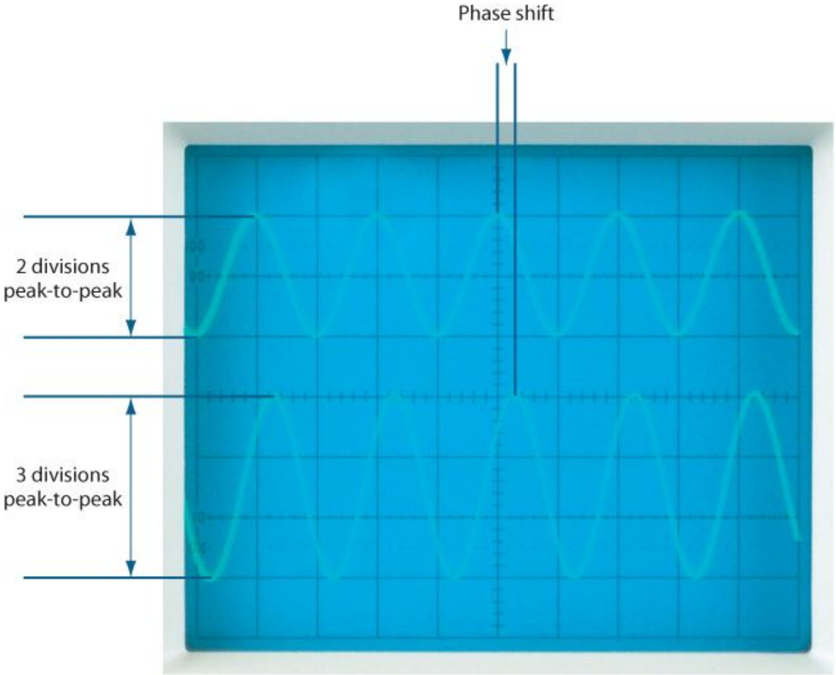
Phase shift
2 divisions peak-to-peak
3 divisions peak-to-peak

Also note that that the peak of the output waveform shifts from the input waveform, a phenomenon called *phase shift* . You learn more about how to calculate phase shift in Chapter 6.

The Inductor in an AC Circuit

16 An *inductor* is a coil of wire, usually wound many times around a piece of soft iron. In some cases, the wire is wound around a nonconducting material.

Questions

A. Is the AC reactance of an inductor high or low? Why? _____

B. Is the DC resistance high or low? _____

C. What is the relationship between the AC reactance and the DC resistance? _____

D. What is the formula for the reactance of an inductor? _____

Answers

A. Its AC reactance (X_L), which can be quite high, is a result of the electromagnetic field that surrounds the coil and induces a current in the opposite direction of the original current.

B. Its DC resistance (r), which is usually quite low, is simply the resistance of the wire that makes up the coil.

C. None

D. $X_L = 2\pi fL$, where L = the value of the inductance in henrys (H). Using this equation, you can expect the reactance of an inductor to increase as the frequency of a signal passing through it increases.

17 Assume the inductance value is 10 henrys (H) and the frequency is 100 Hz.

Question

Find the reactance. _____

Answer

$X_L = 2\pi fL = 2\pi \times 100 \times 10 = 6280$ ohms

18 Now, try these two examples. In each case, find X_{L1} (the reactance of the inductor at 1 kHz) and X_{L2} (the reactance at the frequency given in the question).

Questions

A. L = 1 mH (0.001 H), f = 10 kHz _____

B. L = 0.01 mH, f = 5 MHz _____

Answers

A. $X_{L1} = 6.28 \times 10^3 \times 0.001 = 6.28$ ohms

$X_{L2} = 6.28 \times 10 \times 10^3 \times 0.001 = 62.8$ ohms

B. $X_{L1} = 6.28 \times 10^3 \times 0.01 \times 10^{-3} = 0.0628$ ohms

$X_{L2} = 6.28 \times = \times 10^6 \times 0.01 \times 10^{-3} = 314$ ohms

A circuit containing an inductor in series with a resistor functions as a voltage divider, just as a circuit containing a capacitor in series with a resistor does. Again, the relationship between the input and output voltages is not as simple as a resistive divider. The circuit is discussed in Chapter 6.

Resonance

19 Calculations in previous problems demonstrate that capacitive reactance decreases as frequency increases, and that inductive reactance increases as frequency increases. If a capacitor and an inductor are connected in series, there is one frequency at which their reactance values are equal.

Questions

A. What is this frequency called? _____

B. What is the formula for calculating this frequency? You can find it by setting $X_L = X_C$ and solving for frequency. _____

Answers

A. The resonant frequency

B. $2\pi fL = 1/(2\pi fC)$. Rearranging the terms in this equation to solve for f yields the following formula for the resonant frequency (f_r):

$$f_r = \frac{1}{2\ \sqrt{LC}}$$

20 If a capacitor and an inductor are connected in parallel, there is also a resonant frequency. Analysis of a parallel resonant circuit is not as simple as it is for a series resonant circuit. The reason for this is that inductors always have some internal resistance, which complicates some of the equations. However, under certain conditions, the analysis is similar. For example, if the reactance of the inductor in ohms is more than 10 times greater than its own internal resistance (r), the formula for the resonant frequency is the same as if the inductor and capacitor were connected in series. This is an approximation that you use often.

Questions

For the following inductors, determine if the reactance is more or less than 10 times its internal resistance. A resonant frequency is provided.

A. f_r = 25 kHz, L = 2 mH, r = 20 ohms

_____

B. f_r = 1 kHz, L = 33.5 mH, r = 30 ohms _____

Answers

A. X_L = 314 ohms, which is more than 10 times greater than r

B. X_L = 210 ohms, which is less than 10 times greater than r

Note Chapter 7, “Resonant Circuits,” discusses both series and parallel resonant circuits. At that time, you learn many useful techniques and formulas.

21 Find the resonant frequency (f_r) for the following capacitors and inductors when they are connected both in parallel and in series. Assume r is negligible.

Questions

Determine f_r for the following:

A. C = 1 μF, L = 1 henry _____

B. C = 0.2 μF, L = 3.3 mH _____

Answers

A.
$$f_r = \frac{0.159}{\sqrt{10^{6} \times 1}} = 160\,\text{Hz}$$

B.
$$f_r = \frac{0.159}{\sqrt{3.3 \times 10^{3} \times 0.2 \times 10^{6}}} = 6.2\,\text{kHz}$$

22 Now, try these two final examples.

Questions

Determine f_r:

A. C = 10 μF, L = 1 henry _____

B. C = 0.0033 μF, L = 0.5 mH _____

Answers

A. $f_r = 50$ Hz (approximately)

B. $f_r = 124$ kHz

Understanding resonance is important to understanding certain electronic circuits, such as filters and oscillators.

Filters are electronic circuits that either block a certain band of frequencies, or pass a certain band of frequencies. One common use of filters is in circuits used for radio, television, and other communications applications. *Oscillators* are electronic circuits that generate a continuous output without an input signal. The type of oscillator that uses a resonant circuit produces pure sine waves. (You learn more about oscillators in Chapter 9.)

Summary

Following are the concepts presented in this chapter:

The sine wave is used extensively in AC circuits.

The most common laboratory generator is the function generator.

$$V_p = \sqrt{2} \times V_{rms}, V_{pp} = 2\sqrt{2} = V_{rms}$$

$f = 1/T$

$$I_{pp} = \frac{V_{pp}}{R},\ I_{rms} = \frac{V_{rms}}{R}$$

Capacitive reactance is calculated as follows:

$$X_C = \frac{1}{(2\ \ fC)}$$

Inductive reactance is calculated as follows:

$$X_L = 2\ \ fL$$

Resonant frequency is calculated as follows:

$$f_r = \frac{1}{2\ \ \sqrt{LC}}$$

Self-Test

The following problems test your understanding of the basic concepts presented in this chapter. Use a separate sheet of paper for calculations if necessary. Compare your answers with the answers provided following the test.

1. Convert the following peak or peak-to-peak values to rms values:

A. $V_p = 12$ V

V_{rms} = _____

B. $V_p = 80$ mV

V_{rms} = _____

C. $V_{pp} = 100$ V

V_{rms} = _____

2. Convert the following rms values to the required values shown:

A. $V_{rms} = 120$ V

V_p = _____

B. $V_{rms} = 100$ mV

V_p = _____

C. $V_{rms} = 12$ V

V_{pp} = _____

3. For the given value, find the period or frequency:

A. T = 16.7 ms

f = _____

B. f = 15 kHz

T = _____

4. For the circuit shown in Figure 5.7, find the total current flow and the voltage across R_2, (V_{out}). _____

Figure 5.7

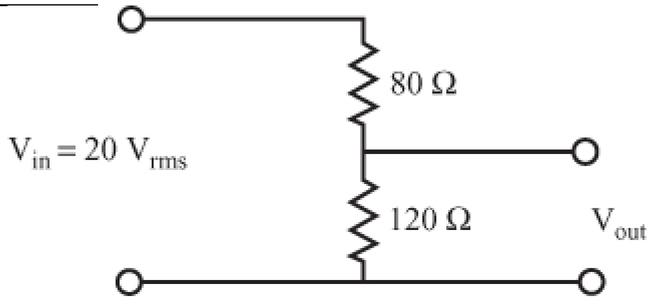

5. Find the reactance of the following components:

A. C = 0.16 μF, f = 12 kHz

X_C = _____

B. L = 5 mH, f = 30 kHz

X_L = _____

6. Find the frequency necessary to cause each reactance shown:

A. C = 1 μF, X_C = 200 ohms

f = _____

B. L = 50 mH, X_L = 320 ohms

f = _____

7. What would be the resonant frequency for the capacitor and inductor values given in A and B of question = if they were connected in series? _____

8. What would be the resonant frequency for the capacitor and inductor values given in A and B of question 6 if they were connected in parallel? What assumption would you need to make? _____

Answers to Self-Test

If your answers do not agree with those provided here, review the problems indicated in parentheses before you go to Chapter 6, “Filters.”

1A. 8.5 V rms (problems 4–6)
1B. 56.6 V rms
1C. 35.4 V rms
2A. 169.7 V_p (problems 4–6)
2B. 141.4 mV_p
2C. 33.9 V_{pp}
3A. 60 Hz (problem 7)
3B. 66.7 μsec
4. I_T = 0.1A rms, V_{out} = 12V rms (problems 9–11)
5A. 82.9 ohms (problems 14 and 17)
5B. 942.5 ohms
6A. 795.8 Hz (problems 14 and 17)
6B. 1.02 kHz
7. 5.63 kHz (problem 19)
8. 711.8 Hz. Assume the internal resistance of the inductor is negligible. (problem 20)

Chapter 6

Filters

Certain types of circuits are found in most electronic devices used to process alternating current (AC) signals. One of the most common of these, *filter circuits* , is covered in this chapter. Filter circuits are formed by resistors and capacitors (RC), or resistors and inductors (RL). These circuits (and their effect on AC signals) play a major part in communications, consumer electronics, and industrial controls.

When you complete this chapter, you will be able to do the following:

Calculate the output voltage of an AC signal after it passes through a high-pass RC filter circuit.

Calculate the output voltage of an AC signal after it passes through a low-pass RC circuit.

Calculate the output voltage of an AC signal after it passes through a high-pass RL circuit.

Calculate the output voltage of an AC signal after it passes through a low-pass RL circuit.

Draw the output waveform of an AC or combined AC-DC signal after it passes through a filter circuit.

Calculate simple phase angles and phase differences.

Capacitors in AC Circuits

1 An AC signal is continually changing, whether it is a pure sine wave or a complex signal made up of many sine waves. If such a signal is applied to one plate of a capacitor, it will be induced on the other plate. To express this another way, a capacitor will "pass" an AC signal, as illustrated in Figure 6.1 .

Figure 6.1

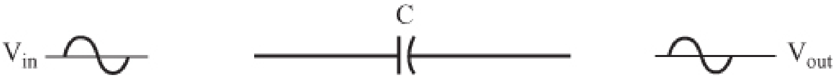

Note Unlike an AC signal, a DC signal is blocked by a capacitor. Equally important is that a capacitor is *not* a short circuit to an AC signal.

Questions

A. What is the main difference in the effect of a capacitor upon an AC signal versus a DC signal? _____

B. Does a capacitor appear as a short or an open circuit to an AC signal? _____

Answers

A. A capacitor will pass an AC signal, whereas it will not pass a DC voltage level.

B. Neither.

2 In general, a capacitor will oppose the flow of an AC current to some degree. As you saw in Chapter 5, "AC Pre-Test and Review," this opposition to current flow is called the *reactance* of the capacitor.

Reactance is similar to resistance, except that the reactance of a capacitor changes when you vary the frequency of a signal. The reactance of a capacitor can be calculated by a formula introduced in Chapter 5.

Question

Write the formula for the reactance of a capacitor. _____

Answer

$$X_C = \frac{1}{2\pi fC}$$

3 From this formula, you can see that the reactance changes when the frequency of the input signal changes.

Question

If the frequency increases, what happens to the reactance? _____

Answer

It decreases.

If you had difficulty with these first three problems, you should review the examples in Chapter 5.

Capacitors and Resistors in Series

4 For simplicity, consider all inputs at this time to be pure sine waves. The circuit shown in Figure 6.2 shows a sine wave as the input signal to a capacitor.

Figure 6.2

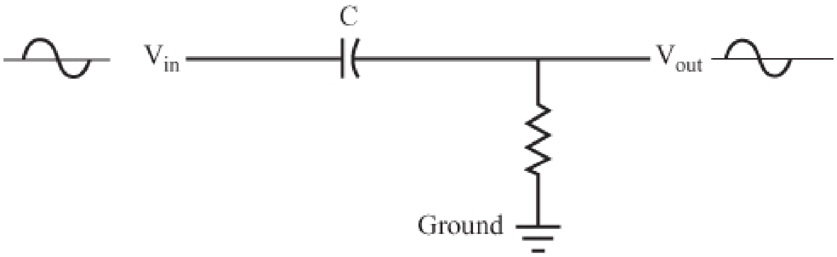

Question

If the input is a pure sine wave, what is the output? _____

Answer

A pure sine wave

5 The output sine wave has the same frequency as the input sine wave. A capacitor cannot change the frequency of the signal. But remember, with an AC input, the capacitor behaves in a manner similar to a resistor in that the capacitor does have some level of opposition to the flow of alternating current. The level of opposition depends upon the value of the capacitor and the frequency of the signal. Therefore, the output amplitude of a sine wave will be less than the input amplitude.

Question

With an AC input to a simple circuit like the one described here, what does the capacitor appear to behave like? _____

Answer

It appears to have opposition to alternating current similar to the behavior of a resistor.

6 If you connect a capacitor and resistor in series (as shown in Figure 6.3), the circuit functions as a voltage divider.

Figure 6.3

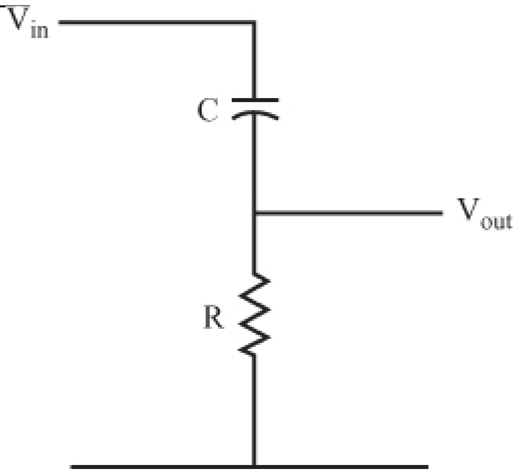

Question

What formula would you use to calculate the output voltage for a voltage divider formed by connecting two resistors in series? _____

Answer

$$V_{out} = V_{in} \times \frac{R_2}{R_1 + R_2}$$

7 You can calculate a total resistance to the flow of electric current for a circuit containing two resistors in series.

Question

What is the formula for this total resistance? _____

Answer

$$R_T = R_1 + R_2$$

8 You can also calculate the total opposition to the flow of electric current for a circuit containing a capacitor and resistor in series. This parameter is called *impedance* , and you can calculate it using the following formula:

$$Z = \sqrt{X_C^2 + R^2}$$

In this equation:

Z = The impedance of the circuit in ohms

X C = The reactance of the capacitor in ohms

R = The resistance of the resistor in ohms

Questions

Use the following steps to calculate the impedance of the circuit, and the current flowing through the circuit, as shown in Figure 6.4 .

Figure 6.4

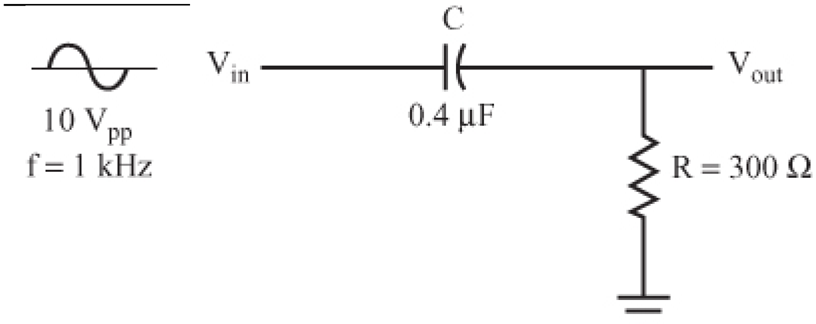

A. $X_C = \dfrac{1}{2\pi fC} =$ _____

B. $Z = \sqrt{X_C^2 + R^2} =$ _____

C. $I = \dfrac{V}{Z} =$ _____

Answers

A. 400 ohms

B. 500 ohms

C. 20 mA pp

9 Now, for the circuit shown in Figure 6.4, calculate the impedance and current using the values provided.

Questions

A. C = 530 μF, R = 12 ohms, V in = 26 V pp, f = 60 Hz _____

B. C = 1.77 μF, R = 12 ohms, V in = 150 V pp, f = 10 kHz _____

Answers

A. Z = 13 ohms, I = 2 A pp

B. Z = 15 ohms, I = 10 A pp

10 You can calculate V out for the circuit shown in Figure 6.5 with a formula similar to the formula used in Chapter 5 to calculate V out for a voltage divider composed of two resistors.

Figure 6.5

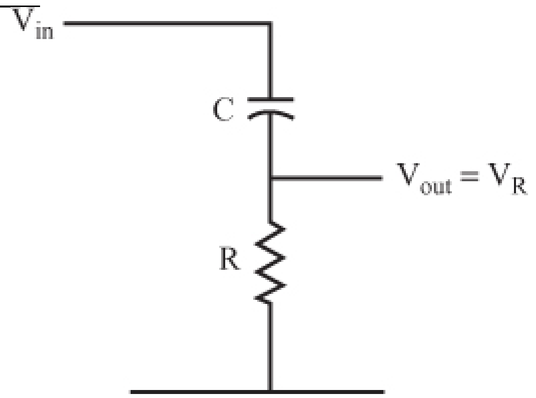

The formula to calculate the output voltage for this circuit is as follows:

$$V_{out} = V_{in} \times \frac{R}{Z}$$

Questions

Calculate the output voltage in this circuit using the component values and input signal voltage and frequency listed on the circuit diagram shown in Figure 6.6 .

Figure 6.6

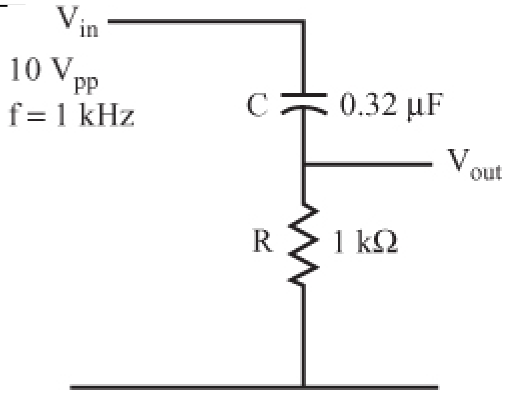

A. Find X_C : _____

B. Find Z: _____

C. Use the formula to find V_{out} : _____

Answers

A. X_C = 500 ohms (rounded off)

B. Z = 1120 ohms (rounded off)

C. V_{out} = 8.9 V_{pp}

11 Now, find V_{out} for the circuit in Figure 6.5 using the given component values, signal voltage, and frequency.

Questions

A. C = 0.16 μF, R = 1 kΩ, V_{in} = 10 V_{pp} , f = 1 kHz _____

B. C = 0.08 μF, R = 1 kΩ, V_{in} = 10 V_{pp} , f = 1 kHz _____

Answers

A. $V_{out} = 7.1 V_{pp}$

B. $V_{out} = 4.5 V_{pp}$

Note Hereafter, you can assume that the answer is a peak-to-peak value if the given value is a peak-to-peak value.

12 The output voltage is said to be *attenuated* in the voltage divider calculations, as shown in the calculations in problems 10 and 11. Compare the input and output voltages in problems 10 and 11.

Question

What does *attenuated* mean? _____

Answer

To reduce in amplitude or magnitude (that is, V_{out} is smaller than V_{in}.).

13 When you calculated V_{out} in the examples in problems 10 and 11, you first had to find X_C. However, X_C changes as the frequency changes, while the resistance remains constant. Therefore, as the frequency changes, the impedance Z changes and also so does the amplitude of the output voltage V_{out}.

If V_{out} is plotted against frequency on a graph, the curve looks like that shown in Figure 6.7 .

Figure 6.7

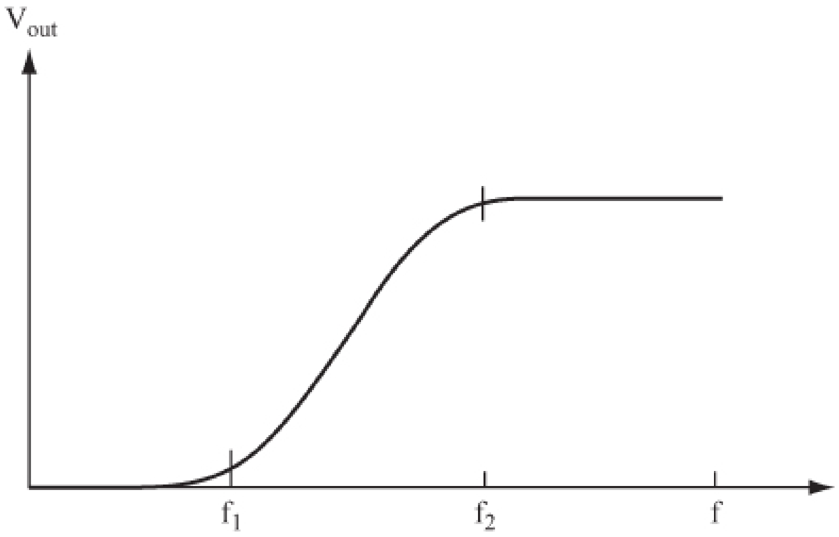

The frequencies of f_1 (at which the curve starts to rise) and f_2 (where it starts to level off) depend on the values of the capacitor and the resistor.

Questions

Calculate the output voltage for the circuit shown in Figure 6.8 for frequencies of 100 Hz, 1 kHz, 10 kHz, and 100 kHz.

Figure 6.8

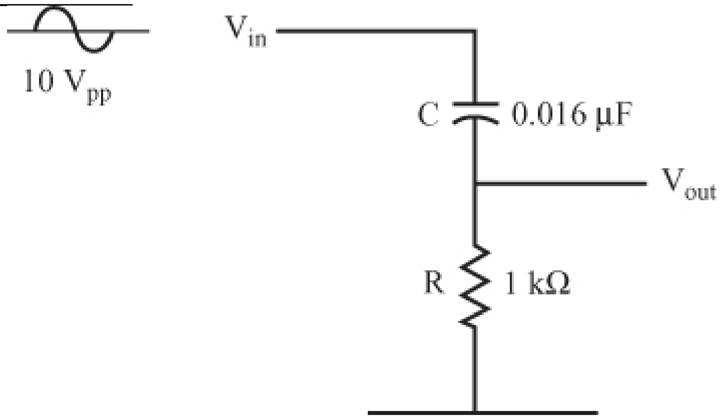

A. 100 Hz: _____

B. 1 kHz: _____

C. 10 kHz: _____

D. 100 kHz: _____

E. Plot these values for V out against f, and draw a curve to fit the points. Use a separate sheet of paper to draw your graph.

Answers

A. V out = 0.1 volt

B. V out = 1 volt

C. V out = 7.1 volts

D. V out = 10 volts

E. The curve is shown in Figure 6.9 .

Figure 6.9

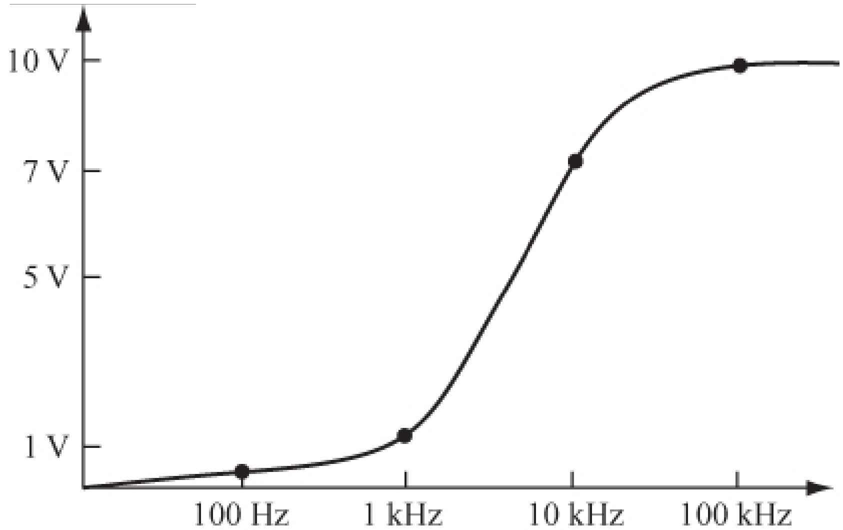

(Note that this is a logarithmic frequency scale.)

Note You can see that V out is equal to V in for the highest frequency and at nearly zero for the lowest frequency. You call this type of circuit a *high-pass filter* because it will pass high-frequency signals with little attenuation and block low-frequency signals.

Project 6.1: The High-Pass Filter

Objective

The objective of this project is to determine how V_{out} changes as the frequency of the input signal changes for a high-pass filter.

General Instructions

When the circuit is set up, measure V_{out} for each frequency; you will also calculate X_C for each frequency value to show the relationship between the output voltage and the reactance of the capacitor.

Parts List

You need the following equipment and supplies:

- One 1 kΩ, 0.25-watt resistor.
- One 0.016 μF capacitor. (You'll probably find 0.016 μF capacitors listed as polypropylene film capacitors. A *polypropylene film capacitor* is made with different material than the more typical ceramic capacitor but performs the same function. If your supplier doesn't carry 0.016 μF capacitors, you can use the closest value the supplier carries. Your results will be changed slightly but will show the same effect.)
- One function generator.
- One oscilloscope. (You can substitute a multimeter and measure V_{out} in rms voltage rather than peak-to-peak voltage.)

One breadboard.

Step-by-Step Instructions

Set up the circuit shown in Figure 6.10. If you have some experience in building circuits, this schematic (along with the previous parts list) should provide all the information you need to build the circuit. If you need a bit more help building the circuit, look at the photos of the completed circuit in the "Expected Results" section.

Figure 6.10

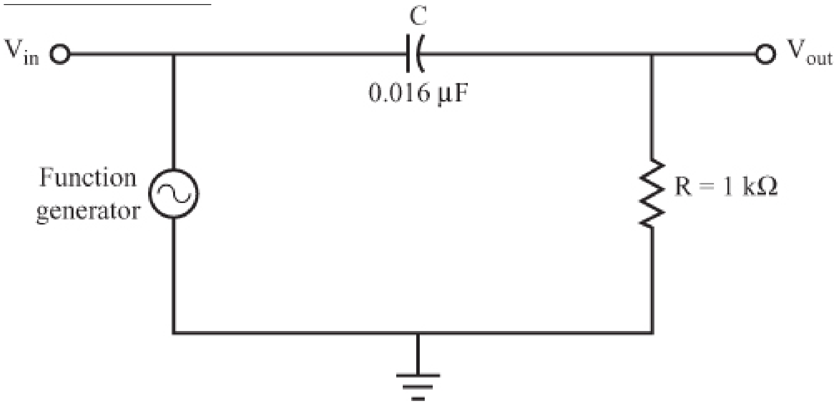

Carefully check your circuit against the diagram.

After you have checked your circuit, follow these steps, and record your measurements in the blank table following the steps:

1. Connect the oscilloscope probe for channel 2 to a jumper wire connected to V_{in}, and connect the ground clip to a jumper wire attached to the ground bus.
2. Connect the oscilloscope probe for channel 1 to a jumper wire connected to V_{out}, and

connect the ground clip to a jumper wire attached to the ground bus.

3. Set the function generator to generate a 10 V_{pp}, 25 Hz sine wave.
4. Measure and record V_{out}.
5. Adjust the function generator to the frequency shown in the next row of the table.
6. Measure and record V_{out}.
7. Repeat steps 5 and 6 until you have recorded V_{out} for the last row of the table.
8. Calculate the values of X_C for each row and enter them in the table.

f_{in}	X_C	V_{out}
25 Hz		
50 Hz		
100 Hz		
250 Hz		
500 Hz		
1 kHz		
3 kHz		
5 kHz		
7 kHz		
10 kHz		
20 kHz		
30 kHz		
50 kHz		
100 kHz		
200 kHz		
500 kHz		
1 MHz		

9. In the blank graph shown in Figure 6.11,

plot V_{out} versus f_{in} with the voltage on the vertical axis and the frequency on the X axis. The curve should have the same shape as the curve shown in Figure 6.8, but don't worry if your curve is shifted slightly to the right or left.

Figure 6.11

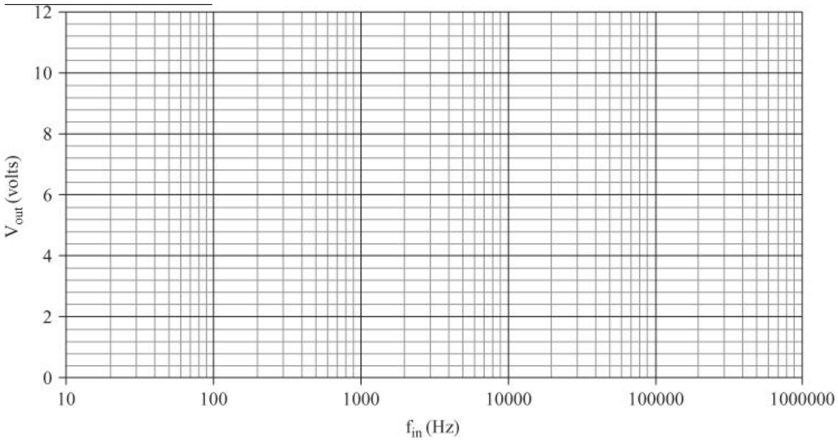

Expected Results

Figure 6.12 shows the breadboarded circuit for this project.

Figure 6.12

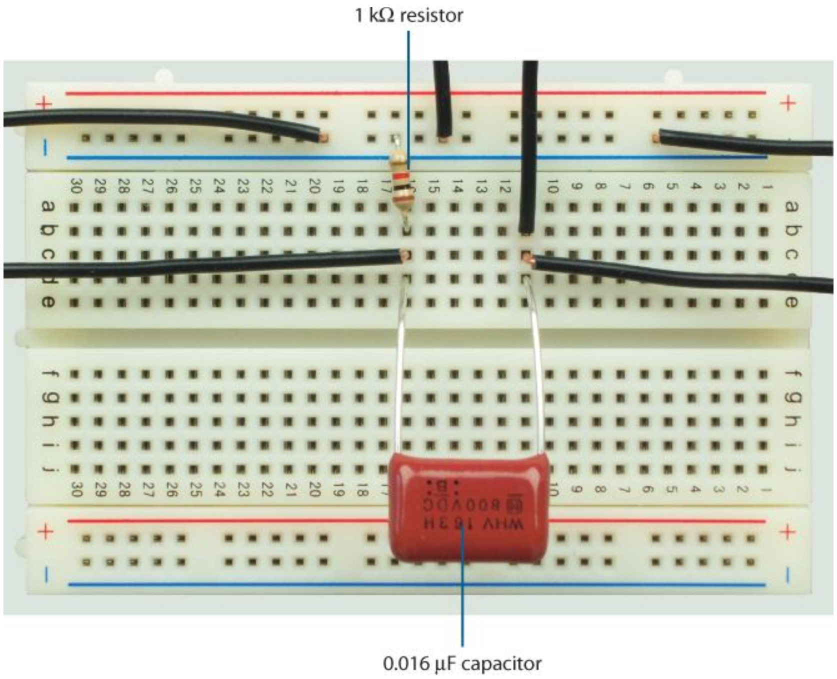
1 kΩ resistor
0.016 µF capacitor

Figure 6.13 shows a function generator and oscilloscope attached to the circuit.

Figure 6.13

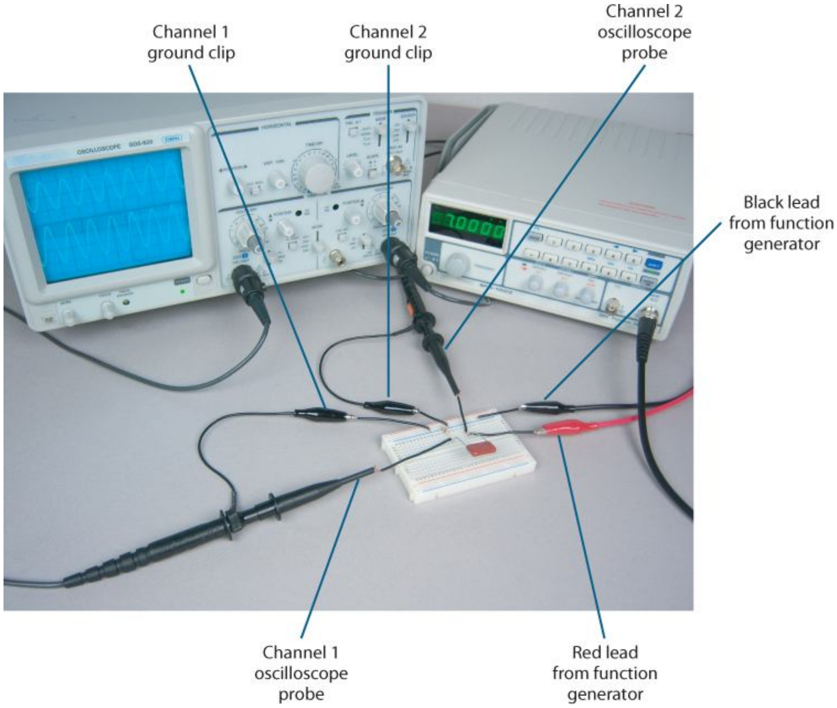
Channel 1 ground clip
Channel 2 ground clip
Channel 2 oscilloscope probe
Black lead from function generator
Channel 1 oscilloscope probe
Red lead from function generator

The input signal is represented by the upper sine wave shown in Figure 6.14, and the output signal is represented by the lower sine wave.

Figure 6.14

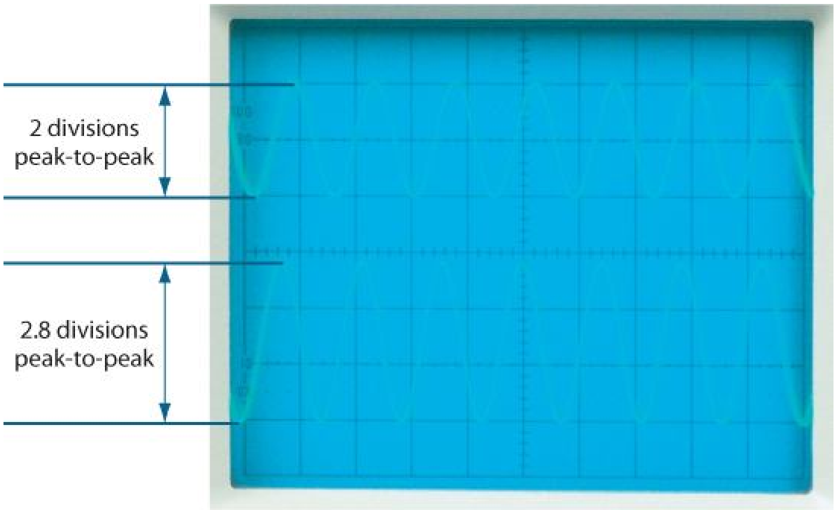
2 divisions peak-to-peak
2.8 divisions peak-to-peak

As you change f_{in}, you may need to adjust the TIME/DIV, VOLTS/DIV, and vertical POSITION controls. The controls shown in Figure 6.15 are adjusted to measure V_{out} when f_{in} = 7 kHz.

Figure 6.15

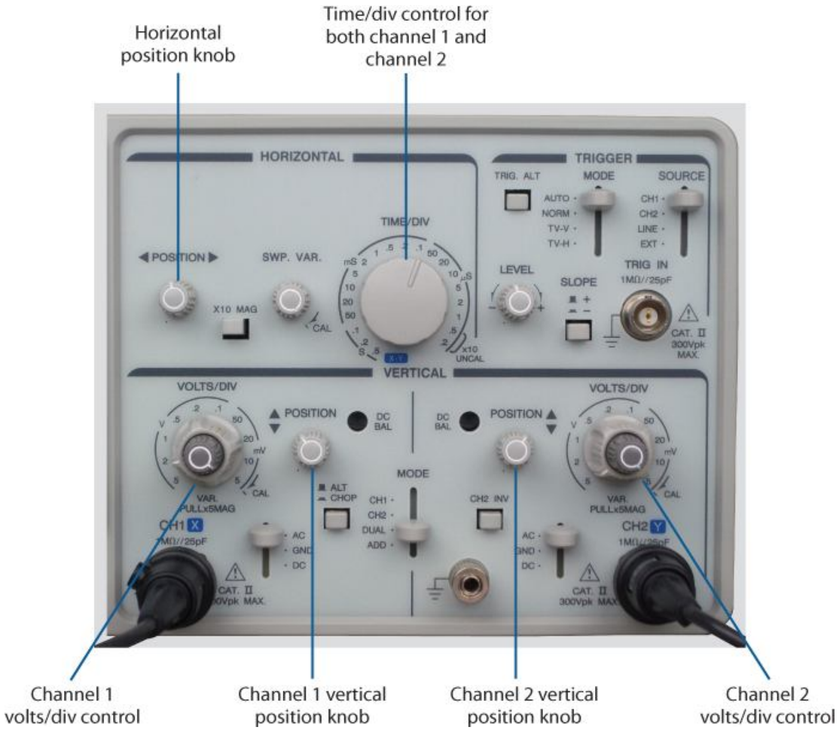
Horizontal position knob
Time/div control for both channel 1 and channel 2
HORIZONTAL
TRIGGER
TIME/DIV
VERTICAL
VOLTS/DIV
Channel 1 volts/div control
Channel 1 vertical position knob
Channel 2 vertical position knob
Channel 2 volts/div control

Your values should be close to those shown in the following table, and the curve should be similar to Figure 6.16 :

f in	X C	V out
25 Hz	400 kΩ	0.025 volts
50 Hz	200 kΩ	0.05 volts
100 Hz	100 kΩ	0.1 volts
250 Hz	40 kΩ	0.25 volts
500 Hz	20 kΩ	0.5 volts
1 kHz	10 kΩ	1 volts
3 kHz	3.3 kΩ	2.9 volts
5 kHz	2 kΩ	4.5 volts
7 kHz	1.4 kΩ	5.6 volts
10 kHz	1 kΩ	7.1 volts

20 kHz	500 Ω	8.9 volts
30 kHz	330 Ω	9.5 volts
50 kHz	200 Ω	9.8 volts
100 kHz	100 Ω	10 volts
200 kHz	50 Ω	10 volts
500 kHz	20 Ω	10 volts
1 MHz	10 Ω	10 volts

Figure 6.16

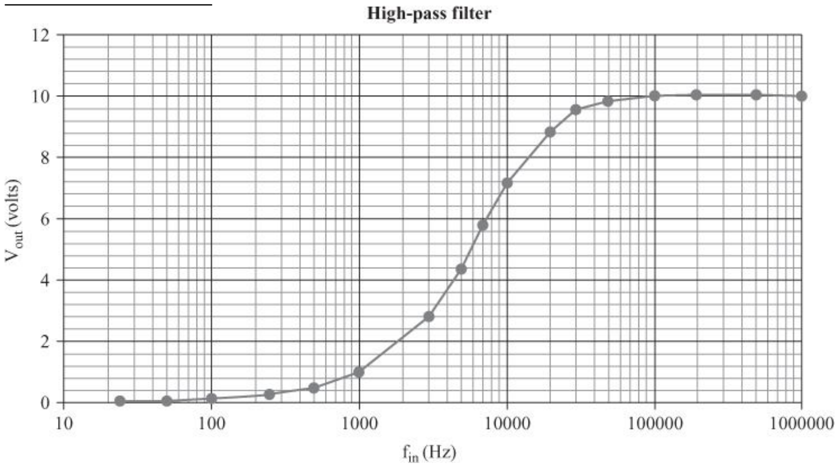

Notice the relationship between X_C and V_{out} in this circuit. Low values of V_{out} and the voltage drop across the resistor in this circuit occur at frequencies for which X_C is high. When X_C is high, more voltage is dropped across the capacitor, and less voltage is dropped across the resistor. (Remember that X_C changes with frequency, while the value of the resistor stays constant.) Similarly, when X_C is low, less voltage is dropped across the capacitor, and more voltage is dropped across the resistor, resulting in a higher V_{out} .

14 Refer to the curve you drew in Project 6.1 for the following question.

Question

What would cause your curve to be moved slightly to the right or the left of the curve shown in Figure 6.16 ? _____

Answer

Slightly different values for the resistor and capacitor that you used versus the resistor and capacitor used to produce the curve in Figure 6.16 . Variations in resistor and capacitor values are to be expected, given the tolerance allowed for standard components.

15 The circuit shown in Figure 6.17 is used in many electronic devices.

Figure 6.17

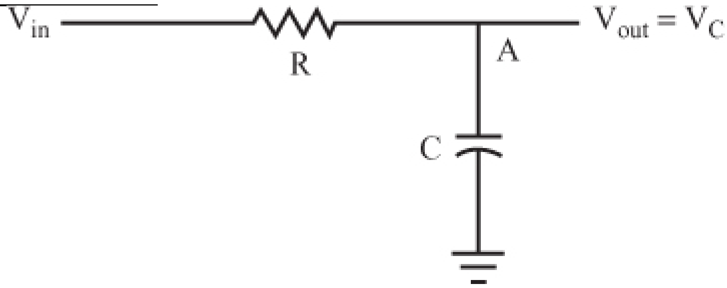

For this circuit, you measure the output voltage across the capacitor instead of across the resistor (between point A and ground).

The impedance of this circuit is the same as that of the circuit used in the last few problems. It still behaves like a voltage divider, and you can calculate the output voltage with an equation similar to the one you used for the high-pass filter circuit discussed in the last few problems. However, by switching the positions of the resistor and capacitor to create the circuit shown in Figure 6.17 , you switch which frequencies will be attenuated, and which will not be attenuated, making the new circuit a low-pass filter, whose characteristics you explore in the next few problems.

Questions

A. What is the impedance formula for the circuit? _____

B. What is the formula for the output voltage? _____

Answers

A. $Z = \sqrt{X_C^2 + R^2}$

B. $V_{out} = V_{in} \times \frac{X_C}{Z}$

16 Refer to the circuit shown in Figure 6.17 and the following values:

$$V_{in} = 10\,V_{pp},\ f = 2\,kHz$$

$$C = 0.1\mu F,\ R = 1k\Omega$$

Questions

Find the following:

A. X_C : _____

B. Z: _____

C. V_{out} : _____

Answers

A. 795 ohms

B. 1277 ohms

C. 6.24 volts

17 Again, refer to the circuit shown in Figure 6.17 to answer the following question.

Question

Calculate the voltage across the resistor using the values given in problem 16, along with the calculated impedance value. _____

Answer

$$V_R = V_{in} \quad \frac{R}{Z} = 10 \quad \frac{1000}{1277} = 7.83V_{pp}$$

18 Use the information from problems 16 and 17 to answer the following question.

Question

What is the formula to calculate V in using the voltages across the capacitor and the resistor? _____

Answer

The formula is $V_{in}^2 = V_C^2 + V_R^2$.

19 V out of the circuit shown in Figure 6.17 changes as the frequency of the input signal changes. Figure 6.18 shows the graph of V out versus frequency for this circuit.

Figure 6.18

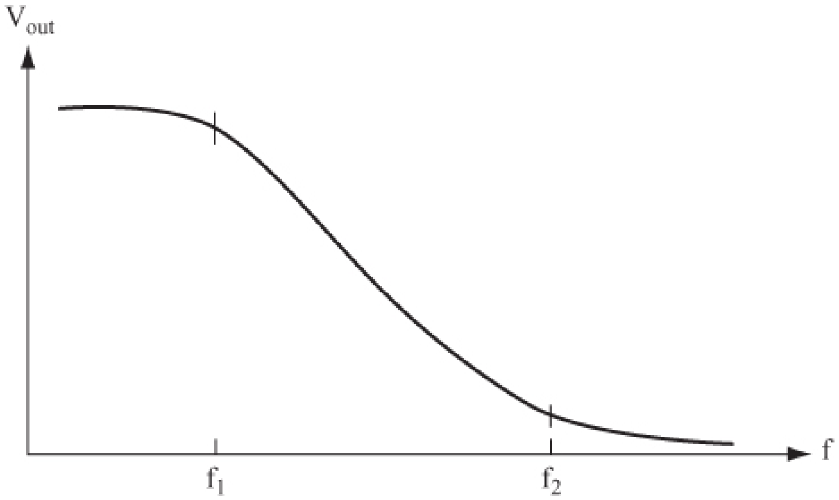

Question

What parameters determine f_1 and f_2?

—————

Answer

The values of the capacitor and the resistor

Note You can see in Figure 6.10 that V_{out} is large for the lowest frequency and nearly zero for the highest frequency. This type of circuit is called a *low-pass filter* because it will pass low frequency signals with little attenuation, while blocking high-frequency signals.

Project 6.2: The Low-Pass Filter

Objective

The objective of this project is to determine how V_{out} changes as the frequency of the input signal changes for a low-pass filter.

General Instructions

After the circuit is set up, measure V_{out} for each frequency. You also calculate X_C for each frequency value to show the relationship between the output voltage and the reactance of the capacitor.

Parts List

You need the following equipment and supplies:

- One 1 kΩ, 0.25-watt resistor. (You can use the same resistor that you used in Project 6.1.)
- One 0.016 μF capacitor. (You can use the same capacitor that you used in Project 6.1.)
- One function generator.
- One oscilloscope. (You can substitute a multimeter and measure V out in rms voltage rather than peak-to-peak voltage.)
- One breadboard.

Step-by-Step Instructions

Set up the circuit shown in Figure 6.19 . If you have some experience in building circuits, this schematic (along with the previous parts list) should provide all the information you need to build the circuit. If you need a bit more help building the circuit, look at the photos of the completed circuit in the "Expected Results" section.

Figure 6.19

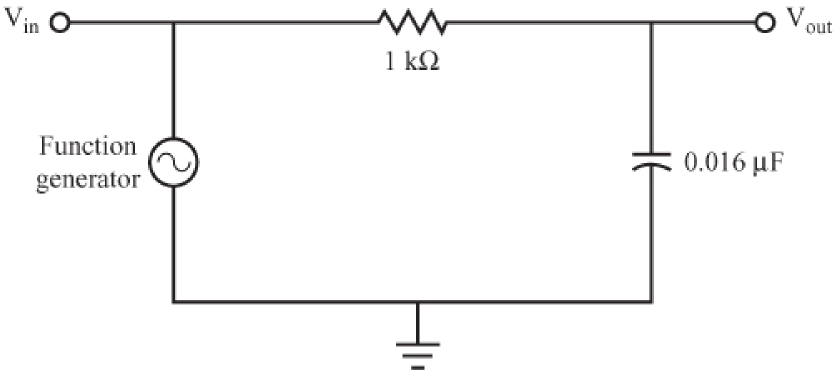

Carefully check your circuit against the diagram.

After you have checked your circuit, follow these steps, and record your measurements in the blank table following the steps:

1. Connect the oscilloscope probe for channel 2 to a jumper wire connected to V_{in}, and connect the ground clip to a jumper wire attached to the ground bus.
2. Connect the oscilloscope probe for channel 1 to a jumper wire connected to V_{out}, and connect the ground clip to a jumper wire attached to the ground bus.
3. Set the function generator to generate a 10 V_{pp}, 25 Hz sine wave.
4. Measure and record V_{out}.
5. Adjust the function generator to the frequency shown in the next row of the table.
6. Measure and record V_{out}.
7. Repeat steps 5 and 6 until you have recorded V_{out} for the last row of the table.
8. Enter the values of X_C for each row in the table. (Because you used the same

capacitor and resistor in Project 6.1, you can take the values X_C from the table in Project 6.1.)

f_{in}	X_C	V_{OUT}
25 Hz		
50 Hz		
100 Hz		
250 Hz		
500 Hz		
1 kHz		
3 kHz		
5 kHz		
7 kHz		
10 kHz		
20 kHz		
30 kHz		
50 kHz		
100 kHz		
200 kHz		
500 kHz		
1 MHz		

9. In the blank graph shown in Figure 6.20, plot V_{out} versus f_{in} with the voltage on the vertical axis and the frequency on the X axis. The curve should have the same shape as the curve shown in Figure 6.18.

Figure 6.20

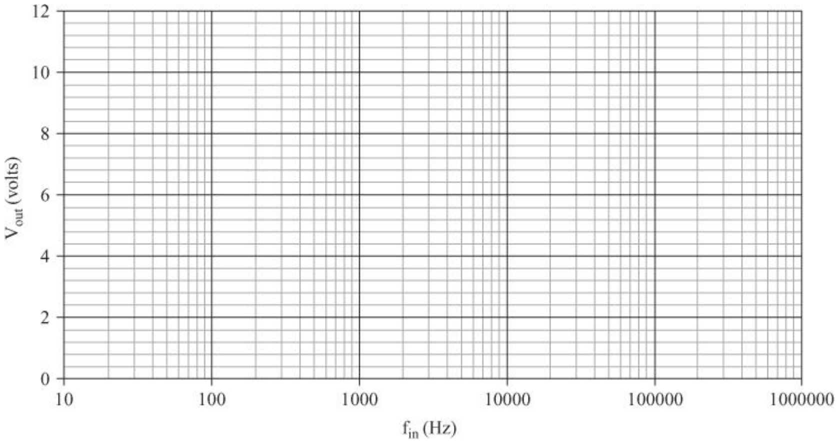

Expected Results

Figure 6.21 shows the breadboarded circuit for this project.

Figure 6.21

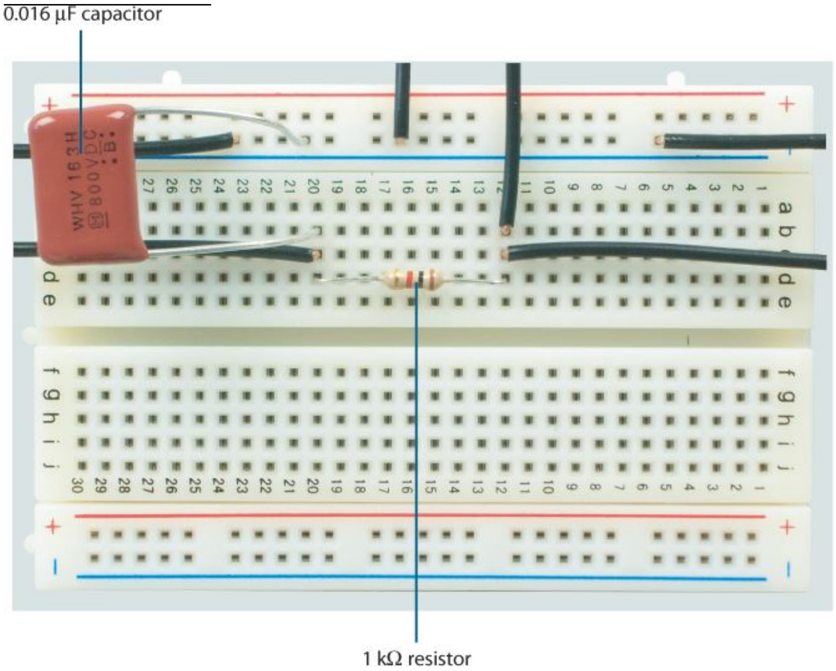
0.016 µF capacitor
1 kΩ resistor

Figure 6.22 shows a function generator and oscilloscope attached to the circuit.

Figure 6.22

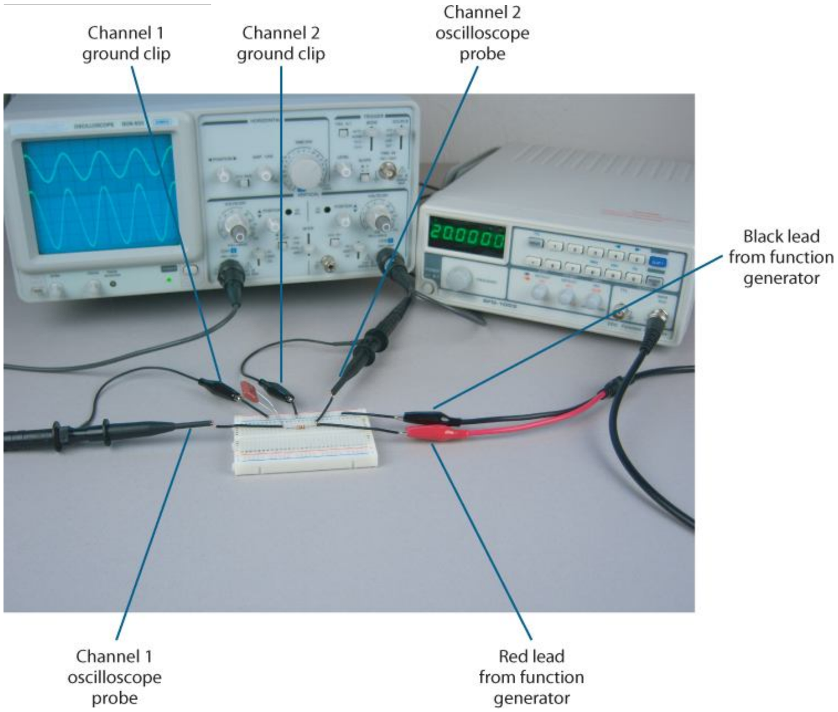
Channel 1 ground clip
Channel 2 ground clip
Channel 2 oscilloscope probe
Black lead from function generator
Channel 1 oscilloscope probe
Red lead from function generator

The input signal is represented by the upper sine wave, as shown in Figure 6.23, and the output signal is represented by the lower sine wave. Reading the number of divisions for the peak-to-peak output sine wave and multiplying it by the corresponding VOLTS/DIV setting allows to you measure V out .

Figure 6.23

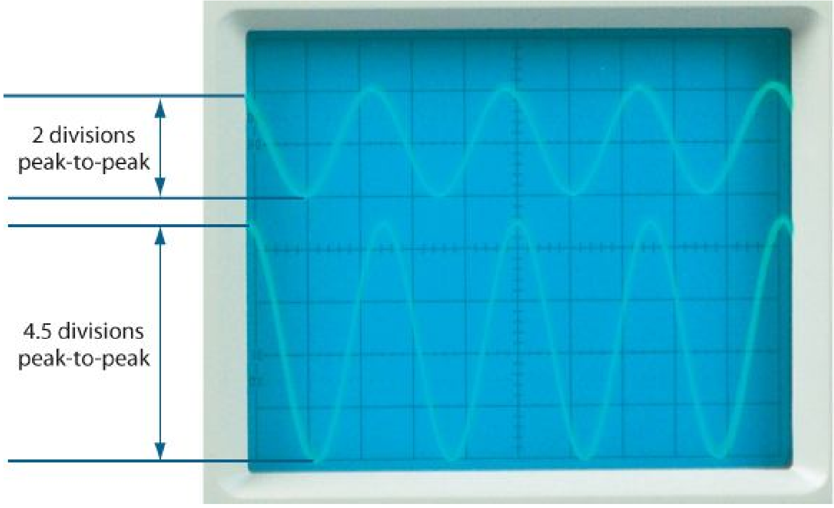
2 divisions
peak-to-peak
4.5 divisions
peak-to-peak

As you change f_{in} adjustments in the TIME/DIV control, the VOLTS/DIV and vertical POSITION controls for channel 1 may be needed. The controls shown in Figure 6.24 are adjusted to measure V_{out} when f_{in} = 20 kHz.

Figure 6.24

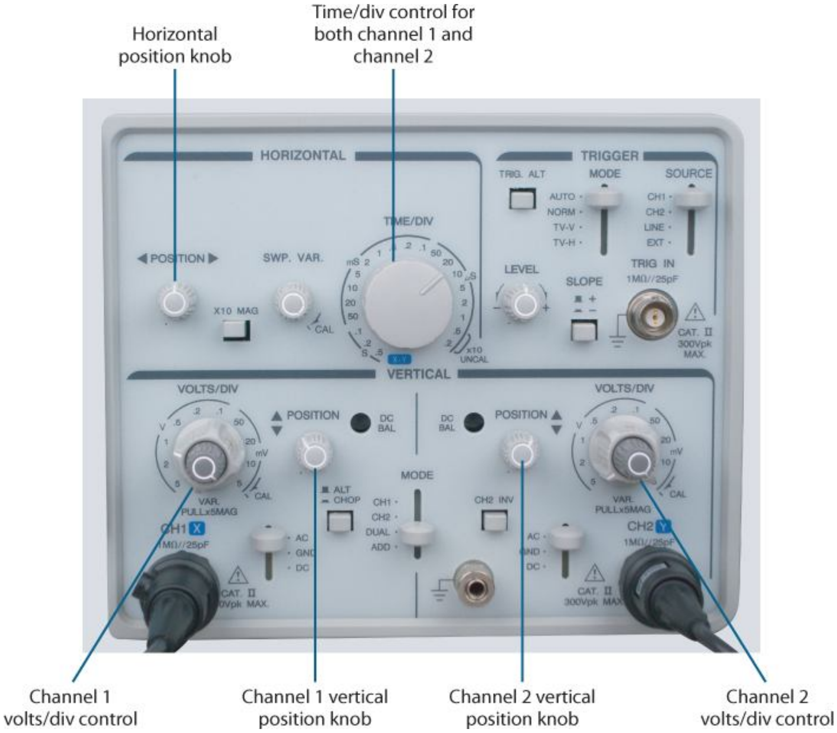
Horizontal position knob
Time/div control for both channel 1 and channel 2
Channel 1 volts/div control
Channel 1 vertical position knob
Channel 2 vertical position knob
Channel 2 volts/div control

Your values should be close to those shown in the following table, and the curve should be similar to Figure 6.25.

f_{in}	X_C	V_{out}
25 Hz	400 kΩ	10 volts
50 Hz	200 kΩ	10 volts
100 Hz	100 kΩ	10 volts
250 Hz	40 kΩ	10 volts
500 Hz	20 kΩ	10 volts
1 kHz	10 kΩ	10 volts
3 kHz	3.3 kΩ	9.4 volts
5 kHz	2 kΩ	9.1 volts
7 kHz	1.4 kΩ	8.2 volts
10 kHz	1 kΩ	7.1 volts

20 kHz	500 Ω	4.5 volts
30 kHz	330 Ω	2.9 volts
50 kHz	200 Ω	2.0 volts
100 kHz	100 Ω	1 volt
200 kHz	50 Ω	0.5 volt
500 kHz	20 Ω	0.2 volt
1 MHz	10 Ω	0.1 volt

Figure 6.25

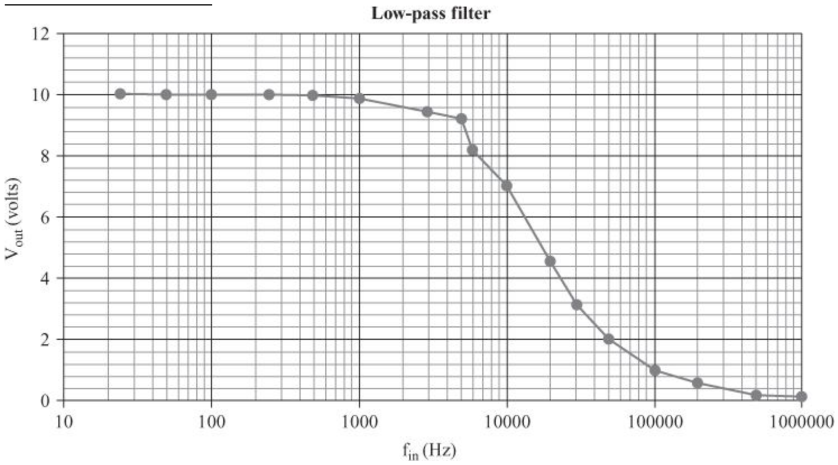

Notice the relationship between X_C and V_{out} in this circuit. Low values of V_{out} (The voltage drop across the capacitor in this circuit.) occur at frequencies for which X_C is also low. When X_C is low, more voltage is dropped across the resistor and less across the capacitor. (Remember that X_C changes with frequency, whereas the value of the resistor stays constant.) Similarly, when X_C is high, less voltage is dropped across the resistor, and more voltage is dropped across the capacitor, resulting in a higher V_{out}.

Phase Shift of an RC Circuit

20 In both of the circuits shown in Figure 6.26 , the output voltage is different from the input voltage.

Figure 6.26

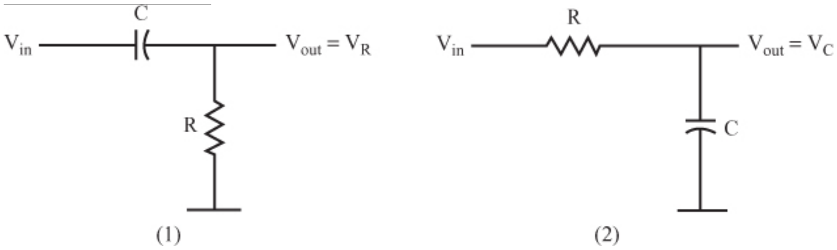

Question

In what ways do they differ? _____

Answer

The signal is attenuated, or reduced. The amount of attenuation depends upon the frequency of the signal. Circuit 1 will pass high-frequency signals while blocking low-frequency signals. Circuit 2 will pass low-frequency signals while blocking high-frequency signals.

21 The voltage is also changed in another way. The voltage across a capacitor rises and falls at the same frequency as the input signal, but it does not reach its peak at the same time, nor does it pass through zero at the same time. You can see this when you compare the V out curves to the V in curves in Figure 6.27 .

Note The numbered graphs in Figure 6.27 are produced by the corresponding numbered circuits in Figure 6.26 .

Figure 6.27

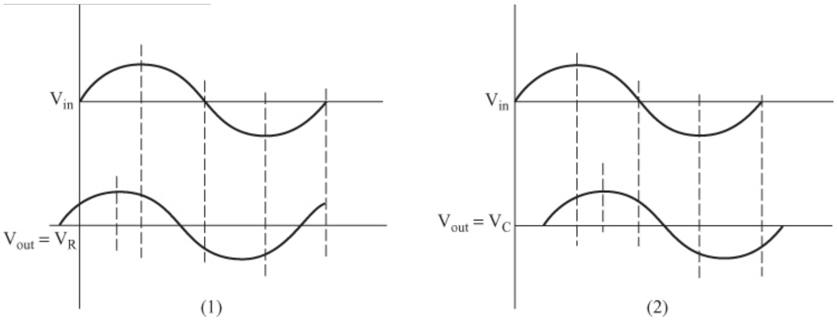

Questions

A. Examine graph (1). Is the output voltage peak displaced to the right or the left? _____

B. Examine graph (2). Is the output voltage peak displaced to the right or the left? _____

Answers

A. To the left

B. To the right

22 The output voltage waveform in graph (1) of Figure 6.27 is said to *lead the input voltage waveform* . The output waveform in graph (2) is said to *lag the input waveform* . The amount that V out leads or lags V in is measured in degrees. There are 90 degrees between the peak of a sine wave and a point at which the sine wave crosses zero volts. You can use this information to estimate the number of degrees V out is leading or lagging V in . The difference between these two waveforms is called a *phase shift* or *phase difference* .

Questions

A. What is the approximate phase shift of the two waveforms shown in the graphs? _____

B. Do you think that the phase shift depends on the value of frequency? _____

C. Will an RC voltage divider with the voltage taken across the capacitor produce a lead or a lag in the phase shift of the output voltage?

_____

Answers

A. Approximately 35 degrees.

B. It does depend upon frequency because the values of the reactance and impedance depend upon frequency.

C. A lag as shown in graph (2).

23 The current through a capacitor is out of phase with the voltage across the capacitor. The current leads the voltage by 90 degrees. The current and voltage across a resistor are in phase. (That is, they have no phase difference.)

Figure 6.28 shows the vector diagram for a series RC circuit. θ is the phase angle by which V R leads V in . ϕ is the phase angle by which V C lags V in .

Figure 6.28

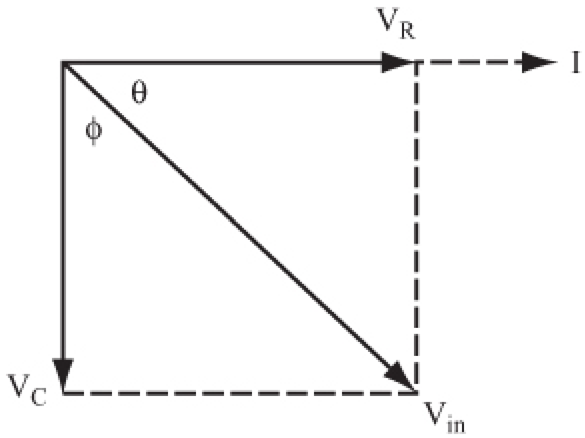

Note Although the voltage across a resistor is in phase with the current through the resistor, both are out of phase with the applied voltage.

You can calculate the phase angle by using this formula:

$$\tan\theta = \frac{V_C}{V_R} = \frac{1}{2\pi fRC} = \frac{X_C}{R}$$

As an example, calculate the phase angle when 160 Hz is applied to a 3.9 kΩ resistor in series with a 0.1 μF capacitor.

$$\tan\theta = \frac{1}{2\times\pi\times160\times3.9\times10^{3}\times0.1\times10^{\ 6}} = 2.564$$

You can calculate the inverse tangent of 2.564 on your calculator and find that the phase angle is 68.7 degrees, which means that V_R leads V_{in} by 68.7 degrees. This also means that V_C lags the input by 21.3 degrees.

In electronics, the diagram shown in Figure 6.28 is called a *phasor diagram* , but the mathematics involved are the same as for vector diagrams, with which you should be familiar.

Question

Sketch a phasor diagram using the angles θ and ϕ resulting from the calculations in this problem. Use a separate sheet of paper for your diagram.

Answer

See Figure 6.29. Note that the phasor diagram shows that the magnitude of V_C is greater than V_R.

Figure 6.29

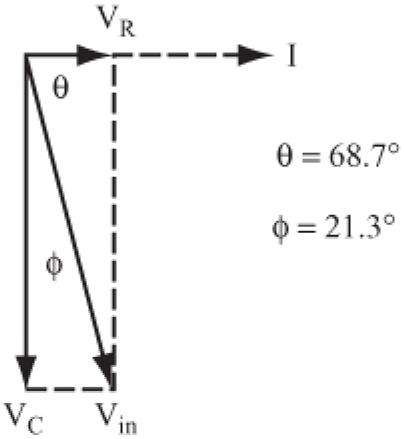

24 Using the component values and input signal shown in Figure 6.30, answer the following questions.

Figure 6.30

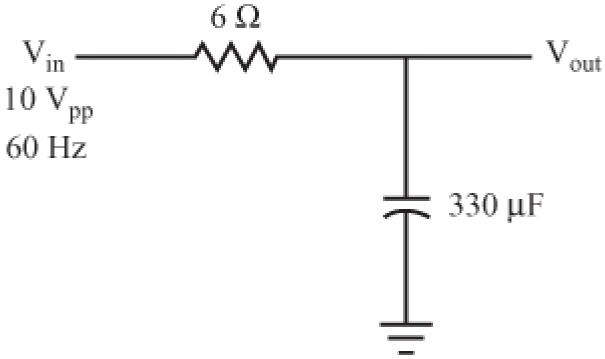

Questions

Find the following:

A. X_C: _____

B. Z: _____

C. V_{out}: _____

D. V_R: _____

E. The current flowing through the circuit: _____

F. The phase angle: _____

Answers

A. $X_C = \frac{1}{2\pi fC} = 8\,\text{ohms}$

B. $Z = \sqrt{8^2 + 6^2} = 10\,\text{ohms}$

C. $V_{out} = V_C = V_{in} \times \frac{X_C}{Z} = 8\text{ volts}$

D. $V_R = V_{in} \times \frac{R}{Z} = 6\text{ volts}$

E. $I = \frac{V}{Z} = \frac{10\,V_{pp}}{10\,\Omega} = 1\text{ ampere}$

F. $\tan\theta = \frac{X_C}{R} = \frac{8\,\Omega}{6\,\Omega} = 1.33.$

Therefore, θ = 53.13 degrees.

25 Use the circuit shown in Figure 6.31 to answer the following questions.

Figure 6.31

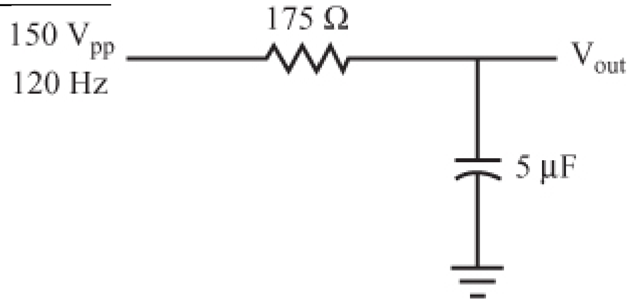

Questions

Calculate the following parameters:

A. X_C: _____

B. Z: _____

C. V_{out}: _____

D. V_R: _____

E. The current flowing through the circuit: _____

F. The phase angle: _____

Answers

A. X_C = 265 ohms

B. $Z = \sqrt{175^2 + 265^2} = 317.57\,\Omega$

C. V_C = 125 volts

D. V_R = 83 volts

E. I = 0.472 ampere

F. $\tan\theta = \frac{265\,\Omega}{175\,\Omega} = 1.5$

Therefore, θ = 56.56 degrees.

Resistor and Capacitor in Parallel

26 The circuit shown in Figure 6.32 is a common variation on the low-pass filter circuit introduced in problem 15.

Figure 6.32

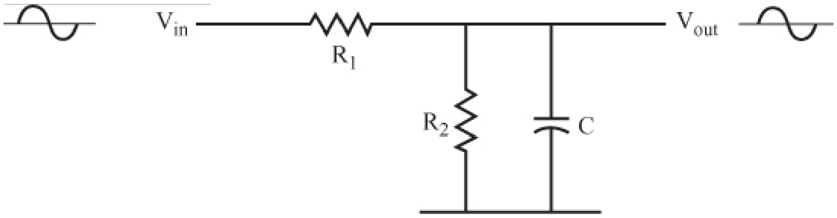

Because a DC signal will not pass through the capacitor, this circuit functions like the circuit shown in Figure 6.33 for DC input signals.

Figure 6.33

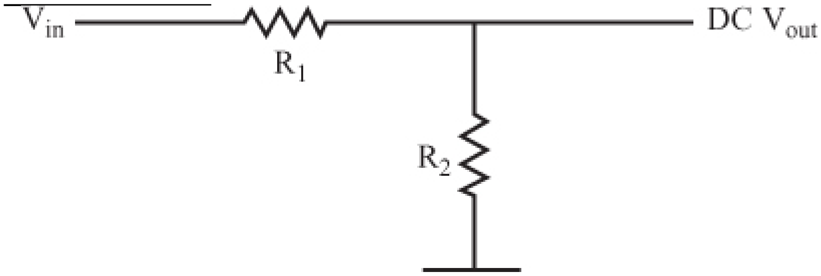

An AC signal will pass through both the capacitor and R_2. You can treat the circuit as if it had a resistor with a value of r (where r is the parallel equivalent of R_2 and X_C) in place of the parallel capacitor and resistor. This is shown in Figure 6.34.

Figure 6.34

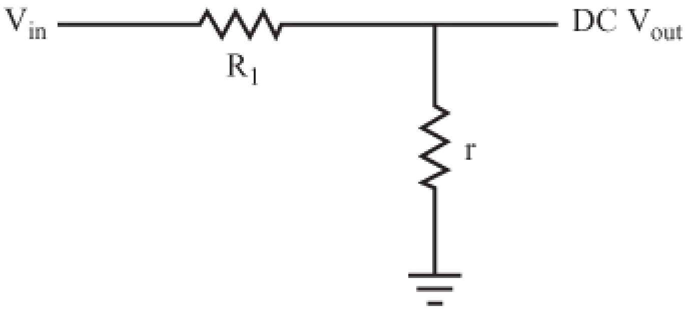

Calculating the exact parallel equivalent (r) is complicated and beyond the scope of this book. However, to demonstrate the usefulness of this circuit, you can make a major simplification. Consider a circuit where X_C is only about one-tenth the value of R_2 or less. This circuit has many practical applications, because it attenuates the AC and the DC differently.

The following example can help to clarify this. For the following circuit, calculate the AC and DC output voltages separately.

For the circuit shown in Figure 6.35 , you can calculate the AC and DC output voltages separately by following the steps outlined in the following questions.

Figure 6.35

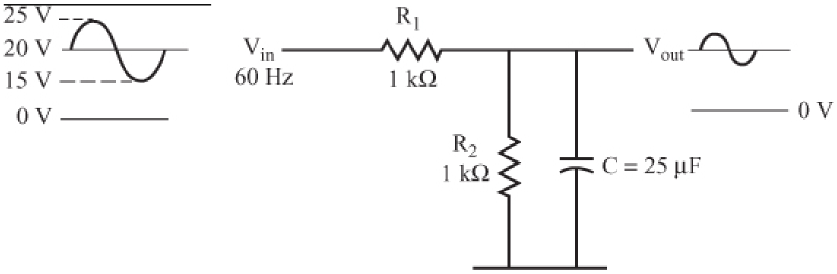

Questions

A. Find X_C. Check that it is less than one-tenth of R_2.

$$X_C =$$

B. For the circuit in Figure 6.35, determine through which circuit components DC signals will flow. Then use the voltage divider formula to find DC V_{out}.

$$\text{DC } V_{out} =$$

C. For the circuit in Figure 6.35 determine which circuit components AC signals will flow through. Then use the voltage divider formula to find AC V_{out}.

$$\text{AC } V_{out} =$$

D. Compare the AC and DC input and output voltages. _____

Answers

A. X_C = 106 ohms and R_2 = 1000 ohms, so X_C is close enough to one-tenth of R_2.

B. Figure 6.36 shows the portion of the circuit that a DC signal passes through.

$$V_{out} = 20 \times \frac{1\text{k}\Omega}{1\text{k}\Omega + 1\text{k}\Omega} = 10 \text{ volts}$$

Figure 6.36

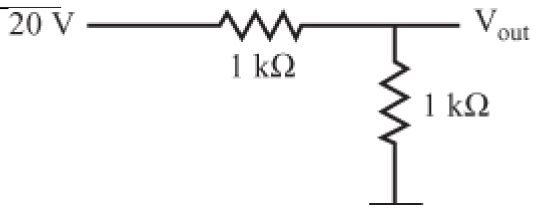

C. Figure 6.37 shows the portion of the circuit that an AC signal passes through.

$$V_{out} = 10 \times \frac{106}{\sqrt{(1000)^2 + (106)^2}} = 1.05 \text{ volts}$$

Figure 6.37

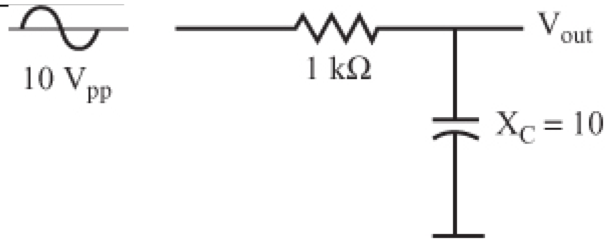

D. Figure 6.38 shows the input waveform on the left and the output waveform on the right. You can see from the waveforms that the DC voltage has dropped from 20 volts to 10 volts and that the AC voltage has dropped from 10 volts to 1.05 volts.

Figure 6.38

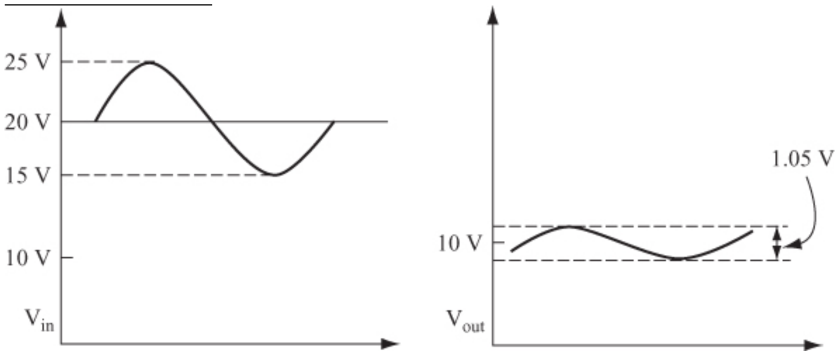

27 Figure 6.39 shows two versions of the circuit discussed in problem 26 with changes to the value of the capacitor or the frequency of the input signal. The DC input voltage is 20 volts, and the AC input voltage is 10 V pp. Use the same steps shown in problem 26 to find and compare the output voltages with the input voltages for the two circuits shown in Figure 6.39 .

Figure 6.39

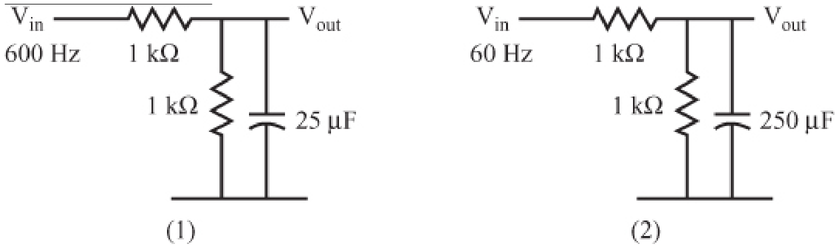
V_{in}
600 Hz
1 kΩ
1 kΩ
25 µF
V_{out}
(1)
V_{in}
60 Hz
1 kΩ
1 kΩ
250 µF
V_{out}
(2)

Questions

1.

A. X_C = _____

B. DC V_{out} = _____

C. AC V_{out} = _____

D. Attenuation: _____

2.

A. X_C = _____

B. DC V_{out} = _____

C. AC V_{out} = _____

D. Attenuation: _____

Answers

1.

A. X_C = 10.6 ohms.

B. DC V_{out} = 10 volts.

C. AC V_{out} = 0.1 volts.

D. Here, the DC attenuation is the same as the example in problem 26, but the AC output voltage is reduced because of the higher frequency.

2.

A. X_C = 10.6 ohms.

B. DC V_{out} = 10 volts.

C. AC V out = 0.1 volts.

D. The DC attenuation is still the same, but the AC output voltage is reduced because of the larger capacitor.

Inductors in AC Circuits

28 Figure 6.40 shows a voltage divider circuit using an inductor, rather than a capacitor.

Figure 6.40

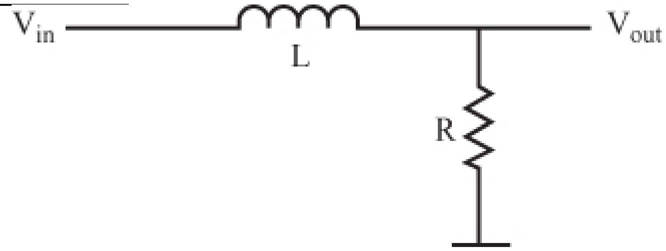

As with previous problems, consider all the inputs to be pure sine waves. Like the capacitor, the inductor cannot change the frequency of a sine wave, but it can reduce the amplitude of the output voltage.

The simple circuit, as shown in Figure 6.40 , opposes current flow.

Questions

A. What is the opposition to current flow called? _____

B. What is the formula for the reactance of the inductor? _____

C. Write out the formula for the opposition to the current flow for this circuit. _____

Answers

A. Impedance

B. $X_L = 2\pi fL$.

C. $Z = \sqrt{X_L^2 + R^2}$

In many cases, the DC resistance of the inductor is low, so assume that it is 0 ohms. For the next two problems, make that assumption in performing your calculations.

29 You can calculate the voltage output for the circuit shown in Figure 6.41 with the voltage divider formula.

Figure 6.41

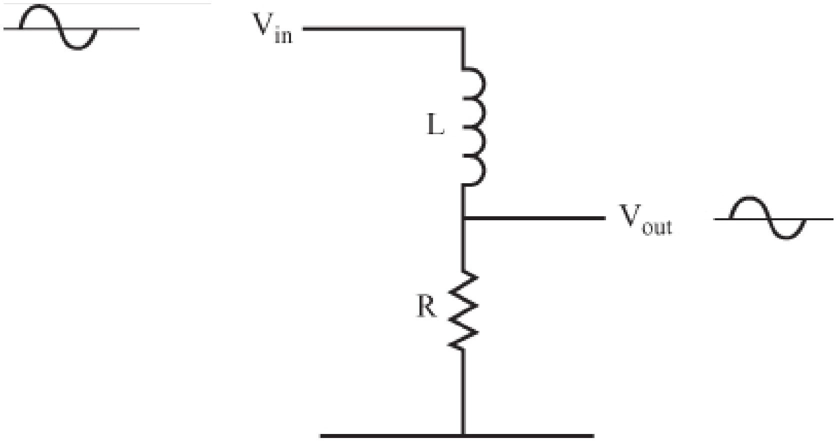

Question

What is the formula for V_{out} ? _____

Answer

$$V_{out} = V_{in} \times \frac{R}{Z}$$

30 Find the output voltage for the circuit shown in Figure 6.42 .

Figure 6.42

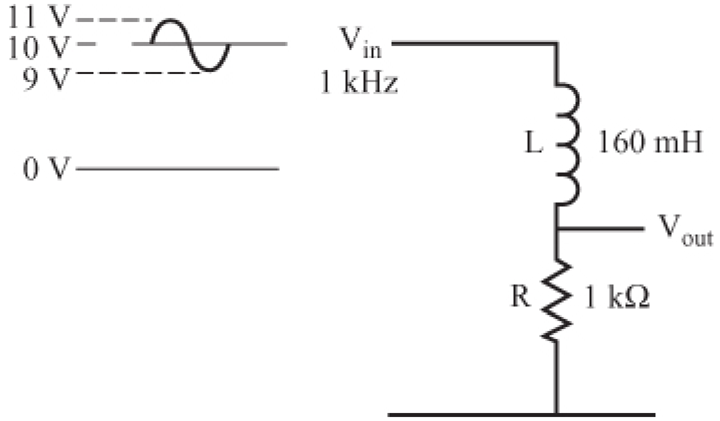

Use the steps in the following questions to perform the calculation.

Questions

A. Find the DC output voltage. Use the DC voltage divider formula.

$$\mathrm{DC}\ V_{out} =$$

B. Find the reactance of the inductor.

$$X_L =$$

C. Find the AC impedance.

$$Z =$$

D. Find the AC output voltage.

$$\mathrm{AC}\ V_{out} =$$

E. Combine the outputs to find the actual output. Draw the output waveform and label the voltage levels of the waveform on the blank graph shown in Figure 6.43 .

Figure 6.43

Answers

A.

$$\text{DC V}_{\text{out}} = 10 \text{ volts} \times \frac{1\text{k}\Omega}{1\text{k}\Omega + 0} = 10 \text{ volts}$$

B. X_L = 1 kΩ (approximately).

C.

$$Z = \sqrt{1^2 + 1^2} = \sqrt{2} = 1.414\,\text{k}\Omega$$

D.

$$\text{AC V}_{\text{out}} = 2\text{V}_{\text{pp}} \times \frac{1\text{k}\Omega}{1.414\,\text{k}\Omega} = 1.414\,\text{V}_{\text{pp}}$$

E. The output waveform is shown in Figure 6.44 .

Figure 6.44

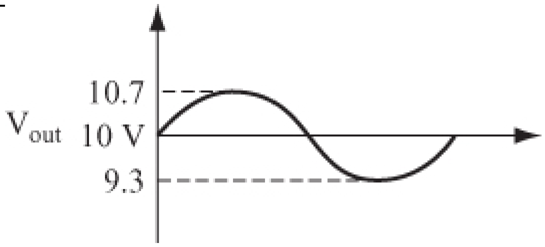

31 For the circuit shown in Figure 6.45 , the DC resistance of the inductor is large enough that you should include that value in your calculations.

Figure 6.45

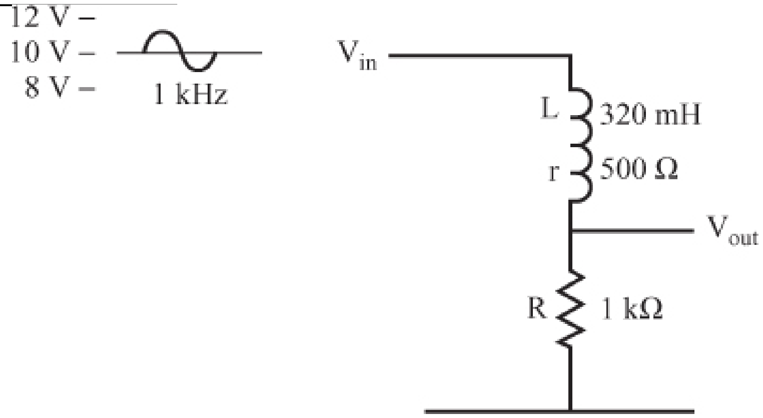

Questions

For the circuit shown in Figure 6.45, calculate the DC and AC output voltages, using the steps listed in problem 30.

A. DC V_{out} = _____

B. X_L = _____

C. Z = _____

D. AC V_{out} = _____

E. Draw the output waveform and label the voltage levels of the waveform on the blank graph in Figure 6.46.

Figure 6.46

Answers

A.

$$\text{DC V}_{out} = 10 \text{ volts} = \frac{1\text{k}\Omega}{(1\text{k}\Omega + 500\,\Omega)} = 6.67 \text{ volts}$$

Note The 500 Ω DC resistance of the inductor has been added to the 1 kΩ resistor value in this calculation.

B. X_L = 2 kΩ,

C. $Z = \sqrt{1.5^2 + 2^2} = 2.5\text{k}\Omega$

Note

The 500 Ω DC resistance of the inductor has been added to the 1 kΩ resistor value in this calculation.

D. AC V_{out} = 1.6 V_{pp},

E. See Figure 6.47.

Figure 6.47

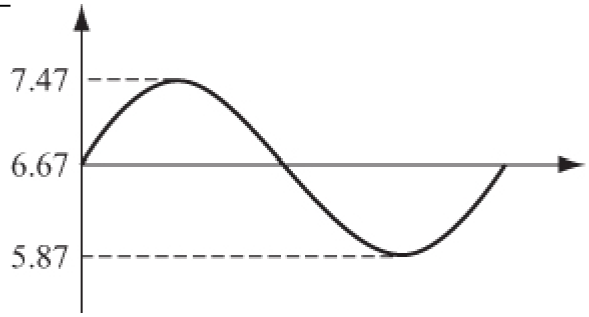

32 To calculate V_{out}, in problems 30 and 31, you also had to calculate X_L. However, because X_L changes with the frequency of the input signal, the impedance and the amplitude of V_{out} also change with the frequency of the input signal. If you plot the output voltage V_{out} against frequency, you should see the curve shown in Figure 6.48.

Figure 6.48

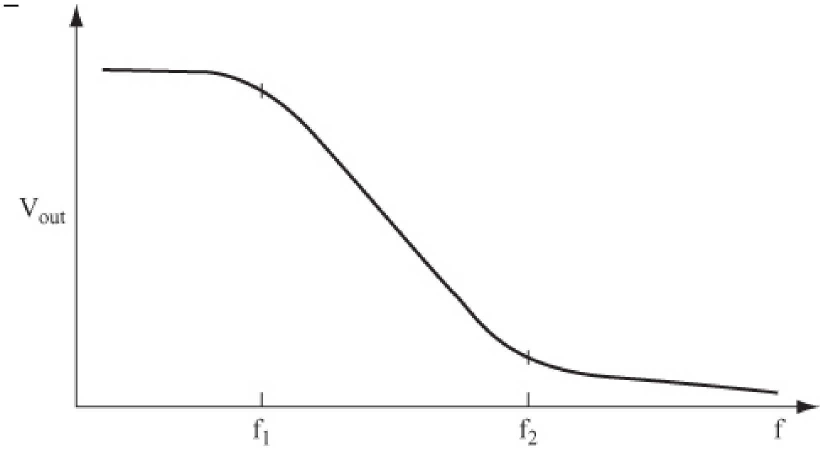

The values of the inductor and resistor determine the frequency at which V out starts to drop (f 1), and the frequency at which V out levels off (f 2).

The curve in Figure 6.48 shows that using an inductor and resistor in a circuit such as the one shown in Figure 6.42 produces a low-pass filter similar to the one discussed in problems 15 through 19.

Question

What values control f 1 and f 2 ? _____

Answer

The values of the inductor and the resistor

33 You can also create a circuit as shown in Figure 6.49 , in which the output voltage is equal to the voltage drop across the inductor.

Figure 6.49

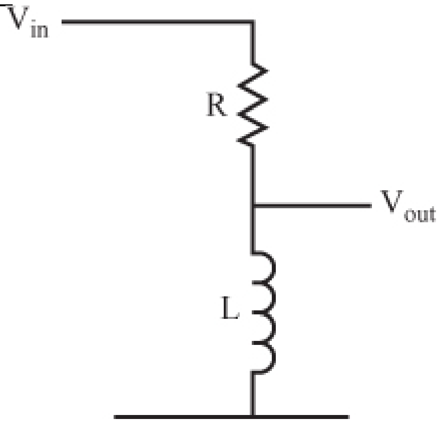

Questions

A. What formula would you use to find V out ? _____

B. If you plot the output voltage versus the frequency, what would you expect the curve to be? Use a separate sheet of paper to draw your answer. _____

Answers

A. $V_{out} = V_{in} \times \frac{X_L}{Z}$

B. See Figure 6.50 .

Figure 6.50

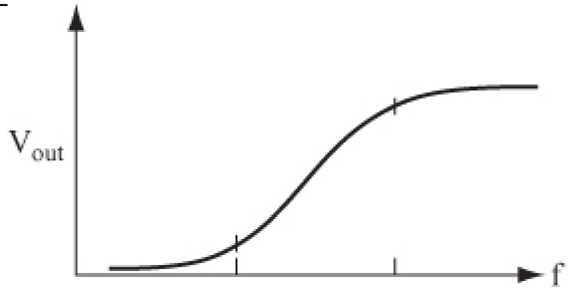

The curve in Figure 6.50 demonstrates that using an inductor and resistor in a circuit, such as the one shown in Figure 6.49 , produces a high-pass filter similar to the one discussed in problems 6 through 13.

Higher-Order Filters

Filter circuits that contain one capacitor or inductor are called *first-order filters* . Filter order numbers reflect the number of capacitors, inductors, or operational amplifiers (a component discussed in Chapter 8, “Transistor Amplifiers”) in the filter. For example, a filter that contains four capacitors is a fourth-order filter, whereas a filter that contains six capacitors is a sixth-order filter.

If you want a sharper drop-off between frequencies, you can connect first-order filters in series. This effect is demonstrated in the following figure.

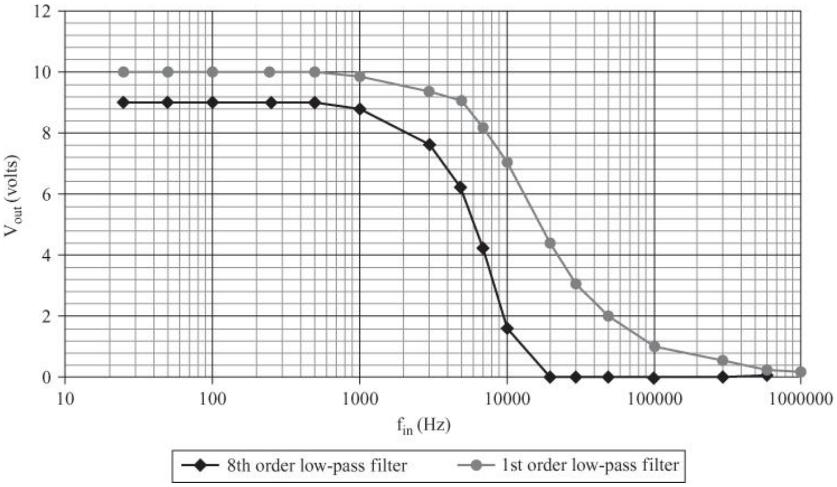

This graph shows how V out changes as f in changes for the first-order low-pass filter used in Project 6.2, and for an eighth-order low-pass filter. V out for the eighth-order filter drops by 80 percent at approximately 10 kHz, whereas V out for the first-order filter doesn't drop by 80 percent until the frequency reaches approximately 50 kHz.

Phase Shift for an RL Circuit

34 Filter circuits that use inductors (such as those shown in Figure 6.51) produce a phase shift in the output signal, just as filter circuits containing capacitors do. You can see the shifts for the circuits shown in Figure 6.51 by comparing the input and output waveforms shown below the circuit diagrams.

Figure 6.51

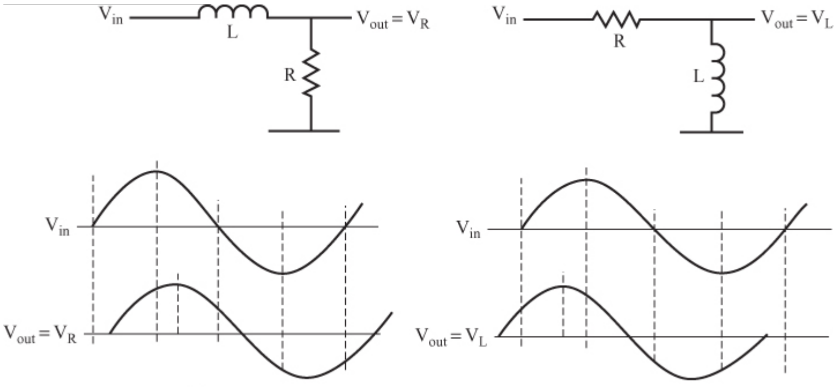
V_{in}
L
$V_{out} = V_R$
R
V_{in}
R
$V_{out} = V_L$
L
V_{in}
$V_{out} = V_R$
V_{in}
$V_{out} = V_L$

(1) (2)

Question

In which circuit does the output voltage lead the input voltage? _____

Answer

In graph (1), the output voltage lags the input voltage, and in graph (2), the output voltage leads.

35 Figure 6.52 shows a vector diagram for both the circuits shown in Figure 6.51 . The current through the inductor lags the voltage across the inductor by 90 degrees.

Figure 6.52

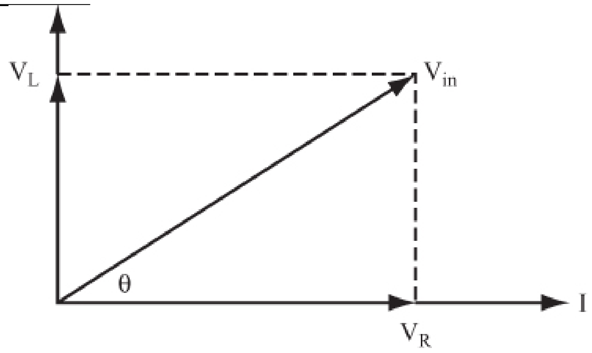

The phase angle is easily found:

$$\tan\theta = \frac{V_L}{V_R} = \frac{X_L}{R} = \frac{2\pi fL}{R}$$

Question

Calculate the phase angle for the circuit discussed in problem 30. _____

Answer

45 degrees

36 Refer to the circuit discussed in problem 31.

Question

Calculate the phase angle. _____

Answer

$$\tan\theta = \frac{X_L}{R} = \frac{2\,k\Omega}{1.5\,k\Omega} = 1.33$$

Therefore, θ = 53.1 degrees.

Summary

This chapter has discussed the uses of capacitors, resistors, and inductors in voltage divider and filter circuits. You learned how to determine the following:

- The output voltage of an AC signal after it passes through a high-pass RC filter circuit
- The output voltage of an AC signal after it passes through a low-pass RC circuit
- The output voltage of an AC signal after it passes through a high-pass RL circuit
- The output voltage of an AC signal after it passes through a low-pass RL circuit

The output waveform of an AC or combined AC-DC signal after it passes through a filter circuit

Simple phase angles and phase differences

Self-Test

These questions test your understanding of this chapter. Use a separate sheet of paper for your calculations. Compare your answers with the answers provided following the test.

For questions 1–3, calculate the following parameters for the circuit shown in each question.

A. X_C

B. Z

C. V_{out}

D. I

E. tan θ and θ

1. Use the circuit shown in Figure 6.53.

Figure 6.53

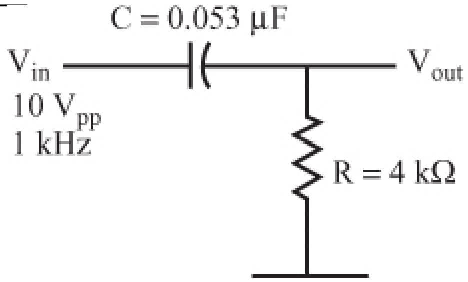

A. _____

B. _____

C. _____

D. _____

E. _____

2. Use the circuit shown in Figure 6.54 .

Figure 6.54

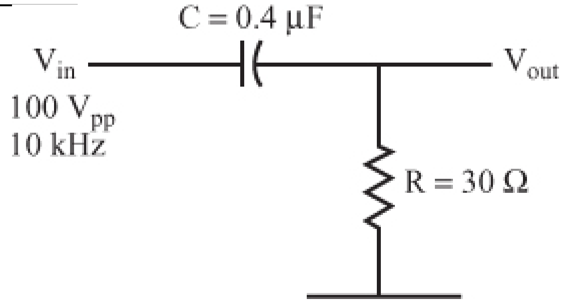
C = 0.4 µF
V_{in}
100 V_{pp}
10 kHz
V_{out}
R = 30 Ω

A. _____

B. _____

C. _____

D. _____

E. _____

3. Use the circuit shown in Figure 6.55 .

Figure 6.55

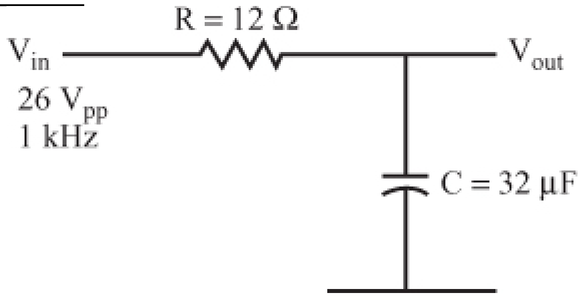
R = 12 Ω
V_{in}
26 V_{pp}
1 kHz
V_{out}
C = 32 µF

A. _____

B. _____

C. _____

D. _____

E. _____

For questions 4–6, calculate the following parameters for the circuit shown in each question.

A. X_C

B. AC V out

C. DC V out

4. Use the circuit shown in Figure 6.56 .

Figure 6.56

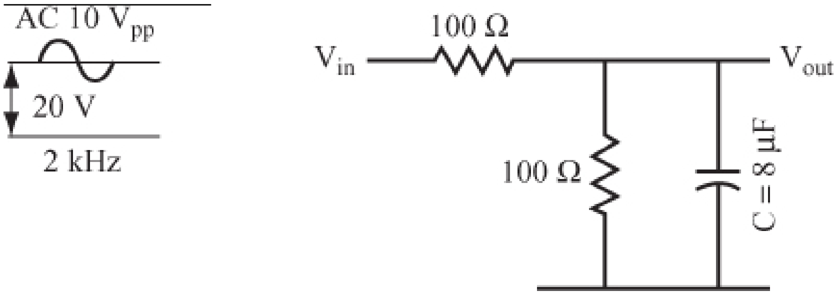
AC 10 V_{pp}
20 V
2 kHz
V_{in}
100 Ω
V_{out}
100 Ω
C = 8 µF

A. _____

B. _____

C. _____

5. Use the circuit shown in Figure 6.57 .

Figure 6.57

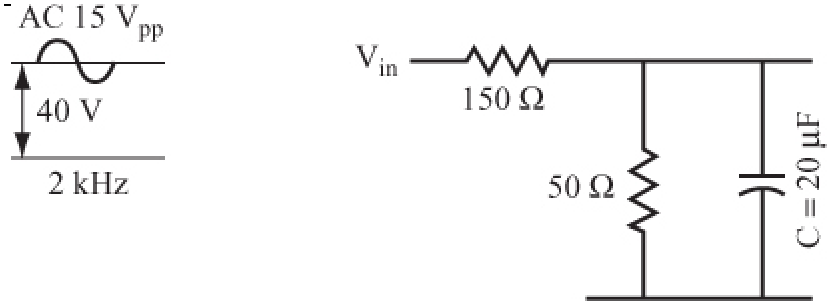
AC 15 V_{pp}
40 V
2 kHz
V_{in}
150 Ω
50 Ω
C = 20 µF

A. _____

B. _____

C. _____

6. Use the circuit shown in Figure 6.58 .

Figure 6.58

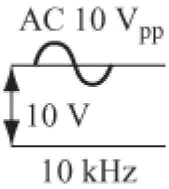
AC 10 Vpp
10 V
10 kHz

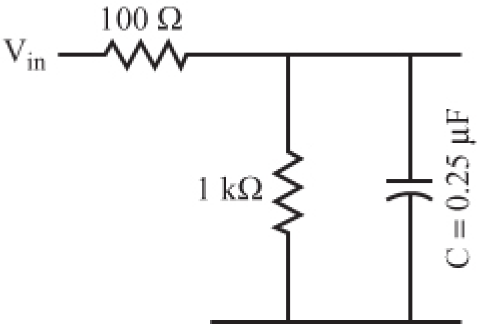
100 Ω
Vin
1 kΩ
C = 0.25 μF

A. _____

B. _____

C. _____

For questions 7–9, calculate the following parameters for the circuit shown in each question.

A. DC V out

B. X L

C. Z

D. AC V out

E. tan θ and θ

7. Use the circuit shown in Figure 6.59 .

Figure 6.59

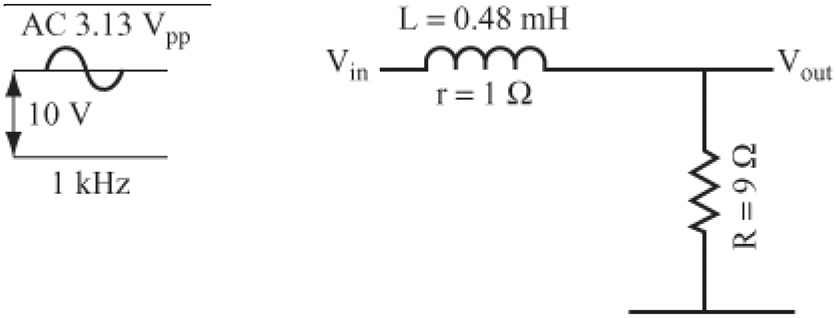
AC 3.13 Vpp
10 V
1 kHz
L = 0.48 mH
Vin
r = 1 Ω
Vout
R = 9 Ω

A. _____

B. _____

C. _____

D. _____

E. _____

8. Use the circuit shown in Figure 6.60.

Figure 6.60

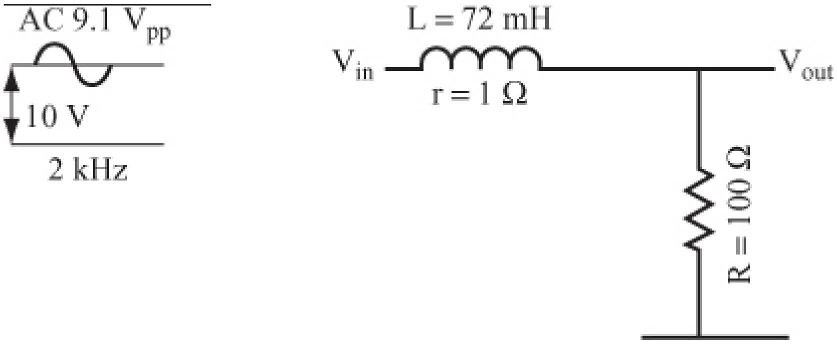
AC 9.1 V_{pp}
10 V
2 kHz
L = 72 mH
V_{in}
r = 1 Ω
V_{out}
R = 100 Ω

A. _____

B. _____

C. _____

D. _____

E. _____

9. Use the circuit shown in Figure 6.61.

Figure 6.61

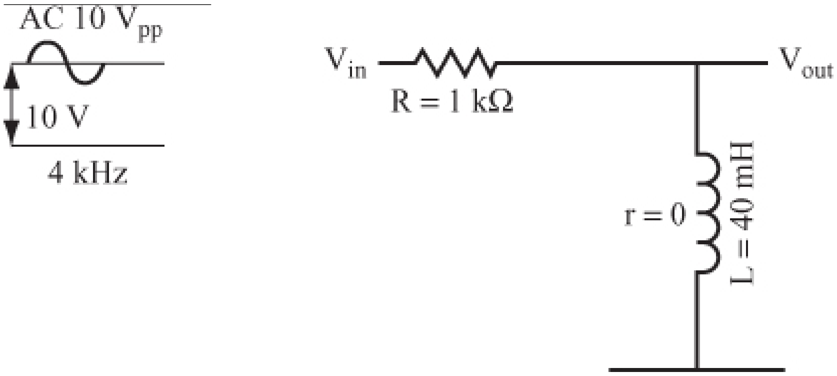
AC 10 V_{pp}
10 V
4 kHz
V_{in}
R = 1 kΩ
V_{out}
r = 0
L = 40 mH

A. _____

B. _____

C. _____

D. _____

E. _____

Answers to Self-Test

If your answers do not agree with those provided here, review the applicable problems in this chapter before you go to Chapter 7.

1A. 3 kΩ problems 8, 9, 10, 23
1B. 5 kΩ
1C. 8 volts
1D. 2 amperes
1E. 36.87 degrees
2A. 40 ohms problems 8, 9, 23
2B. 50 ohms
2C. 60 volts
2D. 2 amperes
2.E 53.13 degrees
3A. 5 ohms problems 8, 9, 23
3B. 13 ohms
3C. 10 volts
3D. 2 amperes
3E. 22.63 degrees
4A. 10 ohms problems 26 and 27
4B. 1 volt
4C. 10 volts
5A. 4 ohms problems 26 and 27
5B. 0.4 volt
5C. 10 volts
6A. 64 ohms problems 26 and 27
6B. 5.4 volts
6C. 9.1 volts
7A. 9 volts problems 28–30, 35

7B. 3 ohms
7C. 10.4 ohms
7D. 2.7 volts
7E. 16.7 degrees
8A. 10 volts problems 28–30, 35
8B. 904 ohms
8C. 910 ohms
8D. 1 volt
8E. 83.69 degrees
9A. 0 volts problems 28–30, 35
9B. 1 kΩ
9C. 1.414 kΩ
9D. 7 volts
9E. 45 degrees

Chapter 7

Resonant Circuits

You have seen how the inductor and the capacitor each present an opposition to the flow of an AC current, and how the magnitude of this reactance depends upon the frequency of the applied signal.

When inductors and capacitors are used together in a circuit (referred to as an LC circuit), a useful phenomenon called *resonance* occurs. Resonance is the frequency at which the reactance of the capacitor and the inductor is equal.

In this chapter, you learn about some of the properties of resonant circuits, and concentrate on those properties that lead to the study of oscillators (which is touched upon in the last few problems in this chapter and covered in more depth in Chapter 9, “Oscillators”).

After completing this chapter, you will be able to do the following:

Find the impedance of a series LC circuit.

Calculate the series LC circuit's resonant frequency.

Sketch a graph of the series LC circuit's output voltage.

Find the impedance of a parallel LC circuit.

Calculate the parallel LC circuit's resonant frequency.

Sketch a graph of the parallel LC circuit's output voltage.

Calculate the bandwidth and the quality factor (Q) of simple series and parallel LC circuits.

Calculate the frequency of an oscillator.

The Capacitor and Inductor in Series

1 Many electronic circuits contain a capacitor and an inductor placed in series, as shown in Figure 7.1 .

Figure 7.1

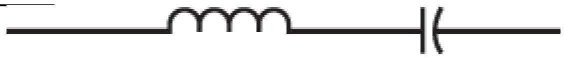

You can combine a capacitor and an inductor in series with a resistor to form voltage divider circuits, such as the two circuits shown in Figure 7.2 . A circuit that contains resistance (R), inductance (L), and capacitance (C) is referred to as an RLC circuit. Although the order of the capacitor and inductor differs in the two circuits shown in Figure 7.2 , they have the same effect on electrical signals.

Figure 7.2

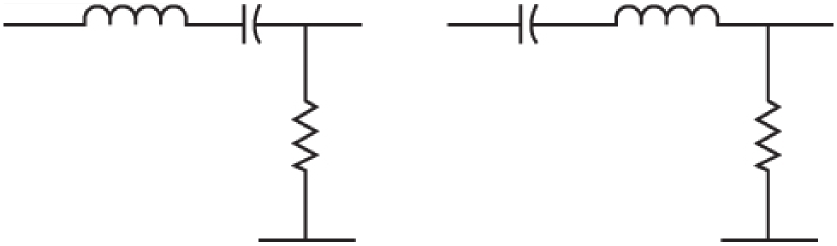

To simplify your calculations in the next few problems, you can assume that the small DC resistance of the inductor is much less than the resistance of the resistor R, and you can, therefore, ignore DC resistance in your calculations.

When you apply an AC signal to the circuits in Figure 7.2, both the inductor's and the capacitor's reactance value depends on the frequency.

Questions

A. What formula would you use to calculate the inductor's reactance? _____

B. What formula would you use to calculate the capacitor's reactance? _____

Answers

A. $X_L = 2\pi fL$

B. $$X_C = \frac{1}{2\pi fC}$$

2 You can calculate the net reactance (X) of a capacitor and inductor in series by using the following formula:

$$X = X_L \quad X_C$$

You can calculate the impedance of the RLC circuits shown in Figure 7.2 by using the following formula:

$$Z = \sqrt{R^2 + X^2}$$

In the formula, keep in mind that X^2 is $(X_L - X_C)^2$.

Calculate the net reactance and impedance for an RLC series circuit, such as those shown in Figure 7.2, with the following values:

1. $f = 1$ kHz
2. $L = 100$ mH
3. $C = 1$ μF
4. $R = 500$ ohms

Questions

Follow these steps to calculate the following:

A. Find X_L : _____

B. Find X_C : _____

C. Use $X = X_L - X_C$ to find the net reactance: _____

D. Use $Z = \sqrt{X^2 + R^2}$ to find the impedance: _____

Answers

A. $X_L = 628$ ohms

B. $X_C = 160$ ohms

C. $X = 468$ ohms (inductive)

D. $Z = 685$ ohms

3 Calculate the net reactance and impedance for an RLC series circuit, such as those shown in Figure 7.2, using the following values:

1. $f = 100$ Hz
2. $L = 0.5$ H
3. $C = 5$ μF
4. $R = 8$ ohms

Questions

Follow the steps outlined in problem 2 to calculate the following parameters:

A. X_L = _____

B. X_C = _____

C. X = _____

D. Z = _____

Answers

A. X_L = 314 ohms

B. X_C = 318 ohms

C. X = −4 ohms (capacitive)

D. Z = 9 ohms

By convention, the net reactance is negative when it is capacitive.

4 Calculate the net reactance and impedance for an RLC series circuit, such as those shown in Figure 7.2, using the values in the following questions.

Questions

f = 10 kHz, L = 15 mH, C = 0.01 μF, R = 494 ohms

X = _____

Z = _____

f = 2 MHz, L = 8 μH, C = 0.001 μF, R = 15 ohms

X = _____

Z = _____

Answers

A. X = −650 ohms (capacitive), Z = 816 ohms

B. X = 21 ohms (inductive), Z = 25.8 ohms

5 For the circuit shown in Figure 7.3, the output voltage is the voltage drop across the resistor.

Figure 7.3

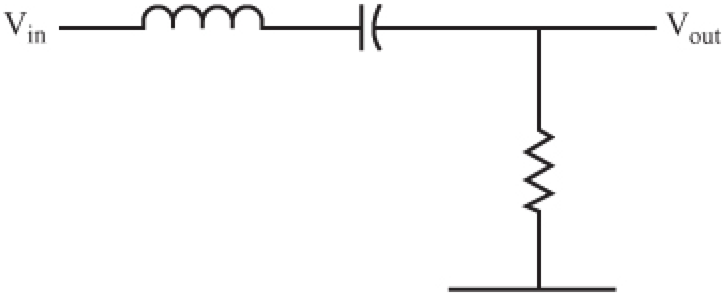

In problems 1 through 4, the net reactance of the series inductor and capacitor changes as the frequency changes. Therefore, as the frequency changes, the voltage drop across the resistor changes and so does the amplitude of the output voltage V out .

If you plot V out against frequency on a graph for the circuit shown in Figure 7.3 , the curve looks like the one shown in Figure 7.4 .

Figure 7.4

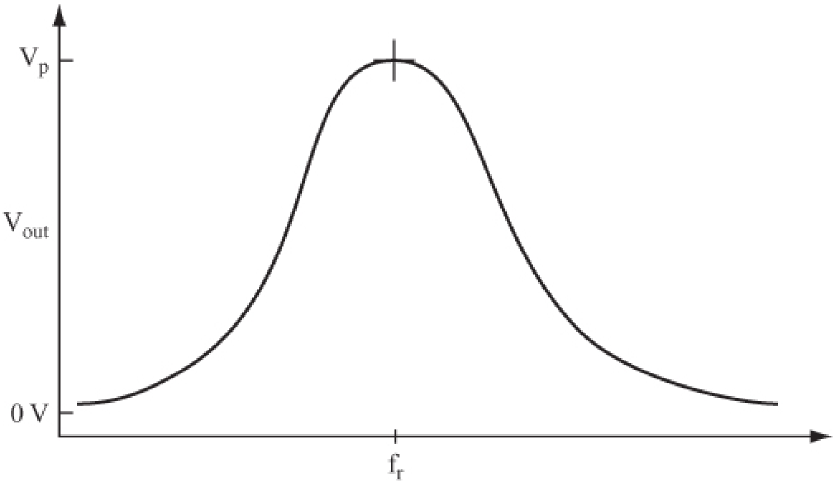

The maximum output voltage (or peak voltage) shown in this curve, V p , is slightly less than V in . This slight attenuation of the peak voltage from the input voltage is because of the DC resistance of the inductor.

The output voltage peaks at a frequency, f_r, where the net reactance of the inductor and capacitor in series is at its lowest value. At this frequency, there is little voltage drop across the inductor and capacitor. Therefore, most of the input voltage is applied across the resistor, and the output voltage is at its highest value.

Question

Under ideal conditions, if X_C were 10.6 ohms, what value of X_L results in a net reactance (X) of 0 for the circuit shown in Figure 7.3 ?

_____

Answer

$X = X_L - X_C = 0$, therefore:

$X_L = X_C + X = 10.6\ \Omega + 0 = 10.6\ \Omega$

6 You can find the frequency at which $X_L - X_C = 0$ by setting the formula for X_L equal to the formula for X_C and solving for f:

$$2\pi fL = \frac{1}{2\pi fC}$$

Therefore,

$$f_r = \frac{1}{2\pi\sqrt{LC}}$$

where f_r is the *resonant frequency* of the circuit.

Question

What effect does the value of the resistance have on the resonant frequency? _____

Answer

It has no effect at all.

7 Calculate the resonant frequency for the circuit shown in Figure 7.3 using the capacitor and inductor values given in the following questions.

Questions

A. C = 1 μF, L = 1 mH _____

f_r = _____

B. C = 16 μF, L = 1.6 mH _____

f_r = _____

Answers

A.
$$f_r = \frac{1}{2\pi\sqrt{1 \quad 10^{\;3} \quad 1 \quad 10^{\;6}}} = 5.0\,\text{kHz}$$

B.
$$f_r = \frac{1}{2\pi\sqrt{16 \quad 10^{\;6} \quad 1.6 \quad 10^{\;3}}} = 1\,\text{kHz}$$

8 Calculate the resonant frequency for the circuit shown in Figure 7.3 using the capacitor and inductor values given in the following questions.

Questions

A. C = 0.1 μF, L = 1 mH _____

f_r = _____

B. C = 1 μF, L = 2 mH _____

f_r = _____

Answers

A. f_r = 16 kHz

B. f_r = 3.6 kHz

9 For the RLC circuit shown in Figure 7.5, the output voltage is the voltage drop across the capacitor and inductor.

Figure 7.5

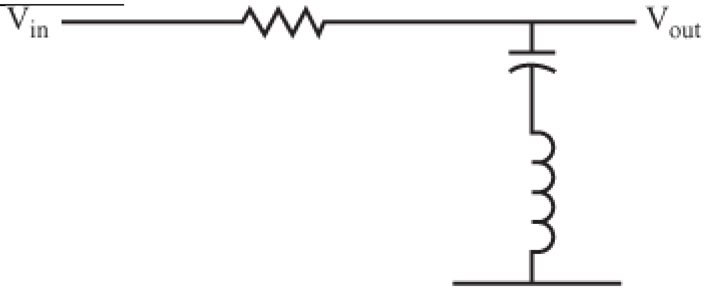

If V out is plotted on a graph against the frequency for the circuit shown in Figure 7.5, the curve looks like that shown in Figure 7.6.

Figure 7.6

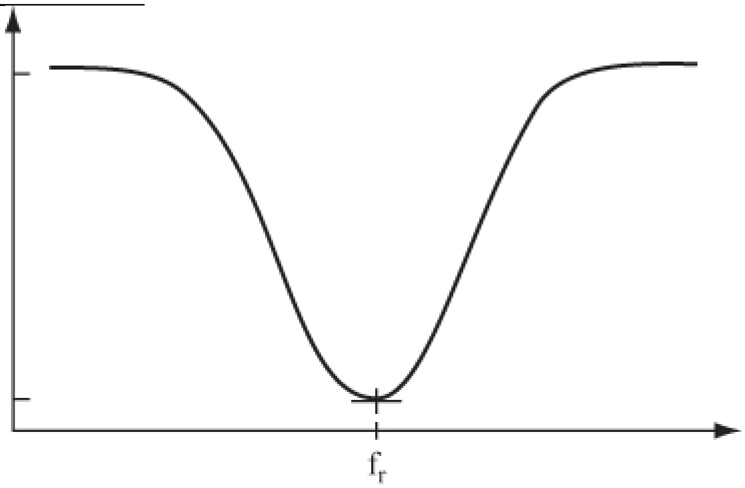

The output voltage drops to its minimum value at the resonant frequency for the circuit, which you can calculate with the formula provided in problem 6. At the resonant frequency, the net reactance of the inductor and capacitor in series is at a minimum. Therefore, there is little voltage drop across the inductor and capacitor, and the output voltage is at its minimum value. A circuit with this type of output (such as the circuit shown in Figure 7.5) is called a *notch filter*, or *band-reject filter*.

Question

What would you expect the minimum output voltage to be? _____

Answer

0 volts, or close to it

Project 7.1: The Notch Filter

Objective

The objective of this project is to determine how V out changes as the frequency of the input signal changes for a notch filter.

General Instructions

After the circuit is set up, you measure V out for each frequency. You can also generate a graph to show the relationship between the output voltage and the input frequency.

Parts List

You need the following equipment and supplies:

- One 100 Ω, 0.25-watt resistor.
- One 1000 pF capacitor. (1000 pF is also sometimes stated by suppliers as 0.001 μF.)
- One 100 μH inductor. You'll often find inductors that use a numerical code to indicate the value of the inductor. The first

two numbers in this code are the first and second significant digits of the inductance value. The third number is the multiplier, and the units are μH. Therefore, an inductor marked with 101 has a value of 100 μH.)

- One function generator.
- One oscilloscope.
- One breadboard.

Step-by-Step Instructions

Set up the circuit shown in Figure 7.7 . If you have some experience in building circuits, this schematic (along with the previous parts list) should provide all the information you need to build the circuit. If you need a bit more help building the circuit, look at the photos of the completed circuit in the "Expected Results" section.

Figure 7.7

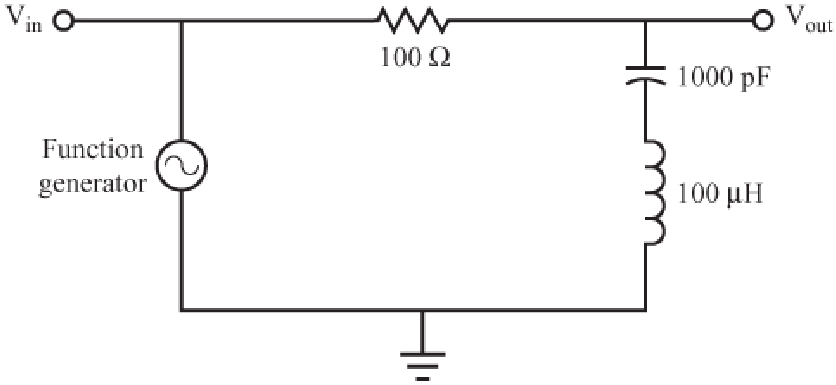

Carefully check your circuit against the diagram.

After you check your circuit, follow these steps, and record your measurements in the blank table following the steps.

1. Connect the oscilloscope probe for channel 2 to a jumper wire connected to V_{in}, and connect the ground clip to a jumper wire attached to the ground bus.
2. Connect the oscilloscope probe for channel 1 to a jumper wire connected to V_{out}, and connect the ground clip to a jumper wire attached to the ground bus.
3. Set the function generator to generate a 5 V_{pp}, 100 kHz sine wave.
4. Measure and record V_{out}.
5. Adjust the function generator to the frequency shown in the next row of the table (labeled 150 kHz in this instance). Each time you change the frequency, check V_{in} and adjust the amplitude knob on the function generator to maintain V_{in} at 5 V_{pp} if needed. (If you leave the amplitude knob in one position, the voltage of the signal provided by the function generator will change as the net reactance of the circuit changes.)
6. Measure and record V_{out}.
7. Repeat steps 5 and 6 for the remaining values until you have recorded V_{out} in all rows of the table.

f_{in} (kHz)	V_{out} (volts)
100	
150	
200	

250
300
350
400
450
500
550
600
650
700
750
800
850
900

8. In the blank graph shown in Figure 7.8, plot Vout vs fin with the voltage on the vertical axis and the frequency on the X axis. The curve should have the same shape as the curve shown in Figure 7.6.

Figure 7.8

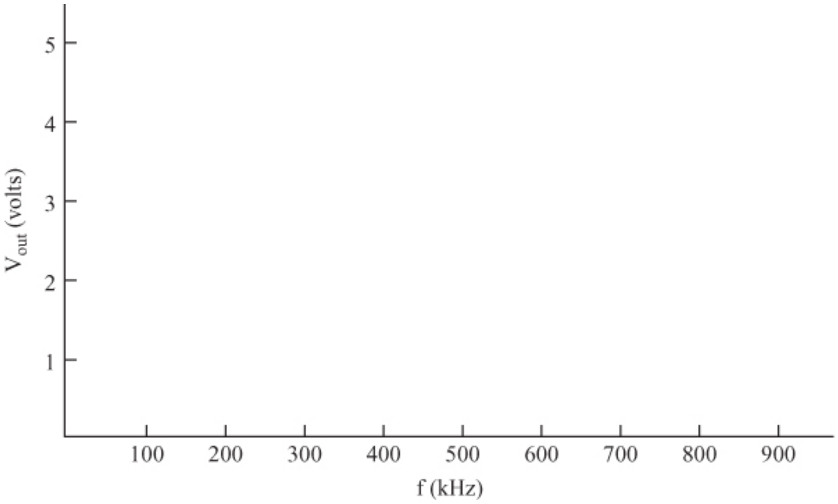

Expected Results

Figure 7.9 shows the breadboarded circuit for this project.

Figure 7.9

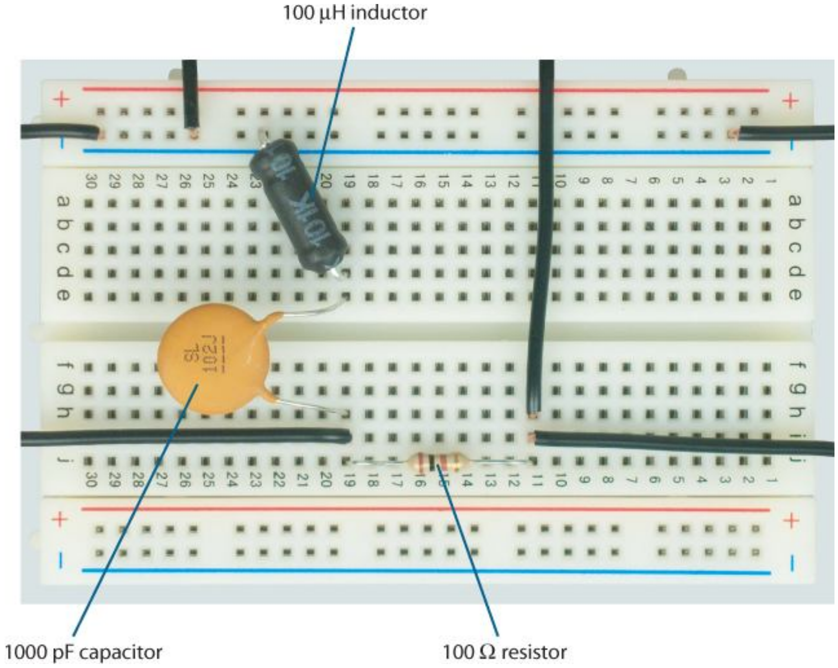
100 µH inductor
1000 pF capacitor
100 Ω resistor

Figure 7.10 shows a function generator and oscilloscope attached to the circuit.

Figure 7.10

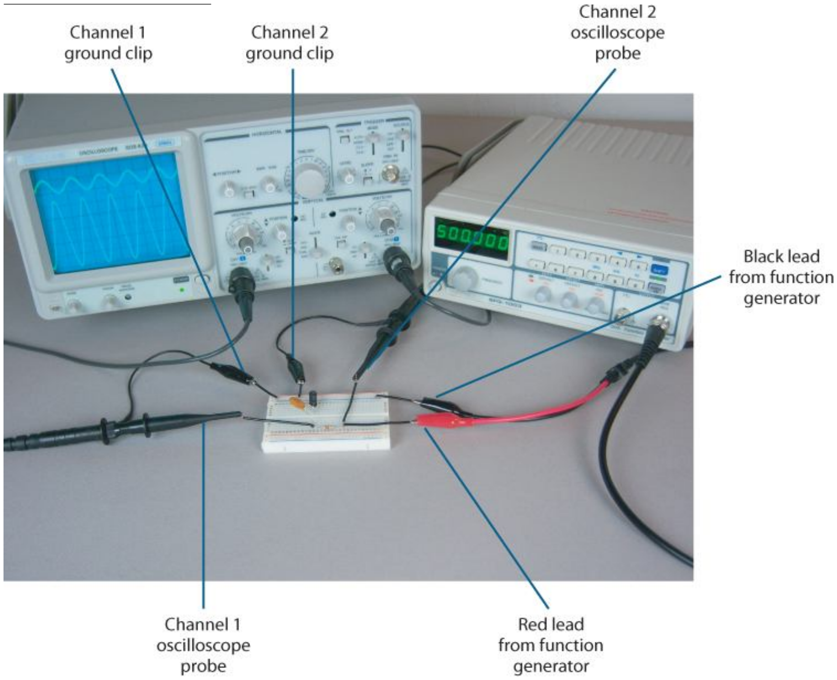
Channel 1
ground clip
Channel 2
ground clip
Channel 2
oscilloscope
probe
Black lead
from function
generator
Channel 1
oscilloscope
probe
Red lead
from function
generator

The input signal is represented by the upper sine wave shown in Figure 7.11, and the output signal is represented by the lower sine wave. Read the number of divisions for the peak-to-peak output sine wave, and multiply it by the corresponding VOLTS/DIV setting to determine V_{out}.

Figure 7.11

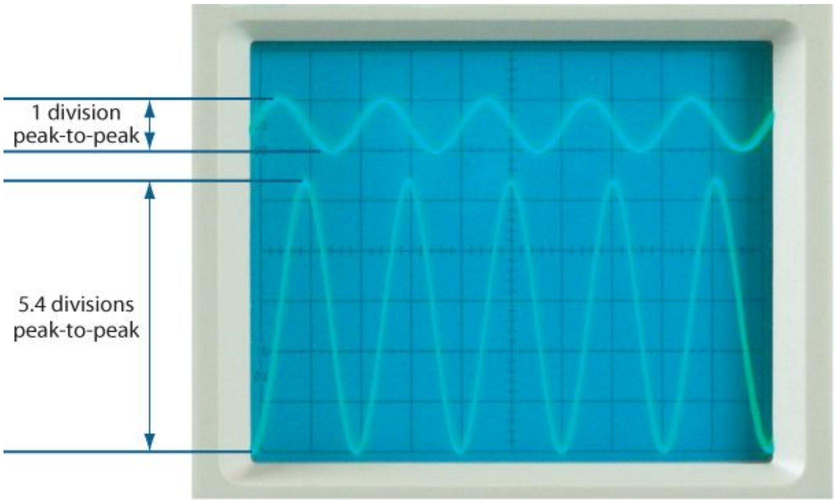
1 division
peak-to-peak
5.4 divisions
peak-to-peak

As you set f in to a new value on the function generator, you may also need to adjust the TIME/DIV control, the VOLTS/DIV control, and vertical POSITION controls on the oscilloscope. The controls shown in Figure 7.12 are adjusted to measure V out when f in = 500 kHz.

Figure 7.12

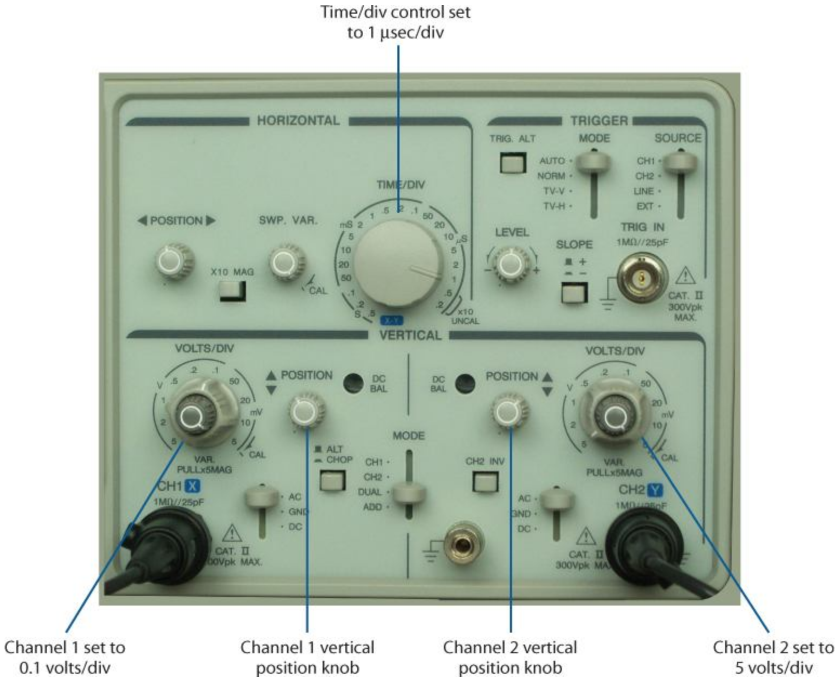
Time/div control set to 1 µsec/div
HORIZONTAL
TRIGGER
TIME/DIV
VERTICAL
VOLTS/DIV
POSITION
MODE
CH1 X
CH2 Y
Channel 1 set to 0.1 volts/div
Channel 1 vertical position knob
Channel 2 vertical position knob
Channel 2 set to 5 volts/div

Your values should be close to those shown in the following table, and the curve should be similar to that shown in Figure 7.13 .

f in (kHz)	V out (volts)
100	4.9
150	4.9
200	4.9
250	4.8
300	4.7
350	4.4
400	4.0
450	2.8
500	0.5
550	2.3
600	3.6

650	4.2
700	4.4
750	4.7
800	4.7
850	4.8
900	4.8

Figure 7.13

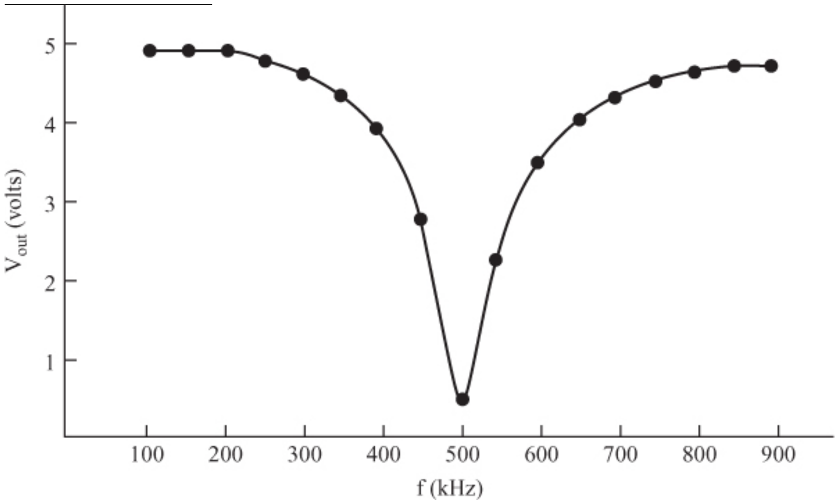

Notice the extra data points shown in the graph near the minimum V_{out}. These extra data points help you to determine the frequency at which the minimum V_{out} occurs. In this graph, the minimum V_{out} occurs at a frequency of 505 kHz, which is close to the calculated resonance frequency of 503 kHz for this circuit.

10 You can connect the capacitor and inductor in parallel, as shown in Figure 7.14.

Figure 7.14

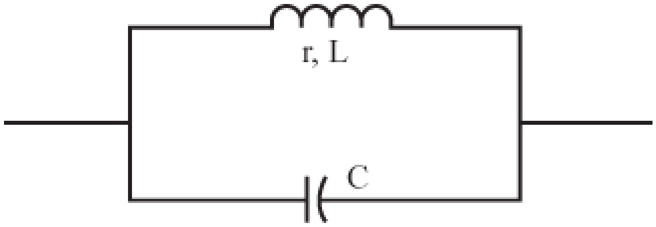

You can calculate the resonance frequency of this circuit using the following formula:

$$f_r = \frac{1}{2\pi\sqrt{LC}}\sqrt{1 \quad \frac{r^2C}{L}}$$

In this formula, r is the DC resistance of the inductor. However, if the reactance of the inductor is equal to, or more than, 10 times the DC resistance of the inductor, you can use the following simpler formula. This is the same formula that you used in problems 7 and 8 for the series circuit.

$$f_r = \frac{1}{2\pi\sqrt{LC}}$$

Q, the quality factor of the circuit, is equal to X_L/r. Therefore, you can use this simple equation to calculate f_r if Q is equal to, or greater than, 10.

Questions

A. Which formula should you use to calculate the resonant frequency of a parallel circuit if the Q of the coil is 20? _____

B. If the Q is 8? _____

Answers

A. $$f_r = \frac{1}{2\pi\sqrt{LC}}$$

B.

$$f_r = \frac{1}{2\pi\sqrt{LC}}\sqrt{1 \quad \frac{r^2C}{L}}$$

Note Here is another version of the resonance frequency formula that is helpful when Q is known:

$$f_r = \frac{1}{2\pi\sqrt{LC}}\sqrt{\frac{Q^2}{1+Q^2}}$$

11 You can calculate the total opposition (impedance) of an inductor and capacitor connected in parallel to the flow of current by using the following formulas for a circuit at resonance:

$$Z_p = Q^2r\text{, if Q is equal to or greater than 10}$$

$$Z_p = \frac{L}{rC}\text{, for any value of Q}$$

At resonance, the impedance of an inductor and capacitor in parallel is at its maximum.

You can use an inductor and capacitor in parallel in a voltage divider circuit, as shown in Figure 7.15 .

Figure 7.15

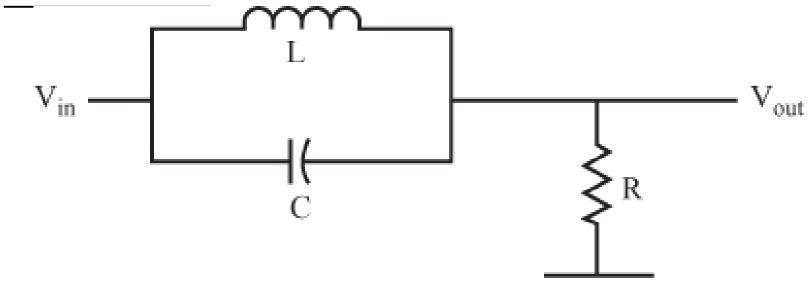

If V_{out} is plotted against frequency on a graph for the circuit shown in Figure 7.15, the curve looks like that shown in Figure 7.16.

Figure 7.16

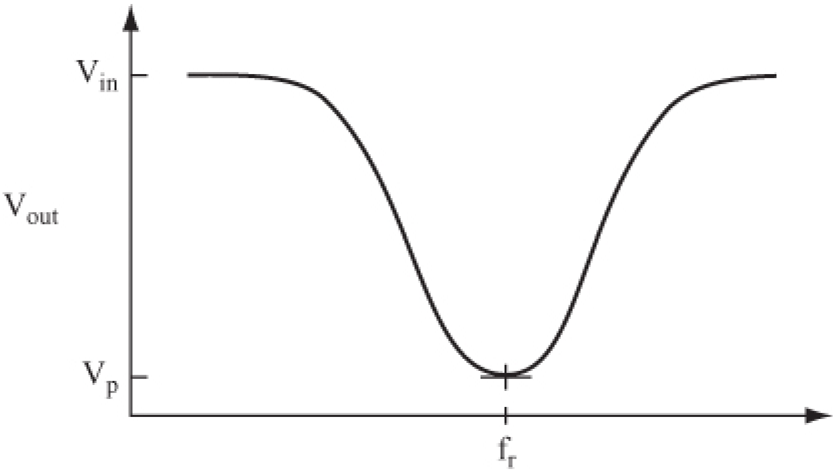

Questions

A. What would be the total impedance formula for the voltage divider circuit at resonance? _____

B. What is the frequency called at the point where the curve is at its lowest point? _____

C. Why is the output voltage at a minimum value at resonance? _____

Answers

A. $Z_T = Z_p + R$

Note The relationship shown by this formula is true only at resonance. At all other frequencies, Z_T is a complicated formula or calculation found by considering a series r, L circuit in parallel with a capacitor.

B. The parallel resonant frequency.

C. The output voltage is at its lowest value at the resonant frequency. This is because the impedance of the parallel resonant circuit is at its highest value at this frequency.

12 For the circuit shown in Figure 7.17, the output voltage equals the voltage drop across the inductor and capacitor.

Figure 7.17

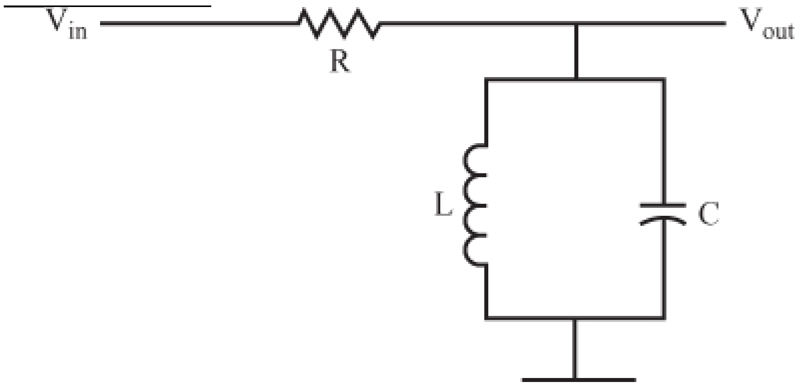

If V out is plotted on a graph against frequency for the circuit shown in Figure 7.17, the curve looks like that shown in Figure 7.18. At the resonance frequency, the impedance of the parallel inductor and capacitor is at its maximum value. Therefore, the voltage drop across the parallel inductor and capacitor (which is also the output voltage) is at its maximum value.

Figure 7.18

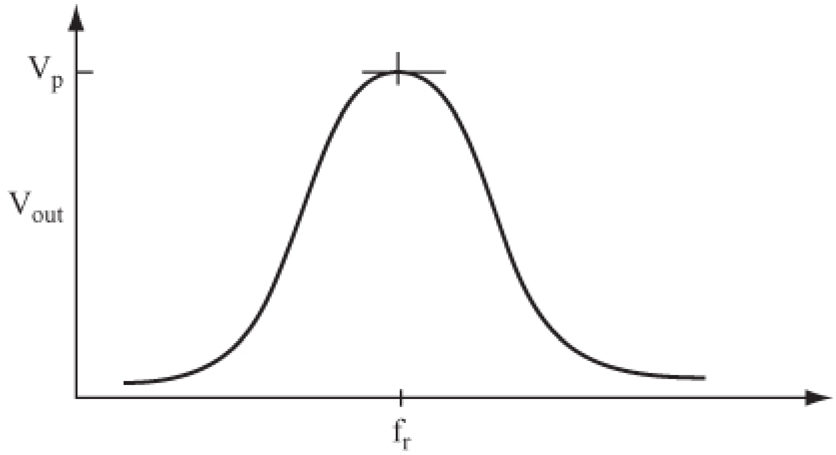

Question

What formula would you use to calculate the resonant frequency? _____

Answer

$$f_r = \frac{1}{2\pi\sqrt{LC}} \text{ if Q is equal to or greater than 10}$$

$$f_r = \frac{1}{2\pi\sqrt{LC}}\sqrt{1 \quad \frac{r^2C}{L}} \text{ if Q is less than 10}$$

13 Find the resonant frequency in these two examples, where the capacitor and the inductor are in parallel. (Q is greater than 10.)

Questions

A. L = 5 mH, C = 5 μF _____ fr = _____

B. L = 1 mH, C = 10 μF _____ fr = _____

Answers

A. fr = 1 kHz (approximately)

B. fr = 1600 Hz (approximately)

The Output Curve

14 Now it's time to look at the output curve in a little more detail. Take a look at the curve shown in Figure 7.19 for an example.

Figure 7.19

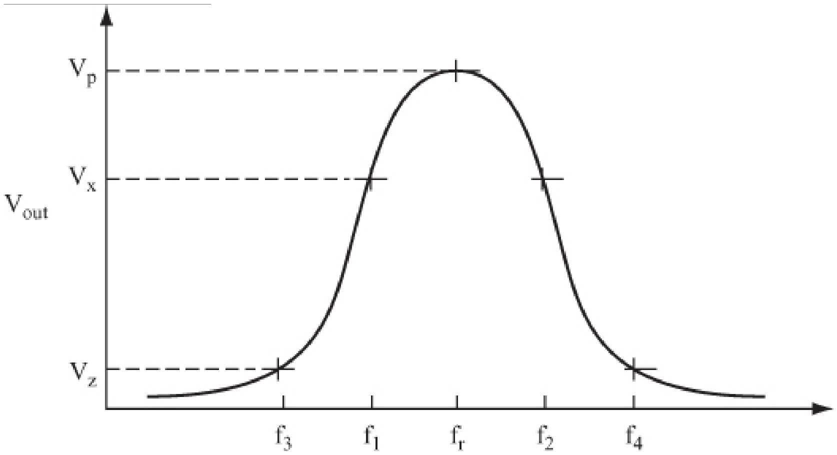

An input signal at the resonant frequency, f_r, passes through a circuit with minimum attenuation, and with its output voltage equal to the peak output voltage, V_p, shown on this curve.

The two frequencies f_1 and f_2 are “passed” almost as well as f_r is passed. That is, signals at those frequencies have a high output voltage, almost as high as the output of a signal at f_r. The graph shows this voltage as V_x.

Signals at frequencies f_3 and f_4 have a low output voltage.

These two frequencies are not passed but are said to be *blocked* or *rejected* by the circuit. This output voltage is shown on the graph as V_z.

The output or frequency response curve for a resonant circuit (series or parallel) has a symmetrical shape for a high value of Q. You can make the assumption that the output curve is symmetrical when Q is greater than 10.

Questions

A. What is meant by a frequency that is passed? _____

B. Why are f_1 and f_2 passed almost as well as f_r? _____

C. What is meant by a frequency that is blocked? _____

D. Which frequencies shown on the previous output curve are blocked? _____

E. Does the output curve shown appear to be symmetrical? What does this mean for the circuit? _____

Answers

A. It appears at the output with minimum attenuation.

B. Because their frequencies are close to f_r.

C. It has a low output voltage.

D. f_3 and f_4 (as well as all frequencies below f_3 and above f_4).

E. It does appear to be symmetrical. This means that the coil has a Q greater than 10.

15 Somewhere between f_r and f_3, and between f_r and f_4, there is a point at which

frequencies are said to be either passed or reduced to such a level that they are effectively blocked. The dividing line is at the level at which the power output of the circuit is half as much as the power output at peak value. This happens to occur at a level that is 0.707, or 70.7 percent of the peak value.

For the output curve shown in problem 14, this occurs at a voltage level of 0.707 V_p. The two corresponding frequencies taken from the graph are called the *half power frequencies* or *half power points*. These are common expressions used in the design of resonant circuits and frequency response graphs.

If a certain frequency results in an output voltage that is equal to or greater than the half power point, it is said to be passed or accepted by the circuit. If it is lower than the half power point, it is said to be blocked or rejected by the circuit.

Question

Suppose V_P = 10 volts. What is the minimum voltage level of all frequencies that are passed by the circuit? _____

Answer

$V = 10 \text{ volts} \times 0.707 = 7.07 \text{ volts}$

(If a frequency has an output voltage above 7.07 volts, you would say it is passed by the circuit.)

16 Assume the output voltage at the

resonant frequency in a circuit is 5 volts. Another frequency has an output of 3.3 volts.

Question

Is this second frequency passed or blocked by the circuit? _____

Answers

$V = V_p \times 0.707 = 5 \times 0.707 = 3.535$ volts

3.3 volts is less than 3.535 volts, so this frequency is blocked.

17 In these examples, find the voltage level at the half power points.

Questions

A. $V_p = 20$ volts _____

B. $V_p = 100$ volts _____

C. $V_p = 3.2$ volts _____

Answers

A. 14.14 volts

B. 70.70 volts

C. 2.262 volts

18 Although this discussion started off by talking about the resonance frequency, a few other frequencies have been introduced. At this point, the discussion is dealing with a band or a range of frequencies.

Two frequencies correspond to the half power points on the curve. Assume these frequencies are f_1 and f_2. The difference you find when you subtract f_1 from f_2 is important because this gives the range of frequencies that are passed by the circuit. This range is called the *bandwidth* of the circuit and can be calculated using the following equation:

$$\mathrm{BW} = \mathrm{f}_2 \quad \mathrm{f}_1$$

All frequencies within the bandwidth are passed by the circuit, whereas all frequencies outside the bandwidth are blocked. A circuit with this type of output (such as the circuit shown in Figure 7.17) is referred to as a *bandpass filter* .

Question

Indicate which of the following pairs of values represent a wider range of frequencies, or, in other words, the wider bandwidth.

A. $f_2 = 200$ Hz, $f_1 = 100$ Hz

B. $f_2 = 20$ Hz, $f_1 = 10$ Hz

Answer

The bandwidth is wider for the frequencies given in A.

When playing a radio, you listen to one station at a time, not to the adjacent stations on the dial. Thus, your radio tuner must have a narrow bandwidth so that it can select only the frequency of that one station.

The amplifiers in a television set, however, must pass frequencies from 30 Hz up to approximately 4.5 MHz, which requires a wider bandwidth. The application or use to which you'll put a circuit determines the bandwidth that you should design the circuit to provide.

19 The output curve for a circuit that

passes a band of frequencies around the resonance frequency (such as the curve shown in Figure 7.20) was discussed in the last few problems.

Figure 7.20

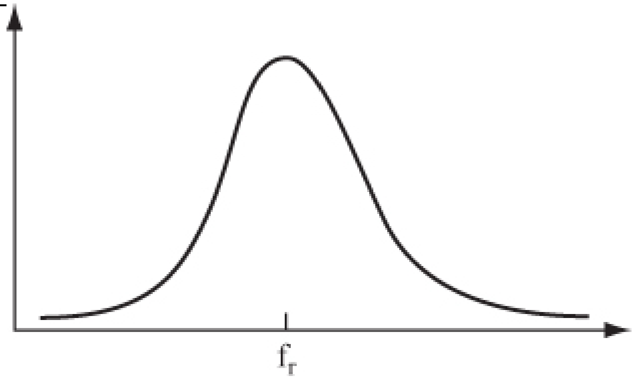

The same principles and equations apply to the output curve for a circuit that blocks a band of frequencies around the resonance frequency, as is the case with the curve shown in Figure 7.21 .

Figure 7.21

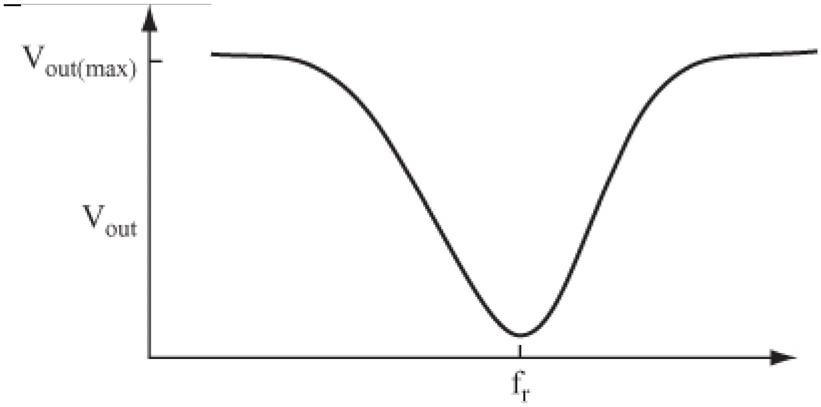

Questions

A. What points on the curve shown in Figure

7.21 would you use to determine the circuit's bandwidth? _____

B. Would the output voltage at the resonant frequency be above or below these points? _____

Answers

A. The half power points (0.707 V out(max)).

B. The output voltage at the resonant frequency is the minimum point on the curve, which is below the level for the half power points.

Project 7.2: The Band Pass Filter

Objective

The objective of this project is to determine how V out changes as the frequency of the input signal changes for a bandpass filter.

General Instructions

When the circuit is set up, you measure V out for each frequency. You also generate a graph to show the relationship between the output voltage and the input frequency.

Parts List

You need the following equipment and supplies:

- One 100 Ω, 0.25-watt resistor.
- One 1000 pF capacitor. (1000 pF is also sometimes stated by suppliers as 0.001 μF.)
- One 100 μH inductor.
- One function generator.
- One oscilloscope.
- One breadboard.

Step-by-Step Instructions

Set up the circuit shown in Figure 7.22. If you have some experience in building circuits, this schematic (along with the previous parts list) should provide all the information you need to build the circuit. If you need a bit more help building the circuit, look at the photos of the completed circuit in the "Expected Results" section.

Figure 7.22

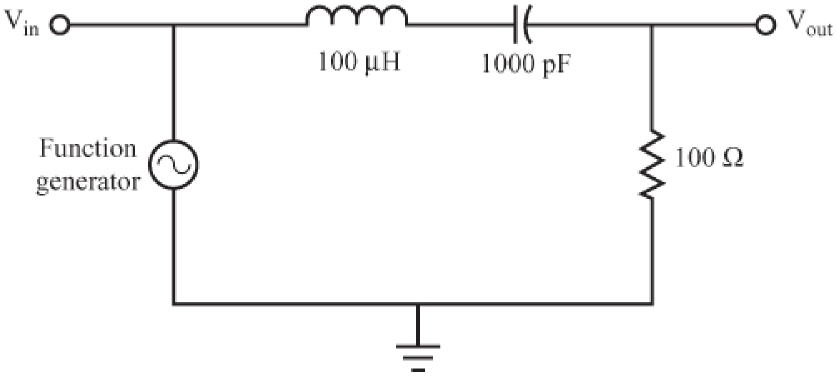

Carefully check your circuit against the diagram.

After you check your circuit, follow these steps, and record your measurements in the blank table following the steps.

1. Connect the oscilloscope probe for channel 2 to a jumper wire connected to V_{in}, and connect the ground clip to a jumper wire attached to the ground bus.
2. Connect the oscilloscope probe for channel 1 to a jumper wire connected to V_{out}, and connect the ground clip to a jumper wire attached to the ground bus.

3. Set the function generator to generate a 5 V_{pp}, 100 kHz sine wave.
4. Measure and record V_{out}.
5. Adjust the function generator to the frequency shown in the next row of the table (labeled 150 kHz in this instance). Each time you change the frequency, check V_{in}, and adjust the amplitude knob on the function generator to maintain V_{in} at 5 V_{pp} if needed. (If you leave the amplitude knob in one position, the voltage of the signal provided by the function generator will change as the net reactance of the circuit changes.)
6. Measure and record V_{out}.
7. Repeat steps 5 and 6 until you have recorded V_{out} for the last row of the table.

f_{in} (kHz)	V_{out} (volts)
100	
150	
200	
250	
300	
350	
400	
450	
500	
550	
600	
650	
700	
750	

800

850

900

8. In the blank graph shown in Figure 7.23, plot V_{out} versus f_{in} with the voltage on the vertical axis and the frequency on the X axis. The curve should have the same shape as the curve shown in Figure 7.20.

Figure 7.23

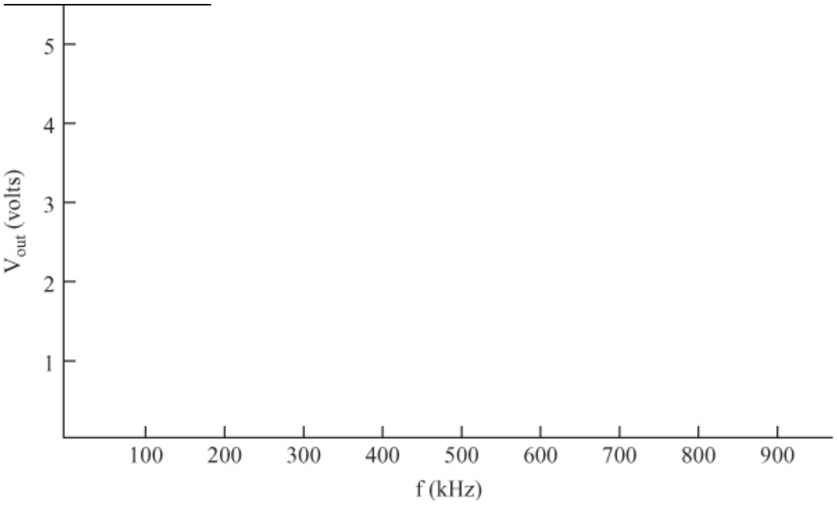

Expected Results

Figure 7.24 shows the breadboarded circuit for this project.

Figure 7.24

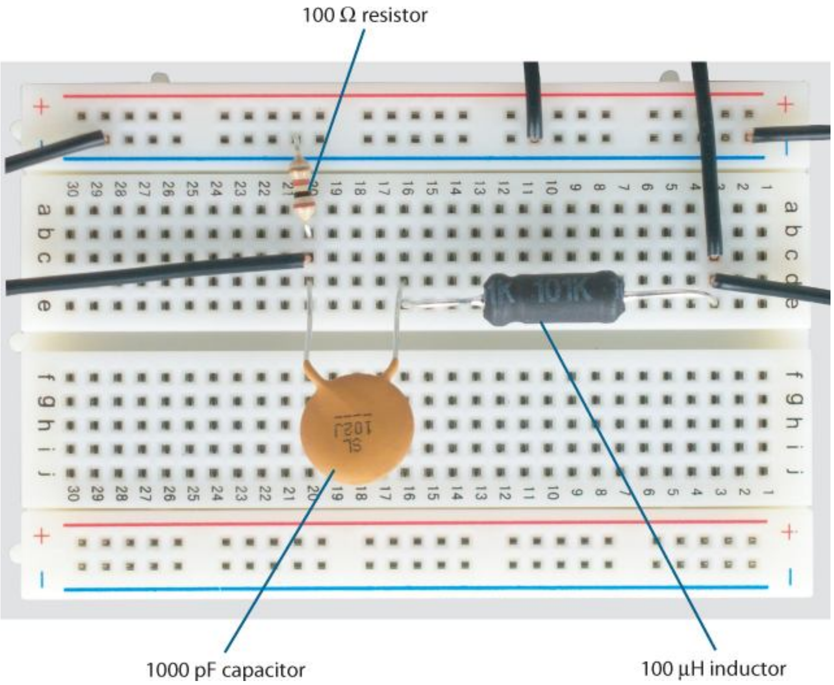

100 Ω resistor
1000 pF capacitor
100 µH inductor

Figure 7.25 shows a function generator and oscilloscope attached to the circuit.

Figure 7.25

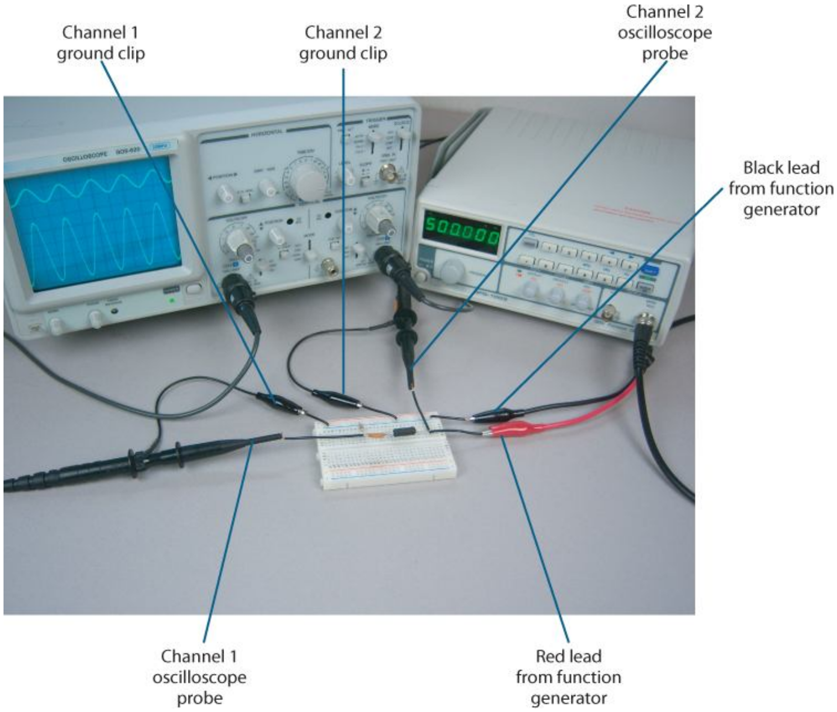
Channel 1 ground clip
Channel 2 ground clip
Channel 2 oscilloscope probe
Black lead from function generator
Channel 1 oscilloscope probe
Red lead from function generator

The input signal is represented by the upper sine wave shown in Figure 7.26, and the output signal is represented by the lower sine wave. Read the number of divisions for the peak-to-peak output sine wave, and multiply it by the corresponding VOLTS/DIV setting to determine V_{out}.

Figure 7.26

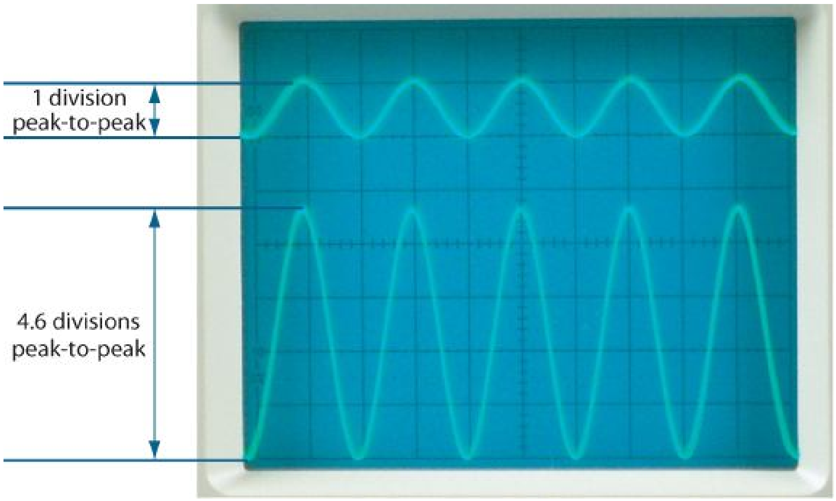
1 division
peak-to-peak
4.6 divisions
peak-to-peak

As you set f_{in} to a new value on the function generator, you may also need to adjust the TIME/DIV control, the VOLTS/DIV control, and vertical POSITION controls on the oscilloscope. The controls shown in Figure 7.27 are adjusted to measure V_{out} when f_{in} = 500 kHz.

Figure 7.27

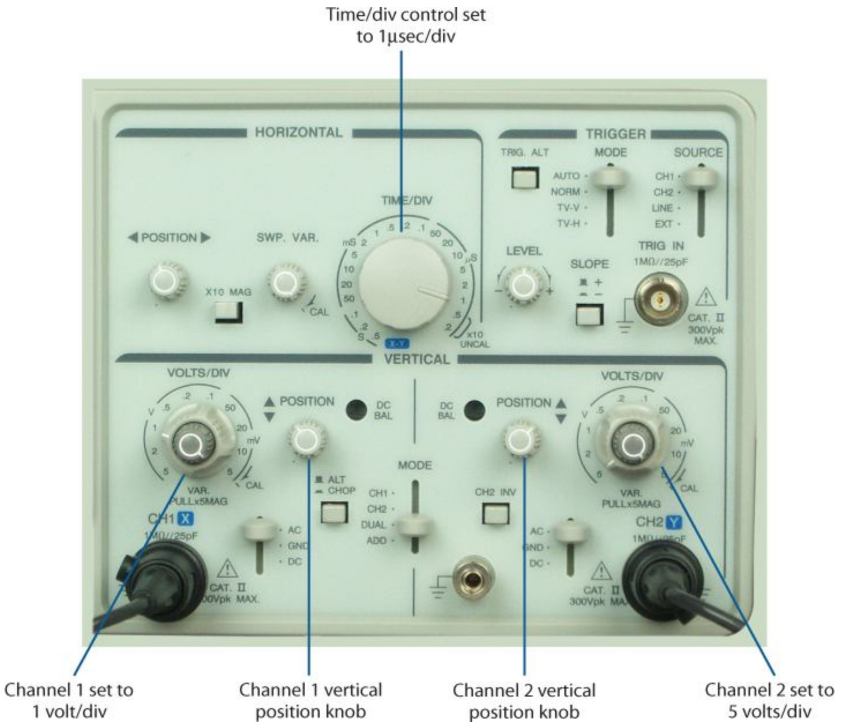
HORIZONTAL
TRIGGER
TIME/DIV
POSITION
SWP. VAR.
X10 MAG
CAL
TRIG. ALT
MODE
SOURCE
AUTO
NORM
TV-V
TV-H
CH1
CH2
LINE
EXT
LEVEL
SLOPE
TRIG IN
VERTICAL
VOLTS/DIV
DC BAL
ALT
CHOP
CH2 INV
DUAL
ADD
AC
GND
DC
CH1 X
CH2 Y
VAR. PULLx5MAG
CAT. II 300Vpk MAX.

Time/div control set to 1µsec/div

Channel 1 set to 1 volt/div

Channel 1 vertical position knob

Channel 2 vertical position knob

Channel 2 set to 5 volts/div

Your values should be close to those shown in the following table, and the curve should be similar to that shown Figure 7.28 .

f in (kHz)	V out (volts)
100	0.3
150	0.5
200	0.7
250	1.0
300	1.4
350	1.8
400	2.6
450	3.6
500	4.6
550	4.2

600	3.4
650	2.7
700	2.2
750	1.8
800	1.6
850	1.4
900	1.3

Because Q = 3.2 (well below 10), the curve for this circuit is not perfectly symmetrical.

Figure 7.28

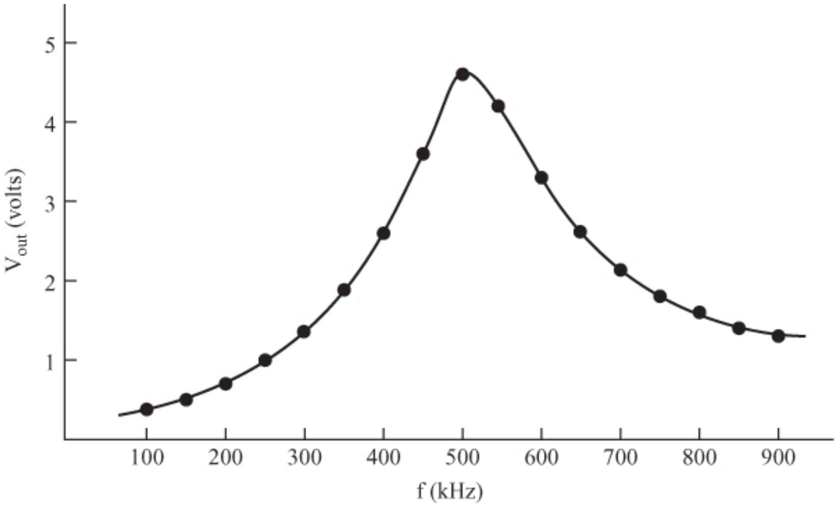

20 You can find the bandwidth of a circuit by measuring the frequencies (f1 and f2) at which the half power points occur and then using the following formula:

$$BW = f_2 \quad f_1$$

Or you can calculate the bandwidth of a circuit using this formula:

$$BW = \frac{f_r}{Q}$$

where:

$$Q = \frac{X_L}{R}$$

The formula used to calculate bandwidth indicates that, for two circuits with the same resonant frequency, the circuit with the larger Q will have the smaller bandwidth.

When you calculate Q for a circuit containing a capacitor and inductor in series (such as that shown in Figure 7.29), use the total DC resistance—the sum of the DC resistance (r) of the inductor and the value of the resistor (R)—to calculate Q.

Figure 7.29

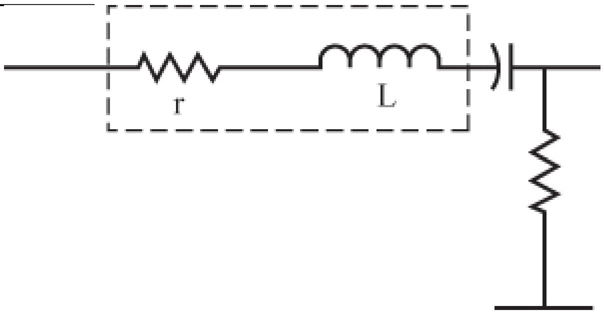

When you calculate Q for a circuit containing an inductor and capacitor in parallel, as with the circuit shown in Figure 7.30 , you do not include the value of the resistor (R) in the calculation. The only resistance you use in the calculation is the DC resistance (r) of the inductor.

Figure 7.30

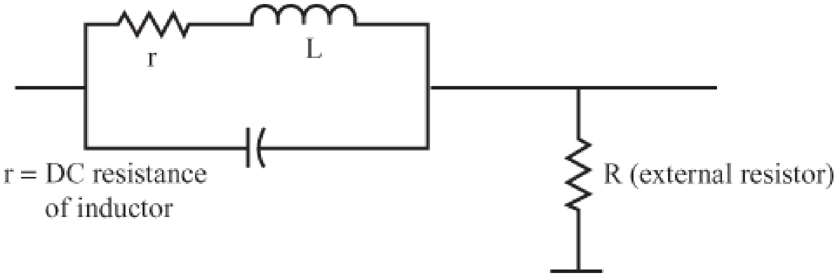
r
L
r = DC resistance of inductor
R (external resistor)

When you calculate Q for a circuit containing an inductor, a capacitor, and a resistor in parallel (as with the two circuits shown in Figure 7.31), include the value of the resistor (R) in the calculation.

Figure 7.31

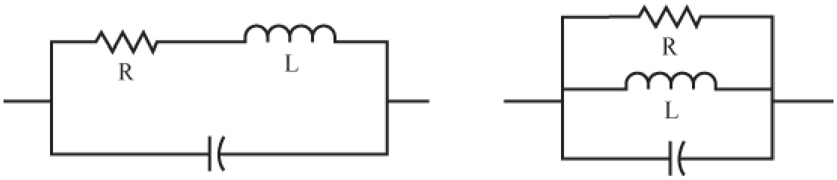

Questions

For the circuit shown in Figure 7.32 , all the component values are provided in the diagram. Find f_r , Q, and BW.

Figure 7.32

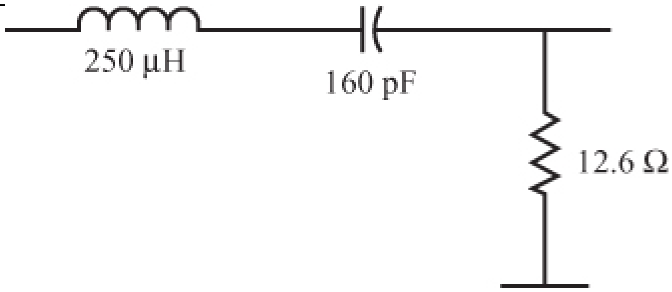
250 µH
160 pF
12.6 Ω

A. f_r = _____

B. Q = _____

C. BW = _____

Answers

A.

$$f_r = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{250 \quad 10^{6} \quad 160 \quad 10^{12}}} = 796\,\text{kHZ}$$

B.

$$Q = \frac{X_L}{R} = \frac{2\pi fL}{R} = \frac{2\pi \quad 796\,\text{kHz} \quad 250\,\mu\text{H}}{12.6\,\Omega} = 99.2$$

C.

$$BW = \frac{f_r}{Q} = \frac{796\,\text{kHz}}{99.2} = 8\,\text{kHz}$$

21 Use the circuit and component values shown in Figure 7.33 to answer the following questions.

Figure 7.33

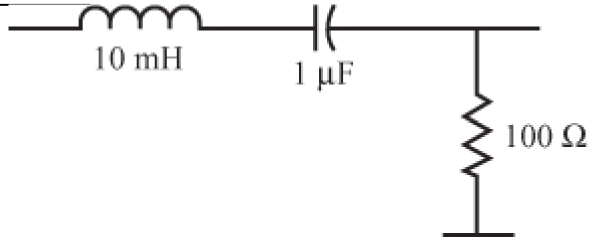

Questions

Find f_r, Q, and BW. Then, on a separate sheet of paper, draw an output curve showing the range of frequencies that are passed and blocked.

f_r = _____

Q = _____

BW = _____

Answers

f_r = 1590 Hz; Q = 1; BW = 1590 Hz

The output curve is shown in Figure 7.34 .

Figure 7.34

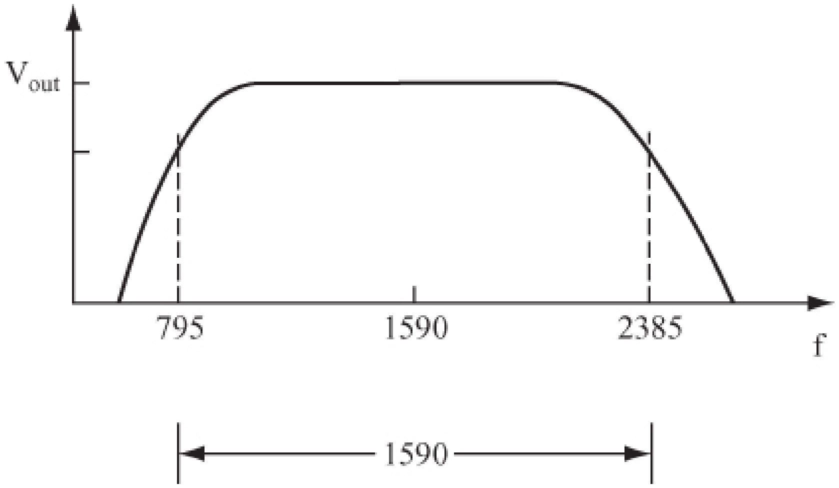

22 Use the circuit and component values shown in Figure 7.35 to answer the following questions.

Figure 7.35

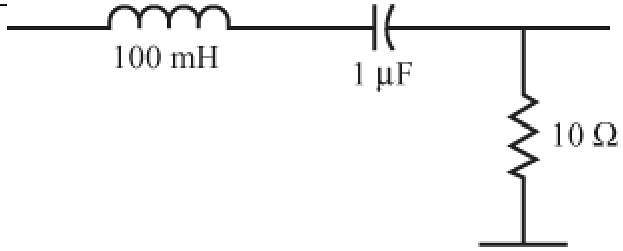

Questions

Find f r , Q, and BW for this circuit. Then draw the output curve on a separate sheet of paper.

f r = _____

Q = _____

BW = _____

Answers

f r = 500 Hz; Q = 31.4; BW = 16 Hz

The output curve is shown in Figure 7.36 .

Figure 7.36

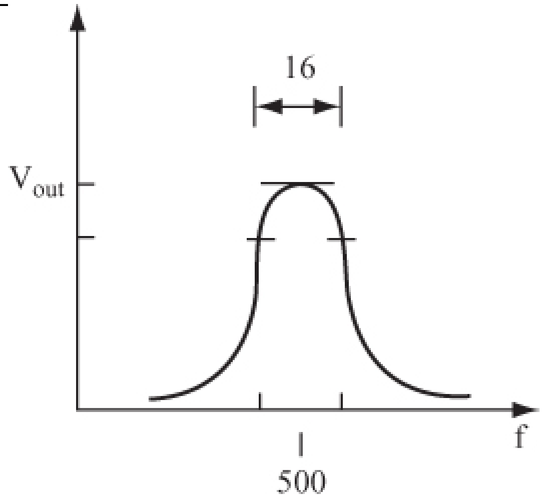

23 Use the circuit shown in Figure 7.37 for this problem. In this case, the resistor value is 10 ohms. However, the inductor and capacitor values are not given.

Figure 7.37

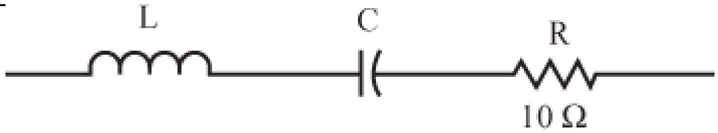

Questions

Find BW and the values of L and C required to give the circuit a resonant frequency of 1200 Hz and a Q of 80.

A. BW = _____

B. L = _____

C. C = _____

Answers

A. BW = 15 Hz

B. L = 106 mH

C. C = 0.166 μF

You can check these values by using the values of L and C to find f_r.

24 Use the circuit shown in Figure 7.37 for this problem. In this case, the resistor value is given as 10 ohms. However, the inductor and capacitor values are not given.

Questions

Calculate the values of Q, L, and C required to give the circuit a resonant frequency of 300 kHz with a bandwidth of 80 kQ.

A. Q = _____

B. L = _____

C. C = _____

Answers

A. Q = 3.75

B. L = 20 μH

C. C = 0.014 μF

25 A circuit that passes (or blocks) only a narrow range of frequencies is called a *high Q* circuit. Figure 7.38 shows the output curve for a high Q circuit.

Figure 7.38

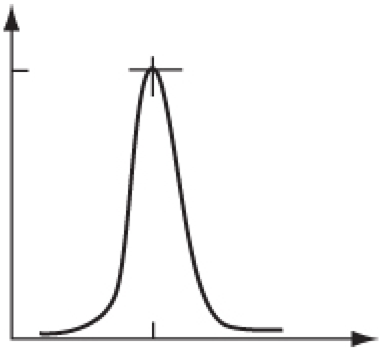

Because of the narrow range of frequencies it passes, a high Q circuit is said to be *selective* in the frequencies it passes.

A circuit that passes (or blocks) a wide range of frequencies is called a *low Q circuit* . Figure 7.39 shows the output curve for a low Q circuit.

Figure 7.39

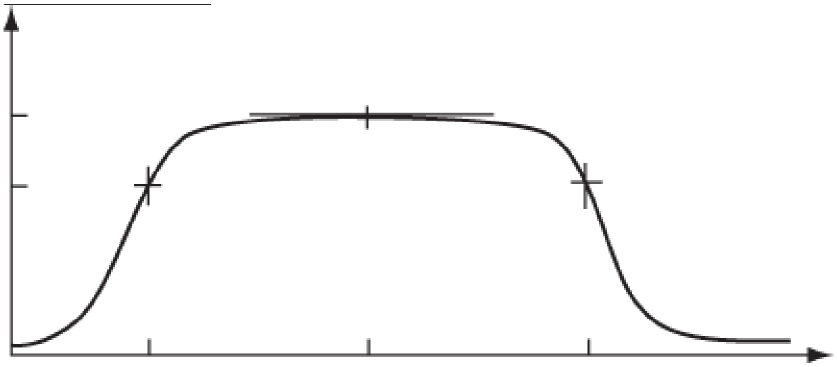

Recall the discussion in problem 18 (comparing the bandwidths of radio tuners and television amplifiers) to help you answer the following questions.

Questions

A. Which is the more selective, the radio tuner or the television amplifier? _____

B. Which would require a lower Q circuit, the radio tuner or the television amplifier? _____

Answers

A. The radio tuner

B. The television amplifier

26 The inductor and capacitor shown in Figure 7.40 are connected in parallel, rather than in series. However, you can use the same formulas you used for the series circuit

in problem 20 to calculate f_r, Q, and BW for parallel LC circuits.

Figure 7.40

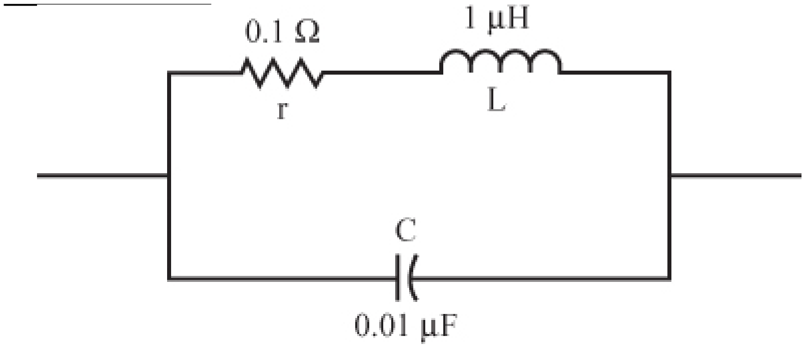

Questions

Find f_r, Q, and BW for the circuit shown in Figure 7.40.

A. f_r = _____

B. Q = _____

C. BW = _____

Answers

A. f_r = 1.6 MHz

B. X_L = 10 ohms, so Q = 10/0.1 = 100 (The only resistance here is the small DC resistance of the inductor.)

C. BW = 16 kHz (This is a fairly high Q circuit.)

27 In the last few problems, you learned how to calculate f_r, BW, and Q, for a given circuit, or conversely, to calculate the component values that would produce a circuit with specified f_r, BW, and Q values.

When you know the resonant frequency and bandwidth for a circuit, you can sketch an approximate output curve. With the simple calculations listed in this problem, you can draw a curve that is accurate to within 1 percent of its true value.

The curve that results from the calculations used in this problem is sometimes called the *general resonance curve*.

You can determine the output voltage at several frequencies by following these steps:

1. Assume the peak output voltage V_p at the resonant frequency f_r to be 100 percent. This is point A on the curve shown in Figure 7.41.
2. The output voltage at f_1 and f_2 is 0.707 of 100 percent. On the graph, these are the two points labeled B in Figure 7.41. Note that $f_2 - f_1 = BW$. Therefore, at half a bandwidth above and below f_r, the output is 70.7 percent of V_p.
3. At f_3 and f_4 (the two points labeled C in Figure 7.41), the output voltage is 44.7 percent of V_p. Note that $f_4 - f_3 = 2\ BW$. Therefore, at 1 bandwidths above and below f_r, the output is 44.7 percent of maximum.
4. At f_5 and f_6 (the two points labeled D in Figure 7.41), the output voltage is 32 percent of V_p. Note that $f_6 - f_5 = 3\ BW$. Therefore, at 1.5 bandwidths above and below f_r, the output is 32 percent of maximum.
5. At f_7 and f_8 (the two points labeled E in

Figure 7.41), the output voltage is 24 percent of V_p. Note that $f_8 - f_7 = 4$ BW. Therefore, at 2 bandwidths above and below f_r, the output is 24 percent of maximum.

6. At f_{10} and f_9 (the two points labeled F in Figure 7.41), the output is 13 percent of V_p. Note that $f_{10} - f_9 = 8$ BW. Therefore, at 4 bandwidths above and below f_r, , the output is 13 percent of maximum.

Figure 7.41

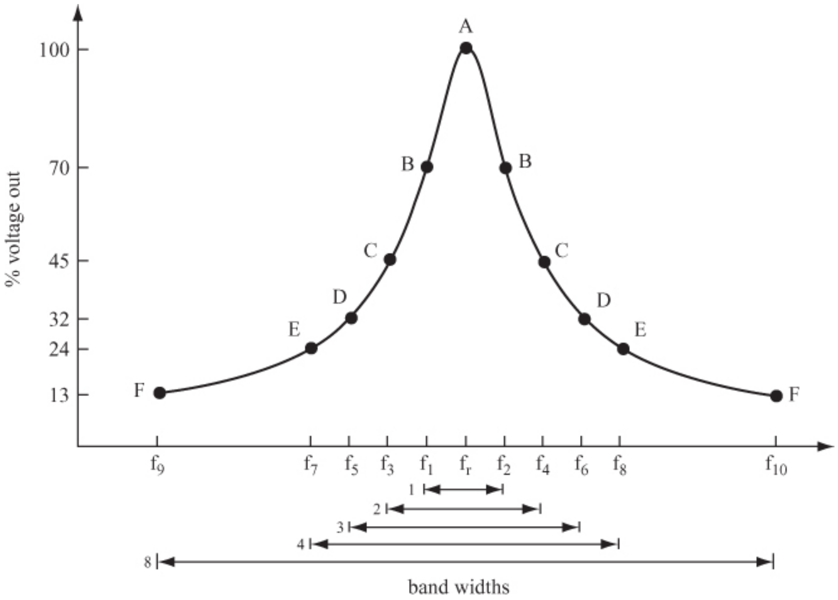

Questions

Calculate f_r, X_L, Q, and BW for the circuit shown in Figure 7.42 .

Figure 7.42

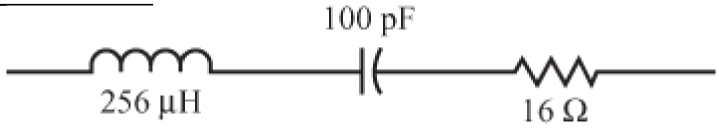

A. f_r = _____

B. X_L = _____

C. Q = _____

D. BW = _____

Answers

A. f_r = 1 MHz

B. X_L = 1607 ohms

C. Q = 100

D. BW = 10 kHz

28 Now, calculate the frequencies that correspond with each percentage of the peak output voltage listed in steps 1 through 6 of problem 27. (Refer to the graph in Figure 7.41 as needed.)

Questions

A. At what frequency will the output level be maximum? _____

B. At what frequencies will the output level be 70 percent of V_p? _____

C. At what frequencies will the output level be 45 percent of V_p? _____

D. At what frequencies will the output level be 32 percent of V_p? _____

E. At what frequencies will the output level be 24 percent of V_p? _____

F. At what frequencies will the output level be 13 percent of V_p? _____

Answers

A. 1 MHz

B. 995 kHz and 1005 kHz (1 MHz – 5 kHz and + 5 kHz)

C. 990 kHz and 1010 kHz
D. 985 kHz and 1015 kHz
E. 980 kHz and 1020 kHz
F. 960 kHz and 1040 kHz

29 You can calculate the output voltage at each frequency in the answers to problem 28 by multiplying the peak voltage by the related percentage for each frequency.

Questions

Calculate the output voltage for the frequencies given here, assuming that the peak output voltage is 5 volts.

A. What is the output voltage level at 995 kHz? _____

B. What is the output voltage level at 980 kHz? _____

Answers

A. V = 5 volts × 0.70 = 3.5 volts

B. V = 5 volts × 0.24 = 1.2 volts

Figure 7.43 shows the output curve generated by plotting the frequencies calculated in problem 28 and the corresponding output voltages calculated in this problem.

Figure 7.43

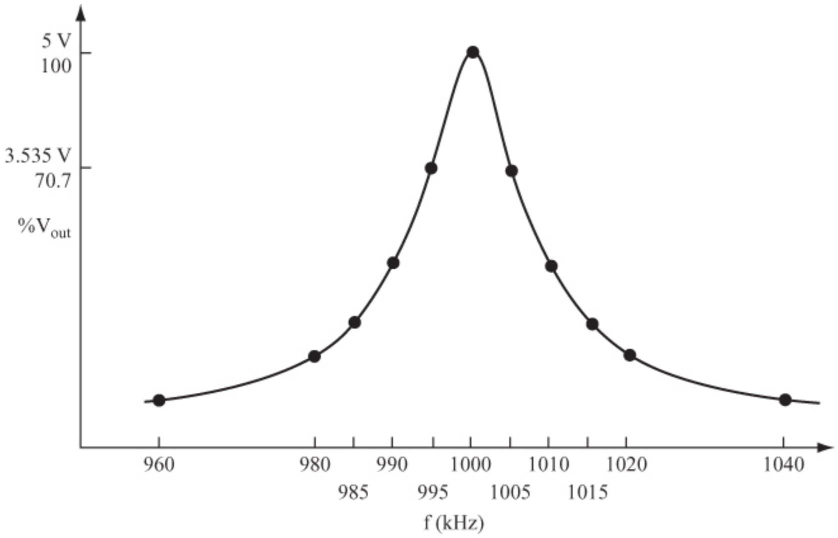

Introduction to Oscillators

In addition to their use in circuits used to filter input signals, capacitors and inductors are used in circuits called *oscillators* .

Oscillators are circuits that generate waveforms at particular frequencies. Many oscillators use a tuned parallel LC circuit to produce a sine wave output. This section is an introduction to the use of parallel capacitors and inductors in oscillators.

30 When the switch in the circuit shown in drawing (1) of Figure 7.44 is closed, current flows through both sides of the parallel LC circuit in the direction shown.

Figure 7.44

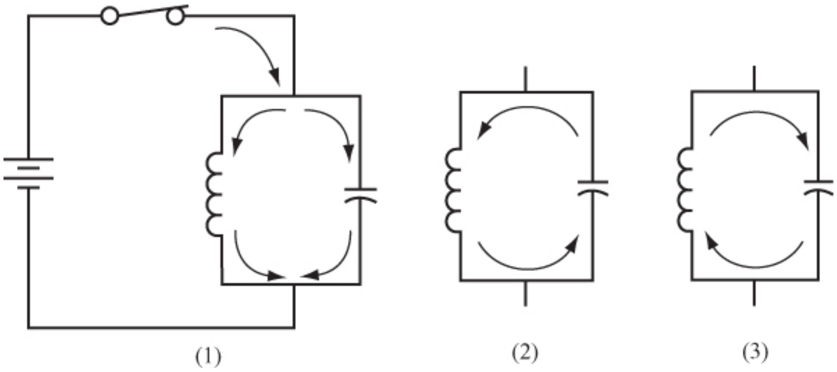

It is difficult for the current to flow through the inductor initially because the inductor opposes any changes in current flow. Conversely, it is easy for the current to flow into the capacitor initially because with no charge on the plates of the capacitor there is no opposition to the flow.

As the charge on the capacitor increases, the current flow in the capacitor side of the circuit decreases. However, more current flows through the inductor. Eventually, the capacitor is fully charged, so current stops flowing in the capacitor side of the circuit, and a steady current flows through the inductor.

Question

When you open the switch, what happens to the charge on the capacitor? _____

Answer

It discharges through the inductor. (Note the current direction, shown in drawing [2] of Figure 7.44 .)

---

31 With the switch open, current continues to flow until the capacitor is fully discharged.

Question

When the capacitor is fully discharged, how much current is flowing through the inductor? _____

Answer

None.

32 Because there is no current in the inductor, its magnetic field collapses. The collapsing of the magnetic field induces a current to flow in the inductor, and this current flows in the same direction as the original current through the inductor (remember that an inductor resists any change in current flow), which is shown in drawing (2) of Figure 7.44 . This current now charges the capacitor to a polarity that is opposite from the polarity that the battery induced.

Question

When the magnetic field of the inductor has fully collapsed, how much current will be flowing? _____

Answer

None.

33 Next, the capacitor discharges through the inductor again, but this time the current flows in the opposite direction, as shown in drawing (3) of Figure 7.44 . The change in current direction builds a magnetic field of the opposite polarity. The magnetic field stops growing when the capacitor is fully

discharged.

Because there is no current flowing through the inductor, its magnetic field collapses and induces current to flow in the direction shown in drawing (3) of Figure 7.44.

Question

What do you think the current generated by the magnetic field in the inductor will do to the capacitor? _____

Answer

It charges it to the original polarity.

34 When the field has fully collapsed, the capacitor stops charging. It now begins to discharge again, causing current to flow through the inductor in the direction shown in drawing (2) of Figure 7.44. This “seesaw” action of current will continue indefinitely.

As the current flows through the inductor, a voltage drop occurs across the inductor. The magnitude of this voltage drop will increase and decrease as the magnitude of the current changes.

Question

What would you expect the voltage across the inductor to look like when you view it on an oscilloscope? _____

Answer

A sine wave

35 In a perfect circuit, this oscillation continues and produces a continuous sine wave. In practice, a small amount of power is

lost in the DC resistance of the inductor and the other wiring. As a result, the sine wave gradually decreases in amplitude and dies out to nothing after a few cycles, as shown in Figure 7.45 .

Figure 7.45

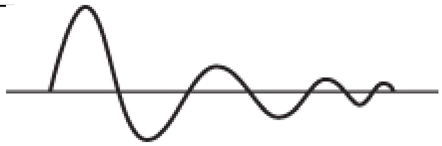

Question

How might you prevent this fade-out? _____

Answer

By replacing a small amount of energy in each cycle.

This lost energy can be injected into the circuit by momentarily closing and opening the switch at the correct time. (See drawing [1] of Figure 7.44 .) This would sustain the oscillations indefinitely.

An electronic switch (such as a transistor) could be connected to the inductor as shown in Figure 7.46 . Changes in the voltage drop across the inductor would turn the electronic switch on or off, thereby opening or closing the switch.

Figure 7.46

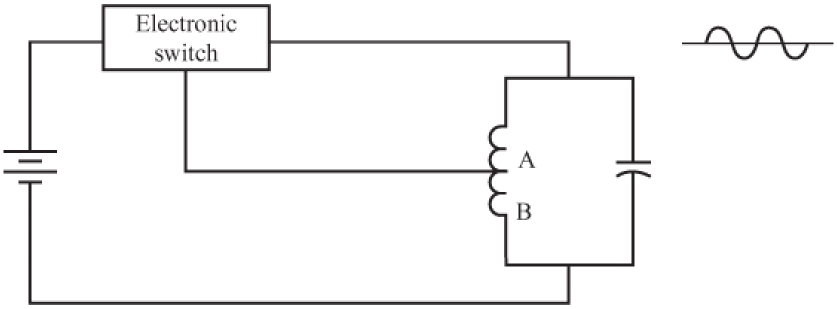

The small voltage drop across the few turns of the inductor (also referred to as a *coil*), between point B at the end of the coil, and point A about halfway along the coil, is used to operate the electronic switch. These points are shown in Figure 7.30 .

Using a small part of an output voltage in this way is called *feedback* because the voltage is "fed back" to an earlier part of the circuit to make it operate correctly.

When you properly set up such a circuit, it produces a continuous sine wave output of constant amplitude and constant frequency. This circuit is called an *oscillator* . You can calculate the frequency of the sine waves generated by an oscillator with the following formula for determining resonant frequency:

$$f = \frac{1}{2\pi\sqrt{LC}}$$

The principles you learned in the last few problems are used in practical oscillator circuits, such as those presented in Chapter 9.

Summary

In this chapter, you learned about the following topics related to resonant circuits:

How the impedance of a series LC circuit and a parallel LC circuit changes with changes in frequency.

At resonant frequency for a parallel LC circuit, the impedance is at its highest; whereas for a series LC circuit, impedance is at its lowest.

The concept of bandwidth enables you to easily calculate the output voltage at various frequencies and draw an accurate output curve.

The principles of bandpass filters and notch (or band-reject) filters.

The fundamental concepts integral to understanding how an oscillator functions.

Self-Test

These questions test your understanding of the concepts covered in this chapter. Use a separate sheet of paper for your drawings or calculations. Compare your answers with the answers provided following the test.

1. What is the formula for the impedance of a series LC circuit? _____

2. What is the formula for the impedance of a series RLC circuit (a circuit containing resistance, inductance, and capacitance)?

_____

3. What is the relationship between X_C and X_L at the resonant frequency? _____

4. What is the voltage across the resistor in a series RLC circuit at the resonant frequency? _____

5. What is the voltage across a resistor in series with a parallel LC circuit at the resonant frequency? _____

6. What is the impedance of a series circuit at resonance? _____

7. What is the formula for the impedance of a parallel circuit at resonance? _____

8. What is the formula for the resonant frequency of a circuit? _____

9. What is the formula for the bandwidth of a circuit? _____

10. What is the formula for the Q of a circuit? _____

Questions 11–13 use a series LC circuit. In each case, the values of the L, C, and R are given. Find f_r, X_L, X_C, Z, Q, and BW. Draw an output curve for each answer.

11. L = 0.1 mH, C = 0.01 *m* F, R = 10 ohms _____

12. L = 4 mH, C = 6.4 *m* F, R = 0.25 ohms _____

13. L = 16 mH, C = 10 *m* F, R = 20 ohms _____

Questions 14 and 15 use a parallel LC circuit. No R is used; r is given. Find f_r, X_L, X_C, Z, Q, and BW.

14. L = 6.4 mH, C = 10 *m* F, r = 8 ohms _____

15. L = 0.7 mH, C = 0.04 *m* F, r = 1.3 ohms _____

16. Use the output curve shown in Figure 7.47 to answer the following questions.

Figure 7.47

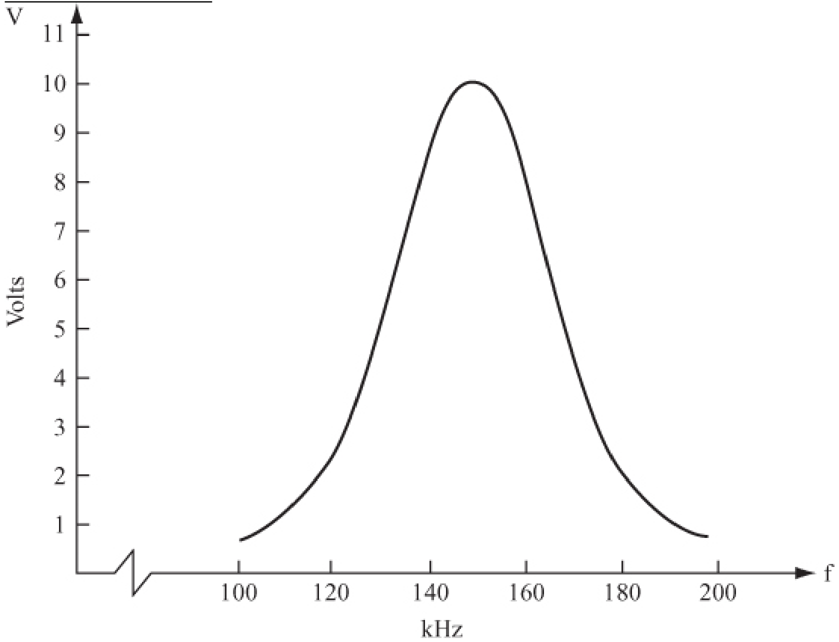

A. What is the peak value of the output curve? _____

B. What is the resonant frequency? _____

C. What is the voltage level at the half power points? _____

D. What are the half power frequencies? _____

E. What is the bandwidth? _____

F. What is the Q of the circuit? _____

Answers to Self-Test

If your answers do not agree with those given here, review the problems indicated in parentheses before you go on to Chapter 8, “Transistor Amplifiers.”

1. $Z = X_L - X_C$ (problem 2)
2. $Z = \sqrt{(X_L \quad X_C)^2 + R^2}$ (problem 2)
3. $X_L = X_C$ (problem 5)
4. Maximum output (problem 5)
5. Minimum output (problem 11)
6. Z = minimum. Ideally, it is equal to the resistance. (problem 5)
7. $Z = \frac{L}{Cr}$ (problem 10)

   In this formula, r is the resistance of the coil.
8. $f_r = \frac{1}{2\pi\sqrt{LC}}$ (problem 6)
9. $BW = \frac{f_r}{Q}$ (problem 20)
10. $Q = \frac{X_L}{R}$ (problem 20)

or

$$\frac{X_L}{r}$$

To draw the output curves for Questions 11–13, use the graph in Figure 7.41 as a guide and insert the appropriate bandwidth and frequency values. (problems 21–29)

11. $f_r = 160$ kHz, $X_L = X_C = 100$ ohms, $Q = 10$, $BW = 16$ kHz, $Z = 10$ ohms (problems 21–29)

12. $f_r = 1$ kHz, $X_L = X_C = 25$ ohms, $Q = 100$, $BW = 10$ Hz, $Z = 0.25$ ohms (problems 21–29)

13. $f_r = 400$ Hz, $X_L = X_C = 40$ ohms, $Q = 2$, $BW = 200$ Hz, $Z = 20$ ohms (problems 21–29)

14. $f_r = 600$ Hz, $X_L = 24$ ohms, $X_C = 26.5$, $Q = 3$, $BW = 200$ Hz, $Z = 80$ ohms (problems 21–29)

Because Q is not given, you should use the more complicated of the two formulas shown in problem 10 to calculate the resonant frequency.

15. $f_r = 30$ kHz, $X_L = 132$ ohms, $X_C = 132$, $Q = 101.5$, $BW = 300$ Hz, $Z = 13.4$ ohms (problems 21–29)

16A. 10.1 volts (problems 27 and 28)

16B. 148 kHz (problems 27 and 28)

16C. 10.1 × 0.707 = 7.14 volts (problems 27 and 28)

16D. Approximately 135 kHz and 160 kHz (not quite symmetrical) (problems 27 and 28)

16E. BW = 25 kHz (problems 27 and 28)

16F. $Q = \frac{f_r}{BW} =$ about 5.9 (problems 27 and 28)

Chapter 8

Transistor Amplifiers

Many of the AC signals you'll work with in electronics are small. For example, the signal that an optical detector reads from a DVD disk cannot drive a speaker, and the signal from a microphone's output is too weak to send out as a radio signal. In cases such as these, you must use an amplifier to boost the signal.

The best way to demonstrate the basics of amplifying a weak signal to a usable level is by starting with a one-transistor amplifier. When you understand a one-transistor amplifier, you can grasp the building block that makes up amplifier circuits used in electronic devices such as cellphones, MP3 players, and home entertainment centers.

Many amplifier circuit configurations are possible. The simplest and most basic of amplifying circuits are used in this chapter to demonstrate how a transistor amplifies a signal. You can also see the steps to design an amplifier.

The emphasis in this chapter is on the bipolar junction transistor (BJT), just as it was in Chapter 3, “Introduction to the Transistor,” and Chapter 4, “The Transistor Switch,” which dealt primarily with the application of transistors in switching circuits. Two other types of devices used as amplifiers are also examined: the junction field effect transistor (JFET) (introduced in Chapters 3 and 4), and an integrated circuit called the *operational amplifier* (*op-amp*).

When you complete this chapter, you will be able to do the following:

Calculate the voltage gain for an amplifier.

Calculate the DC output voltage for an amplifier circuit.

Select the appropriate resistor values to provide the required gain to an amplifier circuit.

Identify several ways to increase the gain of a one-transistor amplifier.

Distinguish between the effects of a standard one-transistor amplifier and an emitter follower circuit.

Design a simple emitter follower circuit.

Analyze a simple circuit to find the DC level out and the AC gain.

Design a simple common source (JFET) amplifier.

Analyze a JFET amplifier to find the AC gain.

Recognize an op-amp and its connections.

Working with Transistor Amplifiers

1 In Chapter 3 you learned how to turn transistors ON and OFF. You also learned how to calculate the value of resistors in amplifier circuits to set the collector DC voltage to half the power supply voltage. To review this concept, examine the circuit shown in Figure 8.1.

Figure 8.1

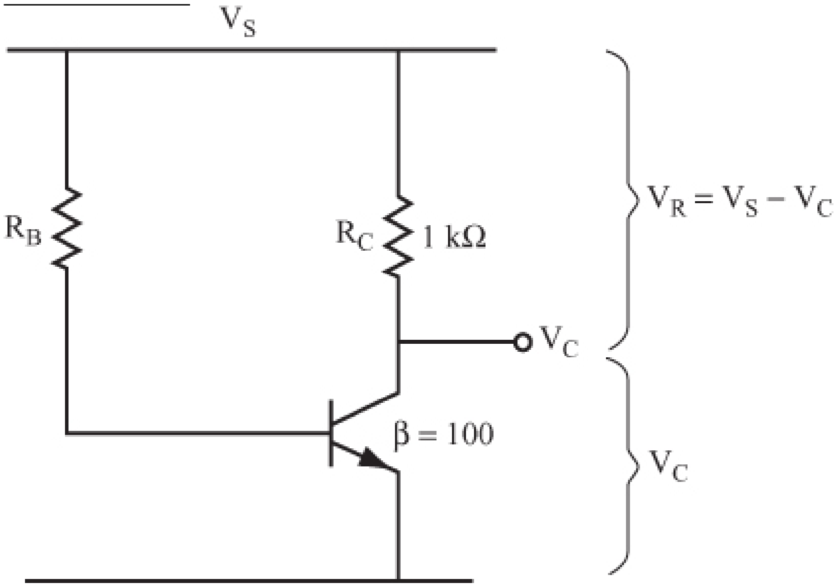

Use the following steps to find the value of R_B that will set the collector DC voltage (V_C) to half the supply voltage (V_S):

1. Find I_C by using the following equation:

$$I_C = \frac{V_R}{R_C} = \frac{V_S - V_C}{R_C}$$

2. Find I_B by using the following equation:

$$I_B = \frac{I_C}{\beta}$$

3. Find R_B by using the following equation:

$$R_B = \frac{V_S}{I_B}$$

Questions

Find the value of R_B that will set the collector voltage to 5 volts, using steps 123 and the following values for the circuit:

$$V_S = 10\,\text{volts}, \quad R_C = 1\,\text{k}\Omega, \quad \beta = 100$$

A. I_C = _____

B. I_B = _____

C. R_B = _____

Answers

A. $$I_C = \frac{5\,\text{volts}}{1\,\text{k}\Omega} = 5\,\text{mA}$$

B. $$I_B = \frac{5}{100} = 0.05\,\text{mA}$$

C. $$R_B = \frac{10\,\text{volts}}{0.05\,\text{mA}} = 200\,\text{k}\Omega$$

2 You have seen that using a 200 kΩ resistor for R_B gives an output level of 5 volts at the collector. This procedure of setting the output DC level is called *biasing* . In problem 1, you biased the transistor to a 5-volt DC output.

Use the circuit shown in Figure 8.1 and the formulas given in problem 1 to answer the following questions.

Questions

A. If you decrease the value of R_B, how do I_B, I_C, V_R, and the bias point V_C change? _____

B. If you increase the value of R_B, how do I_B, I_C, V_R, and V_C change? _____

Answers

A. I_B increases, I_C increases, V_R increases, and so the bias point V_C decreases.

B. I_B decreases, I_C decreases, V_R decreases, and so the bias point V_C increases.

3 In problem 2, you found that changing the value of R_B in the circuit shown in Figure 8.1 changes the value of I_B.

The transistor amplifies slight variations in I_B. Therefore, the amount I_C fluctuates is β times the change in value in I_B.

The variations in I_C cause changes in the voltage drop V_R across R_C. Therefore, the output voltage measured at the collector also changes.

Questions

For the circuit shown in Figure 8.1, calculate the following parameters when R_B = 168 kΩ and V_S = 10 volts:

A. $I_B = \frac{V_C}{R_B} =$ _____

B. $I_C = \beta I_B =$ _____

C. $V_R = I_C R_C =$ _____

D. $V_C = V_S 2V_R =$ _____

Answers

A. $I_B = \frac{10\,\text{volts}}{168\,\text{k}\Omega} = 0.059\,\text{mA}$

B. I_C = 100 × 0.059 = 5.9 mA

C. V_R = 1 kΩ × 5.9 mA = 5.9 volts

D. V_C = 10 volts 2 5.9 volts = 4.1 volts

4 Use the circuit shown in Figure 8.1 to answer the following questions when V_S = 10 volts.

Questions

Calculate V_C for each of the following values of R_B:

A. 100 kΩ _____

B. 10 MΩ _____

C. 133 kΩ _____

D. 400 kΩ _____

Answers

A. I_B = 0.1 mA, I_C = 10 mA, V_C = 0 volts

B. I_B = 1 μA, I_C = 0.1 mA, V_C = 10 volts (approximately)

C. I_B = 0.075 mA, I_C = 7.5 mA, V_C = 2.5 volts

D. I_B = 0.025 mA, I_C = 2.5 mA, V_C = 7.5 volts

5 The values of I_C and V_C that you calculated in problems 1 and 4 are plotted on the graph on the left side of Figure 8.2. The straight line connecting these points on the graph is called the *load line*.

Figure 8.2

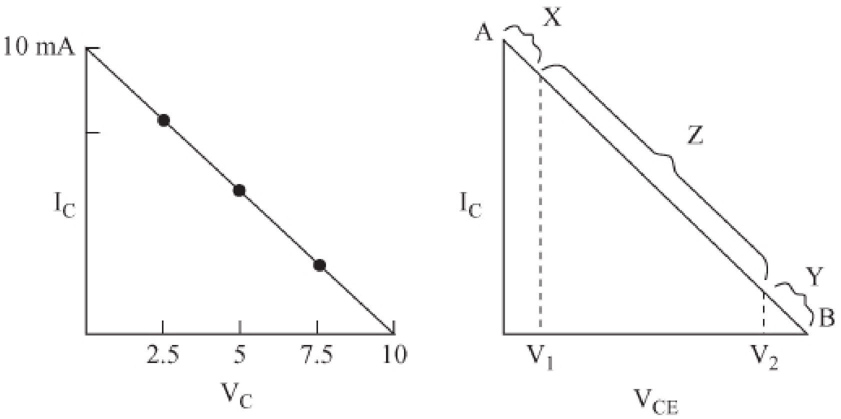

The axis labeled V_C represents the voltage between the collector and the emitter of the transistor, and not the voltage between the collector and ground. Therefore, this axis should correctly be labeled V_{CE}, as shown in the graph on the right of the figure. (For this circuit, $V_{CE} = V_C$ because there is no resistor between the emitter and ground.)

Questions

A. At point A in the graph on the right, is the transistor ON or OFF? _____

B. Is it ON or OFF at point B? _____

Answers

A. ON because full current flows, and the transistor acts like a short circuit. The voltage drop across the transistor is 0 volts.

B. OFF because essentially no current flows, and the transistor acts like an open circuit. The voltage drop across the transistor is at its maximum (10 volts, in this case).

6 Point A on the graph shown in Figure 8.2 is called the *saturated point* (or the *saturation point*) because it is at that point that the collector current is at its maximum.

Point B on the graph shown in Figure 8.2 is often called the *cutoff* point because, at that point, the transistor is OFF and no collector current flows.

In regions X and Y, the gain (β) is not constant, so these are called the *nonlinear regions*. Note that $\beta = I_C/I_B$. Therefore, β is the slope of the line shown in the graph.

As a rough guide, V_1 is approximately 1 volt, and V_2 is approximately 1 volt less than the voltage at point B.

Question

What is the value of V_{CE} at point B? _____

Answer

$V_{CE} = V_S$, which is 10 volts in this case.

7 In region Z of the graph shown in Figure 8.2, β (that is, the slope of the graph) is constant. Therefore, this is called the *linear region*. Operating the transistor in the linear region results in an output signal that is free of distortion.

Question

Which values of I_C and V_C would result in an undistorted output in the circuit shown in Figure 8.1?

A. I_C = 9 mA, V_C = 1 volt

B. I_C = 1 mA, V_C = 9 volts

C. I_C = 6 mA, V_C = 4.5 volts _____

Answer

C is the only one. A and B fall into nonlinear regions.

8 If you apply a small AC signal to the base of the transistor after it has been biased, the small voltage variations of the AC signal (shown in Figure 8.3 as a sine wave) cause small variations in the base current.

Figure 8.3

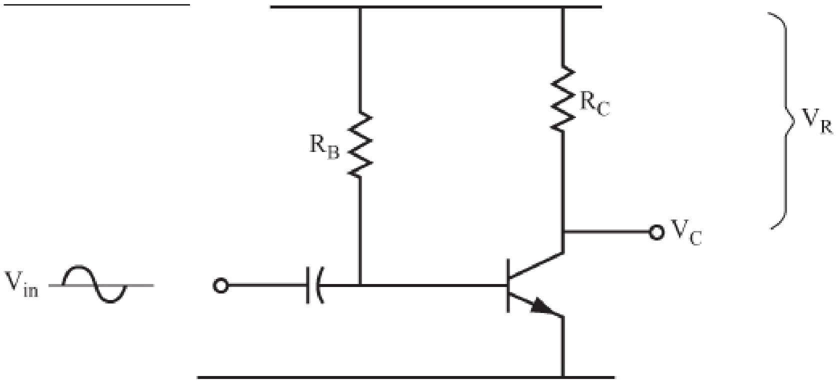

These variations in the base current will be amplified by a factor of β and will cause corresponding variations in the collector current. The variations in the collector current, in turn, will cause similar variations in the collector voltage.

The β used for AC gain calculations is different from the β used in calculating DC variations. The AC β is the value of the common emitter AC forward current transfer ratio, which is listed as h fe in the manufacturer's data sheets for the transistor. Use the AC β whenever you need to calculate the AC output for a given AC input, or to determine an AC current variation. Use the DC β to calculate the base or collector DC current values. You must know which β to use, and remember that one is used for DC, and the other is used for AC variations. The DC β is sometimes called h_{FE} or β_{dc}.

As V_{in} increases, the base current increases, which causes the collector current to increase. An increase in the collector current increases the voltage drop across R_C, which causes V_C to decrease.

Note The capacitor shown at the input blocks DC (infinite reactance) and easily passes AC (low reactance). This is a common isolation technique used at the input and output of AC circuits.

Questions

A. If the input signal decreases, what happens to the collector voltage? _____

B. If you apply a sine wave to the input, what waveform would you expect at the collector? _____

Answers

A. The collector voltage, V_C, increases.

B. A sine wave, but inverted as shown in Figure 8.4.

Figure 8.4

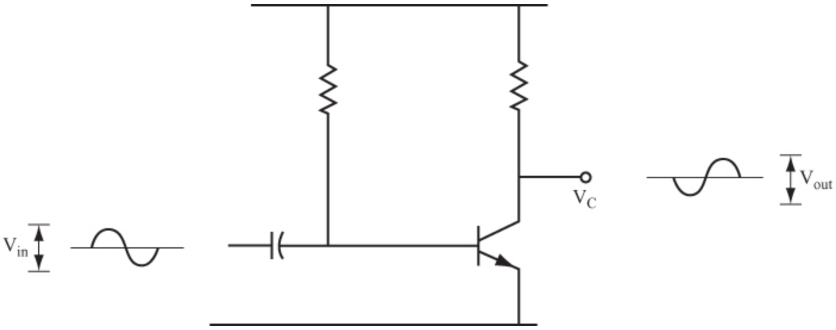

9 Figure 8.4 shows the input and output sine waves for an amplifier circuit.

The input voltage V_{in} is applied to the base. (Strictly speaking, it is applied across the base-emitter diode.) The voltage variations at the collector are centered on the DC bias point V_C, and they will be larger than variations in the input voltage. Therefore, the output sine wave is larger than the input sine wave (that is, amplified).

This amplified output signal at the collector can be used to drive a load (such as a speaker).

To distinguish these AC variations in output from the DC bias level, you indicate the AC output voltage by V_{out}. In most cases, V_{out} is a peak-to-peak value.

Questions

A. What is meant by V_C? _____

B. What is meant by V_{out}? _____

Answers

A. Collector DC voltage, or the bias point

B. AC output voltage

The ratio of the output voltage to the input voltage is called the *voltage gain* of the amplifier.

$$\text{Voltage gain} = A_V = \frac{V_{out}}{V_{in}}$$

To calculate the voltage gain of an amplifier, you can measure V_{in} and V_{out} with an oscilloscope. Measure peak-to-peak voltages for this calculation.

10 For the circuit shown in Figure 8.4 you can calculate the voltage gain using the following formula:

$$A_V = \beta \times \frac{R_L}{R_{in}}$$

In this equation:

R_L is the *load resistance* . In this circuit, the collector resistor, R_C, is the load resistance.

R_{in} is the *input resistance* of the transistor. You can find R_{in} (often called h_{ie}) on the data or specification sheets from the manufacturer. In most transistors, input resistance is approximately 1 kΩ to 2 kΩ.

You can find V_{out} by combining these two voltage gain equations:

$$A_V = \frac{V_{out}}{V_{in}} \quad \text{and} \quad A_V = \beta \times \frac{R_L}{R_{in}}$$

$$\text{Therefore, } \frac{V_{out}}{V_{in}} = \beta \times \frac{R_L}{R_{in}}$$

Solving this for V_{out} results in the following equation. Here, the values of R_{in} = 1 kΩ, V_{in} = 1 mV, R_C = 1 kΩ, and β = 100 were used to perform this sample calculation.

$$V_{out} = V_{in} \times \beta \times \frac{R_L}{R_{in}}$$

$$= 1\,\text{mV} \times 100 \times \frac{1\,\text{k}\Omega}{1\,\text{k}\Omega}$$

$$= 100\,\text{mV}$$

Questions

A. Calculate V_{out} if R_{in} = 2 kΩ, V_{in} = 1 mV, R_C = 1 kΩ, and β = 100. _____

B. Find the voltage gain in both cases. _____

Answers

A. V_{out} = 50 mV

B. A_V = 100 and A_V = 50

This simple amplifier can provide voltage gains of up to approximately 500. But it does have several faults that limit its practical usefulness.

Because of variations in β between transistors, V_C changes if the transistor is changed. To compensate for this, you must adjust R_B.

R_{in} or h_{ie} varies greatly from transistor to transistor. This variation, combined with variations in β, means that you cannot guarantee the gain from one transistor amplifier to another.

Both R_{in} and β change greatly with temperature; hence the gain is temperature-dependent. For example, a simple amplifier circuit like that discussed in this problem was designed to work in the desert in July. It would fail completely in Alaska in the winter. If the amplifier worked perfectly in the lab, it probably would not work outdoors on either a hot or cold day.

Note An amplifier whose gain and DC level bias point change as described in this problem is said to be *unstable* . For reliable operation, an amplifier should be as stable as possible. In later problems, you see how to design a stable amplifier.

A Stable Amplifier

11 You can overcome the instability of the transistor amplifier discussed in the first ten problems of this chapter by adding two resistors to the circuit. Figure 8.5 shows an amplifier circuit to which resistors R_E and R_2 have been added. R_2 , along with R_1 (labeled R_B in the previous circuits), ensures the stability of the DC bias point.

Figure 8.5

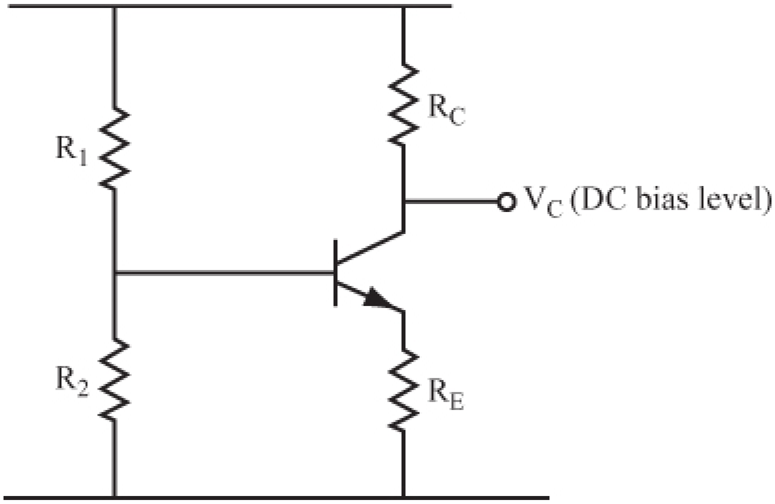
R1
RC
VC (DC bias level)
R2
RE

By adding the emitter resistor R E, you ensure the stability of the AC gain.

The labels in Figure 8.6 identify the DC currents and voltages present in the circuit. These parameters are used in the next several problems.

Figure 8.6

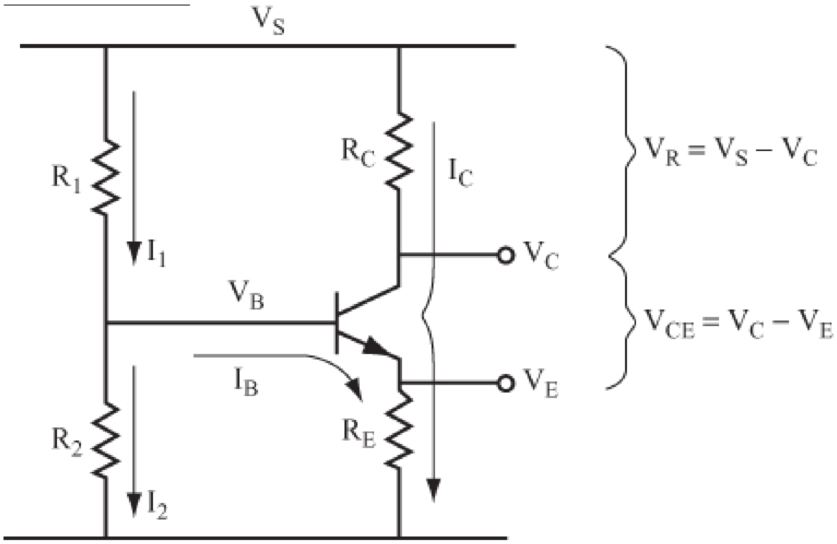
VS
R1
I1
RC
IC
VR = VS – VC
VC
VB
VCE = VC – VE
IB
VE
R2
RE
I2

Question

In designing an amplifier circuit and choosing the resistor values, there are two goals. What are they? _____

Answer

A stable DC bias point, and a stable AC gain

12 Look at the gain first. The gain formula for the circuit shown in Figure 8.6 is as follows:

$$A_V = \frac{V_{out}}{V_{in}} = \frac{R_C}{R_E}$$

This is a slight variation on the formula shown in problem 10. (The complex mathematical justification for this is not important right here.) Here, the AC gain is not affected by transistor β and transistor input resistance, so the AC gain will be constant, regardless of variations in these parameters.

Questions

Use the circuit shown in Figure 8.6 with R_C = 10 kΩ and R_E = 1 kΩ to answer the following questions:

A. What is the AC voltage gain for a transistor if its β = 100? _____

B. What is the gain if β = 500? _____

Answers

A. 10

B. 10

13 This problem provides a couple of examples that can help you understand how to calculate voltage gain and the resulting output voltage.

Questions

A. Calculate the voltage gain (A_V) of the amplifier circuit shown in Figure 8.6 if $R_C = 10\ k\Omega$ and $R_E = 1\ k\Omega$. Then, use A_V to calculate the output voltage if the input signal is $2\ mV_{pp}$. _____

B. Calculate the voltage gain if $R_C = 1\ k\Omega$ and $R_E = 250$ ohms. Then, use A_V to calculate the output voltage if the input signal is $1\ V_{pp}$. _____

Answers

A.
$$A_V = \frac{R_C}{R_E} = \frac{10\,k\Omega}{1\,k\Omega} = 10$$

$$V_{out} = 10 = V_{in} = 20\,mV$$

B.
$$A_V = \frac{1\,k\Omega}{250\,ohms} = 4$$

$$V_{out} = 4\,V_{pp}$$

Although the amplifier circuit shown in Figure 8.6 produces stable values of voltage gain, it does not produce high values of voltage gain. For various reasons, this circuit is limited to voltage gains of 50 or less. Later, this chapter discusses an amplifier circuit that can produce higher values of voltage gain.

14 Before you continue, look at the current relationships in the amplifier circuit shown in Figure 8.6 and an approximation that is often made. You can calculate the current flowing through the emitter resistor with the following equation:

$$I_E = I_B + I_C$$

In other words, the emitter current is the sum of the base and the collector currents.

I_C is much larger than I_B. You can, therefore, assume that the emitter current is equal to the collector current.

$$I_E = I_C$$

Question

Calculate V_C, V_E, and A_V for the circuit shown in Figure 8.7 with V_S = 10 volts, I_C = 1 mA, R_C = 1 kΩ, and R_E = 100 ohms.

—————

Figure 8.7

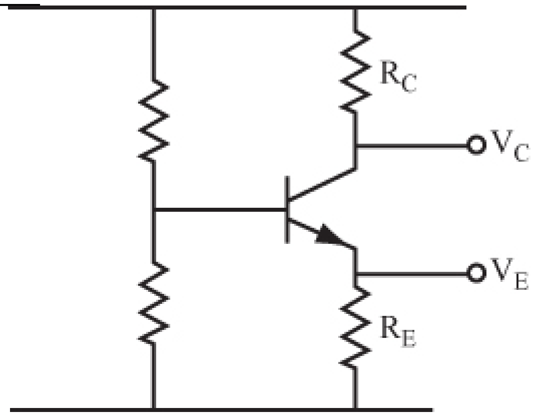

Answer

$$V_R = 1k\Omega \times 1mA = 1 \text{ volt}$$

$$V_C = V_S - V_R = 10 - 1 = 9 \text{ volts}$$

$$V_E = 100 \text{ ohms} \times 1mA = 0.1 \text{ volt}$$

$$A_V = \frac{R_C}{R_E} = \frac{1k\Omega}{100 \text{ ohms}} = 10$$

15 For this problem, use the circuit shown in Figure 8.7 with $V_S = 10$ volts, $I_C = 1$ mA, $R_C = 2$ kΩ, and $R_E = 1$ kΩ.

Question

Calculate V_C, V_E, and A_V. _____

Answers

$$V_R = 2k\Omega \times 1mA = 2 \text{ volts}$$

$$V_C = 10 - 2 = 8 \text{ volts}$$

$$V_E = 1k\Omega \times 1mA = 1 \text{ volt}$$

$$A_V = \frac{R_C}{R_E} = \frac{2k\Omega}{1k\Omega} = 2$$

16 For this problem, use the circuit shown in Figure 8.7 with $V_S = 10$ volts and $I_C = 1$ mA.

Questions

Find V_C, V_E, and A_V for the following values of R_C and R_E:

A. $R_C = 5$ kΩ, $R_E = 1$ kΩ _____

B. R_C = 4.7 kΩ, R_E = 220 ohms _____

Answers

A.

$V_R = 5$ volts, $V_C = 5$ volts, $V_E = 1$ volt, $A_V = 5$

B.

$V_R = 4.7$ volts, $V_C = 5.3$ volts, $V_E = 0.22$ volts, $A_V = 21.36$

Biasing

17 In this problem, you see the steps used to calculate the resistor values needed to bias the amplifier circuit shown in Figure 8.8 .

Figure 8.8

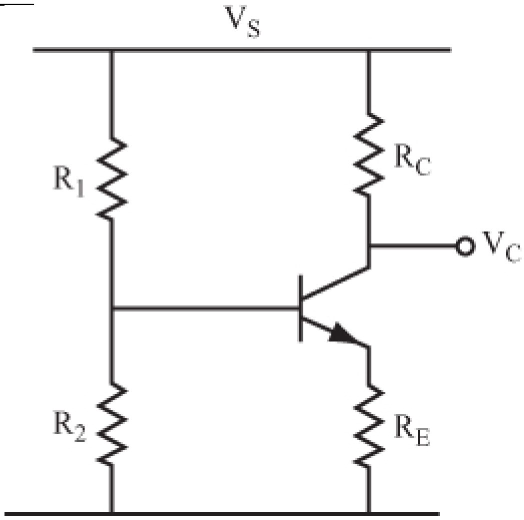

You can determine values for R_1, R_2, and R_E that bias the circuit to a specified DC output voltage and a specified AC voltage gain by using the following steps.

Read the following procedure and the relevant formulas first, and then you will work through an example.

1. Find R_E by using the following:

$$A_V = \frac{R_C}{R_E}.$$

2. Find V_E by using the following:

$$A_V = \frac{V_R}{V_E} = \frac{V_S - V_C}{V_E}.$$

3. Find V_B by using the following:

$$V_B = V_E + 0.7\,\text{volt}$$

4. Find I_C by using the following:

$$I_C = \frac{V_S - V_C}{R_C}.$$

5. Find I_B by using the following:

$$I_B = \frac{I_C}{\beta}$$

6. Find I_2 where I_2 is $10I_B$. (Refer to the circuit shown in Figure 8.6.) This is a convenient rule of thumb that is a crucial step in providing stability to the DC bias point.

7. Find R_2 by using the following:

$$R_2 = \frac{V_B}{I_2}.$$

8. Find R_1 by using the following:

$$R_1 = \frac{V_S - V_B}{I_2 + I_B}.$$

9. Steps 7 and 8 might produce nonstandard values for the resistors, so choose the nearest standard values.

10. Use the voltage divider formula to see if the standard values you chose in step 9

result in a voltage level close to V_B found in step 3. ("Close" means within 10 percent of the ideal.)

This procedure produces an amplifier that works, and results in a DC output voltage and AC gain that are close to those specified at the beginning of the problem.

Questions

Find the values of the parameters specified in each of the following questions for the circuit shown in Figure 8.9 if $A_V = 10$, $V_C = 5$ volts, $R_C = 1$ kΩ, β = 100, and $V_S = 10$ volts.

Figure 8.9

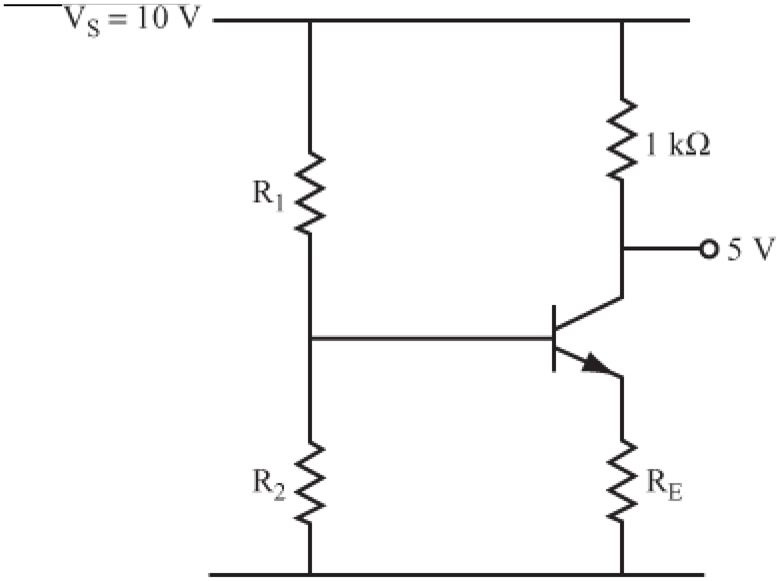

Work through steps 1–10, referring to the steps in this problem for formulas as necessary.

1. Find R_E.

$$A_V = \frac{R_C}{R_E}.\ \text{So}\ R_E = \frac{R_C}{A_V} = \frac{1\text{k}\Omega}{10} = 100\,\text{ohms.}$$

2. V_E = _____
3. V_B = _____
4. I_C = _____
5. I_B = _____
6. I_2 = _____
7. R_2 = _____
8. R_1 = _____
9. Choose the standard resistance values that are closest to the calculated values for R_1 and R_2.

R_1 = _____

R_2 = _____

10. Using the standard resistance values for R_1 and R_2, find V_B.

V_B = _____

Answers

You should have found values close to the following:

1. 100 ohms
2. 0.5 volt
3. 1.2 volts
4. 5 mA
5. 0.05 mA
6. 0.5 mA
7. 2.4 kΩ
8. 16 kΩ
9. 2.4 kΩ and 16 kΩ are standard values. (They are 5 percent values.) Alternative acceptable values would be 2.2 kΩ and 15

kΩ.

10. With 2.4 kΩ and 16 kΩ, V_B = 1.3 volts. With 2.2 kΩ and 15 kΩ, V_B = 1.28 volts. Either value of V_B is within 10 percent of the 1.2 volts calculated for V_B in step 3.

Figure 8.10 shows an amplifier circuit using the values you calculated in this problem for R_1, R_2, and R_E.

Figure 8.10

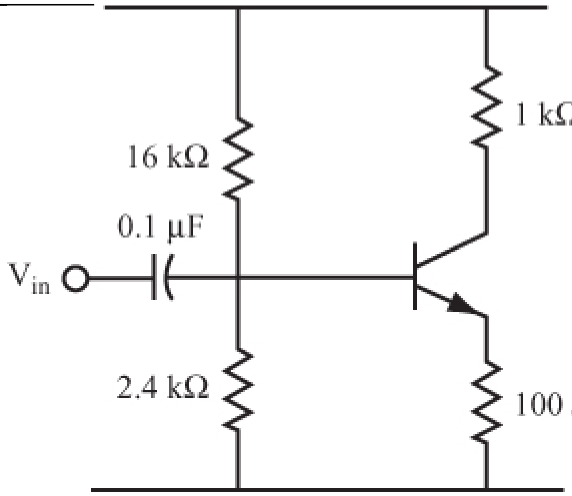

18 Follow the steps in problem 17 to answer the following questions.

Questions

Find the values of the parameters specified in each question for the circuit shown in Figure 8.9 if A_V = 15, V_C = 6 volts, β = 100, R_C = 3.3 kΩ, and V_S = 10 volts.

1. R_E = _____

2. V_E = _____

3. V_B = _____

4. I_C = _____

5. I_B = _____

6. I_2 = _____

7. R_2 = _____

8. R_1 = _____

9. Choose the standard resistance values that are closest to the calculated values for R_1 and R_2.

R_1 = _____

R_2 = _____

10. Using the standard resistance values for R_1 and R_2, find V_B.

V_B = _____

Answers

Following are the values you should have found:

1. 220 ohms
2. 0.27 volt
3. 0.97 volt (You can use 1 volt if you want.)
4. 1.2 mA
5. 0.012 mA
6. 0.12 mA
7. 8.3 kΩ
8. 68.2 kΩ
9. These are close to the standard values of 8.2 kΩ and 68 kΩ.
10. 1.08 volts using the standard values. This is close enough to the value of V_B calculated in question 3.

Project 8.1: The Transistor Amplifier

Objective

The objective of this project is to demonstrate how AC voltage gain changes when you use resistors of different values and transistors with different current gain in a transistor amplifier circuit.

General Instructions

When the circuit is set up, you measure V_{out} for each set of resistors, and find A_V, using the ratio V_{out}/V_{in}. You also determine a calculated A_V using the ratio R_C/R_E in each case to determine how close the calculated A_V is to the measured A_V. You repeat this measurement with a second transistor for each set of resistors.

Parts List

You need the following equipment and supplies:

- One 1 kΩ, 0.25-watt resistor
- One 100 Ω, 0.25-watt resistor
- One 15 kΩ, 0.25-watt resistor
- One 2.2 kΩ, 0.25-watt resistor
- One 3.3 kΩ, 0.25-watt resistor
- One 220 Ω, 0.25-watt resistor
- One 68 kΩ, 0.25-watt resistor
- One 8.2 kΩ, 0.25-watt resistor
- One 0.1 μF capacitor
- One lab type power supply or 9-volt battery
- One function generator.
- One oscilloscope
- One breadboard
- One 2N3904 transistor
- One PN2222 transistor

Figure 8.11 shows the pinout diagram for 2N3904 and PN2222 transistors.

Figure 8.11

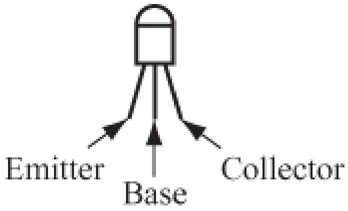

Step-by-Step Instructions

Set up the circuit shown in Figure 8.12 using the components listed for Circuit # 1 in the following table. If you have some experience in building circuits, this schematic (along with the previous parts list) should provide all the information you need to build the circuit. If you need a bit more help building the circuit, look at the photos of the completed circuit in the "Expected Results" section. (If you don't have a lab type power supply to provide 10 volts as indicated on the schematic, use a 9-volt battery.)

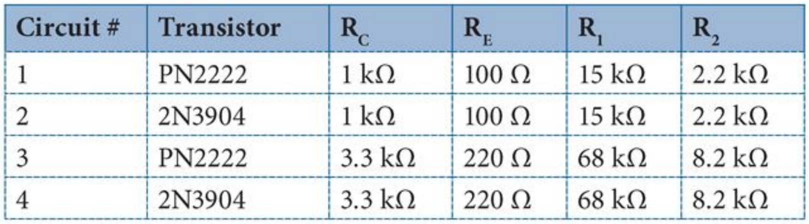

Circuit #	Transistor	R_C	R_E	R_1	R_2
1	PN2222	1 kΩ	100 Ω	15 kΩ	2.2 kΩ
2	2N3904	1 kΩ	100 Ω	15 kΩ	2.2 kΩ
3	PN2222	3.3 kΩ	220 Ω	68 kΩ	8.2 kΩ
4	2N3904	3.3 kΩ	220 Ω	68 kΩ	8.2 kΩ

Figure 8.12

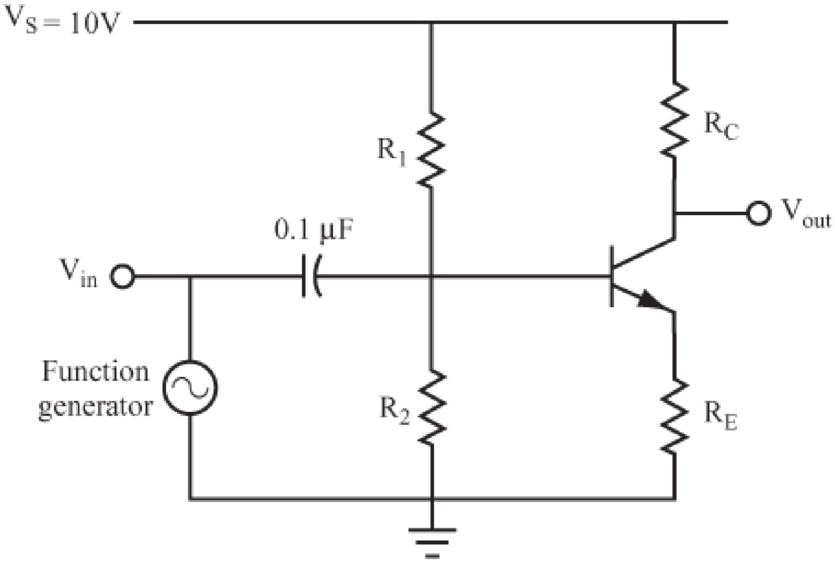

Carefully check your circuit against the diagram.

After you check your circuit, follow these steps, and record your measurements in the blank table following the steps.

1. Connect the oscilloscope probe for channel 2 to a jumper wire connected to V_{in}, and then connect the ground clip to a jumper wire attached to the ground bus.
2. Connect the oscilloscope probe for channel 1 to a jumper wire connected to V_{out}, and then connect the ground clip to a jumper wire attached to the ground bus.
3. Set the function generator to generate a 10 kHz sine wave with approximately 0.2 V_{pp}.
4. Measure and record V_{in} and V_{out}.
5. Change the components to those listed in

the next row of the table (Circuit # 2 in this case.) You should turn off the power to the circuit before changing components to avoid shorting leads together.

6. Measure and record V in and V out .

7. Repeat steps 5 and 6 until you have recorded V in and V out in the last row of the table.

8. Determine β for each of the transistors used in this project. Insert the transistors one at a time into the circuit you built in Project 3-1 to take this measurement.

9. For each transistor, record β in the following table.

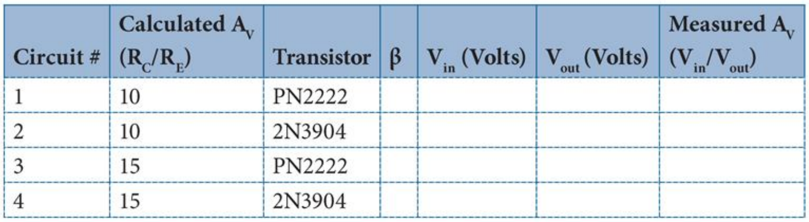

Circuit #	Calculated A_V (R_C/R_E)	Transistor	β	V_{in} (Volts)	V_{out} (Volts)	Measured A_V (V_{in}/V_{out})
1	10	PN2222				
2	10	2N3904				
3	15	PN2222				
4	15	2N3904				

Expected Results

Figure 8.13 shows the breadboarded Circuit # 1.

Figure 8.13

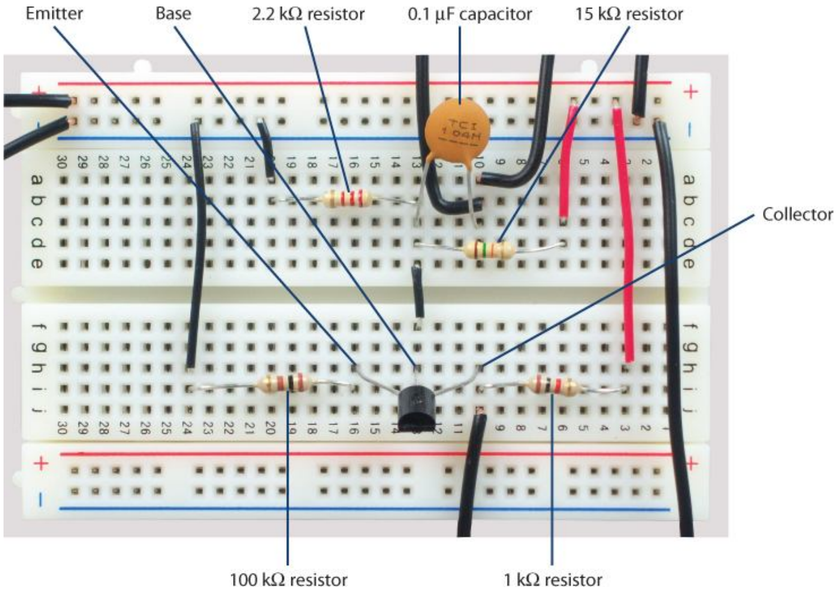
Emitter
Base
2.2 kΩ resistor
0.1 µF capacitor
15 kΩ resistor
Collector
100 kΩ resistor
1 kΩ resistor

Figure 8.14 shows the breadboarded Circuit # 3.

Figure 8.14

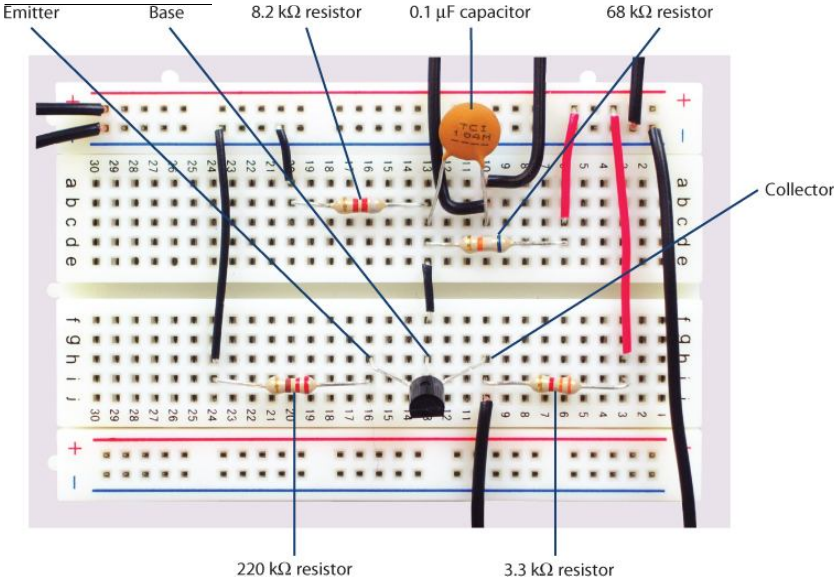
Emitter
Base
8.2 kΩ resistor
0.1 µF capacitor
68 kΩ resistor
Collector
220 kΩ resistor
3.3 kΩ resistor

Figure 8.15 shows a function generator and oscilloscope attached to the circuit.

Figure 8.15

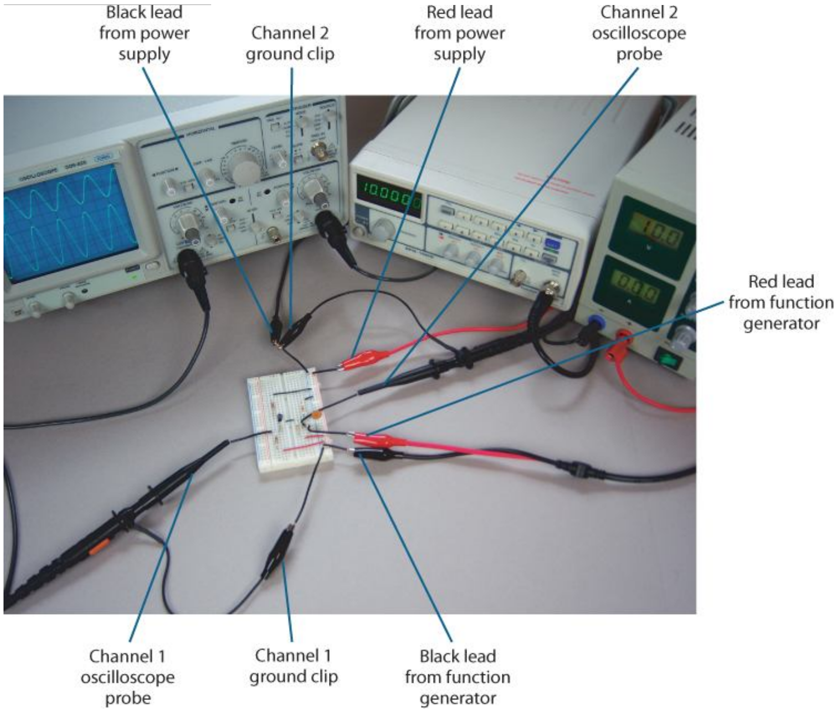
Black lead from power supply
Channel 2 ground clip
Red lead from power supply
Channel 2 oscilloscope probe
Red lead from function generator
Channel 1 oscilloscope probe
Channel 1 ground clip
Black lead from function generator

The input signal is represented by the upper sine wave shown in Figure 8.16 , and the output signal is represented by the lower sine wave. Read the number of divisions for the peak-to-peak output sine wave, and multiply it by the corresponding VOLTS/DIV setting to determine V_{out} .

Figure 8.16

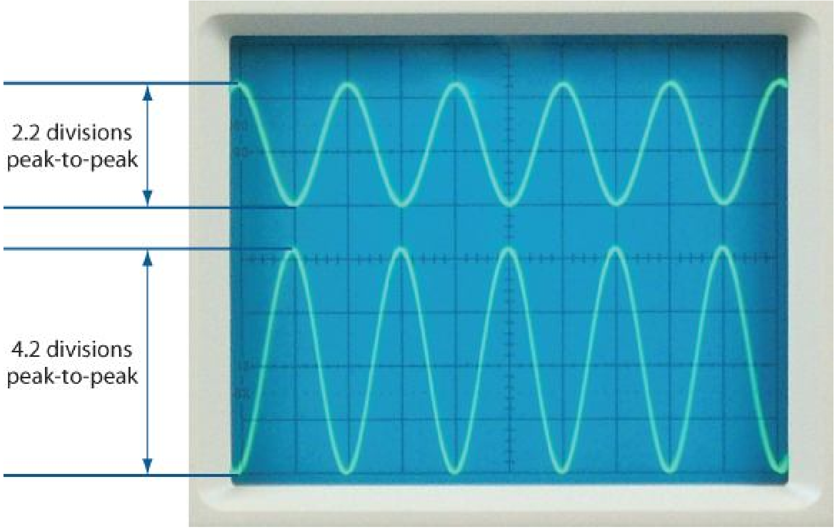
2.2 divisions
peak-to-peak
4.2 divisions
peak-to-peak

As you measure V in and V out for each circuit, you may need to adjust the TIME/DIV control, VOLTS/DIV control, and vertical POSITION controls on the oscilloscope. The controls shown in Figure 8.17 are adjusted to measure V in and V out for Circuit # 2.

Figure 8.17

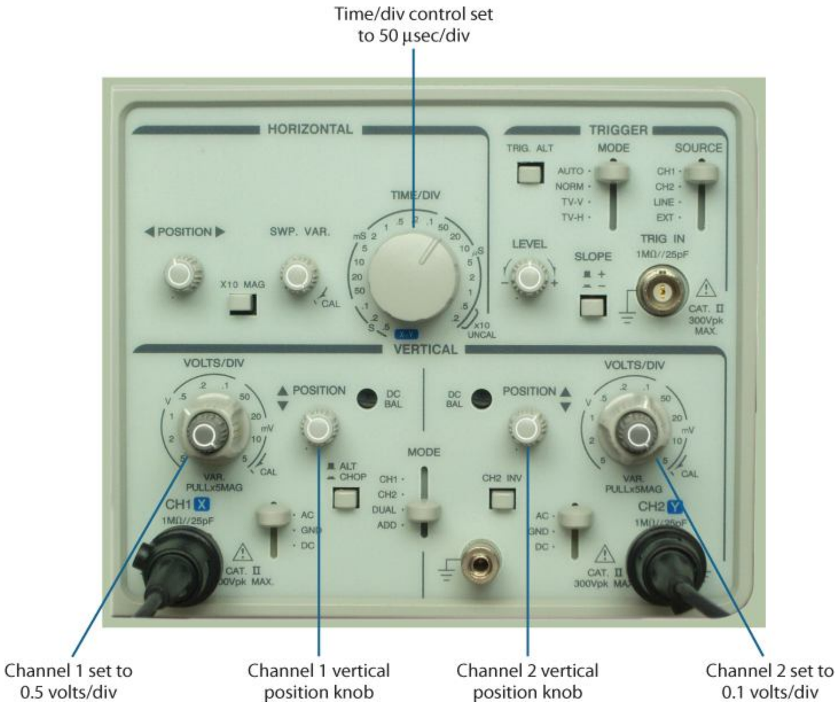
Time/div control set to 50 µsec/div
HORIZONTAL
TRIGGER
VERTICAL
Channel 1 set to 0.5 volts/div
Channel 1 vertical position knob
Channel 2 vertical position knob
Channel 2 set to 0.1 volts/div

Your values should be close to those shown in the following table.

Circuit #	Calculated A_V (R_C/R_E)	Transistor	β	V_{in} (Volts)	V_{out} (Volts)	Measured A_V (V_{in}/V_{out})
1	10	PN2222	235	0.22	2.1	9.5
2	10	2N3904	174	0.22	2.1	9.5
3	15	PN2222	235	0.22	3.2	14.5
4	15	2N3904	174	0.22	3.2	14.5

The measured values of A_V are quite close to the calculated values of A_V, well within variations that could be caused by the ± 5 percent tolerance specified for the resistor values. Also, the variation in transistor β had no effect on the measured values of A_V.

19 The AC voltage gain for the circuit discussed in problem 18 was 15. Earlier, you learned that the maximum practical gain of the amplifier circuit shown in Figure 8.9 is approximately 50.

However, in problem 10, you learned that AC voltage gains of up to 500 are possible for the amplifier circuit shown in Figure 8.4 . Therefore, by ensuring the stability of the DC bias point, the amplifier has much lower gain than is possible with the transistor amplifier circuit shown in Figure 8.4 .

You can make an amplifier with stable bias points without giving up high AC voltage gain by placing a capacitor in parallel with the emitter resistor, as shown in Figure 8.18 .

Figure 8.18

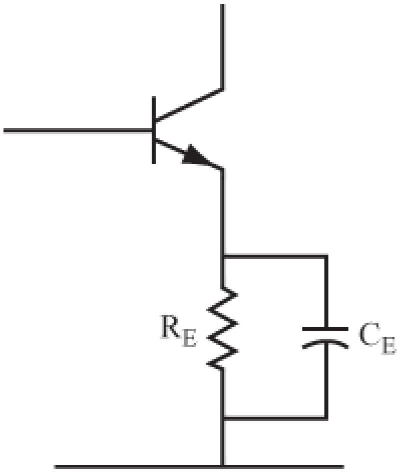

If the reactance of this capacitor for an AC signal is significantly smaller than R E , the AC signal passes through the capacitor rather than the resistor. Therefore, the capacitor is called an *emitter bypass capacitor* . The AC signal “sees” a different circuit from the DC, which is blocked by the capacitor and must flow through the resistor. Figure 8.19 shows the different circuits seen by AC and DC signals.

Figure 8.19

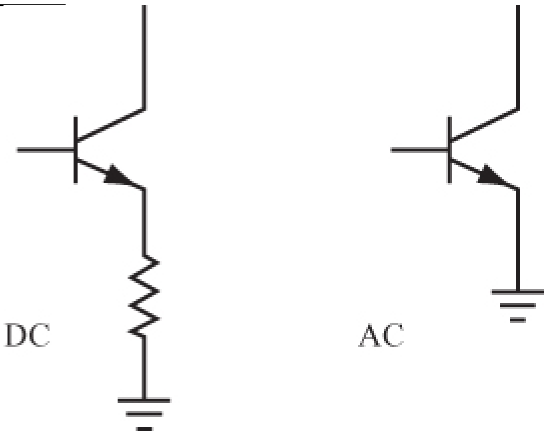

The AC voltage gain is now close to that of the amplifier circuit discussed in problems 1–10.

Questions

A. What effect does the emitter bypass capacitor have on an AC signal? _____

B. What effect does the emitter bypass capacitor have on the AC voltage gain? _____

C. What is the AC voltage gain formula with an emitter bypass capacitor included in the circuit? _____

Answers

A. It makes the emitter look like a ground and effectively turns the circuit into the circuit shown in Figure 8.4 .

B. It increases the gain.

C. The same formula used in problem 10:

$$A_V = \beta \times \frac{R_C}{R_{in}}$$

20 You can use the circuit shown in Figure 8.18 when you need as much AC voltage gain as possible. When high AC voltage gain is your priority, predicting the actual amount of gain is usually not important, so the fact that the equation is inexact is unimportant. If you need an accurate amount of gain, then you must use a different type of amplifier circuit that produces lower amounts of gain.

You can find the value of the capacitor C E using the following steps:

1. Determine the lowest frequency at which the amplifier must operate.
2. Calculate X C with the following formula:

$$X_C = \frac{R_E}{10}$$

3. Calculate C E with the following formula using the lowest frequency at which the amplifier must operate (determined in step 1):

$$X_C = \frac{1}{2\pi fC}$$

For the following question, use the circuit shown in Figure 8.10 , with an emitter bypass capacitor added, as shown in Figure 8.20 .

Figure 8.20

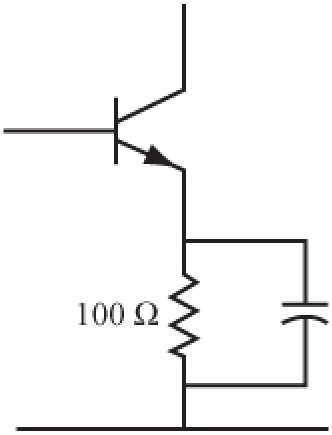

Questions

Follow the previous steps to calculate the value of C_E required if the lowest operating frequency of the amplifier is 50 Hz.

1. 50 Hz is the lowest frequency at which the amplifier must operate.
2. X_C = _____
3. C_E = _____

Answers

X_C = 10 ohms

C_E = 320 μF (approximately)

The AC voltage gain formula for an amplifier with an emitter bypass capacitor (Circuit 2 in Figure 8.21) is the same as the AC voltage gain formula for the amplifiers discussed in problems 1–10, where the emitter is directly connected to ground (Circuit 1 in Figure 8.21).

Figure 8.21

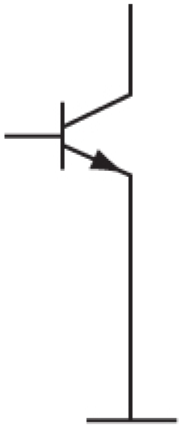

(1)

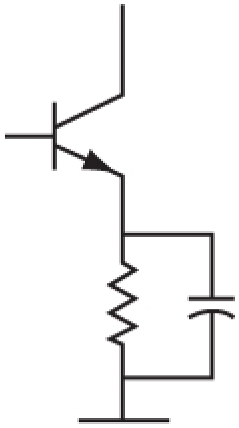

(2)

The AC voltage gain formula for an amplifier is as follows:

$$A_V = \beta \times \frac{R_C}{R_{in} + R_E}$$

(R_C is used instead of R_L because the collector resistor is the total load on the amplifier.)

Circuit 1 —Here, R_E = zero, so the AC voltage gain formula is as follows:

$$A_V = \beta \times \frac{R_C}{R_{in}}.$$

Circuit 2 —Here, R_E = zero for an AC signal because the AC signal is grounded by the capacitor, and R_E is out of the AC circuit. Thus, the AC voltage gain formula is as follows:

$$A_V = \beta \times \frac{R_C}{R_{in}}.$$

21 To obtain even larger voltage gains, two transistor amplifiers can be *cascaded* . That is, you can feed the output of the first amplifier into the input of the second amplifier. Figure 8.22 shows a two-transistor amplifier circuit, also called a *two-stage amplifier* .

Figure 8.22

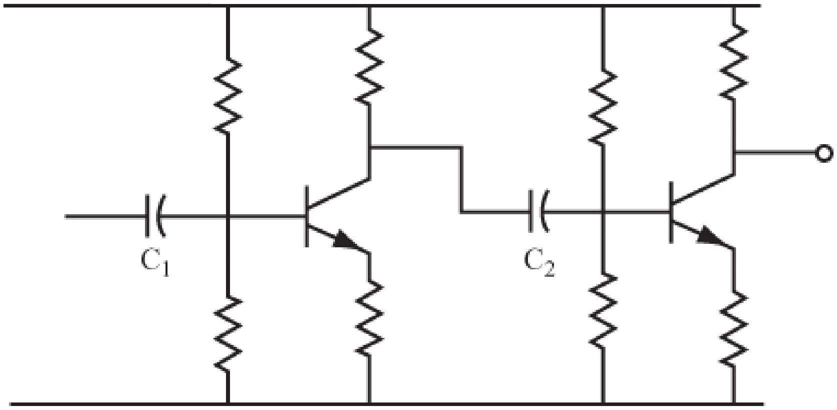

You find the total AC voltage gain by multiplying the individual gains. For example, if the first amplifier has an AC voltage gain of 10, and the second has an AC voltage gain of 10, then the overall AC voltage gain is 100.

Questions

A. Suppose you cascade an amplifier with a gain of 15 with one that has a gain of 25. What is the overall gain? _____

B. What is the overall gain if the individual gains are 13 and 17? _____

Answers

A. 375

B. 221

22 Two-stage amplifiers can achieve large

AC voltage gains if each amplifier uses an emitter bypass capacitor.

Question

What is the total AC voltage gain if each stage of a two-transistor amplifier has a gain of 100? _____

Answer

10,000

The Emitter Follower

23 Figure 8.23 shows another type of amplifier circuit.

Figure 8.23

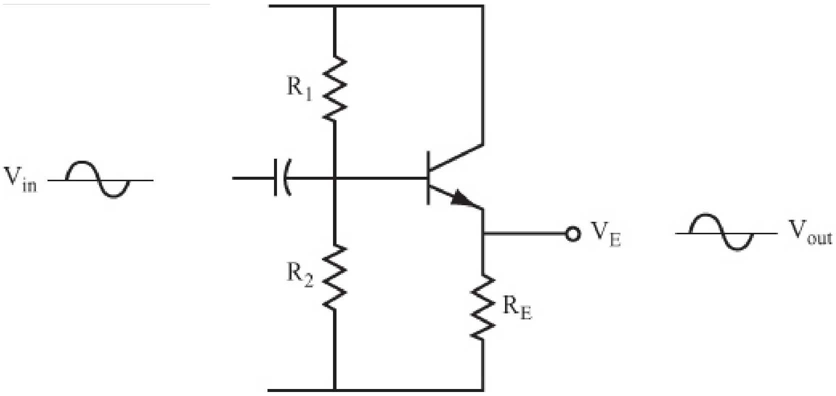

Question

How is the circuit shown in Figure 8.23 different from the amplifier circuit discussed in problems 11–18? _____

Answer

There is no collector resistor, and the output signal is taken from the emitter.

24 The circuit shown in Figure 8.23 is called an *emitter follower* amplifier. (In some cases, it is also called the *common collector* amplifier.)

The output signal has some interesting features:

- The peak-to-peak value of the output signal is almost the same as the input signal. In other words, the circuit gain is slightly less than 1; although in practice it is often considered to be 1.
- The output signal has the same phase as the input signal. It is not inverted; the output is simply considered to be the same as the input.
- The amplifier has a high input resistance. Therefore, it draws little current from the signal source.
- The amplifier has a low output resistance. Therefore, the signal at the emitter appears to be emanating from a battery or signal generator with a low internal resistance.

Questions

A. What is the voltage gain of an emitter follower amplifier? _____

B. Is the output signal inverted? _____

C. What is the input resistance of the emitter follower amplifier? _____

D. What is its output resistance? _____

Answers

A. 1

B. No

C. High

D. Low

25 The example in this problem demonstrates the importance of the emitter follower circuit. The circuit shown in Figure 8.24 contains a small AC motor with 100 ohms resistance that is driven by a 10 V pp signal from a generator. The 50-ohm resistor labeled R G is the internal resistance of the generator. In this circuit, only 6.7 V pp is applied to the motor; the rest of the voltage is dropped across R G .

Figure 8.24

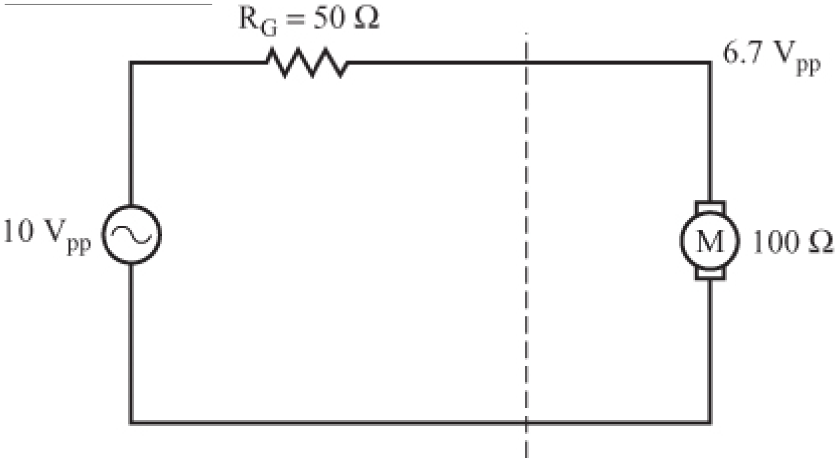

Figure 8.25 shows the same circuit, with a transistor connected between the generator and the motor in an emitter follower configuration.

Figure 8.25

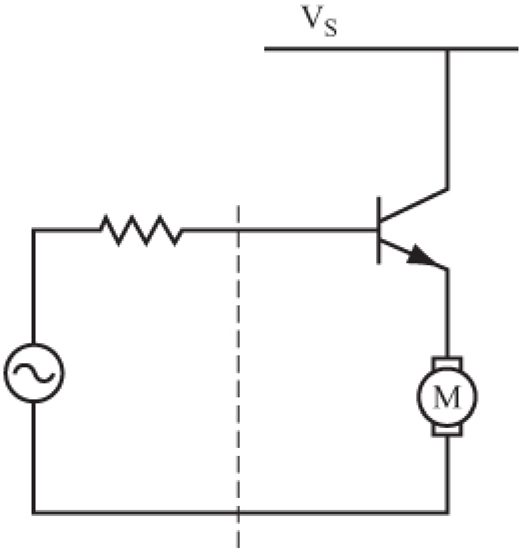

You can use the following formula to calculate the approximate input resistance of the transistor:

$$R_{in} = \beta \times R_E = 100 \times 100\,\Omega = 10{,}000\,\Omega \text{ (assuming that } \beta = 100)$$

The 10 V pp from the generator is divided between the 10,000-ohm input resistance of the transistor and the 50-ohm internal resistance of the generator. Therefore, there is no significant voltage drop across R G , and the full 10 V pp is applied to the base of the transistor. The emitter voltage remains at 10 V pp .

Also, the current through the motor is now produced by the power supply and not the generator, and the transistor looks like a generator with a low internal resistance.

This internal resistance (R O) is called the *output impedance* of the emitter follower. You can calculate it using this formula:

$$R_O = \frac{\text{internal resistance of generator}}{\beta}$$

For the circuit shown in Figure 8.25, if $R_G = 50$ ohms and $\beta = 100$, $R_O = 0.5$ ohms. Therefore, the circuit shown in Figure 8.25 is effectively a generator with an internal resistance of only 0.5 ohms driving a motor with a resistance of 100 ohms. Therefore, the output voltage of 10 V_{pp} is maintained across the motor.

Questions

A. What is the emitter follower circuit used for in this example? _____

B. Which two properties of the emitter follower are useful in circuits? _____

Answers

A. To drive a load that could not be driven directly by a generator

B. High input resistance and its low output resistance

26 The questions in this problem apply to the emitter follower circuit discussed in problems 23–25.

Questions

A. What is the approximate gain of an emitter follower circuit? _____

B. What is the phase of the output signal compared to the phase of the input signal? _____

C. Which has the higher value, the input resistance or the output resistance? _____

D. Is the emitter follower more effective at amplifying signals or at isolating loads? _____

Answers

A. 1

B. The same phase

C. The input resistance

D. Isolating loads

27 You can design an emitter follower circuit using the following steps:

1. Specify V_E. This is a DC voltage level, which is usually specified as half the supply voltage.
2. Find V_B. Use $V_B = V_E + 0.7$ volt.
3. Specify R_E. Often this is a given factor, especially if it is a motor or other load that is being driven.
4. Find I_E by using the following formula:

$$I_E = \frac{V_E}{R_E}$$

5. Find I_B by using the following formula:

$$I_B = \frac{I_E}{\beta}$$

6. Find I_2 by using $I_2 = 10 I_B$.
7. Find R_2 by using the following formula:

$$R_2 = \frac{V_B}{I_2}$$

8. Find R_1 by using the following formula:

$$R_1 = \frac{V_S - V_B}{I_2 + I_B}$$

Usually, I_B is small enough to be dropped from this formula.

9. Choose the nearest standard values for R_1 and R_2.
10. Check that these standard values give a voltage close to V_B. Use the voltage divider formula.

A simple design example illustrates this procedure. Use the values shown in the circuit in Figure 8.26 for this problem.

Figure 8.26

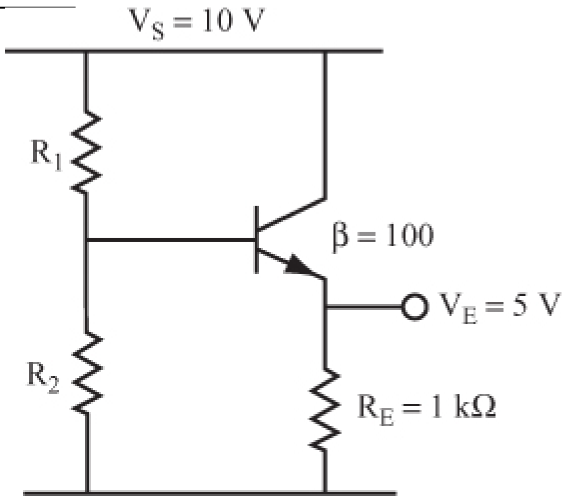

Questions

Work through Steps 1–10 to find the values of the two bias resistors.

1. V_E = _____
2. V_B = _____
3. R_E = _____
4. I_E = _____
5. I_B = _____
6. I_2 = _____

7. R_2 = _____
8. R_1 = _____
9. The nearest standard values are as follows:
R_1 = _____
R_2 = _____
10. V_B = _____

Answers

Your answers should be close to the following values:

1. 5 volts (This was given in Figure 8.26.)
2. 5.7 volts
3. 1 kΩ (This was given in Figure 8.26.)
4. 5 mA
5. 0.05 mA
6. 0.5 mA
7. 11.4 kΩ
8. 7.8 kΩ
9. The nearest standard values are 8.2 kΩ and 12 kΩ.
10. The standard resistor values result in V_B = 5.94 volts. This is a little higher than the V_B calculated in Step 2, but it is acceptable.

V_E is set by the biasing resistors. Therefore, it is not dependent upon the value of R_E. Almost any value of R_E can be used in this circuit. The minimum value for R_E is obtained by using this simple equation:

$$R_E = \frac{10R_2}{\beta}$$

Analyzing an Amplifier

28 Up to now, the emphasis has been on designing a simple amplifier and an emitter follower. This section shows how to “analyze” a circuit that has already been designed. In this case, to “analyze” means to calculate the collector DC voltage (the bias point) and find the AC gain. This procedure is basically the reverse of the design procedure.

Start with the circuit shown in Figure 8.27 .

Figure 8.27

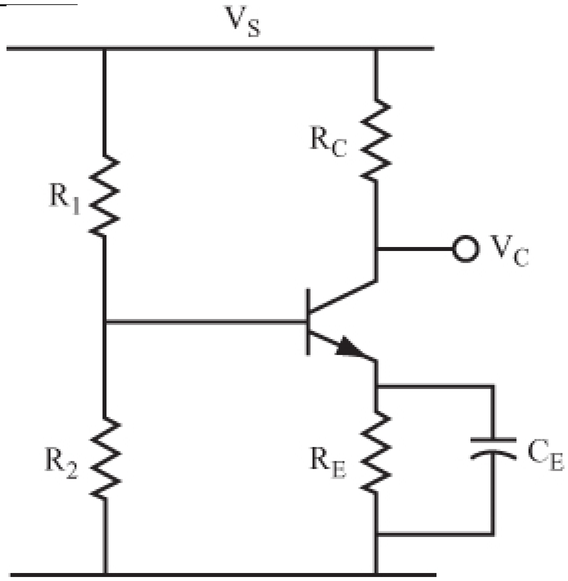

Following are the steps you use to analyze a circuit:

1. Find V_B by using the following equation:

$$V_B = V_S \times \frac{R_2}{R_1 + R_2}.$$

2. Find V_E by using $V_E = V_B - 0.7$ volt.

3. Find I_C by using the following equation:

$$I_C = \frac{V_E}{R_E}.$$

Note that I_C = I_E .

4. Find V_R by using V_R = R_C × I_C .

5. Find V_C by using V_C = V_S 2V_R . This is the bias point.

6. Find A_V by using the following equation:

$$A_V = \frac{R_C}{R_E}, \text{ or } A_V = \beta \times \frac{R_C}{R_{in}}$$

When you use the second formula, you must find the value of R_{in} (or h_{ie}) on the data sheets for the transistor from the manufacturer.

Use the circuit shown in Figure 8.28 for the following questions. For these questions, use $\beta = 100$, $R_{in} = 2$ kΩ and the values given in the circuit drawing.

Figure 8.28

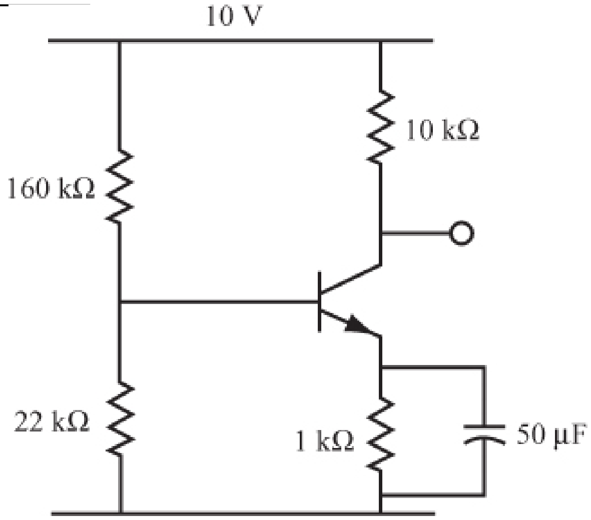

Questions

Calculate V_B, V_E, I_C, V_R, V_C, and A_V using Steps 126 of this problem.

1. V_B = _____
2. V_E = _____
3. I_C = _____
4. V_R = _____
5. V_C = _____
6. A_V = _____

Answers

1.
$$V_B = 10 \times \frac{22\,\text{k}\Omega}{160\,\text{k}\Omega + 22\,\text{k}\Omega} = 1.2\,\text{volts}$$

2. V_E = 1.2 2 0.7 = 0.5 volt

3.
$$I_C = \frac{0.5\,\text{V}}{1\,\text{k}\Omega} = 0.5\,\text{mA}$$

4. V_R = 10 kΩ × 0.5 mA = 5 volts

5. V_C = 10 volts 2 5 volts = 5 volts (This is the bias point.)

6. With the capacitor:
$$A_V = 100 \times \frac{10\,\text{k}\Omega}{2\,\text{k}\Omega} = 500\,(\text{a large gain})$$

7. Without the capacitor:
$$A_V = \frac{10\,\text{k}\Omega}{1\,\text{k}\Omega} = 10\ (\text{a small gain})$$

29 You can determine the lowest frequency the amplifier will satisfactorily pass by following these simple steps:

1. Determine the value of R_E.
2. Calculate the frequency at which $X_C = R_E/10$. Use the capacitor reactance formula.

(This is one of those "rules of thumb" that *can* be mathematically justified and gives reasonably accurate results in practice.)

Questions

For the circuit shown in Figure 8.28, find the following.

A. R E = _____

B. f = _____

Answers

A. R E = 1 kΩ (given in the circuit diagram)

B. So, you set X C = 100 ohms, and use this formula:

$$X_C = \frac{1}{2\pi fC}$$

$$100\,\text{ohms} = \frac{0.16}{f \times 50\,\mu\text{F}} \quad \text{since } 0.16 = \frac{1}{2\pi}$$

So, the following is the result:

$$f = \frac{0.16}{100 \times 50 \times 10^{-6}} = 32\,\text{Hz}$$

30 For the circuit shown in Figure 8.29, follow the steps given in problems 28 and 29 to answer the following questions.

Figure 8.29

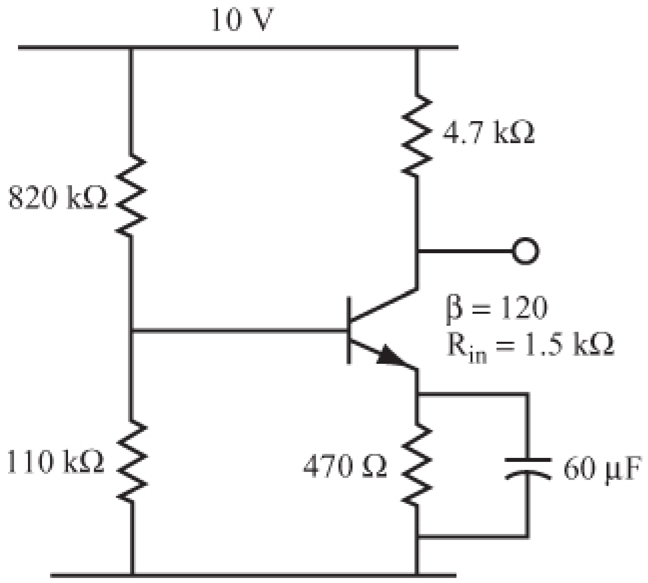
10 V
4.7 kΩ
820 kΩ
β = 120
R_in = 1.5 kΩ
110 kΩ
470 Ω
60 μF

Questions

1. V_B = _____
2. V_E = _____
3. I_C = _____
4. V_R = _____
5. V_C = _____
6. With capacitor:

A_V = _____

Without capacitor:

A_V = _____

7. Low frequency check:

f = _____

Answers

Your answers should be close to these.

1. 1.18 volts
2. 0.48 volts
3. 1 mA
4. 4.7 volts

5. 5.3 volts (bias point)
6. With capacitor: 376
Without capacitor: 10
7. 57 Hz (approximately)

The JFET as an Amplifier

31 Chapter 3 discussed the JFET in problems 28231, and Chapter 4 discussed the JFET in problems 37241. You may want to review these problems before answering the questions in this problem. Figure 8.30 shows a typical biasing circuit for a JFET.

Figure 8.30

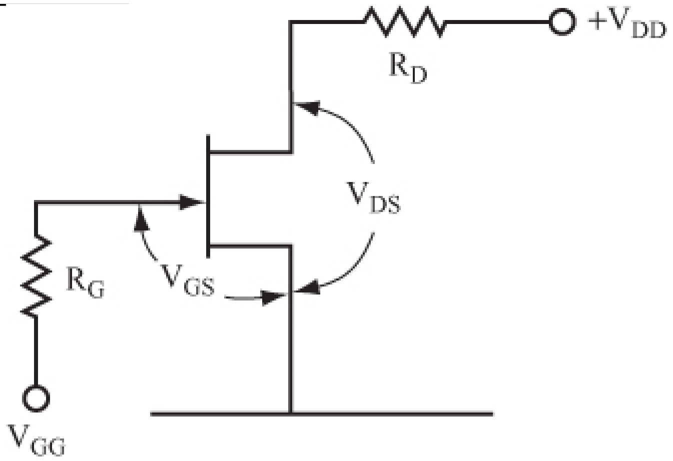

Questions

A. What type of JFET is depicted in the circuit? _____

B. What value of V GS would you need to turn the JFET completely ON? _____

C. What drain current flows when the JFET is completely ON? _____

D. What value of V_{GS} would you need to turn the JFET completely OFF? _____

E. When a JFET is alternately turned completely ON and OFF in a circuit, what type of component are you using the JFET as? _____

Answers

A. N-channel JFET.

B. $V_{GS} = 0$ V to turn the JFET completely ON.

C. Drain saturation current (I_{DSS}).

D. V_{GS} should be a negative voltage for the N-channel JFET to turn it completely OFF. The voltage must be larger than or equal to the cutoff voltage.

E. The JFET is being used as a switch.

32 You can use a JFET to amplify AC signals by biasing the JFET with a gate to source voltage about halfway between the ON and OFF states. You can find the drain current that flows in a JFET biased to a particular V_{GS} by using the following equation for the transfer curve:

$$I_D = I_{DSS}\ 1 - \frac{V_{GS}}{V_{GS(\text{off})}}{}^{2}$$

In this equation, I_{DSS} is the value of the drain saturation current, and $V_{GS(\text{off})}$ is the gate to source voltage at cutoff. Both of these are indicated on the transfer curve shown in Figure 8.31 .

Figure 8.31

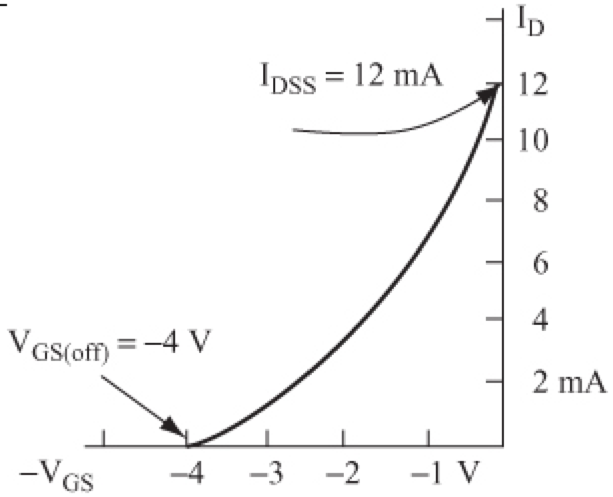

For the transfer curve shown in Figure 8.31, I_{DSS} = 12 mA and $V_{GS(off)}$ = −4 volts. Setting the bias voltage at V_{GS} = −2 volts returns the following value for the drain current:

$$I_D = 12\,\text{mA} \times 1 - \frac{-2}{-4}\ {}^{2} = 12\,\text{mA} = (0.5)^2 = 3\,\text{mA}$$

Questions

Calculate the drain current for the following:

A. V_{GS} = 21.5 _____

B. V_{GS} = 20.5 volts _____

Answers

A. 4.7 mA

B. 9.2 mA

Note Data sheets give a wide range of possible I_{DSS} and $V_{GS(off)}$) values for a given JFET. You may need to resort to actually measuring these with the method shown in Project 4-2.

33 For the circuit shown in Figure 8.30, you choose the value of the drain to source voltage, V_{DS}, and then calculate the value of the load resistor, R_D, by using the following equation:

$$R_D = \frac{(V_{DD} - V_{DS})}{I_D}$$

For this problem, use I_D = 3 mA, and a drain supply voltage (V_{DD}) of 24 volts. Calculate the value of R_D that results in the specified value of V_{DS}; this is also the DC output voltage of the amplifier.

Question

Calculate the value of R_D that will result in V_{DS} = 10 volts. _____

Answers

$$R_D = \frac{(V_{DD} - V_{DS})}{I_D} = \frac{(24\,\text{volts} = 10\,\text{volts})}{3\,\text{mA}} = \frac{14\,\text{volts}}{3\,\text{mA}} = 4.67\,\text{k}\Omega$$

34 The circuit shown in Figure 8.32 (which is referred to as a *JFET common source amplifier*) applies a 0.5 V_{pp} sine wave to the gate of the JFET and produces an amplified sine wave output from the drain.

Figure 8.32

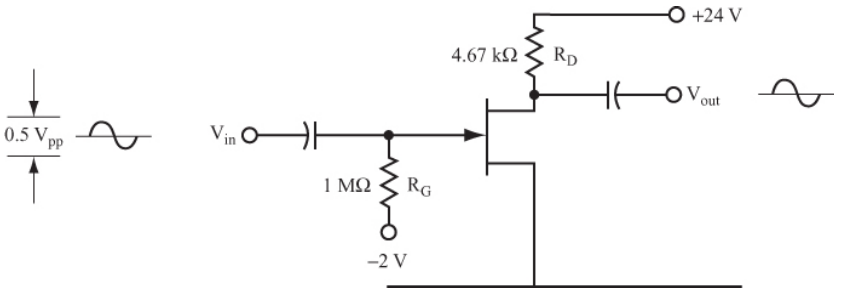

The input sine wave is added to the −2 volt bias applied to the gate of the JFET. Therefore, V_{GS} varies from −1.75 to −2.25 volts.

Question

Using the formula in problem 32, calculate I_D for the maximum and minimum values of V_{GS}.

Answer

For V_{GS} = 21.75 volts, I_D = 3.8 mA

For V_{GS} = 22.25 volts, I_D = 2.3 mA

35 As the drain current changes, V_{RD} (the voltage drop across resistor R_D) also changes.

Question

For the circuit shown in Figure 8.32, calculate the values of V_{RD} for the maximum and minimum values of I_D you calculated in problem 34. _____

Answer

$$\text{For } I_D = 3.8\,\text{mA}, V_{RD} = 3.8\,\text{mA} \times 4.67\,\text{k}\Omega = 17.7\,\text{volts}$$

$$\text{For } I_D = 2.3\,\text{mA}, V_{RD} = 2.3\,\text{mA} = 4.67\,\text{k}\Omega = 10.7\,\text{volts}$$

This corresponds to a 7 V_{pp} sine wave.

36 As the voltage drop across R_D

changes, the output voltage also changes.

Question

For the circuit shown in Figure 8.32, calculate the values of V out for the maximum and minimum values of V RD you calculated in problem 35. _____

Answer

For $V_{RD} = 17.7\text{ volts}, V_{out} = V_{DD} - V_{RD} = 24\text{ volts} - 17.7\text{ volts} = 6.3\text{ volts}$

For $V_{RD} = 10.7\text{ volts}, V_{out} = V_{DD} - V_{RD} = 24\text{ volts} = 10.7\text{ volts} = 13.3\text{ volts}$

Therefore, the output signal is a 7 V pp sine wave.

37 Table 8.1 shows the results of the calculations made in problems 34–36 including the DC bias point.

Table 8.1 Calculation Results

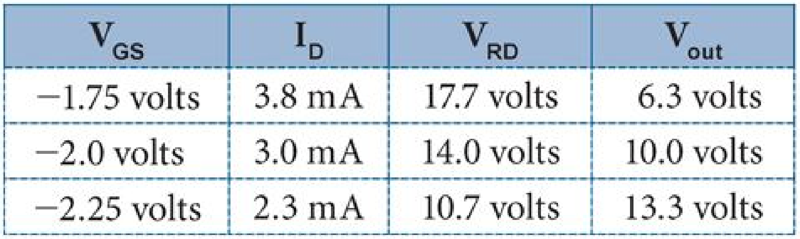

V_{GS}	I_D	V_{RD}	V_{out}
−1.75 volts	3.8 mA	17.7 volts	6.3 volts
−2.0 volts	3.0 mA	14.0 volts	10.0 volts
−2.25 volts	2.3 mA	10.7 volts	13.3 volts

Question

What are some characteristics of the AC output signal? _____

Answer

The output signal is a 7 V pp sine wave with the same frequency as the input sine wave. As the input voltage on V GS increases (toward 0 volts), the output decreases. As the input voltage decreases (becomes more negative), the output voltage increases. This means that the output is 180 degrees out of phase with the input.

38 You can calculate the AC voltage gain for the amplifier discussed in problems 34–37 by using the following formula:

$$A_v = \frac{-V_{out}}{V_{in}}$$

The negative sign in this formula indicates that the output signal is 180 degrees out of phase from the input signal.

Question

Calculate the AC voltage gain for the amplifier discussed in problems 34–37. _____

Answer

$$A_V = \frac{-7\,V_{pp}}{0.5\,V_{pp}} = -14$$

39 You can also calculate the AC voltage gain by using the following formula:

$$A_v = -(g_m)(R_D)$$

In this equation, g_m is the *transconductance* and is a property of the JFET. It is also called the *forward transfer admittance*. A typical value for g_m is usually provided for JFETs in the data sheet from the manufacturers. You can also use the data in Table 8.1 to calculate g_m using the following formula:

$$g_m = \frac{\Delta I_D}{\Delta V_{GS}}$$

In this equation, Δ indicates the change or variation in V_{GS} and the corresponding drain current. The unit for transconductance is mhos.

Questions

A. Using the data from Table 8.1, what is the value of g_m for the JFET used in the amplifier? _____

B. What is the corresponding AC voltage gain? _____

Answers

A.

$$gm = \frac{1.5\,\text{mA}}{0.5\,\text{V}} = 0.003\ \text{mhos}$$

B. $A_v = 2(0.003)(4670) = 214$, the same result you found in problem 38

40 Design a JFET common source amplifier using a JFET with $I_{DSS} = 14.8$ mA and $V_{GS(off)} = 23.2$ volts. The input signal is 40 mV_{pp}. The drain supply is 24 volts.

Questions

A. Determine the value of V_{GS} that will bias the JFET at a voltage near the middle of the transfer curve. _____

B. Calculate the drain current when V_{GS} is at the value determined in step A, using the formula in problem 32. _____

C. Choose a value of V_{DS} and calculate the value of R_D using the formula in problem 33. _____

D. Calculate the maximum and minimum values of V_{GS} that result from the input signal, and the corresponding values of drain current using the procedure in problem 34. _____

E. Calculate the maximum and minimum values of V_{out} that result from the input signal using the procedures in problems 35 and 36. _____

F. Calculate the gain of the amplifier. _____

Answers

A. V_{GS} = 21.6 volts

B. I_D = 3.7 mA

C. For V_{DS} = 10 volts,

$$R_D = \frac{14\,\text{volts}}{3.7\,\text{mA}} = 3780\ \text{ohms}$$

D. V_{GS} will vary from 21.58 to 21.62 volts. Use the formula to calculate values of drain current. I_D will vary from 3.79 to 3.61 mA.

E. V_{RD} will vary from 14.3 to 13.6 volts. Therefore, V_{out} will vary from 9.7 to 10.4 volts.

F. $$A_v = \frac{-0.7}{0.04} = -17.5$$

41 Use the results of problem 40, question D, to answer the following question.

Questions

Calculate the transconductance of the JFET and the AC voltage gain using the formulas discussed in problem 39. _____

Answers

$$g_m = \frac{\Delta I_D}{\Delta V_{GS}} = \frac{0.18\,\text{mA}}{40\,\text{mV}} = 0.0045\text{ mhos}$$

$$A_v = -(g_m)(R_D) = -(0.0045)(3780) = -17$$

This is close to the value you found in problem 40, question F.

42 Figure 8.33 shows a JFET amplifier circuit that uses one power supply, rather than separate power supplies for the drain and gate used in the amplifier discussed in problems 34–41.

Figure 8.33

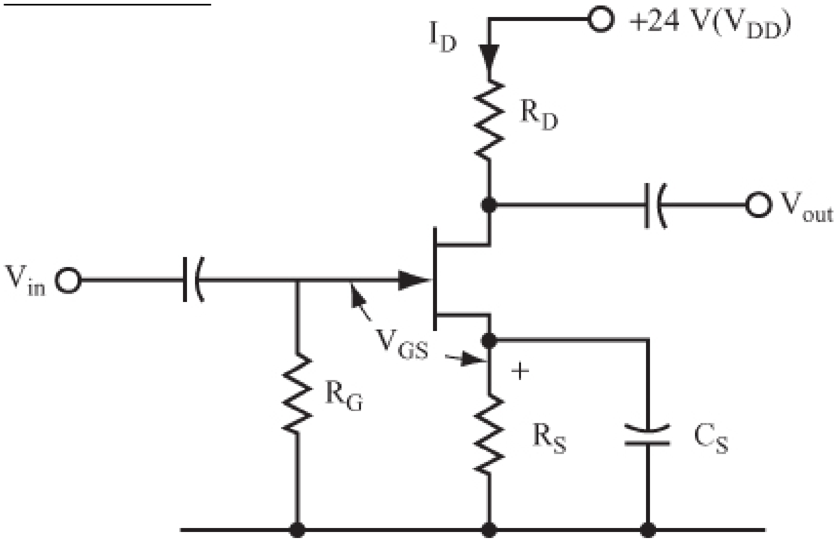

The DC voltage level of the gate is zero because the gate is tied to ground through R_G. Therefore, the voltage drop across R_S becomes the gate to source voltage. To design the circuit, you must find values for both R_S and R_D. Use the same bias point for this problem as you used for the amplifier discussed in problems 34–41: $V_{GS} = 22$ volts and $I_D = 3$ mA. Follow these steps:

1. Calculate R_S, using the following formula, recognizing that $V_{RS} = V_{GS}$:

$$R_S = \frac{V_{RS}}{I_D} = \frac{V_{GS}}{I_D}$$

2. Calculate R_D using the following formula, using $V_{DS} = 10$ volts, the same value you used for the amplifier discussed in problems 34–41:

$$R_D = \frac{(V_{DD} - V_{DS} - V_S)}{I_D}$$

3. Calculate X_{CS} using the following formula:

$$X_{CS} = \frac{R_S}{10}$$

Then, calculate C_S using the following formula:

$$X_{CS} = \frac{1}{2\ \ fC_S}$$

4. Calculate the peak-to-peak output voltage using the procedures shown in problems 34–36.

5. Calculate the AC voltage gain using this formula:

$$A_V = \frac{-V_{out}}{V_{in}}$$

Note Choose the value of C_S so that its reactance is less than 10 percent of R_S at the lowest frequency you need to amplify.

The DC load for the JFET is R_D plus R_S. The AC load is R_D only because C_S bypasses the AC signal around R_S, which keeps the DC operating point stable. The use of C_S reduces the gain slightly because you now use a smaller R_D to calculate the AC voltage swings at the output.

Questions

A. What is the value of R_S? _____

B. What is the value of R_D? _____

C. What is the value of C_S? Assume f = 1 kHz. _____

D. Calculate the peak-to-peak V_{out} for V_{in} = 0.5 V_{pp}. _____

E. What is the voltage gain? _____

Answers

A. $$R_S = \frac{2\,\text{volts}}{3\,\text{mA}} = 667\,\text{ohms}$$

B. $$R_D = \frac{12\,\text{volts}}{3\,\text{mA}} = 4\,\text{k}\Omega$$

C. X_{CS} = 66.7 ohms, C_S = 2.4 μF

D. The AC drain current will still vary from 3.8 to 2.3 mA, as in problem 37. The voltage across R_D is now 6 V_{pp} because R_D is 4 kΩ. The output voltage is also 6 V_{pp}.

E. $$A_V = \frac{-6}{0.5} = -12$$ The gain is 12.

The Operational Amplifier

43 The operational amplifier (op-amp) in use today is actually an integrated circuit (IC). This means that the device has numerous transistors and other components constructed on a small silicon chip. These IC op-amps are much smaller and, therefore, more practical than an amplifier with equivalent performance that is made with discrete components.

You can purchase op-amps in different case configurations. Some of these configurations are the Transistor Outline (TO) metal package, the flat pack, and the dual in-line pin (DIP) package. You can also find two op-amps (dual) or four op-amps (quad) in a single IC.

Their size, low cost, and wide range of applications have made op-amps so common today that they are thought of as a circuit device or component in and of themselves, even though a typical op-amp may contain 20 or more transistors in its design. The characteristics of op-amps closely resemble those of an ideal amplifier. Following are these characteristics:

High input impedance (does not require input current)

High gain (used for amplifying small signal levels)

Low output impedance (not affected by the load)

Questions

A. What are the advantages of using

op-amps? _____

B. Why are op-amps manufactured using IC techniques? _____

Answers

A. Small size, low cost, wide range of applications, high input impedance, high gain, and low output impedance.

B. Because of the large numbers of transistors and components that are required in the design of an op-amp, they must be constructed on a single, small silicon chip using IC manufacturing techniques to be of a reasonable size.

44 Figure 8.34 shows the schematic symbol for an op-amp.

Figure 8.34

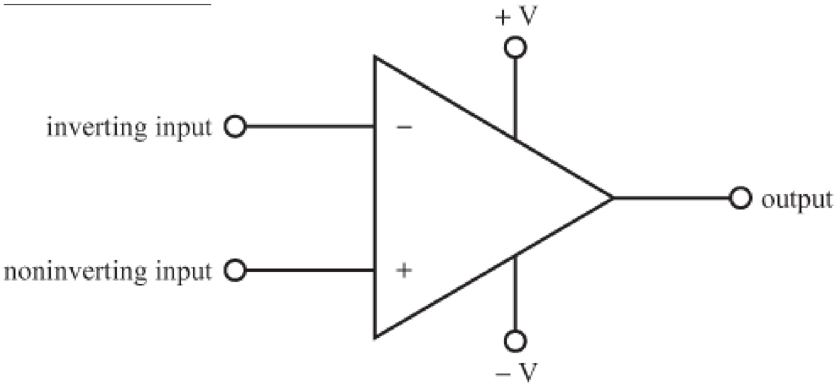

An input at the inverting input results in an output that is 180 degrees out of phase with the input. An input at the noninverting input results in an output that is in phase with the input. Both positive and negative voltage supplies are required, and the data sheet will specify their values for the particular op-amp you use. Datasheets usually contain circuit diagrams showing how you should connect external components to the op-amp for specific applications. These circuit diagrams (showing how a particular op-amp can be used for various applications) can be useful to the designer or the hobbyist.

Questions

A. How many terminals does the op-amp require, and what are their functions? _____

B. How is the output related to the input when the input is connected to the inverting input? _____

Answers

A. Five—two input terminals, one output terminal, two power supply terminals.

B. The output is 180 degrees out of phase with the input.

45 Figure 8.35 shows a basic op-amp circuit. The input signal is connected to an inverting input, as indicated by the negative sign. Therefore, the output signal will be 180 degrees out of phase with the input.

Figure 8.35

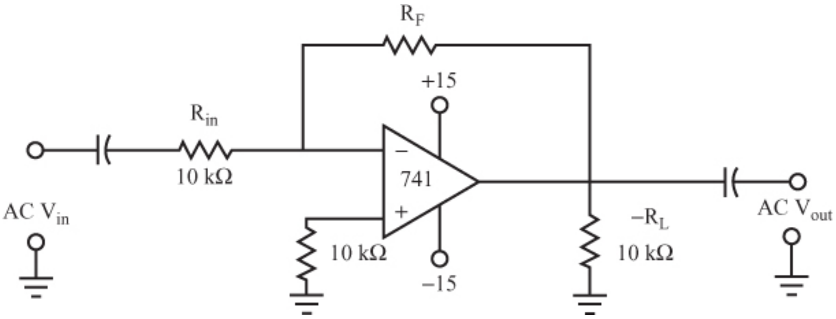

You can find the AC voltage gain for the circuit using the following equation:

$$A_v = \frac{-R_F}{R_{in}}$$

Resistor R F is called a *feedback resistor* because it forms a feedback path from the output to the input. Many op-amp circuits use a feedback loop. Because the op-amp has such a high gain, it is easy to saturate it (at maximum gain) with small voltage differences between the two input terminals. The feedback loop allows the operation of the op-amp at lower gains, allowing a wider range of input voltages. When designing a circuit, you can choose the value of the feedback resistor to achieve a specific voltage gain. The role of the capacitors in the diagram is to block DC voltages.

Questions

A. Calculate the value of R F that would give the amplifier an AC voltage gain of 120. _____

B. Calculate A C V out if AC V in is 5 mV rms. _____

Answers

A. $R_F = 120 \times 10\,k\Omega = 1.2\,M\Omega$

B. $V_{out} = 120 \times 5\,mV = 0.6\,V_{rms}$

The output signal is inverted with respect to the input signal.

46 Use the op-amp circuit shown in Figure 8.35 to build an amplifier with an output voltage of 12 V_{pp}, an AC voltage gain of 50, and with R_{in} = 6.8 kΩ.

Questions

A. Calculate the value of R_F. _____

B. Calculate the value of V_{in} required to produce the output voltage specified earlier. _____

Answers

A. $R_F = 50 \times 6.8\,k\Omega = 340\,k\Omega$

B. $$V_{in} = \frac{12\,V_{pp}}{50} = 0.24\,V_{pp} \text{ or } 0.168\,V_{rms}$$

Project 8.2: The Operational Amplifier

Objective

The objective of this project is to demonstrate how AC voltage gain changes when you use feedback resistors of different values in an op-amp circuit.

General Instructions

After the circuit is set up, you measure V_{out} for each value of R_F, and find A_V, using the ratio V_{out}/V_{in}. You also determine a calculated A_V using the ratio R_F/R_{in} in each case to determine how close the calculated A_V is to the measured A_V.

Parts List

You need the following equipment and supplies:

- One 0.1 μF capacitor.
- Two 10 kΩ, 0.25-watt resistors.
- One 51 kΩ, 0.25-watt resistor.
- One 100 kΩ, 0.25-watt resistor.
- One 150 kΩ, 0.25-watt resistor.
- One 220 kΩ, 0.25-watt resistor.
- One 270 kΩ, 0.25-watt resistor.
- One 330 kΩ, 0.25-watt resistor.
- One 380 kΩ, 0.25-watt resistor.
- Two terminal blocks.
- Two 6-volt battery packs (4 AA batteries each).
- One function generator.
- One oscilloscope.
- One breadboard.
- One OPA134 operational amplifier. This op-amp comes in a few different packages; get the 8-pin dual in-line (DIP) version. Figure 8.36 shows the pinout diagram for the OPA134. When you insert the op-amp into the breadboard, try not to bend any of the leads. The leads on dual in-line packages are fragile and will break off if you bend them more than once or twice.

Figure 8.36

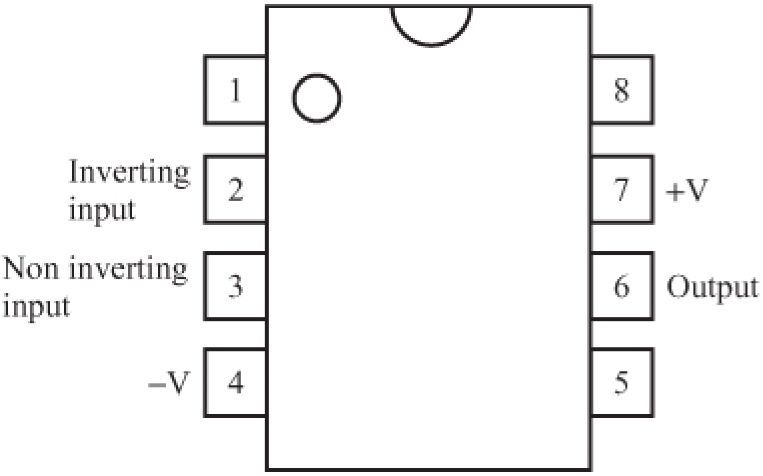

Step-by-Step Instructions

Set up the circuit shown in Figure 8.37 using the 51 kΩ resistor for R_F. Figure 8.38 shows the battery connections. If you have some experience in building circuits, this schematic (along with the previous parts list) should provide all the information you need to build the circuit. If you need a bit more help, look at the photos of the completed circuit in the “Expected Results” section. One unusual aspect of this circuit you may want to look for in the photos is how the 2V bus of one 6-volt battery pack should be connected to the +V bus of the other 6-volt battery pack.

Figure 8.37

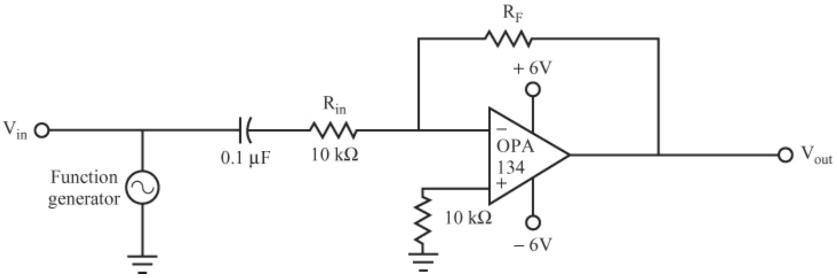
R_F
+ 6V
R_{in}
V_{in}
0.1 µF
10 kΩ
OPA 134
V_{out}
Function generator
10 kΩ
− 6V

Figure 8.38

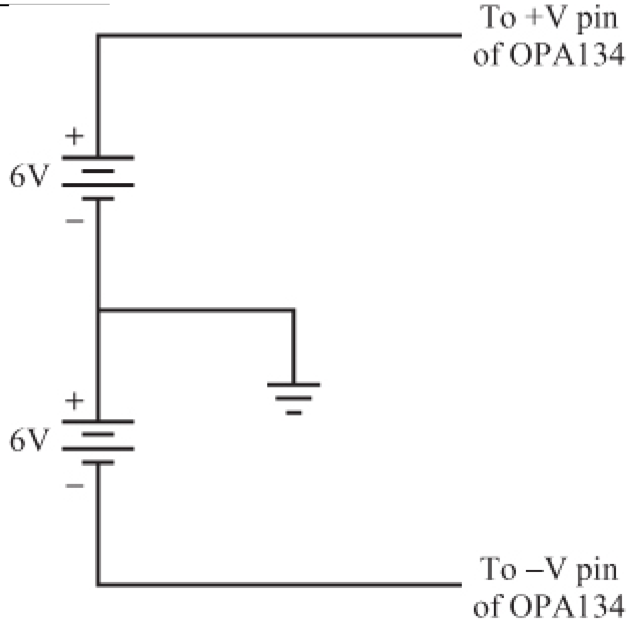
To +V pin of OPA134
+
6V
−
+
6V
−
To −V pin of OPA134

Carefully check your circuit against the diagram.

After you check your circuit, follow these steps, and record your measurements in the blank table following the steps.

1. Connect the oscilloscope probe for channel 2 to a jumper wire connected to V_{in}. Connect the ground clip to a jumper wire attached to the ground bus.

2. Connect the oscilloscope probe for channel 1 to a jumper wire connected to V out , and then connect the ground clip to a jumper wire attached to the ground bus.

3. Set the function generator to generate a 10 kHz sine wave with approximately 0.2 V pp .

4. Measure and record V out and V in .

5. Change the feedback resistor to the value shown in the next row of the table (labeled 100 kΩ in this instance). Each time you change the resistor, it's advisable to disconnect the batteries to avoid shorting wires.

6. Measure and record V out and V in .

7. Repeat steps 5 and 6 until you have recorded V out and V in for the last row of the table.

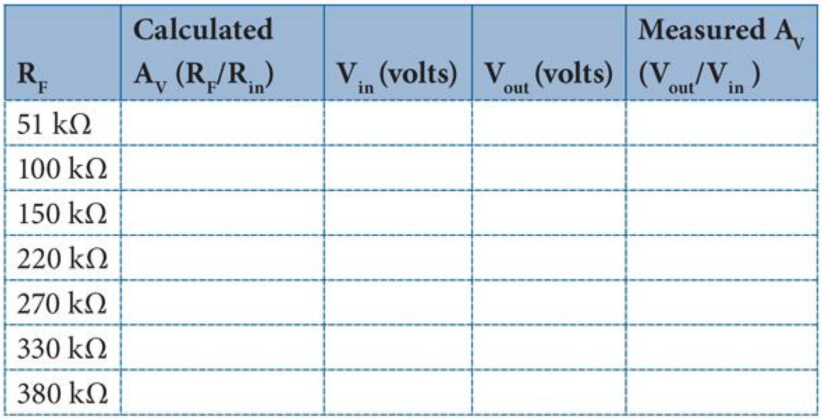

R_F	Calculated A_V (R_F/R_{in})	V_{in} (volts)	V_{out} (volts)	Measured A_V (V_{out}/V_{in})
51 kΩ				
100 kΩ				
150 kΩ				
220 kΩ				
270 kΩ				
330 kΩ				
380 kΩ				

8. Determine the calculated A V and the measured A V , and record these values in each row of the table.

Expected Results

Figure 8.39 shows the breadboarded circuit for this project.

Figure 8.39

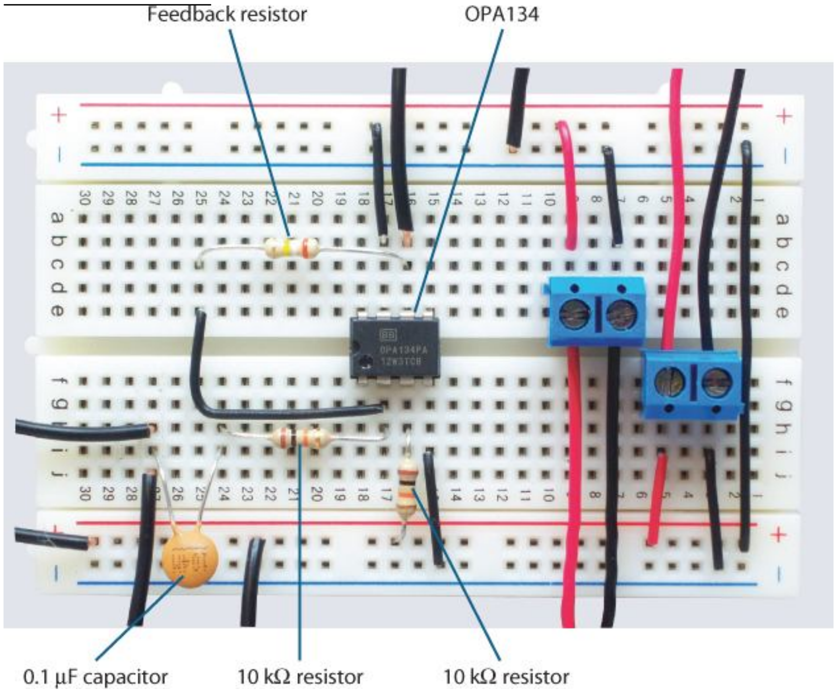
Feedback resistor
OPA134
0.1 µF capacitor
10 kΩ resistor
10 kΩ resistor

Figure 8.40 shows a function generator and oscilloscope attached to the circuit.

Figure 8.40

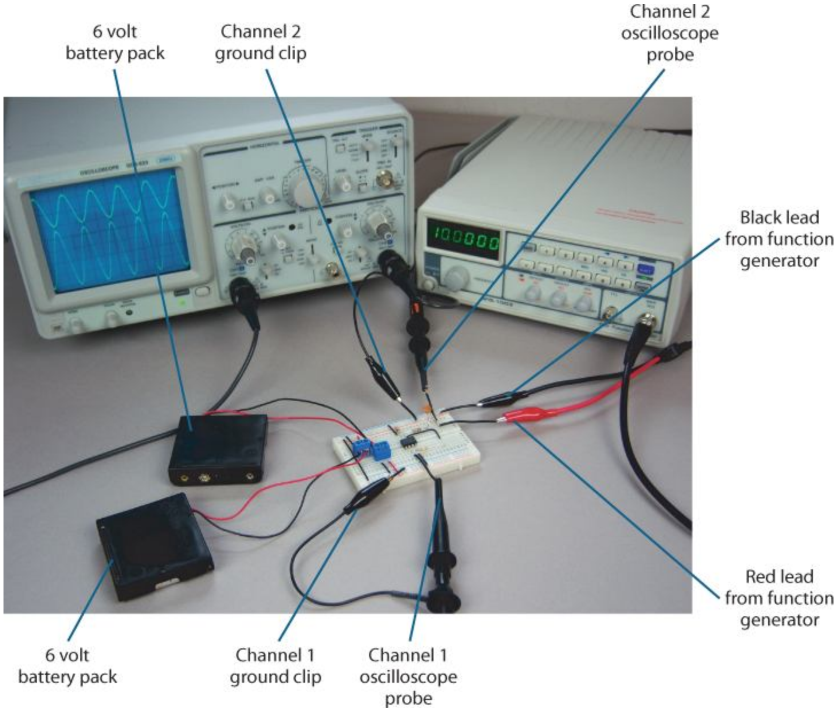
6 volt battery pack
Channel 2 ground clip
Channel 2 oscilloscope probe
Black lead from function generator
Red lead from function generator
6 volt battery pack
Channel 1 ground clip
Channel 1 oscilloscope probe

The input signal is represented by the upper sine wave shown in Figure 8.41, and the output signal is represented by the lower sine wave. Count the number of divisions for the peak-to-peak output sine wave, and multiply that number by the corresponding VOLTS/DIV setting to determine V_{out} and V_{in}.

Figure 8.41

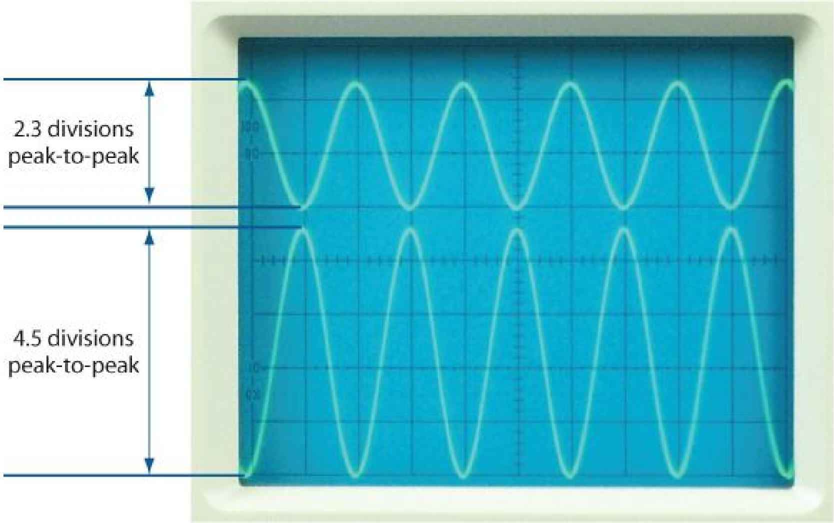
2.3 divisions peak-to-peak
4.5 divisions peak-to-peak

As you measure V_{in} and V_{out}, you may need to adjust the TIME/DIV control, the VOLTS/DIV control, and vertical POSITION controls on the oscilloscope. The controls shown in Figure 8.42 are adjusted to measure V_{out} when R_F = 380 kΩ.

Figure 8.42

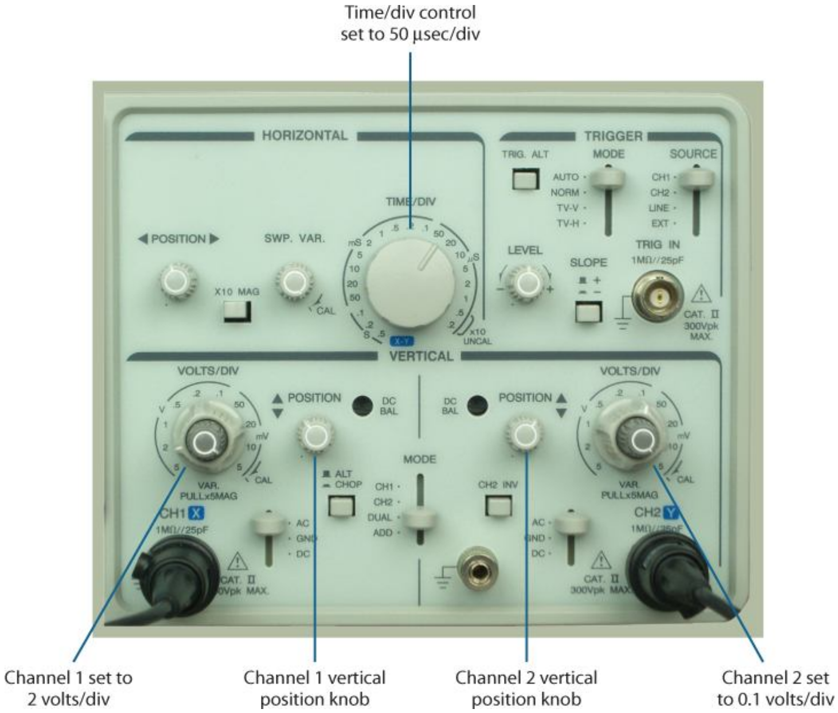
Time/div control
set to 50 µsec/div
HORIZONTAL
TRIGGER
TIME/DIV
VERTICAL
VOLTS/DIV
POSITION
MODE
CH1 X
CH2 Y
Channel 1 set to
2 volts/div
Channel 1 vertical
position knob
Channel 2 vertical
position knob
Channel 2 set
to 0.1 volts/div

Your values should be close to those shown in the following table.

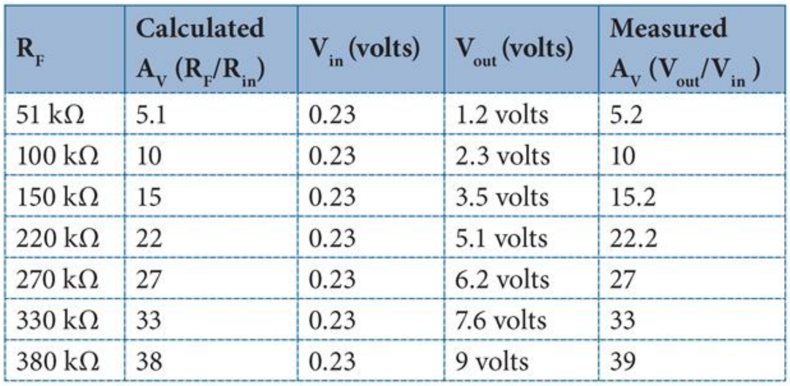

R_F	Calculated A_V (R_F/R_{in})	V_{in} (volts)	V_{out} (volts)	Measured A_V (V_{out}/V_{in})
51 kΩ	5.1	0.23	1.2 volts	5.2
100 kΩ	10	0.23	2.3 volts	10
150 kΩ	15	0.23	3.5 volts	15.2
220 kΩ	22	0.23	5.1 volts	22.2
270 kΩ	27	0.23	6.2 volts	27
330 kΩ	33	0.23	7.6 volts	33
380 kΩ	38	0.23	9 volts	39

The measured values of A V are quite close to the calculated values of A V , well within variations that could be caused by the ± 5 percent tolerance specified for resistor values.

Summary

This chapter introduced the most common types of amplifiers in use today: the common emitter BJT, the common source JFET, and the op-amp. At best, this chapter has scratched only the surface of the world of amplifiers. Actually, there are many variations and types of amplifiers. Still, the terminology and design approach you learned here should give you a basic foundation for further study.

Following are the key skills you gained in this chapter:

How to design a simple amplifier when the bias point and the gain are specified

How to do the same for an emitter follower

How to analyze a simple amplifier circuit

Self-Test

These questions test your understanding of the material presented in this chapter. Use a separate sheet of paper for your diagrams or calculations. Compare your answers with the answers provided following the test.

1. What is the main problem with the amplifier circuit shown in Figure 8.1 ? _____
2. What is the gain formula for that circuit? _____
3. Does it have a high or low gain? _____

Use the circuit shown in Figure 8.43 for questions 428.

Figure 8.43

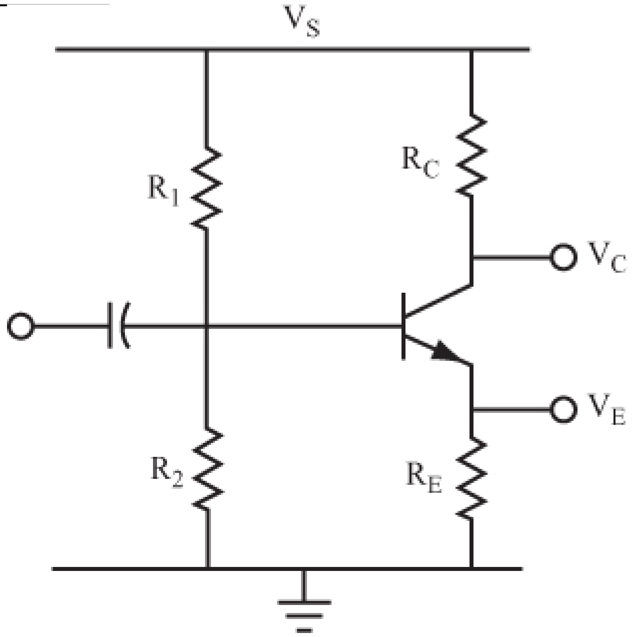

4. Design an amplifier so that the bias point is 5 volts, and the AC voltage gain is 15. Assume β = 75, R_{in} = 1.5 kΩ, V_S = 10 volts, and R_C = 2.4 kΩ. Add capacitor C_E to the circuit and calculate a suitable value to maintain maximum AC voltage gain at 50 Hz. What is the approximate value of this gain? _____

5. Repeat question 4 with these values: V_S = 28 volts, β = 80, R_{in} = 1 kΩ, and R_C = 10 kΩ. The bias point should be 14 volts and the AC voltage gain 20. _____

6. Repeat question 4 with these values: V_S = 14 volts, β = 250, R_{in} = 1 kΩ, and R_C

= 15 kΩ. The bias point should be 7 volts and the AC voltage gain 50. _____

7. Design an emitter follower amplifier given that V_S = 12 volts, R_E = 100 ohms, β = 35, V_E = 7 volts, and R_C = 0 ohms. _____

8. Design an emitter follower amplifier given that V_S = 28 volts, R_E = 100 ohms, β = 35, V_E = 7 volts, and R_C = 0 ohms. _____

In questions 9211, the resistance and β values are given. Analyze the circuit to find the bias point and the gain.

9. R_1 = 16 kΩ, R_2 = 2.2 kΩ, R_E = 100 ohms, R_C = 1 kΩ, β = 100, and V_S = 10 volts _____

10. R_1 = 36 kΩ, R_2 = 3.3 kΩ, R_E = 110 ohms, R_C = 2.2 kΩ, β = 50, and V_S = 12 volts _____

11. R_1 = 2.2 kΩ, R_2 = 90 kΩ, R_E = 20 ohms, R_C = 300 kΩ, β = 30, and V_S = 50 volts _____

12. The circuits from questions 4 and 5 are connected to form a two-stage amplifier. What is the gain when there is an emitter bypass capacitor for both transistors? When the capacitor is not used in either of them? _____

13. Design a JFET amplifier using the circuit shown in Figure 8.32 . The characteristics of the JFET are I_{DSS} = 20 mA and $V_{GS(off)}$ = 24.2 volts. The desired value of V_{DS} is 14 volts. Find the value of R_D . _____

14. If the transconductance of the JFET used in question 13 is 0.0048 mhos, what is the voltage gain? _____

15. If the desired output is 8 V_{pp} for the JFET of questions 13 and 14, what should the input be? _____

16. Design a JFET amplifier using the circuit in Figure 8.33. The JFET characteristics are I_{DSS} = 16 mA and $V_{GS(off)}$ = 22.8 volts. Using a V_{DS} of 10 volts, find the values of R_S, C_S, and R_D. _____

17. If the input to the JFET in question 16 is 20 mV_{pp}, what is the AC output voltage, and what is the gain? _____

18. For the op-amp circuit shown in Figure 8.35, what is the output voltage if the input is 50 mV and the feedback resistor is 750 kΩ? _____

Answers to Self-Test

If your answers do not agree with those provided here, review the problems in parentheses before you go on to Chapter 9, “Oscillators.”

1. Its bias point is unstable, and its gain varies with temperature. Also, you cannot guarantee what the gain will be. (problem 10)
2. $A_v = \beta \times \frac{R_c}{R_{in}}$ (problem 10)
3. Usually the gain is quite high. (problem

10)

For Numbers 426, suitable values are given. Yours should be close to these.

4. R_1 = 29 kΩ, R_2 = 3.82 kΩ, R_E = 160 ohms, C_E = 200 μF, A_V = 120 (problems 17222)

5. R_1 = 138 kΩ, R_2 = 8 kΩ, R_E = 500 ohms, C_E = 64 μF, A_V = 800 (problems 17222)

6. R_1 = 640 kΩ, R_2 = 45 kΩ, R_E = 300 ohms, C_E = 107 μF, A_V = 750 (problems 17222)

7. R_1 = 8 kΩ; R_2 = 11.2 kΩ (problem 27)

8. R_1 = 922 ohms; R_2 = 385 ohms (problem 27)

9. V_C = 5 volts, A_V = 10 (problems 28230)

10. V_C = 6 volts, A_V = 20 (problems 28230)

11. V_C = 30 volts, A_V = 15 (problems 28230)

12. When the capacitor is used, A_V = 120 × 800 = 96,000. (problems 17222)

When the capacitor is not used, A_V = 15 × 20 = 300.

13. Use V_{GS} = 22.1 volts, then I_D = 5 mA, R_D = 2 kΩ. (problems 31233)

14. A_v = 29.6 mV pp (problem 39)

15. V_{in} = 83 mV pp (problem 38)

16. Use V_{GS} = 21.4 volts, then I_D = 4

mA. (problem 42)

R_S = 350 ohms

C_S = 4.5 μF (assume f = 1 kHz)

R_D = 3.15 kΩ

17. V_{GS} varies from 21.39 to 21.41 volts, I_D varies from 4.06 to 3.94 mA, V_{out} will be 400 mV_{pp},

$A_v = \frac{-400}{20} = -20$ (problem 42)

18. A_v = 275, V_{out} = 3.75 volts (problem 45)

Chapter 9

Oscillators

This chapter introduces you to oscillators. An *oscillator* is a circuit that produces a continuous output signal. There are many types of oscillator circuits used extensively in electronic devices. Oscillators can produce a variety of different output signals, such as sine waves, square waves, or triangle waves.

When the output signal of an oscillator is a sine wave of constant frequency, the circuit is called a *sine wave oscillator* . Radio and television signals are sine waves transmitted through the air, and the 120-volts AC from the wall plug is a sine wave, as are many test signals used in electronics.

This chapter introduces three basic sine wave oscillators. They all rely on resonant LC circuits as described in Chapter 7, “Resonant Circuits,” to set the frequency of the sine wave.

When you complete this chapter, you will be able to do the following:

- Recognize the main elements of an oscillator.
- Differentiate between positive and negative feedback.
- Specify the type of feedback that causes a circuit to oscillate.
- Specify at least two methods of obtaining feedback in an oscillator circuit.

- Understand how resonant LC circuits set the frequency of an oscillator.
- Design a simple oscillator circuit.

Understanding Oscillators

1 An oscillator can be divided into three definite sections:

- An amplifier
- The feedback connections
- The components that set frequency

The amplifier replaces the switch in the basic oscillator circuit, introduced in problem 35 of Chapter 7 (Figure 7.46).

Question

Draw an oscillator circuit, and label the parts. Use a separate sheet of paper for your drawing.

Answer

See Figure 9.1 .

Figure 9.1

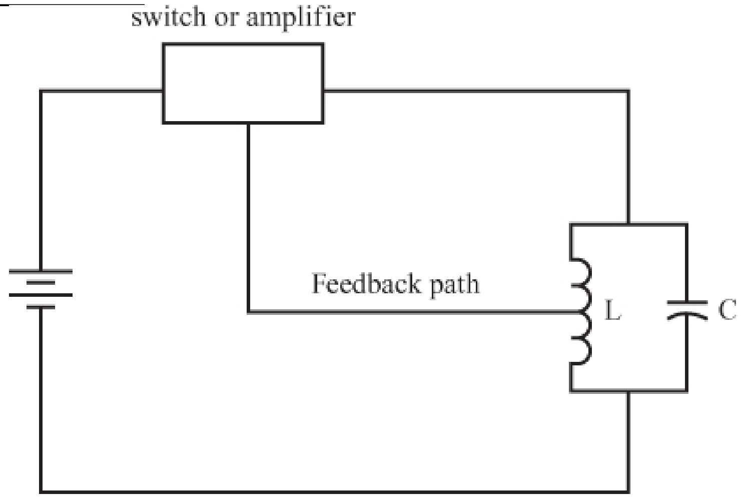

2 When you connect the output of an amplifier to its input, you get feedback. If the feedback is "out of phase" with the input, as shown in Figure 9.2, then the feedback is *negative*.

Figure 9.2

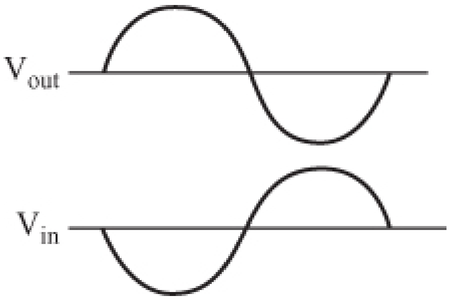

When the signal from the collector is fed back to the base of the transistor through a feedback resistor (R_f), as in the circuit shown in Figure 9.3, the feedback signal is out of phase with the input signal. Therefore, the feedback is negative.

Figure 9.3

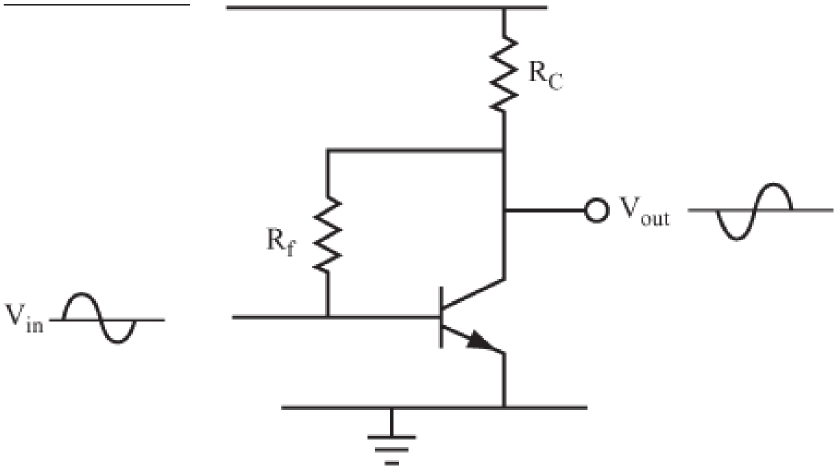

Negative feedback is used to stabilize the operation of an amplifier by doing the following:

- Preventing the DC bias point and gain of an amplifier from being affected by changes in temperature
- Reducing distortion in amplifiers, thereby improving the quality of the sound

Questions

A. Why would feedback signals be used in quality audio amplifiers? _____

B. What kind of feedback do they have? _____

Answers

A. To reduce distortion

B. Negative feedback

3 If the feedback from the output is in phase with the input, as shown in Figure 9.4, the circuit's feedback is *positive*.

Figure 9.4

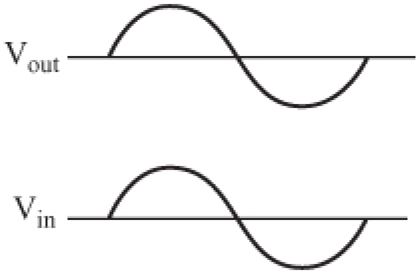

In the circuit shown in Figure 9.5, the collector of the second transistor is connected to the base of the first transistor. Because the output signal at the collector of the second transistor is in phase with the input signal at the base of the first transistor, this circuit has positive feedback.

Figure 9.5

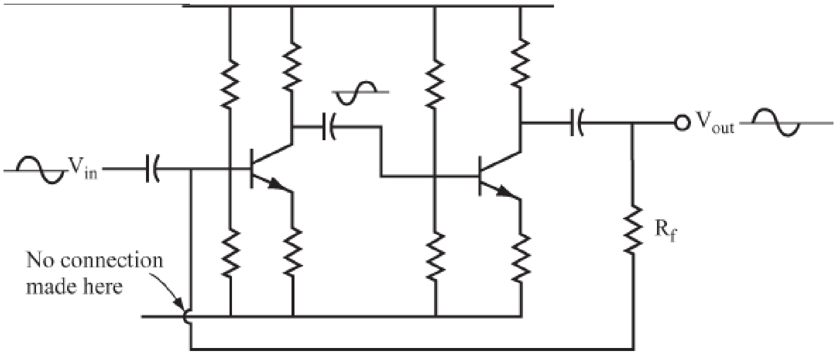

Positive feedback can cause an amplifier to oscillate even when there is *no external input.*

Questions

A. What type of feedback is used to stabilize an amplifier? _____

B. What type of feedback is used in oscillators? _____

C. What parts of an amplifier do you connect to produce feedback? _____

Answers

A. Negative feedback.

B. Positive feedback.

C. Connect the output of an amplifier to its input.

4 The amplifier shown in Figure 9.6 is the same type of amplifier that was discussed in problems 11–18 of Chapter 8, “Transistor Amplifiers.” It is called a *common emitter amplifier* .

Figure 9.6

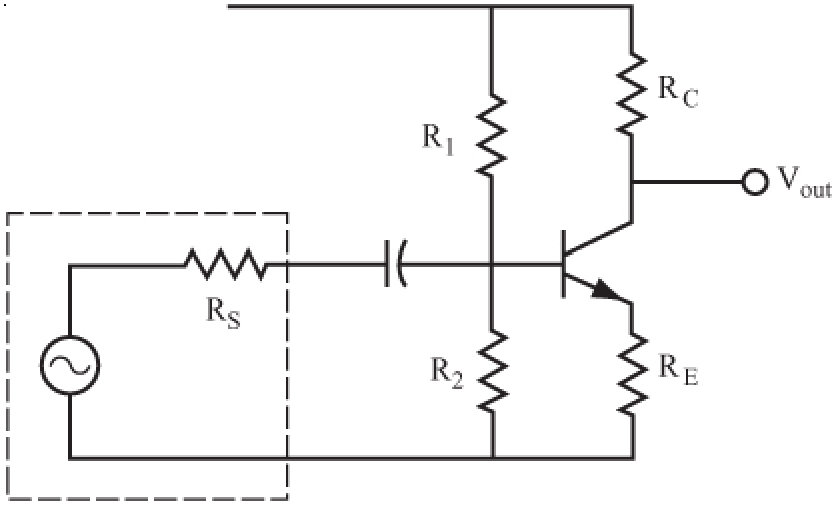

Questions

A. What effect would negative feedback have on this amplifier? _____

B. What effect would positive feedback have on this amplifier? _____

Answers

A. Stabilize it, reduce gain, and reduce distortion

B. Cause it to oscillate

5 In the circuit shown in Figure 9.6 , an input signal applied to the base will be amplified.

Questions

A. What is the basic formula for an amplifier's voltage gain? _____

B. What is the voltage gain formula for the amplifier circuit shown in Figure 9.6 ? _____

Answers

A.

$$A_V = \beta \times \frac{R_L}{R_{in}}$$

B. $$A_V = \frac{R_L}{R_E} = \frac{R_C}{R_E}$$

C. (as discussed in problem 12 of Chapter 8)

6 In the circuit shown in Figure 9.7, an input signal is applied to the emitter of the transistor instead of the base. This circuit is called a *common base amplifier*.

Figure 9.7

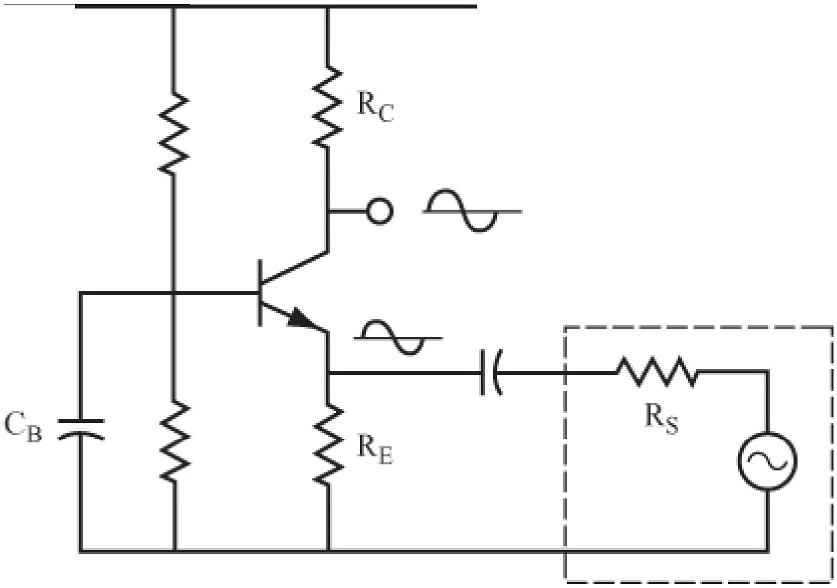

Note When you apply a signal to the emitter, it changes the voltage drop across the base-emitter diode, just as an input signal applied to the base does. Therefore, a signal applied to the emitter changes the base current and the collector current, just as if you had applied a signal to the base.

The voltage gain formula for this type of amplifier can be simplified because the input impedance to the amplifier is so low when the signal is fed into the emitter that you can discount it. This results in the following voltage gain formula for the common base amplifier:

$$A_V = \frac{R_L}{R_S}$$

R_S is the output resistance or impedance of the source or generator. It is also called the *internal impedance* of the source.

Question

What is the voltage gain formula for the circuit shown in Figure 9.7 ? _____

Answer

$$A_V = \frac{R_L}{R_S} = \frac{R_C}{R_S}, (R_C \text{ is the load in this circuit.})$$

7 Notice that the input and output sine waves in Figure 9.7 are in phase. Although the signal is amplified, it is not inverted.

Questions

A. What happens to the input signal to the amplifier when you apply it to the emitter instead of the base? _____

B. Is the input impedance of the common base amplifier high or low compared to the common emitter amplifier? _____

C. What is the gain formula for the common base amplifier? _____

Answers

A. Amplified and not inverted

B. Low

C. $A_V = \frac{R_L}{R_S} = \frac{R_C}{R_S}$

8 Figure 9.8 shows an amplifier circuit with a parallel inductor and capacitor connected between the collector of the transistor and ground. A parallel inductor and capacitor circuit is sometimes called a *tuned* (or *resonant*) *load*.

Figure 9.8

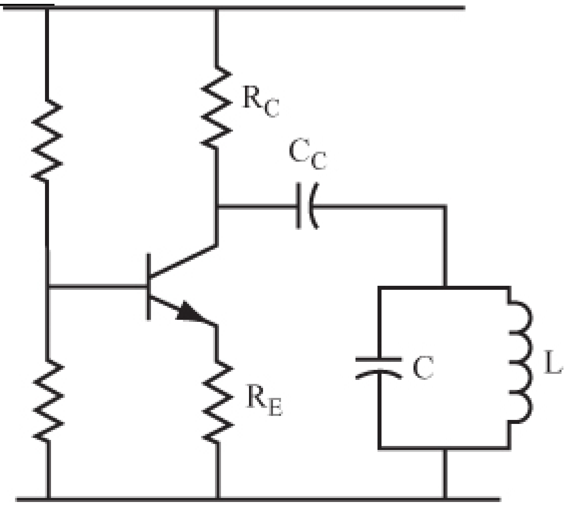

In this circuit, the inductor has a small DC resistance, which could pull the collector DC voltage down to near 0 volts. Therefore, you include capacitor C_C in the circuit to allow AC signals to pass through the LC circuit while preventing the collector DC voltage from being pulled down to 0 volts.

Questions

A. What term would you use to describe the load in this circuit? _____

B. Does the circuit contain all three components of an oscillator at this point? _____

Answers

A. Resonant or tuned.

B. No, the feedback connections are missing.

Note The circuit shown in Figure 9.8 does not have an input signal either to the emitter or to the base. By adding a feedback connection to a parallel LC circuit, you provide an input signal to the emitter or base, as explained later in this chapter.

9 Write the voltage gain formulas for the following circuits. Refer to the circuits and voltage gain formulas in problems 4–6, if necessary.

Questions

A. Common emitter circuit _____

B. Common base circuit _____

Answers

A. $A_V = \frac{R_C}{R_E}$

B. $A_V = \frac{R_C}{R_S}$

10 You can use common emitter and common base amplifier circuits in oscillators, and in each case, you would usually also include an extra capacitor.

In a common emitter amplifier, you can add a capacitor (C_E) between the emitter and ground, as discussed in problems 19 and 20 of Chapter 8.

In a common base circuit, you can add a capacitor (C_B) between the base and the ground, as is shown in Figure 9.7.

Question

What is the general effect in both cases? _____

Answer

An increase in the gain of the amplifier

The gain is increased to the point where you can consider it “large enough” to use the amplifier as an oscillator. When these capacitors are used in either a common emitter or common base amplifier, it is not usually necessary to calculate the gain of the amplifier.

11 An LC circuit has a resonance frequency that you can determine using the methods discussed in problems 6–12 of Chapter 7. When you use an LC circuit in an oscillator, the output signal of the oscillator will be at the resonance frequency of the LC circuit.

Question

What is the formula for the oscillation (or resonant) frequency? _____

Answer

$$f_r = \frac{1}{2\pi\sqrt{LC}}$$

In practice, the actual measured frequency is never quite the same as the calculated frequency. The capacitor and inductor values are not exact, and other stray capacitances in the circuit affect the frequency. When you need to set an exact frequency, use an adjustable capacitor or inductor.

12 Figure 9.9 shows the parallel LC circuit connected between the collector and the supply voltage, rather than between the collector and ground (as in Figure 9.8).

Figure 9.9

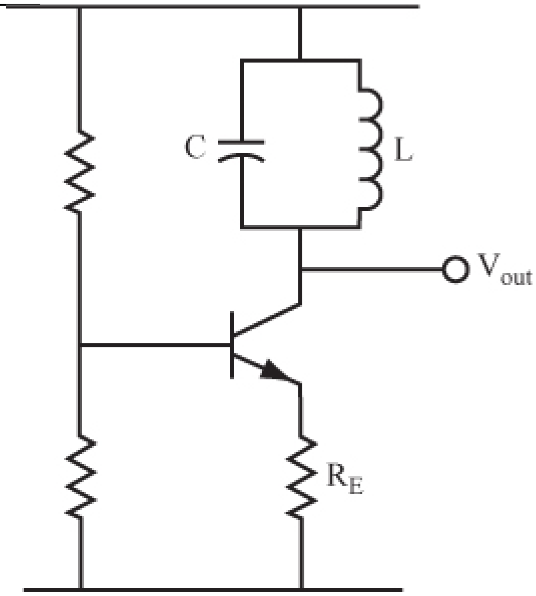

You can use this circuit and the circuit shown in Figure 9.8 to selectively amplify one frequency far more than others.

Questions

A. What would you expect this one frequency to be? _____

B. Write the formula for the impedance of the circuit at the resonance frequency. _____

C. What is the AC voltage gain at this frequency? _____

Answers

A. The resonance frequency

B. $$Z = \frac{L}{C \times r}$$

where r is the DC resistance of the coil

C. $$A_V = \frac{Z}{R_E}$$

13 Because of the low DC resistance of the coil, the DC voltage at the collector is usually close to the supply voltage (V_S). In addition, the AC output voltage positive peaks can exceed the DC level of the supply voltage. With large AC output, the positive peaks can actually reach $2V_S$, as shown in Figure 9.10.

Figure 9.10

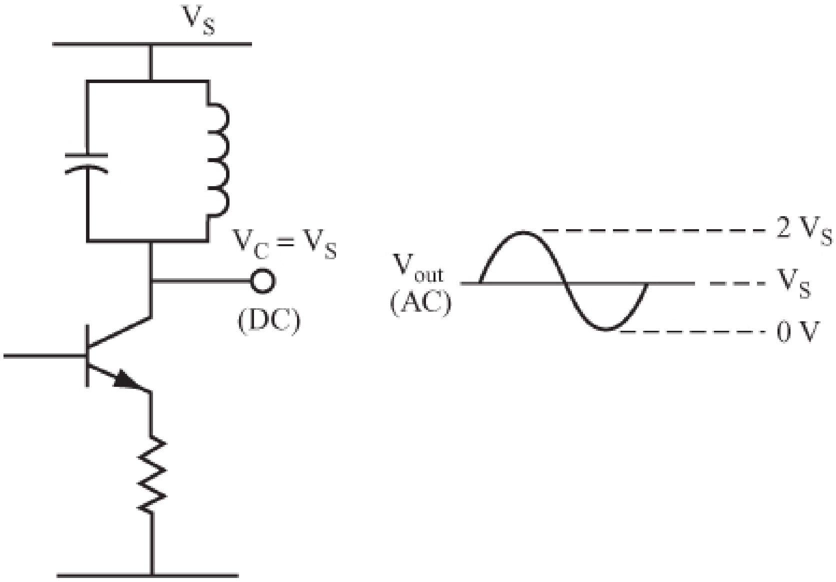

Question

Indicate which of the following is an accurate description of the circuit in Figure 9.10 :

A. Oscillator
B. Tuned amplifier
C. Common base circuit
D. Common emitter circuit

Answer

B. Tuned amplifier

Feedback

14 To convert an amplifier into an oscillator, you must connect a portion of the output signal to the input. This feedback signal must be in phase with the input signal to induce oscillations.

Figure 9.11 shows three methods you can use to provide a feedback signal from a parallel LC circuit. Each is named for its inventor.

Figure 9.11

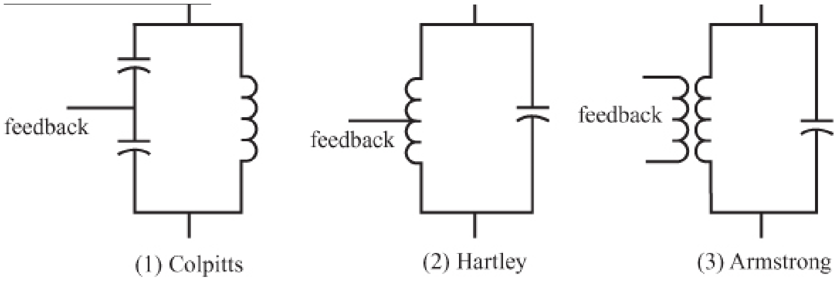

(1) Colpitts (2) Hartley (3) Armstrong

In the *Colpitts* method, the feedback signal is taken from a connection between two capacitors that form a voltage divider. In the *Hartley* method, the feedback signal is taken from a tap partway down the coil, or from a connection between two inductors. Therefore, an inductive voltage divider determines the feedback voltage. The *Armstrong* method uses a step-down transformer (an inductor with an extra coil with fewer turns than the main coil). In all three of these methods, between one-tenth and one-half of the output must be used as feedback.

Questions

A. Where is the feedback taken from in a Colpitts oscillator? _____

B. What type of oscillator uses a tap on the coil for the feedback voltage? _____

C. What type does not use a voltage divider? _____

Answers

A. A capacitive voltage divider

B. Hartley

C. Armstrong

15 The output voltage appears at one end of the parallel LC circuit shown in Figure 9.12 , and the other end is effectively at ground. The feedback voltage V_f is taken between the junction of the two capacitors.

Figure 9.12

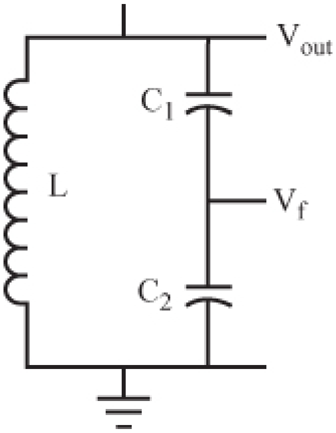

Question

Using the voltage divider formula, what is V_f? _____

Answer

$$V_f = \frac{V_{out} X_{C2}}{(X_{C1} + X_{C2})}$$

which becomes

$$V_f = \frac{V_{out} C_1}{(C_1 + C_2)}$$

16 To find the resonance frequency in this circuit, first find the equivalent total capacitance C_T of the two series capacitors.

You then use C_T in the resonance frequency formula.

Questions

A. What is the formula for C_T? _____

B. What is the resonance frequency formula for the Colpitts oscillator?_____

Answers

A. $$C_T = \frac{C_1 C_2}{C_1 + C_2}$$

B. $$f_r = \frac{1}{2\pi\sqrt{LC_T}},$$

if Q is equal to or greater than 10

Note If Q is less than 10, you can use one of the following two formulas to calculate the resonance frequency for a parallel LC circuit:

$$f_r = \frac{1}{2\pi\sqrt{LC}}\sqrt{1-\frac{r^2C}{L}} \text{ or } f_r = \frac{1}{2\pi\sqrt{LC}}\sqrt{\frac{Q^2}{1+Q^2}}$$

17 Figure 9.13 shows a parallel LC circuit in which the feedback voltage is taken from a tap N_1 turns from one end of a coil, and N_2 turns from the other end.

Figure 9.13

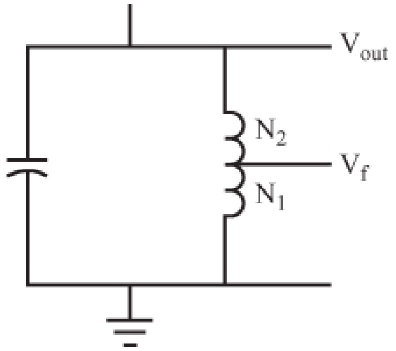

You can calculate the feedback voltage with a voltage divider formula that uses the number of turns in each part of the coil.

$$V_f = V_{out} \times \frac{N_1}{N_1 + N_2}$$

The manufacturer should specify N_1 and N_2.

Questions

A. Who invented this feedback method? _____

B. When you divide V_f by V_{out}, what is the result?

Answers

A. Hartley

B. Between one-tenth and one-half

18 Figure 9.14 shows a parallel LC circuit in which the feedback voltage is taken from the secondary coil of a transformer. The formula used to calculate the output voltage of a secondary coil is covered in problem 6 in Chapter 10, "The Transformer."

Figure 9.14

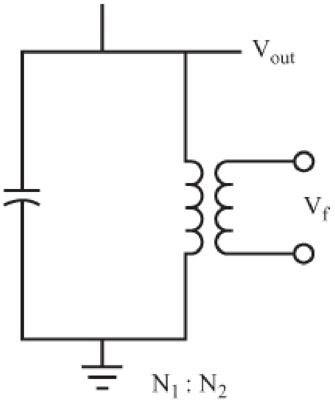

Question

Who invented this type of oscillator? _____

Answer

Armstrong

19 For each of the feedback methods described in the last few problems, the voltage fed back from the output to the input is a fraction of the total output voltage ranging between one-tenth and one-half of V_{out}.

To ensure oscillations, the product of the feedback voltage and the amplifier voltage gain must be greater than 1.

$$A_V \times V_f > 1$$

It is usually easy to achieve this because A_v is much greater than 1.

No external input is applied to the oscillator. Its input is the small part of the output signal that is fed back. If this feedback is of the correct phase and amplitude, the oscillations start spontaneously and continue as long as power is supplied to the circuit.

The transistor amplifier amplifies the feedback signal to sustain the oscillations and converts the DC power from the battery or power supply into the AC power of the oscillations.

Questions

A. What makes an amplifier into an oscillator? _____

B. What input does an amplifier need to

become an oscillator?_____
Answer

A. A resonant LC circuit with feedback of the correct phase and amount.

B. None. Oscillations happen spontaneously if the feedback is correct.

Inside the Inductor

When you use inductors, you should know how to deal with the different ways manufacturers label them.

Inductors are simply a coil of wire wrapped around a core, and some manufacturers leave them just like that. These inductors come with no markings,so you must keep them with the label from the packaging to identify them.

You can also find inductors that have a plastic coating around the wire coil. That coating is often marked with a numerical code that identifies the value of the inductor. The first two numbers are the first and second significant digits of the inductance value; the third number is the multiplier. (The units are µH, so an inductor marked with 101 has a value of 100 µH.)

Another method to mark inductors involves the same color code used for resistors. With this method,an inductor is marked with four color bands to show its value and tolerance.

Some inductors have a wide silver band

(about twice the width of the other bands) at the front of the color code bands. This wider band indicates that the component was built to a U.S. military specification and is not used to determine the inductance value.

The value of each color used in the bands is shown in the following table (with units in μH):

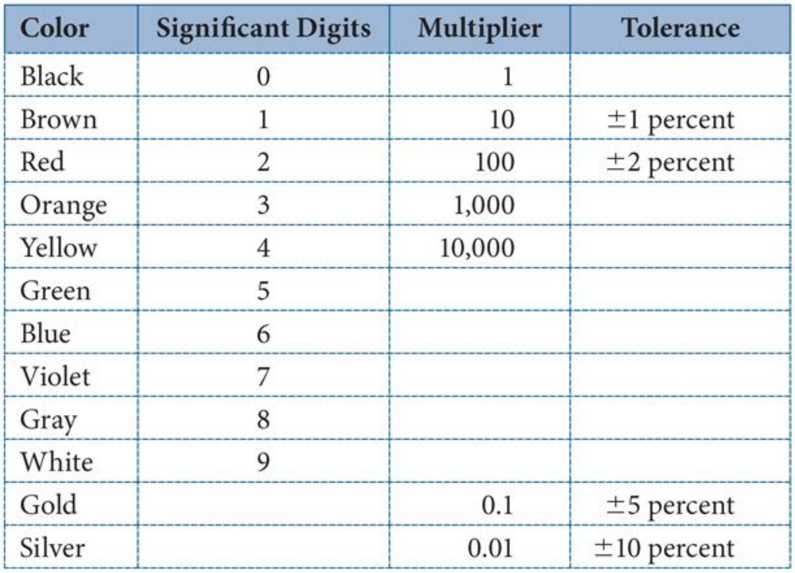

Color	Significant Digits	Multiplier	Tolerance
Black	0	1	
Brown	1	10	±1 percent
Red	2	100	±2 percent
Orange	3	1,000	
Yellow	4	10,000	
Green	5		
Blue	6		
Violet	7		
Gray	8		
White	9		
Gold		0.1	±5 percent
Silver		0.01	±10 percent

The first two colored bands are the first and second significant digits of the inductance value. The third band is the multiplier and the fourth band represent the tolerance. For example, if an inductor is marked with blue, gray, red, and silver bands, its nominal inductance value is 6800 μH (6.8 mH) with a tolerance of 610 percent .

Finally, you see inductors that have the value simply printed on them. This generally occurs on higher-value inductors, which are also physically larger.

Your best bet is to save the label on the packaging that the inductor comes in until you can check out the markings on the component.

The Colpitts Oscillator

20 Figure 9.15 shows a Colpitts oscillator circuit, the simplest of the LC oscillators to build.

Figure 9.15

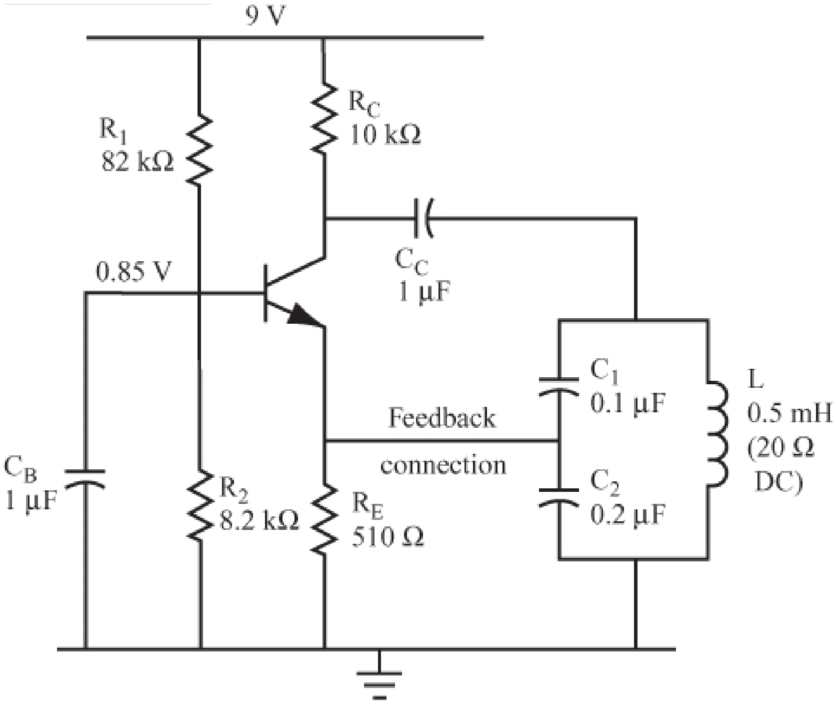
9 V
R1
82 kΩ
RC
10 kΩ
CC
1 µF
0.85 V
C1
0.1 µF
L
0.5 mH
(20 Ω
DC)
Feedback
connection
C2
0.2 µF
CB
1 µF
R2
8.2 kΩ
RE
510 Ω

The feedback signal is taken from the capacitive voltage divider and fed to the emitter. This connection provides a feedback signal to the emitter in the phase required to provide positive feedback.

In this circuit, the reactance of capacitor C_B is low enough for the AC signal to pass through it, rather than passing through R_2. Capacitor C_B should have a reactance, X_{CB}, of less than 160 ohms at the oscillation frequency. If R_2 happens to be smaller than 1.6 kΩ, choose a value of X_{CB} that is less than one-tenth of R2.

Question

For the circuit shown in Figure 9.15, what is your first estimate for C_B? Assume that f_r is equal to 1 kHz, and that X_c equals 160 ohms._____

Answer

$$X_{CB} = 160\,\text{ohms} = \frac{1}{2\pi f C_B} = \frac{1}{2\times\pi\times10^3\times C_B}$$

Therefore, C_B = 1 μF; larger values of C_B also work.

21 Use the Colpitts oscillator component values shown in Figure 9.15 to answer the following questions.

Questions

A. What is the effective total capacitance of the two series capacitors in the tuned circuit?

C_T = _____

B. What is the oscillator frequency?

f_r = _____

C. What is the impedance of the tuned circuit at this frequency?

Z = _____

D. What fraction of the output voltage is fed back?

V_f = _____

E. What is the reactance of C_B at the frequency of oscillation?

X_{CB} = _____

Answers

A. C_T = 0.067 μF.

B. Because Q is not known, use the formula that includes the resistance of the coil (see problem 16):

$$f_r = 26.75\ \text{kHz}.$$

If you use the calculated value of f_r to calculate Q, as in problem 20 of Chapter 7, you find that Q = 4.2. Therefore, it is appropriate to use the formula that includes the resistance of the coil to calculate f_r.

C. Use the following:

$$Z_T = \frac{L}{rC},\ Z_T = 373\ \text{ohms}.$$

D. Use a voltage divider with the capacitor values:

$$V_f = V_{out}\frac{C_1}{(C_1 + C_2)} = \frac{V_{out}}{3}.$$

E. X_{CB} = about 6 ohms, which is a good value (much less than the 8200 ohm value of R_2).

22 Figure 9.16 shows a Colpitts oscillator circuit that uses a different method for making feedback connections between the parallel LC circuit and the transistor.

Figure 9.16

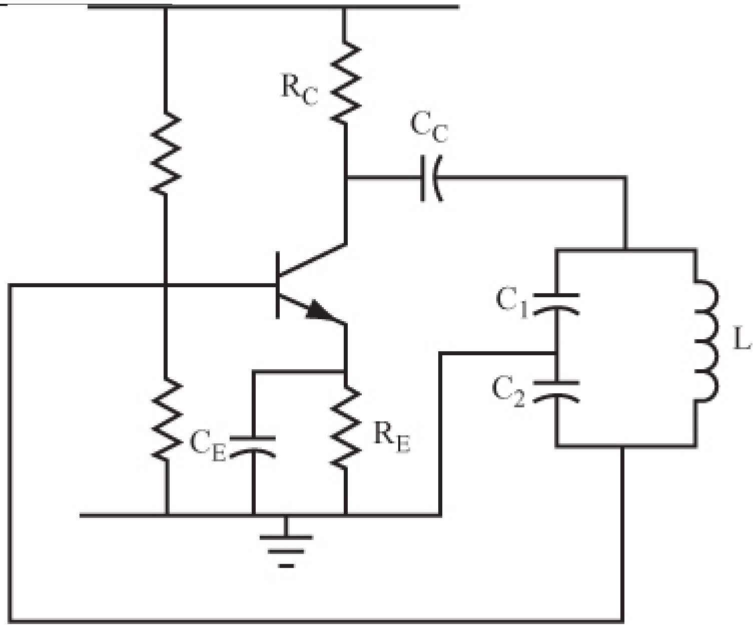

Question

List the differences between this circuit and the one shown in Figure 9.15 . _____

Answer

The feedback is connected to the base instead of the emitter, and the ground is connected to the center of the capacitive voltage divider. The capacitor C E has been added. (This connection provides a feedback signal to the base in the correct phase to provide positive feedback.)

23 In the circuit shown in Figure 9.16 , capacitor C E should have a reactance of less than 160 ohms at the oscillation frequency. If the emitter resistor R_E is smaller in value than 1.6 kΩ, then C_E should have a reactance that is less than $R_E/10$ at the oscillation frequency.

Question

If you use an emitter resistor of 510 ohms in a 1 kHz oscillator, what value of capacitor should you use for C_E?_____

Answers

$$X_C = \frac{510}{10} = \frac{1}{2\pi f C_E} = \frac{0.16}{10^3 \times C_E}$$

So, $C_E = 3.2$ μF. Thus, you should use a capacitor larger than 3 μF.

Project 9.1: The Colpitts Oscillator

Objective

The objective of this project is to demonstrate that an oscillator generates a sine wave when feedback is applied to either the emitter or base.

General Instructions

When the Colpitts oscillator circuit with feedback to the emitter is set up, you use your oscilloscope to measure the period of the waveform. Then you change the circuit to provide feedback to the base, and again use your oscilloscope to measure the period of the waveform. This data enables you to calculate the frequency of the sine wave generated in each case.

Parts List

You need the following equipment and

supplies:

- One 10 kΩ, 0.25-watt resistor.
- One 510 Ω, 0.25-watt resistor.
- One 82 kΩ, 0.25-watt resistor.
- One 8.2 kΩ, 0.25-watt resistor.
- Two 1 μF capacitors (This value of capacitor is available in either polarized or unpolarized versions.You should get unpolarized capacitors for this application.)
- One 0.1 μF capacitor.
- One 0.22 μF capacitor.
- One 4.7 μF capacitor. (This value of capacitor is usually polarized, which is fine for this position in the circuit.)
- One 0.5 mH inductor. (Suppliers may also refer to this value as 500 μH.)
- One 9-volt battery pack.
- One breadboard.
- One oscilloscope.
- One PN2222 transistor. Figure 9.17 shows the pinout diagram for PN2222 transistors.

Figure 9.17

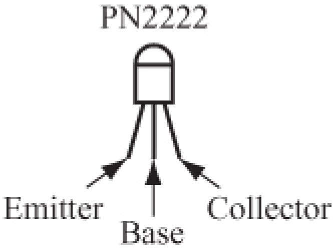

Step-by-Step Instructions

Set up Circuit #1, the Colpitts oscillator circuit with feedback to the emitter, as shown in Figure 9.18 . If you have some experience in building circuits, this schematic (along with the previous parts list) should provide all the information you need to build the circuit. If you need a bit more help building the circuit, look at the photos of the completed circuit in the “Expected Results” section.

Figure 9.18

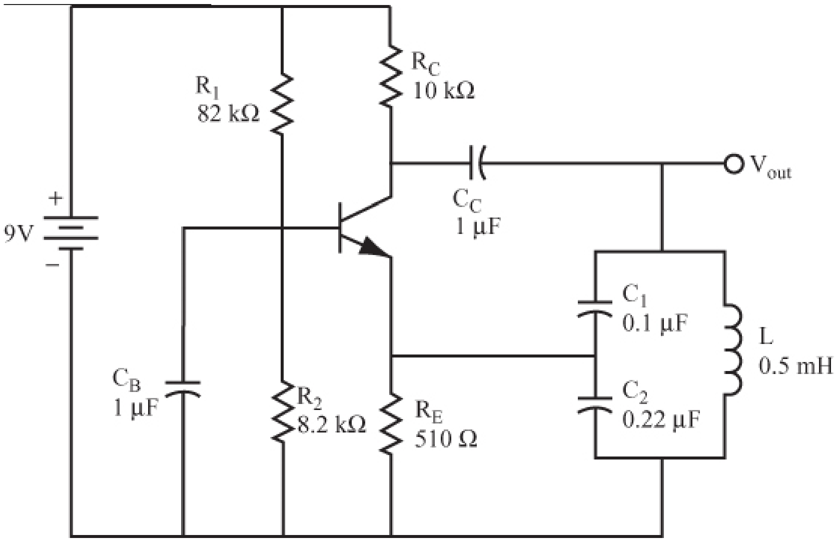

Carefully check your circuit against the diagram.

After you have checked your circuit, follow these steps and record your measurements in the blank table following the steps.

1. Connect the oscilloscope probe for Channel 1 to a jumper wire connected to V out . Connect the ground clip to a jumper wire attached to the ground bus.
2. Measure and record the period of the sine wave.
3. Disconnect the battery from the circuit, and

make the changes required to set up Circuit #2, the Colpitts oscillator circuit with feedback to the base, as shown in Figure 9.19 .

Figure 9.19

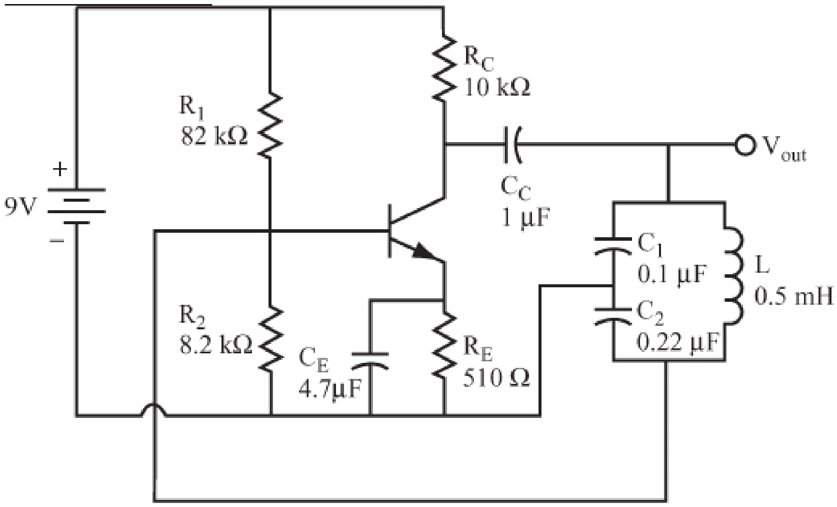
R$_C$
10 kΩ
R$_1$
82 kΩ
+
9V
−
C$_C$
1 µF
V$_{out}$
C$_1$
0.1 µF
L
0.5 mH
C$_2$
0.22 µF
R$_2$
8.2 kΩ
C$_E$
4.7µF
R$_E$
510 Ω

4. Repeat steps 1 and 2.

Circuit #	Period (µsec)	Frequency (kHz)
1		
2		

Expected Results

Figure 9.20 shows the breadboarded Colpitts oscillator with feedback to the emitter (Circuit# 1).

Figure 9.20

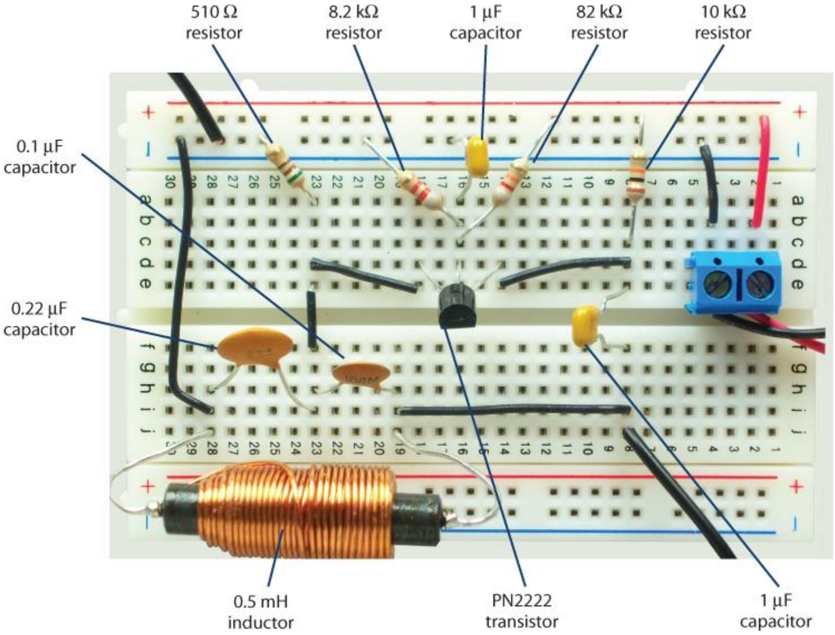
510 Ω resistor
8.2 kΩ resistor
1 µF capacitor
82 kΩ resistor
10 kΩ resistor
0.1 µF capacitor
0.22 µF capacitor
0.5 mH inductor
PN2222 transistor
1 µF capacitor

Figure 9.21 shows an oscilloscope attached to the circuit.

Figure 9.21

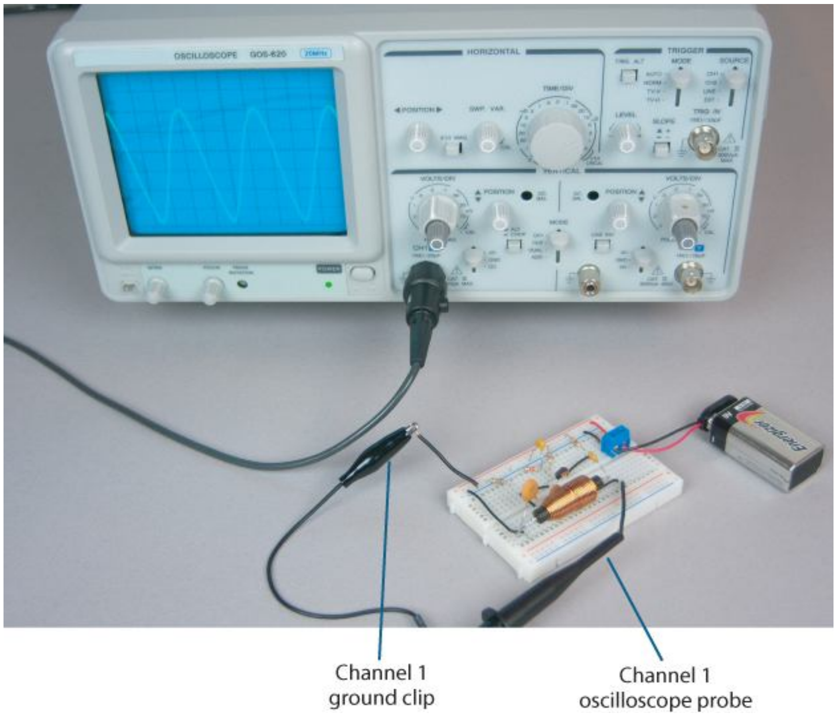
Channel 1
ground clip
Channel 1
oscilloscope probe

Figure 9.22 shows the sine wave generated by the Colpitts oscillator with feedback to the emitter. You can determine the period of this waveform by counting the number of horizontal divisions the waveform takes to complete one cycle, and then multiplying the number of divisions by the TIME/DIV setting.

Figure 9.22

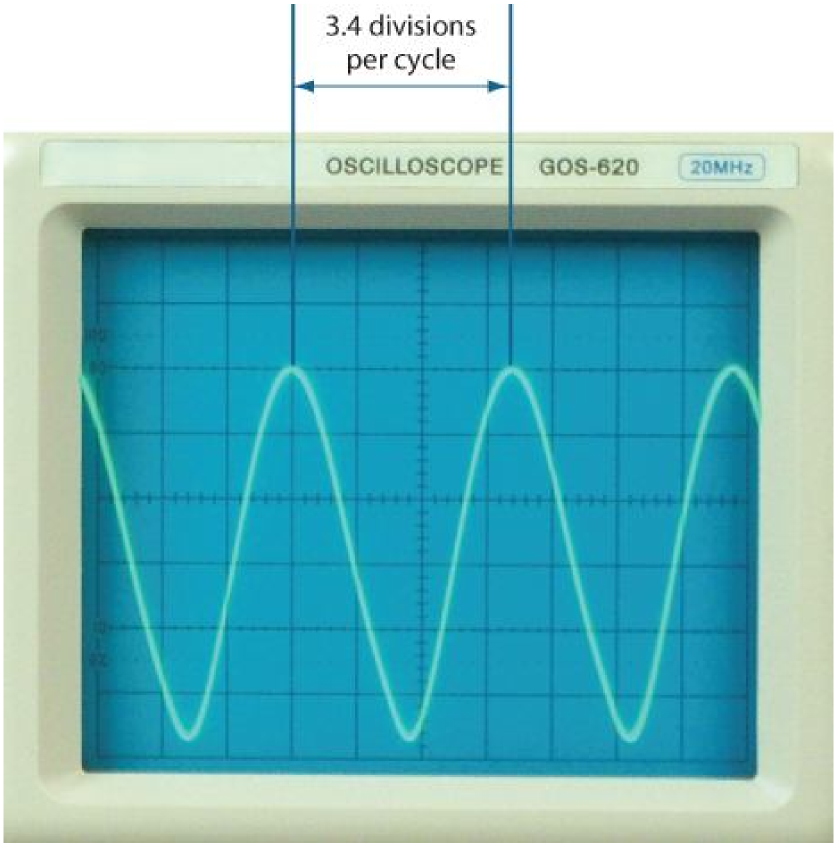
3.4 divisions per cycle
OSCILLOSCOPE
GOS-620
20MHz

As you measure the period, you may need to adjust the TIME/DIV, the horizontal POSITION, and the vertical POSITION controls on the oscilloscope. The controls shown in Figure 9.23 are adjusted to measure the period for the Colpitts oscillator.

Figure 9.23

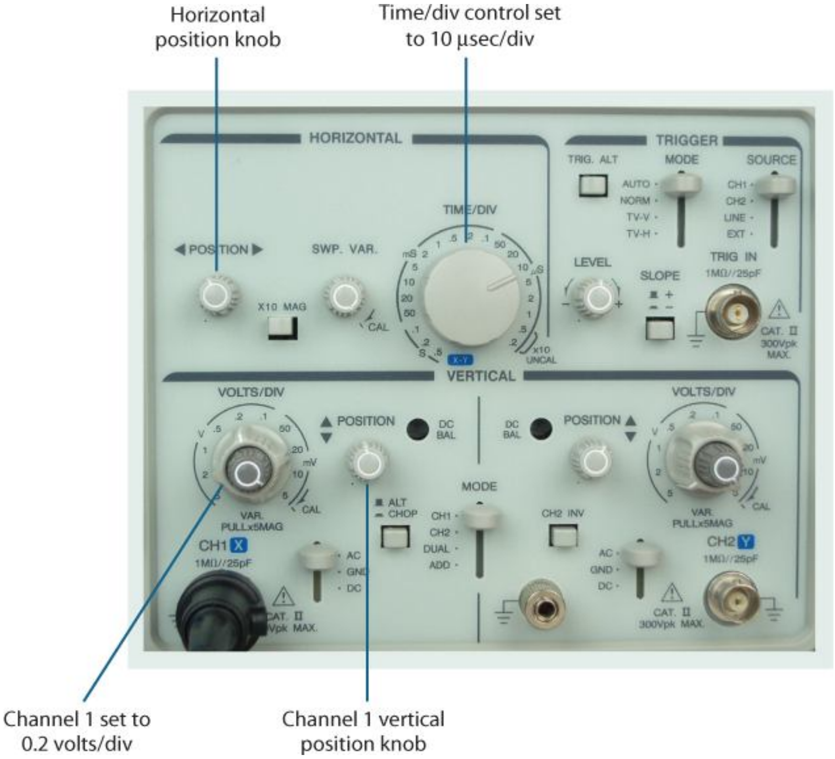
Horizontal position knob
Time/div control set to 10 µsec/div
HORIZONTAL
TRIGGER
VERTICAL
Channel 1 set to 0.2 volts/div
Channel 1 vertical position knob

Figure 9.24 shows the breadboarded Colpitts oscillator with feedback to the base (Circuit# 2).

Figure 9.24

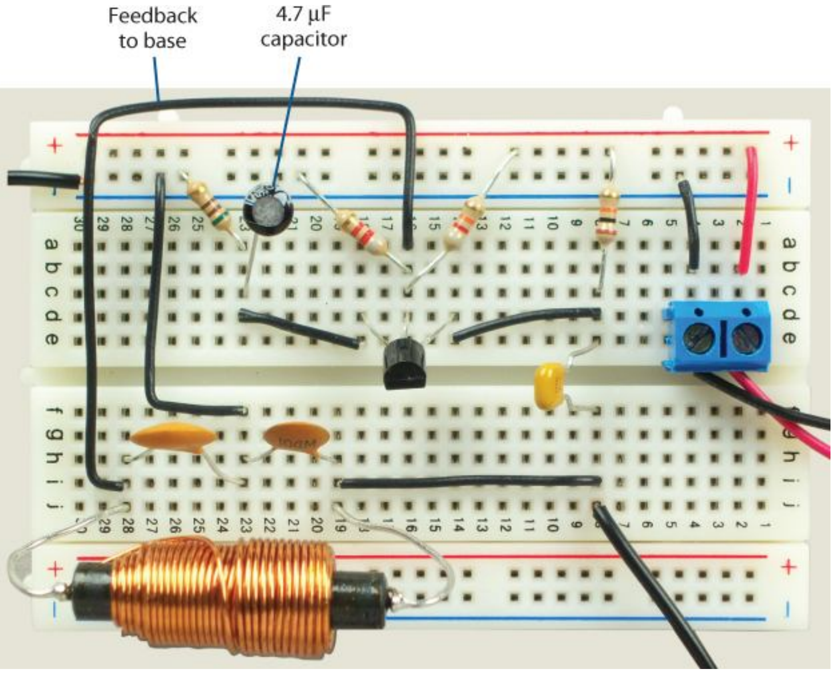

Figure 9.25 shows the sine wave generated by the Colpitts oscillator with feedback to the base. You can determine the period of this waveform by counting the number of horizontal divisions the waveform takes to complete one cycle, and then multiplying the number of divisions by the TIME/DIV setting. The oscilloscope connections and oscilloscope control panel settings for the Colpitts oscillator with feedback to the base are not shown. They are the same as the oscilloscope connections and oscilloscope control panel for the Colpitts oscillator with feedback to the emitter.

Figure 9.25

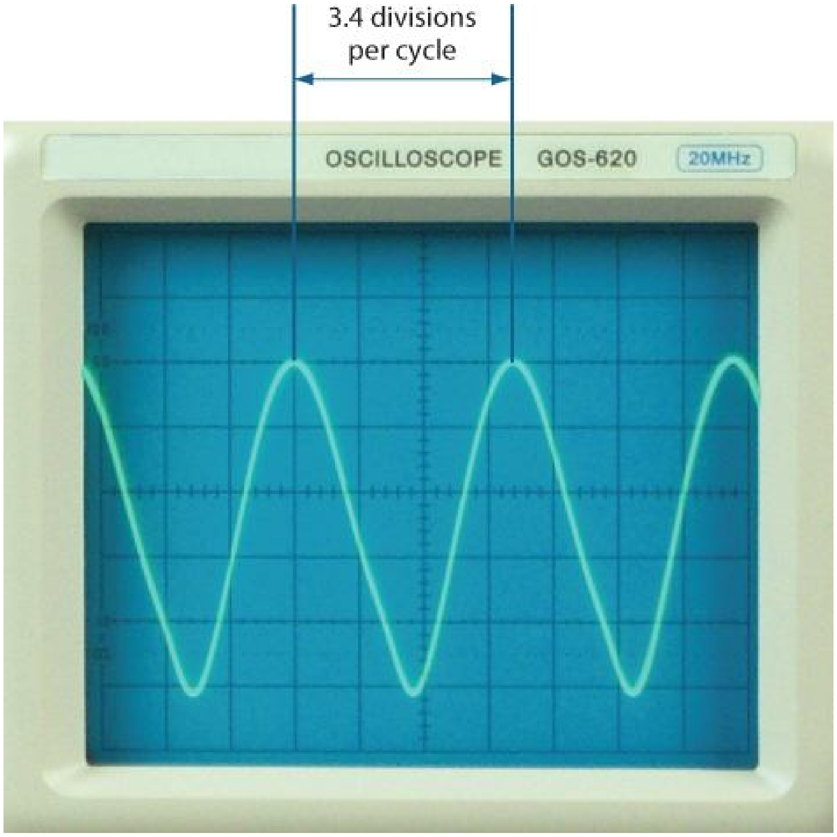
3.4 divisions per cycle
OSCILLOSCOPE GOS-620
20MHz

Your values should be close to those shown in the following table:

Circuit #	Period (μsec)	Frequency (kHz)
1	34	29.4
2	34	29.4

Notice that the frequency of the sine wave generated by both circuits is the same. This demonstrates that an oscillator can function with feedback to either the emitter or base of the transistor.

---

24 Figure 9.26 shows a Colpitts oscillator with the parallel LC circuit connected between the collector and the supply voltage. As with the circuits shown in Figures 9.15 and 9.16, this circuit provides a feedback signal to the transistor (in this case, the emitter) in the correct phase to provide positive feedback.

Figure 9.26

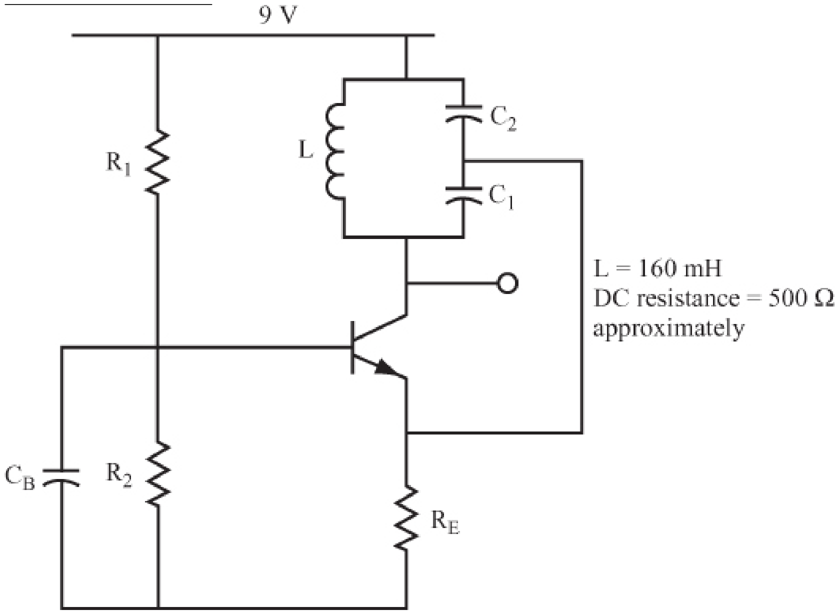
9 V
R1
L
C2
C1
L = 160 mH
DC resistance = 500 Ω
approximately
CB
R2
RE

The following table shows possible values you might use for C 1 and C 2 in the circuit shown in Figure 9.26:

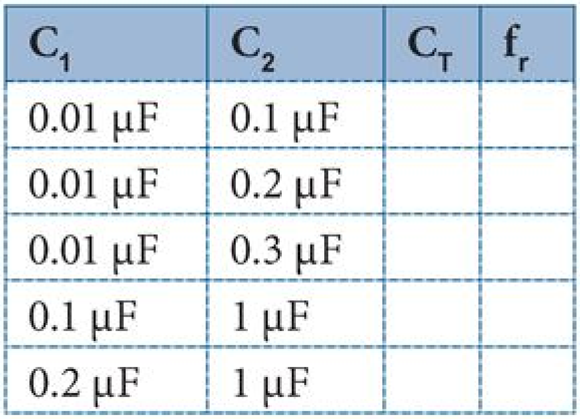

C_1	C_2	C_T	f_r
0.01 µF	0.1 µF		
0.01 µF	0.2 µF		
0.01 µF	0.3 µF		
0.1 µF	1 µF		
0.2 µF	1 µF		

Questions

A. Calculate C_T and f_r for each row of the preceding table._____

B. Does increasing C_2, while holding C_1 constant, increase or decrease the resonance frequency?_____

C. What effect does increasing C_1 have on the resonance frequency?_____

D. What is the condition that results in the highest possible resonance frequency?_____

E. What would be the highest resonance frequency if C_1 is fixed at 0.01 µF, and C_2 can vary from 0.005 µF to 0.5 µF?_____

Answers

A. The following table shows the values of C_T and f_r:

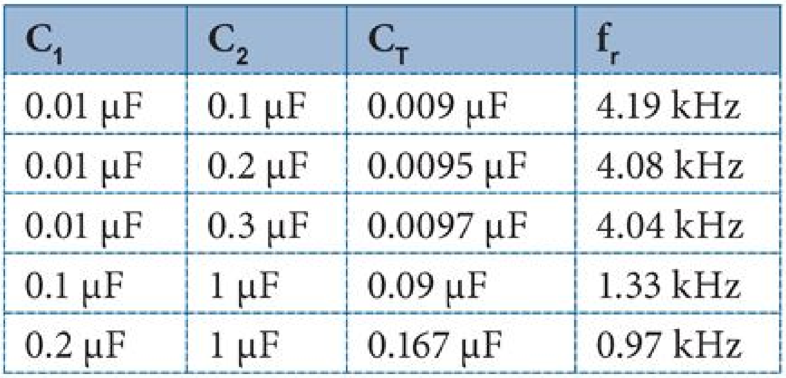

C_1	C_2	C_T	f_r
0.01 μF	0.1 μF	0.009 μF	4.19 kHz
0.01 μF	0.2 μF	0.0095 μF	4.08 kHz
0.01 μF	0.3 μF	0.0097 μF	4.04 kHz
0.1 μF	1 μF	0.09 μF	1.33 kHz
0.2 μF	1 μF	0.167 μF	0.97 kHz

B. Increasing C_2 decreases the resonance frequency, and, therefore, decreases the output frequency of the oscillator.

C. Increasing C_1 also decreases the resonance frequency and the output frequency of the oscillator.

D. When C_T is at its lowest possible value.

E. When C_2 is 0.005 μF, C_T will be 0.0033 μF, which is its lowest possible value. Therefore, the frequency is at the highest possible value, or approximately 6.9 kHz. The lowest frequency occurs when C_2 is at its highest setting of 0.5 μF.

The Hartley Oscillator

25 Figure 9.27 shows a Hartley oscillator circuit. In this type of circuit, the feedback is taken from a tap on the coil, or from a connection between two inductors.

Figure 9.27

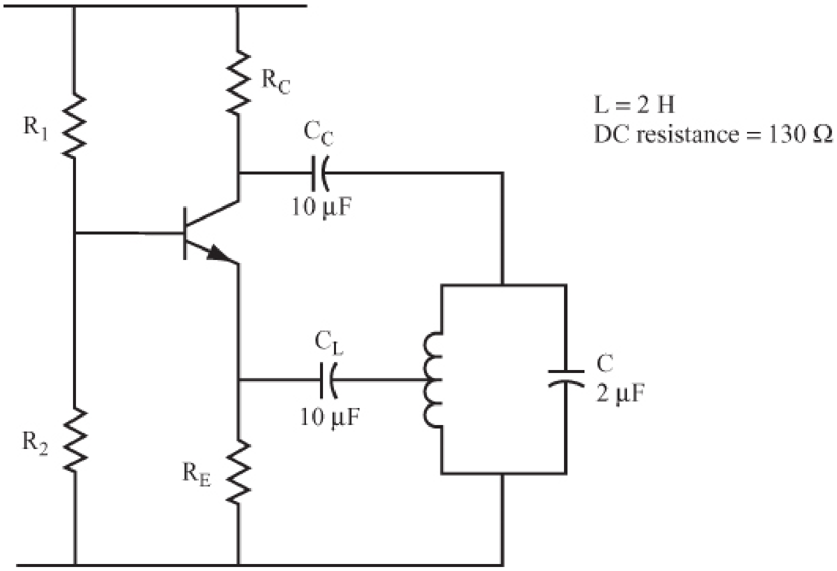

Capacitor C_L stops the emitter DC voltage from being pulled down to 0 volts through the coil. C_L should have a reactance of less than $R_E/10$, or less than 160 ohms at the oscillator frequency.

Questions

Work through the following calculations:

A. What is the resonance frequency?

f_r = _____

B. What is the approximate impedance of the load? Z = _____

C. What missing information prevents you from calculating the fraction of the voltage drop across the coil that is fed back to the emitter?

Answers

A. 80 Hz (approximately).

B. 7.7 kΩ (approximately).

C. The number of turns in the coil and the position of the tap are not known.

Figure 9.28 shows a Hartley oscillator with the parallel LC circuit connected between the collector and the supply voltage. As with the circuit shown in Figure 9.27, this circuit provides a feedback signal to the emitter from a tap in the coil, in the correct phase to provide positive feedback.

Figure 9.28

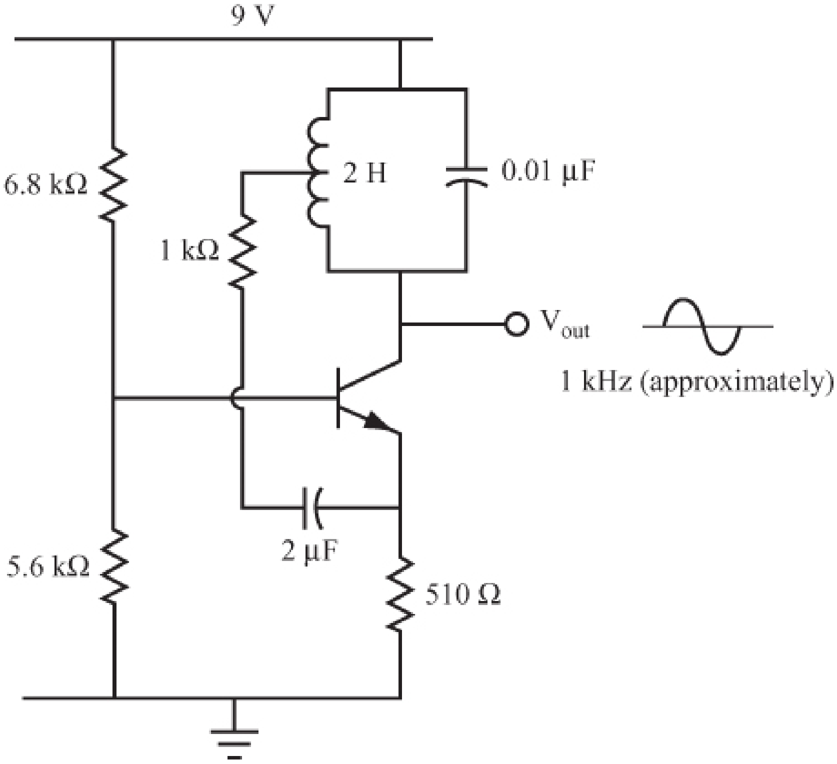
9 V
6.8 kΩ
2 H
0.01 μF
1 kΩ
V_{out}
1 kHz (approximately)
2 μF
5.6 kΩ
510 Ω

Project 9.2: The Hartley Oscillator

Objective

The objective of this project is to demonstrate a Hartley oscillator using two inductors in series.

General Instructions

After the Hartley oscillator circuit is set up, you use your oscilloscope to measure the period of the waveform, from which you can calculate the frequency of the oscillator. You also calculate the frequency from the inductance and capacitance used in the parallel LC circuit. Note that when two inductors in series are used, rather than a tapped coil, the total inductance is found by adding the individual inductance values, using the following equation:

$L_T = L_1 + L_2$

Parts List

You need the following equipment and supplies:

- One 10 kΩ, 0.25-watt resistor.
- One 510 Ω, 0.25-watt resistor.
- One 82 kΩ, 0.25-watt resistor.
- One 8.2 kΩ, 0.25-watt resistor.
- Three 1 μF capacitors. (This value of capacitor is available in either polarizied or unpolarized versions.You should get unpolarized capacitors for this application.)
- One 0.01 μF capacitor.
- One 6.8 mH inductor.
- One 3.1 mH inductor.
- One 9-volt battery pack.

- One breadboard.
- One oscilloscope.
- One PN2222 transistor. Figure 9.29 shows the pinout diagram for PN2222 transistors.

Figure 9.29

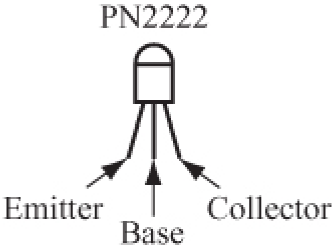

Step-by-Step Instructions

Set up the Hartley oscillator circuit shown in Figure 9.30 . If you have some experience in building circuits, this schematic (along with the previous parts list) should provide all the information you need to build the circuit. If you need a bit more help building the circuit, look at the photos of the completed circuit in the “Expected Results” section.

Figure 9.30

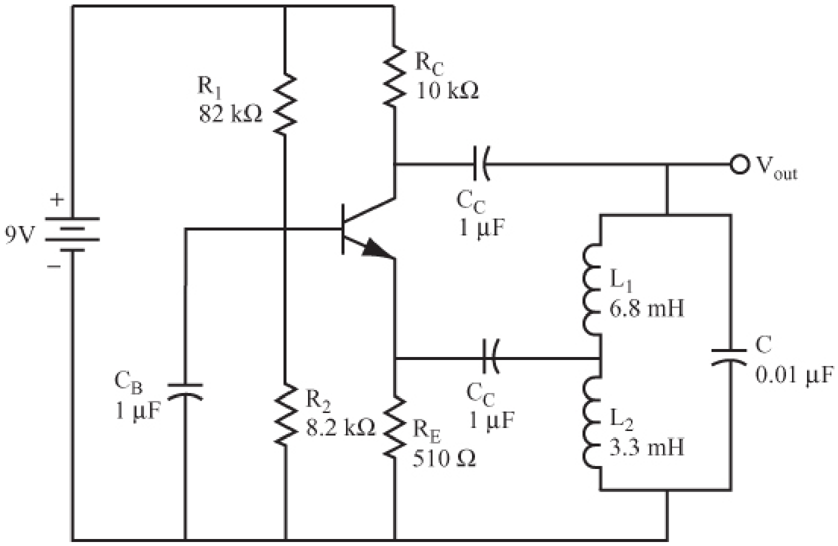
R$_1$
82 kΩ
R$_C$
10 kΩ
V$_{out}$
+
9V
−
C$_C$
1 µF
L$_1$
6.8 mH
C
0.01 µF
C$_B$
1 µF
R$_2$
8.2 kΩ
R$_E$
510 Ω
C$_C$
1 µF
L$_2$
3.3 mH

Carefully check your circuit against the diagram.

When you have checked your circuit, follow these steps.

1. Connect the oscilloscope probe for Channel 1 to a jumper wire connected to V out . Connect the ground clip to a jumper wire attached to the ground bus.

2. Measure and record the period of the sine wave.

Period = _____

3. Calculate the frequency of the sine wave.

Frequency = _____

4. Calculate the expected resonance frequency from the value of the capacitor and inductors used in the parallel LC circuit using the following equation:

$$f_r = \frac{1}{2\pi\sqrt{L_T C}}$$

f_r = _____

Expected Results

Figure 9.31 shows the breadboarded Hartley oscillator.

Figure 9.31

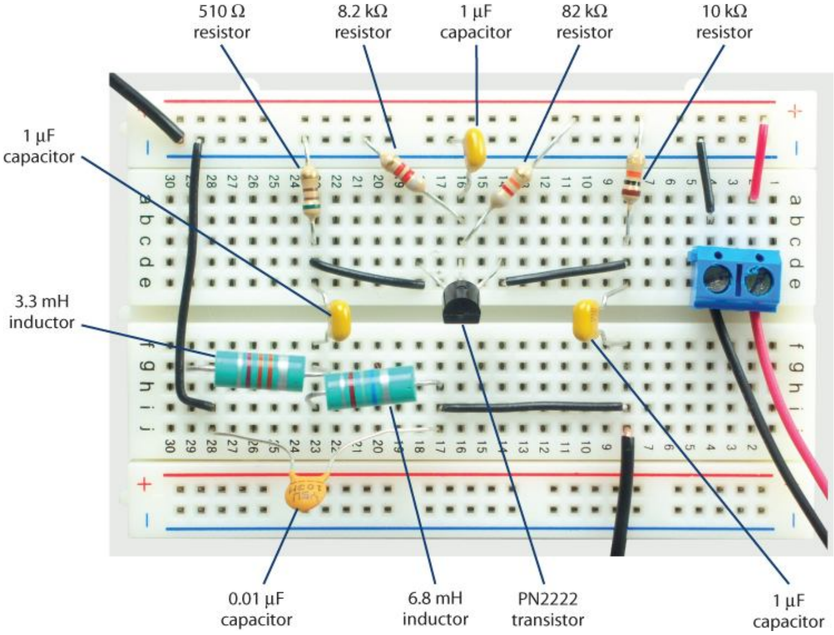
510 Ω resistor
8.2 kΩ resistor
1 μF capacitor
82 kΩ resistor
10 kΩ resistor
1 μF capacitor
3.3 mH inductor
0.01 μF capacitor
6.8 mH inductor
PN2222 transistor
1 μF capacitor

Figure 9.32 shows an oscilloscope attached to the circuit.

Figure 9.32

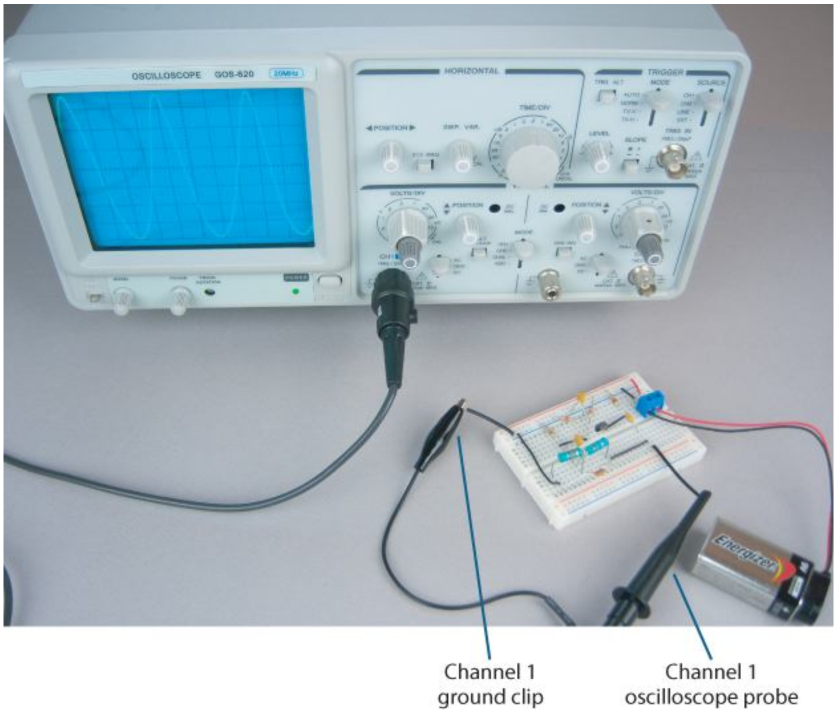
Channel 1
ground clip
Channel 1
oscilloscope probe

Figure 9.33 shows the sine wave generated by the Hartley oscillator. You can determine the period of this waveform by counting the number of horizontal divisions the waveform takes to complete one cycle, and then multiplying the number of divisions by the TIME/DIV setting.

Figure 9.33

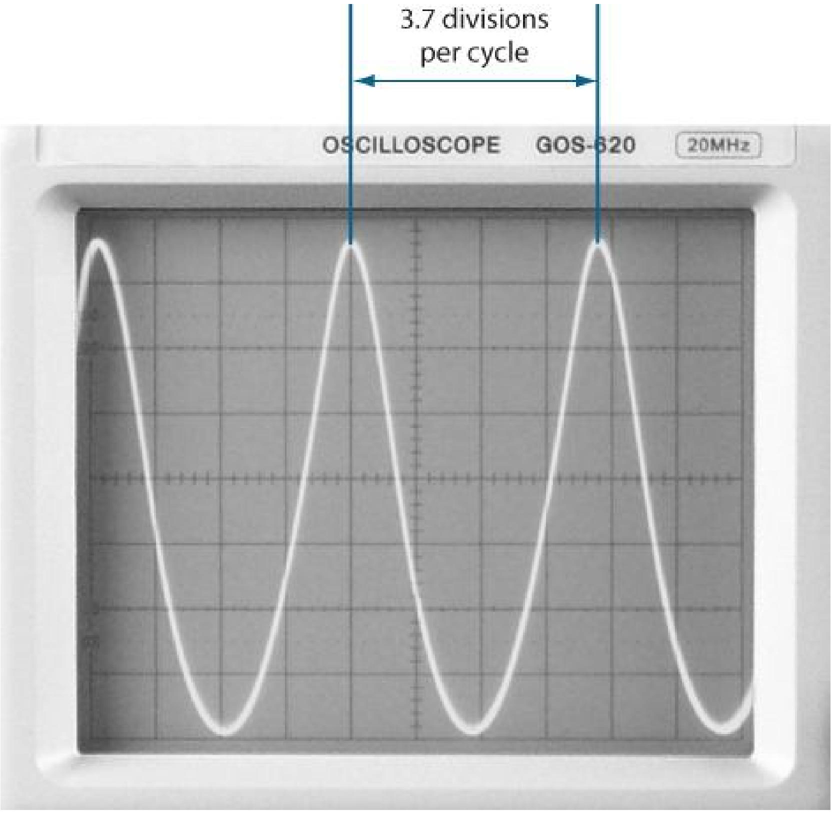
3.7 divisions
per cycle
OSCILLOSCOPE GOS-620
20MHz

As you measure the period, you may need to adjust the TIME/DIV, the horizontal POSITION, and the vertical POSITION controls on the oscilloscope. The controls shown in Figure 9.34 are adjusted to measure the period for the Hartley oscillator.

Your values should be close to those shown here:

Period = 74 μsec

Frequency = 13.5 kHz

This measured frequency is close to the calculated resonance frequency of 15.8 kHz.

Figure 9.34

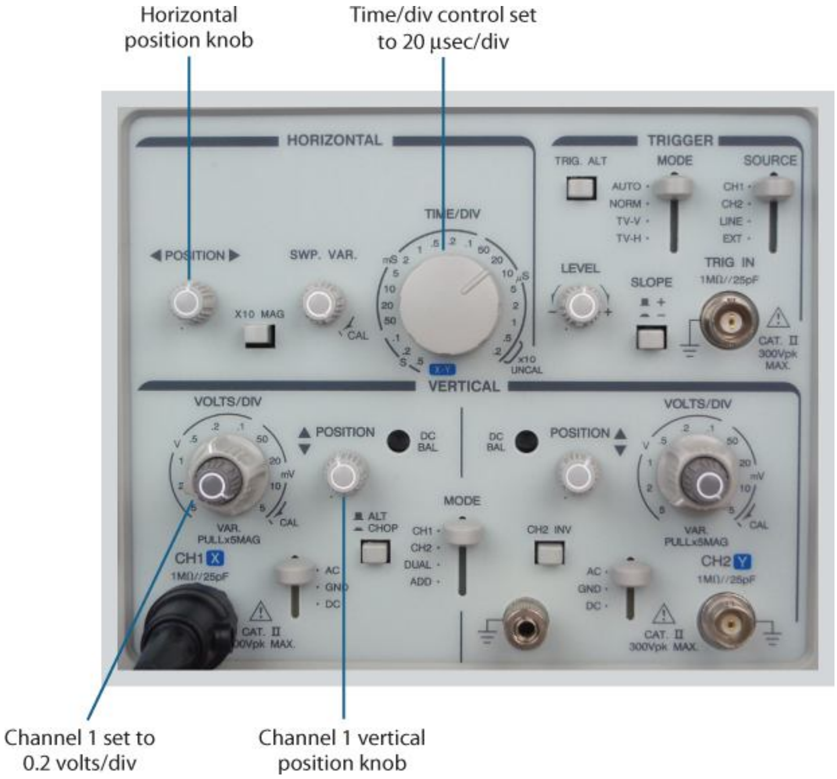
Horizontal position knob
Time/div control set to 20 μsec/div
HORIZONTAL
TRIGGER
VERTICAL
Channel 1 set to 0.2 volts/div
Channel 1 vertical position knob

The Armstrong Oscillator

The Armstrong oscillator shown in Figure 9.35 is somewhat more difficult to design and build. Here, the oscillations depend more on the extra winding on the coil than on any other factor.

Figure 9.35

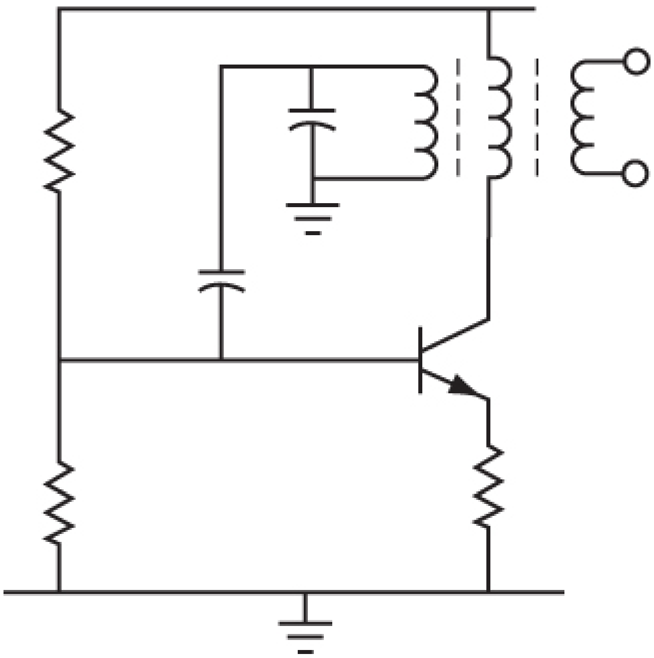

Because of the large variety of transformers and coils available, it is almost impossible to give you a simple procedure for designing an Armstrong oscillator. Instead, the manufacturer specifies the number of turns required on the coils, which guarantees that the oscillator will work in its most common operation, at high radio frequencies.

Because of the practical difficulties, the Armstrong oscillator and its variations are not explored any further.

Practical Oscillator Design

26 This section briefly covers some practical problems with oscillators.

Before you proceed, review the important points of this chapter by answering the following questions.

Questions

A. What three elements must an oscillator have present to work? _____

B. What determines the frequency of an oscillator's output signal?

C. What provides the feedback?

D. How many feedback methods for oscillators have been discussed?

E. What do you need to start the oscillations once the circuit has been built?

Answers

A. An amplifier, a resonant LC circuit (or some other frequency determining components), and feedback.

B. The frequency of the output signal is the same as the resonance frequency.

C. A voltage divider on the resonant circuit.

D. Three: the Colpitts, Hartley, and the Armstrong.

E. Nothing: The oscillations should start spontaneously if the component values in the circuit are correct.

The main practical problem with building oscillators is selecting the coil. For mass production, a manufacturer can specify and purchase the exact coil required. But in a lab or workshop (where you are building only a single circuit), it is often difficult or impossible to find the exact inductor specified in a circuit design. What usually happens is that you use the most readily available coil, and design the rest of the circuit around it. This presents three possible problems:

You may not know the exact value of the inductance.

The inductance value may not be the best for the wanted frequency range.

The coil may or may not have tap points or extra windings, and this may cause a change in the circuit design. For example, if there are no taps, then you cannot build a Hartley oscillator.

Because Colpitts is the easiest oscillator to make work in practice, and provides an easy way around some of the practical difficulties, you can focus on that oscillator.

You can use almost any coil when building a Colpitts oscillator, provided it is suitable for the frequency range you want. For example, a coil from the tuner section of a television set would not be suitable for a 1-kHz audio oscillator because its inductance value is outside the range best suited to a low-frequency audio circuit.

Simple Oscillator Design Procedure

27 Following is a simple step-by-step procedure for the design of a Colpitts oscillator. The Colpitts can work over a wide frequency range. (A Hartley can be designed using a similar set of steps.)

By following this procedure, you can design an oscillator that works in the majority of cases. There is a procedure you can use that guarantees that the oscillator will work, but it is far more complex.

Follow these steps:

1. Choose the frequency of the oscillator output signal.
2. Choose a suitable coil. This step presents the greatest practical difficulty. Some values of coil are often not available, so you must use whatever is readily available. Fortunately, you can use a wide range of inductance values and still obtain the desired resonance frequency by adjusting the value of the capacitor.
3. If you know the value of the inductance, calculate the capacitor value using this formula:

$$f_r = \frac{1}{2\pi\sqrt{LC}}$$

Use this value of capacitor for C 1 in the next steps.

4. If you don't know the inductance value, choose any value of capacitance and call this C 1 . This may produce a frequency considerably different from what you require. However, at this stage, the main thing is to get the circuit oscillating. You can adjust

values later.

5. Choose a capacitor C_2 that is between 3 and 10 times the value of C_1. Figure 9.36 shows the two capacitors and the coil connected in a parallel circuit, with the two capacitors acting as a voltage divider.

Figure 9.36

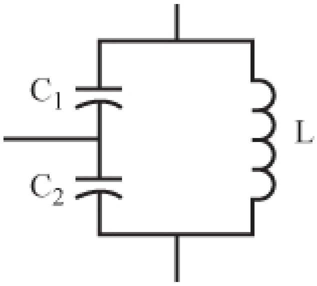

At this point, stop and make some assumptions. Suppose you need a frequency of 10 kHz and have a coil with a 16 mH inductance.

Questions

A. What approximate value of C_1 do you need?_____

B. What value of C_2 do you need?_____

Answers

A. C_1 = 0.016 μF

B. C_2 = 0.048 μF to 0.16 μF

28 Now, continue with the design procedure by following the next steps.

6. Design an amplifier with a common emitter gain of about 20. Choose a collector DC voltage that is about half the supply voltage. The main point to keep in mind here is that the collector resistor R_C should be about

one-tenth the value of the impedance of the LC circuit at the resonance frequency. This is often a difficult choice to make, especially if you don't know the coil value. Usually, you have to make an assumption, so R_C is an arbitrary choice.

7. Draw the circuit.

8. Calculate the value of C_C. Do this by making X_C 160 ohms at the desired frequency. This is another “rule of thumb” that happens to work, and you can justify it mathematically. Use the following formula:

$$C_C = \frac{1}{2\pi f_r X_C}$$

Question

Substitute the values given so far into the formula to calculate C_C._____

Answer

$$C_C = \frac{1}{2\pi \times 10\,\text{kHz} \times 160\,\Omega} = 0.1\,\mu\text{F}$$

29 Now, complete one last step.

9. Calculate the value of C_B. Again, choose a value so that X_C is 160 ohms at the desired frequency.

Question

What is the value of C_B?_____

Answer

$$C_B = 0.1\,\mu\text{F}$$

30 Continue the design procedure steps.

10. After you build an oscillator, apply power to the circuit and look at the output signal on

an oscilloscope. If the output signal is oscillating, check the frequency. If the frequency varies significantly from the desired frequency, then change C_1 until you get the wanted frequency. Change C_2 to keep the ratio of the capacitance values about the same as discussed in step 5. C_2 affects the output level.

11. If the circuit does not oscillate, go through the steps outlined in the troubleshooting checklist that follows.

Oscillator Troubleshooting Checklist

If an oscillator does not work, most often the trouble is with the feedback connections. A little experimenting (as outlined in steps 2 through 6 of the following checklist) should produce the right results. This is especially true when you use an unknown coil that may have several taps or windings. However, you should try each of the following steps if you have trouble.

1. Ensure that C_B, C_C, and C_E are all large enough to have a reactance value less than 160 ohms. Ensure that C_E is less than one-tenth of R_E.
2. Check the C_1/C_2 ratio. It should be between 3:1 and 10:1.
3. Swap out C_1 and C_2. They may be connected to the wrong end of the LC circuit.

4. Check that you made the feedback connection to and from the correct place.
5. Check both ends of the LC circuit to see that they are connected to the correct place.
6. Check the DC voltage level of the collector, base, and emitter.
7. Check the capacitor values of the LC circuit. If necessary, try some other values until the circuit oscillates.
8. If none of the previous actions produce oscillations, check to see if any of the components are defective. The coil may be opened or shorted. The capacitor may be shorted. The transistor may be dead, or its β may be too low. Check the circuit wiring carefully.

In most cases, one or more of these steps produces oscillations.

When an oscillator works, it may still have one or two main faults, including the following:

Distorted output waveform —This can happen when C_B, C_C, or C_E are not low enough in value, or when an output amplitude is too high.

Output level too low —When this happens, the sine wave is usually “clean” and “pure.” In a Colpitts oscillator, changing the ratio of C_1 and C_2 often helps raise the output level. If not, you can use another transistor as an amplifier after the oscillator, as discussed in Chapter 8, problem 21.

31 Now, work through a design example. Design an oscillator with an output frequency of 25 kHz using a coil with a value of 4 mH, and address each of the steps in problems 27–30 as described in these questions.

Questions

1. The value of f_r is given as 25 kHz.
2. L is given as 4 mH.
3. Use the formula to find C_1.

C_1 = _____

4. You do not need this step.
5. Choose C_2.

C_2 = _____

6. The procedure to design amplifiers is shown in Chapter 8.
7. The circuit is shown in Figure 9.37.
8. Find C_C.

C_C = _____

9. Find C_B.

C_B = _____

Figure 9.37

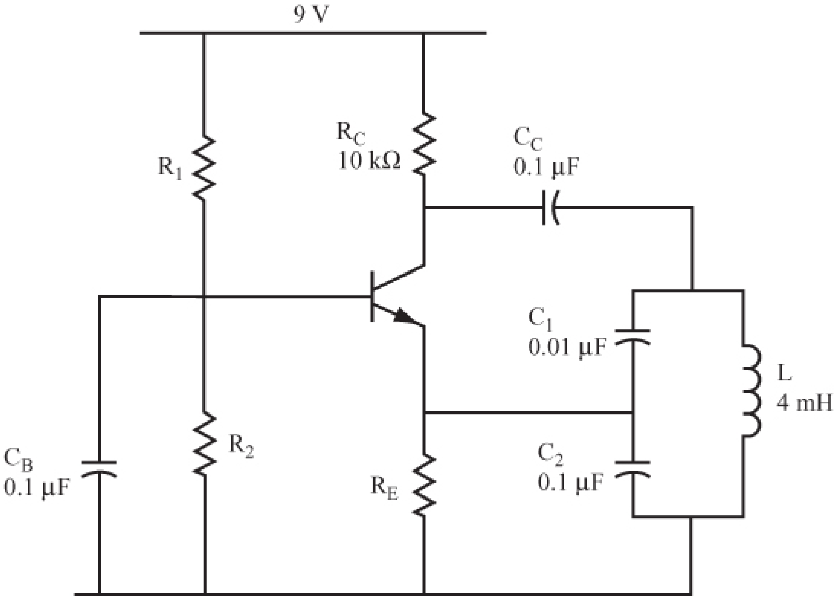

Answers

C_1 = 0.01 μF

C_2 = 0.1 μF

C_C = 0.047 μF (use 0.1 μF)

C_B = 0.047 μF (use 0.1 μF)

Steps 10–11 are the procedure you use to ensure that the oscillator works. If you built this circuit, go through steps 10–11. You don't need to do them if you didn't actually build the circuit.

32 Figure 9.37 shows the circuit designed in problem 31.

Measurements of the output signal of this oscillator confirm a frequency close to 25 kHz.

Question

Find the impedance of the LC circuit at resonance. Note that r (the DC resistance of the inductor) is 12 ohms._____

Answer

$$Z = \frac{L}{C \times r} = \frac{4 \times 10^{-3}}{0.01 \times 10^{-6} \times 12} = 33\,k\Omega \text{ (approximately)}$$

This is about three times the value used for R_C, instead of being 10 times the value of R_C, as suggested in step 6 of problem 28.

33 If you want, work through this second oscillator design example. Design an oscillator with an output frequency of 250 kHz using a coil with a value of 500 μH.

Questions

1. f_r = 250 kHz
2. L = 500 μH = 0.5 mH
3. Find C_1.

C_1 = _____

4. You do not need this step.
5. Find C_2.

C_2 = _____

6. Use the same amplifier as in the last example.
7. The circuit is shown in Figure 9.38.
8. Find C_C.

C_C = _____

9. Find C_B.

C_B = _____

Answer

C_1 = 0.0008 μF; therefore, choose a standard value of 0.001 μF.

C_2 = 0.0047 μF, which is a standard value.

C_B = C_C = 0.004 μF (minimum).

34 The circuit you designed in problem 33 is shown in Figure 9.38 .

Figure 9.38

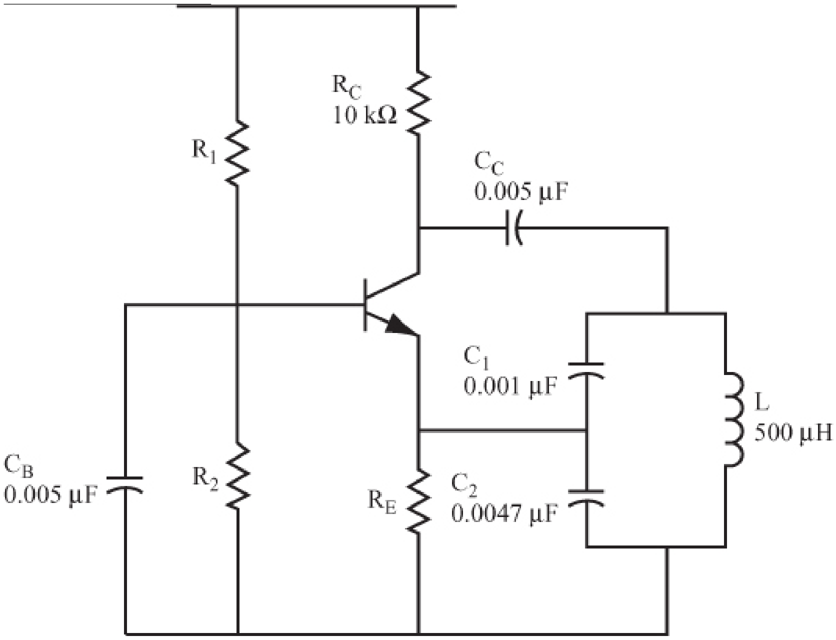
R$_C$
10 kΩ
R$_1$
C$_C$
0.005 μF
C$_1$
0.001 μF
L
500 μH
C$_B$
0.005 μF
R$_2$
R$_E$
C$_2$
0.0047 μF

Measurements of the output signal of this oscillator confirm a frequency close to 250 kHz.

Question

Find the impedance of the LC circuit at resonance. Note that r (the DC resistance of the inductor) is 20 ohms._____

Answer

Z = 30 kΩ

This is about 3 times the value of R C , rather than 10 times the value of R C , as suggested in step 6 of problem 28.

35 Figure 9.39 shows several other

oscillator circuits. Calculate the expected output frequency for each circuit and build as many as you want. Check the measured oscillator output frequency against the calculated values for each circuit you build.

Figure 9.39

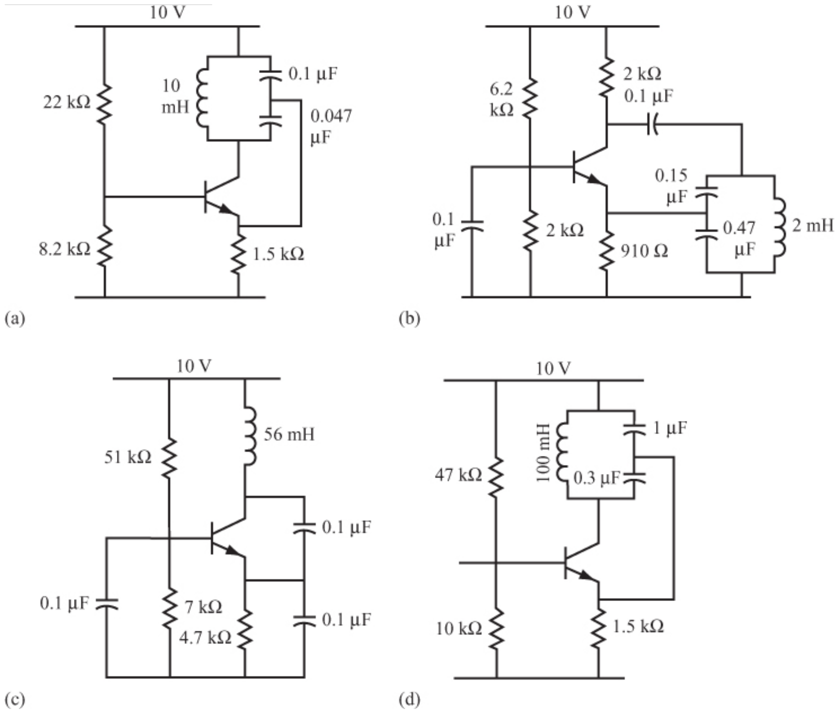
10 V
22 kΩ
10 mH
0.1 μF
0.047 μF
8.2 kΩ
1.5 kΩ
(a)
10 V
6.2 kΩ
2 kΩ
0.1 μF
0.15 μF
0.1 μF
2 kΩ
910 Ω
0.47 μF
2 mH
(b)
10 V
51 kΩ
56 mH
0.1 μF
0.1 μF
7 kΩ
4.7 kΩ
0.1 μF
(c)
10 V
47 kΩ
100 mH
1 μF
0.3 μF
10 kΩ
1.5 kΩ
(d)

Questions

What is the output frequency for each circuit?

A. f = _____

B. f = _____

C. f = _____

D. f = _____

Answers

A. 8.8 kHz

B. 10 kHz

C. 3 kHz
D. 1 kHz

Summary and Applications

This chapter covered the following topics related to oscillators:

- The main elements that make up an oscillator
- How to differentiate between positive and negative feedback
- The type of feedback that causes a circuit to oscillate
- Two methods to obtain feedback in an oscillator circuit
- How resonant LC circuits set the frequency of an oscillator

You also practiced designing a simple oscillator circuit to solidify your understanding of its elements and operation.

Self-Test

These questions test your understanding of the concepts and equations presented in this chapter. Use a separate sheet of paper for your diagrams or calculations. Compare your answers with the answers provided following the test.

1. What are the three sections necessary in an oscillator?_____
2. What is the difference between positive and negative feedback?_____

3. What type of feedback is required in an oscillator?_____

4. What is the formula for the frequency of an oscillator?_____

5. Draw the circuit for a Colpitts oscillator.

6. Draw the circuit for a Hartley oscillator.

7. Draw the circuit for an Armstrong oscillator.

8. Problems 27–30 give a design procedure for oscillators. How well do the circuits in problem 35 fulfill the criteria for that procedure? In other words, check the values of V_f, A_V (for a common emitter amplifier), C_1/C_2 ratio, R_C/Z ratio, and the frequency.

A. _____

B. _____

C. _____

D. _____

9. For the circuit shown in Figure 9.38, calculate the values of C_1, C_2, C_C, and C_B for an oscillator with an output frequency of 10 kHz using a 100 mH coil.

Answers to Self-Test

If your answers do not agree with those provided here, review the problems indicated in parentheses before you go on to Chapter 10.

1. An amplifier, feedback, and a resonant load. (problem 1)

2. Positive feedback is "in phase" with the input, and negative feedback is "out of phase" with the input. (problems 2–3)

3. Positive feedback. (problem 3)

4. $f_r = \frac{1}{2\pi\sqrt{LC}}$ (problem 11)

5. See Figure 9.15. (problem 20)

6. See Figure 9.27. (problem 25)

7. See Figure 9.35. (problem 25)

8A. $V_f = \frac{0.047}{0.147}$ (problems 27–30)

A_V cannot be calculated.

$C_1/C_2 = 0.047/0.1 = 0.47$

Z cannot be calculated because r is unknown.

$f_r = 8.8$ kHz (approximately).

8B. $V_f = \frac{0.15}{0.62}$

$A_V = 2.2$ (approximately).

$C_1/C_2 = 1/3$ (approximately).

Z cannot be calculated.

$f_r = 10$ kHz (approximately).

8C. $V_f = \frac{0.1}{0.2}$

A_v cannot be calculated.

$C_1/C_2 = 1$

Z cannot be calculated.

$f_r = 3$ kHz

8D. $V_f = \frac{0.3}{1}$

A_v cannot be calculated.

$C_1/C_2 = 0.3$

Z cannot be calculated.

$f_r = 1$ kHz (approximately)

9. $C_1 = 0.0033$ μF; $C_2 = 0.01$ μF; $C_B = C_C = 0.1$ μF (problems 26–30)

Chapter 10

The Transformer

Transformers are used to “transform” an AC voltage to a higher or lower level. When you charge your cellphone, you use a transformer to reduce the 120 volts supplied by the wall outlet to the 5 volts or so needed to charge your cellphone's battery. Most electrical devices that you plug into wall outlets use transformers to reduce power coming from an outlet to that required by the electrical components in the device.

You can also use transformers to increase voltage. For example, some of the equipment used to manufacture integrated circuits requires thousands of volts to operate. Transformers are used to increase the 240 volts supplied by the power company to the required voltage.

When you complete this chapter, you will be able to do the following:

Recognize a transformer in a circuit.
Explain and correctly apply the concepts of *turns ratio* and *impedance matching*.
Recognize two types of transformer.
Do simple calculations involving transformers.

Transformer Basics

1 Consider two coils placed close to each other, as shown in Figure 10.1. If you apply an AC voltage to the first (or primary) coil, the alternating current flowing through the coil creates a fluctuating magnetic field that surrounds the coil. As the strength and polarity of this magnetic field changes, it induces an alternating current and a corresponding AC voltage in the second (or secondary) coil. The AC signal induced in the secondary coil is at the same frequency as the AC signal applied to the primary coil.

Figure 10.1

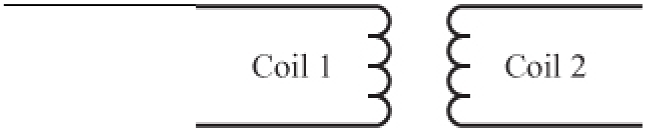

Both transformer coils are usually wound around a core made of a magnetic material such as iron or ferrite to increase the strength of the magnetic field.

Questions

A. When the two coils are wound around the same core, are they connected electrically? _____

B. What type of device consists of two wire coils wound around an iron or ferrite core? _____

C. If you apply an AC voltage to the terminals of the primary coil, what occurs in the secondary coil? _____

Answers

A. No.

B. A transformer.

C. An alternating current is induced in the secondary coil, which produces an AC voltage between the terminals of the secondary coil.

2 A transformer is used only with alternating currents. A fluctuating magnetic field (such as that generated by alternating current flowing through a primary coil) is required to induce current in a secondary coil. The stationary magnetic field generated by direct current flowing through a primary coil will not induce any current or voltage in a secondary coil.

When a sine wave signal is applied to a primary coil, you can observe a sine wave of the same frequency across the secondary coil, as shown in Figure 10.2 .

Figure 10.2

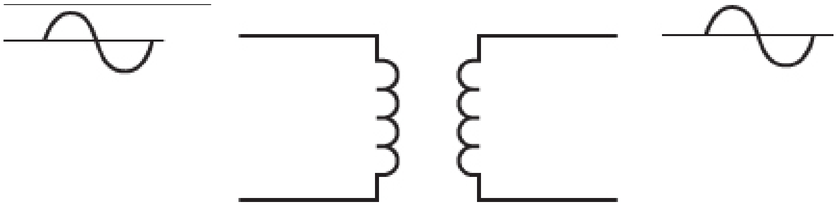

Questions

A. What will be the difference in frequency between a signal applied to a primary coil and the signal induced in a secondary coil? _____

B. What will be the voltage difference across a secondary coil if 10 volts DC is applied to the primary coil? _____

Answers

A. No difference. The frequencies will be the same.

B. Zero volts. When a DC voltage is applied to the primary coil, there is no voltage or current induced in the secondary coil. You can summarize this by saying that DC does not pass through a transformer.

3 You can compare the output waveform measured between the terminals of the secondary coil to the output waveform measured between the terminals of the primary coil. If the output goes positive when the input goes positive, as shown in Figure 10.3, then they are said to be *in phase*.

Figure 10.3

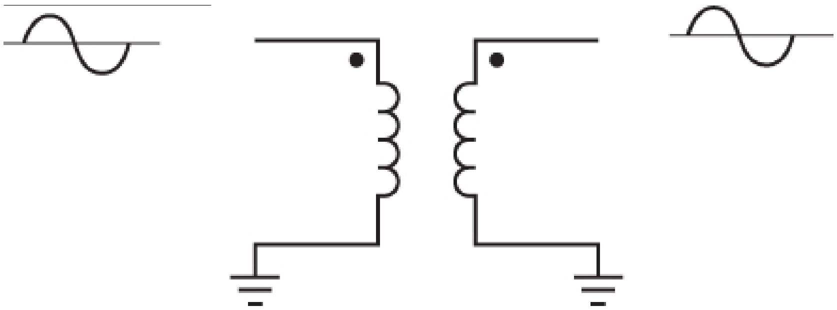

The dots on the coils in Figure 10.3 indicate the corresponding end of each coil. If one coil is reversed, then the output will be inverted from the input. The output is said to be *out of phase* with the input, and a dot is placed at the opposite end of the coil.

Question

In Figure 10.4, the output sine wave is out of phase with the input sine wave. Place a dot in the correct location in the secondary coil to show that it is out of phase.

Figure 10.4

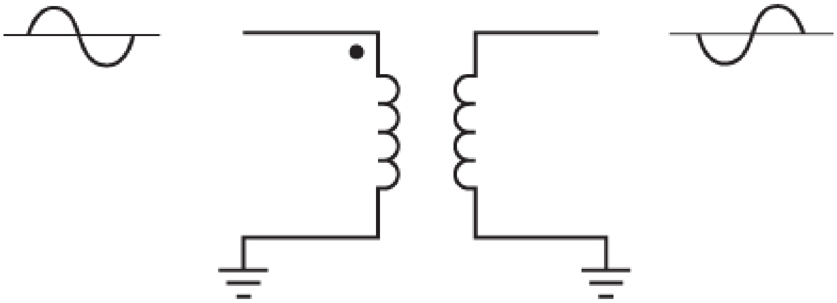

Answer

The dot should be at the lower end of the right coil.

4 The transformer shown in the right side of Figure 10.5 has three terminals. The additional terminal, in the middle of the coil, is called a *center tap*.

Figure 10.5

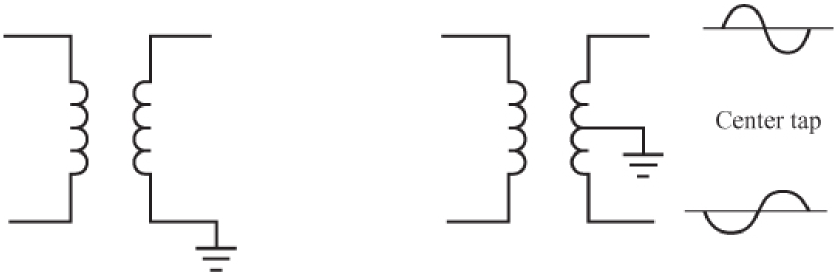

Question

What is the difference between the two output waveforms shown for the transformer on the right side of Figure 10.5? _____

Answer

The two waveforms are 180 degrees out of phase. That is, the positive peak of the upper output occurs at the same time as the negative peak of the lower waveform.

5 In a transformer, the output voltage from the secondary coil is directly proportional to the number of turns of wire in the secondary coil. If you increase the number of turns of wire in the secondary coil, a larger output voltage is induced across the secondary coil. If you decrease the number of turns of wire in the secondary coil, a smaller output voltage is induced across the secondary coil.

Question

How does increasing the number of turns of wire in a secondary coil affect the output voltage across the secondary coil? _____

Answer

It increases the output voltage across the secondary coil.

6 Figure 10.6 shows the number of turns in the primary and secondary coils as N p and N s .

Figure 10.6

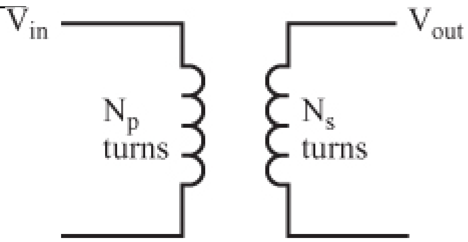

Question

The ratio of the input to output voltage is the same as the ratio of the number of turns in the primary coil to the number of turns in the secondary coil. Write a simple formula to express this. _____

Answer

$$\frac{V_{in}}{V_{out}} = \frac{N_p}{N_s}$$

Note The ratio of primary turns to secondary turns is called the *turns ratio (TR)* :

$$TR = \frac{N_p}{N_s} = \frac{V_{in}}{V_{out}}$$

7 Use the formula from problem 6 to answer the following question.

Question

Calculate the output voltage of a transformer with a 2 to 1 (2:1) turns ratio when you apply a 10 V pp sine wave to the primary coil. _____

Answer

$$\frac{V_{in}}{V_{out}} = \frac{N_p}{N_s} = TR$$

$$V_{out} = V_{in}\frac{N_s}{N_p} = V_{in} \times \frac{1}{TR}$$

$$V_{out} = V_{in} \times \frac{1}{TR} = 10 \times \frac{1}{2} = 5\ V_{pp}$$

8 Use the input voltage and turns ratio for a transformer to answer the following questions.

Questions

Calculate V out in the following:

A. V in = 20 V pp , turns ratio = 5:1.

V out = _____

B. V in = 1 V pp , turns ratio = 1:10.

V out = _____

C. V in = 100 V rms . Find V out when the primary and secondary coil have an equal number of turns.

V out = _____

Answers

A. 4 V pp (This is a *step-down transformer.*)

B. 10 V pp (This is a *step-up transformer* .)

C. 100 V rms (This is an *isolation transformer,* which is used to separate or isolate the voltage source from the load electrically.)

9 Almost all electronic equipment operated from 120 volts AC house current requires a transformer to convert the 120 volts AC to a more suitable, lower voltage. Figure 10.7 shows a transformer that steps down 120 volts AC to 28 volts AC.

Figure 10.7

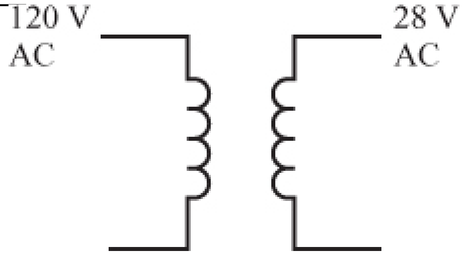

Question

Calculate the turns ratio for this transformer.

_____

Answer

$$TR = \frac{N_p}{N_s} = \frac{120}{28} = 4.3 : 1$$

10 Figure 10.8 shows an oscilloscope trace of the output waveform from the 28-volt secondary coil.

Figure 10.8

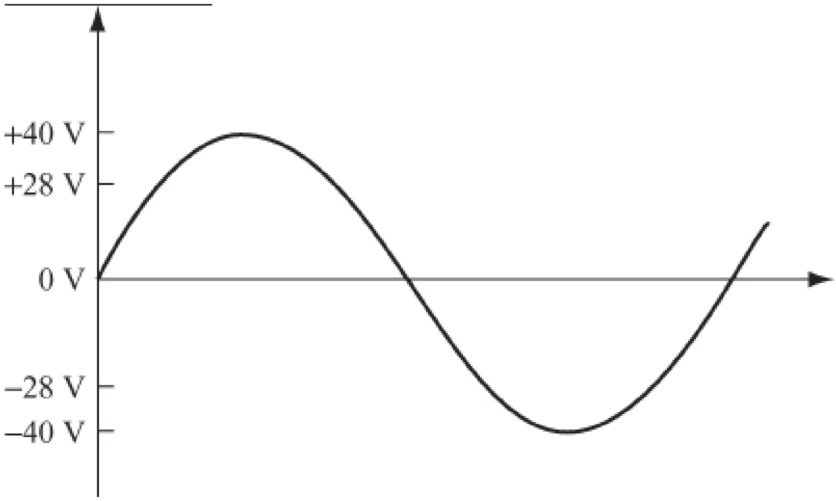

Questions

A. Is 28 volts a peak-to-peak or an rms value? _____

B. What is the peak-to-peak value of the 28 volts across the secondary coil? _____

Answers

A. rms

B. 2 × 1.414 × 28 = 79.184 volts

11 Like the 28-volt transformer output value, the 120-volt wall plug value is an rms measurement.

Question

What is the peak-to-peak value of the voltage from the wall plug? _____

Answer

Approximately 340 volts

12 The actual voltage measured across the secondary coil of a transformer depends upon where and how you make the measurement. Figure 10.9 illustrates different ways to measure voltage across a 20 V pp secondary coil that has a center tap.

Figure 10.9

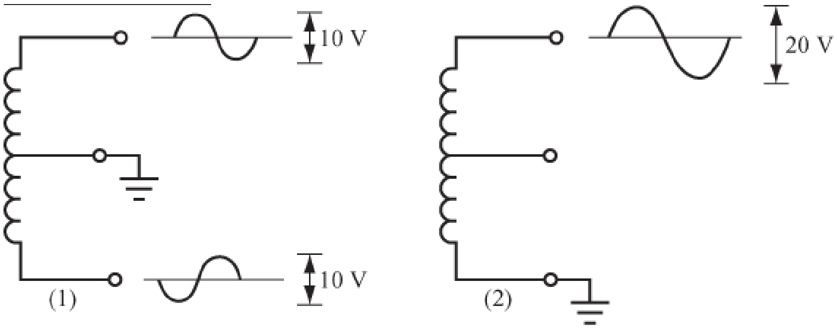

If the center tap is grounded as shown in diagram (1) of Figure 10.9 , then there is 10 V pp AC between each terminal and ground. You can see that the two output waveforms in diagram (1) are out of phase (180 degrees out of phase, in this case) by comparing the two sine waves shown next to the two terminals. If the bottom terminal is grounded as it is in diagram (2) of Figure 10.9 and the center tap is not used, then there is 20 V pp between the top terminal and ground.

Questions

A. Assume a center-tapped secondary coil is rated at 28 V rms referenced to the center tap. What is the rms voltage output when the center tap is grounded? _____

B. Assume the 28 V rms is the total output voltage across the entire secondary winding. What will be the output voltage between each end of the coil and the center tap? _____

C. Assume the output voltage of a center-tapped secondary coil is 15 V rms between each end of the coil and the center tap. What is the peak-to-peak output voltage when the center tap is not connected? _____

Answers

A. 28 V rms between each end of the coil and the center tap.

B. 14 V rms (one half of the total V out).

C. When the center tap is not connected, the output is 30 V rms . Therefore, V pp = 2 × 1.414 × 30 = 84.84 volts.

13 When the magnetic field induces an AC signal on the secondary coil, there is some loss of power. The percentage of power out of the transformer versus the input power is called the *efficiency* of the transformer. For the sake of this discussion, assume the transformer has an efficiency of 100 percent. Therefore, the output power of the secondary coil equals the power into the primary coil.

Power in = Power out (or P in = P out)

However, P = VI. Therefore, the following is true:

$$V_{in} I_{in} = V_{out} I_{out}$$

You can rearrange this to come up with the following formula:

$$\frac{I_{\text{out}}}{I_{\text{in}}} = \frac{V_{\text{in}}}{V_{\text{out}}} = \text{TR}$$

Questions

A. What would be the input current for a transformer if the input power was 12 watts at a voltage of 120 V rms ? _____

B. What would be the transformer's output voltage if the turns ratio was 5:1? _____

C. What would be the output current? _____

D. What would be the output power? _____

Answers

In AC power calculations, you must use the rms values of current and voltage.

A. $$I_{\text{in}} = \frac{P_{\text{in}}}{V_{\text{in}}} = \frac{12}{120} = 0.1\,A_{\text{rms}}$$

B. $$V_{\text{out}} = \frac{V_{\text{in}}}{\text{TR}} = \frac{120}{5} = 24\,V_{\text{rms}}$$

C. I out = I in (TR) = 0.1 × = = 0.5 A rms

D. P out = V in I out = 24 × 0.5 = 12 watts (same as the power in)

Inside the Transformer

In addition to the turns ratio discussed in this chapter, the design of transformers incorporates a few more aspects. The frequency at which a transformer is expected to operate has a big impact on the design and composition of the core. Transformers with iron cores work well at low frequencies, such as 50 or 60 Hz household AC, and even at audio frequencies. Transformers used at these frequencies are often made of laminated sheets of iron, instead of one solid piece of iron, which increases the electrical resistance of the core, which reduces the eddy current. The *eddy current* is an electrical current induced in the core by the fluctuating magnetic field that reduces the efficiency of transformers.

Reducing the eddy current is especially important at high frequencies. A transformer designed to work in the MHz range requires a core with higher electrical resistance to reduce the eddy current. Therefore, high-frequency transformers have cores made of different materials, such as iron oxides (called *ferrites*) or powdered iron. Transformers are rated for a particular frequency range, which you can find either in the supplier's catalog or in the manufacturer's data sheet.

The maximum power that can pass through a transformer is stated as a *VA rating* . VA stands for “volts x amperes” and is dependent upon factors such as the gauge of wire used. The VA rating makes it easy to calculate the maximum amperage when you know the voltage your circuit requires.

Transformers may also be rated by their input and output impedance at a particular frequency stated in the data sheet.

Transformers in Communications Circuits

14 In communications circuits, an input signal is often received via a long interconnecting wire (usually called a *line*) that normally has an impedance of 600 ohms. A typical example is a telephone line between two cities.

Question

Communications equipment works best when connected to a load that has the same impedance as the output of the equipment. What output impedance should communications equipment have? _____

Answer

600 ohms output impedance, to be connected to a 600-ohm line

15 Because most electronic equipment does not have a 600-ohm output impedance, a transformer is often used to connect such equipment to a line. Often, the transformer is built into the equipment for convenience. The transformer is used to "match" the equipment to the line, as shown in Figure 10.10 .

Figure 10.10

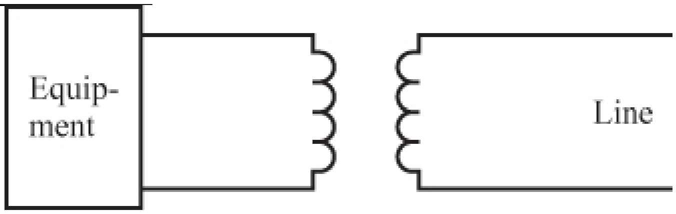

To work correctly, the output of the transformer secondary coil should have a 600-ohm impedance to match the line. The output impedance of the transformer (measured at the secondary winding) is governed by two things. One of these is the output impedance of the equipment.

Question

What would you expect the other governing factor to be? _____

Answer

The turns ratio of the transformer. (The DC resistance of each coil has no effect, and you can ignore it.)

16 Figure 10.11 shows a signal generator with an output impedance of Z G connected to the primary coil of a transformer. A load impedance of Z L is connected to the secondary coil.

Figure 10.11

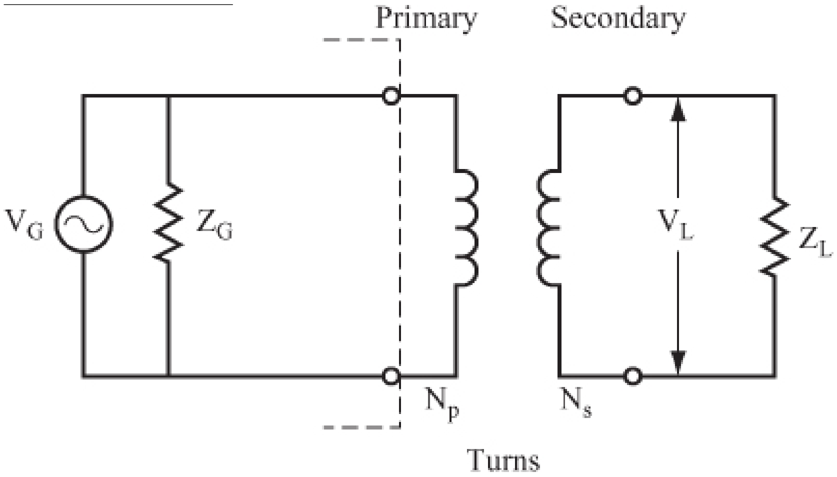

You know that $P_{in} = P_{out}$ and that $P = V^2/Z$. Therefore, you can write an equation equating the power of the generator to the power of the load in terms of V and Z, as shown here:

$$\frac{V_G{}^2}{Z_G} = \frac{V_L{}^2}{Z_L}$$

You can rearrange this equation to give the ratio of the voltages, as shown here:

$$\frac{Z_G}{Z_L} = {\frac{V_G}{V_L}}^{2}$$

And, because $V_G = V_{in}$, and $V_L = V_{out}$, and $V_{in}/V_{out} = N_p/N_s$, the following is true:

$$\frac{Z_G}{Z_L} = {\frac{V_{in}}{V_{out}}}^{2} = {\frac{N_p}{N_s}}^{2} = (TR)^2$$

Therefore, the ratio of the input impedance to the output impedance of a transformer is equal to the square of the turns ratio. As you can see in the following question A, you can determine the turns ratio for a transformer that matches impedances between a generator and a load. In this way, the generator “sees” an impedance equal to its own impedance, and the load also “sees” an impedance equal to its own impedance.

For the following problem, a generator has an output impedance of 10 kΩ and produces a 10 V_{pp} (3.53 V_{rms}) signal. It will be connected to a 600-ohm line.

Questions

A. To properly match the generator to the line, what turns ratio is required? _____

B. Find the output voltage across the load. _____

C. Find the load current and power. _____

Answers

A.
$$\mathrm{TR} = \sqrt{\frac{Z_G}{Z_L}} = \sqrt{\frac{10{,}000\,\Omega}{600\,\Omega}} = \frac{4.08}{1} \text{ or } 4.08:1$$

B. $V_L = \frac{V_G}{\mathrm{TR}} = \frac{10}{4.08} = 2.45\,V_{pp}$, which is 0.866V rms

C.
$$P_{in} = \frac{{V_G}^2}{Z_G} = \frac{(3.53)^2}{10{,}000} = 1.25\,\mathrm{mW}$$

Note For the power calculation, you must use the rms value of the voltage.

$$I_{in} = \frac{P_{in}}{V_{in}} = \frac{1.25\,\mathrm{mW}}{3.53\,V_{rms}} = 0.354\,\mathrm{mA}_{rms}, \text{ which is } 1\,\mathrm{mA}_{pp}$$

$I_L = I_{in}$ (TR) = 0.354 × 4.08 = 1.445 mA $_{rms}$, which is 4.08 mA $_{pp}$

$$P_L = \frac{{V_L}^2}{Z_L} = \frac{(0.866)^2}{600} = 1.25\,\mathrm{mW},$$

which is the same as the input power. This circuit is shown in Figure 10.12.

Figure 10.12

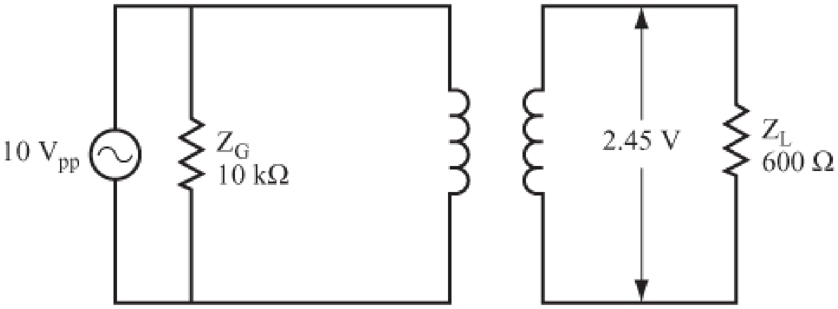

Note The generator now sees 10 kΩ when it looks toward the load, rather than the actual 600-ohm load. By the same token, the load now sees 600 ohms when it looks toward the source. This condition allows the optimum transfer of power to take place between the source and the load. In practice, however, the optimum condition as calculated here rarely exists. Because it may be impossible to obtain a transformer with a turns ratio of 4.08:1, you would have to select the closest available value, which might be a turns ratio of 4:1. The difference in the turns ratio affects the conditions at the load side, but only slightly.

17 In this problem, you use a transformer to match a generator to a load.

Questions

A. What turns ratio is required to match a generator that has a 2 kΩ output to a 600-ohm line? _____

B. If the generator produces 1 V pp , what is the voltage across the load? _____

Answers

A. TR = 1.83

B. V_L = 0.55 V pp

18 In this problem you use a transformer to match a generator to a 2 kΩ load.

Questions

A. What turns ratio is required to match a 2 kΩ load with a source that has an output impedance of = kΩ? _____

B. If the load requires a power of 20 mW, what should the source be? (First, find the voltage across the load.) _____

C. What are the primary and secondary currents and the power supplied by the source to the primary side of the transformer? _____

Answers

A. TR = 1.58

B.

$$V_L = \sqrt{P_L \times Z_L} = \sqrt{20\,\text{mW} \times 2\,\text{k}\Omega} = 6.32\,V_{\text{rms}};\text{and}$$

$$V_G = V_L \times TR = 6.32\,V_{\text{rms}} \times 1.58 = 10\,V_{\text{rms}}$$

C. I_L = 3.16 mA rms , I_p = 2 mA rms , P_{in} = 20 mW

Summary and Applications

In this chapter, you learned about the following topics related to transformers:

The principles that allow an AC signal to be induced in a secondary coil

How the AC voltage across the secondary coil can be stepped up or down depending upon the turns ratio of the transformer

The use of a center tap to produce various voltages from a transformer

The use of transformers to match impedances between a generator and a load

That transformers can cause the output signal to be inverted (out of phase) from the input signal

Self-Test

These questions test your understanding of the material in this chapter. Use a separate sheet of paper for your diagrams or calculations. Compare your answers with the answers provided following the test.

1. How is a transformer constructed? _____

2. What type of signal is used as an input to a transformer? _____

3. If a sine wave is fed into a transformer shown in Figure 10.13 , what does the output waveform look like? _____

Figure 10.13

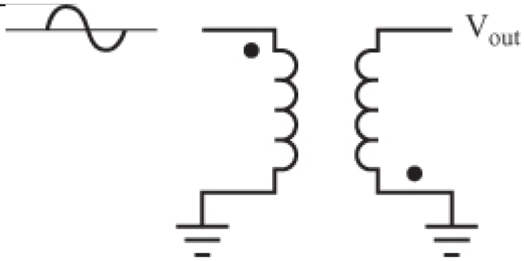

4. What is meant by the term *turns ratio* ? _____

5. If V in = 1 V pp and TR = 2, what is the output voltage?

V out _____

6. V in = 10 V pp and V out = 7 V pp , what is the turns ratio?

TR = _____

7. In the center-tapped secondary winding shown in Figure 10.14 , the voltage between points A and B may be expressed as V A-B = 28 V pp . What is the voltage between C and A? _____

Figure 10.14

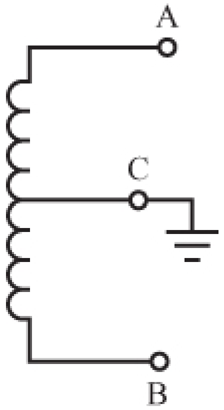

8. In the center-tapped secondary winding shown in Figure 10.14 , the voltage between points B and C is V B-C = = V rms . What is the peak-to-peak voltage between A and B?

_____

9. If I in = 0.5 A rms and I out = 2.0 A rms , what is the turns ratio? _____

10. Is the transformer in problem 9 a step-up or a step-down transformer? _____

11. If Z L = 600 ohms and Z G = 6 kΩ, find

the turns ratio.

TR = _____

12. If Z_L = 1 kΩ and the turns ratio is 10:1, what is the generator impedance?

Z_G = _____

Answers to Self-Test

If your answers do not agree with those given here, review the problems indicated in parentheses before you go to Chapter 11.

1. Two coils of wire wound around a magnetic core (such as iron or ferrite). (problem 1)
2. An AC voltage—DC does not work. (problem 2)
3. An inverted sine wave. (problem 3)
4. The ratio of the turns in the primary winding to the number of turns in the secondary coil. (problem 6)
5. V_{out} = 0.5 volts. (problem 7)
6. TR = 1.43:1. (problem 7)
7. $V_{C\text{-}A}$ = 14 V pp. (problem 12)
8. $V_{A\text{-}B}$ = 14.14 V pp. (problem 12)
9. TR = 4:1. (problem 13)
10. It is a step-down transformer. The voltage is lower (stepped down) in the secondary coil than in the primary coil if the current in the secondary coil is higher than the current in the primary coil. This maintains the same power on either side of the transformer. (problem 13)

11. TR = 3.2:1. (problem 16)

12. Z_G = 100 kΩ. (problem 16)

Chapter 11

Power Supply Circuits

A power supply is incorporated into many electronic devices. Power supplies convert the 120 volts AC from a wall plug to a DC voltage, providing power for all types of electronic circuits.

Power supply circuits are simple in principle, and those shown in this chapter have been around for many years. Because power supplies incorporate many of the features covered in this book, they make an excellent conclusion to your study of basic electronics.

Diodes are a major component in power supplies. Learning how AC signals are affected by diodes is fundamental to your understanding of how power supplies work. Therefore, this chapter begins with a brief discussion of diodes in AC circuits.

Throughout this chapter, diagrams show how AC signals are affected by diodes and other components in power supply circuits. If you have an oscilloscope, you can breadboard the circuits and observe these waveforms.

When you complete this chapter, you will be able to do the following:

Describe the function of diodes in AC circuits.

Identify at least two ways to rectify an AC signal.

Draw the output waveforms from rectifier and smoothing circuits.

Calculate the output voltage from a power supply circuit.

Determine the appropriate component values for a power supply circuit.

Diodes in AC Circuits Produce Pulsating DC

1 You can use diodes for several purposes in AC circuits, where their characteristic of conducting in only one direction is useful.

Questions

Assume that you apply + 20 volts DC at point A of the circuit, as shown in Figure 11.1 .

A. What is the output voltage at point B? _____

B. Suppose that you now apply + 10 volts DC at point B. What is the voltage at point A? _____

Figure 11.1

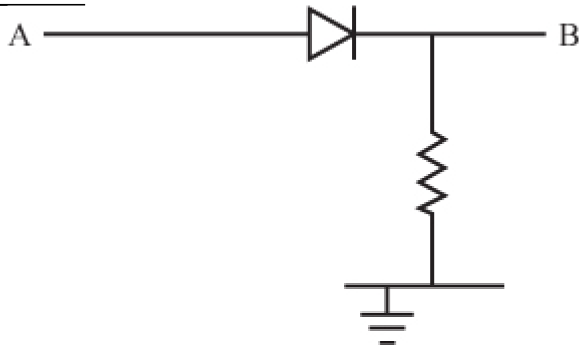

Answers

A. 20 volts DC (Ignore, for now, the voltage drop of 0.7 volt across the diode.)

B. 0 volts (The diode is reverse-biased.)

2 Figure 11.2 shows the circuit in Figure 11.1 with a 20 V pp AC input signal centered at + 20 volts DC.

Figure 11.2

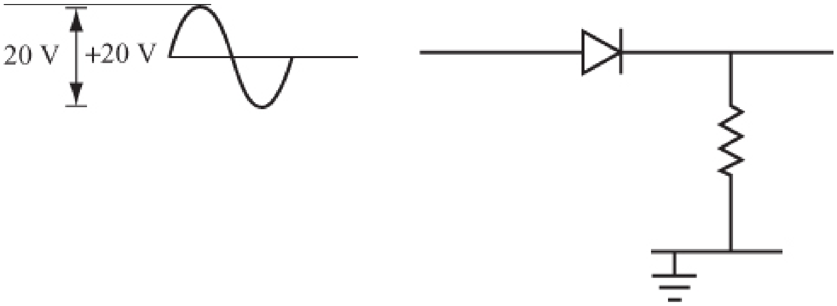

Questions

A. What are the positive and negative peak voltages of the input signal? _____

B. What is the output waveform of this circuit? _____

Answers

A. Positive peak voltage is 20 volts + 10 volts = 30 volts. Negative peak voltage is 20 volts − 10 volts = 10 volts.

B. The diode is always forward-biased, so it always conducts. Thus, the output waveform is exactly the same as the input waveform.

3 Figure 11.3 shows a circuit with 20 V pp AC input signal centered at 0 volts DC.

Figure 11.3

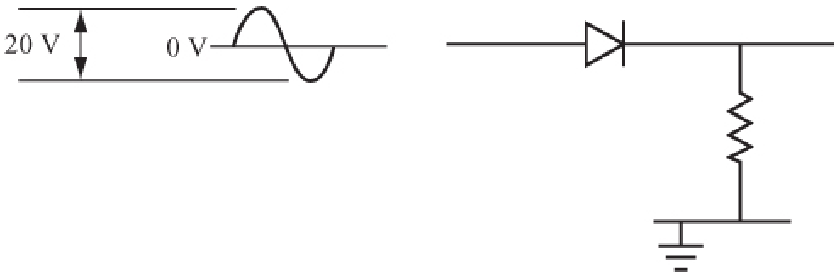

Questions

A. What are the positive and negative peak voltages of the input signal? _____

B. For the positive half wave of the input, draw the output waveform on the blank graph provided in Figure 11.4 .

Figure 11.4

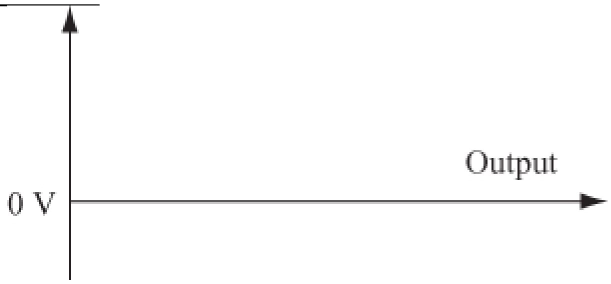

Answers

A. Positive peak voltage is +10 volts. Negative peak voltage is –10 volts.

B. See Figure 11.5 .

Figure 11.5

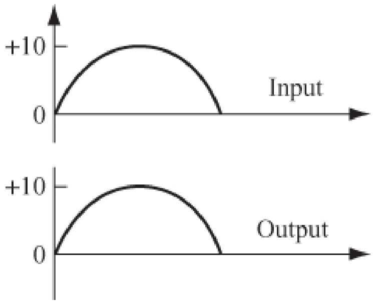

4 When the input is negative, the diode in the circuit shown in Figure 11.3 is reverse-biased. Therefore, the output voltage remains at 0 volts.

Question

Figure 11.6 shows the input waveform for the circuit shown in Figure 11.3 . Draw the output waveform on the blank graph provided in Figure 11.6 .

Figure 11.6

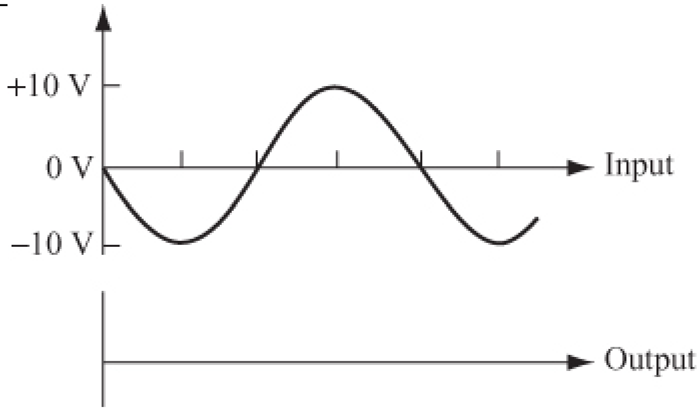

Answer

See Figure 11.7 .

Figure 11.7

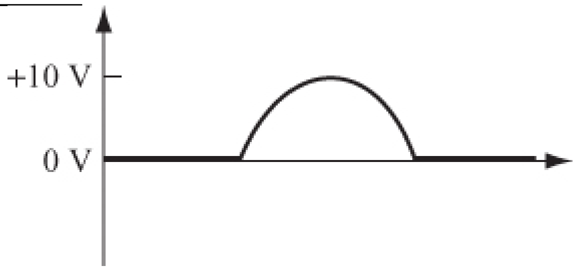

5 Figure 11.7 shows the output waveform of the circuit shown in Figure 11.3 , for one complete cycle of the input waveform.

Question

Now, draw the output waveform for three complete cycles of the input waveform shown in Figure 11.6. Use a separate sheet of paper for your drawing.

Answer

See Figure 11.8.

Figure 11.8

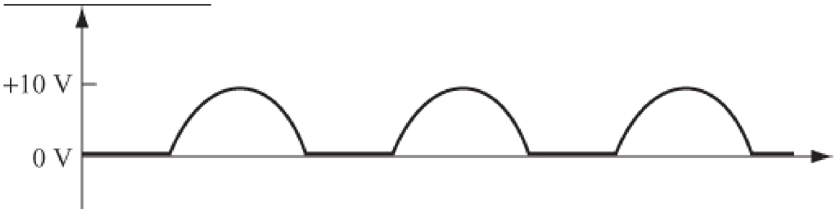

6 When the diode is connected in the opposite direction, it is forward-biased and, therefore, conducts current when the input signal is negative. In this case, the diode is reverse-biased when the input signal is positive. Therefore, the output waveform is inverted from the output waveform shown in Figure 11.8.

Question

On a separate sheet of paper, draw the output waveform for three input cycles, assuming that the diode is connected in the opposite direction from the diode shown in Figure 11.3.

Answer

See Figure 11.9.

Figure 11.9

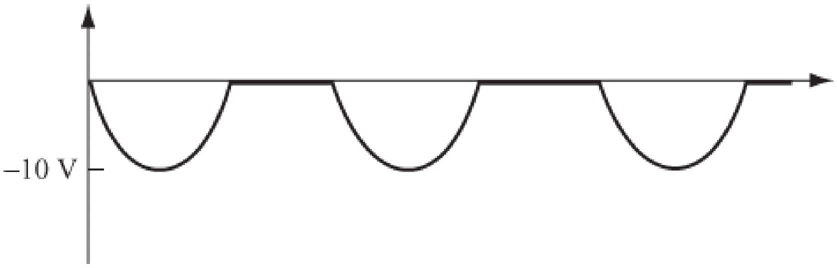

7 Figure 11.10 shows a circuit with a 20 V pp AC input signal centered at –20 volts DC.

Figure 11.10

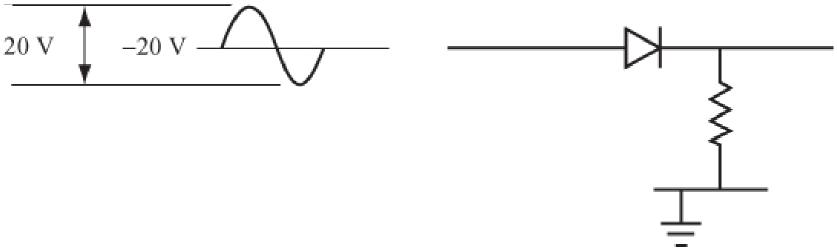

Questions

A. When is the diode forward-biased? _____

B. What is the output voltage? _____

Answers

A. Never because the voltage that results from adding the AC and DC signals ranges from – 10 volts to – 30 volts. Therefore, the diode is always reverse-biased.

B. A constant 0 volts.

8 As you have seen, a diode passes either the positive or negative portion of an AC voltage waveform, depending on how you connect it in a circuit. Therefore, the AC input signal is converted to a pulsed DC output signal, a process called *rectification* . A circuit that converts either the positive or negative portion of an AC voltage waveform to a pulsed DC output signal is called a *half-wave rectifier* .

Question

Refer to the output waveforms shown in Figures 11.8 and 11.9 . Do these waveforms represent positive DC voltage pulses or negative DC voltage pulses? _____

Answer

The waveform in Figure 11.8 represents positive pulses of DC voltage. The waveform in Figure 11.9 represents negative pulses of DC voltage.

9 The circuit shown in Figure 11.11 shows a diode connected to the secondary coil of a transformer.

Figure 11.11

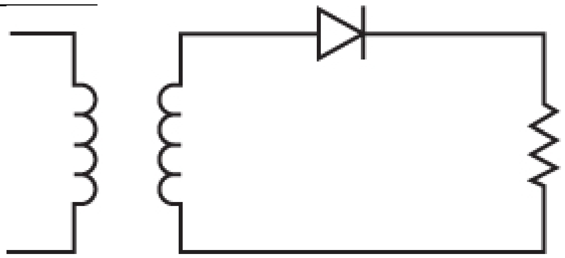

Questions

A. How does the diode affect the AC signal? _____

B. Draw the waveform of the voltage across the load for the circuit shown in Figure 11.11 if the secondary coil of the transformer has a 30 V pp AC output signal centered at 0 volts DC. Use a separate sheet of paper for your drawing.

Answers

A. The AC signal is rectified.

B. See Figure 11.12. This type of circuit (called a *half-wave rectifier*) produces an output waveform containing either the positive or negative portion of the input waveform.

Figure 11.12

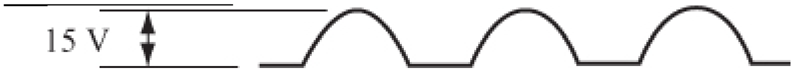

10 Figure 11.13 shows the waveforms at each end of a center-tap transformer secondary coil. Diode D 1 rectifies the waveform shown at point A, and diode D 2 rectifies the waveform shown at point B.

Figure 11.13

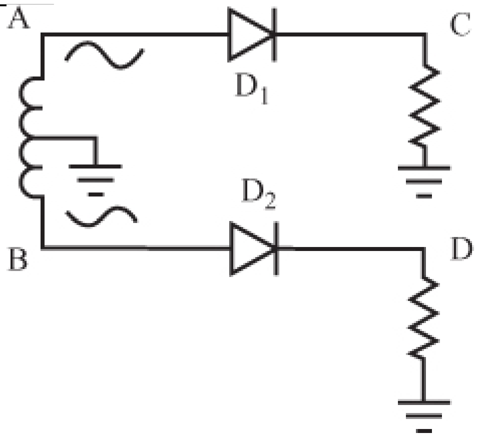

Questions

A. Which diode conducts during the first half of the cycle? _____

B. Which diode conducts during the second half of the cycle? _____

C. Draw the input waveforms (points A and

B), and underneath draw each output waveform (points C and D). Use a separate sheet of paper for your drawing.

Answers

A. During the first half of the cycle, D 1 is forward-biased and conducts current. D 2 is reverse-biased and does not conduct current.

B. During the second half of the cycle, D 2 is forward-biased and conducts current. D 1 is reversed-biased and does not conduct current.

C. See Figure 11.14 .

Figure 11.14

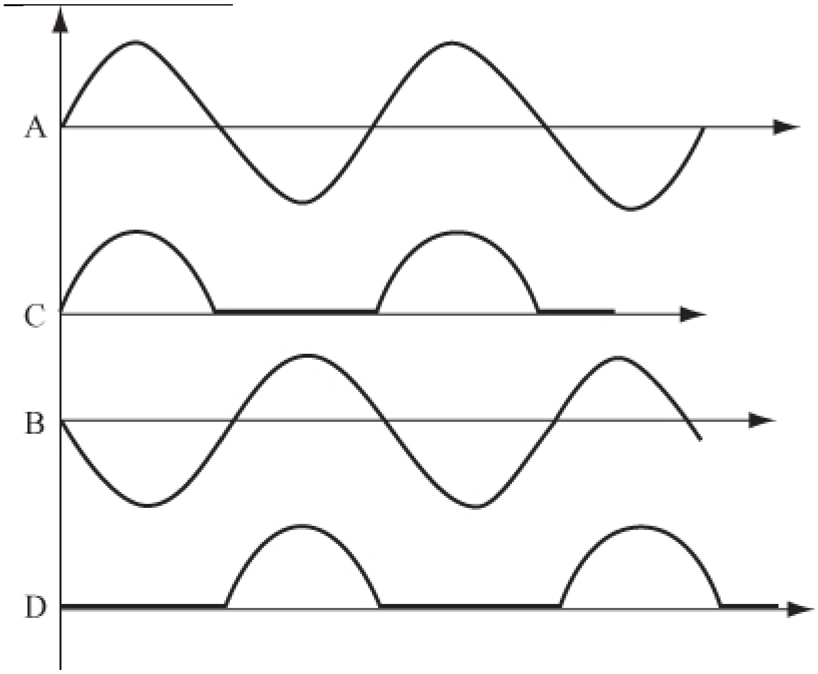

11 Figure 11.15 shows a circuit in which diodes connected to the ends of a center-tap transformer are connected to ground through a single resistor. The output voltage waveforms from both diodes are therefore applied across one load resistor. This type of circuit is called a *full-wave rectifier*.

Figure 11.15

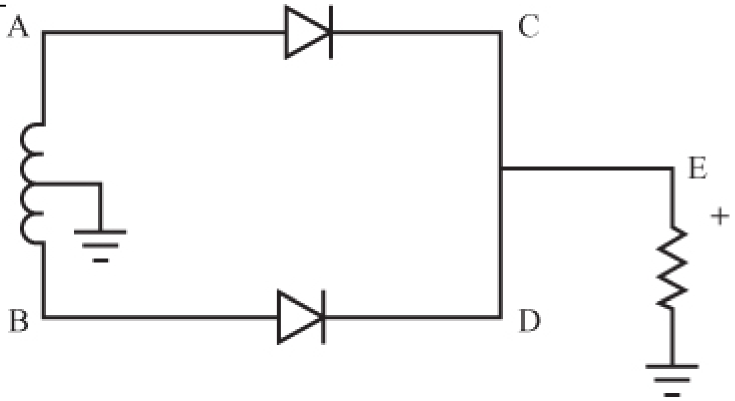

Question

On a separate sheet of paper, draw the waveform representing the voltage at point E in the circuit, as shown in Figure 11.15. (This waveform is a combination of the waveforms at points C and D shown in Figure 11.14.)

Answer

See Figure 11.16.

Figure 11.16

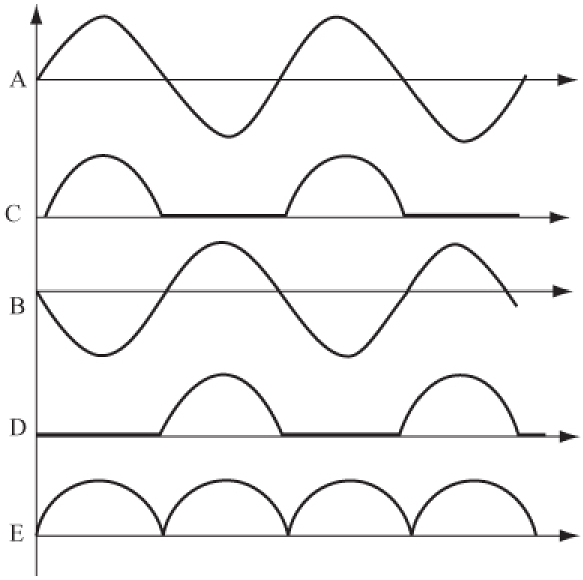

This is called full-wave rectification.

12 Full-wave rectification of AC allows a much "smoother" conversion of AC to DC than half-wave rectification.

Figure 11.17 shows a full-wave rectifier circuit that uses a transformer with a two-terminal secondary coil, rather than a center-tapped secondary coil. This type of circuit is called a *bridge rectifier*.

Figure 11.17

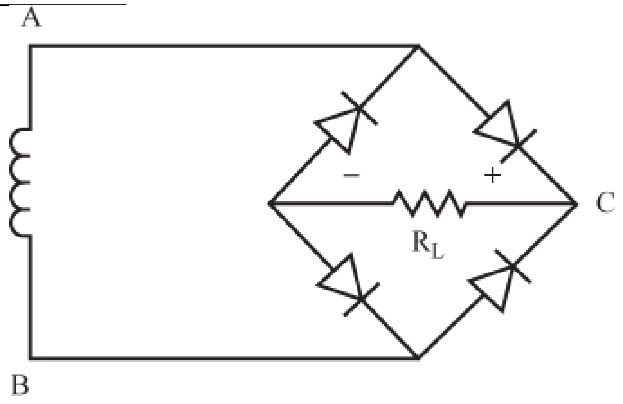

Question

How does this circuit differ from the circuit shown in Figure 11.15 ? _____

Answer

This circuit has no center tap on the secondary coil, and it uses four diodes.

13 Figure 11.18 shows the direction of current flow when the voltage at point A is positive.

Figure 11.18

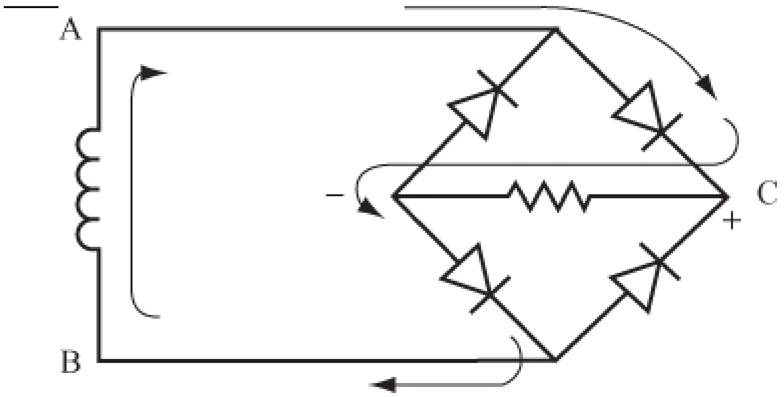

Figure 11.19 shows the direction of current flow when the voltage at point B is positive.

Figure 11.19

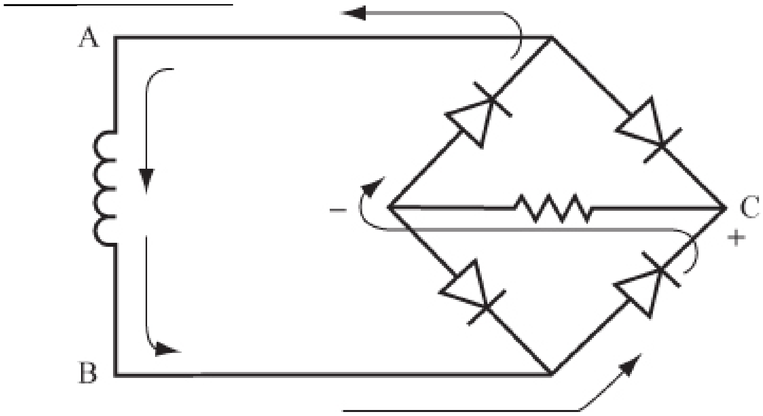

Notice that the direction of current through the load resistor is the same in both cases.

Questions

A. Through how many diodes does the current travel in each conduction path? _____

B. Draw the voltage waveform at point C. Use a separate sheet of paper for your drawing.

Answers

A. Two diodes in each case.

B. See Figure 11.20 .

Figure 11.20

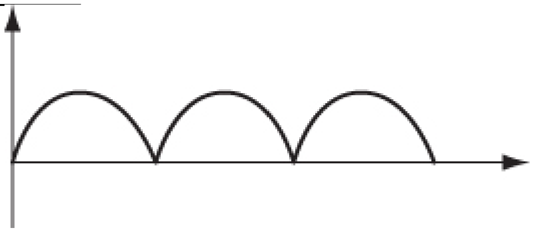

Project 11.1: The Full-Wave Rectifier

Objective

The objective of this project is to compare the outputs of the two types of full-wave rectifiers.

General Instructions

You set up two circuits. One of the circuits is a full-wave rectifier containing a center-tapped transformer and two diodes. The other circuit is a bridge rectifier containing a transformer and four diodes. After each circuit is set up, you apply a 20 V pp signal to the primary side of the transformer. Then you use your oscilloscope to look at the waveform across the load resistor, and measure the peak voltage of each waveform.

Parts List

You need the following equipment and supplies:

- Six 1N4001 diodes.
- Two 10 kΩ, 0.25-watt resistors.
- One audio transformer with impedance of both the primary and secondary coil rated at 600 Ω (with equal impedance for the primary and secondary coils, the turns ratio will be 1:1), and a center-tapped secondary coil. You can use a transformer with a center tap on both the primary and secondary coil; just don't connect a center tap that's not called for in the schematic.
- Two breadboards.
- One function generator.
- One oscilloscope.

Step-by-Step Instructions

Set up Circuit # 1, the full-wave rectifier circuit shown in Figure 11.21 . If you have some experience in building circuits, this schematic (along with the previous parts list) should provide all the information you need to build the circuit. If you need a bit more help building the circuit, look at the photos of the completed circuit in the “Expected Results” section.

Figure 11.21

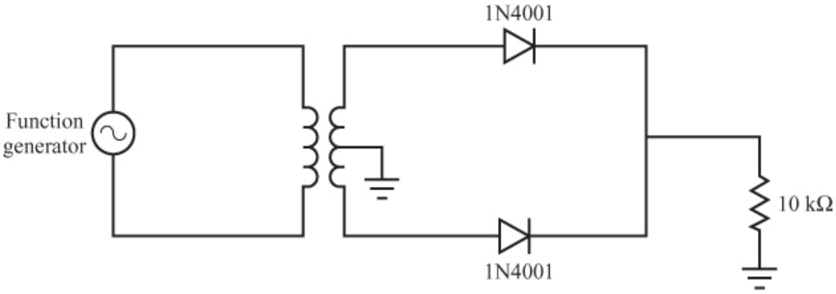

Carefully check your circuit against the diagram.

After you check your circuit, follow these steps and record your measurements in the blank table following the steps.

1. Connect the oscilloscope probe for Channel 1 to a jumper wire connected to the end of the resistor nearest the diodes. Connect the ground clip to a jumper wire attached to the ground bus.
2. Connect the oscilloscope probe for Channel 2 to a jumper wire that is connected to one end of the primary coil. This should be the same end to which you've connected the red lead from the function generator. Connect the ground clip for Channel 2 to a jumper wire that is connected to the other end of the primary coil. This should be the end to which you've connected the black lead from the function generator.
3. Set the function generator to generate a 1-kHz sine wave with 20 V_{pp}.
4. Measure and record V_p for the signal across the resistor.

5. Set up Circuit # 2; the bridge rectifier circuit shown in Figure 11.22 . Use the same transformer you used in Circuit #1. You do not connect the center tap on the secondary coil in this circuit.

6. Connect the oscilloscope probe for Channel 1 to a jumper wire connected to one end of the resistor. Connect the ground clip for Channel 1 to a jumper wire connected to the other end of the resistor.

7. Repeat steps 2 through 4.

Circuit #	V pp (Primary Coil)	V p (Load Resistor)
1		
2		

Figure 11.22

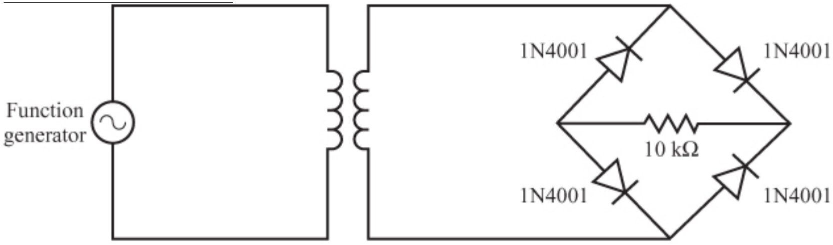

Expected Results

Figure 11.23 shows the breadboarded full-wave rectifier (Circuit # 1).

Figure 11.23

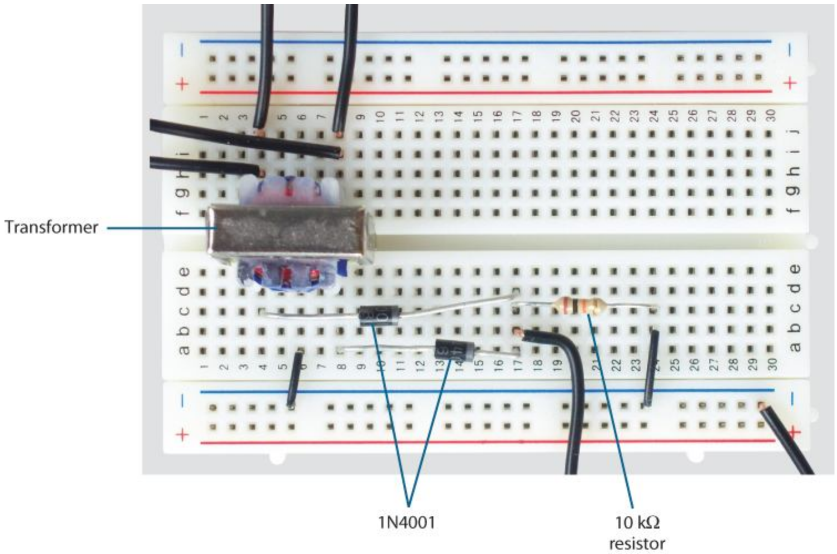
Transformer
1N4001
10 kΩ
resistor

Figure 11.24 shows a function generator and oscilloscope attached to Circuit #1.

Figure 11.24

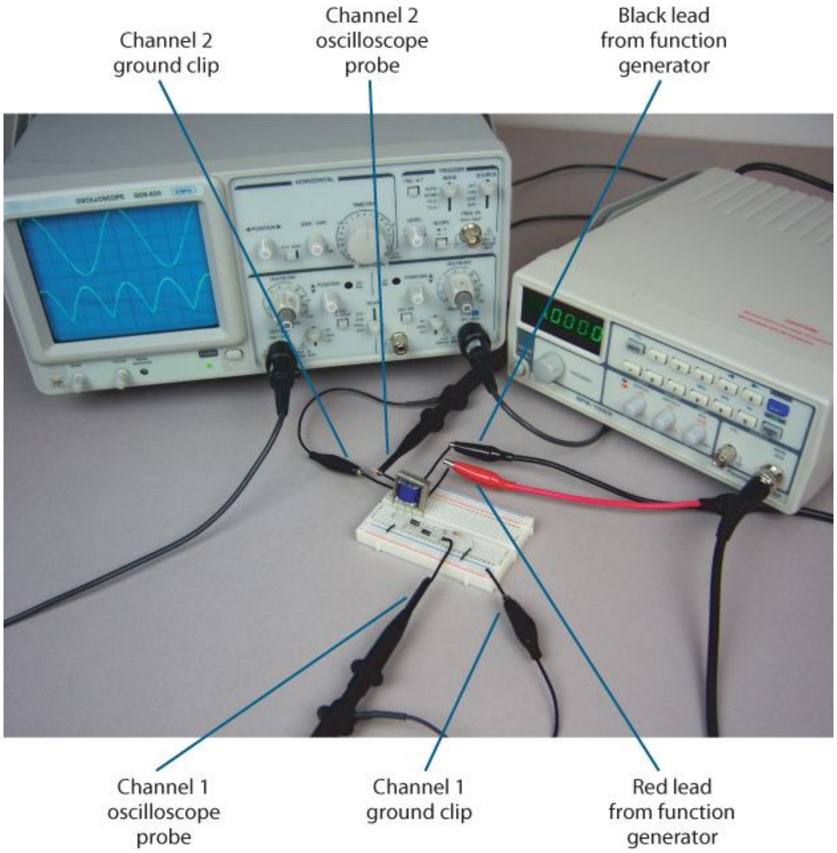
Channel 2 ground clip
Channel 2 oscilloscope probe
Black lead from function generator
Channel 1 oscilloscope probe
Channel 1 ground clip
Red lead from function generator

In Figure 11.25, the signal across the primary coil is represented by the upper waveform, and the signal across the resistor is represented by the lower waveform. Read the number of divisions for the peak-to-peak voltage of the upper waveform, and multiply it by the corresponding VOLTS/DIV setting to determine V pp for the signal across the primary coil. Read the number of divisions for the peak voltage of the lower waveform, and multiply it by the corresponding VOLTS/DIV setting to determine V_p for the signal across the resistor.

Figure 11.25

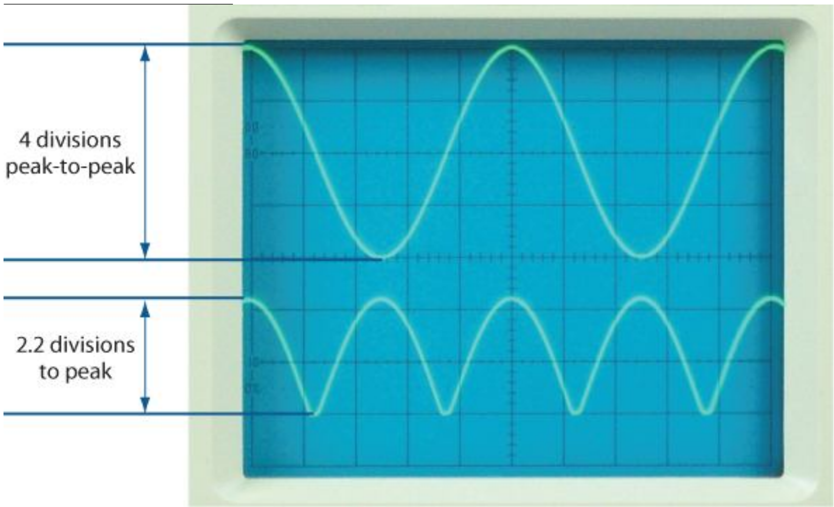

As you measure V_{pp} and V_p, you may need to adjust the TIME/DIV, VOLTS/DIV, and vertical POSITION controls on the oscilloscope. The controls shown in Figure 11.26 are adjusted to measure V_{pp} in the primary coil and V_p across the load resistor for Circuit # 1.

Figure 11.26

Time/div control set to 0.2 msec/div

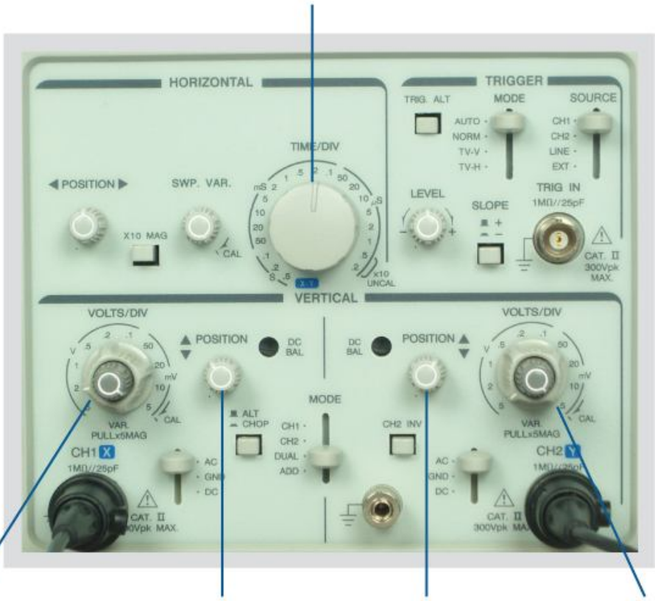
HORIZONTAL
TRIGGER
TRIG. ALT
MODE
SOURCE
AUTO
NORM
TV-V
TV-H
CH1
CH2
LINE
EXT
POSITION
SWP. VAR.
TIME/DIV
X10 MAG
CAL
LEVEL
SLOPE
TRIG IN
1MΩ//25pF
CAT. II
300Vpk
MAX.
x10
UNCAL
VERTICAL
VOLTS/DIV
POSITION
DC
BAL
MODE
ALT
CHOP
CH1
CH2
DUAL
ADD
CH2 INV
AC
GND
DC
VAR.
PULLx5MAG
CH1 X
CH2 Y
1MΩ//25pF
CAT. II
300Vpk MAX.

Channel 1 set to 2 volts/div

Channel 1 vertical position knob

Channel 2 vertical position knob

Channel 2 set to 5 volts/div

Figure 11.27 shows the breadboarded bridge rectifier (Circuit # 2).

Figure 11.27

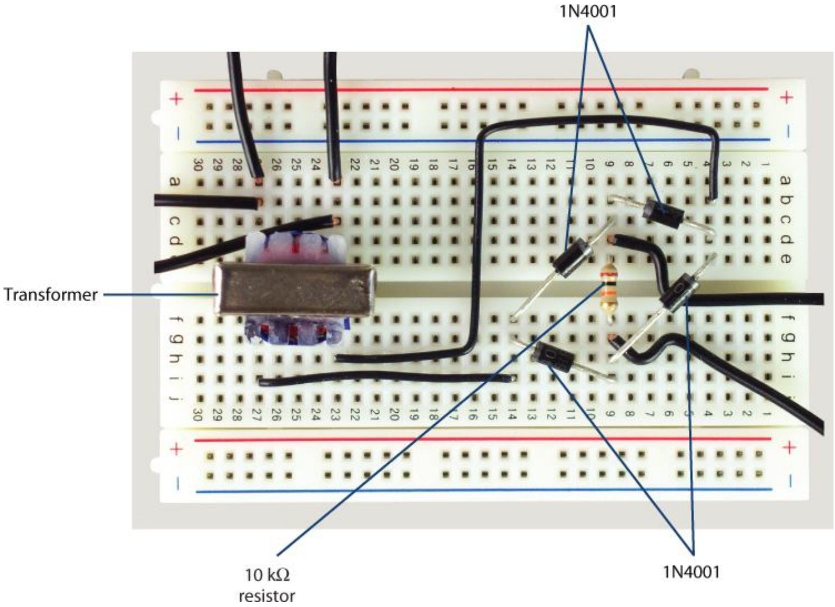
1N4001
Transformer
10 kΩ resistor
1N4001

Figure 11.28 shows a function generator and oscilloscope attached to Circuit #2.

Figure 11.28

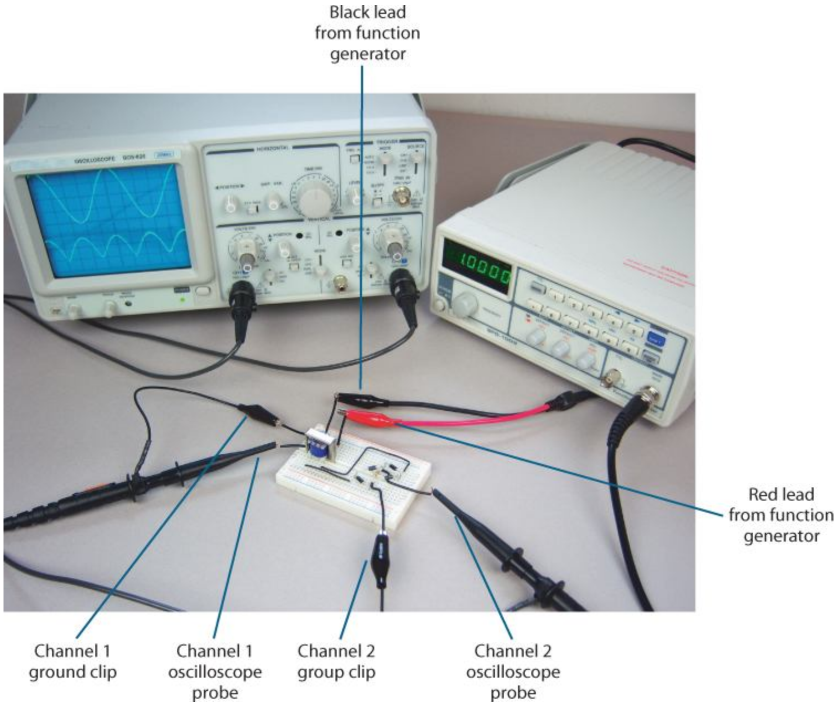
Black lead from function generator
Red lead from function generator
Channel 1 ground clip
Channel 1 oscilloscope probe
Channel 2 group clip
Channel 2 oscilloscope probe

In Figure 11.29 , the signal across the primary coil is represented by the upper waveform, and the signal across the resistor is represented by the lower waveform. Read the number of divisions for the peak-to-peak voltage of the upper waveform, and multiply it by the corresponding VOLTS/DIV setting to determine V pp for the signal across the primary coil. Read the number of divisions for the peak voltage of the lower waveform, and multiply it by the corresponding VOLTS/DIV setting to determine V p for the signal across the resistor.

Figure 11.29

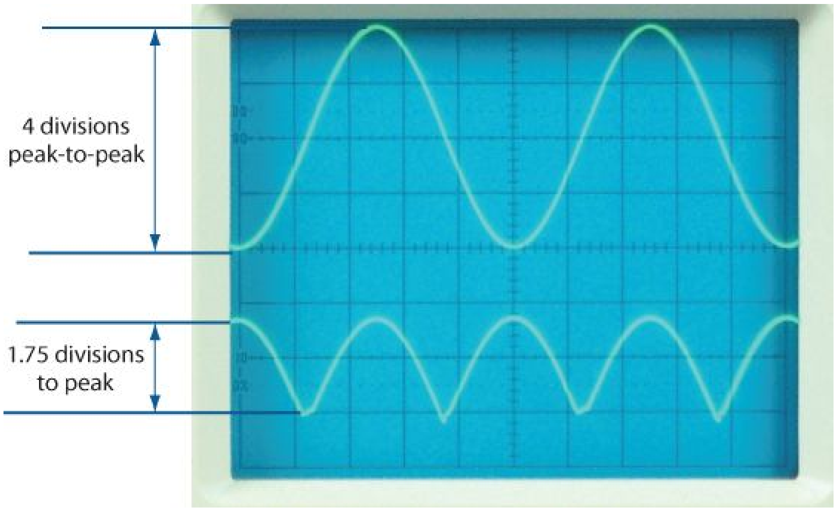
4 divisions peak-to-peak
1.75 divisions to peak

As you measure V pp and V p , you may need to adjust the TIME/DIV, the VOLTS/DIV, and the vertical POSITION controls on the oscilloscope. The controls shown in Figure 11.30 are adjusted to measure V pp in the primary coil and V p across the load resistor for Circuit # 2.

Figure 11.30

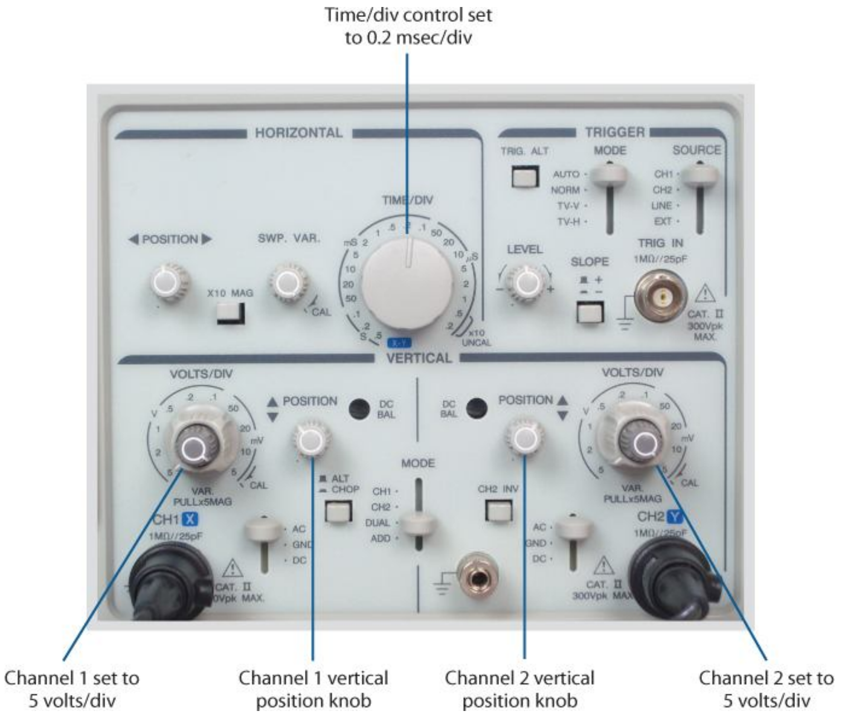
Time/div control set to 0.2 msec/div
HORIZONTAL
TRIGGER
VERTICAL
Channel 1 set to 5 volts/div
Channel 1 vertical position knob
Channel 2 vertical position knob
Channel 2 set to 5 volts/div

Your values should be close to those shown in the following table.

Circuit #	V pp (Primary Coil)	V p (Load Resistor)
1	20 volts	4.4 volts
2	20 volts	8.8 volts

V p across the load resistor for Circuit # 2 is twice the value of V p for Circuit #1. That's because Circuit #1 uses a center-tapped secondary coil that supplies a 10 V pp sine wave to the diodes, whereas the secondary coil in Circuit #2 supplies a 20 V pp sine wave. In each circuit, the 0.6 to 0.7 voltage drop that occurs as the signal passes through each diode causes V p across the load resistor to be slightly lower than half the V pp supplied by the secondary coil.

What this chapter has explored to this point is how AC is turned into pulsating DC. In fact, rectified AC is often called *pulsating DC*. The next step in your understanding of power supplies is to learn how you turn pulsating DC into level DC.

Level DC (Smoothing Pulsating DC)

14 A basic power supply circuit can be divided into four sections, as shown in Figure 11.31 .

Figure 11.31

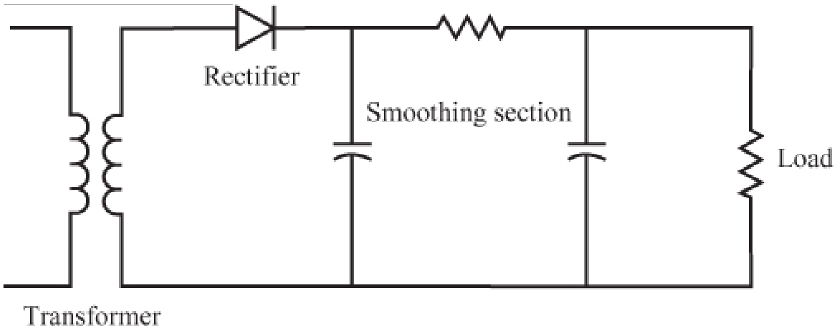

Questions

A. If you use a center-tap transformer in a power supply, how many diodes would you need to produce a full-wave rectified output? _____

B. Will the power supply circuit shown in Figure 11.31 result in full- or half-wave rectification? _____

C. What type of output will the rectifier section of the power supply circuit shown in Figure 11.31 produce? _____

Answers

A. Two

B. Half-wave

C. Pulsating DC

15 The function of the smoothing section of a power supply circuit is to take the pulsating DC (PDC) and convert it to a “pure” DC with as little AC “ripple” as possible. The smoothed DC voltage, shown in the illustration on the right in Figure 11.32, is then applied to the load.

Figure 11.32

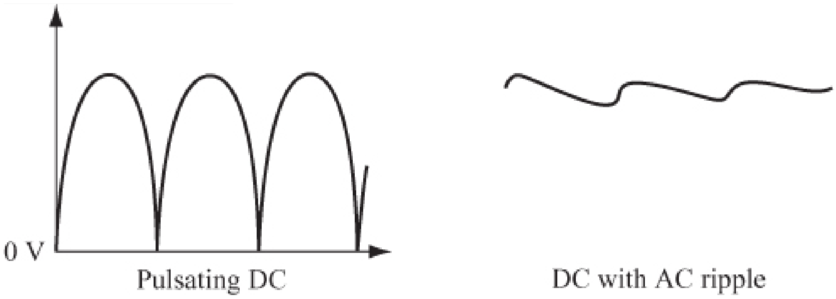

The load (which is “driven” by the power supply) can be a simple lamp or a complex electronic circuit. Whatever load you use, it requires a certain voltage across its terminals and draws a current. Therefore, the load has a resistance.

Usually, the voltage and current required by the load (and, hence, its resistance) are known, and you must design the power supply to provide that voltage and current.

To simplify the circuit diagrams, you can treat the load as a simple resistor.

Questions

A. What does the smoothing section of a power supply do? _____

B. What is connected to a power supply, and what can you treat it like? _____

Answers

A. The smoothing section converts the pulsating DC to a “pure” DC.

B. A load such as a lamp or an electronic circuit is connected to a power supply. In most cases, you can treat the load as you do a resistor.

16 Figure 11.33 shows a power supply circuit with a resistor as the load.

Figure 11.33

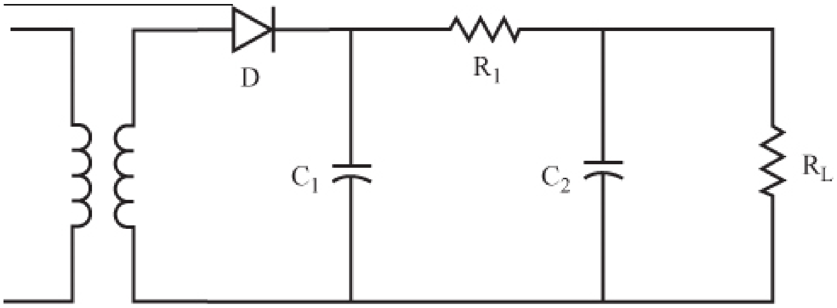

Questions

Look at the circuit shown in Figure 11.33 and answer the following questions:

A. What type of secondary coil is used? _____

B. What type of rectifier is used? _____

C. What components make up the smoothing section? _____

D. What output would you expect from the rectifier section? _____

Answers

A. A secondary coil with no center tap

B. A single-diode, half-wave rectifier

C. A resistor and two capacitors (R_1, C_1, and C_2)

D. Half-wave pulsating DC

17 Figure 11.34 shows the output waveform from the rectifier portion of the power supply circuit shown in Figure 11.33.

Figure 11.34

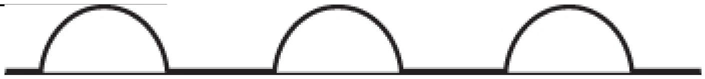

This waveform is the input to the smoothing section of the power supply circuit. Use one of the DC pulses (shown in Figure 11.35) to analyze the effect of the smoothing section on the waveform.

Figure 11.35

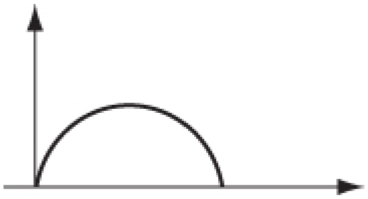

As the voltage level of the DC pulse rises to its peak, the capacitor C_1 is charged to the peak voltage of the DC pulse.

When the input DC pulse drops from its peak voltage back to 0 volts, the electrons stored on capacitor C_1 discharge through the circuit. This maintains the voltage across the load resistor at close to its peak value, as shown in Figure 11.36 . The DC pulse to the right of the diode stays at the peak voltage, even though V_{in} drops to zero.

Figure 11.36

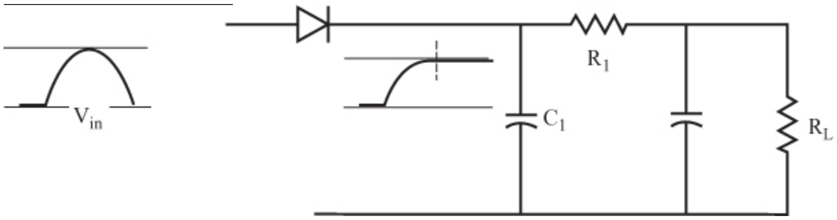

Question

What discharge path is available for the capacitor C_1? _____

Answer

The diode is not conducting, so the capacitor cannot discharge through the diode. The only possible discharge path is through R_1 and the load R_L.

18 If no further pulses pass through the diode, the voltage level drops as the capacitor discharges, resulting in the waveform shown in Figure 11.37 .

Figure 11.37

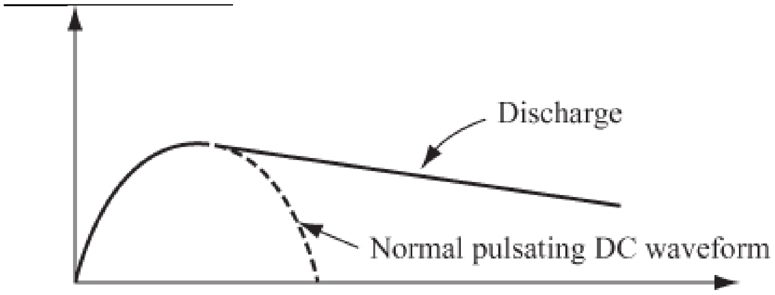

However, if another pulse passes through the diode before the capacitor is discharged, the resulting waveform looks like that shown in Figure 11.38 .

Figure 11.38

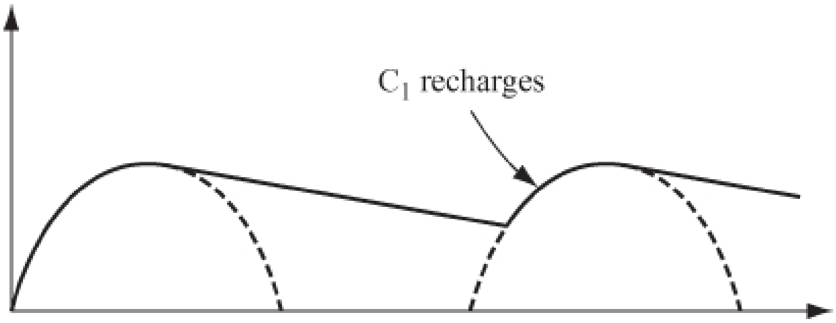

The capacitor discharges only briefly before the second pulse recharges it to peak value. Therefore, the voltage applied to the load resistor drops only a small amount.

Applying further pulses can produce this same recharging effect again and again. Figure 11.39 shows the resulting waveform.

Figure 11.39

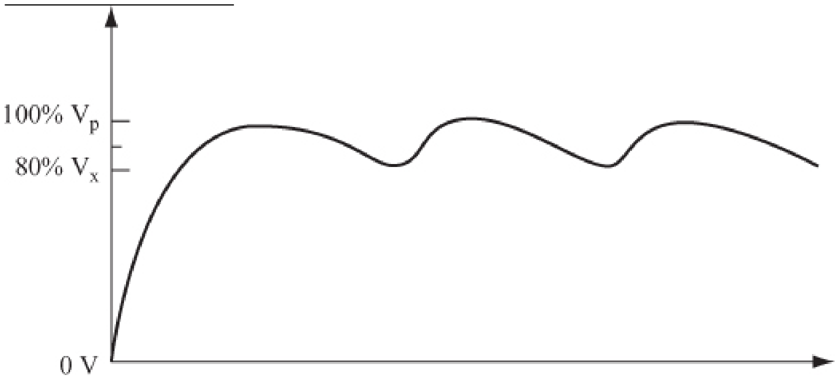

The waveform in Figure 11.39 has a DC level with an AC ripple, which varies between

V_p and V_x. If you choose values of C_1, R_1, and R_L that produce a discharge time constant for C_1 equal to about 10 times the duration of an input pulse, V_x will be approximately 80 percent of V_p.

If the discharge time you select is greater than 10 times the duration of an input pulse, the smoothing effect minimizes the AC ripple. A time constant of 10 times the pulse duration results in practical design values that are used throughout this chapter.

Note The smoothing section of a power supply circuit is sometimes referred to as a *low-pass filter*. Though such a circuit can function as a low-pass filter, in the case of a power supply circuit converting AC to DC, it is the release of electrons by the capacitor that is primarily responsible for leveling out the pulsating DC. For that reason, this discussion uses the term *smoothing section*.

Question

Estimate the average DC output level of the waveform shown in Figure 11.39. _____

Answer

Approximately 90 percent of V_p

19 The output from the secondary coil of the circuit shown in Figure 11.40 is a 28 V_{rms}, 60 Hz sine wave. For this circuit, you need 10 volts DC across the 100-ohm load resistor.

Figure 11.40

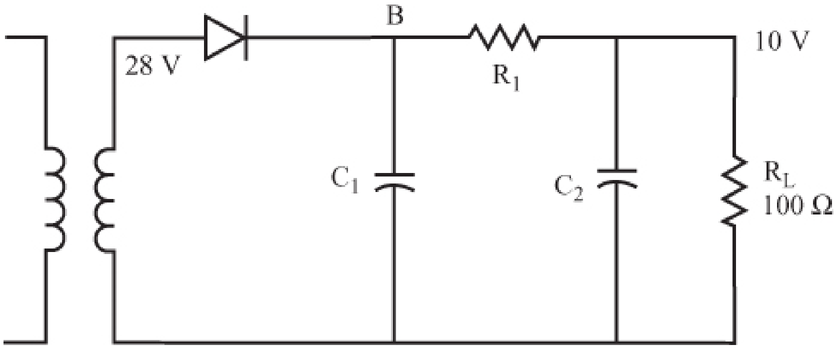

Question

What is the peak voltage out of the rectifier? _____

Answers

The transformer secondary coil delivers 28 V rms , so

$$V_p = \sqrt{2} \times V_{rms} = 1.414 \times 28V_{rms} = 39.59 \text{ volts}$$

or about 40 volts.

20 Figure 11.41 shows the waveform after the diode has rectified the sine wave for the halfway rectifier circuit shown in Figure 11.40 .

Figure 11.41

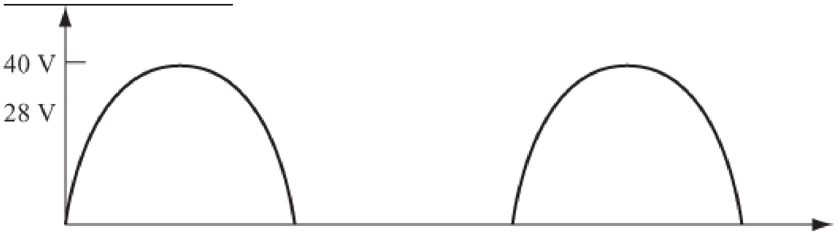

Question

Calculate the duration of one pulse. _____

Answer

60 Hz represents 60 cycles (that is, *wavelengths*) in 1 second. Therefore, one wavelength lasts for 1/60 second.

1/60 second = 1000/60 milliseconds = 16.67 ms

Therefore, the duration of a pulse, which is half a wavelength, is 8.33 ms.

21 The average DC voltage at point B in the circuit shown in Figure 11.40 is approximately 90 percent of the peak value of the sine wave from the secondary coil, or $V_B = 0.9 \times 40$ volts $= 36$ volts. R_1 and R_L act as a voltage divider to reduce the 36-volt DC level to the required 10 volts DC at the output.

Question

Using the voltage divider formula, calculate the value of R_1 that will result in 10 volts DC across the 100-ohm load resistor. _____

Answer

$$V_{out} = \frac{V_{in} R_L}{(R_1 + R_L)}$$

$$10 = \frac{36 \times 100}{(R_1 + 100)}$$

Therefore, $R_1 = 260$ ohms

22 Figure 11.42 shows the half-wave rectifier circuit with the 260-ohm value you calculated for R_1.

Figure 11.42

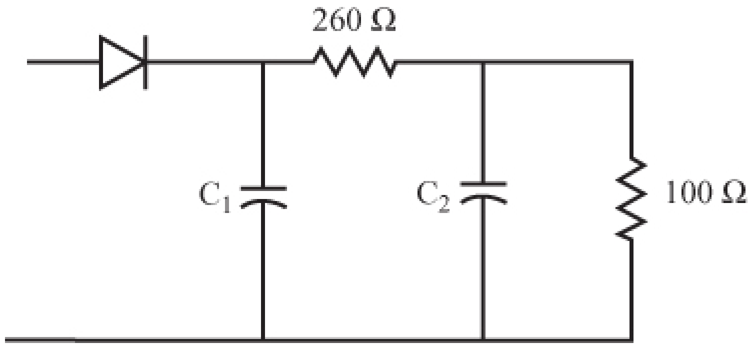

Now, choose a value for C_1 that produces a discharge time through the two resistors equal to 10 times the input wave duration.

Questions

A. How long should the discharge time constant be for the circuit in Figure 11.42 ? Refer to problems 18 and 20. _____

B. Given the time constant, calculate the value of C_1 . _____

Answers

A. The time constant should be 10 times the pulse duration (8.33 ms), so

B. τ = 10 × 8.33 ms = 83.3 ms or 0.083 seconds

C. $\tau = R \times C = (R_1 + R_L) \times C_1 = 360 \times C_1$

D. Therefore, $0.0833 = 360 \times C_1$, or $C_1 = 230$ μF

23 Figure 11.43 shows voltage waveforms at various points in the half-wave rectifier circuit.

Figure 11.43

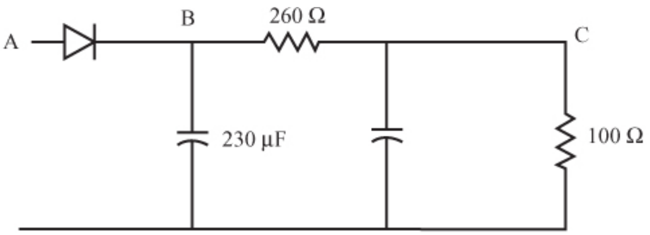

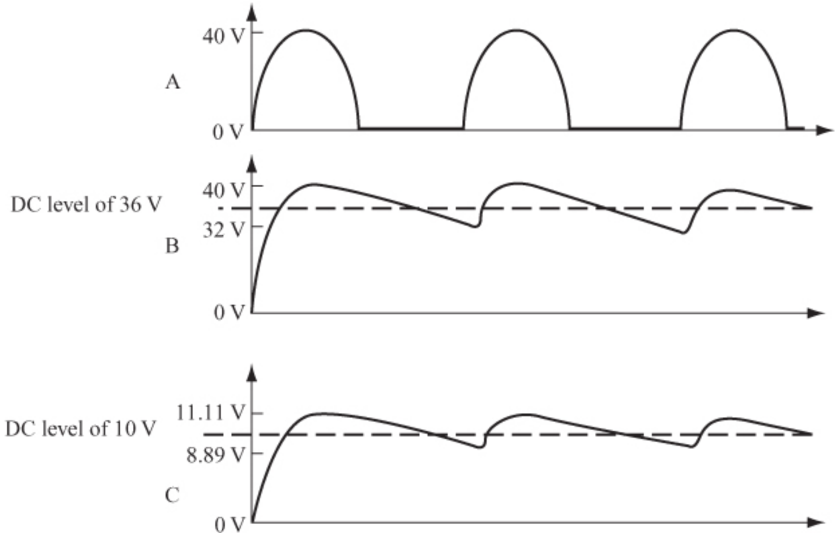

Questions

A. What happens to the DC output voltage between points B and C in this circuit? _____

B. What happens to the waveform between points A and C in the circuit? _____

Answers

A. The voltage has been reduced from 36 volts to 10 volts.

B. The waveform has changed from pulsating DC to a 10-volt DC level with an AC ripple.

24 In most cases, the level of the AC ripple is still too high, and further smoothing is required. Figure 11.44 shows the portion of the half-wave rectifier circuit that forms a voltage divider using R_1, and the parallel combination of R_L and C_2. This voltage divider reduces the AC ripple and the DC voltage level.

Figure 11.44

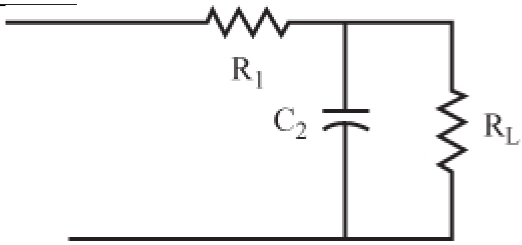

Choose a value for C_2 that causes the capacitor's reactance (X_{C2}) to be equal to or less than one-tenth of the resistance of the load resistor. C_2, R_1, and R_2 form an AC voltage divider. As discussed in problem 26 of Chapter 6, “Filters,” choosing such a value for C_2 simplifies the calculations for an AC voltage divider circuit containing a parallel resistor and capacitor.

Questions

A. What should the value of X_{C2} be? _____

B. What is the formula for the reactance of a capacitor? _____

C. What is the frequency of the AC ripple? _____

D. Calculate the value of the capacitor C_2. _____

Answers

A. $X_{C2} = R_L/10 = 100/10 = 10\ \Omega$ or less

B.

$$X_C = \frac{1}{2\ \ fC}$$

C. 60 Hz. This is identical to the frequency of the sine wave output from the transformer's secondary coil.

D. Solving the reactance formula for C results in the following:

$$C_2 = \frac{1}{2\ \ fX_c} = \frac{1}{2 \times \quad \times 60\,\text{Hz} \times 10\,\Omega} = 265\,\mu\text{F}$$

25 Figure 11.45 shows the half-wave rectifier circuit with all capacitor and resistor values.

Figure 11.45

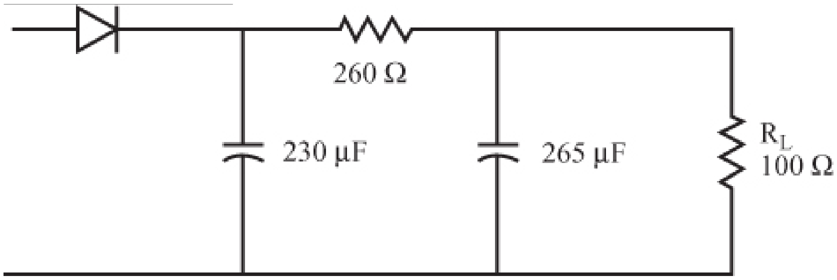

Because X C2 is one-tenth of R L , you can ignore R L in AC voltage divider calculations. Figure 11.46 shows the resulting AC voltage divider circuit.

Figure 11.46

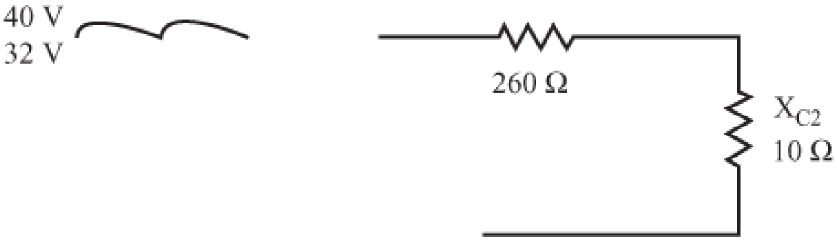

Questions

A. What is the peak-to-peak voltage at the input to the AC voltage divider? _____

B. Find the AC ripple output across R_L using the AC voltage divider formula discussed in problem 26 of Chapter 6. _____

Answers

A. $V_{pp} = V_p - V_x = 40$ volts $-$ 32 volts $= 8\ V_{PP}$

B.
$$\text{AC}\,V_{out} = (\text{AC}\,V_{in}) \times \frac{X_{C2}}{\sqrt{{X_{C2}}^2 + {R_1}^2}}$$
$$\text{AC}\,V_{out} = 8 \times \frac{10}{\sqrt{(10^2 + 260^2)}} = 0.31\,V_{pp}$$

Note This result means that the addition of C_2 lowers the AC ripple shown by curve C in Figure 11.43 , with peak values of 11.11 and 8.89, to values of 10.155 and 9.845 volts. This represents a lower ripple at the output. Hence, C_2 aids the smoothing of the 10 volts DC at the output.

26 You can apply the calculations you performed for a half-wave rectifier circuit in the last few problems to a full-wave rectifier circuit. In the next few problems, you calculate the values of R_1, C_1, and C_2 required to provide 10 volts DC across a 100-ohm load for a full-wave rectifier circuit with a 28 V_{rms} sine wave supplied by the secondary coil of a transformer.

Figure 11.47 shows the output waveform

from the rectifier section of the circuit.

Figure 11.47

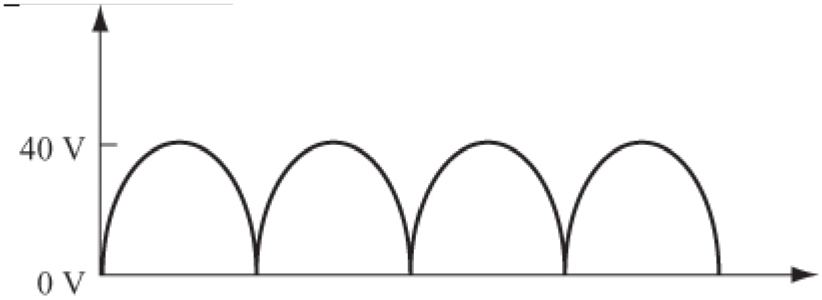

Figure 11.48 shows the waveform that results from using a smoothing capacitor.

Figure 11.48

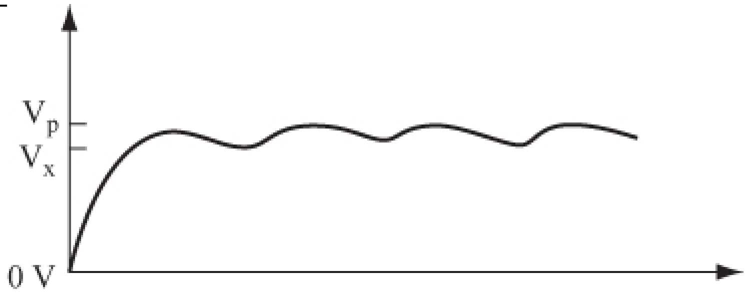

If the discharge time constant C 1 is 10 times the period of the waveform, V x is approximately 90 percent of V p . The average DC level is approximately 95 percent of V p .

Questions

A. What is the average DC level for the half-wave rectifier at point B in Figure 11.43 ? _____

B. What is the average DC level for the waveform in Figure 11.48 given that V p = 40 volts? _____

C. Why does a full-wave rectifier have a higher average DC level than a half-wave rectifier? _____

Answers

A. 36 volts, which is 90 percent of V_p.

B. 38 volts, which is 95 percent of V_p.

C. The slightly higher values occur because the capacitor does not discharge as far with full-wave rectification, and, as a result, there is slightly less AC ripple. Therefore, V_x is higher and the average DC level is higher.

27 You can use the method for calculating the value of R_1 for a half-wave rectifier (see problem 21) to calculate the value of R_1 for a full-wave rectifier.

Question

Calculate the value of R_1 when R_L = 100 Ω, V_{in} = 38 volts, and the required voltage across R_L is 10 volts. _____

Answer

$$V_{out} = 10\,\text{volts} = \frac{V_{in} R_L}{(R_1 + R_L)} = \frac{38 \times 100}{(R_1 + 100)}$$

Therefore, R_1 = 280 ohms.

28 You can also use the method for calculating the value of C_1 for a half-wave rectifier (see problem 22) to calculate the value of C_1 for a full-wave rectifier.

Question

Calculate the value of C_1. _____

Answer

With a time constant of τ = 83.3 ms. and a discharge resistance of $R_1 + R_L$ = 380 ohms, $C_1 = 220\ \mu F$.

29 You can use the voltage divider equation to find the amount of AC ripple across the load resistor for a full-wave rectifier with $R_1 = 280\ \Omega$ and $R_L = 100\ \Omega$. For $V_p = 40$ volts, the calculation results in 10.52 volts. For $V_x = 36$ volts, the calculation results in 9.47 volts. Therefore, the voltage levels at the load resistor vary between 10.52 volts and 9.47 volts, with an average DC level of 10 volts. You can reduce the AC ripple by adding a second capacitor in parallel with the load resistor.

Questions

Use the method for calculating the value of C_2 for the half-wave rectifier in problem 24.

A. Calculate the reactance of the second capacitor (C_2). _____

B. Calculate the value of C_2. (The frequency of the AC ripple for the full-wave rectifier is 120 Hz.) _____

Answers

A. The reactance should be one-tenth (or less) of the load resistance. Therefore, it should be 10 ohms or less.

B.

$$C_2 = \frac{1}{2\ \ fX_C} = \frac{1}{2 \times \quad \times 120\,\mathrm{Hz} \times 10\,\Omega} = 135\,\mu\mathrm{F}$$

30 The AC ripple at the first smoothing capacitor ranges from 36 volts to 40 volts. The AC ripple at the load ranges from 9.47

volts to 10.52 volts when there is only one capacitor in the circuit.

Question

Calculate the upper and lower values of the AC ripple at the output if you use a second capacitor with a value of 135 μF in parallel with the load resistor. You can use the same formulas as those for the half-wave rectifier in problem 25. X_{C2} is 10 Ω from problem 29; R_1 = 280 from problem 27; and AC V_{in} = $V_p - V_x$ = (40 volts − 36 volts) = 4 V_{pp}.

—————

Answer

$$\text{AC } V_{out} = (\text{AC } V_{in}) \times \frac{X_{C2}}{\sqrt{(R_1^{\,2} + X_{C2}^{\,2})}} = 0.143\, V_{pp}$$

The result of approximately 0.14 V_{pp} means that the output will now vary from 10.07 to 9.93 volts. This shows that the second capacitor lowers the ripple significantly. The AC ripple is less than half of the ripple shown for the half-wave rectifier in problem 25. In other words, a full-wave rectifier produces a smoother DC output than a half-wave rectifier.

31 Figure 11.49 shows a full-wave rectifier circuit with an output voltage of 5 volts across a 50 Ω load resistor. Use the following steps to calculate the values of the other components.

Figure 11.49

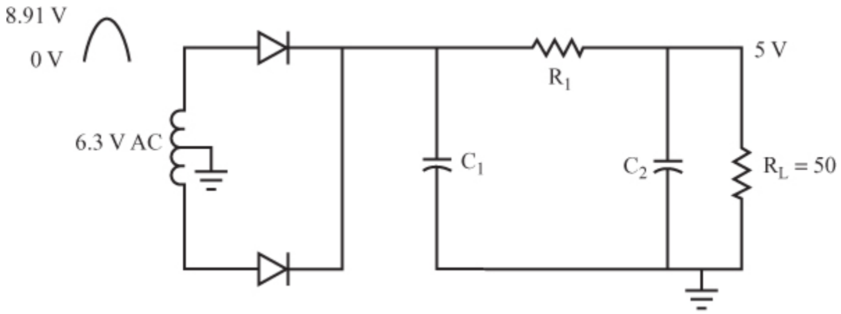

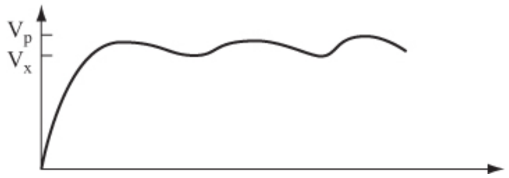

Questions

A. What are V_p, V_x, and the DC level at the first capacitor? _____

B. Calculate the value of R_1 required to make the DC level at the output 5 volts. _____

C. Calculate the value of C_1. _____

D. Calculate the value of C_2. _____

E. What is the amount of AC ripple at the output? _____

F. Draw the final circuit showing the calculated values. Use a separate sheet of paper for your drawing.

Answers

A.

$$V_p = 6.3 \times \sqrt{2} = 8.91\,\text{volts}, V_x = 90 \text{ percent of } V_p = 8.02\,\text{volts.}$$

The DC level is 95 percent of V_p, which is 8.46 volts.

B. About 35 ohms.

C. 980 µF.

D. Using X_{C2} = 5 ohms and 120 Hz, C_2 = 265 µF.

E. At the input to the smoothing section, the AC variation is 8.91 to 8.02, or 0.89 V_{pp}. Using the AC voltage divider equation with R_1 = 35 ohms and X_{C2} = 5 ohms, AC V_{out} equals approximately 0.13 V_{pp}. Therefore, the AC variation at the output is 5.065 to 4.935 volts, a small AC ripple.

F. See Figure 11.40.

Figure 11.50

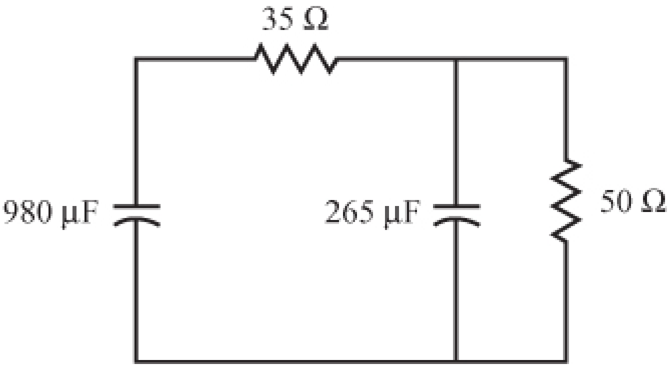

Using the simple procedure shown here always produces a working power supply circuit. This is not the only design procedure you can use for power supplies, but it is one of the simplest and most effective.

Summary

This chapter introduced the following concepts and calculations related to power supplies:

The effects of diodes on AC signals

Methods of rectifying an AC signal

Half-wave and full-wave rectifier circuit designs

The calculations you can use to determine component values for half-wave and full-wave rectifier power supply circuits

Self-Test

These questions test your understanding of the information presented in this chapter. Use a separate sheet of paper for your diagrams or calculations. Compare your answers with the answers provided following the test.

In questions 1 through 5, draw the output waveform of each circuit. The input is given in each case.

1. See Figure 11.51 .

Figure 11.51

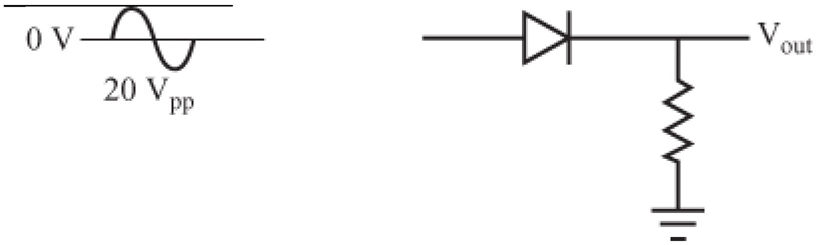

2. See Figure 11.52 .

Figure 11.52

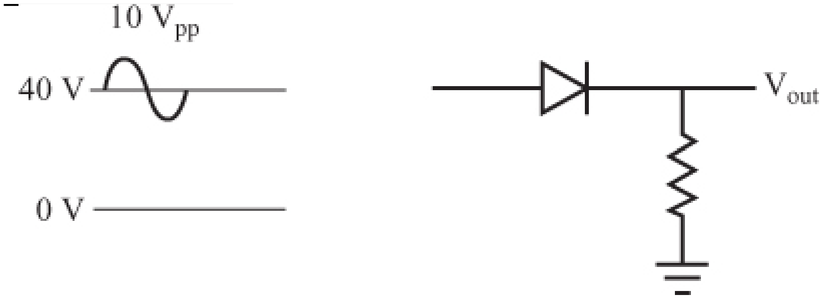

3. See Figure 11.53 .

Figure 11.53

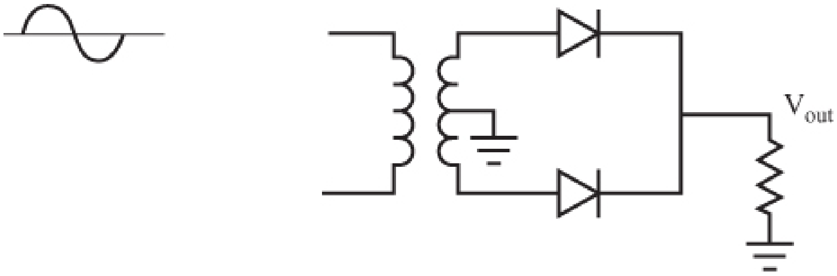

4. See Figure 11.54 .

Figure 11.54

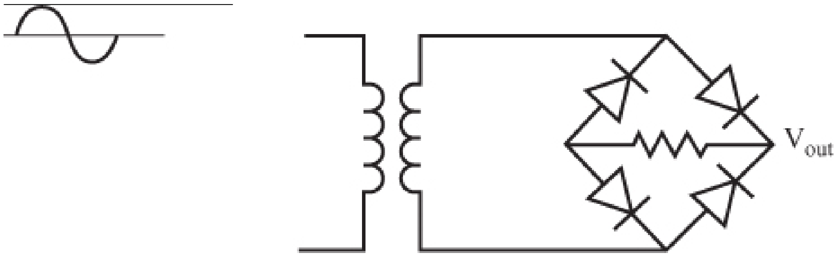

5. See Figure 11.55 .

Figure 11.55

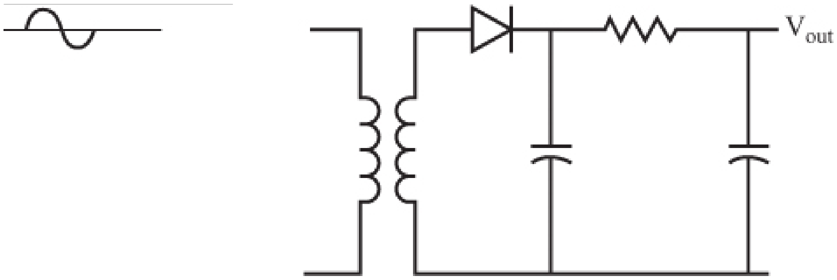

6. In the circuit shown in Figure 11.56 , 100 V rms at 60 Hz appears at the secondary coil of the transformer; 28 volts DC with as little AC ripple as possible is required across the 220-ohm load. Find R 1 , C 1 , and C 2 . Find the approximate AC ripple.

Figure 11.56

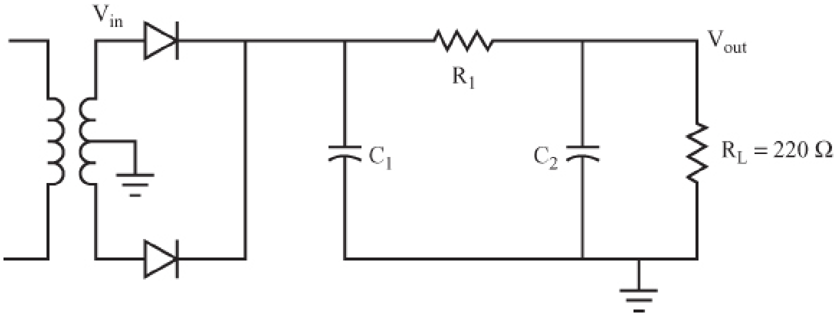

Answers to Self-Test

If your answers do not agree with those given here, review the problems indicated in parentheses.

1. See Figure 11.57 .

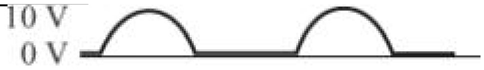

Figure 11.57

(problems 1–5)

2. See Figure 11.58 .

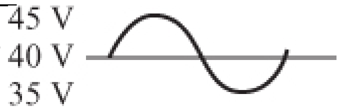

Figure 11.58

(problem 2)

3. See Figure 11.59 .

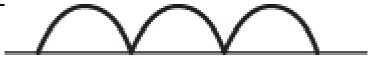

Figure 11.59

(problem 11)

4. See Figure 11.60 .

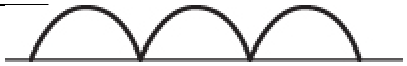

Figure 11.60

(problem 13)

5. See Figure 11.61 .

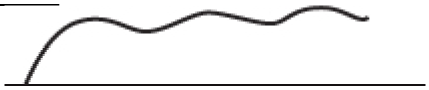

Figure 11.61

(problems 15–18)

6. R 1 = 833 ohms, C 1 = 79 µF: let X C2

= 22 ohms and then C 2 = 60 μF

$$\mathrm{AC\,V_{out}} = 14 \times \frac{22}{\sqrt{(22^2 \times 833^2)}} = 0.37\,\mathrm{V_{pp}}$$

(problems 26–30)

Chapter 12

Conclusion and Final Self-Test

In this book, you have discovered basic concepts and formulas that provide a foundation for your studies in modern electronics, whether you become a dedicated hobbyist or study electrical or electronics engineering.

Conclusion

Having read this book, you should now know enough to read intermediate-level electronics books and articles intelligently, to build electronics circuits and projects, and to pursue electronics to whatever depth and for whatever reason you want. Specifically, you should now be able to do the following:

- Recognize all the important, discrete electronics components in a schematic diagram.
- Understand how circuits that use discrete components work.
- Calculate the component values needed for circuits to function efficiently.
- Design simple circuits.
- Build simple circuits and electronics projects.

To see how much you have learned, you may want to take the final self-test at the end of this chapter. It tests your comprehension of the concepts and formulas presented throughout this book.

When you complete the following self-test and feel confident that you have mastered the information in this book, refer to Appendix E, “Supplemental Resources,” for additional resources for further learning, including the following:

- Books such as *The Art of Electronics* by Paul Horowitz and Winfield Hill (New York: Cambridge University Press, 1989) provide a great next step in further electronics study.
- Magazines such as *Everyday Practical Electronics* offer interesting projects in each issue.
- You can browse websites for electronics project ideas.For example, Earl Boysen's website, www.buildinggadgets.com, provides tips, ideas, and links to a variety of great online resources.

Note For those interested in more serious study, you should be aware that there is a difference between the path you take to become an electrician (or technician) and an electrical (or electronics) engineer. Training for electronics technicians is available in military trade schools, public and private vocational schools, and in many high schools. Engineers are required to understand the mathematical details in more depth and must take at least a 4-year curriculum at an accredited college or university.

Whatever your goal, you can feel confident that this book has given you a solid grounding for your future studies. Wherever you go in electronics, good luck!

Final Self-Test

This final test allows you to assess your overall knowledge of electronics. Answers and review references follow the test. Use a separate sheet of paper for your calculations and drawings.

1. If R = 1 MΩ and I = 2 μA, find the voltage. _____
2. If V = 5 volts and R = 10 kV, find the current. _____
3. If V = 28 volts and I = 4 amperes, find the resistance. _____
4. If 330 ohms and 220 ohms are connected in parallel, find the equivalent resistance. _____
5. If V = 28 volts and I = 5 mA, find the power. _____
6. If the current through a 220-ohm resistor is 30.2 mA, what is the power dissipated by the resistor? _____
7. If the power rating of a 1000-ohm resistor is 0.5 watts, what is the maximum current that can safely flow through the resistor? _____
8. If a 10-ohm resistor is in series with a 32-ohm resistor, and the combination is across a 12-volt supply, what is the voltage

drop across each resistor, and what will the two voltage drops add up to? _____

9. A current of 1 ampere splits between 6-ohm and 12-ohm resistors in parallel. Find the current through each. _____

10. A current of 273 mA splits between 330-ohm and 660-ohm resistors in parallel. Find the current through each resistor. _____

11. If R = 10 kV and C = 1 μF, find the time constant. _____

12. If R = 1 MV and C = 250 μF, find the time constant. _____

13. Three capacitors of 1 μF, 2 μF, and 3 μF are connected in parallel. Find the total capacitance. _____

14. Three capacitors of 100 μF, 220 μF, and 220 μF are connected in series. Find the total capacitance. _____

15. Three capacitors of 22 pF, 22 pF, and 33 pF are connected in series. Find the total capacitance. _____

16. What is the knee voltage for a germanium diode? _____

17. What is the knee voltage for a silicon diode? _____

18. In the circuit shown in Figure 12.1 , V S = 5 volts and R = 1 kV. Find the current through the diode, I D . _____

19. For the circuit shown in Figure 12.1 , V = 12 volts and R = 100 ohms. Find I D . _____

20. For the circuit shown in Figure 12.2 , V S

= 100 volts, R_1 = 7.2 kV, R_2 = 4 kV, and V_Z = 28 volts. Find the current through the zener diode, I_Z. _____

21. For the circuit in Figure 12.2, V_S = 10 volts, R_1 = 1 kV, R_2 = 10 kV, and V_Z = 6.3 volts. Find I_Z. _____

22. Using the circuit shown in Figure 12.3, find the DC collector voltage, V_C, if V_S = 28 volts, β = 10, R_B = 200 kV, and R_C = 10 kV. _____

23. Again, using the circuit shown in Figure 12.3, find R_B if V_S = 12 volts, β = 250, R_C = 2.2 kV, and V_C = 6 volts. _____

24. Using the circuit shown in Figure 12.3, find β if V_S = 10 volts, R_B = 100 kV, R_C = 1 kV, and V_C = 5 volts. _____

25. What are the three terminals for a JFET called, and which one controls the operation of the JFET? _____

26. Using the circuit shown in Figure 12.4, find the value of R_B required to turn the transistor ON if V_S = 14 volts, R_C = 10 kV, and β = 50. _____

27. Again, using the circuit shown in Figure 12.4, find the value of R_B required to turn the transistor ON if V_S = 5 volts, R_C = 4.7 kV, and β = 100. _____

28. Using the circuit shown in Figure 12.5 ,find the values of R_1, R_2, and R_3 that can enable the switch to turn Q_2 ON and OFF , if V_S = 10 volts, β_1 = 50, β_2 = 20,

and R_4 = 2.2 kV. _____

29. Again, using the circuit shown in Figure 12.5, find the values of R_1, R_2, and R_3 that can enable the switch to turn Q_2 ON and OFF if V_S = 28 volts, β_1 = 30, β_2 = 10, and R_4 = 220 V. _____

30. An N-channel JFET has a drain saturation current of I_{DSS} = 14 mA. If a 28-volt drain supply is used, calculate the drain resistance, R_D. _____

31. Draw one cycle of a sine wave. _____

32. Mark in V_{pp}, V_{rms}, and the period of the waveform on your drawing for question 31. _____

33. If V_{pp} = 10 volts, find V_{rms}. _____

34. If V_{rms} = 120 volts, find V_{pp}. _____

35. If the frequency of a sine wave is 14.5 kHz, what is the period of the waveform? _____

36. Find the reactance X_C for a 200 μF capacitor when the frequency is 60 Hz. _____

37. Find the value of the capacitance that gives a 50-ohm reactance at a frequency of 10 kHz. _____

38. Find the inductive reactance X_L for a 10-mH inductor when the frequency is 440 Hz. _____

39. Find the value of the inductance that has 100 ohms reactance when the frequency is 1 kHz. _____

40. Find the series and parallel resonant frequency of a 0.1 μF capacitor and a 4-mH inductor that has negligible internal resistance. _____

41. Using the circuit shown in Figure 12.6, find X_C, Z, V_{out}, I, tan θ, and θ, if V_{in} = 10 V_{pp}, f = 1 kHz, C = 0.1 μF, and R = 1600 ohms. _____

42. Again, using the circuit shown in Figure 12.6, find X_C, Z, V_{out}, I, tan θ, and θ, if V_{in} = 120 V_{rms}, f = 60 Hz, C = 0.33 μF, and R = 6 kV. _____

43. Using the circuit shown in Figure 12.7, find X_C, AC V_{out}, and DC V_{out}, if V_{in} = 1 V_{pp} AC, riding on a 5-volt DC level; f = 10 kHz; R_1 = 10 kV; R_2 = 10 kV; and C = 0.2 μF. _____

44. Again, using the circuit shown in Figure 12.7, find X_C, AC V_{out}, and DC V_{out}, if V_{in} = 0.5 V_{pp} AC, riding on a 10-volt DC level; f = 120 Hz; R_1 = 80 ohms; R_2 = 20 ohms; and C = 1000 μF. _____

45. In the circuit shown in Figure 12.8, V_{in} = 10 V_{pp} AC, riding on a 5-volt DC level; f = 1 kHz; L = 10 mH; r = 9 ohms; and R = 54 ohms. Find AC V_{out}, DC V_{out}, X_L, Z, tan θ, and θ. _____

46. In the circuit shown in Figure 12.9, L = 1 mH, C = 0.1 μF, and R = 10 ohms. Find f_r, X_L, X_C, Z, Q, and the bandwidth. _____

47. In the circuit shown in Figure 12.10, L =

10 mH, C = 0.02 μF, and r = 7 ohms.Find f_r, X_L, X_C, Z, Q, and the bandwidth. _____

48. If the voltage across the resonant circuit of question 47 is at a peak value of 8 volts at the resonant frequency, what is the voltage at the half-power points and what are the half-power frequencies? _____

49. Using the amplifier circuit shown in Figure 12.11, find the values of R_1, R_2, and R_E that can provide the amplifier with a voltage gain of 10. Use V_S = 28 volts, R_C = 1 kV, and β = 100. _____

50. Again, using the circuit shown in Figure 12.11, find the values of R_1, R_2, and R_E that can provide the amplifier a voltage gain of 20. Use V_S = 10 volts, R_C = 2.2 kV, and β = 50. _____

51. Using the circuit shown in Figure 12.11, how would you modify the amplifier in question 50 to obtain a maximum gain? Assume that the lowest frequency it has to pass is 50 Hz. _____

52. Using the JFET amplifier circuit shown in problem 42 of Chapter 8, “Transistor Amplifiers,” with a bias point of V_{GS} = −2.8 volts, a drain current of I_D = 2.7 mA, and V_{DS} = 12 volts, find the values of R_S and R_D. _____

53. If the transconductance of the JFET used in question 52 is 4000 μmhos, what is the AC voltage gain? _____

54. A certain op-amp circuit uses an input resistance of 8 kV to an inverting input. For the op-amp circuit to have a gain of 85, what should the value of the feedback resistance be? _____

55. If the input to the op-amp circuit of question 54 is 2 mV, what is the output? _____

56. What is an oscillator? _____

57. Why is positive feedback rather than negative feedback necessary in an oscillator? _____

58. What feedback method is used in a Colpitts oscillator? _____

59. What feedback method is used in a Hartley oscillator? _____

60. Draw the circuit of a Colpitts oscillator. _____

61. Draw the circuit of a Hartley oscillator. _____

62. What is the formula used to calculate the output frequency of an oscillator? _____

63. Draw the circuit symbol for a transformer with a center tap. _____

64. Name the two main coils used on a transformer. _____

65. What is the equation that shows the relationship between the input voltage, the output voltage, and the number of turns in each coil of a transformer? _____

66. What is the equation that shows the

relationship between the turns ratio and the currents in the primary and secondary coils of the transformer? _____

67. What is the equation that shows the relationship between the impedance of the primary coil, the impedance of the secondary coil, and the number of turns in each coil of a transformer? _____

68. What are the two main uses for transformers? _____

69. Draw a simple half-wave rectifier circuit with a smoothing filter at the output. _____

70. Draw a simple full-wave rectifier circuit using a center tap transformer and a smoothing filter at the output. _____

71. Given a 10 V rms input to a full-wave rectified power supply, calculate the values of R1, C1 and C2 (see Figure 12.12) that results in a 5-volt DC output across a 50-ohm load. _____

Figure 12.1

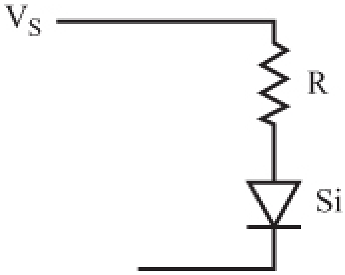

Figure 12.2

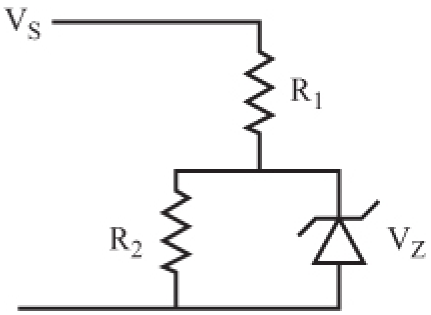

Figure 12.3

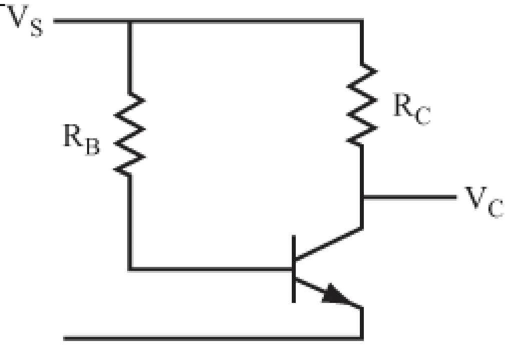

Figure 12.4

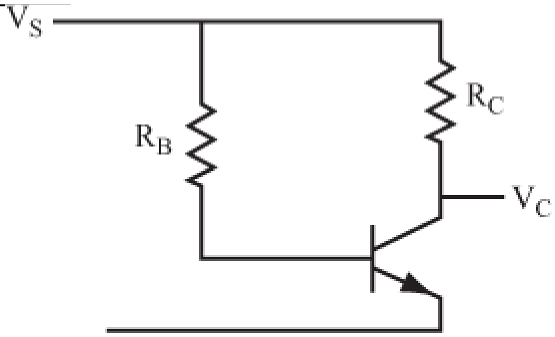

Figure 12.5

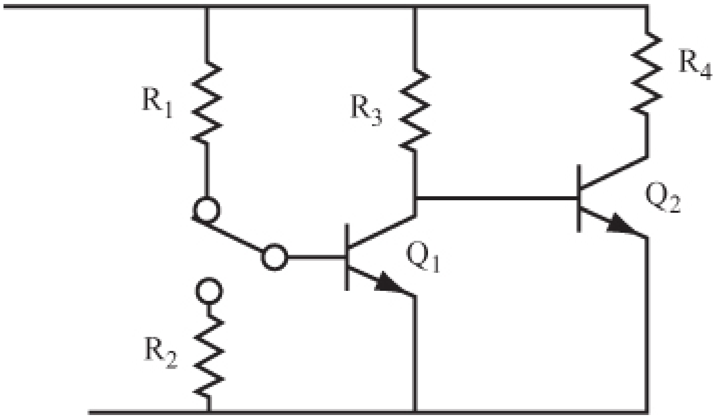
R$_4$
R$_1$
R$_3$
Q$_2$
Q$_1$
R$_2$

Figure 12.6

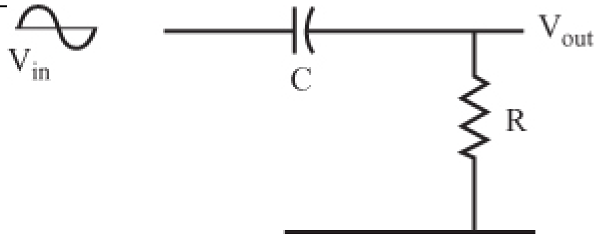

Figure 12.7

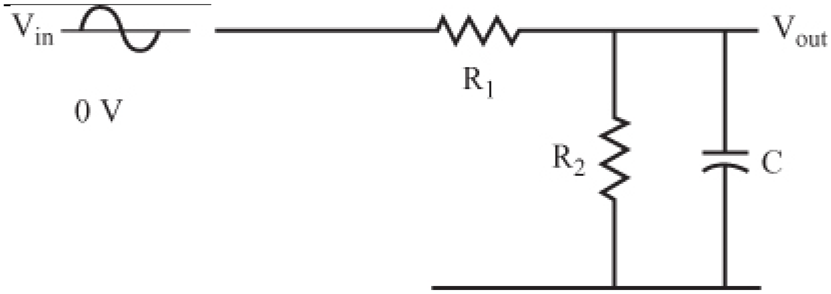
V$_{in}$
V$_{out}$
R$_1$
0 V
R$_2$
C

Figure 12.8

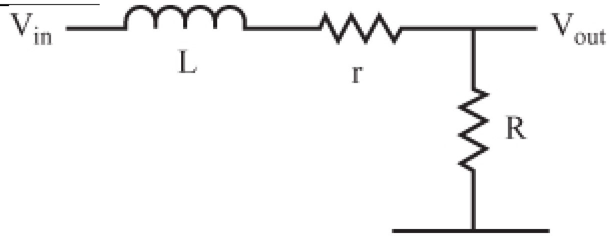
V$_{in}$
V$_{out}$
L
r
R

Figure 12.9

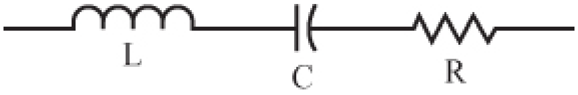

Figure 12.10

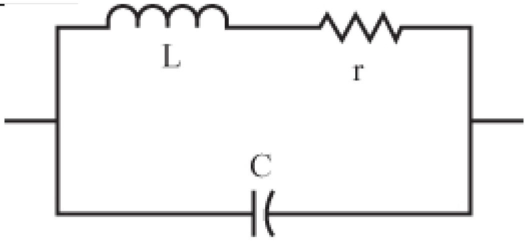

Figure 12.11

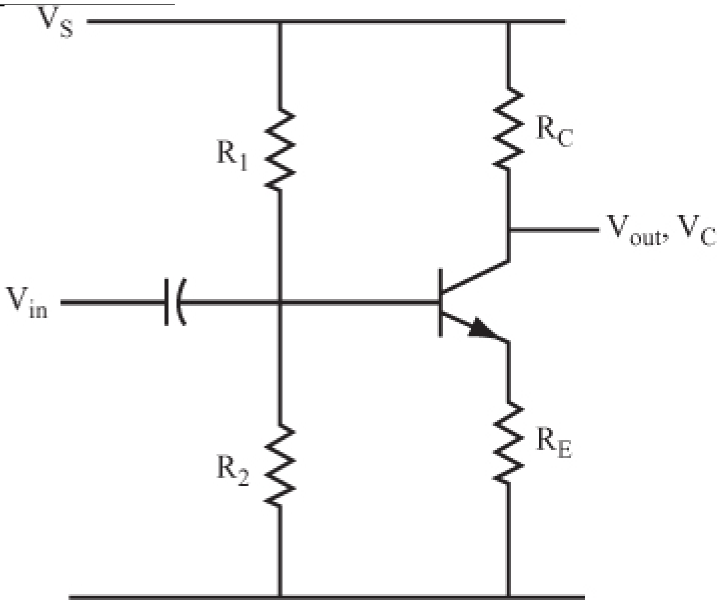
V_S
R_1
R_C
V_{out}, V_C
V_{in}
R_2
R_E

Figure 12.12

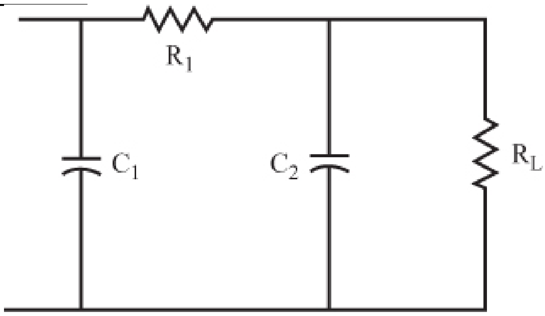

Answers to Final Self-Test

The references in parentheses to the right of the answers give you the chapter and problem number where the material is introduced so that you can easily review any concepts covered in the test.

1. $V = 2$ volts (Chapter 1, problem 5)
2. $I = 0.5$ mA (Chapter 1, problem 6)
3. $R = 7$ ohms (Chapter 1, problem 7)
4. 132 ohms (Chapter 1, problem 10)
5. $P = 140$ milliwatts or 0.14 watts (Chapter 1, problems 13 and 14)
6. 0.2 watts (Chapter 1, problems 13 and 15)
7. 22.36 mA (Chapter 1, problems 13 and 16)
8. 2.86 volts, 9.14 volts, 12 volts (Chapter 1, problems 23 and 26)
9. 2/3 ampere through the 6-ohm resistor; 1/3 ampere through the 12-ohm resistor (Chapter 1, problem 28 or 29)
10. 91 mA through the 660-ohm resistor; 182 mA through the 330-ohm resistor (Chapter 1, problem 28 or 29)
11. $\tau = 0.01$ seconds (Chapter 1, problem 34)
12. $\tau = 250$ seconds (Chapter 1, problem 34)
13. 6 μF (Chapter 1, problem 40)
14. 52.4 μF (Chapter 1, problem 41)
15. 8.25 μF (Chapter 1, problem 41)

16. Approximately 0.3 volts (Chapter 2, problem 9)
17. Approximately 0.7 volts (Chapter 2, problem 9)
18. I_D = 4.3 mA (Chapter 2, problem 12)
19. I_D = 120 mA (Chapter 2, problem 12)
20. I_Z = 3 mA (Chapter 2, problem 29)
21. I_Z = 3.07 mA (Chapter 2, problem 29)
22. V_C = 14 volts (Chapter 3, problems 20–23)
23. R_B = 1.1 MΩ (Chapter 3, problems 20–23)
24. β = 50 (Chapter 3, problems 20–23)
25. Drain, source, and gate, with the gate acting to control the JFET (Chapter 3, problem 28)
26. R_B = 500 kΩ (Chapter 4, problems 8)
27. R_B = 470 kΩ (Chapter 4, problems 4–8)
28. R_3 = 44 kΩ, R_1 = 2.2 kΩ, R_2 = 2.2 kΩ (Chapter 4, problems 19–23)
29. R_3 = 2.2 kΩ, R_1 = 66 kΩ, R_2 = 66 kΩ (Chapter 4, problems 19–23)
30. R_D = 2 kΩ (Chapter 4, problem 39)
31. See Figure 12.13 .

Figure 12.13

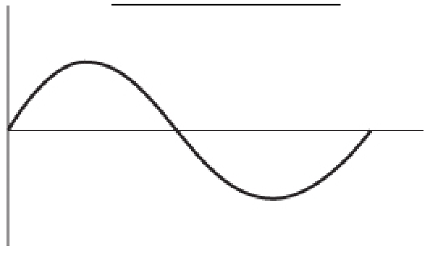

(Chapter 5, problem 7)

32. See Figure 12.14 .

Figure 12.14

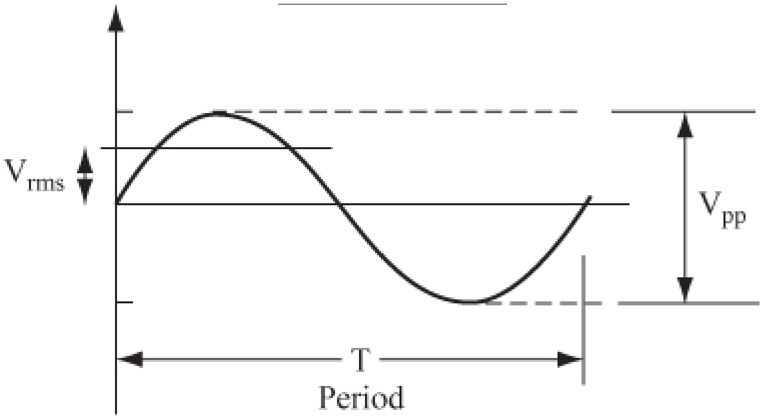

(Chapter 5, problems 3 and 7)

33. 3.535 volts (Chapter 5, problem 4)

34. 340 volts (Chapter 5, problem 5)

35. 69 μsec (Chapter 5, problem 7)

36. 13.3 ohms (Chapter 5, problem 14)

37. 0.32 μF (Chapter 5, problem 14)

38. 27.6 ohms (Chapter 5, problem 17)

39. 16 mH (Chapter 5, problem 17)

40. 8 kHz (Chapter 5, problems 19 and 21)

41. X C = 1.6 k Ω, Z = 2263 ohms, V out = 7.07 volts, I = 4.4 mA, tan θ = 1, θ = 45 degrees (Chapter 6, problems 10 and 23)

42. X C = 8 k Ω, Z = 10 k Ω, V out = 72 volts, I = 12 mA, tan θ = 1.33, θ = 53.13 degrees (Chapter 6, problems 10 and 23)

43. X C = 80 ohms, AC V out = 8 mV, DC V out = 2.5 volts (Chapter 6, problem 26)

44. X C = 1.33 ohms, AC V out = 8.3 mV,

DC V_{out} = 2 volts (Chapter 6, problem 26)

45. X_L = 62.8 ohms, Z = 89 ohms, AC V_{out} = 6.07 volts, DC V_{out} = 4.3 volts, $\tan\theta = 1$, θ = 45 degrees (Chapter 6, problems 31 and 35)

46. f_r = 16 kHz, X_L = 100 ohms, X_C = 100 ohms, Z = 10 ohms, Q = 10, BW = 1.6 kHz (Chapter 7, problems 2, 6, and 20)

47. f_r = 11,254 Hz, $X_L = X_C$ = 707 ohms, Z = 71.4 kΩ, Q = 101, BW = 111 Hz (Chapter 7, problems 10, 11, and 20)

48. V_{hp} = 5.656 volts, f_{1hp} = 11,198 Hz, f_{2hp} = 11,310 Hz (Chapter 7, problem 27)

49. Your values should be close to the following: R_E = 100 ohms, V_C = 14 volts, V_E = 1.4 volts, V_B = 2.1 volts, R_2 = 1.5 kΩ, R_1 = 16.8 kΩ (Chapter 8, problem 17)

50. R_E = 110 ohms, V_C = 5 volts, V_E = 0.25 volts, V_B = 0.95 volts, R_2 = 2.2 kΩ, R_1 = 18.1 kΩ (Chapter 8, problem 17)

51. The gain can be increased by using a capacitor to bypass the emitter resistor R_E; C_E = 300 μF (approximately). (Chapter 8, problem 20)

52. R_S = 1.04 kΩ, R_D = 3.41 kΩ (Chapter 8, problem 42)

53. A_v = –13.6 (Chapter 8, problem 39)

54. R_F = 680 kΩ (Chapter 8, problem 45)

55. V_{out} = 170 mV and is inverted

(Chapter 8, problem 45)

56. An oscillator is a circuit that emits a continuous sine wave output without requiring an input signal. Other types of oscillators exist that do not have sine wave outputs, but they are not discussed in this book. (Chapter 9, introduction)

57. Positive feedback causes the amplifier to sustain an oscillation or sine wave at the output. Negative feedback causes the amplifier to stabilize, which reduces oscillations at the output. (Chapter 9, problems 2 and 3)

58. A capacitive voltage divider (Chapter 9, problem 14)

59. An inductive voltage divider (Chapter 9, problem 14)

60. See Figure 12.15 .

Figure 12.15

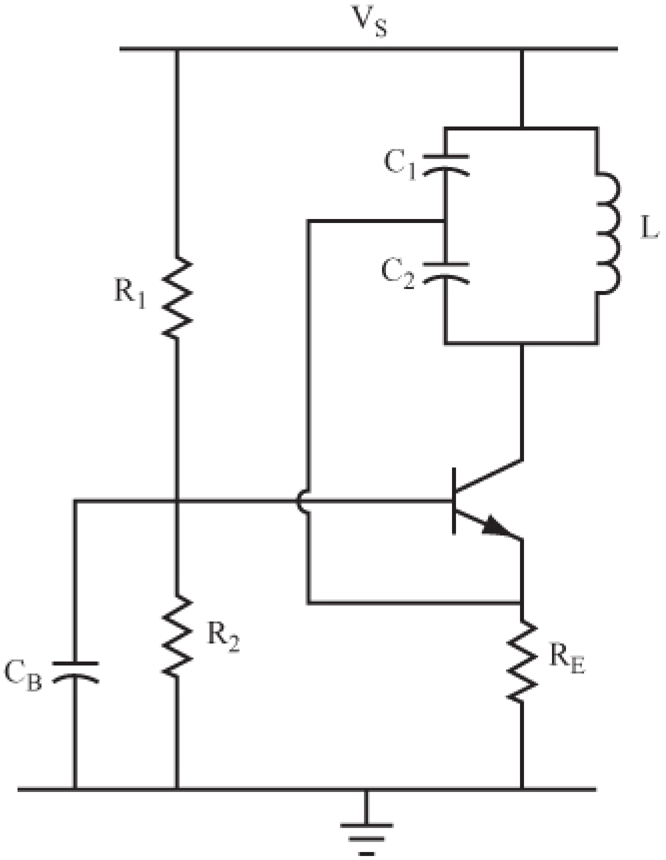
V_S
C_1
L
C_2
R_1
R_2
C_B
R_E

(Chapter 9, problem 24)

61. See Figure 12.16 .

Figure 12.16

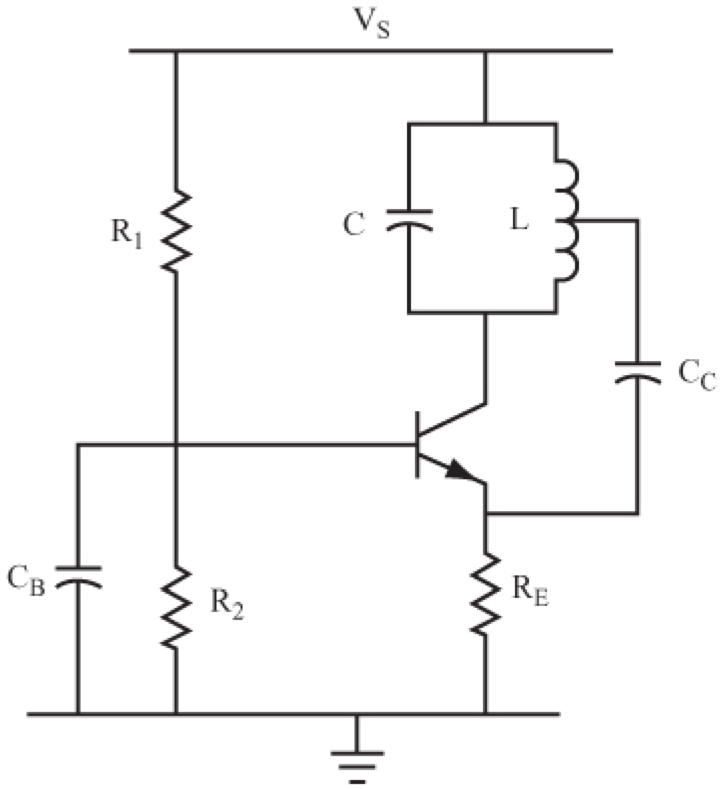
V_S
R_1
C
L
C_C
C_B
R_2
R_E

(Chapter 9, problem 25)

62. $$f_r = \frac{1}{2\pi\sqrt{LC}}$$ (Chapter 9, problem 11)

63. See Figure 12.17 .

Figure 12.17

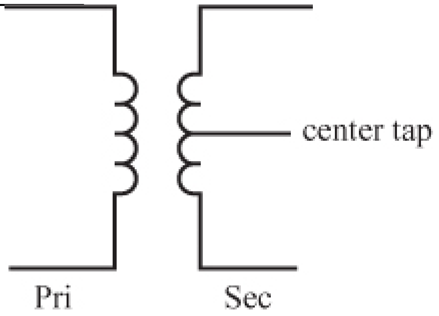
center tap
Pri
Sec

(Chapter 10, problem 4)

64. Primary and secondary (Chapter 10, problem 2)

65. $V_{in}/V_{out} = V_P/V_S = N_P/N_S = TR$ (Chapter 10, problem 6)

66. $I_{out}/I_{in} = I_S/I_P = N_P/N_S = TR$ (Chapter 10, problem 13)

67. $Z_{in}/Z_{out} = (N_P/N_S)^2$, or impedance ratio, is the square of the turns ratio. (Chapter 10, problem 16)

68. They are used for stepping up or stepping down an AC voltage, and to match impedances between a generator and a load. (Chapter 10, introduction)

69. See Figure 12.18 .

Figure 12.18

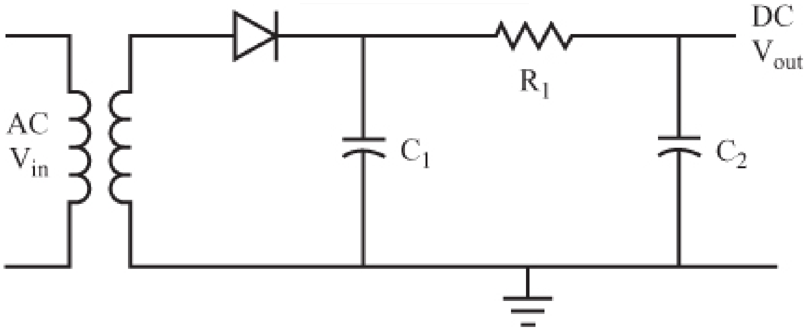

(Chapter 11, problem 14)

70. See Figure 12.19 .

Figure 12.19

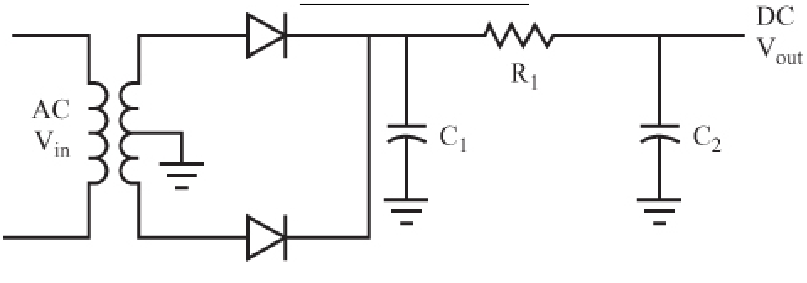

(Chapter 11, problem 31)

71. R_1 = 84 ohms, C_1 = 622 μF, C_2 = 265 μF (Chapter 11, problems 26–29)

Appendix A

Glossary

Ampere (A)

The unit of measurement of electric current.

Amplifier

Electronic device or circuit that produces an output signal with greater power, voltage, or current than that provided by the input signal.

Capacitance (C)

The capability of a component to store an electric charge when voltage is applied across the component, measured in farads.

Capacitor

A component that stores electric charge when voltage is applied to it. It can return the charge to a circuit in the form of an electric current when the voltage is removed.

Current (I)

The amount of electric charge flowing through a circuit, measured in amperes.

Diode

A component that conducts current in one direction only.

Discrete components

Individual electronic parts such as resistors, diodes, capacitors, and transistors.

Farad (F)

The unit of measurement of capacitance.

Feedback

A connection from the output of an amplifier back to the input. In some instances, a portion of the output voltage is used to control, stabilize, or modify the operation of the amplifier. However, in some instances, unwanted feedback can cause a squealing noise or can cancel another signal.

Filter

Electronic circuit that can either block or pass frequencies that fall within a certain range.

Frequency (f)

Number of cycles of a waveform that occurs in a given time period, measured in hertz (cycles per second).

Ground

Zero volts. This is the arbitrary reference point in a circuit from which all voltage measurements are made.

Henry (H)

The unit of measurement of inductance.

Impedance (Z)

Total opposition (resistance and reactance) of a circuit to alternating current (AC) flow, measured in ohms.

Inductance (L)

The property of a component that opposes any change in an existing current, measured in henrys.

Inductor

A coil of wire whose magnetic field opposes changes in current flow when the voltage across the coil is changed.

Integrated circuit (IC)

Electronic component in the form of a small silicon chip in which numerous transistors and other components have been built to form a circuit.

Kirchhoff's laws

A set of formulas that form the basis for direct current (DC) and alternating current (AC) circuit analysis. This includes *Kirchhoff's current law (KCL)* , which states that the sum of all currents at a junction equals zero; and *Kirchhoff's voltage law (KVL)* , which states that the sum of all voltages in a loop equals zero.

Ohm (Ω)

The unit of measurement of resistance.

Ohm's law

A formula used to calculate the relationship between voltage, current, and resistance, expressed as $V= IR$. Also expressed as $E = IR$.

Operational amplifier (op-amp)

An integrated circuit, multistage amplifier. An op-amp is much smaller and, therefore, more practical than an equivalent amplifier made with discrete components.

Oscillator

An electronic circuit that produces a continuous output signal such as a sine wave or square wave.

Phase angle

For a signal, the angle of lead or lag between the current waveform and the voltage waveform, expressed in degrees.

Phase shift

The change in a phase of a signal as it passes through a circuit, such as in an amplifier.

Pinout

The configuration of leads used to connect an electronic component to a circuit.

Power

The expenditure of energy over time, measured in watts.

Reactance (X)

The degree of opposition of a component to the flow of alternating current (AC), measured in ohms. There are two types of reactance: capacitive reactance (X_C) exhibited by capacitors and inductive reactance (X_L) exhibited by inductors.

Rectification

The process to change alternating current (AC) to direct current (DC).

Resistance (R)

The degree of opposition of a component to the flow of electric current, measured in ohms.

Resistor

A component whose value is determined by the amount of opposition it has to the flow of electric current.

Resonance frequency

The frequency at which the reactance of a capacitor and inductor connected together in a circuit are equal.

Saturated transistor

A transistor that is completely turned on.

Semiconductor

A material that has electrical characteristics of a conductor or an insulator, depending on how it is treated. Silicon is the semiconductor material most commonly used in electronic components.

Transformer

A component that transforms an input alternating current (AC) voltage to either a higher level (step-up transformer) or a lower level (step-down transformer) AC voltage.

Transistor, BJT

A bipolar junction transistor (BJT) is a semiconductor component that can either be used as a switch or an amplifier. In either case, a small input signal controls the transistor, producing a much larger output signal.

Transistor, JFET

A junction field effect transistor (JFET), which, like the bipolar junction transistor, can be used either as a switch or an amplifier.

Transistor, MOSFET

Like the BJT and JFET, a metal oxide silicon field effect transistor (MOSFET) that can be used either as a switch or an amplifier. The

MOSFET is the most commonly used transistor in integrated circuits.

Turns ratio (TR)

The ratio of the number of turns in the primary or input winding of a transformer to the number of turns in the secondary (or output) winding.

Volt (V)

The unit of measurement for the potential difference that causes a current to flow through a conductor.

Voltage (V)

The potential difference of energy that, when applied to an electrical circuit, causes current to flow, measured in volts.

Watt (W)

Unit of electric power dissipated as heat when 1 amp of current flows through a component that has 1 volt applied across it.

Zener

A particular type of diode that enables the flow of current at a definite reverse-bias voltage level.

Appendix B

List of Symbols and Abbreviations

The following table lists common symbols and abbreviations.

Symbol/Abbreviation	Meaning
A	Ampere
AC	Alternating current
A_{pp}	Peak-to-peak amperes
A_{rms}	Root mean square amperes
A_V	AC voltage gain
β (beta)	Current gain
BW	Bandwidth
C	Capacitor
DC	Direct current
F	Farad
g_m	Transconductance
f	Frequency
f_r	Resonant frequency
H	Henry
Hz	Hertz
I	Electric current
I_B	Base current
I_C	Collector current
I_D	Drain current of a field effect transistor (FET); also current through a diode
I_{DSS}	Saturation current
I_{pp}	Peak-to-peak current
I_{rms}	Root mean square current
kHz	Kilohertz
kΩ	Kilohm
kW	Kilowatt

L	Inductor
LC	Inductor-capacitor circuit
mA	Milliampere
mH	Millihenry
MΩ	Megohm
μA	Microampere
μF	Microfarad
μH	Microhenry
μs	Microsecond
ms	Millisecond
mV	Millivolt
N_p	Number of turns in a primary coil
N_S	Number of turns in a secondary coil
Ω	Ohms
pF	Picofarad
P	Power
Q	Transistor; also the Q value of a resonant circuit
R	Resistor
R_{in}	Input resistance of a transistor
r	DC resistance of an inductor
T	Period of a waveform
τ	Time constant
TR	Turns ratio
θ	Phase angle
μV	Microvolt
V	Voltage
V_C	Voltage at the collector of a transistor
V_{DD}	Drain supply voltage
V_E	Voltage at the emitter of a transistor
V_{GG}	Gate supply voltage

V_{GS} Gate to source voltage
$V_{GS(off)}$ Gate to source cutoff voltage
V_{in} AC voltage of an input signal
V_{out} AC output voltage
V_p Peak voltage
V_{pp} Peak-to-peak voltage
V_{rms} Root mean square voltage
V_S Supply voltage
W Watts
X_C Reactance of a capacitor
X_L Reactance of an inductor
Z Impedance

Appendix C

Powers of Ten and Engineering Prefixes

The following table shows powers of the number 10, decimal equivalents, prefixes used to denote the value, symbols used, and typical usages.

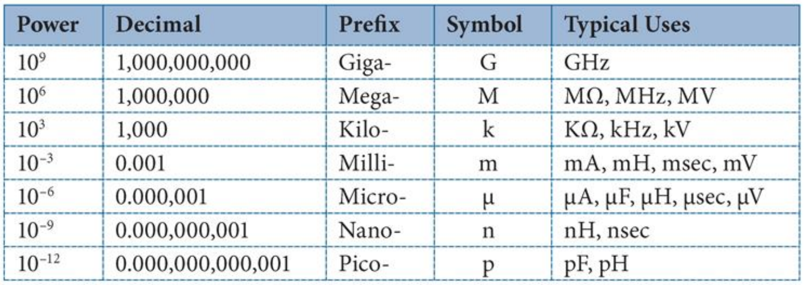

Power	Decimal	Prefix	Symbol	Typical Uses
10^9	1,000,000,000	Giga-	G	GHz
10^6	1,000,000	Mega-	M	MΩ, MHz, MV
10^3	1,000	Kilo-	k	KΩ, kHz, kV
10^{-3}	0.001	Milli-	m	mA, mH, msec, mV
10^{-6}	0.000,001	Micro-	μ	μA, μF, μH, μsec, μV
10^{-9}	0.000,000,001	Nano-	n	nH, nsec
10^{-12}	0.000,000,000,001	Pico-	p	pF, pH

Appendix D

Standard Composition Resistor Values

The most commonly used type of resistor is the carbon film resistor with a ± 5 percent tolerance and either a 1/4 or 1/2 watt power rating. The standard resistance values for this type of resistor are listed in the following table (in ohms). You should purchase resistors at any of these values through the online distributors listed in Appendix E, "Supplemental Resources." Power resistors are available with fewer resistance values, which you can find in the catalogs of various suppliers.

Note In the following table, "k" represents kilo-ohms, so 7.5 k translates into 7,500 ohms. Similarly, "M" stands for megohms, so a value of 3.6 M represents 3,600,000 ohms.

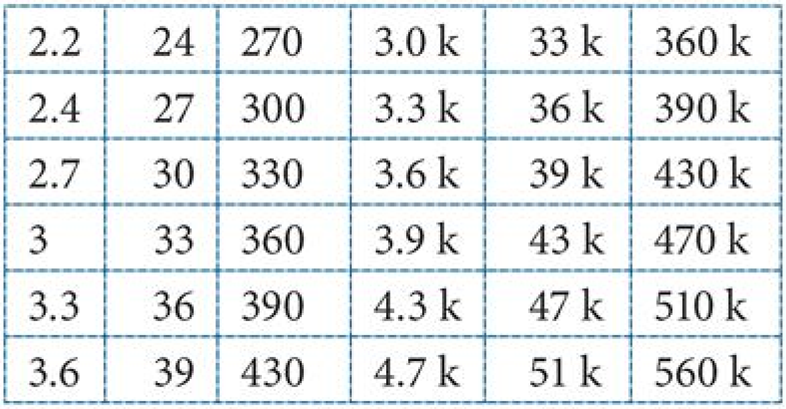

2.2	24	270	3.0 k	33 k	360 k
2.4	27	300	3.3 k	36 k	390 k
2.7	30	330	3.6 k	39 k	430 k
3	33	360	3.9 k	43 k	470 k
3.3	36	390	4.3 k	47 k	510 k
3.6	39	430	4.7 k	51 k	560 k

3.9	43	470	5.1 k	56 k	620 k
4.3	47	510	5.6 k	62 k	680 k
4.7	51	560	6.2 k	68 k	750 k
5.1	56	620	6.8 k	75 k	820 k
5.6	62	680	7.5 k	82 k	910 k
6.2	68	750	8.2 k	91 k	1.0 M
6.8	75	820	9.1 k	100 k	1.2 M
7.5	82	910	10 k	110 k	1.5 M
8.2	91	1.0 k	11 k	120 k	1.8 M
9.1	100	1.1 k	12 k	130 k	2.2 M
10	110	1.2 k	13 k	150 k	2.4 M
11	120	1.3 k	15 k	160 k	2.7 M
12	130	1.5 k	16 k	180 k	3.3 M
13	150	1.6 k	18 k	200 k	3.6 M
15	160	1.8 k	20 k	220 k	3.9 M
16	180	2.0 k	22 k	240 k	4.7 M
18	200	2.2 k	24 k	270 k	5.6 M
20	220	2.4 k	27 k	300 k	6.8 M
22	240	2.7 k	30 k	330 k	8.2 M

Appendix E

Supplemental Resources

This appendix provides a list of websites, books, magazines, tutorials, and electronics suppliers that should be of interest if you want more knowledge about basic electronics concepts, reference material for circuit design, or the supplies needed to build circuits.

Web Sites

Following are some websites that may prove useful:

- *Building Gadgets* (www.buildinggadgets.com/) — This is an electronics reference site maintained by Earl Boysen (one of the authors of this book). There are lots of handy links to electronics tutorials, discussion forums, suppliers, and interesting electronics projects here. This website includes support web pages for each project in the *Complete Electronics Self-Teaching Guide,* including the following:
  - Data sheets for key components.
  - Parts lists with links to the appropriate locations in online catalogues for suppliers.
  - Color photos showing details of circuit construction and testing.
- *All About Circuits* (www.allaboutcircuits.com/) — This site includes an online book on electronics theory and circuits, as well as discussion forums on electronics projects,

microcontrollers, and general electronics issues.

Williamson Labs (www.williamson-labs.com/) — This site includes tutorials on electronics components and circuits. Many of these tutorials include animated illustrations that can help you understand how each circuit functions.

Electro Tech online (www.electro-tech-online.com/) — This is a discussion forum on electronics projects and general electronics issues.

Electronics Lab (www.electronics-lab.com/index.html) — This is a collection of a few hundred interesting electronics projects.

Discover Circuits (www.discovercircuits.com/) — This is a collection of thousands of electronic circuits.

Books

Following are some books that may prove useful:

Electronics For Dummies, Second Edition , *by Cathleen Shamieh and Gordon McComb (Indianapolis: Wiley, 2009)—* This is a good book to start with. It provides an introduction to electronics concepts, components, circuits, and methods.

The Art of Electronics, Second Edition , *by Paul Horowitz and Winfield Hill (New York:*

Cambridge University Press, 1989)— This is a useful reference book for designing circuits, as well as understanding the functionality of existing circuits.

ARRL Handbook for Radio Communications 2012 *(Newington, Connecticut: American Radio Relay League, 2012)*— Although this is intended for ham radio enthusiasts, it is also a useful reference book for understanding circuit design. This handbook is updated every year, but if you buy the current edition, it should be good for several years.

Magazines

Following are some magazines that may prove useful:

Everyday Practical Electronics Magazine (www.epemag3.com/) — This magazine provides interesting detailed electronics projects for hobbyists.

Nuts and Volts Magazine (www.nutsvolts.com/) — This magazine provides information on new components for hobbyists and projects, focusing on circuits using microcontrollers.

EDN Magazine (www.edn.com/) — This magazine includes articles on new components/designs for the engineering community.

Suppliers

This section shows retail stores and online distributors.

Retail Stores

Following are some retail stores that may prove useful:

- *Radio Shack* (www.radioshack.com/) — This retail chain carries electronic components and has stores in most U.S. cities.
- *Fry's Electronics* (www.frys.com/) — This retail chain carries electronics components and has stores in nine states.

Online Distributors

Following are some online distributors that may prove useful:

- *Jameco Electronics* (www.jameco.com/) — This is a medium-sized distributor that carries most of the components you'll need, with a reasonably sized catalog that you'll find easy to use to find components.
- *Mouser Electronics* (www.mouser.com/) — This is a large distributor that carries a wide range of components with a nice ordering system on its website that lets you put together separate orders for different projects, which is handy if you're planning multiple projects. This distributor also does a good job of packaging, clearly labeling

components for shipment.

Digi-key (www.digikey.com/) — This is another large distributor with a broad selection of components. Digi-key may carry components that are difficult to find at smaller suppliers.

Appendix F

Equation Reference

The following table provides a quick reference to common equations.

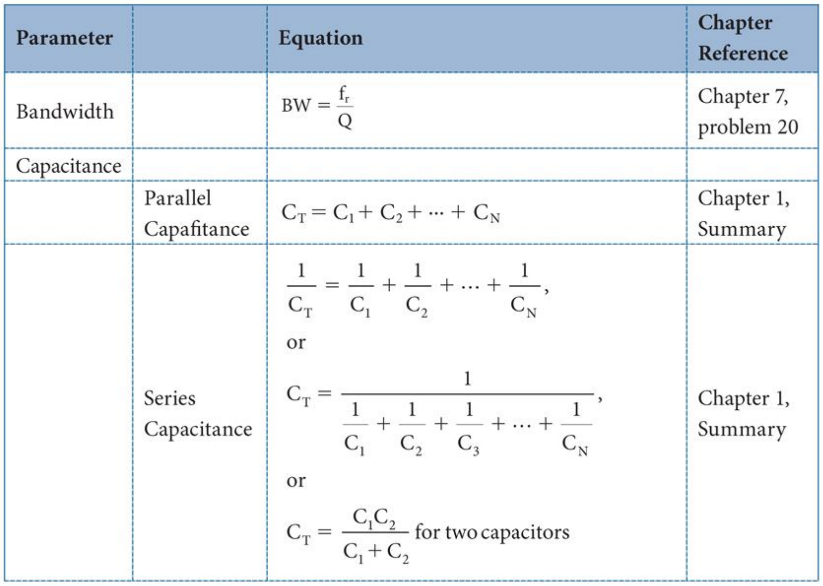

Parameter		Equation	Chapter Reference
Bandwidth		$BW = \frac{f_r}{Q}$	Chapter 7, problem 20
Capacitance			
	Parallel Capafitance	$C_T = C_1 + C_2 + \cdots + C_N$	Chapter 1, Summary
	Series Capacitance	$\frac{1}{C_T} = \frac{1}{C_1} + \frac{1}{C_2} + \cdots + \frac{1}{C_N}$, or $C_T = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots + \frac{1}{C_N}}$, or $C_T = \frac{C_1 C_2}{C_1 + C_2}$ for two capacitors	Chapter 1, Summary

Parameter		Equation	Chapter Reference
Frequency		$f = \frac{1}{T}$	Chapter 5, problem 7
	Resonance Frequency (series LC circuit)	$f_r = \frac{1}{2\pi\sqrt{LC}}$	Chapter 7, problem 6
	Resonance Frequency (parallel LC circuit)	$f_r = \frac{1}{2\pi\sqrt{LC}}\sqrt{1 - \frac{r^2C}{L}}$ if Q is less than 10, or $f_r = \frac{1}{2\pi\sqrt{LC}}$, if $Q \geq 10$	Chapter 7, problem 10
Gain			
	Voltage Gain	$A_v = \frac{V_{out}}{V_{in}}$	Chapter 8, problem 9
	Current Gain	$\beta = \frac{I_C}{I_B}$	Chapter 3, problem 17
Impedance		$Z = \sqrt{X_C{}^2 + R^2}$	Chapter 6, problem 8
Phase Shift			
	Phase Angle (RC circuit)	$\tan\theta = \frac{X_C}{R} = \frac{1}{2\pi fRC}$	Chapter 6, problem 23
	Phase Angle (LC circuit)	$\tan\theta = \frac{X_L}{R} = \frac{2\pi fL}{R}$	Chapter 6, problem 35
Q Value		$Q = \frac{X_L}{R}$	Chapter 7, problem 20

Parameter		**Equation**	**Chapter Reference**
Resistance			
	Parallel Resistance	$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots + = \frac{1}{R_N}$, or $R_T = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots + = \frac{1}{R_N}}$, or $RT = \frac{R_1 R_2}{R_1 + R_2}$ for two resistors	Chapter 1, Summary
	Series Resistance	$R_T = R_1 + R_2 + \cdots + R_N$	Chapter 1, Summary
Power		$P = VI$, or $P = I^2R$, or $P = \frac{V^2}{R}$	Chapter 1, Summary
Reactance			
	Capacitive Reactance	$X_C = \frac{1}{2\pi fC}$	Chapter 5, problem 13
	Inductive Reactance	$X_L = 2\pi fL$	Chapter 5, problem 16
Time Constant		$\tau = RC$	Chapter 1, Summary
Turns Ratio		$TR = \frac{N_P}{N_S}$	Chapter 10, problem 6

Parameter		Equation	Chapter Reference
Voltage			
	Ohm's law (DC)	$V = IR$	Chapter 1, Summary
	Ohm's law (AC)	$V = IZ$	Chapter 6, problem 8
	Voltage divider	$V_1 = \frac{V_S R_1}{R_T}$	Chapter 1, Summary
	Peak-to-Peak Voltage (sine wave)	$V_{pp} = 2V_p = 2 \times \sqrt{2} \times V_{rms}$ $= 2.828 \times V_{rms}$	Chapter 5, problem 4
	RMS Voltage (sine wave)	$V_{rms} = \frac{1}{\sqrt{2}} \times V_p = \frac{1}{\sqrt{2}} \times \frac{V_{pp}}{2}$ $= 0.707 \times \frac{V_{pp}}{2}$	Chapter 5, problem 4
	Transformer output voltage	$V_{out} = \frac{V_{in} N_s}{N_p}$	Chapter 10, problem 7

Appendix G

Schematic Symbols Used in This Book

The following table shows schematic symbols used in this book.

Component		Symbol
Battery		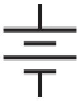
Buzzer		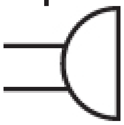
Capacitors	Capacitor	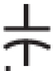
	Variable Capacitor	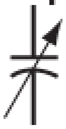
Diodes	Diode	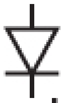
	Zener Diode	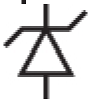
	Light Emitting Diode	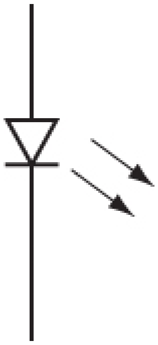
	Photodiode	

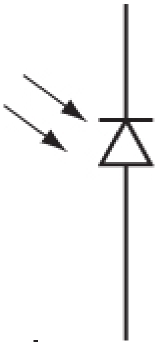

Generator (DC)

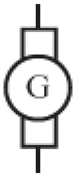

Ground

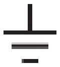

Inductors Inductor

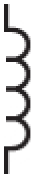

Tapped Inductor

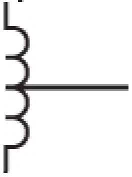

Lamp

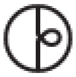

Meters Meter

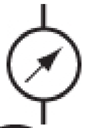

Ammeter

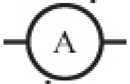

Voltmeter

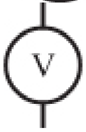

Motor

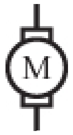

Operational Amplifier

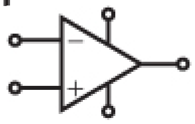

Resistors

Resistor

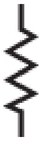

Two-Contact Variable Resistor (Potentiometer)

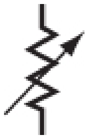

Three-Contact Variable Resistor (Potentiometer)

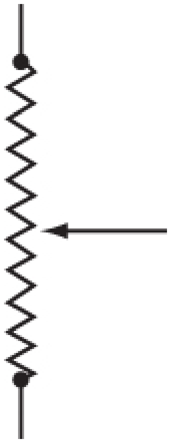

Signal Generator (Sine Wave)

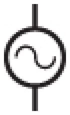

Switches

Single-pole, Single-Throw (SPST) Switch

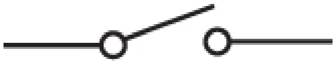

Single-pole, Double-Throw (SPDT) Switch

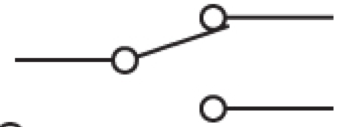

Test Point

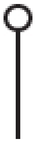

Transformers Transformer

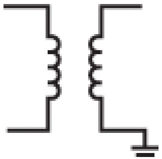

Center Tap Transformer

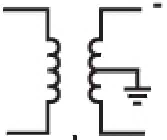

Transistors NPN BJT

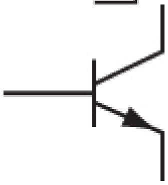

PNP BJT

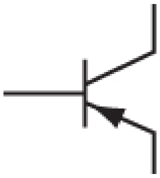

N-Channel JFET

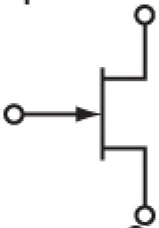

P-Channel JFET

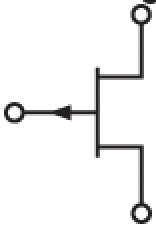

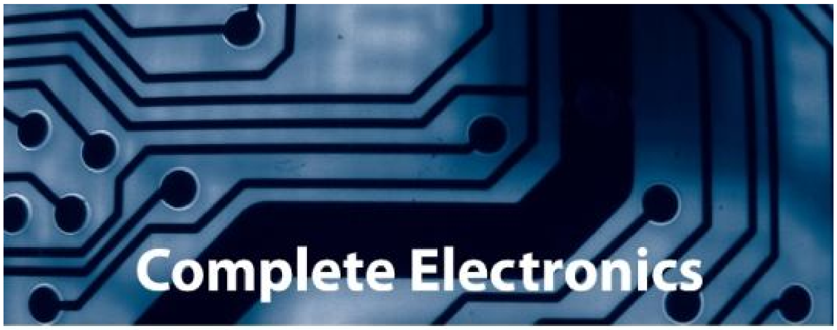

Complete Electronics

SELF-TEACHING GUIDE WITH PROJECTS

Earl Boysen | Harry Kybett

WILEY
John Wiley & Sons, Inc.

Complete Electronics
Executive Editor: Carol Long
Project Editor: Kevin Shafer
Technical Editor: Rex Miller
Production Editor: Kathleen Wisor
Copy Editor: San Dee Phillips
Editorial Manager: Mary Beth Wakefield
Freelancer Editorial Manager: Rosemarie Graham
Associate Director of Marketing: David Mayhew
Marketing Manager: Ashley Zurcher
Business Manager: Amy Knies
Production Manager: Tim Tate
Vice President and Executive Group Publisher: Richard Swadley
Vice President and Executive Publisher: Neil Edde
Associate Publisher: Jim Minatel
Project Coordinator, Cover: Katie Crocker
Proofreader: Nancy Carrasco
Indexer: Jack Lewis
Cover Image: Earl Boysen
Cover Designer: Ryan Sneed
Published by
John Wiley & Sons, Inc.
10475 Crosspoint Boulevard
Indianapolis, IN 46256
www.wiley.com
Copyright © 2012 by Earl Boyson and Harry Kybett

Published by John Wiley & Sons, Inc., Indianapolis, Indiana
Published simultaneously in Canada
ISBN: 978-1-118-21732-0 ISBN: 978-1-118-28232-8 (ebk) ISBN: 978-1-118-28319-6 (ebk) ISBN: 978-1-118-28469-8 (ebk)

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http:// http://www.wiley.com/go/permissions.

Limit of Liability/Disclaimer of Warranty: The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation warranties of fitness for a particular purpose. No warranty may be created or extended by sales or promotional materials. The advice and strategies contained herein may not be suitable for every situation. This work is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If professional assistance is required, the services of a competent professional person should be sought. Neither the publisher nor the author shall be liable for damages arising herefrom. The fact that an organization or Web site is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or website may provide or recommendations it may make. Further, readers should be aware that Internet websites listed in this work may have changed or disappeared between when this work was written and when it is read. For general information on our other products and services please contact our Customer Care Department within the United States at (877) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002. Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at http://booksupport.wiley.com. For more information about Wiley products, visit www.wiley.com.

Library of Congress Control Number: 2012940030

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates, in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

To my science and engineering teachers. I'd particularly like to thank Jim Giovando, my physics and chemistry teacher at Petaluma Senior High School, who, even decades later, I remember as having been an inspiration. I also dedicate this book to the physics and chemistry faculty of Sonoma State University in the 1970s, where the small class size and personal guidance by the professors made for a great learning environment.

About the Author

Earl Boysen spent 20 years as an engineer in the semiconductor industry, and currently runs two websites, BuildingGadgets.com (dedicated to electronics) and UnderstandingNano.com (covering nanotechnology topics). Boysen holds a Masters degree in Engineering Physics from the University of Virginia. He is the co-author of three other books: *Electronics Projects For Dummies* (Indianapolis: Wiley, 2006), *Nanotechnology For Dummies* (Indianapolis: Wiley, 2011), and the first edition of *Electronics For Dummies* (Indianapolis: Wiley, 2005). He lives with his wonderful wife, Nancy, and two cats.

About the Technical Editor

Rex Miller was a Professor of Industrial Technology at The State University of New York, College at Buffalo for more than 35 years. He has taught on the technical school, high school, and college level for more than 40 years. He is the author or co-author of more than 100 textbooks ranging from electronics through carpentry and sheet metal work. He has contributed more than 50 magazine articles over the years to technical publications. He is also the author of seven civil war regimental histories.

Acknowledgments

I want to first thank Harry Kybett for authoring the original version of this book many years ago. It's an honor to take over such a classic book in the electronics field. Thanks also to Carol Long for bringing me on board with the project, and Kevin Shafer for his able project management of the book. My appreciation to Rex Miller for his excellent technical editing, and to San Dee Phillips for handling all the mechanics of spelling and grammar in a thorough copy edit. Finally, thanks to my wonderful wife, Nancy, for her advice and support throughout the writing of this book.

Introduction

The rapid growth of modern electronics is truly a phenomenon. Electronic devices (including cell phones, personal computers, portable MP3 players, and digital cameras) are a big part of many of our daily lives. Many industries have been founded, and older industries have been revamped, because of the availability and application of modern electronics in manufacturing processes, as well as in electronic products. Electronic products are constantly evolving, and their impact on our lives, and even the way we socialize, is substantial.

What This Book Teaches

Complete Electronics Self-Teaching Guide with Projects is for anyone who has a basic understanding of electronics concepts and wants to understand the operation of components found in the most common discrete circuits. The chapters focus on circuits that are the building blocks for many common electronic devices, and on the very few (but important) principles you need to know to work with electronics.

The arrangement and approach is completely different from any other book on electronics in that it uses a question-and-answer approach to help you understand how electronic circuits work. This book steps you through calculations for every example in an easy-to-understand fashion, and you do not need to have a mathematical background beyond first-year algebra to follow along.

For many of you, the best way to understand new concepts is by doing, rather than reading or listening. This book reinforces your understanding of electronic concepts by leading you through the calculations and concepts for key circuits, as well as the construction of circuits. Projects interspersed throughout the material enable you to get hands-on practice. You build many of the circuits and observe or measure how they work.

Helpful sidebars are interspersed throughout the book to provide more information about how components work, and how to choose the right component. Other sidebars provide discussions of techniques for building and testing circuits. If you want this additional information, be sure to read these.

Understanding the circuits composed of discrete components and the applicable calculations discussed is useful not only in building and designing circuits, but it also helps you to work with integrated circuits (ICs). That's because ICs use miniaturized components (such as transistors, diodes, capacitors, and resistors) that function based on the same rules as discrete components (along with some specific rules necessitated by the extremely small size of IC components).

How This Book Is Organized

This book is organized with sets of problems that challenge you to think through a concept or procedure, and then provides answers so that you can constantly check your progress and understanding. Specifically, the chapters in this book are organized as follows:

Chapter 1 DC Review and Pre-Test—This chapter provides a review and pre-test on the basic concepts, components, and calculations that are useful when working with direct current (DC) circuits.

Chapter 2 The Diode—This chapter teaches you about the diode, including how you use diodes in DC circuits, the main characteristics of diodes, and calculations you can use to determine current, voltage, and power.

Chapter 3 Introduction to the Transistor—This chapter explores the transistor and how it's used in circuits. You also discover how bipolar junction transistors (BJTs) and junction field effect transistors (JFETs) control the flow of electric current.

Chapter 4 The Transistor Switch—This chapter examines the simplest and most widespread application of the transistor:

switching. In addition to learning how to design a transistor circuit to drive a particular load, you also compare the switching action of a JFET and a BJT.

Chapter 5 AC Pre-Test and Review—This chapter examines the basic concepts and equations for alternating current (AC) circuits. You discover how to use resistors and capacitors in AC circuits, and learn related calculations.

Chapter 6 Filters—This chapter looks at how resistors, capacitors, and inductors are used in high-pass filters and low-pass filters to pass or block AC signals above or below a certain frequency.

Chapter 7 Resonant Circuits—This chapter examines the use of capacitors, inductors, and resistors in bandpass filters and band-reject filters to pass or block AC signals in a band of frequencies. You also learn how to calculate the resonance frequency and bandwidth of these circuits. This chapter also introduces the use of resonant circuits in oscillators.

Chapter 8 Transistor Amplifiers—This chapter explores the use of transistor amplifiers to amplify electrical signals. In addition to examining the fundamental steps used to design BJT-based amplifiers, you learn how to use JFETs and operational amplifiers (op-amps) in amplifier circuits.

Chapter 9 Oscillators—This chapter introduces you to the oscillator, a circuit that produces a continuous AC output signal. You learn how an oscillator works and step through the procedure to design and build an oscillator.

Chapter 10 The Transformer—This chapter discusses how a transformer converts AC voltage to a higher or lower voltage. You learn how a transformer makes this conversion and how to calculate the resulting output voltage.

Chapter 11 Power Supply Circuits—This chapter examines how power supplies convert AC to DC with a circuit made up of transformers, diodes, capacitors, and resistors. You also learn how to calculate the values of components that produce a specified DC output voltage for a power supply circuit.

Chapter 12 Conclusion and Final Self-Test—This chapter enables you to check your overall knowledge of electronics concepts presented in this book through the use of a final self-test.

In addition, this book contains the following appendixes for easy reference:

Appendix A Glossary—This appendix provides key electronics terms and their definitions.

Appendix B List of Symbols and Abbreviations—This appendix gives you a handy reference of commonly used symbols and abbreviations.

Appendix C Powers of Ten and Engineering Prefixes—This appendix lists prefixes commonly used in electronics, along with their corresponding values.

Appendix D Standard Resistor Values—This appendix provides standard resistance values for the carbon film resistor, the most commonly used type of resistor.

Appendix E Supplemental Resources—This appendix provides references to helpful websites, books, and magazines.

Appendix F Equation Reference—This appendix serves as a quick guide to commonly used equations, along with chapter and problem references showing you where they are first introduced in this book.

Appendix G Schematic Symbols Used in This Book—This appendix provides a listing of schematic symbols used in the problems found throughout the book.

Conventions Used in This Book

As you study electronics, you will find that there is some variation in terminology and the way that circuits are drawn. Following are three conventions followed in this book that you should be aware of:

- The discussions use “V” to stand for voltage, versus “E,” which you see used in some other books.

In all circuit diagrams, intersecting lines indicate an electrical connection. (Some other books use a dot at the intersection of lines to indicate a connection.) If a semicircle appears at the intersection of two lines, it indicates that there is no connection. See Figure 9.5 for an example of this.

The discussions in this book use conventional current flow to determine the flow of electric current (from positive voltage to negative voltage), whereas some other books use electron flow (from negative voltage to positive voltage).

How to Use This Book

This book assumes that you have some knowledge of basic electronics such as Ohm's law and current flow. If you have read a textbook or taken a course on electronics, or if you have worked with electronics, you probably have the prerequisite knowledge. If not, you should read a book such as *Electronics for Dummies* (Indianapolis: Wiley, 2009) to get the necessary background for this book. You can also go to the author's Website (www.BuildingGadgets.com) and use the Tutorial links to find useful online lessons in electronics. In addition, Chapters 1 and 5 enable you to test your knowledge and review the necessary basics of electronics.

You should read the chapters in order because often later material depends on concepts and skills covered in earlier chapters.

Complete Electronics Self-Teaching Guide with Projects is presented in a self-teaching format that enables you to learn easily, and at your own pace. The material is presented in numbered sections called *problems* . Each problem presents some new information and gives you questions to answer. To learn most effectively, you should cover up the answers with a sheet of paper and try to answer each question. Then, compare your answer with the correct answer that follows. If you miss a question, correct your answer and then go on. If you miss many in a row, go back and review the previous section, or you may miss the point of the material that follows.

Be sure to try to do all the projects. They are not difficult, and they help reinforce your learning of the subject matter. If you don't have the equipment to work through a project, simply reading through it can help you to better understand the concepts it demonstrates.

Each project includes a schematic, parts list, step-by-step instructions, and detailed photos of the completed circuit. Working through these projects, you can test your skill by building the circuit using just the schematic and parts list. If you want additional help, check the photos showing the details of how the components are connected. A Camera icon in the margin as shown here indicates that there is a color version of the figure in a special insert in the paperback version of this book. If you purchased an electronic version of this book, and have an e-reader without color capabilities, you can find the color photos on the author's website at www.buildinggadgets.com/complete-electronics.htm.

This website also provides project pages that include links to suppliers. These pages are kept up-to-date with supplier part numbers for the components you need.

When you reach the end of a chapter, evaluate your learning by taking the Self-Test. If you miss any questions, review the related parts of the chapter again. If you do well on the Self-Test, you're ready to go to the next chapter. You may also find the Self-Test useful as a review before you start the next chapter. At the end of the book, there is a Final Self-Test that enables you to assess your overall learning.

You can work through this book alone, or you can use it with a course. If you use the book alone, it serves as an introduction to electronics but is not a complete course. For that reason, at the end of the book are some suggestions for further reading and online resources. Also, at the back of the book is a table of symbols and abbreviations for reference and review.

Now you're ready to learn *electronics* !

Teach Yourself®

COMPLETE MATHEMATICS

Figure out numbers (page 1) Measure and draw angles (page 18) Add and subtract fractions (page 34) Construct triangles (page 48) Convert between decimals and fractions (page 66) Collect statistical data (page 81) Draw a pictogram, pie chart or line graph (page 82) Use coordinates (page 113) Understand the metric system (page 125) Find and use the areas of many shapes (page 136) Simplify algebraic expressions (page 151) Solve linear equations (page 183) Write formulae (page 209) Use ratio (page 265) Pythagoras' theorem (page 273) Mode, median and mean (page 306) Factorize quadratic expressions (page 345) Learn more online (www.teachyourself.com)

PACKED WITH EXAMPLES AND EXERCISES

Complete Mathematics

This page intentionally left blank

Teach
Yourself®

Complete Mathematics

Trevor Johnson and
Hugh Neill

For UK order enquiries: please contact Bookpoint Ltd,
130 Milton Park, Abingdon, Oxon OX14 4SB.
Telephone: +44 (0) 1235 827720. Fax: +44 (0) 1235 400454.
Lines are open 09.00–17.00, Monday to Saturday, with a 24-hour message answering service. Details about our titles and how to order are available at www.teachyourself.com

For USA order enquiries: please contact McGraw-Hill Customer Services, PO Box 545, Blacklick, OH 43004-0545, USA.
Telephone: 1-800-722-4726. Fax: 1-614-755-5645.

For Canada order enquiries: please contact McGraw-Hill Ryerson Ltd, 300 Water St, Whitby, Ontario L1N 9B6, Canada.
Telephone: 905 430 5000. Fax: 905 430 5020.

Long renowned as the authoritative source for self-guided learning – with more than 50 million copies sold worldwide – the Teach Yourself series includes over 500 titles in the fields of languages, crafts, hobbies, business, computing and education.

British Library Cataloguing in Publication Data: a catalogue record for this title is available from the British Library.

Library of Congress Catalog Card Number: on file.

First published in UK 2001 by Hodder Education, part of Hachette UK, 338 Euston Road, London NW1 3BH.

First published in US 2001 by The McGraw-Hill Companies, Inc.

This edition published 2010.

Previously published as Teach Yourself Mathematics.

The Teach Yourself name is a registered trade mark of Hodder Headline.

Copyright © 2001, 2003, 2008, 2010 Hugh Neill and Trevor Johnson

In UK: All rights reserved. Apart from any permitted use under UK copyright law, no part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information, storage and retrieval system, without permission in writing from the publisher or under licence from the Copyright Licensing Agency Limited. Further details of such licences (for reprographic reproduction) may be obtained from the Copyright Licensing Agency Limited, of Saffron House, 6–10 Kirby Street, London EC1N 8TS.

In US: All rights reserved. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

Typeset by MPS Limited, A Macmillan Company.

Printed in Great Britain for Hodder Education, an Hachette UK Company, 338 Euston Road, London NW1 3BH, by CPI Cox & Wyman, Reading, Berkshire RG1 8EX.

The publisher has used its best endeavours to ensure that the URLs for external websites referred to in this book are correct and active at the time of going to press. However, the publisher and the author have no responsibility for the websites and can make no guarantee that a site will remain live or that the content will remain relevant, decent or appropriate.

Hachette UK's policy is to use papers that are natural, renewable and recyclable products and made from wood grown in sustainable forests. The logging and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

Impression number 10 9 8 7 6 5 4 3 2 1
Year 2014 2013 2012 2011 2010

Contents

Personal introduction **xi**
Only got a minute? **xii**
Only got five minutes? **xiv**
Only got ten minutes? **xvi**
1 Number **1**
  1.1 Introduction – place value 1
  1.2 Arithmetic – the four operations 3
  1.3 Order of operations 4
  1.4 Problems which use arithmetic 7
  1.5 Special numbers 8
  1.6 Multiples, factors and primes 10
2 Angles **15**
  2.1 Introduction 15
  2.2 Angles 16
  2.3 Measuring and drawing angles 18
  2.4 Using angle facts 20
  2.5 Parallel lines 22
  2.6 Bearings 25
3 Fractions **29**
  3.1 Introduction 29
  3.2 What is a fraction? 29
  3.3 Which fraction is bigger? 31
  3.4 Simplifying fractions 31
  3.5 Improper fractions 33
  3.6 Adding and subtracting fractions 34
  3.7 Multiplication of fractions 36
  3.8 Fractions of quantities 37
  3.9 Division of fractions 38
  3.10 A number as a fraction of another number 40
4 Two-dimensional shapes **44**
  4.1 Introduction 44
  4.2 Triangles 44
  4.3 Constructing triangles 48

4.4 Quadrilaterals 50
4.5 Polygons 52
4.6 Interior and exterior angles 53
4.7 Symmetries of regular polygons 55
4.8 Congruent shapes 55
4.9 Tessellations 57

5 Decimals 63
5.1 Introduction 63
5.2 Place value 64
5.3 Converting decimals to fractions 66
5.4 Converting fractions to decimals 1 67
5.5 Addition and subtraction 68
5.6 Multiplication of decimals 69
5.7 Division of decimals 72
5.8 Converting fractions to decimals 2 74

6 Statistics 1 80
6.1 Introduction 80
6.2 Collection of data 81
6.3 Pictograms 82
6.4 Bar charts 83
6.5 Pie charts 87
6.6 Line graphs 89
6.7 Scatter graphs 93
6.8 Discrete and continuous data 96
6.9 Grouping data 97

7 Directed numbers 104
7.1 Introduction 104
7.2 Ordering directed numbers 105
7.3 Addition and subtraction 106
7.4 Multiplication and division 108
7.5 Using a calculator 111

8 Graphs 1 113
8.1 Coordinates 113
8.2 Straight-line graphs 116
8.3 Lines parallel to the axes 121

9 Measurement 125
9.1 The metric system 125
9.2 Imperial units 127

9.3 Converting between metric and imperial units 129
9.4 Choosing suitable units 130
10 Perimeter and area 134
10.1 Perimeter 134
10.2 Area 136
10.3 Area of a rectangle 138
10.4 Area of a parallelogram 141
10.5 Area of a triangle 141
10.6 Area of a trapezium 144
11 Algebraic expressions 148
11.1 Introduction – what is algebra? 148
11.2 Writing expressions 148
11.3 Simplifying expressions 151
11.4 Evaluating expressions 153
11.5 Squaring 155
11.6 Brackets 155
11.7 Factorizing expressions 158
11.8 Indices 159
11.9 Laws of indices 160
11.10 Simplifying expressions with indices 163
12 Approximation 169
12.1 Introduction 169
12.2 Rounding whole numbers 169
12.3 Rounding with decimals 171
12.4 Significant figures 174
12.5 Estimates 176
12.6 Rounding in practical problems 178
12.7 Accuracy of measurements 179
13 Equations 1 182
13.1 Introduction 182
13.2 Finding missing numbers 183
13.3 Solving linear equations 183
13.4 Equations with brackets 188
13.5 Solving problems using equations 189
13.6 Solving inequalities 191
14 Percentages 195
14.1 Introduction 195
14.2 Percentages, decimals, fractions 195

14.3 Percentages of quantities 199
14.4 Increasing and decreasing quantities 201
14.5 One quantity as a percentage of another quantity 202
14.6 Percentage increases 203
14.7 Using multipliers 205
15 Formulae **209**
15.1 What is a formula? 209
15.2 Evaluating terms other than the subject 211
15.3 Changing the subject of a formula 213
16 Circles **216**
16.1 Introduction 216
16.2 Circumference of a circle 218
16.3 Area of a circle 221
16.4 Two properties of circles 224
17 Probability **229**
17.1 Introduction 229
17.2 Relative frequency 230
17.3 Probability of a single event 231
17.4 Two events 234
17.5 Tree diagrams 239
17.6 Expected frequency 243
18 Three-dimensional shapes **246**
18.1 Introduction 246
18.2 Nets and surface area 249
18.3 Volume of a cuboid 251
18.4 Volume of a prism 253
18.5 Weight of a prism 256
19 Ratio and proportion **260**
19.1 What is a ratio? 260
19.2 Scales 262
19.3 Using ratio 265
19.4 Direct proportion 267
19.5 Inverse proportion 269
20 Pythagoras' theorem and trigonometry **273**
20.1 Pythagoras' theorem 273
20.2 Using Pythagoras' theorem 274
20.3 Proof of Pythagoras' theorem 278
20.4 Pythagoras' theorem problems 279

20.5 Trigonometry 280
20.6 The tangent ratio 281
20.7 Values of the tangent 282
20.8 Using tangents 283
20.9 Sine and cosine 285
21 Indices and standard form 292
21.1 Indices 292
21.2 Laws of indices 293
21.3 Prime factors 294
21.4 Highest common factor 295
21.5 Lowest common multiple 297
21.6 Standard form – large numbers 298
21.7 Standard form – small numbers 301
21.8 Standard form calculations 302
22 Statistics 2 306
22.1 Averages 306
22.2 The mode 306
22.3 The median 307
22.4 The mean 309
22.5 The range 312
22.6 Frequency tables 313
22.7 Grouped frequency tables 315
23 Graphs 2 322
23.1 Equations of straight lines 322
23.2 Drawing straight-line graphs 325
23.3 The gradient of a straight line 328
23.4 Curved graphs 331
24 Equations 2 337
24.1 Simultaneous equations 337
24.2 Algebraic methods 339
24.3 Quadratic expressions 343
24.4 Factorizing quadratic expressions 345
24.5 Quadratic equations 348
24.6 Solution by factorizing 349
24.7 Cubic equations 352
Answers **358**
Taking it further **388**
Index **393**

Credits

Front cover: © Digifoto Silver/Alamy

Back cover: © Jakub Semeniuk/iStockphoto.com, © Royalty-Free/Corbis, © agencyby/iStockphoto.com, © Andy Cook/iStockphoto.com, © Christopher Ewing/iStockphoto.com, © zebicho – Fotolia.com, © Geoffrey Holman/iStockphoto.com, © Photodisc/Getty Images, © James C. Pruitt/iStockphoto.com, © Mohamed Saber – Fotolia.com

Welcome to *Complete Mathematics*!

Complete Mathematics aims to give you a broad mathematical experience and a firm foundation for further study. The book will also be a useful source of reference for homework or revision for students who are studying a mathematics course.

This book is primarily aimed at students who do not have teacher support. Consequently, the explanations are detailed and there are numerous worked examples. Throughout this new edition are author insight boxes which provide additional support, including memory aids and tips on how to avoid common pitfalls. Also new to this book are end of chapter summaries of all the key points and formulae.

Access to a calculator has been assumed throughout the book. Generally, a basic calculator is adequate but, for Chapters 20 and 21, a scientific calculator, that is, a calculator which includes the trigonometric functions of sine, cosine and tangent, is essential. Very little else, other than some knowledge of basic arithmetic, has been assumed.

To derive maximum benefit from *Complete Mathematics*, work through and don't just read through the book. This applies not only to the exercises but also to the examples. Doing this will help your mathematical confidence grow.

The authors would like to thank the staff of Hodder Education for the valuable advice they have given.

Trevor Johnson, Hugh Neill, 2010

Only got a minute?

Mathematics is probably the most versatile subject there is. It has applications in a wide range of other subjects from science and geography through to art. In science, without mathematics there would be no formulae or equations describing the laws of physics which have enabled the technology that we use each day to develop. Mathematics enables psychologists to find out more about human behaviour, looking for correlations between sets of statistics. Doctors and pharmaceutical companies use mathematics to interpret the data resulting from a drug trial. An understanding of mathematics is essential in architecture and engineering. Mathematics is also used to model situations such as the stock market or the flow of traffic on motorways.

As well as being a useful subject, mathematics is also intrinsically beautiful. Many famous artists like Kandinsky and Escher used mathematics as a basis for their work – Kandinsky made extensive use of geometry in his paintings and Escher is famed for his complicated tessellations and optical illusions. Mathematics also occurs naturally – in nature there are many examples of mathematical patterns and symmetry. For example, the patterns on the wings of a butterfly or moth are symmetrical, and the pattern of seeds on a sunflower head, the scales of a pinecone and the spiral of a snail's shell all follow the same sequence of numbers – called the Fibonacci sequence.

Maths really is ubiquitous and having an understanding of the basics of mathematics is invaluable to many – arguably all – aspects of modern life.

5 Only got five minutes?

The word 'mathematics' originally comes from the Greek for knowledge and science. To most people maths means arithmetic – addition, subtraction, multiplication and division – which are the building blocks for maths. Without these we couldn't interpret large sets of data, solve an equation or work out the area of a shape.

We live in a technological age and mathematics underpins all of this technology – without mathematics none of the luxuries of modern life that we all take for granted would be possible. It is hard to think of an aspect of modern life that is not reliant in some way on technology and therefore mathematics. Mathematics is also intrinsic to many careers: civil engineers rely on mathematics to work out whether the bridge they are building will actually support a given load and surveyors use trigonometry to calculate the distance between two points. Also, insurance companies use complex probability tables for their life assurance policies, the self-employed will need to use percentages to file their tax returns, while pharmacists use formulae to work out the correct dosage for a particular drug and sportspeople are constantly using forms of intuitive mathematics.

Many people lack confidence in mathematics and feel that they are not very good in this subject. For some reason it seems acceptable to say, 'Oh, I'm hopeless at maths', in a way that would not be accepted about reading or writing. This lack of confidence is often rooted in negative experiences when at school – not 'getting' a difficult concept, new idea or solution to a problem. However, without realising it, every day we are using maths. Working out which box of cereal is best value for money involves proportion. Placing a bet on a horse requires an understanding of probability. Calculating the best mobile phone tariff involves using simple formulae. Working out how much tax you should pay for your tax return needs a good grasp of percentages. When reading the newspaper we are bombarded with statistics – results of surveys, charts and graphs – which we need to be able to interpret so that we can make informed choices about our daily lives.

Maths is also fun, a fact which is often forgotten! Most newspapers run a daily maths type puzzle: Sudoku, Futoshiki and Kakuro are three popular puzzles which involve numbers, logical thinking and problem solving skills. So millions of people ‘do’ maths everyday purely for enjoyment and the feeling of satisfaction you get from solving a difficult puzzle. Researchers are discovering more and more health benefits from solving the ubiquitous brain training maths style puzzles. Many of these puzzles are not only fun but are also positively good for you. Recent research suggests that keeping the mind stimulated with activities such as Sudoku and crossword puzzles may help prevent Alzheimer’s disease and can improve your cognitive abilities.

In today’s society being numerate is at least as important a skill as being literate. The key mathematics skills – such as basic numeracy and problem solving – are essential to many careers and highly valued by employers. Being proficient in mathematics has a positive impact on career development and employability, as well as self-esteem. *Complete Mathematics* takes you through worked examples and has carefully graded exercises designed to improve confidence and problem solving skills.

10 Only got ten minutes?

Was mathematics already there just waiting to be discovered or did we invent it? Mathematicians continue to debate this seemingly simple question – many believing that mathematics, like science, is discovered and not simply made up; that there are universal laws which govern mathematics which hold true no matter what system you use for writing down the numbers. Other mathematicians believe that we manufactured mathematics and then discovered patterns and laws based on the rules that we had already set.

Whichever view you hold there is no doubting that mathematics has a long history. Even the most basic communication between people may involve some form of mathematical discourse. When prehistoric man came back to his tribe with news of a herd of mammoths, he would have to relate the number in the herd, the distance, the direction, where they were travelling, at what speed – all of these relied upon his mathematical skills as well as those of his tribe. This vital communication could literally be a matter of life or death.

Over the centuries, more formal mathematics was needed and many different systems of writing numbers have been used. We assume that our use of 'base 10' is almost absolute, whereas, of course, it isn't. An intelligent alien species with two or twelve or even fifteen digits would invariably use a different system. In fact, base 2 or binary – a number system which only uses the digits 1 and 0 – is behind the programming used by all modern computers.

As mathematics developed, so the number system changed. For example, the number 60 is written as LX using Roman numerals while the Babylonians used base 60, but with only two symbols, one meaning 1 and one meaning 10. The Babylonians also began to develop the use of zero as a place holder, as did to some extent the Ancient Greeks. However, it was in India that the number system that we use today, including the use of zero as an indicator of an empty place (e.g. in 4086) as well as a value in its own right, was first used. Hindu mathematicians were also the first to really begin to develop the concept of negative numbers.

As the human race became more evolved so our need for more complicated numbers developed – prehistoric man would not have had much use for decimals or negative numbers. As well as integers, the idea of parts of a whole has developed over many hundreds of years. The Babylonians created a system of fractions in about 2000 BC as did the Ancient Egyptians four hundred years later. The word 'fraction' comes from the Latin *fractus* (broken) and fractions were often called 'broken numbers'. The Romans used fractions which converted easily to hundredths to calculate taxes; a natural progression from this type of fraction is percentages – a percentage is a fraction expressed in hundredths. The word 'percentage' is derived from the Latin *per centum*, which means 'per hundred'. Percentages today are used in a wide variety of contexts from finance to nutritional guidelines.

Once a system for writing numbers was found it was natural to then go on to solve more complicated problems including those which might lead to an equation to solve or a formula to write down. Mathematicians in Ancient Egypt and Babylon were able to solve many types of equation, which were expressed in words or drawings. In Greece, Diophantus (*c.* 275 BC), who has been called the 'Father of Algebra', used a form of symbolism, although the notation in use today did not appear until the seventeenth century.

Another important branch of maths is geometry, which is the study of angles, points, lines, surfaces and solids. Geometry probably originated in Ancient Egypt, where it was applied to land surveying and navigation. Indeed, the name 'geometry' is derived from the Greek for 'earth measure'. It was the Greek mathematician, Euclid, who developed the foundations for geometry – he first defined basic geometrical concepts including angles. Early civilizations had an understanding of angles, and trigonometry which was used in navigation, surveying and astronomy. There is evidence that the Egyptians used trigonometry to help them build their pyramids and today it is used in a variety of fields such as surveying, engineering and even music.

One of the most important modern applications of mathematics is in statistics and probability. Statistics deals with the collection, organization and analysis of data, often, but not always, numerical data. Starting with pictorial records on the walls of cave dwellings, the collection and recording of data is as old as civilization itself. Before 3000 BC, the Babylonians gathered information about crops and trade and so, centuries later, did the Egyptians and the Chinese. The word ‘statistics’ itself is derived from the Latin word *statisticus*, meaning ‘of state affairs’. Today, an understanding of how data is handled and interpreted enables us to make more informed decisions in our everyday lives.

The study of probability has its origins in the mid-sixteenth century, when mathematicians attempted to answer questions arising in gambling. One hundred years later the Frenchmen Blaise Pascal and Pierre de Fermat developed probability theory – in effect, the laws of chance. Today, probability is widely used in science and industry, its applications ranging from genetics and quantum mechanics to insurance.

Today mathematics is more relevant than ever – its applications are wide ranging and impact on most aspects of our lives. *Complete Mathematics* aims to help you improve your knowledge and understanding of mathematics. The key to improving your maths is practice – this is a book to be worked through and not simply read. Have the confidence to try out different strategies, make jottings and re-read the examples. Don’t be put off by large or ‘difficult’ numbers – change the numbers to simplify the problem and then ask yourself how you should approach this new simpler problem. Working through this book will improve your mathematical confidence, knowledge and understanding. Whatever your reasons for studying mathematics, the authors hope you enjoy the process!

1

Number

In this chapter you will learn:

- ***about place value***
- ***about the four operations of arithmetic***
- ***about the order in which arithmetic operations should be carried out***
- ***about some special numbers***.

1.1 Introduction – place value

Over the centuries, many systems of writing numbers have been used. For example, the number which we write as 17 is written as XVII using Roman numerals.

Our system of writing numbers is called the **decimal system**, because it is based on ten, the number of fingers and thumbs we have.

The decimal system uses the **digits** 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

The place of the digit in a number tells you the value of that digit.

When you write a number, the value of the digit in each of the first four columns, starting from the right, is

thousands	**hundreds**	**tens**	**units**
3	6	5	2

Thus, in the number 3652, the value of the 2 is two units and the value of the 6 is six hundreds.

The decimal system uses zero to show that a column is empty. So 308 means three hundreds and eight units; that is, there are no tens. The number five thousand and twenty-three would be written 5023, using the zero to show that there are no hundreds.

Extra columns may be added to deal with larger numbers, the value of each column being ten times greater than that of the column on its immediate right. It is usual, however, to split large numbers up into groups of three digits.

You read 372 891 as three hundred and seventy-two thousand, eight hundred and ninety-one.

In words, 428 763 236 is four hundred and twenty-eight million, seven hundred and sixty-three thousand, two hundred and thirty-six.

EXERICE 1.1

1 What is the value of the 5 in each of the following numbers?
(a) 357 **(b)** 598 **(c)** 5842 **(d)** 6785

2 Write these numbers in figures.
(a) seventeen **(b)** seventy
(c) ninety-seven **(d)** five hundred and forty-six
(e) six hundred and three **(f)** eight hundred and ten
(g) four hundred and fifty **(h)** ten thousand
(i) eight thousand, nine hundred and thirty-four
(j) six thousand, four hundred and eighty
(k) three thousand and six

3 Write these numbers completely in words.
(a) 52 **(b)** 871 **(c)** 5624 **(d)** 980
(e) 7001 **(f)** 35 013 **(g)** 241 001 **(h)** 1 001 312

4 Write down the largest number and the smallest number you can make using
(a) 5, 9 and 7 **(b)** 7, 2, 1 and 9.

5 Write these numbers completely in words.
  (a) 342 785 **(b)** 3 783 194
  (c) 17 021 209 **(d)** 305 213 097

6 Write these numbers in figures.
  (a) five hundred and sixteen thousand, two hundred and nineteen
  (b) two hundred and six thousand, and twenty-four
  (c) twenty-one million, four hundred and thirty-seven thousand, eight hundred and sixty-nine
  (d) seven million, six hundred and four thousand, and thirteen

---

1.2 Arithmetic – the four operations

In this book, the authors assume that you have access to a calculator, and that you are able to carry out simple examples of the four operations of arithmetic – addition, subtraction, multiplication and division – either in your head or on paper. If you have difficulty with this, you may wish to consult *Teach Yourself Basic Mathematics*.

Each of the operations has a special symbol, and a name for the result.

Adding 4 and 3 is written $4 + 3$, and the result, 7, is called the **sum** of 4 and 3.

Subtracting a number, 4, from a larger number, 7, is written $7 - 4$, and the result, 3, is their **difference**.

Addition and subtraction are ‘reverse’ processes:

$$4 + 3 = 7 \text{ and } 7 - 4 = 3.$$

Multiplying 8 by 4 is written 8×4, and the result, 32, is called the **product** of 8 by 4.

Dividing 8 by 4 is written $8 \div 4$, or $\frac{8}{4}$, and the result, 2, is called their **quotient**.

Multiplication and division are 'reverse' processes:

$$4 \times 2 = 8 \text{ and } 8 \div 4 = 2.$$

Sometimes divisions are not exact. For example, if you divide 7 by 4, the quotient is 1 and the **remainder** is 3, because 4 goes into 7 once and there is 3 left over.

Exercise 1.2 gives you practice with the four operations. You should be able to do all of this exercise without a calculator.

EXERCISE 1.2

1 Work out the results of the following additions.
(a) $23 + 6$ **(b)** $33 + 15$ **(c)** $45 + 9$ **(d)** $27 + 34$
2 Find the sum of
(a) 65, 24 and 5 **(b)** 27, 36 and 51.
3 Work out the following differences.
(a) $73 - 9$ **(b)** $49 - 35$ **(c)** $592 - 76$
(d) $128 - 43$ **(e)** $124 - 58$ **(f)** $171 - 93$
4 Find the difference between
(a) 152 and 134 **(b)** 317 and 452.
5 Work out
(a) 5×9 **(b)** 43×7 **(c)** 429×6 **(d)** 508×7
(e) 27×11 **(f)** 12×28 **(g)** 13×14 **(h)** 21×21
6 Find the product of 108 and 23.
7 Work out
(a) $375 \div 5$ **(b)** $846 \div 3$ **(c)** $1701 \div 7$ **(d)** $1752 \div 8$
8 Find the quotients and the remainders for the following divisions.
(a) $8 \div 3$ **(b)** $23 \div 5$ **(c)** $101 \div 8$

1.3 Order of operations

Sometimes you will see a calculation in the form $8 - (2 \times 3)$, where brackets, that is the symbols (and), are put around part of the calculation. Brackets are there to tell you to carry out the calculation inside them first. So 8 – (2 × 3) = 8 – 6 = 2.

If you had to find the answer to 8 + 5 – 3, you might wonder whether to do the addition first or the subtraction first.

If you do the addition first, 8 + 5 – 3 = 13 – 3 = 10.

If you do the subtraction first, 8 + 5 – 3 = 8 + 2 = 10.

In this case, it makes no difference which operation you do first.

If, however, you had to find the answer to 3 + 4 × 2, your answer will depend on whether you do the addition first or the multiplication first.

If you do the addition first, 3 + 4 × 2 = 7 × 2 = 14.

If you do the multiplication first, 3 + 4 × 2 = 3 + 8 = 11.

To avoid confusion, mathematicians have decided that, if there are no brackets, multiplication and division are carried out before addition and subtraction. Such a rule is called a **convention**: clearly the opposite decision could be made, but it has been agreed that the calculation 3 + 4 × 2 should be carried out as 3 + 4 × 2 = 3 + 8 = 11.

There are a number of these conventions, often remembered by the word ‘BiDMAS’, which stands for

Brackets,
Indicies,
Divide,
Multiply,
Add,
Subtract.

This is the order in which arithmetic operations must be carried out. Most calculators have the ‘BiDMAS’ convention programmed into them, and so automatically carry out operations in the correct order.

Insight

When a calculation involves just multiplication/division or addition/subtraction, then you can carry it out in any order. It often helps to reorder a calculation to make it easier to perform. You must keep the operation with its original number.

For example, $2 \times 24 \times 5 \div 3 = 2 \times 5 \times 24 \div 3 = 10 \times 24 \div 3 = 10 \times 8 = 80$

Example 1.3.1

Carry out the following calculations.

(a) $5 \times (6 - 2)$ **(b)** $(10 + 2) \div 4$ **(c)** $20 - 3 \times 5$
(d) $24 \div 6 + 2$ **(e)** $4 \times (8 - 3) + 2$

(a) $5 \times (6 - 2) = 5 \times 4$ brackets first
$= 20$
(b) $(10 + 2) \div 4 = 12 \div 4$ brackets first
$= 3$
(c) $20 - 3 \times 5 = 20 - 15$ multiply before subtracting
$= 5$
(d) $24 \div 6 + 2 = 4 + 2 = 6$ division before addition
(e) $4 \times (8 - 3) + 2 = 4 \times 5 + 2$ brackets first
$= 20 + 2$ multiply before adding
$= 22$

Try the calculations in Example 1.3.1 on your calculator to see if it uses the ‘BiDMAS’ convention and carries them out correctly.

EXERCISE

EXERCISE 1.3

In each part of question 1, use the ‘BiDMAS’ convention to carry out the calculation.

1 **(a)** $(3 + 4) \times 5$ **(b)** $3 + 4 \times 5$ **(c)** $(7 - 3) \times 2$
(d) $7 - 3 \times 2$ **(e)** $(12 \div 6) \div 2$ **(f)** $12 \div (6 \div 2)$
(g) $20 \div 4 + 1$ **(h)** $3 \times 5 + 4 \times 2$ **(i)** $4 \times (7 - 2) + 6$
(j) $6 \times 4 - 2 \times 8$ **(k)** $24 \div (4 + 2) - 1$ **(l)** $(7 + 3) \times (9 - 2)$
(m) $10 + (8 + 4) \div 3$ **(n)** $9 + 5 \times (8 - 2)$ **(o)** $15 \div 3 - 16 \div 8$
(p) $5 + 6 \times 3 - 1$ **(q)** $(5 + 6) \times 3 - 1$ **(r)** $5 + 6 \times (3 - 1)$

1.4 Problems which use arithmetic

In Exercise 1.4, you have to decide which arithmetic operation to use. If you are not sure, it can sometimes be helpful to think of a similar question with smaller numbers.

Example 1.4.1

An egg box holds six eggs. How many eggs boxes are needed to pack 500 eggs, and how many eggs are there in the last box?

You need to find how many sixes there are in 500. If you use a calculator you get 500 ÷ 6 = 83.333 333 3.

This means that 84 boxes are needed. 83 full boxes hold 83 × 6 eggs, which is 498 eggs. So there are 2 eggs in the last box.

Insight

Whether you need to round up or down depends on the context of the question. In the last example, you rounded up as you needed 84 boxes altogether. The answer to the question: 'How many pens costing £6 each can be bought for £500?' is 83, as 83 × £6 = £498. So there is not enough money to buy 84 pens.

EXERCISE 1.4

1. In a school, 457 of the pupils are boys and 536 are girls. Work out the number of pupils in the school.
2. How many hours are there in 7 days?
3. The sum of two numbers is 3472. One of the numbers is 1968. Find the other number.
4. The product of two numbers is 161. One of the numbers is 7. Find the other number.
5. A car travels 34 miles on one gallon of petrol. How far will it travel on 18 gallons?

(Contd)

6 How many pieces of metal 6 cm long can be cut from a bar 117 cm long? What length of metal is left over?
7 A box holds 24 jars of coffee. How many boxes will be needed to hold 768 jars?
8 At the start of a journey, a car’s milometer read 34 652. After the journey, it read 34 841. How long was the journey?
9 There are 365 days in a year. How many days are there in 28 years?
10 A coach can carry 48 passengers. How many coaches are needed to carry 1100 passengers? How many seats will be unoccupied?

1.5 Special numbers

The numbers 0, 1, 2, 3, 4, . . . are called **whole numbers**.

The numbers 1, 2, 3, 4, . . . are called **natural numbers**. They are also called **counting numbers** or **positive whole numbers**.

The numbers . . . , –3, –2, –1, 0, 1, 2, 3, . . . are called **integers**. These are the positive and negative whole numbers, together with 0.

The natural numbers can be split into even and odd numbers.

An **even number** is one into which 2 divides exactly. The first six even numbers are 2, 4, 6, 8, 10 and 12. You can tell whether a number is even by looking at its last digit. If the last digit is 2, 4, 6, 8 or 0, then the number is even. The last digit in the number 784 is 4, so 784 is even.

An **odd number** is one into which 2 does not divide exactly; that is, it is a natural number which is not even. The first six odd numbers are 1, 3, 5, 7, 9 and 11. You can tell whether a number is odd by looking at its last digit. If the last digit is 1, 3, 5, 7 or 9, then the number is odd. In other words, if the last digit is odd, then the number is odd. The last digit in the number 837 is odd, so 837 is odd.

A **square number** is a natural number multiplied by itself. It is sometimes called a **perfect square** or simply a **square**, and can be shown as a square of dots. The first four square numbers are 1 × 1 = 1, 2 × 2 = 4, 3 × 3 = 9 and 4 × 4 = 16 (Figure 1.1).

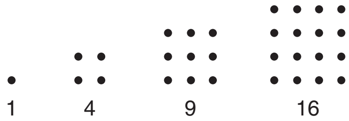

Figure 1.1

To get a square number you multiply a whole number by itself. So the fifth square number is 5 × 5 = 25.

A **triangle number** can be illustrated as a triangle of dots (Figure 1.2). The first four triangle numbers are 1, 3, 6 and 10.

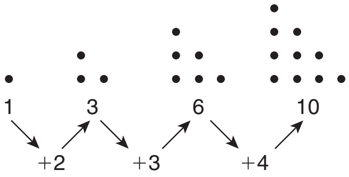

Figure 1.2

Each triangle number is obtained from the previous one by adding one more row of dots down the diagonal. The fifth triangle number is found by adding 5 to the fourth triangle number, so the fifth triangle number is 15.

A **rectangle number** can be shown as a rectangle of dots. A single dot or a line of dots is not regarded as a rectangle. Figure 1.3 shows the rectangle number 12 illustrated in two ways.

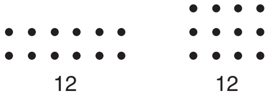

Figure 1.3

A **cube number** can be shown as a cube of dots (Figure 1.4). The first three cube numbers are 1, 8 and 27 which can be written 1 = 1 × 1 × 1, 8 = 2 × 2 × 2 and 27 = 3 × 3 × 3. To get a cube number you multiply a whole number by itself, and then by itself again. So the fourth cube number is 4 × 4 × 4 = 64.

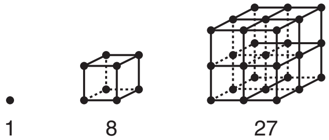

Figure 1.4

EXERCISE 1.5

1 From the list 34, 67, 112, 568, 741 and 4366, write down
 (a) all the even numbers,
 (b) all the odd numbers.
2 Write down the first ten square numbers.
3 Write down the first ten triangle numbers.
4 Write down the first ten rectangle numbers.
5 Write down the first ten cube numbers.
6 From the list of numbers 20, 25, 28, 33, 36 and 37, write down
 (a) all the square numbers,
 (b) all the triangle numbers,
 (c) all the rectangle numbers.
7 The difference between consecutive square numbers is always the same type of number. What type of number is it?
8 The sum of the first two triangle numbers is 4 (=1 + 3).
 (a) Find the sums of three more pairs of consecutive triangle numbers.
 (b) What do you notice about your answers?
9 Find the sums of the first two, three, four and five cube numbers. What do you notice about your answers?
10 **(a)** Write down a rectangle number, other than 12, for which there are two different patterns of dots. Draw the patterns.
 (b) Write down a rectangle number for which there are three different patterns of dots.
 (c) Write down a rectangle number for which there are four different patterns of dots.

1.6 Multiples, factors and primes

Figure 1.5 shows the first five numbers in the 8 times table. The numbers on the right, 8, 16, 24, 32 and 40, are called **multiples** of 8. Notice that 8 itself, which is 1 × 8 or 8 × 1, is a multiple of 8.

$1 \times 8 = 8$
$2 \times 8 = 16$
$3 \times 8 = 24$
$4 \times 8 = 32$
$5 \times 8 = 40$

Figure 1.5

The **factors** of a number are the numbers which divide exactly into it. So, the factors of 18 are 1, 2, 3, 6, 9 and 18. Notice that 1 is a factor of every number and that every number is a factor of itself.

Example 1.6.1

Write down all the factors of **(a)** 12, **(b)** 11.

(a) The factors of 12 are 1, 2, 3, 4, 6 and 12.
(b) The factors of 11 are 1 and 11.

Insight

It is easy to find all the factors of a number by looking for pairs of numbers that multiply to give that number. For example, 1 × 18 = 18, 2 × 9 = 18, 3 × 6 = 18, so the factors of 18 are 1, 2, 3, 6, 9 and 18.

The number 11 has only two factors and is an example of a **prime number**. A prime number, sometimes simply called a **prime**, is a number greater than 1 which has exactly two factors, itself and 1.

Thus, the first eight prime numbers are 2, 3, 5, 7, 11, 13, 17 and 19.

Insight

2 is the only even prime number.

Is 1 a prime number? The number 1 has only one factor (itself), and so it is not a prime number since the definition states that a prime number must have two factors.

EXERCISE 1.6

1 Write down the first five multiples of 7.
2 **(a)** Write down the first ten multiples of 5.
(b) How can you tell that 675 is also a multiple of 5?

(Contd)

3 Write down all the factors of 20.

4 **(a)** Find the sum of the factors of 28, apart from 28 itself.

(b) Comment on your answer.

5 Write down all the factors of 23.

6 **(a)** The number 4 has three factors (1, 2 and 4). Find five more numbers, each of which has an odd number of factors.

(b) What do you notice about your answers to part **(a)**?

7 Write down the four prime numbers between 20 and 40.

8 2 is the only even prime number. Explain why there cannot be any others.

9 Write down the largest number which is a factor of 20 and is also a factor of 30.

10 Write down the smallest number which is a multiple of 6 and is also a multiple of 8.

11 Write down the factors of 30 which are also prime numbers. (These are called the prime factors of 30.)

12 Write down the **prime factors** of 35.

13 The prime factors of 6 are 2 and 3. Find another number which has 2 and 3 as its only prime factors.

THINGS TO REMEMBER

- *The decimal system uses the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.*
- Sum *means the result of adding two or more numbers together.*
- Difference *means the result of subtracting a smaller number from a larger number.*
- Product *means the result of multiplying two or more numbers together.*
- Quotient *means the whole number resulting from the division of two numbers i.e. when 14 is divided by 3, the* quotient *is 4 and the* remainder *is 2.*
- *The order operations should be performed in can be remembered using the mnemonic BiDMAS, which stands for:*
  - *Brackets } Work out any brackets first...*
  - *Divide* } *... then any division or multiplication...*
  - *Multiply* }
  - *Add* } *... lastly, any addition or subtraction.*
  - *Subtract* }
- *Natural numbers or counting numbers or positive whole numbers are the numbers 1, 2, 3, 4, . . .*
- *Integers are the numbers . . . , –3, –2, –1, 0, 1, 2, 3, . . .*
- *Even numbers are exactly divisible by 2, and are the numbers 2, 4, 6, 8, . . .*
- *Odd numbers are not exactly divisible by 2, and are the numbers 1, 3, 5, 7, . .*
- *Square numbers are the result of multiplying a natural number by itself. They include the numbers 1, 4, 9, 16, . . .*

- *Triangle numbers can be shown as a triangle of dots. They include the numbers 1, 3, 6, 10, . . .*
- *Rectangle numbers are any number that can be shown as a rectangle of dots.*
- *Cube numbers can be shown as a cube of dots. The first three cube numbers are 1, 8 and 27.*
- *The multiples of a number are the answers to that number's times table. So the multiples of 4 are 4, 8, 12, 16, 20, . . .*
- *A factor of a number is any number which divides exactly into that number. Every number has 1 and itself as factors. The factors of 10 are 1, 2, 5 and 10.*
- *Prime numbers are numbers with exactly two factors. The first five prime numbers are 2, 3, 5, 7 and 11.*

2

Angles

In this chapter you will learn:

- ***about different types of angles***
- ***how to measure and draw angles***
- ***angle facts and how to use them***
- ***about parallel lines***
- ***about bearings.***

2.1 Introduction

The study of angles is one branch of **geometry**, which is concerned also with points, lines, surfaces and solids. Geometry probably originated in Ancient Egypt, where it was applied, for example, to land surveying and navigation. Indeed, the name ‘geometry’ is derived from the Greek for ‘earth measure’.

It was Greek mathematicians, notably Euclid, who developed a theoretical foundation for geometry. He began his classic treatise *Elements* with definitions of basic geometrical concepts including angles, the subject of this chapter.

2.2 Angles

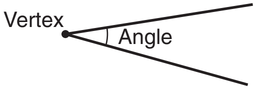

Figure 2.1

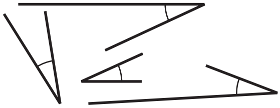

Figure 2.2

An **angle** is formed when two straight lines meet, as in Figure 2.1. The point where the lines meet is called the **vertex** of the angle. The size of the angle is the amount of turn from one line to the other, and is not affected by the lengths of the lines. The angles in Figure 2.2 are all equal.

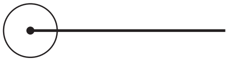

Figure 2.3

If you begin by facing in a certain direction and then turn round until you are facing in the original direction again, as in Figure 2.3, you will have made a complete turn (or revolution). One complete turn is divided into 360 **degrees**, which is written as 360°.

In a half turn, Figure 2.4, therefore, there are 180°. In a quarter turn there are 90°, which is called a **right angle**. It is often marked on diagrams with a small square, as shown in Figure 2.5. If the angle between two lines is a right angle, the lines are said to be **perpendicular**.

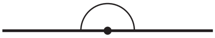

Figure 2.4

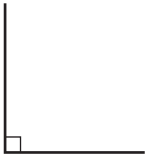

Figure 2.5

Apart from angles of 90° and 180°, all other sizes of angles fall into one of three categories, shown in Figure 2.6.

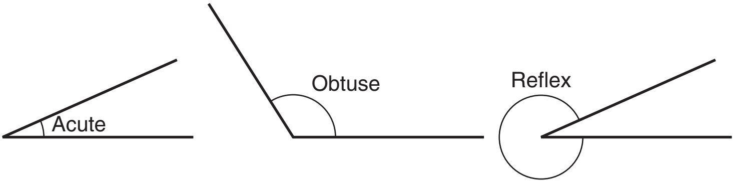

Figure 2.6

An angle which is less than 90° is called an **acute** angle.
An angle which is between 90° and 180° is called an **obtuse** angle.
An angle which is greater than 180° is called a **reflex** angle.

Insight

A useful memory aid is 'a-cute little angle' to remind you that angles less than 90° are acute.

There are three ways of naming angles.

One way is to use a letter inside the angle, as with the angle marked x in Figure 2.7. Usually, a small letter, rather than a capital, is used.

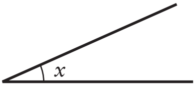

Figure 2.7

In Figure 2.8, the angle could be called angle B, but you should do this only if there is no risk of confusion. The symbols $\angle$ and $\wedge$ are both used to represent 'angle', so angle B could be shortened to $\angle B$ or $\hat{B}$. It could be argued that there are two angles at B, one of them acute and one of them reflex, but you should assume that the acute angle is intended unless you are told otherwise.

Alternatively, the angle in Figure 2.8 could be called angle ABC (or angle CBA). You could shorten this to $\angle ABC$ or $\widehat{ABC}$. The middle letter is always the vertex of the angle.

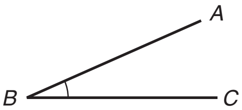

Figure 2.8

EXERCISE 2.1

1 How many degrees does the hour hand of a clock turn in 20 minutes?

2 Work out the size of the angle the second hand of a clock turns through in 1 second.

3 Work out the size of the angle the hour hand of a clock turns through in 35 minutes.

4 Work out the size of the angle the second hand of a clock turns through in 55 seconds.

5 Work out the size of the angle between South and West when it is measured **(a)** clockwise, **(b)** anticlockwise.

6 Work out the size of the angle between West and North West when it is measured **(a)** clockwise, **(b)** anticlockwise.

7 State whether each angle drawn below is acute, obtuse or reflex.

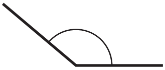

(a)

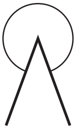

(b)

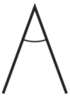

(c)

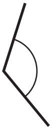

(d)

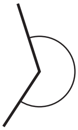

(e)

8 State whether each of the following angles is acute, obtuse or reflex.
(a) 172° **(b)** 203° **(c)** 302° **(d)** 23°

9 Use capital letters to name the angles *x*, *y*, *z* and *t* in the figure.

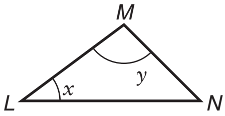

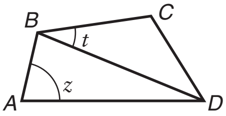

2.3 Measuring and drawing angles

It is a good idea to estimate the size of an angle before you try to measure it, as this will prevent you from giving answers which are not sensible. For this purpose, it is really only necessary to decide whether an angle is less than 90°, between 90° and 180°, between

180° and 270°, or greater than 270°. With practice, you will probably be able to estimate an angle to within about 20°.

To estimate the size of the angle in Figure 2.9, notice that it is less than 90°. It is also more than half a right angle (45°), but nearer to 45° than 90°. A reasonable estimate is, therefore, 50° or 60°.

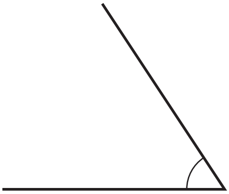

Figure 2.9

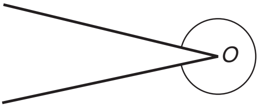

Figure 2.10

The reflex angle at *O* in Figure 2.10 is between 270° and 360°. It is easier to estimate the size of the acute angle and then to subtract it from 360°.

The acute angle at *O* is about a quarter of a right angle, about 20° or 30°, so an estimate for the reflex angle at *O* is 330° or 340°.

To measure or draw an angle accurately, you use a protractor. The outer scale goes clockwise from 0° to 180° and the inner scale goes anticlockwise from 0° to 180°.

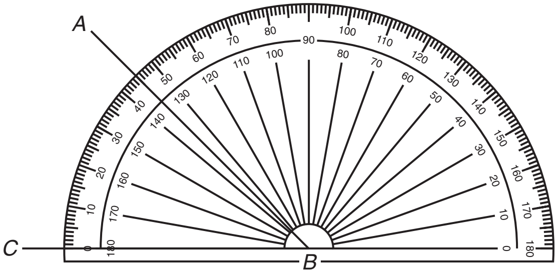

Figure 2.11

To measure the angle *ABC* in Figure 2.11, place the base line of the protractor on the line *BC* with the centre of the protractor over the vertex *B*. Read from the scale which has its zero on *BC*, that is, the outer scale. The size of angle *ABC* is 45°.

Insight

Always estimate the angle first. So if you mistakenly use the inner scale and give 135° as your measurement, you should see that this is not sensible, as angle *ABC* is acute.

EXERCISE 2.2

1 Estimate the size of each of the angles and then measure them. Trace the angles and extend the lines so that you can use your protractor to measure the angles as accurately as possible.

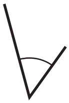
(a)

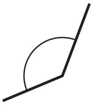
(b)

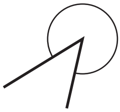
(c)

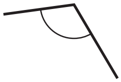
(d)

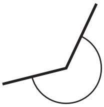
(e)

2 Draw angles with each of the following sizes.
(a) 73° **(b)** 246° **(c)** 109° **(d)** 281° **(e)** 172°

2.4 Using angle facts

You should be able to measure and draw angles to within one or two degrees. Note, however, that many of the diagrams in the rest of this chapter are not drawn accurately but they are still helpful, even when you find solutions by carrying out a calculation.

This section deals with situations which use angle facts, most of which you met in Section 2.2.

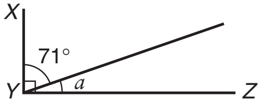

Figure 2.12

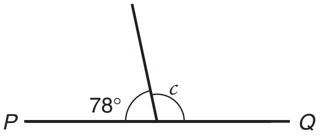

Figure 2.13

There are 90° in a right angle. In Figure 2.12, $\angle XYZ = 90°$, so $a + 71° = 90°$ giving $a = 90° - 71° = 19°$. Two angles whose sum is 90° are **complementary**.

The angle made on a straight line is 180°. If there are two or more angles on a straight line, then their sum must also be 180°. This fact is usually stated as **angles on a straight line add up to 180°**. In Figure 2.13, *PQ* is a straight line, so *c* and 78° add up to 180°. Thus, $c = 180° - 78° = 102°$. Two angles which add up to 180° are called **supplementary**, so 78° and 102° are supplementary angles.

There are 360° in a complete turn. The angles in Figure 2.14, when added together, make a complete turn. They are called **angles at a point**. So **the angles at a point add up to 360°**.

Therefore 117°, 122° and *e* add to 360°. $117° + 122° = 239°$, $e = 360° - 239° = 121°$.

When two straight lines cross, as in Figure 2.15, four angles, *a, b, c* and *d*, are formed. Since *a* added to *b* makes 180°, and *c* added to *b* also makes 180°, it follows that $a = c$. Similarly, $b = d$.

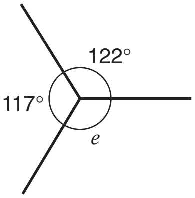

Figure 2.14

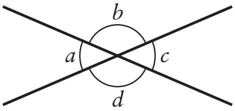

Figure 2.15

The angles *a* and *c* are called **vertically opposite**. The angles *b* and *d* are also vertically opposite. The argument in the previous paragraph shows that **vertically opposite angles are equal**.

Insight

You can think of a pair of scissors: as the angle between the handles increases, so does the angle between the blades.

EXERCISE 2.3

1 Find the size of each angle marked with a letter. Where they occur, *AB* and *CD* are straight lines.

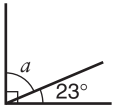

(a)

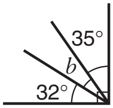

(b)

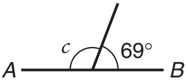

(c)

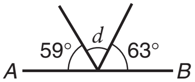

(d)

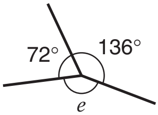

(e)

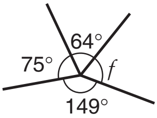

(f)

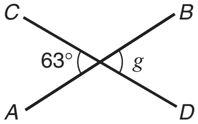

(g)

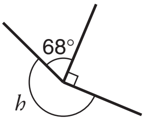

(h)

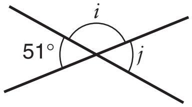

(i)

2 In the diagram, is ∠*AOD* a straight line? Give a reason for your answer.

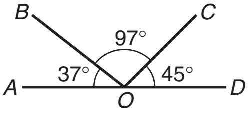
B
C
97°
37°
45°
A
O
D

2.5 Parallel lines

The word 'parallel', derived from the Greek, means 'alongside one another'. Straight lines are **parallel** if they are always the same distance apart. Parallel lines never meet, no matter how far they are extended. Straight railway lines, for example, are parallel.

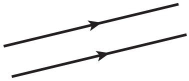

Figure 2.16

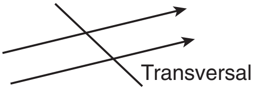
Transversal

Figure 2.17

In diagrams, arrows are often used to show that lines are parallel. (See Figure 2.16.) A straight line which crosses two or more parallel lines is called a **transversal**. Figure 2.17 shows a transversal.

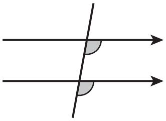

Figure 2.18

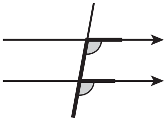

Figure 2.19

In Figure 2.18, a transversal cuts a pair of parallel lines. The pair of shaded angles are called **corresponding angles** and are equal to each other.

Insight

You will see from Figure 2.19 that the corresponding angles form an 'F-shape'. Looking for such a shape may help you to find pairs of corresponding angles.

Other pairs of corresponding angles have been shaded in Figure 2.20.

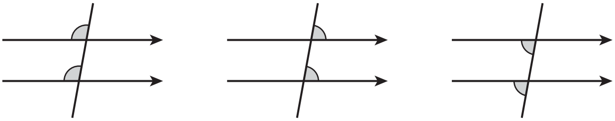

Figure 2.20

In Figure 2.21, a transversal cuts a pair of parallel lines. The pair of shaded angles are **alternate angles**, and are equal to each other.

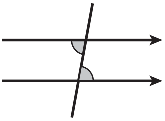

Figure 2.21

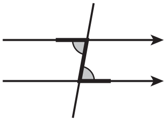

Figure 2.22

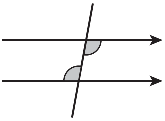

Figure 2.23

Insight

The characteristic 'Z-shape' in Figure 2.22 can help you spot alternate angles.

In Figure 2.24, the angles marked a and b are equal because they are corresponding, and the angles marked b and c are equal because they are vertically opposite. So the alternate angles a and c are equal.

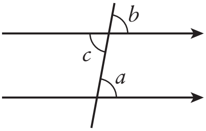

Figure 2.24

Example 2.5.1

Find, with reasons, the sizes of the angles marked a and b.

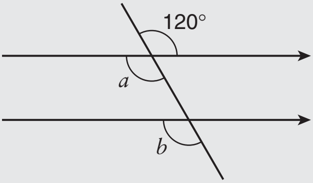

Angles a and 120° are vertically opposite, so $a = 120°$.

Angles a and b are corresponding, so $b = 120°$.

EXERCISE 2.4

1 Write down the letter of the angle which is **(i)** corresponding to the shaded angle, **(ii)** alternate to the shaded angle.

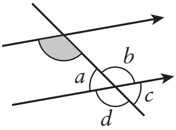

(a)

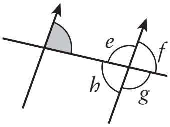

(b)

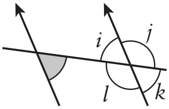

(c)

2 Find the size of each of the marked angles.

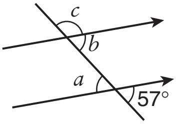

(a)

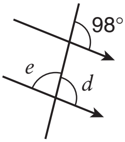

(b)

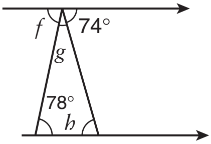

(c)

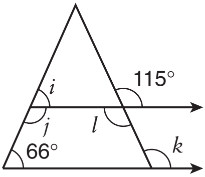

(d)

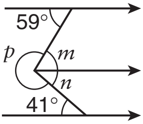

(e)

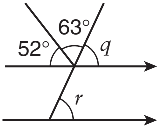

(f)

2.6 Bearings

In some situations, compass bearings, based on North, East, South and West, are used to describe directions. For navigation, **three-figure bearings** are used. A three-figure bearing is the angle always measured clockwise from North. When the angle is less than 100°, one or two zeros are inserted in front of the angle, so that the bearing still has three figures. Thus, East as a bearing is 090° and South is 180°. The directions of the four bearings 032°, 143°, 227° and 315° are shown in Figure 2.25.

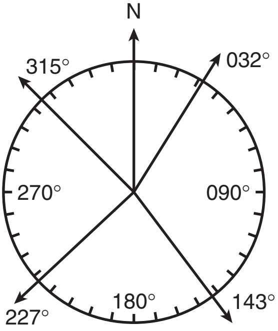

Figure 2.25

Example 2.6.1

The bearing of *B* from *A* is 074°. Find the bearing of *A* from *B*.

It is important to draw a rough sketch.

Start by putting *A* on the paper with the line due North through *A*, and then put *B* on the diagram, together with the line due North through *B*.

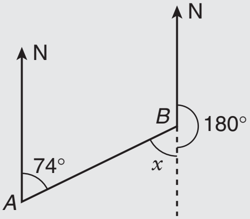

The bearing of *A* from *B* is the reflex angle at *B* and is therefore 180° + *x*. But the angle *x* is alternate to the angle of 74°, so *x* = 74°.

Therefore the bearing of *A* from *B* is 180° + *x* = 254°.

Insight

There are plenty of mnemonics to help you remember the order of the compass points. Never Eat Slimy Worms is one example.

EXERCISE 2.5

1 Make a tracing of the figure and extend the lines. Then, using your protractor, find the bearings of *A*, *B*, *C* and *D* from *O*.

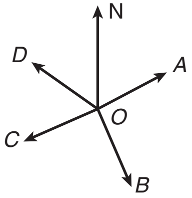

2 Draw accurate diagrams to illustrate these bearings.
(a) 083° **(b)** 126° **(c)** 229° **(d)** 284°

3 Express the following directions as three-figure bearings.
(a) West **(b)** North East **(c)** South West **(d)** North West

4 Express the following directions as three-figure bearings.

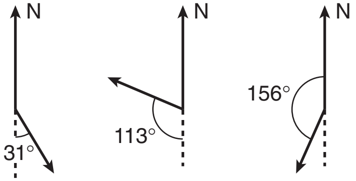

(a) **(b)** **(c)**

5 The bearing of *B* from *A* is 139°. Find the bearing of *A* from *B*.

6 The bearing of *D* from C is 205°. Find the bearing of *C* from *D*.

THINGS TO REMEMBER

- *An angle is formed when two straight lines meet.*
- *The vertex of an angle is the point where the lines meet.*
- *A right angle is a quarter turn, which is 90°.*
- *An acute angle is an angle which is less than 90°.*
- *An obtuse angle is an angle which is between 90° and 180°.*
- *A reflex angle is an angle which is greater than 180°.*
- *A protractor is used to measure angles accurately.*
- *Complementary angles are two angles whose sum is 90°.*
- *Supplementary angles are two angles whose sum is 180°.*
- *Angles on a straight line add up to 180°.*
- *Angles at a point add up to 360°.*
- *Vertically opposite angles are equal.*
- *Perpendicular lines cross at right angles.*
- *Parallel lines are straight lines which are always the same distance apart. Parallel lines never meet.*
- *A transversal is a straight line which crosses two or more parallel lines.*

- *Corresponding angles on a transversal are equal to each other. These are sometimes called F-angles.*

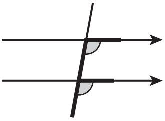

- *Alternate angles on a transversal are equal to each other. These are sometimes called Z-angles.*

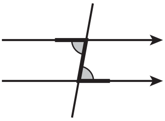

- *A three-figure bearing is the angle measured clockwise from North. So an angle of 27° from North corresponds to a bearing of 027°.*

3

Fractions

In this chapter you will learn:

- ***about equivalent fractions***
- ***how to compare fractions***
- ***how to add and subtract fractions***
- ***how to multiply and divide fractions.***

3.1 Introduction

The word ‘fraction’ comes from the Latin *fractus* (broken) and fractions were often called ‘broken numbers’. For many centuries, the numbers 1, 2, 3, 4, etc. met humanity’s needs and, although the Babylonians developed a system of fractions in about 2000 BC, it was 400 years later that the Ancient Egyptians produced the first thorough treatment.

3.2 What is a fraction?

A **fraction** is one or more equal parts of a whole.

In Figure 3.1 overleaf, the left circle has been split into 4 equal parts (quarters) and 3 of them are shaded. The fraction of the circle which is shaded is $\frac{3}{4}$. The top number, 3, is called the **numerator** and the bottom number, 4, is called the **denominator**. One of the four parts is unshaded, so the fraction of the circle which is unshaded is $\frac{1}{4}$.

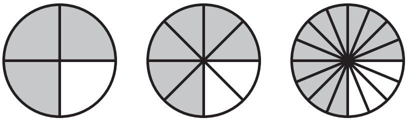

Figure 3.1

If you were asked what fraction of the circle is shaded in the middle circle in Figure 3.1, you might give $\frac{6}{8}$ as your answer, as the circle has been split into 8 equal parts and 6 of them are shaded. Similarly, in the third circle, the fraction $\frac{12}{16}$ is shaded.

You might say that it is obvious from Figure 3.1 that the three fractions $\frac{3}{4}$, $\frac{6}{8}$ and $\frac{12}{16}$ are equal. You would be right: they are equal in value and they are called **equivalent fractions**. You can write $\frac{3}{4} = \frac{6}{8} = \frac{12}{16}$.

You could get $\frac{6}{8}$ from $\frac{3}{4}$ by multiplying both the numerator and the denominator of $\frac{3}{4}$ by 2, that is, $\frac{3 \times 2}{4 \times 2} = \frac{6}{8}$. You can get $\frac{12}{16}$ by multiplying the numerator and the denominator of $\frac{3}{4}$ by 4. Similarly, $\frac{9}{12}$, $\frac{15}{20}$ and $\frac{18}{24}$ are three more fractions equivalent to $\frac{3}{4}$. In general, if you multiply the numerator and the denominator of a fraction by the same number, you will obtain a fraction equivalent to the original.

EXERCISE 3.1

1 What fractions of these shapes are **(a)** shaded, **(b)** unshaded?

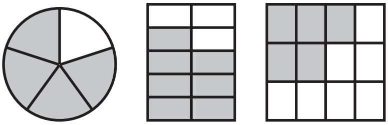

2 If $\frac{5}{6}$ of a shape is shaded, what fraction is unshaded?

3 If $\frac{3}{7}$ of a shape is shaded, what fraction is unshaded?

4 From $\frac{4}{5}$, $\frac{3}{8}$, $\frac{5}{6}$ and $\frac{8}{9}$, write down a fraction with
(a) a numerator of 5, **(b)** a denominator of 8.

5 Write down five fractions which are equivalent to
(a) $\frac{1}{3}$ (b) $\frac{3}{4}$ (c) $\frac{2}{5}$ (d) $\frac{5}{9}$

6 Complete $\frac{2}{3} = \frac{}{24}$ and $\frac{5}{8} = \frac{}{24}$. Is $\frac{2}{3}$ larger than $\frac{5}{8}$?

3.3 Which fraction is bigger?

If two fractions have equal denominators, it is easy to see which fraction is the larger. For example, $\frac{5}{8}$ is clearly larger than $\frac{3}{8}$. You only have to compare their numerators and see that 5 is larger than 3.

To find which is the larger of $\frac{2}{3}$ and $\frac{3}{4}$, make lists of fractions equivalent to $\frac{2}{3}$ and to $\frac{3}{4}$. Such lists are $\frac{2}{3} = \frac{4}{6} = \frac{6}{9} = \frac{8}{12} = \frac{10}{15} = \ldots$ and $\frac{3}{4} = \frac{6}{8} = \frac{9}{12} = \frac{12}{16} = \frac{15}{20} = \ldots$ In each list there is a fraction with a denominator of 12: $\frac{2}{3} = \frac{8}{12}$ and $\frac{3}{4} = \frac{9}{12}$. So $\frac{3}{4}$ is larger than $\frac{2}{3}$.

A quicker method is to convert $\frac{2}{3}$ and $\frac{3}{4}$ to equivalent fractions with the same denominator, in this case 12, because 12 is the lowest number which is a multiple of both 3 and 4. This number, 12, is called the **lowest common multiple** of 3 and 4.

Insight

If the lowest common multiple of the denominator is not obvious, you can find a common denominator by multiplying together the two denominators. For example, to compare $\frac{7}{10}$ and $\frac{5}{8}$ you can convert them to equivalent fractions with denominators of $10 \times 8 = 80$. So $\frac{7}{10} = \frac{7 \times 8}{10 \times 8} = \frac{56}{80}$ and $\frac{5}{8} = \frac{5 \times 10}{8 \times 10} = \frac{50}{80}$, therefore $\frac{7}{10}$ is greater than $\frac{5}{8}$.

3.4 Simplifying fractions

The fractions $\frac{2}{3}$ and $\frac{8}{12}$ are equivalent. To show this, start with $\frac{2}{3}$ and multiply both the numerator and the denominator by 4. Alternatively, start with $\frac{8}{12}$ and divide both its numerator and its denominator by 4. In general, by dividing both the numerator and the denominator of a fraction by the same number, you obtain a fraction which is equivalent to the original fraction. This process is called **cancelling**.

If the numerator and denominator of a fraction have no common factors, the fraction is said to be in its **simplest form** or in its **lowest terms**. So $\frac{2}{3}$ is the simplest form of $\frac{16}{24}$ but $\frac{4}{10}$ is not the simplest form of $\frac{12}{30}$, because 2 is a factor of both 4 and of 10. If you divide the numerator and the denominator of $\frac{4}{10}$ by 2, you obtain $\frac{2}{5}$, which is in its simplest form.

To find the simplest form of $\frac{12}{30}$ in a single step, you must find the highest number which is a factor of both 12 and 30, which is 6, and then divide both the numerator and the denominator by 6.

EXERCISE 3.2

1 Which fraction is larger?

(a) $\frac{3}{5}$ or $\frac{4}{5}$ **(b)** $\frac{2}{5}$ or $\frac{3}{10}$ **(c)** $\frac{2}{3}$ or $\frac{7}{9}$ **(d)** $\frac{1}{4}$ or $\frac{5}{16}$
(e) $\frac{3}{4}$ or $\frac{5}{6}$ **(f)** $\frac{5}{8}$ or $\frac{7}{12}$ **(g)** $\frac{1}{6}$ or $\frac{2}{9}$ **(h)** $\frac{3}{5}$ or $\frac{2}{3}$

2 Put these fractions in order of size starting with the smallest.

(a) $\frac{1}{2}, \frac{5}{8}, \frac{7}{12}$ **(b)** $\frac{3}{4}, \frac{4}{5}, \frac{7}{10}$

3 Complete each of the following.

(a) $\frac{4}{8} = \frac{}{4}$ **(b)** $\frac{6}{18} = \frac{}{6}$ **(c)** $\frac{20}{24} = \frac{}{6}$

4 Complete each of the following.

(a) $\frac{30}{40} = \frac{6}{}$ **(b)** $\frac{24}{36} = \frac{4}{}$ **(c)** $\frac{18}{27} = \frac{2}{}$

5 Find the simplest form of each of the following.

(a) $\frac{4}{6}$ **(b)** $\frac{9}{12}$ **(c)** $\frac{5}{20}$ **(d)** $\frac{8}{10}$
(e) $\frac{18}{21}$ **(f)** $\frac{24}{45}$ **(g)** $\frac{7}{28}$ **(h)** $\frac{33}{55}$

6 Find the simplest form of each of the following.

(a) $\frac{4}{12}$ **(b)** $\frac{10}{20}$ **(c)** $\frac{6}{18}$ **(d)** $\frac{24}{40}$
(e) $\frac{45}{54}$ **(f)** $\frac{36}{60}$ **(g)** $\frac{40}{100}$ **(h)** $\frac{45}{75}$

3.5 Improper fractions

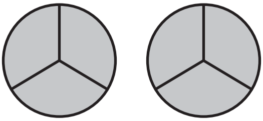

Figure 3.2

Figure 3.2 shows that there are 6 thirds in 2 wholes, that is, $\frac{6}{3} = 2$. Notice also that $\frac{6}{3}$ simplifies to $\frac{2}{1}$, so $\frac{2}{1} = 2$. You can think of the fraction line as a division sign, so $6 \div 3 = \frac{6}{3}$. Hence $6 \div 3 = \frac{6}{3} = \frac{2}{1} = 2$.

A fraction such as $\frac{6}{3}$, in which the numerator is greater than the denominator, is called an **improper fraction**. All the fractions you have met so far in this chapter have been **proper** fractions, in which the numerator is less than the denominator.

Insight

Improper fractions are also called 'top-heavy' fractions.

If the denominator divides exactly into the numerator the improper fraction is equal to a whole number, but this doesn't always happen.

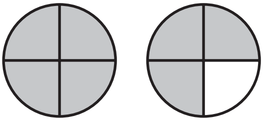

Figure 3.3

Figure 3.3 shows that there are 7 quarters in $1\frac{3}{4}$, that is, $\frac{7}{4} = 1\frac{3}{4}$. The number $1\frac{3}{4}$ is called a **mixed number**, because it is a mixture of a whole number and a proper fraction. To change $\frac{7}{4}$ to a mixed number, think of it as $7 \div 4$. Since 4 divides into 7 once with a remainder of 3, the whole number is 1 and the numerator of the fraction is 3. The denominator, 4, of the fraction is the same as in the improper fraction.

Example 3.5.1

Change $\frac{14}{3}$ to a mixed number.

Think of $\frac{14}{3}$ as $14 \div 3$. Then divide 14 by 3 to get 4 with a remainder of 2. Hence $\frac{14}{3} = 14 \div 3 = 4\frac{2}{3}$.

If you want to change an improper fraction which is not in its simplest form to a mixed number, express it in its simplest form and then change it to a mixed number.

To change a mixed number such as $3\frac{2}{5}$ to an improper fraction, you need to work out how many fifths there are in $3\frac{2}{5}$. There are 5 fifths in 1, so there are $3 \times 5 = 15$ fifths in 3. Adding the extra 2 fifths gives 17 fifths, which is $\frac{17}{5}$. So $3\frac{2}{5} = \frac{17}{5}$.

You would write down this process as $3\frac{2}{5} = \frac{(3 \times 5) + 2}{5} = \frac{17}{5}$.

EXERCISE 3.3

1 Write each of the following as a whole number.
(a) $\frac{30}{6}$ **(b)** $\frac{70}{10}$ **(c)** $\frac{42}{7}$ **(d)** $\frac{56}{8}$

2 Write the following as improper fractions in their simplest forms.
(a) 3 **(b)** 12

3 Change the following to mixed numbers.
(a) $13 \div 5$ **(b)** $19 \div 4$ **(c)** $25 \div 6$ **(d)** $15 \div 8$
(e) $\frac{20}{3}$ **(f)** $\frac{18}{5}$ **(g)** $\frac{17}{8}$ **(h)** $\frac{29}{9}$

4 Change the following to mixed numbers in their simplest forms.
(a) $\frac{14}{4}$ **(b)** $\frac{21}{6}$ **(c)** $\frac{25}{15}$ **(d)** $\frac{35}{14}$

5 Change the following to improper fractions.
(a) $2\frac{4}{5}$ **(b)** $4\frac{2}{3}$ **(c)** $3\frac{1}{2}$ **(d)** $5\frac{3}{4}$

3.6 Adding and subtracting fractions

Just as it is easy to order fractions with equal denominators, it is easy to add and subtract them.

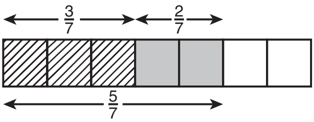

Figure 3.4

Figure 3.4 shows that $\frac{3}{7} + \frac{2}{7} = \frac{5}{7}$; that is, you add the numerators but you do not add the denominators. $\frac{3}{7} + \frac{2}{7}$ is *not* equal to $\frac{3}{7} + \frac{2}{7} = \frac{5}{14}$. The fraction $\frac{5}{7}$ is in its simplest form but sometimes you will be able to simplify your answer. You should do so if you can.

You can also subtract fractions with equal denominators. For example, $\frac{7}{9} - \frac{2}{9} = \frac{5}{9}$; that is, you subtract the second numerator from the first but, as with addition, the denominator stays the same.

If fractions have different denominators, convert them to equivalent fractions with equal denominators. To add $\frac{3}{5}$ and $\frac{1}{4}$, for example, convert both to equivalent fractions with denominator 20, the lowest common multiple of the denominators 5 and 4. (See Section 3.3.)

Then $\frac{3}{5} + \frac{1}{4} = \frac{12}{20} + \frac{5}{20} = \frac{17}{20}$.

Example 3.6.1

Work out $5\frac{2}{3} + 2\frac{1}{2}$.

Adding the whole numbers, $5 + 2 = 7$.

The lowest common multiple of the denominators 2 and 3 is 6.

Then $\frac{2}{3} + \frac{1}{2} = \frac{4}{6} + \frac{3}{6} = \frac{4+3}{6} = \frac{7}{6}$, and, as a mixed number $\frac{7}{6} = 1\frac{1}{6}$.

Finally, adding the two sums $7 + 1\frac{1}{6} = 8\frac{1}{6}$.

You can use the same method for *some* subtractions.

Example 3.6.2

Work out $4\frac{2}{3} - 1\frac{3}{5}$.

Subtracting the whole numbers, $4 - 1 = 3$.

The lowest common multiple of the denominators is 15.

Then $\frac{2}{3} - \frac{3}{5} = \frac{10}{15} - \frac{9}{15} = \frac{10-9}{15} = \frac{1}{15}$, and $4\frac{2}{3} - 1\frac{3}{5} = 3 + \frac{1}{15} = 3\frac{1}{15}$.

If you try to use this method with $4\frac{3}{5} - 1\frac{2}{3}$, the subtraction $\frac{9}{15} - \frac{10}{15}$ arises, so that you have to subtract a larger fraction from a smaller. There are ways around this difficulty, but the safest is to start by changing both mixed numbers to improper fractions. The disadvantage of this method is that the numerators can get large.

Example 3.6.3

Work out $4\frac{3}{5} - 1\frac{2}{3}$.

As improper fractions $4\frac{3}{5} = \frac{23}{5}$ and $1\frac{2}{3} = \frac{5}{3}$.

The lowest common multiple of the denominators is 15.

Then $\frac{23}{5} - \frac{5}{3} = \frac{69}{15} - \frac{25}{15} = \frac{69-25}{15} = \frac{44}{15} = 2\frac{14}{15}$.

EXERCISE 3.4

Work out the following. Give every answer in its simplest form.

1 **(a)** $\frac{3}{5} + \frac{1}{5}$ **(b)** $\frac{4}{7} + \frac{2}{7}$ **(c)** $\frac{3}{10} + \frac{1}{10}$ **(d)** $\frac{7}{12} + \frac{1}{12}$
(e) $\frac{7}{9} + \frac{4}{9}$ **(f)** $\frac{7}{8} + \frac{3}{8}$ **(g)** $\frac{3}{10} + \frac{7}{10}$ **(h)** $\frac{9}{10} + \frac{7}{10}$

2 **(a)** $\frac{4}{5} - \frac{1}{5}$ **(b)** $\frac{5}{7} - \frac{3}{7}$ **(c)** $\frac{5}{8} - \frac{3}{8}$ **(d)** $\frac{5}{6} - \frac{1}{6}$

3 **(a)** $\frac{1}{3} + \frac{2}{5}$ **(b)** $\frac{3}{8} + \frac{1}{4}$ **(c)** $\frac{1}{6} + \frac{1}{2}$ **(d)** $\frac{2}{3} + \frac{5}{6}$
(e) $\frac{2}{3} + \frac{3}{5}$ **(f)** $\frac{1}{6} + \frac{4}{9}$ **(g)** $\frac{7}{8} + \frac{5}{12}$ **(h)** $\frac{5}{6} + \frac{1}{2}$

4 **(a)** $2\frac{1}{4} + 3\frac{4}{5}$ **(b)** $4\frac{3}{5} + 1\frac{3}{10}$ **(c)** $5\frac{1}{2} + 2\frac{3}{10}$ **(d)** $1\frac{5}{6} + 3\frac{3}{4}$

5 **(a)** $3\frac{7}{8} - 1\frac{1}{4}$ **(b)** $4\frac{2}{3} - 2\frac{1}{2}$ **(c)** $3\frac{5}{6} - 2\frac{3}{4}$ **(d)** $4\frac{2}{3} - 1\frac{5}{6}$

3.7 Multiplication of fractions

Multiplication by a whole number is the same as repeated addition. Thus, $3 \times 2 = 3 + 3$ and $\frac{3}{7} \times 2 = \frac{3}{7} + \frac{3}{7} = \frac{6}{7}$. The numerator, 3, has been multiplied by the whole number, 2, but the denominator is unchanged. A common error is to multiply *both* the numerator and the denominator by 2 but this gives a fraction, $\frac{6}{14}$, which is *equivalent* to $\frac{3}{7}$.

Example 3.7.1

Work out $\frac{5}{6} \times 8$.

Multiplying the numerator by the whole number, $\frac{5}{6} \times 8 = \frac{40}{6}$.

As a mixed number $\frac{40}{6} = 6\frac{4}{6}$, which, in its simplest form, is $6\frac{2}{3}$.

The multiplication $\frac{4}{5} \times \frac{2}{3}$ is illustrated in Figure 3.5, where you can see that $\frac{8}{15}$ of the total square is shaded. Thus $\frac{4}{5} \times \frac{2}{3} = \frac{8}{15}$. This suggests that to multiply two fractions you multiply the numerators and then you multiply the denominators.

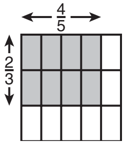

Figure 3.5

Multiplying a fraction by a whole number is similar to multiplying one fraction by another fraction. Writing the whole number as an improper fraction, with the whole number as the numerator and 1 as the denominator, enables you to write, for example, 10 as $\frac{10}{1}$. Then $\frac{3}{4} \times 10 = \frac{3}{4} \times \frac{10}{1} = \frac{30}{4} = \frac{15}{2} = 7\frac{1}{2}$.

To multiply two mixed numbers, change them to improper fractions.

Example 3.7.2

Work out $2\frac{1}{4} \times 1\frac{2}{5}$.

$2\frac{1}{4} \times 1\frac{2}{5} = \frac{9}{4} \times \frac{7}{5} = \frac{9 \times 7}{4 \times 5} = \frac{63}{20} = 3\frac{3}{20}$.

3.8 Fractions of quantities

One half of 10 is obviously 5. Therefore $\frac{1}{2}$ of 10 is 5.

But $\frac{1}{2} \times 10 = \frac{1}{2} \times \frac{10}{1} = \frac{1 \times 10}{2 \times 1} = \frac{10}{2} = 5$.

This suggests that, in mathematics, ‘of’ means multiply. You can use this to find fractions of physical quantities.

Example 3.8.1

Work out $\frac{2}{3}$ of 27 miles.

$\frac{2}{3}$ of $27 = \frac{2}{3} \times 27 = \frac{2}{3} \times \frac{27}{1} = \frac{54}{3} = 18$, so $\frac{2}{3}$ of 27 miles is 18 miles.

EXERCISE 3.5

Work out the following. Give every answer in its simplest form.

1 **(a)** $\frac{2}{3} \times 5$ **(b)** $\frac{4}{5} \times 15$ **(c)** $\frac{3}{4} \times 10$ **(d)** $40 \times \frac{7}{8}$

2 **(a)** $\frac{3}{4} \times \frac{5}{7}$ **(b)** $\frac{2}{3} \times \frac{7}{8}$ **(c)** $\frac{5}{6} \times \frac{9}{10}$ **(d)** $\frac{3}{5} \times \frac{2}{9}$

3 **(a)** $2\frac{1}{2} \times 1\frac{3}{4}$ **(b)** $1\frac{7}{8} \times 2\frac{4}{5}$ **(c)** $2\frac{2}{5} \times 3\frac{1}{3}$ **(d)** $\frac{5}{8} \times 2\frac{1}{3}$

(e) $4\frac{1}{2} \times \frac{4}{9}$ **(f)** $\frac{5}{6} \times 1\frac{1}{2}$ **(g)** $1\frac{2}{5} \times \frac{3}{7}$ **(h)** $4 \times 3\frac{1}{2}$

4 **(a)** $\frac{1}{6}$ of 30 **(b)** $\frac{5}{8}$ of 56 **(c)** $\frac{3}{5}$ of 45 **(d)** $\frac{3}{4}$ of 44

5 **(a)** $\frac{2}{3}$ of 48 kg **(b)** $\frac{3}{10}$ of 70 cm **(c)** $\frac{4}{5}$ of 60 litres

3.9 Division of fractions

The calculation 12 ÷ 3 means ‘what do you multiply three by to get 12?’ Similarly, $5 \div \frac{1}{2}$ means ‘what do you multiply one half by to get 5?’ There are 2 halves in 1; so there are $5 \times 2 = 10$ halves in 5.

By first writing the whole numbers as an improper fraction, $5 \div \frac{1}{2} = \frac{5}{1} \div \frac{1}{2} = 10$.

However, you also know that $5 \times 2 = \frac{5}{1} \times \frac{2}{1} = \frac{10}{1} = 10$. This suggests that dividing by $\frac{1}{2}$ is the same as multiplying by 2.

Now think about $6 \div \frac{3}{4}$. This means 'how many times do you multiply $\frac{3}{4}$ to get 6?' First notice that $\frac{4}{3} \times \frac{3}{4} = \frac{12}{12} = 1$, so you multiply $\frac{3}{4}$ by $\frac{4}{3}$ to get 1, so you must multiply $\frac{3}{4}$ by $6 \times \frac{4}{3}$ to get 6. Therefore

$$6 \div \frac{3}{4} = 6 \times \frac{4}{3} = \frac{6}{1} \times \frac{4}{3} = \frac{24}{3} = 8.$$

This suggests a way to divide one fraction by another. You turn the second fraction upside down and multiply the first fraction by it.

Example 3.9.1

Work out **(a)** $\frac{2}{3} \div \frac{3}{4}$, **(b)** $1 \div \frac{4}{5}$.

(a) $\frac{2}{3} \div \frac{3}{4} = \frac{2}{3} \times \frac{4}{3} = \frac{8}{9}$.

(b) $1 \div \frac{4}{5} = \frac{1}{1} \times \frac{5}{4} = \frac{5}{4} = 1\frac{1}{4}$.

The result of dividing a number into 1 is called the **reciprocal** of the number. So $\frac{5}{4}$ is the reciprocal of $\frac{4}{5}$. You can also work out that $\frac{4}{5}$ is the reciprocal of $\frac{5}{4}$ so $\frac{5}{4}$ and $\frac{4}{5}$ are reciprocals of each other.

Insight

To find the reciprocal of a fraction you can simply turn it upside down. Notice that when you multiply a number by its reciprocal the result is 1. For example, $\frac{5}{4} \times \frac{4}{5} = 1$.

Note that, as with multiplication, mixed numbers must always be changed to improper fractions before dividing.

Example 3.9.2

Work out $3\frac{1}{2} \div \frac{7}{12}$.

$3\frac{1}{2} \div \frac{7}{12} = \frac{7}{2} \times \frac{12}{7} = \frac{84}{14} = 6$.

EXERCISE 3.6

Work out the following. Give every answer in its simplest form.

1 **(a)** $2 \div \frac{1}{3}$ **(b)** $4 \div \frac{3}{4}$ **(c)** $6 \div \frac{2}{3}$ **(d)** $2 \div \frac{3}{5}$

2 **(a)** $\frac{4}{5} \div \frac{3}{8}$ **(b)** $\frac{5}{12} \div \frac{3}{4}$ **(c)** $\frac{7}{9} \div \frac{2}{3}$ **(d)** $\frac{3}{5} \div \frac{8}{9}$
(e) $\frac{2}{5} \div 3$ **(f)** $\frac{7}{10} \div \frac{4}{5}$ **(g)** $\frac{11}{12} \div \frac{3}{8}$ **(h)** $\frac{4}{9} \div \frac{7}{12}$

3 **(a)** $2\frac{1}{2} \div 1\frac{7}{8}$ **(b)** $2\frac{3}{4} \div 3\frac{1}{3}$ **(c)** $2\frac{5}{8} \div \frac{7}{16}$ **(d)** $3 \div 4\frac{1}{2}$
(e) $2\frac{7}{10} \div 1\frac{1}{3}$ **(f)** $1\frac{3}{4} \div 2\frac{1}{4}$ **(g)** $\frac{3}{8} \div 1\frac{4}{5}$ **(h)** $2\frac{3}{5} \div \frac{2}{3}$

3.10 A number as a fraction of another number

Fractions are a useful way of comparing two quantities. For example, 90° is often called a $\frac{1}{4}$ turn, because there are 360° in a complete turn and $\frac{90}{360} = \frac{1}{4}$ in its simplest form. Similarly, 45 minutes is $\frac{3}{4}$ hour, as there are 60 minutes in an hour and $\frac{45}{60} = \frac{3}{4}$ in its simplest form. The numerator and the denominator must be in the same units.

Example 3.10.1

Express, in its simplest form, 36 minutes as a fraction of 2 hours.

There are 120 minutes in 2 hours. The required fraction is $\frac{36}{120} = \frac{3}{10}$.

Insight

When you write one quantity as a fraction of another, you must make sure that they both have the same units of measurement. So in the above example it would be incorrect to say that the answer is $\frac{36}{2}$, since one quantity is in minutes and the other in hours.

EXERCISE 3.7

1 Express the first quantity as a fraction of the second. Give each answer in its simplest form.

(a)	8 minutes, 1 hour	**(b)**	15 hours, 1 day
(c)	5 days, 2 weeks	**(d)**	48 seconds, 2 minutes
(e)	50 minutes, 2 hours	**(f)**	6 days, 3 weeks
(g)	60 cm, 1 metre	**(h)**	18 hours, 4 days
(i)	70 cm, 2 metres	**(j)**	14 hours, 1 week

2 On a jury there are 8 women and 4 men. What fraction of the jury are women?

3 Find a fraction which is halfway between $\frac{3}{5}$ and $\frac{4}{5}$.

4 Two fifths of the employees in a company drive to work, one third travel by bus and the rest walk. Find the fraction who walk.

5 Last Friday, Tony worked for $7\frac{1}{2}$ hours. Express this time as a fraction of a day.

6 The sum of two mixed numbers is $6\frac{3}{20}$. One of the mixed numbers is $2\frac{3}{4}$. Work out the other one.

7 The product of two fractions is $\frac{7}{15}$. One of the fractions is $\frac{4}{5}$. Work out the other one.

8 An angle is $\frac{2}{3}$ of a turn. Express this in degrees.

9 When a petrol tank is $\frac{7}{12}$ full, it contains $5\frac{1}{4}$ gallons. How many gallons does it hold when full?

THINGS TO REMEMBER

- *A fraction is a type of number written as a numerator over a denominator.*

- *Equivalent fractions are fractions that are equal in value.*

- *Cancelling is when you divide both the numerator and denominator of a fraction by the same number, in order to make another fraction equivalent to the original.*

- *A fraction is in its lowest terms or simplest form when it can no longer be cancelled, i.e. when there is no number (other than 1) that divides exactly into both the numerator and the denominator.*

- *The lowest common multiple of two numbers is the smallest multiple that both numbers have in common.*

- *A proper fraction has a numerator which is less than the denominator. For example,* $\frac{5}{7}$.

- *An improper fraction has a numerator which is greater than the denominator. For example,* $\frac{7}{5}$.

- *To add or subtract fractions:*
  Step 1 *Write them as equivalent fractions with the same denominators.*
  Step 2 *Add or subtract the numerators, leaving the denominator as it is.*

- *In mathematics, 'of' means multiply.*

- *To multiply fractions you multiply the numerators and then you multiply the denominators.*
  For example, $\frac{7}{5} \times \frac{2}{3} = \frac{14}{15}$.

- *The reciprocal of a number is the result of dividing that number into 1.*

- *To divide one fraction by another you turn the second fraction upside down and then multiply the first fraction by it. For example,* $\frac{7}{5} \div \frac{2}{3} = \frac{7}{5} \times \frac{3}{2} = \frac{21}{10}$.

4

Two-dimensional shapes

In this chapter you will learn:

- ***about different types of triangles and polygons***
- ***how to use the angle sums of triangles and polygons***
- ***how to construct triangles***
- ***about the symmetries of triangles and polygons***
- ***about tessellations.***

4.1 Introduction

Two-dimensional shapes are flat shapes, such as squares and circles. They are also called **plane** shapes; the geometry which deals with them is called **plane geometry**. Thales of Miletus, a Greek, laid the foundations of plane geometry in the sixth century BC. About 200 years later, Euclid's *Elements* contained a complete, rigorous treatment of it.

4.2 Triangles

The simplest two-dimensional shape is the **triangle**, a three-sided shape with three angles (Figure 4.1). As a triangle made of rods cannot be deformed, it is widely used in buildings and other structures.

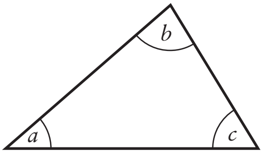

Figure 4.1

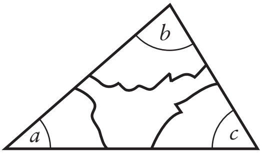

Figure 4.2

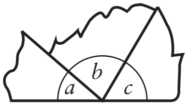

Figure 4.3

Draw a triangle on paper and label its angles *a*, *b* and *c*; see Figure 4.2. Tear off the corners and fit angles *a*, *b* and *c* together, as in Figure 4.3. They fit along a straight line. From Chapter 2, the three angles on a straight line add up to 180°, so $a + b + c = 180°$. This suggests that the angle sum of a triangle is 180°.

However, this demonstration only works for the triangle that you drew. It does not show that the sum of the angles in a triangle is 180° for any triangle. To do that, you need a proof.

In Figure 4.4, the side *AB* of triangle *ABC* has been extended and, at *B*, a line has been drawn parallel to *AC*. Angles *f*, *g* and *h* lie along a straight line, so $f + g + h = 180°$.

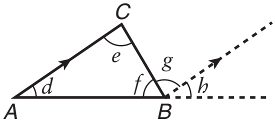

Figure 4.4

As angles *g* and *e* are alternate, $g = e$. As angles *h* and *d* are corresponding, $h = d$. (See Section 2.5.) So $f + e + d = 180°$.

Therefore the angle sum of a triangle is 180°.

The triangle in Figure 4.5 is **acute-angled**, as its angles are all less than 90°. An **obtuse-angled triangle** (Figure 4.6) has an obtuse angle, and a **right-angled triangle** (Figure 4.7) has a right angle.

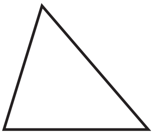

Figure 4.5

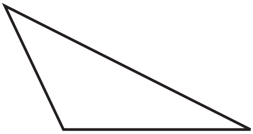

Figure 4.6

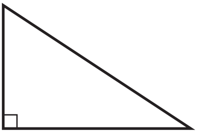

Figure 4.7

There are three other types of triangle. An **equilateral triangle** has three equal sides and angles. As the angle sum is 180°, each angle is 180° ÷ 3 = 60°. In Figure 4.8, the marked sides show that they are equal.

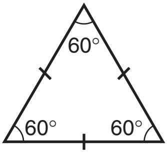

Figure 4.8

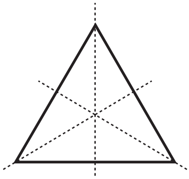

Figure 4.9

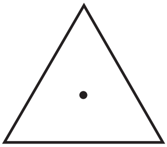

Figure 4.10

An equilateral triangle has 3 **lines of symmetry**, shown dotted in Figure 4.9. If the triangle is folded along one of its lines of symmetry, one half of the triangle will fit exactly on top of the other half. An alternative name for a line of symmetry is a **mirror line**, as one half of the triangle is a mirror image, or **reflection**, of the other half.

An equilateral triangle has **rotational symmetry** of **order** 3. This is the number of ways a tracing of the shape will fit on top of itself without turning it over. In Figure 4.10, the point of the triangle about which the tracing paper turns is called the **centre of rotational symmetry**.

A triangle which has two equal sides is called an **isosceles triangle**. In an isosceles triangle the angles opposite the equal sides are also equal (see Figure 4.11). The reverse is also true: if two angles of a triangle are equal, then the sides opposite them are also equal, so the triangle is isosceles. 'Isosceles' is a Greek word, meaning literally 'equal legs'.

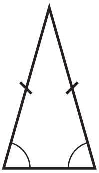

Figure 4.11

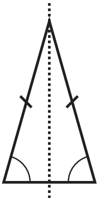

Figure 4.12

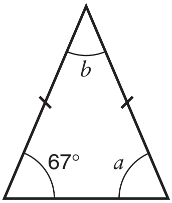

Figure 4.13

An isosceles triangle generally has one line of symmetry (Figure 4.12) and rotational symmetry of order 1. That is, a tracing of the triangle will fit on top of itself in only one way. In the special case when the isosceles triangle is equilateral, it has three lines of symmetry.

Example 4.2.1

In Figure 4.13, find the sizes of the angles marked *a* and *b*.

The triangle is isosceles, so the angles opposite the equal sides are equal, and $a = 67°$.

As the angle sum of the triangle is 180°, $b = 180° - 2 \times 67° = 46°$.

Triangles with no equal sides and no equal angles are **scalene triangles**. They have no lines of symmetry and rotational symmetry of order 1.

EXERCISE 4.1

In questions 1–9, find the size of each of the angles marked with letters. In some questions you will need to use facts about angles from Chapter 2. The triangles are not drawn to scale.

1

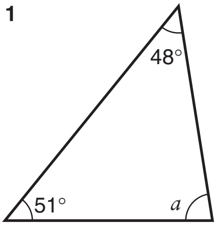

2

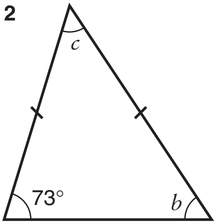

3

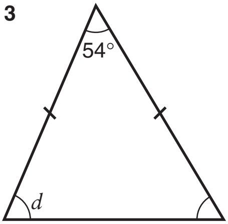

4

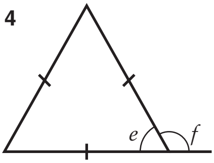

5

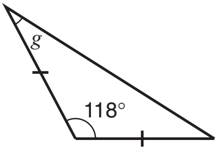

6

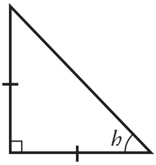

(Contd)

7

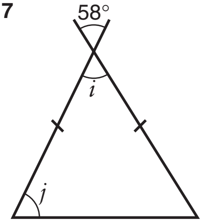

8

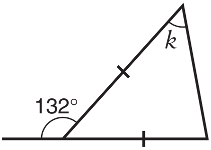

9

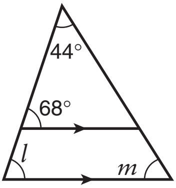

10 One of the angles of an isosceles triangle is 74°. The other two angles are x and y. Find two possible pairs of values for x and y.

11 One of the equal sides of an isosceles triangle is extended, and the angle between the extended side and the other equal side is 40°. Find the angles in the triangle.

4.3 Constructing triangles

Given three suitable measurements, you can make an accurate drawing of a triangle with a ruler, a pair of compasses and a protractor. If the corners, or **vertices**, of the triangle are labelled A, B and C, as in Figure 4.14, the length of the side joining A and B is denoted by AB.

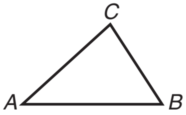

Figure 4.14

Example 4.3.1

In triangle ABC, $AB = 3$ cm, $AC = 2.5$ cm and $BC = 2.7$ cm. Make an accurate drawing of the triangle.

Choose one of the sides, say AB, as the base, draw a line 3 cm long and label its ends A and B. Set your compasses to a radius of 2.5 cm and, with the point on A, draw an arc (part of a circle). With the point of the compasses on B draw a second arc of radius 2.7 cm (see Figure 4.15) to cut the first arc at C. Draw a line from A to C and a line from B to C to complete triangle ABC. See Figure 4.16.

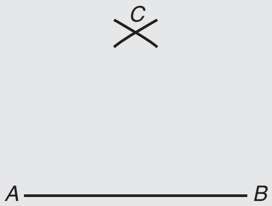

Figure 4.15

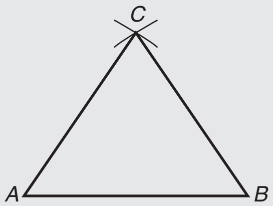

Figure 4.16

You can construct a triangle if you are given the length of one side and the sizes of two angles. If necessary, you can, of course, work out the size of the third angle of the triangle.

You can also construct a triangle if you are given the lengths of two sides and the size of the angle between the two given sides.

If you are given two sides and an angle other than the included angle, you can sometimes construct two different triangles from the information. (See Exercise 4.2, question 9.)

If you are given three angles whose sum is 180° and no side, you can construct an infinite number of triangles from the information.

EXERCISE 4.2

In questions 1–7, make an accurate drawing of the triangle ABC.

1. $AB = 4.2$ cm, $AC = 3.7$ cm, $BC = 3.1$ cm.
2. $AB = 3.6$ cm, $AC = 2.8$ cm, $BC = 5.2$ cm.
3. $AB = 4.3$ cm, angle $A = 59°$, angle $B = 42°$.
4. $AB = 4.8$ cm, angle $A = 124°$, angle $B = 21°$.
5. $AB = 4.3$ cm, angle $A = 25°$, angle $C = 27°$.
6. $AB = 5.1$ cm, $AC = 4.8$ cm, angle, $A = 43°$.
7. $AB = 4.5$ cm, $BC = 3.9$ cm, angle $B = 136°$.
8. What happens if you try to construct a triangle ABC with $AB = 6$ cm, $AC = 3$ cm and $BC = 2$ cm? Find a rule which tells you whether, given three lengths, it is possible to construct a triangle with sides of these lengths.

(Contd)

9 Triangle ABC has $AB = 5.7$ cm, angle $A = 46°$ and $BC = 4.6$ cm. Construct **two** different triangles from this information.

10 Construct a triangle ABC with angle $A = 68°$ and angle $B = 49°$.

4.4 Quadrilaterals

A **quadrilateral** is a shape with four straight sides and four angles. You can investigate the sum of the angles of a quadrilateral in a way similar to the one you used for a triangle.

Draw a quadrilateral and label its angles a, b, c and d (Figure 4.17). Tear off the corners and fit the angles together at a point (Figure 4.18). From Chapter 2, the angles at a point add up to 360°. This suggests that $a + b + c + d = 360°$ and that the angle sum of a quadrilateral is 360°.

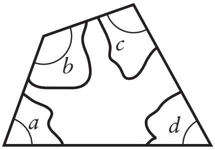

Figure 4.17

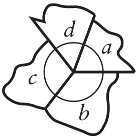

Figure 4.18

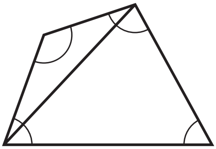

Figure 4.19

To *prove* this, draw a diagonal of a quadrilateral, dividing it into two triangles (see Figure 4.19). The angle sum of each triangle is 180°, so **the angle sum of a quadrilateral is 360°.**

A **square** is a quadrilateral with four equal sides and four right angles (Figure 4.20). A square has 4 lines of symmetry (Figure 4.21), and rotational symmetry of order 4.

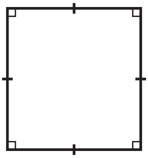

Figure 4.20

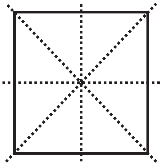

Figure 4.21

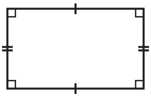

Figure 4.22

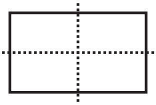

Figure 4.23

A **rectangle** has four right angles and its opposite sides are equal in length (Figure 4.22). A rectangle has 2 lines of symmetry (Figure 4.23), and rotational symmetry of order 2.

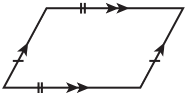
Figure 4.24

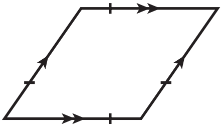
Figure 4.25

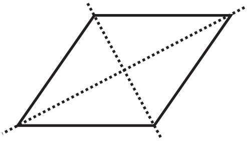
Figure 4.26

A **parallelogram** has opposite sides which are both parallel and equal in length (see Figure 4.24). A parallelogram has no lines of symmetry, and rotational symmetry of order 2.

A **rhombus** (Figure 4.25) is a parallelogram with equal sides. It has 2 lines of symmetry (Figure 4.26), and rotational symmetry of order 2.

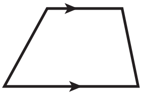
Figure 4.27

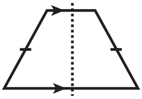
Figure 4.28

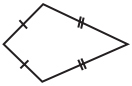
Figure 4.29

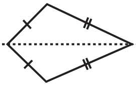
Figure 4.30

A **trapezium** (Figure 4.27) has one pair of parallel sides. It generally has no lines of symmetry and has rotational symmetry of order 1. An **isosceles trapezium** (Figure 4.28), in which the non-parallel sides are equal in length, has one line of symmetry.

A **kite** (Figure 4.29) has two pairs of adjacent sides with equal lengths, like two isosceles triangles with their bases joined. A kite has 1 line of symmetry (Figure 4.30), and rotational symmetry of order 1.

EXERCISE 4.3

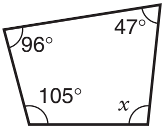

1. In the figure, work out the size of the angle marked x.
2. Write down the names of two quadrilaterals with rotational symmetry of order 2.
3. Write down the names of three quadrilaterals whose diagonals cross at right angles.
4. A quadrilateral has m lines of symmetry. Write down all the possible values of m.
5. A quadrilateral has rotational symmetry of order n. Write down all the possible values of n.

(Contd)

6 Construct an accurate drawing of quadrilateral *ABCD* from the information given below. All lengths are in centimetres.
(a) $AB = 3.7$, $\angle A = 78°$, $AD = 2.1$, $BC = 3.1$, $CD = 4.6$.
(b) $AB = 4.9$, $\angle A = 114°$, $BD = 5.7$, $CD = 3.9$, $BC = 3.2$.
(c) $AB = 4.1$, $\angle BAD = 98°$, $\angle ABD = 34°$, $CD = 2.9$, $BC = 4.6$.
(d) $AB = 3.8$, $AD = 2.9$, $BD = 4.5$, $\angle BDC = 31°$, $CD = 4.7$.

4.5 Polygons

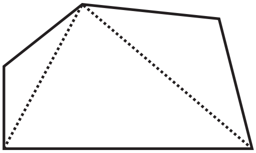

Figure 4.31

A **polygon** is a shape with three or more straight sides. To calculate the angle sum of a polygon, split it into triangles. For the five-sided polygon in Figure 4.31, choose any vertex, draw as many diagonals as you can from that vertex, in this case two, splitting the five-sided polygon into three triangles. The angle sum of a triangle is 180°, so the angle sum of the five-sided polygon is 3 × 180°, or 540°.

In general, the number of triangles into which you can split a polygon in this way is two less than the number of sides of the polygon.

The table shows the angle sums and the names of some polygons.

Number of sides	*Name*	*Number of triangles*	*Angle sum in degrees*
3	Triangle	1	180
4	Quadrilateral	2	360
5	Pentagon	3	540
6	Hexagon	4	720
8	Octagon	6	1080
10	Decagon	8	1440

A polygon whose sides and angles are equal is called **regular**. You can find the size of each angle of a regular polygon by dividing the sum of its angles by the number of its angles.

Insight

Be careful: when a question says 'pentagon' it means a five-sided shape and not necessarily a regular pentagon.

4.6 Interior and exterior angles

So far in your study of polygons, the angles you have been given or have had to find have been *inside* the polygon. These angles are called **interior angles**. If you extend a side, as in Figure 4.32, the angle *outside* the shape is called an **exterior angle**.

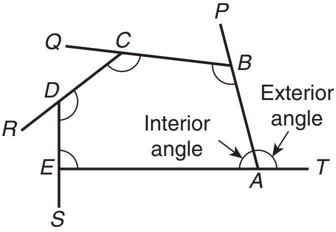

Figure 4.32

There is a surprising fact about the sum of the exterior angles of a polygon. Suppose that you are walking round the polygon, starting from *A* and facing *B*. Walk to *B*, and when you get there, turn from the direction *BP* to the direction *BC*. Now walk to *C*, and when you get there, turn from the direction *CQ* to the direction *CD*. Continue in this way, until you come back to *A*, and you turn from *AT* to face *B*, and you are back where you started, and facing in the same direction. On the way, you have turned through an angle equal to the angle sum of the exterior angles, but you have also turned through 360° because you are facing the same direction as before. Moreover, this argument applies to any polygon, not just the pentagon in Figure 4.32.

Therefore, the sum of the exterior angles of any polygon is 360°.

This fact is often useful in questions involving regular polygons, because the exterior angles of a regular polygon are all equal.

Insight

In Figure 4.32 you can see that the interior angle and exterior angle of a polygon lie on a straight line. It follows that at a vertex the exterior angle + the interior angle = 180°.

Example 4.6.1

Work out the size of
(a) each exterior angle,
(b) each interior angle of a regular pentagon.

(a) A regular pentagon has five equal exterior angles whose sum is 360°, so each exterior angle is $360° \div 5 = 72°$.
(b) At each vertex, interior angle + exterior angle = 180°. Thus every interior angle is $180° - 72° = 108°$.

Example 4.6.2

The size of each exterior angle of a regular polygon is 24°.
(a) Calculate the interior angle of the polygon.
(b) How many sides has the polygon?

(a) As each exterior angle is 24°, the interior angles are each $(180 - 24)°$, which is 156°.
(b) The sum of exterior angles is 360° and each exterior angle is 24°, so the number of sides is $360 \div 24 = 15$.

EXERCISE 4.4

In questions 1–3, work out the sizes of the angles marked with letters.

1

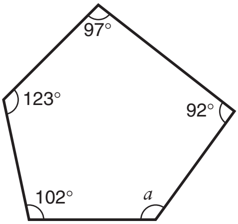

2

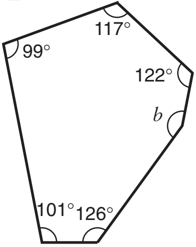

3

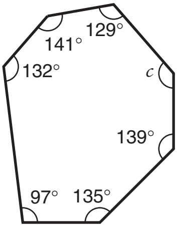

4 Work out the sum of the angles of a 12-sided polygon.
5 Work out the sum of the angles of a 15-sided polygon.

6 Work out the size of each angle of a regular pentagon.

7 Work out the size of each angle of a regular decagon.

8 Work out the size of each angle of a regular 18-sided polygon.

9 The size of an interior angle of a polygon is 109°. Work out the size of the exterior angle at the same vertex.

10 Work out the size of each exterior angle of a regular hexagon.

11 Work out the size of **(a)** each exterior angle and **(b)** each interior angle of a regular 12-sided polygon.

12 The size of each interior angle of a regular polygon is 140°. How many sides has the polygon?

13 The size of each interior angle of a regular polygon is 168°. How many sides has the polygon?

4.7 Symmetries of regular polygons

You already know that an equilateral triangle (a regular polygon with 3 sides) has 3 lines of symmetry and that a square (a regular polygon with 4 sides) has 4 lines of symmetry. Figure 4.33 shows the lines of symmetry drawn on regular polygons with 5, 6, 7 and 8 sides.

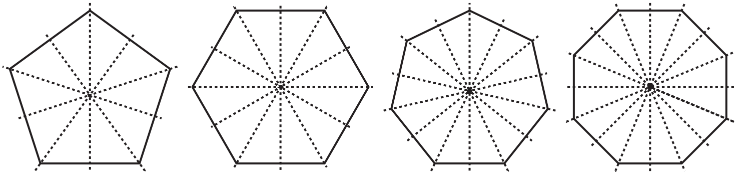

Figure 4.33

In general, the number of lines of symmetry of a regular polygon and the order of rotational symmetry of a regular polygon are both the same as the number of its sides.

4.8 Congruent shapes

If you make a tracing of one of the triangles in Figure 4.34 overleaf you will find that the tracing will fit exactly on top of each of the other triangles.

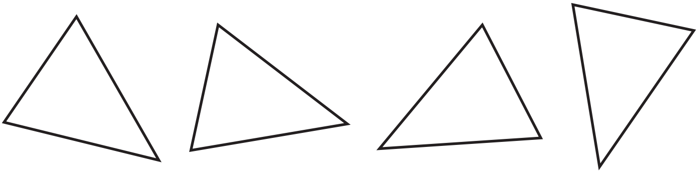

Figure 4.34

The triangles in Figure 4.34 are exactly the same shape and size and they are said to be **congruent.**

If you turn the tracing paper over and the tracing fits exactly on top of a triangle, that triangle is also congruent to the original one. Use tracing paper to check that the triangles in Figure 4.35 are also congruent.

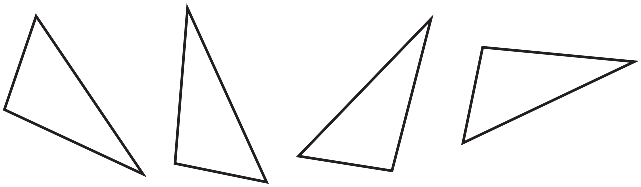

Figure 4.35

Shapes other than triangles may also be congruent to each other.

EXERCISE

EXERCISE 4.5

1 Write down the letters of the pairs of congruent shapes in the diagram.

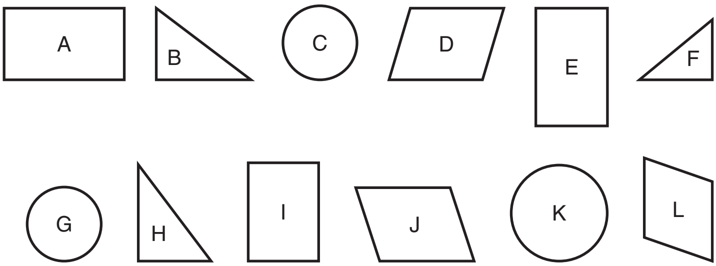

4.9 Tessellations

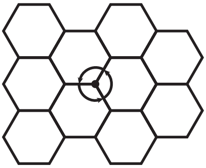

Figure 4.36

Figure 4.36 shows how regular hexagons can fit together without any gaps and without overlaps. Bees' honeycombs are an example of this.

Shapes which fit together like this are said to **tessellate** and the tiling patterns they make are called **tessellations**.

'Tessellate' and 'tessellations' are derived from the Latin word for the small square stones used to make mosaic patterns.

Each interior angle of a regular hexagon is 120° so three corners fit together to make 360°.

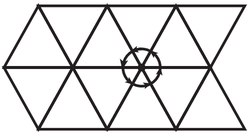

Figure 4.37

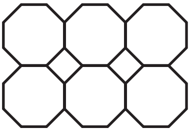

Figure 4.38

Figure 4.37 shows how equilateral triangles tessellate. (In fact, *all* triangles tessellate.) Each angle of an equilateral triangle is 60° so six corners fit together to make 360°.

Regular octagons do not tessellate but a combination of regular octagons and squares will tessellate, as Figure 4.38 shows. Each interior angle of a regular octagon is 135° and two of these corners fit together with a right angle to make 360°. This does not guarantee that the regular octagons and squares will tessellate, but if it were not true, they certainly wouldn't tessellate.

Insight

There is a wide variety of beautiful tessellations, ranging from intricate Islamic mosaics to the work of the Dutch artist M. C. Escher.

EXERCISE 4.6

1 Draw diagrams to show how each of these shapes tessellates. Show at least six shapes in each diagram.
   (a) isosceles triangle **(b)** square
   (c) parallelogram **(d)** kite

2 Without drawing, explain why rectangles tessellate.

3 Without drawing, explain why regular nonagons (nine-sided polygons) do not tessellate.

4 Draw a diagram with at least ten shapes to show a tessellation using a combination of regular hexagons and equilateral triangles.

5 Explain, without a drawing, why it may be possible to make a tessellation using a combination of equilateral triangles and regular 12-sided polygons. Then, using a drawing, show that the tessellation is possible.

THINGS TO REMEMBER

Shape	*Shape name*	*Properties*
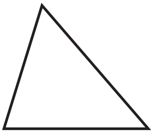	Scalene or acute-angled triangle	All angles less than 90º
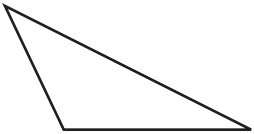	Scalene or obtuse-angled triangle	One obtuse angle
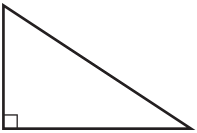	Scalene or right-angled triangle	One right angle
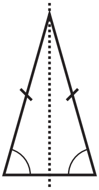	Isosceles triangle	2 equal sides and angles 1 line of symmetry
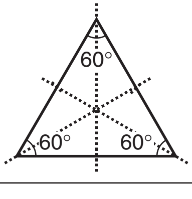	Equilateral triangle	3 equal sides All angles 60º 3 lines of symmetry Rotational symmetry of order 3

(Contd)

Shape	*Shape name*	*Properties*
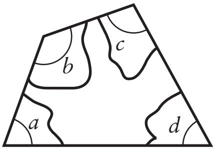	Quadrilateral	Any 4-sided shape
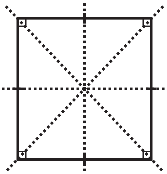	Square	4 equal sides 4 right angles 4 lines of symmetry Rotational symmetry of order 4
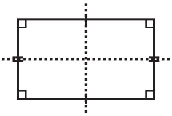	Rectangle	Opposite sides equal in length 4 right angles 2 lines of symmetry Rotational symmetry of order 2
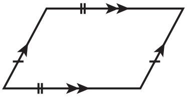	Parallelogram	Opposite sides parallel and equal in length No lines of symmetry Rotational symmetry of order 2

Shape	*Shape name*	*Properties*
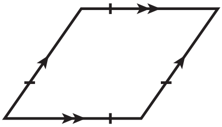	Rhombus	4 equal sides Opposite sides parallel 2 lines of symmetry Rotational symmetry of order 2
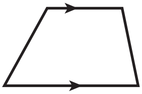	Trapezium	One pair of parallel sides No lines of symmetry Rotational symmetry of order 1
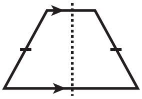	Isosceles trapezium	One pair of parallel sides Non-parallel sides are equal in length 1 line of symmetry Rotational symmetry of order 1
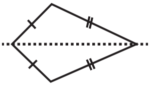	Kite	Adjacent sides with equal lengths 1 line of symmetry Rotational symmetry of order 1

- *This table shows the angle sums and the names of some polygons.*

Number of sides	*Name*	*Number of triangles*	*Angle sum in degrees*
3	Triangle	1	180
4	Quadrilateral	2	360
5	Pentagon	3	540
6	Hexagon	4	720
8	Octagon	6	1080
10	Decagon	8	1440

- *A regular polygon has all sides equal and all angles equal.*
- *The sum of the exterior angles of any polygon is 360°.*
- *The sum of the exterior angle and interior angle at a vertex is 180°.*
- *Congruent shapes are exactly the same shape and size.*
- *Shapes which fit together without overlapping or leaving any gaps tessellate and the resulting tiling pattern is called a tessellation.*

5

Decimals

In this chapter you will learn:
- ***about place value in decimal numbers***
- ***how to convert between decimals and fractions***
- ***how to add and subtract decimals***
- ***how to multiply and divide decimals***
- ***about terminating and recurring decimals.***

5.1 Introduction

The invention of decimal fractions, usually called simply decimals, was a gradual process. It started around the twelfth century and was completed in 1585, when Simon Stevin, a Flemish scientist, published a treatment called *La Disme* (English title: *The Art of Tenths*).

Stevin's symbolism was, however, very complicated and, although numbers written with a decimal point had first appeared in print in 1492, decimals were not generally adopted until the late seventeenth century. In fact, even today, the symbolism is not standardized. While the United Kingdom and the United States use a decimal point, many European countries use a comma instead.

Although Lord Randolph Churchill, Sir Winston's father, complained that he 'never could make out what those damned dots meant', the use of decimals brought many benefits, including making commercial processes easier and, in mathematics, the development of logarithms.

5.2 Place value

In Chapter 1, you saw how, in the decimal system, the place of a digit in a number tells you the value of that digit. The value of the digit in each of the first three columns is

hundreds tens units

Each value is $\frac{1}{10}$ of the value on its left. You can continue this pattern to the right of the units column with column values of **tenths**, **hundredths** and **thousandths**.

The value of each digit in the number 654.739, which is read as six hundred and fifty-four **point** seven three nine, is shown in the table.

Hundreds	Tens	Units		tenths	hundredths	thousandths
6	5	4	.	7	3	9

The **decimal point** acts as a marker which separates the whole-number part from the part which is less than 1.

For example, the value of the 7 is $\frac{7}{10}$ and the value of the 9 is $\frac{9}{1000}$.

The number 654.739 has three **decimal places**. The digit 7 is in the first decimal place; 3 is in the second decimal place and 9 is in the third.

With decimals which are less than 1, it is usual, but not essential, to write a zero on the left of the decimal point. Thus, you write 0.53 rather than .53 to draw attention to the decimal point.

If two decimal numbers have equal whole-number parts, you can find which is larger by comparing their tenths digits, their hundredths digits and so on. For example, 4.23, which has a tenths digit of 2, is greater than 4.19, which has a tenths digit of 1. It does not matter that 4.19 has a greater hundredths digit than 4.23.

Example 5.2.1

Which decimal fraction is greater, 0.506 or 0.512?

Start from the decimal point and work to the right. The tenths digits are equal. 0.506 has a hundredths digit of 0 and 0.512 has a hundredths digit of 1. As 1 is greater than 0, 0.512 is greater than 0.506.

The symbol > means ‘is greater than’. Example 5.2.1 shows that 0.512 > 0.506. The symbol < means ‘is less than’. Thus 0.506 < 0.512.

Insight

It can be confusing to decide whether you should use < or > between two numbers. Remember that the smaller end of the symbol points towards the smaller number, so 6 < 7 or 9 > 5. Or, think of the symbol being an ‘open mouth’ which opens towards the greater amount.

EXERCISE 5.1

1 Write these numbers in a table like that on the opposite page and, in each case, state the value of the digit 3.
(a) 174.783 **(b)** 936.214 **(c)** 472.329 **(d)** 579.038

2 Write down the value of the underlined digit in each of these decimal numbers.
(a) $2\underline{3}.045$ **(b)** $7.\underline{8}1$ **(c)** $0.8\underline{9}2$ **(d)** $0.0\underline{1}9$

3 **(a)** How many decimal places are there in the number 231.67?
(b) Write down the digit which is in the second decimal place.

4 **(a)** How many decimal places are there in the number 0.609?
(b) Write down the digit which is in the second decimal place.

5 Find the greater decimal fraction of each of these pairs.
(a) 0.899, 0.901 **(b)** 0.702, 0.72
(c) 0.01, 0.009 **(d)** 0.101, 0.909

6 Write down the largest number from this list.
2.365 5.362 5.632 5.236 3.625

7 Write down the smallest number from this list.
0.786 0.876 0.678 0.687 0.768

(Contd)

8 Write these numbers in order of size. Start with the smallest.
1.32 1.302 1.203 2.31 1.23

9 Write these numbers in order of size. Start with the smallest.
0.190 0.910 0.109 0.901 0.019

10 Insert either < or > between each of these pairs of number to give a true statement.
(a) 4.76, 6.47 **(b)** 5.72, 5.27
(c) 0.39, 0.4 **(d)** 0.03, 0.029

5.3 Converting decimals to fractions

You can use place value to convert decimals to fractions.

For example, $0.7 = \frac{7}{10}$ and $0.4 = \frac{4}{10}$.

The fraction $\frac{7}{10}$ is already in its simplest form, but you can simplify $\frac{4}{10}$ to get $0.4 = \frac{2}{5}$. (See Section 3.4.)

You can extend this method to numbers, such as 0.37, with two digits on the right of the decimal point.

$$0.37 = \frac{3}{10} + \frac{7}{100} = \frac{30}{100} + \frac{7}{100} = \frac{37}{100}$$

You can write decimals of this type straight down as ‘hundredths’. You should always give your final answer in its simplest form. The fraction $\frac{37}{100}$ is already in its simplest form, but the simplest form can be a fraction with a denominator smaller than 100. For example, $0.45 = \frac{45}{100}$, which is $\frac{9}{20}$ in its simplest form.

Similarly, numbers with three digits on the right of the decimal point may be written straight down as ‘thousandths’ and then simplified, if possible. For example, $0.139 = \frac{139}{1000}$, which cannot be simplified.

When you convert a decimal number greater than 1 to a mixed number, the whole-number part stays the same and you convert the decimal fraction to a fraction. For example, $5.3 = 5\frac{3}{10}$.

5.4 Converting fractions to decimals 1

If the fraction has a denominator which is 10, 100, 1000, etc., you can use place value to convert the fraction directly to a decimal. For example, $\frac{27}{100} = 0.27$.

Even if the fraction does not have 10 or 100 or 1000 as its denominator, you can sometimes convert it to an equivalent fraction which does. (See Section 3.2.) For example, $\frac{3}{20} = \frac{15}{100} = 0.15$.

When you convert a mixed number to a decimal, the whole-number part stays the same and you convert the fraction to a decimal fraction. For example, $4\frac{3}{50} = 4\frac{6}{100} = 4.06$.

The methods you have met so far in this section apply only to fractions with a denominator of 10 or 100 or 1000 or with a denominator which is a factor of one of these numbers. Later in this chapter you will meet a method which may be used to convert any fraction to a decimal.

EXERCISE 5.2

1 Convert each of these decimals to a fraction in its simplest form.
(a) 0.6 **(b)** 0.1 **(c)** 0.48 **(d)** 0.39
(e) 0.09 **(f)** 0.179 **(g)** 0.125 **(h)** 0.003

2 Convert each of the following to a mixed number with the fraction in its simplest form.
(a) 3.8 **(b)** 4.22 **(c)** 54.75 **(d)** 23.048

3 Convert each of these fractions to a decimal.
(a) $\frac{3}{10}$ **(b)** $\frac{71}{100}$ **(c)** $\frac{7}{100}$ **(d)** $\frac{387}{1000}$
(e) $\frac{17}{1000}$ **(f)** $\frac{9}{1000}$ **(g)** $\frac{2}{5}$ **(h)** $\frac{19}{20}$
(i) $\frac{17}{25}$ **(j)** $\frac{11}{50}$ **(k)** $\frac{237}{500}$ **(l)** $\frac{7}{200}$

4 Convert each of these mixed numbers to a decimal.
(a) $6\frac{1}{10}$ **(b)** $9\frac{3}{100}$ **(c)** $5\frac{41}{1000}$ **(d)** $1\frac{9}{20}$
(e) $8\frac{8}{25}$ **(f)** $12\frac{37}{200}$

5.5 Addition and subtraction

If you line up the decimal points underneath each other, adding and subtracting decimals is very similar to adding and subtracting whole numbers.

For example,

with whole numbers

$$\begin{array}{rl} 354 & \\ \underline{2432} & + \\ \underline{2786} & \end{array}$$

and with decimals

$$\begin{array}{rl} 3.54 & \\ \underline{24.32} & + \\ \underline{27.86} & \end{array}$$

Example 5.5.1

Work out **(a)** $9.7 + 6.83$ **(b)** $12.4 + 0.72$ **(c)** $0.71 + 6.092$

(a)
$$\begin{array}{rl} 9.7 & \\ \underline{6.83} & + \\ \underline{16.53} & \end{array}$$

(b)
$$\begin{array}{rl} 12.4 & \\ \underline{0.72} & + \\ \underline{13.12} & \end{array}$$

(c)
$$\begin{array}{ll} 0.71 & \\ \underline{6.092} & + \\ \underline{6.802} & \end{array}$$

Insight

You may find it helpful to put in zeros to the right of the last digit, so that all the numbers have the same number of decimal places. For example, in Example 5.5.1 part (a), you can write 9.7 as 9.70, which helps you to align the decimals.

Example 5.5.2

Work out **(a)** $6.9 - 2.4$ **(b)** $12.3 - 4.7$ **(c)** $19.04 - 6.2$

(a)
$$\begin{array}{rl} 6.9 & \\ \underline{2.4} & - \\ \underline{4.5} & \end{array}$$

(b)
$$\begin{array}{rl} 12.3 & \\ \underline{4.7} & - \\ \underline{7.6} & \end{array}$$

(c)
$$\begin{array}{rl} 19.04 & \\ \underline{6.2} & - \\ \underline{12.84} & \end{array}$$

If the calculations are more difficult than those given, you should use your calculator.

EXERCISE 5.3

1 Work out the answers to the following questions.

(a) $3.2 + 4.7$ **(b)** $8.6 + 4.52$ **(c)** $9.82 + 0.48$
(d) $23.8 + 0.27$ **(e)** $9.7 - 3.4$ **(f)** $24.3 - 7.5$
(g) $26.37 - 19.2$ **(h)** $32.4 - 17.57$ **(i)** $6 + 4.7$
(j) $9 - 6.8$ **(k)** $12 - 0.12$ **(l)** $7.9 - 4.53$
(m) $7.6 - 3.749$ **(n)** $3.86 - 0.372$ **(o)** $8.4 + 3.685$

2 Work out the difference between 8.43 and 9.2.

3 The judges awarded a skater the scores 5.8, 5.7, 6, 5.4, 6 and 5.9. Work out the skater's total score.

4 The runner-up in a 100-metres race had a time of 10.37 seconds, which was 0.09 seconds behind the winner. Find the winner's time.

5.6 Multiplication of decimals

To multiply a decimal by a whole number less than 10, you should set out your working so that the decimal point in the answer is directly below the decimal point in the question. Then you can do the multiplication as if you were multiplying two whole numbers together. For example, the working for 3.6 × 4 is shown on the right.

$$\begin{array}{r} 3.6 \\ 4\ \times \\ \hline 14.4 \\ \hline \end{array}$$

To multiply decimals by 10 or 100 or 1000, there are some simple rules, which you will remember more easily if you see how they are obtained. The diagram illustrates 32.69 × 10.

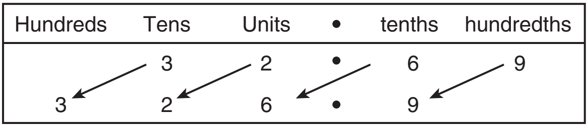

The digits have moved one place to the left, as their place values are 10 times bigger but, when the result is written down without the column headings as 32.69 × 10 = 326.9, it looks as if the decimal point has moved one place to the right.

Using one of these rules, you can write down the result of multiplying a number by 10 without working. For example, 5.732 × 10 = 57.32. Your answer may be a whole number. For example, 38.7 × 10 = 387.

Multiplication of a decimal by 100 may be approached in a similar way. For example, 4.173 × 100 = 417.3.

Insight

You may have learnt the rule 'to multiply by 10, add zero', but this rule works with whole numbers and not with decimals. For example, 0.5 × 10 = 5 not 0.50. Moving the digits one place to the left works for both whole numbers and decimals.

The multiplication of two decimal numbers may be introduced by converting them both to fractions. For example, to work out 0.3 × 0.7, write it as $\frac{3}{10} \times \frac{7}{10}$. Using the method you met in Section 3.7, the result is $\frac{21}{100}$ which as a decimal is 0.21; that is, 0.3 × 0.7 = 0.21.

Notice that 3 × 7 = 21. Notice also that 0.3 and 0.7 each have one decimal place and that 0.21 has two decimal places; that is, the number of decimal places in the answer is the same as the total number of decimal places in the question. This approach enables you to write down the answers to similar products, such as 0.6 × 0.5 = 0.30, which may be written as 0.3, and 0.3 × 0.2 = 0.06 (not 0.6).

Example 5.6.1

Write down the answer to

(a) 0.6 × 0.9 **(b)** 0.5 × 0.8 **(c)** 0.3 × 0.3

(a) 0.6 × 0.9 = 0.54 **(b)** 0.5 × 0.8 = 0.40 (or 0.4)
(c) 0.3 × 0.3 = 0.09

To work out a more difficult product, such as 0.4 × 0.27, you could convert each decimal to a fraction and multiply them: $\frac{4}{10} \times \frac{27}{100} = \frac{108}{1000}$, which is 0.108 as a decimal, so 0.4 × 0.27 = 0.108.

There is a better method: 4 × 27 = 108. Also, 0.4 has one decimal place, 0.27 has two decimal places and 0.108 has three decimal places. So, again, the number of decimal places in the answer is the same as the total number of decimal places in the question.

This suggests a general method for multiplication involving decimals.

Step 1 Ignore the decimals and multiply the numbers together.
Step 2 Find the total number of decimal places in the numbers you are multiplying.
Step 3 The answer must have the same number of decimal places.

Here the steps are used to calculate 0.07 × 0.008.

Step 1 The calculation is 7 × 8 = 56.
Step 2 0.07 has 2 decimal places and 0.008 has 3 decimal places, so the answer must have 2 + 3 = 5 decimal places.
Step 3 The answer has five decimal places, therefore 0.07 × 0.008 = 0.000 56.

Example 5.6.2

Write the answers to **(a)** 0.02 × 0.04 **(b)** 0.05 × 0.8

(a) 0.02 × 0.04 = 0.0008 **(b)** 0.05 × 0.8 = 0.040 or 0.04

If one of the numbers you want to multiply is a whole number, you must remember that a whole number has no decimal places.

If the calculations are any more complicated than those given so far, you are advised to use a calculator.

EXERCISE 5.4

1 Work out
(a) 4.7 × 6 **(b)** 17.8 × 5 **(c)** 18.37 × 8 **(d)** 0.403 × 3

2 Write down the answers to
(a) 8.21 × 10 **(b)** 0.6 × 10 **(c)** 0.47 × 10
(d) 0.702 × 10 **(e)** 28.3 × 10 **(f)** 0.0041 × 10
(g) 5.218 × 100 **(h)** 0.83 × 100 **(i)** 0.4 × 100
(j) 0.0007 × 100 **(k)** 46.36 × 1000 **(l)** 3.9 × 1000

3 Complete these.
(a) 0.21 × ... = 21 **(b)** 6.71 × ... = 67.1
(c) 7.98 × ... = 7980

4 Write down the answers to
(a) 0.9 × 0.7 **(b)** 0.6 × 0.5 **(c)** 0.2 × 0.4
(d) 0.1 × 0.3 **(e)** 0.07 × 0.3 **(f)** 0.03 × 0.02

5 Work out
(a) 3.7 × 1.4 **(b)** 0.28 × 0.41 **(c)** 6.8 × 0.37
(d) 0.084 × 0.47 **(e)** 9.3 × 37 **(f)** 0.049 × 23

6 1 litre = 1.76 pints. How many pints are there in 6 litres?

7 1 gallon = 4.5 litres. How many litres are there in 8 gallons?

8 1 metre = 100 centimetres. How many centimetres are there in 8.32 metres?

5.7 Division of decimals

To divide a decimal by a whole number less than 10, you should, as with multiplication, set out your working so that the decimal point in the answer is directly below the decimal point in the question.

Then you can do the division as if you were dividing one whole number by another. For example, the working for 13.5 ÷ 5 is shown on the right.

$$\begin{array}{r} 5)\,\underline{13.5} \\ 2.7 \end{array}$$

To divide decimals by 10 or 100 or 1000 you can use simple rules which are the opposite of the ones you used for multiplication by these numbers. The diagram below illustrates 574.6 ÷ 10.

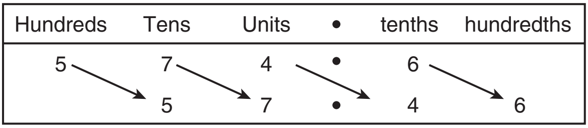

The *digits* have moved one place to the *right,* as their place values are one tenth of what they were, but when the result is written down without the column headings as 574.6 ÷ 10 = 57.46 it looks as if the *decimal point* has moved one place to the *left.*

Using one of these rules, you should be able to write down the result of dividing a decimal by 10 without any working. For example, 34.19 ÷ 10 = 3.419. You may need to write zeros to show that columns are empty. For example, 0.93 ÷ 10 = 0.093.

To divide a decimal by 100 you move the decimal point *two* places to the *left,* and to divide a decimal by 1000 you move the decimal point *three* places to the *left.*

To divide by a decimal, write the division as a fraction, as the line in a fraction may be thought of as a division sign. (See Section 3.5.) Then, by multiplying by 10 or 100 or 1000, convert it to an equivalent fraction with a whole number for the denominator, as you already know how to divide by a whole number. The numerator does not have to be a whole number.

For example, to divide 0.12 by 0.3, write it as a fraction $\frac{0.12}{0.3}$. Then multiply both the numerator and denominator by 10 and obtain $\frac{1.2}{3}$, which you can evaluate to get the answer 0.4.

If the calculations are any more complicated than those given so far, you are advised to use a calculator.

Example 5.7.1

Work out 2.8 ÷ 0.04.

As a fraction, $2.8 \div 0.04 = \frac{2.8}{0.04}$. Multiplying both the numerator and denominator by 100, $\frac{2.8}{0.04} = \frac{280}{4}$. Finally, $\frac{280}{4} = 70$.

EXERCISE 5.5

1 Work out
(a) 9.2 ÷ 4 **(b)** 19.23 ÷ 3 **(c)** 2.382 ÷ 6 **(d)** 0.469 ÷ 7

2 Write down the answers to
(a) 37.5 ÷ 10 **(b)** 457.2 ÷ 10 **(c)** 0.093 ÷ 10
(d) 7.13 ÷ 10 **(e)** 0.34 ÷ 10 **(f)** 652.8 ÷100
(g) 2.5 ÷ 100 **(h)** 0.37 ÷ 100 **(i)** 0.003 ÷ 100
(j) 17.2 ÷ 100 **(k)** 33.8 ÷ 1000 **(l)** 0.071 ÷ 1000

3 Complete these.
(a) 3.7 ÷ ... = 0.37 **(b)** 0.9 ÷ ... = 0.009
(c) 7.9 ÷ ... = 0.079

4 Work out
(a) 39.92 ÷ 2 **(b)** 199.5 ÷ 7 **(c)** 0.624 ÷ 4
(d) 7.28 ÷ 8 **(e)** 257.6 ÷ 70 **(f)** 10.8 ÷ 300
(g) 43.75 ÷ 50 **(h)** 2.618 ÷ 400

5 Work out
(a) 0.36 ÷ 0.4 **(b)** 3 ÷ 0.2 **(c)** 4.8 ÷ 0.06
(d) 0.021 ÷ 0.7 **(e)** 9 ÷ 0.03 **(f)** 4 ÷ 0.08
(g) 7.38 ÷ 0.3 **(h)** 3.384 ÷ 0.09

6 A length of cable 18.48 metres long is cut into 4 equal pieces. Work out the length of each piece.

7 A pile of 100 sheets of paper is 2.1 centimetres high. Work out the thickness of each sheet of paper.

8 In completing one lap, a runner covers 0.4 kilometre. How many laps does she complete in a 10-kilometre race?

5.8 Converting fractions to decimals 2

In Section 5.4, you saw how to convert certain types of fractions to decimals. There is a general method which you can use to convert *any* fraction to a decimal. You have to think of the line in a fraction as a division sign. So, for example, $\frac{3}{4}$ means 3 ÷ 4.

You can write as many zeros as you like on the right of the decimal point.

$$\begin{array}{r} 4)\,\underline{3.00} \\ 0.75 \end{array}$$

The decimal 0.75 is called a **terminating decimal**, because it divides out exactly and the decimal terminates or stops, after two decimal places in this case.

To convert a mixed number to a decimal, you leave the whole number unchanged and use the division method to convert the fraction.

Example 5.8.1

Convert $3\frac{7}{8}$ to a decimal.

If you divide 7 by 8 you get 0.875, so $\frac{7}{8} = 0.875$.

Therefore $3\frac{7}{8} = 3 + 0.875 = 3.875$.

Not all fractions give terminating decimals when you convert them. Using the division method with $\frac{2}{3}$, for example, you find that the decimal repeats itself, and $\frac{2}{3} = 0.666...$, where the three dots means that the sixes repeat for ever.

This kind of decimal, which keeps repeating, is called a **recurring decimal.** To avoid writing down the same figure repeatedly, you can write it in a shortened form, using dot notation, with a dot above the repeating figure. The decimal 0.666... then becomes $0.\dot{6}$.

Example 5.8.2

Convert $\frac{5}{6}$ to a recurring decimal and write it in dot notation.

$$\begin{array}{r} 6)\,\underline{5.0000} \\ 0.8333 \end{array}$$

$\frac{5}{6} = 0.8\dot{3}$.

Sometimes, more than one number repeats. When this happens, dots are placed above the first and last figures in the repeating group. The fraction $\frac{3}{11}$, for example, written as a decimal is 0.272 727 ..., which you write as $0.\dot{2}\dot{7}$. Similarly, $\frac{17}{37} = 0.459\ 459\ 459... = 0.\dot{4}5\dot{9}$.

When writing recurring decimals in dot notation, you must make sure that you place the dots over the first and last of the group of numbers that repeat. So, for example, 0.58216216216... is written as $0.58\dot{2}1\dot{6}$.

When you convert a mixed number to a recurring decimal, you place the dots above the decimal part and leave the whole number unchanged, so that, in dot notation, 7.7777... is $7.\dot{7}$.

In Section 3.3, you saw how to compare the sizes of fractions using equivalent fractions. Although this method will work for any fractions, it would not be ideal for comparing fractions with large denominators such as $\frac{15}{19}$ and $\frac{18}{23}$. It is much easier to convert each of these to a decimal using a calculator. The full calculator reading for 15 ÷ 19 is 0.789 473 684, and for 18 ÷ 23 it is 0.782 608 696, but you only need to go as far as the third decimal place to see that the first decimal is greater than the second, so $\frac{15}{19} > \frac{18}{23}$.

EXERCISE 5.6

1 Convert each of these fractions to a decimal and state whether it is a terminating decimal or a recurring decimal.

(a) $\frac{1}{4}$ **(b)** $\frac{1}{3}$ **(c)** $\frac{3}{8}$ **(d)** $\frac{4}{9}$ **(e)** $\frac{7}{8}$

2 Using a calculator, convert each of these fractions to a recurring decimal and write it in dot notation.

(a) $\frac{9}{11}$ **(b)** $\frac{1}{11}$ **(c)** $\frac{5}{12}$ **(d)** $\frac{1}{12}$

(e) $\frac{19}{33}$ **(f)** $\frac{3}{37}$ **(g)** $\frac{23}{66}$ **(h)** $\frac{39}{74}$

3 Write each of these recurring decimals in dot notation.

(a) 0.8888... **(b)** 0.712 121 2... **(c)** 0.367 367...

(d) 0.456 896 89... **(e)** 18.018 18... **(f)** 0.564 356 43...

(g) 9.3999... **(h)** 263.6363...

4 Insert < or > between each of these pairs of fractions to make a true statement. You should use a calculator.

(a) $\frac{7}{11}, \frac{8}{13}$ **(b)** $\frac{14}{19}, \frac{17}{23}$ **(c)** $\frac{6}{13}, \frac{7}{15}$ **(d)** $\frac{13}{17}, \frac{16}{21}$

5 Use a calculator to put these numbers in order, smallest first.

(a) 5.8, $5\frac{7}{9}$, 5.77, 5.78, $5\frac{17}{22}$

(b) 3.1, $3\frac{1}{11}$, 3.01, 3.09, $3\frac{2}{21}$

(c) $4\frac{1}{12}$, 4.084, 4.083, $4\frac{2}{23}$, 4.085

6 Find a decimal with three decimal places which is between $\frac{16}{31}$ and $\frac{17}{33}$. Convert your decimal to a fraction in its simplest form.

7 **(a)** Convert these fractions to recurring decimals.
$\frac{1}{7}, \frac{2}{7}, \frac{3}{7}, \frac{4}{7}, \frac{5}{7}, \frac{6}{7}$
(b) How many figures are there in each of the recurring groups?
(c) Comment on the figures in the recurring groups.

8 **(a)** Which denominators of the fractions $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots, \frac{1}{24}, \frac{1}{25}$, when converted to decimals, give rise to terminating decimals?
(b) Comment on the numbers in your answer to part (a).
(Hint: find the factors of each denominator.)

THINGS TO REMEMBER

- *Each digit in a number has a place value, as shown in this table:*

Hundreds	Tens	Units		tenths	hundredths	thousandths
1	2	4	.	1	0	7

- *The decimal point acts as a marker which separates the whole-number part from the part which is less than 1.*

- *The symbol > means 'is greater than'. The symbol < means 'is less than'.*

- *You can use place value to convert a decimal to a fraction. For example,* $0.708 = \frac{708}{1000} = \frac{177}{250}$.

- *To convert a fraction to a decimal, divide the numerator by the denominator. So* $\frac{3}{4} = 3 \div 4 = 0.75$.

- *When adding or subtracting decimals, remember to line up the decimal points underneath each other. You may find it helpful to write in zeros to the right of the last digit, so that all the numbers have the same number of decimal places.*

- *To multiply decimals:*
  Step 1 *Ignore the decimal points and multiply the numbers together.*
  Step 2 *Find the total number of decimal places in the numbers you are multiplying.*
  Step 3 *The answer must have the same number of decimal places.*

- *To divide by a decimal:*
  Step 1 *Write the division as a fraction.*
  Step 2 *Multiply by 10 or 100 or 1000 to convert to an equivalent fraction with a whole number for the denominator.*
  Step 3 *Divide the numerator by the denominator.*

- *A terminating decimal has a finite number of decimal places, so the decimal ends or terminates.*
  For example, 0.25 and 0.245 556 are terminating decimals.

- *A recurring decimal has digits that keep repeating without terminating. You can use dot over the repeating digits to write down a recurring decimal.*
  For example, $\frac{17}{37} = 0.459\,459\,459\ldots = 0.\dot{4}5\dot{9}$.

6

Statistics 1

In this chapter you will learn:

- ***about the collection of statistical data***
- ***how to draw and use pictograms, bar charts, pie charts and line graphs***
- ***how to draw and use scatter graphs***
- ***about correlation***
- ***about discrete and continuous data.***

6.1 Introduction

Statistics is the branch of mathematics concerned with the collection, organization and analysis of data, often, but not always, numerical data. This chapter deals with the first two of these three aspects of statistics, with particular emphasis on the presentation of data. The analysis of data, using a variety of statistical averages, will be found in Chapter 22.

Starting with pictorial records on the walls of cave dwellings, the collection and recording of data is as old as civilization itself. Before 3000 BC, the Babylonians gathered information about crops and trade and so, centuries later, did the Egyptians and the Chinese. The Greeks held censuses, for taxation purposes, as early as 594 BC but the Romans were the first to gather a wide range of population, financial and agricultural data about their empire. The word ‘statistics’ itself is derived from the Latin word *statisticus,* meaning ‘of state affairs’.

The first real statistical study of population, *Observations on the London Bills of Mortality,* was published in 1662 and statistical methods subsequently assumed increasing importance, especially in the natural and social sciences.

Today, important decisions in many fields are made on the basis of statistics, although this reliance does have its critics. Benjamin Disraeli believed ‘There are three kinds of lies: lies, damned lies and statistics.’ Rex Todhunter Stout held a similar view: ‘There are two kinds of statistics, the kind you look up and the kind you make up.’

6.2 Collection of data

You can collect statistical data in three ways: by referring to existing source material, such as newspapers; by carrying out an experiment; and by conducting a **survey**. A survey might involve a whole population, as in a national census, but, to save time and money, it is more likely that you would survey part of the whole population, called a **sample**.

Insight

In statistics, it is important that a sample should be as representative of the whole population as possible in order to avoid bias. For example, if you wanted to find out what proportion of the population is trying to reduce their carbon footprint, then conducting a survey of people on a climate change rally would produce a bias in your results.

A **data collection sheet** or **tally chart** is used to organize the data from a survey. You can use it to collate data from completed questionnaires or you can record information directly onto it.

For example, you can record the colours of 30 cars in a table.

Colour	Tally	Frequency
Red	𝍸 𝍸 |||	13
Blue	||||	4
White	𝍸	5
Black	𝍸 |||	8

To make the tally marks easier to add up, group them in fives, drawing the fifth mark across the first four. The total of the tally marks for each colour is called the **frequency**, which is shown in the right-hand column. With this extra column, the table is called a **frequency table**. A frequency table need not have a column for the tally marks.

6.3 Pictograms

Data are often easier to understand if you present them in diagrammatic form. One example is a **pictogram** (or pictograph), in which you use pictures to show information. The frequency table below shows the numbers of cars using a car park on five days. You can illustrate the data in a pictogram by using the symbol 🚗 to represent 10 cars.

You can find how many symbols you need to represent the number of cars on each day by dividing the number of cars by 10 to get 3, 5, 4.5, 4.7 and 2.2. The pictogram is shown in Figure 6.1.

Day	Number of cars
Monday	30
Tuesday	50
Wednesday	45
Thursday	47
Friday	22

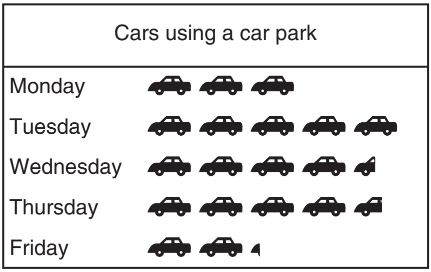

Figure 6.1

Only complete symbols are needed to represent the numbers of cars on Monday and Tuesday but, for each of the other days, parts of a symbol are needed in addition to complete symbols. You must use your judgement to draw the parts of symbols as accurately as possible.

6.4 Bar charts

Pictograms are attractive but time-consuming to draw and it can be awkward to draw parts of a symbol. A **bar chart** is a simpler way of showing information. Figure 6.2 is a bar chart showing the information about the parked cars in which the bars are drawn vertically. You could, if you wish, draw the bars horizontally instead as shown in Figure 6.3.

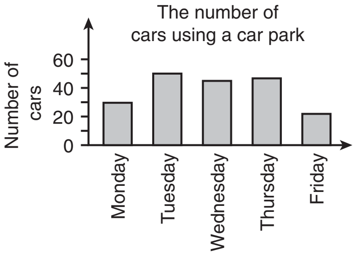

Figure 6.2

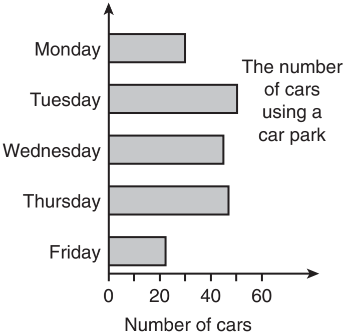

Figure 6.3

These two bar charts therefore show the same information.

Which version you choose may ultimately depend on which you prefer, and the space you have available.

EXERCISE 6.1

1 Thirty-two men were asked how they travelled to work. Their results are shown in the table. Complete the table.

How they travel	Tally	Frequency
Walk	𝍸 𝍸 ||||	
Car	||||	
Bus	𝍸	
Train	𝍸 ||||	

2 The numbers of children in 26 families are given below.

3	5	1	1	2	4	1	2	2	3	1	4	1
3	2	5	2	1	2	3	1	4	3	1	2	4

Collect this information in a frequency table.

3 The dress sizes of 25 women are shown in the frequency table.

Dress size	10	12	14	16	18	20
Number of women	1	5	7	6	4	2

Draw a pictogram to show this information. Use a symbol to represent one woman.

4 The numbers of students in three classes are shown in the table.

Class	Number of students
History	20
Geography	23
Art	31

Draw a pictogram to show this information. Use one symbol to represent 5 students.

5 The numbers of televisions in 100 houses are shown in the table.

Number of televisions	Number of houses
0	2
1	35
2	39
3	24

Draw a pictogram to show this information. Use a symbol to represent 10 houses.

6 The pictogram shows numbers of aircraft in four airlines.

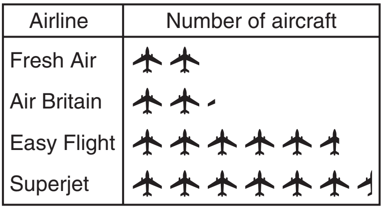

✈ represents 10 aircraft. Find the number of aircraft in each fleet.

(Contd)

7 The table shows the numbers of goals scored by ice hockey teams.

Number of goals	0	1	2	3	4	5
Number of teams	3	6	5	3	2	1

Draw a bar chart to show this information.

8 The passenger capacities of five types of Boeing aircraft are shown in the table. Draw a bar chart to illustrate this information.

Type of aircraft	737	747	757	767	777
Passenger capacity	106	426	195	252	281

9 100 men and 100 women in each of six age groups were asked if they wore spectacles. The numbers of men and women who said yes are shown in the table. Draw a bar chart to show this information.

Age group (years)	15–24	25–34	35–44	45–54	55–64	65+
Number of men	31	33	38	74	88	90
Number of women	37	42	44	79	87	89

10 The bar chart shows the lengths, in years, of the reigns of six British monarchs. Find the length of each of their reigns.

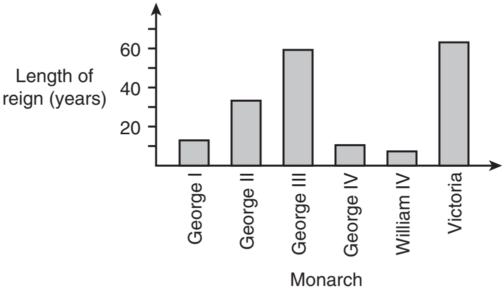

6.5 Pie charts

Information is often illustrated by means of pie charts. **Pie charts** show the proportions in which a quantity is shared out. You can draw a pie chart for the data about the car colours on page 82. The 30 cars are shown by a full circle, with an angle of 360° at its centre. One car is shown by an angle of 360° ÷ 30 = 12°. The 13 red cars are shown by an angle of 13 × 12° = 156°. You find the other angles similarly, getting 48°, 60° and 96° for the blue, white and black cars. With a protractor, draw the pie chart in Figure 6.4 with these angles at its centre. The 'slices' into which the circle is divided are called **sectors**.

The pie chart on its own does not tell you how many cars there were of each colour; it shows the proportion with each colour.

An alternative method for finding the angles is to use fractions. For example, 13 of the 30 cars are red, so the angle representing these cars is $\frac{13}{30}$ of 360° = $\frac{13}{30} \times 360° = 156°$.

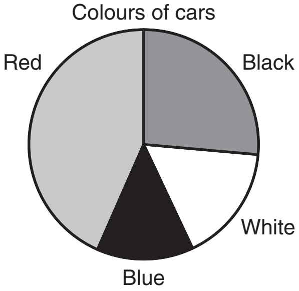

Figure 6.4

The first method is more straightforward to use if the total number of items is a factor of 360. If the numbers are more awkward the fraction method is easier. Where appropriate, give angles to the nearest degree.

Insight

The three-dimensional pie charts and bar charts available in many software packages may look attractive, but are often misleading. A 'slice' at the front of a three-dimensional pie chart may appear to have a greater volume than it should, while a 'slice' at the back will be foreshortened and appear too small.

Example 6.5.1

The pie chart shows information about eye colours of 150 people.

(a) What fraction of the people have grey eyes? Give your fraction in its simplest form.

(b) How many people have brown eyes?

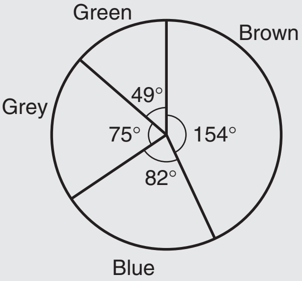

(a) The fraction is $\frac{75}{360}$ which is $\frac{5}{24}$ in its simplest form.

(b) The number of people with brown eyes is $\frac{154}{360} \times 150 = 64$.

This answer is given to the nearest whole number.

EXERCISE 6.2

1 The table shows how a woman spent 24 hours. Draw a pie chart to show this information.

Activity	Sleep	Work	Leisure	Meals	Travel
Number of hours	8	7	6	1	2

2 The favourite types of food of 60 people are shown in the table. Draw a pie chart to show this information.

Type of food	English	Italian	Indian	Chinese	Greek
Number of people	11	9	19	14	7

3 A football team played 38 matches. They won 20, drew 11 and lost 7. Draw a pie chart to show this information.

4 Helen watched television for 3 hours 15 minutes. The figure gives information about the programmes she watched.
(a) How long did the news last?
(b) How long did the film last?
(c) For what fraction of her time did Helen watch sport? Give your answer in its simplest form.

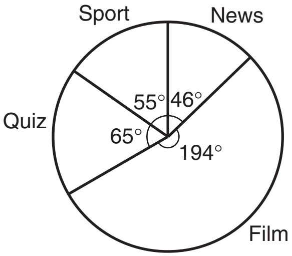

6.6 Line graphs

The points plotted on the grid in Figure 6.5 show the temperature every four hours throughout a day.

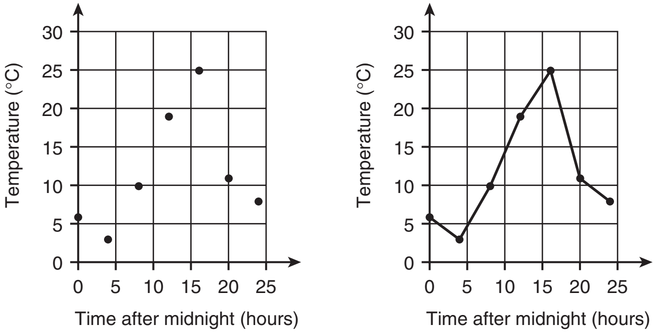

Figure 6.5 *Figure 6.6*

As the temperature changes gradually, you can join the points with straight lines. You then obtain the **line graph** in Figure 6.6.

Line graphs are often used to show data obtained over a period – these are sometimes called **time-series graphs**. You can use a line graph to read off values between those actually plotted. For example, from Figure 6.6, you can estimate the temperature at 2 p.m. as 22°C, but you must recognize that this can only be an estimate.

Joining points with straight lines is only appropriate when a quantity is changing gradually. You have to be even more careful about predicting future data by extending such a line graph.

The points plotted on the grid in Figure 6.7 show the average number of hours of sunshine each day in Budapest in six months of the year.

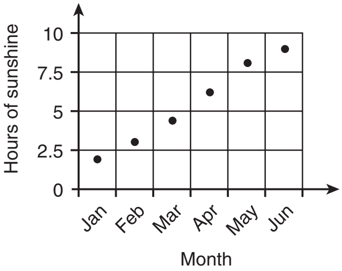

Figure 6.7

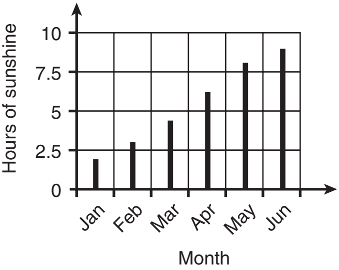

Figure 6.8

Points in between the plotted points have no meaning. A bar chart or a **vertical line graph**, as in Figure 6.8, would be a suitable way of showing this information.

When a quantity is changing gradually, rather than fluctuating, the plotted points may be joined either with straight lines or with a curve.

Example 6.6.1

The temperature, in °C, at five-minute intervals of a pan of hot water cooling down is shown in the table.

Time (min)	0	5	10	15	20	25	30
Temperature (°C)	100	60	42	33	28	25	23

(a) Plot the values in the table and join the points with a curve.
(b) Use your graph to estimate
  (i) the temperature of the water after 9 minutes,
  (ii) the time when the temperature was 66°C.

(a) The graph is an example of a **cooling curve**.
(b) (i) Reading up from 9 on the time axis and across to the temperature axis, the temperature at 9 minutes was 44°C.
  (ii) Reading across from 66 on the temperature axis and down to the time axis, the temperature was 66°C after 4 minutes.

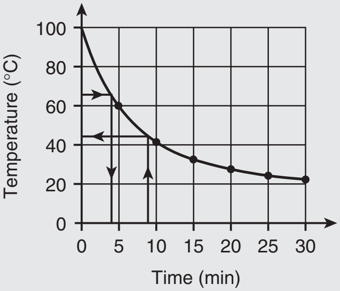

EXERCISE 6.3

1 The height, in centimetres, measured every two days over a 12-day period, of a plant grown from seed is shown in the table.

Time (days)	0	2	4	6	8	10	12
Height (cm)	0	2	5	8	11	13	14

(a) Draw a line graph to show this information.
(b) Use your graph to estimate the height of the plant after five days.

2 A tray of water was placed in the freezing compartment of a fridge. In the next hour, the temperature, in °C, of the water measured every ten minutes. The temperatures are shown in the table.

Time (min)	0	10	20	30	40	50	60
Temperature (°C)	21	13	7	5	4	3	2

(Contd)

(a) Draw a cooling curve for the water.
(b) Use your graph to estimate
  (i) the temperature after 5 minutes,
  (ii) the time when the temperature was 9°C.

3 A girl's height, in centimetres, measured every two years up to the age of fourteen is shown in the table.

Age (years)	0	2	4	6	8	10	12	14
Height (cm)	54	67	89	105	115	120	124	137

(a) Draw a line graph to show this information.
(b) Use your graph to estimate her height when she was aged five.

4 The population, to the nearest million, of the USA every 20 years from 1900 to 1980 is shown in the table.

Year	1900	1920	1940	1960	1980
Population (millions)	76	106	132	179	227

(a) Draw a line graph to show this information.
(b) Use your graph to estimate
  (i) the population in 1930,
  (ii) the year in which the population reached 200 million.

5 London's average monthly rainfall, in millimetres, in the first six months of the year is shown in the table.

Month	Jan	Feb	Mar	Apr	May	Jun
Average monthly rainfall (mm)	45	28	40	39	50	48

Draw a vertical line graph to show this information.

6.7 Scatter graphs

Scatter graphs show if there might be a relationship between two sets of data. For example, to investigate the relationship between women's heights and their weights, the data shown in the table were collected.

Height (cm)	167	152	175	163	164	169	155	171	160	156
Weight (kg)	66	54	72	63	67	69	57	72	64	61

You can record the data in the table on a grid, using height on one axis and weight on the other. This kind of graph, shown in Figure 6.9, is called a **scatter graph**.

Notice the zig-zags on the axes. They are put there to show that the axes are not graduated evenly up from zero.

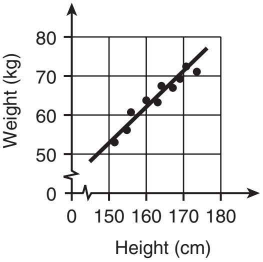

Figure 6.9

You can, by eye, draw a straight line, called a **line of best fit,** which passes near all these points. This line does not have to go through any of the points, although it can. There should be roughly equal numbers of points on each side of the line.

If you can draw a line of best fit, then there may be a relationship, called a **correlation**, between women's heights and their weights.

Weight increases as height increases; when one quantity increases as the other increases, the correlation is called **positive correlation.**

Correlation in which one quantity decreases as the other increases is called negative correlation. Figure 6.10 shows **negative correlation.**

Scatter graphs can show not only whether there is any correlation but also, if there is, how close the correlation is. On the scatter graph Figure 6.10, for example, all the points are close to the line of best fit and so this could be described as **high**, or strong, negative correlation.

When the scatter diagram has no pattern to it, as in Figure 6.11, there is no correlation between the two sets of data.

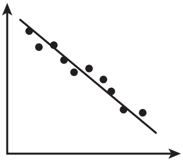

Figure 6.10

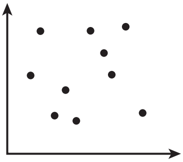

Figure 6.11

If all the points lie on a line, the correlation is called **perfect.** If the points are not close to the line of best fit, the correlation is called **low**, or weak. Figure 6.12 and Figure 6.13 show low positive correlation and low negative correlation.

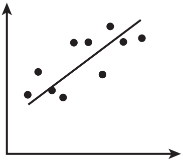

Figure 6.12

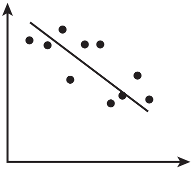

Figure 6.13

Example 6.7.1

The table shows the engine size in cubic centimetres (cc) and the fuel consumption in miles per gallon (mpg) of ten cars.

Engine size (cc)	1000	1100	1300	1400	1600
Fuel consumption (mpg)	61	59	54	52	51
Engine size (cc)	1800	2000	2500	2700	2900
Fuel consumption (mpg)	46	43	40	37	35

(a) Draw a scatter graph to show this information.
(b) Draw a line of best fit on your scatter graph.
(c) Describe the correlation.
(d) Estimate the fuel consumption of a car with a 2300 cc engine.

(a) and **(b)** These are in the diagram.
(c) There is a high negative correlation.
(d) Using the line of best fit, and following the arrow shown in the diagram, the petrol consumption of a car with a 2300 cc engine is about 42 mpg.

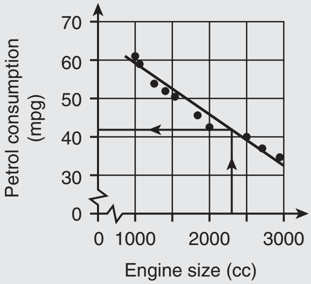

EXERCISE 6.4

1 The table shows the engine size and the top speed of ten cars in miles per hour (mph).

Engine size (cc)	1600	1800	2000	2500	2900
Top speed (mph)	111	121	129	130	140
Engine size (cc)	1400	1300	1100	1000	2700
Top speed (mph)	105	95	89	80	136

(a) Draw a scatter graph to show this information.
(b) Draw a line of best fit on your scatter graph.
(c) Describe the correlation.

2 The outdoors temperature, in °C, at noon on ten days and the number of units of electricity used in heating a house on each of those days are shown in the table.

Noon temperature (°C)	7	11	9	2	4	7	0	10	5	3
Units of electricity used	32	20	27	37	32	28	41	23	33	36

(a) Draw a scatter graph to show this information.
(b) Draw a line of best fit on your scatter graph.
(c) Describe the correlation.
(d) Use your line of best fit to estimate the number of units of electricity used in heating a house on a day when the outdoors temperature at noon was 8°C. *(Contd)*

3 The table shows the number of goals scored and the number of goals conceded by ten soccer teams. What can you deduce from it?

Number of goals scored	13	16	18	19	22	24	26	30	33	34
Number of goals conceded	20	26	15	11	23	18	11	25	12	19

6.8 Discrete and continuous data

There are two types of numerical statistical data: discrete data and continuous data.

Discrete data can take only exact values. Discrete data are often obtained by counting, in which case the values will be whole numbers. Examples of discrete data are the number of goals scored in an ice hockey match and the number of baked beans in a tin.

Continuous data are obtained by measuring, using an instrument such as a ruler or a thermometer, and cannot, therefore, be exact. Examples of continuous data are the length of a room and the weight of an apple.

Insight

Non-numerical data, such as eye colour or favourite film genre, is called **categorical data**. Categorical data is displayed using a bar chart, pie chart or pictogram.

EXERCISE

EXERCISE 6.5

1 State whether each of these types of data is discrete or continuous.

(a) The number of brothers and sisters you have.

(b) The time it takes you to walk a mile.

(c) The temperature in New York at noon.
(d) The score on dice.
(e) The number of seeds in a packet.
(f) Your weight.
(g) Your height.
(h) Your shoe size.
(i) The volume of wine in a bottle.
(j) The number of words on a page in this book.

6.9 Grouping data

Here are the marks gained by 40 candidates in an examination.

4	9	11	14	18	20	24	25	28	31
32	32	35	38	39	40	41	41	43	44
46	46	48	49	52	55	56	58	58	60
65	67	69	74	75	78	80	83	86	94

A vertical line graph shows the frequency of every individual mark but it does not give a helpful picture of the results, as it shows too much detail.

Grouping the marks gives a clearer picture. The marks could be grouped 0–19, 20–39, 40–59 and so on. These are called the **class intervals**. These intervals must not overlap. You could not, for example, use class intervals of 0–20, 20–40, 40–60, etc., as you would not know where to put a mark of 20. It could go in more than one class interval.

The number of candidates with marks in each class interval is shown in the **grouped frequency table**.

Mark	0–19	20–39	40–59	60–79	80–99
Frequency	5	10	14	7	4

You can show this information on the bar chart shown in Figure 6.14.

You could also draw without gaps between the bars, as in Figure 6.15.

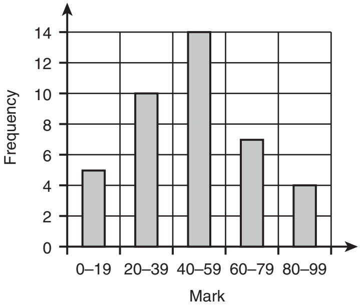

Figure 6.14

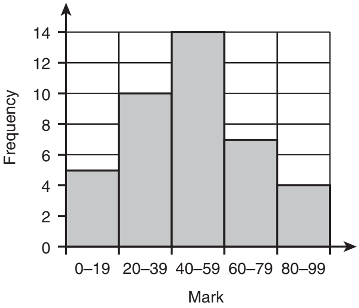

Figure 6.15

Another way to show the information is with a **frequency polygon,** obtained by joining the mid-points of the tops of the bars with straight lines, as in Figure 6.16. Figure 6.17 shows the frequency polygon.

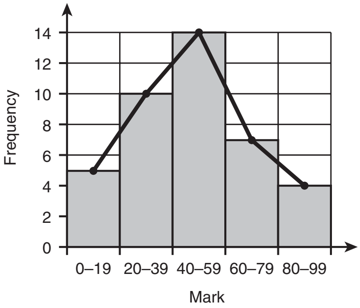

Figure 6.16

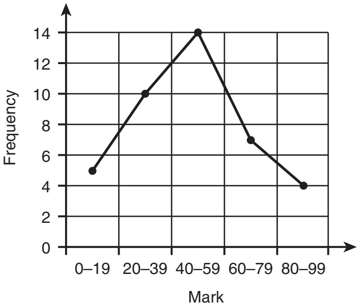

Figure 6.17

Examination marks are discrete data. Grouping continuous data is different in two ways. One of these is the way in which class intervals are defined. For example, with class intervals which have gaps, such as 60–64, 65–69 and so on, where would you put a value of 64.8?

Extending the 60–64 interval up to but not including 65 overcomes this problem. One way of writing such a class interval is $60 \leq w < 65$, where w represents the quantity being measured. $60 \leq w < 65$ is read as 'w is greater than or equal to 60 and less than 65'.

The other difference concerns bar charts which show continuous data. The numbers on the horizontal axis are written in the usual way and there must not be any gaps between the bars.

Example 6.9.1

The table gives information about the weights, in kg, of 50 men.

Weight (w kg)	Frequency
$60 \leq w < 65$	4
$65 \leq w < 70$	6
$70 \leq w < 75$	8
$75 \leq w < 80$	11
$80 \leq w < 85$	15
$85 \leq w < 90$	6

Show this information on
(a) a bar chart **(b)** a frequency polygon.

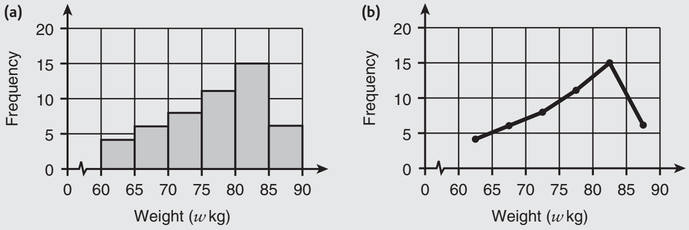

EXERCISE 6.6

1 Here are the numbers of days 30 people were absent from work in a month.

0 0 0 0 0 1 1 1 2 2 3 3 3 4 5
5 6 6 6 8 9 10 10 11 13 14 16 18 19 22

(a) Using groups of 0–4, 5–9 etc., show this data in a grouped frequency table.
(b) Draw a bar chart and a frequency polygon to show the data.

2 The numbers of letters in each of the first 25 words in a book are listed below.

2 6 8 1 11 3 7 3 5 9 8 2 6
2 3 4 2 12 2 4 8 10 6 2 5

(a) Using groups of 1–3, 4–6, 7–9 and 10–12, show this information in a grouped frequency table.
(b) Draw a bar chart and a frequency polygon to show the grouped data.

3 The grouped frequency table gives information about the times, in minutes, 25 people take to travel to work.

Time (t minutes)	Frequency
$0 \le t < 5$	4
$5 \le t < 10$	9
$10 \le t < 15$	7
$15 \le t < 20$	3
$20 \le t < 25$	2

Use the data to draw
(a) a bar chart, **(b)** a frequency polygon.

4 Here are the number of hours of sunshine recorded on each of 20 days in a town.

7.3 2.1 5.2 0.7 2.0 6.4 6.0 9.8 1.8 4.0
3.7 3.0 4.1 3.9 4.9 8.0 1.9 3.1 5.0 0.1

(a) Use the data to make a grouped frequency table using intervals $0 \leq t < 2$, $2 \leq t < 4$, etc.
(b) Draw a bar chart to show the grouped data.

THINGS TO REMEMBER

- *A survey is one method of data collection. Normally a sample of a whole population is taken.*

- *The results of a survey can be shown in a frequency table. For each group in the table the frequency (how many times that particular group occurs in the sample) is shown.*

- *You can display non-numerical or categorical data using:*
  - *a pictogram (or pictograph), in which pictures are used to show the frequency of each group*
  - *a bar chart in which the height of each bar shows the frequency*
  - *a pie chart in which the area of each 'slice' is proportional to the frequency.*

- *Line graphs are often used to show data obtained over a period. The points should only be joined when the points in between the plotted points have meaning, otherwise a vertical line graph is more appropriate.*

- *Scatter graphs show if there might be a relationship between two sets of data.*

- *A scatter graph may have correlation.*
  - *This scatter graph shows high negative correlation.*

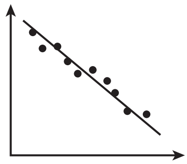

  - *This scatter graph shows weak positive correlation.*

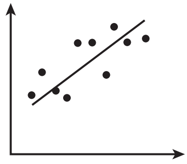

- ▷ *This scatter graph shows no correlation.*

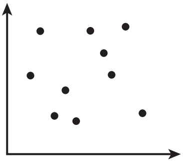

- *A line of best fit is a straight line which passes near all of the points on a scatter graph. There should be roughly equal numbers of points on each side of the line.*

- *Discrete data are obtained by counting and can take only exact values, for example the number of people watching a film.*

- *Continuous data are obtained by measuring and is not exact, for example the height of a person.*

- *Discrete and continuous data can be grouped and organized in a grouped frequency table and displayed in a bar chart.*

- *A frequency polygon is obtained from a bar chart by joining the mid-points of the tops of the bars with straight lines.*

7

Directed numbers

In this chapter you will learn:

- ***about the meaning of negative numbers and directed numbers***
- ***how to order directed numbers***
- ***how to add and subtract directed numbers***
- ***how to multiply and divide directed numbers.***

7.1 Introduction

Numbers greater than zero are called **positive**. You can write them with a + sign in front of them but they are usually written without a sign and assumed to be positive. Numbers less than zero are called **negative** numbers. You *must* write a – sign in front of them. Collectively, they are called **directed numbers**. Positive numbers may be represented in one direction (usually to the right or upwards on a number line), and negative numbers in the opposite direction.

Hindu mathematicians were, from the seventh century AD, the first to make a detailed study of negative numbers. It was not, however, until the Renaissance, 1000 years later, that, through the work of Fermat, Descartes and others, the concept was fully understood.

7.2 Ordering directed numbers

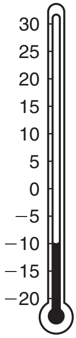

Figure 7.1

Negative numbers are used on thermometers to show temperatures below freezing, that is, below 0°C. The thermometer in Figure 7.1 shows a temperature of –10°C.

You can see from the scale that –5°C is a higher temperature than –10°C, that is, $-5 > -10$, and that –20°C is a lower temperature than –10°C, that is, $-20 < -10$.

A number line like that in Figure 7.2 might help you decide, for example, whether –1 is larger than –3 or not. As –1 is to the right of –3 on the number line, it is larger than –3, so $-1 > -3$.

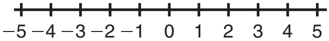

Figure 7.2

Insight

You should say ‘negative 3’ or ‘minus 3’ for –3 and ‘positive 5’ or ‘plus 5’ for +5.

EXERCISE 7.1

1 Which temperature is lower, –7°C or –5°C?

2 Which number is greater, –9 or –10?

3 Insert either < or > appropriately between these pairs of numbers.
(a) –3, –1 **(b)** 0, –2 **(c)** –12, –15

4 Which of the temperatures –8°C, 0°C, –6°C, 2°C, 1°C is
(a) the highest, **(b)** the lowest?

5 Write these numbers in order of size, starting with the lowest.
2, –1, 5, 0, –4, –3

7.3 Addition and subtraction

If a temperature of –3°C increases by 5 degrees, the new temperature is 2°C. You can express this as an addition, –3 + 5 = 2, and illustrate it on a number line as in Figure 7.3. You start at –3 and go 5 to the right.

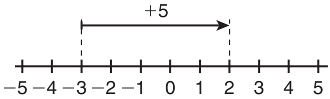

Figure 7.3

You can also obtain similar results such as –3 + 2 = –1 and –3 + 3 = 0. If a temperature of 2°C falls by 3 degrees, the new temperature is –1°C. You can express this as a subtraction, 2 – 3 = –1, and illustrate it on a number line as in Figure 7.4. You start at 2 and go 3 to the left.

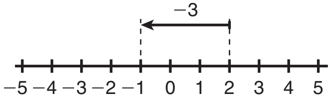

Figure 7.4

In a similar way, you could obtain results such as –2 – 1 = –3.

An addition like 4 + (–1) has to be written in a different form before you can apply this method. These three simple additions can be used to introduce the different form.

4 + 2 = 6 4 + 1 = 5 4 + 0 = 4

Continuing the pattern, 4 + (–1) = 3. However, as 4 – 1 is also 3, this suggests that adding –1 is the same as subtracting 1.

Example 7.3.1

Work out –2 + (–5).

Adding –5 is the same as subtracting 5.
So –2 + (–5) = –2 – 5 = –7.

You can devise a method for a subtraction like 5 – (–1) using a similar approach but in this case you start with three simple subtractions.

$$5 - 2 = 3 \qquad 5 - 1 = 4 \qquad 5 - 0 = 5$$

Continuing the pattern, 5 – (–1) = 6. However, as 5 + 1 is also 6, this suggests that subtracting –1 is the same as adding 1.

Insight

You can think of adding and subtracting negative numbers in the same way as adding and subtracting debt; when you add debt (a negative) the amount in your bank account decreases and when you take away some debt (subtract a negative) your net worth increases.

Example 7.3.2

Work out –7 – (–3).

Subtracting –3 is the same as adding 3.
So –7 – (–3) = –7 + 3 = –4.

Insight

When you add or subtract directed numbers, look at the signs in the middle of the calculation:

Same signs → add	so 6 + (+4) = 6 + 4 = 10 and 6 – (–4) = 6 + 4 = 10
Different signs → subtract	so 6 + (–4) = 6 – 4 = 2 and 6 – (+4) = 6 – 4 = 2

EXERCISE 7.2

1 The temperature rises by 6 degrees from –1°C. Work out the new temperature.

2 The temperature falls by 7 degrees from 3°C. Find the new temperature.

(Contd)

3 The temperature rises from –8°C to 3°C. By how many degrees does the temperature rise?

4 The temperature falls from –2°C to –9°C. By how many degrees does the temperature fall?

5 Work out these additions and subtractions.

(a) –3 + 4 **(b)** –5 + 2 **(c)** –4 + 4 **(d)** 4 – 7 **(e)** –3 – 2
(f) –6 + 5 **(g)** –1 – 8 **(h)** –2 + 5 – 8 **(i)** –4 – 5 + 9 **(j)** 7 – 8 – 3

6 Work out these additions and subtractions.

(a) 5 + (–4) **(b)** 4 – (–3) **(c)** –2 + (–7) **(d)** –6 – (–5)
(e) –3 – (–9) **(f)** 5 + (–5) **(g)** –8 + (–2) **(h)** –8 – (–1)

7.4 Multiplication and division

You may have used a table like the one below to find the product (see Section 1.2) of multiplying two numbers together.

4	4	8	12	16
3	3	6	9	12
2	2	4	6	8
1	1	2	3	4
×	1	2	3	4

You can extend the table to include negative numbers and zero. Then you can complete the rest of the table by continuing the number patterns. For example, the products in the right-hand column are 16, 12, 8 and 4, where each number is 4 less than the one above it. So the next five products will be 0, –4, –8, –12 and –16. You can complete the top left and bottom right sections of the table similarly.

The numbers in the left-hand column of the extended table are –16, –12, –8, –4 and 0; that is, each number is 4 more than the one above it. So the next four numbers will be 4, 8, 12 and 16. You can now complete the table with positive numbers in the bottom left section.

4	–16	–12	–8	–4	0	4	8	12	16
3	–12	–9	–6	–3	0	3	6	9	12
2	–8	–6	–4	–2	0	2	4	6	8
1	–4	–3	–2	–1	0	1	2	3	4
0	0	0	0	0	0	0	0	0	0
–1	4	3	2	1	0	–1	–2	–3	–4
–2	8	6	4	2	0	–2	–4	–6	–8
–3	12	9	6	3	0	–3	–6	–9	–12
–4	16	12	8	4	0	–4	–8	–12	–16
×	–4	–3	–2	–1	0	1	2	3	4

Reading from the table, (–3) × (–2) = 6, 3 × (–2) = –6 and (–3) × 2 = –6.

This suggests that if two numbers have the **same** sign (both positive or both negative) their product is **positive**.

If two numbers have **different** signs, their product is **negative**.

To multiply directed numbers, use these rules to decide if the product is positive or negative; then multiply the numbers, 'ignoring' the signs.

Example 7.4.1

Work out **(a)** 6 × (–5) **(b)** (–9) × (–4) **(c)** (–7) × 0

(a) The numbers have different signs (the 6 is the same as +6) so their product is negative. Ignoring the signs, 6 × 5 = 30, so 6 × (–5) = –30.

(b) As the numbers have the same sign, their product is positive. Ignoring the signs, 9 × 4 = 36, so (–9) × (–4) = 36.

(c) 0 multiplied by *any* number is still 0, so (–7) × 0 = 0.

To carry out a division such as 20 ÷ 4 you need to ask 'What number do you multiply 4 by to get 20?' Then 20 ÷ 4 = 5 because 4 × 5 = 20. The quotient (see Section 1.2) is 5.

In the same way, (–20) ÷ (–4) = 5 and –20 ÷ 5 = –4, because 5 × (–4) = –20. Similarly 20 ÷ (–4) = –5, because (–4) × (–5) = 20.

The rules for dividing directed numbers are the same as those for multiplying. If two numbers have the **same** sign, their quotient is **positive**. If the signs are **different**, their quotient is **negative**.

To divide directed numbers, use these rules to decide if the quotient is positive or negative. Then do the division, 'ignoring' the signs.

Example 7.4.2

Work out **(a)** (–24) ÷ (–3) **(b)** $\frac{-28}{4}$ **(c)** 0 ÷ (–5)

(a) As the numbers have the same sign, the quotient is positive. Ignoring the signs, 24 ÷ 3 = 8, so (–24) ÷ (–3) = 8.
(b) $\frac{-28}{4}$ is the same as (–28) ÷ 4. The numbers have different signs so the quotient is negative. Ignoring signs, 28 ÷ 4 = 7, so $\frac{-28}{4} = -7$.
(c) 0 divided by *any* number is still 0, so 0 ÷ (–5) = 0.

Insight

You can use this memory aid to help you remember the rules for multiplying and dividing a pair of directed numbers:

SAME signs → POSITIVE answer
DIFFERENT signs → NEGATIVE answer

EXERCISE 7.3

EXERCISE

1 Use the table to write down these products.
(a) 2 × (–4) **(b)** (–4) × (–3) **(c)** 0 × (–2)
(d) (–3) × 4 **(e)** (–4) × (–4)

2 Work out these multiplications.

(a) (–5) × 4 **(b)** (–6) × (–4) **(c)** (–7) × 3
(d) (–9) × 0 **(e)** (–6) × (–6) **(f)** 8 × (–8)
(g) (–5) × (–7) **(h)** 0 × 3 **(i)** (–8) × 9

3 Work out these divisions.

(a) 18 ÷ (–3) **(b)** (–32) ÷ (–4) **(c)** $\frac{-21}{7}$
(d) $\frac{-15}{-3}$ **(e)** $\frac{0}{-8}$ **(f)** $\frac{30}{-5}$
(g) (–32) ÷ 8 **(h)** (–36) ÷ (–9) **(i)** 49 ÷ (–7)

7.5 Using a calculator

You can use a calculator to answer questions involving negative numbers, and a calculator is particularly helpful when the numbers are either large or not whole numbers. Calculators vary in the way they deal with negative numbers, but to key in a negative number, you usually have to key in the figures first and then press the +/– key.

Referring to the instructions if necessary, use your calculator to answer some of the questions in earlier exercises in this chapter. When you have done this successfully, use your calculator to carry out the calculations in Exercise 7.4.

EXERCISE 7.4

1 Carry out the following calculations.

(a) 23 – 431 **(b)** –37 – (–93) **(c)** –96 + (–43)
(d) (–442) ÷ 17 **(e)** –1.7 – 3.4 **(f)** –4.3 – (–2.7)
(g) 3.8 × (–0.7) **(h)** (–2.88) ÷ (–0.6)

THINGS TO REMEMBER

- *Positive numbers are greater than zero. They can be written with a + sign in front of them.*
- *Negative numbers are less than zero. They are written with a – sign in front of them, e.g. –3.*
- *Directed numbers are positive and negative numbers.*
- *When a number is written without a sign, it is positive.*
- *You can visualize a thermometer scale to help you order negative numbers. A temperature of –7°C is lower than a temperature of –4°C, so* $-4 > -7$ *or* $-7 < -4$.
- *When you add or subtract directed numbers you can look at the signs in the middle of the calculation:*
  Same signs → add so $6 + (+4) = 6 + 4 = 10$
  and $6 - (-4) = 6 + 4 = 10$
  Different signs → subtract so $6 + (-4) = 6 - 4 = 2$
  and $6 - (+4) = 6 - 4 = 2$
- *When you multiply or divide directed numbers:*
  Step 1: *Carry out the multiplication or division* ignoring *the signs.*
  Step 2: *Look at the signs:*
  *differe*N*t signs then the answer is* N*egative*
  S*ame signs then the answer is po*S*itive*
- *When you use a calculator for negative numbers, you usually have to key in the figures first and then press the +/– key.*

8

Graphs 1

In this chapter you will learn:

- ***how to use coordinates***
- ***how to draw and use straight-line graphs***
- ***about lines parallel to the axes***
- ***about regions and inequalities.***

8.1 Coordinates

Coordinates are used in mathematics to describe the position of a point. The most common system of coordinates uses rectangular or cartesian coordinates.

In Figure 8.1, the horizontal line is called the **x-axis** and the vertical line is called the **y-axis**. The two axes meet at the point O, which is called the **origin**. If you start from the origin and go 3 units to the right and then 2 units up, you reach the point A, which is said to have **coordinates** (3, 2). The first number, 3, is the **x-coordinate** of A and the second number, 2, is the **y-coordinate** of A.

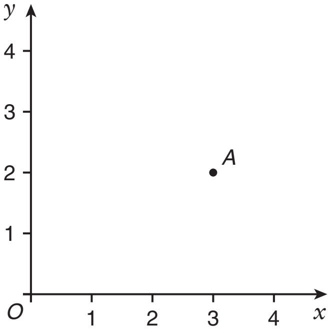

Figure 8.1

The x-coordinate and the y-coordinate of the origin are both 0 so the coordinates of the origin are (0, 0).

The concept of coordinates is ancient. Egyptian surveyors laid out towns in the form of a rectangular grid and the Greeks used the idea of coordinates to study curves. Most significantly, however, it was a coordinate system which enabled René Descartes (1596–1650) to use algebra for solving geometrical problems, thus creating analytical geometry. It is from his surname that the adjective 'cartesian' is derived.

Squared paper or graph paper is normally used for plotting points whose coordinates are given. You may omit some numbers from the scales on the axes. This saves time and avoids cluttering up the diagram. To describe the positions of points left of the y-axis or below the x-axis, you have to extend both scales below zero using negative numbers (see Figure 8.2).

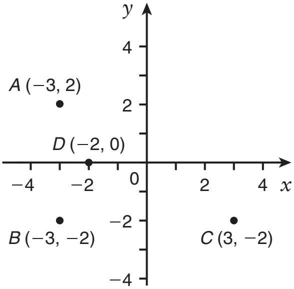

Figure 8.2

You can give the coordinates of a point by giving first its x-coordinate and then its y-coordinate. For example, in Figure 8.2, the x-coordinate of A is –3 and its y-coordinate is 2. So the coordinates of A are (–3, 2). Similarly, the coordinates of B are (–3, –2), the coordinates of C are (3, –2) and D is the point (–2, 0).

Insight

There are several different ways to remember the right order for coordinates:

'Along the corridor and then up the stairs' reminds you that the first number in the coordinate pair is along from the origin and the second number is up (or down).

'*x* is before *y* in the alphabet' reminds you that the first number in the coordinate pair is the *x*-coordinate (horizontal axis) and the second number is the *y*-coordinate (vertical axis).

EXERCISE 8.1

1 Write down the coordinates of the points shown in the diagram.

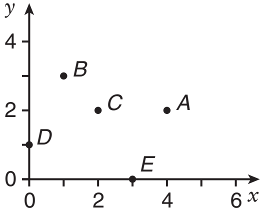

2 **(a)** Draw *x*- and *y*-axes from 0 to 5.
(b) Plot and label the points *A* (2, 3), *B* (5, 1), *C* (1, 0) and *D* (0, 4).

3 **(a)** Draw *x*-and *y*-axes from 0 to 6.
(b) Plot and label the points *E* (2, 1), *F* (2, 4) and *G* (6, 4).
(c) *EFGH* is a rectangle. Find the coordinates of the point *H*.

4 Write down the coordinates of the points shown in the diagram.

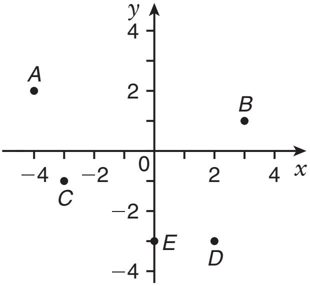

In questions 5–8, draw *x*- and *y*-axes from –5 to 5.

5 Plot and label the points *A* (–3, 4), *B* (–2, –3), *C* (4, –5), *D* (0, –4) and *E* (–4, 0).

6 **(a)** Plot and label the points *D* (2, 2), *E* (–3, –1), *F* (2, –4) and *G* (4, –1). Join the points in order with straight lines.
(b) What type of quadrilateral is *DEFG*?

7 **(a)** Plot and label the points *P* (3, 2), *Q* (–2, 2) and *R* (–4, –2).
(b) *PQRS* is a parallelogram. Find the coordinates of *S*.

8 Find the coordinates of the mid-point of the line joining the points (–5, 4) and (3, 2).

8.2 Straight-line graphs

You can use straight-line graphs to show a relationship between two quantities. For example, the amounts of petrol used by a car in travelling certain distances are given in the table.

Distance travelled in km	0	100	200	300
Petrol used in litres	0	12	24	36

You can draw a graph to show this information by plotting points with coordinates (0, 0), (100, 12), (200, 24) and (300, 36) (see Figure 8.3).

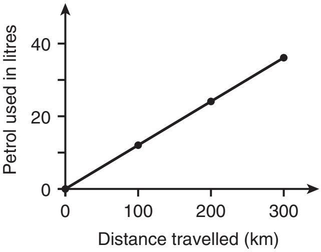

Figure 8.3

Insight

In fact, you only need two points to draw a straight-line graph. However, it is useful to plot at least three in case of mistakes.

You can read information from the graph but you have to be careful reading the scales.

For example, to estimate the amount of petrol used to travel 220 km, read up to the line from 220 on the distance axis and then read across to the petrol axis. It is about 26 litres. Reverse the process to estimate the distance the car will travel on 14 litres. This is about 117 km.

You can also use straight-line graphs to convert systems of units. The graph in Example 8.2.1 converts metric to imperial units. This type of graph is called a **conversion graph**.

You can also use a conversion graph to convert one currency to another.

Example 8.2.1

The diagram is a conversion graph between kilograms (kg) and pounds (lb).

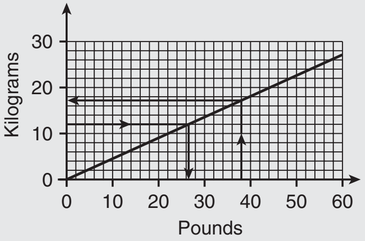

(a) Convert 38 lb to kilograms.
(b) Convert 12 kg to pounds.

(a) Read from 38 on the pounds axis up to the line and then follow the arrow across to the kilograms axis to the number 17. Thus, 38 lb is about 17 kg.
(b) Read across to the line from 12 on the kilograms axis and then follow the arrow down to the pounds axis. So 12 kg is about 26 lb.

Notice that the answers are approximate, that is, they are not exact.

In this book, most graphs are drawn without the small grid of squares. This will enable you to see more clearly what is happening. For your own work, you should use graph paper.

Another type of straight-line graph is the **distance–time** graph (sometimes called a travel graph) which is used to represent a journey. Use the horizontal axis for time and the vertical axis for distance. Figure 8.4 shows a distance–time graph for a cycle ride.

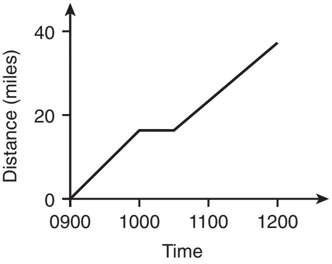

Figure 8.4

Between 0900 and 1000, the cyclist travels 16 miles.

Insight

Notice that the line between 0900 and 1000 is steeper than the line between 1030 and 1200. A steeper line shows that the cyclist has travelled a greater distance in a given time, hence the steeper the line the greater the speed.

Between 1000 and 1030 (the horizontal part of the graph), the cyclist is stationary, resting perhaps.

Finally, between 1030 and 1200, the cyclist travels a further 20 miles.

You can work out the **average speed** for the whole journey using

$$\text{average speed} = \frac{\text{distance travelled}}{\text{time taken}}$$

The cyclist travelled a distance of 36 miles in 3 hours, so the average speed is 36 ÷ 3 = 12. As the distance is measured in miles, and the time in hours, the speed is in miles per hour (mph).

EXERCISE 8.2

1 The diagram shows the number of electricity units used by a 60 watt light bulb left on for up to 30 hours.

(a) How many units of electricity are used in

(i) 7 hours,

(ii) 23 hours?

(b) For how many hours was the bulb left on if it used

(i) 0.7 units,

(ii) 1.1 units?

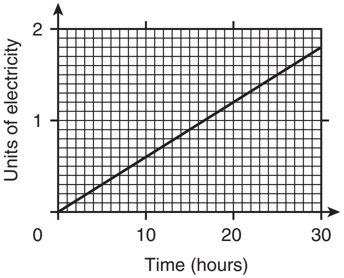

2 The graph converts miles to kilometres and vice versa.

(a) Convert to kilometres
  (i) 28 miles,
  (ii) 42 miles.

(b) Convert to miles
  (i) 35 km,
  (ii) 88 km.

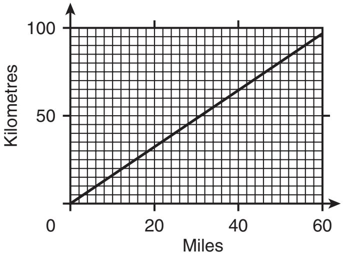

3 This graph is used to convert between pounds (£) and euros (€).

(a) Convert to euros
  (i) £38, **(ii)** £26.

(b) Convert to pounds
  (i) €75, **(ii)** €48.

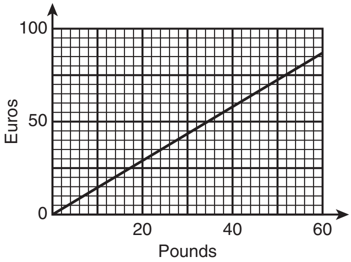

4 Mrs Ford drove from London to Brighton and back. The diagram shows the distance–time graph for her journey.

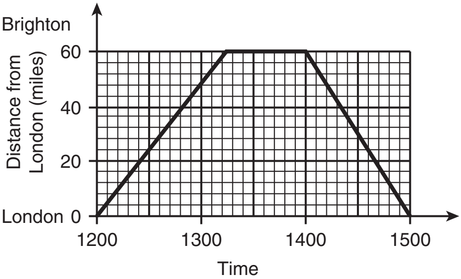

(a) At what time did she reach Brighton?

(b) For how long did she stay in Brighton?

(c) Work out her average speed for the whole journey.

(d) Work out her average speed for the journey from London to Brighton. (Hint: You do *not* work out 60 ÷ 1.15.)

(Contd)

5 **(a)** Using a scale of 1 cm to 5 gallons on the horizontal axis and a scale of 1 cm to 25 litres on the vertical axis, plot the points given by the pairs of values in the table. Join the points with a straight line.

Gallons	0	10	20	30	40
Litres	0	45	90	135	180

(b) Convert 18 gallons to litres.
(c) Convert 120 litres to gallons.

6 **(a)** Using £20 = $39 and £40 = $78, draw a graph which you can use to convert between pounds and dollars for amounts up to £50.
(b) Convert £17 to dollars.
(c) Convert $53 to pounds.

7 A cyclist set off from home at 8 a.m., and at 9.30 a.m. had covered 30 km. He rested for half an hour and then cycled a further 15 km in an hour. Finally, he cycled back home, arriving at 1 p.m.
(a) Draw a distance–time graph for the journey.
(b) Calculate his average speed for the whole journey.
(c) Calculate his average speed for the first 30 km of his journey.

8 The time, in minutes, needed to cook a chicken is given by this rule 'Multiply the weight in pounds by 20 and then add 20.'
(a) Use the rule to complete the table.

Weight in pounds	1	2	3	4	5	6
Time in minutes		60				140

(b) Draw a graph to show this information. (You should *not* join the straight line to the origin.)
(c) Find the time needed to cook a 4.2 pound chicken.

8.3 Lines parallel to the axes

Points with coordinates (3, 4), (3, 1) and (3, –2) all lie on the vertical line shown in Figure 8.5.

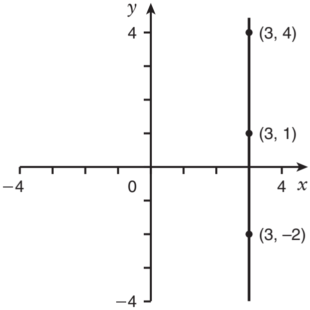

Figure 8.5

The three points have the same x-coordinate, 3. In fact, *every* point on that vertical line has 3 as its x-coordinate; $x = 3$ is called the **equation** of the line. *Any* vertical line, therefore, has an equation of the form $x = c$, where c is the value of x where the line crosses the x-axis. The equation of the y-axis is $x = 0$.

You can use a similar approach to find the equation of a horizontal line, except that, in this case, it will be the y-coordinates which are equal. So the equation of a horizontal line has the form $y = d$, where d is the value of y where the line crosses the y-axis. The equation of the x-axis is $y = 0$.

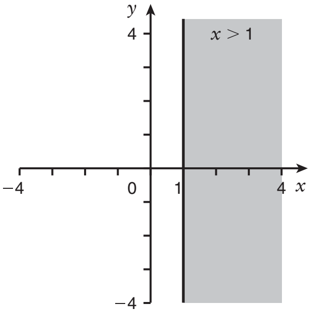

Figure 8.6

In Figure 8.6 the **region** to the right of the line $x = 1$ has been shaded. The x-coordinate of all the points in the shaded region is greater than 1.

The line $x = 1$ is drawn as a solid line, without gaps, which tells you that the region includes the line $x = 1$. (If the line had not been part of the region, it would be a dotted line.) The x-coordinate of every point in the region is, therefore, equal to or greater than 1. The **inequality** $x \geq 1$ describes the region.

In a similar way, you can also describe regions which have one or more edges parallel to the x-axis using inequalities.

Example 8.3.1

Draw regions to illustrate the inequalities

(a) $y < 2$ **(b)** $-1 \leq y < 2$

(a)

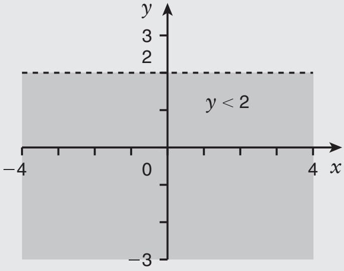

Start by drawing the line $y = 2$ but it must be dotted, as points on the line $y = 2$ are not in the region. Points below the line have a y-coordinate less than 2, so you shade the region below the line as shown above.

(b) The inequality $-1 \leq y < 2$ is a combination of the inequalities $-1 \leq y$ and $y < 2$. As $-1 \leq y$ is the same as $y \geq -1$ the region which represents $-1 \leq y < 2$ combines the regions which represent $y \geq -1$ and $y < 2$. This is shown below.

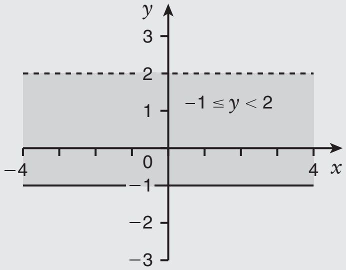

Insight

In this chapter, required regions are shown by shading but, sometimes, *unwanted* regions are shaded. Always make sure that you know which convention is being followed.

EXERCISE 8.3

1 Write down the equations of lines *P*, *Q*, *R* and *S* shown in the diagram.

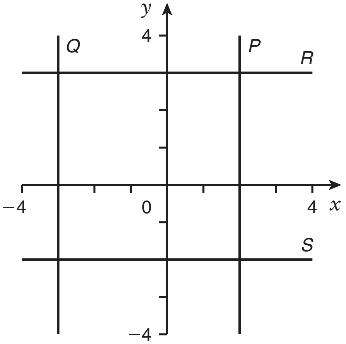

2 Draw *x*- and *y*-axes from –5 to 5, and draw the following lines.
(a) $x = -4$ **(b)** $y = 1$ **(c)** $x = 1$ **(d)** $y = -1$

3 Use inequalities to describe the shaded region in each figure.

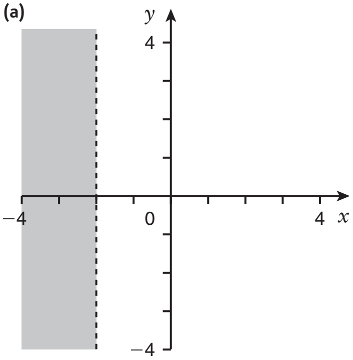

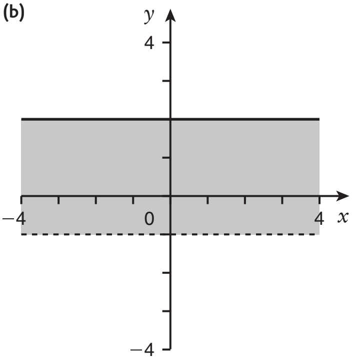

4 Draw regions to illustrate the following inequalities.
(a) $y \geq 2$ **(b)** $x \leq 0$
(c) $-3 < x \leq 1$ **(d)** $-3 \leq y \leq 0$

THINGS TO REMEMBER

- *The position of a point is given by an ordered pair of numbers called coordinates, for example (4, –1). The first number, 4, in the pair is the x-coordinate and the second number, –1, is the y-coordinate.*

- *The x-axis is the horizontal axis and the vertical axis is the y-axis. The plural of axis is axes.*

- *The origin, O, is the point (0, 0) where the two axes meet.*

- *A straight-line graph can be used to show a relationship between two quantities.*

- *A conversion graph is used to convert systems of units, for example to convert between miles and kilometres.*

- *On a distance–time graph or travel graph the horizontal axis is time and the vertical axis is distance. The steeper the line, the greater the speed of the object. A horizontal line shows that the object is stationary.*

$$\text{average speed} = \frac{\text{distance travelled}}{\text{time taken}}$$

- *A vertical line has an equation of the form $x = c$, where c is the value of x where the line crosses the x-axis. The equation of the y-axis is $x = 0$.*

- *A horizontal line has an equation of the form $y = d$, where d is the value of y where the line crosses the y-axis. The equation of the x-axis is $y = 0$.*

- *Regions can be shown using inequalities. The inequality $x \geq 2$ describes the region to the right of, and including, the line $x = 2$. The inequality $2 < y < 5$ describes the region between the lines $y = 2$ and $y = 5$, not including the lines.*

9

Measurement

In this chapter you will learn:

- ***about the metric system***
- ***how to convert between metric units***
- ***about imperial units and how to convert between them***
- ***how to convert between metric and imperial units***
- ***how to choose suitable units.***

9.1 The metric system

The metric system is a system of units based on the metre (*meter* in the USA), which is derived from the Greek *metron,* meaning 'measure'. It was proposed in 1670 by Gabriel Mouton, a priest in Lyons. In the 1790s, France became the first country to adopt the metric system when the revolutionary government introduced it to bring order to the chaotic diversity of traditional units then in use. Subsequently, the metric system of weights and measures has been adopted by the majority of countries and is used by scientists throughout the world.

Insight

A metre was originally defined as one ten-millionth of the distance around the Earth from the North Pole to the equator. It has been redefined several times since and is currently defined as the distance travelled by light in $\frac{1}{299\,792\,458}$ of a second.

The symbol ‘m’ after a number means that the number is a measurement in metres. Thus 100 m means 100 metres.

Latin prefixes are used to express fractions of a metre. The prefix ‘milli’ means $\frac{1}{1000}$. Thus, 1 millimetre = $\frac{1}{1000}$ metre. The short form for a millimetre is mm so 1 mm = $\frac{1}{1000}$ m. Thus 1 m = 1000 mm.

The prefix ‘centi’ means $\frac{1}{100}$. Thus 1 centimetre = $\frac{1}{100}$ metre so 1 metre = 100 centimetres. The short form of centimetre is cm, so 1 cm = $\frac{1}{100}$ m, and 1 m = 100 cm.

Since 1 m = 1000 mm and 1 m = 100 cm, 10 mm = 1 cm.

Greek prefixes are used to express multiples of a metre. For example, 1 kilometre, written 1 km, is 1000 metres, so 1 km = 1000 m.

These prefixes are also used with other metric units.

The main metric units of mass, often called weight, are **grams** (g), **milligrams** (mg) and **kilograms** (kg). For large masses, the **tonne** (t) (= 1000 kg) is used.

The main metric units of capacity, which is the amount a container can hold, are **litres** (l), **centilitres** (cl) and **millilitres** (ml).

To convert from one metric unit to another, you multiply or divide by 10 or 100 or 1000, as appropriate.

Insight

To convert to a *larger* unit, you *divide*, and to convert to a *smaller* unit, you *multiply*.

For example, $\text{mm} \xrightarrow{\div 10} \text{cm} \xrightarrow{\div 100} \text{m} \xrightarrow{\div 1000} \text{km}$ and

$\text{km} \xrightarrow{\times 1000} \text{m} \xrightarrow{\times 100} \text{cm} \xrightarrow{\times 10} \text{mm}$

Example 9.1.1

Convert the following, using appropriate abbreviations for the units.
(a) 38 millimetres to centimetres **(b)** 2.41 metres to centimetres
(c) 8753 grams to kilograms **(d)** 5.7 litres to millilitres

(a) 38 mm = 3.8 cm **(b)** 2.41 m = 241 cm
(c) 8753 g = 8.753 kg **(d)** 5.7 l = 5700 ml

EXERCISE 9.1

1 Convert these quantities.
(a) 79 millimetres to centimetres **(b)** 850 cm to metres
(c) 3240 grams to kilograms **(d)** 4125 millilitres to litres
(e) 1.8 kilometres to metres **(f)** 400 metres to kilometres
(g) 9.4 centimetres to millimetres **(h)** 0.94 litres to millilitres
(i) 2.3 kilograms to grams **(j)** 3.72 tonnes to kilograms

2 Measure the length of this line. Write your answer in
(a) centimetres **(b)** millimetres.

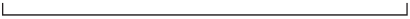

3 A dishwasher uses 25 g of powder for each wash. For how many washes will a 3 kg box of powder last?

4 A paper towel is 242 mm long. There are 230 paper towels on a roll. Find the length, in metres, of the roll.

5 At birth, a baby weighed 3.29 kg. She gained 160 g each week. Find her weight, in kilograms, after 13 weeks.

6 Claire takes 3 teaspoons of medicine twice a day. A teaspoon holds 5 ml. How many days will 1 litre of medicine last?

9.2 Imperial units

English weights and measures, usually called **imperial units**, are gradually being superseded by metric units. Indeed, at the time of writing (2010), the only imperial units which still have legal status in the United Kingdom are miles and pints, although several others are still in everyday use. Here are the most common ones and the relationships between them.

Length 12 inches (in) = 1 foot (ft)
3 feet (ft) = 1 yard (yd)
1760 yards (yd) = 1 mile
Mass 16 ounces (oz) = 1 pound (lb)
Capacity 8 pints (pt) = 1 gallon

Some imperial units have interesting backgrounds. If you stretch your arm out sideways, you will find that the distance from the end of your nose to the tip of your fingers is almost exactly a yard, a fact used by King Henry I when he fixed this measurement, taken from his own body, as the standard yard.

'Mile' comes from the Latin *mille,* meaning a thousand. A mile was the distance covered in a thousand paces, a pace being *two* strides.

Insight

Over the centuries, the British made adjustments to imperial units to correct inconsistencies but these changes were not always implemented in their colonies. Consequently, colonial weights and measures were sometimes slightly different from those in Britain. For example, a British gallon is equal to 1.201 US gallons.

EXERCISE 9.2

1 Express the given quantity in the new unit.
(a) 8 feet to inches
(b) 40 pints to gallons
(c) 48 ounces to pounds
(d) 7 gallons to pints
(e) 24 feet to yards
(f) 3 miles to yards
(g) 5 ft 10 in to inches
(h) 3 lb 7 oz to ounces
(i) 43 inches to feet and inches
(j) 57 oz to pounds and ounces

2 Paul is 5 ft 10 in tall. Peter is 4 inches taller than Paul. Find Peter's height in feet and inches.

3 Bill is 6 ft 3 in tall and Ann is 5 ft 7 in. How many inches taller is Bill than Ann?

4 Margaret drinks a pint of milk every day. How many gallons of milk will she drink in 16 weeks?

5 The biggest beer tankard in the world holds 615 gallons. How many pints is this?

6 My stride is 32 inches long. How many strides do I make in walking 1 mile?

9.3 Converting between metric and imperial units

It is useful to be able to convert between metric and imperial units in common use. For most practical purposes, approximate equivalents are adequate. Here are the main ones.

Length 1 inch = 2.5 centimetres
1 metre = 3 feet
1 mile = 1.6 kilometres
i.e. 5 miles = 8 kilometres
Mass 1 kilogram = 2.2 pounds
1 ounce = 25 grams
Capacity 1 litre = $1\frac{3}{4}$ pints
1 gallon = 4.5 litres

Example 9.3.1

(a) Convert 15 centimetres to inches.
(b) Convert 30 miles to kilometres.

(a) There are 2.5 cm in 1 inch. In 15 cm there are 15 ÷ 2.5 in = 6 in.
(b) One mile is 1.6 km, so 30 miles is 30 × 1.6 km = 48 km.
Alternatively, you could use 5 miles = 8 km.
You must multiply 5 by 6 to get 30, so 30 miles = 6 × 8 km = 48 km.

EXERCISE 9.3

1 Convert the first quantity to the second.
   - **(a)** 6 inches to centimetres
   - **(b)** 35 centimetres to inches
   - **(c)** 4 metres to feet
   - **(d)** 45 miles to kilometres
   - **(e)** 150 grams to ounces
   - **(f)** 5 kilograms to pounds
   - **(g)** 12 litres to pints
   - **(h)** 54 litres to gallons
   - **(i)** 96 kilometres to miles
   - **(j)** 14 pints to litres

2 The waist size of a pair of trousers is 32 inches. How many centimetres is this?

3 A tin contains 425 g of soup. How many ounces is this?

4 The chest size of a sweater is 105 cm. How many inches is this?

5 The capacity of a car's petrol tank is 36 litres. How many gallons is this?

6 David weighs 85 kg. Express his weight in pounds.

7 The distance from London to Birmingham is 120 miles. Express this in kilometres.

8 Barbara's height is 175 cm. Express this in feet and inches.

9 The distance from London to Cardiff is 248 km. Express this in miles.

10 Delia needs 10 lb of sugar to make jam. Sugar is sold in 1 kg bags. How many bags must she buy?

11 A car travels 45 miles on one gallon of petrol. How many kilometres will it travel on one litre of petrol?

9.4 Choosing suitable units

Before you measure something, you must decide which units to use. To measure the width of this page in metric units, for example, centimetres or millimetres are the most appropriate. Expressed in these units, the width is 12.7 cm or 127 mm. If you used metres or kilometres, you would need numbers less than 1 for your answers.

On the other hand, the most sensible metric units for the weight of a car are kilograms or tonnes. 2200 kg or 2.2 tonnes are reasonable ways of expressing the weight of a Ford Mondeo, but if you used grams or milligrams you would require very large numbers for your answers.

EXERCISE 9.4

1 Write down the name of
 (i) an appropriate metric unit and
 (ii) an appropriate imperial unit
 for measuring each of the following.
 (a) the height of the Eiffel Tower
 (b) the weight of a coin
 (c) the amount of water in a kettle
 (d) the radius of the Earth
 (e) the amount of water in a pond
 (f) the weight of a television

2 Complete each if these sentences by writing an appropriate metric unit to make the statement true.
 (a) A tin of soup weighs 425 ....
 (b) A large bottle of lemonade holds 3 ....
 (c) A ball of wool weighs 25 ....
 (d) There are about 5 .... of blood in the human body
 (e) A teaspoon holds about 5 ....

THINGS TO REMEMBER

- *The main metric units of length are millimetres (mm), centimetres (cm), metres (m) and kilometres (km).*

- *The conversions for metric units of length are:*
  - *1 cm = 10 mm*
  - *1 m = 100 cm = 1000 mm*
  - *1 km = 1000 m*

$$\text{mm} \xrightarrow{\div 10} \text{cm} \xrightarrow{\div 100} \text{m} \xrightarrow{\div 1000} \text{km}$$
$$\text{km} \xrightarrow{\times 1000} \text{m} \xrightarrow{\times 100} \text{cm} \xrightarrow{\times 10} \text{mm}$$

- *The main metric units of mass (weight) are milligrams (mg), grams (g), kilograms (kg) and tonnes (t).*

- *The conversions for metric units of mass are:*
  - *1 g = 1000 mg*
  - *1 kg = 1000 g*
  - *1 tonne = 1000 kg*

$$\text{mg} \xrightarrow{\div 1000} \text{g} \xrightarrow{\div 1000} \text{kg} \xrightarrow{\div 1000} \text{tonne}$$
$$\text{tonne} \xrightarrow{\times 1000} \text{kg} \xrightarrow{\times 1000} \text{g} \xrightarrow{\times 1000} \text{mg}$$

- *The main metric units of capacity (volume) are millilitres (ml), centilitres (cl) and litres (l).*

- *The conversions for metric units of capacity are:*
  - *1 l = 1000 ml*
  - *1 cl = 10 ml*
  - *1 l = 100 cl*

$$\text{ml} \xrightarrow{\div 10} \text{cl} \xrightarrow{\div 100} \text{l}$$
$$\text{l} \xrightarrow{\times 100} \text{cl} \xrightarrow{\times 10} \text{ml}$$

- *The main imperial units are:*

  Length 12 inches (in) = 1 foot (ft)
  3 feet (ft) = 1 yard (yd)
  1760 yards (yd) = 1 mile
  Mass 16 ounces (oz) = 1 pound (lb)
  Capacity 8 pints (pt) = 1 gallon

- *Approximate conversions from imperial to metric are:*

  Length 1 inch = 2.5 centimetres
  1 metre = 3 feet
  1 mile = 1.6 kilometres
  i.e. 5 miles = 8 kilometres
  Mass 1 kilogram = 2.2 pounds
  1 ounce = 25 grams
  Capacity 1 litre = 1$\frac{3}{4}$ pints
  1 gallon = 4.5 litres

10

Perimeter and area

In this chapter you will learn:

- ***how to find and use the perimeter of a rectangle***
- ***about the meaning of area***
- ***how to find and use the area of rectangles, triangles, parallelograms and trapeziums.***

10.1 Perimeter

The **perimeter** of a shape is the length of its boundary. The word 'perimeter' is derived from the Greek *peri* and *metron,* meaning 'around' and 'measure'.

For example, to find the perimeter of a triangle with sides of length 2 cm, 3 cm and 4 cm you add the three lengths. So the perimeter of this triangle is (2 + 3 + 4) cm = 9 cm.

To find the perimeter of a square with sides of 3 cm, you could work out the sum 3 + 3 + 3 + 3 but it is quicker to find the product (4 × 3) cm = 12 cm. Similarly, to find the perimeter of a rectangle 6 cm long and 5 cm wide, you could work out the sum 6 + 5 + 6 + 5. Alternatively, you could double both the length and width and then add the two results, giving a perimeter of (2 × 6 + 2 × 5) cm = (12 + 10) cm = 22 cm.

In general,

perimeter of a rectangle = 2 × length + 2 × width.

Insight

To work out the perimeter of a more complicated shape, cross off each side as you go round the shape.

EXERCISE 10.1

1 The lengths of the sides of a triangle are 8 cm, 5 cm and 4 cm. Find its perimeter.

2 The length of each side of a triangle is 9 cm. Find its perimeter.

3 Find the perimeter of a square with sides 9.3 cm long.

4 A rectangle is 6.7 cm long and 4.8 cm wide. Find its perimeter.

5 Find the perimeter of a regular pentagon with sides 7.2 cm long.

6 For international matches, the length of a soccer pitch must be between 100 m and 110 m. Its width must be between 64 m and 73 m. Find the perimeter which is

(a) the least possible, **(b)** the greatest possible.

7 On separate diagrams, plot the following points and join them in the order given. Find the perimeter of each shape.

(a) (1, 1), (1, 4), (3, 4), (3, 3), (5, 3), (5, 2), (4, 2), (4, 1)

(b) (1, 2), (1, 3), (3, 3), (3, 5), (5, 5), (5, 1), (4, 1), (4, 2)

8 A soccer pitch is 100 m long and 75 m wide. A man runs around the pitch, travelling 7 km. How many times did he run round?

9 The figure shows six squares shaded on a centimetre grid measuring 4 cm by 2 cm.

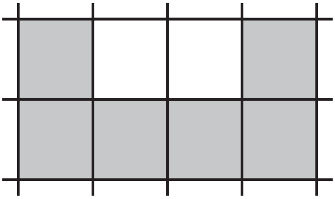

(a) Find the perimeter of the shaded area.

(b) Draw as many more arrangements of six squares on this grid as you can; find the perimeter of each one.

(c) What is the greatest perimeter you can get?

(Contd)

10 The table shows the number of rolls of wallpaper needed for a room, if the perimeter and the height of the room are known.

Height (metres)	Perimeter not over							
	10 m	12 m	14 m	16 m	18 m	20 m	22 m	24 m
2.15–2.30	5	5	6	7	8	9	10	10
2.30–2.45	5	6	7	8	8	9	10	11
2.45–2.60	5	6	7	8	9	10	11	12
2.60–2.75	5	6	7	8	9	10	11	12
2.75–2.90	6	7	8	9	10	11	12	13
2.90–3.05	6	7	8	9	10	11	12	13
3.05–3.20	7	8	9	10	12	13	14	15

Find the number of rolls of wallpaper needed for a room with
(a) a height of 2.40 m and a perimeter of 21 m,
(b) a height of 2.65 m and a perimeter of 18 m,
(c) a height of 2.53 m and a perimeter of 16.15 m.

11 Use the table in question 10 to find the number of rolls of wallpaper needed for a rectangular room which is
(a) 5 m long, 4.5 m wide and 2.84 m high,
(b) 6.24 m long, 4.86 m wide and 2.95 m high.

12 The figure, which is not to scale, shows the plan of a room. The height of the room is 2.5 m. How many rolls of wallpaper are needed for it?

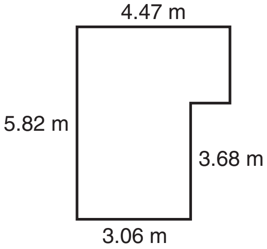

10.2 Area

Area is a measurement of amount of surface. To find the area of a surface, you need to cover it with squares without leaving any gaps, and then count the number of squares used. A square with sides 1 cm long has an area of 1 square centimetre, which is written as 1 cm^2.

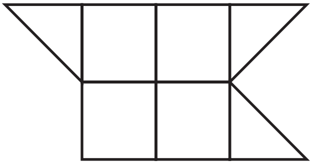

Figure 10.1

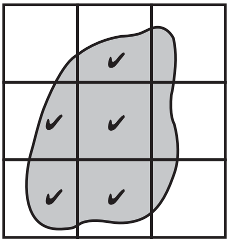

Figure 10.2

For example, the shape shown in Figure 10.1 contains 4 full squares and 3 half squares. So its area is $(4 + 1\frac{1}{2})$ cm^2 = $5\frac{1}{2}$ cm^2.

You cannot always find the area of a shape exactly. For awkward shapes, like that in Figure 10.2, you have to approximate. One way is to count the number of complete squares inside the shape (one in this case) and those which are half or more inside (four). If less than half a square is covered by the shape, ignore it. So the approximate area is 5 cm^2.

EXERCISE 10.2

1 Find the areas of the shapes shown in parts (a) to (f).

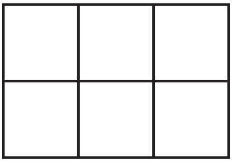

(a)

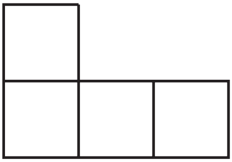

(b)

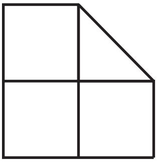

(c)

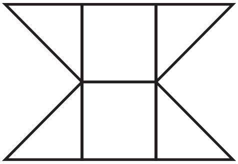

(d)

(Contd)

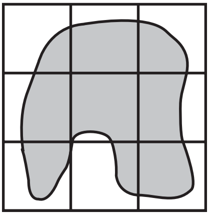

(e)

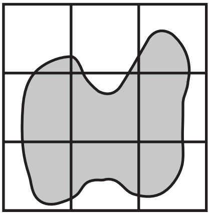

(f)

For questions 2 and 3 use centimetre squared paper.

2 Plot the following points and join them in the order given. Find the area of each shape.

(a) (1, 1), (1, 3), (4, 3), (4, 1)	**(b)** (1, 1), (3, 3), (6, 3), (4, 1)
(c) (1, 2), (4, 5), (7, 2)	**(d)** (2, 1), (4, 5), (7, 2)
(e) (1, 2), (2, 5), (6, 6), (5, 1)	**(f)** (2, 1), (2, 5), (3, 5), (5, 3)

3 Draw a circle with a radius of 3 cm. (Set your compasses to 3 cm.) Find its approximate area.

10.3 Area of a rectangle

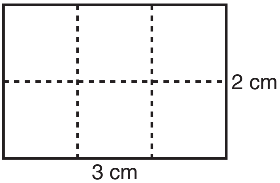

Figure 10.3

The rectangle in Figure 10.3 is 3 cm long and 2 cm wide. It contains 6 centimetre squares, so its area is 6 cm^2. But you could count the squares differently, by saying that there are two rows of three squares, that is $3 \times 2 = 6$.

This is equivalent to multiplying the length of the rectangle by its width. Thus the area of the rectangle is (3×2) cm^2 = 6 cm^2.

In general:

area of a rectangle = length × width.

You must use the same units for both length and width: if they are both centimetres, the units of the area will be cm^2; if they are both metres, the units of the area will be m^2.

A **square** is a rectangle whose length and width are equal.

So the formula for area becomes:

area of a square = length of side × length of side.

You can sometimes find the area of a shape by dividing it into rectangles.

Example 10.3.1

Find the area of the shape in the figure. The dotted lines divide the shape into three rectangles *A*, *B* and *C*.

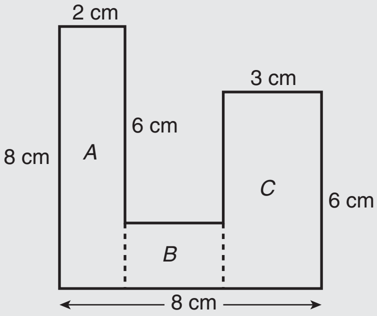

Area $A = (8 \times 2)\ \text{cm}^2 = 16\ \text{cm}^2$
Area $B = (3 \times 2)\ \text{cm}^2 = 6\ \text{cm}^2$
Area $C = (6 \times 3)\ \text{cm}^2 = 18\ \text{cm}^2$
By adding, the total area is $40\ \text{cm}^2$.

Insight

Sometimes a shape has a hole cut out. In that case the total area = area of outside shape – area of the hole.

EXERCISE 10.3

1 A rectangle is 12 cm long and 7 cm wide. Find its area.
2 A rectangle is 8 m long and 6 m wide. Find its area.

(Contd)

3 Each side of a square is 9 cm long. Find its area.

4 A rectangle is 2.3 m long and 60 cm wide. Find its area
(a) in m^2, **(b)** in cm^2.

5 Each side of a square is 100 cm long. Find its area
(a) in m^2, **(b)** in cm^2.

6 Find the areas of the shapes shown in the diagrams. The lengths are in centimetres, and the diagrams are not drawn to scale.

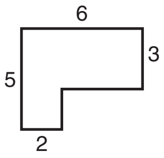

(a)

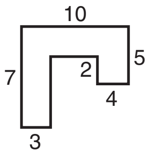

(b)

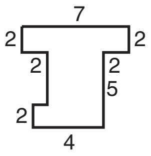

(c)

7 Find the areas of the shaded regions in the diagrams. The lengths are in centimetres, and the diagrams are not drawn to scale.

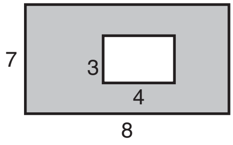

(a)

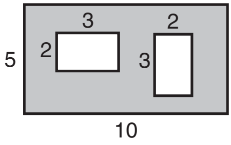

(b)

8 A room is 3 m high and its floor is a rectangle 5 m by 4 m.
- **(a)** Find the total area of the walls.
- **(b)** Find the area of the ceiling.
- **(c)** One litre of paint covers 12 m^2. How many litres of paint are needed for the walls and the ceiling?

9 A carpet which is 3 m by 2 m is placed on a floor which is 5 m by 3 m. What area of the floor is not covered by the carpet?

10 A rectangular lawn is 9 m by 7 m. There is a path which is 1 m wide all around the outside of the lawn. Find the area of the path.

11 The length of each side of a square is 100 m. The square has an area of 1 hectare. Find its area in m^2.

12 The floor of a room is a rectangle 3.5 m by 3 m. How many square carpet tiles, each 50 cm by 50 cm, are needed to cover the floor?

10.4 Area of a parallelogram

To find the area of the parallelogram in Figure 10.4, remove the shaded triangle and replace it at the other end to make a rectangle. This rectangle must have the same area as the original parallelogram. So

$$\text{area} = (3 \times 2)\ \text{cm}^2 = 6\ \text{cm}^2.$$

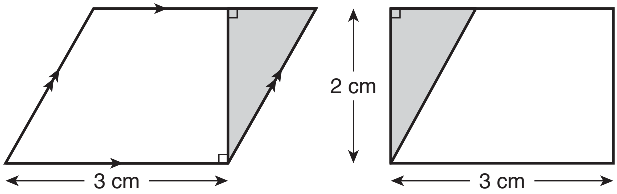

Figure 10.4

In general,

$$\text{area of parallelogram} = \text{base} \times \text{height}.$$

Remember that the 'height' is the vertical (or perpendicular) height, which is sometimes called the **altitude**. In other words, it is the distance between the parallel lines.

10.5 Area of a triangle

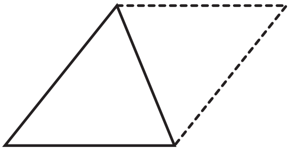

Figure 10.5

Starting with any triangle, you can construct a parallelogram which has twice the area of the original triangle (see Figure 10.5). So

$$\text{area of triangle} = \tfrac{1}{2} \times \text{base} \times \text{height}.$$

Again, remember that 'height' means *vertical* height.

To find the area of triangle *ABC* in Figure 10.6, for example, measure the length of the base *AB* (3 cm) and measure the length of *CD*, the perpendicular distance from *C* on to *AB* (1 cm).

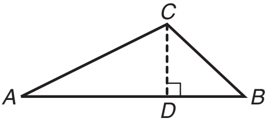

Figure 10.6

$$\text{Area of triangle } ABC = (\tfrac{1}{2} \times 3 \times 1)\text{ cm}^2 = 1\tfrac{1}{2}\text{ cm}^2$$

Alternatively, you could use *AC* or *BC* as the 'base'. If you used *AC* as the base, you would have to extend *AC*, before you could draw and measure *BE*, the 'vertical' height from *B* to *AC* (see Figure 10.7).

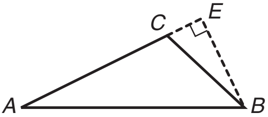

Figure 10.7

Then $AC = 2.2$ cm and $BE = 1.35$ cm.

So area of triangle $ABC \approx \frac{1}{2} \times 2.2 \times 1.35\text{ cm}^2 = 1.485\text{ cm}^2 \approx 1.5\text{ cm}^2$

The symbol ≈ means 'is approximately equal to'.

You can sometimes find the area of a shape by dividing it into rectangles and triangles, or by 'boxing' it.

Example 10.5.1

Find the area of the trapezium in the figure.
The dotted line divides the trapezium into a rectangle *A* and a triangle *B*.

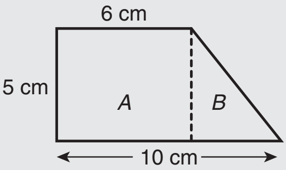

Area of rectangle $A = (6 \times 5)\text{ cm}^2 = 30\text{ cm}^2$
Area of triangle $B = (\frac{1}{2} \times 4 \times 5)\text{ cm}^2 = 10\text{ cm}^2$
So total area = 40 cm^2.

Example 10.5.2

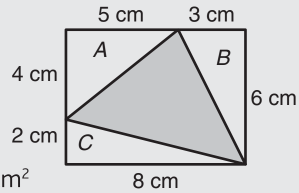

Find the area of the shaded triangle.
Area of rectangle $(8 \times 6)\ \text{cm}^2 = 48\ \text{cm}^2$

Area of triangle $A = (\frac{1}{2} \times 4 \times 5)\ \text{cm}^2 = 10\ \text{cm}^2$

Area of triangle $B = (\frac{1}{2} \times 6 \times 3)\ \text{cm}^2 = 9\ \text{cm}^2$

Area of triangle $C = (\frac{1}{2} \times 8 \times 2)\ \text{cm}^2 = 8\ \text{cm}^2$

Sum of areas of triangles A, B and $C = (10 + 9 + 8)\ \text{cm}^2 = 27\ \text{cm}^2$
Area of shaded triangle $= (48 - 27)\ \text{cm}^2 = 21\ \text{cm}^2$

EXERCISE 10.4

1 Find the areas of the following shapes. The units are centimetres.

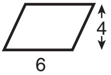

(a)

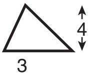

(b)

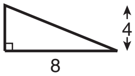

(c)

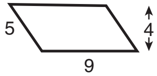

(d)

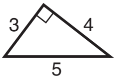

(e)

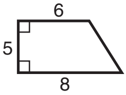

(f)

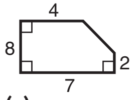

(g)

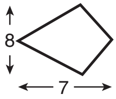

(h)

2 Make appropriate measurements to find the areas of these shapes.

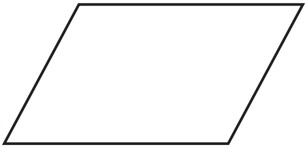

(a)

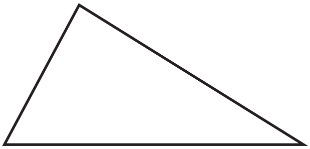

(b)

(Contd)

3 Find the shaded areas. The units are centimetres.

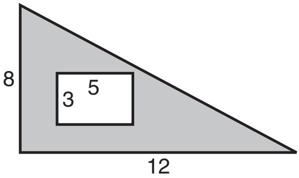

(a)

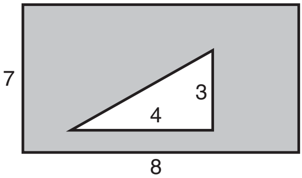

(b)

4 Using the method of Example 10.5.2, find the area of the triangle with its vertices at (0, 1), (3, 3) and (4, 0).

5 The diagonals of a rhombus are 10 cm and 8 cm long. Find its area. (Hint: the diagonals of a rhombus cross at right angles.)

6 The diagonals of a kite are 12 cm and 10 cm long. Find its area.

10.6 Area of a trapezium

To find the area of the trapezium shown in Figure 10.8, join it to a congruent trapezium to form a parallelogram. (See Figure 10.9.)

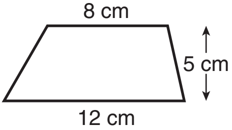

Figure 10.8

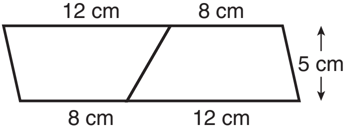

Figure 10.9

Insight

Congruent shapes are exactly the same shape and size.

The length of the base of the parallelogram is

$$(12 + 8)\text{ cm} = 20\text{ cm}$$

So the area of the parallelogram is

$$(20 \times 5)\text{ cm}^2 = 100\text{ cm}^2$$

The area of the trapezium is half that of the parallelogram, that is

$$(\tfrac{1}{2} \times 20 \times 5)\text{ cm}^2 = 50\text{ cm}^2$$

Area of a trapezium
$= \frac{1}{2} \times$ sum of parallel sides $\times$ vertical distance between them

Example 10.6.1

Find the area of the trapezium in the figure.

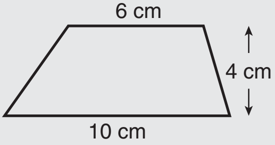

$$\begin{aligned} \text{Area} &= \tfrac{1}{2} \times (10 + 6) \times 4\ \text{cm}^2 \\ &= 32\ \text{cm}^2 \end{aligned}$$

You can sometimes find the area of a shape by dividing it into trapeziums.

Example 10.6.2

Find the area of the pentagon in the figure.

Divide the pentagon into trapeziums A and B. Then:

Area of $A = \frac{1}{2} \times (3 + 9) \times 4\ \text{cm}^2 = 24\ \text{cm}^2$

Area of $B = \frac{1}{2} \times (9 + 5) \times 8\ \text{cm}^2 = 56\ \text{cm}^2$

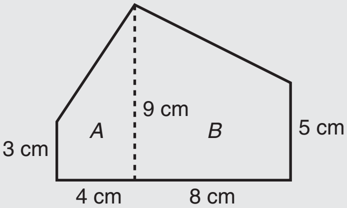

Area of pentagon $= (24 + 56)\ \text{cm}^2 = 80\ \text{cm}^2$

EXERCISE 10.5

1 Find the area of each of the trapeziums shown. The lengths are in centimetres.

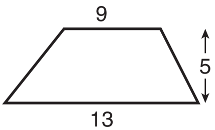

(a)

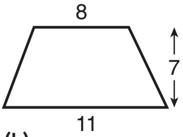

(b)

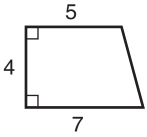

(c)

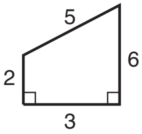

(d)

2 By taking measurements, find the area of this trapezium.

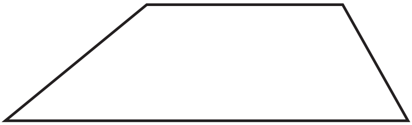

3 Find the areas of trapeziums with vertices at

(a) (1, 1), (2, 3), (5, 3), (8, 1) **(b)** (2, 2), (2, 4), (5, 5), (5, 0)

(c) (2, 3), (2, 6), (6, 6), (6, 1)

4 Find the area of these polygons. The units are centimetres.

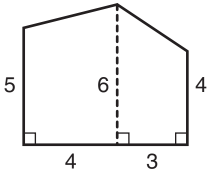

(a)

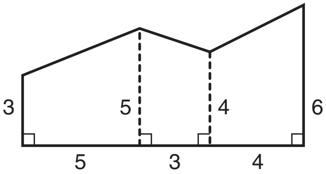

(b)

5 Two of the vertices of an isosceles trapezium are at (1, 1) and (3, 2). The line with equation $y = 3$ is a line of symmetry. Find the area of the trapezium.

6 Three of the vertices of a hexagon are at (1, 1), (3, 2) and (1, 3). The line with equation $x = 4$ is a line of symmetry. Find the area of the hexagon.

THINGS TO REMEMBER

- *The perimeter of a shape is the length of its boundary.*
- *Perimeter of a rectangle = 2 × length + 2 × width*
- *Area is a measurement of amount of surface.*
- *Area of a rectangle = length × width*
- *Area of a square = length of side × length of side*
- *Area of parallelogram = base × height*
- *Area of triangle =* $\frac{1}{2}$ *× base × height*
- *Area of a trapezium =* $\frac{1}{2}$ *× sum of parallel sides × vertical distance between them.*

11

Algebraic expressions

In this chapter you will learn:

- ***how to write and simplify algebraic expressions***
- ***how to evaluate algebraic expressions***
- ***how to use brackets***
- ***about indices***
- ***how to use the laws of indices.***

11.1 Introduction – what is algebra?

The Arabic word *algebra* originally meant the study of equations but has come to mean the whole branch of mathematics in which letters and symbols are used to generalize mathematical relations.

Many of the techniques of solving equations were known to the ancient Babylonians about 4000 years ago but modern, symbolic algebraic notation was not invented until the seventeenth century.

11.2 Writing expressions

Suppose that you are driving along the road, and that you are going at 50 mph. In one hour you will go 50 miles, and in 3 hours you will go 150 miles. To find the distance you travel, you multiply the speed by the time. So if you travel for t hours, then you would travel

$50 \times t$ miles. The letter t stands for a number. Using the expression $50 \times t$, you can calculate the distance travelled for any time.

As letters stand for numbers, you will want to add, subtract, multiply and divide them. You can do all these processes with letters but, sometimes, the answers are written in shortened forms.

Insight

A common mistake is to think of the letters standing for words. So you can't use *a* and *b* to stand for apples and bananas, but they could represent the cost in pence of an apple and a banana.

If y stands for a number and you were asked to write down a number which is 3 more than y, think how you would answer the question if it were asked with numbers instead of letters. Just as $4 + 3$ is 3 more than 4, so $y + 3$ (or $3 + y$) is 3 more than y. Similarly, the number which is 3 less than y is $y - 3$. This is *not* the same as $3 - y$, which is the number that is y less than 3.

The number which is 5 times greater than y could be written as $5 \times y$ (or $y \times 5$), but the multiplication sign is usually omitted and this would be written as $5y$. Notice that the number is written in front of the letter. In the same way, the distance travelled by car earlier in the section would be written as $50t$ miles. The number which is $\frac{1}{3}$ of y is written $\frac{y}{3}$ or $\frac{1}{3}y$. It is not usual, but it is not wrong, to have division signs in algebraic expressions.

$3a$ means $3 \times a$ ab means $a \times b$ $3ab$ means $3 \times a \times b$.

$\frac{a}{3}$ or $\frac{1}{3}a$ means $a \div 3$. $\frac{a}{b}$ means $a \div b$.

Example 11.2.1

The sum of two numbers is 9. One of the numbers is x. Write down an expression for the other number.

The other number must be x less than 9, which is $9 - x$.

Example 11.2.2

A man drives at s miles per hour for t hours. Write down an expression for the distance he travels.

To find the distance he travels, multiply his speed by the time that he drives. The distance he travels is $s \times t$ miles, or st miles.

Notice that $s \times t$ is acceptable as an answer but you would normally omit the multiplication sign and write the answer as st miles.

In general, you write numbers first and then letters, so $3a$ and not $a3$ for $3 \times a$. You normally write letters in alphabetical order, so ab not ba.

EXERCISE 11.1

In each case, write your answer in the form without multiplication or division signs.

1 x is a number. Write down an expression for a number which is
(a) 2 more than x **(b)** 1 less than x
(c) 4 times x **(d)** $\frac{1}{2}$ of x

2 p and q are two numbers. Write down an expression for
(a) the sum of p and q, **(b)** the product of p and q.

3 The product of two numbers is 8. One of the numbers is x. Write down an expression for the other number.

4 a and b are two numbers and a is greater than b. Write down an expression for the difference between the numbers.

5 There are 30 students in a class and b of them are boys. Write down an expression for the number of girls in the class.

6 Write these expressions in their simplest forms.
(a) $6 \times x$ **(b)** $y \div 5$ **(c)** $x \times y$ **(d)** $3 \times x \times y$
(e) $x \div y$ **(f)** $5 \times x \div y$ **(g)** $x \times y \div 5$ **(h)** $x \times y \times z$

7 Write down an expression for the number of days in w weeks.
8 Write down the number of metres in d centimetres.
9 A railway carriage has p seats. Write down an expression for the number of seats in n carriages.
10 When a apples are shared among b children, none are left over. Write down the number of apples each child receives.

11.3 Simplifying expressions

It is useful to be able to make algebraic expressions simpler if possible. For example:

$a + a + a + a$ is written as $4a$, so $a + a + a + a = 4a$.
$3b + 2b$ is written as $5b$, so $3b + 2b = 5b$.
$7c - 4c$ is written as $3c$, so $7c - 4c = 3c$.

Insight

You can see this works for any number because:

$5 + 5 + 5 + 5 = 4$ lots of 5 or 4×5
3 lots of 7 + 2 lots of 7 = 5 lots of 7
7 lots of 8 – 4 lots of 8 = 3 lots of 8

However, in the expression $3d + 5e + 4d - 2e$, the d's and the e's must be dealt with separately. As $3d + 4d = 7d$ and $5e - 2e = 3e$, $3d + 5e + 4d - 2e = 7d + 3e$.

Similarly, $5a - 2b - 3a + b = 2a - b$. Notice that $1b = b$; that is, it is conventional to leave out the 1.

In the expression $7d + 3e$, $7d$ and $3e$ are called **terms**. Terms such as $3b$ and $2b$ are called **like** terms, and terms such as $7d$ and $3e$ are called **unlike** terms.

Thus, in the expression $2a + 3b + 4a - 2c$, there are four terms. Two of them, $2a$ and $4a$, are like each other, and the other terms are unlike each other and unlike the terms $2a$ and $4a$.

Like terms can be combined by adding and subtracting them. This process is called **collecting like terms.** When the like terms have been collected, the expression is said to be **simplified.**

Unlike terms cannot be collected together in the same way.

Example 11.3.1

Simplify **(a)** $6f + 5 + 2f - 3$, **(b)** $3ab + 2bc - ab - abc$.

(a) Collecting like terms gives $6f + 2f = 8f$ and $5 - 3 = 2$, so $6f + 5 + 2f - 3 = 8f + 2$. You will usually write this straight down without intermediate steps.

(b) The only like terms are $3ab$ and $-ab$ so $3ab + 2bc - ab - abc = 2ab + 2bc - abc$.

Example 11.3.2

The length of each side of an equilateral triangle is d centimetres. Write down, in its simplest form, an expression for its perimeter.

The perimeter is $d + d + d$, but it is not in its simplest form, which is $d + d + d = 3d$.

EXERCISE 11.2

EXERCISE

1 The length of each side of a regular pentagon is l centimetres. Write down and simplify an expression for its perimeter.

2 The length of a rectangle is a centimetres and its width is b centimetres. Write down an expression, in its simplest form, for **(a)** its perimeter, **(b)** its area.

3 Write down an expression for the perimeter of the hexagon in the figure, giving your answer in its simplest form.

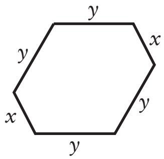

4 The lengths, in centimetres, of the sides of a trapezium are a, a, $2a$ and b. Find and simplify an expression for its perimeter.

5 Write each of these expressions in its simplest form.

(a) $x + x$
(b) $x + x + x + x$
(c) $x + x + x + x + x + x + x$
(d) $x + x - x + x + x - x + x$
(e) $4x + 2x + 3x$
(f) $3y + 2y - 4y$
(g) $4xy - 7xy + xy$
(h) $3abc + 2abc - 6abc$
(i) $5x + 3y + 4x + 2y$
(j) $4x + 7 - 2x - 4$
(k) $6x + 3 - x + 2x$
(l) $3x + 4y + 2x - 3y - y$
(m) $7xy + 5x - 4x + 2xy - 3$
(n) $4a + 5b + 3a + 6 - 7b$
(o) $a + 3ab + 6 + 7a - 2ab - 1$
(p) $3a + 6 + 2a - 4 + a - 2$
(q) $8 + 3pq - 5 - 7pq - 1 + r$
(r) $5p + q - 3r + 2r - 4p + r$

11.4 Evaluating expressions

If you know the values of the letter or letters used in an expression, you can calculate its value. For example, if $x = 2$, then $3x - 1 = 6 - 1 = 5$. Substituting numbers for letters and calculating the value of an expression is called **evaluating** the expression.

Insight

The rules for the order in which the operations must be carried out were introduced in Section 1.3. Multiplication and division are carried out before addition and subtraction. Remember: BiDMAS.

Example 11.4.1

If $x = 5$, evaluate **(a)** $4x + 1$, **(b)** $6 - 2x$, **(c)** $\frac{x+1}{2}$.

(a) $4x + 1 = 4 \times 5 + 1$ Multiplication comes before addition
$= 20 + 1$
$= 21$

(b) $6 - 2x = 6 - 2 \times 5$ Multiplication comes before subtraction
$= 6 - 10$
$= -4$

(c) $\frac{x+1}{2} = \frac{5+1}{2} = \frac{6}{2}$ Evaluate the numerator before dividing
$= 3$

You may want to leave out some steps as you become confident.

Example 11.4.2

If $x = 6$ and $y = 3$, evaluate **(a)** $2xy$, **(b)** $5x - 7y$, **(c)** $\frac{x}{y}$.

(a) $2xy = 2 \times 6 \times 3 = 36$
(b) $5x - 7y = 5 \times 6 - 7 \times 3 = 30 - 21 = 9$
(c) $\frac{x}{y} = \frac{6}{3} = 2$

Insight

Another way of asking part (a) of Example 11.4.2 is: 'Substitute $x = 6$ and $y = 3$ into the expression $2xy$.'

EXERCISE 11.3

EXERCISE

1 If $x = 3$, evaluate the following expressions.
(a) $5x + 4$ **(b)** $2x - 1$ **(c)** $10 - 3x$ **(d)** $1 - 2x$
2 If $x = 4$ and $y = 7$, evaluate the following expressions.
(a) $2x + y$ **(b)** $5x - 2y$ **(c)** $4x + 5y$ **(d)** $3x - 4y$
3 If $a = 2$, $b = 3$ and $c = 10$, evaluate the following expressions.
(a) $ab + c$ **(b)** abc **(c)** $5abc$ **(d)** $2a + bc$

11.5 Squaring

The notation x^2 is short for $x \times x$ and is read as 'x squared'.

The notation $4x^2$ means $4 \times x \times x$. (Only the x is squared, not the 4.)

Example 11.5.1

If $x = 5$, evaluate **(a)** x^2, **(b)** $3x^2$.

(a) $x^2 = 5 \times 5 = 25$ **(b)** $3x^2 = 3 \times 5 \times 5 = 3 \times 25 = 75$

Example 11.5.2

If $x = -3$, evaluate $6x^2 - 7$.

$6x^2 - 7 = 6 \times (-3)^2 - 7 = 6 \times 9 - 7 = 54 - 7 = 47$

EXERCISE 11.4

1 If $x = 3$, evaluate the following.
(a) x^2 **(b)** $x^2 + 1$ **(c)** $5x^2$ **(d)** $2x^2 - 21$

2 If $y = -4$, evaluate the following.
(a) y^2 **(b)** $y^2 - 3$ **(c)** $2y^2 + 5$ **(d)** $2y^2 - 2y$

3 If $a = 5$, $b = -2$, $c = 6$ and $d = 2$ evaluate the following.
(a) $4a^2 - 3a$ **(b)** $3b^2 - 5b$ **(c)** $2c^2 + 4$ **(d)** $d^2 + 2d + 1$

11.6 Brackets

In expressions, calculations inside brackets are always carried out first. The expression $a(b + c)$ means $a \times (b + c)$.

Insight

Operations in algebra follow the same order as in arithmetic, so BiDMAS (Brackets, then Division and Multiplication, lastly Addition and Subtraction) still applies.

So, if you knew the value of x and had to evaluate the expression $6(x + 1)$, you would first evaluate the bracket and then multiply that value by 6. So, if $x = 3$, $6(x + 1) = 6(3 + 1) = 6 \times 4 = 24$.

Example 11.6.1

If $a = 8$ and $b = -5$, evaluate **(a)** $(2b)^2$, **(b)** $(3a + 4b)^2$.

(a) $(2b)^2 = (2 \times (-5))^2 = (-10)^2 = 100$

(b) $(3a + 4b)^2 = (3 \times 8 + 4 \times (-5))^2 = (24 - 20)^2 = 4^2 = 16$

The approach of Example 11.6.1 is fine if you know the values of the letters, but, if you don't, you need to be able to write expressions such as $4(3y + 2)$ and $4(3y - 2)$ in an alternative form.

Suppose you needed to calculate $2(6 + 3)$; you can see that the answer is $2 \times 9 = 18$. But this is the same as $2 \times 6 + 2 \times 3 = 12 + 6 = 18$. This suggests that $2(6 + 3) = 2 \times 6 + 2 \times 3$, or, in general,

$$a(b + c) = ab + ac$$

A similar argument would convince you that

$$a(b - c) = ab - ac$$

If you look at this expression, you can see that each term inside the brackets is multiplied by the term outside.

So $4(3y + 2) = 4 \times (3y) + 4 \times 2 = 12y + 8$, and $4(3y - 2) = 12y - 8$. This process is called **expanding** the brackets.

Suppose now that the sign outside the bracket is negative, and you want to calculate $24 - 2(6 + 3)$. This is $24 - 2 \times 9 = 24 - 18 = 6$. But $24 - 2 \times 6 - 2 \times 3 = 24 - 12 - 6 = 6$, so this suggests that

$$-a(b + c) = -ab - ac$$

Finally, $24 - 2(6 - 3)$, is equal to $24 - 2 \times 3 = 18$. But you find that $24 - 2 \times 6 + 2 \times 3 = 24 - 12 + 6 = 18$. This suggests that

$$-a(b - c) = -ab + ac$$

The – sign outside the brackets changes the sign of every term inside the brackets.

Example 11.6.2

Expand the brackets in these expressions.

(a) $5(x - 2)$ **(b)** $2(4x - 3)$ **(c)** $4x(2x - 3)$ **(d)** $-3(5x - 2)$

(a) $5(x - 2) = 5x - 10$ **(b)** $2(4x - 3) = 8x - 6$

(c) $4x(2x - 3) = 8x^2 - 12x$ **(d)** $-3(5x - 2) = -15x + 6$

Example 11.6.3

Expand and simplify these expressions.

(a) $3(7x + 4) - 2(5x - 6)$ **(b)** $2(4x + 5) + (3x - 7)$

(a) $3(7x + 4) - 2(5x - 6) = 21x + 12 - 10x + 12 = 11x + 24$

(b) $2(4x + 5) + (3x - 7)$ means $2(4x + 5) + 1(3x - 7)$.

So $2(4x + 5) + (3x - 7) = 8x + 10 + 3x - 7 = 11x + 3$

EXERCISE 11.5

1 Expand the brackets in these expressions.

(a) $6(x + 3)$ **(b)** $4(x - 1)$ **(c)** $5(2x + 3)$
(d) $2(3x - 5)$ **(e)** $-4(x + 3)$ **(f)** $-7(x - 2)$
(g) $-8(5x + 4)$ **(h)** $-6(3x - 2)$ **(i)** $x(x + 2)$
(j) $x(4x - 3)$ **(k)** $3x(2x - 1)$ **(l)** $-5x(3x + 4)$

2 Expand the brackets in these expressions and simplify the results.

(a) $3(x - 7) + 6x$ **(b)** $8 - 2(5x - 1)$
(c) $4(x + 2) + 5(x - 3)$ **(d)** $3(x + 4) - 2(x - 5)$
(e) $2(6x - 5) + 3(2x - 1)$ **(f)** $5(4x + 3) - 3(7x - 2)$

(Contd)

(g) $3(2x + 5) + 5(4x - 3)$ **(h)** $4(2x - 1) - 2(4x - 5)$
(i) $x(x + 2) - 6(x + 4)$ **(j)** $x(x + 3) - 3(x - 8)$
(k) $x(x - 7) + 4(2x - 3)$ **(l)** $3x(2x - 5) + 7(3x + 4)$
(m) $3(6x + 2) + (5x - 4)$ **(n)** $8(3x + 1) - (6x + 7)$
(o) $5(2x - 3) - (10x - 9)$ **(p)** $3 + (5x - 2)$
(q) $6x - (6x + 5)$ **(r)** $2(4x + 3) - (7x + 6)$

11.7 Factorizing expressions

Factorizing is the opposite process to expanding.

To factorize $7x - 21$, the first step is to find a term which is a factor of both $7x$ and 21 (7 in this case) and this term goes outside the brackets. Then complete the brackets with an expression which, when multiplied by 7, gives $7x - 21$. That is, $7x - 21 = 7(x - 3)$.

Example 11.7.1

Factorize **(a)** $5x + 20$, **(b)** $6x - 9$, **(c)** $bx - 7b$, **(d)** $x^2 + 8x$.

(a) $5x + 20 = 5(x + 4)$ **(b)** $6x - 9 = 3(2x - 3)$
(c) $bx - 7b = b(x - 7)$ **(d)** $x^2 + 8x = x(x + 8)$

In these examples, each pair of terms has only one common factor. Sometimes the terms in the expression you want to factorize will have more than one common factor.

In $6x^2 + 15x$, for example, 3, x and $3x$ are all factors of both $6x^2$ and $15x$. It might seem, therefore, that, if asked to factorize $6x^2 + 15x$, you could give three possible answers:

$$6x^2 + 15x = 3(2x^2 + 5x) \text{ or } x(6x + 15) \text{ or } 3x(2x + 5)$$

In each of the first two cases, the terms inside the brackets have a common factor but, in the last case, they have not and $6x^2 + 15x$ is said to have been **completely factorized** or **fully factorized**.

Example 11.7.2

Factorize completely **(a)** $8x^2 - 12x$ **(b)** $20ab + 15bc$

(a) $8x^2 - 12x = 4x(2x - 3)$ **(b)** $20ab + 15bc = 5b(4a + 3c)$

EXERCISE 11.6

1 Factorize these expressions completely.

(a) $3x + 9$ **(b)** $14x - 21$ **(c)** $20x + 15$ **(d)** $6x - 15y$
(e) $ax - 7a$ **(f)** $x^2 + 4x$ **(g)** $12x - 18$ **(h)** $15x - 12$
(i) $4ax + 6bx$ **(j)** $9x^2 + 12x$ **(k)** $20x - 8x^2$ **(l)** $4bx - 2bx^2$
(m) $x^2y + xy^2$ **(n)** $6ax^2 - 15a^2x$ **(o)** $8xy + 12xy^2$

11.8 Indices

In Section 11.5, you saw how $x \times x$ is written as x^2. In a similar way, $x \times x \times x$ is written as x^3, which could be read as 'x to the power 3' but is usually read as 'x cubed'. The 3 is called the **index** or **power**.

$x \times x \times x \times x$ is written as x^4, which is read as 'x to the power 4'.

Higher powers are written as x^5, x^6 and so on. $4x^3$ is algebraic shorthand for $4 \times x \times x \times x$. Only the x is cubed, not the 4. In other words, you cube *before* you multiply.

Example 11.8.1

Write more simply

(a) $c \times c \times c \times c \times c \times c$ **(b)** $6 \times d \times d \times d \times d \times d$
(a) $c \times c \times c \times c \times c \times c = c^6$ **(b)** $6 \times d \times d \times d \times d \times d = 6d^5$

Example 11.8.2

If $p = 6$, evaluate **(a)** p^3, **(b)** $5p^4$.

(a) $p^3 = 6 \times 6 \times 6 = 216$
(b) $5p^4 = 5 \times 6 \times 6 \times 6 \times 6 = 5 \times 1296 = 6480$

You can find powers of numbers using the y^x key on your calculator. (Some calculators have differently labelled keys for finding powers of numbers. You may need to look at your calculator handbook.)

Example 11.8.3

If $q = -5$, evaluate $4q^3$.

$4q^3 = 4 \times (-5) \times (-5) \times (-5) = 4 \times (-125) = -500$

You can use the ± key on your calculator when finding powers of negative numbers.

EXERCISE

EXERCISE 11.7

1 Write the following in a simpler form.
  (a) $m \times m \times m$ **(b)** $n \times n \times n \times n \times n$
  (c) $5 \times p \times p \times p \times p \times p$ **(d)** $3 \times q \times q \times q \times q \times q \times q$
2 If $p = 3$, evaluate **(a)** p^5 **(b)** $5p^3$
3 If $q = 10$, evaluate **(a)** q^6 **(b)** $7q^4$
4 If $u = -4$, evaluate **(a)** u^3 **(b)** $3u^5$
5 If $v = -5$, evaluate **(a)** $4v^3$ **(b)** $6v^4$

11.9 Laws of indices

To multiply a^3 by a^4, you could begin by writing down what each expression means. $a^3 = a \times a \times a$ and $a^4 = a \times a \times a \times a$, so

$$a^3 \times a^4 = (a \times a \times a) \times (a \times a \times a \times a) = a^7$$

The index 7 in the answer is the sum 3 + 4 of the indices 3 and 4 in the question. It suggests the general result

$$a^m \times a^n = a^{m+n}$$

This result enables you to write down products of powers of the same letter without intermediate working.

Example 11.9.1

Simplify $a^2 \times a^6$.

Adding the indices, $a^2 \times a^6 = a^{2+6} = a^8$

You cannot add the indices unless the letters are the same. You cannot, for example, simplify expressions like $a^3 \times b^4$ by adding the indices.

To divide a^6 by a^2, note that $a^6 = a \times a \times a \times a \times a \times a$ and $a^2 = a \times a$, so $a^6 \div a^2 = \dfrac{a \times a \times a \times a \times a \times a}{a \times a}$

You can now 'cancel' the 'fraction' in a way similar to that in which numerical fractions were simplified in Section 3.4.

$$\frac{a \times a \times a \times a \times \cancel{a} \times \cancel{a}}{\cancel{a} \times \cancel{a}} = a^4, \text{ that is, } a^6 \div a^2 = a^{6-2} = a^4.$$

The index 4 in the answer is the difference 6 – 2 between the indices in the question. It suggests the general result

$$a^m \div a^n = a^{m-n}, \text{ provided that } m \text{ is greater than } n$$

Similarly, if you were dividing a^2 by a^6, you would find that

$$a^2 \div a^6 = \frac{a \times a}{a \times a \times a \times a \times a \times a} = \frac{1}{a^4}.$$

This suggests the general result

$$a^m \div a^n = \frac{1}{a^{n-m}}, \text{ provided that } n \text{ is greater than } m$$

You can also say that $a^2 \div a^6 = a^{2-6} = a^{-4}$. Since $a^2 \div a^6 = \frac{1}{a^4}$ and a^{-4}, you can see that $\frac{1}{a^4} = a^{-4}$. In general,

$$\frac{1}{a^n} = a^{-n}$$

Using these general results, you can divide powers of the same letter.

Example 11.9.2

Simplify **(a)** $a^9 \div a^4$, **(b)** $a^3 \div a^8$.

(a) Subtracting the indices, $a^9 \div a^4 = a^5$

(b) $a^3 \div a^8 = \frac{1}{a^{8-3}} = \frac{1}{a^5}$

Using the general result to simplify $a^4 \div a^3$ gives the answer a^1, but

$$a^4 \div a^3 = \frac{a \times a \times a \times a}{a \times a \times a} = a \quad \text{so} \quad a^1 = a$$

Example 11.9.3

Simplify **(a)** $a \times a^3$, **(b)** $a^3 \div a$.

(a) $a \times a^3 = a^1 \times a^3 = a^4$ **(b)** $a^3 \div a = a^3 \div a^1 = a^2$

To simplify $(a^4)^3$, write it as $(a^4)^3 = a^4 \times a^4 \times a^4 = a^{4\times 3} = a^{12}$.

This suggests the general result

$$(a^m)^n = a^{mn}$$

Example 11.9.4

Simplify $(a^3)^5$.

Multiplying the indices, $(a^3)^5 = a^{3\times 5} = a^{15}$

EXERCISE 11.8

1 Simplify the following where possible.

(a) $x^3 \times x^4$ **(b)** $x^7 \div x^2$ **(c)** $(x^4)^3$ **(d)** $x \times x^5$

(e) $x^9 \div x^4$ **(f)** $(x^5)^2$ **(g)** $x^4 \times x^6$ **(h)** $x^3 \times y^4$

(i) $(x^2)^2$ **(j)** $x^6 \div x$ **(k)** $x^5 \div x^3$ **(l)** $x^3 \times x$

(m) $x^6 \div x^5$ **(n)** $x^7 \times x^3$ **(o)** $x^3 \div x$ **(p)** $(x^3)^3$

11.10 Simplifying expressions with indices

To simplify $4a^3 \times 5a^2$ consider first what it means.

$$4a^3 \times 5a^2 = (4 \times a^3) \times (5 \times a^2) = (4 \times 5) \times (a^3 \times a^2) = 20a^5$$

This result shows that you can consider the numbers and the powers of a separately. Since $4 \times 5 = 20$ and $a^3 \times a^2 = a^5$, $4a^3 \times 5a^2 = 20a^5$. You can use this method to simplify expressions with more letters.

Example 11.10.1

Simplify

(a) $3a^2b^4 \times 6a^5b^2$ **(b)** $3a^5 \times 4a$ **(c)** $7a^3 \times a^2$ **(d)** $8a^5 \times a$

(a) $3a^2b^4 \times 6a^5b^2 = (3 \times 6) \times (a^2 \times a^5) \times (b^4 \times b^2) = 18a^7b^6$

(b) $3a^5 \times 4a = 12a^6$ (Remember $4a$ means $4a^1$.)

(c) $7a^3 \times a^2 = 7a^5$

(d) $8a^5 \times a = 8a^6$

To simplify $12a^5 \div 3a^2$, you can again consider the numbers and the powers of a independently.

$$12a^5 \div 3a^2 = \frac{12 \times a^5}{3 \times a^2} = \frac{12}{3} \times \frac{a^5}{a^2} = 4a^3$$

Example 11.10.2

Simplify **(a)** $24a^6 \div 8a^4$, **(b)** $\frac{18a^8}{3a^5}$.

(a) $24a^6 \div 8a^4 = 3a^2$ **(b)** $\frac{18a^8}{3a^5} = 6a^3$

To simplify $\dfrac{10a^4b^5}{2a^3b^2}$

$$\frac{10a^4b^5}{2a^3b^2} = \frac{10}{2} \times \frac{a^4}{a^3} \times \frac{b^5}{b^2} = 5ab^3$$

Example 11.10.3

Simplify **(a)** $\frac{8a^5b^3}{2a^3b}$, **(b)** $5a^6b^7 \div a^4b^3$, **(c)** $\frac{15a^5b^2}{3a^2b^3}$.

(a) $\frac{8a^5b^3}{2a^3b} = \frac{8}{2} \times \frac{a^5}{a^3} \times \frac{b^3}{b} = 4a^2b^2$

(b) $5a^6b^7 \div a^4b^3 = \frac{5a^6b^7}{a^4b^3} = \frac{5}{1} \times \frac{a^6}{a^4} \times \frac{b^7}{b^3} = 5a^2b^4$

(c) $\frac{15a^5b^2}{3a^2b^3} = \frac{15}{3} \times \frac{a^5}{a^2} \times \frac{b^2}{b^3} = \frac{5a^3}{b}$

You may find it helpful at first to write the expressions out in full.

Example 11.10.4

Simplify **(a)** $\frac{4a^2}{2a^5}$, **(b)** $\frac{6a^8b^3}{2a^5b^7}$, **(c)** $\frac{3a^5b^2}{6a^4b^4}$, **(d)** $\frac{14a^3b^2}{2b^2c}$.

(a) $\frac{4a^2}{2a^5} = \frac{4}{2} \times \frac{a^2}{a^5} = \frac{2}{a^3}$

(b) $\frac{6a^8b^3}{2a^5b^7} = \frac{6}{2} \times \frac{a^8}{a^5} \times \frac{b^3}{b^7} = \frac{3a^3}{b^4}$

(c) $\frac{3a^5b^2}{6a^4b^4} = \frac{3}{6} \times \frac{a^5}{a^4} \times \frac{b^2}{b^4} = \frac{a}{2b^2}$

(d) $\frac{14a^3b^2}{2b^2c} = \frac{14}{2} \times \frac{a^3}{1} \times \frac{b^2}{b^2} \times \frac{1}{c} = \frac{7a^3}{c}$, as $\frac{b^2}{b^2} = 1$

To simplify $(3a^4)^2$, a good method is to write it as $3a^4 \times 3a^4$, which is $9a^8$; both the 3 and the a^4 have been squared. With practice, you should be able to write the answer straight down.

With higher powers, it can be easier to use the index law $(a^m)^n = a^{mn}$. Thus, $(5a^2b)^3 = 5^3 \times (a^2)^3 \times b^3 = 125a^6b^3$. Note that each term inside the bracket, 5, a^2 and b, has been cubed.

Insight

To check whether a simplified expression could be right, substitute the same values for the letters into both the simplified and unsimplified expressions; if you have simplified correctly you will get the same answer for both.

Example 11.10.5

Simplify **(a)** $(5a)^2$, **(b)** $(4a^2b^3)^2$, **(c)** $(2a^4)^3$, **(d)** $(4a^2b^5)^3$.

(a) $(5a)^2 = 5a \times 5a = 25a^2$
(b) $(4a^2b^3)^2 = 4a^2b^3 \times 4a^2b^3 = 16a^4b^6$
(c) $(2a^4)^3 = 2^3 \times (a^4)^3 = 8a^{12}$
(d) $(4a^2b^5)^3 = 4^3 \times (a^2)^3 \times (b^5)^3 = 64a^6b^{15}$

EXERCISE 11.9

1 Simplify the following.
(a) $3x^2 \times 2x^3$ **(b)** $5x \times 4x^2$ **(c)** $7x^2y^4 \times 2x^3y^2$
(d) $9x^7 \times x$ **(e)** $6x^4 \times x^5$ **(f)** $8x^3y^3 \times x^4y^2$
(g) $12x^5 \div 4x^3$ **(h)** $\frac{21x^7}{3x^4}$ **(i)** $\frac{10x^6y^4}{2x^4y^2}$

(Contd)

EXERCISE

2 Simplify the following.

(a) $\frac{24x^5y^4}{8x^4y}$ **(b)** $6x^7y^6 \div x^2y^4$ **(c)** $\frac{10x^7y^3}{2x^3z^2}$

(d) $\frac{a^4}{a^9}$ **(e)** $\frac{16x^2}{2x^6}$ **(f)** $\frac{8x^6y^3}{2x^2y^5}$

(g) $\frac{5x^3y^6}{15x^4y^2}$ **(h)** $\frac{20x^3y^4}{4x^3}$ **(i)** $(6x)^2$

(j) $(5xy)^2$ **(k)** $(x^5y^3)^2$ **(l)** $(5x^3y^4)^2$

(m) $(10x)^3$ **(n)** $(4x^5)^3$ **(o)** $(5x^4y^3)^3$

THINGS TO REMEMBER

- *In algebra, letters can be used to represent numbers.*
  - $5y$ *means* $5 \times y$ *or* $y + y + y + y + y$.
  - $1y$ *is written as* y.
  - mn *means* $m \times n$.
  - $4pq$ *means* $4 \times p \times q$.
  - $\frac{x}{2}$ *or* $\frac{1}{2}x$ *means* $x \div 2$.
  - $\frac{a}{b}$ *means* $a \div b$.
  - x^2 *means* $x \times x$ *and is read as 'x squared'.*
  - $4b^2$ *means* $4 \times b^2$ *or '4b squared'.*
  - x^3 *means* $x \times x \times x$ *and is read as 'x to the power 3' or 'x cubed'.*

- *An expression is made up of terms. The expression* $4 + 2a - 3b$ *has three terms, 4, 2a and 3b. Like terms contain the same letters or combination of letters.*

- *To simplify an expression you should collect like terms and then combine them. Unlike terms cannot be simplified.*

- *You can evaluate an expression by substituting (or replacing) numbers for letters and then calculating the value of the resulting expression.*

- *You expand brackets by multiplying them out: multiply each term inside the brackets by the term outside the brackets.*
  - $a(b + c) = ab + ac$
  - $a(b - c) = ab - ac$
  - $-a(b + c) = -ab - ac$
  - $-a(b - c) = -ab + ac$

- *Factorizing is the opposite process to expanding, so you need to rewrite the expressions using brackets. Look for the highest common factor of the numbers and letters separately. For example 20ab + 15bc factorized fully is 5b(4a + 3c).*

- *The laws of indices are*
    - $a^m \times a^n = a^{m+n}$
    - $a^m \div a^n = a^{m-n}$, *if m is bigger than n*
    - $a^m \div a^n = \frac{1}{a^{n-m}}$, *if n is bigger than m*
    - $(a^m)^n = a^{mn}$

12

Approximation

In this chapter you will learn:

- ***how to round whole numbers and decimals***
- ***about significant figures***
- ***how to find estimates***
- ***how to use rounding in problems***
- ***about the accuracy of measurements.***

12.1 Introduction

Sometimes exact answers to a problem are unnecessary or even impossible, and a sensible *approximate* answer is needed. For example, in a census, the population of Coventry was given as 304 426 but, for most practical purposes, 300 000 would be an adequate approximation.

You will meet a variety of ways of approximating numbers. Then you will use approximate values to find estimates to calculations, to solve problems and see how approximation and measurement are related.

12.2 Rounding whole numbers

One method of approximating is to give, or **round**, the number to the nearest ten, hundred, thousand, etc.

For example, 22 is between 20 and 30 but it is nearer to 20 than 30. So, when you round 22 to the nearest ten, you **round down** to get 20. When you round 67 to the nearest ten, however, you **round up** to get 70, as 67 is nearer to 70 than to 60.

If a number is exactly halfway between two multiples of 10, the convention is to **round up**. For example, 85 is exactly halfway between 80 and 90 so, to round 85 to the nearest ten, you round up to 90.

In other words, if the digit in the units column is less than 5, you do not alter any of the previous digits, but if the digit in the units column is 5 or more, you increase the tens digit by one. In both cases, your answer will have a zero in the units column.

Example 12.2.1

Round each of the following to the nearest ten.
(a) 392 **(b)** 4697

(a) 392 is between 390 and 400. The 2 tells you to round down, so 392 to the nearest ten is 390.
(b) 4698 is between 4690 and 4700. The 8 tells you to round up, so 4698 to the nearest ten is 4700.

Similarly, you can round numbers to the nearest hundred, when the tens digit tells you whether to round up or down, or to the nearest thousand, when it is the hundreds digit which tells you.

Example 12.2.2

Round to the nearest hundred: **(a)** 349 **(b)** 2861 **(c)** 31 750

(a) 349 is between 300 and 400. The 4 tells you to round down, so 349 to the nearest hundred is 300.
(b) 2861 is between 2800 and 2900. The 6 tells you to round up, so 2861 to the nearest hundred is 2900.
(c) 31 750 is between 31 700 and 31 800. The 5 tells you to round up, so 31 750 to the nearest hundred is 31 800.

EXERCISE 12.1

1 Write each of these numbers to the nearest ten.
(a) 34 (b) 45 (c) 76
(d) 751 (e) 8465 (f) 7396

2 Write each of these numbers to the nearest hundred.
(a) 675 (b) 849 (c) 350
(d) 4351 (e) 2974 (f) 19 483

3 Write each of these numbers to the nearest thousand.
(a) 7682 (b) 3429 (c) 5500
(d) 37 743 (e) 42 499 (f) 99 682

4 The attendance at a Manchester United soccer match was 61 267. Round this number to the nearest hundred.

5 In a census, the population of Wolverhampton was given as 242 187. Round this number
(a) to the nearest thousand,
(b) to the nearest hundred thousand.

6 To the nearest hundred, the population of a town is 7600. Find
(a) the smallest, (b) the largest,
possible population the town could have.

12.3 Rounding with decimals

If you travel 70 miles in $1\frac{1}{2}$ hours, you can find your average speed (see page 118) by dividing 70 by 1.5. My calculator gives 46.66666667 as the answer to 70 ÷ 1.5 but it would not be sensible to write down the whole calculator display. An answer to the nearest **whole number** is appropriate, so as 46.666 666 67 is between 46 and 47, but nearer to 47, an average speed of 47 mph is a reasonable answer.

If the digit in the first decimal place is less than 5, you round down; that is, you leave the whole number unchanged. If it is 5 or more, you round up; that is, you increase the whole number by 1.

Example 12.3.1

Write each of the following to the nearest whole number.

(a) 4.369 **(b)** 64.7132 **(c)** 368.5

(a) 4.369 is between 4 and 5. The 3 tells you to round down, so 4.369 to the nearest whole number is 4.

(b) 64.7132 is between 64 and 65. The 7 tells you to round up, so 64.7132 to the nearest whole number is 65.

(c) 368.5 is halfway between 368 and 369. The 5 tells you to round up, so 368.5 to the nearest whole number is 369.

For greater accuracy, you can give numbers correct to one decimal place (1 d.p.). 7.346, for example, is between 7.3 and 7.4, but nearer to 7.3, so 7.346 = 7.3 (to 1 d.p.).

If the digit in the *second* decimal place (4 in 7.346) is less than 5, round down, leaving all the preceding digits unchanged. If it is 5 or more, round up, increasing the digit in the first decimal place by one.

Example 12.3.2

Write each of these numbers correct to one decimal place.

(a) 7.483 **(b)** 0.65 **(c)** 23.0481

(a) 7.483 is between 7.4 and 7.5; the 8 tells you to round up, so 7.483 = 7.5 (to 1 d.p.).

(b) 0.65 is halfway between 0.6 and 0.7; the 5 tells you to round up, so 0.65 = 0.7 (to 1 d.p.).

(c) The 4 in 23.0481 tells you that 23.0481 = 23.0 (to 1 d.p.).

Insight

In the answer 23.0, the zero *is* necessary; 23 alone is correct only to the nearest whole number.

When you express a number correct to one decimal place, you are giving it correct to the nearest *tenth*.

In a similar way, you can approximate to any number of decimal places. To give a number correct to two decimal places (nearest hundredth), look at the digit in the *third* decimal place; to give a number correct to three decimal places (nearest thousandth), look at the digit in the *fourth* decimal place, and so on.

Example 12.3.3

(a) Write 24.6472 correct to two decimal places.
(b) Write 0.063 279 correct to three decimal places.

(a) The digit in the *third* decimal place is 7 so, rounding up, 24.6472 = 24.65 (to 2 d.p.).
(b) The digit in the *fourth* decimal place is 2 so, rounding down, 0.063 279 = 0.063 (to 3 d.p.).

Example 12.3.4

Write 3.704 891 correct to **(a)** to 3 d.p. **(b)** to 2 d.p.

(a) 3.704 891 = 3.705 (to 3 d.p.)
(b) 3.704 891 = 3.70 (to 2 d.p.)

Notice that, if you give the answer to part (a) correct to two decimal places, you get 3.71, which is *not* the correct answer to part (b).

EXERCISE 12.2

1 Write each of the following to the nearest whole number.
(a) 8.74 **(b)** 21.241 **(c)** 67.5
(d) 260.2841 **(e)** 399.72

2 Write each of these numbers correct to one decimal place.
(a) 9.927 **(b)** 0.8643 **(c)** 17.9821
(d) 76.85 **(e)** 400.0389

3 Write each of these numbers correct to two decimal places.
(a) 4.2491 **(b)** 8.465 **(c)** 0.748 36
(d) 0.061 794

(Contd)

4 Write each of these numbers correct to three decimal places.
(a) 8.678 23 **(b)** 7.608 52 **(c)** 13.419 61
(d) 0.009 472

5 Calculate the value of 86 ÷ 9 and give your answer to the nearest whole number.

6 Calculate the value of 8.3 ÷ 3.9 and give your answer correct to the nearest tenth.

7 Calculate the value of 4.68 × 3.72 and give your answer correct to the nearest hundredth.

8 Convert $\frac{3}{7}$ to a decimal and give your answer correct to three decimal places.

9 Correct to one decimal place, a number is 7.3. Write down the smallest possible value of the number.

10 A number which is divisible by 5 has three decimal places. Correct to one decimal place, the number is 5.8 and, correct to two decimal places, it is 5.85. Find the number.

12.4 Significant figures

Giving a number to a specified number of **significant figures** is another method of approximating. For example, in the number 7483, the most *significant,* or important, figure is 7, as its value is 7000. To give 7483 correct to one significant figure (1 s.f.), you have to choose between 7000 and 8000, the rounded-up value.

To decide which it is, use the same rules you have used earlier in this chapter; you look at the *second* figure, 4, which tells you to leave the 7 unchanged. So 7483 = 7000 (to 1 s.f.), the same answer as giving it to the nearest thousand. Notice that, when you give a number correct to one significant figure, there is only one non-zero figure in your answer.

The second most significant figure in the number 7483 is 4, as its value is 400. To give 7483 correct to two significant figures (2 s.f.), you have to choose between 7400 and 7500. The third figure, 8, tells you to round up. So 7483 = 7500 (to 2 s.f.), the same answer as giving it to the nearest hundred.

When you give a number correct to two significant figures, there are usually two non-zero figures in your answer but there could be only one. For example, 698 = 700 (to 2 s.f.).

For numbers less than 1, zeros at the start of the number are *not* significant. For example, to give 0.0543 correct to one significant figure, the 4 tells you not to round up and so 0.0543 = 0.05 (to 1 s.f.), the same answer as giving it to two decimal places. As before, there is only one non-zero figure in your answer.

Example 12.4.1

Write each of these numbers correct to one significant figure.
(a) 5.23 **(b)** 2500 **(c)** 0.783 **(d)** 0.000 624

(a) 5.23 = 5 (to 1 s.f.) **(b)** 2500 = 3000 (to 1 s.f.)
(c) 0.783 = 0.8 (to 1 s.f.) **(d)** 0.000 624 = 0.0006 (to 1 s.f.)

Example 12.4.2

Write each of these numbers correct to three significant figures.
(a) 23.8486 **(b)** 0.006 083 4 **(c)** 0.000 379 62

(a) 23.8486 = 23.8 (to 3 s.f.)
(b) 0.006 083 4 = 0.006 08 (to 3 s.f.)
(c) 0.000 379 62 = 0.000 380 (to 3 s.f.)

EXERCISE 12.3

1 Write each of these numbers correct to one significant figure.
(a) 23.6 **(b)** 375 **(c)** 0.187
(d) 0.0072 **(e)** 0.085

2 Write each of these numbers correct to two significant figures.
(a) 8673 **(b)** 742.9 **(c)** 6.85
(d) 0.084 72 **(e)** 0.003 971

(Contd)

3 Write each of these numbers correct to three significant figures.
(a) 23 679 **(b)** 376.47 **(c)** 13.7819
(d) 8.925 **(e)** 0.084 382 **(f)** 0.000 436 5
(g) 0.090 237 **(h)** 0.003 697 4

4 Calculate the value of 43.7 × 36.8 and give your answer correct to one significant figure.

5 Calculate the value of 41 ÷ 7 and give your answer correct to two significant figures.

6 Calculate the value of
(a) 3.82 × 2.97 **(b)** 4.72 ÷ 9.84
Give your answers correct to three significant figures.

7 Convert $\frac{1}{11}$ to a decimal and give your answer correct to three significant figures.

8 Write 599.9 correct to
(a) 1 s.f. **(b)** 2 s.f. **(c)** 3 s.f.

9 Write 0.059 99 correct to
(a) 1 s.f. **(b)** 2 s.f. **(c)** 3 s.f.

12.5 Estimates

You should always make an estimate to check that your answer to a calculation is sensible. To do this, approximate each of the numbers involved and do the calculation mentally using these approximations.

For example, to estimate the answer to 27.36 × 0.241, you could say that 27.36 is near 28 and 0.241 is near $\frac{1}{4}$. The product, therefore, is near $28 \times \frac{1}{4}$, which is 7. The exact answer must be less than 7, because 28 and $\frac{1}{4}$ are both greater than the original numbers.

The usual method of estimating, however, is to give each number correct to one significant figure.

So $27.36 \times 0.241 \approx 30 \times 0.2 = 6$ (Recall that the symbol ≈ means ‘is approximately equal to’.)

Insight

Make sure that the approximate calculation is easy to work out in your head. If you have to reach for your calculator in order to work it out then there is no point in using the approximation!

Example 12.5.1

Find an estimate for **(a)** $76.48 \div 2.32$ **(b)** $\dfrac{8.67 \times 0.235}{0.058}$

(a) $76.48 \div 2.32 \approx 80 \div 2 = 40$

(b) $\dfrac{8.67 \times 0.235}{0.058} \approx \dfrac{9 \times 0.2}{0.06} = \dfrac{1.8}{0.06} = \dfrac{180}{6} = 30$

EXERCISE 12.4

1 Find an estimate for each of these calculations and then use a calculator to find the answer. Where necessary, give your answer correct to three significant figures.

(a) 2.39×3.74
(b) 34.91×2.83
(c) $31 \div 5.83$
(d) 247.1×47.8
(e) $543 \div 97.4$
(f) $76.4 \div 23.2$
(g) $32.8 \times 3.74 \times 66.8$
(h) $\frac{41.7 \times 8.92}{27.4}$
(i) 67.3×0.347
(j) $0.628 \div 1.84$
(k) $56.4 \div 0.286$
(l) 376×0.0354
(m) $0.0673 \times 0.002\,36$
(n) $8.32 \div 0.381$
(o) $0.0437 \div 0.176$
(p) $0.783 \div 0.0154$
(q) $0.0643 \div 2.816$
(r) $\frac{4.79 \times 0.0783}{0.247}$
(s) $\frac{0.42 \times 0.673}{0.0732}$
(t) $\frac{369 \times 0.527}{98.3 \times 0.0162}$

12.6 Rounding in practical problems

So far you have used rules to decide whether to round up or down but there are many real-life situations in which these rules are not relevant and you have to use common sense.

For example, if you had 375 eggs and wanted to know how many boxes, each holding 6 eggs, you could fill, you would probably use your calculator to work out 375 ÷ 6, obtaining the answer 62.5. To the nearest whole number 62.5 is 63, but there are not enough eggs to fill the sixty-third box and so you round down to 62.

The 62 full boxes contain 372 eggs (62 × 6), so there are 3 eggs left.

EXERCISE 12.5

1. How many boxes, each holding 6 eggs, can be filled with 250 eggs? How many eggs are left over?
2. How many pieces of metal 7 cm long can be cut from bar 60 cm long? What length of metal is left?
3. The emulsion paint in a 1 litre tin covers an area of about 12 m^2. How many tins of paint are needed to cover an area of 45 m^2?
4. The maximum number of people allowed in a lift is 30. Find the minimum number of trips the lift must make to carry 65 people.
5. A decorator needs 27 litres of masonry paint for the outside of a house. The paint is sold in 5 litre tins. How many tins must he buy?
6. How many cars, each 4.5 m long, would fit end to end on a ferry deck which is 31 m long?
7. 190 tiles are needed for a bathroom. The tiles are sold in boxes of 20. How many boxes are needed?
8. A railway carriage has seats for 56 passengers. How many carriages are needed for 400 passengers?

12.7 Accuracy of measurements

All measurements are approximate, even if you are very careful when making them and use sophisticated measuring instruments. For example, giving the length of a line as '7 cm correct to the nearest centimetre' means that its length lies between 6.5 cm and 7.5 cm, that is, 6.5 cm $\le$ actual length $<$ 7.5 cm. So the maximum possible error is 0.5 cm, which is half the unit being used for the measurement.

The length of the line might be given more accurately to the nearest 0.1 centimetre, or millimetre, as 7.3 cm. This means that its actual length lies between 7.25 cm and 7.35 cm and so the maximum possible error is 0.05 cm, which is again half the unit of measurement.

Even if the degree of accuracy is not explicitly stated, it can be implied by the number of decimal places given in the measurement.

Insight

Saying that the length of a line is 7.0 cm implies that the measurement is accurate to the nearest mm; however, 7 cm implies that the measurement is accurate to the nearest centimetre.

Example 12.7.1

The world record for the women's 100 metres is 10.49 seconds.

(a) To what degree of accuracy is this time given?

(b) Write down the minimum time and the maximum time this could be.

(c) What is the maximum possible error?

(a) 10.49 has two decimal places and so the time is correct to the nearest 0.01 second.

(b) The minimum time is 10.485 s; the maximum time is 10.495 s.

(c) The maximum possible error is 0.005 s.

Note that in part (b) you may think, reasonably, that the maximum time should not be 10.495 s because that would correct to 10.50 s, and that it should be the largest number which is just less than 10.495. However, there is no such number. No matter how close a number to 10.495 you choose, say $10.494\overbrace{999\ldots9}^{\text{100 of these}}$, there is always a closer number to 10.495, obtained by taking one more 9. Thus it is conventional to say that the maximum is 10.495, although strictly this is not true.

EXERCISE 12.6

1 The distance from London to Liverpool is 325 km, correct to the nearest kilometre. Write down the minimum distance and the maximum distance this could be.

2 The largest butterfly in the United Kingdom has a wingspan of 12.7 cm, correct to the nearest 0.1 cm. Write down the minimum length and the maximum length this could be.

3 The world record for the men's 100 metres is 9.85 seconds, correct to the nearest 0.01 second. Write down the minimum time and the maximum time this could be.

4 The heaviest baby on record weighed 10.6 kg at birth.

(a) To what degree of accuracy is this weight given?

(b) Write down the minimum and the maximum weight this could be.

(c) What is the maximum possible error in recording the weight as 10.6 kg?

5 The world record for the women's 400 metres is 47.60 seconds.

(a) To what degree of accuracy is this time given?

(b) What is the maximum possible error?

THINGS TO REMEMBER

- *To round to the nearest hundred, look at the digit in the tens place:*
  5 or more – round up
  4 or less – round down
  Use a similar method for rounding to the nearest unit, ten or thousand.
- *To give a number correct to one decimal place (1 d.p.), look at the digit in the second decimal place:*
  5 or more – round up
  4 or less – round down
- *The 1st significant figure in a number is the first non-zero digit in the number, the 2nd significant figure is the next digit and so on. So in the number 304 567 the 1st significant figure is 3 and the 2nd significant figure is 0. In 0.002 735 the 1st significant figure is 2 and the 2nd significant figure is 7.*
- *To give a number correct to two significant figures (2 s.f.), look at the third significant figure:*
  5 or more – round up
  4 or less – round down
- *When answering real-life problems you may need to use common sense when deciding to round up or down.*
- *To estimate the answer to a calculation you can round each number to one significant figure.*
- *All measurements are approximate and have an error of half a unit; the degree of accuracy is implied by the number of decimal places given in the measurement.*

13

Equations 1

In this chapter you will learn:

- ***how to solve linear equations***
- ***how to solve problems using equations***
- ***how to solve inequalities.***

13.1 Introduction

An **equation** is a mathematical statement that two expressions are equal, such as 3 × 2 = 6. Sometimes, missing numbers are denoted by symbols or letters, for example, $\square + 2 = 7$ or $x - 3 = 2$. In both cases, the equation is **satisfied**, or true, if you replace the missing number by 5, which is called the **solution**, or **root**, of the equation.

Mathematicians in Ancient Egypt and Babylon were able to solve many types of equation, which were expressed in words or drawings. In Greece, Diophantus (c. 275 BC), who has been called the 'Father of Algebra', used a form of symbolism, although the notation in use today did not appear until the seventeenth century.

In the Islamic world of the Middle Ages, equation solving was known as 'the science of restoration and balancing'. The Arabic word for restoration, *al-jabru,* is the root of 'algebra'. Even up to the sixteenth century, algebra was the name given to the art of bonesetting.

13.2 Finding missing numbers

‘$\square + 7 = 16$. What number makes this equation true?’

‘I think of a number and add 7 to it. The result is 16. Find the number.’

‘$x + 7 = 16$. What number does x stand for?’

These are three different ways of asking the same question. You can spot by mental arithmetic that the unknown number in each case is 9.

EXERCISE 13.1

1 Find the numbers which make these equations true.
(a) $\square + 8 = 13$ **(b)** $\square - 5 = 6$ **(c)** $\square \times 4 = 36$
(d) $\square \div 3 = 4$ **(e)** $10 - \square = 3$ **(f)** $(2 \times \square) + 3 = 11$
(g) $3 \times (\square + 1) = 9$ **(h)** $(\square - 3) \times 5 = 5$ **(i)** $(8 + \square) \div 5 = 3$

2 Find the number in each of the following.
(a) I think of a number and add 9 to it. The result is 22.
(b) I think of a number and subtract 5 from it. The result is 12.
(c) I think of a number and multiply it by 6. The result is 42.
(d) I think of a number and divide it by 3. The result is 9.
(e) I think of a number, double it and add 3. The result is 15.

3 Find the number which x stands for in each of these equations.
(a) $x + 3 = 7$ **(b)** $x - 2 = 4$ **(c)** $5x = 20$ **(d)** $\frac{1}{4}x = 3$

13.3 Solving linear equations

In this context, **linear** means that the equation does not involve higher powers of x, such as x^2 or x^3.

If you had to solve the equation $x + 4 = 9$ – in other words, find the number which x stands for – you would probably just say ‘$x = 5$, because $5 + 4 = 9$’. With more complicated equations, however, you will not be able to write answers straight down; more systematic methods are needed.

In this case, you can subtract 4 from both sides of the equation $x + 4 = 9$. This gives $x = 9 - 4$, which leads to the correct answer.

You may find it helpful to think of an equation as a balance, with $x + 4$ in one scale pan and 9 in the other. If you subtract 4 from both sides of the equation $x + 4 = 9$ you get $x = 9 - 4$, giving $x = 5$. The idea of a balance is key, and may be helpful in deciding how to tackle a problem.

Example 13.3.1

Solve the equation $x - 5 = 3$.

Add 5 to each side: $x = 5 + 3$

$x = 8$

Insight

You should always check your solution by substituting it back into the original equation. In this case, $8 - 5$ does equal 3.

The solution of the equation $3x = 15$ is $x = 5$. To obtain this, you could divide both sides of the equation by 3. This gives $x = \frac{15}{3}$ which is equal to 5. In terms of the balance, if two masses, $3x$ and 15, are equal, then $\frac{1}{3}$ of each of them will also be equal. Checking the solution, $3 \times 5 = 15$.

Example 13.3.2

Solve the equation $4x = 11$ and check your answer.

Divide both sides by 4: $x = \frac{11}{4}$

$x = 2\frac{3}{4}$

Check: $4 \times 2\frac{3}{4} = 11$

Multiplication and division are the inverse operations of each other.

Example 13.3.3

Solve the equation $\frac{1}{6}x = 4$ and check your answer.

Multiply both sides by 6: $x = 4 \times 6$

$x = 24$

Check: $\frac{1}{6} \times 24 = 4$

In general, you must do the same to both sides of an equation so that the balance of the equation is not destroyed.

EXERCISE 13.2

1 Solve these equations and check your answers.

(a) $x + 4 = 7$	**(b)** $x - 3 = 5$	**(c)** $4x = 24$	**(d)** $\frac{1}{5}x = 2$
(e) $x + 6 = 1$	**(f)** $5x = 3$	**(g)** $x + 4 = 4$	**(h)** $3x = 7$
(i) $7x = 0$	**(j)** $3 + x = -2$	**(k)** $x - 3 = -1$	**(l)** $\frac{1}{6}x = -1$

With a more complicated equation, you should aim to rearrange it so that the x term is on its own on one side of the equation and there is a number on the other side. For example, to solve $4x + 9 = 1$, *subtract* 9 from each side, so $4x = 1 - 9$ or $4x = -8$. Then *divide* both sides by 4 to obtain the solution $x = -\frac{8}{4}$, or $x = -2$.

In the equation $4x + 9 = 1$, the number, 4, multiplying x is called the **coefficient** of x.

Insight

You can think of solving an equation as a process of 'undoing'. So to 'undo' adding 9 you need to subtract 9, and to 'undo' multiplying by 4 you need to divide by 4.

Example 13.3.4

Solve the equation $6x - 1 = 3$ and check your answer.

Add 1 to each side: $6x = 3 + 1$

$6x = 4$

Divide both sides by 6: $x = \frac{4}{6} = \frac{2}{3}$

Check: $(6 \times \frac{2}{3}) - 1 = 4 - 1 = 3$

Sometimes an equation may have terms involving x on both sides. To solve such an equation, start by collecting all the x terms on one side and then carry on in the usual way.

Example 13.3.5

Solve the equation $5x + 1 = 2x - 2$ and check your answer.

$5x + 1 = 2x - 2$

Subtract $2x$ from both sides: $5x - 2x + 1 = -2$

$3x + 1 = -2$

Subtract 1 from both sides: $3x = -2 - 1$

$3x = -3$

Divide both sides by 3: $x = -1$

Check: LHS (left-hand side) $= (5 \times (-1)) + 1 = -5 + 1 = -4$

RHS (right-hand side) $= (2 \times (-1)) - 2 = -2 - 2 = -4$

Note that you could have started by subtracting $5x$ from both sides of the equation. In that case you would have reached the equation $-3x = 3$. Dividing both sides by -3 gives $x = -1$, as before.

To solve an equation with a negative x term, make it zero by *adding* an appropriate x term to both sides. To solve $10 - 3x = 5$, for example, start by adding $3x$ to both sides.

Example 13.3.6

Solve $10 - 3x = 5$.

Add $3x$ to both sides: $10 = 3x + 5$

Subtract 5 from both sides: $5 = 3x$

Divide both sides by 3: $x = \frac{5}{3} = 1\frac{2}{3}$

Check: $10 - (3 \times 1\frac{2}{3}) = 10 - 5 = 5$

Alternatively, you could start by subtracting 10 from both sides giving $-3x = -5$ and then divide both sides by -3. This is a correct method for solving equations but it has the disadvantage that arithmetical errors are more likely with negative numbers.

For an equation with x terms on both sides but just one negative coefficient, you can use the same approach as in Example 13.3.6.

Example 13.3.7

Solve $1 - 2x = 3x + 8$ and check your answer.

Add $2x$ to both sides: $1 = 5x + 8$
Subtract 8 from both sides: $-7 = 5x$
Divide both sides by 5: $x = -\frac{7}{5} = -1\frac{2}{5}$

Check: LHS $= 1 - \left(2 \times \left(-1\frac{2}{5}\right)\right) = 1 + 2\frac{4}{5} = 3\frac{4}{5}$

RHS $= \left(3 \times \left(-1\frac{2}{5}\right)\right) + 8 = -4\frac{1}{5} + 8 = 3\frac{4}{5}$

To solve equations with more than one negative coefficient, add to each side an x term which is big enough to give you an equation with no negative coefficients. For example, to solve $8 - 3x = 2 - 5x$, you start by adding $5x$ to both sides. If you added only $3x$, the equation you obtain still has a negative coefficient.

Example 13.3.8

Solve $8 - 3x = 2 - 5x$.

Add $5x$ to both sides: $8 + 2x = 2$
Subtract 8 from both sides: $2x = -6$
Divide both sides by 2: $x = -3$
Check: LHS $= 8 - (3 \times (-3)) = 8 - (-9) = 8 + 9 = 17$
RHS $= 2 - (5 \times (-3)) = 2 - (-15) = 2 + 15 = 17$

EXERCISE 13.3

1 Solve these equations and check your answers.

(a) $2x + 1 = 9$ **(b)** $3x - 2 = 7$ **(c)** $4x + 3 = 5$
(d) $5x + 8 = 3$ **(e)** $6x - 5 = 9$ **(f)** $8x + 9 = 3$
(g) $5 + 3x = 17$ **(h)** $7x + 9 = 9$ **(i)** $2x + 3 = 8$
(j) $6x - 1 = 4$ **(k)** $4x - 1 = -9$ **(l)** $3x + 7 = -1$

(Contd)

2 Solve these equations and check your answers.

(a) $3x + 1 = 2x + 5$	**(b)** $5x - 3 = 3x + 7$
(c) $6x - 2 = 4x - 1$	**(d)** $7x + 3 = 3x - 5$
(e) $5x - 2 = 8x - 7$	**(f)** $2x + 1 = 7x + 3$
(g) $9x + 4 = 5x - 6$	**(h)** $8x + 1 = 2x + 5$
(i) $7x + 1 = 5x + 1$	**(j)** $5x + 6 = 2x - 3$
(k) $7x + 1 = x - 3$	**(l)** $10x - 9 = 2x + 13$

3 Solve these equations and check your answers.

(a) $8 - x = 3$	**(b)** $7 - x = 12$
(c) $8 - 3x = 2$	**(d)** $9 - 5x = 6$
(e) $1 - 4x = 9$	**(f)** $8 - 3x = 1$
(g) $3 - 2x = 4$	**(h)** $3 - 4x = 9$
(i) $x + 4 = 7 - 2x$	**(j)** $2x + 3 = 7 - 3x$
(k) $3 - 4x = 5x + 12$	**(l)** $4 - 7x = 3x + 4$
(m) $8 - 3x = 2x - 5$	**(n)** $7x + 6 = 2 - 3x$
(o) $4 - x = 7 - 2x$	**(p)** $9 - 2x = 1 - 4x$

13.4 Equations with brackets

To solve an equation which has brackets, the usual first step is to expand the brackets. Sometimes, this step alone transforms the equation into one of the types you have already met.

Example 13.4.1

Solve $4(2x - 3) = 20$.

$$4(2x - 3) = 20$$

Expand the brackets: $8x - 12 = 20$

Add 12 to both sides: $8x = 32$

Divide both sides by 8: $x = 4$

Check: $4(2 \times 4 - 3) = 4(8 - 3) = 4 \times 5 = 20$

If necessary, after expanding the brackets, simplify both sides of the equation by collecting like terms.

Example 13.4.2

Solve $7 + 2(3x - 1) = 8x - 3(2x + 5)$.

$$7 + 2(3x - 1) = 8x - 3(2x + 5)$$

Expand the brackets: $7 + 6x - 2 = 8x - 6x - 15$

Collect like terms: $5 + 6x = 2x - 15$

Subtract $2x$ from both sides: $5 + 4x = -15$

Subtract 5 from both sides: $4x = -20$

Divide both sides by 4: $x = -5$

Check: LHS $= 7 + 2(3 \times (-5) - 1) = 7 + 2 \times (-16) = -25$

RHS $= 8 \times (-5) - 3(2 \times (-5) + 5) = -40 - 3 \times (-5) = -25$

EXERCISE 13.4

1 Solve these equations and check your answers.

(a) $5(x + 4) = 35$
(b) $3(4x - 5) = 9$
(c) $4(x + 6) = 12$
(d) $3(2x - 1) = 1$
(e) $6(x - 2) = 5x$
(f) $2(5x - 3) = 6x$
(g) $4(x + 3) = x + 16$
(h) $5(3x + 2) = 7x + 4$
(i) $2(x + 5) = 4 - x$
(j) $5(2x - 3) = 9 - 2x$
(k) $3(2 - x) = 14$
(l) $4(3 - 2x) = 5$
(m) $6(2x - 3) = 4(x - 2)$
(n) $4(2x - 7) = 3(4x - 9)$
(o) $3(2x + 3) = 4(6 - x)$
(p) $3(2x + 5) = 2(4x - 5) + 3x$
(q) $2(3x + 5) - 3(x - 4) = 1$
(r) $9x - 4(2x - 3) = 5(x + 2)$
(s) $8x - (3x - 7) = 2(x + 3) + 1$
(t) $5(3x + 4) = 9 + (6x - 7)$

13.5 Solving problems using equations

You can sometimes solve problems by using a letter to represent the unknown number, expressing the given information as an equation and then solving it.

Example 13.5.1

I think of a number, multiply it by 5 and subtract 4. I get the same answer if I add 2 to the number and multiply the result by 3. Find the number.

Let x be the number.
The equation is: $5x - 4 = 3(x + 2)$
Expand the brackets: $5x - 4 = 3x + 6$
Subtract $3x$ from both sides: $2x - 4 = 6$
Add 4 to both sides: $2x = 10$
Divide both sides by 2: $x = 5$
The number is 5.
Check in the original problem.
$5 \times 5 = 25$ and $25 - 4 = 21$; $5 + 2 = 7$ and $3 \times 7 = 21$.

Example 13.5.2

Trevor is 8 years younger than his sister, Gwen. The sum of their ages is 110 years. Find Trevor's age.

Let Trevor's age be x years. Then Gwen's age is $x + 8$ years.
The equation is: $x + (x + 8) = 110$
Remove brackets: $x + x + 8 = 110$
Collect like terms: $2x + 8 = 110$
Subtract 8 from both sides: $2x = 102$
Divide both sides by 2: $x = 51$
Trevor's age is 51 years.

Check: If Trevor is 51, Gwen is $51 + 8 = 59$. The sum of their ages is $51 + 59 = 110$.

EXERCISE 13.5

1 I think of a number, multiply it by 4 and add 7. The answer is 55. Find the number.
2 I think of a number, add 7 to it and multiply the result by 6. The answer is 90. Find the number.

3 I think of a number, multiply it by 3 and add 7. I get the same answer if I multiply the number by 4 and subtract 12. Find it.
4 I think of a number, multiply it by 7 and subtract 12. I get the same answer if I add 2 to the number and multiply the result by 5. Find the number.
5 John is 7 years older than his wife, Anne. The sum of their ages is 63 years. Find John's age.
6 Rachel is four times as old as her daughter. The sum of their ages is 40 years. Find their ages.
7 The length of a rectangle is 3 cm greater than its width. The perimeter of the rectangle is 34 cm. Find its length.
8 One of the angles of an isosceles triangle is 114°. Find the size of the other two angles.
9 The sum of three consecutive numbers is 27. Find the numbers.
10 In a soccer league, teams get 3 points for a win, 1 point for a draw and 0 points for a defeat. A team with 18 points has won two more games than it has drawn. How many games has it won?

13.6 Solving inequalities

To solve inequalities, you may use almost all the techniques you have used to solve equations. You may add the same amount to both sides, subtract the same amount from both sides and multiply or divide both sides by the same *positive* number. You can verify all of these operations by thinking of a balance where the scales do not balance.

You may *not*, however, multiply or divide both sides by the same *negative* number.

To show this, start with a true inequality such as $3 > -6$. If you multiply both sides by a negative number, say -2, you get $-6 > 12$, which is *false*. Similarly, if you divide both sides of $3 > -6$ by the same negative number, say -3, you get $-1 > 2$, which is also false.

To solve the inequality $-5x < -10$, first rewrite it as $10 < 5x$ by adding $10 + 5x$ to both sides. Now divide by 5, getting $x > 2$. You can see that the left and right sides of the original inequality $-5x < -10$ have been divided by -5, but the sign of the inequality has been reversed.

Insight

Another way around the problem of multiplying and dividing by a negative is to remember to reverse the inequality sign. So to solve $-3x \leq 12$ you can divide by -3 and reverse the inequality sign to get $x \geq -4$.

Example 13.6.1

Solve $3x - 4 > 14$.

Add 4 to both sides: $3x > 18$

Divide both sides by 3: $x > 6$

Insight

You can check you are correct by substituting some values into the original inequality. For the above example, when $x = 7$, $3 \times 7 - 4 = 17 > 14$; when $x = 5$, $3 \times 5 - 4 = 11 < 14$ and when $x = 6$, $3 \times 6 - 4 = 14$. So you can see that $x > 6$ is correct.

Example 13.6.2

Solve $3(2x + 7) \leq 9 - 2x$.

Expand the brackets: $6x + 21 \leq 9 - 2x$

Add $2x$ to both sides: $8x + 21 \leq 9$

Subtract 21 from both sides: $8x \leq -12$

Divide both sides by 8: $x \leq -1\frac{1}{2}$

Example 13.6.3

Solve $8 - 3x \leq 7 - x$.

Add $3x$ to both sides: $8 \leq 7 + 2x$

Subtract 7 from both sides: $1 \leq 2x$

Divide both sides by 2: $\frac{1}{2} \leq x$

Rewrite with x on the left: $x \geq \frac{1}{2}$

Notice how, when the inequality is rewritten to have x on the left, the sign of the inequality needs to be reversed.

EXERCISE 13.6

1 Solve these inequalities.

(a) $4x < 12$
(b) $x + 7 \geq 2$
(c) $5x - 3 \leq 12$
(d) $7 - 2x \geq 1$
(e) $4x + 5 \geq 1$
(f) $3(x - 2) > 15$
(g) $2(x + 4) < 6$
(h) $x + 1 > 3x$
(i) $8x + 1 \leq 4x - 3$
(j) $4(x + 2) < 5x - 1$
(k) $2(x - 2) \geq 5(x + 1)$
(l) $5x + 1 > 10 - x$
(m) $6x - 5 \leq 15 - 2x$
(n) $4(2x + 3) < 7 - 2x$
(o) $1 - 8x \geq 11 - 4x$

THINGS TO REMEMBER

- *An equation is a mathematical statement that two expressions are equal, such as $3 \times 2 = 6$.*
- *Equations may contain unknown numbers which are denoted by symbols or letters.*
- *The solution (or root) of an equation is the value of the unknown that satisfies the equation.*
- *To solve an equation you must do the same to both sides of the equation.*
- *Always check your solution by substituting it back into the original equation.*
- *When an equation contains brackets, start by expanding them and then simplify if necessary.*
- *When an equation has an unknown on both sides, start by collecting all the unknowns on one side and then carry on in the usual way.*
- *You can solve an inequality using the same techniques as for an ordinary equation. Take extra care when dealing with negatives; either rewrite the inequality (by adding the same terms to both sides) so that it no longer includes the negative, or remember to reverse the inequality sign when dividing or multiplying by a negative number.*

14

Percentages

In this chapter you will learn:

- ***how to convert between percentages and fractions and decimals***
- ***how to find percentages of quantities***
- ***how to express one quantity as a percentage of another***
- ***how to find percentage changes.***

14.1 Introduction

A percentage is a fraction expressed in hundredths. The name is derived from the Latin *per centum,* which means 'per hundred'. The symbol % is often used for 'per cent'; so, for example, 19% means $\frac{19}{100}$.

Although the Romans did not use percentages, they used fractions which converted easily to hundredths to calculate taxes. From the Middle Ages, as commerce grew, so did the use of percentages. By the beginning of the seventeenth century their use was widespread, especially for the computation of interest and profit and loss.

14.2 Percentages, decimals, fractions

You can easily write any percentage as a fraction with 100 as its denominator, for example 37% = $\frac{37}{100}$. This fraction is already in its simplest form but, in other cases, you can simplify the fraction. For example, 48% = $\frac{48}{100}$, which, in its simplest form, is $\frac{12}{25}$.

Example 14.2.1

Write each percentage as a fraction in its simplest form.

(a) 7% **(b)** 60% **(c)** 200% **(d)** 350%

(a) 7% = $\frac{7}{100}$ **(b)** 60% = $\frac{60}{100} = \frac{3}{5}$

(c) 200% = $\frac{200}{100} = 2$ **(d)** 350% = $\frac{350}{100} = 3\frac{1}{2}$

When the percentage includes a fraction, for example $17\frac{1}{2}$%, first write the percentage as a fraction with a denominator of 100. Then multiply numerator and denominator by a number which will give you a numerator which is a whole number. Finally, if necessary, give the fraction in its simplest form.

Example 14.2.2

Write $17\frac{1}{2}$% as a fraction in its simplest form.

As a fraction with a denominator of 100: $17\frac{1}{2}\% = \frac{17\frac{1}{2}}{100}$

Multiply numerator and denominator by 2: $\frac{17\frac{1}{2} \times 2}{100 \times 2} = \frac{35}{200}$

In its simplest form: $\frac{35}{200} = \frac{7}{40}$

You can write a percentage straight down as a decimal. For example, 23% = $\frac{23}{100}$ which, as a decimal, is 0.23.

Insight

You can think of $\frac{23}{100}$ as 23 ÷ 100. Take care with percentages which are less than 10% or greater than 100%. For example, $4\frac{1}{2}$% = 4.5 ÷ 100 = 0.045 and 110% = 110 ÷ 100 = 1.1.

Example 14.2.3

Write each percentage as a decimal.

(a) 47% **(b)** 3% **(c)** 60% **(d)** 213% **(e)** 120%
(f) 7.8% **(g)** $23\frac{1}{2}$%

(a) 47% = 0.47 **(b)** 3% = 0.03 **(c)** 60% = 0.60 = 0.6
(d) 213% = 2.13 **(e)** 120% = 1.20 = 1.2
(f) 7.8% = $\frac{7.8}{100}$ = 7.8 ÷ 100 = 0.078 **(g)** $23\frac{1}{2}$% = 23.5% = 0.235

Reversing this process, you can write a decimal as a percentage.

Insight

So to write a decimal as a percentage you multiply it by 100.

Example 14.2.4

Write each decimal as a percentage.

(a) 0.39 **(b)** 0.07 **(c)** 0.9 **(d)** 1.08 **(e)** 0.475 **(f)** 0.043

(a) 0.39 = 39% **(b)** 0.07 = 7% **(c)** 0.9 = 90%
(d) 1.08 = 108% **(e)** 0.475 = 47.5% **(f)** 0.043 = 4.3%

You can write a fraction with a denominator of 100 straight down as a percentage. For example, $\frac{29}{100}$ = 29%. For a fraction with a denominator which is a factor or a multiple of 100, convert it to an equivalent fraction with a denominator of 100 and then express it as a percentage.

Example 14.2.5

Express each fraction as a percentage.

(a) $\frac{7}{25}$ **(b)** $\frac{135}{500}$

(a) Multiply numerator and denominator by 4: $\frac{7}{25} = \frac{28}{100}$ = 28%

(b) Divide numerator and denominator by 5: $\frac{135}{500} = \frac{27}{100}$ = 27%

Alternatively, to convert any fraction to a percentage, think of the line in the fraction as a division sign and then, by hand or with a calculator, work out the result as a decimal, and then convert it to a percentage.

Insight

You can think of this as 'top' ÷ 'bottom' × 100.

Example 14.2.6

Express each fraction as a percentage.

(a) $\frac{3}{4}$ **(b)** $\frac{7}{8}$

(a) $\frac{3}{4} = 3 \div 4 = 0.75 = 75\%$ **(b)** $\frac{7}{8} = 7 \div 8 = 0.875 = 87.5\%$

EXERCISE 14.1

1 Write each percentage as a fraction in its simplest form.
(a) 49% **(b)** 50% **(c)** 25% **(d)** 10%
(e) 65% **(f)** 1% **(g)** 3% **(h)** 100%
(i) 250% **(j)** $12\frac{1}{2}\%$ **(k)** $33\frac{1}{3}\%$ **(l)** $13\frac{3}{4}\%$

2 Write each percentage as a decimal.
(a) 39% **(b)** 7% **(c)** 30% **(d)** 112%
(e) 3.8% **(f)** $62\frac{1}{2}\%$ **(g)** $3\frac{1}{2}\%$ **(h)** $12\frac{1}{4}\%$

3 Write each decimal as a percentage.
(a) 0.71 **(b)** 0.04 **(c)** 0.8 **(d)** 1.75
(e) 1.9 **(f)** 0.027 **(g)** 0.0325 **(h)** 0.009

4 Write each fraction as a percentage.
(a) $\frac{43}{100}$ **(b)** $\frac{9}{100}$ **(c)** $\frac{7}{10}$ **(d)** $\frac{31}{50}$
(e) $\frac{17}{25}$ **(f)** $\frac{3}{4}$ **(g)** $\frac{165}{300}$ **(h)** $\frac{3}{8}$

5 Approximately 52% of babies born are boys.
(a) Write 52%
(i) as a decimal **(ii)** as a fraction.
(b) What percentage of babies born are girls?

6 Approximately 45% of people who take their car driving test pass.
(a) What percentage fail? **(b)** What fraction fail?

14.3 Percentages of quantities

You can find a percentage of a quantity mentally when the percentage has a well-known fractional equivalent and the arithmetic is simple. For example, to find 50% of 18, use the fact that 50% = $\frac{1}{2}$ and then find $\frac{1}{2}$ of 18; in other words $\frac{1}{2} \times 18 = 9$.

Example 14.3.1

Find 75% of 24 km.

75% = $\frac{3}{4}$, so $\frac{3}{4}$ of 24 = $\frac{3}{4} \times 24 = 18$

In general, to find a percentage of a quantity, write the percentage as a fraction or as a decimal; replace 'of' by '×' and carry out the multiplication, either by hand or using a calculator.

Example 14.3.2

Find 36% of 175.

As a fraction: 36% = $\frac{36}{100}$
Replace 'of' by '×': 36% of 175 = $\frac{36}{100} \times 175 = 63$

Example 14.3.3

Find 47% of 260 litres.

As a decimal: 47% = 0.47
Replace 'of' by '×': 47% of 260 = 0.47 × 260 = 122.2
Therefore 47% of 260 litres is 122.2 litres.

When you find a percentage of an amount of money, you sometimes have to round your answer to the nearest penny.

Example 14.3.4

Find 37% of £6.53 and give your answer to the nearest penny.

$0.37 \times 6.53 = 2.4161$, so 37% of £6.53 = £2.4161
To the nearest penny, 37% of £6.53 = £2.42.

EXERCISE 14.2

1 Without using a calculator, find
- **(a)** 50% of 36
- **(b)** 25% of 48
- **(c)** 75% of 36 kg
- **(d)** 10% of 250 m
- **(e)** 20% of 45 cm
- **(f)** $33\frac{1}{3}$% of 27 litres

2 Using a calculator if necessary, find
- **(a)** 30% of 80
- **(b)** 46% of 250
- **(c)** 76% of 32
- **(d)** 8% of 475 m
- **(e)** 12% of 540
- **(f)** 120% of 85
- **(g)** 250% of 386
- **(h)** 115% of 680 kg
- **(i)** 15.6% of 800 litres
- **(j)** 9.3% of 30 km
- **(k)** $27\frac{1}{2}$% of 72 cm
- **(l)** $7\frac{1}{2}$% of 620 m

3 There are 1200 students in a school. 57% of them are girls.
- **(a)** Work out the number of girls.
- **(b)** What percentage of the students are boys?
- **(c)** Work out the number of boys.

4 A test was marked out of 80. Mary scored 85% of the marks. How many marks did she score?

5 Find 55% of 1 day and give your answer in minutes.

6 Find
- **(a)** 92% of £865
- **(b)** 78% of £426
- **(c)** 45% of £364

7 Find 19% of £9.36 and give your answer to the nearest penny.

8 David pays a deposit of 15% on a car which costs £8200.
- **(a)** How much deposit does he pay?
- **(b)** What percentage of the cost has he still to pay?

14.4 Increasing and decreasing quantities

Sometimes you have to find a percentage of a quantity and then add it to or subtract it from the original amount. For example, if you put £600 in a bank account and receive 5% interest after one year, the amount in your account after one year is not 5% of £600 (£30). You must add the £30 to the original £600. There is £630 in your account.

Example 14.4.1

(a) Increase 350 by 6%. **(b)** Decrease 480 by 5%.

(a) 6% of 350 = 0.06 × 350 = 21. The new amount is 350 + 21 = 371.
(b) 5% of 480 = 0.05 × 480 = 24.The new amount is 480 – 24 = 456.

EXERCISE 14.3

1 Increase
(a) 65 by 20%, **(b)** 400 by 9%, **(c)** 840 by 4%,
(d) 5600 by 14%, **(e)** 70 by 12%, **(f)** 750 by 3%.

2 Decrease
(a) 98 by 50%, **(b)** 430 by 10%, **(c)** 240 by 5%,
(d) 3200 by 15%, **(e)** 820 by 24%, **(f)** 45 by 18%.

3 The number of people employed by a company increased by 16% from 650. How many people are employed by the company now?

4 The normal price of a computer is £550. In a sale, normal prices are reduced by 15%. Find the sale price of the computer.

5 Julie’s salary was £2350 per month before she received a pay rise of 4%. Find her new monthly salary.

6 A year ago, Richard bought a new car for £12 680. Since then, it has lost 12% of its value. Find his car’s current value.

7 Chris invests £630 in a bank account at an interest rate of 4.5% per annum. (Per annum means per year.) After one year, the interest is added to her account. How much is in her account after one year?

8 Tony buys a camera for £130 and later sells it, making a loss of 12.5%. Find the selling price.

14.5 One quantity as a percentage of another quantity

To express one quantity as a percentage of another quantity, first express it as a fraction of the other quantity. Then convert the fraction to a decimal and finally to a percentage.

Example 14.5.1

Elaine scores 52 out of 80 in a test. Express her mark as a percentage.

As a fraction, 52 out of 80 is $\frac{52}{80}$.
Find 52 ÷ 80 to convert $\frac{52}{80}$ to a decimal, $\frac{52}{80} = 0.65$.
As a percentage 0.65 = 65%.

Insight

So to write 52 out of 80 as a percentage you need to work out (52 ÷ 80) × 100. You can think of 'out of' as meaning 'divide by'.

If both quantities are expressed in the same units, the method is the same as if both quantities were just numbers.

If the quantities are expressed in different units, however, you must express them in the same units, usually the smaller of the two.

Example 14.5.2

(a) Express 84 m as a percentage of 230 m. Give your answer correct to one decimal place.
(b) Express 39 mm as a percentage of 15 cm.

(a) As a fraction, you require $\frac{84}{230}$.
As a decimal, this is $84 \div 230 = 0.3652\ldots$
As a percentage, this is 36.5% (correct to 1 d.p.).
(b) In the same units, the quantities are 39 mm and 150 mm.
As a fraction, you require $\frac{39}{150}$.
As a decimal, this is $39 \div 150 = 0.26$.
As a percentage, this is 26%.

EXERCISE 14.4

1 Tony scores 63 out of 75 in a test. Express this as a percentage.
2 In a class of 35 students, 21 are girls. What percentage of the students are girls and what percentage are boys?
3 A team played 38 matches and won 21. What percentage of its matches did it win? Give your answer correct to one decimal place.
4 The area of the surface of the Earth is 510 066 000 km^2, of which 361 743 000 km^2 is water. What percentage of the Earth's surface is water? Give your answer correct to the nearest whole number.
5 Express the first quantity as a percentage of the second. Where necessary, give your answer correct to one decimal place.
(a) 9 cm, 25 cm **(b)** 24 m, 60 m **(c)** 18 hrs, 24 hrs
(d) 29 kg, 82 kg **(e)** 18 mm, 4 cm **(f)** 85 cm, 5 m
(g) 14 hrs, 3 days **(h)** 350 m, 3 km **(i)** 83p, £7

14.6 Percentage increases

When a quantity changes, you can find the percentage increase in a quantity by expressing the change as a percentage of the *original* quantity. The same applies to decreases.

Example 14.6.1

The population of a village increases from 675 to 783. Find the percentage increase.

The increase is 783 – 675 = 108, on an original total of 675.

As a percentage, this is $\frac{\text{increase}}{\text{original}} = \frac{108}{675} = 0.16 = 16\%$.

The percentage increase is 16%.

Percentage profit and **percentage loss** are important types of percentage change. For example, if you buy a car for £8000 (the cost price) and sell it for £6800 (the selling price), you make a loss of £(8000 – 6800) = £1200. To find the percentage loss, express the loss as a percentage of the cost price, that is, $\frac{1200}{8000} = 0.15 = 15\%$.

Example 14.6.2

Joan bought a house for £240 000 and later sold it for £288 000. Find her percentage profit.

Profit = £(288 000 – 240 000) = £48 000

As a percentage, this is $\frac{\text{profit}}{\text{cost price}} = \frac{48\,000}{240\,000} = 0.2 = 20\%$

The percentage profit is 20%.

You can just remember this general formula:

$$\text{Percentage change} = \frac{\text{change}}{\text{original}} \times 100$$

EXERCISE 14.5

EXERCISE

1. The number of workers employed by a company increased from 225 to 243. Find the percentage increase.
2. The population of a town fell from 23 650 to 20 812. Find the percentage decrease.
3. Helen's annual salary was increased from £22 500 to £23 400. Find the percentage increase.
4. The original price of a bicycle was £184. In a sale, its price was £161. Find the percentage reduction.
5. Bill put £5500 in a bank account. A year later, interest was added and the amount in his account increased to £5885. Find the rate of interest.
6. The cost price of a pair of shoes was £24 and the selling price was £42. Find the percentage profit.

7 Kim bought a computer for £750 and later sold it for £495. Find her percentage loss.

8 An art dealer buys a painting for £3500 and sells it for £4480. Find his percentage profit.

14.7 Using multipliers

If normal prices are reduced by 15% in a sale, the sale price of an item is 85%, or (100% – 15%) of its normal price. Thus, to find the sale price of a television, the normal price of which is £360, you find 85% of £360. As 0.85 × 360 = 306, the sale price is £306.

Similarly, if the number of employees in a company increases by 12%, the new number is 112% = 100% + 12% of the original number. So, if the number of employees was originally 350, the new number is 112% of 350. As 112% of 350 = 1.12 × 350 = 392, the new number of employees is 392.

The numbers 0.85 and 1.12 are called **multipliers**.

Example 14.7.1

Write down the multipliers needed for these changes.

(a) a decrease of 23% **(b)** an increase of 20%

(c) a decrease of 9%

(a) 100% – 23% = 77%, so the multiplier is 0.77.

(b) 100% + 20% = 120%, so the multiplier is 1.20 or 1.2.

(c) 100% – 9% = 91%, so the multiplier is 0.91.

Multipliers are especially useful for calculating **compound interest**, the type of interest paid by banks and building societies. Compound interest is paid on the *total* of the amount invested (called the **principal**) and the interest earned in previous years.

For example, if you invest £5000 at 6% p.a. compound interest, the amount in your account after one year will be £5000 × 1.06 = £5300. After two years, the new amount increases to £5000 × 1.06 × 1.06 = £5000 × 1.06^2 = £5618. After three years, the amount will be £5000 × 1.06^3 = £5955.08, and so on.

Example 14.7.2

Jan invests £10 000 at 4.5% p.a. compound interest. Work out
(a) the amount in her account after 3 years,
(b) the amount of interest earned.

(a) 10 000 × 1.045^3 = 11 411.661, so the amount in her account is £11 411.66, to the nearest penny.
(b) Amount of interest earned = £11 411.66 – £10 000 = £1411.66

You can also use multipliers to find the original quantity *before* a percentage change, if you know the percentage change and the new quantity *after* the change.

For example, if normal prices are reduced by 12% in a sale, you multiply the normal price of an item by 0.88 to find its sale price.

Thus, normal price × 0.88 = sale price.

Working back, if you know the sale price, divide it by 0.88 (the inverse operation to multiplying by 0.88) to find the normal price. If the sale price is £396, find 396 ÷ 0.88 = 450, giving a normal price of £450.

Example 14.7.3

The cost of an airline ticket is increased by 14%. After the increase, the cost of the ticket is £399. Find its cost before the increase.

As cost before increase × 1.14 = cost after increase
so cost before increase = cost after increase ÷ 1.14

As 399 ÷ 1.14 = 350, the ticket cost £350 before the increase.

EXERCISE 14.6

1 Write down the multiplier needed for each of these changes.
  (a) a decrease of 31% **(b)** an increase of 43%
  (c) a decrease of 30% **(d)** an increase of 7%
  (e) an increase of 17.5% **(f)** a decrease of 3.4%

2 Normal prices are reduced by 35% in a sale. The normal price of a video recorder is £260. Find its sale price.

3 The population of the United Kingdom was 50.3 million in 1951. By 1991, it had increased by 15%. Find the population in 1991. Give your answer correct to the nearest 0.1 million.

4 Joe invests £8000 at 7% p.a. compound interest. Work out the amount in his account after 2 years.

5 Jean invests £500 at 3.5% p.a. compound interest. Work out
  (a) the amount in her account after 3 years,
  (b) the amount of interest earned.

6 Janet invests a sum of money at an interest rate of 8% p.a. After one year, she has £810 in her account. How much did she invest?

7 In a sale, normal prices are reduced by 20%. The sale price of a microwave oven is £140. Find its normal price.

8 The price of a new car is increased by 18% to £14 750. Find its cost before the increase.

THINGS TO REMEMBER

- *The symbol % is used for 'per cent'. Per cent means per hundred.*

- *To write a percentage as a fraction, write the percentage as a fraction with 100 as its denominator. For example,* $61\% = \frac{61}{100}$.

- *To write a percentage as a decimal, divide the percentage by 100. For example,* $61\% = \frac{61}{100} = 0.61$ *or* $61\% = 61 \div 100 = 0.61$.

- *To write a decimal as a percentage, multiply it by 100.*

- *To write a fraction as a percentage, either:*
  - *write it as a fraction with a denominator of 100. For example,* $\frac{3}{20} = \frac{15}{100} = 15\%$.

Or:

  - *write the fraction as a decimal by dividing the numerator (top) by the denominator (bottom) and then multiply by 100 to get a percentage.*

- *To find a percentage of a quantity, write the percentage as a decimal; replace 'of' by '×' and carry out the multiplication, either by hand or using a calculator.*

- *To increase or decrease an amount by a percentage, work out the percentage and then add it to or subtract it from the original amount.*

$$\text{Percentage change} = \frac{\text{change}}{\text{original}} \times 100$$

- *You can use multipliers to work out a percentage increase or decrease directly. For example, to increase an amount by 15% you use the multiplier 1.15 and to decrease an amount by 15% you use the multiplier 0.85.*

15

Formulae

In this chapter you will learn:

- ***how to write formulae***
- ***how to evaluate and use formulae***
- ***how to change the subject of a formula.***

15.1 What is a formula?

In mathematics and science, a **formula** is a quick way of expressing a rule or a fact. You have already met some formulae (the plural of formula) expressed in words: area of parallelogram = base × height, for example, in Chapter 10. If you use A to stand for the area of a parallelogram, b to stand for its base and h to stand for its height, you can write this result more concisely as the formula $A = bh$.

A is called the **subject** of this formula, because it appears on its own on the left-hand side and does not appear on the right-hand side.

A multiplication sign is not necessary in the formula $A = bh$. This is one of the many ideas from Chapter 11 you will need in this chapter. You will, for example, evaluate formulae in exactly the same ways as you evaluated algebraic expressions. A formula is not, however, the same as an expression; for example, a formula has an equals sign in it, whereas an expression does not.

Example 15.1.1

Each side of a square is l centimetres long. The perimeter of the square is P centimetres.

(a) Write down a formula for P in terms of l. Give your answer as simply as possible.
(b) Evaluate P when $l = 3.7$.

(a) Perimeter of a square = 4 × length of each side, so $P = 4l$
(b) When $l = 3.7$, $P = 4 \times 3.7 = 14.8$

Insight

The letters in a formula represent numbers and not words or quantities. The letters in a formula are called **variables** because they represent numbers which can vary.

EXERCISE 15.1

1 Area of a triangle = $\frac{1}{2}$ × base × height.
- **(a)** Write down a formula for the area, A, of a triangle in terms of its base, b, and its height, h.
- **(b)** Evaluate A when $b = 7$ and $h = 8$.

2 Average speed = distance travelled ÷ time taken.
- **(a)** Write down a formula for the average speed, s miles per hour, of a motorist who travels d miles in t hours.
- **(b)** Evaluate s when $d = 126$ and $t = 2.25$.

3 To find the angle sum, in degrees, of a polygon, subtract 2 from the number of sides it has and then multiply the result by 180.
- **(a)** Find the angle sum, S, of a polygon with n sides.
- **(b)** Evaluate S when $n = 9$.

4 Each side of a regular octagon is d centimetres long. The perimeter of the octagon is P centimetres.
- **(a)** Write down a formula for P in terms of d.
- **(b)** Evaluate P when $d = 3.7$.

5 Each side of a square is l centimetres long. The area of the square is A square centimetres.

(a) Write down a formula for A in terms of l.

(b) Evaluate A when $l = 8.6$.

6 a, b and c are three numbers. The mean of the numbers is m.

(a) Write down a formula for m in terms of a, b and c.

(b) Evaluate m when $a = 27$, $b = 41$ and $c = 37$.

7 The lengths of the parallel sides of a trapezium are a centimetres and b centimetres. The parallel sides are h centimetres apart. The area of the trapezium is A square centimetres.

(a) Write down a formula for A in terms of a, b and h.

(b) Evaluate A when $a = 12.6$, $b = 7.4$ and $h = 5.8$.

8 A regular polygon has n sides. The size of each of its exterior angles is $x°$.

(a) Write down a formula for x in terms of n.

(b) Evaluate x when $n = 9$.

9 $T = \frac{1}{2}n(n + 1)$. Evaluate T when $n = 10$. (This value is the tenth triangle number.)

10 $v = u + at$. Evaluate v when $u = 20$, $a = -3$ and $t = 4$.

11 $s = \frac{1}{2}gt^2$ Evaluate s when $g = -10$ and $t = 6$.

12 $I = \frac{1}{100} PRT$. Evaluate I when $P = 1200$, $R = 3\frac{1}{2}$ and $T = 8$.

15.2 Evaluating terms other than the subject

In each question of Exercise 15.1, the term you had to evaluate was the subject of a formula. Sometimes you will have to find the value of a term which is not the subject. To do this, you substitute into the formula the values you are given and then solve the resulting equation.

If, for example, in the formula $A = bh$, you are told that $A = 54$ and $b = 6$, substituting these values gives the equation $6h = 54$, which has the solution $h = 9$.

Example 15.2.1

$v = u + at$. Given that $u = 6$, $v = 38$ and $a = 4$, find the value of t.

Substituting the values of u, v and a gives $38 = 6 + 4t$.
Subtracting 6 from both sides gives $4t = 32$.
Dividing both sides by 4 gives $t = 8$.

Example 15.2.2

In the formula $a = kv^2$, $a = 45$ and $k = 5$. Find the value of v.

Substituting the values of a and k gives $45 = 5v^2$.
Dividing both sides by 5 gives $v^2 = 9$.
Therefore $v = 3$ or $v = -3$.

The number 3 is the positive **square root** of 9, written $\sqrt{9}$, because $3^2 = 9$, and is positive.

In general, x is the **square root** of a, written $\sqrt{a}$, if $x^2 = a$ and $x \geq 0$.

Insight

All numbers have two square roots: one positive and one negative. So –3 is also the square root of 9 as $(-3)^2 = 9$. However, strictly speaking, the symbol $\sqrt{}$ means the positive square root only. If you need the positive square root of 9 you should write $\sqrt{9} = 3$ and if you need both square roots you should write $\pm\sqrt{9} = \pm 3$ (read as 'plus or minus the square root of 9').

EXERCISE 15.2

EXERCISE

1 $P = 3d$. Find the value of d when $P = 42$.
2 $A = lb$. Find the value of b when $A = 108$ and $l = 9$.
3 $y = 4x - 3$. Find the value of x when $y = 7$.
4 $P = 2x + y$. Find the value of y when $P = 23$ and $x = 8$.

5 $P = 2l + 2b$. Find the value of b when $P = 28$ and $l = 8$.
6 $I = \frac{1}{100}PRT$. Find the value of R when $I = 360$, $P = 1500$ and $T = 6$.
7 $s = \frac{d}{t}$. Find the value of d when $s = 84$ and $t = 1.5$.
8 $v = u + at$. Find the value of a when $v = 13$, $u = 25$ and $t = 4$.
9 $P = 2(l + b)$. Find the value of l when $P = 16$ and $b = 3$.
10 $s = ut + \frac{1}{2}at^2$. Find the value of u when $s = 52$, $t = 4$ and $a = 5$.
11 $a = rw^2$. Find the value of w when $a = 75$ and $r = 3$.
12 $s = \frac{1}{2}gt^2$. Find the value of t when $s = 80$ and $g = 10$.
13 $P = \frac{V^2}{R}$. Find the value of V when $P = 20$ and $R = 5$.
14 $F = \frac{mv^2}{r}$. Find the value of v when $F = 24$, $m = 8$ and $r = 12$.

15.3 Changing the subject of a formula

The techniques needed to change the subject of a formula are the same as those you used to solve equations in Chapter 13. Formulae are really equations. You must do the same to both sides of a formula, just as you did the same to both sides of an equation.

For example, to make l the subject of the formula $A = lb$, divide both sides by b, obtaining $\frac{A}{b} = l$, which is $l = \frac{A}{b}$. Similarly, to make b the subject, divide both sides by l, obtaining $b = \frac{A}{l}$.

Example 15.3.1

Make P the subject of the formula $\frac{PV}{T} = k$.

Multiply both sides by T: $PV = kT$
Divide both sides by V: $P = \frac{kT}{V}$

Example 15.3.2

Make h the subject of the formula $A = \frac{1}{2}bh$.

Multiply both sides by 2: $2A = bh$
Divide both sides by b: $h = \frac{2A}{b}$

Example 15.3.3

Make t the subject of the formula $v = u + at$.
Subtract u from both sides: $v - u = at$
Divide both sides by a: $t = \frac{V - u}{a}$

Example 15.3.4

Make u the subject of the formula $I = m(v - u)$.

Expand the brackets: $I = mv - mu$
Add mu to both sides: $mu + I = mv$
Subtract I from both sides: $mu = mv - I$
Divide both sides by m: $u = \frac{mv - 1}{m}$

Insight

A common mistake is to only divide part of the right-hand side by m. However, for the formula to balance, both terms must be divided by m.

EXERCISE 15.3

1 In each part, make the letter in brackets the subject of the formula.

(a) $V = IR$ (R)
(b) $V = bh$ (b)
(c) $E = mgh$ (h)
(d) $P = 2x + y$ (y)
(e) $S = \frac{D}{T}$ (T)
(f) $V = \frac{1}{3}Ah$ (h)
(g) $I = \frac{PRT}{100}$ (R)
(h) $\frac{PV}{T} = k$ (V)
(i) $p = a + b + c$ (a)
(j) $P = 2x + y$ (x)
(k) $v = u - gt$ (u)
(l) $v = u + at$ (a)
(m) $v = u - gt$ (t)
(n) $P = 2(l + b)$ (l)
(o) $I = m(v - u)$ (v)
(p) $S = 180n - 360$ (n)
(q) $2S = a + b + c$ (a)
(r) $A = \frac{h(a+b)}{2}$ (a)

THINGS TO REMEMBER

- *A formula is a rule expressed in symbols for working something out.*
- *The plural of formula is formulae.*
- *You don't write × symbols in a formula. So* $A = lw$ *means* $A = l \times w$.
- *The subject of a formula appears on its own on the left-hand side and does not appear on the right-hand side. For example, in the formula* $v = u + at$, v *is the subject.*
- *You can evaluate the subject of a formula by substituting the values of the other variables (letters or symbols) into the formula.*
- *You can evaluate another term which is not the subject by substituting into the formula the values you are given and then solving the resulting equation.*
- $\sqrt{16}$ *means the positive square root of 16; so* $\sqrt{16} = 4$. *16 has two square roots 4 and –4. The context of the problem determines whether or not the negative square root is needed.*
- *You can change the subject of a formula by rearranging it using the same techniques as when solving an equation. You must make sure that you do the same thing to both sides of the formula so that it balances.*

16

Circles

In this chapter you will learn:

- ***the names of parts of a circle***
- ***how to find the circumference of a circle***
- ***how to find the area of a circle***
- ***how to use two circle properties.***

16.1 Introduction

A **circle** is a plane curve, every point of which is the same distance from a fixed point, called the **centre** of the circle. This distance from the centre to a point on the circle is the **radius.** (See Figure 16.1.)

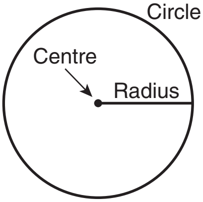

Figure 16.1

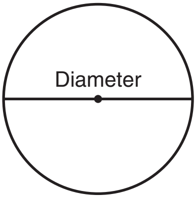

Figure 16.2

A straight line which joins two points on the circle and passes through the centre is a **diameter**. (See Figure 16.2.) So a diameter divides a circle exactly in half, or, the diameter **bisects** the circle.

This was one of the fundamental propositions of Thales (c. 600 BC), on which classical geometry was based.

A diameter divides a circle into two **semicircles** and is a **line of symmetry**. As you can draw an infinite number of diameters on a circle, you can also draw an infinite number of lines of symmetry. Similarly, the **order of rotational symmetry** of a circle is also infinite, as a tracing of a circle can be rotated to fit on top of itself in an infinite number of different positions.

A diameter is twice as long as a radius. If you use d to stand for the diameter and r to stand for the radius, then $d = 2r$.

An **arc** is a part of a circle. The word 'arc' comes from the Latin for 'bow'. The shaded area in Figure 16.3 formed by an arc and two radii is a **sector**. You met this term in Chapter 6 as the mathematical name for the 'slice' of a pie chart.

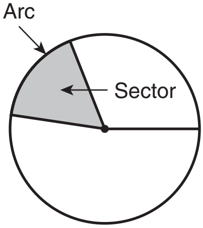

Figure 16.3

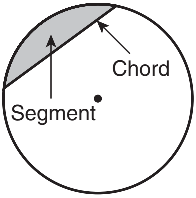

Figure 16.4

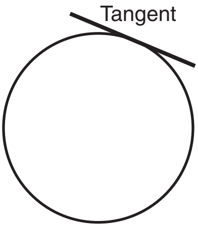

Figure 16.5

A straight line joining two points on a circle is a **chord**. A diameter is, therefore, a special type of chord.

The shaded region in Figure 16.4 formed by an arc and a chord is called a **segment**. The straight line in Figure 16.5 which just touches a circle is a **tangent**, from the Latin for 'touch'. A line is a tangent to a circle if, when extended in either direction, it meets the circle only once.

16.2 Circumference of a circle

Circumference is the special name of the perimeter of a circle, that is, the distance all around it. The word is derived from the Latin *circum*, meaning 'round', and *fero*, meaning 'carry'.

If you measure the circumferences and diameters of some circular objects and then, for each one, work out the value of $\frac{\text{circumference}}{\text{diameter}}$, you should find that your answer is always just over 3. In fact, the value of $\frac{\text{circumference}}{\text{diameter}}$ is the same for every circle and is denoted by the Greek letter π, pronounced 'pie'. So for every circle

$$\frac{\text{circumference}}{\text{diameter}} = \pi .$$

William Jones, an Englishman, was, in 1706, the first mathematician to use π in this way, probably because π is the first letter of the Greek word *peripheria*, meaning 'periphery' or 'boundary'.

Archimedes (c. 225 BC) showed that the value of π is between $3\frac{1}{7}$ and $3\frac{10}{71}$. Using modern computing techniques, its value has been calculated to over six billion decimal places but it cannot be found exactly as the decimal never recurs or terminates.

To perform calculations involving π, use the π key on your calculator. If you do not have a π key, 3.14 or 3.142 is sufficiently accurate for all practical purposes.

Insight

There are many memory aids to help you remember the first few digits of π. One such example is: 'May I draw a round perimeter?' Count the number of letters in each word and you get 3.14159.

Using C to stand for the circumference of a circle and d to stand for its diameter, you can write the result $\frac{\text{circumference}}{\text{diameter}} = \pi$ as $\frac{C}{d} = \pi$. Multiplying both sides by d, you obtain the formula $C = \pi d$.

So, to find the circumference of a circle, you multiply its diameter by π. The same units must be used for the circumference and the diameter.

Example 16.2.1

The diameter of a circle is 6.3 cm. Calculate its circumference. Give your answer correct to three significant figures.

If C cm is the circumference, then $C = \pi \times 6.3 = 19.8$, so the circumference is 19.8 cm, correct to three significant figures.

If you are told the radius of a circle and want to find its circumference, you could first double the radius to obtain the diameter and then multiply the result by π.

Alternatively, using r to stand for the radius of the circle, you can replace d by $2r$ in the formula $C = \pi d$. This gives $C = \pi \times 2r$, usually written as $C = 2\pi r$. In this formula, the same units must be used for the circumference and the radius.

Example 16.2.2

The radius of a circle is 1.96 m. Calculate its circumference. Give your answer correct to three significant figures.

If C m is the circumference, then $C = 2 \times \pi \times 1.96 = 12.3$, so the circumference is 12.3 m, correct to three significant figures.

If you know the circumference of a circle and want to find its diameter, you could substitute the value of C in the formula $C = \pi d$ and then solve the resulting equation.

Alternatively, you can make d the subject of the formula $C = \pi d$.

Dividing both sides by π gives $d = \frac{C}{\pi}$. So, to find the diameter of a circle, you divide its circumference by π.

Example 16.2.3

The circumference of a circle is 57.8 cm. Calculate its diameter. Give your answer correct to three significant figures.

As $d = \frac{C}{\pi}$, $d = \frac{57.8}{\pi} = 18.4$. The diameter is 18.4 cm, correct to three significant figures.

If you know the circumference of a circle and want to find its radius, you could find the diameter of the circle and then halve your answer.

Alternatively, you could make r the subject of the formula $C = 2\pi r$.

Dividing both sides of the formula by 2π gives $r = \frac{C}{2\pi}$. So, to find the radius of a circle, you divide its circumference by 2π.

EXERCISE 16.1

Give all answers correct to three significant figures.

1 The diameter of a circle is 3.7 cm. Calculate its circumference.
2 The radius of a circle is 6.34 m. Calculate its circumference.
3 The circumference of a circle is 76 cm. Calculate its diameter.
4 The circumference of a circle is 97 m. Calculate its radius.
5 The diameter of the Earth at the equator is 12 756 km. Calculate the circumference of the Earth to the nearest hundred km.
6 The minute hand of Big Ben is 3.35 m long. How far does the tip of this hand move in 30 minutes?
7 The diameter of a bicycle wheel is 71 cm.
  (a) Calculate the circumference of the wheel.
  (b) Calculate the number of revolutions the wheel makes when the bicycle travels 1 kilometre.
8 The largest Big Wheel in the world has a circumference of 200 m. Calculate its radius.

16.3 Area of a circle

One of the most famous problems of antiquity was to construct, using only compasses and a straight edge, a square whose area is equal to that of a given circle. In fact, the task is impossible, precisely what is meant today when someone is said to be trying to 'square the circle'.

In attempting this problem, however, the Greeks developed methods of finding the area of a circle. One of these was the 'method of exhaustion', in which a regular polygon was drawn inside the circle and its area calculated. The more sides the regular polygon has, the nearer its area is to the circle; the area between the polygon and the circle is 'exhausted'. (See Figure 16.6.)

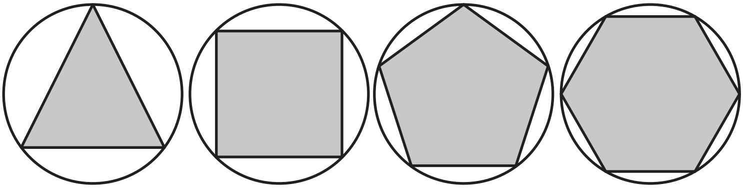

Figure 16.6

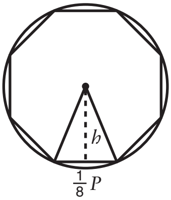

Figure 16.7

The area of a polygon was found by splitting it up into triangles by drawing lines from the centre of the circle to each vertex of the polygon. The area of each triangle, and thus the area of the polygon, were calculated. Figure 16.7 shows a regular octagon with perimeter P. It has been split into eight triangles with vertical height h. The length of the base of each triangle is $\frac{1}{8}P$. The area of each triangle is $\frac{1}{2} \times \frac{1}{8}P \times h$.

So the area of the octagon is $8 \times \frac{1}{2} \times \frac{1}{8}P \times h = \frac{1}{2}Ph$ or, in words, area of polygon = $\frac{1}{2}$ × its perimeter × vertical height of triangle.

The same result is true for any regular polygon. As the number of sides of the polygon increases, the polygon becomes more and more like a circle. Its perimeter is almost equal to the circumference of the circle and the vertical height of each triangle is almost equal to the radius. Thus,

$$\text{area of a circle} = \frac{1}{2} \times \text{circumference} \times \text{radius}.$$

Using A for the area of a circle, r for its radius and $2\pi r$ for the circumference, you get $A = \frac{1}{2} \times 2\pi r \times r$. This simplifies to $A = \pi r^2$.

Example 16.3.1

The radius of a circle is 6.7 cm. Calculate its area. Give your answer correct to three significant figures.

If A cm^2 is the area, $A = \pi \times 6.7^2 = 141$, so the area of the circle is 141 cm^2, correct to three significant figures.

Insight

It is important to recognize that the area of the circle is *not* $(\pi \times 6.7)^2$; you must find 6.7^2 first, and then multiply by π.

You can avoid major errors like this by making an estimate of the answer. In this case, $\pi \times 6.7^2 \approx 3 \times 7^2 = 3 \times 49 \approx 150$.

If you are told the diameter of a circle and want to find its area, you must first find the radius by halving the diameter.

Example 16.3.2

The diameter of a circle is 9.56 m. Calculate its area. Give your answer correct to three significant figures.

Let r m be the radius and A m^2 be the area. Then $r = \frac{1}{2} \times 9.56 = 4.78$, and $A = \pi \times 4.78^2 = 71.8$. So the area of the circle is 71.8 m^2, correct to three significant figures.

If you know the area of a circle and want to find its radius, you substitute the value of A in $A = \pi r^2$ and solve the resulting equation.

Example 16.3.3

The area of a circle is 70 cm^2. Calculate its radius. Give your answer correct to three significant figures.

In $A = \pi r^2$, substituting 70 for A gives $70 = \pi r^2$.

Dividing both sides by π gives $r^2 = \frac{70}{\pi} = 22.28\ldots$.

Then $r = \sqrt{22.28\ldots} = 4.72$ so the radius is 4.72 cm, correct to three significant figures.

Note that $r = -\sqrt{22.28\ldots}$ also solves the equation $70 = \pi r^2$ but the radius of a circle must be positive.

EXERCISE 16.2

Give all answers correct to three significant figures.

1 The radius of a circle is 26.3 cm. Calculate its area.
2 The diameter of a circle is 28 m. Calculate its area.
3 The area of a circle is 100 cm^2. Calculate its radius.
4 The area of a circle is 58 cm^2. Calculate its diameter.
5 The radius of a semicircle is 7.31 m. Calculate its area.
6 The sides of a square piece of metal are 22 mm long. A hole with a diameter of 8 mm is drilled in the piece of metal. Calculate the area of metal which remains.
7 Two concentric circles (circles having the same centre) have radii of 4.7 cm and 6.9 cm. Calculate the area of the ring between the circles. (It is called an *annulus*.)
8 The diameter of a circular pond is 8 m. The pond has a path, which is 1 m wide, around it. Calculate the area of the path.
9 The radius of a circle is 10 cm. How many times greater is its area when its radius is doubled?
10 A farmer has 200 m of fencing. He arranges it in a circle. Calculate the area he encloses.

EXERCISE

16.4 Two properties of circles

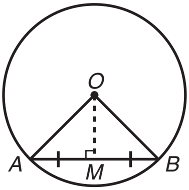

Figure 16.8

In Figure 16.8, *AB* is a chord of a circle centre *O*.

The mid-point of *AB* is *M*. The line *OM* is a line of symmetry so it is perpendicular to the chord *AB*.

In general, the line joining the centre of a circle to the mid-point of a chord is perpendicular to the chord.

The *converse*, or opposite, of this is also true: a line drawn from the centre of a circle perpendicular to a chord bisects the chord.

Insight

Remember two lines are *perpendicular* when they are at right angles, and *bisect* means to cut exactly in half.

Example 16.4.1

In the figure, *O* is the centre of the circle and *R* is the mid-point of the chord *PQ*. Angle *OPR* is 54°. Find the size of the angle marked *x*.

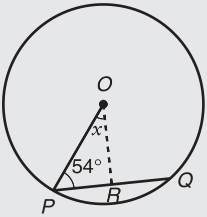

Angle *PRO* = 90° (*OR* is perpendicular to *PQ*)
$x + 90° + 54° = 180°$ (angles in a triangle), so
$x + 144° = 180°$, giving $x = 36°$

Alternatively, you could find angle *POQ* which is $180° - 2 \times 54° = 72°$, and then use the fact that *OR* is a line of symmetry. *OR* therefore bisects angle *POQ*, and it follows that angle $POR = \frac{1}{2} \times 72° = 36°$.

In Figure 16.9, *O* is the centre of a circle. A tangent touches the circle at *T*. The radius, *OT*, is a line of symmetry, so it is perpendicular to the tangent.

In general, a radius drawn to the point where a tangent touches a circle is perpendicular to the tangent.

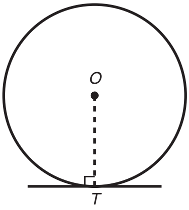

Figure 16.9

Example 16.4.2

In the diagram, *PQ* is the tangent at *T* to a circle centre *O*. Angle *PTR* = 47°. Calculate the size of the angle marked *y*.

Angle *PTO* = 90° (radius is perpendicular to tangent)
Angle *RTO* = 90° – 47° = 43°
Angle *TRO* = 43° (triangle *ORT* is isosceles)
43° + 43° + *y* = 180° (angle sum of triangle)
giving *y* + 86° = 180°, so *y* = 180° – 86° = 94°

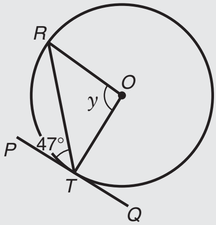

EXERCISE 16.3

Find the size of each of the angles marked with letters. *O* is the centre of the circle in every question, and in questions 3–6 a tangent is drawn to the circle.

1

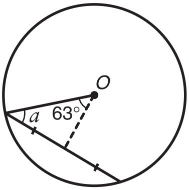

2

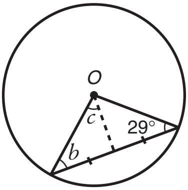

(Contd)

3

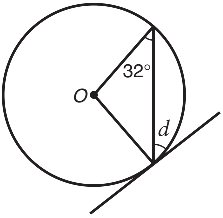

4

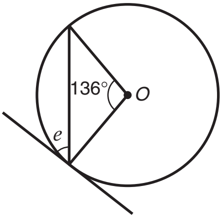

5

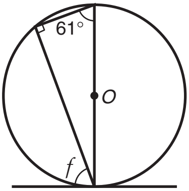

6

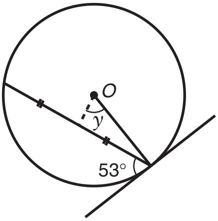

THINGS TO REMEMBER

- *The radius, r, is the distance from the centre of a circle to a point on the circle. The plural of radius is radii.*

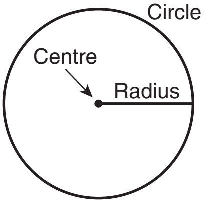

- *A diameter divides a circle into two semicircles and is a line of symmetry.*

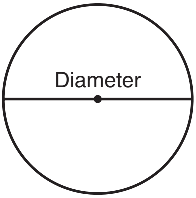

- *The diameter, d, is twice as long as a radius, r. So* $d = 2r$.
- *An arc is a part of a circle. The shaded area formed by an arc and two radii is a sector.*

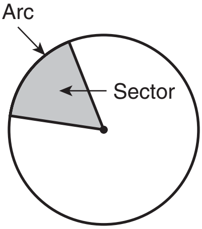

- *A straight line joining two points on a circle is a chord. The region formed by an arc and a chord is called a segment.*

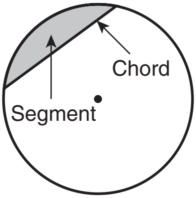

- *The circumference is the perimeter of a circle.*
- *The circumference, C, of a circle is found using the formulae:* $C = \pi d$ *or* $C = 2\pi r$.

- *The area, A, of a circle is found using the formula A =* πr^2.
- *A tangent is a straight line that touches a circle only once.*

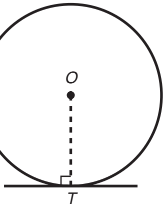

- *A radius and a tangent meet at right angles.*
- *A line drawn from the centre of a circle to the mid-point of a chord cuts the chord at right angles.*

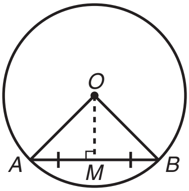

- *A line drawn from the centre of a circle perpendicular to a chord bisects the chord, i.e. it cuts the chord exactly in half.*

17

Probability

In this chapter you will learn:
- ***the meaning of 'probability'***
- ***how to find and use relative frequency***
- ***how to find the probability of single events and of two events***
- ***how to find and use expected frequency.***

17.1 Introduction

The **probability** of an event is a measure of the likelihood that the event will occur. Probability ranges from 0, when the event is impossible, to 1, when the event is certain to happen, although, as Benjamin Franklin wrote, 'In this world nothing can be said to be certain but death and taxes.'

Sometimes information from the past is used to predict the probabilities of future events. For example, since about 1700, life insurance companies have used mortality tables to estimate life expectancies and thus to set their premiums.

The study of probability has its origins in the mid-sixteenth century, when mathematicians, notably Cardan and Tartaglia, attempted to answer questions arising in gambling. It was to be another 100 years, however, before the Frenchmen Blaise Pascal and Pierre de Fermat developed probability theory – in effect, the laws of chance.

Today, probability is widely used in science and industry, its applications ranging from genetics and quantum mechanics to insurance. Laplace, himself a leading figure in the field of probability, called it a science which began with play but evolved into ‘the most important object of human knowledge’.

You can show the probability of an event on the **probability scale** in Figure 17.1.

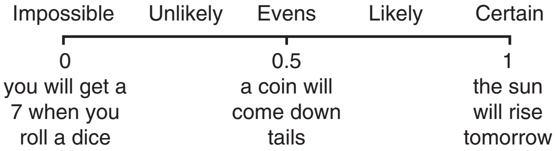
Impossible
Unlikely
Evens
Likely
Certain
0
you will get a 7 when you roll a dice
0.5
a coin will come down tails
1
the sun will rise tomorrow

Figure 17.1

17.2 Relative frequency

The relative frequency of an event is defined as

$$\frac{\text{number of times event occurs}}{\text{total number of trials}}.$$

One way of estimating the probability of an event is to carry out an experiment, or **trial**, a number of times and to find the relative frequency of the event.

For example, if you had a bag of coloured beads and wanted to estimate the probability that you will pick a red bead, you could, without looking, pick a bead from the bag, record its colour and replace it. Suppose that you carried out this experiment 20 times and picked a red bead on 9 occasions. Then

$$\text{relative frequency} = \frac{\text{number of times event occurs}}{\text{total number of trials}} = \frac{9}{20}$$

The relative frequency, $\frac{9}{20}$, estimates the probability of picking a red bead. This result may also be given as a decimal or a percentage but not in any other form. The more times you carry out the experiment, the more accurate your estimate of the probability is likely to be.

EXERCISE 17.1

1 John spun a coin 50 times. It came down tails 29 times. Find the relative frequency of it coming down tails. Give your answer as
(a) a fraction, **(b)** a decimal, **(c)** a percentage.

2 Kate spun a coin 80 times. It came down heads 34 times. Find the relative frequency of it coming down
(a) heads, **(b)** tails.

3 David rolled a dice 100 times. He got a six 22 times. Find the relative frequency of him getting a six.

4 Without looking, Sharon picked a bead from a bag of coloured beads, recorded its colour and then replaced it. She did this 60 times and picked a blue bead on 12 occasions. Find an estimate for the probability that she will pick a blue bead.

5 Paul picked a card from a pack, recorded its suit and then replaced it. He did this 40 times and picked a heart on 12 occasions. Find an estimate for the probability that he will pick a heart.

17.3 Probability of a single event

You can find the probability of an event using

$$\text{Probability} = \frac{\text{number of successful outcomes}}{\text{total number of possible outcomes}}$$

If you spin a coin, there are two possible outcomes, heads or tails. If the coin is **fair**, or **unbiased**, each of these outcomes is equally likely. Taking a head as the 'successful' outcome,

$$\text{probability of a head} = \tfrac{1}{2}$$

Instead of repeatedly writing the word 'probability', you can use the letter P and write simply, $P(\text{head}) = \frac{1}{2}$.

As probability is expressed as a fraction, decimal or percentage, you could also write $P(\text{head}) = 0.5$ or $P(\text{head}) = 50\%$.

Example 17.3.1

A fair dice is rolled. Find the probability that the number obtained is

(a) a 5, **(b)** an odd number,

(c) a 2 or a 3, **(d)** less than 7.

There are six possible outcomes and they are all equally likely.

(a) As there is only one 5 on the dice, $P(5) = \frac{1}{6}$.

(b) There are three odd numbers on the dice, so $P(\text{odd no.}) = \frac{3}{6} = \frac{1}{2}$.

(c) A 2 or a 3 is a 'success', so $P(2 \text{ or } 3) = \frac{2}{6} = \frac{1}{3}$.

(d) As all the numbers on a dice are less than 7, the number obtained will be less than 7. Thus, $P(\text{number less than } 7) = 1$.

In part (c) of Example 17.3.1, $P(2) = \frac{1}{6}$ and $P(3) = \frac{1}{6}$, so, in this case, $P(2 \text{ or } 3) = P(2) + P(3)$, and you could have added the probabilities. This is because the two events cannot both occur and are said to be **mutually exclusive.**

Insight

Two events are mutually exclusive when once one event has occurred it excludes the possibility of the other event occurring and vice versa. For example, you can't roll a dice and get a 3 *and* an even number at the same time; the event 'getting a 3' excludes the possibility of the event 'getting an even number' and vice versa.

If the events are not mutually exclusive, adding the probabilities would be incorrect.

For example, if you wanted to find the probability of an odd number or a multiple of 3, then the numbers 1, 3, 5 and 6 satisfy the criteria, so
$P(\text{odd number or a multiple of } 3) = \frac{4}{6} = \frac{2}{3}$
But
$P(\text{odd number}) = \frac{3}{6} = \frac{1}{2}$
and
$P(\text{multiple of } 3) = \frac{2}{6} = \frac{1}{3}$
so
$P\ (\text{odd number or a multiple of } 3) \neq P(\text{odd number}) + P(\text{multiple of } 3)$
(The sign ≠ means 'is not equal to.')

This is because, the number 3 satisfies both criteria, and the events 'odd number' and 'multiple of 3' are not mutually exclusive.

In the following example, the term 'at random' means that every bead has an equal chance of being selected.

Example 17.3.2

There are 9 beads in a bag. 2 beads are red, 3 beads are green and 4 beads are blue. A bead is picked at random. Find the probability it will be
(a) red, **(b)** red or green, **(c)** black, **(d)** not red.

(a) There are 2 red beads, so P(red) = $\frac{2}{9}$.
(b) 5 beads are red or green, so P(red or green) = $\frac{5}{9}$. (Note that these two events are mutually exclusive.)
(c) There are no black beads in the bag so it is impossible for one to be picked. Therefore P(black) = 0.
(d) 7 of the beads are not red, so P(not red) = $\frac{7}{9}$.

In part (d), P(not red) = 1 – P(red). In general, if the probability that an event will occur is p, then the probability that the event will not occur is $1 - p$. For example, if the probability that it will rain tomorrow is 0.7, the probability that it will not rain tomorrow is $1 - 0.7 = 0.3$.

Insight

When an event is certain it has a probability of 1. Any event will either happen or not happen so
P(event happens) + P(event doesn't happen) = 1.

EXERCISE 17.2

1 A fair dice is rolled. Find the probability of obtaining
(a) a 6, **(b)** an even number,
(c) a prime number, **(d)** a 4 or a 5.

(Contd)

2 A fair spinner has the numbers 1, 2, 3, 4 and 5 on it. Find the probability that it will land on
(a) a 1, **(b)** an odd number, **(c)** a 3 or a 4,
(d) a 6, **(e)** an even number or a prime number,
(f) a number other than 5.

3 There are 10 beads in a bag. 5 beads are black, 3 beads are white and 2 beads are red. A bead is picked at random. Find the probability that it will be
(a) white, **(b)** black or red,
(c) not white, **(d)** not blue.

4 A card is picked at random from a pack of 52. Find the probability that it will be
(a) the queen of hearts, **(b)** a king,
(c) a club, **(d)** a red card,
(e) a king or a queen, **(f)** an ace or a spade.

5 A letter is chosen at random from the word MATHEMATICS. Find the probability that it will be
(a) an H, **(b)** an M, **(c)** a B,
(d) a vowel, **(e)** an S or a T, **(f)** other than an A.

6 In a raffle, 500 tickets are sold. George buys one ticket and Julia buys five. There is one winning ticket. Find the probability that
(a) George will win, **(b)** Julia will win.

7 The probability that Sally will be late for work tomorrow is $\frac{2}{9}$. What is the probability that she will not be late for work?

8 My drawer contains grey socks, brown socks and black socks. I pick a sock at random. The probability that it will be a grey sock is 0.5 and the probability that it will be a brown sock is 0.2. What is the probability that it will be a black sock?

17.4 Two events

If you spin two coins, you can get two heads, two tails or one of each. If you spin two coins 100 times, you will find that you get two heads about 25 times, two tails about 25 times and one of each about 50 times. To explain this, there are four possible outcomes which you can show in a list.

HH HT TH TT

Alternatively, you can show the possible outcomes on a diagram like Figure 17.2; each cross represents one of the four possible outcomes.

When you spin two coins, the probabilities that you will get two heads, two tails and a head and a tail are respectively $\frac{1}{4}$, $\frac{1}{4}$ and $\frac{2}{4} = \frac{1}{2}$. The sum of these three probabilities is 1, as one of them is certain to happen.

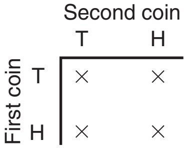

Figure 17.2

When you spin two coins, the outcome for one coin has no effect on the outcome for the other coin. The events are said to be **independent**.

Insight

If a fair coin lands head nine times in a row, the probability of throwing a head on the tenth throw is still 0.5 because the events are independent. Although, of course the event ‘throw 10 heads in a row’ is itself very unlikely – but that would be a different question altogether!

Rolling two dice is another example of two independent events. You can show the 36 possible outcomes as a list, rather like coordinates.

(1, 1)	(2, 1)	(3, 1)	(4, 1)	(5, 1)	(6, 1)
(1, 2)	(2, 2)	(3, 2)	(4, 2)	(5, 2)	(6, 2)
(1, 3)	(2, 3)	(3, 3)	(4, 3)	(5, 3)	(6, 3)
(1, 4)	(2, 4)	(3, 4)	(4, 4)	(5, 4)	(6, 4)
(1, 5)	(2, 5)	(3, 5)	(4, 5)	(5, 5)	(6, 5)
(1, 6)	(2, 6)	(3, 6)	(4, 6)	(5, 6)	(6, 6)

You can also show the outcomes on a diagram, where each cross represents one of the 36 possible outcomes. See Figure 17.3. Using either the list or the figure, you can find, for example, that the probability of a double six is $\frac{1}{36}$ or the probability that the sum of the numbers on the two dice is 4 is $\frac{3}{36}$ or $\frac{1}{12}$. In the latter case the table in Figure 17.4 showing the sum of the two numbers for all 36 possible outcomes is helpful.

	Second dice					
First dice	1	2	3	4	5	6
1	×	×	×	×	×	×
2	×	×	×	×	×	×
3	×	×	×	×	×	×
4	×	×	×	×	×	×
5	×	×	×	×	×	×
6	×	×	×	×	×	×

Figure 17.3

Insight

Diagrams like these are sometimes called probability space diagrams or sample space diagrams. Always work systematically when listing all possible outcomes.

You can use the table in Figure 17.4 to help you find, for example, the probability that the sum of the numbers on the two dice is more than 10. Of the 36 sums in the table, 3 are more than 10, so P(sum >10) = $\frac{3}{36} = \frac{1}{12}$.

	Second dice					
First dice +	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Figure 17.4

There are 6 possible outcomes for one dice and 36 (= 6 × 6) possible outcomes for two dice. Similarly, there are 2 possible outcomes for one coin and 4 (= 2 × 2) possible outcomes for two coins. This suggests that you multiply the number of possible outcomes for each of the separate items to find the number of possible outcomes for a combination of the two items.

Example 17.4.1

A coin is spun and a dice is rolled.

(a) How many possible outcomes are there?

(b) Show all the possible outcomes on a diagram.

(c) Find the probability of obtaining **(i)** a head and a 5, **(ii)** a tail and an even number.

(a) There are two possible outcomes for a coin and six possible outcomes for a dice. So, for the coin and the dice together, there are $2 \times 6 = 12$ possible outcomes.

(b) See the figure below.

(c) **(i)** $P(H, 5) = P(H \text{ and } 5) = \frac{1}{12}$.

(ii) Three possible outcomes have a tail and an even number: they are (T, 2), (T, 4) and (T, 6).

So $P(\text{tail and even number}) = \frac{3}{12} = \frac{1}{4}$.

		Dice					
		1	2	3	4	5	6
Coin	H	×	×	×	×	×	×
	T	×	×	×	×	×	×

EXERCISE 17.3

1 Two dice are rolled. Find the probability that

(a) the numbers on the dice are the same,

(b) there is a 6 on at least one of the dice,

(c) the sum of the numbers is less than 5.

2 Two dice are rolled. The sum of the numbers on the dice is found.

(a) Complete the table to show the probability of obtaining each of the possible sums.

Sum	2	3	4	5	6	7	8	9	10	11	12
Probability							$\frac{5}{36}$				$\frac{1}{36}$

(Contd)

(b) What is the most likely sum?

(c) Find the sum of the probabilities. Explain your answer.

3 A coin is spun and a dice rolled. Find the probability of obtaining

(a) a tail and a 5,

(b) a head and 2 or a head and 3,

(c) a head and a prime number,

(d) a tail and a number less than 4.

4 Two dice are rolled. The difference between the numbers on the dice is found.

(a) Show all the possible outcomes in a table.

(b) Find the probability that the difference will be **(i)** 2, **(ii)** 5.

(c) What is the most likely difference?

5 A spinner has the numbers 1, 2 and 3 on it. A coin and this spinner are spun.

(a) How many possible outcomes are there?

(b) List all the possible outcomes.

(c) Draw a diagram to show all the possible outcomes.

(d) Find the probability of obtaining

(i) a tail and a 3, **(ii)** a head and an odd number,

(iii) a tail and a factor of 6.

6 Two spinners are spun. Each has the numbers 1, 2, 3, 4 and 5 on it. The sum of the two numbers on the spinners is found.

(a) How many possible outcomes are there?

(b) Show all the possible outcomes in a table.

(c) Find the probability that the sum will be **(i)** 5, **(ii)** 3 or 4.

(d) What is the most likely sum?

7 Out of three beads in a bag, one bead is red, one is green and one is blue. A dice is thrown and a bead is picked at random from the bag.

(a) How many possible outcomes are there?

(b) Find the probability of obtaining

(i) a red bead and a 4, **(ii)** a blue bead and a 5 or a 6,

(iii) a green bead and a number greater than 4.

8 One spinner has the numbers 1, 2, 3 and 4 on it. Another spinner has the numbers 2, 3 and 5 on it. The two spinners are spun. Find the most likely sum of the two numbers.

17.5 Tree diagrams

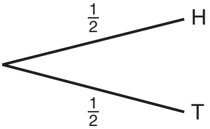
$\frac{1}{2}$
H
$\frac{1}{2}$
T

Figure 17.5

Another way of showing the possible outcomes when a coin is spun is to use a tree diagram (see Figure 17.5). The probability of each outcome is written on the branch leading to it. The sum of the probabilities on the branches is 1, because you are certain to get one of the outcomes. You can extend the diagram to show the possible outcomes when two coins are spun. (See Figure 17.6.)

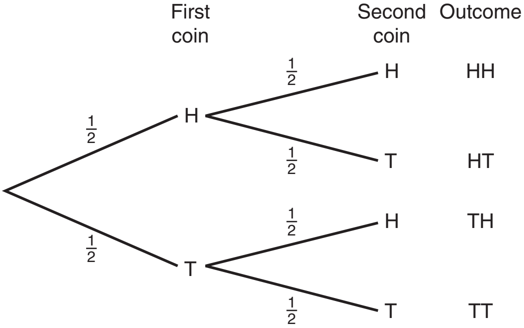
First coin
Second coin
Outcome
$\frac{1}{2}$
H
HH
$\frac{1}{2}$
H
$\frac{1}{2}$
T
HT
$\frac{1}{2}$
H
TH
$\frac{1}{2}$
T
$\frac{1}{2}$
T
TT

Figure 17.6

At the start of Section 17.4, you saw that the probability of getting two heads is $\frac{1}{4}$. You can find this probability by working out $\frac{1}{2} \times \frac{1}{2}$, *multiplying* the probabilities on the two branches leading to the outcome HH. This only works for independent events.

In general, you can find the probability that two independent events both occur by multiplying their probabilities.

Example 17.5.1

There are 5 marbles in a bag: 2 of the marbles are red and the other 3 marbles are blue. A marble is picked at random. Its colour is noted and the marble is replaced in the bag. A marble is again picked at random and its colour noted.

(a) Draw a tree diagram showing all the outcomes and the probability of each one.

(b) Find the probability that **(i)** both marbles will be the same colour, **(ii)** at least one of the marbles will be red.

(a)

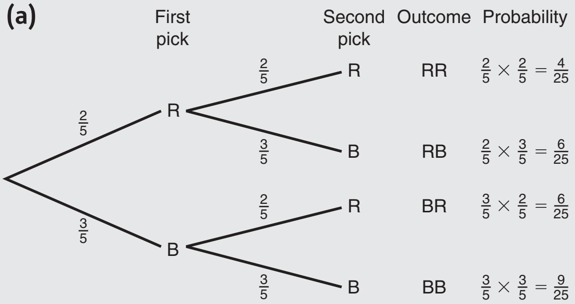
First pick
Second pick
Outcome
Probability
$\frac{2}{5}$
R
$\frac{3}{5}$
B
$\frac{2}{5}$
R
$\frac{3}{5}$
B
$\frac{2}{5}$
R
$\frac{3}{5}$
B
RR
RB
BR
BB
$\frac{2}{5} \times \frac{2}{5} = \frac{4}{25}$
$\frac{2}{5} \times \frac{3}{5} = \frac{6}{25}$
$\frac{3}{5} \times \frac{2}{5} = \frac{6}{25}$
$\frac{3}{5} \times \frac{3}{5} = \frac{9}{25}$

(Notice that the sum of the probabilities is 1, because you are certain to get one of the four listed outcomes.)

(b) (i) P(both marbles will be the same colour) = P(RR or BB)
As the outcomes RR and BB are mutually exclusive,
P(RR or BB) = P(RR) + P(BB), so P(RR) + P(BB) = $\frac{4}{25} + \frac{9}{25} = \frac{13}{25}$.
(ii) 'At least' one of the marbles will be red means that one or both of the marbles will be red, so
P(at least one of the marbles will be red)
= P(RR or RB or BR) = P(RR) + P(RB) + P(BR) = $\frac{4}{25} + \frac{6}{25} + \frac{6}{25} = \frac{16}{25}$.

Alternatively, P(at least one of the marbles will be red)
= 1 – P(BB) = 1 – $\frac{9}{25} = \frac{16}{25}$.

Insight

Notice that the final probabilities at the end of the branches must sum to 1, as the only possible outcomes are RR, RB, BR and BB. So

$$\begin{aligned}&P(R,R)+P(R,B)+P(B,R)+P(B,B)\\&=\tfrac{4}{25}+\tfrac{6}{25}+\tfrac{6}{25}+\tfrac{9}{25}\\&=\tfrac{25}{25}\\&=1\end{aligned}$$

This serves as a useful check once you have drawn a tree diagram.

It is not always necessary to draw a complete tree diagram. Sometimes not all the branches are relevant to the problem.

Example 17.5.2

I have to go through two sets of traffic lights on my way to work. The probability that the first set will be on green is 0.7 and the probability that the second set will be on green is 0.8. Find the probability that the first set will be on green and the second set will not be on green.

P(second set will not be on green) = 1 – 0.8 = 0.2

The relevant part of the tree diagram is shown in the diagram.

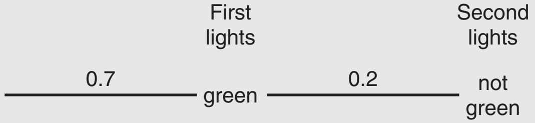

The probability that the first set will be on green and the second set will not be on green is $0.7 \times 0.2 = 0.14$.

EXERCISE 17.4

1 There are 8 marbles in a bag; 5 of the marbles are red and the other 3 marbles are green. A marble is picked at random. Its colour is noted and the marble is replaced in the bag. A marble is again picked at random and its colour noted.

(a) Draw a tree diagram showing all the outcomes and the probability of each one.

(b) Find the probability that

  (i) the marbles will be different colours,

  (ii) neither marble will be green.

(Contd)

EXERCISE

2 Sharon has a biased coin. The probability that it will come down heads is $\frac{2}{3}$. Sharon spins the coin twice.

(a) Draw a tree diagram showing all the outcomes and the probability of each one.

(b) Find the probability that she will get

(i) at least one head, **(ii)** either two heads or two tails.

3 The probability that Tom will pass a science test is 0.7 and the probability that he will pass a French test is 0.4. Find the probability that he will

(a) pass both tests, **(b)** fail both tests,

(c) pass at least one of the tests.

4 A card is picked at random from a pack. It is replaced and a card is again picked at random. Find the probability of picking

(a) two red cards,

(b) two aces,

(c) first a heart and second a king,

(d) a heart and a king (in either order),

(e) at least one diamond.

5 The letters in the word LONDON are put in one bag and the letters in the word LIVERPOOL are put in another bag. A letter is picked at random from each bag. Find the probability of picking

(a) two Os, **(b)** two vowels, **(c)** at least one L.

6 There are 9 beads in a bag, 2 of the beads are red, 3 are blue and the rest are black. A bead is picked at random and then replaced. A bead is again picked at random. Find the probability that

(a) both beads will be red,

(b) both beads will be the same colour,

(c) exactly one of the beads will be black.

7 Holly catches a bus to work every day. The bus may be on time or early or late. The probability that it will be on time is 0.5 and the probability that it will be early is 0.3. Find the probability that, in the next two days, the bus will be

(a) early twice, **(b)** late twice, **(c)** on time at least once.

8 The probability that Melchester Rovers will win a soccer match is 0.6 and the probability that they will draw is 0.3. Find the probability that they will

(a) win the next two matches,

(b) lose the next two matches,

(c) draw at least one of the next two matches.

17.6 Expected frequency

The laws of chance are of most use when applied to a large number of trials rather than to a particular trial. For example, if you spin a coin 100 times, you would expect to get heads about 50 times. This is the **expected frequency**. It may be calculated by finding the product of the probability, $\frac{1}{2}$, and the number of trials, 100.

Expected frequency = probability × number of trials

Similarly, it is not the fact that the probability that a 30-year-old man will live to the age of 70 is 0.4514 which is of interest to a life assurance company. What this means to them is that about 4500 of every 10 000 30-year-old men will live to the age of 70. This aspect of probability is known as the law of large numbers.

Example 17.6.1

A dice is rolled 600 times. Find the expected frequency of a 5.

$P(5) = \frac{1}{6}$. The number of trials is 600. Using the formula
expected frequency = probability × number of trials,
expected frequency $= \frac{1}{6} \times 600 = 100$

Example 17.6.2

Two dice are rolled 180 times. How many times would you expect their sum to be **(a)** 2, **(b)** 11 or 12?

(a) P(sum is 2) $= \frac{1}{36}$. The number of trials is 180, so expected frequency $= \frac{1}{36} \times 180 = 5$.

(b) P(sum is 11 or 12) $= \frac{3}{36} = \frac{1}{12}$, so expected frequency $= \frac{1}{12} \times 180 = 15$

EXERCISE 17.5

1 A coin is spun 500 times. How many heads would you expect?

2 A dice is rolled 300 times. How many times would you expect
(a) a 4, **(b)** a 2 or a 3, **(c)** an even number?

3 There are 8 counters in a bag; 5 are green, 2 are red and 1 is black. A counter is picked at random from the bag and then replaced. In 200 picks, how many times would you expect to get
(a) a green counter,
(b) a red counter or a black counter,
(c) a counter which is not black?

4 The probability that a baby will be a boy is 0.52. How many boys would you expect in 1000 births?

5 The probability that it will rain on any day in a month is 40%. How many rainy days would you expect in a month of 30 days?

6 Two coins are spun 200 times. How many times would you expect to get
(a) two tails, **(b)** a head and a tail, **(c)** at least one head?

7 Two dice are rolled 720 times. How many times would you expect their sum to be
(a) 4, **(b)** more than 9, **(c)** 7 or 8?

8 A coin is spun and a dice is rolled. This is done 300 times. How many times would you expect to get
(a) a tail and a 4, **(b)** a head and an odd number?

9 A card is picked at random from a pack and a coin is spun. This is done 260 times. How many times would you expect an ace and a head?

10 A card is picked at random from a pack and a dice is spun. This is done 400 times. How many times would you expect a diamond and a prime number?

THINGS TO REMEMBER

- *The probability of an event is a measure of the likelihood that the event will occur. Probability ranges from 0, when the event is impossible, to 1, when the event is certain to happen.*

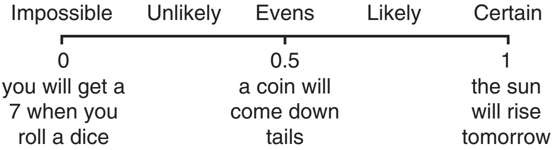
Impossible
Unlikely
Evens
Likely
Certain
0
you will get a 7 when you roll a dice
0.5
a coin will come down tails
1
the sun will rise tomorrow

- $$\text{Relative frequency} = \frac{\text{number of times event occurs}}{\text{total number of trials}}$$

- $$\text{Probability} = \frac{\text{number of successful outcomes}}{\text{total number of possible outcomes}}$$

- *A coin, dice or spinner is fair (or unbiased) when each possible outcome is equally likely.*

- *Two events, A and B, are said to be mutually exclusive if they cannot both occur at the same time.* $P(A \text{ or } B) = P(A) + P(B)$.

- *Given that the probability that an event will occur is* p, *then the probability that the event will not occur is* $1 - p$.

- *Two events are said to be independent when the outcome of one event has no effect on the outcome of the other event. The probability that two independent events both occur is found by multiplying their probabilities.*

- *You can list all the possible outcomes of two or more events on a diagram.*

- *A tree diagram can be used to show all the possible outcomes of two or more independent events. You can find a particular outcome by multiplying the probabilities along the branches.*

- *Expected frequency = probability × number of trials*

18

Three-dimensional shapes

In this chapter you will learn:

- ***the names of solids***
- ***how to find the surface area of a solid***
- ***how to find and use the volume of a cuboid***
- ***how to find the volume of a prism.***

18.1 Introduction

Three-dimensional shapes, or solids, have length, breadth and height (or thickness). In other words, they take up space in all directions. An example is the **cube**, shown in Figure 18.1.

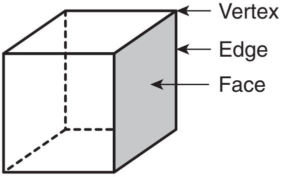

Figure 18.1

One face has been shaded. A cube has 6 faces altogether, all of them squares. The line where two faces meet is called an **edge**; the cube has 12 edges. Edges meet at a **vertex**, or corner; the cube has 8 **vertices**.

The **cuboid**, or rectangular box, shown in Figure 18.2 has the same numbers of faces, edges and vertices as the cube but its faces are rectangular rather than square.

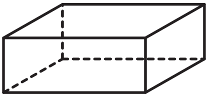

Figure 18.2

There are several categories of three-dimensional shapes. A **prism** is a solid with a constant cross-section. That is, if you cut through a prism parallel to one of its ends, you get the same shape all along its length. This shape is called the **cross-section** of the prism. Both a cube and a cuboid are examples of prisms. The cross-section of a prism must be a polygon and these polygons are joined by parallelograms (often rectangles). A prism is usually defined in terms of the shape of its cross-section. Two prisms are shown in Figure 18.3.

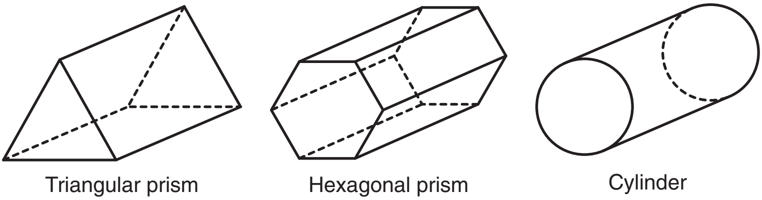

Figure 18.3

Insight

A **cylinder** has a constant circular cross-section. However, although it is not a prism – as a circle is not a polygon – you can treat it as a prism when you work out its volume (see Section 18.4).

The term **pyramid** probably evokes thoughts of the square-based pyramids built over 4000 years ago by the Ancient Egyptians. In fact, the base may be any shape but, whatever shape the base is, lines drawn from the edge of the base must meet at a point, called the **apex**. If the apex is vertically above the centre of the base, the pyramid is called a **right pyramid**.

A pyramid is usually defined in terms of the shape of its base, which is a polygon; the other faces are triangles which meet at the apex. A pyramid with a triangular base is called a **tetrahedron**; if each face is an equilateral triangle, it is a regular tetrahedron. Two pyramids are shown in Figure 18.4 overleaf.

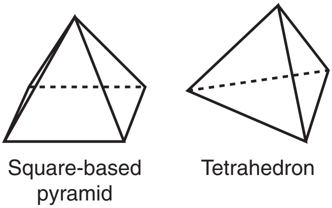

Figure 18.4

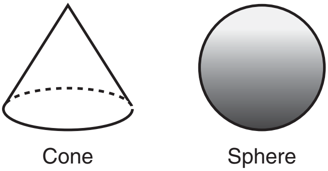

Figure 18.5

Other common three-dimensional shapes are a **cone** and a **sphere**, like a ball, is circular in every direction (Figure 18.5, above).

If every face of a solid is a polygon, the solid is called a **polyhedron**. Of the solids illustrated so far, a cube, a cuboid, a triangular prism, a hexagonal prism, a square-based pyramid and a tetrahedron are polyhedra but a cylinder, a cone and a sphere are not.

If all the edges of a solid are the same length and all its faces are the same shape and all its angles are the same, it is called a **regular** solid. The cube and the regular tetrahedron are the only regular solids illustrated so far but there are five regular solids altogether. Often called **Platonic solids**, they have fascinated mathematicians for 3000 years.

There is an interesting relationship between the number of faces (F), the number of vertices (V) and the number of edges (E) of a polyhedron.

Expressed as a formula, it is

$$F + V = E + 2$$

This result is often known as Euler's theorem. It may have been known to Archimedes (c. 225 BC), but Descartes, in about 1635, was the first to state it in this form. The Swiss mathematician Leonard Euler proved the result in 1752.

EXERCISE 18.1

1 How many faces, vertices and edges do each of these of these polyhedra have?
   (a) triangular prism **(b)** hexagonal prism
   (c) square-based pyramid **(d)** tetrahedron

2 **(a)** Draw a sketch of a prism which has a regular pentagon as its cross-section.
   (b) How many faces, vertices and edges does the prism have?

3 A polyhedron has 10 faces and 16 vertices. Work out how many edges it has.

4 Find out the names of the other three Platonic solids and the number of faces each of them has.

18.2 Nets and surface area

A **net** of a solid shape is a flat shape which can be cut out and folded to make the solid shape. Figure 18.6 shows a sketch of a cuboid and one of its possible nets.

This is not the only possible net of the cuboid; there are several others. Nets are helpful in finding the surface area of a three-dimensional shape. For example, to find the surface area of the cuboid in Figure 18.6, you can find the total area of the six rectangles which make up its net. You can shorten the working by using the fact that there are three pairs of congruent rectangles in the net.

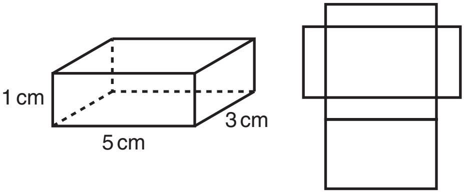

Figure 18.6

The surface area in $\text{cm}^2 = 2 \times (5 \times 3) + 2 \times (5 \times 1) + 2 \times (3 \times 1)$
$= 46\ \text{cm}^2$

In general, the surface area, A, of a cuboid with length l, breadth b and height h is given, in appropriate units, by the formula

$$A = 2lb + 2hl + 2bh$$

Example 18.2.1

The figure shows a solid cylinder. The radius of the circular end of the cylinder is r and the height of the cylinder is h.

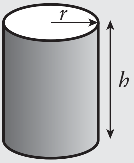

(a) Draw a net of the cylinder.

(b) Find formulae for **(i)** the area of the curved surface (A) and **(ii)** the total surface area (S) of the cylinder.

(c) Use your formula for S to find the total surface area of a cylinder with an end radius of 3.6 cm and a height of 8.3 cm. Give your answer correct to three significant figures.

(a) The net of the curved surface of a cylinder is a rectangle, the shape that you get if you remove the label from a tin of soup and flatten it. The length of the rectangle is equal to the circumference of the circular end of the cylinder. To complete the net you need the two circular ends.

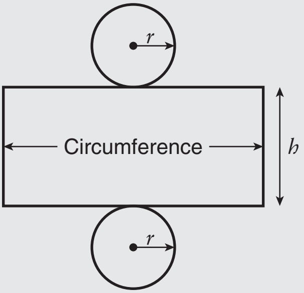

(b) (i) The area A of the rectangle (circumference × h) gives the area of the cylinder's curved surface. Writing the circumference as $2\pi r$, the area $A = 2\pi rh$.

(ii) The total surface area S is the sum of the area of the curved surface and the areas of the two circular ends. Thus $S = 2\pi rh + 2\pi r^2$. You can factorize this expression if you wish, to get $S = 2\pi r(h + r)$.

(c) To find the total surface area, you substitute 3.6 for r and 8.3 for h in one of the formulae for S. Using the factorized form,

$$\begin{aligned} S = 2\pi r(h + r) &= 2 \times \pi \times 3.6 \times (8.3 + 3.6) \\ &= 2 \times \pi \times 3.6 \times 11.9 = 269 \text{ (to 3sf)} \end{aligned}$$

The total surface area is 269 cm^2.

EXERCISE 18.2

1 The sides of a cube are 7 cm long.
   (a) Draw a sketch of a net of the cube.
   (b) Find the surface area of the cube.

2 The length of a cuboid is 10 cm. Its breadth is 6 cm and its height is 4 cm. Find its surface area.

3 The figure shows a triangular prism. The cross-section of the prism is a right-angled triangle with sides 3 cm, 4 cm and 5 cm long. The length of the prism is 8 cm.
   (a) Draw a sketch of the net of the prism.
   (b) Find the surface area of the prism.

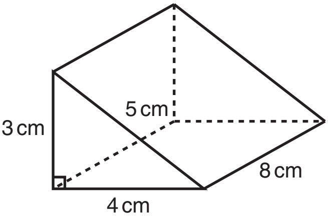

4 A pyramid has a square base with sides 8 cm long. Its other faces are isosceles triangles with a vertical height of 10 cm.
   (a) Draw a sketch of the net of the pyramid.
   (b) Find the surface area of the pyramid.

5 The radius of the ends of a solid cylinder is 4.7 cm and its length is 7.2 cm. Find its total surface area.

6 A hollow cylinder is open at one end. Its radius is 5.3 cm and its length is 9.1 cm.
   (a) Draw a net of the cylinder.
   (b) Find the total surface area of the cylinder.

18.3 Volume of a cuboid

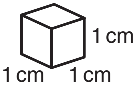

Figure 18.7

The volume of a three-dimensional shape is a measure of the amount of space it occupies.

Figure 18.7 shows a cube with edges 1 cm long. It is called a centimetre cube and its volume is 1 cubic centimetre (written 1 cm^3).

The volume of the cuboid in Figure 18.8 is the number of centimetre cubes it contains.

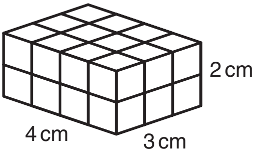

Figure 18.8

There are 4 × 3 = 12 centimetre cubes in each layer and there are two layers. So the number of centimetre cubes is 4 × 3 × 2 = 24. The volume of the cuboid is 24 cm^3.

In general,

volume of a cuboid = length × breadth × height

The volume, V, of a cuboid with length l, breadth b and height h is given by the formula $V = lbh$.

For the volumes of larger cuboids, you could use cubic metres, written m^3, and, for smaller cuboids, cubic millimetres written mm^3. You must ensure, however, that you use the same units for the length, the breadth and the height.

Example 18.3.1

The length of a cuboid is 5 m; its breadth is 3 m and its height is 40 cm. Find its volume in m^3.

The units of the measurements must be the same, so the height is 0.4 m. If V m^3 is the volume, $V = 5 \times 3 \times 0.4 = 6$; the volume is therefore 6 m^3.

Sometimes, the volume of a cuboid will be given and you will have to find one of its dimensions.

Example 18.3.2

The volume of a cuboid is 336 cm^3. Its length is 8 cm and its breadth is 7 cm. Find its height.

Let the height of the cuboid be h cm. Then $336 = 8 \times 7 \times h$, giving $56h = 336$ and $h = \frac{336}{56} = 6$. The height of the cuboid is 6 cm.

EXERCISE 18.3

1 The edges of a cube are 9 cm long. Find its volume.

2 Find the volume of cuboids with these dimensions.
(a) Length 8 cm, breadth 5 cm and height 4 cm,
(b) Length 10 m, breadth 7 m and height 3 m.

3 The length of a cuboid is 2 m; its breadth is 30 cm and its height is 20 cm. Find its volume in cm^3.

4 Find the number of cubic metres of concrete needed for a path 12 m long, 1.5 m wide and 10 cm thick.

5 The volume of a cuboid is 60 cm^3. Its length is 5 cm and its breadth is 4 cm. Find its height.

6 The volume of a cuboid is 400 cm^3. Its length is 10 cm and its height is 8 cm. Find its breadth.

7 A water tank measures 1.4 m by 80 cm by 60 cm. Find its capacity (the volume of water it can hold) in litres.

8 A room measures 10 m by 6 m by 3 m. The volume of airspace needed for one person is 5 m^3. Find the maximum number of people who may use the room at the same time.

9 How many cubic centimetres are there in a cubic metre?

10 How many packets of tea measuring 12 cm by 4 cm by 4 cm may be packed in a box measuring 48 cm by 24 cm by 16 cm?

18.4 Volume of a prism

Figure 18.9 shows a prism with a right-angled triangle as its cross-section. Its volume will be half the volume of a cuboid measuring 4 cm by 5 cm by 9 cm.

If V cm^3 is the volume, then
$V = \frac{1}{2} \times 5 \times 4 \times 9 = 90$ cm^3.

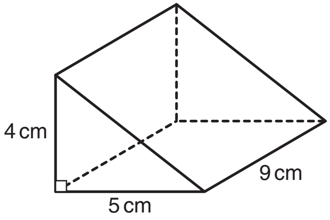

Figure 18.9

$\frac{1}{2} \times 5 \times 4$ cm^2 is the area of the end of the prism, that is, its area of cross-section. This demonstrates the result

volume of a prism = area of cross-section × length

This is sometimes written as

volume of a prism = end area × length

Example 18.4.1

The figure shows a prism with a trapezium as its cross-section. Find the volume of the prism.

Let area of cross-section be A cm^2.
Then $A = \frac{1}{2} \times (\text{sum of parallel sides}) \times (\text{distance between them})$

$= \frac{1}{2} \times (8 + 6) \times 4 = 28$

If V cm^3 is the volume, $V = 28 \times 10 = 280$
The volume is 280 cm^3.

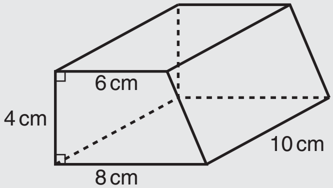

A cylinder with radius r and height h has an area of cross-section of πr^2. So you can find the volume, V, of a cylinder using this formula.

$$V = \pi r^2 \times h = \pi r^2 h$$

Example 18.4.2

The radius of the end of a cylinder is 3.8 cm and its height is 8.4 cm. Find the volume of the cylinder, correct to three significant figures.

Volume = $((\pi \times 3.8^2) \times 8.4)$ cm^3 = 381.06 cm^3
The volume is 381 cm^3, correct to three significant figures.

EXERCISE 18.4

1 Find the volume of the triangular prism shown in Figure 18.10 below.

2 Figure 18.11 below shows a prism with a trapezium as its cross-section. Find its volume.

Figure 18.10

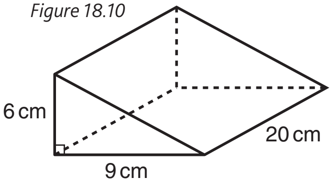
6 cm
20 cm
9 cm

Figure 18.11

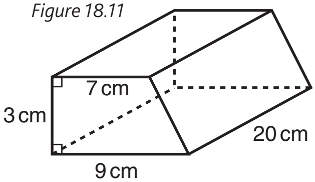
7 cm
3 cm
20 cm
9 cm

3 The radius of the end of a cylinder is 5.7 cm. Its height is 13.6 cm. Find the volume of the cylinder correct to three significant figures.

4 Find the volume of **(a)** and **(b)** prisms below.

(a)

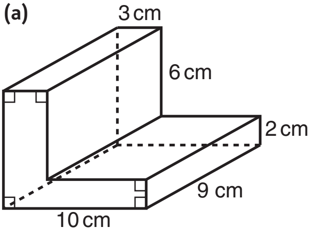
3 cm
6 cm
2 cm
9 cm
10 cm

(b)

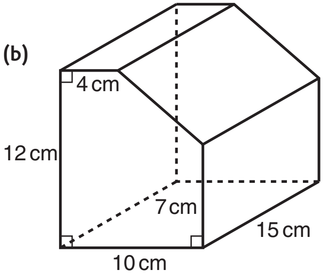
4 cm
12 cm
7 cm
15 cm
10 cm

5 A prism has a volume of 120 cm³. Its area of cross-section is 15 cm². Find the length of the prism.

6 Figure 18.12 below shows the cross-section of a swimming pool which is 25 m long. Find the volume of water in the pool when it is full.

7 Figure 18.13 shows the cross-section of a barn which is 12 m long. Find the volume of the barn.

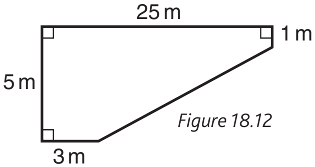

Figure 18.12

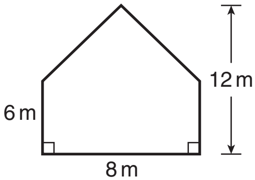

Figure 18.13

(Contd)

8 Figure 18.14 below shows the cross-section of a circular pipe. The internal and external diameters are 10 cm and 12 cm. Find the volume of material used to make a 2-metre length of pipe.

9 Figure 18.15 below shows the cross-section of a prism. It is a regular hexagon. The prism is 20 cm long. Find its volume.

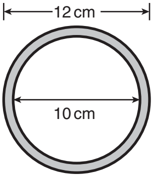

Figure 18.14

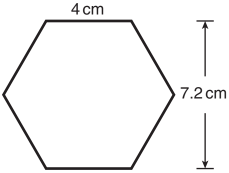

Figure 18.15

10 The volume of a cylinder is 300 cm^3. Its radius is 5 cm. Find the length of the cylinder correct to three significant figures.

18.5 Weight of a prism

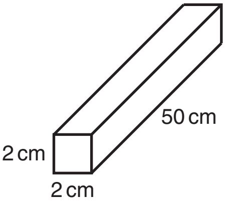

Figure 18.16

Figure 18.16 shows a steel rod. The volume of the rod is 200 cm^3. You can find the **weight** of the rod if you know the weight of 1 cubic centimetre of steel. In fact, steel weighs 7.7 grams per cm^3 (the **density** of steel).

You can then find the weight of the rod by multiplying the volume of the rod by its density.

Its weight in grams is 200 × 7.7 = 1540.
The weight is 1540 g or 1.54 kg.

Insight

There is a problem with the use of the word ‘weight’. Strictly speaking, what has been found is the **mass** of the rod, and this is the term which is used in science. However, in everyday language people use the word ‘weight’ rather than ‘mass’, so, even though it is incorrect, ‘weight’ is the word which will be used in this book.

In general,

weight = volume × density

Density may also be measured in other units, for example, kilograms per m³.

EXERCISE 18.5

1 The volume of a brick is 1400 cm³. The density of the brick is 1.6 grams per cm³. Find the weight of the brick.

2 A concrete block in the shape of a cuboid measures 10 cm by 21.5 cm by 44 cm. The density of concrete is 2.3 grams per cm³. Find the weight of the block.

3 A solid cylinder of gold has a radius of 2 cm and a height of 5 cm. The density of gold is 19.3 grams per cm³. Find the weight of the cylinder correct to three significant figures.

4 The figure shows the cross-section of a steel girder. The density of the steel in the girder is 7700 kilograms per m³. The length of the girder is 5 m. Find its weight.

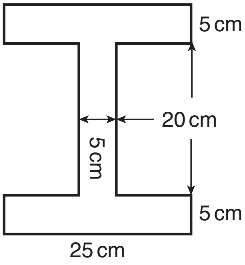

5 Each side of a square brass plate is 4 cm long. The plate is 5 mm thick. A circular hole with a diameter of 2 cm is drilled in the plate. The density of brass is 8.3 grams per cm³. Find the weight of brass which remains. Give your answer correct to the nearest gram.

THINGS TO REMEMBER

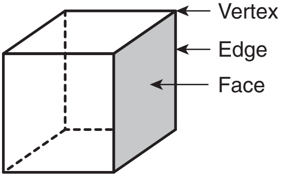

- *A cube has 6 square faces and a cuboid has 6 rectangular faces.*

- *Both a cube and cuboid have 12 edges and 8 vertices (corners).*

- *Some common solid shapes are:*

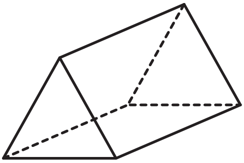
Triangular prism

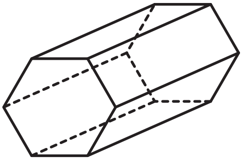
Hexagonal prism

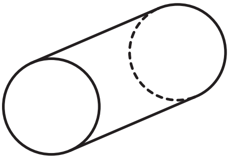
Cylinder

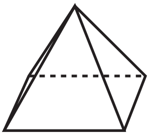
Square-based pyramid

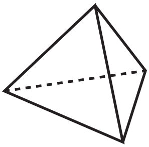
Tetrahedron

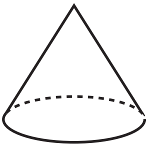
Cone

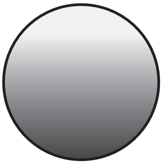
Sphere

- *A prism is a three-dimensional solid shape which has a constant cross-section.*

- *A pyramid is a solid whose base is a polygon and its other faces are triangles which meet at the apex.*

- *A polyhedron is a solid with faces that are all polygons (triangles, quadrilaterals, pentagons and so on).*

- *Euler's formula is* $F + V = E + 2$ *where* F *= number of faces,* V *= number of vertices and* E *= number of edges.*

- *A net of a solid shape is a flat shape which can be cut out and folded to make the solid shape.*

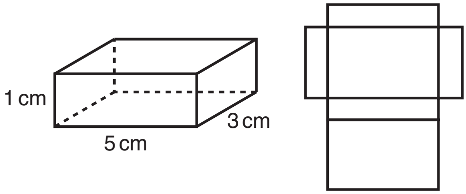

- *The surface area,* A*, of a cuboid with length* l*, breadth* b *and height* h *is* $A = 2lb + 2hl + 2bh$.

- *Surface area of a cylinder,* $S = 2\pi rh + 2\pi r^2$

- *Volume of a cuboid = length × breadth × height*

- *Volume of a prism = area of cross-section × length*

- *Weight = volume × density, where density is the weight (mass) of 1 cubic unit.*

19

Ratio and proportion

In this chapter you will learn:
- ***the meaning of 'ratio'***
- ***how to simplify ratios***
- ***how to use scales***
- ***how to use ratio***
- ***how to use direct proportion and inverse proportion***.

19.1 What is a ratio?

1:72 is the scale of a model aeroplane. 1:10 appears on a road sign. 4:3:2 is printed on a sack of fertilizer. Each of these is a ratio.

You use a ratio to compare quantities. For example, if there are 9 men and 7 women in a room, the ratio of men to women is 9:7. The order of the numbers is important, as 7:9 is the ratio of women to men. You may also write these ratios as 9 to 7 and 7 to 9 respectively but, in this chapter, colons will be used.

The numbers 9 and 7 have no common factors and so you cannot simplify the ratio 9:7. If, however, the numbers in a ratio do have a common factor, you can simplify the ratio and find its simplest form using a method similar to the one you used to find the simplest form of a fraction in Section 3.4.

To find the simplest form of the ratio 24:30, for example, you divide both numbers by 6, the highest common factor of 24 and 30. Thus, the simplest form of 24:30 is 4:5.

If the quantities in the ratio have units and the units are the same, such as 25 cm:15 cm, you remove the units, obtaining 25:15 in this case. Dividing both numbers by 5 gives its simplest form as 5:3.

If the units used for the quantities in the ratio are not the same, you must start by expressing both quantities in the same units.

Example 19.1.1

Find the simplest form of the ratio 27 mins:$1\frac{1}{2}$ hours.

Change $1\frac{1}{2}$ hours to minutes.	27 mins:90 mins
Remove the units.	27:90
Divide both numbers by 9.	3:10 is the simplest form.

To compare ratios, it is sometimes useful to express them in the form 1:*n* or *n*:1. For example, if one school has 808 students and 47 teachers, while another has 1025 students and 61 teachers, it is not obvious which school has the smaller ratio of students to teachers. For the first school, the ratio of students to teachers is 808:47. If you divide both numbers by 47, you obtain, 17.2:1, correct to one decimal place. For the second school, the ratio of students to teachers is 1025:61. If you divide both numbers by 61, you obtain 16.8:1, correct to one decimal place. So the second school has the smaller ratio of students to teachers.

Insight

It is easy to make mistakes when comparing ratios and fractions. For example, a class where there is 1 boy for every 2 girls has a ratio of boys to girls of 1:2. However, the fraction of the class that are boys is $\frac{1}{3}$ not $\frac{1}{2}$, as out of every group of 3 children 1 child will be a boy.

EXERCISE 19.1

1 Find the simplest form of each of these ratios.
(a) 8:12 **(b)** 6:3 **(c)** 20:15 **(d)** 18:30

2 Find the simplest form of each of these ratios.
(a) 24 cm:18 cm **(b)** 9 days:27 days
(c) 60 cm:1.5 m **(d)** $1\frac{1}{4}$ h:30 min
(e) 40 s:3 min **(f)** 32 mm:8 cm
(g) 1.4 kg:420 g **(h)** 720 ml:1.2 l

3 The lengths of the sides of two squares are 4 cm and 6 cm. Find, in its simplest form, the ratio of
(a) their perimeters, **(b)** their areas.

4 The lengths of the edges of two cubes are 2 cm and 4 cm. Find, in the form 1: *n*, the ratio of
(a) the total lengths of their edges,
(b) their surface areas, **(c)** their volumes.

5 Express these ratios in the form 1:*n*.
(a) 3:12 **(b)** 2:3 **(c)** 4:9 **(d)** 10:8

6 Express these ratios in the form *n*:1.
(a) 18:3 **(b)** 7:2 **(c)** 23:10 **(d)** 12:16

7 Sinton School has 723 students and 45 teachers. Tanville School has 936 students and 57 teachers. For each school, calculate, correct to one decimal place, the ratio of students to teachers in the form *n*:1. Which school has the smaller ratio of students to teachers?

19.2 Scales

If a model of a car is made to a scale of 1:50, the model is $\frac{1}{50}$ of the size of the real car – that is, you divide the real car's measurements by 50 to obtain those of the model car. Thus, if the real length of the car is 4.5 m, the length of the model car is 4.5 m ÷ 50 = 0.09 m or 9 cm.

To find the real car's measurements, you multiply the model car's measurements by 50. So, if the height of the model car is 2.8 cm, the real height of the car is 50 × 2.8 cm = 140 cm or 1.4 m.

The scales of maps are sometimes expressed as ratios. The scale tells you the relationship between distances on the map and real distances.

Example 19.2.1

The scale of a map is 1:20 000.

(a) Find the real distance represented by a distance of 6 cm on the map.

(b) What distance on the map represents a real distance of 5 km?

(a) 1 cm represents 20 000 cm.
So 6 cm represents 6 × 20 000 cm = 120 000 cm.
Convert 120 000 cm to km by dividing by 100 000.
Thus, 120 000 cm = 1.2 km.
6 cm on the map represents a real distance of 1.2 km.

(b) 5 km = 500 000 cm.
Distance on the map = $\frac{500\,000}{20\,000}$ cm = 25 cm.
25 cm on the map represents a real distance of 5 km.

Insight

Notice that a ratio has no units; so a scale of 1:20 000 means that 1 mm on the map represents 20 000 mm in real life and a distance of 1 m on the map represents a distance of 20 000 m in real life. You can use any units you like, so long as you use the same units on the map and in real life.

If you know a measurement on a model or a map and the real measurement it represents, you can find the scale being used. For example, if a length of 4 cm on a model represents a real length of 1.44 m, the ratio of the lengths is 4 cm: 1.44 m. Converting 1.44 m to centimetres, the ratio becomes 4 cm: 144 cm. You can then remove the units and divide both numbers by 4 to obtain 1:36 as the scale.

Example 19.2.2

The distance between two towns is 4.5 km. On a map, the distance between the towns is 9 cm. Find the scale of the map.

The ratio of the distances is 9 cm:4.5 km.
Convert 4.5 km to cm by multiplying 4.5 by 100 000.
The ratio is 9 cm:450 000 cm.

Remove the units and divide both sides by 9.

The scale is 1:50 000.

EXERCISE 19.2

1 The scale of a model car is 1:100.
  - **(a)** The length of the model car is 3.7 cm. Find, in metres, the real length of the car.
  - **(b)** The real width of the car is 1.61 m. Find, in centimetres, the width of the model car.

2 A plan of a house is drawn to a scale of 1:50.
  - **(a)** On the plan, the dining room is 14 cm long and 12.4 cm wide. Find, in metres, the real length and width of the dining room.
  - **(b)** The lounge is 8 m long and 6.8 m wide. Find, in centimetres, its length and its width on the plan.

3 The scale of a model aeroplane is 1:72.
  - **(a)** The length of the model aeroplane is 86 cm. Find, in metres, the real length of the aeroplane.
  - **(b)** The wingspan of the aeroplane is 25.2 m. Find, in centimetres, the wingspan of the model.

4 The scale of a map is 1:40 000.
  - **(a)** Find, in kilometres, the real distance represented by a distance of 8 cm on the map.
  - **(b)** What distance on the map represents a real distance of 8 km?

5 The distance between two towns is 7 km. On a map, the distance between the towns is 28 cm. Find, as a ratio, the scale of the map.

19.3 Using ratio

You can use ratio to solve a wide variety of problems, many of which fall into one of two categories. The first of these is when you are told the ratio of two quantities and the amount of one of the quantities and have to find the amount of remaining quantity.

For example, a type of mortar is made by mixing sand and cement in the ratio, by weight, of 4:1. This means the weight of sand is four times that of cement. So, if you have 24 kg of sand the weight of cement you have to mix with it is 24 kg divided by 4, that is, 6 kg.

Example 19.3.1

Brass is an alloy (mixture) containing copper and zinc in the ratio 5:2.

(a) What weight of copper is mixed with 26 g of zinc?

(b) What weight of brass contains 40 g of copper?

(a) The ratio is equivalent to $\frac{5}{2}$:1, so there are $\frac{5}{2}$ grams of copper to every gram of zinc.
The weight of copper is $\frac{5}{2} \times 26$ g $= 65$ g.

(b) The ratio is equivalent to 1:$\frac{2}{5}$, so there are $\frac{2}{5}$ grams of zinc to every gram of copper.
The weight of zinc is $\frac{2}{5} \times 40$ g $= 16$ g.
The weight of bronze is $(40 + 16)$ g $= 56$ g.

The second type of problem is sharing a quantity in a given ratio. For example, to share £60 in the ratio 3:2, first add the 3 and 2 to obtain 5, the number of parts into which the £60 is divided. Each part is worth £(60 ÷ 5) = £12. The larger share is 3 × £12 = £36 and the smaller share is 2 × £12 = £24. The sum of the two shares is, of course, £60.

As a fraction of the total amount, the larger share is $\frac{36}{60} = \frac{3}{5}$, in its simplest form, and the smaller share is $\frac{24}{60} = \frac{2}{5}$. You can use fractions to divide a quantity in a given ratio.

Example 19.3.2

A company has 63 employees. The ratio of men to women is 4:5.
Find **(a)** the number of men, **(b)** the number of women.

(a) $\frac{4}{4+5} = \frac{4}{9}$ of the employees are men.
So the number of men is $\frac{4}{9} \times 63 = \frac{4}{9} \times \frac{63}{1} = 28$. (See Section 3.8.)
(b) $\frac{5}{4+5} = \frac{5}{9}$ of the employees are women.
So the number of women is $\frac{5}{9} \times 63 = \frac{5}{9} \times \frac{63}{1} = 35$.

Insight

You can find the number of women by subtracting the number of men from 63. The disadvantage of using this method is that, if your first answer is wrong, so is your second answer. It is better to evaluate each share and then check that the sum of the shares is correct.

You can use both methods to divide a quantity in a ratio which has more than two numbers. For example, to share 72 kg in the ratio 3:4:5 using the first method, add the 3, 4 and 5 to obtain 12, the number of parts. Each part is worth 72 kg ÷ 12 = 6 kg, so the shares are 3 × 6 kg, 4 × 6 kg and 5 × 6 kg, that is, 18 kg, 24 kg and 30 kg. Alternatively, you can use the fractions $\frac{3}{3+4+5}$, $\frac{4}{3+4+5}$ and $\frac{5}{3+4+5}$; that is, $\frac{3}{12}$ or $\frac{1}{4}$, $\frac{4}{12}$ or $\frac{1}{3}$, and $\frac{5}{12}$.

EXERCISE 19.3

EXERCISE

1 Bronze is an alloy containing copper and tin in the ratio 3:1.
   (a) What weight of copper is mixed with 50 g of tin?
   (b) What weight of tin is mixed with 180 g of copper?
   (c) What weight of bronze contains 480 g of copper?
   (d) Work out the weight of copper and the weight of tin in 800 g of bronze.

2 An orange drink is made by mixing orange juice and water in the ratio 1:5.
   (a) How much water is mixed with 3 litres of orange juice?
   (b) How much water is in 24 litres of orange drink?

3 The ratio of Kevin's age to Michelle's age is 2:3. Kevin is 24 years old. How old is Michelle?

4 The ratio of the length of a rectangle to its width is 5:4. Its length is 55 cm. Work out its width.

5 The lengths of the sides of a triangle are in the ratio 2:3:4. The shortest side is 14 cm long. Find the lengths of the other sides.

6 Anna and Brian share £80 in the ratio 7:3. Work out the amount each of them receives.

7 Cheryl and David share $420 in the ratio 2:5. Work out the amount each of them receives.

8 The ratio of Debbie's age to her father's age is 3:5. The sum of their ages is 96 years. Work out Debbie's age.

9 The angles of a triangle are in the ratio 6:5:7. Work out the size of each of its angles.

10 The perimeter of a triangle is 240 cm. The lengths of the sides are in the ratio 3:4:5. Work out the length of the longest side.

19.4 Direct proportion

If you walk at a steady speed of 4 mph, you will cover 4 miles in 1 hour, 8 miles in 2 hours, 12 miles in 3 hours and so on. If you walk for 5 hours, your distance–time graph will look like Figure 19.1.

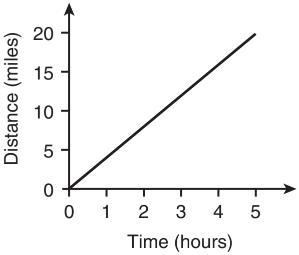

Figure 19.1

The distance you walk obviously depends on the time, but there is a more precise relationship between the two quantities. For example, you walk twice as far in 2 hours as you walk in 1 hour; you walk three times as far in 3 hours as you walk in 1 hour. When two quantities are related like this, they are said to be **directly proportional** (or proportional) to each other. If two quantities are directly proportional to each other, the graph showing the relationship between them is a straight line through the origin.

There are many examples of direct proportion in everyday life. The cost of petrol at a pump is directly proportional to the quantity of petrol you put in your tank. The cost of running an electric fire is directly proportional to the time for which it is on. When exchanging pounds for dollars, the number of dollars you get is directly proportional to the number of pounds you exchange.

There are two methods for answering questions on direct proportion. One is the **unitary method**, in which you find the value of one unit of the quantity. For example, if 8 cereal bars weigh 208 g, and you have to find the weight of 5 of these bars, the first step is to find the weight of one bar. To do this, you divide 208 g by 8, which gives 26 g. To find the weight of 5 bars, you then multiply 26 g by 5, which is 130 g. The other method is the **multiplier method**. To answer the above example using this method, you find $\frac{5}{8}$ of 208 g; that is, $\frac{5}{8} \times \frac{208}{1} = 130$ g. Alternatively, you can convert $\frac{5}{8}$ to a decimal (0.625) and work out 0.625×208.

Insight

A common mistake made when using the multiplier method is to write the fraction ‘upside down’; for example, writing $\frac{8}{5}$ instead of $\frac{5}{8}$. Always decide whether the answer should be greater or less than the original quantity: if it should be less, then you need a multiplier which is less than 1 ($\frac{5}{8}$); if it should be greater, then you need a multiplier which is greater than 1 ($\frac{8}{5}$).

Example 19.4.1

4 cm^3 of tin weighs 29.2 g. Using **(a)** the unitary method and **(b)** the multiplier method, work out the weight of 9 cm^3 of tin.

(a) Weight of 1 cm^3 of tin = 29.2 g ÷ 4 = 7.3 g
Weight of 9 cm^3 of tin = 9 × 7.3 g = 65.7 g

(b) 9 cm^3 of tin weighs more than 4 cm^3 of tin. So the multiplier is $\frac{9}{4}$ or 2.25. Weight of 9 cm^3 of tin = 2.25 × 29.2 g = 65.7 g

EXERCISE 19.4

1 A car travels 43 miles on one gallon of petrol. How far will it travel on 7 gallons?

2 £1 is worth 1.95 dollars. How many dollars is £9 worth?

3 A motorist drives at a steady speed and goes 144 miles in 3 hours.
   (a) What distance will she cover in 5 hours?
   (b) How long will it take her to cover 120 miles?

4 5 litres of paint covers an area of 60 m^2.
   (a) What area will 15 litres of paint cover?
   (b) How much paint is needed to cover an area of 240 m^2?

5 An electric fire uses 10 units of electricity in 4 hours.
   (a) How many units does it use in 3 hours?
   (b) For how long will it run on 17.5 units?

6 3 cm^3 of aluminium weighs 8.1 g. Work out the weight of 10 cm^3 of aluminium.

7 A recipe for 6 people requires 300 g of sugar and 1.5 kg of flour. Work out the weight of sugar and the weight of flour required for
   (a) 12 people, **(b)** 9 people, **(c)** 4 people.

8 20 kg of cement is mixed with 60 kg of sand to make mortar.
   (a) What weight of sand should be mixed with 16 kg of cement?
   (b) What weight of cement should be mixed with 72 kg of sand?

9 There are 330 calories in 420 g of baked beans. How many calories are there in 700 g of baked beans?

10 5000 Japanese yen is worth £21.
   (a) How many pounds is 3000 yen worth?
   (b) To the nearest yen, how many yen is £40 worth?

19.5 Inverse proportion

The table shows the time taken to travel 120 miles at various steady speeds.

Speed (mph)	10	20	30	40	50	60	80
Time (hours)	12	6	4	3	2.4	2	1.5

For example, when a speed of 10 mph is multiplied by 4, becoming 40 mph, the time of 12 hours is divided by 4. Similarly, when a time of 6 hours is divided by 3, becoming 2 hours, the speed of 20 mph is multiplied by 3. When two quantities are related like this, they are said to be **inversely proportional** to each other.

The graph in Figure 19.2 shows the information from the table. Unlike direct proportion, when two quantities are inversely proportional to each other, an increase in one of the quantities causes a decrease in the other quantity.

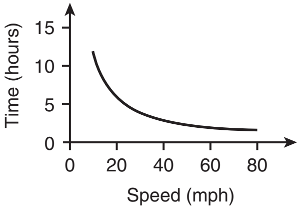

Figure 19.2

As with direct proportion, you can answer questions on inverse proportion using either the unitary method or the multiplier method.

For example, if 3 men build a wall in 8 days, and you have to find how long 4 men would take, the first step using the unitary method is to work out how long one man would take: 3 × 8 days = 24 days.

To find how long 4 men would take, you then divide 24 by 4 to obtain the answer of 6 days.

If you use the multiplier method for this question, the multiplier is $\frac{3}{4}$, not $\frac{4}{3}$, as 4 men will take less time than 3 men. Thus, 4 men take $\frac{3}{4}$ of 8 days, which is 6 days.

Insight

A common mistake is to end up with 4 men taking longer than 3 men, so whichever method you use always think carefully about whether your final answer is sensible.

Example 19.5.1

A field provides grazing for 20 cows for 6 days. Using
(a) the unitary method and **(b)** the multiplier method,
find the number of days grazing the field would provide for 15 cows.

(a) One cow could graze for 6 × 20 days = 120 days.
So 15 cows would have grazing for $\frac{120}{15} = 8$ days.

(b) The field will provide 15 cows with grazing for more than 6 days.
So the multiplier is $\frac{20}{15}$ or $\frac{4}{3}$.
So 15 cows would have grazing for $\frac{4}{3} \times 6$ days = 8 days.

EXERCISE 19.5

1 It takes 8 men 9 hours to dig a trench. How long would it take 6 men to dig the trench?
2 It takes 12 people 15 days to harvest a crop of raspberries. How long would it take 18 people?
3 Using 12 taps a tank is filled in 8 hours. How long would it take to fill the tank using 16 taps?
4 A field provides grazing for 18 sheep for 8 days. How many days grazing would it provide for 24 sheep?
5 It takes an hour to mow a lawn using a mower with blades 14 inches wide. How long would it take using a mower with blades 12 inches wide?

THINGS TO REMEMBER

- *Ratios are used to compare quantities. The order a ratio is written in is important; so if there are 5 black balls and 3 red balls in a bag the ratio of black balls to red balls is 5:3.*

- *You can simplify a ratio by dividing both numbers in the ratio by their highest common factor. For example, 25:35 simplifies to 5:7 (by dividing both 25 and 35 by 5).*

- *A ratio has no units, so if a length of wood is cut into two pieces of length 3 m and 4 m, the ratio of the lengths is 3:4 and not 3 m:4 m. However, if the units used for the quantities in the ratio are different (for example, centimetres and metres), you must start by expressing both quantities in the same units.*

- *You can compare two ratios by expressing them in the form 1:n or n:1.*

- *Ratios can be used to show the scale of a map. A scale of 1:10 000 means that 1 cm on the map represents a distance of 10 000 cm in real life.*

- *There are two main types of ratio problems:*
  - *Given the ratio of two quantities and the amount of one of the quantities then find the amount of remaining quantity.*
  - *Share a quantity in a given ratio.*

- *Two quantities are directly proportional (or proportional) to each other if the graph showing the relationship between them is a straight line through the origin. So, if one quantity is doubled, then the other quantity is also doubled.*

- *Two quantities are inversely proportional to each other if when one quantity is increased then the other quantity decreases. So, if one quantity is doubled, then the other quantity is halved.*

20

Pythagoras' theorem and trigonometry

In this chapter you will learn:

- ***how to use Pythagoras' theorem to find lengths***
- ***how to use Pythagoras' theorem to solve problems***
- ***how to use trigonometry (tangents, sines and cosines) to find lengths and angles***
- ***how to use trigonometry to solve problems.***

20.1 Pythagoras' theorem

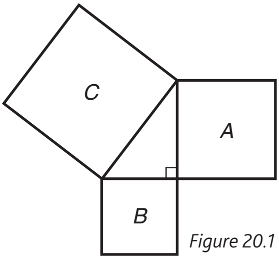

Figure 20.1

Pythagoras' theorem applies to **right-angled** triangles. The longest side of a right-angled triangle, which is opposite the right angle, is called the **hypotenuse**.

Pythagoras' theorem states that:

in a right-angled triangle, the area of the square drawn on the hypotenuse is equal to the sum of the areas of the squares drawn on the other two sides.

In Figure 20.1, this becomes

area of square C = area of square A + area of square B

This result was known to the Babylonians over 4000 years ago. In Ancient Egypt, surveyors known as 'rope stretchers' knew that a triangle with sides of length 3, 4 and 5 units has a right angle between the sides of length 3 and 4. They used a knotted rope for marking right angles on the ground.

Pythagoras (c. 580–500 BC), or one of his followers, may have been the first to prove the result. He settled in Crotona, a Greek colony in southern Italy, in about 530 BC. There he founded a mathematical and philosophical community which made major contributions in many areas of mathematics, especially number, to which they attached mystical qualities, and the harmonics of stringed instruments.

A proof of the theorem is in Section 20.3: the next section shows how to use the result.

20.2 Using Pythagoras' theorem

If you know the areas of the squares drawn on two sides of a right-angled triangle, you can use Pythagoras' theorem to find the area of the square drawn on the third side.

For example, in Figure 20.2, if the area of square A is 20 cm^2 and the area of square B is 13 cm^2, then the area of square C is $(20 + 13)$ cm^2 = 33 cm^2.

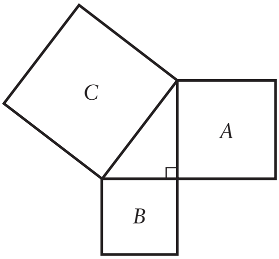

Figure 20.2

In practice, when you use Pythagoras' theorem, it will usually be to find the length of the third side of a right-angled triangle, if you know the lengths of the other two sides.

For this purpose, it is more useful if the theorem is stated in terms of the lengths of the sides. (See Figure 20.3.)

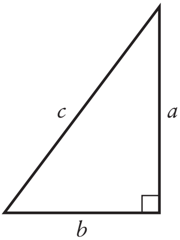

Figure 20.3

$$c^2 = a^2 + b^2$$

In Figure 20.4, the value of a is 8 and the value of b is 6, the units being centimetres. As x takes the place of c, using Pythagoras' theorem gives

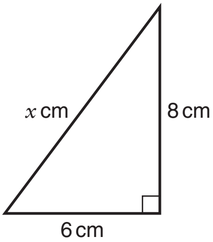

Figure 20.4

$$x^2 = 8^2 + 6^2 = 64 + 36 = 100$$

As x is positive and $x^2 = 100$, $x = 10$, so the third side is 10 cm.

As the square root of a square number is a whole number, it was easy to find x, but in other examples you must use your calculator to find square roots. You should give your result to an appropriate degree of accuracy, but three significant figures will usually suffice. Example 20.2.1 illustrates this point.

Example 20.2.1

Calculate the length marked x cm in the right-angled triangle, giving your answer correct to three significant figures.

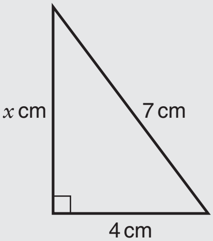

From Pythagoras' theorem,
$7^2 = x^2 + 4^2$
$49 = x^2 + 16$
$x^2 = 49 - 16 = 33$
$x = \sqrt{33} = 5.74$, correct to three significant figures.

The length is 5.74 cm, correct to three significant figures.

Insight

Mistakes with rearranging $c^2 = a^2 + b^2$ are easy to make; remember that you need to square the two given sides and then add them together to find the square of the longest side, and subtract to find the square of a shorter side.

You can split isosceles and equilateral triangles into two right-angled triangles by a line of symmetry and then use Pythagoras' theorem.

Example 20.2.2

The diagram shows an isosceles triangle with a base of 24 cm and a vertical height of 9 cm. Calculate the length of the sides marked l cm.

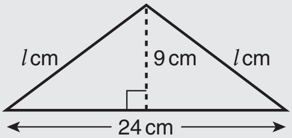

From Pythagoras' theorem in the right-angled triangle,
$l^2 = 12^2 + 9^2 = 144 + 81 = 225$ so $l = \sqrt{225} = 15$.
The length of the side is 15 cm.

Although it will not be proved in this book, Pythagoras' theorem works in reverse. If the sum of the areas of squares on two of the sides of a triangle is equal to the area of the third square, then the triangle is right-angled. You will need this property in Exercise 20.1, question 2.

EXERCISE 20.1

1 Complete the table, which refers to Figure 20.1.

Area of square *A*	Area of square *B*	Area of square *C*
12 cm^2	7 cm^2	
9 cm^2		26 cm^2
	13 cm^2	29 cm^2

2 Three squares have areas of 12 cm^2, 18 cm^2 and 30 cm^2. If these squares are arranged as in Figure 20.1, will the triangle in the centre be right-angled? Give a reason for your answer.

In the rest of this exercise, give your answers to three significant figures where necessary.

3 Calculate the length of each of the sides marked with a letter. The units are centimetres.

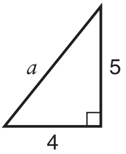

(a)

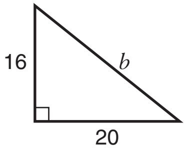

(b)

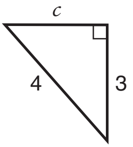

(c)

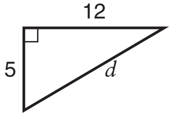

(d)

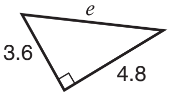

(e)

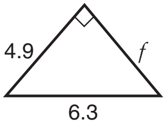

(f)

4 The sides of a square are 5 cm long. Calculate the length of a diagonal.

5 A rectangle is 10 cm long and 7 cm wide. Calculate the length of a diagonal.

6 In the isosceles triangle *ABC*, *AC* = *BC* = 8 cm. The vertical height, *CD,* of the triangle is 5 cm. Calculate the length of the base, *AB*.

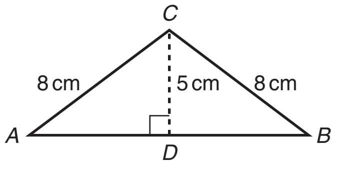

7 The sides of an equilateral triangle are 6 cm long. Calculate the vertical height of the triangle.

8 The diagram shows a trapezium *ABCD* with right angles at *B* and *C*. *AB* = 9 cm, *BC* = 4 cm and *CD* = 6 cm. Calculate the perimeter of the trapezium.

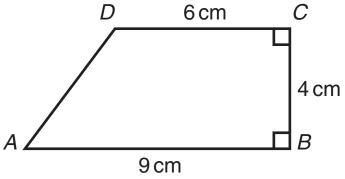

20.3 Proof of Pythagoras' theorem

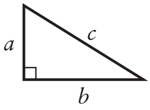

Figure 20.5

This proof of Pythagoras' theorem is based on a dissection. Figure 20.5 shows a right-angled triangle with adjacent sides labelled a, b and hypotenuse c.

The first diagram in Figure 20.6 shows four copies of the triangle arranged in the form of a square surrounding a smaller square. The area of the larger square is c^2.

If you look at the square in the centre, its side is of length x; from the dimensions in the figure, $x + a = b$. This is the same as $b - x = a$.

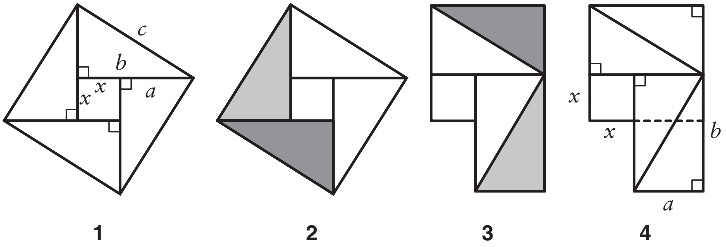

Figure 20.6

The second figure is identical to the first, but with two of the triangles shaded. The third figure shows how the two shaded triangles have been moved. The fourth figure is the same as the third, but with an additional dotted line drawn across it.

Note that the fourth figure is made up of the same shapes as the first, so its area is the same as the area of the first square, which is c^2.

Below the dotted line is a rectangular shape whose breadth is a, and whose height is $b - x$, which is the same as a, so its area is a^2.

Above the dotted line is a rectangular shape whose breadth is b, and whose height is $x + a$, which is the same as b, so its area is b^2.

So the area of the fourth figure is $a^2 + b^2$. So $a^2 + b^2 = c^2$.

20.4 Pythagoras' theorem problems

When solving problems on Pythagoras' theorem, always draw a diagram to show the information you are given. You may introduce a letter to represent the required distance but you should not use it in the final statement of the answer. Include units with your answer, if necessary.

Example 20.4.1

A ladder 6 m long leans against a wall so that the top of the ladder is 5 m from the ground. How far is the foot of the ladder from the wall? Give your answer correct to three significant figures.

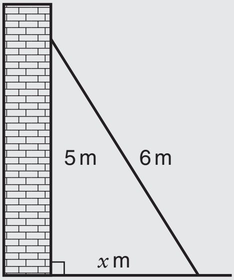

Let x m be the distance of the foot of the ladder from the wall.

Then, using Pythagoras' theorem,

$$6^2 = x^2 + 5^2.$$

So $x^2 = 36 - 25 = 11$,
giving $x = \sqrt{11} = 3.32$ m, correct to three significant figures.

The foot of the ladder is 3.32 m, correct to three significant figures, from the wall.

EXERCISE 20.2

Give your answers to three significant figures where necessary.

1 The foot of a ladder is 2 m from a wall and the top of the ladder is 6 m above the ground. Calculate the length of the ladder.

(Contd)

2 A ladder 7 m long leans against a wall so that the foot of the ladder is 2 m from the wall. How far is the top of the ladder above the ground?

3 Oxford is 48 miles South of Birmingham and 19 miles East of Birmingham. Calculate the straight-line distance between Oxford and Birmingham.

4 The lengths of the diagonals of a rhombus are 16 cm and 12 cm. Calculate the length of the sides of the rhombus. (Hint: the diagonals of a rhombus bisect each other at right angles.)

5 The sides of a rhombus are 8 cm long. One of its diagonals is 8 cm long. Calculate the length of the other diagonal.

6 The lengths of two of the sides of a right-angled triangle are 4 cm and 5 cm. Calculate the two possible lengths of the third side.

7 The diagonals of a square are 6 cm long. Calculate the length of the sides of the square.

8 *A* and *B* are points with coordinates (2, 4) and (6, 7) respectively. The units are centimetres. Calculate the distance *AB*.

20.5 Trigonometry

Trigonometry is the branch of mathematics about the relationships between the sides and angles of triangles. The word ‘trigonometry’ is derived from the Greek *trigonon* (triangle) and *metron* (measure).

Trigonometry was used in navigation, surveying and astronomy, often when the problem was to find an inaccessible distance, such as that between the Earth and the Moon. The origins of trigonometry are probably the works of the Greek astronomer and mathematician Hipparchus (c. 190–120 BC).

Thereafter, the development of trigonometry had many strands, initially Indian and Arab and, later, European. The first major Western European work on trigonometry *De trinaglis omni modis* was written by the German mathematician Regiomontanus in 1464. In the seventeenth century, through the improvements in algebraic symbolism, trigonometry became the analytic science it is today.

20.6 The tangent ratio

Figure 20.7 shows a right-angled triangle *ABC,* in which ∠A = 35°. The length, *AB*, of the side **adjacent** to the 35° angle is drawn to be 2 cm. If you measure the length of the side *BC* **opposite** the 35° angle you will find that it is very close to 1.4 cm. (Figure 20.7 is drawn accurately, so you can actually measure the length of *BC* on the diagram.)

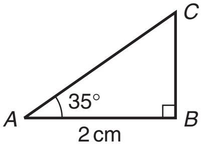

Figure 20.7

If you enlarge the triangle in Figure 20.7 to get the triangle in Figure 20.8 so that the length of the side adjacent to the 35° angle is 3 cm, and you again measure the length of the side opposite to the 35° angle, you will find that the length is very close to 2.1 cm.

If you calculate the ratio $\frac{\text{opposite side}}{\text{adjacent side}} = \frac{BC}{AB}$ in each case, you will find that, in both cases, $\frac{BC}{AB} = 0.7$.

In fact, for all right-angled triangles with an angle of 35°, no matter what the length of the adjacent side, the ratio $\frac{\text{opposite side}}{\text{adjacent side}}$ is 0.7, correct to one decimal place.

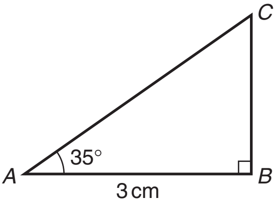

Figure 20.8

The ratio $\frac{\text{opposite side}}{\text{adjacent side}}$, usually abbreviated to $\frac{\text{opposite}}{\text{adjacent}}$, is called the **tangent** of the angle, written as **tan**. In the case above, tan 35° = 0.7.

In general, using the Greek letter θ (theta) to represent any angle,

$$\tan\theta = \frac{\text{opposite}}{\text{adjacent}}$$

(See Figure 20.9.)

Remember that the sides ‘opposite’ and ‘adjacent’ must relate to the angle you are considering.

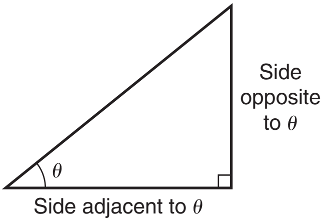

Figure 20.9

Insight

The angle θ is between two sides: the hypotenuse and the adjacent.

Remember that the hypotenuse is the longest side.

20.7 Values of the tangent

You can find the value of the tangent of an angle using your calculator. After making sure that it is in degree mode, use your calculator to check that tan 35° = 0.700 207…. You may need to consult your handbook.

You should also find that tan 45° = 1. Tangents continue to increase as the angle increases from 45° towards 90°, and tangents of angles close to 90° are very large. The ratio tan 90°, however, does not exist. If θ were 90° in Figure 20.9, you would no longer have a triangle, so tan 90° is not defined. Try finding tan 90° on your calculator; what happens?

If you know the tangent of an angle and want to find the angle, you use the **inverse tangent**. This is often written as **$\tan^{-1}$**, or sometimes as **arctan**. For example, if you need to know which angle has a tangent of 1.9, you should find that $\tan^{-1} 1.9 = 62.241\ldots°$. If you have problems, consult your calculator handbook.

EXERCISE 20.3

1 Use your calculator to find the values, correct to three significant figures, of the following.
(a) tan 40° **(b)** tan 65° **(c)** tan 2°
(d) tan 32.7° **(e)** tan 0.7° **(f)** tan 74.83°
(g) tan 89.9° **(h)** tan 89.99°

2 Use your calculator to find the values of θ, giving your answers correct to one decimal place.
(a) $\tan \theta = 0.6$ **(b)** $\tan \theta = 3$ **(c)** $\tan \theta = 0.25$
(d) $\tan \theta = 5$ **(e)** $\tan \theta = 0.08$ **(f)** $\tan \theta = 0.893$
(g) $\tan \theta = 9.372$ **(h)** $\tan \theta = 23.4$

20.8 Using tangents

You can use tangents to calculate angles and lengths in right-angled triangles.

Example 20.8.1

Calculate the size of the angle marked θ in the diagram. Give your answer correct to one decimal place.

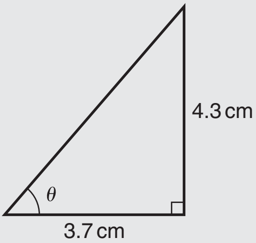

$\tan\,\theta = \frac{\text{opposite}}{\text{adjacent}} = \frac{4.3}{3.7} = 1.1621\ldots$

$\theta = \tan^{-1} 1.1621$
$\theta = 49.3°$, correct to one decimal place

The triangle may sometimes be in a position where identifying the opposite and adjacent sides to the relevant angle takes some care.

Example 20.8.2

Calculate the length of the side marked *a* cm in the diagram. Give your answer correct to three significant figures.

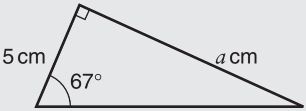

If you focus on the angle 67°, you will see that *a* cm is opposite the 67° angle and the 5 cm side is adjacent to it.

$\tan\,\theta = \frac{\text{opposite}}{\text{adjacent}} = \frac{a}{5}$, so $\tan 67° = \frac{a}{5}$

Looking at this as an equation and multiplying both sides by 5 gives

$a = 5 \tan 67° = 5 \times 2.3558\ldots = 11.779\ldots$

The length of the side is 11.8 cm, correct to three significant figures.

You can also use tangents to solve problems.

Example 20.8.3

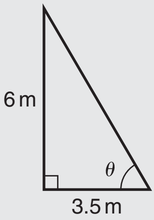

A flag pole is 6 m high. A guy rope is attached to the top of the flag pole and to a peg in the ground 3.5 m from the foot of the flag pole. Calculate the angle the guy rope makes with the ground. Give your answer correct to one decimal place.

Let θ be the angle which the guy rope makes with the ground. Then

$$\tan\theta = \frac{\text{opposite}}{\text{adjacent}} = \frac{6}{3.5}, 1.7142... = \frac{a}{5}.$$

$$\theta = \tan^{-1} 1.7142\ldots = 59.743\ldots°$$

The angle which the guy rope makes with the ground is 59.7° (correct to one decimal place).

Insight

Always draw a diagram. If a diagram is given, you may find it helps to turn the page around so that the triangle is like ◿ or ◺.

EXERCISE 20.4

1 Find the marked angles or sides, giving lengths correct to three significant figures and angles correct to one decimal place.

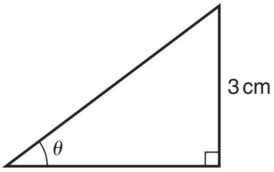

(a)

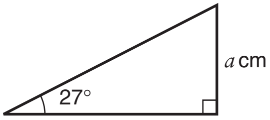

(b)

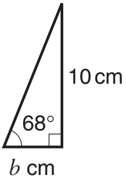

(c)

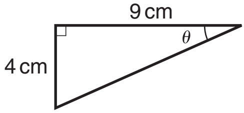

(d)

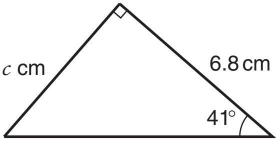

(e)

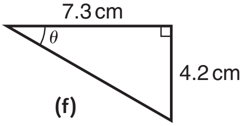

(f)

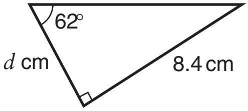

(g)

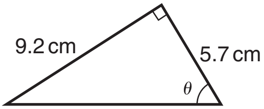

(h)

2 A ladder resting against a wall makes an angle of 63° with the ground. The foot of the ladder is 4.7 m from the wall. Calculate the height of the top of the ladder above the ground.

3 The lengths of two adjacent sides of a rectangle are 9 cm and 7 cm. Calculate the size of the angle a diagonal of the rectangle makes with a 9 cm side.

4 The bearing of York from Leeds is 058°. York is 18 miles East of Leeds. Calculate the distance York is North of Leeds.

5 A man 1.8 m tall stands 100 m away from the foot of the Eiffel Tower on level ground. To see the top, he has to look up at an angle of 71.5° above the horizontal. (This is called an angle of elevation.) Calculate the height of the Eiffel Tower.

20.9 Sine and cosine

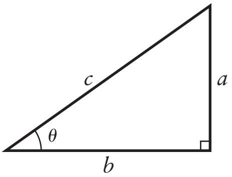

Figure 20.10

Figure 20.10 shows a right-angled triangle. The side a is opposite angle θ; b is adjacent to θ and c is the hypotenuse. You already know that, for a fixed angle θ, the ratio $\frac{a}{b}$ is constant (and equal to tan θ).

The ratio $\frac{a}{c} = \frac{\text{opposite}}{\text{hypotenuse}}$ is also constant. This ratio is the **sine** of the angle θ and is written sin θ.

Similarly, the ratio $\frac{b}{c} = \frac{\text{adjacent}}{\text{hypotenuse}}$, is constant. This ratio is the **cosine** of the angle θ and is written cos θ.

Thus:

$$\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}} = \frac{a}{c}; \quad \cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}} = \frac{b}{c}$$

Finding the value of the sine or cosine of an angle using your calculator is similar to finding the tangent of an angle. Sines and cosines, however, cannot be greater than 1. You can use $\sin^{-1}$ and $\cos^{-1}$ to find the inverse sine and the inverse cosine in the same way that you used $\tan^{-1}$ to find the inverse tangent.

You can use sines and cosines to find angles and lengths in right-angled triangles in a way similar to the one used for tangents.

For example, in Figure 20.11, the 7 cm side is opposite angle θ and the 8 cm side is the hypotenuse. The ratio relating the opposite side and the hypotenuse is sine. So,

$$\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}} = \frac{7}{8} = 0.875$$

Then using your calculator to find $\sin^{-1}$ 0.875 gives $\theta = 61.04449...°$.

Thus, correct to one decimal place, $\theta = 61.0°$.

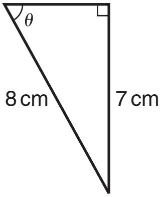

Figure 20.11

Insight

You can use the word '*sohcahtoa*' to help you remember which ratio you need. The middle letter of *soh*, *cah* and *toa* is on the numerator (top) of each ratio. So S^oh gives $\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}}$, c^ah gives $\cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}}$, and t^oa gives $\tan\theta = \frac{\text{opposite}}{\text{adjacent}}$.

For example, if you know the *a*djacent and the *h*ypotenuse then you have *cah* so use $\cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}}$.

Example 20.9.1

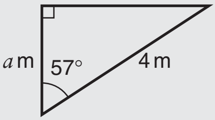

Calculate the length of the side marked ***a*** m in the figure. Give your answer correct to three significant figures.

$\cos 57° = \frac{\text{adjacent}}{\text{hypotenuse}} = \frac{a}{4}$

so $a = 4\cos 57° = 4 \times 0.5446\ldots = 2.1785\ldots$

Thus the length is 2.18 m, correct to three significant figures.

Example 20.9.2

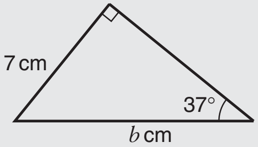

Calculate the length of the side marked ***b*** cm in the figure. Give your answer correct to three significant figures.

$\sin 37° = \frac{\text{opposite}}{\text{hypotenuse}} = \frac{7}{b}$, so $\sin 37° = \frac{7}{b}$.

Multiplying both sides by ***b*** gives ***b*** sin 37° = 7, so, dividing by sin 37°, gives

$b = \frac{7}{\sin 37°} = \frac{7}{0.6017\ldots} = 11.6314\ldots.$

Thus the length is 11.6 cm, correct to three significant figures.

You can use sines and cosines to solve problems. As with tangents, always draw a diagram.

Insight

You can use these cover-up triangles to help you rearrange the trigonometric ratios. The triangles help you work out whether you need to multiply or divide.

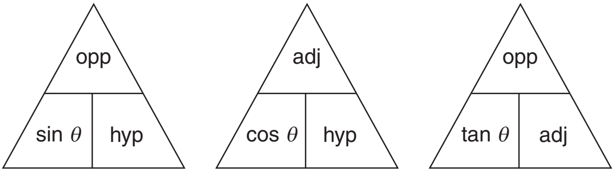

For example, should you need to find the hypotenuse and know both the opposite and the value of θ then cover 'hyp' in the first triangle. (The first triangle is the only one that has both the opposite and the hypotenuse.) So hyp $= \frac{\text{opp}}{\sin\theta}$.

Example 20.9.3

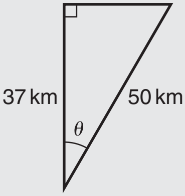

An aircraft flies 50 km in a straight line. It is then 37 km North of its starting point. Find the bearing on which it flew. Give your answer correct to the nearest degree.

In the figure, let θ be the required bearing.

$$\cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}} = \frac{37}{50} = 0.74$$

$$\theta = \cos^{-1} 0.74 = 42.26\ldots°$$

The aircraft flew on the bearing 042°, correct to the nearest degree.

EXERCISE 20.5

1 Find the marked angles or sides. Give lengths correct to three significant figures and angles correct to one decimal place.

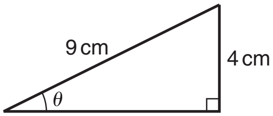

(a)

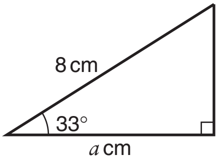

(b)

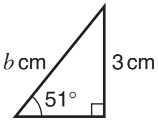

(c)

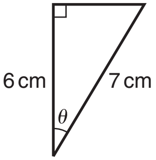

(d)

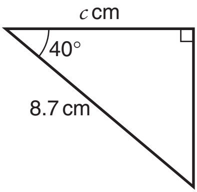

(e)

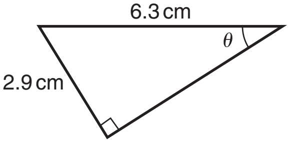

(f)

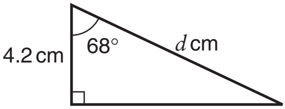

(g)

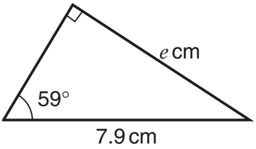

(h)

2 An aircraft flies for 400 m, climbing at an angle of 9° above the horizontal. Calculate its increase in height.

3 A boy flies a kite with a piece of string 25 m long. Calculate the size of the angle the string makes with the horizontal when the kite is 20 m above the ground.

(Contd)

4 A ladder leans against a wall. It makes an angle of 63° with the ground and the foot of the ladder is 3 m from the wall. Calculate the length of the ladder.

5 A wire rope is attached to the top of a TV mast and to a point on the ground. The length of the rope is 183 m and the height of the mast is 128 m. Calculate the angle between the rope and the ground.

6 The diagonal of a rectangle is 10 cm long. It makes an angle of 36° with one of the sides. Calculate the area of the rectangle.

7 The two equal sides of an isosceles triangle are 7.5 cm long. The length of the other side is 10 cm. Calculate the size of each of the angles of the triangle.

8 A regular pentagon is drawn inside a circle of radius 8 cm. Calculate the perimeter of the pentagon.

THINGS TO REMEMBER

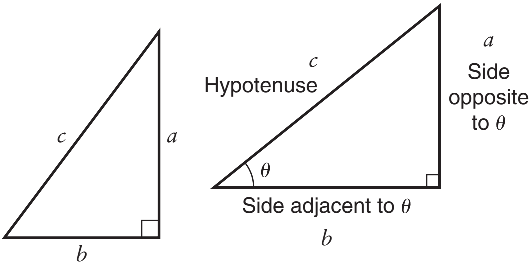
c
a
b
c
Hypotenuse
a
Side opposite to θ
θ
Side adjacent to θ
b

- *Pythagoras' theorem states that for a right-angled triangle:* $c^2 = a^2 + b^2$

- *The longest side of a right-angled triangle is called the hypotenuse. The adjacent is next to the angle θ. The opposite is opposite the angle θ.*

- $\tan\theta = \frac{\text{opposite}}{\text{adjacent}}$

  $\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}}$

  $\cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}}$

- *You can use sohcahtoa to help you remember the trig ratios.*

- *You can use these cover-up triangles to help you rearrange the trig ratios.*

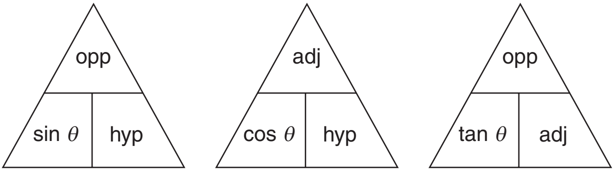
opp
sin θ
hyp
adj
cos θ
hyp
opp
tan θ
adj

21

Indices and standard form

In this chapter you will learn:

- ***how to use indices and the laws of indices***
- ***how to express a number as the product of powers of its prime factors***
- ***how to find the highest common factor (HCF) of two numbers***
- ***how to find the lowest common multiple (LCM) of two numbers***
- ***how to use standard form.***

21.1 Indices

You saw in Chapter 11 how indices are a shorthand for repeated multiplication. For example, you can write $a \times a \times a \times a \times a$ as a^5, read as 'a to the power 5', where 5 is the index or power.

In a similar way, you can use indices (the plural of index) to express powers of numbers. For example, 2^5 is shorthand for $2 \times 2 \times 2 \times 2 \times 2$, the value of which, when you carry out the multiplications, is 32.

You may be able to evaluate 2^5 using your calculator without carrying out repeated multiplication: some calculators use a y^x key; others use a key marked $\wedge$.

EXERCISE 21.1

1 Write the following in index form.
(a) 5×5 **(b)** $7 \times 7 \times 7$ **(c)** $2 \times 2 \times 2 \times 2$

2 Find the value of each of the following.
(a) 3^2 **(b)** 5^3 **(c)** 2^4 **(d)** 10^6

3 Find the values of x which satisfy the following equations.
(a) $2^x = 8$ **(b)** $3^x = 81$ **(c)** $10^x = 10\,000$

4 Find the value of each of the following.
(a) $2^3 \times 3^2$ **(b)** $2^4 \times 5^2$ **(c)** $2^2 \times 3^4 \times 5$

21.2 Laws of indices

In Chapter 11, you used the laws of indices to simplify algebraic expressions. For convenience, some of the laws are restated below.

$$a^m \times a^n = a^{m+n} \qquad a^m \div a^n = a^{m-n}$$
$$(a^m)^n = a^{mn} \qquad a^1 = a$$

You can, for example, use the first law to express $2^3 \times 2^5$ as a single power of 2. You add the indices to obtain $2^3 \times 2^5 = 2^{3+5} = 2^8$.

Example 21.2.1

Express each of the following as a single power of 3.
(a) $3^7 \div 3^2$ **(b)** $3^8 \div 3$ **(c)** $(3^5)^2$

(a) Subtracting the indices gives $3^7 \div 3^2 = 3^{7-2} = 3^5$.
(b) Write 3 as 3^1. Then $3^8 \div 3 = 3^8 \div 3^1 = 3^{8-1} = 3^7$.
(c) Multiplying the indices gives $(3^5)^2 = 3^{5\times 2} = 3^{10}$.

You cannot use the laws of indices to express expressions like $2^5 \times 3^4$ as a single power of a number, because 2 and 3 are different numbers.

EXERCISE 21.2

1 Write each of the following as a single power of the number.

(a) $2^4 \times 2^5$ **(b)** $5^8 \div 5^2$ **(c)** $(3^4)^2$

(d) $10^3 \times 10$ **(e)** $5^9 \div 5$ **(f)** $2^5 \times 2^3 \times 2^4$

(g) $3^7 \times 3^4 \div 3^5$ **(h)** $\frac{10^4 \times 10^8}{10^7}$

21.3 Prime factors

The factors of 20 are 1, 2, 4, 5, 10 and 20. Of these factors, 2 and 5 are **prime** numbers, so 2 and 5 are called the **prime factors** of 20.

Every number can be expressed as the product of powers of its prime factors. You can express 20, for example, as $2^2 \times 5$. In a simple case like this, you might be able to do the working in your head but in more complicated cases it is easier to use repeated division.

The first prime number is 2, which is a factor of 20. You can divide 20 exactly by 2 twice. $20 \div 2 = 10$ and $10 \div 2 = 5$.

The second prime number is 3, but 3 is not a factor of 5.

The third prime number is 5, which is a factor of 5.

$5 \div 5 = 1$, which completes the dividing.

You can set out the working like this.

2	20
2	10
5	5
	1

So $20 = 2 \times 2 \times 5 = 2^2 \times 5$.

The next example illustrates this method with a larger number.

Example 21.3.1

Express 360 as the product of powers of its prime factors.

2	360
2	180
2	90
3	45
3	15
5	5
	1

So $360 = 2 \times 2 \times 2 \times 3 \times 3 \times 5 = 2^3 \times 3^2 \times 5$

Insight

All whole numbers can be written as a product of prime factors. It is a good idea to evaluate your product of primes to check it actually does equal the original number.

EXERCISE 21.3

1 Write these numbers as products of powers of prime factors.
(a) 12 **(b)** 18 **(c)** 36 **(d)** 45 **(e)** 48
(f) 60 **(g)** 108 **(h)** 225 **(i)** 280 **(j)** 2200

21.4 Highest common factor

The **highest common factor** (HCF) of two numbers is the largest number which will divide exactly into both of them.

The factors of 24 are 1, 2, 3, 4, 6, 8, 12, 24.
The factors of 36 are 1, 2, 3, 4, 6, 9, 12, 18, 36.
Six numbers divide exactly into both 24 and 36: 1, 2, 3, 4, 6 and 12.

The highest factor 24 and 36 have in common is 12, so 12 is the highest common factor of 24 and 36.

In this case, the lists of factors are short and you may be able to find the HCF mentally. However, to find the HCF of larger numbers such as 540 and 1008, listing the factors would be tedious, so it would be useful to have a way of finding the HCF more directly.

In prime factor form, $24 = 2^3 \times 3$, $36 = 2^2 \times 3^2$. If you look at the powers of 2, you see that 2^2 is the highest power of 2 which divides both 24 and 36. Similarly, 3 is the highest power of 3 which divides both of them. So $2^2 \times 3 = 12$ is the largest number which divides both 24 and 36.

In general, the HCF of two numbers is the product of the **lowest** powers of the prime factors they have in common. You can use this method to find the HCF of more than two numbers.

Example 21.4.1

Find in index form the HCF of $2^3 \times 3^2 \times 5^4$ and $2^4 \times 5^2 \times 7^3$.

The common prime factors are 2 and 5.

The lowest power of 2 which divides both is 2^3 and the lowest power of 5 which divides both is 5^2, so the HCF is $2^3 \times 5^2$.

Example 21.4.2

Find the HCF of 540 and 1008.

In prime factor form, $540 = 2^2 \times 3^3 \times 5$ and $1008 = 2^4 \times 3^2 \times 7$.

The common prime factors are 2 and 3.

The lowest power of 2 which divides both is 2^2 and the lowest power of 3 which divides both is 3^2. So the HCF is $2^2 \times 3^2 = 36$.

Insight

Always check that your answer is actually a factor of both numbers.

EXERCISE 21.4

1 Write down the highest common factor of the following pairs of numbers.
(a) 6 and 9 **(b)** 14 and 21 **(c)** 18 and 24 **(d)** 20 and 30

2 Write down in index form the highest common factor of the following sets of numbers.
(a) $2^3 \times 3^2, 2^2 \times 3^4$ **(b)** $2^4 \times 3^3 \times 5^2, 2^3 \times 3^5 \times 7^2$
(c) $2 \times 3^2 \times 5^3, 2^3 \times 5^2 \times 11$ **(d)** $2^4 \times 5, 2^3 \times 5^4, 2^5 \times 5^2 \times 7$

3 Find the highest common factor of the following sets of numbers.
(a) 120 and 168 **(b)** 300 and 420 **(c)** 120, 280 and 360

21.5 Lowest common multiple

The **lowest common multiple** (LCM) of two numbers is the lowest number which is a multiple of the two numbers.

The multiples of 12 are 12, 24, 36, 48, 60, 72, 84, 96, 108, 120, …

The multiples of 15 are 15, 30, 45, 60, 75, 90, 105, 120, 135, …

The multiples which 12 and 15 have in common are 60, 120, …

The lowest multiple 12 and 15 have in common is 60, so 60 is the lowest common multiple of 12 and 15.

In prime factor form, $12 = 2^2 \times 3$, $15 = 3 \times 5$. The highest powers of 2, 3 and 5 respectively are 2^2, 3 and 5, so both 12 and 15 divide into $2^2 \times 3 \times 5 = 60$. Thus, 60 is the LCM of the two numbers.

In general, the LCM of two numbers is the product of the highest powers of all their prime factors, not just the ones they have in common.

Example 21.5.1

Write down in index form the lowest common multiple of $2^3 \times 5^2 \times 7$ and $2^2 \times 3^4 \times 5^3$.

The highest power of each of the prime factors is 2^3, 3^4, 5^3 and 7.
The LCM is $2^3 \times 3^4 \times 5^3 \times 7$.

Example 21.5.2

Find the LCM of 45 and 60.

In prime factor form, $45 = 3^2 \times 5$ and $60 = 2^2 \times 3 \times 5$.

The highest power of each of the prime factors is 2^2, 3^2 and 5.

In index form, the LCM is $2^2 \times 3^2 \times 5 = 180$.

EXERCISE

EXERCISE 21.5

1 Write down the LCM of the following sets of numbers.
   (a) 3 and 5   **(b)** 4 and 6   **(c)** 8 and 12   **(d)** 2, 3 and 4

2 Write down in index form the LCM of the following numbers.
   (a) $2^3 \times 3^4$, $2^2 \times 3^5$ **(b)** $2^4 \times 3^2 \times 5$, $2^5 \times 3 \times 7^3$
   (c) $2^3 \times 5^2$, $2^6 \times 5 \times 11^2$ **(d)** $2^4 \times 3^3$, $3^2 \times 5^3$, $2^3 \times 5 \times 7$

3 Find the lowest common multiple of the following numbers.
   (a) 40 and 50   **(b)** 24 and 30   **(c)** 18, 20 and 27

21.6 Standard form – large numbers

In this section, you will be using powers of 10. Here are some of them.

$10^4 = 10 \times 10 \times 10 \times 10 = 10\,000$
$10^3 = 10 \times 10 \times 10 = 1000$
$10^2 = 10 \times 10 = 100$
$10^1 = 10 = 10$

Note that the value of 10^4 is 1 followed by 4 zeros; the value of 10^3 is 1 followed by 3 zeros and so on. You can extend this to higher powers of 10, so 10^7 is 1 followed by 7 zeros, that is, $10^7 = 10\ 000\ 000$.

The value of 2×10^3 is $2 \times 1000 = 2000$. Again, the index is the same as the number of zeros.

The number 2×10^3 is an example of a number written in **standard form**, or **standard index form**.

Standard form consists of a number between 1 and 10 multiplied by a power of 10. The number between 1 and 10 need not be a whole number; the number 7.1×10^4 is in standard form. The value of 7.1×10^4 is $7.1 \times 10\ 000$, which is 71 000.

To write a number in standard form, write down the number between 1 and 10 and then decide what power of 10 it has to be multiplied by. With 657 000 the number between 1 and 10 is 6.57, which has to be multiplied by 100 000, which is 10^5. So $657\ 000 = 6.57 \times 10^5$.

Standard form is often used to write very large numbers. For example, the distance from Earth to the star Sirius is 50 000 000 000 000 miles. It is much more convenient to express this distance in standard form as 5×10^{13} miles, rather than count all those zeros.

Example 21.6.1

Write in standard form

(a) 5730 **(b)** 600 000 **(c)** 2.8 million

(a) $5730 = 5.73 \times 1000$, so $5730 = 5.73 \times 10^3$
(b) $600\ 000 = 6 \times 100\ 000 = 6 \times 10^5$
(c) 2.8 million $= 2\ 800\ 000 = 2.8 \times 1\ 000\ 000 = 2.8 \times 10^6$

Example 21.6.2

Write 6.47×10^4 as an ordinary number.

$6.47 \times 10^4 = 6.47 \times 10\,000 = 64\,700$

Try using your calculator to check the answers to the last two examples. You may have to use the manual to find out how to do it. The number of figures on your calculator display will limit the size of the number it can show as an ordinary number.

Insight

We are so used to hearing about very large numbers that it is easy to not appreciate exactly how large numbers like a billion or a trillion really are. To count to a million at a rate of one number per second would take about eleven and a half days; to count to a billion at the same rate would take over 31 years 8 months, and to count to a trillion would take nearly 31 710 years.

EXERCISE

EXERCISE 21.6

1 Write the following numbers in standard form.
(a) 900 **(b)** 3000 **(c)** 700 000 **(d)** 40 000 000
(e) 32 000 **(f)** 170 000 **(g)** 4520 **(h)** 275 000 000
(i) 52 million **(j)** 4.6 million

2 Write the following numbers as ordinary numbers.
(a) 5×10^2 **(b)** 8×10^4 **(c)** 9×10^7 **(d)** 8.7×10^5
(e) 9.1×10^3 **(f)** 2.3×10^6 **(g)** 1.97×10^4 **(h)** 4.21×10^{10}

3 Light travels 300 000 000 metres in one second. Write this number in standard form.

4 The age of the Sun is approximately 7500 million years. Write this number in standard form.

5 The average distance of the Earth from the Sun is 9.37×10^7 miles. Write this as an ordinary number.

21.7 Standard form – small numbers

Here are some numbers expressed in standard form.

$$\begin{aligned} 37\,000 &= 3.7 \times 10^4 \\ 3700 &= 3.7 \times 10^3 \\ 370 &= 3.7 \times 10^2 \\ 37 &= 3.7 \times 10^1 \end{aligned}$$

You can obtain each number on the left by dividing the one above it by 10 and each index is one less than the one above it. You can use these relationships to continue the pattern.

$$\begin{aligned} 3.7 &= 3.7 \times 10^0 \\ 0.37 &= 3.7 \times 10^{-1} \\ 0.037 &= 3.7 \times 10^{-2} \\ 0.0037 &= 3.7 \times 10^{-3} \end{aligned}$$

This shows that, when you express a number less than 1 in standard form, the power of 10 is negative. One way of finding the index is to count the number of zeros before the first significant figure and then make this number negative. For example, 0.000 72 has 4 zeros before the first significant figure, 7, and so the index is –4. So in standard form, 0.000 72 is 7.2×10^{-4}.

Example 21.7.1

Write in standard form **(a)** 0.000 003 **(b)** 0.0912

(a) There are 6 zeros before the first significant figure.
So the index is –6. Then $0.000\,003 = 3 \times 10^{-6}$
(b) There are 2 zeros before the first significant figure.
So the index is –2. Then $0.0912 = 9.12 \times 10^{-2}$

Example 21.7.2

Write 1.9×10^{-5} as an ordinary number.

The index, –5, tells you that there are 5 zeros before the 1.

Thus $1.9 \times 10^{-5} = 0.000\,019$.

Try using your calculator to check the answers to the last two examples. You may need to consult your manual.

EXERCISE 21.7

1 Write the following numbers in standard form.
(a) 0.07 **(b)** 0.0004
(c) 0.3 **(d)** 0.000 000 09
(e) 0.0084 **(f)** 0.000 029
(g) 0.007 91 **(h)** 0.739

2 Write the following numbers in standard form as ordinary numbers.
(a) 4×10^{-3} **(b)** 8×10^{-5}
(c) 9×10^{-1} **(d)** 3×10^{-9}
(e) 1.3×10^{-2} **(f)** 9.7×10^{-4}
(g) 4.73×10^{-6} **(h)** 5.08×10^{-7}

3 The wavelength of visible light is 0.0005 mm. Write this number in standard form.

4 The coefficient of linear expansion of copper is 0.000 016 9 per degree C. Write this number in standard form.

5 The lifetime in seconds of Mu Minus, an atomic particle, is 2.2×10^{-6}. Write this as an ordinary number.

21.8 Standard form calculations

You can do simple standard form calculations mentally with the laws of indices. For more awkward calculations, use your calculator.

Example 21.8.1

Work out $(3 \times 10^6) \times (5 \times 10^4)$. Give your answer in standard form.

You can change the order of the multiplication without affecting it.

So $(3 \times 10^6) \times (5 \times 10^4) = 3 \times 5 \times 10^6 \times 10^4 = 15 \times 10^{10}$

But this is not in standard form; write 15 in standard form as 1.5×10^1.

Then $15 \times 10^{10} = 1.5 \times 10^1 \times 10^{10} = 1.5 \times 10^{11}$, so

$(3 \times 10^6) \times (5 \times 10^4) = 1.5 \times 10^{11}$

Example 21.8.2

Work out $\frac{8 \times 10^7}{2 \times 10^4}$.

You can write $\frac{8 \times 10^7}{2 \times 10^4}$ as $\frac{8 \times 10^7}{2 \times 10^4}$.

Then $\frac{8 \times 10^7}{2 \times 10^4} = \frac{8 \times 10^7}{2 \times 10^4} = 4 \times 10^3$.

Example 21.8.3

Work out $(5.4 \times 10^4) + (3.6 \times 10^3)$. Give your answer in standard form.

In this case, you have to evaluate each of the terms first.

$5.4 \times 10^4 = 54\,000$ and $3.6 \times 10^3 = 3600$

Adding the numbers gives $54\,000 + 3600 = 57\,600$.

In standard form, $57\,600 = 5.76 \times 10^4$.

So $(5.4 \times 10^4) + (3.6 \times 10^3) = 5.76 \times 10^4$

EXERCISE 21.8

1 Work out the following without using a calculator. Give your answers in standard form.

(a) $(3 \times 10^5) \times (2 \times 10^6)$ **(b)** $(6 \times 10^3) \times (4 \times 10^4)$

(c) $\frac{9 \times 10^8}{3 \times 10^5}$ **(d)** $\frac{4 \times 10^6}{5 \times 10^2}$

(e) $(4 \times 10^4) + (3 \times 10^3)$ **(f)** $(5 \times 10^6) - (7 \times 10^4)$

(g) $(3.1 \times 10^5) + (4.7 \times 10^3)$ **(h)** $(6.3 \times 10^3) - (3.8 \times 10^2)$

2 Using a calculator, work out the following. Give your answers in standard form, correct to three significant figures.

(a) $(6.71 \times 10^3) \times (4.29 \times 10^6)$

(b) $\frac{9.7 \times 10^5}{3.8 \times 10^3}$

(c) $(4.82 \times 10^4) + (8.73 \times 10^3)$

(d) $(9.82 \times 10^7) - (6.31 \times 10^6)$

3 A computer performs 400 million calculations per second. Work out the number of calculations it performs in 2 hours. Give your answer in standard form.

4 The distance of the Earth from the Sun is 1.5×10^8 km. The distance of Uranus from the Sun is 2.8×10^9. Approximately how many times further from the Sun is Uranus than the Earth?

5 Astronomical distances are often measured in light years.
1 light year = 9.46×10^{12} km

(a) The centre of the Milky Way is 2.6×10^4 light years from Earth. How many kilometres is this? Give your answer in standard form, correct to two significant figures.

(b) The Andromeda Galaxy is 1.4×10^{19} km from Earth. Express this in light years. Give your answer in standard form, correct to two significant figures.

THINGS TO REMEMBER

- *The laws of indices are:*

  $a^m \times a^n = a^{m+n}$ $\quad$ $a^m \div a^n = a^{m-n}$

  $(a^m)^n = a^{mn}$ $\quad$ $a^1 = a$

- *The factors of a number are numbers which divide exactly into that number. For example, the factors of 70 are 1, 2, 5, 7, 10, 14, 35 and 70.*

- *Prime factors are the factors of a number which are prime. Every whole number can be written as a product of prime factors. For example, the prime factors of 70 are 2, 5 and 7 and* $70 = 2 \times 5 \times 7$.

- *The highest common factor (HCF) of two numbers is the largest number which will divide exactly into both of them.*

- *The lowest common multiple (LCM) of two numbers is the lowest number which is a multiple of the two numbers.*

- *Standard form is used to write very large or very small numbers. It consists of a number between 1 and 10 multiplied by a power of 10.*

- *To write a number which is greater than 1 in standard form: count how many digits there are after the first significant figure to give you the power of 10. For example,* $906\,000 = 9.06 \times 10^5$.

- *To write a number which is less than 1 in standard form: count the number of zeros before the first significant figure and then make this number negative to give you the power of 10. For example,* $0.000\,0906 = 9.06 \times 10^{-5}$.

- *You can do simple standard form calculations mentally using the laws of indices.*

22

Statistics 2

In this chapter you will learn:

- ***how to find the mode, median, mean and range from lists and tables***
- ***how to solve problems involving the mode, median, mean or range***
- ***how to find an estimate for the mean from grouped frequency tables***
- ***how to draw and use cumulative frequency graphs.***

22.1 Averages

Chapter 6 dealt with the collection and organization of data. This chapter deals with the analysis of data. One way in which you can do this is to find an average, usually a single value, which is typical or representative of the set of data. The three averages most commonly used are the mode, the median and the arithmetic mean, usually just called the mean.

22.2 The mode

The **mode** is the item which occurs most often. In everyday speech, 'mode' means 'fashion'; you can think of the mode as the most fashionable or popular item. Thus, if the shoe sizes of seven men are 8, 6, 9, 7, 9, 7 and 9, the mode, or modal size, is 9, because it appears three times and no other size appears more than twice.

If the above information about shoe sizes is shown on a bar chart (see Figure 22.1), the mode is the shoe size with the highest bar.

To find the mode of longer lists of numbers, you might find it helpful to use tally marks and draw up a frequency table.

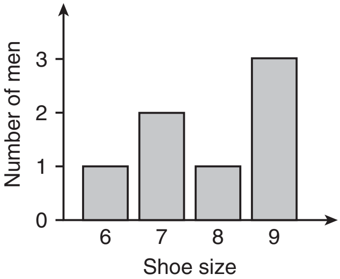

Figure 22.1

The mode is not always a number; it could, for example, be a colour. The mode has the advantages of being easy to obtain and being unaffected by extreme values but it is possible for data to have more than one mode or, if all the items are all different, no mode.

Example 22.2.1

Find the mode of

(a) 4, 6, 7, 5, 4, 7, 6, 8, 6, 8 **(b)** 2, 3, 5, 3, 1, 6, 5

(a) 6 occurs three times. 4, 7 and 8 each occur twice. 5 occurs once, so the mode is 6.

(b) 3 and 5 each occur twice. 1, 2 and 6 each occur once. So the modes are 3 and 5. (The data is said to be **bi-modal**.)

22.3 The median

The **median** is the middle value, after the data have been arranged in order of size. For example, to find the median of 7, 3, 6, 4, 2, write the numbers in order of size (2, 3, 4, 6, 7) and find the middle number, which is 4. So the median is 4.

For five values, the median is the third value in the data, after they have been arranged in order of size. If there were seven values in the data, the median would be the fourth value and so on. In general, after you have put the values in order of size, you can find the position of the median by adding one to the number of values and dividing the result by 2. For example, if there are 17 values, find $\frac{1}{2}(17 + 1)$; so the median is the ninth value.

A problem arises with the median of an even number of values, for example 7, 3, 8, 1, 9, 2. When you write the numbers in order of size (1, 2, 3, 7, 8, 9), there are two numbers in the middle, 3 and 7. The median is the number halfway between 3 and 7, that is, 5. Notice that $5 = \frac{1}{2}(3 + 7)$ and that 5 did not appear in the original data.

In general, after putting the values in order of size, the median of an odd number of values is simply the middle value and the median of an even number of values is halfway between the middle two values.

Example 22.3.1

Find the median of

(a) 2, 4, 5, 1, 4, 1, 3, 0, 2 **(b)** 13, 4, 12, 27, 3, 7

(a) Write the numbers 0, 1, 1, 2, 2, 3, 4, 4, 5 in order of size. The middle number is 2, so the median is 2.

(b) Write the numbers 3, 4, 7, 12, 13, 27 in order of size. There are two numbers in the middle, 7 and 12. Adding 7 and 12 and dividing the result by 2 gives $\frac{1}{2}(7 + 12) = \frac{1}{2} \times 19 = 9\frac{1}{2}$, so the median is $9\frac{1}{2}$.

From part (b) the median, like the mode, is not affected by extreme values.

22.4 The mean

The **mean** is the sum of the values divided by the number of values, that is,

$$\text{mean} = \frac{\text{sum of values}}{\text{number of values}}$$

For example, to find the mean of 9, 12, 2 and 5, add the numbers $(9 + 12 + 2 + 5 = 28)$ and then divide 28 by 4, the number of values. Thus the mean is $\frac{9+12+2+5}{4} = \frac{28}{4} = 7$. The mean, therefore, need not be one of the numbers in the original data. Also, it need not be a whole number, even if all the values in the data are whole numbers.

Example 22.4.1

Ten boys were asked how many sisters each of them had. The numbers were 3, 2, 0, 2, 1, 4, 0, 1, 2, 1. Work out the mean number of sisters.

$$\text{Mean} = \frac{3+2+0+2+1+4+0+1+2+1}{10} = \frac{16}{10} = 1.6$$

Of course, you cannot have 0.6 of a sister, but it is still correct to give this answer as 1.6.

The 'average' referred to in everyday speech is usually the mean. The term has already been used in this book; *average number of hours of sunshine* and *average monthly rainfall* occurred in Chapter 6 and *average speed* appeared in Chapter 8. Each of these is a mean value.

An advantage of the mean is that it makes use of all the data; a disadvantage is that it may be affected a great deal by extreme values.

Example 22.4.2

The rainfall, in millimetres, each day during a week is 0, 1, 2, 17, 1, 0, 0. Work out the mean daily rainfall and comment on your answer.

(Contd)

$$\text{Mean} = \frac{0+1+2+17+1+0+0}{7} = \frac{21}{7} = 3$$

The mean rainfall is 3 mm.

The mean is not a very representative value, as, on six of the seven days, the rainfall is less than 3 mm. The median, 1 mm, might be a more appropriate representative value for certain purposes.

At the start of this section, you saw that, if the mean of four numbers is 7, their sum is 28, which is 7 × 4. In general,

sum of values = mean × number of values.

Example 22.4.3

The mean number of points scored by a rugby team in their first four matches was 17. After five matches, their mean was 19 points. How many points did they score in the fifth match?

Number of points scored in first four matches = 17 × 4 = 68
Number of points scored in all five matches = 19 × 5 = 95
Number of points scored in the fifth match = 95 – 68 = 27

You often need to decide on which average would be the most appropriate one to use – this table can help you decide:

Average	*Advantages*
Mode	Easy to spot – especially when data has been ordered. Can be used for non-numerical data; for example, the most popular colour is blue.
Median	Useful when the data set has a few values which are much higher or lower than most of the other values. Useful for making comparison; for example half of the class got 60% or above.
Mean	Takes into account all of the data values.

EXERCISE 22.1

1 Find the mode, median and mean of the following sets of numbers.
   (a) 7, 7, 8, 11, 12 **(b)** 9, 5, 13, 5
   (c) 9, 0, 8, 9, 4 **(d)** 7, 1, 6, 7, 6, 1, 7
   (e) 12, 5, 7, 17, 7, 15 **(f)** 5, –1, –4, 0, 5

2 The number of children in each of ten families was 2, 3, 1, 3, 4, 2, 1, 3, 2, 3. Find the mode, median and mean.

3 The marks of five candidates in an examination were 76, 84, 72, 84 and 59. Find the mode, median and mean.

4 The numbers of goals scored by a soccer team in the first eight matches of the season were 3, 7, 1, 0, 2, 1, 5 and 1. Find the mode, median and mean.

5 The minimum daily temperatures, in °C, during a winter week were –1, –8, 0, –6, –1, –2 and –3. Find the mode, median and mean.

6 The lengths of six films are 2 h 4 min, 2 h 14 min, 1 h 43 min, 2 h 14 min, 2 h 29 min and 1 h 52 min. Find, in hours and minutes, the mode, median and mean.

7 The ages of four children are 3 years 7 months, 5 years 4 months, 6 years 8 months and 8 years 9 months. Find, in years and months, their mean age.

8 **(a)** A man drives 252 miles in 6 hours. Find his average speed.
   (b) On the next day, he drives 105 miles at the same average speed. How long will this journey take him?

9 The mean of ten numbers is 8.3. Work out their sum.

10 The mean of 19, 37, 17, x, 29 and 15 is 24. Find the value of x.

11 The mode of five numbers is 1, the median is 5 and the mean is 4. Find the numbers.

12 The mean weight of seven people is 61 kg. Another person joins the group and the mean falls to 59 kg. Work out the weight of the eighth person.

13 The mean height of the 15 boys in a class is 162 cm and the mean height of the 10 girls is 152 cm. Work out the mean height of the 25 pupils.

14 A cyclist travels for 3 hours at 20 mph and then for 2 hours at 15 mph. Work out her average speed for the whole journey.

22.5 The range

In addition to finding averages, it is useful to know how spread out data are. For example, if in an examination six boys scored 54, 55, 58, 62, 62, 63 and seven girls scored 32, 50, 54, 60, 62, 62, 93, the boys' mode (62), median (60) and mean (59) are the same as those of the girls but the spread, or dispersion, of each set of marks is quite different.

The **range** is one measure of the spread of data. The range is the difference between the highest value and the lowest value, that is,

range = highest value – lowest value

Thus, the range of the boys' marks is 63 – 54 = 9 and the range of the girls' marks is 93 – 32 = 61. Notice that, in mathematics, the range is a single value, even though, in everyday speech, you might describe the boys' range, for example, as 54 to 63.

Insight

When the range is low it means that the data values are close together; however, it doesn't give you any information about the actual data values. So, in the above example you can tell that the boys are more consistent as their range of marks is lower; but the range alone doesn't tell you whether the boys have consistently good marks or consistently bad marks.

The range has the advantage of being simple to work out but it is obviously affected by extreme values which are very different from other values in the data.

Example 22.5.1

In Mathsville, the number of hours of sunshine each day last week was 6, 9, 4, 3, 11, 6 and 7. Find the range.

The highest value is 11 and the lowest value is 3.

Highest value – lowest value = 11 – 3 = 8, so the range is 8 hours.

EXERCISE 22.2

1 Find the range of the following sets of numbers.
(a) 2, 4, 7, 8, 9 **(b)** 24, 17, 42, 29, 37, 19
(c) 3, –4, 0, –2, 1, –3

2 The range of a set of data is 9. The lowest value is 12. Work out the highest value.

3 In a science examination, six students scored 42, 76, 51, 64, 48, 59 and, in a history examination, they scored 86, 24, 93, 17, 41, 65. Find the range of their marks for each examination.

4 One day last winter, I measured the temperature every four hours. My readings, in °C, were –6, –3, 0, 3, 1, –2. Find the range.

5 The median of five numbers is 8; their mode is 3; their range is 9; their mean is 7. Find the numbers.

22.6 Frequency tables

You can find averages and the range from a frequency table. For example, the table shows information about the number of goals scored by 20 soccer teams.

Number of goals	0	1	2	3	4	5
Number of teams	2	6	5	4	1	2

The mode, or modal number of goals, is 1, because more teams, 6, scored 1 goal than any other number of goals.

To find the median, you could first write down the numbers of goals from the table 0, 0, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 5, 5.

The median lies halfway between the tenth and eleventh values (see Section 22.3) and, as these are both 2, the median is 2.

You can, however, find the median using the table without writing out a list. The first two teams scored no goals, so the second team scored no goals. The next six teams scored 1 goal, so the eighth team scored 1 goal. The next five teams scored 2 goals, so the thirteenth team scored 2 goals. The tenth and eleventh teams must each have scored 2 goals, so the median is 2.

The numbers 2, 8, 13 are called the **running total** or **cumulative frequency**. The next two are 17 (13 + 4) and 18 (17 + 1). The last one is 20, the total number of teams.

To work out the mean, you need to find the total number of goals and divide it by the number of teams, 20. Two teams scored no goals. Six teams each with 1 goal scored a total of 6 goals (1 × 6). Five teams each with 2 goals scored a total of 10 goals (2 × 5) and so on.

Thus, the total number of goals

$$= (0 \times 2) + (1 \times 6) + (2 \times 5) + (3 \times 4) + (4 \times 1) + (5 \times 2)$$
$$= 0 + 6 + 10 + 12 + 4 + 10 = 42$$

So the mean is $\frac{42}{20} = 2.1$.

The highest number of goals is 5 and the lowest is 0. The range is 5 − 0 = 5, so the range is 5 goals.

EXERCISE 22.3

Find the mode, median, mean and range from the following tables.

1 The table shows the number of goals scored by 25 hockey teams.

Number of goals	0	1	2	3	4	5	6
Number of teams	1	2	4	7	8	1	2

2 The table shows the number of days 30 people were absent from work last week.

Number of days	0	1	2	3	4	5
Number of people	9	8	4	5	3	1

3 The table shows the numbers of people living in houses in a street.

Number of people	1	2	3	4	5	6	7
Number of houses	3	6	5	2	3	0	1

4 According to the label on a box of screws, the 'average' contents are 15. John bought some of these boxes of screws and counted the number of screws in each box. The table shows his results.

Number of screws	13	14	15	16	17
Number of boxes	1	1	7	6	3

22.7 Grouped frequency tables

In Section 6.9, you saw how to use class intervals to show data in grouped frequency tables. Even if you do not have the original data, you can use a grouped frequency table to obtain a variety of statistical information.

The marks scored by 50 students in a test were used to complete this grouped frequency table.

Mark	1–5	6–10	11–15	16–20	21–25
Frequency	7	9	12	17	5

You cannot find the mode from the table, but you can find which class interval has the highest frequency. This is called the **modal class interval**. So, in this example, the modal class interval is 16–20. The mode does not necessarily lie in the modal class interval.

Nor can you use the table to find the exact values of the mean and the median but you can use it to estimate them.

At one extreme, the seven students in the interval 1–5 could each have a mark of 1; at the other, they could each have a mark of 5.

It is more likely, though, that their marks are distributed throughout the interval. To estimate the mean, you need only an approximate sum of the seven marks, so assume each of the seven students scored a mark in the middle of the interval. The halfway mark of the interval 1–5 is 3. So the total mark of the seven students is approximately 3 × 7 = 21. Continuing like this, you approximate the sum of all the students' marks.

$$\begin{aligned}\text{Sum} &\approx 3 \times 7 + 8 \times 9 + 13 \times 12 + 18 \times 17 + 23 \times 5 \\ &= 21 + 72 + 156 + 306 + 115 = 670\end{aligned}$$

Finally, to obtain an estimate of the mean, divide the approximate sum, 670, by the number, 50, of students.

$$\text{Mean} \approx \frac{670}{50} = 13.4$$

The median lies halfway between the 25th and 26th values. You can find the interval in which the median lies using an approach similar to that in Section 22.5.

The first seven students scored marks in the interval 1–5, so the seventh student scored a mark of 5 or less. The next nine students scored marks from 6–10, so the 16th student scored a mark of 10 or less. The next 12 students scored marks from 11–15, so the 28th student scored a mark of 15 or less. The 25th and 26th student must each have scored marks in the interval 11–15, so the median lies in this interval. In this case, the median lies in the middle interval, but this will not always be so.

The numbers 7, 16 and 28 are called **cumulative frequencies**. The next one is 45 (28 + 17) and the last one is 50, the total number of students. You can show the cumulative frequency in a table.

Mark	*up to 5*	*up to 10*	*up to 15*	*up to 20*	*up to 25*
Cumulative frequency	7	16	28	45	50

You can use the values in the table to draw a **cumulative frequency graph**. (See Figure 22.2.) A cumulative frequency of 7 is plotted above a mark of 5, because you cannot be sure that all seven students have been included until you reach a mark of 5. A cumulative frequency of 16 is plotted above a mark of 10 and so on. You can join the points either with a smooth curve, as shown, or with a series of straight lines.

As this data is discrete and the marks are whole numbers, you plot the cumulative frequencies above 5, 10, 15, 20 and 25. Had the data been continuous, you should plot above 5.5, 10.5, 15.5, 20.5 and 25.5. However, in practice, this makes little difference in the results.

You can use the graph to obtain an estimate for the median by finding the mark which corresponds to a cumulative frequency of 25. (Strictly speaking, it is a cumulative frequency of $25\frac{1}{2}$ but 25 is near enough for an estimate.) The dotted lines with arrows in Figure 22.2 show how this method gives 14 as an estimate of the median.

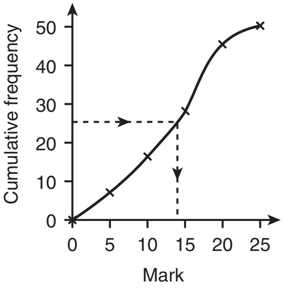

Figure 22.2

The median is the mark with a cumulative frequency 25, which is half of the total cumulative frequency, 50. The mark with a cumulative frequency which is one-quarter of the total cumulative frequency, $12\frac{1}{2}$, is called the **lower quartile** or **first quartile**.

From the graph, the lower quartile is 8. You can think of the median as the second quartile. Then the mark with a cumulative frequency of three-quarters of the total cumulative frequency, that is, $37\frac{1}{2}$, is the **upper quartile** or **third quartile**. From the graph, the upper quartile is 18.

The difference between the lower quartile and upper quartile is called the **interquartile range**, which is a measure of the spread of the marks. In this case, therefore, the interquartile range = 18 – 8 = 10.

Insight

The interquartile range tells you about the spread of the middle 50% of the data.

You can also use the graph to find how many candidates scored less than a certain mark and, by subtraction from 50, how many scored more than that mark. For example, reading up to the curve from a mark of 19, you find that 42 students scored 19 marks or less so 50 – 42 = 8 eight students scored more than 19 marks.

In the next example, the class intervals for continuous data are written in a way you met in Section 6.9.

Example 22.7.1

The grouped frequency table is about the heights of 80 women.

Height (h cm)	$140 \le h < 150$	$150 \le h < 160$
Frequency	4	29
Height (h cm)	$160 \le h < 170$	$170 \le h < 180$
Frequency	38	9

(a) State the modal class interval.
(b) Work out an estimate of the mean.
(c) Draw a cumulative frequency graph and use it to estimate the median, the lower quartile, the upper quartile and the interquartile range.

(a) The modal class interval is $160 \le h < 170$.

(b) Mean in (cm) $\approx \dfrac{145 \times 4 + 155 \times 29 + 165 \times 38 + 175 \times 9}{80}$

$$= \frac{580 + 4495 + 6270 + 1575}{80} = \frac{12\,920}{80}$$

$$= 161.5$$

The mean is approximately 161.5 cm.

(c)

Height (h cm)	140 ≤ *h* < 150	140 ≤ *h* < 160
Cumulative frequency	4	33
Height (h cm)	140 ≤ *h* < 170	140 ≤ *h* < 180
Cumulative frequency	71	80

Notice how the entries in the row marked 'heights' differ from the ones in the earlier table in this example. This is typical of tables which are constructed for drawing cumulative frequency graphs.

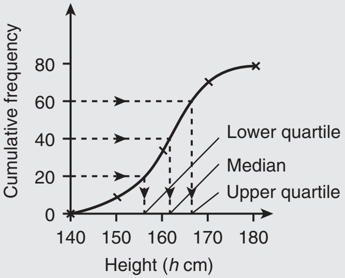

From the cumulative frequency graph, the median is 162 cm; the lower quartile is 157 cm and the upper quartile is 167 cm. The interquartile range is, therefore, 167 cm – 157 cm = 10 cm.

EXERCISE 22.4

1 The table gives information about the number of words in 40 sentences in a book.

Number of words	1–5	6–10	11–15	16–20
Frequency	5	9	19	7

(a) State the modal class interval.
(b) Find an estimate of the mean.
(c) Draw a cumulative frequency graph to estimate the median, the lower and upper quartiles and the interquartile range.

(Contd)

2 The table gives information about the weights, in kilograms, of 100 men.

Weight (w kg)	Frequency
$60 \leq w < 70$	8
$70 \leq w < 80$	14
$80 \leq w < 90$	23
$90 \leq w < 100$	39
$100 \leq w < 110$	16

(a) State the modal class interval.
(b) Find an estimate of the mean.
(c) From a cumulative frequency graph, estimate the median, the lower and upper quartiles and the interquartile range.
(d) How many men weighed more than 76 kg?

THINGS TO REMEMBER

- *The mode is the data item which occurs most often. When a set of data has two modes it is bi-modal. When data is organized into a grouped frequency table the group with the highest frequency is called the modal class interval.*

- *The median is the middle value, after the data have been arranged in order of size.*

- $\text{Mean} = \dfrac{\text{sum of values}}{\text{number of values}}$

- *Sum of values = mean × number of values*

- *Range = highest value – lowest value*

- *The range is a measure of the spread of data.*

- *Cumulative frequency is the running total of the frequencies once the data has been organized in a frequency table.*

- *A cumulative frequency graph is drawn by plotting the cumulative frequencies against the upper bound of each class interval.*

- *The lower quartile or first quartile is the data value which is one-quarter of the way through the ordered data set. The upper quartile or third quartile is the data value which is three-quarters of the way through the ordered data set.*

- *Interquartile range = upper quartile – lower quartile*

23

Graphs 2

In this chapter you will learn:
- ***how to find equations of straight lines***
- ***how to draw straight-line graphs***
- ***how to find and use the gradient of a straight-line graph***
- ***how to draw quadratic, cubic and reciprocal graphs*.**

23.1 Equations of straight lines

You saw in Section 8.3 how to find the equation of a straight line which is parallel to one of the axes. You can sometimes find the equation of a line which is not parallel to one of the axes by finding the coordinates of some points on the line and looking for a relationship between the x-coordinates and the y-coordinates.

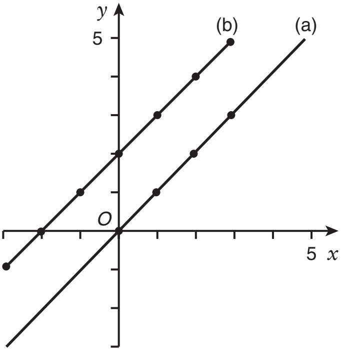

Figure 23.1

For example, in Figure 23.1, the coordinates of some of the points on line (a) are (0, 0), (1, 1), (2, 2) and (3, 3). In general, the y-coordinate is equal to the x-coordinate, so the equation of line (a) is $y = x$ or $x = y$. The coordinates of some of the points on line (b) are (–3, –1), (–2, 0), (–1, 1), (0, 2), (1, 3), (2, 4) and (3, 5). The relationship may be clearer if you show the coordinates in a table.

x	–3	–2	–1	0	1	2	3
y	–1	0	1	2	3	4	5

The y-coordinate is 2 more than the x-coordinate, so the equation of line (b) is $y = x + 2$. It is parallel to the line $y = x$ and is that line moved up 2 units. In a similar way, you could show that the line $y = x - 2$ is the line $y = x$ moved down 2 units.

The coordinates of some of the points on line (c) in Figure 23.2 are shown in the table.

x	–1	0	1	2
y	–2	0	2	4

The y-coordinate is twice the x-coordinate, so the equation of line (c) is $y = 2x$.

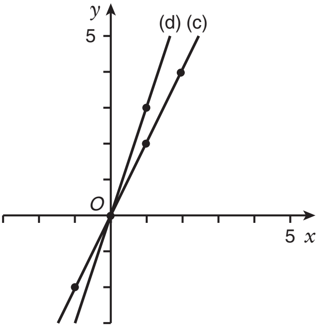

Figure 23.2

This table shows the coordinates of some of the points on line (d) in Figure 23.2.

x	–1	0	1	2
y	–3	0	3	6

The y-coordinate is 3 times the x-coordinate, so the equation of line (d) is $y = 3x$. In general, a line whose equation is $y = mx$ passes through the origin. The coefficient of x, 3 in the case of $y = 3x$, is related to the steepness of the line. The coefficient of x may be negative or a fraction.

Example 23.1.1

Find the equation of each of the lines shown in the diagram.

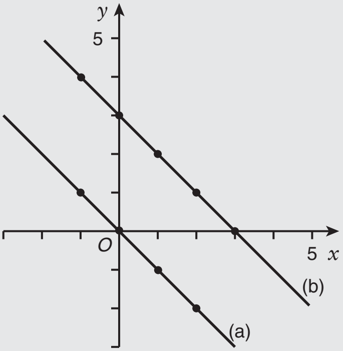

The table shows the coordinates of some of the points on line (a).

x	–1	0	1	2
y	1	0	–1	–2

The y-coordinate is –1 times the x-coordinate, so the equation of line (a) is $y = -x$ or $x + y = 0$.

The table shows the coordinates of some of the points on line (b).

x	–1	0	1	2	3
y	4	3	2	1	0

The sum of the x-coordinate and the y-coordinate is 3, so the equation of line (b) is $x + y = 3$.

Alternatively, you could say that line (b) is the line with equation $y = -x$ moved up 3 units; so the equation of line (b) is $y = -x + 3$, which you obtain when you make y the subject of the equation $x + y = 3$.

Notice that in both $y = -x$ and $y = -x + 3$, the coefficient of x is negative (–1) and that the lines are sloping in the opposite direction to those on the previous page, which have a positive coefficient of x in their equations.

EXERCISE 23.1

1 Find the equation of each of the lines shown in the diagrams.

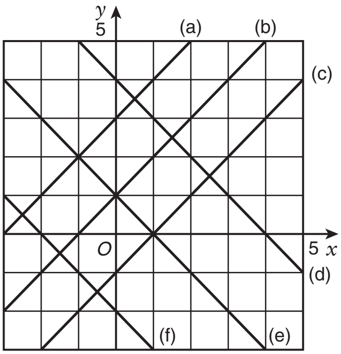

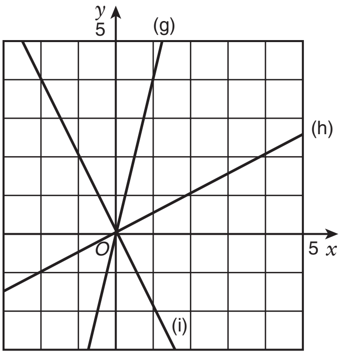

23.2 Drawing straight-line graphs

To draw a graph when you are told its equation, you have to substitute values of x into the equation and find the corresponding y-values. This gives you the coordinates of points on the graph.

For example, when you substituting the value $x = 3$ into the equation $y = 2x + 1$, $y = 2 \times 3 + 1 = 7$. Thus (3, 7) is a point on the graph of $y = 2x + 1$. The coordinates are often shown in a table of values. This table shows the x- and y-coordinates of points on the graph of $y = 2x + 1$ for values of x between $x = -2$ and $x = 4$.

x	−2	−1	0	1	2	3	4
y	−3	−1	1	3	5	7	9

You can then plot the points and join them with a straight line (see Figure 23.3 overleaf).

The line cuts the y-axis at (0, 1). This point is called its **intercept** on the y-axis or **y-intercept**.

In general, the graph of an equation of the form $y = mx + c$ is a straight line with $(0, c)$ as its intercept on the y-axis. For example, when $m = 3$ and $c = -2$, the equation is $y = 3x - 2$. Thus, the graph of $y = 3x - 2$ is a straight line and its intercept on the y-axis is (0, –2).

If you are not asked to complete a table of values like the one above, you need plot only two points on a line to be able to draw it, but a third point is a useful check.

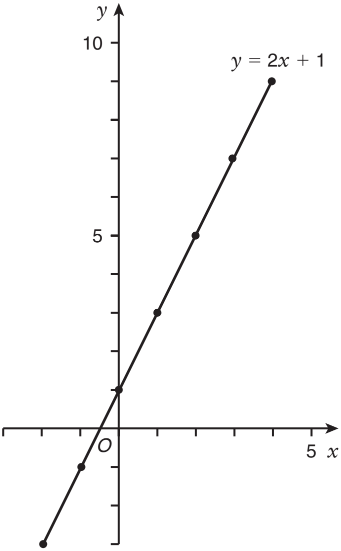

Figure 23.3

Example 23.2.1

(a) Draw the line whose equation is $y = -2x + 5$ for values of x between $x = -2$ and $x = 4$.

(b) Write down the coordinates of its intercept on the y-axis.

(a) Choose three suitable x values, say $x = 0$, $x = 1$ and $x = 2$ and substitute these values into the equation $y = -2x + 5$.

When $x = 0$, $y = 0 + 5 = 5$; when $x = 1$, $y = -2 + 5 = 3$; when $x = 2$, $y = -4 + 5 = 1$.

With practice, you should be able to do this working mentally and put the y-value obtained using each x-value straight into a table of values.

x	0	1	2
y	5	3	1

Plot the points with coordinates (0, 5), (1, 3) and (2, 1). Join them with a straight line and extend it in both directions to cover the range of x-values specified, in this case, $x = -2$ to $x = 4$. (See diagram.)

(b) The intercept on the y-axis is (0, 5).

Although equations of lines are often written in the form $y = mx + c$, you can use algebra to express them in other ways. The equation $y = -2x + 5$, for example, may be rearranged and written as $y = 5 - 2x$ or $2x + y = 5$ or any other equivalent form.

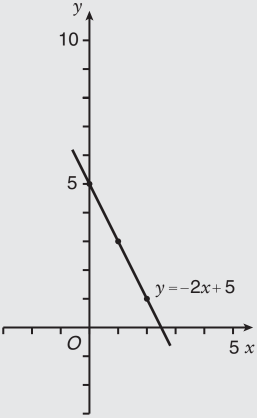

EXERCISE 23.2

1 Complete the table for the equation $y = 3x - 2$ and draw its graph.

x	–2	–1	0	1	2	3	4
y		–5				7	

2 Complete the table for the equation $y = \frac{1}{2}x + 1$ and draw its graph.

x	–6	–4	–2	0	2	4	6
y	–2					3	

3 Complete the table for the equation $y = -2x - 3$ and draw its graph.

x	–2	–1	0	1	2	3	4
y		–1					

4 Draw lines with the following equations for values of x between $x = -4$ and $x = 4$. For each line, write down the coordinates of its intercept on the y-axis.

(a) $y = 4x + 3$ **(b)** $y = -3x + 4$ **(c)** $y = \frac{1}{3}x - 2$
(d) $y = 3 - \frac{1}{2}x$ **(e)** $x + y = 2$ **(f)** $3x + y = 6$

23.3 The gradient of a straight line

Figure 23.4 shows the lines with equations $y = 2x + 3$, $y = 2x + 1$ and $y = 2x - 1$. You saw in Section 23.1 that the coefficient of x affects the steepness of a line. In this case, the coefficient, 2, of x is the same in all three equations and the lines are parallel.

You saw in Section 23.2 that, using the +3, +1 and –1 in the equations, you can write down the intercepts on the y-axis of the lines as (0, 3), (0, 1) and (0, –1) respectively.

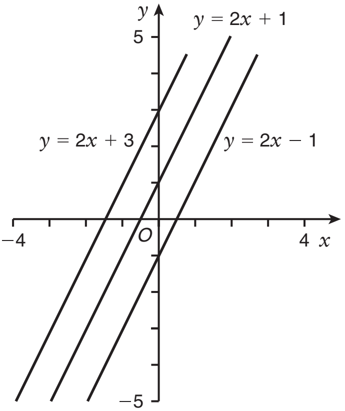

Figure 23.4

Thus, you can write down the equation of any line parallel to these lines, if you know its intercept on the y-axis. For example, if its intercept on the y-axis is (0, –5), its equation is $y = 2x - 5$.

Example 23.3.1

A line passes through the point with coordinates (2, 10) and is parallel to the line whose equation is $y = 3x - 1$. Find the equation of the line and its intercept on the y-axis.

The equation of the line is $y = 3x + c$. When $x = 2$, $y = 10$, so $10 = 3 \times 2 + c$, that is $c = 4$. Thus, the equation is $y = 3x + 4$ and the intercept is (0, 4).

To find a more precise interpretation than 'steepness' of the coefficient, m, of x in the equation $y = mx + c$, consider the line with equation $y = 2x + 1$, shown in Figure 23.5.

A, *B* and *C* are points on the line.

The value of $\frac{BE}{AE}$ is $\frac{2}{1} = 2$.

The value of $\frac{CD}{BD}$ is $\frac{4}{2} = 2$.

In fact, for any two points on the line, $\frac{\text{the } y\text{-step}}{\text{the } x\text{-step}}$ is always 2. This is the **gradient** of the line. (It is also the tangent of the angle the line makes with the horizontal.)

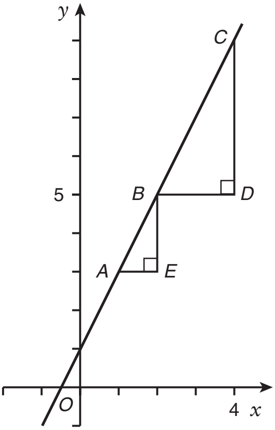

Figure 23.5

In the equation $y = mx + c$, m is the gradient of the line and its intercept on the y-axis is $(0, c)$. For example, the line whose equation is $y = 6x - 5$ has a gradient of 6 and its intercept on the y-axis is $(0, -5)$.

In a similar way, the line with equation $y = -2x + 5$ (see Example 23.2.1) has a gradient of -2 and its intercept is $(0, 5)$. You can see from its graph the direction of the slope of a line with a negative gradient. A horizontal line has a gradient of zero.

To find the gradient of a line joining a pair of points, start by drawing a sketch. It need not be accurate but the points must be in the correct relative positions. This will enable you to decide whether the gradient of the line joining them is positive or negative.

In general, to find the gradient of the line joining the points with coordinates (a, b) and (c, d), you should use

$$\text{gradient} = \frac{d-b}{c-a} \text{ or } = \frac{b-d}{a-c}$$

Example 23.3.2

Find the gradient of the line joining the points with coordinates **(a)** (5, 3) and (2, 1) **(b)** (3, 8) and (5, 2)

(a) From the sketch, which is not to scale, the gradient is positive. The y-step is $3 - 1 = 2$ and the x-step is $5 - 2 = 3$.

$$\text{Gradient} = \frac{\text{the } y\text{-step}}{\text{the } x\text{-step}} = \frac{2}{3}$$

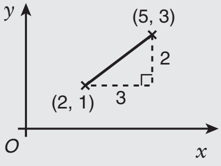

(b) When you sketch the points, you find that the line joining them slopes in the other way and the gradient is therefore negative. The y-step is $8 - 2 = 6$ and the x-step is $3 - 5 = -2$.

$$\text{Gradient} = \frac{\text{the } y\text{-step}}{\text{the } x\text{-step}} = \frac{6}{-2} = -3$$

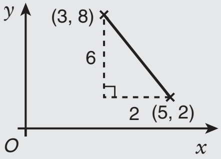

Insight

Lines that are 'uphill' (from left to right) have a positive gradient and lines that are 'downhill' have a negative gradient.

Example 23.3.3

Find the equation of the line joining the points (3, 8) and (5, 2).

From Example 23.3.2(b), the gradient is –3.
Using $y = mx + c$, the equation of the line is $y = -3x + c$.
When $x = 3$, $y = 8$ so $8 = -3 \times 3 + c$, that is, $c = 17$.
Therefore the equation of the line is $y = -3x + 17$.

As a check, notice that the values $x = 5$, $y = 2$ satisfy this equation.

Sometimes you may have to rearrange the equation of the line in the form $y = mx + c$, that is, make y the subject.

Example 23.3.4

Find **(a)** the gradient and **(b)** the intercept on the y-axis of the line whose equation is $3x - 2y = 12$.

Rearrange the equation as $2y = 3x - 12$.

Make y the subject $\quad y = \frac{3}{2}x - 6$

The gradient is $\frac{3}{2}$ and the intercept on the y-axis is $(0, -6)$.

EXERCISE 23.3

1 The equation of a line is $y = 7x - 4$. Write down
(a) the gradient of the line, **(b)** its intercept on the y-axis.

2 The gradient of a line is -5 and its intercept on the y-axis is $(0, 4)$. Write down the equation of the line.

3 A line passes through the point with coordinates $(4, 13)$ and is parallel to the line whose equation is $y = 2x - 3$. Find the equation of the line and its intercept on the y-axis.

4 Find the gradient of the line joining the points with coordinates
(a) $(3, 2), (5, 10)$ **(b)** $(-3, 5), (5, 1)$ **(c)** $(-2, 4), (5, 4)$

5 Find the equation of the line joining the points with coordinates
(a) $(0, 3), (3, 15)$ **(b)** $(4, 5), (10, 8)$ **(c)** $(1, 7), (4, -8)$

6 Find the gradient and the intercept on the y-axis of the line
(a) $2x + y = 3$ **(b)** $2x - 5y = 15$ **(c)** $3x + 4y = 24$

23.4 Curved graphs

Not all graphs are straight lines. The graph of $y = x^2$, for example, is curved. Here is a table of values for the equation $y = x^2$.

x	-3	-2	-1	0	1	2	3
y	9	4	1	0	1	4	9

Figure 23.6 overleaf shows the points from the table and the graph of $y = x^2$, formed by joining the points with a curve, called a **parabola.**

This is the simplest **quadratic** graph. The word ‘quadratic’ comes from the Latin *quadratus*, meaning ‘square’. In the equation of a quadratic graph, the highest power of x is x^2. Thus, $y = x^2 + 3$, $y = 3x^2 - 2x + 1$ and $y = 4x^2 - 3x$ are all equations of quadratic graphs.

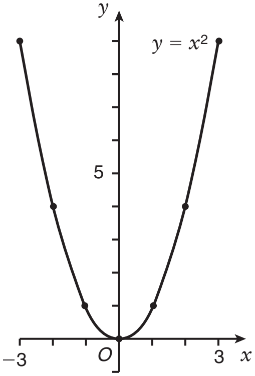

Figure 23.6

Note that although Figure 23.6 is drawn with the same scale on both axes, it can be difficult to see all the features of the graph. It is usual to use different scales on the axes if necessary.

Example 23.4.1

Draw the graph of $y = x^2 - 2x - 1$ for values of x from –2 to 4.

When $x = -2$, $y = (-2)^2 - 2 \times (-2) - 1 = 4 + 4 - 1 = 7$.
Making similar calculations, you can complete the table of values.

x	–2	–1	0	1	2	3	4
y	7	2	–1	–2	–1	2	7

The diagram on the right shows the graph of $y = x^2 - 2x - 1$.

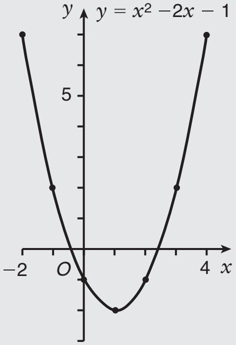

Insight

When the coefficient of the x^2 term is positive (e.g. $y = x^2 - 2x$) then the parabola looks like $\cup$ (a ‘smile’) and when the coefficient of the x^2 term is negative (e.g. $y = 2x - x^2$) then the parabola looks like $\cap$ (a ‘frown’).

Graphs of equations in which the highest power of x is x^3 are called **cubic** graphs. For example, $y = 2x^3 - 9$ and $y = x^3 + 4x^2 - 2x + 7$ are equations of cubic graphs. The simplest cubic graph is $y = x^3$.

Here is a table of values for the equation $y = x^3$.

x	–3	–2	–1	0	1	2	3
y	–27	–8	–1	0	1	8	27

Figure 23.7 shows the graph of $y = x^3$.

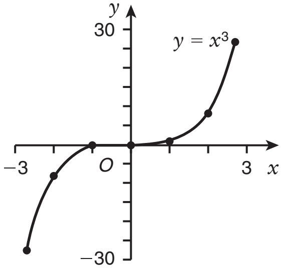

Figure 23.7

Example 23.4.2

Draw the graph of $y = x^3 + x^2 - 8x$ for values of x from –3 to 3. First, draw up a table of values.

x	–3	–2	–1	0	1	2	3
y	6	12	8	0	–6	–4	12

The diagram shows the graph of $y = x^3 + x^2 - 8x$.

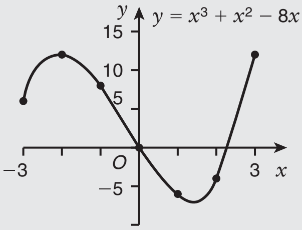

Reciprocal graphs are the final type included in this chapter. You find the reciprocal of a number by dividing it into 1. For example, the reciprocal of 2 is $\frac{1}{2}$. So a reciprocal graph has an equation such as $\frac{2}{x}$ or $\frac{4}{x}$. The simplest reciprocal graph is $\frac{1}{x}$.

Here is a table of values for $\frac{1}{x}$. As division by 0 is undefined, y is undefined when $x = 0$ and so values of x *near* 0 are used in the table.

x	–3	–2	–1	–0.5	–0.2	0.2	0.5	1	2	3
y	–0.3	–0.5	–1	–2	–5	5	2	1	0.5	0.3

Figure 23.8 shows the graph of $\frac{1}{x}$. It is unlike the graphs you have met so far in this chapter in that there is a break in the graph at $x = 0$. The x- and y-axes have the property that the graph gets infinitely close to them, but never actually touches them.

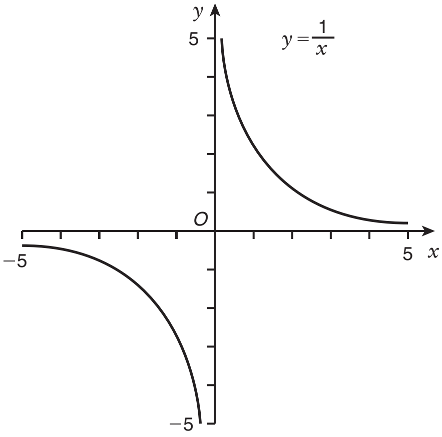

Figure 23.8

EXERCISE 23.4

The graphs in the text were drawn without grid lines because the features stand out more clearly that way. To answer the following questions you will need to use graph paper with a 2 mm grid.

1 Draw the following quadratic graphs for values of x from –4 to 4.

(a) $y = x^2 + 3$ **(b)** $y = 5x^2$ **(c)** $y = \frac{1}{2}x^2$

(d) $y = -x^2$ **(e)** $y = -2x^2$ **(f)** $y = x^2 - 2x$

(g) $y = (x - 1)^2$ **(h)** $y = x^2 - 4x - 1$ **(i)** $y = 2x^2 + 3x - 5$

2 Draw the following cubic graphs. Take values of x from –3 to 3.

(a) $y = x^3 + 8$ **(b)** $y = 2x^3$ **(c)** $y = \frac{1}{2}x^3$

(d) $y = -x^3$ **(e)** $y = -2x^3$ **(f)** $y = (x - 1)^3$

(g) $y = x^3 - x^2$ **(h)** $y = x^3 - x^2 - 2x$ **(i)** $y = x^3 + 3x - 2$

3 Draw the following reciprocal graphs. Use the same x values as those used for the graph of $y = \frac{1}{x}$.

(a) $y = \frac{2}{x}$ **(b)** $y = -\frac{1}{x}$ **(c)** $y = -\frac{2}{x}$ **(d)** $y = \frac{1}{x} + 5$

EXERCISE

THINGS TO REMEMBER

- *A line whose equation is $y = mx$ is a straight line through the origin.*

- *The point where a line cuts the y-axis is called the y-intercept.*

- *You can make a table of values to help you draw a graph. Choose x values and then work out the corresponding y values.*

x	–3	–2	–1	0	1	2	3
y							

- *An equation in the form $y = mx + c$ produces a straight line with $(0, c)$ as its intercept on the y-axis and gradient m.*

- *Gradient* $= \frac{\text{the } y\text{-step}}{\text{the } x\text{-step}}$

- *Graphs of equations in which the highest power of x is x^2 are called quadratic graphs. The resulting curve is called a parabola.*

- *Graphs of equations in which the highest power of x is x^3 are called cubic graphs.*

- *The reciprocal of a number is found by dividing it into 1. Reciprocal graphs have equations like $y = \frac{1}{x}$ or $y = 2 + \frac{3}{x}$.*

24

Equations 2

In this chapter you will learn:

- ***how to solve simultaneous equations***
- ***how to expand and simplify products which give quadratic expressions***
- ***how to factorize quadratic expressions***
- ***how to solve quadratic equations by factorization***
- ***how to solve cubic equations by trial and improvement.***

24.1 Simultaneous equations

If you were asked to find values of x and y which satisfy the equation $2x + y = 4$, you might say '$x = 0$, $y = 4$'. This is correct but it is not the only possible pair of values. Two more are $x = 1$, $y = 2$ and $x = 2$, $y = 0$. In fact, the number of pairs of values of x and y which satisfy $2x + y = 4$ is infinite.

From Chapter 23, you know that the graph of $2x + y = 4$ is a straight line (see Figure 24.1). The x -coordinate and y-coordinate of every point on this line represent a solution to the equation $2x + y = 4$.

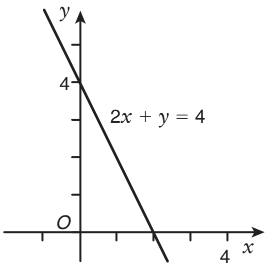

Figure 24.1

Similarly, the equation $x + y = 3$ has an infinite number of solutions. For example, $x = 3$, $y = 0$ and $x = 1$, $y = 2$ satisfy $x + y = 3$.

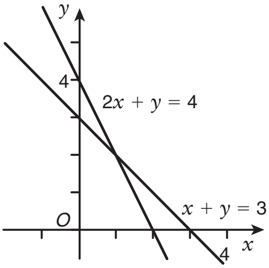

Figure 24.2

There is, however, only one pair of values which satisfy both $2x + y = 4$ and $x + y = 3$ at the same time, or **simultaneously**. This pair is $x = 1$, $y = 2$, the x- and y-coordinates of the point of intersection of the lines whose equations are $2x + y = 4$ and $x + y = 3$. These equations are called **simultaneous equations**. Their solution is $x = 1$, $y = 2$. To check, substitute this pair of values into the equations. In $2x + y = 4$, $2 \times 1 + 2 = 4$ and, in $x + y = 3$, $1 + 2 = 3$.

Insight

You need to have three simultaneous equations to find the value of three unknowns. In fact, you need the same number of simultaneous equations as there are unknowns to find.

Example 24.1.1

Solve the simultaneous equations $3x + y = 7$ and $2x - y = 8$.

Draw up tables of values of $3x + y = 7$ and $2x - y = 8$.

x	0	1	2
y	7	4	1

x	4	5	6
y	0	2	4

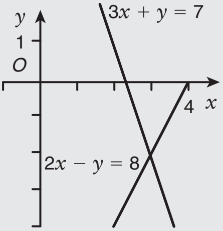

The diagram shows the lines with equations $3x + y = 7$ and $2x - y = 8$. The lines intersect at the point with coordinates $(3, -2)$, so the solution is $x = 3$, $y = -2$.

Checking in the original equations, $3 \times 3 + (-2) = 7$ and $2 \times 3 - (-2) = 8$.

EXERCISE 24.1

1 Find three pairs of values of x and y which satisfy these equations.
(a) $x + y = 5$ **(b)** $x - y = 3$ **(c)** $2x + 3y = 6$

2 Which of these pairs of values satisfy the equation $2x + 5y = 20$?
(a) $x = 5, y = 2$ **(b)** $x = 4, y = 0$ **(c)** $x = -5, y = 6$

3 Which of these pairs of values satisfies the simultaneous equations $3x + 4y = 8$, $2x - 3y = 11$?
(a) $x = 0, y = 2$ **(b)** $x = 4, y = -1$ **(c)** $x = 8, y = -4$

4 Draw appropriate straight lines to solve these simultaneous equations. Check your answers.
(a) $x + y = 4, x - y = 2$ **(b)** $x + y = 5, 2x + y = 7$
(c) $x + 2y = 8, x - y = 2$ **(d)** $3x + 2y = 6, x - 2y = 10$
(e) $x + y = 2, 2x + 3y = 8$ **(f)** $x - y = 1, 3x - 2y = 1$

5 Try to solve the simultaneous equations $x + y = 3, x + y = 5$ by drawing appropriate straight lines. What happens?

24.2 Algebraic methods

Graphical methods are time-consuming and may give only approximate solutions. Using algebra, you can **eliminate** one of the unknowns so that you are left with an equation in only one unknown.

Sometimes you can achieve this simply by adding the equations. For example, to solve the simultaneous equations $2x + 3y = 11$, $5x - 3y = 17$, you can add the two left-hand sides
$2x + 3y + (5x - 3y) = 2x + 3y + 5x - 3y = 7x$
and then add the two right-hand sides
$11 + 17 = 28$.
Thus, $7x = 28$, so $x = 4$.

To find the value of y, substitute $x = 4$ into either of the original equations. If you substitute $x = 4$ into $2x + 3y = 11$, $2 \times 4 + 3y = 11$, that is, $3y + 8 = 11$, giving $y = 1$. The solution is $x = 4$, $y = 1$.

Checking by substituting $x = 4$, $y = 1$ into the equation $5x - 3y = 17$ gives $5 \times 4 - 3 \times 1 = 20 - 3 = 17$, so the solution is correct.

You have seen that, if the coefficients of an unknown differ only in sign, you can eliminate that unknown by adding two simultaneous equations. If, however, the coefficients of an unknown are equal, you can eliminate that unknown by subtracting one of the equations from the other.

Example 24.2.1

Solve the simultaneous equations $4x + 5y = 2$, $4x - 3y = 18$.

Label the equations A and B, and write them under each other.

$$\begin{aligned} 4x + 5y &= 2 \quad \text{A} \\ 4x - 3y &= 18 \quad \text{B} \end{aligned}$$

Then you can use these labels to explain your method. Subtracting equation B from equation A gives

$$\begin{aligned} 4x + 5y - (4x - 3y) &= 2 - 18 \\ 4x + 5y - 4x + 3y &= -16 \\ 8y &= -16 \\ y &= -2 \end{aligned}$$

Substituting $y = -2$ in A gives $4x - 10 = 2$ so $x = 3$. The solution is $x = 3, y = -2$.

Check in B: $4 \times 3 - (3 \times (-2)) = 12 - (-6) = 18$

In this example, you could subtract equation A from equation B. The disadvantage of this is that it gives you a negative coefficient of y.

If you cannot eliminate one of the unknowns just by adding or subtracting the equations, you will have to multiply one or both of the equations by an appropriate number and then add or subtract.

Example 24.2.2

Solve the simultaneous equations $4x + 3y = 10$, $5x + 6y = 8$.

$$\begin{aligned} 4x + 3y &= 10 \quad \text{A} \\ 5x + 6y &= 8 \quad \text{B} \end{aligned}$$

Multiply both sides of A by 2 to make the coefficients of y equal.
Label the new equation C.

$$8x + 6y = 20 \quad \text{C}$$

C – B gives $3x = 12$, that is, $x = 4$.
Substitute $x = 4$ in A giving $4 \times 4 + 3y = 10$, or $y = -2$.
The solution is $x = 4$, $y = -2$.
Check in B: $5 \times 4 + 6 \times (-2) = 20 - 12 = 8$

Notice that you should make the check in one of the original two equations. You know these are correct but there could be an error in one of those you have obtained by multiplying.

In the next example, you have to multiply both equations.
It also shows the usual way in which the steps in the working are described.

Example 24.2.3

Solve the simultaneous equations $5x + 2y = 7$, $4x - 3y = 24$.

$$\begin{aligned} 5x + 2y &= 7 \quad \text{A} \\ 4x - 3y &= 24 \quad \text{B} \end{aligned}$$

Multiplying *A* by 3 and *B* by 2 and re-labelling, gives

$$\begin{aligned} 15x + 6y &= 21 \quad \text{C} \\ 8x - 6y &= 48 \quad \text{D} \end{aligned}$$

(Contd)

Adding C and D gives $15x + 6y + (8x - 6y) = 21 + 48$, which is $23x = 69$, that is, $x = 3$.

Substituting $x = 3$ in A gives $5 \times 3 + 2y = 7$, or $y = -4$.

The solution is $x = 3$, $y = -4$.

Check in B: $4 \times 3 - (3 \times (-4)) = 12 - (-12) = 12 + 12 = 24$

In the above example, you could have started by multiplying A by 4 and *B* by 5. Then, by subtraction, you could have eliminated *x*.

You can use the elimination methods to solve problems in which the information can be expressed as a pair of simultaneous equations.

Example 24.2.4

The sum of two numbers is 59 and their difference is 15. Find the numbers.

Let x and y be the numbers. Then you can write the given information as two simultaneous equations in the form

$$\begin{aligned} x + y &= 59 \quad \text{A} \\ x - y &= 15 \quad \text{B} \end{aligned}$$

Adding A and B gives $2x = 74$, or $x = 37$.

Substituting in A, $37 + y = 59$, giving $y = 22$.

The numbers are 37 and 22.

Check in original problem: $37 + 22 = 59$ and $37 - 22 = 15$

EXERCISE 24.2

1 Solve the following pairs of simultaneous equations.

(a) $3x + y = 17$
$2x - y = 8$

(b) $2x + 3y = 7$
$2x + 7y = 11$

(c) $5x + 4y = 11$
$3x - 4y = 13$

2 Solve the following pairs of simultaneous equations.

(a) $3x - 2y = 1$
$4x + y = 5$

(b) $5x + 3y = 7$
$7x + 9y = 5$

(c) $2x + 3y = 12$
$3x - 4y = 1$

(d) $3x + 5y = 6$
$2x + 3y = 5$

(e) $4x - 3y = 4$
$6x + 5y = 25$

(f) $8x - 3y = 16$
$6x - 5y = 23$

3 The sum of two numbers is 74 and their difference is 24. Find the numbers.

4 Cinema tickets for two adults and a child cost £13. The cost for four adults and three children is £29. Find the cost of a child's ticket.

5 The sum of two numbers is 122. The difference between twice one number and three times the other is 49. Find the numbers.

6 The equation of a straight line is $y = mx + c$. When $x = 4$, $y = -1$ and when $x = 10$, $y = 2$. Find the values of m and c.

7 The points with coordinates $(4, -1)$ and $(10, 2)$ lie on the line whose equation is $ax + by = 6$. Find the values of a and b.

24.3 Quadratic expressions

A quadratic expression is one in which the highest power of x is x^2. Examples of quadratic expressions are $x^2 + 2x - 3$, $3x^2 + 5x$ and $4x^2 - 1$. Some quadratic expressions can be written as a product of factors, such as $2x(x - 3)$ and $(x + 4)(x + 3)$. You saw in Section 11.6 how to expand expressions like $2x(x - 3)$ and the method used there can be extended to expand expressions like $(x + 4)(x + 3)$.

It can be helpful to think of $(x + 4)(x + 3)$ as x 'lots' of $x + 3$ and 4 'lots' of $x + 3$. Then

$$\begin{aligned}(x + 4)(x + 3) &= x(x + 3) + 4(x + 3)\\ &= x^2 + 3x + 4x + 12\\ &= x^2 + 7x + 12\end{aligned}$$

Figure 24.3 illustrates the result $(x+4)(x+3) = x^2 + 7x + 12$.

You can think of the product $(x+4)(x+3)$, as the area of a rectangle of length $x+4$ and width $x+3$.

The total area $(x+4)(x+3)$ is the sum of the areas of the four smaller rectangles.

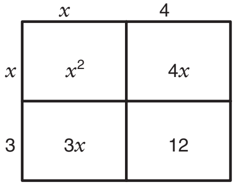

Figure 24.3

Notice that, before simplification, there are four terms in the expansion.

Insight

You need to multiply each term in the second bracket by each term in the first bracket.

Example 24.3.1

Expand and simplify $(x+2)(x+5)$.

$$\begin{aligned}(x+2)(x+5) &= x(x+5) + 2(x+5)\\ &= x^2 + 5x + 2x + 10\\ &= x^2 + 7x + 10\end{aligned}$$

With practice, you should be able to omit the first line of working.

Example 24.3.2

Expand and simplify the products

(a) $(x+7)(x+2)$ **(b)** $(x-5)(x+3)$ **(c)** $(x-1)^2$

(d) $(x+4)(x-4)$ **(e)** $(4x-5)(2x+3)$ **(f)** $(3x+1)(5-x)$

(a) $$\begin{aligned}(x+7)(x+2) &= x^2 + 2x + 7x + 14\\ &= x^2 + 9x + 14\end{aligned}$$

(b) $$\begin{aligned}(x-5)(x+3) &= x^2 + 3x - 5x - 15\\ &= x^2 - 2x - 15\end{aligned}$$

(c) $(x-1)^2 = (x-1)(x-1)$
$= x^2 - x - x + 1$
$= x^2 - 2x + 1$

(d) $(x+4)(x-4) = x^2 - 4x + 4x - 16$
$= x^2 - 16$

This is an example of the **difference of two squares**.

(e) $(4x-5)(2x+3) = 8x^2 + 12x - 10x - 15$
$= 8x^2 + 2x - 15$

(f) $(3x+1)(5-x) = 15x - 3x^2 + 5 - x$
$= 5 + 14x - 3x^2$

EXERCISE 24.3

1 Expand and simplify these products.

(a) $(x+3)(x+2)$ **(b)** $(x+4)(x-5)$ **(c)** $(x-5)(x+1)$
(d) $(x-6)(x-3)$ **(e)** $(x+5)^2$ **(f)** $(x-2)^2$
(g) $(x+7)(x-7)$ **(h)** $(x-4)(x+4)$ **(i)** $(x+3)(6-x)$
(j) $(5-x)(x+1)$ **(k)** $(2x+5)(3x+4)$ **(l)** $(5x+1)(4x-3)$
(m) $(6x-1)(2x-7)$ **(n)** $(3x+2)^2$ **(o)** $(2x-5)^2$
(p) $(9x+4)(9x-4)$ **(q)** $(3x-7)(3x+7)$ **(r)** $(5x+2)(6-x)$
(s) $(4x+3)(2-3x)$ **(t)** $(5-3x)^2$ **(u)** $(ax-b)(ax+b)$

24.4 Factorizing quadratic expressions

You saw in Section 11.7 how to factorize quadratic expressions with a common factor. For example, $x^2 - 5x = x(x-5)$.

A different approach is, however, needed to factorize other types of quadratic expression.

To factorize $x^2 + 3x + 2$, you would seek two bracketed expressions with a product of $x^2 + 3x + 2$.

The first term in each bracket must be x.

As the number term, called the **constant term**, is positive, +2, the signs in the brackets must be the same, either both + or both –.

The coefficient of x is positive, so the signs in the brackets must both be +.

The factors are $(x + ...)(x + ...)$, where the product of the missing numbers must be 2.

The only possible pair of numbers is 1 and 2, so the factors are $(x + 1)(x + 2)$.

You should check that the answer is correct by finding the product of the factors: $(x + 1)(x + 2) = x^2 + 2x + x + 2 = x^2 + 3x + 2$.

Example 24.4.1

Factorize $x^2 - x - 12$.

As the sign of the constant term is negative, the signs inside the brackets must be different.

The factors are $(x + \ldots)(x - \ldots)$, where the product of the missing numbers is 12.

The possible pairs of numbers are 1 and 12, 2 and 6, and 3 and 4.

The factors must be one of these pairs.

$(x + 1)(x - 12)$ $(x + 2)(x - 6)$ $(x + 3)(x - 4)$
$(x - 1)(x + 12)$ $(x - 2)(x + 6)$ $(x - 3)(x + 4)$

By a process of trying these products one by one, you find that
$x^2 - x - 12 = (x + 3)(x - 4)$

Check: $(x + 3)(x - 4) = x^2 - 4x + 3x - 12 = x^2 - x - 12$

You should have noticed that the process of factorization is a matter of following up the clues which are there. There is no sure-fire process for factorizing in this way.

Insight

A quick way of factorizing quadratic expressions is to look for a pair of numbers that multiply to give the constant term and add to give the coefficient of x. For example, for $x^2 - 4x + 3$ you need to find a pair of numbers which multiply to give +3 and add to give −4. Now, $-1 \times -3 = +3$ and $-1 + -3 = -4$, so $x^2 - 4x + 3 = (x - 1)(x - 3)$.

Example 24.4.2

Factorize $x^2 - 36$.

This is a difference of two squares. Example 24.3.2(d) suggests that $x^2 - 36 = (x + 6)(x - 6)$.

Check: $(x + 6)(x - 6) = x^2 - 6x + 6x - 36 = x^2 - 36$

In general, the difference of two squares is

$$x^2 - a^2 = (x + a)(x - a)$$

EXERCISE 24.4

1 One factor of each quadratic expression is given. Find the other factor.

(a) $x^2 + 6x + 8$, $(x + 2)$
(b) $x^2 - 2x - 8$, $(x - 4)$
(c) $x^2 + 2x - 8$, $(x - 2)$
(d) $x^2 - 6x + 8$, $(x - 2)$
(e) $x^2 + 9x + 8$, $(x + 1)$
(f) $x^2 - 7x - 8$, $(x + 1)$
(g) $x^2 - 7x + 12$, $(x - 3)$
(h) $x^2 - 4$, $(x + 2)$
(i) $x^2 - 100$, $(x - 10)$
(j) $x^2 + 2x$, (x)

2 Factorize the following quadratic expressions.

(a) $x^2 - 2x - 15$
(b) $x^2 + 9x + 14$
(c) $x^2 - 24x - 25$
(d) $x^2 - 10x - 24$
(e) $x^2 - 49$
(f) $x^2 + 8x$
(g) $x^2 + 10x - 56$
(h) $x^2 - 1$
(i) $x^2 - 8x - 20$
(j) $x^2 - 9x$

24.5 Quadratic equations

A quadratic equation is one in which the highest power of x is x^2, for example, $x^2 - 4 = 0$. If you were asked to solve this equation, you might say '$x = 2$'. This certainly satisfies $x^2 - 4 = 0$, as $2^2 - 4 = 0$. But there is a second solution, $x = -2$, because $(-2)^2 - 4 = 0$ also.

Figure 24.4 is the graph of $y = x^2 - 4$. It crosses the x-axis, that is, $y = 0$, at points whose x-coordinates are 2 and -2, the solutions of $x^2 - 4 = 0$. This suggests that you can solve any quadratic equation by drawing the related parabola and finding the x-coordinate of each of the points where it cuts the x-axis. There are usually two solutions.

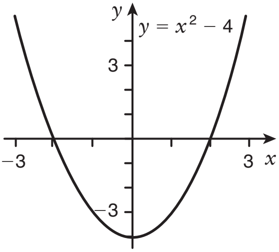

Figure 24.4

Example 24.5.1

Draw the graph of $y = x^2 - 2x - 1$ and use it to solve the equation $x^2 - 2x - 1 = 0$. Take values of x from -2 to 4.

x	-2	-1	0	1	2	3	4
y	7	2	-1	-2	-1	2	7

The diagram shows the graph of $y = x^2 - 2x - 1$. The x-coordinates of the points where the graph crosses the x-axis are $x = 2.4$ and $x = -0.4$. Thus, the solutions of the equation $x^2 - 2x - 1 = 0$ are, correct to one decimal place, $x = 2.4$ and $x = -0.4$.

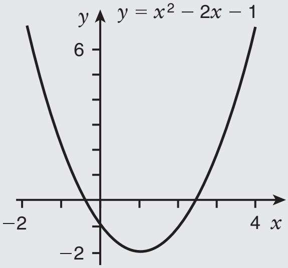

To use this method, you must have 0 on one side of the quadratic equation. The equation $x^2 = 3x + 5$, for example, would have to be rearranged as $x^2 - 3x - 5 = 0$ and then solved using the graph of $y = x^2 - 3x - 5$. Alternatively, you could draw the graph of $y = x^2$, a parabola, and the graph of $y = 3x + 5$, a straight line. The solutions of $x^2 - 3x - 5 = 0$ are then the x-coordinates of the points of intersection of the two graphs.

EXERCISE 24.5

1 Solve each of these equations by drawing a graph, using the range of values of x given in brackets. Where necessary, give your solutions correct to one decimal place.
(a) $x^2 - 7 = 0$ (–4 to 4) **(b)** $x^2 - 2x = 0$ (–2 to 4)
(c) $x^2 - 4x + 3 = 0$ (–2 to 4) **(d)** $x^2 + x - 3 = 0$ (–3 to 3)
(e) $2x^2 - 3x - 7 = 0$ (–4 to 4) **(f)** $x^2 = 3x + 2$ (–2 to 4)

2 **(a)** Solve $x^2 - 4x + 4 = 0$, by drawing a graph and taking values of x from –2 to 4. In what way is this equation different from those in question 1?
(b) Draw the graph of $y = x^2 + x + 1$, taking values of x from –3 to 3. Explain why you cannot use your graph to solve the equation $x^2 + x + 1 = 0$.

24.6 Solution by factorizing

You can solve some quadratic equations by factorization. This depends on the fact that, if a and b are two numbers such that $ab = 0$, then either $a = 0$ or $b = 0$. So, if $(x - 2)(x - 4) = 0$, then $x - 2 = 0$ or $x - 4 = 0$, that is, $x = 2$ or $x = 4$. To use it, you must have 0 on one side of the equation and be able to factorize the expression on the other.

Thus, to solve the equation $x^2 - 4x + 3 = 0$, factorize to obtain $(x - 3)(x - 1) = 0$. Either $x - 3 = 0$ or $x - 1 = 0$, that is, $x = 3$ or $x = 1$.

Example 24.6.1

Solve these quadratic equations.

(a) $x^2 + 4x - 5 = 0$ **(b)** $x^2 - 2x = 0$ **(c)** $x^2 - 4x + 4 = 0$

(d) $x^2 - 4 = 0$ **(e)** $x^2 + 9x - 8 = 2x$

(a) $x^2 + 4x - 5 = 0$

$(x + 5)(x - 1) = 0$

Either $x + 5 = 0$ or $x - 1 = 0$, so $x = -5$ or $x = 1$.

(b) $x^2 - 2x = 0$

$x(x - 2) = 0$

Either $x = 0$ or $x - 2 = 0$, so $x = 0$ or $x = 2$.

(c) $x^2 - 4x + 4 = 0$

$(x - 2)(x - 2) = 0$

Either $x - 2 = 0$ or $x - 2 = 0$, so $x = 2$ or $x = 2$.

In this case there is only one root of the equation $x^2 - 4x + 4 = 0$, $x = 2$, although it is sometimes given as $x = 2$ repeated. See Exercise 24.5, question 2(a). The parabola $y = x^2 - 4x + 4$ touches the x-axis.

(d) $x^2 - 4 = 0$

$(x - 2)(x + 2) = 0$

Either $x - 2 = 0$ or $x + 2 = 0$, so $x = 2$ or $x = -2$.

(e) Rearrange $x^2 + 9x - 8 = 2x$ to get 0 on the right-hand side.

$x^2 + 9x - 8 = 2x$ so $x^2 + 7x - 8 = 0$

$(x - 1)(x + 8) = 0$

Either $x - 1 = 0$ or $x + 8 = 0$, so $x = 1$ or $x = -8$.

You can solve some problems by expressing the information as a quadratic equation and solving it.

Example 24.6.2

The length of a rectangle is 4 cm greater than its width. The area of the rectangle is 96 cm^2. Find its width.

Let x cm be the width of the rectangle. The length of the rectangle is, therefore, $(x + 4)$ cm and its area is $x(x + 4)$ cm^2. So $x(x + 4) = 96$.

$$\begin{aligned} x(x + 4) &= 96 \\ x^2 + 4x &= 96 \\ x^2 + 4x - 96 &= 0 \\ (x - 8)(x + 12) &= 0 \end{aligned}$$

Either $x - 8 = 0$ or $x + 12 = 0$, so $x = 8$ or $x = -12$.

As the width of a rectangle cannot be negative, $x = 8$. The width of the rectangle is 8 cm.

Check: The length of the rectangle is 12 cm and its area is 96 cm^2.

EXERCISE 24.6

1 Write down the solutions to each of these quadratic equations.

(a) $(x - 6)(x - 1) = 0$ **(b)** $(x + 4)(x - 3) = 0$
(c) $(x + 7)(x - 7) = 0$ **(d)** $(x + 4)(x + 5) = 0$
(e) $x(x - 6) = 0$ **(f)** $x(x + 8) = 0$

2 Solve these quadratic equations.

(a) $x^2 - 3x + 2 = 0$ **(b)** $x^2 + x - 2 = 0$
(c) $x^2 + 7x + 12 = 0$ **(d)** $x^2 - 25 = 0$
(e) $x^2 - 10x + 25 = 0$ **(f)** $x^2 + 10x = 0$
(g) $x^2 - x - 20 = 0$ **(h)** $x^2 - 64 = 0$
(i) $x^2 - 7x = 0$ **(j)** $x^2 - x - 9 = 2x + 1$

3 The length of a rectangle is 3 cm greater than its width. The area of the rectangle is 28 cm^2. Find its length.

4 The square of a number is 27 more than 6 times the number. Find the number.

(Contd)

5 The sum of the square of a number and the number itself is 72. Find the number.

6 Two numbers have a difference of 2 and a product of 63. Find the numbers.

7 The product of two consecutive counting numbers is 56. Find the numbers.

24.7 Cubic equations

A cubic equation is one in which the highest power of x is x^3. For example, $x^3 - 4x^2 - 2x - 3 = 0$, $x^3 = 3x + 7$ and $2x^3 + 5x^2 - 10 = 0$ are cubic equations.

You can solve simple cubic equations, such as $x^3 = 8$, in your head. A solution to this equation is $x = 2$, because $2^3 = 8$. Notice, however, that $x = -2$ is not a solution, as $(-2)^3 = -8$.

You can also find approximate solutions to equations like $x^3 = 10$, in which the right-hand side is not the cube of an integer. In this case, x is the **cube root** of 10, written $\sqrt[3]{10}$. Using your calculator, you can find the value of $\sqrt[3]{10}$; it is 2.15, correct to two decimal places. Similarly, an approximate solution to $x^3 = -10$ is $x = -2.15$.

Sometimes you can use factorization to solve a cubic equation. For example, factorizing the left-hand side of the equation $x^3 - 4x = 0$, $x(x^2 - 4) = 0$. Factorizing again, $x(x - 2)(x + 2) = 0$ and so $x = 0$, $x = 2$ or $x = -2$.

Usually, though, you will have to solve cubic equations either by drawing an appropriate graph or by trial and improvement.

Example 24.7.1

Draw the graph of $y = x^3 - 4x^2 - 2x + 3$ and use it to solve the equation $x^3 - 4x^2 - 2x + 3 = 0$. Take values of x from –2 to 5.

x	–2	–1	0	1	2	3	4	5
y	–17	0	3	–2	–9	–12	–5	18

The diagram shows the graph with equation $y = x^3 - 4x^2 - 2x + 3$. The x-coordinates of the points where the graph crosses the x-axis are $x = -1$, $x = 0.7$ and $x = 4.3$. Thus, the solutions of the equation $x^3 - 4x^2 - 2x + 3 = 0$ are $x = -1$, $x = 0.7$ and $x = 4.3$. The solution $x = -1$ is exact and the other two are correct to one decimal place.

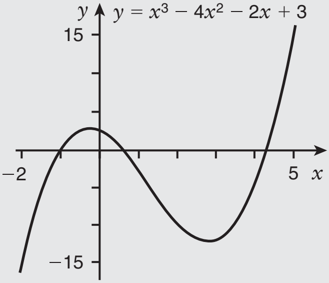

The above cubic equation has three solutions. This is the highest possible number of solutions a cubic equation can have, because a cubic graph can cross the x-axis at most three times. A cubic equation must have at least one solution, because a cubic graph must cross the x-axis at least once.

You can use trial and improvement to solve an equation to any degree of accuracy. For example, you can use it to find, correct to one decimal place, the higher positive solution to the equation $x^3 - 4x^2 - 2x + 3 = 0$, which was solved graphically in Example 24.7.1.

From the table of values and the graph, it is clear that there is a solution between $x = 4$ and $x = 5$. To determine which of these two values the solution is nearer to, evaluate $x^3 - 4x^2 - 2x + 3$ when $x = 4.5$.

x	4	4.5	5
$x^3 - 4x^2 - 2x + 3$	–5	4.125	18

The change in the sign of $x^3 - 4x^2 - 2x + 3$ between $x = 4$ and $x = 4.5$ shows that the solution lies between these two values of x. The next step is to evaluate $x^3 - 4x^2 - 2x + 3$ when $x = 4.1$, 4.2 and so on until there is a sign change.

x	4.1	4.2	4.3	4.4
$x^3 - 4x^2 - 2x + 3$	–3.519	–1.872	–0.053	1.944

The sign change occurs between $x = 4.3$ and $x = 4.4$, so the solution lies between these two values of x. The value of $x^3 - 4x^2 - 2x + 3$ when $x = 4.3$ is very close to 0, suggesting that, correct to one decimal place, the solution is $x = 4.3$ but you must evaluate $x^3 - 4x^2 - 2x + 3$ when $x = 4.35$ to confirm this.

x	4.3	4.35	4.4
$x^3 - 4x^2 - 2x + 3$	–0.053	0.922	1.944

Thus, there is a solution between $x = 4.3$ and $x = 4.35$, so, correct to one decimal place, the solution is $x = 4.3$.

With trial and improvement, it is not necessary to have 0 on one side of the equation.

Example 24.7.2

Solve the equation $x^3 + 2x = 57$, correct to one decimal place.

x	$x^3 + 2x$	*greater or smaller than 57*
1	3	Smaller
2	12	Smaller
3	33	Smaller
4	72	Greater
3.5	49.875	Smaller
3.6	53.856	Smaller
3.7	58.053	Greater
3.65	55.927…	Smaller

From the table, $x = 3.7$ (to one decimal place).

You could obtain a solution correct to two decimal places in the above example by evaluating $x^3 + 2x$ when $x = 3.66$, 3.67 etc. There is a solution between $x = 3.67$ and $x = 3.68$. Finally, evaluating $x^3 + 2x$ when $x = 3.675$ shows that, correct to two decimal places, $x = 3.68$.

EXERCISE 24.7

1 Solve each of these equations by drawing a graph, using the range of values of x given in brackets. Where necessary, give your solutions correct to one decimal place.

(a) $x^3 - 7x^2 + 10x = 0$ (–1 to 6)
(b) $x^3 - 2x^2 - 5x + 6 = 0$ (–3 to 5)
(c) $x^3 - 7x - 5 = 0$ (–4 to 4)
(d) $x^3 - x^2 - 4x + 4 = 0$ (–3 to 3)
(e) $x^3 - 4x^2 + 4x = 0$ (–2 to 4)
(f) $x^3 - 2x^2 - 2 = 0$ (–2 to 3)

2 There is a solution to the equation $x^3 + x - 200 = 0$ between $x = 5$ and $x = 6$. Find this solution correct to one decimal place.

3 There is a solution to the equation $x^3 + 2x^2 - 120 = 0$ between $x = 4$ and $x = 5$. Find this solution correct to one decimal place.

4 Find, correct to one decimal place, a positive solution to the equation $x^3 - 3x - 15 = 0$.

5 Solve the equation $x^3 + 6x = 15$ correct to one decimal place.

6 Solve the equation $x^3 + 4x^2 = 32$ correct to one decimal place.

THINGS TO REMEMBER

- *A pair of equations like* $x + y = 5$ *and* $2x - y = 7$ *are called simultaneous equations. There is only one pair of values (*$x = 4$ *and* $y = 1$*) which satisfy both equations.*

- *You can solve simultaneous equations by plotting the graphs of both equations and finding the coordinates of the point where the graphs intersect.*

- *You can solve a pair of simultaneous equations algebraically by eliminating one of the variables.*

- *A quadratic expression is one in which the highest power of* x *is* x^2*.*

- *To expand a pair of brackets you need to multiply each term in the second bracket by both terms in the first bracket. This will give you an expression with four terms which you can then simplify.*

- *To factorize a quadratic, look for a pair of numbers that:*
  add to give the coefficient of x
  multiply to give the constant term.

  So $x^2 + x - 6 = (x - 2)(x + 3)$ *since* $-2 + 3 = +1$ *and* $-2 \times 3 = -6$

- *A quadratic in the form* $x^2 - a^2$ *is called the difference of two squares.*

  $x^2 - a^2 = (x + a)(x - a)$

- *You can find an estimate for the solution of a quadratic equation by drawing its curve and finding the x-coordinate of each of the points where it cuts the x-axis. There are at most two solutions.*

- *To solve a quadratic equation you should:*
  make sure that the quadratic is equal to 0
  factorize the quadratic into two brackets
  set each bracket equal to 0 and solve to find x.
- *A cubic equation is one in which the highest power of x is* x^3.
  A cubic equation has at most three solutions.
- *You can use trial and improvement to solve an equation.*

Answers

Exercise 1.1 (page 2)

1 **(a)** fifty **(b)** five hundred **(c)** five thousand **(d)** five

2 **(a)** 17 **(b)** 70 **(c)** 97 **(d)** 546 **(e)** 603 **(f)** 810 **(g)** 450 **(h)** 10 000 **(i)** 8934 **(j)** 6480 **(k)** 3006

3 **(a)** fifty-two **(b)** eight hundred and seventy-one **(c)** five thousand, six hundred and twenty-four **(d)** nine hundred and eighty **(e)** seven thousand, and one **(f)** thirty-five thousand, and thirteen **(g)** two hundred and forty-one thousand, and one **(h)** one million, one thousand, three hundred and twelve

4 **(a)** 975, 579 **(b)** 9721, 1279

5 **(a)** three hundred and forty-two thousand, seven hundred and eighty-five **(b)** three million, seven hundred and eighty-three thousand, one hundred and ninety-four **(c)** seventeen million, twenty-one thousand, two hundred and nine **(d)** three hundred and five million, two hundred and thirteen thousand, and ninety-seven

6 **(a)** 516 219 **(b)** 206 024 **(c)** 21 437 869 **(d)** 7 604 013

Exercise 1.2 (page 4)

1 **(a)** 29 **(b)** 48 **(c)** 54 **(d)** 61

2 **(a)** 94 **(b)** 114

3 **(a)** 64 **(b)** 14 **(c)** 516 **(d)** 85 **(e)** 66 **(f)** 78

4 **(a)** 18 **(b)** 135

5 **(a)** 45 **(b)** 301 **(c)** 2574 **(d)** 3556 **(e)** 297 **(f)** 336 **(g)** 182 **(h)** 441

6 2484

7 **(a)** 75 **(b)** 282 **(c)** 243 **(d)** 219

8 **(a)** 2 r 2 **(b)** 4 r 3 **(c)** 12 r 5

Exercise 1.3 (page 6)

1 **(a)** 35 **(b)** 23 **(c)** 8 **(d)** 1 **(e)** 1 **(f)** 4 **(g)** 6 **(h)** 23 **(i)** 26 **(j)** 8 **(k)** 3 **(l)** 70 **(m)** 14 **(n)** 39 **(o)** 3 **(p)** 22 **(q)** 32 **(r)** 17

Exercise 1.4 (page 7)

1 993 pupils
2 168 hours
3 1504
4 23
5 612 miles
6 19 pieces, 3 cm
7 32 boxes
8 189 miles
9 10 220 days
10 23 coaches, 4 seats

Exercise 1.5 (page 10)

1 **(a)** 34, 112, 568, 4366 **(b)** 67, 741
2 1, 4, 9, 16, 25, 36, 49, 64, 81, 100
3 1, 3, 6, 10, 15, 21, 28, 36, 45, 55
4 4, 6, 8, 9, 10, 12, 14, 15, 16, 18
5 1, 8, 27, 64, 125, 216, 343, 512, 729, 1000
6 **(a)** 25, 36 **(b)** 28, 36 **(c)** 20, 25, 28, 33, 36
7 odd
8 **(a)** e.g. $3 + 6 = 9$, $6 + 10 = 16$, $10 + 15 = 25$ **(b)** They are squares.
9 9, 36, 100, 225; they are square numbers.
10 **(a)** e.g. 16 **(b)** e.g. 24 **(c)** e.g. 36

Exercise 1.6 (page 11)

1 7, 14, 21, 28, 35
2 **(a)** 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 **(b)** ends in 5
3 1, 2, 4, 5, 10, 20
4 **(a)** $1 + 2 + 4 + 7 + 14 = 28$ **(b)** same as original number
5 1, 23
6 **(a)** e.g. 9, 16, 25, 36, 49 **(b)** square numbers
7 23, 29, 31, 37
8 Other even numbers have factor 2.
9 10
10 24
11 2, 3, 5
12 5, 7
13 e.g. 12

Exercise 2.1 (page 18)

1 10
2 6°
3 $17\frac{1}{2}$°
4 330°
5 **(a)** 90° **(b)** 270°
6 **(a)** 45° **(b)** 315°
7 **(a)** obtuse **(b)** reflex **(c)** acute **(d)** obtuse **(e)** reflex
8 **(a)** obtuse **(b)** reflex **(c)** reflex **(d)** acute
9 $\angle L$ or $\angle MLN$; $\angle M$ or $\angle LMN$; $\angle A$ or $\angle BAD$; $\angle CBD$.

Exercise 2.2 (page 20)

1 **(a)** 50° **(b)** 132° **(c)** 315° **(d)** 111° **(e)** 231°

Exercise 2.3 (page 22)

1 **(a)** 67° **(b)** 23° **(c)** 111° **(d)** 58° **(e)** 152° **(f)** 72° **(g)** 63° **(h)** 202° **(i)** 129°, 51°
2 No; 37° + 97° + 45° = 179°, not 180°.

Exercise 2.4 (page 24)

1 **(a)** **(i)** d **(ii)** b **(b)** **(i)** f **(ii)** h **(c)** **(i)** k **(ii)** i
2 **(a)** $a = 57°$, $b = 57°$, $c = 123°$ **(b)** $d = 98°$, $e = 82°$ **(c)** $f = 78°$, $g = 28°$, $h = 74°$ **(d)** $i = 66°$, $j = 114°$, $k = 115°$, $l = 115°$ **(e)** $m = 59°$, $n = 41°$, $p = 260°$ **(f)** $q = 65°$, $r = 65°$

Exercise 2.5 (page 26)

1 063°, 158°, 248°, 305°
3 **(a)** 270° **(b)** 045° **(c)** 225° **(d)** 315°
4 **(a)** 149° **(b)** 293° **(c)** 204°
5 319°
6 025°

Exercise 3.1 (page 30)

1 **(a)** $\frac{4}{5}$, $\frac{7}{10}$, $\frac{5}{12}$ **(b)** $\frac{1}{5}$, $\frac{3}{10}$, $\frac{7}{12}$
2 $\frac{1}{6}$
3 $\frac{4}{7}$
4 **(a)** $\frac{5}{6}$ **(b)** $\frac{3}{8}$

5 **(a)** e.g. $\frac{2}{6}, \frac{3}{9}, \frac{4}{12}, \frac{5}{15}, \frac{6}{18}$ **(b)** e.g. $\frac{6}{8}, \frac{9}{12}, \frac{12}{16}, \frac{15}{20}, \frac{18}{24}$
(c) e.g. $\frac{4}{10}, \frac{6}{15}, \frac{8}{20}, \frac{10}{25}, \frac{12}{30}$ **(d)** e.g. $\frac{10}{18}, \frac{15}{27}, \frac{20}{36}, \frac{25}{45}, \frac{30}{54}$

6 $\frac{2}{3} = \frac{16}{24}$ and $\frac{5}{8} = \frac{15}{24}$ so $\frac{2}{3} > \frac{5}{8}$

Exercise 3.2 (page 32)

1 **(a)** $\frac{4}{5}$ **(b)** $\frac{2}{5}$ **(c)** $\frac{7}{9}$ **(d)** $\frac{5}{16}$ **(e)** $\frac{5}{6}$ **(f)** $\frac{5}{8}$ **(g)** $\frac{2}{9}$ **(h)** $\frac{2}{3}$

2 **(a)** $\frac{1}{2}, \frac{7}{12}, \frac{5}{8}$ **(b)** $\frac{7}{10}, \frac{3}{4}, \frac{4}{5}$

3 **(a)** $\frac{4}{8} = \frac{2}{4}$ **(b)** $\frac{6}{18} = \frac{2}{6}$ **(c)** $\frac{20}{24} = \frac{5}{6}$

4 **(a)** $\frac{30}{40} = \frac{6}{8}$ **(b)** $\frac{24}{36} = \frac{4}{6}$ **(c)** $\frac{18}{27} = \frac{2}{3}$

5 **(a)** $\frac{2}{3}$ **(b)** $\frac{3}{4}$ **(c)** $\frac{1}{4}$ **(d)** $\frac{4}{5}$ **(e)** $\frac{6}{7}$ **(f)** $\frac{8}{15}$ **(g)** $\frac{1}{4}$ **(h)** $\frac{3}{5}$

6 **(a)** $\frac{1}{3}$ **(b)** $\frac{1}{2}$ **(c)** $\frac{1}{3}$ **(d)** $\frac{3}{5}$ **(e)** $\frac{5}{6}$ **(f)** $\frac{3}{5}$ **(g)** $\frac{2}{5}$ **(h)** $\frac{3}{5}$

Exercise 3.3 (page 34)

1 **(a)** 5 **(b)** 7 **(c)** 6 **(d)** 7

2 **(a)** $\frac{3}{1}$ **(b)** $\frac{12}{1}$

3 **(a)** $2\frac{3}{5}$ **(b)** $4\frac{3}{4}$ **(c)** $4\frac{1}{6}$ **(d)** $1\frac{7}{8}$ **(e)** $6\frac{2}{3}$ **(f)** $3\frac{3}{5}$ **(g)** $2\frac{1}{8}$ **(h)** $3\frac{2}{9}$

4 **(a)** $3\frac{1}{2}$ **(b)** $3\frac{1}{2}$ **(c)** $1\frac{2}{3}$ **(d)** $2\frac{1}{2}$

5 **(a)** $\frac{14}{5}$ **(b)** $\frac{14}{3}$ **(c)** $\frac{7}{2}$ **(d)** $\frac{23}{4}$

Exercise 3.4 (page 36)

1 **(a)** $\frac{4}{5}$ **(b)** $\frac{6}{7}$ **(c)** $\frac{2}{5}$ **(d)** $\frac{2}{3}$ **(e)** $1\frac{2}{9}$ **(f)** $1\frac{1}{4}$ **(g)** 1 **(h)** $1\frac{3}{5}$

2 **(a)** $\frac{3}{5}$ **(b)** $\frac{2}{7}$ **(c)** $\frac{1}{4}$ **(d)** $\frac{2}{3}$

3 **(a)** $\frac{11}{15}$ **(b)** $\frac{5}{8}$ **(c)** $\frac{2}{3}$ **(d)** $1\frac{1}{2}$ **(e)** $1\frac{4}{15}$ **(f)** $\frac{11}{18}$ **(g)** $1\frac{7}{24}$ **(h)** $1\frac{1}{3}$

4 **(a)** $6\frac{1}{20}$ **(b)** $5\frac{9}{10}$ **(c)** $7\frac{4}{5}$ **(d)** $5\frac{7}{12}$

5 **(a)** $2\frac{5}{8}$ **(b)** $2\frac{1}{6}$ **(c)** $1\frac{1}{12}$ **(d)** $2\frac{5}{6}$

Exercise 3.5 (page 38)

1 **(a)** $3\frac{1}{3}$ **(b)** 12 **(c)** $7\frac{1}{2}$ **(d)** 35

2 **(a)** $\frac{15}{28}$ **(b)** $\frac{7}{12}$ **(c)** $\frac{3}{4}$ **(d)** $\frac{2}{15}$

3 **(a)** $4\frac{3}{8}$ **(b)** $5\frac{1}{4}$ **(c)** 8 **(d)** $1\frac{11}{24}$ **(e)** 2 **(f)** $1\frac{1}{4}$ **(g)** $\frac{3}{5}$ **(h)** 14

4 **(a)** 5 **(b)** 35 **(c)** 27 **(d)** 33

5 **(a)** 32 kg **(b)** 21 cm **(c)** 48 litres

Exercise 3.6 (page 40)

1 **(a)** 6 **(b)** $5\frac{1}{3}$ **(c)** 9 **(d)** $3\frac{1}{3}$

2 **(a)** $2\frac{2}{15}$ **(b)** $\frac{5}{9}$ **(c)** $1\frac{1}{6}$ **(d)** $\frac{27}{40}$ **(e)** $\frac{2}{15}$ **(f)** $\frac{7}{8}$ **(g)** $2\frac{4}{9}$ **(h)** $\frac{16}{21}$

3 **(a)** $1\frac{1}{3}$ **(b)** $\frac{33}{40}$ **(c)** 6 **(d)** $\frac{2}{3}$ **(e)** $2\frac{1}{40}$ **(f)** $\frac{7}{9}$ **(g)** $\frac{5}{24}$ **(h)** $3\frac{9}{10}$

Exercise 3.7 (page 41)

1 **(a)** $\frac{2}{15}$ **(b)** $\frac{5}{8}$ **(c)** $\frac{5}{14}$ **(d)** $\frac{2}{5}$ **(e)** $\frac{5}{12}$ **(f)** $\frac{2}{7}$ **(g)** $\frac{3}{5}$ **(h)** $\frac{3}{16}$ **(i)** $\frac{7}{20}$ **(j)** $\frac{1}{12}$

2 $\frac{2}{3}$

3 $\frac{7}{10}$

4 $\frac{4}{15}$

5 $\frac{5}{16}$

6 $3\frac{2}{5}$

7 $\frac{7}{12}$

8 240°

9 9

Exercise 4.1 (page 47)

1 $a = 81°$

2 $b = 73°$, $c = 34°$

3 $d = 63°$

4 $e = 60°$, $f = 120°$

5 $g = 31°$

6 $h = 45°$

7 $i = 58°$, $j = 61°$

8 $k = 66°$

9 $l = 68°$, $m = 68°$

10 $x = 74°$, $y = 32°$ (interchangeable)* or $x = 53°$, $y = 53°$

11 140°, 20°, 20°

Exercise 4.2 (page 49)

8 Cannot be done – arcs do not cross. The sum of the lengths of the two smallest sides must be greater than the length of the largest side.

Exercise 4.3 (page 51)

1 112°

2 two from rectangle, parallelogram, rhombus

3 square, rhombus, kite

4 0, 1, 2, 4

5 1, 2, 4

---

* $x = 32°$, $y = 74°$ is the same pair

Exercise 4.4 (page 54)

1 126°
2 155°
3 127°
4 1800°
5 2340°
6 108°
7 144°
8 160°
9 71°
10 60°
11 **(a)** 30° **(b)** 150°
12 9
13 30

Exercise 4.5 (page 56)

1 A and E, B and H, C and G, D and J.

Exercise 4.6 (page 58)

2 Each interior angle of a rectangle is 90°, so four corners fit to make 360°.
3 Each interior angle of a regular nonagon, 140°; 140 does not divide 360 exactly.
5 Each interior angle of an equilateral triangle is 60° and of a regular 12-gon is 150°. $60 + 2 \times 150 = 360$, so the two shapes might tessellate. They do.

Exercise 5.1 (page 65)

1 **(a)** $\frac{3}{1000}$ **(b)** 30 **(c)** $\frac{3}{10}$ **(d)** $\frac{3}{100}$
2 **(a)** 3 **(b)** $\frac{8}{10}$ **(c)** $\frac{9}{100}$ **(d)** $\frac{1}{100}$
3 **(a)** 2 **(b)** 7
4 **(a)** 3 **(b)** 0
5 **(a)** 0.901 **(b)** 0.72 **(c)** 0.01 **(d)** 0.909
6 5.632
7 0.678
8 1.203, 1.23, 1.302, 1.32, 2.31
9 0.019, 0.109, 0.190, 0.901, 0.910
10 **(a)** $4.76 < 6.47$ **(b)** $5.72 > 5.27$ **(c)** $0.39 < 0.4$ **(d)** $0.03 > 0.029$

Exercise 5.2 (page 67)

1 **(a)** $\frac{3}{5}$ **(b)** $\frac{1}{10}$ **(c)** $\frac{12}{25}$ **(d)** $\frac{39}{100}$ **(e)** $\frac{9}{100}$ **(f)** $\frac{179}{1000}$ **(g)** $\frac{1}{8}$ **(h)** $\frac{3}{1000}$
2 **(a)** $3\frac{4}{5}$ **(b)** $4\frac{11}{50}$ **(c)** $54\frac{3}{4}$ **(d)** $23\frac{6}{125}$
3 **(a)** 0.3 **(b)** 0.71 **(c)** 0.07 **(d)** 0.387 **(e)** 0.017 **(f)** 0.009 **(g)** 0.4 **(h)** 0.95 **(i)** 0.68 **(j)** 0.22 **(k)** 0.474 **(l)** 0.035
4 **(a)** 6.1 **(b)** 9.03 **(c)** 5.041 **(d)** 1.45 **(e)** 8.32 **(f)** 12.185

Exercise 5.3 (page 69)

1 **(a)** 7.9 **(b)** 13.12 **(c)** 10.3 **(d)** 24.07 **(e)** 6.3 **(f)** 16.8 **(g)** 7.17 **(h)** 14.83 **(i)** 10.7 **(j)** 2.2 **(k)** 11.88 **(l)** 3.37 **(m)** 3.851 **(n)** 3.488 **(o)** 12.085
2 0.77
3 34.8
4 10.28 s

Exercise 5.4 (page 72)

1 **(a)** 28.2 **(b)** 89 **(c)** 146.96 **(d)** 1.209
2 **(a)** 82.1 **(b)** 6 **(c)** 4.7 **(d)** 7.02 **(e)** 283 **(f)** 0.041 **(g)** 521.8 **(h)** 83 **(i)** 40 **(j)** 0.07 **(k)** 46 360 **(l)** 3900
3 **(a)** $0.21 \times 100 = 21$ **(b)** $6.71 \times 10 = 67.1$ **(c)** $7.98 \times 1000 = 7980$
4 **(a)** 0.63 **(b)** 0.3 **(c)** 0.08 **(d)** 0.03 **(e)** 0.021 **(f)** 0.0006
5 **(a)** 5.18 **(b)** 0.1148 **(c)** 2.516 **(d)** 0.039 48 **(e)** 344.1 **(f)** 1.127
6 10.56 pints
7 36 litres
8 832 cm

Exercise 5.5 (page 74)

1 **(a)** 2.3 **(b)** 6.41 **(c)** 0.397 **(d)** 0.067
2 **(a)** 3.75 **(b)** 45.72 **(c)** 0.0093 **(d)** 0.713 **(e)** 0.034 **(f)** 6.528 **(g)** 0.025 **(h)** 0.0037 **(i)** 0.000 03 **(j)** 0.172 **(k)** 0.0338 **(l)** 0.000 071
3 **(a)** $3.7 \div 10 = 0.37$ **(b)** $0.9 \div 100 = 0.009$ **(c)** $7.9 \div 100 = 0.079$
4 **(a)** 19.96 **(b)** 28.5 **(c)** 0.156 **(d)** 0.91 **(e)** 3.68 **(f)** 0.036 **(g)** 0.875 **(h)** 0.006545
5 **(a)** 0.9 **(b)** 15 **(c)** 80 **(d)** 0.03 **(e)** 300 **(f)** 50 **(g)** 24.6 **(h)** 37.6
6 4.62 m
7 0.021 cm (or 0.21 mm)
8 25

Exercise 5.6 (page 76)

1 **(a)** 0.25 terminating **(b)** $0.\dot{3}$ recurring **(c)** 0.375 terminating **(d)** $0.\dot{4}$ recurring **(e)** 0.875 terminating

2 **(a)** $0.\dot{8}\dot{1}$ **(b)** $0.\dot{0}\dot{9}$ **(c)** $0.41\dot{6}$ **(d)** $0.08\dot{3}$ **(e)** $0.\dot{5}\dot{7}$ **(f)** $0.\dot{0}8\dot{1}$ **(g)** $0.3\dot{4}\dot{8}$ **(h)** $0.5\dot{2}7\dot{0}$

3 **(a)** $0.\dot{8}$ **(b)** $0.7\dot{1}\dot{2}$ **(c)** $0.\dot{3}6\dot{7}$ **(d)** $0.45\dot{6}8\dot{9}$ **(e)** $18.0\dot{1}\dot{8}$ **(f)** $0.\dot{5}64\dot{3}$ **(g)** $9.3\dot{9}$ **(h)** $263.\dot{6}\dot{3}$

4 **(a)** $\frac{7}{11} > \frac{8}{13}$ **(b)** $\frac{14}{19} < \frac{17}{23}$ **(c)** $\frac{6}{13} < \frac{7}{15}$ **(d)** $\frac{13}{17} > \frac{16}{21}$

5 **(a)** 5.77, $5\frac{17}{22}$, $5\frac{7}{9}$, 5.78, 5.8 **(b)** 3.01, 3.09, $3\frac{1}{11}$, $3\frac{2}{21}$, 3.1 **(c)** 4.083, $4\frac{1}{12}$, 4.084, $4\frac{2}{23}$, 4.085

6 0.516, $\frac{129}{250}$

7 **(a)** $0.\dot{1}4285\dot{7}$, $0.\dot{2}8571\dot{4}$, $0.\dot{4}2857\dot{1}$, $0.\dot{5}7142\dot{8}$, $0.\dot{7}1428\dot{5}$, $0.\dot{8}5714\dot{2}$ **(b)** 6 **(c)** Same six digits in each recurring group.

8 **(a)** 2, 4, 5, 8, 10, 16, 20, 25 **(b)** Each denominator is a power of 2, or a power of 5, or a product of powers of 2 and 5.

Exercise 6.1 (page 84)

1 14, 4, 5, 9

2 8, 7, 5, 4, 2

6 20, 21, 56, 65

10 13, 33, 59, 10, 7, 63

Exercise 6.2 (page 88)

1 120°, 105°, 90°, 15°, 30°

2 66°, 54°, 114°, 84°, 42°

3 189°, 104°, 66°

4 **(a)** 25 min **(b)** 105 min **(c)** $\frac{11}{72}$

Exercise 6.3 (page 91)

1 **(b)** 6.5 cm

2 **(b)(i)** 17°C **(ii)** 15–16 min

3 **(b)** 97 cm

4 **(b)(i)** 120 million **(ii)** 1970

Exercise 6.4 (page 95)

1 **(c)** low positive

2 **(c)** high negative **(d)** 27 units

3 no correlation

Exercise 6.5 (page 96)

1 discrete: **(a)**, **(d)**, **(e)**, **(h)**, **(j)**
continuous: **(b)**, **(c)**, **(f)**, **(g)**, **(i)**

Exercise 6.6 (page 100)

1 **(a)** 14, 7, 5, 3, 1
2 **(a)** 10, 7, 5, 3
4 **(a)** 4, 6, 5, 3, 2

Exercise 7.1 (page 105)

1 –7°C
2 –9
3 **(a)** $-3 < -1$ **(b)** $0 > -2$ **(c)** $-12 > -15$
4 **(a)** 2°C **(b)** –8°C
5 –4, –3 –1, 0, 2, 5

Exercise 7.2 (page 107)

1 5°C
2 –4°C
3 11
4 7
5 **(a)** 1 **(b)** –3 **(c)** 0 **(d)** –3 **(e)** –5 **(f)** –1 **(g)** –9 **(h)** –5 **(i)** 0 **(j)** –4
6 **(a)** 1 **(b)** 7 **(c)** –9 **(d)** –1 **(e)** 6 **(f)** 0 **(g)** –10 **(h)** –7

Exercise 7.3 (page 110)

1 **(a)** –8 **(b)** 12 **(c)** 0 **(d)** –12 **(e)** 16
2 **(a)** –20 **(b)** 24 **(c)** –21 **(d)** 0 **(e)** 36 **(f)** –64 **(g)** 35 **(h)** 0 **(i)** –72
3 **(a)** –6 **(b)** 8 **(c)** –3 **(d)** 5 **(e)** 0 **(f)** –6 **(g)** –4 **(h)** 4 **(i)** –7

Exercise 7.4 (page 111)

1 **(a)** –408 **(b)** 56 **(c)** –139 **(d)** –26 **(e)** –5.1 **(f)** –1.6 **(g)** –2.66 **(h)** 4.8

Exercise 8.1 (page 115)

1 A (4, 2), B (1, 3), C (2, 2), D (0, 1), E (3, 0)
3 **(c)** (6, 1)
4 A (–4, 2), B (3, 1), C (–3, –1), D (2, –3), E (0, –3)
6 **(b)** kite
7 **(b)** (1, –2)
8 (–1, 3)

Exercise 8.2 (page 118)

1 **(a)(i)** 0.4 **(ii)** 1.4 **(b)(i)** 12 **(ii)** 18
2 **(a)(i)** 45 km **(ii)** 67 km **(b)(i)** 22 miles **(ii)** 55 miles
3 **(a)(i)** €55 **(ii)** €38 **(b)(i)** £52 **(ii)** £33
4 **(a)** 1315 **(b)** 45 mins **(c)** 40 mph **(d)** 48 mph
5 **(b)** 81 gallons **(c)** 27 gallons
6 **(b)** $33 **(c)** £27
7 **(b)** 18 km/h **(c)** 20 km/h
8 **(a)** 40 60 80 100 120 140 **(c)** 104 mins

Exercise 8.3 (page 123)

1 $P{:}x = 2$; $Q{:}x = -3$; $R{:}y = 3$; $S{:}y = -2$
3 **(a)** $x < -2$ **(b)** $-1 < y \leq 2$

Exercise 9.1 (page 127)

1 **(a)** 7.9 cm **(b)** 8.5 m **(c)** 3.24 kg **(d)** 4.125 l **(e)** 1800 m **(f)** 0.4 km **(g)** 94 mm **(h)** 940 ml **(i)** 2300 g **(j)** 3720 kg
2 **(a)** 5.3 cm **(b)** 53 mm
3 120
4 55.66 m
5 5.37 kg
6 33 days

Exercise 9.2 (page 128)

1 **(a)** 96 in **(b)** 5 gallons **(c)** 3 lb **(d)** 56 pints **(e)** 8 yd **(f)** 5280 yd **(g)** 70 in **(h)** 55 oz **(i)** 3 ft 7 in **(j)** 3 lb 9 oz
2 6 ft 2 in
3 8 in
4 14 gallons
5 4920 pt
6 1980

Exercise 9.3 (page 130)

1 **(a)** 15 cm **(b)** 14 in **(c)** 12 ft **(d)** 72 km **(e)** 6 oz **(f)** 11 lb **(g)** 21 pt **(h)** 12 gallons **(i)** 60 miles **(j)** 8 l
2 80 cm
3 17 oz
4 42 in
5 8 gallons

6 187 lb
7 192 km
8 5 ft 10 in
9 155 miles
10 5
11 16 km

Exercise 9.4 (page 131)

1 **(a)(i)** metres **(ii)** feet **(b)(i)** grams **(ii)** ounces **(c)(i)** litres **(ii)** pints **(d)(i)** kilometres **(ii)** miles **(e)(i)** litres **(ii)** gallons **(f)(i)** kilograms **(ii)** pounds
2 **(a)** grams **(b)** litres **(c)** grams **(d)** litres **(e)** millilitres

Exercise 10.1 (page 135)

1 17 cm
2 27 cm
3 37.2 cm
4 23 cm
5 36 cm
6 **(a)** 328 m **(b)** 366 m
7 **(a)** 14 **(b)** 16
8 20
9 **(a)** 14 cm **(b)** 10 cm, 12 cm, 14 cm **(c)** 14 cm
10 **(a)** 10 **(b)** 9 **(c)** 9
11 **(a)** 11 **(b)** 13
12 11

Exercise 10.2 (page 137)

1 **(a)** 6 cm^2 **(b)** 4 cm^2 **(c)** $3\frac{1}{2}$ cm^2 **(d)** 4 cm^2 **(e)** 6 cm^2 **(f)** 5 cm^2
2 **(a)** 6 cm^2 **(b)** 6 cm^2 **(c)** 9 cm^2 **(d)** 9 cm^2 **(e)** 16 cm^2 **(f)** 7 cm^2
3 28 cm^2 approximately

Exercise 10.3 (page 139)

1 84 cm^2
2 48 m^2
3 81 cm^2
4 **(a)** 1.38 m^2 **(b)** 13 800 cm^2
5 **(a)** 1 m^2 **(b)** 10 000 cm^2
6 **(a)** 22 cm^2 **(b)** 50 cm^2 **(c)** 31 cm^2
7 **(a)** 44 cm^2 **(b)** 38 cm^2

8 **(a)** 54 m^2 **(b)** 20 m^2 **(c)** 7 litres
9 9 m^2
10 36 m^2
11 10 000 m^2
12 42

Exercise 10.4 (page 143)

1 **(a)** 24 cm^2 **(b)** 6 cm^2 **(c)** 16 cm^2 **(d)** 36 cm^2 **(e)** 6 cm^2 **(f)** 35 cm^2 **(g)** 47 cm^2 **(h)** 28 cm^2
2 **(a)** 5.4 cm^2 **(b)** 3.6 cm^2
3 **(a)** 33 cm^2 **(b)** 50 cm^2
4 $5\frac{1}{2}$ cm^2
5 40 cm^2
6 60 cm^2

Exercise 10.5 (page 146)

1 **(a)** 55 cm^2 **(b)** 66.5 cm^2 **(c)** 24 cm^2 **(d)** 12 cm^2
2 6 cm^2
3 **(a)** 10 **(b)** 10.5 **(c)** 16
4 **(a)** 37 cm^2 **(b)** 53.5
5 6
6 8

Exercise 11.1 (page 150)

1 **(a)** $x + 2$ **(b)** $x - 1$ **(c)** $4x$ **(d)** $\frac{1}{2}x$
2 **(a)** $p + q$ **(b)** pq
3 $\frac{8}{x}$
4 $a - b$
5 $30 - b$
6 **(a)** $6x$ **(b)** $\frac{1}{5}y$ **(c)** xy **(d)** $3xy$ **(e)** $\frac{x}{y}$ **(f)** $\frac{5x}{y}$ **(g)** $\frac{1}{5}xy$ **(h)** xyz
7 $7w$
8 $\frac{1}{100}d$ or $\frac{d}{100}$
9 np
10 $\frac{a}{b}$

Exercise 11.2 (page 152)

1 $5l$
2 **(a)** $2a + 2b$ **(b)** ab

3 $2x + 4y$
4 $4a + b$
5 **(a)** $2x$ **(b)** $4x$ **(c)** $7x$ **(d)** $3x$ **(e)** $9x$ **(f)** y **(g)** $-2xy$ **(h)** $-abc$
(i) $9x + 5y$ **(j)** $2x + 3$ **(k)** $7x + 3$ **(l)** $5x$ **(m)** $9xy + x - 3$
(n) $7a - 2b + 6$ **(o)** $8a + ab + 5$ **(p)** $6a$ **(q)** $2 - 4pq + r$ **(r)** $p + q$

Exercise 11.3 (page 154)

1 **(a)** 19 **(b)** 5 **(c)** 1 **(d)** –5
2 **(a)** 15 **(b)** 6 **(c)** 51 **(d)** –16
3 **(a)** 16 **(b)** 60 **(c)** 300 **(d)** 34

Exercise 11.4 (page 155)

1 **(a)** 9 **(b)** 10 **(c)** 45 **(d)** –3
2 **(a)** 16 **(b)** 13 **(c)** 37 **(d)** 40
3 **(a)** 85 **(b)** 22 **(c)** 76 **(d)** 9

Exercise 11.5 (page 157)

1 **(a)** $6x + 18$ **(b)** $4x - 4$ **(c)** $10x + 15$ **(d)** $6x - 10$ **(e)** $-4x - 12$
(f) $-7x + 14$ **(g)** $-40x - 32$ **(h)** $-18x + 12$ **(i)** $x^2 + 2x$
(j) $4x^2 - 3x$ **(k)** $6x^2 - 3x$ **(l)** $-15x^2 - 20x$
2 **(a)** $9x - 21$ **(b)** $10 - 10x$ **(c)** $9x - 7$ **(d)** $x + 22$ **(e)** $18x - 13$
(f) $21 - x$ **(g)** $26x$ **(h)** 6 **(i)** $x^2 - 4x - 24$ **(j)** $x^2 + 24$
(k) $x^2 + x - 12$ **(l)** $6x^2 + 6x + 28$ **(m)** $23x + 2$ **(n)** $18x + 1$
(o) –6 **(p)** $1 + 5x$ **(q)** –5 **(r)** x

Exercise 11.6 (page 159)

1 **(a)** $3(x + 3)$ **(b)** $7(2x - 3)$ **(c)** $5(4x + 3)$ **(d)** $3(2x - 5y)$
(e) $a(x - 7)$ **(f)** $x(x + 4)$ **(g)** $6(2x - 3)$ **(h)** $3(5x - 4)$
(i) $2x(2a + 3b)$ **(j)** $3x(3x + 4)$ **(k)** $4x(5 - 2x)$ **(l)** $2bx(2 - x)$
(m) $xy(x + y)$ **(n)** $3ax(2x - 5a)$ **(o)** $4xy(2 + 3y)$

Exercise 11.7 (page 160)

1 **(a)** m^3 **(b)** n^5 **(c)** $5p^5$ **(d)** $3q^6$
2 **(a)** 243 **(b)** 135
3 **(a)** 1 000 000 **(b)** 70 000
4 **(a)** –64 **(b)** –3072
5 **(a)** –500 **(b)** 3750

Exercise 11.8 (page 163)

1 **(a)** x^7 **(b)** x^5 **(c)** x^{12} **(d)** x^6 **(e)** x^5 **(f)** x^{10} **(g)** x^{10} **(h)** x^3y^4 **(i)** x^4 **(j)** x^5 **(k)** x^2 **(l)** x^4 **(m)** x **(n)** x^{10} **(o)** x^2 **(p)** x^9

Exercise 11.9 (page 165)

1 **(a)** $6x^5$ **(b)** $20x^3$ **(c)** $14x^5y^6$ **(d)** $9x^8$ **(e)** $6x^9$ **(f)** $8x^7y^5$ **(g)** $3x^2$ **(h)** $7x^3$ **(i)** $5x^2y^2$
2 **(a)** $3xy^3$ **(b)** $6x^5y^2$ **(c)** $\frac{5x^4y^3}{z^2}$ **(d)** $\frac{1}{a^5}$ **(e)** $\frac{8}{x^4}$ **(f)** $\frac{4x^4}{y^2}$ **(g)** $\frac{y^4}{3x}$ **(h)** $5y^4$ **(i)** $36x^2$ **(j)** $25x^2y^2$ **(k)** $x^{10}y^6$ **(l)** $25x^6y^8$ **(m)** $1000x^3$ **(n)** $64x^{15}$ **(o)** $125x^{12}y^9$

Exercise 12.1 (page 171)

1 **(a)** 30 **(b)** 50 **(c)** 80 **(d)** 750 **(e)** 8470 **(f)** 7400
2 **(a)** 700 **(b)** 800 **(c)** 400 **(d)** 4400 **(e)** 3000 **(f)** 19 500
3 **(a)** 8000 **(b)** 3000 **(c)** 6000 **(d)** 38 000 **(e)** 42 000 **(f)** 100 000
4 61 300
5 **(a)** 242 000 **(b)** 200 000
6 **(a)** 7550 **(b)** 7649

Exercise 12.2 (page 173)

1 **(a)** 9 **(b)** 21 **(c)** 68 **(d)** 260 **(e)** 400
2 **(a)** 9.9 **(b)** 0.9 **(c)** 18.0 **(d)** 76.9 **(e)** 400.0
3 **(a)** 4.25 **(b)** 8.47 **(c)** 0.75 **(d)** 0.06
4 **(a)** 8.678 **(b)** 7.609 **(c)** 13.420 **(d)** 0.009
5 10
6 2.1
7 17.41
8 0.429
9 7.25
10 5.845

Exercise 12.3 (page 175)

1 **(a)** 20 **(b)** 400 **(c)** 0.2 **(d)** 0.007 **(e)** 0.09
2 **(a)** 8700 **(b)** 740 **(c)** 6.9 **(d)** 0.085 **(e)** 0.0040
3 **(a)** 23 700 **(b)** 376 **(c)** 13.8 **(d)** 8.93 **(e)** 0.0844 **(f)** 0.000 437 **(g)** 0.0902 **(h)** 0.003 70
4 2000
5 5.9
6 **(a)** 11.3 **(b)** 0.480

7 0.0909
8 **(a)** 600 **(b)** 600 **(c)** 600
9 **(a)** 0.06 **(b)** 0.060 **(c)** 0.0600

Exercise 12.4 (page 177)

1 **(a)** 8, 8.9386 **(b)** 90, 98.7953 **(c)** 5, 5.32 **(d)** 10 000, 11 811.38
(e) 5, 5.57 **(f)** 4, 3.29 **(g)** 8400, 8194.4896
(h) 12, 13.6 **(i)** 21, 23.3531 **(j)** 0.3, 0.341 **(k)** 200, 197
(l) 16, 13.3104 **(m)** 0.000 14, 0.000 159 **(n)** 20, 21.8
(o) 0.2, 0.248 **(p)** 40, 50.8 **(q)** 0.02, 0.0228 **(r)** 2, 1.52
(s) 4, 3.86 **(t)** 100, 122

Exercise 12.5 (page 178)

1 41, 4
2 8, 4 cm
3 4
4 3
5 6
6 6
7 10
8 8

Exercise 12.6 (page 180)

1 324.5 km, 325.5 km
2 12.65 cm, 12.75 cm
3 9.845 s, 9.855 s
4 **(a)** 0.1 kg **(b)** 10.55 kg, 10.65 kg **(c)** 0.05 kg
5 **(a)** 0.01 s **(b)** 0.005 s

Exercise 13.1 (page 183)

1 **(a)** 5 **(b)** 11 **(c)** 9 **(d)** 12 **(e)** 7 **(f)** 4 **(g)** 2 **(h)** 4 **(i)** 7
2 **(a)** 13 **(b)** 17 **(c)** 7 **(d)** 27 **(e)** 6
3 **(a)** 4 **(b)** 6 **(c)** 4 **(d)** 12

Exercise 13.2 (page 185)

1 **(a)** 3 **(b)** 8 **(c)** 6 **(d)** 10 **(e)** –5 **(f)** $\frac{3}{5}$ **(g)** 0 **(h)** $2\frac{1}{3}$ **(i)** 0 **(j)** –5
(k) 2 **(l)** –6

Exercise 13.3 (page 187)

1 **(a)** 4 **(b)** 3 **(c)** $\frac{1}{2}$ **(d)** -1 **(e)** $2\frac{1}{3}$ **(f)** $-\frac{3}{4}$ **(g)** 4 **(h)** 0 **(i)** $2\frac{1}{2}$ **(j)** $\frac{5}{6}$ **(k)** -2 **(l)** $-2\frac{2}{3}$

2 **(a)** 4 **(b)** 5 **(c)** $\frac{1}{2}$ **(d)** -2 **(e)** $1\frac{2}{3}$ **(f)** $-\frac{2}{5}$ **(g)** $-2\frac{1}{2}$ **(h)** $\frac{2}{3}$ **(i)** 0 **(j)** -3 **(k)** $-\frac{2}{3}$ **(l)** $2\frac{3}{4}$

3 **(a)** 5 **(b)** -5 **(c)** 2 **(d)** $\frac{3}{5}$ **(e)** -2 **(f)** $2\frac{1}{3}$ **(g)** $-\frac{1}{2}$ **(h)** $-1\frac{1}{2}$ **(i)** 1 **(j)** $\frac{4}{5}$ **(k)** -1 **(l)** 0 **(m)** $2\frac{3}{5}$ **(n)** $-\frac{2}{5}$ **(o)** 3 **(p)** -4

Exercise 13.4 (page 189)

1 **(a)** 3 **(b)** 2 **(c)** -3 **(d)** $\frac{2}{3}$ **(e)** 12 **(f)** $1\frac{1}{2}$ **(g)** $1\frac{1}{3}$ **(h)** $-\frac{3}{4}$ **(i)** -2 **(j)** 2 **(k)** $-2\frac{2}{3}$ **(l)** $\frac{7}{8}$ **(m)** $1\frac{1}{4}$ **(n)** $-\frac{1}{4}$ **(o)** $1\frac{1}{2}$ **(p)** 5 **(q)** -7 **(r)** $\frac{1}{2}$ **(s)** 0 **(t)** -2

Exercise 13.5 (page 190)

1 12
2 8
3 19
4 11
5 35
6 32, 8
7 10 cm
8 33°
9 8, 9, 10
10 5

Exercise 13.6 (page 193)

1 **(a)** $x < 3$ **(b)** $x \geq -5$ **(c)** $x \leq 3$ **(d)** $x \leq 3$ **(e)** $x \geq -1$ **(f)** $x > 7$ **(g)** $x < -1$ **(h)** $x < \frac{1}{2}$ **(i)** $x \leq -1$ **(j)** $x > 9$ **(k)** $x \leq -3$ **(l)** $x > 1\frac{1}{2}$ **(m)** $x \leq 2\frac{1}{2}$ **(n)** $x < -\frac{1}{2}$ **(o)** $x \leq -2\frac{1}{2}$

Exercise 14.1 (page 198)

1 **(a)** $\frac{49}{100}$ **(b)** $\frac{1}{2}$ **(c)** $\frac{1}{4}$ **(d)** $\frac{1}{10}$ **(e)** $\frac{13}{20}$ **(f)** $\frac{1}{100}$ **(g)** $\frac{3}{100}$ **(h)** 1 **(i)** $2\frac{1}{2}$ **(j)** $\frac{1}{8}$ **(k)** $\frac{1}{3}$ **(l)** $\frac{11}{80}$

2 **(a)** 0.39 **(b)** 0.07 **(c)** 0.3 **(d)** 1.12 **(e)** 0.038 **(f)** 0.625 **(g)** 0.035 **(h)** 0.1225

3 **(a)** 71% **(b)** 4% **(c)** 80% **(d)** 175% **(e)** 190% **(f)** 2.7% **(g)** 3.25% **(h)** 0.9%

4 **(a)** 43% **(b)** 9% **(c)** 70% **(d)** 62% **(e)** 68% **(f)** 75% **(g)** 55% **(h)** 37.5%

5 **(a)(i)** 0.52 **(ii)** $\frac{13}{25}$ **(b)** 48%

6 **(a)** 55% **(b)** $\frac{11}{20}$

Exercise 14.2 (page 200)

1 **(a)** 18 **(b)** 12 **(c)** 27 kg **(d)** 25 m **(e)** 9 cm **(f)** 9 l

2 **(a)** 24 **(b)** 115 **(c)** 24.32 **(d)** 38 **(e)** 64.8 **(f)** 102 **(g)** 965 **(h)** 782 kg **(i)** 124.8 l **(j)** 2.79 km **(k)** 19.8 cm **(l)** 46.5 m

3 **(a)** 684 **(b)** 43% **(c)** 516

4 68

5 792 min

6 **(a)** £795.80 **(b)** £332.28 **(c)** £163.80

7 £1.78

8 **(a)** £1230 **(b)** 85%

Exercise 14.3 (page 201)

1 **(a)** 78 **(b)** 436 **(c)** 873.6 **(d)** 6384 **(e)** 78.4 **(f)** 772.5

2 **(a)** 49 **(b)** 387 **(c)** 228 **(d)** 2720 **(e)** 623.2 **(f)** 36.9

3 754

4 £467.50

5 £2444

6 £11 158.40

7 £658.35

8 £113.75

Exercise 14.4 (page 203)

1 84%

2 60%, 40%

3 55.3%

4 71%

5 **(a)** 36% **(b)** 40% **(c)** 75% **(d)** 35.4% **(e)** 45% **(f)** 17% **(g)** 19.4% **(h)** 11.7% **(i)** 11.9%

Exercise 14.5 (page 204)

1 8%
2 12%
3 4%
4 12.5%
5 7%
6 75%
7 34%
8 28%

Exercise 14.6 (page 207)

1 **(a)** 0.69 **(b)** 1.43 **(c)** 0.7 **(d)** 1.07 **(e)** 1.175 **(f)** 0.966
2 £169
3 57.8 million
4 £9159.20
5 **(a)** £554.36 **(b)** £54.36
6 £750
7 £175
8 £12 500

Exercise 15.1 (page 210)

1 **(a)** $A = \frac{1}{2}bh$ **(b)** 28
2 **(a)** $s = \frac{d}{t}$ **(b)** 56
3 **(a)** $S = 180(n - 2)$ **(b)** 1260
4 **(a)** $P = 8d$ **(b)** 29.6
5 **(a)** $A = l^2$ **(b)** 73.96 cm^2
6 **(a)** $m = \frac{1}{3}(a + b + c)$ **(b)** 35
7 **(a)** $A = \frac{1}{2}(a + b)h$ **(b)** 58 cm^2
8 **(a)** $x = \frac{360}{x}$ **(b)** 40
9 55
10 8
11 –180
12 336

Exercise 15.2 (page 212)

1 14
2 12
3 2.5
4 7
5 6
6 4
7 126
8 –3
9 5
10 3
11 5 or –5
12 4 or –4
13 10 or –10
14 6 or –6

Exercise 15.3 (page 214)

In some parts, your answer could take a different form and still be correct.

1 **(a)** $R = \frac{V}{I}$ **(b)** $b = \frac{V}{h}$ **(c)** $h = \frac{E}{mg}$ **(d)** $y = P - 2x$ **(e)** $T = \frac{D}{S}$
(f) $h = \frac{3V}{A}$ **(g)** $R = \frac{100I}{PT}$ **(h)** $V = \frac{kT}{P}$ **(i)** $a = p - b - c$
(j) $x = \frac{1}{2}(P-y)$ **(k)** $u = v + gt$ **(l)** $a = \frac{v-u}{t}$ **(m)** $t = \frac{u-v}{g}$
(n) $l = \frac{1}{2}(P-2b)$ **(o)** $v = \frac{I+mu}{m}$ **(p)** $n = \frac{1}{180}(S+360)$
(q) $a = 2s - b - c$ **(r)** $a = \frac{2A-bh}{h}$

Exercise 16.1 (page 220)

1 11.6 cm
2 39.8 m
3 24.2 cm
4 15.4 m
5 40 100 km
6 10.5 m
7 **(a)** 2.23 m **(b)** 448
8 31.8 m

Exercise 16.2 (page 223)

1 $2170\ \text{cm}^2$

2 $616\ \text{m}^2$

3 5.64 cm

4 8.59 cm

5 $83.9\ \text{m}^2$

6 $434\ \text{mm}^2$

7 $80.2\ \text{cm}^2$

8 $28.3\ \text{m}^2$

9 4

10 $3180\ \text{m}^2$

Exercise 16.3 (page 225)

1 27°

2 $b = 29°$, $c = 61°$

3 58°

4 68°

5 61°

6 53°

Exercise 17.1 (page 231)

1 **(a)** $\frac{29}{50}$ **(b)** 0.58 **(c)** 58%

2 **(a)** $\frac{17}{40}$ **(b)** $\frac{23}{40}$

3 $\frac{11}{50}$

4 $\frac{1}{5}$

5 $\frac{3}{10}$

Exercise 17.2 (page 233)

1 **(a)** $\frac{1}{6}$ **(b)** $\frac{1}{2}$ **(c)** $\frac{1}{2}$ **(d)** $\frac{1}{3}$

2 **(a)** $\frac{1}{5}$ **(b)** $\frac{3}{5}$ **(c)** $\frac{2}{5}$ **(d)** 0 **(e)** $\frac{4}{5}$ **(f)** $\frac{4}{5}$

3 **(a)** $\frac{3}{10}$ **(b)** $\frac{7}{10}$ **(c)** $\frac{7}{10}$ **(d)** 1

4 **(a)** $\frac{1}{52}$ **(b)** $\frac{1}{13}$ **(c)** $\frac{1}{4}$ **(d)** $\frac{1}{2}$ **(e)** $\frac{2}{13}$ **(f)** $\frac{4}{13}$

5 **(a)** $\frac{1}{11}$ **(b)** $\frac{2}{11}$ **(c)** 0 **(d)** $\frac{4}{11}$ **(e)** $\frac{3}{11}$ **(f)** $\frac{9}{11}$

6 **(a)** $\frac{1}{500}$ **(b)** $\frac{1}{100}$

7 $\frac{7}{9}$

8 0.3

Exercise 17.3 (page 237)

1 **(a)** $\frac{1}{6}$ **(b)** $\frac{11}{36}$ **(c)** $\frac{1}{6}$

2 **(a)** $\frac{1}{36}, \frac{2}{36}, \frac{3}{36}, \frac{4}{36}, \frac{5}{36}, \frac{6}{36}, \frac{5}{36}, \frac{4}{36}, \frac{3}{36}, \frac{2}{36}, \frac{1}{36}$ **(b)** 7 **(c)** Sum of probabilities is 1. The sum of the numbers on the dice is certain to be a whole number between 2 and 12 inclusive.

3 **(a)** $\frac{1}{12}$ **(b)** $\frac{1}{6}$ **(c)** $\frac{1}{4}$

4 **(b)** **(i)** $\frac{2}{9}$ **(ii)** $\frac{1}{18}$ **(c)** 1

5 **(a)** 6 **(b)** H1 H2 H3 T1 T2 T3 **(d)(i)** $\frac{1}{6}$ **(ii)** $\frac{1}{3}$ **(iii)** $\frac{1}{2}$

6 **(a)** 25 **(c)(i)** $\frac{4}{25}$ **(ii)** $\frac{1}{5}$ **(d)** 6

7 **(a)** 18 **(b)(i)** $\frac{1}{18}$ **(ii)** $\frac{1}{9}$ **(iii)** $\frac{1}{9}$

8 6

Exercise 17.4 (page 241)

1 **(b)(i)** $\frac{15}{32}$ **(ii)** $\frac{25}{64}$

2 **(b)(i)** $\frac{8}{9}$ **(ii)** $\frac{5}{9}$

3 **(a)** 0.28 **(b)** 0.18 **(c)** 0.82

4 **(a)** $\frac{1}{4}$ **(b)** $\frac{1}{169}$ **(c)** $\frac{1}{52}$ **(d)** $\frac{1}{26}$ **(e)** $\frac{7}{16}$

5 **(a)** $\frac{2}{27}$ **(b)** $\frac{4}{27}$ **(c)** $\frac{19}{54}$

6 **(a)** $\frac{4}{81}$ **(b)** $\frac{29}{81}$ **(c)** $\frac{40}{81}$

7 **(a)** 0.09 **(b)** 0.04 **(c)** 0.75

8 **(a)** 0.36 **(b)** 0.01 **(c)** 0.51

Exercise 17.5 (page 244)

1 250

2 **(a)** 50 **(b)** 100 **(c)** 150

3 **(a)** 125 **(b)** 75 **(c)** 175

4 520

5 12

6 **(a)** 50 **(b)** 100 **(c)** 150

7 **(a)** 60 **(b)** 120 **(c)** 220

8 **(a)** 25 **(b)** 75

9 10

10 50

Exercise 18.1 (page 249)

1 **(a)** 5, 6, 9 **(b)** 8, 12, 18 **(c)** 5, 5, 8 **(d)** 4, 4, 6
2 **(b)** 7, 10, 15
3 24
4 regular octahedron (8 faces), regular dodecahedron (12), regular icosahedron (20)

Exercise 18.2 (page 251)

1 **(b)** 294 cm^2
2 248 cm^2
3 **(b)** 108 cm^2
4 **(b)** 224 cm^2
5 351 cm^2
6 **(b)** 391 cm^2

Exercise 18.3 (page 253)

1 729 cm^3
2 **(a)** 160 cm^3 **(b)** 210 m^3
3 120 000 cm^3
4 1.8 m^3
5 3 cm
6 5 cm
7 672 l
8 36
9 1000 000 cm^3
10 96

Exercise 18.4 (page 255)

1 540 cm^3
2 480 cm^3
3 1390 cm^3
4 **(a)** 342 cm^3 **(b)** 1575 cm^3
5 8 cm
6 2025 m^3
7 864 m^3
8 6910 cm^3 (3 s.f.)
9 864 cm^3
10 3.82 cm

Exercise 18.5 (page 257)

1 2240 g or 2.24 kg
2 21 758 g or 21.758 kg
3 1210 g or 1.21 kg
4 1347.5 kg
5 53 g

Exercise 19.1 (page 262)

1 **(a)** 2:3 **(b)** 2:1 **(c)** 4:3 **(d)** 3:5
2 **(a)** 4:3 **(b)** 1:3 **(c)** 2:5 **(d)** 5:2 **(e)** 2:9 **(f)** 2:5 **(g)** 10:3 **(h)** 3:5
3 **(a)** 2:3 **(b)** 4:9
4 **(a)** 1:2 **(b)** 1:4 **(c)** 1:8
5 **(a)** 1:4 **(b)** 1:1.5 **(c)** 1:2.25 **(d)** 1:0.8
6 **(a)** 6:1 **(b)** 3.5:1 **(c)** 2.3:1 **(d)** 0.75:1
7 Sinton 16.1:1, Tanville 16.4:1, Sinton School

Exercise 19.2 (page 264)

1 **(a)** 3.7 m **(b)** 1.61 cm
2 **(a)** 7 m, 6.2 m **(b)** 16 cm, 13.6 cm
3 **(a)** 61.92 m **(b)** 35 cm
4 **(a)** 3.2 km **(b)** 20 cm
5 1:25 000

Exercise 19.3 (page 266)

1 **(a)** 150 g **(b)** 60 g **(c)** 640 g **(d)** 600 g, 200 g
2 **(a)** 15 l **(b)** 20 l
3 36
4 44 cm
5 21 cm, 28 cm
6 £56, £24
7 $120, $300
8 36
9 60°, 50°, 70°
10 100 cm

Exercise 19.4 (page 269)

1 301 miles
2 $17.55

3 **(a)** 240 miles **(b)** 2.5 h
4 **(a)** 180 m^2 **(b)** 20 l
5 **(a)** 7.5 **(b)** 7 h
6 27 g
7 **(a)** 600 g, 3 kg **(b)** 450 g, 2.25 kg **(c)** 200 g, 1 kg
8 **(a)** 48 kg **(b)** 24 kg
9 550
10 **(a)** £12.60 **(b)** 9524 yen

Exercise 19.5 (page 271)

1 12 h
2 10 days
3 6 h
4 6 days
5 1 h 10 min

Exercise 20.1 (page 276)

1 19 cm^2, 17 cm^2, 16 cm^2
2 Yes. $12 + 18 = 30$
3 **(a)** 6.40 cm **(b)** 25.6 cm **(c)** 2.65 cm **(d)** 13 cm **(e)** 6 cm **(f)** 3.96 cm
4 7.07 cm
5 12.2 cm
6 12.5 cm
7 5.20 cm
8 24 cm

Exercise 20.2 (page 279)

1 6.32 m
2 6.71 m
3 51.6 miles
4 10 cm
5 13.9 cm
6 3 cm, 6.40 cm
7 4.24 cm
8 5

Exercise 20.3 (page 282)

1 **(a)** 0.839 **(b)** 2.14 **(c)** 0.0349 **(d)** 0.642 **(e)** 0.0122 **(f)** 3.69 **(g)** 573 **(h)** 5730

2 **(a)** 31.0° **(b)** 71.6° **(c)** 14.0° **(d)** 78.7° **(e)** 4.6° **(f)** 41.8° **(g)** 83.9° **(h)** 87.6°

Exercise 20.4 (page 284)

1 **(a)** 36.9° **(b)** 2.04 cm **(c)** 4.04 cm **(d)** 24.0° **(e)** 5.91 cm **(f)** 29.9° **(g)** 4.47 cm **(h)** 58.2°

2 9.22 m

3 37.9°

4 11.2 miles

5 301 m

Exercise 20.5 (page 289)

1 **(a)** 26.4° **(b)** 6.71 cm **(c)** 3.86 cm **(d)** 31.0° **(e)** 6.66 cm **(f)** 27.4° **(g)** 11.2 cm **(h)** 6.77 cm

2 62.6 m

3 53.1°

4 6.61 m

5 44.4°

6 47.6 cm^2

7 48.2°, 48.2°, 83.6°

8 47.0 cm

Exercise 21.1 (page 293)

1 **(a)** 5^2 **(b)** 7^3 **(c)** 2^4

2 **(a)** 9 **(b)** 125 **(c)** 16 **(d)** 1 000 000

3 **(a)** 3 **(b)** 4 **(c)** 4

4 **(a)** 72 **(b)** 400 **(c)** 1620

Exercise 21.2 (page 294)

1 **(a)** 2^9 **(b)** 5^6 **(c)** 3^8 **(d)** 10^4 **(e)** 5^8 **(f)** 2^{12} **(g)** 3^6 **(h)** 10^5

Exercise 21.3 (page 295)

1 **(a)** $2^2 \times 3$ **(b)** 2×3^2 **(c)** $2^2 \times 3^2$ **(d)** $3^2 \times 5$ **(e)** $2^4 \times 3$ **(f)** $2^2 \times 3 \times 5$ **(g)** $2^2 \times 3^3$ **(h)** $3^2 \times 5^2$ **(i)** $2^3 \times 5 \times 7$ **(j)** $2^3 \times 5^2 \times 11$

Exercise 21.4 (page 297)

1 **(a)** 3 **(b)** 7 **(c)** 6 **(d)** 10

2 **(a)** $2^2 \times 3^2$ **(b)** $2^3 \times 3^3$ **(c)** 2×5^2 **(d)** $2^3 \times 5$

3 **(a)** 24 **(b)** 60 **(c)** 40

Exercise 21.5 (page 298)

1 **(a)** 15 **(b)** 12 **(c)** 24 **(d)** 12

2 **(a)** $2^3 \times 3^5$ **(b)** $2^5 \times 3^2 \times 5 \times 7^3$ **(c)** $2^6 \times 5^2 \times 11^2$
(d) $2^4 \times 3^3 \times 5^3 \times 7$

3 **(a)** 200 **(b)** 120 **(c)** 540

Exercise 21.6 (page 300)

1 **(a)** 9×10^2 **(b)** 3×10^3 **(c)** 7×10^5 **(d)** 4×10^7
(e) 3.2×10^4 **(f)** 1.7×10^5 **(g)** 4.52×10^3 **(h)** 2.75×10^8
(i) 5.2×10^7 **(j)** 4.6×10^6

2 **(a)** 500 **(b)** 80 000 **(c)** 90 000 000 **(d)** 870 000 **(e)** 9100
(f) 2 300 000 **(g)** 19 700 **(h)** 42 100 000 000

3 3×10^8

4 7.5×10^9

5 93 700 000

Exercise 21.7 (page 302)

1 **(a)** 7×10^{-2} **(b)** 4×10^{-4} **(c)** 3×10^{-1} **(d)** 9×10^{-8}
(e) 8.4×10^{-3} **(f)** 2.9×10^{-5} **(g)** 7.91×10^{-3} **(h)** 7.39×10^{-1}

2 **(a)** 0.004 **(b)** 0.000 08 **(c)** 0.9 **(d)** 0.000 000 003 **(e)** 0.013
(f) 0.000 97 **(g)** 0.000 004 73 **(h)** 0.000 000 508

3 5×10^{-4}

4 1.69×10^{-5}

5 0.000 002 2

Exercise 21.8 (page 304)

1 **(a)** 6×10^{11} **(b)** 2.4×10^8 **(c)** 3×10^3 **(d)** 8×10^3 **(e)** 4.3×10^4
(f) 4.93×10^6
(g) 3.147×10^5 **(h)** 5.92×10^3

2 **(a)** 2.88×10^{10} **(b)** 2.55×10^2 **(c)** 5.69×10^4 **(d)** 9.19×10^7

3 2.88×10^{12}

4 20 (1 sf)

5 **(a)** 2.5×10^{17} **(b)** 1.5×10^6

Exercise 22.1 (page 311)

1 **(a)** 7, 8, 9 **(b)** 5, 7, 8 **(c)** 9, 8, 6 **(d)** 7, 6, 5 **(e)** 7, 9.5, 10.5 **(f)** 5, 0, 1
2 3, 2.5, 2.4
3 84, 76, 75
4 1, 1.5, 2.5
5 –1°C, –2°C, –3°C
6 2 h 14 min, 2 h 9 min, 2 h 6 min
7 6 years 1 month
8 **(a)** 42 mph **(b)** $2\frac{1}{2}$ h
9 83
10 27
11 1, 1, 5, 6, 7
12 45 kg
13 158 cm
14 18 mph

Exercise 22.2 (page 313)

1 **(a)** 7 **(b)** 25 **(c)** 7
2 21
3 34, 76
4 9 degrees
5 3, 3, 8, 9, 12

Exercise 22.3 (page 314)

1 4, 3, 3.2, 6
2 0, 1, 1.6, 5
3 2, 3, 3, 6
4 15, 15, 15.5, 4

Exercise 22.4 (page 319)

Answers obtained from your cumulative frequency graphs may differ slightly from those given here.

1 **(a)** 11–15 **(b)** 11.5 **(c)** 12, 8, 14, 6
2 **(a)** $90 \leq w < 100$ **(b)** 89.1 kg **(c)** 91 kg, 81 kg, 98 kg, 17 kg **(d)** 85

Exercise 23.1 (page 325)

1 **(a)** $y = x + 3$ **(b)** $y = x + 1$ **(c)** $y = x - 1$ **(d)** $x + y = 4$ **(e)** $x + y = 1$ **(f)** $x + y = -2$ **(g)** $y = 4x$ **(h)** $y = \frac{1}{2}x$ **(i)** $y = -2x$

Exercise 23.2 (page 327)

1 –8, –5, –2, 1, 4, 7, 10

2 –2, –1, 0, 1, 2, 3, 4

3 1, –1, –3, –5, –7, –9, –11

4 **(a)** (0, 3) **(b)** (0, 4) **(c)** (0, –2) **(d)** (0, 3) **(e)** (0, 2) **(f)** (0, 6)

Exercise 23.3 (page 331)

1 **(a)** 7 **(b)** (0, –4)

2 $y = -5x + 4$

3 $y = 2x + 5$, (0, 5)

4 **(a)** 4 **(b)** $-\frac{1}{2}$ **(c)** 0

5 **(a)** $y = 4x + 3$ **(b)** $y = \frac{1}{2}x + 3$ **(c)** $y = -5x + 12$

6 **(a)** –2, (0, 3) **(b)** $\frac{2}{5}$, (0, –3) **(c)** $-\frac{3}{4}$, (0,6)

Exercise 23.4 (page 335)

1 **(a)** 19, 12, 7, 4, 3, 4, 3, 7, 12, 19 **(b)** 80, 45, 20, 5, 0, 5, 20, 45, 80 **(c)** 8, 4.5, 2, 0.5, 0, 0.5, 2, 4.5, 8 **(d)** –16, –9, –4, –1, 0, –1, –4, –9, –16 **(e)** –32, –18, –8, –2, 0, –2, –8, –18, –32 **(f)** 24, 15, 8, 3, 0, –1, 0, 3, 8 **(g)** 25, 16, 9, 4, 1, 0, 1, 4, 9 **(h)** 31, 20, 11, 4, –1, –4, –5, –4, –1 **(i)** 15, 4, –3, –6, –5, 0, 9, 22, 39

2 **(a)** –19, 0, 7, 8, 9, 16, 35 **(b)** –54, –16, –2, 0, 2, 16, 54 **(c)** –13.5, –4, –0.5, 0, 0.5, 4, 13.5 **(d)** 27, 8, 1, 0, –1, –8, –27 **(e)** 54, 16, 2, 0, –2, –16, –54 **(f)** –64, –27, –8, –1, 0, 1, 8 **(g)** –36, –12, –2, 0, 0, 4, 18 **(h)** –30, –8, 0, 0, –2, 0, 12 **(i)** –38, –16, –6, –2, 2, 12, 34

3 **(a)** –0.7, –1, –2, –4, –10, 10, 4, 2, 1, 0.7 **(b)** 0.3, 0.5, 1, 2, 5, –5, –2, –1, –0.5, –0.3 **(c)** 0.7, 1, 2, 4, 10, –10, –4, –2, –1, –0.7 **(d)** 4.7, 4.5, 4, 3, 0, 10, 7, 6, 5.5, 5.3

Exercise 24.1 (page 339)

1 **(a)** e.g. $x = 2, y = 3$ **(b)** $x = 3, y = 0$ **(c)** $x = 3, y = 0$

2 **(a)**, **(c)**

3 **(b)**
4 **(a)** $x = 3, y = 1$ **(b)** $x = 2, y = 3$ **(c)** $x = 4, y = 2$ **(d)** $x = 4, y = -3$
(e) $x = -2, y = 4$ **(f)** $x = -1, y = -2$
5 The lines are parallel so they do not meet.

Exercise 24.2 (page 343)

1 **(a)** $x = 5, y = 2$ **(b)** $x = 2, y = 1$ **(c)** $x = 3, y = -1$
2 **(a)** $x = 1, y = 1$ **(b)** $x = 2, y = -1$ **(c)** $x = 3, y = 2$
(d) $x = 7, y = -3$ **(e)** $x = 2\frac{1}{2}, y = 2$ **(f)** $x = \frac{1}{2}, y = -4$
3 49, 25
4 £3
5 39, 83
6 $m = \frac{1}{2}, c = -3$
7 $a = 1, b = -2$

Exercise 24.3 (page 345)

1 **(a)** $x^2 + 5x + 6$ **(b)** $x^2 - x - 20$ **(c)** $x^2 - 4x - 5$
(d) $x^2 - 9x + 18$ **(e)** $x^2 + 10x + 25$ **(f)** $x^2 - 4x + 4$ **(g)** $x^2 - 49$
(h) $x^2 - 16$ **(i)** $18 + 3x - x^2$ **(j)** $5 + 4x - x^2$ **(k)** $6x^2 + 23x + 20$
(l) $20x^2 - 11x - 3$ **(m)** $12x^2 - 44x + 7$ **(n)** $9x^2 + 12x + 4$
(o) $4x^2 - 20x + 25$ **(p)** $81x^2 - 16$ **(q)** $9x^2 - 49$ **(r)** $12 + 28x - 5x^2$
(s) $6 - x - 12x^2$ **(t)** $25 - 30x + 9x^2$ **(u)** $a^2x^2 - b^2$

Exercise 24.4 (page 347)

1 **(a)** $x + 4$ **(b)** $x + 2$ **(c)** $x + 4$ **(d)** $x - 4$ **(e)** $x + 8$ **(f)** $x - 8$
(g) $x - 4$ **(h)** $x - 2$ **(i)** $x + 10$ **(j)** $x + 2$
2 **(a)** $(x - 5)(x + 3)$ **(b)** $(x + 2)(x + 7)$ **(c)** $(x + 1)(x - 25)$
(d) $(x + 2)(x - 12)$ **(e)** $(x + 7)(x - 7)$ **(f)** $x(x + 8)$ **(g)** $(x - 4)(x + 14)$
(h) $(x - 1)(x + 1)$ **(i)** $(x - 10)(x + 2)$ **(j)** $x(x - 9)$

Exercise 24.5 (page 349)

1 **(a)** 2.65, –2.65 **(b)** 0, 2 **(c)** 1, 3 **(d)** 1.3, –2.3 **(e)** 2.8, –1.3
(f) 3.6, –0.6
2 **(a)** $x = 2$ graph touches x-axis but does not cross it
(b) no solution as graph does not cross the x-axis

Exercise 24.6 (page 351)

1 **(a)** 1, 6 **(b)** –4, 3 **(c)** –7, 7 **(d)** –5, –4 **(e)** 0, 6 **(f)** 0, –8

2 **(a)** 1, 2 **(b)** –2, 1 **(c)** –4, –3 **(d)** –5, 5 **(e)** 5 repeated **(f)** 0, –10 **(g)** –4, 5 **(h)** –8, 8 **(i)** 0, 7 **(j)** –2, 5

3 7 cm

4 9 or –3

5 8 or –9

6 7, 9 (or –7, –9)

7 7, 8

Exercise 24.7 (page 355)

1 **(a)** 0, 2, 5 **(b)** –2, 1, 3 **(c)** –2.2, –0.8, 2.9 **(d)** –2, 1, 2 **(e)** 0, 2 repeated **(f)** 2.4

2 5.8

3 4.3

4 2.9

5 1.7

6 2.3

Taking it further

Complete Mathematics is a general course that covers and integrates different branches of mathematics. To go on from here, you could either pursue a more advanced general course in mathematics, or decide that you would like to know more about the individual subjects, algebra, trigonometry, calculus and statistics which make up *Complete Mathematics*. Whichever line you decide to take, you should think carefully about buying a graphical display calculator if you do not have one already. These calculators are not cheap, but they are important tools for learning trigonometry, calculus and statistics. Recommended graphical display calculators at the time of writing (2010) are Texas Instruments TI84 and Casio 9850+.

Websites and organizations

A+B Books, who write and publish mathematics books for use with a graphics calculator.
www.aplusb.co.uk

Association of Teachers of Mathematics (UK)
www.atm.org.uk

Autograph: software for learning advanced mathematics
www.autograph-math.com

Bournemouth University applets (computer animations) demonstrating mathematical principles
mathinsite.bmth.ac.uk

Coventry University Mathematics Support Centre for Mathematics Education
www.mis.cov.ac.uk/maths_centre

Free-Standing Mathematics Qualifications
www.fsmq.org

International Baccalaureate: an international qualification
www.ibo.org

iCT Training Centre, with many useful resources and links to other sites worldwide
www.tsm-resources.com

Mathematical Association (UK)
www.m-a.org.uk

Mathpuzzle contains a range of puzzles and other maths resources
www.mathpuzzle.com

MathsNet contains a wide range of puzzles, books and software
http://www.mathsnet.net

National Council of Teachers of Mathematics (USA)
www.nctm.org

Open University Mathematics courses
www.mathematics.open.ac.uk

Oundle School site, with many useful resources and links *to* other sites worldwide
http://www.argonet.co.uk/oundlesch/mlink.html

National Council of Teachers of Mathematics (USA)
www.nctm.org

UK National Statistics on line
http://www.statistics.gov.uk/

University of Plymouth Mathematics support materials
www.tech.plym.ac.uk/maths/resources/PDFLaTeX/mathaid.html

Reading list

Abbott, P. (2010) *Understand Algebra,* Hodder Education, London.

Abbott, P. (2010) *Understand Trigonometry,* Hodder Education, London.

Abbott, P. (2010) *Understand Calculus,* Hodder Education, London.

Barrow, John D. (1993) *Pi in the Sky: Counting, Thinking and Being,* Penguin, London. An exploration of where maths comes from and how it is performed.

Eastaway, Rob & Wyndham, Jeremy (1998) *Why Do Buses Come in Threes?,* Robson Books, London. Practical uses for various mathematical topics, including probability, Venn diagrams and prime numbers.

Eastaway, Rob & Wyndham, Jeremy (2002) *How Long is a Piece of String?,* Robson Books, London. Examples of mathematics in everyday life.

Eastaway, Rob & Wells, David (2005) *Mindbenders and Brainteasers,* Robson Books, London. A collection of 100 puzzles and conundrums, old and new.

Flannery, Sarah (2000) *In Code: A Mathematical Journey,* Profile Books, London. A collection of problems with solutions and explanations, based on the author's experiences of growing up in a mathematical home.

Graham, Alan (2010) *Understand Statistics,* Hodder Education, London. A straightforward and accessible account of the big ideas of statistics with a minimum of hard mathematics.

Haighton, June et al. (2004) *Maths the Basic Skills,* Nelson Thornes, Cheltenham. A traditional textbook on basic mathematics based on the Adult Numeracy Core Curriculum.

Huntley, H.E. (1970) *The Divine Proportion, Study in Mathematical Beauty,* Dover, New York. Applications in art and nature of the 'Golden Ratio'.

Ifrah, Georges (1998) *The Universal History of Numbers,* The Harvill Press, London. A detailed book (translated from French) about the history of numbers and counting from prehistory to the age of the computer.

Neill, H. and Quadling, D. *Mathematics for the IB Diploma,* Standard Level, Cambridge University Press.

Paulos, John Allen (1990) *Innumeracy – Mathematical Illiteracy and its Consequences,* Penguin, London. Real-world examples of innumeracy, including stock scams, risk perception and election statistics.

Pólya, G. (1990) *How to Solve It,* Penguin, London. A classic text on mathematical problem solving that is well-known around the world.

Potter, Lawrence (2006) *Mathematics Minus Fear,* Marion Boyars Publishers Ltd, London. A romp through school mathematics that takes in puzzles and gambling.

Seiter, Charles (1995) *Everyday Math for Dummies,* For Dummies, New York. A friendly guide to basic maths, written in an entertaining style.

Singh, Simon (2000) *The Code Book,* Fourth Estate, London. A history of codes and ciphers and their modern applications in electronic security.

Singh, Simon (1998) *Fermat's Last Theorem,* Fourth Estate, London. An account of Andrew Wiles' proof of Fermat's Last Theorem, but also outlining some problems that have interested mathematicians over many centuries.

Stewart, Ian (1996) *From Here to Infinity,* Oxford University Press, Oxford. An introduction to how mathematical ideas are developing today.

Stewart, Ian (1997) *Does God Play Dice?,* Penguin, London. An introduction to the theory and practice of chaos and fractals.

Stewart, Ian (2006) *Letters to a Young Mathematician,* Basic Books, New York. What the author wishes he knew about mathematics when he was a student.

Index

acute-angled triangles, *45*
acute angles, *17, 19*
addition, *3, 4*
  decimals, *68–9*
  directed numbers, *106–8*
  fractions, *34–6*
algebra
  defining, *148*
  methods of solving equations, *339–43*
  origins of the word, *182*
  ***see also* equations**
algebraic expressions, *148–68*
  brackets, *155–8*
  evaluating, *153–5*
  factorizing, *158–9, 345–7*
  and formulae, *209*
  indices, *159–65, 292–8*
  simplifying, *151–3, 163–6*
  writing, *148–51*
alternate angles, *23–4*
altitude, *141*
angles, *15–28*
  acute angles, *17, 19*
  at a point, *21*
  bearings, *25–6, 28*
  facts, *20–1*
  interior and exterior, *53–5*
  measuring and drawing, *18–20*
  naming, *17*
  obtuse angles, *17*
  parallel lines, *22–4*
  reflex angles, *17, 19*
  right angles, *16, 21*
  on a straight line, *21*
  vertex, *16*
  vertically opposite, *21, 24*
angle sum of a quadrilateral, *50*
apex, *247*
approximation, *169–81*
  accuracy of measurements, *179–80*
  estimates, *176–7*
  rounding
    with decimals, *171–4*
    in practical problems, *178*
    whole numbers, *169–71*
  significant figures, *174–6*
arc, *217*
arctan, *282*
area, *136–47*
  parallelograms, *141, 144*
  rectangles, *138–40*
  trapeziums, *144–6*
  triangles, *141–3*
area of a circle, *221–3*
arithmetic
  the four operations, *3–4*
  order of operations, *4–6*
  problems which use, *7–6*
averages, *306–311*

bar charts, *83, 86, 87, 99*
bearings, *25–6, 28*
BiDMAS convention, *5–6*
  in algebraic expressions, *153, 156*
bi-modal data, *307*
bisecting a circle, *216–17*
brackets
  in algebraic expressions, *155–8*
  equations with, *188–9*

calculators, and negative numbers, *111*
cancelling fractions, *32*
categorical data, *96*
centilitres, *126*
centre of a circle, *216*
centre of rotational symmetry, *46*
circles, *216–28*
  arc, *217*
  area, *221–3*
  centre, *216*
  chords, *217, 224*
  circumference, *218–20*
  diameter, *216–17, 218–20, 222*
  properties of, *224–6*
  radius, *216*
  sector, *217*

circumference of a circle, *218–20*
class intervals, *97*
complementary angles, *21*
completely factorized algebraic expressions, *158–9*
compound interest, *205–6*
cones, *248*
congruent shapes, *55–6, 144*
constant term, *346*
continuous data, *96–7*
conversion graphs, *116–17*
cooling curves, *91*
coordinates, *113–15, 322–4*
correlation in scatter graphs, *93–4*
cosines, *285–8*
counting numbers, *8*
cross-sections, of three-dimensional shapes, *247*
cube numbers, *9*
cube root, *352*
cubes, *246*
cubic equations, *352–5*
cubic graphs, *333*
cuboids, *246, 251–3*
cumulative frequencies, *314, 316–17*
curved graphs, *331–4*
cylinders, *247, 250*

data collection, *81–2*
decagons, *52*
decimal points, *63, 64*
decimals, *63–79*
  addition and subtraction, *68–9*
  converting fractions to, *67, 74–7*
  converting to fractions, *66*
  division, *72–4*
  multiplication, *69–72*
  place value, *64–5*
  rounding with, *171–4*
  writing percentages as, *195–9*
decimal system of numbers, *1–3*
denominators, *29–30*
Descartes, René, *114*
diagrams
  probability/sample space, *236*
  tree diagrams, *239–41*
diameter, *216–17, 218–20, 222*
difference of a number, *3*
different-sign numbers, *109, 110*
digits, *1*
Diophantus, *182*
directed numbers, *104–9*
  addition and subtraction, *106–8*
  ordering, *105*
direct proportion, *267–9*
discrete data, *96–7*
distance–time graphs, *117–18, 119*
divided numbers, multiplication and division, *108–111*
division, *3, 4*
  of decimals, *72–4*
  divided numbers, *108–111*
  of fractions, *38–40*

elimination (of unknowns), *339–40*
equations, *182–93, 337–57*
  algebraic methods, *339–43*
  with brackets, *188–9*
  cubic, *352–5*
  finding missing numbers, *183*
  quadratic, *348–9*
    solution by factorizing, *349–52*
  quadratic expressions, *343–7*
  simultaneous, *337–9, 340, 341–3*
  solving inequalities, *191–3*
  solving linear, *183–8*
  solving problems using, *189–91*
equation of straight lines, *121, 322–5*
equilateral triangles, *46, 57, 59*
equivalent fractions, *30, 31*
estimates, *176–7*
Euclid, *Elements*, *15, 44*
Euler, Leonard, *248*
even numbers, *8*
expected frequency, *243–4*
exterior angles, *53–5*

factorization
  algebraic expressions, *158–9*
  cubic equations, *352*
  quadratic expressions, *345–7*
factors of a number, *11, 12*
  highest common factor (HCF), *295–6*
  prime factors, *294–5*
fair dice, and probability, *231, 232, 233–4, 235*
F-angles, *28*
first quartiles, *317*
formulae, *209–15*
  evaluating terms other than the subject, *211–13*

    - **subject of,** *209*
        - **changing,** *213–14*
    - **writing,** *209–11*
- **fractions,** *29–43*
    - **adding and subtracting,** *34–6*
    - **cancelling,** *32*
    - **comparing with ratios,** *261*
    - **converting decimals to,** *66*
    - **converting to decimals,** *67, 74–7*
    - **division of,** *38–40*
    - **equivalent,** *30, 31*
    - **finding the larger of two,** *31*
    - **improper,** *33–4, 39*
    - **lowest common multiple,** *31*
    - **lowest terms,** *32*
    - **mixed numbers,** *33–4, 39, 76*
    - **multiplication of,** *36–8*
    - **numbers as a fraction of another number,** *40–1*
    - **of quantities,** *37–8*
    - **simplest form,** *32*
    - **simplifying,** *31–2*
    - **writing percentages as,** *195–9*
    - ***see also* decimals**
- **frequency/frequency tables,** *82, 84, 313–20*
    - **grouped,** *315–19*
- **frequency polygons,** *97*
- **fully factorized algebraic expressions,** *158–9*

- **geometry, origins of,** *15*
- **gradient of a straight line,** *328–31*
- **grams,** *126*
- **graphs,** *113–24, 322–36*
    - **coordinates,** *113–15, 322–4*
    - **curved,** *331–4*
    - **drawing straight-line,** *325–7*
    - **equations of straight lines,** *121, 322–5*
    - **gradient of a straight line,** *328–31*
    - **line graphs,** *89–92*
    - **lines parallel to the axes,** *121–3*
    - **scatter graphs,** *93–6*
    - **straight-line,** *116–20*
- **grouped frequency tables,** *97, 315–19*
- **grouping data,** *97–9*

- **HCF (highest common factor),** *295–7*
- **hexagons,** *52, 57*
- **highest common factor (HCF),** *295–7*
- **high negative correlation,** *94*
- **hundreds,** *1*
- **hundredths,** *64*
- **hypotenuse,** *273, 285–6*

- **imperial units of measurment,** *127–9*
    - **converting between metric and,** *129–30*
- **improper fractions,** *33–4, 39*
- **index,** *159*
- **indices,** *159–66, 292–8*
    - **highest common factor (HCF),** *295–7*
    - **laws of,** *160–3, 293–4*
    - **lowest common multiple (LCM),** *297–8*
    - **prime factors,** *294–5*
    - **simplying expressions with,** *163–6*
- **inequalities**
    - **of lines,** *121–2, 123*
    - **solving,** *191–3*
- **intecept,** *326*
- **integers,** *8*
- **interior angles,** *53–5*
- **interquartile range,** *317–18*
- **inverse proportion,** *269–71*
- **inverse tangents,** *282*
- **isoceles triangles,** *46–7, 59*
- **isosceles trapeziums,** *51, 61, 146*

- **kilograms,** *126*
- **kites,** *51, 61*

- **large numbers,** *298–300*
- **LCM (lowest common multiple),** *31, 297–8*
- **like terms,** *151, 152*
- **linear equations, solving,** *183–8*
- **line of best fit,** *93*
- **line graphs,** *89–92*
- **lines of symmetry,** *46, 51, 217*
- **logarithms,** *63*
- **lower quartiles,** *317*
- **lowest common multiple (LCM),** *31, 297–8*
- **lowest powers of prime factors,** *296*
- **lowest terms of fractions,** *32*

- **mean,** *309–10, 311*
- **measurement,** *125–33*
    - **accuracy of measurements,** *179–80*
    - **choosing suitable units,** *130–1*
    - **converting between metric and imperial units,** *129–30*
    - **imperial units,** *127–9*
    - **metric system,** *125–7*
- **median,** *307–8, 311, 313–14*

metric system, *125–7*
  converting between metric and imperial units, *129–30*
  using suitable units, *130–1*
mililitres, *126*
milligrams, *126*
mirror lines, *46*
mixed numbers, *33–4, 39, 76*
modal class interval, *315*
mode, *306–7, 311, 313*
multiples, *10, 11*
multiplication, *3, 4*
  decimals, *69–72*
  divided numbers, *108–111*
  of fractions, *36–8*
multiplier method, *268*
multipliers, *205–7*
mutually exclusive events, *232, 233*

natural numbers, *8*
negative correlation, *93, 94*
negative numbers, *104, 109, 110*
  and calculators, *111*
nets and surface area, *249–51*
number, *1–14*
  arithmetic
    the four operations, *3–4*
    order of operations, *4–6*
    problems which use, *7–6*
  place value, *1–3*
  special numbers, *8–9*
numbers
  factors, *11, 12*
  as a fraction of another number, *40–1*
  multiples, *10, 11*
  standard form, *298–304*
  ***see also* whole numbers**
numerators, *29–30*

obtuse-angled triangles, *45, 59*
obtuse angles, *17*
octagons, *52, 57*
odd numbers, *8*
order of rotational symmetry, *46, 51, 217*
origins, *113*

parabolas, *331–2*
parallel lines, *22–4*
parallelograms, *51, 60*
  area of, *141, 144*
pentagons, *52, 53*
percentages, *195–208*
  increases, *203–5*
  of quantities, *199–200*
    increasing and decreasing, *201*
    one quantity as a percentage of another quantity, *202–3*
  using multipliers, *205–7*
  writing as fractions or decimals, *195–9*
perfect correlation, *94*
perfect squares, *9*
perimeter, *134–6*
perpendicular lines, *16*
pictograms, *82–3, 85*
pie charts, *87–9*
place value, *1–3*
  decimals, *64–6*
plane geometry, *44*
plane shapes, *44*
polygons, *52–5*
  finding the area of a circle, *221–2*
polyhedrons, *248*
positive correlation, *93, 94*
positive numbers, *104, 109, 110*
positive whole numbers, *8*
power, *159, 160, 292, 293–4, 296*
prime factors, *12, 294–5*
prime numbers, *11, 12, 294*
principal, and compound interest, *205–6*
prisms, *247*
  volume, *253–6*
  weight, *256–7*
probability, *229–45*
  expected frequency, *243–4*
  and relative frequency, *230–1*
  of a single event, *231–4*
  tree diagrams, *239–41*
  two events, *234–8*
probability/sample space diagrams, *236*
probability scale, *230*
problem solving
  Pythagoras' theorem problems, *279–80*
  rounding in practical problems, *178*
  using equations, *189–91*
  using ratio, *265–7*
product of a number, *3*
proportion, *267–71*
  direct, *267–9*
  inverse, *269–71*

protractors, *19–20*
pyramids, *247–8, 249*
Pythagoras' theorem, *273–80*
  problems, *279*
  proof of, *278*
  using, *274–7*

quadratic equations, *348–9*
  solution by factorizing, *349–52*
quadratic expressions, *343–7*
  factorizing, *345–7*
quadratic graphs, *332*
quadrilaterals, *50–2, 60*
quantities
  fractions of, *37–8*
  percentages of, *199–200*
    increases, *203–5*
    increasing and decreasing, *201*
    one quantity as a percentage of another quantity, *202–3*
  using ratio, *265–7*
quotient of a number, *3, 4*

radius, *216*
range (in statistics), *312–13*
ratio, *260–7*
  comparing with fractions, *261*
  scales, *262–4*
  simplying, *260–2*
  using, *265–7*
reciprocal of a fraction, *39*
reciprocal graphs, *334*
rectangles, *51, 60*
  area of, *138–40*
rectangular numbers, *9*
recurring decimals, *75–6*
reflex angles, *17, 19*
region of lines, *121*
regular solids, *248*
relative frequency, *230–1*
rhombus, *51, 61*
right-angled triangles, *45*
  Pythagoras' theorem, *273–80*
right angles, *16, 21*
right pyramids, *247*
Roman numerals, *1*
root of equations, *182*
rotational symmetry, *46, 51, 217*
rounding
  with decimals, *171–4*
  in practical problems, *178*
  whole numbers, *169–71*
rounding down, *170*
rounding up, *170*
running totals, *314*

same-sign numbers, *109, 110*
samples, *81*
satisfied equations, *182*
scalene triangles, *47, 59*
scales, *262–4*
scatter graphs, *93–6*
sector of a circle, *217*
segments, *217*
semicircles, *217*
significant figures, *174–6*
simplest form of fractions, *32*
simultaneous equations, *337–9, 340, 341–3*
sines, *285–8*
small numbers, *301–2*
solids ***see*** **three-dimensional shapes**
solutions of equations, *182*
  cubic equations, *353–5*
  quadratic equations and fractorization, *349–52*
special numbers, *8–9*
spheres, *248*
square numbers, *9*
squares (quadrilaterals), *50, 60*
  area of, *139*
  perimeter of, *134, 135*
standard form, *298–304*
  calculations, *302–4*
  large numbers, *298–300*
  small numbers, *301–2*
standard index form, *299*
statistics, *80–103, 306–21*
  averages, *306–311*
  bar charts, *83, 86, 87, 99*
  collection of data, *81–2*
  continuous data, *96–7*
  discrete data, *96–7*
  frequency tables, *313–20*
  grouping data, *97–9*
  line graphs, *89–92*
  pictograms, *82–3, 85*
  pie charts, *87–9*
  range, *312–13*
  scatter graphs, *93–6*
Stevin, Simon, *63*

straight-line graphs, *116–20*
  gradient of a straight line, *328–31*
subject of formulae, *209*
  changing, *213–14*
subtraction, *3, 4*
  decimals, *68–9*
  directed numbers, *106–8*
  fractions, *34–6*
sum of a number, *3*
supplementary angles, *21*
surveys, *81–2*
symmetries
  lines of symmetry, *46, 51, 217*
  of quadrilaterals, *51*
  of regular polygons, *55*
  of triangles, *46*

tally charts, *81–2, 84*
tangents, *217, 225*
  ratio, *281*
  using, *283–5*
  values of, *282*
tan (tangent), *281, 282*
tens, *1*
tenths, *64*
terminating decimals, *75*
terms, *151, 152*
tessellations, *57–8*
tetrahedron, *247*
Thales of Miletus, *44*
third quartiles, *317*
thousands, *1*
thousandths, *64*
three-dimensional shapes, *246–59*
  cubes, *246*
  cuboids, *246, 251–3*
  nets and surface area, *249–51*
  prisms, *247, 253–6*
  pyramids, *247–8, 249*
  regular solids, *248*
  volume of, *251–6*
three-figure bearings, *25, 28*
time-series graphs, *89–90*
tonnes, *126*
top-heavy (improper) fractions, *33–4*
transversals, *23*
trapeziums, *51, 61*
  area, *144–6*
tree diagrams, *239–41*
triangles, *44–50, 52, 59*
  area of, *141–3*
  congruent, *56*
  constructing, *48–50*
  perimeter of, *134, 135*
  Pythagoras' theorem, *273–80*
  symmetry, *46*
  types of, *45–8*
  ***see also* trigonometry**
triangular numbers, *9*
trigonometry, *280–90*
  sines and cosines, *285–8*
  tangents, *281–5*
two-dimensional shapes, *44–62*
  congruent, *55–6*
  polygons, *52–5*
  quadrilaterals, *50–2*
  tessellations, *57–8*
  triangles, *44–50, 52, 59*

unbiased, *231*
unitary method, *268*
units, *1*
unlike terms, *151, 152*
upper quartiles, *317*

vertex, *16, 246*
vertical line graphs, *90*
vertically opposite angles, *21, 24*
vertices, *246*

whole numbers, *8*
  and prime factors, *295*
  rounding, *169–71*
  rounding with decimals, *171–2*

***x*-axis,** *113, 114*
***x*-coordinates,** *113, 114, 115, 322–3, 324*
***y*-coordinates,** *113, 114, 115, 322–3, 324*
***y*-intecept,** *326*

Z-angles, *28*
zero, *2*

CRC Concise Encyclopedia
of
MATHEMATICS

CRC Concise Encyclopedia of MATHEMATICS

Eric W. Weisstein

CRC Press
Boca Raton London New York Washington, D.C.

Library of Congress Cataloging-in-Publication Data

Weisstein, Eric W.
The CRC concise encyclopedia of mathematics / Eric W. Weisstein.
p. cm.
Includes bibliographical references and index.
ISBN 0-8493-9640-9 (alk. paper)
1. Mathematics– –Encyclopedias. I. Title.
QA5.W45 1998
510′.3—DC21 98-22385
CIP

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

Neither this book nor any part may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming, and recording, or by any information storage or retrieval system, without prior permission in writing from the publisher.

The consent of CRC Press LLC does not extend to copying for general distribution, for promotion, for creating new works, or for resale. Specific permission must be obtained in writing from CRC Press LLC for such copying.

Direct all inquiries to CRC Press LLC, 2000 Corporate Blvd., N.W., Boca Raton, Florida 33431.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are only used for identification and explanation, without intent to infringe.

© 1999 by CRC Press LLC

No claim to original U.S. Government works

International Standard Book Number 0-8493-9640-9
Library of Congress Card Number 98-22385
Printed in the United States of America 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

Introduction

The CRC Concise Encyclopedia of Mathematics is a compendium of mathematical definitions, formulas, figures, tabulations, and references. It is written in an informal style intended to make it accessible to a broad spectrum of readers with a wide range of mathematical backgrounds and interests. Although mathematics is a fascinating subject, it all too frequently is clothed in specialized jargon and dry formal exposition that make many interesting and useful mathematical results inaccessible to laypeople. This problem is often further compounded by the difficulty in locating concrete and easily understood examples. To give perspective to a subject, I find it helpful to learn why it is useful, how it is connected to other areas of mathematics and science, and how it is actually implemented. While a picture may be worth a thousand words, explicit examples are worth at least a few hundred! This work attempts to provide enough details to give the reader a flavor for a subject without getting lost in minutiae. While absolute rigor may suffer somewhat, I hope the improvement in usefulness and readability will more than make up for the deficiencies of this approach.

The format of this work is somewhere between a handbook, a dictionary, and an encyclopedia. It differs from existing dictionaries of mathematics in a number of important ways. First, the entire text and all the equations and figures are available in searchable electronic form on CD-ROM. Second, the entries are extensively cross-linked and cross-referenced, not only to related entries but also to many external sites on the Internet. This makes locating information very convenient. It also provides a highly efficient way to "navigate" from one related concept to another, a feature that is especially powerful in the electronic version. Standard mathematical references, combined with a few popular ones, are also given at the end of most entries to facilitate additional reading and exploration. In the interests of offering abundant examples, this work also contains a large number of explicit formulas and derivations, providing a ready place to locate a particular formula, as well as including the framework for understanding where it comes from.

The selection of topics in this work is more extensive than in most mathematical dictionaries (e.g., Borowski and Borwein's *HarperCollins Dictionary of Mathematics* and Jeans and Jeans' *Mathematics Dictionary*). At the same time, the descriptions are more accessible than in "technical" mathematical encyclopedias (e.g., Hazewinkel's *Encyclopaedia of Mathematics* and Iyanaga's *Encyclopedic Dictionary of Mathematics*). While the latter remain models of accuracy and rigor, they are not terribly useful to the undergraduate, research scientist, or recreational mathematician. In this work, the most useful, interesting, and entertaining (at least to my mind) aspects of topics are discussed in addition to their technical definitions. For example, in my entry for pi (π), the definition in terms of the diameter and circumference of a circle is supplemented by a great many formulas and series for pi, including some of the amazing discoveries of Ramanujan. These formulas are comprehensible to readers with only minimal mathematical background, and are interesting to both those with and without formal mathematics training. However, they have not previously been collected in a single convenient location. For this reason, I hope that, in addition to serving as a reference source, this work has some of the same flavor and appeal of Martin Gardner's delightful *Scientific American* columns.

Everything in this work has been compiled by me alone. I am an astronomer by training, but have picked up a fair bit of mathematics along the way. It never ceases to amaze me how mathematical connections weave their way through the physical sciences. It frequently transpires that some piece of recently acquired knowledge turns out to be just what I need to solve some apparently unrelated problem. I have therefore developed the habit of picking up and storing away odd bits of information for future use. This work has provided a mechanism for organizing what has turned out to be a fairly large collection of mathematics. I have also found it very difficult to find clear yet accessible explanations of technical mathematics unless I already have some familiarity with the subject. I hope this encyclopedia will provide jumping-off points for people who are interested in the subjects listed here but who, like me, are not necessarily experts.

The encyclopedia has been compiled over the last 11 years or so, beginning in my college years and continuing during graduate school. The initial document was written in *Microsoft Word*® on a Mac Plus® computer, and had reached about 200 pages by the time I started graduate school in 1990. When Andrew Treverrow made his OzTEX program available for the Mac, I began the task of converting all my documents to TEX, resulting in a vast improvement in readability. While undertaking the *Word* to TEX conversion, I also began cross-referencing entries, anticipating that eventually I would be able to convert the entire document to hypertext. This hope was realized beginning in 1995, when the Internet explosion was in full swing and I learned of Nikos Drakos's excellent TEX to HTML converter, LATEX2HTML. After some additional effort, I was able to post an HTML version of my encyclopedia to the World Wide Web, currently located at `www.astro.virginia.edu/~eww6n/math/`.

The selection of topics included in this compendium is not based on any fixed set of criteria, but rather reflects my own random walk through mathematics. In truth, there is no good way of selecting topics in such a work. The mathematician James Sylvester may have summed up the situation most aptly. According to Sylvester (as quoted in the introduction to Ian Stewart's book *From Here to Infinity*), "Mathematics is not a book confined within a cover and bound between brazen clasps, whose contents it needs only patience to ransack; it is not a mine, whose treasures may take long to reduce into possession, but which fill only a limited number of veins and lodes; it is not a soil, whose fertility can be exhausted by the yield of successive harvests; it is not a continent or an ocean, whose area can be mapped out and its contour defined; it is as limitless as that space which it finds too narrow for its aspiration; its possibilities are as infinite as the worlds which are forever crowding in and multiplying upon the astronomer's gaze; it is as incapable of being restricted within assigned boundaries or being reduced to definitions of permanent validity, as the consciousness of life."

Several of Sylvester's points apply particularly to this undertaking. As he points out, mathematics itself cannot be confined to the pages of a book. The results of mathematics, however, are shared and passed on primarily through the printed (and now electronic) medium. While there is no danger of mathematical results being lost through lack of dissemination, many people miss out on fascinating and useful mathematical results simply because they are not aware of them. Not only does collecting many results in one place provide a single starting point for mathematical exploration, but it should also lessen the aggravation of encountering explanations for new concepts which themselves use unfamiliar terminology. In this work, the reader is only a cross-reference (or a mouse click) away from the necessary background material. As to Sylvester's second point, the very fact that the quantity of mathematics is so great means that any attempt to catalog it with any degree of completeness is doomed to failure. This certainly does not mean that it's not worth trying. Strangely, except for relatively small works usually on particular subjects, there do not appear to have been any substantial attempts to collect and display in a place of prominence the treasure trove of mathematical results that have been discovered (invented?) over the years (one notable exception being Sloane and Plouffe's *Encyclopedia of Integer Sequences*). This work, the product of the "gazing" of a single astronomer, attempts to fill that omission.

Finally, a few words about logistics. Because of the alphabetical listing of entries in the encyclopedia, neither table of contents nor index are included. In many cases, a particular entry of interest can be located from a cross-reference (indicated in SMALL CAPS TYPEFACE in the text) in a related article. In addition, most articles are followed by a "see also" list of related entries for quick navigation. This can be particularly useful if you are looking for a specific entry (say, "Zeno's Paradoxes"), but have forgotten the exact name. By examining the "see also" list at bottom of the entry for "Paradox," you will likely recognize Zeno's name and thus quickly locate the desired entry.

The alphabetization of entries contains a few peculiarities which need mentioning. All entries beginning with a numeral are ordered by increasing value and appear before the first entry for "A." In multiple-word entries containing a space or dash, the space or dash is treated as a character which precedes "a," so entries appear in the following order: "Sum," "Sum P...," "Sum-P...," and "Summary." One exception is that in a series of entries where a trailing "s" appears in some and not others, the trailing "s" is ignored in the alphabetization. Therefore, entries involving Euclid would be alphabetized as follows: "Euclid's Axioms," "Euclid Number," "Euclidean Algorithm." Because of the non-standard nomenclature that ensues from naming mathematical results after their discoverers, an important result such as the "Pythagorean Theorem" is written variously as "Pythagoras's Theorem," the "Pythagoras Theorem," etc. In this encyclopedia, I have endeavored to use the most widely accepted form. I have also tried to consistently give entry titles in the singular (e.g., "Knot" instead of "Knots").

In cases where the same word is applied in different contexts, the context is indicated in parentheses or appended to the end. Examples of the first type are "Crossing Number (Graph)" and "Crossing Number (Link)." Examples of the second type are "Convergent Sequence" and "Convergent Series." In the case of an entry like "Euler Theorem," which may describe one of three or four different formulas, I have taken the liberty of adding descriptive words ("Euler's *Something* Theorem") to all variations, or kept the standard name for the most commonly used variant and added descriptive words for the others. In cases where specific examples are derived from a general concept, em dashes (—) are used (for example, "Fourier Series," "Fourier Series—Power Series," "Fourier Series—Square Wave," "Fourier Series—Triangle"). The decision to put a possessive 's at the end of a name or to use a lone trailing apostrophe is based on whether the final "s" is pronounced. "Gauss's Theorem" is therefore written out, whereas "Archimedes' Recurrence Formula" is not. Finally, given the absence of a definitive stylistic convention, plurals of numerals are written without an apostrophe (e.g., 1990s instead of 1990's).

In an endeavor of this magnitude, errors and typographical mistakes are inevitable. The blame for these lies with me alone. Although the current length makes extensive additions in a printed version problematic, I plan to continue updating, correcting, and improving the work.

Eric Weisstein

Charlottesville, Virginia
August 8, 1998

Acknowledgments

Although I alone have compiled and typeset this work, many people have contributed indirectly and directly to its creation. I have not yet had the good fortune to meet Donald Knuth of Stanford University, but he is unquestionably the person most directly responsible for making this work possible. Before his mathematical typesetting program TeX, it would have been impossible for a single individual to compile such a work as this. Had Prof. Bateman owned a personal computer equipped with TeX, perhaps his shoe box of notes would not have had to await the labors of Erdelyi, Magnus, and Oberhettinger to become a three-volume work on mathematical functions. Andrew Trevorrow's shareware implementation of TeX for the Macintosh, OzTeX (`www.kagi.com/authors/akt/oztex.html`), was also of fundamental importance. Nikos Drakos and Ross Moore have provided another building block for this work by developing the LaTeX2HTML program (`www-dsed.llnl.gov/files/programs/unix/latex2html/manual/manual.html`), which has allowed me to easily maintain and update an on-line version of the encyclopedia long before it existed in book form.

I would like to thank Steven Finch of MathSoft, Inc., for his interesting on-line essays about mathematical constants (`www.mathsoft.com/asolve/constant/constant.html`), and also for his kind permission to reproduce excerpts from some of these essays. I hope that Steven will someday publish his detailed essays in book form. Thanks also to Neil Sloane and Simon Plouffe for compiling and making available the printed and on-line (`www.research.att.com/~njas/sequences/`) versions of the *Encyclopedia of Integer Sequences*, an immensely valuable compilation of useful information which represents a truly mind-boggling investment of labor.

Thanks to Robert Dickau, Simon Plouffe, and Richard Schroeppel for reading portions of the manuscript and providing a number of helpful suggestions and additions. Thanks also to algebraic topologist Ryan Budney for sharing some of his expertise, to Charles Walkden for his helpful comments about dynamical systems theory, and to Lambros Lambrou for his contributions. Thanks to David W. Wilson for a number of helpful comments and corrections. Thanks to Dale Rolfsen, compiler James Bailey, and artist Ali Roth for permission to reproduce their beautiful knot and link diagrams. Thanks to Gavin Theobald for providing diagrams of his masterful polygonal dissections. Thanks to Wolfram Research, not only for creating an indispensable mathematical tool in *Mathematica*®, but also for permission to include figures from the *Mathematica*® book and *MathSource* repository for the braid, conical spiral, double helix, Enneper's surfaces, Hadamard matrix, helicoid, helix, Henneberg's minimal surface, hyperbolic polyhedra, Klein bottle, Maeder's "owl" minimal surface, Penrose tiles, polyhedron, and Scherk's minimal surfaces entries.

Sincere thanks to Judy Schroeder for her skill and diligence in the monumental task of proofreading the entire document for syntax. Thanks also to Bob Stern, my executive editor from *CRC Press*, for his encouragement, and to Mimi Williams of CRC Press for her careful reading of the manuscript for typographical and formatting errors. As this encyclopedia's entry on Proofreading Mistakes shows, the number of mistakes that are expected to remain after three independent proofreadings is much lower than the original number, but unfortunately still nonzero. Many thanks to the library staff at the University of Virginia, who have provided invaluable assistance in tracking down many an obscure citation. Finally, I would like to thank the hundreds of people who took the time to e-mail me comments and suggestions while this work was in its formative stages. Your continued comments and feedback are very welcome.

Numerals

0

see ZERO

1

The number one (1) is the first POSITIVE INTEGER. It is an ODD NUMBER. Although the number 1 used to be considered a PRIME NUMBER, it requires special treatment in so many definitions and applications involving primes greater than or equal to 2 that it is usually placed into a class of its own. The number 1 is sometimes also called "unity," so the nth roots of 1 are often called the nth ROOTS OF UNITY. FRACTIONS having 1 as a NUMERATOR are called UNIT FRACTIONS. If only one root, solution, etc., exists to a given problem, the solution is called UNIQUE.

The GENERATING FUNCTION have all COEFFICIENTS 1 is given by

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \ldots.$$

see also 2, 3, EXACTLY ONE, ROOT OF UNITY, UNIQUE, UNIT FRACTION, ZERO

2

The number two (2) is the second POSITIVE INTEGER and the first PRIME NUMBER. It is EVEN, and is the only EVEN PRIME (the PRIMES other than 2 are called the ODD PRIMES). The number 2 is also equal to its FACTORIAL since $2! = 2$. A quantity taken to the POWER 2 is said to be SQUARED. The number of times k a given BINARY number $b_n \cdots b_2 b_1 b_0$ is divisible by 2 is given by the position of the first $b_k = 1$, counting from the right. For example, $12 = 1100$ is divisible by 2 twice, and $13 = 1101$ is divisible by 2 0 times.

see also 1, BINARY, 3, SQUARED, ZERO

$2x$ mod 1 Map

Let x_0 be a REAL NUMBER in the CLOSED INTERVAL $[0, 1]$, and generate a SEQUENCE using the MAP

$$x_{n+1} \equiv 2x_n \pmod{1}. \tag{1}$$

Then the number of periodic ORBITS of period p (for p PRIME) is given by

$$N_p = \frac{2^p - 2}{p}. \tag{2}$$

Since a typical ORBIT visits each point with equal probability, the NATURAL INVARIANT is given by

$$\rho(x) = 1. \tag{3}$$

see also TENT MAP

References

Ott, E. *Chaos in Dynamical Systems.* Cambridge: Cambridge University Press, pp. 26–31, 1993.

3

3 is the only INTEGER which is the sum of the preceding POSITIVE INTEGERS ($1 + 2 = 3$) and the only number which is the sum of the FACTORIALS of the preceding POSITIVE INTEGERS ($1! + 2! = 3$). It is also the first ODD PRIME. A quantity taken to the POWER 3 is said to be CUBED.

see also 1, 2, $3x+1$ MAPPING, CUBED, PERIOD THREE THEOREM, SUPER-3 NUMBER, TERNARY, THREE-COLORABLE, ZERO

$3x + 1$ Mapping

see COLLATZ PROBLEM

10

The number 10 (ten) is the basis for the DECIMAL system of notation. In this system, each "decimal place" consists of a DIGIT 0–9 arranged such that each DIGIT is multiplied by a POWER of 10, decreasing from left to right, and with a decimal place indicating the $10^0 = 1$s place. For example, the number 1234.56 specifies

$$1 \times 10^3 + 2 \times 10^2 + 3 \times 10^1 + 4 \times 10^0 + 5 \times 10^{-1} + 6 \times 10^{-2}.$$

The decimal places to the left of the decimal point are 1, 10, 100, 1000, 10000, 10000, 100000, 10000000, 100000000, ... (Sloane's A011557), called one, ten, HUNDRED, THOUSAND, ten thousand, hundred thousand, MILLION, 10 million, 100 million, and so on. The names of subsequent decimal places for LARGE NUMBERS differ depending on country.

Any POWER of 10 which can be written as the PRODUCT of two numbers not containing 0s must be of the form $2^n \cdot 5^n = 10^n$ for n an INTEGER such that neither 2^n nor 5^n contains any ZEROS. The largest known such number is

$$\begin{aligned} 10^{33} &= 2^{33} \cdot 5^{33} \\ &= 8,589,934,592 \cdot 116,415,321,826,934,814,453,125. \end{aligned}$$

A complete list of known such numbers is

$$\begin{aligned} 10^1 &= 2^1 \cdot 5^1 \\ 10^2 &= 2^2 \cdot 5^2 \\ 10^3 &= 2^3 \cdot 5^3 \\ 10^4 &= 2^4 \cdot 5^4 \\ 10^5 &= 2^5 \cdot 5^5 \\ 10^6 &= 2^6 \cdot 5^6 \\ 10^7 &= 2^7 \cdot 5^7 \\ 10^9 &= 2^9 \cdot 5^9 \\ 10^{18} &= 2^{18} \cdot 5^{18} \\ 10^{33} &= 2^{33} \cdot 5^{33} \end{aligned}$$

(Madachy 1979). Since all POWERS of 2 with exponents $n \leq 4.6 \times 10^7$ contain at least one ZERO (M. Cook), no other POWER of ten less than 46 million can be written as the PRODUCT of two numbers not containing 0s.

see also BILLION, DECIMAL, HUNDRED, LARGE NUMBER, MILLIARD, MILLION, THOUSAND, TRILLION, ZERO

References
Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 127–128, 1979.
Pickover, C. A. *Keys to Infinity.* New York: W. H. Freeman, p. 135, 1995.
Sloane, N. J. A. Sequence A011557 in "An On-Line Version of the Encyclopedia of Integer Sequences."

12

One DOZEN, or a twelfth of a GROSS.

see also DOZEN, GROSS

13

A NUMBER traditionally associated with bad luck. A so-called BAKER'S DOZEN is equal to 13. Fear of the number 13 is called TRISKAIDEKAPHOBIA.

see also BAKER'S DOZEN, FRIDAY THE THIRTEENTH, TRISKAIDEKAPHOBIA

15

see 15 PUZZLE, FIFTEEN THEOREM

15 Puzzle

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	

A puzzle introduced by Sam Loyd in 1878. It consists of 15 squares numbered from 1 to 15 which are placed in a 4×4 box leaving one position out of the 16 empty. The goal is to rearrange the squares from a given arbitrary starting arrangement by sliding them one at a time into the configuration shown above. For some initial arrangements, this rearrangement is possible, but for others, it is not.

To address the solubility of a given initial arrangement, proceed as follows. If the SQUARE containing the number i appears "before" (reading the squares in the box from left to right and top to bottom) n numbers which are less than i, then call it an inversion of order n, and denote it n_i. Then define

$$N \equiv \sum_{i=1}^{15} n_i = \sum_{i=2}^{15} n_i,$$

where the sum need run only from 2 to 15 rather than 1 to 15 since there are no numbers less than 1 (so n_1 must equal 0). If N is EVEN, the position is possible, otherwise it is not. This can be formally proved using ALTERNATING GROUPS. For example, in the following arrangement

2	1	3	4
5	6	7	8
9	10	11	12
13	14	15	

$n_2 = 1$ (2 precedes 1) and all other $n_i = 0$, so $N = 1$ and the puzzle cannot be solved.

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 312–316, 1987.
Bogomolny, A. "Sam Loyd's Fifteen." `http://www.cut-the-knot.com/pythagoras/fifteen.html`.
Bogomolny, A. "Sam Loyd's Fifteen [History]." `http://www.cut-the-knot.com/pythagoras/history15.html`.
Johnson, W. W. "Notes on the '15 Puzzle. I.'" *Amer. J. Math.* **2**, 397–399, 1879.
Kasner, E. and Newman, J. R. *Mathematics and the Imagination.* Redmond, WA: Tempus Books, pp. 177–180, 1989.
Kraitchik, M. "The 15 Puzzle." §12.2.1 in *Mathematical Recreations.* New York: W. W. Norton, pp. 302–308, 1942.
Story, W. E. "Notes on the '15 Puzzle. II.'" *Amer. J. Math.* **2**, 399–404, 1879.

16-Cell

A finite regular 4-D POLYTOPE with SCHLÄFLI SYMBOL {3, 3, 4} and VERTICES which are the PERMUTATIONS of (± 1, 0, 0, 0).

see also 24-CELL, 120-CELL, 600-CELL, CELL, POLYTOPE

17

17 is a FERMAT PRIME which means that the 17-sided REGULAR POLYGON (the HEPTADECAGON) is CONSTRUCTIBLE using COMPASS and STRAIGHTEDGE (as proved by Gauss).

see also CONSTRUCTIBLE POLYGON , FERMAT PRIME, HEPTADECAGON

References
Carr, M. "Snow White and the Seven(teen) Dwarfs." `http://www.math.harvard.edu/~hmb/issue2.1/SEVENTEEN/seventeen.html`.
Fischer, R. "Facts About the Number 17." `http://tempo.harvard.edu/~rfischer/hcssim/17_facts/kelly/kelly.html`.
Lefevre, V. "Properties of 17." `http://www.ens-lyon.fr/~vlefevre/d17_eng.html`.
Shell Centre for Mathematical Education. "Number 17." `http://acorn.educ.nottingham.ac.uk/ShellCent/Number/Num17.html`.

18-Point Problem

Place a point somewhere on a LINE SEGMENT. Now place a second point and number it 2 so that each of the points is in a different half of the LINE SEGMENT. Continue, placing every Nth point so that all N points are on different $(1/N)$th of the LINE SEGMENT. Formally, for a given N, does there exist a sequence of real numbers x_1, x_2, …, x_N such that for every $n \in \{1, \ldots, N\}$ and every $k \in \{1, \ldots, n\}$, the inequality

$$\frac{k-1}{n} \leq x_i < \frac{k}{n}$$

holds for some $i \in \{1, \ldots, n\}$? Surprisingly, it is only possible to place 17 points in this manner (Berlekamp and Graham 1970, Warmus 1976).

Steinhaus (1979) gives a 14-point solution (0.06, 0.55, 0.77, 0.39, 0.96, 0.28, 0.64, 0.13, 0.88, 0.48, 0.19, 0.71, 0.35, 0.82), and Warmus (1976) gives the 17-point solution

$$\begin{gathered}\tfrac{4}{7} \leq x_1 < \tfrac{7}{12}, \tfrac{2}{7} \leq x_2 < \tfrac{5}{17}, \tfrac{16}{17} \leq x_3 < 1, \tfrac{1}{14} \leq x_4 < \tfrac{1}{13}, \\ \tfrac{8}{11} \leq x_5 < \tfrac{11}{15}, \tfrac{5}{11} \leq x_6 < \tfrac{6}{13}, \tfrac{1}{7} \leq x_7 < \tfrac{2}{13}, \tfrac{14}{17} \leq x_8 < \tfrac{5}{6}, \\ \tfrac{3}{8} \leq x_9 < \tfrac{5}{13}, \tfrac{11}{17} \leq x_{10} < \tfrac{2}{3}, \tfrac{3}{14} \leq x_{11} < \tfrac{3}{13}, \\ \tfrac{15}{17} \leq x_{12} < \tfrac{11}{12}, \tfrac{1}{2} \leq x_{12} < \tfrac{9}{17}, 0 \leq x_{14} < \tfrac{1}{17}, \\ \tfrac{13}{17} \leq x_{15} < \tfrac{4}{5}, \tfrac{5}{16} \leq x_{16} < \tfrac{6}{17}, \tfrac{10}{17} \leq x_{17} < \tfrac{11}{17}.\end{gathered}$$

Warmus (1976) states that there are 768 patterns of 17-point solutions (counting reversals as equivalent).

see also Discrepancy Theorem, Point Picking

References

Berlekamp, E. R. and Graham, R. L. "Irregularities in the Distributions of Finite Sequences." *J. Number Th.* **2**, 152–161, 1970.

Gardner, M. *The Last Recreations: Hydras, Eggs, and Other Mathematical Mystifications.* New York: Springer-Verlag, pp. 34–36, 1997.

Steinhaus, H. "Distribution on Numbers" and "Generalization." Problems 6 and 7 in *One Hundred Problems in Elementary Mathematics.* New York: Dover, pp. 12–13, 1979.

Warmus, M. "A Supplementary Note on the Irregularities of Distributions." *J. Number Th.* **8**, 260–263, 1976.

24-Cell

A finite regular 4-D Polytope with Schläfli Symbol $\{3, 4, 3\}$. Coxeter (1969) gives a list of the Vertex positions. The Even coefficients of the D_4 lattice are 1, 24, 24, 96, ... (Sloane's A004011), and the 24 shortest vectors in this lattice form the 24-cell (Coxeter 1973, Conway and Sloane 1993, Sloane and Plouffe 1995).

see also 16-Cell, 120-Cell, 600-Cell, Cell, Polytope

References

Conway, J. H. and Sloane, N. J. A. *Sphere-Packings, Lattices and Groups, 2nd ed.* New York: Springer-Verlag, 1993.

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, p. 404, 1969.

Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, 1973.

Sloane, N. J. A. Sequences A004011/M5140 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sloane, N. J. A. and Plouffe, S. Extended entry in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.

42

According to Adams, 42 is the ultimate answer to life, the universe, and everything, although it is left as an exercise to the reader to determine the actual question leading to this result.

References

Adams, D. *The Hitchhiker's Guide to the Galaxy.* New York: Ballantine Books, 1997.

72 Rule

see Rule of 72

120-Cell

A finite regular 4-D Polytope with Schläfli Symbol $\{5, 3, 3\}$ (Coxeter 1969).

see also 16-Cell, 24-Cell, 600-Cell, Cell, Polytope

References

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, p. 404, 1969.

144

A Dozen Dozen, also called a Gross. 144 is a Square Number and a Sum-Product Number.

see also Dozen

196-Algorithm

Take any Positive Integer of two Digits or more, reverse the Digits, and add to the original number. Now repeat the procedure with the Sum so obtained. This procedure quickly produces Palindromic Numbers for most Integers. For example, starting with the number 5280 produces (5280, 6105, 11121, 23232). The end results of applying the algorithm to 1, 2, 3, ... are 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 11, 33, 44, 55, 66, 77, 88, 99, 121, ... (Sloane's A033865). The value for 89 is especially large, being 8813200023188.

The first few numbers not known to produce Palindromes are 196, 887, 1675, 7436, 13783, ... (Sloane's A006960), which are simply the numbers obtained by iteratively applying the algorithm to the number 196. This number therefore lends itself to the name of the Algorithm.

The number of terms $a(n)$ in the iteration sequence required to produce a Palindromic Number from n (i.e., $a(n) = 1$ for a Palindromic Number, $a(n) = 2$ if a Palindromic Number is produced after a single iteration of the 196-algorithm, etc.) for $n = 1, 2, \ldots$ are 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 2, 2, 2, 2, 2, 2, 2, 3, 2, 2, 1, ... (Sloane's A030547). The smallest numbers which require $n = 0, 1, 2, \ldots$ iterations to reach a palindrome are 0, 10, 19, 59, 69, 166, 79, 188, ... (Sloane's A023109).

see also Additive Persistence, Digitadition, Multiplicative Persistence, Palindromic Number, Palindromic Number Conjecture, RATS Sequence, Recurring Digital Invariant

References

Gardner, M. *Mathematical Circus: More Puzzles, Games, Paradoxes and Other Mathematical Entertainments from Scientific American.* New York: Knopf, pp. 242–245, 1979.

Gruenberger, F. "How to Handle Numbers with Thousands of Digits, and Why One Might Want to." *Sci. Amer.* **250**, 19–26, Apr. 1984.

Sloane, N. J. A. Sequences A023109, A030547, A033865, and A006960/M5410 in "An On-Line Version of the Encyclopedia of Integer Sequences."

239

Some interesting properties (as well as a few arcane ones not reiterated here) of the number 239 are discussed in Beeler *et al.* (1972, Item 63). 239 appears in MACHIN'S FORMULA

$$\tfrac{1}{4}\pi = 4\tan(\tfrac{1}{5}) - \tan^{-1}(\tfrac{1}{239}),$$

which is related to the fact that

$$2 \cdot 13^4 - 1 = 239^2,$$

which is why 239/169 is the 7th CONVERGENT of $\sqrt{2}$. Another pair of INVERSE TANGENT FORMULAS involving 239 is

$$\begin{aligned}\tan^{-1}(\tfrac{1}{239}) &= \tan^{-1}(\tfrac{1}{70}) - \tan^{-1}(\tfrac{1}{99}) \\ &= \tan^{-1}(\tfrac{1}{408}) + \tan^{-1}(\tfrac{1}{577}).\end{aligned}$$

239 needs 4 SQUARES (the maximum) to express it, 9 CUBES (the maximum, shared only with 23) to express it, and 19 fourth POWERS (the maximum) to express it (see WARING'S PROBLEM). However, 239 doesn't need the maximum number of fifth POWERS (Beeler *et al.* 1972, Item 63).

References
Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

257-gon

257 is a FERMAT PRIME, and the 257-gon is therefore a CONSTRUCTIBLE POLYGON using COMPASS and STRAIGHTEDGE, as proved by Gauss. An illustration of the 257-gon is not included here, since its 257 segments so closely resemble a CIRCLE. Richelot and Schwendenwein found constructions for the 257-gon in 1832 (Coxeter 1969). De Temple (1991) gives a construction using 150 CIRCLES (24 of which are CARLYLE CIRCLES) which has GEOMETROGRAPHY symbol $94S_1 + 47S_2 + 275C_1 + 0C_2 + 150C_3$ and SIMPLICITY 566.

see also 65537-GON, CONSTRUCTIBLE POLYGON, FERMAT PRIME, HEPTADECAGON, PENTAGON

References
Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, 1969.
De Temple, D. W. "Carlyle Circles and the Lemoine Simplicity of Polygonal Constructions." *Amer. Math. Monthly* **98**, 97–108, 1991.
Dixon, R. *Mathographics.* New York: Dover, p. 53, 1991.
Rademacher, H. *Lectures on Elementary Number Theory.* New York: Blaisdell, 1964.

600-Cell

A finite regular 4-D POLYTOPE with SCHLÄFLI SYMBOL $\{3,3,5\}$. For VERTICES, see Coxeter (1969).

see also 16-CELL, 24-CELL, 120-CELL, CELL, POLYTOPE

References
Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, p. 404, 1969.

666

A number known as the BEAST NUMBER appearing in the *Bible* and ascribed various numerological properties.

see also APOCALYPTIC NUMBER, BEAST NUMBER, LEVIATHAN NUMBER

References
Hardy, G. H. *A Mathematician's Apology, reprinted with a foreword by C. P. Snow.* New York: Cambridge University Press, p. 96, 1993.

2187

The digits in the number 2187 form the two VAMPIRE NUMBERS: 21 × 87 = 1827 and 2187 = 27 × 81.

References
Gardner, M. "Lucky Numbers and 2187." *Math. Intell.* **19**, 26–29, Spring 1997.

65537-gon

65537 is the largest known FERMAT PRIME, and the 65537-gon is therefore a CONSTRUCTIBLE POLYGON using COMPASS and STRAIGHTEDGE, as proved by Gauss. The 65537-gon has so many sides that it is, for all intents and purposes, indistinguishable from a CIRCLE using any reasonable printing or display methods. Hermes spent 10 years on the construction of the 65537-gon at Göttingen around 1900 (Coxeter 1969). De Temple (1991) notes that a GEOMETRIC CONSTRUCTION can be done using 1332 or fewer CARLYLE CIRCLES.

see also 257-GON, CONSTRUCTIBLE POLYGON, HEPTADECAGON, PENTAGON

References
Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, 1969.
De Temple, D. W. "Carlyle Circles and the Lemoine Simplicity of Polygonal Constructions." *Amer. Math. Monthly* **98**, 97–108, 1991.
Dixon, R. *Mathographics.* New York: Dover, p. 53, 1991.

A

A-Integrable

A generalization of the LEBESGUE INTEGRAL. A MEASURABLE FUNCTION $f(x)$ is called A-integrable over the CLOSED INTERVAL $[a, b]$ if

$$m\{x : |f(x)| > n\} = \mathcal{O}(n^{-1}), \tag{1}$$

where m is the LEBESGUE MEASURE, and

$$I = \lim_{n\to\infty} \int_a^b [f(x)]_n \, dx \tag{2}$$

exists, where

$$[f(x)]_n = \begin{cases} f(x) & \text{if } |f(x)| \leq n \\ 0 & \text{if } |f(x)| > n. \end{cases} \tag{3}$$

References

Titmarsch, E. G. "On Conjugate Functions." *Proc. London Math. Soc.* **29**, 49–80, 1928.

A-Sequence

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

An INFINITE SEQUENCE of POSITIVE INTEGERS a_i satisfying

$$1 \leq a_1 < a_2 < a_3 < \ldots \tag{1}$$

is an A-sequence if no a_k is the SUM of two or more distinct earlier terms (Guy 1994). Erdős (1962) proved

$$S(A) \equiv \sup_{\text{all } A \text{ sequences}} \sum_{k=1}^{\infty} \frac{1}{a_k} < 103. \tag{2}$$

Any A-sequence satisfies the CHI INEQUALITY (Levine and O'Sullivan 1977), which gives $S(A) < 3.9998$. Abbott (1987) and Zhang (1992) have given a bound from below, so the best result to date is

$$2.0649 < S(A) < 3.9998. \tag{3}$$

Levine and O'Sullivan (1977) conjectured that the sum of RECIPROCALS of an A-sequence satisfies

$$S(A) \leq \sum_{k=1}^{\infty} \frac{1}{\chi_k} = 3.01\ldots, \tag{4}$$

where χ_i are given by the LEVINE-O'SULLIVAN GREEDY ALGORITHM.

see also B_2-SEQUENCE, MIAN-CHOWLA SEQUENCE

References

Abbott, H. L. "On Sum-Free Sequences." *Acta Arith.* **48**, 93–96, 1987.
Erdős, P. "Remarks on Number Theory III. Some Problems in Additive Number Theory." *Mat. Lapok* **13**, 28–38, 1962.
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/erdos/erdos.html`.
Guy, R. K. "B_2-Sequences." §E28 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 228–229, 1994.
Levine, E. and O'Sullivan, J. "An Upper Estimate for the Reciprocal Sum of a Sum-Free Sequence." *Acta Arith.* **34**, 9–24, 1977.
Zhang, Z. X. "A Sum-Free Sequence with Larger Reciprocal Sum." Unpublished manuscript, 1992.

AAA Theorem

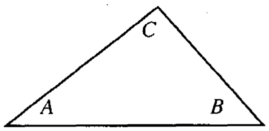

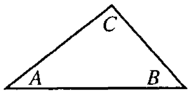

Specifying three ANGLES A, B, and C does not uniquely define a TRIANGLE, but any two TRIANGLES with the same ANGLES are SIMILAR. Specifying two ANGLES of a TRIANGLE automatically gives the third since the sum of ANGLES in a TRIANGLE sums to 180° (π RADIANS), i.e.,

$$C = \pi - A - B.$$

see also AAS THEOREM, ASA THEOREM, ASS THEOREM, SAS THEOREM, SSS THEOREM, TRIANGLE

AAS Theorem

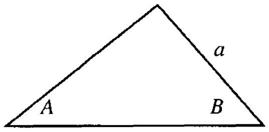

Specifying two angles A and B and a side a uniquely determines a TRIANGLE with AREA

$$K = \frac{a^2 \sin B \sin C}{2 \sin A} = \frac{a^2 \sin B \sin(\pi - A - B)}{2 \sin A}. \tag{1}$$

The third angle is given by

$$C = \pi - A - B, \tag{2}$$

since the sum of angles of a TRIANGLE is 180° (π RADIANS). Solving the LAW OF SINES

$$\frac{a}{\sin A} = \frac{b}{\sin B} \tag{3}$$

for b gives

$$b = a \frac{\sin B}{\sin A}. \tag{4}$$

Finally,

$$\begin{aligned} c &= b \cos A + a \cos B = a(\sin B \cot A + \cos B) \quad &(5) \\ &= a \sin B(\cot A + \cot B). \quad &(6) \end{aligned}$$

see also AAA THEOREM, ASA THEOREM, ASS THEOREM, SAS THEOREM, SSS THEOREM, TRIANGLE

Abacus

A mechanical counting device consisting of a frame holding a series of parallel rods on each of which beads are strung. Each bead represents a counting unit, and each rod a place value. The primary purpose of the abacus is not to perform actual computations, but to provide a quick means of storing numbers during a calculation. Abaci were used by the Japanese and Chinese, as well as the Romans.

see also ROMAN NUMERAL, SLIDE RULE

References

Boyer, C. B. and Merzbach, U. C. "The Abacus and Decimal Fractions." *A History of Mathematics, 2nd ed.* New York: Wiley, pp. 199–201, 1991.

Fernandes, L. "The Abacus: The Art of Calculating with Beads." `http://www.ee.ryerson.ca:8080/~elf/abacus`.

Gardner, M. "The Abacus." Ch. 18 in *Mathematical Circus: More Puzzles, Games, Paradoxes and Other Mathematical Entertainments from Scientific American.* New York: Knopf, pp. 232–241, 1979.

Pappas, T. "The Abacus." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 209, 1989.

Smith, D. E. "Mechanical Aids to Calculation: The Abacus." Ch. 3 §1 in *History of Mathematics, Vol. 2.* New York: Dover, pp. 156–196, 1958.

abc Conjecture

A CONJECTURE due to J. Oesterlé and D. W. Masser. It states that, for any INFINITESIMAL $\epsilon > 0$, there exists a CONSTANT C_ϵ such that for any three RELATIVELY PRIME INTEGERS a, b, c satisfying

$$a + b = c,$$

the INEQUALITY

$$\max\{|a|, |b|, |c|\} \leq C_\epsilon \prod_{p|abc} p^{1+\epsilon}$$

holds, where $p|abc$ indicates that the PRODUCT is over PRIMES p which DIVIDE the PRODUCT abc. If this CONJECTURE were true, it would imply FERMAT'S LAST THEOREM for sufficiently large POWERS (Goldfeld 1996). This is related to the fact that the abc conjecture implies that there are at least $C \ln x$ WIEFERICH PRIMES $\leq x$ for some constant C (Silverman 1988, Vardi 1991).

see also FERMAT'S LAST THEOREM, MASON'S THEOREM, WIEFERICH PRIME

References

Cox, D. A. "Introduction to Fermat's Last Theorem." *Amer. Math. Monthly* **101**, 3–14, 1994.

Goldfeld, D. "Beyond the Last Theorem." *The Sciences*, 34–40, March/April 1996.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 75–76, 1994.

Silverman, J. "Wieferich's Criterion and the abc Conjecture." *J. Number Th.* **30**, 226–237, 1988.

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 66, 1991.

Abelian

see ABELIAN CATEGORY, ABELIAN DIFFERENTIAL, ABELIAN FUNCTION, ABELIAN GROUP, ABELIAN INTEGRAL, ABELIAN VARIETY, COMMUTATIVE

Abelian Category

An Abelian category is an abstract mathematical CATEGORY which displays some of the characteristic properties of the CATEGORY of all ABELIAN GROUPS.

see also ABELIAN GROUP, CATEGORY

Abel's Curve Theorem

The sum of the values of an INTEGRAL of the "first" or "second" sort

$$\int_{x_0,y_0}^{x_1,y_1} \frac{P\,dx}{Q} + \ldots + \int_{x_0,y_0}^{x_N,y_N} \frac{P\,dx}{Q} = F(z)$$

and

$$\frac{P(x_1,y_1)}{Q(x_1,y_1)}\frac{dx_1}{dz} + \ldots + \frac{P(x_N,y_N)}{Q(x_N,y_N)}\frac{dx_N}{dz} = \frac{dF}{dz},$$

from a FIXED POINT to the points of intersection with a curve depending rationally upon any number of parameters is a RATIONAL FUNCTION of those parameters.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 277, 1959.

Abelian Differential

An Abelian differential is an ANALYTIC or MEROMORPHIC DIFFERENTIAL on a COMPACT or closed RIEMANN SURFACE.

Abelian Function

An INVERSE FUNCTION of an ABELIAN INTEGRAL. Abelian functions have two variables and four periods. They are a generalization of ELLIPTIC FUNCTIONS, and are also called HYPERELLIPTIC FUNCTIONS.

see also ABELIAN INTEGRAL, ELLIPTIC FUNCTION

References

Baker, H. F. *Abelian Functions: Abel's Theorem and the Allied Theory, Including the Theory of the Theta Functions.* New York: Cambridge University Press, 1995.

Baker, H. F. *An Introduction to the Theory of Multiply Periodic Functions.* London: Cambridge University Press, 1907.

Abel's Functional Equation

Let $\mathrm{Li}_2(x)$ denote the DILOGARITHM, defined by

$$\mathrm{Li}_2(x) = \sum_{n=1}^{\infty} \frac{x^n}{n^2},$$

then

$$\mathrm{Li}_2(x) + \mathrm{Li}_2(y) + \mathrm{Li}_2(xy) + \mathrm{Li}_2\left(\frac{x(1-y)}{1-xy}\right) + \mathrm{Li}_2\left(\frac{y(1-x)}{1-xy}\right) = 3\,\mathrm{Li}_2(1).$$

see also DILOGARITHM, POLYLOGARITHM, RIEMANN ZETA FUNCTION

Abelian Group

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

A GROUP for which the elements COMMUTE (i.e., $AB = BA$ for all elements A and B) is called an Abelian group. All CYCLIC GROUPS are Abelian, but an Abelian group is not necessarily CYCLIC. All SUBGROUPS of an Abelian group are NORMAL. In an Abelian group, each element is in a CONJUGACY CLASS by itself, and the CHARACTER TABLE involves POWERS of a single element known as a GENERATOR.

No general formula is known for giving the number of nonisomorphic FINITE GROUPS of a given ORDER. However, the number of nonisomorphic Abelian FINITE GROUPS $a(n)$ of any given ORDER n is given by writing n as

$$n = \prod_i {p_i}^{\alpha_i}, \tag{1}$$

where the p_i are distinct PRIME FACTORS, then

$$a(n) = \prod_i P(\alpha_i), \tag{2}$$

where P is the PARTITION FUNCTION. This gives 1, 1, 1, 2, 1, 1, 1, 3, 2, ... (Sloane's A000688). The smallest orders for which $n = 1, 2, 3, \ldots$ nonisomorphic Abelian groups exist are 1, 4, 8, 36, 16, 72, 32, 900, 216, 144, 64, 1800, 0, 288, 128, ... (Sloane's A046056), where 0 denotes an impossible number (i.e., not a product of partition numbers) of nonisomorphic Abelian, groups. The "missing" values are 13, 17, 19, 23, 26, 29, 31, 34, 37, 38, 39, 41, 43, 46, ... (Sloane's A046064). The incrementally largest numbers of Abelian groups as a function of order are 1, 2, 3, 5, 7, 11, 15, 22, 30, 42, 56, 77, 101, ... (Sloane's A046054), which occur for orders 1, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192, ... (Sloane's A046055).

The KRONECKER DECOMPOSITION THEOREM states that every FINITE Abelian group can be written as a DIRECT PRODUCT of CYCLIC GROUPS of PRIME POWER ORDERS. If the ORDERS of a FINITE GROUP is a PRIME p, then there exists a single Abelian group of order p (denoted Z_p) and no non-Abelian groups. If the ORDERS is a prime squared p^2, then there are two Abelian groups (denoted Z_{p^2} and $Z_p \otimes Z_p$. If the ORDERS is a prime cubed p^3, then there are three Abelian groups (denoted $Z_p \otimes Z_p \otimes Z_p$, $Z_p \otimes Z_{p^2}$, and Z_{p^3}), and five groups total. If the order is a PRODUCT of two primes p and q, then there exists exactly one Abelian group of order pq (denoted $Z_p \otimes Z_q$).

Another interesting result is that if $a(n)$ denotes the number of nonisomorphic Abelian groups of ORDER n, then

$$\sum_{n=1}^{\infty} a(n) n^{-s} = \zeta(s)\zeta(2s)\zeta(3s)\cdots, \tag{3}$$

where $\zeta(s)$ is the RIEMANN ZETA FUNCTION. Srinivasan (1973) has also shown that

$$\sum_{n=1}^{N} a(n) = A_1 N + A_2 N^{1/2} + A_3 N^{1/3} + \mathcal{O}[x^{105/407}(\ln x)^2], \tag{4}$$

where

$$A_k \equiv \prod_{\substack{j=1\\ j\neq k}} \zeta\left(\frac{j}{k}\right) = \begin{cases} 2.294856591\ldots & \text{for } k=1 \\ -14.6475663\ldots & \text{for } k=2 \\ 118.6924619\ldots & \text{for } k=3, \end{cases} \tag{5}$$

and ζ is again the RIEMANN ZETA FUNCTION. [Richert (1952) incorrectly gave $A_3 = 114$.] DeKoninck and Ivic (1980) showed that

$$\sum_{n=1}^{N} \frac{1}{a(n)} = BN + \mathcal{O}[\sqrt{N}(\ln N)^{-1/2}], \tag{6}$$

where

$$B \equiv \prod \left\{1 - \sum_{k=2}^{\infty} \left[\frac{1}{P(k-2)} - \frac{1}{P(k)}\right] \frac{1}{p^k}\right\} = 0.752\ldots \tag{7}$$

is a product over PRIMES. Bounds for the number of nonisomorphic non-Abelian groups are given by Neumann (1969) and Pyber (1993).

see also FINITE GROUP, GROUP THEORY, KRONECKER DECOMPOSITION THEOREM, PARTITION FUNCTION P, RING

References

DeKoninck, J.-M. and Ivic, A. *Topics in Arithmetical Functions: Asymptotic Formulae for Sums of Reciprocals of Arithmetical Functions and Related Fields.* Amsterdam, Netherlands: North-Holland, 1980.

Erdős, P. and Szekeres, G. "Über die Anzahl abelscher Gruppen gegebener Ordnung und über ein verwandtes zahlentheoretisches Problem." *Acta Sci. Math. (Szeged)* **7**, 95–102, 1935.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/abel/abel.html`.

Kendall, D. G. and Rankin, R. A. "On the Number of Abelian Groups of a Given Order." *Quart. J. Oxford* **18**, 197–208, 1947.

Kolesnik, G. "On the Number of Abelian Groups of a Given Order." *J. Reine Angew. Math.* **329**, 164–175, 1981.

Neumann, P. M. "An Enumeration Theorem for Finite Groups." *Quart. J. Math. Ser. 2* **20**, 395–401, 1969.
Pyber, L. "Enumerating Finite Groups of Given Order." *Ann. Math.* **137**, 203–220, 1993.
Richert, H.-E. "Über die Anzahl abelscher Gruppen gegebener Ordnung I." *Math. Zeitschr.* **56**, 21–32, 1952.
Sloane, N. J. A. Sequence A000688/M0064 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Srinivasan, B. R. "On the Number of Abelian Groups of a Given Order." *Acta Arith.* **23**, 195–205, 1973.

Abel's Identity

Given a homogeneous linear SECOND-ORDER ORDINARY DIFFERENTIAL EQUATION,

$$y'' + P(x)y' + Q(x)y = 0, \tag{1}$$

call the two linearly independent solutions $y_1(x)$ and $y_2(x)$. Then

$$y_1''(x) + P(x)y_1'(x) + Q(x)y_1 = 0 \tag{2}$$

$$y_2''(x) + P(x)y_2'(x) + Q(x)y_2 = 0. \tag{3}$$

Now, take $y_1 \times (3) - y_2 \times (2)$,

$$\begin{aligned} y_1[y_2'' + P(x)y_2' + Q(x)y_2] \\ -y_2[y_1'' + P(x)y_1' + Q(x)y_1] = 0 \end{aligned} \tag{4}$$

$$(y_1y_2''-y_2y_1'')+P(y_1y_2'-y_1'y_2)+Q(y_1y_2-y_1y_2) = 0 \tag{5}$$

$$(y_1y_2'' - y_2y_1'') + P(y_1y_2' - y_1'y_2) = 0. \tag{6}$$

Now, use the definition of the WRONSKIAN and take its DERIVATIVE,

$$W \equiv y_1y_2' - y_1'y_2 \tag{7}$$

$$\begin{aligned} W' &= (y_1'y_2' + y_1y_2'') - (y_1'y_2' + y_1''y_2) \\ &= y_1y_2'' - y_1''y_2. \end{aligned} \tag{8}$$

Plugging W and W' into (6) gives

$$W' + PW = 0. \tag{9}$$

This can be rearranged to yield

$$\frac{dW}{W} = -P(x)\,dx \tag{10}$$

which can then be directly integrated to

$$\ln W = -C_1 \int P(x)\,dx, \tag{11}$$

where $\ln x$ is the NATURAL LOGARITHM. A second integration then yields Abel's identity

$$W(x) = C_2 e^{-\int P(x)\,dx}, \tag{12}$$

where C_1 is a constant of integration and $C_2 \equiv e^{C_1}$.

see also ORDINARY DIFFERENTIAL EQUATION—SECOND-ORDER

References

Boyce, W. E. and DiPrima, R. C. *Elementary Differential Equations and Boundary Value Problems, 4th ed.* New York: Wiley, pp. 118, 262, 277, and 355, 1986.

Abel's Impossibility Theorem

In general, POLYNOMIAL equations higher than fourth degree are incapable of algebraic solution in terms of a finite number of ADDITIONS, MULTIPLICATIONS, and ROOT extractions.

see also CUBIC EQUATION, GALOIS'S THEOREM, POLYNOMIAL, QUADRATIC EQUATION, QUARTIC EQUATION, QUINTIC EQUATION

References

Abel, N. H. "Démonstration de l'impossibilité de la résolution algébraique des équations générales qui dépassent le quatrième degré." *Crelle's J.* **1**, 1826.

Abel's Inequality

Let $\{f_n\}$ and $\{a_n\}$ be SEQUENCES with $f_n \geq f_{n+1} > 0$ for $n = 1, 2, \ldots$, then

$$\left|\sum_{n=1}^{m} a_n f_n\right| \leq A f_1,$$

where

$$A = \max\{|a_1|, |a_1 + a_2|, \ldots, |a_1 + a_2 + \ldots + a_m|\}.$$

Abelian Integral

An INTEGRAL of the form

$$\int_0^x \frac{dt}{\sqrt{R(t)}},$$

where $R(t)$ is a POLYNOMIAL of degree > 4. They are also called HYPERELLIPTIC INTEGRALS.

see also ABELIAN FUNCTION, ELLIPTIC INTEGRAL

Abel's Irreducibility Theorem

If one ROOT of the equation $f(x) = 0$, which is irreducible over a FIELD K, is also a ROOT of the equation $F(x) = 0$ in K, then all the ROOTS of the irreducible equation $f(x) = 0$ are ROOTS of $F(x) = 0$. Equivalently, $F(x)$ can be divided by $f(x)$ without a REMAINDER,

$$F(x) = f(x)F_1(x),$$

where $F_1(x)$ is also a POLYNOMIAL over K.

see also ABEL'S LEMMA, KRONECKER'S POLYNOMIAL THEOREM, SCHOENEMANN'S THEOREM

References

Abel, N. H. "Mémoir sur une classe particulière d'équations résolubles algébraiquement." *Crelle's J.* **4**, 1829.
Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, p. 120, 1965.

Abel's Lemma

The pure equation

$$x^p = C$$

of Prime degree p is irreducible over a Field when C is a number of the Field but not the pth Power of an element of the Field.

see also Abel's Irreducibility Theorem, Gauss's Polynomial Theorem, Kronecker's Polynomial Theorem, Schoenemann's Theorem

References

Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, p. 118, 1965.

Abel's Test

see Abel's Uniform Convergence Test

Abel's Theorem

Given a Taylor Series

$$F(z) = \sum_{n=0}^{\infty} C_n z^n = \sum_{n=0}^{\infty} C_n r^n e^{in\theta}, \tag{1}$$

where the Complex Number z has been written in the polar form $z = re^{i\theta}$, examine the Real and Imaginary Parts

$$u(r,\theta) = \sum_{n=0}^{\infty} C_n r^n \cos(n\theta) \tag{2}$$

$$v(r,\theta) = \sum_{n=0}^{\infty} C_n r^n \sin(n\theta). \tag{3}$$

Abel's theorem states that, if $u(1,\theta)$ and $v(1,\theta)$ are Convergent, then

$$u(1,\theta) + iv(1,\theta) = \lim_{r\to 1} f(re^{i\theta}). \tag{4}$$

Stated in words, Abel's theorem guarantees that, if a Real Power Series Converges for some Positive value of the argument, the Domain of Uniform Convergence extends at least up to and including this point. Furthermore, the continuity of the sum function extends at least up to and including this point.

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 773, 1985.

Abel Transform

The following Integral Transform relationship, known as the Abel transform, exists between two functions $f(x)$ and $g(t)$ for $0 < \alpha < 1$,

$$f(x) = \int_0^x \frac{g(t)\,dt}{(x-t)^\alpha} \tag{1}$$

$$g(t) = -\frac{\sin(\pi\alpha)}{\pi}\frac{d}{dt}\int_0^t \frac{f(x)\,dx}{(x-t)^{1-\alpha}} \tag{2}$$

$$= -\frac{\sin(\pi\alpha)}{\pi}\left[\int_0^t \frac{df}{dx}\frac{dx}{(t-x)^{1-\alpha}} + \frac{f(0)}{t^{1-\alpha}}\right]. \tag{3}$$

The Abel transform is used in calculating the radial mass distribution of galaxies and inverting planetary radio occultation data to obtain atmospheric information.

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 875–876, 1985.

Binney, J. and Tremaine, S. *Galactic Dynamics.* Princeton, NJ: Princeton University Press, p. 651, 1987.

Bracewell, R. *The Fourier Transform and Its Applications.* New York: McGraw-Hill, pp. 262–266, 1965.

Abel's Uniform Convergence Test

Let $\{u_n(x)\}$ be a Sequence of functions. If

1. $u_n(x)$ can be written $u_n(x) = a_n f_n(x)$,
2. $\sum a_n$ is Convergent,
3. $f_n(x)$ is a Monotonic Decreasing Sequence (i.e., $f_{n+1}(x) \leq f_n(x)$) for all n, and
4. $f_n(x)$ is Bounded in some region (i.e., $0 \leq f_n(x) \leq M$ for all $x \in [a,b]$)

then, for all $x \in [a,b]$, the Series $\sum u_n(x)$ Converges Uniformly.

see also Convergence Tests

References

Bromwich, T. J. I'a and MacRobert, T. M. *An Introduction to the Theory of Infinite Series, 3rd ed.* New York: Chelsea, p. 59, 1991.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, p. 17, 1990.

Abelian Variety

An Abelian variety is an algebraic Group which is a complete Algebraic Variety. An Abelian variety of Dimension 1 is an Elliptic Curve.

see also Albanese Variety

References

Murty, V. K. *Introduction to Abelian Varieties.* Providence, RI: Amer. Math. Soc., 1993.

Abhyankar's Conjecture

For a Finite Group G, let $p(G)$ be the Subgroup generated by all the Sylow p-Subgroups of G. If X is a projective curve in characteristic $p > 0$, and if $x_0, \ldots, x_t$ are points of X (for $t > 0$), then a Necessary and Sufficient condition that G occur as the Galois Group of a finite covering Y of X, branched only at the points $x_0, \ldots, x_t$, is that the Quotient Group $G/p(G)$ has $2g + t$ generators.

Raynaud (1994) solved the Abhyankar problem in the crucial case of the affine line (i.e., the projective line with a point deleted), and Harbater (1994) proved the full Abhyankar conjecture by building upon this special solution.

see also Finite Group, Galois Group, Quotient Group, Sylow p-Subgroup

References

Abhyankar, S. "Coverings of Algebraic Curves." *Amer. J. Math.* **79**, 825–856, 1957.

American Mathematical Society. "Notices of the AMS, April 1995, 1995 Frank Nelson Cole Prize in Algebra." `http://www.ams.org/notices/199504/prize-cole.html`.

Harbater, D. "Abhyankar's Conjecture on Galois Groups Over Curves." *Invent. Math.* **117**, 1–25, 1994.

Raynaud, M. "Revêtements de la droite affine en caractéristique $p > 0$ et conjecture d'Abhyankar." *Invent. Math.* **116**, 425–462, 1994.

Ablowitz-Ramani-Segur Conjecture

The Ablowitz-Ramani-Segur conjecture states that a nonlinear PARTIAL DIFFERENTIAL EQUATION is solvable by the INVERSE SCATTERING METHOD only if every nonlinear ORDINARY DIFFERENTIAL EQUATION obtained by exact reduction has the PAINLEVÉ PROPERTY.

see also INVERSE SCATTERING METHOD

References

Tabor, M. *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, p. 351, 1989.

Abscissa

The x- (horizontal) axis of a GRAPH.

see also AXIS, ORDINATE, REAL LINE, x-AXIS, y-AXIS, z-AXIS

Absolute Convergence

A SERIES $\sum_n u_n$ is said to CONVERGE absolutely if the SERIES $\sum_n |u_n|$ CONVERGES, where $|u_n|$ denotes the ABSOLUTE VALUE. If a SERIES is absolutely convergent, then the sum is independent of the order in which terms are summed. Furthermore, if the SERIES is multiplied by another absolutely convergent series, the product series will also converge absolutely.

see also CONDITIONAL CONVERGENCE, CONVERGENT SERIES, RIEMANN SERIES THEOREM

References

Bromwich, T. J. I'a and MacRobert, T. M. "Absolute Convergence." Ch. 4 in *An Introduction to the Theory of Infinite Series, 3rd ed.* New York: Chelsea, pp. 69–77, 1991.

Absolute Deviation

Let $\bar{u}$ denote the MEAN of a SET of quantities u_i, then the absolute deviation is defined by

$$\Delta u_i \equiv |u_i - \bar{u}|.$$

see also DEVIATION, MEAN DEVIATION, SIGNED DEVIATION, STANDARD DEVIATION

Absolute Error

The DIFFERENCE between the measured or inferred value of a quantity x_0 and its actual value x, given by

$$\Delta x \equiv x_0 - x$$

(sometimes with the ABSOLUTE VALUE taken) is called the absolute error. The absolute error of the SUM or DIFFERENCE of a number of quantities is less than or equal to the SUM of their absolute errors.

see also ERROR PROPAGATION, PERCENTAGE ERROR, RELATIVE ERROR

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 14, 1972.

Absolute Geometry

GEOMETRY which depends only on the first four of EUCLID'S POSTULATES and not on the PARALLEL POSTULATE. Euclid himself used only the first four postulates for the first 28 propositions of the *Elements*, but was forced to invoke the PARALLEL POSTULATE on the 29th.

see also AFFINE GEOMETRY, *Elements*, EUCLID'S POSTULATES, GEOMETRY, ORDERED GEOMETRY, PARALLEL POSTULATE

References

Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, pp. 90–91, 1989.

Absolute Pseudoprime

see CARMICHAEL NUMBER

Absolute Square

Also known as the squared NORM. The absolute square of a COMPLEX NUMBER z is written $|z|^2$ and is defined as

$$|z|^2 \equiv zz^*, \tag{1}$$

where z^* denotes the COMPLEX CONJUGATE of z. For a REAL NUMBER, (1) simplifies to

$$|z|^2 = z^2. \tag{2}$$

If the COMPLEX NUMBER is written $z = x + iy$, then the absolute square can be written

$$|x + iy|^2 = x^2 + y^2. \tag{3}$$

An important identity involving the absolute square is given by

$$\begin{aligned} |a \pm be^{-i\delta}|^2 &= (a \pm be^{-i\delta})(a \pm be^{i\delta}) \\ &= a^2 + b^2 \pm ab(e^{i\delta} + e^{-i\delta}) \\ &= a^2 + b^2 \pm 2ab\cos\delta. \end{aligned} \tag{4}$$

If $a = 1$, then (4) becomes

$$\begin{aligned}|1 \pm be^{-i\delta}|^2 &= 1 + b^2 \pm 2b\cos\delta \\ &= 1 + b^2 \pm 2b[1 - 2\sin^2(\tfrac{1}{2}\delta)] \\ &= 1 \pm 2b + b^2 \mp 4b\sin^2(\tfrac{1}{2}\delta) \\ &= (1 \pm b)^2 \mp 4b\sin^2(\tfrac{1}{2}\delta). \qquad (5)\end{aligned}$$

If $a = 1$, and $b = 1$, then

$$|1 - e^{-i\delta}|^2 = (1-1)^2 + 4 \cdot 1\sin^2(\tfrac{1}{2}\delta) = 4\sin^2(\tfrac{1}{2}\delta). \qquad (6)$$

Finally,

$$\begin{aligned}|e^{i\phi_1} + e^{i\phi_2}|^2 &= (e^{i\phi_1} + e^{i\phi_2})(e^{-i\phi_1} + e^{-i\phi_2}) \\ &= 2 + e^{i(\phi_2 - \phi_1)} + e^{-i(\phi_2 - \phi_1)} \\ &= 2 + 2\cos(\phi_2 - \phi_1) = 2[1 + \cos(\phi_2 - \phi_1)] \\ &= 4\cos^2(\phi_2 - \phi_1). \qquad (7)\end{aligned}$$

Absolute Value

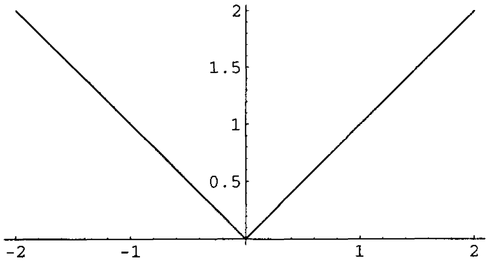

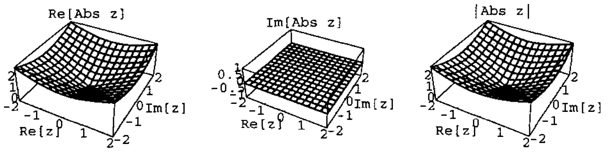

The absolute value of a REAL NUMBER x is denoted $|x|$ and given by

$$|x| = x\,\mathrm{sgn}(x) = \begin{cases} -x & \text{for } x \le 0 \\ x & \text{for } x \ge 0, \end{cases}$$

where SGN is the sign function.

The same notation is used to denote the MODULUS of a COMPLEX NUMBER $z = x + iy$, $|z| \equiv \sqrt{x^2 + y^2}$, a p-ADIC absolute value, or a general VALUATION. The NORM of a VECTOR $\mathbf{x}$ is also denoted $|\mathbf{x}|$, although $||\mathbf{x}||$ is more commonly used.

Other NOTATIONS similar to the absolute value are the FLOOR FUNCTION $\lfloor x \rfloor$, NINT function $[x]$, and CEILING FUNCTION $\lceil x \rceil$.

see also ABSOLUTE SQUARE, CEILING FUNCTION, FLOOR FUNCTION, MODULUS (COMPLEX NUMBER), NINT, SGN, TRIANGLE FUNCTION, VALUATION

Absolutely Continuous

Let μ be a POSITIVE MEASURE on a SIGMA ALGEBRA M and let λ be an arbitrary (real or complex) MEASURE on M. Then λ is absolutely continuous with respect to μ, written $\lambda \ll \mu$, if $\lambda(E) = 0$ for every $E \in M$ for which $\mu(E) = 0$.

see also CONCENTRATED, MUTUALLY SINGULAR

References
Rudin, W. *Functional Analysis.* New York: McGraw-Hill, pp. 121–125, 1991.

Absorption Law

The law appearing in the definition of a BOOLEAN ALGEBRA which states

$$a \wedge (a \vee b) = a \vee (a \wedge b) = a$$

for binary operators $\vee$ and $\wedge$ (which most commonly are logical OR and logical AND).

see also BOOLEAN ALGEBRA, LATTICE

References
Birkhoff, G. and Mac Lane, S. *A Survey of Modern Algebra, 3rd ed.* New York: Macmillian, p. 317, 1965.

Abstraction Operator

see LAMBDA CALCULUS

Abundance

The abundance of a number n is the quantity

$$A(n) \equiv \sigma(n) - 2n,$$

where $\sigma(n)$ is the DIVISOR FUNCTION. Kravitz has conjectured that no numbers exist whose abundance is an ODD SQUARE (Guy 1994).

The following table lists special classifications given to a number n based on the value of $A(n)$.

$A(n)$	Number
< 0	deficient number
-1	almost perfect number
0	perfect number
1	quasiperfect number
> 0	abundant number

see also DEFICIENCY

References
Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 45–46, 1994.

Abundant Number

An abundant number is an INTEGER n which is not a PERFECT NUMBER and for which

$$s(n) \equiv \sigma(n) - n > n, \tag{1}$$

where $\sigma(n)$ is the DIVISOR FUNCTION. The quantity $\sigma(n) - 2n$ is sometimes called the ABUNDANCE. The first few abundant numbers are 12, 18, 20, 24, 30, 36, ... (Sloane's A005101). Abundant numbers are sometimes called EXCESSIVE NUMBERS.

There are only 21 abundant numbers less than 100, and they are all EVEN. The first ODD abundant number is

$$945 = 3^3 \cdot 7 \cdot 5. \tag{2}$$

That 945 is abundant can be seen by computing

$$s(945) = 975 > 945. \tag{3}$$

Any multiple of a PERFECT NUMBER or an abundant number is also abundant. Every number greater than 20161 can be expressed as a sum of two abundant numbers.

Define the density function

$$A(x) \equiv \lim_{n\to\infty} \frac{|\{n : \sigma(n) \geq xn\}|}{n} \tag{4}$$

for a POSITIVE REAL NUMBER x, then Davenport (1933) proved that $A(x)$ exists and is continuous for all x, and Erdős (1934) gave a simplified proof (Finch). Wall (1971) and Wall *et al.* (1977) showed that

$$0.2441 < A(2) < 0.2909, \tag{5}$$

and Deléglise showed that

$$0.2474 < A(2) < 0.2480. \tag{6}$$

A number which is abundant but for which all its PROPER DIVISORS are DEFICIENT is called a PRIMITIVE ABUNDANT NUMBER (Guy 1994, p. 46).

see also ALIQUOT SEQUENCE, DEFICIENT NUMBER, HIGHLY ABUNDANT NUMBER, MULTIAMICABLE NUMBERS, PERFECT NUMBER, PRACTICAL NUMBER, PRIMITIVE ABUNDANT NUMBER, WEIRD NUMBER

References

Deléglise, M. "Encadrement de la densité des nombres abondants." Submitted.

Dickson, L. E. *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, pp. 3–33, 1952.

Erdős, P. "On the Density of the Abundant Numbers." *J. London Math. Soc.* **9**, 278–282, 1934.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/abund/abund.html`.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 45–46, 1994.

Singh, S. *Fermat's Enigma: The Epic Quest to Solve the World's Greatest Mathematical Problem.* New York: Walker, pp. 11 and 13, 1997.

Sloane, N. J. A. Sequence A005101/M4825 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Wall, C. R. "Density Bounds for the Sum of Divisors Function." In *The Theory of Arithmetic Functions* (Ed. A. A. Gioia and D. L. Goldsmith). New York: Springer-Verlag, pp. 283–287, 1971.

Wall, C. R.; Crews, P. L.; and Johnson, D. B. "Density Bounds for the Sum of Divisors Function." *Math. Comput.* **26**, 773–777, 1972.

Wall, C. R.; Crews, P. L.; and Johnson, D. B. "Density Bounds for the Sum of Divisors Function." *Math. Comput.* **31**, 616, 1977.

Acceleration

Let a particle travel a distance $s(t)$ as a function of time t (here, s can be thought of as the ARC LENGTH of the curve traced out by the particle). The SPEED (the SCALAR NORM of the VECTOR VELOCITY) is then given by

$$\frac{ds}{dt} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}. \tag{1}$$

The acceleration is defined as the time DERIVATIVE of the VELOCITY, so the SCALAR acceleration is given by

$$\begin{aligned} a &\equiv \frac{dv}{dt} && (2) \\ &= \frac{d^2s}{dt^2} && (3) \\ &= \frac{\frac{dx}{dt}\frac{d^2x}{dt^2} + \frac{dy}{dt}\frac{d^2y}{dt^2} + \frac{dz}{dt}\frac{d^2z}{dt^2}}{\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}} && (4) \\ &= \frac{dx}{ds}\frac{d^2x}{dt^2} + \frac{dy}{ds}\frac{d^2y}{dt^2} + \frac{dz}{ds}\frac{d^2z}{dt^2} && (5) \\ &= \frac{d\mathbf{r}}{ds} \cdot \frac{d^2\mathbf{r}}{dt^2}. && (6) \end{aligned}$$

The VECTOR acceleration is given by

$$\mathbf{a} \equiv \frac{d\mathbf{v}}{dt} = \frac{d^2\mathbf{r}}{dt^2} = \frac{d^2s}{dt^2}\hat{\mathbf{T}} + \kappa\left(\frac{ds}{dt}\right)^2 \hat{\mathbf{N}}, \tag{7}$$

where $\hat{\mathbf{T}}$ is the UNIT TANGENT VECTOR, κ the CURVATURE, s the ARC LENGTH, and $\hat{\mathbf{N}}$ the UNIT NORMAL VECTOR.

Let a particle move along a straight LINE so that the positions at times t_1, t_2, and t_3 are s_1, s_2, and s_3, respectively. Then the particle is uniformly accelerated with acceleration a IFF

$$a \equiv 2\left[\frac{(s_2 - s_3)t_1 + (s_3 - s_1)t_2 + (s_1 - s_2)t_3}{(t_1 - t_2)(t_2 - t_3)(t_3 - t_1)}\right] \tag{8}$$

is a constant (Klamkin 1995, 1996).

Consider the measurement of acceleration in a rotating reference frame. Apply the ROTATION OPERATOR

$$\tilde{R} \equiv \left(\frac{d}{dt}\right)_{\text{body}} + \boldsymbol{\omega}\times \tag{9}$$

twice to the RADIUS VECTOR $\mathbf{r}$ and suppress the *body* notation,

$$\begin{aligned}\mathbf{a}_{\text{space}} &= \tilde{R}^2\mathbf{r} = \left(\frac{d}{dt} + \boldsymbol{\omega}\times\right)^2 \mathbf{r}\\ &= \left(\frac{d}{dt} + \boldsymbol{\omega}\times\right)\left(\frac{d\mathbf{r}}{dt} + \boldsymbol{\omega}\times\mathbf{r}\right)\\ &= \frac{d^2\mathbf{r}}{dt^2} + \frac{d}{dt}(\boldsymbol{\omega}\times\mathbf{r}) + \boldsymbol{\omega}\times\frac{d\mathbf{r}}{dt} + \boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r})\\ &= \frac{d^2\mathbf{r}}{dt^2} + \boldsymbol{\omega}\times\frac{d\mathbf{r}}{dt} + \mathbf{r}\times\frac{d\boldsymbol{\omega}}{dt} + \boldsymbol{\omega}\times\frac{d\mathbf{r}}{dt}\\ &\quad + \boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r}).\end{aligned} \tag{10}$$

Grouping terms and using the definitions of the VELOCITY $\mathbf{v} \equiv d\mathbf{r}/dt$ and ANGULAR VELOCITY $\boldsymbol{\alpha} \equiv d\boldsymbol{\omega}/dt$ give the expression

$$\mathbf{a}_{\text{space}} = \frac{d^2\mathbf{r}}{dt^2} + 2\boldsymbol{\omega}\times\mathbf{v} + \boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r}) + \mathbf{r}\times\boldsymbol{\alpha}. \tag{11}$$

Now, we can identify the expression as consisting of three terms

$$\mathbf{a}_{\text{body}} \equiv \frac{d^2\mathbf{r}}{dt^2}, \tag{12}$$

$$\mathbf{a}_{\text{Coriolis}} \equiv 2\boldsymbol{\omega}\times\mathbf{v}, \tag{13}$$

$$\mathbf{a}_{\text{centrifugal}} \equiv \boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r}), \tag{14}$$

a "body" acceleration, centrifugal acceleration, and Coriolis acceleration. Using these definitions finally gives

$$\mathbf{a}_{\text{space}} = \mathbf{a}_{\text{body}} + \mathbf{a}_{\text{Coriolis}} + \mathbf{a}_{\text{centrifugal}} + \mathbf{r}\times\boldsymbol{\alpha}, \tag{15}$$

where the fourth term will vanish in a uniformly rotating frame of reference (i.e., $\boldsymbol{\alpha} = \mathbf{0}$). The centrifugal acceleration is familiar to riders of merry-go-rounds, and the Coriolis acceleration is responsible for the motions of hurricanes on Earth and necessitates large trajectory corrections for intercontinental ballistic missiles.

see also ANGULAR ACCELERATION, ARC LENGTH, JERK, VELOCITY

References

Klamkin, M. S. "Problem 1481." *Math. Mag.* **68**, 307, 1995.

Klamkin, M. S. "A Characteristic of Constant Acceleration." Solution to Problem 1481. *Math. Mag.* **69**, 308, 1996.

Accidental Cancellation

see ANOMALOUS CANCELLATION

Accumulation Point

An accumulation point is a POINT which is the limit of a SEQUENCE, also called a LIMIT POINT. For some MAPS, periodic orbits give way to CHAOTIC ones beyond a point known as the accumulation point.

see also CHAOS, LOGISTIC MAP, MODE LOCKING, PERIOD DOUBLING

Achilles and the Tortoise Paradox

see ZENO'S PARADOXES

Ackermann Function

The Ackermann function is the simplest example of a well-defined TOTAL FUNCTION which is COMPUTABLE but not PRIMITIVE RECURSIVE, providing a counterexample to the belief in the early 1900s that every COMPUTABLE FUNCTION was also PRIMITIVE RECURSIVE (Dötzel 1991). It grows faster than an exponential function, or even a multiple exponential function. The Ackermann function $A(x,y)$ is defined by

$$A(x,y) \equiv \begin{cases} y+1 & \text{if } x = 0\\ A(x-1,1) & \text{if } y = 0\\ A(x-1,A(x,y-1)) & \text{otherwise.}\end{cases} \tag{1}$$

Special values for INTEGER x include

$$A(0,y) = y+1 \tag{2}$$

$$A(1,y) = y+2 \tag{3}$$

$$A(2,y) = 2y+3 \tag{4}$$

$$A(3,y) = 2^{y+3} - 3 \tag{5}$$

$$A(4,y) = \underbrace{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}_{y+3} - 3. \tag{6}$$

Expressions of the latter form are sometimes called POWER TOWERS. $A(0,y)$ follows trivially from the definition. $A(1,y)$ can be derived as follows,

$$\begin{aligned}A(1,y) &= A(0,A(1,y-1)) = A(1,y-1)+1\\ &= A(0,A(1,y-2))+1 = A(1,y-2)+2\\ &= \ldots = A(1,0)+y = A(0,1)+y = y+2.\end{aligned} \tag{7}$$

$A(2,y)$ has a similar derivation,

$$\begin{aligned}A(2,y) &= A(1,A(2,y-1)) = A(2,y-1)+2\\ &= A(1,A(2,y-2))+2 = A(2,y-2)+4 = \ldots\\ &= A(2,0)+2y = A(1,1)+2y = 2y+3.\end{aligned} \tag{8}$$

Buck (1963) defines a related function using the same fundamental RECURRENCE RELATION (with arguments flipped from Buck's convention)

$$F(x,y) = F(x-1,F(x,y-1)), \tag{9}$$

but with the slightly different boundary values

$$F(0,y) = y + 1 \tag{10}$$

$$F(1,0) = 2 \tag{11}$$

$$F(2,0) = 0 \tag{12}$$

$$F(x,0) = 1 \qquad \text{for } x = 3, 4, \ldots. \tag{13}$$

Buck's recurrence gives

$$F(1,y) = 2 + y \tag{14}$$

$$F(2,y) = 2y \tag{15}$$

$$F(3,y) = 2^y \tag{16}$$

$$F(4,y) = \underbrace{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}_{y}. \tag{17}$$

Taking $F(4,n)$ gives the sequence 1, 2, 4, 16, 65536, 2^{65536}, ... Defining $\psi(x) = F(x,x)$ for $x = 0, 1, \ldots$ then gives 1, 3, 4, 8, 65536, $\underbrace{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}_{m}$, ... (Sloane's A001695), where $m = \underbrace{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}_{65536}$, a truly huge number!

see also ACKERMANN NUMBER, COMPUTABLE FUNCTION, GOODSTEIN SEQUENCE, POWER TOWER, PRIMITIVE RECURSIVE FUNCTION, TAK FUNCTION, TOTAL FUNCTION

References

Buck, R. C. "Mathematical Induction and Recursive Definitions." *Amer. Math. Monthly* **70**, 128–135, 1963.

Dötzel, G. "A Function to End All Functions." *Algorithm: Recreational Programming* **2.4**, 16–17, 1991.

Kleene, S. C. *Introduction to Metamathematics.* New York: Elsevier, 1971.

Péter, R. *Rekursive Funktionen.* Budapest: Akad. Kiado, 1951.

Reingold, E. H. and Shen, X. "More Nearly Optimal Algorithms for Unbounded Searching, Part I: The Finite Case." *SIAM J. Comput.* **20**, 156–183, 1991.

Rose, H. E. *Subrecursion, Functions, and Hierarchies.* New York: Clarendon Press, 1988.

Sloane, N. J. A. Sequence A001695/M2352 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Smith, H. J. "Ackermann's Function." `http://www.netcom.com/~hjsmith/Ackerman.html`.

Spencer, J. "Large Numbers and Unprovable Theorems." *Amer. Math. Monthly* **90**, 669–675, 1983.

Tarjan, R. E. *Data Structures and Network Algorithms.* Philadelphia PA: SIAM, 1983.

Vardi, I. *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 11, 227, and 232, 1991.

Ackermann Number

A number of the form $\underbrace{n \uparrow \cdots \uparrow n}_{n}$, where ARROW NOTATION has been used. The first few Ackermann numbers are $1 \uparrow 1 = 1$, $2 \uparrow\uparrow 2 = 4$, and $3 \uparrow\uparrow\uparrow 3 = \underbrace{3^{3^{\cdot^{\cdot^{\cdot^{3}}}}}}_{7,625,597,484,987}$.

see also ACKERMANN FUNCTION, ARROW NOTATION, POWER TOWER

References

Ackermann, W. "Zum hilbertschen Aufbau der reellen Zahlen." *Math. Ann.* **99**, 118–133, 1928.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 60–61, 1996.

Crandall, R. E. "The Challenge of Large Numbers." *Sci. Amer.* **276**, 74–79, Feb. 1997.

Vardi, I. *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 11, 227, and 232, 1991.

Acnode

Another name for an ISOLATED POINT.

see also CRUNODE, SPINODE, TACNODE

Acoptic Polyhedron

A term invented by B. Grünbaum in an attempt to promote concrete and precise POLYHEDRON terminology. The word "coptic" derives from the Greek for "to cut," and acoptic polyhedra are defined as POLYHEDRA for which the FACES do not intersect (cut) themselves, making them 2-MANIFOLDS.

see also HONEYCOMB, NOLID, POLYHEDRON, SPONGE

Action

Let $M(X)$ denote the GROUP of all invertible MAPS $X \to X$ and let G be any GROUP. A HOMOMORPHISM $\theta : G \to M(X)$ is called an action of G on X. Therefore, θ satisfies

1. For each $g \in G$, $\theta(g)$ is a MAP $X \to X : x \mapsto \theta(g)x$,
2. $\theta(gh)x = \theta(g)(\theta(h)x)$,
3. $\theta(e)x = x$, where e is the group identity in G,
4. $\theta(g^{-1})x = \theta(g)^{-1}x$.

see also CASCADE, FLOW, SEMIFLOW

Acute Angle

An ANGLE of less than $\pi/2$ RADIANS (90°) is called an acute angle.

see also ANGLE, OBTUSE ANGLE, RIGHT ANGLE, STRAIGHT ANGLE

Acute Triangle

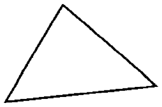

A TRIANGLE in which all three ANGLES are ACUTE ANGLES. A TRIANGLE which is neither acute nor a RIGHT TRIANGLE (i.e., it has an OBTUSE ANGLE) is called an OBTUSE TRIANGLE. A SQUARE can be dissected into as few as 8 acute triangles.

see also OBTUSE TRIANGLE, RIGHT TRIANGLE

Adams-Bashforth-Moulton Method

see ADAMS' METHOD

Adams' Method

Adams' method is a numerical METHOD for solving linear FIRST-ORDER ORDINARY DIFFERENTIAL EQUATIONS of the form

$$\frac{dy}{dx} = f(x,y). \qquad (1)$$

Let

$$h = x_{n+1} - x_n \qquad (2)$$

be the step interval, and consider the MACLAURIN SERIES of y about x_n,

$$y_{n+1} = y_n + \left(\frac{dy}{dx}\right)_n (x - x_n) + \frac{1}{2}\left(\frac{d^2y}{dx^2}\right)_n (x - x_n)^2 + \ldots \qquad (3)$$

$$\left(\frac{dy}{dx}\right)_{n+1} = \left(\frac{dy}{dx}\right)_n + \left(\frac{d^2y}{dx^2}\right)_n (x - x_n)^2 + \ldots. \qquad (4)$$

Here, the DERIVATIVES of y are given by the BACKWARD DIFFERENCES

$$q_n \equiv \left(\frac{dy}{dx}\right)_n = \frac{\Delta y_n}{x_{n+1} - x_n} = \frac{y_{n+1} - y_n}{h} \qquad (5)$$

$$\nabla q_n \equiv \left(\frac{d^2y}{dx^2}\right)_n = q_n - q_{n-1} \qquad (6)$$

$$\nabla^2 q_n \equiv \left(\frac{d^3y}{dx^3}\right)_n = \nabla q_n - \nabla q_{n-1}, \qquad (7)$$

etc. Note that by (1), q_n is just the value of $f(x_n, y_n)$.

For first-order interpolation, the method proceeds by iterating the expression

$$y_{n+1} = y_n + q_n h \qquad (8)$$

where $q_n \equiv f(x_n, y_n)$. The method can then be extended to arbitrary order using the finite difference integration formula from Beyer (1987)

$$\int_0^1 f_p\,dp = (1 + \tfrac{1}{2}\nabla + \tfrac{5}{12}\nabla^2 + \tfrac{3}{8}\nabla^3 + \tfrac{251}{720}\nabla^4 + \tfrac{95}{288}\nabla^5 + \tfrac{19087}{60480}\nabla^6 + \ldots)f_p \qquad (9)$$

to obtain

$$y_{n+1} - y_n = h(q_n + \tfrac{1}{2}\nabla q_{n-1} + \tfrac{5}{12}\nabla^2 q_{n-2} + \tfrac{3}{8}\nabla^3 q_{n-3} + \tfrac{251}{720}\nabla^4 q_{n-4} + \tfrac{95}{288}\nabla^5 q_{n-5} + \ldots). \qquad (10)$$

Note that von Kármán and Biot (1940) confusingly use the symbol normally used for FORWARD DIFFERENCES Δ to denote BACKWARD DIFFERENCES ∇.

see also GILL'S METHOD, MILNE'S METHOD, PREDICTOR-CORRECTOR METHODS, RUNGE-KUTTA METHOD

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 896, 1972.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 455, 1987.

Kármán, T. von and Biot, M. A. *Mathematical Methods in Engineering: An Introduction to the Mathematical Treatment of Engineering Problems.* New York: McGraw-Hill, pp. 14–20, 1940.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 741, 1992.

Addend

A quantity to be ADDED to another, also called a SUMMAND. For example, in the expression $a+b+c$, a, b, and c are all addends. The first of several addends, or "the one to which the others are added" (a in the previous example), is sometimes called the AUGEND.

see also ADDITION, AUGEND, PLUS, RADICAND

Addition

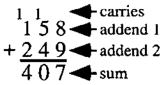

The combining of two or more quantities using the PLUS operator. The individual numbers being combined are called ADDENDS, and the total is called the SUM. The first of several ADDENDS, or "the one to which the others are added," is sometimes called the AUGEND. The opposite of addition is SUBTRACTION.

While the usual form of adding two n-digit INTEGERS (which consists of summing over the columns right to left and "CARRYING" a 1 to the next column if the sum exceeds 9) requires n operations (plus carries), two n-digit INTEGERS can be added in about $2\lg n$ steps by n processors using carry-lookahead addition (McGeoch 1993). Here, $\lg x$ is the LG function, the LOGARITHM to the base 2.

see also ADDEND, AMENABLE NUMBER, AUGEND, CARRY, DIFFERENCE, DIVISION, MULTIPLICATION, PLUS, SUBTRACTION, SUM

References

McGeoch, C. C. "Parallel Addition." *Amer. Math. Monthly* **100**, 867–871, 1993.

Addition Chain

An addition chain for a number n is a SEQUENCE $1 = a_0 < a_1 < \ldots < a_r = n$, such that each member after a_0 is the SUM of the two earlier (not necessarily distinct) ones. The number r is called the length of the addition chain. For example,

$$1, 1+1=2, 2+2=4, 4+2=6, 6+2=8, 8+6=14$$

is an addition chain for 14 of length $r = 5$ (Guy 1994).

see also BRAUER CHAIN, HANSEN CHAIN, SCHOLZ CONJECTURE

References

Guy, R. K. "Addition Chains. Brauer Chains. Hansen Chains." §C6 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 111–113, 1994.

Addition-Multiplication Magic Square

46	81	117	102	15	76	200	203
19	60	232	175	54	69	153	78
216	161	17	52	171	90	58	75
135	114	50	87	184	189	13	68
150	261	45	38	91	136	92	27
119	104	108	23	174	225	57	30
116	25	133	120	51	26	162	207
39	34	138	243	100	29	105	152

200	87	95	42	99	1	46	108	170
14	44	10	184	81	85	150	261	19
138	243	17	50	116	190	56	33	5
57	125	232	9	7	66	68	230	54
4	70	22	51	115	216	171	25	174
153	23	162	76	250	58	3	35	88
145	152	75	11	6	63	270	34	92
110	2	28	135	136	69	29	114	225
27	102	207	290	38	100	55	8	21

200	87	95	42	99	1	46	108	170
14	44	10	184	81	85	150	261	19
138	243	17	50	116	190	56	33	5
57	125	232	9	7	66	68	230	54
4	70	22	51	115	216	171	25	174
153	23	162	76	250	58	3	35	88
145	152	75	11	6	63	270	34	92
110	2	28	135	136	69	29	114	225
27	102	207	290	38	100	55	8	21

A square which is simultaneously a MAGIC SQUARE and MULTIPLICATION MAGIC SQUARE. The three squares shown above (the top square has order eight and the bottom two have order nine) have addition MAGIC CONSTANTS (840, 848, 1200) and multiplicative magic constants (2,058,068,231,856,000; 5,804,807,833,440,000; 1,619,541,385,529,760,000), respectively (Hunter and Madachy 1975, Madachy 1979).

see also MAGIC SQUARE

References

Hunter, J. A. H. and Madachy, J. S. "Mystic Arrays." Ch. 3 in *Mathematical Diversions.* New York: Dover, pp. 30–31, 1975.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 89–91, 1979.

Additive Persistence

Consider the process of taking a number, adding its DIGITS, then adding the DIGITS of number derived from it, etc., until the remaining number has only one DIGIT. The number of additions required to obtain a single DIGIT from a number n is called the additive persistence of n, and the DIGIT obtained is called the DIGITAL ROOT of n.

For example, the sequence obtained from the starting number 9876 is (9876, 30, 3), so 9876 has an additive persistence of 2 and a DIGITAL ROOT of 3. The additive persistences of the first few positive integers are 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, ... (Sloane's A031286). The smallest numbers of additive persistence n for $n = 0, 1, \ldots$ are 0, 10, 19, 199, 19999999999999999999999, ... (Sloane's A006050). There is no number $< 10^{50}$ with additive persistence greater than 11.

It is conjectured that the maximum number lacking the DIGIT 1 with persistence 11 is

$$77777733332222222222222222222$$

There is a stronger conjecture that there is a maximum number lacking the DIGIT 1 for each persistence ≥ 2.

The maximum additive persistence in base 2 is 1. It is conjectured that all powers of $2 > 2^{15}$ contain a 0 in base 3, which would imply that the maximum persistence in base 3 is 3 (Guy, 1994).

see also DIGITADITION, DIGITAL ROOT, MULTIPLICATIVE PERSISTENCE, NARCISSISTIC NUMBER, RECURRING DIGITAL INVARIANT

References

Guy, R. K. "The Persistence of a Number." §F25 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 262–263, 1994.

Hinden, H. J. "The Additive Persistence of a Number." *J. Recr. Math.* **7**, 134–135, 1974.

Sloane, N. J. A. Sequences A031286 and A006050/M4683 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sloane, N. J. A. "The Persistence of a Number." *J. Recr. Math.* **6**, 97–98, 1973.

Adéle

An element of an ADÉLE GROUP, sometimes called a REPARTITION in older literature. Adéles arise in both NUMBER FIELDS and FUNCTION FIELDS. The adéles of a NUMBER FIELD are the additive SUBGROUPS of all elements in $\prod k_\nu$, where ν is the PLACE, whose ABSOLUTE VALUE is < 1 at all but finitely many νs.

Let F be a FUNCTION FIELD of algebraic functions of one variable. Then a MAP r which assigns to every PLACE P of F an element $r(P)$ of F such that there are only a finite number of PLACES P for which $\nu_P(r(P)) < 0$.

see also IDELE

References

Chevalley, C. C. *Introduction to the Theory of Algebraic Functions of One Variable.* Providence, RI: Amer. Math. Soc., p. 25, 1951.

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Adéle Group

The restricted topological DIRECT PRODUCT of the GROUP G_{k_ν} with distinct invariant open subgroups G_{0_ν}.

References

Weil, A. *Adéles and Algebraic Groups.* Princeton, NJ: Princeton University Press, 1961.

Adem Relations

Relations in the definition of a STEENROD ALGEBRA which state that, for $i < 2j$,

$$Sq^i \circ Sq^j(x) = \sum_{k=0}^{\lfloor i \rfloor} \binom{j-k-1}{i-2k} Sq^{i+j-k} \circ Sq^k(x),$$

where $f \circ g$ denotes function COMPOSITION and $\lfloor i \rfloor$ is the FLOOR FUNCTION.

see also STEENROD ALGEBRA

Adequate Knot

A class of KNOTS containing the class of ALTERNATING KNOTS. Let $c(K)$ be the CROSSING NUMBER. Then for KNOT SUM $K_1 \# K_2$ which is an adequate knot,

$$c(K_1 \# K_2) = c(K_1) + c(K_2).$$

This relationship is postulated to hold true for all KNOTS.

see also ALTERNATING KNOT, CROSSING NUMBER (LINK)

Adiabatic Invariant

A property of motion which is conserved to exponential accuracy in the small parameter representing the typical rate of change of the gross properties of the body.

see also ALGEBRAIC INVARIANT, LYAPUNOV CHARACTERISTIC NUMBER

Adjacency Matrix

The adjacency matrix of a simple GRAPH is a MATRIX with rows and columns labelled by VERTICES, with a 1 or 0 in position (v_i, v_j) according to whether v_i and v_j are ADJACENT or not.

see also INCIDENCE MATRIX

References

Chartrand, G. *Introductory Graph Theory.* New York: Dover, p. 218, 1985.

Adjacency Relation

The SET E of EDGES of a GRAPH (V, E), being a set of unordered pairs of elements of V, constitutes a RELATION on V. Formally, an adjacency relation is any RELATION which is IRREFLEXIVE and SYMMETRIC.

see also IRREFLEXIVE, RELATION, SYMMETRIC

Adjacent Fraction

Two FRACTIONS are said to be adjacent if their difference has a unit NUMERATOR. For example, 1/3 and 1/4 are adjacent since $1/3 - 1/4 = 1/12$, but 1/2 and 1/5 are not since $1/2 - 1/5 = 3/10$. Adjacent fractions can be adjacent in a FAREY SEQUENCE.

see also FAREY SEQUENCE, FORD CIRCLE, FRACTION, NUMERATOR

References

Pickover, C. A. *Keys to Infinity.* New York: W. H. Freeman, p. 119, 1995.

Adjacent Value

The value nearest to but still inside an inner FENCE.

References

Tukey, J. W. *Explanatory Data Analysis.* Reading, MA: Addison-Wesley, p. 667, 1977.

Adjacent Vertices

In a GRAPH G, two VERTICES are adjacent if they are joined by an EDGE.

Adjoint Curve

A curve which has at least multiplicity $r_i - 1$ at each point where a given curve (having only ordinary singular points and cusps) has a multiplicity r_i is called the adjoint to the given curve. When the adjoint curve is of order $n - 3$, it is called a special adjoint curve.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 30, 1959.

Adjoint Matrix

The adjoint matrix, sometimes also called the ADJUGATE MATRIX, is defined by

$$\mathsf{A}^\dagger \equiv (\mathsf{A}^{\mathrm{T}})^*, \tag{1}$$

where the ADJOINT OPERATOR is denoted † and $^{\mathrm{T}}$ denotes the TRANSPOSE. If a MATRIX is SELF-ADJOINT, it is said to be HERMITIAN. The adjoint matrix of a MATRIX product is given by

$$(ab)^\dagger{}_{ij} \equiv [(ab)^{\mathrm{T}}]^*_{ij}\,. \tag{2}$$

Using the property of transpose products that

$$\begin{aligned}[(ab)^{\mathrm{T}}]^*_{ij} &= (b^{\mathrm{T}} a^{\mathrm{T}})^*_{ij} = (b^{\mathrm{T}}_{ik} a^{\mathrm{T}}_{kj})^* = (b^{\mathrm{T}})^*_{ik} (a^{\mathrm{T}})^*_{kj} \\ &= b^\dagger_{ik} a^\dagger_{kj} = (b^\dagger a^\dagger)_{ij}\,,\end{aligned} \tag{3}$$

it follows that

$$(\mathsf{AB})^\dagger = \mathsf{B}^\dagger \mathsf{A}^\dagger. \tag{4}$$

Adjoint Operator

Given a SECOND-ORDER ORDINARY DIFFERENTIAL EQUATION

$$\tilde{\mathcal{L}} u(x) \equiv p_0 \frac{du^2}{dx^2} + p_1 \frac{du}{dx} + p_2 u, \tag{1}$$

where $p_i \equiv p_i(x)$ and $u \equiv u(x)$, the adjoint operator $\tilde{\mathcal{L}}^\dagger$ is defined by

$$\begin{aligned}\tilde{\mathcal{L}}^\dagger u &\equiv \frac{d}{dx^2}(p_0 u) - \frac{d}{dx}(p_1 u) + p_2 u \\ &= p_0 \frac{d^2 u}{dx^2} + (2 p_0{}' - p_1) \frac{du}{dx} + (p_0{}'' - p_1{}' + p_2) u.\end{aligned} \tag{2}$$

Write the two LINEARLY INDEPENDENT solutions as $y_1(x)$ and $y_2(x)$. Then the adjoint operator can also be written

$$\tilde{\mathcal{L}}^\dagger u = \int (y_2\tilde{\mathcal{L}}y_1 - y_1\tilde{\mathcal{L}}y_2)\,dx = \left[\frac{p_1}{p_0}(y_1{}'y_2 - y_1y_2{}')\right]. \tag{3}$$

see also SELF-ADJOINT OPERATOR, STURM-LIOUVILLE THEORY

Adjugate Matrix

see ADJOINT MATRIX

Adjunction

If a is an element of a FIELD F over the PRIME FIELD P, then the set of all RATIONAL FUNCTIONS of a with COEFFICIENTS in P is a FIELD derived from P by adjunction of a.

Adleman-Pomerance-Rumely Primality Test

A modified MILLER'S PRIMALITY TEST which gives a guarantee of PRIMALITY or COMPOSITENESS. The ALGORITHM's running time for a number N has been proved to be as $\mathcal{O}((\ln N)^{c\ln\ln\ln N})$ for some $c > 0$. It was simplified by Cohen and Lenstra (1984), implemented by Cohen and Lenstra (1987), and subsequently optimized by Bosma and van der Hulst (1990).

References

Adleman, L. M.; Pomerance, C.; and Rumely, R. S. "On Distinguishing Prime Numbers from Composite Number." *Ann. Math.* **117**, 173–206, 1983.

Bosma, W. and van der Hulst, M.-P. "Faster Primality Testing." In *Advances in Cryptology, Proc. Eurocrypt '89, Houthalen, April 10–13, 1989* (Ed. J.-J. Quisquater). New York: Springer-Verlag, 652–656, 1990.

Brillhart, J.; Lehmer, D. H.; Selfridge, J.; Wagstaff, S. S. Jr.; and Tuckerman, B. *Factorizations of* $b^n \pm 1$, $b = 2, 3, 5, 6, 7, 10, 11, 12$ *Up to High Powers, rev. ed.* Providence, RI: Amer. Math. Soc., pp. lxxxiv–lxxxv, 1988.

Cohen, H. and Lenstra, A. K. "Primality Testing and Jacobi Sums." *Math. Comput.* **42**, 297–330, 1984.

Cohen, H. and Lenstra, A. K. "Implementation of a New Primality Test." *Math. Comput.* **48**, 103–121, 1987.

Mihailescu, P. "A Primality Test Using Cyclotomic Extensions." In *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes* (Proc. AAECC-6, Rome, July 1988). New York: Springer-Verlag, pp. 310–323, 1989.

Adleman-Rumely Primality Test

see ADLEMAN-POMERANCE-RUMELY PRIMALITY TEST

Admissible

A string or word is said to be admissible if that word appears in a given SEQUENCE. For example, in the SEQUENCE *aabaabaabaabaab*..., *a*, *aa*, *baab* are all admissible, but *bb* is inadmissible.

see also BLOCK GROWTH

Affine Complex Plane

The set $\mathbb{A}^2$ of all ordered pairs of COMPLEX NUMBERS.

see also AFFINE CONNECTION, AFFINE EQUATION, AFFINE GEOMETRY, AFFINE GROUP, AFFINE HULL, AFFINE PLANE, AFFINE SPACE, AFFINE TRANSFORMATION, AFFINITY, COMPLEX PLANE, COMPLEX PROJECTIVE PLANE

Affine Connection

see CONNECTION COEFFICIENT

Affine Equation

A nonhomogeneous LINEAR EQUATION or system of nonhomogeneous LINEAR EQUATIONS is said to be affine.

see also AFFINE COMPLEX PLANE, AFFINE CONNECTION, AFFINE GEOMETRY, AFFINE GROUP, AFFINE HULL, AFFINE PLANE, AFFINE SPACE, AFFINE TRANSFORMATION, AFFINITY

Affine Geometry

A GEOMETRY in which properties are preserved by PARALLEL PROJECTION from one PLANE to another. In an affine geometry, the third and fourth of EUCLID'S POSTULATES become meaningless. This type of GEOMETRY was first studied by Euler.

see also ABSOLUTE GEOMETRY, AFFINE COMPLEX PLANE, AFFINE CONNECTION, AFFINE EQUATION, AFFINE GROUP, AFFINE HULL, AFFINE PLANE, AFFINE SPACE, AFFINE TRANSFORMATION, AFFINITY, ORDERED GEOMETRY

References

Birkhoff, G. and Mac Lane, S. "Affine Geometry." §9.13 in *A Survey of Modern Algebra, 3rd ed.* New York: Macmillan, pp. 268–275, 1965.

Affine Group

The set of all nonsingular AFFINE TRANSFORMATIONS of a TRANSLATION in SPACE constitutes a GROUP known as the affine group. The affine group contains the full linear group and the group of TRANSLATIONS as SUBGROUPS.

see also AFFINE COMPLEX PLANE, AFFINE CONNECTION, AFFINE EQUATION, AFFINE GEOMETRY, AFFINE HULL, AFFINE PLANE, AFFINE SPACE, AFFINE TRANSFORMATION, AFFINITY

References

Birkhoff, G. and Mac Lane, S. *A Survey of Modern Algebra, 3rd ed.* New York: Macmillan, p. 237, 1965.

Affine Hull

The IDEAL generated by a SET in a VECTOR SPACE.

see also AFFINE COMPLEX PLANE, AFFINE CONNECTION, AFFINE EQUATION, AFFINE GEOMETRY, AFFINE GROUP, AFFINE PLANE, AFFINE SPACE, AFFINE TRANSFORMATION, AFFINITY, CONVEX HULL, HULL

Affine Plane

A 2-D Affine Geometry constructed over a Finite Field. For a Field F of size n, the affine plane consists of the set of points which are ordered pairs of elements in F and a set of lines which are themselves a set of points. Adding a Point at Infinity and Line at Infinity allows a Projective Plane to be constructed from an affine plane. An affine plane of order n is a Block Design of the form $(n^2, n, 1)$. An affine plane of order n exists Iff a Projective Plane of order n exists.

see also Affine Complex Plane, Affine Connection, Affine Equation, Affine Geometry, Affine Group, Affine Hull, Affine Space, Affine Transformation, Affinity, Projective Plane

References

Lindner, C. C. and Rodger, C. A. *Design Theory.* Boca Raton, FL: CRC Press, 1997.

Affine Scheme

A technical mathematical object defined as the Spectrum $\sigma(A)$ of a set of Prime Ideals of a commutative Ring A regarded as a local ringed space with a structure sheaf.

see also Scheme

References

Iyanaga, S. and Kawada, Y. (Eds.). "Schemes." §18E in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 69, 1980.

Affine Space

Let V be a Vector Space over a Field K, and let A be a nonempty Set. Now define addition $p + \mathbf{a} \in A$ for any Vector $\mathbf{a} \in V$ and element $p \in A$ subject to the conditions

1. $p + \mathbf{0} = p$,
2. $(p + \mathbf{a}) + \mathbf{b} = p + (\mathbf{a} + \mathbf{b})$,
3. For any $q \in A$, there Exists a unique Vector $\mathbf{a} \in V$ such that $q = p + \mathbf{a}$.

Here, $\mathbf{a}, \mathbf{b} \in V$. Note that (1) is implied by (2) and (3). Then A is an affine space and K is called the Coefficient Field.

In an affine space, it is possible to fix a point and coordinate axis such that every point in the Space can be represented as an n-tuple of its coordinates. Every ordered pair of points A and B in an affine space is then associated with a Vector AB.

see also Affine Complex Plane, Affine Connection, Affine Equation, Affine Geometry, Affine Group, Affine Hull, Affine Plane, Affine Space, Affine Transformation, Affinity

Affine Transformation

Any Transformation preserving Collinearity (i.e., all points lying on a Line initially still lie on a Line after Transformation). An affine transformation is also called an Affinity. An affine transformation of $\mathbb{R}^n$ is a Map $F : \mathbb{R}^n \to \mathbb{R}^n$ of the form

$$F(\mathbf{p}) = A\mathbf{p} + \mathbf{q} \tag{1}$$

for all $p \in \mathbb{R}^n$, where A is a linear transformation of $\mathbb{R}^n$. If $\det(A) = 1$, the transformation is Orientation-Preserving; if $\det(A) = -1$, it is Orientation-Reversing.

Dilation (Contraction, Homothecy), Expansion, Reflection, Rotation, and Translation are all affine transformations, as are their combinations. A particular example combining Rotation and Expansion is the rotation-enlargement transformation

$$\begin{aligned} \begin{bmatrix} x' \\ y' \end{bmatrix} &= s \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} \\ &= s \begin{bmatrix} \cos\alpha(x - x_0) + \sin\alpha(y - y_0) \\ -\sin\alpha(x - x_0) + \cos\alpha(y - y_0) \end{bmatrix}. \end{aligned} \tag{2}$$

Separating the equations,

$$x' = (s\cos\alpha)x + (s\sin\alpha)y - s(x_0\cos\alpha + y_0\sin\alpha) \tag{3}$$

$$y' = (-s\sin\alpha)x + (s\cos\alpha)y + s(x_0\sin\alpha - y_0\cos\alpha). \tag{4}$$

This can be also written as

$$x' = ax + by + c \tag{5}$$

$$y' = bx + ay + d, \tag{6}$$

where

$$a = s\cos\alpha \tag{7}$$

$$b = -s\sin\alpha. \tag{8}$$

The scale factor s is then defined by

$$s \equiv \sqrt{a^2 + b^2}, \tag{9}$$

and the rotation Angle by

$$\alpha = \tan^{-1}\left(-\frac{b}{a}\right). \tag{10}$$

see also Affine Complex Plane, Affine Connection, Affine Equation, Affine Geometry, Affine Group, Affine Hull, Affine Plane, Affine Space, Affine Transformation, Affinity, Equiaffinity, Euclidean Motion

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 105, 1993.

Affinity

see AFFINE TRANSFORMATION

Affix

In the archaic terminology of Whittaker and Watson (1990), the COMPLEX NUMBER z representing $x + iy$.

References

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Aggregate

An archaic word for infinite SETS such as those considered by Georg Cantor.

see also CLASS (SET), SET

AGM

see ARITHMETIC-GEOMETRIC MEAN

Agnesi's Witch

see WITCH OF AGNESI

Agnésienne

see WITCH OF AGNESI

Agonic Lines

see SKEW LINES

Ahlfors-Bers Theorem

The RIEMANN'S MODULI SPACE gives the solution to RIEMANN'S MODULI PROBLEM, which requires an ANALYTIC parameterization of the compact RIEMANN SURFACES in a fixed HOMEOMORPHISM.

Airy Differential Equation

Some authors define a general Airy differential equation as

$$y'' \pm k^2 xy = 0. \tag{1}$$

This equation can be solved by series solution using the expansions

$$y = \sum_{n=0}^{\infty} a_n x^n \tag{2}$$

$$y' = \sum_{n=0}^{\infty} n a_n x^{n-1} = \sum_{n=1}^{\infty} n a_n x^{n-1}$$

$$= \sum_{n=0}^{\infty} (n+1) a_{n+1} x^n \tag{3}$$

$$y'' = \sum_{n=0}^{\infty} (n+1) n a_{n+1} x^{n-1} = \sum_{n=1}^{\infty} (n+1) n a_{n+1} x^{n-1}$$

$$= \sum_{n=0}^{\infty} (n+2)(n+1) a_{n+2} x^n. \tag{4}$$

Specializing to the "conventional" Airy differential equation occurs by taking the MINUS SIGN and setting $k^2 = 1$. Then plug (4) into

$$y'' - xy = 0 \tag{5}$$

to obtain

$$\sum_{n=0}^{\infty} (n+2)(n+1) a_{n+2} x^n - x \sum_{n=0}^{\infty} a_n x^n = 0 \tag{6}$$

$$\sum_{n=0}^{\infty} (n+2)(n+1) a_{n+2} x^n - \sum_{n=0}^{\infty} a_n x^{n+1} = 0 \tag{7}$$

$$2a_2 + \sum_{n=1}^{\infty} (n+2)(n+1) a_{n+2} x^n - \sum_{n=1}^{\infty} a_{n-1} x^n = 0 \tag{8}$$

$$2a_2 + \sum_{n=1}^{\infty} [(n+2)(n+1) a_{n+2} - a_{n-1}] x^n = 0. \tag{9}$$

In order for this equality to hold for all x, each term must separately be 0. Therefore,

$$a_2 = 0 \tag{10}$$

$$(n+2)(n+1) a_{n+2} = a_{n-1}. \tag{11}$$

Starting with the $n = 3$ term and using the above RECURRENCE RELATION, we obtain

$$5 \cdot 4 a_5 = 20 a_5 = a_2 = 0. \tag{12}$$

Continuing, it follows by INDUCTION that

$$a_2 = a_5 = a_8 = a_{11} = \ldots a_{3n-1} = 0 \tag{13}$$

for $n = 1, 2, \ldots$. Now examine terms of the form a_{3n}.

$$a_3 = \frac{a_0}{3 \cdot 2} \tag{14}$$

$$a_6 = \frac{a_3}{6 \cdot 5} = \frac{a_0}{(6 \cdot 5)(3 \cdot 2)} \tag{15}$$

$$a_9 = \frac{a_6}{9 \cdot 8} = \frac{a_0}{(9 \cdot 8)(6 \cdot 5)(3 \cdot 2)}. \tag{16}$$

Again by INDUCTION,

$$a_{3n} = \frac{a_0}{[(3n)(3n-1)][(3n-3)(3n-4)] \cdots [6 \cdot 5][3 \cdot 2]} \tag{17}$$

for $n = 1, 2, \ldots$. Finally, look at terms of the form a_{3n+1},

$$a_4 = \frac{a_1}{4 \cdot 3} \tag{18}$$

$$a_7 = \frac{a_4}{7 \cdot 6} = \frac{a_1}{(7 \cdot 6)(4 \cdot 3)} \tag{19}$$

$$a_{10} = \frac{a_7}{10 \cdot 9} = \frac{a_1}{(10 \cdot 9)(7 \cdot 6)(4 \cdot 3)}. \tag{20}$$

By INDUCTION,

$$a_{3n+1} = \frac{a_1}{[(3n+1)(3n)][(3n-2)(3n-3)]\cdots[7\cdot 6][4\cdot 3]} \tag{21}$$

for $n = 1, 2, \ldots$. The general solution is therefore

$$y = a_0\left[1 + \sum_{n=1}^{\infty}\frac{x^{3n}}{(3n)(3n-1)(3n-3)(3n-4)\cdots 3\cdot 2}\right] + a_1\left[x + \sum_{n=1}^{\infty}\frac{x^{3n+1}}{(3n+1)(3n)(3n-2)(3n-3)\cdots 4\cdot 3}\right]. \tag{22}$$

For a general k^2 with a MINUS SIGN, equation (1) is

$$y'' - k^2xy = 0, \tag{23}$$

and the solution is

$$y(x) = \tfrac{1}{3}\sqrt{x}\left[AI_{-1/3}\left(\tfrac{2}{3}kx^{3/2}\right) - BI_{1/3}\left(\tfrac{2}{3}kx^{3/2}\right)\right], \tag{24}$$

where I is a MODIFIED BESSEL FUNCTION OF THE FIRST KIND. This is usually expressed in terms of the AIRY FUNCTIONS $\operatorname{Ai}(x)$ and $\operatorname{Bi}(x)$

$$y(x) = A'\operatorname{Ai}(k^{2/3}x) + B'\operatorname{Bi}(k^{2/3}x). \tag{25}$$

If the PLUS SIGN is present instead, then

$$y'' + k^2xy = 0 \tag{26}$$

and the solutions are

$$y(x) = \tfrac{1}{3}\sqrt{x}\left[AJ_{-1/3}\left(\tfrac{2}{3}kx^{3/2}\right) + BJ_{1/3}\left(\tfrac{2}{3}kx^{3/2}\right)\right], \tag{27}$$

where $J(z)$ is a BESSEL FUNCTION OF THE FIRST KIND.

see also AIRY-FOCK FUNCTIONS, AIRY FUNCTIONS, BESSEL FUNCTION OF THE FIRST KIND, MODIFIED BESSEL FUNCTION OF THE FIRST KIND

Airy-Fock Functions

The three Airy-Fock functions are

$$\nu(z) = \tfrac{1}{2}\sqrt{\pi}\operatorname{Ai}(z) \tag{1}$$

$$w_1(z) = 2e^{i\pi/6}\nu(\omega z) \tag{2}$$

$$w_2(z) = 2e^{-i\pi/6}\nu(\omega^{-1}z), \tag{3}$$

where $\operatorname{Ai}(z)$ is an AIRY FUNCTION. These functions satisfy

$$\nu(z) = \frac{w_1(z) - w_2(z)}{2i} \tag{4}$$

$$[w_1(z)]^* = w_2(z^*), \tag{5}$$

where z^* is the COMPLEX CONJUGATE of z.

see also AIRY FUNCTIONS

References

Hazewinkel, M. (Managing Ed.). *Encyclopaedia of Mathematics: An Updated and Annotated Translation of the Soviet "Mathematical Encyclopaedia."* Dordrecht, Netherlands: Reidel, p. 65, 1988.

Airy Functions

Watson's (1966, pp. 188–190) definition of an Airy function is the solution to the AIRY DIFFERENTIAL EQUATION

$$\Phi'' \pm k^2\Phi x = 0 \tag{1}$$

which is FINITE at the ORIGIN, where Φ' denotes the DERIVATIVE $d\Phi/dx$, $k^2 = 1/3$, and either SIGN is permitted. Call these solutions $(1/\pi)\Phi(\pm k^2, x)$, then

$$\frac{1}{\pi}\Phi(\pm\tfrac{1}{3};x) \equiv \int_0^\infty \cos(t^3 \pm xt)\,dt \tag{2}$$

$$\Phi(\tfrac{1}{3};x) = \tfrac{1}{3}\pi\sqrt{\tfrac{x}{3}}\left[J_{-1/3}\left(\frac{2x^{3/2}}{3^{3/2}}\right) + J_{1/3}\left(\frac{2x^{3/2}}{3^{3/2}}\right)\right] \tag{3}$$

$$\Phi(-\tfrac{1}{3};x) = \tfrac{1}{3}\pi\sqrt{\tfrac{x}{3}}\left[I_{-1/3}\left(\frac{2x^{3/2}}{3^{3/2}}\right) - I_{1/3}\left(\frac{2x^{3/2}}{3^{3/2}}\right)\right], \tag{4}$$

where $J(z)$ is a BESSEL FUNCTION OF THE FIRST KIND and $I(z)$ is a MODIFIED BESSEL FUNCTION OF THE FIRST KIND. Using the identity

$$K_n(x) = \frac{\pi}{2}\frac{I_{-n}(x) - I_n(x)}{\sin(n\pi)}, \tag{5}$$

where $K(z)$ is a MODIFIED BESSEL FUNCTION OF THE SECOND KIND, the second case can be re-expressed

$$\Phi(-\tfrac{1}{3};x) = \tfrac{1}{3}\pi\sqrt{\tfrac{x}{3}}\left(\frac{2}{\pi}\right)\sin\left(\tfrac{1}{3}\pi\right)K_{1/3}\left(\frac{2x^{3/2}}{3^{3/2}}\right) \tag{6}$$

$$= \frac{\pi}{3}\sqrt{\tfrac{1}{3}x}\,\frac{2}{\pi}\frac{\sqrt{3}}{2}K_{1/3}\left(\frac{2x^{3/2}}{3^{3/2}}\right) \tag{7}$$

$$= \tfrac{1}{3}\sqrt{x}K_{1/3}\left(\frac{2x^{3/2}}{3^{3/2}}\right). \tag{8}$$

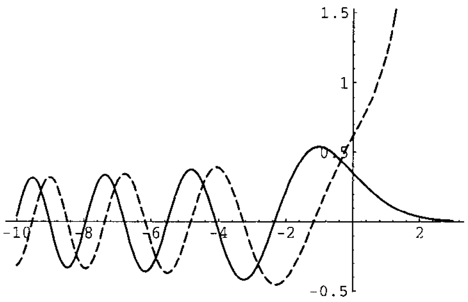

A more commonly used definition of Airy functions is given by Abramowitz and Stegun (1972, pp. 446–447) and illustrated above. This definition identifies the $\operatorname{Ai}(x)$ and $\operatorname{Bi}(x)$ functions as the two LINEARLY INDEPENDENT solutions to (1) with $k^2 = 1$ and a MINUS SIGN,

$$y'' - yz = 0. \tag{9}$$

The solutions are then written

$$y(z) = A\,\mathrm{Ai}(z) + B\,\mathrm{Bi}(z), \tag{10}$$

where

$$\begin{aligned}\mathrm{Ai}(z) &\equiv \frac{1}{\pi}\Phi(-1,z)\\ &= \tfrac{1}{3}\sqrt{x}\,[I_{-1/3}(\tfrac{2}{3}z^{3/2}) - I_{1/3}(\tfrac{2}{3}z^{3/2})]\\ &= \sqrt{\frac{z}{3\pi}}\,K_{1/3}(\tfrac{2}{3}z^{3/2}) \qquad (11)\end{aligned}$$

$$\mathrm{Bi}(z) \equiv \sqrt{\frac{z}{3}}\,[I_{-1/3}(\tfrac{2}{3}z^{3/2}) + I_{1/3}(\tfrac{2}{3}z^{3/2})]. \tag{12}$$

In the above plot, Ai(z) is the solid curve and Bi(z) is dashed. For zero argument,

$$\mathrm{Ai}(0) = -\frac{3^{-2/3}}{\Gamma(\frac{2}{3})}, \tag{13}$$

where $\Gamma(z)$ is the GAMMA FUNCTION. This means that Watson's expression becomes

$$(3a)^{-1/3}\pi\,\mathrm{Ai}(\pm(3a)^{-1/3}x) = \int_0^\infty \cos(at^3 \pm xt)\,dt. \tag{14}$$

A generalization has been constructed by Hardy.

The ASYMPTOTIC SERIES of Ai(z) has a different form in different QUADRANTS of the COMPLEX PLANE, a fact known as the STOKES PHENOMENON. Functions related to the Airy functions have been defined as

$$\mathrm{Gi}(z) \equiv \frac{1}{\pi}\int_0^\infty \sin(\tfrac{1}{3}t^3 + zt)\,dt \tag{15}$$

$$\mathrm{Hi}(z) \equiv \frac{1}{\pi}\int_0^\infty \exp(-\tfrac{1}{3}t^3 + zt)\,dt. \tag{16}$$

see also AIRY-FOCK FUNCTIONS

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Airy Functions." §10.4 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 446–452, 1972.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Bessel Functions of Fractional Order, Airy Functions, Spherical Bessel Functions." §6.7 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 234–245, 1992.

Spanier, J. and Oldham, K. B. "The Airy Functions Ai(x) and Bi(x)." Ch. 56 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 555–562, 1987.

Watson, G. N. *A Treatise on the Theory of Bessel Functions, 2nd ed.* Cambridge, England: Cambridge University Press, 1966.

Airy Projection

A MAP PROJECTION. The inverse equations for ϕ are computed by iteration. Let the ANGLE of the projection plane be θ_b. Define

$$a = \begin{cases} 0 & \text{for } \theta_b = \frac{1}{2}\pi \\ \frac{\ln[\frac{1}{2}\cos(\frac{1}{2}\pi-\theta_b)]}{\tan[\frac{1}{2}(\frac{1}{2}\pi-\theta_b)]} & \text{otherwise.}\end{cases} \tag{1}$$

For proper convergence, let $x_i = \pi/6$ and compute the initial point by checking

$$x_i = \left|\exp[-(\sqrt{x^2+y^2} + a\tan x_i)\tan x_i]\right|. \tag{2}$$

As long as $x_i > 1$, take $x_{i+1} = x_i/2$ and iterate again. The first value for which $x_i < 1$ is then the starting point. Then compute

$$x_i = \cos^{-1}\{\exp[-(\sqrt{x^2+y^2} + a\tan x_i)\tan x_i]\} \tag{3}$$

until the change in x_i between evaluations is smaller than the acceptable tolerance. The (inverse) equations are then given by

$$\phi = \tfrac{1}{2}\pi - 2x_i \tag{4}$$

$$\lambda = \tan^{-1}\left(-\frac{x}{y}\right). \tag{5}$$

Aitken's δ^2 Process

An ALGORITHM which extrapolates the partial sums s_n of a SERIES $\sum_n a_n$ whose CONVERGENCE is approximately geometric and accelerates its rate of CONVERGENCE. The extrapolated partial sum is given by

$$s_n{}' \equiv s_{n+1} - \frac{(s_{n+1}-s_n)^2}{s_{n+1} - 2s_n + s_{n-1}}.$$

see also EULER'S SERIES TRANSFORMATION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 18, 1972.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 160, 1992.

Aitken Interpolation

An algorithm similar to NEVILLE'S ALGORITHM for constructing the LAGRANGE INTERPOLATING POLYNOMIAL. Let $f(x|x_0, x_1, \ldots, x_k)$ be the unique POLYNOMIAL of kth ORDER coinciding with $f(x)$ at $x_0, \ldots, x_k$. Then

$$\begin{aligned} f(x|x_0,x_1) &= \frac{1}{x_1-x_0}\begin{vmatrix} f_0 & x_0-x \\ f_1 & x_1-x \end{vmatrix} \\ f(x|x_0,x_2) &= \frac{1}{x_2-x_0}\begin{vmatrix} f_0 & x_0-x \\ f_2 & x_2-x \end{vmatrix} \\ f(x|x_0,x_1,x_2) &= \frac{1}{x_2-x_1}\begin{vmatrix} f(x|x_0,x_1) & x_1-x \\ f(x|x_0,x_2) & x_2-x \end{vmatrix} \\ f(x|x_0,x_1,x_2,x_3) &= \frac{1}{x_3-x_2}\begin{vmatrix} f(x|x_0,x_1,x_2) & x_2-x \\ f(x|x_0,x_1,x_3) & x_3-x \end{vmatrix}. \end{aligned}$$

see also LAGRANGE INTERPOLATING POLYNOMIAL

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 879, 1972.

Acton, F. S. *Numerical Methods That Work, 2nd printing.* Washington, DC: Math. Assoc. Amer., pp. 93–94, 1990.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 102, 1992.

Ajima-Malfatti Points

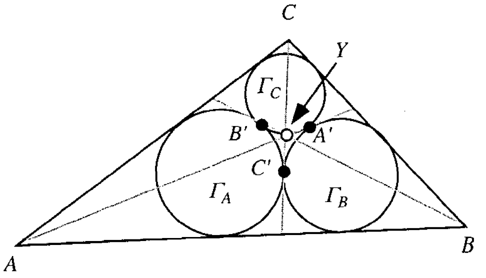

The lines connecting the vertices and corresponding circle-circle intersections in MALFATTI'S TANGENT TRIANGLE PROBLEM coincide in a point Y called the first Ajima-Malfatti point (Kimberling and MacDonald 1990, Kimberling 1994). Similarly, letting A'', B'', and C'' be the excenters of ABC, then the lines $A'A''$, $B'B''$, and $C'C''$ are coincident in another point called the second Ajima-Malfatti point. The points are sometimes simply called the MALFATTI POINTS (Kimberling 1994).

References

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Kimberling, C. "1st and 2nd Ajima-Malfatti Points." `http://www.evansville.edu/~ck6/tcenters/recent/ajmalf.html`.

Kimberling, C. and MacDonald, I. G. "Problem E 3251 and Solution. " *Amer. Math. Monthly* **97**, 612–613, 1990.

Albanese Variety

An ABELIAN VARIETY which is canonically attached to an ALGEBRAIC VARIETY which is the solution to a certain universal problem. The Albanese variety is dual to the PICARD VARIETY.

References

Hazewinkel, M. (Managing Ed.). *Encyclopaedia of Mathematics: An Updated and Annotated Translation of the Soviet "Mathematical Encyclopaedia."* Dordrecht, Netherlands: Reidel, pp. 67–68, 1988.

Albers Conic Projection

see ALBERS EQUAL-AREA CONIC PROJECTION

Albers Equal-Area Conic Projection

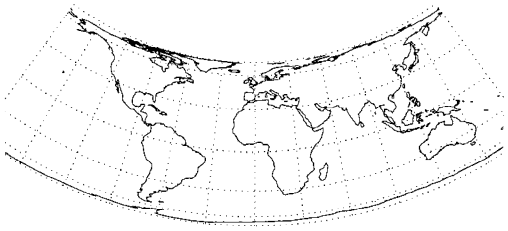

Let ϕ_0 be the LATITUDE for the origin of the CARTESIAN COORDINATES and λ_0 its LONGITUDE. Let ϕ_1 and ϕ_2 be the standard parallels. Then

$$x = \rho\sin\theta \tag{1}$$
$$y = \rho_0 - \rho\cos\theta, \tag{2}$$

where

$$\rho = \frac{\sqrt{C - 2n\sin\phi}}{n} \tag{3}$$
$$\theta = n(\lambda - \lambda_0) \tag{4}$$
$$\rho_0 = \frac{\sqrt{C - 2n\sin\phi_0}}{n} \tag{5}$$
$$C = \cos^2\phi_1 + 2n\sin\phi_1 \tag{6}$$
$$n = \tfrac{1}{2}(\sin\phi_1 + \sin\phi_2). \tag{7}$$

The inverse FORMULAS are

$$\phi = \sin^{-1}\left(\frac{C - \rho^2 n^2}{2n}\right) \tag{8}$$
$$\lambda = \lambda_0 + \frac{\theta}{n}, \tag{9}$$

where

$$\rho = \sqrt{x^2 + (\rho_0 - y)^2} \tag{10}$$
$$\theta = \tan^{-1}\left(\frac{x}{\rho_0 - y}\right). \tag{11}$$

References

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 98–103, 1987.

Alcuin's Sequence

The INTEGER SEQUENCE 1, 0, 1, 1, 2, 1, 3, 2, 4, 3, 5, 4, 7, 5, 8, 7, 10, 8, 12, 10, 14, 12, 16, 14, 19, 16, 21, 19, ... (Sloane's A005044) given by the COEFFICIENTS of the MACLAURIN SERIES for $1/(1-x^2)(1-x^3)(1-x^4)$. The number of different TRIANGLES which have INTEGRAL sides and PERIMETER n is given by

$$T(n) = P_3(n) - \sum_{1\leq j\leq\lfloor n/2\rfloor} P_2(j) \tag{1}$$

$$= \left[\frac{n^2}{12}\right] - \left\lfloor\frac{n}{4}\right\rfloor\left\lfloor\frac{n+2}{4}\right\rfloor \tag{2}$$

$$= \begin{cases} \left[\frac{n^2}{48}\right] & \text{for } n \text{ even} \\ \left[\frac{(n+3)^2}{48}\right] & \text{for } n \text{ odd,} \end{cases} \tag{3}$$

where $P_2(n)$ and $P_3(n)$ are PARTITION FUNCTIONS, with $P_k(n)$ giving the number of ways of writing n as a sum of k terms, $[x]$ is the NINT function, and $\lfloor x\rfloor$ is the FLOOR FUNCTION (Jordan *et al.* 1979, Andrews 1979, Honsberger 1985). Strangely enough, $T(n)$ for $n = 3, 4, \ldots$ is precisely Alcuin's sequence.

see also PARTITION FUNCTION P, TRIANGLE

References
Andrews, G. "A Note on Partitions and Triangles with Integer Sides." *Amer. Math. Monthly* **86**, 477, 1979.
Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 39–47, 1985.
Jordan, J. H.; Walch, R.; and Wisner, R. J. "Triangles with Integer Sides." *Amer. Math. Monthly* **86**, 686–689, 1979.
Sloane, N. J. A. Sequence A005044/M0146 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Aleksandrov-Čech Cohomology

A theory which satisfies all the EILENBERG-STEENROD AXIOMS with the possible exception of the LONG EXACT SEQUENCE OF A PAIR AXIOM, as well as a certain additional continuity CONDITION.

References
Hazewinkel, M. (Managing Ed.). *Encyclopaedia of Mathematics: An Updated and Annotated Translation of the Soviet "Mathematical Encyclopaedia."* Dordrecht, Netherlands: Reidel, p. 68, 1988.

Aleksandrov's Uniqueness Theorem

A convex body in EUCLIDEAN n-space that is centrally symmetric with center at the ORIGIN is determined among all such bodies by its brightness function (the VOLUME of each projection).

see also TOMOGRAPHY

References
Gardner, R. J. "Geometric Tomography." *Not. Amer. Math. Soc.* **42**, 422–429, 1995.

Aleph

The SET THEORY symbol ($\aleph$) for the CARDINALITY of an INFINITE SET.

see also ALEPH-0 ($\aleph_0$), ALEPH-1 ($\aleph_1$), COUNTABLE SET, COUNTABLY INFINITE SET, FINITE, INFINITE, TRANSFINITE NUMBER, UNCOUNTABLY INFINITE SET

Aleph-0 ($\aleph_0$)

The SET THEORY symbol for a SET having the same CARDINAL NUMBER as the "small" INFINITE SET of INTEGERS. The ALGEBRAIC NUMBERS also belong to $\aleph_0$. Rather surprising properties satisfied by $\aleph_0$ include

$${\aleph_0}^r = \aleph_0 \tag{1}$$

$$r\aleph_0 = \aleph_0 \tag{2}$$

$$\aleph_0 + f = \aleph_0, \tag{3}$$

where f is any FINITE SET. However,

$${\aleph_0}^{\aleph_0} = C, \tag{4}$$

where C is the CONTINUUM.

see also ALEPH-1, CARDINAL NUMBER, CONTINUUM, CONTINUUM HYPOTHESIS, COUNTABLY INFINITE SET, FINITE, INFINITE, TRANSFINITE NUMBER, UNCOUNTABLY INFINITE SET

Aleph-1 ($\aleph_1$)

The SET THEORY symbol for the smallest INFINITE SET larger than ALPHA-0 ($\aleph_0$). The CONTINUUM HYPOTHESIS asserts that $\aleph_1 = c$, where c is the CARDINALITY of the "large" INFINITE SET of REAL NUMBERS (called the CONTINUUM in SET THEORY). However, the truth of the CONTINUUM HYPOTHESIS depends on the version of SET THEORY you are using and so is UNDECIDABLE.

Curiously enough, n-D SPACE has the same number of points (c) as 1-D SPACE, or any FINITE INTERVAL of 1-D SPACE (a LINE SEGMENT), as was first recognized by Georg Cantor.

see also ALEPH-0 ($\aleph_0$), CONTINUUM, CONTINUUM HYPOTHESIS, COUNTABLY INFINITE SET, FINITE, INFINITE, TRANSFINITE NUMBER, UNCOUNTABLY INFINITE SET

Alethic

A term in LOGIC meaning pertaining to TRUTH and FALSEHOOD.

see also FALSE, PREDICATE, TRUE

Alexander-Conway Polynomial

see CONWAY POLYNOMIAL

Alexander's Horned Sphere

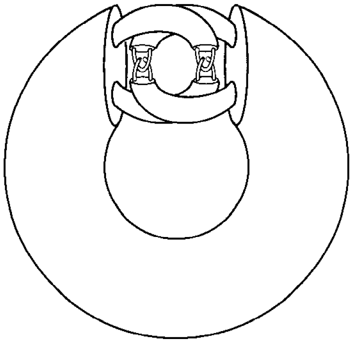

The above solid, composed of a countable UNION of COMPACT SETS, is called Alexander's horned sphere. It is HOMEOMORPHIC with the BALL $\mathbb{B}^3$, and its boundary is therefore a SPHERE. It is therefore an example of a wild embedding in $\mathbb{E}^3$. The outer complement of the solid is not SIMPLY CONNECTED, and its fundamental GROUP is not finitely generated. Furthermore, the set of nonlocally flat ("bad") points of Alexander's horned sphere is a CANTOR SET.

The complement in $\mathbb{R}^3$ of the bad points for Alexander's horned sphere is SIMPLY CONNECTED, making it inequivalent to ANTOINE'S HORNED SPHERE. Alexander's horned sphere has an uncountable infinity of WILD POINTS, which are the limits of the sequences of the horned sphere's branch points (roughly, the "ends" of the horns), since any NEIGHBORHOOD of a limit contains a horned complex.

A humorous drawing by Simon Frazer (Guy 1983, Schroeder 1991, Albers 1994) depicts mathematician John H. Conway with Alexander's horned sphere growing from his head.

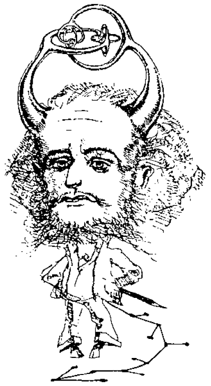

see also ANTOINE'S HORNED SPHERE

References
Albers, D. J. Illustration accompanying "The Game of 'Life'." *Math Horizons,* p. 9, Spring 1994.
Guy, R. "Conway's Prime Producing Machine." *Math. Mag.* **56**, 26–33, 1983.
Hocking, J. G. and Young, G. S. *Topology.* New York: Dover, 1988.
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 80–81, 1976.
Schroeder, M. *Fractals, Chaos, Power Law: Minutes from an Infinite Paradise.* New York: W. H. Freeman, p. 58, 1991.

Alexander Ideal

The order IDEAL in Λ, the RING of integral LAURENT POLYNOMIALS, associated with an ALEXANDER MATRIX for a KNOT K. Any generator of a principal Alexander ideal is called an ALEXANDER POLYNOMIAL. Because the ALEXANDER INVARIANT of a TAME KNOT in $\mathbb{S}^3$ has a SQUARE presentation MATRIX, its Alexander ideal is PRINCIPAL and it has an ALEXANDER POLYNOMIAL $\Delta(t)$.

see also ALEXANDER INVARIANT, ALEXANDER MATRIX, ALEXANDER POLYNOMIAL

References
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 206–207, 1976.

Alexander Invariant

The Alexander invariant $H_*(\tilde{X})$ of a KNOT K is the HOMOLOGY of the INFINITE cyclic cover of the complement of K, considered as a MODULE over Λ, the RING of integral LAURENT POLYNOMIALS. The Alexander invariant for a classical TAME KNOT is finitely presentable, and only H_1 is significant.

For any KNOT K^n in $\mathbb{S}^{n+2}$ whose complement has the homotopy type of a FINITE COMPLEX, the Alexander invariant is finitely generated and therefore finitely presentable. Because the Alexander invariant of a TAME KNOT in $\mathbb{S}^3$ has a SQUARE presentation MATRIX, its ALEXANDER IDEAL is PRINCIPAL and it has an ALEXANDER POLYNOMIAL denoted $\Delta(t)$.

see also ALEXANDER IDEAL, ALEXANDER MATRIX, ALEXANDER POLYNOMIAL

References
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 206–207, 1976.

Alexander Matrix

A presentation matrix for the ALEXANDER INVARIANT $H_1(\tilde{X})$ of a KNOT K. If V is a SEIFERT MATRIX for a TAME KNOT K in $\mathbb{S}^3$, then $V^{\mathrm{T}} - tV$ and $V - tV^{\mathrm{T}}$ are Alexander matrices for K, where V^{T} denotes the MATRIX TRANSPOSE.

see also ALEXANDER IDEAL, ALEXANDER INVARIANT, ALEXANDER POLYNOMIAL, SEIFERT MATRIX

References
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 206–207, 1976.

Alexander Polynomial

A POLYNOMIAL invariant of a KNOT discovered in 1923 by J. W. Alexander (Alexander 1928). In technical language, the Alexander polynomial arises from the HOMOLOGY of the infinitely cyclic cover of a KNOT's complement. Any generator of a PRINCIPAL ALEXANDER IDEAL is called an Alexander polynomial (Rolfsen 1976). Because the ALEXANDER INVARIANT of a TAME KNOT in $\mathbb{S}^3$ has a SQUARE presentation MATRIX, its ALEXANDER IDEAL is PRINCIPAL and it has an Alexander polynomial denoted $\Delta(t)$.

Let Ψ be the MATRIX PRODUCT of BRAID WORDS of a KNOT, then

$$\frac{\det(\mathsf{I} - \Psi)}{1 + t + \ldots + t^{n-1}} = \Delta_L, \tag{1}$$

where Δ_L is the Alexander polynomial and det is the DETERMINANT. The Alexander polynomial of a TAME KNOT in $\mathbb{S}^3$ satisfies

$$\Delta(t) = \det(V^{\mathrm{T}} - tV), \tag{2}$$

where V is a SEIFERT MATRIX, det is the DETERMINANT, and V^{T} denotes the MATRIX TRANSPOSE. The Alexander polynomial also satisfies

$$\Delta(1) = \pm 1. \tag{3}$$

The Alexander polynomial of a splittable link is always 0. Surprisingly, there are known examples of nontrivial KNOTS with Alexander polynomial 1. An example is the $(-3, 5, 7)$ PRETZEL KNOT.

The Alexander polynomial remained the *only* known KNOT POLYNOMIAL until the JONES POLYNOMIAL was discovered in 1984. Unlike the Alexander polynomial, the more powerful JONES POLYNOMIAL *does,* in most cases, distinguish HANDEDNESS. A normalized form of the Alexander polynomial symmetric in t and t^{-1} and satisfying

$$\Delta(\text{unknot}) = 1 \tag{4}$$

was formulated by J. H. Conway and is sometimes denoted ∇_L. The NOTATION $[a + b + c + \ldots$ is an abbreviation for the Conway-normalized Alexander polynomial of a KNOT

$$a + b(x + x^{-1}) + c(x^2 + x^{-2}) + \ldots. \tag{5}$$

For a description of the NOTATION for LINKS, see Rolfsen (1976, p. 389). Examples of the Conway-Alexander polynomials for common KNOTS include

$$\nabla_{\mathrm{TK}} = [1 - 1 = -x^{-1} + 1 - x \tag{6}$$

$$\nabla_{\mathrm{FEK}} = [3 - 1 = -x^{-1} + 3 - x \tag{7}$$

$$\nabla_{\mathrm{SSK}} = [1 - 1 + 1 = x^{-2} - x^{-1} + 1 - x + x^2 \tag{8}$$

for the TREFOIL KNOT, FIGURE-OF-EIGHT KNOT, and SOLOMON'S SEAL KNOT, respectively. Multiplying through to clear the NEGATIVE POWERS gives the usual Alexander polynomial, where the final SIGN is determined by convention.

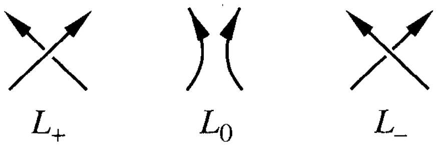

Let an Alexander polynomial be denoted Δ, then there exists a SKEIN RELATIONSHIP (discovered by J. H. Conway)

$$\Delta_{L_+}(t) - \Delta_{L_-}(t) + (t^{-1/2} - t^{1/2})\Delta_{L_0}(t) = 0 \tag{9}$$

corresponding to the above LINK DIAGRAMS (Adams 1994). A slightly different SKEIN RELATIONSHIP convention used by Doll and Hoste (1991) is

$$\nabla_{L_+} - \nabla_{L_-} = z\nabla_{L_0}. \tag{10}$$

These relations allow Alexander polynomials to be constructed for arbitrary knots by building them up as a sequence of over- and undercrossings.

For a KNOT,

$$\Delta_K(-1) \equiv \begin{cases} 1 \pmod 8 & \text{if } \mathrm{Arf}(K) = 0. \\ 5 \pmod 8 & \text{if } \mathrm{Arf}(K) = 1, \end{cases} \tag{11}$$

where Arf is the ARF INVARIANT (Jones 1985). If K is a KNOT and

$$|\Delta_K(i)| > 3, \tag{12}$$

then K cannot be represented as a closed 3-BRAID. Also, if

$$\Delta_K(e^{2\pi i/5}) > \tfrac{13}{2}, \tag{13}$$

then K cannot be represented as a closed 4-braid (Jones 1985).

The HOMFLY POLYNOMIAL $P(a, z)$ generalizes the Alexander polynomial (as well at the JONES POLYNOMIAL) with

$$\nabla(z) = P(1, z) \tag{14}$$

(Doll and Hoste 1991).

Rolfsen (1976) gives a tabulation of Alexander polynomials for KNOTS up to 10 CROSSINGS and LINKS up to 9 CROSSINGS.

see also BRAID GROUP, JONES POLYNOMIAL, KNOT, KNOT DETERMINANT, LINK, SKEIN RELATIONSHIP

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 165–169, 1994.

Alexander, J. W. "Topological Invariants of Knots and Links." *Trans. Amer. Math. Soc.* **30**, 275–306, 1928.

Alexander, J. W. "A Lemma on a System of Knotted Curves." *Proc. Nat. Acad. Sci. USA* **9**, 93–95, 1923.
Doll, H. and Hoste, J. "A Tabulation of Oriented Links." *Math. Comput.* **57**, 747–761, 1991.
Jones, V. "A Polynomial Invariant for Knots via von Neumann Algebras." *Bull. Amer. Math. Soc.* **12**, 103–111, 1985.
Rolfsen, D. "Table of Knots and Links." Appendix C in *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 280–287, 1976.
Stoimenow, A. "Alexander Polynomials." `http://www.informatik.hu-berlin.de/~stoimeno/ptab/a10.html`.
Stoimenow, A. "Conway Polynomials." `http://www.informatik.hu-berlin.de/~stoimeno/ptab/c10.html`.

Alexander-Spanier Cohomology

A fundamental result of de Rham Cohomology is that the kth de Rham Cohomology Vector Space of a Manifold M is canonically isomorphic to the Alexander-Spanier cohomology Vector Space $H^k(M;\mathbb{R})$ (also called cohomology with compact support). In the case that M is Compact, Alexander-Spanier cohomology is exactly "singular" Cohomology.

Alexander's Theorem

Any Link can be represented by a closed Braid.

Algebra

The branch of mathematics dealing with Group Theory and Coding Theory which studies number systems and operations within them. The word "algebra" is a distortion of the Arabic title of a treatise by Al-Khwarizmi about algebraic methods. Note that mathematicians refer to the "school algebra" generally taught in middle and high school as "Arithmetic," reserving the word "algebra" for the more advanced aspects of the subject.

Formally, an algebra is a Vector Space V, over a Field F with a Multiplication which turns it into a Ring defined such that, if $f \in F$ and $x, y \in V$, then

$$f(xy) = (fx)y = x(fy).$$

In addition to the usual algebra of Real Numbers, there are ≈ 1151 additional Consistent algebras which can be formulated by weakening the Field Axioms, at least 200 of which have been *rigorously* proven to be self-Consistent (Bell 1945).

Algebras which have been investigated and found to be of interest are usually named after one or more of their investigators. This practice leads to exotic-sounding (but unenlightening) names which algebraists frequently use with minimal or nonexistent explanation.

see also Alternate Algebra, Alternating Algebra, B^*-Algebra, Banach Algebra, Boolean Algebra, Borel Sigma Algebra, C^*-Algebra, Cayley Algebra, Clifford Algebra, Commutative Algebra, Exterior Algebra, Fundamental Theorem of Algebra, Graded Algebra, Grassmann Algebra, Hecke Algebra, Heyting Algebra, Homological Algebra, Hopf Algebra, Jordan Algebra, Lie Algebra, Linear Algebra, Measure Algebra, Nonassociative Algebra, Quaternion, Robbins Algebra, Schur Algebra, Semisimple Algebra, Sigma Algebra, Simple Algebra, Steenrod Algebra, von Neumann Algebra

References
Artin, M. *Algebra.* Englewood Cliffs, NJ: Prentice-Hall, 1991.
Bell, E. T. *The Development of Mathematics, 2nd ed.* New York: McGraw-Hill, pp. 35–36, 1945.
Bhattacharya, P. B.; Jain, S. K.; and Nagpu, S. R. (Eds.). *Basic Algebra, 2nd ed.* New York: Cambridge University Press, 1994.
Birkhoff, G. and Mac Lane, S. *A Survey of Modern Algebra, 5th ed.* New York: Macmillan, 1996.
Brown, K. S. "Algebra." `http://www.seanet.com/~ksbrown/ialgebra.htm`.
Cardano, G. *Ars Magna or The Rules of Algebra.* New York: Dover, 1993.
Chevalley, C. C. *Introduction to the Theory of Algebraic Functions of One Variable.* Providence, RI: Amer. Math. Soc., 1951.
Chrystal, G. *Textbook of Algebra, 2 vols.* New York: Dover, 1961.
Dickson, L. E. *Algebras and Their Arithmetics.* Chicago, IL: University of Chicago Press, 1923.
Dickson, L. E. *Modern Algebraic Theories.* Chicago, IL: H. Sanborn, 1926.
Edwards, H. M. *Galois Theory, corrected 2nd printing.* New York: Springer-Verlag, 1993.
Euler, L. *Elements of Algebra.* New York: Springer-Verlag, 1984.
Gallian, J. A. *Contemporary Abstract Algebra, 3rd ed.* Lexington, MA: D. C. Heath, 1994.
Grove, L. *Algebra.* New York: Academic Press, 1983.
Hall, H. S. and Knight, S. R. *Higher Algebra, A Sequel to Elementary Algebra for Schools.* London: Macmillan, 1960.
Harrison, M. A. "The Number of Isomorphism Types of Finite Algebras." *Proc. Amer. Math. Soc.* **17**, 735–737, 1966.
Herstein, I. N. *Noncommutative Rings.* Washington, DC: Math. Assoc. Amer., 1996.
Herstein, I. N. *Topics in Algebra, 2nd ed.* New York: Wiley, 1975.
Jacobson, N. *Basic Algebra II, 2nd ed.* New York: W. H. Freeman, 1989.
Kaplansky, I. *Fields and Rings, 2nd ed.* Chicago, IL: University of Chicago Press, 1995.
Lang, S. *Undergraduate Algebra, 2nd ed.* New York: Springer-Verlag, 1990.
Pedersen, J. "Catalogue of Algebraic Systems." `http://tarski.math.usf.edu/algctlg/`.
Uspensky, J. V. *Theory of Equations.* New York: McGraw-Hill, 1948.
van der Waerden, B. L. *Algebra, Vol. 2.* New York: Springer-Verlag, 1991.
van der Waerden, B. L. *Geometry and Algebra in Ancient Civilizations.* New York: Springer-Verlag, 1983.
van der Waerden, B. L. *A History of Algebra: From Al-Khwarizmi to Emmy Noether.* New York: Springer-Verlag, 1985.
Varadarajan, V. S. *Algebra in Ancient and Modern Times.* Providence, RI: Amer. Math. Soc., 1998.

Algebraic Closure

The algebraic closure of a FIELD K is the "smallest" FIELD containing K which is algebraically closed. For example, the FIELD of COMPLEX NUMBERS $\mathbb{C}$ is the algebraic closure of the FIELD of REALS $\mathbb{R}$.

Algebraic Coding Theory

see CODING THEORY

Algebraic Curve

An algebraic curve over a FIELD K is an equation $f(X,Y)=0$, where $f(X,Y)$ is a POLYNOMIAL in X and Y with COEFFICIENTS in K. A nonsingular algebraic curve is an algebraic curve over K which has no SINGULAR POINTS over K. A point on an algebraic curve is simply a solution of the equation of the curve. A K-RATIONAL POINT is a point (X,Y) on the curve, where X and Y are in the FIELD K.

see also ALGEBRAIC GEOMETRY, ALGEBRAIC VARIETY, CURVE

References

Griffiths, P. A. *Introduction to Algebraic Curves.* Providence, RI: Amer. Math. Soc., 1989.

Algebraic Function

A function which can be constructed using only a finite number of ELEMENTARY FUNCTIONS together with the INVERSES of functions capable of being so constructed.

see also ELEMENTARY FUNCTION, TRANSCENDENTAL FUNCTION

Algebraic Function Field

A finite extension $K=\mathbb{Z}(z)(w)$ of the FIELD $\mathbb{C}(z)$ of RATIONAL FUNCTIONS in the indeterminate z, i.e., w is a ROOT of a POLYNOMIAL $a_0+a_1\alpha+a_2\alpha^2+\ldots+a_n\alpha^n$, where $a_i \in \mathbb{C}(z)$.

see also ALGEBRAIC NUMBER FIELD, RIEMANN SURFACE

Algebraic Geometry

The study of ALGEBRAIC CURVES, ALGEBRAIC VARIETIES, and their generalization to n-D.

see also ALGEBRAIC CURVE, ALGEBRAIC VARIETY, COMMUTATIVE ALGEBRA, DIFFERENTIAL GEOMETRY, GEOMETRY, PLANE CURVE, SPACE CURVE

References

Abhyankar, S. S. *Algebraic Geometry for Scientists and Engineers.* Providence, RI: Amer. Math. Soc., 1990.
Cox, D.; Little, J.; and O'Shea, D. *Ideals, Varieties, and Algorithms: An Introduction to Algebraic Geometry and Commutative Algebra, 2nd ed.* New York: Springer-Verlag, 1996.
Eisenbud, D. *Commutative Algebra with a View Toward Algebraic Geometry.* New York: Springer-Verlag, 1995.
Griffiths, P. and Harris, J. *Principles of Algebraic Geometry.* New York: Wiley, 1978.
Hartshorne, R. *Algebraic Geometry, rev. ed.* New York: Springer-Verlag, 1997.
Lang, S. *Introduction to Algebraic Geometry.* New York: Interscience, 1958.
Pedoe, D. and Hodge, W. V. *Methods of Algebraic Geometry, Vol. 1.* Cambridge, England: Cambridge University Press, 1994.
Pedoe, D. and Hodge, W. V. *Methods of Algebraic Geometry, Vol. 2.* Cambridge, England: Cambridge University Press, 1994.
Pedoe, D. and Hodge, W. V. *Methods of Algebraic Geometry, Vol. 3.* Cambridge, England: Cambridge University Press, 1994.
Seidenberg, A. (Ed.). *Studies in Algebraic Geometry.* Washington, DC: Math. Assoc. Amer., 1980.
Weil, A. *Foundations of Algebraic Geometry, enl. ed.* Providence, RI: Amer. Math. Soc., 1962.

Algebraic Integer

If r is a ROOT of the POLYNOMIAL equation

$$x^n+a_{n-1}x^{n-1}+\ldots+a_1x+a_0=0,$$

where the a_is are INTEGERS and r satisfies no similar equation of degree $< n$, then r is an algebraic INTEGER of degree n. An algebraic INTEGER is a special case of an ALGEBRAIC NUMBER, for which the leading COEFFICIENT a_n need not equal 1. RADICAL INTEGERS are a subring of the ALGEBRAIC INTEGERS.

A SUM or PRODUCT of algebraic integers is again an algebraic integer. However, ABEL'S IMPOSSIBILITY THEOREM shows that there are algebraic integers of degree ≥ 5 which are not expressible in terms of ADDITION, SUBTRACTION, MULTIPLICATION, DIVISION, and the extraction of ROOTS on REAL NUMBERS.

The GAUSSIAN INTEGER are are algebraic integers of $\mathbb{Q}(\sqrt{-1})$, since $a+bi$ are roots of

$$z^2-2az+a^2+b^2=0.$$

see also ALGEBRAIC NUMBER, EUCLIDEAN NUMBER, RADICAL INTEGER

References

Hancock, H. *Foundations of the Theory of Algebraic Numbers, Vol. 1: Introduction to the General Theory.* New York: Macmillan, 1931.
Hancock, H. *Foundations of the Theory of Algebraic Numbers, Vol. 2: The General Theory.* New York: Macmillan, 1932.
Pohst, M. and Zassenhaus, H. *Algorithmic Algebraic Number Theory.* Cambridge, England: Cambridge University Press, 1989.
Wagon, S. "Algebraic Numbers." §10.5 in *Mathematica in Action.* New York: W. H. Freeman, pp. 347–353, 1991.

Algebraic Invariant

A quantity such as a DISCRIMINANT which remains unchanged under a given class of algebraic transformations. Such invariants were originally called HYPERDETERMINANTS by Cayley.

see also DISCRIMINANT (POLYNOMIAL), INVARIANT, QUADRATIC INVARIANT

References
Grace, J. H. and Young, A. *The Algebra of Invariants.* New York: Chelsea, 1965.
Gurevich, G. B. *Foundations of the Theory of Algebraic Invariants.* Groningen, Netherlands: P. Noordhoff, 1964.
Hermann, R. and Ackerman, M. *Hilbert's Invariant Theory Papers.*rookline, MA: Math Sci Press, 1978.
Hilbert, D. *Theory of Algebraic Invariants.* Cambridge, England: Cambridge University Press, 1993.
Mumford, D.; Fogarty, J.; and Kirwan, F. *Geometric Invariant Theory, 3rd enl. ed.* New York: Springer-Verlag, 1994.

Algebraic Knot

A single component ALGEBRAIC LINK.

see also ALGEBRAIC LINK, KNOT, LINK

Algebraic Link

A class of fibered knots and links which arises in ALGEBRAIC GEOMETRY. An algebraic link is formed by connecting the NW and NE strings and the SW and SE strings of an ALGEBRAIC TANGLE (Adams 1994).

see also ALGEBRAIC TANGLE, FIBRATION, TANGLE

References
Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 48–49, 1994.
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 335, 1976.

Algebraic Number

If r is a ROOT of the POLYNOMIAL equation

$$a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n = 0, \qquad (1)$$

where the a_is are INTEGERS and r satisfies no similar equation of degree $< n$, then r is an algebraic number of degree n. If r is an algebraic number and $a_0 = 1$, then it is called an ALGEBRAIC INTEGER. It is also true that if the c_is in

$$c_0x^n + c_1x^{n-1} + \ldots + c_{n-1}x + c_n = 0 \qquad (2)$$

are algebraic numbers, then any ROOT of this equation is also an algebraic number.

If α is an algebraic number of degree n satisfying the POLYNOMIAL

$$a(x-\alpha)(x-\beta)(x-\gamma)\cdots, \qquad (3)$$

then there are $n-1$ other algebraic numbers β, γ, ... called the conjugates of α. Furthermore, if α satisfies any other algebraic equation, then its conjugates also satisfy the same equation (Conway and Guy 1996).

Any number which is not algebraic is said to be TRANSCENDENTAL.

see also ALGEBRAIC INTEGER, EUCLIDEAN NUMBER, HERMITE-LINDEMANN THEOREM, RADICAL INTEGER, SEMIALGEBRAIC NUMBER, TRANSCENDENTAL NUMBER

References
Conway, J. H. and Guy, R. K. "Algebraic Numbers." In *The Book of Numbers.* New York: Springer-Verlag, pp. 189–190, 1996.
Courant, R. and Robbins, H. "Algebraic and Transcendental Numbers." §2.6 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 103–107, 1996.
Hancock, H. *Foundations of the Theory of Algebraic Numbers. Vol. 1: Introduction to the General Theory.* New York: Macmillan, 1931.
Hancock, H. *Foundations of the Theory of Algebraic Numbers. Vol. 2: The General Theory.* New York: Macmillan, 1932.
Wagon, S. "Algebraic Numbers." §10.5 in *Mathematica in Action.* New York: W. H. Freeman, pp. 347–353, 1991.

Algebraic Number Field

see NUMBER FIELD

Algebraic Surface

The set of ROOTS of a POLYNOMIAL $f(x,y,z) = 0$. An algebraic surface is said to be of degree $n = \max(i+j+k)$, where n is the maximum sum of powers of all terms $a_m x^{i_m} y^{j_m} z^{k_m}$. The following table lists the names of algebraic surfaces of a given degree.

Order	Surface
3	cubic surface
4	quartic surface
5	quintic surface
6	sextic surface
7	heptic surface
8	octic surface
9	nonic surface
10	decic surface

see also BARTH DECIC, BARTH SEXTIC, BOY SURFACE, CAYLEY CUBIC, CHAIR, CLEBSCH DIAGONAL CUBIC, CUSHION, DERVISH, ENDRASS OCTIC, HEART SURFACE, KUMMER SURFACE, ORDER (ALGEBRAIC SURFACE), ROMAN SURFACE, SURFACE, TOGLIATTI SURFACE

References
Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, p. 7, 1986.

Algebraic Tangle

Any TANGLE obtained by ADDITIONS and MULTIPLICATIONS of rational TANGLES (Adams 1994).

see also ALGEBRAIC LINK

References
Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 41–51, 1994.

Algebraic Topology

The study of intrinsic qualitative aspects of spatial objects (e.g., SURFACES, SPHERES, TORI, CIRCLES, KNOTS, LINKS, configuration spaces, etc.) that remain invariant under both-directions continuous ONE-TO-ONE (HOMEOMORPHIC) transformations. The discipline of algebraic topology is popularly known as "RUBBER-SHEET GEOMETRY" and can also be viewed as the study of DISCONNECTIVITIES. Algebraic topology has a great deal of mathematical machinery for studying different kinds of HOLE structures, and it gets the prefix "algebraic" since many HOLE structures are represented best by algebraic objects like GROUPS and RINGS.

A technical way of saying this is that algebraic topology is concerned with FUNCTORS from the topological CATEGORY of GROUPS and HOMOMORPHISMS. Here, the FUNCTORS are a kind of filter, and given an "input" SPACE, they spit out something else in return. The returned object (usually a GROUP or RING) is then a representation of the HOLE structure of the SPACE, in the sense that this algebraic object is a vestige of what the original SPACE was like (i.e., much information is lost, but some sort of "shadow" of the SPACE is retained—just enough of a shadow to understand some aspect of its HOLE-structure, but no more). The idea is that FUNCTORS give much simpler objects to deal with. Because SPACES by themselves are very complicated, they are unmanageable without looking at particular aspects.

COMBINATORIAL TOPOLOGY is a special type of algebraic topology that uses COMBINATORIAL methods.

see also CATEGORY, COMBINATORIAL TOPOLOGY, DIFFERENTIAL TOPOLOGY, FUNCTOR, HOMOTOPY THEORY

References

Dieudonné, J. *A History of Algebraic and Differential Topology: 1900–1960.* Boston, MA: Birkhäuser, 1989.

Algebraic Variety

A generalization to n-D of ALGEBRAIC CURVES. More technically, an algebraic variety is a reduced SCHEME of FINITE type over a FIELD K. An algebraic variety V is defined as the SET of points in the REALS $\mathbb{R}^n$ (or the COMPLEX NUMBERS $\mathbb{C}^n$) satisfying a system of POLYNOMIAL equations $f_i(x_1, \ldots, x_n) = 0$ for $i = 1, 2, \ldots$. According to the HILBERT BASIS THEOREM, a FINITE number of equations suffices.

see also ABELIAN VARIETY, ALBANESE VARIETY, BRAUER-SEVERI VARIETY, CHOW VARIETY, PICARD VARIETY

References

Ciliberto, C.; Laura, E.; and Somese, A. J. (Eds.). *Classification of Algebraic Varieties.* Providence, RI: Amer. Math. Soc., 1994.

Algebroidal Function

An ANALYTIC FUNCTION $f(z)$ satisfying the irreducible algebraic equation

$$A_0(z)f^k + A_1(z)f^{k-1} + \ldots + A_k(z) = 0$$

with single-valued MEROMORPHIC functions $A_j(z)$ in a COMPLEX DOMAIN G is called a k-algebroidal function in G.

References

Iyanaga, S. and Kawada, Y. (Eds.). "Algebroidal Functions." §19 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 86–88, 1980.

Algorithm

A specific set of instructions for carrying out a procedure or solving a problem, usually with the requirement that the procedure terminate at some point. Specific algorithms sometimes also go by the name METHOD, PROCEDURE, or TECHNIQUE. The word "algorithm" is a distortion of Al-Khwarizmi, an Arab mathematician who wrote an influential treatise about algebraic methods.

see also 196-ALGORITHM, ALGORITHMIC COMPLEXITY, ARCHIMEDES ALGORITHM, BHASKARA-BROUCKNER ALGORITHM, BORCHARDT-PFAFF ALGORITHM, BRELAZ'S HEURISTIC ALGORITHM, BUCHBERGER'S ALGORITHM, BULIRSCH-STOER ALGORITHM, BUMPING ALGORITHM, CLEAN ALGORITHM, COMPUTABLE FUNCTION, CONTINUED FRACTION FACTORIZATION ALGORITHM, DECISION PROBLEM, DIJKSTRA'S ALGORITHM, EUCLIDEAN ALGORITHM, FERGUSON-FORCADE ALGORITHM, FERMAT'S ALGORITHM, FLOYD'S ALGORITHM, GAUSSIAN APPROXIMATION ALGORITHM, GENETIC ALGORITHM, GOSPER'S ALGORITHM, GREEDY ALGORITHM, HASSE'S ALGORITHM, HJLS ALGORITHM, JACOBI ALGORITHM, KRUSKAL'S ALGORITHM, LEVINE-O'SULLIVAN GREEDY ALGORITHM, LLL ALGORITHM, MARKOV ALGORITHM, MILLER'S ALGORITHM, NEVILLE'S ALGORITHM, NEWTON'S METHOD, PRIME FACTORIZATION ALGORITHMS, PRIMITIVE RECURSIVE FUNCTION, PROGRAM, PSLQ ALGORITHM, PSOS ALGORITHM, QUOTIENT-DIFFERENCE ALGORITHM, RISCH ALGORITHM, SCHRAGE'S ALGORITHM, SHANKS' ALGORITHM, SPIGOT ALGORITHM, SYRACUSE ALGORITHM, TOTAL FUNCTION, TURING MACHINE, ZASSENHAUS-BERLEKAMP ALGORITHM, ZEILBERGER'S ALGORITHM

References

Aho, A. V.; Hopcroft, J. E.; and Ullman, J.D. *The Design and Analysis of Computer Algorithms.* Reading, MA: Addison-Wesley, 1974.

Baase, S. *Computer Algorithms.* Reading, MA: Addison-Wesley, 1988.

Brassard, G. and Bratley, P. *Fundamentals of Algorithmics.* Englewood Cliffs, NJ: Prentice-Hall, 1995.

Cormen, T. H.; Leiserson, C. E.; and Rivest, R. L. *Introduction to Algorithms.* Cambridge, MA: MIT Press, 1990.

Greene, D. H. and Knuth, D. E. *Mathematics for the Analysis of Algorithms, 3rd ed.* Boston: Birkhäuser, 1990.
Harel, D. *Algorithmics: The Spirit of Computing, 2nd ed.* Reading, MA: Addison-Wesley, 1992.
Knuth, D. E. *The Art of Computer Programming, Vol. 1: Fundamental Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, 1973.
Knuth, D. E. *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, 1981.
Knuth, D. E. *The Art of Computer Programming, Vol. 3: Sorting and Searching, 2nd ed.* Reading, MA: Addison-Wesley, 1973.
Kozen, D. C. *Design and Analysis and Algorithms.* New York: Springer-Verlag, 1991.
Shen, A. *Algorithms and Programming.* Boston: Birkhäuser, 1996.
Skiena, S. S. *The Algorithm Design Manual.* New York: Springer-Verlag, 1997.
Wilf, H. *Algorithms and Complexity.* Englewood Cliffs, NJ: Prentice Hall, 1986. `http://www.cis.upenn.edu/~wilf/`.

Algorithmic Complexity

see BIT COMPLEXITY, KOLMOGOROV COMPLEXITY

Alhazen's Billiard Problem

In a given CIRCLE, find an ISOSCELES TRIANGLE whose LEGS pass through two given POINTS inside the CIRCLE. This can be restated as: from two POINTS in the PLANE of a CIRCLE, draw LINES meeting at the POINT of the CIRCUMFERENCE and making equal ANGLES with the NORMAL at that POINT.

The problem is called the billiard problem because it corresponds to finding the POINT on the edge of a circular "BILLIARD" table at which a cue ball at a given POINT must be aimed in order to carom once off the edge of the table and strike another ball at a second given POINT. The solution leads to a BIQUADRATIC EQUATION of the form

$$H(x^2-y^2)-2Kxy+(x^2+y^2)(hy-kx)=0.$$

The problem is equivalent to the determination of the point on a spherical mirror where a ray of light will reflect in order to pass from a given source to an observer. It is also equivalent to the problem of finding, given two points and a CIRCLE such that the points are both inside or outside the CIRCLE, the ELLIPSE whose FOCI are the two points and which is tangent to the given CIRCLE.

The problem was first formulated by Ptolemy in 150 AD, and was named after the Arab scholar Alhazen, who discussed it in his work on optics. It was not until 1997 that Neumann proved the problem to be insoluble using a COMPASS and RULER construction because the solution requires extraction of a CUBE ROOT. This is the same reason that the CUBE DUPLICATION problem is insoluble.

see also BILLIARDS, BILLIARD TABLE PROBLEM, CUBE DUPLICATION

References
Dörrie, H. "Alhazen's Billiard Problem." §41 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 197–200, 1965.
Hogendijk, J. P. "Al-Mutaman's Simplified Lemmas for Solving 'Alhazen's Problem'." *From Baghdad to Barcelona/De Bagdad a Barcelona, Vol. I, II (Zaragoza, 1993)*, pp. 59–101, Anu. Filol. Univ. Barc., XIX B-2, Univ. Barcelona, Barcelona, 1996.
Lohne, J. A. "Alhazens Spiegelproblem." *Nordisk Mat. Tidskr.* **18**, 5–35, 1970.
Neumann, P. Submitted to *Amer. Math. Monthly.*
Riede, H. "Reflexion am Kugelspiegel. Oder: das Problem des Alhazen." *Praxis Math.* **31**, 65–70, 1989.
Sabra, A. I. "ibn al-Haytham's Lemmas for Solving 'Alhazen's Problem'." *Arch. Hist. Exact Sci.* **26**, 299-324, 1982.

Alhazen's Problem

see ALHAZEN'S BILLIARD PROBLEM

Alias' Paradox

Choose between the following two alternatives:

1. 90% chance of an unknown amount x and a 10% chance of \$1 million, or
2. 89% chance of the same unknown amount x, 10% chance of \$2.5 million, and 1% chance of nothing.

The PARADOX is to determine which choice has the larger expectation value, $0.9x+\$100,000$ or $0.89x+\$250,000$. However, the best choice depends on the unknown amount, even though it is the same in both cases! This appears to violate the INDEPENDENCE AXIOM.

see also INDEPENDENCE AXIOM, MONTY HALL PROBLEM, NEWCOMB'S PARADOX

Aliasing

Given a power spectrum (a plot of power vs. frequency), aliasing is a false translation of power falling in some frequency range $(-f_c, f_c)$ outside the range. Aliasing can be caused by discrete sampling below the NYQUIST FREQUENCY. The sidelobes of any INSTRUMENT FUNCTION (including the simple SINC SQUARED function obtained simply from FINITE sampling) are also a form of aliasing. Although sidelobe contribution at large offsets can be minimized with the use of an APODIZATION FUNCTION, the tradeoff is a widening of the response (i.e., a lowering of the resolution).

see also APODIZATION FUNCTION, NYQUIST FREQUENCY

Aliquant Divisor

A number which does not DIVIDE another exactly. For instance, 4 and 5 are aliquant divisors of 6. A number which is not an aliquant divisor (i.e., one that *does* DIVIDE another exactly) is said to be an ALIQUOT DIVISOR.

see also ALIQUOT DIVISOR, DIVISOR, PROPER DIVISOR

Aliquot Cycle

see SOCIABLE NUMBERS

Aliquot Divisor

A number which DIVIDES another exactly. For instance, 1, 2, 3, and 6 are aliquot divisors of 6. A number which is not an aliquot divisor is said to be an ALIQUANT DIVISOR. The term "aliquot" is frequently used to specifically mean a PROPER DIVISOR, i.e., a DIVISOR of a number other than the number itself.

see also ALIQUANT DIVISOR, DIVISOR, PROPER DIVISOR

Aliquot Sequence

Let

$$s(n) \equiv \sigma(n) - n,$$

where $\sigma(n)$ is the DIVISOR FUNCTION and $s(n)$ is the RESTRICTED DIVISOR FUNCTION. Then the SEQUENCE of numbers

$$s^0(n) \equiv n, s^1(n) = s(n), s^2(n) = s(s(n)), \ldots$$

is called an aliquot sequence. If the SEQUENCE for a given n is bounded, it either ends at $s(1) = 0$ or becomes periodic.

1. If the SEQUENCE reaches a constant, the constant is known as a PERFECT NUMBER.
2. If the SEQUENCE reaches an alternating pair, it is called an AMICABLE PAIR.
3. If, after k iterations, the SEQUENCE yields a cycle of minimum length t of the form $s^{k+1}(n)$, $s^{k+2}(n)$, ..., $s^{k+t}(n)$, then these numbers form a group of SOCIABLE NUMBERS of order t.

It has not been proven that all aliquot sequences eventually terminate and become period. The smallest number whose fate is not known is 276, which has been computed up to $s^{487}(276)$ (Guy 1994).

see also 196-ALGORITHM, ADDITIVE PERSISTENCE, AMICABLE NUMBERS, MULTIAMICABLE NUMBERS, MULTIPERFECT NUMBER, MULTIPLICATIVE PERSISTENCE, PERFECT NUMBER, SOCIABLE NUMBERS, UNITARY ALIQUOT SEQUENCE

References
Guy, R. K. "Aliquot Sequences." §B6 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 60–62, 1994.
Guy, R. K. and Selfridge, J. L. "What Drives Aliquot Sequences." *Math. Comput.* **29**, 101–107, 1975.
Sloane, N. J. A. Sequences A003023/M0062 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Sloane, N. J. A. and Plouffe, S. Extended entry in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.

All-Poles Model

see MAXIMUM ENTROPY METHOD

Alladi-Grinstead Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let $N(n)$ be the number of ways in which the FACTORIAL $n!$ can be decomposed into n FACTORS of the form ${p_k}^{b_k}$ arranged in nondecreasing order. Also define

$$m(n) \equiv \max({p_1}^{b_1}), \tag{1}$$

i.e., $m(n)$ is the LEAST PRIME FACTOR raised to its appropriate POWER in the factorization. Then define

$$\alpha(n) \equiv \frac{\ln m(n)}{\ln n} \tag{2}$$

where $\ln(x)$ is the NATURAL LOGARITHM. For instance,

$$\begin{aligned} 9! &= 2\cdot 2\cdot 2\cdot 2\cdot 2\cdot 2^2\cdot 5\cdot 7\cdot 3^4 \\ &= 2\cdot 2\cdot 2\cdot 2\cdot 3\cdot 5\cdot 7\cdot 2^3\cdot 3^3 \\ &= 2\cdot 2\cdot 2\cdot 2\cdot 5\cdot 7\cdot 2^3\cdot 3^2\cdot 3^2 \\ &= 2\cdot 2\cdot 2\cdot 3\cdot 2^2\cdot 2^2\cdot 5\cdot 7\cdot 3^3 \\ &= 2\cdot 2\cdot 2\cdot 2^2\cdot 2^2\cdot 5\cdot 7\cdot 3^2\cdot 3^2 \\ &= 2\cdot 2\cdot 2\cdot 3\cdot 3\cdot 5\cdot 7\cdot 3^2\cdot 2^4 \\ &= 2\cdot 2\cdot 3\cdot 3\cdot 2^2\cdot 5\cdot 7\cdot 2^3\cdot 3^2 \\ &= 2\cdot 2\cdot 3\cdot 3\cdot 3\cdot 3\cdot 5\cdot 7\cdot 2^5 \\ &= 2\cdot 3\cdot 3\cdot 2^2\cdot 2^2\cdot 2^2\cdot 5\cdot 7\cdot 3^2 \\ &= 2\cdot 3\cdot 3\cdot 3\cdot 3\cdot 2^2\cdot 5\cdot 7\cdot 2^4 \\ &= 2\cdot 3\cdot 3\cdot 3\cdot 3\cdot 5\cdot 7\cdot 2^3\cdot 2^3 \\ &= 3\cdot 3\cdot 3\cdot 3\cdot 2^2\cdot 2^2\cdot 5\cdot 7\cdot 2^3, \end{aligned} \tag{3}$$

so

$$\alpha(9) = \frac{\ln 3}{\ln 9} = \frac{\ln 3}{2\ln 3} = \frac{1}{2}. \tag{4}$$

For large n,

$$\lim_{n\to\infty} \alpha(n) = e^{c-1} = 0.809394020534\ldots, \tag{5}$$

where

$$c \equiv \sum_{k=2}^{\infty} \frac{1}{k} \ln\left(\frac{k}{k-1}\right). \tag{6}$$

References
Alladi, K. and Grinstead, C. "On the Decomposition of $n!$ into Prime Powers." *J. Number Th.* **9**, 452–458, 1977.
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/aldgrns/aldgrns.html`.
Guy, R. K. "Factorial n as the Product of n Large Factors." §B22 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 79, 1994.

Allegory

A technical mathematical object which bears the same resemblance to binary relations as CATEGORIES do to FUNCTIONS and SETS.

see also CATEGORY

References
Freyd, P. J. and Scedrov, A. *Categories, Allegories.* Amsterdam, Netherlands: North-Holland, 1990.

Allometric

Mathematical growth in which one population grows at a rate PROPORTIONAL to the POWER of another population.

References

Cofrey, W. J. *Geography Towards a General Spatial Systems Approach.* London: Routledge, Chapman & Hall, 1981.

Almost All

Given a property P, if $P(x) \sim x$ as $x \to \infty$ (so the number of numbers less than x not satisfying the property P is $o(x)$), then P is said to hold true for almost all numbers. For example, almost all positive integers are COMPOSITE NUMBERS (which is not in conflict with the second of EUCLID'S THEOREMS that there are an infinite number of PRIMES).

see also FOR ALL, NORMAL ORDER

References

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, p. 8, 1979.

Almost Alternating Knot

An ALMOST ALTERNATING LINK with a single component.

Almost Alternating Link

Call a projection of a LINK an almost alternating projection if one crossing change in the projection makes it an alternating projection. Then an almost alternating link is a LINK with an almost alternating projection, but no alternating projection. Every ALTERNATING KNOT has an almost alternating projection. A PRIME KNOT which is almost alternating is either a TORUS KNOT or a HYPERBOLIC KNOT. Therefore, no SATELLITE KNOT is an almost alternating knot.

All nonalternating 9-crossing PRIME KNOTS are almost alternating. Of the 393 nonalternating with 11 or fewer crossings, all but five are known to be nonalternating (3 of these have 11 crossings). The fate of the remaining five is not known. The $(2,q)$, $(3,4)$, and $(3,5)$-TORUS KNOTS are almost alternating.

see also ALTERNATING KNOT, LINK

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 139–146, 1994.

Almost Everywhere

A property of X is said to hold almost everywhere if the SET of points in X where this property fails has MEASURE 0.

see also MEASURE

References

Sansone, G. *Orthogonal Functions, rev. English ed.* New York: Dover, p. 1, 1991.

Almost Integer

A number which is very close to an INTEGER. One surprising example involving both e and PI is

$$e^{\pi} - \pi = 19.999099979\ldots, \tag{1}$$

which can also be written as

$$(\pi + 20)^{i} = -0.9999999992 - 0.0000388927i \approx -1 \tag{2}$$

$$\cos(\ln(\pi + 20)) \approx -0.9999999992. \tag{3}$$

Applying COSINE a few more times gives

$$\begin{aligned}&\cos(\pi \cos(\pi \cos(\ln(\pi + 20))))\\ &\qquad \approx -1 + 3.9321609261 \times 10^{-35}.\end{aligned} \tag{4}$$

This curious near-identity was apparently noticed almost simultaneously around 1988 by N. J. A. Sloane, J. H. Conway, and S. Plouffe, but no satisfying explanation as to "why" it has been true has yet been discovered.

An interesting near-identity is given by

$$\begin{aligned}&\frac{1}{4}\left[\cos(\tfrac{1}{10}) + \cosh(\tfrac{1}{10}) + 2\cos(\tfrac{1}{20}\sqrt{2}\,)\cosh(\tfrac{1}{20}\sqrt{2}\,)\right]\\ &\qquad = 1 + 2.480\ldots \times 10^{-13}\end{aligned} \tag{5}$$

(W. Dubuque). Other remarkable near-identities are given by

$$\frac{5(1+\sqrt{5}\,)[\Gamma(\frac{3}{4})]^2}{e^{5\pi/6}\sqrt{\pi}} = 1 + 4.5422\ldots \times 10^{-14}, \tag{6}$$

where $\Gamma(z)$ is the GAMMA FUNCTION (S. Plouffe), and

$$e^6 - \pi^4 - \pi^5 = 0.000017673\ldots \tag{7}$$

(D. Wilson).

A whole class of IRRATIONAL "almost integers" can be found using the theory of MODULAR FUNCTIONS, and a few rather spectacular examples are given by Ramanujan (1913–14). Such approximations were also studied by Hermite (1859), Kronecker (1863), and Smith (1965). They can be generated using some amazing (and very deep) properties of the j-FUNCTION. Some of the numbers which are closest approximations to INTEGERS are $e^{\pi\sqrt{163}}$ (sometimes known as the RAMANUJAN CONSTANT and which corresponds to the field $\mathbb{Q}(\sqrt{-163}\,)$ which has CLASS NUMBER 1 and is the IMAGINARY quadratic field of maximal discriminant), $e^{\pi\sqrt{22}}$, $e^{\pi\sqrt{37}}$, and $e^{\pi\sqrt{58}}$, the latter three of which have CLASS NUMBER 2 and are due to Ramanujan (Berndt 1994, Waldschmidt 1988).

The properties of the j-FUNCTION also give rise to the spectacular identity

$$\left[\frac{\ln(640320^3+744)}{\pi}\right]^2 = 163 + 2.32167\ldots \times 10^{-29} \quad (8)$$

(Le Lionnais 1983, p. 152).

The list below gives numbers of the form $x \equiv e^{\pi\sqrt{n}}$ for $n \leq 1000$ for which $\lceil x \rceil - x \leq 0.01$.

$$\begin{aligned}
e^{\pi\sqrt{6}} &= 2,197.990\,869\,543\ldots\\
e^{\pi\sqrt{17}} &= 422,150.997\,675\,680\ldots\\
e^{\pi\sqrt{18}} &= 614,551.992\,885\,619\ldots\\
e^{\pi\sqrt{22}} &= 2,508,951.998\,257\,553\ldots\\
e^{\pi\sqrt{25}} &= 6,635,623.999\,341\,134\ldots\\
e^{\pi\sqrt{37}} &= 199,148,647.999\,978\,046\,551\ldots\\
e^{\pi\sqrt{43}} &= 884,736,743.999\,777\,466\ldots\\
e^{\pi\sqrt{58}} &= 24,591,257,751.999\,999\,822\,213\ldots\\
e^{\pi\sqrt{59}} &= 30,197,683,486.993\,182\,260\ldots\\
e^{\pi\sqrt{67}} &= 147,197,952,743.999\,998\,662\,454\ldots\\
e^{\pi\sqrt{74}} &= 54,551,812,208.999\,917\,467\,885\ldots\\
e^{\pi\sqrt{149}} &= 45,116,546,012,289,599.991\,830\,287\ldots\\
e^{\pi\sqrt{163}} &= 262,537,412,640,768,743.999\,999\,999\,999\,250\,072\ldots\\
e^{\pi\sqrt{177}} &= 1,418,556,986,635,586,485.996\,179\,355\ldots\\
e^{\pi\sqrt{232}} &= 604,729,957,825,300,084,759.999\,992\,171\,526\ldots\\
e^{\pi\sqrt{267}} &= 19,683,091,854,079,461,001,445.992\,737\,040\ldots
\end{aligned}$$

$$\begin{aligned}
e^{\pi\sqrt{326}} &= 4,309,793,301,730,386,363,005,719.996\,011\,651\ldots\\
e^{\pi\sqrt{386}} &= 639,355,180,631,208,421,\cdots\\
&\qquad \cdots 212,174,016.997\,669\,832\ldots\\
e^{\pi\sqrt{522}} &= 14,871,070,263,238,043,663,567,\cdots\\
&\qquad \cdots 627,879,007.999\,848\,726\ldots\\
e^{\pi\sqrt{566}} &= 288,099,755,064,053,264,917,867,\cdots\\
&\qquad \cdots 975,825,573.993\,898\,311\ldots\\
e^{\pi\sqrt{638}} &= 28,994,858,898,043,231,996,779,\cdots\\
&\qquad \cdots 771,804,797,161.992\,372\,939\ldots\\
e^{\pi\sqrt{719}} &= 3,842,614,373,539,548,891,490,\cdots\\
&\qquad \cdots 294,277,805,829,192.999\,987\,249\ldots\\
e^{\pi\sqrt{790}} &= 223,070,667,213,077,889,794,379,\cdots\\
&\qquad \cdots 623,183,838,336,437.992\,055\,118\ldots\\
e^{\pi\sqrt{792}} &= 249,433,117,287,892,229,255,125,\cdots\\
&\qquad \cdots 388,685,911,710,805.996\,097\,323\ldots\\
e^{\pi\sqrt{928}} &= 365,698,321,891,389,219,219,142,\cdots\\
&\qquad \cdots 531,076,638,716,362,775.998\,259\,747\ldots\\
e^{\pi\sqrt{986}} &= 6,954,830,200,814,801,770,418,837,\cdots\\
&\qquad 940,281,460,320,666,108.994\,649\,611\ldots.
\end{aligned}$$

Gosper noted that the expression

$$\begin{aligned}1 - 262537412640768744e^{-\pi\sqrt{163}} - 196884e^{-2\pi\sqrt{163}}\\ +103378831900730205293632e^{-3\pi\sqrt{163}}. \quad (9)\end{aligned}$$

differs from an INTEGER by a mere 10^{-59}.

see also CLASS NUMBER, j-FUNCTION, PI

References

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 90–91, 1994.

Hermite, C. "Sur la théorie des équations modulaires." *C. R. Acad. Sci. (Paris)* **48**, 1079–1084 and 1095–1102, 1859.

Hermite, C. "Sur la théorie des équations modulaires." *C. R. Acad. Sci. (Paris)* **49**, 16–24, 110–118, and 141–144, 1859.

Kronecker, L. "Über die Klassenzahl der aus Werzeln der Einheit gebildeten komplexen Zahlen." *Monatsber. K. Preuss. Akad. Wiss. Berlin*, 340–345. 1863.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, 1983.

Ramanujan, S. "Modular Equations and Approximations to π." *Quart. J. Pure Appl. Math.* **45**, 350–372, 1913–1914.

Smith, H. J. S. *Report on the Theory of Numbers.* New York: Chelsea, 1965.

Waldschmidt, M. "Some Transcendental Aspects of Ramanujan's Work." In *Ramanujan Revisited: Proceedings of the Centenary Conference* (Ed. G. E. Andrews, B. C. Berndt, and R. A. Rankin). New York: Academic Press, pp. 57–76, 1988.

Almost Perfect Number

A number n for which the DIVISOR FUNCTION satisfies $\sigma(n) = 2n - 1$ is called almost perfect. The only known almost perfect numbers are the POWERS of 2, namely 1, 2, 4, 8, 16, 32, ... (Sloane's A000079). Singh (1997) calls almost perfect numbers SLIGHTLY DEFECTIVE.

see also QUASIPERFECT NUMBER

References

Guy, R. K. "Almost Perfect, Quasi-Perfect, Pseudoperfect, Harmonic, Weird, Multiperfect and Hyperperfect Numbers." §B2 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 16 and 45–53, 1994.

Singh, S. *Fermat's Enigma: The Epic Quest to Solve the World's Greatest Mathematical Problem.* New York: Walker, p. 13, 1997.

Sloane, N. J. A. Sequence A000079/M1129 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Almost Prime

A number n with prime factorization

$$n = \prod_{i=1}^{r} {p_i}^{a_i}$$

is called k-almost prime when the sum of the POWERS $\sum_{i=1}^{r} a_i = k$. The set of k-almost primes is denoted P_k.

The PRIMES correspond to the "1-almost prime" numbers 2, 3, 5, 7, 11, ... (Sloane's A000040). The 2-almost prime numbers correspond to SEMIPRIMES 4, 6, 9, 10, 14, 15, 21, 22, ... (Sloane's A001358). The first few 3-almost primes are 8, 12, 18, 20, 27, 28, 30, 42, 44, 45, 50, 52, 63, 66, 68, 70, 75, 76, 78, 92, 98, 99, ... (Sloane's A014612). The first few 4-almost primes are 16, 24, 36, 40, 54, 56, 60, 81, 84, 88, 90, 100, ... (Sloane's A014613). The first few 5-almost primes are 32, 48, 72, 80, ... (Sloane's A014614).

see also Chen's Theorem, Prime Number, Semiprime

References
Sloane, N. J. A. Sequences A014612, A014613, A014614, A000040/M0652, and A001358/M3274 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Alpha

A financial measure giving the difference between a fund's actual return and its expected level of performance, given its level of risk (as measured by Beta). A Positive alpha indicates that a fund has performed better than expected based on its Beta, whereas a Negative alpha indicates poorer performance

see also Beta, Sharpe Ratio

Alpha Function

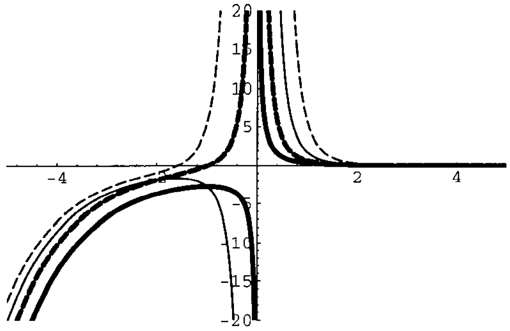

$$\alpha_n(z) \equiv \int_1^\infty t^n e^{-zt}\,dt = n! z^{-(n+1)} e^{-z} \sum_{k=0}^{n} \frac{z^k}{k!}.$$

The alpha function satisfies the Recurrence Relation

$$z\alpha_n(z) = e^{-z} + n\alpha_{n-1}(z).$$

see also Beta Function (Exponential)

Alpha Value

An alpha value is a number $0 \leq \alpha \leq 1$ such that $P(z \geq z_{\text{observed}}) \leq \alpha$ is considered "Significant," where P is a P-Value.

see also Confidence Interval, P-Value, Significance

Alphabet

A Set (usually of letters) from which a Subset is drawn. A sequence of letters is called a Word, and a set of Words is called a Code.

see also Code, Word

Alphamagic Square

A Magic Square for which the number of letters in the word for each number generates another Magic Square. This definition depends, of course, on the language being used. In English, for example,

$$\begin{array}{ccc} 5 & 22 & 18 \\ 28 & 15 & 2 \\ 12 & 8 & 25 \end{array} \qquad \begin{array}{ccc} 4 & 9 & 8 \\ 11 & 7 & 3 \\ 6 & 5 & 10 \end{array},$$

where the Magic Square on the right corresponds to the number of letters in

$$\begin{array}{ccc} \text{five} & \text{twenty-two} & \text{eighteen} \\ \text{twenty-eight} & \text{fifteen} & \text{two} \\ \text{twelve} & \text{eight} & \text{twenty-five} \end{array}.$$

References
Sallows, L. C. F. "Alphamagic Squares." *Abacus* 4, 28–45, 1986.
Sallows, L. C. F. "Alphamagic Squares. 2." *Abacus* 4, 20–29 and 43, 1987.
Sallows, L. C. F. "Alpha Magic Squares." In *The Lighter Side of Mathematics* (Ed. R. K. Guy and R. E. Woodrow). Washington, DC: Math. Assoc. Amer., 1994.

Alphametic

A Cryptarithm in which the letters used to represent distinct Digits are derived from related words or meaningful phrases. The term was coined by Hunter in 1955 (Madachy 1979, p. 178).

References
Brooke, M. *One Hundred & Fifty Puzzles in Crypt-Arithmetic.* New York: Dover, 1963.
Hunter, J. A. H. and Madachy, J. S. "Alphametics and the Like." Ch. 9 in *Mathematical Diversions.* New York: Dover, pp. 90–95, 1975.
Madachy, J. S. "Alphametics." Ch. 7 in *Madachy's Mathematical Recreations.* New York: Dover, pp. 178–200 1979.

Alternate Algebra

Let A denote an $\mathbb{R}$-Algebra, so that A is a Vector Space over R and

$$A \times A \to A \tag{1}$$

$$(x, y) \mapsto x \cdot y. \tag{2}$$

Then A is said to be alternate if, for all $x, y \in A$,

$$(x \cdot y) \cdot y = x \cdot (y \cdot y) \tag{3}$$

$$(x \cdot x) \cdot y = x \cdot (x \cdot y). \tag{4}$$

Here, Vector Multiplication $x \cdot y$ is assumed to be Bilinear.

References
Finch, S. "Zero Structures in Real Algebras." `http://www.mathsoft.com/asolve/zerodiv/zerodiv.html`.
Schafer, R. D. *An Introduction to Non-Associative Algebras.* New York: Dover, 1995.

Alternating Algebra

see EXTERIOR ALGEBRA

Alternating Group

EVEN PERMUTATION GROUPS A_n which are NORMAL SUBGROUPS of the PERMUTATION GROUP of ORDER $n!/2$. They are FINITE analogs of the families of simple LIE GROUPS. The lowest order alternating group is 60. Alternating groups with $n \geq 5$ are non-ABELIAN SIMPLE GROUPS. The number of conjugacy classes in the alternating groups A_n for $n = 2, 3, \ldots$ are 1, 3, 4, 5, 7, 9, ... (Sloane's A000702).

see also 15 PUZZLE, FINITE GROUP, GROUP, LIE GROUP, SIMPLE GROUP, SYMMETRIC GROUP

References

Sloane, N. J. A. Sequence A000702/M2307 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas#alt`.

Alternating Knot

An alternating knot is a KNOT which possesses a knot diagram in which crossings alternate between under- and overpasses. Not all knot diagrams of alternating knots need be alternating diagrams.

The TREFOIL KNOT and FIGURE-OF-EIGHT KNOT are alternating knots. One of TAIT'S KNOT CONJECTURES states that the number of crossings is the same for any diagram of a reduced alternating knot. Furthermore, a reduced alternating projection of a knot has the least number of crossings for any projection of that knot. Both of these facts were proved true by Kauffman (1988), Thistlethwaite (1987), and Murasugi (1987).

If K has a reduced alternating projection of n crossings, then the SPAN of K is $4n$. Let $c(K)$ be the CROSSING NUMBER. Then an alternating knot $K_1 \# K_2$ (a KNOT SUM) satisfies

$$c(K_1 \# K_2) = c(K_1) + c(K_2).$$

In fact, this is true as well for the larger class of ADEQUATE KNOTS and postulated for all KNOTS. The number of PRIME alternating knots of n crossing for $n = 1, 2, \ldots$ are 0, 0, 1, 1, 2, 3, 7, 18, 41, 123, 367, ... (Sloane's A002864).

see also ADEQUATE KNOT, ALMOST ALTERNATING LINK, ALTERNATING LINK, FLYPING CONJECTURE

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 159–164, 1994.

Arnold, B.; Au, M.; Candy, C.; Erdener, K.; Fan, J.; Flynn, R.; Muir, J.; Wu, D.; and Hoste, J. "Tabulating Alternating Knots through 14 Crossings." `ftp://chs.cusd.claremont.edu/pub/knot/paper.TeX.txt` and `ftp://chs.cusd.claremont.edu/pub/knot/AltKnots/`.

Erdener, K. and Flynn, R. "Rolfsen's Table of all Alternating Diagrams through 9 Crossings." `ftp://chs.cusd.claremont.edu/pub/knot/Rolfsen_table.final`.

Kauffman, L. "New Invariants in the Theory of Knots." *Amer. Math. Monthly* **95**, 195–242, 1988.

Murasugi, K. "Jones Polynomials and Classical Conjectures in Knot Theory." *Topology* **26**, 297–307, 1987.

Sloane, N. J. A. Sequence A002864/M0847 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Thistlethwaite, M. "A Spanning Tree Expansion for the Jones Polynomial." *Topology* **26**, 297–309, 1987.

Alternating Knot Diagram

A KNOT DIAGRAM which has alternating under- and overcrossings as the KNOT projection is traversed. The first KNOT which does not have an alternating diagram has 8 crossings.

Alternating Link

A LINK which has a LINK DIAGRAM with alternating underpasses and overpasses.

see also ALMOST ALTERNATING LINK

References

Menasco, W. and Thistlethwaite, M. "The Classification of Alternating Links." *Ann. Math.* **138**, 113–171, 1993.

Alternating Permutation

An arrangement of the elements $c_1, \ldots, c_n$ such that no element c_i has a magnitude between c_{i-1} and c_{i+1} is called an alternating (or ZIGZAG) permutation. The determination of the number of alternating permutations for the set of the first n INTEGERS $\{1, 2, \ldots, n\}$ is known as ANDRÉ'S PROBLEM. An example of an alternating permutation is (1, 3, 2, 5, 4).

As many alternating permutations among n elements begin by rising as by falling. The magnitude of the c_ns does not matter; only the number of them. Let the number of alternating permutations be given by $Z_n = 2A_n$. This quantity can then be computed from

$$2na_n = \sum a_r a_s, \tag{1}$$

where r and s pass through all INTEGRAL numbers such that

$$r + s = n - 1, \tag{2}$$

$a_0 = a_1 = 1$, and

$$A_n = n!a_n. \tag{3}$$

The numbers A_n are sometimes called the EULER ZIGZAG NUMBERS, and the first few are given by 1, 1, 1, 2, 5, 16, 61, 272, ... (Sloane's A000111). The ODD-numbered A_ns are called EULER NUMBERS, SECANT NUMBERS, or ZIG NUMBERS, and the EVEN-numbered ones are sometimes called TANGENT NUMBERS or ZAG NUMBERS.

Curiously enough, the SECANT and TANGENT MACLAURIN SERIES can be written in terms of the A_ns as

$$\sec x = A_0 + A_2\frac{x^2}{2!} + A_4\frac{x^4}{4!} + \ldots \qquad (4)$$

$$\tan x = A_1 x + A_3\frac{x^3}{3!} + A_5\frac{x^5}{5!} + \ldots, \qquad (5)$$

or combining them,

$$\sec x + \tan x$$
$$= A_0 + A_1 x + A_2\frac{x^2}{2!} + A_3\frac{x^3}{3!} + A_4\frac{x^4}{4!} + A_5\frac{x^5}{5!} + \ldots. \qquad (6)$$

see also ENTRINGER NUMBER, EULER NUMBER, EULER ZIGZAG NUMBER, SECANT NUMBER, SEIDEL-ENTRINGER-ARNOLD TRIANGLE, TANGENT NUMBER

References

André, D. "Developments de $\sec x$ et $\tan x$." *C. R. Acad. Sci. Paris* **88**, 965–967, 1879.

André, D. "Memoire sur le permutations alternées." *J. Math.* **7**, 167–184, 1881.

Arnold, V. I. "Bernoulli-Euler Updown Numbers Associated with Function Singularities, Their Combinatorics and Arithmetics." *Duke Math. J.* **63**, 537–555, 1991.

Arnold, V. I. "Snake Calculus and Combinatorics of Bernoulli, Euler, and Springer Numbers for Coxeter Groups." *Russian Math. Surveys* **47**, 3–45, 1992.

Bauslaugh, B. and Ruskey, F. "Generating Alternating Permutations Lexicographically." *BIT* **30**, 17–26, 1990.

Conway, J. H. and Guy, R. K. In *The Book of Numbers.* New York: Springer-Verlag, pp. 110–111, 1996.

Dörrie, H. "André's Deviation of the Secant and Tangent Series." §16 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 64–69, 1965.

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 69–75, 1985.

Knuth, D. E. and Buckholtz, T. J. "Computation of Tangent, Euler, and Bernoulli Numbers." *Math. Comput.* **21**, 663–688, 1967.

Millar, J.; Sloane, N. J. A.; and Young, N. E. "A New Operation on Sequences: The Boustrophedon Transform." *J. Combin. Th. Ser. A* **76**, 44–54, 1996.

Ruskey, F. "Information of Alternating Permutations." `http:// sue . csc . uvic . ca / ~ cos / inf / perm / Alternating.html`.

Sloane, N. J. A. Sequence A000111/M1492 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Alternating Series

A SERIES of the form

$$\sum_{k=1}^{\infty}(-1)^{k+1}a_k$$

or

$$\sum_{k=1}^{\infty}(-1)^k a_k.$$

see also SERIES

References

Arfken, G. "Alternating Series." §5.3 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 293–294, 1985.

Bromwich, T. J. I'a and MacRobert, T. M. "Alternating Series." §19 in *An Introduction to the Theory of Infinite Series, 3rd ed.* New York: Chelsea, pp. 55–57, 1991.

Pinsky, M. A. "Averaging an Alternating Series." *Math. Mag.* **51**, 235–237, 1978.

Alternating Series Test

Also known as the LEIBNIZ CRITERION. An ALTERNATING SERIES CONVERGES if $a_1 \geq a_2 \geq \ldots$ and

$$\lim_{k\to\infty} a_k = 0.$$

see also CONVERGENCE TESTS

Alternative Link

A category of LINK encompassing both ALTERNATING KNOTS and TORUS KNOTS.

see also ALTERNATING KNOT, LINK, TORUS KNOT

References

Kauffman, L. "Combinatorics and Knot Theory." *Contemp. Math.* **20**, 181–200, 1983.

Altitude

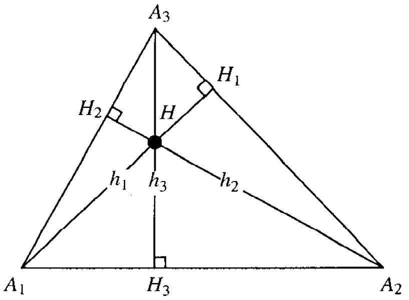

The altitudes of a TRIANGLE are the CEVIANS A_iH_i which are PERPENDICULAR to the LEGS A_jA_k opposite A_i. They have lengths $h_i \equiv \overline{A_iH_i}$ given by

$$h_i = a_{i+1}\sin\alpha_{i+2} = a_{i+2}\sin\alpha_{i+1} \qquad (1)$$

$$h_1 = \frac{2\sqrt{s(s-a_1)(s-a_2)(s-a_3)}}{a_1}, \qquad (2)$$

where s is the SEMIPERIMETER and $a_i \equiv \overline{A_jA_k}$. Another interesting FORMULA is

$$h_1h_2h_3 = 2s\Delta \qquad (3)$$

(Johnson 1929, p. 191), where Δ is the AREA of the TRIANGLE. The three altitudes of any TRIANGLE are CONCURRENT at the ORTHOCENTER H. This fundamental fact did not appear anywhere in Euclid's *Elements.*

Other formulas satisfied by the altitude include

$$\frac{1}{h_1} + \frac{1}{h_2} + \frac{1}{h_3} = \frac{1}{r} \qquad (4)$$

$$\frac{1}{r_1} = \frac{1}{h_2} + \frac{1}{h_3} - \frac{1}{h_1} \tag{5}$$

$$\frac{1}{r_2} + \frac{1}{r_3} = \frac{1}{r} - \frac{1}{r_1} = \frac{2}{h_1}, \tag{6}$$

where r is the Inradius and r_i are the Exradii (Johnson 1929, p. 189). In addition,

$$HA_1 \cdot HH_1 = HA_2 \cdot HH_2 = HA_3 \cdot HH_3 \tag{7}$$

$$HA_1 \cdot HH_1 = \tfrac{1}{2}({a_1}^2 + {a_2}^2 + {a_3}^2) - 4R^2, \tag{8}$$

where R is the Circumradius.

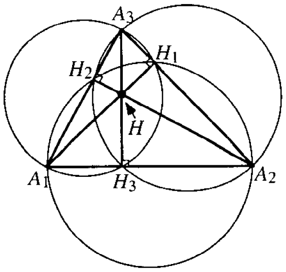

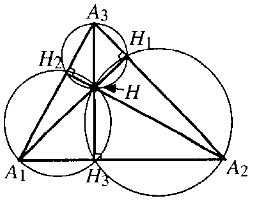

The points A_1, A_3, H_1, and H_3 (and their permutations with respect to indices) all lie on a Circle, as do the points A_3, H_3, H, and H_1 (and their permutations with respect to indices). Triangles $\Delta A_1A_2A_3$ and $\Delta A_1H_2H_3$ are inversely similar.

The triangle $H_1H_2H_3$ has the minimum Perimeter of any Triangle inscribed in a given Acute Triangle (Johnson 1929, pp. 161–165). The Perimeter of $\Delta H_1H_2H_3$ is $2\Delta/R$ (Johnson 1929, p. 191). Additional properties involving the Feet of the altitudes are given by Johnson (1929, pp. 261–262).

see also Cevian, Foot, Orthocenter, Perpendicular, Perpendicular Foot

References
Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 9 and 36–40, 1967.
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, 1929.

Alysoid

see Catenary

Ambient Isotopy

An ambient isotopy from an embedding of a Manifold M in N to another is a Homotopy of self Diffeomorphisms (or Isomorphisms, or piecewise-linear transformations, etc.) of N, starting at the Identity Map, such that the "last" Diffeomorphism compounded with the first embedding of M is the second embedding of M. In other words, an ambient isotopy is like an Isotopy except that instead of distorting the embedding, the whole ambient Space is being stretched and distorted and the embedding is just "coming along for the ride." For Smooth Manifolds, a Map is Isotopic Iff it is ambiently isotopic.

For Knots, the equivalence of Manifolds under continuous deformation is independent of the embedding Space. Knots of opposite Chirality have ambient isotopy, but not Regular Isotopy.

see also Isotopy, Regular Isotopy

References
Hirsch, M. W. *Differential Topology.* New York: Springer-Verlag, 1988.

Ambiguous

An expression is said to be ambiguous (or poorly defined) if its definition does not assign it a unique interpretation or value. An expression which is *not* ambiguous is said to be Well-Defined.

see also Well-Defined

Ambrose-Kakutani Theorem

For every ergodic Flow on a nonatomic Probability Space, there is a Measurable Set intersecting almost every orbit in a discrete set.

Amenable Number

A number n which can be built up from Integers a_1, a_2, ..., a_k by either Addition or Multiplication such that

$$\sum_{i=1}^{k} a_i = \prod_{i=1}^{k} a_i = n.$$

The numbers $\{a_1, \ldots, a_n\}$ in the Sum are simply a Partition of n. The first few amenable numbers are

$$\begin{aligned} 2+2 &= 2\times 2 = 4 \\ 1+2+3 &= 1\times 2\times 3 = 6 \\ 1+1+2+4 &= 1\times 1\times 2\times 4 = 8 \\ 1+1+2+2+2 &= 1\times 1\times 2\times 2\times 2 = 8. \end{aligned}$$

In fact, all Composite Numbers are amenable.

see also Composite Number, Partition, Sum

References
Tamvakis, H. "Problem 10454." *Amer. Math. Monthly* **102**, 463, 1995.

Amicable Numbers

see Amicable Pair, Amicable Quadruple, Amicable Triple, Multiamicable Numbers

Amicable Pair

An amicable pair consists of two INTEGERS m, n for which the sum of PROPER DIVISORS (the DIVISORS excluding the number itself) of one number equals the other. Amicable pairs are occasionally called FRIENDLY PAIRS, although this nomenclature is to be discouraged since FRIENDLY PAIRS are defined by a different, if related, criterion. Symbolically, amicable pairs satisfy

$$s(m) = n \tag{1}$$
$$s(n) = m, \tag{2}$$

where $s(n)$ is the RESTRICTED DIVISOR FUNCTION or, equivalently,

$$\sigma(m) = \sigma(n) = s(m) + s(n) = m + n, \tag{3}$$

where $\sigma(n)$ is the DIVISOR FUNCTION. The smallest amicable pair is (220, 284) which has factorizations

$$220 = 11 \cdot 5 \cdot 2^2 \tag{4}$$
$$284 = 71 \cdot 2^2 \tag{5}$$

giving RESTRICTED DIVISOR FUNCTIONS

$$\begin{aligned} s(220) &= \sum\{1, 2, 4, 5, 10, 11, 20, 22, 44, 55, 110\} \\ &= 284 \end{aligned} \tag{6}$$
$$\begin{aligned} s(284) &= \sum\{1, 2, 4, 71, 142\} \\ &= 220. \end{aligned} \tag{7}$$

The quantity

$$\sigma(m) = \sigma(n) = s(m) + s(n), \tag{8}$$

in this case, $220 + 284 = 504$, is called the PAIR SUM.

In 1636, Fermat found the pair (17296, 18416) and in 1638, Descartes found (9363584, 9437056). By 1747, Euler had found 30 pairs, a number which he later extended to 60. There were 390 known as of 1946 (Scott 1946). There are a total of 236 amicable pairs below 10^8 (Cohen 1970), 1427 below 10^{10} (te Riele 1986), 3340 less than 10^{11} (Moews and Moews 1993), [illegible] less than 2.01×10^{11} (Moews and Moews), and 5001 less than $\approx 3.06 \times 10^{11}$ (Moews and Moews).

The first few amicable pairs are (220, 284), (1184, 1210), (2620, 2924) (5020, 5564), (6232, 6368), (10744, 10856), (12285, 14595), (17296, 18416), (63020, 76084), ... (Sloane's A002025 and A002046). An exhaustive tabulation is maintained by D. Moews.

Let an amicable pair be denoted (m, n) with $m < n$. (m, n) is called a regular amicable pair of type (i, j) if

$$(m, n) = (gM, gN), \tag{9}$$

where $g \equiv \mathrm{GCD}(m, n)$ is the GREATEST COMMON DIVISOR,

$$\mathrm{GCD}(g, M) = \mathrm{GCD}(g, N) = 1, \tag{10}$$

M and N are SQUAREFREE, then the number of PRIME factors of M and N are i and j. Pairs which are not regular are called irregular or exotic (te Riele 1986). There are no regular pairs of type $(1, j)$ for $j \geq 1$. If $m \equiv 0 \pmod 6$ and

$$n = \sigma(m) - m \tag{11}$$

is EVEN, then (m, n) cannot be an amicable pair (Lee 1969). The minimal and maximal values of m/n found by te Riele (1986) were

$$938304290/1344480478 = 0.697893577\ldots \tag{12}$$

and

$$4000783984/4001351168 = 0.9998582519\ldots. \tag{13}$$

te Riele (1986) also found 37 pairs of amicable pairs having the same PAIR SUM. The first such pair is (609928, 686072) and (643336, 652664), which has the PAIR SUM

$$\sigma(m) = \sigma(n) = m + n = 1{,}296{,}000. \tag{14}$$

te Riele (1986) found no amicable n-tuples having the same PAIR SUM for $n > 2$. However, Moews and Moews found a triple in 1993, and te Riele found a quadruple in 1995. In November 1997, a quintuple and sextuple were discovered. The sextuple is (1953433861918, 2216492794082), (1968039941816, 2201886714184), (1981957651366, 2187969004634), (1993501042130, 2176425613870), (2046897812505, 2123028843495), (2068113162038, 2101813493962), all having PAIR SUM 4169926656000. Amazingly, the sextuple is smaller than any known quadruple or quintuple, and is likely smaller than any quintuple.

On October 4, 1997, Mariano Garcia found the largest known amicable pair, each of whose members has 4829 DIGITS. The new pair is

$$N_1 = CM[(P + Q)P^{89} - 1] \tag{15}$$
$$N_2 = CQ[(P - M)P^{89} - 1], \tag{16}$$

where

$$C = 2^{11}P^{89} \tag{17}$$
$$M = 287155430510003638403359267 \tag{18}$$
$$P = 574451143340278962374313859 \tag{19}$$
$$Q = 136272576607912041393307632916794623. \tag{20}$$

P, Q, $(P + Q)P^{89} - 1$, and $(P - M)P^{89} - 1$ are PRIME.

Pomerance (1981) has proved that

$$[\text{amicable numbers } \leq n] < ne^{-[\ln(n)]^{1/3}} \qquad (21)$$

for large enough n (Guy 1994). No nonfinite lower bound has been proven.

see also AMICABLE QUADRUPLE, AMICABLE TRIPLE, AUGMENTED AMICABLE PAIR, BREEDER, CROWD, EULER'S RULE, FRIENDLY PAIR, MULTIAMICABLE NUMBERS, PAIR SUM, QUASIAMICABLE PAIR, SOCIABLE NUMBERS, UNITARY AMICABLE PAIR

References

Alanen, J.; Ore, Ø.; and Stemple, J. "Systematic Computations on Amicable Numbers." *Math. Comput.* **21**, 242–245, 1967.

Battiato, S. and Borho, W. "Are there Odd Amicable Numbers not Divisible by Three?" *Math. Comput.* **50**, 633–637, 1988.

Beeler, M.; Gosper, R. W.; and Schroeppel, R. Item 62 in *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Borho, W. and Hoffmann, H. "Breeding Amicable Numbers in Abundance." *Math. Comput.* **46**, 281–293, 1986.

Bratley, P.; Lunnon, F.; and McKay, J. "Amicable Numbers and Their Distribution." *Math. Comput.* **24**, 431–432, 1970.

Cohen, H. "On Amicable and Sociable Numbers." *Math. Comput.* **24**, 423–429, 1970.

Costello, P. "Amicable Pairs of Euler's First Form." *J. Rec. Math.* **10**, 183–189, 1977–1978.

Costello, P. "Amicable Pairs of the Form $(i,1)$." *Math. Comput.* **56**, 859–865, 1991.

Dickson, L. E. *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, pp. 38–50, 1952.

Erdős, P. "On Amicable Numbers." *Publ. Math. Debrecen* **4**, 108–111, 1955–1956.

Erdős, P. "On Asymptotic Properties of Aliquot Sequences." *Math. Comput.* **30**, 641–645, 1976.

Gardner, M. "Perfect, Amicable, Sociable." Ch. 12 in *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, pp. 160–171, 1978.

Guy, R. K. "Amicable Numbers." §B4 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 55–59, 1994.

Lee, E. J. "Amicable Numbers and the Bilinear Diophantine Equation." *Math. Comput.* **22**, 181–197, 1968.

Lee, E. J. "On Divisibility of the Sums of Even Amicable Pairs." *Math. Comput.* **23**, 545–548, 1969.

Lee, E. J. and Madachy, J. S. "The History and Discovery of Amicable Numbers, I." *J. Rec. Math.* **5**, 77–93, 1972.

Lee, E. J. and Madachy, J. S. "The History and Discovery of Amicable Numbers, II." *J. Rec. Math.* **5**, 153–173, 1972.

Lee, E. J. and Madachy, J. S. "The History and Discovery of Amicable Numbers, III." *J. Rec. Math.* **5**, 231–249, 1972.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 145 and 155–156, 1979.

Moews, D. and Moews, P. C. "A Search for Aliquot Cycles and Amicable Pairs." *Math. Comput.* **61**, 935–938, 1993.

Moews, D. and Moews, P. C. "A List of Amicable Pairs Below 2.01×10^{11}." Rev. Jan. 8, 1993. `http://xraysgi.ims.uconn.edu:8080/amicable.txt`.

Moews, D. and Moews, P. C. "A List of the First 5001 Amicable Pairs." Rev. Jan. 7, 1996. `http://xraysgi.ims.uconn.edu:8080/amicable2.txt`.

Ore, Ø. *Number Theory and Its History.* New York: Dover, pp. 96–100, 1988.

Pedersen, J. M. "Known Amicable Pairs." `http://www.vejlehs.dk/staff/jmp/aliquot/knwnap.htm`.

Pomerance, C. "On the Distribution of Amicable Numbers." *J. reine angew. Math.* **293/294**, 217–222, 1977.

Pomerance, C. "On the Distribution of Amicable Numbers, II." *J. reine angew. Math.* **325**, 182–188, 1981.

Scott, E. B. E. "Amicable Numbers." *Scripta Math.* **12**, 61–72, 1946.

Sloane, N. J. A. Sequences A002025/M5414 and A002046/M5435 in "An On-Line Version of the Encyclopedia of Integer Sequences."

te Riele, H. J. J. "On Generating New Amicable Pairs from Given Amicable Pairs." *Math. Comput.* **42**, 219–223, 1984.

te Riele, H. J. J. "Computation of All the Amicable Pairs Below 10^{10}." *Math. Comput.* **47**, 361–368 and S9–S35, 1986.

te Riele, H. J. J.; Borho, W.; Battiato, S.; Hoffmann, H.; and Lee, E. J. "Table of Amicable Pairs Between $10^1$0 and 10^{52}." Centrum voor Wiskunde en Informatica, Note NM-N8603. Amsterdam: Stichting Math. Centrum, 1986.

te Riele, H. J. J. "A New Method for Finding Amicable Pairs." In *Mathematics of Computation 1943–1993: A Half-Century of Computational Mathematics (Vancouver, BC, August 9–13, 1993)* (Ed. W. Gautschi). Providence, RI: Amer. Math. Soc., pp. 577–581, 1994.

Weisstein, E. W. "Sociable and Amicable Numbers." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Sociable.m`.

Amicable Quadruple

An amicable quadruple as a QUADRUPLE (a,b,c,d) such that

$$\sigma(a) = \sigma(b) = \sigma(c) = \sigma(d) = a + b + c + d,$$

where $\sigma(n)$ is the DIVISOR FUNCTION.

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 59, 1994.

Amicable Triple

Dickson (1913, 1952) defined an amicable triple to be a TRIPLE of three numbers (l,m,n) such that

$$\begin{aligned} s(l) &= m + n \\ s(m) &= l + n \\ s(n) &= l + m, \end{aligned}$$

where $s(n)$ is the RESTRICTED DIVISOR FUNCTION (Madachy 1979). Dickson (1913, 1952) found eight sets of amicable triples with two equal numbers, and two sets with distinct numbers. The latter are (123228768, 103340640, 124015008), for which

$$\begin{aligned} s(12322876) &= 103340640 + 124015008 = 227355648 \\ s(103340640) &= 123228768 + 124015008 = 24724377 \\ s(124015008) &= 123228768 + 10334064 = 226569408, \end{aligned}$$

and (1945330728960, 2324196638720, 2615631953920), for which

$$\begin{aligned} s(1945330728960) &= 2324196638720 + 2615631953920 \\ &= 4939828592640 \\ s(2324196638720) &= 1945330728960 + 2615631953920 \\ &= 4560962682880 \\ s(2615631953920) &= 1945330728960 + 2324196638720 \\ &= 4269527367680. \end{aligned}$$

A second definition (Guy 1994) defines an amicable triple as a TRIPLE (a, b, c) such that

$$\sigma(a) = \sigma(b) = \sigma(c) = a + b + c,$$

where $\sigma(n)$ is the DIVISOR FUNCTION. An example is $(2^2 3^2 5 \cdot 11,\ 2^5 3^2 7,\ 2^2 3^2 71)$.

see also AMICABLE PAIR, AMICABLE QUADRUPLE

References
Dickson, L. E. "Amicable Number Triples." *Amer. Math. Monthly* **20**, 84–92, 1913.
Dickson, L. E. *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, p. 50, 1952.
Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 59, 1994.
Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, p. 156, 1979.
Mason, T. E. "On Amicable Numbers and Their Generalizations." *Amer. Math. Monthly* **28**, 195–200, 1921.
Weisstein, E. W. "Sociable and Amicable Numbers." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Sociable.m`.

Amortization

The payment of a debt plus accrued INTEREST by regular payments.

Ampersand Curve

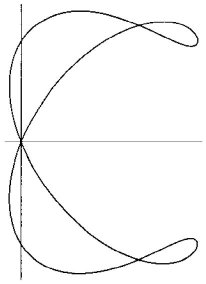

The PLANE CURVE with Cartesian equation

$$(y^2 - x^2)(x - 1)(2x - 3) = 4(x^2 + y^2 - 2x)^2.$$

References
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 72, 1989.

Amphichiral

An object is amphichiral (also called REFLEXIBLE) if it is superposable with its MIRROR IMAGE (i.e., its image in a plane mirror).

see also AMPHICHIRAL KNOT, CHIRAL, DISYMMETRIC, HANDEDNESS, MIRROR IMAGE

Amphichiral Knot

An amphichiral knot is a KNOT which is capable of being continuously deformed into its own MIRROR IMAGE. The amphichiral knots having ten or fewer crossings are 04_{001} (FIGURE-OF-EIGHT KNOT), 06_{003}, 08_{003}, 08_{009}, 08_{012}, 08_{017}, 08_{018}, 10_{017}, 10_{033}, 10_{037}, 10_{043}, 10_{045}, 10_{079}, 10_{081}, 10_{088}, 10_{099}, 10_{109}, 10_{115}, 10_{118}, and 10_{123} (Jones 1985). The HOMFLY POLYNOMIAL is good at identifying amphichiral knots, but sometimes fails to identify knots which are not. No complete invariant (an invariant which always definitively determines if a KNOT is AMPHICHIRAL) is known.

Let b_+ be the SUM of POSITIVE exponents, and b_- the SUM of NEGATIVE exponents in the BRAID GROUP B_n. If

$$b_+ - 3b_- - n + 1 > 0,$$

then the KNOT corresponding to the closed BRAID b is not amphichiral (Jones 1985).

see also AMPHICHIRAL, BRAID GROUP, INVERTIBLE KNOT, MIRROR IMAGE

References
Burde, G. and Zieschang, H. *Knots.* Berlin: de Gruyter, pp. 311–319, 1985.
Jones, V. "A Polynomial Invariant for Knots via von Neumann Algebras." *Bull. Amer. Math. Soc.* **12**, 103–111, 1985.
Jones, V. "Hecke Algebra Representations of Braid Groups and Link Polynomials." *Ann. Math.* **126**, 335–388, 1987.

Amplitude

The variable ϕ used in ELLIPTIC FUNCTIONS and ELLIPTIC INTEGRALS, which can be defined by

$$\phi = \operatorname{am} u \equiv \int \operatorname{dn} u \, du,$$

where $\operatorname{dn}(u)$ is a JACOBI ELLIPTIC FUNCTION. The term "amplitude" is also used to refer to the maximum offset of a function from its baseline level.

see also ARGUMENT (ELLIPTIC INTEGRAL), CHARACTERISTIC (ELLIPTIC INTEGRAL), DELTA AMPLITUDE, ELLIPTIC FUNCTION, ELLIPTIC INTEGRAL, JACOBI ELLIPTIC FUNCTIONS, MODULAR ANGLE, MODULUS (ELLIPTIC INTEGRAL), NOME, PARAMETER

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 590, 1972.
Fischer, G. (Ed.). Plate 132 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 129, 1986.

Anallagmatic Curve

A curve which is invariant under INVERSION. Examples include the CARDIOID, CARTESIAN OVALS, CASSINI OVALS, LIMAÇON, STROPHOID, and MACLAURIN TRISECTRIX.

Anallagmatic Pavement

see HADAMARD MATRIX

Analogy

Inference of the TRUTH of an unknown result obtained by noting its similarity to a result already known to be TRUE. In the hands of a skilled mathematician, analogy can be a very powerful tool for suggesting new and extending old results. However, subtleties can render results obtained by analogy incorrect, so rigorous PROOF is still needed.

see also INDUCTION

Analysis

The study of how continuous mathematical structures (FUNCTIONS) vary around the NEIGHBORHOOD of a point on a SURFACE. Analysis includes CALCULUS, DIFFERENTIAL EQUATIONS, etc.

see also ANALYSIS SITUS, CALCULUS, COMPLEX ANALYSIS, FUNCTIONAL ANALYSIS, NONSTANDARD ANALYSIS, REAL ANALYSIS

References

Bottazzini, U. *The "Higher Calculus": A History of Real and Complex Analysis from Euler to Weierstraß.* New York: Springer-Verlag, 1986.

Bressoud, D. M. *A Radical Approach to Real Analysis.* Washington, DC: Math. Assoc. Amer., 1994.

Ehrlich, P. *Real Numbers, Generalization of the Reals, & Theories of Continua.* Norwell, MA: Kluwer, 1994.

Hairer, E. and Wanner, G. *Analysis by Its History.* New York: Springer-Verlag, 1996.

Royden, H. L. *Real Analysis, 3rd ed.* New York: Macmillan, 1988.

Wheeden, R. L. and Zygmund, A. *Measure and Integral: An Introduction to Real Analysis.* New York: Dekker, 1977.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Analysis Situs

An archaic name for TOPOLOGY.

Analytic Continuation

A process of extending the region in which a COMPLEX FUNCTION is defined.

see also MONODROMY THEOREM, PERMANENCE OF ALGEBRAIC FORM, PERMANENCE OF MATHEMATICAL RELATIONS PRINCIPLE

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 378–380, 1985.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 389–390 and 392–398, 1953.

Analytic Function

A FUNCTION in the COMPLEX NUMBERS $\mathbb{C}$ is analytic on a region R if it is COMPLEX DIFFERENTIABLE at every point in R. The terms HOLOMORPHIC FUNCTION and REGULAR FUNCTION are sometimes used interchangeably with "analytic function." If a FUNCTION is analytic, it is infinitely DIFFERENTIABLE.

see also BERGMAN SPACE, COMPLEX DIFFERENTIABLE, DIFFERENTIABLE, PSEUDOANALYTIC FUNCTION, SEMIANALYTIC, SUBANALYTIC

References

Morse, P. M. and Feshbach, H. "Analytic Functions." §4.2 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 356–374, 1953.

Analytic Geometry

The study of the GEOMETRY of figures by algebraic representation and manipulation of equations describing their positions, configurations, and separations. Analytic geometry is also called COORDINATE GEOMETRY since the objects are described as n-tuples of points (where $n = 2$ in the PLANE and 3 in SPACE) in some COORDINATE SYSTEM.

see also ARGAND DIAGRAM, CARTESIAN COORDINATES, COMPLEX PLANE, GEOMETRY, PLANE, QUADRANT, SPACE, x-AXIS, y-AXIS, z-AXIS

References

Courant, R. and Robbins, H. "Remarks on Analytic Geometry." §2.3 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 72–77, 1996.

Analytic Set

A DEFINABLE SET, also called a SOUSLIN SET.

see also COANALYTIC SET, SOUSLIN SET

Anarboricity

Given a GRAPH G, the anarboricity is the maximum number of line-disjoint nonacyclic SUBGRAPHS whose UNION is G.

see also ARBORICITY

Anchor

An anchor is the BUNDLE MAP ρ from a VECTOR BUNDLE A to the TANGENT BUNDLE TB satisfying

1. $[\rho(X), \rho(Y)] = \rho([X, Y])$ and
2. $[X, \phi Y] = \phi[X, Y] + (\rho(X) \cdot \phi)Y$,

where X and Y are smooth sections of A, ϕ is a smooth function of B, and the bracket is the "Jacobi-Lie bracket" of a VECTOR FIELD.

see also LIE ALGEBROID

References

Weinstein, A. "Groupoids: Unifying Internal and External Symmetry." *Not. Amer. Math. Soc.* **43**, 744–752, 1996.

Anchor Ring

An archaic name for the TORUS.

References
Eisenhart, L. P. *A Treatise on the Differential Geometry of Curves and Surfaces.* New York: Dover, p. 314, 1960.
Stacey, F. D. *Physics of the Earth, 2nd ed.* New York: Wiley, p. 239, 1977.
Whittaker, E. T. *A Treatise on the Analytical Dynamics of Particles & Rigid Bodies, 4th ed.* Cambridge, England: Cambridge University Press, p. 21, 1959.

And

A term (PREDICATE) in LOGIC which yields TRUE if one or more conditions are TRUE, and FALSE if any condition is FALSE. A AND B is denoted $A\&B$, $A \wedge B$, or simply AB. The BINARY AND operator has the following TRUTH TABLE:

A	B	$A \wedge B$
F	F	F
F	T	F
T	F	F
T	T	T

A PRODUCT of ANDs (the AND of n conditions) is called a CONJUNCTION, and is denoted

$$\bigwedge_{k=1}^{n} A_k.$$

Two binary numbers can have the operation AND performed bitwise with 1 representing TRUE and 0 FALSE. Some computer languages denote this operation on A, B, and C as `A&&B&&C` or `logand(A,B,C)`.

see also BINARY OPERATOR, INTERSECTION, NOT, OR, PREDICATE, TRUTH TABLE, XOR

Anderson-Darling Statistic

A statistic defined to improve the KOLMOGOROV-SMIRNOV TEST in the TAIL of a distribution.

see also KOLMOGOROV-SMIRNOV TEST, KUIPER STATISTIC

References
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 621, 1992.

André's Problem

The determination of the number of ALTERNATING PERMUTATIONS having elements $\{1, 2, \ldots, n\}$

see also ALTERNATING PERMUTATION

André's Reflection Method

A technique used by André (1887) to provide an elegant solution to the BALLOT PROBLEM (Hilton and Pederson 1991).

References
André, D. "Solution directe du problème résolu par M. Bertrand." *Comptes Rendus Acad. Sci. Paris* **105**, 436–437, 1887.
Comtet, L. *Advanced Combinatorics.* Dordrecht, Netherlands: Reidel, p. 22, 1974.
Hilton, P. and Pederson, J. "Catalan Numbers, Their Generalization, and Their Uses." *Math. Intel.* **13**, 64–75, 1991.
Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 185, 1991.

Andrew's Sine

The function

$$\psi(z) = \begin{cases} \sin\left(\frac{z}{c}\right) & |z| < c\pi \\ 0, & |z| > c\pi \end{cases}$$

which occurs in estimation theory.

see also SINE

References
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 697, 1992.

Andrews Cube

see SEMIPERFECT MAGIC CUBE

Andrews-Curtis Link

The LINK of 2-spheres in $\mathbb{R}^4$ obtained by SPINNING intertwined arcs. The link consists of a knotted 2-sphere and a SPUN TREFOIL KNOT.

see also SPUN KNOT, TREFOIL KNOT

References
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 94, 1976.

Andrews-Schur Identity

$$\begin{aligned} &\sum_{k=0}^{n} q^{k^2+ak} \begin{bmatrix} 2n-k+a \\ k \end{bmatrix} \\ &= \sum_{k=-\infty}^{\infty} q^{10k^2+(4a-1)k} \begin{bmatrix} 2n+2a+2 \\ n-5k \end{bmatrix} \frac{[10k+2a+2]}{[2n+2a+2]}, \end{aligned} \tag{1}$$

where $[x]$ is a GAUSSIAN POLYNOMIAL. It is a POLYNOMIAL identity for $a = 0, 1$ which implies the ROGERS-RAMANUJAN IDENTITIES by taking $n \to \infty$ and applying the JACOBI TRIPLE PRODUCT identity. A variant of this equation is

$$\sum_{k=-\lfloor a/2\rfloor}^{n} q^{k^2+2ak}\begin{bmatrix} n+k+a\\ n-k\end{bmatrix} = \sum_{-\lfloor(n+2a+2)/5\rfloor}^{\lfloor n/5\rfloor} q^{15k^2+(6a+1)k}\begin{bmatrix}2n+2a+2\\ 5-5k\end{bmatrix} \times\frac{[10k+2a+2]}{[2n+2a+2]}, \qquad (2)$$

where the symbol $\lfloor x\rfloor$ in the SUM limits is the FLOOR FUNCTION (Paule 1994). The RECIPROCAL of the identity is

$$\sum_{k=0}^{\infty}\frac{q^{k^2+2ak}}{(q;q)_{2k+a}} = \prod_{j=0}^{\infty}\frac{1}{(1-q^{2j+1})(1-q^{20j+4a+4})(1-q^{20j-4a+16})} \qquad (3)$$

for $a = 0, 1$ (Paule 1994). For $q = 1$, (1) and (2) become

$$\sum_{-\lfloor a/2\rfloor}^{n}\binom{n+k+a}{n-k} = \sum_{-\lfloor(n+2a+2)/5\rfloor}^{\lfloor n/5\rfloor}\binom{2n+2a+2}{n-5k}\frac{5k+a+1}{n+a+1}. \qquad (4)$$

References

Andrews, G. E. "A Polynomial Identity which Implies the Rogers-Ramanujan Identities." *Scripta Math.* **28**, 297–305, 1970.

Paule, P. "Short and Easy Computer Proofs of the Rogers-Ramanujan Identities and of Identities of Similar Type." *Electronic J. Combinatorics* **1**, R10, 1–9, 1994. `http://www.combinatorics.org/Volume_1/volume1.html#R10`.

Andrica's Conjecture

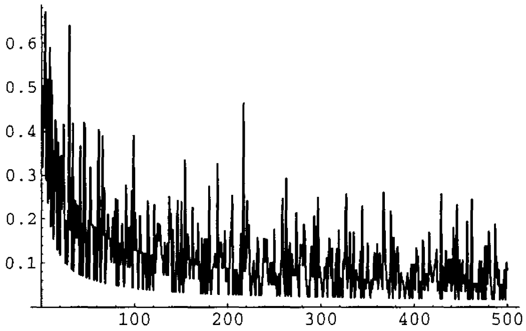

Andrica's conjecture states that, for p_n the nth PRIME NUMBER, the INEQUALITY

$$A_n \equiv \sqrt{p_{n+1}} - \sqrt{p_n} < 1$$

holds, where the discrete function A_n is plotted above. The largest value among the first 1000 PRIMES is for $n = 4$, giving $\sqrt{11} - \sqrt{7} \approx 0.670873$. Since the Andrica function falls asymptotically as n increases so a PRIME GAP of increasing size is needed at large n, it seems likely the CONJECTURE is true. However, it has not yet been proven.

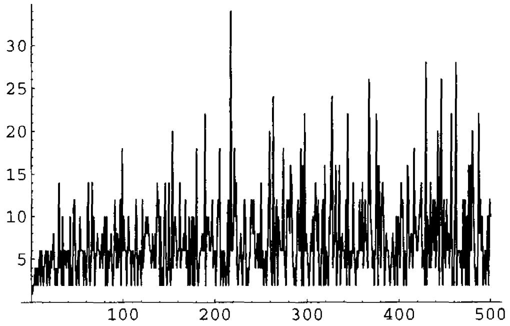

A_n bears a strong resemblance to the PRIME DIFFERENCE FUNCTION, plotted above, the first few values of which are 1, 2, 2, 4, 2, 4, 2, 4, 6, 2, 6, ... (Sloane's A001223).

see also BROCARD'S CONJECTURE, GOOD PRIME, FORTUNATE PRIME, PÓLYA CONJECTURE, PRIME DIFFERENCE FUNCTION, TWIN PEAKS

References

Golomb, S. W. "Problem E2506: Limits of Differences of Square Roots." *Amer. Math. Monthly* **83**, 60–61, 1976.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 21, 1994.

Rivera, C. "Problems & Puzzles (Conjectures): Andrica's Conjecture." `http://www.sci.net.mx/~crivera/ppp/conj_008.htm`.

Sloane, N. J. A. Sequence A001223/M0296 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Anger Function

A generalization of the BESSEL FUNCTION OF THE FIRST KIND defined by

$$\mathcal{J}_\nu(z) \equiv \frac{1}{\pi}\int_0^\pi \cos(\nu\theta - z\sin\theta)\,d\theta.$$

If ν is an INTEGER n, then $\mathcal{J}_n(z) = J_n(z)$, where $J_n(z)$ is a BESSEL FUNCTION OF THE FIRST KIND. Anger's original function had an upper limit of 2π, but the current NOTATION was standardized by Watson (1966).

see also BESSEL FUNCTION, MODIFIED STRUVE FUNCTION, PARABOLIC CYLINDER FUNCTION, STRUVE FUNCTION, WEBER FUNCTIONS

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Anger and Weber Functions." §12.3 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 498–499, 1972.

Watson, G. N. *A Treatise on the Theory of Bessel Functions, 2nd ed.* Cambridge, England: Cambridge University Press, 1966.

Angle

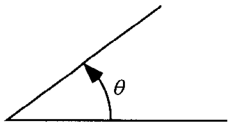

Given two intersecting Lines or Line Segments, the amount of Rotation about the point of intersection (the Vertex) required to bring one into correspondence with the other is called the angle θ between them. Angles are usually measured in Degrees (denoted °), Radians (denoted rad, or without a unit), or sometimes Gradians (denoted grad).

One full rotation in these three measures corresponds to 360°, 2π rad, or 400 grad. Half a full Rotation is called a Straight Angle, and a Quarter of a full rotation is called a Right Angle. An angle less than a Right Angle is called an Acute Angle, and an angle greater than a Right Angle is called an Obtuse Angle.

The use of Degrees to measure angles harks back to the Babylonians, whose Sexagesimal number system was based on the number 60. 360° likely arises from the Babylonian year, which was composed of 360 days (12 months of 30 days each). The Degree is further divided into 60 Arc Minutes, and an Arc Minute into 60 Arc Seconds. A more natural measure of an angle is the Radian. It has the property that the Arc Length around a Circle is simply given by the radian angle measure times the Circle Radius. The Radian is also the most useful angle measure in Calculus because the Derivative of Trigonometric functions such as

$$\frac{d}{dx}\sin x = \cos x$$

does not require the insertion of multiplicative constants like $\pi/180$. Gradians are sometimes used in surveying (they have the nice property that a Right Angle is exactly 100 Gradians), but are encountered infrequently, if at all, in mathematics.

The concept of an angle can be generalized from the Circle to the Sphere. The fraction of a Sphere subtended by an object is measured in Steradians, with the entire Sphere corresponding to 4π Steradians.

A ruled Semicircle used for measuring and drawing angles is called a Protractor. A Compass can also be used to draw circular Arcs of some angular extent.

see also Acute Angle, Arc Minute, Arc Second, Central Angle, Complementary Angle, Degree, Dihedral Angle, Directed Angle, Euler Angles, Gradian, Horn Angle, Inscribed Angle, Oblique Angle, Obtuse Angle, Perigon, Protractor, Radian, Right Angle, Solid Angle, Steradian, Straight Angle, Subtend, Supplementary Angle, Vertex Angle

References
Dixon, R. *Mathographics.* New York: Dover, pp. 99–100, 1991.

Angle Bisector

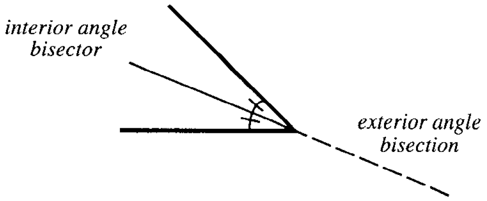

The (interior) bisector of an Angle is the Line or Line Segment which cuts it into two equal Angles on the same "side" as the Angle.

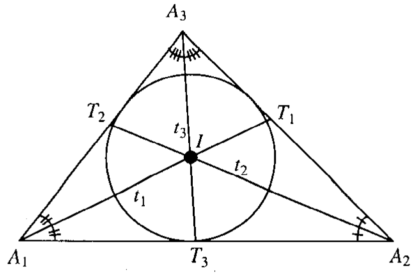

The length of the bisector of Angle A_1 in the above Triangle $\Delta A_1A_2A_3$ is given by

$$t_1{}^2 = a_2a_3\left[1 - \frac{a_1{}^2}{(a_2+a_3)^2}\right],$$

where $t_i \equiv \overline{A_iT_i}$ and $a_i \equiv \overline{A_jA_k}$. The angle bisectors meet at the Incenter I, which has Trilinear Coordinates 1:1:1.

see also Angle Bisector Theorem, Cyclic Quadrangle, Exterior Angle Bisector, Isodynamic Points, Orthocentric System, Steiner-Lehmus Theorem, Trisection

References
Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 9–10, 1967.
Dixon, R. *Mathographics.* New York: Dover, p. 19, 1991.
Mackay, J. S. "Properties Concerned with the Angular Bisectors of a Triangle." *Proc. Edinburgh Math. Soc.* **13**, 37–102, 1895.

Angle Bisector Theorem

The Angle Bisector of an Angle in a Triangle divides the opposite side in the same Ratio as the sides adjacent to the Angle.

Angle Bracket

The combination of a Bra and Ket (bra+ket = bracket) which represents the Inner Product of two functions or vectors,

$$\langle f|g\rangle = \int f(x)g(x)\,dx$$

$$\langle \mathbf{v}|\mathbf{w}\rangle = \mathbf{v}\cdot\mathbf{w}.$$

By itself, the Bra is a Covariant 1-Vector, and the Ket is a Covariant One-Form. These terms are commonly used in quantum mechanics.

see also Bra, Differential k-Form, Ket, One-Form

Angle of Parallelism

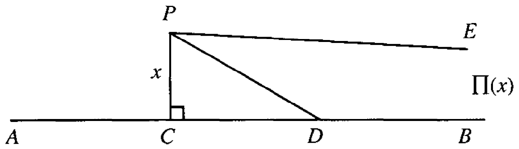

Given a point P and a LINE AB, draw the PERPENDICULAR through P and call it PC. Let PD be any other line from P which meets CB in D. In a HYPERBOLIC GEOMETRY, as D moves off to infinity along CB, then the line PD approaches the limiting line PE, which is said to be parallel to CB at P. The angle $\angle CPE$ which PE makes with PC is then called the angle of parallelism for perpendicular distance x, and is given by

$$\Pi(x) = 2\tan^{-1}(e^{-x}).$$

This is known as LOBACHEVSKY'S FORMULA.

see also HYPERBOLIC GEOMETRY, LOBACHEVSKY'S FORMULA

References

Manning, H. P. *Introductory Non-Euclidean Geometry.* New York: Dover, pp. 31–32 and 58, 1963.

Angle Trisection

see TRISECTION

Angular Acceleration

The angular acceleration $\boldsymbol{\alpha}$ is defined as the time DERIVATIVE of the ANGULAR VELOCITY $\boldsymbol{\omega}$,

$$\boldsymbol{\alpha} \equiv \frac{d\boldsymbol{\omega}}{dt} = \frac{d^2\theta}{dt^2}\hat{\mathbf{z}} = \frac{\mathbf{a}}{r}.$$

see also ACCELERATION, ANGULAR DISTANCE, ANGULAR VELOCITY

Angular Defect

The DIFFERENCE between the SUM of face ANGLES A_i at a VERTEX of a POLYHEDRON and 2π,

$$\delta = 2\pi - \sum_i A_i.$$

see also DESCARTES TOTAL ANGULAR DEFECT, JUMP ANGLE

Angular Distance

The angular distance traveled around a CIRCLE is the number of RADIANS the path subtends,

$$\theta \equiv \frac{\ell}{2\pi r}2\pi = \frac{\ell}{r}.$$

see also ANGULAR ACCELERATION, ANGULAR VELOCITY

Angular Velocity

The angular velocity $\boldsymbol{\omega}$ is the time DERIVATIVE of the ANGULAR DISTANCE θ with direction $\hat{\mathbf{z}}$ PERPENDICULAR to the plane of angular motion,

$$\boldsymbol{\omega} \equiv \frac{d\theta}{dt}\hat{\mathbf{z}} = \frac{\mathbf{v}}{r}.$$

see also ANGULAR ACCELERATION, ANGULAR DISTANCE

Anharmonic Ratio

see CROSS-RATIO

Anisohedral Tiling

A k-anisohedral tiling is a tiling which permits no n-ISOHEDRAL TILING with $n < k$.

References

Berglund, J. "Is There a k-Anisohedral Tile for $k \geq 5$?" *Amer. Math. Monthly* **100**, 585-588, 1993.
Klee, V. and Wagon, S. *Old and New Unsolved Problems in Plane Geometry and Number Theory.* Washington, DC: Math. Assoc. Amer., 1991.

Annihilator

The term annihilator is used in several different ways in various aspects of mathematics. It is most commonly used to mean the SET of all functions satisfying a given set of conditions which is zero on every member of a given SET.

Annulus

The region in common to two concentric CIRCLES of RADII a and b. The AREA of an annulus is

$$A_{\text{annulus}} = \pi(b^2 - a^2).$$

An interesting identity is as follows. In the figure,

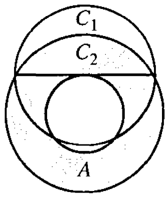

the AREA of the shaded region A is given by

$$A = C_1 + C_2.$$

see also CHORD, CIRCLE, CONCENTRIC CIRCLES, LUNE (PLANE), SPHERICAL SHELL

References

Pappas, T. "The Amazing Trick." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 69, 1989.

Annulus Conjecture

see ANNULUS THEOREM

Annulus Theorem

Let K_1^n and K_2^n be disjoint bicollared knots in $\mathbb{R}^{n+1}$ or $\mathbb{S}^{n+1}$ and let U denote the open region between them. Then the closure of U is a closed annulus $\mathbb{S}^n \times [0,1]$. Except for the case $n = 3$, the theorem was proved by Kirby (1969).

References

Kirby, R. C. "Stable Homeomorphisms and the Annulus Conjecture." *Ann. Math.* **89**, 575–582, 1969.

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 38, 1976.

Anomalous Cancellation

The simplification of a FRACTION a/b which gives a correct answer by "canceling" DIGITS of a and b. There are only four such cases for NUMERATOR and DENOMINATORS of two DIGITS in base 10: $64/16 = 4/1 = 4$, $98/49 = 8/4 = 2$, $95/19 = 5/1 = 5$, and $65/26 = 5/2$ (Boas 1979).

The concept of anomalous cancellation can be extended to arbitrary bases. PRIME bases have no solutions, but there is a solution corresponding to each PROPER DIVISOR of a COMPOSITE b. When $b - 1$ is PRIME, this type of solution is the only one. For base 4, for example, the only solution is $32_4/13_4 = 2_4$. Boas gives a table of solutions for $b \leq 39$. The number of solutions is EVEN unless b is an EVEN SQUARE.

b	N	b	N
4	1	26	4
6	2	27	6
8	2	28	10
9	2	30	6
10	4	32	4
12	4	34	6
14	2	35	6
15	6	36	21
16	7	38	2
18	4	39	6
20	4		
21	10		
22	6		
24	6		

see also FRACTION, PRINTER'S ERRORS, REDUCED FRACTION

References

Boas, R. P. "Anomalous Cancellation." Ch. 6 in *Mathematical Plums* (Ed. R. Honsberger). Washington, DC: Math. Assoc. Amer., pp. 113–129, 1979.

Ogilvy, C. S. and Anderson, J. T. *Excursions in Number Theory.* New York: Dover, pp. 86–87, 1988.

Anomalous Number

see BENFORD'S LAW

Anonymous

A term in SOCIAL CHOICE THEORY meaning invariance of a result under permutation of voters.

see also DUAL VOTING, MONOTONIC VOTING

Anosov Automorphism

A HYPERBOLIC linear map $\mathbb{R}^n \to \mathbb{R}^n$ with INTEGER entries in the transformation MATRIX and DETERMINANT ± 1 is an ANOSOV DIFFEOMORPHISM of the n-TORUS, called an Anosov automorphism (or HYPERBOLIC AUTOMORPHISM). Here, the term automorphism is used in the GROUP THEORY sense.

Anosov Diffeomorphism

An Anosov diffeomorphism is a C^1 DIFFEOMORPHISM ϕ such that the MANIFOLD M is HYPERBOLIC with respect to ϕ. Very few classes of Anosov diffeomorphisms are known. The best known is ARNOLD'S CAT MAP.

A HYPERBOLIC linear map $\mathbb{R}^n \to \mathbb{R}^n$ with INTEGER entries in the transformation MATRIX and DETERMINANT ± 1 is an Anosov diffeomorphism of the n-TORUS. Not every MANIFOLD admits an Anosov diffeomorphism. Anosov diffeomorphisms are EXPANSIVE, and there are no Anosov diffeomorphisms on the CIRCLE.

It is conjectured that if $\phi : M \to M$ is an Anosov diffeomorphism on a COMPACT RIEMANNIAN MANIFOLD and the NONWANDERING SET $\Omega(\phi)$ of ϕ is M, then ϕ is TOPOLOGICALLY CONJUGATE to a FINITE-TO-ONE FACTOR of an ANOSOV AUTOMORPHISM of a NILMANIFOLD. It has been proved that any Anosov diffeomorphism on the n-TORUS is TOPOLOGICALLY CONJUGATE to an ANOSOV AUTOMORPHISM, and also that Anosov diffeomorphisms are C^1 STRUCTURALLY STABLE.

see also ANOSOV AUTOMORPHISM, AXIOM A DIFFEOMORPHISM, DYNAMICAL SYSTEM

References

Anosov, D. V. "Geodesic Flow on Closed Riemannian Manifolds with Negative Curvature." *Proc. Steklov Inst., A. M. S.* 1969.

Smale, S. "Differentiable Dynamical Systems." *Bull. Amer. Math. Soc.* **73**, 747–817, 1967.

Anosov Flow

A FLOW defined analogously to the ANOSOV DIFFEOMORPHISM, except that instead of splitting the TANGENT BUNDLE into two invariant sub-BUNDLES, they are split into three (one exponentially contracting, one expanding, and one which is 1-dimensional and tangential to the flow direction).

see also DYNAMICAL SYSTEM

Anosov Map

An important example of a ANOSOV DIFFEOMORPHISM.

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix},$$

where x_{n+1}, y_{n+1} are computed mod 1.

see also ARNOLD'S CAT MAP

ANOVA

"Analysis of Variance." A STATISTICAL TEST for heterogeneity of MEANS by analysis of group VARIANCES. To apply the test, assume random sampling of a variate y with equal VARIANCES, independent errors, and a NORMAL DISTRIBUTION. Let n be the number of REPLICATES (sets of identical observations) within each of K FACTOR LEVELS (treatment groups), and y_{ij} be the jth observation within FACTOR LEVEL i. Also assume that the ANOVA is "balanced" by restricting n to be the same for each FACTOR LEVEL.

Now define the sum of square terms

$$\begin{aligned} \mathrm{SST} &\equiv \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{\bar{y}})^2 && (1) \\ &= \sum_{i=1}^{k} \sum_{j=1}^{n} {y_{ij}}^2 - \frac{\left(\sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij}\right)^2}{Kn} && (2) \\ \mathrm{SSA} &\equiv \frac{1}{n} \sum_{i=1}^{k} \left(\sum_{j=1}^{n} y_{ij}\right)^2 - \frac{1}{Kn} \left(\sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij}\right)^2 && (3) \\ \mathrm{SSE} &\equiv \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2 && (4) \\ &= \mathrm{SST} - \mathrm{SSA}, && (5) \end{aligned}$$

which are the total, treatment, and error sums of squares. Here, $\bar{y}_i$ is the mean of observations within FACTOR LEVEL i, and $\bar{\bar{y}}$ is the "group" mean (i.e., mean of means). Compute the entries in the following table, obtaining the P-VALUE corresponding to the calculated F-RATIO of the mean squared values

$$F = \frac{\mathrm{MSA}}{\mathrm{MSE}}. \qquad (6)$$

Category	SS	°Freedom	Mean Squared	F-Ratio
treatment	SSA	$K-1$	$\mathrm{MSA} \equiv \frac{\mathrm{SSA}}{K-1}$	$\frac{\mathrm{MSA}}{\mathrm{MSE}}$
error	SSE	$K(n-1)$	$\mathrm{MSE} \equiv \frac{\mathrm{SSE}}{K(n-1)}$	
total	SST	$Kn-1$	$\mathrm{MST} \equiv \frac{\mathrm{SST}}{Kn-1}$	

If the P-VALUE is small, reject the NULL HYPOTHESIS that all MEANS are the same for the different groups.

see also FACTOR LEVEL, REPLICATE, VARIANCE

Anthropomorphic Polygon

A SIMPLE POLYGON with precisely two EARS and one MOUTH.

References

Toussaint, G. "Anthropomorphic Polygons." *Amer. Math. Monthly* **122**, 31–35, 1991.

Anthyphairetic Ratio

An archaic word for a CONTINUED FRACTION.

References

Fowler, D. H. *The Mathematics of Plato's Academy: A New Reconstruction.* New York: Oxford University Press, 1987.

Antiautomorphism

If a MAP $f : G \to G'$ from a GROUP G to a GROUP G' satisfies $f(ab) = f(a)f(b)$ for all $a, b \in G$, then f is said to be an antiautomorphism.

see also AUTOMORPHISM

Anticevian Triangle

Given a center $\alpha : \beta : \gamma$, the anticevian triangle is defined as the TRIANGLE with VERTICES $-\alpha : \beta : \gamma$, $\alpha : -\beta : \gamma$, and $\alpha : \beta : -\gamma$. If $A'B'C'$ is the CEVIAN TRIANGLE of X and $A''B''C''$ is an anticevian triangle, then X and A'' are HARMONIC CONJUGATE POINTS with respect to A and A'.

see also CEVIAN TRIANGLE

References

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Antichain

Let P be a finite PARTIALLY ORDERED SET. An antichain in P is a set of pairwise incomparable elements (a family of SUBSETS such that, for any two members, one is not the SUBSET of another). The WIDTH of P is the maximum CARDINALITY of an ANTICHAIN in P. For a PARTIAL ORDER, the size of the longest ANTICHAIN is called the WIDTH.

see also CHAIN, DILWORTH'S LEMMA, PARTIALLY ORDERED SET, WIDTH (PARTIAL ORDER)

References

Sloane, N. J. A. Sequence A006826/M2469 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Anticlastic

When the GAUSSIAN CURVATURE K is everywhere NEGATIVE, a SURFACE is called anticlastic and is saddle-shaped. A SURFACE on which K is everywhere POSITIVE is called SYNCLASTIC. A point at which the GAUSSIAN CURVATURE is NEGATIVE is called a HYPERBOLIC POINT.

see also ELLIPTIC POINT, GAUSSIAN QUADRATURE, HYPERBOLIC POINT, PARABOLIC POINT, PLANAR POINT, SYNCLASTIC

Anticommutative

An OPERATOR $*$ for which $a * b = -b * a$ is said to be anticommutative.

see also COMMUTATIVE

Anticommutator

For OPERATORS $\tilde{A}$ and $\tilde{B}$, the anticommutator is defined by

$$\{\tilde{A}, \tilde{B}\} \equiv \tilde{A}\tilde{B} + \tilde{B}\tilde{A}.$$

see also COMMUTATOR, JORDAN ALGEBRA

Anticomplementary Triangle

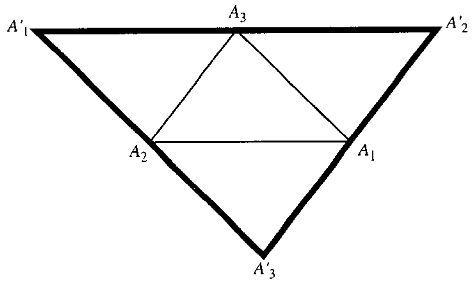

A TRIANGLE $\Delta A'B'C'$ which has a given TRIANGLE ΔABC as its MEDIAL TRIANGLE. The TRILINEAR COORDINATES of the anticomplementary triangle are

$$\begin{aligned} A' &= -a^{-1} : b^{-1} : c^{-1} \\ B' &= a^{-1} : -b^{-1} : c^{-1} \\ C' &= a^{-1} : b^{-1} : -c^{-1}. \end{aligned}$$

see also MEDIAL TRIANGLE

Antiderivative

see INTEGRAL

Antidifferentiation

see INTEGRATION

Antigonal Points

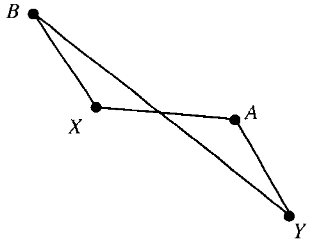

Given $\angle AXB + \angle AYB = \pi$ RADIANS in the above figure, then X and Y are said to be antigonal points with respect to A and B.

Antihomography

A CIRCLE-preserving TRANSFORMATION composed of an ODD number of INVERSIONS.

see also HOMOGRAPHY

Antihomologous Points

Two points which are COLLINEAR with respect to a SIMILITUDE CENTER but are not HOMOLOGOUS POINTS. Four interesting theorems from Johnson (1929) follow.

1. Two pairs of antihomologous points form inversely similar triangles with the HOMOTHETIC CENTER.
2. The PRODUCT of distances from a HOMOTHETIC CENTER to two antihomologous points is a constant.
3. Any two pairs of points which are antihomologous with respect to a SIMILITUDE CENTER lie on a CIRCLE.
4. The tangents to two CIRCLES at antihomologous points make equal ANGLES with the LINE through the points.

see also HOMOLOGOUS POINTS, HOMOTHETIC CENTER, SIMILITUDE CENTER

References
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 19–21, 1929.

Antilaplacian

The antilaplacian of u with respect to x is a function whose LAPLACIAN with respect to x equals u. The antilaplacian is never unique.

see also LAPLACIAN

Antilinear Operator

An antilinear OPERATOR satisfies the following two properties:

$$\begin{gathered} \tilde{A}[f_1(x) + f_2(x)] = \tilde{A}f_1(x) + \tilde{A}f_2(x) \\ \tilde{A}cf(x) = c^*\tilde{A}f(x), \end{gathered}$$

where c^* is the COMPLEX CONJUGATE of c.

see also LINEAR OPERATOR

Antilogarithm

The INVERSE FUNCTION of the LOGARITHM, defined such that

$$\log_b(\operatorname{antilog}_b z) = z = \operatorname{antilog}_b(\log_b z).$$

The antilogarithm in base b of z is therefore b^z.

see also COLOGARITHM, LOGARITHM, POWER

Antimagic Graph

A GRAPH with e EDGES labeled with distinct elements $\{1, 2, \ldots, e\}$ so that the SUM of the EDGE labels at each VERTEX differ.

see also MAGIC GRAPH

References
Hartsfield, N. and Ringel, G. *Pearls in Graph Theory: A Comprehensive Introduction.* San Diego, CA: Academic Press, 1990.

Antimagic Square

15	2	12	4
1	14	10	5
8	9	3	16
11	13	6	7

21	18	6	17	4
7	3	13	16	24
5	20	23	11	1
15	8	19	2	25
14	12	9	22	10

10	25	32	13	16	9
22	7	3	24	21	30
20	27	18	26	11	6
1	31	23	33	17	8
19	5	36	12	15	29
34	14	2	4	35	28

14	3	34	21	47	29	22
43	16	13	25	6	26	44
30	48	24	8	12	9	45
10	5	11	38	49	46	19
4	41	37	36	33	27	1
39	17	40	20	7	35	23
31	42	18	32	28	2	15

49	16	50	10	19	28	24	56
42	43	11	15	44	38	55	5
25	21	48	46	9	37	6	63
29	47	8	40	51	30	52	1
45	22	54	23	20	34	2	62
14	59	18	33	41	26	61	13
36	12	58	32	27	64	3	35
17	39	7	57	53	4	60	31

52	19	81	22	29	15	42	31	76
61	10	67	23	54	79	25	33	16
57	9	71	24	38	1	51	47	75
26	78	7	69	66	77	13	27	12
39	21	74	20	37	17	49	55	64
8	65	4	62	50	34	73	41	40
56	68	2	63	14	72	35	44	6
53	30	60	32	36	3	46	43	58
11	70	5	59	48	80	28	45	18

An antimagic square is an $n \times n$ ARRAY of integers from 1 to n^2 such that each row, column, and main diagonal produces a different sum such that these sums form a SEQUENCE of consecutive integers. It is therefore a special case of a HETEROSQUARE.

Antimagic squares of orders one and two are impossible, and it is believed that there are also no antimagic squares of order three. There are 18 families of antimagic squares of order four. Antimagic squares of orders 4–9 are illustrated above (Madachy 1979).

see also HETEROSQUARE, MAGIC SQUARE, TALISMAN SQUARE

References

Abe, G. "Unsolved Problems on Magic Squares." *Disc. Math.* **127**, 3–13, 1994.

Madachy, J. S. "Magic and Antimagic Squares." Ch. 4 in *Madachy's Mathematical Recreations.* New York: Dover, pp. 103–113, 1979.

Weisstein, E. W. "Magic Squares." `http://www.astro.virginia.edu/~eww6n/math/notebooks/MagicSquares.m`.

Antimorph

A number which can be represented both in the form ${x_0}^2 - D{y_0}^2$ and in the form $D{x_1}^2 - {y_1}^2$. This is only possible when the PELL EQUATION

$$x^2 - Dy^2 = -1$$

is solvable. Then

$$\begin{aligned} x^2 - Dy^2 &= -(x_0 - D{y_0}^2)({x_n}^2 - D{y_n}^2) \\ &= D(x_0 y_n - y_0 x_n)^2 - (x_0 x_n - D y_0 y_n)^2. \end{aligned}$$

see also IDONEAL NUMBER, POLYMORPH

References

Beiler, A. H. *Recreations in the Theory of Numbers: The Queen of Mathematical Entertains.* New York: Dover, 1964.

Antimorphic Number

see ANTIMORPH

Antinomy

A PARADOX or contradiction.

Antiparallel

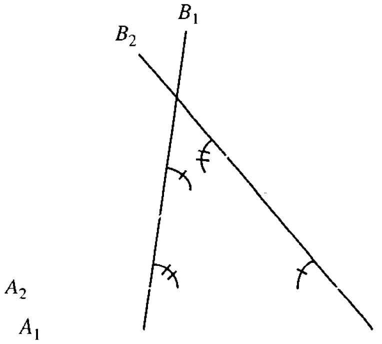

A pair of LINES B_1, B_2 which make the same ANGLES but in opposite order with two other given LINES A_1 and A_2, as in the above diagram, are said to be antiparallel to A_1 and A_2.

see also HYPERPARALLEL, PARALLEL

References

Phillips, A. W. and Fisher, I. *Elements of Geometry.* New York: American Book Co., 1896.

Antipedal Triangle

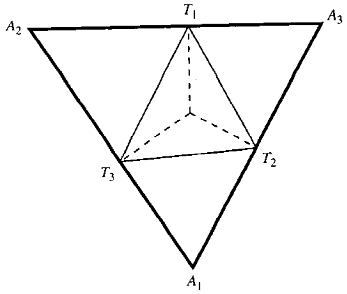

The antipedal triangle A of a given TRIANGLE T is the TRIANGLE of which T is the PEDAL TRIANGLE. For a TRIANGLE with TRILINEAR COORDINATES $\alpha : \beta : \gamma$ and ANGLES A, B, and C, the antipedal triangle has VERTICES with TRILINEAR COORDINATES

$$\begin{aligned} &-(\beta+\alpha\cos C)(\gamma+\alpha\cos B) : (\gamma+\alpha\cos B)(\alpha+\beta\cos C) : \\ &\qquad (\beta+\alpha\cos C)(\alpha+\gamma\cos B) \\ &(\gamma+\beta\cos A)(\beta+\alpha\cos C) : -(\gamma+\beta\cos A)(\alpha+\beta\cos C) : \\ &\qquad (\alpha+\beta\cos C)(\beta+\gamma\cos A) \\ &(\beta+\gamma\cos A)(\gamma+\alpha\cos B) : (\alpha+\gamma\cos B)(\gamma+\beta\cos A) : \\ &\qquad -(\alpha+\gamma\cos B)(\beta+\gamma\cos A). \end{aligned}$$

The ISOGONAL CONJUGATE of the ANTIPEDAL TRIANGLE of a given TRIANGLE is HOMOTHETIC with the original TRIANGLE. Furthermore, the PRODUCT of their AREAS equals the SQUARE of the AREA of the original TRIANGLE (Gallatly 1913).

see also PEDAL TRIANGLE

References
Gallatly, W. *The Modern Geometry of the Triangle, 2nd ed.* London: Hodgson, pp. 56–58, 1913.

Antipersistent Process

A FRACTAL PROCESS for which $H < 1/2$, so $r < 0$.

see also PERSISTENT PROCESS

Antipodal Map

The MAP which takes points on the surface of a SPHERE $\mathbb{S}^2$ to their ANTIPODAL POINTS.

Antipodal Points

Two points are antipodal (i.e., each is the ANTIPODE of the other) if they are diametrically opposite. Examples include endpoints of a LINE SEGMENT, or poles of a SPHERE. Given a point on a SPHERE with LATITUDE δ and LONGITUDE λ, the antipodal point has LATITUDE $-\delta$ and LONGITUDE $\lambda \pm 180°$ (where the sign is taken so that the result is between $-180°$ and $+180°$).

see also ANTIPODE, DIAMETER, GREAT CIRCLE, SPHERE

Antipode

Given a point A, the point B which is the ANTIPODAL POINT of A is said to be the antipode of A.

see also ANTIPODAL POINTS

Antiprism

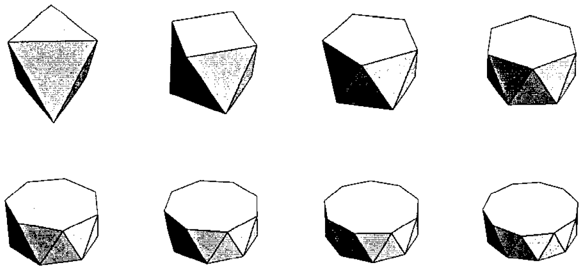

A SEMIREGULAR POLYHEDRON constructed with 2 n-gons and $2n$ TRIANGLES. The 3-antiprism is simply the OCTAHEDRON. The DUALS are the TRAPEZOHEDRA. The SURFACE AREA of a n-gonal antiprism is

$$\begin{aligned} S &= 2A_{n\text{-gon}} + 2nA_\Delta \\ &= 2\left[\tfrac{1}{4}na^2 \cot\left(\frac{\pi}{n}\right)\right] + 2n(\tfrac{1}{4}\sqrt{3}\,a^2) \\ &= \tfrac{1}{2}na^2\left[\cot\left(\frac{\pi}{n}\right) + \sqrt{3}\right]. \end{aligned}$$

see also OCTAHEDRON, PRISM, PRISMOID, TRAPEZOHEDRON

References
Ball, W. W. R. and Coxeter, H. S. M. "Polyhedra." Ch. 5 in *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 130, 1987.
Cromwell, P. R. *Polyhedra.* New York: Cambridge University Press, pp. 85–86, 1997.
Weisstein, E. W. "Prisms and Antiprisms." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Prism.m`.

Antiquity

see GEOMETRIC PROBLEMS OF ANTIQUITY

Antisnowflake

see KOCH ANTISNOWFLAKE

Antisquare Number

A number of the form $p^a \cdot A$ is said to be an antisquare if it fails to be a SQUARE NUMBER for the two reasons that a is ODD and A is a nonsquare modulo p.

see also SQUARE NUMBER

Antisymmetric

A quantity which changes SIGN when indices are reversed. For example, $A_{ij} \equiv a_i - a_j$ is antisymmetric since $A_{ij} = -A_{ji}$.

see also ANTISYMMETRIC MATRIX, ANTISYMMETRIC TENSOR, SYMMETRIC

Antisymmetric Matrix

An antisymmetric matrix is a MATRIX which satisfies the identity

$$\mathsf{A} = -\mathsf{A}^{\mathrm{T}} \tag{1}$$

where A^{T} is the MATRIX TRANSPOSE. In component notation, this becomes

$$a_{ij} = -a_{ji}. \tag{2}$$

Letting $k = i = j$, the requirement becomes

$$a_{kk} = -a_{kk}, \tag{3}$$

so an antisymmetric matrix must have zeros on its diagonal. The general 3×3 antisymmetric matrix is of the form

$$\begin{bmatrix} 0 & a_{12} & a_{13} \\ -a_{12} & 0 & a_{23} \\ -a_{13} & -a_{23} & 0 \end{bmatrix}. \tag{4}$$

Applying A^{-1} to both sides of the antisymmetry condition gives

$$-\mathsf{A}^{-1}\mathsf{A}^{\mathrm{T}} = \mathsf{I}. \tag{5}$$

Any SQUARE MATRIX can be expressed as the sum of symmetric and antisymmetric parts. Write

$$\mathsf{A} = \tfrac{1}{2}(\mathsf{A} + \mathsf{A}^{\mathrm{T}}) + \tfrac{1}{2}(\mathsf{A} - \mathsf{A}^{\mathrm{T}}). \tag{6}$$

But

$$\mathsf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \tag{7}$$

$$\mathsf{A}^{\mathrm{T}} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix}, \tag{8}$$

so

$$\mathsf{A}+\mathsf{A}^{\mathrm{T}} = \begin{bmatrix} 2a_{11} & a_{12}+a_{21} & \cdots & a_{1n}+a_{n1} \\ a_{12}+a_{21} & 2a_{22} & \cdots & a_{2n}+a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n}+a_{n1} & a_{2n}+a_{n2} & \cdots & 2a_{nn} \end{bmatrix}, \tag{9}$$

which is symmetric, and

$$\mathsf{A}-\mathsf{A}^{\mathrm{T}} = \begin{bmatrix} 0 & a_{12}-a_{21} & \cdots & a_{1n}-a_{n1} \\ -(a_{12}-a_{21}) & 0 & \cdots & a_{2n}-a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ -(a_{1n}-a_{n1}) & -(a_{2n}-a_{n2}) & \cdots & 0 \end{bmatrix}, \tag{10}$$

which is antisymmetric.

see also SKEW SYMMETRIC MATRIX, SYMMETRIC MATRIX

Antisymmetric Relation

A RELATION R on a SET S is antisymmetric provided that distinct elements are never both related to one another. In other words xRy and yRx together imply that $x = y$.

Antisymmetric Tensor

An antisymmetric tensor is defined as a TENSOR for which

$$A^{mn} = -A^{nm}. \tag{1}$$

Any TENSOR can be written as a sum of SYMMETRIC and antisymmetric parts as

$$A^{mn} = \tfrac{1}{2}(A^{mn}+A^{nm}) + \tfrac{1}{2}(A^{mn}-A^{nm}). \tag{2}$$

The antisymmetric part is sometimes denoted using the special notation

$$A^{[ab]} = \tfrac{1}{2}(A^{ab}-A^{ba}). \tag{3}$$

For a general TENSOR,

$$A^{[a_1\cdots a_n]} \equiv \frac{1}{n!}\epsilon_{a_1\cdots a_n} \sum_{\text{permutations}} A^{a_1\cdots a_n}, \tag{4}$$

where $\epsilon_{a_1\cdots a_n}$ is the LEVI-CIVITA SYMBOL, a.k.a. the PERMUTATION SYMBOL.

see also SYMMETRIC TENSOR

Antoine's Horned Sphere

A topological 2-sphere in 3-space whose exterior is not SIMPLY CONNECTED. The outer complement of Antoine's horned sphere is not SIMPLY CONNECTED. Furthermore, the group of the outer complement is not even finitely generated. Antoine's horned sphere is inequivalent to ALEXANDER'S HORNED SPHERE since the complement in $\mathbb{R}^3$ of the bad points for ALEXANDER'S HORNED SPHERE is SIMPLY CONNECTED.

see also ALEXANDER'S HORNED SPHERE

References

Alexander, J. W. "An Example of a Simply-Connected Surface Bounding a Region which is not Simply-Connected." *Proc. Nat. Acad. Sci.* **10**, 8–10, 1924.

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 76–79, 1976.

Antoine's Necklace

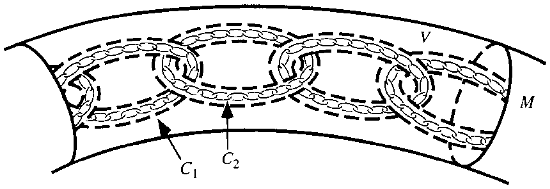

Construct a chain C of $2n$ components in a solid TORUS V. Now form a chain C_1 of $2n$ solid tori in V, where

$$\pi_1(V-C_1) \cong \pi_1(V-C)$$

via inclusion. In each component of C_1, construct a smaller chain of solid tori embedded in that component. Denote the union of these smaller solid tori C_2. Continue this process a countable number of times, then the intersection

$$A = \bigcap_{i=1}^{\infty} C_i$$

which is a nonempty compact SUBSET of $\mathbb{R}^3$ is called Antoine's necklace. Antoine's necklace is HOMEOMORPHIC with the CANTOR SET.

see also ALEXANDER'S HORNED SPHERE, NECKLACE

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 73–74, 1976.

Apeirogon

The REGULAR POLYGON essentially equivalent to the CIRCLE having an infinite number of sides and denoted with SCHLÄFLI SYMBOL $\{\infty\}$.

see also CIRCLE, REGULAR POLYGON

References

Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, 1973.

Schwartzman, S. *The Words of Mathematics: An Etymological Dictionary of Mathematical Terms Used in English.* Washington, DC: Math. Assoc. Amer., 1994.

Apéry's Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Apéry's constant is defined by

$$\zeta(3) = 1.2020569\ldots, \tag{1}$$

(Sloane's A002117) where $\zeta(z)$ is the RIEMANN ZETA FUNCTION. Apéry (1979) proved that $\zeta(3)$ is IRRATIONAL, although it is not known if it is TRANSCENDENTAL. The CONTINUED FRACTION for $\zeta(3)$ is [1, 4, 1, 18, 1, 1, 1, 4, 1, ...] (Sloane's A013631). The positions at which the numbers 1, 2, ... occur in the continued fraction are 1, 12, 25, 2, 64, 27, 17, 140, 10, ...

Sums related to $\zeta(3)$ are

$$\zeta(3) = \frac{5}{2}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n^3\binom{2n}{n}} = \frac{5}{2}\sum_{k=1}^{\infty}\frac{(-1)^{k+1}(k!)^2}{(2k)!k^3} \tag{2}$$

(used by Apéry), and

$$\lambda(3) = \sum_{k=0}^{\infty}\frac{1}{(2k+1)^3} = \tfrac{7}{8}\zeta(3) \tag{3}$$

$$\sum_{k=0}^{\infty}\frac{1}{(3k+1)^3} = \frac{2\pi^3}{81\sqrt{3}} + \tfrac{13}{27}\zeta(3) \tag{4}$$

$$\sum_{k=0}^{\infty}\frac{1}{(4k+1)^3} = \frac{\pi^3}{64} + \tfrac{7}{16}\zeta(3) \tag{5}$$

$$\sum_{k=0}^{\infty}\frac{1}{(6k+1)^3} = \frac{\pi^3}{36\sqrt{3}} + \tfrac{91}{216}\zeta(3), \tag{6}$$

where $\lambda(z)$ is the DIRICHLET LAMBDA FUNCTION. The above equations are special cases of a general result due to Ramanujan (Berndt 1985). Apéry's proof relied on showing that the sum

$$a(n) \equiv \sum_{k=0}^{n}\binom{n}{k}^2\binom{n+k}{k}^2, \tag{7}$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT, satisfies the RECURRENCE RELATION

$$(n+1)^3a(n+1) - (34n^3+51n^2+27n+5)a(n) + n^3a(n-1) = 0 \tag{8}$$

(van der Poorten 1979, Zeilberger 1991).

Apéry's constant is also given by

$$\zeta(3) = \sum_{n=1}^{\infty}\frac{S_{n,2}}{n!n}, \tag{9}$$

where $S_{n,m}$ is a STIRLING NUMBER OF THE FIRST KIND. This can be rewritten as

$$\sum_{n=1}^{\infty}\frac{H_n}{n^2} = 2\zeta(3), \tag{10}$$

where H_n is the nth HARMONIC NUMBER. Yet another expression for $\zeta(3)$ is

$$\zeta(3) = \frac{1}{2}\sum_{n=1}^{\infty}\frac{1}{n^2}\left(1+\frac{1}{2}+\ldots+\frac{1}{n}\right) \tag{11}$$

(Castellanos 1988).

INTEGRALS for $\zeta(3)$ include

$$\zeta(3) = \frac{1}{2}\int_0^{\infty}\frac{t^2}{e^t-1}\,dt \tag{12}$$

$$= \frac{8}{7}\left[\tfrac{1}{4}\pi^2\ln 2 + 2\int_0^{\pi/4}x\ln(\sin x)\,dx\right]. \tag{13}$$

Gosper (1990) gave

$$\zeta(3) = \frac{1}{4}\sum_{k=1}^{\infty}\frac{30k-11}{(2k-1)k^3\binom{2k}{k}^2}. \tag{14}$$

A CONTINUED FRACTION involving Apéry's constant is

$$\frac{6}{\zeta(3)} = 5 - \frac{1^6}{117-}\,\frac{2^6}{535-}\cdots\frac{n^6}{34n^3+51n^2+27n+5-}\cdots \tag{15}$$

(Apéry 1979, Le Lionnais 1983). Amdeberhan (1996) used WILF-ZEILBERGER PAIRS (F, G) with

$$F(n,k) = \frac{(-1)^k k!^2(sn-k-1)!}{(sn+k+1)!(k+1)}, \tag{16}$$

$s = 1$ to obtain

$$\zeta(3) = \frac{5}{2}\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{\binom{2n}{n}n^3}. \tag{17}$$

For $s = 2$,

$$\zeta(3) = \frac{1}{4}\sum_{n=1}^{\infty}(-1)^{n-1}\frac{56n^2-32+5}{(2n-1)^2}\,\frac{1}{\binom{3n}{n}\binom{2n}{n}n^3} \tag{18}$$

and for $s = 3$,

$$\zeta(3) = \sum_{n=0}^{\infty}\frac{(-1)^n}{72\binom{4n}{n}\binom{3n}{n}} \times\frac{6120n+5265n^4+13761n^2+13878n^3+1040}{(4n+1)(4n+3)(n+1)(3n+1)^2(3n+2)^2} \tag{19}$$

(Amdeberhan 1996). The corresponding $G(n,k)$ for $s = 1$ and 2 are

$$G(n,k) = \frac{2(-1)^k k!^2 (n-k)!}{(n+k+1)!(n+1)^2} \tag{20}$$

and

$$G(n,k) = \frac{(-1)^k k!^2 (2n-k)!(3+4n)(4n^2+6n+k+3)}{2(2n+k+2)!(n+1)^2(2n+1)^2}. \tag{21}$$

Gosper (1996) expressed $\zeta(3)$ as the MATRIX PRODUCT

$$\lim_{N\to\infty} \prod_{n=1}^{N} M_n = \begin{bmatrix} 0 & \zeta(3) \\ 0 & 1 \end{bmatrix}, \tag{22}$$

where

$$M_n \equiv \begin{bmatrix} \frac{(n+1)^4}{4096(n+\frac{5}{4})^2(n+\frac{7}{4})^2} & \frac{24570n^4+64161n^3+62152n^2+26427n+4154}{31104(n+\frac{1}{3})(n+\frac{1}{2})(n+\frac{2}{3})} \\ 0 & 1 \end{bmatrix} \tag{23}$$

which gives 12 bits per term. The first few terms are

$$M_1 = \begin{bmatrix} \frac{1}{19600} & \frac{2077}{1728} \\ 0 & 1 \end{bmatrix} \tag{24}$$

$$M_2 = \begin{bmatrix} \frac{1}{9801} & \frac{7561}{4320} \\ 0 & 1 \end{bmatrix} \tag{25}$$

$$M_3 = \begin{bmatrix} \frac{9}{67600} & \frac{50501}{20160} \\ 0 & 1 \end{bmatrix}, \tag{26}$$

which gives

$$\zeta(3) \approx \tfrac{423203577229}{352066176000} = 1.20205690315732\ldots. \tag{27}$$

Given three INTEGERS chosen at random, the probability that no common factor will divide them all is

$$[\zeta(3)]^{-1} \approx 1.202^{-1} = 0.832\ldots. \tag{28}$$

B. Haible and T. Papanikolaou computed $\zeta(3)$ to 1,000,000 DIGITS using a WILF-ZEILBERGER PAIR identity with

$$F(n,k) = (-1)^k \frac{n!^6(2n-k-1)!k!^3}{2(n+k+1)!^2(2n)!^3}, \tag{29}$$

$s = 1$, and $t = 1$, giving the rapidly converging

$$\zeta(3) = \sum_{n=0}^{\infty} (-1)^n \frac{n!^{10}(205n^2+250n+77)}{64(2n+1)!^5} \tag{30}$$

(Amdeberhan and Zeilberger 1997). The record as of Aug. 1998 was 64 million digits (Plouffe).

see also RIEMANN ZETA FUNCTION, WILF-ZEILBERGER PAIR

References

Amdeberhan, T. "Faster and Faster Convergent Series for $\zeta(3)$." *Electronic J. Combinatorics* **3**, R13, 1–2, 1996. `http://www.combinatorics.org/Volume_3/volume3.html#R13`.

Amdeberhan, T. and Zeilberger, D. "Hypergeometric Series Acceleration via the WZ Method." *Electronic J. Combinatorics* **4**, No. 2, R3, 1–3, 1997. `http://www.combinatorics.org/Volume_4/wilftoc.html#R03`. Also available at `http://www.math.temple.edu/~zeilberg/mamarim/mamarimhtml/accel.html`.

Apéry, R. "Irrationalité de $\zeta(2)$ et $\zeta(3)$." *Astérisque* **61**, 11–13, 1979.

Berndt, B. C. *Ramanujan's Notebooks: Part I.* New York: Springer-Verlag, 1985.

Beukers, F. "A Note on the Irrationality of $\zeta(3)$." *Bull. London Math. Soc.* **11**, 268–272, 1979.

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, 1987.

Castellanos, D. "The Ubiquitous Pi. Part I." *Math. Mag.* **61**, 67–98, 1988.

Conway, J. H. and Guy, R. K. "The Great Enigma." In *The Book of Numbers.* New York: Springer-Verlag, pp. 261–262, 1996.

Ewell, J. A. "A New Series Representation for $\zeta(3)$." *Amer. Math. Monthly* **97**, 219–220, 1990.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/apery/apery.html`.

Gosper, R. W. "Strip Mining in the Abandoned Orefields of Nineteenth Century Mathematics." In *Computers in Mathematics* (Ed. D. V. Chudnovsky and R. D. Jenks). New York: Marcel Dekker, 1990.

Haible, B. and Papanikolaou, T. "Fast Multiprecision Evaluation of Series of Rational Numbers." Technical Report TI-97-7. Darmstadt, Germany: Darmstadt University of Technology, Apr. 1997.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 36, 1983.

Plouffe, S. "Plouffe's Inverter: Table of Current Records for the Computation of Constants." `http://lacim.uqam.ca/pi/records.html`.

Plouffe, S. "32,000,279 Digits of Zeta(3)." `http://lacim.uqam.ca/piDATA/Zeta3.txt`.

Sloane, N. J. A. Sequences A013631 and A002117/M0020 in "An On-Line Version of the Encyclopedia of Integer Sequences."

van der Poorten, A. "A Proof that Euler Missed... Apéry's Proof of the Irrationality of $\zeta(3)$." *Math. Intel.* **1**, 196–203, 1979.

Zeilberger, D. "The Method of Creative Telescoping." *J. Symb. Comput.* **11**, 195–204, 1991.

Apoapsis

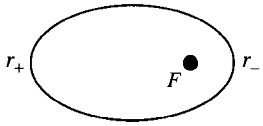

The greatest radial distance of an ELLIPSE as measured from a FOCUS. Taking $v = \pi$ in the equation of an ELLIPSE

$$r = \frac{a(1-e^2)}{1+e\cos v}$$

gives the apoapsis distance

$$r_+ = a(1+e).$$

Apoapsis for an orbit around the Earth is called apogee, and apoapsis for an orbit around the Sun is called aphelion.

see also ECCENTRICITY, ELLIPSE, FOCUS, PERIAPSIS

Apocalypse Number

A number having 666 DIGITS (where 666 is the BEAST NUMBER) is called an apocalypse number. The FIBONACCI NUMBER F_{3184} is an apocalypse number.

see also BEAST NUMBER, LEVIATHAN NUMBER

References

Pickover, C. A. *Keys to Infinity.* New York: Wiley, pp. 97–102, 1995.

Apocalyptic Number

A number of the form 2^n which contains the digits 666 (the BEAST NUMBER) is called an APOCALYPTIC NUMBER. 2^{157} is an apocalyptic number. The first few such powers are 157, 192, 218, 220, ... (Sloane's A007356).

see also APOCALYPSE NUMBER, LEVIATHAN NUMBER

References

Pickover, C. A. *Keys to Infinity.* New York: Wiley, pp. 97–102, 1995.

Sloane, N. J. A. Sequences A007356/M5405 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sloane, N. J. A. and Plouffe, S. Extended entry in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.

Apodization

The application of an APODIZATION FUNCTION.

Apodization Function

A function (also called a TAPERING FUNCTION) used to bring an interferogram smoothly down to zero at the edges of the sampled region. This suppresses sidelobes which would otherwise be produced, but at the expense of widening the lines and therefore decreasing the resolution.

The following are apodization functions for symmetrical (2-sided) interferograms, together with the INSTRUMENT FUNCTIONS (or APPARATUS FUNCTIONS) they produce and a blowup of the INSTRUMENT FUNCTION sidelobes. The INSTRUMENT FUNCTION $I(k)$ corresponding to a given apodization function $A(x)$ can be computed by taking the finite FOURIER COSINE TRANSFORM,

$$I(k) = \int_{-a}^{a} \cos(2\pi kx) A(x)\, dx. \qquad (1)$$

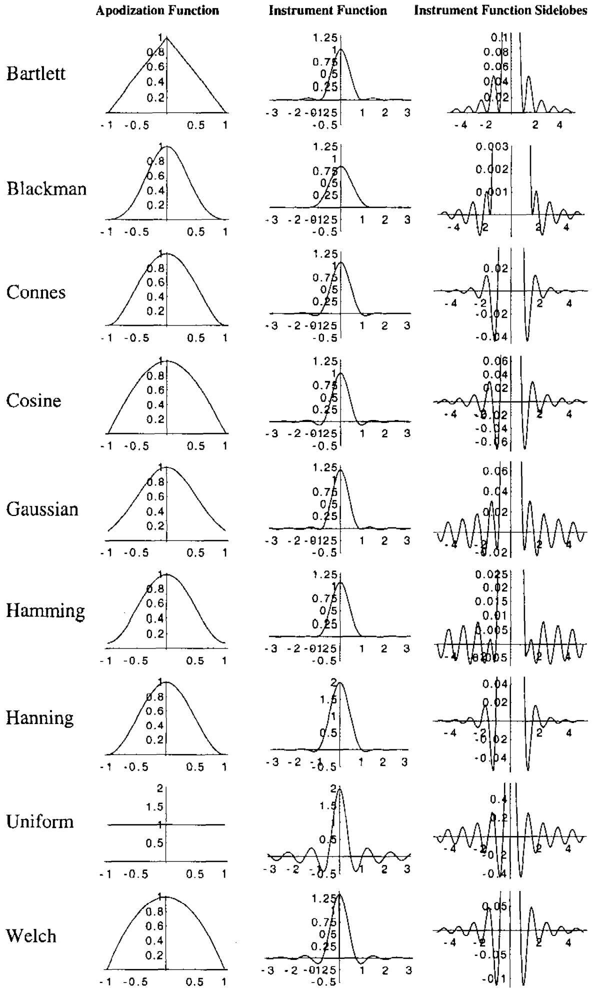

Type	Apodization Function	Instrument Function
Bartlett	$1-\frac{\lvert x\rvert}{a}$	$a\,\mathrm{sinc}^2(\pi ka)$
Blackman	$B_A(x)$	$B_I(k)$
Connes	$\left(1-\frac{x^2}{a^2}\right)^2$	$8a\sqrt{2\pi}\,\frac{J_{5/2}(2\pi ka)}{(2\pi ka)^{5/2}}$
Cosine	$\cos\left(\frac{\pi x}{2a}\right)$	$\frac{4a\cos(2\pi ak)}{\pi(1-16a^2k^2)}$
Gaussian	$e^{-x^2/(2\sigma^2)}$	$2\int_0^a \cos(2\pi kx)e^{-x^2/(2\sigma^2)}\,dx$
Hamming	$Hm_A(x)$	$Hm_I(k)$
Hanning	$Hn_A(x)$	$Hn_I(k)$
Uniform	1	$2a\,\mathrm{sinc}\,(2\pi ka)$
Welch	$1-\frac{x^2}{a^2}$	$W_I(k)$

where

$$B_A(x) = 0.42 + 0.5\cos\left(\frac{\pi x}{a}\right) + 0.08\cos\left(\frac{2\pi x}{a}\right) \qquad (2)$$

$$B_I(k) = \frac{a(0.84 - 0.36a^2k^2 - 2.17\times 10^{-19}a^4k^4)\,\mathrm{sinc}(2\pi ak)}{(1-a^2k^2)(1-4a^2k^2)} \qquad (3)$$

$$Hm_A(x) = 0.54 + 0.46\cos\left(\frac{\pi x}{a}\right) \qquad (4)$$

$$Hm_I(k) = \frac{a(1.08 - 0.64a^2k^2)\,\mathrm{sinc}(2\pi ak)}{1-4a^2k^2} \qquad (5)$$

$$Hn_A(x) = \cos^2\left(\frac{\pi x}{2a}\right) \tag{6}$$

$$= \frac{1}{2}\left[1 + \cos\left(\frac{\pi x}{a}\right)\right] \tag{7}$$

$$Hn_I(k) = \frac{a\,\mathrm{sinc}\,(2\pi ak)}{1 - 4a^2k^2} \tag{8}$$

$$= a[\mathrm{sinc}(2\pi ka) + \tfrac{1}{2}\,\mathrm{sinc}(2\pi ka - \pi) + \tfrac{1}{2}\,\mathrm{sinc}(2\pi ka + \pi)] \tag{9}$$

$$W_I(k) = a2\sqrt{2\pi}\,\frac{J_{3/2}(2\pi ka)}{(2\pi ka)^{3/2}} \tag{10}$$

$$= a\frac{\sin(2\pi ka) - 2\pi ak\cos(2\pi ak)}{2a^3k^3\pi^3}. \tag{11}$$

Type	IF FWHM	IF Peak	$\frac{\text{Peak } (-) \text{ S.L.}}{\text{Peak}}$	$\frac{\text{Peak } (+) \text{ S.L.}}{\text{Peak}}$
Bartlett	1.77179	1	0.00000000	0.0471904
Blackman	2.29880	0.84	−0.00106724	0.00124325
Connes	1.90416	$\frac{16}{15}$	−0.0411049	0.0128926
Cosine	1.63941	$\frac{4}{\pi}$	−0.0708048	0.0292720
Gaussian	—	1	—	—
Hamming	1.81522	1.08	−0.00689132	0.00734934
Hanning	2.00000	1	−0.0267076	0.00843441
Uniform	1.20671	2	−0.217234	0.128375
Welch	1.59044	$\frac{4}{3}$	−0.0861713	0.356044

A general symmetric apodization function $A(x)$ can be written as a FOURIER SERIES

$$A(x) = a_0 + 2\sum_{n=1}^{\infty} a_n \cos\left(\frac{n\pi x}{b}\right), \tag{12}$$

where the COEFFICIENTS satisfy

$$a_0 + 2\sum_{n=1}^{\infty} a_n = 1. \tag{13}$$

The corresponding apparatus function is

$$I(t) \equiv \int_{-b}^{b} A(x)e^{-2\pi ikx}\,dx = 2b\Big\{a_0\,\mathrm{sinc}(2\pi kb) + \sum_{n=1}^{\infty}[\mathrm{sinc}(2\pi kb + n\pi) + \mathrm{sinc}(2\pi kb - n\pi)]\Big\}. \tag{14}$$

To obtain an APODIZATION FUNCTION with zero at $ka = 3/4$, use

$$a_0\,\mathrm{sinc}(\tfrac{3}{2}\pi) + a_1[\mathrm{sinc}(\tfrac{5}{2}\pi) + \mathrm{sinc}(\tfrac{1}{2}\pi) = 0. \tag{15}$$

Plugging in (13),

$$-(1 - 2a_1)\frac{2}{3\pi} + a_1\left(\frac{2}{5\pi} + \frac{2}{\pi}\right) = -\tfrac{1}{3}(1 - 2a_1) + a_1(\tfrac{1}{5} + 1) = 0 \tag{16}$$

$$a_1(\tfrac{6}{5} + \tfrac{2}{3}) = \tfrac{1}{3} \tag{17}$$

$$a_1 = \frac{\frac{1}{3}}{\frac{6}{5} + \frac{2}{3}} = \frac{5}{6\cdot 3 + 2\cdot 5} = \tfrac{5}{28} \tag{18}$$

$$a_0 = 1 - 2a_1 = \frac{28 - 2\cdot 5}{28} = \tfrac{18}{28} = \tfrac{9}{14}. \tag{19}$$

The HAMMING FUNCTION is close to the requirement that the APPARATUS FUNCTION goes to 0 at $ka = 5/4$, giving

$$a_0 = \tfrac{25}{46} \approx 0.5435 \tag{20}$$

$$a_1 = \tfrac{21}{92} \approx 0.2283. \tag{21}$$

The BLACKMAN FUNCTION is chosen so that the APPARATUS FUNCTION goes to 0 at $ka = 5/4$ and $9/4$, giving

$$a_0 = \tfrac{3969}{9304} \approx 0.4266 \tag{22}$$

$$a_1 = \tfrac{1155}{4652} \approx 0.2483 \tag{23}$$

$$a_2 = \tfrac{715}{18608} \approx 0.0384. \tag{24}$$

see also BARTLETT FUNCTION, BLACKMAN FUNCTION, CONNES FUNCTION, COSINE APODIZATION FUNCTION, FULL WIDTH AT HALF MAXIMUM, GAUSSIAN FUNCTION, HAMMING FUNCTION, HANN FUNCTION, HANNING FUNCTION, MERTZ APODIZATION FUNCTION, PARZEN APODIZATION FUNCTION, UNIFORM APODIZATION FUNCTION, WELCH APODIZATION FUNCTION

References

Ball, J. A. "The Spectral Resolution in a Correlator System" §4.3.5 in *Methods of Experimental Physics* **12C** (Ed. M. L. Meeks). New York: Academic Press, pp. 55–57, 1976.

Blackman, R. B. and Tukey, J. W. "Particular Pairs of Windows." In *The Measurement of Power Spectra, From the Point of View of Communications Engineering.* New York: Dover, pp. 95–101, 1959.

Brault, J. W. "Fourier Transform Spectrometry." In *High Resolution in Astronomy: 15th Advanced Course of the Swiss Society of Astronomy and Astrophysics* (Ed. A. Benz, M. Huber, and M. Mayor). Geneva Observatory, Sauverny, Switzerland, pp. 31–32, 1985.

Harris, F. J. "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform." *Proc. IEEE* **66**, 51–83, 1978.

Norton, R. H. and Beer, R. "New Apodizing Functions for Fourier Spectroscopy." *J. Opt. Soc. Amer.* **66**, 259–264, 1976.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 547–548, 1992.

Schnopper, H. W. and Thompson, R. I. "Fourier Spectrometers." In *Methods of Experimental Physics* **12A** (Ed. M. L. Meeks). New York: Academic Press, pp. 491–529, 1974.

Apollonius Circles

There are two completely different definitions of the so-called Apollonius circles:

1. The set of all points whose distances from two fixed points are in a constant ratio $1 : \mu$ (Ogilvy 1990).

2. The eight CIRCLES (two of which are nondegenerate) which solve APOLLONIUS' PROBLEM for three CIRCLES.

Given one side of a TRIANGLE and the ratio of the lengths of the other two sides, the LOCUS of the third VERTEX is the Apollonius circle (of the first type) whose CENTER is on the extension of the given side. For a given TRIANGLE, there are three circles of Apollonius.

Denote the three Apollonius circles (of the first type) of a TRIANGLE by k_1, k_2, and k_3, and their centers L_1, L_2, and L_3. The center L_1 is the intersection of the side A_2A_3 with the tangent to the CIRCUMCIRCLE at A_1. L_1 is also the pole of the SYMMEDIAN POINT K with respect to CIRCUMCIRCLE. The centers L_1, L_2, and L_3 are COLLINEAR on the POLAR of K with regard to its CIRCUMCIRCLE, called the LEMOINE LINE. The circle of Apollonius k_1 is also the locus of a point whose PEDAL TRIANGLE is ISOSCELES such that $\overline{P_1P_2} = \overline{P_1P_3}$.

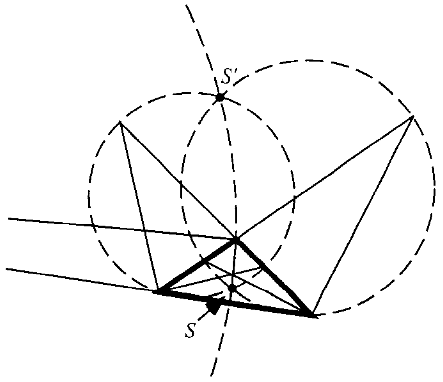

Let U and V be points on the side line BC of a TRIANGLE ΔABC met by the interior and exterior ANGLE BISECTORS of ANGLES A. The CIRCLE with DIAMETER UV is called the A-Apollonian circle. Similarly, construct the B- and C-Apollonian circles. The Apollonian circles pass through the VERTICES A, B, and C, and through the two ISODYNAMIC POINTS S and S'. The VERTICES of the D-TRIANGLE lie on the respective Apollonius circles.

see also APOLLONIUS' PROBLEM, APOLLONIUS PURSUIT PROBLEM, CASEY'S THEOREM, HART'S THEOREM, ISODYNAMIC POINTS, SODDY CIRCLES

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 40 and 294–299, 1929.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 14–23, 1990.

Apollonius Point

Consider the EXCIRCLES Γ_A, Γ_B, and Γ_C of a TRIANGLE, and the CIRCLE Γ internally TANGENT to all three. Denote the contact point of Γ and Γ_A by A', etc. Then the LINES AA', BB', and CC' CONCUR in this point. It has TRIANGLE CENTER FUNCTION

$$\alpha = \sin^2 A \cos^2[\tfrac{1}{2}(B-C)].$$

References

Kimberling, C. "Apollonius Point." `http://www.evansville.edu/~ck6/tcenters/recent/apollon.html`.

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Kimberling, C.; Iwata, S.; and Hidetosi, F. "Problem 1091 and Solution." *Crux Math.* **13**, 128–129 and 217–218, 1987.

Apollonius' Problem

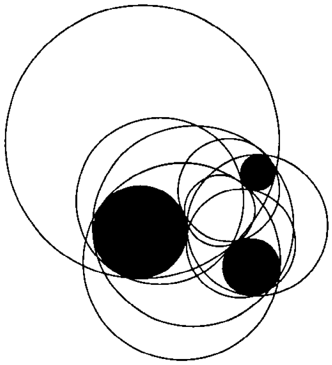

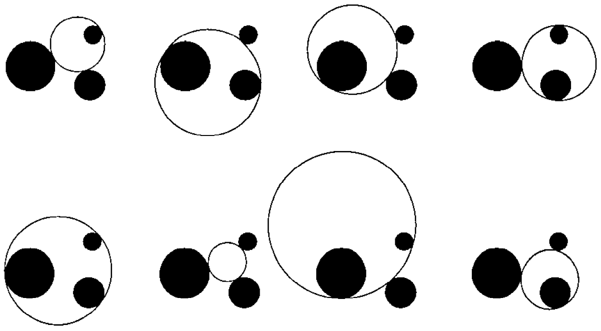

Given three objects, each of which may be a POINT, LINE, or CIRCLE, draw a CIRCLE that is TANGENT to each. There are a total of ten cases. The two easiest involve three points or three LINES, and the hardest involves three CIRCLES. Euclid solved the two easiest cases in his *Elements*, and the others (with the exception of the three CIRCLE problem), appeared in the *Tangencies* of Apollonius which was, however, lost. The general problem is, in principle, solvable by STRAIGHTEDGE and COMPASS alone.

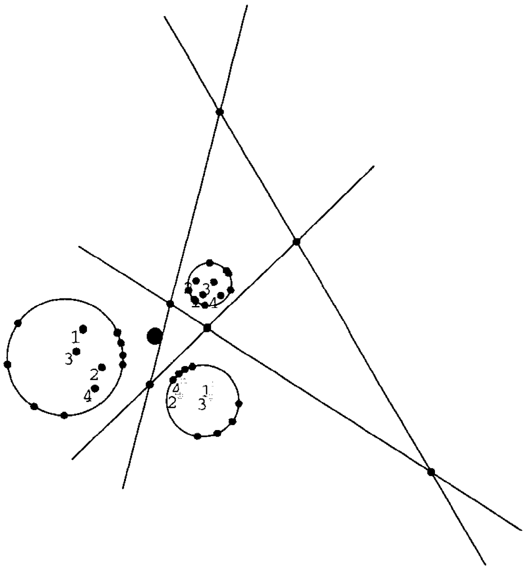

The three-CIRCLE problem was solved by Viète (Boyer 1968), and the solutions are called APOLLONIUS CIRCLES. There are eight total solutions. The simplest solution is obtained by solving the three simultaneous quadratic equations

$$(x-x_1)^2+(y-y_1)^2-(r\pm r_1)^2=0 \qquad (1)$$

$$(x-x_2)^2+(y-y_2)^2-(r\pm r_2)^2=0 \qquad (2)$$

$$(x-x_3)^2+(y-y_3)^2-(r\pm r_3)^2=0 \qquad (3)$$

in the three unknowns x, y, r for the eight triplets of signs (Courant and Robbins 1996). Expanding the equations gives

$$(x^2+y^2-r^2)-2xx_i-2yy_i\pm 2rr_i+({x_i}^2+{y_i}^2-{r_i}^2)=0 \qquad (4)$$

for $i=1$, 2, 3. Since the first term is the same for each equation, taking (2) − (1) and (3) − (1) gives

$$ax+by+cr=d \qquad (5)$$

$$a'x+b'y+c'r=d', \qquad (6)$$

where

$$a=2(x_1-x_2) \qquad (7)$$

$$b=2(y_1-y_2) \qquad (8)$$

$$c=\mp 2(r_1-r_2) \qquad (9)$$

$$d=({x_2}^2+{y_2}^2-{r_2}^2)-({x_1}^2+{y_1}^2-{r_1}^2) \qquad (10)$$

and similarly for a', b', c' and d' (where the 2 subscripts are replaced by 3s). Solving these two simultaneous linear equations gives

$$x=\frac{b'd-bd'-b'cr+bc'r}{ab'-ba'} \qquad (11)$$

$$y=\frac{-a'd+ad'+a'cr-ac'r}{ab'-a'b}, \qquad (12)$$

which can then be plugged back into the QUADRATIC EQUATION (1) and solved using the QUADRATIC FORMULA.

Perhaps the most elegant solution is due to Gergonne. It proceeds by locating the six HOMOTHETIC CENTERS (three internal and three external) of the three given CIRCLES. These lie three by three on four lines (illustrated above). Determine the POLES of one of these with respect to each of the three CIRCLES and connect the POLES with the RADICAL CENTER of the CIRCLES. If the connectors meet, then the three pairs of intersections are the points of tangency of two of the eight circles (Johnson 1929, Dörrie 1965). To determine *which* two of the eight Apollonius circles are produced by the three pairs, simply take the two which intersect the original three CIRCLES only in a single point of tangency. The procedure, when repeated, gives the other three pairs of CIRCLES.

If the three CIRCLES are mutually tangent, then the eight solutions collapse to two, known as the SODDY CIRCLES.

see also APOLLONIUS PURSUIT PROBLEM, BEND (CURVATURE), CASEY'S THEOREM, DESCARTES CIRCLE THEOREM, FOUR COINS PROBLEM, HART'S THEOREM, SODDY CIRCLES

References

Boyer, C. B. *A History of Mathematics.* New York: Wiley, p. 159, 1968.

Courant, R. and Robbins, H. "Apollonius' Problem." §3.3 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 117 and 125–127, 1996.

Dörrie, H. "The Tangency Problem of Apollonius." §32 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 154–160, 1965.

Gauss, C. F. *Werke, Vol. 4.* New York: George Olms, p. 399, 1981.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 118–121, 1929.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 48–51, 1990.

Pappas, T. *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 151, 1989.

Simon, M. *Über die Entwicklung der Elementargeometrie im XIX Jahrhundert.* Berlin, pp. 97–105, 1906.

Weisstein, E. W. "Plane Geometry." `http://www.astro.virginia.edu/~eww6n/math/notebooks/PlaneGeometry.m`.

Apollonius Pursuit Problem

Given a ship with a known constant direction and speed v, what course should be taken by a chase ship in pursuit (traveling at speed V) in order to intersect the other ship in as short a time as possible? The problem can be solved by finding all points which can be simultaneously reached by both ships, which is an APOLLONIUS CIRCLE with $\mu = v/V$. If the CIRCLE cuts the path of the pursued ship, the intersection is the point towards which the pursuit ship should steer. If the CIRCLE does not cut the path, then it cannot be caught.

see also APOLLONIUS CIRCLES, APOLLONIUS' PROBLEM, PURSUIT CURVE

References

Ogilvy, C. S. Solved by M. S. Klamkin. "A Slow Ship Intercepting a Fast Ship." Problem E991. *Amer. Math. Monthly* **59**, 408, 1952.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, p. 17, 1990.

Steinhaus, H. *Mathematical Snapshots, 3rd American ed.* New York: Oxford University Press, pp. 126–138, 1983.

Apollonius Theorem

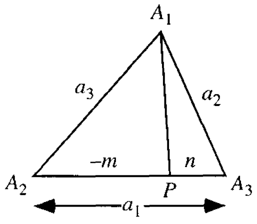

$$m{a_2}^2+n{a_3}^2=(m+n)\overline{A_1P}^2+m\overline{PA_3}^2+n\overline{PA_2}^2.$$

Apothem

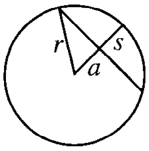

Given a CIRCLE, the PERPENDICULAR distance a from the MIDPOINT of a CHORD to the CIRCLE's center is called the apothem. It is also equal to the RADIUS r minus the SAGITTA s,

$$a=r-s.$$

see also CHORD, RADIUS, SAGITTA, SECTOR, SEGMENT

Apparatus Function

see INSTRUMENT FUNCTION

Appell Hypergeometric Function

A formal extension of the HYPERGEOMETRIC FUNCTION to two variables, resulting in four kinds of functions (Appell 1925),

$$F_1(\alpha;\beta,\beta';\gamma;x,y)=\sum_{m=0}^\infty\sum_{n=0}^\infty\frac{(\alpha)_{m+n}(\beta)_m(\beta')_n}{m!n!(\gamma)_{m+n}}x^my^n$$

$$F_2(\alpha;\beta,\beta';\gamma,\gamma';x,y)=\sum_{m=0}^\infty\sum_{n=0}^\infty\frac{(\alpha)_{m+n}(\beta)_m(\beta')_n}{m!n!(\gamma)_m(\gamma')_n}x^my^n$$

$$F_3(\alpha,\alpha';\beta,\beta';\gamma;x,y)=\sum_{m=0}^\infty\sum_{n=0}^\infty\frac{(\alpha)_m(\alpha')_n(\beta)_m(\beta')_n}{m!n!(\gamma)_{m+n}}x^my^n$$

$$F_4(\alpha;\beta;\gamma,\gamma';x,y)=\sum_{m=0}^\infty\sum_{n=0}^\infty\frac{(\alpha)_{m+n}(\beta)_{m+n}}{m!n!(\gamma)_m(\gamma')_n}x^my^n.$$

Appell defined the functions in 1880, and Picard showed in 1881 that they may all be expressed by INTEGRALS of the form

$$\int_0^1 u^\alpha(1-u)^\beta(1-xu)^\gamma(1-yu)^\delta\,du.$$

References

Appell, P. "Sur les fonctions hypergéométriques de plusieurs variables." In *Mémoir. Sci. Math.* Paris: Gauthier-Villars, 1925.

Bailey, W. N. *Generalised Hypergeometric Series.* Cambridge, England: Cambridge University Press, p. 73, 1935.

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1461, 1980.

Appell Polynomial

A type of POLYNOMIAL which includes the BERNOULLI POLYNOMIAL, HERMITE POLYNOMIAL, and LAGUERRE POLYNOMIAL as special cases. The series of POLYNOMIALS $\{A_n(z)\}_{n=0}^\infty$ is defined by

$$A(t)e^{zt}=\sum_{n=0}^\infty A_n(z)t^n,$$

where

$$A(t)=\sum_{k=0}^\infty a_kt^k$$

is a formal POWER series with $k=0,\,1,\,\ldots$ and $a_0\neq 0$.

References

Hazewinkel, M. (Managing Ed.). *Encyclopaedia of Mathematics: An Updated and Annotated Translation of the Soviet "Mathematical Encyclopaedia."* Dordrecht, Netherlands: Reidel, pp. 209–210, 1988.

Appell Transformation

A HOMOGRAPHIC transformation

$$x_1=\frac{ax+by+c}{a''x+b''y+c}$$
$$y_1=\frac{a'x+b'y+c'}{a''x+b''y+c''}$$

with t_1 substituted for t according to

$$k\,dt_1=\frac{dt}{(a''x+b''y+c'')^2}.$$

References

Hazewinkel, M. (Managing Ed.). *Encyclopaedia of Mathematics: An Updated and Annotated Translation of the Soviet "Mathematical Encyclopaedia."* Dordrecht, Netherlands: Reidel, pp. 210–211, 1988.

Apple

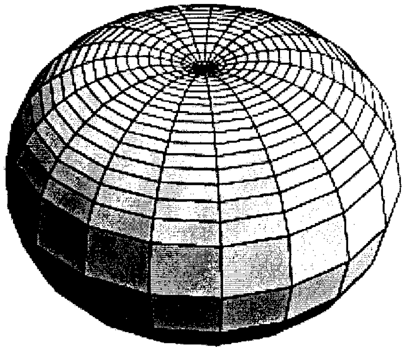

A SURFACE OF REVOLUTION defined by Kepler. It consists of more than half of a circular ARC rotated about an axis passing through the endpoints of the ARC. The equations of the upper and lower boundaries in the x-z PLANE are

$$z_{\pm} = \pm\sqrt{R^2 - (x-r)^2}$$

for $R > r$ and $x \in [-(r+R), r+R]$. It is the outside surface of a SPINDLE TORUS.

see also BUBBLE, LEMON, SPHERE-SPHERE INTERSECTION, SPINDLE TORUS

Approximately Equal

If two quantities A and B are approximately equal, this is written $A \approx B$.

see also DEFINED, EQUAL

Approximation Theory

The mathematical study of how given quantities can be approximated by other (usually simpler) ones under appropriate conditions. Approximation theory also studies the size and properties of the ERROR introduced by approximation. Approximations are often obtained by POWER SERIES expansions in which the higher order terms are dropped.

see also LAGRANGE REMAINDER

References

Achieser, N. I. and Hyman, C. J. *Theory of Approximation.* New York: Dover, 1993.

Akheizer, N. I. *Theory of Approximation.* New York: Dover, 1992.

Cheney, E. W. *Introduction to Approximation Theory.* New York: McGraw-Hill, 1966.

Golomb, M. *Lectures on Theory of Approximation.* Argonne, IL: Argonne National Laboratory, 1962.

Jackson, D. *The Theory of Approximation.* New York: Amer. Math. Soc., 1930.

Natanson, I. P. *Constructive Function Theory, Vol. 1: Uniform Approximation.* New York: Ungar, 1964.

Petrushev, P. P. and Popov, V. A. *Rational Approximation of Real Functions.* New York: Cambridge University Press, 1987.

Rivlin, T. J. *An Introduction to the Approximation of Functions.* New York: Dover, 1981.

Timan, A. F. *Theory of Approximation of Functions of a Real Variable.* New York: Dover, 1994.

Arakelov Theory

A formal mathematical theory which introduces "components at infinity" by defining a new type of divisor class group of INTEGERS of a NUMBER FIELD. The divisor class group is called an "arithmetic surface."

see also ARITHMETIC GEOMETRY

Arbelos

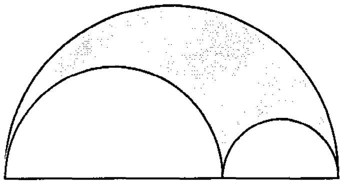

The term "arbelos" means SHOEMAKER'S KNIFE in Greek, and this term is applied to the shaded AREA in the above figure which resembles the blade of a knife used by ancient cobblers (Gardner 1979). Archimedes himself is believed to have been the first mathematician to study the mathematical properties of this figure. The position of the central notch is arbitrary and can be located anywhere along the DIAMETER.

The arbelos satisfies a number of unexpected identities (Gardner 1979).

1. Call the radii of the left and right SEMICIRCLES a and b, respectively, with $a + b \equiv R$. Then the arc length along the bottom of the arbelos is

   $$L = 2\pi a + 2\pi b = 2\pi(a+b) = 2\pi R,$$

   so the arc lengths along the top and bottom of the arbelos are the same.

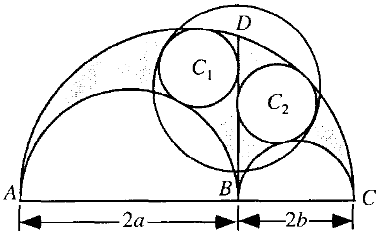

2. Draw the PERPENDICULAR BD from the tangent of the two SEMICIRCLES to the edge of the large CIRCLE. Then the AREA of the arbelos is the same as the AREA of the CIRCLE with DIAMETER BD.
3. The CIRCLES C_1 and C_2 inscribed on each half of BD on the arbelos (called ARCHIMEDES' CIRCLES) each have DIAMETER $(AB)(BC)/(AC)$. Furthermore, the smallest CIRCUMCIRCLE of these two circles has an area equal to that of the arbelos.
4. The line tangent to the semicircles AB and BC contains the point E and F which lie on the lines AD and CD, respectively. Furthermore, BD and EF bisect each other, and the points B, D, E, and F are CONCYCLIC.

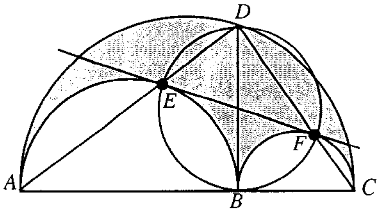

5. In addition to the ARCHIMEDES' CIRCLES C_1 and C_2 in the arbelos figure, there is a third circle C_3 called the BANKOFF CIRCLE which is congruent to these two.

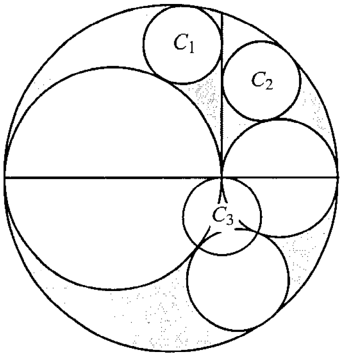

6. Construct a chain of TANGENT CIRCLES starting with the CIRCLE TANGENT to the two small ones and large one. The centers of the CIRCLES lie on an ELLIPSE, and the DIAMETER of the nth CIRCLE C_n is $(1/n)$th PERPENDICULAR distance to the base of the SEMICIRCLE. This result is most easily proven using INVERSION, but was known to Pappus, who referred to it as an ancient theorem (Hood 1961, Cadwell 1966, Gardner 1979, Bankoff 1981). If $r \equiv AB/AC$, then the radius of the nth circle in the PAPPUS CHAIN is

$$r_n = \frac{(1-r)r}{2[n^2(1-r)^2+r]}.$$

This general result simplifies to $r_n = 1/(6+n^2)$ for $r = 2/3$ (Gardner 1979). Further special cases when $AC = 1 + AB$ are considered by Gaba (1940).

7. If B divides AC in the GOLDEN RATIO ϕ, then the circles in the chain satisfy a number of other special properties (Bankoff 1955).

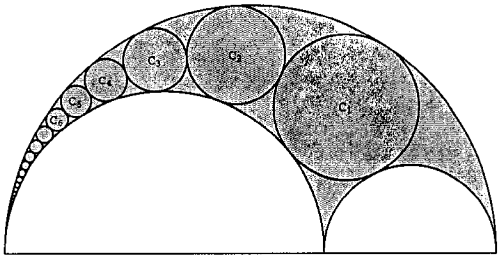

see also ARCHIMEDES' CIRCLES, BANKOFF CIRCLE, COXETER'S LOXODROMIC SEQUENCE OF TANGENT CIRCLES, GOLDEN RATIO, INVERSION, PAPPUS CHAIN, STEINER CHAIN

References
Bankoff, L. "The Fibonacci Arbelos." *Scripta Math.* **20**, 218, 1954.
Bankoff, L. "The Golden Arbelos." *Scripta Math.* **21**, 70–76, 1955.
Bankoff, L. "Are the Twin Circles of Archimedes Really Twins?" *Math. Mag.* **47**, 214–218, 1974.
Bankoff, L. "How Did Pappus Do It?" In *The Mathematical Gardner* (Ed. D. Klarner). Boston, MA: Prindle, Weber, and Schmidt, pp. 112–118, 1981.
Bankoff, L. "The Marvelous Arbelos." In *The Lighter Side of Mathematics* (Ed. R. K. Guy and R. E. Woodrow). Washington, DC: Math. Assoc. Amer., 1994.
Cadwell, J. H. *Topics in Recreational Mathematics.* Cambridge, England: Cambridge University Press, 1966.
Gaba, M. G. "On a Generalization of the Arbelos." *Amer. Math. Monthly* **47**, 19–24, 1940.
Gardner, M. "Mathematical Games: The Diverse Pleasures of Circles that Are Tangent to One Another." *Sci. Amer.* **240**, 18–28, Jan. 1979.
Heath, T. L. *The Works of Archimedes with the Method of Archimedes.* New York: Dover, 1953.
Hood, R. T. "A Chain of Circles." *Math. Teacher* **54**, 134–137, 1961.
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 116–117, 1929.
Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 54–55, 1990.

Arborescence

A DIGRAPH is called an arborescence if, from a given node x known as the ROOT, there is exactly one elementary path from x to every other node y.

see also ARBORICITY

Arboricity

Given a GRAPH G, the arboricity is the MINIMUM number of line-disjoint acyclic SUBGRAPHS whose UNION is G.

see also ANARBORICITY

Arc

In general, any smooth curve joining two points. In particular, any portion (other than the entire curve) of a CIRCLE or ELLIPSE.

see also APPLE, CIRCLE-CIRCLE INTERSECTION, FIVE DISKS PROBLEM, FLOWER OF LIFE, LEMON, LENS, PIECEWISE CIRCULAR CURVE, REULEAUX POLYGON, REULEAUX TRIANGLE, SALINON, SEED OF LIFE, TRIANGLE ARCS, VENN DIAGRAM, YIN-YANG

Arc Length

Arc length is defined as the length along a curve,

$$s \equiv \int_a^b |d\boldsymbol{\ell}|. \tag{1}$$

Defining the line element $ds^2 \equiv |d\boldsymbol{\ell}|^2$, parameterizing the curve in terms of a parameter t, and noting that

ds/dt is simply the magnitude of the VELOCITY with which the end of the RADIUS VECTOR $\mathbf{r}$ moves gives

$$s = \int_a^b ds = \int_a^b \frac{ds}{dt}\,dt = \int_a^b |\mathbf{r}'(t)|\,dt. \tag{2}$$

In POLAR COORDINATES,

$$d\boldsymbol{\ell} = \hat{\mathbf{r}}\,dr + r\hat{\boldsymbol{\theta}}\,d\theta = \left(\frac{dr}{d\theta}\hat{\mathbf{r}} + r\hat{\boldsymbol{\theta}}\right)d\theta, \tag{3}$$

so

$$ds = |d\boldsymbol{\ell}| = \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}\,d\theta \tag{4}$$

$$s = \int |d\boldsymbol{\ell}| = \int_{\theta_1}^{\theta_2} \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}\,d\theta. \tag{5}$$

In CARTESIAN COORDINATES,

$$d\boldsymbol{\ell} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} \tag{6}$$

$$ds = \sqrt{dx^2 + dy^2} = \sqrt{\left(\frac{dy}{dx}\right)^2 + 1}\,dx. \tag{7}$$

Therefore, if the curve is written

$$\mathbf{r}(x) = x\hat{\mathbf{x}} + f(x)\hat{\mathbf{y}}, \tag{8}$$

then

$$s = \int_a^b \sqrt{1 + f'^2(x)}\,dx. \tag{9}$$

If the curve is instead written

$$\mathbf{r}(t) = x(t)\hat{\mathbf{x}} + y(t)\hat{\mathbf{y}}, \tag{10}$$

then

$$s = \int_a^b \sqrt{x'^2(t) + y'^2(t)}\,dt. \tag{11}$$

Or, in three dimensions,

$$\mathbf{r}(t) = x(t)\hat{\mathbf{x}} + y(t)\hat{\mathbf{y}} + z(t)\hat{\mathbf{z}}, \tag{12}$$

so

$$s = \int_a^b \sqrt{x'^2(t) + y'^2(t) + z'^2(t)}\,dt. \tag{13}$$

see also CURVATURE, GEODESIC, NORMAL VECTOR, RADIUS OF CURVATURE, RADIUS OF TORSION, SPEED, SURFACE AREA, TANGENTIAL ANGLE, TANGENT VECTOR, TORSION (DIFFERENTIAL GEOMETRY), VELOCITY

Arc Minute

A unit of ANGULAR measure equal to 60 ARC SECONDS, or 1/60 of a DEGREE. The arc minute is denoted ′ (not to be confused with the symbol for feet).

Arc Second

A unit of ANGULAR measure equal to 1/60 of an ARC MINUTE, or 1/3600 of a DEGREE. The arc second is denoted ″ (not to be confused with the symbol for inches).

Arccosecant

see INVERSE COSECANT

Arccosine

see INVERSE COSINE

Arccotangent

see INVERSE COTANGENT

Arch

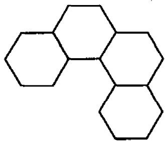

A 4-POLYHEX.

References

Gardner, M. *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, p. 147, 1978.

Archimedes Algorithm

Successive application of ARCHIMEDES' RECURRENCE FORMULA gives the Archimedes algorithm, which can be used to provide successive approximations to π (PI). The algorithm is also called the BORCHARDT-PFAFF ALGORITHM. Archimedes obtained the first rigorous approximation of π by CIRCUMSCRIBING and INSCRIBING $n = 6\cdot 2^k$-gons on a CIRCLE. From ARCHIMEDES' RECURRENCE FORMULA, the CIRCUMFERENCES a and b of the circumscribed and inscribed POLYGONS are

$$a(n) = 2n\tan\left(\frac{\pi}{n}\right) \tag{1}$$

$$b(n) = 2n\sin\left(\frac{\pi}{n}\right), \tag{2}$$

where

$$b(n) < C = 2\pi r = 2\pi\cdot 1 = 2\pi < a(n). \tag{3}$$

For a HEXAGON, $n = 6$ and

$$a_0 \equiv a(6) = 4\sqrt{3} \tag{4}$$

$$b_0 \equiv b(6) = 6, \tag{5}$$

where $a_k \equiv a(6\cdot 2^k)$. The first iteration of ARCHIMEDES' RECURRENCE FORMULA then gives

$$a_1 = \frac{2\cdot 6\cdot 4\sqrt{3}}{6 + 4\sqrt{3}} = \frac{24\sqrt{3}}{3 + 2\sqrt{3}} = 24(2 - \sqrt{3}) \tag{6}$$

$$\begin{aligned} b_1 &= \sqrt{24(2-\sqrt{3})\cdot 6} = 12\sqrt{2-\sqrt{3}} \\ &= 6(\sqrt{6} - \sqrt{2}). \end{aligned} \tag{7}$$

Additional iterations do not have simple closed forms, but the numerical approximations for $k = 0, 1, 2, 3, 4$ (corresponding to 6-, 12-, 24-, 48-, and 96-gons) are

$$3.00000 < \pi < 3.46410 \tag{8}$$

$$3.10583 < \pi < 3.21539 \tag{9}$$

$$3.13263 < \pi < 3.15966 \tag{10}$$

$$3.13935 < \pi < 3.14609 \tag{11}$$

$$3.14103 < \pi < 3.14271. \tag{12}$$

By taking $k = 4$ (a 96-gon) and using strict inequalities to convert irrational bounds to rational bounds at each step, Archimedes obtained the slightly looser result

$$\tfrac{223}{71} = 3.14084\ldots < \pi < \tfrac{22}{7} = 3.14285\ldots. \tag{13}$$

References

Miel, G. "Of Calculations Past and Present: The Archimedean Algorithm." *Amer. Math. Monthly* **90**, 17–35, 1983.

Phillips, G. M. "Archimedes in the Complex Plane." *Amer. Math. Monthly* **91**, 108–114, 1984.

Archimedes' Axiom

An AXIOM actually attributed to Eudoxus (Boyer 1968) which states that

$$a/b = c/d$$

IFF the appropriate one of following conditions is satisfied for INTEGERS m and n:

1. If $ma < nb$, then $mc < md$.
2. If $ma = nd$, then $mc = nd$.
3. If $ma > nd$, then $mc > nd$.

ARCHIMEDES' LEMMA is sometimes also known as Archimedes' axiom.

References

Boyer, C. B. *A History of Mathematics.* New York: Wiley, p. 99, 1968.

Archimedes' Cattle Problem

Also called the BOVINUM PROBLEMA. It is stated as follows: "The sun god had a herd of cattle consisting of bulls and cows, one part of which was white, a second black, a third spotted, and a fourth brown. Among the bulls, the number of white ones was one half plus one third the number of the black greater than the brown; the number of the black, one quarter plus one fifth the number of the spotted greater than the brown; the number of the spotted, one sixth and one seventh the number of the white greater than the brown. Among the cows, the number of white ones was one third plus one quarter of the total black cattle; the number of the black, one quarter plus one fifth the total of the spotted cattle; the number of spotted, one fifth plus one sixth the total of the brown cattle; the number of the brown, one sixth plus one seventh the total of the white cattle. What was the composition of the herd?"

Solution consists of solving the simultaneous DIOPHANTINE EQUATIONS in INTEGERS W, X, Y, Z (the number of white, black, spotted, and brown bulls) and w, x, y, z (the number of white, black, spotted, and brown cows),

$$W = \tfrac{5}{6}X + Z \tag{1}$$

$$X = \tfrac{9}{20}Y + Z \tag{2}$$

$$Y = \tfrac{13}{42}W + Z \tag{3}$$

$$w = \tfrac{7}{12}(X + x) \tag{4}$$

$$x = \tfrac{9}{20}(Y + y) \tag{5}$$

$$y = \tfrac{11}{30}(Z + z) \tag{6}$$

$$z = \tfrac{13}{42}(W + w). \tag{7}$$

The smallest solution in INTEGERS is

$$W = 10{,}366{,}482 \tag{8}$$

$$X = 7{,}460{,}514 \tag{9}$$

$$Y = 7{,}358{,}060 \tag{10}$$

$$Z = 4{,}149{,}387 \tag{11}$$

$$w = 7{,}206{,}360 \tag{12}$$

$$x = 4{,}893{,}246 \tag{13}$$

$$y = 3{,}515{,}820 \tag{14}$$

$$z = 5{,}439{,}213. \tag{15}$$

A more complicated version of the problem requires that $W+X$ be a SQUARE NUMBER and $Y+Z$ a TRIANGULAR NUMBER. The solution to this PROBLEM are numbers with 206544 or 206545 digits.

References

Amthor, A. and Krumbiegel B. "Das Problema bovinum des Archimedes." *Z. Math. Phys.* **25**, 121–171, 1880.

Archibald, R. C. "Cattle Problem of Archimedes." *Amer. Math. Monthly* **25**, 411–414, 1918.

Beiler, A. H. *Recreations in the Theory of Numbers: The Queen of Mathematics Entertains.* New York: Dover, pp. 249–252, 1966.

Bell, A. H. "Solution to the Celebrated Indeterminate Equation $x^2 - ny^2 = 1$." *Amer. Math. Monthly* **1**, 240, 1894.

Bell, A. H. "'Cattle Problem.' By Archimedes 251 BC." *Amer. Math. Monthly* **2**, 140, 1895.

Bell, A. H. "Cattle Problem of Archimedes." *Math. Mag.* **1**, 163, 1882–1884.

Calkins, K. G. "Archimedes' *Problema Bovinum.*" `http://www.andrews.edu/~calkins/cattle.html`.

Dörrie, H. "Archimedes' *Problema Bovinum.*" §1 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 3–7, 1965.

Grosjean, C. C. and de Meyer, H. E. "A New Contribution to the Mathematical Study of the Cattle-Problem of Archimedes." In *Constantin Carathéodory: An International Tribute, Vols. 1 and 2* (Ed. T. M. Rassias). Teaneck, NJ: World Scientific, pp. 404–453, 1991.

Merriman, M. "Cattle Problem of Archimedes." *Pop. Sci. Monthly* **67**, 660, 1905.

Rorres, C. "The Cattle Problem." `http://www.mcs.drexel.edu/~crorres/Archimedes/Cattle/Statement.html`.

Vardi, I. "Archimedes' Cattle Problem." *Amer. Math. Monthly* **105**, 305–319, 1998.

Archimedes' Circles

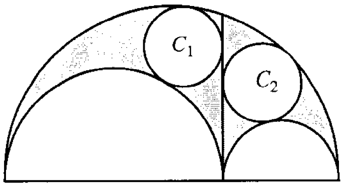

Draw the PERPENDICULAR LINE from the intersection of the two small SEMICIRCLES in the ARBELOS. The two CIRCLES C_1 and C_2 TANGENT to this line, the large SEMICIRCLE, and each of the two SEMICIRCLES are then congruent and known as Archimedes' circles.

see also ARBELOS, BANKOFF CIRCLE, SEMICIRCLE

Archimedes' Constant

see PI

Archimedes' Hat-Box Theorem

Enclose a SPHERE in a CYLINDER and slice PERPENDICULARLY to the CYLINDER's axis. Then the SURFACE AREA of the of SPHERE slice is equal to the SURFACE AREA of the CYLINDER slice.

Archimedes' Lemma

Also known as the continuity axiom, this LEMMA survives in the writings of Eudoxus (Boyer 1968). It states that, given two magnitudes having a ratio, one can find a multiple of either which will exceed the other. This principle was the basis for the EXHAUSTION METHOD which Archimedes invented to solve problems of AREA and VOLUME.

see also CONTINUITY AXIOMS

References
Boyer, C. B. *A History of Mathematics.* New York: Wiley, p. 100, 1968.

Archimedes' Midpoint Theorem

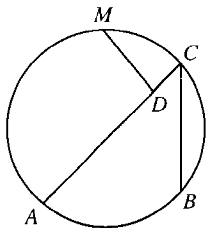

Let M be the MIDPOINT of the ARC AMB. Pick C at random and pick D such that $MD \perp AC$ (where $\perp$ denotes PERPENDICULAR). Then

$$AD = DC + BC.$$

see also MIDPOINT

References
Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., pp. 31–32, 1991.

Archimedes' Postulate

see ARCHIMEDES' LEMMA

Archimedes' Problem

Cut a SPHERE by a PLANE in such a way that the VOLUMES of the SPHERICAL SEGMENTS have a given RATIO.

see also SPHERICAL SEGMENT

Archimedes' Recurrence Formula

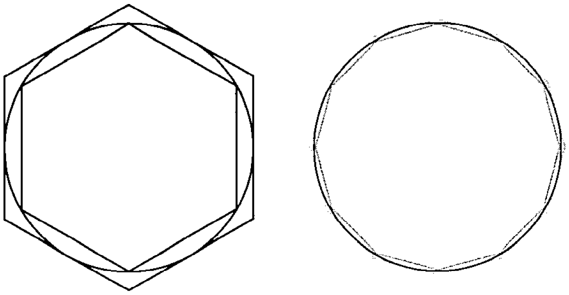

Let a_n and b_n be the PERIMETERS of the CIRCUMSCRIBED and INSCRIBED n-gon and a_{2n} and b_{2n} the PERIMETERS of the CIRCUMSCRIBED and INSCRIBED $2n$-gon. Then

$$a_{2n} = \frac{2a_nb_n}{a_n+b_n} \tag{1}$$

$$b_{2n} = \sqrt{a_{2n}b_n}\,. \tag{2}$$

The first follows from the fact that side lengths of the POLYGONS on a CIRCLE of RADIUS $r = 1$ are

$$s_R = 2\tan\left(\frac{\pi}{n}\right) \tag{3}$$

$$s_r = 2\sin\left(\frac{\pi}{n}\right), \tag{4}$$

so

$$a_n = 2n\tan\left(\frac{\pi}{n}\right) \tag{5}$$

$$b_n = 2n\sin\left(\frac{\pi}{n}\right). \tag{6}$$

But

$$\frac{2a_nb_n}{a_n+b_n} = \frac{2\cdot 2n\tan\left(\frac{\pi}{n}\right)\cdot 2n\sin\left(\frac{\pi}{n}\right)}{2n\tan\left(\frac{\pi}{n}\right)+2n\sin\left(\frac{\pi}{n}\right)} = 4n\frac{\tan\left(\frac{\pi}{n}\right)\sin\left(\frac{\pi}{n}\right)}{\tan\left(\frac{\pi}{n}\right)+\sin\left(\frac{\pi}{n}\right)}. \tag{7}$$

Using the identity

$$\tan(\tfrac{1}{2}x) = \frac{\tan x\sin x}{\tan x+\sin x} \tag{8}$$

then gives

$$\frac{2a_nb_n}{a_n+b_n} = 4n\tan\left(\frac{\pi}{2n}\right) = a_{2n}. \tag{9}$$

The second follows from

$$\sqrt{a_{2n}b_n} = \sqrt{4n\tan\left(\frac{\pi}{2n}\right)\cdot 2n\sin\left(\frac{\pi}{n}\right)} \qquad (10)$$

Using the identity

$$\sin x = 2\sin(\tfrac{1}{2}x)\cos(\tfrac{1}{2}x) \qquad (11)$$

gives

$$\begin{aligned}\sqrt{a_{2n}b_n} &= 2n\sqrt{2\tan\left(\frac{\pi}{2n}\right)\cdot 2\sin\left(\frac{\pi}{2n}\right)\cos\left(\frac{\pi}{2n}\right)}\\ &= 4n\sqrt{\sin^2\left(\frac{\pi}{2n}\right)} = 4n\sin\left(\frac{\pi}{2n}\right) = b_{2n}. \qquad (12)\end{aligned}$$

Successive application gives the ARCHIMEDES ALGORITHM, which can be used to provide successive approximations to PI (π).

see also ARCHIMEDES ALGORITHM, PI

References

Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, p. 186, 1965.

Archimedean Solid

The Archimedean solids are convex POLYHEDRA which have a similar arrangement of nonintersecting regular plane CONVEX POLYGONS of two or more different types about each VERTEX with all sides the same length. The Archimedean solids are distinguished from the PRISMS, ANTIPRISMS, and ELONGATED SQUARE GYROBICUPOLA by their symmetry group: the Archimedean solids have a spherical symmetry, while the others have "dihedral" symmetry. The Archimedean solids are sometimes also referred to as the SEMIREGULAR POLYHEDRA.

Pugh (1976, p. 25) points out the Archimedean solids are all capable of being circumscribed by a regular TETRAHEDRON so that four of their faces lie on the faces of that TETRAHEDRON. A method of constructing the Archimedean solids using a method known as "expansion" has been enumerated by Stott (Stott 1910; Ball and Coxeter 1987, pp. 139–140).

Let the cyclic sequence $S = (p_1, p_2, \ldots, p_q)$ represent the degrees of the faces surrounding a vertex (i.e., S is a list of the number of sides of all polygons surrounding any vertex). Then the definition of an Archimedean solid requires that the sequence must be the same for each vertex to within ROTATION and REFLECTION. Walsh (1972) demonstrates that S represents the degrees of the faces surrounding each vertex of a semiregular convex polyhedron or TESSELLATION of the plane IFF

1. $q \geq 3$ and every member of S is at least 3,
2. $\sum_{i=1}^q \frac{1}{p_i} \geq \frac{1}{2}q - 1$, with equality in the case of a plane TESSELLATION, and
3. for every ODD NUMBER $p \in S$, S contains a subsequence (b, p, b).

Condition (1) simply says that the figure consists of two or more polygons, each having at least three sides. Condition (2) requires that the sum of interior angles at a vertex must be equal to a full rotation for the figure to lie in the plane, and less than a full rotation for a solid figure to be convex.

The usual way of enumerating the semiregular polyhedra is to eliminate solutions of conditions (1) and (2) using several classes of arguments and then prove that the solutions left are, in fact, semiregular (Kepler 1864, pp. 116–126; Catalan 1865, pp. 25–32; Coxeter 1940, p. 394; Coxeter *et al.* 1954; Lines 1965, pp. 202–203; Walsh 1972). The following table gives all possible regular and semiregular polyhedra and tessellations. In the table, 'P' denotes PLATONIC SOLID, 'M' denotes a PRISM or ANTIPRISM, 'A' denotes an Archimedean solid, and 'T' a plane tessellation.

S	Fg.	Solid	Schläfli
(3, 3, 3)	P	tetrahedron	$\{3,3\}$
(3, 4, 4)	M	triangular prism	$\mathrm{t}\{2,3\}$
(3, 6, 6)	A	truncated tetrahedron	$\mathrm{t}\{3,3\}$
(3, 8, 8)	A	truncated cube	$\mathrm{t}\{4,3\}$
(3, 10, 10)	A	truncated dodecahedron	$\mathrm{t}\{5,3\}$
(3, 12, 12)	T	(plane tessellation)	$\mathrm{t}\{6,3\}$
(4, 4, n)	M	n-gonal Prism	$\mathrm{t}\{2,n\}$
(4, 4, 4)	P	cube	$\{4,3\}$
(4, 6, 6)	A	truncated octahedron	$\mathrm{t}\{3,4\}$
(4, 6, 8)	A	great rhombicuboct.	$\mathrm{t}\begin{Bmatrix}3\\4\end{Bmatrix}$
(4, 6, 10)	A	great rhombicosidodec.	$\mathrm{t}\begin{Bmatrix}3\\5\end{Bmatrix}$
(4, 6, 12)	T	(plane tessellation)	$\mathrm{t}\begin{Bmatrix}3\\6\end{Bmatrix}$
(4, 8, 8)	T	(plane tessellation)	$\mathrm{t}\{4,4\}$
(5, 5, 5)	P	dodecahedron	$\{5,3\}$
(5, 6, 6)	A	truncated icosahedron	$\mathrm{t}\{3,5\}$
(6, 6, 6)	T	(plane tessellation)	$\{6,3\}$
(3, 3, 3, n)	M	n-gonal antiprism	$\mathrm{s}\begin{Bmatrix}2\\n\end{Bmatrix}$
(3, 3, 3, 3)	P	octahedron	$\{3,4\}$
(3, 4, 3, 4)	A	cuboctahedron	$\begin{Bmatrix}3\\4\end{Bmatrix}$
(3, 5, 3, 5)	A	icosidodecahedron	$\begin{Bmatrix}3\\5\end{Bmatrix}$
(3, 6, 3, 6)	T	(plane tessellation)	$\begin{Bmatrix}3\\6\end{Bmatrix}$
(3, 4, 4, 4)	A	small rhombicuboct.	$\mathrm{r}\begin{Bmatrix}3\\4\end{Bmatrix}$
(3, 4, 5, 4)	A	small rhombicosidodec.	$\mathrm{r}\begin{Bmatrix}3\\5\end{Bmatrix}$
(3, 4, 6, 4)	T	(plane tessellation)	$\mathrm{r}\begin{Bmatrix}3\\6\end{Bmatrix}$
(4, 4, 4, 4)	T	(plane tessellation)	$\{4,4\}$
(3, 3, 3, 3, 3)	P	icosahedron	$\{3,5\}$
(3, 3, 3, 3, 4)	A	snub cube	$\mathrm{s}\begin{Bmatrix}3\\4\end{Bmatrix}$
(3, 3, 3, 3, 5)	A	snub dodecahedron	$\mathrm{s}\begin{Bmatrix}3\\5\end{Bmatrix}$
(3, 3, 3, 3, 6)	T	(plane tessellation)	$\mathrm{s}\begin{Bmatrix}3\\6\end{Bmatrix}$
(3, 3, 3, 4, 4)	T	(plane tessellation)	—
(3, 3, 4, 3, 4)	T	(plane tessellation)	$\mathrm{s}\begin{Bmatrix}4\\4\end{Bmatrix}$
(3, 3, 3, 3, 3)	T	(plane tessellation)	$\{3,6\}$

As shown in the above table, there are exactly 13 Archimedean solids (Walsh 1972, Ball and Coxeter 1987).

They are called the Cuboctahedron, Great Rhombicosidodecahedron, Great Rhombicuboctahedron, Icosidodecahedron, Small Rhombicosidodecahedron, Small Rhombicuboctahedron, Snub Cube, Snub Dodecahedron, Truncated Cube, Truncated Dodecahedron, Truncated Icosahedron (soccer ball), Truncated Octahedron, and Truncated Tetrahedron. The Archimedean solids satisfy

$$(2\pi - \sigma)V = 4\pi,$$

where σ is the sum of face-angles at a vertex and V is the number of vertices (Steinitz and Rademacher 1934, Ball and Coxeter 1987).

Here are the Archimedean solids shown in alphabetical order (left to right, then continuing to the next row).

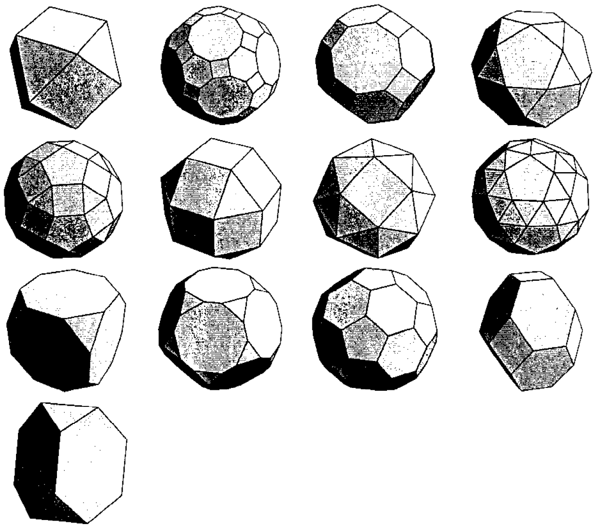

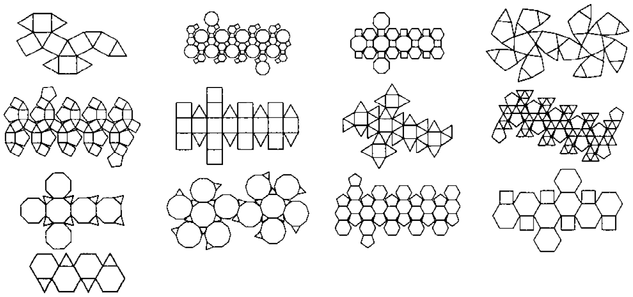

The following table lists the symbol and number of faces of each type for the Archimedean solids (Wenninger 1989, p. 9).

Solid	Schläfli	Wythoff	C&R
cuboctahedron	$\begin{Bmatrix}3\\4\end{Bmatrix}$	$2 \mid 3\ 4$	$(3.4)^2$
great rhombicosidodecahedron	$t\begin{Bmatrix}3\\5\end{Bmatrix}$	$2\ 3\ 5 \mid$	
great rhombicuboctahedron	$t\begin{Bmatrix}3\\4\end{Bmatrix}$	$2\ 3\ 4 \mid$	
icosidodecahedron	$\begin{Bmatrix}3\\5\end{Bmatrix}$	$2 \mid 3\ 5$	$(3.5)^2$
small rhombicosidodecahedron	$t\begin{Bmatrix}3\\5\end{Bmatrix}$	$3\ 5 \mid 2$	3.4.5.4
small rhombicuboctahedron	$r\begin{Bmatrix}3\\4\end{Bmatrix}$	$3\ 4 \mid 2$	3.4^3
snub cube	$s\begin{Bmatrix}3\\4\end{Bmatrix}$	$\mid 2\ 3\ 4$	$3^4.4$
snub dodecahedron	$s\begin{Bmatrix}3\\5\end{Bmatrix}$	$\mid 2\ 3\ 5$	$3^4.5$
truncated cube	$t\{4,3\}$	$2\ 3 \mid 4$	3.8^2
truncated dodecahedron	$t\{5,3\}$	$2\ 3 \mid 5$	3.10^2
truncated icosahedron	$t\{3,5\}$	$2\ 5 \mid 3$	5.6^2
truncated octahedron	$t\{3,4\}$	$2\ 4 \mid 3$	4.6^2
truncated tetrahedron	$t\{3,3\}$	$2\ 3 \mid 3$	3.6^2

Solid	v	e	f_3	f_4	f_5	f_6	f_8	f_{10}
cuboctahedron	12	24	8	6				
great rhombicos.	120	180		30		20		12
great rhombicub.	48	72		12		8	6	
icosidodecahedron	30	60	20		12			
small rhombicos.	60	120	20	30	12			
small rhombicub.	24	48	8	18				
snub cube	24	60	32	6				
snub dodecahedron	60	150	80		12			
trunc. cube	24	36	8				6	
trunc. dodec.	60	90	20					12
trunc. icosahedron	60	90			12	20		
trunc. octahedron	24	36		6		8		
trunc. tetrahedron	12	18	4			4		

Let r be the Inradius, ρ the Midradius, and R the Circumradius. The following tables give the analytic and numerical values of r, ρ, and R for the Archimedean solids with Edges of unit length.

Solid	r
cuboctahedron	$\frac{3}{4}$
great rhombicosidodecahedron	$\frac{1}{241}(105+6\sqrt{5})\sqrt{31+12\sqrt{5}}$
great rhombicuboctahedron	$\frac{3}{97}(14+\sqrt{2})\sqrt{13+6\sqrt{2}}$
icosidodecahedron	$\frac{1}{8}(5+3\sqrt{5})$
small rhombicosidodecahedron	$\frac{1}{41}(15+2\sqrt{5})\sqrt{11+4\sqrt{5}}$
small rhombicuboctahedron	$\frac{1}{17}(6+\sqrt{2})\sqrt{5+2\sqrt{2}}$
snub cube	*
snub dodecahedron	*
truncated cube	$\frac{1}{17}(5+2\sqrt{2})\sqrt{7+4\sqrt{2}}$
truncated dodecahedron	$\frac{5}{488}(17\sqrt{2}+3\sqrt{10})\sqrt{37+15\sqrt{5}}$
truncated icosahedron	$\frac{9}{872}(21+\sqrt{5})\sqrt{58+18\sqrt{5}}$
truncated octahedron	$\frac{9}{20}\sqrt{10}$
truncated tetrahedron	$\frac{9}{44}\sqrt{22}$

Solid	ρ	R
cuboctahedron	$\frac{1}{2}\sqrt{3}$	1
great rhombicosidodecahedron	$\frac{1}{2}\sqrt{30+12\sqrt{5}}$	$\frac{1}{2}\sqrt{31+12\sqrt{5}}$
great rhombicuboctahedron	$\frac{1}{2}\sqrt{12+6\sqrt{2}}$	$\frac{1}{2}\sqrt{13+6\sqrt{2}}$
icosidodecahedron	$\frac{1}{2}\sqrt{5+2\sqrt{5}}$	$\frac{1}{2}(1+\sqrt{5})$
small rhombicosidodecahedron	$\frac{1}{2}\sqrt{10+4\sqrt{5}}$	$\frac{1}{2}\sqrt{11+4\sqrt{5}}$
small rhombicuboctahedron	$\frac{1}{2}\sqrt{4+2\sqrt{2}}$	$\frac{1}{2}\sqrt{5+2\sqrt{2}}$
snub cube	*	*
snub dodecahedron	*	*
truncated cube	$\frac{1}{2}(2+\sqrt{2})$	$\frac{1}{2}\sqrt{7+4\sqrt{2}}$
truncated dodecahedron	$\frac{1}{4}(5+3\sqrt{5})$	$\frac{1}{4}\sqrt{74+30\sqrt{5}}$
truncated icosahedron	$\frac{3}{4}(1+\sqrt{5})$	$\frac{1}{4}\sqrt{58+18\sqrt{5}}$
truncated octahedron	$\frac{3}{2}$	$\frac{1}{2}\sqrt{10}$
truncated tetrahedron	$\frac{3}{4}\sqrt{2}$	$\frac{1}{4}\sqrt{22}$

*The complicated analytic expressions for the CIRCUMRADII of these solids are given in the entries for the SNUB CUBE and SNUB DODECAHEDRON.

Solid	r	ρ	R
cuboctahedron	0.75	0.86603	1
great rhombicosidodecahedron	3.73665	3.76938	3.80239
great rhombicuboctahedron	2.20974	2.26303	2.31761
icosidodecahedron	1.46353	1.53884	1.61803
small rhombicosidodecahedron	2.12099	2.17625	2.23295
small rhombicuboctahedron	1.22026	1.30656	1.39897
snub cube	1.15766	1.24722	1.34371
snub dodecahedron	2.03987	2.09705	2.15583
truncated cube	1.63828	1.70711	1.77882
truncated dodecahedron	2.88526	2.92705	2.96945
truncated icosahedron	2.37713	2.42705	2.47802
truncated octahedron	1.42302	1.5	1.58114
truncated tetrahedron	0.95940	1.06066	1.17260

The DUALS of the Archimedean solids, sometimes called the CATALAN SOLIDS, are given in the following table.

Archimedean Solid	Dual
rhombicosidodecahedron	deltoidal hexecontahedron
small rhombicuboctahedron	deltoidal icositetrahedron
great rhombicuboctahedron	disdyakis dodecahedron
great rhombicosidodecahedron	disdyakis triacontahedron
truncated icosahedron	pentakis dodecahedron
snub dodecahedron (laevo)	pentagonal hexecontahedron (dextro)
snub cube (laevo)	pentagonal icositetrahedron (dextro)
cuboctahedron	rhombic dodecahedron
icosidodecahedron	rhombic triacontahedron
truncated octahedron	tetrakis hexahedron
truncated dodecahedron	triakis icosahedron
truncated cube	triakis octahedron
truncated tetrahedron	triakis tetrahedron

Here are the Archimedean DUALS (Holden 1971, Pearce 1978) displayed in alphabetical order (left to right, then continuing to the next row).

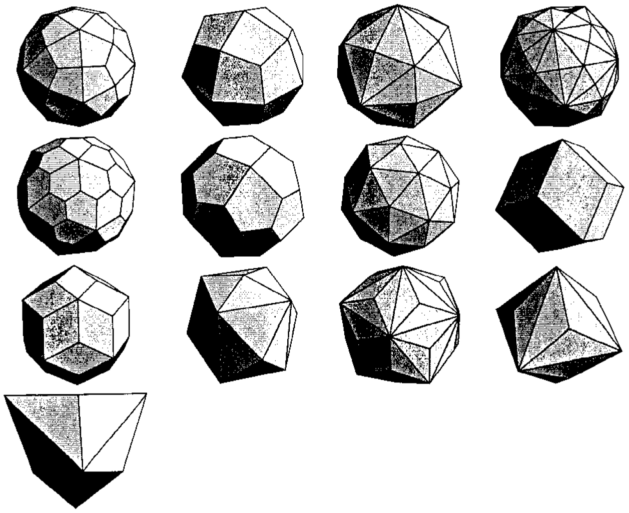

Here are the Archimedean solids paired with their DUALS.

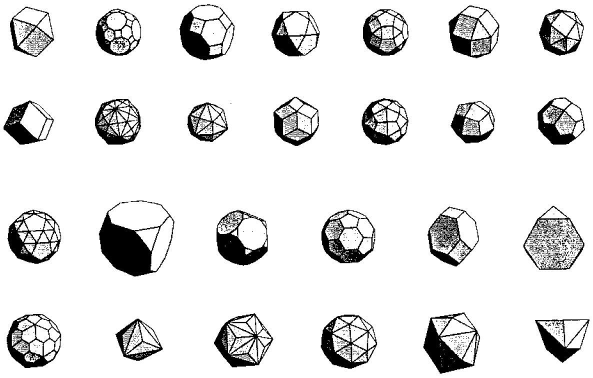

The Archimedean solids and their DUALS are all CANONICAL POLYHEDRA.

see also ARCHIMEDEAN SOLID STELLATION, CATALAN SOLID, DELTAHEDRON, JOHNSON SOLID, KEPLER-POINSOT SOLID, PLATONIC SOLID, SEMIREGULAR POLYHEDRON, UNIFORM POLYHEDRON

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 136, 1987.

Behnke, H.; Bachman, F.; Fladt, K.; and Kunle, H. (Eds.). *Fundamentals of Mathematics, Vol. 2.* Cambridge, MA: MIT Press, pp. 269–286, 1974.

Catalan, E. "Mémoire sur la Théorie des Polyèdres." *J. l'École Polytechnique (Paris)* **41**, 1–71, 1865.

Coxeter, H. S. M. "The Pure Archimedean Polytopes in Six and Seven Dimensions." *Proc. Cambridge Phil. Soc.* **24**, 1–9, 1928.

Coxeter, H. S. M. "Regular and Semi-Regular Polytopes I." *Math. Z.* **46**, 380–407, 1940.

Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, 1973.

Coxeter, H. S. M.; Longuet-Higgins, M. S.; and Miller, J. C. P. "Uniform Polyhedra." *Phil. Trans. Roy. Soc. London Ser. A* **246**, 401–450, 1954.

Critchlow, K. *Order in Space: A Design Source Book.* New York: Viking Press, 1970.

Cromwell, P. R. *Polyhedra.* New York: Cambridge University Press, pp. 79–86, 1997.
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., 1989.
Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, p. 54, 1991.
Kepler, J. "Harmonice Mundi." *Opera Omnia, Vol. 5.* Frankfurt, pp. 75–334, 1864.
Kraitchik, M. *Mathematical Recreations.* New York: W. W. Norton, pp. 199–207, 1942.
Le, Ha. "Archimedean Solids." `http://daisy.uwaterloo.ca/~hqle/archimedean.html`.
Pearce, P. *Structure in Nature is a Strategy for Design.* Cambridge, MA: MIT Press, pp. 34–35, 1978.
Pugh, A. *Polyhedra: A Visual Approach.* Berkeley: University of California Press, p. 25, 1976.
Rawles, B. A. "Platonic and Archimedean Solids—Faces, Edges, Areas, Vertices, Angles, Volumes, Sphere Ratios." `http://www.intent.com/sg/polyhedra.html`.
Rorres, C. "Archimedean Solids: Pappus." `http://www.mcs.drexel.edu/~crorres/Archimedes/Solids/Pappus.html`.
Steinitz, E. and Rademacher, H. *Vorlesungen über die Theorie der Polyheder.* Berlin, p. 11, 1934.
Stott, A. B. *Verhandelingen der Konniklijke Akad. Wetenschappen, Amsterdam* **11**, 1910.
Walsh, T. R. S. "Characterizing the Vertex Neighbourhoods of Semi-Regular Polyhedra." *Geometriae Dedicata* **1**, 117–123, 1972.
Wenninger, M. J. *Polyhedron Models.* New York: Cambridge University Press, 1989.

Archimedean Solid Stellation

A large class of POLYHEDRA which includes the DODECADODECAHEDRON and GREAT ICOSIDODECAHEDRON. No complete enumeration (even with restrictive uniqueness conditions) has been worked out.

References

Coxeter, H. S. M.; Longuet-Higgins, M. S.; and Miller, J. C. P. "Uniform Polyhedra." *Phil. Trans. Roy. Soc. London Ser. A* **246**, 401–450, 1954.
Wenninger, M. J. *Polyhedron Models.* New York: Cambridge University Press, pp. 66–72, 1989.

Archimedean Spiral

A SPIRAL with POLAR equation

$$r = a\theta^{1/m},$$

where r is the radial distance, θ is the polar angle, and m is a constant which determines how tightly the spiral is "wrapped." The CURVATURE of an Archimedean spiral is given by

$$\kappa = \frac{n\theta^{1-1/n}(1+n+n^2\theta^2)}{a(1+n^2\theta^2)^{3/2}}.$$

Various special cases are given in the following table.

Name	m
lituus	-2
hyperbolic spiral	-1
Archimedes' spiral	1
Fermat's spiral	2

see also ARCHIMEDES' SPIRAL, DAISY, FERMAT'S SPIRAL, HYPERBOLIC SPIRAL, LITUUS, SPIRAL

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 69–70, 1993.
Lauweirer, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 59–60, 1991.
Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 186 and 189, 1972.
Lee, X. "Archimedean Spiral." `http://www.best.com/~xah/Special Plane Curves _ dir / Archimedean Spiral _ dir / archimedeanSpiral.html`.
Lockwood, E. H. *A Book of Curves.* Cambridge, England: Cambridge University Press, p. 175, 1967.
MacTutor History of Mathematics Archive. "Spiral of Archimedes." `http: // www - groups . dcs . st - and . ac . uk / ~history/Curves/Spiral.html`.
Pappas, T. "The Spiral of Archimedes." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 149, 1989.

Archimedean Spiral Inverse Curve

The INVERSE CURVE of the ARCHIMEDEAN SPIRAL

$$r = a\theta^{1/m}$$

with INVERSION CENTER at the origin and inversion RADIUS k is the ARCHIMEDEAN SPIRAL

$$r = ka\theta^{1/m}.$$

Archimedes' Spiral

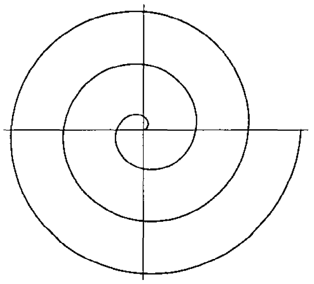

An ARCHIMEDEAN SPIRAL with POLAR equation

$$r = a\theta.$$

This spiral was studied by Conon, and later by Archimedes in *On Spirals* about 225 BC. Archimedes was able to work out the lengths of various tangents to the spiral.

Archimedes' spiral can be used for COMPASS and STRAIGHTEDGE division of an ANGLE into n parts (including ANGLE TRISECTION) and can also be used for CIRCLE SQUARING. In addition, the curve can be used as a cam to convert uniform circular motion into uniform linear motion. The cam consists of one arch of the spiral above the x-AXIS together with its reflection in the x-AXIS. Rotating this with uniform angular velocity about its center will result in uniform linear motion of the point where it crosses the y-AXIS.

see also ARCHIMEDEAN SPIRAL

References

Gardner, M. *The Unexpected Hanging and Other Mathematical Diversions.* Chicago, IL: Chicago University Press, pp. 106–107, 1991.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 69–70, 1993.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 186–187, 1972.

Lockwood, E. H. *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 173–164, 1967.

Archimedes' Spiral Inverse

Taking the ORIGIN as the INVERSION CENTER, ARCHIMEDES' SPIRAL $r = a\theta$ inverts to the HYPERBOLIC SPIRAL $r = a/\theta$.

Archimedean Valuation

A VALUATION for which $|x| \leq 1$ IMPLIES $|1+x| \leq C$ for the constant $C = 1$ (independent of x). Such a VALUATION *does not* satisfy the strong TRIANGLE INEQUALITY

$$|x+y| \leq \max(|x|, |y|).$$

Arcsecant

see INVERSE SECANT

Arcsine

see INVERSE SINE

Arctangent

see INVERSE TANGENT

Area

The AREA of a SURFACE is the amount of material needed to "cover" it completely. The AREA of a TRIANGLE is given by

$$A_\Delta = \tfrac{1}{2}lh, \tag{1}$$

where l is the base length and h is the height, or by HERON'S FORMULA

$$A_\Delta = \sqrt{s(s-a)(s-b)(s-c)}, \tag{2}$$

where the side lengths are a, b, and c and s the SEMIPERIMETER. The AREA of a RECTANGLE is given by

$$A_{\text{rectangle}} = ab, \tag{3}$$

where the sides are length a and b. This gives the special case of

$$A_{\text{square}} = a^2 \tag{4}$$

for the SQUARE. The AREA of a regular POLYGON with n sides and side length s is given by

$$A_{n\text{-gon}} = \tfrac{1}{4}ns^2 \cot\left(\frac{\pi}{n}\right). \tag{5}$$

CALCULUS and, in particular, the INTEGRAL, are powerful tools for computing the AREA between a curve $f(x)$ and the x-AXIS over an INTERVAL $[a, b]$, giving

$$A = \int_a^b f(x)\,dx. \tag{6}$$

The AREA of a POLAR curve with equation $r = r(\theta)$ is

$$A = \tfrac{1}{2}\int r^2\,d\theta. \tag{7}$$

Written in CARTESIAN COORDINATES, this becomes

$$A = \frac{1}{2}\int \left(x\frac{dy}{dt} - y\frac{dx}{dt}\right) dt \tag{8}$$

$$= \frac{1}{2}\int (x\,dy - y\,dx). \tag{9}$$

For the AREA of special surfaces or regions, see the entry for that region. The generalization of AREA to 3-D is called VOLUME, and to higher DIMENSIONS is called CONTENT.

see also ARC LENGTH, AREA ELEMENT, CONTENT, SURFACE AREA, VOLUME

References

Gray, A. "The Intuitive Idea of Area on a Surface." §13.2 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 259–260, 1993.

Area Element

The area element for a SURFACE with RIEMANNIAN METRIC

$$ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2$$

is

$$dA = \sqrt{EG - F^2}\,du \wedge dv,$$

where $du \wedge dv$ is the WEDGE PRODUCT.

see also AREA, LINE ELEMENT, RIEMANNIAN METRIC, VOLUME ELEMENT

References

Gray, A. "The Intuitive Idea of Area on a Surface." §13.2 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 259–260, 1993.

Area-Preserving Map

A MAP F from $\mathbb{R}^n$ to $\mathbb{R}^n$ is AREA-preserving if

$$m(F(A)) = m(A)$$

for every subregion A of $\mathbb{R}^n$, where $m(A)$ is the n-D MEASURE of A. A linear transformation is AREA-preserving if its corresponding DETERMINANT is equal to 1.

see also CONFORMAL MAP, SYMPLECTIC MAP

Area Principle

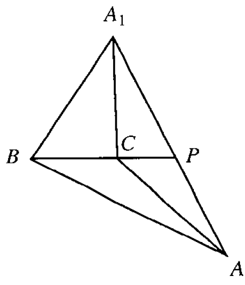

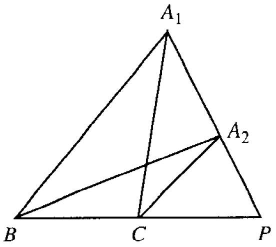

The "AREA principle" states that

$$\frac{|A_1P|}{|A_2P|} = \frac{|A_1BC|}{|A_2BC|}. \tag{1}$$

This can also be written in the form

$$\left[\frac{A_1P}{A_2P}\right] = \left[\frac{A_1BC}{A_2BC}\right], \tag{2}$$

where

$$\left[\frac{AB}{CD}\right] \tag{3}$$

is the ratio of the lengths $[A, B]$ and $[C, D]$ for $AB||CD$ with a PLUS or MINUS SIGN depending on if these segments have the same or opposite directions, and

$$\left[\frac{ABC}{DEFG}\right] \tag{4}$$

is the RATIO of signed AREAS of the TRIANGLES. Grünbaum and Shepard show that CEVA'S THEOREM, HOEHN'S THEOREM, and MENELAUS' THEOREM are the consequences of this result.

see also CEVA'S THEOREM, HOEHN'S THEOREM, MENELAUS' THEOREM, SELF-TRANSVERSALITY THEOREM

References
Grünbaum, B. and Shepard, G. C. "Ceva, Menelaus, and the Area Principle." *Math. Mag.* **68**, 254–268, 1995.

Areal Coordinates

TRILINEAR COORDINATES normalized so that

$$t_1 + t_2 + t_3 = 1.$$

When so normalized, they become the AREAS of the TRIANGLES PA_1A_2, PA_1A_3, and PA_2A_3, where P is the point whose coordinates have been specified.

Arf Invariant

A LINK invariant which always has the value 0 or 1. A KNOT has ARF INVARIANT 0 if the KNOT is "pass equivalent" to the UNKNOT and 1 if it is pass equivalent to the TREFOIL KNOT. If K_+, K_-, and L are projections which are identical outside the region of the crossing diagram, and K_+ and K_- are KNOTS while L is a 2-component LINK with a nonintersecting crossing diagram where the two left and right strands belong to the different LINKS, then

$$a(K_+) = a(K_-) + l(L_1, L_2), \tag{1}$$

where l is the LINKING NUMBER of L_1 and L_2. The Arf invariant can be determined from the ALEXANDER POLYNOMIAL or JONES POLYNOMIAL for a KNOT. For Δ_K the ALEXANDER POLYNOMIAL of K, the Arf invariant is given by

$$\Delta_K(-1) \equiv \begin{cases} 1 \pmod 8 & \text{if } \operatorname{Arf}(K) = 0 \\ 5 \pmod 8 & \text{if } \operatorname{Arf}(K) = 1 \end{cases} \tag{2}$$

(Jones 1985). For the JONES POLYNOMIAL W_K of a KNOT K,

$$\operatorname{Arf}(K) = W_K(i) \tag{3}$$

(Jones 1985), where i is the IMAGINARY NUMBER.

References
Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 223–231, 1994.
Jones, V. "A Polynomial Invariant for Knots via von Neumann Algebras." *Bull. Amer. Math. Soc.* **12**, 103–111, 1985.
Weisstein, E. W. "Knots." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Knots.m`.

Argand Diagram

A plot of COMPLEX NUMBERS as points

$$z = x + iy$$

using the x-AXIS as the REAL axis and y-AXIS as the IMAGINARY axis. This is also called the COMPLEX PLANE or ARGAND PLANE.

Argand Plane

see ARGAND DIAGRAM

Argoh's Conjecture

Let B_k be the kth BERNOULLI NUMBER. Then does

$$nB_{n-1} \equiv -1 \pmod n$$

IFF n is PRIME? For example, for $n = 1, 2, \ldots, nB_{n-1} \pmod n$ is 0, −1, −1, 0, −1, 0, −1, 0, −3, 0, −1, ... There are no counterexamples less than $n = 5,600$. Any counterexample to Argoh's conjecture would be a contradiction to GIUGA'S CONJECTURE, and vice versa.

see also BERNOULLI NUMBER, GIUGA'S CONJECTURE

References
Borwein, D.; Borwein, J. M.; Borwein, P. B.; and Girgensohn, R. "Giuga's Conjecture on Primality." *Amer. Math. Monthly* **103**, 40–50, 1996.

Argument Addition Relation

A mathematical relationship relating $f(x+y)$ to $f(x)$ and $f(y)$.

see also ARGUMENT MULTIPLICATION RELATION, RECURRENCE RELATION, REFLECTION RELATION, TRANSLATION RELATION

Argument (Complex Number)

A COMPLEX NUMBER z may be represented as

$$z \equiv x + iy = |z|e^{i\theta}, \tag{1}$$

where $|z|$ is called the MODULUS of z, and θ is called the argument

$$\arg(x+iy) \equiv \tan^{-1}\left(\frac{y}{x}\right). \tag{2}$$

Therefore,

$$\begin{aligned}\arg(zw) &= \arg(|z|e^{i\theta_z}|w|e^{i\theta_w}) = \arg(e^{i\theta_z}e^{i\theta_w})\\ &= \arg[e^{i(\theta_z+\theta_w)}] = \arg(z) + \arg(w).\end{aligned} \tag{3}$$

Extending this procedure gives

$$\arg(z^n) = n\arg(z). \tag{4}$$

The argument of a COMPLEX NUMBER is sometimes called the PHASE.

see also AFFIX, COMPLEX NUMBER, DE MOIVRE'S IDENTITY, EULER FORMULA, MODULUS (COMPLEX NUMBER), PHASE, PHASOR

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 16, 1972.

Argument (Elliptic Integral)

Given an AMPLITUDE ϕ in an ELLIPTIC INTEGRAL, the argument u is defined by the relation

$$\phi \equiv \operatorname{am} u.$$

see also AMPLITUDE, ELLIPTIC INTEGRAL

Argument (Function)

An argument of a FUNCTION $f(x_1, \ldots, x_n)$ is one of the n parameters on which the function's value depends. For example, the SINE $\sin x$ is a one-argument function, the BINOMIAL COEFFICIENT $\binom{n}{m}$ is a two-argument function, and the HYPERGEOMETRIC FUNCTION ${}_2F_1(a, b; c; z)$ is a four-argument function.

Argument Multiplication Relation

A mathematical relationship relating $f(nx)$ to $f(x)$ for INTEGER n.

see also ARGUMENT ADDITION RELATION, RECURRENCE RELATION, REFLECTION RELATION, TRANSLATION RELATION

Argument Principle

If $f(z)$ is MEROMORPHIC in a region R enclosed by a curve γ, let N be the number of COMPLEX ROOTS of $f(z)$ in γ, and P be the number of POLES in γ, then

$$N - P = \frac{1}{2\pi i}\int_\gamma \frac{f'(z)\,dz}{f(z)}.$$

Defining $w \equiv f(z)$ and $\sigma \equiv f(\gamma)$ gives

$$N - P = \frac{1}{2\pi i}\int_\sigma \frac{dw}{w}.$$

see also VARIATION OF ARGUMENT

References

Duren, P.; Hengartner, W.; and Laugessen, R. S. "The Argument Principle for Harmonic Functions." *Math. Mag.* **103**, 411–415, 1996.

Argument Variation

see VARIATION OF ARGUMENT

Aristotle's Wheel Paradox

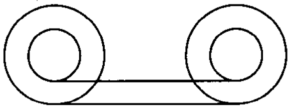

A PARADOX mentioned in the Greek work *Mechanica,* dubiously attributed to Aristotle. Consider the above diagram depicting a wheel consisting of two concentric CIRCLES of different DIAMETERS (a wheel within a wheel). There is a 1:1 correspondence of points on the large CIRCLE with points on the small CIRCLE, so the wheel should travel the same distance regardless of whether it is rolled from left to right on the top straight line or on the bottom one. This seems to imply that the two CIRCUMFERENCES of different sized CIRCLES are equal, which is impossible.

The fallacy lies in the assumption that a 1:1 correspondence of points means that two curves must have the same length. In fact, the CARDINALITIES of points in a LINE SEGMENT of any length (or even an INFINITE LINE, a PLANE, a 3-D SPACE, or an infinite dimensional EUCLIDEAN SPACE) are all the same: $\aleph_1$ (ALEPH-1), so the points of any of these can be put in a ONE-TO-ONE correspondence with those of any other.

see also ZENO'S PARADOXES

References

Ballew, D. "The Wheel of Aristotle." *Math. Teacher* **65**, 507–509, 1972.

Costabel, P. "The Wheel of Aristotle and French Consideration of Galileo's Arguments." *Math. Teacher* **61**, 527–534, 1968.

Drabkin, I. "Aristotle's Wheel: Notes on the History of the Paradox." *Osiris* **9**, 162–198, 1950.

Gardner, M. *Wheels, Life, and other Mathematical Amusements.* New York: W. H. Freeman, pp. 2–4, 1983.

Pappas, T. "The Wheel of Paradox Aristotle." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 202, 1989.

vos Savant, M. *The World's Most Famous Math Problem.* New York: St. Martin's Press, pp. 48–50, 1993.

Arithmetic

The branch of mathematics dealing with INTEGERS or, more generally, numerical computation. Arithmetical operations include ADDITION, CONGRUENCE calculation, DIVISION, FACTORIZATION, MULTIPLICATION, POWER computation, ROOT extraction, and SUBTRACTION.

The FUNDAMENTAL THEOREM OF ARITHMETIC, also called the UNIQUE FACTORIZATION THEOREM, states that any POSITIVE INTEGER can be represented in exactly one way as a PRODUCT of PRIMES.

The LÖWENHEIMER-SKOLEM THEOREM, which is a fundamental result in MODEL THEORY, establishes the existence of "nonstandard" models of arithmetic.

see also ALGEBRA, CALCULUS, FUNDAMENTAL THEOREM OF ARITHMETIC, GROUP THEORY, HIGHER ARITHMETIC, LINEAR ALGEBRA, LÖWENHEIMER-SKOLEM THEOREM, MODEL THEORY, NUMBER THEORY, TRIGONOMETRY

References

Karpinski, L. C. *The History of Arithmetic.* Chicago, IL: Rand, McNally, & Co., 1925.

Maxfield, J. E. and Maxfield, M. W. *Abstract Algebra and Solution by Radicals.* Philadelphia, PA: Saunders, 1992.

Thompson, J. E. *Arithmetic for the Practical Man.* New York: Van Nostrand Reinhold, 1973.

Arithmetic-Geometric Mean

The arithmetic-geometric mean (AGM) $M(a,b)$ of two numbers a and b is defined by starting with $a_0 \equiv a$ and $b_0 \equiv b$, then iterating

$$a_{n+1} = \tfrac{1}{2}(a_n + b_n) \tag{1}$$

$$b_{n+1} = \sqrt{a_n b_n} \tag{2}$$

until $a_n = b_n$. a_n and b_n converge towards each other since

$$\begin{aligned} a_{n+1} - b_{n+1} &= \tfrac{1}{2}(a_n + b_n) - \sqrt{a_n b_n} \\ &= \frac{a_n - 2\sqrt{a_n b_n} + b_n}{2}. \end{aligned} \tag{3}$$

But $\sqrt{b_n} < \sqrt{a_n}$, so

$$2b_n < 2\sqrt{a_n b_n}. \tag{4}$$

Now, add $a_n - b_n - 2\sqrt{a_n b_n}$ to each side

$$a_n + b_n - 2\sqrt{a_n b_n} < a_n - b_n, \tag{5}$$

so

$$a_{n+1} - b_{n+1} < \tfrac{1}{2}(a_n - b_n). \tag{6}$$

The AGM is very useful in computing the values of complete ELLIPTIC INTEGRALS and can also be used for finding the INVERSE TANGENT. The special value $1/M(1,\sqrt{2}\,)$ is called GAUSS'S CONSTANT.

The AGM has the properties

$$\lambda M(a,b) = M(\lambda a, \lambda b) \tag{7}$$

$$M(a,b) = M\left(\tfrac{1}{2}(a+b), \sqrt{ab}\right) \tag{8}$$

$$M(1, \sqrt{1-x^2}\,) = M(1+x, 1-x) \tag{9}$$

$$M(1,b) = \frac{1+b}{2} M\left(1, \frac{2\sqrt{b}}{1+b}\right). \tag{10}$$

The Legendre form is given by

$$M(1,x) = \prod_{n=0}^{\infty} \tfrac{1}{2}(1+k_n), \tag{11}$$

where $k_0 \equiv x$ and

$$k_{n+1} \equiv \frac{2\sqrt{k_n}}{1+k_n}. \tag{12}$$

Solutions to the differential equation

$$(x^3 - x)\frac{d^2y}{dx^2} + (3x^2 - 1)\frac{dy}{dx} + xy = 0 \tag{13}$$

are given by $[M(1+x, 1-x)]^{-1}$ and $[M(1,x)]^{-1}$.

A generalization of the ARITHMETIC-GEOMETRIC MEAN is

$$I_p(a,b) = \int_0^\infty \frac{x^{p-2}\,dx}{(x^p + a^p)^{1/p}(x^p + b^p)^{(p-1)/p}}, \tag{14}$$

which is related to solutions of the differential equation

$$x(1-x^p)Y'' + [1-(p+1)x^p]Y' - (p-1)x^{p-1}Y = 0. \tag{15}$$

When $p = 2$ or $p = 3$, there is a modular transformation for the solutions of (15) that are bounded as $x \to 0$. Letting $J_p(x)$ be one of these solutions, the transformation takes the form

$$J_p(\lambda) = \mu J_p(x), \tag{16}$$

where

$$\lambda = \frac{1-u}{1+(p-1)u} \tag{17}$$

$$\mu = \frac{1+(p-1)u}{p} \tag{18}$$

and

$$x^p + u^p = 1. \tag{19}$$

The case $p = 2$ gives the ARITHMETIC-GEOMETRIC MEAN, and $p = 3$ gives a cubic relative discussed by Borwein and Borwein (1990, 1991) and Borwein (1996) in which, for $a, b > 0$ and $I(a,b)$ defined by

$$I(a,b) = \int_0^\infty \frac{t\,dt}{[(a^3+t^3)(b^3+t^3)^2]^{1/3}}, \tag{20}$$

$$I(a,b) = I\left(\frac{a+2b}{3}, \left[\frac{b}{3}(a^2+ab+b^2)\right]\right). \qquad (21)$$

For iteration with $a_0 = a$ and $b_0 = b$ and

$$a_{n+1} = \frac{a_n + 2b_n}{3} \qquad (22)$$

$$b_{n+1} = \frac{b_n}{3}({a_n}^2 + a_n b_n + {b_n}^2), \qquad (23)$$

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = \frac{I(1,1)}{I(a,b)}. \qquad (24)$$

Modular transformations are known when $p = 4$ and $p = 6$, but they do not give identities for $p = 6$ (Borwein 1996).

see also ARITHMETIC-HARMONIC MEAN

References

Abramowitz, M. and Stegun, C. A. (Eds.). "The Process of the Arithmetic-Geometric Mean." §17.6 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 571 ad 598–599, 1972.

Borwein, J. M. Problem 10281. "A Cubic Relative of the AGM." *Amer. Math. Monthly* **103**, 181–183, 1996.

Borwein, J. M. and Borwein, P. B. "A Remarkable Cubic Iteration." In *Computational Method & Function Theory: Proc. Conference Held in Valparaiso, Chile, March 13–18, 1989*0387527680 (Ed. A. Dold, B. Eckmann, F. Takens, E. B Saff, S. Ruscheweyh, L. C. Salinas, L. C., and R. S. Varga). New York: Springer-Verlag, 1990.

Borwein, J. M. and Borwein, P. B. "A Cubic Counterpart of Jacobi's Identity and the AGM." *Trans. Amer. Math. Soc.* **323**, 691–701, 1991.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 906–907, 1992.

Arithmetic Geometry

A vaguely defined branch of mathematics dealing with VARIETIES, the MORDELL CONJECTURE, ARAKELOV THEORY, and ELLIPTIC CURVES.

References

Cornell, G. and Silverman, J. H. (Eds.). *Arithmetic Geometry.* New York: Springer-Verlag, 1986.

Lorenzini, D. *An Invitation to Arithmetic Geometry.* Providence, RI: Amer. Math. Soc., 1996.

Arithmetic-Harmonic Mean

Let

$$a_{n+1} = \tfrac{1}{2}(a_n + b_n) \qquad (1)$$

$$b_{n+1} = \frac{2a_n b_2}{a_n + b_n}. \qquad (2)$$

Then

$$A(a_0, b_0) = \lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = \sqrt{a_0 b_0}, \qquad (3)$$

which is just the GEOMETRIC MEAN.

Arithmetic-Logarithmic-Geometric Mean Inequality

$$\frac{a+b}{2} > \frac{b-a}{\ln b - \ln a} > \sqrt{ab}.$$

see also NAPIER'S INEQUALITY

References

Nelson, R. B. "Proof without Words: The Arithmetic-Logarithmic-Geometric Mean Inequality." *Math. Mag.* **68**, 305, 1995.

Arithmetic Mean

For a CONTINUOUS DISTRIBUTION function, the arithmetic mean of the population, denoted μ, $\bar{x}$, $\langle x \rangle$, or $A(x)$, is given by

$$\mu = \langle f(x) \rangle \equiv \int_{-\infty}^{\infty} P(x) f(x)\, dx, \qquad (1)$$

where $\langle x \rangle$ is the EXPECTATION VALUE. For a DISCRETE DISTRIBUTION,

$$\mu = \langle f(x) \rangle \equiv \frac{\sum_{n=0}^{N} P(x_n) f(x_n)}{\sum_{n=0}^{N} P(x_n)} = \sum_{n=0}^{N} P(x_n) f(x_n). \qquad (2)$$

The population mean satisfies

$$\langle f(x) + g(x) \rangle = \langle f(x) \rangle + \langle g(x) \rangle \qquad (3)$$

$$\langle c f(x) \rangle = c \langle f(x) \rangle, \qquad (4)$$

and

$$\langle f(x) g(y) \rangle = \langle f(x) \rangle \langle g(y) \rangle \qquad (5)$$

if x and y are INDEPENDENT STATISTICS. The "sample mean," which is the mean estimated from a statistical sample, is an UNBIASED ESTIMATOR for the population mean.

For small samples, the mean is more efficient than the MEDIAN and approximately $\pi/2$ less (Kenney and Keeping 1962, p. 211). A general expression which often holds approximately is

$$\text{mean} - \text{mode} \approx 3(\text{mean} - \text{median}). \qquad (6)$$

Given a set of samples $\{x_i\}$, the arithmetic mean is

$$A(x) \equiv \bar{x} \equiv \mu \equiv \langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i. \qquad (7)$$

Hoehn and Niven (1985) show that

$$A(a_1 + c, a_2 + c, \ldots, a_n + c) = c + A(a_1, a_2, \ldots, a_n) \qquad (8)$$

for any POSITIVE constant c. The arithmetic mean satisfies

$$A \geq G \geq H, \qquad (9)$$

where G is the GEOMETRIC MEAN and H is the HARMONIC MEAN (Hardy *et al.* 1952; Mitrinović 1970; Beckenbach and Bellman 1983; Bullen *et al.* 1988; Mitrinović *et al.* 1993; Alzer 1996). This can be shown as follows. For $a, b > 0$,

$$\left(\frac{1}{\sqrt{a}} - \frac{1}{\sqrt{b}}\right)^2 \geq 0 \tag{10}$$

$$\frac{1}{a} - \frac{2}{\sqrt{ab}} + \frac{1}{b} \geq 0 \tag{11}$$

$$\frac{1}{a} + \frac{1}{b} \geq \frac{2}{\sqrt{ab}} \tag{12}$$

$$\frac{2}{\frac{1}{a} + \frac{1}{b}} \geq \sqrt{ab} \tag{13}$$

$$H \geq G, \tag{14}$$

with equality IFF $b = a$. To show the second part of the inequality,

$$(\sqrt{a} - \sqrt{b}\,)^2 = a - 2\sqrt{ab} + b \geq 0 \tag{15}$$

$$\frac{a+b}{2} \geq \sqrt{ab} \tag{16}$$

$$A \geq H, \tag{17}$$

with equality IFF $a = b$. Combining (14) and (17) then gives (9).

Given n independent random GAUSSIAN DISTRIBUTED variates x_i, each with population mean $\mu_i = \mu$ and VARIANCE ${\sigma_i}^2 = \sigma^2$,

$$\bar{x} \equiv \frac{1}{N} \sum_{i=1}^{N} x_i \tag{18}$$

$$\begin{aligned} \langle \bar{x} \rangle &= \frac{1}{N} \left\langle \sum_{i=1}^{N} x_i \right\rangle = \frac{1}{N} \sum_{i=1}^{N} \langle x_i \rangle \\ &= \frac{1}{N} \sum_{i=1}^{N} \mu = \frac{1}{N}(N\mu) = \mu, \end{aligned} \tag{19}$$

so the sample mean is an UNBIASED ESTIMATOR of population mean. However, the distribution of $\bar{x}$ depends on the sample size. For large samples, $\bar{x}$ is approximately NORMAL. For small samples, STUDENT'S t-DISTRIBUTION should be used.

The VARIANCE of the population mean is independent of the distribution.

$$\begin{aligned} \operatorname{var}(\bar{x}) &= \operatorname{var}\left(\frac{1}{n} \sum_{i=1}^{N} x_i\right) = \frac{1}{N^2} \operatorname{var}\left(\sum_{i=1}^{N} x_i\right) \\ &= \frac{1}{N^2} \sum_{i=1}^{n} \operatorname{var}(x_i) = \left(\frac{1}{N^2}\right) \sum_{i=1}^{N} \sigma^2 = \frac{\sigma^2}{N}. \end{aligned} \tag{20}$$

From k-STATISTICS for a GAUSSIAN DISTRIBUTION, the UNBIASED ESTIMATOR for the VARIANCE is given by

$$\sigma^2 = \frac{N}{N-1} s^2, \tag{21}$$

where

$$s \equiv \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2, \tag{22}$$

so

$$\operatorname{var}(\bar{x}) = \frac{s^2}{N-1}. \tag{23}$$

The SQUARE ROOT of this,

$$\sigma_x = \frac{s}{\sqrt{N-1}}, \tag{24}$$

is called the STANDARD ERROR.

$$\operatorname{var}(\bar{x}) \equiv \left\langle \bar{x}^2 \right\rangle - \langle \bar{x} \rangle^2, \tag{25}$$

so

$$\left\langle \bar{x}^2 \right\rangle = \operatorname{var}(\bar{x}) + \langle \bar{x} \rangle^2 = \frac{\sigma^2}{N} + \mu^2. \tag{26}$$

see also ARITHMETIC-GEOMETRIC MEAN, ARITHMETIC-HARMONIC MEAN, CARLEMAN'S INEQUALITY, CUMULANT, GENERALIZED MEAN, GEOMETRIC MEAN, HARMONIC MEAN, HARMONIC-GEOMETRIC MEAN, KURTOSIS, MEAN, MEAN DEVIATION, MEDIAN (STATISTICS), MODE, MOMENT, QUADRATIC MEAN, ROOT-MEAN-SQUARE, SAMPLE VARIANCE, SKEWNESS, STANDARD DEVIATION, TRIMEAN, VARIANCE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 10, 1972.

Alzer, H. "A Proof of the Arithmetic Mean–Geometric Mean Inequality." *Amer. Math. Monthly* **103**, 585, 1996.

Beckenbach, E. F. and Bellman, R. *Inequalities.* New York: Springer-Verlag, 1983.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 471, 1987.

Bullen, P. S.; Mitrinović, D. S.; and Vasić, P. M. *Means & Their Inequalities.* Dordrecht, Netherlands: Reidel, 1988.

Hardy, G. H.; Littlewood, J. E.; and Pólya, G. *Inequalities.* Cambridge, England: Cambridge University Press, 1952.

Hoehn, L. and Niven, I. "Averages on the Move." *Math. Mag.* **58**, 151–156, 1985.

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, 1962.

Mitrinović, D. S.; Pečarić, J. E.; and Fink, A. M. *Classical and New Inequalities in Analysis.* Dordrecht, Netherlands: Kluwer, 1993.

Vasic, P. M. and Mitrinović, D. S. *Analytic Inequalities.* New York: Springer-Verlag, 1970.

Arithmetic Progression

see ARITHMETIC SERIES

Arithmetic Sequence

A SEQUENCE of n numbers $\{d_0 + kd\}_{k=0}^{n-1}$ such that the differences between successive terms is a constant d.

see also ARITHMETIC SERIES, SEQUENCE

Arithmetic Series

An arithmetic series is the SUM of a SEQUENCE $\{a_k\}$, $k = 1, 2, \ldots$, in which each term is computed from the previous one by adding (or subtracting) a constant. Therefore, for $k > 1$,

$$a_k = a_{k-1} + d = a_{k-2} + 2d = \ldots = a_1 + d(k-1). \quad (1)$$

The sum of the sequence of the first n terms is then given by

$$\begin{aligned} S_n \equiv \sum_{k=1}^{n} a_k &= \sum_{k=1}^{n} [a_1 + (k-1)d] \\ &= na_1 + d\sum_{k=1}^{n}(k-1) = na_1 + d\sum_{k=2}^{n}(k-1) \\ &= na_1 + d\sum_{k=1}^{n-1} k \end{aligned} \quad (2)$$

Using the SUM identity

$$\sum_{k=1}^{n} = \tfrac{1}{2}n(n+1) \quad (3)$$

then gives

$$S_n = na_1 + \tfrac{1}{2}d(n-1) = \tfrac{1}{2}n[2a_1 + d(n-1)]. \quad (4)$$

Note, however, that

$$a_1 + a_n = a_1 + [a_1 + d(n-1)] = 2a_1 + d(n-1), \quad (5)$$

so

$$S_n = \tfrac{1}{2}n(a_1 + a_n), \quad (6)$$

or n times the AVERAGE of the first and last terms! This is the trick Gauss used as a schoolboy to solve the problem of summing the INTEGERS from 1 to 100 given as busy-work by his teacher. While his classmates toiled away doing the ADDITION longhand, Gauss wrote a single number, the correct answer

$$\tfrac{1}{2}(100)(1 + 100) = 50 \cdot 101 = 5050 \quad (7)$$

on his slate. When the answers were examined, Gauss's proved to be the only correct one.

see also ARITHMETIC SEQUENCE, GEOMETRIC SERIES, HARMONIC SERIES, PRIME ARITHMETIC PROGRESSION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 10, 1972.

Beyer, W. H. (Ed.). *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 8, 1987.

Courant, R. and Robbins, H. "The Arithmetical Progression." §1.2.2 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 12–13, 1996.

Pappas, T. *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 164, 1989.

Armstrong Number

The n-digit numbers equal to sum of nth powers of their digits (a finite sequence), also called PLUS PERFECT NUMBERS. They first few are given by 1, 2, 3, 4, 5, 6, 7, 8, 9, 153, 370, 371, 407, 1634, 8208, 9474, 54748, ... (Sloane's A005188).

see also NARCISSISTIC NUMBER

References

Sloane, N. J. A. Sequence A005188/M0488 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Arnold's Cat Map

The best known example of an ANOSOV DIFFEOMORPHISM. It is given by the TRANSFORMATION

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix}, \quad (1)$$

where x_{n+1} and y_{n+1} are computed mod 1. The Arnold cat mapping is non-Hamiltonian, nonanalytic, and mixing. However, it is AREA-PRESERVING since the DETERMINANT is 1. The LYAPUNOV CHARACTERISTIC EXPONENTS are given by

$$\begin{vmatrix} 1-\sigma & 1 \\ 1 & 2-\sigma \end{vmatrix} = \sigma^2 - 3\sigma + 1 = 0, \quad (2)$$

so

$$\sigma_\pm = \tfrac{1}{2}(3 \pm \sqrt{5}\,). \quad (3)$$

The EIGENVECTORS are found by plugging $\sigma_\pm$ into the MATRIX EQUATION

$$\begin{bmatrix} 1-\sigma_\pm & 1 \\ 1 & 2-\sigma_\pm \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (4)$$

For σ_+, the solution is

$$y = \tfrac{1}{2}(1 + \sqrt{5}\,)x \equiv \phi x, \quad (5)$$

where ϕ is the GOLDEN RATIO, so the unstable (normalized) EIGENVECTOR is

$$\boldsymbol{\xi}_+ = \tfrac{1}{10}\sqrt{50 - 10\sqrt{5}} \begin{bmatrix} 1 \\ \frac{1}{2}(1 + \sqrt{5}\,) \end{bmatrix}. \quad (6)$$

Similarly, for σ_-, the solution is

$$y = -\tfrac{1}{2}(\sqrt{5} - 1)x \equiv \phi^{-1}x, \quad (7)$$

so the stable (normalized) EIGENVECTOR is

$$\boldsymbol{\xi}_- = \tfrac{1}{10}\sqrt{50 + 10\sqrt{5}} \begin{bmatrix} 1 \\ \frac{1}{2}(1 - \sqrt{5}\,) \end{bmatrix}. \quad (8)$$

see also ANOSOV MAP

Arnold Diffusion

The nonconservation of ADIABATIC INVARIANTS which arises in systems with three or more DEGREES OF FREEDOM.

Arnold Tongue

Consider the CIRCLE MAP. If K is NONZERO, then the motion is periodic in some FINITE region surrounding each rational Ω. This execution of periodic motion in response to an irrational forcing is known as MODE LOCKING. If a plot is made of K versus Ω with the regions of periodic MODE-LOCKED parameter space plotted around rational Ω values (the WINDING NUMBERS), then the regions are seen to widen upward from 0 at $K = 0$ to some FINITE width at $K = 1$. The region surrounding each RATIONAL NUMBER is known as an ARNOLD TONGUE.

At $K = 0$, the Arnold tongues are an isolated set of MEASURE zero. At $K = 1$, they form a general CANTOR SET of dimension $d \approx 0.8700$. In general, an Arnold tongue is defined as a resonance zone emanating out from RATIONAL NUMBERS in a two-dimensional parameter space of variables.

see also CIRCLE MAP

Aronhold Process

The process used to generate an expression for a covariant in the first degree of any one of the equivalent sets of COEFFICIENTS for a curve.

see also CLEBSCH-ARONHOLD NOTATION, JOACHIMSTHAL'S EQUATION

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 74, 1959.

Aronson's Sequence

The sequence whose definition is: "t is the first, fourth, eleventh, ... letter of this sentence." The first few values are 1, 4, 11, 16, 24, 29, 33, 35, 39, ... (Sloane's A005224).

References

Hofstadter, D. R. *Metamagical Themas: Questing of Mind and Pattern.* New York: BasicBooks, p. 44, 1985.
Sloane, N. J. A. Sequence A005224/M3406 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Arrangement

In general, an arrangement of objects is simply a grouping of them. The number of "arrangements" of n items is given either by a COMBINATION (order is ignored) or PERMUTATION (order is significant).

The division of SPACE into cells by a collection of HYPERPLANES is also called an arrangement.

see also COMBINATION, CUTTING, HYPERPLANE, ORDERING, PERMUTATION

Arrangement Number

see PERMUTATION

Array

An array is a "list of lists" with the length of each level of list the same. The size (sometimes called the "shape") of a d-dimensional array is then indicated as $\underbrace{m \times n \times \cdots \times p}_{d}$. The most common type of array encountered is the 2-D $m \times n$ rectangular array having m columns and n rows. If $m = n$, a square array results. Sometimes, the order of the elements in an array is significant (as in a MATRIX), whereas at other times, arrays which are equivalent modulo reflections (and rotations, in the case of a square array) are considered identical (as in a MAGIC SQUARE or PRIME ARRAY).

In order to exhaustively list the number of distinct arrays of a given shape with each element being one of k possible choices, the naive algorithm of running through each case and checking to see whether it's equivalent to an earlier one is already just about as efficient as can be. The running time must be at least the number of answers, and this is so close to $k^{mn\cdots p}$ that the difference isn't significant.

However, finding the *number* of possible arrays of a given shape is much easier, and an exact formula can be obtained using the POLYA ENUMERATION THEOREM. For the simple case of an $m \times n$ array, even this proves unnecessary since there are only a few possible symmetry types, allowing the possibilities to be counted explicitly. For example, consider the case of m and n EVEN and distinct, so only reflections need be included. To take a specific case, let $m = 6$ and $n = 4$ so the array looks like

$$\begin{array}{ccc|ccc} a & b & c & d & e & f \\ g & h & i & j & k & l \\ \hline m & n & o & p & q & r \\ s & t & u & v & w & x, \end{array}$$

where each $a, b, \ldots, x$ can take a value from 1 to k. The total number of possible arrangements is k^{24} (k^{mn} in general). The number of arrangements which are equivalent to their left-right mirror images is k^{12} (in general, $k^{mn/2}$), as is the number equal to their up-down mirror images, or their rotations through 180°. There are also k^6 arrangements (in general, $k^{mn/4}$) with full symmetry.

In general, it is therefore true that

$$\begin{cases} k^{mn/4} & \text{with full symmetry} \\ k^{mn/2} - k^{mn/4} & \text{with } \textit{only} \text{ left-right reflection} \\ k^{mn/2} - k^{mn/4} & \text{with } \textit{only} \text{ up-down reflection} \\ k^{mn/2} - k^{mn/4} & \text{with } \textit{only} \text{ } 180^\circ \text{ rotation,} \end{cases}$$

so there are

$$k^{mn} - 3k^{mn/2} + 2k^{mn/4}$$

arrangements with no symmetry. Now dividing by the number of images of each type, the result, for $m \neq n$ with m, n EVEN, is

$$\begin{aligned} N(m,n,k) &= \tfrac{1}{4}k^{mn} + (\tfrac{1}{2})(3)(k^{mn/2} - k^{mn/4}) \\ &\quad + \tfrac{1}{4}(k^{mn} - 3k^{mn/2} + 2k^{mn/4}) \\ &= \tfrac{1}{4}k^{mn} + \tfrac{3}{4}k^{mn/2} + \tfrac{1}{2}k^{mn/4}. \end{aligned}$$

The number is therefore of order $\mathcal{O}(k^{mn}/4)$, with "correction" terms of much smaller order.

see also ANTIMAGIC SQUARE, EULER SQUARE, KIRKMAN'S SCHOOLGIRL PROBLEM, LATIN RECTANGLE, LATIN SQUARE, MAGIC SQUARE, MATRIX, MRS. PERKINS' QUILT, MULTIPLICATION TABLE, ORTHOGONAL ARRAY, PERFECT SQUARE, PRIME ARRAY, QUOTIENT-DIFFERENCE TABLE, ROOM SQUARE, STOLARSKY ARRAY, TRUTH TABLE, WYTHOFF ARRAY

Arrow Notation

A NOTATION invented by Knuth (1976) to represent LARGE NUMBERS in which evaluation proceeds from the right (Conway and Guy 1996, p. 60).

$$\begin{array}{ll} m \uparrow n & \underbrace{m \cdot m \cdots m}_{n} \\ m \uparrow\uparrow n & \underbrace{m \uparrow m \uparrow \cdots \uparrow m}_{n} \\ m \uparrow\uparrow\uparrow n & \underbrace{m \uparrow\uparrow m \uparrow\uparrow \cdots \uparrow\uparrow m}_{n} \end{array}$$

For example,

$$m \uparrow n = m^n \tag{1}$$

$$m \uparrow\uparrow n = \underbrace{m \uparrow \cdots \uparrow m}_{n} = \underbrace{m^{m^{\cdot^{\cdot^{\cdot^{m}}}}}}_{n}$$

$$m \uparrow\uparrow 2 = \underbrace{m \uparrow m}_{2} = m \uparrow m = m^m \tag{2}$$

$$\begin{aligned} m \uparrow\uparrow 3 &= \underbrace{m \uparrow m \uparrow m}_{3} = m \uparrow (m \uparrow m) \\ &= m \uparrow m^m = m^{m^m} \end{aligned} \tag{3}$$

$$m \uparrow\uparrow\uparrow 2 = \underbrace{m \uparrow\uparrow m}_{2} = m \uparrow\uparrow m = \underbrace{m^{m^{\cdot^{\cdot^{\cdot^{m}}}}}}_{m} \tag{4}$$

$$\begin{aligned} m \uparrow\uparrow\uparrow 3 &= \underbrace{m \uparrow\uparrow m \uparrow\uparrow m}_{3} = m \uparrow\uparrow \underbrace{m^{m^{\cdot^{\cdot^{\cdot^{m}}}}}}_{m} \\ &= \underbrace{m \uparrow \cdots \uparrow m}_{\underbrace{m^{m^{\cdot^{\cdot^{\cdot^{m}}}}}}_{m}} = \underbrace{m^{m^{\cdot^{\cdot^{\cdot^{m}}}}}}_{\underbrace{m^{m^{\cdot^{\cdot^{\cdot^{m}}}}}}_{m}}. \end{aligned} \tag{5}$$

$m \uparrow\uparrow n$ is sometimes called a POWER TOWER. The values $\underbrace{n \uparrow \cdots \uparrow}_{n} n$ are called ACKERMANN NUMBERS.

see also ACKERMANN NUMBER, CHAINED ARROW NOTATION, DOWN ARROW NOTATION, LARGE NUMBER, POWER TOWER, STEINHAUS-MOSER NOTATION

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 59–62, 1996.

Guy, R. K. and Selfridge, J. L. "The Nesting and Roosting Habits of the Laddered Parenthesis." *Amer. Math. Monthly* **80**, 868–876, 1973.

Knuth, D. E. "Mathematics and Computer Science: Coping with Finiteness. Advances in Our Ability to Compute are Bringing Us Substantially Closer to Ultimate Limitations." *Science* **194**, 1235–1242, 1976.

Vardi, I. *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 11 and 226–229, 1991.

Arrow's Paradox

Perfect democratic voting is, not just in practice but *in principle,* impossible.

References

Gardner, M. *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, p. 56, 1988.

Arrowhead Curve

see SIERPIŃSKI ARROWHEAD CURVE

Art Gallery Theorem

Also called CHVÁTAL'S ART GALLERY THEOREM. If the walls of an art gallery are made up of n straight LINES SEGMENTS, then the entire gallery can always be supervised by $\lfloor n/3 \rfloor$ watchmen placed in corners, where $\lfloor x \rfloor$ is the FLOOR FUNCTION. This theorem was proved by V. Chvátal in 1973. It is conjectured that an art gallery with n walls and h HOLES requires $\lfloor (n+h)/3 \rfloor$ watchmen.

see also ILLUMINATION PROBLEM

References

Honsberger, R. "Chvátal's Art Gallery Theorem." Ch. 11 in *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., pp. 104–110, 1976.

O'Rourke, J. *Art Gallery Theorems and Algorithms.* New York: Oxford University Press, 1987.

Stewart, I. "How Many Guards in the Gallery?" *Sci. Amer.* **270**, 118–120, May 1994.

Tucker, A. "The Art Gallery Problem." *Math Horizons,* pp. 24-26, Spring 1994.

Wagon, S. "The Art Gallery Theorem." §10.3 in *Mathematica in Action.* New York: W. H. Freeman, pp. 333–345, 1991.

Articulation Vertex

A VERTEX whose removal will disconnect a GRAPH, also called a CUT-VERTEX.

see also BRIDGE (GRAPH)

References

Chartrand, G. "Cut-Vertices and Bridges." §2.4 in *Introductory Graph Theory.* New York: Dover, pp. 45–49, 1985.

Artin Braid Group

see BRAID GROUP

Artin's Conjecture

There are at least two statements which go by the name of Artin's conjecture. The first is the RIEMANN HYPOTHESIS. The second states that every INTEGER not equal to -1 or a SQUARE NUMBER is a primitive root modulo p for infinitely many p and proposes a density for the set of such p which are always rational multiples of a constant known as ARTIN'S CONSTANT. There is an analogous theorem for functions instead of numbers which has been proved by Billharz (Shanks 1993, p. 147).

see also ARTIN'S CONSTANT, RIEMANN HYPOTHESIS

References

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 31, 80–83, and 147, 1993.

Artin's Constant

If $n \neq -1$ and n is not a PERFECT SQUARE, then Artin conjectured that the SET $S(n)$ of all PRIMES for which n is a PRIMITIVE ROOT is infinite. Under the assumption of the EXTENDED RIEMANN HYPOTHESIS, Artin's conjecture was solved in 1967 by C. Hooley. If, in addition, n is not an rth POWER for any $r > 1$, then Artin conjectured that the density of $S(n)$ relative to the PRIMES is C_{Artin} (independent of the choice of n), where

$$C_{\text{Artin}} = \prod_{q \text{ prime}} \left[1 - \frac{1}{q(q-1)}\right] = 0.3739558136\ldots,$$

and the PRODUCT is over PRIMES. The significance of this constant is more easily seen by describing it as the fraction of PRIMES p for which $1/p$ has a maximal DECIMAL EXPANSION (Conway and Guy 1996).

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 169, 1996.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/artin/artin.html`.

Hooley, C. "On Artin's Conjecture." *J. reine angew. Math.* **225**, 209–220, 1967.

Ireland, K. and Rosen, M. *A Classical Introduction to Modern Number Theory, 2nd ed.* New York: Springer-Verlag, 1990.

Ribenboim, P. *The Book of Prime Number Records.* New York: Springer-Verlag, 1989.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 80–83, 1993.

Wrench, J. W. "Evaluation of Artin's Constant and the Twin Prime Constant." *Math. Comput.* **15**, 396–398, 1961.

Artin L-Function

An Artin L-function over the RATIONALS $\mathbb{Q}$ encodes in a GENERATING FUNCTION information about how an irreducible monic POLYNOMIAL over $\mathbb{Z}$ factors when reduced modulo each PRIME. For the POLYNOMIAL x^2+1, the Artin L-function is

$$L(s, \mathbb{Q}(i)/\mathbb{Q}, \text{sgn}) = \prod_{p \text{ odd prime}} \frac{1}{1 - \left(\frac{-1}{p}\right) p^{-s}},$$

where $(-1/p)$ is a LEGENDRE SYMBOL, which is equivalent to the EULER L-FUNCTION. The definition over arbitrary POLYNOMIALS generalizes the above expression.

see also LANGLANDS RECIPROCITY

References

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Artin Reciprocity

see ARTIN'S RECIPROCITY THEOREM

Artin's Reciprocity Theorem

A general RECIPROCITY THEOREM for all orders. If R is a NUMBER FIELD and R' a finite integral extension, then there is a SURJECTION from the group of fractional IDEALS prime to the discriminant, given by the Artin symbol. For some cycle c, the kernel of this SURJECTION contains each PRINCIPAL fractional IDEAL generated by an element congruent to 1 mod c.

see also LANGLANDS PROGRAM

Artinian Group

A GROUP in which any decreasing CHAIN of distinct SUBGROUPS terminates after a FINITE number.

Artinian Ring

A noncommutative SEMISIMPLE RING satisfying the "descending chain condition."

see also GORENSTEIN RING, SEMISIMPLE RING

References

Artin, E. "Zur Theorie der hyperkomplexer Zahlen." *Hamb. Abh.* **5**, 251-260, 1928.

Artin, E. "Zur Arithmetik hyperkomplexer Zahlen." *Hamb. Abh.* **5**, 261–289, 1928.

Artistic Series

A SERIES is called artistic if every three consecutive terms have a common three-way ratio

$$P[a_i, a_{i+1}, a_{i+2}] = \frac{(a_i + a_{i+1} + a_{i+2})a_{i+1}}{a_i a_{i+2}}.$$

A SERIES is also artistic IFF its BIAS is a constant. A GEOMETRIC SERIES with RATIO $r > 0$ is an artistic series with

$$P = \frac{1}{r} + 1 + r \geq 3.$$

see also BIAS (SERIES), GEOMETRIC SERIES, MELODIC SERIES

References

Duffin, R. J. "On Seeing Progressions of Constant Cross Ratio." *Amer. Math. Monthly* **100**, 38–47, 1993.

ASA Theorem

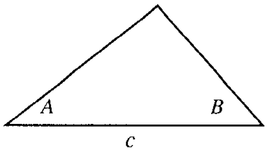

Specifying two adjacent ANGLES A and B and the side between them c uniquely determines a TRIANGLE with AREA

$$K = \frac{c^2}{2(\cot A + \cot B)}. \tag{1}$$

The angle C is given in terms of A and B by

$$C = \pi - A - B, \tag{2}$$

and the sides a and b can be determined by using the LAW OF SINES

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C} \tag{3}$$

to obtain

$$a = \frac{\sin A}{\sin(\pi - A - B)} c \tag{4}$$

$$b = \frac{\sin B}{\sin(\pi - A - B)} c. \tag{5}$$

see also AAA THEOREM, AAS THEOREM, ASS THEOREM, SAS THEOREM, SSS THEOREM, TRIANGLE

Aschbacher's Component Theorem

Suppose that $E(G)$ (the commuting product of all components of G) is SIMPLE and G contains a SEMISIMPLE INVOLUTION. Then there is some SEMISIMPLE INVOLUTION x such that $C_G(x)$ has a NORMAL SUBGROUP K which is either QUASISIMPLE or ISOMORPHIC to $O^+(4,q)'$ and such that $Q = C_G(K)$ is TIGHTLY EMBEDDED.

see also INVOLUTION (GROUP), ISOMORPHIC GROUPS, NORMAL SUBGROUP, QUASISIMPLE GROUP, SIMPLE GROUP, TIGHTLY EMBEDDED

ASS Theorem

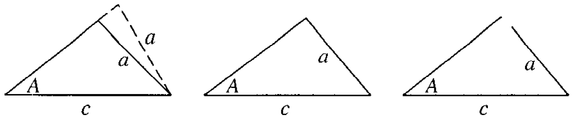

Specifying two adjacent side lengths a and b of a TRIANGLE (taking $a > b$) and one ACUTE ANGLE A opposite a does not, in general, uniquely determine a triangle. If $\sin A < a/c$, there are two possible TRIANGLES satisfying the given conditions. If $\sin A = a/c$, there is one possible TRIANGLE. If $\sin A > a/c$, there are no possible TRIANGLES. Remember: don't try to prove congruence with the ASS theorem or you will make make an ASS out of yourself.

see also AAA THEOREM, AAS THEOREM, SAS THEOREM, SSS THEOREM, TRIANGLE

Associative

In simple terms, let x, y, and z be members of an ALGEBRA. Then the ALGEBRA is said to be associative if

$$x \cdot (y \cdot z) = (x \cdot y) \cdot z, \tag{1}$$

where $\cdot$ denotes MULTIPLICATION. More formally, let A denote an $\mathbb{R}$-algebra, so that A is a VECTOR SPACE over $\mathbb{R}$ and

$$A \times A \to A \tag{2}$$

$$(x, y) \mapsto x \cdot y. \tag{3}$$

Then A is said to be m-associative if there exists an m-D SUBSPACE S of A such that

$$(y \cdot x) \cdot z = y \cdot (x \cdot z) \tag{4}$$

for all $y, z \in A$ and $x \in S$. Here, VECTOR MULTIPLICATION $x \cdot y$ is assumed to be BILINEAR. An n-D n-associative ALGEBRA is simply said to be "associative."

see also COMMUTATIVE, DISTRIBUTIVE

References

Finch, S. "Zero Structures in Real Algebras." `http://www.mathsoft.com/asolve/zerodiv/zerodiv.html`.

Associative Magic Square

1	15	24	8	17
23	7	16	5	14
20	4	13	22	6
12	21	10	19	3
9	18	2	11	25

An $n \times n$ MAGIC SQUARE for which every pair of numbers symmetrically opposite the center sum to $n^2 + 1$. The LO SHU is associative but not PANMAGIC. Order four squares can be PANMAGIC or associative, but not both. Order five squares are the smallest which can be both associative and PANMAGIC, and 16 distinct associative PANMAGIC SQUARES exist, one of which is illustrated above (Gardner 1988).

see also MAGIC SQUARE, PANMAGIC SQUARE

References

Gardner, M. "Magic Squares and Cubes." Ch. 17 in *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, 1988.

Astroid

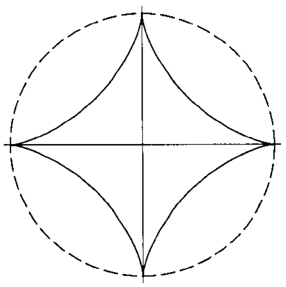

A 4-cusped HYPOCYCLOID which is sometimes also called a TETRACUSPID, CUBOCYCLOID, or PARACYCLE. The parametric equations of the astroid can be obtained by plugging in $n \equiv a/b = 4$ or $4/3$ into the equations for a general HYPOCYCLOID, giving

$$x = 3b\cos\phi + b\cos(3\phi) = 4b\cos^3\phi = a\cos^3\phi \quad (1)$$
$$y = 3b\sin\phi - b\sin(3\phi) = 4b\sin^3\phi = a\sin^3\phi. \quad (2)$$

In CARTESIAN COORDINATES,

$$x^{2/3} + y^{2/3} = a^{2/3}. \quad (3)$$

In PEDAL COORDINATES with the PEDAL POINT at the center, the equation is

$$r^2 + 3p^2 = a^2. \quad (4)$$

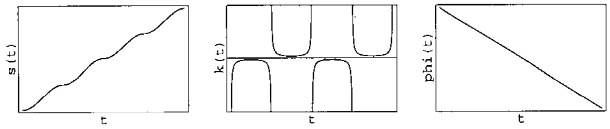

The ARC LENGTH, CURVATURE, and TANGENTIAL ANGLE are

$$s(t) = \tfrac{3}{2}\int_0^t |\sin(2t')|\,dt' = \tfrac{3}{2}\sin^2 t \quad (5)$$
$$\kappa(t) = -\tfrac{2}{3}\csc(2t) \quad (6)$$
$$\phi(t) = -t. \quad (7)$$

As usual, care must be taken in the evaluation of $s(t)$ for $t > \pi/2$. Since (5) comes from an integral involving the ABSOLUTE VALUE of a function, it must be monotonic increasing. Each QUADRANT can be treated correctly by defining

$$n = \left\lfloor \frac{2t}{\pi} \right\rfloor + 1, \quad (8)$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION, giving the formula

$$s(t) = (-1)^{1+[n \pmod 2]}\tfrac{3}{2}\sin^2 t + 3\left\lfloor \tfrac{1}{2}n \right\rfloor. \quad (9)$$

The overall ARC LENGTH of the astroid can be computed from the general HYPOCYCLOID formula

$$s_n = \frac{8a(n-1)}{n} \quad (10)$$

with $n = 4$,

$$s_4 = 6a. \quad (11)$$

The AREA is given by

$$A_n = \frac{(n-1)(n-2)}{n^2}\pi a^2 \quad (12)$$

with $n = 4$,

$$A_4 = \tfrac{3}{8}\pi a^2. \quad (13)$$

The EVOLUTE of an ELLIPSE is a stretched HYPOCYCLOID. The gradient of the TANGENT T from the point with parameter p is $-\tan p$. The equation of this TANGENT T is

$$x\sin p + y\cos p = \tfrac{1}{2}a\sin(2p) \quad (14)$$

(MacTutor Archive). Let T cut the x-AXIS and the y-AXIS at X and Y, respectively. Then the length XY is a constant and is equal to a.

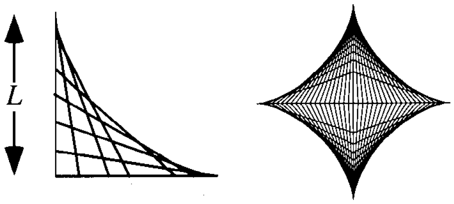

The astroid can also be formed as the ENVELOPE produced when a LINE SEGMENT is moved with each end on one of a pair of PERPENDICULAR axes (e.g., it is the curve enveloped by a ladder sliding against a wall or a garage door with the top corner moving along a vertical track; left figure above). The astroid is therefore a GLISSETTE. To see this, note that for a ladder of length L, the points of contact with the wall and floor are $(x_0, 0)$ and $(0, \sqrt{L^2 - x_0{}^2})$, respectively. The equation of the LINE made by the ladder with its foot at $(x_0, 0)$ is therefore

$$y - 0 = \frac{\sqrt{L^2 - x_0{}^2}}{-x_0}(x - x_0) \quad (15)$$

which can be written

$$U(x, y, x_0) = y + \frac{\sqrt{L^2 - x_0{}^2}}{x_0}(x - x_0). \quad (16)$$

The equation of the ENVELOPE is given by the simultaneous solution of

$$\begin{cases} U(x, y, x_0) = y + \frac{\sqrt{L^2 - x_0{}^2}}{x_0}(x - x_0) = 0 \\ \frac{\partial U}{\partial x_0} = \frac{x_0{}^2 - Lx^2}{x_0{}^2\sqrt{L^2 - x_0{}^2}} = 0, \end{cases} \quad (17)$$

which is

$$x = \frac{x_0{}^3}{L^2} \quad (18)$$
$$y = \frac{(L^2 - x_0{}^2)^{3/2}}{L^2}. \quad (19)$$

Noting that

$$x^{2/3} = \frac{{x_0}^2}{L^{4/3}} \tag{20}$$

$$y^{2/3} = \frac{L^2 - {x_0}^2}{L^{4/3}} \tag{21}$$

allows this to be written implicitly as

$$x^{2/3} + y^{2/3} = L^{2/3}, \tag{22}$$

the equation of the astroid, as promised.

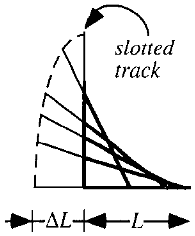

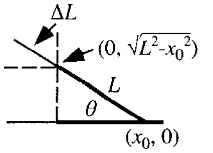

The related problem obtained by having the "garage door" of length L with an "extension" of length ΔL move up and down a slotted track also gives a surprising answer. In this case, the position of the "extended" end for the foot of the door at horizontal position x_0 and ANGLE θ is given by

$$x = -\Delta L \cos\theta \tag{23}$$

$$y = \sqrt{L^2 - {x_0}^2} + \Delta L \sin\theta. \tag{24}$$

Using

$$x_0 = L\cos\theta \tag{25}$$

then gives

$$x = -\frac{\Delta L}{L} x_0 \tag{26}$$

$$y = \sqrt{L^2 - {x_0}^2}\left(1 + \frac{\Delta L}{L}\right). \tag{27}$$

Solving (26) for x_0, plugging into (27) and squaring then gives

$$y^2 = L^2 - \frac{L^2 x^2}{(\Delta L)^2}\left(1 + \frac{\Delta L}{L}\right)^2. \tag{28}$$

Rearranging produces the equation

$$\frac{x^2}{(\Delta L)^2} + \frac{y^2}{(L + \Delta L)^2} = 1, \tag{29}$$

the equation of a (QUADRANT of an) ELLIPSE with SEMIMAJOR and SEMIMINOR AXES of lengths ΔL and $L + \Delta L$.

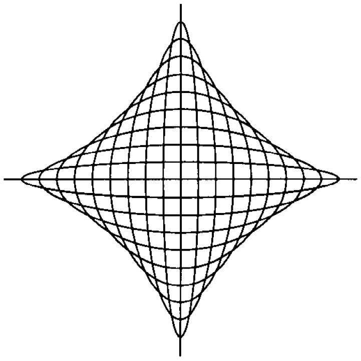

The astroid is also the ENVELOPE of the family of ELLIPSES

$$\frac{x^2}{c^2} + \frac{y^2}{(1-c)^2} - 1 = 0, \tag{30}$$

illustrated above.

see also DELTOID, ELLIPSE ENVELOPE, LAMÉ CURVE, NEPHROID, RANUNCULOID

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 172–175, 1972.

Lee, X. "Astroid." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Astroid_dir/astroid.html`.

Lockwood, E. H. "The Astroid." Ch. 6 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 52–61, 1967.

MacTutor History of Mathematics Archive. "Astroid." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Astroid.html`.

Yates, R. C. "Astroid." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 1–3, 1952.

Astroid Evolute

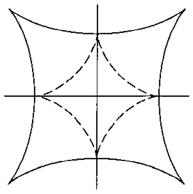

A HYPOCYCLOID EVOLUTE for $n = 4$ is another ASTROID scaled by a factor $n/(n-2) = 4/2 = 2$ and rotated $1/(2 \cdot 4) = 1/8$ of a turn.

Astroid Involute

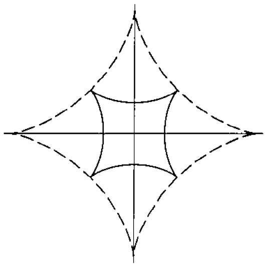

A HYPOCYCLOID INVOLUTE for $n = 4$ is another ASTROID scaled by a factor $(n-2)/2 = 2/4 = 1/2$ and rotated $1/(2 \cdot 4) = 1/8$ of a turn.

Astroid Pedal Curve

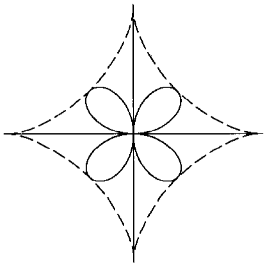

The PEDAL CURVE of an ASTROID with PEDAL POINT at the center is a QUADRIFOLIUM.

Astroid Radial Curve

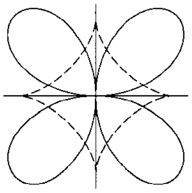

The QUADRIFOLIUM

$$x = x_0 + 3a\cos t - 3a\cos(3t)$$
$$y = y_0 + 3a\sin t + 3a\sin(3t).$$

Astroidal Ellipsoid

The surface which is the inverse of the ELLIPSOID in the sense that it "goes in" where the ELLIPSOID "goes out." It is given by the parametric equations

$$x = (a\cos u\cos v)^3$$
$$y = (b\sin u\cos v)^3$$
$$z = (c\sin v)^3$$

for $u \in [-\pi/2, \pi/2]$ and $v \in [-\pi, \pi]$. The special case $a = b = c = 1$ corresponds to the HYPERBOLIC OCTAHEDRON.

see also ELLIPSOID, HYPERBOLIC OCTAHEDRON

References
Nordstrand, T. "Astroidal Ellipsoid." `http://www.uib.no/people/nfytn/asttxt.htm`.

Asymptosy

ASYMPTOTIC behavior. A useful yet endangered word, found rarely outside the captivity of the *Oxford English Dictionary*.

see also ASYMPTOTE, ASYMPTOTIC

Asymptote

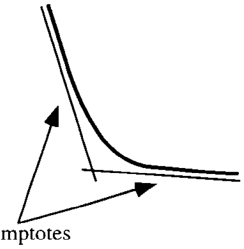

A curve approaching a given curve arbitrarily closely, as illustrated in the above diagram.

see also ASYMPTOSY, ASYMPTOTIC, ASYMPTOTIC CURVE

References
Giblin, P. J. "What is an Asymptote?" *Math. Gaz.* **56**, 274–284, 1972.

Asymptotic

Approaching a value or curve arbitrarily closely (i.e., as some sort of LIMIT is taken). A CURVE A which is asymptotic to given CURVE C is called the ASYMPTOTE of C.

see also ASYMPTOSY, ASYMPTOTE, ASYMPTOTIC CURVE, ASYMPTOTIC DIRECTION, ASYMPTOTIC SERIES, LIMIT

Asymptotic Curve

Given a REGULAR SURFACE M, an asymptotic curve is formally defined as a curve $\mathbf{x}(t)$ on M such that the NORMAL CURVATURE is 0 in the direction $\mathbf{x}'(t)$ for all t in the domain of $\mathbf{x}$. The differential equation for the parametric representation of an asymptotic curve is

$$eu'^2 + 2fu'v' + gv'^2 = 0, \tag{1}$$

where e, f, and g are second FUNDAMENTAL FORMS. The differential equation for asymptotic curves on a MONGE PATCH $(u, v, h(u, v))$ is

$$h_{uu}u'^2 + 2h_{uu}u'v' + h_{vv}v'^2 = 0, \tag{2}$$

and on a polar patch $(r\cos\theta, r\sin\theta, h(r))$ is

$$h''(r)r'^2 + h'(r)r\theta'^2 = 0. \tag{3}$$

The images below show asymptotic curves for the ELLIPTIC HELICOID, FUNNEL, HYPERBOLIC PARABOLOID, and MONKEY SADDLE.

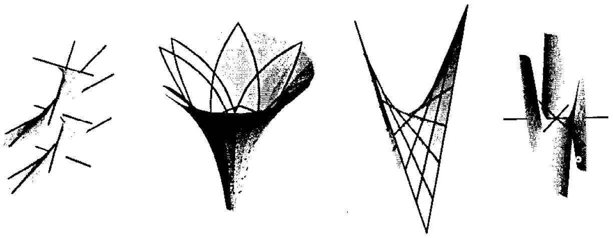

see also RULED SURFACE

References

Gray, A. "Asymptotic Curves," "Examples of Asymptotic Curves," "Using Mathematica to Find Asymptotic Curves." §16.1, 16.2, and 16.3 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 320–331, 1993.

Asymptotic Direction

An asymptotic direction at a point **p** of a REGULAR SURFACE $M \in \mathbb{R}^3$ is a direction in which the NORMAL CURVATURE of M vanishes.

1. There are no asymptotic directions at an ELLIPTIC POINT.
2. There are exactly two asymptotic directions at a HYPERBOLIC POINT.
3. There is exactly one asymptotic direction at a PARABOLIC POINT.
4. Every direction is asymptotic at a PLANAR POINT.

see also ASYMPTOTIC CURVE

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.*Boca Raton, FL: CRC Press, pp. 270 and 320, 1993.

Asymptotic Notation

Let n be a integer variable which tends to infinity and let x be a continuous variable tending to some limit. Also, let $\phi(n)$ or $\phi(x)$ be a positive function and $f(n)$ or $f(x)$ any function. Then Hardy and Wright (1979) define

1. $f = \mathcal{O}(\phi)$ to mean that $|f| < A\phi$ for some constant A and all values of n and x,
2. $f = o(\phi)$ to mean that $f/\phi \to 0$,
3. $f \sim \phi$ to mean that $f/\phi \to 1$,
4. $f \prec \phi$ to mean the same as $f = o(\phi)$,
5. $f \succ \phi$ to mean $f/\phi \to \infty$, and
6. $f \asymp \phi$ to mean $A_1\phi < f < A_2$ for some positive constants A_1 and A_2.

$f = o(\phi)$ implies and is stronger than $f = \mathcal{O}(\phi)$.

References

Hardy, G. H. and Wright, E. M. "Some Notation." §1.6 in *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, pp. 7–8, 1979.

Asymptotic Series

An asymptotic series is a SERIES EXPANSION of a FUNCTION in a variable x which may converge or diverge (Erdelyi 1987, p. 1), but whose partial sums can be made an arbitrarily good approximation to a given function for large enough x. To form an asymptotic series $R(x)$ of $f(x)$, written

$$f(x) \sim R(x), \qquad (1)$$

take

$$x^n R_n(x) = x^n[f(x) - S_n(x)], \qquad (2)$$

where

$$S_n(x) \equiv a_0 + \frac{a_1}{x} + \frac{a_2}{x^2} + \ldots + \frac{a_n}{x^n}. \qquad (3)$$

The asymptotic series is defined to have the properties

$$\lim_{x\to\infty} x^n R_n(x) = 0 \qquad \text{for fixed } n \qquad (4)$$

$$\lim_{n\to\infty} x^n R_n(x) = \infty \qquad \text{for fixed } x. \qquad (5)$$

Therefore,

$$f(x) \approx \sum_{n=0}^{\infty} a_n x^{-n} \qquad (6)$$

in the limit $x \to \infty$. If a function has an asymptotic expansion, the expansion is unique. The symbol $\sim$ is also used to mean directly SIMILAR.

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 15, 1972.
Arfken, G. "Asymptotic of Semiconvergent Series." §5.10 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 339–346, 1985.
Bleistein, N. and Handelsman, R. A. *Asymptotic Expansions of Integrals.* New York: Dover, 1986.
Copson, E. T. *Asymptotic Expansions.* Cambridge, England: Cambridge University Press, 1965.
de Bruijn, N. G. *Asymptotic Methods in Analysis, 2nd ed.* New York: Dover, 1982.
Dingle, R. B. *Asymptotic Expansions: Their Derivation and Interpretation.* London: Academic Press, 1973.
Erdelyi, A. *Asymptotic Expansions.* New York: Dover, 1987.
Morse, P. M. and Feshbach, H. "Asymptotic Series; Method of Steepest Descent." §4.6 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 434–443, 1953.
Olver, F. W. J. *Asymptotics and Special Functions.* New York: Academic Press, 1974.
Wasow, W. R. *Asymptotic Expansions for Ordinary Differential Equations.* New York: Dover, 1987.

Atiyah-Singer Index Theorem

A theorem which states that the analytic and topological "indices" are equal for any elliptic differential operator on an n-D COMPACT DIFFERENTIABLE $\mathbb{C}^\infty$ boundaryless MANIFOLD.

see also COMPACT MANIFOLD, DIFFERENTIABLE MANIFOLD

References

Atiyah, M. F. and Singer, I. M. "The Index of Elliptic Operators on Compact Manifolds." *Bull. Amer. Math. Soc.* **69**, 322–433, 1963.
Atiyah, M. F. and Singer, I. M. "The Index of Elliptic Operators I, II, III." *Ann. Math.* **87**, 484–604, 1968.
Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, p. 4, 1996.

Atkin-Goldwasser-Kilian-Morain Certificate

A recursive PRIMALITY CERTIFICATE for a PRIME p. The certificate consists of a list of

1. A point on an ELLIPTIC CURVE C

$$y^2 = x^3 + g_2 x + g_3 \pmod{p}$$

for some numbers g_2 and g_3.

2. A PRIME q with $q > (p^{1/4}+1)^2$, such that for some other number k and $m = kq$ with $k \neq 1$, $mC(x, y, g_2, g_3, p)$ is the identity on the curve, but $kC(x, y, g_2, g_3, p)$ is not the identity. This guarantees PRIMALITY of p by a theorem of Goldwasser and Kilian (1986).
3. Each q has its recursive certificate following it. So if the smallest q is known to be PRIME, all the numbers are certified PRIME up the chain.

A PRATT CERTIFICATE is quicker to generate for small numbers. The *Mathematica*® (Wolfram Research, Champaign, IL) task `ProvablePrime[n]` therefore generates an Atkin-Goldwasser-Kilian-Morain certificate only for numbers above a certain limit (10^{10} by default), and a PRATT CERTIFICATE for smaller numbers.

see also ELLIPTIC CURVE PRIMALITY PROVING, ELLIPTIC PSEUDOPRIME, PRATT CERTIFICATE, PRIMALITY CERTIFICATE, WITNESS

References
Atkin, A. O. L. and Morain, F. "Elliptic Curves and Primality Proving." *Math. Comput.* **61**, 29–68, 1993.
Bressoud, D. M. *Factorization and Prime Testing.* New York: Springer-Verlag, 1989.
Goldwasser, S. and Kilian, J. "Almost All Primes Can Be Quickly Certified." *Proc. 18th STOC.* pp. 316-329, 1986.
Morain, F. "Implementation of the Atkin-Goldwasser-Kilian Primality Testing Algorithm." Rapport de Recherche 911, INRIA, Octobre 1988.
Schoof, R. "Elliptic Curves over Finite Fields and the Computation of Square Roots mod p." *Math. Comput.* **44**, 483–494, 1985.
Wunderlich, M. C. "A Performance Analysis of a Simple Prime-Testing Algorithm." *Math. Comput.* **40**, 709–714, 1983.

Atomic Statement

In LOGIC, a statement which cannot be broken down into smaller statements.

Attraction Basin

see BASIN OF ATTRACTION

Attractor

An attractor is a SET of states (points in the PHASE SPACE), invariant under the dynamics, towards which neighboring states in a given BASIN OF ATTRACTION asymptotically approach in the course of dynamic evolution. An attractor is defined as the smallest unit which cannot be itself decomposed into two or more attractors with distinct BASINS OF ATTRACTION. This restriction is necessary since a DYNAMICAL SYSTEM may have multiple attractors, each with its own BASIN OF ATTRACTION.

Conservative systems do not have attractors, since the motion is periodic. For dissipative DYNAMICAL SYSTEMS, however, volumes shrink exponentially so attractors have 0 volume in n-D phase space.

A stable FIXED POINT surrounded by a dissipative region is an attractor known as a SINK. Regular attractors (corresponding to 0 LYAPUNOV CHARACTERISTIC EXPONENTS) act as LIMIT CYCLES, in which trajectories circle around a limiting trajectory which they asymptotically approach, but never reach. STRANGE ATTRACTORS are bounded regions of PHASE SPACE (corresponding to POSITIVE LYAPUNOV CHARACTERISTIC EXPONENTS) having zero MEASURE in the embedding PHASE SPACE and a FRACTAL DIMENSION. Trajectories within a STRANGE ATTRACTOR appear to skip around randomly.

see also BARNSLEY'S FERN, BASIN OF ATTRACTION, CHAOS GAME, FRACTAL DIMENSION, LIMIT CYCLE, LYAPUNOV CHARACTERISTIC EXPONENT, MEASURE, SINK (MAP), STRANGE ATTRACTOR

Auction

A type of sale in which members of a group of buyers offer ever increasing amounts. The bidder making the last bid (for which no higher bid is subsequently made within a specified time limit: "going once, going twice, sold") must then purchase the item in question at this price. Variants of simple bidding are also possible, as in a VICKERY AUCTION.

see also VICKERY AUCTION

Augend

The first of several ADDENDS, or "the one to which the others are added," is sometimes called the augend. Therefore, while a, b, and c are ADDENDS in $a+b+c$, a is the augend.

see also ADDEND, ADDITION

Augmented Amicable Pair

A PAIR of numbers m and n such that

$$\sigma(m) = \sigma(n) = m + n - 1,$$

where $\sigma(m)$ is the DIVISOR FUNCTION. Beck and Najar (1977) found 11 augmented amicable pairs.

see also AMICABLE PAIR, DIVISOR FUNCTION, QUASIAMICABLE PAIR

References
Beck, W. E. and Najar, R. M. "More Reduced Amicable Pairs." *Fib. Quart.* **15**, 331–332, 1977.
Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 59, 1994.

Augmented Dodecahedron

see JOHNSON SOLID

Augmented Hexagonal Prism

see JOHNSON SOLID

Augmented Pentagonal Prism

see JOHNSON SOLID

Augmented Polyhedron

A UNIFORM POLYHEDRON with one or more other solids adjoined.

Augmented Sphenocorona

see JOHNSON SOLID

Augmented Triangular Prism

see JOHNSON SOLID

Augmented Tridiminished Icosahedron

see JOHNSON SOLID

Augmented Truncated Cube

see JOHNSON SOLID

Augmented Truncated Dodecahedron

see JOHNSON SOLID

Augmented Truncated Tetrahedron

see JOHNSON SOLID

Aureum Theorema

Gauss's name for the QUADRATIC RECIPROCITY THEOREM.

Aurifeuillean Factorization

A factorization of the form

$$2^{4n+2}+1=(2^{2n+1}-2^{n+1}+1)(2^{2n+1}+2^{n+1}+1). \quad (1)$$

The factorization for $n = 14$ was discovered by Aurifeuille, and the general form was subsequently discovered by Lucas. The large factors are sometimes written as L and M as follows

$$2^{4k-2}+1=(2^{2k-1}-2^k+1)(2^{2k-1}+2^k+1) \quad (2)$$

$$3^{6k-3}+1=(3^{2k-1}+1)(3^{2k-1}-3^k+1)(3^{2k-1}+3^k+1), \quad (3)$$

which can be written

$$2^{2h}+1=L_{2h}M_{2h} \quad (4)$$

$$3^{3h}+1=(3^h+1)L_{3h}M_{3h} \quad (5)$$

$$5^{5h}-1=(5^h-1)L_{5h}M_{5h}, \quad (6)$$

where $h \equiv 2k-1$ and

$$L_{2h}, M_{2h}=2^h+1\mp 2^k \quad (7)$$

$$L_{3h}, M_{3h}=3^h+1\mp 3^k \quad (8)$$

$$L_{5h}, M_{5h}=5^{2h}+3\cdot 5^h+1\mp 5^k(5^h+1). \quad (9)$$

see also GAUSS'S FORMULA

References

Brillhart, J.; Lehmer, D. H.; Selfridge, J.; Wagstaff, S. S. Jr.; and Tuckerman, B. *Factorizations of* $b^n \pm 1$, $b = 2, 3, 5, 6, 7, 10, 11, 12$ *Up to High Powers, rev. ed.* Providence, RI: Amer. Math. Soc., pp. lxviii–lxxii, 1988.

Wagstaff, S. S. Jr. "Aurifeullian Factorizations and the Period of the Bell Numbers Modulo a Prime." *Math. Comput.* **65**, 383–391, 1996.

Ausdehnungslehre

see EXTERIOR ALGEBRA

Authalic Latitude

An AUXILIARY LATITUDE which gives a SPHERE equal SURFACE AREA relative to an ELLIPSOID. The authalic latitude is defined by

$$\beta=\sin^{-1}\left(\frac{q}{q_p}\right), \quad (1)$$

where

$$q\equiv(1-e^2)\left[\frac{\sin\phi}{1-e^2\sin^2\phi}-\frac{1}{2e}\ln\left(\frac{1-e\sin\phi}{1+e\sin\phi}\right)\right], \quad (2)$$

and q_p is q evaluated at the north pole ($\phi = 90°$). Let R_q be the RADIUS of the SPHERE having the same SURFACE AREA as the ELLIPSOID, then

$$R_q=a\sqrt{\frac{q_p}{2}}. \quad (3)$$

The series for β is

$$\begin{aligned}\beta=\phi&-(\tfrac{1}{3}e^2+\tfrac{31}{180}e^4+\tfrac{59}{560}e^6+\ldots)\sin(2\phi)\\&+(\tfrac{17}{360}e^4+\tfrac{61}{1260}e^6+\ldots)\sin(4\phi)\\&-(\tfrac{383}{45360}e^6+\ldots)\sin(6\phi)+\ldots.\end{aligned} \quad (4)$$

The inverse FORMULA is found from

$$\Delta\phi=\frac{(1-e^2\sin^2\phi)^2}{2\cos\phi}\left[\frac{q}{1-e^2}-\frac{\sin\phi}{1-e^2\sin^2\phi}+\frac{1}{2e}\ln\left(\frac{1-e\sin\phi}{1+e\sin\phi}\right)\right], \quad (5)$$

where

$$q=q_p\sin\beta \quad (6)$$

and $\phi_0 = \sin^{-1}(q/2)$. This can be written in series form as

$$\begin{aligned}\phi &= \beta + (\tfrac{1}{3}e^2 + \tfrac{31}{180}e^4 + \tfrac{517}{5040}e^6 + \ldots)\sin(2\beta)\\ &\quad + (\tfrac{23}{360}e^4 + \tfrac{251}{3780}e^6 + \ldots)\sin(4\beta)\\ &\quad + (\tfrac{761}{45360}e^6 + \ldots)\sin(6\beta) + \ldots. \qquad (7)\end{aligned}$$

see also LATITUDE

References

Adams, O. S. "Latitude Developments Connected with Geodesy and Cartography with Tables, Including a Table for Lambert Equal-Area Meridional Projections." Spec. Pub. No. 67. U. S. Coast and Geodetic Survey, 1921.

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, p. 16, 1987.

Autocorrelation

The autocorrelation function is defined by

$$C_f(t) \equiv f \star f = f^*(-t) * f(t) = \int_{-\infty}^{\infty} f^*(\tau) f(t+\tau)\, d\tau, \qquad (1)$$

where $*$ denotes CONVOLUTION and $\star$ denotes CROSS-CORRELATION. A finite autocorrelation is given by

$$\begin{aligned}C_f(\tau) &\equiv \langle [y(t)-\bar{y}][y(t+\tau)-\bar{y}]\rangle \qquad (2)\\ &= \lim_{T\to\infty} \int_{-T/2}^{T/2} [y(t)-\bar{y}][y(t+\tau)-\bar{y}]\, dt. \qquad (3)\end{aligned}$$

If f is a REAL FUNCTION,

$$f^* = f, \qquad (4)$$

and an EVEN FUNCTION so that

$$f(-\tau) = f(\tau), \qquad (5)$$

then

$$C_f(t) = \int_{-\infty}^{\infty} f(\tau) f(t+\tau)\, d\tau. \qquad (6)$$

But let $\tau' \equiv -\tau$, so $d\tau' = -d\tau$, then

$$\begin{aligned}C_f(t) &= \int_{\infty}^{-\infty} f(-\tau) f(t-\tau)\,(-d\tau)\\ &= \int_{-\infty}^{\infty} f(-\tau) f(t-\tau)\, d\tau\\ &= \int_{-\infty}^{\infty} f(\tau) f(t-\tau)\, d\tau = f * f. \qquad (7)\end{aligned}$$

The autocorrelation discards phase information, returning only the POWER. It is therefore not reversible.

There is also a somewhat surprising and extremely important relationship between the autocorrelation and the FOURIER TRANSFORM known as the WIENER-KHINTCHINE THEOREM. Let $\mathcal{F}[f(x)] = F(k)$, and F^* denote the COMPLEX CONJUGATE of F, then the FOURIER TRANSFORM of the ABSOLUTE SQUARE of $F(k)$ is given by

$$\mathcal{F}[|F(k)|^2] = \int_{-\infty}^{\infty} f^*(\tau) f(\tau + x)\, d\tau. \qquad (8)$$

The autocorrelation is a HERMITIAN OPERATOR since $C_f(-t) = C_f{}^*(t)$. $f \star f$ is MAXIMUM at the ORIGIN. In other words,

$$\int_{-\infty}^{\infty} f(u) f(u+x)\, du \leq \int_{-\infty}^{\infty} f^2(u)\, du. \qquad (9)$$

To see this, let ϵ be a REAL NUMBER. Then

$$\int_{-\infty}^{\infty} [f(u) + \epsilon f(u+x)]^2\, du > 0 \qquad (10)$$

$$\int_{-\infty}^{\infty} f^2(u)\, du + 2\epsilon \int_{-\infty}^{\infty} f(u) f(u+x)\, du + \epsilon^2 \int_{-\infty}^{\infty} f^2(u+x)\, du > 0 \qquad (11)$$

$$\int_{-\infty}^{\infty} f^2(u)\, du + 2\epsilon \int_{-\infty}^{\infty} f(u) f(u+x)\, du + \epsilon^2 \int_{-\infty}^{\infty} f^2(u)\, du > 0. \qquad (12)$$

Define

$$\begin{aligned}a &\equiv \int_{-\infty}^{\infty} f^2(u)\, du \qquad (13)\\ b &\equiv 2\int_{-\infty}^{\infty} f(u) f(u+x)\, du. \qquad (14)\end{aligned}$$

Then plugging into above, we have $a\epsilon^2 + b\epsilon + c > 0$. This QUADRATIC EQUATION does not have any REAL ROOT, so $b^2 - 4ac \leq 0$, i.e., $b/2 \leq a$. It follows that

$$\int_{-\infty}^{\infty} f(u) f(u+x)\, du \leq \int_{-\infty}^{\infty} f^2(u)\, du, \qquad (15)$$

with the equality at $x = 0$. This proves that $f \star f$ is MAXIMUM at the ORIGIN.

see also CONVOLUTION, CROSS-CORRELATION, QUANTIZATION EFFICIENCY, WIENER-KHINTCHINE THEOREM

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Correlation and Autocorrelation Using the FFT." §13.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 538–539, 1992.

Automorphic Function

An automorphic function $f(z)$ of a COMPLEX variable z is one which is analytic (except for POLES) in a domain D and which is invariant under a DENUMERABLY INFINITE group of LINEAR FRACTIONAL TRANSFORMATIONS (also known as MÖBIUS TRANSFORMATIONS)

$$z' = \frac{az+b}{cz+d}.$$

Automorphic functions are generalizations of TRIGONOMETRIC FUNCTIONS and ELLIPTIC FUNCTIONS.

see also MODULAR FUNCTION, MÖBIUS TRANSFORMATIONS, ZETA FUCHSIAN

Automorphic Number

A number k such that nk^2 has its last digits equal to k is called n-automorphic. For example, $1 \cdot \underline{5}^2 = 2\underline{5}$ and $1 \cdot \underline{6}^2 = 3\underline{6}$ are 1-automorphic and $2 \cdot \underline{8}^2 = 12\underline{8}$ and $2 \cdot \underline{88}^2 = 154\underline{88}$ are 2-automorphic. de Guerre and Fairbairn (1968) give a history of automorphic numbers.

The first few 1-automorphic numbers are 1, 5, 6, 25, 76, 376, 625, 9376, 90625, ... (Sloane's A003226, Wells 1986, p. 130). There are two 1-automorphic numbers with a given number of digits, one ending in 5 and one in 6 (except that the 1-digit automorphic numbers include 1), and each of these contains the previous number with a digit prepended. Using this fact, it is possible to construct automorphic numbers having more than 25,000 digits (Madachy 1979). The first few 1-automorphic numbers ending with 5 are 5, 25, 625, 0625, 90625, ... (Sloane's A007185), and the first few ending with 6 are 6, 76, 376, 9376, 09376, ... (Sloane's A016090). The 1-automorphic numbers $a(n)$ ending in 5 are IDEMPOTENT (mod 10^n) since

$$[a(n)]^2 \equiv a(n) \pmod{10^n}$$

(Sloane and Plouffe 1995).

The following table gives the 10-digit n-automorphic numbers.

n	n-Automorphic Numbers	Sloane
1	0000000001, 8212890625, 1787109376	—, A007185, A016090
2	0893554688	A030984
3	6666666667, 7262369792, 9404296875	—, A030985, A030986
4	0446777344	A030987
5	3642578125	A030988
6	3631184896	A030989
7	7142857143, 4548984375, 1683872768	A030990, A030991, A030992
8	0223388672	A030993
9	5754123264, 3134765625, 8888888889	A030994, A030995, —

see also IDEMPOTENT, NARCISSISTIC NUMBER, NUMBER PYRAMID, TRIMORPHIC NUMBER

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. Item 59 in *HAKMEM*. Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Fairbairn, R. A. "More on Automorphic Numbers." *J. Recr. Math.* **2**, 170–174, 1969.

Fairbairn, R. A. Erratum to "More on Automorphic Numbers." *J. Recr. Math.* **2**, 245, 1969.

de Guerre, V. and Fairbairn, R. A. "Automorphic Numbers." *J. Recr. Math.* **1**, 173–179, 1968.

Hunter, J. A. H. "Two Very Special Numbers." *Fib. Quart.* **2**, 230, 1964.

Hunter, J. A. H. "Some Polyautomorphic Numbers." *J. Recr. Math.* **5**, 27, 1972.

Kraitchik, M. "Automorphic Numbers." §3.8 in *Mathematical Recreations.* New York: W. W. Norton, pp. 77–78, 1942.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 34–54 and 175–176, 1979.

Sloane, N. J. A. Sequences A016090, A003226/M3752, and A007185/M3940 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Wells, D. *The Penguin Dictionary of Curious and Interesting Numbers.* Middlesex: Penguin Books, pp. 171, 178, 191–192, 1986.

Automorphism

An ISOMORPHISM of a system of objects onto itself.

see also ANOSOV AUTOMORPHISM

Automorphism Group

The GROUP of functions from an object G to itself which preserve the structure of the object, denoted $\mathrm{Aut}(G)$. The automorphism group of a GROUP preserves the MULTIPLICATION table, the automorphism group of a GRAPH the INCIDENCE MATRICES, and that of a FIELD the ADDITION and MULTIPLICATION tables.

see also OUTER AUTOMORPHISM GROUP

Autonomous

A differential equation or system of ORDINARY DIFFERENTIAL EQUATIONS is said to be autonomous if it does not explicitly contain the independent variable (usually denoted t). A second-order autonomous differential equation is of the form $F(y, y', y'') = 0$, where $y' \equiv dy/dt \equiv v$. By the CHAIN RULE, y'' can be expressed as

$$y'' = v' = \frac{dv}{dt} = \frac{dv}{dy}\frac{dy}{dt} = \frac{dv}{dy}v.$$

For an autonomous ODE, the solution is independent of the time at which the initial conditions are applied. This means that all particles pass through a given point in phase space. A nonautonomous system of n first-order ODEs can be written as an autonomous system of $n+1$ ODEs by letting $t \equiv x_{n+1}$ and increasing the dimension of the system by 1 by adding the equation

$$\frac{dx_{n+1}}{dt} = 1.$$

Autoregressive Model

see MAXIMUM ENTROPY METHOD

Auxiliary Circle

The CIRCUMCIRCLE of an ELLIPSE, i.e., the CIRCLE whose center corresponds with that of the ELLIPSE and whose RADIUS is equal to the ELLIPSE's SEMIMAJOR AXIS.

see also CIRCLE, ECCENTRIC ANGLE, ELLIPSE

Auxiliary Latitude

see AUTHALIC LATITUDE, CONFORMAL LATITUDE, GEOCENTRIC LATITUDE, ISOMETRIC LATITUDE, LATITUDE, PARAMETRIC LATITUDE, RECTIFYING LATITUDE, REDUCED LATITUDE

Auxiliary Triangle

see MEDIAL TRIANGLE

Average

see MEAN

Average Absolute Deviation

$$\alpha \equiv \frac{1}{N}\sum_{i=1}^{N}|x_i-\mu| = \langle |x_i-\mu|\rangle .$$

see also ABSOLUTE DEVIATION, DEVIATION, STANDARD DEVIATION, VARIANCE

Average Function

If f is CONTINUOUS on a CLOSED INTERVAL $[a,b]$, then there is at least one number x^* in $[a,b]$ such that

$$\int_a^b f(x)dx = f(x^*)(b-a).$$

The average value of the FUNCTION $(\bar{f})$ on this interval is then given by $f(x^*)$.

see MEAN-VALUE THEOREM

Average Seek Time

see POINT-POINT DISTANCE—1-D

Ax-Kochen Isomorphism Theorem

Let P be the SET of PRIMES, and let $\mathbb{Q}_p$ and $Z_p(t)$ be the FIELDS of p-ADIC NUMBERS and formal POWER series over $Z_p = (0,1,\ldots,p-1)$. Further, suppose that D is a "nonprincipal maximal filter" on P. Then $\prod_{p\in P}\mathbb{Q}_p/D$ and $\prod_{p\in P}Z_p(t)/D$ are ISOMORPHIC.

see also HYPERREAL NUMBER, NONSTANDARD ANALYSIS

Axial Vector

see PSEUDOVECTOR

Axiom

A PROPOSITION regarded as self-evidently TRUE without PROOF. The word "axiom" is a slightly archaic synonym for POSTULATE. Compare CONJECTURE or HYPOTHESIS, both of which connote apparently TRUE *but not self-evident* statements.

see also ARCHIMEDES' AXIOM, AXIOM OF CHOICE, AXIOMATIC SYSTEM, CANTOR-DEDEKIND AXIOM, CONGRUENCE AXIOMS, CONJECTURE, CONTINUITY AXIOMS, COUNTABLE ADDITIVITY PROBABILITY AXIOM, DEDEKIND'S AXIOM, DIMENSION AXIOM, EILENBERG-STEENROD AXIOMS, EUCLID'S AXIOMS, EXCISION AXIOM, FANO'S AXIOM, FIELD AXIOMS, HAUSDORFF AXIOMS, HILBERT'S AXIOMS, HOMOTOPY AXIOM, INACCESSIBLE CARDINALS AXIOM, INCIDENCE AXIOMS, INDEPENDENCE AXIOM, INDUCTION AXIOM, LAW, LEMMA, LONG EXACT SEQUENCE OF A PAIR AXIOM, ORDERING AXIOMS, PARALLEL AXIOM, PASCH'S AXIOM, PEANO'S AXIOMS, PLAYFAIR'S AXIOM, PORISM, POSTULATE, PROBABILITY AXIOMS, PROCLUS' AXIOM, RULE, T2-SEPARATION AXIOM, THEOREM, ZERMELO'S AXIOM OF CHOICE, ZERMELO-FRAENKEL AXIOMS

Axiom A Diffeomorphism

Let $\phi : M \to M$ be a C^1 DIFFEOMORPHISM on a compact RIEMANNIAN MANIFOLD M. Then ϕ satisfies Axiom A if the NONWANDERING set $\Omega(\phi)$ of ϕ is hyperbolic and the PERIODIC POINTS of ϕ are DENSE in $\Omega(\phi)$. Although it was conjectured that the first of these conditions implies the second, they were shown to be independent in or around 1977. Examples include the ANOSOV DIFFEOMORPHISMS and SMALE HORSESHOE MAP.

In some cases, Axiom A can be replaced by the condition that the DIFFEOMORPHISM is a hyperbolic diffeomorphism on a hyperbolic set (Bowen 1975, Parry and Pollicott 1990).

see also ANOSOV DIFFEOMORPHISM, AXIOM A FLOW, DIFFEOMORPHISM, DYNAMICAL SYSTEM, RIEMANNIAN MANIFOLD, SMALE HORSESHOE MAP

References

Bowen, R. *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms.* New York: Springer-Verlag, 1975.

Ott, E. *Chaos in Dynamical Systems.* New York: Cambridge University Press, p. 143, 1993.

Parry, W. and Pollicott, M. "Zeta Functions and the Periodic Orbit Structure of Hyperbolic Dynamics." *Astérisque* No. 187–188, 1990.

Smale, S. "Differentiable Dynamical Systems." *Bull. Amer. Math. Soc.* **73**, 747–817, 1967.

Axiom A Flow

A FLOW defined analogously to the AXIOM A DIFFEOMORPHISM, except that instead of splitting the TANGENT BUNDLE into two invariant sub-BUNDLES, they are split into three (one exponentially contracting, one expanding, and one which is 1-dimensional and tangential to the flow direction).

see also DYNAMICAL SYSTEM

Axiom of Choice

An important and fundamental result in SET THEORY sometimes called ZERMELO'S AXIOM OF CHOICE. It was formulated by Zermelo in 1904 and states that, given any SET of mutually exclusive nonempty SETS, there exists at least one SET that contains exactly one element in common with each of the nonempty SETS.

It is related to HILBERT'S PROBLEM 1B, and was proved to be consistent with other AXIOMS in SET THEORY in 1940 by Gödel. In 1963, Cohen demonstrated that the axiom of choice is independent of the other AXIOMS in Cantorian SET THEORY, so the AXIOM cannot be proved within the system (Boyer and Merzbacher 1991, p. 610).

see also HILBERT'S PROBLEMS, SET THEORY, WELL-ORDERED SET, ZERMELO-FRAENKEL AXIOMS, ZORN'S LEMMA

References

Boyer, C. B. and Merzbacher, U. C. *A History of Mathematics, 2nd ed.* New York: Wiley, 1991.

Cohen, P. J. "The Independence of the Continuum Hypothesis." *Proc. Nat. Acad. Sci. U. S. A.* **50**, 1143–1148, 1963.

Cohen, P. J. "The Independence of the Continuum Hypothesis. II." *Proc. Nat. Acad. Sci. U. S. A.* **51**, 105–110, 1964.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 274–276, 1996.

Moore, G. H. *Zermelo's Axiom of Choice: Its Origin, Development, and Influence.* New York: Springer-Verlag, 1982.

Axiomatic Set Theory

A version of SET THEORY in which axioms are taken as uninterpreted rather than as formalizations of pre-existing truths.

see also NAIVE SET THEORY, SET THEORY

Axiomatic System

A logical system which possesses an explicitly stated SET of AXIOMS from which THEOREMS can be derived.

see also COMPLETE AXIOMATIC THEORY, CONSISTENCY, MODEL THEORY, THEOREM

Axis

A LINE with respect to which a curve or figure is drawn, measured, rotated, etc. The term is also used to refer to a LINE SEGMENT through a RANGE (Woods 1961).

see also ABSCISSA, ORDINATE, x-AXIS, y-AXIS, z-AXIS

References

Woods, F. S. *Higher Geometry: An Introduction to Advanced Methods in Analytic Geometry.* New York: Dover, p. 8, 1961.

Axonometry

A METHOD for mapping 3-D figures onto the PLANE.

see also CROSS-SECTION, MAP PROJECTION, POHLKE'S THEOREM, PROJECTION, STEREOLOGY

References

Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, p. 313, 1973.

Hazewinkel, M. (Managing Ed.). *Encyclopaedia of Mathematics: An Updated and Annotated Translation of the Soviet "Mathematical Encyclopaedia."* Dordrecht, Netherlands: Reidel, pp. 322–323, 1988.

Azimuthal Equidistant Projection

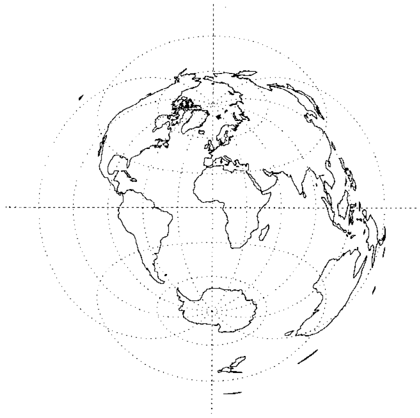

An AZIMUTHAL PROJECTION which is neither equal-AREA nor CONFORMAL. Let ϕ_1 and λ_0 be the LATITUDE and LONGITUDE of the center of the projection, then the transformation equations are given by

$$x = k' \cos\phi \sin(\lambda - \lambda_0) \tag{1}$$

$$y = k'[\cos\phi_1 \sin\phi - \sin\phi_1 \cos\phi \cos(\lambda - \lambda_0)]. \tag{2}$$

Here,

$$k' = \frac{c}{\sin c} \tag{3}$$

and

$$\cos c = \sin\phi_1 \sin\phi + \cos\phi_1 \cos\phi \cos(\lambda - \lambda_0), \tag{4}$$

where c is the angular distance from the center. The inverse FORMULAS are

$$\phi = \sin^{-1}\left(\cos c \sin\phi_1 + \frac{y \sin c \cos\phi_1}{c}\right) \tag{5}$$

and

$$\lambda = \begin{cases} \lambda_0 + \tan^{-1}\left(\frac{x \sin c}{c \cos\phi_1 \cos c - y \sin\phi_1 \sin c}\right) \\ \quad \text{for } \phi_1 \neq \pm 90^\circ \\ \lambda_0 + \tan^{-1}\left(-\frac{x}{y}\right) \\ \quad \text{for } \phi_1 = 90^\circ \\ \lambda_0 + \tan^{-1}\left(\frac{x}{y}\right), \\ \quad \text{for } \phi_1 = -90^\circ, \end{cases} \tag{6}$$

with the angular distance from the center given by

$$c = \sqrt{x^2 + y^2}. \tag{7}$$

References

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 191–202, 1987.

Azimuthal Projection

see AZIMUTHAL EQUIDISTANT PROJECTION, LAMBERT AZIMUTHAL EQUAL-AREA PROJECTION, ORTHOGRAPHIC PROJECTION, STEREOGRAPHIC PROJECTION

B

B^*-Algebra

A BANACH ALGEBRA with an ANTIAUTOMORPHIC INVOLUTION $*$ which satisfies

$$x^{**} = x \tag{1}$$

$$x^* y^* = (yx)^* \tag{2}$$

$$x^* + y^* = (x+y)^* \tag{3}$$

$$(cx)^* = \bar{c}x^* \tag{4}$$

and whose NORM satisfies

$$||xx^*|| = ||x||^2. \tag{5}$$

A C^*-ALGEBRA is a special type of B^*-algebra.

see also BANACH ALGEBRA, C^*-ALGEBRA

B_2-Sequence

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Also called a SIDON SEQUENCE. An INFINITE SEQUENCE of POSITIVE INTEGERS

$$1 \le b_1 < b_2 < b_3 < \ldots \tag{1}$$

such that all pairwise sums

$$b_i + b_j \tag{2}$$

for $i \le j$ are distinct (Guy 1994). An example is 1, 2, 4, 8, 13, 21, 31, 45, 66, 81, ... (Sloane's A005282).

Zhang (1993, 1994) showed that

$$S(B2) \equiv \sup_{\text{all B2 sequences}} \sum_{k=1}^{\infty} \frac{1}{b_k} > 2.1597. \tag{3}$$

The definition can be extended to B_n-sequences (Guy 1994).

see also A-SEQUENCE, MIAN-CHOWLA SEQUENCE

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/erdos/erdos.html`.

Guy, R. K. "Packing Sums of Pairs," "Three-Subsets with Distinct Sums," and "B_2-Sequences," and B_2-Sequences Formed by the Greedy Algorithm." §C9, C11, E28, and E32 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 115–118, 121–123, 228–229, and 232–233, 1994.

Sloane, N. J. A. Sequence A005282/M1094 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Zhang, Z. X. "A B2-Sequence with Larger Reciprocal Sum." *Math. Comput.* **60**, 835–839, 1993.

Zhang, Z. X. "Finding Finite B2-Sequences with Larger $m - a_m{}^{1/2}$." *Math. Comput.* **63**, 403–414, 1994.

B_p-Theorem

If $O_{p'}(G) = 1$ and if x is a p-element of G, then

$$L_{p'}(C_G(x) \le E(C_G(x)),$$

where $L_{p'}$ is the p-LAYER.

B-Spline

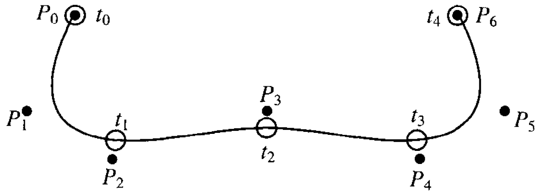

A generalization of the BÉZIER CURVE. Let a vector known as the KNOT VECTOR be defined

$$\mathbf{T} = \{t_0, t_1, \ldots, t_m\}, \tag{1}$$

where $\mathbf{T}$ is a nondecreasing SEQUENCE with $t_i \in [0,1]$, and define control points $\mathbf{P}_0, \ldots, \mathbf{P}_n$. Define the degree as

$$p \equiv m - n - 1. \tag{2}$$

The "knots" $t_{p+1}, \ldots, t_{m-p-1}$ are called INTERNAL KNOTS.

Define the basis functions as

$$N_{i,0}(t) = \begin{cases} 1 & \text{if } t_i \le t < t_{i+1} \text{ and } t_i < t_{i+1} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

$$N_{i,p}(t) = \frac{t - t_i}{t_{i+p} - t_i} N_{i,p-1}(t) + \frac{t_{i+p+1} - t}{t_{i+p+1} - t_{i+1}} N_{i+1,p-1}(t). \tag{4}$$

Then the curve defined by

$$\mathbf{C}(t) = \sum_{i=0}^{n} \mathbf{P}_i N_{i,p}(t) \tag{5}$$

is a B-spline. Specific types include the nonperiodic B-spline (first $p+1$ knots equal 0 and last $p+1$ equal to 1) and uniform B-spline (INTERNAL KNOTS are equally spaced). A B-Spline with no INTERNAL KNOTS is a BÉZIER CURVE.

The degree of a B-spline is independent of the number of control points, so a low order can always be maintained for purposes of numerical stability. Also, a curve is $p-k$ times differentiable at a point where k duplicate knot values occur. The knot values determine the extent of the control of the control points.

A nonperiodic B-spline is a B-spline whose first $p+1$ knots are equal to 0 and last $p+1$ knots are equal to 1. A uniform B-spline is a B-spline whose INTERNAL KNOTS are equally spaced.

see also BÉZIER CURVE, NURBS CURVE

B-Tree

B-trees were introduced by Bayer (1972) and McCreight. They are a special m-ary balanced tree used in databases because their structure allows records to be inserted, deleted, and retrieved with guaranteed worst-case performance. An n-node *B*-tree has height $\mathcal{O}(\lg 2)$, where LG is the LOGARITHM to base 2. The Apple® Macintosh® (Apple Computer, Cupertino, CA) HFS filing system uses *B*-trees to store disk directories (Benedict 1995). A *B*-tree satisfies the following properties:

1. The ROOT is either a LEAF (TREE) or has at least two CHILDREN.
2. Each node (except the ROOT and LEAVES) has between $\lceil m/2 \rceil$ and m CHILDREN, where $\lceil x \rceil$ is the CEILING FUNCTION.
3. Each path from the ROOT to a LEAF (TREE) has the same length.

Every 2-3 TREE is a *B*-tree of order 3. The number of *B*-trees of order $n = 1, 2, \ldots$ are 0, 1, 1, 1, 2, 2, 3, 4, 5, 8, 14, 23, 32, 43, 63, ... (Ruskey, Sloane's A014535).

see also RED-BLACK TREE

References
Aho, A. V.; Hopcroft, J. E.; and Ullmann, J. D. *Data Structures and Algorithms.* Reading, MA: Addison-Wesley, pp. 369–374, 1987.
Benedict, B. *Using Norton Utilities for the Macintosh.* Indianapolis, IN: Que, pp. B-17–B-33, 1995.
Beyer, R. "Symmetric Binary *B*-Trees: Data Structures and Maintenance Algorithms." *Acta Informat.* **1**, 290–306, 1972.
Ruskey, F. "Information on B-Trees." `http://sue.csc.uvic.ca/~cos/inf/tree/BTrees.html`.
Sloane, N. J. A. Sequence A014535 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Baby Monster Group

Also known as FISCHER'S BABY MONSTER GROUP. The SPORADIC GROUP B. It has ORDER

$$2^{41} \cdot 3^{13} \cdot 5^6 \cdot 7^2 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 23 \cdot 31 \cdot 47.$$

see also MONSTER GROUP

References
Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas/BM.html`.

BAC-CAB Identity

The VECTOR TRIPLE PRODUCT identity

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B}).$$

This identity can be generalized to n-D

$$\mathbf{a}_2 \times \cdots \times \mathbf{a}_{n-1} \times (\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}) = (-1)^{n+1} \begin{vmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_{n-1} \\ \mathbf{a}_2 \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_2 \cdot \mathbf{b}_{n-1} \\ \vdots & \ddots & \vdots \\ \mathbf{a}_{n-1} \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_{n-1} \cdot \mathbf{b}_{n-1} \end{vmatrix}.$$

see also LAGRANGE'S IDENTITY

BAC-CAB Rule

see BAC-CAB IDENTITY

Bachelier Function

see BROWN FUNCTION

Bachet's Conjecture

see LAGRANGE'S FOUR-SQUARE THEOREM

Bachet Equation

The DIOPHANTINE EQUATION

$$x^2 + k = y^3,$$

which is also an ELLIPTIC CURVE. The general equation is still the focus of ongoing study.

Backhouse's Constant

Let $P(x)$ be defined as the POWER series whose nth term has a COEFFICIENT equal to the nth PRIME,

$$P(x) \equiv \sum_{k=0}^{\infty} p_k x^k = 1 + 2x + 3x^2 + 5x^3 + 7x^4 + 11x^5 + \ldots,$$

and let $Q(x)$ be defined by

$$Q(x) = \frac{1}{P(x)} = \sum_{k=0}^{\infty} q_k x^k.$$

Then N. Backhouse conjectured that

$$\lim_{n\to\infty} \left| \frac{q_{n+1}}{q_n} \right| = 1.456074948582689671399595351116\ldots.$$

The constant was subsequently shown to exist by P. Flajolet.

References
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/backhous/backhous.html`.

Bäcklund Transformation

A method for solving classes of nonlinear PARTIAL DIFFERENTIAL EQUATIONS.

see also INVERSE SCATTERING METHOD

References
Infeld, E. and Rowlands, G. *Nonlinear Waves, Solitons, and Chaos.* Cambridge, England: Cambridge University Press, p. 196, 1990.
Miura, R. M. (Ed.) *Bäcklund Transformations, the Inverse Scattering Method, Solitons, and Their Applications.* New York: Springer-Verlag, 1974.

Backtracking

A method of drawing FRACTALS by appropriate numbering of the corresponding tree diagram which does not require storage of intermediate results.

Backus-Gilbert Method

A method which can be used to solve some classes of Integral Equations and is especially useful in implementing certain types of data inversion. It has been applied to invert seismic data to obtain density profiles in the Earth.

References

Backus, G. and Gilbert, F. "The Resolving Power of Growth Earth Data." *Geophys. J. Roy. Astron. Soc.* **16**, 169–205, 1968.

Backus, G. E. and Gilbert, F. "Uniqueness in the Inversion of Inaccurate Gross Earth Data." *Phil. Trans. Roy. Soc. London Ser. A* **266**, 123–192, 1970.

Loredo, T. J. and Epstein, R. I. "Analyzing Gamma-Ray Burst Spectral Data." *Astrophys. J.* **336**, 896–919, 1989.

Parker, R. L. "Understanding Inverse Theory." *Ann. Rev. Earth Planet. Sci.* **5**, 35–64, 1977.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Backus-Gilbert Method." §18.6 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 806–809, 1992.

Backward Difference

The backward difference is a Finite Difference defined by

$$\nabla_p \equiv \nabla f_p \equiv f_p - f_{p-1}. \qquad (1)$$

Higher order differences are obtained by repeated operations of the backward difference operator, so

$$\begin{aligned} \nabla_p^2 &= \nabla(\nabla p) = \nabla(f_p - f_{p-1}) = \nabla f_p - \nabla f_{p-1} && (2) \\ &= (f_p - f_{p-1}) - (f_{p-1} - f_{p-2}) \\ &= f_p - 2f_{p-1} + f_{p-2} && (3) \end{aligned}$$

In general,

$$\nabla_p^k \equiv \nabla^k f_p \equiv \sum_{m=0}^{k} (-1)^m \binom{k}{m} f_{p-k+m}, \qquad (4)$$

where $\binom{k}{m}$ is a Binomial Coefficient.

Newton's Backward Difference Formula expresses f_p as the sum of the nth backward differences

$$f_p = f_0 + p\nabla_0 + \tfrac{1}{2!}p(p+1)\nabla_0^2 + \tfrac{1}{3!}p(p+1)(p+2)\nabla_0^3 + \ldots, \qquad (5)$$

where ∇_0^n is the first nth difference computed from the difference table.

see also Adams' Method, Difference Equation, Divided Difference, Finite Difference, Forward Difference, Newton's Backward Difference Formula, Reciprocal Difference

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 429 and 433, 1987.

Bader-Deuflhard Method

A generalization of the Bulirsch-Stoer Algorithm for solving Ordinary Differential Equations.

References

Bader, G. and Deuflhard, P. "A Semi-Implicit Mid-Point Rule for Stiff Systems of Ordinary Differential Equations." *Numer. Math.* **41**, 373–398, 1983.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 730, 1992.

Baguenaudier

A Puzzle involving disentangling a set of rings from a looped double rod (also called Chinese Rings). The minimum number of moves needed for n rings is

$$\begin{cases} \frac{1}{3}(2^{n+1} - 2) & n \text{ even} \\ \frac{1}{3}(2^{n+1} - 1) & n \text{ odd.} \end{cases}$$

By simultaneously moving the two end rings, the number of moves can be reduced to

$$\begin{cases} 2^{n-1} - 1 & n \text{ even} \\ 2^{n-1} & n \text{ odd.} \end{cases}$$

The solution of the baguenaudier is intimately related to the theory of Gray Codes.

References

Dubrovsky, V. "Nesting Puzzles, Part II: Chinese Rings Produce a Chinese Monster." *Quantum* **6**, 61–65 (Mar.) and 58–59 (Apr.), 1996.

Gardner, M. "The Binary Gray Code." In *Knotted Doughnuts and Other Mathematical Entertainments.* New York: W. H. Freeman, pp. 15–17, 1986.

Kraitchik, M. "Chinese Rings." §3.12.3 in *Mathematical Recreations.* New York: W. W. Norton, pp. 89–91, 1942.

Steinhaus, H. *Mathematical Snapshots, 3rd American ed.* New York: Oxford University Press, p. 268, 1983.

Bailey's Method

see Lambert's Method

Bailey's Theorem

Let $\Gamma(z)$ be the Gamma Function, then

$$\left[\frac{\Gamma(m+\frac{1}{2})}{\Gamma(m)}\right]^2 \underbrace{\left[\frac{1}{m} + \left(\frac{1}{2}\right)^2 \frac{1}{m+1} + \left(\frac{1\cdot 3}{2\cdot 4}\right)^2 \frac{1}{m+2} + \ldots\right]}_{n}$$

$$= \left[\frac{\Gamma(n+\frac{1}{2})}{\Gamma(n)}\right]^2 \underbrace{\left[\frac{1}{n} + \left(\frac{1}{2}\right)^2 \frac{1}{n+1} + \left(\frac{1\cdot 3}{2\cdot 4}\right)^2 \frac{1}{n+2} + \ldots\right]}_{m}.$$

Baire Category Theorem

A nonempty complete Metric Space cannot be represented as the Union of a Countable family of nowhere Dense Subsets.

Baire Space

A TOPOLOGICAL SPACE X in which each SUBSET of X of the "first category" has an empty interior. A TOPOLOGICAL SPACE which is HOMEOMORPHIC to a complete METRIC SPACE is a Baire space.

Bairstow's Method

A procedure for finding the quadratic factors for the COMPLEX CONJUGATE ROOTS of a POLYNOMIAL $P(x)$ with REAL COEFFICIENTS.

$$[x-(a+ib)][x-(a-ib)] = x^2+2ax+(a^2+b^2) \equiv x^2+Bx+C. \quad (1)$$

Now write the original POLYNOMIAL as

$$P(x) = (x^2+Bx+C)Q(x)+Rx+S \quad (2)$$

$$R(B+\delta B, C+\delta C) \approx R(B,C)+\frac{\partial R}{\partial B}\,dB+\frac{\partial R}{\partial C}\,dC \quad (3)$$

$$S(B+\delta B, C+\delta C) \approx S(B,C)+\frac{\partial S}{\partial B}\,dB+\frac{\partial S}{\partial C}\,dC \quad (4)$$

$$\frac{\partial P}{\partial C} = 0 = (x^2+Bx+C)\frac{\partial Q}{\partial C}+Q(x)+\frac{\partial R}{\partial C}+\frac{\partial S}{\partial C} \quad (5)$$

$$-Q(x) = (x^2+Bx+C)\frac{\partial Q}{\partial C}+\frac{\partial R}{\partial C}+\frac{\partial S}{\partial C} \quad (6)$$

$$\frac{\partial P}{\partial B} = 0 = (x^2+Bx+C)\frac{\partial Q}{\partial B}+xQ(x)+\frac{\partial R}{\partial B}+\frac{\partial S}{\partial B} \quad (7)$$

$$-xQ(x) = (x^2+Bx+C)\frac{\partial Q}{\partial B}+\frac{\partial R}{\partial B}+\frac{\partial S}{\partial B}. \quad (8)$$

Now use the 2-D NEWTON'S METHOD to find the simultaneous solutions.

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge, England: Cambridge University Press, pp. 277 and 283–284, 1989.

Baker's Dozen

The number 13.

see also 13, DOZEN

Baker's Map

The MAP

$$x_{n+1} = 2\mu x_n, \quad (1)$$

where x is computed modulo 1. A generalized Baker's map can be defined as

$$x_{n+1} = \begin{cases} \lambda_a x_n & y_n < \alpha \\ (1-\lambda_b)+\lambda_b x_n & y_n > \alpha \end{cases} \quad (2)$$

$$y_{n+1} = \begin{cases} \frac{y_n}{\alpha} & y_n < \alpha \\ \frac{y_n-\alpha}{\beta} & y_n > \alpha, \end{cases} \quad (3)$$

where $\beta \equiv 1-\alpha$, $\lambda_a+\lambda_b \leq 1$, and x and y are computed mod 1. The $q=1$ q-DIMENSION is

$$D_1 = 1+\frac{\alpha\ln\left(\frac{1}{\alpha}\right)+\beta\ln\left(\frac{1}{\beta}\right)}{\alpha\ln\left(\frac{1}{\lambda_a}\right)+\beta\ln\left(\frac{1}{\lambda_b}\right)}. \quad (4)$$

If $\lambda_a = \lambda_b$, then the general q-DIMENSION is

$$D_q = 1+\frac{1}{q-1}\frac{\ln(\alpha^q+\beta^q)}{\ln\lambda_a}. \quad (5)$$

References

Lichtenberg, A. and Lieberman, M. *Regular and Stochastic Motion.* New York: Springer-Verlag, p. 60, 1983.

Ott, E. *Chaos in Dynamical Systems.* Cambridge, England: Cambridge University Press, pp. 81–82, 1993.

Rasband, S. N. *Chaotic Dynamics of Nonlinear Systems.* New York: Wiley, p. 32, 1990.

Balanced ANOVA

An ANOVA in which the number of REPLICATES (sets of identical observations) is restricted to be the same for each FACTOR LEVEL (treatment group).

see also ANOVA

Balanced Incomplete Block Design

see BLOCK DESIGN

Ball

The n-ball, denoted $\mathbb{B}^n$, is the interior of a SPHERE $\mathbb{S}^{n-1}$, and sometimes also called the n-DISK. (Although physicists often use the term "SPHERE" to mean the solid ball, mathematicians definitely do not!) Let $\mathrm{Vol}(\mathbb{B}^n)$ denote the volume of an n-D ball of RADIUS r. Then

$$\sum_{n=0}^{\infty} \mathrm{Vol}(B^n) = e^{\pi r^2}[1+\mathrm{erf}(r\sqrt{\pi})],$$

where $\mathrm{erf}(x)$ is the ERF function.

see also ALEXANDER'S HORNED SPHERE, BANACH-TARSKI PARADOX, BING'S THEOREM, BISHOP'S INEQUALITY, BOUNDED, DISK, HYPERSPHERE, SPHERE, WILD POINT

References

Freden, E. Problem 10207. "Summing a Series of Volumes." *Amer. Math. Monthly* **100**, 882, 1993.

Ball Triangle Picking

The determination of the probability for obtaining an OBTUSE TRIANGLE by picking 3 points at random in the unit DISK was generalized by Hall (1982) to the n-D BALL. Buchta (1986) subsequently gave closed form evaluations for Hall's integrals, with the first few solutions being

$$\begin{aligned} P_2 &= \frac{9}{8} - \frac{4}{\pi^2} \approx 0.72 \\ P_3 &= \tfrac{37}{70} \approx 0.53 \\ P_4 &\approx 0.39 \\ P_5 &\approx 0.29. \end{aligned}$$

The case P_2 corresponds to the usual DISK case.

see also CUBE TRIANGLE PICKING, OBTUSE TRIANGLE

References
Buchta, C. "A Note on the Volume of a Random Polytope in a Tetrahedron." *Ill. J. Math.* **30**, 653–659, 1986.
Hall, G. R. "Acute Triangles in the n-Ball." *J. Appl. Prob.* **19**, 712–715, 1982.

Ballantine

see BORROMEAN RINGS

Ballieu's Theorem

For any set $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)$ of POSITIVE numbers with $\mu_0 = 0$ and

$$M_\mu = \max_{0 \leq k \leq n-1} \frac{\mu_k + \mu_n |b_{n-k}|}{\mu_{k+1}}.$$

Then all the EIGENVALUES λ satisfying $P(\lambda) = 0$, where $P(\lambda)$ is the CHARACTERISTIC POLYNOMIAL, lie on the DISK $|z| \leq M_\mu$.

References
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1119, 1979.

Ballot Problem

Suppose A and B are candidates for office and there are $2n$ voters, n voting for A and n for B. In how many ways can the ballots be counted so that A is always ahead of or tied with B? The solution is a CATALAN NUMBER C_n.

A related problem also called "the" ballot problem is to let A receive a votes and B b votes with $a > b$. This version of the ballot problem then asks for the probability that A stays ahead of B as the votes are counted (Vardi 1991). The solution is $(a-b)/(a+b)$, as first shown by M. Bertrand (Hilton and Pedersen 1991). Another elegant solution was provided by André (1887) using the so-called ANDRÉ'S REFLECTION METHOD.

The problem can also be generalized (Hilton and Pedersen 1991). Furthermore, the TAK FUNCTION is connected with the ballot problem (Vardi 1991).

see also ANDRÉ'S REFLECTION METHOD, CATALAN NUMBER, TAK FUNCTION

References
André, D. "Solution directe du problème résolu par M. Bertrand." *Comptes Rendus Acad. Sci. Paris* **105**, 436–437, 1887.
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 49, 1987.
Carlitz, L. "Solution of Certain Recurrences." *SIAM J. Appl. Math.* **17**, 251–259, 1969.
Comtet, L. *Advanced Combinatorics.* Dordrecht, Netherlands: Reidel, p. 22, 1974.
Feller, W. *An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd ed.* New York: Wiley, pp. 67–97, 1968.
Hilton, P. and Pedersen, J. "The Ballot Problem and Catalan Numbers." *Nieuw Archief voor Wiskunde* **8**, 209–216, 1990.
Hilton, P. and Pedersen, J. "Catalan Numbers, Their Generalization, and Their Uses." *Math. Intel.* **13**, 64–75, 1991.
Kraitchik, M. "The Ballot-Box Problem." §6.13 in *Mathematical Recreations.* New York: W. W. Norton, p. 132, 1942.
Motzkin, T. "Relations Between Hypersurface Cross Ratios, and a Combinatorial Formula for Partitions of a Polygon, for Permanent Preponderance, and for Non-Associative Products." *Bull. Amer. Math. Soc.* **54**, 352–360, 1948.
Vardi, I. *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 185–187, 1991.

Banach Algebra

An ALGEBRA A over a FIELD F with a NORM that makes A into a COMPLETE METRIC SPACE, and therefore, a BANACH SPACE. F is frequently taken to be the COMPLEX NUMBERS in order to assure that the SPECTRUM fully characterizes an OPERATOR (i.e., the spectral theorems for normal or compact normal operators do not, in general, hold in the SPECTRUM over the REAL NUMBERS).

see also B^*-ALGEBRA

Banach Fixed Point Theorem

Let f be a contraction mapping from a closed SUBSET F of a BANACH SPACE E into F. Then there exists a unique $z \in F$ such that $f(z) = z$.

see also FIXED POINT THEOREM

References
Debnath, L. and Mikusiński, P. *Introduction to Hilbert Spaces with Applications.* San Diego, CA: Academic Press, 1990.

Banach-Hausdorff-Tarski Paradox

see BANACH-TARSKI PARADOX

Banach Measure

An "AREA" which can be defined for every set—even those without a true geometric AREA—which is rigid and finitely additive.

Banach Space

A normed linear SPACE which is COMPLETE in the norm-determined METRIC. A HILBERT SPACE is always a Banach space, but the converse need not hold.

see also BESOV SPACE, HILBERT SPACE, SCHAUDER FIXED POINT THEOREM

Banach-Steinhaus Theorem

see UNIFORM BOUNDEDNESS PRINCIPLE

Banach-Tarski Paradox

First stated in 1924, this theorem demonstrates that it is possible to dissect a BALL into six pieces which can be reassembled by rigid motions to form two balls of the same size as the original. The number of pieces was subsequently reduced to five. However, the pieces are extremely complicated. A generalization of this theorem is that any two bodies in $\mathbb{R}^3$ which do not extend to infinity and each containing a ball of arbitrary size can be dissected into each other (they are are EQUIDECOMPOSABLE).

References

Stromberg, K. "The Banach-Tarski Paradox." *Amer. Math. Monthly* **86**, 3, 1979.

Wagon, S. *The Banach-Tarski Paradox.* New York: Cambridge University Press, 1993.

Bang's Theorem

The lines drawn to the VERTICES of a face of a TETRAHEDRON from the point of contact of the FACE with the INSPHERE form three ANGLES at the point of contact which are the same three ANGLES in each FACE.

References

Brown, B. H. "Theorem of Bang. Isosceles Tetrahedra." *Amer. Math. Monthly* **33**, 224–226, 1926.

Honsberger, R. *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., p. 93, 1976.

Bankoff Circle

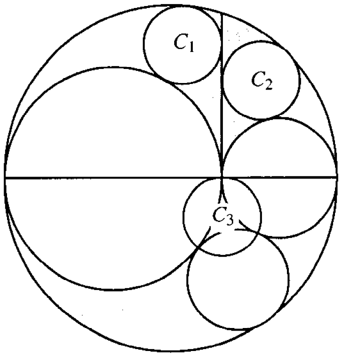

In addition to the ARCHIMEDES' CIRCLES C_1 and C_2 in the ARBELOS figure, there is a third circle C_3 congruent to these two as illustrated in the above figure.

see also ARBELOS

References

Bankoff, L. "Are the Twin Circles of Archimedes Really Twins?" *Math. Mag.* **47**, 214–218, 1974.

Gardner, M. "Mathematical Games: The Diverse Pleasures of Circles that Are Tangent to One Another." *Sci. Amer.* **240**, 18–28, Jan. 1979.

Banzhaf Power Index

The number of ways in which a group of n with weights $\sum_{i=1}^{n} w_i = 1$ can change a losing coalition (one with $\sum w_i < 1/2$) to a winning one, or vice versa. It was proposed by the lawyer J. F. Banzhaf in 1965.

References

Paulos, J. A. *A Mathematician Reads the Newspaper.* New York: BasicBooks, pp. 9–10, 1995.

Bar (Edge)

The term in rigidity theory for the EDGES of a GRAPH.

see also CONFIGURATION, FRAMEWORK

Bar Polyhex

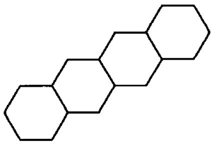

A POLYHEX consisting of HEXAGONS arranged along a line.

see also BAR POLYIAMOND

References

Gardner, M. *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, p. 147, 1978.

Bar Polyiamond

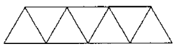

A POLYIAMOND consisting of EQUILATERAL TRIANGLES arranged along a line.

see also BAR POLYHEX

References

Golomb, S. W. *Polyominoes: Puzzles, Patterns, Problems, and Packings, 2nd ed.* Princeton, NJ: Princeton University Press, p. 92, 1994.

Barber Paradox

A man of Seville is shaved by the Barber of Seville IFF the man does not shave himself. Does the barber shave himself? Proposed by Bertrand Russell.

Barbier's Theorem

All CURVES OF CONSTANT WIDTH of width w have the same PERIMETER πw.

Bare Angle Center

The TRIANGLE CENTER with TRIANGLE CENTER FUNCTION

$$\alpha = A.$$

References

Kimberling, C. "Major Centers of Triangles." *Amer. Math. Monthly* **104**, 431–438, 1997.

Barnes G-Function

see G-FUNCTION

Barnes' Lemma

If a CONTOUR in the COMPLEX PLANE is curved such that it separates the increasing and decreasing sequences of POLES, then

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\Gamma(\alpha+s)\Gamma(\beta+s)\Gamma(\gamma-s)\Gamma(\delta-s)\,ds = \frac{\Gamma(\alpha+\gamma)\Gamma(\alpha+\delta)\Gamma(\beta+\gamma)\Gamma(\beta+\delta)}{\Gamma(\alpha+\beta+\gamma+\delta)},$$

where $\Gamma(z)$ is the GAMMA FUNCTION.

Barnes-Wall Lattice

A lattice which can be constructed from the LEECH LATTICE Λ_{24}.

see also COXETER-TODD LATTICE, LATTICE POINT, LEECH LATTICE

References

Barnes, E. S. and Wall, G. E. "Some Extreme Forms Defined in Terms of Abelian Groups." *J. Austral. Math. Soc.* **1**, 47–63, 1959.

Conway, J. H. and Sloane, N. J. A. "The 16-Dimensional Barnes-Wall Lattice Λ_{16}." §4.10 in *Sphere Packings, Lattices, and Groups, 2nd ed.* New York: Springer-Verlag, pp. 127–129, 1993.

Barnsley's Fern

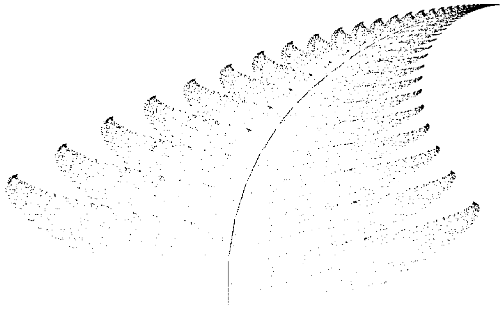

The ATTRACTOR of the ITERATED FUNCTION SYSTEM given by the set of "fern functions"

$$f_1(x,y) = \begin{bmatrix} 0.85 & 0.04 \\ -0.04 & 0.85 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0.00 \\ 1.60 \end{bmatrix} \tag{1}$$

$$f_2(x,y) = \begin{bmatrix} -0.15 & 0.28 \\ 0.26 & 0.24 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0.00 \\ 0.44 \end{bmatrix} \tag{2}$$

$$f_3(x,y) = \begin{bmatrix} 0.20 & -0.26 \\ 0.23 & 0.22 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0.00 \\ 1.60 \end{bmatrix} \tag{3}$$

$$f_4(x,y) = \begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.16 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \tag{4}$$

(Barnsley 1993, p. 86; Wagon 1991). These AFFINE TRANSFORMATIONS are contractions. The tip of the fern (which resembles the black spleenwort variety of fern) is the fixed point of f_1, and the tips of the lowest two branches are the images of the main tip under f_2 and f_3 (Wagon 1991).

see also DYNAMICAL SYSTEM, FRACTAL, ITERATED FUNCTION SYSTEM

References

Barnsley, M. *Fractals Everywhere, 2nd ed.* Boston, MA: Academic Press, pp. 86, 90, 102 and Plate 2, 1993.

Gleick, J. *Chaos: Making a New Science.* New York: Penguin Books, p. 238, 1988.

Wagon, S. "Biasing the Chaos Game: Barnsley's Fern." §5.3 in *Mathematica in Action.* New York: W. H. Freeman, pp. 156–163, 1991.

Barrier

A number n is called a barrier of a number-theoretic function $f(m)$ if, for all $m < n$, $m + f(m) \leq n$. Neither the TOTIENT FUNCTION $\phi(n)$ nor the DIVISOR FUNCTION $\sigma(n)$ has barriers.

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 64–65, 1994.

Barth Decic

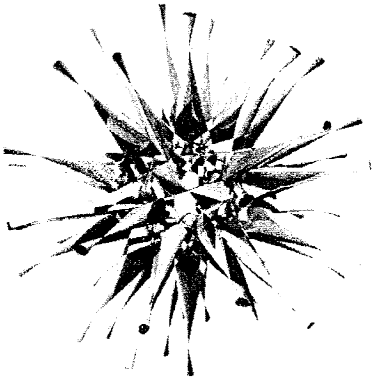

The Barth decic is a Decic Surface in complex three-dimensional projective space having the maximum possible number of Ordinary Double Points (345). It is given by the implicit equation

$$\begin{aligned}&8(x^2-\phi^4y^2)(y^2-\phi^4z^2)(z^2-\phi^4x^2)\\&\quad\times(x^4+y^4+z^4-2x^2y^2-2x^2z^2-2y^2z^2)\\&+(3+5\phi)(x^2+y^2+z^2-w^2)^2[x^2+y^2+z^2-(2-\phi)w^2]^2w^2\\&\qquad=0,\end{aligned}$$

where ϕ is the Golden Mean and w is a parameter (Endraß, Nordstrand), taken as $w=1$ in the above plot. The Barth decic is invariant under the Icosahedral Group.

see also Algebraic Surface, Barth Sextic, Decic Surface, Ordinary Double Point

References

Barth, W. "Two Projective Surfaces with Many Nodes Admitting the Symmetries of the Icosahedron." *J. Alg. Geom.* **5**, 173–186, 1996.

Endraß, S. "Flächen mit vielen Doppelpunkten." *DMV-Mitteilungen* **4**, 17–20, 4/1995.

Endraß, S. "Barth's Decic." `http://www.mathematik.uni-mainz.de/AlgebraischeGeometrie/docs/Ebarthdecic.shtml`.

Nordstrand, T. "Batch Decic." `http://www.uib.no/people/nfytn/bdectxt.htm`.

Barth Sextic

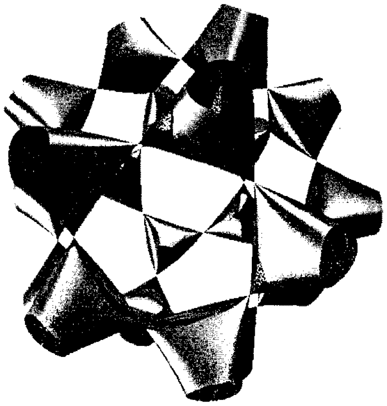

The Barth-sextic is a Sextic Surface in complex three-dimensional projective space having the maximum possible number of Ordinary Double Points (65). It is given by the implicit equation

$$\begin{aligned}&4(\phi^2x^2-y^2)(\phi^2y^2-z^2)(\phi^2z^2-x^2)\\&\qquad-(1+2\phi)(x^2+y^2+z^2-w^2)^2w^2=0.\end{aligned}$$

where ϕ is the Golden Mean, and w is a parameter (Endraß, Nordstrand), taken as $w=1$ in the above plot. The Barth sextic is invariant under the Icosahedral Group. Under the map

$$(x,y,z,w)\to(x^2,y^2,z^2,w^2),$$

the surface is the eightfold cover of the Cayley Cubic (Endraß).

see also Algebraic Surface, Barth Decic, Cayley Cubic, Ordinary Double Point, Sextic Surface

References

Barth, W. "Two Projective Surfaces with Many Nodes Admitting the Symmetries of the Icosahedron." *J. Alg. Geom.* **5**, 173–186, 1996.

Endraß, S. "Flächen mit vielen Doppelpunkten." *DMV-Mitteilungen* **4**, 17–20, 4/1995.

Endraß, S. "Barth's Sextic." `http://www.mathematik.uni-mainz.de/AlgebraischeGeometrie/docs/Ebarthsextic.shtml`.

Nordstrand, T. "Barth Sextic." `http://www.uib.no/people/nfytn/sexttxt.htm`.

Bartlett Function

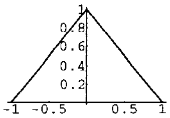

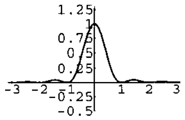

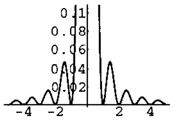

The Apodization Function

$$f(x)=1-\frac{|x|}{a}\tag{1}$$

which is a generalization of the one-argument Triangle Function. Its Full Width at Half Maximum is a. It has Instrument Function

$$\begin{aligned}I(x)&=\int_{-a}^{a}e^{-2\pi ikx}\left(1-\frac{|x|}{a}\right)dx\\&=\int_{-a}^{0}e^{-2\pi ikx}\left(1+\frac{x}{a}\right)dx\\&\quad+\int_{0}^{a}e^{-2\pi ikx}\left(1-\frac{x}{a}\right)dx.\end{aligned}\tag{2}$$

Letting $x'\equiv -x$ in the first part therefore gives

$$\begin{aligned}\int_{-a}^{0}e^{-2\pi ikx}\left(1+\frac{x}{a}\right)dx&=\int_{a}^{0}e^{2\pi ikx'}\left(1-\frac{x'}{a}\right)(-dx')\\&=\int_{0}^{a}e^{2\pi ikx}\left(1-\frac{x}{a}\right)dx.\end{aligned}\tag{3}$$

Rewriting (2) using (3) gives

$$\begin{aligned}I(x)&=(e^{2\pi ikx}+e^{-2\pi ikx})\left(1-\frac{x}{a}\right)dx\\&=2\int_{0}^{a}\cos(2\pi kx)\left(1-\frac{x}{a}\right)dx.\end{aligned}\tag{4}$$

Integrating the first part and using the integral

$$\int x\cos(bx)\,dx=\frac{1}{b^2}\cos(bx)+\frac{x}{b}\sin(bx)\tag{5}$$

for the second part gives

$$\begin{aligned} I(x) &= 2\left[\frac{\sin(2\pi kx)}{2\pi k}\right. \\ &\quad \left.-\frac{1}{a}\left\{\frac{1}{4\pi^2k^2}\cos(2\pi kx)+\frac{x}{2\pi k}\sin(2\pi kx)\right\}\right]_0^a \\ &= 2\left\{\left[\frac{\sin(2\pi ka)}{2\pi k}-0\right]\right. \\ &\quad \left.-\frac{1}{a}\left[\frac{\cos(2\pi ka)-1}{4\pi^2k^2}+\frac{a\sin(2\pi ka)}{2\pi k}\right]\right\} \\ &= \frac{1}{2\pi^2ak^2}[\cos(2\pi ka)-1] = a\frac{\sin^2(\pi ka)}{\pi^2k^2a^2} \\ &= a\operatorname{sinc}^2(\pi ka), \end{aligned} \tag{6}$$

where $\operatorname{sinc} x$ is the SINC FUNCTION. The peak (in units of a) is 1. The function $I(x)$ is always positive, so there are no NEGATIVE sidelobes. The extrema are given by letting $\beta \equiv \pi ka$ and solving

$$\frac{d}{d\beta}\left(\frac{\sin\beta}{\beta}\right)^2 = 2\frac{\sin\beta}{\beta}\frac{\sin\beta-\beta\cos\beta}{\beta^2} = 0 \tag{7}$$

$$\sin\beta(\sin\beta-\beta\cos\beta) = 0 \tag{8}$$

$$\sin\beta-\beta\cos\beta = 0 \tag{9}$$

$$\tan\beta = \beta. \tag{10}$$

Solving this numerically gives $\beta = 4.49341$ for the first maximum, and the peak POSITIVE sidelobe is 0.047190. The full width at half maximum is given by setting $x \equiv \pi ka$ and solving

$$\operatorname{sinc}^2 x = \tfrac{1}{2} \tag{11}$$

for $x_{1/2}$, yielding

$$x_{1/2} = \pi k_{1/2}a = 1.39156. \tag{12}$$

Therefore, with $L \equiv 2a$,

$$\text{FWHM} = 2k_{1/2} = \frac{0.885895}{a} = \frac{1.77179}{L}. \tag{13}$$

see also APODIZATION FUNCTION, PARZEN APODIZATION FUNCTION, TRIANGLE FUNCTION

References

Bartlett, M. S. "Periodogram Analysis and Continuous Spectra." *Biometrika* **37**, 1–16, 1950.

Barycentric Coordinates

Also known as HOMOGENEOUS COORDINATES or TRILINEAR COORDINATES.

see TRILINEAR COORDINATES

Base Curve

see DIRECTRIX (RULED SURFACE)

Base (Logarithm)

The number used to define a LOGARITHM, which is then written $\log_b$. The symbol $\log x$ is an abbreviation for $\log_{10} x$, $\ln x$ for $\log_e x$ (the NATURAL LOGARITHM), and $\lg x$ for $\log_2 x$.

see also E, LG, LN, LOGARITHM, NAPIERIAN LOGARITHM, NATURAL LOGARITHM

Base (Neighborhood System)

A base for a neighborhood system of a point x is a collection N of OPEN SETS such that x belongs to every member of N, and any OPEN SET containing x also contains a member of N as a SUBSET.

Base (Number)

A REAL NUMBER x can be represented using any INTEGER number b as a base (sometimes also called a RADIX or SCALE). The choice of a base yields to a representation of numbers known as a NUMBER SYSTEM. In base b, the DIGITS 0, 1, ..., $b-1$ are used (where, by convention, for bases larger than 10, the symbols A, B, C, ... are generally used as symbols representing the DECIMAL numbers 10, 11, 12, ...).

Base	Name
2	binary
3	ternary
4	quaternary
5	quinary
6	senary
7	septenary
8	octal
9	nonary
10	decimal
11	undenary
12	duodecimal
16	hexadecimal
20	vigesimal
60	sexagesimal

Let the base b representation of a number x be written

$$(a_n\, a_{n-1}\, \ldots\, a_0 . a_{-1}\, \ldots)_b, \tag{1}$$

(e.g., 123.456_{10}), then the index of the leading DIGIT needed to represent the number is

$$n \equiv \lfloor \log_b x \rfloor, \tag{2}$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION. Now, recursively compute the successive DIGITS

$$a_i = \left\lfloor \frac{r_i}{b^i} \right\rfloor, \tag{3}$$

where $r_n \equiv x$ and

$$r_{i-1} = r_i - a_i b^i \tag{4}$$

for $i = n, n-1, \ldots, 1, 0, \ldots$. This gives the base b representation of x. Note that if x is an INTEGER, then i need only run through 0, and that if x has a fractional part, then the expansion may or may not terminate. For example, the HEXADECIMAL representation of 0.1 (which terminates in DECIMAL notation) is the infinite expression $0.19999\ldots_h$.

Some number systems use a mixture of bases for counting. Examples include the Mayan calendar and the old British monetary system (in which ha'pennies, pennies, threepence, sixpence, shillings, half crowns, pounds, and guineas corresponded to units of 1/2, 1, 3, 6, 12, 30, 240, and 252, respectively).

Knuth has considered using TRANSCENDENTAL bases. This leads to some rather unfamiliar results, such as equating π to 1 in "base π," $\pi = 1_\pi$.

see also BINARY, DECIMAL, HEREDITARY REPRESENTATION, HEXADECIMAL, OCTAL, QUATERNARY, SEXAGESIMAL, TERNARY, VIGESIMAL

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 28, 1972.

Bogomolny, A. "Base Converter." `http://www.cut-the-knot.com/binary.html`.

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 6–11, 1991.

Weisstein, E. W. "Bases." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Bases.m`.

Base Space

The SPACE B of a FIBER BUNDLE given by the MAP $f: E \to B$, where E is the TOTAL SPACE of the FIBER BUNDLE.

see also FIBER BUNDLE, TOTAL SPACE

Baseball

The numbers 3 and 4 appear prominently in the game of baseball. There are $3 \cdot 3 = 9$ innings in a game, and three strikes are an out. However, 4 balls are needed for a walk. The number of bases can either be regarded as 3 (excluding HOME PLATE) or 4 (including it).

see BASEBALL COVER, HOME PLATE

Baseball Cover

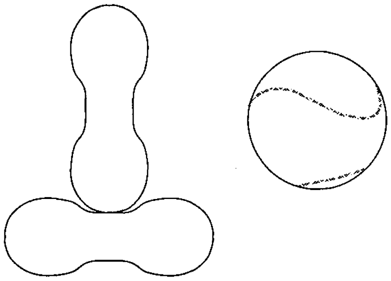

A pair of identical plane regions (mirror symmetric about two perpendicular lines through the center) which can be stitched together to form a baseball (or tennis ball). A baseball has a CIRCUMFERENCE of 9 1/8 inches. The practical consideration of separating the regions far enough to allow the pitcher a good grip requires that the "neck" distance be about 1 3/16 inches. The baseball cover was invented by Elias Drake as a boy in the 1840s. (Thompson's attribution of the current design to trial and error development by C. H. Jackson in the 1860s is apparently unsubstantiated, as discovered by George Bart.)

One way to produce a baseball cover is to draw the regions on a SPHERE, then cut them out. However, it is difficult to produce two identical regions in this manner. Thompson (1996) gives mathematical expressions giving baseball cover curves both in the plane and in 3-D. J. H. Conway has humorously proposed the following "baseball curve conjecture:" no two definitions of "the" baseball curve will give the same answer unless their equivalence was obvious from the start.

see also BASEBALL, HOME PLATE, TENNIS BALL THEOREM, YIN-YANG

References

Thompson, R. B. "Designing a Baseball Cover. 1860's: Patience, Trial, and Error. 1990's: Geometry, Calculus, and Computation." `http://www.mathsoft.com/asolve/baseball/baseball.html`. Rev. March 5, 1996.

Basin of Attraction

The set of points in the space of system variables such that initial conditions chosen in this set dynamically evolve to a particular ATTRACTOR.

see also WADA BASIN

Basis

A (vector) basis is any SET of n LINEARLY INDEPENDENT VECTORS capable of generating an n-dimensional SUBSPACE of $\mathbb{R}^n$. Given a HYPERPLANE defined by

$$x_1 + x_2 + x_3 + x_4 + x_5 = 0,$$

a basis is found by solving for x_1 in terms of x_2, x_3, x_4, and x_5. Carrying out this procedure,

$$x_1 = -x_2 - x_3 - x_4 - x_5,$$

so

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix},$$

and the above Vector form an (unnormalized) Basis. Given a Matrix A with an orthonormal basis, the Matrix corresponding to a new basis, expressed in terms of the original $\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_n$ is

$$\mathsf{A}' = [\,\mathsf{A}\hat{\mathbf{x}}_1 \quad \ldots \quad \mathsf{A}\hat{\mathbf{x}}_n\,].$$

see also Bilinear Basis, Modular System Basis, Orthonormal Basis, Topological Basis

Basis Theorem

see Hilbert Basis Theorem

Basler Problem

The problem of analytically finding the value of $\zeta(2)$, where ζ is the Riemann Zeta Function.

References

Castellanos, D. "The Ubiquitous Pi. Part I." *Math. Mag.* **61**, 67–98, 1988.

Basset Function

see Modified Bessel Function of the Second Kind

Batch

A set of values of similar meaning obtained in any manner.

References

Tukey, J. W. *Explanatory Data Analysis.* Reading, MA: Addison-Wesley, p. 667, 1977.

Bateman Function

$$k_\nu(x) \equiv \frac{e^{-x}}{\Gamma(1+\frac{1}{2}\nu)} U(-\tfrac{1}{2}\nu, 0, 2x)$$

for $x > 0$, where U is a Confluent Hypergeometric Function of the Second Kind.

see also Confluent Hypergeometric Differential Equation, Hypergeometric Function

Batrachion

A class of Curve defined at Integer values which hops from one value to another. Their name derives from the word *batrachion*, which means "frog-like." Many batrachions are Fractal. Examples include the Blancmange Function, Hofstadter-Conway $10,000 Sequence, Hofstadter's Q-Sequence, and Mallow's Sequence.

References

Pickover, C. A. "The Crying of Fractal Batrachion 1,489." Ch. 25 in *Keys to Infinity.* New York: W. H. Freeman, pp. 183–191, 1995.

Bauer's Identical Congruence

Let $t(m)$ denote the set of the $\phi(m)$ numbers less than and Relatively Prime to m, where $\phi(n)$ is the Totient Function. Define

$$f_m(x) \equiv \prod_{t(m)} (x - t). \tag{1}$$

A theorem of Lagrange states that

$$f_m(x) \equiv x^{\phi(m)} - 1 \pmod{m}. \tag{2}$$

This can be generalized as follows. Let p be an Odd Prime Divisor of m and p^a the highest Power which divides m, then

$$f_m(x) \equiv (x^{p-1} - 1)^{\phi(m)/(p-1)} \pmod{p^a} \tag{3}$$

and, in particular,

$$f_{p^a}(x) \equiv (x^{p-1} - 1)^{p^{a-1}} \pmod{p^a}. \tag{4}$$

Furthermore, if $m > 2$ is Even and 2^a is the highest Power of 2 that divides m, then

$$f_m(x) \equiv (x^2 - 1)^{\phi(m)/2} \pmod{2^a} \tag{5}$$

and, in particular,

$$f_{2^a}(x) \equiv (x^2 - 1)^{2^{a-2}} \pmod{2^a}. \tag{6}$$

see also Leudesdorf Theorem

References

Hardy, G. H. and Wright, E. M. "Bauer's Identical Congruence." §8.5 in *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, pp. 98–100, 1979.

Bauer's Theorem

see Bauer's Identical Congruence

Bauspiel

A construction for the Rhombic Dodecahedron.

References

Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, pp. 26 and 50, 1973.

Bayes' Formula

Let A and B_j be Sets. Conditional Probability requires that

$$P(A \cap B_j) = P(A)P(B_j|A), \tag{1}$$

where $\cap$ denotes Intersection ("and"), and also that

$$P(A \cap B_j) = P(B_j \cap A) = P(B_j)P(A|B_j) \tag{2}$$

and

$$P(B_j \cap A) = P(B_j)P(A|B_j). \tag{3}$$

Since (2) and (3) must be equal,

$$P(A \cap B_j) = P(B_j \cap A). \tag{4}$$

From (2) and (3),

$$P(A \cap B_j) = P(B_j)P(A|B_j). \tag{5}$$

Equating (5) with (2) gives

$$P(A)P(B_j|A) = P(B_j)P(A|B_j), \tag{6}$$

so

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{P(A)}. \tag{7}$$

Now, let

$$S \equiv \bigcup_{i=1}^{N} A_i, \tag{8}$$

so A_i is an event is S and $A_i \cap A_j = \varnothing$ for $i \neq j$, then

$$A = A \cap S = A \cap \left(\bigcup_{i=1}^{N} A_i\right) = \bigcup_{i=1}^{N}(A \cap A_i) \tag{9}$$

$$P(A) = P\left(\bigcup_{i=1}^{N}(A \cap A_i)\right) = \sum_{i=1}^{N} P(A \cap A_i). \tag{10}$$

From (5), this becomes

$$P(A) = \sum_{i=1}^{N} P(A_i)P(E|A_i), \tag{11}$$

so

$$P(A_i|A) = \frac{P(A_i)P(A|A_i)}{\sum_{j=1}^{N} P(A_j)P(A|A_j)}. \tag{12}$$

see also CONDITIONAL PROBABILITY, INDEPENDENT STATISTICS

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 810, 1992.

Bayes' Theorem

see BAYES' FORMULA

Bayesian Analysis

A statistical procedure which endeavors to estimate parameters of an underlying distribution based on the observed distribution. Begin with a "PRIOR DISTRIBUTION" which may be based on anything, including an assessment of the relative likelihoods of parameters or the results of non-Bayesian observations. In practice, it is common to assume a UNIFORM DISTRIBUTION over the appropriate range of values for the PRIOR DISTRIBUTION.

Given the PRIOR DISTRIBUTION, collect data to obtain the observed distribution. Then calculate the LIKELIHOOD of the observed distribution as a function of parameter values, multiply this likelihood function by the PRIOR DISTRIBUTION, and normalize to obtain a unit probability over all possible values. This is called the POSTERIOR DISTRIBUTION. The MODE of the distribution is then the parameter estimate, and "probability intervals" (the Bayesian analog of CONFIDENCE INTERVALS) can be calculated using the standard procedure. Bayesian analysis is somewhat controversial because the validity of the result depends on how valid the PRIOR DISTRIBUTION is, and this cannot be assessed statistically.

see also MAXIMUM LIKELIHOOD, PRIOR DISTRIBUTION, UNIFORM DISTRIBUTION

References

Hoel, P. G.; Port, S. C.; and Stone, C. J. *Introduction to Statistical Theory.* New York: Houghton Mifflin, pp. 36–42, 1971.
Iversen, G. R. *Bayesian Statistical Inference.* Thousand Oaks, CA: Sage Pub., 1984.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 799–806, 1992.
Sivia, D. S. *Data Analysis: A Bayesian Tutorial.* New York: Oxford University Press, 1996.

Bays' Shuffle

A shuffling algorithm used in a class of RANDOM NUMBER generators.

References

Knuth, D. E. §3.2 and 3.3 in *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, 1981.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 270–271, 1992.

Beam Detector

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

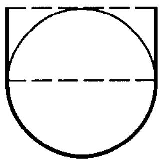

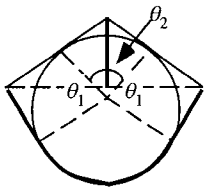

A "beam detector" for a given curve C is defined as a curve (or set of curves) through which every Line tangent to or intersecting C passes. The shortest 1-arc beam detector, illustrated in the upper left figure, has length $L_1 = \pi + 2$. The shortest known 2-arc beam detector, illustrated in the right figure, has angles

$$\theta_1 \approx 1.286 \text{ rad} \tag{1}$$

$$\theta_2 \approx 1.191 \text{ rad}, \tag{2}$$

given by solving the simultaneous equations

$$2\cos\theta_1 - \sin(\tfrac{1}{2}\theta_2) = 0 \tag{3}$$

$$\tan(\tfrac{1}{2}\theta_1)\cos(\tfrac{1}{2}\theta_2) + \sin(\tfrac{1}{2}\theta_2)[\sec^2(\tfrac{1}{2}\theta_2) + 1] = 2. \tag{4}$$

The corresponding length is

$$\begin{aligned} L_2 &= 2\pi - 2\theta_1 - \theta_2 + 2\tan(\tfrac{1}{2}\theta_1) + \sec(\tfrac{1}{2}\theta_2) \\ &- \cos(\tfrac{1}{2}\theta_2) + \tan(\tfrac{1}{2}\theta_1)\sin(\tfrac{1}{2}\theta_2) = 4.8189264563\ldots. \end{aligned} \tag{5}$$

A more complicated expression gives the shortest known 3-arc length $L_3 = 4.799891547\ldots$. Finch defines

$$L = \inf_{n\geq 1} L_n \tag{6}$$

as the beam detection constant, or the Trench Diggers' Constant. It is known that $L \geq \pi$.

References

Croft, H. T.; Falconer, K. J.; and Guy, R. K. §A30 in *Unsolved Problems in Geometry.* New York: Springer-Verlag, 1991.

Faber, V.; Mycielski, J.; and Pedersen, P. "On the Shortest Curve which Meets All Lines which Meet a Circle." *Ann. Polon. Math.* **44**, 249–266, 1984.

Faber, V. and Mycielski, J. "The Shortest Curve that Meets All Lines that Meet a Convex Body." *Amer. Math. Monthly* **93**, 796–801, 1986.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/beam/beam.html`.

Makai, E. "On a Dual of Tarski's Plank Problem." In *Diskrete Geometrie.* 2 Kolloq., Inst. Math. Univ. Salzburg, 127–132, 1980.

Stewart, I. "The Great Drain Robbery." *Sci. Amer.*, 206–207, 106, and 125, Sept. 1995, Dec. 1995, and Feb. 1996.

Bean Curve

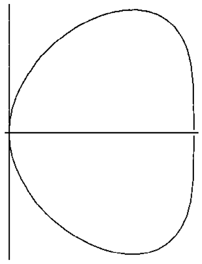

The Plane Curve given by the Cartesian equation

$$x^4 + x^2y^2 + y^4 = x(x^2 + y^2).$$

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., 1989.

Beast Number

The occult "number of the beast" associated in the Bible with the Antichrist. It has figured in many numerological studies. It is mentioned in Revelation 13:13: "Here is wisdom. Let him that hath understanding count the number of the beast: for it is the number of a man; and his number is 666."

The beast number has several interesting properties which numerologists may find particularly interesting (Keith 1982–83). In particular, the beast number is equal to the sum of the squares of the first 7 Primes

$$2^2 + 3^2 + 5^2 + 7^2 + 11^2 + 13^2 + 17^2 = 666, \tag{1}$$

satisfies the identity

$$\phi(666) = 6 \cdot 6 \cdot 6, \tag{2}$$

where ϕ is the Totient Function, as well as the sum

$$\sum_{i=1}^{6\cdot 6} i = 666. \tag{3}$$

The number 666 is a sum and difference of the first three 6th Powers,

$$666 = 1^6 - 2^6 + 3^6 \tag{4}$$

(Keith). Another curious identity is that there are exactly two ways to insert "+" signs into the sequence 123456789 to make the sum 666, and exactly one way for the sequence 987654321,

$$666 = 1 + 2 + 3 + 4 + 567 + 89 = 123 + 456 + 78 + 9 \tag{5}$$

$$666 = 9 + 87 + 6 + 543 + 21 \tag{6}$$

(Keith). 666 is a Repdigit, and is also a Triangular Number

$$T_{6\cdot 6} = T_{36} = 666. \tag{7}$$

In fact, it is the largest Repdigit Triangular Number (Bellew and Weger 1975–76). 666 is also a Smith Number. The first 144 Digits of $\pi - 3$, where π is Pi, add to 666. In addition $144 = (6+6) \times (6+6)$ (Blatner 1997).

A number of the form 2^i which contains the digits of the beast number "666" is called an Apocalyptic Number, and a number having 666 digits is called an Apocalypse Number.

see also APOCALYPSE NUMBER, APOCALYPTIC NUMBER, BIMONSTER, MONSTER GROUP

References

Bellew, D. W. and Weger, R. C. "Repdigit Triangular Numbers." *J. Recr. Math.* **8**, 96–97, 1975–76.

Blatner, D. *The Joy of Pi.* New York: Walker, back jacket, 1997.

Castellanos, D. "The Ubiquitous π." *Math. Mag.* **61**, 153–154, 1988.

Hardy, G. H. *A Mathematician's Apology, reprinted with a foreword by C. P. Snow.* New York: Cambridge University Press, p. 96, 1993.

Keith, M. "The Number of the Beast." `http://users.aol.com/s6sj7gt/mike666.htm`.

Keith, M. "The Number 666." *J. Recr. Math.* **15**, 85–87, 1982–1983.

Beatty Sequence

The Beatty sequence is a SPECTRUM SEQUENCE with an IRRATIONAL base. In other words, the Beatty sequence corresponding to an IRRATIONAL NUMBER θ is given by $\lfloor\theta\rfloor$, $\lfloor 2\theta\rfloor$, $\lfloor 3\theta\rfloor$, ..., where $\lfloor x\rfloor$ is the FLOOR FUNCTION. If α and β are POSITIVE IRRATIONAL NUMBERS such that

$$\frac{1}{\alpha}+\frac{1}{\beta}=1,$$

then the Beatty sequences $\lfloor\alpha\rfloor$, $\lfloor 2\alpha\rfloor$, ... and $\lfloor\beta\rfloor$, $\lfloor 2\beta\rfloor$, ... together contain all the POSITIVE INTEGERS without repetition.

References

Gardner, M. *Penrose Tiles and Trapdoor Ciphers... and the Return of Dr. Matrix, reissue ed.* New York: W. H. Freeman, p. 21, 1989.

Graham, R. L.; Lin, S.; and Lin, C.-S. "Spectra of Numbers." *Math. Mag.* **51**, 174–176, 1978.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 227, 1994.

Sloane, N. J. A. *A Handbook of Integer Sequences.* Boston, MA: Academic Press, pp. 29–30, 1973.

Beauzamy and Dégot's Identity

For P, Q, R, and S POLYNOMIALS in n variables

$$[P\cdot Q,R\cdot S]=\sum_{i_1,\ldots,i_n\geq 0}\frac{A}{i_1!\cdots i_n!},$$

where

$$A\equiv[R^{(i_1,\ldots,i_n)}(D_1,\ldots,D_n)Q(x_1,\ldots,x_n)\\ \times P^{(i_1,\ldots,i_n)}(D_1,\ldots,D_n)S(x_1,\ldots,x_n)]$$

$D_i=\partial/\partial x_i$ is the DIFFERENTIAL OPERATOR, $[X,Y]$ is the BOMBIERI INNER PRODUCT, and

$$P^{(i_1,\ldots,i_n)}=D_1^{i_1}\cdots D_n^{i_n}P.$$

see also REZNIK'S IDENTITY

Bee

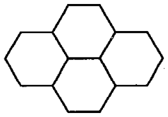

A 4-POLYHEX.

References

Gardner, M. *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, p. 147, 1978.

Behrens-Fisher Test

see FISHER-BEHRENS PROBLEM

Behrmann Cylindrical Equal-Area Projection

A CYLINDRICAL AREA-PRESERVING projection which uses 30° N as the no-distortion parallel.

References

Dana, P. H. "Map Projections." `http://www.utexas.edu/depts/grg/gcraft/notes/mapproj/mapproj.html`.

Bei

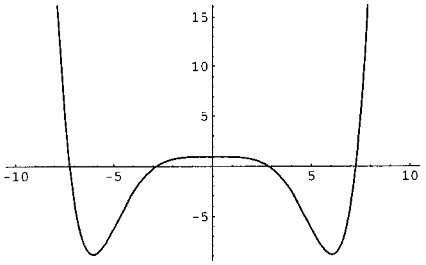

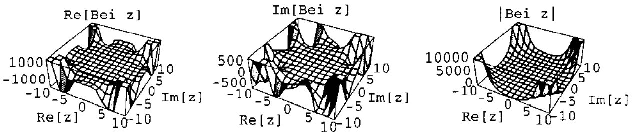

The IMAGINARY PART of

$$J_\nu(xe^{3\pi i/4})=\mathrm{ber}_\nu(x)+i\,\mathrm{bei}_\nu(x). \tag{1}$$

The special case $\nu=0$ gives

$$J_0(i\sqrt{i}\,x)\equiv\mathrm{ber}(x)+i\,\mathrm{bei}(x), \tag{2}$$

where $J_0(z)$ is the zeroth order BESSEL FUNCTION OF THE FIRST KIND.

$$\mathrm{bei}(x)\equiv\sum_{n=0}^{\infty}\frac{(-1)^n\left(\frac{x}{2}\right)^{4n}}{[(2n)!]^2}. \tag{3}$$

see also BER, BESSEL FUNCTION, KEI, KELVIN FUNCTIONS, KER

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Kelvin Functions." §9.9 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 379–381, 1972.

Spanier, J. and Oldham, K. B. "The Kelvin Functions." Ch. 55 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 543–554, 1987.

Bell Curve

see GAUSSIAN DISTRIBUTION, NORMAL DISTRIBUTION

Bell Number

The number of ways a SET of n elements can be PARTITIONED into nonempty SUBSETS is called a BELL NUMBER and is denoted B_n. For example, there are five ways the numbers {1, 2, 3} can be partitioned: {{1}, {2}, {3}}, {{1, 2}, {3}}, {{1, 3}, {2}}, {{1}, {2, 3}}, and {{1, 2, 3}}, so $B_3 = 5$. $B_0 = 1$ and the first few Bell numbers for $n = 1, 2, \ldots$ are 1, 2, 5, 15, 52, 203, 877, 4140, 21147, 115975, ... (Sloane's A000110). Bell numbers are closely related to CATALAN NUMBERS.

The diagram below shows the constructions giving $B_3 = 5$ and $B_4 = 15$, with line segments representing elements in the same SUBSET and dots representing subsets containing a single element (Dickau).

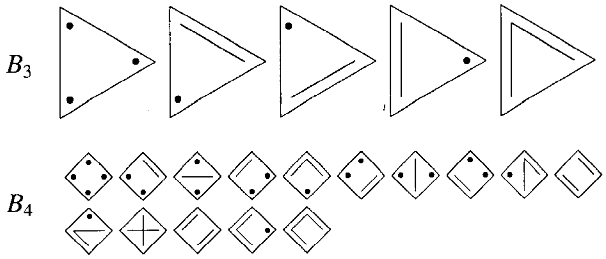

The INTEGERS B_n can be defined by the sum

$$B_n = \sum_{k=1}^{n} \left\{ {n \atop k} \right\}, \tag{1}$$

where $S_n^{(k)} = \left\{ {n \atop k} \right\}$ is a STIRLING NUMBER OF THE SECOND KIND, or by the generating function

$$e^{e^x - 1} = \sum_{n=0}^{\infty} \frac{B_n}{n!} x^n. \tag{2}$$

The Bell numbers can also be generated using the BELL TRIANGLE, using the RECURRENCE RELATION

$$B_{n+1} = \sum_{k=0}^{n} B_k \binom{n}{k}, \tag{3}$$

where $\binom{a}{b}$ is a BINOMIAL COEFFICIENT, or using the formula of Comtet (1974)

$$B_n = \left\lceil e^{-1} \sum_{m=1}^{2n} \frac{m^n}{m!} \right\rceil, \tag{4}$$

where $\lceil x \rceil$ denotes the CEILING FUNCTION.

The Bell number B_n is also equal to $\phi_n(1)$, where $\phi_n(x)$ is a BELL POLYNOMIAL. DOBIŃSKI'S FORMULA gives the nth Bell number

$$B_n = \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^n}{k!}. \tag{5}$$

Lovász (1993) showed that this formula gives the asymptotic limit

$$B_n \sim n^{-1/2} [\lambda(n)]^{n+1/2} e^{\lambda(n)-n-1}, \tag{6}$$

where $\lambda(n)$ is defined implicitly by the equation

$$\lambda(n) \log[\lambda(n)] = n. \tag{7}$$

A variation of DOBIŃSKI'S FORMULA gives

$$B_k = \sum_{m=1}^{n} \frac{m^k}{m!} \sum_{s=0}^{n-m} \frac{(-1)^s}{s} \tag{8}$$

for $1 \leq k \leq n$ (Pitman 1997). de Bruijn (1958) gave the asymptotic formula

$$\begin{aligned} \frac{\ln B_n}{n} &= \ln n - \ln \ln n - 1 + \frac{\ln \ln n}{\ln n} + \frac{1}{lnn} \\ &\quad + \frac{1}{2} \left(\frac{\ln \ln n}{\ln n} \right)^2 + \mathcal{O} \left[\frac{\ln \ln n}{(\ln n)^2} \right]. \end{aligned} \tag{9}$$

TOUCHARD'S CONGRUENCE states

$$B_{p+k} \equiv B_k + B_{k+1} \pmod{p}, \tag{10}$$

when p is PRIME. The only PRIME Bell numbers for $n \leq 1000$ are B_2, B_3, B_7, B_{13}, B_{42}, and B_{55}. The Bell numbers also have the curious property that

$$\begin{vmatrix} B_0 & B_1 & B_2 & \cdots & B_n \\ B_1 & B_2 & B_3 & \cdots & B_{n+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ B_n & B_{n+1} & B_{n+2} & \cdots & B_{2n} \end{vmatrix} = \prod_{i=1}^{n} n! \tag{11}$$

(Lenard 1986).

see also BELL POLYNOMIAL, BELL TRIANGLE, DOBIŃSKI'S FORMULA, STIRLING NUMBER OF THE SECOND KIND, TOUCHARD'S CONGRUENCE

References

Bell, E. T. "Exponential Numbers." *Amer. Math. Monthly* **41**, 411–419, 1934.

Comtet, L. *Advanced Combinatorics.* Dordrecht, Netherlands: Reidel, 1974.

Conway, J. H. and Guy, R. K. In *The Book of Numbers.* New York: Springer-Verlag, pp. 91–94, 1996.

de Bruijn, N. G. *Asymptotic Methods in Analysis.* New York: Dover, pp. 102–109, 1958.

Dickau, R. M. "Bell Number Diagrams." `http:// forum . swarthmore.edu/advanced/robertd/bell.html`.

Gardner, M. "The Tinkly Temple Bells." Ch. 2 in *Fractal Music, Hypercards, and More Mathematical Recreations from Scientific American Magazine.* New York: W. H. Freeman, 1992.

Gould, H. W. *Bell & Catalan Numbers: Research Bibliography of Two Special Number Sequences, 6th ed.* Morgantown, WV: Math Monongliae, 1985.

Lenard, A. In *Fractal Music, Hypercards, and More Mathematical Recreations from Scientific American Magazine.* (M. Gardner). New York: W. H. Freeman, pp. 35–36, 1992.

Levine, J. and Dalton, R. E. "Minimum Periods, Modulo p, of First Order Bell Exponential Integrals." *Math. Comput.* **16**, 416–423, 1962.

Lovász, L. *Combinatorial Problems and Exercises, 2nd ed.* Amsterdam, Netherlands: North-Holland, 1993.

Pitman, J. "Some Probabilistic Aspects of Set Partitions." *Amer. Math. Monthly* **104**, 201–209, 1997.

Rota, G.-C. "The Number of Partitions of a Set." *Amer. Math. Monthly* **71**, 498–504, 1964.

Sloane, N. J. A. Sequence A000110/M1484 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Bell Polynomial

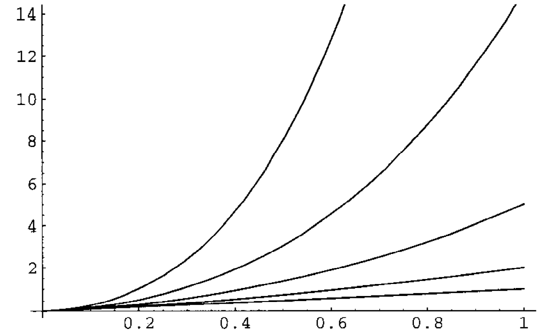

Two different GENERATING FUNCTIONS for the Bell polynomials for $n > 0$ are given by

$$\phi_n(x) \equiv e^{-x} \sum_{k=1}^{\infty} \frac{k^{n-1} x^k}{(k-1)!}$$

or

$$\phi_n(x) = x \sum_{k=1}^{n-1} \binom{n-1}{k-1} \phi_{k-1}(x),$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT.

The Bell polynomials are defined such that $\phi_n(1) = B_n$, where B_n is a BELL NUMBER. The first few Bell polynomials are

$$\begin{aligned}
\phi_0(x) &= 1 \\
\phi_1(x) &= x \\
\phi_2(x) &= x + x^2 \\
\phi_3(x) &= x + 3x^2 + x^3 \\
\phi_4(x) &= x + 7x^2 + 6x^3 + x^4 \\
\phi_5(x) &= x + 15x^2 + 25x^3 + 10x^4 + x^5 \\
\phi_6(x) &= x + 31x^2 + 90x^3 + 65x^4 + 15x^5 + x^6.
\end{aligned}$$

see also BELL NUMBER

References

Bell, E. T. "Exponential Polynomials." *Ann. Math.* **35**, 258–277, 1934.

Bell Triangle

$$\begin{array}{c}
1\ \ 2\ \ 5\ \ 15\ \ 52\ \ 203\ \ 877 \cdots \\
1\ \ 3\ \ 10\ \ 37\ \ 151\ \ 674\ \ddots \\
2\ \ 7\ \ 27\ \ 114\ \ 523\ \ddots \\
5\ \ 20\ \ 87\ \ 409\ \ddots \\
15\ \ 67\ \ 322\ \ddots \\
52\ \ 255\ \ddots \\
203\ \ddots \\
\ddots
\end{array}$$

A triangle of numbers which allow the BELL NUMBERS to be computed using the RECURRENCE RELATION

$$B_{n+1} = \sum_{k=0}^{n} B_k \binom{n}{k}.$$

see also BELL NUMBER, CLARK'S TRIANGLE, LEIBNIZ HARMONIC TRIANGLE, NUMBER TRIANGLE, PASCAL'S TRIANGLE, SEIDEL-ENTRINGER-ARNOLD TRIANGLE

Bellows Conjecture

see FLEXIBLE POLYHEDRON

Beltrami Differential Equation

For a measurable function μ, the Beltrami differential equation is given by

$$f_{z^*} = \mu f_z,$$

where f_z is a PARTIAL DERIVATIVE and z^* denotes the COMPLEX CONJUGATE of z.

see also QUASICONFORMAL MAP

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1087, 1980.

Beltrami Field

A VECTOR FIELD **u** satisfying the vector identity

$$\mathbf{u} \times (\nabla \times \mathbf{u}) = \mathbf{0}$$

where $\mathbf{A} \times \mathbf{B}$ is the CROSS PRODUCT and $\nabla \times \mathbf{A}$ is the CURL is said to be a Beltrami field.

see also DIVERGENCELESS FIELD, IRROTATIONAL FIELD, SOLENOIDAL FIELD

Beltrami Identity

An identity in CALCULUS OF VARIATIONS discovered in 1868 by Beltrami. The EULER-LAGRANGE DIFFERENTIAL EQUATION is

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y_x}\right) = 0. \tag{1}$$

Now, examine the DERIVATIVE of x

$$\frac{df}{dx} = \frac{\partial f}{\partial y} y_x + \frac{\partial f}{\partial y_x} y_{xx} + \frac{\partial f}{\partial x}. \tag{2}$$

Solving for the $\partial f / \partial y$ term gives

$$\frac{\partial f}{\partial y} y_x = \frac{df}{dx} - \frac{\partial f}{\partial y_x} y_{xx} - \frac{\partial f}{\partial x}. \tag{3}$$

Now, multiplying (1) by y_x gives

$$y_x \frac{\partial f}{\partial y} - y_x \frac{d}{dx}\left(\frac{\partial f}{\partial y_x}\right) = 0. \tag{4}$$

Substituting (3) into (4) then gives

$$\frac{df}{dx} - \frac{\partial f}{\partial y_x} y_{xx} - \frac{\partial f}{\partial x} - y_x \frac{d}{dx}\left(\frac{\partial f}{\partial y_x}\right) = 0 \tag{5}$$

$$-\frac{\partial f}{\partial x} + \frac{d}{dx}\left(f - y_x \frac{\partial f}{\partial y_x}\right) = 0. \tag{6}$$

This form is especially useful if $f_x = 0$, since in that case

$$\frac{d}{dx}\left(f - y_x \frac{\partial f}{\partial y_x}\right) = 0, \tag{7}$$

which immediately gives

$$f - y_x \frac{\partial f}{\partial y_x} = C, \tag{8}$$

where C is a constant of integration.

The Beltrami identity greatly simplifies the solution for the minimal AREA SURFACE OF REVOLUTION about a given axis between two specified points. It also allows straightforward solution of the BRACHISTOCHRONE PROBLEM.

see also BRACHISTOCHRONE PROBLEM, CALCULUS OF VARIATIONS, EULER-LAGRANGE DIFFERENTIAL EQUATION, SURFACE OF REVOLUTION

Bend (Curvature)

Given four mutually tangent circles, their bends are defined as the signed CURVATURES of the CIRCLES. If the contacts are all external, the signs are all taken as POSITIVE, whereas if one circle surrounds the other three, the sign of this circle is taken as NEGATIVE (Coxeter 1969).

see also CURVATURE, DESCARTES CIRCLE THEOREM, SODDY CIRCLES

References

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 13–14, 1969.

Bend (Knot)

A KNOT used to join the ends of two ropes together to form a longer length.

References

Owen, P. *Knots.* Philadelphia, PA: Courage, p. 49, 1993.

Benford's Law

Also called the FIRST DIGIT LAW, FIRST DIGIT PHENOMENON, or LEADING DIGIT PHENOMENON. In listings, tables of statistics, etc., the DIGIT 1 tends to occur with PROBABILITY ~ 30%, much greater than the expected 10%. This can be observed, for instance, by examining tables of LOGARITHMS and noting that the first pages are much more worn and smudged than later pages. The table below, taken from Benford (1938), shows the distribution of first digits taken from several disparate sources. Of the 54 million real constants in Plouffe's "Inverse Symbolic Calculator" database, 30% begin with the DIGIT 1.

Title	First Digit									#
	1	2	3	4	5	6	7	8	9	
Rivers, Area	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3259
Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
Newspapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
Specific Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1389
Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
H.P. Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	690
Mol. Wgt.	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2	1800
Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	159
Atomic Wgt.	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	91
n^{-1}, $\sqrt{n}$	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	5000
Design	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	560
Reader's Dig.	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	308
Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	741
X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	707
Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1458
Blackbody	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	1165
Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
n^1, $n^2 \cdots n!$	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	900
Death Rate	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	418
Average	30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	1011
Prob. Error	0.8	0.4	0.4	0.3	0.2	0.2	0.2	0.2	0.3	

In fact, the first SIGNIFICANT DIGIT seems to follow a LOGARITHMIC DISTRIBUTION, with

$$P(n) \approx \log(n+1) - \log n$$

for $n = 1, \ldots, 9$. One explanation uses CENTRAL LIMIT-like theorems for the MANTISSAS of random variables under MULTIPLICATION. As the number of variables increases, the density function approaches that of a LOGARITHMIC DISTRIBUTION.

References

Benford, F. "The Law of Anomalous Numbers." *Proc. Amer. Phil. Soc.* **78**, 551–572, 1938.

Boyle, J. "An Application of Fourier Series to the Most Significant Digit Problem." *Amer. Math. Monthly* **101**, 879–886, 1994.

Hill, T. P. "Base-Invariance Implies Benford's Law." *Proc. Amer. Math. Soc.* **12**, 887–895, 1995.

Hill, T. P. "The Significant-Digit Phenomenon." *Amer. Math. Monthly* **102**, 322–327, 1995.

Hill, T. P. "A Statistical Derivation of the Significant-Digit Law." *Stat. Sci.* **10**, 354–363, 1996.

Hill, T. P. "The First Digit Phenomenon." *Amer. Sci.* **86**, 358–363, 1998.

Ley, E. "On the Peculiar Distribution of the U.S. Stock Indices Digits." *Amer. Stat.* **50**, 311–313, 1996.

Newcomb, S. "Note on the Frequency of the Use of Digits in Natural Numbers." *Amer. J. Math.* **4**, 39–40, 1881.

Nigrini, M. "A Taxpayer Compliance Application of Benford's Law." *J. Amer. Tax. Assoc.* **18**, 72–91, 1996.

Plouffe, S. "Inverse Symbolic Calculator." `http://www.cecm.sfu.ca/projects/ISC/`.

Raimi, R. A. "The Peculiar Distribution of First Digits." *Sci. Amer.* **221**, 109–119, Dec. 1969.

Raimi, R. A. "The First Digit Phenomenon." *Amer. Math. Monthly* **83**, 521–538, 1976.

Benham's Wheel

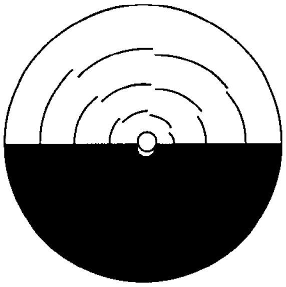

An optical ILLUSION consisting of a spinnable top marked in black with the pattern shown above. When the wheel is spun (especially slowly), the black broken lines appear as green, blue, and red colored bands!

References

Cohen, J. and Gordon, D. A. "The Prevost-Fechner-Benham Subjective Colors." *Psycholog. Bull.* **46**, 97–136, 1949.

Festinger, L.; Allyn, M. R.; and White, C. W. "The Perception of Color with Achromatic Stimulation." *Vision Res.* **11**, 591–612, 1971.

Fineman, M. *The Nature of Visual Illusion.* New York: Dover, pp. 148–151, 1996.

Trolland, T. L. "The Enigma of Color Vision." *Amer. J. Physiology* **2**, 23–48, 1921.

Bennequin's Conjecture

A BRAID with M strands and R components with P positive crossings and N negative crossings satisfies

$$|P - N| \leq 2U + M - R \leq P + N,$$

where U is the UNKNOTTING NUMBER. While the second part of the INEQUALITY was already known to be true (Boileau and Weber, 1983, 1984) at the time the conjecture was proposed, the proof of the entire conjecture was completed using results of Kronheimer and Mrowka on MILNOR'S CONJECTURE (and, independently, using MENASCO'S THEOREM).

see also BRAID, MENASCO'S THEOREM, MILNOR'S CONJECTURE, UNKNOTTING NUMBER

References

Bennequin, D. "L'instanton gordien (d'après P. B. Kronheimer et T. S. Mrowka)." *Astérisque* **216**, 233–277, 1993.

Birman, J. S. and Menasco, W. W. "Studying Links via Closed Braids. II. On a Theorem of Bennequin." *Topology Appl.* **40**, 71–82, 1991.

Boileau, M. and Weber, C. "Le problème de J. Milnor sur le nombre gordien des nœuds algébriques." *Enseign. Math.* **30**, 173–222, 1984.

Boileau, M. and Weber, C. "Le problème de J. Milnor sur le nombre gordien des nœuds algébriques." In *Knots, Braids and Singularities (Plans-sur-Bex, 1982).* Geneva, Switzerland: Monograph. Enseign. Math. Vol. 31, pp. 49–98, 1983.

Cipra, B. *What's Happening in the Mathematical Sciences, Vol. 2.* Providence, RI: Amer. Math. Soc., pp. 8–13, 1994.

Kronheimer, P. B. "The Genus-Minimizing Property of Algebraic Curves." *Bull. Amer. Math. Soc.* **29**, 63–69, 1993.

Kronheimer, P. B. and Mrowka, T. S. "Gauge Theory for Embedded Surfaces. I." *Topology* **32**, 773–826, 1993.

Kronheimer, P. B. and Mrowka, T. S. "Recurrence Relations and Asymptotics for Four-Manifold Invariants." *Bull. Amer. Math. Soc.* **30**, 215–221, 1994.

Menasco, W. W. "The Bennequin-Milnor Unknotting Conjectures." *C. R. Acad. Sci. Paris Sér. I Math.* **318**, 831–836, 1994.

Benson's Formula

An equation for a LATTICE SUM with $n = 3$

$$-b_3(1) = \sum_{i,\,j,\,k=-\infty}^{\infty}{}' \frac{(-1)^{i+j+k+1}}{\sqrt{i^2+j^2+k^2}}$$

$$= 12\pi \sum_{m,\,n=1,\,3,\,\ldots}^{\infty} \operatorname{sech}^2\left(\tfrac{1}{2}\pi\sqrt{m^2+n^2}\,\right).$$

Here, the prime denotes that summation over $(0, 0, 0)$ is excluded. The sum is numerically equal to $-1.74756\ldots$, a value known as "the" MADELUNG CONSTANT.

see also MADELUNG CONSTANTS

References

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, p. 301, 1987.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/mdlung/mdlung.html`.

Ber

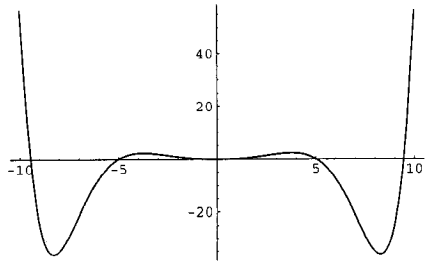

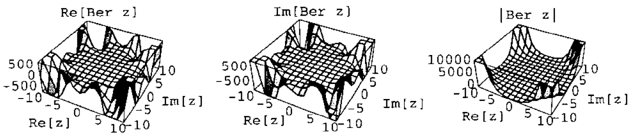

The REAL PART of

$$J_\nu(xe^{3\pi i/4}) = \operatorname{ber}_\nu(x) + i\operatorname{bei}_\nu(x). \qquad (1)$$

The special case $\nu = 0$ gives

$$J_0(i\sqrt{i}\,x) \equiv \operatorname{ber}(x) + i\operatorname{bei}(x), \qquad (2)$$

where J_0 is the zeroth order BESSEL FUNCTION OF THE FIRST KIND.

$$\operatorname{ber}(x) \equiv \sum_{n=0}^{\infty} \frac{(-1)^n \left(\frac{x}{2}\right)^{2+4n}}{[(2n+1)!]^2}. \qquad (3)$$

see also BEI, BESSEL FUNCTION, KEI, KELVIN FUNCTIONS, KER

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Kelvin Functions." §9.9 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 379–381, 1972.

Spanier, J. and Oldham, K. B. "The Kelvin Functions." Ch. 55 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 543–554, 1987.

Beraha Constants

The nth Beraha constant is given by

$$\mathrm{Be}_n \equiv 2 + 2\cos\left(\frac{2\pi}{n}\right).$$

The first few are

$$\begin{aligned} \mathrm{Be}_1 &= 4 \\ \mathrm{Be}_2 &= 0 \\ \mathrm{Be}_3 &= 1 \\ \mathrm{Be}_4 &= 2 \\ \mathrm{Be}_5 &= \tfrac{1}{2}(3+\sqrt{5}\,) \approx 2.618 \\ \mathrm{Be}_6 &= 3 \\ \mathrm{Be}_7 &= 2 + 2\cos(\tfrac{2}{7}\pi) \approx 3.247\ldots. \end{aligned}$$

They appear to be ROOTS of the CHROMATIC POLYNOMIALS of planar triangular GRAPHS. Be_4 is $\phi + 1$, where ϕ is the GOLDEN RATIO, and Be_7 is the SILVER CONSTANT.

References

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 143, 1983.

Berger-Kazdan Comparison Theorem

Let M be a compact n-D MANIFOLD with INJECTIVITY radius $\operatorname{inj}(M)$. Then

$$\operatorname{Vol}(M) \geq \frac{c_n \operatorname{inj}(M)}{\pi},$$

with equality IFF M is ISOMETRIC to the standard round SPHERE S^n with RADIUS $\operatorname{inj}(M)$, where $c_n(r)$ is the VOLUME of the standard n-HYPERSPHERE of RADIUS r.

see also BLASCHKE CONJECTURE, HYPERSPHERE, INJECTIVE, ISOMETRY

References

Chavel, I. *Riemannian Geometry: A Modern Introduction.* New York: Cambridge University Press, 1994.

Bergman Kernel

A Bergman kernel is a function of a COMPLEX VARIABLE with the "reproducing kernel" property defined for any DOMAIN in which there exist NONZERO ANALYTIC FUNCTIONS of class $L_2(D)$ with respect to the LEBESGUE MEASURE dV.

References

Hazewinkel, M. (Managing Ed.). *Encyclopaedia of Mathematics: An Updated and Annotated Translation of the Soviet "Mathematical Encyclopaedia."* Dordrecht, Netherlands: Reidel, pp. 356–357, 1988.

Bergman Space

Let G be an open subset of the COMPLEX PLANE $\mathbb{C}$, and let $L_a^2(G)$ denote the collection of all ANALYTIC FUNCTIONS $f : G \to C$ whose MODULUS is square integrable with respect to AREA measure. Then $L_a^2(G)$, sometimes also denoted $A^2(G)$, is called the Bergman space for G. Thus, the Bergman space consists of all the ANALYTIC FUNCTIONS in $L^2(G)$. The Bergman space can also be generalized to $L_a^p(G)$, where $0 < p < \infty$.

Bernoulli Differential Equation

$$\frac{dy}{dx} + p(x)y = q(x)y^n. \qquad (1)$$

Let $v \equiv y^{1-n}$ for $n \neq 1$, then

$$\frac{dv}{dx} = (1-n)y^{-n}\frac{dy}{dx}. \qquad (2)$$

Rewriting (1) gives

$$y^{-n}\frac{dy}{dx} = q(x) - p(x)y^{1-n} = q(x) - vp(x). \qquad (3)$$

Plugging (3) into (2),

$$\frac{dv}{dx} = (1-n)[q(x) - vp(x)]. \qquad (4)$$

Now, this is a linear FIRST-ORDER ORDINARY DIFFERENTIAL EQUATION of the form

$$\frac{dv}{dx} + vP(x) = Q(x), \tag{5}$$

where $P(x) \equiv (1-n)p(x)$ and $Q(x) \equiv (1-n)q(x)$. It can therefore be solved analytically using an INTEGRATING FACTOR

$$\begin{aligned} v &= \frac{\int e^{\int P(x)\,dx} Q(x)\,dx + C}{e^{\int P(x)\,dx}} \\ &= \frac{(1-n)\int e^{(1-n)\int p(x)\,dx} q(x)\,dx + C}{e^{(1-n)\int p(x)\,dx}}, \end{aligned} \tag{6}$$

where C is a constant of integration. If $n = 1$, then equation (1) becomes

$$\frac{dy}{dx} = y(q-p) \tag{7}$$

$$\frac{dy}{y} = (q-p)\,dx \tag{8}$$

$$y = C_2 e^{\int [q(x)-p(x)]\,dx}. \tag{9}$$

The general solution is then, with C_1 and C_2 constants,

$$y = \begin{cases} \left[\dfrac{(1-n)\int e^{(1-n)\int p(x)\,dx} q(x)\,dx + C_1}{e^{(1-n)\int p(x)\,dx}}\right]^{1/(1-n)} \\ \quad \text{for } n \neq 1 \\ C_2 e^{\int [q(x)-p(x)]\,dx} \\ \quad \text{for } n = 1. \end{cases} \tag{10}$$

Bernoulli Distribution

A DISTRIBUTION given by

$$\begin{aligned} P(n) &= \begin{cases} q \equiv 1-p & \text{for } n = 0 \\ p & \text{for } n = 1 \end{cases} && (1) \\ &= p^n (1-p)^{1-n} \qquad \text{for } n = 0, 1. && (2) \end{aligned}$$

The distribution of heads and tails in COIN TOSSING is a Bernoulli distribution with $p = q = 1/2$. The GENERATING FUNCTION of the Bernoulli distribution is

$$M(t) = \langle e^{tn} \rangle = \sum_{n=0}^{1} e^{tn} p^n (1-p)^{1-n} = e^0(1-p) + e^t p, \tag{3}$$

so

$$\begin{aligned} M(t) &= (1-p) + pe^t && (4) \\ M'(t) &= pe^t && (5) \\ M''(t) &= pe^t && (6) \\ M^{(n)}(t) &= pe^t, && (7) \end{aligned}$$

and the MOMENTS about 0 are

$$\begin{aligned} \mu_1' &= \mu = M'(0) = p && (8) \\ \mu_2' &= M''(0) = p && (9) \\ \mu_n' &= M^{(n)}(0) = p. && (10) \end{aligned}$$

The MOMENTS about the MEAN are

$$\begin{aligned} \mu_2 &= \mu_2' - (\mu_1')^2 = p - p^2 = p(1-p) && (11) \\ \mu_3 &= \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3 = p - 3p^2 + 2p^3 \\ &= p(1-p)(1-2p) && (12) \\ \mu_4 &= \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4 \\ &= p - 4p^2 + 6p^3 - 3p^4 \\ &= p(1-p)(3p^2 - 3p + 1). && (13) \end{aligned}$$

The MEAN, VARIANCE, SKEWNESS, and KURTOSIS are then

$$\begin{aligned} \mu &= \mu_1' = p && (14) \\ \sigma^2 &= \mu_2 = p(1-p) && (15) \\ \gamma_1 &= \frac{\mu_3}{\sigma^3} = \frac{p(1-p)(1-2p)}{[p(1-p)]^{3/2}} \\ &= \frac{1-2p}{\sqrt{p(1-p)}} && (16) \\ \gamma_2 &= \frac{\mu_4}{\sigma^4} - 3 = \frac{p(1-2p)(2p^2-2p+1)}{p^2(1-p)^2} - 3 \\ &= \frac{6p^2 - 6p + 1}{p(1-p)}. && (17) \end{aligned}$$

To find an estimator for a population mean,

$$\begin{aligned} \langle p \rangle &= \sum_{Np=0}^{N} p \binom{N}{Np} \theta^{Np} (1-\theta)^{Nq} \\ &= \theta \sum_{Np=1}^{N} \binom{N-1}{Np-1} \theta^{Np-1} (1-\theta)^{Nq} \\ &= \theta[\theta + (1-\theta)]^{N-1} = \theta, \end{aligned} \tag{18}$$

so $\langle p \rangle$ is an UNBIASED ESTIMATOR for θ. The probability of Np successes in N trials is then

$$\binom{N}{Np} \theta^{Np} (1-\theta)^{Nq}, \tag{19}$$

where

$$p = \frac{[\text{number of successes}]}{N} \equiv \frac{n}{N}. \tag{20}$$

see also BINOMIAL DISTRIBUTION

Bernoulli Function

see BERNOULLI POLYNOMIAL

Bernoulli Inequality

$$(1+x)^n > 1+nx, \tag{1}$$

where $x \in \mathbb{R} > -1 \neq 0$, $n \in \mathbb{Z} > 1$. This inequality can be proven by taking a MACLAURIN SERIES of $(1+x)^n$,

$$(1+x)^n = 1+nx+\tfrac{1}{2}n(n-1)x^2+\tfrac{1}{6}n(n-1)(n-2)x^3+\ldots. \tag{2}$$

Since the series terminates after a finite number of terms for INTEGRAL n, the Bernoulli inequality for $x > 0$ is obtained by truncating after the first-order term. When $-1 < x < 0$, slightly more finesse is needed. In this case, let $y = |x| = -x > 0$ so that $0 < y < 1$, and take

$$(1-y)^n = 1-ny+\tfrac{1}{2}n(n-1)y^2-\tfrac{1}{6}n(n-1)(n-2)y^3+\ldots. \tag{3}$$

Since each POWER of y multiplies by a number < 1 and since the ABSOLUTE VALUE of the COEFFICIENT of each subsequent term is smaller than the last, it follows that the sum of the third order and subsequent terms is a POSITIVE number. Therefore,

$$(1-y)^n > 1-ny, \tag{4}$$

or

$$(1+x)^n > 1+nx, \qquad \text{for } -1 < x < 0, \tag{5}$$

completing the proof of the INEQUALITY over all ranges of parameters.

Bernoulli Lemniscate

see LEMNISCATE

Bernoulli Number

There are two definitions for the Bernoulli numbers. The older one, no longer in widespread use, defines the Bernoulli numbers B_n^* by the equations

$$\begin{aligned}\frac{x}{e^x-1}+\frac{x}{2}-1 &\equiv \sum_{n=1}^{\infty} \frac{(-1)^{n-1}B_n^* x^{2n}}{(2n)!} \\ &= \frac{B_1^* x^2}{2!} - \frac{B_2^* x^4}{4!} + \frac{B_3^* x^6}{6!} + \ldots\end{aligned} \tag{1}$$

for $|x| < 2\pi$, or

$$\begin{aligned}1-\frac{x}{2}\cot\left(\frac{x}{2}\right) &\equiv \sum_{n=1}^{\infty} \frac{B_n^* x^{2n}}{(2n)!} \\ &= \frac{B_1^* x^2}{2!} + \frac{B_2^* x^4}{4!} + \frac{B_3^* x^6}{6!} + \ldots\end{aligned} \tag{2}$$

for $|x| < \pi$ (Whittaker and Watson 1990, p. 125). Gradshteyn and Ryzhik (1979) denote these numbers B_n^*, while Bernoulli numbers defined by the newer (National Bureau of Standards) definition are denoted B_n. The B_n^* Bernoulli numbers may be calculated from the integral

$$B_n^* = 4n \int_0^{\infty} \frac{t^{2n-1}\,dt}{e^{2\pi t}-1}, \tag{3}$$

and analytically from

$$B_n^* = \frac{2(2n)!}{(2\pi)^{2n}} \sum_{p=1}^{\infty} p^{-2n} = \frac{2(2n)!}{(2\pi)^{2n}}\zeta(2n) \tag{4}$$

for $n = 1, 2, \ldots$, where $\zeta(z)$ is the RIEMANN ZETA FUNCTION.

The first few Bernoulli numbers B_n^* are

$$\begin{aligned}
B_1^* &= \tfrac{1}{6} \\
B_2^* &= \tfrac{1}{30} \\
B_3^* &= \tfrac{1}{42} \\
B_4^* &= \tfrac{1}{30} \\
B_5^* &= \tfrac{5}{66} \\
B_6^* &= \tfrac{691}{2{,}730} \\
B_7^* &= \tfrac{7}{6} \\
B_8^* &= \tfrac{3{,}617}{510} \\
B_9^* &= \tfrac{43{,}867}{798} \\
B_{10}^* &= \tfrac{174{,}611}{330} \\
B_{11}^* &= \tfrac{854{,}513}{138}.
\end{aligned}$$

Bernoulli numbers defined by the modern definition are denoted B_n and also called "EVEN-index" Bernoulli numbers. These are the Bernoulli numbers returned by the *Mathematica*® (Wolfram Research, Champaign, IL) function `BernoulliB[n]`. These Bernoulli numbers are a superset of the archaic ones B_n^* since

$$B_n \equiv \begin{cases} 1 & \text{for } n = 0 \\ -\frac{1}{2} & \text{for } n = 1 \\ (-1)^{(n/2)-1}B_{n/2}^* & \text{for } n \text{ even} \\ 0 & \text{for } n \text{ odd.} \end{cases} \tag{5}$$

The B_n can be defined by the identity

$$\frac{x}{e^x-1} \equiv \sum_{n=0}^{\infty} \frac{B_n x^n}{n!}. \tag{6}$$

These relationships can be derived using the generating function

$$F(x,t) = \sum_{n=0}^{\infty} \frac{B_n(x)t^n}{n!}, \tag{7}$$

which converges uniformly for $|t| < 2\pi$ and all x (Castellanos 1988). Taking the partial derivative gives

$$\frac{\partial F(x,t)}{\partial x} = \sum_{n=0}^{\infty} \frac{B_{n-1}(x)t^n}{(n-1)!} = t\sum_{n=0}^{\infty} \frac{B_n(x)t^n}{n!} = tF(x,t). \tag{8}$$

The solution to this differential equation is

$$F(x,t) = T(t)e^{xt}, \tag{9}$$

so integrating gives

$$\begin{aligned}\int_0^1 F(x,t)\,dx = T(t)\int_0^1 e^{xt}\,dx = T(t)\frac{e^t-1}{t} \\ = \sum_{n=0}^{\infty}\frac{t^n}{n!}\int_0^1 B_n(x)\,dx \\ = 1 + \sum_{n=1}^{\infty}\frac{t^n}{n!}\int_0^1 B_n(x)\,dx = 1\end{aligned} \tag{10}$$

or

$$\frac{te^{xt}}{e^t-1} = \sum_{n=0}^{\infty}\frac{B_n(x)t^n}{n!} \tag{11}$$

(Castellanos 1988). Setting $x = 0$ and adding $t/2$ to both sides then gives

$$\tfrac{1}{2}t\coth(\tfrac{1}{2}t) = \sum_{n=0}^{\infty}\frac{B_{2n}t^{2n}}{(2n)!}. \tag{12}$$

Letting $t = 2ix$ then gives

$$x\cot x = \sum_{n=0}^{\infty}(-1)^n B_{2n}\frac{2x^{2n}}{(2n)!} \tag{13}$$

for $x \in [-\pi, \pi]$. The Bernoulli numbers may also be calculated from the integral

$$B_n = \frac{n!}{2\pi i}\int \frac{z}{e^z-1}\frac{dz}{z^{n+1}}, \tag{14}$$

or from

$$B_n = \left[\frac{d^n}{dx^n}\frac{x}{e^x-1}\right]_{x=0}. \tag{15}$$

The Bernoulli numbers satisfy the identity

$$\binom{k+1}{1}B_k + \binom{k+1}{2}B_{k-1} + \ldots + \binom{k+1}{k}B_1 + B_0 = 0, \tag{16}$$

where $\binom{n}{k}$ is a Binomial Coefficient. An asymptotic Formula is

$$\lim_{n\to\infty}|B_{2n}| \sim 4\sqrt{\pi n}\left(\frac{n}{\pi e}\right)^{2n}. \tag{17}$$

Bernoulli numbers appear in expressions of the form $\sum_{k=1}^{n} k^p$, where $p = 1, 2, \ldots$. Bernoulli numbers also appear in the series expansions of functions involving $\tan x$, $\cot x$, $\csc x$, $\ln|\sin x|$, $\ln|\cos x|$, $\ln|\tan x|$, $\tanh x$, $\coth x$, and $\operatorname{csch} x$. An analytic solution exists for Even orders,

$$B_{2n} = \frac{(-1)^{n-1}2(2n)!}{(2\pi)^2 n}\sum_{p=1}^{\infty}p^{-2n} = \frac{(-1)^{n-1}2(2n)!}{(2\pi)^{2n}}\zeta(2n) \tag{18}$$

for $n = 1, 2, \ldots$, where $\zeta(2n)$ is the Riemann Zeta Function. Another intimate connection with the Riemann Zeta Function is provided by the identity

$$B_n = (-1)^{n+1}n\zeta(1-n). \tag{19}$$

The Denominator of B_{2k} is given by the von Staudt-Clausen Theorem

$$\operatorname{denom}(B_{2k}) = \prod_{\substack{p \text{ prime} \\ (p-1)|2k}}^{2k+1} p, \tag{20}$$

which also implies that the Denominator of B_{2k} is Squarefree (Hardy and Wright 1979). Another curious property is that the fraction part of B_n in Decimal has a Decimal Period which divides n, and there is a single digit before that period (Conway 1996).

$$\begin{aligned} B_0 &= 1 \\ B_1 &= -\tfrac{1}{2} \\ B_2 &= \tfrac{1}{6} \\ B_4 &= -\tfrac{1}{30} \\ B_6 &= \tfrac{1}{42} \\ B_8 &= -\tfrac{1}{30} \\ B_{10} &= \tfrac{5}{66} \\ B_{12} &= -\tfrac{691}{2{,}730} \\ B_{14} &= \tfrac{7}{6} \\ B_{16} &= -\tfrac{3{,}617}{510} \\ B_{18} &= \tfrac{43{,}867}{798} \\ B_{20} &= -\tfrac{174{,}611}{330} \\ B_{22} &= \tfrac{854{,}513}{138} \end{aligned}$$

(Sloane's A000367 and A002445). In addition,

$$B_{2n+1} = 0 \tag{21}$$

for $n = 1, 2, \ldots$.

Bernoulli first used the Bernoulli numbers while computing $\sum_{k=1}^{n} k^p$. He used the property of the Figurate Number Triangle that

$$\sum_{i=0}^{n} a_{ij} = \frac{(n+1)a_{nj}}{j+1}, \tag{22}$$

along with a form for a_{nj} which he derived inductively to compute the sums up to $n = 10$ (Boyer 1968, p. 85). For $p \in \mathbb{Z} > 0$, the sum is given by

$$\sum_{k=1}^{n} k^p = \sum_{k=1}^{n} \frac{(B+n+1)^{[p+1]} - B^{p+1}}{p+1}, \tag{23}$$

where the NOTATION $B^{[k]}$ means the quantity in question is raised to the appropriate POWER k, and all terms of the form B^m are replaced with the corresponding Bernoulli numbers B_m. Written explicitly in terms of a sum of POWERS,

$$\sum_{k=1}^{n} k^p = \frac{B_k p!}{k!(p-k+1)!} n^{p-k+1}. \tag{24}$$

It is also true that the COEFFICIENTS of the terms in such an expansion sum to 1 (which Bernoulli stated without proof). Ramanujan gave a number of curious infinite sum identities involving Bernoulli numbers (Berndt 1994).

G. J. Fee and S. Plouffe have computed $B_{200,000}$, which has $\sim 800,000$ DIGITS (Plouffe). Plouffe and collaborators have also calculated B_n for n up to 72,000.

see also ARGOH'S CONJECTURE, BERNOULLI FUNCTION, BERNOULLI POLYNOMIAL, DEBYE FUNCTIONS, EULER-MACLAURIN INTEGRATION FORMULAS, EULER NUMBER, FIGURATE NUMBER TRIANGLE, GENOCCHI NUMBER, PASCAL'S TRIANGLE, RIEMANN ZETA FUNCTION, VON STAUDT-CLAUSEN THEOREM

References
Abramowitz, M. and Stegun, C. A. (Eds.). "Bernoulli and Euler Polynomials and the Euler-Maclaurin Formula." §23.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 804–806, 1972.
Arfken, G. "Bernoulli Numbers, Euler-Maclaurin Formula." §5.9 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 327–338, 1985.
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 71, 1987.
Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 81–85, 1994.
Boyer, C. B. *A History of Mathematics.* New York: Wiley, 1968.
Castellanos, D. "The Ubiquitous Pi. Part I." *Math. Mag.* **61**, 67–98, 1988.
Conway, J. H. and Guy, R. K. In *The Book of Numbers.* New York: Springer-Verlag, pp. 107–110, 1996.
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, 1980.
Hardy, G. H. and Wright, W. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Oxford University Press, pp. 91–93, 1979.
Ireland, K. and Rosen, M. "Bernoulli Numbers." Ch. 15 in *A Classical Introduction to Modern Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 228–248, 1990.
Knuth, D. E. and Buckholtz, T. J. "Computation of Tangent, Euler, and Bernoulli Numbers." *Math. Comput.* **21**, 663–688, 1967.
Plouffe, S. "Plouffe's Inverter: Table of Current Records for the Computation of Constants." `http://lacim.uqam.ca/pi/records.html`.
Ramanujan, S. "Some Properties of Bernoulli's Numbers." *J. Indian Math. Soc.* **3**, 219–234, 1911.
Sloane, N. J. A. Sequences A000367/M4039 and A002445/M4189 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Spanier, J. and Oldham, K. B. "The Bernoulli Numbers, B_n." Ch. 4 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 35–38, 1987.
Wagstaff, S. S. Jr. "Ramanujan's Paper on Bernoulli Numbers." *J. Indian Math. Soc.* **45**, 49–65, 1981.
Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Bernoulli's Paradox

Suppose the HARMONIC SERIES converges to h:

$$\sum_{k=1}^{\infty} \frac{1}{k} = h.$$

Then rearranging the terms in the sum gives

$$h - 1 = h,$$

which is a contradiction.

References
Boas, R. P. "Some Remarkable Sequences of Integers." Ch. 3 in *Mathematical Plums* (Ed. R. Honsberger). Washington, DC: Math. Assoc. Amer., pp. 39–40, 1979.

Bernoulli Polynomial

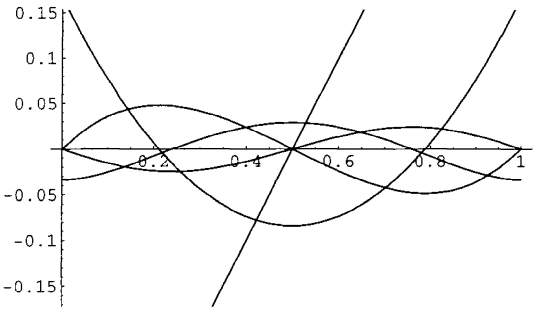

There are two definitions of Bernoulli polynomials in use. The nth Bernoulli polynomial is denoted here by $B_n(x)$, and the archaic Bernoulli polynomial by $B_n^*(x)$. These definitions correspond to the BERNOULLI NUMBERS evaluated at 0,

$$B_n \equiv B_n(0) \tag{1}$$

$$B_n^* \equiv B_n^*(0). \tag{2}$$

They also satisfy

$$B_n(1) = (-1)^n B_n(0) \tag{3}$$

and

$$B_n(1-x) = (-1)^n B_n(x) \tag{4}$$

(Lehmer 1988). The first few Bernoulli POLYNOMIALS are

$$\begin{aligned}
B_0(x) &= 1\\
B_1(x) &= x - \tfrac{1}{2}\\
B_2(x) &= x^2 - x + \tfrac{1}{6}\\
B_3(x) &= x^3 - \tfrac{3}{2}x^2 + \tfrac{1}{2}x\\
B_4(x) &= x^4 - 2x^3 + x^2 - \tfrac{1}{30}\\
B_5(x) &= x^5 - \tfrac{5}{2}x^4 + \tfrac{5}{3}x^3 - \tfrac{1}{6}x\\
B_6(x) &= x^6 - 3x^5 + \tfrac{5}{2}x^4 - \tfrac{1}{2}x^2 + \tfrac{1}{42}.
\end{aligned}$$

Bernoulli (1713) defined the POLYNOMIALS in terms of sums of the POWERS of consecutive integers,

$$\sum_{k=0}^{m-1} k^{n-1} = \frac{1}{n}[B_n(m) - B_n(0)]. \tag{5}$$

Euler (1738) gave the Bernoulli POLYNOMIALS $B_n(x)$ in terms of the generating function

$$\frac{te^{tx}}{e^t - 1} \equiv \sum_{n=0}^{\infty} B_n(x)\frac{t^n}{n!}. \tag{6}$$

They satisfy recurrence relation

$$\frac{dB_n}{dx} = nB_{n-1}(x) \tag{7}$$

(Appell 1882), and obey the identity

$$B_n(x) = (B + x)^n, \tag{8}$$

where B^k is interpreted here as $B_k(x)$. Hurwitz gave the FOURIER SERIES

$$B_n(x) = -\frac{n!}{(2\pi i)^n} \sum_{k=-\infty}^{\infty} k^{-n} e^{2\pi ikx}, \tag{9}$$

for $0 < x < 1$, and Raabe (1851) found

$$\frac{1}{m}\sum_{k=0}^{m-1} B_n\left(x + \frac{k}{m}\right) = m^{-n}B_n(mx). \tag{10}$$

A sum identity involving the Bernoulli POLYNOMIALS is

$$\begin{aligned}
&\sum_{k=0}^{m} \binom{m}{k} B_k(\alpha)B_{m-k}(\beta)\\
&= -(m-1)B_m(\alpha+\beta) + m(\alpha+\beta-1)B_{m-1}(\alpha+\beta)
\end{aligned} \tag{11}$$

for an INTEGER m and arbitrary REAL NUMBERS α and β.

see also BERNOULLI NUMBER, EULER-MACLAURIN INTEGRATION FORMULAS, EULER POLYNOMIAL

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Bernoulli and Euler Polynomials and the Euler-Maclaurin Formula." §23.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 804–806, 1972.

Appell, P. E. "Sur une classe de polynomes." *Annales d'École Normal Superieur, Ser. 2* **9**, 119–144, 1882.

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 330, 1985.

Bernoulli, J. *Ars conjectandi.* Basel, Switzerland, p. 97, 1713. Published posthumously.

Euler, L. "Methodus generalis summandi progressiones." *Comment. Acad. Sci. Petropol.* **6**, 68–97, 1738.

Lehmer, D. H. "A New Approach to Bernoulli Polynomials." *Amer. Math. Monthly.* **95**, 905–911, 1988.

Lucas, E. Ch. 14 in *Théorie des Nombres.* Paris, 1891.

Raabe, J. L. "Zurückführung einiger Summen und bestimmten Integrale auf die Jakob Bernoullische Function." *J. reine angew. Math.* **42**, 348–376, 1851.

Spanier, J. and Oldham, K. B. "The Bernoulli Polynomial $B_n(x)$." Ch. 19 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 167–173, 1987.

Bernoulli's Theorem

see WEAK LAW OF LARGE NUMBERS

Bernoulli Trial

An experiment in which s TRIALS are made of an event, with probability p of success in any given TRIAL.

Bernstein-Bézier Curve

see BÉZIER CURVE

Bernstein's Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let $E_n(f)$ be the error of the best uniform approximation to a REAL function $f(x)$ on the INTERVAL $[-1, 1]$ by REAL POLYNOMIALS of degree at most n. If

$$\alpha(x) = |x|, \tag{1}$$

then Bernstein showed that

$$0.267\ldots < \lim_{n\to\infty} 2nE_{2n}(\alpha) < 0.286. \tag{2}$$

He conjectured that the lower limit (β) was $\beta = 1/(2\sqrt{\pi})$. However, this was disproven by Varga and Carpenter (1987) and Varga (1990), who computed

$$\beta = 0.2801694990\ldots. \tag{3}$$

For rational approximations $p(x)/q(x)$ for p and q of degree m and n, D. J. Newman (1964) proved

$$\tfrac{1}{2}e^{-9\sqrt{n}} \leq E_{n,n}(\alpha) \leq 3e^{-\sqrt{n}} \tag{4}$$

for $n \geq 4$. Gonchar (1967) and Bulanov (1975) improved the lower bound to

$$e^{-\pi\sqrt{n+1}} \leq E_{n,n}(\alpha) \leq 3e^{-\sqrt{n}}. \tag{5}$$

Vjacheslavo (1975) proved the existence of POSITIVE constants m and M such that

$$m \leq e^{\pi\sqrt{n}} E_{n,n}(\alpha) < M \tag{6}$$

(Petrushev 1987, pp. 105–106). Varga *et al.* (1993) conjectured and Stahl (1993) proved that

$$\lim_{n\to\infty} e^{\pi\sqrt{2n}} E_{2n,2n}(\alpha) = 8. \tag{7}$$

References

Bulanov, A. P. "Asymptotics for the Best Rational Approximation of the Function Sign x." *Mat. Sbornik* **96**, 171–178, 1975.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/brnstn/brnstn.html`.

Gonchar, A. A. "Estimates for the Growth of Rational Functions and their Applications." *Mat. Sbornik* **72**, 489–503, 1967.

Newman, D. J. "Rational Approximation to $|x|$." *Michigan Math. J.* **11**, 11-14, 1964.

Petrushev, P. P. and Popov, V. A. *Rational Approximation of Real Functions.* New York: Cambridge University Press, 1987.

Stahl, H. "Best Uniform Rational Approximation of $|x|$ on $[-1,1]$." *Russian Acad. Sci. Sb. Math.* **76**, 461–487, 1993.

Varga, R. S. *Scientific Computations on Mathematical Problems and Conjectures.* Philadelphia, PA: SIAM, 1990.

Varga, R. S. and Carpenter, A. J. "On a Conjecture of S. Bernstein in Approximation Theory." *Math. USSR Sbornik* **57**, 547–560, 1987.

Varga, R. S.; Ruttan, A.; and Carpenter, A. J. "Numerical Results on Best Uniform Rational Approximations to $|x|$ on $[-1,+1]$. *Math. USSR Sbornik* **74**, 271–290, 1993.

Vjacheslavo, N. S. "On the Uniform Approximation of $|x|$ by Rational Functions." *Dokl. Akad. Nauk SSSR* **220**, 512–515, 1975.

Bernstein's Inequality

Let P be a POLYNOMIAL of degree n with derivative P'. Then

$$||P'||_\infty \leq n||P||_\infty,$$

where

$$||P||_\infty \equiv \max_{|z|=1} |P(z)|.$$

Bernstein Minimal Surface Theorem

If a MINIMAL SURFACE is given by the equation $z = f(x,y)$ and f has CONTINUOUS first and second PARTIAL DERIVATIVES for all REAL x and y, then f is a PLANE.

References

Hazewinkel, M. (Managing Ed.). *Encyclopaedia of Mathematics: An Updated and Annotated Translation of the Soviet "Mathematical Encyclopaedia."* Dordrecht, Netherlands: Reidel, p. 369, 1988.

Bernstein Polynomial

The POLYNOMIALS defined by

$$B_{i,n}(t) = \binom{n}{i} t^i (1-t)^{n-i},$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT. The Bernstein polynomials of degree n form a basis for the POWER POLYNOMIALS of degree n.

see also BÉZIER CURVE

Bernstein's Polynomial Theorem

If $g(\theta)$ is a trigonometric POLYNOMIAL of degree m satisfying the condition $|g(\theta)| \leq 1$ where θ is arbitrary and real, then $g'(\theta) \leq m$.

References

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., p. 5, 1975.

Bernstein-Szegő Polynomials

The POLYNOMIALS on the interval $[-1,1]$ associated with the WEIGHT FUNCTIONS

$$w(x) = (1-x^2)^{-1/2}$$
$$w(x) = (1-x^2)^{1/2}$$
$$w(x) = \sqrt{\frac{1-x}{1+x}},$$

also called BERNSTEIN POLYNOMIALS.

References

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., pp. 31–33, 1975.

Berry-Osseen Inequality

Gives an estimate of the deviation of a DISTRIBUTION FUNCTION as a SUM of independent RANDOM VARIABLES with a NORMAL DISTRIBUTION.

References
Hazewinkel, M. (Managing Ed.). *Encyclopaedia of Mathematics: An Updated and Annotated Translation of the Soviet "Mathematical Encyclopaedia."* Dordrecht, Netherlands: Reidel, p. 369, 1988.

Berry Paradox

There are several versions of the Berry paradox, the original version of which was published by Bertrand Russell and attributed to Oxford University librarian Mr. G. Berry. In one form, the paradox notes that the number "one million, one hundred thousand, one hundred and twenty one" can be named by the description: "the first number not nameable in under ten words." However, this latter expression has only nine words, so the number *can* be named in under ten words, so there is an inconsistency in naming it in this manner!

References
Chaitin, G. J. "The Berry Paradox." *Complexity* **1**, 26–30, 1995.

Bertelsen's Number

An erroneous value of $\pi(10^9)$, where $\pi(x)$ is the PRIME COUNTING FUNCTION. Bertelsen's value of 50,847,478 is 56 lower than the correct value of 50,847,534.

References
Brown, K. S. "Bertelsen's Number." `http://www.seanet.com/~ksbrown/kmath049.htm`.

Bertini's Theorem

The general curve of a system which is LINEARLY INDEPENDENT on a certain number of given irreducible curves will not have a singular point which is not fixed for all the curves of the system.

References
Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 115, 1959.

Bertrand Curves

Two curves which, at any point, have a common principal NORMAL VECTOR are called Bertrand curves. The product of the TORSIONS of Bertrand curves is a constant.

Bertrand's Paradox

see BERTRAND'S PROBLEM

Bertrand's Postulate

If $n > 3$, there is always at least one PRIME between n and $2n - 2$. Equivalently, if $n > 1$, then there is always at least one PRIME between n and $2n$. It was proved in 1850–51 by Chebyshev, and is therefore sometimes known as CHEBYSHEV'S THEOREM. An elegant proof was later given by Erdős. An extension of this result is that if $n > k$, then there is a number containing a PRIME divisor $> k$ in the sequence $n, n+1, \ldots, n+k-1$. (The case $n = k + 1$ then corresponds to Bertrand's postulate.) This was first proved by Sylvester, independently by Schur, and a simple proof was given by Erdős.

A related problem is to find the least value of θ so that there exists at least one PRIME between n and $n+\mathcal{O}(n^\theta)$ for sufficiently large n (Berndt 1994). The smallest known value is $\theta = 6/11 + \epsilon$ (Lou and Yao 1992).

see also CHOQUET THEORY, DE POLIGNAC'S CONJECTURE, PRIME NUMBER

References
Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, p. 135, 1994.
Erdős, P. "Ramanujan and I." In *Proceedings of the International Ramanujan Centenary Conference held at Anna University, Madras, Dec. 21, 1987.* (Ed. K. Alladi). New York: Springer-Verlag, pp. 1–20, 1989.
Lou, S. and Yau, Q. "A Chebyshev's Type of Prime Number Theorem in a Short Interval (II)." *Hardy-Ramanujan J.* **15**, 1–33, 1992.

Bertrand's Problem

What is the PROBABILITY that a CHORD drawn at RANDOM on a CIRCLE of RADIUS r has length $\geq r$? The answer, it turns out, depends on the interpretation of "two points drawn at RANDOM." In the usual interpretation that ANGLES θ_1 and θ_2 are picked at RANDOM on the CIRCUMFERENCE,

$$P = \frac{\pi - \frac{\pi}{3}}{\pi} = \frac{2}{3}.$$

However, if a point is instead placed at RANDOM on a RADIUS of the CIRCLE and a CHORD drawn PERPENDICULAR to it,

$$P = \frac{\frac{\sqrt{3}}{2} r}{r} = \frac{\sqrt{3}}{2}.$$

The latter interpretation is more satisfactory in the sense that the result remains the same for a rotated CIRCLE, a slightly smaller CIRCLE INSCRIBED in the first, or for a CIRCLE of the same size but with its center slightly offset. Jaynes (1983) shows that the interpretation of "RANDOM" as a continuous UNIFORM DISTRIBUTION over the RADIUS is the only one possessing all these three invariances.

References
Bogomolny, A. "Bertrand's Paradox." `http://www.cut-the-knot.com/bertrand.html`.
Jaynes, E. T. *Papers on Probability, Statistics, and Statistical Physics.* Dordrecht, Netherlands: Reidel, 1983.
Pickover, C. A. *Keys to Infinity.* New York: Wiley, pp. 42–45, 1995.

Bertrand's Test

A CONVERGENCE TEST also called DE MORGAN'S AND BERTRAND'S TEST. If the ratio of terms of a SERIES $\{a_n\}_{n=1}^{\infty}$ can be written in the form

$$\frac{a_n}{a_{n+1}} = 1 + \frac{1}{n} + \frac{\rho_n}{n \ln n},$$

then the series converges if $\underline{\lim}_{n\to\infty}\rho_n > 1$ and diverges if $\overline{\lim}_{n\to\infty}\rho_n < 1$, where $\underline{\lim}_{n\to\infty}$ is the LOWER LIMIT and $\overline{\lim}_{n\to\infty}$ is the UPPER LIMIT.

see also KUMMER'S TEST

References

Bromwich, T. J. I'a and MacRobert, T. M. *An Introduction to the Theory of Infinite Series, 3rd ed.* New York: Chelsea, p. 40, 1991.

Bertrand's Theorem

see BERTRAND'S POSTULATE

Besov Space

A type of abstract SPACE which occurs in SPLINE and RATIONAL FUNCTION approximations. The Besov space $B_{p,q}^{\alpha}$ is a complete quasinormed space which is a BANACH SPACE when $1 \le p,\ q \le \infty$ (Petrushev and Popov 1987).

References

Bergh, J. and Löfström, J. *Interpolation Spaces.* New York: Springer-Verlag, 1976.

Peetre, J. *New Thoughts on Besov Spaces.* Durham, NC: Duke University Press, 1976.

Petrushev, P. P. and Popov, V. A. "Besov Spaces." §7.2 in *Rational Approximation of Real Functions.* New York: Cambridge University Press, pp. 201–203, 1987.

Triebel, H. *Interpolation Theory, Function Spaces, Differential Operators.* New York: Elsevier, 1978.

Bessel's Correction

The factor $(N-1)/N$ in the relationship between the VARIANCE σ and the EXPECTATION VALUES of the SAMPLE VARIANCE,

$$\langle s^2 \rangle = \frac{N-1}{N}\sigma^2, \tag{1}$$

where

$$s^2 \equiv \langle x^2 \rangle - \langle x \rangle^2. \tag{2}$$

For two samples,

$$\hat{\sigma}^2 = \frac{N_1 {s_1}^2 + N_2 {s_2}^2}{N_1 + N_2 - 2}. \tag{3}$$

see also SAMPLE VARIANCE, VARIANCE

Bessel Differential Equation

$$x^2 \frac{d^2y}{dx^2} + x\frac{dy}{dx} + (x^2 - m^2)y = 0. \tag{1}$$

Equivalently, dividing through by x^2,

$$\frac{d^2y}{dx^2} + \frac{1}{x}\frac{dy}{dx} + \left(1 - \frac{m^2}{x^2}\right)y = 0. \tag{2}$$

The solutions to this equation define the BESSEL FUNCTIONS. The equation has a regular SINGULARITY at 0 and an irregular SINGULARITY at ∞.

A transformed version of the Bessel differential equation given by Bowman (1958) is

$$x^2 \frac{d^2y}{dx^2} + (2p+1)x\frac{dy}{dx} + (\alpha^2 x^{2r} + \beta^2)y = 0. \tag{3}$$

The solution is

$$y = x^{-p}\left[C_1 J_{q/r}\left(\frac{\alpha}{r}x^r\right) + C_2 Y_{q/r}\left(\frac{\alpha}{r}x^r\right)\right], \tag{4}$$

where

$$q \equiv \sqrt{p^2 - \beta^2}, \tag{5}$$

J and Y are the BESSEL FUNCTIONS OF THE FIRST and SECOND KINDS, and C_1 and C_2 are constants. Another form is given by letting $y = x^{\alpha} J_n(\beta x^{\gamma})$, $\eta = y x^{-\alpha}$, and $\xi = \beta x^{\gamma}$ (Bowman 1958, p. 117), then

$$\frac{d^2y}{dx^2} - \frac{2\alpha - 1}{x}\frac{dy}{dx} + \left(\beta^2\gamma^2 x^{2\gamma-2} + \frac{\alpha^2 - n^2\gamma^2}{x^2}\right)y = 0. \tag{6}$$

The solution is

$$y = \begin{cases} x^{\alpha}[A J_n(\beta x^{\gamma}) + B Y_n(\beta x^{\gamma})] & \text{for integral } n \\ A J_n(\beta x^{\gamma}) + B J_{-n}(\beta x^{\gamma})] & \text{for nonintegral } n. \end{cases} \tag{7}$$

see also AIRY FUNCTIONS, ANGER FUNCTION, BEI, BER, BESSEL FUNCTION, BOURGET'S HYPOTHESIS, CATALAN INTEGRALS, CYLINDRICAL FUNCTION, DINI EXPANSION, HANKEL FUNCTION, HANKEL'S INTEGRAL, HEMISPHERICAL FUNCTION, KAPTEYN SERIES, LIPSCHITZ'S INTEGRAL, LOMMEL DIFFERENTIAL EQUATION, LOMMEL FUNCTION, LOMMEL'S INTEGRALS, NEUMANN SERIES (BESSEL FUNCTION), PARSEVAL'S INTEGRAL, POISSON INTEGRAL, RAMANUJAN'S INTEGRAL, RICCATI DIFFERENTIAL EQUATION, SONINE'S INTEGRAL, STRUVE FUNCTION, WEBER FUNCTIONS, WEBER'S DISCONTINUOUS INTEGRALS

References

Bowman, F. *Introduction to Bessel Functions.* New York: Dover, 1958.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 550, 1953.

Bessel's Finite Difference Formula

An INTERPOLATION formula also sometimes known as

$$f_p = f_0 + p\delta_{1/2} + B_2(\delta_0^2 + \delta_1^2) + B_3\delta_{1/2}^3 + B_4(\delta_0^4 + \delta_1^4) + B_5\delta_{1/2}^5 + \ldots, \qquad (1)$$

for $p \in [0, 1]$, where δ is the CENTRAL DIFFERENCE and

$$B_{2n} \equiv \tfrac{1}{2}G_{2n} \equiv \tfrac{1}{2}(E_{2n} + F_{2n}) \qquad (2)$$

$$B_{2n+1} \equiv G_{2n+1} - \tfrac{1}{2}G_{2n} \equiv \tfrac{1}{2}(F_{2n} - E_{2n}) \qquad (3)$$

$$E_{2n} \equiv G_{2n} - G_{2n+1} \equiv B_{2n} - B_{2n+1} \qquad (4)$$

$$F_{2n} \equiv G_{2n+1} \equiv B_{2n} + B_{2n+1}, \qquad (5)$$

where G_k are the COEFFICIENTS from GAUSS'S BACKWARD FORMULA and GAUSS'S FORWARD FORMULA and E_k and F_k are the COEFFICIENTS from EVERETT'S FORMULA. The B_ks also satisfy

$$B_{2n}(p) = B_{2n}(q) \qquad (6)$$

$$B_{2n+1}(p) = -B_{2n+1}(q), \qquad (7)$$

for

$$q \equiv 1 - p. \qquad (8)$$

see also EVERETT'S FORMULA

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 880, 1972.

Acton, F. S. *Numerical Methods That Work, 2nd printing.* Washington, DC: Math. Assoc. Amer., pp. 90–91, 1990.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 433, 1987.

Bessel's First Integral

$$J_n(x) = \frac{1}{\pi}\int_0^\pi \cos(n\theta - x\sin\theta)\,d\theta,$$

where $J_n(x)$ is a BESSEL FUNCTION OF THE FIRST KIND.

Bessel's Formula

see BESSEL'S FINITE DIFFERENCE FORMULA, BESSEL'S INTERPOLATION FORMULA, BESSEL'S STATISTICAL FORMULA

Bessel Function

A function $Z(x)$ defined by the RECURRENCE RELATIONS

$$Z_{m+1} + Z_{m-1} = \frac{2m}{x}Z_m$$

and

$$Z_{m+1} - Z_{m-1} = -2\frac{dZ_m}{dx}.$$

The Bessel functions are more frequently defined as solutions to the DIFFERENTIAL EQUATION

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + (x^2 - m^2)y = 0.$$

There are two classes of solution, called the BESSEL FUNCTION OF THE FIRST KIND J and BESSEL FUNCTION OF THE SECOND KIND Y. (A BESSEL FUNCTION OF THE THIRD KIND is a special combination of the first and second kinds.) Several related functions are also defined by slightly modifying the defining equations.

see also BESSEL FUNCTION OF THE FIRST KIND, BESSEL FUNCTION OF THE SECOND KIND, BESSEL FUNCTION OF THE THIRD KIND, CYLINDER FUNCTION, HEMICYLINDRICAL FUNCTION, MODIFIED BESSEL FUNCTION OF THE FIRST KIND, MODIFIED BESSEL FUNCTION OF THE SECOND KIND, SPHERICAL BESSEL FUNCTION OF THE FIRST KIND, SPHERICAL BESSEL FUNCTION OF THE SECOND KIND

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Bessel Functions of Integer Order," "Bessel Functions of Fractional Order," and "Integrals of Bessel Functions." Chs. 9–11 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 355–389, 435–456, and 480–491, 1972.

Arfken, G. "Bessel Functions." Ch. 11 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 573–636, 1985.

Bickley, W. G. *Bessel Functions and Formulae.* Cambridge, England: Cambridge University Press, 1957.

Bowman, F. *Introduction to Bessel Functions.* New York: Dover, 1958.

Gray, A. and Matthews, G. B. *A Treatise on Bessel Functions and Their Applications to Physics, 2nd ed.* New York: Dover, 1966.

Luke, Y. L. *Integrals of Bessel Functions.* New York: McGraw-Hill, 1962.

McLachlan, N. W. *Bessel Functions for Engineers, 2nd ed. with corrections.* Oxford, England: Clarendon Press, 1961.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Bessel Functions of Integral Order" and "Bessel Functions of Fractional Order, Airy Functions, Spherical Bessel Functions." §6.5 and 6.7 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 223–229 and 234–245, 1992.

Watson, G. N. *A Treatise on the Theory of Bessel Functions, 2nd ed.* Cambridge, England: Cambridge University Press, 1966.

Bessel Function of the First Kind

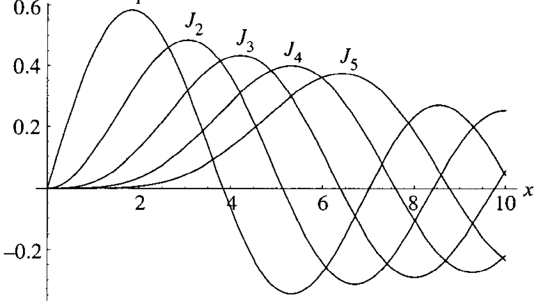

The Bessel functions of the first kind $J_n(x)$ are defined as the solutions to the BESSEL DIFFERENTIAL EQUATION

$$x^2\frac{d^2y}{dx^2}+x\frac{dy}{dx}+(x^2-m^2)y=0 \tag{1}$$

which are nonsingular at the origin. They are sometimes also called CYLINDER FUNCTIONS or CYLINDRICAL HARMONICS. The above plot shows $J_n(x)$ for $n = 1, 2, \ldots, 5$.

To solve the differential equation, apply FROBENIUS METHOD using a series solution of the form

$$y=x^k\sum_{n=0}^{\infty}a_nx^n=\sum_{n=0}^{\infty}a_nx^{n+k}. \tag{2}$$

Plugging into (1) yields

$$\begin{aligned}x^2\sum_{n=0}^{\infty}(k+n)(k+n-1)a_nx^{k+n-2}&\\+x\sum_{n=0}^{\infty}(k+n)a_nx^{k+n-1}+x^2\sum_{n=0}^{\infty}a_nx^{k+n}&\\-m^2\sum_{n=0}^{\infty}a_nx^{n+k}&=0\end{aligned} \tag{3}$$

$$\begin{aligned}\sum_{n=0}^{\infty}(k+n)(k+n-1)a_nx^{k+n}+\sum_{n=0}^{\infty}(k+n)a_nx^{k+n}&\\+\sum_{n=2}^{\infty}a_{n-2}x^{k+n}-m^2\sum_{n=0}^{\infty}a_nx^{n+k}&=0.\end{aligned} \tag{4}$$

The INDICIAL EQUATION, obtained by setting $n = 0$, is

$$a_0[k(k-1)+k-m^2]=a_0(k^2-m^2)=0. \tag{5}$$

Since a_0 is defined as the first NONZERO term, $k^2-m^2=0$, so $k=\pm m$. Now, if $k=m$,

$$\begin{aligned}\sum_{n=0}^{\infty}[(m+n)(m+n-1)+(m+n)-m^2]a_nx^{m+n}&\\+\sum_{n=2}^{\infty}a_{n-2}x^{m+n}&=0\end{aligned} \tag{6}$$

$$\sum_{n=0}^{\infty}[(m+n)^2-m^2]a_nx^{m+n}+\sum_{n=2}^{\infty}a_{n-2}x^{m+n}=0 \tag{7}$$

$$\sum_{n=0}^{\infty}n(2m+n)a_nx^{m+n}+\sum_{n=2}^{\infty}a_{n-2}x^{m+n}=0 \tag{8}$$

$$a_1(2m+1)+\sum_{n=2}^{\infty}[a_nn(2m+n)+a_{n-2}]x^{m+n}=0. \tag{9}$$

First, look at the special case $m=-1/2$, then (9) becomes

$$\sum_{n=2}^{\infty}[a_nn(n-1)+a_{n-2}]x^{m+n}=0, \tag{10}$$

so

$$a_n=-\frac{1}{n(n-1)}a_{n-2}. \tag{11}$$

Now let $n\equiv 2l$, where $l=1, 2, \ldots$.

$$\begin{aligned}a_{2l}&=-\frac{1}{2l(2l-1)}a_{2l-2}\\&=\frac{(-1)^l}{[2l(2l-1)][2(l-1)(2l-3)]\cdots[2\cdot1\cdot1]}a_0\\&=\frac{(-1)^l}{2^ll!(2l-1)!!}a_0,\end{aligned} \tag{12}$$

which, using the identity $2^ll!(2l-1)!!=(2l)!$, gives

$$a_{2l}=\frac{(-1)^l}{(2l)!}a_0. \tag{13}$$

Similarly, letting $n\equiv 2l+1$

$$\begin{aligned}a_{2l+1}&=-\frac{1}{(2l+1)(2l)}a_{2l-1}\\&=\frac{(-1)^l}{[2l(2l+1)][2(l-1)(2l-1)]\cdots[2\cdot1\cdot3][1]}a_1,\end{aligned} \tag{14}$$

which, using the identity $2^ll!(2l+1)!!=(2l+1)!$, gives

$$a_{2l+1}=\frac{(-1)^l}{2^ll!(2l+1)!!}a_1=\frac{(-1)^l}{(2l+1)!}a_1. \tag{15}$$

Plugging back into (2) with $k=m=-1/2$ gives

$$\begin{aligned}y&=x^{-1/2}\sum_{n=0}^{\infty}a_nx^n\\&=x^{-1/2}\left[\sum_{n=1,3,5,\ldots}^{\infty}a_nx^n+\sum_{n=0,2,4,\ldots}^{\infty}a_nx^n\right]\\&=x^{-1/2}\left[\sum_{l=0}^{\infty}a_{2l}x^{2l}+\sum_{l=0}^{\infty}a_{2l+1}x^{2l+1}\right]\\&=x^{-1/2}\left[a_0\sum_{l=0}^{\infty}\frac{(-1)^l}{(2l)!}x^{2l}+a_1\sum_{l=0}^{\infty}\frac{(-1)^l}{(2l+1)!}x^{2l+1}\right]\\&=x^{-1/2}(a_0\cos x+a_1\sin x).\end{aligned} \tag{16}$$

The BESSEL FUNCTIONS of order $\pm1/2$ are therefore defined as

$$J_{-1/2}(x)\equiv\sqrt{\frac{2}{\pi x}}\cos x \tag{17}$$

$$J_{1/2}(x)\equiv\sqrt{\frac{2}{\pi x}}\sin x, \tag{18}$$

so the general solution for $m = \pm 1/2$ is

$$y = a_0' J_{-1/2}(x) + a_1' J_{1/2}(x). \tag{19}$$

Now, consider a general $m \neq -1/2$. Equation (9) requires

$$a_1(2m+1) = 0 \tag{20}$$

$$[a_n n(2m+n) + a_{n-2}]x^{m+n} = 0 \tag{21}$$

for $n = 2, 3, \ldots$, so

$$a_1 = 0 \tag{22}$$

$$a_n = -\frac{1}{n(2m+n)} a_{n-2} \tag{23}$$

for $n = 2, 3, \ldots$. Let $n \equiv 2l+1$, where $l = 1, 2, \ldots$, then

$$\begin{aligned} a_{2l+1} &= -\frac{1}{(2l+1)[2(m+1)+1]} a_{2l-1} \\ &= \ldots = f(n,m) a_1 = 0, \end{aligned} \tag{24}$$

where $f(n,m)$ is the function of l and m obtained by iterating the recursion relationship down to a_1. Now let $n \equiv 2l$, where $l = 1, 2, \ldots$, so

$$\begin{aligned} a_{2l} &= -\frac{1}{2l(2m+2l)} a_{2l-2} = -\frac{1}{4l(m+l)} a_{2l-2} \\ &= \frac{(-1)^l}{[4l(m+l)][4(l-1)(m+l-1)]\cdots[4\cdot(m+1)]} a_0. \end{aligned} \tag{25}$$

Plugging back into (9),

$$\begin{aligned} y &= \sum_{n=0}^{\infty} a_n x^{n+m} = \sum_{n=1,3,5,\ldots}^{\infty} a_n x^{n+m} + \sum_{n=0,2,4,\ldots}^{\infty} a_n x^{n+m} \\ &= \sum_{l=0}^{\infty} a_{2l+1} x^{2l+m+1} + \sum_{l=0}^{\infty} a_{2l} x^{2l+m} \\ &= a_0 \sum_{l=0}^{\infty} \frac{(-1)^l}{[4l(m+l)][4(l-1)(m+l-1)]\cdots[4\cdot(m+1)]} x^{2l+m} \\ &= a_0 \sum_{l=0}^{\infty} \frac{[(-1)^l m(m-1)\cdots 1]x^{2l+m}}{[4l(m+l)][4(l-1)(m+l-1)]\cdots[m(m-1)\cdots 1]} \\ &= a_0 \sum_{l=0}^{\infty} \frac{(-1)^l m!}{4^l l!(m+l)!} = a_0 \sum_{l=0}^{\infty} \frac{(-1)^l m!}{2^{2l} l!(m+l)!}. \end{aligned} \tag{26}$$

Now define

$$J_m(x) \equiv \sum_{l=0}^{\infty} \frac{(-1)^l}{2^{2l+m} l!(m+l)!} x^{2l+m}, \tag{27}$$

where the factorials can be generalized to GAMMA FUNCTIONS for nonintegral m. The above equation then becomes

$$y = a_0 2^m m! J_m(x) = a_0' J_m(x). \tag{28}$$

Returning to equation (5) and examining the case $k = -m$,

$$a_1(1-2m) + \sum_{n=2}^{\infty} [a_n n(n-2m) + a_{n-2}]x^{n-m} = 0. \tag{29}$$

However, the sign of m is arbitrary, so the solutions must be the same for $+m$ and $-m$. We are therefore free to replace $-m$ with $-|m|$, so

$$a_1(1+2|m|) + \sum_{n=2}^{\infty} [a_n n(n+2|m|) + a_{n-2}]x^{|m|+n} = 0, \tag{30}$$

and we obtain the same solutions as before, but with m replaced by $|m|$.

$$J_m(x) = \begin{cases} \sum_{l=0}^{\infty} \frac{(-1)^l}{2^{2l+|m|} l!(|m|+l)!} x^{2l+|m|} & \text{for } |m| \neq -\frac{1}{2} \\ \sqrt{\frac{2}{\pi x}} \cos x & \text{for } m = -\frac{1}{2} \\ \sqrt{\frac{2}{\pi x}} \sin x & \text{for } m = \frac{1}{2}. \end{cases} \tag{31}$$

We can relate J_m and J_{-m} (when m is an INTEGER) by writing

$$J_{-m}(x) = \sum_{l=0}^{\infty} \frac{(-1)^l}{2^{2l-m} l!(l-m)!} x^{2l-m}. \tag{32}$$

Now let $l \equiv l' + m$. Then

$$\begin{aligned} J_{-m}(x) &= \sum_{l'+m=0}^{\infty} \frac{(-1)^{l'+m}}{2^{2l'+m}(l'+m)!l!} x^{2l'+m} \\ &= \sum_{l'=-m}^{-1} \frac{(-1)^{l'+m}}{2^{2l'+m} l'!(l'+m)!} x^{2l'+m} \\ &\quad + \sum_{l'=0}^{\infty} \frac{(-1)^{l'+m}}{2^{2l'+m} l'!(l'+m)!} x^{2l'+m}. \end{aligned} \tag{33}$$

But $l'! = \infty$ for $l' = -m, \ldots, -1$, so the DENOMINATOR is infinite and the terms on the right are zero. We therefore have

$$J_{-m}(x) = \sum_{l=0}^{\infty} \frac{(-1)^{l+m}}{2^{2l+m} l!(l+m)!} x^{2l+m} = (-1)^m J_m(x). \tag{34}$$

Note that the BESSEL DIFFERENTIAL EQUATION is second-order, so there must be two linearly independent solutions. We have found both only for $|m| = 1/2$. For a general nonintegral order, the independent solutions are J_m and J_{-m}. When m is an INTEGER, the general (real) solution is of the form

$$Z_m \equiv C_1 J_m(x) + C_2 Y_m(x), \tag{35}$$

where J_m is a Bessel function of the first kind, Y_m (a.k.a. N_m) is the Bessel Function of the Second Kind (a.k.a. Neumann Function or Weber Function), and C_1 and C_2 are constants. Complex solutions are given by the Hankel Functions (a.k.a. Bessel Functions of the Third Kind).

The Bessel functions are Orthogonal in $[0, 1]$ with respect to the weight factor x. Except when $2n$ is a Negative Integer,

$$J_m(z) = \frac{z^{-1/2}}{2^{2m+1/2} i^{m+1/2} \Gamma(m+1)} M_{0,m}(2iz), \tag{36}$$

where $\Gamma(x)$ is the Gamma Function and $M_{0,m}$ is a Whittaker Function.

In terms of a Confluent Hypergeometric Function of the First Kind, the Bessel function is written

$$J_\nu(z) = \frac{(\frac{1}{2}z)^\nu}{\Gamma(\nu+1)} {}_0F_1(\nu+1; -\tfrac{1}{4}z^2). \tag{37}$$

A derivative identity for expressing higher order Bessel functions in terms of $J_0(x)$ is

$$J_n(x) = i^n T_n\left(i\frac{d}{dx}\right) J_0(x), \tag{38}$$

where $T_n(x)$ is a Chebyshev Polynomial of the First Kind. Asymptotic forms for the Bessel functions are

$$J_m(x) \approx \frac{1}{\Gamma(m+1)} \left(\frac{x}{2}\right)^m \tag{39}$$

for $x \ll 1$ and

$$J_m(x) \approx \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{m\pi}{2} - \frac{\pi}{4}\right) \tag{40}$$

for $x \gg 1$. A derivative identity is

$$\frac{d}{dx}[x^m J_m(x)] = x^m J_{m-1}(x). \tag{41}$$

An integral identity is

$$\int_0^u u' J_0(u')\, du' = u J_1(u). \tag{42}$$

Some sum identities are

$$1 = [J_0(x)]^2 + 2[J_1(x)]^2 + 2[J_2(x)]^2 + \ldots \tag{43}$$

$$1 = J_0(x) + 2J_2(x) + 2J_4(x) + \ldots \tag{44}$$

and the Jacobi-Anger Expansion

$$e^{iz\cos\theta} = \sum_{n=-\infty}^{\infty} i^n J_n(z) e^{in\theta}, \tag{45}$$

which can also be written

$$e^{iz\cos\theta} = J_0(z) + 2\sum_{n=1}^{\infty} i^n J_n(z)\cos(n\theta). \tag{46}$$

The Bessel function addition theorem states

$$J_n(y+z) = \sum_{m=-\infty}^{\infty} J_m(y) J_{n-m}(z). \tag{47}$$

Roots of the Function $J_n(x)$ are given in the following table.

zero	$J_0(x)$	$J_1(x)$	$J_2(x)$	$J_3(x)$	$J_4(x)$	$J_5(x)$
1	2.4048	3.8317	5.1336	6.3802	7.5883	8.7715
2	5.5201	7.0156	8.4172	9.7610	11.0647	12.3386
3	8.6537	10.1735	11.6198	13.0152	14.3725	15.7002
4	11.7915	13.3237	14.7960	16.2235	17.6160	18.9801
5	14.9309	16.4706	17.9598	19.4094	20.8269	22.2178

Let x_n be the nth Root of the Bessel function $J_0(x)$, then

$$\sum_{n=1}^{\infty} \frac{1}{x_n J_0(x_n)} = 0.38479\ldots \tag{48}$$

(Le Lionnais 1983).

The Roots of its Derivatives are given in the following table.

zero	$J_0'(x)$	$J_1'(x)$	$J_2'(x)$	$J_3'(x)$	$J_4'(x)$	$J_5'(x)$
1	3.8317	1.8412	3.0542	4.2012	5.3175	6.4156
2	7.0156	5.3314	6.7061	8.0152	9.2824	10.5199
3	10.1735	8.5363	9.9695	11.3459	12.6819	13.9872
4	13.3237	11.7060	13.1704	14.5858	15.9641	17.3128
5	16.4706	14.8636	16.3475	17.7887	19.1960	20.5755

Various integrals can be expressed in terms of Bessel functions

$$J_0(z) = \frac{1}{2\pi}\int_0^{2\pi} e^{iz} \cos\phi\, d\phi \tag{49}$$

$$J_n(z) = \frac{1}{\pi}\int_0^{\pi} \cos(z\sin\theta - n\theta)\, d\theta, \tag{50}$$

which is Bessel's First Integral,

$$J_n(z) = \frac{i^{-n}}{\pi}\int_0^{\pi} e^{iz\cos\theta}\cos(n\theta)\, d\theta \tag{51}$$

$$J_n(z) = \frac{1}{2\pi i^n}\int_0^{2\pi} e^{iz\cos\phi} e^{in\phi}\, d\phi \tag{52}$$

for $n = 1, 2, \ldots,$

$$J_n(z) = \frac{2}{\pi}\frac{x^n}{(2m-1)!!}\int_0^{\pi/2} \sin^{2n} u \cos(x\cos u)\, du \tag{53}$$

for $n = 1, 2, \ldots,$

$$J_n(x) = \frac{1}{2\pi i}\int_\gamma e^{(x/2)(z-1/z)} z^{-n-1}\, dz \tag{54}$$

for $n > -1/2$. Integrals involving $J_1(x)$ include

$$\int_0^\infty J_1(x)\,dx = 1 \tag{55}$$

$$\int_0^\infty \left[\frac{J_1(x)}{x}\right]^2 dx = \frac{4}{3\pi} \tag{56}$$

$$\int_0^\infty \left[\frac{J_1(x)}{x}\right]^2 x\,dx = \frac{1}{2}. \tag{57}$$

see also BESSEL FUNCTION OF THE SECOND KIND, DEBYE'S ASYMPTOTIC REPRESENTATION, DIXON-FERRAR FORMULA, HANSEN-BESSEL FORMULA, KAPTEYN SERIES, KNESER-SOMMERFELD FORMULA, MEHLER'S BESSEL FUNCTION FORMULA, NICHOLSON'S FORMULA, POISSON'S BESSEL FUNCTION FORMULA, SCHLÄFLI'S FORMULA, SCHLÖMILCH'S SERIES, SOMMERFELD'S FORMULA, SONINE-SCHAFHEITLIN FORMULA, WATSON'S FORMULA, WATSON-NICHOLSON FORMULA, WEBER'S DISCONTINUOUS INTEGRALS, WEBER'S FORMULA, WEBER-SONINE FORMULA, WEYRICH'S FORMULA

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Bessel Functions J and Y." §9.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 358–364, 1972.

Arfken, G. "Bessel Functions of the First Kind, $J_\nu(x)$" and "Orthogonality." §11.1 and 11.2 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 573–591 and 591–596, 1985.

Lehmer, D. H. "Arithmetical Periodicities of Bessel Functions." *Ann. Math.* **33**, 143–150, 1932.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 25, 1983.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 619–622, 1953.

Spanier, J. and Oldham, K. B. "The Bessel Coefficients $J_0(x)$ and $J_1(x)$" and "The Bessel Function $J_\nu(x)$." Chs. 52–53 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 509–520 and 521–532, 1987.

Bessel Function Fourier Expansion

Let $n \geq 1/2$ and α_1, α_2, ... be the POSITIVE ROOTS of $J_n(x) = 0$. An expansion of a function in the interval (0,1) in terms of BESSEL FUNCTIONS OF THE FIRST KIND

$$f(x) = \sum_{l=1}^{\infty} A_r J_n(x\alpha_r), \tag{1}$$

has COEFFICIENTS found as follows:

$$\int_0^1 xf(x)J_n(x\alpha_l)\,dx = \sum_{r=1}^{\infty} A_r \int_0^1 xJ_n(x\alpha_r)J_n(x\alpha_l)\,dx. \tag{2}$$

But ORTHOGONALITY of BESSEL FUNCTION ROOTS gives

$$\int_0^1 xJ_n(x\alpha_l)J_n(x\alpha_r)\,dx = \tfrac{1}{2}\delta_{l,r}{J_{n+1}}^2(\alpha_r) \tag{3}$$

(Bowman 1958, p. 108), so

$$\begin{aligned}\int_0^1 xf(x)J_n(x\alpha_l)\,dx &= \tfrac{1}{2}\sum_{r=1}^{\infty} A_r\delta_{l,r}{J_{n+1}}^2(x\alpha_r)\\ &= \tfrac{1}{2}A_l{J_{n+1}}^2(\alpha_l),\end{aligned} \tag{4}$$

and the COEFFICIENTS are given by

$$A_l = \frac{2}{{J_{n+1}}^2(\alpha_l)}\int_0^1 xf(x)J_n(x\alpha_l)\,dx. \tag{5}$$

References

Bowman, F. *Introduction to Bessel Functions.* New York: Dover, 1958.

Bessel Function of the Second Kind

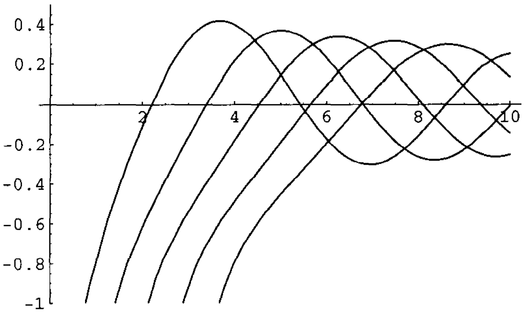

A Bessel function of the second kind $Y_n(x)$ is a solution to the BESSEL DIFFERENTIAL EQUATION which is singular at the origin. Bessel functions of the second kind are also called NEUMANN FUNCTIONS or WEBER FUNCTIONS. The above plot shows $Y_n(x)$ for $n = 1, 2, \ldots, 5$.

Let $v \equiv J_m(x)$ be the first solution and u be the other one (since the BESSEL DIFFERENTIAL EQUATION is second-order, there are two LINEARLY INDEPENDENT solutions). Then

$$xu'' + u' + xu = 0 \tag{1}$$

$$xv'' + v' + xv = 0. \tag{2}$$

Take $v \times (1) - u \times (2)$,

$$x(u''v - uv'') + u'v - uv' = 0 \tag{3}$$

$$\frac{d}{dx}[x(u'v - uv')] = 0, \tag{4}$$

so $x(u'v - uv') = B$, where B is a constant. Divide by xv^2,

$$\frac{u'v - uv'}{v^2} = \frac{d}{dx}\left(\frac{u}{v}\right) = \frac{B}{xv^2} \tag{5}$$

$$\frac{u}{v} = A + B\int\frac{dx}{xv^2}. \tag{6}$$

Rearranging and using $v \equiv J_m(x)$ gives

$$u = AJ_m(x) + BJ_m(x)\int \frac{dx}{xJ_m{}^2(x)}$$
$$\equiv A'J_m(x) + B'Y_m(x), \qquad (7)$$

where the Bessel function of the second kind is defined by

$$Y_m(x) = \frac{J_m(x)\cos(m\pi) - J_{-m}(x)}{\sin(m\pi)}$$
$$= \frac{1}{\pi}\sum_{k=1}^{\infty} \frac{(-1)^k x^{m+2k}}{2^{m+2k}k!(m+k)!}\left[2\ln\left(\frac{x}{2}\right) + 2\gamma - b_{m+k} - b_k\right]$$
$$-\frac{1}{\pi}\sum_{k=0}^{m-1} \frac{x^{-m+2k}(m-k-1)!}{2^{-m+2k}k!} \qquad (8)$$

$m = 0,\ 1,\ 2,\ \ldots,\ \gamma$ is the EULER-MASCHERONI CONSTANT, and

$$b_k \equiv \begin{cases} 0 & k = 0, \\ \sum_{n=0}^{k} \frac{1}{n} & k \neq 0. \end{cases} \qquad (9)$$

The function is given by

$$Y_n(z) = \frac{1}{\pi}\int_0^{\pi} \sin(z\sin\theta - n\theta)\,d\theta$$
$$-\frac{1}{\pi}\int_0^{\infty} [e^{nt} + e^{-nt}(-1)^n]e^{-z\sinh t}\,dt. \qquad (10)$$

Asymptotic equations are

$$Y_m(x) = \begin{cases} \frac{2}{\pi}\left[\ln(\frac{1}{2}x) + \gamma\right] & m = 0, x \ll 1 \\ -\frac{\Gamma(m)}{\pi}\left(\frac{2}{x}\right)^m & m \neq 0, x \ll 1 \end{cases} \qquad (11)$$

$$Y_m(x) = \sqrt{\frac{2}{\pi x}}\sin\left(x - \frac{m\pi}{2} - \frac{\pi}{4}\right) \qquad x \gg 1, \qquad (12)$$

where $\Gamma(z)$ is a GAMMA FUNCTION.

see also BESSEL FUNCTION OF THE FIRST KIND, BOURGET'S HYPOTHESIS, HANKEL FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Bessel Functions J and Y." §9.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 358–364, 1972.

Arfken, G. "Neumann Functions, Bessel Functions of the Second Kind, $N_\nu(x)$." §11.3 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 596–604, 1985.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 625-627, 1953.

Spanier, J. and Oldham, K. B. "The Neumann Function $Y_\nu(x)$." Ch. 54 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 533-542, 1987.

Bessel Function of the Third Kind

see HANKEL FUNCTION

Bessel's Inequality

If $f(x)$ is piecewise CONTINUOUS and has a general FOURIER SERIES

$$\sum_i a_i\phi_i(x) \qquad (1)$$

with WEIGHTING FUNCTION $w(x)$, it must be true that

$$\int\left[f(x) - \sum_i a_i\phi_i(x)\right]^2 w(x)\,dx \geq 0 \qquad (2)$$

$$\int f^2(x)w(x)\,dx - 2\sum_i a_i\int f(x)\phi_i(x)w(x)\,dx$$
$$+\sum_i a_i{}^2\int \phi_i{}^2(x)w(x)\,dx \geq 0. \qquad (3)$$

But the COEFFICIENT of the generalized FOURIER SERIES is given by

$$a_m \equiv \int f(x)\phi_m(x)w(x)\,dx, \qquad (4)$$

so

$$\int f^2(x)w(x)\,dx - 2\sum_i a_i{}^2 + \sum_i a_i{}^2 \geq 0 \qquad (5)$$

$$\int f^2(x)w(x)\,dx \geq \sum_i a_i{}^2. \qquad (6)$$

Equation (6) is an inequality if the functions ϕ_i are not COMPLETE. If they are COMPLETE, then the inequality (2) becomes an equality, so (6) becomes an equality and is known as PARSEVAL'S THEOREM. If $f(x)$ has a simple FOURIER SERIES expansion with COEFFICIENTS a_0, a_1, …, a_n and b_1, …, b_n, then

$$\tfrac{1}{2}a_0{}^2 + \sum_{k=1}^{\infty}(a_k{}^2 + b_k{}^2) \leq \frac{1}{\pi}\int_{-\pi}^{\pi}[f(x)]^2\,dx. \qquad (7)$$

The inequality can also be derived from SCHWARZ'S INEQUALITY

$$|\langle f|g\rangle|^2 \leq \langle f|f\rangle\langle g|g\rangle \qquad (8)$$

by expanding g in a superposition of EIGENFUNCTIONS of f, $g = \sum_i a_i f_i$. Then

$$\langle f|g\rangle = \sum_i a_i\langle f|f_i\rangle \leq \sum_i a_i \qquad (9)$$

$$|\langle f|g\rangle|^2 \leq \left|\sum_i a_i\right|^2 = \left(\sum_i a_i\right)\left(\sum_i a_i{}^*\right)$$
$$= \sum_i a_i a_i{}^* \leq \langle f|f\rangle\langle g|g\rangle. \qquad (10)$$

If g is normalized, then $\langle g|g\rangle = 1$ and

$$\langle f|f\rangle \geq \sum_i a_i a_i^*. \tag{11}$$

see also SCHWARZ'S INEQUALITY, TRIANGLE INEQUALITY

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 526–527, 1985.

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1102, 1980.

Bessel's Interpolation Formula

see BESSEL'S FINITE DIFFERENCE FORMULA

Bessel Polynomial

see BESSEL FUNCTION

Bessel's Second Integral

see POISSON INTEGRAL

Bessel's Statistical Formula

$$t = \frac{\bar{w}-\omega}{\sigma_w/\sqrt{N}} = \frac{\bar{w}-\omega}{\sqrt{\frac{\sum_{i=1}^n (w_i-\bar{w})^2}{N(N-1)}}}, \tag{1}$$

where

$$\bar{w} \equiv \hat{x}_1 - \bar{x}_2 \tag{2}$$

$$\omega \equiv \mu_{(1)} - \mu_{(2)} \tag{3}$$

$$N \equiv N_1 + N_2. \tag{4}$$

Beta

A financial measure of a fund's sensitivity to market movements which measures the relationship between a fund's excess return over Treasury Bills and the excess return of a benchmark index (which, by definition, has $\beta = 1$). A fund with a beta of β has performed $r = (\beta - 1) \times 100\%$ better (or $|r|$ worse if $r < 0$) than its benchmark index (after deducting the T-bill rate) in up markets and $|r|$ worse (or $|r|$ better if $r < 0$) in down markets.

see also ALPHA, SHARPE RATIO

Beta Distribution

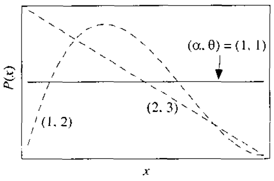

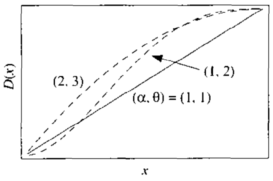

A general type of statistical DISTRIBUTION which is related to the GAMMA DISTRIBUTION. Beta distributions have two free parameters, which are labeled according to one of two notational conventions. The usual definition calls these α and β, and the other uses $\beta' \equiv \beta - 1$ and $\alpha' \equiv \alpha - 1$ (Beyer 1987, p. 534). The above plots are for $(\alpha, \beta) = (1, 1)$ [solid], (1, 2) [dotted], and (2, 3) [dashed]. The probability function $P(x)$ and DISTRIBUTION FUNCTION $D(x)$ are given by

$$\begin{aligned} P(x) &= \frac{(1-x)^{\beta-1}x^{\alpha-1}}{B(\alpha,\beta)} \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}(1-x)^{\beta-1}x^{\alpha-1} \qquad (1) \\ D(x) &= I(x;a,b), \qquad (2) \end{aligned}$$

where $B(a, b)$ is the BETA FUNCTION, $I(x; a, b)$ is the REGULARIZED BETA FUNCTION, and $0 < x < 1$ where α, $\beta > 0$. The distribution is normalized since

$$\begin{aligned} \int_0^1 P(x)\,dx &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}B(\alpha,\beta) = 1. \qquad (3) \end{aligned}$$

The CHARACTERISTIC FUNCTION is

$$\phi(t) = {}_1F_1(a, a+b, it). \tag{4}$$

The MOMENTS are given by

$$M_r = \int_0^1 (x-\mu)^r\,dx = \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+r)}{\Gamma(\alpha+\beta+r)\Gamma(\alpha)}. \tag{5}$$

The MEAN is

$$\begin{aligned} \mu &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}x\,dx \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}B(\alpha+1,\beta) \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} = \frac{\alpha}{\alpha+\beta}, \qquad (6) \end{aligned}$$

and the VARIANCE, SKEWNESS, and KURTOSIS are

$$\begin{aligned} \sigma^2 &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \qquad (7) \\ \gamma_1 &= \frac{2(\sqrt{\beta}-\sqrt{\alpha})(\sqrt{\alpha}+\sqrt{\beta})\sqrt{1+\alpha+\beta}}{\sqrt{\alpha\beta}\,(\alpha+\beta+2)} \qquad (8) \\ \gamma_2 &= \frac{6(\alpha^2+\alpha^3-4\alpha\beta-2\alpha^2\beta+\beta^2-2\alpha\beta^2+\beta^3)}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)}. \qquad (9) \end{aligned}$$

The MODE of a variate distributed as $\beta(\alpha, \beta)$ is

$$\hat{x} = \frac{\alpha-1}{\alpha+\beta-2}. \tag{10}$$

In "normal" form, the distribution is written

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \tag{11}$$

and the MEAN, VARIANCE, SKEWNESS, and KURTOSIS are

$$\mu = \frac{\alpha}{\alpha+\beta} \tag{12}$$

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(1+\alpha+\beta)} \tag{13}$$

$$\gamma_1 = \frac{2(\sqrt{\alpha}-\sqrt{\beta})(\sqrt{\alpha}+\sqrt{\beta})\sqrt{1+\alpha+\beta}}{\sqrt{\alpha\beta}(\alpha+\beta+2)} \tag{14}$$

$$\gamma_2 = \frac{3(1+\alpha+\beta)(2\alpha^2-2\alpha\beta+\alpha^2\beta+2\beta^2+\alpha\beta^2)}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)}. \tag{15}$$

see also GAMMA DISTRIBUTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 944–945, 1972.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 534–535, 1987.

Beta Function

The beta function is the name used by Legendre and Whittaker and Watson (1990) for the EULERIAN INTEGRAL OF THE SECOND KIND. To derive the integral representation of the beta function, write the product of two FACTORIALS as

$$m!n! = \int_0^\infty e^{-u}u^m\,du \int_0^\infty e^{-v}v^n\,dv. \tag{1}$$

Now, let $u \equiv x^2$, $v \equiv y^2$, so

$$\begin{aligned} m!n! &= 4\int^\infty e^{-x^2}x^{2m+1}\,dx \int_0^\infty e^{-y^2}y^{2n+1}\,dy \\ &= 4\int_{-\infty}^\infty \int_{-\infty}^\infty e^{-(x^2+y^2)}x^{2m+1}y^{2m+1}\,dx\,dy. \end{aligned} \tag{2}$$

Transforming to POLAR COORDINATES with $x = r\cos\theta$, $y = r\sin\theta$

$$\begin{aligned} m!n! &= 4\int_0^{\pi/2}\int_0^\infty e^{-r^2}(r\cos\theta)^{2m+1}(r\sin\theta)^{2n+1}r\,dr\,d\theta \\ &= 4\int_0^\infty e^{-r^2}r^{2m+2n+3}\,dr \int_0^{\pi/2}\cos^{2m+1}\theta\sin^{2n+1}\theta\,d\theta \\ &= 2(m+n+1)!\int_0^{\pi/2}\cos^{2m+1}\theta\sin^{2n+1}\theta\,d\theta. \end{aligned} \tag{3}$$

The beta function is then defined by

$$\begin{aligned} B(m+1,n+1) &= B(n+1,m+1) \\ &\equiv 2\int_0^{\pi/2}\cos^{2m+1}\theta\sin^{2n+1}\theta\,d\theta = \frac{m!n!}{(m+n+1)!}. \end{aligned} \tag{4}$$

Rewriting the arguments,

$$B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} = \frac{(p-1)!(q-1)!}{(p+q-1)!}. \tag{5}$$

The general trigonometric form is

$$\int_0^{\pi/2}\sin^n x\cos^m x\,dx = \tfrac{1}{2}B(n+\tfrac{1}{2}, m+\tfrac{1}{2}). \tag{6}$$

Equation (6) can be transformed to an integral over POLYNOMIALS by letting $u \equiv \cos^2\theta$,

$$B(m+1,n+1) \equiv \frac{m!n!}{(m+n+1)!} = \int_0^1 u^m(1-u)^n\,du \tag{7}$$

$$B(m,n) \equiv \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)} = \int_0^1 u^{m-1}(1-u)^{n-1}\,du. \tag{8}$$

To put it in a form which can be used to derive the LEGENDRE DUPLICATION FORMULA, let $x \equiv \sqrt{u}$, so $u = x^2$ and $du = 2x\,dx$, and

$$\begin{aligned} B(m,n) &= \int_0^1 x^{2(m-1)}(1-x^2)^{n-1}(2x\,dx) \\ &= 2\int_0^1 x^{2m-1}(1-x^2)^{n-1}\,dx. \end{aligned} \tag{9}$$

To put it in a form which can be used to develop integral representations of the BESSEL FUNCTIONS and HYPERGEOMETRIC FUNCTION, let $u \equiv x/(1+x)$, so

$$B(m+1,n+1) = \int_0^\infty \frac{u^m\,du}{(1+u)^{m+n+2}}. \tag{10}$$

Various identities can be derived using the GAUSS MULTIPLICATION FORMULA

$$\begin{aligned} B(np,nq) &= \frac{\Gamma(np)\Gamma(nq)}{\Gamma[n(p+q)]} \\ &= n^{-nq}\frac{B(p,q)B(p+\frac{1}{n},q)\cdots B(p+\frac{n-1}{n},q)}{B(q,q)B(2q,q)\cdots B([n-1]q,q)}. \end{aligned} \tag{11}$$

Additional identities include

$$\begin{aligned} B(p,q+1) &= \frac{\Gamma(p)\Gamma(q+1)}{\Gamma(p+q+1)} = \frac{q}{p}\frac{\Gamma(p+1)\Gamma(q)}{\Gamma([p+1]q)} \\ &= \frac{q}{p}B(p+1,q) \end{aligned} \tag{12}$$

$$B(p,q) = B(p+1,q) + B(p,q+1) \tag{13}$$

$$B(p,q+1)=\frac{q}{p+q}B(p,q). \tag{14}$$

If n is a POSITIVE INTEGER, then

$$B(p,n+1)=\frac{1\cdot 2\cdots n}{p(p+1)\cdots(p+n)} \tag{15}$$

$$B(p,p)B(p+\tfrac{1}{2},p+\tfrac{1}{2})=\frac{\pi}{2^{4p-1}p} \tag{16}$$

$$B(p+q)B(p+q,r)=B(q,r)B(q+r,p). \tag{17}$$

A generalization of the beta function is the incomplete beta function

$$\begin{aligned}B(t;x,y)&\equiv\int_0^t u^{x-1}(1-u)^{y-1}\,du\\&=t^p\left[\frac{1}{x}+\frac{1-y}{x+1}t+\ldots+\frac{(1-y)\cdots(n-y)}{n!(x+n)}t^n+\ldots\right].\end{aligned} \tag{18}$$

see also CENTRAL BETA FUNCTION, DIRICHLET INTEGRALS, GAMMA FUNCTION, REGULARIZED BETA FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Beta Function" and "Incomplete Beta Function." §6.2 and 6.6 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 258 and 263, 1972.

Arfken, G. "The Beta Function." §10.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 560–565, 1985.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 425, 1953.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Gamma Function, Beta Function, Factorials, Binomial Coefficients" and "Incomplete Beta Function, Student's Distribution, F-Distribution, Cumulative Binomial Distribution." §6.1 and 6.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 206–209 and 219–223, 1992.

Spanier, J. and Oldham, K. B. "The Incomplete Beta Function $B(\nu;\mu;x)$." Ch. 58 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 573–580, 1987.

Whittaker, E. T. and Watson, G. N. *A Course of Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Beta Function (Exponential)

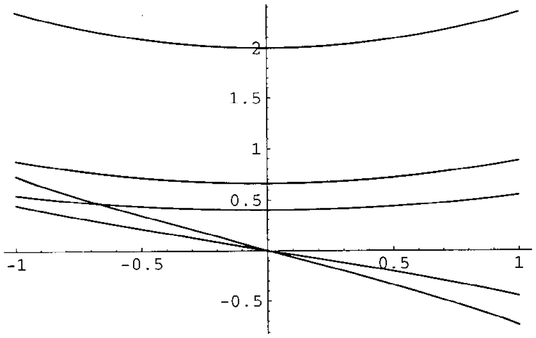

Another "BETA FUNCTION" defined in terms of an integral is the "exponential" beta function, given by

$$\beta_n(z)\equiv\int_{-1}^{1}t^n e^{-zt}\,dt \tag{1}$$

$$=n!z^{-(n+1)}\left[e^z\sum_{k=0}^{n}\frac{(-1)^k z^k}{k!}-e^{-z}\sum_{k=0}^{n}\frac{z^k}{k!}\right]. \tag{2}$$

The exponential beta function satisfies the RECURRENCE RELATION

$$z\beta_n(z)=(-1)^n e^z-e^{-z}+n\beta_{n-1}(z). \tag{3}$$

The first few integral values are

$$\beta_0(z)=\frac{2\sinh z}{z} \tag{4}$$

$$\beta_1(z)=\frac{2(\sinh z-z\cosh z)}{z^2} \tag{5}$$

$$\beta_2(a)=\frac{2(2+z^2)\sinh z-4z\cosh z}{z^3}. \tag{6}$$

see also ALPHA FUNCTION

Beta Prime Distribution

A distribution with probability function

$$P(x)=\frac{x^{\alpha-1}(1+x)^{-\alpha-\beta}}{B(\alpha,\beta)},$$

where B is a BETA FUNCTION. The MODE of a variate distributed as $\beta'(\alpha,\beta)$ is

$$\hat{x}=\frac{\alpha-1}{\beta+1}.$$

If x is a $\beta'(\alpha,\beta)$ variate, then $1/x$ is a $\beta'(\beta,\alpha)$ variate. If x is a $\beta(\alpha,\beta)$ variate, then $(1-x)/x$ and $x/(1-x)$ are $\beta'(\beta,\alpha)$ and $\beta'(\alpha,\beta)$ variates. If x and y are $\gamma(\alpha_1)$ and $\gamma(\alpha_2)$ variates, then x/y is a $\beta'(\alpha_1,\alpha_2)$ variate. If $x^2/2$ and $y^2/2$ are $\gamma(1/2)$ variates, then $z^2\equiv(x/y)^2$ is a $\beta'(1/2,1/2)$ variate.

Bethe Lattice

see CAYLEY TREE

Betrothed Numbers

see QUASIAMICABLE PAIR

Betti Group

The free part of the HOMOLOGY GROUP with a domain of COEFFICIENTS in the GROUP of INTEGERS (if this HOMOLOGY GROUP is finitely generated).

References

Hazewinkel, M. (Managing Ed.). *Encyclopaedia of Mathematics: An Updated and Annotated Translation of the Soviet "Mathematical Encyclopaedia."* Dordrecht, Netherlands: Reidel, p. 380, 1988.

Betti Number

Betti numbers are topological objects which were proved to be invariants by Poincaré, and used by him to extend the POLYHEDRAL FORMULA to higher dimensional spaces. The nth Betti number is the rank of the nth HOMOLOGY GROUP. Let p_r be the RANK of the HOMOLOGY GROUP H_r of a TOPOLOGICAL SPACE K. For a closed, orientable surface of GENUS g, the Betti numbers are $p_0 = 1$, $p_1 = 2g$, and $p_2 = 1$. For a nonorientable surface with k CROSS-CAPS, the Betti numbers are $p_0 = 1$, $p_1 = k - 1$, and $p_2 = 0$.

see also EULER CHARACTERISTIC, POINCARÉ DUALITY

Bézier Curve

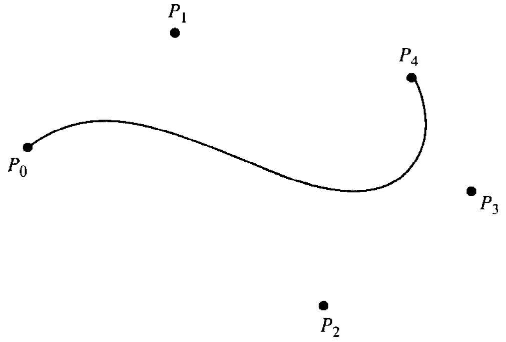

Given a set of n control points, the corresponding Bézier curve (or BERNSTEIN-BÉZIER CURVE) is given by

$$\mathbf{C}(t) = \sum_{i=0}^{n} \mathbf{P}_i B_{i,n}(t),$$

where $B_{i,n}(t)$ is a BERNSTEIN POLYNOMIAL and $t \in [0,1]$.

A "rational" Bézier curve is defined by

$$\mathbf{C}(t) = \frac{\sum_{i=0}^{n} B_{i,p}(t) w_i \mathbf{P}_i}{\sum_{i=0}^{n} B_{i,p}(t) w_i},$$

where p is the order, $B_{i,p}$ are the BERNSTEIN POLYNOMIALS, $\mathbf{P}_i$ are control points, and the weight w_i of $\mathbf{P}_i$ is the last ordinate of the homogeneous point $\mathbf{P}_i^w$. These curves are closed under perspective transformations, and can represent CONIC SECTIONS exactly.

The Bézier curve always passes through the first and last control points and lies within the CONVEX HULL of the control points. The curve is tangent to $\mathbf{P}_1 - \mathbf{P}_0$ and $\mathbf{P}_n - \mathbf{P}_{n-1}$ at the endpoints. The "variation diminishing property" of these curves is that no line can have more intersections with a Bézier curve than with the curve obtained by joining consecutive points with straight line segments. A desirable property of these curves is that the curve can be translated and rotated by performing these operations on the control points.

Undesirable properties of Bézier curves are their numerical instability for large numbers of control points, and the fact that moving a single control point changes the global shape of the curve. The former is sometimes avoided by smoothly patching together low-order Bézier curves. A generalization of the Bézier curve is the B-SPLINE.

see also B-SPLINE, NURBS CURVE

Bézier Spline

see BÉZIER CURVE, SPLINE

Bezout Numbers

Integers (λ, μ) for a and b such that

$$\lambda a + \mu b = \mathrm{GCD}(a, b).$$

For INTEGERS $a_1, \ldots, a_n$, the Bezout numbers are a set of numbers $k_1, \ldots, k_n$ such that

$$k_1 a_1 + k_2 a_2 + \ldots + k_n a_n = d,$$

where d is the GREATEST COMMON DIVISOR of $a_1, \ldots, a_n$.

see also GREATEST COMMON DIVISOR

Bezout's Theorem

In general, two algebraic curves of degrees m and n intersect in $m{\cdot}n$ points and cannot meet in more than $m{\cdot}n$ points unless they have a component in common (i.e., the equations defining them have a common factor). This can also be stated: if P and Q are two POLYNOMIALS with no roots in common, then there exist two other POLYNOMIALS A and B such that $AP + BQ = 1$. Similarly, given N POLYNOMIAL equations of degrees n_1, n_2, $\ldots n_N$ in N variables, there are in general $n_1 n_2 \cdots n_N$ common solutions.

see also POLYNOMIAL

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 10, 1959.

Bhargava's Theorem

Let the nth composition of a function $f(x)$ be denoted $f^{(n)}(x)$, such that $f^{(0)}(x) = x$ and $f^{(1)}(x) = f(x)$. Denote $f \circ g(x) = f(g(x))$, and define

$$\sum F(a,b,c) = F(a,b,c) + F(b,c,a) + F(c,a,b). \quad (1)$$

Let

$$u \equiv (a,b,c) \quad (2)$$

$$|u| \equiv a + b + c \quad (3)$$

$$||u|| \equiv a^4 + b^4 + c^4, \quad (4)$$

and

$$\begin{aligned} f(u) &= (f_1(u), f_2(u), f_3(u)) && (5)\\ &= (a(b-c), b(c-a), c(a-b)) && (6)\\ g(u) &= (g_1(u), g_2(u), g_3(u))\\ &= \left(\sum a^2b, \sum ab^2, 3abc\right). && (7) \end{aligned}$$

Then if $|u| = 0$,

$$\begin{aligned} \|f^{(m)} \circ g^{(n)}(u)\| &= 2(ab+bc+ca)^{2^{m+1}3^n}\\ &= \|g^{(n)} \circ f^{(m)}(u)\|, && (8) \end{aligned}$$

where $m, n \in \{0,\ 1,\ \dots\}$ and composition is done in terms of components.

see also DIOPHANTINE EQUATION—QUARTIC, FORD'S THEOREM

References
Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 97–100, 1994.
Bhargava, S. "On a Family of Ramanujan's Formulas for Sums of Fourth Powers." *Ganita* **43**, 63–67, 1992.

Bhaskara-Brouckner Algorithm

see SQUARE ROOT

Bi-Connected Component

A maximal SUBGRAPH of an undirected graph such that any two edges in the SUBGRAPH lie on a common simple cycle.

see also STRONGLY CONNECTED COMPONENT

Bianchi Identities

The RIEMANN TENSOR is defined by

$$R_{\lambda\mu\nu\kappa;\eta} = \frac{1}{2}\frac{\partial}{\partial x^\eta}\left(\frac{\partial^2 g_{\lambda\nu}}{\partial x^\kappa \partial x^\mu} - \frac{\partial^2 g_{\mu\nu}}{\partial x^\kappa \partial x^\lambda} - \frac{\partial^2 g_{\lambda\kappa}}{\partial x^\mu \partial x^\nu} + \frac{\partial^2 g_{\mu\kappa}}{\partial x^\nu \partial x^\lambda}\right).$$

Permuting ν, κ, and η (Weinberg 1972, pp. 146–147) gives the Bianchi identities

$$R_{\lambda\mu\nu\kappa;\eta} + R_{\lambda\mu\eta\nu;\kappa} + R_{\lambda\mu\kappa\eta;\nu} = 0.$$

see also BIANCHI IDENTITIES (CONTRACTED), RIEMANN TENSOR

References
Weinberg, S. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity.* New York: Wiley, 1972.

Bianchi Identities (Contracted)

CONTRACTING λ with ν in the BIANCHI IDENTITIES

$$R_{\lambda\mu\nu\kappa;\eta} + R_{\lambda\mu\eta\nu;\kappa} + R_{\lambda\mu\kappa\eta;\nu} = 0 \tag{1}$$

gives

$$R_{\mu\kappa;\eta} - R_{\mu\eta;\kappa} + R^\nu{}_{\mu\kappa\eta;\nu} = 0. \tag{2}$$

CONTRACTING again,

$$R_{;\eta} - R^\mu{}_{\eta;\mu} - R^\nu{}_{\eta;\nu} = 0, \tag{3}$$

or

$$(R^\mu{}_\eta - \tfrac{1}{2}\delta^\mu{}_\eta R)_{;\mu} = 0, \tag{4}$$

or

$$(R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R)_{;\mu} = 0. \tag{5}$$

Bias (Estimator)

The bias of an ESTIMATOR $\tilde{\theta}$ is defined as

$$B(\tilde{\theta}) \equiv \langle\tilde{\theta}\rangle - \theta.$$

It is therefore true that

$$\tilde{\theta} - \theta = (\tilde{\theta} - \langle\tilde{\theta}\rangle) + (\langle\tilde{\theta}\rangle - \theta) = (\tilde{\theta} - \langle\tilde{\theta}\rangle) + B(\tilde{\theta}).$$

An ESTIMATOR for which $B = 0$ is said to be UNBIASED.

see also ESTIMATOR, UNBIASED

Bias (Series)

The bias of a SERIES is defined as

$$Q[a_i, a_{i+1}, a_{i+2}] \equiv \frac{a_i a_{i+2} - {a_{i+1}}^2}{a_i a_{i+1} a_{i+2}}.$$

A SERIES is GEOMETRIC IFF $Q = 0$. A SERIES is ARTISTIC IFF the bias is constant.

see also ARTISTIC SERIES, GEOMETRIC SERIES

References
Duffin, R. J. "On Seeing Progressions of Constant Cross Ratio." *Amer. Math. Monthly* **100**, 38–47, 1993.

Biased

An ESTIMATOR which exhibits BIAS.

Biaugmented Pentagonal Prism

see JOHNSON SOLID

Biaugmented Triangular Prism

see JOHNSON SOLID

Biaugmented Truncated Cube

see JOHNSON SOLID

BIBD

see BLOCK DESIGN

Bicentric Polygon

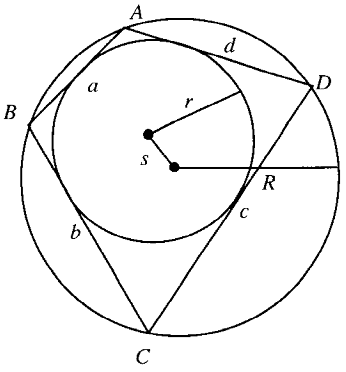

A POLYGON which has both a CIRCUMCIRCLE and an INCIRCLE, both of which touch all VERTICES. All TRIANGLES are bicentric with

$$R^2 - s^2 = 2Rr, \tag{1}$$

where R is the CIRCUMRADIUS, r is the INRADIUS, and s is the separation of centers. In 1798, N. Fuss characterized bicentric POLYGONS of $n = 4$, 5, 6, 7, and 8 sides. For bicentric QUADRILATERALS (FUSS'S PROBLEM), the CIRCLES satisfy

$$2r^2(R^2 - s^2) = (R^2 - s^2)^2 - 4r^2s^2 \tag{2}$$

(Dörrie 1965) and

$$r = \frac{\sqrt{abcd}}{s} \tag{3}$$

$$R = \frac{1}{4}\sqrt{\frac{(ac+bd)(ad+bc)(ab+cd)}{abcd}} \tag{4}$$

(Beyer 1987). In addition,

$$\frac{1}{(R-s)^2} + \frac{1}{(R+s)^2} = \frac{1}{r^2} \tag{5}$$

and

$$a + c = b + d. \tag{6}$$

The AREA of a bicentric quadrilateral is

$$A = \sqrt{abcd}. \tag{7}$$

If the circles permit successive tangents around the INCIRCLE which close the POLYGON for one starting point on the CIRCUMCIRCLE, then they do so for all points on the CIRCUMCIRCLE.

see also PONCELET'S CLOSURE THEOREM

References
Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 124, 1987.
Dörrie, H. "Fuss' Problem of the Chord-Tangent Quadrilateral." §39 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 188–193, 1965.

Bicentric Quadrilateral

A 4-sided BICENTRIC POLYGON, also called a CYCLIC-INSCRIPTABLE QUADRILATERAL.

References
Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 124, 1987.

Bichromatic Graph

A GRAPH with EDGES of two possible "colors," usually identified as red and blue. For a bichromatic graph with R red EDGES and B blue EDGES,

$$R + B \geq 2.$$

see also BLUE-EMPTY GRAPH, EXTREMAL COLORING, EXTREMAL GRAPH, MONOCHROMATIC FORCED TRIANGLE, RAMSEY NUMBER

Bicollared

A SUBSET $X \subset Y$ is said to be bicollared in Y if there exists an embedding $b : X \times [-1, 1] \to Y$ such that $b(x, 0) = x$ when $x \in X$. The MAP b or its image is then said to be the bicollar.

References
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 34–35, 1976.

Bicorn

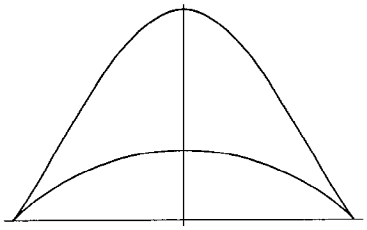

The bicorn is the name of a collection of QUARTIC CURVES studied by Sylvester in 1864 and Cayley in 1867 (MacTutor Archive). The bicorn is given by the parametric equations

$$x = a\sin t$$
$$y = \frac{a\cos^2 t(2+\cos t)}{3\sin^2 t}.$$

The graph is similar to that of the COCKED HAT CURVE.

References
Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 147–149, 1972.
MacTutor History of Mathematics Archive. "Bicorn." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Bicorn.html`.

Bicubic Spline

A bicubic spline is a special case of bicubic interpolation which uses an interpolation function of the form

$$y(x_1,x_2)=\sum_{i=1}^{4}\sum_{j=1}^{4}c_{ij}t^{i-1}u^{j-1}$$

$$y_{x_1}(x_1,x_2)=\sum_{i=1}^{4}\sum_{j=1}^{4}(i-1)c_{ij}t^{i-2}u^{j-1}$$

$$y_{x_2}(x_1,x_2)=\sum_{i=1}^{4}\sum_{j=1}^{4}(j-1)c_{ij}t^{i-1}u^{j-2}$$

$$y_{x_1x_2}=\sum_{i=1}^{4}\sum_{j=1}^{4}(i-1)(j-1)c_{ij}t^{i-2}u^{j-2},$$

where c_{ij} are constants and u and t are parameters ranging from 0 to 1. For a bicubic spline, however, the partial derivatives at the grid points are determined globally by 1-D SPLINES.

see also B-SPLINE, SPLINE

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 118–122, 1992.

Bicupola

Two adjoined CUPOLAS.

see also CUPOLA, ELONGATED GYROBICUPOLA, ELONGATED ORTHOBICUPOLA, GYROBICUPOLA, ORTHOBICUPOLA

Bicuspid Curve

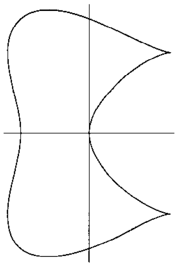

The PLANE CURVE given by the Cartesian equation

$$(x^2-a^2)(x-a)^2+(y^2-a^2)^2=0.$$

Bicylinder

see STEINMETZ SOLID

Bidiakis Cube

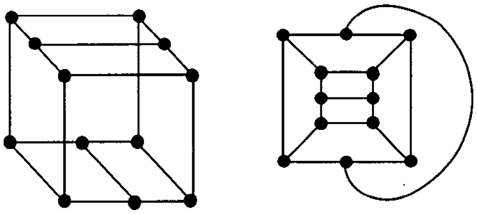

The 12-VERTEX graph consisting of a CUBE in which two opposite faces (say, top and bottom) have edges drawn across them which connect the centers of opposite sides of the faces in such a way that the orientation of the edges added on top and bottom are PERPENDICULAR to each other.

see also BISLIT CUBE, CUBE, CUBICAL GRAPH

Bieberbach Conjecture

The nth COEFFICIENT in the POWER series of a UNIVALENT FUNCTION should be no greater than n. In other words, if

$$f(z)=a_0+a_1z+a_2z^2+\ldots+a_nz^n+\ldots$$

is a conformal transformation of a unit disk on any domain, then$|a_n|\leq n|a_1|$. In more technical terms, "geometric extremality implies metric extremality." The conjecture had been proven for the first six terms (the cases $n=2$, 3, and 4 were done by Bieberbach, Lowner, and Shiffer and Garbedjian, respectively), was known to be false for only a finite number of indices (Hayman 1954), and true for a convex or symmetric domain (Le Lionnais 1983). The general case was proved by Louis de Branges (1985). De Branges proved the MILIN CONJECTURE, which established the ROBERTSON CONJECTURE, which in turn established the Bieberbach conjecture (Stewart 1996).

References

de Branges, L. "A Proof of the Bieberbach Conjecture." *Acta Math.* **154**, 137–152, 1985.

Hayman, W. K. *Multivalent Functions, 2nd ed.* Cambridge, England: Cambridge University Press, 1994.

Hayman, W. K. and Stewart, F. M. "Real Inequalities with Applications to Function Theory." *Proc. Cambridge Phil. Soc.* **50**, 250–260, 1954.

Kazarinoff, N. D. "Special Functions and the Bieberbach Conjecture." *Amer. Math. Monthly* **95**, 689–696, 1988.

Korevaar, J. "Ludwig Bieberbach's Conjecture and its Proof." *Amer. Math. Monthly* **93**, 505–513, 1986.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 53, 1983.

Pederson, R. N. "A Proof of the Bieberbach Conjecture for the Sixth Coefficient." *Arch. Rational Mech. Anal.* **31**, 331–351, 1968/1969.

Pederson, R. and Schiffer, M. "A Proof of the Bieberbach Conjecture for the Fifth Coefficient." *Arch. Rational Mech. Anal.* **45**, 161–193, 1972.

Stewart, I. "The Bieberbach Conjecture." In *From Here to Infinity: A Guide to Today's Mathematics.* Oxford, England: Oxford University Press, pp. 164–166, 1996.

Bienaymé-Chebyshev Inequality

see CHEBYSHEV INEQUALITY

Bifoliate

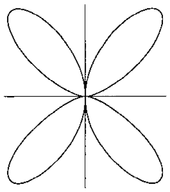

The PLANE CURVE given by the Cartesian equation

$$x^4 + y^4 = 2axy^2.$$

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 72, 1989.

Bifolium

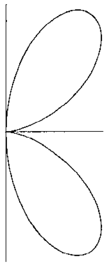

A FOLIUM with $b = 0$. The bifolium is the PEDAL CURVE of the DELTOID, where the PEDAL POINT is the MIDPOINT of one of the three curved sides. The Cartesian equation is

$$(x^2 + y^2)^2 = 4axy^2$$

and the POLAR equation is

$$r = 4a\sin^2\theta\cos\theta.$$

see also FOLIUM, QUADRIFOLIUM, TRIFOLIUM

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 152–153, 1972.

MacTutor History of Mathematics Archive. "Double Folium." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Double.html`.

Bifurcation

A period doubling, quadrupling, etc., that accompanies the onset of CHAOS. It represents the sudden appearance of a qualitatively different solution for a nonlinear system as some parameter is varied. Bifurcations come in four basic varieties: FLIP BIFURCATION, FOLD BIFURCATION, PITCHFORK BIFURCATION, and TRANSCRITICAL BIFURCATION (Rasband 1990).

see also CODIMENSION, FEIGENBAUM CONSTANT, FEIGENBAUM FUNCTION, FLIP BIFURCATION, HOPF BIFURCATION, LOGISTIC MAP, PERIOD DOUBLING, PITCHFORK BIFURCATION, TANGENT BIFURCATION, TRANSCRITICAL BIFURCATION

References

Guckenheimer, J. and Holmes, P. "Local Bifurcations." Ch. 3 in *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields, 2nd pr., rev. corr.* New York: Springer-Verlag, pp. 117–165, 1983.

Lichtenberg, A. J. and Lieberman, M. A. "Bifurcation Phenomena and Transition to Chaos in Dissipative Systems." Ch. 7 in *Regular and Chaotic Dynamics, 2nd ed.* New York: Springer-Verlag, pp. 457–569, 1992.

Rasband, S. N. "Asymptotic Sets and Bifurcations." §2.4 in *Chaotic Dynamics of Nonlinear Systems.* New York: Wiley, pp. 25–31, 1990.

Wiggins, S. "Local Bifurcations." Ch. 3 in *Introduction to Applied Nonlinear Dynamical Systems and Chaos.* New York: Springer-Verlag, pp. 253–419, 1990.

Bifurcation Theory

The study of the nature and properties of BIFURCATIONS.

see also CHAOS, DYNAMICAL SYSTEM

Bigraph

see BIPARTITE GRAPH

Bigyrate Diminished Rhombicosidodecahedron

see JOHNSON SOLID

Biharmonic Equation

The differential equation obtained by applying the BIHARMONIC OPERATOR and setting to zero.

$$\nabla^4\phi = 0. \tag{1}$$

In CARTESIAN COORDINATES, the biharmonic equation is

$$\begin{aligned}\nabla^4\phi &= \nabla^2(\nabla^2)\phi\\ &= \left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\phi\\ &= \frac{\partial^4\phi}{\partial x^4}+\frac{\partial^4\phi}{\partial y^4}+\frac{\partial^4\phi}{\partial z^4}+2\frac{\partial^4\phi}{\partial x^2\partial y^2}\\ &\quad +2\frac{\partial^4\phi}{\partial y^2\partial z^2}+2\frac{\partial^4\phi}{\partial x^2\partial z^2}=0. \end{aligned} \tag{2}$$

In POLAR COORDINATES (Kaplan 1984, p. 148)

$$\begin{aligned}\nabla^4\phi &= \phi_{rrrr}+\frac{2}{r^2}\phi_{rr\theta\theta}+\frac{1}{r^4}\phi_{\theta\theta\theta\theta}+\frac{2}{r}\phi_{rrr}\\ &\quad -\frac{2}{r^3}\phi_{r\theta\theta}-\frac{1}{r^2}\phi_{rr}+\frac{4}{r^4}\phi_{\theta\theta}+\frac{1}{r^3}\phi_r=0. \end{aligned} \tag{3}$$

For a radial function $\phi(r)$, the biharmonic equation becomes

$$\begin{aligned}\nabla^4\phi &= \frac{1}{r}\frac{d}{dr}\left\{r\frac{d}{dr}\left[\frac{1}{r}\frac{d}{dr}\left(r\frac{d\phi}{dr}\right)\right]\right\}\\ &= \phi_{rrrr} + \frac{2}{r}\phi_{rrr} - \frac{1}{r^2}\phi_{rr} + \frac{1}{r^3}\phi_r = 0.\end{aligned} \tag{4}$$

Writing the inhomogeneous equation as

$$\nabla^4\phi = 64\beta, \tag{5}$$

we have

$$64\beta r\,dr = d\left\{r\frac{d}{dr}\left[\frac{1}{r}\frac{d}{dr}\left(r\frac{d\phi}{dr}\right)\right]\right\} \tag{6}$$

$$32\beta r^2 + C_1 = r\frac{d}{dr}\left[\frac{1}{r}\frac{d}{dr}\left(r\frac{d\phi}{dr}\right)\right] \tag{7}$$

$$\left(32\beta r + \frac{C_1}{r}\right)dr = d\left[\frac{1}{r}\frac{d}{dr}\left(r\frac{d\phi}{dr}\right)\right] \tag{8}$$

$$16\beta r^2 + C_1\ln r + C_2 = \frac{1}{r}\frac{d}{dr}\left(r\frac{d\phi}{dr}\right) \tag{9}$$

$$(16\beta r^3 + C_1 r\ln r + C_2 r)\,dr = d\left(r\frac{d\phi}{dr}\right). \tag{10}$$

Now use

$$\int r\ln r\,dr = \tfrac{1}{2}r^2\ln r - \tfrac{1}{4}r^2 \tag{11}$$

to obtain

$$4\beta r^4 + C_1(\tfrac{1}{2}r^2\ln r - \tfrac{1}{4}r^2) + \tfrac{1}{2}C_2r^2 + C_3 = r\frac{d\phi}{dr} \tag{12}$$

$$\left(4\beta r^3 + C_1' r\ln r + C_2' r + \frac{C_3}{r}\right)dr = d\phi \tag{13}$$

$$\begin{aligned}\phi(r) &= \beta r^4 + C_1'\left(\tfrac{1}{2}r^2\ln r - \tfrac{1}{4}r^2\right)\\ &\quad + \tfrac{1}{2}C_2'r^2 + C_3\ln r + C_4\\ &= \beta r^4 + ar^2 + b + (cr^2 + d)\ln\left(\frac{r}{R}\right).\end{aligned} \tag{14}$$

The homogeneous biharmonic equation can be separated and solved in 2-D BIPOLAR COORDINATES.

References

Kaplan, W. *Advanced Calculus, 4th ed.* Reading, MA: Addison-Wesley, 1991.

Biharmonic Operator

Also known as the BILAPLACIAN.

$$\nabla^4 \equiv (\nabla^2)^2.$$

In n-D space,

$$\nabla^4\left(\frac{1}{r}\right) = \frac{3(15 - 8n + n^2)}{r^5}.$$

see also BIHARMONIC EQUATION

Bijection

A transformation which is ONE-TO-ONE and ONTO.

see also ONE-TO-ONE, ONTO, PERMUTATION

Bilaplacian

see BIHARMONIC OPERATOR

Bilinear

A function of two variables is bilinear if it is linear with respect to each of its variables. The simplest example is $f(x,y) = xy$.

Bilinear Basis

A bilinear basis is a BASIS, which satisfies the conditions

$$(a\mathbf{x} + b\mathbf{y})\cdot\mathbf{z} = a(\mathbf{x}\cdot\mathbf{z}) + b(\mathbf{y}\cdot\mathbf{z})$$

$$\mathbf{z}\cdot(a\mathbf{x} + b\mathbf{y}) = a(\mathbf{z}\cdot\mathbf{x}) + b(\mathbf{z}\cdot\mathbf{y}).$$

see also BASIS

Billiard Table Problem

Given a billiard table with only corner pockets and sides of INTEGER lengths m and n, a ball sent at a 45° angle from a corner will be pocketed in a corner after $m+n-2$ bounces.

see also ALHAZEN'S BILLIARD PROBLEM, BILLIARDS

Billiards

The game of billiards is played on a RECTANGULAR table (known as a billiard table) upon which balls are placed. One ball (the "cue ball") is then struck with the end of a "cue" stick, causing it to bounce into other balls and REFLECT off the sides of the table. Real billiards can involve spinning the ball so that it does not travel in a straight LINE, but the mathematical study of billiards generally consists of REFLECTIONS in which the reflection and incidence angles are the same. However, strange table shapes such as CIRCLES and ELLIPSES are often considered. Many interesting problems can arise.

For example, ALHAZEN'S BILLIARD PROBLEM seeks to find the point at the edge of a circular "billiards" table at which a cue ball at a given point must be aimed in order to carom once off the edge of the table and strike another ball at a second given point. It was not until 1997 that Neumann proved that the problem is insoluble using a COMPASS and RULER construction.

On an ELLIPTICAL billiard table, the ENVELOPE of a trajectory is a smaller ELLIPSE, a HYPERBOLA, a LINE through the FOCI of the ELLIPSE, or periodic curve (e.g., DIAMOND-shape) (Wagon 1991).

see also ALHAZEN'S BILLIARD PROBLEM, BILLIARD TABLE PROBLEM, REFLECTION PROPERTY

References

Davis, D.; Ewing, C.; He, Z.; and Shen, T. "The Billiards Simulation." `http://serendip.brynmawr.edu/chaos/home.html`.

Dullin, H. R.; Richter, P.H.; and Wittek, A. "A Two-Parameter Study of the Extent of Chaos in a Billiard System." *Chaos* **6**, 43–58, 1996.

Madachy, J. S. "Bouncing Billiard Balls." In *Madachy's Mathematical Recreations.* New York: Dover, pp. 231–241, 1979.

Neumann, P. Submitted to *Amer. Math. Monthly.*

Pappas, T. "Mathematics of the Billiard Table." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 43, 1989.

Peterson, I. "Billiards in the Round." `http://www.sciencenews.org/sn_arc97/3_1_97/mathland.htm`.

Wagon, S. "Billiard Paths on Elliptical Tables." §10.2 in *Mathematica in Action.* New York: W. H. Freeman, pp. 330–333, 1991.

Billion

The word billion denotes different numbers in American and British usage. In the American system, one billion equals 10^9. In the British, French, and German systems, one billion equals 10^{12}.

see also LARGE NUMBER, MILLIARD, MILLION, TRILLION

Bilunabirotunda

see JOHNSON SOLID

Bimagic Square

16	41	36	5	27	62	55	18
26	63	54	19	13	44	33	8
1	40	45	12	22	51	58	31
23	50	59	30	4	37	48	9
38	3	10	47	49	24	29	60
52	21	32	57	39	2	11	46
43	14	7	34	64	25	20	53
61	28	17	56	42	15	6	35

If replacing each number by its square in a MAGIC SQUARE produces another MAGIC SQUARE, the square is said to be a bimagic square. The first bimagic square (shown above) has order 8 with magic constant 260 for addition and 11,180 after squaring. Bimagic squares are also called DOUBLY MAGIC SQUARES, and are 2-MULTIMAGIC SQUARES.

see also MAGIC SQUARE, MULTIMAGIC SQUARE, TRIMAGIC SQUARE

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 212, 1987.

Hunter, J. A. H. and Madachy, J. S. "Mystic Arrays." Ch. 3 in *Mathematical Diversions.* New York: Dover, p. 31, 1975.

Kraitchik, M. "Multimagic Squares." §7.10 in *Mathematical Recreations.* New York: W. W. Norton, pp. 176–178, 1942.

Bimedian

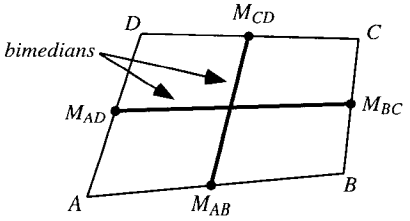

A LINE SEGMENT joining the MIDPOINTS of opposite sides of a QUADRILATERAL.

see also MEDIAN (TRIANGLE), VARIGNON'S THEOREM

Bimodal Distribution

A DISTRIBUTION having two separated peaks.

see also UNIMODAL DISTRIBUTION

Bimonster

The wreathed product of the MONSTER GROUP by Z_2. The bimonster is a quotient of the COXETER GROUP with the following COXETER-DYNKIN DIAGRAM.

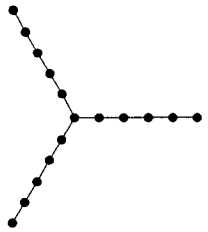

This had been conjectured by Conway, but was proven around 1990 by Ivanov and Norton. If the parameters p, q, r in Coxeter's NOTATION $[3^{p,q,r}]$ are written side by side, the bimonster can be denoted by the BEAST NUMBER 666.

Bin

An interval into which a given data point does or does not fall.

see also HISTOGRAM

Binary

The BASE 2 method of counting in which only the digits 0 and 1 are used. In this BASE, the number 1011 equals $1\cdot2^0+1\cdot2+0\cdot2^2+1\cdot2^3=11$. This BASE is used in computers, since all numbers can be simply represented as a string of electrically pulsed ons and offs. A NEGATIVE $-n$ is most commonly represented as the complement of the POSITIVE number $n-1$, so $-11=00001011_2$ would be written as the complement of $10=00001010_2$, or 11110101. This allows addition to be carried out with the usual carrying and the left-most digit discarded, so $17-11=6$ gives

$$\begin{array}{lr} 00010001 & 17 \\ \underline{11110101} & \underline{-11} \\ 00000110 & 6 \end{array}$$

The number of times k a given binary number $b_n \cdots b_2 b_1 b_0$ is divisible by 2 is given by the position of the first $b_k = 1$ counting from the right. For example, $12 = 1100$ is divisible by 2 twice, and $13 = 1101$ is divisible by 2 0 times.

Unfortunately, the storage of binary numbers in computers is not entirely standardized. Because computers store information in 8-bit bytes (where a bit is a single binary digit), depending on the "word size" of the machine, numbers requiring more than 8 bits must be stored in multiple bytes. The usual `FORTRAN77` integer size is 4 bytes long. However, a number represented as (byte1 byte2 byte3 byte4) in a VAX would be read and interpreted as (byte4 byte3 byte2 byte1) on a Sun. The situation is even worse for floating point (real) numbers, which are represented in binary as a MANTISSA and CHARACTERISTIC, and worse still for long (8-byte) reals!

Binary multiplication of single bit numbers (0 or 1) is equivalent to the AND operation, as can be seen in the following MULTIPLICATION TABLE.

×	0	1
0	0	0
1	0	1

see also BASE (NUMBER), DECIMAL, HEXADECIMAL, OCTAL, QUATERNARY, TERNARY

References
Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 6–9, 1991.
Pappas, T. "Computers, Counting, & Electricity." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 24–25, 1989.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Error, Accuracy, and Stability" and "Diagnosing Machine Parameters." §1.2 and §20.1 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 18–21, 276, and 881–886, 1992.
Weisstein, E. W. "Bases." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Bases.m`.

Binary Bracketing

A binary bracketing is a BRACKETING built up entirely of binary operations. The number of binary bracketings of n letters (CATALAN'S PROBLEM) are given by the CATALAN NUMBERS C_{n-1}, where

$$C_n \equiv \frac{1}{n+1}\binom{2n}{n} = \frac{1}{n+1}\frac{(2n)!}{n!^2} = \frac{(2n)!}{(n+1)!n!},$$

where $\binom{2n}{n}$ denotes a BINOMIAL COEFFICIENT and $n!$ is the usual FACTORIAL, as first shown by Catalan in 1838. For example, for the four letters a, b, c, and d there are five possibilities: $((ab)c)d$, $(a(bc))d$, $(ab)(cd)$, $a((bc)d)$, and $a(b(cd))$, written in shorthand as $((xx)x)x$, $(x(xx))x$, $(xx)(xx)$, $x((xx)x)$, and $x(x(xx))$.

see also BRACKETING, CATALAN NUMBER, CATALAN'S PROBLEM

References
Schröder, E. "Vier combinatorische Probleme." *Z. Math. Physik* **15**, 361–376, 1870.
Sloane, N. J. A. Sequences A000108/M1459 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Sloane, N. J. A. and Plouffe, S. Extended entry in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.
Stanley, R. P. "Hipparchus, Plutarch, Schröder, and Hough." *Amer. Math. Monthly* **104**, 344–350, 1997.

Binary Operator

An OPERATOR which takes two mathematical objects as input and returns a value is called a binary operator. Binary operators are called compositions by Rosenfeld (1968). Sets possessing a binary multiplication operation include the GROUP, GROUPOID, MONOID, QUASIGROUP, and SEMIGROUP. Sets possessing both a binary multiplication and a binary addition operation include the DIVISION ALGEBRA, FIELD, RING, RINGOID, SEMIRING, and UNIT RING.

see also AND, BOOLEAN ALGEBRA, CLOSURE, DIVISION ALGEBRA, FIELD, GROUP, GROUPOID, MONOID, OPERATOR, OR, MONOID, NOT, QUASIGROUP, RING, RINGOID, SEMIGROUP, SEMIRING, XOR, UNIT RING

References
Rosenfeld, A. *An Introduction to Algebraic Structures.* New York: Holden-Day, 1968.

Binary Quadratic Form

A 2-variable QUADRATIC FORM of the form

$$Q(x,y) = a_{11}x^2 + 2a_{12}xy + a_{22}y^2.$$

see also QUADRATIC FORM, QUADRATIC INVARIANT

Binary Remainder Method

An ALGORITHM for computing a UNIT FRACTION (Stewart 1992).

References
Stewart, I. "The Riddle of the Vanishing Camel." *Sci. Amer.* **266**, 122–124, June 1992.

Binary Tree

A TREE with two BRANCHES at each FORK and with one or two LEAVES at the end of each BRANCH. (This definition corresponds to what is sometimes known as an "extended" binary tree.) The height of a binary tree is the number of levels within the TREE. For a binary tree of height H with n nodes,

$$H \leq n \leq 2^H - 1.$$

These extremes correspond to a balanced tree (each node except the LEAVES has a left and right CHILD, and all LEAVES are at the same level) and a degenerate tree (each node has only one outgoing BRANCH), respectively. For a search of data organized into a binary tree, the number of search steps $S(n)$ needed to find an item is bounded by

$$\lg n \leq S(n) \leq n.$$

Partial balancing of an arbitrary tree into a so-called AVL binary search tree can improve search speed.

The number of binary trees with n internal nodes is the CATALAN NUMBER C_n (Sloane's A000108), and the number of binary trees of height b is given by Sloane's A001699.

see also B-TREE, QUADTREE, QUATERNARY TREE, RED-BLACK TREE, STERN-BROCOT TREE, WEAKLY BINARY TREE

References
Lucas, J.; Roelants van Baronaigien, D.; and Ruskey, F. "Generating Binary Trees by Rotations." *J. Algorithms* **15**, 343–366, 1993.
Ranum, D. L. "On Some Applications of Fibonacci Numbers." *Amer. Math. Monthly* **102**, 640–645, 1995.
Ruskey, F. "Information on Binary Trees." `http://sue.csc.uvic.ca/~cos/inf/tree/BinaryTrees.html`.
Ruskey, F. and Proskurowski, A. "Generating Binary Trees by Transpositions." *J. Algorithms* **11**, 68–84, 1990.
Sloane, N. J. A. Sequences A000108/M1459 and A001699/M3087 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Binet Forms

The two RECURRENCE SEQUENCES

$$U_n = mU_{n-1} + U_{n-2} \tag{1}$$
$$V_n = mV_{n-1} + V_{n-2} \tag{2}$$

with $U_0 = 0$, $U_1 = 1$ and $V_0 = 2$, $V_1 = m$, can be solved for the individual U_n and V_n. They are given by

$$U_n = \frac{\alpha^n - \beta^n}{\Delta} \tag{3}$$
$$V_n = \alpha^n + \beta^n, \tag{4}$$

where

$$\Delta \equiv \sqrt{m^2 + 4} \tag{5}$$
$$\alpha \equiv \frac{m + \Delta}{2} \tag{6}$$
$$\beta \equiv \frac{m - \Delta}{2}. \tag{7}$$

A useful related identity is

$$U_{n-1} + U_{n+1} = V_n. \tag{8}$$

BINET'S FORMULA is a special case of the Binet form for U_n corresponding to $m = 1$.

see also FIBONACCI Q-MATRIX

Binet's Formula

A special case of the U_n BINET FORM with $m = 0$, corresponding to the nth FIBONACCI NUMBER,

$$F_n = \frac{(1+\sqrt{5}\,)^n - (1-\sqrt{5}\,)^n}{2^n\sqrt{5}}.$$

It was derived by Binet in 1843, although the result was known to Euler and Daniel Bernoulli more than a century earlier.

see also BINET FORMS, FIBONACCI NUMBER

Bing's Theorem

If M^3 is a closed oriented connected 3-MANIFOLD such that every simple closed curve in M lies interior to a BALL in M, then M is HOMEOMORPHIC with the HYPERSPHERE, $\mathbb{S}^3$.

see also BALL, HYPERSPHERE

References
Bing, R. H. "Necessary and Sufficient Conditions that a 3-Manifold be S^3." *Ann. Math.* **68**, 17–37, 1958.
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 251–257, 1976.

Binomial

A POLYNOMIAL with 2 terms.

see also MONOMIAL, POLYNOMIAL, TRINOMIAL

Binomial Coefficient

The number of ways of picking n unordered outcomes from N possibilities. Also known as a COMBINATION. The binomial coefficients form the rows of PASCAL'S TRIANGLE. The symbols ${}_NC_n$ and

$$\binom{N}{n} \equiv \frac{N!}{(N-n)!n!} \tag{1}$$

are used, where the latter is sometimes known as N CHOOSE n. The number of LATTICE PATHS from the ORIGIN $(0,0)$ to a point (a,b) is the BINOMIAL COEFFICIENT $\binom{a+b}{a}$ (Hilton and Pedersen 1991).

For POSITIVE integer n, the BINOMIAL THEOREM gives

$$(x+a)^n = \sum_{k=0}^{n} \binom{n}{k} x^k a^{n-k}. \tag{2}$$

The FINITE DIFFERENCE analog of this identity is known as the CHU-VANDERMONDE IDENTITY. A similar formula holds for NEGATIVE INTEGRAL n,

$$(x+a)^{-n} = \sum_{k=0}^{\infty} \binom{-n}{k} x^k a^{-n-k}. \tag{3}$$

A general identity is given by

$$\frac{(a+b)^n}{a} = \sum_{j=0}^{n} \binom{n}{j} (a-jc)^{j-1}(b+jc)^{n-j} \tag{4}$$

(Prudnikov *et al.* 1986), which gives the BINOMIAL THEOREM as a special case with $c = 0$.

The binomial coefficients satisfy the identities:

$$\binom{n}{0} = \binom{n}{n} = 1 \tag{5}$$

$$\binom{n}{k} = \binom{n}{n-k} = (-1)^k \binom{k-n-1}{k} \tag{6}$$

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}. \tag{7}$$

Sums of powers include

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n \tag{8}$$

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0 \tag{9}$$

$$\sum_{k=0}^{n} \binom{n}{k} r^k = (1+r)^n \tag{10}$$

(the BINOMIAL THEOREM), and

$$\begin{aligned} \sum_{n=0}^{\infty} \binom{2n+s}{n} x^n &= {}_2F_1(\tfrac{1}{2}(s+1), \tfrac{1}{2}(s+2); s+1, 4x) \\ &= \frac{2^s}{(\sqrt{1-4x}+1)^s \sqrt{1-4x}}, \end{aligned} \tag{11}$$

where ${}_2F_1(a, b; c; z)$ is a HYPERGEOMETRIC FUNCTION (Abramowitz and Stegun 1972, p. 555; Graham *et al.* 1994, p. 203). For NONNEGATIVE INTEGERS n and r with $r \leq n+1$,

$$\begin{aligned} \sum_{k=0}^{n} \frac{(-1)^k}{k+1} \binom{n}{k} \Bigg[&\sum_{j=0}^{r-1} (-1)^j \binom{n}{j} (r-j)^{n-k} \\ &+ \sum_{j=0}^{n-r} (-1)^j \binom{n}{j} (n+1-r-j)^{n-k} \Bigg] = n!. \end{aligned} \tag{12}$$

Taking $n = 2r - 1$ gives

$$\sum_{k=0}^{n} \frac{(-1)^k}{k+1} \binom{n}{k} \sum_{j=0}^{r-1} \binom{n}{j} (r-j)^{n-k} = \tfrac{1}{2} n!. \tag{13}$$

Another identity is

$$\sum_{k=0}^{n} \binom{n+k}{k} [x^{n+1}(1-x)^k + (1-x)^{n+1} x^k] = 1 \tag{14}$$

(Beeler *et al.* 1972, Item 42).

RECURRENCE RELATIONS of the sums

$$s_p \equiv \sum_{k=0}^{n} \binom{n}{k}^p \tag{15}$$

are given by

$$2s_1(n) - s_1(n+1) = 0 \tag{16}$$

$$-2(2n+1)s_2(n) + (n+1)s_2(n) = 0 \tag{17}$$

$$\begin{aligned} -8(n+1)^2 s_3(n) + (-16 - 21n - 7n^2) s_3(n+1) \\ + (n+2)^2 s_3(n+2) = 0 \end{aligned} \tag{18}$$

$$\begin{aligned} -4(n+1)(4n+3)(4n+5)s_4(n) \\ -2(2n+3)(3n^2+9n+7)s_4(n+1) \\ +(n+2)^3 s_4(n+2) = 0. \end{aligned} \tag{19}$$

This sequence for s_3 cannot be expressed as a fixed number of hypergeometric terms (Petkovšek *et al.* 1996, p. 160).

A fascinating series of identities involving binomial coefficients times small powers are

$$\sum_{n=1}^{\infty} \frac{1}{\binom{2n}{n}} = \tfrac{1}{27}(2\pi\sqrt{3} + 9) = 0.7363998587\ldots \tag{20}$$

$$\sum_{n=1}^{\infty} \frac{1}{n\binom{2n}{n}} = \tfrac{1}{9}\pi\sqrt{3} = 0.6045997881\ldots \tag{21}$$

$$\sum_{n=1}^{\infty} \frac{1}{n^2\binom{2n}{n}} = \tfrac{1}{3}\zeta(2) = \tfrac{1}{8}\pi^2 \tag{22}$$

$$\sum_{n=1}^{\infty} \frac{1}{n^4\binom{2n}{n}} = \tfrac{17}{36}\zeta(4) = \tfrac{17}{3240}\pi^4 \tag{23}$$

(Comtet 1974, p. 89) and

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^3\binom{2n}{n}} = \tfrac{2}{5}\zeta(3), \tag{24}$$

where $\zeta(z)$ is the RIEMANN ZETA FUNCTION (Le Lionnais 1983, pp. 29, 30, 41, 36, and 35; Guy 1994, p. 257).

As shown by Kummer in 1852, the exact POWER of p dividing $\binom{a+b}{a}$ is equal to

$$\epsilon_0 + \epsilon_1 + \ldots + \epsilon_t, \tag{25}$$

where this is the number of carries in performing the addition of a and b written in base b (Graham *et al.* 1989, Exercise 5.36; Ribenboim 1989; Vardi 1991, p. 68). Kummer's result can also be stated in the form that the exponent of a PRIME p dividing $\binom{n}{m}$ is given by the number of integers $j \geq 0$ for which

$$\text{frac}(m/p^j) > \text{frac}(n/p^j), \tag{26}$$

where $\text{frac}(x)$ denotes the FRACTIONAL PART of x. This inequality may be reduced to the study of the exponential sums $\sum_n \Lambda(n)e(x/n)$, where $\Lambda(n)$ is the MANGOLDT FUNCTION. Estimates of these sums are given by Jutila (1974, 1975), but recent improvements have been made by Granville and Ramare (1996).

R. W. Gosper showed that

$$f(n) = \binom{n-1}{\frac{1}{2}(n-1)} \equiv (-1)^{(n-1)/2} \pmod{n} \tag{27}$$

for all PRIMES, and conjectured that it holds *only* for PRIMES. This was disproved when Skiena (1990) found it also holds for the COMPOSITE NUMBER $n = 3 \times 11 \times 179$. Vardi (1991, p. 63) subsequently showed that $n = p^2$ is a solution whenever p is a WIEFERICH PRIME and that if $n = p^k$ with $k > 3$ is a solution, then so is $n = p^{k-1}$. This allowed him to show that the only solutions for COMPOSITE $n < 1.3 \times 10^7$ are 5907, 1093^2, and 3511^2, where 1093 and 3511 are WIEFERICH PRIMES.

Consider the binomial coefficients $\binom{2n-1}{n}$, the first few of which are 1, 3, 10, 35, 126, ... (Sloane's A001700). The GENERATING FUNCTION is

$$\frac{1}{2}\left[\frac{1}{\sqrt{1-4x}} - 1\right] = x + 3x^2 + 10x^3 + 35x^4 + \ldots. \tag{28}$$

These numbers are SQUAREFREE only for $n = 2$, 3, 4, 6, 9, 10, 12, 36, ... (Sloane's A046097), with no others less than $n = 10{,}000$. Erdős showed that the binomial coefficient $\binom{n}{k}$ is never a POWER of an INTEGER for $n \geq 3$ where $k \neq 0$, 1, $n-1$, and n (Le Lionnais 1983, p. 48).

The binomial coefficients $\binom{n}{\lfloor n/2 \rfloor}$ are called CENTRAL BINOMIAL COEFFICIENTS, where $\lfloor x \rfloor$ is the FLOOR FUNCTION, although the subset of coefficients $\binom{2n}{n}$ is sometimes also given this name. Erdős and Graham (1980, p. 71) conjectured that the CENTRAL BINOMIAL COEFFICIENT $\binom{2n}{n}$ is *never* SQUAREFREE for $n > 4$, and this is sometimes known as the ERDŐS SQUAREFREE CONJECTURE. SÁRKÖZY'S THEOREM (Sárközy 1985) provides a partial solution which states that the BINOMIAL COEFFICIENT $\binom{2n}{n}$ is never SQUAREFREE for all sufficiently large $n \geq n_0$ (Vardi 1991). Granville and Ramare (1996) proved that the *only* SQUAREFREE values are $n = 2$ and 4. Sander (1992) subsequently showed that $\binom{2n \pm d}{n}$ are also never SQUAREFREE for sufficiently large n as long as d is not "too big."

For p, q, and r distinct PRIMES, then the above function satisfies

$$f(pqr)f(p)f(q)f(r) \equiv f(pq)f(pr)p(qr) \pmod{pqr} \tag{29}$$

(Vardi 1991, p. 66).

The binomial coefficient $\binom{m}{n}$ mod 2 can be computed using the XOR operation n XOR m, making PASCAL'S TRIANGLE mod 2 very easy to construct.

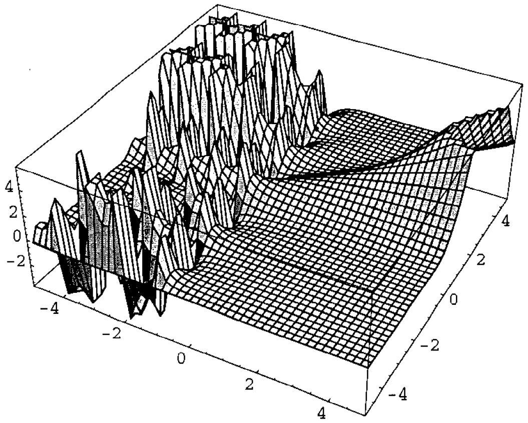

The binomial coefficient "function" can be defined as

$$C(x,y) \equiv \frac{x!}{y!(x-y)!} \tag{30}$$

(Fowler 1996), shown above. It has a very complicated GRAPH for NEGATIVE x and y which is difficult to render using standard plotting programs.

see also BALLOT PROBLEM, BINOMIAL DISTRIBUTION, BINOMIAL THEOREM, CENTRAL BINOMIAL COEFFICIENT, CHU-VANDERMONDE IDENTITY, COMBINATION, DEFICIENCY, ERDŐS SQUAREFREE CONJECTURE, GAUSSIAN COEFFICIENT, GAUSSIAN POLYNOMIAL, KINGS PROBLEM, MULTINOMIAL COEFFICIENT, PERMUTATION, ROMAN COEFFICIENT, SÁRKÖZY'S THEOREM, STREHL IDENTITY, WOLSTENHOLME'S THEOREM

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Binomial Coefficients." §24.1.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 10 and 822–823, 1972.

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Comtet, L. *Advanced Combinatorics.* Amsterdam, Netherlands: Kluwer, 1974.

Conway, J. H. and Guy, R. K. In *The Book of Numbers.* New York: Springer-Verlag, pp. 66–74, 1996.

Erdős, P.; Graham, R. L.; Nathanson, M. B.; and Jia, X. *Old and New Problems and Results in Combinatorial Number Theory.* New York: Springer-Verlag, 1998.

Fowler, D. "The Binomial Coefficient Function." *Amer. Math. Monthly* **103**, 1–17, 1996.

Graham, R. L.; Knuth, D. E.; and Patashnik, O. "Binomial Coefficients." Ch. 5 in *Concrete Mathematics: A Foundation for Computer Science.* Reading, MA: Addison-Wesley, pp. 153–242, 1990.

Granville, A. and Ramare, O. "Explicit Bounds on Exponential Sums and the Scarcity of Squarefree Binomial Coefficients." *Mathematika* **43**, 73–107, 1996.

Guy, R. K. "Binomial Coefficients," "Largest Divisor of a Binomial Coefficient," and "Series Associated with the ζ-Function." §B31, B33, and F17 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 84–85, 87–89, and 257–258, 1994.

Harborth, H. "Number of Odd Binomial Coefficients." *Not. Amer. Math. Soc.* **23**, 4, 1976.

Hilton, P. and Pedersen, J. "Catalan Numbers, Their Generalization, and Their Uses." *Math. Intel.* **13**, 64–75, 1991.

Jutila, M. "On Numbers with a Large Prime Factor." *J. Indian Math. Soc.* **37**, 43–53, 1973.

Jutila, M. "On Numbers with a Large Prime Factor. II." *J. Indian Math. Soc.* **38**, 125–130, 1974.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, 1983.

Ogilvy, C. S. "The Binomial Coefficients." *Amer. Math. Monthly* **57**, 551–552, 1950.

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, 1996.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Gamma Function, Beta Function, Factorials, Binomial Coefficients." §6.1 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 206–209, 1992.

Prudnikov, A. P.; Marichev, O. I.; and Brychkow, Yu. A. Formula 41 in *Integrals and Series, Vol. 1: Elementary Functions.* Newark, NJ: Gordon & Breach, p. 611, 1986.

Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, pp. 23–24, 1989.

Riordan, J. "Inverse Relations and Combinatorial Identities." *Amer. Math. Monthly* **71**, 485–498, 1964.

Sander, J. W. "On Prime Divisors of Binomial Coefficients." *Bull. London Math. Soc.* **24**, 140–142, 1992.

Sárközy, A. "On the Divisors of Binomial Coefficients, I." *J. Number Th.* **20**, 70–80, 1985.

Skiena, S. *Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica.* Reading, MA: Addison-Wesley, p. 262, 1990.

Sloane, N. J. A. Sequences A046097 and A001700/M2848 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Spanier, J. and Oldham, K. B. "The Binomial Coefficients $\binom{\nu}{m}$." Ch. 6 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 43–52, 1987.

Sved, M. "Counting and Recounting." *Math. Intel.* **5**, 21–26, 1983.

Vardi, I. "Application to Binomial Coefficients," "Binomial Coefficients," "A Class of Solutions," "Computing Binomial Coefficients," and "Binomials Modulo and Integer." §2.2, 4.1, 4.2, 4.3, and 4.4 in *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 25–28 and 63–71, 1991.

Wolfram, S. "Geometry of Binomial Coefficients." *Amer. Math. Monthly* **91**, 566–571, 1984.

Binomial Distribution

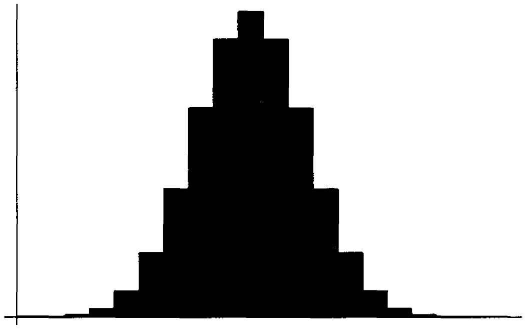

The probability of n successes in N BERNOULLI TRIALS is

$$P(n|N) = \binom{N}{n} p^n (1-p)^{N-n} = \frac{N!}{n!(N-n)!} p^n q^{N-n}. \tag{1}$$

The probability of obtaining more successes than the n observed is

$$P = \sum_{k=n+1}^{N} \binom{N}{k} p^k (1-p)^{N-k} = I_p(n+1, N-N), \tag{2}$$

where

$$I_x(a,b) \equiv \frac{B(x;a,b)}{B(a,b)}, \tag{3}$$

$B(a,b)$ is the BETA FUNCTION, and $B(x;a,b)$ is the incomplete BETA FUNCTION. The CHARACTERISTIC FUNCTION is

$$\phi(t) = (q + pe^{it})^n. \tag{4}$$

The MOMENT-GENERATING FUNCTION M for the distribution is

$$\begin{aligned} M(t) = \langle e^{tn} \rangle &= \sum_{n=0}^{N} e^{tn} \binom{N}{n} p^n q^{N-n} \\ &= \sum_{n=0}^{N} \binom{N}{n} (pe^t)^n (1-p)^{N-n} \\ &= [pe^t + (1-p)]^N \qquad (5) \\ M'(t) &= N[pe^t + (1-p)]^{N-1}(pe^t) \qquad (6) \\ M''(t) &= N(N-1)[pe^t + (1-p)]^{N-2}(pe^t)^2 \\ &\quad + N[pe^t + (1-p)]^{N-1}(pe^t). \qquad (7) \end{aligned}$$

The MEAN is

$$\mu = M'(0) = N(p + 1 - p)p = Np. \tag{8}$$

The MOMENTS about 0 are

$$\begin{aligned} \mu'_1 &= \mu = Np \qquad (9) \\ \mu'_2 &= Np(1 - p + Np) \qquad (10) \\ \mu'_3 &= Np(1 - 3p + 3Np + 2p^2 - 3NP^2 + N^2p^2) \qquad (11) \\ \mu'_4 &= Np(1 - 7p + 7Np + 12p^2 - 18Np^2 + 6N^2p^2 \\ &\quad - 6p^3 + 11Np^3 - 6N^2p^3 + N^3p^3), \qquad (12) \end{aligned}$$

so the MOMENTS about the MEAN are

$$\begin{aligned} \mu_2 = \sigma^2 &= [N(N-1)p^2 + Np] - (Np)^2 \\ &= N^2p^2 - Np^2 + Np - N^2p^2 \\ &= Np(1-p) = Npq \qquad (13) \\ \mu_3 &= \mu'_3 - 3\mu'_2\mu'_1 + 2(\mu_1)^3 \\ &= Np(1-p)(1-2p) \qquad (14) \\ \mu_4 &= \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2(\mu'_1)^2 - 3(\mu_1)^4 \\ &= Np(1-p)[3p^2(2-N) + 3p(N-2) + 1]. \qquad (15) \end{aligned}$$

The SKEWNESS and KURTOSIS are

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{Np(1-p)(1-2p)}{[Np(1-p)]^{3/2}} = \frac{1-2p}{\sqrt{Np(1-p)}} = \frac{q-p}{\sqrt{Npq}} \tag{16}$$

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{6p^2-6p+1}{Np(1-p)} = \frac{1-6pq}{Npq}. \tag{17}$$

An approximation to the Bernoulli distribution for large N can be obtained by expanding about the value $\tilde{n}$ where $P(n)$ is a maximum, i.e., where $dP/dn = 0$. Since the LOGARITHM function is MONOTONIC, we can instead choose to expand the LOGARITHM. Let $n \equiv \tilde{n} + \eta$, then

$$\ln[P(n)] = \ln[P(\tilde{n})] + B_1\eta + \tfrac{1}{2}B_2\eta^2 + \tfrac{1}{3!}B_3\eta^3 + \ldots, \tag{18}$$

where

$$B_k \equiv \left[\frac{d^k \ln[P(n)]}{dn^k}\right]_{n=\tilde{n}}. \tag{19}$$

But we are expanding about the maximum, so, by definition,

$$B_1 = \left[\frac{d\ln[P(n)]}{dn}\right]_{n=\tilde{n}} = 0. \tag{20}$$

This also means that B_2 is negative, so we can write $B_2 = -|B_2|$. Now, taking the LOGARITHM of (1) gives

$$\ln[P(n)] = \ln N! - \ln n! - \ln(N-n)! + n\ln p + (N-n)\ln q. \tag{21}$$

For large n and $N-n$ we can use STIRLING'S APPROXIMATION

$$\ln(n!) \approx n\ln n - n, \tag{22}$$

so

$$\frac{d[\ln(n!)]}{dn} \approx (\ln n + 1) - 1 = \ln n \tag{23}$$

$$\begin{aligned}\frac{d[\ln(N-n)!]}{dn} &\approx \frac{d}{dn}[(N-n)\ln(N-n) - (N-n)] \\ &= \left[-\ln(N-n) + (N-n)\frac{-1}{N-n} + 1\right] \\ &= -\ln(N-n),\end{aligned} \tag{24}$$

and

$$\frac{d\ln[P(n)]}{dn} \approx -\ln n + \ln(N-n) + \ln p - \ln q. \tag{25}$$

To find $\tilde{n}$, set this expression to 0 and solve for n,

$$\ln\left(\frac{N-\tilde{n}}{\tilde{n}}\frac{p}{q}\right) = 0 \tag{26}$$

$$\frac{N-\tilde{n}}{\tilde{n}}\frac{p}{q} = 1 \tag{27}$$

$$(N-\tilde{n})p = \tilde{n}q \tag{28}$$

$$\tilde{n}(q+p) = \tilde{n} = Np, \tag{29}$$

since $p + q = 1$. We can now find the terms in the expansion

$$\begin{aligned}B_2 &\equiv \left[\frac{d^2\ln[P(n)]}{dn^2}\right]_{n=\tilde{n}} = -\frac{1}{\tilde{n}} - \frac{1}{N-\tilde{n}} \\ &= -\frac{1}{Np} - \frac{1}{N(1-p)} = -\frac{1}{N}\left(\frac{1}{p}+\frac{1}{q}\right) \\ &= -\frac{1}{N}\left(\frac{p+q}{pq}\right) = -\frac{1}{Npq} = -\frac{1}{N(1-p)}\end{aligned} \tag{30}$$

$$\begin{aligned}B_3 &\equiv \left[\frac{d^3\ln[P(n)]}{dn^3}\right]_{n=\tilde{n}} = \frac{1}{\tilde{n}^2} - \frac{1}{(N-\tilde{n})^2} \\ &= \frac{1}{N^2p^2} - \frac{1}{N^2q^2} = \frac{q^2-p^2}{N^2p^2q^2} \\ &= \frac{(1-2p+p^2)-p^2}{N^2p^2(1-p)^2} = \frac{1-2p}{N^2p^2(1-p)^2}\end{aligned} \tag{31}$$

$$\begin{aligned}B_4 &\equiv \left[\frac{d^4\ln[P(n)]}{dn^4}\right]_{n=\tilde{n}} = -\frac{2}{\tilde{n}^3} - \frac{2}{(n-\tilde{n})^3} \\ &= -2\left(\frac{1}{N^3p^3} + \frac{1}{N^3q^3}\right) = \frac{2(p^3+q^3)}{N^3p^3q^3} \\ &= \frac{2(p^2-pq+q^2)}{N^3p^3q^3} \\ &= \frac{2[p^2-p(1-p)+(1-2p+p^2)]}{N^3p^3(1-p^3)} \\ &= \frac{2(3p^2-3p+1)}{N^3p^3(1-p^3)}.\end{aligned} \tag{32}$$

Now, treating the distribution as continuous,

$$\lim_{N\to\infty}\sum_{n=0}^{N} P(n) \approx \int P(n)\,dn = \int_{-\infty}^{\infty} P(\tilde{n}+\eta)\,d\eta = 1. \tag{33}$$

Since each term is of order $1/N \sim 1/\sigma^2$ smaller than the previous, we can ignore terms higher than B_2, so

$$P(n) = P(\tilde{n})e^{-|B_2|\eta^2/2}. \tag{34}$$

The probability must be normalized, so

$$\int_{-\infty}^{\infty} P(\tilde{n})e^{-|B_2|\eta^2/2}\,d\eta = P(\tilde{n})\sqrt{\frac{2\pi}{|B_2|}} = 1, \tag{35}$$

and

$$\begin{aligned}P(n) &= \sqrt{\frac{|B_2|}{2\pi}}e^{-|B_2|(n-\tilde{n})^2/2} \\ &= \frac{1}{\sqrt{2\pi Npq}}\exp\left[-\frac{(n-Np)^2}{2Npq}\right].\end{aligned} \tag{36}$$

Defining $\sigma^2 \equiv 2Npq$,

$$P(n) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(n-\tilde{n})^2}{2\sigma^2}\right], \tag{37}$$

which is a GAUSSIAN DISTRIBUTION. For $p \ll 1$, a different approximation procedure shows that the binomial distribution approaches the POISSON DISTRIBUTION. The first CUMULANT is

$$\kappa_1 = np, \tag{38}$$

and subsequent CUMULANTS are given by the RECURRENCE RELATION

$$\kappa_{r+1} = pq\frac{d\kappa_r}{dp}. \tag{39}$$

Let x and y be independent binomial RANDOM VARIABLES characterized by parameters n, p and m, p. The CONDITIONAL PROBABILITY of x given that $x + y = k$ is

$$\begin{aligned} P(x = i|x + y = k) &= \frac{P(x = i, x + y = k)}{P(x + y = k)} \\ &= \frac{P(x = i, y = k - i)}{P(x + y = k)} = \frac{P(x = i)P(y = k - i)}{P(x + y = k)} \\ &= \frac{\binom{n}{i}p^i(1-p)^{n-i}\binom{m}{k-i}p^{k-i}(1-p)^{m-(k-i)}}{\binom{n+m}{k}p^k(1-p)^{n+m-k}} \\ &= \frac{\binom{n}{i}\binom{m}{k-i}}{\binom{n+m}{k}}. \end{aligned} \tag{40}$$

Note that this is a HYPERGEOMETRIC DISTRIBUTION!

see also DE MOIVRE-LAPLACE THEOREM, HYPERGEOMETRIC DISTRIBUTION, NEGATIVE BINOMIAL DISTRIBUTION

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 531, 1987.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Incomplete Beta Function, Student's Distribution, F-Distribution, Cumulative Binomial Distribution." §6.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 219–223, 1992.

Spiegel, M. R. *Theory and Problems of Probability and Statistics.* New York: McGraw-Hill, p. 108–109, 1992.

Binomial Expansion

see BINOMIAL SERIES

Binomial Formula

see BINOMIAL SERIES, BINOMIAL THEOREM

Binomial Number

A number of the form $a^n \pm b^n$, where a, b, and n are INTEGERS. They can be factored algebraically

$$a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + \ldots + ab^{n-2} + b^{n-1}) \tag{1}$$

$$a^n + b^n = (a+b)(a^{n-1} - a^{n-2}b + \ldots - ab^{n-2} + b^{n-1}) \tag{2}$$

$$a^{nm} - b^{nm} = (a^m - b^m)[a^{m(n-1)} + a^{m(n-2)}b^m + \ldots + b^{m(n-1)}]. \tag{3}$$

In 1770, Euler proved that if $(a, b) = 1$, then every FACTOR of

$$a^{2^n} + b^{2^n} \tag{4}$$

is either 2 or of the form $2^{n+1}K + 1$. If p and q are PRIMES, then

$$\frac{a^{pq} - 1)(a-1)}{(a^p - 1)(a^q - 1)} - 1 \tag{5}$$

is DIVISIBLE by every PRIME FACTOR of a^{p-1} not dividing $a^q - 1$.

see also CUNNINGHAM NUMBER, FERMAT NUMBER, MERSENNE NUMBER, RIESEL NUMBER, SIERPIŃSKI NUMBER OF THE SECOND KIND

References

Guy, R. K. "When Does $2^a - 2^b$ Divide $n^a - n^b$." §B47 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 102, 1994.

Qi, S and Ming-Zhi, Z. "Pairs where $2^a - a^b$ Divides $n^a - n^b$ for All n." *Proc. Amer. Math. Soc.* **93**, 218–220, 1985.

Schinzel, A. "On Primitive Prime Factors of $a^n - b^n$." *Proc. Cambridge Phil. Soc.* **58**, 555–562, 1962.

Binomial Series

For $|x| < 1$,

$$\begin{aligned} (1+x)^n &= \sum_{k=0}^{n} \binom{n}{k} x^k \qquad (1) \\ &= \binom{n}{0}x^0 + \binom{n}{1}x^1 + \binom{n}{2}x^2 + \ldots \qquad (2) \\ &= 1 + \frac{n!}{1!(n-1)!}x + \frac{n!}{(n-2)!2!}x^2 + \ldots \qquad (3) \\ &= 1 + nx + \frac{n(n-1)}{2}x^2 + \ldots. \qquad (4) \end{aligned}$$

The binomial series also has the CONTINUED FRACTION representation

$$(1+x)^n = \cfrac{1}{1 - \cfrac{nx}{1 + \cfrac{\frac{1\cdot(1+n)}{1\cdot 2}x}{1 + \cfrac{\frac{1\cdot(1-n)}{2\cdot 3}x}{1 + \cfrac{\frac{2(2+n)}{3\cdot 4}x}{1 + \cfrac{\frac{2(2-n)}{4\cdot 5}x}{1 + \cfrac{\frac{3(3+n)}{5\cdot 6}x}{1 + \ldots}}}}}}}. \tag{5}$$

see also BINOMIAL THEOREM, MULTINOMIAL SERIES, NEGATIVE BINOMIAL SERIES

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 14–15, 1972.

Pappas, T. "Pascal's Triangle, the Fibonacci Sequence & Binomial Formula." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 40–41, 1989.

Binomial Theorem

The theorem that, for INTEGRAL POSITIVE n,

$$(x+a)^n = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} x^k a^{n-k} = \sum_{k=0}^{n} \binom{n}{k} x^k a^{n-k},$$

the so-called BINOMIAL SERIES, where $\binom{n}{k}$ are BINOMIAL COEFFICIENTS. The theorem was known for the case $n = 2$ by Euclid around 300 BC, and stated in its modern form by Pascal in 1665. Newton (1676) showed that a similar formula (with INFINITE upper limit) holds for NEGATIVE INTEGRAL n,

$$(x+a)^{-n} = \sum_{k=0}^{\infty} \binom{-n}{k} x^k a^{-n-k},$$

the so-called NEGATIVE BINOMIAL SERIES, which converges for $|x| > |a|$.

see also BINOMIAL COEFFICIENT, BINOMIAL SERIES, CAUCHY BINOMIAL THEOREM, CHU-VANDERMONDE IDENTITY, LOGARITHMIC BINOMIAL FORMULA, NEGATIVE BINOMIAL SERIES, q-BINOMIAL THEOREM, RANDOM WALK

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 10, 1972.

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 307–308, 1985.

Conway, J. H. and Guy, R. K. "Choice Numbers Are Binomial Coefficients." In *The Book of Numbers.* New York: Springer-Verlag, pp. 72–74, 1996.

Coolidge, J. L. "The Story of the Binomial Theorem." *Amer. Math. Monthly* **56**, 147–157, 1949.

Courant, R. and Robbins, H. "The Binomial Theorem." §1.6 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 16–18, 1996.

Binomial Triangle

see PASCAL'S TRIANGLE

Binormal Developable

A RULED SURFACE M is said to be a binormal developable of a curve $\mathbf{y}$ if M can be parameterized by $\mathbf{x}(u,v) = \mathbf{y}(u) + v\hat{\mathbf{B}}(u)$, where $\mathbf{B}$ is the BINORMAL VECTOR.

see also NORMAL DEVELOPABLE, TANGENT DEVELOPABLE

References

Gray, A. "Developables." §17.6 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 352–354, 1993.

Binormal Vector

$$\hat{\mathbf{B}} \equiv \hat{\mathbf{T}} \times \hat{\mathbf{N}} \tag{1}$$

$$= \frac{\mathbf{r}' \times \mathbf{r}''}{|\mathbf{r}' \times \mathbf{r}''|}, \tag{2}$$

where the unit TANGENT VECTOR $\mathbf{T}$ and unit "principal" NORMAL VECTOR $\mathbf{N}$ are defined by

$$\hat{\mathbf{T}} \equiv \frac{\mathbf{r}'(s)}{|\mathbf{r}'(s)|} \tag{3}$$

$$\hat{\mathbf{N}} \equiv \frac{\mathbf{r}''(s)}{|\mathbf{r}''(s)|} \tag{4}$$

Here, $\mathbf{r}$ is the RADIUS VECTOR, s is the ARC LENGTH, τ is the TORSION, and κ is the CURVATURE. The binormal vector satisfies the remarkable identity

$$[\dot{\mathbf{B}}, \ddot{\mathbf{B}}, \dddot{\mathbf{B}}] = \tau^5 \frac{d}{ds}\left(\frac{\kappa}{\tau}\right). \tag{5}$$

see also FRENET FORMULAS, NORMAL VECTOR, TANGENT VECTOR

References

Kreyszig, E. "Binormal. Moving Trihedron of a Curve." §13 in *Differential Geometry.* New York: Dover, p. 36–37, 1991.

Bioche's Theorem

If two complementary PLÜCKER CHARACTERISTICS are equal, then each characteristic is equal to its complement except in four cases where the sum of order and class is 9.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 101, 1959.

Biotic Potential

see LOGISTIC EQUATION

Bipartite Graph

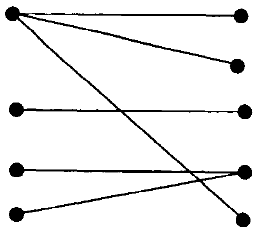

A set of VERTICES decomposed into two disjoint sets such that no two VERTICES within the same set are adjacent. A bigraph is a special case of a k-PARTITE GRAPH with $k = 2$.

see also COMPLETE BIPARTITE GRAPH, k-PARTITE GRAPH, KÖNIG-EGEVÁRY THEOREM

References

Chartrand, G. *Introductory Graph Theory.* New York: Dover, p. 116, 1985.

Saaty, T. L. and Kainen, P. C. *The Four-Color Problem: Assaults and Conquest.* New York: Dover, p. 12, 1986.

Biplanar Double Point

see ISOLATED SINGULARITY

Bipolar Coordinates

Bipolar coordinates are a 2-D system of coordinates. There are two commonly defined types of bipolar coordinates, the first of which is defined by

$$x = \frac{a \sinh v}{\cosh v - \cos u} \tag{1}$$

$$y = \frac{a \sin u}{\cosh v - \cos u}, \tag{2}$$

where $u \in [0, 2\pi)$, $v \in (-\infty, \infty)$. The following identities show that curves of constant u and v are CIRCLES in xy-space.

$$x^2 + (y - a \cot u)^2 = a^2 \csc^2 u \tag{3}$$

$$(x - a \coth v)^2 + y^2 = a^2 \operatorname{csch}^2 v. \tag{4}$$

The SCALE FACTORS are

$$h_u = \frac{a}{\cosh v - \cos u} \tag{5}$$

$$h_v = \frac{a}{\cosh v - \cos u}. \tag{6}$$

The LAPLACIAN is

$$\nabla^2 = \frac{(\cosh v - \cos u)^2}{a^2} \left(\frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2} \right). \tag{7}$$

LAPLACE'S EQUATION is separable.

Two-center bipolar coordinates are two coordinates giving the distances from two fixed centers r_1 and r_2, sometimes denoted r and r'. For two-center bipolar coordinates with centers at $(\pm c, 0)$,

$$r_1{}^2 = (x + c)^2 + y^2 \tag{8}$$

$$r_2{}^2 = (x - c)^2 + y^2. \tag{9}$$

Combining (8) and (9) gives

$$r_1{}^2 - r_2{}^2 = 4cx. \tag{10}$$

Solving for CARTESIAN COORDINATES x and y gives

$$x = \frac{r_1{}^2 - r_2{}^2}{4c} \tag{11}$$

$$y = \pm \frac{1}{4c} \sqrt{16c^2 r_1{}^2 - (r_1{}^2 - r_2{}^2 + 4c^2)}. \tag{12}$$

Solving for POLAR COORDINATES gives

$$r = \sqrt{\frac{r_1{}^2 + r_2{}^2 - 2c^2}{2}} \tag{13}$$

$$\theta = \tan^{-1} \left[\sqrt{\frac{8c^2 (r_1{}^2 + r_2{}^2 - 2c^2)}{r_1{}^2 - r_2{}^2} - 1} \right]. \tag{14}$$

References

Lockwood, E. H. "Bipolar Coordinates." Ch. 25 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 186–190, 1967.

Bipolar Cylindrical Coordinates

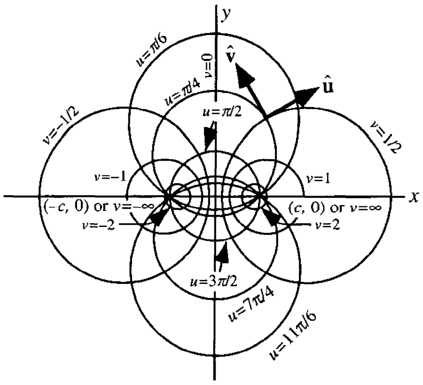

A set of CURVILINEAR COORDINATES defined by

$$x = \frac{a \sinh v}{\cosh v - \cos u} \tag{1}$$

$$y = \frac{a \sin u}{\cosh v - \cos u} \tag{2}$$

$$z = z, \tag{3}$$

where $u \in [0, 2\pi)$, $v \in (-\infty, \infty)$, and $z \in (-\infty, \infty)$. There are several notational conventions, and whereas (u, v, z) is used in this work, Arfken (1970) prefers

(η, ξ, z). The following identities show that curves of constant u and v are CIRCLES in xy-space.

$$x^2 + (y - a\cot u)^2 = a^2\csc^2 u \tag{4}$$

$$(x - a\coth v)^2 + y^2 = a^2\operatorname{csch}^2 v. \tag{5}$$

The SCALE FACTORS are

$$h_u = \frac{a}{\cosh v - \cos u} \tag{6}$$

$$h_v = \frac{a}{\cosh v - \cos u} \tag{7}$$

$$h_z = 1. \tag{8}$$

The LAPLACIAN is

$$\nabla^2 = \frac{(\cosh v - \cos u)^2}{a^2}\left(\frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2}\right) + \frac{\partial^2}{\partial z^2}. \tag{9}$$

LAPLACE'S EQUATION is not separable in BIPOLAR CYLINDRICAL COORDINATES, but it is in 2-D BIPOLAR COORDINATES.

References

Arfken, G. "Bipolar Coordinates (ξ, η, z)." §2.9 in *Mathematical Methods for Physicists, 2nd ed.* Orlando, FL: Academic Press, pp. 97–102, 1970.

Biprism

Two slant triangular PRISMS fused together.

see also PRISM, SCHMITT-CONWAY BIPRISM

Bipyramid

see DIPYRAMID

Biquadratefree

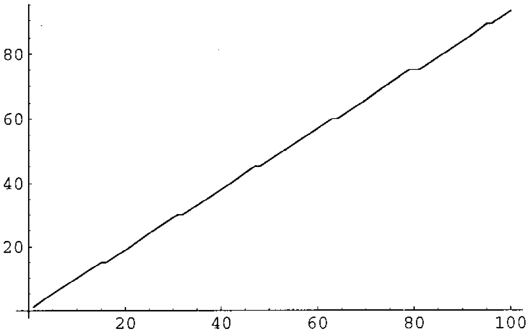

A number is said to be biquadratefree if its PRIME decomposition contains no tripled factors. All PRIMES are therefore trivially biquadratefree. The biquadratefree numbers are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, ... (Sloane's A046100). The biquadrateful numbers (i.e., those that contain at least one biquadrate) are 16, 32, 48, 64, 80, 81, 96, ... (Sloane's A046101). The number of biquadratefree numbers less than 10, 100, 1000, ... are 10, 93, 925, 9240, 92395, 923939, ..., and their asymptotic density is $1/\zeta(4) = 90/\pi^4 \approx 0.923938$, where $\zeta(n)$ is the RIEMANN ZETA FUNCTION.

see also CUBEFREE, PRIME NUMBER, RIEMANN ZETA FUNCTION, SQUAREFREE

References

Sloane, N. J. A. Sequences A046100 and A046101 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Biquadratic Equation

see QUARTIC EQUATION

Biquadratic Number

A biquadratic number is a fourth POWER, n^4. The first few biquadratic numbers are 1, 16, 81, 256, 625, ... (Sloane's A000583). The minimum number of squares needed to represent the numbers 1, 2, 3, ... are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1, 2, 3, 4, 5, ... (Sloane's A002377), and the number of distinct ways to represent the numbers 1, 2, 3, ... in terms of biquadratic numbers are 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, ... A brute-force algorithm for enumerating the biquadratic permutations of n is repeated application of the GREEDY ALGORITHM.

Every POSITIVE integer is expressible as a SUM of (at most) $g(4) = 19$ biquadratic numbers (WARING'S PROBLEM). Davenport (1939) showed that $G(4) = 16$, meaning that all sufficiently large integers require only 16 biquadratic numbers. The following table gives the first few numbers which require 1, 2, 3, ..., 19 biquadratic numbers to represent them as a sum, with the sequences for 17, 18, and 19 being finite.

#	Sloane	Numbers
1	000290	1, 16, 81, 256, 625, 1296, 2401, 4096, ...
2	003336	2, 17, 32, 82, 97, 162, 257, 272, ...
3	003337	3, 18, 33, 48, 83, 98, 113, 163, ...
4	003338	4, 19, 34, 49, 64, 84, 99, 114, 129, ...
5	003339	5, 20, 35, 50, 65, 80, 85, 100, 115, ...
6	003340	6, 21, 36, 51, 66, 86, 96, 101, 116, ...
7	003341	7, 22, 37, 52, 67, 87, 102, 112, 117, ...
8	003342	8, 23, 38, 53, 68, 88, 103, 118, 128, ...
9	003343	9, 24, 39, 54, 69, 89, 104, 119, 134, ...
10	003344	10, 25, 40, 55, 70, 90, 105, 120, 135, ...
11	003345	11, 26, 41, 56, 71, 91, 106, 121, 136, ...
12	003346	12, 27, 42, 57, 72, 92, 107, 122, 137, ...

The following table gives the numbers which can be represented in n different ways as a sum of k biquadrates.

k	n	Sloane	Numbers
1	1	000290	1, 16, 81, 256, 625, 1296, 2401, 4096, ...
2	2		635318657, 3262811042, 8657437697, ...

The numbers 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 18, 19, 20, 21, ... (Sloane's A046039) cannot be represented using distinct biquadrates.

see also CUBIC NUMBER, SQUARE NUMBER, WARING'S PROBLEM

References

Davenport, H. "On Waring's Problem for Fourth Powers." *Ann. Math.* **40**, 731–747, 1939.

Biquadratic Reciprocity Theorem

$$x^4 \equiv q \pmod{p}. \tag{1}$$

This was solved by Gauss using the GAUSSIAN INTEGERS as

$$\left(\frac{\pi}{\sigma}\right)_4 \left(\frac{\sigma}{\pi}\right)_4 = (-1)^{[(N(\pi)-1)/4][(N(\sigma)-1)/4]}, \tag{2}$$

where π and σ are distinct GAUSSIAN INTEGER PRIMES,

$$N(a+bi) = \sqrt{a^2+b^2} \tag{3}$$

and N is the norm.

$$\left(\frac{\alpha}{\pi}\right)_4 = \begin{cases} 1 & \text{if } x^4 \equiv \alpha \pmod{\pi} \text{ is solvable} \\ -1, i, \text{ or } -i & \text{otherwise,} \end{cases} \tag{4}$$

where solvable means solvable in terms of GAUSSIAN INTEGERS.

see also RECIPROCITY THEOREM

Biquaternion

A QUATERNION with COMPLEX coefficients. The ALGEBRA of biquaternions is isomorphic to a full matrix ring over the complex number field (van der Waerden 1985).

see also QUATERNION

References

Clifford, W. K. "Preliminary Sketch of Biquaternions." *Proc. London Math. Soc.* **4**, 381–395, 1873.

Hamilton, W. R. *Lectures on Quaternions: Containing a Systematic Statement of a New Mathematical Method.* Dublin: Hodges and Smith, 1853.

Study, E. "Von den Bewegung und Umlegungen." *Math. Ann.* **39**, 441–566, 1891.

van der Waerden, B. L. *A History of Algebra from al-Khwarizmi to Emmy Noether.* New York: Springer-Verlag, pp. 188–189, 1985.

Birational Transformation

A transformation in which coordinates in two SPACES are expressed rationally in terms of those in another.

see also RIEMANN CURVE THEOREM, WEBER'S THEOREM

Birch Conjecture

see SWINNERTON-DYER CONJECTURE

Birch–Swinnerton-Dyer Conjecture

see SWINNERTON-DYER CONJECTURE

Birkhoff's Ergodic Theorem

Let T be an ergodic ENDOMORPHISM of the PROBABILITY SPACE X and let $f : X \to \mathbb{R}$ be a real-valued MEASURABLE FUNCTION. Then for ALMOST EVERY $x \in X$, we have

$$\frac{1}{n}\sum_{j=1}^{n} f \circ F^j(x) \to \int f\,dm$$

as $n \to \infty$. To illustrate this, take f to be the characteristic function of some SUBSET A of X so that

$$f(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

The left-hand side of (-1) just says how often the orbit of x (that is, the points x, Tx, T^2x, ...) lies in A, and the right-hand side is just the MEASURE of A. Thus, for an ergodic ENDOMORPHISM, "space-averages = time-averages almost everywhere." Moreover, if T is continuous and uniquely ergodic with BOREL PROBABILITY MEASURE m and f is continuous, then we can replace the ALMOST EVERYWHERE convergence in (-1) to everywhere.

Birotunda

Two adjoined ROTUNDAS.

see also BILUNABIROTUNDA, CUPOLAROTUNDA, ELONGATED GYROCUPOLAROTUNDA, ELONGATED ORTHOCUPOLAROTUNDA, ELONGATED ORTHOBIROTUNDA, GYROCUPOLAROTUNDA, GYROELONGATED ROTUNDA, ORTHOBIROTUNDA, TRIANGULAR HEBESPHENOROTUNDA

Birthday Attack

Birthday attacks are a class of brute-force techniques used in an attempt to solve a class of cryptographic hash function problems. These methods take advantage of functions which, when supplied with a random input, return one of k equally likely values. By repeatedly evaluating the function for different inputs, the same output is expected to be obtained after about $1.2\sqrt{k}$ evaluations.

see also BIRTHDAY PROBLEM

References

RSA Laboratories. "Question 95. What is a Birthday Attack." `http://www.rsa.com/rsalabs/newfaq/q95.html`. "Question 96. How Does the Length of a Hash Value Affect Security?" `http://www.rsa.com/rsalabs/newfaq/q96.html`.

van Oorschot, P. and Wiener, M. "A Known Plaintext Attack on Two-Key Triple Encryption." In *Advances in Cryptology—Eurocrypt '90.* New York: Springer-Verlag, pp. 366–377, 1991.

Yuval, G. "How to Swindle Rabin." *Cryptologia* **3**, 187–189, Jul. 1979.

Birthday Problem

Consider the probability $Q_1(n,d)$ that *no two people* out of a group of n will have matching birthdays out of d equally possible birthdays. Start with an arbitrary person's birthday, then note that the probability that the second person's birthday is different is $(d-1)/d$, that the third person's birthday is different from the first two is $[(d-1)/d][(d-2)/d]$, and so on, up through the nth person. Explicitly,

$$\begin{aligned} Q_1(n,d) &= \frac{d-1}{d}\frac{d-2}{d}\cdots\frac{d-(n-1)}{d} \\ &= \frac{(d-1)(d-2)\cdots[d-(n-1)]}{d^n}. \qquad (1) \end{aligned}$$

But this can be written in terms of FACTORIALS as

$$Q_1(n,d) = \frac{d!}{(d-n)!d^n}, \qquad (2)$$

so the probability $P_2(n,365)$ that two people out of a group of n *do have the same birthday* is therefore

$$P_2(n,d) = 1 - Q_1(n,d) = 1 - \frac{d!}{(d-n)!d^n}. \qquad (3)$$

If 365-day years have been assumed, i.e., the existence of leap days is ignored, then the number of people needed for there to be at least a 50% chance that two share birthdays is the smallest n such that $P_2(n,365) \geq 1/2$. This is given by $n = 23$, since

$$\begin{aligned} &P_2(23,365) = \\ &\frac{38093904702297390785243708291056390518886454060947061}{75091883268515350125426207425223147563269805908203125} \\ &\approx 0.507297. \qquad (4) \end{aligned}$$

The number of people needed to obtain $P_2(n,365) \geq 1/2$ for $n = 1, 2, \ldots,$ are 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, ... (Sloane's A033810).

The probability $P_2(n,d)$ can be estimated as

$$\begin{aligned} P_2(n,d) &\approx 1 - e^{-n(n-1)/2d} \qquad (5) \\ &\approx 1 - \left(1 - \frac{n}{2d}\right)^{n-1}, \qquad (6) \end{aligned}$$

where the latter has error

$$\epsilon < \frac{n^3}{6(d-n+1)^2} \qquad (7)$$

(Sayrafiezadeh 1994).

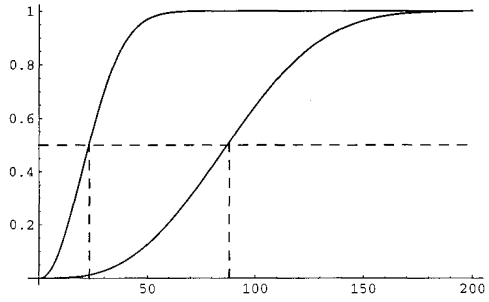

In general, let $Q_i(n,d)$ denote the probability that a birthday is shared by exactly i (and no more) people out of a group of n people. Then the probability that a birthday is shared by k *or more* people is given by

$$P_k(n,d) = 1 - \sum_{i=1}^{k-1} Q_i(n,d). \qquad (8)$$

Q_2 can be computed explicitly as

$$\begin{aligned} Q_2(n,d) &= \frac{n!}{d^n}\sum_{i=1}^{\lfloor n/2\rfloor}\frac{1}{2^i}\binom{d}{i}\binom{d-i}{n-2i} \\ &= \frac{n!}{d^n}\sum_{i=1}^{\lfloor n/2\rfloor}\frac{d!}{2^i i!(n-2i)!(d-n+i)!} \\ &= \frac{(-1)^n}{d^n}\left[2^{-n/2}\Gamma(1+n)P_n^{(-d)}(\tfrac{1}{2}\sqrt{2}) - \frac{\Gamma(1+d)}{\Gamma(1+d-n)}\right], \qquad (9) \end{aligned}$$

where $\binom{n}{m}$ is a BINOMIAL COEFFICIENT, $\Gamma(n)$ is a GAMMA FUNCTION, and $P_n^{(\lambda)}(x)$ is an ULTRASPHERICAL POLYNOMIAL. This gives the explicit formula for $P_3(n,d)$ as

$$\begin{aligned} P_3(n,d) &= 1 - Q_1(n,d) - Q_2(n,d) \\ &= 1 + \frac{(-1)^{n+1}\Gamma(n+1)P_n^{(-d)}(2^{-1/2})}{2^{n/2}d^n}. \qquad (10) \end{aligned}$$

$Q_3(n,d)$ cannot be computed in entirely closed form, but a partially reduced form is

$$\begin{aligned} Q_3(n,d) = \frac{\Gamma(d+1)}{d^n}&\left[\frac{(-1)^n F(\frac{9}{8}) - F(-\frac{9}{8})}{\Gamma(d-n+1)}\right. \\ &\left. +(-1)^n\Gamma(1+n)\sum_{i=1}^{\lfloor n/3\rfloor}\frac{(-3)^{-i}2^{(i-n)/2}P_{n-3i}^{(i-d)}(\frac{1}{2}\sqrt{2})}{\Gamma(d-i+1)\Gamma(i+1)}\right], \qquad (11) \end{aligned}$$

where

$$F = F(n,d,a) \equiv 1 - {}_3F_2\left[\begin{matrix}\frac{1}{3}(1-n), \frac{1}{3}(2-n), -\frac{1}{3} \\ \frac{1}{2}(d-n+1), \frac{1}{2}(d-n+2)\end{matrix}; a\right] \qquad (12)$$

and ${}_3F_2(a,b,c;d,e;z)$ is a GENERALIZED HYPERGEOMETRIC FUNCTION.

In general, $Q_k(n,d)$ can be computed using the RECURRENCE RELATION

$$\begin{aligned} Q_k(n,d) = \sum_{i=1}^{\lfloor n/k\rfloor}&\left[\frac{n!d!}{d^{ik}i!(k!)^i(n-ik)!(d-i)!}\right. \\ &\left.\times\sum_{j=1}^{k-1}Q_j(n-k,d-i)\frac{(d-i)^{n-ik}}{d^{n-ik}}\right] \qquad (13) \end{aligned}$$

(Finch). However, the time to compute this recursive function grows exponentially with k and so rapidly becomes unwieldy. The minimal number of people to give a 50% probability of having at least n coincident birthdays is 1, 23, 88, 187, 313, 460, 623, 798, 985, 1181, 1385, 1596, 1813, ... (Sloane's A014088; Diaconis and Mosteller 1989).

A good approximation to the number of people n such that $p = P_k(n, d)$ is some given value can given by solving the equation

$$ne^{-n/(dk)} = \left[d^{k-1}k! \ln\left(\frac{1}{1-p}\right)\left(1 - \frac{n}{d(k+1)}\right)\right]^{1/k} \tag{14}$$

for n and taking $\lceil n \rceil$, where $\lceil n \rceil$ is the CEILING FUNCTION (Diaconis and Mosteller 1989). For $p = 0.5$ and $k = 1, 2, 3, \ldots$, this formula gives $n = 1$, 23, 88, 187, 313, 459, 722, 797, 983, 1179, 1382, 1592, 1809, ..., which differ from the true values by from 0 to 4. A much simpler but also poorer approximation for n such that $p = 0.5$ for $k < 20$ is given by

$$n = 47(k - 1.5)^{3/2} \tag{15}$$

(Diaconis and Mosteller 1989), which gives 86, 185, 307, 448, 606, 778, 965, 1164, 1376, 1599, 1832, ... for $k = 3$, 4, ...

The "almost" birthday problem, which asks the number of people needed such that two have a birthday within a day of each other, was considered by Abramson and Moser (1970), who showed that 14 people suffice. An approximation for the minimum number of people needed to get a 50-50 chance that two have a match within k days out of d possible is given by

$$n(k, d) = 1.2\sqrt{\frac{d}{2k+1}} \tag{16}$$

(Sevast'yanov 1972, Diaconis and Mosteller 1989).

see also BIRTHDAY ATTACK, COINCIDENCE, SMALL WORLD PROBLEM

References

Abramson, M. and Moser, W. O. J. "More Birthday Surprises." *Amer. Math. Monthly* **77**, 856–858, 1970.
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 45–46, 1987.
Bloom, D. M. "A Birthday Problem." *Amer. Math. Monthly* **80**, 1141–1142, 1973.
Bogomolny, A. "Coincidence." `http://www.cut-the-knot.com/do_you_know/coincidence.html`.
Clevenson, M. L. and Watkins, W. "Majorization and the Birthday Inequality." *Math. Mag.* **64**, 183–188, 1991.
Diaconis, P. and Mosteller, F. "Methods of Studying Coincidences." *J. Amer. Statist. Assoc.* **84**, 853–861, 1989.
Feller, W. *An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd ed.* New York: Wiley, pp. 31–32, 1968.
Finch, S. "Puzzle #28 [June 1997]: Coincident Birthdays." `http://www.mathsoft.com/mathcad/library/puzzle/soln28/soln28.html`.
Gehan, E. A. "Note on the 'Birthday Problem.'" *Amer. Stat.* **22**, 28, Apr. 1968.
Heuer, G. A. "Estimation in a Certain Probability Problem." *Amer. Math. Monthly* **66**, 704–706, 1959.
Hocking, R. L. and Schwertman, N. C. "An Extension of the Birthday Problem to Exactly k Matches." *College Math. J.* **17**, 315–321, 1986.
Hunter, J. A. H. and Madachy, J. S. *Mathematical Diversions.* New York: Dover, pp. 102–103, 1975.
Klamkin, M. S. and Newman, D. J. "Extensions of the Birthday Surprise." *J. Combin. Th.* **3**, 279–282, 1967.
Levin, B. "A Representation for Multinomial Cumulative Distribution Functions." *Ann. Statistics* **9**, 1123–1126, 1981.
McKinney, E. H. "Generalized Birthday Problem." *Amer. Math. Monthly* **73**, 385–387, 1966.
Mises, R. von. "Über Aufteilungs—und Besetzungs-Wahrscheinlichkeiten." *Revue de la Faculté des Sciences de l'Université d'Istanbul, N. S.* **4**, 145–163, 1939. Reprinted in *Selected Papers of Richard von Mises, Vol. 2* (Ed. P. Frank, S. Goldstein, M. Kac, W. Prager, G. Szegő, and G. Birkhoff). Providence, RI: Amer. Math. Soc., pp. 313–334, 1964.
Riesel, H. *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, pp. 179–180, 1994.
Sayrafiezadeh, M. "The Birthday Problem Revisited." *Math. Mag.* **67**, 220–223, 1994.
Sevast'yanov, B. A. "Poisson Limit Law for a Scheme of Sums of Dependent Random Variables." *Th. Prob. Appl.* **17**, 695–699, 1972.
Sloane, N. J. A. Sequences A014088 and A033810 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Stewart, I. "What a Coincidence!" *Sci. Amer.* **278**, 95–96, June 1998.
Tesler, L. "Not a Coincidence!" `http://www.nomodes.com/coincidence.html`.

Bisected Perimeter Point

see NAGEL POINT

Bisection Procedure

Given an interval $[a, b]$, let a_n and b_n be the endpoints at the nth iteration and r_n be the nth approximate solution. Then, the number of iterations required to obtain an error smaller than ϵ is found as follows.

$$b_n - a_n = \frac{1}{2^{n-1}}(b - a) \tag{1}$$

$$r_n \equiv \tfrac{1}{2}(a_n + b_n) \tag{2}$$

$$|r_n - r| \leq \tfrac{1}{2}(b_n - a_n) = 2^{-n}(b - a) < \epsilon \tag{3}$$

$$-n \ln 2 < \ln \epsilon - \ln(b - a), \tag{4}$$

so

$$n > \frac{\ln(b - a) - \ln \epsilon}{\ln 2}. \tag{5}$$

see also ROOT

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 964–965, 1985.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Bracketing and Bisection." §9.1 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 343–347, 1992.

Bisector

Bisection is the division of a given curve or figure into two equal parts (halves).

see also ANGLE BISECTOR, BISECTION PROCEDURE, EXTERIOR ANGLE BISECTOR, HALF, HEMISPHERE, LINE BISECTOR, PERPENDICULAR BISECTOR, TRISECTION

Bishop's Inequality

Let $V(r)$ be the volume of a BALL of radius r in a complete n-D RIEMANNIAN MANIFOLD with RICCI CURVATURE $\geq (n-1)\kappa$. Then $V(r) \geq V_\kappa(r)$, where V_κ is the volume of a BALL in a space having constant SECTIONAL CURVATURE. In addition, if equality holds for some BALL, then this BALL is ISOMETRIC to the BALL of radius r in the space of constant SECTIONAL CURVATURE κ.

References

Chavel, I. *Riemannian Geometry: A Modern Introduction.* New York: Cambridge University Press, 1994.

Bishops Problem

B							
B							B
B							B
B							B
B							B
B							B
B							B
B							

Find the maximum number of bishops $B(n)$ which can be placed on an $n \times n$ CHESSBOARD such that no two attack each other. The answer is $2n-2$ (Dudeney 1970, Madachy 1979), giving the sequence 2, 4, 6, 8, ... (the EVEN NUMBERS) for $n = 2, 3, \ldots$. One maximal solution for $n = 8$ is illustrated above. The number of distinct maximal arrangements of bishops for $n = 1, 2, \ldots$ are 1, 4, 26, 260, 3368, ... (Sloane's A002465). The number of rotationally and reflectively distinct solutions on an $n \times n$ board for $n \geq 2$ is

$$B(n) = \begin{cases} 2^{(n-4)/2}[2^{(n-2)/2}+1] & \text{for } n \text{ even} \\ 2^{(n-3)/2}[2^{(n-3)/2}+1] & \text{for } n \text{ odd} \end{cases}$$

(Dudeney 1970, p. 96; Madachy 1979, p. 45; Pickover 1995). An equivalent formula is

$$B(n) = 2^{n-3} + 2^{\lfloor (n-1)/2 \rfloor - 1},$$

where $\lfloor n \rfloor$ is the FLOOR FUNCTION, giving the sequence for $n = 1, 2, \ldots$ as 1, 1, 2, 3, 6, 10, 20, 36, ... (Sloane's A005418).

			B				
			B				
			B				
			B				
			B				
			B				
			B				
			B				

The minimum number of bishops needed to occupy or attack all squares on an $n \times n$ CHESSBOARD is n, arranged as illustrated above.

see also CHESS, KINGS PROBLEM, KNIGHTS PROBLEM, QUEENS PROBLEM, ROOKS PROBLEM

References

Ahrens, W. *Mathematische Unterhaltungen und Spiele, Vol. 1, 3rd ed.* Leipzig, Germany: Teubner, p. 271, 1921.
Dudeney, H. E. "Bishops—Unguarded" and "Bishops—Guarded." §297 and 298 in *Amusements in Mathematics.* New York: Dover, pp. 88–89, 1970.
Guy, R. K. "The n Queens Problem." §C18 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 133–135, 1994.
Madachy, J. *Madachy's Mathematical Recreations.* New York: Dover, pp. 36–46, 1979.
Pickover, C. A. *Keys to Infinity.* New York: Wiley, pp. 74–75, 1995.
Sloane, N. J. A. Sequences A002465/M3616 and A005418/M0771 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Bislit Cube

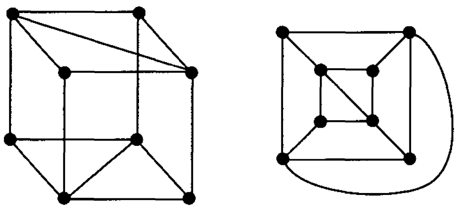

The 8-VERTEX graph consisting of a CUBE in which two opposite faces have DIAGONALS oriented PERPENDICULAR to each other.

see also BIDIAKIS CUBE, CUBE, CUBICAL GRAPH

Bispherical Coordinates

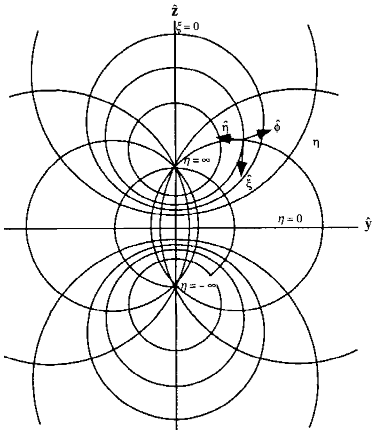

A system of CURVILINEAR COORDINATES defined by

$$x = \frac{a \sin\xi \cos\phi}{\cosh\eta - \cos\xi} \tag{1}$$

$$y = \frac{a \sin\xi \sin\phi}{\cosh\eta - \cos\xi} \tag{2}$$

$$z = \frac{a \sinh\eta}{\cosh\eta - \cos\xi}. \tag{3}$$

The SCALE FACTORS are

$$h_\xi = \frac{a}{\cos\eta - \cos\xi} \tag{4}$$

$$h_\eta = \frac{a}{\cosh\eta - \cos\xi} \tag{5}$$

$$h_\phi = \frac{a \sin\xi}{\cosh\eta - \cos\xi}. \tag{6}$$

The LAPLACIAN is

$$\begin{aligned}\nabla^2 = \bigg(&\frac{-\cos u \cot^2 u + 3\cosh v \cot^2 u}{\cosh v - \cos u} \\ &+ \frac{-3\cosh^2 v \cot u \csc u + \cosh^3 v \csc^2 u}{\cosh v - \cos u}\bigg) \frac{\partial}{\partial\phi^2} \\ &+ (\cos u - \cosh v) \sinh v \frac{\partial}{\partial v} + (\cosh^2 v - \cos u)^2 \frac{\partial^2}{\partial v^2} \\ &+ (\cosh v - \cos u)(\cosh v \cot u - \sin u - \cos u \cot u) \frac{\partial}{\partial u} \\ &+ (\cosh^2 v - \cos u)^2 \frac{\partial^2}{\partial u^2}.\end{aligned} \tag{7}$$

In bispherical coordinates, LAPLACE'S EQUATION is separable, but the HELMHOLTZ DIFFERENTIAL EQUATION is not.

see also LAPLACE'S EQUATION—BISPHERICAL COORDINATES, TOROIDAL COORDINATES

References

Arfken, G. "Bispherical Coordinates (ξ, η, ϕ)." §2.14 in *Mathematical Methods for Physicists, 2nd ed.* Orlando, FL: Academic Press, pp. 115–117, 1970.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 665–666, 1953.

Bit Complexity

The number of single operations (of ADDITION, SUBTRACTION, and MULTIPLICATION) required to complete an algorithm.

see also STRASSEN FORMULAS

References

Borodin, A. and Munro, I. *The Computational Complexity of Algebraic and Numeric Problems.* New York: American Elsevier, 1975.

Bitangent

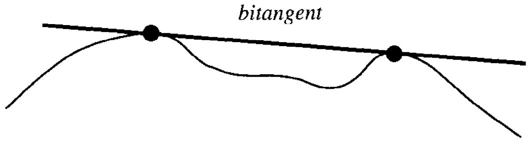

A LINE which is TANGENT to a curve at two distinct points.

see also KLEIN'S EQUATION, PLÜCKER CHARACTERISTICS, SECANT LINE, SOLOMON'S SEAL LINES, TANGENT LINE

Bivariate Distribution

see GAUSSIAN BIVARIATE DISTRIBUTION

Bivector

An antisymmetric TENSOR of second RANK (a.k.a. 2-form).

$$\vec{X} = X_{ab}\omega^a \wedge \omega^b,$$

where $\wedge$ is the WEDGE PRODUCT (or OUTER PRODUCT).

Biweight

see TUKEY'S BIWEIGHT

Black-Scholes Theory

The theory underlying financial derivatives which involves "stochastic calculus" and assumes an uncorrelated LOG NORMAL DISTRIBUTION of continuously varying prices. A simplified "binomial" version of the theory was subsequently developed by Sharpe *et al.* (1995) and Cox *et al.* (1979). It reproduces many results of the full-blown theory, and allows approximation of options for which analytic solutions are not known (Price 1996).

see also GARMAN-KOHLHAGEN FORMULA

References

Black, F. and Scholes, M. S. "The Pricing of Options and Corporate Liabilities." *J. Political Econ.* **81**, 637–659, 1973.

Cox, J. C.; Ross, A.; and Rubenstein, M. "Option Pricing: A Simplified Approach." *J. Financial Economics* **7**, 229–263, 1979.

Price, J. F. "Optional Mathematics is Not Optional." *Not. Amer. Math. Soc.* **43**, 964–971, 1996.

Sharpe, W. F.; Alexander, G. J.; and Bailey, J. V. *Investments, 5th ed.* Englewood Cliffs, NJ: Prentice-Hall, 1995.

Black Spleenwort Fern

see BARNSLEY'S FERN

Blackman Function

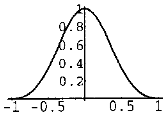

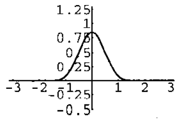

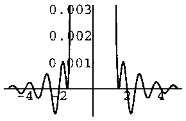

An APODIZATION FUNCTION given by

$$A(x) = 0.42 + 0.5\cos\left(\frac{\pi x}{a}\right) + 0.08\cos\left(\frac{2\pi x}{a}\right). \quad (1)$$

Its FULL WIDTH AT HALF MAXIMUM is $0.810957a$. The APPARATUS FUNCTION is

$$I(k) = \frac{a(0.84 - 0.36a^2k^2 - 2.17\times 10^{-19}a^4k^4)\sin(2\pi ak)}{(1-a^2k^2)(1-4a^2k^2)}. \quad (2)$$

The COEFFICIENTS are approximations to

$$a_0 = \frac{3969}{9304} \quad (3)$$

$$a_1 = \frac{1155}{4652} \quad (4)$$

$$a_2 = \frac{715}{18608}, \quad (5)$$

which would have produced zeros of $I(k)$ at $k = (7/4)a$ and $k = (9/4)a$.

see also APODIZATION FUNCTION

References

Blackman, R. B. and Tukey, J. W. "Particular Pairs of Windows." In *The Measurement of Power Spectra, From the Point of View of Communications Engineering.* New York: Dover, pp. 98-99, 1959.

Blancmange Function

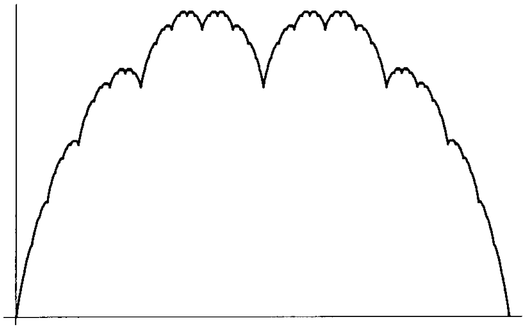

A CONTINUOUS FUNCTION which is nowhere DIFFERENTIABLE. The iterations towards the continuous function are BATRACHIONS resembling the HOFSTADTER-CONWAY \$10,000 SEQUENCE. The first six iterations are illustrated below. The dth iteration contains $N+1$ points, where $N = 2^d$, and can be obtained by setting $b(0) = b(N) = 0$, letting

$$b(m + 2^{n-1}) = 2^n + \tfrac{1}{2}[b(m) + b(m+2^n)],$$

and looping over $n = d$ to 1 by steps of -1 and $m = 0$ to $N-1$ by steps of 2^n.

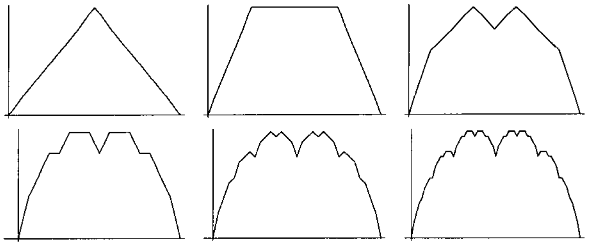

Peitgen and Saupe (1988) refer to this curve as the TAKAGI FRACTAL CURVE.

see also HOFSTADTER-CONWAY \$10,000 SEQUENCE, WEIERSTRAß FUNCTION

References

Dixon, R. *Mathographics.* New York: Dover, pp. 175–176 and 210, 1991.

Peitgen, H.-O. and Saupe, D. (Eds.). "Midpoint Displacement and Systematic Fractals: The Takagi Fractal Curve, Its Kin, and the Related Systems." §A.1.2 in *The Science of Fractal Images.* New York: Springer-Verlag, pp. 246–248, 1988.

Takagi, T. "A Simple Example of the Continuous Function without Derivative." *Proc. Phys. Math. Japan* **1**, 176–177, 1903.

Tall, D. O. "The Blancmange Function, Continuous Everywhere but Differentiable Nowhere." *Math. Gaz.* **66**, 11–22, 1982.

Tall, D. "The Gradient of a Graph." *Math. Teaching* **111**, 48–52, 1985.

Blaschke Conjecture

The only WIEDERSEHEN MANIFOLDS are the standard round spheres. The conjecture has been proven by combining the BERGER-KAZDAN COMPARISON THEOREM with A. Weinstein's results for n EVEN and C. T. Yang's for n ODD.

References

Chavel, I. *Riemannian Geometry: A Modern Introduction.* New York: Cambridge University Press, 1994.

Blaschke's Theorem

A convex planar domain in which the minimal length is > 1 always contains a CIRCLE of RADIUS 1/3.

References

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 25, 1983.

Blecksmith-Brillhart-Gerst Theorem

A generalization of SCHRÖTER'S FORMULA.

References

Berndt, B. C. *Ramanujan's Notebooks, Part III.* New York: Springer-Verlag, p. 73, 1985.

Blichfeldt's Lemma

see BLICHFELDT'S THEOREM

Blichfeldt's Theorem

Published in 1914 by Hans Blichfeldt. It states that any bounded planar region with POSITIVE AREA $> A$ placed in any position of the UNIT SQUARE LATTICE can be TRANSLATED so that the number of LATTICE POINTS inside the region will be at least $A + 1$. The theorem can be generalized to n-D.

BLM/Ho Polynomial

A 1-variable unoriented KNOT POLYNOMIAL $Q(x)$. It satisfies

$$Q_{\text{unknot}} = 1 \tag{1}$$

and the SKEIN RELATIONSHIP

$$Q_{L_+} + Q_{L_-} = x(Q_{L_0} + Q_{L_\infty}). \tag{2}$$

It also satisfies

$$Q_{L_1 \# L_2} = Q_{L_1} Q_{L_2}, \tag{3}$$

where # is the KNOT SUM and

$$Q_{L^*} = Q_L, \tag{4}$$

where L^* is the MIRROR IMAGE of L. The BLM/Ho polynomials of MUTANT KNOTS are also identical. Brandt *et al.* (1986) give a number of interesting properties. For any LINK L with ≥ 2 components, $Q_L - 1$ is divisible by $2(x - 1)$. If L has c components, then the lowest POWER of x in $Q_L(x)$ is $1 - c$, and

$$\lim_{x \to 0} x^{c-1} Q_L(x) = \lim_{(\ell, m) \to (1,0)} (-m)^{c-1} P_L(\ell, m), \tag{5}$$

where P_L is the HOMFLY POLYNOMIAL. Also, the degree of Q_L is less than the CROSSING NUMBER of L. If L is a 2-BRIDGE KNOT, then

$$Q_L(z) = 2z^{-1} V_L(t) V_L(t^{-1} + 1 - 2z^{-1}), \tag{6}$$

where $z \equiv -t - t^{-1}$ (Kanenobu and Sumi 1993).

The POLYNOMIAL was subsequently extended to the 2-variable KAUFFMAN POLYNOMIAL $F(a, z)$, which satisfies

$$Q(x) = F(1, x). \tag{7}$$

Brandt *et al.* (1986) give a listing of Q POLYNOMIALS for KNOTS up to 8 crossings and links up to 6 crossings.

References

Brandt, R. D.; Lickorish, W. B. R.; and Millett, K. C. "A Polynomial Invariant for Unoriented Knots and Links." *Invent. Math.* **84**, 563–573, 1986.

Ho, C. F. "A New Polynomial for Knots and Links—Preliminary Report." *Abstracts Amer. Math. Soc.* **6**, 300, 1985.

Kanenobu, T. and Sumi, T. "Polynomial Invariants of 2-Bridge Knots through 22-Crossings." *Math. Comput.* **60**, 771–778 and S17–S28, 1993.

Stoimenow, A. "Brandt-Lickorish-Millett-Ho Polynomials." `http://www.informatik.hu-berlin.de/~stoimeno/ptab/blmh10.html`.

Weisstein, E. W. "Knots." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Knots.m`.

Bloch Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let F be the set of COMPLEX analytic functions f defined on an open region containing the closure of the unit disk $D = \{z : |z| < 1\}$ satisfying $f(0) = 0$ and $df/dz(0) = 1$. For each f in F, let $b(f)$ be the SUPREMUM of all numbers r such that there is a disk S in D on which f is ONE-TO-ONE and such that $f(S)$ contains a disk of radius r. In 1925, Bloch (Conway 1978) showed that $b(f) \geq 1/72$. Define Bloch's constant by

$$B \equiv \inf\{b(f) : f \in F\}.$$

Ahlfors and Grunsky (1937) derived

$$\begin{aligned} 0.433012701\ldots = \tfrac{1}{4}\sqrt{3} &\leq B \\ &< \frac{1}{\sqrt{1+\sqrt{3}}} \frac{\Gamma(\frac{1}{3})\Gamma(\frac{11}{12})}{\Gamma(\frac{1}{4})} < 0.4718617. \end{aligned}$$

They also conjectured that the upper limit is actually the value of B,

$$\begin{aligned} B &= \frac{1}{\sqrt{1+\sqrt{3}}} \frac{\Gamma(\frac{1}{3})\Gamma(\frac{11}{12})}{\Gamma(\frac{1}{4})} \\ &= \sqrt{\pi}\, 2^{1/4} \frac{\Gamma(\frac{1}{3})}{\Gamma(\frac{1}{4})} \sqrt{\frac{\Gamma(\frac{11}{12})}{\Gamma(\frac{1}{12})}} \\ &= 0.4718617\ldots \end{aligned}$$

[illegible]

(Le Lionnais 1983).

see also LANDAU CONSTANT

References

Conway, J. B. *Functions of One Complex Variable, 2nd ed.* New York: Springer-Verlag, 1989.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/bloch/bloch.html`.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 25, 1983.

Minda, C. D. "Bloch Constants." *J. d'Analyse Math.* **41**, 54–84, 1982.

Bloch-Landau Constant

see LANDAU CONSTANT

Block

see also BLOCK DESIGN, SQUARE POLYOMINO

Block Design

An incidence system (v, k, λ, r, b) in which a set X of v points is partitioned into a family A of b subsets (blocks) in such a way that any two points determine λ blocks, there are k points in each block, and each point is contained in r different blocks. It is also generally required that $k < v$, which is where the "incomplete" comes from in the formal term most often encountered for block designs, BALANCED INCOMPLETE BLOCK DESIGNS (BIBD). The five parameters are not independent, but satisfy the two relations

$$vr = bk \tag{1}$$

$$\lambda(v-1) = r(k-1). \tag{2}$$

A BIBD is therefore commonly written as simply (v, k, λ), since b and r are given in terms of v, k, and λ by

$$b = \frac{v(v-1)\lambda}{k(k-1)} \tag{3}$$

$$r = \frac{\lambda(v-1)}{k-1}. \tag{4}$$

A BIBD is called SYMMETRIC if $b = v$ (or, equivalently, $r = k$).

Writing $X = \{x_i\}_{i=1}^{v}$ and $A = \{A_j\}_{j=1}^{b}$, then the INCIDENCE MATRIX of the BIBD is given by the $v \times b$ MATRIX M defined by

$$m_{ij} = \begin{cases} 1 & \text{if } x_i \in A \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

This matrix satisfies the equation

$$\mathsf{M}\mathsf{M}^{\mathrm{T}} = (r-\lambda)\mathsf{I} + \lambda\mathsf{J}, \tag{6}$$

where I is a $v \times v$ IDENTITY MATRIX and J is a $v \times v$ matrix of 1s (Dinitz and Stinson 1992).

Examples of BIBDs are given in the following table.

Block Design	(v, k, λ)
affine plane	(n^2, n, 1)
Fano plane	(7, 3, 1))
Hadamard design	symmetric ($4n+3$, $2n+1$, n)
projective plane	symmetric (n^2+n+1, $n+1$, 1)
Steiner triple system	(v, 3, 1)
unital	(q^3+1, $q+1$, 1)

see also AFFINE PLANE, DESIGN, FANO PLANE, HADAMARD DESIGN, PARALLEL CLASS, PROJECTIVE PLANE, RESOLUTION, RESOLVABLE, STEINER TRIPLE SYSTEM, SYMMETRIC BLOCK DESIGN, UNITAL

References

Dinitz, J. H. and Stinson, D. R. "A Brief Introduction to Design Theory." Ch. 1 in *Contemporary Design Theory: A Collection of Surveys* (Ed. J. H. Dinitz and D. R. Stinson). New York: Wiley, pp. 1–12, 1992.

Ryser, H. J. "The (b, v, r, k, λ)-Configuration." §8.1 in *Combinatorial Mathematics.* Buffalo, NY: Math. Assoc. Amer., pp. 96–102, 1963.

Block Growth

Let $(x_0x_1x_2\ldots)$ be a sequence over a finite ALPHABET A (all the entries are elements of A). Define the block growth function $B(n)$ of a sequence to be the number of ADMISSIBLE words of length n. For example, in the sequence $aabaabaabaabaab\ldots$, the following words are ADMISSIBLE

Length	Admissible Words
1	a, b
2	aa, ab, ba
3	aab, aba, baa
4	$aaba, abaa, baab$

so $B(1) = 2$, $B(2) = 3$, $B(3) = 3$, $B(4) = 3$, and so on. Notice that $B(n) \leq B(n+1)$, so the block growth function is always nondecreasing. This is because any ADMISSIBLE word of length n can be extended rightwards to produce an ADMISSIBLE word of length $n+1$. Moreover, suppose $B(n) = B(n+1)$ for some n. Then each admissible word of length n extends to a *unique* ADMISSIBLE word of length $n+1$.

For a SEQUENCE in which each substring of length n uniquely determines the next symbol in the SEQUENCE, there are only finitely many strings of length n, so the process must eventually cycle and the SEQUENCE must be eventually periodic. This gives us the following theorems:

1. If the SEQUENCE is eventually periodic, with least period p, then $B(n)$ is strictly increasing until it reaches p, and $B(n)$ is constant thereafter.
2. If the SEQUENCE is not eventually periodic, then $B(n)$ is strictly increasing and so $B(n) \geq n+1$ for all n. If a SEQUENCE has the property that $B(n) = n+1$ for all n, then it is said to have minimal block growth, and the SEQUENCE is called a STURMIAN SEQUENCE.

The block growth is also called the GROWTH FUNCTION or the COMPLEXITY of a SEQUENCE.

Block Matrix

A square DIAGONAL MATRIX in which the diagonal elements are SQUARE MATRICES of any size (possibly even 1×1), and the off-diagonal elements are 0.

Block (Set)

One of the disjoint SUBSETS making up a SET PARTITION. A block containing n elements is called an n-block. The partitioning of sets into blocks can be denoted using a RESTRICTED GROWTH STRING.

see also BLOCK DESIGN, RESTRICTED GROWTH STRING, SET PARTITION

Blow-Up

A common mechanism which generates SINGULARITIES from smooth initial conditions.

Blue-Empty Coloring

see BLUE-EMPTY GRAPH

Blue-Empty Graph

An EXTREMAL GRAPH in which the forced TRIANGLES are all the same color. Call R the number of red MONOCHROMATIC FORCED TRIANGLES and B the number of blue MONOCHROMATIC FORCED TRIANGLES, then a blue-empty graph is an EXTREMAL GRAPH with $B = 0$. For EVEN n, a blue-empty graph can be achieved by coloring red two COMPLETE SUBGRAPHS of $n/2$ points (the RED NET method). There is no blue-empty coloring for ODD n except for $n = 7$ (Lorden 1962).

see also COMPLETE GRAPH, EXTREMAL GRAPH, MONOCHROMATIC FORCED TRIANGLE, RED NET

References

Lorden, G. "Blue-Empty Chromatic Graphs." *Amer. Math. Monthly* **69**, 114–120, 1962.

Sauvé, L. "On Chromatic Graphs." *Amer. Math. Monthly* **68**, 107–111, 1961.

Board

A subset of $\mathbf{d} \times \mathbf{d}$, where $\mathbf{d} = \{1, 2, \ldots, d\}$.

see also ROOK NUMBER

Boatman's Knot

see CLOVE HITCH

Bochner Identity

For a smooth HARMONIC MAP $u : M \to N$,

$$\begin{aligned}\Delta(|\nabla u|^2) = |\nabla(du)|^2 &+ \langle \mathrm{Ric}_M \nabla u, \nabla u\rangle \\ &- \langle \mathrm{Riem}_N(u)(\nabla u, \nabla u)\nabla u, \nabla u\rangle,\end{aligned}$$

where ∇ is the GRADIENT, Ric is the RICCI TENSOR, and Riem is the RIEMANN TENSOR.

References

Eels, J. and Lemaire, L. "A Report on Harmonic Maps." *Bull. London Math. Soc.* **10**, 1–68, 1978.

Bochner's Theorem

Among the continuous functions on $\mathbb{R}^n$, the POSITIVE DEFINITE FUNCTIONS are those functions which are the FOURIER TRANSFORMS of finite measures.

Bode's Rule

$$\int_{x_1}^{x_5} f(x)\,dx = \tfrac{2}{45}h(7f_1 + 32f_2 + 12f_3 + 32f_4 + 7f_5) - \tfrac{8}{945}h^7 f^{(6)}(\xi).$$

see also HARDY'S RULE, NEWTON-COTES FORMULAS, SIMPSON'S 3/8 RULE, SIMPSON'S RULE, TRAPEZOIDAL RULE, WEDDLE'S RULE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 886, 1972.

Bogdanov Map

A 2-D MAP which is conjugate to the HÉNON MAP in its nondissipative limit. It is given by

$$\begin{aligned}x' &= x + y' \\ y' &= y + \epsilon y + kx(x-1) + \mu xy.\end{aligned}$$

see also HÉNON MAP

References

Arrowsmith, D. K.; Cartwright, J. H. E.; Lansbury, A. N.; and Place, C. M. "The Bogdanov Map: Bifurcations, Mode Locking, and Chaos in a Dissipative System." *Int. J. Bifurcation Chaos* **3**, 803–842, 1993.

Bogdanov, R. "Bifurcations of a Limit Cycle for a Family of Vector Fields on the Plane." *Selecta Math. Soviet* **1**, 373–388, 1981.

Bogomolov-Miyaoka-Yau Inequality

Relates invariants of a curve defined over the INTEGERS. If this inequality were proven true, then FERMAT'S LAST THEOREM would follow for sufficiently large exponents. Miyaoka claimed to have proven this inequality in 1988, but the proof contained an error.

see also FERMAT'S LAST THEOREM

References

Cox, D. A. "Introduction to Fermat's Last Theorem." *Amer. Math. Monthly* **101**, 3–14, 1994.

Bohemian Dome

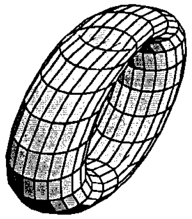

A QUARTIC SURFACE which can be constructed as follows. Given a CIRCLE C and PLANE E PERPENDICULAR to the PLANE of C, move a second CIRCLE K of the same RADIUS as C through space so that its CENTER always lies on C and it remains PARALLEL to E. Then K sweeps out the Bohemian dome. It can be given by the parametric equations

$$\begin{aligned}x &= a\cos u \\ y &= b\cos v + a\sin u \\ z &= c\sin v\end{aligned}$$

where $u, v \in [0, 2\pi)$. In the above plot, $a = 0.5$, $b = 1.5$, and $c = 1$.

see also QUARTIC SURFACE

References
Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, pp. 19–20, 1986.
Fischer, G. (Ed.). Plate 50 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 50, 1986.
Nordstrand, T. "Bohemian Dome." `http://www.uib.no/people/nfytn/bodtxt.htm`.

Bohr-Favard Inequalities

If f has no spectrum in $[-\lambda, \lambda]$, then

$$||f||_\infty \leq \frac{\pi}{2\lambda}||f'||_\infty$$

(Bohr 1935). A related inequality states that if A_k is the class of functions such that

$$f(x) = f(x + 2\pi), f(x), f'(x), \ldots, f^{(k-1)}(x)$$

are absolutely continuous and $\int_0^{2\pi} f(x)\,dx = 0$, then

$$||f||_\infty \leq \frac{4}{\pi}\sum_{\nu=0}^{\infty} \frac{(-1)^{\nu(k+1)}}{(2\nu+1)^{k+1}}||f^{(k)}(x)||_\infty$$

(Northcott 1939). Further, for each value of k, there is always a function $f(x)$ belonging to A_k and not identically zero, for which the above inequality becomes an inequality (Favard 1936). These inequalities are discussed in Mitrinovic *et al.* (1991).

References
Bohr, H. "Ein allgemeiner Satz über die Integration eines trigonometrischen Polynoms." *Prace Matem.-Fiz.* **43**, 1935.
Favard, J. "Application de la formule sommatoire d'Euler à la démonstration de quelques propriétés extrémales des intégrale des fonctions périodiques ou presquepériodiques." *Mat. Tidsskr. B*, 81–94, 1936. [Reviewed in *Zentralblatt f. Math.* **16**, 58–59, 1939.]
Mitrinovic, D. S.; Pecaric, J. E.; and Fink, A. M. *Inequalities Involving Functions and Their Integrals and Derivatives.* Dordrecht, Netherlands: Kluwer, pp. 71–72, 1991.
Northcott, D. G. "Some Inequalities Between Periodic Functions and Their Derivatives." *J. London Math. Soc.* **14**, 198–202, 1939.
Tikhomirov, V. M. "Approximation Theory." In *Analysis II* (Ed. R. V. Gamrelidze). New York: Springer-Verlag, pp. 93–255, 1990.

Bolyai-Gerwein Theorem

see WALLACE-BOLYAI-GERWEIN THEOREM

Bolza Problem

Given the functional

$$U = \int_{t_0}^{t_1} f(y_1, \ldots, y_n; y_1', \ldots, y_n')\,dt + G(y_{10}, \ldots, y_{nr}; y_{11}, \ldots, y_{n1}),$$

find in a class of arcs satisfying p differential and q finite equations

$$\phi_\alpha(y_1, \ldots, y_n; y_1', \ldots, y_n') = 0 \qquad \text{for } \alpha = 1, \ldots, p$$
$$\psi_\beta(y_1, \ldots, y_n) = 0 \qquad \text{for } \beta = 1, \ldots, q$$

as well as the r equations on the endpoints

$$\chi_\gamma(y_{10}, \ldots, y_{nr}; y_{11}, \ldots, y_{n1}) = 0 \qquad \text{for } \gamma = 1, \ldots, r,$$

one which renders U a minimum.

References
Goldstine, H. H. *A History of the Calculus of Variations from the 17th through the 19th Century.* New York: Springer-Verlag, p. 374, 1980.

Bolzano Theorem

see BOLZANO-WEIERSTRAß THEOREM

Bolzano-Weierstraß Theorem

Every BOUNDED infinite set in $\mathbb{R}^n$ has an ACCUMULATION POINT. For $n = 1$, the theorem can be stated as follows: If a SET in a METRIC SPACE, finite-dimensional EUCLIDEAN SPACE, or FIRST-COUNTABLE SPACE has infinitely many members within a finite interval $x \in [a, b]$, then it has at least one LIMIT POINT x such that $x \in [a, b]$. The theorem can be used to prove the INTERMEDIATE VALUE THEOREM.

Bombieri's Inequality

For HOMOGENEOUS POLYNOMIALS P and Q of degree m and n, then

$$[P \cdot Q]_2 \geq \sqrt{\frac{m!n!}{(m+n)!}}[P]_2[Q]_2,$$

where $[P \cdot Q]_2$ is the BOMBIERI NORM. If $m = n$, this becomes

$$[P \cdot Q]_2 \geq [P]_2[Q]_2.$$

see also BEAUZAMY AND DÉGOT'S IDENTITY, REZNIK'S IDENTITY

Bombieri Inner Product

For HOMOGENEOUS POLYNOMIALS P and Q of degree n,

$$[P, Q] \equiv \sum_{i_1, \ldots, i_n \geq 0} (i_1! \cdots i_n!)(a_{i_1,\ldots,i_n} b_{i_1,\ldots,i_n}).$$

Bombieri Norm

For HOMOGENEOUS POLYNOMIALS P of degree m,

$$[P]_2 \equiv \sqrt{[P,P]} = \left(\sum_{|\alpha|=m} \frac{\alpha!}{m!}|a_\alpha|^2\right)^2.$$

see also POLYNOMIAL BAR NORM

Bombieri's Theorem

Define

$$E(x;q,a) \equiv \psi(x;q,a) - \frac{x}{\phi(q)}, \qquad (1)$$

where

$$\psi(x;q,a) = \sum_{\substack{n\leq x\\ n\equiv a \pmod q}} \Lambda(n) \qquad (2)$$

(Davenport 1980, p. 121), $\Lambda(n)$ is the MANGOLDT FUNCTION, and $\phi(q)$ is the TOTIENT FUNCTION. Now define

$$E(x;q) = \max_{\substack{a\\(a,q)=1}} |E(x;q,a)| \qquad (3)$$

where the sum is over a RELATIVELY PRIME to q, $(a,q)=1$, and

$$E^*(x,q) = \max_{y\leq x} E(y,q). \qquad (4)$$

Bombieri's theorem then says that for $A > 0$ fixed,

$$\sum_{q\leq Q} E^*(x,q) \ll \sqrt{x}\,Q(\ln x)^5, \qquad (5)$$

provided that $\sqrt{x}\,(\ln x)^{-4} \leq Q \leq \sqrt{x}$.

References

Bombieri, E. "On the Large Sieve." *Mathematika* **12**, 201–225, 1965.

Davenport, H. "Bombieri's Theorem." Ch. 28 in *Multiplicative Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 161–168, 1980.

Bond Percolation

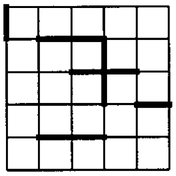

bond percolation

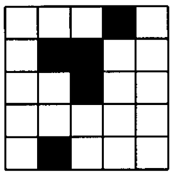

site percolation

A PERCOLATION which considers the lattice edges as the relevant entities (left figure).

see also PERCOLATION THEORY, SITE PERCOLATION

Bonferroni Correction

The Bonferroni correction is a multiple-comparison correction used when several independent STATISTICAL TESTS are being performed simultaneously (since while a given ALPHA VALUE α may be appropriate for each individual comparison, it is not for the set of *all* comparisons). In order to avoid a lot of spurious positives, the ALPHA VALUE needs to be lowered to account for the number of comparisons being performed.

The simplest and most conservative approach is the Bonferroni correction, which sets the ALPHA VALUE for the entire *set* of n comparisons equal to α by taking the ALPHA VALUE for *each* comparison equal to α/n. Explicitly, given n tests T_i for hypotheses H_i $(1 \leq i \leq n)$ under the assumption H_0 that all hypotheses H_i are false, and if the individual test critical values are $\leq \alpha/n$, then the experiment-wide critical value is $\leq \alpha$. In equation form, if

$$P(T_i \text{ passes } |H_0) \leq \frac{\alpha}{n}$$

for $1 \leq i \leq n$, then

$$P(\text{some } T_i \text{ passes } |H_0) \leq \alpha,$$

which follows from BONFERRONI'S INEQUALITY.

Another correction instead uses $1-(1-\alpha)^{1/n}$. While this choice is applicable for two-sided hypotheses, multivariate normal statistics, and positive orthant dependent statistics, it is not, in general, correct (Shaffer 1995).

see also ALPHA VALUE, HYPOTHESIS TESTING, STATISTICAL TEST

References

Bonferroni, C. E. "Il calcolo delle assicurazioni su gruppi di teste." In *Studi in Onore del Professore Salvatore Ortu Carboni.* Rome: Italy, pp. 13–60, 1935.

Bonferroni, C. E. "Teoria statistica delle classi e calcolo delle probabilità." *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* **8**, 3–62, 1936.

Dewey, M. "Carlo Emilio Bonferroni: Life and Works." `http://www.nottingham.ac.uk/~mhzmd/life.html`.

Miller, R. G. Jr. *Simultaneous Statistical Inference.* New York: Springer-Verlag, 1991.

Perneger, T. V. "What's Wrong with Bonferroni Adjustments." *Brit. Med. J.* **316**, 1236–1238, 1998.

Shaffer, J. P. "Multiple Hypothesis Testing." *Ann. Rev. Psych.* **46**, 561–584, 1995.

Bonferroni's Inequality

Let $P(E_i)$ be the probability that E_i is true, and $P\left(\bigcup_{i=1}^n E_i\right)$ be the probability that E_1, E_2, …, E_n are all true. Then

$$P\left(\bigcup_{i=1}^n E_i\right) \leq \sum_{i=1}^n P(E_i).$$

Bonferroni Test

see BONFERRONI CORRECTION

Bonne Projection

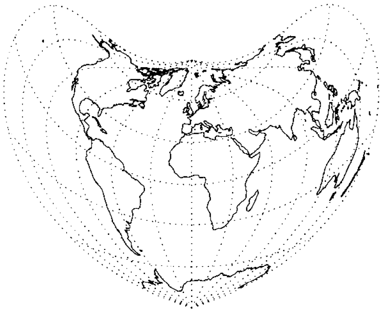

A MAP PROJECTION which resembles the shape of a heart. Let ϕ_1 be the standard parallel and λ_0 the central meridian. Then

$$x = \rho \sin E \tag{1}$$
$$y = R \cot \phi_1 - \rho \cos R, \tag{2}$$

where

$$\rho = \cot \phi_1 + \phi_1 - \phi \tag{3}$$
$$E = \frac{(\lambda - \lambda_0)\cos\phi}{\rho}. \tag{4}$$

The inverse FORMULAS are

$$\phi = \cot\phi_1 + \phi_1 - \rho \tag{5}$$
$$\lambda = \lambda_0 + \frac{\rho}{\cos\phi}\tan^{-1}\left(\frac{x}{\cot\phi_1 - y}\right), \tag{6}$$

where

$$\rho = \pm\sqrt{x^2 + (\cot\phi_1 - y)^2}. \tag{7}$$

References

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 138–140, 1987.

Book Stacking Problem

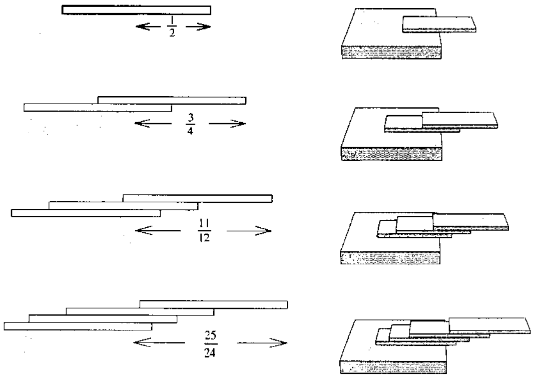

How far can a stack of n books protrude over the edge of a table without the stack falling over? It turns out that the maximum overhang possible d_n for n books (in terms of book lengths) is half the nth partial sum of the HARMONIC SERIES, given explicitly by

$$d_n = \frac{1}{2}\sum_{k=1}^{n}\frac{1}{k} = \tfrac{1}{2}[\gamma + \Psi(1+n)]$$

where $\Psi(z)$ is the DIGAMMA FUNCTION and γ is the EULER-MASCHERONI CONSTANT. The first few values are

$$d_1 = \frac{1}{2} = 0.5$$
$$d_2 = \tfrac{3}{4} = 0.75$$
$$d_3 = \tfrac{11}{12} \approx 0.91667$$
$$d_4 = \tfrac{25}{24} \approx 1.04167,$$

(Sloane's A001008 and A002805).

In order to find the number of stacked books required to obtain d book-lengths of overhang, solve the d_n equation for d, and take the CEILING FUNCTION. For $n = 1, 2, \ldots$ book-lengths of overhang, 4, 31, 227, 1674, 12367, 91380, 675214, 4989191, 36865412, 272400600, ... (Sloane's A014537) books are needed.

References

Dickau, R. M. "The Book-Stacking Problem." `http://www.prairienet.org/~pops/BookStacking.html`.

Eisner, L. "Leaning Tower of the Physical Review." *Amer. J. Phys.* **27**, 121, 1959.

Gardner, M. *Martin Gardner's Sixth Book of Mathematical Games from Scientific American.* New York: Scribner's, p. 167, 1971.

Graham, R. L.; Knuth, D. E.; and Patashnik, O. *Concrete Mathematics: A Foundation for Computer Science.* Reading, MA: Addison-Wesley, pp. 272–274, 1990.

Johnson, P. B. "Leaning Tower of Lire." *Amer. J. Phys.* **23**, 240, 1955.

Sharp, R. T. "Problem 52." *Pi Mu Epsilon J.* **1**, 322, 1953.

Sharp, R. T. "Problem 52." *Pi Mu Epsilon J.* **2**, 411, 1954.

Sloane, N. J. A. Sequences A014537, A001008/M2885, and A002805/M1589 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Boole's Inequality

$$P\left(\bigcup_{i=1}^{N} E_i\right) \leq \sum_{i=1}^{N} P(E_i).$$

If E_i and E_j are MUTUALLY EXCLUSIVE for all i and j, then the INEQUALITY becomes an equality.

Boolean Algebra

A mathematical object which is similar to a BOOLEAN RING, but which uses the meet and join operators instead of the usual addition and multiplication operators. A Boolean algebra is a set B of elements a, b, ... with BINARY OPERATORS + and $\cdot$ such that

1a. If a and b are in the set B, then $a + b$ is in the set B.

1b. If a and b are in the set B, then $a \cdot b$ is in the set B.

2a. There is an element Z (zero) such that $a + Z = a$ for every element a.

2b. There is an element U (unity) such that $a \cdot U = a$ for every element a.

3a. $a + b = b + a$

3b. $a \cdot b = b \cdot a$

4a. $a + b \cdot c = (a + b)(a + c)$

4b. $a \cdot (b + c) = a \cdot b + a \cdot c$

5. For every element a there is an element a' such that $a + a' = U$ and $a \cdot a' = Z$.

6. There are are least two distinct elements in the set B.

(Bell 1937, p. 444).

In more modern terms, a Boolean algebra is a SET B of elements a, b, ... with the following properties:

1. B has two binary operations, $\wedge$ (WEDGE) and $\vee$ (VEE), which satisfy the IDEMPOTENT laws
$$a \wedge a = a \vee a = a,$$
the COMMUTATIVE laws
$$a \wedge b = b \wedge a$$
$$a \vee b = b \vee a,$$
and the ASSOCIATIVE laws
$$a \wedge (b \wedge c) = (a \wedge b) \wedge c$$
$$a \vee (b \vee c) = (a \vee b) \vee c.$$
2. The operations satisfy the ABSORPTION LAW
$$a \wedge (a \vee b) = a \vee (a \wedge b) = a.$$
3. The operations are mutually distributive
$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$$
$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c).$$
4. B contains universal bounds O, I which satisfy
$$O \wedge a = O$$
$$O \vee a = a$$
$$I \wedge a = a$$
$$I \vee a = I.$$
5. B has a unary operation $a \to a'$ of complementation which obeys the laws
$$a \wedge a' = O$$
$$a \vee a' = I$$

(Birkhoff and Mac Lane 1965). Under intersection, union, and complement, the subsets of any set I form a Boolean algebra.

Huntington (1933a, b) presented the following basis for Boolean algebra,

1. Commutivity. $x + y = y + x$.
2. Associativity. $(x + y) + z = x + (y + z)$.
3. HUNTINGTON EQUATION. $n(n(x) + y) + n(n(x) + n(y)) = x$.

H. Robbins then conjectured that the HUNTINGTON EQUATION could be replaced with the simpler ROBBINS EQUATION,
$$n(n(x + y) + n(x + n(y))) = x.$$

The ALGEBRA defined by commutivity, associativity, and the ROBBINS EQUATION is called ROBBINS ALGEBRA. Computer theorem proving demonstrated that every ROBBINS ALGEBRA satisfies the second WINKLER CONDITION, from which it follows immediately that all ROBBINS ALGEBRAS are Boolean.

References
Bell, E. T. *Men of Mathematics.* New York: Simon and Schuster, 1986.
Birkhoff, G. and Mac Lane, S. *A Survey of Modern Algebra, 3rd ed.* New York: Macmillian, p. 317, 1965.
Halmos, P. *Lectures on Boolean Algebras.* Princeton, NJ: Van Nostrand, 1963.
Huntington, E. V. "New Sets of Independent Postulates for the Algebra of Logic." *Trans. Amer. Math. Soc.* **35**, 274–304, 1933a.
Huntington, E. V. "Boolean Algebras: A Correction." *Trans. Amer. Math. Soc.* **35**, 557–558, 1933.
McCune, W. "Robbins Algebras are Boolean." `http://www.mcs.anl.gov/~mccune/papers/robbins/`.

Boolean Connective

One of the LOGIC operators AND $\wedge$, OR $\vee$, and NOT $\neg$.

see also QUANTIFIER

Boolean Function

A Boolean function in n variables is a function
$$f(x_1, \ldots, x_n),$$
where each x_i can be 0 or 1 and f is 0 or 1. Determining the number of monotone Boolean functions of n variables is known as DEDEKIND'S PROBLEM. The number of monotonic increasing Boolean functions of n variables is given by 2, 3, 6, 20, 168, 7581, 7828354, ... (Sloane's A000372, Beeler *et al.* 1972, Item 17). The number of inequivalent monotone Boolean functions of n variables is given by 2, 3, 5, 10, 30, ... (Sloane's A003182).

Let $M(n, k)$ denote the number of distinct monotone Boolean functions of n variables with k mincuts. Then

$$M(n,0) = 1$$
$$M(n,1) = 2^n$$
$$M(n,2) = 2^{n-1}(2^n - 1) - 3^n + 2^n$$
$$M(n,3) = \tfrac{1}{6}(2^n)(2^n - 1)(2^n - 2) - 6^n + 5^n + 4^n - 3^n.$$

References
Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.
Sloane, N. J. A. Sequences A003182/M0729 and A000372/M0817 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Boolean Ring

A RING with a unit element in which every element is IDEMPOTENT.

see also BOOLEAN ALGEBRA

Borchardt-Pfaff Algorithm

see ARCHIMEDES ALGORITHM

Border Square

40	1	2	3	42	41	46
38	31	13	14	32	35	12
39	30	26	21	28	20	11
43	33	27	25	23	17	7
6	16	22	29	24	34	44
5	15	37	36	18	19	45
4	49	48	47	8	9	10

31	13	14	32	35
30	26	21	28	20
33	27	25	23	17
16	22	29	24	34
15	37	36	18	19

26	21	28
27	25	23
22	29	24

A MAGIC SQUARE that remains magic when its border is removed. A nested magic square remains magic after the border is successively removed one ring at a time. An example of a nested magic square is the order 7 square illustrated above (i.e., the order 7, 5, and 3 squares obtained from it are all magic).

see also MAGIC SQUARE

References

Kraitchik, M. "Border Squares." §7.7 in *Mathematical Recreations.* New York: W. W. Norton, pp. 167–170, 1942.

Bordism

A relation between COMPACT boundaryless MANIFOLDS (also called closed MANIFOLDS). Two closed MANIFOLDS are bordant IFF their disjoint union is the boundary of a compact $(n+1)$-MANIFOLD. Roughly, two MANIFOLDS are bordant if together they form the boundary of a MANIFOLD. The word bordism is now used in place of the original term COBORDISM.

References

Budney, R. "The Bordism Project." `http://math.cornell.edu/~rybu/bordism/bordism.html`.

Bordism Group

There are bordism groups, also called COBORDISM GROUPS or COBORDISM RINGS, and there are singular bordism groups. The bordism groups give a framework for getting a grip on the question, "When is a compact boundaryless MANIFOLD the boundary of another MANIFOLD?" The answer is, precisely when all of its STIEFEL-WHITNEY CLASSES are zero. Singular bordism groups give insight into STEENROD'S REALIZATION PROBLEM: "When can homology classes be realized as the image of fundamental classes of manifolds?" That answer is known, too.

The machinery of the bordism group winds up being important for HOMOTOPY THEORY as well.

References

Budney, R. "The Bordism Project." `http://math.cornell.edu/~rybu/bordism/bordism.html`.

Borel-Cantelli Lemma

Let $\{A_n\}_{n=0}^{\infty}$ be a SEQUENCE of events occurring with a certain probability distribution, and let A be the event consisting of the occurrence of a finite number of events A_n, $n = 1, \ldots$. Then if

$$\sum_{n=1}^{\infty} P(A_n) < \infty,$$

then

$$P(A) = 1.$$

References

Hazewinkel, M. (Managing Ed.). *Encyclopaedia of Mathematics: An Updated and Annotated Translation of the Soviet "Mathematical Encyclopaedia."* Dordrecht, Netherlands: Reidel, pp. 435–436, 1988.

Borel Determinacy Theorem

Let T be a tree defined on a metric over a set of paths such that the distance between paths p and q is $1/n$, where n is the number of nodes shared by p and q. Let A be a Borel set of paths in the topology induced by this metric. Suppose two players play a game by choosing a path down the tree, so that they alternate and each time choose an immediate successor of the previously chosen point. The first player wins if the chosen path is in A. Then one of the players has a winning STRATEGY in this GAME.

see also GAME THEORY, STRATEGY

Borel's Expansion

Let $\phi(t) = \sum_{n=0}^{\infty} A_n t^n$ be any function for which the integral

$$I(x) \equiv \int_0^{\infty} e^{-tx} t^p \phi(t)\, dt$$

converges. Then the expansion

$$I(x) = \frac{\Gamma(p+1)}{x^{p+1}} \left[A_0 + (p+1)\frac{A_1}{x} + (p+1)(p+2)\frac{A_2}{x^2} + \ldots \right],$$

where $\Gamma(z)$ is the GAMMA FUNCTION, is usually an ASYMPTOTIC SERIES for $I(x)$.

Borel Measure

If F is the BOREL SIGMA ALGEBRA on some TOPOLOGICAL SPACE, then a MEASURE $m : F \to \mathbb{R}$ is said to be a Borel measure (or BOREL PROBABILITY MEASURE). For a Borel measure, all continuous functions are MEASURABLE.

Borel Probability Measure

see BOREL MEASURE

Borel Set

A DEFINABLE SET derived from the REAL LINE by removing a FINITE number of intervals. Borel sets are measurable and constitute a special type of SIGMA ALGEBRA called a BOREL SIGMA ALGEBRA.

see also STANDARD SPACE

Borel Sigma Algebra

A SIGMA ALGEBRA which is related to the TOPOLOGY of a SET. The Borel *sigma*-algebra is defined to be the SIGMA ALGEBRA generated by the OPEN SETS (or equivalently, by the CLOSED SETS).

see also BOREL MEASURE

Borel Space

A SET equipped with a SIGMA ALGEBRA of SUBSETS.

Borromean Rings

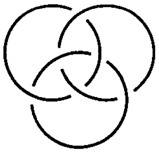

Three mutually interlocked rings named after the Italian Renaissance family who used them on their coat of arms. No two rings are linked, so if one of the rings is cut, all three rings fall apart. They are given the LINK symbol 06^{02}_{03}, and are also called the BALLANTINE. The Borromean rings have BRAID WORD $\sigma_1{}^{-1}\sigma_2\sigma_1{}^{-1}\sigma_2\sigma_1{}^{-1}\sigma_2$ and are also the simplest BRUNNIAN LINK.

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 58–59, 1989.

Gardner, M. *The Unexpected Hanging and Other Mathematical Diversions.* Chicago, IL: University of Chicago Press, 1991.

Jablan, S. "Borromean Triangles." `http://members.tripod.com/~modularity/links.htm`.

Pappas, T. "Trinity of Rings—A Topological Model." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./ Tetra, p. 31, 1989.

Borrow

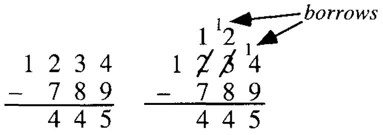

The procedure used in SUBTRACTION to "borrow" 10 from the next higher DIGIT column in order to obtain a POSITIVE DIFFERENCE in the column in question.

see also CARRY

Borsuk's Conjecture

Borsuk conjectured that it is possible to cut an n-D shape of DIAMETER 1 into $n+1$ pieces each with diameter smaller than the original. It is true for $n = 2$, 3 and when the boundary is "smooth." However, the minimum number of pieces required has been shown to increase as $\sim 1.1^{\sqrt{n}}$. Since $1.1^{\sqrt{n}} > n+1$ at $n = 9162$, the conjecture becomes false at high dimensions. In fact, the limit has been pushed back to ~ 2000.

see also DIAMETER (GENERAL), KELLER'S CONJECTURE, LEBESGUE MINIMAL PROBLEM

References

Borsuk, K. "Über die Zerlegung einer Euklidischen n-dimensionalen Vollkugel in n Mengen." *Verh. Internat. Math.-Kongr. Zürich* **2**, 192, 1932.

Borsuk, K. "Drei Sätze über die n-dimensionale euklidische Sphäre." *Fund. Math.* **20**, 177–190, 1933.

Cipra, B. "If You Can't See It, Don't Believe It...." *Science* **259**, 26–27, 1993.

Cipra, B. *What's Happening in the Mathematical Sciences, Vol. 1.* Providence, RI: Amer. Math. Soc., pp. 21–25, 1993.

Grünbaum, B. "Borsuk's Problem and Related Questions." In *Convexity, Proceedings of the Seventh Symposium in Pure Mathematics of the American Mathematical Society, Held at the University of Washington, Seattle, June 13–15, 1961.* Providence, RI: Amer. Math. Soc., pp. 271–284, 1963.

Kalai, J. K. G. "A Counterexample to Borsuk's Conjecture." *Bull. Amer. Math. Soc.* **329**, 60–62, 1993. Listernik, L. and Schnirelmann, L. *Topological Methods in Variational Problems.* Moscow, 1930.

Borwein Conjectures

Use the definition of the q-SERIES

$$(a;q)_n \equiv \prod_{j=0}^{n-1}(1-aq^j) \tag{1}$$

and define

$$\begin{bmatrix} N \\ M \end{bmatrix} \equiv \frac{(q^{N-M+1};q)_M}{(q;q)_m}. \tag{2}$$

Then P. Borwein has conjectured that (1) the POLYNOMIALS $A_n(q)$, $B_n(q)$, and $C_n(q)$ defined by

$$(q;q^3)_n(q^2;q^3)_n = A_n(q^3) - qB_n(q^3) - q^2C_n(q^3) \tag{3}$$

have NONNEGATIVE COEFFICIENTS, (2) the POLYNOMIALS $A_n^\star(q)$, $B_n^\star(q)$, and $C_n^\star(q)$ defined by

$$(q;q^3)_n^2(q^2;q^3)_n^2 = A_n^\star(q^3) - qB_n^\star(q^3) - q^2C_n^\star(q^3) \tag{4}$$

have NONNEGATIVE COEFFICIENTS, (3) the POLYNOMIALS $A_n^\star(q)$, $B_n^\star(q)$, $C_n^\star(q)$, $D_n^\star(q)$, and $E_n^\star(q)$ defined by

$$(q;q^5)_n(q^2;q^5)_n(q^3;q^5)_n(q^4;q^5)_n =$$
$$A_n^\star(q^5) - qB_n^\star(q^5) - q^2C_n^\star(q^5) - q^3D_n^\star(q^5) - q^4E_n^\star(q^5) \tag{5}$$

have NONNEGATIVE COEFFICIENTS, (4) the POLYNOMIALS $A_n^\dagger(m,n,t,q)$, $B_n^\dagger(m,n,t,q)$, and $C_n^\dagger(m,n,t,q)$ defined by

$$(q;q^3)_m(q^2;q^3)_m(zq;q^3)_n(zq^2;q^3)_n = \sum_{t=0}^{2m} z^t[A^\dagger(m,n,t,q^3) - qB^\dagger(m,n,t,q^3) - q^2C^\dagger(m,n,t,q^3)] \qquad (6)$$

have NONNEGATIVE COEFFICIENTS, (5) for k ODD and $1 \leq a \leq k/2$, consider the expansion

$$(q^a;q^k)_m(q^{k-a};q^k)_n = \sum_{\nu=(1-k)/2}^{(k-1)/2} (-1)^\nu q^{k(\nu^2+\nu)/2-a\nu}F_\nu(q^k) \qquad (7)$$

with

$$F_\nu(q) = \sum_{j=-\infty}^{\infty} (-1)^j q^{j(k^2j+2k\nu+k-2a)/2} \begin{bmatrix} m+n \\ m+\nu+kj \end{bmatrix}, \qquad (8)$$

then if a is RELATIVELY PRIME to k and $m = n$, the COEFFICIENTS of $F_\nu(q)$ are NONNEGATIVE, and (6) given $\alpha + \beta < 2K$ and $-K + \beta \leq n - m \leq K - \alpha$, consider

$$G(\alpha,\beta,K;q) = \sum_q (-1)^j q^{j[K(\alpha+\beta)j+K(\alpha+\beta)]/2} \times \begin{bmatrix} m+n \\ m+Kj \end{bmatrix}, \qquad (9)$$

the GENERATING FUNCTION for partitions inside an $m \times n$ rectangle with hook difference conditions specified by α, β, and K. Let α and β be POSITIVE RATIONAL NUMBERS and $K > 1$ an INTEGER such that αK and βK are integers. Then if $1 \leq \alpha+\beta \leq 2K-1$ (with strict inequalities for $K = 2$) and $-K+\beta \leq n-m \leq K-\alpha$, then $G(\alpha,\beta,K;q)$ has NONNEGATIVE COEFFICIENTS.

see also q-SERIES

References

Andrews, G. E. *et al.* "Partitions with Prescribed Hook Differences." *Europ. J. Combin.* **8**, 341–350, 1987.

Bressoud, D. M. "The Borwein Conjecture and Partitions with Prescribed Hook Differences." *Electronic J. Combinatorics* **3**, No. 2, R4, 1–14, 1996. `http://www.combinatorics.org/Volume_3/volume3_2.html#R4`.

Bouligand Dimension

see MINKOWSKI-BOULIGAND DIMENSION

Bound

see GREATEST LOWER BOUND, INFIMUM, LEAST UPPER BOUND, SUPREMUM

Bound Variable

An occurrence of a variable in a LOGIC which is not FREE.

Boundary

The set of points, known as BOUNDARY POINTS, which are members of the CLOSURE of a given set S and the CLOSURE of its complement set. The boundary is sometimes called the FRONTIER.

see also SURGERY

Boundary Conditions

There are several types of boundary conditions commonly encountered in the solution of PARTIAL DIFFERENTIAL EQUATIONS.

1. DIRICHLET BOUNDARY CONDITIONS specify the value of the function on a surface $T = f(\mathbf{r}, t)$.
2. NEUMANN BOUNDARY CONDITIONS specify the normal derivative of the function on a surface,

$$\frac{\partial T}{\partial n} = \hat{\mathbf{n}} \cdot \nabla T = f(\mathbf{r}, y).$$

3. CAUCHY BOUNDARY CONDITIONS specify a weighted average of first and second kinds.
4. ROBIN BOUNDARY CONDITIONS. For an elliptic partial differential equation in a region Ω, Robin boundary conditions specify the sum of αu and the normal derivative of $u = f$ at all points of the boundary of Ω, with α and f being prescribed.

see also BOUNDARY VALUE PROBLEM, DIRICHLET BOUNDARY CONDITIONS, INITIAL VALUE PROBLEM, NEUMANN BOUNDARY CONDITIONS, PARTIAL DIFFERENTIAL EQUATION, ROBIN BOUNDARY CONDITIONS

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 502–504, 1985.

Morse, P. M. and Feshbach, H. "Boundary Conditions and Eigenfunctions." Ch. 6 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 495–498 and 676–790, 1953.

Boundary Map

The MAP $H_n(X, A) \to H_{n-1}(A)$ appearing in the LONG EXACT SEQUENCE OF A PAIR AXIOM.

see also LONG EXACT SEQUENCE OF A PAIR AXIOM

Boundary Point

A point which is a member of the CLOSURE of a given set S and the CLOSURE of its complement set. If A is a subset of $\mathbb{R}^n$, then a point $\mathbf{x} \in \mathbb{R}^n$ is a boundary point of A if every NEIGHBORHOOD of $\mathbf{x}$ contains at least one point in A and at least one point not in A.

see also BOUNDARY

Boundary Set

A (symmetrical) boundary set of RADIUS r and center $\mathbf{x}_0$ is the set of all points $\mathbf{x}$ such that

$$|\mathbf{x} - \mathbf{x}_0| = r.$$

Let $\mathbf{x}_0$ be the ORIGIN. In $\mathbb{R}^1$, the boundary set is then the pair of points $x = r$ and $x = -r$. In $\mathbb{R}^2$, the boundary set is a CIRCLE. In $\mathbb{R}^3$, the boundary set is a SPHERE.

see also CIRCLE, DISK, OPEN SET, SPHERE

Boundary Value Problem

A boundary value problem is a problem, typically an ORDINARY DIFFERENTIAL EQUATION or a PARTIAL DIFFERENTIAL EQUATION, which has values assigned on the physical boundary of the DOMAIN in which the problem is specified. For example,

$$\begin{cases} \frac{\partial^2 u}{\partial t^2} - \nabla^2 u = f & \text{in } \Omega \\ u(0,t) = u_1 & \text{on } \partial\Omega \\ \frac{\partial u}{\partial t}(0,t) = u_2 & \text{on } \partial\Omega, \end{cases}$$

where $\partial\Omega$ denotes the boundary of Ω, is a boundary problem.

see also BOUNDARY CONDITIONS, INITIAL VALUE PROBLEM

References

Eriksson, K.; Estep, D.; Hansbo, P.; and Johnson, C. *Computational Differential Equations.* Lund: Studentlitteratur, 1996.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Two Point Boundary Value Problems." Ch. 17 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 745–778, 1992.

Bounded

A SET in a METRIC SPACE (X, d) is bounded if it has a FINITE diameter, i.e., there is an $R < \infty$ such that $d(x, y) \leq R$ for all $x, y \in X$. A SET in $\mathbb{R}^n$ is bounded if it is contained inside some BALL ${x_1}^2 + \ldots + {x_n}^2 \leq R^2$ of FINITE RADIUS R (Adams 1994).

see also BOUND, FINITE

References

Adams, R. A. *Calculus: A Complete Course.* Reading, MA: Addison-Wesley, p. 707, 1994.

Bounded Variation

A FUNCTION $f(x)$ is said to have bounded variation if, over the CLOSED INTERVAL $x \in [a, b]$, there exists an M such that

$$|f(x_i) - f(a)| + |f(x_2) - f(x_1)| + \ldots + |f(b) - f(x_{n-1})| \leq M$$

for all $a < x_1 < x_2 < \ldots < x_{n-1} < b$.

Bourget Function

$$\begin{aligned} J_{n,k}(z) &= \frac{1}{\pi i} \int t^{-n-1} \left(t + \frac{1}{t}\right)^k \exp\left[\tfrac{1}{2} z \left(t - \frac{1}{t}\right)\right] dt \\ &= \frac{1}{\pi} \int_0^\pi (2\cos\theta)^k \cos(n\theta - z\sin\theta)\, d\theta. \end{aligned}$$

see also BESSEL FUNCTION OF THE FIRST KIND

References

Hazewinkel, M. (Managing Ed.). *Encyclopaedia of Mathematics: An Updated and Annotated Translation of the Soviet "Mathematical Encyclopaedia."* Dordrecht, Netherlands: Reidel, p. 465, 1988.

Bourget's Hypothesis

When n is an INTEGER ≥ 0, then $J_n(z)$ and $J_{n+m}(z)$ have no common zeros other than at $z = 0$ for m an INTEGER ≥ 1, where $J_n(z)$ is a BESSEL FUNCTION OF THE FIRST KIND. The theorem has been proved true for m=1 2, 3, and 4.

References

Watson, G. N. *A Treatise on the Theory of Bessel Functions, 2nd ed.* Cambridge, England: Cambridge University Press, 1966.

Boustrophedon Transform

The boustrophedon ("ox-plowing") transform $\mathbf{b}$ of a sequence $\mathbf{a}$ is given by

$$b_n = \sum_{k=0}^{n} \binom{n}{k} a_k E_{n-k} \tag{1}$$

$$a_n = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} b_k E_{n-k} \tag{2}$$

for $n \geq 0$, where E_n is a SECANT NUMBER or TANGENT NUMBER defined by

$$\sum_{n=0}^{\infty} E_n \frac{x^n}{n!} = \sec x + \tan x. \tag{3}$$

The exponential generating functions of $\mathbf{a}$ and $\mathbf{b}$ are related by

$$\mathcal{B}(x) = (\sec x + \tan x)\mathcal{A}(x), \tag{4}$$

where the exponential generating function is defined by

$$\mathcal{A}(x) = \sum_{n=0}^{\infty} A_n \frac{x^n}{n!}. \tag{5}$$

see also ALTERNATING PERMUTATION, ENTRINGER NUMBER, SECANT NUMBER, SEIDEL-ENTRINGER-ARNOLD TRIANGLE, TANGENT NUMBER

References

Millar, J.; Sloane, N. J. A.; and Young, N. E. "A New Operation on Sequences: The Boustrophedon Transform." *J. Combin. Th. Ser. A* **76**, 44–54, 1996.

Bovinum Problema

see ARCHIMEDES' CATTLE PROBLEM

Bow

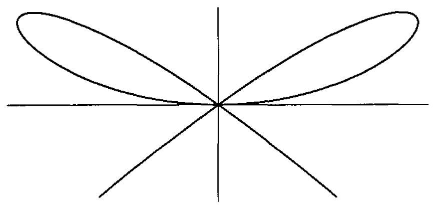

$$x^4 = x^2 y - y^3.$$

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 72, 1989.

Bowditch Curve

see LISSAJOUS CURVE

Bowley Index

The statistical INDEX

$$P_B \equiv \tfrac{1}{2}(P_L + P_P),$$

where P_L is LASPEYRES' INDEX and P_P is PAASCHE'S INDEX.

see also INDEX

References

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, p. 66, 1962.

Bowley Skewness

Also known as QUARTILE SKEWNESS COEFFICIENT,

$$\frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1)} = \frac{Q_1 - 2Q_2 + Q_3}{Q_3 - Q_1},$$

where the Qs denote the INTERQUARTILE RANGES.

see also SKEWNESS

Bowling

Bowling is a game played by rolling a heavy ball down a long narrow track and attempting to knock down ten pins arranged in the form of a TRIANGLE with its vertex oriented towards the bowler. The number 10 is, in fact, the TRIANGULAR NUMBER $T_4 = 4(4+1)/2 = 10$.

Two "bowls" are allowed per "frame." If all the pins are knocked down in the two bowls, the score for that frame is the number of pins knocked down. If some or none of the pins are knocked down on the first bowl, then all the pins knocked down on the second, it is called a "spare," and the number of points tallied is 10 plus the number of pins knocked down on the bowl of the next frame. If all of the pins are knocked down on the first bowl, the number of points tallied is 10 plus the number of pins knocked down on the next two bowls. Ten frames are bowled, unless the last frame is a strike or spare, in which case an additional bowl is awarded.

The maximum number of points possible, corresponding to knocking down all 10 pins on every bowl, is 300.

References

Cooper, C. N. and Kennedy, R. E. "A Generating Function for the Distribution of the Scores of All Possible Bowling Games." In *The Lighter Side of Mathematics* (Ed. R. K. Guy and R. E. Woodrow). Washington, DC: Math. Assoc. Amer., 1994.

Cooper, C. N. and Kennedy, R. E. "Is the Mean Bowling Score Awful?" In *The Lighter Side of Mathematics* (Ed. R. K. Guy and R. E. Woodrow). Washington, DC: Math. Assoc. Amer., 1994.

Box

see CUBOID

Box-and-Whisker Plot

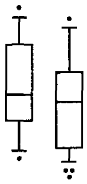

A HISTOGRAM-like method of displaying data invented by J. Tukey (1977). Draw a box with ends at the QUARTILES Q_1 and Q_3. Draw the MEDIAN as a horizontal line in the box. Extend the "whiskers" to the farthest points. For every point that is more than 3/2 times the INTERQUARTILE RANGE from the end of a box, draw a dot on the corresponding top or bottom of the whisker. If two dots have the same value, draw them side by side.

References

Tukey, J. W. *Explanatory Data Analysis.* Reading, MA: Addison-Wesley, pp. 39–41, 1977.

Box Counting Dimension

see CAPACITY DIMENSION

Box Fractal

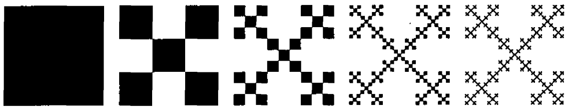

A FRACTAL which can be constructed using STRING REWRITING by creating a matrix with 3 times as many entries as the current matrix using the rules

line 1: `"*"->"* *"," "->"   "`
line 2: `"*"->" * "," "->"   "`
line 3: `"*"->"* *"," "->"   "`

Let N_n be the number of black boxes, L_n the length of a side of a white box, and A_n the fractional AREA of black boxes after the nth iteration.

$$N_n = 5^n \tag{1}$$

$$L_n = (\tfrac{1}{3})^n = 3^{-n} \tag{2}$$

$$A_n = {L_n}^2 N_n = (\tfrac{5}{9})^n. \tag{3}$$

The CAPACITY DIMENSION is therefore

$$\begin{aligned} d_{\text{cap}} &= -\lim_{n\to\infty} \frac{\ln N_n}{\ln L_n} = -\lim_{n\to\infty} \frac{\ln(5^n)}{\ln(3^{-n})} \\ &= \frac{\ln 5}{\ln 3} = 1.464973521\ldots. \end{aligned} \tag{4}$$

see also CANTOR DUST, SIERPIŃSKI CARPET, SIERPIŃSKI SIEVE

References

Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m`.

Box-Muller Transformation

A transformation which transforms from a 2-D continuous UNIFORM DISTRIBUTION to a 2-D GAUSSIAN BIVARIATE DISTRIBUTION (or COMPLEX GAUSSIAN DISTRIBUTION). If x_1 and x_2 are uniformly and independently distributed between 0 and 1, then z_1 and z_2 as defined below have a GAUSSIAN DISTRIBUTION with MEAN $\mu = 0$ and VARIANCE $\sigma^2 = 1$.

$$z_1 = \sqrt{-2\ln x_1}\cos(2\pi x_2) \tag{1}$$

$$z_2 = \sqrt{-2\ln x_1}\sin(2\pi x_2). \tag{2}$$

This can be verified by solving for x_1 and x_2,

$$x_1 = e^{-({z_1}^2+{z_2}^2)/2} \tag{3}$$

$$x_2 = \frac{1}{2\pi}\tan^{-1}\left(\frac{z_2}{z_1}\right). \tag{4}$$

Taking the JACOBIAN yields

$$\begin{aligned} \frac{\partial(x_1,x_2)}{\partial(z_1,z_2)} &= \begin{vmatrix} \frac{\partial x_1}{\partial z_1} & \frac{\partial x_1}{\partial z_2} \\ \frac{\partial x_2}{\partial z_1} & \frac{\partial x_2}{\partial z_2} \end{vmatrix} \\ &= -\left[\frac{1}{\sqrt{2\pi}}e^{-{z_1}^2/2}\right]\left[\frac{1}{\sqrt{2\pi}}e^{-{z_2}^2/2}\right]. \end{aligned} \tag{5}$$

Box-Packing Theorem

The number of "prime" boxes is always finite, where a set of boxes is prime if it cannot be built up from one or more given configurations of boxes.

see also CONWAY PUZZLE, CUBOID, DE BRUIJN'S THEOREM, KLARNER'S THEOREM, SLOTHOUBER-GRAATSMA PUZZLE

References

Honsberger, R. *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., p. 74, 1976.

Boxcar Function

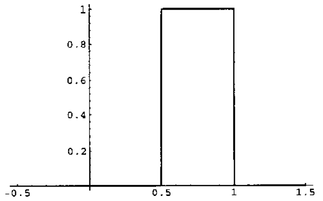

$$y = c[H(x-a) - H(x-b)],$$

where H is the HEAVISIDE STEP FUNCTION.

References

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 324, 1993.

Boxcars

A roll of two 6s (the highest roll possible) on a pair of 6-sided DICE. The probability of rolling boxcars is 1/36, or 2.777...%.

see also DICE, DOUBLE SIXES, SNAKE EYES

Boy Surface

A NONORIENTABLE SURFACE which is one of the three possible SURFACES obtained by sewing a MÖBIUS STRIP to the edge of a DISK. The other two are the CROSS-CAP and ROMAN SURFACE. The Boy surface is a model of the PROJECTIVE PLANE without singularities and is a SEXTIC SURFACE.

The Boy surface can be described using the general method for NONORIENTABLE SURFACES, but this was not known until the analytic equations were found by Apéry (1986). Based on the fact that it had been proven impossible to describe the surface using quadratic polynomials, Hopf had conjectured that quartic polynomials were also insufficient (Pinkall 1986). Apéry's IMMERSION proved this conjecture wrong, giving the equations explicitly in terms of the standard form for a NONORIENTABLE SURFACE,

$$\begin{aligned} f_1(x,y,z) = \tfrac{1}{2}[&(2x^2-y^2-z^2)(x^2+y^2+z^2) \\ &+ 2yz(y^2-z^2) + zx(x^2-z^2) \\ &+ xy(y^2-x^2)] \end{aligned} \tag{1}$$

$$\begin{aligned} f_2(x,y,z) = \tfrac{1}{2}\sqrt{3}[&(y^2-z^2)(x^2+y^2+z^2) \\ &+ zx(z^2-x^2) + xy(y^2-x^2)] \end{aligned} \tag{2}$$

$$\begin{aligned} f_3(x,y,z) = \tfrac{1}{8}(x+y+z)[&(x+y+z)^3 \\ &+ 4(y-x)(z-y)(x-z)]. \end{aligned} \tag{3}$$

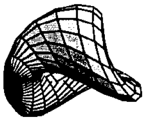

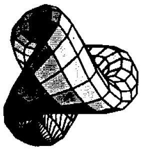

Plugging in

$$x = \cos u \sin v \tag{4}$$

$$y = \sin u \sin v \tag{5}$$

$$z = \cos v \tag{6}$$

and letting $u \in [0, \pi]$ and $v \in [0, \pi]$ then gives the Boy surface, three views of which are shown above.

The $\mathbb{R}^3$ parameterization can also be written as

$$x = \frac{\sqrt{2}\cos^2 v \cos(2u) + \cos u \sin(2v)}{2 - \sqrt{2}\sin(3u)\sin(2v)} \tag{7}$$

$$y = \frac{\sqrt{2}\cos^2 v \sin(2u) + \cos u \sin(2v)}{2 - \sqrt{2}\sin(3u)\sin(2v)} \tag{8}$$

$$z = \frac{3\cos^2 v}{2 - \sqrt{2}\sin(3u)\sin(2v)} \tag{9}$$

(Nordstrand) for $u \in [-\pi/2, \pi/2]$ and $v \in [0, \pi]$.

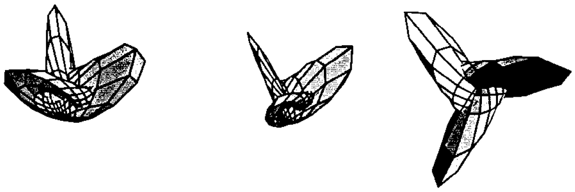

Three views of the surface obtained using this parameterization are shown above.

In fact, a HOMOTOPY (smooth deformation) between the ROMAN SURFACE and Boy surface is given by the equations

$$x(u,v) = \frac{\sqrt{2}\cos(2u)\cos^2 v + \cos u \sin(2v)}{2 - \alpha\sqrt{2}\sin(3u)\sin(2v)} \tag{10}$$

$$y(u,v) = \frac{\sqrt{2}\sin(2u)\cos^2 v - \sin u \sin(2v)}{2 - \alpha\sqrt{2}\sin(3u)\sin(2v)} \tag{11}$$

$$z(u,v) = \frac{3\cos^2 v}{2 - \alpha\sqrt{2}\sin(3u)\sin(2v)} \tag{12}$$

as α varies from 0 to 1, where $\alpha = 0$ corresponds to the ROMAN SURFACE and $\alpha = 1$ to the Boy surface (Wang), shown below.

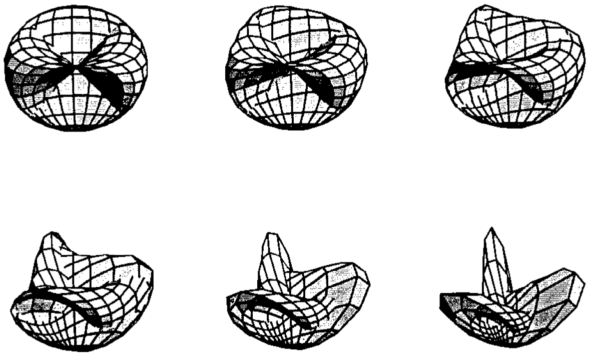

In $\mathbb{R}^4$, the parametric representation is

$$x_0 = 3[(u^2 + v^2 + w^2)(u^2 + v^2) - \sqrt{2}\,vw(3u^2 - v^2)] \tag{13}$$

$$x_1 = \sqrt{2}\,(u^2 + v^2)(u^2 - v^2 + \sqrt{2}\,uw) \tag{14}$$

$$x_2 = \sqrt{2}\,(u^2 + v^2)(2uv - \sqrt{2}\,vw) \tag{15}$$

$$x_3 = 3(u^2 + v^2)^2, \tag{16}$$

and the algebraic equation is

$$\begin{aligned} &64(x_0 - x_3)^3 x_3{}^3 - 48(x_0 - x_3)^2 x_3{}^2(3x_1{}^2 + 3x_2{}^2 + 2x_3{}^2) \\ &\quad +12(x_0 - x_3)x_3[27(x_1{}^2 + x_2{}^2)^2 - 24x_3{}^2(x_1{}^2 + x_2{}^2) \\ &\quad +36\sqrt{2}x_2x_3(x_2{}^2 - 3x_1{}^2) + x_3{}^4] \\ &\quad +(9x_1{}^2 + 9x_2{}^2 - 2x_3{}^2) \\ &\qquad \times[-81(x_1{}^2 + x_2{}^2)^2 - 72x_3{}^2(x_1{}^2 + x_2{}^2) \\ &\qquad +108\sqrt{2}x_1x_3(x_1{}^2 - 3x_2{}^2) + 4x_3{}^4] = 0 \end{aligned} \tag{17}$$

(Apéry 1986). Letting

$$x_0 = 1 \tag{18}$$

$$x_1 = x \tag{19}$$

$$x_2 = y \tag{20}$$

$$x_3 = z \tag{21}$$

gives another version of the surface in $\mathbb{R}^3$.

see also CROSS-CAP, IMMERSION, MÖBIUS STRIP, NONORIENTABLE SURFACE, REAL PROJECTIVE PLANE, ROMAN SURFACE, SEXTIC SURFACE

References

Apéry, F. "The Boy Surface." *Adv. Math.* **61**, 185–266, 1986.

Boy, W. "Über die Curvatura integra und die Topologie geschlossener Flächen." *Math. Ann* **57**, 151–184, 1903.

Brehm, U. "How to Build Minimal Polyhedral Models of the Boy Surface." *Math. Intell.* **12**, 51–56, 1990.

Carter, J. S. "On Generalizing Boy Surface—Constructing a Generator of the 3rd Stable Stem." *Trans. Amer. Math. Soc.* **298**, 103–122, 1986.

Fischer, G. (Ed.). Plates 115–120 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 110–115, 1986.

Geometry Center. "Boy's Surface." `http://www.geom.umn.edu/zoo/toptype/pplane/boy/`.

Hilbert, D. and Cohn-Vossen, S. §46–47 in *Geometry and the Imagination.* New York: Chelsea, 1952.

Nordstrand, T. "Boy's Surface." `http://www.uib.no/people/nfytn/boytxt.htm`.

Petit, J.-P. and Souriau, J. "Une représentation analytique de la surface de Boy." *C. R. Acad. Sci. Paris Sér. 1 Math* **293**, 269–272, 1981.

Pinkall, U. *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 64–65, 1986.

Stewart, I. *Game, Set and Math.* New York: Viking Penguin, 1991.

Wang, P. "Renderings." `http://www.ugcs.caltech.edu/~peterw/portfolio/renderings/`.

Bra

A (COVARIANT) 1-VECTOR denoted $\langle\psi|$. The bra is DUAL to the CONTRAVARIANT KET, denoted $|\psi\rangle$. Taken together, the bra and KET form an ANGLE BRACKET (bra+ket = bracket). The bra is commonly encountered in quantum mechanics.

see also ANGLE BRACKET, BRACKET PRODUCT, COVARIANT VECTOR, DIFFERENTIAL k-FORM, KET, ONE-FORM

Brachistochrone Problem

Find the shape of the CURVE down which a bead sliding from rest and ACCELERATED by gravity will slip (without friction) from one point to another in the least time. This was one of the earliest problems posed in the CALCULUS OF VARIATIONS. The solution, a segment of a CYCLOID, was found by Leibniz, L'Hospital, Newton, and the two Bernoullis.

The time to travel from a point P_1 to another point P_2 is given by the INTEGRAL

$$t_{12} = \int_1^2 \frac{ds}{v}. \tag{1}$$

The VELOCITY at any point is given by a simple application of energy conservation equating kinetic energy to gravitational potential energy,

$$\tfrac{1}{2}mv^2 = mgy, \tag{2}$$

so

$$v = \sqrt{2gy}. \tag{3}$$

Plugging this into (1) then gives

$$t_{12} = \int_1^2 \frac{\sqrt{1+y'^2}}{\sqrt{2gy}}\,dx = \int_1^2 \sqrt{\frac{1+y'^2}{2gy}}\,dx. \tag{4}$$

The function to be varied is thus

$$f = (1+y'^2)^{1/2}(2gy)^{-1/2}. \tag{5}$$

To proceed, one would normally have to apply the full-blown EULER-LAGRANGE DIFFERENTIAL EQUATION

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = 0. \tag{6}$$

However, the function $f(y, y', x)$ is particularly nice since x does not appear explicitly. Therefore, $\partial f/\partial x = 0$, and we can immediately use the BELTRAMI IDENTITY

$$f - y'\frac{\partial f}{\partial y'} = C. \tag{7}$$

Computing

$$\frac{\partial f}{\partial y'} = y'(1+y'^2)^{-1/2}(2gy)^{-1/2}, \tag{8}$$

subtracting $y'(\partial f/\partial y')$ from f, and simplifying then gives

$$\frac{1}{\sqrt{2gy}\sqrt{1+y'^2}} = C. \tag{9}$$

Squaring both sides and rearranging slightly results in

$$\left[1+\left(\frac{dy}{dx}\right)^2\right]y = \frac{1}{2gC^2} = k^2, \tag{10}$$

where the square of the old constant C has been expressed in terms of a new (POSITIVE) constant k^2. This equation is solved by the parametric equations

$$x = \tfrac{1}{2}k^2(\theta - \sin\theta) \tag{11}$$

$$y = \tfrac{1}{2}k^2(1 - \cos\theta), \tag{12}$$

which are—lo and behold—the equations of a CYCLOID.

If kinetic friction is included, the problem can also be solved analytically, although the solution is significantly messier. In that case, terms corresponding to the normal component of weight and the normal component of the ACCELERATION (present because of path CURVATURE) must be included. Including both terms requires a constrained variational technique (Ashby *et al.* 1975), but including the normal component of weight only gives an elementary solution. The TANGENT and NORMAL VECTORS are

$$\mathbf{T} = \frac{dx}{ds}\hat{\mathbf{x}} + \frac{dy}{ds}\hat{\mathbf{y}} \tag{13}$$

$$\mathbf{N} = -\frac{dy}{ds}\hat{\mathbf{x}} + \frac{dx}{ds}\hat{\mathbf{y}}, \tag{14}$$

gravity and friction are then

$$\mathbf{F}_{\text{gravity}} = mg\hat{\mathbf{y}} \tag{15}$$

$$\mathbf{F}_{\text{friction}} = -\mu(\mathbf{F}_{\text{gravity}}\dot{\mathbf{N}})\mathbf{T} = -\mu mg\frac{dx}{ds}\mathbf{T}, \tag{16}$$

and the components along the curve are

$$\mathbf{F}_{\text{gravity}}\dot{\mathbf{T}} = mg\frac{dy}{ds} \tag{17}$$

$$\mathbf{F}_{\text{friction}}\dot{\mathbf{T}} = -\mu mg\frac{dx}{ds}, \tag{18}$$

so Newton's Second Law gives

$$m\frac{dv}{dt} = mg\frac{dy}{ds} - \mu mg\frac{dx}{ds}. \tag{19}$$

But

$$\frac{dv}{dt} = v\frac{dv}{ds} = \frac{1}{2}\frac{d}{ds}(v^2) \tag{20}$$

$$\tfrac{1}{2}v^2 = g(y - \mu x) \tag{21}$$

$$v = \sqrt{2g(y - \mu x)}, \tag{22}$$

so

$$t=\int\sqrt{\frac{1+(y')^2}{2g(y-\mu x)}}\,dx. \tag{23}$$

Using the EULER-LAGRANGE DIFFERENTIAL EQUATION gives

$$[1+y'^2](1+\mu y')+2(y-\mu x)y''=0. \tag{24}$$

This can be reduced to

$$\frac{1+(y')^2}{(1+\mu y')^2}=\frac{C}{y-\mu x}. \tag{25}$$

Now letting

$$y'=\cot(\tfrac{1}{2}\theta), \tag{26}$$

the solution is

$$x=\tfrac{1}{2}k^2[(\theta-\sin\theta)+\mu(1-\cos\theta)] \tag{27}$$

$$y=\tfrac{1}{2}k^2[(1-\cos\theta)+\mu(\theta+\sin\theta)]. \tag{28}$$

see also CYCLOID, TAUTOCHRONE PROBLEM

References

Ashby, N.; Brittin, W. E.; Love, W. F.; and Wyss, W. "Brachistochrone with Coulomb Friction." *Amer. J. Phys.* **43**, 902–905, 1975.

Haws, L. and Kiser, T. "Exploring the Brachistochrone Problem." *Amer. Math. Monthly* **102**, 328–336, 1995.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 60–66 and 385–389, 1991.

Bracket

see ANGLE BRACKET, BRA, BRACKET POLYNOMIAL, BRACKET PRODUCT, IVERSON BRACKET, KET, LAGRANGE BRACKET, POISSON BRACKET

Bracket Polynomial

A one-variable KNOT POLYNOMIAL related to the JONES POLYNOMIAL. The bracket polynomial, however, is *not* a topological invariant, since it is changed by type I REIDEMEISTER MOVES. However, the SPAN of the bracket polynomial is a knot invariant. The bracket polynomial is occasionally given the grandiose name REGULAR ISOTOPY INVARIANT. It is defined by

$$\langle L\rangle(A,B,d)\equiv\sum_{\sigma}\langle L|\sigma\rangle\,d^{||\sigma||}, \tag{1}$$

where A and B are the "splitting variables," σ runs through all "states" of L obtained by SPLITTING the LINK, $\langle L|\sigma\rangle$ is the product of "splitting labels" corresponding to σ, and

$$||\sigma||\equiv N_L-1, \tag{2}$$

where N_L is the number of loops in σ. Letting

$$B=A^{-1} \tag{3}$$

$$d=-A^2-A^{-2} \tag{4}$$

gives a KNOT POLYNOMIAL which is invariant under REGULAR ISOTOPY, and normalizing gives the KAUFFMAN POLYNOMIAL X which is invariant under AMBIENT ISOTOPY. The bracket POLYNOMIAL of the UNKNOT is 1. The bracket POLYNOMIAL of the MIRROR IMAGE K^* is the same as for K but with A replaced by A^{-1}. In terms of the one-variable KAUFFMAN POLYNOMIAL X, the two-variable KAUFFMAN POLYNOMIAL F and the JONES POLYNOMIAL V,

$$X(A)=(-A^3)^{-w(L)}\langle L\rangle, \tag{5}$$

$$\langle L\rangle(A)=F(-A^3,A+A^{-1}) \tag{6}$$

$$\langle L\rangle(A)=V(A^{-4}), \tag{7}$$

where $w(L)$ is the WRITHE of L.

see also SQUARE BRACKET POLYNOMIAL

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 148–155, 1994.

Kauffman, L. "New Invariants in the Theory of Knots." *Amer. Math. Monthly* **95**, 195–242, 1988.

Kauffman, L. *Knots and Physics.* Teaneck, NJ: World Scientific, pp. 26–29, 1991.

Weisstein, E. W. "Knots and Links." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Knots.m`.

Bracket Product

The INNER PRODUCT in an L_2 SPACE represented by an ANGLE BRACKET.

see also ANGLE BRACKET, BRA, KET, L_2 SPACE, ONE-FORM

Bracketing

Take x itself to be a bracketing, then recursively define a bracketing as a sequence $B=(B_1,\ldots,B_k)$ where $k\geq 2$ and each B_i is a bracketing. A bracketing can be represented as a parenthesized string of xs, with parentheses removed from any single letter x for clarity of notation (Stanley 1997). Bracketings built up of binary operations only are called BINARY BRACKETINGS. For example, four letters have 11 possible bracketings:

$$\begin{array}{cccc} xxxx & (xx)xx & x(xx)x & xx(xx) \\ (xxx)x & x(xxx) & ((xx)x)x & (x(xx))x \\ (xx)(xx) & x((xx)x) & x(x(xx)), & \end{array}$$

the last five of which are binary.

The number of bracketings on n letters is given by the GENERATING FUNCTION

$$\tfrac{1}{4}(1+x-\sqrt{1-6x+x^2})=x+x^2+3x^3+11x^4+45x^5$$

(Schröder 1870, Stanley 1997) and the RECURRENCE RELATION

$$s_n=\frac{3(2n-3)s_{n-1}-(n-3)s_{n-2}}{n}$$

(Sloane), giving the sequence for s_n as 1, 1, 3, 11, 45, 197, 903, ... (Sloane's A001003). The numbers are also given by

$$s_n = \sum_{i_1+\ldots+i_k=n} s(i_1)\cdots s(i_k)$$

for $n \geq 2$ (Stanley 1997).

The first PLUTARCH NUMBER 103,049 is equal to s_{10} (Stanley 1997), suggesting that Plutarch's problem of ten compound propositions is equivalent to the number of bracketings. In addition, Plutarch's second number 310,954 is given by $(s_{10} + s_{11})/2 = 310,954$ (Habsieger *et al.* 1998).

see also BINARY BRACKETING, PLUTARCH NUMBERS

References
Habsieger, L.; Kazarian, M.; and Lando, S. "On the Second Number of Plutarch." *Amer. Math. Monthly* **105**, 446, 1998.
Schröder, E. "Vier combinatorische Probleme." *Z. Math. Physik* **15**, 361–376, 1870.
Sloane, N. J. A. Sequence A001003/M2898 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Stanley, R. P. "Hipparchus, Plutarch, Schröder, and Hough." *Amer. Math. Monthly* **104**, 344–350, 1997.

Bradley's Theorem

Let

$$S(\alpha,\beta,m;z) \equiv m\sum_{j=0}^{\infty} \frac{\Gamma(m+j(z+1))\Gamma(\beta+1+jz)}{\Gamma(m+jz+1)\Gamma(\alpha+\beta+1+j(z+1))} \frac{(\alpha)+j}{j!}$$

and α be a NEGATIVE INTEGER. Then

$$S(\alpha,\beta,m;z) = \frac{\Gamma(\beta+1-m)}{\Gamma(\alpha+\beta+1-m)},$$

where $\Gamma(z)$ is the GAMMA FUNCTION.

References
Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 346–348, 1994.
Bradley, D. "On a Claim by Ramanujan about Certain Hypergeometric Series." *Proc. Amer. Math. Soc.* **121**, 1145–1149, 1994.

Brahmagupta's Formula

For a QUADRILATERAL with sides of length a, b, c, and d, the AREA K is given by

$$K = \sqrt{(s-a)(s-b)(s-c)(s-d) - abcd\cos^2[\tfrac{1}{2}(A+B)]}, \tag{1}$$

where

$$s \equiv \tfrac{1}{2}(a+b+c+d) \tag{2}$$

is the SEMIPERIMETER, A is the ANGLE between a and d, and B is the ANGLE between b and c. For a CYCLIC QUADRILATERAL (i.e., a QUADRILATERAL inscribed in a CIRCLE), $A + B = \pi$, so

$$K = \sqrt{(s-a)(s-b)(s-c)(s-d)} \tag{3}$$

$$= \frac{\sqrt{(bc+ad)(ac+bd)(ab+cd)}}{4R}, \tag{4}$$

where R is the RADIUS of the CIRCUMCIRCLE. If the QUADRILATERAL is INSCRIBED in one CIRCLE and CIRCUMSCRIBED on another, then the AREA FORMULA simplifies to

$$K = \sqrt{abcd}. \tag{5}$$

see also BRETSCHNEIDER'S FORMULA, HERON'S FORMULA

References
Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 56–60, 1967.
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 81–82, 1929.

Brahmagupta Identity

Let

$$\beta \equiv |B| = x^2 - ty^2,$$

where B is the BRAHMAGUPTA MATRIX, then

$$\det[B(x_1,y_1)B(x_2,y_2)] = \det[B(x_1,y_1)]\det[B(x_2,y_2)] = \beta_1\beta_2.$$

References
Suryanarayan, E. R. "The Brahmagupta Polynomials." *Fib. Quart.* **34**, 30–39, 1996.

Brahmagupta Matrix

$$B(x,y) = \begin{bmatrix} x & y \\ \pm ty & \pm x \end{bmatrix}.$$

It satisfies

$$B(x_1,y_1)B(x_2,y_2) = B(x_1x_2 \pm ty_1y_2, x_1y_2 \pm y_1x_2).$$

Powers of the matrix are defined by

$$B^n = \begin{bmatrix} x & y \\ ty & x \end{bmatrix}^n = \begin{bmatrix} x_n & y_n \\ ty_n & x_n \end{bmatrix} \equiv B_n.$$

The x_n and y_n are called BRAHMAGUPTA POLYNOMIALS. The Brahmagupta matrices can be extended to NEGATIVE INTEGERS

$$B^{-n} = \begin{bmatrix} x & y \\ ty & x \end{bmatrix}^{-n} = \begin{bmatrix} x_{-n} & y_{-n} \\ ty_{-n} & x_{-n} \end{bmatrix} \equiv B_{-n}.$$

see also BRAHMAGUPTA IDENTITY

References
Suryanarayan, E. R. "The Brahmagupta Polynomials." *Fib. Quart.* **34**, 30–39, 1996.

Brahmagupta Polynomial

One of the POLYNOMIALS obtained by taking POWERS of the BRAHMAGUPTA MATRIX. They satisfy the recurrence relation

$$x_{n+1} = xx_n + tyy_n \qquad (1)$$

$$y_{n+1} = xy_n + yx_n. \qquad (2)$$

A list of many others is given by Suryanarayan (1996). Explicitly,

$$x_n = x^n + t\binom{n}{2}x^{n-2}y^2 + t^2\binom{n}{4}x^{n-4}y^4 + \ldots \qquad (3)$$

$$y_n = nx^{n-1}y + t\binom{n}{3}x^{n-3}y^3 + t^2\binom{n}{5}x^{n-5}y^5 + \ldots. \qquad (4)$$

The Brahmagupta POLYNOMIALS satisfy

$$\frac{\partial x_n}{\partial x} = \frac{\partial y_n}{\partial y} = nx_{n-1} \qquad (5)$$

$$\frac{\partial x_n}{\partial y} = t\frac{\partial y_n}{\partial y} = nty_{n-1}. \qquad (6)$$

The first few POLYNOMIALS are

$$\begin{aligned} x_0 &= 0 \\ x_1 &= x \\ x_2 &= x^2 + ty^2 \\ x_3 &= x^3 + 3txy^2 \\ x_4 &= x^4 + 6tx^2y^2 + t^2y^4 \end{aligned}$$

and

$$\begin{aligned} y_0 &= 0 \\ y_1 &= y \\ y_2 &= 2xy \\ y_3 &= 3x^2y + ty^3 \\ y_4 &= 4x^3y + 4txy^3. \end{aligned}$$

Taking $x = y = 1$ and $t = 2$ gives y_n equal to the PELL NUMBERS and x_n equal to half the Pell-Lucas numbers. The Brahmagupta POLYNOMIALS are related to the MORGAN-VOYCE POLYNOMIALS, but the relationship given by Suryanarayan (1996) is incorrect.

References

Suryanarayan, E. R. "The Brahmagupta Polynomials." *Fib. Quart.* **34**, 30–39, 1996.

Brahmagupta's Problem

Solve the PELL EQUATION

$$x^2 - 92y^2 = 1$$

in INTEGERS. The smallest solution is $x = 1151$, $y = 120$.

see also DIOPHANTINE EQUATION, PELL EQUATION

Braid

An intertwining of strings attached to top and bottom "bars" such that each string never "turns back up." In other words, the path of a braid in something that a falling object could trace out if acted upon only by gravity and horizontal forces.

see also BRAID GROUP

References

Christy, J. "Braids." `http://www.mathsource.com/cgi-bin/MathSource/Applications/Mathematics/0202-228`.

Braid Group

Also called ARTIN BRAID GROUPS. Consider n strings, each oriented vertically from a lower to an upper "bar." If this is the least number of strings needed to make a closed braid representation of a LINK, n is called the BRAID INDEX. Now enumerate the possible braids in a group, denoted B_n. A general n-braid is constructed by iteratively applying the σ_i ($i = 1, \ldots, n-1$) operator, which switches the lower endpoints of the ith and $(i+1)$th strings—keeping the upper endpoints fixed—with the $(i+1)$th string brought *above* the ith string. If the $(i+1)$th string passes *below* the ith string, it is denoted σ_i^{-1}.

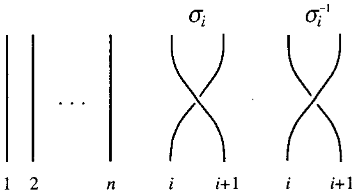

Topological equivalence for different representations of a BRAID WORD $\prod_i \sigma_i$ and $\prod_i \sigma_i'$ is guaranteed by the conditions

$$\begin{cases} \sigma_i\sigma_j = \sigma_j'\sigma_i' & \text{for } |i-j| \geq 2 \\ \sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}'\sigma_i'\sigma_{i+1}' & \text{for all } i \end{cases}$$

as first proved by E. Artin. Any n-braid is expressed as a BRAID WORD, e.g., $\sigma_1\sigma_2\sigma_3\sigma_2^{-1}\sigma_1$ is a BRAID WORD for the braid group B_3. When the opposite ends of the braids are connected by nonintersecting lines, KNOTS are formed which are identified by their braid group and BRAID WORD. The BURAU REPRESENTATION gives a matrix representation of the braid groups.

References

Birman, J. S. "Braids, Links, and the Mapping Class Groups." *Ann. Math. Studies*, No. 82. Princeton, NJ: Princeton University Press, 1976.

Birman, J. S. "Recent Developments in Braid and Link Theory." *Math. Intell.* **13**, 52–60, 1991.

Christy, J. "Braids." `http://www.mathsource.com/cgi-bin/MathSource/Applications/Mathematics/0202-228`.

Jones, V. F. R. "Hecke Algebra Representations of Braid Groups and Link Polynomials." *Ann. Math.* **126**, 335–388, 1987.

Weisstein, E. W. "Knots and Links." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Knots.m`.

Braid Index

The least number of strings needed to make a closed braid representation of a LINK. The braid index is equal to the least number of SEIFERT CIRCLES in any projection of a KNOT (Yamada 1987). Also, for a nonsplittable LINK with CROSSING NUMBER $c(L)$ and braid index $i(L)$,

$$c(L) \geq 2[i(L) - 1]$$

(Ohyama 1993). Let E be the largest and e the smallest POWER of ℓ in the HOMFLY POLYNOMIAL of an oriented LINK, and i be the braid index. Then the MORTON-FRANKS-WILLIAMS INEQUALITY holds,

$$i \geq \tfrac{1}{2}(E - e) + 1$$

(Franks and Williams 1987). The inequality is sharp for all PRIME KNOTS up to 10 crossings with the exceptions of 09_{042}, 09_{049}, 10_{132}, 10_{150}, and 10_{156}.

References
Franks, J. and Williams, R. F. "Braids and the Jones Polynomial." *Trans. Amer. Math. Soc.* **303**, 97–108, 1987.
Jones, V. F. R. "Hecke Algebra Representations of Braid Groups and Link Polynomials." *Ann. Math.* **126**, 335–388, 1987.
Ohyama, Y. "On the Minimal Crossing Number and the Brad Index of Links." *Canad. J. Math.* **45**, 117–131, 1993.
Yamada, S. "The Minimal Number of Seifert Circles Equals the Braid Index of a Link." *Invent. Math.* **89**, 347–356, 1987.

Braid Word

Any n-braid is expressed as a braid word, e.g., $\sigma_1\sigma_2\sigma_3\sigma_2^{-1}\sigma_1$ is a braid word for the BRAID GROUP B_3. By ALEXANDER'S THEOREM, any LINK is representable by a closed braid, but there is no general procedure for reducing a braid word to its simplest form. However, MARKOV'S THEOREM gives a procedure for identifying different braid words which represent the same LINK.

Let b_+ be the sum of POSITIVE exponents, and b_- the sum of NEGATIVE exponents in the BRAID GROUP B_n. If

$$b_+ - 3b_- - n + 1 > 0,$$

then the closed braid b is not AMPHICHIRAL (Jones 1985).

see also BRAID GROUP

References
Jones, V. F. R. "A Polynomial Invariant for Knots via von Neumann Algebras." *Bull. Amer. Math. Soc.* **12**, 103–111, 1985.
Jones, V. F. R. "Hecke Algebra Representations of Braid Groups and Link Polynomials." *Ann. Math.* **126**, 335–388, 1987.

Braikenridge-Maclaurin Construction

The converse of PASCAL'S THEOREM. Let A_1, B_2, C_1, A_2, and B_1 be the five points on a CONIC. Then the CONIC is the LOCUS of the point

$$C_2 = A_1(z \cdot C_1A_2) \cdot B_1(z \cdot C_1B_2),$$

where z is a line through the point $A_1B_2 \cdot B_1A_2$.

see also PASCAL'S THEOREM

Branch

The segments of a TREE between the points of connection (FORKS).

see also FORK, LEAF (TREE)

Branch Cut

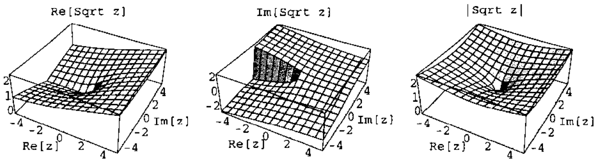

A line in the COMPLEX PLANE across which a FUNCTION is discontinuous.

function	branch cut(s)
$\cos^{-1} z$	$(-\infty, -1)$ and $(1, \infty)$
$\cosh^{-1}$	$(-\infty, 1)$
$\cot^{-1} z$	$(-i, i)$
$\coth^{-1}$	$[-1, 1]$
$\csc^{-1} z$	$(-1, 1)$
csch^{-1}	$(-i, i)$
$\ln z$	$(-\infty, 0]$
$\sec^{-1} z$	$(-1, 1)$
sech^{-1}	$(\infty, 0]$ and $(1, \infty)$
$\sin^{-1} z$	$(-\infty, -1)$ and $(1, \infty)$
$\sinh^{-1}$	$(-i\infty, -i)$ and $(i, i\infty)$
$\sqrt{z}$	$(-\infty, 0)$
$\tan^{-1} z$	$(-i\infty, -i)$ and $(i, i\infty)$
$\tanh^{-1}$	$(-\infty, -1]$ and $[1, \infty)$
$z^n, n \notin \mathbb{Z}$	$(-\infty, 0)$ for $\Re[n] \leq 0$; $(-\infty, 0]$ for $\Re[n] > 0$

see also BRANCH POINT

References
Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 399–401, 1953.

Branch Line

see BRANCH CUT

Branch Point

An argument at which identical points in the COMPLEX PLANE are mapped to different points. For example, consider

$$f(z) = z^a.$$

Then $f(e^{0i}) = f(1) = 1$, but $f(e^{2\pi i}) = e^{2\pi ia}$, despite the fact that $e^{i0} = e^{2\pi i}$. PINCH POINTS are also called branch points.

see also BRANCH CUT, PINCH POINT

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 397–399, 1985.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 391–392 and 399–401, 1953.

Brauer Chain

A Brauer chain is an ADDITION CHAIN in which each member uses the previous member as a summand. A number n for which a shortest chain exists which is a Brauer chain is called a BRAUER NUMBER.

see also ADDITION CHAIN, BRAUER NUMBER, HANSEN CHAIN

References

Guy, R. K. "Addition Chains. Brauer Chains. Hansen Chains." §C6 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 111–113, 1994.

Brauer Group

The GROUP of classes of finite dimensional central simple ALGEBRAS over k with respect to a certain equivalence.

References

Hazewinkel, M. (Managing Ed.). *Encyclopaedia of Mathematics: An Updated and Annotated Translation of the Soviet "Mathematical Encyclopaedia."* Dordrecht, Netherlands: Reidel, p. 479, 1988.

Brauer Number

A number n for which a shortest chain exists which is a BRAUER CHAIN is called a Brauer number. There are infinitely many non-Brauer numbers.

see also BRAUER CHAIN, HANSEN NUMBER

References

Guy, R. K. "Addition Chains. Brauer Chains. Hansen Chains." §C6 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 111–113, 1994.

Brauer-Severi Variety

An ALGEBRAIC VARIETY over a FIELD K that becomes ISOMORPHIC to a PROJECTIVE SPACE.

References

Hazewinkel, M. (Managing Ed.). *Encyclopaedia of Mathematics: An Updated and Annotated Translation of the Soviet "Mathematical Encyclopaedia."* Dordrecht, Netherlands: Reidel, pp. 480–481, 1988.

Brauer's Theorem

If, in the GERŜGORIN CIRCLE THEOREM for a given m,

$$|a_{jj} - a_{mm}| > \Lambda_j + \Lambda_m$$

for all $j \neq m$, then exactly one EIGENVALUE of A lies in the DISK Γ_m.

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1121, 1979.

Braun's Conjecture

Let $B = \{b_1, b_2, \ldots\}$ be an INFINITE Abelian SEMIGROUP with linear order $b_1 < b_2 < \ldots$ such that b_1 is the unit element and $a < b$ IMPLIES $ac < bc$ for $a, b, c \in B$. Define a MÖBIUS FUNCTION μ on B by $\mu(b_1) = 1$ and

$$\sum_{b_d | b_n} \mu(b_d) = 0$$

for $n = 2, 3, \ldots$. Further suppose that $\mu(b_n) = \mu(n)$ (the true MÖBIUS FUNCTION) for all $n \geq 1$. Then Braun's conjecture states that

$$b_{mn} = b_m b_n$$

for all $m, n \geq 1$.

see also MÖBIUS PROBLEM

References

Flath, A. and Zulauf, A. "Does the Möbius Function Determine Multiplicative Arithmetic?" *Amer. Math. Monthly* **102**, 354–256, 1995.

Breeder

A pair of POSITIVE INTEGERS (a_1, a_2) such that the equations

$$a_1 + a_2 x = \sigma(a_1) = \sigma(a_2)(x + 1)$$

have a POSITIVE INTEGER solution x, where $\sigma(n)$ is the DIVISOR FUNCTION. If x is PRIME, then $(a_1, a_2 x)$ is an AMICABLE PAIR (te Riele 1986). (a_1, a_2) is a "special" breeder if

$$a_1 = au$$
$$a_2 = a,$$

where a and u are RELATIVELY PRIME, $(a, u) = 1$. If regular amicable pairs of type $(i, 1)$ with $i \geq 2$ are of the form (au, ap) with p PRIME, then (au, a) are special breeders (te Riele 1986).

References

te Riele, H. J. J. "Computation of All the Amicable Pairs Below 10^{10}." *Math. Comput.* **47**, 361–368 and S9–S35, 1986.

Brelaz's Heuristic Algorithm

An ALGORITHM which can be used to find a good, but not necessarily minimal, EDGE or VERTEX coloring for a GRAPH.

see also CHROMATIC NUMBER

Brent's Factorization Method

A modification of the POLLARD ρ FACTORIZATION METHOD which uses

$$x_{i+1} = {x_i}^2 - c \pmod{n}.$$

References

Brent, R. "An Improved Monte Carlo Factorization Algorithm." *Nordisk Tidskrift for Informationsbehandlung (BIT)* **20**, 176–184, 1980.

Brent's Method

A ROOT-finding ALGORITHM which combines root bracketing, bisection, and INVERSE QUADRATIC INTERPOLATION. It is sometimes known as the VAN WIJNGAARDEN-DEKER-BRENT METHOD.

Brent's method uses a LAGRANGE INTERPOLATING POLYNOMIAL of degree 2. Brent (1973) claims that this method will always converge as long as the values of the function are computable within a given region containing a ROOT. Given three points x_1, x_2, and x_3, Brent's method fits x as a quadratic function of y, then uses the interpolation formula

$$\begin{aligned} x = {} & \frac{[y - f(x_1)][y - f(x_2)]x_3}{[f(x_3) - f(x_1)][f(x_3) - f(x_2)]} \\ & + \frac{[y - f(x_2)][y - f(x_3)]x_1}{[f(x_1) - f(x_2)][f(x_1) - f(x_3)]} \\ & + \frac{[y - f(x_3)][y - f(x_1)]x_2}{[f(x_2) - f(x_3)][f(x_2) - f(x_1)]}. \end{aligned} \tag{1}$$

Subsequent root estimates are obtained by setting $y = 0$, giving

$$x = b + \frac{P}{Q}, \tag{2}$$

where

$$P = S[R(R - T)(x_3 - x_2) - (1 - R)(x_2 - x_1)] \tag{3}$$

$$Q = (T - 1)(R - 1)(S - 1) \tag{4}$$

with

$$R \equiv \frac{f(x_2)}{f(x_3)} \tag{5}$$

$$S \equiv \frac{f(x_2)}{f(x_1)} \tag{6}$$

$$T \equiv \frac{f(x_1)}{f(x_3)} \tag{7}$$

(Press *et al.* 1992).

References

Brent, R. P. Ch. 3–4 in *Algorithms for Minimization Without Derivatives.* Englewood Cliffs, NJ: Prentice-Hall, 1973.

Forsythe, G. E.; Malcolm, M. A.; and Moler, C. B. §7.2 in *Computer Methods for Mathematical Computations.* Englewood Cliffs, NJ: Prentice-Hall, 1977.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Van Wijngaarden-Dekker-Brent Method." §9.3 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 352–355, 1992.

Brent-Salamin Formula

A formula which uses the ARITHMETIC-GEOMETRIC MEAN to compute PI. It has quadratic convergence and is also called the GAUSS-SALAMIN FORMULA and SALAMIN FORMULA. Let

$$a_{n+1} = \tfrac{1}{2}(a_n + b_n) \tag{1}$$

$$b_{n+1} = \sqrt{a_n b_n} \tag{2}$$

$$c_{n+1} = \tfrac{1}{2}(a_n - b_n) \tag{3}$$

$$d_n \equiv {a_n}^2 - {b_n}^2, \tag{4}$$

and define the initial conditions to be $a_0 = 1$, $b_0 = 1/\sqrt{2}$. Then iterating a_n and b_n gives the ARITHMETIC-GEOMETRIC MEAN, and π is given by

$$\pi = \frac{4[M(1, 2^{-1/2})]^2}{1 - \sum_{j=1}^{\infty} 2^{j+1} d_j} \tag{5}$$

$$= \frac{4[M(1, 2^{-1/2})]^2}{1 - \sum_{j=1}^{\infty} 2^{j+1} {c_j}^2}. \tag{6}$$

King (1924) showed that this formula and the LEGENDRE RELATION are equivalent and that either may be derived from the other.

see also ARITHMETIC-GEOMETRIC MEAN, PI

References

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, pp. 48–51, 1987.

Castellanos, D. "The Ubiquitous Pi. Part II." *Math. Mag.* **61**, 148–163, 1988.

King, L. V. *On the Direct Numerical Calculation of Elliptic Functions and Integrals.* Cambridge, England: Cambridge University Press, 1924.

Lord, N. J. "Recent Calculations of π: The Gauss-Salamin Algorithm." *Math. Gaz.* **76**, 231–242, 1992.

Salamin, E. "Computation of π Using Arithmetic-Geometric Mean." *Math. Comput.* **30**, 565-570, 1976.

Bretschneider's Formula

Given a general QUADRILATERAL with sides of lengths a, b, c, and d (Beyer 1987), the AREA is given by

$$A_{\text{quadrilateral}} = \tfrac{1}{4}\sqrt{4p^2q^2 - (b^2 + d^2 - a^2 - c^2)^2},$$

where p and q are the diagonal lengths.

see also BRAHMAGUPTA'S FORMULA, HERON'S FORMULA

References

Beyer, W. H. (Ed.). *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 123, 1987.

Brianchon Point

The point of CONCURRENCE of the joins of the VERTICES of a TRIANGLE and the points of contact of a CONIC SECTION INSCRIBED in the TRIANGLE. A CONIC INSCRIBED in a TRIANGLE has an equation of the form

$$\frac{f}{u}+\frac{g}{v}+\frac{h}{w}=0,$$

[redacted] $(1/f, 1/g, 1/h)$. For KIEPERT'S PARABOLA, the Branchion point has TRIANGLE CENTER FUNCTION

$$\alpha=\frac{1}{a(b^2-c^2)},$$

which is the STEINER POINT.

Brianchon's Theorem

The DUAL of PASCAL'S THEOREM. It states that, given a 6-sided POLYGON CIRCUMSCRIBED on a CONIC SECTION, the lines joining opposite VERTICES (DIAGONALS) meet in a single point.

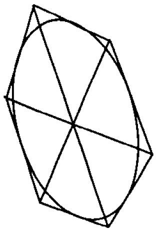

see also DUALITY PRINCIPLE, PASCAL'S THEOREM

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 77–79, 1967.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, p. 110, 1990.

Brick

see EULER BRICK, HARMONIC BRICK, RECTANGULAR PARALLELEPIPED

Bride's Chair

One name for the figure used by Euclid to prove the PYTHAGOREAN THEOREM.

see also PEACOCK'S TAIL, WINDMILL

Bridge Card Game

Bridge is a CARD game played with a normal deck of 52 cards. The number of possible distinct 13-card hands is

$$N=\binom{52}{13}=635{,}013{,}559{,}600.$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT. While the chances of being dealt a hand of 13 CARDS (out of 52) of the same suit are

$$\frac{4}{\binom{52}{13}}=\frac{1}{158{,}753{,}389{,}900},$$

the chance that one of four players will receive a hand of a single suit is

$$\frac{1}{39{,}688{,}347{,}497}.$$

There are special names for specific types of hands. A ten, jack, queen, king, or ace is called an "honor." Get- [redacted] suits and the ace, king, and queen, and jack of the remaining suit is called 13 top honors. Getting all cards of the same suit is called a 13-card suit. Getting 12 cards of same suit with ace high and the 13th card *not* an ace is called 2-card suit, ace high. Getting *no* honors is called a Yarborough.

The probabilities of being dealt 13-card bridge hands of a given type are given below. As usual, for a hand with probability P, the ODDS against being dealt it are $(1/P)-1:1$.

Hand	Exact	Probability
13 top honors	$\frac{4}{N}$	$\frac{1}{158{,}753{,}389{,}900}$
13-card suit	$\frac{4}{N}$	$\frac{1}{158{,}753{,}389{,}900}$
12-card suit, ace high	$\frac{4\cdot 12\cdot 36}{N}$	$\frac{4}{1{,}469{,}938{,}795}$
Yarborough	$\frac{\binom{32}{13}}{N}$	$\frac{5{,}394}{9{,}860{,}459}$
four aces	$\frac{\binom{48}{9}}{N}$	$\frac{11}{4{,}165}$
nine honors	$\frac{\binom{20}{9}\binom{32}{4}}{N}$	$\frac{888{,}212}{93{,}384{,}347}$

Hand	Probability	Odds
13 top honors	6.30×10^{-12}	158,753,389,899:1
13-card suit	6.30×10^{-12}	158,753,389,899:1
12-card suit, ace high	2.72×10^{-9}	367,484,697.8:1
Yarborough	5.47×10^{-4}	1,827.0:1
four aces	2.64×10^{-3}	377.6:1
nine honors	9.51×10^{-3}	104.1:1

see also CARDS, POKER

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 48–49, 1987.

Kraitchik, M. "Bridge Hands." §6.3 in *Mathematical Recreations.* New York: W. W. Norton, pp. 119–121, 1942.

Bridge (Graph)

The bridges of a GRAPH are the EDGES whose removal disconnects the GRAPH.

see also ARTICULATION VERTEX

References

Chartrand, G. "Cut-Vertices and Bridges." §2.4 in *Introductory Graph Theory.* New York: Dover, pp. 45–49, 1985.

Bridge Index

A numerical Knot invariant. For a Tame Knot K, the bridge index is the least Bridge Number of all planar representations of the Knot. The bridge index of the Unknot is defined as 1.

see also Bridge Number, Crookedness

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 114, 1976.

Schubert, H. "Über eine numerische Knotteninvariante." *Math. Z.* **61**, 245–288, 1954.

Bridge of Königsberg

see Königsberg Bridge Problem

Bridge Knot

An n-bridge knot is a knot with Bridge Number n. The set of 2-bridge knots is identical to the set of rational knots. If L is a 2-Bridge Knot, then the BLM/Ho Polynomial Q and Jones Polynomial V satisfy

$$Q_L(z) = 2z^{-1}V_L(t)V_L(t^{-1}+1-2z^{-1}),$$

where $z \equiv -t-t^{-1}$ (Kanenobu and Sumi 1993). Kanenobu and Sumi also give a table containing the number of distinct 2-bridge knots of n crossings for $n=10$ to 22, both not counting and counting Mirror Images as distinct.

n	K_n	$K_n+K_n^*$
3	0	0
4	0	0
5		
6		
7		
8		
9		
10	45	85
11	91	182
12	176	341
13	352	704
14	693	1365
15	1387	2774
16	2752	5461
17	5504	11008
18	10965	21845
19	21931	43862
20	43776	87381
21	87552	175104
22	174933	349525

References

Kanenobu, T. and Sumi, T. "Polynomial Invariants of 2-Bridge Knots through 22-Crossings." *Math. Comput.* **60**, 771–778 and S17–S28, 1993.

Schubert, H. "Knotten mit zwei Brücken." *Math. Z.* **65**, 133–170, 1956.

Bridge Number

The least number of unknotted arcs lying above the plane in any projection. The knot 05_{05} has bridge number 2. Such knots are called 2-Bridge Knots. There is a one-to-one correspondence between 2-Bridge Knots and rational knots. The knot 08_{010} is a 3-bridge knot. A knot with bridge number b is an n-Embeddable Knot where $n \leq b$.

see also Bridge Index

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 64–67, 1994.

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 115, 1976.

Brill-Noether Theorem

If the total group of the canonical series is divided into two parts, the difference between the number of points in each part and the double of the dimension of the complete series to which it belongs is the same.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 263, 1959.

Bring-Jerrard Quintic Form

A Tschirnhausen Transformation can be used to algebraically transform a general Quintic Equation to the form

$$z^5 + c_1 z + c_0 = 0. \tag{1}$$

In practice, the general quintic is first reduced to the Principal Quintic Form

$$y^5 + b_2 y^2 + b_1 y + b_0 = 0 \tag{2}$$

before the transformation is done. Then, we require that the sum of the third Powers of the Roots vanishes, so $s_3(y_j) = 0$. We assume that the Roots z_i of the Bring-Jerrard quintic are related to the Roots y_i of the Principal Quintic Form by

$$z_i = \alpha {y_i}^4 + \beta {y_i}^3 + \gamma {y_i}^2 + \delta y_i + \epsilon. \tag{3}$$

In a similar manner to the Principal Quintic Form transformation, we can express the Coefficients c_j in terms of the b_j.

see also Bring Quintic Form, Principal Quintic Form, Quintic Equation

Bring Quintic Form

A Tschirnhausen Transformation can be used to take a general Quintic Equation to the form

$$x^5 - x - a = 0,$$

where a may be Complex.

see also Bring-Jerrard Quintic Form, Quintic Equation

References

Ruppert, W. M. "On the Bring Normal Form of a Quintic in Characteristic 5." *Arch. Math.* **58**, 44–46, 1992.

Brioschi Formula

For a curve with METRIC

$$ds^2 = E\,du^2 + F\,du\,dv + G\,dv^2, \tag{1}$$

where E, F, and G is the first FUNDAMENTAL FORM, the GAUSSIAN CURVATURE is

$$K = \frac{M_1 + M_2}{(EG - F^2)^2}, \tag{2}$$

where

$$M_1 \equiv \begin{vmatrix} -\frac{1}{2}E_{vv} + F_{uv} - \frac{1}{2}G_{uu} & \frac{1}{2}E_u & F_u - \frac{1}{2}E_v \\ F_v - \frac{1}{2}G_u & E & F \\ \frac{1}{2}G_v & F & G \end{vmatrix} \tag{3}$$

$$M_2 \equiv \begin{vmatrix} 0 & \frac{1}{2}E_v & \frac{1}{2}G_u \\ \frac{1}{2}E_v & E & F \\ \frac{1}{2}G_u & F & G \end{vmatrix}, \tag{4}$$

which can also be written

$$K = -\frac{1}{\sqrt{EG}}\left[\frac{\partial}{\partial u}\left(\frac{1}{\sqrt{E}}\frac{\partial\sqrt{G}}{\partial u}\right) + \frac{\partial}{\partial v}\left(\frac{1}{\sqrt{G}}\frac{\partial\sqrt{E}}{\partial v}\right)\right] \tag{5}$$

$$= -\frac{1}{2\sqrt{EG}}\left[\frac{\partial}{\partial u}\left(\frac{G_u}{\sqrt{EG}}\right) + \frac{\partial}{\partial v}\left(\frac{E_v}{\sqrt{EG}}\right)\right]. \tag{6}$$

see also FUNDAMENTAL FORMS, GAUSSIAN CURVATURE

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 392–393, 1993.

Briot-Bouquet Equation

An ORDINARY DIFFERENTIAL EQUATION of the form

$$x^m y' = f(x, y),$$

where m is a POSITIVE INTEGER, f is ANALYTIC at $x = y = 0$, $f(0,0) = 0$, and $f'_y(0,0) \neq 0$.

References

Hazewinkel, M. (Managing Ed.). *Encyclopaedia of Mathematics: An Updated and Annotated Translation of the Soviet "Mathematical Encyclopaedia."* Dordrecht, Netherlands: Reidel, pp. 481–482, 1988.

Brocard Angle

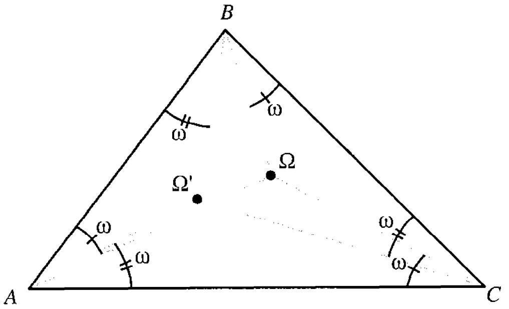

Define the first BROCARD POINT as the interior point Ω of a TRIANGLE for which the ANGLES $\angle\Omega AB$, $\angle\Omega BC$, and $\angle\Omega CA$ are equal. Similarly, define the second BROCARD POINT as the interior point Ω' for which the ANGLES $\angle\Omega' AC$, $\angle\Omega' CB$, and $\angle\Omega' BA$ are equal. Then the ANGLES in both cases are equal, and this angle is called the Brocard angle, denoted ω.

The Brocard angle ω of a TRIANGLE ΔABC is given by the formulas

$$\begin{aligned} \cot\omega &= \cot A + \cot B + \cot C && (1) \\ &= \left(\frac{a^2 + b^2 + c^2}{4\Delta}\right) && (2) \\ &= \frac{1 + \cos\alpha_1\cos\alpha_2\cos\alpha_3}{\sin\alpha_1\sin\alpha_2\sin\alpha_3} && (3) \\ &= \frac{\sin^2\alpha_1 + \sin^2\alpha_2 + \sin^2\alpha_3}{2\sin\alpha_1\sin\alpha_2\sin\alpha_3} && (4) \\ &= \frac{a_1\sin\alpha_1 + a_2\sin\alpha_2 + a_3\sin\alpha_3}{a_1\cos\alpha_1 + a_2\cos\alpha_2 + a_3\cos\alpha_3} && (5) \\ \csc^2\omega &= \csc^2\alpha_1 + \csc^2\alpha_2 + \csc^2\alpha_3 && (6) \\ \sin\omega &= \frac{2\Delta}{\sqrt{{a_1}^2{a_2}^2 + {a_2}^2{a_3}^2 + {a_3}^2{a_1}^2}}, && (7) \end{aligned}$$

where Δ is the TRIANGLE AREA, A, B, and C are ANGLES, and a, b, and c are side lengths.

If an ANGLE α of a TRIANGLE is given, the maximum possible Brocard angle is given by

$$\cot\omega = \tfrac{3}{2}\tan(\tfrac{1}{2}\alpha) + \tfrac{1}{2}\cos(\tfrac{1}{2}\alpha). \tag{8}$$

Let a TRIANGLE have ANGLES A, B, and C. Then

$$\sin A \sin B \sin C \leq kABC, \tag{9}$$

where

$$k = \left(\frac{3\sqrt{3}}{2\pi}\right)^3 \tag{10}$$

(Le Lionnais 1983). This can be used to prove that

$$8\omega^3 < ABC \tag{11}$$

(Abi-Khuzam 1974).

see also BROCARD CIRCLE, BROCARD LINE, EQUI-BROCARD CENTER, FERMAT POINT, ISOGONIC CENTERS

References

Abi-Khuzam, F. "Proof of Yff's Conjecture on the Brocard Angle of a Triangle." *Elem. Math.* **29**, 141–142, 1974.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 263–286 and 289–294, 1929.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 28, 1983.

Brocard Axis

The LINE KO passing through the LEMOINE POINT K and CIRCUMCENTER O of a TRIANGLE. The distance $\overline{OK}$ is called the BROCARD DIAMETER. The Brocard axis is PERPENDICULAR to the LEMOINE AXIS and is the ISOGONAL CONJUGATE of KIEPERT'S HYPERBOLA. It has equations

$$\sin(B-C)\alpha + \sin(C-A)\beta + \sin(A-B)\gamma = 0$$

$$bc(b^2-c^2)\alpha + ca(c^2-a^2)\beta + ab(a^2-b^2)\gamma = 0.$$

The LEMOINE POINT, CIRCUMCENTER, ISODYNAMIC POINTS, and BROCARD MIDPOINT all lie along the Brocard axis. Note that the Brocard axis is *not* equivalent to the BROCARD LINE.

see also BROCARD CIRCLE, BROCARD DIAMETER, BROCARD LINE

Brocard Circle

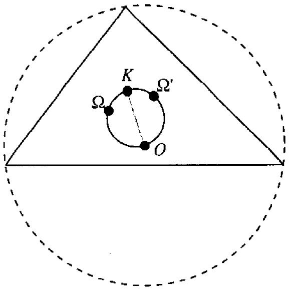

The CIRCLE passing through the first and second BROCARD POINTS Ω and Ω', the LEMOINE POINT K, and the CIRCUMCENTER O of a given TRIANGLE. The BROCARD POINTS Ω and Ω' are symmetrical about the LINE $\overleftrightarrow{KO}$, which is called the BROCARD LINE. The LINE SEGMENT $\overline{KO}$ is called the BROCARD DIAMETER, and it has length

$$\overline{OK} = \frac{\overline{O\Omega}}{\cos\omega} = \frac{R\sqrt{1-4\sin^2\omega}}{\cos\omega},$$

where R is the CIRCUMRADIUS and ω is the BROCARD ANGLE. The distance between either of the BROCARD POINTS and the LEMOINE POINT is

$$\overline{\Omega K} = \overline{\Omega' K} = \overline{\Omega O}\tan\omega.$$

see also BROCARD ANGLE, BROCARD DIAMETER, BROCARD POINTS

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, p. 272, 1929.

Brocard's Conjecture

$$\pi({p_{n+1}}^2) - \pi({p_n}^2) \geq 4$$

for $n \geq 2$ where π is the PRIME COUNTING FUNCTION.

see also ANDRICA'S CONJECTURE

Brocard Diameter

The LINE SEGMENT $\overline{KO}$ joining the LEMOINE POINT K and CIRCUMCENTER O of a given TRIANGLE. It is the DIAMETER of the TRIANGLE'S BROCARD CIRCLE, and lies along the BROCARD AXIS. The Brocard diameter has length

$$\overline{OK} = \frac{\overline{O\Omega}}{\cos\omega} = \frac{R\sqrt{1-4\sin^2\omega}}{\cos\omega},$$

where Ω is the first BROCARD POINT, R is the CIRCUMRADIUS, and ω is the BROCARD ANGLE.

see also BROCARD AXIS, BROCARD CIRCLE, BROCARD LINE, BROCARD POINTS

Brocard Line

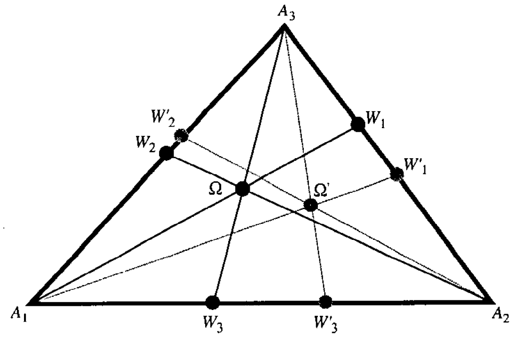

A LINE from any of the VERTICES A_i of a TRIANGLE to the first Ω or second Ω' BROCARD POINT. Let the ANGLE at a VERTEX A_i also be denoted A_i, and denote the intersections of $A_1\Omega$ and $A_1\Omega'$ with A_2A_3 as W_1 and W_2. Then the ANGLES involving these points are

$$\angle A_1\Omega W_3 = A_1 \tag{1}$$

$$\angle W_3\Omega A_2 = A_3 \tag{2}$$

$$\angle A_2\Omega W_1 = A_2. \tag{3}$$

Distances involving the points W_i and W_i' are given by

$$\overline{A_2\Omega} = \frac{a_3}{\sin A_2}\sin\omega \tag{4}$$

$$\frac{\overline{A_2\Omega}}{\overline{A_3\Omega}} = \frac{a_3{}^2}{a_1a_2} = \frac{\sin(A_3-\omega)}{\sin\omega} \tag{5}$$

$$\frac{\overline{W_3A_1}}{\overline{W_3A_2}} = \frac{a_2\sin\omega}{a_1\sin(A_3-\omega)} = \left(\frac{a_2}{a_3}\right)^2, \tag{6}$$

where ω is the BROCARD ANGLE (Johnson 1929, pp. 267–268).

The Brocard line, MEDIAN M, and LEMOINE POINT K are concurrent, with $A_1\Omega_1$, A_2K, and A_3M meeting at a point P. Similarly, $A_1\Omega'$, A_2M, and A_3K meet at a point which is the ISOGONAL CONJUGATE point of P (Johnson 1929, pp. 268–269).

see also BROCARD AXIS, BROCARD DIAMETER, BROCARD POINTS, ISOGONAL CONJUGATE, LEMOINE POINT, MEDIAN (TRIANGLE)

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 263–286, 1929.

Brocard Midpoint

The MIDPOINT of the BROCARD POINTS. It has TRIANGLE CENTER FUNCTION

$$\alpha = a(b^2+c^2) = \sin(A+\omega),$$

where ω is the BROCARD ANGLE. It lies on the BROCARD AXIS.

References

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Brocard Points

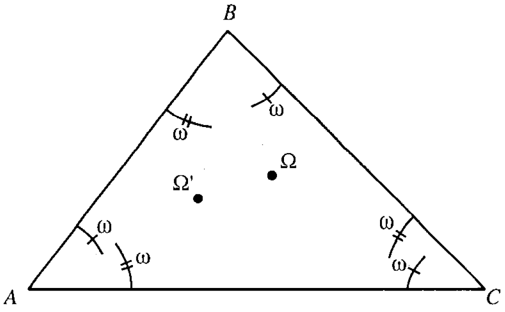

The first Brocard point is the interior point Ω (or τ_1 or Z_1) of a TRIANGLE for which the ANGLES $\angle\Omega AB$, $\angle\Omega BC$, and $\angle\Omega CA$ are equal. The second Brocard point is the interior point Ω' (or τ_2 or Z_2) for which the ANGLES $\angle\Omega' AC$, $\angle\Omega' CB$, and $\angle\Omega' BA$ are equal. The ANGLES in both cases are equal to the BROCARD ANGLE ω,

$$\begin{aligned}\omega &= \angle\Omega AB = \angle\Omega BC = \angle\Omega CA\\ &= \angle\Omega' AC = \angle\Omega' CB = \angle\Omega' BA.\end{aligned}$$

The first two Brocard points are ISOGONAL CONJUGATES (Johnson 1929, p. 266).

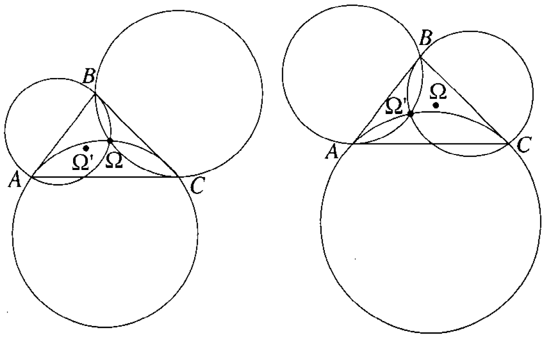

Let C_{BC} be the CIRCLE which passes through the vertices B and C and is TANGENT to the line AC at C, and similarly for C_{AB} and C_{BC}. Then the CIRCLES C_{AB}, C_{BC}, and C_{AC} intersect in the first Brocard point Ω. Similarly, let C'_{BC} be the CIRCLE which passes through the vertices B and C and is TANGENT to the line AB at B, and similarly for C'_{AB} and C'_{AC}. Then the CIRCLES C'_{AB}, C'_{BC}, and C'_{AC} intersect in the second Brocard points Ω' (Johnson 1929, pp. 264–265).

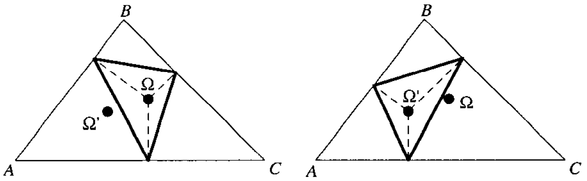

The PEDAL TRIANGLES of Ω and Ω' are congruent, and SIMILAR to the TRIANGLE ΔABC (Johnson 1929, p. 269). Lengths involving the Brocard points include

$$\overline{O\Omega} = \overline{O\Omega'} = R\sqrt{1-4\sin^2\omega} \tag{1}$$

$$\overline{\Omega\Omega'} = 2R\sin\omega\sqrt{1-4\sin^2\omega}. \tag{2}$$

Brocard's third point is related to a given TRIANGLE by the TRIANGLE CENTER FUNCTION

$$\alpha = a^{-3} \tag{3}$$

(Casey 1893, Kimberling 1994). The third Brocard point Ω'' (or τ_3 or Z_3) is COLLINEAR with the SPIEKER CENTER and the ISOTOMIC CONJUGATE POINT of its TRIANGLE'S INCENTER.

see also BROCARD ANGLE, BROCARD MIDPOINT, EQUI-BROCARD CENTER, YFF POINTS

References

Casey, J. *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions, with Numerous Examples, 2nd ed., rev. enl.* Dublin: Hodges, Figgis, & Co., p. 66, 1893.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 263–286, 1929.

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Stroeker, R. J. "Brocard Points, Circulant Matrices, and Descartes' Folium." *Math. Mag.* **61**, 172–187, 1988.

Brocard's Problem

Find the values of n for which $n!+1$ is a SQUARE NUMBER m^2, where $n!$ is the FACTORIAL (Brocard 1876, 1885). The only known solutions are $n = 4$, 5, and 7, and there are no other solutions < 1027. The pairs of numbers (m, n) are called BROWN NUMBERS.

see also BROWN NUMBERS, FACTORIAL, SQUARE NUMBER

References

Brocard, H. Question 166. *Nouv. Corres. Math.* **2**, 287, 1876.

Brocard, H. Question 1532. *Nouv. Ann. Math.* **4**, 391, 1885.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 193, 1994.

Brocard Triangles

Let the point of intersection of $A_2\Omega$ and $A_3\Omega'$ be B_1, where Ω and Ω' are the BROCARD POINTS, and similarly define B_2 and B_3. $B_1B_2B_3$ is the first Brocard triangle, and is inversely similar to $A_1A_2A_3$. It is inscribed in the BROCARD CIRCLE drawn with OK as the DIAMETER. The triangles $B_1A_2A_3$, $B_2A_3A_1$, and $B_3A_1A_2$ are ISOSCELES TRIANGLES with base angles ω, where ω is the BROCARD ANGLE. The sum of the areas of the ISOSCELES TRIANGLES is Δ, the AREA of TRIANGLE $A_1A_2A_3$. The first Brocard triangle is in perspective with the given TRIANGLE, with A_1B_1, A_2B_2, and A_3B_3 CONCURRENT. The MEDIAN POINT of the first Brocard triangle is the MEDIAN POINT M of the original triangle. The Brocard triangles are in perspective at M.

Let c_1, c_2, and c_3 and c_1', c_2', and c_3' be the CIRCLES intersecting in the BROCARD POINTS Ω and Ω', respectively. Let the two circles c_1 and c_1' tangent at A_1 to A_1A_2 and A_1A_3, and passing respectively through A_3 and A_2, meet again at C_1. The triangle $C_1C_2C_3$ is the second Brocard triangle. Each VERTEX of the second Brocard triangle lies on the second BROCARD CIRCLE.

The two Brocard triangles are in perspective at M.

see also STEINER POINTS, TARRY POINT

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 277–281, 1929.

Bromwich Integral

The inverse of the LAPLACE TRANSFORM, given by

$$F(t) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{st} f(s)\, ds,$$

where γ is a vertical CONTOUR in the COMPLEX PLANE chosen so that all singularities of $f(s)$ are to the left of it.

References

Arfken, G. "Inverse Laplace Transformation." §15.12 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 853–861, 1985.

Brothers

A PAIR of consecutive numbers.

see also PAIR, SMITH BROTHERS, TWINS

Brouwer Fixed Point Theorem

Any continuous FUNCTION $G : D^n \to D^n$ has a FIXED POINT, where

$$D^n = \{\mathbf{x} \in \mathbb{R}^n : {x_1}^2 + \ldots + {x_n}^2 \leq 1\}$$

is the unit n-BALL.

see also FIXED POINT THEOREM

References

Milnor, J. W. *Topology from the Differentiable Viewpoint.* Princeton, NJ: Princeton University Press, p. 14, 1965.

Browkin's Theorem

For every POSITIVE INTEGER n, there exists a SQUARE in the plane with exactly n LATTICE POINTS in its interior. This was extended by Schinzel and Kulikowski to *all* plane figures of a given shape. The generalization of the SQUARE in 2-D to the CUBE in 3-D was also proved by Browkin.

see also CUBE, SCHINZEL'S THEOREM, SQUARE

References

Honsberger, R. *Mathematical Gems I.* Washington, DC: Math. Assoc. Amer., pp. 121–125, 1973.

Brown's Criterion

A SEQUENCE $\{\nu_i\}$ of nondecreasing POSITIVE INTEGERS is COMPLETE IFF

1. $\nu_1 = 1$.
2. For all $k = 2, 3, \ldots,$

$$s_{k-1} = \nu_1 + \nu_2 + \ldots + \nu_{k-1} \geq \nu_k - 1.$$

A corollary states that a SEQUENCE for which $\nu_1 = 1$ and $\nu_{k+1} \leq 2\nu_k$ is COMPLETE (Honsberger 1985).

see also COMPLETE SEQUENCE

References

Brown, J. L. Jr. "Notes on Complete Sequences of Integers." *Amer. Math. Monthly*, 557–560, 1961.

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 123–130, 1985.

Brown Function

For a FRACTAL PROCESS with values $y(t-\Delta t)$ and $y(t+\Delta t)$, the correlation between these two values is given by the Brown function

$$r = 2^{2H-1} - 1,$$

also known as the BACHELIER FUNCTION, LÉVY FUNCTION, or WIENER FUNCTION.

Brown Numbers

Brown numbers are PAIRS (m, n) of INTEGERS satisfying the condition of BROCARD'S PROBLEM, i.e., such that

$$n! + 1 = m^2$$

where $n!$ is the FACTORIAL and m^2 is a SQUARE NUMBER. Only three such PAIRS of numbers are known: (5,4), (11,5), (71,7), and Erdős conjectured that these are the only three such PAIRS. Le Lionnais (1983) points out that there are 3 numbers less than 200,000 for which

$$(n-1)! + 1 \equiv 0 \pmod{n^2},$$

namely 5, 13, and 563.

see also BROCARD'S PROBLEM, FACTORIAL, SQUARE NUMBER

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 193, 1994.
Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 56, 1983.
Pickover, C. A. *Keys to Infinity.* New York: W. H. Freeman, p. 170, 1995.

Broyden's Method

An extension of the secant method of root finding to higher dimensions.

References

Broyden, C. G. "A Class of Methods for Solving Nonlinear Simultaneous Equations." *Math. Comput.* **19**, 577–593, 1965.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 382–385, 1992.

Bruck-Ryser-Chowla Theorem

If $n \equiv 1, 2 \pmod 4$, and the SQUAREFREE part of n is divisible by a PRIME $p \equiv 3 \pmod 4$, then no DIFFERENCE SET of ORDER n exists. Equivalently, if a PROJECTIVE PLANE of order n exists, and $n = 1$ or $2 \pmod 4$, then n is the sum of two SQUARES.

Dinitz and Stinson (1992) give the theorem in the following form. If a symmetric (v, k, λ)-BLOCK DESIGN exists, then

1. If v is EVEN, then $k - \lambda$ is a SQUARE NUMBER,
2. If v is ODD, the the DIOPHANTINE EQUATION

$$x^2 = (k-\lambda)y^2 + (-1)^{(v-1)/2}\lambda z^2$$

has a solution in integers, not all of which are 0.

see also BLOCK DESIGN, FISHER'S BLOCK DESIGN INEQUALITY

References

Dinitz, J. H. and Stinson, D. R. "A Brief Introduction to Design Theory." Ch. 1 in *Contemporary Design Theory: A Collection of Surveys* (Ed. J. H. Dinitz and D. R. Stinson). New York: Wiley, pp. 1–12, 1992.
Gordon, D. M. "The Prime Power Conjecture is True for $n < 2{,}000{,}000$." *Electronic J. Combinatorics* **1**, R6, 1–7, 1994. `http://www.combinatorics.org/Volume_1/volume1.html#R6`.
Ryser, H. J. *Combinatorial Mathematics.* Buffalo, NY: Math. Assoc. Amer., 1963.

Bruck-Ryser Theorem

see BRUCK-RYSER-CHOWLA THEOREM

Brun's Constant

The number obtained by adding the reciprocals of the TWIN PRIMES,

$$B \equiv \left(\tfrac{1}{3} + \tfrac{1}{5}\right) + \left(\tfrac{1}{5} + \tfrac{1}{7}\right) + \left(\tfrac{1}{11} + \tfrac{1}{13}\right) + \left(\tfrac{1}{17} + \tfrac{1}{19}\right) + \cdots, \quad (1)$$

By BRUN'S THEOREM, the constant converges to a definite number as $p \to \infty$. Any finite sum underestimates B. Shanks and Wrench (1974) used all the TWIN PRIMES among the first 2 million numbers. Brent (1976) calculated all TWIN PRIMES up to 100 billion and obtained (Ribenboim 1989, p. 146)

$$B \approx 1.90216054, \quad (2)$$

assuming the truth of the first HARDY-LITTLEWOOD CONJECTURE. Using TWIN PRIMES up to 10^{14}, Nicely (1996) obtained

$$B \approx 1.9021605778 \pm 2.1 \times 10^{-9} \quad (3)$$

(Cipra 1995, 1996), in the process discovering a bug in Intel's® Pentium™ microprocessor. The value given by Le Lionnais (1983) is incorrect.

see also TWIN PRIMES, TWIN PRIME CONJECTURE, TWIN PRIMES CONSTANT

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 64, 1987.
Brent, R. P. "Tables Concerning Irregularities in the Distribution of Primes and Twin Primes Up to 10^{11}." *Math. Comput.* **30**, 379, 1976.
Cipra, B. "How Number Theory Got the Best of the Pentium Chip." *Science* **267**, 175, 1995.
Cipra, B. "Divide and Conquer." *What's Happening in the Mathematical Sciences, 1995–1996, Vol. 3.* Providence, RI: Amer. Math. Soc., pp. 38–47, 1996.
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/brun/brun.html`.
Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 41, 1983.
Nicely, T. "Enumeration to 10^{14} of the Twin Primes and Brun's Constant." *Virginia J. Sci.* **46**, 195–204, 1996.
Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, 1989.
Shanks, D. and Wrench, J. W. "Brun's Constant." *Math. Comput.* **28**, 293–299, 1974.
Wolf, M. "Generalized Brun's Constants." `http://www.ift.uni.wroc.pl/~mwolf/`.

Brunn-Minkowski Inequality

The nth root of the CONTENT of the set sum of two sets in Euclidean n-space is greater than or equal to the sum of the nth roots of the CONTENTS of the individual sets.

see also TOMOGRAPHY

References

Cover, T. M. "The Entropy Power Inequality and the Brunn-Minkowski Inequality" §5.10 in In *Open Problems in Communications and Computation.* (Ed. T. M. Cover and B. Gopinath). New York: Springer-Verlag, p. 172, 1987.

Schneider, R. *Convex Bodies: The Brunn-Minkowski Theory.* Cambridge, England: Cambridge University Press, 1993.

Brun's Sum

see BRUN'S CONSTANT

Brun's Theorem

The series producing BRUN'S CONSTANT CONVERGES even if there are an infinite number of TWIN PRIMES. Proved in 1919 by V. Brun.

Brunnian Link

A Brunnian link is a set of n linked loops such that each proper sublink is trivial, so that the removal of any component leaves a set of trivial unlinked UNKNOTS. The BORROMEAN RINGS are the simplest example and have $n = 3$.

see also BORROMEAN RINGS

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, 1976.

Brute Force Factorization

see DIRECT SEARCH FACTORIZATION

Bubble

A bubble is a MINIMAL SURFACE of the type that is formed by soap film. The simplest bubble is a single SPHERE. More complicated forms occur when multiple bubbles are joined together. Two outstanding problems involving bubbles are to find the arrangements with the smallest PERIMETER (planar problem) or SURFACE AREA (AREA problem) which enclose and separate n given unit areas or volumes in the plane or in space. For $n = 2$, the problems are called the DOUBLE BUBBLE CONJECTURE and the solution to both problems is known to be the DOUBLE BUBBLE.

see also DOUBLE BUBBLE, MINIMAL SURFACE, PLATEAU'S LAWS, PLATEAU'S PROBLEM

References

Morgan, F. "Mathematicians, Including Undergraduates, Look at Soap Bubbles." *Amer. Math. Monthly* **101**, 343–351, 1994.

Pappas, T. "Mathematics & Soap Bubbles." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 219, 1989.

Buchberger's Algorithm

The algorithm for the construction of a GRÖBNER BASIS from an arbitrary ideal basis.

see also GRÖBNER BASIS

References

Becker, T. and Weispfenning, V. *Gröbner Bases: A Computational Approach to Commutative Algebra.* New York: Springer-Verlag, pp. 213–214, 1993.

Buchberger, B. "Theoretical Basis for the Reduction of Polynomials to Canonical Forms." *SIGSAM Bull.* **39**, 19–24, Aug. 1976.

Cox, D.; Little, J.; and O'Shea, D. *Ideals, Varieties, and Algorithms: An Introduction to Algebraic Geometry and Commutative Algebra, 2nd ed.* New York: Springer-Verlag, 1996.

Buckminster Fuller Dome

see GEODESIC DOME

Buffon-Laplace Needle Problem

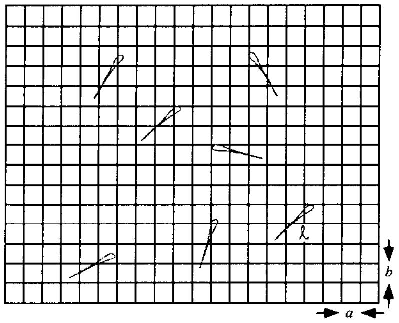

Find the probability $P(\ell, a, b)$ that a needle of length ℓ will land on a line, given a floor with a grid of equally spaced PARALLEL LINES distances a and b apart, with $\ell > a, b$.

$$P(\ell, a, b) = \frac{2\ell(a+b) - \ell^2}{\pi ab}.$$

see also BUFFON'S NEEDLE PROBLEM

Buffon's Needle Problem

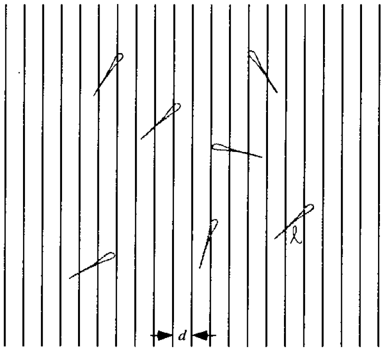

Find the probability $P(\ell, d)$ that a needle of length ℓ will land on a line, given a floor with equally spaced PARALLEL LINES a distance d apart.

$$\begin{aligned} P(\ell, d) &= \int_0^{2\pi} \frac{\ell |\cos\theta|}{d} \frac{d\theta}{2\pi} = \frac{\ell}{2\pi d} 4 \int_0^{\pi/2} \cos\theta\, d\theta \\ &= \frac{2\ell}{\pi d} [\sin\theta]_0^{\pi/2} = \frac{2\ell}{\pi d}. \end{aligned}$$

Several attempts have been made to experimentally determine π by needle-tossing. For a discussion of the relevant statistics and a critical analysis of one of the more accurate (and least believable) needle-tossings, see Badger (1994).

see also BUFFON-LAPLACE NEEDLE PROBLEM

References

Badger, L. "Lazzarini's Lucky Approximation of π." *Math. Mag.* **67**, 83–91, 1994.

Dörrie, H. "Buffon's Needle Problem." §18 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 73–77, 1965.

Kraitchik, M. "The Needle Problem." §6.14 in *Mathematical Recreations.* New York: W. W. Norton, p. 132, 1942.

Wegert, E. and Trefethen, L. N. "From the Buffon Needle Problem to the Kreiss Matrix Theorem." *Amer. Math. Monthly* **101**, 132–139, 1994.

Bulirsch-Stoer Algorithm

An algorithm which finds RATIONAL FUNCTION extrapolations of the form

$$R_{i(i+1)\cdots(i+m)} = \frac{P_\mu(x)}{P_\nu(x)} = \frac{p_0 + p_1 x + \ldots + p_\mu x^\mu}{q_0 + q_1 x + \ldots + q_\nu x^\nu}$$

and can be used in the solution of ORDINARY DIFFERENTIAL EQUATIONS.

References

Bulirsch, R. and Stoer, J. §2.2 in *Introduction to Numerical Analysis.* New York: Springer-Verlag, 1991.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Richardson Extrapolation and the Bulirsch-Stoer Method." §16.4 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 718–725, 1992.

Bullet Nose

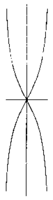

A plane curve with implicit equation

$$\frac{a^2}{x^2} - \frac{b^2}{y^2} = 1. \tag{1}$$

In parametric form,

$$x = a\cos t \tag{2}$$

$$y = b\cot t. \tag{3}$$

The CURVATURE is

$$\kappa = \frac{3ab\cot t\csc t}{(b^2\csc^4 t + a^2\sin^2 t)^{3/2}} \tag{4}$$

and the TANGENTIAL ANGLE is

$$\phi = \tan^{-1}\left(\frac{b\csc^3 t}{a}\right). \tag{5}$$

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 127–129, 1972.

Bumping Algorithm

Given a PERMUTATION $\{p_1, p_2, \ldots, p_n\}$ of $\{1, \ldots, n\}$, the bumping algorithm constructs a standard YOUNG TABLEAU by inserting the p_i one by one into an already constructed YOUNG TABLEAU. To apply the bumping algorithm, start with $\{\{p_1\}\}$, which is a YOUNG TABLEAU. If p_1 through p_k have already been inserted, then in order to insert p_{k+1}, start with the first line of the already constructed YOUNG TABLEAU and search for the first element of this line which is greater than p_{k+1}. If there is no such element, append p_{k+1} to the first line and stop. If there is such an element (say, p_p), exchange p_p for p_{k+1}, search the second line using p_p, and so on.

see also YOUNG TABLEAU

References

Skiena, S. *Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica.* Reading, MA: Addison-Wesley, 1990.

Bundle

see FIBER BUNDLE

Burau Representation

Gives a MATRIX representation b_i of a BRAID GROUP in terms of $(n-1)\times(n-1)$ MATRICES. A $-t$ always appears in the (i, i) position.

$$\mathsf{b}_1 = \begin{bmatrix} -t & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & \cdots & 1 \end{bmatrix} \tag{1}$$

$$\mathsf{b}_i = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -t & 0 & \cdots & 0 \\ 0 & \cdots & -t & 0 & \cdots & 0 \\ 0 & \cdots & -1 & 1 & \cdots & 0 \\ 0 & \ddots & 0 & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{2}$$

$$\mathsf{b}_{n-1} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & -t \\ 0 & 0 & \cdots & 0 & -t \end{bmatrix}. \tag{3}$$

Let Ψ be the MATRIX PRODUCT of BRAID WORDS, then

$$\frac{\det(\mathsf{I} - \Psi)}{1 + t + \cdots + t^{n-1}} = \Delta_L, \tag{4}$$

where Δ_L is the ALEXANDER POLYNOMIAL and det is the DETERMINANT.

References

Burau, W. "Über Zopfgruppen und gleichsinnig verdrilte Verkettungen." *Abh. Math. Sem. Hanischen Univ.* **11**, 171–178, 1936.

Jones, V. "Hecke Algebra Representation of Braid Groups and Link Polynomials." *Ann. Math.* **126**, 335–388, 1987.

Burkhardt Quartic

The VARIETY which is an invariant of degree four and is given by the equation

$$y_0^4 - y_0(y_1^3 + y_2^3 + y_3^3 + y_4^3) + 3y_1y_2y_3y_4 = 0.$$

References

Burkhardt, H. "Untersuchungen aus dem Gebiet der hyperelliptischen Modulfunctionen. II." *Math. Ann.* **38**, 161–224, 1890.

Burkhardt, H. "Untersuchungen aus dem Gebiet der hyperelliptischen Modulfunctionen. III." *Math. Ann.* **40**, 313–343, 1892.

Hunt, B. "The Burkhardt Quartic." Ch. 5 in *The Geometry of Some Special Arithmetic Quotients.* New York: Springer-Verlag, pp. 168–221, 1996.

Burnside's Conjecture

Every non-ABELIAN SIMPLE GROUP has EVEN ORDER.

see also ABELIAN GROUP, SIMPLE GROUP

Burnside's Lemma

Let J be a FINITE GROUP and the image $R(J)$ be a representation which is a HOMEOMORPHISM of J into a PERMUTATION GROUP $S(X)$, where $S(X)$ is the GROUP of all permutations of a SET X. Define the orbits of $R(J)$ as the equivalence classes under $x \sim y$, which is true if there is some permutation p in $R(J)$ such that $p(x) = y$. Define the fixed points of p as the elements x of X for which $p(x) = x$. Then the AVERAGE number of FIXED POINTS of permutations in $R(J)$ is equal to the number of orbits of $R(J)$.

The LEMMA was apparently known by Cauchy (1845) in obscure form and Frobenius (1887) prior to Burnside's (1900) rediscovery. It was subsequently extended and refined by Pólya (1937) for applications in COMBINATORIAL counting problems. In this form, it is known as PÓLYA ENUMERATION THEOREM.

References

Pólya, G. "Kombinatorische Anzahlbestimmungen für Gruppen, Graphen, und chemische Verbindungen." *Acta Math.* **68**, 145–254, 1937.

Burnside Problem

A problem originating with W. Burnside (1902), who wrote, "A still undecided point in the theory of discontinuous groups is whether the ORDER of a GROUP may be not finite, while the order of every operation it contains is finite." This question would now be phrased as "Can a finitely generated group be infinite while every element in the group has finite order?" (Vaughan-Lee 1990). This question was answered by Golod (1964) when he constructed finitely generated infinite p-GROUPS. These GROUPS, however, do not have a finite exponent.

Let F_r be the FREE GROUP of RANK r and let N be the SUBGROUP generated by the set of nth POWERS $\{g^n | g \in F_r\}$. Then N is a normal subgroup of F_r. We define $B(r,n) = F_r/N$ to be the QUOTIENT GROUP. We call $B(r,n)$ the r-generator Burnside group of exponent n. It is the largest r-generator group of exponent n, in the sense that every other such group is a HOMEOMORPHIC image of $B(r,n)$. The Burnside problem is usually stated as: "For which values of r and n is $B(r,n)$ a FINITE GROUP?"

An answer is known for the following values. For $r = 1$, $B(1,n)$ is a CYCLIC GROUP of ORDER n. For $n = 2$, $B(r,2)$ is an elementary ABELIAN 2-group of ORDER 2^n. For $n = 3$, $B(r,3)$ was proved to be finite by Burnside. The ORDER of the $B(r,3)$ groups was established by Levi and van der Waerden (1933), namely 3^a where

$$a \equiv r + \binom{r}{2} + \binom{r}{3}, \tag{1}$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT. For $n = 4$, $B(r,4)$ was proved to be finite by Sanov (1940). Groups of exponent four turn out to be the most complicated for which a POSITIVE solution is known. The precise nilpotency class and derived length are known, as are bounds for the ORDER. For example,

$$|B(2,4)| = 2^{12} \tag{2}$$

$$|B(3,4)| = 2^{69} \tag{3}$$

$$|B(4,4)| = 2^{422} \tag{4}$$

$$|B(5,4)| = 2^{2728}, \tag{5}$$

while for larger values of r the exact value is not yet known. For $n = 6$, $B(r,6)$ was proved to be finite by Hall (1958) with ORDER 2^a3^b, where

$$a \equiv 1 + (r-1)3^c \tag{6}$$

$$b \equiv 1 + (r-1)2^r \tag{7}$$

$$c \equiv r + \binom{r}{2} + \binom{r}{3}. \tag{8}$$

No other Burnside groups are known to be finite. On the other hand, for $r > 2$ and $n \geq 665$, with n ODD,

$B(r,n)$ is infinite (Novikov and Adjan 1968). There is a similar fact for $r > 2$ and n a large POWER of 2.

E. Zelmanov was awarded a FIELDS MEDAL in 1994 for his solution of the "restricted" Burnside problem.

see also FREE GROUP

References

Burnside, W. "On an Unsettled Question in the Theory of Discontinuous Groups." *Quart. J. Pure Appl. Math.* **33**, 230–238, 1902.
Golod, E. S. "On Nil-Algebras and Residually Finite p-Groups." *Isv. Akad. Nauk SSSR Ser. Mat.* **28**, 273–276, 1964.
Hall, M. "Solution of the Burnside Problem for Exponent Six." *Ill. J. Math.* **2**, 764–786, 1958.
Levi, F. and van der Waerden, B. L. "Über eine besondere Klasse von Gruppen." *Abh. Math. Sem. Univ. Hamburg* **9**, 154–158, 1933.
Novikov, P. S. and Adjan, S. I. "Infinite Periodic Groups I, II, III." *Izv. Akad. Nauk SSSR Ser. Mat.* **32**, 212–244, 251–524, and 709–731, 1968.
Sanov, I. N. "Solution of Burnside's problem for exponent four." *Leningrad State Univ. Ann. Math. Ser.* **10**, 166–170, 1940.
Vaughan-Lee, M. *The Restricted Burnside Problem, 2nd ed.* New York: Clarendon Press, 1993.

Busemann-Petty Problem

If the section function of a centered convex body in Euclidean n-space ($n \geq 3$) is smaller than that of another such body, is its volume also smaller?

References

Gardner, R. J. "Geometric Tomography." *Not. Amer. Math. Soc.* **42**, 422–429, 1995.

Busy Beaver

A busy beaver is an n-state, 2-symbol, 5-tuple TURING MACHINE which writes the maximum possible number $BB(n)$ of 1s on an initially blank tape before halting. For $n = 0, 1, 2, \ldots$, $BB(n)$ is given by 0, 1, 4, 6, 13, $\geq$ 4098, $\geq$ 136612, ... The busy beaver sequence is also known as RADO'S SIGMA FUNCTION.

see also HALTING PROBLEM, TURING MACHINE

References

Chaitin, G. J. "Computing the Busy Beaver Function." §4.4 in *Open Problems in Communication and Computation* (Ed. T. M. Cover and B. Gopinath). New York: Springer-Verlag, pp. 108–112, 1987.
Dewdney, A. K. "A Computer Trap for the Busy Beaver, the Hardest-Working Turing Machine." *Sci. Amer.* **251**, 19–23, Aug. 1984.
Marxen, H. and Buntrock, J. "Attacking the Busy Beaver 5." *Bull. EATCS* **40**, 247–251, Feb. 1990.
Sloane, N. J. A. Sequence A028444 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Butterfly Catastrophe

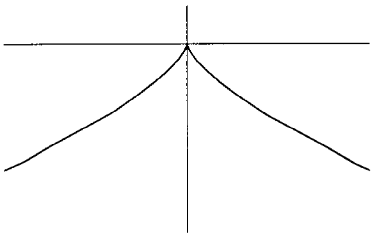

A CATASTROPHE which can occur for four control factors and one behavior axis. The equations

$$x = c(8at^3 + 24t^5)$$
$$y = c(-6at^2 - 15t^4)$$

display such a catastrophe (von Seggern 1993).

References

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 94, 1993.

Butterfly Curve

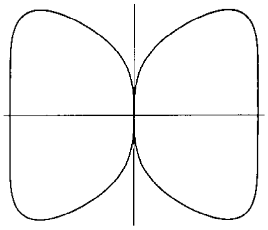

A PLANE CURVE given by the implicit equation

$$y^6 = (x^2 - x^6).$$

see also DUMBBELL CURVE, EIGHT CURVE, PIRIFORM

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 72, 1989.

Butterfly Effect

Due to nonlinearities in weather processes, a butterfly flapping its wings in Tahiti can, in theory, produce a tornado in Kansas. This strong dependence of outcomes on very slightly differing initial conditions is a hallmark of the mathematical behavior known as CHAOS.

see also CHAOS, LORENZ SYSTEM

Butterfly Fractal

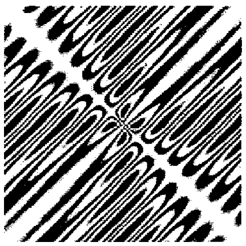

The FRACTAL-like curve generated by the 2-D function

$$f(x,y) = \frac{(x^2 - y^2)\sin\left(\frac{x+y}{a}\right)}{x^2 + y^2}.$$

Butterfly Polyiamond

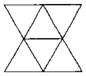

A 6-POLYIAMOND.

References

Golomb, S. W. *Polyominoes: Puzzles, Patterns, Problems, and Packings, 2nd ed.* Princeton, NJ: Princeton University Press, p. 92, 1994.

Butterfly Theorem

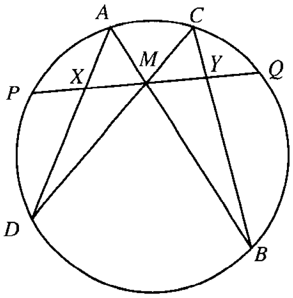

Given a CHORD PQ of a CIRCLE, draw any other two CHORDS AB and CD passing through its MIDPOINT. Call the points where AD and BC meet PQ X and Y. Then M is the MIDPOINT of XY.

see also CHORD, CIRCLE, MIDPOINT

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 45–46, 1967.

C

$\mathbb{C}$

The FIELD of COMPLEX NUMBERS, denoted $\mathbb{C}$.

see also $\mathbb{C}^*$, COMPLEX NUMBER, $\mathbb{I}$, $\mathbb{N}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{Z}$

$\mathbb{C}^*$

The RIEMANN SPHERE $\mathbb{C} \cup \{\infty\}$.

see also $\mathbb{C}$, COMPLEX NUMBER, $\mathbb{Q}$, $\mathbb{R}$, RIEMANN SPHERE, $\mathbb{Z}$

C^*-Algebra

A special type of B^*-ALGEBRA in which the INVOLUTION is the ADJOINT OPERATOR in a HILBERT SPACE.

see also B^*-ALGEBRA, k-THEORY

References

Davidson, K. R. *C^*-Algebras by Example.* Providence, RI: Amer. Math. Soc., 1996.

C-Curve

see LÉVY FRACTAL

C-Determinant

A DETERMINANT appearing in PADÉ APPROXIMANT identities:

$$C_{r/s} = \begin{vmatrix} a_{r-s+1} & a_{r-s+2} & \cdots & a_r \\ \vdots & \vdots & \ddots & \vdots \\ a_r & a_{r+1} & \cdots & a_{r+s-1} \end{vmatrix}.$$

see also PADÉ APPROXIMANT

C-Matrix

Any SYMMETRIC MATRIX ($\mathsf{A}^{\mathrm{T}} = \mathsf{A}$) or SKEW SYMMETRIC MATRIX ($\mathsf{A}^{\mathrm{T}} = -\mathsf{A}$) C_n with diagonal elements 0 and others ± 1 satisfying

$$\mathsf{C}\mathsf{C}^{\mathrm{T}} = (n-1)\mathsf{I},$$

where I is the IDENTITY MATRIX, is known as a C-matrix (Ball and Coxeter 1987). Examples include

$$\mathsf{C}_4 = \begin{bmatrix} 0 & + & + & + \\ - & 0 & - & + \\ - & + & 0 & - \\ - & - & + & 0 \end{bmatrix}$$

$$\mathsf{C}_6 = \begin{bmatrix} 0 & + & + & + & + & + \\ + & 0 & + & - & - & + \\ + & + & 0 & + & + & - \\ + & - & + & 0 & + & - \\ + & - & - & + & 0 & + \\ + & + & - & - & + & 0 \end{bmatrix}.$$

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 308–309, 1987.

C-Table

see C-DETERMINANT

Cable Knot

Let K_1 be a TORUS KNOT. Then the SATELLITE KNOT with COMPANION KNOT K_2 is a cable knot on K_2.

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, p. 118, 1994.

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 112 and 283, 1976.

Cactus Fractal

A MANDELBROT SET-like FRACTAL obtained by iterating the map

$$z_{n+1} = {z_n}^3 + (z_0 - 1)z_n - z_0.$$

see also FRACTAL, JULIA SET, MANDELBROT SET

Cake Cutting

It is always possible to "fairly" divide a cake among n people using only vertical cuts. Furthermore, it is possible to cut and divide a cake such that each person believes that *everyone* has received $1/n$ of the cake according to his own measure. Finally, if there is some piece on which two people disagree, then there is a way of partitioning and dividing a cake such that each participant believes that he has obtained more than $1/n$ of the cake according to his own measure.

Ignoring the height of the cake, the cake-cutting problem is really a question of fairly dividing a CIRCLE into n equal AREA pieces using cuts in its plane. One method of proving fair cake cutting to always be possible relies on the FROBENIUS-KÖNIG THEOREM.

see also CIRCLE CUTTING, CYLINDER CUTTING, ENVYFREE, FROBENIUS-KÖNIG THEOREM, HAM SANDWICH THEOREM, PANCAKE THEOREM, PIZZA THEOREM, SQUARE CUTTING, TORUS CUTTING

References

Brams, S. J. and Taylor, A. D. "An Envy-Free Cake Division Protocol." *Amer. Math. Monthly* **102**, 9–19, 1995.

Brams, S. J. and Taylor, A. D. *Fair Division: From Cake-Cutting to Dispute Resolution.* New York: Cambridge University Press, 1996.

Dubbins, L. and Spanier, E. "How to Cut a Cake Fairly." *Amer. Math. Monthly* **68**, 1–17, 1961.

Gale, D. "Dividing a Cake." *Math. Intel.* **15**, 50, 1993.

Jones, M. L. "A Note on a Cake Cutting Algorithm of Banach and Knaster." *Amer. Math. Monthly* **104**, 353–355, 1997.

Rebman, K. "How to Get (At Least) a Fair Share of the Cake." In *Mathematical Plums* (Ed. R. Honsberger). Washington, DC: Math. Assoc. Amer., pp. 22–37, 1979.

Steinhaus, H. "Sur la division progmatique." *Ekonometrika (Supp.)* **17**, 315–319, 1949.
Stromquist, W. "How to Cut a Cake Fairly." *Amer. Math. Monthly* **87**, 640–644, 1980.

Cal

see WALSH FUNCTION

Calabi's Triangle

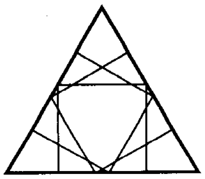

Equilateral Triangle

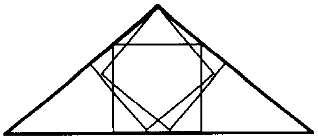

Calabi's Triangle

The one TRIANGLE in addition to the EQUILATERAL TRIANGLE for which the largest inscribed SQUARE can be inscribed in three different ways. The ratio of the sides to that of the base is given by $x = 1.55138752455\ldots$ (Sloane's A046095), where

$$x = \frac{1}{3} + \frac{(-23 + 3i\sqrt{237})^{1/3}}{3 \cdot 2^{2/3}} + \frac{11}{3[2(-23 + 3i\sqrt{237})]^{1/3}}$$

is the largest POSITIVE ROOT of

$$2x^3 - 2x^2 - 3x + 2 = 0,$$

which has CONTINUED FRACTION [1, 1, 1, 4, 2, 1, 2, 1, 5, 2, 1, 3, 1, 1, 390, ...] (Sloane's A046096).

see also GRAHAM'S BIGGEST LITTLE HEXAGON

References
Conway, J. H. and Guy, R. K. "Calabi's Triangle." In *The Book of Numbers.* New York: Springer-Verlag, p. 206, 1996.
Sloane, N. J. A. Sequences A046095 and A046096 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Calabi-Yau Space

A structure into which the 6 extra DIMENSIONS of 10-D string theory curl up.

Calculus

In general, "a" calculus is an abstract theory developed in a purely formal way.

"The" calculus, more properly called ANALYSIS (or REAL ANALYSIS or, in older literature, INFINITESIMAL ANALYSIS) is the branch of mathematics studying the rate of change of quantities (which can be interpreted as SLOPES of curves) and the length, AREA, and VOLUME of objects. The CALCULUS is sometimes divided into DIFFERENTIAL and INTEGRAL CALCULUS, concerned with DERIVATIVES

$$\frac{d}{dx} f(x)$$

and INTEGRALS

$$\int f(x)\, dx,$$

respectively.

While ideas related to calculus had been known for some time (Archimedes' EXHAUSTION METHOD was a form of calculus), it was not until the independent work of Newton and Leibniz that the modern elegant tools and ideas of calculus were developed. Even so, many years elapsed until the subject was put on a mathematically rigorous footing by mathematicians such as Weierstraß.

see also ARC LENGTH, AREA, CALCULUS OF VARIATIONS, CHANGE OF VARIABLES THEOREM, DERIVATIVE, DIFFERENTIAL CALCULUS, ELLIPSOIDAL CALCULUS, EXTENSIONS CALCULUS, FLUENT, FLUXION, FRACTIONAL CALCULUS, FUNCTIONAL CALCULUS, FUNDAMENTAL THEOREMS OF CALCULUS, HEAVISIDE CALCULUS, INTEGRAL, INTEGRAL CALCULUS, JACOBIAN, LAMBDA CALCULUS, KIRBY CALCULUS, MALLIAVIN CALCULUS, PREDICATE CALCULUS, PROPOSITIONAL CALCULUS, SLOPE, TENSOR CALCULUS, UMBRAL CALCULUS, VOLUME

References
Anton, H. *Calculus with Analytic Geometry, 5th ed.* New York: Wiley, 1995.
Apostol, T. M. *Calculus, 2nd ed., Vol. 1: One-Variable Calculus, with an Introduction to Linear Algebra.* Waltham, MA: Blaisdell, 1967.
Apostol, T. M. *Calculus, 2nd ed., Vol. 2: Multi-Variable Calculus and Linear Algebra, with Applications to Differential Equations and Probability.* Waltham, MA: Blaisdell, 1969.
Apostol, T. M. *A Century of Calculus, 2 vols. Pt. 1: 1894–1968. Pt. 2: 1969–1991.* Washington, DC: Math. Assoc. Amer., 1992.
Ayres, F. Jr. and Mendelson, E. *Schaum's Outline of Theory and Problems of Differential and Integral Calculus, 3rd ed.* New York: McGraw-Hill, 1990.
Borden, R. S. *A Course in Advanced Calculus.* New York: Dover, 1998.
Boyer, C. B. *A History of the Calculus and Its Conceptual Development.* New York: Dover, 1989.
Brown, K. S. "Calculus and Differential Equations." `http://www.seanet.com/~ksbrown/icalculu.htm`.
Courant, R. and John, F. *Introduction to Calculus and Analysis, Vol. 1.* New York: Springer-Verlag, 1990.
Courant, R. and John, F. *Introduction to Calculus and Analysis, Vol. 2.* New York: Springer-Verlag, 1990.
Hahn, A. *Basic Calculus: From Archimedes to Newton to Its Role in Science.* New York: Springer-Verlag, 1998.
Kaplan, W. *Advanced Calculus, 4th ed.* Reading, MA: Addison-Wesley, 1992.
Marsden, J. E. and Tromba, A. J. *Vector Calculus, 4th ed.* New York: W. H. Freeman, 1996.
Strang, G. *Calculus.* Wellesley, MA: Wellesley-Cambridge Press, 1991.

Calculus of Variations

A branch of mathematics which is a sort of generalization of CALCULUS. Calculus of variations seeks to find the path, curve, surface, etc., for which a given FUNCTION has a STATIONARY VALUE (which, in physical problems, is usually a MINIMUM or MAXIMUM). Mathematically, this involves finding STATIONARY VALUES of integrals of the form

$$I = \int_b^a f(y, \dot{y}, x)\,dx. \tag{1}$$

I has an extremum only if the EULER-LAGRANGE DIFFERENTIAL EQUATION is satisfied, i.e., if

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial \dot{y}}\right) = 0. \tag{2}$$

The FUNDAMENTAL LEMMA OF CALCULUS OF VARIATIONS states that, if

$$\int_a^b M(x)h(x)\,dx = 0 \tag{3}$$

for all $h(x)$ with CONTINUOUS second PARTIAL DERIVATIVES, then

$$M(x) = 0 \tag{4}$$

on (a, b).

see also BELTRAMI IDENTITY, BOLZA PROBLEM, BRACHISTOCHRONE PROBLEM, CATENARY, ENVELOPE THEOREM, EULER-LAGRANGE DIFFERENTIAL EQUATION, ISOPERIMETRIC PROBLEM, ISOVOLUME PROBLEM, LINDELOF'S THEOREM, PLATEAU'S PROBLEM, POINT-POINT DISTANCE—2-D, POINT-POINT DISTANCE—3-D, ROULETTE, SKEW QUADRILATERAL, SPHERE WITH TUNNEL, UNDULOID, WEIERSTRAß-ERDMAN CORNER CONDITION

References

Arfken, G. "Calculus of Variations." Ch. 17 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 925–962, 1985.

Bliss, G. A. *Calculus of Variations.* Chicago, IL: Open Court, 1925.

Forsyth, A. R. *Calculus of Variations.* New York: Dover, 1960.

Fox, C. *An Introduction to the Calculus of Variations.* New York: Dover, 1988.

Isenberg, C. *The Science of Soap Films and Soap Bubbles.* New York: Dover, 1992.

Menger, K. "What is the Calculus of Variations and What are Its Applications?" In *The World of Mathematics* (Ed. K. Newman). Redmond, WA: Microsoft Press, pp. 886–890, 1988.

Sagan, H. *Introduction to the Calculus of Variations.* New York: Dover, 1992.

Todhunter, I. *History of the Calculus of Variations During the Nineteenth Century.* New York: Chelsea, 1962.

Weinstock, R. *Calculus of Variations, with Applications to Physics and Engineering.* New York: Dover, 1974.

Calcus

$$1 \text{ calcus} \equiv \tfrac{1}{2304}.$$

see also HALF, QUARTER, SCRUPLE, UNCIA, UNIT FRACTION

Calderón's Formula

$$f(x) = C_\psi \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \langle f, \psi^{a,b} \rangle\, \psi^{a,b}(x) a^{-2}\,da\,db,$$

where

$$\psi^{a,b}(x) = |a|^{-1/2} \psi\left(\frac{x-b}{a}\right).$$

This result was originally derived using HARMONIC ANALYSIS, but also follows from a WAVELETS viewpoint.

Caliban Puzzle

A puzzle in LOGIC in which one or more facts must be inferred from a set of given facts.

Calvary Cross

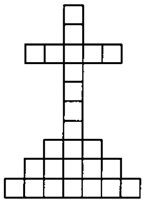

see also CROSS

Cameron's Sum-Free Set Constant

A set of POSITIVE INTEGERS S is sum-free if the equation $x + y = z$ has no solutions x, y, $z \in S$. The probability that a random sum-free set S consists entirely of ODD INTEGERS satisfies

$$0.21759 \leq c \leq 0.21862.$$

References

Cameron, P. J. "Cyclic Automorphisms of a Countable Graph and Random Sum-Free Sets." *Graphs and Combinatorics* 1, 129–135, 1985.

Cameron, P. J. "Portrait of a Typical Sum-Free Set." In *Surveys in Combinatorics 1987* (Ed. C. Whitehead). New York: Cambridge University Press, 13–42, 1987.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/cameron/cameron.html`.

Cancellation

see ANOMALOUS CANCELLATION

Cancellation Law

If $bc \equiv bd \pmod{a}$ and $(b, a) = 1$ (i.e., a and b are RELATIVELY PRIME), then $c \equiv d \pmod{a}$.

see also CONGRUENCE

References

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 36, 1996.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 56, 1993.

Cannonball Problem

Find a way to stack a Square of cannonballs laid out on the ground into a Square Pyramid (i.e., find a Square Number which is also Square Pyramidal). This corresponds to solving the Diophantine Equation

$$\sum_{i=1}^{k} i^2 = \tfrac{1}{6}k(1+k)(1+2k) = N^2$$

for some pyramid height k. The only solution is $k = 24$, $N = 70$, corresponding to 4900 cannonballs (Ball and Coxeter 1987, Dickson 1952), as conjectured by Lucas (1875, 1876) and proved by Watson (1918).

see also Sphere Packing, Square Number, Square Pyramid, Square Pyramidal Number

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 59, 1987.
Dickson, L. E. *History of the Theory of Numbers, Vol. 2: Diophantine Analysis.* New York: Chelsea, p. 25, 1952.
Lucas, É. Question 1180. *Nouvelles Ann. Math. Ser. 2* **14**, 336, 1875.
Lucas, É. Solution de Question 1180. *Nouvelles Ann. Math. Ser. 2* **15**, 429–432, 1876.
Ogilvy, C. S. and Anderson, J. T. *Excursions in Number Theory.* New York: Dover, pp. 77 and 152, 1988.
Pappas, T. "Cannon Balls & Pyramids." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 93, 1989.
Watson, G. N. "The Problem of the Square Pyramid." *Messenger. Math.* **48**, 1–22, 1918.

Canonical Form

A clear-cut way of describing every object in a class in a One-to-One manner.

see also Normal Form, One-to-One

References
Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, p. 7, 1996.

Canonical Polyhedron

A Polyhedron is said to be canonical if all its Edges touch a Sphere and the center of gravity of their contact points is the center of that Sphere. Each combinatorial type of (Genus zero) polyhedron contains just one canonical version. The Archimedean Solids and their Duals are all canonical.

References
Conway, J. H. "Re: polyhedra database." Posting to `geometry.forum` newsgroup, Aug. 31, 1995.

Canonical Transformation

see Symplectic Diffeomorphism

Cantor Comb

see Cantor Set

Cantor-Dedekind Axiom

The points on a line can be put into a One-to-One correspondence with the Real Numbers.

see also Cardinal Number, Continuum Hypothesis, Dedekind Cut

Cantor Diagonal Slash

A clever and rather abstract technique used by Georg Cantor to show that the Integers and Reals cannot be put into a One-to-One correspondence (i.e., the Infinity of Real Numbers is "larger" than the Infinity of Integers). It proceeds by constructing a new member S' of a Set from already known members S by arranging its nth term to differ from the nth term of the nth member of S. The tricky part is that this is done in such a way that the Set including the new member has a larger Cardinality than the original Set S.

see also Cardinality, Continuum Hypothesis, Denumerable Set

References
Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 81–83, 1996.
Penrose, R. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics.* Oxford, England: Oxford University Press, pp. 84–85, 1989.

Cantor Dust

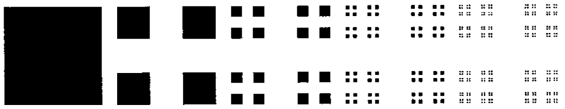

A Fractal which can be constructed using String Rewriting by creating a matrix three times the size of the current matrix using the rules

```
line 1: "*"->"* *"," "->"   "
line 2: "*"->"   "," "->"   "
line 3: "*"->"* *"," "->"   "
```

Let N_n be the number of black boxes, L_n the length of a side of a white box, and A_n the fractional Area of black boxes after the nth iteration.

$$N_n = 5^n \tag{1}$$

$$L_n = (\tfrac{1}{3})^n = 3^{-n} \tag{2}$$

$$A_n = {L_n}^2 N_n = (\tfrac{5}{9})^n. \tag{3}$$

The Capacity Dimension is therefore

$$\begin{aligned} d_{\text{cap}} &= -\lim_{n\to\infty} \frac{\ln N_n}{\ln L_n} = -\lim_{n\to\infty} \frac{\ln(5^n)}{\ln(3^{-n})} \\ &= \frac{\ln 5}{\ln 3} = 1.464973521\ldots. \end{aligned} \tag{4}$$

see also Box Fractal, Sierpiński Carpet, Sierpiński Sieve

References
Dickau, R. M. "Cantor Dust." `http://forum.swarthmore.edu/advanced/robertd/cantor.html`.
Ott, E. *Chaos in Dynamical Systems.* New York: Cambridge University Press, pp. 103–104, 1993.
Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m`.

Cantor's Equation

$$\omega^\epsilon = \epsilon,$$

where ω is an ORDINAL NUMBER and ϵ is an INACCESSIBLE CARDINAL.

see also INACCESSIBLE CARDINAL, ORDINAL NUMBER

References
Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 274, 1996.

Cantor Function

The function whose values are

$$\frac{1}{2}\left(\frac{c_1}{2} + \ldots + \frac{c_{m-1}}{2^{m-1}} + \frac{2}{2^m}\right)$$

for any number between

$$a \equiv \frac{c_1}{3} + \ldots + \frac{c_{m-1}}{3^{m-1}} + \frac{1}{3^m}$$

and

$$b \equiv \frac{c_1}{3} + \ldots + \frac{c_{m-1}}{3^{m-1}} + \frac{2}{3^m}.$$

Chalice (1991) shows that any real-values function $F(x)$ on [0, 1] which is MONOTONE INCREASING and satisfies

1. $F(0) = 0$,
2. $F(x/3) = F(x)/2$,
3. $F(1-x) = 1 - F(x)$

is the Cantor function.

see also CANTOR SET, DEVIL'S STAIRCASE

References
Chalice, D. R. "A Characterization of the Cantor Function." *Amer. Math. Monthly* **98**, 255–258, 1991.
Wagon, S. "The Cantor Function" and "Complex Cantor Sets." §4.2 and 5.1 in *Mathematica in Action.* New York: W. H. Freeman, pp. 102–108 and 143–149, 1991.

Cantor's Paradox

The SET of all SETS is its own POWER SET. Therefore, the CARDINALITY of the SET of all SETS must be bigger than itself.

see also CANTOR'S THEOREM, POWER SET

Cantor Set

The Cantor set (T_∞) is given by taking the interval [0,1] (set T_0), removing the middle third (T_1), removing the middle third of each of the two remaining pieces (T_2), and continuing this procedure ad infinitum. It is therefore the set of points in the INTERVAL [0,1] whose ternary expansions do not contain 1, illustrated below.

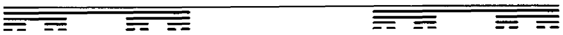

This produces the SET of REAL NUMBERS $\{x\}$ such that

$$x = \frac{c_1}{3} + \ldots + \frac{c_n}{3^n} + \ldots, \tag{1}$$

where c_n may equal 0 or 2 for each n. This is an infinite, PERFECT SET. The total length of the LINE SEGMENTS in the nth iteration is

$$\ell_n = \left(\frac{2}{3}\right)^n, \tag{2}$$

and the number of LINE SEGMENTS is $N_n = 2^n$, so the length of each element is

$$\epsilon_n \equiv \frac{\ell}{N} = \left(\frac{1}{3}\right)^n \tag{3}$$

and the CAPACITY DIMENSION is

$$\begin{aligned} d_{\text{cap}} &\equiv -\lim_{\epsilon\to 0^+} \frac{\ln N}{\ln \epsilon} = -\lim_{n\to\infty} \frac{n \ln 2}{-n \ln 3} \\ &= \frac{\ln 2}{\ln 3} = 0.630929\ldots. \end{aligned} \tag{4}$$

The Cantor set is nowhere DENSE, so it has LEBESGUE MEASURE 0.

A general Cantor set is a CLOSED SET consisting entirely of BOUNDARY POINTS. Such sets are UNCOUNTABLE and may have 0 or POSITIVE LEBESGUE MEASURE. The Cantor set is the only totally disconnected, perfect, COMPACT METRIC SPACE up to a HOMEOMORPHISM (Willard 1970).

see also ALEXANDER'S HORNED SPHERE, ANTOINE'S NECKLACE, CANTOR FUNCTION

References
Boas, R. P. Jr. *A Primer of Real Functions.* Washington, DC: Amer. Math. Soc., 1996.
Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 15–20, 1991.
Willard, S. §30.4 in *General Topology.* Reading, MA: Addison-Wesley, 1970.

Cantor Square Fractal

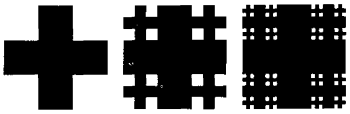

A FRACTAL which can be constructed using STRING REWRITING by creating a matrix three times the size of the current matrix using the rules

line 1: `"*"->"***"," "->"   "`
line 2: `"*"->"* *"," "->"   "`
line 3: `"*"->"***"," "->"   "`

The first few steps are illustrated above.

The size of the unit element after the nth iteration is

$$L_n = \left(\frac{1}{3}\right)^n$$

and the number of elements is given by the RECURRENCE RELATION

$$N_n = 4N_{n-1} + 5(9^n)$$

where $N_1 \equiv 5$, and the first few numbers of elements are 5, 65, 665, 6305, ... Expanding out gives

$$N_n = 5\sum_{k=0}^{n} 4^{n-k}9^{k-1} = 9^n - 4^n.$$

The CAPACITY DIMENSION is therefore

$$\begin{aligned} D &= -\lim_{n\to\infty}\frac{\ln N_n}{\ln L_n} = -\lim_{n\to\infty}\frac{\ln(9^n - 4^n)}{\ln(3^{-n})} \\ &= -\lim_{n\to\infty}\frac{\ln(9^n)}{\ln(3^{-n})} = \frac{\ln 9}{\ln 3} = \frac{2\ln 3}{\ln 3} = 2. \end{aligned}$$

Since the DIMENSION of the filled part is 2 (i.e., the SQUARE is completely filled), Cantor's square fractal is not a true FRACTAL.

see also BOX FRACTAL, CANTOR DUST

References

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 82–83, 1991.

Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m`.

Cantor's Theorem

The CARDINAL NUMBER of any set is lower than the CARDINAL NUMBER of the set of all its subsets. A COROLLARY is that there is no highest ℵ (ALEPH).

see also CANTOR'S PARADOX

Cap

see CROSS-CAP, SPHERICAL CAP

Capacity

see TRANSFINITE DIAMETER

Capacity Dimension

A DIMENSION also called the FRACTAL DIMENSION, HAUSDORFF DIMENSION, and HAUSDORFF-BESICOVITCH DIMENSION in which nonintegral values are permitted. Objects whose capacity dimension is different from their TOPOLOGICAL DIMENSION are called FRACTALS. The capacity dimension of a compact METRIC SPACE X is a REAL NUMBER d_{capacity} such that if $n(\epsilon)$ denotes the minimum number of open sets of diameter less than or equal to ϵ, then $n(\epsilon)$ is proportional to ϵ^{-D} as $\epsilon \to 0$. Explicitly,

$$d_{\text{capacity}} \equiv -\lim_{\epsilon\to 0^+}\frac{\ln N}{\ln \epsilon}$$

(if the limit exists), where N is the number of elements forming a finite COVER of the relevant METRIC SPACE and ϵ is a bound on the diameter of the sets involved (informally, ϵ is the size of each element used to cover the set, which is taken to to approach 0). If each element of a FRACTAL is equally likely to be visited, then $d_{\text{capacity}} = d_{\text{information}}$, where $d_{\text{information}}$ is the INFORMATION DIMENSION. The capacity dimension satisfies

$$d_{\text{correlation}} \leq d_{\text{information}} \leq d_{\text{capacity}}$$

where $d_{\text{correlation}}$ is the CORRELATION DIMENSION, and is conjectured to be equal to the LYAPUNOV DIMENSION.

see also CORRELATION EXPONENT, DIMENSION, HAUSDORFF DIMENSION, KAPLAN-YORKE DIMENSION

References

Nayfeh, A. H. and Balachandran, B. *Applied Nonlinear Dynamics: Analytical, Computational, and Experimental Methods.* New York: Wiley, pp. 538–541, 1995.

Peitgen, H.-O. and Richter, D. H. *The Beauty of Fractals: Images of Complex Dynamical Systems.* New York: Springer-Verlag, 1986.

Wheeden, R. L. and Zygmund, A. *Measure and Integral: An Introduction to Real Analysis.* New York: M. Dekker, 1977.

Carathéodory Derivative

A function f is Carathéodory differentiable at a if there exists a function ϕ which is CONTINUOUS at a such that

$$f(x) - f(a) = \phi(x)(x - a).$$

Every function which is Carathéodory differentiable is also FRÉCHET DIFFERENTIABLE.

see also DERIVATIVE, FRÉCHET DERIVATIVE

Carathéodory's Fundamental Theorem

Each point in the CONVEX HULL of a set S in $\mathbb{R}^n$ is in the convex combination of $n+1$ or fewer points of S.

see also CONVEX HULL, HELLY'S THEOREM

Cardano's Formula

see CUBIC EQUATION

Cardinal Number

In informal usage, a cardinal number is a number used in counting (a COUNTING NUMBER), such as 1, 2, 3, ...

Formally, a cardinal number is a type of number defined in such a way that any method of counting SETS using it gives the same result. (This is not true for the ORDINAL NUMBERS.) In fact, the cardinal numbers are obtained by collecting all ORDINAL NUMBERS which are obtainable by counting a given set. A set has $\aleph_0$ (ALEPH-0) members if it can be put into a ONE-TO-ONE correspondence with the finite ORDINAL NUMBERS.

Two sets are said to have the same cardinal number if all the elements in the sets can be paired off ONE-TO-ONE. An INACCESSIBLE CARDINAL cannot be expressed in terms of a smaller number of smaller cardinals.

see also ALEPH, ALEPH-0 ($\aleph_0$), ALEPH-1 ($\aleph_1$), CANTOR-DEDEKIND AXIOM, CANTOR DIAGONAL SLASH, CONTINUUM, CONTINUUM HYPOTHESIS, EQUIPOLLENT, INACCESSIBLE CARDINALS AXIOM, INFINITY, ORDINAL NUMBER, POWER SET, SURREAL NUMBER, UNCOUNTABLE SET

References

Cantor, G. *Über unendliche, lineare Punktmannigfaltigkeiten, Arbeiten zur Mengenlehre aus dem Jahren 1872–1884.* Leipzig, Germany: Teubner, 1884.

Conway, J. H. and Guy, R. K. "Cardinal Numbers." In *The Book of Numbers.* New York: Springer-Verlag, pp. 277–282, 1996.

Courant, R. and Robbins, H. "Cantor's 'Cardinal Numbers.'" §2.4.3 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 83–86, 1996.

Cardinality

see CARDINAL NUMBER

Cardioid

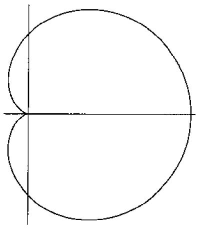

The curve given by the POLAR equation

$$r = a(1 + \cos\theta), \tag{1}$$

sometimes also written

$$r = 2b(1 + \cos\theta), \tag{2}$$

where $b \equiv a/2$, the CARTESIAN equation

$$(x^2 + y^2 - ax)^2 = a^2(x^2 + y^2), \tag{3}$$

and the parametric equations

$$x = a\cos t(1 + \cos t) \tag{4}$$

$$y = a\sin t(1 + \cos t). \tag{5}$$

The cardioid is a degenerate case of the LIMAÇON. It is also a 1-CUSPED EPICYCLOID (with $r = R$) and is the CAUSTIC formed by rays originating at a point on the circumference of a CIRCLE and reflected by the CIRCLE.

The name cardioid was first used by de Castillon in *Philosophical Transactions of the Royal Society* in 1741. Its ARC LENGTH was found by La Hire in 1708. There are exactly three PARALLEL TANGENTS to the cardioid with any given gradient. Also, the TANGENTS at the ends of any CHORD through the CUSP point are at RIGHT ANGLES. The length of any CHORD through the CUSP point is $2a$.

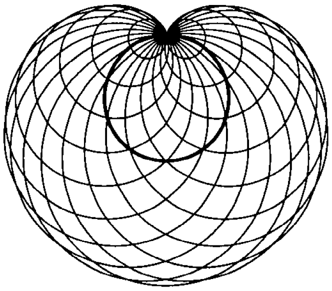

The cardioid may also be generated as follows. Draw a CIRCLE C and fix a point A on it. Now draw a set of CIRCLES centered on the CIRCUMFERENCE of C and passing through A. The ENVELOPE of these CIRCLES is then a cardioid (Pedoe 1995). Let the CIRCLE C be centered at the origin and have RADIUS 1, and let the fixed point be $A = (1, 0)$. Then the RADIUS of a CIRCLE centered at an ANGLE θ from (1, 0) is

$$\begin{aligned} r^2 &= (0 - \cos\theta)^2 + (1 - \sin\theta)^2 \\ &= \cos^2\theta + 1 - 2\sin\theta + \sin^2\theta \\ &= 2(1 - \sin\theta). \end{aligned} \tag{6}$$

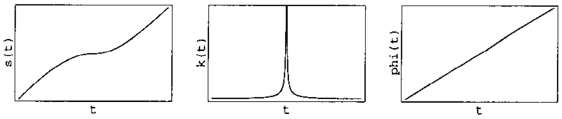

The ARC LENGTH, CURVATURE, and TANGENTIAL ANGLE are

$$s = \int_0^t 2|\cos(\tfrac{1}{2}t)|\,dt = 4a\sin(\tfrac{1}{2}\theta) \tag{7}$$

$$\kappa = \frac{3|\sec(\frac{1}{2}\theta)|}{4a} \tag{8}$$

$$\phi = \tfrac{3}{2}\theta. \tag{9}$$

As usual, care must be taken in the evaluation of $s(t)$ for $t > \pi$. Since (7) comes from an integral involving the

ABSOLUTE VALUE of a function, it must be monotonic increasing. Each QUADRANT can be treated correctly by defining

$$n = \left\lfloor \frac{t}{\pi} \right\rfloor + 1, \tag{10}$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION, giving the formula

$$s(t) = (-1)^{1+[n \pmod 2]} 4 \sin(\tfrac{1}{2}t) + 8 \left\lfloor \tfrac{1}{2}n \right\rfloor. \tag{11}$$

The PERIMETER of the curve is

$$\begin{aligned} L &= \int_0^{2\pi} |2a\cos(\tfrac{1}{2}\theta)|\,d\theta = 4a \int_0^{\pi} \cos(\tfrac{1}{2}\theta)\,d\theta \\ &= 4a \int_0^{\pi/2} \cos\phi (2\,d\phi) = 8a \int_0^{\pi/2} \cos\phi\,d\phi \\ &= 8a[\sin\phi]_0^{\pi/2} = 8a. \end{aligned} \tag{12}$$

The AREA is

$$\begin{aligned} A &= \tfrac{1}{2} \int_0^{2\pi} r^2\,d\theta = \tfrac{1}{2}a^2 \int_0^{2\pi} (1 + 2\cos\theta + \cos^2\theta)\,d\theta \\ &= \tfrac{1}{2}a^2 \int_0^{2\pi} \{1 + 2\cos\theta + \tfrac{1}{2}[1 + \cos(2\theta)]\}\,d\theta \\ &= \tfrac{1}{2}a^2 \int_0^{2\pi} [\tfrac{3}{2} + 2\cos\theta + \tfrac{1}{2}\cos(2\theta)]\,d\theta \\ &= \tfrac{1}{2}a^2[\tfrac{3}{2}\theta + 2\sin\theta + \tfrac{1}{4}\sin(2\theta)]_0^{2\pi} = \tfrac{3}{2}\pi a^2. \end{aligned} \tag{13}$$

see also CIRCLE, CISSOID, CONCHOID, EQUIANGULAR SPIRAL, LEMNISCATE, LIMAÇON, MANDELBROT SET

References

Gray, A. "Cardioids." §3.3 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 41–42, 1993.
Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 118–121, 1972.
Lee, X. "Cardioid." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Cardioid_dir/cardioid.html`.
Lee, X. "Cardioid." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Cardioid_dir/cardioidGG.html`.
Lockwood, E. H. "The Cardioid." Ch. 4 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 34–43, 1967.
MacTutor History of Mathematics Archive. "Cardioid." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Cardioid.html`.
Pedoe, D. *Circles: A Mathematical View, rev. ed.* Washington, DC: Math. Assoc. Amer., pp. xxvi–xxvii, 1995.
Yates, R. C. "The Cardioid." *Math. Teacher* **52**, 10–14, 1959.
Yates, R. C. "Cardioid." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 4–7, 1952.

Cardioid Caustic

The CATACAUSTIC of a CARDIOID for a RADIANT POINT at the CUSP is a NEPHROID. The CATACAUSTIC for PARALLEL rays crossing a CIRCLE is a CARDIOID.

Cardioid Evolute

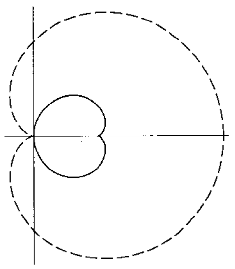

$$\begin{aligned} x &= \tfrac{2}{3}a + \tfrac{1}{3}a\cos\theta(1 - \cos\theta) \\ y &= \tfrac{1}{3}a\sin\theta(1 - \cos\theta). \end{aligned}$$

This is a mirror-image CARDIOID with $a' = a/3$.

Cardioid Inverse Curve

If the CUSP of the cardioid is taken as the INVERSION CENTER, the cardioid inverts to a PARABOLA.

Cardioid Involute

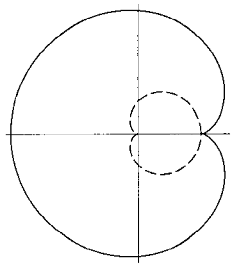

$$\begin{aligned} x &= 2a + 3a\cos\theta(1 - \cos\theta) \\ y &= 3a\sin\theta(1 - \cos\theta). \end{aligned}$$

This is a mirror-image CARDIOID with $a' = 3a$.

Cardioid Pedal Curve

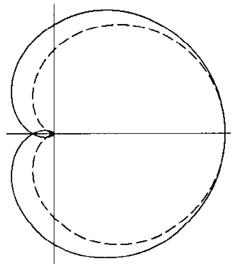

The PEDAL CURVE of the CARDIOID where the PEDAL POINT is the CUSP is CAYLEY'S SEXTIC.

Cards

Cards are a set of n rectangular pieces of cardboard with markings on one side and a uniform pattern on the other. The collection of all cards is called a "deck," and a normal deck of cards consists of 52 cards of four different "suits." The suits are called clubs (♣), diamonds (♢), hearts (♡), and spades (♠). Spades and clubs are colored black, while hearts and diamonds are colored red. The cards of each suit are numbered 1 through 13, where the special terms ace (1), jack (11), queen (12), and king (13) are used instead of numbers 1 and 11–13.

The randomization of the order of cards in a deck is called SHUFFLING. Cards are used in many gambling games (such as POKER), and the investigation of the probabilities of various outcomes in card games was one of the original motivations for the development of modern PROBABILITY theory.

see also BRIDGE CARD GAME, CLOCK SOLITAIRE, COIN, COIN TOSSING, DICE, POKER, SHUFFLE

Carleman's Inequality

Let $\{a_i\}_{i=1}^n$ be a SET of POSITIVE numbers. Then the GEOMETRIC MEAN and ARITHMETIC MEAN satisfy

$$\sum_{i=1}^n (a_1 a_2 \cdots a_i)^{1/i} \leq \frac{e}{n} \sum_{i=1}^n a_i.$$

Here, the constant e is the best possible, in the sense that counterexamples can be constructed for any stricter INEQUALITY which uses a smaller constant.

see also ARITHMETIC MEAN, e, GEOMETRIC MEAN

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1094, 1979.

Hardy, G. H.; Littlewood, J. E.; and Pólya, G. *Inequalities, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 249–250, 1988.

Carlson-Levin Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Assume that f is a NONNEGATIVE REAL function on $[0, \infty)$ and that the two integrals

$$\int_0^\infty x^{p-1-\lambda}[f(x)]^p\,dx \tag{1}$$

$$\int_0^\infty x^{q-1+\mu}[f(x)]^q\,dx \tag{2}$$

exist and are FINITE. If $p = q = 2$ and $\lambda = \mu = 1$, Carlson (1934) determined

$$\int_0^\infty f(x)\,dx \leq \sqrt{\pi}\left(\int_0^\infty [f(x)]^2\,dx\right)^{1/4} \times \left(\int_0^\infty x^2[f(x)]^2\,dx\right)^{1/4} \tag{3}$$

and showed that $\sqrt{\pi}$ is the best constant (in the sense that counterexamples can be constructed for any stricter INEQUALITY which uses a smaller constant). For the general case

$$\int_0^\infty f(x)\,dx \leq C\left(\int_0^\infty x^{p-1-\lambda}[f(x)]^p\,dx\right)^s \times \left(\int_0^\infty x^{q-1+\mu}[f(x)]^q\,dx\right)^t, \tag{4}$$

and Levin (1948) showed that the best constant

$$C = \frac{1}{(ps)^s(qt)^t}\left[\frac{\Gamma\left(\frac{s}{\alpha}\right)\Gamma\left(\frac{t}{\alpha}\right)}{(\lambda+\mu)\Gamma\left(\frac{s+t}{\alpha}\right)}\right]^\alpha, \tag{5}$$

where

$$s \equiv \frac{\mu}{p\mu + q\lambda} \tag{6}$$

$$t \equiv \frac{\lambda}{p\mu + q\lambda} \tag{7}$$

$$\alpha \equiv 1 - s - t \tag{8}$$

and $\Gamma(z)$ is the GAMMA FUNCTION.

References

Beckenbach, E. F.; and Bellman, R. *Inequalities.* New York: Springer-Verlag, 1983.

Boas, R. P. Jr. Review of Levin, V. I. "Exact Constants in Inequalities of the Carlson Type." *Math. Rev.* **9**, 415, 1948.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/crlslvn/crlslvn.html`.

Levin, V. I. "Exact Constants in Inequalities of the Carlson Type." *Doklady Akad. Nauk. SSSR (N. S.)* **59**, 635–638, 1948. English review in Boas (1948).

Mitrinovic, D. S.; Pecaric, J. E.; and Fink, A. M. *Inequalities Involving Functions and Their Integrals and Derivatives.* Kluwer, 1991.

Carlson's Theorem

If $f(z)$ is regular and of the form $\mathcal{O}(e^{k|z|})$ where $k < \pi$, for $\Re[z] \geq 0$, and if $f(z) = 0$ for $z = 0, 1, \ldots$, then $f(z)$ is identically zero.

see also GENERALIZED HYPERGEOMETRIC FUNCTION

References

Bailey, W. N. "Carlson's Theorem." §5.3 in *Generalised Hypergeometric Series.* Cambridge, England: Cambridge University Press, pp. 36–40, 1935.

Carlyle Circle

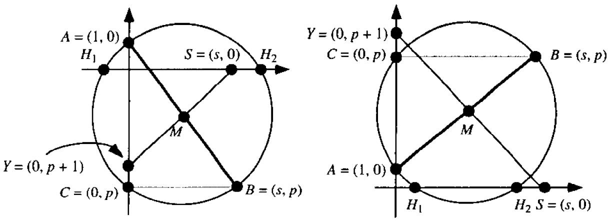

Consider a QUADRATIC EQUATION $x^2 - sx + p = 0$ where s and p denote *signed* lengths. The CIRCLE which has the points $A = (0, 1)$ and $B = (s, p)$ as a DIAMETER is then called the Carlyle circle $C_{s,p}$ of the equation. The CENTER of $C_{s,p}$ is then at the MIDPOINT of AB, $M = (s/2, (1 + p)/2)$, which is also the MIDPOINT of $S = (s, 0)$ and $Y = (0, 1 + p)$. Call the points at which $C_{s,p}$ crosses the x-AXIS $H_1 = (x_1, 0)$ and $H_2 = (x_2, 0)$ (with $x_1 \geq x_2$). Then

$$s = x_1 + x_2$$

$$p = x_1 x_2$$

$$(x - x_1)(x - x_2) = x^2 - sx + p,$$

so x_1 and x_2 are the ROOTS of the quadratic equation.

see also 257-GON, 65537-GON, HEPTADECAGON, PENTAGON

References

De Temple, D. W. "Carlyle Circles and the Lemoine Simplicity of Polygonal Constructions." *Amer. Math. Monthly* **98**, 97–108, 1991.

Eves, H. *An Introduction to the History of Mathematics, 6th ed.* Philadelphia, PA: Saunders, 1990.

Leslie, J. *Elements of Geometry and Plane Trigonometry with an Appendix and Very Copious Notes and Illustrations, 4th ed., improved and exp.* Edinburgh: W. & G. Tait, 1820.

Carmichael Condition

A number n satisfies the Carmichael condition IFF $(p - 1) | (n/p - 1)$ for all PRIME DIVISORS p of n. This is equivalent to the condition $(p - 1) | (n - 1)$ for all PRIME DIVISORS p of n.

see also CARMICHAEL NUMBER

References

Borwein, D.; Borwein, J. M.; Borwein, P. B.; and Girgensohn, R. "Giuga's Conjecture on Primality." *Amer. Math. Monthly* **103**, 40–50, 1996.

Carmichael's Conjecture

Carmichael's conjecture asserts that there are an INFINITE number of CARMICHAEL NUMBERS. This was proven by Alford *et al.* (1994).

see also CARMICHAEL NUMBER, CARMICHAEL'S TOTIENT FUNCTION CONJECTURE

References

Alford, W. R.; Granville, A.; and Pomerance, C. "There Are Infinitely Many Carmichael Numbers." *Ann. Math.* **139**, 703–722, 1994.

Cipra, B. *What's Happening in the Mathematical Sciences, Vol. 1.* Providence, RI: Amer. Math. Soc., 1993.

Guy, R. K. "Carmichael's Conjecture." §B39 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 94, 1994.

Pomerance, C.; Selfridge, J. L.; and Wagstaff, S. S. Jr. "The Pseudoprimes to $25 \cdot 10^9$." *Math. Comput.* **35**, 1003–1026, 1980.

Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, pp. 29–31, 1989.

Schlafly, A. and Wagon, S. "Carmichael's Conjecture on the Euler Function is Valid Below $10^{10,000,000}$." *Math. Comput.* **63**, 415–419, 1994.

Carmichael Function

$\lambda(n)$ is the LEAST COMMON MULTIPLE (LCM) of all the FACTORS of the TOTIENT FUNCTION $\phi(n)$, except that if $8|n$, then $2^{\alpha-2}$ is a FACTOR instead of $2^{\alpha-1}$.

$$\lambda(n) = \begin{cases} \phi(n) & \\ \quad \text{for } n = p^\alpha, p = 2 \text{ and } \alpha \leq 2, \text{ or } p \geq 3 \\ \frac{1}{2}\phi(n) & \\ \quad \text{for } n = 2^\alpha \text{ and } \alpha \geq 3 \\ \text{LCM}[\lambda({p_i}^{\alpha_i})]_i & \\ \quad \text{for } n = \prod_i {p_i}^{\alpha_i}. \end{cases}$$

Some special values are

$$\lambda(1) = 1$$

$$\lambda(2) = 1$$

$$\lambda(4) = 2$$

$$\lambda(2^r) = 2^{r-2}$$

for $r \geq 3$, and

$$\lambda(p^r) = \phi(p^r)$$

for p an ODD PRIME and $r \geq 1$. The ORDER of a (mod n) is at most $\lambda(n)$ (Ribenboim 1989). The values of $\lambda(n)$ for the first few n are 1, 1, 2, 2, 4, 2, 6, 4, 10, 2, 12, ... (Sloane's A011773).

see also MODULO MULTIPLICATION GROUP

References

Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, p. 27, 1989.

Riesel, H. "Carmichael's Function." *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, pp. 273–275, 1994.

Sloane, N. J. A. Sequence A011773 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Vardi, I. *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, p. 226, 1991.

Carmichael Number

A Carmichael number is an ODD COMPOSITE NUMBER n which satisfies FERMAT'S LITTLE THEOREM

$$a^{n-1} - 1 \equiv 0 \pmod{n}$$

for *every* choice of a satisfying $(a, n) = 1$ (i.e., a and n are RELATIVELY PRIME) with $1 < a < n$. A Carmichael number is therefore a PSEUDOPRIMES to any base. Carmichael numbers therefore cannot be found to be COMPOSITE using FERMAT'S LITTLE THEOREM. However, if $(a, n) \neq 1$, the congruence of FERMAT'S LITTLE THEOREM is sometimes NONZERO, thus identifying a Carmichael number n as COMPOSITES.

Carmichael numbers are sometimes called ABSOLUTE PSEUDOPRIMES and also satisfy KORSELT'S CRITERION. R. D. Carmichael first noted the existence of such numbers in 1910, computed 15 examples, and conjectured that there were infinitely many (a fact finally proved by Alford *et al.* 1994).

The first few Carmichael numbers are 561, 1105, 1729, 2465, 2821, 6601, 8911, 10585, 15841, 29341, ... (Sloane's A002997). Carmichael numbers have at least three PRIME FACTORS. For Carmichael numbers with exactly three PRIME FACTORS, once one of the PRIMES has been specified, there are only a finite number of Carmichael numbers which can be constructed. Numbers of the form $(6k+1)(12k+1)(18k+1)$ are Carmichael numbers if each of the factors is PRIME (Korselt 1899, Ore 1988, Guy 1994). This can be seen since for

$$N \equiv (6k+1)(12k+1)(18k+1) = 1296k^3+396k^2+36k+1,$$

$N-1$ is a multiple of $36k$ and the LEAST COMMON MULTIPLE of $6k$, $12k$, and $18k$ is $36k$, so $a^{N-1} \equiv 1$ modulo each of the PRIMES $6k+1$, $12k+1$, and $18k+1$, hence $a^{N-1} \equiv 1$ modulo their product. The first few such Carmichael numbers correspond to $k = 1$, 6, 35, 45, 51, 55, 56, ... and are 1729, 294409, 56052361, 118901521, ... (Sloane's A046025). The largest known Carmichael number of this form was found by H. Dubner in 1996 and has 1025 digits.

The smallest Carmichael numbers having 3, 4, ... factors are $561 = 3 \times 11 \times 17$, $41041 = 7 \times 11 \times 13 \times 41$, 825265, 321197185, ... (Sloane's A006931). In total, there are only 43 Carmichael numbers $< 10^6$, 2163 $\leq 2.5\times10^{10}$, 105,212 $\leq 10^{15}$, and 246,683 $\leq 10^{16}$ (Pinch 1993). Let $C(n)$ denote the number of Carmichael numbers less than n. Then, for sufficiently large n ($n \sim 10^7$ from numerical evidence),

$$C(n) \sim n^{2/7}$$

(Alford *et al.* 1994).

The Carmichael numbers have the following properties:

1. If a PRIME p divides the Carmichael number n, then $n \equiv 1 \pmod{p-1}$ implies that $n \equiv p \pmod{p(p-1)}$.
2. Every Carmichael number is SQUAREFREE.
3. An ODD COMPOSITE SQUAREFREE number n is a Carmichael number IFF n divides the DENOMINATOR of the BERNOULLI NUMBER B_{n-1}.

see also CARMICHAEL CONDITION, PSEUDOPRIME

References

Alford, W. R.; Granville, A.; and Pomerance, C. "There are Infinitely Many Carmichael Numbers." *Ann. Math.* **139**, 703–722, 1994.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 87, 1987.

Guy, R. K. "Carmichael Numbers." §A13 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 30–32, 1994.

Korselt, A. "Problème chinois." *L'intermédiaire math.* **6**, 143–143, 1899.

Ore, Ø. *Number Theory and Its History.* New York: Dover, 1988.

Pinch, R. G. E. "The Carmichael Numbers up to 10^{15}." *Math. Comput.* **55**, 381–391, 1993.

Pinch, R. G. E. `ftp://emu.pmms.cam.ac.uk/pub/Carmichael/`.

Pomerance, C.; Selfridge, J. L.; and Wagstaff, S. S. Jr. "The Pseudoprimes to $25 \cdot 10^9$." *Math. Comput.* **35**, 1003–1026, 1980.

Riesel, H. *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Basel: Birkhäuser, pp. 89–90 and 94–95, 1994.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 116, 1993.

Sloane, N. J. A. Sequences A002997/M5462 and A006931/M5463 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Carmichael Sequence

A FINITE, INCREASING SEQUENCE of INTEGERS $\{a_1, \ldots, a_m\}$ such that

$$(a_i - 1) | (a_1 \cdots a_{i-1})$$

for $i = 1, \ldots, m$, where $m|n$ indicates that m DIVIDES n. A Carmichael sequence has exclusive EVEN or ODD elements. There are infinitely many Carmichael sequences for every order.

see also GIUGA SEQUENCE

References

Borwein, D.; Borwein, J. M.; Borwein, P. B.; and Girgensohn, R. "Giuga's Conjecture on Primality." *Amer. Math. Monthly* **103**, 40–50, 1996.

Carmichael's Theorem

If a and n are RELATIVELY PRIME so that the GREATEST COMMON DENOMINATOR $\mathrm{GCD}(a, n) = 1$, then

$$a^{\lambda(n)} \equiv 1 \pmod{n},$$

where λ is the CARMICHAEL FUNCTION.

Carmichael's Totient Function Conjecture

It is thought that the TOTIENT VALENCE FUNCTION $N_\phi(m) \geq 2$ (i.e., the TOTIENT VALENCE FUNCTION never takes the value 1). This assertion is called Carmichael's totient function conjecture and is equivalent to the statement that there exists an $m \neq n$ such that $\phi(n) = \phi(m)$ (Ribenboim 1996, pp. 39–40). Any counterexample to the conjecture must have more than 10,000 DIGITS (Conway and Guy 1996). Recently, the conjecture was reportedly proven by F. Saidak in November, 1997 with a proof short enough to fit on a postcard.

see also TOTIENT FUNCTION, TOTIENT VALENCE FUNCTION

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 155, 1996.

Ribenboim, P. *The New Book of Prime Number Records.* New York: Springer-Verlag, 1996.

Carnot's Polygon Theorem

If P_1, P_2, ..., are the VERTICES of a finite POLYGON with no "minimal sides" and the side P_iP_j meets a curve in the POINTS P_{ij1} and P_{ij2}, then

$$\frac{\prod_i \overline{P_1P_{12i}}\prod_i \overline{P_2P_{23i}}\cdots\prod_i \overline{P_NP_{N1i}}}{\prod_i \overline{P_NP_{N1i}}\cdots\prod_i \overline{P_2P_{2i1}}} = 1,$$

where $\overline{AB}$ denotes the DISTANCE from POINT A to B.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 190, 1959.

Carnot's Theorem

Given any TRIANGLE $A_1A_2A_3$, the signed sum of PERPENDICULAR distances from the CIRCUMCENTER O to the sides is

$$OO_1 + OO_2 + OO_3 = R + r,$$

where r is the INRADIUS and R is the CIRCUMRADIUS. The sign of the distance is chosen to be POSITIVE IFF the entire segment OO_i lies outside the TRIANGLE.

see also JAPANESE TRIANGULATION THEOREM

References

Eves, H. W. *A Survey of Geometry, rev. ed.* Boston, MA: Allyn and Bacon, pp. 256 and 262, 1972.

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., p. 25, 1985.

Carotid-Kundalini Fractal

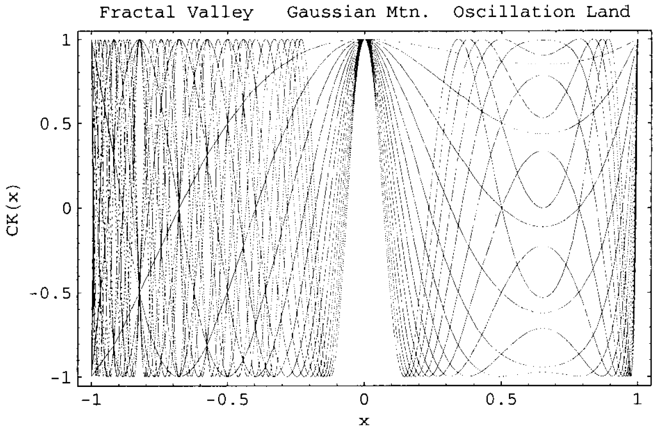

A fractal-like structure is produced for $x < 0$ by superposing plots of CAROTID-KUNDALINI FUNCTIONS CK_n of different orders n. The region $-1 < x < 0$ is called FRACTAL LAND by Pickover (1995), the central region the GAUSSIAN MOUNTAIN RANGE, and the region $x > 0$ OSCILLATION LAND. The plot above shows $n = 1$ to 25. Gaps in FRACTAL LAND occur whenever

$$x \cos^{-1} x = 2\pi \frac{p}{q}$$

for p and q RELATIVELY PRIME INTEGERS. At such points x, the functions assume the $\lceil (q+1)/2 \rceil$ values $\cos(2\pi r/q)$ for $r = 0, 1, \ldots, \lfloor q/2 \rfloor$, where $\lceil z \rceil$ is the CEILING FUNCTION and $\lfloor z \rfloor$ is the FLOOR FUNCTION.

References

Pickover, C. A. "Are Infinite Carotid-Kundalini Functions Fractal?" Ch. 24 in *Keys to Infinity.* New York: W. H. Freeman, pp. 179–181, 1995.

Carotid-Kundalini Function

The FUNCTION given by

$$CK_n(x) \equiv \cos(nx \cos^{-1} x),$$

where n is an INTEGER and $-1 < x < 1$.

see also CAROTID-KUNDALINI FRACTAL

Carry

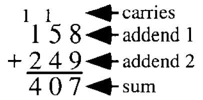

The operating of shifting the leading DIGITS of an ADDITION into the next column to the left when the SUM of that column exceeds a single DIGIT (i.e., 9 in base 10).

see also ADDEND, ADDITION, BORROW

Carrying Capacity

see LOGISTIC GROWTH CURVE

Cartan Matrix

A MATRIX used in the presentation of a LIE ALGEBRA.

References

Jacobson, N. *Lie Algebras.* New York: Dover, p. 121, 1979.

Cartan Relation

The relationship $Sq^i(x \smile y) = \Sigma_{j+k=i} Sq^j(x) \smile Sq^k(y)$ encountered in the definition of the STEENROD ALGEBRA.

Cartan Subgroup

A type of maximal Abelian SUBGROUP.

References

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Cartan Torsion Coefficient

The ANTISYMMETRIC parts of the CONNECTION COEFFICIENT $\Gamma^\lambda{}_{\mu\nu}$.

Cartesian Coordinates

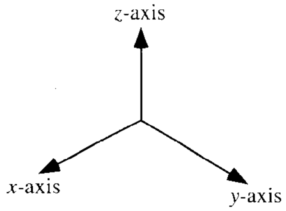

Cartesian coordinates are rectilinear 2-D or 3-D coordinates (and therefore a special case of CURVILINEAR COORDINATES) which are also called RECTANGULAR COORDINATES. The three axes of 3-D Cartesian coordinates, conventionally denoted the x-, y-, and z-AXES (a NOTATION due to Descartes) are chosen to be linear and mutually PERPENDICULAR. In 3-D, the coordinates x, y, and z may lie anywhere in the INTERVAL $(-\infty, \infty)$.

The SCALE FACTORS of Cartesian coordinates are all unity, $h_i = 1$. The LINE ELEMENT is given by

$$d\mathbf{s} = dx\,\hat{\mathbf{x}} + dy\,\hat{\mathbf{y}} + dz\,\hat{\mathbf{z}}, \tag{1}$$

and the VOLUME ELEMENT by

$$dV = dx\,dy\,dz. \tag{2}$$

The GRADIENT has a particularly simple form,

$$\nabla \equiv \hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{y}}\frac{\partial}{\partial y} + \hat{\mathbf{z}}\frac{\partial}{\partial z}, \tag{3}$$

as does the LAPLACIAN

$$\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{4}$$

The LAPLACIAN is

$$\begin{aligned}\nabla^2\mathbf{F} \equiv \nabla\cdot(\nabla\mathbf{F}) &= \frac{\partial^2\mathbf{F}}{\partial x^2} + \frac{\partial^2\mathbf{F}}{\partial y^2} + \frac{\partial^2\mathbf{F}}{\partial z^2}\\ &= \hat{\mathbf{x}}\left(\frac{\partial^2 F_x}{\partial x^2} + \frac{\partial^2 F_x}{\partial y^2} + \frac{\partial^2 F_x}{\partial z^2}\right)\\ &\quad + \hat{\mathbf{y}}\left(\frac{\partial^2 F_y}{\partial x^2} + \frac{\partial^2 F_y}{\partial y^2} + \frac{\partial^2 F_y}{\partial z^2}\right)\\ &\quad + \hat{\mathbf{z}}\left(\frac{\partial^2 F_z}{\partial x^2} + \frac{\partial^2 F_z}{\partial y^2} + \frac{\partial^2 F_z}{\partial z^2}\right).\end{aligned} \tag{5}$$

The DIVERGENCE is

$$\nabla\cdot\mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}, \tag{6}$$

and the CURL is

$$\begin{aligned}\nabla\times\mathbf{F} &= \begin{vmatrix}\hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}}\\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z}\\ F_x & F_y & F_z\end{vmatrix}\\ &= \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right)\hat{\mathbf{x}} + \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}\right)\hat{\mathbf{y}}\\ &\quad + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right)\hat{\mathbf{z}}.\end{aligned} \tag{7}$$

The GRADIENT of the DIVERGENCE is

$$\begin{aligned}\nabla(\nabla\cdot\mathbf{u}) &= \begin{bmatrix}\frac{\partial}{\partial x}\left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z}\right)\\ \frac{\partial}{\partial y}\left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z}\right)\\ \frac{\partial}{\partial z}\left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z}\right)\end{bmatrix}\\ &= \begin{bmatrix}\frac{\partial}{\partial x}\\ \frac{\partial}{\partial y}\\ \frac{\partial}{\partial z}\end{bmatrix}\left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z}\right).\end{aligned} \tag{8}$$

LAPLACE'S EQUATION is separable in Cartesian coordinates.

see also COORDINATES, HELMHOLTZ DIFFERENTIAL EQUATION—CARTESIAN COORDINATES

References

Arfken, G. "Special Coordinate Systems—Rectangular Cartesian Coordinates." §2.3 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 94–95, 1985.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 656, 1953.

Cartesian Ovals

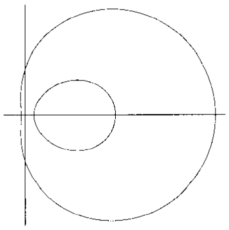

A curve consisting of two ovals which was first studied by Descartes in 1637. It is the locus of a point P whose distances from two FOCI F_1 and F_2 in two-center BIPOLAR COORDINATES satisfy

$$mr \pm nr' = k, \tag{1}$$

where m, n are POSITIVE INTEGERS, k is a POSITIVE real, and r and r' are the distances from F_1 and F_2. If $m = n$, the oval becomes an an ELLIPSE. In CARTESIAN COORDINATES, the Cartesian ovals can be written

$$m\sqrt{(x-a)^2 + y^2} + n\sqrt{(x+a)^2 + y^2} = k^2 \tag{2}$$

$$\begin{aligned}(x^2 + y^2 + a^2)(m^2 - n^2) - 2ax(m^2 + n^2) - k^2\\ = -2n\sqrt{(x+a)^2 + y^2},\end{aligned} \tag{3}$$

$$\begin{aligned}&[(m^2 - n^2)(x^2 + y^2 + a^2) - 2ax(m^2 + n^2)]^2\\ &= 2(m^2 + n^2)(n^2 + y^2 + a^2) - 4ax(m^2 - n^2) - k^2.\end{aligned} \tag{4}$$

Now define

$$b \equiv m^2 - n^2 \tag{5}$$

$$c \equiv m^2 + n^2, \tag{6}$$

and set $a = 1$. Then

$$[b(x^2+y^2)-2cx+b]^2+4bx+k^2-2c = 2c(x^2+y^2). \quad (7)$$

If c' is the distance between F_1 and F_2, and the equation

$$r + mr' = a \quad (8)$$

is used instead, an alternate form is

$$[(1-m^2)(x^2+y^2)+2m^2c'x+a'^2-m^2c'^2]^2 = 4a'^2(x^2+y^2). \quad (9)$$

The curves possess three FOCI. If $m = 1$, one Cartesian oval is a central CONIC, while if $m = a/c$, then the curve is a LIMAÇON and the inside oval touches the outside one. Cartesian ovals are ANALLAGMATIC CURVES.

References
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 35, 1989.
Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 155–157, 1972.
Lockwood, E. H. *A Book of Curves.* Cambridge, England: Cambridge University Press, p. 188, 1967.
MacTutor History of Mathematics Archive. "Cartesian Oval." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Cartesian.html`.

Cartesian Product

see DIRECT PRODUCT (SET)

Cartesian Trident

see TRIDENT OF DESCARTES

Cartography

The study of MAP PROJECTIONS and the making of geographical maps.

see also MAP PROJECTION

Cascade

A $\mathbb{Z}$-ACTION or $\mathbb{N}$-ACTION. A cascade and a single MAP $X \to X$ are essentially the same, but the term "cascade" is preferred by many Russian authors.

see also ACTION, FLOW

Casey's Theorem

Four CIRCLES are TANGENT to a fifth CIRCLE or a straight LINE IFF

$$t_{12}t_{34} \pm t_{13}t_{42} \pm t_{14}t_{23} = 0,$$

where t_{ij} is a common TANGENT to CIRCLES i and j.

see also PURSER'S THEOREM

References
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 121–127, 1929.

Casimir Operator

An OPERATOR

$$\Gamma = \sum_{i=1}^{m} e_i^R u^{iR}$$

on a representation R of a LIE ALGEBRA.

References
Jacobson, N. *Lie Algebras.* New York: Dover, p. 78, 1979.

Cassini Ellipses

see CASSINI OVALS

Cassini's Identity

For F_n the nth FIBONACCI NUMBER,

$$F_{n-1}F_{n+1} - F_n{}^2 = (-1)^n.$$

see also FIBONACCI NUMBER

References
Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, p. 12, 1996.

Cassini Ovals

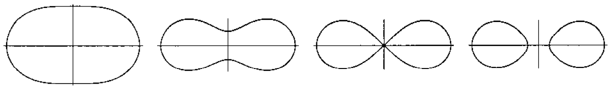

The curves, also called CASSINI ELLIPSES, described by a point such that the product of its distances from two fixed points a distance $2a$ apart is a constant b^2. The shape of the curve depends on b/a. If $a < b$, the curve is a single loop with an OVAL (left figure above) or dog bone (second figure) shape. The case $a = b$ produces a LEMNISCATE (third figure). If $a > b$, then the curve consists of two loops (right figure). The curve was first investigated by Cassini in 1680 when he was studying the relative motions of the Earth and the Sun. Cassini believed that the Sun traveled around the Earth on one of these ovals, with the Earth at one FOCUS of the oval.

Cassini ovals are ANALLAGMATIC CURVES. The Cassini ovals are defined in two-center BIPOLAR COORDINATES by the equation

$$r_1 r_2 = b^2, \quad (1)$$

with the origin at a FOCUS. Even more incredible curves are produced by the locus of a point the product of whose distances from 3 or more fixed points is a constant.

The Cassini ovals have the CARTESIAN equation

$$[(x-a)^2+y^2][(x+a)^2+y^2] = b^4 \quad (2)$$

or the equivalent form

$$(x^2+y^2+a^2)^2 - 4a^2x^2 = b^4 \quad (3)$$

and the polar equation

$$r^4 + a^4 - 2a^2r^2\cos(2\theta) = b^4. \qquad (4)$$

Solving for r^2 using the QUADRATIC EQUATION gives

$$\begin{aligned} r^2 &= \frac{2a^2\cos(2\theta) + \sqrt{4a^4\cos^2(2\theta) - 4(a^4 - b^4)}}{2} \\ &= a^2\cos(2\theta) + \sqrt{a^4\cos^2(2\theta) + b^4 - a^4} \\ &= a^2\cos(2\theta)\sqrt{a^4[\cos^2(2\theta) - 1] + b^4} \\ &= a^2\cos(2\theta) + \sqrt{b^4 - a^4\sin^2(2\theta)} \\ &= a^2\left[\cos(2\theta) + \sqrt{\left(\frac{b}{a}\right)^4 - \sin^2(2\theta)}\,\right]. \qquad (5) \end{aligned}$$

If $a < b$, the curve has AREA

$$A = \tfrac{1}{2}r^2\,d\theta = 2(\tfrac{1}{2})\int_{-\pi/4}^{\pi/4} r^2\,d\theta = a^2 + b^2E\left(\frac{a^4}{b^4}\right), \qquad (6)$$

where the integral has been done over half the curve and then multiplied by two and $E(x)$ is the complete ELLIPTIC INTEGRAL OF THE SECOND KIND. If $a = b$, the curve becomes

$$r^2 = a^2\left[\cos(2\theta) + \sqrt{1 - \sin^2\theta}\,\right] = 2a^2\cos(2\theta), \qquad (7)$$

which is a LEMNISCATE having AREA

$$A = 2a^2 \qquad (8)$$

(two loops of a curve $\sqrt{2}$ the linear scale of the usual lemniscate $r^2 = a^2\cos(2\theta)$, which has area $A = a^2/2$ for each loop). If $a > b$, the curve becomes two disjoint ovals with equations

$$r = \pm a\sqrt{\cos(2\theta) \pm \sqrt{\left(\frac{b}{a}\right)^4 - \sin^2(2\theta)}}\,, \qquad (9)$$

where $\theta \in [-\theta_0, \theta_0]$ and

$$\theta_0 \equiv \tfrac{1}{2}\sin^{-1}\left[\left(\frac{b}{a}\right)^2\right]. \qquad (10)$$

see also CASSINI SURFACE, LEMNISCATE, MANDELBROT SET, OVAL

References

Gray, A. "Cassinian Ovals." §4.2 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 63–65, 1993.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 153–155, 1972.

Lee, X. "Cassinian Oval." `http://www.best.com/~xah/SpecialPlaneCurves_dir/CassinianOval_dir/cassinianOval.html`.

Lockwood, E. H. *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 187–188, 1967.

MacTutor History of Mathematics Archive. "Cassinian Ovals." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Cassinian.html`.

Yates, R. C. "Cassinian Curves." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 8–11, 1952.

Cassini Projection

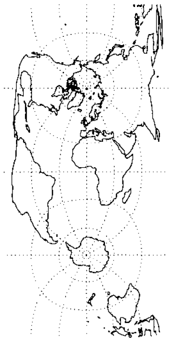

A MAP PROJECTION.

$$x = \sin^{-1} B \qquad (1)$$

$$y = \tan^{-1}\left[\frac{\tan\phi}{\cos(\lambda - \lambda_0)}\right], \qquad (2)$$

where

$$B = \cos\phi\sin(\lambda - \lambda_0). \qquad (3)$$

The inverse FORMULAS are

$$\phi = \sin^{-1}(\sin D\cos x) \qquad (4)$$

$$\lambda = \lambda_0 + \tan^{-1}\left(\frac{\tan x}{\cos D}\right), \qquad (5)$$

where

$$D = y + \phi_0. \qquad (6)$$

References

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 92–95, 1987.

Cassini Surface

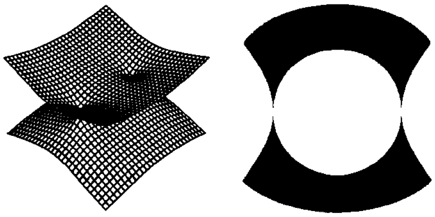

The QUARTIC SURFACE obtained by replacing the constant c in the equation of the CASSINI OVALS

$$[(x-a)^2 + y^2][(x+a)^2 + y^2] = c^2 \qquad (1)$$

by $c = z^2$, obtaining

$$[(x-a)^2 + y^2][(x+a)^2 + y^2] = z^4. \qquad (2)$$

As can be seen by letting $y = 0$ to obtain

$$(x^2 - a^2)^2 = z^4 \qquad (3)$$

$$x^2 + z^2 = a^2, \qquad (4)$$

the intersection of the surface with the $y = 0$ PLANE is a CIRCLE of RADIUS a.

References
Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, p. 20, 1986.
Fischer, G. (Ed.). Plate 51 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 51, 1986.

Castillon's Problem

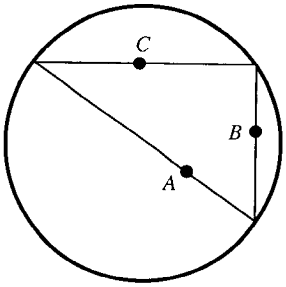

Inscribe a TRIANGLE in a CIRCLE such that the sides of the TRIANGLE pass through three given POINTS A, B, and C.

References
Dörrie, H. "Castillon's Problem." §29 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 144–147, 1965.

Casting Out Nines

An elementary check of a MULTIPLICATION which makes use of the CONGRUENCE $10^n \equiv 1 \pmod 9$ for $n \geq 2$. From this CONGRUENCE, a MULTIPLICATION $ab = c$ must give

$$a \equiv \sum a_i = a^*$$

$$b \equiv \sum b_i = b^*$$

$$c \equiv \sum c_i = c^*,$$

so $ab \equiv a^*b^*$ must be $\equiv c^* \pmod 9$. Casting out nines is sometimes also called "the HINDU CHECK."

References
Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 28–29, 1996.

Cat Map

see ARNOLD'S CAT MAP

Catacaustic

The curve which is the ENVELOPE of reflected rays.

Curve	Source	Catacaustic
cardioid	cusp	nephroid
circle	not on circumf.	limaçon
circle	on circumf.	cardioid
circle	point at ∞	nephroid
cissoid of Diocles	focus	cardioid
1 arch of a cycloid	rays $\perp$ axis	2 arches of a cycloid
deltoid	point at ∞	astroid
$\ln x$	rays $\|\|$ axis	catenary
logarithmic spiral	origin	equal logarithmic spiral
parabola	rays $\perp$ axis	Tschirnhausen cubic
quadrifolium	center	astroid
Tschirnhausen cubic	focus	semicubical parabola

see also CAUSTIC, DIACAUSTIC

References
Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 60 and 207, 1972.

Catalan's Conjecture

8 and 9 (2^3 and 3^2) are the only consecutive POWERS (excluding 0 and 1), i.e., the only solution to CATALAN'S DIOPHANTINE PROBLEM. Solutions to this problem (CATALAN'S DIOPHANTINE PROBLEM) are equivalent to solving the simultaneous DIOPHANTINE EQUATIONS

$$X^2 - Y^3 = 1$$

$$X^3 - Y^2 = 1.$$

This CONJECTURE has not yet been proved or refuted, although it has been shown to be decidable in a FINITE (but more than astronomical) number of steps. In particular, if n and $n+1$ are POWERS, then $n < \exp\exp\exp\exp 730$ (Guy 1994, p. 155), which follows from R. Tijdeman's proof that there can be only a FINITE number of exceptions should the CONJECTURE not hold.

Hyyrő and Mąkowski proved that there do not exist three consecutive POWERS (Ribenboim 1996), and it is also known that 8 and 9 are the only consecutive CUBIC and SQUARE NUMBERS (in either order).

see also CATALAN'S DIOPHANTINE PROBLEM

References
Guy, R. K. "Difference of Two Power." §D9 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 155–157, 1994.
Ribenboim, P. *Catalan's Conjecture.* Boston, MA: Academic Press, 1994.
Ribenboim, P. "Catalan's Conjecture." *Amer. Math. Monthly* **103**, 529–538, 1996.
Ribenboim, P. "Consecutive Powers." *Expositiones Mathematicae* **2**, 193–221, 1984.

Catalan's Constant

A constant which appears in estimates of combinatorial functions. It is usually denoted K, $\beta(2)$, or G. It is not known if K is IRRATIONAL. Numerically,

$$K = 0.915965594177\ldots \tag{1}$$

(Sloane's A006752). The CONTINUED FRACTION for K is [0, 1, 10, 1, 8, 1, 88, 4, 1, 1, ...] (Sloane's A014538). K can be given analytically by the following expressions,

$$\begin{aligned} K &\equiv \beta(2) & (2)\\ &= \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)^2} = \frac{1}{1^2} - \frac{1}{3^2} + \frac{1}{5^2} + \ldots & (3)\\ &= 1 + \sum_{n=1}^{\infty} \frac{1}{(4n+1)^2} - \frac{1}{9} - \sum_{n=1}^{\infty} \frac{1}{(4n+3)^2} & (4)\\ &= \int_0^1 \frac{\tan^{-1} x\,dx}{x} & (5)\\ &= -\int_0^1 \frac{\ln x\,dx}{1+x^2}, & (6) \end{aligned}$$

where $\beta(z)$ is the DIRICHLET BETA FUNCTION. In terms of the POLYGAMMA FUNCTION $\Psi_1(x)$,

$$\begin{aligned} K &= \tfrac{1}{16}\Psi_1\left(\tfrac{1}{4}\right) - \tfrac{1}{16}\Psi_1\left(\tfrac{3}{4}\right) & (7)\\ &= \tfrac{1}{80}\Psi_1(\tfrac{5}{12}) + \tfrac{1}{80}\Psi_1(\tfrac{1}{12}) - \tfrac{1}{10}\pi^2 & (8)\\ &= \tfrac{1}{32}\Psi_1(\tfrac{1}{8}) - \tfrac{1}{32}\Psi_1(\tfrac{3}{8}) - \tfrac{1}{16}\sqrt{2}. & (9) \end{aligned}$$

Applying CONVERGENCE IMPROVEMENT to (3) gives

$$K = \frac{1}{16} \sum_{m=1}^{\infty} (m+1) \frac{3^m - 1}{4^m} \zeta(m+2), \tag{10}$$

where $\zeta(z)$ is the RIEMANN ZETA FUNCTION and the identity

$$\frac{1}{(1-3z)^2} - \frac{1}{(1-z)^2} = \sum_{m=1}^{\infty} (m+1) \frac{3^m - 1}{4^m} z^m \tag{11}$$

has been used (Flajolet and Vardi 1996). The Flajolet and Vardi algorithm also gives

$$K = \frac{1}{\sqrt{2}} \prod_{k=1}^{\infty} \left[\left(1 - \frac{1}{2^{2^k}}\right) \frac{\zeta(2^k)}{\beta(2^k)}\right]^{1/(2^{k+1})}, \tag{12}$$

where $\beta(z)$ is the DIRICHLET BETA FUNCTION. Glaisher (1913) gave

$$K = 1 - \sum_{n=1}^{\infty} \frac{n\zeta(2n+1)}{16^n} \tag{13}$$

(Vardi 1991, p. 159). W. Gosper used the related FORMULA

$$K = \frac{1}{\sqrt{2}} \left[\frac{1}{\Psi(2) - 1}\right]^{2^{1/2}} \prod_{k=2}^{\infty} \left[\frac{1}{-\Psi(2^k) - 1}\right]^{1/(2^{k+1})}, \tag{14}$$

where

$$\Psi(m) = \frac{m\psi_{m-1}(\frac{1}{4})}{\pi^m (2^m - 1) 4^{m-1} B_m}, \tag{15}$$

where B_n is a BERNOULLI NUMBER and $\psi(x)$ is a POLYGAMMA FUNCTION (Finch). The Catalan constant may also be defined by

$$K \equiv \tfrac{1}{2} \int_0^1 K(k)\,dk, \tag{16}$$

where $K(k)$ (not to be confused with Catalan's constant itself, denoted K) is a complete ELLIPTIC INTEGRAL OF THE FIRST KIND.

$$K = \frac{\pi \ln 2}{8} + \sum_{i=1}^{\infty} \frac{a_i}{2^{\lfloor (i+1)/2 \rfloor} i^2}, \tag{17}$$

where

$$\{a_i\} = \{\overline{1, 1, 1, 0, -1, -1, -1, 0}\} \tag{18}$$

is given by the periodic sequence obtained by appending copies of {1, 1, 1, 0, −1, −1, −1, 0} (in other words, $a_i \equiv a_{[(i-1) \pmod 8)]+1}$ for $i > 8$) and $\lfloor x \rfloor$ is the FLOOR FUNCTION (Nielsen 1909).

see also DIRICHLET BETA FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 807–808, 1972.

Adamchik, V. "32 Representations for Catalan's Constant." `http://www.wolfram.com/~victor/articles/catalan/catalan.html`.

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 551–552, 1985.

Fee, G. J. "Computation of Catalan's Constant using Ramanujan's Formula." *ISAAC '90. Proc. Internat. Symp. Symbolic Algebraic Comp., Aug. 1990.* Reading, MA: Addison-Wesley, 1990.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/catalan/catalan.html`.

Flajolet, P. and Vardi, I. "Zeta Function Expansions of Classical Constants." Unpublished manuscript. 1996. `http://pauillac.inria.fr/algo/flajolet/Publications/landau.ps`.

Glaisher, J. W. L. "Numerical Values of the Series $1 - 1/3^n + 1/5^n - 1/7^n + 1/9^n - \&c$ for $n = 2, 4, 6$." *Messenger Math.* **42**, 35–58, 1913.

Gosper, R. W. "A Calculus of Series Rearrangements." In *Algorithms and Complexity: New Directions and Recent Results* (Ed. J. F. Traub). New York: Academic Press, 1976.

Nielsen, N. *Der Eulersche Dilogarithms.* Leipzig, Germany: Halle, pp. 105 and 151, 1909.

Plouffe, S. "Plouffe's Inverter: Table of Current Records for the Computation of Constants." `http://lacim.uqam.ca/pi/records.html`.
Sloane, N. J. A. Sequences A014538 and A006752/M4593 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Srivastava, H. M. and Miller, E. A. "A Simple Reducible Case of Double Hypergeometric Series involving Catalan's Constant and Riemann's Zeta Function." *Int. J. Math. Educ. Sci. Technol.* **21**, 375–377, 1990.
Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 159, 1991.
Yang, S. "Some Properties of Catalan's Constant *G*." *Int. J. Math. Educ. Sci. Technol.* **23**, 549–556, 1992.

Catalan's Diophantine Problem

Find consecutive POWERS, i.e., solutions to

$$a^b - c^d = 1,$$

excluding 0 and 1. CATALAN'S CONJECTURE is that the only solution is $3^2 - 2^3 = 1$, so 8 and 9 (2^3 and 3^2) are the only consecutive POWERS (again excluding 0 and 1).

see also CATALAN'S CONJECTURE

References

Cassels, J. W. S. "On the Equation $a^x - b^y = 1$. II." *Proc. Cambridge Phil. Soc.* **56**, 97–103, 1960.
Inkeri, K. "On Catalan's Problem." *Acta Arith.* **9**, 285–290, 1964.

Catalan Integrals

Special cases of general FORMULAS due to Bessel.

$$J_0(\sqrt{z^2-y^2}\,) = \frac{1}{\pi}\int_0^\pi e^{y\cos\theta}\cos(z\sin\theta)\,d\theta,$$

where J_0 is a BESSEL FUNCTION OF THE FIRST KIND. Now, let $z \equiv 1 - z'$ and $y \equiv 1 + z'$. Then

$$J_0(2i\sqrt{z}\,) = \frac{1}{\pi}\int_0^\pi e^{(1+z)\cos\theta}\cos[(1-z)\sin\theta]\,d\theta.$$

Catalan Number

The Catalan numbers are an INTEGER SEQUENCE $\{C_n\}$ which appears in TREE enumeration problems of the type, "In how many ways can a regular n-gon be divided into $n-2$ TRIANGLES if different orientations are counted separately?" (EULER'S POLYGON DIVISION PROBLEM). The solution is the Catalan number C_{n-2} (Dörrie 1965, Honsberger 1973), as graphically illustrated below (Dickau).

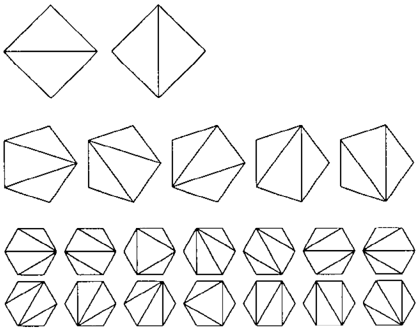

The first few Catalan numbers are 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, ... (Sloane's A000108). The only ODD Catalan numbers are those of the form c_{2^k-1}, and the last DIGIT is five for $k = 9$ to 15. The only PRIME Catalan numbers for $n \leq 2^{15} - 1$ are $C_2 = 2$ and $C_3 = 5$.

The Catalan numbers turn up in many other related types of problems. For instance, the Catalan number C_{n-1} gives the number of BINARY BRACKETINGS of n letters (CATALAN'S PROBLEM). The Catalan numbers also give the solution to the BALLOT PROBLEM, the number of trivalent PLANTED PLANAR TREES (Dickau), the number of states possible in an n-FLEXAGON, the number of different diagonals possible in a FRIEZE PATTERN with $n+1$ rows, the number of ways of forming an n-fold exponential, the number of rooted planar binary trees with n internal nodes, the number of rooted plane bushes with n EDGES, the number of extended BINARY TREES with n internal nodes, the number of mountains which can be drawn with n upstrokes and n downstrokes, the number of noncrossing handshakes possible across a round table between n pairs of people (Conway and Guy 1996), and the number of SEQUENCES with NONNEGATIVE PARTIAL SUMS which can be formed from n 1s and n -1s (Bailey 1996, Buraldi 1992)!

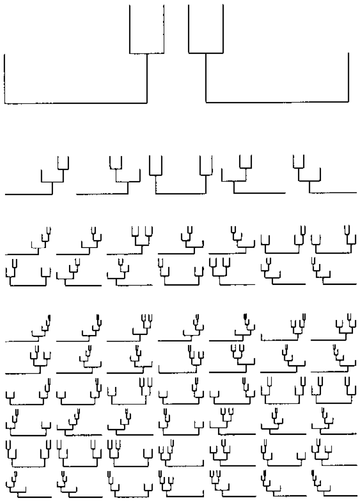

An explicit formula for C_n is given by

$$C_n \equiv \frac{1}{n+1}\binom{2n}{n} = \frac{1}{n+1}\frac{(2n)!}{n!^2} = \frac{(2n)!}{(n+1)!n!}, \qquad (1)$$

where $\binom{2n}{n}$ denotes a BINOMIAL COEFFICIENT and $n!$ is the usual FACTORIAL. A RECURRENCE RELATION for C_n is obtained from

$$\begin{aligned}\frac{C_{n+1}}{C_n} &= \frac{(2n+2)!}{(n+2)[(n+1)!]^2}\frac{(n+1)(n!)^2}{(2n)!} \\ &= \frac{(2n+2)(2n+1)(n+1)}{(n+2)(n+1)^2} \\ &= \frac{2(2n+1)(n+1)^2}{(n+1)^2(n+2)} = \frac{2(2n+1)}{n+2}, \qquad (2)\end{aligned}$$

so

$$C_{n+1} = \frac{2(2n+1)}{n+2} C_n. \tag{3}$$

Other forms include

$$\begin{aligned} C_n &= \frac{2 \cdot 6 \cdot 10 \cdots (4n-2)}{(n+1)!} && (4)\\ &= \frac{2^n (2n-1)!!}{(n+1)!} && (5)\\ &= \frac{(2n)!}{n!(n+1)!}. && (6) \end{aligned}$$

SEGNER'S RECURRENCE FORMULA, given by Segner in 1758, gives the solution to EULER'S POLYGON DIVISION PROBLEM

$$E_n = E_2 E_{n-1} + E_3 E_{n-2} + \ldots + E_{n-1} E_2. \tag{7}$$

With $E_1 = E_2 = 1$, the above RECURRENCE RELATION gives the Catalan number $C_{n-2} = E_n$.

The GENERATING FUNCTION for the Catalan numbers is given by

$$\frac{1 - \sqrt{1-4x}}{2x} = \sum_{n=0}^{\infty} C_n x^n = 1 + x + 2x^2 + 5x^3 + \ldots. \tag{8}$$

The asymptotic form for the Catalan numbers is

$$C_k \sim \frac{4^k}{\sqrt{\pi}\, k^{3/2}} \tag{9}$$

(Vardi 1991, Graham *et al.* 1994).

A generalization of the Catalan numbers is defined by

$$_p d_k = \frac{1}{k} \binom{pk}{k-1} = \frac{1}{(p-1)k+1} \binom{pk}{k} \tag{10}$$

for $k \geq 1$ (Klarner 1970, Hilton and Pederson 1991). The usual Catalan numbers $C_k = {}_2 d_k$ are a special case with $p = 2$. $_p d_k$ gives the number of p-ary TREES with k source-nodes, the number of ways of associating k applications of a given p-ary OPERATOR, the number of ways of dividing a convex POLYGON into k disjoint $(p+1)$-gons with nonintersecting DIAGONALS, and the number of p-GOOD PATHS from $(0, -1)$ to $(k, (p-1)k-1)$ (Hilton and Pederson 1991).

A further generalization is obtained as follows. Let p be an INTEGER > 1, let $P_k = (k, (p-1)k - 1)$ with $k \geq 0$, and $q \leq p - 1$. Then define $_p d_{q0} = 1$ and let $_p d_{qk}$ be the number of p-GOOD PATHS from $(1, q-1)$ to P_k (Hilton and Pederson 1991). Formulas for $_p d_{qi}$ include the generalized JONAH FORMULA

$$\binom{n-q}{k-1} = \sum_{i=1}^{k} {}_p d_{qi} \binom{n-pi}{k-i} \tag{11}$$

and the explicit formula

$$_p d_{qk} = \frac{p-q}{pk-q} \binom{pk-q}{k-1}. \tag{12}$$

A RECURRENCE RELATION is given by

$$_p d_{qk} = \sum_{i,j} {}_p d_{p-r,i}\, {}_p d_{q+r,j} \tag{13}$$

where $i, j, r \geq 1$, $k \geq 1$, $q < p - r$, and $i + j = k + 1$ (Hilton and Pederson 1991).

see also BALLOT PROBLEM, BINARY BRACKETING, BINARY TREE, CATALAN'S PROBLEM, CATALAN'S TRIANGLE, DELANNOY NUMBER, EULER'S POLYGON DIVISION PROBLEM, FLEXAGON, FRIEZE PATTERN, MOTZKIN NUMBER, p-GOOD PATH, PLANTED PLANAR TREE, SCHRÖDER NUMBER, SUPER CATALAN NUMBER

References

Alter, R. "Some Remarks and Results on Catalan Numbers." *Proc. 2nd Louisiana Conf. Comb., Graph Th., and Comput.*, 109–132, 1971.

Alter, R. and Kubota, K. K. "Prime and Prime Power Divisibility of Catalan Numbers." *J. Combin. Th.* **15**, 243–256, 1973.

Bailey, D. F. "Counting Arrangements of 1's and −1's." *Math. Mag.* **69**, 128–131, 1996.

Brualdi, R. A. *Introductory Combinatorics, 3rd ed.* New York: Elsevier, 1997.

Campbell, D. "The Computation of Catalan Numbers." *Math. Mag.* **57**, 195–208, 1984.

Chorneyko, I. Z. and Mohanty, S. G. "On the Enumeration of Certain Sets of Planted Trees." *J. Combin. Th. Ser. B* **18**, 209–221, 1975.

Chu, W. "A New Combinatorial Interpretation for Generalized Catalan Numbers." *Disc. Math.* **65**, 91–94, 1987.

Conway, J. H. and Guy, R. K. In *The Book of Numbers.* New York: Springer-Verlag, pp. 96–106, 1996.

Dershowitz, N. and Zaks, S. "Enumeration of Ordered Trees." *Disc. Math.* **31**, 9–28, 1980.

Dickau, R. M. "Catalan Numbers." `http://forum.swarthmore.edu/advanced/robertd/catalan.html`.

Dörrie, H. "Euler's Problem of Polygon Division." §7 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 21–27, 1965.

Eggleton, R. B. and Guy, R. K. "Catalan Strikes Again! How Likely is a Function to be Convex?" *Math. Mag.* **61**, 211–219, 1988.

Gardner, M. "Catalan Numbers." Ch. 20 in *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, 1988.

Gardner, M. "Catalan Numbers: An Integer Sequence that Materializes in Unexpected Places." *Sci. Amer.* **234**, 120–125, June 1976.

Gould, H. W. *Bell & Catalan Numbers: Research Bibliography of Two Special Number Sequences, 6th ed.* Morgantown, WV: Math Monongliae, 1985.

Graham, R. L.; Knuth, D. E.; and Patashnik, O. Exercise 9.8 in *Concrete Mathematics: A Foundation for Computer Science, 2nd ed.* Reading, MA: Addison-Wesley, 1994.

Guy, R. K. "Dissecting a Polygon Into Triangles." *Bull. Malayan Math. Soc.* **5**, 57–60, 1958.

Hilton, P. and Pederson, J. "Catalan Numbers, Their Generalization, and Their Uses." *Math. Int.* **13**, 64–75, 1991.

Honsberger, R. *Mathematical Gems I.* Washington, DC: Math. Assoc. Amer., pp. 130–134, 1973.

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 146–150, 1985.
Klarner, D. A. "Correspondences Between Plane Trees and Binary Sequences." *J. Comb. Th.* **9**, 401–411, 1970.
Rogers, D. G. "Pascal Triangles, Catalan Numbers and Renewal Arrays." *Disc. Math.* **22**, 301–310, 1978.
Sands, A. D. "On Generalized Catalan Numbers." *Disc. Math.* **21**, 218–221, 1978.
Singmaster, D. "An Elementary Evaluation of the Catalan Numbers." *Amer. Math. Monthly* **85**, 366–368, 1978.
Sloane, N. J. A. *A Handbook of Integer Sequences.* Boston, MA: Academic Press, pp. 18–20, 1973.
Sloane, N. J. A. Sequences A000108/M1459 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Sloane, N. J. A. and Plouffe, S. Extended entry in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.
Vardi, I. *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 187–188 and 198–199, 1991.
Wells, D. G. *The Penguin Dictionary of Curious and Interesting Numbers.* London: Penguin, pp. 121–122, 1986.

Catalan's Problem

The problem of finding the number of different ways in which a PRODUCT of n different ordered FACTORS can be calculated by pairs (i.e., the number of BINARY BRACKETINGS of n letters). For example, for the four FACTORS a, b, c, and d, there are five possibilities: $((ab)c)d$, $(a(bc))d$, $(ab)(cd)$, $a((bc)d)$, and $a(b(cd))$. The solution was given by Catalan in 1838 as

$$C'_n = \frac{2\cdot 6\cdot 10\cdot (4n-6)}{n!}$$

and is equal to the CATALAN NUMBER $C_{n-1} = C'_n$.

see also BINARY BRACKETING, CATALAN'S DIOPHANTINE PROBLEM, EULER'S POLYGON DIVISION PROBLEM

References
Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, p. 23, 1965.

Catalan Solid

The DUAL POLYHEDRA of the ARCHIMEDEAN SOLIDS, given in the following table.

Archimedean Solid	Dual
rhombicosidodecahedron	deltoidal hexecontahedron
small rhombicuboctahedron	deltoidal icositetrahedron
great rhombicuboctahedron	disdyakis dodecahedron
great rhombicosidodecahedron	disdyakis triacontahedron
truncated icosahedron	pentakis dodecahedron
snub dodecahedron (laevo)	pentagonal hexecontahedron (dextro)
snub cube (laevo)	pentagonal icositetrahedron (dextro)
cuboctahedron	rhombic dodecahedron
icosidodecahedron	rhombic triacontahedron
truncated octahedron	tetrakis hexahedron
truncated dodecahedron	triakis icosahedron
truncated cube	triakis octahedron
truncated tetrahedron	triakis tetrahedron

Here are the ARCHIMEDEAN DUALS (Holden 1971, Pearce 1978) displayed in alphabetical order (left to right, then continuing to the next row).

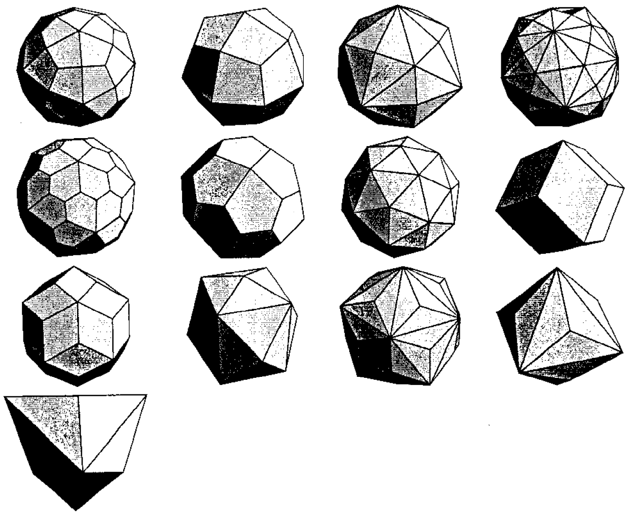

Here are the Archimedean solids paired with the corresponding Catalan solids.

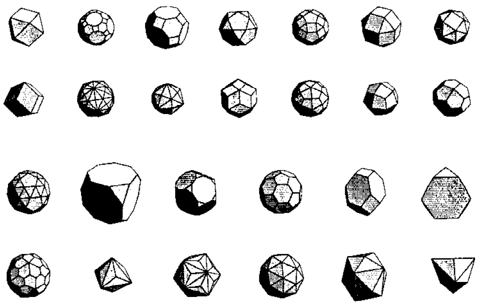

see also ARCHIMEDEAN SOLID, DUAL POLYHEDRON, SEMIREGULAR POLYHEDRON

References
Catalan, E. "Mémoire sur la Théorie des Polyèdres." *J. l'École Polytechnique (Paris)* **41**, 1–71, 1865.
Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, 1991.

Catalan's Surface

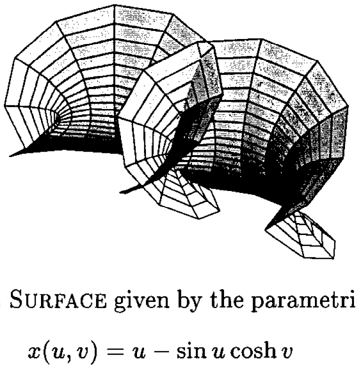

A MINIMAL SURFACE given by the parametric equations

$$x(u,v) = u - \sin u \cosh v \tag{1}$$

$$y(u,v) = 1 - \cos u \cosh v \tag{2}$$

$$z(u,v) = 4\sin(\tfrac{1}{2}u)\sinh(\tfrac{1}{2}v) \tag{3}$$

(Gray 1993), or

$$x(r,\phi) = a\sin(2\phi) - 2a\phi + \tfrac{1}{2}av^2\cos(2\phi) \tag{4}$$

$$y(r,\phi) = -a\cos(2\phi) - \tfrac{1}{2}av^2\cos(2\phi) \tag{5}$$

$$z(r,\phi) = 2av\sin\phi, \tag{6}$$

where

$$v = -r + \frac{1}{r} \tag{7}$$

(do Carmo 1986).

References

Catalan, E. "Mémoir sur les surfaces dont les rayons de courburem en chaque point, sont égaux et des signes contraires." *C. R. Acad. Sci. Paris* **41**, 1019–1023, 1855.

do Carmo, M. P. "Catalan's Surface" §3.5D in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 45–46, 1986.

Fischer, G. (Ed.). Plates 94–95 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 90–91, 1986.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.*Boca Raton, FL: CRC Press, pp. 448–449, 1993.

Catalan's Triangle

A triangle of numbers with entries given by

$$c_{nm} = \frac{(n+m)!(n-m+1)}{m!(n+1)!}$$

for $0 \leq m \leq n$, where each element is equal to the one above plus the one to the left. Furthermore, the sum of each row is equal to the last element of the next row and also equal to the CATALAN NUMBER C_n.

1						
1	1					
1	2	2				
1	3	5	5			
1	4	9	14	14		
1	5	14	28	42	42	
1	6	20	48	90	132	132

(Sloane's A009766).

see also BELL TRIANGLE, CLARK'S TRIANGLE, EULER'S TRIANGLE, LEIBNIZ HARMONIC TRIANGLE, NUMBER TRIANGLE, PASCAL'S TRIANGLE, PRIME TRIANGLE, SEIDEL-ENTRINGER-ARNOLD TRIANGLE

References

Sloane, N. J. A. Sequence A009766 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Catalan's Trisectrix

see TSCHIRNHAUSEN CUBIC

Catastrophe

see BUTTERFLY CATASTROPHE, CATASTROPHE THEORY, CUSP CATASTROPHE, ELLIPTIC UMBILIC CATASTROPHE, FOLD CATASTROPHE, HYPERBOLIC UMBILIC CATASTROPHE, PARABOLIC UMBILIC CATASTROPHE, SWALLOWTAIL CATASTROPHE

Catastrophe Theory

Catastrophe theory studies how the qualitative nature of equation solutions depends on the parameters that appear in the equations. Subspecializations include bifurcation theory, nonequilibrium thermodynamics, singularity theory, synergetics, and topological dynamics. For any system that seeks to minimize a function, only seven different local forms of catastrophe "typically" occur for four or fewer variables: (1) FOLD CATASTROPHE, (2) CUSP CATASTROPHE, (3) SWALLOWTAIL CATASTROPHE, (4) BUTTERFLY CATASTROPHE, (5) ELLIPTIC UMBILIC CATASTROPHE, (6) HYPERBOLIC UMBILIC CATASTROPHE, (7) PARABOLIC UMBILIC CATASTROPHE.

More specifically, for any system with fewer than five control factors and fewer than three behavior axes, these are the only seven catastrophes possible. The following tables gives the possible catastrophes as a function of control factors and behavior axes (Goetz).

Control Factors	1 Behavior Axis	2 Behavior Axes
1	fold	—
2	cusp	—
3	swallowtail	hyperbolic umbilic, elliptic umbilic
4	butterfly	parabolic umbilic

References

Arnold, V. I. *Catastrophe Theory, 3rd ed.* Berlin: Springer-Verlag, 1992.

Gilmore, R. *Catastrophe Theory for Scientists and Engineers.* New York: Dover, 1993.

Goetz, P. "Phil's Good Enough Complexity Dictionary." `http://www.cs.buffalo.edu/~goetz/dict.html`.

Saunders, P. T. *An Introduction to Catastrophe Theory.* Cambridge, England: Cambridge University Press, 1980.

Stewart, I. *The Problems of Mathematics, 2nd ed.* Oxford, England: Oxford University Press, p. 211, 1987.

Thom, R. *Structural Stability and Morphogenesis: An Outline of a General Theory of Models.* Reading, MA: Reading, MA: Addison-Wesley, 1993.

Thompson, J. M. T. *Instabilities and Catastrophes in Science and Engineering.* New York: Wiley, 1982.

Woodcock, A. E. R. and Davis, M. *Catastrophe Theory.* New York: E. P. Dutton, 1978.

Zeeman, E. C. *Catastrophe Theory—Selected Papers 1972–1977.* Reading, MA: Addison-Wesley, 1977.

Categorical Game

A GAME in which no draw is possible.

Categorical Variable

A variable which belongs to exactly one of a finite number of CATEGORIES.

Category

A category consists of two things: an OBJECT and a MORPHISM (sometimes called an "arrow"). An OBJECT is some mathematical structure (e.g., a GROUP, VECTOR SPACE, or DIFFERENTIABLE MANIFOLD) and a MORPHISM is a MAP between two OBJECTS. The MORPHISMS are then required to satisfy some fairly natural conditions; for instance, the IDENTITY MAP between any object and itself is always a MORPHISM, and the composition of two MORPHISMS (if defined) is always a MORPHISM.

One usually requires the MORPHISMS to preserve the mathematical structure of the objects. So if the objects are all groups, a good choice for a MORPHISM would be a group HOMOMORPHISM. Similarly, for vector spaces, one would choose linear maps, and for differentiable manifolds, one would choose differentiable maps.

In the category of TOPOLOGICAL SPACES, homomorphisms are usually continuous maps between topological spaces. However, there are also other category structures having TOPOLOGICAL SPACES as objects, but they are not nearly as important as the "standard" category of TOPOLOGICAL SPACES and continuous maps.

see also ABELIAN CATEGORY, ALLEGORY, EILENBERG-STEENROD AXIOMS, GROUPOID, HOLONOMY, LOGOS, MONODROMY, TOPOS

References

Freyd, P. J. and Scedrov, A. *Categories, Allegories.* Amsterdam, Netherlands: North-Holland, 1990.

Category Theory

The branch of mathematics which formalizes a number of algebraic properties of collections of transformations between mathematical objects (such as binary relations, groups, sets, topological spaces, etc.) of the same type, subject to the constraint that the collections contain the identity mapping and are closed with respect to compositions of mappings. The objects studied in category theory are called CATEGORIES.

see also CATEGORY

Catenary

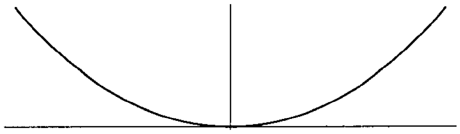

The curve a hanging flexible wire or chain assumes when supported at its ends and acted upon by a uniform gravitational force. The word catenary is derived from the Latin word for "chain." In 1669, Jungius disproved Galileo's claim that the curve of a chain hanging under gravity would be a PARABOLA (MacTutor Archive). The curve is also called the ALYSOID and CHAINETTE. The equation was obtained by Leibniz, Huygens, and Johann Bernoulli in 1691 in response to a challenge by Jakob Bernoulli.

Huygens was the first to use the term catenary in a letter to Leibniz in 1690, and David Gregory wrote a treatise on the catenary in 1690 (MacTutor Archive). If you roll a PARABOLA along a straight line, its FOCUS traces out a catenary. As proved by Euler in 1744, the catenary is also the curve which, when rotated, gives the surface of minimum SURFACE AREA (the CATENOID) for the given bounding CIRCLE.

The CARTESIAN equation for the catenary is given by

$$y = \tfrac{1}{2}a(e^{x/a} + e^{-x/a}) = a\cosh\left(\frac{x}{a}\right), \tag{1}$$

and the CESÀRO EQUATION is

$$(s^2 + a^2)\kappa = -a. \tag{2}$$

The catenary gives the shape of the road over which a regular polygonal "wheel" can travel smoothly. For a regular n-gon, the corresponding catenary is

$$y = -A\cosh\left(\frac{x}{A}\right), \tag{3}$$

where

$$A \equiv R\cos\left(\frac{\pi}{n}\right). \tag{4}$$

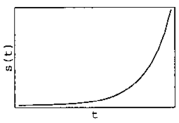

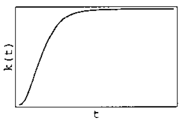

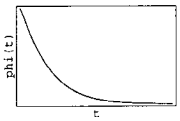

The ARC LENGTH, CURVATURE, and TANGENTIAL ANGLE are

$$s = a\sinh\left(\frac{t}{a}\right), \tag{5}$$

$$\kappa = -\frac{1}{a}\operatorname{sech}^2\left(\frac{t}{a}\right), \tag{6}$$

$$\phi = -2\tan^{-1}\left[\tanh\left(\frac{t}{2a}\right)\right]. \tag{7}$$

The slope is proportional to the ARC LENGTH as measured from the center of symmetry.

see also CALCULUS OF VARIATIONS, CATENOID, LINDELOF'S THEOREM, SURFACE OF REVOLUTION

References

Geometry Center. "The Catenary." `http://www.geom.umn.edu/zoo/diffgeom/surfspace/catenoid/catenary.html`.

Gray, A. "The Evolute of a Tractrix is a Catenary." §5.3 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 80–81, 1993.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 195 and 199–200, 1972.

Lockwood, E. H. "The Tractrix and Catenary." Ch. 13 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 118–124, 1967.

MacTutor History of Mathematics Archive. "Catenary." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Catenary.html`.

Pappas, T. "The Catenary & the Parabolic Curves." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 34, 1989.

Yates, R. C. "Catenary." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 12–14, 1952.

Catenary Evolute

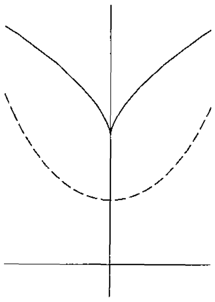

$$x = a[x - \tfrac{1}{2}\sinh(2t)]$$
$$y = 2a\cosh t.$$

Catenary Involute

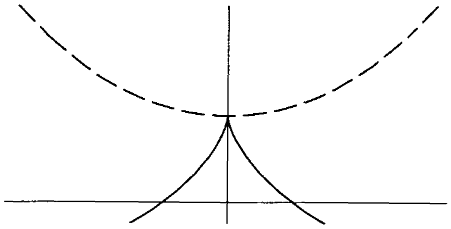

The parametric equation for a CATENARY is

$$\mathbf{r}(t) = a\begin{bmatrix} t \\ \cosh t \end{bmatrix}, \qquad (1)$$

so

$$\frac{d\mathbf{r}}{dt} = a\begin{bmatrix} 1 \\ \sinh t \end{bmatrix} \qquad (2)$$
$$\left|\frac{d\mathbf{r}}{dt}\right| = a\sqrt{1+\sinh^2 t} = a\cosh t \qquad (3)$$

and

$$\hat{\mathbf{T}} = \frac{\frac{d\mathbf{r}}{dt}}{\left|\frac{d\mathbf{r}}{dt}\right|} = \begin{bmatrix} \operatorname{sech} t \\ \tanh t \end{bmatrix} \qquad (4)$$
$$ds^2 = |d\mathbf{r}^2| = a^2(1+\sinh^2 t)\,dt^2 = a^2\cosh^2\,dt^2 \qquad (5)$$
$$\frac{ds}{dt} = a\cosh t. \qquad (6)$$

Therefore,

$$s = a\int \cosh t\,dt = a\sinh t \qquad (7)$$

and the equation of the INVOLUTE is

$$x = a(t - \tanh t) \qquad (8)$$
$$y = a\operatorname{sech} t. \qquad (9)$$

This curve is called a TRACTRIX.

Catenary Radial Curve

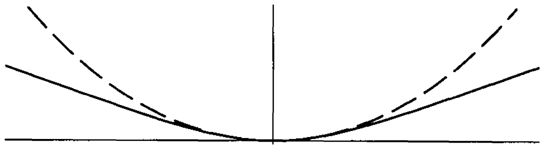

The KAMPYLE OF EUDOXUS.

Catenoid

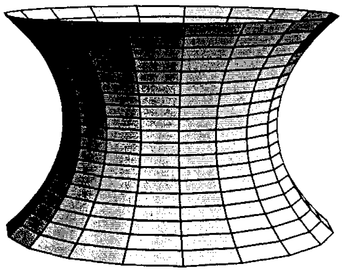

A CATENARY of REVOLUTION. The catenoid and PLANE are the only SURFACES OF REVOLUTION which are also MINIMAL SURFACES. The catenoid can be given by the parametric equations

$$x = c\cosh\left(\frac{v}{c}\right)\cos u \qquad (1)$$
$$y = c\cosh\left(\frac{v}{c}\right)\sin u \qquad (2)$$
$$z = v, \qquad (3)$$

where $u \in [0, 2\pi)$. The differentials are

$$dx = \sinh\left(\frac{v}{c}\right)\cos u\,dv - \cosh\left(\frac{v}{c}\right)\sin u\,du \qquad (4)$$
$$dy = \sinh\left(\frac{v}{c}\right)\sin u\,dv + \cosh\left(\frac{v}{c}\right)\cos u\,du \qquad (5)$$
$$dz = du, \qquad (6)$$

so the LINE ELEMENT is

$$\begin{aligned} ds^2 &= dx^2 + dy^2 + dz^2 \\ &= \left[\sinh^2\left(\frac{v}{c}\right) + 1\right] dv^2 + \cosh^2\left(\frac{v}{c}\right) du^2 \\ &= \cosh^2\left(\frac{v}{c}\right) dv^2 + \cosh^2\left(\frac{v}{c}\right) du^2. \end{aligned} \qquad (7)$$

The PRINCIPAL CURVATURES are

$$\kappa_1 = -\frac{1}{c}\operatorname{sech}^2\left(\frac{v}{c}\right) \qquad (8)$$
$$\kappa_2 = \frac{1}{c}\operatorname{sech}^2\left(\frac{v}{c}\right). \qquad (9)$$

The MEAN CURVATURE of the catenoid is

$$H = 0 \qquad (10)$$

and the Gaussian Curvature is

$$K = -\frac{1}{c^2}\operatorname{sech}^4\left(\frac{v}{c}\right). \tag{11}$$

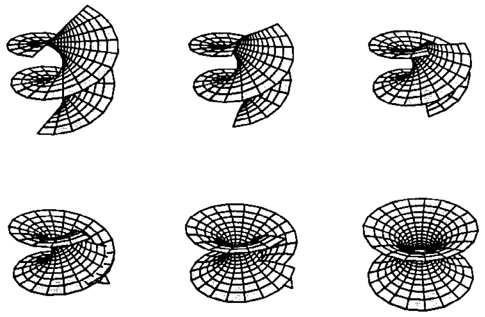

The Helicoid can be continuously deformed into a catenoid with $c = 1$ by the transformation

$$x(u,v) = \cos\alpha \sinh v \sin u + \sin\alpha \cosh v \cos u \tag{12}$$

$$y(u,v) = -\cos\alpha \sinh v \cos u + \sin\alpha \cosh v \sin u \tag{13}$$

$$z(u,v) = u\cos\alpha + v\sin\alpha, \tag{14}$$

where $\alpha = 0$ corresponds to a Helicoid and $\alpha = \pi/2$ to a catenoid.

see also Catenary, Costa Minimal Surface, Helicoid, Minimal Surface, Surface of Revolution

References

do Carmo, M. P. "The Catenoid." §3.5A in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, p. 43, 1986.

Fischer, G. (Ed.). Plate 90 in *Mathematische Modelle/ Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 86, 1986.

Geometry Center. "The Catenoid." `http://www.geom.umn.edu/zoo/diffgeom/surfspace/catenoid/`.

Gray, A. "The Catenoid." §18.4 *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 367–369, 1993.

Meusnier, J. B. "Mémoire sur la courbure des surfaces." *Mém. des savans étrangers* **10** (lu 1776), 477–510, 1785.

Caterpillar Graph

A Tree with every Node on a central stalk or only one Edge away from the stalk.

References

Gardner, M. *Wheels, Life, and other Mathematical Amusements.* New York: W. H. Freeman, p. 160, 1983.

Cattle Problem of Archimedes

see Archimedes' Cattle Problem

Cauchy Binomial Theorem

$$\sum_{m=0}^{n} y^m q^{m(m+1)/2} \binom{n}{m}_q = \prod_{k=1}^{n}(1+yq^k),$$

where $\binom{n}{m}_q$ is a Gaussian Coefficient.

see also q-Binomial Theorem

Cauchy Boundary Conditions

Boundary Conditions of a Partial Differential Equation which are a weighted Average of Dirichlet Boundary Conditions (which specify the value of the function on a surface) and Neumann Boundary Conditions (which specify the normal derivative of the function on a surface).

see also Boundary Conditions, Cauchy Problem, Dirichlet Boundary Conditions, Neumann Boundary Conditions

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 678–679, 1953.

Cauchy's Cosine Integral Formula

$$\int_{-\pi/2}^{\pi/2} \cos^{\mu+\nu-2}\theta\, e^{i\theta(\mu-\nu+2\xi)}\,d\theta = \frac{\pi\Gamma(\mu+\nu-1)}{2^{\mu+\nu-2}\Gamma(\mu+\xi)\Gamma(\nu-\xi)},$$

where $\Gamma(z)$ is the Gamma Function.

Cauchy Criterion

A Necessary and Sufficient condition for a Sequence S_i to Converge. The Cauchy criterion is satisfied when, for all $\epsilon > 0$, there is a fixed number N such that $|S_j - S_i| < \epsilon$ for all $i, j > N$.

Cauchy Distribution

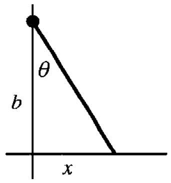

The Cauchy distribution, also called the Lorentzian Distribution, describes resonance behavior. It also describes the distribution of horizontal distances at which a Line Segment tilted at a random Angle cuts the x-Axis. Let θ represent the Angle that a line, with fixed point of rotation, makes with the vertical axis, as shown above. Then

$$\tan\theta = \frac{x}{b} \tag{1}$$

$$\theta = \tan^{-1}\left(\frac{x}{b}\right) \tag{2}$$

$$d\theta = -\frac{1}{1+\frac{x^2}{b^2}}\frac{dx}{b} = -\frac{b\,dx}{b^2+x^2}, \tag{3}$$

so the distribution of Angle θ is given by

$$\frac{d\theta}{\pi} = -\frac{1}{\pi}\frac{b\,dx}{b^2+x^2}. \tag{4}$$

This is normalized over all angles, since

$$\int_{-\pi/2}^{\pi/2} \frac{d\theta}{\pi} = 1 \tag{5}$$

and

$$\begin{aligned} -\int_{-\infty}^{\infty} \frac{1}{\pi} \frac{b\,dx}{b^2+x^2} &= \frac{1}{\pi}\left[\tan^{-1}\left(\frac{b}{x}\right)\right]_{-\infty}^{\infty} \\ &= \frac{1}{\pi}[\tfrac{1}{2}\pi - (-\tfrac{1}{2}\pi)] = 1. \end{aligned} \tag{6}$$

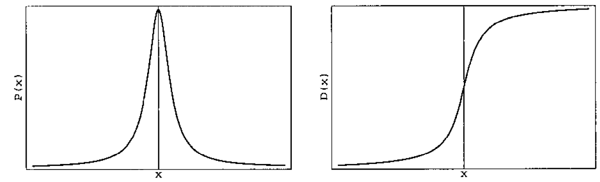

The general Cauchy distribution and its cumulative distribution can be written as

$$P(x) = \frac{1}{\pi} \frac{\frac{1}{2}\Gamma}{(x-\mu)^2 + (\frac{1}{2}\Gamma)^2} \tag{7}$$

$$D(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x-\mu}{b}\right), \tag{8}$$

where Γ is the FULL WIDTH AT HALF MAXIMUM ($\Gamma = 2b$ in the above example) and μ is the MEAN ($\mu = 0$ in the above example). The CHARACTERISTIC FUNCTION is

$$\begin{aligned} \phi(t) &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{it(\Gamma x/2-\mu)}}{1+x^2}\,dx \\ &= \frac{e^{-i\mu t}}{\pi}\int_{-\infty}^{\infty} \frac{\cos(\Gamma t x/2)}{1+(\Gamma x/2)^2}\,dx \\ &= e^{-i\mu t - \Gamma|t|/2}. \end{aligned} \tag{9}$$

The MOMENTS are given by

$$\mu_2 = \sigma^2 = \infty \tag{10}$$

$$\mu_3 = \begin{cases} 0 & \text{for } \mu = 0 \\ \infty & \text{for } \mu \neq 0 \end{cases} \tag{11}$$

$$\mu_4 = \infty, \tag{12}$$

and the STANDARD DEVIATION, SKEWNESS, and KURTOSIS by

$$\sigma^2 = \infty \tag{13}$$

$$\gamma_1 = \begin{cases} 0 & \text{for } \mu = 0 \\ \infty & \text{for } \mu \neq 0 \end{cases} \tag{14}$$

$$\gamma_2 = \infty. \tag{15}$$

If X and Y are variates with a NORMAL DISTRIBUTION, then $Z \equiv X/Y$ has a Cauchy distribution with MEAN $\mu = 0$ and full width

$$\Gamma = \frac{2\sigma_y}{\sigma_x}. \tag{16}$$

see also GAUSSIAN DISTRIBUTION, NORMAL DISTRIBUTION

References

Spiegel, M. R. *Theory and Problems of Probability and Statistics.* New York: McGraw-Hill, pp. 114–115, 1992.

Cauchy Equation

see EULER EQUATION

Cauchy's Formula

The GEOMETRIC MEAN is smaller than the ARITHMETIC MEAN,

$$\left(\prod_{i=1}^{N} n_i\right)^{1/N} < \frac{\sum_{i=1}^{N} n_i}{N}.$$

Cauchy Functional Equation

The fifth of HILBERT'S PROBLEMS is a generalization of this equation.

Cauchy-Hadamard Theorem

The RADIUS OF CONVERGENCE of the TAYLOR SERIES

$$a_0 + a_1 z + a_2 z^2 + \ldots$$

is

$$r = \frac{1}{\overline{\lim\limits_{n\to\infty}}\,(|a_n|)^{1/n}}.$$

see also RADIUS OF CONVERGENCE, TAYLOR SERIES

Cauchy Inequality

A special case of the HÖLDER SUM INEQUALITY with $p = q = 2$,

$$\left(\sum_{k=1}^{n} a_k b_k\right)^2 \leq \left(\sum_{k=1}^{n} {a_k}^2\right)\left(\sum_{k=1}^{n} {b_k}^2\right), \tag{1}$$

where equality holds for $a_k = cb_k$. In 2-D, it becomes

$$(a^2+b^2)(c^2+a^2) \geq (ac+bd)^2. \tag{2}$$

It can be proven by writing

$$\sum_{i=1}^{n} (a_i x + b_i)^2 = \sum_{i=1}^{n} {a_i}^2 \left(x + \frac{b_i}{a_i}\right)^2 = 0. \tag{3}$$

If b_i/a_i is a constant c, then $x = -c$. If it is not a constant, then all terms cannot simultaneously vanish for REAL x, so the solution is COMPLEX and can be found using the QUADRATIC EQUATION

$$x = \frac{-2\sum a_i b_i \pm \sqrt{4\left(\sum a_i b_i\right)^2 - 4\sum {a_i}^2 \sum {b_i}^2}}{2\sum {a_i}^2}. \tag{4}$$

In order for this to be COMPLEX, it must be true that

$$\left(\sum_i a_i b_i\right)^2 \leq \left(\sum_i {a_i}^2\right)\left(\sum_i {b_i}^2\right), \qquad (5)$$

with equality when b_i/a_i is a constant. The VECTOR derivation is much simpler,

$$(\mathbf{a}\cdot\mathbf{b})^2 = a^2 b^2 \cos^2\theta \leq a^2 b^2, \qquad (6)$$

where

$$a^2 \equiv \mathbf{a}\cdot\mathbf{a} = \sum_i {a_i}^2, \qquad (7)$$

and similarly for b.

see also CHEBYSHEV INEQUALITY, HÖLDER SUM INEQUALITY

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 11, 1972.

Cauchy Integral Formula

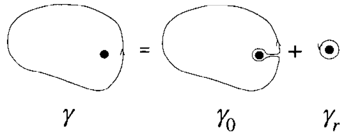

Given a CONTOUR INTEGRAL of the form

$$\int_\gamma \frac{f(z)\,dz}{z-z_0}, \qquad (1)$$

define a path γ_0 as an infinitesimal CIRCLE around the point z_0 (the dot in the above illustration). Define the path γ_r as an arbitrary loop with a cut line (on which the forward and reverse contributions cancel each other out) so as to go around z_0.

The total path is then

$$\gamma = \gamma_0 + \gamma_r, \qquad (2)$$

so

$$\int_\gamma \frac{f(z)\,dz}{z-z_0} = \int_{\gamma_0} \frac{f(z)\,dz}{z-z_0} + \int_{\gamma_r} \frac{f(z)\,dz}{z-z_0}. \qquad (3)$$

From the CAUCHY INTEGRAL THEOREM, the CONTOUR INTEGRAL along any path not enclosing a POLE is 0. Therefore, the first term in the above equation is 0 since γ_0 does not enclose the POLE, and we are left with

$$\int_\gamma \frac{f(z)\,dz}{z-z_0} = \int_{\gamma_r} \frac{f(z)\,dz}{z-z_0}. \qquad (4)$$

Now, let $z \equiv z_0 + re^{i\theta}$, so $dz = ire^{i\theta}\,d\theta$. Then

$$\int_\gamma \frac{f(z)\,dz}{z-z_0} = \int_{\gamma_r} \frac{f(z_0+re^{i\theta})}{re^{i\theta}} ire^{i\theta}\,d\theta = \int_{\gamma_r} f(z_0+re^{i\theta})i\,d\theta. \qquad (5)$$

But we are free to allow the radius r to shrink to 0, so

$$\int_\gamma \frac{f(z)\,dz}{z-z_0} = \lim_{r\to 0}\int_{\gamma_r} f(z_0+re^{i\theta})i\,d\theta = \int_{\gamma_r} f(z_0)i\,d\theta = if(z_0)\int_{\gamma_r} d\theta = 2\pi i f(z_0), \qquad (6)$$

and

$$f(z_0) = \frac{1}{2\pi i}\int_\gamma \frac{f(z)\,dz}{z-z_0}. \qquad (7)$$

If multiple loops are made around the POLE, then equation (7) becomes

$$n(\gamma, z_0)f(z_0) = \frac{1}{2\pi i}\int_\gamma \frac{f(z)\,dz}{z-z_0}, \qquad (8)$$

where $n(\gamma, z_0)$ is the WINDING NUMBER.

A similar formula holds for the derivatives of $f(z)$,

$$\begin{aligned} f'(z_0) &= \lim_{h\to 0}\frac{f(z_0+h)-f(h)}{h} \\ &= \lim_{h\to 0}\frac{1}{2\pi i h}\left(\int_\gamma \frac{f(z)\,dz}{z-z_0-h} - \int_\gamma \frac{f(z)\,dz}{z-z_0}\right) \\ &= \lim_{h\to 0}\frac{1}{2\pi i h}\int_\gamma \frac{f(z)[(z-z_0)-(z-z_0-h)]\,dz}{(z-z_0-h)(z-z_0)} \\ &= \lim_{h\to 0}\frac{1}{2\pi i h}\int_\gamma \frac{hf(z)\,dz}{(z-z_0-h)(z-z_0)} \\ &= \frac{1}{2\pi i}\int_\gamma \frac{f(z)\,dz}{(z-z_0)^2}. \end{aligned} \qquad (9)$$

Iterating again,

$$f''(z_0) = \frac{2}{2\pi i}\int_\gamma \frac{f(z)\,dz}{(z-z_0)^3}. \qquad (10)$$

Continuing the process and adding the WINDING NUMBER n,

$$n(\gamma, z_0)f^{(r)}(z_0) = \frac{r!}{2\pi i}\int_\gamma \frac{f(z)\,dz}{(z-z_0)^{r+1}}. \qquad (11)$$

see also MORERA'S THEOREM

References
Arfken, G. "Cauchy's Integral Formula." §6.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 371–376, 1985.
Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 367–372, 1953.

Cauchy Integral Test

see INTEGRAL TEST

Cauchy Integral Theorem

If f is continuous and finite on a simply connected region R and has only finitely many points of nondifferentiability in R, then

$$\oint_\gamma f(z)\,dz = 0 \tag{1}$$

for any closed CONTOUR γ completely contained in R. Writing z as

$$z \equiv x + iy \tag{2}$$

and $f(z)$ as

$$f(z) \equiv u + iv \tag{3}$$

then gives

$$\begin{aligned}\oint_\gamma f(z)\,dz &= \int_\gamma (u+iv)(dx+i\,dy)\\ &= \int_\gamma u\,dx - v\,dy + i\int_\gamma v\,dx + u\,dy. \qquad (4)\end{aligned}$$

From GREEN'S THEOREM,

$$\int_\gamma f(x,y)\,dx - g(x,y)\,dy = -\iint \left(\frac{\partial g}{\partial x} + \frac{\partial f}{\partial y}\right) dx\,dy, \tag{5}$$

$$\int_\gamma f(x,y)\,dx + g(x,y)\,dy = \iint \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\right) dx\,dy \tag{6}$$

so (4) becomes

$$\begin{aligned}\oint_\gamma f(z)\,dz = &-\iint \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right) dx\,dy\\ &+ i\iint \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) dx\,dy. \qquad (7)\end{aligned}$$

But the CAUCHY-RIEMANN EQUATIONS require that

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \tag{8}$$

$$\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}, \tag{9}$$

so

$$\oint_\gamma f(z)\,dz = 0, \tag{10}$$

Q. E. D.

For a MULTIPLY CONNECTED region,

$$\int_{C_1} f(z)\,dz = \int_{C_2} f(z)\,dz. \tag{11}$$

see also CAUCHY INTEGRAL THEOREM, MORERA'S THEOREM, RESIDUE THEOREM (COMPLEX ANALYSIS)

References

Arfken, G. "Cauchy's Integral Theorem." §6.3 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 365–371, 1985.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 363–367, 1953.

Cauchy-Kovalevskaya Theorem

The theorem which proves the existence and uniqueness of solutions to the CAUCHY PROBLEM.

see also CAUCHY PROBLEM

Cauchy-Lagrange Identity

$$\begin{aligned}&({a_1}^2 + {a_2}^2 + \ldots + {a_n}^2)({b_1}^2 + {b_2}^2 + \ldots + {b_n}^2)\\ &\quad = (a_1b_2 - a_2b_1)^2 + (a_1b_3 - a_3b_1)^2 + \ldots\\ &\qquad + (a_{n-1}b_n - a_nb_{n-1})^2.\end{aligned}$$

From this identity, the n-D CAUCHY INEQUALITY follows.

Cauchy-Maclaurin Theorem

see MACLAURIN-CAUCHY THEOREM

Cauchy Mean Theorem

For numbers > 0, the GEOMETRIC MEAN $<$ the ARITHMETIC MEAN.

Cauchy Principal Value

$$PV\int_{-\infty}^{\infty} f(x)\,dx \equiv \lim_{R\to\infty}\int_{-R}^{R} f(x)\,dx$$

$$PV\int_a^b f(x)\,dx \equiv \lim_{\epsilon\to 0}\left[\int_a^{c-\epsilon} f(x)\,dx + \int_{c+\epsilon}^b f(x)\,dx\right],$$

where $\epsilon > 0$ and $a \leq c \leq b$.

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 401–403, 1985.

Sansone, G. *Orthogonal Functions, rev. English ed.* New York: Dover, p. 158, 1991.

Cauchy Problem

If $f(x,y)$ is an ANALYTIC FUNCTION in a NEIGHBORHOOD of the point (x_0, y_0) (i.e., it can be expanded in a series of NONNEGATIVE INTEGER POWERS of $(x - x_0)$ and $(y - y_0)$), find a solution $y(x)$ of the DIFFERENTIAL EQUATION

$$\frac{dy}{dx} = f(x),$$

with initial conditions $y = y_0$ and $x = x_0$. The existence and uniqueness of the solution were proven by Cauchy and Kovalevskaya in the CAUCHY-KOVALEVSKAYA THEOREM. The Cauchy problem amounts to determining the shape of the boundary and type of equation which yield unique and reasonable solutions for the CAUCHY BOUNDARY CONDITIONS.

see also CAUCHY BOUNDARY CONDITIONS

Cauchy Ratio Test

see RATIO TEST

Cauchy Remainder Form

The remainder of n terms of a TAYLOR SERIES is given by

$$R_n = \frac{(x-c)^{n-1}(x-a)}{(n-1)!} f^{(n)}(c),$$

where $a \leq c \leq x$.

Cauchy-Riemann Equations

Let

$$f(x,y) \equiv u(x,y) + iv(x,y), \tag{1}$$

where

$$z \equiv x + iy, \tag{2}$$

so

$$dz = dx + i\,dy. \tag{3}$$

The total derivative of f with respect to z may then be computed as follows.

$$y = \frac{z-x}{i} \tag{4}$$

$$x = z - iy, \tag{5}$$

so

$$\frac{\partial y}{\partial z} = \frac{1}{i} = -i \tag{6}$$

$$\frac{\partial x}{\partial z} = 1, \tag{7}$$

and

$$\frac{df}{dz} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial z} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial z} = \frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}. \tag{8}$$

In terms of u and v, (8) becomes

$$\begin{aligned} \frac{df}{dz} &= \left(\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}\right) - i\left(\frac{\partial u}{\partial y} + i\frac{\partial v}{\partial y}\right) \\ &= \left(\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}\right) + \left(-i\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}\right). \end{aligned} \tag{9}$$

Along the real, or x-AXIS, $\partial f/\partial y = 0$, so

$$\frac{df}{dz} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}. \tag{10}$$

Along the imaginary, or y-axis, $\partial f/\partial x = 0$, so

$$\frac{df}{dz} = -i\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}. \tag{11}$$

If f is COMPLEX DIFFERENTIABLE, then the value of the derivative must be the same for a given dz, regardless of its orientation. Therefore, (10) must equal (11), which requires that

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \tag{12}$$

and

$$\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}. \tag{13}$$

These are known as the Cauchy-Riemann equations. They lead to the condition

$$\frac{\partial^2 u}{\partial x \partial y} = -\frac{\partial^2 v}{\partial x \partial y}. \tag{14}$$

The Cauchy-Riemann equations may be concisely written as

$$\begin{aligned} \frac{df}{dz^*} &= \frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y} = \left(\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}\right) + i\left(\frac{\partial u}{\partial y} + i\frac{\partial v}{\partial y}\right) \\ &= \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) + i\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) = 0. \end{aligned} \tag{15}$$

In POLAR COORDINATES,

$$f(re^{i\theta}) \equiv R(r,\theta)e^{i\Theta(r,\theta)}, \tag{16}$$

so the Cauchy-Riemann equations become

$$\frac{\partial R}{\partial r} = \frac{R}{r}\frac{\partial \Theta}{\partial \theta} \tag{17}$$

$$\frac{1}{r}\frac{\partial R}{\partial \theta} = -R\frac{\partial \Theta}{\partial r}. \tag{18}$$

If u and v satisfy the Cauchy-Riemann equations, they also satisfy LAPLACE'S EQUATION in 2-D, since

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{\partial}{\partial x}\left(\frac{\partial v}{\partial y}\right) + \frac{\partial}{\partial y}\left(-\frac{\partial v}{\partial x}\right) = 0 \tag{19}$$

$$\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = \frac{\partial}{\partial x}\left(-\frac{\partial u}{\partial y}\right) + \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial x}\right) = 0. \tag{20}$$

By picking an arbitrary $f(z)$, solutions can be found which automatically satisfy the Cauchy-Riemann equations and LAPLACE'S EQUATION. This fact is used to find so-called CONFORMAL SOLUTIONS to physical problems involving scalar potentials such as fluid flow and electrostatics.

see also CAUCHY INTEGRAL THEOREM, CONFORMAL SOLUTION, MONOGENIC FUNCTION, POLYGENIC FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 17, 1972.

Arfken, G. "Cauchy-Riemann Conditions." §6.2 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 3560–365, 1985.

Cauchy's Rigidity Theorem

see RIGIDITY THEOREM

Cauchy Root Test

see ROOT TEST

Cauchy-Schwarz Integral Inequality

Let $f(x)$ and $g(x)$ by any two Real integrable functions of $[a,b]$, then

$$\left[\int_a^b f(x)g(x)\,dx\right]^2 \leq \left[\int_a^b f^2(x)\,dx\right]\left[\int_a^b g^2(x)\,dx\right],$$

with equality Iff $f(x) = kg(x)$ with k real.

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1099, 1993.

Cauchy-Schwarz Sum Inequality

$$|\mathbf{a}\cdot\mathbf{b}| \leq |\mathbf{a}|\,|\mathbf{b}|.$$

$$\left(\sum_{k=1}^{n} a_k b_k\right)^2 \leq \left(\sum_{k=1}^{n} {a_k}^2\right)\left(\sum_{k=1}^{n} {b_k}^2\right).$$

Equality holds Iff the sequences a_1, a_2, ... and b_1, b_2, ... are proportional.

see also Fibonacci Identity

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1092, 1979.

Cauchy Sequence

A Sequence $a_1, a_2, \ldots$ such that the Metric $d(a_m, a_n)$ satisfies

$$\lim_{\min(m,n)\to\infty} d(a_m, a_n) = 0.$$

Cauchy sequences in the rationals do not necessarily Converge, but they do Converge in the Reals.

Real Numbers can be defined using either Dedekind Cuts or Cauchy sequences.

see also Dedekind Cut

Cauchy Test

see Ratio Test

Caustic

The curve which is the Envelope of reflected (Catacaustic) or refracted (Diacaustic) rays of a given curve for a light source at a given point (known as the Radiant Point). The caustic is the Evolute of the Orthotomic.

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, p. 60, 1972.

Lee, X. "Caustics." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Caustics_dir/caustics.html`.

Lockwood, E. H. "Caustic Curves." Ch. 24 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 182–185, 1967.

Yates, R. C. "Caustics." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 15-20, 1952.

Cavalieri's Principle

1. If the lengths of every one-dimensional slice are equal for two regions, then the regions have equal Areas.
2. If the Areas of every two-dimensional slice (Cross-Section) are equal for two Solids, then the Solids have equal Volumes.

see also Cross-Section, Pappus's Centroid Theorem

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 126 and 132, 1987.

Cayley Algebra

The only Nonassociative Division Algebra with Real Scalars. There is an 8-square identity corresponding to this algebra. The elements of a Cayley algebra are called Cayley Numbers or Octonions.

References

Kurosh, A. G. *General Algebra.* New York: Chelsea, pp. 226–28, 1963.

Cayley-Bacharach Theorem

Let $X_1, X_2 \subset \mathbb{P}^2$ be Cubic plane curves meeting in nine points p_1, ..., p_9. If $X \subset \mathbb{P}^2$ is any Cubic containing p_1, ..., p_8, then X contains p_9 as well. It is related to Gorenstein Rings, and is a generalization of Pappus's Hexagon Theorem and Pascal's Theorem.

References

Eisenbud, D.; Green, M.; and Harris, J. "Cayley-Bacharach Theorems and Conjectures." *Bull. Amer. Math. Soc.* **33**, 295–324, 1996.

Cayley Cubic

A Cubic Ruled Surface (Fischer 1986) in which the director line meets the director Conic Section. Cayley's surface is the unique cubic surface having four Ordinary Double Points (Hunt), the maximum possible for Cubic Surface (Endraß). The Cayley cubic is invariant under the Tetrahedral Group and contains exactly nine lines, six of which connect the four nodes pairwise and the other three of which are coplanar (Endraß).

If the Ordinary Double Points in projective 3-space are taken as (1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1), then the equation of the surface in projective coordinates is

$$\frac{1}{x_0} + \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} = 0$$

(Hunt). Defining "affine" coordinates with plane at infinity $v = x_0 + x_1 + x_2 + 2x_3$ and

$$x = \frac{x_0}{v}$$
$$y = \frac{x_1}{v}$$
$$z = \frac{x_2}{v}$$

then gives the equation

$$-5(x^2y+x^2z+y^2x+y^2z+z^2y+z^2x)+2(xy+xz+yz) = 0$$

plotted in the left figure above (Hunt). The slightly different form

$$4(x^3+y^3+z^3+w^3)-(x+y+z+w)^3 = 0$$

is given by Endraß which, when rewritten in TETRAHEDRAL COORDINATES, becomes

$$x^2+y^2-x^2z+y^2z+z^2-1 = 0,$$

plotted in the right figure above.

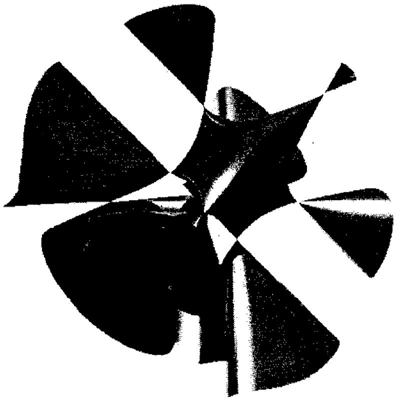

The Hessian of the Cayley cubic is given by

$$\begin{aligned}0 = {}& x_0{}^2(x_1x_2+x_1x_3+x_2x_3)+x_1^2(x_0x_2+x_0x_3+x_2x_3)\\ &+x_2^2(x_0x_1+x_0x_3+x_1x_3)+x_3^2(x_0x_1+x_0x_2+x_1x_2).\end{aligned}$$

in homogeneous coordinates x_0, x_1, x_2, and x_3. Taking the plane at infinity as $v = 5(x_0+x_1+x_2+2x_3)/2$ and setting x, y, and z as above gives the equation

$$\begin{aligned}&25[x^3(y+z)+y^3(x+z)+z^3(x+y)]+50(x^2y^2+x^2z^2+y^2z^2)\\ &-125(x^2yz+y^2xz+z^2xy)+60xyz-4(xy+xz+yz) = 0,\end{aligned}$$

plotted above (Hunt). The Hessian of the Cayley cubic has 14 ORDINARY DOUBLE POINTS, four more than a the general Hessian of a smooth CUBIC SURFACE (Hunt).

References

Endraß, S. "Flächen mit vielen Doppelpunkten." *DMV-Mitteilungen* **4**, 17–20, Apr. 1995.

Endraß, S. "The Cayley Cubic." `http://www.mathematik.uni-mainz.de/AlgebraischeGeometrie/docs/Ecayley.shtml`.

Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, p. 14, 1986.

Fischer, G. (Ed.). Plate 33 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 33, 1986.

Hunt, B. "Algebraic Surfaces." `http://www.mathematik.uni-kl.de/~wwwagag/Galerie.html`.

Hunt, B. *The Geometry of Some Special Arithmetic Quotients.* New York: Springer-Verlag, pp. 115–122, 1996.

Nordstrand, T. "The Cayley Cubic." `http://www.uib.no/people/nfytn/cleytxt.htm`.

Cayley Graph

The representation of a GROUP as a network of directed segments, where the vertices correspond to elements and the segments to multiplication by group generators and their inverses.

see also CAYLEY TREE

References

Grossman, I. and Magnus, W. *Groups and Their Graphs.* New York: Random House, p. 45, 1964.

Cayley's Group Theorem

Every FINITE GROUP of order n can be represented as a PERMUTATION GROUP on n letters, as first proved by Cayley in 1878 (Rotman 1995).

see also FINITE GROUP, PERMUTATION GROUP

References

Rotman, J. J. *An Introduction to the Theory of Groups, 4th ed.* New York: Springer-Verlag, p. 52, 1995.

Cayley-Hamilton Theorem

Given

$$\begin{vmatrix} a_{11}-x & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22}-x & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm}-x \end{vmatrix} = x^m + c_{m-1}x^{m-1} + \ldots + c_0, \quad (1)$$

then

$$\mathsf{A}^m + c_{m-1}\mathsf{A}^{m-1} + \ldots + c_0\mathsf{I} = 0, \quad (2)$$

where I is the IDENTITY MATRIX. Cayley verified this identity for $m = 2$ and 3 and postulated that it was true for all m. For $m = 2$, direct verification gives

$$\begin{aligned}\begin{vmatrix} a-x & b \\ c & d-x \end{vmatrix} &= (a-x)(d-x)-bc \\ &= x^2-(a+d)x+(ad-bc) \equiv x^2+c_1x+c_2 \quad (3)\end{aligned}$$

$$\mathsf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{4}$$

$$\begin{aligned} \mathsf{A}^2 &= \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \\ &= \begin{bmatrix} a^2+bc & ab+bd \\ ac+cd & bc+d^2 \end{bmatrix} \end{aligned} \tag{5}$$

$$-(a+d)\mathsf{A} = \begin{bmatrix} -a^2-ad & -ab-bd \\ -ac-dc & -ad-d^2 \end{bmatrix} \tag{6}$$

$$(ad-bc)\mathsf{I} = \begin{bmatrix} ad-bc & 0 \\ 0 & ad-bc \end{bmatrix}, \tag{7}$$

so

$$\mathsf{A}^2 - (a+d)\mathsf{A} + (ad-bc)\mathsf{I} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \tag{8}$$

The Cayley-Hamilton theorem states that a $n \times n$ MATRIX A is annihilated by its CHARACTERISTIC POLYNOMIAL $\det(x\mathsf{I} - \mathsf{A})$, which is monic of degree n.

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1117, 1979.

Segercrantz, J. "Improving the Cayley-Hamilton Equation for Low-Rank Transformations." *Amer. Math. Monthly* **99**, 42–44, 1992.

Cayley's Hypergeometric Function Theorem

If

$$(1-z)^{a+b-c}\,{}_2F_1(2a,2b;2c;z) = \sum_{n=0}^{\infty} a_n z^n,$$

then

$$\begin{aligned} &{}_2F_1(a,b;c+\tfrac{1}{2};z)\,{}_2F_1(c-a,c-b;c\tfrac{1}{2};z) \\ &\qquad = \sum_{n=0}^{\infty} \frac{(c)_n}{(c+\frac{1}{2})} a_n z^n, \end{aligned}$$

where ${}_2F_1(a,b;c;z)$ is a HYPERGEOMETRIC FUNCTION.

see also HYPERGEOMETRIC FUNCTION

Cayley-Klein Parameters

The parameters α, β, γ, and δ which, like the three EULER ANGLES, provide a way to uniquely characterize the orientation of a solid body. These parameters satisfy the identities

$$\alpha\alpha^* + \gamma\gamma^* = 1 \tag{1}$$

$$\alpha\alpha^* + \beta\beta^* = 1 \tag{2}$$

$$\beta\beta^* + \delta\delta^* = 1 \tag{3}$$

$$\alpha^*\beta + \gamma^*\delta = 0 \tag{4}$$

$$\alpha\delta - \beta\gamma = 1 \tag{5}$$

and

$$\beta = -\gamma^* \tag{6}$$

$$\delta = \alpha^*, \tag{7}$$

where z^* denotes the COMPLEX CONJUGATE. In terms of the EULER ANGLES θ, ϕ, and ψ, the Cayley-Klein parameters are given by

$$\alpha = e^{i(\psi+\phi)/2}\cos(\tfrac{1}{2}\theta) \tag{8}$$

$$\beta = ie^{i(\psi-\phi)/2}\sin(\tfrac{1}{2}\theta) \tag{9}$$

$$\gamma = ie^{-i(\psi-\phi)/2}\sin(\tfrac{1}{2}\theta) \tag{10}$$

$$\delta = e^{-(\psi+\phi)/2}\cos(\tfrac{1}{2}\theta) \tag{11}$$

(Goldstein 1960, p. 155).

The transformation matrix is given in terms of the Cayley-Klein parameters by

$$\mathsf{A} = \begin{bmatrix} \frac{1}{2}(\alpha^2-\gamma^2+\delta^2-\beta^2) & \frac{1}{2}i(\gamma^2-\alpha^2+\delta^2-\beta^2) & \gamma\delta-\alpha\beta \\ \frac{1}{2}i(\alpha^2+\gamma^2-\beta^2-\delta^2) & \frac{1}{2}(\alpha^2+\gamma^2+\beta^2+\delta^2) & -i(\alpha\beta+\gamma\delta) \\ \beta\delta-\alpha\gamma & i(\alpha\gamma+\beta\delta) & \alpha\delta+\beta\gamma \end{bmatrix} \tag{12}$$

(Goldstein 1960, p. 153).

The Cayley-Klein parameters may be viewed as parameters of a matrix (denoted Q for its close relationship with QUATERNIONS)

$$\mathsf{Q} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \tag{13}$$

which characterizes the transformations

$$u' = \alpha u + \beta v \tag{14}$$

$$v' = \gamma u + \delta v. \tag{15}$$

of a linear space having complex axes. This matrix satisfies

$$\mathsf{Q}^\dagger\mathsf{Q} = \mathsf{Q}\mathsf{Q}^\dagger = \mathsf{I}, \tag{16}$$

where I is the IDENTITY MATRIX and $\mathsf{A}^\dagger$ the MATRIX TRANSPOSE, as well as

$$|\mathsf{Q}|^*|\mathsf{Q}| = 1. \tag{17}$$

In terms of the EULER PARAMETERS e_i and the PAULI MATRICES σ_i, the Q-matrix can be written as

$$\mathsf{Q} = e_0\mathsf{I} + i(e_1\sigma_1 + e_2\sigma_2 + e_3\sigma_3) \tag{18}$$

(Goldstein 1980, p. 156).

see also EULER ANGLES, EULER PARAMETERS, PAULI MATRICES, QUATERNION

References

Goldstein, H. "The Cayley-Klein Parameters and Related Quantities." §4-5 in *Classical Mechanics, 2nd ed.* Reading, MA: Addison-Wesley, pp. 148–158, 1980.

Cayley-Klein-Hilbert Metric

The METRIC of Felix Klein's model for HYPERBOLIC GEOMETRY,

$$\begin{aligned} g_{11} &= \frac{a^2(1-x_2{}^2)}{(1-x_1{}^2-x_2{}^2)^2} \\ g_{12} &= \frac{a^2 x_1 x_2}{(1-x_1{}^2-x_2{}^2)^2} \\ g_{22} &= \frac{a^2(1-x_1{}^2)}{(1-x_1{}^2-x_2{}^2)^2}. \end{aligned}$$

see also HYPERBOLIC GEOMETRY

Cayley Number

There are two completely different definitions of Cayley numbers. The first type Cayley numbers is one of the eight elements in a CAYLEY ALGEBRA, also known as an OCTONION. A typical Cayley number is of the form

$$a + bi_0 + ci_1 + di_2 + ei_3 + fi_4 + gi_5 + hi_6,$$

where each of the triples (i_0, i_1, i_3), (i_1, i_2, i_4), (i_2, i_3, i_5), (i_3, i_4, i_6), (i_4, i_5, i_0), (i_5, i_6, i_1), (i_6, i_0, i_2) behaves like the QUATERNIONS (i, j, k). Cayley numbers are *not* ASSOCIATIVE. They have been used in the study of 7- and 8-D space, and a general rotation in 8-D space can be written

$$x' \to ((((((xc_1)c_2)c_3)c_4)c_5)c_6)c_7.$$

The second type of Cayley number is a quantity which describes a DEL PEZZO SURFACE.

see also COMPLEX NUMBER, DEL PEZZO SURFACE, QUATERNION, REAL NUMBER

References

Conway, J. H. and Guy, R. K. "Cayley Numbers." In *The Book of Numbers.* New York: Springer-Verlag, pp. 234–235, 1996.

Okubo, S. *Introduction to Octonion and Other Non-Associative Algebras in Physics.* New York: Cambridge University Press, 1995.

Cayley's Ruled Surface

see CAYLEY CUBIC

Cayley's Sextic

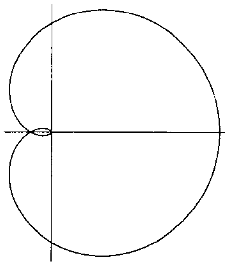

A plane curve discovered by Maclaurin but first studied in detail by Cayley. The name Cayley's sextic is due to R. C. Archibald, who attempted to classify curves in a paper published in Strasbourg in 1900 (MacTutor Archive). Cayley's sextic is given in POLAR COORDINATES by

$$r = a\cos^3(\tfrac{1}{3}\theta), \tag{1}$$

or

$$r = 4b\cos^3(\tfrac{1}{3}\theta), \tag{2}$$

where $b \equiv a/4$. In the latter case, the CARTESIAN equation is

$$4(x^2+y^2-bx)^3 = 27a^2(x^2+y^2)^2. \tag{3}$$

The parametric equations are

$$x(t) = 4a\cos^4(\tfrac{1}{2}t)(2\cos t - 1) \tag{4}$$

$$y(t) = 4a\cos^3(\tfrac{1}{2}t)\sin(\tfrac{3}{2}t). \tag{5}$$

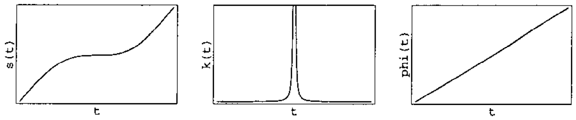

The ARC LENGTH, CURVATURE, and TANGENTIAL ANGLE are

$$s(t) = 3(t + \sin t), \tag{6}$$

$$\kappa(t) = \tfrac{1}{3}\sec^2(\tfrac{1}{2}t), \tag{7}$$

$$\phi(t) = 2t. \tag{8}$$

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 178 and 180, 1972.

MacTutor History of Mathematics Archive. "Cayley's Sextic." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Cayleys.html`.

Cayley's Sextic Evolute

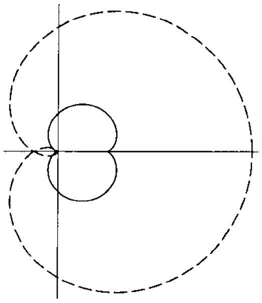

The EVOLUTE of Cayley's sextic is

$$\begin{aligned} x &= \tfrac{1}{8}a + \tfrac{1}{16}a[3\cos(\tfrac{2}{3}t) - \cos(2t)] \\ y &= \tfrac{1}{16}a[3\sin(\tfrac{2}{3}t) - \sin(2t)], \end{aligned}$$

which is a NEPHROID.

Cayley Tree

A TREE in which each NODE has a constant number of branches. The PERCOLATION THRESHOLD for a Cayley tree having z branches is

$$p_c = \frac{1}{z-1}.$$

see also CAYLEY GRAPH

Cayleyian Curve

The ENVELOPE of the lines connecting corresponding points on the JACOBIAN CURVE and STEINERIAN CURVE. The Cayleyian curve of a net of curves of order n has the same GENUS (CURVE) as the JACOBIAN CURVE and STEINERIAN CURVE and, in general, the class $3n(n-1)$.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 150, 1959.

Čech Cohomology

The direct limit of the COHOMOLOGY groups with COEFFICIENTS in an ABELIAN GROUP of certain coverings of a TOPOLOGICAL SPACE.

Ceiling Function

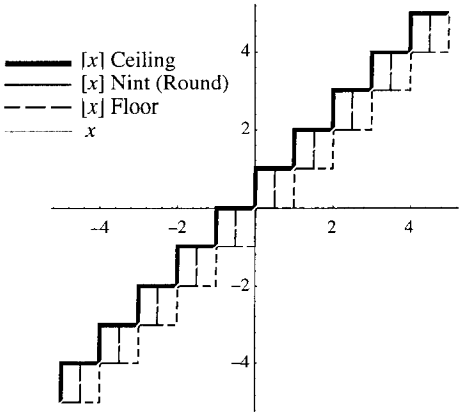

The function $\lceil x \rceil$ which gives the smallest INTEGER $\geq x$, shown as the thick curve in the above plot. Schroeder (1991) calls the ceiling function symbols the "GALLOWS" because of the similarity in appearance to the structure used for hangings. The name and symbol for the ceiling function were coined by K. E. Iverson (Graham *et al.* 1990). It can be implemented as `ceil(x)=-int(-x)`, where `int(x)` is the INTEGER PART of x.

see also FLOOR FUNCTION, INTEGER PART, NINT

References

Graham, R. L.; Knuth, D. E.; and Patashnik, O. "Integer Functions." Ch. 3 in *Concrete Mathematics: A Foundation for Computer Science.* Reading, MA: Addison-Wesley, pp. 67–101, 1990.

Iverson, K. E. *A Programming Language.* New York: Wiley, p. 12, 1962.

Schroeder, M. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise.* New York: W. H. Freeman, p. 57, 1991.

Cell

A finite regular POLYTOPE.

see also 16-CELL, 24-CELL, 120-CELL, 600-CELL

Cellular Automaton

A grid (possibly 1-D) of cells which evolves according to a set of rules based on the states of surrounding cells. von Neumann was one of the first people to consider such a model, and incorporated a cellular model into his "universal constructor." von Neumann proved that an automaton consisting of cells with four orthogonal neighbors and 29 possible states would be capable of simulating a TURING MACHINE for some configuration of about 200,000 cells (Gardner 1983, p. 227).

1-D automata are called "elementary" and are represented by a row of pixels with states either 0 or 1. These can be represented with an 8-bit binary number, as shown by Stephen Wolfram. Wolfram further restricted the number from $2^8 = 256$ to 32 by requiring certain symmetry conditions.

The most well-known cellular automaton is Conway's game of LIFE, popularized in Martin Gardner's *Scientific American* columns. Although the computation of successive LIFE generations was originally done by hand, the computer revolution soon arrived and allowed more extensive patterns to be studied and propagated.

see LIFE, LANGTON'S ANT

References

Adami, C. *Artificial Life.* Cambridge, MA: MIT Press, 1998.

Buchi, J. R. and Siefkes, D. (Eds.). *Finite Automata, Their Algebras and Grammars: Towards a Theory of Formal Expressions.* New York: Springer-Verlag, 1989.

Burks, A. W. (Ed.). *Essays on Cellular Automata.* Urbana-Champaign, IL: University of Illinois Press, 1970.

Cipra, B. "Cellular Automata Offer New Outlook on Life, the Universe, and Everything." In *What's Happening in the Mathematical Sciences, 1995–1996, Vol. 3.* Providence, RI: Amer. Math. Soc., pp. 70–81, 1996.

Dewdney, A. K. *The Armchair Universe: An Exploration of Computer Worlds.* New York: W. H. Freeman, 1988.

Gardner, M. "The Game of Life, Parts I–III." Chs. 20–22 in *Wheels, Life, and Other Mathematical Amusements.* New York: W. H. Freeman, pp. 219 and 222, 1983.

Gutowitz, H. (Ed.). *Cellular Automata: Theory and Experiment.* Cambridge, MA: MIT Press, 1991.

Levy, S. *Artificial Life: A Report from the Frontier Where Computers Meet Biology.* New York: Vintage, 1993.

Martin, O.; Odlyzko, A.; and Wolfram, S. "Algebraic Aspects of Cellular Automata." *Communications in Mathematical Physics* **93**, 219–258, 1984.

McIntosh, H. V. "Cellular Automata." `http://www.cs.cinvestav.mx/mcintosh/cellular.html`.

Preston, K. Jr. and Duff, M. J. B. *Modern Cellular Automata: Theory and Applications.* New York: Plenum, 1985.

Sigmund, K. *Games of Life: Explorations in Ecology, Evolution and Behaviour.* New York: Penguin, 1995.

Sloane, N. J. A. Sequences A006977/M2497 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sloane, N. J. A. and Plouffe, S. Extended entry in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.

Toffoli, T. and Margolus, N. *Cellular Automata Machines: A New Environment for Modeling.* Cambridge, MA: MIT Press, 1987.
Wolfram, S. "Statistical Mechanics of Cellular Automata." *Rev. Mod. Phys.* **55**, 601–644, 1983.
Wolfram, S. (Ed.). *Theory and Application of Cellular Automata.* Reading, MA: Addison-Wesley, 1986.
Wolfram, S. *Cellular Automata and Complexity: Collected Papers.* Reading, MA: Addison-Wesley, 1994.
Wuensche, A. and Lesser, M. *The Global Dynamics of Cellular Automata: An Atlas of Basin of Attraction Fields of One-Dimensional Cellular Automata.* Reading, MA: Addison-Wesley, 1992.

Cellular Space

A HAUSDORFF SPACE which has the structure of a so-called CW-COMPLEX.

Center

A special POINT which usually has some symmetric placement with respect to points on a curve or in a SOLID. The center of a CIRCLE is equidistant from all points on the CIRCLE and is the intersection of any two distinct DIAMETERS. The same holds true for the center of a SPHERE.

see also CENTER (GROUP), CENTER OF MASS, CIRCUMCENTER, CURVATURE CENTER, ELLIPSE, EQUI-BROCARD CENTER, EXCENTER, HOMOTHETIC CENTER, INCENTER, INVERSION CENTER, ISOGONIC CENTERS, MAJOR TRIANGLE CENTER, NINE-POINT CENTER, ORTHOCENTER, PERSPECTIVE CENTER, POINT, RADICAL CENTER, SIMILITUDE CENTER, SPHERE, SPIEKER CENTER, TAYLOR CENTER, TRIANGLE CENTER, TRIANGLE CENTER FUNCTION, YFF CENTER OF CONGRUENCE

Center Function

see TRIANGLE CENTER FUNCTION

Center of Gravity

see CENTER OF MASS

Center (Group)

The center of a GROUP is the set of elements which commute with every member of the GROUP. It is equal to the intersection of the CENTRALIZERS of the GROUP elements.

see also ISOCLINIC GROUPS, NILPOTENT GROUP

Center of Mass

see CENTROID (GEOMETRIC)

Centered Cube Number

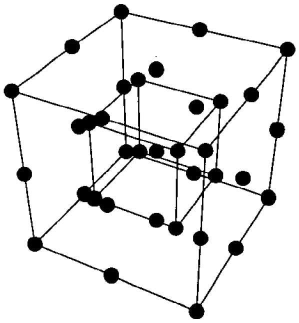

A FIGURATE NUMBER of the form,

$$CCub_n = n^3 + (n-1)^3 = (2n-1)(n^2-n+1).$$

The first few are 1, 9, 35, 91, 189, 341, ... (Sloane's A005898). The GENERATING FUNCTION for the centered cube numbers is

$$\frac{x(x^3+5x^2+5x+1)}{(x-1)^4} = x + 9x^2 + 35x^3 + 91x^4 + \ldots.$$

see also CUBIC NUMBER

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 51, 1996.
Sloane, N. J. A. Sequence A005898/M4616 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Centered Hexagonal Number

see HEX NUMBER

Centered Pentagonal Number

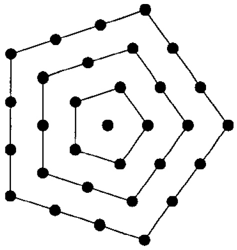

A CENTERED POLYGONAL NUMBER consisting of a central dot with five dots around it, and then additional dots in the gaps between adjacent dots. The general term is $(5n^2 - 5n + 2)/2$, and the first few such numbers are 1, 6, 16, 31, 51, 76, ... (Sloane's A005891). The GENERATING FUNCTION of the centered pentagonal numbers is

$$\frac{x(x^2+3x+1)}{(x-1)^3} = x + 6x^2 + 16x^3 + 31x^4 + \ldots.$$

see also CENTERED SQUARE NUMBER, CENTERED TRIANGULAR NUMBER

References

Sloane, N. J. A. Sequence A005891/M4112 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Centered Polygonal Number

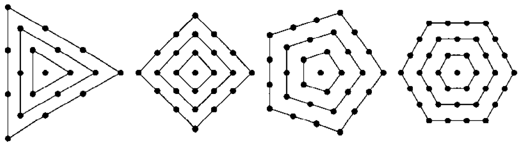

A FIGURATE NUMBER in which layers of POLYGONS are drawn centered about a point instead of with the point at a VERTEX.

see also CENTERED PENTAGONAL NUMBER, CENTERED SQUARE NUMBER, CENTERED TRIANGULAR NUMBER

References

Sloane, N. J. A. and Plouffe, S. Extended entry for sequence M3826 in *The Encyclopedia of Integer Sequences.* San Diego, CA: Academic Press, 1995.

Centered Square Number

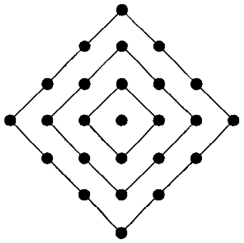

A CENTERED POLYGONAL NUMBER consisting of a central dot with four dots around it, and then additional dots in the gaps between adjacent dots. The general term is $n^2 + (n-1)^2$, and the first few such numbers are 1, 5, 13, 25, 41, ... (Sloane's A001844). Centered square numbers are the sum of two consecutive SQUARE NUMBERS and are congruent to 1 (mod 4). The GENERATING FUNCTION giving the centered square numbers is

$$\frac{x(x+1)^2}{(1-x)^3} = x + 5x^2 + 13x^3 + 25x^4 + \ldots.$$

see also CENTERED PENTAGONAL NUMBER, CENTERED POLYGONAL NUMBER, CENTERED TRIANGULAR NUMBER, SQUARE NUMBER

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 41, 1996.

Sloane, N. J. A. Sequence A001844/M3826 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Centered Triangular Number

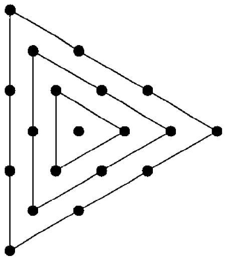

A CENTERED POLYGONAL NUMBER consisting of a central dot with three dots around it, and then additional dots in the gaps between adjacent dots. The general term is $(3n^2 - 3n + 2)/2$, and the first few such numbers are 1, 4, 10, 19, 31, 46, 64, ... (Sloane's A005448). The GENERATING FUNCTION giving the centered triangular numbers is

$$\frac{x(x^2+x+1)}{(1-x)^3} = x + 4x^2 + 10x^3 + 19x^4 + \ldots.$$

see also CENTERED PENTAGONAL NUMBER, CENTERED SQUARE NUMBER

References

Sloane, N. J. A. Sequence A005448/M3378 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Centillion

In the American system, 10^{303}.

see also LARGE NUMBER

Central Angle

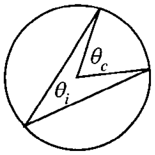

An ANGLE having its VERTEX at a CIRCLE's center which is formed by two points on the CIRCLE'S CIRCUMFERENCE. For angles with the same endpoints,

$$\theta_c = 2\theta_i,$$

where θ_i is the INSCRIBED ANGLE.

References

Pedoe, D. *Circles: A Mathematical View, rev. ed.* Washington, DC: Math. Assoc. Amer., pp. xxi–xxii, 1995.

Central Beta Function

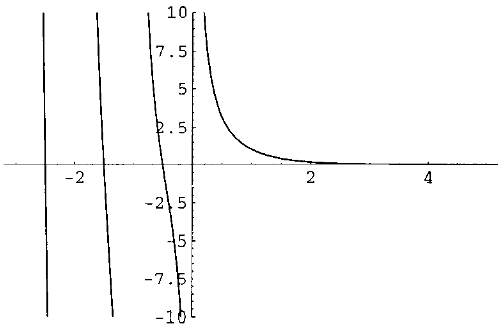

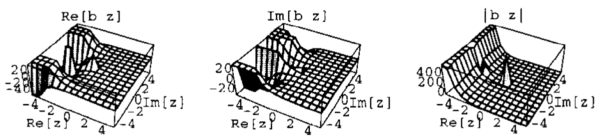

The central beta function is defined by

$$\beta(p) \equiv B(p,p), \tag{1}$$

where $B(p, q)$ is the BETA FUNCTION. It satisfies the identities

$$\beta(p) = 2^{1-2p} B(p, \tfrac{1}{2}) \tag{2}$$

$$= 2^{1-2p} \cos(\pi p) B(\tfrac{1}{2} - p, p) \tag{3}$$

$$= \int_0^1 \frac{t^p\,dt}{(1+t)^{2p}} \tag{4}$$

$$= \frac{2}{p} \prod_{n=1}^{\infty} \frac{n(n+2p)}{(n+p)(n+p)}. \tag{5}$$

With $p = 1/2$, the latter gives the WALLIS FORMULA. When $p = a/b$,

$$b\beta(a/b) = 2^{1-2a/b} J(a, b), \tag{6}$$

where

$$J(a, b) \equiv \int_0^1 \frac{t^{a-1}\,dt}{\sqrt{1-t^b}}. \tag{7}$$

The central beta function satisfies

$$(2 + 4x)\beta(1 + x) = x\beta(x) \tag{8}$$

$$(1 - 2x)\beta(1 - x)\beta(x) = 2\pi \cot(\pi x) \tag{9}$$

$$\beta(\tfrac{1}{2} - x) = 2^{4x-1} \tan(\pi x)\beta(x) \tag{10}$$

$$\beta(x)\beta(x + \tfrac{1}{2}) = 2^{4x+1} \pi \beta(2x)\beta(2x + \tfrac{1}{2}). \tag{11}$$

For p an ODD POSITIVE INTEGER, the central beta function satisfies the identity

$$\beta(px) = \frac{1}{\sqrt{p}} \prod_{k=1}^{(p-1)/2} \frac{2x + \frac{2k-1}{p}}{2\pi} \prod_{k=0}^{p-1} \beta\left(x + \frac{k}{p}\right). \tag{12}$$

see also BETA FUNCTION, REGULARIZED BETA FUNCTION

References

Borwein, J. M. and Zucker, I. J. "Elliptic Integral Evaluation of the Gamma Function at Rational Values of Small Denominators." *IMA J. Numerical Analysis* **12**, 519–526, 1992.

Central Binomial Coefficient

The nth central binomial coefficient is defined as $\binom{n}{\lfloor n/2 \rfloor}$, where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT and $\lfloor n \rfloor$ is the FLOOR FUNCTION. The first few values are 1, 2, 3, 6, 10, 20, 35, 70, 126, 252, ... (Sloane's A001405). The central binomial coefficients have GENERATING FUNCTION

$$\frac{1 - 4x^2 - \sqrt{1 - 4x^2}}{2(2x^3 - x^2)} = 1 + 2x + 3x^2 + 6x^3 + 10x^4 + \ldots.$$

The central binomial coefficients are SQUAREFREE only for $n = 1, 2, 3, 4, 5, 7, 8, 11, 17, 19, 23, 71, \ldots$ (Sloane's A046098), with no others less than 1500.

The above coefficients are a superset of the alternative "central" binomial coefficients

$$\binom{2n}{n} = \frac{(2n)!}{(n!)^2},$$

which have GENERATING FUNCTION

$$\frac{1}{\sqrt{1-4x}} = 1 + 2x + 6x^2 + 20x^3 + 70x^4 + \ldots.$$

The first few values are 2, 6, 20, 70, 252, 924, 3432, 12870, 48620, 184756, ... (Sloane's A000984).

Erdős and Graham (1980, p. 71) conjectured that the central binomial coefficient $\binom{2n}{n}$ is *never* SQUAREFREE for $n > 4$, and this is sometimes known as the ERDŐS SQUAREFREE CONJECTURE. SÁRKÖZY'S THEOREM (Sárközy 1985) provides a partial solution which states that the BINOMIAL COEFFICIENT $\binom{2n}{n}$ is never SQUAREFREE for all sufficiently large $n \geq n_0$ (Vardi 1991). Granville and Ramare (1996) proved that the *only* SQUAREFREE values are $n = 2$ and 4. Sander (1992) subsequently showed that $\binom{2n \pm d}{n}$ are also never SQUAREFREE for sufficiently large n as long as d is not "too big."

see also BINOMIAL COEFFICIENT, CENTRAL TRINOMIAL COEFFICIENT, ERDŐS SQUAREFREE CONJECTURE, SÁRKÖZY'S THEOREM, QUOTA SYSTEM

References

Granville, A. and Ramare, O. "Explicit Bounds on Exponential Sums and the Scarcity of Squarefree Binomial Coefficients." *Mathematika* **43**, 73–107, 1996.

Sander, J. W. "On Prime Divisors of Binomial Coefficients." *Bull. London Math. Soc.* **24**, 140–142, 1992.

Sárközy, A. "On Divisors of Binomial Coefficients. I." *J. Number Th.* **20**, 70–80, 1985.

Sloane, N. J. A. Sequences A046098, A000984/M1645, and A001405/M0769 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Vardi, I. "Application to Binomial Coefficients," "Binomial Coefficients," "A Class of Solutions," "Computing Binomial Coefficients," and "Binomials Modulo and Integer." §2.2, 4.1, 4.2, 4.3, and 4.4 in *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 25–28 and 63–71, 1991.

Central Conic

An ELLIPSE or HYPERBOLA.

see also CONIC SECTION

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 146–150, 1967.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, p. 77, 1990.

Central Difference

The central difference for a function tabulated at equal intervals f_i is defined by

$$\delta(f_{n+1/2}) = \delta_{n+1/2} = \delta^1_{n+1/2} \equiv f_{n+1} - f_n. \qquad (1)$$

Higher order differences may be computed for EVEN and ODD powers,

$$\delta^{2k}_{n+1/2} = \sum_{j=0}^{2k} (-1)^j \binom{2k}{j} f_{n+k-j} \qquad (2)$$

$$\delta^{2k+1}_{n+1/2} = \sum_{j=0}^{2k+1} (-1)^j \binom{2k+1}{j} f_{n+k+1-j}. \qquad (3)$$

see also BACKWARD DIFFERENCE, DIVIDED DIFFERENCE, FORWARD DIFFERENCE

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Differences." §25.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 877–878, 1972.

Central Limit Theorem

Let $x_1, x_2, \ldots, x_N$ be a set of N INDEPENDENT random variates and each x_i have an *arbitrary* probability distribution $P(x_1, \ldots, x_N)$ with MEAN μ_i and a finite VARIANCE ${\sigma_i}^2$. Then the normal form variate

$$X_{\text{norm}} \equiv \frac{\sum_{i=1}^N x_i - \sum_{i=1}^N \mu_i}{\sqrt{\sum_{i=1}^N {\sigma_i}^2}} \qquad (1)$$

has a limiting distribution which is NORMAL (GAUSSIAN) with MEAN $\mu = 0$ and VARIANCE $\sigma^2 = 1$. If conversion to normal form is not performed, then the variate

$$X \equiv \frac{1}{N} \sum_{i=1}^N x_i \qquad (2)$$

is NORMALLY DISTRIBUTED with $\mu_X = \mu_x$ and $\sigma_X = \sigma_x/\sqrt{N}$. To prove this, consider the INVERSE FOURIER TRANSFORM of $P_X(f)$.

$$\begin{aligned}
\mathcal{F}^{-1}[P_X(f)] &\equiv \int_{-\infty}^{\infty} e^{2\pi i f X} p(X)\,dX \\
&= \int_{-\infty}^{\infty} \sum_{n=0}^{\infty} \frac{(2\pi i f X)^n}{n!} p(X)\,dX \\
&= \sum_{n=0}^{\infty} \frac{(2\pi i f)^n}{n!} \int_{-\infty}^{\infty} X^n p(X)\,dx \\
&= \sum_{n=0}^{\infty} \frac{(2\pi i f)^n}{n!} \langle x \rangle^n. \qquad (3)
\end{aligned}$$

Now write

$$\begin{aligned}
\langle X^n \rangle &= \left\langle N^{-n}(x_1 + x_2 + \ldots + x_N)^n \right\rangle \\
&= \int_{-\infty}^{\infty} N^{-n}(x_1 + \ldots + x_N)^n p(x_1)\cdots p(x_N)\,dx_1 \cdots dx_N, \qquad (4)
\end{aligned}$$

so we have

$$\begin{aligned}
\mathcal{F}^{-1}[P_X(f)] &= \sum_{n=0}^{\infty} \frac{(2\pi i f)^n}{n!} \langle X^n \rangle \\
&= \sum_{n=0}^{\infty} \frac{(2\pi i f)^n}{n!} \int_{-\infty}^{\infty} N^{-n}(x_1 + \ldots + x_N)^n \\
&\qquad \times p(x_1)\cdots p(x_N)\,dx_1 \cdots dx_N \\
&= \int_{-\infty}^{\infty} \sum_{n=0}^{\infty} \left[\frac{2\pi i f(x_1 + \ldots + x_N)}{N}\right]^n \frac{1}{n!} \\
&\qquad \times p(x_1)\cdots p(x_N)\,dx_1 \cdots dx_N \\
&= \int_{-\infty}^{\infty} e^{2\pi i f(x_1 + \ldots + x_N)/N} p(x_1)\cdots p(x_N)\,dx_1 \cdots dx_N \\
&= \left[\int_{-\infty}^{\infty} e^{2\pi i f x_1/N} p(x_1)\,dx_1\right] \\
&\qquad \times \cdots \times \left[\int_{-\infty}^{\infty} e^{2\pi i f x_N/N} p(x_N)\,dx_N\right] \\
&= \left[\int_{-\infty}^{\infty} e^{2\pi i f x/N} p(x)\,dx\right]^N \\
&= \left\{\int_{-\infty}^{\infty} \left[1 + \left(\frac{2\pi i f}{N}\right) x + \frac{1}{2}\left(\frac{2\pi i f}{N}\right)^2 x^2 + \ldots\right] p(x)\,dx\right\}^N \\
&= \left[\int_{-\infty}^{\infty} p(x)\,dx + \frac{2\pi i f}{N} \int_{-\infty}^{\infty} x p(x)\,dx \right. \\
&\qquad \left. - \frac{(2\pi f)^2}{2N^2} \int_{-\infty}^{\infty} x^2 p(x)\,dx + \mathcal{O}(N^{-3})\right]^N \\
&= \left[1 + \frac{2\pi i f}{N} \langle x \rangle - \frac{(2\pi f)^2}{2N^2} \left\langle x^2 \right\rangle + \mathcal{O}(N^{-3})\right]^N \\
&= \exp\left\{N \ln\left[1 + \frac{2\pi i f}{N} \langle x \rangle - \frac{(2\pi f)^2}{2N^2} \left\langle x^2 \right\rangle + \mathcal{O}(N^{-3})\right]\right\}. \qquad (5)
\end{aligned}$$

Now expand

$$\ln(1 + x) = x - \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 + \ldots, \qquad (6)$$

so

$$\begin{aligned}
&\mathcal{F}^{-1}[P_X(f)] \\
&\quad \approx \exp\left\{N\left[\frac{2\pi i f}{N} \langle x \rangle - \frac{(2\pi f)^2}{2N^2} \left\langle x^2 \right\rangle \right.\right. \\
&\qquad \left.\left. + \frac{1}{2} \frac{(2\pi i f)^2}{N^2} \langle x \rangle^2 + \mathcal{O}(N^{-3})\right]\right\} \\
&\quad = \exp\left[2\pi i f \langle x \rangle - \frac{(2\pi f)^2(\left\langle x^2 \right\rangle - \langle x \rangle^2)}{2N} + \mathcal{O}(N^{-2})\right] \\
&\quad \approx \exp\left[2\pi i f \mu_x - \frac{(2\pi f)^2 {\sigma_x}^2}{2N}\right], \qquad (7)
\end{aligned}$$

since

$$\mu_x \equiv \langle x \rangle \tag{8}$$

$$\sigma_x{}^2 \equiv \left\langle x^2 \right\rangle - \langle x \rangle^2 . \tag{9}$$

Taking the FOURIER TRANSFORM,

$$\begin{aligned} P_X &\equiv \int_{-\infty}^{\infty} e^{-2\pi i f x} \mathcal{F}^{-1}[P_X(f)]\, df \\ &= \int_{-\infty}^{\infty} e^{2\pi i f(\mu_x - x) - (2\pi f)^2 \sigma_x{}^2/2N}\, df. \end{aligned} \tag{10}$$

This is of the form

$$\int_{-\infty}^{\infty} e^{iaf - bf^2}\, df, \tag{11}$$

where $a \equiv 2\pi(\mu_x - x)$ and $b \equiv (2\pi\sigma_x)^2/2N$. But, from Abramowitz and Stegun (1972, p. 302, equation 7.4.6),

$$\int_{-\infty}^{\infty} e^{iaf - bf^2}\, df = e^{-a^2/4b} \sqrt{\frac{\pi}{b}}. \tag{12}$$

Therefore,

$$\begin{aligned} P_X &= \sqrt{\frac{\pi}{\frac{(2\pi\sigma_x)^2}{2N}}} \exp\left\{ \frac{-[2\pi(\mu_x - x)]^2}{4\frac{(2\pi\sigma_x)^2}{2N}} \right\} \\ &= \sqrt{\frac{2\pi N}{4\pi^2 \sigma_x{}^2}} \exp\left[-\frac{4\pi^2(\mu_x - x)^2 2N}{4 \cdot 4\pi^2 \sigma_x{}^2} \right] \\ &= \frac{\sqrt{N}}{\sigma_x \sqrt{2\pi}} e^{-(\mu_x - x)^2 N/2\sigma_x{}^2}. \end{aligned} \tag{13}$$

But $\sigma_X = \sigma_x/\sqrt{N}$ and $\mu_X = \mu_x$, so

$$P_X = \frac{1}{\sigma_X \sqrt{2\pi}} e^{-(\mu_X - x)^2/2\sigma_X{}^2}. \tag{14}$$

The "fuzzy" central limit theorem says that data which are influenced by many small and unrelated random effects are approximately NORMALLY DISTRIBUTED.

see also LINDEBERG CONDITION, LINDEBERG-FELLER CENTRAL LIMIT THEOREM, LYAPUNOV CONDITION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, 1972.

Spiegel, M. R. *Theory and Problems of Probability and Statistics.* New York: McGraw-Hill, pp. 112–113, 1992.

Zabell, S. L. "Alan Turing and the Central Limit Theorem." *Amer. Math. Monthly* **102**, 483–494, 1995.

Central Trinomial Coefficient

The nth central binomial coefficient is defined as the coefficient of x^n in the expansion of $(1+x+x^2)^n$. The first few are 1, 3, 7, 19, 51, 141, 393, ... (Sloane's A002426). This sequence cannot be expressed as a fixed number of hypergeometric terms (Petkovšek *et al.* 1996, p. 160). The GENERATING FUNCTION is given by

$$f(x) = \frac{1}{\sqrt{(1+x)(1-3x)}} = 1 + x + 3x^2 + 7x^3 + \ldots.$$

see also CENTRAL BINOMIAL COEFFICIENT

References

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, 1996.

Sloane, N. J. A. Sequence A002426/M2673 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Centralizer

The centralizer of a FINITE non-ABELIAN SIMPLE GROUP G is an element z of order 2 such that

$$C_G(z) = \{g \in G : gz = zg\}.$$

see also CENTER (GROUP), NORMALIZER

Centrode

$$\mathbf{C} \equiv \tau \mathbf{T} + \kappa \mathbf{B},$$

where τ is the TORSION, κ is the CURVATURE, $\mathbf{T}$ is the TANGENT VECTOR, and $\mathbf{B}$ is the BINORMAL VECTOR.

Centroid (Function)

By analogy with the GEOMETRIC CENTROID, the centroid of an arbitrary function $f(x)$ is defined as

$$\langle x \rangle = \frac{\int_{-\infty}^{\infty} x f(x)\, dx}{\int_{-\infty}^{\infty} f(x)\, dx}.$$

References

Bracewell, R. *The Fourier Transform and Its Applications.* New York: McGraw-Hill, pp. 139–140 and 156, 1965.

Centroid (Geometric)

The CENTER OF MASS of a 2-D planar LAMINA or a 3-D solid. The mass of a LAMINA with surface density function $\sigma(x, y)$ is

$$M = \iint \sigma(x, y)\, dA. \tag{1}$$

The coordinates of the centroid (also called the CENTER OF GRAVITY) are

$$\bar{x} = \frac{\iint x\sigma(x, y)\, dA}{M} \tag{2}$$

$$\bar{y} = \frac{\iint y\sigma(x,y)\,dA}{M}. \tag{3}$$

The centroids of several common laminas along the non-symmetrical axis are summarized in the following table.

Figure	$\bar{y}$
parabolic segment	$\frac{3}{5}h$
semicircle	$\frac{4r}{3\pi}$

In 3-D, the mass of a solid with density function $\rho(x,y,z)$ is

$$M = \iiint \rho(x,y,z)\,dV, \tag{4}$$

and the coordinates of the center of mass are

$$\bar{x} = \frac{\iiint x\rho(x,y,z)\,dV}{M} \tag{5}$$

$$\bar{y} = \frac{\iiint y\rho(x,y,z)\,dV}{M} \tag{6}$$

$$\bar{z} = \frac{\iiint z\rho(x,y,z)\,dV}{M}. \tag{7}$$

Figure	$\bar{z}$
cone	$\frac{1}{4}h$
conical frustum	$\frac{h({R_1}^2+2R_1R_2+3{R_2}^2)}{4({R_1}^2+R_1R_2+{R_2}^2)}$
hemisphere	$\frac{3}{8}R$
paraboloid	$\frac{2}{3}h$
pyramid	$\frac{1}{4}h$

see also PAPPUS'S CENTROID THEOREM

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 132, 1987.

McLean, W. G. and Nelson, E. W. "First Moments and Centroids." Ch. 9 in *Schaum's Outline of Theory and Problems of Engineering Mechanics: Statics and Dynamics, 4th ed.* New York: McGraw-Hill, pp. 134–162, 1988.

Centroid (Orthocentric System)

The centroid of the four points constituting an ORTHOCENTRIC SYSTEM is the center of the common NINE-POINT CIRCLE (Johnson 1929, p. 249). This fact automatically guarantees that the centroid of the INCENTER and EXCENTERS of a TRIANGLE is located at the CIRCUMCENTER.

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, 1929.

Centroid (Triangle)

The centroid (CENTER OF MASS) of the VERTICES of a TRIANGLE is the point M (or G) of intersection of the TRIANGLE'S three MEDIANS, also called the MEDIAN POINT (Johnson 1929, p. 249). The centroid is always in the interior of the TRIANGLE, and has TRILINEAR COORDINATES

$$\frac{1}{a} : \frac{1}{b} : \frac{1}{c}, \tag{1}$$

or

$$\csc A : \csc B : \csc C. \tag{2}$$

If the sides of a TRIANGLE are divided so that

$$\frac{\overline{A_2P_1}}{\overline{P_1A_3}} = \frac{\overline{A_3P_2}}{\overline{P_2A_1}} = \frac{\overline{A_1P_2}}{\overline{P_3A_2}} = \frac{p}{q}, \tag{3}$$

the centroid of the TRIANGLE $\Delta P_1P_2P_3$ is M (Johnson 1929, p. 250).

Pick an interior point X. The TRIANGLES BXC, CXA, and AXB have equal areas IFF X corresponds to the centroid. The centroid is located one third of the way from each VERTEX to the MIDPOINT of the opposite side. Each median divides the triangle into two equal areas; all the medians together divide it into six equal parts, and the lines from the MEDIAN POINT to the VERTICES divide the whole into three equivalent TRIANGLES. In general, for any line in the plane of a TRIANGLE ABC,

$$d = \tfrac{1}{3}(d_A + d_B + d_C), \tag{4}$$

where d, d_A, d_B, and d_C are the distances from the centroid and VERTICES to the line. A TRIANGLE will balance at the centroid, and along any line passing through the centroid. The TRILINEAR POLAR of the centroid is called the LEMOINE AXIS. The PERPENDICULARS from the centroid are proportional to ${s_i}^{-1}$,

$$a_1p_2 = a_2p_2 = a_3p_3 = \tfrac{2}{3}\Delta, \tag{5}$$

where Δ is the AREA of the TRIANGLE. Let P be an arbitrary point, the VERTICES be A_1, A_2, and A_3, and the centroid M. Then

$$\overline{PA_1}^2+\overline{PA_2}^2+\overline{PA_3}^2 = \overline{MA_1}^2+\overline{MA_2}^2+\overline{MA_3}^2+3\overline{PM}^2. \tag{6}$$

If O is the CIRCUMCENTER of the triangle's centroid, then

$$\overline{OM}^2 = R^2 - \tfrac{1}{9}(a^2+b^2+c^2). \tag{7}$$

The centroid lies on the EULER LINE.

The centroid of the PERIMETER of a TRIANGLE is the triangle's SPIEKER CENTER (Johnson 1929, p. 249).

see also CIRCUMCENTER, EULER LINE, EXMEDIAN POINT, INCENTER, ORTHOCENTER

References

Carr, G. S. *Formulas and Theorems in Pure Mathematics, 2nd ed.* New York: Chelsea, p. 622, 1970.

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., p. 7, 1967.
Dixon, R. *Mathographics.* New York: Dover, pp. 55–57, 1991.
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 173–176 and 249, 1929.
Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.
Kimberling, C. "Centroid." `http://www.evansville.edu/~ck6/tcenters/class/centroid.html`.

Certificate of Compositeness

see COMPOSITENESS CERTIFICATE

Certificate of Primality

see PRIMALITY CERTIFICATE

Cesàro Equation

An INTRINSIC EQUATION which expresses a curve in terms of its ARC LENGTH s and RADIUS OF CURVATURE R (or equivalently, the CURVATURE κ).

see also ARC LENGTH, INTRINSIC EQUATION, NATURAL EQUATION, RADIUS OF CURVATURE, WHEWELL EQUATION

References

Yates, R. C. "Intrinsic Equations." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 123–126, 1952.

Cesàro Fractal

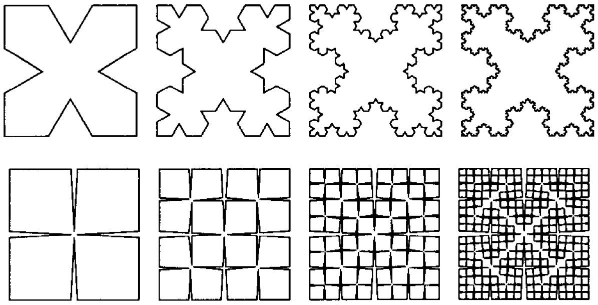

A FRACTAL also known as the TORN SQUARE FRACTAL. The base curves and motifs for the two fractals illustrated above are show below.

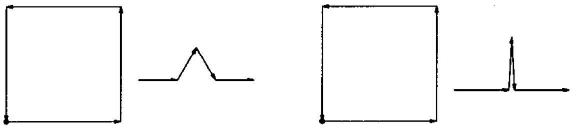

see also FRACTAL, KOCH SNOWFLAKE

References

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, p. 43, 1991.
Pappas, T. *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 79, 1989.
Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m`.

Cesàro Mean

see FEJES TÓTH'S INTEGRAL

Ceva's Theorem

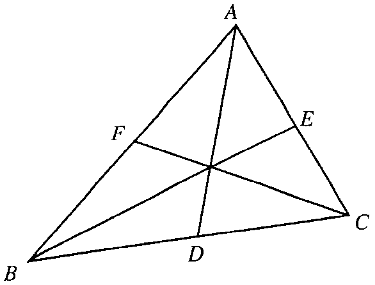

Given a TRIANGLE with VERTICES A, B, and C and points along the sides D, E, and F, a NECESSARY and SUFFICIENT condition for the CEVIANS AD, BE, and CF to be CONCURRENT (intersect in a single point) is that

$$BD \cdot CE \cdot AF = DC \cdot EA \cdot FB. \qquad (1)$$

Let $P = [V_1, \ldots, V_n]$ be an arbitrary n-gon, C a given point, and k a POSITIVE INTEGER such that $1 \leq k \leq n/2$. For $i = 1, \ldots, n$, let W_i be the intersection of the lines CV_i and $V_{i-k}V_{i+k}$, then

$$\prod_{i=1}^{n}\left[\frac{V_{i-k}W_i}{W_iV_{i+k}}\right] = 1. \qquad (2)$$

Here, $AB||CD$ and

$$\left[\frac{AB}{CD}\right] \qquad (3)$$

is the RATIO of the lengths $[A, B]$ and $[C, D]$ with a plus or minus sign depending on whether these segments have the same or opposite directions (Grünbaum and Shepard 1995).

Another form of the theorem is that three CONCURRENT lines from the VERTICES of a TRIANGLE divide the opposite sides in such fashion that the product of three nonadjacent segments equals the product of the other three (Johnson 1929, p. 147).

see also HOEHN'S THEOREM, MENELAUS' THEOREM

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 122, 1987.
Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 4–5, 1967.
Grünbaum, B. and Shepard, G. C. "Ceva, Menelaus, and the Area Principle." *Math. Mag.* **68**, 254–268, 1995.
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 145–151, 1929.
Pedoe, D. *Circles: A Mathematical View, rev. ed.* Washington, DC: Math. Assoc. Amer., p. xx, 1995.

Cevian

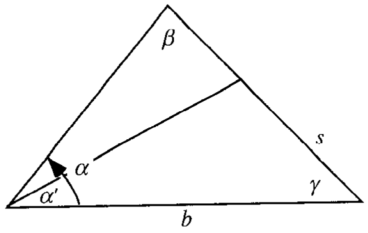

A line segment which joins a VERTEX of a TRIANGLE with a point on the opposite side (or its extension). In the above figure,

$$s = \frac{b \sin \alpha'}{\sin(\gamma + \alpha')}.$$

References

Thébault, V. "On the Cevians of a Triangle." *Amer. Math. Monthly* **60**, 167–173, 1953.

Cevian Conjugate Point

see ISOTOMIC CONJUGATE POINT

Cevian Transform

Vandeghen's (1965) name for the transformation taking points to their ISOTOMIC CONJUGATE POINTS.

see also ISOTOMIC CONJUGATE POINT

References

Vandeghen, A. "Some Remarks on the Isogonal and Cevian Transforms. Alignments of Remarkable Points of a Triangle." *Amer. Math. Monthly* **72**, 1091–1094, 1965.

Cevian Triangle

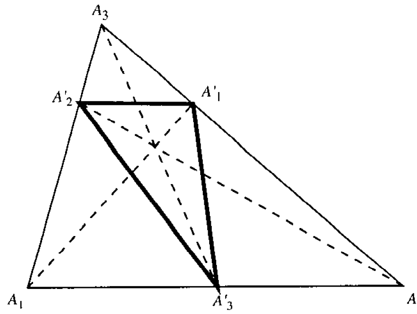

Given a center $\alpha : \beta : \gamma$, the cevian triangle is defined as that with VERTICES $0 : \beta : \gamma$, $\alpha : 0 : \gamma$, and $\alpha : \beta : 0$. If $A'B'C'$ is the CEVIAN TRIANGLE of X and $A''B''C''$ is the ANTICEVIAN TRIANGLE, then X and A'' are HARMONIC CONJUGATE POINTS with respect to A and A'.

see also ANTICEVIAN TRIANGLE

Chain

Let P be a finite PARTIALLY ORDERED SET. A chain in P is a set of pairwise comparable elements (i.e., a TOTALLY ORDERED subset). The WIDTH of P is the maximum CARDINALITY of an ANTICHAIN in P. For a PARTIAL ORDER, the size of the longest CHAIN is called the WIDTH.

see also ADDITION CHAIN, ANTICHAIN, BRAUER CHAIN, CHAIN (GRAPH), DILWORTH'S LEMMA, HANSEN CHAIN

Chain Fraction

see CONTINUED FRACTION

Chain (Graph)

A chain of a GRAPH is a SEQUENCE $\{x_1, x_2, \ldots, x_n\}$ such that (x_1, x_2), (x_2, x_3), ..., (x_{n-1}, x_n) are EDGES of the GRAPH.

Chain Rule

If $g(x)$ is DIFFERENTIABLE at the point x and $f(x)$ is DIFFERENTIABLE at the point $g(x)$, then $f \circ g$ is DIFFERENTIABLE at x. Furthermore, let $y = f(g(x))$ and $u = g(x)$, then

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}. \tag{1}$$

There are a number of related results which also go under the name of "chain rules." For example, if $z = f(x,y)$, $x = g(t)$, and $y = h(t)$, then

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}. \tag{2}$$

The "general" chain rule applies to two sets of functions

$$\begin{aligned} y_1 &= f_1(u_1, \ldots, u_p) \\ &\vdots (3) \\ y_m &= f_m(u_1, \ldots, u_p) \end{aligned}$$

and

$$\begin{aligned} u_1 &= g_1(x_1, \ldots, x_n) \\ &\vdots (4) \\ u_p &= g_p(x_1, \ldots, x_n). \end{aligned}$$

Defining the $m \times n$ JACOBI MATRIX by

$$\left(\frac{\partial y_i}{\partial x_j}\right) = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}, \tag{5}$$

and similarly for $(\partial y_i / \partial u_j)$ and $(\partial u_i / \partial x_j)$ then gives

$$\left(\frac{\partial y_i}{\partial x_j}\right) = \left(\frac{\partial y_i}{\partial u_j}\right)\left(\frac{\partial u_i}{\partial x_j}\right). \tag{6}$$

In differential form, this becomes

$$dy_1 = \left(\frac{\partial y_1}{\partial u_1}\frac{\partial u_1}{\partial x_1} + \ldots + \frac{\partial y_1}{\partial u_p}\frac{\partial u_p}{\partial x_1}\right) dx_1 + \left(\frac{\partial y_1}{\partial u_1}\frac{\partial u_1}{\partial x_2} + \ldots + \frac{\partial y_1}{\partial u_p}\frac{\partial u_p}{\partial x_2}\right) dx_2 + \ldots \quad (7)$$

(Kaplan 1984).

see also DERIVATIVE, JACOBIAN, POWER RULE, PRODUCT RULE

References

Anton, H. *Calculus with Analytic Geometry, 2nd ed.* New York: Wiley, p. 165, 1984.

Kaplan, W. "Derivatives and Differentials of Composite Functions" and "The General Chain Rule." §2.8 and 2.9 in *Advanced Calculus, 3rd ed.* Reading, MA: Addison-Wesley, pp. 101–105 and 106–110, 1984.

Chained Arrow Notation

A NOTATION which generalizes ARROW NOTATION and is defined as

$$\underbrace{a\uparrow\cdots\uparrow b}_{c} \equiv a \to b \to c.$$

see also ARROW NOTATION

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 61, 1996.

Chainette

see CATENARY

Chair

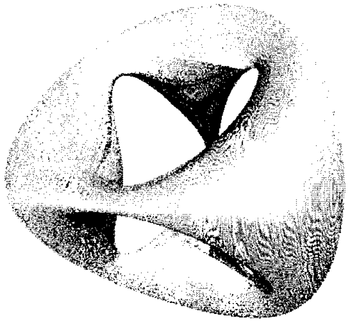

A SURFACE with tetrahedral symmetry which, according to Nordstrand, looks like an inflatable chair from the 1970s. It is given by the implicit equation

$$(x^2+y^2+z^2-ak^2)^2-b[(z-k)^2-2x^2][(z+k)^2-2y^2]=0.$$

see also BRIDE'S CHAIR

References

Nordstrand, T. "Chair." `http://www.uib.no/people/nfytn/chairtxt.htm`.

Chaitin's Constant

An IRRATIONAL NUMBER Ω which gives the probability that for any set of instructions, a UNIVERSAL TURING MACHINE will halt. The digits in Ω are random and cannot be computed ahead of time.

see also HALTING PROBLEM, TURING MACHINE, UNIVERSAL TURING MACHINE

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/chaitin/chaitin.html`.

Gardner, M. "The Random Number Ω Bids Fair to Hold the Mysteries of the Universe." *Sci. Amer.* **241**, 20–34, Nov. 1979.

Gardner, M. "Chaitin's Omega." Ch. 21 in *Fractal Music, Hypercards, and More Mathematical Recreations from Scientific American Magazine.* New York: W. H. Freeman, 1992.

Kobayashi, K. "Sigma(N)O-Complete Properties of Programs and Lartin-Lof Randomness." *Information Proc. Let.* **46**, 37–42, 1993.

Chaitin's Number

see CHAITIN'S CONSTANT

Chaitin's Omega

see CHAITIN'S CONSTANT

Champernowne Constant

Champernowne's number 0.1234567891011... (Sloane's A033307) is the decimal obtained by concatenating the POSITIVE INTEGERS. It is NORMAL in base 10. In 1961, Mahler showed it to also be TRANSCENDENTAL.

The CONTINUED FRACTION of the Champernowne constant is [0, 8, 9, 1, 149083, 1, 1, 1, 4, 1, 1, 1, 3, 4, 1, 1, 1, 15,

$$4575401113910310764836466282429561185996039 39\cdots$$
$$7104575550006620043930902626592563149379532 07\cdots$$
$$7471286563138641209375503552094607183089984 57\cdots$$
$$58014698631488335921417830 10987,$$

6, 1, 1, 21, 1, 9, 1, 1, 2, 3, 1, 7, 2, 1, 83, 1, 156, 4, 58, 8, 54, ...] (Sloane's A030167). The next term of the CONTINUED FRACTION is huge, having 2504 digits. In fact, the coefficients eventually become unbounded, making the continued fraction difficult to calculate for too many more terms. Large terms greater than 10^5 occur at positions 5, 19, 41, 102, 163, 247, 358, 460, ... and have 6, 166, 2504, 140, 33102, 109, 2468, 136, ... digits (Plouffe). Interestingly, the COPELAND-ERDŐS CONSTANT, which is the decimal obtained by concatenating the PRIMES, has a well-behaved CONTINUED FRACTION which does not show the "large term" phenomenon.

see also COPELAND-ERDŐS CONSTANT, SMARANDACHE SEQUENCES

References
Champernowne, D. G. "The Construction of Decimals Normal in the Scale of Ten." *J. London Math. Soc.* **8**, 1933.
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/cntfrc/cntfrc.html`.
Sloane, N. J. A. Sequences A030167 and A033307 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Change of Variables Theorem

A theorem which effectively describes how lengths, areas, volumes, and generalized n-dimensional volumes (CONTENTS) are distorted by DIFFERENTIABLE FUNCTIONS. In particular, the change of variables theorem reduces the whole problem of figuring out the distortion of the content to understanding the infinitesimal distortion, i.e., the distortion of the DERIVATIVE (a linear MAP), which is given by the linear MAP's DETERMINANT. So $f : \mathbb{R}^n \to \mathbb{R}^n$ is an AREA-PRESERVING linear MAP IFF $|\det(f)| = 1$, and in more generality, if S is any subset of $\mathbb{R}^n$, the CONTENT of its image is given by $|\det(f)|$ times the CONTENT of the original. The change of variables theorem takes this infinitesimal knowledge, and applies CALCULUS by breaking up the DOMAIN into small pieces and adds up the change in AREA, bit by bit.

The change of variable formula persists to the generality of DIFFERENTIAL FORMS on MANIFOLDS, giving the formula

$$\int_M (f^*\omega) = \int_W (\omega)$$

under the conditions that M and W are compact connected oriented MANIFOLDS with nonempty boundaries, $f : M \to W$ is a smooth map which is an orientation-preserving DIFFEOMORPHISM of the boundaries.

In 2-D, the explicit statement of the theorem is

$$\int_R f(x,y)\,dx\,dy = \int_{R^*} f[x(u,v),y(u,v)] \left|\frac{\partial(x,y)}{\partial(u,v)}\right| du\,dv$$

and in 3-D, it is

$$\int_R f(x,y,z)\,dx\,dy\,dz = \int_{R^*} f[x(u,v,w),y(u,v,w),z(u,v,w)] \left|\frac{\partial(x,y,z)}{\partial(u,v,w)}\right| du\,dv\,dw,$$

where $R = f(R^*)$ is the image of the original region R^*,

$$\left|\frac{\partial(x,y,z)}{\partial(u,v,w)}\right|$$

is the JACOBIAN, and f is a global orientation-preserving DIFFEOMORPHISM of R and R^* (which are open subsets of $\mathbb{R}^n$).

The change of variables theorem is a simple consequence of the CURL THEOREM and a little DE RHAM COHOMOLOGY. The generalization to n-D requires no additional assumptions other than the regularity conditions on the boundary.

see also IMPLICIT FUNCTION THEOREM, JACOBIAN

References
Kaplan, W. "Change of Variables in Integrals." §4.6 in *Advanced Calculus, 3rd ed.* Reading, MA: Addison-Wesley, pp. 238–245, 1984.

Chaos

A DYNAMICAL SYSTEM is chaotic if it

1. Has a DENSE collection of points with periodic orbits,
2. Is sensitive to the initial condition of the system (so that initially nearby points can evolve quickly into very different states), and
3. Is TOPOLOGICALLY TRANSITIVE.

Chaotic systems exhibit irregular, unpredictable behavior (the BUTTERFLY EFFECT). The boundary between linear and chaotic behavior is characterized by PERIOD DOUBLING, following by quadrupling, etc.

An example of a simple physical system which displays chaotic behavior is the motion of a magnetic pendulum over a plane containing two or more attractive magnets. The magnet over which the pendulum ultimately comes to rest (due to frictional damping) is highly dependent on the starting position and velocity of the pendulum (Dickau). Another such system is a double pendulum (a pendulum with another pendulum attached to its end).

see also ACCUMULATION POINT, ATTRACTOR, BASIN OF ATTRACTION, BUTTERFLY EFFECT, CHAOS GAME, FEIGENBAUM CONSTANT, FRACTAL DIMENSION, GINGERBREADMAN MAP, HÉNON-HEILES EQUATION, HÉNON MAP, LIMIT CYCLE, LOGISTIC EQUATION, LYAPUNOV CHARACTERISTIC EXPONENT, PERIOD THREE THEOREM, PHASE SPACE, QUANTUM CHAOS, RESONANCE OVERLAP METHOD, ŠARKOVSKII'S THEOREM, SHADOWING THEOREM, SINK (MAP), STRANGE ATTRACTOR

References
Bai-Lin, H. *Chaos.* Singapore: World Scientific, 1984.
Baker, G. L. and Gollub, J. B. *Chaotic Dynamics: An Introduction, 2nd ed.* Cambridge: Cambridge University Press, 1996.
Cvitanovic, P. *Universality in Chaos: A Reprint Selection, 2nd ed.* Bristol: Adam Hilger, 1989.
Dickau, R. M. "Magnetic Pendulum." `http://forum.swarthmore.edu/advanced/robertd/magneticpendulum.html`.
Drazin, P. G. *Nonlinear Systems.* Cambridge, England: Cambridge University Press, 1992.
Field, M. and Golubitsky, M. *Symmetry in Chaos: A Search for Pattern in Mathematics, Art and Nature.* Oxford, England: Oxford University Press, 1992.
Gleick, J. *Chaos: Making a New Science.* New York: Penguin, 1988.

Guckenheimer, J. and Holmes, P. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields, 3rd ed.* New York: Springer-Verlag, 1997.

Lichtenberg, A. and Lieberman, M. *Regular and Stochastic Motion, 2nd ed.* New York: Springer-Verlag, 1994.

Lorenz, E. N. *The Essence of Chaos.* Seattle, WA: University of Washington Press, 1996.

Ott, E. *Chaos in Dynamical Systems.* New York: Cambridge University Press, 1993.

Ott, E.; Sauer, T.; and Yorke, J. A. *Coping with Chaos: Analysis of Chaotic Data and the Exploitation of Chaotic Systems.* New York: Wiley, 1994.

Peitgen, H.-O.; Jürgens, H.; and Saupe, D. *Chaos and Fractals: New Frontiers of Science.* New York: Springer-Verlag, 1992.

Poon, L. "Chaos at Maryland." `http://www-chaos.umd.edu.`

Rasband, S. N. *Chaotic Dynamics of Nonlinear Systems.* New York: Wiley, 1990.

Strogatz, S. H. *Nonlinear Dynamics and Chaos, with Applications to Physics, Biology, Chemistry, and Engineering.* Reading, MA: Addison-Wesley, 1994.

Tabor, M. *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, 1989.

Tufillaro, N.; Abbott, T. R.; and Reilly, J. *An Experimental Approach to Nonlinear Dynamics and Chaos.* Redwood City, CA: Addison-Wesley, 1992.

Wiggins, S. *Global Bifurcations and Chaos: Analytical Methods.* New York: Springer-Verlag, 1988.

Wiggins, S. *Introduction to Applied Nonlinear Dynamical Systems and Chaos.* New York: Springer-Verlag, 1990.

Chaos Game

Pick a point at random inside a regular n-gon. Then draw the next point a fraction r of the distance between it and a VERTEX picked at random. Continue the process (after throwing out the first few points). The result of this "chaos game" is sometimes, but not always, a FRACTAL. The case $(n, r) = (4, 1/2)$ gives the interior of a SQUARE with all points visited with equal probability.

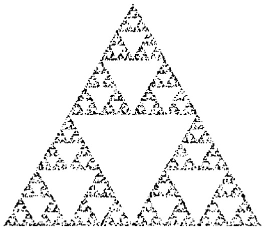

$(3, 1/2)$

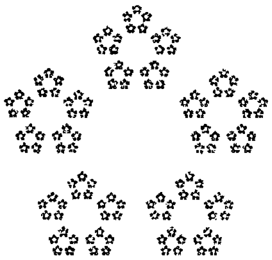

$(5, 1/3)$

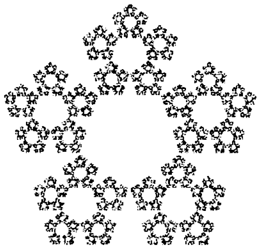

$(5, 3/8)$

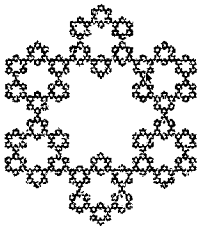

$(6, 1/3)$

see also BARNSLEY'S FERN

References

Barnsley, M. F. and Rising, H. *Fractals Everywhere, 2nd ed.* Boston, MA: Academic Press, 1993.

Dickau, R. M. "The Chaos Game." `http://forum.swarthmore.edu/advanced/robertd/chaos_game.html.`

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 149–163, 1991.

Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m.`

Character (Group)

The GROUP THEORY term for what is known to physicists as the TRACE. All members of the same CONJUGACY CLASS in the same representation have the same character. Members of other CONJUGACY CLASSES may also have the same character, however. An (abstract) GROUP can be uniquely identified by a listing of the characters of its various representations, known as a CHARACTER TABLE. Some of the SCHÖNFLIES SYMBOLS denote different sets of symmetry operations but correspond to the same abstract GROUP and so have the same CHARACTER TABLES.

Character (Multiplicative)

A continuous HOMEOMORPHISM of a GROUP into the NONZERO COMPLEX NUMBERS. A multiplicative character ω gives a REPRESENTATION on the 1-D SPACE $\mathbb{C}$ of COMPLEX NUMBERS, where the REPRESENTATION action by $g \in G$ is multiplication by $\omega(g)$. A multiplicative character is UNITARY if it has ABSOLUTE VALUE 1 everywhere.

References

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Character (Number Theory)

A number theoretic function $\chi_k(n)$ for POSITIVE integral n is a character modulo k if

$$\begin{aligned} \chi_k(1) &= 1 \\ \chi_k(n) &= \chi_k(n+k) \\ \chi_k(m)\chi_k(n) &= \chi_k(mn) \end{aligned}$$

for all m, n, and

$$\chi_k(n) = 0$$

if $(k, n) \neq 1$. χ_k can only assume values which are $\phi(k)$ ROOTS OF UNITY, where ϕ is the TOTIENT FUNCTION.

see also DIRICHLET L-SERIES

Character Table

C_1	E
A	1

C_s	E	σ_h		
A	1	1	x, y, R_z	x^2, y^2, z^2, xy
B	1	-1	z, R_x, R_y	yz, xz

C_i	E	i		
A_g	1	1	R_x, R_y, R_z	$x^2, y^2, z^2, xy, xz, yz$
A_u	1	−1	x, y, z	

C_2	E	C_2		
A	1	1	z, R_z	x^2, y^2, z^2, xy
B	1	−1	x, y, R_x, R_y	yz, xz

C_3	E	C_3	$C_3{}^2$		$\varepsilon = \exp(2\pi i/3)$
A	1	1	1	z, R_z	x^2, y^2, z^2, xy
E	1	ε	ε^*	$(x, y)(R_x, R_y)$	$(x^2 - y^2, xy)(yz, xz)$
	1	ε^*	ε		

C_4	E	C_3	C_2	$C_4{}^3$		
A	1	1	1	1	z, R_z	$x^2 + y^2, z^2$
B	1	−1	1	−1		$x^2 - y^2, xy$
E	1	i	−1	$-i$	$(x, y)(R_x, R_y)$	(yz, xz)
	1	$-i$	1	i		

C_5	E	C_5	$C_5{}^2$	$C_5{}^3$	$C_5{}^4$		$\varepsilon = \exp(2\pi i/5)$
A	1	1	1	1	1	z, R_z	$x^2 + y^2, z^2$
E_1	1	ε	ε^2	ε^{2*}	ε^*	$(x, y)(R_x, R_y)$	(yz, xz)
	1	ε^*	ε^{2*}	ε^2	ε		
E_2	1	ε^2	ε^*	ε	ε^{2*}		$(x^2 - y^2, xy)$
	1	ε^{2*}	ε	ε^*	ε^2		

C_6	E	C_6	C_3	C_2	$C_3{}^2$	$C_6{}^5$		$\varepsilon = \exp(2\pi i/6)$
A	1	1	1	1	1	1	z, R_z	$x^2 + y^2, z^2$
B	1	−1	1	−1	1	−1		
E_1	1	ε	$-\varepsilon^*$	−1	$-\varepsilon$	ε^*	(x, y) (R_x, R_y)	(yz, xz)
	1	ε^*	$-\varepsilon$	−1	$-\varepsilon^*$	ε		
E_2	1	$-\varepsilon$	$-\varepsilon$	1	$-\varepsilon^*$	$-\varepsilon$		$(x^2 - y^2, xy)$
	1	$-\varepsilon^*$	$-\varepsilon^*$	1	$-\varepsilon$	ε^*		

D_2	E	$C_2(z)$	$C_2(y)$	$C_2(x)$		
A_1	1	1	1	1		$x^2 + y^2, z^2$
B_1	1	1	−1	−1	z, R_z	xy
B_2	1	−1	1	−1	y, R_y	xz
B_3	1	−1	−1	1	z, R_z	yz

D_3	E	$2C_3$	$3C_2$		
A_1	1	1	1		$x^2 + y^2, z^2$
A_2	1	1	−1	z, R_z	xy
E	2	−1	0	$(x, y)(R_x, R_y)$	$(x^2 - y^2, xy)(xz, yz)$

D_4	E	$2C_4$	C_2	$2C_2'$	$2C_2''$		
A_1	1	1	1	1	1		$x^2 + y^2, z^2$
A_2	1	1	1	−1	−1	z, R_z	
B_1	1	−1	1	1	−1		$x^2 - y^2$
B_2	1	−1	1	−1	1		xy
E	2	0	−2	0	0	$(x, y)(R_x, R_y)$	(xz, yz)

D_5	E	$2C_5$	$2C_5{}^2$	$5C_2$		
A_1	1	1	1	1		$x^2 + y^2, z^2$
B_1	1	1	1	−1	z, R_z	
B_2	2	2 cos 72°	2 cos 144°	0	$(x, y)(R_x, R_y)$	(xz, yz)
B_3	2	2 cos 144°	2 cos 72°	0		$(x^2 - y^2, xy)$

D_6	E	$2C_6$	$2C_3$	C_2	$3C_2'$	$3C_2''$		
A_1	1	1	1	1	1	1		$x^2 + y^2, z^2$
A_2	1	1	1	1	−1	−1	z, R_z	
B_1	1	−1	1	−1	1	−1		
B_2	1	−1	1	−1	−1	1	$(x, y)(R_x, R_y)$	
E_1	2	1	−1	−2	0	0		(xz, yz)
E_2	2	−1	−1	2	0	0		$(x^2 - y^2, xy)$

C_{2v}	E	C_2	$\sigma_v(xz)$	$\sigma_v'(yz)$		
A_1	1	1	1	1	z	x^2, y^2, z^2
A_2	1	1	−1	−1	R_z	xy
B_1	1	−1	1	−1	x, R_y	xz
B_2	1	−1	−1	1	y, R_x	yz

C_{3v}	E	$2C_3$	$3\sigma_v$		
A_1	1	1	1	z	$x^2 + y^2, z^2$
A_2	1	1	−1	R_z	
E	2	−1	0	$(x, y)(R_x, R_y)$	$(x^2 - y^2, xy)(xz, yz)$

C_{4v}	E	$2C_4$	C_2	$2\sigma_v$	$2\sigma_d$		
A_1	1	1	1	1	1	z	$x^2 + y^2, z^2$
A_2	1	1	1	−1	−1	R_z	
B_1	1	−1	1	1	−1		$x^2 - y^2$
B_2	1	−1	1	−1	1		xy
E	2	0	−2	0	0	$(x, y)(R_x, R_y)$	(xz, yz)

C_{5v}	E	$2C_5$	$2C_5{}^2$	$5\sigma_v$		
A_1	1	1	1	1	z	$x^2 + y^2, z^2$
B_1	1	1	1	−1	R_z	
B_2	2	2 cos 72°	2 cos 144°	0	$(x, y)(R_x, R_y)$	(xz, yz)
B_3	2	2 cos 144°	2 cos 72°	0		$(x^2 - y^2, xy)$

C_{6v}	E	$2C_6$	$2C_3$	C_2	$3\sigma_v$	$3\sigma_d$		
A_1	1	1	1	1	1	1	z	$x^2 + y^2, z^2$
A_2	1	1	1	1	−1	−1	R_z	
B_1	1	−1	1	−1	1	−1		
B_2	1	−1	1	−1	−1	1		
E_1	2	1	−1	−2	0	0	$(x, y)(R_x, R_y)$	(xz, yz)
E_2	2	−1	−1	2	0	0		$(x^2 - y^2, xy)$

$C_{\infty v}$	E	$C_\infty{}^\Phi$	$\ldots$	$\infty\sigma_v$		
$A_1 \equiv \Sigma^+$	1	1	$\ldots$	1	z	$x^2 + y^2, z^2$
$A_2 \equiv \Sigma^-$	1	1	$\ldots$	−1	R_z	
$E_1 \equiv \Pi$	2	2 cos Φ	$\ldots$	0	$(x, y); (R_x, R_y)$	(xz, yz)
$E_2 \equiv \Delta$	2	2 cos 2Φ	$\ldots$	0		$(x^2 - y^2, xy)$
$E_3 \equiv \Phi$	2	2 cos 3Φ	$\ldots$	0		
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$		

References

Bishop, D. M. "Character Tables." Appendix 1 in *Group Theory and Chemistry.* New York: Dover, pp. 279–288, 1993.

Cotton, F. A. *Chemical Applications of Group Theory, 3rd ed.* New York: Wiley, 1990.

Iyanaga, S. and Kawada, Y. (Eds.). "Characters of Finite Groups." Appendix B, Table 5 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 1496–1503, 1980.

Characteristic Class

Characteristic classes are COHOMOLOGY classes in the BASE SPACE of a VECTOR BUNDLE, defined through OBSTRUCTION theory, which are (perhaps partial) obstructions to the existence of k everywhere linearly independent vector FIELDS on the VECTOR BUNDLE. The most common examples of characteristic classes are the CHERN, PONTRYAGIN, and STIEFEL-WHITNEY CLASSES.

Characteristic (Elliptic Integral)

A parameter n used to specify an ELLIPTIC INTEGRAL OF THE THIRD KIND.

see also AMPLITUDE, ELLIPTIC INTEGRAL, MODULAR ANGLE, MODULUS (ELLIPTIC INTEGRAL), NOME, PARAMETER

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 590, 1972.

Characteristic Equation

The equation which is solved to find a MATRIX's EIGENVALUES, also called the CHARACTERISTIC POLYNOMIAL. Given a 2×2 system of equations with MATRIX

$$\mathsf{M} \equiv \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \tag{1}$$

the MATRIX EQUATION is

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = t \begin{bmatrix} x \\ y \end{bmatrix}, \tag{2}$$

which can be rewritten

$$\begin{bmatrix} a-t & b \\ c & d-t \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = t \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{3}$$

M can have no MATRIX INVERSE, since otherwise

$$\begin{bmatrix} x \\ y \end{bmatrix} = \mathsf{M}^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{4}$$

which contradicts our ability to pick arbitrary x and y. Therefore, M has no inverse, so its DETERMINANT is 0. This gives the characteristic equation

$$\begin{vmatrix} a-t & b \\ c & d-t \end{vmatrix} = 0, \tag{5}$$

where $|\mathsf{A}|$ denotes the DETERMINANT of A. For a general $k \times k$ MATRIX

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix}, \tag{6}$$

the characteristic equation is

$$\begin{vmatrix} a_{11}-t & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22}-t & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk}-t \end{vmatrix} = 0. \tag{7}$$

see also BALLIEU'S THEOREM, CAYLEY-HAMILTON THEOREM, PARODI'S THEOREM, ROUTH-HURWITZ THEOREM

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 1117–1119, 1979.

Characteristic (Euler)

see EULER CHARACTERISTIC

Characteristic Factor

A characteristic factor is a factor in a particular factorization of the TOTIENT FUNCTION $\phi(n)$ such that the product of characteristic factors gives the representation of a corresponding abstract GROUP as a DIRECT PRODUCT. By computing the characteristic factors, any ABELIAN GROUP can be expressed as a DIRECT PRODUCT of CYCLIC SUBGROUPS, for example, $Z_2 \otimes Z_4$ or $Z_2 \otimes Z_2 \otimes Z_2$. There is a simple algorithm for determining the characteristic factors of MODULO MULTIPLICATION GROUPS.

see also CYCLIC GROUP, DIRECT PRODUCT (GROUP), MODULO MULTIPLICATION GROUP, TOTIENT FUNCTION

References

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 94, 1993.

Characteristic (Field)

For a FIELD K with multiplicative identity 1, consider the numbers $2 = 1 + 1$, $3 = 1 + 1 + 1$, $4 = 1 + 1 + 1 + 1$, etc. Either these numbers are all different, in which case we say that K has characteristic 0, or two of them will be equal. In this case, it is straightforward to show that, for some number p, we have $\underbrace{1 + 1 + \ldots + 1}_{p \text{ times}} = 0$.

If p is chosen to be as small as possible, then p will be a PRIME, and we say that K has characteristic p. The FIELDS $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, and the p-ADIC NUMBERS $\mathbb{Q}_p$ have characteristic 0. For p a PRIME, the GALOIS FIELD $\mathrm{GF}(p^n)$ has characteristic p.

If H is a SUBFIELD of K, then H and K have the same characteristic.

see also FIELD, SUBFIELD

Characteristic Function

The characteristic function $\phi(t)$ is defined as the FOURIER TRANSFORM of the PROBABILITY DENSITY FUNCTION,

$$\begin{aligned}\phi(t) = \mathcal{F}[P(x)] &= \int_{-\infty}^{\infty} e^{itx} P(x)\,dx && (1)\\ &= \int_{-\infty}^{\infty} P(x)\,dx + it\int_{-\infty}^{\infty} xP(x)\,dx\\ &\quad + \tfrac{1}{2}(it)^2\int_{-\infty}^{\infty} x^2 P(x)\,dx + \ldots && (2)\\ &= \sum_{k=0}^{\infty}\frac{(it)^k}{k!}\mu'_k && (3)\\ &= 1 + it\mu'_1 - \tfrac{1}{2}t^2\mu'_2 - \tfrac{1}{3!}it^3\mu'_3 + \tfrac{1}{4!}t^4\mu'_4 + \ldots, && (4)\end{aligned}$$

where μ'_n (sometimes also denoted ν_n) is the nth MOMENT about 0 and $\mu'_0 \equiv 1$. The characteristic function can therefore be used to generate MOMENTS about 0,

$$\phi^{(n)}(0) \equiv \left[\frac{d^n\phi}{dt^n}\right]_{t=0} = i^n\mu'_n \qquad (5)$$

or the CUMULANTS κ_n,

$$\ln\phi(t) \equiv \sum_{n=0}^{\infty}\kappa_n\frac{(it)^n}{n!}. \qquad (6)$$

A DISTRIBUTION is not uniquely specified by its MOMENTS, but is uniquely specified by its characteristic function.

see also CUMULANT, MOMENT, MOMENT-GENERATING FUNCTION, PROBABILITY DENSITY FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 928, 1972.

Kenney, J. F. and Keeping, E. S. "Moment-Generating and Characteristic Functions," "Some Examples of Moment-Generating Functions," and "Uniqueness Theorem for Characteristic Functions." §4.6–4.8 in *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, pp. 72–77, 1951.

Characteristic (Partial Differential Equation)

Paths in a 2-D plane used to transform PARTIAL DIFFERENTIAL EQUATIONS into systems of ORDINARY DIFFERENTIAL EQUATIONS. They were invented by Riemann. For an example of the use of characteristics, consider the equation

$$u_t - 6uu_x = 0.$$

Now let $u(s) = u(x(s), t(s))$. Since

$$\frac{du}{ds} = \frac{dx}{ds}u_x + \frac{dt}{ds}u_t,$$

it follows that $dt/ds = 1$, $dx/ds = -6u$, and $du/ds = 0$. Integrating gives $t(s) = s$, $x(s) = -6su_0(x)$, and $u(s) = u_0(x)$, where the constants of integration are 0 and $u_0(x) = u(x, 0)$.

Characteristic Polynomial

The expanded form of the CHARACTERISTIC EQUATION.

$$\det(x\mathsf{I} - \mathsf{A}),$$

where A is an $n \times n$ MATRIX and I is the IDENTITY MATRIX.

see also CAYLEY-HAMILTON THEOREM

Characteristic (Real Number)

For a REAL NUMBER x, $\lfloor x \rfloor = \texttt{int(x)}$ is called the characteristic. Here, $\lfloor x \rfloor$ is the FLOOR FUNCTION.

see also MANTISSA, SCIENTIFIC NOTATION

Charlier's Check

A check which can be used to verify correct computation of MOMENTS.

Chasles-Cayley-Brill Formula

The number of coincidences of a (ν, ν') correspondence of value γ on a curve of GENUS p is given by

$$\nu + \nu' + 2p\gamma.$$

see also ZEUTHEN'S THEOREM

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 129, 1959.

Chasles's Contact Theorem

If a one-parameter family of curves has index N and class M, the number tangent to a curve of order n_1 and class m_1 in general position is

$$m_1 N + n_1 M.$$

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 436, 1959.

Chasles's Polars Theorem

If the TRILINEAR POLARS of the VERTICES of a TRIANGLE are distinct from the respectively opposite sides, they meet the sides in three COLLINEAR points.

see also COLLINEAR, TRIANGLE, TRILINEAR POLAR

Chasles's Theorem

If two projective PENCILS of curves of orders n and n' have no common curve, the LOCUS of the intersections of corresponding curves of the two is a curve of order $n+n'$ through all the centers of either PENCIL. Conversely, if a curve of order $n+n'$ contains all centers of a PENCIL of order n to the multiplicity demanded by NOETHER'S FUNDAMENTAL THEOREM, then it is the LOCUS of the intersections of corresponding curves of this PENCIL and one of order n' projective therewith.

see also NOETHER'S FUNDAMENTAL THEOREM, PENCIL

References
Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 33, 1959.

Chebyshev Approximation Formula

Using a CHEBYSHEV POLYNOMIAL OF THE FIRST KIND T, define

$$\begin{aligned} c_j &\equiv \frac{2}{N}\sum_{k=1}^{N} f(x_k)T_j(x_k) \\ &= \frac{2}{N}\sum_{k=1}^{N} f\left[\cos\left\{\frac{\pi(k-\frac{1}{2})}{N}\right\}\right]\cos\left\{\frac{\pi j(k-\frac{1}{2})}{N}\right\}. \end{aligned}$$

Then

$$f(x) \approx \sum_{k=0}^{N-1} c_k T_k(x) - \tfrac{1}{2}c_0.$$

It is exact for the N zeros of $T_N(x)$. This type of approximation is important because, when truncated, the error is spread smoothly over $[-1,1]$. The Chebyshev approximation formula is very close to the MINIMAX POLYNOMIAL.

References
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Chebyshev Approximation," "Derivatives or Integrals of a Chebyshev-Approximated Function," and "Polynomial Approximation from Chebyshev Coefficients." §5.8, 5.9, and 5.10 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 184–188, 189–190, and 191–192, 1992.

Chebyshev Constants

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

The constants

$$\lambda_{m,n} = \inf_{r\in R_{m,n}} \sup_{x\geq 0} |e^{-x} - r(x)|,$$

where

$$r(x) = \frac{p(x)}{q(x)},$$

p and q are mth and nth order POLYNOMIALS, and $R_{m,n}$ is the set all RATIONAL FUNCTIONS with REAL coefficients.

see also ONE-NINTH CONSTANT, RATIONAL FUNCTION

References
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/onenin/onenin.html`.
Petrushev, P. P. and Popov, V. A. *Rational Approximation of Real Functions.* New York: Cambridge University Press, 1987.
Varga, R. S. *Scientific Computations on Mathematical Problems and Conjectures.* Philadelphia, PA: SIAM, 1990.
Philadelphia, PA: SIAM, 1990.

Chebyshev Deviation

$$\max_{a\leq x\leq b}\{|f(x) - \rho(x)|w(x)\}.$$

References
Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., p. 41, 1975.

Chebyshev Differential Equation

$$(1-x^2)\frac{d^2y}{dx^2} - x\frac{dy}{dx} + m^2y = 0 \qquad (1)$$

for $|x| < 1$. The Chebyshev differential equation has regular SINGULARITIES at -1, 1, and ∞. It can be solved by series solution using the expansions

$$y = \sum_{n=0}^{\infty} a_n x^n \qquad (2)$$

$$\begin{aligned} y' &= \sum_{n=0}^{\infty} n a_n x^{n-1} = \sum_{n=1}^{\infty} n a_n x^{n-1} \\ &= \sum_{n=0}^{\infty}(n+1)a_{n+1}x^n \qquad (3) \end{aligned}$$

$$\begin{aligned} y'' &= \sum_{n=0}^{\infty}(n+1)n a_{n+1}x^{n-1} = \sum_{n=1}^{\infty}(n+1)n a_{n+1}x^{n-1} \\ &= \sum_{n=0}^{\infty}(n+2)(n+1)a_{n+2}x^n. \qquad (4) \end{aligned}$$

Now, plug (2-4) into the original equation (1) to obtain

$$\begin{aligned} &(1-x^2)\sum_{n=0}^{\infty}(n+2)(n+1)a_{n+2}x^n \\ &\qquad -x\sum_{n=0}^{\infty}(n+1)n_{n+1}x^n + m^2\sum_{n=0}^{\infty}a_n x^n = 0 \qquad (5) \end{aligned}$$

$$\begin{aligned} &\sum_{n=0}^{\infty}(n+2)(n+1)a_{n+2}x^n - \sum_{n=0}^{\infty}(n+2)(n+1)a_{n+2}x^{n+2} \\ &\qquad - \sum_{n=0}^{\infty}(n+1)a_{n+1}x^{n+1} + m^2\sum_{n=0}^{\infty}a_n x^n = 0 \qquad (6) \end{aligned}$$

$$\sum_{n=0}^{\infty}(n+2)(n+1)a_{n+2}x^n - \sum_{n=2}^{\infty}n(n-1)a_nx^{n+2} - \sum_{n=1}^{\infty}na_nx^n + m^2\sum_{n=0}^{\infty}a_nx^n = 0 \quad (7)$$

$$2\cdot 1a_2 + 3\cdot 2a_3x - 1\cdot ax + m^2a_0 + m^2a_1x + \sum_{n=2}^{\infty}[(n+2)(n+1)a_{n+2} - n(n-1)a_n - na_n + m^2a_n]x^n = 0 \quad (8)$$

$$(2a_2 + m^2a_0) + [(m^2-1)a_1 + 6a_3]x + \sum_{n=2}^{\infty}[(n+2)(n+1)a_{n+2} + (m^2-n^2)a_n]x^n = 0, \quad (9)$$

so

$$2a_2 + m^2a_0 = 0 \quad (10)$$

$$(m^2-1)a_1 + 6a_3 = 0 \quad (11)$$

$$a_{n+2} = \frac{n^2-m^2}{(n+1)(n+2)}a_n \qquad \text{for } n = 2, 3, \ldots. \quad (12)$$

The first two are special cases of the third, so the general recurrence relation is

$$a_{n+2} = \frac{n^2-m^2}{(n+1)(n+2)}a_n \qquad \text{for } n = 0, 1, \ldots. \quad (13)$$

From this, we obtain for the EVEN COEFFICIENTS

$$a_2 = -\tfrac{1}{2}m^2a_0 \quad (14)$$

$$a_4 = \frac{2^2-m^2}{3\cdot 4}a_2 = \frac{(2^2-m^2)(-m^2)}{1\cdot 2\cdot 3\cdot 4}a_0 \quad (15)$$

$$a_{2n} = \frac{[(2n)^2-m^2][(2n-2)^2-m^2]\cdots[-m^2]}{(2n)!}a_0, \quad (16)$$

and for the ODD COEFFICIENTS

$$a_3 = \frac{1-m^2}{6}a_0 \quad (17)$$

$$a_5 = \frac{3^2-m^2}{4\cdot 5}a_3 = \frac{(3^2-m^2)(1^2-m^2)}{5!}a_1 \quad (18)$$

$$a_{2n-1} = \frac{[(2n-1)^2-m^2][(2n-3)^2-m^2]\cdots[1^2-m^2]}{(2n+1)!}a_1. \quad (19)$$

So the general solution is

$$y = a_0\left[1 + \sum_{k=2,4,\ldots}^{\infty}\frac{[k^2-m^2][(k-2)^2-m^2]\cdots[-m^2]}{k!}x^k\right] + a_1\left[x + \sum_{k=3,5,\ldots}^{\infty}\frac{[(k-2)^2-m^2][(k-2)^2-m^2]\cdots[1^2-m^2]}{k!}x^k\right] \quad (20)$$

If n is EVEN, then y_1 terminates and is a POLYNOMIAL solution, whereas if n is ODD, then y_2 terminates and is a POLYNOMIAL solution. The POLYNOMIAL solutions defined here are known as CHEBYSHEV POLYNOMIALS OF THE FIRST KIND. The definition of the CHEBYSHEV POLYNOMIAL OF THE SECOND KIND gives a similar, but distinct, recurrence relation

$$a'_{n+2} = \frac{(n+1)^2-m^2}{(n+2)(n+3)}a'_n \qquad \text{for } n = 0, 1, \ldots. \quad (21)$$

Chebyshev Function

$$\theta(x) \equiv \sum_{p\leq x}\ln p,$$

where the sum is over PRIMES p, so

$$\lim_{x\to\infty}\frac{x}{\theta(x)} = 1.$$

Chebyshev-Gauss Quadrature

Also called CHEBYSHEV QUADRATURE. A GAUSSIAN QUADRATURE over the interval $[-1,1]$ with WEIGHTING FUNCTION $W(x) = 1/\sqrt{1-x^2}$. The ABSCISSAS for quadrature order n are given by the roots of the CHEBYSHEV POLYNOMIAL OF THE FIRST KIND $T_n(x)$, which occur symmetrically about 0. The WEIGHTS are

$$w_i = -\frac{A_{n+1}\gamma_n}{A_nT'_n(x_i)T_{n+1}(x_i)} = \frac{A_n}{A_{n-1}}\frac{\gamma_{n-1}}{T_{n-1}(x_i)T'_n(x_i)}, \quad (1)$$

where A_n is the COEFFICIENT of x^n in $T_n(x)$. For HERMITE POLYNOMIALS,

$$A_n = 2^{n-1}, \quad (2)$$

so

$$\frac{A_{n+1}}{A_n} = 2. \quad (3)$$

Additionally,

$$\gamma_n = \tfrac{1}{2}\pi, \quad (4)$$

so

$$w_i = -\frac{\pi}{T_{n+1}(x_i)T'_n(x_i)}. \quad (5)$$

Since

$$T_n(x) = \cos(n\cos^{-1}x), \quad (6)$$

the ABSCISSAS are given explicitly by

$$x_i = \cos\left[\frac{(2i-1)\pi}{2n}\right]. \quad (7)$$

Since

$$T'_n(x_i) = \frac{(-1)^{i+1}n}{\alpha_i} \quad (8)$$

$$T_{n+1}(x_i) = (-1)^i\sin\alpha_i, \quad (9)$$

where

$$\alpha_i = \frac{(2i-1)\pi}{2n}, \tag{10}$$

all the WEIGHTS are

$$w_i = \frac{\pi}{n}. \tag{11}$$

The explicit FORMULA is then

$$\int_{-1}^{1} \frac{f(x)\,dx}{\sqrt{1-x^2}} = \frac{\pi}{n}\sum_{k=1}^{n} f\left[\cos\left(\frac{2k-1}{2n}\pi\right)\right] + \frac{2\pi}{2^{2n}(2n)!} f^{(2n)}(\xi). \tag{12}$$

n	x_i	w_i
2	±0.707107	1.5708
3	0	1.0472
	±0.866025	1.0472
4	±0.382683	0.785398
	±0.92388	0.785398
5	0	0.628319
	±0.587785	0.628319
	±0.951057	0.628319

References

Hildebrand, F. B. *Introduction to Numerical Analysis.* New York: McGraw-Hill, pp. 330–331, 1956.

Chebyshev Inequality

Apply MARKOV'S INEQUALITY with $a \equiv k^2$ to obtain

$$P[(x-\mu)^2 \geq k^2] \leq \frac{\langle (x-\mu)^2 \rangle}{k^2} = \frac{\sigma^2}{k^2}. \tag{1}$$

Therefore, if a RANDOM VARIABLE x has a finite MEAN μ and finite VARIANCE σ^2, then $\forall\ k \geq 0$,

$$P(|x-\mu| \geq k) \leq \frac{\sigma^2}{k^2} \tag{2}$$

$$P(|x-\mu| \geq k\sigma) \leq \frac{1}{k^2}. \tag{3}$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 11, 1972.

Chebyshev Integral

$$\int x^p (1-x)^q \, dx.$$

Chebyshev Integral Inequality

$$\int_a^b f_1(x)\,dx \int_a^b f_2(x)\,dx \cdots \int_a^b f_n(x)\,dx \leq (b-a)^{n-1} \int_a^b f(x_1) f(x_2) \cdots f_n(x)\,dx,$$

where f_1, f_2, …, f_n are NONNEGATIVE integrable functions on $[a,b]$ which are monotonic increasing or decreasing.

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1092, 1979.

Chebyshev Phenomenon

see PRIME QUADRATIC EFFECT

Chebyshev Polynomial of the First Kind

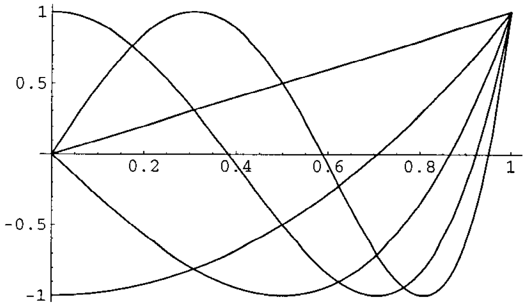

A set of ORTHOGONAL POLYNOMIALS defined as the solutions to the CHEBYSHEV DIFFERENTIAL EQUATION and denoted $T_n(x)$. They are used as an approximation to a LEAST SQUARES FIT, and are a special case of the ULTRASPHERICAL POLYNOMIAL with $\alpha = 0$. The Chebyshev polynomials of the first kind $T_n(x)$ are illustrated above for $x \in [0,1]$ and $n = 1, 2, \ldots, 5$.

The Chebyshev polynomials of the first kind can be obtained from the generating functions

$$g_1(t,x) \equiv \frac{1-t^2}{1-2xt+t^2} = T_0(x) + 2\sum_{n=1}^{\infty} T_n(x) t^n \tag{1}$$

and

$$g_2(t,x) \equiv \frac{1-xt}{1-2xt+t^2} = \sum_{n=0}^{\infty} T_n(x) t^n \tag{2}$$

for $|x| \leq 1$ and $|t| < 1$ (Beeler *et al.* 1972, Item 15). (A closely related GENERATING FUNCTION is the basis for the definition of CHEBYSHEV POLYNOMIAL OF THE SECOND KIND.) They are normalized such that $T_n(1) = 1$. They can also be written

$$T_n(x) = \frac{n}{2} \sum_{r=0}^{\lfloor n/2 \rfloor} \frac{(-1)^r}{n-r} \binom{n-r}{r} (2x)^{n-2r}, \tag{3}$$

or in terms of a DETERMINANT

$$T_n = \begin{vmatrix} x & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 2x & 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 2x & 1 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 2x & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 2x \end{vmatrix}. \tag{4}$$

In closed form,

$$T_n(x) = \cos(n\cos^{-1} x) = \sum_{m=0}^{\lfloor n/2 \rfloor} \binom{n}{2m} x^{n-2m}(x^2-1)^m, \tag{5}$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT and $\lfloor x \rfloor$ is the FLOOR FUNCTION. Therefore, zeros occur when

$$x = \cos\left[\frac{\pi(k-\frac{1}{2})}{n}\right] \tag{6}$$

for $k = 1, 2, \ldots, n$. Extrema occur for

$$x = \cos\left(\frac{\pi k}{n}\right), \tag{7}$$

where $k = 0, 1, \ldots, n$. At maximum, $T_n(x) = 1$, and at minimum, $T_n(x) = -1$. The Chebyshev POLYNOMIALS are ORTHONORMAL with respect to the WEIGHTING FUNCTION $(1-x^2)^{-1/2}$

$$\int_{-1}^{1} \frac{T_m(x)T_n(x)\,dx}{\sqrt{1-x^2}} = \begin{cases} \frac{1}{2}\pi\delta_{nm} & \text{for } m \neq 0,\ n \neq 0 \\ \pi & \text{for } m = n = 0, \end{cases} \tag{8}$$

where δ_{mn} is the KRONECKER DELTA. Chebyshev polynomials of the first kind satisfy the additional discrete identity

$$\sum_{k=1}^{m} T_i(x_k)T_j(x_k) = \begin{cases} \frac{1}{2}m\delta_{ij} & \text{for } i \neq 0,\ j \neq 0 \\ m & \text{for } i = j = 0, \end{cases} \tag{9}$$

where x_k for $k = 1, \ldots, m$ are the m zeros of $T_m(x)$. They also satisfy the RECURRENCE RELATIONS

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \tag{10}$$

$$T_{n+1}(x) = xT_n(x) - \sqrt{(1-x^2)\{1-[T_n(x)]^2\}} \tag{11}$$

for $n \geq 1$. They have a COMPLEX integral representation

$$T_n(x) = \frac{1}{4\pi i}\int_\gamma \frac{(1-z^2)z^{-n-1}\,dz}{1-2xz+z^2} \tag{12}$$

and a Rodrigues representation

$$T_n(x) = \frac{(-1)^n\sqrt{\pi}\,(1-x^2)^{1/2}}{2n(n-\frac{1}{2})!}\frac{d^n}{dx^n}[(1-x^2)^{n-1/2}]. \tag{13}$$

Using a FAST FIBONACCI TRANSFORM with multiplication law

$$(A,B)(C,D) = (AD + BC + 2xAC, BD - AC) \tag{14}$$

gives

$$(T_{n+1}(x), -T_n(x)) = (T_1(x), -T_0(x))(1,0)^n. \tag{15}$$

Using GRAM-SCHMIDT ORTHONORMALIZATION in the range (−1,1) with WEIGHTING FUNCTION $(1-x^2)^{(-1/2)}$ gives

$$p_0(x) = 1 \tag{16}$$

$$\begin{aligned} p_1(x) &= \left[x - \frac{\int_{-1}^{1} x(1-x^2)^{-1/2}\,dx}{\int_{-1}^{1}(1-x^2)^{-1/2}\,dx}\right] \\ &= x - \frac{[-(1-x^2)^{1/2}]_{-1}^{1}}{[\sin^{-1} x]_{-1}^{1}} = x \end{aligned} \tag{17}$$

$$\begin{aligned} p_2(x) &= \left[x - \frac{\int_{-1}^{1} x^3(1-x^2)^{-1/2}\,dx}{\int_{-1}^{1} x^2(1-x^2)^{-1/2}\,dx}\right]x \\ &\quad - \left[\frac{\int_{-1}^{1} x^2(1-x^2)^{-1/2}\,dx}{\int_{-1}^{1}(1-x^2)^{-1/2}\,dx}\right]\cdot 1 \\ &= [x-0]x - \frac{\frac{\pi}{2}}{\pi} = x^2 - \tfrac{1}{2}, \end{aligned} \tag{18}$$

etc. Normalizing such that $T_n(1) = 1$ gives

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_2(x) &= 2x^2 - 1 \\ T_3(x) &= 4x^3 - 3x \\ T_4(x) &= 8x^4 - 8x^2 + 1 \\ T_5(x) &= 16x^5 - 20x^3 + 5x \\ T_6(x) &= 32x^6 - 48x^4 + 18x^2 - 1. \end{aligned}$$

The Chebyshev polynomial of the first kind is related to the BESSEL FUNCTION OF THE FIRST KIND $J_n(x)$ and MODIFIED BESSEL FUNCTION OF THE FIRST KIND $I_n(x)$ by the relations

$$J_n(x) = i^n T_n\left(i\frac{d}{dx}\right) J_0(x) \tag{19}$$

$$I_n(x) = T_n\left(\frac{d}{dx}\right) I_0(x). \tag{20}$$

Letting $x \equiv \cos\theta$ allows the Chebyshev polynomials of the first kind to be written as

$$T_n(x) = \cos(n\theta) = \cos(n\cos^{-1} x). \tag{21}$$

The second linearly dependent solution to the transformed differential equation

$$\frac{d^2T_n}{d\theta^2}+n^2T_n=0 \tag{22}$$

is then given by

$$V_n(x)=\sin(n\theta)=\sin(n\cos^{-1}x), \tag{23}$$

which can also be written

$$V_n(x)=\sqrt{1-x^2}\,U_{n-1}(x), \tag{24}$$

where U_n is a CHEBYSHEV POLYNOMIAL OF THE SECOND KIND. Note that $V_n(x)$ is therefore *not* a POLYNOMIAL.

The POLYNOMIAL

$$x^n-2^{1-n}T_n(x) \tag{25}$$

(of degree $n-2$) is the POLYNOMIAL of degree $< n$ which stays closest to x^n in the interval $(-1,1)$. The maximum deviation is 2^{1-n} at the $n+1$ points where

$$x=\cos\left(\frac{k\pi}{n}\right), \tag{26}$$

for $k=0, 1, \ldots, n$ (Beeler *et al.* 1972, Item 15).

see also CHEBYSHEV APPROXIMATION FORMULA, CHEBYSHEV POLYNOMIAL OF THE SECOND KIND

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Orthogonal Polynomials." Ch. 22 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 771–802, 1972.

Arfken, G. "Chebyshev (Tschebyscheff) Polynomials" and "Chebyshev Polynomials—Numerical Applications." §13.3 and 13.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 731–748, 1985.

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Iyanaga, S. and Kawada, Y. (Eds.). "Čebyšev (Tschebyscheff) Polynomials." Appendix A, Table 20.II in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 1478–1479, 1980.

Rivlin, T. J. *Chebyshev Polynomials.* New York: Wiley, 1990.

Spanier, J. and Oldham, K. B. "The Chebyshev Polynomials $T_n(x)$ and $U_n(x)$." Ch. 22 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 193–207, 1987.

Chebyshev Polynomial of the Second Kind

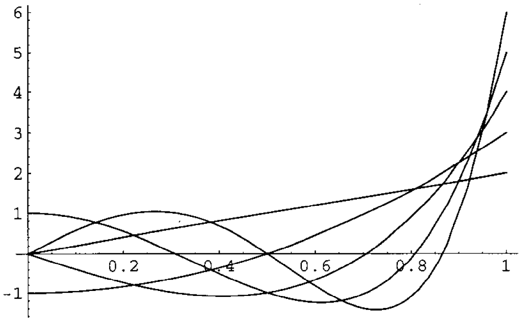

A modified set of Chebyshev POLYNOMIALS defined by a slightly different GENERATING FUNCTION. Used to develop four-dimensional SPHERICAL HARMONICS in angular momentum theory. They are also a special case of the ULTRASPHERICAL POLYNOMIAL with $\alpha=1$. The Chebyshev polynomials of the second kind $U_n(x)$ are illustrated above for $x\in[0,1]$ and $n=1, 2, \ldots, 5$.

The defining GENERATING FUNCTION of the Chebyshev polynomials of the second kind is

$$g_2(t,x)=\frac{1}{1-2xt+t^2}=\sum_{n=0}^{\infty}U_n(x)t^n \tag{1}$$

for $|x|<1$ and $|t|<1$. To see the relationship to a CHEBYSHEV POLYNOMIAL OF THE FIRST KIND (T), take $\partial g/\partial t$,

$$\begin{aligned}\frac{\partial g}{\partial t}&=-(1-2xt+t^2)^{-2}(-2x+2t)\\&=2(t-x)(1-2xt+t^2)^{-2}\\&=\sum_{n=0}^{\infty}nU_n(x)t^{n-1}.\end{aligned} \tag{2}$$

Multiply (2) by t,

$$(2t^2-2xt)(1-2xt+t^2)^{-2}=\sum_{n=0}^{\infty}nU_n(x)t^n \tag{3}$$

and take (3)−(2),

$$\begin{aligned}\frac{(2t^2-2tx)-(1-2xt+t^2)}{(1-2xt+t^2)^2}&=\frac{t^2-1}{(1-2xt+t)^2}\\&=\sum_{n=0}^{\infty}(n-1)U_n(x)t^n.\end{aligned} \tag{4}$$

The Rodrigues representation is

$$U_n(x)=\frac{(-1)^n(n+1)\sqrt{\pi}}{2^{n+1}(n+\frac{1}{2})!(1-x^2)^{1/2}}\frac{d^n}{dx^n}[(1-x^2)^{n+1/2}]. \tag{5}$$

The polynomials can also be written

$$\begin{aligned}U_n(x)&=\sum_{r=0}^{\lfloor n/2\rfloor}(-1)^r\binom{n-r}{r}(2x)^{n-2r}\\&=\sum_{m=0}^{\lceil n/2\rceil}\binom{n+1}{2m+1}x^{n-2m}(x^2-1)^m,\end{aligned} \tag{6}$$

where $\lfloor x\rfloor$ is the FLOOR FUNCTION and $\lceil x\rceil$ is the CEILING FUNCTION, or in terms of a DETERMINANT

$$U_n=\begin{vmatrix}2x & 1 & 0 & 0 & \cdots & 0 & 0\\ 0 & 2x & 1 & 0 & \cdots & 0 & 0\\ 0 & 1 & 2x & 1 & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots\\ 0 & 0 & 0 & 0 & \cdots & 1 & 2x\end{vmatrix}. \tag{7}$$

The first few POLYNOMIALS are

$$\begin{aligned} U_0(x) &= 1 \\ U_1(x) &= 2x \\ U_2(x) &= 4x^2 - 1 \\ U_3(x) &= 8x^3 - 4x \\ U_4(x) &= 16x^4 - 12x^2 + 1 \\ U_5(x) &= 32x^5 - 32x^3 + 6x \\ U_6(x) &= 64x^6 - 80x^4 + 24x^2 - 1. \end{aligned}$$

Letting $x \equiv \cos\theta$ allows the Chebyshev polynomials of the second kind to be written as

$$U_n(x) = \frac{\sin[(n+1)\theta]}{\sin\theta}. \tag{8}$$

The second linearly dependent solution to the transformed differential equation is then given by

$$W_n(x) = \frac{\cos[(n+1)\theta]}{\sin\theta}, \tag{9}$$

which can also be written

$$W_n(x) = (1-x^2)^{-1/2} T_{n+1}(x), \tag{10}$$

where T_n is a CHEBYSHEV POLYNOMIAL OF THE FIRST KIND. Note that $W_n(x)$ is therefore *not* a POLYNOMIAL.

see also CHEBYSHEV APPROXIMATION FORMULA, CHEBYSHEV POLYNOMIAL OF THE FIRST KIND, ULTRASPHERICAL POLYNOMIAL

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Orthogonal Polynomials." Ch. 22 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 771–802, 1972.

Arfken, G. "Chebyshev (Tschebyscheff) Polynomials" and "Chebyshev Polynomials—Numerical Applications." §13.3 and 13.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 731–748, 1985.

Rivlin, T. J. *Chebyshev Polynomials.* New York: Wiley, 1990.

Spanier, J. and Oldham, K. B. "The Chebyshev Polynomials $T_n(x)$ and $U_n(x)$." Ch. 22 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 193–207, 1987.

Chebyshev Quadrature

A GAUSSIAN QUADRATURE-like FORMULA for numerical estimation of integrals. It uses WEIGHTING FUNCTION $W(x) = 1$ in the interval $[-1,1]$ and forces all the weights to be equal. The general FORMULA is

$$\int_{-1}^{1} f(x)\,dx = \frac{2}{n}\sum_{i=1}^{n} f(x_i).$$

The ABSCISSAS are found by taking terms up to y^n in the MACLAURIN SERIES of

$$s_n(y) = \exp\left\{\tfrac{1}{2}n\left[-2 + \ln(1-y)\left(1-\frac{1}{y}\right) + \ln(1+y)\left(1+\frac{1}{y}\right)\right]\right\},$$

and then defining

$$G_n(x) \equiv x^n s_n\left(\frac{1}{x}\right).$$

The ROOTS of $G_n(x)$ then give the ABSCISSAS. The first few values are

$$\begin{aligned} G_0(x) &= 1 \\ G_1(x) &= x \\ G_2(x) &= \tfrac{1}{3}(3x^2 - 1) \\ G_3(x) &= \tfrac{1}{2}(2x^3 - x) \\ G_4(x) &= \tfrac{1}{45}(45x^4 - 30x^2 + 1) \\ G_5(x) &= \tfrac{1}{72}(72x^5 - 60x^3 + 7x) \\ G_6(x) &= \tfrac{1}{105}(105x^6 - 105x^4 + 21x^2 - 1) \\ G_7(x) &= \tfrac{1}{6480}(6480x^7 - 7560x^5 + 2142x^3 - 149x) \\ G_8(x) &= \tfrac{1}{42525}(42525x^8 - 56700x^6 + 20790x^4 \\ &\quad - 2220x^2 - 43) \\ G_9(x) &= \tfrac{1}{22400}(22400x^9 - 33600x^7 + 15120x^5 \\ &\quad - 2280x^3 + 53x). \end{aligned}$$

Because the ROOTS are all REAL for $n \leq 7$ and $n = 9$ only (Hildebrand 1956), these are the only permissible orders for Chebyshev quadrature. The error term is

$$E_n = \begin{cases} c_n \frac{f^{(n+1)}(\xi)}{(n+1)!} & n \text{ odd} \\ c_n \frac{f^{(n+2)}(\xi)}{(n+2)!} & n \text{ even,} \end{cases}$$

where

$$c_n = \begin{cases} \int_{-1}^{1} x G_n(x)\,dx & n \text{ odd} \\ \int_{-1}^{1} x^2 G_n(x)\,dx & n \text{ even.} \end{cases}$$

The first few values of c_n are 2/3, 8/45, 1/15, 32/945, 13/756, and 16/1575 (Hildebrand 1956). Beyer (1987) gives abscissas up to $n = 7$ and Hildebrand (1956) up to $n = 9$.

n	x_i
2	± 0.57735
3	0
	± 0.707107
4	± 0.187592
	± 0.794654
5	0
	± 0.374541
	± 0.832497
6	± 0.266635
	± 0.422519
	± 0.866247
7	0
	± 0.323912
	± 0.529657
	± 0.883862
9	0
	± 0.167906
	± 0.528762
	± 0.601019
	± 0.911589

The ABSCISSAS and weights can be computed analytically for small n.

n	x_i
2	$\pm\frac{1}{3}\sqrt{3}$
3	0
	$\pm\frac{1}{2}\sqrt{2}$
4	$\pm\sqrt{\frac{\sqrt{5}-2}{3\sqrt{5}}}$
	$\pm\sqrt{\frac{\sqrt{5}+2}{3\sqrt{5}}}$
5	0
	$\pm\frac{1}{2}\sqrt{\frac{5-\sqrt{11}}{3}}$
	$\pm\frac{1}{2}\sqrt{\frac{5+\sqrt{11}}{3}}$

see also CHEBYSHEV QUADRATURE, LOBATTO QUADRATURE

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 466, 1987.

Hildebrand, F. B. *Introduction to Numerical Analysis.* New York: McGraw-Hill, pp. 345–351, 1956.

Chebyshev-Radau Quadrature

A GAUSSIAN QUADRATURE-like FORMULA over the interval $[-1, 1]$ which has WEIGHTING FUNCTION $W(x) = x$. The general FORMULA is

$$\int_{-1}^{1} x f(x)\,dx = \sum_{i=1}^{n} w_i[f(x_i) - f(-x_i)].$$

n	x_i	w_i
1	0.7745967	0.4303315
2	0.5002990	0.2393715
	0.8922365	0.2393715
3	0.4429861	0.1599145
	0.7121545	0.1599145
	0.9293066	0.1599145
4	0.3549416	0.1223363
	0.6433097	0.1223363
	0.7783202	0.1223363
	0.9481574	0.1223363

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 466, 1987.

Chebyshev Sum Inequality

If

$$a_1 \geq a_2 \geq \ldots \geq a_n$$

$$b_1 \geq b_2 \geq \ldots \geq b_n,$$

then

$$n\sum_{k=1}^{n} a_k b_k \geq \left(\sum_{k=1}^{n} a_k\right)\left(\sum_{k=1}^{n} b_k\right).$$

This is true for *any* distribution.

see also CAUCHY INEQUALITY, HÖLDER SUM INEQUALITY

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1092, 1979.

Hardy, G. H.; Littlewood, J. E.; and Pólya, G. *Inequalities, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 43–44, 1988.

Chebyshev-Sylvester Constant

In 1891, Chebyshev and Sylvester showed that for sufficiently large x, there exists at least one prime number p satisfying

$$x < p < (1+\alpha)x,$$

where $\alpha = 0.092\ldots$. Since the PRIME NUMBER THEOREM shows the above inequality is true for all $\alpha > 0$ for sufficiently large x, this constant is only of historical interest.

References

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 22, 1983.

Chebyshev's Theorem

see BERTRAND'S POSTULATE

Checker-Jumping Problem

Seeks the minimum number of checkers placed on a board required to allow pieces to move by a sequence of horizontal or vertical jumps (removing the piece jumped over) n rows beyond the forward-most initial checker. The first few cases are 2, 4, 8, 20. It is, however, impossible to reach level 5.

References

Honsberger, R. *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., pp. 23–28, 1976.

Checkerboard

see Chessboard

Checkers

Beeler *et al.* (1972, Item 93) estimated that there are about 10^{12} possible positions. However, this disagrees with the estimate of Jon Schaeffer of 5×10^{20} plausible positions, with 10^{18} reachable under the rules of the game. Because "solving" checkers may require only the Square Root of the number of positions in the search space (i.e., 10^9), so there is hope that some day checkers may be solved (i.e., it may be possible to guarantee a win for the first player to move before the game is even started; Dubuque 1996).

Depending on how they are counted, the number of Eulerian Circuits on an $n \times n$ checkerboard are either 1, 40, 793, 12800, 193721, ... (Sloane's A006240) or 1, 13, 108, 793, 5611, 39312, ... (Sloane's A006239).

see also Checkerboard, Checker-Jumping Problem

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Dubuque, W. "Re: number of legal chess positions." math-fun@cs.arizona.edu posting, Aug 15, 1996.

Kraitchik, M. "Chess and Checkers" and "Checkers (Draughts)." §12.1.1 and 12.1.10 in *Mathematical Recreations.* New York: W. W. Norton, pp. 267–276 and 284–287, 1942.

Schaeffer, J. *One Jump Ahead: Challenging Human Supremacy in Checkers.* New York: Springer-Verlag, 1997.

Sloane, N. J. A. Sequences A006239/M4909 and A006240/M5271 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Checksum

A sum of the digits in a given transmission modulo some number. The simplest form of checksum is a parity bit appended on to 7-bit numbers (e.g., ASCII characters) such that the total number of 1s is always Even ("even parity") or Odd ("odd parity"). A significantly more sophisticated checksum is the Cyclic Redundancy Check (or CRC), which is based on the algebra of polynomials over the integers (mod 2). It is substantially more reliable in detecting transmission errors, and is one common error-checking protocol used in modems.

see also Cyclic Redundancy Check, Error-Correcting Code

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Cyclic Redundancy and Other Checksums." Ch. 20.3 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 888–895, 1992.

Cheeger's Finiteness Theorem

Consider the set of compact n-Riemannian Manifolds M with diameter$(M) \leq d$, Volume$(M) \geq V$, and $|\mathcal{K}| \leq \kappa$ where κ is the Sectional Curvature. Then there is a bound on the number of Diffeomorphisms classes of this set in terms of the constants n, d, V, and κ.

References

Chavel, I. *Riemannian Geometry: A Modern Introduction.* New York: Cambridge University Press, 1994.

Chefalo Knot

A fake Knot created by tying a Square Knot, then looping one end twice through the Knot such that when both ends are pulled, the Knot vanishes.

Chen's Theorem

Every "large" Even Integer may be written as $2n = p + m$ where p is a Prime and $m \in P_2$ is the Set of Semiprimes (i.e., 2-Almost Primes).

see also Almost Prime, Prime Number, Semiprime

References

Rivera, C. "Problems & Puzzles (Conjectures): Chen's Conjecture." `http://www.sci.net.mx/~crivera/ppp/conj_002.htm`.

Chern Class

A Gadget defined for Complex Vector Bundles. The Chern classes of a Complex Manifold are the Chern classes of its Tangent Bundle. The ith Chern class is an Obstruction to the existence of $(n - i + 1)$ everywhere Complex linearly independent Vector Fields on that Vector Bundle. The ith Chern class is in the $(2i)$th cohomology group of the base Space.

see also Obstruction, Pontryagin Class, Stiefel-Whitney Class

Chern Number

The Chern number is defined in terms of the Chern Class of a Manifold as follows. For any collection Chern Classes such that their cup product has the same Dimension as the Manifold, this cup product can be evaluated on the Manifold's Fundamental Class. The resulting number is called the Chern number for that combination of Chern classes. The most important aspect of Chern numbers is that they are Cobordism invariant.

see also Pontryagin Number, Stiefel-Whitney Number

Chernoff Face

A way to display n variables on a 2-D surface. For instance, let x be eyebrow slant, y be eye size, z be nose length, etc.

References

Gonick, L. and Smith, W. *The Cartoon Guide to Statistics.* New York: Harper Perennial, p. 212, 1993.

Chess

Chess is a game played on an 8×8 board, called a CHESSBOARD, of alternating black and white squares. Pieces with different types of allowed moves are placed on the board, a set of black pieces in the first two rows and a set of white pieces in the last two rows. The pieces are called the bishop (2), king (1), knight (2), pawn (8), queen (1), and rook (2). The object of the game is to capture the opponent's king. It is believed that chess was played in India as early as the sixth century AD.

In a game of 40 moves, the number of possible board positions is at least 10^{120} according to Peterson (1996). However, this value does not agree with the 10^{40} possible positions given by Beeler *et al.* (1972, Item 95). This value was obtained by estimating the number of pawn positions (in the no-captures situation, this is 15^8), times all pieces in all positions, dividing by 2 for each of the (rook, knight) which are interchangeable, dividing by 2 for each pair of bishops (since half the positions will have the bishops on the same color squares). There are more positions with one or two captures, since the pawns can then switch columns (Schroeppel 1996). Shannon (1950) gave the value

$$P(40) \approx \frac{64!}{32!(8!)^2(2!)^6} \approx 10^{43}.$$

The number of chess games which end in exactly n plies (including games that mate in fewer than n plies) for $n = 1, 2, 3, \ldots$ are 20, 400, 8902, 197742, 4897256, 119060679, 3195913043, ... (K. Thompson, Sloane's A007545). Rex Stout's fictional detective Nero Wolfe quotes the number of possible games after ten moves as follows: "Wolfe grunted. One hundred and sixty-nine million, five hundred and eighteen thousand, eight hundred and twenty-nine followed by twenty-one ciphers. The number of ways the first ten moves, both sides, may be played" (Stout 1983). The number of chess positions after n moves for $n = 1, 2, \ldots$ are 20, 400, 5362, 71852, 809896?, 9132484?, ... (Schwarzkopf 1994, Sloane's A019319).

Cunningham (1889) incorrectly found 197,299 games and 71,782 positions after the fourth move. C. Flye St. Marie was the first to find the correct number of positions after four moves: 71,852. Dawson (1946) gives the source as *Intermediare des Mathematiques* (1895), but K. Fabel writes that Flye St. Marie corrected the number 71,870 (which he found in 1895) to 71,852 in 1903. The history of the determination of the chess sequences is discussed in Schwarzkopf (1994).

Two problems in recreational mathematics ask

1. How many pieces of a given type can be placed on a CHESSBOARD without any two attacking.
2. What is the smallest number of pieces needed to occupy or attack every square.

The answers are given in the following table (Madachy 1979).

Piece	Max.	Min.
bishops	14	8
kings	16	9
knights	32	12
queens	8	5
rooks	8	8

see also BISHOPS PROBLEM, CHECKERBOARD, CHECKERS, FAIRY CHESS, GO, GOMORY'S THEOREM, HARD HEXAGON ENTROPY CONSTANT, KINGS PROBLEM, KNIGHT'S TOUR, MAGIC TOUR, QUEENS PROBLEM, ROOKS PROBLEM, TOUR

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 124–127, 1987.
Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.
Dawson, T. R. "A Surprise Correction." *The Fairy Chess Review* **6**, 44, 1946.
Dickins, A. "A Guide to Fairy Chess." p. 28, 1967/1969/1971.
Dudeney, H. E. "Chessboard Problems." *Amusements in Mathematics.* New York: Dover, pp. 84–109, 1970.
Fabel, K. "Nüsse." *Die Schwalbe* **84**, 196, 1934.
Fabel, K. "Weihnachtsnüsse." *Die Schwalbe* **190**, 97, 1947.
Fabel, K. "Weihnachtsnüsse." *Die Schwalbe* **195**, 14, 1948.
Fabel, K. "Eröffnungen." *Am Rande des Schachbretts*, 34–35, 1947.
Fabel, K. "Die ersten Schritte." *Rund um das Schachbrett*, 107–109, 1955.
Fabel, K. "Eröffnungen." *Schach und Zahl* **8**, 1966/1971.
Hunter, J. A. H. and Madachy, J. S. *Mathematical Diversions.* New York: Dover, pp. 86–89, 1975.
Kraitchik, M. "Chess and Checkers." §12.1.1 in *Mathematical Recreations.* New York: W. W. Norton, pp. 267–276, 1942.
Madachy, J. S. "Chessboard Placement Problems." Ch. 2 in *Madachy's Mathematical Recreations.* New York: Dover, pp. 34–54, 1979.
Peterson, I. "The Soul of a Chess Machine: Lessons Learned from a Contest Pitting Man Against Computer." *Sci. News* **149**, 200–201, Mar. 30, 1996.
Petković, M. *Mathematics and Chess.* New York: Dover, 1997.
Schroeppel, R. "Reprise: Number of legal chess positions." tech-news@cs.arizona.edu posting, Aug. 18, 1996.
Schwarzkopf, B. "Die ersten Züge." *Problemkiste*, 142–143, No. 92, Apr. 1994.
Shannon, C. "Programming a Computer for Playing Chess." *Phil. Mag.* **41**, 256–275, 1950.
Sloane, N. J. A. Sequences A019319 and A007545/M5100 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Stout, R. "Gambit." In *Seven Complete Nero Wolfe Novels.* New York: Avenic Books, p. 475, 1983.

Chessboard

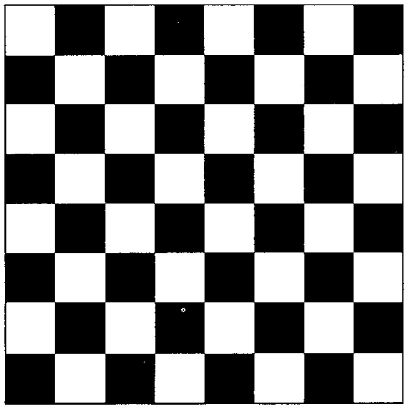

A board containing 8 × 8 squares alternating in color between black and white on which the game of CHESS is played. The checkerboard is identical to the chessboard except that chess's black and white squares are colored red and white in CHECKERS. It is impossible to cover a chessboard from which two opposite corners have been removed with DOMINOES.

see also CHECKERS, CHESS, DOMINO, GOMORY'S THEOREM, WHEAT AND CHESSBOARD PROBLEM

References

Pappas, T. "The Checkerboard." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 136 and 232, 1989.

Chevalley Groups

Finite SIMPLE GROUPS of LIE-TYPE. They include four families of linear SIMPLE GROUPS: $PSL(n,q)$, $PSU(n,q)$, $PSp(2n,q)$, or $P\Omega^\epsilon(n,q)$.

see also TWISTED CHEVALLEY GROUPS

References

Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas#chev`.

Chevalley's Theorem

Let $f(x)$ be a member of a FINITE FIELD $F[x_1, x_2, \ldots, x_n]$ and suppose $f(0,0,\ldots,0) = 0$ and n is greater than the degree of f, then f has at least two zeros in $A^n(F)$.

References

Chevalley, C. "Démonstration d'une hypothèse de M. Artin." *Abhand. Math. Sem. Hamburg* **11**, 73–75, 1936.
Ireland, K. and Rosen, M. "Chevalley's Theorem." §10.2 in *A Classical Introduction to Modern Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 143–144, 1990.

Chevron

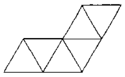

A 6-POLYIAMOND.

References

Golomb, S. W. *Polyominoes: Puzzles, Patterns, Problems, and Packings, 2nd ed.* Princeton, NJ: Princeton University Press, p. 92, 1994.

Chi

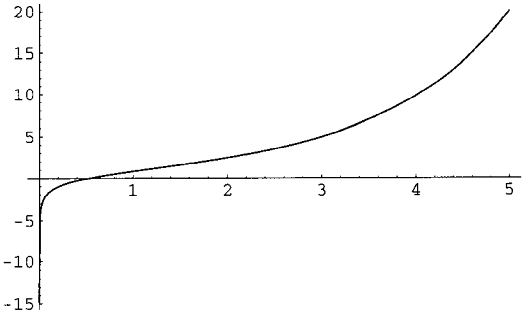

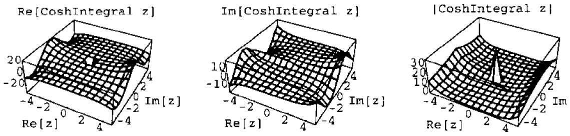

$$\operatorname{Chi}(z) = \gamma + \ln z + \int_0^z \frac{\cosh t - 1}{t}\,dt,$$

where γ is the EULER-MASCHERONI CONSTANT. The function is given by the *Mathematica*® (Wolfram Research, Champaign, IL) command `CoshIntegral[z]`.

see also COSINE INTEGRAL, SHI, SINE INTEGRAL

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Sine and Cosine Integrals." §5.2 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 231–233, 1972.

Chi Distribution

The probability density function and cumulative distribution function are

$$P_n(x) = \frac{2^{1-n/2} x^{n-1} e^{-x^2/2}}{\Gamma(\frac{1}{2}n)} \tag{1}$$

$$D_n(x) = Q(\tfrac{1}{2}n, \tfrac{1}{2}x^2), \tag{2}$$

where Q is the REGULARIZED GAMMA FUNCTION.

$$\mu = \frac{\sqrt{2}\,\Gamma(\frac{1}{2}(n+1))}{\Gamma(\frac{1}{2}n)} \tag{3}$$

$$\sigma^2 = \frac{2[\Gamma(\frac{1}{2}n)\Gamma(1+\frac{1}{2}n) - \Gamma^2(\frac{1}{2}(n+1))]}{\Gamma^2(\frac{1}{2}n)} \tag{4}$$

$$\gamma_1 = \frac{2\Gamma^3(\frac{1}{2}(n+1)) - 3\Gamma(\frac{1}{2}n)\Gamma(\frac{1}{2}(n+1))\Gamma(1+\frac{1}{2}n)}{[\Gamma(\frac{1}{2}n)\Gamma(1+\frac{1}{2}n) - \Gamma^2(\frac{1}{2}(n+1))]^{3/2}}$$

$$+\frac{\Gamma^2(\frac{1}{2}n)\Gamma(\frac{3+n}{2})}{[\Gamma(\frac{1}{2}n)\Gamma(1+\frac{1}{2}n)-\Gamma^2(\frac{1}{2}(n+1))]^{3/2}} \tag{5}$$

$$\gamma_2 = \frac{-3\Gamma^4(\frac{1}{2}(n+1))+6\Gamma(\frac{1}{2}n)+\Gamma^2(\frac{1}{2}(n+1))\Gamma(1+\frac{1}{2}n)}{[\Gamma(\frac{1}{2}n)\Gamma(\frac{2+n}{2})-\Gamma^2(\frac{1}{2}(n+1))]^2}$$
$$+\frac{-4\Gamma^2(\frac{1}{2}n)\Gamma(\frac{1}{2}(n+1))\Gamma(\frac{3+n}{2})+\Gamma^3(\frac{1}{2}n)\Gamma(\frac{4+n}{2})}{[\Gamma(\frac{1}{2}n)\Gamma(\frac{2+n}{2})-\Gamma^2(\frac{1}{2}(n+1))]^2}, \tag{6}$$

where μ is the Mean, σ^2 the Variance, γ_1 the Skewness, and γ_2 the Kurtosis. For $n = 1$, the χ distribution is a Half-Normal Distribution with $\theta = 1$. For $n = 2$, it is a Rayleigh Distribution with $\sigma = 1$.

see also Chi-Squared Distribution, Half-Normal Distribution, Rayleigh Distribution

Chi Inequality

The inequality

$$(j+1)a_j + a_i \geq (j+1)i,$$

which is satisfied by all A-Sequences.

References

Levine, E. and O'Sullivan, J. "An Upper Estimate for the Reciprocal Sum of a Sum-Free Sequence." *Acta Arith.* **34**, 9–24, 1977.

Chi-Squared Distribution

A χ^2 distribution is a Gamma Distribution with $\theta \equiv 2$ and $\alpha \equiv r/2$, where r is the number of Degrees of Freedom. If Y_i have Normal Independent distributions with Mean 0 and Variance 1, then

$$\chi^2 \equiv \sum_{i=1}^{n} {Y_i}^2 \tag{1}$$

is distributed as χ^2 with n Degrees of Freedom. If ${\chi_i}^2$ are independently distributed according to a χ^2 distribution with n_1, n_2, …, n_k Degrees of Freedom, then

$$\sum_{j=1}^{k} {\chi_j}^2 \tag{2}$$

is distributed according to χ^2 with $n \equiv \sum_{j=1}^{n} n_j$ Degrees of Freedom.

$$P_n(x) = \begin{cases} \frac{x^{r/2-1}e^{-x/2}}{\Gamma(\frac{1}{2}r)2^{r/2}} & \text{for } 0 \leq x < \infty \\ 0 & \text{for } x < 0. \end{cases} \tag{3}$$

The cumulative distribution function is then

$$D_n(\chi^2) = \int_0^{\chi^2} \frac{t^{r/2-1}e^{-t/2}\,dt}{\Gamma(\frac{1}{2}r)2^{r/2}}$$
$$= \frac{\gamma(\frac{1}{2}n, \frac{1}{2}\chi^2)}{\Gamma(\frac{1}{2}n)} = P(\tfrac{1}{2}n, \tfrac{1}{2}\chi^2), \tag{4}$$

where $P(a,z)$ is a Regularized Gamma Function. The Confidence Intervals can be found by finding the value of x for which $D_n(x)$ equals a given value. The Moment-Generating Function of the χ^2 distribution is

$$M(t) = (1-2t)^{-r/2} \tag{5}$$
$$R(t) \equiv \ln M(t) = -\tfrac{1}{2}r\ln(1-2t) \tag{6}$$
$$R'(t) = \frac{r}{1-2t} \tag{7}$$
$$R''(t) = \frac{2r}{(1-2t)^2}, \tag{8}$$

so

$$\mu = R'(0) = r \tag{9}$$
$$\sigma^2 = R''(0) = 2r \tag{10}$$
$$\gamma_1 = 2\sqrt{\frac{2}{r}} \tag{11}$$
$$\gamma_2 = \frac{12}{r}. \tag{12}$$

The nth Moment about zero for a distribution with n Degrees of Freedom is

$$m'_n = 2^n \frac{\Gamma(n+\frac{1}{2}r)}{\Gamma(\frac{1}{2}r)} = r(r+2)\cdots(r+2n-2), \tag{13}$$

and the moments about the Mean are

$$\mu_2 = 2r \tag{14}$$
$$\mu_3 = 8r \tag{15}$$
$$\mu_4 = 12n^2 + 48n. \tag{16}$$

The nth Cumulant is

$$\kappa_n = 2^n\Gamma(n)(\tfrac{1}{2}r) = 2^{n-1}(n-1)!r. \tag{17}$$

The Moment-Generating Function is

$$M(t) = e^{rt/\sqrt{2r}}\left(1-\frac{2t}{\sqrt{2r}}\right)^{-r/2}$$
$$= \left[e^{t\sqrt{2/r}}\left(1-\sqrt{\frac{2}{r}}\,t\right)\right]^{-r/2}$$
$$= \left[1-\frac{t^2}{r}-\frac{1}{3}\left(\frac{2}{r}\right)^{3/2}t^3-\ldots\right]^{-r/2}. \tag{18}$$

As $r \to \infty$,

$$\lim_{r\to\infty} M(t) = e^{t^2/2}, \tag{19}$$

so for large r,

$$\sqrt{2\chi^2} = \sqrt{\sum_i \frac{(x_i-\mu_i)^2}{{\sigma_i}^2}} \tag{20}$$

is approximately a GAUSSIAN DISTRIBUTION with MEAN $\sqrt{2r}$ and VARIANCE $\sigma^2 = 1$. Fisher showed that

$$\frac{\chi^2 - r}{\sqrt{2r-1}} \tag{21}$$

is an improved estimate for moderate r. Wilson and Hilferty showed that

$$\left(\frac{\chi^2}{r}\right)^{1/3} \tag{22}$$

is a nearly GAUSSIAN DISTRIBUTION with MEAN $\mu = 1 - 2/(9r)$ and VARIANCE $\sigma^2 = 2/(9r)$.

In a GAUSSIAN DISTRIBUTION,

$$P(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}\,dx, \tag{23}$$

let

$$z \equiv (x-\mu)^2/\sigma^2. \tag{24}$$

Then

$$dz = \frac{2(x-\mu)}{\sigma^2}\,dx = \frac{2\sqrt{z}}{\sigma}\,dx \tag{25}$$

so

$$dx = \frac{\sigma}{2\sqrt{z}}\,dz. \tag{26}$$

But

$$P(z)\,dz = 2P(x)\,dx, \tag{27}$$

so

$$P(x)\,dx = 2\frac{1}{\sigma\sqrt{2\pi}} e^{-z/2}\,dz = \frac{1}{\sigma\sqrt{\pi}} e^{-z/2}\,dz. \tag{28}$$

This is a χ^2 distribution with $r = 1$, since

$$P(z)\,dz = \frac{z^{1/2-1}e^{-z/2}}{\Gamma(\frac{1}{2})2^{1/2}}\,dz = \frac{x^{-1/2}e^{-1/2}}{\sqrt{2\pi}}\,dz. \tag{29}$$

If X_i are independent variates with a NORMAL DISTRIBUTION having MEANS μ_i and VARIANCES ${\sigma_i}^2$ for $i = 1, \ldots, n$, then

$$\tfrac{1}{2}\chi^2 \equiv \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{2{\sigma_i}^2} \tag{30}$$

is a GAMMA DISTRIBUTION variate with $\alpha = n/2$,

$$P(\tfrac{1}{2}\chi^2)\,d(\tfrac{1}{2}\chi^2) = \frac{1}{\Gamma(\frac{1}{2}n)} e^{-\chi^2/2}(\tfrac{1}{2}\chi^2)^{(n/2)-1}\,d(\tfrac{1}{2}\chi^2). \tag{31}$$

The noncentral chi-squared distribution is given by

$$P(x) = 2^{-n/2}e^{-(\lambda+x)/2}x^{n/2-1}F(\tfrac{1}{2}n, \tfrac{1}{4}\lambda x), \tag{32}$$

where

$$F(a,z) \equiv \frac{{}_0F_1(;a;z)}{\Gamma(a)}, \tag{33}$$

${}_0F_1$ is the CONFLUENT HYPERGEOMETRIC LIMIT FUNCTION and Γ is the GAMMA FUNCTION. The MEAN, VARIANCE, SKEWNESS, and KURTOSIS are

$$\mu = \lambda + n \tag{34}$$

$$\sigma^2 = 2(2\lambda + n) \tag{35}$$

$$\gamma_1 = \frac{2\sqrt{2}\,(3\lambda + n)}{(2\lambda + n)^{3/2}} \tag{36}$$

$$\gamma_2 = \frac{12(4\lambda + n)}{(2\lambda + n)^2}. \tag{37}$$

see also CHI DISTRIBUTION, SNEDECOR'S F-DISTRIBUTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 940–943, 1972.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 535, 1987.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Incomplete Gamma Function, Error Function, Chi-Square Probability Function, Cumulative Poisson Function." §6.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 209–214, 1992.

Spiegel, M. R. *Theory and Problems of Probability and Statistics.* New York: McGraw-Hill, pp. 115–116, 1992.

Chi-Squared Test

Let the probabilities of various classes in a distribution be $p_1, p_2, \ldots, p_k$. The expected frequency

$${\chi_s}^2 = \sum_{i=1}^{k} \frac{(m_i - Np_i)^2}{Np_i}$$

is a measure of the deviation of a sample from expectation. Karl Pearson proved that the limiting distribution of ${\chi_s}^2$ is χ^2 (Kenney and Keeping 1951, pp. 114–116).

$$\begin{aligned}
\Pr(\chi^2 \geq {\chi_s}^2) &= \int_{{\chi_s}^2}^{\infty} f(\chi^2)\,d(\chi^2) \\
&= \frac{1}{2}\int_{{\chi_s}^2}^{\infty} \frac{\left(\frac{\chi^2}{2}\right)^{(k-3)/2}}{\Gamma\left(\frac{k-1}{2}\right)} e^{-\chi^2/2}\,d(\chi^2) \\
&= 1 - \frac{\Gamma\left(\frac{1}{2}{\chi_s}^2, \frac{k-1}{2}\right)}{\Gamma\left(\frac{k-1}{2}\right)} \\
&= 1 - I\left(\frac{{\chi_s}^2}{\sqrt{2(k-1)}}, \frac{k-3}{2}\right),
\end{aligned}$$

where $I(x,n)$ is PEARSON'S FUNCTION. There are some subtleties involved in using the χ^2 test to fit curves (Kenney and Keeping 1951, pp. 118–119).

When fitting a one-parameter solution using χ^2, the best-fit parameter value can be found by calculating χ^2

at three points, plotting against the parameter values of these points, then finding the minimum of a Parabola fit through the points (Cuzzi 1972, pp. 162–168).

References

Cuzzi, J. *The Subsurface Nature of Mercury and Mars from Thermal Microwave Emission.* Ph.D. Thesis. Pasadena, CA: California Institute of Technology, 1972.

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, 1951.

Child

A node which is one Edge further away from a given Edge in a Rooted Tree.

see also Root (Tree), Rooted Tree, Sibling

Chinese Hypothesis

A Prime p always satisfies the condition that $2^p - 2$ is divisible by p. However, this condition is not true *exclusively* for Prime (e.g., $2^{341} - 2$ is divisible by $341 = 11 \cdot 31$). Composite Numbers n (such as 341) for which $2^n - 2$ is divisible by n are called Poulet Numbers, and are a special class of Fermat Pseudoprimes. The Chinese hypothesis is a special case of Fermat's Little Theorem.

see also Carmichael Number, Euler's Theorem, Fermat's Little Theorem, Fermat Pseudoprime, Poulet Number, Pseudoprime

References

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 19–20, 1993.

Chinese Remainder Theorem

Let r and s be Positive Integers which are Relatively Prime and let a and b be any two Integers. Then there is an Integer N such that

$$N \equiv a \pmod{r} \tag{1}$$

and

$$N \equiv b \pmod{s}. \tag{2}$$

Moreover, N is uniquely determined modulo rs. An equivalent statement is that if $(r, s) = 1$, then every pair of Residue Classes modulo r and s corresponds to a simple Residue Class modulo rs.

The theorem can also be generalized as follows. Given a set of simultaneous Congruences

$$x \equiv a_i \pmod{m_i} \tag{3}$$

for $i = 1, \ldots, r$ and for which the m_i are pairwise Relatively Prime, the solution of the set of Congruences is

$$x = a_1 b_1 \frac{M}{m_1} + \ldots + a_r b_r \frac{M}{m_r} \pmod{M}, \tag{4}$$

where

$$M \equiv m_1 m_2 \cdots m_r \tag{5}$$

and the b_i are determined from

$$b_i \frac{M}{m_i} \equiv 1 \pmod{m_i}. \tag{6}$$

References

Ireland, K. and Rosen, M. "The Chinese Remainder Theorem." §3.4 in *A Classical Introduction to Modern Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 34–38, 1990.

Uspensky, J. V. and Heaslet, M. A. *Elementary Number Theory.* New York: McGraw-Hill, pp. 189–191, 1939.

Wagon, S. "The Chinese Remainder Theorem." §8.4 in *Mathematica in Action.* New York: W. H. Freeman, pp. 260–263, 1991.

Chinese Rings

see Baguenaudier

Chiral

Having forms of different Handedness which are not mirror-symmetric.

see also Disymmetric, Enantiomer, Handedness, Mirror Image, Reflexible

Choice Axiom

see Axiom of Choice

Choice Number

see Combination

Cholesky Decomposition

Given a symmetric Positive Definite Matrix A, the Cholesky decomposition is an upper Triangular Matrix U such that

$$\mathsf{A} = \mathsf{U}^\mathsf{T}\mathsf{U}.$$

see also LU Decomposition, QR Decomposition

References

Nash, J. C. "The Choleski Decomposition." Ch. 7 in *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation, 2nd ed.* Bristol, England: Adam Hilger, pp. 84–93, 1990.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Cholesky Decomposition." §2.9 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 89–91, 1992.

Choose

An alternative term for a Binomial Coefficient, in which $\binom{n}{k}$ is read as "n choose k." R. K. Guy suggested this pronunciation around 1950, when the notations ${}^nC_r$ and ${}_nC_r$ were commonly used. Leo Moser liked the pronunciation and he and others spread it around. It got the final seal of approval from Donald Knuth when he incorporated it into the TeX mathematical typesetting language as `{n\choose k}`.

Choquet Theory

Erdős proved that there exist at least one Prime of the form $4k+1$ and at least one Prime of the form $4k+3$ between n and $2n$ for all $n>6$.

see also Equinumerous, Prime Number

Chord

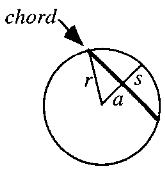

The Line Segment joining two points on a curve. The term is often used to describe a Line Segment whose ends lie on a Circle. In the above figure, r is the Radius of the Circle, a is called the Apothem, and s the Sagitta.

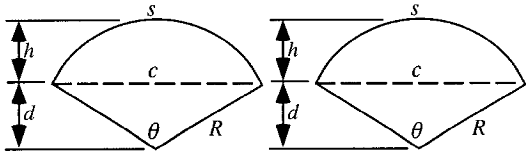

The shaded region in the left figure is called a Sector, and the shaded region in the right figure is called a Segment.

All Angles inscribed in a Circle and subtended by the same chord are equal. The converse is also true: The Locus of all points from which a given segment subtends equal Angles is a Circle.

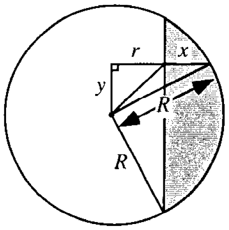

Let a Circle of Radius R have a Chord at distance r. The Area enclosed by the Chord, shown as the shaded region in the above figure, is then

$$A = 2\int_0^{\sqrt{R^2-r^2}} x(y)\,dy. \tag{1}$$

But

$$y^2 + (r+x)^2 = R^2, \tag{2}$$

so

$$x(y) = \sqrt{R^2-y^2} - r \tag{3}$$

and

$$\begin{aligned} A &= 2\int_0^{\sqrt{R^2-r^2}} (\sqrt{R^2-y^2} - r)\,dy \\ &= \left[y\sqrt{R^2-y^2} + R^2 \tan^{-1}\left(\frac{y}{\sqrt{R^2-y^2}}\right) \right. \\ &\qquad \left. - 2ry \vphantom{\frac{y}{\sqrt{R^2}}}\right]_0^{\sqrt{R^2-r^2}} \\ &= r\sqrt{R^2-r^2} + R^2\tan^{-1}\left[\left(\frac{R}{r}\right)^2 - 1\right] - 2r\sqrt{R^2-r^2} \\ &= R^2\tan^{-1}\left[\left(\frac{R}{r}\right)^2 - 1\right] - r\sqrt{R^2-r^2}. \end{aligned} \tag{4}$$

Checking the limits, when $r=R$, $A=0$ and when $r\to 0$,

$$A = \tfrac{1}{2}\pi R^2, \tag{5}$$

see also Annulus, Apothem, Bertrand's Problem, Concentric Circles, Radius, Sagitta, Sector, Segment

Chordal

see Radical Axis

Chordal Theorem

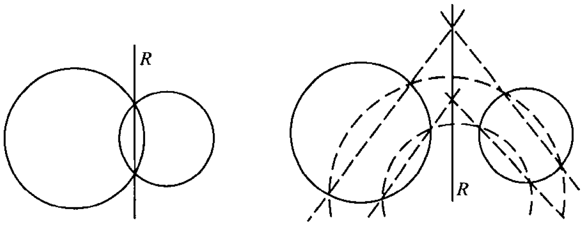

The Locus of the point at which two given Circles possess the same Power is a straight line Perpendicular to the line joining the Midpoints of the Circle and is known as the chordal (or Radical Axis) of the two Circles.

References

Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, p. 153, 1965.

Chow Coordinates

A generalization of Grassmann Coordinates to m-D varieties of degree d in P^n, where P^n is an n-D projective space. To define the Chow coordinates, take the intersection of a m-D Variety Z of degree d by an $(n-m)$-D Subspace U of P^n. Then the coordinates of the d points of intersection are algebraic functions of the Grassmann Coordinates of U, and by taking a symmetric function of the algebraic functions, a hHomogeneous Polynomial known as the Chow form of Z is obtained. The Chow coordinates are then the COEFFICIENTS of the Chow form. Chow coordinates can generate the smallest field of definition of a divisor.

References
Chow, W.-L. and van der Waerden., B. L. "Zur algebraische Geometrie IX." *Math. Ann.* **113**, 692–704, 1937.
Wilson, W. S.; Chern, S. S.; Abhyankar, S. S.; Lang, S.; and Igusa, J.-I. "Wei-Liang Chow." *Not. Amer. Math. Soc.* **43**, 1117–1124, 1996.

Chow Ring

The intersection product for classes of rational equivalence between cycles on an ALGEBRAIC VARIETY.

References
Chow, W.-L. "On Equivalence Classes of Cycles in an Algebraic Variety." *Ann. Math.* **64**, 450–479, 1956.
Wilson, W. S.; Chern, S. S.; Abhyankar, S. S.; Lang, S.; and Igusa, J.-I. "Wei-Liang Chow." *Not. Amer. Math. Soc.* **43**, 1117–1124, 1996.

Chow Variety

The set $C_{n,m,d}$ of all m-D varieties of degree d in an n-D projective space P^n into an M-D projective space P^M.

References
Wilson, W. S.; Chern, S. S.; Abhyankar, S. S.; Lang, S.; and Igusa, J.-I. "Wei-Liang Chow." *Not. Amer. Math. Soc.* **43**, 1117–1124, 1996.

Christoffel-Darboux Formula

For three consecutive ORTHOGONAL POLYNOMIALS

$$p_n(x) = (A_n x + B_n)p_{n-1}x - C_n p_{n-2}(x) \qquad (1)$$

for $n = 2, 3, \ldots$, where $A_n > 0$, B_n, and $C_n > 0$ are constants. Denoting the highest COEFFICIENT of $p_n(x)$ by k_n,

$$A_n = \frac{k_n}{k_{n-1}} \qquad (2)$$

$$C_n = \frac{A_n}{A_{n-1}} = \frac{k_n k_{n-2}}{{k_{n-1}}^2}. \qquad (3)$$

Then

$$p_0(x)p_0(y) + \ldots + p_n(x)p_n(y) = \frac{k_n}{k_{n+1}} \frac{p_{n+1}(x)p_n(y) - p_n(x)p_{n+1}(y)}{x - y}. \qquad (4)$$

In the special case of $x = y$, (4) gives

$$[p_0(x)]^2 + \ldots + [p_n(x)]^2 = \frac{k_n}{k_{n+1}}[p'_{n+1}(x)p_n(x) - p'_n(x)p_{n+1}(x)]. \qquad (5)$$

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 785, 1972.
Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., pp. 42–44, 1975.

Christoffel-Darboux Identity

$$\sum_{k=0}^{\infty} \frac{\phi_k(x)\phi_k(y)}{\gamma_k} = \frac{\phi_{m+1}(x)\phi_m(y) - \phi_m(x)\phi_{m+1}(y)}{a_m \gamma_m (x - y)}, \qquad (1)$$

where $\phi_k(x)$ are ORTHOGONAL POLYNOMIALS with WEIGHTING FUNCTION $W(x)$,

$$\gamma_m \equiv \int [\phi_m(x)]^2 W(x)\, dx, \qquad (2)$$

and

$$a_k \equiv \frac{A_{k+1}}{A_k} \qquad (3)$$

where A_k is the COEFFICIENT of x^k in $\phi_k(x)$.

References
Hildebrand, F. B. *Introduction to Numerical Analysis.* New York: McGraw-Hill, p. 322, 1956.

Christoffel Formula

Let $\{p_n(x)\}$ be orthogonal POLYNOMIALS associated with the distribution $d\alpha(x)$ on the interval $[a,b]$. Also let

$$\rho \equiv c(x - x_1)(x - x_2)\cdots(x - x_l)$$

(for $c \neq 0$) be a POLYNOMIAL of order l which is NONNEGATIVE in this interval. Then the orthogonal POLYNOMIALS $\{q(x)\}$ associated with the distribution $\rho(x)\, d\alpha(x)$ can be represented in terms of the POLYNOMIALS $p_n(x)$ as

$$\rho(x)q_n(x) = \begin{vmatrix} p_n(x) & p_{n+1}(x) & \cdots & p_{n+l}(x) \\ p_n(x_1) & p_{n+1}(x_1) & \cdots & p_{n+l}(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ p_n(x_l) & p_{n+1}(x_l) & \cdots & p_{n+l}(x_l) \end{vmatrix}.$$

In the case of a zero x_k of multiplicity $m > 1$, we replace the corresponding rows by the derivatives of order 0, 1, 2, ..., $m - 1$ of the POLYNOMIALS $p_n(x_l)$, ..., $p_{n+l}(x_l)$ at $x = x_k$.

References
Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., pp. 29–30, 1975.

Christoffel Number

One of the quantities λ_i appearing in the GAUSS-JACOBI MECHANICAL QUADRATURE. They satisfy

$$\lambda_1 + \lambda_2 + \ldots + \lambda_n = \int_a^b d\alpha(x) = \alpha(b) - \alpha(a) \qquad (1)$$

and are given by

$$\lambda_\nu = \int_a^b \left[\frac{p_n(x)}{p'_n(x_\nu)(x-x_\nu)}\right]^2 d\alpha(x) \tag{2}$$

$$\lambda_\nu = -\frac{k_{n+1}}{k_n}\frac{1}{p_{n+1}(x_\nu)p'_n(x_\nu)} \tag{3}$$

$$= \frac{k_n}{k_{n-1}}\frac{1}{p_{n-1}(x_\nu)P'_n(x_\nu)} \tag{4}$$

$$(\lambda_\nu)^{-1} = [p_0(x_\nu)]^2 + \ldots + [p_n(x_\nu)]^2, \tag{5}$$

where k_n is the higher COEFFICIENT of $p_n(x)$.

References

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., pp. 47–48, 1975.

Christoffel Symbol of the First Kind

Variously denoted $[ij,k]$, $\left[\begin{smallmatrix} i & & j \\ & k & \end{smallmatrix}\right]$, Γ_{abc}, or $\{ab,c\}$.

$$[ij,k] = \begin{bmatrix} i & & j \\ & k & \end{bmatrix} \equiv g_{mk}\Gamma^m_{ij} = g_{mk}\vec{e}^{\,m} \cdot \frac{\partial \vec{e}_i}{\partial q^j} = \vec{e}_k \cdot \frac{\partial \vec{e}_i}{\partial q^j}, \tag{1}$$

where g_{mk} is the METRIC TENSOR and

$$\vec{e}_i \equiv \frac{\partial \vec{r}}{\partial q^i} = h_i \hat{e}_i. \tag{2}$$

But

$$\frac{\partial g_{ij}}{\partial q^k} = \frac{\partial}{\partial q^k}(\vec{e}_i \cdot \vec{e}_j) = \frac{\partial \vec{e}_i}{\partial q^k} \cdot \vec{e}_j + \vec{e}_i \cdot \frac{\partial \vec{e}_j}{\partial q^k}$$

$$= [ik,j] + [jk,i], \tag{3}$$

so

$$[ab,c] = \tfrac{1}{2}(g_{ac,b} + g_{bc,a} - g_{ab,c}). \tag{4}$$

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 160–167, 1985.

Christoffel Symbol of the Second Kind

Variously denoted $\left\{\begin{smallmatrix} & m & \\ i & & j \end{smallmatrix}\right\}$ or Γ^m_{ij}.

$$\left\{\begin{matrix} & m & \\ i & & j \end{matrix}\right\} \equiv \Gamma^m_{ij} = \vec{e}^{\,m} \cdot \frac{\partial \vec{e}_i}{\partial q^j} = g^{km}[ij,k]$$

$$= \frac{1}{2}g^{km}\left(\frac{\partial g_{ik}}{\partial q^j} + \frac{\partial g_{jk}}{\partial q^i} - \frac{\partial g_{ij}}{\partial q^k}\right), \tag{1}$$

where Γ^m_{ij} is a CONNECTION COEFFICIENT and $\{bc,d\}$ is a CHRISTOFFEL SYMBOL OF THE FIRST KIND.

$$\left\{\begin{matrix} & a & \\ b & & c \end{matrix}\right\} = g_{ad}\{bc,d\}. \tag{2}$$

The Christoffel symbols are given in terms of the first FUNDAMENTAL FORM E, F, and G by

$$\Gamma^1_{11} = \frac{GE_u - 2FF_u + FE_v}{2(EG-F^2)} \tag{3}$$

$$\Gamma^1_{12} = \frac{GE_v - FG_u}{2(EG-F^2)} \tag{4}$$

$$\Gamma^1_{22} = \frac{2GF_v - GG_u - FG_v}{2(EG-F^2)} \tag{5}$$

$$\Gamma^2_{11} = \frac{2EF_u - EE_v - FE_u}{2(EG-F^2)} \tag{6}$$

$$\Gamma^2_{12} = \frac{EG_u - FE_v}{2(EG-F^2)} \tag{7}$$

$$\Gamma^2_{22} = \frac{EG_v - 2FF_v + FG_u}{2(EG-F^2)}, \tag{8}$$

and $\Gamma^1_{21} = \Gamma^1_{12}$ and $\Gamma^2_{21} = \Gamma^2_{12}$. If $F = 0$, the Christoffel symbols of the second kind simplify to

$$\Gamma^1_{11} = \frac{E_u}{2E} \tag{9}$$

$$\Gamma^1_{12} = \frac{E_v}{2E} \tag{10}$$

$$\Gamma^1_{22} = -\frac{G_u}{2E} \tag{11}$$

$$\Gamma^2_{11} = -\frac{E_v}{2G} \tag{12}$$

$$\Gamma^2_{12} = \frac{G_u}{2G} \tag{13}$$

$$\Gamma^2_{22} = \frac{G_v}{2G} \tag{14}$$

(Gray 1993).

The following relationships hold between the Christoffel symbols of the second kind and coefficients of the first FUNDAMENTAL FORM,

$$\Gamma^1_{11}E + \Gamma^2_{11}F = \tfrac{1}{2}E_u \tag{15}$$

$$\Gamma^1_{12}E + \Gamma^2_{12}F = \tfrac{1}{2}E_v \tag{16}$$

$$\Gamma^1_{22}E + \Gamma^2_{22}F = F_v - \tfrac{1}{2}G_u \tag{17}$$

$$\Gamma^1_{11}F + \Gamma^2_{11}G = F_u - \tfrac{1}{2}E_v \tag{18}$$

$$\Gamma^1_{12}F + \Gamma^2_{12}G = \tfrac{1}{2}G_u \tag{19}$$

$$\Gamma^1_{22}F + \Gamma^2_{22}G = \tfrac{1}{2}G_v \tag{20}$$

$$\Gamma^1_{11} + \Gamma^2_{12} = (\ln\sqrt{EG-F^2}\,)_u \tag{21}$$

$$\Gamma^1_{12} + \Gamma^2_{22} = (\ln\sqrt{EG-F^2}\,)_v \tag{22}$$

(Gray 1993).

For a surface given in MONGE'S FORM $z = F(x,y)$,

$$\Gamma^k_{ij} = \frac{z_{ij}z_k}{1 + z_1{}^2 + z_2{}^2}. \tag{23}$$

see also CHRISTOFFEL SYMBOL OF THE FIRST KIND, CONNECTION COEFFICIENT, GAUSS EQUATIONS

References
Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 160–167, 1985.
Gray, A. "Christoffel Symbols." §20.3 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 397–400, 1993.
Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 47–48, 1953.

Chromatic Number

The fewest number of colors $\gamma(G)$ necessary to color a GRAPH or surface. The chromatic number of a surface of GENUS g is given by the HEAWOOD CONJECTURE,

$$\gamma(g) = \left\lfloor \tfrac{1}{2}(7+\sqrt{48g+1})\right\rfloor,$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION. $\gamma(g)$ is sometimes also denoted $\chi(g)$. For $g = 0, 1, \ldots$, the first few values of $\chi(g)$ are 4, 7, 8, 9, 10, 11, 12, 12, 13, 13, 14, 15, 15, 16, ... (Sloane's A000934).

The fewest number of colors necessary to color each EDGE of a GRAPH so that no two EDGES incident on the same VERTEX have the same color is called the "EDGE chromatic number."

see also BRELAZ'S HEURISTIC ALGORITHM, CHROMATIC POLYNOMIAL, EDGE-COLORING, EULER CHARACTERISTIC, HEAWOOD CONJECTURE, MAP COLORING, TORUS COLORING

References
Chartrand, G. "A Scheduling Problem: An Introduction to Chromatic Numbers." §9.2 in *Introductory Graph Theory.* New York: Dover, pp. 202–209, 1985.
Eppstein, D. "The Chromatic Number of the Plane." `http://www.ics.uci.edu/~eppstein/junkyard/plane-color/`.
Sloane, N. J. A. Sequence A000934/M3292 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Chromatic Polynomial

A POLYNOMIAL $P(z)$ of a graph g which counts the number of ways to color g with exactly z colors. Tutte (1970) showed that the chromatic POLYNOMIALS of planar triangular graphs possess a ROOT close to $\phi^2 = 2.618033\ldots$, where ϕ is the GOLDEN MEAN. More precisely, if n is the number of VERTICES of G, then

$$P_G(\phi^2) \leq \phi^{5-n}$$

(Le Lionnais 1983).

References
Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 46, 1983.
Tutte, W. T. "On Chromatic Polynomials and the Golden Ratio." *J. Combin. Th.* **9**, 289–296, 1970.

Chu Identity

see CHU-VANDERMONDE IDENTITY

Chu Space

A Chu space is a binary relation from a SET A to an antiset X which is defined as a SET which transforms via converse functions.

References
Stanford Concurrency Group. "Guide to Papers on Chu Spaces." `http://boole.stanford.edu/chuguide.html`.

Chu-Vandermonde Identity

$$(x+a)_n = \sum_{k=0}^{\infty} \binom{n}{k} (a)_k (x)_{n-k}$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT and $(a)_n \equiv a(a-1)\cdots(a-n+1)$ is the POCHHAMMER SYMBOL. A special case gives the identity

$$\sum_{l=0}^{\max(k,n)} \binom{m}{k-l}\binom{n}{l} = \binom{m+n}{k}.$$

see also BINOMIAL THEOREM, UMBRAL CALCULUS

References
Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, pp. 130 and 181–182, 1996.

Church's Theorem

No decision procedure exists for ARITHMETIC.

Church's Thesis

see CHURCH-TURING THESIS

Church-Turing Thesis

The TURING MACHINE concept defines what is meant mathematically by an algorithmic procedure. Stated another way, a function f is effectively COMPUTABLE IFF it can be computed by a TURING MACHINE.

see also ALGORITHM, COMPUTABLE FUNCTION, TURING MACHINE

References
Penrose, R. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics.* Oxford, England: Oxford University Press, pp. 47–49, 1989.

Chvátal's Art Gallery Theorem

see ART GALLERY THEOREM

Chvátal's Theorem

Let the GRAPH G have VERTICES with VALENCES $d_1 \leq \ldots \leq d_m$. If for every $i < n/2$ we have either $d_i \geq i+1$ or $d_{n-i} \geq n-i$, then the GRAPH is HAMILTONIAN.

ci

see COSINE INTEGRAL

Ci

see COSINE INTEGRAL

Cigarettes

It is possible to place 7 cigarettes in such a way that each touches the other if $l/d > 7\sqrt{3}/2$ (Gardner 1959, p. 115).

References

Gardner, M. *The Scientific American Book of Mathematical Puzzles & Diversions.* New York: Simon and Schuster, 1959.

Cin

see COSINE INTEGRAL

Circle

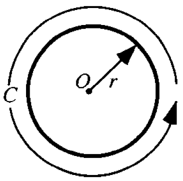

A circle is the set of points equidistant from a given point O. The distance r from the CENTER is called the RADIUS, and the point O is called the CENTER. Twice the RADIUS is known as the DIAMETER $d = 2r$. The PERIMETER C of a circle is called the CIRCUMFERENCE, and is given by

$$C = \pi d = 2\pi r. \tag{1}$$

The circle is a CONIC SECTION obtained by the intersection of a CONE with a PLANE PERPENDICULAR to the CONE's symmetry axis. A circle is the degenerate case of an ELLIPSE with equal semimajor and semiminor axes (i.e., with ECCENTRICITY 0). The interior of a circle is called a DISK. The generalization of a circle to 3-D is called a SPHERE, and to n-D for $n \geq 4$ a HYPERSPHERE.

The region of intersection of two circles is called a LENS. The region of intersection of three symmetrically placed circles (as in a VENN DIAGRAM), in the special case of the center of each being located at the intersection of the other two, is called a REULEAUX TRIANGLE.

The parametric equations for a circle of RADIUS a are

$$x = a\cos t \tag{2}$$
$$y = a\sin t. \tag{3}$$

For a body moving uniformly around the circle,

$$x' = -a\sin t \tag{4}$$
$$y' = a\cos t, \tag{5}$$

and

$$x'' = -a\cos t \tag{6}$$
$$y'' = -a\sin t. \tag{7}$$

When normalized, the former gives the equation for the unit TANGENT VECTOR of the circle, $(-\sin t, \cos t)$. The circle can also be parameterized by the rational functions

$$x = \frac{1-t^2}{t(1+t)} \tag{8}$$
$$y = \frac{2t}{1+t^2}, \tag{9}$$

but an ELLIPTIC CURVE cannot. The following plots show a sequence of NORMAL and TANGENT VECTORS for the circle.

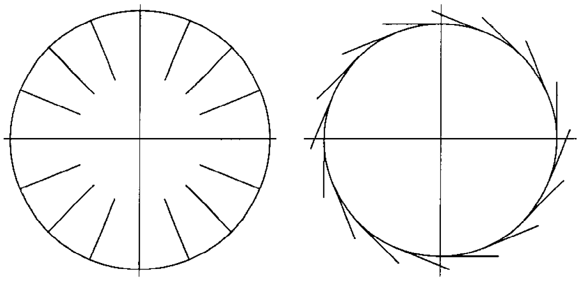

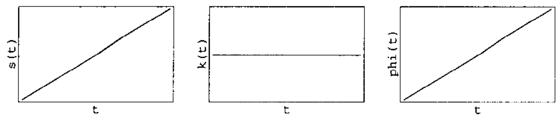

The ARC LENGTH s, CURVATURE κ, and TANGENTIAL ANGLE ϕ of the circle are

$$s(t) = \int ds = \int \sqrt{x'^2 + y'^2}\, dt = at \tag{10}$$
$$\kappa(t) = \frac{x'y'' - y'x''}{(x'^2 + y'^2)^{3/2}} = \frac{1}{a} \tag{11}$$
$$\phi(t) = \int \kappa(t)\, dt = \frac{t}{a}. \tag{12}$$

The CESÀRO EQUATION is

$$\kappa = \frac{1}{a}. \tag{13}$$

In POLAR COORDINATES, the equation of the circle has a particularly simple form.

$$r = a \tag{14}$$

is a circle of RADIUS a centered at ORIGIN,

$$r = 2a\cos\theta \tag{15}$$

is circle of RADIUS a centered at $(a, 0)$, and

$$r = 2a\sin\theta \tag{16}$$

is a circle of RADIUS a centered on $(0, a)$. In CARTESIAN COORDINATES, the equation of a circle of RADIUS a centered on (x_0, y_0) is

$$(x - x_0)^2 + (y - y_0)^2 = a^2. \qquad (17)$$

In PEDAL COORDINATES with the PEDAL POINT at the center, the equation is

$$pa = r^2. \qquad (18)$$

The circle having P_1P_2 as a diameter is given by

$$(x - x_1)(x - x_2) + (y - y_1)(y - y_2) = 0. \qquad (19)$$

The equation of a circle passing through the three points (x_i, y_i) for $i = 1, 2, 3$ (the CIRCUMCIRCLE of the TRIANGLE determined by the points) is

$$\begin{vmatrix} x^2 + y^2 & x & y & 1 \\ {x_1}^2 + {y_1}^2 & x_1 & y_1 & 1 \\ {x_2}^2 + {y_2}^2 & x_2 & y_2 & 1 \\ {x_3}^2 + {y_3}^2 & x_3 & y_3 & 1 \end{vmatrix} = 0. \qquad (20)$$

The CENTER and RADIUS of this circle can be identified by assigning coefficients of a QUADRATIC CURVE

$$ax^2 + cy^2 + dx + ey + f = 0, \qquad (21)$$

where $a = c$ and $b = 0$ (since there is no xy cross term). COMPLETING THE SQUARE gives

$$a\left(x + \frac{d}{2a}\right)^2 + a\left(y + \frac{e}{2a}\right)^2 + f - \frac{d^2 + e^2}{4a} = 0. \qquad (22)$$

The CENTER can then be identified as

$$x_0 = -\frac{d}{2a} \qquad (23)$$

$$y_0 = -\frac{e}{2a} \qquad (24)$$

and the RADIUS as

$$r = \sqrt{\frac{d^2 + e^2}{4a^2} - \frac{f}{a}}, \qquad (25)$$

where

$$a = \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} \qquad (26)$$

$$d = -\begin{vmatrix} {x_1}^2 + {y_1}^2 & y_1 & 1 \\ {x_2}^2 + {y_2}^2 & y_2 & 1 \\ {x_3}^2 + {y_3}^2 & y_3 & 1 \end{vmatrix} \qquad (27)$$

$$e = \begin{vmatrix} {x_1}^2 + {y_1}^2 & x_1 & 1 \\ {x_2}^2 + {y_2}^2 & x_2 & 1 \\ {x_3}^2 + {y_3}^2 & x_3 & 1 \end{vmatrix} \qquad (28)$$

$$f = -\begin{vmatrix} {x_1}^2 + {y_1}^2 & x_1 & y_1 \\ {x_2}^2 + {y_2}^2 & x_2 & y_2 \\ {x_3}^2 + {y_3}^2 & x_3 & y_3 \end{vmatrix}. \qquad (29)$$

Four or more points which lie on a circle are said to be CONCYCLIC. Three points are trivially concyclic since three noncollinear points determine a circle.

The CIRCUMFERENCE-to-DIAMETER ratio C/d for a circle is constant as the size of the circle is changed (as it must be since scaling a plane figure by a factor s increases its PERIMETER by s), and d also scales by s. This ratio is denoted π (PI), and has been proved TRANSCENDENTAL. With d the DIAMETER and r the RADIUS,

$$C = \pi d = 2\pi r. \qquad (30)$$

Knowing C/d, we can then compute the AREA of the circle either geometrically or using CALCULUS. From CALCULUS,

$$A = \int_0^{2\pi} d\theta \int_0^r r\,dr = (2\pi)(\tfrac{1}{2}r^2) = \pi r^2. \qquad (31)$$

Now for a few geometrical derivations. Using concentric strips, we have

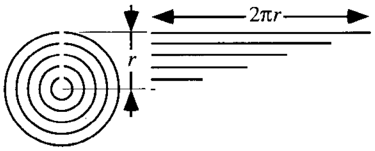

As the number of strips increases to infinity, we are left with a TRIANGLE on the right, so

$$A = \tfrac{1}{2}(2\pi r)r = \pi r^2. \qquad (32)$$

This derivation was first recorded by Archimedes in *Measurement of a Circle* (ca. 225 BC). If we cut the circle instead into wedges,

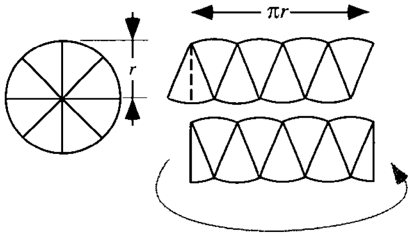

As the number of wedges increases to infinity, we are left with a RECTANGLE, so

$$A = (\pi r)r = \pi r^2. \qquad (33)$$

see also ARC, BLASCHKE'S THEOREM, BRAHMAGUPTA'S FORMULA, BROCARD CIRCLE, CASEY'S THEOREM, CHORD, CIRCUMCIRCLE, CIRCUMFERENCE, CLIFFORD'S CIRCLE THEOREM, CLOSED DISK, CONCENTRIC CIRCLES, COSINE CIRCLE, COTES CIRCLE PROPERTY, DIAMETER, DISK, DROZ-FARNY CIRCLES, EULER TRIANGLE FORMULA, EXCIRCLE, FEUERBACH'S THEOREM, Five Disks Problem, Flower of Life, Ford Circle, Fuhrmann Circle, Gerŝgorin Circle Theorem, Hopf Circle, Incircle, Inversive Distance, Johnson Circle, Kinney's Set, Lemoine Circle, Lens, Magic Circles, Malfatti Circles, McCay Circle, Midcircle, Monge's Theorem, Moser's Circle Problem, Neuberg Circles, Nine-Point Circle, Open Disk, *P*-Circle, Parry Circle, Pi, Polar Circle, Power (Circle), Prime Circle, Ptolemy's Theorem, Purser's Theorem, Radical Axis, Radius, Reuleaux Triangle, Seed of Life, Seifert Circle, Semicircle, Soddy Circles, Sphere, Taylor Circle, Triangle Inscribing in a Circle, Triplicate-Ratio Circle, Tucker Circles, Unit Circle, Venn Diagram, Villarceau Circles, Yin-Yang

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 125 and 197, 1987.

Casey, J. "The Circle." Ch. 3 in *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions, with Numerous Examples, 2nd ed., rev. enl.* Dublin: Hodges, Figgis, & Co., pp. 96–150, 1893.

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 74–75, 1996.

Dunham, W. "Archimedes' Determination of Circular Area." Ch. 4 in *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 84–112, 1990.

Eppstein, D. "Circles and Spheres." `http://www.ics.uci.edu/~eppstein/junkyard/sphere.html`.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 65–66, 1972.

MacTutor History of Mathematics Archive. "Circle." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Circle.html`.

Pappas, T. "Infinity & the Circle" and "Japanese Calculus." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 68 and 139, 1989.

Pedoe, D. *Circles: A Mathematical View, rev. ed.* Washington, DC: Math. Assoc. Amer., 1995.

Yates, R. C. "The Circle." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 21–25, 1952.

Circles-and-Squares Fractal

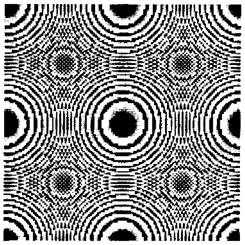

A Fractal produced by iteration of the equation

$$z_{n+1} = {z_n}^2 \pmod{m}$$

which results in a Møiré-like pattern.

see also Fractal, Møiré Pattern

Circle Caustic

Consider a point light source located at a point $(\mu, 0)$. The Catacaustic of a unit Circle for the light at $\mu = \infty$ is the Nephroid

$$x = \tfrac{1}{4}[3\cos t - \cos(3t)] \tag{1}$$

$$y = \tfrac{1}{4}[3\sin t - \sin(3t)]. \tag{2}$$

The Catacaustic for the light at a finite distance $\mu > 1$ is the curve

$$x = \frac{\mu(1 - 3\mu\cos t + 2\mu\cos^3 t)}{-(1 + 2\mu^2) + 3\mu\cos t} \tag{3}$$

$$y = \frac{2\mu^2 \sin^3 t}{1 + 2\mu^2 - 3\mu\cos t}, \tag{4}$$

and for the light on the Circumference of the Circle $\mu = 1$ is the Cardioid

$$x = \tfrac{2}{3}\cos t(1 + \cos t) - \tfrac{1}{3} \tag{5}$$

$$y = \tfrac{2}{3}\sin t(1 + \cos t). \tag{6}$$

If the point is inside the circle, the catacaustic is a discontinuous two-part curve. These four cases are illustrated below.

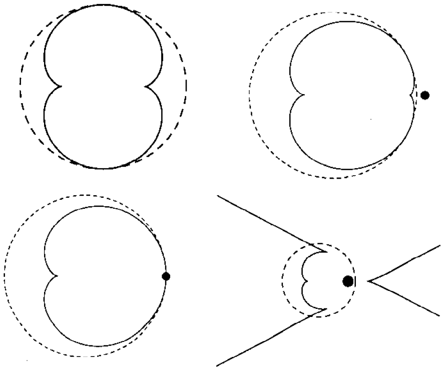

The Catacaustic for Parallel rays crossing a Circle is a Cardioid.

see also Catacaustic, Caustic

Circle-Circle Intersection

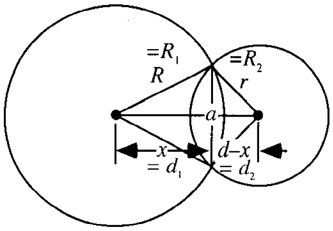

Let two Circles of Radii R and r and centered at $(0, 0)$ and $(d, 0)$ intersect in a Lens-shaped region. The equations of the two circles are

$$x^2 + y^2 = R^2 \tag{1}$$

$$(x - d)^2 + y^2 = r^2. \tag{2}$$

Combining (1) and (2) gives

$$(x-d)^2+(R^2-x^2)=r^2. \tag{3}$$

Multiplying through and rearranging gives

$$x^2-2dx+d^2-x^2=r^2-R^2. \tag{4}$$

Solving for x results in

$$x=\frac{d^2-r^2+R^2}{2d}. \tag{5}$$

The line connecting the cusps of the LENS therefore has half-length given by plugging x back in to obtain

$$\begin{aligned} y^2=R^2-x^2&=R^2-\left(\frac{d^2-r^2+R^2}{2d}\right)^2\\ &=\frac{4d^2R^2-(d^2-r^2+R^2)^2}{4d^2}, \end{aligned} \tag{6}$$

giving a length of

$$\begin{aligned} a&=\frac{1}{d}\sqrt{4d^2R^2-(d^2-r^2+R^2)^2}\\ &=\frac{1}{d}[(-d+r-R)(-d-r+R)\\ &\quad\times[(-d+r+R)(d+r+R)]^{1/2}. \end{aligned} \tag{7}$$

This same formulation applies directly to the SPHERE-SPHERE INTERSECTION problem.

To find the AREA of the asymmetric "LENS" in which the CIRCLES intersect, simply use the formula for the circular SEGMENT of radius $R'$and triangular height d'

$$A(R',d')=R'^2\cos^{-1}\left(\frac{d'}{R'}\right)-d'\sqrt{R'^2-d'^2} \tag{8}$$

twice, one for each half of the "LENS." Noting that the heights of the two segment triangles are

$$d_1=x=\frac{d^2-r^2+R^2}{2d} \tag{9}$$

$$d_2=d-x=\frac{d^2+r^2-R^2}{2d}. \tag{10}$$

The result is

$$\begin{aligned} A&=A(R_1,d_1)+A(R_2,d_2)\\ &=r^2\cos^{-1}\left(\frac{d^2+r^2-R^2}{2dr}\right)\\ &\quad+R^2\cos^{-1}\left(\frac{d^2+R^2-r^2}{2dR}\right)\\ &\quad-\tfrac{1}{2}\sqrt{(d-r-R)(d+r-R)(d-r+R)(d+r+R)}. \end{aligned} \tag{11}$$

The limiting cases of this expression can be checked to give 0 when $d=R+r$ and

$$A=2R^2\cos^{-1}\left(\frac{d}{2R}\right)-\tfrac{1}{2}d\sqrt{4R^2-d^2} \tag{12}$$

$$=2A(\tfrac{1}{2}d,R) \tag{13}$$

when $r=R$, as expected. In order for half the area of two UNIT DISKS ($R=1$) to overlap, set $A=\pi R^2/2=\pi/2$ in the above equation

$$\tfrac{1}{2}\pi=2\cos^{-1}(\tfrac{1}{2}d)-\tfrac{1}{2}d\sqrt{4-d^2} \tag{14}$$

and solve numerically, yielding $d\approx 0.807946$.

see also LENS, SEGMENT, SPHERE-SPHERE INTERSECTION

Circle Cutting

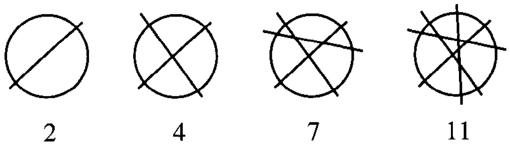

Determining the maximum number of pieces in which it is possible to divide a CIRCLE for a given number of cuts is called the circle cutting, or sometimes PANCAKE CUTTING, problem. The minimum number is always $n+1$, where n is the number of cuts, and it is always possible to obtain any number of pieces between the minimum and maximum. The first cut creates 2 regions, and the nth cut creates n new regions, so

$$f(1)=2 \tag{1}$$

$$f(2)=2+f(1) \tag{2}$$

$$f(n)=n+f(n-1). \tag{3}$$

Therefore,

$$\begin{aligned} f(n)&=n+[(n-1)+f(n-2)]\\ &=n+(n-1)+\ldots+2+f(1)=\sum_{k=2}^{n}kf(1)\\ &=\sum_{k=1}^{n}k-1+f(1)=\tfrac{1}{2}n(n+1)-1+2\\ &=\tfrac{1}{2}(n^2+n+2). \end{aligned} \tag{4}$$

Evaluating for $n=1, 2, \ldots$ gives 2, 4, 7, 11, 16, 22, ... (Sloane's A000124).

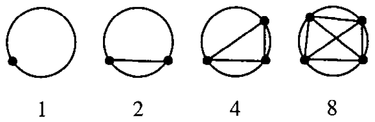

A related problem, sometimes called MOSER'S CIRCLE PROBLEM, is to find the number of pieces into which a CIRCLE is divided if n points on its CIRCUMFERENCE are joined by CHORDS with no three CONCURRENT. The answer is

$$g(n) = \binom{n}{4} + \binom{n}{2} + 1 \tag{5}$$

$$= \tfrac{1}{24}(n^4 - 6n^3 + 23n^2 - 18n + 24), \tag{6}$$

(Yaglom and Yaglom 1987, Guy 1988, Conway and Guy 1996, Noy 1996), where $\binom{n}{m}$ is a BINOMIAL COEFFICIENT. The first few values are 1, 2, 4, 8, 16, 31, 57, 99, 163, 256, ... (Sloane's A000127). This sequence and problem are an example of the danger in making assumptions based on limited trials. While the series starts off like 2^{n-1}, it begins differing from this GEOMETRIC SERIES at $n = 6$.

see also CAKE CUTTING, CYLINDER CUTTING, HAM SANDWICH THEOREM, PANCAKE THEOREM, PIZZA THEOREM, SQUARE CUTTING, TORUS CUTTING

References
Conway, J. H. and Guy, R. K. "How Many Regions." In *The Book of Numbers.* New York: Springer-Verlag, pp. 76–79, 1996.
Guy, R. K. "The Strong Law of Small Numbers." *Amer. Math. Monthly* **95**, 697–712, 1988.
Noy, M. "A Short Solution of a Problem in Combinatorial Geometry." *Math. Mag.* **69**, 52–53, 1996.
Sloane, N. J. A. Sequences A000124/M1041 and A000127/M1119 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Yaglom, A. M. and Yaglom, I. M. Problem 47. *Challenging Mathematical Problems with Elementary Solutions, Vol. 1.* New York: Dover, 1987.

Circle Evolute

$$x = \cos t \qquad x' = -\sin t \qquad x'' = -\cos t \tag{1}$$

$$y = \sin t \qquad y' = \cos t \qquad y'' = -\sin t, \tag{2}$$

so the RADIUS OF CURVATURE is

$$R = \frac{(x'^2 + y'^2)^{3/2}}{y''x' - x''y'}$$

$$= \frac{(\sin^2 t + \cos^2 t)^{3/2}}{(-\sin t)(-\sin t) - (-\cos t)\cos t} = 1, \tag{3}$$

and the TANGENT VECTOR is

$$\hat{\mathbf{T}} = \begin{bmatrix} -\sin t \\ \cos t \end{bmatrix}. \tag{4}$$

Therefore,

$$\cos\tau = \hat{\mathbf{T}} \cdot \hat{\mathbf{x}} = -\sin t \tag{5}$$

$$\sin\tau = \hat{\mathbf{T}} \cdot \hat{\mathbf{y}} = \cos t, \tag{6}$$

so

$$\xi(t) = x - R\sin\tau = \cos t - 1 \cdot \cos t = 0 \tag{7}$$

$$\eta(t) = y + R\cos\tau = \sin t + 1 \cdot (-\sin t) = 0, \tag{8}$$

and the EVOLUTE degenerates to a POINT at the ORIGIN.

see also CIRCLE INVOLUTE

References
Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 77, 1993.
Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 55–59, 1991.

Circle Inscribing

If r is the RADIUS of a CIRCLE inscribed in a RIGHT TRIANGLE with sides a and b and HYPOTENUSE c, then

$$r = \tfrac{1}{2}(a + b - c).$$

see INSCRIBED, POLYGON

Circle Involute

First studied by Huygens when he was considering clocks without pendula for use on ships at sea. He used the circle involute in his first pendulum clock in an attempt to force the pendulum to swing in the path of a CYCLOID.

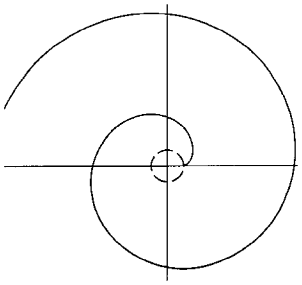

For a CIRCLE with $a = 1$, the parametric equations of the circle and their derivatives are given by

$$x = \cos t \qquad x' = -\sin t \qquad x'' = -\cos t \tag{1}$$

$$y = \sin t \qquad y' = \cos t \qquad y'' = -\sin t. \tag{2}$$

The TANGENT VECTOR is

$$\hat{\mathbf{T}} = \begin{bmatrix} -\sin t \\ \cos t \end{bmatrix} \tag{3}$$

and the ARC LENGTH along the circle is

$$s = \int \sqrt{x'^2 + y'^2}\, dt = \int dt = t, \tag{4}$$

so the involute is given by

$$\mathbf{r}_i = \mathbf{r} - s\hat{\mathbf{T}} = \begin{bmatrix} \cos t \\ \sin t \end{bmatrix} - t \begin{bmatrix} -\sin t \\ \cos t \end{bmatrix} = \begin{bmatrix} \cos t + t\sin t \\ \sin t - t\cos t \end{bmatrix}, \tag{5}$$

or

$$x = a(\cos t + t\sin t) \tag{6}$$

$$y = a(\sin t - t\cos t). \tag{7}$$

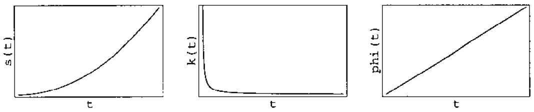

The ARC LENGTH, CURVATURE, and TANGENTIAL ANGLE are

$$s = \int ds = \int \sqrt{x'^2 + y'^2}\,dt = \tfrac{1}{2}at^2 \qquad (8)$$

$$\kappa = \frac{1}{at} \qquad (9)$$

$$\phi = t. \qquad (10)$$

The CESÀRO EQUATION is

$$\kappa = \frac{1}{\sqrt{as}}. \qquad (11)$$

see also CIRCLE, CIRCLE EVOLUTE, ELLIPSE INVOLUTE, INVOLUTE

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 83, 1993.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 190–191, 1972.

MacTutor History of Mathematics Archive. "Involute of a Circle." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Involute.html`.

Circle Involute Pedal Curve

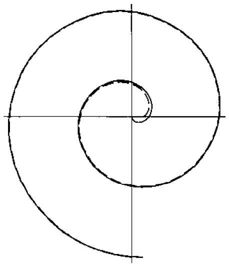

The PEDAL CURVE of CIRCLE INVOLUTE

$$f = \cos t + t \sin t$$
$$g = \sin t - t \cos t$$

with the center as the PEDAL POINT is the ARCHIMEDES' SPIRAL

$$x = t \sin t$$
$$y = -t \cos t.$$

Circle Lattice Points

For every POSITIVE INTEGER n, there exists a CIRCLE which contains exactly n lattice points in its interior. H. Steinhaus proved that for every POSITIVE INTEGER n, there exists a CIRCLE of AREA n which contains exactly n lattice points in its interior.

SCHINZEL'S THEOREM shows that for every POSITIVE INTEGER n, there exists a CIRCLE in the PLANE having exactly n LATTICE POINTS on its CIRCUMFERENCE. The theorem also explicitly identifies such "SCHINZEL CIRCLES" as

$$\begin{cases} (x - \frac{1}{2})^2 + y^2 = \frac{1}{4}5^{k-1} & \text{for } n = 2k \\ (x - \frac{1}{3})^2 + y^2 = \frac{1}{9}5^{2k} & \text{for } n = 2k+1. \end{cases} \qquad (1)$$

Note, however, that these solutions do not necessarily have the smallest possible RADIUS. For example, while the SCHINZEL CIRCLE centered at (1/3, 0) and with RADIUS 625/3 has nine lattice points on its CIRCUMFERENCE, so does the CIRCLE centered at (1/3, 0) with RADIUS 65/3.

Let r be the smallest INTEGER RADIUS of a CIRCLE centered at the ORIGIN (0, 0) with $L(r)$ LATTICE POINTS. In order to find the number of lattice points of the CIRCLE, it is only necessary to find the number in the first octant, i.e., those with $0 \leq y \leq \lfloor r/\sqrt{2} \rfloor$, where $\lfloor z \rfloor$ is the FLOOR FUNCTION. Calling this $N(r)$, then for $r \geq 1$, $L(r) = 8N(r) - 4$, so $L(r) \equiv 4 \pmod 8$. The multiplication by eight counts all octants, and the subtraction by four eliminates points on the axes which the multiplication counts twice. (Since $\sqrt{2}$ is IRRATIONAL, the MIDPOINT of a are is never a LATTICE POINT.)

GAUSS'S CIRCLE PROBLEM asks for the number of lattice points *within* a CIRCLE of RADIUS r

$$N(r) = 1 + 4\lfloor r \rfloor + 4 \sum_{i=1}^{\lfloor r \rfloor} \left\lfloor \sqrt{r^2 - i^2} \right\rfloor. \qquad (2)$$

Gauss showed that

$$N(r) = \pi r^2 + E(r), \qquad (3)$$

where

$$|E(r)| \leq 2\sqrt{2}\,\pi r. \qquad (4)$$

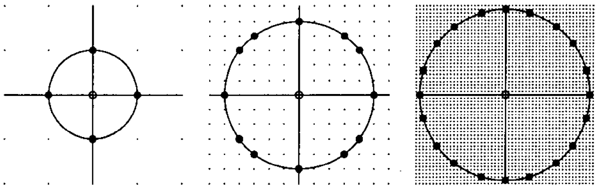

The number of lattice points on the CIRCUMFERENCE of circles centered at (0, 0) with radii 0, 1, 2, ... are 1, 4, 4, 4, 4, 12, 4, 4, 4, 4, 12, 4, 4, ... (Sloane's A046109). The following table gives the smallest RADIUS $r \leq 111{,}000$ for a circle centered at (0, 0) having a given number of LATTICE POINTS $L(r)$. Note that the high water mark radii are always multiples of five.

$L(r)$	r
1	0
4	1
12	5
20	25
28	125
36	65
44	3,125
52	15,625
60	325
68	$\leq 390,625$
76	$\leq 1,953,125$
84	1,625
92	$\leq 48,828,125$
100	4,225
108	1,105
132	40,625
140	21,125
180	5,525
252	27,625
300	71,825
324	32,045

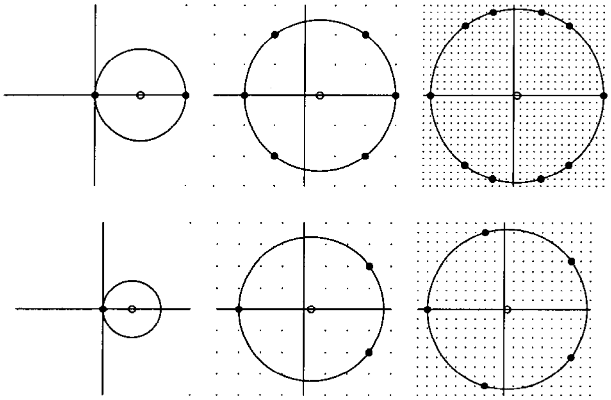

If the CIRCLE is instead centered at (1/2, 0), then the CIRCLES of RADII 1/2, 3/2, 5/2, ... have 2, 2, 6, 2, 2, 2, 6, 6, 6, 2, 2, 2, 10, 2, ... (Sloane's A046110) on their CIRCUMFERENCES. If the CIRCLE is instead centered at (1/3, 0), then the number of lattice points on the CIRCUMFERENCE of the CIRCLES of RADIUS 1/3, 2/3, 4/3, 5/3, 7/3, 8/3, ... are 1, 1, 1, 3, 1, 1, 3, 1, 3, 1, 1, 3, 1, 3, 1, 1, 5, 3, ... (Sloane's A046111).

Let

1. a_n be the RADIUS of the CIRCLE centered at (0, 0) having $8n+4$ lattice points on its CIRCUMFERENCE,
2. $b_n/2$ be the RADIUS of the CIRCLE centered at (1/2, 0) having $4n+2$ lattice points on its CIRCUMFERENCE,
3. $c_n/3$ be the RADIUS of CIRCLE centered at (1/3, 0) having $2n+1$ lattice points on its CIRCUMFERENCE.

Then the sequences $\{a_n\}$, $\{b_n\}$, and $\{c_n\}$ are equal, with the exception that $b_n = 0$ if $2|n$ and $c_n = 0$ if $3|n$. However, the sequences of *smallest* radii having the above numbers of lattice points are equal in the three cases and given by 1, 5, 25, 125, 65, 3125, 15625, 325, ... (Sloane's A046112).

KULIKOWSKI'S THEOREM states that for every POSITIVE INTEGER n, there exists a 3-D SPHERE which has exactly n LATTICE POINTS on its surface. The SPHERE is given by the equation

$$(x-a)^2 + (y-b)^2 + (z-\sqrt{2}\,)^2 = c^2 + 2,$$

where a and b are the coordinates of the center of the so-called SCHINZEL CIRCLE and c is its RADIUS (Honsberger 1973).

see also CIRCLE, CIRCUMFERENCE, GAUSS'S CIRCLE PROBLEM, KULIKOWSKI'S THEOREM, LATTICE POINT, SCHINZEL CIRCLE, SCHINZEL'S THEOREM

References

Honsberger, R. "Circles, Squares, and Lattice Points." Ch. 11 in *Mathematical Gems I.* Washington, DC: Math. Assoc. Amer., pp. 117–127, 1973.

Kulikowski, T. "Sur l'existence d'une sphère passant par un nombre donné aux coordonnées entières." *L'Enseignement Math. Ser. 2* **5**, 89–90, 1959.

Schinzel, A. "Sur l'existence d'un cercle passant par un nombre donné de points aux coordonnées entières." *L'Enseignement Math. Ser. 2* **4**, 71–72, 1958.

Sierpiński, W. "Sur quelques problèmes concernant les points aux coordonnées entières." *L'Enseignement Math. Ser. 2* **4**, 25–31, 1958.

Sierpiński, W. "Sur un problème de H. Steinhaus concernant les ensembles de points sur le plan." *Fund. Math.* **46**, 191–194, 1959.

Sierpiński, W. *A Selection of Problems in the Theory of Numbers.* New York: Pergamon Press, 1964.

Weisstein, E. W. "Circle Lattice Points." `http://www.astro.virginia.edu/~eww6n/math/notebooks/CircleLatticePoints.m`.

Circle Lattice Theorem

see GAUSS'S CIRCLE PROBLEM

Circle Map

A 1-D MAP which maps a CIRCLE onto itself

$$\theta_{n+1} = \theta_n + \Omega - \frac{K}{2\pi}\sin(2\pi\theta_n), \tag{1}$$

where θ_{n+1} is computed mod 1. Note that the circle map has two parameters: Ω and K. Ω can be interpreted as an externally applied frequency, and K as a strength of nonlinearity. The 1-D JACOBIAN is

$$\frac{\partial\theta_{n+1}}{\partial\theta_n} = 1 - K\cos(2\pi\theta_n), \tag{2}$$

so the circle map is not AREA-PRESERVING. It is related to the STANDARD MAP

$$I_{n+1} = I_n + \frac{K}{2\pi}\sin(2\pi\theta_n) \tag{3}$$

$$\theta_{n+1} = \theta_n + I_{n+1}, \tag{4}$$

for I and θ computed mod 1. Writing θ_{n+1} as

$$\theta_{n+1} = \theta_n + I_n + \frac{K}{2\pi}\sin(2\pi\theta_n) \tag{5}$$

gives the circle map with $I_n = \Omega$ and $K = -K$. The unperturbed circle map has the form

$$\theta_{n+1} = \theta_n + \Omega. \tag{6}$$

If Ω is RATIONAL, then it is known as the map WINDING NUMBER, defined by

$$\Omega = W \equiv \frac{p}{q}, \tag{7}$$

and implies a periodic trajectory, since θ_n will return to the same point (at most) every q ORBITS. If Ω is IRRATIONAL, then the motion is quasiperiodic. If K is NONZERO, then the motion may be periodic in some finite region surrounding each RATIONAL Ω. This execution of periodic motion in response to an IRRATIONAL forcing is known as MODE LOCKING.

If a plot is made of K vs. Ω with the regions of periodic MODE-LOCKED parameter space plotted around RATIONAL Ω values (WINDING NUMBERS), then the regions are seen to widen upward from 0 at $K = 0$ to some finite width at $K = 1$. The region surrounding each RATIONAL NUMBER is known as an ARNOLD TONGUE. At $K = 0$, the ARNOLD TONGUES are an isolated set of MEASURE zero. At $K = 1$, they form a CANTOR SET of DIMENSION $d \approx 0.08700$. For $K > 1$, the tongues overlap, and the circle map becomes noninvertible. The circle map has a FEIGENBAUM CONSTANT

$$\delta \equiv \lim_{n\to\infty} \frac{\theta_n - \theta_{n-1}}{\theta_{n+1} - \theta_n} = 2.833. \tag{8}$$

see also ARNOLD TONGUE, DEVIL'S STAIRCASE, MODE LOCKING, WINDING NUMBER (MAP)

Circle Method

see PARTITION FUNCTION P

Circle Negative Pedal Curve

The NEGATIVE PEDAL CURVE of a circle is an ELLIPSE if the PEDAL POINT is inside the CIRCLE, and a HYPERBOLA if the PEDAL POINT is outside the CIRCLE.

Circle Notation

A NOTATION for LARGE NUMBERS due to Steinhaus (1983) in which ⓝ is defined in terms of STEINHAUS-MOSER NOTATION as n in n SQUARES. The particular number known as the MEGA is then defined as follows.

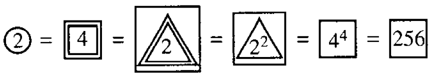

see also MEGA, MEGISTRON, STEINHAUS-MOSER NOTATION

References

Steinhaus, H. *Mathematical Snapshots, 3rd American ed.* New York: Oxford University Press, pp. 28–29, 1983.

Circle Order

A POSET P is a circle order if it is ISOMORPHIC to a SET of DISKS ordered by containment.

see also ISOMORPHIC POSETS, PARTIALLY ORDERED SET

Circle Orthotomic

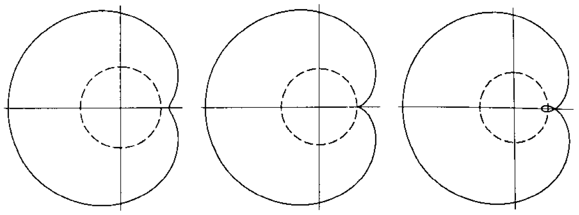

The ORTHOTOMIC of the CIRCLE represented by

$$x = \cos t \tag{1}$$
$$y = \sin t \tag{2}$$

with a source at (x, y) is

$$x = x\cos(2t) - y\sin(2t) + 2\sin t \tag{3}$$
$$y = -x\sin(2t) - y\cos(2t) + 2\cos t. \tag{4}$$

Circle Packing

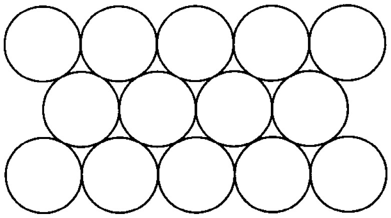

The densest packing of spheres in the PLANE is the hexagonal lattice of the bee's honeycomb (illustrated above), which has a PACKING DENSITY of

$$\eta = \frac{\pi}{2\sqrt{3}} = 0.9068996821\ldots.$$

Gauss proved that the hexagonal lattice is the densest plane *lattice* packing, and in 1940, L. Fejes Tóth proved that the hexagonal lattice is indeed the densest of *all* possible plane packings.

Solutions for the smallest diameter CIRCLES into which n UNIT CIRCLES can be packed have been proved optimal for $n = 1$ through 10 (Kravitz 1967). The best known results are summarized in the following table.

n	d exact	d approx.
1	1	1.00000
2	2	2.00000
3	$1+\frac{2}{3}\sqrt{3}$	2.15470...
4	$1+\sqrt{2}$	2.41421...
5	$1+\sqrt{2(1+1/\sqrt{5})}$	2.70130...
6	3	3.00000
7	3	3.00000
8	$1+\csc(\pi/7)$	3.30476...
9	$1+\sqrt{2(2+\sqrt{2})}$	3.61312...
10		3.82...
11		
12		4.02...

For CIRCLE packing inside a SQUARE, proofs are known only for $n = 1$ to 9.

n	d exact	d approx.
1	1	1.000
2		0.58...
3		0.500...
4	$\frac{1}{2}$	0.500
5		0.41...
6		0.37...
7		0.348...
8		0.341...
9	$\frac{1}{3}$	0.333...
10		0.148204...

The smallest SQUARE into which two UNIT CIRCLES, one of which is split into two pieces by a chord, can be packed is not known (Goldberg 1968, Ogilvy 1990).

see also HYPERSPHERE PACKING, MALFATTI'S RIGHT TRIANGLE PROBLEM, MERGELYAN-WESLER THEOREM, SPHERE PACKING

References
Conway, J. H. and Sloane, N. J. A. *Sphere Packings, Lattices, and Groups, 2nd ed.* New York: Springer-Verlag, 1992.
Eppstein, D. "Covering and Packing." `http://www.ics.uci.edu/~eppstein/junkyard/cover.html`.
Folkman, J. H. and Graham, R. "A Packing Inequality for Compact Convex Subsets of the Plane." *Canad. Math. Bull.* **12**, 745–752, 1969.
Gardner, M. "Mathematical Games: The Diverse Pleasures of Circles that Are Tangent to One Another." *Sci. Amer.* **240**, 18–28, Jan. 1979.
Gardner, M. "Tangent Circles." Ch. 10 in *Fractal Music, Hypercards, and More Mathematical Recreations from Scientific American Magazine.* New York: W. H. Freeman, 1992.
Goldberg, M. "Problem E1924." *Amer. Math. Monthly* **75**, 195, 1968.
Goldberg, M. "The Packing of Equal Circles in a Square." *Math. Mag.* **43**, 24–30, 1970.
Goldberg, M. "Packing of 14, 16, 17, and 20 Circles in a Circle." *Math. Mag.* **44**, 134–139, 1971.
Graham, R. L. and Luboachevsky, B. D. "Repeated Patterns of Dense Packings of Equal Disks in a Square." *Electronic J. Combinatorics* **3**, R16, 1–17, 1996. `http://www.combinatorics.org/Volume_3/volume3.html#R16`.
Kravitz, S. "Packing Cylinders into Cylindrical Containers." *Math. Mag.* **40**, 65–70, 1967.
McCaughan, F. "Circle Packings." `http://www.pmms.cam.ac.uk/~gjm11/cpacking/info.html`.
Molland, M. and Payan, Charles. "A Better Packing of Ten Equal Circles in a Square." *Discrete Math.* **84**, 303–305, 1990.
Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, p. 145, 1990.
Reis, G. E. "Dense Packing of Equal Circle within a Circle." *Math. Mag.* **48**, 33–37, 1975.
Schaer, J. "The Densest Packing of Nine Circles in a Square." *Can. Math. Bul.* **8**, 273–277, 1965.
Schaer, J. "The Densest Packing of Ten Equal Circles in a Square." *Math. Mag.* **44**, 139–140, 1971.
Valette, G. "A Better Packing of Ten Equal Circles in a Square." *Discrete Math.* **76**, 57–59, 1989.

Circle Pedal Curve

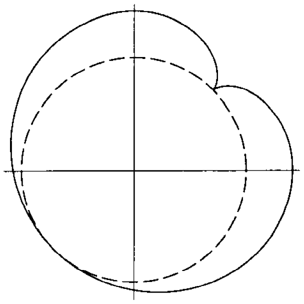

The PEDAL CURVE of a CIRCLE is a CARDIOID if the PEDAL POINT is taken on the CIRCUMFERENCE,

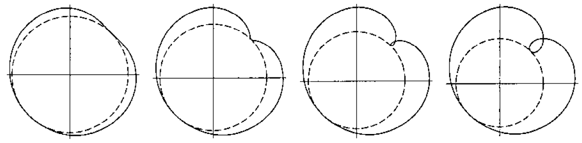

and otherwise a LIMAÇON.

Circle-Point Midpoint Theorem

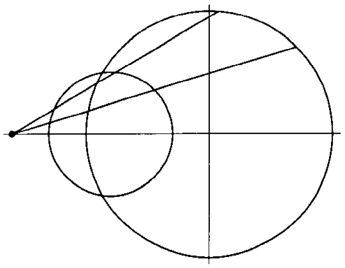

Taking the locus of MIDPOINTS from a fixed point to a circle of radius r results in a circle of radius $r/2$. This follows trivially from

$$\begin{aligned}\mathbf{r}(\theta) &= \begin{bmatrix}-x\\0\end{bmatrix}+\frac{1}{2}\left(\begin{bmatrix}r\cos\theta\\r\sin\theta\end{bmatrix}-\begin{bmatrix}-x\\0\end{bmatrix}\right)\\&=\begin{bmatrix}\frac{1}{2}r\cos\theta-\frac{1}{2}x\\\frac{1}{2}\sin\theta\end{bmatrix}.\end{aligned}$$

References
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, p. 17, 1929.

Circle Radial Curve

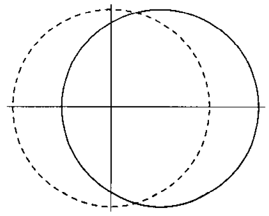

The RADIAL CURVE of a unit CIRCLE from a RADIAL POINT $(x, 0)$ is another CIRCLE with parametric equations

$$\begin{aligned} x(t) &= x - \cos t \\ y(t) &= -\sin t. \end{aligned}$$

Circle Squaring

Construct a SQUARE equal in AREA to a CIRCLE using only a STRAIGHTEDGE and COMPASS. This was one of the three GEOMETRIC PROBLEMS OF ANTIQUITY, and was perhaps first attempted by Anaxagoras. It was finally proved to be an impossible problem when PI was proven to be TRANSCENDENTAL by Lindemann in 1882.

However, approximations to circle squaring are given by constructing lengths close to $\pi = 3.1415926\ldots$. Ramanujan (1913–14) and Olds (1963) give geometric constructions for $355/113 = 3.1415929\ldots$. Gardner (1966, pp. 92–93) gives a geometric construction for $3 + 16/113 = 3.1415929\ldots$. Dixon (1991) gives constructions for $6/5(1+\phi) = 3.141640\ldots$ and $\sqrt{4 + [3 - \tan(30°)]} = 3.141533\ldots$.

While the circle cannot be squared in EUCLIDEAN SPACE, it *can* in GAUSS-BOLYAI-LOBACHEVSKY SPACE (Gray 1989).

see also GEOMETRIC CONSTRUCTION, QUADRATURE, SQUARING

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 190–191, 1996.

Dixon, R. *Mathographics.* New York: Dover, pp. 44–49 and 52–53, 1991.

Dunham, W. "Hippocrates' Quadrature of the Lune." Ch. 1 in *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 20–26, 1990.

Gardner, M. "The Transcendental Number Pi." Ch. 8 in *Martin Gardner's New Mathematical Diversions from Scientific American.* New York: Simon and Schuster, 1966.

Gray, J. *Ideas of Space.* Oxford, England: Oxford University Press, 1989.

Meyers, L. F. "Update on William Wernick's 'Triangle Constructions with Three Located Points.'" *Math. Mag.* **69**, 46–49, 1996.

Olds, C. D. *Continued Fractions.* New York: Random House, pp. 59–60, 1963.

Ramanujan, S. "Modular Equations and Approximations to π." *Quart. J. Pure. Appl. Math.* **45**, 350–372, 1913–1914.

Circle Strophoid

The STROPHOID of a CIRCLE with pole at the center and fixed point on the CIRCUMFERENCE is a FREETH'S NEPHROID.

Circle Tangents

There are four CIRCLES that touch all the sides of a given TRIANGLE. These are all touched by the CIRCLE through the intersection of the ANGLE BISECTORS of the TRIANGLE, known as the NINE-POINT CIRCLE.

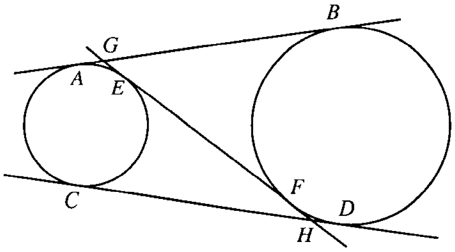

Given the above figure, $GE = FH$, since

$$\begin{aligned} AB &= AG + GB = GE + GF = GE + (GE + EF) \\ &= 2G + EF \\ CD &= CH + HD = EH + FH = FH + (FH + EF) \\ &= EF + 2FH. \end{aligned}$$

Because $AB = CD$, it follows that $GE = FH$.

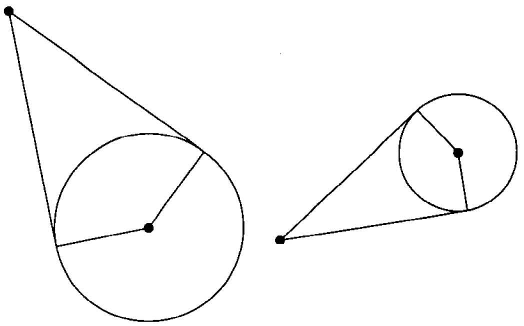

The line tangent to a CIRCLE of RADIUS a centered at (x, y)

$$\begin{aligned} x' &= x + a\cos t \\ y' &= y + a\sin t \end{aligned}$$

through $(0, 0)$ can be found by solving the equation

$$\begin{bmatrix} x + a\cos t \\ y + a\sin t \end{bmatrix} \cdot \begin{bmatrix} a\cos t \\ a\sin t \end{bmatrix} = 0,$$

giving

$$t = \pm\cos^{-1}\left(\frac{-ax \pm y\sqrt{x^2 + y^2 - a^2}}{x^2 + y^2}\right).$$

Two of these four solutions give tangent lines, as illustrated above.

see also KISSING CIRCLES PROBLEM, MIQUEL POINT, MONGE'S PROBLEM, PEDAL CIRCLE, TANGENT LINE, TRIANGLE

References
Dixon, R. *Mathographics.* New York: Dover, p. 21, 1991.
Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., pp. 4–5, 1991.

Circuit

see CYCLE (GRAPH)

Circuit Rank

Also known as the CYCLOMATIC NUMBER. The circuit rank is the smallest number of EDGES γ which must be removed from a GRAPH of N EDGES and n nodes such that no CIRCUIT remains.

$$\gamma = N - n + 1.$$

Circulant Determinant

Gradshteyn and Ryzhik (1970) define circulants by

$$\begin{vmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ x_n & x_1 & x_2 & \cdots & x_{n-1} \\ x_{n-1} & x_n & x_1 & \cdots & x_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_2 & x_3 & x_4 & \cdots & x_1 \end{vmatrix} = \prod_{j=1} (x_1 + x_2\omega_j + x_3{\omega_j}^2 + \ldots + x_n{\omega_j}^{n-1}), \quad (1)$$

where ω_j is the nth ROOT OF UNITY. The second-order circulant determinant is

$$\begin{vmatrix} x_1 & x_2 \\ x_2 & x_1 \end{vmatrix} = (x_1 + x_2)(x_1 - x_2), \quad (2)$$

and the third order is

$$\begin{vmatrix} x_1 & x_2 & x_3 \\ x_3 & x_1 & x_2 \\ x_2 & x_3 & x_1 \end{vmatrix} = (x_1 + x_2 + x_3)(x_1 + \omega x_2 + \omega^2 x_3)(x_1 + \omega^2 x_2 + \omega x_3), \quad (3)$$

where ω and ω^2 are the COMPLEX CUBE ROOTS of UNITY.

The EIGENVALUES λ of the corresponding $n \times n$ circulant matrix are

$$\lambda_j = x_1 + x_2\omega_j + x_3{\omega_j}^2 + \ldots + x_n{\omega_j}^{n-1}. \quad (4)$$

see also CIRCULANT MATRIX

References
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 1111–1112, 1979.
Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 114, 1991.

Circulant Graph

A GRAPH of n VERTICES in which the ith VERTEX is adjacent to the $(i+j)$th and $(i-j)$th VERTICES for each j in a list l.

Circulant Matrix

An $n \times n$ MATRIX C defined as follows,

$$\mathsf{C} = \begin{bmatrix} 1 & \binom{n}{1} & \binom{n}{2} & \cdots & \binom{n}{n-1} \\ \binom{n}{n-1} & 1 & \binom{n}{1} & \cdots & \binom{n}{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \binom{n}{1} & \binom{n}{2} & \binom{n}{3} & \cdots & 1 \end{bmatrix}$$

$$\mathsf{C} = \prod_{j=0}^{n-1} [(1 + \omega_j)^n - 1],$$

where $\omega_0 \equiv 1$, ω_1, ..., ω_{n-1} are the nth ROOTS OF UNITY. Circulant matrices are examples of LATIN SQUARES.

see also CIRCULANT DETERMINANT

References
Davis, P. J. *Circulant Matrices, 2nd ed.* New York: Chelsea, 1994.
Stroeker, R. J. "Brocard Points, Circulant Matrices, and Descartes' Folium." *Math. Mag.* **61**, 172–187, 1988.
Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 114, 1991.

Circular Cylindrical Coordinates

see CYLINDRICAL COORDINATES

Circular Functions

The functions describing the horizontal and vertical positions of a point on a CIRCLE as a function of ANGLE (COSINE and SINE) and those functions derived from them:

$$\cot x \equiv \frac{1}{\tan x} = \frac{\cos x}{\sin x} \quad (1)$$

$$\csc x \equiv \frac{1}{\sin x} \quad (2)$$

$$\sec x \equiv \frac{1}{\cos x} \quad (3)$$

$$\tan x \equiv \frac{\sin x}{\cos x}. \quad (4)$$

The study of circular functions is called TRIGONOMETRY.

see also COSECANT, COSINE, COTANGENT, ELLIPTIC FUNCTION, GENERALIZED HYPERBOLIC FUNCTIONS, HYPERBOLIC FUNCTIONS, SECANT, SINE, TANGENT, TRIGONOMETRY

References
Abramowitz, M. and Stegun, C. A. (Eds.). "Circular Functions." §4.3 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 71–79, 1972.

Circular Permutation

The number of ways to arrange n distinct objects along a CIRCLE is

$$P_n = (n-1)!.$$

The number is $(n-1)!$ instead of the usual FACTORIAL $n!$ since all CYCLIC PERMUTATIONS of objects are equivalent because the CIRCLE can be rotated.

see also PERMUTATION, PRIME CIRCLE

Circumcenter

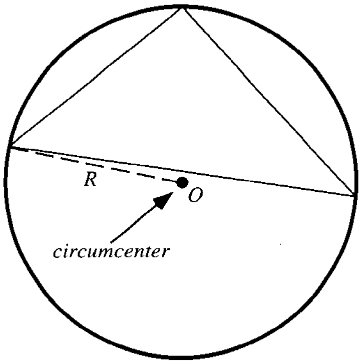

The center O of a TRIANGLE'S CIRCUMCIRCLE. It can be found as the intersection of the PERPENDICULAR BISECTORS. If the TRIANGLE is ACUTE, the circumcenter is in the interior of the TRIANGLE. In a RIGHT TRIANGLE, the circumcenter is the MIDPOINT of the HYPOTENUSE.

$$\overline{OO_1} + \overline{OO_2} + \overline{OO_3} = R + r, \tag{1}$$

where O_i are the MIDPOINTS of sides A_i, R is the CIRCUMRADIUS, and r is the INRADIUS (Johnson 1929, p. 190). The TRILINEAR COORDINATES of the circumcenter are

$$\cos A : \cos B : \cos C, \tag{2}$$

and the exact trilinears are therefore

$$R\cos A : R\cos B : R\cos C. \tag{3}$$

The AREAL COORDINATES are

$$(\tfrac{1}{2}a\cot A, \tfrac{1}{2}b\cot B, \tfrac{1}{2}c\cot C). \tag{4}$$

The distance between the INCENTER and circumcenter is $\sqrt{R(R-2r)}$. Given an interior point, the distances to the VERTICES are equal IFF this point is the circumcenter. It lies on the BROCARD AXIS.

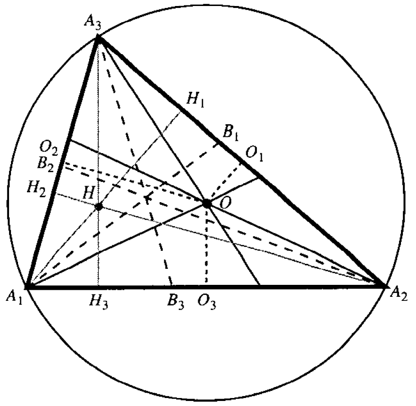

The circumcenter O and ORTHOCENTER H are ISOGONAL CONJUGATES.

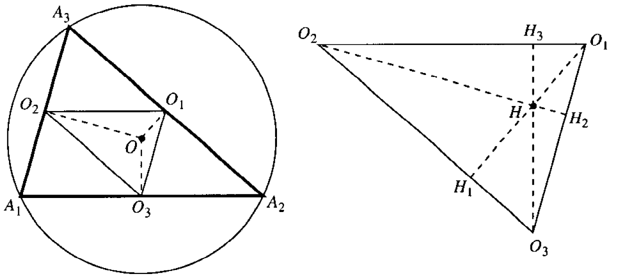

The ORTHOCENTER H of the PEDAL TRIANGLE $\Delta O_1O_2O_3$ formed by the CIRCUMCENTER O concurs with the circumcenter O itself, as illustrated above. The circumcenter also lies on the EULER LINE.

see also BROCARD DIAMETER, CARNOT'S THEOREM, CENTROID (TRIANGLE), CIRCLE, EULER LINE, INCENTER, ORTHOCENTER

References

Carr, G. S. *Formulas and Theorems in Pure Mathematics, 2nd ed.* New York: Chelsea, p. 623, 1970.

Dixon, R. *Mathographics.* New York: Dover, p. 55, 1991.

Eppstein, D. "Circumcenters of Triangles." `http://www.ics.uci.edu/~eppstein/junkyard/circumcenter.html`.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, 1929.

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Kimberling, C. "Circumcenter." `http://www.evansville.edu/~ck6/tcenters/class/ccenter.html`.

Circumcircle

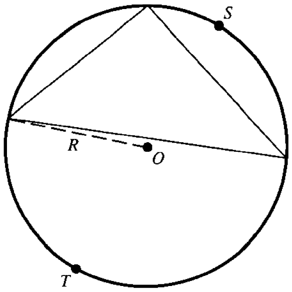

A TRIANGLE'S circumscribed circle. Its center O is called the CIRCUMCENTER, and its RADIUS R the CIRCUMRADIUS. The circumcircle can be specified using TRILINEAR COORDINATES as

$$\beta\gamma a + \gamma\alpha b + \alpha\beta c = 0. \tag{1}$$

The STEINER POINT S and TARRY POINT T lie on the circumcircle.

A GEOMETRIC CONSTRUCTION for the circumcircle is given by Pedoe (1995, pp. xii–xiii). The equation for the circumcircle of the TRIANGLE with VERTICES (x_i, y_i) for $i = 1, 2, 3$ is

$$\begin{vmatrix} x^2+y^2 & x & y & 1 \\ {x_1}^2+{y_1}^2 & x_1 & y_1 & 1 \\ {x_2}^2+{y_2}^2 & x_2 & y_2 & 1 \\ {x_3}^2+{y_3}^2 & x_3 & y_3 & 1 \end{vmatrix} = 0. \tag{2}$$

Expanding the DETERMINANT,

$$a(x^2+y^2) + 2dx + 2fy + g = 0, \tag{3}$$

where

$$a = \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} \tag{4}$$

$$d = -\tfrac{1}{2}\begin{vmatrix} {x_1}^2+{y_1}^2 & y_1 & 1 \\ {x_2}^2+{y_2}^2 & y_2 & 1 \\ {x_3}^2+{y_3}^2 & y_3 & 1 \end{vmatrix} \tag{5}$$

$$f = \tfrac{1}{2}\begin{vmatrix} {x_1}^2+{y_1}^2 & x_1 & 1 \\ {x_2}^2+{y_2}^2 & x_2 & 1 \\ {x_3}^2+{y_3}^2 & x_3 & 1 \end{vmatrix} \tag{6}$$

$$g = -\begin{vmatrix} {x_1}^2+{y_1}^2 & x_1 & y_1 \\ {x_2}^2+{y_2}^2 & x_2 & y_2 \\ {x_3}^2+{y_3}^2 & x_3 & y_3 \end{vmatrix}. \tag{7}$$

COMPLETING THE SQUARE gives

$$a\left(x+\frac{d}{a}\right)^2 + a\left(y+\frac{f}{a}\right)^2 - \frac{d^2}{a} - \frac{f^2}{a} + g = 0 \tag{8}$$

which is a CIRCLE of the form

$$(x-x_0)^2 + (y-y_0)^2 = r^2, \tag{9}$$

with CIRCUMCENTER

$$x_0 = -\frac{d}{a} \tag{10}$$

$$y_0 = -\frac{f}{a} \tag{11}$$

and CIRCUMRADIUS

$$r = \sqrt{\frac{f^2+d^2}{a^2} - \frac{g}{a}}. \tag{12}$$

see also CIRCLE, CIRCUMCENTER, CIRCUMRADIUS, EXCIRCLE, INCIRCLE, PARRY POINT, PURSER'S THEOREM, STEINER POINTS, TARRY POINT

References

Pedoe, D. *Circles: A Mathematical View, rev. ed.* Washington, DC: Math. Assoc. Amer., 1995.

Circumference

The PERIMETER of a CIRCLE. For RADIUS r or DIAMETER $d = 2r$,

$$C = 2\pi r = \pi d,$$

where π is PI.

see also CIRCLE, DIAMETER, PERIMETER, PI, RADIUS

Circuminscribed

Given two closed curves, the circuminscribed curve is simultaneously INSCRIBED in the outer one and CIRCUMSCRIBED on the inner one.

see also PONCELET'S CLOSURE THEOREM

Circumradius

The radius of a TRIANGLE'S CIRCUMCIRCLE or of a POLYHEDRON'S CIRCUMSPHERE, denoted R. For a TRIANGLE,

$$R = \frac{abc}{\sqrt{(a+b+c)(b+c-a)(c+a-b)(a+b-c)}}, \tag{1}$$

where the side lengths of the TRIANGLE are a, b, and c.

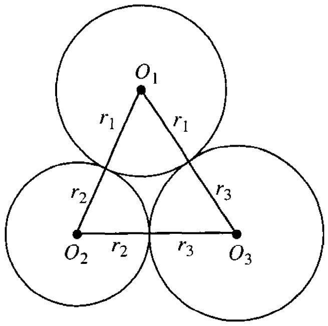

This equation can also be expressed in terms of the RADII of the three mutually tangent CIRCLES centered at the TRIANGLE'S VERTICES. Relabeling the diagram for the SODDY CIRCLES with VERTICES O_1, O_2, and O_3 and the radii r_1, r_2, and r_3, and using

$$a = r_1 + r_2 \tag{2}$$

$$b = r_2 + r_3 \tag{3}$$

$$c = r_1 + r_3 \tag{4}$$

then gives

$$R = \frac{(r_1+r_2)(r_1+r_3)(r_2+r_3)}{4\sqrt{r_1 r_2 r_3 (r_1+r_2+r_3)}}. \tag{5}$$

If O is the CIRCUMCENTER and M is the triangle CENTROID, then

$$\overline{OM}^2 = R^2 - \tfrac{1}{9}(a^2+b^2+c^2). \tag{6}$$

$$Rr = \frac{a_1 a_2 a_3}{4s} \tag{7}$$

$$\cos\alpha_1 + \cos\alpha_2 + \cos\alpha_3 = 1 + \frac{r}{R} \tag{8}$$

$$r = 2R\cos\alpha_1\cos\alpha_2\cos\alpha_3 \tag{9}$$

$$a_1{}^2 + a_2{}^2 + a_3{}^2 = 4rR + 8R^2 \tag{10}$$

(Johnson 1929, pp. 189–191). Let d be the distance between INRADIUS r and circumradius R, $d = \overline{rR}$. Then

$$R^2 - d^2 = 2Rr \tag{11}$$

$$\frac{1}{R-d} + \frac{1}{R+d} = \frac{1}{r} \tag{12}$$

(Mackay 1886–87). These and many other identities are given in Johnson (1929, pp. 186–190).

For an ARCHIMEDEAN SOLID, expressing the circumradius in terms of the INRADIUS r and MIDRADIUS ρ gives

$$R = \tfrac{1}{2}(r + \sqrt{r^2 + a^2}\,) \tag{13}$$

$$= \sqrt{\rho^2 + \tfrac{1}{4}a^2} \tag{14}$$

for an ARCHIMEDEAN SOLID.

see also CARNOT'S THEOREM, CIRCUMCIRCLE, CIRCUMSPHERE

References
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, 1929.
Mackay, J. S. "Historical Notes on a Geometrical Theorem and its Developments [18th Century]." *Proc. Edinburgh Math. Soc.* **5**, 62–78, 1886–1887.

Circumscribed

A geometric figure which touches only the VERTICES (or other extremities) of another figure.

see also CIRCUMCENTER, CIRCUMCIRCLE, CIRCUMINSCRIBED, CIRCUMRADIUS, INSCRIBED

Circumsphere

A SPHERE circumscribed in a given solid. Its radius is called the CIRCUMRADIUS.

see also INSPHERE

Cis

$$\operatorname{Cis} x \equiv e^{ix} = \cos x + i \sin x.$$

Cissoid

Given two curves C_1 and C_2 and a fixed point O, let a line from O cut C at Q and C at R. Then the LOCUS of a point P such that $OP = QR$ is the cissoid. The word cissoid means "ivy shaped."

Curve 1	Curve 2	Pole	Cissoid
line	parallel line	any point	line
line	circle	center	conchoid of Nicomedes
circle	tangent line	on C	oblique cissoid
circle	tangent line	on C opp. tangent	cissoid of Diocles
circle	radial line	on C	strophoid
circle	concentric circle	center	circle
circle	same circle	$(a\sqrt{2}, 0)$	lemniscate

see also CISSOID OF DIOCLES

References
Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 53–56 and 205, 1972.
Lee, X. "Cissoid." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Cissoid_dir/cissoid.html`.
Lockwood, E. H. "Cissoids." Ch. 15 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 130–133, 1967.
Yates, R. C. "Cissoid." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 26–30, 1952.

Cissoid of Diocles

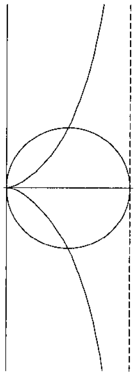

A curve invented by Diocles in about 180 BC in connection with his attempt to duplicate the cube by geometrical methods. The name "cissoid" first appears in the work of Geminus about 100 years later. Fermat and Roberval constructed the tangent in 1634. Huygens and Wallis found, in 1658, that the AREA between the curve and its asymptote was $3a$ (MacTutor Archive). From a given point there are either one or three TANGENTS to the cissoid.

Given an origin O and a point P on the curve, let S be the point where the extension of the line OP intersects the line $x = 2a$ and R be the intersection of the CIRCLE of RADIUS a and center $(a, 0)$ with the extension of OP. Then the cissoid of Diocles is the curve which satisfies $OP = RS$.

The cissoid of Diocles is the ROULETTE of the VERTEX of a PARABOLA rolling on an equal PARABOLA. Newton gave a method of drawing the cissoid of Diocles using two line segments of equal length at RIGHT ANGLES. If they are moved so that one line always passes through a fixed point and the end of the other line segment slides along a straight line, then the MIDPOINT of the sliding line segment traces out a cissoid of Diocles.

The cissoid of Diocles is given by the parametric equations

$$x = 2a\sin^2\theta \tag{1}$$

$$y = \frac{2a\sin^3\theta}{\cos\theta}. \tag{2}$$

Converting these to POLAR COORDINATES gives

$$\begin{aligned} r^2 = x^2 + y^2 &= 4a^2\left(\sin^4\theta + \frac{\sin^6\theta}{\cos^2\theta}\right) \\ &= 4a^2\sin^4\theta(1+\tan^2\theta) = 4a^2\sin^4\theta\sec^2\theta, \end{aligned} \tag{3}$$

so

$$r = 2a\sin^2\theta\sec\theta = 2a\sin\theta\tan\theta. \tag{4}$$

In CARTESIAN COORDINATES,

$$\begin{aligned} \frac{x^3}{2a-x} = \frac{8a^3\sin^6\theta}{2a-2a\sin^2\theta} &= 4a^2\frac{\sin^6\theta}{1-\sin^2\theta} \\ &= 4a^2\frac{\sin^6\theta}{\cos^2\theta} = y^2. \end{aligned} \tag{5}$$

An equivalent form is

$$x(x^2+y^2) = 2ay^2. \tag{6}$$

Using the alternative parametric form

$$x(t) = \frac{2at^2}{1+t^2} \tag{7}$$

$$y(t) = \frac{2at^3}{1+t^2} \tag{8}$$

(Gray 1993), gives the CURVATURE as

$$\kappa(t) = \frac{3}{a|t|(t^2+4)^{3/2}}. \tag{9}$$

References

Gray, A. "The Cissoid of Diocles." §3.4 in *Modern Differential Geometry of Curves and Surfaces.*Boca Raton, FL: CRC Press, pp. 43–46, 1993.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 98–100, 1972.

Lee, X. "Cissoid of Diocles." `http://www.best.com/~xah/SpecialPlaneCurves_dir/CissoidOfDiocles_dir/cissoidOfDiocles.html`.

Lockwood, E. H. *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 130–133, 1967.

MacTutor History of Mathematics Archive. "Cissoid of Diocles." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Cissoid.html`.

Yates, R. C. "Cissoid." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 26–30, 1952.

Cissoid of Diocles Caustic

The CAUSTIC of the cissoid where the RADIANT POINT is taken as $(8a, 0)$ is a CARDIOID.

Cissoid of Diocles Inverse Curve

If the cusp of the CISSOID OF DIOCLES is taken as the INVERSION CENTER, then the cissoid inverts to a PARABOLA.

Cissoid of Diocles Pedal Curve

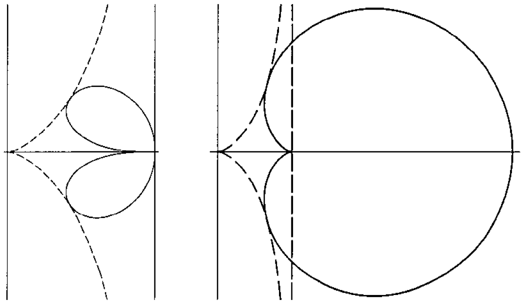

The PEDAL CURVE of the cissoid, when the PEDAL POINT is on the axis beyond the ASYMPTOTE at a distance from the cusp which is four times that of the ASYMPTOTE is a CARDIOID.

Clairaut's Differential Equation

$$y = x\frac{dy}{dx} + f\left(\frac{dy}{dx}\right)$$

or

$$y = px + f(p),$$

where f is a FUNCTION of one variable and $p \equiv dy/dx$. The general solution is $y = cx + f(c)$. The singular solution ENVELOPES are $x = -f'(c)$ and $y = f(c) - cf'(c)$.

see also D'ALEMBERT'S EQUATION

References

Boyer, C. B. *A History of Mathematics.* New York: Wiley, p. 494, 1968.

Clarity

The RATIO of a measure of the size of a "fit" to the size of a "residual."

References

Tukey, J. W. *Explanatory Data Analysis.* Reading, MA: Addison-Wesley, p. 667, 1977.

Clark's Triangle

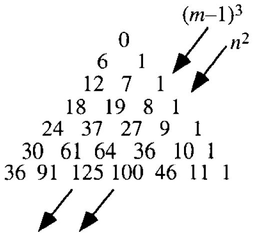

A NUMBER TRIANGLE created by setting the VERTEX equal to 0, filling one diagonal with 1s, the other diagonal with multiples of an INTEGER f, and filling in the remaining entries by summing the elements on either side from one row above. Call the first column $n = 0$ and the last column $m = n$ so that

$$c(m,0) = fm \tag{1}$$

$$c(m,m) = 1, \tag{2}$$

then use the RECURRENCE RELATION

$$c(m,n) = c(m-1,n-1) + c(m-1,n) \tag{3}$$

to compute the rest of the entries. For $n = 1$, we have

$$c(m,1) = c(m-1,0) + c(m-1,1) \tag{4}$$

$$c(m,1) - c(m-1,1) = c(m-1,0) = f(m-1). \tag{5}$$

For arbitrary m, the value can be computed by SUMMING this RECURRENCE,

$$c(m,1) = f\left(\sum_{k=1}^{m-1} k\right) + 1 = \tfrac{1}{2}fm(m-1) + 1. \tag{6}$$

Now, for $n = 2$ we have

$$c(m,2) = c(m-1,1) + c(m-1,2) \tag{7}$$

$$c(m,2) - c(m-1,2) = c(m-1,1) = \tfrac{1}{2}f(m-1)m + 1, \tag{8}$$

so SUMMING the RECURRENCE gives

$$\begin{aligned} c(m,2) &= \sum_{k=1}^{m}[\tfrac{1}{2}fk(k-1)+1] = \sum_{k=1}^{m}(\tfrac{1}{2}fk^2 - \tfrac{1}{2}fk + 1) \\ &= \tfrac{1}{2}f[\tfrac{1}{6}m(m+1)(2m+1)] - \tfrac{1}{2}f[\tfrac{1}{2}m(m+1)] + m \\ &= \tfrac{1}{6}(m-1)(fm^2 - 2fm + 6). \end{aligned} \tag{9}$$

Similarly, for $n = 3$ we have

$$\begin{aligned} c(m,3) - c(m-1,3) &= c(m-1,2) \\ &= \tfrac{1}{6}fm^3 - fm^2 + (\tfrac{11}{6}f + 1)m - (f+2). \end{aligned} \tag{10}$$

Taking the SUM,

$$c(m,3) = \sum_{k=2}^{m} \tfrac{1}{6}fk^3 - fk^2 + (\tfrac{11}{6}f+1)k - (f+2). \tag{11}$$

Evaluating the SUM gives

$$c(m,3) = \tfrac{1}{24}(m-1)(m-2)(fm^2 - 3fm + 12). \tag{12}$$

So far, this has just been relatively boring ALGEBRA. But the amazing part is that if $f = 6$ is chosen as the INTEGER, then $c(m,2)$ and $c(m,3)$ simplify to

$$\begin{aligned} c(m,2) &= \tfrac{1}{6}(m-1)(6m^2 - 12m + 6) \\ &= (m-1)^3 \end{aligned} \tag{13}$$

$$c(m,3) = \tfrac{1}{4}(m-1)^2(m-2)^2, \tag{14}$$

which are consecutive CUBES $(m-1)^3$ and nonconsecutive SQUARES $n^2 = [(m-1)(m-2)/2]^2$.

see also BELL TRIANGLE, CATALAN'S TRIANGLE, EULER'S TRIANGLE, LEIBNIZ HARMONIC TRIANGLE, NUMBER TRIANGLE, PASCAL'S TRIANGLE, SEIDEL-ENTRINGER-ARNOLD TRIANGLE, SUM

References

Clark, J. E. "Clark's Triangle." *Math. Student* **26**, No. 2, p. 4, Nov. 1978.

Class

see CHARACTERISTIC CLASS, CLASS INTERVAL, CLASS (MULTIPLY PERFECT NUMBER), CLASS NUMBER, CLASS (SET), CONJUGACY CLASS

Class (Group)

see CONJUGACY CLASS

Class Interval

The constant bin size in a HISTOGRAM.

see also SHEPPARD'S CORRECTION

Class (Map)

A MAP $u : \mathbb{R}^n \to \mathbb{R}^n$ from a DOMAIN G is called a map of class C^r if each component of

$$u(x) = (u_1(x_1, \ldots, x_n), \ldots, u_m(x_1, \ldots, x_n))$$

is of class C^r $(0 \leq r \leq \infty$ or $r = \omega)$ in G, where C^d denotes a continuous function which is differentiable d times.

Class (Multiply Perfect Number)

The number k in the expression $s(n) = kn$ for a MULTIPLY PERFECT NUMBER is called its class.

Class Number

For any IDEAL I, there is an IDEAL I_i such that

$$II_i = z, \tag{1}$$

where z is a PRINCIPAL IDEAL, (i.e., an IDEAL of rank 1). Moreover, there is a finite list of ideals I_i such that this equation may be satisfied for every I. The size of this list is known as the class number. When the class number is 1, the RING corresponding to a given IDEAL has unique factorization and, in a sense, the class number is a measure of the failure of unique factorization in the original number ring.

A finite series giving exactly the class number of a RING is known as a CLASS NUMBER FORMULA. A CLASS NUMBER FORMULA is known for the full ring of cyclotomic integers, as well as for any subring of the cyclotomic integers. Finding the class number is a computationally difficult problem.

Let $h(d)$ denote the class number of a quadratic ring, corresponding to the BINARY QUADRATIC FORM

$$ax^2 + bxy + cy^2, \tag{2}$$

with DISCRIMINANT

$$d \equiv b^2 - 4ac. \tag{3}$$

Then the class number $h(d)$ for DISCRIMINANT d gives the number of possible factorizations of $ax^2 + bxy + cy^2$ in the QUADRATIC FIELD $\mathbb{Q}(\sqrt{d})$. Here, the factors are of the form $x + y\sqrt{d}$, with x and y half INTEGERS.

Some fairly sophisticated mathematics shows that the class number for discriminant d can be given by the CLASS NUMBER FORMULA

$$h(d) \equiv \begin{cases} -\frac{1}{2\ln\eta}\sum_{r=1}^{d-1}(d|r)\ln\sin\left(\frac{\pi r}{d}\right) & \text{for } d > 0 \\ -\frac{w(d)}{2|d|}\sum_{r=1}^{|d|-1}(d|r)r & \text{for } d < 0, \end{cases} \tag{4}$$

where $(d|r)$ is the KRONECKER SYMBOL, $\eta(d)$ is the FUNDAMENTAL UNIT, $w(d)$ is the number of substitutions which leave the BINARY QUADRATIC FORM unchanged

$$w(d) = \begin{cases} 6 & \text{for } d = -3 \\ 4 & \text{for } d = -4 \\ 2 & \text{otherwise,} \end{cases} \tag{5}$$

and the sums are taken over all terms where the KRONECKER SYMBOL is defined (Cohn 1980). The class number for $d > 0$ can also be written

$$\eta^{2h(d)} = \prod_{r=1}^{d-1} \sin^{-(d|r)}\left(\frac{\pi r}{d}\right) \tag{6}$$

for $d > 0$, where the PRODUCT is taken over terms for which the KRONECKER SYMBOL is defined.

The class number is related to the DIRICHLET L-SERIES by

$$h(d) = \frac{L_d(1)}{\kappa(d)}, \tag{7}$$

where $\kappa(d)$ is the DIRICHLET STRUCTURE CONSTANT.

Wagner (1996) shows that class number $h(-d)$ satisfies the INEQUALITY

$$h(-d) > \frac{1}{55}\prod_{p|d}{}^{*}\left(1 - \frac{\lfloor 2\sqrt{p}\rfloor}{p+1}\right)\ln d, \tag{8}$$

for $-d < 0$, where $\lfloor x\rfloor$ is the FLOOR FUNCTION, the product is over PRIMES dividing d, and the $*$ indicates that the GREATEST PRIME FACTOR of d is omitted from the product.

The *Mathematica*® (Wolfram Research, Champaign, IL) function `NumberTheory'NumberTheoryFunctions' ClassNumber[n]` gives the class number $h(d)$ for d a NEGATIVE SQUAREFREE number of the form $4k + 1$.

GAUSS'S CLASS NUMBER PROBLEM asks to determine a complete list of fundamental DISCRIMINANTS $-d$ such that the CLASS NUMBER is given by $h(-d) = m$ for a given m. This problem has been solved for $n \leq 7$ and ODD $n \leq 23$. Gauss conjectured that the class number $h(-d)$ of an IMAGINARY quadratic field with DISCRIMINANT $-d$ tends to infinity with d, an assertion now known as GAUSS'S CLASS NUMBER CONJECTURE.

The discriminants d having $h(-d) = 1, 2, 3, 4, 5, \ldots$ are Sloane's A014602 (Cohen 1993, p. 229; Cox 1997, p. 271), Sloane's A014603 (Cohen 1993, p. 229), Sloane's A006203 (Cohen 1993, p. 504), Sloane's A013658 (Cohen 1993, p. 229), Sloane's A046002, Sloane's A046003, ... The complete set of negative discriminants having class numbers 1–5 and ODD 7–23 are known. Buell (1977) gives the smallest and largest fundamental class numbers for $d < 4{,}000{,}000$, partitioned into EVEN discriminants, discriminants 1 (mod 8), and discriminants 5 (mod 8). Arno *et al.* (1993) give complete lists of values of d with $h(-d) = k$ for ODD $k = 5, 7, 9, \ldots, 23$. Wagner gives complete lists of values for $k = 5$, 6, and 7.

Lists of NEGATIVE discriminants corresponding to IMAGINARY QUADRATIC FIELDS $\mathbb{Q}(\sqrt{-d(n)})$ having small class numbers $h(-d)$ are given in the table below. In the table, N is the number of "fundamental" values of $-d$ with a given class number $h(-d)$, where "fundamental" means that $-d$ is not divisible by any SQUARE NUMBER s^2 such that $h(-d/s^2) < h(-d)$. For example, although $h(-63) = 2$, -63 is not a fundamental discriminant since $63 = 3^2 \cdot 7$ and $h(-63/3^2) = h(-7) = 1 < h(-63)$. EVEN values $8 \leq h(-d) \leq 18$ have been computed by Weisstein. The number of negative discriminants having class number 1, 2, 3, ... are 9, 18, 16, 54, 25, 51, 31, ... (Sloane's A046125). The largest negative discriminants having class numbers 1, 2, 3, ... are 163, 427, 907, 1555, 2683, ... (Sloane's A038552).

The following table lists the numbers with small class numbers ≤ 11. Lists including larger class numbers are given by Weisstein.

$h(-d)$	N	d
1	9	3, 4, 7, 8, 11, 19, 43, 67, 163
2	18	15, 20, 24, 35, 40, 51, 52, 88, 91, 115, 123, 148, 187, 232, 235, 267, 403, 427
3	16	23, 31, 59, 83, 107, 139, 211, 283, 307, 331, 379, 499, 547, 643, 883, 907

$h(-d)$	N	d
4	54	39, 55, 56, 68, 84, 120, 132, 136, 155, 168, 184, 195, 203, 219, 228, 259, 280, 291, 292, 312, 323, 328, 340, 355, 372, 388, 408, 435, 483, 520, 532, 555, 568, 595, 627, 667, 708, 715, 723, 760, 763, 772, 795, 955, 1003, 1012, 1027, 1227, 1243, 1387, 1411, 1435, 1507, 1555
5	25	47, 79, 103, 127, 131, 179, 227, 347, 443, 523, 571, 619, 683, 691, 739, 787, 947, 1051, 1123, 1723, 1747, 1867, 2203, 2347, 2683
6	51	87, 104, 116, 152, 212, 244, 247, 339, 411, 424, 436, 451, 472, 515, 628, 707, 771, 808, 835, 843, 856, 1048, 1059, 1099, 1108, 1147, 1192, 1203, 1219, 1267, 1315, 1347, 1363, 1432, 1563, 1588, 1603, 1843, 1915, 1963, 2227, 2283, 2443, 2515, 2563, 2787, 2923, 3235, 3427, 3523, 3763
7	31	71, 151, 223, 251, 463, 467, 487, 587, 811, 827, 859, 1163, 1171, 1483, 1523, 1627, 1787, 1987, 2011, 2083, 2179, 2251, 2467, 2707, 3019, 3067, 3187, 3907, 4603, 5107, 5923
8	131	95, 111, 164, 183, 248, 260, 264, 276, 295, 299, 308, 371, 376, 395, 420, 452, 456, 548, 552, 564, 579, 580, 583, 616, 632, 651, 660, 712, 820, 840, 852, 868, 904, 915, 939, 952, 979, 987, 995, 1032, 1043, 1060, 1092, 1128, 1131, 1155, 1195, 1204, 1240, 1252, 1288, 1299, 1320, 1339, 1348, 1380, 1428, 1443, 1528, 1540, 1635, 1651, 1659, 1672, 1731, 1752, 1768, 1771, 1780, 1795, 1803, 1828, 1848, 1864, 1912, 1939, 1947, 1992, 1995, 2020, 2035, 2059, 2067, 2139, 2163, 2212, 2248, 2307, 2308, 2323, 2392, 2395, 2419, 2451, 2587, 2611, 2632, 2667, 2715, 2755, 2788, 2827, 2947, 2968, 2995, 3003, 3172, 3243, 3315, 3355, 3403, 3448, 3507, 3595, 3787, 3883, 3963, 4123, 4195, 4267, 4323, 4387, 4747, 4843, 4867, 5083, 5467, 5587, 5707, 5947, 6307
9	34	199, 367, 419, 491, 563, 823, 1087, 1187, 1291, 1423, 1579, 2003, 2803, 3163, 3259, 3307, 3547, 3643, 4027, 4243, 4363, 4483, 4723, 4987, 5443, 6043, 6427, 6763, 6883, 7723, 8563, 8803, 9067, 10627
10	87	119, 143, 159, 296, 303, 319, 344, 415, 488, 611, 635, 664, 699, 724, 779, 788, 803, 851, 872, 916, 923, 1115, 1268, 1384, 1492, 1576, 1643, 1684, 1688, 1707, 1779, 1819, 1835, 1891, 1923, 2152, 2164, 2363, 2452, 2643, 2776, 2836, 2899, 3028, 3091, 3139, 3147, 3291, 3412, 3508, 3635, 3667, 3683, 3811, 3859, 3928, 4083, 4227, 4372, 4435, 4579, 4627, 4852, 4915, 5131, 5163, 5272, 5515, 5611, 5667, 5803, 6115, 6259, 6403, 6667, 7123, 7363, 7387, 7435, 7483, 7627, 8227, 8947, 9307, 10147, 10483, 13843
11	41	167, 271, 659, 967, 1283, 1303, 1307, 1459, 1531, 1699, 2027, 2267, 2539, 2731, 2851, 2971, 3203, 3347, 3499, 3739, 3931, 4051, 5179, 5683, 6163, 6547, 7027, 7507, 7603, 7867, 8443, 9283, 9403, 9643, 9787, 10987, 13003, 13267, 14107, 14683, 15667

The table below gives lists of POSITIVE fundamental discriminants d having small class numbers $h(d)$, corresponding to REAL quadratic fields. All POSITIVE SQUAREFREE values of $d \leq 97$ (for which the KRONECKER SYMBOL is defined) are included.

$h(d)$	d
1	5, 13, 17, 21, 29, 37, 41, 53, 57, 61, 69, 73, 77
2	65

The POSITIVE d for which $h(d) = 1$ is given by Sloane's A014539.

see also CLASS NUMBER FORMULA, DIRICHLET L-SERIES, DISCRIMINANT (BINARY QUADRATIC FORM), GAUSS'S CLASS NUMBER CONJECTURE, GAUSS'S CLASS NUMBER PROBLEM, HEEGNER NUMBER, IDEAL, j-FUNCTION

References

Arno, S. "The Imaginary Quadratic Fields of Class Number 4." *Acta Arith.* **40**, 321–334, 1992.

Arno, S.; Robinson, M. L.; and Wheeler, F. S. "Imaginary Quadratic Fields with Small Odd Class Number." `http://www.math.uiuc.edu/Algebraic-Number-Theory/0009/`.

Buell, D. A. "Small Class Numbers and Extreme Values of L-Functions of Quadratic Fields." *Math. Comput.* **139**, 786–796, 1977.

Cohen, H. *A Course in Computational Algebraic Number Theory.* New York: Springer-Verlag, 1993.

Cohn, H. *Advanced Number Theory.* New York: Dover, pp. 163 and 234, 1980.

Cox, D. A. *Primes of the Form x^2+ny^2: Fermat, Class Field Theory and Complex Multiplication.* New York: Wiley, 1997.

Davenport, H. "Dirichlet's Class Number Formula." Ch. 6 in *Multiplicative Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 43–53, 1980.

Iyanaga, S. and Kawada, Y. (Eds.). "Class Numbers of Algebraic Number Fields." Appendix B, Table 4 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 1494–1496, 1980.

Montgomery, H. and Weinberger, P. "Notes on Small Class Numbers." *Acta. Arith.* **24**, 529–542, 1974.

Sloane, N. J. A. Sequences A014539, A038552, A046125, and A003657/M2332 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Stark, H. M. "A Complete Determination of the Complex Quadratic Fields of Class Number One." *Michigan Math. J.* **14**, 1–27, 1967.
Stark, H. M. "On Complex Quadratic Fields with Class Number Two." *Math. Comput.* **29**, 289–302, 1975.
Wagner, C. "Class Number 5, 6, and 7." *Math. Comput.* **65**, 785–800, 1996.
Weisstein, E. W. "Class Numbers." `http://www.astro.virginia.edu/~eww6n/math/notebooks/ClassNumbers.m`.

Class Number Formula

A class number formula is a finite series giving exactly the CLASS NUMBER of a RING. For a RING of quadratic integers, the class number is denoted $h(d)$, where d is the discriminant. A class number formula is known for the full ring of cyclotomic integers, as well as for any subring of the cyclotomic integers. This formula includes the quadratic case as well as many cubic and higher-order rings.

see also CLASS NUMBER

Class Representative

A set of class representatives is a SUBSET of X which contains exactly one element from each EQUIVALENCE CLASS.

Class (Set)

A class is a special kind of SET invented to get around RUSSELL'S PARADOX while retaining the arbitrary criteria for membership which leads to difficulty for SETS. The members of classes are SETS, but it is possible to have the class C of "all SETS which are not members of themselves" without producing a paradox (since C is a proper class (and not a SET), it is not a candidate for membership in C).

see also AGGREGATE, RUSSELL'S PARADOX, SET

Classical Groups

The four following types of GROUPS,

1. LINEAR GROUPS,
2. ORTHOGONAL GROUPS,
3. SYMPLECTIC GROUPS, and
4. UNITARY GROUPS,

which were studied before more exotic types of groups (such as the SPORADIC GROUPS) were discovered.

see also GROUP, LINEAR GROUP, ORTHOGONAL GROUP, SYMPLECTIC GROUP, UNITARY GROUP

Classification

The classification of a collection of objects generally means that a list has been constructed with exactly one member from each ISOMORPHISM type among the objects, and that tools and techniques can effectively be used to identify any combinatorially given object with its unique representative in the list. Examples of mathematical objects which have been classified include the finite SIMPLE GROUPS and 2-MANIFOLDS but not, for example, KNOTS.

Classification Theorem

The classification theorem of FINITE SIMPLE GROUPS, also known as the ENORMOUS THEOREM, which states that the FINITE SIMPLE GROUPS can be classified completely into

1. CYCLIC GROUPS $\mathbb{Z}_p$ of PRIME ORDER,
2. ALTERNATING GROUPS A_n of degree at least five,
3. LIE-TYPE CHEVALLEY GROUPS $PSL(n,q)$, $PSU(n,q)$, $PsP(2n,q)$, and $P\Omega^\epsilon(n,q)$,
4. LIE-TYPE (TWISTED CHEVALLEY GROUPS or the TITS GROUP) ${}^3D_4(q)$, $E_6(q)$, $E_7(q)$, $E_8(q)$, $F_4(q)$, ${}^2F_4(2^n)'$, $G_2(q)$, ${}^2G_2(3^n)$, ${}^2B(2^n)$,
5. SPORADIC GROUPS M_{11}, M_{12}, M_{22}, M_{23}, M_{24}, $J_2 = HJ$, Suz, HS, McL, Co_3, Co_2, Co_1, He, Fi_{22}, Fi_{23}, Fi'_{24}, HN, Th, B, M, J_1, $O'N$, J_3, Ly, Ru, J_4.

The "PROOF" of this theorem is spread throughout the mathematical literature and is estimated to be approximately 15,000 pages in length.

see also FINITE GROUP, GROUP, j-FUNCTION, SIMPLE GROUP

References
Cartwright, M. "Ten Thousand Pages to Prove Simplicity." *New Scientist* **109**, 26–30, 1985.
Cipra, B. "Are Group Theorists Simpleminded?" *What's Happening in the Mathematical Sciences, 1995–1996, Vol. 3.* Providence, RI: Amer. Math. Soc., pp. 82–99, 1996.
Cipra, B. "Slimming an Outsized Theorem." *Science* **267**, 794–795, 1995.
Gorenstein, D. "The Enormous Theorem." *Sci. Amer.* **253**, 104–115, Dec. 1985.
Solomon, R. "On Finite Simple Groups and Their Classification." *Not. Amer. Math. Soc.* **42**, 231–239, 1995.

Clausen Formula

Clausen's ${}_4F_3$ identity

$${}_4F_3\left[\begin{matrix} a & b & c & d \\ e & f & g & \end{matrix};1\right] = \frac{(2a)_{|d|}(a+b)_{|d|}(2b)_{|d|}}{(2a+2b)_{|d|}a_{|d|}b_{|d|}},$$

holds for $a+b+c-d=1/2$, $e=a+b+1/2$, $a+f=d+1=b+g$, d a nonpositive integer, and $(a)_n$ is the POCHHAMMER SYMBOL (Petkovšek *et al.* 1996).

Another identity ascribed to Clausen which involves the HYPERGEOMETRIC FUNCTION ${}_2F_1(a,b;c;z)$ and the GENERALIZED HYPERGEOMETRIC FUNCTION ${}_3F_2(a,b,c;d,e;z)$ is given by

$$\left[{}_2F_1\left(\begin{matrix} a,b \\ a+b+\frac{1}{2} \end{matrix};x\right)\right]^2 = {}_3F_2\left(\begin{matrix} 2a, a+b, 2b \\ a+b+\frac{1}{2}, 2a+2b \end{matrix};x\right).$$

see also GENERALIZED HYPERGEOMETRIC FUNCTION, HYPERGEOMETRIC FUNCTION

References
Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B*. Wellesley, MA: A. K. Peters, pp. 43 and 127, 1996.

Clausen Function

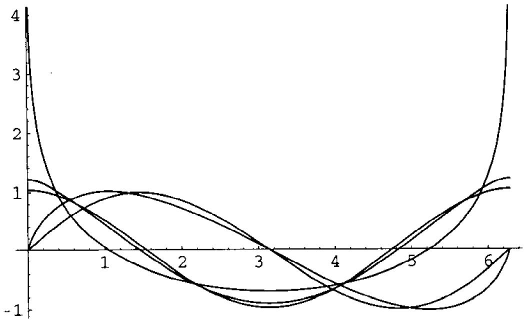

Define

$$S_n(x) \equiv \sum_{k=1}^{\infty} \frac{\sin(kx)}{k^n} \qquad (1)$$

$$C_n(x) \equiv \sum_{k=1}^{\infty} \frac{\cos(kx)}{k^n}, \qquad (2)$$

and write

$$\mathrm{Cl}_n(x) \equiv \begin{cases} S_n(x) = \sum_{k=1}^{\infty} \frac{\sin(kx)}{k^n} & n \text{ even} \\ C_n(x) = \sum_{k=1}^{\infty} \frac{\cos(kx)}{k^n} & n \text{ odd.} \end{cases} \qquad (3)$$

Then the Clausen function $\mathrm{Cl}_n(x)$ can be given symbolically in terms of the POLYLOGARITHM as

$$\mathrm{Cl}_n(x) = \begin{cases} \frac{1}{2}i[\mathrm{Li}_n(e^{-ix}) - \mathrm{Li}_n(e^{ix})] & n \text{ even} \\ \frac{1}{2}[\mathrm{Li}_n(e^{-ix}) + \mathrm{Li}_n(e^{ix})] & n \text{ odd.} \end{cases}$$

For $n = 1$, the function takes on the special form

$$\mathrm{Cl}_1(x) = C_1(x) = -\ln|2\sin(\tfrac{1}{2}x)| \qquad (4)$$

and for $n = 2$, it becomes CLAUSEN'S INTEGRAL

$$\mathrm{Cl}_2(x) = S_2(x) = -\int_0^x \ln[2\sin(\tfrac{1}{2}t)]\,dt. \qquad (5)$$

The symbolic sums of opposite parity are summable symbolically, and the first few are given by

$$C_2(x) = \tfrac{1}{6}\pi^2 - \tfrac{1}{2}\pi x + \tfrac{1}{4}x^2 \qquad (6)$$

$$C_4(x) = \tfrac{1}{90} - \tfrac{1}{12}\pi^2 x^2 + \tfrac{1}{12}\pi x^3 - \tfrac{1}{48}x^4 \qquad (7)$$

$$S_1(x) = \tfrac{1}{2}(\pi - x) \qquad (8)$$

$$S_3(x) = \tfrac{1}{6}\pi^2 x - \tfrac{1}{4}\pi x^2 + \tfrac{1}{12}x^3 \qquad (9)$$

$$S_5(x) = \tfrac{1}{90}\pi^4 x - \tfrac{1}{36}\pi^2 x^3 + \tfrac{1}{48}\pi x^4 - \tfrac{1}{240}x^5 \qquad (10)$$

for $0 \leq x \leq 2\pi$ (Abramowitz and Stegun 1972).

see also CLAUSEN'S INTEGRAL, POLYGAMMA FUNCTION, POLYLOGARITHM

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Clausen's Integral and Related Summations" §27.8 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 1005–1006, 1972.

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 783, 1985.

Clausen, R. "Über die Zerlegung reeller gebrochener Funktionen." *J. reine angew. Math.* **8**, 298–300, 1832.

Grosjean, C. C. "Formulae Concerning the Computation of the Clausen Integral $\mathrm{Cl}_2(\alpha)$." *J. Comput. Appl. Math.* **11**, 331–342, 1984.

Jolley, L. B. W. *Summation of Series.* London: Chapman, 1925.

Wheelon, A. D. *A Short Table of Summable Series.* Report No. SM-14642. Santa Monica, CA: Douglas Aircraft Co., 1953.

Clausen's Integral

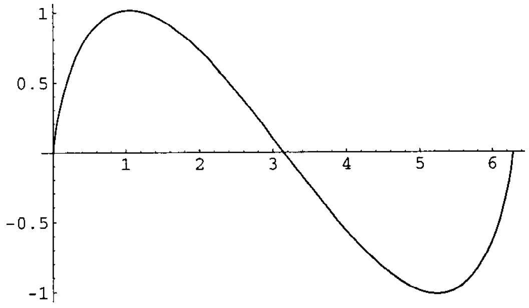

The CLAUSEN FUNCTION

$$\mathrm{Cl}_2(\theta) = -\int_0^\theta \ln[2\sin(\tfrac{1}{2}t)]\,dt.$$

see also CLAUSEN FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 1005–1006, 1972.

Ashour, A. and Sabri, A. "Tabulation of the Function $\psi(\theta) = \sum_{n=1}^{\infty} \frac{\sin(n\theta)}{n^2}$." *Math. Tables Aids Comp.* **10**, 54 and 57–65, 1956.

Clausen, R. "Über die Zerlegung reeller gebrochener Funktionen." *J. reine angew. Math.* **8**, 298–300, 1832.

CLEAN Algorithm

An iterative algorithm which DECONVOLVES a sampling function (the "DIRTY BEAM") from an observed brightness ("DIRTY MAP") of a radio source. This algorithm is of fundamental importance in radio astronomy, where it is used to create images of astronomical sources which are observed using arrays of radio telescopes ("synthesis imaging"). As a result of the algorithm's importance to synthesis imaging, a great deal of effort has gone into optimizing and adjusting the ALGORITHM. CLEAN is a nonlinear algorithm, since linear DECONVOLUTION algorithms such as WIENER FILTERING and inverse filtering are inapplicable to applications with invisible distributions (i.e., incomplete sampling of the spatial frequency plane) such as map obtained in synthesis imaging.

The basic CLEAN method was developed by Högbom (1974). It was originally designed for point sources, but it has been found to work well for extended sources as well when given a reasonable starting model. The Högbom CLEAN constructs discrete approximations I_n to the CLEAN MAP in the (ξ, η) plane from the CONVOLUTION equation

$$b' * I = I', \tag{1}$$

where b' is the DIRTY BEAM, I' is the DIRTY MAP (both in the (ξ, η) PLANE), and $f * g$ denotes a CONVOLUTION.

The CLEAN algorithm starts with an initial approximation $I_0 = 0$. At the nth iteration, it then searches for the largest value in the residual map

$$I_n = I' - b' * I_{n-1}. \tag{2}$$

A DELTA FUNCTION is then centered at the location of the largest residual flux and given an amplitude μ (the so-called "LOOP GAIN") times this value. An antenna's response to the DELTA FUNCTION, the DIRTY BEAM, is then subtracted from I_{n-1} to yield I_n. Iteration continues until a specified iteration limit N is reached, or until the peak residual or ROOT-MEAN-SQUARE residual decreases to some level. The resulting final map is denoted I_N, and the position of each DELTA FUNCTION is saved in a "CLEAN component" table in the CLEAN MAP file. At the point where component subtraction is stopped, it is assumed that the residual brightness distribution consists mainly of NOISE.

To diminish high spatial frequency features which may be spuriously extrapolated from the measured data, each CLEAN component is convolved with the so-called CLEAN BEAM b, which is simply a suitably smoothed version of the sampling function ("DIRTY BEAM"). Usually, a GAUSSIAN is used. A good CLEAN BEAM should:

1. Have a unity FOURIER TRANSFORM inside the sampled region of (u, v) space,
2. Have a FOURIER TRANSFORM which tends to 0 outside the sampled (u, v) region as quickly as possible, and
3. Not have any effects produced by NEGATIVE sidelobes larger than the NOISE level.

A CLEAN MAP is produced when the final residual map is added to the the approximate solution,

$$[\text{clean map}] = I_N * b + [I' - b' * I_N] \tag{3}$$

in order to include the NOISE.

CLEAN will always converge to one (of possibly many) solutions if the following three conditions are satisfied (Schwarz 1978):

1. The beam must be symmetric.
2. The FOURIER TRANSFORM of the DIRTY BEAM is NONNEGATIVE (positive definite or positive semidefinite).
3. There must be no spatial frequencies present in the dirty image which are not also present in the DIRTY BEAM.

These conditions are almost always satisfied in practice. If the number of CLEAN components does not exceed the number of independent (u, v) points, CLEAN converges to a solution which is the least squares fit of the FOURIER TRANSFORMS of the DELTA FUNCTION components to the measured visibility (Thompson *et al.* 1986, p. 347). Schwarz claims that the CLEAN algorithm is equivalent to a least squares fitting of cosine and sine parts in the (u, v) plane of the visibility data. Schwab has produced a NOISE analysis of the CLEAN algorithm in the case of least squares minimization of a noiseless image which involves an $N \times M$ MATRIX. However, no NOISE analysis has been performed for a REAL image.

Poor modulation of short spacings results in an underestimation of the flux, which is manifested in a bowl of negative surface brightness surrounding an object. Providing an estimate of the "zero spacing" flux (the total flux of the source, which cannot be directly measured by an interferometer) can considerably reduce this effect. Modulations or stripes can occur at spatial frequencies corresponding to undersampled parts of the (u, v) plane. This can result in a golf ball-like mottling for disk sources such as planets, or a corrugated pattern of parallel lines of peaks and troughs ("stripes"). A more accurate model can be used to suppress the "golf ball" modulations, but may not eliminate the corrugations. A tapering function which deemphasizes data near $(u, v) = (0, 0)$ can also be used. Stripes can sometimes be eliminated using the Cornwell smoothness-stabilized CLEAN (a.k.a. Prussian helmet algorithm; Thompson *et al.* 1986). CLEANing part way, then restarting the CLEAN also seems to eliminate the stripes, although this fact is more disturbing than reassuring. Stability the the CLEAN algorithm is discussed by Tan (1986).

In order to CLEAN a map of a given dimension, it is necessary to have a beam pattern twice as large so a point source can be subtracted from any point in the map. Because the CLEAN algorithm uses a FAST FOURIER TRANSFORM, the size must also be a POWER of 2.

There are many variants of the basic Högbom CLEAN which extend the method to achieve greater speed and produce more realistic maps. Alternate nonlinear DECONVOLUTION methods, such as the MAXIMUM ENTROPY METHOD, may also be used, but are generally slower than the CLEAN technique. The Astronomical Image Processing Software (AIPS) of the National Radio Astronomical Observatory includes 2-D

DECONVOLUTION algorithms in the tasks **DCONV** and **UVMAP**. Among the variants of the basic Högbom CLEAN are Clark, Cornwell smoothness stabilized (Prussian helmet), Cotton-Schwab, Gerchberg-Saxton (Fienup), Steer, Steer-Dewdney-Ito, and van Cittert iteration.

In the Clark (1980) modification, CLEAN picks out only the largest residual points, and subtracts approximate point source responses in the (ξ, η) plane during minor (Högbom CLEAN) cycles. It only occasionally (during major cycles) computes the full I_n residual map by subtracting the identified point source responses in the (u, v) plane using a FAST FOURIER TRANSFORM for the CONVOLUTION. The ALGORITHM then returns to a minor cycle. This algorithm modifies the Högbom method to take advantage of the array processor (although it also works without one). It is therefore a factor of 2–10 faster than the simple Högbom routine. It is implemented as the AIPS task **APCLN**.

The Cornwell smoothness stabilized variant was developed because, when dealing with two-dimensional extended structures, CLEAN can produce artifacts in the form of low-level high frequency stripes running through the brighter structure. These stripes derive from poor interpolations into unsampled or poorly sampled regions of the (u, v) plane. When dealing with quasi-one-dimensional sources (i.e., jets), the artifacts resemble knots (which may not be so readily recognized as spurious). **APCLN** can invoke a modification of CLEAN that is intended to bias it toward generating smoother solutions to the deconvolution problem while preserving the requirement that the transform of the CLEAN components list fits the data. The mechanism for introducing this bias is the addition to the DIRTY BEAM of a DELTA FUNCTION (or "spike") of small amplitude (**PHAT**) while searching for the CLEAN components. The beam used for the deconvolution resembles the helmet worn by German military officers in World War I, hence the name "Prussian helmet" CLEAN.

The theory underlying the Cornwell smoothness stabilized algorithm is given by Cornwell (1982, 1983), where it is described as the smoothness stabilized CLEAN. It is implemented in the AIPS tasks **APCLN** and **MX**. The spike performs a NEGATIVE feedback into the dirty image, thus suppressing features not required by the data. Spike heights of a few percent and lower than usual loop gains are usually needed. Also according to the **MX** documentation,

$$\mathtt{PHAT} \approx \frac{(\text{noise})^2}{2(\text{signal})^2} = \frac{1}{2(\text{SNR})^2}.$$

Unfortunately, the addition of a Prussian helmet generally has "limited success," so resorting to another deconvolution method such as the MAXIMUM ENTROPY METHOD is sometimes required.

The Cotton-Schwab uses the Clark method, but the major cycle subtractions of CLEAN components are performed on ungridded visibility data. The Cotton-Schwab technique is often faster than the Clark variant. It is also capable of including the w baseline term, thus removing distortions from noncoplanar baselines. It is often faster than the Clark method. The Cotton-Schwab technique is implemented as the AIPS task **MX**.

The Gerchberg-Saxton variant, also called the Fienup variant, is a technique originally introduced for solving the phase problem in electron microscopy. It was subsequently adapted for visibility amplitude measurements only. A Gerchberg-Saxton map is constrained to be NONZERO, and positive. Data and image plane constraints are imposed alternately while transforming to and from the image plane. If the boxes to CLEAN are chosen to surround the source snugly, then the algorithm will converge faster and will have more chance of finding a unique image. The algorithm is slow, but should be comparable to the Clark technique (**APCLN**) if the map contains many picture elements. However, the resolution is data dependent and varies across the map. It is implemented as the AIPS task **APGS** (Pearson 1984).

The Steer variant is a modification of the Clark variant (Cornwell 1982). It is slow, but should be comparable to the Clark algorithm if the map contains many picture elements. The algorithm used in the program is due to David Steer. The principle is similar to Barry Clark's CLEAN except that in the minor cycle only points above the (trim level) × (peak in the residual map) are selected. In the major cycle these are removed using a FAST FOURIER TRANSFORM. If boxes are chosen to surround the source snugly, then the algorithm will converge faster and will have more chance of finding a unique image. It is implemented in AIPS as the experimental program **STEER** and as the Steer-Dewdney-Ito variant combined with the Clark algorithm as **SDCLN**.

The Steer-Dewdney-Ito variant is similar to the Clark variant, but the components are taken as all pixels having residual flux greater than a cutoff value times the current peak residual. This method should avoid the "ripples" produced by the standard CLEAN on extended emission. The AIPS task **SDCLN** does an AP-based CLEAN of the the Clark type, but differs from **APCLN** in that it offers the option to switch to the Steer-Dewdney-Ito method.

Finally, van Cittert iteration consists of two steps:

1. Estimate a correction to add to the current map estimate by multiplying the residuals by some weight. In the classical van Cittert algorithm, this weight is a constant, where as in CLEAN the weight is zero everywhere except at the peak of the residuals.
2. Add the step to the current estimate, and subtract the estimate, convolved with the DIRTY BEAM, from the residuals.

Though it is a simple algorithm, it works well (if slowly) for cases where the Dirty Beam is positive semidefinite (as it is in astronomical observations). The basic idea is that the Dirty Map is a reasonably good estimate of the deconvolved map. The different iterations vary only in the weight to apply to each residual in determining the correction step. van Cittert iteration is implemented as the AIPS task `APVC`, which is a rather experimental and ad hoc procedure. In some limiting cases, it reduces to the standard CLEAN algorithm (though it would be impractically slow).

see also CLEAN Beam, CLEAN Map, Dirty Beam, Dirty Map

References

Christiansen, W. N. and Högbom, J. A. *Radiotelescopes, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 214–216, 1985.

Clark, B. G. "An Efficient Implementation of the Algorithm 'CLEAN'." *Astron. Astrophys.* **89**, 377–378, 1980.

Cornwell, T. J. "Can CLEAN be Improved?" VLA Scientific Memorandum No. 141, 1982.

Cornwell, T. J. "Image Restoration (and the CLEAN Technique)." Lecture 9. NRAO VLA Workshop on Synthesis Mapping, p. 113, 1982.

Cornwell, T. J. "A Method of Stabilizing the CLEAN Algorithm." *Astron. Astrophys.* **121**, 281–285, 1983.

Cornwell, T. and Braun, R. "Deconvolution." Ch. 8 in *Synthesis Imaging in Radio Astronomy: Third NRAO Summer School, 1988* (Ed. R. A. Perley, F. R. Schwab, and A. H. Bridle). San Francisco, CA: Astronomical Society of the Pacific, pp. 178–179, 1989.

Högbom, J. A. "Aperture Synthesis with a Non-Regular Distribution of Interferometric Baselines." *Astron. Astrophys. Supp.* **15**, 417–426, 1974.

National Radio Astronomical Observatory. Astronomical Image Processing Software (AIPS) software package. `APCLN`, `MX`, and `UVMAP` tasks.

Pearson, T. J. and Readhead, A. C. S. "Image Formation by Self-Calibration in Radio Astronomy." *Ann. Rev. Astron. Astrophys.* **22**, 97–130, 1984.

Schwarz, U. J. "Mathematical-Statistical Description of the Iterative Beam Removing Technique (Method CLEAN)." *Astron. Astrophys.* **65**, 345–356, 1978.

Tan, S. M. "An Analysis of the Properties of CLEAN and Smoothness Stabilized CLEAN—Some Warnings." *Mon. Not. Royal Astron. Soc.* **220**, 971–1001, 1986.

Thompson, A. R.; Moran, J. M.; and Swenson, G. W. Jr. *Interferometry and Synthesis in Radio Astronomy.* New York: Wiley, p. 348, 1986.

CLEAN Beam

An Elliptical Gaussian fit to the Dirty Beam in order to remove sidelobes. The CLEAN beam is convolved with the final CLEAN iteration to diminish spurious high spatial frequencies.

see also CLEAN Algorithm, CLEAN Map, Deconvolution, Dirty Beam, Dirty Map

CLEAN Map

The deconvolved map extracted from a finitely sampled Dirty Map by the CLEAN Algorithm, Maximum Entropy Method, or any other Deconvolution procedure.

see also CLEAN Algorithm, CLEAN Beam, Deconvolution, Dirty Beam, Dirty Map

Clebsch-Aronhold Notation

A notation used to describe curves. The fundamental principle of Clebsch-Aronhold notation states that if each of a number of forms be replaced by a Power of a linear form in the same number of variables equal to the order of the given form, and if a sufficient number of equivalent symbols are introduced by the Aronhold Process so that no actual Coefficient appears except to the first degree, then every identical relation holding for the new specialized forms holds for the general ones.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 79, 1959.

Clebsch Diagonal Cubic

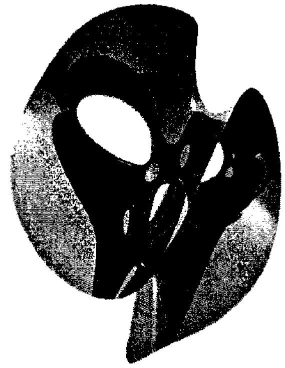

A Cubic Algebraic Surface given by the equation

$$x_0{}^3 + x_1{}^3 + x_2{}^3 + x_3{}^3 + x_4{}^3 = 0, \qquad (1)$$

with the added constraint

$$x_0 + x_1 + x_2 + x_3 + x_4 = 0. \qquad (2)$$

The implicit equation obtained by taking the plane at infinity as $x_0 + x_1 + x_2 + x_3/2$ is

$$\begin{aligned} 81(x^3+y^3+z^3) - 189(x^2y+x^2z+y^2x+y^2z+z^2x+z^2y) \\ +54xyz + 126(xy+xz+yz) - 9(x^2+y^2+z^2) \\ -9(x+y+z)+1 = 0 \quad (3) \end{aligned}$$

(Hunt, Nordstrand). On Clebsch's diagonal surface, all 27 of the complex lines (Solomon's Seal Lines) present on a general smooth Cubic Surface are real. In addition, there are 10 points on the surface where 3 of the 27 lines meet. These points are called Eckardt Points (Fischer 1986, Hunt), and the Clebsch diagonal surface is the unique Cubic Surface containing 10 such points (Hunt).

If one of the variables describing Clebsch's diagonal surface is dropped, leaving the equations

$$x_0{}^3 + x_1{}^3 + x_2{}^3 + x_3{}^3 = 0, \qquad (4)$$

$$x_0 + x_1 + x_2 + x_3 = 0, \tag{5}$$

the equations degenerate into two intersecting PLANES given by the equation

$$(x + y)(x + z)(y + z) = 0. \tag{6}$$

see also CUBIC SURFACE, ECKARDT POINT

References

Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, pp. 9–11, 1986.

Fischer, G. (Ed.). Plates 10–12 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 13–15, 1986.

Hunt, B. *The Geometry of Some Special Arithmetic Quotients.* New York: Springer-Verlag, pp. 122–128, 1996.

Nordstrand, T. "Clebsch Diagonal Surface." `http://www.uib.no/people/nfytn/clebtxt.htm`.

Clebsch-Gordon Coefficient

A mathematical symbol used to integrate products of three SPHERICAL HARMONICS. Clebsch-Gordon coefficients commonly arise in applications involving the addition of angular momentum in quantum mechanics. If products of more than three SPHERICAL HARMONICS are desired, then a generalization known as WIGNER $6j$-SYMBOLS or WIGNER $9j$-SYMBOLS is used. The Clebsch-Gordon coefficients are written

$$C^j_{m_1 m_2} = (j_1 j_2 m_1 m_2 | j_1 j_2 j m) \tag{1}$$

and are defined by

$$\Psi_{JM} = \sum_{M = M_1 + M_2} C^J_{M_1 M_2} \Psi_{M_1 M_2}, \tag{2}$$

where $J \equiv J_1 + J_2$. The Clebsch-Gordon coefficients are sometimes expressed using the related RACAH V-COEFFICIENTS

$$V(j_1 j_2 j; m_1 m_2 m) \tag{3}$$

or WIGNER $3j$-SYMBOLS. Connections among the three are

$$(j_1 j_2 m_1 m_2 | j_1 j_2 m) \\ = (-1)^{-j_1 + j_2 - m} \sqrt{2j + 1} \begin{pmatrix} j_1 & j_2 & j \\ m_1 & m_2 & -m \end{pmatrix} \tag{4}$$

$$(j_1 j_2 m_1 m_2 | j_1 j_2 j m) \\ = (-1)^{j+m} \sqrt{2j + 1} V(j_1 j_2 j; m_1 m_2 \; - m) \tag{5}$$

$$V(j_1 j_2 j; m_1 m_2 m) = (-1)^{-j_1 + j_2 + j} \begin{pmatrix} j_1 & j_2 & j_1 \\ m_2 & m_1 & m_2 \end{pmatrix}. \tag{6}$$

They have the symmetry

$$(j_1 j_2 m_1 m_2 | j_1 j_2 j m) = (-1)^{j_1 + j_2 - j} (j_2 j_1 m_2 m_1 | j_2 j_1 j m), \tag{7}$$

and obey the orthogonality relationships

$$\sum_{j,m} (j_1 j_2 m_1 m_2 | j_1 j_2 j m)(j_1 j_2 j m | j_1 j_2 m_1' m_2') \\ = \delta_{m_1 m_1'} \delta_{m_2 m_2'} \tag{8}$$

$$\sum_{m_1, m_2} (j_1 j_2 m_1 m_2 | j_1 j_2 j m)(j_1 j_2 j' m' | j_1 j_2 m_1 m_2) \\ = \delta_{jj'} \delta_{mm'}. \tag{9}$$

see also RACAH V-COEFFICIENT, RACAH W-COEFFICIENT, WIGNER $3j$-SYMBOL, WIGNER $6j$-SYMBOL, WIGNER $9j$-SYMBOL

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Vector-Addition Coefficients." §27.9 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 1006–1010, 1972.

Cohen-Tannoudji, C.; Diu, B.; and Laloë, F. "Clebsch-Gordon Coefficients." Complement B_X in *Quantum Mechanics, Vol. 2.* New York: Wiley, pp. 1035–1047, 1977.

Condon, E. U. and Shortley, G. §3.6–3.14 in *The Theory of Atomic Spectra.* Cambridge, England: Cambridge University Press, pp. 56–78, 1951.

Fano, U. and Fano, L. *Basic Physics of Atoms and Molecules.* New York: Wiley, p. 240, 1959.

Messiah, A. "Clebsch-Gordon (C.-G.) Coefficients and '$3j$' Symbols." Appendix C.I in *Quantum Mechanics, Vol. 2.* Amsterdam, Netherlands: North-Holland, pp. 1054–1060, 1962.

Shore, B. W. and Menzel, D. H. "Coupling and Clebsch-Gordon Coefficients." §6.2 in *Principles of Atomic Spectra.* New York: Wiley, pp. 268–276, 1968.

Sobel'man, I. I. "Angular Momenta." Ch. 4 in *Atomic Spectra and Radiative Transitions, 2nd ed.* Berlin: Springer-Verlag, 1992.

Clement Matrix

see KAC MATRIX

Clenshaw Recurrence Formula

The downward Clenshaw recurrence formula evaluates a sum of products of indexed COEFFICIENTS by functions which obey a recurrence relation. If

$$f(x) = \sum_{k=0}^{N} c_k F_k(x)$$

and

$$F_{n+1}(x) = \alpha(n, x) F_n(x) + \beta(n, x) F_{n-1}(x),$$

where the c_ks are known, then define

$$\begin{aligned} y_{N+2} &= y_{N+1} = 0 \\ y_k &= \alpha(k,x)y_{k+1} + \beta(k+1,x)y_{k+2} + c_k \end{aligned}$$

for $k = N, N-1, \ldots$ and solve backwards to obtain y_2 and y_1.

$$c_k = y_k - \alpha(k,x)y_{k+1} - \beta(k+1,x)y_{k+2}$$

$$\begin{aligned} f(x) &= \sum_{k=0}^{N} c_k F_k(x) \\ &= c_0F_0(x) + [y_1 - \alpha(1,x)y_2 - \beta(2,x)y_3]F_1(x) \\ &\quad +[y_2 - \alpha(2,x)y_3 - \beta(3,x)y_4]F_2(x) \\ &\quad +[y_3 - \alpha(3,x)y_4 - \beta(4,x)y_5]F_3(x) \\ &\quad +[y_4 - \alpha(4,x)y_5 - \beta(5,x)y_6]F_4(x) + \ldots \\ &= c_0F_0(x) + y_1F_1(x) + y_2[F_2(x) - \alpha(1,x)F_1(x)] \\ &\quad + y_3[F_3(x) - \alpha(2,x)F_2(x) - \beta(2,x)] \\ &\quad + y_4[F_4(x) - \alpha(3,x)F_3(x) - \beta(3,x)] + \ldots \\ &= c_0F_0(x) + y_2[\{\alpha(1,x)F_1(x) + \beta(1,x)F_0(x)\} \\ &\quad - \alpha(1,x)F_1(x)] + y_1F_1(x) \\ &= c_0F_0(x) + y_1F_1(x) + \beta(1,x)F_0(x)y_2. \end{aligned}$$

The upward Clenshaw recurrence formula is

$$y_{-2} = y_{-1} = 0$$

$$y_k = \frac{1}{\beta(k+1,x)}[y_{k-2} - \alpha(k,x)y_{k-1} - c_k]$$

for $k = 0, 1, \ldots, N-1$.

$$f(x) = c_NF_N(x) - \beta(N,x)F_{N-1}(x)y_{N-1} - F_N(x)y_{N-2}.$$

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Recurrence Relations and Clenshaw's Recurrence Formula." §5.5 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 172–178, 1992.

Cliff Random Number Generator

A RANDOM NUMBER generator produced by iterating

$$X_{n+1} = |100 \ln X_n \pmod 1|$$

for a SEED $X_0 = 0.1$. This simple generator passes the NOISE SPHERE test for randomness by showing no structure.

see also RANDOM NUMBER, SEED

References

Pickover, C. A. "Computers, Randomness, Mind, and Infinity." Ch. 31 in *Keys to Infinity.* New York: W. H. Freeman, pp. 233–247, 1995.

Clifford Algebra

Let V be an n-D linear SPACE over a FIELD K, and let Q be a QUADRATIC FORM on V. A Clifford algebra is then defined over the $T(V)/I(Q)$, where $T(V)$ is the tensor algebra over V and I is a particular IDEAL of $T(V)$.

References

Iyanaga, S. and Kawada, Y. (Eds.). "Clifford Algebras." §64 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 220–222, 1980.

Lounesto, P. "Counterexamples to Theorems Published and Proved in Recent Literature on Clifford Algebras, Spinors, Spin Groups, and the Exterior Algebra." `http://www.hit.fi/~lounesto/counterexamples.htm`.

Clifford's Circle Theorem

Let C_1, C_2, C_3, and C_4 be four CIRCLES of GENERAL POSITION through a point P. Let P_{ij} be the second intersection of the CIRCLES C_i and C_j. Let C_{ijk} be the CIRCLE $P_{ij}P_{ik}P_{jk}$. Then the four CIRCLES P_{234}, P_{134}, P_{124}, and P_{123} all pass through the point P_{1234}. Similarly, let C_5 be a fifth CIRCLE through P. Then the five points P_{2345}, P_{1345}, P_{1245}, P_{1235} and P_{1234} all lie on one CIRCLE C_{12345}. And so on.

see also CIRCLE, COX'S THEOREM

Clifford's Curve Theorem

The dimension of a special series can never exceed half its order.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 263, 1959.

Clique

In a GRAPH of N VERTICES, a subset of pairwise adjacent VERTICES is known as a clique. A clique is a fully connected subgraph of a given graph. The problem of finding the size of a clique for a given GRAPH is an NP-COMPLETE PROBLEM. The number of graphs on n nodes having 3 cliques are 0, 0, 1, 4, 12, 31, 67, ... (Sloane's A005289).

see also CLIQUE NUMBER, MAXIMUM CLIQUE PROBLEM, RAMSEY NUMBER, TURÁN'S THEOREM

References

Sloane, N. J. A. Sequence A005289/M3440 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Clique Number

The number of VERTICES in the largest CLIQUE of G, denoted $\omega(G)$. For an arbitrary GRAPH,

$$\omega(G) \geq \sum_{i=1}^{n} \frac{1}{n - d_i},$$

where d_i is the DEGREE of VERTEX i.

References

Aigner, M. "Turán's Graph Theorem." *Amer. Math. Monthly* **102**, 808–816, 1995.

Clock Solitaire

A solitaire game played with CARDS. The chance of winning is 1/13, and the AVERAGE number of CARDS turned up is 42.4.

References
Gardner, M. *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, pp. 244–247, 1978.

Close Packing

see SPHERE PACKING

Closed Curve

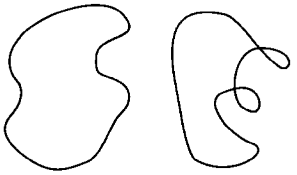

closed curves

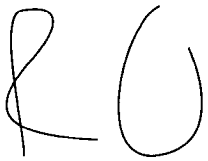

open curves

A CURVE with no endpoints which completely encloses an AREA. A closed curve is formally defined as the continuous IMAGE of a CLOSED SET.

see also SIMPLE CURVE

Closed Curve Problem

Find NECESSARY and SUFFICIENT conditions that determine when the integral curve of two periodic functions $\kappa(s)$ and $\tau(s)$ with the same period L is a CLOSED CURVE.

Closed Disk

An n-D closed disk of RADIUS r is the collection of points of distance $\leq r$ from a fixed point in EUCLIDEAN n-space.

see also DISK, OPEN DISK

Closed Form

A discrete FUNCTION $A(n,k)$ is called closed form (or sometimes "hypergeometric") in two variables if the ratios $A(n+1,k)/A(n,k)$ and $A(n,k+1)/A(n,k)$ are both RATIONAL FUNCTIONS. A pair of closed form functions (F,G) is said to be a WILF-ZEILBERGER PAIR if

$$F(n+1,k)-F(n,k)=G(n,k+1)-G(n,k).$$

see also RATIONAL FUNCTION, WILF-ZEILBERGER PAIR

References
Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, p. 141, 1996.
Zeilberger, D. "Closed Form (Pun Intended!)." *Contemporary Math.* **143**, 579–607, 1993.

Closed Graph Theorem

A linear OPERATOR between two BANACH SPACES is continuous IFF it has a "closed" GRAPH.

see also BANACH SPACE

References
Zeidler, E. *Applied Functional Analysis: Applications to Mathematical Physics.* New York: Springer-Verlag, 1995.

Closed Interval

An INTERVAL which includes its LIMIT POINTS. If the endpoints of the interval are FINITE numbers a and b, then the INTERVAL is denoted $[a,b]$. If one of the endpoints is $\pm\infty$, then the interval still contains all of its LIMIT POINTS, so $[a,\infty)$ and $(-\infty,b]$ are also closed intervals.

see also HALF-CLOSED INTERVAL, OPEN INTERVAL

Closed Set

There are several equivalent definitions of a closed SET. A SET S is closed if

1. The COMPLEMENT of S is an OPEN SET,
2. S is its own CLOSURE,
3. Sequences/nets/filters in S which converge do so within S,
4. Every point outside S has a NEIGHBORHOOD disjoint from S.

The POINT-SET TOPOLOGICAL definition of a closed set is a set which contains all of its LIMIT POINTS. Therefore, a closed set C is one for which, whatever point x is picked outside of C, x can always be isolated in some OPEN SET which doesn't touch C.

see also CLOSED INTERVAL

Closure

A SET S and a BINARY OPERATOR $*$ are said to exhibit closure if applying the BINARY OPERATOR to two elements S returns a value which is itself a member of S.

The term "closure" is also used to refer to a "closed" version of a given set. The closure of a SET can be defined in several equivalent ways, including

1. The SET plus its LIMIT POINTS, also called "boundary" points, the union of which is also called the "frontier,"
2. The unique smallest CLOSED SET containing the given SET,
3. The COMPLEMENT of the interior of the COMPLEMENT of the set,
4. The collection of all points such that every NEIGHBORHOOD of them intersects the original SET in a nonempty SET.

In topologies where the T2-SEPARATION AXIOM is assumed, the closure of a finite SET S is S itself.

see also BINARY OPERATOR, EXISTENTIAL CLOSURE, REFLEXIVE CLOSURE, TIGHT CLOSURE, TRANSITIVE CLOSURE

Clothoid

see also CORNU SPIRAL

Clove Hitch

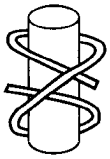

A HITCH also called the BOATMAN'S KNOT or PEG KNOT.

References
Owen, P. *Knots.* Philadelphia, PA: Courage, pp. 24–27, 1993.

Clump

see RUN

Cluster

Given a lattice, a cluster is a group of filled cells which are all connected to their neighbors vertically or horizontally.

see also CLUSTER PERIMETER, PERCOLATION THEORY, s-CLUSTER, s-RUN

References
Stauffer, D. and Aharony, A. *Introduction to Percolation Theory, 2nd ed.* London: Taylor & Francis, 1992.

Cluster Perimeter

The number of empty neighbors of a CLUSTER.

see also PERIMETER POLYNOMIAL

Coanalytic Set

A DEFINABLE SET which is the complement of an ANALYTIC SET.

see also ANALYTIC SET

Coastline Paradox

Determining the length of a country's coastline is not as simple as it first appears, as first considered by L. F. Richardson (1881–1953). In fact, the answer depends on the length of the RULER you use for the measurements. A shorter RULER measures more of the sinuosity of bays and inlets than a larger one, so the estimated length continues to increase as the RULER length decreases.

In fact, a coastline is an example of a FRACTAL, and plotting the length of the RULER versus the measured length of the coastline on a log-log plot gives a straight line, the slope of which is the FRACTAL DIMENSION of the coastline (and will be a number between 1 and 2).

References
Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 29–31, 1991.

Coates-Wiles Theorem

In 1976, Coates and Wiles showed that ELLIPTIC CURVES with COMPLEX MULTIPLICATION having an infinite number of solutions have L-functions which are zero at the relevant fixed point. This is a special case of the SWINNERTON-DYER CONJECTURE.

References
Cipra, B. "Fermat Prover Points to Next Challenges." *Science* **271**, 1668–1669, 1996.

Coaxal Circles

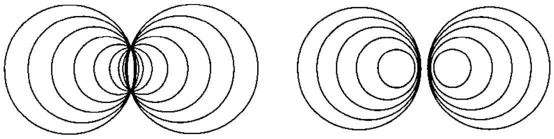

CIRCLES which share a RADICAL LINE with a given circle are said to be coaxal. The centers of coaxal circles are COLLINEAR. It is possible to combine the two types of coaxal systems illustrated above such that the sets are orthogonal.

see also CIRCLE, COAXALOID SYSTEM, GAUSS-BODENMILLER THEOREM, RADICAL LINE

References
Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 35–36 and 122, 1967.
Dixon, R. *Mathographics.* New York: Dover, pp. 68–72, 1991.
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 34–37, 199, and 279, 1929.

Coaxal System

A system of COAXAL CIRCLES.

Coaxaloid System

A system of circles obtained by multiplying each RADIUS in a COAXAL SYSTEM by a constant.

References
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 276–277, 1929.

Cobordant Manifold

Two open MANIFOLDS M and M' are cobordant if there exists a MANIFOLD with boundary W^{n+1} such that an acceptable restrictive relationship holds.

see also COBORDISM, h-COBORDISM THEOREM, MORSE THEORY

Cobordism

see BORDISM, h-COBORDISM

Cobordism Group

see BORDISM GROUP

Cobordism Ring

see BORDISM GROUP

Cochleoid

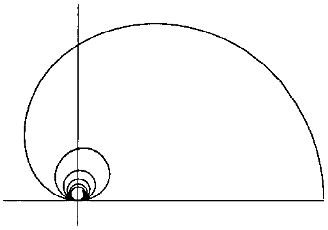

The cochleoid, whose name means "snail-form" in Latin, was first discussed by J. Peck in 1700 (MacTutor Archive). The points of contact of PARALLEL TANGENTS to the cochleoid lie on a STROPHOID.

In POLAR COORDINATES,

$$r = \frac{a \sin\theta}{\theta}. \tag{1}$$

In CARTESIAN COORDINATES,

$$(x^2 + y^2) \tan^{-1}\left(\frac{y}{x}\right) = ay. \tag{2}$$

The CURVATURE is

$$\kappa = \frac{2\sqrt{2}\,\theta^3[2\theta - \sin(2\theta)]}{[1 + 2\theta^2 - \cos(2\theta) - 2\theta\sin(2\theta)]^{3/2}}. \tag{3}$$

see also QUADRATRIX OF HIPPIAS

References
Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 192 and 196, 1972.
MacTutor History of Mathematics Archive. "Cochleoid." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Cochleoid.html`.

Cochleoid Inverse Curve

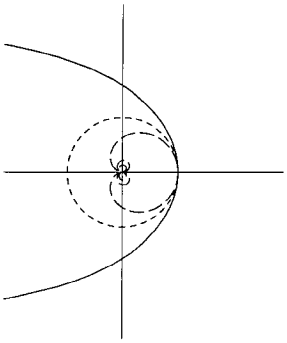

The INVERSE CURVE of the COCHLEOID

$$r = \frac{\sin\theta}{\theta} \tag{1}$$

with INVERSION CENTER at the ORIGIN and inversion radius k is the QUADRATRIX OF HIPPIAS.

$$x = kt\cot\theta \tag{2}$$

$$y = kt. \tag{3}$$

Cochloid

see CONCHOID OF NICOMEDES

Cochran's Theorem

The converse of FISHER'S THEOREM.

Cocked Hat Curve

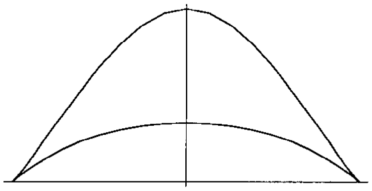

The PLANE CURVE

$$(x^2 + 2ay - a^2)^2 = y^2(a^2 - x^2),$$

which is similar to the BICORN.

References
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 72, 1989.

Cocktail Party Graph

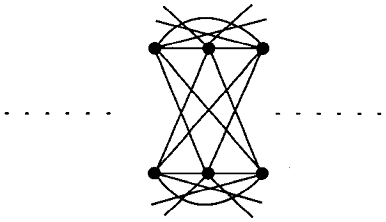

A GRAPH consisting of two rows of paired nodes in which all nodes but the paired ones are connected with an EDGE. It is the complement of the LADDER GRAPH.

Coconut

see MONKEY AND COCONUT PROBLEM

Codazzi Equations

see MAINARDI-CODAZZI EQUATIONS

Code

A code is a set of n-tuples of elements ("WORDS") taken from an ALPHABET.

see also ALPHABET, CODING THEORY, ENCODING, ERROR-CORRECTING CODE, GRAY CODE, HUFFMAN CODING, ISBN, LINEAR CODE, WORD

Codimension

The minimum number of parameters needed to fully describe all possible behaviors near a nonstructurally stable element.

see also BIFURCATION

Coding Theory

Coding theory, sometimes called ALGEBRAIC CODING THEORY, deals with the design of ERROR-CORRECTING CODES for the reliable transmission of information across noisy channels. It makes use of classical and modern algebraic techniques involving FINITE FIELDS, GROUP THEORY, and polynomial algebra. It has connections with other areas of DISCRETE MATHEMATICS, especially NUMBER THEORY and the theory of experimental designs.

see also ENCODING, ERROR-CORRECTING CODE, GALOIS FIELD, HADAMARD MATRIX

References

Alexander, B. "At the Dawn of the Theory of Codes." *Math. Intel.* **15**, 20–26, 1993.

Golomb, S. W.; Peile, R. E.; and Scholtz, R. A. *Basic Concepts in Information Theory and Coding: The Adventures of Secret Agent 00111.* New York: Plenum, 1994.

Humphreys, O. F. and Prest, M. Y. *Numbers, Groups, and Codes.* New York: Cambridge University Press, 1990.

MacWilliams, F. J. and Sloane, N. J. A. *The Theory of Error-Correcting Codes.* New York: Elsevier, 1978.

Roman, S. *Coding and Information Theory.* New York: Springer-Verlag, 1992.

Coefficient

A multiplicative factor (usually indexed) such as one of the constants a_i in the POLYNOMIAL $a_n x^n + a_{n-1} x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0$.

see also BINOMIAL COEFFICIENT, CARTAN TORSION COEFFICIENT, CENTRAL BINOMIAL COEFFICIENT, CLEBSCH-GORDON COEFFICIENT, COEFFICIENT FIELD, COMMUTATION COEFFICIENT, CONNECTION COEFFICIENT, CORRELATION COEFFICIENT, CROSS-CORRELATION COEFFICIENT, EXCESS COEFFICIENT, GAUSSIAN COEFFICIENT, LAGRANGIAN COEFFICIENT, MULTINOMIAL COEFFICIENT, PEARSON'S SKEWNESS COEFFICIENTS, PRODUCT-MOMENT COEFFICIENT OF CORRELATION, QUARTILE SKEWNESS COEFFICIENT, QUARTILE VARIATION COEFFICIENT, RACAH V-COEFFICIENT, RACAH W-COEFFICIENT, REGRESSION COEFFICIENT, ROMAN COEFFICIENT, TRIANGLE COEFFICIENT, UNDETERMINED COEFFICIENTS METHOD, VARIATION COEFFICIENT

Coefficient Field

Let V be a VECTOR SPACE over a FIELD K, and let A be a nonempty SET. For an appropriately defined AFFINE SPACE A, K is called the COEFFICIENT field.

Coercive Functional

A bilinear FUNCTIONAL ϕ on a normed SPACE E is called coercive (or sometimes ELLIPTIC) if there exists a POSITIVE constant K such that

$$\phi(x,x) \geq K\|x\|^2$$

for all $x \in E$.

see also LAX-MILGRAM THEOREM

References

Debnath, L. and Mikusiński, P. *Introduction to Hilbert Spaces with Applications.* San Diego, CA: Academic Press, 1990.

Cofactor

The MINOR of a DETERMINANT is another DETERMINANT $|\mathsf{C}|$ formed by omitting the ith row and jth column of the original DETERMINANT $|\mathsf{M}|$.

$$C_{ij} \equiv (-1)^{i+j} a_i M_{ij}.$$

see also DETERMINANT EXPANSION BY MINORS, MINOR

Cohen-Kung Theorem

Guarantees that the trajectory of LANGTON'S ANT is unbounded.

Cohomology

Cohomology is an invariant of a TOPOLOGICAL SPACE, formally "dual" to HOMOLOGY, and so it detects "holes" in a SPACE. Cohomology has more algebraic structure than HOMOLOGY, making it into a graded ring (multiplication given by "cup product"), whereas HOMOLOGY is just a graded ABELIAN GROUP invariant of a SPACE.

A generalized homology or cohomology theory must satisfy all of the EILENBERG-STEENROD AXIOMS with the exception of the dimension axiom.

see also ALEKSANDROV-ČECH COHOMOLOGY, ALEXANDER-SPANIER COHOMOLOGY, ČECH COHOMOLOGY, DE RHAM COHOMOLOGY, HOMOLOGY (TOPOLOGY)

Cohomotopy Group

Cohomotopy groups are similar to HOMOTOPY GROUPS. A cohomotopy group is a GROUP related to the HOMOTOPY classes of MAPS from a SPACE X into a SPHERE $\mathbb{S}^n$.

see also HOMOTOPY GROUP

Coin

A flat disk which acts as a two-sided DIE.

see BERNOULLI TRIAL, CARDS, COIN PARADOX, COIN TOSSING, DICE, FELLER'S COIN-TOSSING CONSTANTS, FOUR COINS PROBLEM, GAMBLER'S RUIN

References

Brooke, M. *Fun for the Money.* New York: Scribner's, 1963.

Coin Flipping

see COIN TOSSING

Coin Paradox

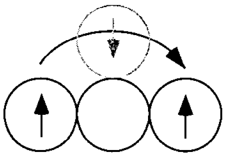

After a half rotation of the coin on the left around the central coin (of the same RADIUS), the coin undergoes a complete rotation.

References

Pappas, T. "The Coin Paradox." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 220, 1989.

Coin Problem

Let there be $n \geq 2$ INTEGERS $0 < a_1 < \ldots < a_n$ with $(a_1, a_2, \ldots, a_n) = 1$ (all RELATIVELY PRIME). For large enough $N = \sum_{i=1}^n a_i x_i$, there is a solution in NONNEGATIVE INTEGERS x_i. The greatest $N = g(a_1, a_2, \ldots a_n)$ for which there is no solution is called the coin problem. Sylvester showed

$$g(a_1, a_2) = (a_1 - 1)(a_2 - 1) - 1,$$

and an explicit solution is known for $n = 3$, but no closed form solution is known for larger N.

References

Guy, R. K. "The Money-Changing Problem." §C7 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 113–114, 1994.

Coin Tossing

An idealized coin consists of a circular disk of zero thickness which, when thrown in the air and allowed to fall, will rest with either side face up ("heads" H or "tails" T) with equal probability. A coin is therefore a two-sided DIE. A coin toss corresponds to a BERNOULLI DISTRIBUTION with $p = 1/2$. Despite slight differences between the sides and NONZERO thickness of actual coins, the distribution of their tosses makes a good approximation to a $p = 1/2$ BERNOULLI DISTRIBUTION.

There are, however, some rather counterintuitive properties of coin tossing. For example, it is twice as likely that the triple *TTH* will be encountered before *THT* than after it, and three times as likely that *THH* will precede *HTT*. Furthermore, it is six times as likely that *HTT* will be the first of *HTT*, *TTH*, and *TTT* to occur (Honsberger 1979). More amazingly still, *spinning* a penny instead of tossing it results in heads only about 30% of the time (Paulos 1995).

Let $w(n)$ be the probability that no RUN of three consecutive heads appears in n independent tosses of a COIN. The following table gives the first few values of $w(n)$.

n	$w(n)$
0	1
1	1
2	1
3	$\frac{7}{8}$
4	$\frac{13}{16}$
5	$\frac{3}{4}$

Feller (1968, pp. 278–279) proved that

$$\lim_{n\to\infty} w(n)\alpha^{n+1} = \beta, \tag{1}$$

where

$$\begin{aligned}\alpha &= \tfrac{1}{3}[(136 + 24\sqrt{33})^{1/3} - 8(136 + 24\sqrt{33})^{-1/3} - 2] \\ &= 1.087378025\ldots\end{aligned} \tag{2}$$

and

$$\beta = \frac{2-\alpha}{4-3\alpha} = 1.236839845\ldots. \tag{3}$$

The corresponding constants for a RUN of $k > 1$ heads are α_k, the smallest POSITIVE ROOT of

$$1 - x + (\tfrac{1}{2}x)^{k+1} = 0, \tag{4}$$

and

$$\beta_k = \frac{2-\alpha}{k+1-k\alpha_k}. \tag{5}$$

These are modified for unfair coins with $P(H) = p$ and $P(T) = q = 1 - p$ to α'_k, the smallest POSITIVE ROOT of

$$1 - x + qp^k x^{k+1} = 0, \tag{6}$$

and

$$\beta'_k = \frac{1 - p\alpha'_k}{(k+1-k\alpha'_k)p} \tag{7}$$

(Feller 1968, pp. 322–325).

see also BERNOULLI DISTRIBUTION, CARDS, COIN, DICE, GAMBLER'S RUIN, MARTINGALE, RUN, SAINT PETERSBURG PARADOX

References

Feller, W. *An Introduction to Probability Theory and Its Application, Vol. 1, 3rd ed.* New York: Wiley, 1968.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/feller/feller.html`.

Ford, J. "How Random is a Coin Toss?" *Physics Today* **36**, 40–47, 1983.

Honsberger, R. "Some Surprises in Probability." Ch. 5 in *Mathematical Plums* (Ed. R. Honsberger). Washington, DC: Math. Assoc. Amer., pp. 100–103, 1979.

Keller, J. B. "The Probability of Heads." *Amer. Math. Monthly* **93**, 191–197, 1986.

Paulos, J. A. *A Mathematician Reads the Newspaper.* New York: BasicBooks, p. 75, 1995.

Peterson, I. *Islands of Truth: A Mathematical Mystery Cruise.* New York: W. H. Freeman, pp. 238–239, 1990.

Spencer, J. "Combinatorics by Coin Flipping." *Coll. Math. J.*, **17**, 407–412, 1986.

Coincidence

A coincidence is a surprising concurrence of events, perceived as meaningfully related, with no apparent causal connection (Diaconis and Mosteller 1989).

see also BIRTHDAY PROBLEM, LAW OF TRULY LARGE NUMBERS, ODDS, PROBABILITY, RANDOM NUMBER

References

Bogomolny, A. "Coincidence." `http://www.cut-the-knot.com/do_you_know/coincidence.html`.

Falk, R. "On Coincidences." *Skeptical Inquirer* **6**, 18–31, 1981–82.

Falk, R. "The Judgment of Coincidences: Mine Versus Yours." *Amer. J. Psych.* **102**, 477–493, 1989.

Falk, R. and MacGregor, D. "The Surprisingness of Coincidences." In *Analysing and Aiding Decision Processes* (Ed. P. Humphreys, O. Svenson, and A. Vári). New York: Elsevier, pp. 489–502, 1984.

Diaconis, P. and Mosteller, F. "Methods of Studying Coincidences." *J. Amer. Statist. Assoc.* **84**, 853–861, 1989.

Jung, C. G. *Synchronicity: An Acausal Connecting Principle.* Princeton, NJ: Princeton University Press, 1973.

Kammerer, P. *Das Gesetz der Serie: Eine Lehre von den Wiederholungen im Lebens—und im Weltgeschehen.* Stuttgart, Germany: Deutsche Verlags-Anstahlt, 1919.

Stewart, I. "What a Coincidence!" *Sci. Amer.* **278**, 95–96, June 1998.

Colatitude

The polar angle on a SPHERE measured from the North Pole instead of the equator. The angle ϕ in SPHERICAL COORDINATES is the COLATITUDE. It is related to the LATITUDE δ by $\phi = 90° - \delta$.

see also LATITUDE, LONGITUDE, SPHERICAL COORDINATES

Colinear

see COLLINEAR

Collatz Problem

A problem posed by L. Collatz in 1937, also called the $3x + 1$ MAPPING, HASSE'S ALGORITHM, KAKUTANI'S PROBLEM, SYRACUSE ALGORITHM, SYRACUSE PROBLEM, THWAITES CONJECTURE, and ULAM'S PROBLEM (Lagarias 1985). Thwaites (1996) has offered a £1000 reward for resolving the CONJECTURE. Let n be an INTEGER. Then the Collatz problem asks if iterating

$$f(n) = \begin{cases} \frac{1}{2}n & \text{for } n \text{ even} \\ 3n+1 & \text{for } n \text{ odd} \end{cases} \tag{1}$$

always returns to 1 for POSITIVE n. This question has been tested and found to be true for all numbers $< 5.6 \times 10^{13}$ (Leavens and Vermeulen 1992), and more recently, 10^{15} (Vardi 1991, p. 129). The members of the SEQUENCE produced by the Collatz are sometimes known as HAILSTONE NUMBERS. Because of the difficulty in solving this problem, Erdős commented that "mathematics is not yet ready for such problems" (Lagarias 1985). If NEGATIVE numbers are included, there are four known cycles (excluding the trivial 0 cycle): (4, 2, 1), (−2, −1), (−5, −7, −10), and (−17, −25, −37, −55, −82, −41, −61, −91, −136, −68, −34). The number of tripling steps needed to reach 1 for $n = 1, 2, \ldots$ are 0, 0, 2, 0, 1, 2, 5, 0, 6, ... (Sloane's A006667).

The Collatz problem was modified by Terras (1976, 1979), who asked if iterating

$$T(x) = \begin{cases} \frac{1}{2}x & \text{for } x \text{ even} \\ \frac{1}{2}(3x+1) & \text{for } x \text{ odd} \end{cases} \tag{2}$$

always returns to 1. If NEGATIVE numbers are included, there are 4 known cycles: (1, 2), (−1), (−5, −7, −10), and (−17, −25, −37, −55, −82, −41, −61, −91, −136, −68, −34). It is a special case of the "generalized Collatz problem" with $d = 2$, $m_0 = 1$, $m_1 = 3$, $r_0 = 0$, and $r_1 = -1$. Terras (1976, 1979) also proved that the set of INTEGERS $S_k \equiv \{n : n \text{ has stopping time} \leq k\}$ has a limiting asymptotic density $F(k)$, so the limit

$$F(k) = \lim_{x \to \infty} \frac{1}{x}, \tag{3}$$

for $\{n : n \leq x \text{ and } \sigma(n) \leq k\}$ exists. Furthermore, $F(k) \to 1$ as $k \to \infty$, so almost all INTEGERS have a finite stopping time. Finally, for all $k \geq 1$,

$$1 - F(k) = \lim_{x \to \infty} \frac{1}{x} \leq 2^{-\eta k}, \tag{4}$$

where

$$\eta = 1 - H(\theta) = 0.05004\ldots \tag{5}$$

$$H(x) = -x \lg x - (1-x) \lg(1-x) \tag{6}$$

$$\theta = \frac{1}{\lg 3} \tag{7}$$

(Lagarias 1985).

Conway proved that the original Collatz problem has no nontrivial cycles of length < 400. Lagarias (1985) showed that there are no nontrivial cycles with length < 275,000. Conway (1972) also proved that Collatz-type problems can be formally UNDECIDABLE.

A generalization of the COLLATZ PROBLEM lets $d \geq 2$ be a POSITIVE INTEGER and $m_0, \ldots, m_{d-1}$ be NONZERO INTEGERS. Also let $r_i \in \mathbb{Z}$ satisfy

$$r_i \equiv i m_i \pmod{d}. \tag{8}$$

Then

$$T(x) = \frac{m_i x - r_i}{d} \tag{9}$$

for $x \equiv i \pmod{d}$ defines a generalized Collatz mapping. An equivalent form is

$$T(x) = \left\lfloor \frac{m_i x}{d} \right\rfloor + X_i \tag{10}$$

for $x \equiv i \pmod{d}$ where $X_0, \ldots, X_{d-1}$ are INTEGERS and $\lfloor r \rfloor$ is the FLOOR FUNCTION. The problem is connected with ERGODIC THEORY and MARKOV CHAINS (Matthews 1995). Matthews (1995) obtained the following table for the mapping

$$T_k(x) = \begin{cases} \frac{1}{2}x & \text{for } x \equiv 0 \pmod{2} \\ \frac{1}{2}(3x+k) & \text{for } x \equiv 1 \pmod{2}, \end{cases} \qquad (11)$$

where $k = T_{5^k}$.

k	# Cycles	Max. Cycle Length
0	5	27
1	10	34
2	13	118
3	17	118
4	19	118
5	21	165
6	23	433

Matthews and Watts (1984) proposed the following conjectures.

1. If $|m_0 \cdots m_{d-1}| < d^d$, then all trajectories $\{T^K(n)\}$ for $n \in \mathbb{Z}$ eventually cycle.
2. If $|m_0 \cdots m_{d-1}| > d^d$, then almost all trajectories $\{T^K(n)\}$ for $n \in \mathbb{Z}$ are divergent, except for an exceptional set of INTEGERS n satisfying
$$\#\{n \in S \mid -X \le n < X\} = o(X).$$
3. The number of cycles is finite.
4. If the trajectory $\{T^K(n)\}$ for $n \in \mathbb{Z}$ is not eventually cyclic, then the iterates are uniformly distribution mod d^α for each $\alpha \ge 1$, with
$$\lim_{N \to \infty} \frac{1}{N+1} \operatorname{card}\{K \le N \mid T^K(n) \equiv j \pmod{d^\alpha}\} = d^{-\alpha} \qquad (12)$$
for $0 \le j \le d^\alpha - 1$.

Matthews believes that the map

$$T(x) = \begin{cases} 7x+3 & \text{for } x \equiv 0 \pmod{3} \\ \frac{1}{3}(7x+2) & \text{for } x \equiv 1 \pmod{3} \\ \frac{1}{3}(x-2) & \text{for } x \equiv 2 \pmod{3} \end{cases} \qquad (13)$$

will either reach 0 (mod 3) or will enter one of the cycles (-1) or $(-2, -4)$, and offers a \$100 (Australian?) prize for a proof.

see also HAILSTONE NUMBER

References

Applegate, D. and Lagarias, J. C. "Density Bounds for the $3x+1$ Problem 1. Tree-Search Method." *Math. Comput.* **64**, 411–426, 1995.

Applegate, D. and Lagarias, J. C. "Density Bounds for the $3x+1$ Problem 2. Krasikov Inequalities." *Math. Comput.* **64**, 427–438, 1995.

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Burckel, S. "Functional Equations Associated with Congruential Functions." *Theor. Comp. Sci.* **123**, 397–406, 1994.

Conway, J. H. "Unpredictable Iterations." *Proc. 1972 Number Th. Conf.*, University of Colorado, Boulder, Colorado, pp. 49–52, 1972.

Crandall, R. "On the '$3x+1$' Problem." *Math. Comput.* **32**, 1281–1292, 1978.

Everett, C. "Iteration of the Number Theoretic Function $f(2n) = n$, $f(2n+1) = f(3n+2)$." *Adv. Math.* **25**, 42–45, 1977.

Guy, R. K. "Collatz's Sequence." §E16 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 215–218, 1994.

Lagarias, J. C. "The $3x+1$ Problem and Its Generalizations." *Amer. Math. Monthly* **92**, 3–23, 1985. `http://www.cecm.sfu.ca/organics/papers/lagarias/`.

Leavens, G. T. and Vermeulen, M. "$3x+1$ Search Programs." *Comput. Math. Appl.* **24**, 79–99, 1992.

Matthews, K. R. "The Generalized $3x+1$ Mapping." `http://www.maths.uq.oz.au/~krm/survey.dvi`. Rev. Sept. 10, 1995.

Matthews, K. R. "A Generalized $3x+1$ Conjecture." [\$100 Reward for a Proof.] `ftp://www.maths.uq.edu.au/pub/krm/gnubc/challenge`.

Matthews, K. R. and Watts, A. M. "A Generalization of Hasses's Generalization of the Syracuse Algorithm." *Acta Arith.* **43**, 167–175, 1984.

Sloane, N. J. A. Sequence A006667/M0019 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Terras, R. "A Stopping Time Problem on the Positive Integers." *Acta Arith.* **30**, 241–252, 1976.

Terras, R. "On the Existence of a Density." *Acta Arith.* **35**, 101–102, 1979.

Thwaites, B. "Two Conjectures, or How to win £1100." *Math. Gaz.* **80**, 35–36, 1996.

Vardi, I. "The $3x+1$ Problem." Ch. 7 in *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 129–137, 1991.

Collinear

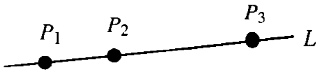

Three or more points P_1, P_2, P_3, ..., are said to be collinear if they lie on a single straight LINE L. (Two points are always collinear.) This will be true IFF the ratios of distances satisfy

$$x_2 - x_1 : y_2 - y_1 : z_2 - z_1 = x_3 - x_1 : y_3 - y_1 : z_3 - z_1.$$

Two points are trivially collinear since two points determine a LINE.

see also CONCYCLIC, DIRECTED ANGLE, N-CLUSTER, SYLVESTER'S LINE PROBLEM

Collineation

A transformation of the plane which transforms COLLINEAR points into COLLINEAR points. A projective collineation transforms every 1-D form projectively, and a perspective collineation is a collineation which leaves all lines through a point and points through a line invariant. In an ELATION, the center and axis are incident; in a HOMOLOGY they are not. For further discussion, see Coxeter (1969, p. 248).

see also AFFINITY, CORRELATION, ELATION, EQUIAFFINITY, HOMOLOGY (GEOMETRY), PERSPECTIVE COLLINEATION, PROJECTIVE COLLINEATION

References

Coxeter, H. S. M. "Collineations and Correlations." §14.6 in *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 247–251, 1969.

Cologarithm

The LOGARITHM of the RECIPROCAL of a number, equal to the NEGATIVE of the LOGARITHM of the number itself,

$$\operatorname{colog} x \equiv \log\left(\frac{1}{x}\right) = -\log x.$$

see also ANTILOGARITHM, LOGARITHM

Colon Product

Let **AB** and **CD** be DYADS. Their colon product is defined by

$$\mathbf{AB} : \mathbf{CD} \equiv \mathbf{C} \cdot \mathbf{AB} \cdot \mathbf{D} = (\mathbf{A} \cdot \mathbf{C})(\mathbf{B} \cdot \mathbf{D}).$$

Colorable

Color each segment of a KNOT DIAGRAM using one of three colors. If

1. at any crossing, either the colors are all different or all the same, and
2. at least two colors are used,

then a KNOT is said to be colorable (or more specifically, THREE-COLORABLE). Colorability is invariant under REIDEMEISTER MOVES, and can be generalized. For instance, for five colors 0, 1, 2, 3, and 4, a KNOT is five-colorable if

1. at any crossing, three segments meet. If the overpass is numbered a and the two underpasses B and C, then $2a \equiv b + c \pmod 5$, and
2. at least two colors are used.

Colorability cannot alway distinguish HANDEDNESS. For instance, three-colorability can distinguish the mirror images of the TREFOIL KNOT but not the FIGURE-OF-EIGHT KNOT. Five-colorability, on the other hand, distinguishes the MIRROR IMAGES of the FIGURE-OF-EIGHT KNOT but not the TREFOIL KNOT.

see also COLORING, THREE-COLORABLE

Coloring

A coloring of plane regions, LINK segments, etc., is an assignment of a distinct labelling (which could be a number, letter, color, etc.) to each component. Coloring problems generally involve TOPOLOGICAL considerations (i.e., they depend on the abstract study of the arrangement of objects), and theorems about colorings, such as the famous FOUR-COLOR THEOREM, can be extremely difficult to prove.

see also COLORABLE, EDGE-COLORING, FOUR-COLOR THEOREM, k-COLORING, POLYHEDRON COLORING, SIX-COLOR THEOREM, THREE-COLORABLE, VERTEX COLORING

References

Eppstein, D. "Coloring." `http://www.ics.uci.edu/~eppstein/junkyard/color.html`.

Saaty, T. L. and Kainen, P. C. *The Four-Color Problem: Assaults and Conquest.* New York: Dover, 1986.

Columbian Number

see SELF NUMBER

Colunar Triangle

Given a SCHWARZ TRIANGLE $(p\ q\ r)$, replacing each VERTEX with its antipodes gives the three colunar SPHERICAL TRIANGLES

$$(p\ q'\ r'), (p'\ q\ r'), (p'\ q'\ r),$$

where

$$\frac{1}{p} + \frac{1}{p'} = 1$$

$$\frac{1}{q} + \frac{1}{q'} = 1$$

$$\frac{1}{r} + \frac{1}{r'} = 1.$$

see also SCHWARZ TRIANGLE, SPHERICAL TRIANGLE

References

Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, p. 112, 1973.

Comb Function

see SHAH FUNCTION

Combination

The number of ways of picking r unordered outcomes from n possibilities. Also known as the BINOMIAL COEFFICIENT or CHOICE NUMBER and read "n choose r."

$${}_nC_r \equiv \binom{n}{r} \equiv \frac{n!}{r!(n-r)!},$$

where $n!$ is a FACTORIAL.

see also BINOMIAL COEFFICIENT, DERANGEMENT, FACTORIAL, PERMUTATION, SUBFACTORIAL

References

Conway, J. H. and Guy, R. K. "Choice Numbers." In *The Book of Numbers.* New York: Springer-Verlag, pp. 67–68, 1996.

Ruskey, F. "Information on Combinations of a Set." `http://sue.csc.uvic.ca/~cos/inf/comb/CombinationsInfo.html`.

Combination Lock

Let a combination of n buttons be a Sequence of disjoint nonempty Subsets of the Set $\{1, 2, \ldots, n\}$. If the number of possible combinations is denoted a_n, then a_n satisfies the Recurrence Relation

$$a_n = \sum_{i=0}^{n-1} \binom{n}{n-i} a_i, \tag{1}$$

with $a_0 = 1$. This can also be written

$$a_n = \frac{d^n}{dx^n}\left(\frac{1}{2-e^x}\right)\bigg|_{x=0} = \tfrac{1}{2}\sum_{k=0}^{\infty} \frac{k^n}{2^k}, \tag{2}$$

where the definition $0^0 = 1$ has been used. Furthermore,

$$a_n = \sum_{k=1}^{n} A_{n,k} 2^{n-k} = \sum_{k=1}^{n} A_{n,k} 2^{k-1}, \tag{3}$$

where $A_{n,k}$ are Eulerian Numbers. In terms of the Stirling Numbers of the Second Kind $s(n,k)$,

$$a_n = \sum_{k=1}^{n} k! s(n,k). \tag{4}$$

a_n can also be given in closed form as

$$a_n = \tfrac{1}{2}\operatorname{Li}_{-n}(\tfrac{1}{2}), \tag{5}$$

where $\operatorname{Li}_n(z)$ is the Polylogarithm. The first few values of a_n for $n = 1, 2, \ldots$ are 1, 3, 13, 75, 541, 4683, 47293, 545835, 7087261, 102247563, ... (Sloane's A000670).

The quantity

$$b_n \equiv \frac{a_n}{n!} \tag{6}$$

satisfies the inequality

$$\frac{1}{2(\ln 2)^n} \leq b_n \leq \frac{1}{(\ln 2)^n}. \tag{7}$$

References

Sloane, N. J. A. Sequence A000670/M2952 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Velleman, D. J. and Call, G. S. "Permutations and Combination Locks." *Math. Mag.* **68**, 243–253, 1995.

Combinatorial Species

see Species

Combinatorial Topology

Combinatorial topology is a special type of Algebraic Topology that uses Combinatorial methods. For example, Simplicial Homology is a combinatorial construction in Algebraic Topology, so it belongs to combinatorial topology.

see also Algebraic Topology, Simplicial Homology, Topology

Combinatorics

The branch of mathematics studying the enumeration, combination, and permutation of sets of elements and the mathematical relations which characterize these properties.

see also Antichain, Chain, Dilworth's Lemma, Diversity Condition, Erdős-Szekeres Theorem, Inclusion-Exclusion Principle, Kirkman's Schoolgirl Problem, Kirkman Triple System, Length (Partial Order), Partial Order, Pigeonhole Principle, Ramsey's Theorem, Schröder-Bernstein Theorem, Schur's Lemma, Sperner's Theorem, Total Order, van der Waerden's Theorem, Width (Partial Order)

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Combinatorial Analysis." Ch. 24 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 821–8827, 1972.

Aigner, M. *Combinatorial Theory.* New York: Springer-Verlag, 1997.

Bellman, R. and Hall, M. *Combinatorial Analysis.* Amer. Math. Soc., 1979.

Biggs, N. L. "The Roots of Combinatorics." *Historia Mathematica* **6**, 109–136, 1979.

Bose, R. C. and Manvel, B. *Introduction to Combinatorial Theory.* New York: Wiley, 1984.

Brown, K. S. "Combinatorics." `http://www.seanet.com/~ksbrown/icombina.htm`.

Cameron, P. J. *Combinatorics: Topics, Techniques, Algorithms.* New York: Cambridge University Press, 1994.

Cohen, D. *Basic Techniques of Combinatorial Theory.* New York: Wiley, 1978.

Cohen, D. E. *Combinatorial Group Theory: A Topological Approach.* New York: Cambridge University Press, 1989.

Colbourn, C. J. and Dinitz, J. H. *CRC Handbook of Combinatorial Designs.* Boca Raton, FL: CRC Press, 1996.

Comtet, L. *Advanced Combinatorics.* Dordrecht, Netherlands: Reidel, 1974.

Coolsaet, K. "Index of Combinatorial Objects." `http://www.hogent.be/~kc/ico/`.

Dinitz, J. H. and Stinson, D. R. (Eds.). *Contemporary Design Theory: A Collection of Surveys.* New York: Wiley, 1992.

Electronic Journal of Combinatorics. `http://www.combinatorics.org/previous_volumes.html`.

Eppstein, D. "Combinatorial Geometry." `http://www.ics.uci.edu/~eppstein/junkyard/combinatorial.html`.

Erickson, M. J. *Introduction to Combinatorics.* New York: Wiley, 1996.

Fields, J. "On-Line Dictionary of Combinatorics." `http://math.uic.edu/~fields/dic/`.

Godsil, C. D. "Problems in Algebraic Combinatorics." *Electronic J. Combinatorics* **2**, F1, 1–20, 1995. `http://www.combinatorics.org/Volume_2/volume2.html#F1`.

Graham, R. L.; Grötschel, M.; and Lovász, L. (Eds.). *Handbook of Combinatorics, 2 vols.* Cambridge, MA: MIT Press, 1996.

Graham, R. L.; Knuth, D. E.; and Patashnik, O. *Concrete Mathematics: A Foundation for Computer Science, 2nd ed.* Reading, MA: Addison-Wesley, 1994.

Hall, M. Jr. *Combinatorial Theory, 2nd ed.* New York: Wiley, 1986.

Knuth, D. E. (Ed.). *Stable Marriage and Its Relation to Other Combinatorial Problems.* Providence, RI: Amer. Math. Soc., 1997.

Kučera, L. *Combinatorial Algorithms.* Bristol, England: Adam Hilger, 1989.

Liu, C. L. *Introduction to Combinatorial Mathematics.* New York: McGraw-Hill, 1968.

MacMahon, P. A. *Combinatory Analysis.* New York: Chelsea, 1960.

Nijenhuis, A. and Wilf, H. *Combinatorial Algorithms for Computers and Calculators, 2nd ed.* New York: Academic Press, 1978.

Riordan, J. *Combinatorial Identities, reprint ed. with corrections.* Huntington, NY: Krieger, 1979.

Riordan, J. *An Introduction to Combinatorial Analysis.* New York: Wiley, 1980.

Roberts, F. S. *Applied Combinatorics.* Englewood Cliffs, NJ: Prentice-Hall, 1984.

Rota, G.-C. (Ed.). *Studies in Combinatorics.* Providence, RI: Math. Assoc. Amer., 1978.

Ruskey, F. "The (Combinatorial) Object Server." `http://sue.csc.uvic.ca/~cos`.

Ryser, H. J. *Combinatorial Mathematics.* Buffalo, NY: Math. Assoc. Amer., 1963.

Skiena, S. S. *Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica.* Reading, MA: Addison-Wesley, 1990.

Sloane, N. J. A. "An On-Line Version of the Encyclopedia of Integer Sequences." `http://www.research.att.com/~njas/sequences/eisonline.html`.

Sloane, N. J. A. and Plouffe, S. *The Encyclopedia of Integer Sequences.* San Diego, CA: Academic Press, 1995.

Street, A. P. and Wallis, W. D. *Combinatorial Theory: An Introduction.* Winnipeg, Manitoba: Charles Babbage Research Center, 1977.

Tucker, A. *Applied Combinatorics, 3rd ed.* New York: Wiley, 1995.

van Lint, J. H. and Wilson, R. M. *A Course in Combinatorics.* New York: Cambridge University Press, 1992.

Wilf, H. S. *Combinatorial Algorithms: An Update.* Philadelphia, PA: SIAM, 1989.

Comma Derivative

$$A_{,k} \equiv \frac{\partial A}{\partial x^k} \equiv \partial_k A$$

$$A^k_{,k} \equiv \frac{1}{g_k} \frac{\partial A^k}{\partial x^k} \equiv \partial_k A^k.$$

see also COVARIANT DERIVATIVE, SEMICOLON DERIVATIVE

Comma of Didymus

The musical interval by which four fifths exceed a seventeenth (i.e., two octaves and a major third),

$$\frac{(\frac{3}{2})^4}{2^2(\frac{5}{4})} = \frac{3^4}{2^4 \cdot 5} = \frac{81}{80} = 1.0125,$$

also called a SYNTONIC COMMA.

see also COMMA OF PYTHAGORAS, DIESIS, SCHISMA

Comma of Pythagoras

The musical interval by which twelve fifths exceed seven octaves,

$$\frac{(\frac{3}{2})^{12}}{2^7} = \frac{3^{12}}{2^{19}} = \frac{531441}{524288} = 1.013643265.$$

Successive CONTINUED FRACTION CONVERGENTS to $\log 2/\log(3/2)$ give increasingly close approximations m/n of m fifths by n octaves as 1, 2, 5/3, 12/7, 41/24, 53/31, 306/179, 665/389, ... (Sloane's A005664 and A046102; Jeans 1968, p. 188), shown in **bold** in the table below. All near-equalities of m fifths and n octaves having

$$R \equiv \frac{(\frac{3}{2})^m}{2^n} = \frac{3^m}{2^{m+n}}$$

with $|R-1| < 0.02$ are given in the following table.

m	n	Ratio	m	n	Ratio
12	**7**	1.013643265	265	155	1.010495356
41	**24**	0.9886025477	294	172	0.9855324037
53	**31**	1.002090314	**306**	**179**	0.9989782832
65	38	1.015762098	318	186	1.012607608
94	55	0.9906690375	347	203	0.9875924759
106	62	1.004184997	359	210	1.001066462
118	69	1.017885359	371	217	1.014724276
147	86	0.9927398469	400	234	0.9896568543
159	93	1.006284059	412	241	1.003159005
188	110	0.9814251419	424	248	1.016845369
200	117	0.994814985	453	265	0.9917255479
212	124	1.008387509	465	272	1.005255922
241	141	0.9834766286	477	279	1.018970895
253	148	0.9968944607	494	289	0.9804224033

see also COMMA OF DIDYMUS, DIESIS, SCHISMA

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 257, 1995.

Guy, R. K. "Small Differences Between Powers of 2 and 3." §F23 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 261, 1994.

Sloane, N. J. A. Sequences A005664 and A046102 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Common Cycloid

see CYCLOID

Common Residue

The value of b, where $a \equiv b \pmod{m}$, taken to be NONNEGATIVE and smaller than m.

see also MINIMAL RESIDUE, RESIDUE (CONGRUENCE)

Commutation Coefficient

A coefficient which gives the difference between partial derivatives of two coordinates with respect to the other coordinate,

$$c^{\mu}_{\alpha\beta}\vec{e}_{\mu} = [\vec{e}_{\alpha}, \vec{e}_{\beta}] = \nabla_{\alpha}\vec{e}_{\beta} - \nabla_{\beta}\vec{e}_{\alpha}.$$

see also CONNECTION COEFFICIENT

Commutative

Let A denote an $\mathbb{R}$-algebra, so that A is a VECTOR SPACE over R and

$$A \times A \to A$$
$$(x, y) \mapsto x \cdot y.$$

Now define

$$Z \equiv \{x \in a : x \cdot y \text{ for some } y \in A \neq 0\},$$

where $0 \in Z$. An ASSOCIATIVE $\mathbb{R}$-algebra is commutative if $x \cdot y = y \cdot x$ for all $x, y \in A$. Similarly, a RING is commutative if the MULTIPLICATION operation is commutative, and a LIE ALGEBRA is commutative if the COMMUTATOR $[A, B]$ is 0 for every A and B in the LIE ALGEBRA.

see also ABELIAN, ASSOCIATIVE, TRANSITIVE

References

Finch, S. "Zero Structures in Real Algebras." `http://www.mathsoft.com/asolve/zerodiv/zerodiv.html`.

MacDonald, I. G. and Atiyah, M. F. *Introduction to Commutative Algebra.* Reading, MA: Addison-Wesley, 1969.

Commutative Algebra

An ALGEBRA in which the + operators and × are COMMUTATIVE.

see also ALGEBRAIC GEOMETRY, GRÖBNER BASIS

References

MacDonald, I. G. and Atiyah, M. F. *Introduction to Commutative Algebra.* Reading, MA: Addison-Wesley, 1969.

Cox, D.; Little, J.; and O'Shea, D. *Ideals, Varieties, and Algorithms: An Introduction to Algebraic Geometry and Commutative Algebra, 2nd ed.* New York: Springer-Verlag, 1996.

Samuel, P. and Zariski, O. *Commutative Algebra, Vol. 2.* New York: Springer-Verlag, 1997.

Commutator

Let $\tilde{A}$, $\tilde{B}$, ... be OPERATORS. Then the commutator of $\tilde{A}$ and $\tilde{B}$ is defined as

$$[\tilde{A}, \tilde{B}] \equiv \tilde{A}\tilde{B} - \tilde{B}\tilde{A}. \tag{1}$$

Let a, b, ... be constants. Identities include

$$[f(x), x] = 0 \tag{2}$$
$$[\tilde{A}, \tilde{A}] = 0 \tag{3}$$
$$[\tilde{A}, \tilde{B}] = -[\tilde{B}, \tilde{A}] \tag{4}$$
$$[\tilde{A}, \tilde{B}\tilde{C}] = [\tilde{A}, \tilde{B}]\tilde{C} + \tilde{B}[\tilde{A}, \tilde{C}] \tag{5}$$
$$[\tilde{A}\tilde{B}, \tilde{C}] = [\tilde{A}, \tilde{C}]\tilde{B} + \tilde{A}[\tilde{B}, \tilde{C}] \tag{6}$$
$$[a + \tilde{A}, b + \tilde{B}] = [\tilde{A}, \tilde{B}] \tag{7}$$
$$[\tilde{A} + \tilde{B}, \tilde{C} + \tilde{D}] = [\tilde{A}, \tilde{C}] + [\tilde{A}, \tilde{D}] + [\tilde{B}, \tilde{C}] + [\tilde{B}, \tilde{D}]. \tag{8}$$

The commutator can be interpreted as the "infinitesimal" of the commutator of a LIE GROUP.

Let A and B be TENSORS. Then

$$[A, B] \equiv \nabla_A B - \nabla_B A. \tag{9}$$

see also ANTICOMMUTATOR, JACOBI IDENTITIES

Compact Group

If the parameters of a LIE GROUP vary over a CLOSED INTERVAL, the GROUP is compact. Every representation of a compact group is equivalent to a UNITARY representation.

Compact Manifold

A MANIFOLD which can be "charted" with finitely many EUCLIDEAN SPACE charts. The CIRCLE is the only compact 1-D MANIFOLD. The SPHERE and n-TORUS are the only compact 2-D MANIFOLDS. It is an open question if the known compact MANIFOLDS in 3-D are complete, and it is not even known what a complete list in 4-D should look like. The following terse table therefore summarizes current knowledge about the number of compact manifolds $N(D)$ of D dimensions.

D	$N(D)$
1	1
2	2

see also TYCHONOF COMPACTNESS THEOREM

Compact Set

The SET S is compact if, from any SEQUENCE of elements X_1, X_2, ... of S, a subsequence can always be extracted which tends to some limit element X of S. Compact sets are therefore closed and bounded.

Compact Space

A TOPOLOGICAL SPACE is compact if every open cover of X has a finite subcover. In other words, if X is the union of a family of open sets, there is a finite subfamily whose union is X. A subset A of a TOPOLOGICAL SPACE X is compact if it is compact as a TOPOLOGICAL SPACE with the relative topology (i.e., every family of open sets of X whose union contains A has a finite subfamily whose union contains A).

Compact Surface

A surface with a finite number of TRIANGLES in its TRIANGULATION. The SPHERE and TORUS are compact, but the PLANE and TORUS minus a DISK are not.

Compactness Theorem

Inside a BALL B in $\mathbb{R}^3$,

$$\{\text{rectifiable currents } S \text{ in } BL \text{ AREA } S \leq c, \quad \text{length } \partial S \leq c\}$$

is compact under the FLAT NORM.

References

Morgan, F. "What Is a Surface?" *Amer. Math. Monthly* **103**, 369–376, 1996.

Companion Knot

Let K_1 be a knot inside a TORUS. Now knot the TORUS in the shape of a second knot (called the companion knot) K_2. Then the new knot resulting from K_1 is called the SATELLITE KNOT K_3.

References
Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 115–118, 1994.

Comparability Graph

The comparability graph of a POSET $P = (X, \leq)$ is the GRAPH with vertex set X for which vertices x and y are adjacent IFF either $x \leq y$ or $y \leq x$ in P.

see also INTERVAL GRAPH, PARTIALLY ORDERED SET

Comparison Test

Let $\sum a_k$ and $\sum b_k$ be a SERIES with POSITIVE terms and suppose $a_1 \leq b_1$, $a_2 \leq b_2$, ...

1. If the bigger series CONVERGES, then the smaller series also CONVERGES.
2. If the smaller series DIVERGES, then the bigger series also DIVERGES.

see also CONVERGENCE TESTS

References
Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 280–281, 1985.

Compass

A tool with two arms joined at their ends which can be used to draw CIRCLES. In GEOMETRIC CONSTRUCTIONS, the classical Greek rules stipulate that the compass cannot be used to mark off distances, so it must "collapse" whenever one of its arms is removed from the page. This results in significant complication in the complexity of GEOMETRIC CONSTRUCTIONS.

see also CONSTRUCTIBLE POLYGON, GEOMETRIC CONSTRUCTION, GEOMETROGRAPHY, MASCHERONI CONSTRUCTION, PLANE GEOMETRY, POLYGON, PONCELET-STEINER THEOREM, RULER, SIMPLICITY, STEINER CONSTRUCTION, STRAIGHTEDGE

References
Dixon, R. "Compass Drawings." Ch. 1 in *Mathographics.* New York: Dover, pp. 1–78, 1991.

Compatible

Let $||\mathsf{A}||$ be the MATRIX NORM associated with the MATRIX A and $||\mathbf{x}||$ be the VECTOR NORM associated with a VECTOR $\mathbf{x}$. Let the product $\mathsf{A}\mathbf{x}$ be defined, then $||\mathsf{A}||$ and $||\mathbf{x}||$ are said to be compatible if

$$||\mathsf{A}\mathbf{x}|| \leq ||\mathsf{A}||\,||\mathbf{x}||.$$

References
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1115, 1980.

Complement Graph

The complement GRAPH $\bar{G}$ of G has the same VERTICES as G but contains precisely those two-element SUBSETS which are not in G.

Complement Knot

see KNOT COMPLEMENT

Complement Set

Given a set S with a subset E, the complement of E is defined as

$$E' \equiv \{F : F \in S, F \notin E\}. \tag{1}$$

If $E = S$, then

$$E' \equiv S' = \varnothing, \tag{2}$$

where $\varnothing$ is the EMPTY SET. Given a single SET, the second PROBABILITY AXIOM gives

$$1 = P(S) = P(E \cup E'). \tag{3}$$

Using the fact that $E \cap E' = \varnothing$,

$$1 = P(E) + P(E') \tag{4}$$

$$P(E') = 1 - P(E). \tag{5}$$

This demonstrates that

$$P(S') = P(\varnothing) = 1 - P(S) = 1 - 1 = 0. \tag{6}$$

Given two SETS,

$$P(E \cap F') = P(E) - P(E \cap F) \tag{7}$$

$$P(E' \cap F') = 1 - P(E) - P(F) + P(E \cap F). \tag{8}$$

Complementary Angle

Two ANGLES α and $\pi/2 - \alpha$ are said to be complementary.

see also ANGLE, SUPPLEMENTARY ANGLE

Complete

see COMPLETE AXIOMATIC THEORY, COMPLETE BIGRAPH, COMPLETE FUNCTIONS, COMPLETE GRAPH, COMPLETE QUADRANGLE, COMPLETE QUADRILATERAL, COMPLETE SEQUENCE, COMPLETE SPACE, COMPLETENESS PROPERTY, WEAKLY COMPLETE SEQUENCE

Complete Axiomatic Theory

An axiomatic theory (such as a GEOMETRY) is said to be complete if each valid statement in the theory is capable of being proven true or false.

see also CONSISTENCY

Complete Bigraph

see COMPLETE BIPARTITE GRAPH

Complete Bipartite Graph

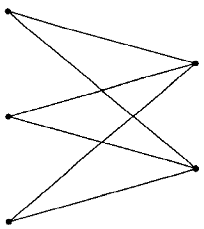

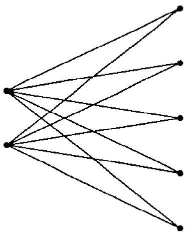

A BIPARTITE GRAPH (i.e., a set of VERTICES decomposed into two disjoint sets such that there are no two VERTICES within the same set are adjacent) such that every pair of VERTICES in the two sets are adjacent. If there are p and q VERTICES in the two sets, the complete bipartite graph (sometimes also called a COMPLETE BIGRAPH) is denoted $K_{p,q}$. The above figures show $K_{3,2}$ and $K_{2,5}$.

see also BIPARTITE GRAPH, COMPLETE GRAPH, COMPLETE k-PARTITE GRAPH, k-PARTITE GRAPH, THOMASSEN GRAPH, UTILITY GRAPH

References

Saaty, T. L. and Kainen, P. C. *The Four-Color Problem: Assaults and Conquest.* New York: Dover, p. 12, 1986.

Complete Functions

A set of ORTHONORMAL FUNCTIONS $\phi_n(x)$ is termed complete in the CLOSED INTERVAL $x \in [a,b]$ if, for every piecewise CONTINUOUS FUNCTION $f(x)$ in the interval, the minimum square error

$$E_n \equiv \|f - (c_1\phi_1 + \ldots + c_n\phi_n)\|^2$$

(where $\|$ denotes the NORM) converges to zero as n becomes infinite. Symbolically, a set of functions is complete if

$$\lim_{m\to\infty} \int_a^b \left[f(x) - \sum_{n=0}^{m} a_n\phi_n(x)\right]^2 w(x)\,dx = 0,$$

where $w(x)$ is a WEIGHTING FUNCTION and the above is a LEBESGUE INTEGRAL.

see also BESSEL'S INEQUALITY, HILBERT SPACE

References

Arfken, G. "Completeness of Eigenfunctions." §9.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 523–538, 1985.

Complete Graph

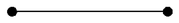

K_2

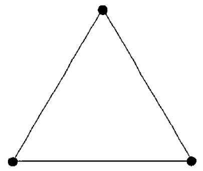

K_3

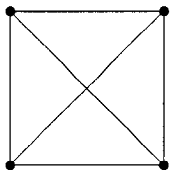

K_4

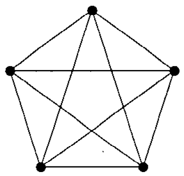

K_5

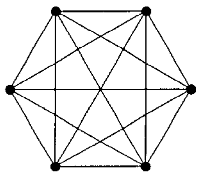

K_6

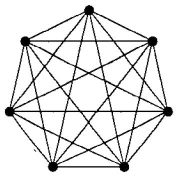

K_7

A GRAPH in which each pair of VERTICES is connected by an EDGE. The complete graph with n VERTICES is denoted K_n. In older literature, complete GRAPHS are called UNIVERSAL GRAPHS.

K_4 is the TETRAHEDRAL GRAPH and is therefore a PLANAR GRAPH. K_5 is nonplanar. Conway and Gordon (1983) proved that every embedding of K_6 is INTRINSICALLY LINKED with at least one pair of linked triangles. They also showed that any embedding of K_7 contains a knotted HAMILTONIAN CYCLE.

The number of EDGES in K_v is $v(v-1)/2$, and the GENUS is $(v-3)(v-4)/12$ for $v \geq 3$. The number of distinct variations for K_n (GRAPHS which cannot be transformed into each other without passing nodes through an EDGE or another node) for $n = 1, 2, \ldots$ are 1, 1, 1, 1, 1, 1, 6, 3, 411, 37, ... The ADJACENCY MATRIX of the complete graph takes the particularly simple form of all 1s with 0s on the diagonal.

It is not known in general if a set of TREES with $1, 2, \ldots, n-1$ EDGES can always be packed into K_n. However, if the choice of TREES is restricted to either the path or star from each family, then the packing can always be done (Zaks and Liu 1977, Honsberger 1985).

References

Chartrand, G. *Introductory Graph Theory.* New York: Dover, pp. 29–30, 1985.

Conway, J. H. and Gordon, C. M. "Knots and Links in Spatial Graphs." *J. Graph Th.* **7**, 445–453, 1983.

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 60–63, 1985.

Saaty, T. L. and Kainen, P. C. *The Four-Color Problem: Assaults and Conquest.* New York: Dover, p. 12, 1986.

Zaks, S. and Liu, C. L. "Decomposition of Graphs into Trees." *Proc. Eighth Southeastern Conference on Combinatorics, Graph Theory, and Computing.* pp. 643–654, 1977.

Complete k-Partite Graph

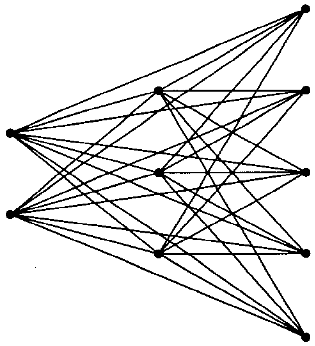

A k-PARTITE GRAPH (i.e., a set of VERTICES decomposed into k disjoint sets such that no two VERTICES within the same set are adjacent) such that every pair of VERTICES in the k sets are adjacent. If there are p, q, ..., r VERTICES in the k sets, the complete k-partite graph is denoted $K_{p,q,\ldots,r}$. The above figure shows $K_{2,3,5}$.

see also COMPLETE GRAPH, COMPLETE k-PARTITE GRAPH, k-PARTITE GRAPH

References

Saaty, T. L. and Kainen, P. C. *The Four-Color Problem: Assaults and Conquest.* New York: Dover, p. 12, 1986.

Complete Metric Space

A complete metric space is a METRIC SPACE in which every CAUCHY SEQUENCE is CONVERGENT. Examples include the REAL NUMBERS with the usual metric and the p-ADIC NUMBERS.

Complete Permutation

see DERANGEMENT

Complete Quadrangle

If the four points making up a QUADRILATERAL are joined pairwise by six distinct lines, a figure known as a complete quadrangle results. Note that a complete quadrilateral is defined differently from a COMPLETE QUADRANGLE.

The midpoints of the sides of any complete quadrangle and the three diagonal points all lie on a CONIC known as the NINE-POINT CONIC. If it is an ORTHOCENTRIC QUADRILATERAL, the CONIC reduces to a CIRCLE. The ORTHOCENTERS of the four TRIANGLES of a complete quadrangle are COLLINEAR on the RADICAL LINE of the CIRCLES on the diameters of a QUADRILATERAL.

see also COMPLETE QUADRANGLE, PTOLEMY'S THEOREM

References

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 230–231, 1969.

Demir, H. "The Compleat [sic] Cyclic Quadrilateral." *Amer. Math. Monthly* **79**, 777–778, 1972.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 61–62, 1929.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 101–104, 1990.

Complete Quadrilateral

The figure determined by four lines and their six points of intersection (Johnson 1929, pp. 61–62). Note that this is different from a COMPLETE QUADRANGLE. The midpoints of the diagonals of a complete quadrilateral are COLLINEAR (Johnson 1929, pp. 152–153).

A theorem due to Steiner (Mention 1862, Johnson 1929, Steiner 1971) states that in a complete quadrilateral, the bisectors of angles are CONCURRENT at 16 points which are the incenters and EXCENTERS of the four TRIANGLES. Furthermore, these points are the intersections of two sets of four CIRCLES each of which is a member of a conjugate coaxal system. The axes of these systems intersect at the point common to the CIRCUMCIRCLES of the quadrilateral.

see also COMPLETE QUADRANGLE, GAUSS-BODENMILLER THEOREM, POLAR CIRCLE

References

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 230–231, 1969.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 61–62, 149, 152–153, and 255–256, 1929.

Mention, M. J. "Démonstration d'un Théorème de M. Steiner." *Nouv. Ann. Math., 2nd Ser.* **1**, 16–20, 1862.

Mention, M. J. "Démonstration d'un Théorème de M. Steiner." *Nouv. Ann. Math., 2nd Ser.* **1**, 65–67, 1862.

Steiner, J. *Gesammelte Werke, 2nd ed, Vol. 1.* New York: Chelsea, p. 223, 1971.

Complete Residue System

A set of numbers a_0, a_1, ..., a_{m-1} (mod m) form a complete set of residues if they satisfy

$$a_i \equiv i \pmod{m}$$

for $i = 0, 1, \ldots, m-1$. In other words, a complete system of residues is formed by a base and a modulus if the residues r_i in $b^i \equiv r_i \pmod{m}$ for $i = 1, \ldots, m-1$ run through the values $1, 2, \ldots, m-1$.

see also HAUPT-EXPONENT

Complete Sequence

A SEQUENCE of numbers $V = \{\nu_n\}$ is complete if every POSITIVE INTEGER n is the sum of some subsequence of V, i.e., there exist $a_i = 0$ or 1 such that

$$n = \sum_{i=1}^{\infty} a_i \nu_i$$

(Honsberger 1985, pp. 123–126). The FIBONACCI NUMBERS are complete. In fact, dropping one number still leaves a complete sequence, although dropping two numbers does not (Honsberger 1985, pp. 123 and 126). The Sequence of Primes with the element $\{1\}$ prepended,

$$\{1, 2, 3, 5, 7, 11, 13, 17, 19, 23, \ldots\}$$

is complete, even if any number of Primes each > 7 are dropped, as long as the dropped terms do not include two consecutive Primes (Honsberger 1985, pp. 127–128). This is a consequence of Bertrand's Postulate.

see also Bertrand's Postulate, Brown's Criterion, Fibonacci Dual Theorem, Greedy Algorithm, Weakly Complete Sequence, Zeckendorf's Theorem

References

Brown, J. L. Jr. "Unique Representations of Integers as Sums of Distinct Lucas Numbers." *Fib. Quart.* **7**, 243–252, 1969.

Hoggatt, V. E. Jr.; Cox, N.; and Bicknell, M. "A Primer for Fibonacci Numbers. XII." *Fib. Quart.* **11**, 317–331, 1973.

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., 1985.

Complete Space

A Space of Complete Functions.

see also Complete Metric Space

Completely Regular Graph

A Polyhedral Graph is completely regular if the Dual Graph is also Regular. There are only five types. Let ρ be the number of Edges at each node, ρ^* the number of Edges at each node of the Dual Graph, V the number of Vertices, E the number of Edges, and F the number of faces in the Platonic Solid corresponding to the given graph. The following table summarizes the completely regular graphs.

Type	ρ	ρ^*	V	E	F
Tetrahedral	3	3	4	6	4
Cubical	3	4	8	12	6
Dodecahedral	3	5	20	39	12
Octahedral	4	3	6	12	8
Icosahedral	5	3	12	30	20

Completeness Property

All lengths can be expressed as Real Numbers.

Completing the Square

The conversion of an equation of the form $ax^2 + bx + c$ to the form

$$a\left(x + \frac{b}{2a}\right)^2 + \left(c - \frac{b^2}{4a}\right),$$

which, defining $B \equiv b/2a$ and $C \equiv c - b^2/4a$, simplifies to

$$a(x + B)^2 + C.$$

Complex

A finite Set of Simplexes such that no two have a common point. A 1-D complex is called a Graph.

see also CW-Complex, Simplicial Complex

Complex Analysis

The study of Complex Numbers, their Derivatives, manipulation, and other properties. Complex analysis is an extremely powerful tool with an unexpectedly large number of practical applications to the solution of physical problems. Contour Integration, for example, provides a method of computing difficult Integrals by investigating the singularities of the function in regions of the Complex Plane near and between the limits of integration.

The most fundamental result of complex analysis is the Cauchy-Riemann Equations, which give the conditions a Function must satisfy in order for a complex generalization of the Derivative, the so-called Complex Derivative, to exist. When the Complex Derivative is defined "everywhere," the function is said to be Analytic. A single example of the unexpected power of complex analysis is Picard's Theorem, which states that an Analytic Function assumes every Complex Number, with possibly one exception, infinitely often in any Neighborhood of an Essential Singularity!

see also Analytic Continuation, Branch Cut, Branch Point, Cauchy Integral Formula, Cauchy Integral Theorem, Cauchy Principal Value, Cauchy-Riemann Equations, Complex Number, Conformal Map, Contour Integration, de Moivre's Identity, Euler Formula, Inside-Outside Theorem, Jordan's Lemma, Laurent Series, Liouville's Conformality Theorem, Monogenic Function, Morera's Theorem, Permanence of Algebraic Form, Picard's Theorem, Pole, Polygenic Function, Residue (Complex Analysis)

References

Arfken, G. "Functions of a Complex Variable I: Analytic Properties, Mapping" and "Functions of a Complex Variable II: Calculus of Residues." Chs. 6–7 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 352–395 and 396–436, 1985.

Boas, R. P. *Invitation to Complex Analysis.* New York: Random House, 1987.

Churchill, R. V. and Brown, J. W. *Complex Variables and Applications, 6th ed.* New York: McGraw-Hill, 1995.

Conway, J. B. *Functions of One Complex Variable, 2nd ed.* New York: Springer-Verlag, 1995.

Forsyth, A. R. *Theory of Functions of a Complex Variable, 3rd ed.* Cambridge, England: Cambridge University Press, 1918.

Lang, S. *Complex Analysis, 3rd ed.* New York: Springer-Verlag, 1993.

Morse, P. M. and Feshbach, H. "Functions of a Complex Variable" and "Tabulation of Properties of Functions of Complex Variables." Ch. 4 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 348–491 and 480–485, 1953.

Complex Conjugate

The complex conjugate of a COMPLEX NUMBER $z \equiv a+bi$ is defined to be $z^* \equiv a-bi$. The complex conjugate is ASSOCIATIVE, $(z_1+z_2)^* = z_1{}^* + z_2{}^*$, since

$$\begin{aligned}(a_1+b_1i)^* + (a_2+b_2i)^* &= a_1 - ib_1 + a_2 - ib_2\\ &= (a_1 - ib_1) + (a_2 - ib_2)\\ &= (a_1+b_1)^* + (a_2+b_2)^*,\end{aligned}$$

and DISTRIBUTIVE, $(z_1z_2)^* = z_1{}^*z_2{}^*$, since

$$\begin{aligned}[(a_1+b_1i)(a_2+b_2i)]^* &= [(a_1a_2 - b_1b_2) + i(a_1b_2 + a_2b_1)]^*\\ &= (a_1a_2 - b_1b_2) - i(a_1b_2 + a_2b_1)\\ &= (a_1 - ib_1)(a_2 - ib_2)\\ &= (a_1 + ib_1)^*(a_2 + ib_2)^*.\end{aligned}$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 16, 1972.

Complex Derivative

A DERIVATIVE of a COMPLEX function, which must satisfy the CAUCHY-RIEMANN EQUATIONS in order to be COMPLEX DIFFERENTIABLE.

see also CAUCHY-RIEMANN EQUATIONS, COMPLEX DIFFERENTIABLE, DERIVATIVE

Complex Differentiable

If the CAUCHY-RIEMANN EQUATIONS are satisfied for a function $f(x) = u(x) + iv(x)$ and the PARTIAL DERIVATIVES of $u(x)$ and $v(x)$ are CONTINUOUS, then the COMPLEX DERIVATIVE df/dz exists.

see also ANALYTIC FUNCTION, CAUCHY-RIEMANN EQUATIONS, COMPLEX DERIVATIVE, PSEUDOANALYTIC FUNCTION

Complex Function

A FUNCTION whose RANGE is in the COMPLEX NUMBERS is said to be a complex function.

see also REAL FUNCTION, SCALAR FUNCTION, VECTOR FUNCTION

Complex Matrix

A MATRIX whose elements may contain COMPLEX NUMBERS. The MATRIX PRODUCT of two 2×2 complex matrices is given by

$$\begin{bmatrix} x_{11}+y_{11}i & x_{12}+y_{12}i\\ x_{21}+y_{21}i & x_{22}+y_{22}i\end{bmatrix}\begin{bmatrix} u_{11}+v_{11}i & u_{12}+v_{12}i\\ u_{21}+v_{21}i & u_{22}+v_{22}i\end{bmatrix} = \begin{bmatrix} R_{11} & R_{12}\\ R_{21} & R_{22}\end{bmatrix} + i\begin{bmatrix} I_{11} & I_{12}\\ I_{21} & I_{22}\end{bmatrix},$$

where

$$\begin{aligned}R_{11} &= u_{11}x_{11} + u_{21}x_{12} - v_{11}y_{11} - v_{21}y_{12}\\ R_{12} &= u_{12}x_{11} + u_{22}x_{12} - v_{12}y_{11} - v_{22}y_{12}\\ R_{21} &= u_{11}x_{21} + u_{21}x_{22} - v_{11}y_{21} - v_{21}y_{22}\\ R_{22} &= u_{12}x_{21} + u_{22}x_{22} - v_{12}y_{21} - v_{22}y_{22}\\ I_{11} &= v_{11}x_{11} + v_{21}x_{12} + u_{11}y_{11} + u_{21}y_{12}\\ I_{12} &= v_{12}x_{11} + v_{22}x_{12} + u_{12}y_{11} + u_{22}y_{12}\\ I_{21} &= v_{11}x_{21} + v_{21}x_{22} + u_{11}y_{21} + u_{21}y_{22}\\ I_{22} &= v_{12}x_{21} + v_{22}x_{22} + u_{12}y_{21} + u_{22}y_{22}.\end{aligned}$$

see also REAL MATRIX

Complex Multiplication

Two COMPLEX NUMBERS $x = a + ib$ and $y = c + id$ are multiplied as follows:

$$\begin{aligned}xy = (a+ib)(c+id) &= ac + ibc + iad - bd\\ &= (ac - bd) + i(ad + bc).\end{aligned}$$

However, the multiplication can be carried out using only three REAL multiplications, ac, bd, and $(a+b)(c+d)$ as

$$\begin{aligned}\Re[(a+ib)(c+id)] &= ac - bd\\ \Im[(a+ib)(c+id)] &= (a+b)(c+d) - ac - bd.\end{aligned}$$

Complex multiplication has a special meaning for ELLIPTIC CURVES.

see also COMPLEX NUMBER, ELLIPTIC CURVE, IMAGINARY PART, MULTIPLICATION, REAL PART

References

Cox, D. A. *Primes of the Form x^2+ny^2: Fermat, Class Field Theory and Complex Multiplication.* New York: Wiley, 1997.

Complex Number

The complex numbers are the FIELD $\mathbb{C}$ of numbers of the form $x + iy$, where x and y are REAL NUMBERS and i is the IMAGINARY NUMBER equal to $\sqrt{-1}$. When a single letter $z = x + iy$ is used to denote a complex number, it is sometimes called an "AFFIX." The FIELD of complex numbers includes the FIELD of REAL NUMBERS as a SUBFIELD.

Through the EULER FORMULA, a complex number

$$z = x + iy \tag{1}$$

may be written in "PHASOR" form

$$z = |z|(\cos\theta + i\sin\theta) = |z|e^{i\theta}. \tag{2}$$

Here, $|z|$ is known as the MODULUS and θ is known as the ARGUMENT or PHASE. The ABSOLUTE SQUARE of

z is defined by $|z|^2 = zz^*$, and the argument may be computed from

$$\arg(z) = \theta = \tan^{-1}\left(\frac{y}{x}\right). \tag{3}$$

DE MOIVRE'S IDENTITY relates POWERS of complex numbers

$$z^n = |z|^n[\cos(n\theta) + i\sin(n\theta)]. \tag{4}$$

Finally, the REAL $\Re(z)$ and IMAGINARY PARTS $\Im(z)$ are given by

$$\Re(z) = \tfrac{1}{2}(z + z^*) \tag{5}$$

$$\Im(z) = \frac{z - z^*}{2i} = -\tfrac{1}{2}i(z - z^*) = \tfrac{1}{2}i(z^* - z). \tag{6}$$

The POWERS of complex numbers can be written in closed form as follows:

$$z^n = \left[x^n - \binom{n}{2}x^{n-2}y^2 + \binom{n}{4}x^{n-4}y^4 - \ldots\right] + i\left[\binom{n}{1}x^{n-1}y - \binom{n}{3}x^{n-3}y^3 + \ldots\right]. \tag{7}$$

The first few are explicitly

$$z^2 = (x^2 - y^2) + i(2xy) \tag{8}$$

$$z^3 = (x^3 - 3xy^2) + i(3x^2y - y^3) \tag{9}$$

$$z^4 = (x^4 - 6x^2y^2 + y^4) + i(4x^3y - 4xy^3) \tag{10}$$

$$z^5 = (x^5 - 10x^3y^2 + 5xy^4) + i(5x^4y - 10x^2y^3 + y^5) \tag{11}$$

(Abramowitz and Stegun 1972).

see also ABSOLUTE SQUARE, ARGUMENT (COMPLEX NUMBER), COMPLEX PLANE, I, IMAGINARY NUMBER, MODULUS, PHASE, PHASOR, REAL NUMBER, SURREAL NUMBER

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 16–17, 1972.
Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 353–357, 1985.
Courant, R. and Robbins, H. "Complex Numbers." §2.5 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 88–103, 1996.
Morse, P. M. and Feshbach, H. "Complex Numbers and Variables." §4.1 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 349–356, 1953.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Complex Arithmetic." §5.4 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 171–172, 1992.

Complex Plane

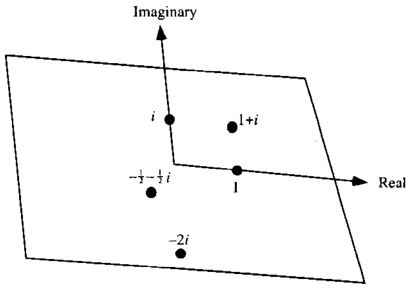

The plane of COMPLEX NUMBERS spanned by the vectors 1 and i, where i is the IMAGINARY NUMBER. Every COMPLEX NUMBER corresponds to a unique POINT in the complex plane. The LINE in the plane with $i = 0$ is the REAL LINE. The complex plane is sometimes called the ARGAND PLANE or GAUSS PLANE, and a plot of COMPLEX NUMBERS in the plane is sometimes called an ARGAND DIAGRAM.

see also AFFINE COMPLEX PLANE, ARGAND DIAGRAM, ARGAND PLANE, BERGMAN SPACE, COMPLEX PROJECTIVE PLANE

References
Courant, R. and Robbins, H. "The Geometric Interpretation of Complex Numbers." §5.2 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 92–97, 1996.

Complex Projective Plane

The set $\mathbb{P}^2$ is the set of all EQUIVALENCE CLASSES $[a, b, c]$ of ordered triples $(a, b, c) \in \mathbb{C}^3 \backslash (0, 0, 0)$ under the equivalence relation $(a, b, c) \sim (a', b', c')$ if $(a, b, c) = (\lambda a', \lambda b', \lambda c')$ for some NONZERO COMPLEX NUMBER λ.

Complex Representation

see PHASOR

Complex Structure

The complex structure of a point $\mathbf{x} = x_1, x_2$ in the PLANE is defined by the linear MAP $J : \mathbb{R}^2 \to \mathbb{R}^2$

$$J(x_1, x_2) = (-x_2, x_1),$$

and corresponds to a clockwise rotation by $\pi/2$. This map satisfies

$$J^2 = -I$$

$$(J\mathbf{x}) \cdot (J\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$$

$$(J\mathbf{x}) \cdot \mathbf{x} = 0,$$

where I is the IDENTITY MAP.

More generally, if V is a 2-D VECTOR SPACE, a linear map $J : V \to V$ such that $J^2 = -I$ is called a complex structure on V. If $V = \mathbb{R}^2$, this collapses to the previous definition.

References
Gray, A. *Modern Differential Geometry of Curves and Surfaces.*Boca Raton, FL: CRC Press, pp. 3 and 229, 1993.

Complexity (Number)

The number of 1s needed to represent an INTEGER using only additions, multiplications, and parentheses are called the integer's complexity. For example,

$$\begin{aligned}
1 &= 1 \\
2 &= 1+1 \\
3 &= 1+1+1 \\
4 &= (1+1)(1+1) = 1+1+1+1 \\
5 &= (1+1)(1+1)+1 = 1+1+1+1+1 \\
6 &= (1+1)(1+1+1) \\
7 &= (1+1)(1+1+1)+1 \\
8 &= (1+1)(1+1)(1+1) \\
9 &= (1+1+1)(1+1+1) \\
10 &= (1+1+1)(1+1+1)+1 \\
&= (1+1)(1+1+1+1+1)
\end{aligned}$$

So, for the first few n, the complexity is 1, 2, 3, 4, 5, 5, 6, 6, 6, 7, 8, 7, 8, ... (Sloane's A005245).

References

Guy, R. K. "Expressing Numbers Using Just Ones." §F26 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 263, 1994.

Guy, R. K. "Some Suspiciously Simple Sequences." *Amer. Math. Monthly* **93**, 186–190, 1986.

Guy, R. K. "Monthly Unsolved Problems, 1969–1987." *Amer. Math. Monthly* **94**, 961–970, 1987.

Guy, R. K. "Unsolved Problems Come of Age." *Amer. Math. Monthly* **96**, 903–909, 1989.

Rawsthorne, D. A. "How Many 1's are Needed?" *Fib. Quart.* **27**, 14–17, 1989.

Sloane, N. J. A. Sequence A005245/M0457 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Complexity (Sequence)

see BLOCK GROWTH

Complexity Theory

Divides problems into "easy" and "hard" categories. A problem is easy and assigned to the P-PROBLEM (POLYNOMIAL time) class if the number of steps needed to solve it is bounded by some POWER of the problem's size. A problem is hard and assigned to the NP-PROBLEM (nondeterministic POLYNOMIAL time) class if the number of steps is not bounded and may grow exponentially.

However, if a solution is known to an NP-PROBLEM, it can be reduced to a single period verification. A problem is NP-COMPLETE if an ALGORITHM for solving it can be translated into one for solving any other NP-PROBLEM. Examples of NP-COMPLETE PROBLEMS include the HAMILTONIAN CYCLE and TRAVELING SALESMAN PROBLEMS. LINEAR PROGRAMMING, thought to be an NP-PROBLEM, was shown to actually be a P-PROBLEM by L. Khachian in 1979. It is not known if all apparently NP-PROBLEMS are actually P-PROBLEMS.

see also BIT COMPLEXITY, NP-COMPLETE PROBLEM, NP-PROBLEM, P-PROBLEM

References

Bridges, D. S. *Computability.* New York: Springer-Verlag, 1994.

Brookshear, J. G. *Theory of Computation: Formal Languages, Automata, and Complexity.* Redwood City, CA: Benjamin/Cummings, 1989.

Cooper, S. B.; Slaman, T. A.; and Wainer, S. S. (Eds.). *Computability, Enumerability, Unsolvability: Directions in Recursion Theory.* New York: Cambridge University Press, 1996.

Garey, M. R. and Johnson, D. S. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* New York: W. H. Freeman, 1983.

Goetz, P. "Phil Goetz's Complexity Dictionary." `http://www.cs.buffalo.edu/~goetz/dict.html`.

Hopcroft, J. E. and Ullman, J. D. *Introduction to Automated Theory, Languages, and Computation.* Reading, MA: Addison-Wesley, 1979.

Lewis, H. R. and Papadimitriou, C. H. *Elements of the Theory of Computation, 2nd ed.* Englewood Cliffs, NJ: Prentice-Hall, 1997.

Sudkamp, T. A. *Language and Machines: An Introduction to the Theory of Computer Science, 2nd ed.* Reading, MA: Addison-Wesley, 1996.

Welsh, D. J. A. *Complexity: Knots, Colourings and Counting.* New York: Cambridge University Press, 1993.

Component

A GROUP L is a component of H if L is a QUASISIMPLE GROUP which is a SUBNORMAL SUBGROUP of H.

see also GROUP, QUASISIMPLE GROUP, SUBGROUP, SUBNORMAL

Composite Knot

A KNOT which is not a PRIME KNOT. Composite knots are special cases of SATELLITE KNOTS.

see also KNOT, PRIME KNOT, SATELLITE KNOT

Composite Number

A POSITIVE INTEGER which is not PRIME (i.e., which has FACTORS other than 1 and itself).

A composite number C can always be written as a PRODUCT in at least two ways (since $1 \cdot C$ is always possible). Call these two products

$$C = ab = cd, \tag{1}$$

then it is obviously the case that $C|ab$ (C divides ab). Set

$$c = mn, \tag{2}$$

where m is the part of C which divides a, and n the part of C which divides n. Then there are p and q such that

$$a = mp \tag{3}$$

$$b = nq. \tag{4}$$

Solving $ab = cd$ for d gives

$$d = \frac{ab}{c} = \frac{(mp)(nq)}{mn} = pq. \tag{5}$$

It then follows that

$$\begin{aligned} S &\equiv a^2 + b^2 + c^2 + d^2 \\ &= m^2p^2 + n^2q^2 + m^2n^2 + p^2q^2 \\ &= (m^2 + q^2)(n^2 + p^2). \end{aligned} \tag{6}$$

It therefore follows that $a^2 + b^2 + c^2 + d^2$ is never Prime! In fact, the more general result that

$$S \equiv a^k + b^k + c^k + d^k \tag{7}$$

is never Prime for k an Integer ≥ 0 also holds (Honsberger 1991).

There are infinitely many integers of the form $\lfloor (3/2)^n \rfloor$ and $\lfloor (4/3)^n \rfloor$ which are composite, where $\lfloor x \rfloor$ is the Floor Function (Forman and Shapiro, 1967; Guy 1994, p. 220). The first few composite $\lfloor (3/2)^n \rfloor$ occur for $n = 8$, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 23, ..., and the the few composite $\lfloor (4/3)^n \rfloor$ occur for $n = 5$, 8, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, ...

see also Amenable Number, Grimm's Conjecture, Highly Composite Number, Prime Factorization Prime Gaps, Prime Number

References

Forman, W. and Shapiro, H. N. "An Arithmetic Property of Certain Rational Powers." *Comm. Pure Appl. Math.* **20**, 561–573, 1967.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, 1994.

Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., pp. 19–20, 1991.

Sloane, N. J. A. Sequence A002808/M3272 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Composite Runs

see Prime Gaps

Compositeness Certificate

A compositeness certificate is a piece of information which guarantees that a given number p is Composite. Possible certificates consist of a Factor of a number (which, in general, is much quicker to check by direct division than to determine initially), or of the determination that either

$$a^{p-1} \not\equiv 1 \pmod{p},$$

(i.e., p violates Fermat's Little Theorem), or

$$a \neq -1, 1 \text{ and } a^2 \equiv 1 \pmod{p}.$$

A quantity a satisfying either property is said to be a Witness to p's compositeness.

see also Adleman-Pomerance-Rumely Primality Test, Fermat's Little Theorem, Miller's Primality Test, Primality Certificate, Witness

Compositeness Test

A test which always identifies Prime numbers correctly, but may incorrectly identify a Composite Number as a Prime.

see also Primality Test

Composition

The combination of two Functions to form a single new Operator. The composition of two functions f and g is denoted $f \circ g$ and is defined by

$$f \circ g = f(g(x))$$

when f and g are both functions of x.

An operation called composition is also defined on Binary Quadratic Forms. For two numbers represented by two forms, the product can then be represented by the composition. For example, the composition of the forms $2x^2 + 15y^2$ and $3x^2 + 10y^2$ is given by $6x^2 + 5y^2$, and in this case, the product of 17 and 13 would be represented as $(6 \cdot 36 + 5 \cdot 1 = 221)$. There are several algorithms for computing binary quadratic form composition, which is the basis for some factoring methods.

see also Adem Relations, Binary Operator, Binary Quadratic Form

Composition Series

Every Finite Group G of order greater than one possesses a finite series of Subgroups, called a composition series, such that

$$I \subset H_s \subset \ldots \subset H_2 \subset H_1 \subset G,$$

where H_{i+1} is a maximal subgroup of H_i. The Quotient Groups G/H_1, H_1/H_2, ..., H_{s-1}/H_s, H_s are called composition quotient groups.

see also Finite Group, Jordan-Hölder Theorem, Quotient Group, Subgroup

References

Lomont, J. S. *Applications of Finite Groups.* New York: Dover, p. 26, 1993.

Composition Theorem

Let

$$Q(x, y) \equiv x^2 + y^2.$$

Then

$$Q(x, y)Q(x', y') = Q(xx' - yy', x'y + xy'),$$

since

$$\begin{aligned} (x^2 + y^2)(x'^2 + y'^2) &= (xx' - yy')^2 + (xy' + x'y)^2 \\ &= x^2x'^2 + y^2y'^2 + x'^2y^2 + x^2y'^2. \end{aligned}$$

see also Genus Theorem

Compound Interest

Let P be the PRINCIPAL (initial investment), r be the annual compounded rate, $i^{(n)}$ the "nominal rate," n be the number of times INTEREST is compounded per year (i.e., the year is divided into n CONVERSION PERIODS), and t be the number of years (the "term"). The INTEREST rate per CONVERSION PERIOD is then

$$r \equiv \frac{i^{(n)}}{n}. \tag{1}$$

If interest is compounded n times at an annual rate of r (where, for example, 10% corresponds to $r = 0.10$), then the effective rate over $1/n$ the time (what an investor would earn if he did not redeposit his interest after each compounding) is

$$(1+r)^{1/n}. \tag{2}$$

The total amount of holdings A after a time t when interest is re-invested is then

$$A = P\left(1 + \frac{i^{(n)}}{n}\right)^{nt} = P(1+r)^{nt}. \tag{3}$$

Note that even if interest is compounded continuously, the return is still finite since

$$\lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n = e, \tag{4}$$

where e is the base of the NATURAL LOGARITHM.

The time required for a given PRINCIPAL to double (assuming $n = 1$ CONVERSION PERIOD) is given by solving

$$2P = P(1+r)^t, \tag{5}$$

or

$$t = \frac{\ln 2}{\ln(1+r)}, \tag{6}$$

where LN is the NATURAL LOGARITHM. This function can be approximated by the so-called RULE OF 72:

$$t \approx \frac{0.72}{r}. \tag{7}$$

see also e, INTEREST, LN, NATURAL LOGARITHM, PRINCIPAL, RULE OF 72, SIMPLE INTEREST

References

Kellison, S. G. *The Theory of Interest, 2nd ed.* Burr Ridge, IL: Richard D. Irwin, pp. 14-16, 1991.

Milanfar, P. "A Persian Folk Method of Figuring Interest." *Math. Mag.* **69**, 376, 1996.

Compound Polyhedron

see POLYHEDRON COMPOUND

Computability

see COMPLEXITY THEORY

Computable Function

Any computable function can be incorporated into a PROGRAM using while-loops (i.e., "while something is true, do something else"). For-loops (which have a fixed iteration limit) are a special case of while-loops, so computable functions could also be coded using a combination of for- and while-loops. The ACKERMANN FUNCTION is the simplest example of a well-defined TOTAL FUNCTION which is computable but not PRIMITIVE RECURSIVE, providing a counterexample to the belief in the early 1900s that every computable function was also primitive recursive (Dötzel 1991).

see also ACKERMANN FUNCTION, CHURCH'S THESIS, COMPUTABLE NUMBER, PRIMITIVE RECURSIVE FUNCTION, TURING MACHINE

References

Dötzel, G. "A Function to End All Functions." *Algorithm: Recreational Programming* **2**, 16–17, 1991.

Computable Number

A number which can be computed to any number of DIGITS desired by a TURING MACHINE. Surprisingly, most IRRATIONALS are not computable numbers!

References

Penrose, R. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics.* Oxford, England: Oxford University Press, 1989.

Computational Complexity

see COMPLEXITY THEORY

Concatenated Number Sequences

see CONSECUTIVE NUMBER SEQUENCES

Concatenation

The concatenation of two strings a and b is the string ab formed by joining a and b. Thus the concatenation of the strings "book" and "case" is the string "bookcase". The concatenation of two strings a and b is often denoted ab, $a||b$, or (in *Mathematica*® (Wolfram Research, Champaign, IL) $a <> b$. Concatenation is an associative operation, so that the concatenation of three or more strings, for example abc, $abcd$, etc., is well-defined.

The concatenation of two or more numbers is the number formed by concatenating their numerals. For example, the concatenation of 1, 234, and 5678 is 12345678. The value of the result depends on the numeric base, which is typically understood from context.

The formula for the concatenation of numbers p and q in base b is

$$p||q = pb^{l(q)} + q,$$

where

$$l(q) = \lfloor \log_b q \rfloor + 1$$

is the LENGTH of q in base b and $\lfloor x \rfloor$ is the FLOOR FUNCTION.

see also CONSECUTIVE NUMBER SEQUENCES, LENGTH (NUMBER), SMARANDACHE SEQUENCES

Concave

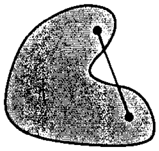

concave

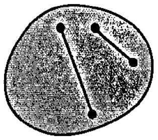

convex

A SET in $\mathbb{R}^d$ is concave if it *does not* contain all the LINE SEGMENTS connecting any pair of its points. If the SET *does* contain all the LINE SEGMENTS, it is called CONVEX.

see also CONNECTED SET, CONVEX FUNCTION, CONVEX HULL, CONVEX OPTIMIZATION THEORY, CONVEX POLYGON, DELAUNAY TRIANGULATION, SIMPLY CONNECTED

Concave Function

A function $f(x)$ is said to be concave on an interval $[a,b]$ if, for any points x_1 and x_2 in $[a,b]$, the function $-f(x)$ is CONVEX on that interval. If the second DERIVATIVE of f

$$f''(x) > 0,$$

on an open interval (a,b) (where $f''(x)$ is the second DERIVATIVE), then f is concave up on the interval. If

$$f''(x) < 0$$

on the interval, then f is concave down on it.

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1100, 1980.

Concentrated

Let μ be a POSITIVE MEASURE on a SIGMA ALGEBRA M, and let λ be an arbitrary (real or complex) MEASURE on M. If there is a SET $A \in M$ such that $\lambda(E) = \lambda(A \cap E)$ for every $E \in M$, then *lambda* is said to be concentrated on A. This is equivalent to requiring that $\lambda(E) = 0$ whenever $E \cap A = \varnothing$.

see also ABSOLUTELY CONTINUOUS, MUTUALLY SINGULAR

References

Rudin, W. *Functional Analysis.* New York: McGraw-Hill, p. 121, 1991.

Concentric

Two geometric figures are said to be concentric if their CENTERS coincide. The region between two concentric CIRCLES is called an ANNULUS.

see also ANNULUS, CONCENTRIC CIRCLES, CONCYCLIC, ECCENTRIC

Concentric Circles

The region between two CONCENTRIC circles of different RADII is called an ANNULUS.

Given two concentric circles with RADII R and $2R$, what is the probability that a chord chosen at random from the outer circle will cut across the inner circle? Depending on how the "random" CHORD is chosen, 1/2, 1/3, or 1/4 could all be correct answers.

1. Picking any two points on the outer circle and connecting them gives 1/3.
2. Picking any random point on a diagonal and then picking the CHORD that perpendicularly bisects it gives 1/2.
3. Picking any point on the large circle, drawing a line to the center, and then drawing the perpendicularly bisected CHORD gives 1/4.

So some care is obviously needed in specifying what is meant by "random" in this problem.

Given an arbitrary CHORD BB' to the larger of two concentric CIRCLES centered on O, the distance between inner and outer intersections is equal on both sides ($AB = A'B'$). To prove this, take the PERPENDICULAR to BB' passing through O and crossing at P. By symmetry, it must be true that PA and PA' are equal. Similarly, PB and PB' must be equal. Therefore, $PB - PA = AB$ equals $PB' - PA' = A'B'$. Incidentally, this is also true for HOMEOIDS, but the proof is nontrivial.

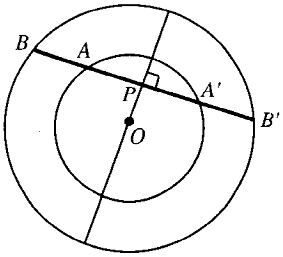

see also ANNULUS

Concho-Spiral

The SPACE CURVE with parametric equations

$$r = \mu^u a$$
$$\theta = u$$
$$z = \mu^u c.$$

see also CONICAL SPIRAL, SPIRAL

Conchoid

A curve whose name means "shell form." Let C be a curve and O a fixed point. Let P and P' be points on a line from O to C meeting it at Q, where $P'Q = QP = k$, with k a given constant. For example, if C is a CIRCLE and O is on C, then the conchoid is a LIMAÇON, while in the special case that k is the DIAMETER of C, then the conchoid is a Cardioid. The equation for a parametrically represented curve $(f(t), g(t))$ with $O = (x_0, y_0)$ is

$$x = f \pm \frac{k(f - x_0)}{\sqrt{(f - x_0)^2 + (g - y_0)^2}}$$
$$y = g \pm \frac{k(g - y_0)}{\sqrt{(f - x_0)^2 + (g - y_0)^2}}.$$

see also Concho-Spiral, Conchoid of de Sluze, Conchoid of Nicomedes, Conical Spiral, Dürer's Conchoid

References
Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 49–51, 1972.
Lee, X. "Conchoid." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Conchoid_dir/conchoid.html`.
Lockwood, E. H. "Conchoids." Ch. 14 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 126–129, 1967.
Yates, R. C. "Conchoid." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 31–33, 1952.

Conchoid of de Sluze

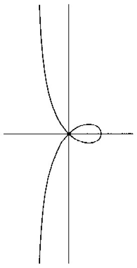

A curve first constructed by René de Sluze in 1662. In Cartesian Coordinates,

$$a(x - a)(x^2 + y^2) = k^2 x^2,$$

and in Polar Coordinates,

$$r = \frac{k^2 \cos\theta}{a} + a \sec\theta.$$

The above curve has $k^2/a = 1$, $a = -0.5$.

Conchoid of Nicomedes

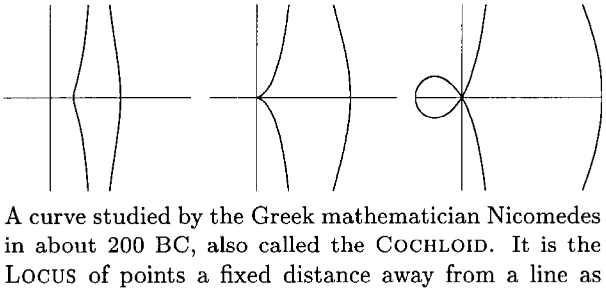

A curve studied by the Greek mathematician Nicomedes in about 200 BC, also called the Cochloid. It is the Locus of points a fixed distance away from a line as measured along a line from the Focus point (MacTutor Archive). Nicomedes recognized the three distinct forms seen in this family. This curve was a favorite with 17th century mathematicians and could be used to solve the problems of Cube Duplication and Angle Trisection.

In Polar Coordinates,

$$r = b + a \sec\theta. \tag{1}$$

In Cartesian Coordinates,

$$(x - a)^2(x^2 + y^2) = b^2 x^2. \tag{2}$$

The conchoid has $x = a$ as an asymptote and the Area between either branch and the Asymptote is infinite. The Area of the loop is

$$A = a\sqrt{b^2 - a^2} - 2ab \ln\left(\frac{b + \sqrt{b^2 - a^2}}{a}\right) + b^2 \cos^{-1}\left(\frac{a}{b}\right). \tag{3}$$

see also Conchoid

References
Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 135–139, 1972.
Lee, X. "Conchoid of Nicomedes." `http://www.best.com/~xah/SpecialPlaneCurves_dir/ConchoidOfNicomedes_dir/conchoidOfNicomedes.html`.
MacTutor History of Mathematics Archive. "Conchoid." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Conchoid.html`.
Pappas, T. "Conchoid of Nicomedes." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 94–95, 1989.
Yates, R. C. "Conchoid." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 31–33, 1952.

Concordant Form

A concordant form is an integer Triple (a, b, N) where

$$\begin{cases} a^2 + b^2 = c^2 \\ a^2 + Nb^2 = d^2, \end{cases}$$

with c and d integers. Examples include

$$\begin{cases} 14663^2 + 111384^2 = 112345^2 \\ 14663^2 + 47 \cdot 111384^2 = 763751^2 \end{cases}$$
$$\begin{cases} 1141^2 + 13260^2 = 13309^2 \\ 1141^2 + 53 \cdot 13260^2 = 96541^2 \end{cases}$$
$$\begin{cases} 2873161^2 + 2401080^2 = 3744361^2 \\ 2873161^2 + 83 \cdot 2401080^2 = 22062761^2. \end{cases}$$

Dickson (1962) states that C. H. Brooks and S. Watson found in *The Ladies' and Gentlemen's Diary* (1857) that $x^2 + y^2$ and $x^2 + Ny^2$ can be simultaneously squares for $N < 100$ only for 1, 7, 10, 11, 17, 20, 22, 23, 24, 27, 30, 31, 34, 41, 42, 45, 49, 50, 52, 57, 58, 59, 60, 61, 68, 71, 72, 74, 76, 77, 79, 82, 85, 86, 90, 92, 93, 94, 97, 99, and 100 (which evidently omits 47, 53, and 83 from above). The list of concordant primes less than 1000 is now complete with the possible exception of the 16 primes 103, 131, 191, 223, 271, 311, 431, 439, 443, 593, 607, 641, 743, 821, 929, and 971 (Brown).

see also Congruum

References

Brown, K. S. "Concordant Forms." `http://www.seanet.com/~ksbrown/kmath286.htm`.

Dickson, L. E. *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, p. 475, 1952.

Concur

Two or more lines which intersect in a Point are said to concur.

see also Concurrent

Concurrent

Two or more Lines are said to be concurrent if they intersect in a single point. Two Lines concur if their Trilinear Coordinates satisfy

$$\begin{vmatrix} l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \\ l_3 & m_3 & n_3 \end{vmatrix} = 0. \tag{1}$$

Three Lines concur if their Trilinear Coordinates satisfy

$$l_1\alpha + m_1\beta + n_1\gamma = 0 \tag{2}$$

$$l_2\alpha + m_2\beta + n_2\gamma = 0 \tag{3}$$

$$l_3\alpha + m_3\beta + n_3\gamma = 0, \tag{4}$$

in which case the point is

$$m_2n_3 - n_2m_3 : n_2l_3 - l_2n_3 : l_2m_3 - m_2l_3. \tag{5}$$

Three lines

$$A_1x + B_1y + C_1 = 0 \tag{6}$$

$$A_2x + B_2y + C_2 = 0 \tag{7}$$

$$A_3x + B_3y + C_3 = 0. \tag{8}$$

are concurrent if their Coefficients satisfy

$$\begin{vmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A_3 & B_3 & C_3 \end{vmatrix} = 0. \tag{9}$$

see also Concyclic, Point

Concyclic

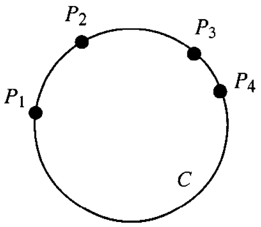

Four or more points P_1, P_2, P_3, P_4, ... which lie on a Circle C are said to be concyclic. Three points are trivially concyclic since three noncollinear points determine a Circle. The number of the n^2 Lattice Points $x, y \in [1, n]$ which can be picked with no four concyclic is $\mathcal{O}(n^{2/3} - \epsilon)$ (Guy 1994).

A theorem states that if any four consecutive points of a Polygon are not concyclic, then its Area can be increased by making them concyclic. This fact arises in some Proofs that the solution to the Isoperimetric Problem is the Circle.

see also Circle, Collinear, Concentric, Cyclic Hexagon, Cyclic Pentagon, Cyclic Quadrilateral, Eccentric, N-Cluster

References

Guy, R. K. "Lattice Points, No Four on a Circle." §F3 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 241, 1994.

Condition

A requirement Necessary for a given statement or theorem to hold. Also called a Criterion.

see also Boundary Conditions, Carmichael Condition, Cauchy Boundary Conditions, Condition Number, Dirichlet Boundary Conditions, Diversity Condition, Feller-Lévy Condition, Hölder Condition, Lichnerowicz Conditions, Lindeberg Condition, Lipschitz Condition, Lyapunov Condition, Neumann Boundary Conditions, Robertson Condition, Robin Boundary Conditions, Taylor's Condition, Triangle Condition, Weierstraß-Erdman Corner Condition, Winkler Conditions

Condition Number

The ratio of the largest to smallest Singular Value of a system. A system is said to be singular if the condition number is Infinite, and ill-conditioned if it is too large.

Conditional Convergence

If the Series

$$\sum_{n=0}^{\infty} u_n$$

Converges, but

$$\sum_{n=0}^{\infty} |u_n|$$

does not, where $|x|$ is the ABSOLUTE VALUE, then the SERIES is said to be conditionally CONVERGENT.

see also ABSOLUTE CONVERGENCE, CONVERGENCE TESTS, RIEMANN SERIES THEOREM, SERIES

Conditional Probability

The conditional probability of A given that B has occurred, denoted $P(A|B)$, equals

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad (1)$$

which can be proven directly using a VENN DIAGRAM. Multiplying through, this becomes

$$P(A|B)P(B) = P(A \cap B), \qquad (2)$$

which can be generalized to

$$P(A \cup B \cup C) = P(A)P(B|A)P(C|A \cup B). \qquad (3)$$

Rearranging (1) gives

$$P(B|A) = \frac{P(B \cap A)}{P(A)}. \qquad (4)$$

Solving (4) for $P(B \cap A) = P(A \cap B)$ and plugging in to (1) gives

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}. \qquad (5)$$

see also BAYES' FORMULA

Condom Problem

see GLOVE PROBLEM

Condon-Shortley Phase

The $(-1)^m$ phase factor in some definitions of the SPHERICAL HARMONICS and associated LEGENDRE POLYNOMIALS. Using the Condon-Shortley convention gives

$$Y_n^m(\theta, \phi) = (-1)^m \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}}\, P_n^m(\cos\theta) e^{im\phi}.$$

see also LEGENDRE POLYNOMIAL, SPHERICAL HARMONIC

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 682 and 692, 1985.

Condon, E. U. and Shortley, G. *The Theory of Atomic Spectra.* Cambridge, England: Cambridge University Press, 1951.

Shore, B. W. and Menzel, D. H. *Principles of Atomic Spectra.* New York: Wiley, p. 158, 1968.

Conductor

see j-CONDUCTOR

Cone

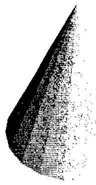

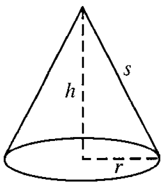

A cone is a PYRAMID with a circular CROSS-SECTION. A right cone is a cone with its vertex above the center of its base. A right cone of height h can be described by the parametric equations

$$x = r(h - z)\cos\theta \qquad (1)$$
$$y = r(h - z)\sin\theta \qquad (2)$$
$$z = z \qquad (3)$$

for $z \in [0, h]$ and $\theta \in [0, 2\pi)$. The VOLUME of a cone is therefore

$$V = \tfrac{1}{3} A_b h, \qquad (4)$$

where A_b is the base AREA and h is the height. If the base is circular, then

$$V = \tfrac{1}{3}\pi r^2 h. \qquad (5)$$

This amazing fact was first discovered by Eudoxus, and other proofs were subsequently found by Archimedes in *On the Sphere and Cylinder* (ca. 225 BC) and Euclid in Proposition XII.10 of his *Elements* (Dunham 1990).

The CENTROID can be obtained by setting $R_2 = 0$ in the equation for the centroid of the CONICAL FRUSTUM,

$$\bar{z} = \frac{\langle z \rangle}{V} = \frac{h({R_1}^2 + R_1 R_2 + {R_2}^2)}{4({R_1}^2 + 2R_1 R_2 + 3{R_2}^2)}, \qquad (6)$$

(Beyer 1987, p. 133) yielding

$$\bar{z} = \tfrac{1}{4} h. \qquad (7)$$

For a right circular cone, the SLANT HEIGHT s is

$$s = \sqrt{r^2 + h^2} \qquad (8)$$

and the surface AREA (not including the base) is

$$S = \pi r s = \pi r \sqrt{r^2 + h^2}. \qquad (9)$$

In discussions of CONIC SECTIONS, the word cone is often used to refer to two similar cones placed apex to apex. This allows the HYPERBOLA to be defined as the intersection of a PLANE with both NAPPES (pieces) of the cone.

The LOCUS of the apex of a variable cone containing an ELLIPSE fixed in 3-space is a HYPERBOLA through the FOCI of the ELLIPSE. In addition, the LOCUS of the apex of a cone containing that HYPERBOLA is the original ELLIPSE. Furthermore, the ECCENTRICITIES of the ELLIPSE and HYPERBOLA are reciprocals.

see also CONIC SECTION, CONICAL FRUSTUM, CYLINDER, NAPPE, PYRAMID, SPHERE

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 129 and 133, 1987.

Dunham, W. *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 76–77, 1990.

Yates, R. C. "Cones." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 34–35, 1952.

Cone Graph

A GRAPH $C_n + \overline{K_m}$, where C_n is a CYCLIC GRAPH and K_m is a COMPLETE GRAPH.

Cone Net

The mapping of a grid of regularly ruled squares onto a CONE with no overlap or misalignment. Cone nets are possible for vertex angles of 90°, 180°, and 270°, and are beautifully illustrated by Steinhaus (1983).

References

Steinhaus, H. *Mathematical Snapshots, 3rd American ed.* New York: Oxford University Press, pp. 224–228, 1983.

Cone (Space)

The JOIN of a TOPOLOGICAL SPACE X and a point P, $C(X) = X * P$.

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 6, 1976.

Cone-Sphere Intersection

Let a CONE of opening parameter c and vertex at $(0,0,0)$ intersect a SPHERE of RADIUS r centered at (x_0, y_0, z_0), with the CONE oriented such that its axis does not pass through the center of the SPHERE. Then the equations of the curve of intersection are

$$\frac{x^2+y^2}{c^2} = z^2 \tag{1}$$

$$(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2 = r^2. \tag{2}$$

Combining (1) and (2) gives

$$(x-x_0)^2+(y-y_0)^2+\frac{x^2+y^2}{c^2}-\frac{2z_0}{c}\sqrt{x^2+y^2}+{z_0}^2 = r^2 \tag{3}$$

$$x^2\left(1+\frac{1}{c^2}\right) - 2x_0x + y^2\left(1+\frac{1}{c^2}\right) - 2y_0y + ({x_0}^2+{y_0}^2+{z_0}^2-r^2) - \frac{2z_0}{c}\sqrt{x^2+y^2} = 0. \tag{4}$$

Therefore, x and y are connected by a complicated QUARTIC EQUATION, and x, y, and z by a QUADRATIC EQUATION.

If the CONE-SPHERE intersection is *on-axis* so that a CONE of opening parameter c and vertex at $(0,0,z_0)$ is oriented with its AXIS along a radial of the SPHERE of radius r centered at $(0,0,0)$, then the equations of the curve of intersection are

$$(z-z_0)^2 = \frac{x^2+y^2}{c^2} \tag{5}$$

$$x^2+y^2+z^2 = r^2. \tag{6}$$

Combining (5) and (6) gives

$$c^2(z-z_0)^2 + z2 = r^2 \tag{7}$$

$$c^2(z^2 - 2z_0z + {z_0}^2) + z^2 = r^2 \tag{8}$$

$$z^2(c^2+1) - 2c^2z_0z + ({z_0}^2c^2 - r^2) = 0. \tag{9}$$

Using the QUADRATIC EQUATION gives

$$\begin{aligned} z &= \frac{2c^2z_0 \pm \sqrt{4c^4{z_0}^2 - 4(c^2+1)({z_0}^2c^2-r^2)}}{2(c^2+1)} \\ &= \frac{c^2z_0 \pm \sqrt{c^2(r^2-{z_0}^2)+r^2}}{c^2+1}. \end{aligned} \tag{10}$$

So the curve of intersection is planar. Plugging (10) into (5) shows that the curve is actually a CIRCLE, with RADIUS given by

$$a = \sqrt{r^2-z^2}. \tag{11}$$

Confidence Interval

The probability that a measurement will fall within a given CLOSED INTERVAL $[a,b]$. For a continuous distribution,

$$\mathrm{CI}(a,b) \equiv \int_b^a P(x)\,dx, \tag{1}$$

where $P(x)$ is the PROBABILITY DISTRIBUTION FUNCTION. Usually, the confidence interval of interest is symmetrically placed around the mean, so

$$\mathrm{CI}(x) \equiv \mathrm{CI}(\mu-x,\mu+x) = \int_{\mu-x}^{\mu+x} P(x)\,dx, \tag{2}$$

where μ is the MEAN. For a GAUSSIAN DISTRIBUTION, the probability that a measurement falls within $n\sigma$ of the mean μ is

$$\begin{aligned} \mathrm{CI}(n\sigma) &\equiv \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu-n\sigma}^{\mu+n\sigma} e^{-(x-\mu)^2/2\sigma^2}\,dx \\ &= \frac{2}{\sigma\sqrt{2\pi}} \int_{0}^{\mu+n\sigma} e^{-(x-\mu)^2/2\sigma^2}\,dx. \end{aligned} \tag{3}$$

Now let $u \equiv (x-\mu)/\sqrt{2}\,\sigma$, so $du = dx/\sqrt{2}\,\sigma$. Then

$$\begin{aligned} \mathrm{CI}(n\sigma) &= \frac{2}{\sigma\sqrt{2\pi}} \sqrt{2}\,\sigma \int_0^{n/\sqrt{2}} e^{-u^2}\,du \\ &= \frac{2}{\sqrt{\pi}} \int_0^{n/\sqrt{2}} e^{-u^2}\,du = \mathrm{erf}\left(\frac{n}{\sqrt{2}}\right), \end{aligned} \tag{4}$$

where $\mathrm{erf}(x)$ is the so-called ERF function. The variate value producing a confidence interval CI is often denoted x_{CI}, so

$$x_{\mathrm{CI}} = \sqrt{2}\,\mathrm{erf}^{-1}(\mathrm{CI}). \tag{5}$$

range	CI
σ	0.6826895
2σ	0.9544997
3σ	0.9973002
4σ	0.9999366
5σ	0.9999994

To find the standard deviation range corresponding to a given confidence interval, solve (4) for n.

$$n = \sqrt{2}\,\mathrm{erf}^{-1}(\mathrm{CI}) \tag{6}$$

CI	range
0.800	$\pm 1.28155\sigma$
0.900	$\pm 1.64485\sigma$
0.950	$\pm 1.95996\sigma$
0.990	$\pm 2.57583\sigma$
0.995	$\pm 2.80703\sigma$
0.999	$\pm 3.29053\sigma$

Configuration

A finite collection of points $p = (p_1, \ldots, p_n)$, $p_i \in \mathbb{R}^d$, where $\mathbb{R}^d$ is a EUCLIDEAN SPACE.

see also BAR (EDGE), EUCLIDEAN SPACE, FRAMEWORK, RIGID

Confluent Hypergeometric Differential Equation

$$xy'' + (b-x)y' - ay = 0, \tag{1}$$

where $y' \equiv dy/dx$ and with boundary conditions

$${}_1F_1(a;b;0) = 1 \tag{2}$$

$$\left[\frac{\partial}{\partial x}{}_1F_1(a;b;x)\right]_{x=0} = \frac{a}{b}. \tag{3}$$

The equation has a REGULAR SINGULAR POINT at 0 and an irregular singularity at ∞. The solutions are called CONFLUENT HYPERGEOMETRIC FUNCTION OF THE FIRST or SECOND KINDS. Solutions of the first kind are denoted ${}_1F_1(a;b;x)$ or $M(a,b,x)$.

see also HYPERGEOMETRIC DIFFERENTIAL EQUATION, WHITTAKER DIFFERENTIAL EQUATION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 504, 1972.

Arfken, G. "Confluent Hypergeometric Functions." §13.6 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 753–758, 1985.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 551–555, 1953.

Confluent Hypergeometric Function

see CONFLUENT HYPERGEOMETRIC FUNCTION OF THE FIRST KIND, CONFLUENT HYPERGEOMETRIC FUNCTION OF THE SECOND KIND

Confluent Hypergeometric Function of the First Kind

The confluent hypergeometric function a degenerate form the HYPERGEOMETRIC FUNCTION ${}_2F_1(a,b;c;z)$ which arises as a solution the the CONFLUENT HYPERGEOMETRIC DIFFERENTIAL EQUATION. It is commonly denoted ${}_1F_1(a;b;z)$, $M(a,b,z)$, or $\Phi(a;b;z)$, and is also known as KUMMER'S FUNCTION of the first kind. An alternate form of the solution to the Confluent Hypergeometric Differential Equation is known as the WHITTAKER FUNCTION.

The confluent hypergeometric function has a HYPERGEOMETRIC SERIES given by

$${}_1F_1(a;b;z) = 1 + \frac{a}{b}z + \frac{a(a+1)}{b(b+1)}\frac{z^2}{2!} + \ldots = \sum_{k=0}^{\infty} \frac{(a)_k}{(b)_k}\frac{z^k}{k!}, \tag{1}$$

where $(a)_k$ and $(b)_k$ are POCHHAMMER SYMBOLS. If a and b are INTEGERS, $a < 0$, and either $b > 0$ or $b < a$, then the series yields a POLYNOMIAL with a finite number of terms. If b is an INTEGER ≤ 0, then ${}_1F_1(a;b;z)$ is undefined. The confluent hypergeometric function also has an integral representation

$${}_1F_1(a;b;z) = \frac{\Gamma(b)}{\Gamma(b-a)\Gamma(a)} \int_0^1 e^{zt} t^{a-1} (1-t)^{b-a-1}\,dt \tag{2}$$

(Abramowitz and Stegun 1972, p. 505).

BESSEL FUNCTIONS, the ERROR FUNCTION, the incomplete GAMMA FUNCTION, HERMITE POLYNOMIAL, LAGUERRE POLYNOMIAL, as well as other are all special cases of this function (Abramowitz and Stegun 1972, p. 509).

KUMMER'S SECOND FORMULA gives

$$ {}_1F_1(\tfrac{1}{2}+m; 2m+1; z) = M_{0,m}(z) = z^{m+1/2} \times \left[1+\sum_{p=1}^{\infty} \frac{z^{2p}}{2^{4p}p!(m+1)(m+2)\cdots(m+p)}\right], \quad (3) $$

where ${}_1F_1$ is the CONFLUENT HYPERGEOMETRIC FUNCTION and $m \neq -1/2, -1, -3/2, \ldots$.

see also CONFLUENT HYPERGEOMETRIC DIFFERENTIAL EQUATION, CONFLUENT HYPERGEOMETRIC FUNCTION OF THE SECOND KIND, CONFLUENT HYPERGEOMETRIC LIMIT FUNCTION, GENERALIZED HYPERGEOMETRIC FUNCTION, HEINE HYPERGEOMETRIC SERIES, HYPERGEOMETRIC FUNCTION, HYPERGEOMETRIC SERIES, KUMMER'S FORMULAS, WEBER-SONINE FORMULA, WHITTAKER FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Confluent Hypergeometric Functions." Ch. 13 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 503–515, 1972.

Arfken, G. "Confluent Hypergeometric Functions." §13.6 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 753–758, 1985.

Iyanaga, S. and Kawada, Y. (Eds.). "Hypergeometric Function of Confluent Type." Appendix A, Table 19.I in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1469, 1980.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 551–554 and 604–605, 1953.

Slater, L. J. *Confluent Hypergeometric Functions.* Cambridge, England: Cambridge University Press, 1960.

Spanier, J. and Oldham, K. B. "The Kummer Function $M(a;c;x)$." Ch. 47 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 459–469, 1987.

Confluent Hypergeometric Function of the Second Kind

Gives the second linearly independent solution to the CONFLUENT HYPERGEOMETRIC DIFFERENTIAL EQUATION. It is also known as the KUMMER'S FUNCTION of the second kind, the TRICOMI FUNCTION, or the GORDON FUNCTION. It is denoted $U(a,b,z)$ and has an integral representation

$$ U(a,b,z) = \frac{1}{\Gamma(a)}\int_0^{\infty} e^{-zt}t^{a-1}(1+t)^{b-a-1}\,dt $$

(Abramowitz and Stegun 1972, p. 505). The WHITTAKER FUNCTIONS give an alternative form of the solution. For small z, the function behaves as z^{1-b}.

see also BATEMAN FUNCTION, CONFLUENT HYPERGEOMETRIC FUNCTION OF THE FIRST KIND, CONFLUENT HYPERGEOMETRIC LIMIT FUNCTION, COULOMB WAVE FUNCTION, CUNNINGHAM FUNCTION, GORDON FUNCTION, HYPERGEOMETRIC FUNCTION, POISSON-CHARLIER POLYNOMIAL, TORONTO FUNCTION, WEBER FUNCTIONS, WHITTAKER FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Confluent Hypergeometric Functions." Ch. 13 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 503–515, 1972.

Arfken, G. "Confluent Hypergeometric Functions." §13.6 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 753–758, 1985.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 671–672, 1953.

Spanier, J. and Oldham, K. B. "The Tricomi Function $U(a;c;x)$." Ch. 48 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 471–477, 1987.

Confluent Hypergeometric Limit Function

$$ {}_0F_1(;a;z) \equiv \lim_{q\to\infty} {}_1F_1\left(q;a;\frac{z}{q}\right). \quad (1) $$

It has a series expansion

$$ {}_0F_1(;a;z) = \sum_{n=0}^{\infty} \frac{z^n}{(a)_n n!} \quad (2) $$

and satisfies

$$ z\frac{d^2y}{dz^2} + a\frac{dy}{dz} - y = 0. \quad (3) $$

A BESSEL FUNCTION OF THE FIRST KIND can be expressed in terms of this function by

$$ J_n(x) = \frac{(\frac{1}{2}x)^n}{n!}\,{}_0F_1(;n+1;-\tfrac{1}{4}x^2) \quad (4) $$

(Petkovšek *et al.* 1996).

see also CONFLUENT HYPERGEOMETRIC FUNCTION, GENERALIZED HYPERGEOMETRIC FUNCTION, HYPERGEOMETRIC FUNCTION

References

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, p. 38, 1996.

Confocal Conics

Confocal conics are CONIC SECTIONS sharing a common FOCUS. Any two confocal CENTRAL CONICS are orthogonal (Ogilvy 1990, p. 77).

see also CONIC SECTION, FOCUS

References

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 77–78, 1990.

Confocal Ellipsoidal Coordinates

The confocal ellipsoidal coordinates (called simply ellipsoidal coordinates by Morse and Feshbach 1953) are given by the equations

$$\frac{x^2}{a^2+\xi}+\frac{y^2}{b^2+\xi}+\frac{z^2}{c^2+\xi}=1 \tag{1}$$

$$\frac{x^2}{a^2+\eta}+\frac{y^2}{b^2+\eta}+\frac{z^2}{c^2+\eta}=1 \tag{2}$$

$$\frac{x^2}{a^2+\zeta}+\frac{y^2}{b^2+\zeta}+\frac{z^2}{c^2+\zeta}=1, \tag{3}$$

where $-c^2<\xi<\infty$, $-b^2<\eta<-c^2$, and $-a^2<\zeta<-b^2$. Surfaces of constant ξ are confocal ELLIPSOIDS, surfaces of constant η are one-sheeted HYPERBOLOIDS, and surfaces of constant ζ are two-sheeted HYPERBOLOIDS. For every (x,y,z), there is a unique set of ellipsoidal coordinates. However, (ξ,η,ζ) specifies eight points symmetrically located in octants. Solving for x, y, and z gives

$$x^2=\frac{(a^2+\xi)(a^2+\eta)(a^2+\zeta)}{(b^2-a^2)(c^2-a^2)} \tag{4}$$

$$y^2=\frac{(b^2+\xi)(b^2+\eta)(b^2+\zeta)}{(a^2-b^2)(c^2-b^2)} \tag{5}$$

$$z^2=\frac{(c^2+\xi)(c^2+\eta)(c^2+\zeta)}{(a^2-c^2)(b^2-c^2)}. \tag{6}$$

The LAPLACIAN is

$$\nabla^2\Psi=(\eta-\zeta)f(\xi)\frac{\partial}{\partial\xi}\left[f(\xi)\frac{\partial\Psi}{\partial\xi}\right]$$
$$+(\zeta-\xi)f(\eta)\frac{\partial}{\partial\eta}\left[f(\eta)\frac{\partial\Psi}{\partial\eta}\right]+(\xi-\eta)f(\zeta)\frac{\partial}{\partial\zeta}\left[f(\zeta)\frac{\partial\Psi}{\partial\zeta}\right], \tag{7}$$

where

$$f(x)\equiv\sqrt{(x+a^2)(x+b^2)(x+c^2)}. \tag{8}$$

Another definition is

$$\frac{x^2}{a^2-\lambda}+\frac{y^2}{b^2-\lambda}+\frac{z^2}{c^2-\lambda}=1 \tag{9}$$

$$\frac{x^2}{a^2-\mu}+\frac{y^2}{b^2-\mu}+\frac{z^2}{c^2-\mu}=1 \tag{10}$$

$$\frac{x^2}{a^2-\nu}+\frac{y^2}{b^2-\nu}+\frac{z^2}{c^2-\nu}=1, \tag{11}$$

where

$$\lambda<c^2<\mu<b^2<\nu<a^2 \tag{12}$$

(Arfken 1970, pp. 117–118). Byerly (1959, p. 251) uses a slightly different definition in which the Greek variables are replaced by their squares, and $a=0$. Equation (9) represents an ELLIPSOID, (10) represents a one-sheeted HYPERBOLOID, and (11) represents a two-sheeted HYPERBOLOID. In terms of CARTESIAN COORDINATES,

$$x^2=\frac{(a^2-\lambda)(a^2-\mu)(a^2-\nu)}{(a^2-b^2)(a^2-c^2)} \tag{13}$$

$$y^2=\frac{(b^2-\lambda)(b^2-\mu)(b^2-\nu)}{(b^2-a^2)(b^2-c^2)} \tag{14}$$

$$z^2=\frac{(c^2-\lambda)(c^2-\mu)(c^2-\nu)}{(c^2-a^2)(c^2-b^2)}. \tag{15}$$

The SCALE FACTORS are

$$h_\lambda=\sqrt{\frac{(\mu-\lambda)(\nu-\lambda)}{4(a^2-\lambda)(b^2-\lambda)(c^2-\lambda)}} \tag{16}$$

$$h_\mu=\sqrt{\frac{(\nu-\mu)(\lambda-\mu)}{4(a^2-\mu)(b^2-\mu)(c^2-\mu)}} \tag{17}$$

$$h_\nu=\sqrt{\frac{(\lambda-\nu)(\mu-\nu)}{4(a^2-\nu)(b^2-\nu)(c^2-\nu)}}. \tag{18}$$

The LAPLACIAN is

$$\begin{aligned}\nabla^2 &= 2\frac{a^2b^2+a^2c^2+b^2c^2-2\nu(a^2+b^2+c^2)+3\nu^2}{(\mu-\nu)(\nu-\lambda)}\frac{\partial}{\partial\nu}\\ &+\frac{4(a^2-\nu)(b^2-\nu)(c^2-\nu)}{(\mu-\nu)(\nu-\lambda)}\frac{\partial^2}{\partial\nu^2}\\ &+2\frac{a^2b^2+a^2c^2+b^2c^2-2\mu(a^2+b^2+c^2)+3\mu^2}{(\nu-\mu)(\mu-\lambda)}\frac{\partial}{\partial\mu}\\ &+\frac{4(a^2-\mu)(b^2-\mu)(c^2-\mu)}{(\mu-\lambda)(\nu-\mu)}\frac{\partial^2}{\partial\mu^2}\\ &+2\frac{-(a^2b^2+a^2c^2+b^2c^2)+2\lambda(a^2+b^2+c^2)-3\lambda^2}{(\mu-\lambda)(\nu-\lambda)}\frac{\partial}{\partial\lambda}\\ &+\frac{4(a^2-\lambda)(b^2-\lambda)(c^2-\lambda)}{(\mu-\lambda)(\nu-\lambda)}\frac{\partial^2}{\partial\lambda^2}.\end{aligned} \tag{19}$$

Using the NOTATION of Byerly (1959, pp. 252–253), this can be reduced to

$$\nabla^2=(\mu^2-\nu^2)\frac{\partial^2}{\partial\alpha^2}+(\lambda^2-\nu^2)\frac{\partial^2}{\partial\beta^2}+(\lambda^2-\mu^2)\frac{\partial^2}{\partial\gamma^2}, \tag{20}$$

where

$$\begin{aligned}\alpha &= c\int_c^\lambda\frac{d\lambda}{\sqrt{(\lambda^2-b^2)(\lambda^2-c^2)}}\\ &= F\left(\frac{b}{c},\frac{\pi}{2}\right)-F\left(\frac{b}{c},\sin^{-1}\left(\frac{c}{\lambda}\right)\right)\end{aligned} \tag{21}$$

$$\begin{aligned}\beta &= c\int_b^\mu\frac{d\mu}{\sqrt{(c^2-\mu^2)(\mu^2-b^2)}}\\ &= F\left[\sqrt{1-\frac{b^2}{c^2}},\sin^{-1}\left(\sqrt{\frac{1-\frac{b^2}{\mu^2}}{1-\frac{b^2}{c^2}}}\right)\right]\end{aligned} \tag{22}$$

$$\begin{aligned}\gamma &= c\int_0^\nu\frac{d\nu}{\sqrt{(b^2-\nu^2)(c^2-\nu^2)}}\\ &= F\left(\frac{b}{c},\sin^{-1}\left(\frac{\nu}{b}\right)\right).\end{aligned} \tag{23}$$

Here, F is an ELLIPTIC INTEGRAL OF THE FIRST KIND. In terms of α, β, and γ,

$$\lambda = c\,\mathrm{dc}\left(\alpha, \frac{b}{c}\right) \tag{24}$$

$$\mu = b\,\mathrm{nd}\left(\beta, \sqrt{1-\frac{b^2}{c^2}}\right) \tag{25}$$

$$\nu = b\,\mathrm{sn}\left(\gamma, \frac{b}{c}\right), \tag{26}$$

where dc, nd and sn are JACOBI ELLIPTIC FUNCTIONS. The HELMHOLTZ DIFFERENTIAL EQUATION is separable in confocal ellipsoidal coordinates.

see also HELMHOLTZ DIFFERENTIAL EQUATION—CONFOCAL ELLIPSOIDAL COORDINATES

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Definition of Elliptical Coordinates." §21.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 752, 1972.

Arfken, G. "Confocal Ellipsoidal Coordinates (ξ_1, ξ_2, ξ_3)." §2.15 in *Mathematical Methods for Physicists, 2nd ed.* New York: Academic Press, pp. 117–118, 1970.

Byerly, W. E. *An Elementary Treatise on Fourier's Series, and Spherical, Cylindrical, and Ellipsoidal Harmonics, with Applications to Problems in Mathematical Physics.* New York: Dover, 1959.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 663, 1953.

Confocal Parabolic Coordinates

see CONFOCAL PARABOLOIDAL COORDINATES

Confocal Paraboloidal Coordinates

$$\frac{x^2}{a^2-\lambda} + \frac{y^2}{b^2-\lambda} = z - \lambda \tag{1}$$

$$\frac{x^2}{a^2-\mu} + \frac{y^2}{b^2-\mu} = z - \mu \tag{2}$$

$$\frac{x^2}{a^2-\nu} + \frac{y^2}{b^2-\nu} = z - \nu, \tag{3}$$

where $\lambda \in (-\infty, b^2)$, $\mu \in (b^2, a^2)$, and $\nu \in (a^2, \infty)$.

$$x^2 = \frac{(a^2-\lambda)(a^2-\mu)(a^2-\nu)}{(b^2-a^2)} \tag{4}$$

$$y^2 = \frac{(b^2-\lambda)(b^2-\mu)(b^2-\nu)}{(a^2-b^2)} \tag{5}$$

$$z = \lambda + \mu + \nu - a^2 - b^2. \tag{6}$$

The SCALE FACTORS are

$$h_\lambda = \sqrt{\frac{(\mu-\lambda)(\nu-\lambda)}{4(a^2-\lambda)(b^2-\lambda)}} \tag{7}$$

$$h_\mu = \sqrt{\frac{(\nu-\mu)(\lambda-\mu)}{4(a^2-\mu)(b^2-\mu)}} \tag{8}$$

$$h_\nu = \sqrt{\frac{(\lambda-\nu)(\mu-\nu)}{16(a^2-\nu)(b^2-\nu)}}. \tag{9}$$

The LAPLACIAN is

$$\begin{aligned}\nabla^2 = {} & \frac{2(a^2+b^2-2\nu)}{(\mu-\nu)(\nu-\lambda)}\frac{\partial}{\partial\nu} + \frac{4(a^2-\nu)(\nu-b^2)}{(\mu-\nu)(\nu-\lambda)}\frac{\partial^2}{\nu^2} \\ & + \frac{2(a^2+b^2-2\mu)}{(\mu-\lambda)(\nu-\mu)}\frac{\partial}{\partial\mu} + \frac{4(a^2-\mu)(\mu-b^2)}{(\mu-\lambda)(\nu-\mu)}\frac{\partial^2}{\partial\mu^2} \\ & + \frac{2(2\lambda-a^2-b^2)}{(\mu-\lambda)(\nu-\lambda)}\frac{\partial}{\partial\lambda} + \frac{4(\lambda-a^2)(\lambda-b^2)}{(\mu-\lambda)(\nu-\lambda)}\frac{\partial^2}{\partial\lambda^2}.\end{aligned} \tag{10}$$

The HELMHOLTZ DIFFERENTIAL EQUATION is SEPARABLE.

see also HELMHOLTZ DIFFERENTIAL EQUATION—CONFOCAL PARABOLOIDAL COORDINATES

References

Arfken, G. "Confocal Parabolic Coordinates (ξ_1, ξ_2, ξ_3)." §2.17 in *Mathematical Methods for Physicists, 2nd ed.* Orlando, FL: Academic Press, pp. 119–120, 1970.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 664, 1953.

Conformal Latitude

An AUXILIARY LATITUDE defined by

$$\begin{aligned}\chi &\equiv 2\tan^{-1}\left\{\tan(\tfrac{1}{4}\pi+\tfrac{1}{2}\phi)\left[\frac{1-e\sin\phi}{1+e\sin\phi}\right]^{e/2}\right\} - \tfrac{1}{2}\pi \\ &= 2\tan^{-1}\left[\frac{1+\sin\phi}{1-\sin\phi}\left(\frac{1-e\sin\phi}{1+e\sin\phi}\right)^{e}\right]^{1/2} - \tfrac{1}{2}\pi \\ &= \phi - (\tfrac{1}{2}e^2 + \tfrac{5}{24}e^4 + \tfrac{3}{32}e^6 + \tfrac{281}{5760}e^8 + \ldots)\sin(2\phi) \\ &\quad + (\tfrac{5}{48}e^4 + \tfrac{7}{80}e^6 + \tfrac{697}{11520}e^8 + \ldots)\sin(4\phi) \\ &\quad - (\tfrac{13}{480}e^6 + \tfrac{461}{13440} + \ldots)\sin(6\phi) \\ &\quad + (\tfrac{1237}{161280}e^8 + \ldots)\sin(8\phi) + \ldots.\end{aligned}$$

The inverse is obtained by iterating the equation

$$\phi = 2\tan^{-1}\left[\tan(\tfrac{1}{4}\pi+\tfrac{1}{2}\chi)\left(\frac{1+e\sin\phi}{1-e\sin\phi}\right)^{e/2}\right] - \tfrac{1}{2}\pi$$

using $\phi = \chi$ as the first trial. A series form is

$$\begin{aligned}\phi = \chi &+ (\tfrac{1}{2}e^2 + \tfrac{5}{24}e^4 + \tfrac{1}{12}e^6 + \tfrac{13}{360}e^8 + \ldots)\sin(2\chi) \\ &+ (\tfrac{7}{48}e^4 + \tfrac{29}{240}e^6 + \tfrac{811}{11520}e^9 + \ldots)\sin(4\chi) \\ &+ (\tfrac{7}{120}e^6 + \tfrac{81}{1120}e^8 + \ldots)\sin(6\chi) \\ &+ (\tfrac{4279}{161280}e^8 + \ldots)\sin(8\chi) + \ldots\end{aligned}$$

The conformal latitude was called the ISOMETRIC LATITUDE by Adams (1921), but this term is now used to refer to a different quantity.

see also AUXILIARY LATITUDE, LATITUDE

References

Adams, O. S. "Latitude Developments Connected with Geodesy and Cartography with Tables, Including a Table for Lambert Equal-Area Meridianal Projections." Spec. Pub. No. 67. U. S. Coast and Geodetic Survey, pp. 18 and 84–85, 1921.

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 15–16, 1987.

Conformal Map

A TRANSFORMATION which preserves ANGLES is known as conformal. For a transformation to be conformal, it must be an ANALYTIC FUNCTION and have a NONZERO DERIVATIVE. Let θ and ϕ be the tangents to the curves γ and $f(\gamma)$ at z_0 and w_0,

$$w - w_0 \equiv f(z) - f(z_0) = \frac{f(z) - f(z_0)}{z - z_0}(z - z_0) \quad (1)$$

$$\arg(w - w_0) = \arg\left[\frac{f(z) - f(z_0)}{z - z_0}\right] + \arg(z - z_0). \quad (2)$$

Then as $w \to w_0$ and $z \to z_0$,

$$\phi = \arg f'(z_0) + \theta \quad (3)$$

$$|w| = |f'(z_0)|\,|z|. \quad (4)$$

see also ANALYTIC FUNCTION, HARMONIC FUNCTION, MÖBIUS TRANSFORMATION, QUASICONFORMAL MAP, SIMILAR

References

Arfken, G. "Conformal Mapping." §6.7 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 392–394, 1985.

Bergman, S. *The Kernel Function and Conformal Mapping.* New York: Amer. Math. Soc., 1950.

Katznelson, Y. *An Introduction to Harmonic Analysis.* New York: Dover, 1976.

Morse, P. M. and Feshbach, H. "Conformal Mapping." §4.7 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 358–362 and 443–453, 1953.

Nehari, Z. *Conformal Map.* New York: Dover, 1982.

Conformal Solution

By letting $w \equiv f(z)$, the REAL and IMAGINARY PARTS of w must satisfy the CAUCHY-RIEMANN EQUATIONS and LAPLACE'S EQUATION, so they automatically provide a scalar POTENTIAL and a so-called stream function. If a physical problem can be found for which the solution is valid, we obtain a solution—which may have been very difficult to obtain directly—by working backwards. Let

$$Az^n = Ar^n e^{in\theta}, \quad (1)$$

the REAL and IMAGINARY PARTS then give

$$\phi = Ar^n \cos(n\theta) \quad (2)$$

$$\psi = Ar^n \sin(n\theta). \quad (3)$$

For $n = -2$,

$$\phi = \frac{A}{r^2}\cos(2\theta) \quad (4)$$

$$\psi = -\frac{A}{r^2}\sin(2\theta), \quad (5)$$

which is a double system of LEMNISCATES (Lamb 1945, p. 69). For $n = -1$,

$$\phi = \frac{A}{r}\cos\theta \quad (6)$$

$$\psi = -\frac{A}{r}\sin\theta. \quad (7)$$

This solution consists of two systems of CIRCLES, and ϕ is the POTENTIAL FUNCTION for two PARALLEL opposite charged line charges (Feynman *et al.* 1989, §7-5; Lamb 1945, p. 69). For $n = 1/2$,

$$\phi = Ar^{1/2}\cos\left(\frac{\theta}{2}\right) = A\sqrt{\frac{\sqrt{x^2+y^2}+x}{2}} \quad (8)$$

$$\psi = Ar^{1/2}\sin\left(\frac{\theta}{2}\right) = A\sqrt{\frac{\sqrt{x^2+y^2}-x}{2}}. \quad (9)$$

ϕ gives the field near the edge of a thin plate (Feynman *et al.* 1989, §7-5). For $n = 1$,

$$\phi = Ar\cos\theta = Ax \quad (10)$$

$$\psi = Ar\sin\theta = Ay. \quad (11)$$

This is two straight lines (Lamb 1945, p. 68). For $n = 3/2$,

$$w = Ar^{3/2}e^{3i\theta/2}. \quad (12)$$

ϕ gives the field near the outside of a rectangular corner (Feynman *et al.* 1989, §7-5). For $n = 2$,

$$w = A(x+iy)^2 = A[(x^2-y^2)+2ixy] \quad (13)$$

$$\phi = A(x^2-y^2) = Ar^2\cos(2\theta) \quad (14)$$

$$\psi = 2Axy = Ar^2\sin(2\theta). \quad (15)$$

These are two PERPENDICULAR HYPERBOLAS, and ϕ is the POTENTIAL FUNCTION near the middle of two point charges or the field on the opening side of a charged RIGHT ANGLE conductor (Feynman 1989, §7-3).

see also CAUCHY-RIEMANN EQUATIONS, CONFORMAL MAP, LAPLACE'S EQUATION

References

Feynman, R. P.; Leighton, R. B.; and Sands, M. *The Feynman Lectures on Physics, Vol. 1.* Redwood City, CA: Addison-Wesley, 1989.

Lamb, H. *Hydrodynamics, 6th ed.* New York: Dover, 1945.

Conformal Tensor

see WEYL TENSOR

Conformal Transformation

see CONFORMAL MAP

Congruence

If $b - c$ is integrally divisible by a, then b and c are said to be congruent with MODULUS a. This is written mathematically as $b \equiv c \pmod{a}$. If $b - c$ is *not* divisible by a, then we say $b \not\equiv c \pmod{a}$. The $(\mathrm{mod}\ a)$ is sometimes omitted when the MODULUS a is understood for a given computation, so care must be taken not to confuse the symbol $\equiv$ with that for an EQUIVALENCE. The quantity b is called the RESIDUE or REMAINDER. The COMMON RESIDUE is taken to be NONNEGATIVE and smaller than m, and the MINIMAL RESIDUE is b or $b - m$, whichever is smaller in ABSOLUTE VALUE. In many computer languages (such as **FORTRAN** or *Mathematica*®), the COMMON RESIDUE of $c \pmod{a}$ is written `mod(c,a)`.

Congruence arithmetic is perhaps most familiar as a generalization of the arithmetic of the clock: 40 minutes past the hour plus 35 minutes gives $40 + 35 \equiv 15 \pmod{60}$, or 15 minutes past the hour, and 10 o'clock a.m. plus five hours gives $10 + 5 \equiv 3 \pmod{12}$, or 3 o'clock p.m. Congruences satisfy a number of important properties, and are extremely useful in many areas of NUMBER THEORY. Using congruences, simple DIVISIBILITY TESTS to check whether a given number is divisible by another number can sometimes be derived. For example, if the sum of a number's digits is divisible by 3 (9), then the original number is divisible by 3 (9).

Congruences also have their limitations. For example, if $a \equiv b$ and $c \equiv d \pmod{n}$, then it follows that $a^x \equiv b^x$, but usually not that $x^c \equiv x^d$ or $a^c \equiv b^d$. In addition, by "rolling over," congruences discard absolute information. For example, knowing the number of minutes past the hour is useful, but knowing the hour the minutes are past is often more useful still.

Let $a \equiv a' \pmod{m}$ and $b \equiv b' \pmod{m}$, then important properties of congruences include the following, where $\Rightarrow$ means "IMPLIES":

1. Equivalence: $a \equiv b \pmod{0} \Rightarrow a = b$.
2. Determination: either $a \equiv b \pmod{m}$ or $a \not\equiv b$ (mom m).
3. Reflexivity: $a \equiv a \pmod{m}$.
4. Symmetry: $a \equiv b \pmod{m} \Rightarrow b \equiv a \pmod{m}$.
5. Transitivity: $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m} \Rightarrow a \equiv c \pmod{m}$.
6. $a + b \equiv a' + b' \pmod{m}$.
7. $a - b \equiv a' - b' \pmod{m}$.
8. $ab \equiv a'b' \pmod{m}$.
9. $a \equiv b \pmod{m} \Rightarrow ka \equiv kb \pmod{m}$.
10. $a \equiv b \pmod{m} \Rightarrow a^n \equiv b^n \pmod{m}$.
11. $a \equiv b \pmod{m_1}$ and $a \equiv b \pmod{m_2} \Rightarrow a \equiv b \pmod{[m_1, m_2]}$, where $[m_1, m_2]$ is the LEAST COMMON MULTIPLE.
12. $ak \equiv bk \pmod{m} \Rightarrow a \equiv b \left(\mathrm{mod}\ \frac{m}{(k,m)}\right)$, where (k, m) is the GREATEST COMMON DIVISOR.
13. If $a \equiv b \pmod{m}$, then $P(a) \equiv P(b) \pmod{m}$, for $P(x)$ a POLYNOMIAL.

Properties (6–8) can be proved simply by defining

$$a \equiv a' + rd \tag{1}$$

$$b \equiv b' + sd, \tag{2}$$

where r and s are INTEGERS. Then

$$a + b = a' + b' + (r + s)d \tag{3}$$

$$a - b = a' - b' + (r - s)d \tag{4}$$

$$ab = a'b' + (a's + b'r + rsd)d, \tag{5}$$

so the properties are true.

Congruences also apply to FRACTIONS. For example, note that (mod 7)

$$2 \times 4 \equiv 1 \qquad 3 \times 3 \equiv 2 \qquad 6 \times 6 \equiv 1 \pmod{7}, \tag{6}$$

so

$$\tfrac{1}{2} \equiv 4 \qquad \tfrac{1}{4} \equiv 2 \qquad \tfrac{2}{3} \equiv 3 \qquad \tfrac{1}{6} \equiv 6 \pmod{7}. \tag{7}$$

To find $p/q \bmod m$, use an ALGORITHM similar to the GREEDY ALGORITHM. Let $q_0 \equiv q$ and find

$$p_0 = \left\lceil \frac{m}{q_0} \right\rceil, \tag{8}$$

where $\lceil x \rceil$ is the CEILING FUNCTION, then compute

$$q_1 \equiv q_0 p_0 \pmod{m}. \tag{9}$$

Iterate until $q_n = 1$, then

$$\frac{p}{q} \equiv p \prod_{i=0}^{n-1} p_i \pmod{m}. \tag{10}$$

This method always works for m PRIME, and sometimes even for m COMPOSITE. However, for a COMPOSITE m, the method can fail by reaching 0 (Conway and Guy 1996).

A LINEAR CONGRUENCE

$$ax \equiv b \pmod{m} \tag{11}$$

is solvable IFF the congruence

$$b \equiv 0 \pmod{(a, m)} \tag{12}$$

is solvable, where $d \equiv (a, m)$ is the GREATEST COMMON DIVISOR, in which case the solutions are x_0, $x_0 + m/d$, $x_0 + 2m/d$, ..., $x_0 + (d - 1)m/d$, where $x_0 < m/d$. If $d = 1$, then there is only one solution.

A general QUADRATIC CONGRUENCE

$$a_2x^2 + a_1x + a_0 \equiv 0 \pmod{n} \tag{13}$$

can be reduced to the congruence

$$x^2 \equiv q \pmod{p} \tag{14}$$

and can be solved using EXCLUDENTS. Solution of the general polynomial congruence

$$a_mx^m + \ldots + a_2x^2 + a_1x + a_0 \equiv 0 \pmod{n} \tag{15}$$

is intractable. Any polynomial congruence will give congruent results when congruent values are substituted.

Two simultaneous congruences

$$x \equiv a \pmod{m} \tag{16}$$

$$x \equiv b \pmod{n} \tag{17}$$

are solvable only when $x \equiv b \pmod{(m,n)}$, and the single solution is

$$x \equiv x_0 \pmod{[m,n]}, \tag{18}$$

where $x_0 < m/d$.

see also CANCELLATION LAW, CHINESE REMAINDER THEOREM, COMMON RESIDUE, CONGRUENCE AXIOMS, DIVISIBILITY TESTS, GREATEST COMMON DIVISOR, LEAST COMMON MULTIPLE, MINIMAL RESIDUE, MODULUS (CONGRUENCE), QUADRATIC RECIPROCITY LAW, RESIDUE (CONGRUENCE)

References

Conway, J. H. and Guy, R. K. "Arithmetic Modulo *p*." In *The Book of Numbers.* New York: Springer-Verlag, pp. 130–132, 1996.

Courant, R. and Robbins, H. "Congruences." §2 in Supplement to Ch. 1 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 31–40, 1996.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 55, 1993.

Weisstein, E. W. "Fractional Congruences." http://www.astro.virginia.edu/~eww6n/math/notebooks/ModFraction.m.

Congruence Axioms

The five of HILBERT'S AXIOMS which concern geometric equivalence.

see also CONGRUENCE AXIOMS, CONTINUITY AXIOMS, HILBERT'S AXIOMS, INCIDENCE AXIOMS, ORDERING AXIOMS, PARALLEL POSTULATE

References

Hilbert, D. *The Foundations of Geometry, 2nd ed.* Chicago, IL: Open Court, 1980.

Iyanaga, S. and Kawada, Y. (Eds.). "Hilbert's System of Axioms." §163B in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 544–545, 1980.

Congruence (Geometric)

Two geometric figures are said to be congruent if they are equivalent to within a ROTATION. This relationship is written $A \cong B$. (Unfortunately, this symbol is also used to denote ISOMORPHIC GROUPS.)

see also SIMILAR

Congruence Transformation

A transformation of the form $g = \mathsf{D}^{\mathrm{T}}\eta\mathsf{D}$, where $\det(\mathsf{D}) \neq 0$ and $\det(\mathsf{D})$ is the DETERMINANT.

see also SYLVESTER'S INERTIA LAW

Congruent

A number a is said to be congruent to b modulo m if $m|a-b$ (m DIVIDES $a-b$).

Congruent Incircles Point

The point Y for which TRIANGLES BYC, CYA, and AYB have congruent INCIRCLES. It is a special case of an ELKIES POINT.

References

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Congruent Isoscelizers Point

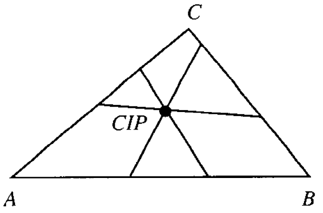

In 1989, P. Yff proved there is a unique configuration of ISOSCELIZERS for a given TRIANGLE such that all three have the same length. Furthermore, these ISOSCELIZERS meet in a point called the congruent isoscelizers point, which has TRIANGLE CENTER FUNCTION

$$\alpha = \cos(\tfrac{1}{2}B) + \cos(\tfrac{1}{2}C) - \cos(\tfrac{1}{2}A).$$

see also CONGRUENT ISOSCELIZERS POINT, ISOSCELIZER

References

Kimberling, C. "Congruent Isoscelizers Point." http://www.evansville.edu/~ck6/tcenters/recent/conisos.html.

Congruent Numbers

A set of numbers (a, x, y, t) such that

$$\begin{cases} x^2 + ay^2 = z^2 \\ x^2 - ay^2 = t^2. \end{cases}$$

They are a generalization of the CONGRUUM PROBLEM, which is the case $y = 1$. For $a = 101$, the smallest solution is

$$\begin{aligned} x &= 2015242462949760001961 \\ y &= 1181714318527794519 00 \\ z &= 2339148435306225006961 \\ t &= 1628124370727269996961. \end{aligned}$$

see also CONGRUUM

References
Guy, R. K. "Congruent Number." §D76 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 195–197, 1994.

Congruum

A number h which satisfies the conditions of the CONGRUUM PROBLEM:

$$x^2 + h = a^2$$

and

$$x^2 - h = b^2.$$

see also CONCORDANT FORM, CONGRUUM PROBLEM

Congruum Problem

Find a SQUARE NUMBER x^2 such that, when a given number h is added or subtracted, new SQUARE NUMBERS are obtained so that

$$x^2 + h = a^2 \tag{1}$$

and

$$x^2 - h = b^2. \tag{2}$$

This problem was posed by the mathematicians Théodore and Jean de Palerma in a mathematical tournament organized by Frederick II in Pisa in 1225. The solution (Ore 1988, pp. 188–191) is

$$x = m^2 + n^2 \tag{3}$$

$$h = 4mn(m^2 - n^2), \tag{4}$$

where m and n are INTEGERS. Fibonacci proved that all numbers h (the CONGRUA) are divisible by 24. FERMAT'S RIGHT TRIANGLE THEOREM is equivalent to the result that a congruum cannot be a SQUARE NUMBER. A table for small m and n is given in Ore (1988, p. 191), and a larger one (for $h \leq 1000$) by Lagrange (1977).

m	n	h	x
2	1	24	5
3	1	96	10
3	2	120	13
4	1	240	17
4	3	336	25

see also CONCORDANT FORM, CONGRUENT NUMBERS, SQUARE NUMBER

References
Alter, R. and Curtz, T. B. "A Note on Congruent Numbers." *Math. Comput.* **28**, 303–305, 1974.
Alter, R.; Curtz, T. B.; and Kubota, K. K. "Remarks and Results on Congruent Numbers." In *Proc. Third Southeastern Conference on Combinatorics, Graph Theory, and Computing, 1972, Boca Raton, FL.* Boca Raton, FL: Florida Atlantic University, pp. 27–35, 1972.
Bastien, L. "Nombres congruents." *Interméd. des Math.* **22**, 231–232, 1915.
Gérardin, A. "Nombres congruents." *Interméd. des Math.* **22**, 52–53, 1915.
Lagrange, J. "Construction d'une table de nombres congruents." *Calculateurs en Math., Bull. Soc. math. France.*, Mémoire 49–50, 125–130, 1977.
Ore, Ø. *Number Theory and Its History.* New York: Dover, 1988.

Conic

see CONIC SECTION

Conic Constant

$$K \equiv -e^2,$$

where e is the ECCENTRICITY of a CONIC SECTION.

see also CONIC SECTION, ECCENTRICITY

Conic Double Point

see ISOLATED SINGULARITY

Conic Equidistant Projection

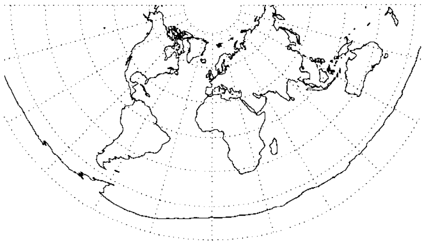

A MAP PROJECTION with transformation equations

$$x = \rho \sin\theta \tag{1}$$

$$y = \rho_0 - \rho\cos\theta, \tag{2}$$

where

$$\rho = (G - \phi) \tag{3}$$

$$\theta = n(\lambda - \lambda_0) \tag{4}$$

$$\rho_0 = (G - \theta_0) \tag{5}$$

$$G = \frac{\cos\phi_1}{n} + \phi_1 \tag{6}$$

$$n = \frac{\cos\phi_1 - \cos\phi_2}{\phi_2 - \phi_1}. \tag{7}$$

The inverse FORMULAS are given by

$$\phi = G - \rho \tag{8}$$

$$\lambda = \lambda_0 + \frac{\theta}{n}, \tag{9}$$

where

$$\rho = \operatorname{sgn}(n)\sqrt{x^2 + (\rho_0 - y)^2} \tag{10}$$

$$\theta = \tan^{-1}\left(\frac{x}{\rho_0 - y}\right). \tag{11}$$

Conic Projection

see ALBERS EQUAL-AREA CONIC PROJECTION, CONIC EQUIDISTANT PROJECTION, LAMBERT AZIMUTHAL EQUAL-AREA PROJECTION, POLYCONIC PROJECTION

Conic Section

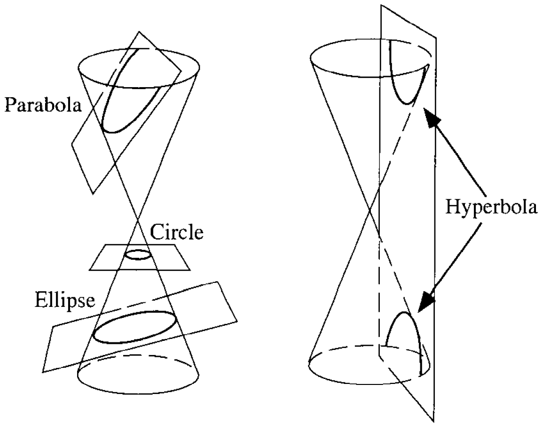

The conic sections are the nondegenerate curves generated by the intersections of a PLANE with one or two NAPPES of a CONE. For a PLANE parallel to a CROSS-SECTION, a CIRCLE is produced. The closed curve produced by the intersection of a single NAPPE with an inclined PLANE is an ELLIPSE or PARABOLA. The curve produced by a PLANE intersecting both NAPPES is a HYPERBOLA. The ELLIPSE and HYPERBOLA are known as CENTRAL CONICS.

Because of this simple geometric interpretation, the conic sections were studied by the Greeks long before their application to inverse square law orbits was known. Apollonius wrote the classic ancient work on the subject entitled *On Conics.* Kepler was the first to notice that planetary orbits were ELLIPSES, and Newton was then able to derive the shape of orbits mathematically using CALCULUS, under the assumption that gravitational force goes as the inverse square of distance. Depending on the energy of the orbiting body, orbit shapes which are any of the four types of conic sections are possible.

A conic section may more formally be defined as the locus of a point P that moves in the PLANE of a fixed point F called the FOCUS and a fixed line d called the DIRECTRIX (with F not on d) such that the ratio of the distance of P from F to its distance from d is a constant e called the ECCENTRICITY. For a FOCUS $(0,0)$ and DIRECTRIX $x = -a$, the equation is

$$x^2 + y^2 = e^2(x+a)^2.$$

If $e = 1$, the conic is a PARABOLA, if $e < 1$, the conic is an ELLIPSE, and if $e > 1$, it is a HYPERBOLA.

In standard form, a conic section is written

$$y^2 = 2Rx - (1 - e^2)x^2,$$

where R is the RADIUS OF CURVATURE and e is the ECCENTRICITY. Five points in a plane determine a conic (Le Lionnais 1983, p. 56).

see also BRIANCHON'S THEOREM, CENTRAL CONIC, CIRCLE, CONE, ECCENTRICITY, ELLIPSE, FERMAT CONIC, HYPERBOLA, NAPPE, PARABOLA, PASCAL'S THEOREM, QUADRATIC CURVE, SEYDEWITZ'S THEOREM, SKEW CONIC, STEINER'S THEOREM

References

Besant, W. H. *Conic Sections, Treated Geometrically, 8th ed. rev.* Cambridge, England: Deighton, Bell, 1890.

Casey, J. "Special Relations of Conic Sections" and "Invariant Theory of Conics." Chs. 9 and 15 in *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions, with Numerous Examples, 2nd ed., rev. enl.* Dublin: Hodges, Figgis, & Co., pp. 307–332 and 462–545, 1893.

Coolidge, J. L. *A History of the Conic Sections and Quadric Surfaces.* New York: Dover, 1968.

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 138–141, 1967.

Downs, J. W. *Conic Sections.* Dale Seymour Pub., 1993.

Iyanaga, S. and Kawada, Y. (Eds.). "Conic Sections." §80 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 271–276, 1980.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 56, 1983.

Lee, X. "Conic Sections." `http://www.best.com/~xah/SpecialPlaneCurves_dir/ConicSections_dir/conicSections.html`.

Ogilvy, C. S. "The Conic Sections." Ch. 6 in *Excursions in Geometry.* New York: Dover, pp. 73–85, 1990.

Pappas, T. "Conic Sections." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 196–197, 1989.

Salmon, G. *Conic Sections, 6th ed.* New York: Chelsea, 1954.

Smith, C. *Geometric Conics.* London: MacMillan, 1894.

Sommerville, D. M. Y. *Analytical Conics, 3rd ed.* London: G. Bell and Sons, 1961.

Yates, R. C. "Conics." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 36–56, 1952.

Conic Section Tangent

Given a CONIC SECTION

$$x^2 + y^2 + 2gx + 2fy + c = 0,$$

the tangent at (x_1, y_1) is given by the equation

$$xx_1 + yy_1 + g(x + x_1) + f(y + y_1) + c = 0.$$

Conical Coordinates

Arfken (1970) and Morse and Feshbach (1953) use slightly different definitions of these coordinates. The system used in *Mathematica*® (Wolfram Research, Inc., Champaign, Illinois) is

$$x = \frac{\lambda\mu\nu}{ab} \tag{1}$$

$$y = \frac{\lambda}{a}\sqrt{\frac{(\mu^2 - a^2)(\nu^2 - a^2)}{a^2 - b^2}} \tag{2}$$

$$z = \frac{\lambda}{b}\sqrt{\frac{(\mu^2 - b^2)(\nu^2 - b^2)}{b^2 - a^2}}, \tag{3}$$

where $b^2 > \mu^2 > c^2 > \nu^2$. The NOTATION of Byerly replaces λ with r, and a and b with b and c. The above equations give

$$x^2 + y^2 + z^2 = \lambda^2 \tag{4}$$

$$\frac{x^2}{\mu^2} + \frac{y^2}{\mu^2 - a^2} + \frac{z^2}{\mu^2 - b^2} = 0 \tag{5}$$

$$\frac{x^2}{\nu^2} + \frac{y^2}{\nu^2 - a^2} + \frac{z^2}{\nu^2 - b^2} = 0. \tag{6}$$

The SCALE FACTORS are

$$h_\lambda = 1 \tag{7}$$

$$h_\mu = \sqrt{\frac{\lambda^2(\mu^2 - \nu^2)}{(\mu^2 - a^2)(b^2 - \mu^2)}} \tag{8}$$

$$h_\nu = \sqrt{\frac{\lambda^2(\mu^2 - \nu^2)}{(\nu^2 - a^2)(\nu^2 - b^2)}}. \tag{9}$$

The LAPLACIAN is

$$\begin{aligned}\nabla^2 = {} & \frac{\nu(2\nu^2 - a^2 - b^2)}{(\mu - \nu)(\mu + \nu)\lambda^2}\frac{\partial}{\partial\nu} \\ & + \frac{(a - \nu)(a + \nu)(\nu - b)(\nu + b)}{(\nu - \mu)(\nu + \mu)\lambda^2}\frac{\partial^2}{\partial\nu^2} \\ & + \frac{\mu(2\mu^2 - a^2 - b^2)}{(\nu - \mu)(\nu + \mu)\lambda^2}\frac{\partial}{\partial\mu} \\ & + \frac{(\mu - b)(\mu + b)(\mu - a)(\mu + a)}{(\nu - \mu)(\nu + \mu)\lambda^2}\frac{\partial^2}{\partial\mu^2} \\ & + \frac{2}{\lambda}\frac{\partial}{\partial\lambda} + \frac{\partial^2}{\partial\lambda^2}.\end{aligned} \tag{10}$$

The HELMHOLTZ DIFFERENTIAL EQUATION is separable in conical coordinates.

see also HELMHOLTZ DIFFERENTIAL EQUATION—CONICAL COORDINATES

References

Arfken, G. "Conical Coordinates (ξ_1, ξ_2, ξ_3)." §2.16 in *Mathematical Methods for Physicists, 2nd ed.* Orlando, FL: Academic Press, pp. 118–119, 1970.

Byerly, W. E. *An Elementary Treatise on Fourier's Series, and Spherical, Cylindrical, and Ellipsoidal Harmonics, with Applications to Problems in Mathematical Physics.* New York: Dover, p. 263, 1959.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 659, 1953.

Spence, R. D. "Angular Momentum in Sphero-Conal Coordinates." *Amer. J. Phys.* **27**, 329–335, 1959.

Conical Frustum

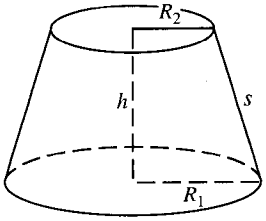

A conical frustum is a FRUSTUM created by slicing the top off a CONE (with the cut made parallel to the base). For a right circular CONE, let s be the slant height and R_1 and R_2 the top and bottom RADII. Then

$$s = \sqrt{(R_1 - R_2)^2 + h^2}. \tag{1}$$

The SURFACE AREA, not including the top and bottom CIRCLES, is

$$A = \pi(R_1 + R_2)s = \pi(R_1 + R_2)\sqrt{(R_1 - R_2)^2 + h^2}. \tag{2}$$

The VOLUME of the frustum is given by

$$V = \pi\int_0^h [r(z)]^2\, dz. \tag{3}$$

But

$$r(z) = R_1 + (R_2 - R_1)\frac{z}{h}, \tag{4}$$

so

$$\begin{aligned} V &= \pi\int_0^h \left[R_1 + (R_2 - R_1)\frac{z}{h}\right]^2 dz \\ &= \tfrac{1}{3}\pi h({R_1}^2 + R_1R_2 + {R_2}^2). \end{aligned} \tag{5}$$

This formula can be generalized to any PYRAMID by letting A_i be the base AREAS of the top and bottom of the frustum. Then the VOLUME can be written as

$$V = \tfrac{1}{3}h(A_1 + A_2 + \sqrt{A_1A_2}\,). \tag{6}$$

The weighted mean of z over the frustum is

$$\langle z\rangle = \pi\int_0^h z[r(z)]^2\, dz = \tfrac{1}{12}h^2({R_1}^2 + 2R_1R_2 + 3{R_2}^2). \tag{7}$$

The CENTROID is then given by

$$\bar z = \frac{\langle z\rangle}{V} = \frac{h({R_1}^2 + R_1R_2 + {R_2}^2)}{4({R_1}^2 + 2R_1R_2 + 3{R_2}^2)} \tag{8}$$

(Beyer 1987, p. 133). The special case of the CONE is given by taking $R_2 = 0$, yielding $\bar z = h/4$.

see also CONE, FRUSTUM, PYRAMIDAL FRUSTUM, SPHERICAL SEGMENT

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 129–130 and 133, 1987.

Conical Function

Functions which can be expressed in terms of LEGENDRE FUNCTIONS OF THE FIRST and SECOND KINDS. See Abramowitz and Stegun (1972, p. 337).

$$\begin{aligned} P^{\mu}_{-1/2+ip}(\cos\theta) &= 1 + \frac{4p^2+1^2}{2^2}\sin^2(\tfrac{1}{2}\theta) \\ &\quad + \frac{(4p^2+1^2)(4p^2+3^2)}{2^2 4^2}\sin^4(\tfrac{1}{2}\theta) + \ldots \\ &= \frac{2}{\pi}\int_0^{\theta} \frac{\cosh(pt)\,dt}{\sqrt{2(\cos t - \cos\theta)}} \end{aligned}$$

$$\begin{aligned} Q^{\mu}_{-1/2\mp ip}(\cos\theta) &= \pm i\sinh(p\pi)\int_0^{\infty} \frac{\cos(pt)\,dt}{\sqrt{2(\cosh t + \cos\theta)}} \\ &\quad + \int_0^{\infty} \frac{\cosh(pt)\,dt}{\sqrt{2(\cos t - \cos\theta)}}. \end{aligned}$$

see also TOROIDAL FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Conical Functions." §8.12 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 337, 1972.

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1464, 1980.

Conical Spiral

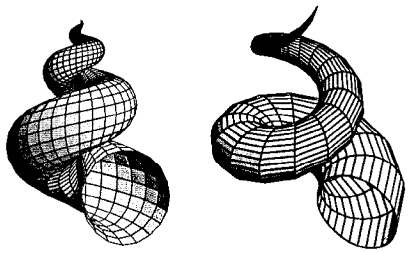

A surface modeled after the shape of a SEASHELL. One parameterization (left figure) is given by

$$x = 2[1 - e^{u/(6\pi)}]\cos u\cos^2(\tfrac{1}{2}v) \tag{1}$$

$$y = 2[-1 + e^{u/(6\pi)}]\cos^2(\tfrac{1}{2}v)\sin u \tag{2}$$

$$z = 1 - e^{u/(3\pi)} - \sin v + e^{u/(6\pi)}\sin v, \tag{3}$$

where $v \in [0, 2\pi)$, and $u \in [0, 6\pi)$ (Wolfram). Nordstrand gives the parameterization

$$x = \left[\left(1 - \frac{v}{2\pi}\right)(1+\cos u) + c\right]\cos(nv) \tag{4}$$

$$x = \left[\left(1 - \frac{v}{2\pi}\right)(1+\cos u) + c\right]\sin(nv) \tag{5}$$

$$z = \frac{bv}{2\pi} + a\sin u\left(1 - \frac{v}{2\pi}\right) \tag{6}$$

for $u, v \in [0, 2\pi]$ (right figure with $a = 0.2$, $b = 1$, $c = 0.1$, and $n = 2$).

References

Gray, A. "Sea Shells." §11.6 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 223–223, 1993.

Nordstrand, T. "Conic Spiral or Seashell." `http://www.uib.no/people/nfytn/shelltxt.htm`.

Wolfram Research "Mathematica Version 2.0 Graphics Gallery." `http://www.mathsource.com/cgi-bin/MathSource/Applications/Graphics/3D/0207-155`.

Conical Wedge

The SURFACE also called the CONOCUNEUS OF WALLIS and given by the parametric equation

$$\begin{aligned} x &= u\cos v \\ y &= u\sin v \\ z &= c(1 - 2\cos^2 v). \end{aligned}$$

see also CYLINDRICAL WEDGE, WEDGE

References

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 302, 1993.

Conjecture

A proposition which is consistent with known data, but has neither been verified nor shown to be false. It is synonymous with HYPOTHESIS.

see also ABC CONJECTURE, ABHYANKAR'S CONJECTURE, ABLOWITZ-RAMANI-SEGUR CONJECTURE, ANDRICA'S CONJECTURE, ANNULUS CONJECTURE, ARGOH'S CONJECTURE, ARTIN'S CONJECTURE, AXIOM, BACHET'S CONJECTURE, BENNEQUIN'S CONJECTURE, BIEBERBACH CONJECTURE, BIRCH CONJECTURE, BLASCHKE CONJECTURE, BORSUK'S CONJECTURE, BORWEIN CONJECTURES, BRAUN'S CONJECTURE, BROCARD'S CONJECTURE, BURNSIDE'S CONJECTURE, CARMICHAEL'S CONJECTURE, CATALAN'S CONJECTURE, CRAMÉR CONJECTURE, DE POLIGNAC'S CONJECTURE, DIESIS, DODECAHEDRAL CONJECTURE, DOUBLE BUBBLE CONJECTURE, EBERHART'S CONJECTURE, EULER'S CONJECTURE, EULER POWER CONJECTURE, EULER QUARTIC CONJECTURE, FEIT-THOMPSON CONJECTURE, FERMAT'S CONJECTURE, FLYPING CONJECTURE, GILBREATH'S CONJECTURE, GIUGA'S CONJECTURE, GOLDBACH CONJECTURE, GRIMM'S CONJECTURE, GUY'S CONJECTURE, HARDY-LITTLEWOOD CONJECTURES, HASSE'S CONJECTURE, HEAWOOD CONJECTURE, HYPOTHESIS, JACOBIAN CONJECTURE, KAPLAN-YORKE CONJECTURE, KELLER'S CONJECTURE, KELVIN'S CONJECTURE, KEPLER CONJECTURE, KREISEL CONJECTURE, KUMMER'S CONJECTURE, LEMMA, LOCAL DENSITY CONJECTURE, MERTENS CONJECTURE, MILIN CONJECTURE, MILNOR'S CONJECTURE, MORDELL CONJECTURE, NETTO'S CONJECTURE, NIRENBERG'S CONJECTURE, ORE'S CONJECTURE, PADÉ CONJECTURE, Palindromic Number Conjecture, Pillai's Conjecture, Poincaré Conjecture, Pólya Conjecture, Porism, Prime k-Tuples Conjecture, Prime Patterns Conjecture, Prime Power Conjecture, Proof, Quillen-Lichtenbaum Conjecture, Ramanujan-Petersson Conjecture, Robertson Conjecture, Safarevich Conjecture, Sausage Conjecture, Schanuel's Conjecture, Schisma, Scholz Conjecture, Seifert Conjecture, Selfridge's Conjecture, Shanks' Conjecture, Smith Conjecture, Swinnerton-Dyer Conjecture, Szpiro's Conjecture, Tait's Hamiltonian Graph Conjecture, Tait's Knot Conjectures, Taniyama-Shimura Conjecture, Tau Conjecture, Theorem, Thurston's Geometrization Conjecture, Thwaites Conjecture, Vojta's Conjecture, Wang's Conjecture, Waring's Prime Conjecture, Waring's Sum Conjecture, Zarankiewicz's Conjecture

References
Rivera, C. "Problems & Puzzles (Conjectures)." `http://www.sci.net.mx/~crivera/ppp/conjectures.htm`.

Conjugacy Class

A complete set of mutually conjugate Group elements. Each element in a Group belongs to exactly one class, and the identity ($I = 1$) element is always in its own class. The Orders of all classes must be integral Factors of the Order of the Group. From the last two statements, a Group of Prime order has one class for each element. More generally, in an Abelian Group, each element is in a conjugacy class by itself. Two operations belong to the same class when one may be replaced by the other in a new Coordinate System which is accessible by a symmetry operation (Cotton 1990, p. 52). These sets correspond directly to the sets of equivalent operation.

Let G be a Finite Group of Order $|G|$. If $|G|$ is Odd, then

$$|G| \equiv s \pmod{16}$$

(Burnside 1955, p. 295). Furthermore, if every Prime p_i Dividing $|G|$ satisfies $p_i \equiv 1 \pmod 4$, then

$$|G| \equiv s \pmod{32}$$

(Burnside 1955, p. 320). Poonen (1995) showed that if every Prime p_i Dividing $|G|$ satisfies $p_i \equiv 1 \pmod m$ for $m \geq 2$, then

$$|G| \equiv s \left(\mathrm{mod}\ 2m^2\right).$$

References
Burnside, W. *Theory of Groups of Finite Order, 2nd ed.* New York: Dover, 1955.
Cotton, F. A. *Chemical Applications of Group Theory, 3rd ed.* New York: Wiley, 1990.
Poonen, B. "Congruences Relating the Order of a Group to the Number of Conjugacy Classes." *Amer. Math. Monthly* **102**, 440–442, 1995.

Conjugate Element

Given a Group with elements A and X, there must be an element B which is a Similarity Transformation of $A, B = X^{-1}AX$ so A and B are conjugate with respect to X. Conjugate elements have the following properties:

1. Every element is conjugate with itself.
2. If A is conjugate with B with respect to X, then B is conjugate to A with respect to X.
3. If A is conjugate with B and C, then B and C are conjugate with each other.

see also Conjugacy Class, Conjugate Subgroup

Conjugate Gradient Method

An Algorithm for calculating the Gradient $\nabla f(\mathbf{P})$ of a function at an n-D point $\mathbf{P}$. It is more robust than the simpler Steepest Descent Method.

References
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 413–417, 1992.

Conjugate Points

see Harmonic Conjugate Points, Isogonal Conjugate, Isotomic Conjugate Point

Conjugate Subgroup

A Subgroup H of an original Group G has elements h_i. Let x be a fixed element of the original Group G which is not a member of H. Then the transformation xh_ix^{-1}, ($i = 1, 2, \ldots$) generates a conjugate Subgroup xHx^{-1}. If, for all x, $xHx^{-1} = H$, then H is a Self-Conjugate (also called Invariant or Normal) Subgroup. All Subgroups of an Abelian Group are invariant.

Conjugation

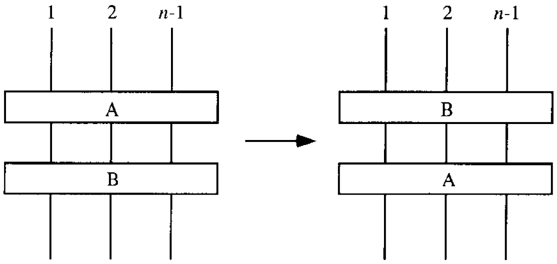

A type I Markov Move.

see also Markov Moves, Stabilization

Conjunction

A product of Ands, denoted

$$\bigwedge_{k+1}^{n} A_k.$$

see also And, Disjunction

Connected Graph

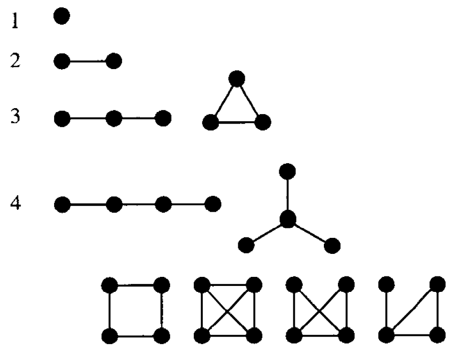

A GRAPH which is connected (as a TOPOLOGICAL SPACE), i.e., there is a path from any point to any other point in the GRAPH. The number of n-VERTEX (unlabeled) connected graphs for $n = 1, 2, \ldots$ are 1, 1, 2, 6, 21, 112, 853, 11117, ... (Sloane's A001349).

References
Chartrand, G. "Connected Graphs." §2.3 in *Introductory Graph Theory.* New York: Dover, pp. 41–45, 1985.
Sloane, N. J. A. Sequence A001349/M1657 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Connected Set

A connected set is a SET which cannot be partitioned into two nonempty SUBSETS which are open in the relative topology induced on the SET. Equivalently, it is a SET which cannot be partitioned into two nonempty SUBSETS such that each SUBSET has no points in common with the closure of the other.

The REAL NUMBERS are a connected set.

see also CLOSED SET, EMPTY SET, OPEN SET, SET, SUBSET

Connected Space

A SPACE D is connected if any two points in D can be connected by a curve lying wholly within D. A SPACE is 0-connected (a.k.a. PATHWISE-CONNECTED) if every MAP from a 0-SPHERE to the SPACE extends continuously to the 1-DISK. Since the 0-SPHERE is the two endpoints of an interval (1-DISK), every two points have a path between them. A space is 1-connected (a.k.a. SIMPLY CONNECTED) if it is 0-connected and if every MAP from the 1-SPHERE to it extends continuously to a MAP from the 2-DISK. In other words, every loop in the SPACE is contractible. A SPACE is n-MULTIPLY CONNECTED if it is $(n-1)$-connected and if every MAP from the n-SPHERE into it extends continuously over the $(n+1)$-DISK.

A theorem of Whitehead says that a SPACE is infinitely connected IFF it is contractible.

see also CONNECTIVITY, LOCALLY PATHWISE-CONNECTED SPACE, MULTIPLY CONNECTED, PATHWISE-CONNECTED, SIMPLY CONNECTED

Connected Sum

The connected sum $M_1 \# M_2$ of n-manifolds M_1 and M_2 is formed by deleting the interiors of n-BALLS B_i^n in M_i^n and attaching the resulting punctured MANIFOLDS $M_i - \dot{B}_i$ to each other by a HOMEOMORPHISM $h : \partial B_2 \to \partial B_1$, so

$$M_1 \# M_2 = (M_1 - \dot{B}_1) \bigcup_h (M_2 - \dot{B}_2).$$

B_i is required to be interior to M_i and ∂B_i bicollared in M_i to ensure that the connected sum is a MANIFOLD.

The connected sum of two KNOTS is called a KNOT SUM.

see also KNOT SUM

References
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 39, 1976.

Connected Sum Decomposition

Every COMPACT 3-MANIFOLD is the CONNECTED SUM of a unique collection of PRIME 3-MANIFOLDS.

see also JACO-SHALEN-JOHANNSON TORUS DECOMPOSITION

Connection

see CONNECTION COEFFICIENT, GAUSS-MANIN CONNECTION

Connection Coefficient

A quantity also known as a CHRISTOFFEL SYMBOL OF THE SECOND KIND. Connection COEFFICIENTS are defined by

$$\Gamma^{\vec{e}_\alpha}_{\vec{e}_\beta \vec{e}_\gamma} \equiv \vec{e}^{\,\alpha} \cdot (\nabla_{\vec{e}_\gamma} \vec{e}_\beta) \qquad (1)$$

(long form) or

$$\Gamma^{\alpha}_{\beta\gamma} \equiv \vec{e}^{\,\alpha} \cdot (\nabla_\gamma \vec{e}_\beta), \qquad (2)$$

(abbreviated form), and satisfy

$$\nabla_{\vec{e}_\gamma} \vec{e}_\beta = \Gamma^{\vec{e}_\alpha}_{\vec{e}_\beta \vec{e}_\gamma} \vec{e}_\alpha \qquad (3)$$

(long form) and

$$\nabla_\gamma \vec{e}_\beta = \Gamma^{\alpha}_{\beta\gamma} \vec{e}_\alpha \qquad (4)$$

(abbreviated form).

Connection COEFFICIENTS are not TENSORS, but have TENSOR-like CONTRAVARIANT and COVARIANT indices. A fully COVARIANT connection COEFFICIENT is given by

$$\Gamma_{\alpha\beta\gamma} \equiv \tfrac{1}{2}(g_{\alpha\beta,\gamma} + g_{\alpha\gamma,\beta} + c_{\alpha\beta\gamma} + c_{\alpha\gamma\beta} - c_{\beta\gamma\alpha}), \qquad (5)$$

where the gs are the METRIC TENSORS, the cs are COMMUTATION COEFFICIENTS, and the commas indicate the

COMMA DERIVATIVE. In an ORTHONORMAL BASIS, $g_{\alpha\beta,\gamma} = 0$ and $g_{\mu\gamma} = \delta_{\mu\gamma}$, so

$$\Gamma_{\alpha\beta\gamma} = \Gamma^{\mu}_{\alpha\beta} g_{\mu\gamma} = \Gamma_{\alpha\beta}{}^{\mu} = \tfrac{1}{2}(c_{\alpha\beta\gamma} + c_{\alpha\gamma\beta} - c_{\beta\gamma\alpha}) \tag{6}$$

and

$$\Gamma_{ijk} = 0 \qquad \text{for } i \neq j \neq k \tag{7}$$

$$\Gamma_{iik} = -\frac{1}{2}\frac{\partial g_{ii}}{\partial x^k} \qquad \text{for } i \neq k \tag{8}$$

$$\Gamma_{iji} = \Gamma_{jii} = \frac{1}{2}\frac{\partial g_{ii}}{\partial x^j} \tag{9}$$

$$\Gamma^k_{ij} = 0 \qquad \text{for } i \neq j \neq k \tag{10}$$

$$\Gamma^k_{ii} = -\frac{1}{2g_{kk}}\frac{\partial g_{ii}}{\partial x^k} \qquad \text{for } i \neq k \tag{11}$$

$$\Gamma^i_{ij} = \Gamma^i_{ji} = \frac{1}{2g_{ii}}\frac{\partial g_{ii}}{\partial x^j} = \frac{1}{2}\frac{\partial \ln g_{ii}}{\partial x^j}. \tag{12}$$

For TENSORS of RANK 3, the connection COEFFICIENTS may be concisely summarized in MATRIX form:

$$\Gamma^{\theta} \equiv \begin{bmatrix} \Gamma^{\theta}_{rr} & \Gamma^{\theta}_{r\theta} & \Gamma^{\theta}_{r\phi} \\ \Gamma^{\theta}_{\theta r} & \Gamma^{\theta}_{\theta\theta} & \Gamma^{\theta}_{\theta\phi} \\ \Gamma^{\theta}_{\phi r} & \Gamma^{\theta}_{\phi\theta} & \Gamma^{\theta}_{\phi\phi} \end{bmatrix}. \tag{13}$$

Connection COEFFICIENTS arise in the computation of GEODESICS. The GEODESIC EQUATION of free motion is

$$d\tau^2 = -\eta_{\alpha\beta}\, d\xi^{\alpha}\, d\xi^{\beta}, \tag{14}$$

or

$$\frac{d^2\xi^{\alpha}}{d\tau^2} = 0. \tag{15}$$

Expanding,

$$\frac{d}{d\tau}\left(\frac{\partial\xi^{\alpha}}{\partial x^{\mu}}\frac{dx^{\mu}}{d\tau}\right) = \frac{\partial\xi^{\alpha}}{\partial x^{\mu}}\frac{d^2x^{\mu}}{d\tau^2} + \frac{\partial^2\xi^{\alpha}}{\partial x^{\mu}\partial x^{\nu}}\frac{dx^{\mu}}{d\tau}\frac{dx^{\nu}}{d\tau} = 0 \tag{16}$$

$$\frac{\partial\xi^{\alpha}}{\partial x^{\mu}}\frac{d^2x^{\mu}}{d\tau^2}\frac{\partial x^{\lambda}}{\partial\xi^{\alpha}} + \frac{\partial^2\xi^{\alpha}}{\partial x^{\mu}\partial x^{\nu}}\frac{dx^{\mu}}{d\tau}\frac{dx^{\nu}}{d\tau}\frac{\partial x^{\lambda}}{\partial\xi^{\alpha}} = 0. \tag{17}$$

But

$$\frac{\partial\xi^{\alpha}}{\partial x^{\nu}}\frac{\partial x^{\lambda}}{\partial\xi^{\alpha}} = \delta^{\lambda}_{\mu}, \tag{18}$$

so

$$\delta^{\lambda}_{\mu}\frac{d^2x^{\mu}}{d\tau^2} + \left[\frac{\partial^2\xi^{\alpha}}{\partial x^{\mu}\partial x^{\nu}}\frac{\partial x^{\lambda}}{\partial\xi^{\alpha}}\right]\frac{dx^{\mu}}{d\tau}\frac{dx^{\nu}}{d\tau} = \frac{d^2x^{\lambda}}{d\tau^2} + \Gamma^{\lambda}_{\mu\nu}\frac{dx^{\mu}}{d\tau}\frac{dx^{\nu}}{d\tau}, \tag{19}$$

where

$$\Gamma^{\lambda}_{\mu\nu} \equiv \frac{\partial^2\xi^{\alpha}}{\partial x^{\mu}\partial x^{\nu}}\frac{\partial x^{\lambda}}{\partial\xi^{\alpha}}. \tag{20}$$

see also CARTAN TORSION COEFFICIENT, CHRISTOFFEL SYMBOL OF THE FIRST KIND, CHRISTOFFEL SYMBOL OF THE SECOND KIND, COMMA DERIVATIVE, COMMUTATION COEFFICIENT, CURVILINEAR COORDINATES, SEMICOLON DERIVATIVE, TENSOR

Connectivity

see CONNECTED SPACE, EDGE CONNECTIVITY, VERTEX CONNECTIVITY

Connes Function

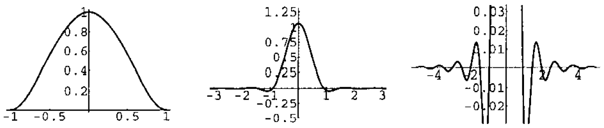

The APODIZATION FUNCTION

$$A(x) = \left(1 - \frac{x^2}{a^2}\right)^2.$$

Its FULL WIDTH AT HALF MAXIMUM is $\sqrt{4 - 2\sqrt{2}}\,a$, and its INSTRUMENT FUNCTION is

$$I(x) = 8a\sqrt{2\pi}\,\frac{J_{5/2}(2\pi ka)}{(2\pi ka)^{5/2}},$$

where $J_n(z)$ is a BESSEL FUNCTION OF THE FIRST KIND.

see also APODIZATION FUNCTION

Conocuneus of Wallis

see CONICAL WEDGE

Conoid

see PLÜCKER'S CONOID, RIGHT CONOID

Consecutive Number Sequences

Consecutive number sequences are sequences constructed by concatenating numbers of a given type. Many of these sequences were considered by Smarandache, so they are sometimes known as SMARANDACHE SEQUENCES.

The nth term of the consecutive integer sequence consists of the concatenation of the first n POSITIVE integers: 1, 12, 123, 1234, ... (Sloane's A007908; Smarandache 1993, Dumitrescu and Seleacu 1994, sequence 1; Mudge 1995; Stephen 1998). This sequence gives the digits of the CHAMPERNOWNE CONSTANT and contains no PRIMES in the first 4,470 terms (Weisstein). This is roughly consistent with simple arguments based on the distribution of prime which suggest that only a single prime is expected in the first 15,000 or so terms. The number of digits of the n term can be computed by noticing the pattern in the following table, where $d = \lfloor \log_{10} n \rfloor + 1$ is the number of digits in n.

d	n Range	Digits
1	1–9	n
2	10–99	$9 + 2(n - 9)$
3	100–999	$9 + 90 \cdot 2 + 3(n - 99)$
4	1000–9999	$9 + 90 \cdot 2 + 900 \cdot 3 + 4(n - 999)$

Therefore, the number of digits $D(n)$ in the nth term can be written

$$\begin{aligned} D(n) &= d(n+1-10^{d-1}) + \sum_{k=1}^{d-1} 9k \cdot 10^{k-1} \\ &= (n+1)d - \frac{10^d - 1}{9}, \end{aligned}$$

where the second term is the REPUNIT R_d.

The nth term of the reverse integer sequence consists of the concatenation of the first n POSITIVE integers written backwards: 1, 21, 321, 4321, ... (Sloane's A000422; Smarandache 1993, Dumitrescu and Seleacu 1994, Stephen 1998). The only PRIME in the first 3,576 terms (Weisstein) of this sequence is the 82nd term 828180...321 (Stephen 1998), which has 155 digits. This is roughly consistent with simple arguments based on the distribution of prime which suggest that a single prime is expected in the first 15,000 or so terms. The terms of the reverse integer sequence have the same number of digits as do the consecutive integer sequence.

The concatenation of the first n PRIMES gives 2, 23, 235, 2357, 235711, ... (Sloane's A019518; Smith 1996, Mudge 1997). This sequence converges to the digits of the COPELAND-ERDŐS CONSTANT and is PRIME for terms 1, 2, 4, 128, 174, 342, 435, 1429, ... (Sloane's A046035; Ibstedt 1998, pp. 78–79), with no others less than 2,305 (Weisstein).

The concatenation of the first n ODD NUMBERS gives 1, 13, 135, 1357, 13579, ... (Sloane's A019519; Smith 1996, Marimutha 1997, Mudge 1997). This sequence is PRIME for terms 2, 10, 16, 34, 49, 2570, ... (Sloane's A046036; Weisstein, Ibstedt 1998, pp. 75–76), with no others less than 2,650 (Weisstein). The 2570th term, given by 1 3 5 7...5137 5139, has 9725 digits and was discovered by Weisstein in Aug. 1998.

The concatenation of the first n EVEN NUMBERS gives 2, 24, 246, 2468, 246810, ... (Sloane's A019520; Smith 1996; Marimutha 1997; Mudge 1997; Ibstedt 1998, pp. 77–78).

The concatenation of the first n SQUARE NUMBERS gives 1, 14, 149, 14916, ... (Sloane's A019521; Marimutha 1997). The only PRIME in the first 2,090 terms is the third term, 149, (Weisstein).

The concatenation of the first n CUBIC NUMBERS gives 1, 18, 1827, 182764, ... (Sloane's A019522; Marimutha 1997). There are no PRIMES in the first 1,830 terms (Weisstein).

see also CHAMPERNOWNE CONSTANT, CONCATENATION, COPELAND-ERDŐS CONSTANT, CUBIC NUMBER, DEMLO NUMBER, EVEN NUMBER, ODD NUMBER, SMARANDACHE SEQUENCES, SQUARE NUMBER

References

Dumitrescu, C. and Seleacu, V. (Ed.). *Some Notions and Questions in Number Theory.* Glendale, AZ: Erhus University Press, 1994.

Ibstedt, H. "Smarandache Concatenated Sequences." Ch. 5 in *Computer Analysis of Number Sequences.* Lupton, AZ: American Research Press, pp. 75–79, 1998.

Marimutha, H. "Smarandache Concatenate Type Sequences." *Bull. Pure Appl. Sci.* **16E**, 225–226, 1997.

Mudge, M. "Top of the Class." *Personal Computer World,* 674–675, June 1995.

Mudge, M. "Not Numerology but Numeralogy!" *Personal Computer World,* 279–280, 1997.

Smarandache, F. *Only Problems, Not Solutions!, 4th ed.* Phoenix, AZ: Xiquan, 1993.

Smith, S. "A Set of Conjectures on Smarandache Sequences." *Bull. Pure Appl. Sci.* **15E**, 101–107, 1996.

Stephen, R. W. "Factors and Primes in Two Smarandache Sequences." *Smarandache Notions J.* **9**, 4–10, 1998. `http://www.tmt.de/~stephen/sm.ps.gz`.

Conservation of Number Principle

A generalization of Poncelet's PERMANENCE OF MATHEMATICAL RELATIONS PRINCIPLE made by H. Schubert in 1874–79. The conservation of number principle asserts that the number of solutions of any determinate algebraic problem in any number of parameters under variation of the parameters is invariant in such a manner that no solutions become INFINITE. Schubert called the application of this technique the CALCULUS of ENUMERATIVE GEOMETRY.

see also DUALITY PRINCIPLE, HILBERT'S PROBLEMS, PERMANENCE OF MATHEMATICAL RELATIONS PRINCIPLE

References

Bell, E. T. *The Development of Mathematics, 2nd ed.* New York: McGraw-Hill, p. 340, 1945.

Conservative Field

The following conditions are equivalent for a conservative VECTOR FIELD:

1. For any oriented simple closed curve C, the LINE INTEGRAL $\oint_C \mathbf{F} \cdot d\mathbf{s} = 0$.
2. For any two oriented simple curves C_1 and C_2 with the same endpoints, $\int_{C_1} \mathbf{F} \cdot d\mathbf{s} = \int_{C_2} \mathbf{F} \cdot d\mathbf{s}$.
3. There exists a SCALAR POTENTIAL FUNCTION f such that $\mathbf{F} = \nabla f$, where ∇ is the GRADIENT.
4. The CURL $\nabla \times \mathbf{F} = \mathbf{0}$.

see also CURL, GRADIENT, LINE INTEGRAL, POTENTIAL FUNCTION, VECTOR FIELD

Consistency

The absence of contradiction (i.e., the ability to prove that a statement and its NEGATIVE are both true) in an AXIOMATIC THEORY is known as consistency.

see also COMPLETE AXIOMATIC THEORY, CONSISTENCY STRENGTH

Consistency Strength

If the Consistency of one of two propositions implies the Consistency of the other, the first is said to have greater consistency strength.

Constant

Any Real Number which is "significant" (or interesting) in some way. In this work, the term "constant" is generally reserved for Real nonintegral numbers of interest, while "Number" is reserved for interesting Integers (e.g., Brun's Constant, but Beast Number).

Certain constants are known to many Decimal Digits and recur throughout many diverse areas of mathematics, often in unexpected and surprising places (e.g., Pi, e, and to some extent, the Euler-Mascheroni Constant γ). Other constants are more specialized and may be known to only a few Digits. S. Plouffe maintains a site about the computation and identification of numerical constants. Plouffe's site also contains a page giving the largest number of Digits computed for the most common constants. S. Finch maintains a delightful, more expository site containing detailed essays and references on constants both common and obscure.

see also Abundant Number, Alladi-Grinstead Constant, Apéry's Constant, Archimedes' Constant, Artin's Constant, Backhouse's Constant, Beraha Constants, Bernstein's Constant, Bloch Constant, Brun's Constant, Cameron's Sum-Free Set Constant, Carlson-Levin Constant, Catalan's Constant, Chaitin's Constant, Champernowne Constant, Chebyshev Constants, Chebyshev-Sylvester Constant, Comma of Didymus, Comma of Pythagoras, Conic Constant, Constant Function, Constant Problem, Continued Fraction Constant, Conway's Constant, Copeland-Erdős Constant, Copson-de Bruijn Constant, de Bruijn-Newman Constant, Delian Constant, Diesis, Du Bois Raymond Constants, e, Ellison–Mendès-France Constant, Erdős Reciprocal Sum Constants, Euler-Mascheroni Constant, Extreme Value Distribution, Favard Constants, Feller's Coin-Tossing Constants, Fransén-Robinson Constant, Freiman's Constant, Gauss's Circle Problem, Gauss's Constant, Gauss-Kuzmin-Wirsing Constant, Gelfond-Schneider Constant, Geometric Probability Constants, Gibbs Constant, Glaisher-Kinkelin Constant, Golden Mean, Golomb Constant, Golomb-Dickman Constant, Gompertz Constant, Grossman's Constant, Grothendieck's Majorant, Hadamard-Vallée Poussin Constants, Hafner-Sarnak-McCurley Constant, Halphen Constant, Hard Square Entropy Constant, Hardy-Littlewood Constants, Hermite Constants, Hilbert's Constants, Infinite Product, Iterated Exponential Constants, Khintchine's Constant, Khintchine-Lévy Constant, Koebe's Constant, Kolmogorov Constant, Lal's Constant, Landau Constant, Landau-Kolmogorov Constants, Landau-Ramanujan Constant, Lebesgue Constants (Fourier Series), Lebesgue Constants (Lagrange Interpolation), Legendre's Constant, Lehmer's Constant, Lengyel's Constant, Lévy Constant, Linnik's Constant, Liouville's Constant, Liouville-Roth Constant, Ludolph's Constant, Madelung Constants, Magic Constant, Magic Geometric Constants, Masser-Gramain Constant, Mertens Constant, Mills' Constant, Moving Sofa Constant, Napier's Constant, Nielsen-Ramanujan Constants, Niven's Constant, Omega Constant, One-Ninth Constant, Otter's Tree Enumeration Constants, Parity Constant, Pi, Pisot-Vijayaraghavan Constants, Plastic Constant, Plouffe's Constant, Polygon Circumscribing Constant, Polygon Inscribing Constant, Porter's Constant, Pythagoras's Constant, Quadratic Recurrence, Quadtree, Rabbit Constant, Ramanujan Constant, Random Walk, Rényi's Parking Constants, Robbin Constant, Salem Constants, Self-Avoiding Walk, Shah-Wilson Constant, Shallit Constant, Shapiro's Cyclic Sum Constant, Sierpiński Constant, Silver Constant, Silverman Constant, Smarandache Constants, Soldner's Constant, Sphere Packing, Stieltjes Constants, Stolarsky-Harborth Constant, Sylvester's Sequence, Thue Constant, Thue-Morse Constant, Totient Function Constants, Traveling Salesman Constants, Tree Searching, Twin Primes Constant, Varga's Constant, W2-Constant, Weierstraß Constant, Whitney-Mikhlin Extension Constants, Wilbraham-Gibbs Constant, Wirtinger-Sobolev Isoperimetric Constants

References

Borwein, J. and Borwein, P. *A Dictionary of Real Numbers.* London: Chapman & Hall, 1990.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/constant.html`.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, 1983.

Plouffe, S. "Inverse Symbolic Calculator Table of Constants." `http://www.cecm.sfu.ca/projects/ISC/I_d.html`.

Plouffe, S. "Plouffe's Inverter." `http://www.lacim.uqam.ca/pi/`.

Plouffe, S. "Plouffe's Inverter: Table of Current Records for the Computation of Constants." `http://lacim.uqam.ca/pi/records.html`.

Wells, D. W. *The Penguin Dictionary of Curious and Interesting Numbers.* Harmondsworth, England: Penguin Books, 1986.

Constant Function

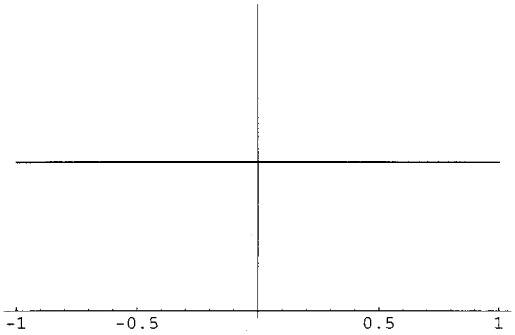

A FUNCTION $f(x) = c$ which does not change as its parameters vary. The GRAPH of a 1-D constant FUNCTION is a straight LINE. The DERIVATIVE of a constant FUNCTION c is

$$\frac{d}{dx}c = 0, \tag{1}$$

and the INTEGRAL is

$$\int c\,dx = cx. \tag{2}$$

The FOURIER TRANSFORM of the constant function $f(x) = 1$ is given by

$$\mathcal{F}[1] = \int_{-\infty}^{\infty} e^{-2\pi ikx}\,dx = \delta(k), \tag{3}$$

where $\delta(k)$ is the DELTA FUNCTION.

see also FOURIER TRANSFORM—1

References

Spanier, J. and Oldham, K. B. "The Constant Function c." Ch. 1 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 11–14, 1987.

Constant Precession Curve

see CURVE OF CONSTANT PRECESSION

Constant Problem

Given an expression involving known constants, integration in finite terms, computation of limits, etc., determine if the expression is equal to ZERO. The constant problem is a very difficult unsolved problem in TRANSCENDENTAL NUMBER theory. However, it is known that the problem is UNDECIDABLE if the expression involves oscillatory functions such as SINE. However, the FERGUSON-FORCADE ALGORITHM is a practical algorithm for determining if there exist integers a_i for given real numbers x_i such that

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0,$$

or else establish bounds within which no relation can exist (Bailey 1988).

see also FERGUSON-FORCADE ALGORITHM, INTEGER RELATION, SCHANUEL'S CONJECTURE

References

Bailey, D. H. "Numerical Results on the Transcendence of Constants Involving π, e, and Euler's Constant." *Math. Comput.* **50**, 275–281, 1988.

Sackell, J. "Zero-Equivalence in Function Fields Defined by Algebraic Differential Equations." *Trans. Amer. Math. Soc.* **336**, 151–171, 1993.

Constant Width Curve

see CURVE OF CONSTANT WIDTH

Constructible Number

A number which can be represented by a FINITE number of ADDITIONS, SUBTRACTIONS, MULTIPLICATIONS, DIVISIONS, and FINITE SQUARE ROOT extractions of integers. Such numbers correspond to LINE SEGMENTS which can be constructed using only STRAIGHTEDGE and COMPASS.

All RATIONAL NUMBERS are constructible, and all constructible numbers are ALGEBRAIC NUMBERS (Courant and Robbins 1996, p. 133). If a CUBIC EQUATION with rational coefficients has no rational root, then none of its roots is constructible (Courant and Robbins, p. 136).

In particular, let F_0 be the FIELD of RATIONAL NUMBERS. Now construct an extension field F_1 of constructible numbers by the adjunction of $\sqrt{k_0}$, where k_0 is in F_0, but $\sqrt{k_0}$ is not, consisting of all numbers of the form $a_0 + b_0\sqrt{k_0}$, where $a_0, b_0 \in F_0$. Next, construct an extension field F_2 of F_1 by the adjunction of $\sqrt{k_1}$, defined as the numbers $a_1 + b_1\sqrt{k_1}$, where $a_1, b_1 \in F_1$, and k_1 is a number in F_1 for which $\sqrt{k_1}$ does not lie in F_1. Continue the process n times. Then constructible numbers are precisely those which can be reached by such a sequence of extension fields F_n, where n is a measure of the "complexity" of the construction (Courant and Robbins 1996).

see also ALGEBRAIC NUMBER, COMPASS, CONSTRUCTIBLE POLYGON, EUCLIDEAN NUMBER, RATIONAL NUMBER, STRAIGHTEDGE

References

Courant, R. and Robbins, H. "Constructible Numbers and Number Fields." §3.2 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 127–134, 1996.

Constructible Polygon

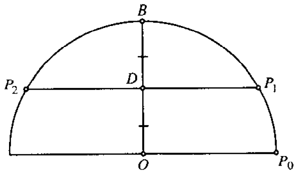

Equilateral Triangle

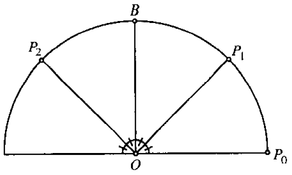

Square

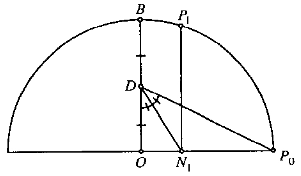

Pentagon

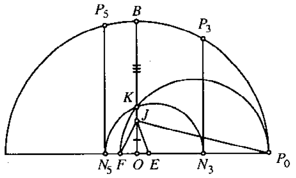

17-gon

COMPASS and STRAIGHTEDGE constructions dating back to Euclid were capable of inscribing regular polygons of 3, 4, 5, 6, 8, 10, 12, 16, 20, 24, 32, 40, 48, 64, ..., sides. However, this listing is not a complete enumeration of "constructible" polygons. A regular n-gon ($n \geq 3$) can be constructed by STRAIGHTEDGE and COMPASS IFF

$$n = 2^k p_1 p_2 \cdots p_s,$$

where k is in INTEGER ≥ 0 and the p_i are distinct FERMAT PRIMES. FERMAT NUMBERS are of the form

$$F_m = 2^{2^m} + 1,$$

where m is an INTEGER ≥ 0. The only known PRIMES of this form are 3, 5, 17, 257, and 65537. The fact that this condition was SUFFICIENT was first proved by Gauss in 1796 when he was 19 years old. That this condition was also NECESSARY was not explicitly proven by Gauss, and the first proof of this fact is credited to Wantzel (1836).

see also COMPASS, CONSTRUCTIBLE NUMBER, GEOMETRIC CONSTRUCTION, GEOMETROGRAPHY, HEPTADECAGON, HEXAGON, OCTAGON, PENTAGON, POLYGON, SQUARE, STRAIGHTEDGE, TRIANGLE

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 94–96, 1987.

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 119, 1996.

De Temple, D. W. "Carlyle Circles and the Lemoine Simplicity of Polygonal Constructions." *Amer. Math. Monthly* **98**, 97–108, 1991.

Dixon, R. "Compass Drawings." Ch. 1 in *Mathographics.* New York: Dover, pp. 1–78, 1991.

Gauss, C. F. §365 and 366 in *Disquisitiones Arithmeticae.* Leipzig, Germany, 1801. Translated by A. A. Clarke. New Haven, CT: Yale University Press, 1965.

Kazarinoff, N. D. "On Who First Proved the Impossibility of Constructing Certain Regular Polygons with Ruler and Compass Alone." *Amer. Math. Monthly* **75**, 647–648, 1968.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 137–138, 1990.

Wantzel, P. L. "Recherches sur les moyens de reconnaître si un Problème de Géométrie peut se résoudre avec la règle et le compas." *J. Math. pures appliq.* **1**, 366–372, 1836.

Construction

see GEOMETRIC CONSTRUCTION

Constructive Dilemma

A formal argument in LOGIC in which it is stated that (1) $P \Rightarrow Q$ and $R \Rightarrow S$ (where $\Rightarrow$ means "IMPLIES"), and (2) either P or R is true, from which two statements it follows that either Q or S is true.

see also DESTRUCTIVE DILEMMA, DILEMMA

Contact Angle

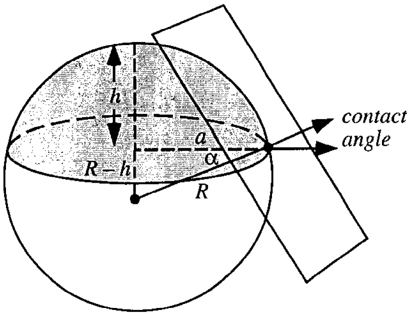

The ANGLE α between the normal vector of a SPHERE (or other geometric object) at a point where a PLANE is tangent to it and the normal vector of the plane. In the above figure,

$$\begin{aligned} \alpha &= \cos^{-1}\left(\frac{a}{R}\right) \\ &= \sin^{-1}\left(\frac{R-h}{R}\right). \end{aligned}$$

see also SPHERICAL CAP

Contact Number

see KISSING NUMBER

Contact Triangle

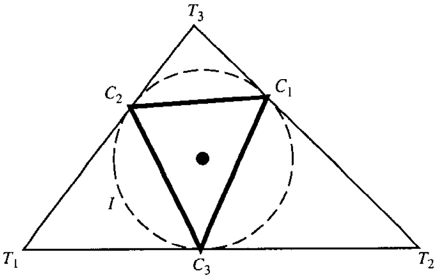

The TRIANGLE formed by the points of intersection of a TRIANGLE T's INCIRCLE with T. This is the PEDAL TRIANGLE of T with the INCENTER as the PEDAL POINT (c.f., TANGENTIAL TRIANGLE). The contact triangle and TANGENTIAL TRIANGLE are perspective from the GERGONNE POINT.

see also GERGONNE POINT, PEDAL TRIANGLE, TANGENTIAL TRIANGLE

References
Oldknow, A. "The Euler-Gergonne-Soddy Triangle of a Triangle." *Amer. Math. Monthly* **103**, 319–329, 1996.

Content

The generalized VOLUME for an n-D object (the "HYPERVOLUME").

see also VOLUME

Contiguous Function

A HYPERGEOMETRIC FUNCTION in which one parameter changes by +1 or −1 is said to be contiguous. There are 26 functions contiguous to ${}_2F_1(a,b,c;x)$ taking one pair at a time. There are 325 taking two or more pairs at a time. See Abramowitz and Stegun (1972, pp. 557–558).

see also HYPERGEOMETRIC FUNCTION

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, 1972.

Continued Fraction

A "general" continued fraction representation of a REAL NUMBER x is of the form

$$x = b_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{b_3 + \dots}}}, \tag{1}$$

which can be written

$$x = b_0 + \frac{a_1}{b_1+}\,\frac{a_2}{b_2+}\cdots. \tag{2}$$

The SIMPLE CONTINUED FRACTION representation of x (which is usually what is meant when the term "continued fraction" is used without qualification) of a number is given by

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \dots}}}, \tag{3}$$

which can be written in a compact abbreviated NOTATION as

$$x = [a_0, a_1, a_2, a_3, \dots]. \tag{4}$$

Here,

$$a_0 = \lfloor x \rfloor \tag{5}$$

is the integral part of σ (where $\lfloor x \rfloor$ is the FLOOR FUNCTION),

$$a_1 = \left\lfloor \frac{1}{x - a_0} \right\rfloor \tag{6}$$

is the integral part of the RECIPROCAL of $x - a_0$, a_2 is the integral part of the reciprocal of the remainder, etc. The quantities a_i are called PARTIAL QUOTIENTS. An archaic word for a continued fraction is ANTHYPHAIRETIC RATIO.

Continued fractions provide, in some sense, a series of "best" estimates for an IRRATIONAL NUMBER. Functions can also be written as continued fractions, providing a series of better and better rational approximations. Continued fractions have also proved useful in the proof of certain properties of numbers such as e and π (PI). Because irrationals which are square roots of RATIONAL NUMBERS have periodic continued fractions, an exact representation for a tabulated numerical value (i.e., 1.414... for PYTHAGORAS'S CONSTANT, $\sqrt{2}$) can sometimes be found.

Continued fractions are also useful for finding near commensurabilities between events with different periods. For example, the Metonic cycle used for calendrical purposes by the Greeks consists of 235 lunar months which very nearly equal 19 solar years, and 235/19 is the sixth CONVERGENT of the ratio of the lunar phase (synodic) period and solar period (365.2425/29.53059). Continued fractions can also be used to calculate gear ratios, and were used for this purpose by the ancient Greeks (Guy 1990).

If only the first few terms of a continued fraction are kept, the result is called a CONVERGENT. Let P_n/Q_n be convergents of a nonsimple continued fraction. Then

$$P_{-1} \equiv 1 \qquad Q_{-1} \equiv 0 \tag{7}$$

$$P_0 \equiv b_0 \qquad Q_0 \equiv 1 \tag{8}$$

and subsequent terms are calculated from the RECURRENCE RELATIONS

$$P_j = b_j P_{j-1} + a_j P_{j-2} \tag{9}$$

$$Q_j = b_j Q_{j-1} + a_j Q_{j-2} \tag{10}$$

for $j = 1, 2, \dots, n$. It is also true that

$$P_n Q_{n-1} - P_{n-1} Q_n = (-1)^{n-1} \prod_{k=1}^{n} a_k. \tag{11}$$

The error in approximating a number by a given CONVERGENT is roughly the MULTIPLICATIVE INVERSE of the square of the DENOMINATOR of the first neglected term.

A *finite simple* continued fraction representation terminates after a finite number of terms. To "round" a continued fraction, truncate the last term unless it is ±1, in which case it should be added to the previous term (Beeler *et al.* 1972, Item 101A). To take one over a continued fraction, add (or possibly delete) an initial 0 term. To negate, take the NEGATIVE of all terms, optionally using the identity

$$[-a, -b, -c, -d, \ldots] = [-a-1, 1, b-1, c, d, \ldots]. \quad (12)$$

A particularly beautiful identity involving the terms of the continued fraction is

$$\frac{[a_0, a_1, \ldots, a_n]}{[a_0, a_1, \ldots, a_{n-1}]} = \frac{[a_n, a_{n-1}, \ldots, a_1, a_0]}{[a_n, a_{n-1}, \ldots, a_1]}. \quad (13)$$

Finite simple fractions represent rational numbers and all rational numbers are represented by finite continued fractions. There are two possible representations for a finite simple fraction:

$$[a_1, \ldots, a_n] = \begin{cases} [a_1, \ldots, a_{n-1}, a_n - 1, 1] & \text{for } a_n > 1 \\ [a_1, \ldots, a_{n-2}, a_{n-1} + 1] & \text{for } a_n = 1. \end{cases} \quad (14)$$

On the other hand, an infinite simple fraction represents a unique IRRATIONAL NUMBER, and each IRRATIONAL NUMBER has a unique infinite continued fraction.

Consider the CONVERGENTS p_n/q_n of a *simple* continued fraction, and define

$$p_{-1} \equiv 0 \qquad q_{-1} \equiv 1 \quad (15)$$

$$p_0 \equiv 1 \qquad q_0 \equiv 0 \quad (16)$$

$$p_1 \equiv a_1 \qquad q_1 \equiv 1. \quad (17)$$

Then subsequent terms can be calculated from the RECURRENCE RELATIONS

$$p_i = a_i p_{i-1} + p_{i-2} \quad (18)$$

$$q_i = a_i q_{i-1} + q_{i-2}. \quad (19)$$

The CONTINUED FRACTION FUNDAMENTAL RECURRENCE RELATION for *simple* continued fractions is

$$p_n q_{n-1} - p_{n-1} q_n = (-1)^n. \quad (20)$$

It is also true that if $a_1 \neq 0$,

$$\frac{p_n}{p_{n-1}} = [a_n, a_{n-1}, \ldots, a_1] \quad (21)$$

$$\frac{q_n}{q_{n-1}} = [a_n, \ldots, a_2]. \quad (22)$$

Furthermore,

$$\frac{p_n}{q_n} = \frac{p_{n+1} - p_{n-1}}{q_{n+1} - q_{n-1}} \quad (23)$$

$$\begin{aligned} p_n = (n-1)p_{n-1} + (n-1)p_{n-2} + (n-2)p_{n-3} \\ + \ldots + 3p_2 + 2p_1 + p_1 + 1. \end{aligned} \quad (24)$$

Also, if $p/q > 1$ and

$$\frac{p}{q} = [a_1, a_2, \ldots, a_n], \quad (25)$$

then

$$\frac{q}{p} = [0, a_1, \ldots, a_n]. \quad (26)$$

Similarly, if $p/q < 1$ so

$$\frac{p}{q} = [0, a_1, \ldots, a_n], \quad (27)$$

then

$$\frac{q}{p} = [a_1, \ldots, a_n]. \quad (28)$$

The convergents also satisfy

$$c_n - c_{n-1} = \frac{(-1)^n}{q_n q_{n-1}} \quad (29)$$

$$c_n - c_{n-2} = \frac{a_n(-1)^{n-1}}{q_n q_{n-2}}. \quad (30)$$

The ODD convergents c_{2n+1} of an infinite simple continued fraction form an INCREASING SEQUENCE, and the EVEN convergents c_{2n} form a DECREASING SEQUENCE (so any ODD convergent is less than any EVEN convergent). Summarizing,

$$\begin{aligned} c_1 < c_3 < c_5 < \cdots < c_{2n+1} < \cdots \\ < c_{2n} < \cdots < c_6 < c_4 < c_2. \end{aligned} \quad (31)$$

Furthermore, each convergent for $n \geq 3$ lies between the two preceding ones. Each convergent is nearer to the value of the infinite continued fraction than the previous one. Let p_n/q_n be the nth continued fraction representation. Then

$$\frac{1}{(a_{n+1}+2){q_n}^2} < \left| \sigma - \frac{p_n}{q_n} \right| < \frac{1}{a_{n+1}{q_n}^2}. \quad (32)$$

The SQUARE ROOT of a SQUAREFREE INTEGER has a periodic continued fraction of the form

$$\sqrt{n} = [a_1, \overline{a_2, \ldots, a_n, 2a_1}] \quad (33)$$

(Rose 1994, p. 130). Furthermore, if D is not a SQUARE NUMBER, then the terms of the continued fraction of $\sqrt{D}$ satisfy

$$0 < a_n < 2\sqrt{D}. \quad (34)$$

In particular,

$$[\bar{a}] = \frac{a + \sqrt{a^2 + 4}}{2} \quad (35)$$

$$[1, \bar{a}] = \frac{-1 + \sqrt{1 + 4a}}{2} \quad (36)$$

$$[a, \overline{2a}] = \sqrt{a^2 + 1} \quad (37)$$

$$[\overline{ac, a}] = \frac{b + \sqrt{b^2 + 4c}}{2} \quad (38)$$

$$\begin{aligned}&[\overline{a_1,\ldots,a_n}]\\&\quad=\frac{-(q_{n-1}-p_n)+\sqrt{(q_{n-1}-p_n)^2+4q_np_{n-1}}}{2q_n}\end{aligned} \tag{39}$$

$$[a_1,\overline{b_1,\ldots,b_n}]=a_1+\frac{1}{[\overline{b_1,\ldots,b_n}]} \tag{40}$$

$$[\overline{b_1,\ldots,b_n}]=\frac{[\overline{b_1,\ldots,b_n}]p_n+p_{n-1}}{[\overline{b_1,\ldots,b_n}]q_n+q_{n-1}}. \tag{41}$$

The first follows from

$$\begin{aligned}\alpha&=n+\cfrac{1}{n+\cfrac{1}{n+\cfrac{1}{n+\ldots}}}\\&=n+\cfrac{1}{n+\left(\cfrac{1}{n+\cfrac{1}{n+\ldots}}\right)}.\end{aligned} \tag{42}$$

Therefore,

$$\alpha-n=\cfrac{1}{n+\cfrac{1}{n+\cfrac{1}{n+\ldots}}}, \tag{43}$$

so plugging (43) into (42) gives

$$\alpha=n+\frac{1}{n+(\alpha-n)}=n+\frac{1}{\alpha}. \tag{44}$$

Expanding

$$\alpha^2-n\alpha-1=0, \tag{45}$$

and solving using the QUADRATIC FORMULA gives

$$\alpha=\frac{n+\sqrt{n^2+4}}{2}. \tag{46}$$

The analog of this treatment in the general case gives

$$\alpha=\frac{\alpha p_n+p_{n-1}}{\alpha q_n+q_{n-1}}. \tag{47}$$

The following table gives the repeating simple continued fractions for the square roots of the first few integers (excluding the trivial SQUARE NUMBERS).

N	$\alpha_{\sqrt{N}}$	N	$\alpha_{\sqrt{N}}$
2	$[1,\overline{2}]$	22	$[4,\overline{1,2,4,2,1,8}]$
3	$[1,\overline{1,2}]$	23	$[4,\overline{1,3,1,8}]$
5	$[2,\overline{4}]$	24	$[4,\overline{1,8}]$
6	$[2,\overline{2,4}]$	26	$[5,\overline{10}]$
7	$[2,\overline{1,1,1,4}]$	27	$[5,\overline{5,10}]$
8	$[2,\overline{1,4}]$	28	$[5,\overline{3,2,3,10}]$
10	$[3,\overline{6}]$	29	$[5,\overline{2,1,1,2,10}]$
11	$[3,\overline{3,6}]$	30	$[5,\overline{2,10}]$
12	$[3,\overline{2,6}]$	31	$[5,\overline{1,1,3,5,3,1,1,10}]$
13	$[3,\overline{1,1,1,1,6}]$	32	$[5,\overline{1,1,1,10}]$
14	$[3,\overline{1,2,1,6}]$	33	$[5,\overline{1,2,1,10}]$
15	$[3,\overline{1,6}]$	34	$[5,\overline{1,4,1,10}]$
17	$[4,\overline{8}]$	35	$[5,\overline{1,10}]$
18	$[4,\overline{4,8}]$	37	$[6,\overline{12}]$
19	$[4,\overline{2,1,3,1,2,8}]$	38	$[6,\overline{6,12}]$
20	$[4,\overline{2,8}]$	39	$[6,\overline{4,12}]$
21	$[4,\overline{1,1,2,1,1,8}]$	40	$[6,\overline{3,12}]$

The periods of the continued fractions of the square roots of the first few nonsquare integers 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, ... (Sloane's A000037) are 1, 2, 1, 2, 4, 2, 1, 2, 2, 5, ... (Sloane's A013943; Williams 1981, Jacobson *et al.* 1995). An upper bound for the length is roughly $\mathcal{O}(\ln D\sqrt{D})$.

An even stronger result is that a continued fraction is periodic IFF it is a ROOT of a quadratic POLYNOMIAL. Calling the portion of a number x remaining after a given convergent the "tail," it must be true that the relationship between the number x and terms in its tail is of the form

$$x=\frac{ax+b}{cd+d}, \tag{48}$$

which can only lead to a QUADRATIC EQUATION.

LOGARITHMS $\log_{b_0} b_1$ can be computed by defining b_2, ... and the POSITIVE INTEGER n_1, ... such that

$${b_1}^{n_1}<b_0<{b_1}^{n_1+1} \qquad b_2=\frac{b_0}{{b_1}^{n_1}} \tag{49}$$

$${b_2}^{n_2}<b_1<{b_2}^{n_2+1} \qquad b_3=\frac{b_1}{{b_2}^{n_2}} \tag{50}$$

and so on. Then

$$\log_{b_0} b_1=[n_1,n_2,n_3,\ldots]. \tag{51}$$

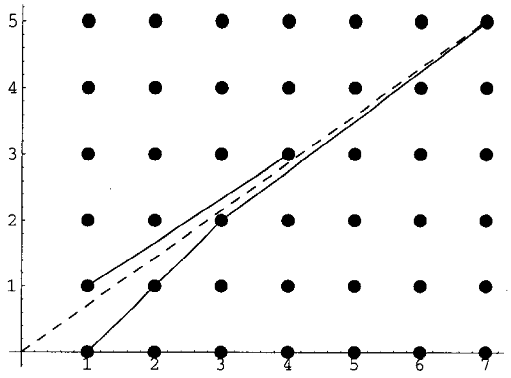

A geometric interpretation for a reduced FRACTION y/x consists of a string through a LATTICE of points with ends at $(1,0)$ and (x,y) (Klein 1907, 1932; Steinhaus 1983; Ball and Coxeter 1987, pp. 86–87; Davenport 1992). This interpretation is closely related to a similar one for the GREATEST COMMON DIVISOR. The pegs it presses against (x_i, y_i) give alternate CONVERGENTS y_i/x_i, while the other CONVERGENTS are obtained from the pegs it presses against with the initial end at $(0,1)$. The above plot is for $e-2$, which has convergents 0, 1, 2/3, 3/4, 5/7, ...

Let the continued fraction for x be written $[a_1, a_2, \ldots, a_n]$. Then the limiting value is *almost always* KHINTCHINE'S CONSTANT

$$K \equiv \lim_{n\to\infty} (a_1 a_2 \cdots a_n)^{1/n} = 2.68545\ldots. \tag{52}$$

Continued fractions can be used to express the POSITIVE ROOTS of any POLYNOMIAL equation. Continued fractions can also be used to solve linear DIOPHANTINE EQUATIONS and the PELL EQUATION. Euler showed that if a CONVERGENT SERIES can be written in the form

$$c_1 + c_1c_2 + c_1c_2c_3 + \ldots, \tag{53}$$

then it is equal to the continued fraction

$$\cfrac{c_1}{1-\cfrac{c_2}{1+c_2-\cfrac{c_3}{1+c_3-\ldots}}}. \tag{54}$$

Gosper has invented an ALGORITHM for performing analytic ADDITION, SUBTRACTION, MULTIPLICATION, and DIVISION using continued fractions. It requires keeping track of eight INTEGERS which are conceptually arranged at the VERTICES of a CUBE. The ALGORITHM has not, however, appeared in print (Gosper 1996).

An algorithm for computing the continued fraction for $(ax+b)/(cx+d)$ from the continued fraction for x is given by Beeler *et al.* (1972, Item 101), Knuth (1981, Exercise 4.5.3.15, pp. 360 and 601), and Fowler (1991). (In line 9 of Knuth's solution, $X_k \leftarrow \lfloor A/C \rfloor$ should be replaced by $X_k \leftarrow \min(\lfloor A/C \rfloor, \lfloor (A+B)/(C+D) \rfloor)$.) Beeler *et al.* (1972) and Knuth (1981) also mention the bivariate case $(axy+bx+cy+d)/(Axy+Bx+Cy+D)$.

see also GAUSSIAN BRACKETS, HURWITZ'S IRRATIONAL NUMBER THEOREM, KHINTCHINE'S CONSTANT, LAGRANGE'S CONTINUED FRACTION THEOREM, LAMÉ'S THEOREM, LÉVY CONSTANT, PADÉ APPROXIMANT, PARTIAL QUOTIENT, PI, QUADRATIC IRRATIONAL NUMBER, QUOTIENT-DIFFERENCE ALGORITHM, SEGRE'S THEOREM

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 19, 1972.

Acton, F. S. "Power Series, Continued Fractions, and Rational Approximations." Ch. 11 in *Numerical Methods That Work, 2nd printing.* Washington, DC: Math. Assoc. Amer., 1990.

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 54–57 and 86–87, 1987.

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, pp. 36–44, Feb. 1972.

Beskin, N. M. *Fascinating Fractions.* Moscow: Mir Publishers, 1980.

Brezinski, C. *History of Continued Fractions and Padé Approximants.* New York: Springer-Verlag, 1980.

Conway, J. H. and Guy, R. K. "Continued Fractions." In *The Book of Numbers.* New York: Springer-Verlag, pp. 176–179, 1996.

Courant, R. and Robbins, H. "Continued Fractions. Diophantine Equations." §2.4 in Supplement to Ch. 1 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 49–51, 1996.

Davenport, H. §IV.12 in *The Higher Arithmetic: An Introduction to the Theory of Numbers, 6th ed.* New York: Cambridge University Press, 1992.

Euler, L. *Introduction to Analysis of the Infinite, Book I.* New York: Springer-Verlag, 1980.

Fowler, D. H. *The Mathematics of Plato's Academy.* Oxford, England: Oxford University Press, 1991.

Guy, R. K. "Continued Fractions" §F20 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 259, 1994.

Jacobson, M. J. Jr.; Lukes, R. F.; and Williams, H. C. "An Investigation of Bounds for the Regulator of Quadratic Fields." *Experiment. Math.* **4**, 211–225, 1995.

Khinchin, A. Ya. *Continued Fractions.* New York: Dover, 1997.

Kimberling, C. "Continued Fractions." `http://www.evansville.edu/~ck6/integer/contfr.html`.

Klein, F. *Ausgewählte Kapitel der Zahlentheorie.* Germany: Teubner, 1907.

Klein, F. *Elementary Number Theory.* New York, p. 44, 1932.

Kline, M. *Mathematical Thought from Ancient to Modern Times.* New York: Oxford University Press, 1972.

Knuth, D. E. *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, p. 316, 1981.

Moore, C. D. *An Introduction to Continued Fractions.* Washington, DC: National Council of Teachers of Mathematics, 1964.

Olds, C. D. *Continued Fractions.* New York: Random House, 1963.

Pettofrezzo, A. J. and Bykrit, D. R. *Elements of Number Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1970.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Evaluation of Continued Fractions." §5.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 163–167, 1992.

Rose, H. E. *A Course in Number Theory, 2nd ed.* Oxford, England: Oxford University Press, 1994.

Rosen, K. H. *Elementary Number Theory and Its Applications.* New York: Addison-Wesley, 1980.

Sloane, N. J. A. Sequences A013943 and A000037/M0613 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Steinhaus, H. *Mathematical Snapshots, 3rd American ed.* New York: Oxford University Press, pp. 39–42, 1983.

Van Tuyl, A. L. "Continued Fractions." `http://www.calvin.edu/academic/math/confrac/`.
Wagon, S. "Continued Fractions." §8.5 in *Mathematica in Action.* New York: W. H. Freeman, pp. 263–271, 1991.
Wall, H. S. *Analytic Theory of Continued Fractions.* New York: Chelsea, 1948.
Williams, H. C. "A Numerical Investigation into the Length of the Period of the Continued Fraction Expansion of $\sqrt{D}$." *Math. Comp.* **36**, 593–601, 1981.

Continued Fraction Constant

A continued fraction with partial quotients which increase in ARITHMETIC PROGRESSION is

$$[A+D, A+2D, A+3D, \ldots] = \frac{I_{A/D}\left(\frac{2}{D}\right)}{I_{1+A/D}\left(\frac{2}{D}\right)},$$

where $I_n(x)$ is a MODIFIED BESSEL FUNCTION OF THE FIRST KIND (Beeler *et al.* 1972, Item 99). A special case is

$$C = 0 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{3 + \cfrac{1}{4 + \cfrac{1}{5 + \ldots}}}}},$$

which has the value

$$C = \frac{I_1(2)}{I_0(2)} = 0.697774658\ldots$$

(Lehmer 1973, Rabinowitz 1990).

see also e, GOLDEN MEAN, MODIFIED BESSEL FUNCTION OF THE FIRST KIND, PI, RABBIT CONSTANT, THUE-MORSE CONSTANT

References
Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/cntfrc/cntfrc.html`.
Guy, R. K. "Review: The Mathematics of Plato's Academy." *Amer. Math. Monthly* **97**, 440–443, 1990.
Lehmer, D. H. "Continued Fractions Containing Arithmetic Progressions." *Scripta Math.* **29**, 17–24, 1973.
Rabinowitz, S. Problem E3264. "Asymptotic Estimates from Convergents of a Continued Fraction." *Amer. Math. Monthly* **97**, 157–159, 1990.

Continued Fraction Factorization Algorithm

A PRIME FACTORIZATION ALGORITHM which uses RESIDUES produced in the CONTINUED FRACTION of $\sqrt{mN}$ for some suitably chosen m to obtain a SQUARE NUMBER. The ALGORITHM solves

$$x^2 \equiv y^2 \pmod{n}$$

by finding an m for which $m^2 \pmod{n}$ has the smallest upper bound. The method requires (by conjecture) about $\exp(\sqrt{2 \log n \log\log n})$ steps, and was the fastest PRIME FACTORIZATION ALGORITHM in use before the QUADRATIC SIEVE FACTORIZATION METHOD, which eliminates the 2 under the SQUARE ROOT (Pomerance 1996), was developed.

see also EXPONENT VECTOR, PRIME FACTORIZATION ALGORITHMS

References
Morrison, M. A. and Brillhart, J. "A Method of Factoring and the Factorization of F_7." *Math. Comput.* **29**, 183–205, 1975.
Pomerance, C. "A Tale of Two Sieves." *Not. Amer. Math. Soc.* **43**, 1473–1485, 1996.

Continued Fraction Fundamental Recurrence Relation

For a SIMPLE CONTINUED FRACTION $\sigma = [a_0, a_1, \ldots]$ with CONVERGENTS p_n/q_n, the fundamental RECURRENCE RELATION is given by

$$p_n q_{n-1} - p_{n-1} q_n = (-1)^n.$$

References
Olds, C. D. *Continued Fractions.* New York: Random House, p. 27, 1963.

Continued Fraction Map

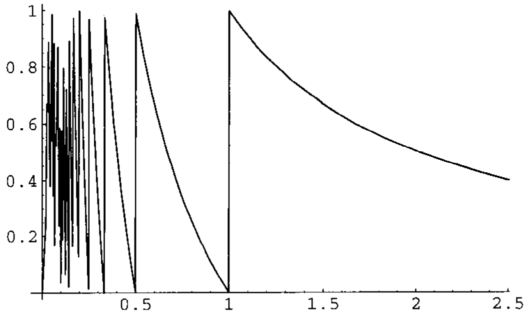

$$f(x) = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor$$

for $x \in [0, 1]$, where $\lfloor x \rfloor$ is the FLOOR FUNCTION. The INVARIANT DENSITY of the map is

$$\rho(y) = \frac{1}{(1+y)\ln 2}.$$

References
Beck, C. and Schlögl, F. *Thermodynamics of Chaotic Systems.* Cambridge, England: Cambridge University Press, pp. 194–195, 1995.

Continued Fraction Unit Fraction Algorithm

An algorithm for computing a UNIT FRACTION, called the FAREY SEQUENCE method by Bleicher (1972).

References
Bleicher, M. N. "A New Algorithm for the Expansion of Continued Fractions." *J. Number Th.* **4**, 342–382, 1972.

Continued Square Root

Expressions of the form

$$\lim_{k\to\infty} x_0+\sqrt{x_1+\sqrt{x_2+\sqrt{\ldots+x_k}}}.$$

Herschfeld (1935) proved that a continued square root of Real Nonnegative terms converges Iff $(x_n)^{2^{-n}}$ is bounded. He extended this result to arbitrary Powers (which include continued square roots and Continued Fractions as well), which is known as Herschfeld's Convergence Theorem.

see also Continued Fraction, Herschfeld's Convergence Theorem, Square Root

References

Herschfeld, A. "On Infinite Radicals." *Amer. Math. Monthly* **42**, 419–429, 1935.

Pólya, G. and Szegő, G. *Problems and Theorems in Analysis, Vol. 1.* New York: Springer-Verlag, 1997.

Sizer, W. S. "Continued Roots." *Math. Mag.* **59**, 23–27, 1986.

Continued Vector Product

see Vector Triple Product

Continuity

The property of being Continuous.

see also Continuity Axioms, Continuity Correction, Continuity Principle, Continuous Distribution, Continuous Function, Continuous Space, Fundamental Continuity Theorem

Continuity Axioms

"The" continuity axiom is an additional Axiom which must be added to those of Euclid's *Elements* in order to guarantee that two equal Circles of Radius r intersect each other if the separation of their centers is less than $2r$ (Dunham 1990). The continuity *axioms* are the three of Hilbert's Axioms which concern geometric equivalence.

Archimedes' Lemma is sometimes also known as "the continuity axiom."

see also Congruence Axioms, Hilbert's Axioms, Incidence Axioms, Ordering Axioms, Parallel Postulate

References

Dunham, W. *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, p. 38, 1990.

Hilbert, D. *The Foundations of Geometry.* Chicago, IL: Open Court, 1980.

Iyanaga, S. and Kawada, Y. (Eds.). "Hilbert's System of Axioms." §163B in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 544–545, 1980.

Continuity Correction

A correction to a discrete Binomial Distribution to approximate a continuous distribution.

$$P(a\leq X\leq b)\approx P\left(\frac{a-\frac{1}{2}-np}{\sqrt{np(1-p)}}\leq z\leq\frac{b-\frac{1}{2}-np}{\sqrt{np(1-p)}}\right),$$

where

$$z\equiv\frac{(x-\mu)}{\sigma}$$

is a continuous variate with a Normal Distribution and X is a variate of a Binomial Distribution.

see also Binomial Distribution, Normal Distribution

Continuity Principle

see Permanence of Mathematical Relations Principle

Continuous

A general mathematical property obeyed by mathematical objects in which all elements are within a Neighborhood of nearby points.

see also Absolutely Continuous, Continuous Distribution, Continuous Function, Continuous Space, Jump

Continuous Distribution

A Distribution for which the variables may take on a continuous range of values. Abramowitz and Stegun (1972, p. 930) give a table of the parameters of most common discrete distributions.

see also Beta Distribution, Bivariate Distribution, Cauchy Distribution, Chi Distribution, Chi-Squared Distribution, Correlation Coefficient, Discrete Distribution, Double Exponential Distribution, Equally Likely Outcomes Distribution, Exponential Distribution, Extreme Value Distribution, F-Distribution, Fermi-Dirac Distribution, Fisher's z-Distribution, Fisher-Tippett Distribution, Gamma Distribution, Gaussian Distribution, Half-Normal Distribution, Laplace Distribution, Lattice Distribution, Lévy Distribution, Logarithmic Distribution, Log-Series Distribution, Logistic Distribution, Lorentzian Distribution, Maxwell Distribution, Normal Distribution, Pareto Distribution, Pascal Distribution, Pearson Type III Distribution, Poisson Distribution, Pólya Distribution, Ratio Distribution, Rayleigh Distribution, Rice Distribution, Snedecor's F-Distribution, Student's t-Distribution, Student's z-Distribution, Uniform Distribution, Weibull Distribution

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 927 and 930, 1972.

Continuous Function

A continuous function is a FUNCTION $f : X \to Y$ where the pre-image of every OPEN SET in Y is OPEN in X. A function $f(x)$ in a single variable x is said to be continuous at point x_0 if

1. $f(x_0)$ is defined, so x_0 is the DOMAIN of f.
2. $\lim_{x\to x_0} f(x)$ exists.
3. $\lim_{x\to x_0} f(x) = f(x_0)$,

where lim denotes a LIMIT. If f is DIFFERENTIABLE at point x_0, then it is also continuous at x_0. If f and g are continuous at x_0, then

1. $f + g$ is continuous at x_0.
2. $f - g$ is continuous at x_0.
3. $f \times g$ is continuous at x_0.
4. $f \div g$ is continuous at x_0 if $g(x_0) \neq 0$ and is discontinuous at x_0 if $g(x_0) = 0$.
5. $f \circ g$ is continuous, where $\circ$ denotes using g as the argument to f.

see also CRITICAL POINT, DIFFERENTIABLE, LIMIT, NEIGHBORHOOD, STATIONARY POINT

Continuous Space

A TOPOLOGICAL SPACE.

see also NET

Continuum

The nondenumerable set of REAL NUMBERS, denoted C. It satisfies

$$\aleph_0 + C = C \tag{1}$$

and

$$C^r = C, \tag{2}$$

where $\aleph_0$ is $\aleph_0$ (ALEPH-0). It is also true that

$$\aleph_0{}^{\aleph_0} = C. \tag{3}$$

However,

$$C^C = F \tag{4}$$

is a SET larger than the continuum. Paradoxically, there are exactly as many points C on a LINE (or LINE SEGMENT) as in a PLANE, a 3-D SPACE, or finite HYPERSPACE, since all these SETS can be put into a ONE-TO-ONE correspondence with each other.

The CONTINUUM HYPOTHESIS, first proposed by Georg Cantor, holds that the CARDINAL NUMBER of the continuum is the same as that of $\aleph_1$. The surprising truth is that this proposition is UNDECIDABLE, since neither it nor its converse contradicts the tenets of SET THEORY.

see also ALEPH-0 ($\aleph_0$), ALEPH-1 ($\aleph_1$), CONTINUUM HYPOTHESIS, DENUMERABLE SET

Continuum Hypothesis

The proposal originally made by Georg Cantor that there is no infinite SET with a CARDINAL NUMBER between that of the "small" infinite SET of INTEGERS $\aleph_0$ and the "large" infinite set of REAL NUMBERS C (the "CONTINUUM"). Symbolically, the continuum hypothesis is that $\aleph_1 = C$. Gödel showed that no contradiction would arise if the continuum hypothesis were added to conventional ZERMELO-FRAENKEL SET THEORY. However, using a technique called FORCING, Paul Cohen (1963, 1964) proved that no contradiction would arise if the *negation* of the continuum hypothesis was added to SET THEORY. Together, Gödel's and Cohen's results established that the validity of the continuum hypothesis depends on the version of SET THEORY being used, and is therefore UNDECIDABLE (assuming the ZERMELO-FRAENKEL AXIOMS together with the AXIOM OF CHOICE).

Conway and Guy (1996) give a generalized version of the Continuum Hypothesis which is also UNDECIDABLE: is $2^{\aleph_\aleph} = \aleph_{\alpha+1}$ for every α?

see also ALEPH-0 ($\aleph_0$), ALEPH-1 ($\aleph_1$), AXIOM OF CHOICE, CARDINAL NUMBER, CONTINUUM, DENUMERABLE SET, FORCING, HILBERT'S PROBLEMS, LEBESGUE MEASURABILITY PROBLEM, UNDECIDABLE, ZERMELO-FRAENKEL AXIOMS, ZERMELO-FRAENKEL SET THEORY

References

Cohen, P. J. "The Independence of the Continuum Hypothesis." *Proc. Nat. Acad. Sci. U. S. A.* **50**, 1143–1148, 1963.

Cohen, P. J. "The Independence of the Continuum Hypothesis. II." *Proc. Nat. Acad. Sci. U. S. A.* **51**, 105–110, 1964.

Cohen, P. J. *Set Theory and the Continuum Hypothesis.* New York: W. A. Benjamin, 1966.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 282, 1996.

Gödel, K. *The Consistency of the Continuum-Hypothesis.* Princeton, NJ: Princeton University Press, 1940.

McGough, N. "The Continuum Hypothesis." `http://www.jazzie.com/ii/math/ch` and `http://www.best.com/~ii/math/ch/`.

Contour

A path in the COMPLEX PLANE over which CONTOUR INTEGRATION is performed.

see also CONTOUR INTEGRATION

Contour Integral

see CONTOUR INTEGRATION

Contour Integration

Let $P(x)$ and $Q(x)$ be POLYNOMIALS of DEGREES n and m with COEFFICIENTS $b_n, \ldots, b_0$ and $c_m, \ldots, c_0$. Take the contour in the upper half-plane, replace x by z, and write $z \equiv Re^{i\theta}$. Then

$$\int_{-\infty}^{\infty} \frac{P(z)\,dz}{Q(z)} = \lim_{R\to\infty} \int_{-R}^{R} \frac{P(z)\,dz}{Q(z)}. \tag{1}$$

Define a path γ_R which is straight along the REAL axis from $-R$ to R and makes a circular arc to connect the two ends in the upper half of the COMPLEX PLANE. The RESIDUE THEOREM then gives

$$\lim_{R\to\infty}\int_{\gamma_R}\frac{P(z)\,dz}{Q(z)}$$
$$=\lim_{R\to\infty}\int_{-R}^{R}\frac{P(z)\,dz}{Q(z)}+\lim_{R\to\infty}\int_0^\pi\frac{P(Re^{i\theta})}{Q(Re^{i\theta})}iRe^{i\theta}\,d\theta$$
$$=2\pi i\sum_{\Im[z]>0}\operatorname{Res}\left[\frac{P(z)}{Q(z)}\right],\qquad(2)$$

where Res denotes the RESIDUES. Solving,

$$\lim_{R\to\infty}\int_{-R}^{R}\frac{P(z)\,dz}{Q(z)}$$
$$=2\pi i\sum_{\Im[z]>0}\operatorname{Res}\frac{P(z)}{Q(z)}-\lim_{R\to\infty}\int_0^\pi\frac{P(Re^{i\theta})}{Q(Re^{i\theta})}iRe^{i\theta}\,d\theta.\qquad(3)$$

Define

$$I_r\equiv\lim_{R\to\infty}\int_0^\pi\frac{P(Re^{i\theta})}{Q(Re^{i\theta})}iRe^{i\theta}\,d\theta$$
$$=\lim_{R\to\infty}\int_0^\pi\frac{b_n(Re^{i\theta})^n+b_{n-1}(Re^{i\theta})^{n-1}+\ldots+b_0}{c_m(Re^{i\theta})^m+c_{m-1}(Re^{i\theta})^{m-1}+\ldots+c_0}iR\,d\theta$$
$$=\lim_{R\to\infty}\int_0^\pi\frac{b_n}{c_m}(Re^{i\theta})^{n-m}iR\,d\theta$$
$$=\lim_{R\to\infty}\int_0^\pi\frac{b_n}{c_m}R^{n+1-m}i(e^{i\theta})^{n-m}\,d\theta\qquad(4)$$

and set

$$\epsilon\equiv-(n+1-m),\qquad(5)$$

then equation (4) becomes

$$I_r\equiv\lim_{R\to\infty}\frac{i}{R^\epsilon}\frac{b_n}{c_m}\int_0^\pi e^{i(n-m)\theta}\,d\theta.\qquad(6)$$

Now,

$$\lim_{R\to\infty}R^{-\epsilon}=0\qquad(7)$$

for $\epsilon>0$. That means that for $-n-1+m\geq 1$, or $m\geq n+2$, $I_R=0$, so

$$\int_{-\infty}^{\infty}\frac{P(z)\,dz}{Q(z)}=2\pi i\sum_{\Im[z]>0}\operatorname{Res}\left[\frac{P(z)}{Q(z)}\right]\qquad(8)$$

for $m\geq n+2$. Apply JORDAN'S LEMMA with $f(x)\equiv P(x)/Q(x)$. We must have

$$\lim_{x\to\infty}f(x)=0,\qquad(9)$$

so we require $m\geq n+1$. Then

$$\int_{-\infty}^{\infty}\frac{P(z)}{Q(z)}e^{iaz}\,dz=2\pi i\sum_{\Im[z]>0}\operatorname{Res}\left[\frac{P(z)}{Q(z)}e^{iaz}\right]\qquad(10)$$

for $m\geq n+1$.

Since this must hold separately for REAL and IMAGINARY PARTS, this result can be extended to

$$\int_{-\infty}^{\infty}\frac{P(x)}{Q(x)}\cos(ax)\,dx=2\pi\Re\left\{\sum_{\Im[z]>0}\operatorname{Res}\left[\frac{P(z)}{Q(z)}e^{iaz}\right]\right\}\qquad(11)$$

$$\int_{-\infty}^{\infty}\frac{P(x)}{Q(x)}\sin(ax)\,dx=2\pi\Im\left\{\sum_{\Im[z]>0}\operatorname{Res}\left[\frac{P(z)}{Q(z)}e^{iaz}\right]\right\}.\qquad(12)$$

It is also true that

$$\int_{-\infty}^{\infty}\frac{P(z)}{Q(z)}\ln(az)\,dz=0.\qquad(13)$$

see also CAUCHY INTEGRAL FORMULA, CAUCHY INTEGRAL THEOREM, INSIDE-OUTSIDE THEOREM, JORDAN'S LEMMA, RESIDUE (COMPLEX ANALYSIS), SINE INTEGRAL

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 353–356, 1953.

Contracted Cycloid

see CURTATE CYCLOID

Contraction

see DILATION

Contraction (Graph)

The merging of nodes in a GRAPH by eliminating segments between two nodes.

Contraction (Tensor)

The contraction of a TENSOR is obtained by setting unlike indices equal and summing according to the EINSTEIN SUMMATION convention. Contraction reduces the RANK of a TENSOR by 2. For a second RANK TENSOR,

$$\operatorname{contr}(B'^{i}_{j})\equiv B'^{i}_{i}$$

$$B'^{i}_{i}=\frac{\partial x'_i}{\partial x_k}\frac{\partial x_l}{\partial x'_i}B^k_l=\frac{\partial x_l}{\partial x_k}B^k_l=\delta^l_kB^k_l=B^k_k.$$

Therefore, the contraction is invariant, and must be a SCALAR. In fact, this SCALAR is known as the TRACE of a MATRIX in MATRIX theory.

References

Arfken, G. "Contraction, Direct Product." §3.2 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 124–126, 1985.

Contradiction Law

No A is not-A.

see also NOT

Contravariant Tensor

A contravariant tensor is a TENSOR having specific transformation properties (c.f., a COVARIANT TENSOR). To examine the transformation properties of a contravariant tensor, first consider a TENSOR of RANK 1 (a VECTOR)

$$d\mathbf{r} = dx_1\hat{\mathbf{x}}_1 + dx_2\hat{\mathbf{x}}_2 + dx_3\hat{\mathbf{x}}_3, \tag{1}$$

for which

$$dx_i' = \frac{\partial x_i'}{\partial x_j} dx_j. \tag{2}$$

Now let $A_i \equiv dx_i$, then any set of quantities A_j which transform according to

$$A_i' = \frac{\partial x_i'}{\partial x_j} A_j, \tag{3}$$

or, defining

$$a_{ij} \equiv \frac{\partial x_i'}{\partial x_j}, \tag{4}$$

according to

$$A_i' = a_{ij} A_j \tag{5}$$

is a contravariant tensor. Contravariant tensors are indicated with raised indices, i.e., a^μ.

COVARIANT TENSORS are a type of TENSOR with differing transformation properties, denoted a_ν. However, in 3-D CARTESIAN COORDINATES,

$$\frac{\partial x_j}{\partial x_i'} = \frac{\partial x_i'}{\partial x_j} \equiv a_{ij} \tag{6}$$

for $i, j = 1, 2, 3$, meaning that contravariant and covariant tensors are equivalent. The two types of tensors do differ in higher dimensions, however. Contravariant FOUR-VECTORS satisfy

$$a^\mu = \Lambda^\mu_\nu a^\nu, \tag{7}$$

where Λ is a LORENTZ TENSOR.

To turn a COVARIANT TENSOR into a contravariant tensor, use the METRIC TENSOR $g^{\mu\nu}$ to write

$$a^\mu \equiv g^{\mu\nu} a_\nu. \tag{8}$$

Covariant and contravariant indices can be used simultaneously in a MIXED TENSOR.

see also COVARIANT TENSOR, FOUR-VECTOR, LORENTZ TENSOR, METRIC TENSOR, MIXED TENSOR, TENSOR

References

Arfken, G. "Noncartesian Tensors, Covariant Differentiation." §3.8 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 158–164, 1985.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 44–46, 1953.

Contravariant Vector

A CONTRAVARIANT TENSOR of RANK 1.

see also CONTRAVARIANT TENSOR, VECTOR

Control Theory

The mathematical study of how to manipulate the parameters affecting the behavior of a system to produce the desired or optimal outcome.

see also KALMAN FILTER, LINEAR ALGEBRA, PONTRYAGIN MAXIMUM PRINCIPLE

References

Zabczyk, J. *Mathematical Control Theory: An Introduction.* Boston, MA: Birkhäuser, 1993.

Convective Acceleration

The acceleration of an element of fluid, given by the CONVECTIVE DERIVATIVE of the VELOCITY $\mathbf{v}$,

$$\frac{D\mathbf{v}}{Dt} = \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v},$$

where ∇ is the GRADIENT operator.

see also ACCELERATION, CONVECTIVE DERIVATIVE, CONVECTIVE OPERATOR

References

Batchelor, G K. *An Introduction to Fluid Dynamics.* Cambridge, England: Cambridge University Press, p. 73, 1977.

Convective Derivative

A DERIVATIVE taken with respect to a moving coordinate system, also called a LAGRANGIAN DERIVATIVE. It is given by

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla,$$

where ∇ is the GRADIENT operator and $\mathbf{v}$ is the VELOCITY of the fluid. This type of derivative is especially useful in the study of fluid mechanics. When applied to $\mathbf{v}$,

$$\frac{D\mathbf{v}}{Dt} = \frac{\partial \mathbf{v}}{\partial t} + (\nabla \times \mathbf{v}) \times \mathbf{v} + \nabla(\tfrac{1}{2}\mathbf{v}^2).$$

see also CONVECTIVE OPERATOR, DERIVATIVE, VELOCITY

References

Batchelor, G K. *An Introduction to Fluid Dynamics.* Cambridge, England: Cambridge University Press, p. 73, 1977.

Convective Operator

Defined for a VECTOR FIELD $\mathbf{A}$ by $(\mathbf{A} \cdot \nabla)$, where ∇ is the GRADIENT operator.

Applied in arbitrary orthogonal 3-D coordinates to a VECTOR FIELD $\mathbf{B}$, the convective operator becomes

$$[(\mathbf{A} \cdot \nabla)\mathbf{B}]_j = \sum_{k=1}^{3} \left[\frac{A_k}{h_k} \frac{\partial B_j}{\partial q_k} + \frac{B_k}{h_k h_j} \left(A_j \frac{\partial h_j}{\partial q_k} - A_k \frac{\partial h_k}{\partial q_j} \right) \right], \tag{1}$$

where the h_is are related to the METRIC TENSORS by $h_i = \sqrt{g_{ii}}$. In CARTESIAN COORDINATES,

$$\begin{aligned}(\mathbf{A}\cdot\nabla)\mathbf{B} &= \left(A_x\frac{\partial B_x}{\partial x} + A_y\frac{\partial B_x}{\partial y} + A_z\frac{\partial B_x}{\partial z}\right)\hat{\mathbf{x}}\\ &+ \left(A_x\frac{\partial B_y}{\partial x} + A_y\frac{\partial B_y}{\partial y} + A_z\frac{\partial B_y}{\partial z}\right)\hat{\mathbf{y}}\\ &+ \left(A_x\frac{\partial B_z}{\partial x} + A_y\frac{\partial B_z}{\partial y} + A_z\frac{\partial B_z}{\partial z}\right)\hat{\mathbf{z}}. \quad (2)\end{aligned}$$

In CYLINDRICAL COORDINATES,

$$\begin{aligned}(\mathbf{A}\cdot\nabla)\mathbf{B} &= \left(A_r\frac{\partial B_r}{\partial r} + \frac{A_\phi}{r}\frac{\partial B_r}{\partial \phi} + A_z\frac{\partial B_r}{\partial z} - \frac{A_\phi B_\phi}{r}\right)\hat{\mathbf{r}}\\ &+ \left(A_r\frac{\partial B_\phi}{\partial r} + \frac{A_\phi}{r}\frac{\partial B_\phi}{\partial \phi} + A_z\frac{\partial B_\phi}{\partial z} + \frac{A_\phi B_r}{r}\right)\hat{\boldsymbol{\phi}}\\ &+ \left(A_r\frac{\partial B_z}{\partial r} + \frac{A_\phi}{r}\frac{\partial B_z}{\partial \phi} + A_z\frac{\partial B_z}{\partial z}\right)\hat{\mathbf{z}}. \quad (3)\end{aligned}$$

In SPHERICAL COORDINATES,

$$\begin{aligned}&(\mathbf{A}\cdot\nabla)\mathbf{B}\\ &= \left(A_r\frac{\partial B_r}{\partial r} + \frac{A_\theta}{r}\frac{\partial B_r}{\partial \theta} + \frac{A_\phi}{r\sin\theta}\frac{\partial B_r}{\partial \phi} - \frac{A_\theta B_\theta + A_\phi B_\phi}{r}\right)\hat{\mathbf{r}}\\ &+ \left(A_r\frac{\partial B_\theta}{\partial r} + \frac{A_\theta}{r}\frac{\partial B_\theta}{\partial \theta} + \frac{A_\phi}{r\sin\theta}\frac{\partial B_\theta}{\partial \phi} + \frac{A_\theta B_r}{r} - \frac{A_\phi B_\phi \cot\theta}{r}\right)\hat{\boldsymbol{\theta}}\\ &+ \left(A_r\frac{\partial B_\phi}{\partial r} + \frac{A_\theta}{r}\frac{\partial B_\phi}{\partial \theta} + \frac{A_\phi}{r\sin\theta}\frac{\partial B_\phi}{\partial \phi} + \frac{A_\phi B_r}{r} + \frac{A_\phi B_\theta \cot\theta}{r}\right)\hat{\boldsymbol{\phi}}. \quad (4)\end{aligned}$$

see also CONVECTIVE ACCELERATION, CONVECTIVE DERIVATIVE, CURVILINEAR COORDINATES, GRADIENT

Convergence Acceleration

see CONVERGENCE IMPROVEMENT

Convergence Improvement

The improvement of the convergence properties of a SERIES, also called CONVERGENCE ACCELERATION, such that a SERIES reaches its limit to within some accuracy with fewer terms than required before. Convergence improvement can be effected by forming a linear combination with a SERIES whose sum is known. Useful sums include

$$\sum_{n=1}^\infty \frac{1}{n(n+1)} = 1 \quad (1)$$

$$\sum_{n=1}^\infty \frac{1}{n(n+1)(n+2)} = \frac{1}{4} \quad (2)$$

$$\sum_{n=1}^\infty \frac{1}{n(n+1)(n+2)(n+3)} = \frac{1}{18} \quad (3)$$

$$\sum_{n=1}^\infty \frac{1}{n(n+1)\cdots(n+p)} = \frac{1}{p\cdot p!}. \quad (4)$$

Kummer's transformation takes a convergent series

$$s = \sum_{k=0}^\infty a_k \quad (5)$$

and another convergent series

$$c = \sum_{k=0}^\infty c_k \quad (6)$$

with known c such that

$$\lim_{k\to\infty} \frac{a_k}{c_k} = \lambda \neq 0. \quad (7)$$

Then a series with more rapid convergence to the same value is given by

$$s = \lambda c + \sum_{k=0}^\infty \left(1 - \lambda\frac{c_k}{a_k}\right) a_k \quad (8)$$

(Abramowitz and Stegun 1972).

EULER'S TRANSFORM takes a convergent alternating series

$$\sum_{k=0}^\infty (-1)^k a_k = a_0 - a_1 + a_2 - \ldots \quad (9)$$

into a series with more rapid convergence to the same value to

$$s = \sum_{k=0}^\infty \frac{(-1)^k \Delta^k a_0}{2^{k+1}}, \quad (10)$$

where

$$\Delta^k a_0 = \sum_{m=0}^k \equiv (-1)^m \binom{k}{m} a_{k-m} \quad (11)$$

(Abramowitz and Stegun 1972; Beeler *et al.* 1972, Item 120).

Given a series of the form

$$S = \sum_{n=1}^\infty f\left(\frac{1}{n}\right), \quad (12)$$

where $f(z)$ is an ANALYTIC at 0 and on the closed unit DISK, and

$$f(z)|_{z\to 0} = \mathcal{O}(z^2), \quad (13)$$

then the series can be rearranged to

$$\begin{aligned}S &= \sum_{n=1}^\infty \sum_{m=2}^\infty f_m \left(\frac{1}{n}\right)^m\\ &= \sum_{m=2}^\infty \sum_{n=1}^\infty f_m \left(\frac{1}{n}\right)^m = \sum_{m=2}^\infty f_m \zeta(m), \quad (14)\end{aligned}$$

where

$$f(z) = \sum_{m=2}^{\infty} f_m z^m \tag{15}$$

is the MACLAURIN SERIES of f and $\zeta(z)$ is the RIEMANN ZETA FUNCTION (Flajolet and Vardi 1996). The transformed series exhibits geometric convergence. Similarly, if $f(z)$ is ANALYTIC in $|z| \leq 1/n_0$ for some POSITIVE INTEGER n_0, then

$$S = \sum_{n=1}^{n_0-1} f\left(\frac{1}{n}\right) + \sum_{m=2}^{\infty} f_m \left[\zeta(m) - \frac{1}{1^m} - \ldots - \frac{1}{(n_0-1)^m}\right], \tag{16}$$

which converges geometrically (Flajolet and Vardi 1996). (16) can also be used to further accelerate the convergence of series (14).

see also EULER'S TRANSFORM, WILF-ZEILBERGER PAIR

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 16, 1972.

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 288–289, 1985.

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Flajolet, P. and Vardi, I. "Zeta Function Expansions of Classical Constants." Unpublished manuscript. 1996. `http://pauillac.inria.fr/algo/flajolet/Publications/landau.ps`.

Convergence Tests

A test to determine if a given SERIES CONVERGES or DIVERGES.

see also ABEL'S UNIFORM CONVERGENCE TEST, BERTRAND'S TEST, D'ALEMBERT RATIO TEST, DIVERGENCE TESTS, ERMAKOFF'S TEST, GAUSS'S TEST, INTEGRAL TEST, KUMMER'S TEST, RAABE'S TEST, RATIO TEST, RIEMANN SERIES THEOREM, ROOT TEST

References

Arfken, G. "Convergence Tests." §5.2 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 280–293, 1985.

Bromwich, T. J. I'a and MacRobert, T. M. *An Introduction to the Theory of Infinite Series, 3rd ed.* New York: Chelsea, pp. 55–57, 1991.

Convergent

The RATIONAL NUMBER obtained by keeping only a limited number of terms in a CONTINUED FRACTION is called a convergent. For example, in the SIMPLE CONTINUED FRACTION for the GOLDEN RATIO,

$$\phi = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \ldots}},$$

the convergents are

$$1, 1 + \frac{1}{1} = \frac{3}{2}, 1 + \frac{1}{1 + \frac{1}{1}} = \frac{5}{3}, \ldots.$$

The word convergent is also used to describe a CONVERGENT SEQUENCE or CONVERGENT SERIES.

see also CONTINUED FRACTION, CONVERGENT SEQUENCE, CONVERGENT SERIES, PARTIAL QUOTIENT, SIMPLE CONTINUED FRACTION

Convergent Sequence

A SEQUENCE S_n converges to the limit S

$$\lim_{n\to\infty} S_n = S$$

if, for any $\epsilon > 0$, there exists an N such that $|S_n - S| < \epsilon$ for $n > N$. If S_n does not converge, it is said to DIVERGE. Every bounded MONOTONIC SEQUENCE converges. Every unbounded SEQUENCE diverges. This condition can also be written as

$$\overline{\lim_{n\to\infty}} S_n = \underline{\lim_{n\to\infty}} S_n = S.$$

see also CONDITIONAL CONVERGENCE, STRONG CONVERGENCE, WEAK CONVERGENCE

Convergent Series

The infinite SERIES $\sum_{n=1}^{\infty} a_n$ is convergent if the SEQUENCE of partial sums

$$S_n = \sum_{k=1}^{n} a_k$$

is convergent. Conversely, a SERIES is divergent if the SEQUENCE of partial sums is divergent. If $\sum u_k$ and $\sum v_k$ are convergent SERIES, then $\sum (u_k + v_k)$ and $\sum (u_k - v_k)$ are convergent. If $c \neq 0$, then $\sum u_k$ and $c \sum u_k$ both converge or both diverge. Convergence and divergence are unaffected by deleting a finite number of terms from the beginning of a series. Constant terms in the denominator of a sequence can usually be deleted without affecting convergence. All but the highest POWER terms in POLYNOMIALS can usually be deleted in both NUMERATOR and DENOMINATOR of a SERIES without affecting convergence. If a SERIES converges absolutely, then it converges.

see also CONVERGENCE TESTS, RADIUS OF CONVERGENCE

References

Bromwich, T. J. I'a. and MacRobert, T. M. *An Introduction to the Theory of Infinite Series, 3rd ed.* New York: Chelsea, 1991.

Conversion Period

The period of time between Interest payments.

see also Compound Interest, Interest, Simple Interest

Convex

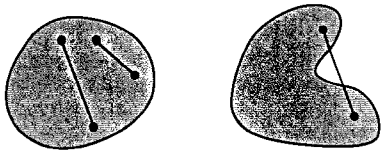

convex *concave*

A Set in Euclidean Space $\mathbb{R}^d$ is convex if it contains *all* the Line Segments connecting any pair of its points. If the Set does not contain all the Line Segments, it is called Concave.

see also Connected Set, Convex Function, Convex Hull, Convex Optimization Theory, Convex Polygon, Delaunay Triangulation, Minkowski Convex Body Theorem, Simply Connected

References

Croft, H. T.; Falconer, K. J.; and Guy, R. K. "Convexity." Ch. A in *Unsolved Problems in Geometry.* New York: Springer-Verlag, pp. 6–47, 1994.

Convex Function

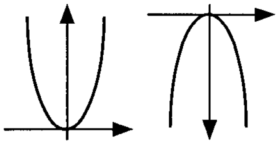

concave up *concave down*

A function whose value at the Midpoint of every Interval in its Domain does not exceed the Average of its values at the ends of the Interval. In other words, a function $f(x)$ is convex on an Interval $[a,b]$ if for any two points x_1 and x_2 in $[a,b]$,

$$f[\tfrac{1}{2}(x_1+x_2)] \leq \tfrac{1}{2}[f(x_1)+f(x_2)].$$

If $f(x)$ has a second Derivative in $[a,b]$, then a Necessary and Sufficient condition for it to be convex on that Interval is that the second Derivative $f''(x) > 0$ for all x in $[a,b]$.

see also Concave Function, Logarithmically Convex Function

References

Eggleton, R. B. and Guy, R. K. "Catalan Strikes Again! How Likely is a Function to be Convex?" *Math. Mag.* **61**, 211–219, 1988.

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1100, 1980.

Convex Hull

The convex hull of a set of points S is the Intersection of all convex sets containing S. For N points $p_1, \ldots, p_N$, the convex hull C is then given by the expression

$$C \equiv \left\{ \sum_{j=1}^{N} \lambda_j p_j : \lambda_j \geq 0 \text{ for all } j \text{ and } \sum_{j=1}^{N} \lambda_j = 1 \right\}.$$

see also Carathéodory's Fundamental Theorem, Cross Polytope, Groemer Packing, Groemer Theorem, Sausage Conjecture, Sylvester's Four-Point Problem

References

Santaló, L. A. *Integral Geometry and Geometric Probability.* Reading, MA: Addison-Wesley, 1976.

Convex Optimization Theory

The problem of maximizing a linear function over a convex polyhedron, also known as Operations Research or Optimization Theory. The general problem of convex optimization is to find the minimum of a convex (or quasiconvex) function f on a Finite-dimensional convex body A. Methods of solution include Levin's algorithm and the method of circumscribed Ellipsoids, also called the Nemirovsky-Yudin-Shor method.

References

Tokhomirov, V. M. "The Evolution of Methods of Convex Optimization." *Amer. Math. Monthly* **103**, 65–71, 1996.

Convex Polygon

A Polygon is Convex if it contains *all* the Line Segments connecting any pair of its points. Let $f(n)$ be the smallest number such that when W is a set of more than $f(n)$ points in General Position (with no three points Collinear) in the plane, all of the Vertices of some convex n-gon are contained in W. The answers for $n = 2$, 3, and 4 are 2, 4, and 8. It is conjectured that $f(n) = 2^{n-2}$, but only proven that

$$2^{n-2} \leq f(n) \leq \binom{2n-4}{n-2},$$

where $\binom{n}{k}$ is a Binomial Coefficient.

Convex Polyhedron

A Polyhedron for which a line connecting any two (noncoplanar) points on the surface always lies in the interior of the polyhedron. The 92 convex polyhedra having only Regular Polygons as faces are called the Johnson Solids, which include the Platonic Solids and Archimedean Solids. No method is known for computing the Volume of a general convex polyhedron (Ogilvy 1990, p. 173).

see also Archimedean Solid, Deltahedron, Johnson Solid, Kepler-Poinsot Solid, Platonic Solid, Regular Polygon

References

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, 1990.

Convolution

A convolution is an integral which expresses the amount of overlap of one function $g(t)$ as it is shifted over another function $f(t)$. It therefore "blends" one function with another. For example, in synthesis imaging, the measured DIRTY MAP is a convolution of the "true" CLEAN MAP with the DIRTY BEAM (the FOURIER TRANSFORM of the sampling distribution). The convolution is sometimes also known by its German name, FALTUNG ("folding"). A convolution over a finite range $[0, t]$ is given by

$$f(t) * g(t) \equiv \int_0^t f(\tau) g(t-\tau)\, d\tau, \tag{1}$$

where the symbol $f*g$ (occasionally also written as $f \otimes g$) denotes convolution of f and g. Convolution is more often taken over an infinite range,

$$f(t)*g(t) \equiv \int_{-\infty}^{\infty} f(\tau)g(t-\tau)\,d\tau = \int_{-\infty}^{\infty} g(\tau)f(t-\tau)\,d\tau. \tag{2}$$

Let f, g, and h be arbitrary functions and a a constant. Convolution has the following properties:

$$f * g = g * f \tag{3}$$

$$f * (g * h) = (f * g) * h \tag{4}$$

$$f * (g + h) = (f * g) + (f * h) \tag{5}$$

$$a(f * g) = (af) * g = f * (ag). \tag{6}$$

The INTEGRAL identity

$$\int_a^x \int_a^x f(t)\,dt\,dx = \int_a^x (x-t) f(t)\,dt \tag{7}$$

also gives a convolution. Taking the DERIVATIVE of a convolution gives

$$\frac{d}{dx}(f * g) = \frac{df}{dx} * g = f * \frac{dg}{dx}. \tag{8}$$

The AREA under a convolution is the product of areas under the factors,

$$\begin{aligned}\int_{-\infty}^{\infty} (f*g)\,dx &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(u)g(x-u)\,du\right] dx \\ &= \int_{-\infty}^{\infty} f(u)\left[\int_{-\infty}^{\infty} g(x-u)\,dx\right] du \\ &= \left[\int_{-\infty}^{\infty} f(u)\,du\right]\left[\int_{-\infty}^{\infty} g(x)\,dx\right].\end{aligned} \tag{9}$$

The horizontal CENTROIDS add

$$\int_{-\infty}^{\infty} \langle x(f*g)\rangle\, dx = \langle xf\rangle + \langle xg\rangle, \tag{10}$$

as do the VARIANCES

$$\int_{-\infty}^{\infty} \langle x^2(f*g)\rangle\, dx = \langle x^2 f\rangle + \langle x^2 g\rangle, \tag{11}$$

where

$$\langle x^n f\rangle \equiv \frac{\int_{-\infty}^{\infty} x^n f(x)\,dx}{\int_{-\infty}^{\infty} f(x)\,dx}. \tag{12}$$

see also AUTOCORRELATION, CONVOLUTION THEOREM, CROSS-CORRELATION, WIENER-KHINTCHINE THEOREM

References

Bracewell, R. "Convolution." Ch. 3 in *The Fourier Transform and Its Applications.* New York: McGraw-Hill, pp. 25–50, 1965.

Hirschman, I. I. and Widder, D. V. *The Convolution Transform.* Princeton, NJ: Princeton University Press, 1955.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 464–465, 1953.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Convolution and Deconvolution Using the FFT." §13.1 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 531–537, 1992.

Convolution Theorem

Let $f(t)$ and $f(t)$ be arbitrary functions of time t with FOURIER TRANSFORMS. Take

$$f(t) = \mathcal{F}^{-1}[F(\nu)] = \int_{-\infty}^{\infty} F(\nu) e^{2\pi i\nu t}\,d\nu \tag{1}$$

$$g(t) = \mathcal{F}^{-1}[G(\nu)] = \int_{-\infty}^{\infty} G(\nu) e^{2\pi i\nu t}\,d\nu, \tag{2}$$

where $\mathcal{F}^{-1}$ denotes the inverse FOURIER TRANSFORM (where the transform pair is defined to have constants $A = 1$ and $B = -2\pi$). Then the CONVOLUTION is

$$\begin{aligned} f * g &\equiv \int_{-\infty}^{\infty} g(t')f(t-t')dt' \\ &= \int_{-\infty}^{\infty} g(t')\left[\int_{-\infty}^{\infty} F(\nu)e^{2\pi i\nu(t-t')}\,d\nu\right] dt'.\end{aligned} \tag{3}$$

Interchange the order of integration,

$$\begin{aligned} f * g &= \int_{-\infty}^{\infty} F(\nu)\left[\int_{-\infty}^{\infty} g(t')e^{-2\pi i\nu t'}\,dt'\right] e^{2\pi i\nu t}\,d\nu \\ &= \int_{-\infty}^{\infty} F(\nu)G(\nu)e^{2\pi i\nu t}\,d\nu \\ &= \mathcal{F}^{-1}[F(\nu)G(\nu)].\end{aligned} \tag{4}$$

So, applying a FOURIER TRANSFORM to each side, we have

$$\mathcal{F}[f*g] = \mathcal{F}[f]\mathcal{F}[g]. \tag{5}$$

The convolution theorem also takes the alternate forms

$$\mathcal{F}[fg] = \mathcal{F}[f] * \mathcal{F}[g] \tag{6}$$

$$\mathcal{F}(\mathcal{F}[f]\mathcal{F}[g]) = f * g \tag{7}$$

$$\mathcal{F}(\mathcal{F}[f] * \mathcal{F}[g]) = fg. \tag{8}$$

see also AUTOCORRELATION, CONVOLUTION, FOURIER TRANSFORM, WIENER-KHINTCHINE THEOREM

References

Arfken, G. "Convolution Theorem." §15.5 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 810–814, 1985.

Bracewell, R. "Convolution Theorem." *The Fourier Transform and Its Applications.* New York: McGraw-Hill, pp. 108–112, 1965.

Conway-Alexander Polynomial

see ALEXANDER POLYNOMIAL

Conway's Constant

The constant

$$\lambda = 1.303577269034296\ldots$$

(Sloane's A014715) giving the asymptotic rate of growth $C\lambda^k$ of the number of DIGITS in the kth term of the LOOK AND SAY SEQUENCE. λ is given by the largest ROOT of the POLYNOMIAL

$$\begin{aligned}0 = {} & x^{71} \\ & - x^{69} - 2x^{68} - x^{67} + 2x^{66} + 2x^{65} + x^{64} - x^{63} - x^{62} - x^{61} \\ & - x^{60} - x^{59} + 2x^{58} + 5x^{57} + 3x^{56} - 2x^{55} - 10x^{54} \\ & - 3x^{53} - 2x^{52} + 6x^{51} + 6x^{50} + x^{49} + 9x^{48} - 3x^{47} \\ & - 7x^{46} - 8x^{45} - 8x^{44} + 10x^{43} + 6x^{42} + 8x^{41} - 4x^{40} \\ & - 12x^{39} + 7x^{38} - 7x^{37} + 7x^{36} + x^{35} - 3x^{34} + 10x^{33} \\ & + x^{32} - 6x^{31} - 2x^{30} - 10x^{29} - 3x^{28} + 2x^{27} + 9x^{26} \\ & - 3x^{25} + 14x^{24} - 8x^{23} - 7x^{21} + 9x^{20} - 3x^{19} - 4x^{18} \\ & - 10x^{17} - 7x^{16} + 12x^{15} + 7x^{14} + 2x^{13} - 12x^{12} \\ & - 4x^{11} - 2x^{10} - 5x^{9} + x^{7} - 7x^{6} \\ & + 7x^{5} - 4x^{4} + 12x^{3} - 6x^{2} + 3x - 6.\end{aligned}$$

The POLYNOMIAL given in Conway (1987, p. 188) contains a misprint. The CONTINUED FRACTION for λ is 1, 3, 3, 2, 2, 54, 5, 2, 1, 16, 1, 30, 1, 1, 1, 2, 2, 1, 14, 1, ... (Sloane's A014967).

see also CONWAY SEQUENCE, COSMOLOGICAL THEOREM, LOOK AND SAY SEQUENCE

References

Conway, J. H. "The Weird and Wonderful Chemistry of Audioactive Decay." §5.11 in *Open Problems in Communications and Computation* (Ed. T. M. Cover and B. Gopinath). New York: Springer-Verlag, pp. 173–188, 1987.

Conway, J. H. and Guy, R. K. "The Look and Say Sequence." In *The Book of Numbers.* New York: Springer-Verlag, pp. 208–209, 1996.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/cnwy/cnwy.html`.

Sloane, N. J. A. Sequence A014967 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 13–14, 1991.

Conway's Game of Life

see LIFE

Conway Groups

The AUTOMORPHISM GROUP Co_1 of the LEECH LATTICE modulo a center of order two is called "the" Conway group. There are 15 exceptional CONJUGACY CLASSES of the Conway group. This group, combined with the GROUPS Co_2 and Co_3 obtained similarly from the LEECH LATTICE by stabilization of the 1-D and 2-D sublattices, are collectively called Conway groups. The Conway groups are SPORADIC GROUPS.

see also LEECH LATTICE, SPORADIC GROUP

References

Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas/Co1.html`, `Co2.html`, `Co3.html`.

Conway's Knot

The KNOT with BRAID WORD

$${\sigma_2}^3\sigma_1{\sigma_3}^{-1}{\sigma_2}^{-2}\sigma_1{\sigma_2}^{-1}\sigma_1{\sigma_3}^{-1}.$$

The JONES POLYNOMIAL of Conway's knot is

$$t^{-4}(-1 + 2t - 2t^2 + 2t^3 + t^6 - 2t^7 + 2t^8 - 2t^9 + t^{10}),$$

the *same* as for the KINOSHITA-TERASAKA KNOT.

Conway's Knot Notation

A concise NOTATION based on the concept of the TANGLE used by Conway (1967) to enumerate KNOTS up to 11 crossings. An ALGEBRAIC KNOT containing no NEGATIVE signs in its Conway knot NOTATION is an ALTERNATING KNOT.

References

Conway, J. H. "An Enumeration of Knots and Links, and Some of Their Algebraic Properties." In *Computation Problems in Abstract Algebra* (Ed. J. Leech). Oxford, England: Pergamon Press, pp. 329–358, 1967.

Conway's Life

see LIFE

Conway Notation

see CONWAY'S KNOT NOTATION, CONWAY POLYHEDRON NOTATION

Conway Polyhedron Notation

A NOTATION for POLYHEDRA which begins by specifying a "seed" polyhedron using a capital letter. The PLATONIC SOLIDS are denoted T (TETRAHEDRON), O (OCTAHEDRON), C (CUBE), I (ICOSAHEDRON), and D (DODECAHEDRON), according to their first letter. Other polyhedra include the PRISMS, Pn, ANTIPRISMS, An, and PYRAMIDS, Yn, where $n \geq 3$ specifies the number of sides of the polyhedron's base.

Operations to be performed on the polyhedron are then specified with lower-case letters preceding the capital letter.

see also POLYHEDRON, SCHLÄFLI SYMBOL, WYTHOFF SYMBOL

References

Hart, G. "Conway Notation for Polyhedra." `http://www.li.net/~george/virtual-polyhedra/conway_notation.html`.

Conway Polynomial

see ALEXANDER POLYNOMIAL

Conway Puzzle

Construct a $5 \times 5 \times 5$ cube from 13 $1 \times 2 \times 4$ blocks, 1 $2 \times 2 \times 2$ block, 1 $1 \times 2 \times 2$ and 3 $1 \times 1 \times 3$ blocks.

see also BOX-PACKING THEOREM, CUBE DISSECTION, DE BRUIJN'S THEOREM, KLARNER'S THEOREM, POLYCUBE, SLOTHOUBER-GRAATSMA PUZZLE

References

Honsberger, R. *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., pp. 77–80, 1976.

Conway Sequence

The LOOK AND SAY SEQUENCE generated from a starting DIGIT of 3, as given by Vardi (1991).

see also CONWAY'S CONSTANT, LOOK AND SAY SEQUENCE

References

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 13–14, 1991.

Conway Sphere

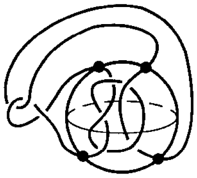

A sphere with four punctures occurring where a KNOT passes through the surface.

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, p. 94, 1994.

Coordinate Geometry

see ANALYTIC GEOMETRY

Coordinate System

A system of COORDINATES.

Coordinates

A set of n variables which fix a geometric object. If the coordinates are distances measured along PERPENDICULAR axes, they are known as CARTESIAN COORDINATES. The study of GEOMETRY using one or more coordinate systems is known as ANALYTIC GEOMETRY.

see also AREAL COORDINATES, BARYCENTRIC COORDINATES, BIPOLAR COORDINATES, BIPOLAR CYLINDRICAL COORDINATES, BISPHERICAL COORDINATES, CARTESIAN COORDINATES, CHOW COORDINATES, CIRCULAR CYLINDRICAL COORDINATES, CONFOCAL ELLIPSOIDAL COORDINATES, CONFOCAL PARABOLOIDAL COORDINATES, CONICAL COORDINATES, CURVILINEAR COORDINATES, CYCLIDIC COORDINATES, CYLINDRICAL COORDINATES, ELLIPSOIDAL COORDINATES, ELLIPTIC CYLINDRICAL COORDINATES, GAUSSIAN COORDINATE SYSTEM, GRASSMANN COORDINATES, HARMONIC COORDINATES, HOMOGENEOUS COORDINATES, OBLATE SPHEROIDAL COORDINATES, ORTHOCENTRIC COORDINATES, PARABOLIC COORDINATES, PARABOLIC CYLINDRICAL COORDINATES, PARABOLOIDAL COORDINATES, PEDAL COORDINATES, POLAR COORDINATES, PROLATE SPHEROIDAL COORDINATES, QUADRIPLANAR COORDINATES, RECTANGULAR COORDINATES, SPHERICAL COORDINATES, TOROIDAL COORDINATES, TRILINEAR COORDINATES

References

Arfken, G. "Coordinate Systems." Ch. 2 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 85–117, 1985.
Woods, F. S. *Higher Geometry: An Introduction to Advanced Methods in Analytic Geometry.* New York: Dover, p. 1, 1961.

Coordination Number

see KISSING NUMBER

Copeland-Erdős Constant

The decimal 0.23571113171923... (Sloane's A033308) obtained by concatenating the PRIMES: 2, 23, 235, 2357, 235711, ... (Sloane's A033308; one of the SMARANDACHE SEQUENCES). In 1945, Copeland and Erdős showed that it is a NORMAL NUMBER. The first few digits of the CONTINUED FRACTION of the Copeland-Erdős are 0, 4, 4, 8, 16, 18, 5, 1, ... (Sloane's A030168). The positions of the first occurrence of n in the CONTINUED FRACTION are 8, 16, 20, 2, 7, 15, 12, 4, 17, 254, ... (Sloane's A033309). The incrementally largest terms are 1, 27, 154, 1601, 2135, ... (Sloane's A033310), which occur at positions 2, 5, 11, 19, 1801, ... (Sloane's A033311).

see also CHAMPERNOWNE CONSTANT, PRIME NUMBER

References

Sloane, N. J. A. Sequences A030168, A033308, A033309, A033310, and A033311 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Coplanar

Three noncollinear points determine a plane and so are trivially coplanar. Four points are coplanar IFF the volume of the TETRAHEDRON defined by them is 0,

$$\begin{vmatrix} x_1 & y_1 & z_1 & 0 \\ x_2 & y_2 & z_2 & 0 \\ x_3 & y_3 & z_3 & 0 \\ x_4 & y_4 & z_4 & 0 \end{vmatrix}.$$

Coprime

see RELATIVELY PRIME

Copson-de Bruijn Constant

see DE BRUIJN CONSTANT

Copson's Inequality

Let $\{a_n\}$ be a NONNEGATIVE SEQUENCE and $f(x)$ a NONNEGATIVE integrable function. Define

$$A_n = \sum_{k=1}^{n} a_k \qquad (1)$$

$$B_n = \sum_{k=n}^{\infty} a_k \qquad (2)$$

and

$$F(x) = \int_0^x f(t)\,dt \qquad (3)$$

$$G(x) = \int_x^\infty f(t)\,dt, \qquad (4)$$

and take $0 < p < 1$. For integrals,

$$\int_0^\infty \left[\frac{G(x)}{x}\right]^p dx > \left(\frac{p}{p-1}\right)^p \int_0^\infty [f(x)]^p\,dx \qquad (5)$$

(unless f is identically 0). For sums,

$$\left(1+\frac{1}{p-1}\right){B_1}^p + \sum_{n=2}^{\infty}\left(\frac{B_n}{n}\right)^p > \left(\frac{p}{p-1}\right)^p \sum_{n=1}^{\infty} {a_n}^p \qquad (6)$$

(unless all $a_n = 0$).

References

Beesack, P. R. "On Some Integral Inequalities of E. T. Copson." In *General Inequalities 2* (Ed. E. F. Beckenbach). Basel: Birkhäuser, 1980.

Copson, E. T. "Some Integral Inequalities." *Proc. Royal Soc. Edinburgh* **75A**, 157–164, 1975–1976.

Hardy, G. H.; Littlewood, J. E.; and Pólya, G. Theorems 326–327, 337–338, and 345 in *Inequalities*. Cambridge, England: Cambridge University Press, 1934.

Mitrinovic, D. S.; Pecaric, J. E.; and Fink, A. M. *Inequalities Involving Functions and Their Integrals and Derivatives.* Dordrecht, Netherlands: Kluwer, 1991.

Copula

A function that joins univariate distribution functions to form multivariate distribution functions. A 2-D copula is a function $C : I^2 \to I$ such that

$$C(0,t) = C(t,0) = 0$$

and

$$C(1,t) = C(t,1) = t$$

for all $t \in I$, and

$$C(u_2,v_2) - C(u_1,v_2) - C(u_2,v_1) + C(u_1,v_1) \geq 0$$

for all $u_1, u_2, v_1, v-2 \in I$ such that $u_1 \leq u_2$ and $v_1 \leq v-2$.

see also SKLAR'S THEOREM

Cork Plug

A 3-D SOLID which can stopper a SQUARE, TRIANGULAR, or CIRCULAR HOLE. There is an infinite family of such shapes. The one with smallest VOLUME has TRIANGULAR CROSS-SECTIONS and $V = \pi r^3$; that with the largest VOLUME is made using two cuts from the top diameter to the EDGE and has VOLUME $V = 4\pi r^3/3$.

see also STEREOLOGY, TRIP-LET

Corkscrew Surface

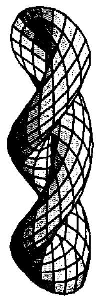

A surface also called the TWISTED SPHERE.

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.*Boca Raton, FL: CRC Press, pp. 493–494, 1993.

Cornish-Fisher Asymptotic Expansion

$$y \approx m + \sigma w,$$

where

$$\begin{aligned} w = x &+ [\gamma_1 h_1(x)] + [\gamma_2 h_2(x) + {\gamma_1}^2 h_{11}(x)] \\ &+ [\gamma_3 h_3(x) + \gamma_1\gamma_2 h_{12}(x) + {\gamma_1}^3 h_{111}(x)] \\ &+ [\gamma_4 h_4(x) + {\gamma_2}^2 h_{22}(x) + \gamma_1\gamma_3 h_{13}(x) \\ &+ {\gamma_1}^2\gamma_2 h_{112}(x) + {\gamma_1}^4 h_{1111}(x)] + \ldots, \end{aligned}$$

where

$$\begin{aligned}
h_1(x) &= \tfrac{1}{6}\,\mathrm{He}_2(x)\\
h_2(x) &= \tfrac{1}{24}\,\mathrm{He}_3(x)\\
h_{11}(x) &= -\tfrac{1}{36}[2\,\mathrm{He}_3(x)+\mathrm{He}_1(x)]\\
h_3(x) &= \tfrac{1}{120}\,\mathrm{He}_4(x)\\
h_{12}(x) &= -\tfrac{1}{24}[\mathrm{He}_4(x)+\mathrm{He}_2(x)]\\
h_{111}(x) &= \tfrac{1}{324}[12\,\mathrm{He}_4(x)+19\,\mathrm{He}_2(x)]\\
h_4(x) &= \tfrac{1}{720}\,\mathrm{He}_5(x)\\
h_{22}(x) &= -\tfrac{1}{384}[3\,\mathrm{He}_5(x)+6\,\mathrm{He}_3(x)+2\,\mathrm{He}_1(x)]\\
h_{13}(x) &= -\tfrac{1}{180}[2\,\mathrm{He}_5+3\,\mathrm{He}_3(x)]\\
h_{112}(x) &= \tfrac{1}{288}[14\,\mathrm{He}_5(x)+37\,\mathrm{He}_3(x)+8\,\mathrm{He}_1(x)]\\
h_{1111}(x) &= -\tfrac{1}{7776}[252\,\mathrm{He}_5(x)+832\,\mathrm{He}_3(x)+227\,\mathrm{He}_1(x)].
\end{aligned}$$

see also EDGEWORTH SERIES, GRAM-CHARLIER SERIES

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 935, 1972.

Cornu Spiral

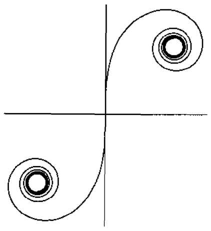

A plot in the COMPLEX PLANE of the points

$$B(z) = C(t) + iS(t) = \int_0^t e^{i\pi x^2/2}\,dx, \qquad (1)$$

where $C(z)$ and $S(z)$ are the FRESNEL INTEGRALS. The Cornu spiral is also known as the CLOTHOID or EULER'S SPIRAL. A Cornu spiral describes diffraction from the edge of a half-plane.

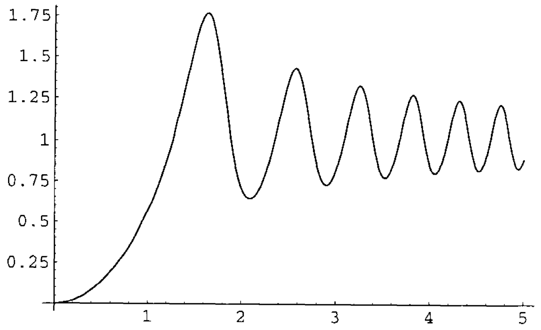

The SLOPE of the Cornu spiral

$$m(t) = \frac{S(t)}{C(t)} \qquad (2)$$

is plotted above.

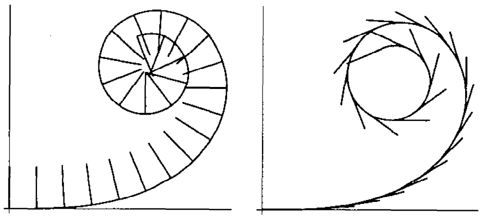

The SLOPE of the curve's TANGENT VECTOR (above right figure) is

$$m_T(t) = \frac{S'(t)}{C'(t)} = \tan(\tfrac{1}{2}\pi t^2), \qquad (3)$$

plotted below.

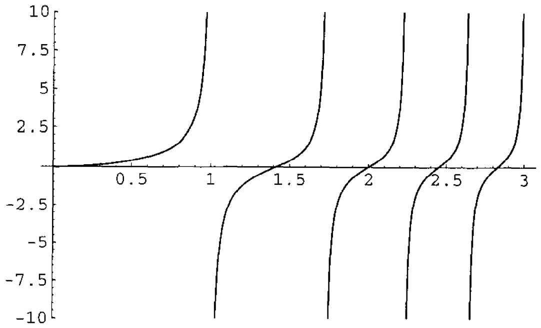

The CESÀRO EQUATION for a Cornu spiral is $\rho = c^2/s$, where ρ is the RADIUS OF CURVATURE and s the ARC LENGTH. The TORSION is $\tau = 0$.

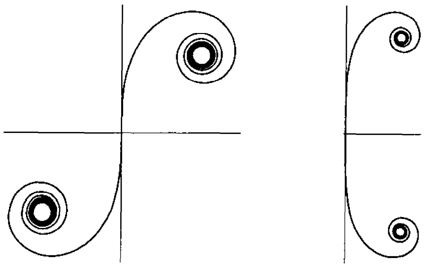

Gray (1993) defines a generalization of the Cornu spiral given by parametric equations

$$x(t) = a\int_0^t \sin\left(\frac{u^{n+1}}{n+1}\right)\,du \qquad (4)$$

$$y(t) = a\int_0^t \cos\left(\frac{u^{n+1}}{n+1}\right)\,du. \qquad (5)$$

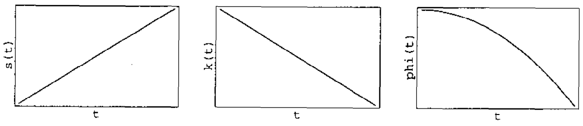

The ARC LENGTH, CURVATURE, and TANGENTIAL ANGLE of this curve are

$$\begin{aligned} s(t) &= at \tag{3}\\ \kappa(t) &= -\frac{t^n}{a} \tag{4}\\ \phi(t) &= -\frac{t^{n+1}}{n+1}. \tag{5}\end{aligned}$$

The CESÀRO EQUATION is

$$\kappa = -\frac{a}{s^n}. \tag{6}$$

Dillen (1990) describes a class of "polynomial spirals" for which the CURVATURE is a polynomial function of the ARC LENGTH. These spirals are a further generalization of the Cornu spiral.

see also FRESNEL INTEGRALS, NIELSEN'S SPIRAL

References

Dillen, F. "The Classification of Hypersurfaces of a Euclidean Space with Parallel Higher Fundamental Form." *Math. Z.* **203**, 635–643, 1990.

Gray, A. "Clothoids." §3.6 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 50–52, 1993.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 190–191, 1972.

Cornucopia

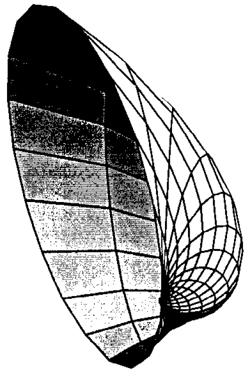

The SURFACE given by the parametric equations

$$\begin{aligned} x &= e^{bv}\cos v + e^{av}\cos u\cos v\\ y &= e^{bv}\sin v + e^{av}\cos u\sin v\\ z &= e^{av}\sin u.\end{aligned}$$

References

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 304, 1993.

Corollary

An immediate consequence of a result already proved. Corollaries usually state more complicated THEOREMS in a language simpler to use and apply.

see also LEMMA, PORISM, THEOREM

Corona (Polyhedron)

see AUGMENTED SPHENOCORONA, HEBESPHENOMEGACORONA, SPHENOCORONA, SPHENOMEGACORONA

Corona (Tiling)

The first corona of a TILE is the set of all tiles that have a common boundary point with that tile (including the original tile itself). The second corona is the set of tiles that share a point with something in the first corona, and so on.

References

Eppstein, D. "Heesch's Problem." `http://www.ics.uci.edu/~eppstein/junkyard/heesch`.

Correlation

see AUTOCORRELATION, CORRELATION COEFFICIENT, CORRELATION (GEOMETRIC), CORRELATION (STATISTICAL), CROSS-CORRELATION

Correlation Coefficient

The correlation coefficient is a quantity which gives the quality of a LEAST SQUARES FITTING to the original data. To define the correlation coefficient, first consider the sum of squared values ss_{xx}, ss_{xy}, and ss_{yy} of a set of n data points (x_i, y_i) about their respective means,

$$\begin{aligned} \mathrm{ss}_{xx} &\equiv \Sigma(x_i-\bar{x})^2 = \Sigma x^2 - 2\bar{x}\Sigma x + \Sigma\bar{x}^2\\ &= \Sigma x^2 - 2n\bar{x}^2 + n\bar{x}^2 = \Sigma x^2 - n\bar{x}^2 \tag{1}\\ \mathrm{ss}_{yy} &\equiv \Sigma(y_i-\bar{y})^2 = \Sigma y^2 - 2\bar{y}\Sigma y + \Sigma\bar{y}^2\\ &= \Sigma y^2 - 2n\bar{y}^2 + n\bar{y}^2 = \Sigma y^2 - n\bar{y}^2 \tag{2}\\ \mathrm{ss}_{xy} &\equiv \Sigma(x_i-\bar{x})(y_i-\bar{y}) = \Sigma(x_iy_i - \bar{x}y_i - x_i\bar{y} + \bar{x}\bar{y})\\ &= \Sigma xy - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y} = \Sigma xy - n\bar{x}\bar{y}. \tag{3}\end{aligned}$$

For linear LEAST SQUARES FITTING, the COEFFICIENT b in

$$y = a + bx \tag{4}$$

is given by

$$b = \frac{n\sum xy - \sum x\sum y}{n\sum x^2 - \left(\sum x\right)^2} = \frac{\mathrm{ss}_{xy}}{\mathrm{ss}_{xx}}, \tag{5}$$

and the COEFFICIENT b' in

$$x = a' + b'y \tag{6}$$

is given by

$$b' = \frac{n\sum xy - \sum x\sum y}{n\sum y^2 - \left(\sum y\right)^2}. \tag{7}$$

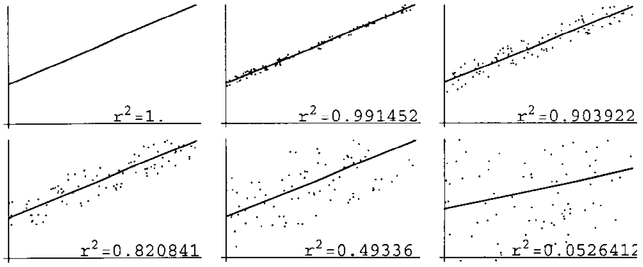

The correlation coefficient r^2 (sometimes also denoted R^2) is then defined by

$$r \equiv \sqrt{bb'} = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right]\left[n\sum y^2 - \left(\sum y\right)^2\right]}}, \tag{8}$$

which can be written more simply as

$$r^2 = \frac{\mathrm{ss}_{xy}{}^2}{\mathrm{ss}_{xx}\mathrm{ss}_{yy}}. \tag{9}$$

The correlation coefficient is also known as the PRODUCT-MOMENT COEFFICIENT OF CORRELATION or PEARSON'S CORRELATION. The correlation coefficients for linear fits to increasingly noise data are shown above.

The correlation coefficient has an important physical interpretation. To see this, define

$$A \equiv (\Sigma x^2 - n\bar{x}^2)^{-1} \tag{10}$$

and denote the "expected" value for y_i as $\hat{y}_i$. Sums of $\hat{y}_i$ are then

$$\begin{aligned}
\hat{y}_i &= a + bx_i = \bar{y} - b\bar{x} + bx_i = \bar{x} + b(x_i - \bar{x}) \\
&= A(\bar{y}\Sigma x^2 - \bar{x}\Sigma xy + x_i\Sigma xy - n\bar{x}\bar{y}x_i) \\
&= A[\bar{y}\Sigma x^2 + (x_i - \bar{x})\Sigma xy - n\bar{x}\bar{y}x_i] && (11) \\
\Sigma\hat{y}_i &= A(n\bar{y}\Sigma x^2 - n^2\bar{x}^2\bar{y}) && (12) \\
\Sigma\hat{y}_i^2 &= A^2[n\bar{y}^2(\Sigma x^2)^2 - n^2\bar{x}^2\bar{y}^2(\Sigma x^2) \\
&\quad - 2n\bar{x}\bar{y}(\Sigma xy)(\Sigma x^2) + 2n^2\bar{x}^3\bar{y}(\Sigma xy) \\
&\quad + (\Sigma x^2)(\Sigma xy)^2 - n\bar{x}^2(\Sigma xy)] && (13) \\
\Sigma y_i\hat{y}_i &= A\Sigma[y_i\bar{y}\Sigma x^2 + y_i(x_i - \bar{x})\Sigma xy - n\bar{x}\bar{y}x_iy_i] \\
&= A[n\bar{y}^2\Sigma x^2 + (\Sigma xy)^2 - n\bar{x}\bar{y}\Sigma xy - n\bar{x}\bar{y}(\Sigma xy)] \\
&= A[n\bar{y}^2\Sigma x^2 + (\Sigma xy)^2 - 2n\bar{x}\bar{y}\Sigma xy]. && (14)
\end{aligned}$$

The sum of squared residuals is then

$$\begin{aligned}
\mathrm{SSR} &\equiv \Sigma(\hat{y}_i - \bar{y})^2 = \Sigma(\hat{y}_i^2 - 2\bar{y}\hat{y}_i + \bar{y}^2) \\
&= A^2(\Sigma xy - n\bar{x}\bar{y})^2(\Sigma x^2 - n\bar{x}^2) = \frac{(\Sigma xy - n\bar{x}\bar{y})^2}{\Sigma x^2 - n\bar{x}^2} \\
&= b\,\mathrm{ss}_{xy} = \frac{\mathrm{ss}_{xy}{}^2}{\mathrm{ss}_{xx}} = \mathrm{ss}_{yy}r^2 = b^2\mathrm{ss}_{xx}, && (15)
\end{aligned}$$

and the sum of squared errors is

$$\begin{aligned}
\mathrm{SSE} &\equiv \Sigma(y_i - \hat{y}_i)^2 = \Sigma(y_i - \bar{y} + b\bar{x} - bx_i)^2 \\
&= \Sigma[y_i - \bar{y} - b(x_i - \bar{x})]^2 \\
&= \Sigma(y_i - \bar{y})^2 + b^2\Sigma(x_i - \bar{x})^2 - 2b\Sigma(x_i - \bar{x})(y_i - \bar{y}) \\
&= \mathrm{ss}_{yy} + b^2\mathrm{ss}_{xx} - 2b\mathrm{ss}_{xy}. && (16)
\end{aligned}$$

But

$$b = \frac{\mathrm{ss}_{xy}}{\mathrm{ss}_{xx}} \tag{17}$$

$$r^2 = \frac{\mathrm{ss}_{xy}{}^2}{\mathrm{ss}_{xx}\mathrm{ss}_{yy}}, \tag{18}$$

so

$$\begin{aligned}
\mathrm{SSE} &= \mathrm{ss}_{yy} + \frac{\mathrm{ss}_{xy}{}^2}{\mathrm{ss}_{xx}{}^2}\mathrm{ss}_{xx} - 2\frac{\mathrm{ss}_{xy}}{\mathrm{ss}_{xx}}\mathrm{ss}_{xy} \\
&= \mathrm{ss}_{yy} - \frac{\mathrm{ss}_{xy}{}^2}{\mathrm{ss}_{xx}} && (19) \\
&= \mathrm{ss}_{yy}\left(1 - \frac{\mathrm{ss}_{xy}{}^2}{\mathrm{ss}_{xx}{}^2}\right) = \mathrm{ss}_{yy}(1 - r^2) \\
&= s_y{}^2 - s_{\hat{y}}{}^2, && (20)
\end{aligned}$$

and

$$\mathrm{SSE} + \mathrm{SSR} = \mathrm{ss}_{yy}(1 - r^2) + \mathrm{ss}_{yy}r^2 = \mathrm{ss}_{yy}. \tag{21}$$

The square of the correlation coefficient r^2 is therefore given by

$$r^2 \equiv \frac{\mathrm{SSR}}{\mathrm{ss}_{yy}} = \frac{\mathrm{ss}_{xy}{}^2}{\mathrm{ss}_{xx}\mathrm{ss}_{yy}} = \frac{(\Sigma xy - n\bar{x}\bar{y})^2}{(\Sigma x^2 - n\bar{x}^2)(\Sigma y^2 - n\bar{y}^2)}. \tag{22}$$

In other words, r^2 is the proportion of ss_{yy} which is accounted for by the regression.

If there is complete correlation, then the lines obtained by solving for best-fit (a, b) and (a', b') coincide (since all data points lie on them), so solving (6) for y and equating to (4) gives

$$y = -\frac{a'}{b'} + \frac{x}{b'} = a + bx. \tag{23}$$

Therefore, $a = -a'/b'$ and $b = 1/b'$, giving

$$r^2 = bb' = 1. \tag{24}$$

The correlation coefficient is independent of both origin and scale, so

$$r(u, v) = r(x, y), \tag{25}$$

where

$$u \equiv \frac{x - x_0}{h} \tag{26}$$

$$v \equiv \frac{y - y_0}{h}. \tag{27}$$

see also CORRELATION INDEX, CORRELATION COEFFICIENT—GAUSSIAN BIVARIATE DISTRIBUTION, CORRELATION RATIO, LEAST SQUARES FITTING, REGRESSION COEFFICIENT

References

Acton, F. S. *Analysis of Straight-Line Data.* New York: Dover, 1966.

Kenney, J. F. and Keeping, E. S. "Linear Regression and Correlation." Ch. 15 in *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, pp. 252–285, 1962.

Gonick, L. and Smith, W. *The Cartoon Guide to Statistics.* New York: Harper Perennial, 1993.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Linear Correlation." §14.5 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 630–633, 1992.

Correlation Coefficient—Gaussian Bivariate Distribution

For a GAUSSIAN BIVARIATE DISTRIBUTION, the distribution of correlation COEFFICIENTS is given by

$$\begin{aligned}P(r) &= \frac{1}{\pi}(N-2)(1-r^2)^{(N-4)/2}(1-\rho^2)^{(N-1)/2} \\ &\quad \times \int_0^\infty \frac{d\beta}{(\cosh\beta-\rho r)^{N-1}} \\ &= \frac{1}{\pi}(N-2)(1-r^2)^{(N-4)/2}(1-\rho^2)^{(N-1)/2}\sqrt{\frac{\pi}{2}}\,\frac{\Gamma(N-1)}{\Gamma(N-\frac{1}{2})} \\ &\quad \times(1-\rho r)^{-(N-3/2)}\,{}_2F_1(\tfrac{1}{2},\tfrac{1}{2},\tfrac{2N-1}{2};\tfrac{\rho r+1}{2}) \\ &= \frac{(N-2)\Gamma(N-1)(1-\rho^2)^{(N-1)/2}(1-r^2)^{(N-4)/2}}{\sqrt{2\pi}\,\Gamma(N-\frac{1}{2})(1-\rho r)^{N-3/2}} \\ &\quad \times\left[1+\frac{1}{4}\frac{\rho r+1}{2N-1}+\frac{9}{16}\frac{(\rho r+1)^2}{(2N-1)(2N+1)}+\cdots\right], \qquad (1)\end{aligned}$$

where ρ is the population correlation COEFFICIENT, ${}_2F_1(a,b;c;x)$ is a HYPERGEOMETRIC FUNCTION, and $\Gamma(z)$ is the GAMMA FUNCTION (Kenney and Keeping 1951, pp. 217–221). The MOMENTS are

$$\langle r\rangle = \rho-\frac{\rho(1-\rho^2)}{2n} \qquad (2)$$

$$\operatorname{var}(r) = \frac{(1-\rho^2)^2}{n}\left(1+\frac{11\rho^2}{2n}+\cdots\right) \qquad (3)$$

$$\gamma_1 = \frac{6\rho}{\sqrt{n}}\left(1+\frac{77\rho^2-30}{12n}+\cdots\right)$$

$$\gamma_2 = \frac{6}{n}(12\rho^2-1)+\ldots, \qquad (4)$$

where $n\equiv N-1$. If the variates are uncorrelated, then $\rho=0$ and

$$\begin{aligned}{}_2F_1(\tfrac{1}{2},\tfrac{1}{2},\tfrac{2N-1}{2};\tfrac{\rho r+1}{2}) &= {}_2F_1(\tfrac{1}{2},\tfrac{1}{2},\tfrac{2N-1}{2};\tfrac{1}{2}) \\ &= \frac{\Gamma(N-\frac{1}{2})2^{3/2-N}\sqrt{\pi}}{[\Gamma(\frac{N}{2})]^2}, \qquad (5)\end{aligned}$$

so

$$\begin{aligned}P(r) &= \frac{(N-2)\Gamma(N-1)}{\sqrt{2\pi}\,\Gamma(N-\frac{1}{2})}(1-r^2)^{(N-4)/2} \\ &\quad \times\frac{\Gamma(N-\frac{1}{2})2^{3/2-N}\sqrt{\pi}}{[\Gamma(\frac{N}{2})]^2} \\ &= \frac{2^{1-N}(N-2)\Gamma(N-1)}{[\Gamma(\frac{N}{2})]^2}(1-r^2)^{(N-4/2)}. \qquad (6)\end{aligned}$$

But from the LEGENDRE DUPLICATION FORMULA,

$$\sqrt{\pi}\,\Gamma(N-1) = 2^{N-2}\Gamma(\tfrac{N}{2})\Gamma(\tfrac{N-1}{2}), \qquad (7)$$

so

$$\begin{aligned}P(r) &= \frac{(2^{1-N})(2^{N-2})(N-2)\Gamma(\frac{N}{2})\Gamma(\frac{N-1}{2})}{\sqrt{\pi}\,[\Gamma(\frac{N}{2})]^2}(1-r^2)^{(N-4)/2} \\ &= \frac{(N-2)\Gamma(\frac{N-1}{2})}{2\sqrt{\pi}\,\Gamma(\frac{N}{2})}(1-r^2)^{(N-4)/2} \\ &= \frac{1}{\sqrt{\pi}}\frac{\frac{\nu}{2}\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}+1)}(1-r^2)^{(\nu-2)/2} \\ &= \frac{1}{\sqrt{\pi}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}(1-r^2)^{(\nu-2)/2}. \qquad (8)\end{aligned}$$

The uncorrelated case can be derived more simply by letting β be the true slope, so that $\eta=\alpha+\beta x$. Then

$$t\equiv(b-\beta)\frac{s_x}{s_y}\sqrt{\frac{N-2}{1-r^2}} = \frac{(b-\beta)r}{b}\sqrt{\frac{N-2}{1-r^2}} \qquad (9)$$

is distributed as STUDENT'S t with $\nu\equiv N-2$ DEGREES OF FREEDOM. Let the population regression COEFFICIENT ρ be 0, then $\beta=0$, so

$$t = r\sqrt{\frac{\nu}{1-r^2}}, \qquad (10)$$

and the distribution is

$$P(t)\,dt = \frac{1}{\sqrt{\nu\pi}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\left(1+\frac{t^2}{\nu}\right)^{(\nu+1)/2}}\,dt. \qquad (11)$$

Plugging in for t and using

$$\begin{aligned}dt &= \sqrt{\nu}\left[\frac{\sqrt{1-r^2}-r(\frac{1}{2})(-2r)(1-r^2)^{-1/2}}{1-r^2}\right]dr \\ &= \sqrt{\frac{\nu}{1-r^2}}\left(\frac{1-r^2+r^2}{1-r^2}\right)dr \\ &= \sqrt{\frac{\nu}{(1-r)^3}}\,dr \qquad (12)\end{aligned}$$

gives

$$P(t)\,dt = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\left[1+\frac{r^2\nu}{(1-r^2)\nu}\right]^{(\nu+1)/2}} \sqrt{\frac{\nu}{(1-r)^3}}\,dr$$
$$= \frac{(1-r^2)^{-3/2}}{\sqrt{\pi}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\left(\frac{1}{1-r^2}\right)^{(\nu+1)/2}}\,dr$$
$$= \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} (1-r^2)^{-3/2}(1-r^2)^{(\nu+1)/2}\,dr$$
$$= \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} (1-r^2)^{(\nu-2)/2}\,dr, \tag{13}$$

so

$$P(r) = \frac{1}{\sqrt{\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} (1-r^2)^{(\nu-2)/2} \tag{14}$$

as before. See Bevington (1969, pp. 122–123) or Pugh and Winslow (1966, §12-8). If we are interested instead in the probability that a correlation Coefficient would be obtained $\geq |r|$, where r is the observed Coefficient, then

$$P_c(r,N) = 2\int_{|r|}^{1} P(r',N)\,dr' = 1 - 2\int_0^{|r|} P(r',N)\,dr'$$
$$= 1 - \frac{2}{\sqrt{\pi}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \int_0^{|r|} (1-r^2)^{(\nu-2)/2}\,dr. \tag{15}$$

Let $I \equiv \frac{1}{2}(\nu-2)$. For Even ν, the exponent I is an Integer so, by the Binomial Theorem,

$$(1-r^2)^I = \sum_{k=0}^{I} \binom{I}{k} (-r^2)^k \tag{16}$$

and

$$P_c(r) = 1 - \frac{2}{\sqrt{\pi}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} (-1)^k \frac{I!}{(I-k)!k!} \int_0^{|r|} \sum_{k=0}^{I} r'^{2k}\,dr'$$
$$= 1 - \frac{2}{\sqrt{\pi}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \sum_{k=0}^{I} \left[(-1)^k \frac{I!}{(I-k)!k!} \frac{|r|^{2k+1}}{2k+1}\right]. \tag{17}$$

For Odd ν, the integral is

$$P_c(r) = 1 - 2\int_0^{|r|} P(r')\,dr'$$
$$= 1 - \frac{2}{\sqrt{\pi}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \int_0^{|r|} (\sqrt{1-r^2}\,)^{\nu-2}\,dr. \tag{18}$$

Let $r \equiv \sin x$ so $dr = \cos x\,dx$, then

$$P_c(r) = 1 - \frac{2}{\sqrt{\pi}} \frac{\Gamma[(\frac{\nu+1}{2})]}{\Gamma(\frac{\nu}{2})} \int_0^{\sin^{-1}|r|} \cos^{\nu-2} x \cos x\,dx$$
$$= 1 - \frac{2}{\sqrt{\pi}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} + \int_0^{\sin^{-1}|r|} \cos^{\nu-1} x\,dx. \tag{19}$$

But ν is Odd, so $\nu - 1 \equiv 2n$ is Even. Therefore

$$\frac{2}{\sqrt{\pi}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} = \frac{2}{\sqrt{\pi}} \frac{\Gamma(n+1)}{\Gamma(n+\frac{1}{2})} = \frac{2}{\sqrt{\pi}} \frac{n!}{\frac{(2n-1)!!\sqrt{\pi}}{2^n}}$$
$$= \frac{2}{\pi} \frac{2^n n!}{(2n-1)!!} = \frac{2}{\pi} \frac{(2n)!!}{(2n-1)!!}. \tag{20}$$

Combining with the result from the Cosine Integral gives

$$P_c(r) = 1 - \frac{2}{\pi} \frac{(2n)!!(2n-1)!!}{(2n-1)!!(2n)!!}$$
$$\times \left[\sin x \sum_{k=0}^{n-1} \frac{(2k)!!}{(2k+1)!!} \cos^{2k+1} x + x\right]_0^{\sin^{-1}|r|}. \tag{21}$$

Use

$$\cos^{2k-1} x = (1-r^2)^{(2k-1)/2} = (1-r^2)^{(k-1/2)}, \tag{22}$$

and define $J \equiv n-1 = (\nu-3)/2$, then

$$P_c(r)$$
$$= 1 - \frac{2}{\pi}\left[\sin^{-1}|r| + |r| \sum_{k=0}^{J} \frac{(2k)!!}{(2k+1)!!} (1-r^2)^{k+1/2}\right]. \tag{23}$$

(In Bevington 1969, this is given incorrectly.) Combining the correct solutions

$$P_c(r) = \begin{cases} 1 - \frac{2}{\sqrt{\pi}} \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)} \sum_{k=0}^{I} \left[(-1)^k \frac{I!}{(I-k)!k!} \frac{|r|^{2k+1}}{2k+1}\right] \\ \quad \text{for } \nu \text{ even} \\ 1 - \frac{2}{\pi}\left[\sin^{-1}|r| + |r| \sum_{k=0}^{J} \frac{(2k)!!}{(2k+1)!!} (1-r^2)^{k+1/2}\right] \\ \quad \text{for } \nu \text{ odd} \end{cases} \tag{24}$$

If $\rho \neq 0$, a skew distribution is obtained, but the variable z defined by

$$z \equiv \tanh^{-1} r \tag{25}$$

is approximately normal with

$$\mu_z = \tanh^{-1} \rho \tag{26}$$
$$\sigma_z{}^2 = \frac{1}{N-3} \tag{27}$$

(Kenney and Keeping 1962, p. 266).

Let b_j be the slope of a best-fit line, then the multiple correlation Coefficient is

$$R^2 \equiv \sum_{j=1}^{n} \left(b_j \frac{s_{jy}{}^2}{s_y{}^2}\right) = \sum_{j=1}^{n} \left(b_j \frac{s_j}{s_y} r_{jy}\right), \tag{28}$$

where s_{jy} is the sample Variance.

On the surface of a SPHERE,

$$r \equiv \frac{\int fg\,d\Omega}{\int f\,d\Omega \int g\,d\Omega}, \tag{29}$$

where $d\Omega$ is a differential SOLID ANGLE. This definition guarantees that $-1 < r < 1$. If f and g are expanded in REAL SPHERICAL HARMONICS,

$$\begin{aligned} f(\theta,\phi) \equiv \sum_{l=0}^{\infty}\sum_{m=0}^{l}[C_l^m Y_l^{m\,c}(\theta,\phi)\sin(m\phi) \\ + S_l^m Y_l^{m\,s}(\theta,\phi)] \end{aligned} \tag{30}$$

$$\begin{aligned} g(\theta,\phi) \equiv \sum_{l=0}^{\infty}\sum_{m=0}^{l}[A_l^m Y_l^{m\,c}(\theta,\phi)\sin(m\phi) \\ + B_l^m Y_l^{m\,s}(\theta,\phi)]. \end{aligned} \tag{31}$$

Then

$$r_l = \frac{\sum_{m=0}^{l}(C_l^m A_l^m + S_l^m B_l^m)}{\sqrt{\sum_{m=0}^{l}(C_l^{m\,2} + S_l^{m\,2})}\sqrt{\sum_{m=0}^{l}(A_l^{m\,2} + B_l^{m\,2})}}. \tag{32}$$

The confidence levels are then given by

$$\begin{aligned} G_1(r) &= r \\ G_2(r) &= r(1 + \tfrac{1}{2}s^2) = \tfrac{1}{2}r(3 - r^2) \\ G_3(r) &= r[1 + \tfrac{1}{2}s^2(1 + \tfrac{3}{4}s^2)] = \tfrac{1}{8}r(15 - 10r^2 + 3r^4) \\ G_4(r) &= r\{1 + \tfrac{1}{2}s^2[1 + \tfrac{3}{4}s^2(1 + \tfrac{5}{6}s^2)]\} \\ &= \tfrac{1}{16}r(35 - 35r^2 + 21r^4 - 5r^6), \end{aligned}$$

where

$$s \equiv \sqrt{1 - r^2} \tag{33}$$

(Eckhardt 1984).

see also FISHER'S z'-TRANSFORMATION, SPEARMAN RANK CORRELATION, SPHERICAL HARMONIC

References

Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences.* New York: McGraw-Hill, 1969.

Eckhardt, D. H. "Correlations Between Global Features of Terrestrial Fields." *Math. Geology* **16**, 155–171, 1984.

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, 1962.

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, 1951.

Pugh, E. M. and Winslow, G. H. *The Analysis of Physical Measurements.* Reading, MA: Addison-Wesley, 1966.

Correlation Dimension

Define the correlation integral as

$$C(\epsilon) \equiv \lim_{n\to\infty}\frac{1}{N^2}\sum_{\substack{i,j=1\\ i\neq j}}^{\infty} H(\epsilon - \|x_i - x_j\|), \tag{1}$$

where H is the HEAVISIDE STEP FUNCTION. When the below limit exists, the correlation dimension is then defined as

$$D_2 \equiv d_{\rm cor} \equiv \lim_{\epsilon,\epsilon'\to 0^+}\frac{\ln\left[\frac{C(\epsilon)}{C(\epsilon')}\right]}{\ln\left(\frac{\epsilon}{\epsilon'}\right)}. \tag{2}$$

If ν is the CORRELATION EXPONENT, then

$$\lim_{\epsilon\to 0}\nu \to D_2. \tag{3}$$

It satisfies

$$d_{\rm cor} \leq d_{\rm inf} \leq d_{\rm cap} \overset{?}{=} d_{\rm Lya}. \tag{4}$$

To estimate the correlation dimension of an M-dimensional system with accuracy $(1-Q)$ requires $N_{\rm min}$ data points, where

$$N_{\rm min} \geq \left[\frac{R(2-Q)}{2(1-Q)}\right]^M, \tag{5}$$

where $R \geq 1$ is the length of the "plateau region." If an ATTRACTOR exists, then an estimate of D_2 saturates above some M given by

$$M \geq 2D + 1, \tag{6}$$

which is sometimes known as the fractal Whitney embedding prevalence theorem.

see also CORRELATION EXPONENT, q-DIMENSION

References

Nayfeh, A. H. and Balachandran, B. *Applied Nonlinear Dynamics: Analytical, Computational, and Experimental Methods.* New York: Wiley, pp. 547–548, 1995.

Correlation Exponent

A measure ν of a STRANGE ATTRACTOR which allows the presence of CHAOS to be distinguished from random noise. It is related to the CAPACITY DIMENSION D and INFORMATION DIMENSION σ, satisfying

$$\nu \leq \sigma \leq D. \tag{1}$$

It satisfies

$$\nu \leq D_{\rm KY}, \tag{2}$$

where $D_{\rm KY}$ is the KAPLAN-YORKE DIMENSION. As the cell size goes to zero,

$$\lim_{\epsilon\to 0}\nu \to D_2, \tag{3}$$

where D_2 is the CORRELATION DIMENSION.

References

Grassberger, P. and Procaccia, I. "Measuring the Strangeness of Strange Attractors." *Physica D* **9**, 189–208, 1983.

Correlation (Geometric)

A point-to-line and line-to-point TRANSFORMATION which transforms points A into lines a' and lines b into points B' such that a' passes through B' IFF A' lies on b.

see also POLARITY

Correlation Index

$$r_c \equiv \frac{s_{y\hat{y}}}{s_y s_{\hat{y}}}$$

$$r_c{}^2 = \frac{s_{\hat{y}}{}^2}{s_y{}^2} = 1 - \frac{\mathrm{SSE}}{s_y{}^2}.$$

see also CORRELATION COEFFICIENT

Correlation Integral

Consider a set of points $\mathbf{X}_i$ on an ATTRACTOR, then the correlation integral is

$$C(l) \equiv \lim_{N\to\infty} \frac{1}{N^2} f,$$

where f is the number of pairs (i,j) whose distance $|\mathbf{X}_i - \mathbf{X}_j| < l$. For small l,

$$C(l) \sim l^\nu,$$

where ν is the CORRELATION EXPONENT.

References

Grassberger, P. and Procaccia, I. "Measuring the Strangeness of Strange Attractors." *Physica D* **9**, 189–208, 1983.

Correlation Ratio

Let there be N_i observations of the ith phenomenon, where $i = 1, \ldots, p$ and

$$N \equiv \sum N_i \qquad (1)$$

$$\bar{y}_i \equiv \frac{1}{N_i} \sum_\alpha y_{i\alpha} \qquad (2)$$

$$\bar{y} \equiv \frac{1}{N} \sum_i \sum_\alpha y_{i\alpha}. \qquad (3)$$

Then

$$E_{yx}{}^2 \equiv \frac{\sum_i N_i(\bar{y}_i - \bar{y})^2}{\sum_i \sum_\alpha (y_{i\alpha} - \bar{y})^2}. \qquad (4)$$

Let η_{yx} be the population correlation ratio. If $N_i = N_j$ for $i \neq j$, then

$$f(E^2) = \frac{e^{-\lambda}(E^2)^{a-1}(1-E^2)^{b-1}{}_1F_1(a,b;\lambda E^2)}{B(a,b)}, \qquad (5)$$

where

$$\lambda \equiv \frac{N\eta^2}{2(1-\eta^2)} \qquad (6)$$

$$a \equiv \frac{n_1}{2} \qquad (7)$$

$$b \equiv \frac{n_2}{2}, \qquad (8)$$

and ${}_1F_1(a,b;z)$ is the CONFLUENT HYPERGEOMETRIC LIMIT FUNCTION. If $\lambda = 0$, then

$$f(E^2) = \beta(a,b) \qquad (9)$$

(Kenney and Keeping 1951, pp. 323–324).

see also CORRELATION COEFFICIENT, REGRESSION COEFFICIENT

References

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, 1951.

Correlation (Statistical)

For two variables x and y,

$$\operatorname{cor}(x,y) \equiv \frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y}, \qquad (1)$$

where σ_x denotes STANDARD DEVIATION and $\operatorname{cov}(x,y)$ is the COVARIANCE of these two variables. For the general case of variables x_i and x_j, where $i,j = 1, 2, \ldots, n$,

$$\operatorname{cor}(x_i,x_j) = \frac{\operatorname{cov}(x_i,x_j)}{\sqrt{V_{ii}V_{jj}}}, \qquad (2)$$

where V_{ii} are elements of the COVARIANCE MATRIX. In general, a correlation gives the strength of the relationship between variables. The variance of any quantity is alway NONNEGATIVE by definition, so

$$\operatorname{var}\left(\frac{x}{\sigma_x} + \frac{y}{\sigma_y}\right) \geq 0. \qquad (3)$$

From a property of VARIANCES, the sum can be expanded

$$\operatorname{var}\left(\frac{x}{\sigma_x}\right) + \operatorname{var}\left(\frac{y}{\sigma_y}\right) + 2\operatorname{cov}\left(\frac{x}{\sigma_x}, \frac{y}{\sigma_y}\right) \geq 0 \qquad (4)$$

$$\frac{1}{\sigma_x{}^2}\operatorname{var}(x) + \frac{1}{\sigma_y{}^2}\operatorname{var}(y) + \frac{2}{\sigma_x\sigma_y}\operatorname{cov}(x,y) \geq 0 \qquad (5)$$

$$1 + 1 + \frac{2}{\sigma_x\sigma_y}\operatorname{cov}(x,y) = 2 + \frac{2}{\sigma_x\sigma_y}\operatorname{cov}(x,y) \geq 0. \qquad (6)$$

Therefore,

$$\operatorname{cor}(x,y) = \frac{\operatorname{cov}(x,y)}{\sigma_x\sigma_y} \geq -1. \qquad (7)$$

Similarly,

$$\operatorname{var}\left(\frac{x}{\sigma_x}\right) - \left(\frac{y}{\sigma_y}\right) \geq 0 \tag{8}$$

$$\operatorname{var}\left(\frac{x}{\sigma_x}\right) + \operatorname{var}\left(-\frac{y}{\sigma_y}\right) + 2\operatorname{cov}\left(\frac{x}{\sigma_x}, -\frac{y}{\sigma_y}\right) \geq 0 \tag{9}$$

$$\frac{1}{{\sigma_x}^2}\operatorname{var}(x) + \frac{1}{{\sigma_y}^2}\operatorname{var}(y) - \frac{2}{\sigma_x\sigma_y}\operatorname{cov}(x,y) \geq 0 \tag{10}$$

$$1 + 1 - \frac{2}{\sigma_x\sigma_y}\operatorname{cov}(x,y) = 2 - \frac{2}{\sigma_x\sigma_y}\operatorname{cov}(x,y) \geq 0. \tag{11}$$

Therefore,

$$\operatorname{cor}(x,y) = \frac{\operatorname{cov}(x,y)}{\sigma_x\sigma_y} \leq 1, \tag{12}$$

so $-1 \leq \operatorname{cor}(x,y) \leq 1$. For a linear combination of two variables,

$$\begin{aligned}\operatorname{var}(y - bx) &= \operatorname{var}(y) + \operatorname{var}(-bx) + 2\operatorname{cov}(y, -bx)\\ &= \operatorname{var}(y) + b^2\operatorname{var}(x) - 2b\operatorname{cov}(x,y)\\ &= {\sigma_y}^2 + {\sigma_x}^2 - 2b\operatorname{cov}(x,y).\end{aligned} \tag{13}$$

Examine the cases where $\operatorname{cor}(x,y) = \pm 1$,

$$\operatorname{cor}(x,y) \equiv \frac{\operatorname{cov}(x,y)}{\sigma_x\sigma_y} = \pm 1 \tag{14}$$

$$\operatorname{var}(y - bx) = b^2{\sigma_x}^2 + {\sigma_y}^2 \mp 2b\sigma_x\sigma_y = (b\sigma_x \mp \sigma_y)^2. \tag{15}$$

The VARIANCE will be zero if $b \equiv \pm\sigma_y/\sigma_x$, which requires that the argument of the VARIANCE is a constant. Therefore, $y - bx = a$, so $y = a + bx$. If $\operatorname{cor}(x,y) = \pm 1$, y is either perfectly correlated ($b > 0$) or perfectly anticorrelated ($b < 0$) with x.

see also COVARIANCE, COVARIANCE MATRIX, VARIANCE

Cosecant

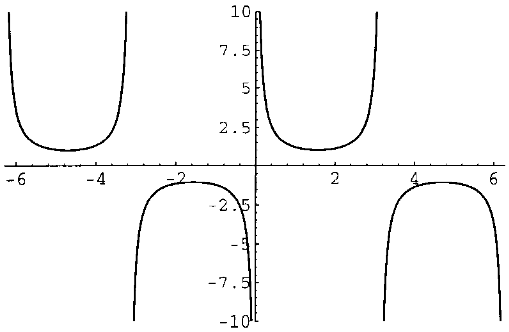

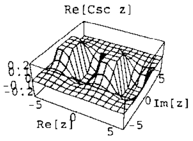

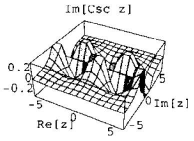

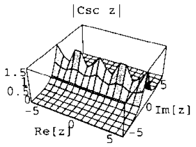

The function defined by $\csc x \equiv 1/\sin x$, where $\sin x$ is the SINE. The MACLAURIN SERIES of the cosecant function is

$$\begin{aligned}\csc x = \frac{1}{x} + \tfrac{1}{6}x + \tfrac{7}{360}x^3 + \tfrac{31}{15120}x^5 + \ldots\\ + \frac{(-1)^{n+1}2(2^{2n-1}-1)B_{2n}}{(2n)!}x^{2n-1} + \ldots,\end{aligned}$$

where B_{2n} is a BERNOULLI NUMBER.

see also INVERSE COSECANT, SECANT, SINE

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Circular Functions." §4.3 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 71–79, 1972.

Spanier, J. and Oldham, K. B. "The Secant sec(x) and Cosecant csc(x) Functions." Ch. 33 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 311–318, 1987.

Coset

Consider a countable SUBGROUP H with ELEMENTS h_i and an element x not in H, then

$$xh_i \tag{1}$$

$$h_ix \tag{2}$$

for $i = 1, 2, \ldots$ are left and right cosets of the SUBGROUP H with respect to x. The coset of a SUBGROUP has the same number of ELEMENTS as the SUBGROUP. The ORDER of any SUBGROUP is a divisor of the ORDER of the GROUP. The original GROUP can be represented by

$$G = H + x_1H + x_2H + \ldots. \tag{3}$$

For G a not necessarily FINITE GROUP with H a SUBGROUP of G, define an EQUIVALENCE RELATION $x \sim y$ if $x = hy$ for some h in H. Then the EQUIVALENCE CLASSES are the left (or right, depending on convention) cosets of H in G, namely the sets

$$\{x \in G : x = ha \text{ for some } h \text{ in } H\}, \tag{4}$$

where a is an element of G.

see also EQUIVALENCE CLASS, GROUP, SUBGROUP

Cosh

see HYPERBOLIC COSINE

Cosine

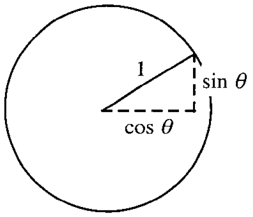

Let θ be an ANGLE measured counterclockwise from the x-axis along the arc of the unit CIRCLE. Then $\cos\theta$ is the horizontal coordinate of the arc endpoint. As a result of this definition, the cosine function is periodic with period 2π.

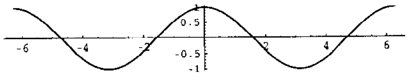

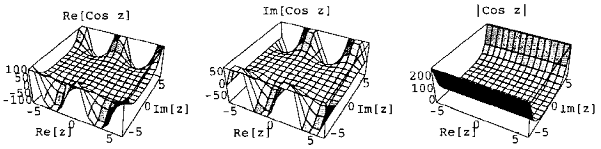

The cosine function has a FIXED POINT at 0.739085.

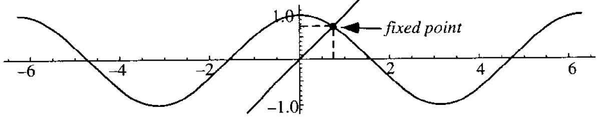

The cosine function can be defined algebraically using the infinite sum

$$\cos x \equiv \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots, \quad (1)$$

or the INFINITE PRODUCT

$$\cos x = \prod_{n=1}^{\infty} \left[1 - \frac{4x^2}{\pi^2(2n-1)^2}\right]. \quad (2)$$

A close approximation to $\cos(x)$ for $x \in [0, \pi/2]$ is

$$\cos\left(\frac{\pi}{2}x\right) \approx 1 - \frac{x^2}{x + (1-x)\sqrt{\frac{2-x}{3}}} \quad (3)$$

(Hardy 1959). The difference between $\cos x$ and Hardy's approximation is plotted below.

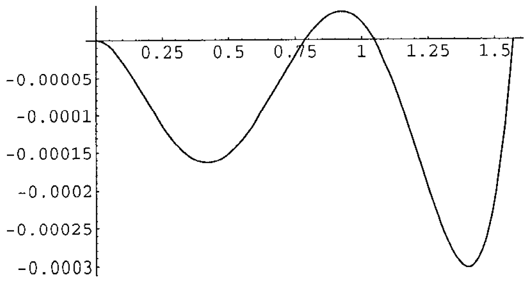

The FOURIER TRANSFORM of $\cos(2\pi k_0 x)$ is given by

$$\begin{aligned}\mathcal{F}[\cos(2\pi k_0 x)] &= \int_{-\infty}^{\infty} e^{-2\pi i k x} \cos(2\pi k_0 x)\,dx \\ &= \tfrac{1}{2}[\delta(k-k_0) + \delta(k+k_0)], \quad (4)\end{aligned}$$

where $\delta(k)$ is the DELTA FUNCTION.

The cosine sum rule gives an expansion of the COSINE function of a multiple ANGLE in terms of a sum of POWERS of sines and cosines,

$$\begin{aligned}\cos(n\theta) &= 2\cos\theta\cos[(n-1)\theta] - \cos[(n-2)\theta] \\ &= \cos^n\theta - \binom{n}{2}\cos^{n-2}\theta\sin^2\theta \\ &\quad + \binom{n}{4}\cos^{n-4}\theta\sin^4\theta - \ldots. \quad (5)\end{aligned}$$

Summing the COSINE of a multiple angle from $n = 0$ to $N-1$ can be done in closed form using

$$\sum_{n=0}^{N-1} \cos(nx) = \Re\left[\sum_{n=0}^{N-1} e^{inx}\right]. \quad (6)$$

The EXPONENTIAL SUM FORMULAS give

$$\begin{aligned}\sum_{n=0}^{N-1} \cos(nx) &= \Re\left[\frac{\sin(\frac{1}{2}Nx)}{\sin(\frac{1}{2}x)} e^{i(N-1)x/2}\right] \\ &= \frac{\sin(\frac{1}{2}Nx)}{\sin(\frac{1}{2}x)} \cos[\tfrac{1}{2}x(N-1)]. \quad (7)\end{aligned}$$

Similarly,

$$\sum_{n=0}^{\infty} p^n \cos(nx) = \Re\left[\sum_{n=0}^{\infty} p^n e^{inx}\right], \quad (8)$$

where $|p| < 1$. The EXPONENTIAL SUM FORMULA gives

$$\begin{aligned}\sum_{n=0}^{\infty} p^n \cos(nx) &= \Re\left[\frac{1 - pe^{-ix}}{1 - 2p\cos x + p^2}\right] \\ &= \frac{1 - p\cos x}{1 - 2p\cos x + p^2}. \quad (9)\end{aligned}$$

Cvijović and Klinowski (1995) note that the following series

$$C_\nu(\alpha) = \sum_{k=0}^{\infty} \frac{\cos(2k+1)\alpha}{(2k+1)^\nu} \quad (10)$$

has closed form for $\nu = 2n$,

$$C_{2n}(\alpha) = \frac{(-1)^n}{4(2n-1)!}\pi^{2n} E_{2n-1}\left(\frac{\alpha}{\pi}\right), \quad (11)$$

where $E_n(x)$ is an EULER POLYNOMIAL.

see also EULER POLYNOMIAL, EXPONENTIAL SUM FORMULAS, FOURIER TRANSFORM—COSINE, HYPERBOLIC COSINE, SINE, TANGENT, TRIGONOMETRIC FUNCTIONS

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Circular Functions." §4.3 in *Handbook of Mathematical Functions with*

Formulas, Graphs, and Mathematical Tables, 9th printing. New York: Dover, pp. 71–79, 1972.
Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, p. 68, 1959.
Cvijović, D. and Klinowski, J. "Closed-Form Summation of Some Trigonometric Series." *Math. Comput.* **64**, 205–210, 1995.
Hansen, E. R. *A Table of Series and Products.* Englewood Cliffs, NJ: Prentice-Hall, 1975.
Project Mathematics! *Sines and Cosines, Parts I–III.* Videotapes (28, 30, and 30 minutes). California Institute of Technology. Available from the Math. Assoc. Amer.
Spanier, J. and Oldham, K. B. "The Sine sin(x) and Cosine cos(x) Functions." Ch. 32 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 295–310, 1987.

Cosine Apodization Function

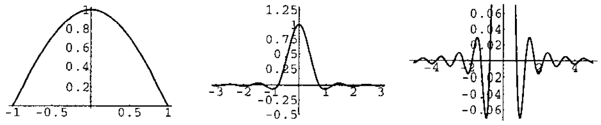

The APODIZATION FUNCTION

$$A(x) = \cos\left(\frac{\pi x}{2a}\right).$$

Its FULL WIDTH AT HALF MAXIMUM is $4a/3$. Its INSTRUMENT FUNCTION is

$$I(k) = \frac{4a\cos(2\pi ak)}{\pi(1-16a^2k^2)}.$$

see also APODIZATION FUNCTION

Cosine Circle

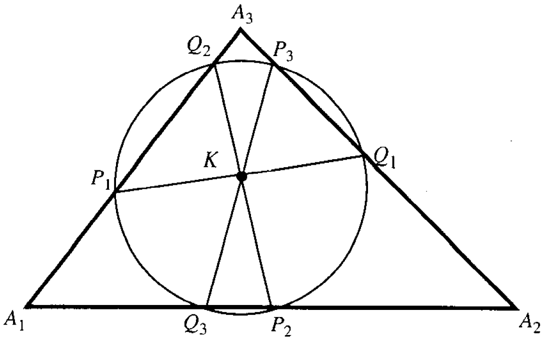

Also called the second LEMOINE CIRCLE. Draw lines through the LEMOINE POINT K and PARALLEL to the sides of the TRIANGLES. The points where the antiparallel lines intersect the sides then lie on a CIRCLE known as the cosine circle with center at K. The CHORDS P_2Q_3, P_3Q_1, and P_1Q_2 are proportional to the COSINES of the ANGLES of $\Delta A_1A_2A_3$, giving the circle its name.

TRIANGLES $P_1P_2P_3$ and $\Delta A_1A_2A_3$ are directly similar, and TRIANGLES $\Delta Q_1Q_2Q_3$ and $A_1A_2A_3$ are similar. The MIQUEL POINT of $\Delta P_1P_2P_3$ is at the BROCARD POINT Ω of $\Delta P_1P_2P_3$.

see also BROCARD POINTS, LEMOINE CIRCLE, MIQUEL POINT, TUCKER CIRCLES

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 271–273, 1929.

Cosine Integral

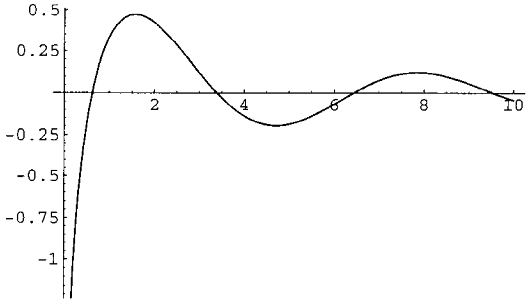

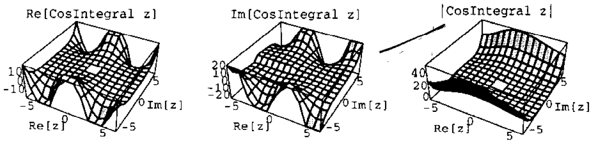

There are (at least) three types of "cosine integrals," denoted ci(x), Ci(x), and Cin(x):

$$\operatorname{ci}(x) \equiv -\int_x^\infty \frac{\cos t\,dt}{t} \quad (1)$$

$$= \tfrac{1}{2}[\operatorname{ei}(ix)+\operatorname{ei}(-ix)] \quad (2)$$

$$= -\tfrac{1}{2}[\mathrm{E}_1(ix)+\mathrm{E}_1(-ix)], \quad (3)$$

$$\operatorname{Ci}(x) \equiv \gamma + \ln z + \int_0^z \frac{\cos t - 1}{t}\,dt \quad (4)$$

$$\operatorname{Cin}(x) \equiv \int_0^z \frac{(1-\cos t)\,dt}{t} \quad (5)$$

$$= -\operatorname{Ci}(x) + \ln x + \gamma. \quad (6)$$

Here, ei(x) is the EXPONENTIAL INTEGRAL, $\mathrm{E}_n(x)$ is the E_n-FUNCTION, and γ is the EULER-MASCHERONI CONSTANT. ci(x) is the function returned by the *Mathematica*® (Wolfram Research, Champaign, IL) command `CosIntegral[x]` and displayed above.

To compute the integral of an EVEN power times a cosine,

$$I \equiv \int x^{2n}\cos(mx)\,dx, \quad (7)$$

use INTEGRATION BY PARTS. Let

$$u = x^{2n} \qquad dv = \cos(mx)\,dx \quad (8)$$

$$du = 2nx^{2n-1}\,dx \qquad v = \frac{1}{m}\sin(mx), \quad (9)$$

so

$$I = \frac{1}{m}x^{2n}\sin(mx) - \frac{2n}{m}\int x^{2n-1}\sin(mx)\,dx. \tag{10}$$

Using INTEGRATION BY PARTS again,

$$u = x^{2n-1} \qquad dv = \sin(mx)\,dx \tag{11}$$

$$du = (2n-1)x^{2n-2}\,dx \qquad v = -\frac{1}{m}\cos(mx), \tag{12}$$

and

$$\begin{aligned}
&\int x^{2n}\cos(mx)\,dx \\
&\quad = \frac{1}{m}x^{2n}\sin(mx) - \frac{2n}{m}\left[-\frac{1}{m}x^{2n-1}\cos(mx)\right. \\
&\qquad\qquad \left. + \frac{2n-1}{m}\int x^{2n-2}\cos(mx)\,dx\right] \\
&\quad = \frac{1}{m}x^{2n}\sin(mx) + \frac{2n}{m^2}x^{2n-1}\cos(mx) \\
&\qquad\qquad - \frac{(2n)(2n-1)}{m^2}\int x^{2n-2}\cos(mx)\,dx \\
&\quad = \frac{1}{m}x^{2n}\sin(mx) + \frac{2n}{m^2}x^{2n-1}\cos(mx) \\
&\qquad\qquad + \ldots + \frac{(2n)!}{m^{2n}}\int x^0\cos(mx)\,dx \\
&\quad = \frac{1}{m}x^{2n}\sin(mx) + \frac{2n}{m^2}x^{2n-1}\cos(mx) \\
&\qquad\qquad + \ldots + \frac{(2n)!}{m^{2n+1}}\sin(mx) \\
&= \sin(mx)\sum_{k=0}^{n}(-1)^{k+1}\frac{(2n)!}{(2n-2k)!m^{2k+1}}x^{2n-2k} \\
&\quad + \cos(mx)\sum_{k=1}^{n}(-1)^{k+1}\frac{(2n)!}{(2k-2n-1)!m^{2k}}x^{2n-2k+1}.
\end{aligned} \tag{13}$$

Letting $k' \equiv n-k$,

$$\begin{aligned}
&\int x^{2n}\cos(mx)\,dx \\
&\quad = \sin(mx)\sum_{k=0}^{n}(-1)^{n-k+1}\frac{(2n)!}{(2k)!m^{2n-2k+1}}x^{2k} \\
&\qquad + \cos(mx)\sum_{k=0}^{n-1}(-1)^{n-k+1}\frac{(2n)!}{(2k-1)!m^{2n-2k}}x^{2k+1} \\
&= (-1)^{n+1}(2n)!\left[\sin(mx)\sum_{k=0}^{n}\frac{(-1)^k}{(2k)!m^{2n-2k+1}}x^{2k}\right. \\
&\quad \left. + \cos(mx)\sum_{k=1}^{n}\frac{(-1)^{k+1}}{(2k-3)!m^{2n-2k+2}}x^{2k-1}\right].
\end{aligned} \tag{14}$$

To find a closed form for an integral power of a cosine function,

$$I \equiv \int \cos^m x\,dx, \tag{15}$$

perform an INTEGRATION BY PARTS so that

$$u = \cos^{m-1}x \qquad dv = \cos x\,dx \tag{16}$$

$$du = -(m-1)\cos^{m-2}x\sin x\,dx \qquad v = \sin x. \tag{17}$$

Therefore

$$\begin{aligned}
I &= \sin x\cos^{m-1}x + (m-1)\int \cos^{m-2}x\sin^2 x\,dx \\
&= \sin x\cos^{m-1}x \\
&\qquad + (m-1)\left[\int \cos^{m-2}x\,dx - \int \cos^m x\,dx\right] \\
&= \sin x\cos^{m-1}x + (m-1)\left[\int \cos^{m-2}x\,dx - I\right],
\end{aligned} \tag{18}$$

so

$$I[1+(m-1)] = \sin x\cos^{m-1}x + (m-1)\int \cos^{m-2}x\,dx \tag{19}$$

$$\begin{aligned}
I &= \int \cos^m x\,dx \\
&= \frac{\sin x\cos^{m-1}x}{m} + \frac{m-1}{m}\int \cos^{m-2}x\,dx.
\end{aligned} \tag{20}$$

Now, if m is EVEN so $m \equiv 2n$, then

$$\begin{aligned}
&\int \cos^{2n}x\,dx \\
&\qquad = \frac{\sin x\cos^{2n-1}x}{2n} + \frac{2n-1}{2n}\int \cos^{2n-2}x\,dx \\
&= \frac{\sin x\cos^{2n-1}x}{2n} + \frac{2n-1}{n}\left[\frac{\sin x\cos^{2n-3}x}{2n-2}\right. \\
&\qquad \left. + \frac{2n-3}{2n-2}\int \cos^{2n-4}x\,dx\right] \\
&= \sin x\left[\frac{1}{2n}\cos^{2n-1}x + \frac{2n-1}{(2n)(2n-2)}\cos^{2n-3}x\right] \\
&\qquad + \frac{(2n-1)(2n-3)}{(2n)(2n-2)}\int \cos^{2n-4}x\,dx \\
&= \sin x\left[\frac{1}{2n}\cos^{2n-1}x\right. \\
&\qquad \left. + \frac{2n-1}{(2n)(2n-2)}\cos^{2n-3}x + \ldots\right] \\
&\qquad + \frac{(2n-1)(2n-3)\cdots 1}{(2n)(2n-2)\cdots 2}\int \cos^0 x\,dx \\
&= \sin x\sum_{k=1}^{n}\frac{(2n-2k)!!}{(2n)!!}\frac{(2n-1)!!}{(2n-2k+1)!!}\cos^{2n-2k+1}x \\
&\qquad + \frac{(2n-1)!!}{(2n)!!}x.
\end{aligned} \tag{21}$$

Now let $k' \equiv n - k + 1$, so $n - k = k' - 1$,

$$\begin{aligned} \int \cos^{2n} x\,dx &= \sin x \sum_{k=1}^{n} \frac{(2k-2)!!}{(2n)!!} \frac{(2n-1)!!}{(2k-1)!!} \cos^{2k-1} x + \frac{(2n-1)!!}{(2n)!!} x \\ &= \frac{(2n-1)!!}{(2n)!!} \left[\sin x \sum_{k=0}^{n-1} \frac{(2k)!!}{(2k+1)!!} \cos^{2k+1} x + x \right]. \end{aligned} \qquad (22)$$

Now if m is ODD so $m \equiv 2n + 1$, then

$$\begin{aligned} \int \cos^{2n+1} x\,dx &= \frac{\sin x \cos^{2n} x}{2n+1} + \frac{2n}{2n+1} \int \cos^{2n-1} x\,dx \\ &= \frac{\sin x \cos^{2n} x}{2n+1} + \frac{2n}{2n+1} \left[\frac{\sin x \cos^{2n-2} x}{2n-1} \right. \\ &\quad \left. + \frac{2n-2}{2n-1} \int \cos^{2n-3} x\,dx \right] \\ &= \sin x \left[\frac{1}{2n+1} \cos^{2n} x + \frac{2n}{(2n+1)(2n-1)} \cos^{2n-2} x \right] \\ &\quad + \frac{(2n)(2n-2)}{(2n+1)(2n-1)} \int \cos^{2n-3} x\,dx \\ &= \sin x \left[\frac{1}{2n+1} \cos^{2n} x \right. \\ &\quad \left. + \frac{2n}{(2n+1)(2n-1)} \cos^{2n-2} x + \ldots \right] \\ &\quad + \frac{(2n)(2n-2)\cdots 2}{(2n+1)(2n-1)\cdots 3} \int \cos x\,dx \\ &= \sin x \sum_{k=0}^{n} \frac{(2n-2k-1)!!}{(2n+1)!!} \frac{(2n)!!}{(2n-2k)!!} \cos^{2n-2k} x. \end{aligned} \qquad (23)$$

Now let $k' \equiv n - k$,

$$\int \cos^{2n} x\,dx = \frac{(2n)!!}{(2n+1)!!} \sin x \sum_{k=0}^{n} \frac{(2k-1)!!}{(2k)!!} \cos^{2k} x. \qquad (24)$$

The general result is then

$$\int \cos^m x\,dx = \begin{cases} \frac{(2n-1)!!}{(2n)!!} \left[\sin x \sum_{k=0}^{n-1} \frac{(2k)!!}{(2k+1)!!} \cos^{2k+1} x + x \right] \\ \qquad \text{for } m = 2n \\ \frac{(2n)!!}{(2n+1)!!} \sin x \sum_{k=0}^{n} \frac{(2k-1)!!}{(2k)!!} \cos^{2k} x \\ \qquad \text{for } m = 2n + 1. \end{cases} \qquad (25)$$

The infinite integral of a cosine times a Gaussian can also be done in closed form,

$$\int_{-\infty}^{\infty} e^{-ax^2} \cos(kx)\,dx = \sqrt{\frac{\pi}{a}}\, e^{-k^2/4a}. \qquad (26)$$

see also CHI, DAMPED EXPONENTIAL COSINE INTEGRAL, NIELSEN'S SPIRAL, SHI, SICI SPIRAL, SINE INTEGRAL

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Sine and Cosine Integrals." §5.2 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 231–233, 1972.

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 342–343, 1985.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Fresnel Integrals, Cosine and Sine Integrals." §6.79 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 248–252, 1992.

Spanier, J. and Oldham, K. B. "The Cosine and Sine Integrals." Ch. 38 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 361–372, 1987.

Cosines Law

see LAW OF COSINES

Cosmic Figure

see PLATONIC SOLID

Cosmological Theorem

There exists an INTEGER N such that every string in the LOOK AND SAY SEQUENCE "decays" in at most N days to a compound of "common" and "transuranic elements."

The table below gives the periodic table of atoms associated with the LOOK AND SAY SEQUENCE as named by Conway (1987). The "abundance" is the average number of occurrences for long strings out of every million atoms. The asymptotic abundances are zero for transuranic elements, and 27.246... for arsenic (As), the next rarest element. The most common element is hydrogen (H), having an abundance of 91,970.383.... The starting element is U, represented by the string "3," and subsequent terms are those giving a description of the current term: one three (13); one one, one three (1113); three ones, one three (3113), etc.

Abundance	n	E_n	E_n is the derivate of E_{n+1}
102.56285249	92	U	3
9883.5986392	91	Pa	12
7581.9047125	90	Th	1113
6926.9352045	89	Ac	3113
5313.7894999	88	Ra	132113
4076.3134078	87	Fr	1113122113
3127.0209328	86	Rn	311311222113
2398.7998311	85	At	Ho.1322113
1840.1669683	84	Po	1113222113
1411.6286100	83	Bi	3113322113
1082.8883285	82	Pb	Pm.123222113
830.70513293	81	Tl	111213322113
637.25039755	80	Hg	31121123222113
488.84742982	79	Au	132112211213322113
375.00456738	78	Pt	111312212221121123222113
287.67344775	77	Ir	3113112211322112211213322113
220.68001229	76	Os	1321132122211322212221121123222113
169.28801808	75	Re	111312211312113221133211322112211213322113
315.56655252	74	W	Ge.Ca.312211322212221121123222113
242.07736666	73	Ta	13112221133211322112211213322113
2669.0970363	72	Hf	11132.Pa.H.Ca.W
2047.5173200	71	Lu	311312
1570.6911808	70	Yb	1321131112
1204.9083841	69	Tm	11131221133112
1098.5955997	68	Er	311311222.Ca.Co
47987.529438	67	Ho	1321132.Pm
36812.186418	66	Dy	111312211312
28239.358949	65	Tb	3113112221131112
21662.972821	64	Gd	Ho.13221133112
20085.668709	63	Eu	1113222.Ca.Co
15408.115182	62	Sm	311332
29820.456167	61	Pm	132.Ca.Zn
22875.863883	60	Nd	111312
17548.529287	59	Pr	31131112
13461.825166	58	Ce	1321133112
10326.833312	57	La	11131.H.Ca.Co
7921.9188284	56	Ba	311311
6077.0611889	55	Cs	13211321
4661.8342720	54	Xe	11131221131211
3576.1856107	53	I	311311222113111221
2743.3629718	52	Te	Ho.1322113312211
2104.4881933	51	Sb	Eu.Ca.3112221
1614.3946687	50	Sn	Pm.13211
1238.4341972	49	In	11131221
950.02745646	48	Cd	3113112211
728.78492056	47	Ag	132113212221
559.06537946	46	Pd	111312211312113211
428.87015041	45	Rh	311311222113111221131221
328.99480576	44	Ru	Ho.132211331222113112211
386.07704943	43	Tc	Eu.Ca.311322113212221
296.16736852	42	Mo	13211322211312113211
227.19586752	41	Nb	1113122113322113111221131221
174.28645997	40	Zr	Er.12322211331222113112211
133.69860315	39	Y	1112133.H.Ca.Tc
102.56285249	38	Sr	3112112.U
78.678000089	37	Rb	1321122112
60.355455682	36	Kr	11131221222112
46.299868152	35	Br	3113112211322112
35.517547944	34	Se	13211321222113222112
27.246216076	33	As	11131221131211322113322112
1887.4372276	32	Ge	31131122211311122113222.Na
1447.8905642	31	Ga	Ho.13221133122211332
23571.391336	30	Zn	Eu.Ca.Ac.H.Ca.312
18082.082203	29	Cu	131112
13871.123200	28	Ni	11133112
45645.877256	27	Co	Zn.32112
35015.858546	26	Fe	13122112
26861.360180	25	Mn	111311222112
20605.882611	24	Cr	31132.Si
15807.181592	23	V	13211312
12126.002783	22	Ti	11131221131112
9302.0974443	21	Sc	3113112221133112
56072.543129	20	Ca	Ho.Pa.H.12.Co
43014.360913	19	K	1112
32997.170122	18	Ar	3112
25312.784218	17	Cl	132112
19417.939250	16	S	1113122112
14895.886658	15	P	311311222112
32032.812960	14	Si	Ho.1322112
24573.006696	13	Al	1113222112
18850.441228	12	Mg	3113322112
14481.448773	11	Na	Pm.123222112
11109.006696	10	Ne	111213322112
8521.9396539	9	F	31121123222112
6537.3490750	8	O	132112211213322112
5014.9302464	7	N	111312212221121123222112
3847.0525419	6	C	3113112211322112211213322112
2951.1503716	5	B	1321132122211322212221121123222112
2263.8860325	4	Be	111312211312113221133211322112211213322112
4220.0665982	3	Li	Ge.Ca.312211322212221121123222122
3237.2968588	2	He	13112221133211322112211213322112
91790.383216	1	H	Hf.Pa.22.Ca.Li

see also Conway's Constant, Look and Say Sequence

References

Conway, J. H. "The Weird and Wonderful Chemistry of Audioactive Decay." §5.11 in *Open Problems in Communication and Computation* (Ed. T. M. Cover and B. Gopinath). New York: Springer-Verlag, pp. 173–188, 1987.

Conway, J. H. "The Weird and Wonderful Chemistry of Audioactive Decay." *Eureka,* 5–18, 1985.

Ekhad, S. B. and Zeilberger, D. "Proof of Conway's Lost Cosmological Theorem." *Electronic Research Announcement of the Amer. Math. Soc.* **3**, 78–82, 1997. `http://www.mathtemple.edu/~zeilberg/mamarim/mamarimhtml/horton.html`.

Costa-Hoffman-Meeks Minimal Surface

see Costa Minimal Surface

Costa Minimal Surface

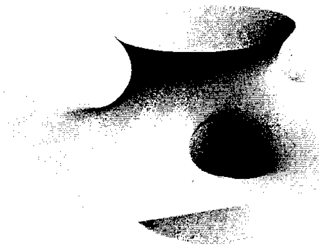

A complete embedded MINIMAL SURFACE of finite topology. It has no BOUNDARY and does not intersect itself. It can be represented parametrically by

$$\begin{aligned} x &= \tfrac{1}{2}\Re\left\{-\zeta(u+iv)+\pi u+\frac{\pi^2}{4e_1}\right. \\ &\quad \left.+\frac{\pi}{2e_1}[\zeta(u+iv-\tfrac{1}{2})-\zeta(u+iv-\tfrac{1}{2}i)]\right\} \\ y &= \tfrac{1}{2}\Re\left\{-i\zeta(u+iv)+\pi v+\frac{\pi^2}{4e_1}\right. \\ &\quad \left.-\frac{\pi}{2e_1}[i\zeta(u+iv-\tfrac{1}{2})-i\zeta(u+iv-\tfrac{1}{2}i)]\right\} \\ z &= \tfrac{1}{4}\sqrt{2\pi}\ln\left|\frac{\wp(u+iv)-e_1}{\wp(u+iv)+e_1}\right|, \end{aligned}$$

where $\zeta(z)$ is the WEIERSTRAß ZETA FUNCTION, $\wp(g_2, g_3; z)$ is the WEIERSTRAß ELLIPTIC FUNCTION, $c = 189.07272$, $e_1 = 6.87519$, and the invariants are given by $g_2 = c$ and $g_3 = 0$.

References
Costa, A. "Examples of a Complete Minimal Immersion in R^3 of Genus One and Three Embedded Ends." *Bil. Soc. Bras. Mat.* **15**, 47–54, 1984.
do Carmo, M. P. *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, p. 43, 1986.
Gray, A. *Modern Differential Geometry of Curves and Surfaces.*Boca Raton, FL: CRC Press, 1993.
Gray, A. Images of the Costa surface. `ftp://bianchi.umd.edu/pub/COSTAPS/`.
Nordstrand, T. "Costa-Hoffman-Meeks Minimal Surface." `http://www.uib.no/people/nfytn/costatxt.htm`.
Peterson, I. "The Song in the Stone: Developing the Art of Telecarving a Minimal Surface." *Sci. News* **149**, 110–111, Feb. 17, 1996.

Cosymmedian Triangles

Extend the SYMMEDIAN LINES of a TRIANGLE $\Delta A_1A_2A_3$ to meet the CIRCUMCIRCLE at P_1, P_2, P_3. Then the LEMOINE POINT K of $\Delta A_1A_2A_3$ is also the LEMOINE POINT of $\Delta P_1P_2P_3$. The TRIANGLES $\Delta A_1A_2A_3$ and $\Delta P_1P_2P_3$ are cosymmedian triangles, and have the same BROCARD CIRCLE, second BROCARD TRIANGLE, BROCARD ANGLE, BROCARD POINTS, and CIRCUMCIRCLE.

see also BROCARD ANGLE, BROCARD CIRCLE, BROCARD POINTS, BROCARD TRIANGLES, CIRCUMCIRCLE, LEMOINE POINT, SYMMEDIAN LINE

Cotangent

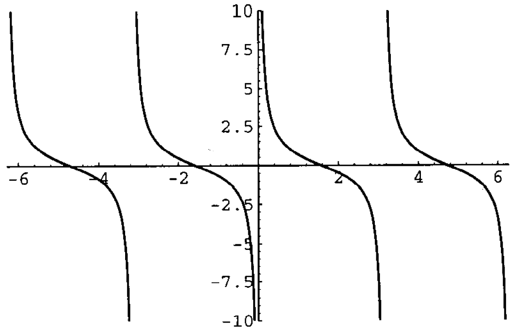

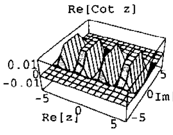

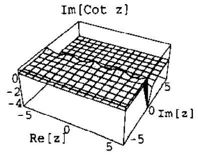

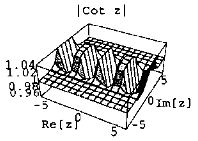

The function defined by $\cot x \equiv 1/\tan x$, where $\tan x$ is the TANGENT. The MACLAURIN SERIES for $\cot x$ is

$$\begin{aligned} \cot x &= \frac{1}{x} - \tfrac{1}{3}x - \tfrac{1}{45}x^3 - \tfrac{2}{945}x^5 - \tfrac{1}{4725}x^7 - \ldots \\ &\quad - \frac{(-1)^{n+1}2^{2n}B_{2n}}{(2n)!} - \ldots, \end{aligned}$$

where B_n is a BERNOULLI NUMBER.

$$\pi\cot(\pi x) = \frac{1}{x} + 2x\sum_{n=1}^{\infty}\frac{1}{x^2-n^2}.$$

It is known that, for $n \geq 3$, $\cot(\pi/n)$ is rational only for $n = 4$.

see also HYPERBOLIC COTANGENT, INVERSE COTANGENT, LEHMER'S CONSTANT, TANGENT

References
Abramowitz, M. and Stegun, C. A. (Eds.). "Circular Functions." §4.3 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 71–79, 1972.
Spanier, J. and Oldham, K. B. "The Tangent tan(x) and Cotangent cot(x) Functions." Ch. 34 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 319–330, 1987.

Cotangent Bundle

The cotangent bundle of a MANIFOLD is similar to the TANGENT BUNDLE, except that it is the set (x, f) where $x \in M$ and f is a dual vector in the TANGENT SPACE to $x \in M$. The cotangent bundle is denoted by T^*M.

see also TANGENT BUNDLE

Cotes Circle Property

$$x^{2n}+1=\left[x^2-2x\cos\left(\frac{\pi}{2n}\right)+1\right]\times\left[x^2-2x\cos\left(\frac{3\pi}{2n}\right)+1\right]\times\cdots\times\times\left[x^2-2x\cos\left(\frac{(2n-1)\pi}{2n}\right)+1\right].$$

Cotes Number

The numbers $\lambda_{\nu n}$ in the GAUSSIAN QUADRATURE formula

$$Q_n(f)=\sum_{\nu=1}^{n}\lambda_{\nu n}f(x_{\nu n}).$$

see also GAUSSIAN QUADRATURE

References
Cajori, F. *A History of Mathematical Notations, Vols. 1–2.* New York: Dover, p. 42, 1993.

Cotes' Spiral

The planar orbit of a particle under a r^{-3} force field. It is an EPISPIRAL.

Coth

see HYPERBOLIC COTANGENT.

Coulomb Wave Function

A special case of the CONFLUENT HYPERGEOMETRIC FUNCTION OF THE FIRST KIND. It gives the solution to the radial Schrödinger equation in the Coulomb potential $(1/r)$ of a point nucleus

$$\frac{d^2W}{d\rho^2}+\left[1-\frac{2\eta}{\rho}-\frac{L(L+1)}{\rho^2}\right]W=0. \tag{1}$$

The complete solution is

$$W=C_1F_L(\eta,\rho)+C_2G_L(\eta,\rho). \tag{2}$$

The Coulomb function of the first kind is

$$F_L(\eta,\rho)=C_L(\eta)\rho^{L+1}e^{-i\rho}{}_1F_1(L+1-i\eta;2L+2;2i\rho), \tag{3}$$

where

$$C_L(\eta)\equiv\frac{2^Le^{-\pi\eta/2}|\Gamma(L+1+i\eta)|}{\Gamma(2L+2)}, \tag{4}$$

${}_1F_1(a;b;z)$ is the CONFLUENT HYPERGEOMETRIC FUNCTION, $\Gamma(z)$ is the GAMMA FUNCTION, and the Coulomb function of the second kind is

$$G_L(\eta,\rho)=\frac{2\eta}{{C_0}^2(\eta)}F_L(\eta,\rho)\left[\ln(2\rho)+\frac{q_L(\eta)}{p_L(\eta)}\right]+\frac{1}{(2L+1)C_L(\eta)}\rho^{-L}\sum_{K=-L}^{\infty}a_k^L(\eta)\rho^{K+L}, \tag{5}$$

where q_L, p_L, and a_k^L are defined in Abramowitz and Stegun (1972, p. 538).

References
Abramowitz, M. and Stegun, C. A. (Eds.). "Coulomb Wave Functions." Ch. 14 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 537–544, 1972.
Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 631–633, 1953.

Count

The largest n such that $|z_n|<4$ in a MANDELBROT SET. Points of different count are often assigned different colors.

Countable Additivity Probability Axiom

For a COUNTABLE SET of n disjoint events E_1, E_2, ..., E_n

$$P\left(\bigcup_{i=1}^{n}E_i\right)=\sum_{i=1}^{n}P(E_i).$$

see also COUNTABLE SET

Countable Set

A SET which is either FINITE or COUNTABLY INFINITE.

see also ALEPH-0, ALEPH-1, COUNTABLY INFINITE SET, FINITE, INFINITE, UNCOUNTABLY INFINITE SET

Countable Space

see FIRST-COUNTABLE SPACE

Countably Infinite Set

Any SET which can be put in a ONE-TO-ONE correspondence with the NATURAL NUMBERS (or INTEGERS), and so has CARDINAL NUMBER $\aleph_0$. Examples of countable sets include the INTEGERS and ALGEBRAIC NUMBERS. Georg Cantor showed that the number of REAL NUMBERS is rigorously larger than a countably infinite set, and the postulate that this number, the "CONTINUUM," is equal to $\aleph_1$ is called the CONTINUUM HYPOTHESIS.

see also ALEPH-0, ALEPH-1, CANTOR DIAGONAL SLASH, CARDINAL NUMBER, CONTINUUM HYPOTHESIS, COUNTABLE SET,

Counting Generalized Principle

If r experiments are performed with n_i possible outcomes for each experiment $i=1,2,\ldots,r$, then there are a total of $\prod_{i=1}^{r}n_i$ possible outcomes.

Counting Number

A POSITIVE INTEGER: 1, 2, 3, 4, ... (Sloane's A000027), also called a NATURAL NUMBER. However, 0 is sometimes also included in the list of counting numbers. Due to lack of standard terminology, the following terms are recommended in preference to "counting number," "NATURAL NUMBER," and "WHOLE NUMBER."

Set	Name	Symbol
$\ldots, -2, -1, 0, 1, 2, \ldots$	integers	$\mathbb{Z}$
$1, 2, 3, 4, \ldots$	positive integers	$\mathbb{Z}^+$
$0, 1, 2, 3, 4 \ldots$	nonnegative integers	$\mathbb{Z}^*$
$-1, -2, -3, -4, \ldots$	negative integers	$\mathbb{Z}^-$

see also NATURAL NUMBER, WHOLE NUMBER, $\mathbb{Z}$, $\mathbb{Z}^-$, $\mathbb{Z}^+$, $\mathbb{Z}^*$

References
Sloane, N. J. A. Sequence A000027/M0472 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Coupon Collector's Problem

Let n objects be picked repeatedly with probability p_i that object i is picked on a given try, with

$$\sum_i p_i = 1.$$

Find the earliest time at which all n objects have been picked at least once.

References
Hildebrand, M. V. "The Birthday Problem." *Amer. Math. Monthly* **100**, 643, 1993.

Covariance

Given n sets of variates denoted $\{x_1\}, \ldots, \{x_n\}$, a quantity called the COVARIANCE MATRIX is defined by

$$\begin{aligned} V_{ij} &= \operatorname{cov}(x_i, x_j) & (1)\\ &\equiv \langle (x_i - \mu_i)(x_j - \mu_j) \rangle & (2)\\ &= \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle , & (3)\end{aligned}$$

where $\mu_i = \langle x_i \rangle$ and $\mu_j = \langle x_j \rangle$ are the MEANS of x_i and x_j, respectively. An individual element V_{ij} of the COVARIANCE MATRIX is called the covariance of the two variates x_i and x_j, and provides a measure of how strongly correlated these variables are. In fact, the derived quantity

$$\operatorname{cor}(x_i, x_j) \equiv \frac{\operatorname{cov}(x_i, x_j)}{\sigma_i \sigma_j}, \qquad (4)$$

where σ_i, σ_j are the STANDARD DEVIATIONS, is called the CORRELATION of x_i and x_j. Note that if x_i and x_j are taken from the same set of variates (say, x), then

$$\operatorname{cov}(x, x) = \langle x^2 \rangle - \langle x \rangle^2 = \operatorname{var}(x), \qquad (5)$$

giving the usual VARIANCE $\operatorname{var}(x)$. The covariance is also symmetric since

$$\operatorname{cov}(x, y) = \operatorname{cov}(y, x). \qquad (6)$$

For two variables, the covariance is related to the VARIANCE by

$$\operatorname{var}(x + y) = \operatorname{var}(x) + \operatorname{var}(y) + 2 \operatorname{cov}(x, y). \qquad (7)$$

For two independent variates $x = x_i$ and $y = x_j$,

$$\operatorname{cov}(x, y) = \langle xy \rangle - \mu_x \mu_y = \langle x \rangle \langle y \rangle - \mu_x \mu_y = 0, \qquad (8)$$

so the covariance is zero. However, if the variables are correlated in some way, then their covariance will be NONZERO. In fact, if $\operatorname{cov}(x, y) > 0$, then y tends to increase as x increases. If $\operatorname{cov}(x, y) < 0$, then y tends to decrease as x increases.

The covariance obeys the identity

$$\begin{aligned} \operatorname{cov}(x + z, y) &= \langle (x + z)y - \langle x + z \rangle \langle y \rangle \rangle \\ &= \langle xy \rangle + \langle zy \rangle - (\langle x \rangle + \langle z \rangle) \langle y \rangle \\ &= \langle xy \rangle - \langle x \rangle \langle y \rangle + \langle zy \rangle - \langle z \rangle \langle y \rangle \\ &= \operatorname{cov}(x, y) + \operatorname{cov}(z, y). & (9)\end{aligned}$$

By induction, it therefore follows that

$$\begin{aligned} \operatorname{cov}\left(\sum_{i=1}^n x_i, y\right) &= \sum_{i=1}^n \operatorname{cov}(x_i, y) & (10)\\ \operatorname{cov}\left(\sum_{i=1}^n x_i, \sum_{j=1}^m y_j\right) &= \sum_{i=1}^n \operatorname{cov}\left(x_i, \sum_{j=1}^m y_j\right) & (11)\\ &= \sum_{i=1}^n \operatorname{cov}\left(\sum_{j=1}^m y_j, x_i\right) & (12)\\ &= \sum_{i=1}^n \sum_{j=1}^m \operatorname{cov}(y_j, x_i) & (13)\\ &= \sum_{i=1}^n \sum_{j=1}^m \operatorname{cov}(x_i, y_j). & (14)\end{aligned}$$

see also CORRELATION (STATISTICAL), COVARIANCE MATRIX, VARIANCE

Covariance Matrix

Given n sets of variates denoted $\{x_1\}, \ldots, \{x_n\}$, the first-order covariance matrix is defined by

$$V_{ij} = \operatorname{cov}(x_i, x_j) \equiv \langle (x_i - \mu_i)(x_j - \mu_j) \rangle ,$$

where μ_i is the MEAN. Higher order matrices are given by

$$V_{ij}^{mn} = \langle (x_i - \mu_i)^m (x_j - \mu_j)^n \rangle .$$

An individual matrix element $V_{ij} = \operatorname{cov}(x_i, x_j)$ is called the COVARIANCE of x_i and x_j.

see also CORRELATION (STATISTICAL), COVARIANCE, VARIANCE

Covariant Derivative

The covariant derivative of a TENSOR A^α (also called the SEMICOLON DERIVATIVE since its symbol is a semicolon)

$$A^\alpha{}_{;\alpha} = \nabla \cdot \mathbf{A} = A^k_{,k} + \Gamma^k_{jk} A^j, \tag{1}$$

and of A_j is

$$A_{j;k} = \frac{1}{g^{kk}} \frac{\partial A_j}{\partial x_k} - \Gamma^i_{jk} A_i, \tag{2}$$

where Γ is a CONNECTION COEFFICIENT.

see also CONNECTION COEFFICIENT, COVARIANT TENSOR, DIVERGENCE

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 48–50, 1953.

Covariant Tensor

A covariant tensor is a TENSOR having specific transformation properties (c.f., a CONTRAVARIANT TENSOR). To examine the transformation properties of a covariant tensor, first consider the GRADIENT

$$\nabla\phi \equiv \frac{\partial\phi}{\partial x_1}\hat{\mathbf{x}}_1 + \frac{\partial\phi}{\partial x_2}\hat{\mathbf{x}}_2 + \frac{\partial\phi}{\partial x_3}\hat{\mathbf{x}}_3, \tag{1}$$

for which

$$\frac{\partial\phi'}{\partial x'_i} = \frac{\partial\phi}{\partial x_j}\frac{\partial x_j}{\partial x'_i}, \tag{2}$$

where $\phi(x_1, x_2, x_3) = \phi'(x'_1, x'_2, x'_3)$. Now let

$$A_i \equiv \frac{\partial\phi}{\partial x_i}, \tag{3}$$

then any set of quantities A_j which transform according to

$$A'_i = \frac{\partial x_j}{\partial x'_j} A'_j \tag{4}$$

or, defining

$$a_{ij} \equiv \frac{\partial x_j}{\partial x'_i}, \tag{5}$$

according to

$$A_i = a_{ij} A'_j \tag{6}$$

is a covariant tensor. Covariant tensors are indicated with lowered indices, i.e., a_μ.

CONTRAVARIANT TENSORS are a type of TENSOR with differing transformation properties, denoted a^ν. However, in 3-D CARTESIAN COORDINATES,

$$\frac{\partial x_j}{\partial x'_i} = \frac{\partial x'_i}{\partial x_j} \equiv a_{ij} \tag{7}$$

for $i, j = 1, 2, 3$, meaning that contravariant and covariant tensors are equivalent. The two types of tensors do differ in higher dimensions, however. Covariant FOUR-VECTORS satisfy

$$a_\mu = \Lambda^\nu_\mu a_\nu, \tag{8}$$

where Λ is a LORENTZ TENSOR.

To turn a CONTRAVARIANT TENSOR into a covariant tensor, use the METRIC TENSOR $g_{\mu\nu}$ to write

$$a_\mu \equiv g_{\mu\nu} a^\nu. \tag{9}$$

Covariant and contravariant indices can be used simultaneously in a MIXED TENSOR.

see also CONTRAVARIANT TENSOR, FOUR-VECTOR, LORENTZ TENSOR, METRIC TENSOR, MIXED TENSOR, TENSOR

References

Arfken, G. "Noncartesian Tensors, Covariant Differentiation." §3.8 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 158–164, 1985.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 44–46, 1953.

Covariant Vector

A COVARIANT TENSOR of RANK 1.

Cover

A group C of SUBSETS of X whose UNION contains the given set X ($\cup\{S : S \in C\} = X$) is called a cover (or a COVERING). A MINIMAL COVER is a cover for which removal of one member destroys the covering property. There are various types of specialized covers, including proper covers, antichain covers, minimal covers, k-covers, and k^*-covers. The number of possible covers for a set of N elements is

$$|C(N)| = \frac{1}{2}\sum_{k=0}^{N} (-1)^k \binom{N}{k} 2^{2^{N-l}},$$

the first few of which are 1, 5, 109, 32297, 2147321017, 9223372023970362989, ... (Sloane's A003465). The number of proper covers for a set of N elements is

$$\begin{aligned} |C'(N)| &= |C(N)| - \tfrac{1}{4} 2^{2^N} \\ &= \frac{1}{2}\sum_{k=0}^{N} (-1)^k \binom{N}{k} 2^{2^{N-l}} - \frac{2^{2^N}}{4}, \end{aligned}$$

the first few of which are 0, 1, 45, 15913, 1073579193, ... (Sloane's A007537).

see also MINIMAL COVER

References

Eppstein, D. "Covering and Packing." `http://www.ics.uci.edu/~eppstein/junkyard/cover.html`.

Macula, A. J. "Covers of a Finite Set." *Math. Mag.* **67**, 141–144, 1994.

Sloane, N. J. A. Sequences A003465/M4024 and A007537/M5287 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Cover Relation

The transitive reflexive reduction of a PARTIAL ORDER. An element z of a POSET $(X, \leq)$ covers another element x provided that there exists no third element y in the poset for which $x \leq y \leq z$. In this case, z is called an "upper cover" of x and x a "lower cover" of z.

Covering

see COVER

Covering Dimension

see LEBESGUE COVERING DIMENSION

Covering System

A system of congruences a_i mod n_i with $1 \leq i \leq k$ is called a covering system if every INTEGER y satisfies $y \equiv a_i \pmod{n}$ for at least one value of i.

see also EXACT COVERING SYSTEM

References

Guy, R. K. "Covering Systems of Congruences." §F13 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 251–253, 1994.

Coversine

$$\operatorname{covers} A \equiv 1 - \sin A,$$

where $\sin A$ is the SINE.

see also EXSECANT, HAVERSINE, SINE, VERSINE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 78, 1972.

Cox's Theorem

Let $\sigma_1, \ldots, \sigma_4$ be four PLANES in GENERAL POSITION through a point P and let P_{ij} be a point on the LINE $\sigma_i \cdot \sigma_j$. Let σ_{ijk} denote the PLANE $P_{ij}P_{ik}P_{jk}$. Then the four PLANES σ_{234}, σ_{134}, σ_{124}, σ_{123} all pass through one point P_{1234}. Similarly, let $\sigma_1, \ldots, \sigma_5$ be five PLANES in GENERAL POSITION through P. Then the five points P_{2345}, P_{1345}, P_{1245}, P_{1235}, and P_{1234} all lie in one PLANE. And so on.

see also CLIFFORD'S CIRCLE THEOREM

Coxeter Diagram

see COXETER-DYNKIN DIAGRAM

Coxeter-Dynkin Diagram

A labeled graph whose nodes are indexed by the generators of a COXETER GROUP having (P_i, P_j) as an EDGE labeled by M_{ij} whenever $M_{ij} > 2$, where M_{ij} is an element of the COXETER MATRIX. Coxeter-Dynkin diagrams are used to visualize COXETER GROUPS. A Coxeter-Dynkin diagram is associated with each RATIONAL DOUBLE POINT (Fischer 1986).

see also COXETER GROUP, DYNKIN DIAGRAM, RATIONAL DOUBLE POINT

References

Arnold, V. I. "Critical Points of Smooth Functions." *Proc. Int. Congr. Math.* **1**, 19–39, 1974.
Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, pp. 12–13, 1986.

Coxeter Graph

see COXETER-DYNKIN DIAGRAM

Coxeter Group

A group generated by the elements P_i for $i = 1, \ldots, n$ subject to

$$(P_iP_j)^{M_{ij}} = 1,$$

where M_{ij} are the elements of a COXETER MATRIX. Coxeter used the NOTATION $[3^{p,q,r}]$ for the Coxeter group generated by the nodes of a Y-shaped COXETER-DYNKIN DIAGRAM whose three arms have p, q, and r EDGES. A Coxeter group of this form is finite IFF

$$\frac{1}{p+q} + \frac{1}{q+1} + \frac{1}{r+1} > 1.$$

see also BIMONSTER

References

Arnold, V. I. "Snake Calculus and Combinatorics of Bernoulli, Euler, and Springer Numbers for Coxeter Groups." *Russian Math. Surveys* **47**, 3–45, 1992.

Coxeter's Loxodromic Sequence of Tangent Circles

An infinite sequence of CIRCLES such that every four consecutive CIRCLES are mutually tangent, and the CIRCLES' RADII $\ldots$, R_{-n}, $\ldots$, R_{-1}, R_0, R_1, R_2, R_3, R_4, $\ldots$, R_n, $R_n + 1$, $\ldots$, are in GEOMETRIC PROGRESSION with ratio

$$k \equiv \frac{R_{n+1}}{R_n} = \phi + \sqrt{\phi},$$

where ϕ is the GOLDEN RATIO (Gardner 1979ab). Coxeter (1968) generalized the sequence to SPHERES.

see also ARBELOS, GOLDEN RATIO, HEXLET, PAPPUS CHAIN, STEINER CHAIN

References

Coxeter, D. "Coxeter on 'Firmament.'" `http://www.bangor.ac.uk/SculMath/image/donald.htm`.
Coxeter, H. S. M. "Loxodromic Sequences of Tangent Spheres." *Aequationes Math.* **1**, 112–117, 1968.

Gardner, M. "Mathematical Games: The Diverse Pleasures of Circles that Are Tangent to One Another." *Sci. Amer.* **240**, 18–28, Jan. 1979a.
Gardner, M. "Mathematical Games: How to be a Psychic, Even if You are a Horse or Some Other Animal." *Sci. Amer.* **240**, 18–25, May 1979b.

Coxeter Matrix

An $n \times n$ SQUARE MATRIX M with

$$M_{ii} = 1$$
$$M_{ij} = M_{ji} > 1$$

for all $i, j = 1, \ldots, n$.

see also COXETER GROUP

Coxeter-Todd Lattice

The complex LATTICE Λ_6^ω corresponding to real lattice K_{12} having the densest HYPERSPHERE PACKING (KISSING NUMBER) in 12-D. The associated AUTOMORPHISM GROUP G_0 was discovered by Mitchell (1914). The order of G_0 is given by

$$|\operatorname{Aut}(\Lambda_6^\omega)| = 2^9 \cdot 3^7 \cdot 5 \cdot 7 = 39,191,040.$$

The order of the AUTOMORPHISM GROUP of K_{12} is given by

$$|\operatorname{Aut}(K_{12})| = 2^{10} \cdot 3^7 \cdot 5 \cdot 7$$

(Conway and Sloane 1983).

see also BARNES-WALL LATTICE, LEECH LATTICE

References
Conway, J. H. and Sloane, N. J. A. "The Coxeter-Todd Lattice, the Mitchell Group and Related Sphere Packings." *Math. Proc. Camb. Phil. Soc.* **93**, 421–440, 1983.
Conway, J. H. and Sloane, N. J. A. "The 12-Dimensional Coxeter-Todd Lattice K_{12}." §4.9 in *Sphere Packings, Lattices, and Groups, 2nd ed.* New York: Springer-Verlag, pp. 127–129, 1993.
Coxeter, H. S. M. and Todd, J. A. "As Extreme Duodenary Form." *Canad. J. Math.* **5**, 384–392, 1953.
Mitchell, H. H. "Determination of All Primitive Collineation Groups in More than Four Variables." *Amer. J. Math.* **36**, 1–12, 1914.
Todd, J. A. "The Characters of a Collineation Group in Five Dimensions." *Proc. Roy. Soc. London Ser. A* **200**, 320–336, 1950.

Cramér Conjecture

An unproven CONJECTURE that

$$\overline{\lim_{n\to\infty}} \frac{p_{n+1} - p_n}{(\ln p_n)^2} = 1,$$

where p_n is the nth PRIME.

References
Cramér, H. "On the Order of Magnitude of the Difference Between Consecutive Prime Numbers." *Acta Arith.* **2**, 23–46, 1936.
Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 7, 1994.
Riesel, H. "The Cramér Conjecture." *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, pp. 79–82, 1994.
Rivera, C. "Problems & Puzzles (Conjectures): Cramer's Conjecture." `http://www.sci.net.mx/~crivera/ppp/conj_007.htm`.

Cramér-Euler Paradox

A curve of order n is generally determined by $n(n+3)/2$ points. So a CONIC SECTION is determined by five points and a CUBIC CURVE should require nine. But the MACLAURIN-BEZOUT THEOREM says that two curves of degree n intersect in n^2 points, so two CUBICS intersect in nine points. This means that $n(n+3)/2$ points do not always uniquely determine a single curve of order n. The paradox was publicized by Stirling, and explained by Plücker.

see also CUBIC CURVE, MACLAURIN-BEZOUT THEOREM

Cramer's Rule

Given a set of linear equations

$$\left\{\begin{array}{l} a_1x + b_1y + c_1z = d_1 \\ a_2x + b_2y + c_2z = d_2 \\ a_3x + b_3y + c_3z = d_3, \end{array}\right. \tag{1}$$

consider the DETERMINANT

$$D \equiv \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}. \tag{2}$$

Now multiply D by x, and use the property of DETERMINANTS that MULTIPLICATION by a constant is equivalent to MULTIPLICATION of each entry in a given row by that constant

$$x \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1x & b_1 & c_1 \\ a_2x & b_2 & c_2 \\ a_3x & b_3 & c_3 \end{vmatrix}. \tag{3}$$

Another property of DETERMINANTS enables us to add a constant times any column to any column and obtain the same DETERMINANT, so add y times column 2 and z times column 3 to column 1,

$$xD = \begin{vmatrix} a_1x + b_1y + c_1z & b_1 & c_1 \\ a_2x + b_2y + c_2z & b_2 & c_2 \\ a_3x + b_3y + c_3z & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}. \tag{4}$$

If $\mathbf{d} = \mathbf{0}$, then (4) reduces to $xD = 0$, so the system has nondegenerate solutions (i.e., solutions other than (0, 0, 0)) only if $D = 0$ (in which case there is a family of solutions). If $\mathbf{d} \neq \mathbf{0}$ and $D = 0$, the system has no unique solution. If instead $\mathbf{d} \neq \mathbf{0}$ and $D \neq 0$, then solutions are given by

$$x = \frac{\begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}}{D}, \tag{5}$$

and similarly for

$$y = \frac{\begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}}{D} \tag{6}$$

$$z = \frac{\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}}{D}. \tag{7}$$

This procedure can be generalized to a set of n equations so, given a system of n linear equations

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix}, \tag{8}$$

let

$$D \equiv \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}. \tag{9}$$

If $\mathbf{d} = \mathbf{0}$, then nondegenerate solutions exist only if $D = 0$. If $\mathbf{d} \neq \mathbf{0}$ and $D = 0$, the system has no unique solution. Otherwise, compute

$$D_k \equiv \begin{vmatrix} a_{11} & \cdots & a_{1(k-1)} & d_1 & a_{1(k+1)} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{n(k-1)} & d_n & a_{n(k+1)} & \cdots & a_{nn} \end{vmatrix}. \tag{10}$$

Then $x_k = D_k/D$ for $1 \leq k \leq n$. In the 3-D case, the VECTOR analog of Cramer's rule is

$$(\mathbf{A}\times\mathbf{B})\times(\mathbf{C}\times\mathbf{D}) = (\mathbf{A}\cdot\mathbf{B}\times\mathbf{D})\mathbf{C} - (\mathbf{A}\cdot\mathbf{B}\times\mathbf{C})\mathbf{D}. \tag{11}$$

see also DETERMINANT, LINEAR ALGEBRA, MATRIX, SYSTEM OF EQUATIONS, VECTOR

Cramér's Theorem

If X and Y are INDEPENDENT variates and $X + Y$ is a GAUSSIAN DISTRIBUTION, then both X and Y must have GAUSSIAN DISTRIBUTIONS. This was proved by Cramér in 1936.

Craps

A game played with two DICE. If the total is 7 or 11 (a "natural"), the thrower wins and retains the DICE for another throw. If the total is 2, 3, or 12 ("craps"), the thrower loses but retains the DICE. If the total is any other number (called the thrower's "point"), the thrower must continue throwing and roll the "point" value again before throwing a 7. If he succeeds, he wins and retains the DICE, but if a 7 appears first, the player loses and passes the dice. The probability of winning is $244/495 \approx 0.493$ (Kraitchik 1942).

References

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, pp. 12–13, 1951.

Kraitchik, M. "Craps." §6.5 in *Mathematical Recreations.* New York: W. W. Norton, pp. 123–126, 1942.

CRC

see CYCLIC REDUNDANCY CHECK

Creative Telescoping

see TELESCOPING SUM, ZEILBERGER'S ALGORITHM

Cremona Transformation

An entire Cremona transformation is a BIRATIONAL TRANSFORMATION of the PLANE. Cremona transformations are MAPS of the form

$$x_{i+1} = f(x_i, y_i)$$
$$y_{i+1} = g(x_i, y_i),$$

in which f and g are POLYNOMIALS. A quadratic Cremona transformation is always factorable.

see also NOETHER'S TRANSFORMATION THEOREM

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, pp. 203–204, 1959.

Cribbage

Cribbage is a game in which each of two players is dealt a hand of six CARDS. Each player then discards two of his six cards to a four-card "crib" which alternates between players. After the discard, the top card in the remaining deck is turned up. Cards are then alternating played out by the two players, with points being scored for pairs, runs, cumulative total of 15 and 31, and playing the last possible card ("go") not giving a total over 31. All face cards are counted as 10 for the purpose of playing out, but the normal values of Jack = 11, Queen = 12, King = 13 are used to determine runs. Aces are always low (ace = 1). After all cards have been played, each player counts the four cards in his hand taken in conjunction with the single top card. Points are awarded for pairs, flushes, runs, and combinations of cards giving 15. A Jack having the same suit as a top card is awarded an additional point for "nobbs." The crib is then also counted and scored. The winner is the first person to "peg" a certain score, as recorded on a "cribbage board."

The best possible score in a hand is 29, corresponding to three 5s and a Jack with a top 5 the same suit as the Jack. Hands with scores of 25, 26, and 27 are not possible.

see also BRIDGE CARD GAME, CARDS, POKER

Criss-Cross Method

A standard form of the LINEAR PROGRAMMING problem of maximizing a linear function over a CONVEX POLYHEDRON is to maximize $\mathbf{c} \cdot \mathbf{x}$ subject to $\mathsf{m}\mathbf{x} \leq \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$, where m is a given $s \times d$ matrix, $\mathbf{c}$ and $\mathbf{b}$ are given d-vector and s-vectors, respectively. The Criss-cross method always finds a VERTEX solution if an optimal solution exists.

see also CONVEX POLYHEDRON, LINEAR PROGRAMMING, VERTEX (POLYHEDRON)

Criterion

A requirement NECESSARY for a given statement or theorem to hold. Also called a CONDITION.

see also BROWN'S CRITERION, CAUCHY CRITERION, EULER'S CRITERION, GAUSS'S CRITERION, KORSELT'S CRITERION, LEIBNIZ CRITERION, POCKLINGTON'S CRITERION, VANDIVER'S CRITERIA, WEYL'S CRITERION

Critical Line

The LINE $\Re(s) = 1/2$ in the COMPLEX PLANE on which the RIEMANN HYPOTHESIS asserts that all nontrivial (COMPLEX) ROOTS of the RIEMANN ZETA FUNCTION $\zeta(s)$ lie. Although it is known that an INFINITE number of zeros lie on the critical line and that these comprise *at least* 40% of all zeros, the RIEMANN HYPOTHESIS is still unproven.

see also RIEMANN HYPOTHESIS, RIEMANN ZETA FUNCTION

References

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 142, 1991.

Critical Point

A FUNCTION $y = f(x)$ has critical points at all points x_0 where $f'(x_0) = 0$ or $f(x)$ is not DIFFERENTIABLE. A FUNCTION $z = f(x, y)$ has critical points where the GRADIENT $\nabla f = \mathbf{0}$ or $\partial f/\partial x$ or the PARTIAL DERIVATIVE $\partial f/\partial y$ is not defined.

see also FIXED POINT, INFLECTION POINT, ONLY CRITICAL POINT IN TOWN TEST, STATIONARY POINT

Critical Strip

see CRITICAL LINE

Crook

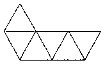

A 6-POLYIAMOND.

References

Golomb, S. W. *Polyominoes: Puzzles, Patterns, Problems, and Packings, 2nd ed.* Princeton, NJ: Princeton University Press, p. 92, 1994.

Crookedness

Let a KNOT K be parameterized by a VECTOR FUNCTION $\mathbf{v}(t)$ with $t \in \mathbb{S}^1$, and let $\mathbf{w}$ be a fixed UNIT VECTOR in $\mathbb{R}^3$. Count the number of RELATIVE MINIMA of the projection function $\mathbf{w} \cdot \mathbf{v}(t)$. Then the MINIMUM such number over all directions $\mathbf{w}$ and all K of the given type is called the crookedness $\mu(K)$. Milnor (1950) showed that $2\pi\mu(K)$ is the INFIMUM of the total curvature of K. For any TAME KNOT K in $\mathbb{R}^3$, $\mu(K) = b(K)$ where $b(K)$ is the BRIDGE INDEX.

see also BRIDGE INDEX

References

Milnor, J. W. "On the Total Curvature of Knots." *Ann. Math.* **52**, 248–257, 1950.

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 115, 1976.

Cross

In general, a cross is a figure formed by two intersecting LINE SEGMENTS. In LINEAR ALGEBRA, a cross is defined as a set of n mutually PERPENDICULAR pairs of VECTORS of equal magnitude from a fixed origin in EUCLIDEAN n-SPACE.

The word "cross" is also used to denote the operation of the CROSS PRODUCT, so $\mathbf{a} \times \mathbf{b}$ would be pronounced "$\mathbf{a}$ cross $\mathbf{b}$."

see also CROSS PRODUCT, DOT, EUTACTIC STAR, GAULLIST CROSS, GREEK CROSS, LATIN CROSS, MALTESE CROSS, PAPAL CROSS, SAINT ANDREW'S CROSS, SAINT ANTHONY'S CROSS, STAR

Cross-Cap

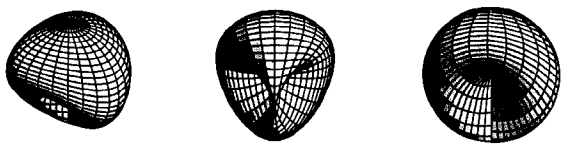

The self-intersection of a one-sided SURFACE. It can be described as a circular HOLE which, when entered, exits from its opposite point (from a topological viewpoint, both singular points on the cross-cap are equivalent). The cross-cap has a segment of double points which terminates at two "PINCH POINTS" known as WHITNEY SINGULARITIES.

The cross-cap can be generated using the general method for NONORIENTABLE SURFACES using the polynomial function

$$\mathbf{f}(x, y, z) = (xz, yz, \tfrac{1}{2}(z^2 - x^2)) \tag{1}$$

(Pinkall 1986). Transforming to SPHERICAL COORDINATES gives

$$x(u, v) = \tfrac{1}{2}\cos u \sin(2v) \tag{2}$$

$$y(u, v) = \tfrac{1}{2}\sin u \sin(2v) \tag{3}$$

$$z(u, v) = \tfrac{1}{2}(\cos^2 v - \cos^2 u \sin^2 v) \tag{4}$$

for $u \in [0, 2\pi)$ and $v \in [0, \pi/2]$. To make the equations slightly simpler, all three equations are normally multiplied by a factor of 2 to clear the arbitrary scaling constant. Three views of the cross-cap generated using this equation are shown above. Note that the middle one looks suspiciously like MAEDER'S OWL MINIMAL SURFACE.

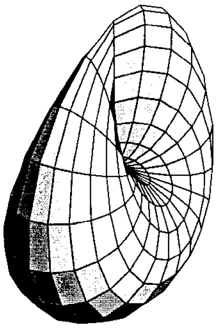

Another representation is

$$\mathbf{f}(x,y,z) = (yz, 2xy, x^2 - y^2), \tag{5}$$

(Gray 1993), giving parametric equations

$$x = \tfrac{1}{2}\sin u \sin(2v) \tag{6}$$
$$y = \sin(2u)\sin^2 v \tag{7}$$
$$z = \cos(2u)\sin^2 v, \tag{8}$$

(Geometry Center) where, for aesthetic reasons, the y- and z-coordinates have been multiplied by 2 to produce a squashed, but topologically equivalent, surface. Nordstrand gives the implicit equation

$$4x^2(x^2 + y^2 + z^2 + z) + y^2(y^2 + z^2 - 1) = 0 \tag{9}$$

which can be solved for z to yield

$$z = \frac{-2x^2 \pm \sqrt{(y^2 + 2x^2)(1 - 4x^2 - y^2)}}{4x^2 + y^2}. \tag{10}$$

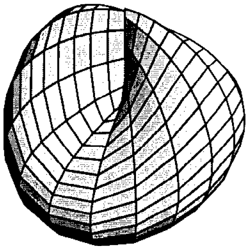

Taking the inversion of a cross-cap such that $(0, 0, -1/2)$ is sent to ∞ gives a CYLINDROID, shown above (Pinkall 1986).

The cross-cap is one of the three possible SURFACES obtained by sewing a MÖBIUS STRIP to the edge of a DISK. The other two are the BOY SURFACE and ROMAN SURFACE.

see also BOY SURFACE, MÖBIUS STRIP, NONORIENTABLE SURFACE, PROJECTIVE PLANE, ROMAN SURFACE

References

Fischer, G. (Ed.). Plate 107 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 108, 1986.

Geometry Center. "The Crosscap." `http://www.geom.umn.edu/zoo/toptype/pplane/cap/`.

Pinkall, U. *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, p. 64, 1986.

Cross-Correlation

Let $\star$ denote cross-correlation. Then the cross-correlation of two functions $f(t)$ and $g(t)$ of a real variable t is defined by

$$f \star g \equiv f^*(-t) * g(t), \tag{1}$$

where $*$ denotes CONVOLUTION and $f^*(t)$ is the COMPLEX CONJUGATE of $f(t)$. The CONVOLUTION is defined by

$$f(t) * g(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau, \tag{2}$$

therefore

$$f \star g \equiv \int_{-\infty}^{\infty} f^*(-\tau)g(t-\tau)\,d\tau. \tag{3}$$

Let $\tau' \equiv -\tau$, so $d\tau' = -d\tau$ and

$$\begin{aligned} f \star g &= \int_{\infty}^{-\infty} f^*(\tau')g(t+\tau')(-d\tau') \\ &= \int_{-\infty}^{\infty} f^*(\tau)g(t+\tau)\,d\tau. \end{aligned} \tag{4}$$

The cross-correlation satisfies the identity

$$(g \star h) \star (g \star h) = (g \star g) \star (h \star h). \tag{5}$$

If f or g is EVEN, then

$$f \star g = f * g, \tag{6}$$

where $*$ denotes CONVOLUTION.

see also AUTOCORRELATION, CONVOLUTION, CROSS-CORRELATION THEOREM

Cross-Correlation Coefficient

The COEFFICIENT ρ in a GAUSSIAN BIVARIATE DISTRIBUTION.

Cross-Correlation Theorem

Let $f \star g$ denote the CROSS-CORRELATION of functions $f(t)$ and $g(t)$. Then

$$\begin{aligned} f \star g &= \int_{-\infty}^{\infty} f^*(\tau)g(t+\tau)\,d\tau \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} F^*(\nu)e^{2\pi i\nu\tau}\,d\nu \int_{-\infty}^{\infty} G(\nu')e^{-2\pi i\nu'(t+\tau)}\,d\nu'\right]d\tau \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F^*(\nu)G(\nu')e^{-2\pi i\tau(\nu'-\nu)}e^{-2\pi i\nu' t}\,d\tau\,d\nu\,d\nu' \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F^*(\nu)G(\nu')e^{-2\pi i\nu' t}\left[\int_{-\infty}^{\infty} e^{-2\pi i\tau(\nu'-\nu)}\,d\tau\right]d\nu\,d\nu' \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F^*(\nu)G(\nu')e^{-2\pi i\nu' t}\delta(\nu'-\nu)\,d\nu'\,d\nu \\ &= \int_{-\infty}^{\infty} F^*(\nu)G(\nu)e^{-2\pi i\nu t}\,d\nu \\ &= \mathcal{F}[F^*(\nu)G(\nu)], \end{aligned} \tag{1}$$

where $\mathcal{F}$ denotes the FOURIER TRANSFORM and

$$f(t) \equiv \mathcal{F}[F(\nu)] = \int_{-\infty}^{\infty} F(\nu)e^{-2\pi i\nu t}\,dt \qquad (2)$$

$$g(t) \equiv \mathcal{F}[G(\nu)] = \int_{-\infty}^{\infty} G(\nu)e^{-2\pi i\nu t}\,dt. \qquad (3)$$

Applying a FOURIER TRANSFORM on each side gives the cross-correlation theorem,

$$f \star g = \mathcal{F}[F^*(\nu)G(\nu)]. \qquad (4)$$

If $F = G$, then the cross-correlation theorem reduces to the WIENER-KHINTCHINE THEOREM.

see also FOURIER TRANSFORM, WIENER-KHINTCHINE THEOREM

Cross Curve

see CRUCIFORM

Cross Fractal

see CANTOR SQUARE FRACTAL

Cross of Lorraine

see GAULLIST CROSS

Cross Polytope

A regular POLYTOPE in n-D (generally assumed to satisfy $n \geq 5$) corresponding to the CONVEX HULL of the points formed by permuting the coordinates ($\pm$ 1, 0, 0, ..., 0). It is denoted β_n and has SCHLÄFLI SYMBOL $\{3^{n-2}, 4\}$. In 3-D, the cross polytope is the OCTAHEDRON.

see also MEASURE POLYTOPE, SIMPLEX

Cross Product

For VECTORS $\mathbf{u}$ and $\mathbf{v}$,

$$\mathbf{u}\times\mathbf{v} = \hat{\mathbf{x}}(u_y v_z - u_z v_y) - \hat{\mathbf{y}}(u_x v_z - u_z v_x) + \hat{\mathbf{z}}(u_x v_y - u_y v_x). \qquad (1)$$

This can be written in a shorthand NOTATION which takes the form of a DETERMINANT

$$\mathbf{u}\times\mathbf{v} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ u_x & u_y & u_z \\ v_x & v_y & v_z \end{vmatrix}. \qquad (2)$$

It is also true that

$$|\mathbf{u}\times\mathbf{v}| = |\mathbf{u}|\,|\mathbf{v}|\sin\theta, \qquad (3)$$
$$= |\mathbf{u}|\,|\mathbf{v}|\sqrt{1-(\hat{\mathbf{u}}\cdot\hat{\mathbf{v}})^2}, \qquad (4)$$

where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$, given by the DOT PRODUCT

$$\cos\theta \equiv \hat{\mathbf{u}}\cdot\hat{\mathbf{v}}. \qquad (5)$$

Identities involving the cross product include

$$\frac{d}{dt}[\mathbf{r}_1(t)\times\mathbf{r}_2(t)] = \mathbf{r}_1(t)\times\frac{d\mathbf{r}_2}{dt} + \frac{d\mathbf{r}_1}{dt}\times\mathbf{r}_2(t) \qquad (6)$$

$$\mathbf{A}\times\mathbf{B} = -\mathbf{B}\times\mathbf{A} \qquad (7)$$

$$\mathbf{A}\times(\mathbf{B}+\mathbf{C}) = \mathbf{A}\times\mathbf{B} + \mathbf{A}\times\mathbf{C} \qquad (8)$$

$$(t\mathbf{A})\times\mathbf{B} = t(\mathbf{A}\times\mathbf{B}). \qquad (9)$$

For a proof that $\mathbf{A}\times\mathbf{B}$ is a PSEUDOVECTOR, see Arfken (1985, pp. 22–23). In TENSOR notation,

$$\mathbf{A}\times\mathbf{B} = \epsilon_{ijk}A^j B^k, \qquad (10)$$

where ϵ_{ijk} is the LEVI-CIVITA TENSOR.

see also DOT PRODUCT, SCALAR TRIPLE PRODUCT

References

Arfken, G. "Vector or Cross Product." §1.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 18–26, 1985.

Cross-Ratio

$$[a,b,c,d] \equiv \frac{(a-b)(c-d)}{(a-d)(c-b)}. \qquad (1)$$

For a MÖBIUS TRANSFORMATION f,

$$[a,b,c,d] = [f(a), f(b), f(c), f(d)]. \qquad (2)$$

There are six different values which the cross-ratio may take, depending on the order in which the points are chosen. Let $\lambda \equiv [a,b,c,d]$. Possible values of the cross-ratio are then λ, $1-\lambda$, $1/\lambda$, $(\lambda-1)/\lambda$, $1/(1-\lambda)$, and $\lambda/(\lambda-1)$.

Given lines a, b, c, and d which intersect in a point O, let the lines be cut by a line l, and denote the points of intersection of l with each line by A, B, C, and D. Let the distance between points A and B be denoted AB, etc. Then the cross-ratio

$$[AB, CD] \equiv \frac{(AB)(CD)}{(BC)(AD)} \qquad (3)$$

is the same for any position of the l (Coxeter and Greitzer 1967). Note that the definitions $(AB/AD)/(BC/CD)$ and $(CA/CB)/(DA/DB)$ are used instead by Kline (1990) and Courant and Robbins (1966), respectively. The identity

$$[AD, BC] + [AB, DC] = 1 \qquad (4)$$

holds IFF $AC//BD$, where $//$ denotes SEPARATION.

The cross-ratio of four points on a radial line of an INVERSION CIRCLE is preserved under INVERSION (Ogilvy 1990, p. 40).

see also MÖBIUS TRANSFORMATION, SEPARATION

References
Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, 1996.
Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 107–108, 1967.
Kline, M. *Mathematical Thought from Ancient to Modern Times, Vol. 1.* Oxford, England: Oxford University Press, 1990.
Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 39–41, 1990.

Cross-Section

The cross-section of a SOLID is a LAMINA obtained by its intersection with a PLANE. The cross-section of an object therefore represents an infinitesimal "slice" of a solid, and may be different depending on the orientation of the slicing plane. While the cross-section of a SPHERE is always a DISK, the cross-section of a CUBE may be a SQUARE, HEXAGON, or other shape.

see also AXONOMETRY, CAVALIERI'S PRINCIPLE, LAMINA, PLANE, PROJECTION, RADON TRANSFORM, STEREOLOGY

Crossed Ladders Problem

Given two crossed LADDERS resting against two buildings, what is the distance between the buildings? Let the height at which they cross be c and the lengths of the LADDERS a and b. The height at which b touches the building k is then obtained by solving

$$k^4 - 2ck^3 + k^2(a^2 - b^2) - 2ck(a^2 - b^2) + c^2(a^2 - b^2) = 0.$$

Call the horizontal distance from the top of a to the crossing u, and the distance from the top of b, v. Call the height at which a touches the building h. There are solutions in which a, b, h, k, u, and v are all INTEGERS. One is $a = 119$, $b = 70$, $c = 30$, and $u + v = 56$.

see also LADDER

References
Gardner, M. *Mathematical Circus: More Puzzles, Games, Paradoxes and Other Mathematical Entertainments from Scientific American.* New York: Knopf, pp. 62–64, 1979.

Crossed Trough

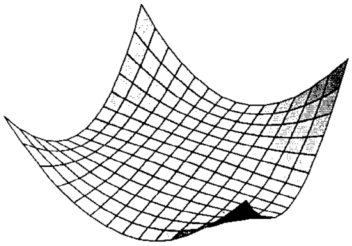

The SURFACE

$$z = cx^2y^2.$$

see also MONKEY SADDLE

References
von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 286, 1993.

Crossing Number (Graph)

Given a "good" GRAPH (i.e., one for which all intersecting EDGES intersect in a single point and arise from four distinct VERTICES), the crossing number is the minimum possible number of crossings with which the GRAPH can be drawn. A GRAPH with crossing number 0 is a PLANAR GRAPH. Garey and Johnson (1983) showed that determining the crossing number is an NP-COMPLETE PROBLEM. GUY'S CONJECTURE suggests that the crossing number for the COMPLETE GRAPH K_n is

$$\frac{1}{4}\left\lfloor\frac{n}{2}\right\rfloor\left\lfloor\frac{n-1}{2}\right\rfloor\left\lfloor\frac{n-2}{2}\right\rfloor\left\lfloor\frac{n-3}{2}\right\rfloor, \qquad (1)$$

which can be rewritten

$$\begin{cases} \frac{1}{64}n(n-2)^2(n-4) & \text{for } n \text{ even} \\ \frac{1}{64}(n-1)^2(n-3)^2 & \text{for } n \text{ odd.} \end{cases} \qquad (2)$$

The first few predicted and known values are given in the following table (Sloane's A000241).

Order	Predicted	Known
1	0	0
2	0	0
3	0	0
4	0	0
5	1	1
6	3	3
7	9	9
8	18	18
9	36	36
10	60	60
11	100	
12	150	
13	225	
14	315	
15	441	
16	588	

ZARANKIEWICZ'S CONJECTURE asserts that the crossing number for a COMPLETE BIGRAPH is

$$\left\lfloor\frac{n}{2}\right\rfloor\left\lfloor\frac{n-1}{2}\right\rfloor\left\lfloor\frac{m}{2}\right\rfloor\left\lfloor\frac{m-1}{2}\right\rfloor. \qquad (3)$$

It has been checked up to $m, n = 7$, and Zarankiewicz has shown that, in general, the FORMULA provides an upper bound to the actual number. The table below gives known results. When the number is not known exactly, the prediction of ZARANKIEWICZ'S CONJECTURE is given in parentheses.

	1	2	3	4	5	6	7
1	0	0	0	0	0	0	0
2		0	0	0	0	0	0
3			1	2	4	6	9
4				4	8	12	18
5					16	24	36
6						36	54
7							77, 79, or (81)

Consider the crossing number for a rectilinear GRAPH G which may have only straight EDGES, denoted $\bar{\nu}(G)$. For a COMPLETE GRAPH of order $n \geq 10$, the rectilinear crossing number is always larger than the general graph crossing number. The first few values for COMPLETE GRAPHS are 0, 0, 0, 0, 1, 3, 9, 19, 36, 61 or 62, ... (Sloane's A014540). The $n = 10$ lower limit is from Singer (1986), who proved that

$$\bar{\nu}(K_n) \leq \tfrac{1}{312}(5n^4 - 39n^3 + 91n^2 - 57n). \qquad (4)$$

Jensen (1971) has shown that

$$\bar{\nu}(K_n) \leq \tfrac{7}{432}n^4 + \mathcal{O}(n^3). \qquad (5)$$

Consider the crossing number for a toroidal GRAPH. For a COMPLETE GRAPH, the first few are 0, 0, 0, 0, 0, 0, 0, 4, 9, 23, 42, 70, 105, 154, 226, 326, ... (Sloane's A014543). The toroidal crossing numbers for a COMPLETE BIGRAPH are given in the following table.

	1	2	3	4	5	6	7
1	0	0	0	0	0	0	
2		0	0	0	0	0	
3			0	0	0	0	
4					2		
5					5	8	
6						12	
7							

see also GUY'S CONJECTURE, ZARANKIEWICZ'S CONJECTURE

References

Gardner, M. "Crossing Numbers." Ch. 11 in *Knotted Doughnuts and Other Mathematical Entertainments.* New York: W. H. Freeman, pp. 133–144, 1986.

Garey, M. R. and Johnson, D. S. "Crossing Number is NP-Complete." *SIAM J. Alg. Discr. Meth.* **4**, 312–316, 1983.

Guy, R. K. "Latest Results on Crossing Numbers." In *Recent Trends in Graph Theory, Proc. New York City Graph Theory Conference, 1st, 1970.* (Ed. New York City Graph Theory Conference Staff). New York: Springer-Verlag, 1971.

Guy, R. K. and Jenkyns, T. "The Toroidal Crossing Number of $K_{m,n}$." *J. Comb. Th.* **6**, 235–250, 1969.

Guy, R. K.; Jenkyns, T.; and Schaer, J. "Toroidal Crossing Number of the Complete Graph." *J. Comb. Th.* **4**, 376–390, 1968.

Jensen, H. F. "An Upper Bound for the Rectilinear Crossing Number of the Complete Graph." *J. Comb. Th. Ser. B* **10**, 212–216, 1971.

Kleitman, D. J. "The Crossing Number of $K_{5,n}$." *J. Comb. Th.* **9**, 315–323, 1970.

Singer, D. Unpublished manuscript "The Rectilinear Crossing Number of Certain Graphs," 1971. Quoted in Gardner, M. *Knotted Doughnuts and Other Mathematical Entertainments.* New York: W. H. Freeman, 1986.

Sloane, N. J. A. Sequences A014540, A014543, and A000241/M2772 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Tutte, W. T. "Toward a Theory of Crossing Numbers." *J. Comb. Th.* **8**, 45–53, 1970.

Crossing Number (Link)

The least number of crossings that occur in any projection of a LINK. In general, it is difficult to find the crossing number of a given LINK.

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 67–69, 1994.

Crout's Method

A ROOT finding technique used in LU DECOMPOSITION. It solves the N^2 equations

$$\begin{array}{ll} i < j & \alpha_{i1}\beta_{1j} + \alpha_{i2}\beta_{2j} + \ldots + \alpha_{ii}\beta_{ij} = a_{ij} \\ i = j & \alpha_{i1}\beta_{1j} + \alpha_{i2}\beta_{2j} + \ldots + \alpha_{ii}\beta_{jj} = a_{ij} \\ i > j & \alpha_{i1}\beta_{1j} + \alpha_{i2}\beta_{2j} + \ldots + \alpha_{ij}\beta_{jj} = a_{ij} \end{array}$$

for the $N^2 + N$ unknowns α_{ij} and β_{ij}.

see also LU DECOMPOSITION

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 36–38, 1992.

Crowd

A group of SOCIABLE NUMBERS of order 3.

Crown

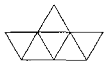

A 6-POLYIAMOND.

References

Golomb, S. W. *Polyominoes: Puzzles, Patterns, Problems, and Packings, 2nd ed.* Princeton, NJ: Princeton University Press, p. 92, 1994.

Crucial Point

The HOMOTHETIC CENTER of the ORTHIC TRIANGLE and the triangular hull of the three EXCIRCLES. It has TRIANGLE CENTER FUNCTION

$$\alpha = \tan A = \sin(2B) + \sin(2C) - \sin(2A).$$

References

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Lyness, R. and Veldkamp, G. R. Problem 682 and Solution. *Crux Math.* **9**, 23–24, 1983.

Cruciform

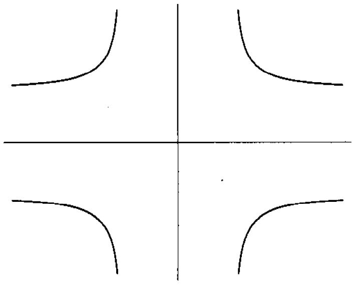

A plane curve also called the CROSS CURVE and POLICEMAN ON POINT DUTY CURVE (Cundy and Rollett 1989). It is given by the equation

$$x^2y^2 - a^2x^2 - a^2y^2 = 0, \tag{1}$$

which is equivalent to

$$1 - \frac{a^2}{y^2} - \frac{a^2}{x^2} = 0 \tag{2}$$

$$\frac{a^2}{x^2} + \frac{b^2}{y^2} = 1, \tag{3}$$

or, rewriting,

$$y^2 = \frac{a^2x^2}{x^2 - a^2}. \tag{4}$$

In parametric form,

$$x = a \sec t \tag{5}$$

$$y = b \csc t. \tag{6}$$

The CURVATURE is

$$\kappa = \frac{3ab \csc^2 t \sec^2 t}{(b^2 \cos^2 t \csc^2 t + a^2 \sec^2 t \tan^2 t)^{3/2}}. \tag{7}$$

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 71, 1989.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 127 and 130–131, 1972.

Crunode

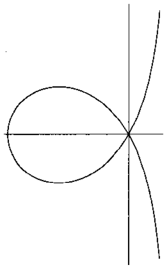

A point where a curve intersects itself so that two branches of the curve have distinct tangent lines. The MACLAURIN TRISECTRIX, shown above, has a crunode at the origin.

see also ACNODE, SPINODE, TACNODE

Cryptarithm

see CRYPTARITHMETIC

Cryptarithmetic

A number PUZZLE in which a group of arithmetical operations has some or all of its DIGITS replaced by letters or symbols, and where the original DIGITS must be found. In such a puzzle, each letter represents a unique digit.

see also ALPHAMETIC, DIGIMETIC, SKELETON DIVISION

References

Bogomolny, A. "Cryptarithms." `http://www.cut-the-knot.com/st_crypto.html`.

Brooke, M. *One Hundred & Fifty Puzzles in Crypt-Arithmetic.* New York: Dover, 1963.

Kraitchik, M. "Cryptarithmetic." §3.11 in *Mathematical Recreations.* New York: W. W. Norton, pp. 79–83, 1942.

Marks, R. W. *The New Mathematics Dictionary and Handbook.* New York: Bantam Books, 1964.

Cryptography

The science and mathematics of encoding and decoding information.

see also CRYPTARITHM, KNAPSACK PROBLEM, PUBLIC-KEY CRYPTOGRAPHY

References

Davies, D. W. *The Security of Data in Networks.* Los Angeles, CA: IEEE Computer Soc., 1981.

Diffie, W. and Hellman, M. "New Directions in Cryptography." *IEEE Trans. Info. Th.* **22**, 644–654, 1976.

Honsberger, R. "Four Clever Schemes in Cryptography." Ch. 10 in *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 151–173, 1985.

Simmons, G. J. "Cryptology, The Mathematics of Secure Communications." *Math. Intel.* **1**, 233–246, 1979.

Crystallography Restriction

If a discrete GROUP of displacements in the plane has more than one center of rotation, then the only rotations that can occur are by 2, 3, 4, and 6. This can be shown as follows. It must be true that the sum of the interior angles divided by the number of sides is a divisor of 360°.

$$\frac{180°(n-2)}{n} = \frac{360°}{m},$$

where m is an INTEGER. Therefore, symmetry will be possible only for

$$\frac{2n}{n-2} = m,$$

where m is an INTEGER. This will hold for 1-, 2-, 3-, 4-, and 6-fold symmetry. That it does not hold for $n > 6$ is seen by noting that $n = 6$ corresponds to $m = 3$. The $m = 2$ case requires that $n = n - 2$ (impossible), and the $m = 1$ case requires that $n = -2$ (also impossible).

see also POINT GROUPS, SYMMETRY

Császár Polyhedron

A POLYHEDRON topologically equivalent to a TORUS discovered in the late 1940s. It has 7 VERTICES, 14 faces, and 21 EDGES, and is the DUAL POLYHEDRON of the SZILASSI POLYHEDRON. Its SKELETON is ISOMORPHIC to the COMPLETE GRAPH K_7.

see also SZILASSI POLYHEDRON, TOROIDAL POLYHEDRON

References

Császár, Á. "A Polyhedron without Diagonals." *Acta Sci. Math.* **13**, 140–142, 1949–1950.

Gardner, M. "The Császár Polyhedron." Ch. 11 in *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, 1988.

Gardner, M. *Fractal Music, Hypercards, and More: Mathematical Recreations from Scientific American Magazine.* New York: W. H. Freeman, pp. 118–120, 1992.

Hart, G. "Toroidal Polyhedra." `http://www.li.net/~george/virtual-polyhedra/toroidal.html`.

Csch

see HYPERBOLIC COSECANT

Cube

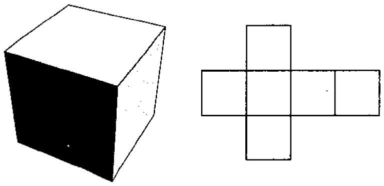

The three-dimensional PLATONIC SOLID (P_3) which is also called the HEXAHEDRON. The cube is composed of six SQUARE faces 6{4} which meet each other at RIGHT ANGLES, and has 8 VERTICES and 12 EDGES. It is described by the SCHLÄFLI SYMBOL {4,3}. It is a ZONOHEDRON. It is also the UNIFORM POLYHEDRON U_6 with WYTHOFF SYMBOL $3\,|\,2\,4$. It has the O_h OCTAHEDRAL GROUP of symmetries. The DUAL POLYHEDRON of the cube is the OCTAHEDRON.

Because the VOLUME of a cube of side length n is given by n^3, a number of the form n^3 is called a CUBIC NUMBER (or sometimes simply "a cube"). Similarly, the operation of taking a number to the third POWER is called CUBING.

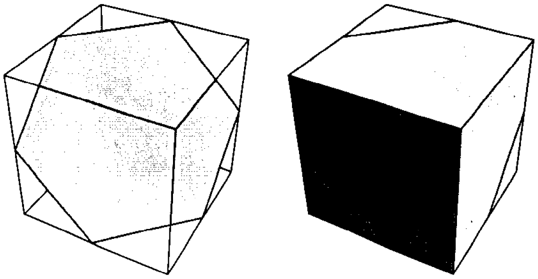

The cube cannot be STELLATED. A PLANE passing through the MIDPOINTS of opposite sides (perpendicular to a C_3 axis) cuts the cube in a regular HEXAGONAL CROSS-SECTION (Gardner 1960; Steinhaus 1983, p. 170; Cundy and Rollett 1989, p. 157; Holden 1991, pp. 22–23). Since there are four such axes, there are four possibly hexagonal cross-sections. If the vertices of the cube are $(\pm 1, \pm 1 \pm 1)$, then the vertices of the inscribed HEXAGON are $(0,-1,-1)$, $(1,0,-1)$, $(1,1,0)$, $(0,1,1)$, $(-1,0,1)$, and $(-1,-1,0)$. The largest SQUARE which will fit inside a cube of side a has each corner a distance 1/4 from a corner of a cube. The resulting SQUARE has side length $3\sqrt{2}\,a/4$, and the cube containing that side is called PRINCE RUPERT'S CUBE.

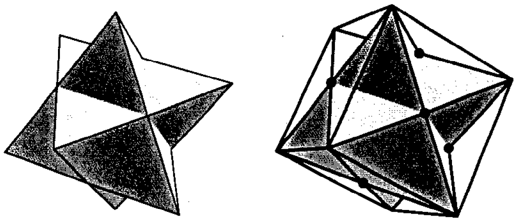

The solid formed by the faces having the sides of the STELLA OCTANGULA (left figure) as DIAGONALS is a cube (right figure; Ball and Coxeter 1987).

The VERTICES of a cube of side length 2 with face-centered axes are given by $(\pm 1, \pm 1, \pm 1)$. If the cube is oriented with a space diagonal along the z-axis, the coordinates are $(0,\ 0,\ \sqrt{3})$, $(0,\ 2\sqrt{2/3},\ 1/\sqrt{3})$, $(\sqrt{2},\ \sqrt{2/3},\ -1/\sqrt{3})$, $(\sqrt{2},\ -\sqrt{2/3},\ 1/\sqrt{3})$, $(0,\ -2\sqrt{2/3},\ -1/\sqrt{3})$, $(-\sqrt{2},\ -\sqrt{2/3},\ 1/\sqrt{3})$, $(-\sqrt{2},\ \sqrt{2/3},\ -1/\sqrt{3})$, and the negatives of these vectors. A FACETED version is the GREAT CUBICUBOCTAHEDRON.

A cube of side length 1 has INRADIUS, MIDRADIUS, and CIRCUMRADIUS of

$$r = \tfrac{1}{2} = 0.5 \qquad (1)$$

$$\rho = \tfrac{1}{2}\sqrt{2} \approx 0.70710 \qquad (2)$$

$$R = \tfrac{1}{2}\sqrt{3} \approx 0.86602. \qquad (3)$$

The cube has a DIHEDRAL ANGLE of

$$\alpha = \tfrac{1}{2}\pi. \qquad (4)$$

The SURFACE AREA and VOLUME of the cube are

$$S = 6a^2 \qquad (5)$$

$$V = a^3. \qquad (6)$$

see also AUGMENTED TRUNCATED CUBE, BIAUGMENTED TRUNCATED CUBE, BIDIAKIS CUBE, BISLIT CUBE, BROWKIN'S THEOREM, CUBE DISSECTION, CUBE DOVETAILING PROBLEM, CUBE DUPLICATION, CUBIC NUMBER, CUBICAL GRAPH, HADWIGER PROBLEM, HYPERCUBE, KELLER'S CONJECTURE, PRINCE

Rupert's Cube, Rubik's Cube, Soma Cube, Stella Octangula, Tesseract

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 127, 1987.

Cundy, H. and Rollett, A. "Hexagonal Section of a Cube." §3.15.1 in *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 157, 1989.

Davie, T. "The Cube (Hexahedron)." `http://www.dcs.st-and.ac.uk/~d/mathrecs/polyhedra/cube.html`.

Eppstein, D. "Rectilinear Geometry." `http://www.ics.uci.edu/~eppstein/junkyard/rect.html`.

Gardner, M. "Mathematical Games: More About the Shapes that Can Be Made with Complex Dominoes." *Sci. Amer.* **203**, 186–198, Nov. 1960.

Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, 1991.

Steinhaus, H. *Mathematical Snapshots, 3rd American ed.* New York: Oxford University Press, 1983.

Cube 2-Compound

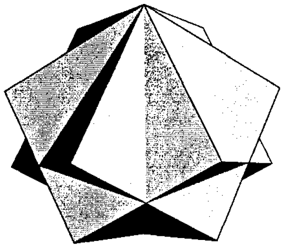

A Polyhedron Compound obtained by allowing two Cubes to share opposite Vertices, then rotating one a sixth of a turn (Holden 1971, p. 34).

see also Cube, Cube 3-Compound, Cube 4-Compound, Cube 5-Compound, Polyhedron Compound

References

Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, 1991.

Cube 3-Compound

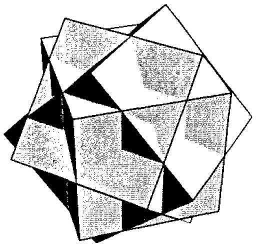

A compound with the symmetry of the Cube which arises by joining three Cubes such that each shares two C_2 axes (Holden 1971, p. 35).

see also Cube, Cube 2-Compound, Cube 4-Compound, Cube 5-Compound, Polyhedron Compound

References

Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, 1991.

Cube 4-Compound

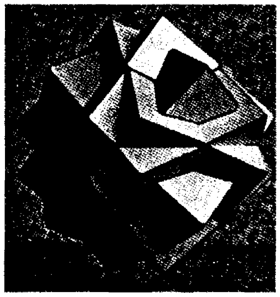

A compound with the symmetry of the Cube which arises by joining four Cubes such that each C_3 axis falls along the C_3 axis of one of the other Cubes (Holden 1971, p. 35).

see also Cube, Cube 2-Compound, Cube 3-Compound, Cube 5-Compound, Polyhedron Compound

References

Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, 1991.

Cube 5-Compound

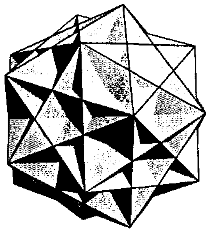

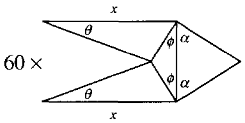

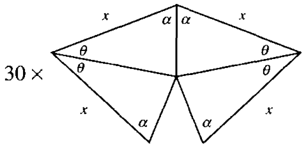

A Polyhedron Compound consisting of the arrangement of five Cubes in the Vertices of a Dodecahedron. In the above figure, let a be the length of a Cube Edge. Then

$$x = \tfrac{1}{2}a(3-\sqrt{5})$$

$$\theta = \tan^{-1}\left(\frac{3-\sqrt{5}}{2}\right) \approx 20^\circ 54'$$

$$\phi = \tan^{-1}\left(\frac{\sqrt{5}-1}{2}\right) \approx 31^\circ 43'$$

$$\psi = 90^\circ - \phi \approx 58^\circ 17'$$

$$\alpha = 90^\circ - \theta \approx 69^\circ 6'.$$

The compound is most easily constructed using pieces like the ones in the above line diagram. The cube 5-compound has the 30 facial planes of the Rhombic Triacontahedron (Ball and Coxeter 1987).

see also Cube, Cube 2-Compound, Cube 3-Compound, Cube 4-Compound, Dodecahedron, Polyhedron Compound, Rhombic Triacontahedron

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 135 and 137, 1987.
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 135–136, 1989.

Cube Dissection

A Cube can be divided into n subcubes for only $n = 1$, 8, 15, 20, 22, 27, 29, 34, 36, 38, 39, 41, 43, 45, 46, and $n \geq 48$ (Sloane's A014544).

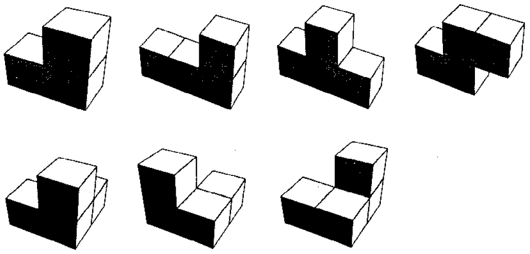

The seven pieces used to construct the $3 \times 3 \times 3$ cube dissection known as the Soma Cube are one 3-Polycube and six 4-Polycubes ($1 \cdot 3 + 6 \cdot 4 = 27$), illustrated above.

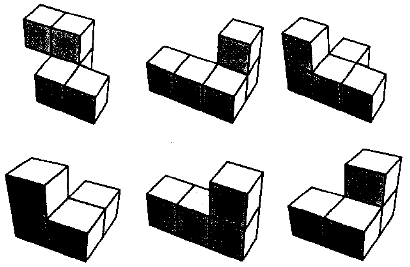

Another $3 \times 3 \times 3$ cube dissection due to Steinhaus uses three 5-Polycubes and three 4-Polycubes ($3 \cdot 5 + 3 \cdot 4 = 27$), illustrated above.

It is possible to cut a 1×3 Rectangle into two identical pieces which will form a Cube (without overlapping) when folded and joined. In fact, an Infinite number of solutions to this problem were discovered by C. L. Baker (Hunter and Madachy 1975).

see also Conway Puzzle, Dissection, Hadwiger Problem, Polycube, Slothouber-Graatsma Puzzle, Soma Cube

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 112–113, 1987.
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 203–205, 1989.
Gardner, M. "Block Packing." Ch. 18 in *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, pp. 227–239, 1988.
Gardner, M. *Fractal Music, Hypercards, and More: Mathematical Recreations from Scientific American Magazine.* New York: W. H. Freeman, pp. 297–298, 1992.
Honsberger, R. *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., pp. 75–80, 1976.
Hunter, J. A. H. and Madachy, J. S. *Mathematical Diversions.* New York: Dover, pp. 69–70, 1975.
Sloane, N. J. A. Sequence A014544 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Cube Dovetailing Problem

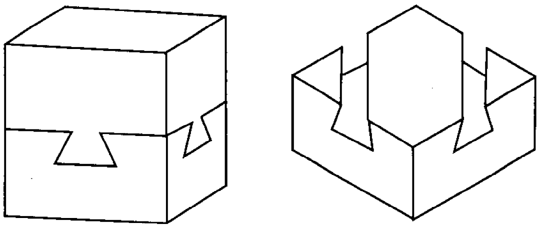

Given the figure on the left (without looking at the solution on the right), determine how to disengage the two slotted Cube halves without cutting, breaking, or distorting.

References
Dudeney, H. E. *Amusements in Mathematics.* New York: Dover, pp. 145 and 249, 1958.
Ogilvy, C. S. *Excursions in Mathematics.* New York: Dover, pp. 57, 59, and 143, 1994.

Cube Duplication

Also called the Delian Problem or Duplication of the Cube. A classical problem of antiquity which, given the Edge of a Cube, requires a second Cube to be constructed having double the Volume of the first using only a Straightedge and Compass.

Under these restrictions, the problem cannot be solved because the Delian Constant $2^{1/3}$ (the required Ratio of sides of the original Cube and that to be constructed) is not a Euclidean Number. The problem can be solved, however, using a Neusis Construction.

see also Alhazen's Billiard Problem, Compass, Cube, Delian Constant, Geometric Problems of Antiquity, Neusis Construction, Straightedge

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 93–94, 1987.
Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 190–191, 1996.
Courant, R. and Robbins, H. "Doubling the Cube" and "A Classical Construction for Doubling the Cube." §3.3.1 and 3.5.1 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 134–135 and 146, 1996.
Dörrie, H. "The Delian Cube-Doubling Problem." §35 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 170–172, 1965.

Cube-Octahedron Compound

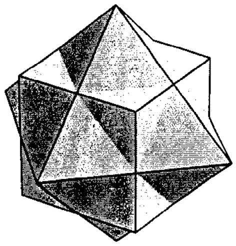

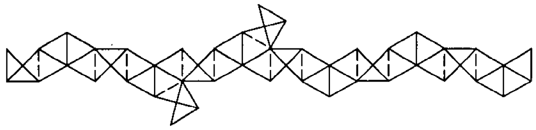

A POLYHEDRON COMPOUND composed of a CUBE and its DUAL POLYHEDRON, the OCTAHEDRON. The 14 vertices are given by (±1, ±1, ±1), (±2, 0, 0), (0, ±2, 0), (0, 0, ±2).

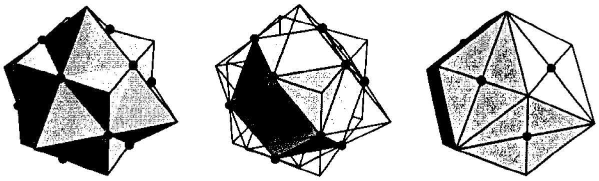

The solid common to both the CUBE and OCTAHEDRON (left figure) in a cube-octahedron compound is a CUBOCTAHEDRON (middle figure). The edges intersecting in the points plotted above are the diagonals of RHOMBUSES, and the 12 RHOMBUSES form a RHOMBIC DODECAHEDRON (right figure; Ball and Coxeter 1987).

see also CUBE, CUBOCTAHEDRON, OCTAHEDRON, POLYHEDRON COMPOUND

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 137, 1987.

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, p. 158, 1969.

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 130, 1989.

Cube Point Picking

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let two points be picked randomly from a unit n-D HYPERCUBE. The expected distance between the points $\Delta(N)$ is then

$$\begin{aligned}
\Delta(1) &= \tfrac{1}{3} \\
\Delta(2) &= \tfrac{1}{15}[\sqrt{2}+2+5\ln(1+\sqrt{2})] = 0.521405433\ldots \\
\Delta(3) &= \tfrac{1}{105}[4+17\sqrt{2}-6\sqrt{3}+21\ln(1+\sqrt{2}) \\
&\quad +42\ln(2+\sqrt{3})-7\pi] = 0.661707182\ldots \\
\Delta(4) &= 0.77766\ldots \\
\Delta(5) &= 0.87852\ldots \\
\Delta(6) &= 0.96895\ldots \\
\Delta(7) &= 1.05159\ldots \\
\Delta(8) &= 1.12817\ldots.
\end{aligned}$$

The function $\Delta(n)$ satisfies

$$\tfrac{1}{3}n^{1/2} \leq \Delta(n) \leq (\tfrac{1}{6}n)^{1/2}\sqrt{\frac{1}{3}\left[1+2\left(1-\frac{3}{5n}\right)^{1/2}\right]}$$

(Anderssen *et al.* 1976).

Pick N points $p_1, \ldots, p_N$ randomly in a unit n-cube. Let C be the CONVEX HULL, so

$$C \equiv \left\{ \sum_{j=1}^{N} \lambda_j p_j : \lambda_j \geq 0 \text{ for all } j \text{ and } \sum_{j=1}^{N} \lambda_j = 1 \right\}.$$

Let $V(n,N)$ be the expected n-D VOLUME (the CONTENT) of C, $S(n,N)$ be the expected $(n-1)$-D SURFACE AREA of C, and $P(n,N)$ the expected number of VERTICES on the POLYGONAL boundary of C. Then

$$\lim_{N\to\infty} \frac{N[1-V(2,N)]}{\ln N} = \frac{8}{3}$$

$$\begin{aligned}
\lim_{N\to\infty} \sqrt{N}[4-S(2,N)] &= \sqrt{2\pi}\left[2-\int_0^1 (\sqrt{1+t^2}-1)t^{-3/2}\,dt\right] \\
&= 4.2472965\ldots,
\end{aligned}$$

and

$$\begin{aligned}
\lim_{N\to\infty} P(2,N) - \tfrac{8}{3}\ln N = \tfrac{8}{3}(\gamma-\ln 2) \\
= -0.309150708\ldots
\end{aligned}$$

(Rényi and Sulanke 1963, 1964). The average DISTANCE between two points chosen at random inside a unit cube is

$$\tfrac{1}{105}(4+17\sqrt{2}-6\sqrt{3}+21\ln(1+\sqrt{2})+42\ln(2+\sqrt{3})-7\pi)$$

(Robbins 1978, Le Lionnais 1983).

Pick n points on a CUBE, and space them as far apart as possible. The best value known for the minimum straight LINE distance between any two points is given in the following table.

n	$d(n)$
5	1.1180339887498
6	1.0606601482100
7	1
8	1
9	0.86602540378463
10	0.74999998333331
11	0.70961617562351
12	0.70710678118660
13	0.70710678118660
14	0.70710678118660
15	0.625

see also Cube Triangle Picking, Discrepancy Theorem, Point Picking

References

Anderssen, R. S.; Brent, R. P.; Daley, D. J.; and Moran, A. P. "Concerning $\int_0^1 \cdots \int_0^1 \sqrt{x_1{}^2 + \ldots + x_k{}^2}\, dx_1 \cdots dx_k$ and a Taylor Series Method." *SIAM J. Appl. Math.* **30**, 22–30, 1976.

Bolis, T. S. Solution to Problem E2629. "Average Distance Between Two Points in a Box." *Amer. Math. Monthly* **85**, 277–278, 1978.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/geom/geom.html`.

Ghosh, B. "Random Distances within a Rectangle and Between Two Rectangles." *Bull. Calcutta Math. Soc.* **43**, 17–24, 1951.

Holshouser, A. L.; King, L. R.; and Klein, B. G. Solution to Problem E3217, "Minimum Average Distance Between Points in a Rectangle." *Amer. Math. Monthly* **96**, 64–65, 1989.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 30, 1983.

Rényi, A. and Sulanke, R. "Über die konvexe Hülle von n zufällig gewählten Punkten, I." *Z. Wahrscheinlichkeits* **2**, 75–84, 1963.

Rényi, A. and Sulanke, R. "Über die konvexe Hülle von n zufällig gewählten Punkten, II." *Z. Wahrscheinlichkeits* **3**, 138–147, 1964.

Robbins, D. "Average Distance Between Two Points in a Box." *Amer. Math. Monthly* **85**, 278, 1978.

Santaló, L. A. *Integral Geometry and Geometric Probability.* Reading, MA: Addison-Wesley, 1976.

Cube Power

A number raised to the third Power. x^3 is read as "x cubed."

see also Cubic Number

Cube Root

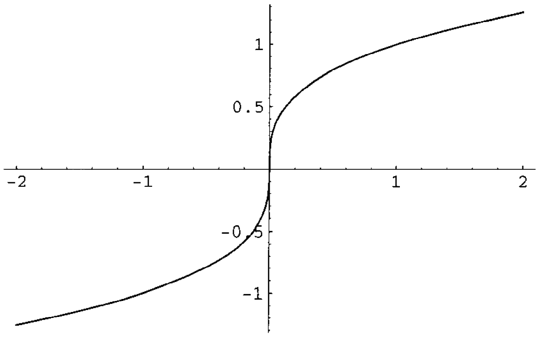

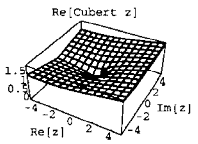

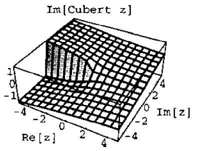

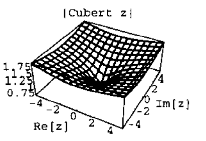

Given a number z, the cube root of z, denoted $\sqrt[3]{z}$ or $z^{1/3}$ (z to the 1/3 Power), is a number a such that $a^3 = z$. There are three (not necessarily distinct) cube roots for any number.

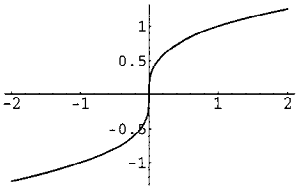

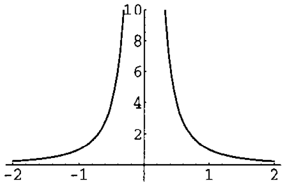

For real arguments, the cube root is an Increasing Function, although the usual derivative test cannot be used to establish this fact at the Origin since the the derivative approaches infinity there (as illustrated above).

see also Cube Duplication, Cubed, Delian Constant, Geometric Problems of Antiquity, k-Matrix, Square Root

Cube Triangle Picking

Pick 3 points at random in the unit n-Hypercube. Denote the probability that the three points form an Obtuse Triangle $\Pi(n)$. Langford (1969) proved

$$\Pi(2) = \tfrac{97}{150} + \tfrac{1}{40}\pi = 0.725206483\ldots.$$

see also Ball Triangle Picking, Cube Point Picking

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/geom/geom.html`.

Langford, E. "The Probability that a Random Triangle is Obtuse." *Biometrika* **56**, 689–690, 1969.

Santaló, L. A. *Integral Geometry and Geometric Probability.* Reading, MA: Addison-Wesley, 1976.

Cubed

A number to the Power 3 is said to be cubed, so that x^3 is called "x cubed."

see also Cube Root, Squared

Cubefree

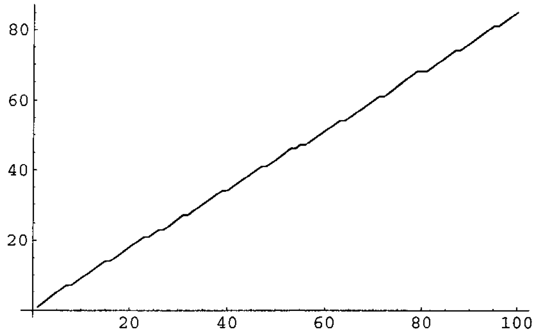

A number is said to be cubefree if its Prime decomposition contains no tripled factors. All Primes are therefore trivially cubefree. The cubefree numbers are 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 17, ... (Sloane's A004709). The cubeful numbers (i.e., those that contain at least one cube) are 8, 16, 24, 27, 32, 40, 48, 54, ... (Sloane's A046099). The number of cubefree numbers less than 10, 100, 1000, ... are 9, 85, 833, 8319, 83190, 831910, ..., and their asymptotic density is $1/\zeta(3) \approx 0.831907$, where $\zeta(n)$ is the RIEMANN ZETA FUNCTION.

see also BIQUADRATEFREE, PRIME NUMBER, RIEMANN ZETA FUNCTION, SQUAREFREE

References
Sloane, N. J. A. Sequences A004709 and A046099 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Cubic Curve

A cubic curve is an ALGEBRAIC CURVE of degree 3. An algebraic curve over a FIELD K is an equation $f(X,Y) = 0$, where $f(X,Y)$ is a POLYNOMIAL in X and Y with COEFFICIENTS in K, and the degree of f is the MAXIMUM degree of each of its terms (MONOMIALS).

Newton showed that all cubics can be generated by the projection of the five divergent cubic parabolas. Newton's classification of cubic curves appeared in the chapter "Curves" in *Lexicon Technicum* by John Harris published in London in 1710. Newton also classified all cubics into 72 types, missing six of them. In addition, he showed that any cubic can be obtained by a suitable projection of the ELLIPTIC CURVE

$$y^2 = ax^3 + bx^2 + cx + d, \tag{1}$$

where the projection is a BIRATIONAL TRANSFORMATION, and the general cubic can also be written as

$$y^2 = x^3 + ax + b. \tag{2}$$

Newton's first class is equations of the form

$$xy^2 + ey = ax^3 + bx^2 + cx + d. \tag{3}$$

This is the hardest case and includes the SERPENTINE CURVE as one of the subcases. The third class was

$$ay^2 = x(x^2 - 2bx + c), \tag{4}$$

which is called NEWTON'S DIVERGING PARABOLAS. Newton's 66th curve was the TRIDENT OF NEWTON. Newton's classification of cubics was criticized by Euler because it lacked generality. Plücker later gave a more detailed classification with 219 types.

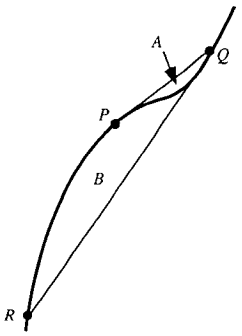

Pick a point P, and draw the tangent to the curve at P. Call the point where this tangent intersects the curve Q. Draw another tangent and call the point of intersection with the curve R. Every curve of third degree has the property that, with the areas in the above labeled figure,

$$B = 16A \tag{5}$$

(Honsberger 1991).

see also CAYLEY-BACHARACH THEOREM, CUBIC EQUATION

References
Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., pp. 114–118, 1991.
Newton, I. *Mathematical Works, Vol. 2.* New York: Johnson Reprint Corp., pp. 135–161, 1967.
Wall, C. T. C. "Affine Cubic Functions III." *Math. Proc. Cambridge Phil. Soc.* **87**, 1–14, 1980.
Westfall, R. S. *Never at Rest: A Biography of Isaac Newton.* New York: Cambridge University Press, 1988.
Yates, R. C. "Cubic Parabola." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 56–59, 1952.

Cubic Equation

A cubic equation is a POLYNOMIAL equation of degree three. Given a general cubic equation

$$z^3 + a_2z^2 + a_1z + a_0 = 0 \tag{1}$$

(the COEFFICIENT a_3 of z^3 may be taken as 1 without loss of generality by dividing the entire equation through by a_3), first attempt to eliminate the a_2 term by making a substitution of the form

$$z \equiv x - \lambda. \tag{2}$$

Then

$$(x-\lambda)^3 + a_2(x-\lambda)^2 + a_1(x-\lambda) + a_0 = 0 \tag{3}$$

$$(x^3 - 3\lambda x^2 + 3\lambda^2 x - \lambda^3) + a_2(x^2 - 2\lambda x + \lambda^2) + a_1(x-\lambda) + a_0 = 0 \tag{4}$$

$$x^3 + x^2(a_2 - 3\lambda) + x(a_1 - 2a_2\lambda + 3\lambda^2) + (a_0 - a_1\lambda + a_2\lambda^2 - \lambda^3) = 0. \tag{5}$$

The x^2 is eliminated by letting $\lambda = a_2/3$, so

$$z \equiv x - \tfrac{1}{3}a_2. \tag{6}$$

Then

$$z^3 = (x - \tfrac{1}{3}a_2)^3 = x^3 - a_2x^2 + \tfrac{1}{3}{a_2}^2x - \tfrac{1}{27}{a_2}^3 \tag{7}$$

$$a_2z^2 = a_2(x - \tfrac{1}{3}a_2)^2 = a_2x^2 - \tfrac{2}{3}{a_2}^2x + \tfrac{1}{9}{a_2}^3 \tag{8}$$

$$a_1z = a_1(x - \tfrac{1}{3}a_2) = a_1x - \tfrac{1}{3}a_1a_2, \tag{9}$$

so equation (1) becomes

$$x^3 + (-a_2 + a_2)x^2 + (\tfrac{1}{3}{a_2}^2 - \tfrac{2}{3}{a_2}^2 + a_1)x - (\tfrac{1}{27}{a_2}^3 - \tfrac{1}{9}{a_2}^3 + \tfrac{1}{3}a_1a_2 - a_0) = 0 \tag{10}$$

$$x^3 + (a_1 - \tfrac{1}{3}{a_2}^2)x - (\tfrac{1}{3}a_1a_2 - \tfrac{2}{27}{a_2}^3 - a_0) = 0 \qquad (11)$$

$$x^3 + 3 \cdot \frac{3a_1 - {a_2}^2}{9} x - 2 \cdot \frac{9a_1a_2 - 27a_0 - 2{a_2}^3}{54} = 0. \qquad (12)$$

Defining

$$p \equiv \frac{3a_1 - {a_2}^2}{3} \qquad (13)$$

$$q \equiv \frac{9a_1a_2 - 27a_0 - 2{a_2}^3}{27} \qquad (14)$$

then allows (12) to be written in the standard form

$$x^3 + px = q. \qquad (15)$$

The simplest way to proceed is to make VIETA'S SUBSTITUTION

$$x = w - \frac{p}{3w}, \qquad (16)$$

which reduces the cubic to the equation

$$w^3 - \frac{p^3}{27w^3} - q = 0, \qquad (17)$$

which is easily turned into a QUADRATIC EQUATION in w^3 by multiplying through by w^3 to obtain

$$(w^3)^2 - q(w^3) - \tfrac{1}{27}p^3 = 0 \qquad (18)$$

(Birkhoff and Mac Lane 1965, p. 106). The result from the QUADRATIC EQUATION is

$$\begin{aligned} w^3 &= \tfrac{1}{2}\left(q \pm \sqrt{q^2 + \tfrac{4}{27}p^3}\right) = \tfrac{1}{2}q \pm \sqrt{\tfrac{1}{4}q^2 + \tfrac{1}{27}p^3} \\ &= R \pm \sqrt{R^2 + Q^3}, \qquad (19) \end{aligned}$$

where Q and R are are sometimes more useful to deal with than are p and q. There are therefore six solutions for w (two corresponding to each sign for each ROOT of w^3). Plugging w back in to (17) gives three pairs of solutions, but each pair is equal, so there are three solutions to the cubic equation.

Equation (12) may also be explicitly factored by attempting to pull out a term of the form $(x - B)$ from the cubic equation, leaving behind a quadratic equation which can then be factored using the QUADRATIC FORMULA. This process is equivalent to making Vieta's substitution, but does a slightly better job of motivating Vieta's "magic" substitution, and also at producing the explicit formulas for the solutions. First, define the intermediate variables

$$Q \equiv \frac{3a_1 - {a_2}^2}{9} \qquad (20)$$

$$R \equiv \frac{9a_2a_1 - 27a_0 - 2{a_2}^3}{54} \qquad (21)$$

(which are identical to p and q up to a constant factor). The general cubic equation (12) then becomes

$$x^3 + 3Qx - 2R = 0. \qquad (22)$$

Let B and C be, for the moment, arbitrary constants. An identity satisfied by PERFECT CUBIC equations is that

$$x^3 - B^3 = (x - B)(x^2 + Bx + B^2). \qquad (23)$$

The general cubic would therefore be directly factorable if it did not have an x term (i.e., if $Q = 0$). However, since in general $Q \neq 0$, add a multiple of $(x - B)$—say $C(x-B)$—to both sides of (23) to give the slightly messy identity

$$(x^3 - B^3) + C(x - B) = (x - B)(x^2 + Bx + B^2 + C) = 0, \qquad (24)$$

which, after regrouping terms, is

$$x^3 + Cx - (B^3 + BC) = (x - B)[x^2 + Bx + (B^2 + C)] = 0. \qquad (25)$$

We would now like to match the COEFFICIENTS C and $-(B^3 + BC)$ with those of equation (22), so we must have

$$C = 3Q \qquad (26)$$

$$B^3 + BC = 2R. \qquad (27)$$

Plugging the former into the latter then gives

$$B^3 + 3QB = 2R. \qquad (28)$$

Therefore, if we can find a value of B satisfying the above identity, we have factored a linear term from the cubic, thus reducing it to a QUADRATIC EQUATION. The trial solution accomplishing this miracle turns out to be the symmetrical expression

$$B = [R + \sqrt{Q^3 + R^2}\,]^{1/3} + [R - \sqrt{Q^3 + R^2}\,]^{1/3}. \qquad (29)$$

Taking the second and third POWERS of B gives

$$\begin{aligned} B^2 &= [R + \sqrt{Q^3 + R^2}]^{2/3} + 2[R^2 - (Q^3 + R^2)]^{1/3} \\ &\quad + [R - \sqrt{Q^3 + R^2}]^{2/3} \\ &= [R + \sqrt{Q^3 + R^2}]^{2/3} + [R - \sqrt{Q^3 + R^2}]^{2/3} - 2Q \qquad (30) \\ B^3 &= -2QB + \left\{[R + \sqrt{Q^3 + R^2}]^{1/3} + [R - \sqrt{Q^3 + R^2}]^{1/3}\right\} \\ &\quad \times \left\{[R + \sqrt{Q^3 + R^2}]^{2/3} + [R - \sqrt{Q^3 + R^2}]^{2/3}\right\} \\ &= [R + \sqrt{Q^3 + R^2}] + [R - \sqrt{Q^3 + R^2}] \\ &\quad + [R - \sqrt{Q^3 + R^2}]^{1/3}[R - \sqrt{Q^3 + R^2}]^{2/3} \\ &\quad + [R - \sqrt{Q^3 + R^2}]^{2/3}[R - \sqrt{Q^3 + R^2}]^{1/3} - 2QB \\ &= -2QB + 2R + [R^2 - (Q^3 + R^2)]^{1/3} \\ &\quad \times \left[\left(R + \sqrt{Q^3 + R^2}\right)^{1/3} + \left(R - \sqrt{Q^3 - R^2}\right)^{1/3}\right] \\ &= -2QB + 2R - QB = -3QB + 2R. \qquad (31) \end{aligned}$$

Plugging B^3 and B into the left side of (28) gives

$$(-3QB+2R)+3QB=2R, \tag{32}$$

so we have indeed found the factor $(x-B)$ of (22), and we need now only factor the quadratic part. Plugging $C=3Q$ into the quadratic part of (25) and solving the resulting

$$x^2+Bx+(B^2+3Q)=0 \tag{33}$$

then gives the solutions

$$\begin{aligned} x &= \tfrac{1}{2}[-B\pm\sqrt{B^2-4(B^2+3Q)}\,]\\ &= -\tfrac{1}{2}B\pm\tfrac{1}{2}\sqrt{-3B^2-12Q}\\ &= -\tfrac{1}{2}B\pm\tfrac{1}{2}\sqrt{3}\,i\sqrt{B^2+4Q}\,. \end{aligned} \tag{34}$$

These can be simplified by defining

$$A\equiv[R+\sqrt{Q^3+R^2}\,]^{1/3}-[R-\sqrt{Q^3+R^2}\,]^{1/3} \tag{35}$$

$$\begin{aligned} A^2 &= [R+\sqrt{Q^3+R^2}\,]^{2/3}-2[R^2-(Q^3+R^2)]^{1/3}\\ &\quad+[R-\sqrt{Q^3+R^2}\,]^{2/3}\\ &= [R+\sqrt{Q^3+R^2}\,]^{2/3}+[R-\sqrt{Q^3+R^2}\,]^{2/3}+2Q\\ &= B^2+4Q, \end{aligned} \tag{36}$$

so that the solutions to the quadratic part can be written

$$x=-\tfrac{1}{2}B\pm\tfrac{1}{2}\sqrt{3}\,iA. \tag{37}$$

Defining

$$D\equiv Q^3+R^2 \tag{38}$$

$$S\equiv\sqrt[3]{R+\sqrt{D}} \tag{39}$$

$$T\equiv\sqrt[3]{R-\sqrt{D}}, \tag{40}$$

where D is the DISCRIMINANT (which is defined slightly differently, including the opposite SIGN, by Birkhoff and Mac Lane 1965) then gives very simple expressions for A and B, namely

$$B=S+T \tag{41}$$

$$A=S-T. \tag{42}$$

Therefore, at last, the ROOTS of the original equation in z are then given by

$$z_1=-\tfrac{1}{3}a_2+(S+T) \tag{43}$$

$$z_2=-\tfrac{1}{3}a_2-\tfrac{1}{2}(S+T)+\tfrac{1}{2}i\sqrt{3}\,(S-T) \tag{44}$$

$$z_3=-\tfrac{1}{3}a_2-\tfrac{1}{2}(S+T)-\tfrac{1}{2}i\sqrt{3}\,(S-T), \tag{45}$$

with a_2 the COEFFICIENT of z^2 in the original equation, and S and T as defined above. These three equations giving the three ROOTS of the cubic equation are sometimes known as CARDANO'S FORMULA. Note that if the equation is in the standard form of Vieta

$$x^3+px=q, \tag{46}$$

in the variable x, then $a_2=0$, $a_1=p$, and $a_0=-q$, and the intermediate variables have the simple form (c.f. Beyer 1987)

$$Q=\tfrac{1}{3}p \tag{47}$$

$$R=\tfrac{1}{2}q \tag{48}$$

$$D\equiv Q^3+R^2=\left(\frac{p}{3}\right)^3+\left(\frac{q}{2}\right)^2. \tag{49}$$

The equation for z_1 in CARDANO'S FORMULA does not have an i appearing in it explicitly while z_2 and z_3 do, but this does not say anything about the number of REAL and COMPLEX ROOTS (since S and T are themselves, in general, COMPLEX). However, determining which ROOTS are REAL and which are COMPLEX can be accomplished by noting that if the DISCRIMINANT $D>0$, one ROOT is REAL and two are COMPLEX CONJUGATES; if $D=0$, all ROOTS are REAL and at least two are equal; and if $D<0$, all ROOTS are REAL and unequal. If $D<0$, define

$$\theta\equiv\cos^{-1}\left(\frac{R}{\sqrt{-Q^3}}\right). \tag{50}$$

Then the REAL solutions are of the form

$$z_1=2\sqrt{-Q}\cos\left(\frac{\theta}{3}\right)-\tfrac{1}{3}a_2 \tag{51}$$

$$z_2=2\sqrt{-Q}\cos\left(\frac{\theta+2\pi}{3}\right)-\tfrac{1}{3}a_2 \tag{52}$$

$$z_3=2\sqrt{-Q}\cos\left(\frac{\theta+4\pi}{3}\right)-\tfrac{1}{3}a_2. \tag{53}$$

This procedure can be generalized to find the REAL ROOTS for any equation in the standard form (46) by using the identity

$$\sin^3\theta-\tfrac{3}{4}\sin\theta+\tfrac{1}{4}\sin(3\theta)=0 \tag{54}$$

(Dickson 1914) and setting

$$x\equiv\sqrt{\frac{4|p|}{3}}\,y \tag{55}$$

(Birkhoff and Mac Lane 1965, pp. 90–91), then

$$\left(\frac{4|p|}{3}\right)^{3/2}y^3+p\sqrt{\frac{4|p|}{3}}\,y=q \tag{56}$$

$$y^3+\tfrac{3}{4}\frac{p}{|p|}y=\left(\frac{3}{4|p|}\right)^{3/2}q \tag{57}$$

$$4y^3 + 3\operatorname{sgn}(p)y = \tfrac{1}{2}q\left(\frac{3}{|p|}\right)^{3/2} \equiv C. \tag{58}$$

If $p > 0$, then use

$$\sinh(3\theta) = 4\sinh^3\theta + 3\sinh\theta \tag{59}$$

to obtain

$$y = \sinh(\tfrac{1}{3}\sinh^{-1}C). \tag{60}$$

If $p < 0$ and $|C| \geq 1$, use

$$\cosh(3\theta) = 4\cosh^3\theta - 3\cosh\theta, \tag{61}$$

and if $p < 0$ and $|C| \leq 1$, use

$$\cos(3\theta) = 4\cos^3\theta - 3\cos\theta, \tag{62}$$

to obtain

$$y = \begin{cases} \cosh(\tfrac{1}{3}\cosh^{-1}C) & \text{for } C \geq 1 \\ -\cosh(\tfrac{1}{3}\cosh^{-1}|C|) & \text{for } C \leq -1 \\ \cos(\tfrac{1}{3}\cos^{-1}C)\ \text{[three solutions]} & \text{for } |C| < 1. \end{cases} \tag{63}$$

The solutions to the original equation are then

$$x_i = 2\sqrt{\frac{|p|}{3}}\,y_i - \tfrac{1}{3}a_2. \tag{64}$$

An alternate approach to solving the cubic equation is to use LAGRANGE RESOLVENTS. Let $\omega \equiv e^{2\pi i/3}$, define

$$(1, x_1) = x_1 + x_2 + x_3 \tag{65}$$

$$(\omega, x_1) = x_1 + \omega x_2 + \omega^2 x_3 \tag{66}$$

$$(\omega^2, x_1) = x_1 + \omega^2 x_2 + \omega x_3, \tag{67}$$

where x_i are the ROOTS of

$$x^3 + px + q = 0, \tag{68}$$

and consider the equation

$$[x - (u_1 + u_2)][x - (\omega u_1 + \omega^2 u_2)][x - (\omega^2 u_1 + \omega u_2)] = 0, \tag{69}$$

where u_1 and u_2 are COMPLEX NUMBERS. The ROOTS are then

$$x_j = \omega^j u_1 + \omega^{2j} u_2 \tag{70}$$

for $j = 0$, 1, 2. Multiplying through gives

$$x^3 - 3u_1u_2x - ({u_1}^3 + {u_2}^3) = 0, \tag{71}$$

or

$$x^3 + px + q = 0, \tag{72}$$

where

$${u_1}^3 + {u_2}^3 = -q \tag{73}$$

$${u_1}^3{u_2}^3 = -\left(\frac{1}{3}\right)^3. \tag{74}$$

The solutions satisfy NEWTON'S IDENTITIES

$$z_1 + z_2 + z_3 = -a_2 \tag{75}$$

$$z_1z_2 + z_2z_3 + z_1z_3 = a_1 \tag{76}$$

$$z_1z_2z_3 = -a_0. \tag{77}$$

In standard form, $a_2 = 0$, $a_1 = p$, and $a_0 = -q$, so we have the identities

$$p = z_1z_2 - {z_3}^2 \tag{78}$$

$$(z_1 - z_2)^2 = -(4p - 3{z_3}^2) \tag{79}$$

$${z_1}^2 + {z_2}^2 + {z_3}^2 = -2p. \tag{80}$$

Some curious identities involving the roots of a cubic equation due to Ramanujan are given by Berndt (1994).

see also QUADRATIC EQUATION, QUARTIC EQUATION, QUINTIC EQUATION, SEXTIC EQUATION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 17, 1972.

Berger, M. §16.4.1–16.4.11.1 in *Geometry I.* New York: Springer-Verlag, 1994.

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 22–23, 1994.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 9–11, 1987.

Birkhoff, G. and Mac Lane, S. *A Survey of Modern Algebra, 3rd ed.* New York: Macmillan, pp. 90–91, 106–107, and 414–417, 1965.

Dickson, L. E. "A New Solution of the Cubic Equation." *Amer. Math. Monthly* **5**, 38–39, 1898.

Dickson, L. E. *Elementary Theory of Equations.* New York: Wiley, pp. 36–37, 1914.

Dunham, W. "Cardano and the Solution of the Cubic." Ch. 6 in *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 133–154, 1990.

Ehrlich, G. §4.16 in *Fundamental Concepts of Abstract Algebra.* Boston, MA: PWS-Kent, 1991.

Jones, J. "Omar Khayyám and a Geometric Solution of the Cubic." `http://jwilson.coe.uga.edu/emt669/Student.Folders/Jones.June/omar/omarpaper.html`.

Kennedy, E. C. "A Note on the Roots of a Cubic." *Amer. Math. Monthly* **40**, 411–412, 1933.

King, R. B. *Beyond the Quartic Equation.* Boston, MA: Birkhäuser, 1996.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Quadratic and Cubic Equations." §5.6 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 178–180, 1992.

Spanier, J. and Oldham, K. B. "The Cubic Function $x^3 + ax^2 + bx + c$ and Higher Polynomials." Ch. 17 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 131–147, 1987.

van der Waerden, B. L. §64 in *Algebra.* New York: Frederick Ungar, 1970.

Cubic Number

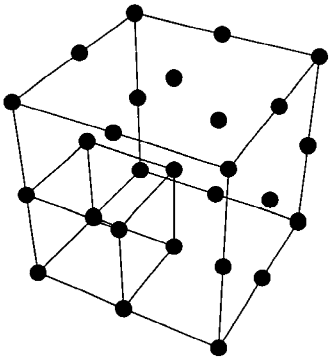

A Figurate Number of the form n^3, for n a Positive Integer. The first few are 1, 8, 27, 64, ... (Sloane's A000578). The Generating Function giving the cubic numbers is

$$\frac{x(x^2+4x+1)}{(x-1)^4} = x + 8x^2 + 27x^3 + \ldots. \quad (1)$$

The Hex Pyramidal Numbers are equivalent to the cubic numbers (Conway and Guy 1996).

The number of *positive* cubes needed to represent the numbers 1, 2, 3, ... are 1, 2, 3, 4, 5, 6, 7, 1, 2, 3, 4, 5, 6, 7, 8, 2, ... (Sloane's A02376), and the number of distinct ways to represent the numbers 1, 2, 3, ... in terms of positive cubes are 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 5, 5, 5, 5, ... (Sloane's A003108). In the early twentieth century, Dickson, Pillai, and Niven proved that every Positive Integer is the sum of not more than nine Cubes (so $g(3) = 9$ in Waring's Problem).

In 1939, Dickson proved that the only Integers requiring nine Cubes are 23 and 239. Wieferich proved that only 15 Integers require eight Cubes: 15, 22, 50, 114, 167, 175, 186, 212, 213, 238, 303, 364, 420, 428, and 454 (Sloane's A018889). The quantity $G(3)$ in Waring's Problem therefore satisfies $G(3) \leq 7$, and the largest number known requiring seven cubes is 8042. The following table gives the first few numbers which require at least $N = 1, 2, 3, \ldots, 9$ (positive) cubes to represent them as a sum.

N	Sloane	Numbers
1	000578	1, 8, 27, 64, 125, 216, 343, 512, ...
2	003325	2, 9, 16, 28, 35, 54, 65, 72, 91, ...
3	003072	3, 10, 17, 24, 29, 36, 43, 55, 62, ...
4	003327	4, 11, 18, 25, 30, 32, 37, 44, 51, ...
5	003328	5, 12, 19, 26, 31, 33, 38, 40, 45, ...
6		6, 13, 20, 34, 39, 41, 46, 48, 53, ...
7	018890	7, 14, 21, 42, 47, 49, 61, 77, ...
8	018889	15, 22, 50, 114, 167, 175, 186, ...,
9	—	23, 239

There is a finite set of numbers which cannot be expressed as the sum of *distinct* cubes: 2, 3, 4, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, ... (Sloane's A001476). The following table gives the numbers which can be represented in W different ways as a sum of N positive cubes. For example,

$$157 = 4^3+4^3+3^3+1^3+1^3 = 5^3+2^3+2^3+2^3+2^3 \quad (2)$$

can be represented in $W = 2$ ways by $N = 5$ cubes. The smallest number representable in $W = 2$ ways as a sum of $N = 2$ cubes,

$$1729 = 1^3 + 12^3 = 9^3 + 10^3, \quad (3)$$

is called the Hardy-Ramanujan Number and has special significance in the history of mathematics as a result of a story told by Hardy about Ramanujan. Sloane's A001235 is defined as the sequence of numbers which are the sum of cubes in two *or more* ways, and so appears identical in the first few terms.

N	W	Sloane	Numbers
1	1	000578	1, 8, 27, 64, 125, 216, 343, 512, ...
2	1	025403	2, 9, 16, 28, 35, 54, 65, 72, 91, ...
2	2		1729, 4104, 13832, 20683, 32832, ...
2	3	003825	87539319, 119824488, 143604279, ...
2	4	003826	6963472309248, 12625136269928, ...
2	5		48988659276962496, ...
2	6		8230545258248091551205888, ...
3	1	025395	3, 10, 17, 24, 29, 36, 43, 55, 62, ...

It is believed to be possible to express any number as a Sum of four (positive or negative) cubes, although this has not been proved for numbers of the form $9n \pm 4$. In fact, all numbers *not* of the form $9n \pm 4$ are known to be expressible as the Sum of *three* (positive or negative) cubes except 30, 33, 42, 52, 74, 110, 114, 156, 165, 195, 290, 318, 366, 390, 420, 435, 444, 452, 462, 478, 501, 530, 534, 564, 579, 588, 600, 606, 609, 618, 627, 633, 732, 735, 758, 767, 786, 789, 795, 830, 834, 861, 894, 903, 906, 912, 921, 933, 948, 964, 969, and 975 (Guy 1994, p. 151).

The following table gives the possible residues (mod n) for cubic numbers for $n = 1$ to 20, as well as the number of distinct residues $s(n)$.

n	$s(n)$	$x^3 \pmod n$
2	2	0, 1
3	3	0, 1, 2
4	3	0, 1, 3
5	5	0, 1, 2, 3, 4
6	6	0, 1, 2, 3, 4, 5
7	3	0, 1, 6
8	5	0, 1, 3, 5, 7
9	3	0, 1, 8
10	10	0, 1, 2, 3, 4, 5, 6, 7, 8, 9
11	11	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
12	9	0, 1, 3, 4, 5, 7, 8, 9, 11
13	5	0, 1, 5, 8, 12
14	6	0, 1, 6, 7, 8, 13
15	15	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14
16	10	0, 1, 3, 5, 7, 8, 9, 11, 13, 15
17	17	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16
18	6	0, 1, 8, 9, 10, 17
19	7	0, 1, 7, 8, 11, 12, 18
20	15	0, 1, 3, 4, 5, 7, 8, 9, 11, 12, 13, 15, 16, 17, 19

Dudeney found two RATIONAL NUMBERS other than 1 and 2 whose cubes sum to 9,

$$\frac{415280564497}{348671682660} \text{ and } \frac{676702467503}{348671682660}. \tag{4}$$

The problem of finding two RATIONAL NUMBERS whose cubes sum to six was "proved" impossible by Legendre. However, Dudeney found the simple solutions 17/21 and 37/21.

The only three consecutive INTEGERS whose cubes sum to a cube are given by the DIOPHANTINE EQUATION

$$3^3 + 4^3 + 5^3 = 6^3. \tag{5}$$

CATALAN'S CONJECTURE states that 8 and 9 (2^3 and 3^2) are the only consecutive POWERS (excluding 0 and 1), i.e., the only solution to CATALAN'S DIOPHANTINE PROBLEM. This CONJECTURE has not yet been proved or refuted, although R. Tijdeman has proved that there can be only a finite number of exceptions should the CONJECTURE not hold. It is also known that 8 and 9 are the only consecutive cubic and SQUARE NUMBERS (in either order).

There are six POSITIVE INTEGERS equal to the sum of the DIGITS of their cubes: 1, 8, 17, 18, 26, and 27 (Moret Blanc 1879). There are four POSITIVE INTEGERS equal to the sums of the cubes of their digits:

$$153 = 1^3 + 5^3 + 3^3 \tag{6}$$

$$370 = 3^3 + 7^3 + 0^3 \tag{7}$$

$$371 = 3^3 + 7^3 + 1^3 \tag{8}$$

$$407 = 4^3 + 0^3 + 7^3 \tag{9}$$

(Ball and Coxeter 1987). There are two SQUARE NUMBERS of the form $n^3 - 4$: $4 = 2^3 - 4$ and $121 = 5^3 - 4$ (Le Lionnais 1983). A cube cannot be the concatenation of two cubes, since if c^3 is the concatenation of a^3 and b^3, then $c^3 = 10^k a^3 + b^3$, where k is the number of digits in b^3. After shifting any powers of 1000 in 10^k into a^3, the original problem is equivalent to finding a solution to one of the DIOPHANTINE EQUATIONS

$$c^3 - b^3 = a^3 \tag{10}$$

$$c^3 - b^3 = 10a^3 \tag{11}$$

$$c^3 - b^3 = 100a^3. \tag{12}$$

None of these have solutions in integers, as proved independently by Sylvester, Lucas, and Pepin (Dickson 1966, pp. 572–578).

see also BIQUADRATIC NUMBER, CENTERED CUBE NUMBER, CLARK'S TRIANGLE, DIOPHANTINE EQUATION—CUBIC, HARDY-RAMANUJAN NUMBER, PARTITION, SQUARE NUMBER

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 14, 1987.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 42–44, 1996.

Davenport, H. "On Waring's Problem for Cubes." *Acta Math.* **71**, 123–143, 1939.

Dickson, L. E. *History of the Theory of Numbers, Vol. 2: Diophantine Analysis.* New York: Chelsea, 1966.

Guy, R. K. "Sum of Four Cubes." §D5 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 151–152, 1994.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 53, 1983.

Sloane, N. J. A. Sequences A000578/M4499, A02376/M0466, and A003108/M0209 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Cubic Reciprocity Theorem

A RECIPROCITY THEOREM for the case $n = 3$ solved by Gauss using "INTEGERS" of the form $a + b\rho$, when ρ is a root if $x^2 + x + 1 = 0$ and a, b are INTEGERS.

see also RECIPROCITY THEOREM

References

Ireland, K. and Rosen, M. "Cubic and Biquadratic Reciprocity." Ch. 9 in *A Classical Introduction to Modern Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 108–137, 1990.

Cubic Spline

A cubic spline is a SPLINE constructed of piecewise third-order POLYNOMIALS which pass through a set of control points. The second DERIVATIVE of each POLYNOMIAL is zero at the endpoints.

References

Burden, R. L.; Faires, J. D.; and Reynolds, A. C. *Numerical Analysis, 6th ed.* Boston, MA: Brooks/Cole, pp. 120–121, 1997.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Cubic Spline Interpolation." §3.3 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 107–110, 1992.

Cubic Surface

An ALGEBRAIC SURFACE of ORDER 3. Schläfli and Cayley classified the singular cubic surfaces. On the general cubic, there exists a curious geometrical structure called DOUBLE SIXES, and also a particular arrangement of 27 (possibly complex) lines, as discovered by Schläfli (Salmon 1965, Fischer 1986) and sometimes called SOLOMON'S SEAL LINES. A nonregular cubic surface can contain 3, 7, 15, or 27 real lines (Segre 1942, Le Lionnais 1983). The CLEBSCH DIAGONAL CUBIC contains all possible 27. The maximum number of ORDINARY DOUBLE POINTS on a cubic surface is four, and the unique cubic surface having four ORDINARY DOUBLE POINTS is the CAYLEY CUBIC.

Schoutte (1910) showed that the 27 lines can be put into a ONE-TO-ONE correspondence with the vertices of a particular POLYTOPE in 6-D space in such a manner that all incidence relations between the lines are mirrored in the connectivity of the POLYTOPE and conversely (Du Val 1931). A similar correspondence can be made between the 28 bitangents of the general plane QUARTIC CURVE and a 7-D POLYTOPE (Coxeter 1928) and between the tritangent planes of the canonical curve of genus 4 and an 8-D POLYTOPE (Du Val 1933).

A smooth cubic surface contains 45 TRITANGENTS (Hunt). The Hessian of smooth cubic surface contains at least 10 ORDINARY DOUBLE POINTS, although the Hessian of the CAYLEY CUBIC contains 14 (Hunt).

see also CAYLEY CUBIC, CLEBSCH DIAGONAL CUBIC, DOUBLE SIXES, ECKARDT POINT, ISOLATED SINGULARITY, NORDSTRAND'S WEIRD SURFACE, SOLOMON'S SEAL LINES, TRITANGENT

References

Bruce, J. and Wall, C. T. C. "On the Classification of Cubic Surfaces." *J. London Math. Soc.* **19**, 245–256, 1979.

Cayley, A. "A Memoir on Cubic Surfaces." *Phil. Trans. Roy. Soc.* **159**, 231–326, 1869.

Coxeter, H. S. M. "The Pure Archimedean Polytopes in Six and Seven Dimensions." *Proc. Cambridge Phil. Soc.* **24**, 7–9, 1928.

Du Val, P. "On the Directrices of a Set of Points in a Plane." *Proc. London Math. Soc. Ser. 2* **35**, 23–74, 1933.

Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, pp. 9–14, 1986.

Fladt, K. and Baur, A. *Analytische Geometrie spezieler Flächen und Raumkurven.* Braunschweig, Germany: Vieweg, pp. 248–255, 1975.

Hunt, B. "Algebraic Surfaces." `http://www.mathematik.uni-kl.de/~wwwagag/Galerie.html`.

Hunt, B. "The 27 Lines on a Cubic Surface" and "Cubic Surfaces." Ch. 4 and Appendix B.4 in *The Geometry of Some Special Arithmetic Quotients.* New York: Springer-Verlag, pp. 108–167 and 302–310, 1996.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 49, 1983.

Rodenberg, C. "Zur Classification der Flächen dritter Ordnung." *Math. Ann.* **14**, 46–110, 1878.

Salmon, G. *Analytic Geometry of Three Dimensions.* New York: Chelsea, 1965.

Schläfli, L. "On the Distribution of Surface of Third Order into Species." *Phil. Trans. Roy. Soc.* **153**, 193–247, 1864.

Schoutte, P. H. "On the Relation Between the Vertices of a Definite Sixdimensional Polytope and the Lines of a Cubic Surface." *Proc. Roy. Acad. Amsterdam* **13**, 375–383, 1910.

Segre, B. *The Nonsingular Cubic Surface.* Oxford, England: Clarendon Press, 1942.

Cubical Conic Section

see CUBICAL ELLIPSE, CUBICAL HYPERBOLA, CUBICAL PARABOLA, SKEW CONIC

Cubical Ellipse

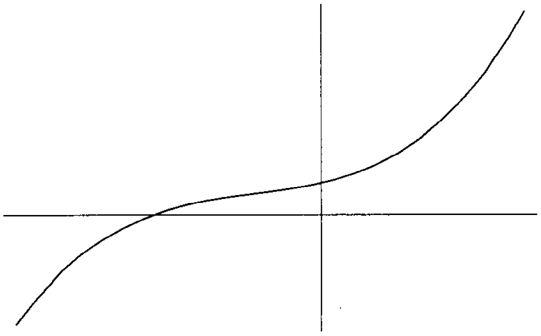

An equation of the form

$$y = ax^3 + bx^2 + cx + d$$

where only one ROOT is real.

see also CUBICAL CONIC SECTION, CUBICAL HYPERBOLA, CUBICAL PARABOLA, CUBICAL PARABOLIC HYPERBOLA, ELLIPSE, SKEW CONIC

Cubical Graph

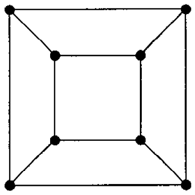

An 8-vertex POLYHEDRAL GRAPH.

see also BIDIAKIS CUBE, BISLIT CUBE, DODECAHEDRAL GRAPH, ICOSAHEDRAL GRAPH, OCTAHEDRAL GRAPH, TETRAHEDRAL GRAPH

Cubical Hyperbola

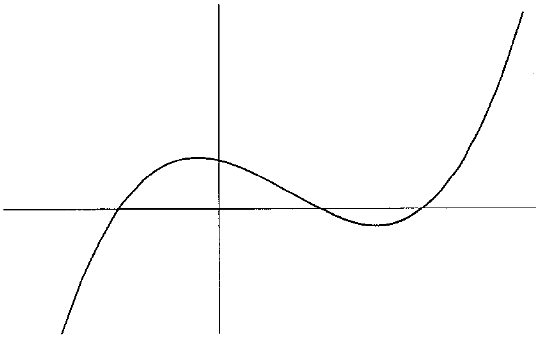

An equation of the form

$$y = ax^3 + bx^2 + cx + d,$$

where the three ROOTS are REAL and distinct, i.e.,

$$\begin{aligned} y &= a(x-r_1)(x-r_2)(x-r_3) \\ &= a[x^3 - (r_1+r_2+r_3)x^2 + (r_1r_2 + r_1r_3 + r_2r_3)x \\ &\quad - r_1r_2r_3]. \end{aligned}$$

see also CUBICAL CONIC SECTION, CUBICAL ELLIPSE, CUBICAL HYPERBOLA, CUBICAL PARABOLA, HYPERBOLA

Cubical Parabola

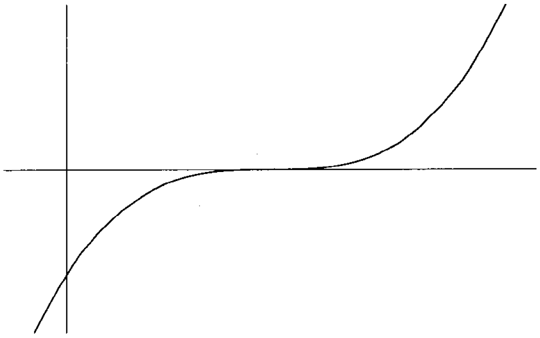

An equation of the form

$$y = ax^3 + bx^2 + cx + d,$$

where the three ROOTS of the equation coincide (and are therefore real), i.e.,

$$y = a(x-r)^3 = a(x^3 - 3rx^2 - 3r^2x - r^3).$$

see also CUBICAL CONIC SECTION, CUBICAL ELLIPSE, CUBICAL HYPERBOLA, CUBICAL PARABOLIC HYPERBOLA, PARABOLA, SEMICUBICAL PARABOLA

Cubical Parabolic Hyperbola

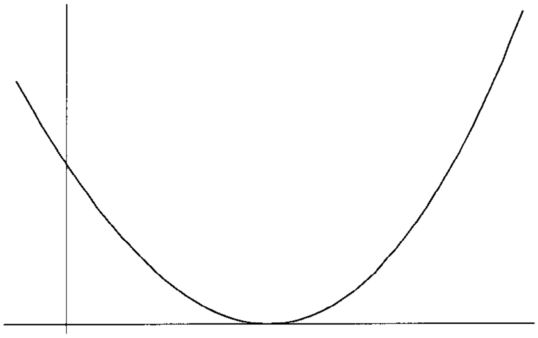

An equation of the form

$$y = ax^3 + bx^2 + cx + d,$$

where two of the ROOTS of the equation coincide (and all three are therefore real), i.e.,

$$\begin{aligned} y &= a(x-r_1)^2(x-r_2) \\ &= a[x^3 - (2r_1+r_2)x^2 + r_1(r_1+2r_2)x - {r_1}^2r_2]. \end{aligned}$$

see also CUBICAL CONIC SECTION, CUBICAL ELLIPSE, CUBICAL HYPERBOLA, CUBICAL PARABOLA, HYPERBOLA

Cubicuboctahedron

see GREAT CUBICUBOCTAHEDRON, SMALL CUBICUBOCTAHEDRON

Cubique d'Agnesi

see WITCH OF AGNESI

Cubitruncated Cuboctahedron

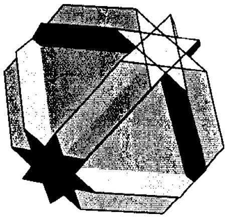

The UNIFORM POLYHEDRON U_{16} whose DUAL is the TETRADYAKIS HEXAHEDRON. It has WYTHOFF SYMBOL $3\,\tfrac{4}{3}\,4\,|$. Its faces are $8\{6\} + 6\{8\} + 6\{\tfrac{8}{3}\}$. It is a FACETED OCTAHEDRON. The CIRCUMRADIUS for a cubitruncated cuboctahedron of unit edge length is

$$R = \tfrac{1}{2}\sqrt{7}.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 113–114, 1971.

Cuboctahedron

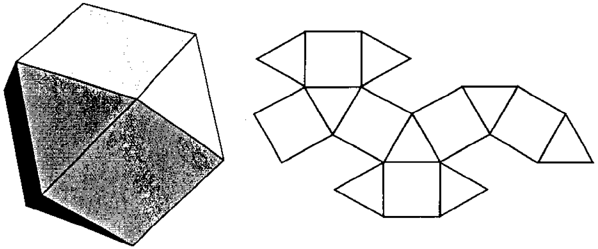

An ARCHIMEDEAN SOLID (also called the DYMAXION or HEPTAPARALLELOHEDRON) whose DUAL is the RHOMBIC DODECAHEDRON. It is one of the two convex QUASIREGULAR POLYHEDRA and has SCHLÄFLI SYMBOL $\left\{\begin{matrix}3\\4\end{matrix}\right\}$. It is also UNIFORM POLYHEDRON U_7 and has WYTHOFF SYMBOL $2\,|\,3\,4$. Its faces are $\{3\} + 6\{4\}$. It has the O_h OCTAHEDRAL GROUP of symmetries.

The VERTICES of a cuboctahedron with EDGE length of $\sqrt{2}$ are $(0, \pm 1, \pm 1)$, $(\pm 1, 0, \pm 1)$, and $(\pm 1, \pm 1, 0)$. The INRADIUS, MIDRADIUS, and CIRCUMRADIUS for $a = 1$ are

$$\begin{aligned} r &= \tfrac{3}{4} = 0.75 \\ \rho &= \tfrac{1}{2}\sqrt{3} \approx 0.86602 \\ R &= 1. \end{aligned}$$

FACETED versions include the CUBOHEMIOCTAHEDRON and OCTAHEMIOCTAHEDRON.

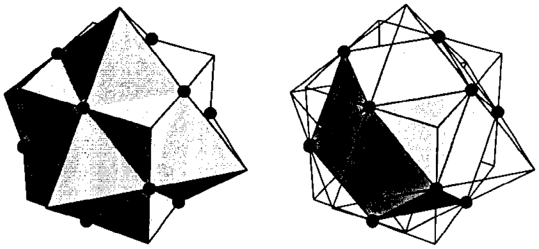

The solid common to both the CUBE and OCTAHEDRON (left figure) in a CUBE-OCTAHEDRON COMPOUND is a CUBOCTAHEDRON (right figure; Ball and Coxeter 1987).

see also ARCHIMEDEAN SOLID, CUBE, CUBE-OCTAHEDRON COMPOUND, CUBOHEMIOCTAHEDRON, OCTAHEDRON, OCTAHEMIOCTAHEDRON, QUASIREGULAR POLYHEDRON, RHOMBIC DODECAHEDRON, RHOMBUS

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 137, 1987.

Ghyka, M. *The Geometry of Art and Life.* New York: Dover, p. 54, 1977.

Cuboctatruncated Cuboctahedron

see CUBITRUNCATED CUBOCTAHEDRON

Cubocycloid

see ASTROID

Cubohemioctahedron

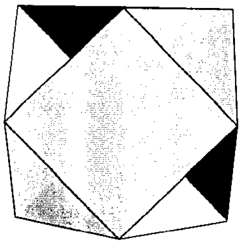

The UNIFORM POLYHEDRON U_{15} whose DUAL is the HEXAHEMIOCTAHEDRON. It has WYTHOFF SYMBOL $\frac{4}{3}\,4\,|\,3$. Its faces are $4\{6\} + 6\{4\}$. It is a FACETED version of the CUBOCTAHEDRON. Its CIRCUMRADIUS for unit edge length is

$$R = 1.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 121–122, 1971.

Cuboid

A rectangular PARALLELEPIPED.

see also EULER BRICK, PARALLELEPIPED, SPIDER AND FLY PROBLEM

Cullen Number

A number of the form

$$C_n = 2^n n + 1.$$

The first few are 3, 9, 25, 65, 161, 385, ... (Sloane's A002064). The only Cullen numbers C_n for $n < 300,000$ which are PRIME are for $n = 1$, 141, 4713, 5795, 6611, 18496, 32292, 32469, 59656, 90825, 262419, ... (Sloane's A005849; Ballinger). Cullen numbers are DIVISIBLE by $p = 2n - 1$ if p is a PRIME of the form $8k \pm 3$.

see also CUNNINGHAM NUMBER, FERMAT NUMBER, SIERPIŃSKI NUMBER OF THE FIRST KIND, WOODALL NUMBER

References

Ballinger, R. "Cullen Primes: Definition and Status." `http://ballingerr.xray.ufl.edu/proths/cullen.html`.

Guy, R. K. "Cullen Numbers." §B20 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 77, 1994.

Keller, W. "New Cullen Primes." *Math. Comput.* **64**, 1733–1741, 1995.

Leyland, P. `ftp://sable.ox.ac.uk/pub/math/factors/cullen`.

Ribenboim, P. *The New Book of Prime Number Records.* New York: Springer-Verlag, pp. 360–361, 1996.

Sloane, N. J. A. Sequences A002064/M2795 and A005849/M5401 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Cumulant

Let $\phi(t)$ be the CHARACTERISTIC FUNCTION, defined as the FOURIER TRANSFORM of the PROBABILITY DENSITY FUNCTION,

$$\phi(t) = \mathcal{F}[P(x)] = \int_{-\infty}^{\infty} e^{itx} P(x)\,dx. \tag{1}$$

Then the cumulants κ_n are defined by

$$\ln \phi(t) \equiv \sum_{n=0}^{\infty} \kappa_n \frac{(it)^n}{n!}. \tag{2}$$

Taking the MACLAURIN SERIES gives

$$\begin{aligned} \ln \phi(t) = {} & (it)\mu_1' + \tfrac{1}{2}(it)^2(\mu_2' - {\mu_1'}^2) \\ & + \tfrac{1}{3!}(it)^3(2{\mu_1'}^3 - 3\mu_1'\mu_2' + \mu_3') \\ & + \tfrac{1}{4!}(it)^4(-6{\mu_1'}^4 + 12{\mu_1'}^2\mu_2' - 3{\mu_2'}^2 - 4\mu_1'\mu_3' + \mu_4') \\ & + \tfrac{1}{5!}(it)^5[-24{\mu_1'}^5 + 60{\mu_1'}^3\mu_2' + 20{\mu_1'}^2\mu_3' + 10\mu_2'\mu_3' \\ & + 5\mu_1'(6{\mu_2'}^2 - \mu_4') + \mu_5'] + \ldots, \end{aligned} \tag{3}$$

where μ'_n are MOMENTS about 0, so

$$\kappa_1 = \mu'_1 \tag{4}$$

$$\kappa_2 = \mu'_2 - {\mu'_1}^2 \tag{5}$$

$$\kappa_3 = 2{\mu'_1}^3 - 3\mu'_1\mu'_2 + \mu'_3 \tag{6}$$

$$\kappa_4 = -6{\mu'_1}^4 + 12{\mu'_1}^2\mu'_2 - 3{\mu'_2}^2 - 4\mu'_1\mu'_3 + \mu'_4 \tag{7}$$

$$\begin{aligned}\kappa_5 = {} & -24{\mu'_1}^5 + 60{\mu'_1}^3\mu'_2 + 20{\mu'_1}^2\mu'_3 + 10\mu'_2\mu'_3 \\ & + 5\mu'_1(6{\mu'_2}^2 - \mu'_4) + \mu'_5.\end{aligned} \tag{8}$$

In terms of the MOMENTS μ_n about the MEAN,

$$\kappa_1 = \mu \tag{9}$$

$$\kappa_2 = \mu_2 = \sigma^2 \tag{10}$$

$$\kappa_3 = \mu_3 \tag{11}$$

$$\kappa_4 = \mu_4 - 3{\mu_2}^2 \tag{12}$$

$$\kappa_5 = \mu_5 - 10\mu_2\mu_3, \tag{13}$$

where μ is the MEAN and $\sigma^2 \equiv \mu_2$ is the VARIANCE.

The k-STATISTICS are UNBIASED ESTIMATORS of the cumulants.

see also CHARACTERISTIC FUNCTION, CUMULANT-GENERATING FUNCTION, k-STATISTIC, KURTOSIS, MEAN, MOMENT, SHEPPARD'S CORRECTION, SKEWNESS, VARIANCE

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 928, 1972.
Kenney, J. F. and Keeping, E. S. "Cumulants and the Cumulant-Generating Function," "Additive Property of Cumulants," and "Sheppard's Correction." §4.10–4.12 in *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, pp. 77–82, 1951.

Cumulant-Generating Function

Let $M(h)$ be the MOMENT-GENERATING FUNCTION. Then

$$K(h) \equiv \ln M(h) = \kappa_1 h + \tfrac{1}{2!}h^2\kappa_2 + \tfrac{1}{3!}h^3\kappa_3 + \ldots.$$

If

$$L = \sum_{j=1}^{M} c_j x_j$$

is a function of N independent variables, the cumulant generating function for L is then

$$K(h) = \sum_{j=1}^{N} K_j(c_j h).$$

see also CUMULANT, MOMENT-GENERATING FUNCTION

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 928, 1972.
Kenney, J. F. and Keeping, E. S. "Cumulants and the Cumulant-Generating Function" and "Additive Property of Cumulants." §4.10–4.11 in *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, pp. 77–80, 1951.

Cumulative Distribution Function

see DISTRIBUTION FUNCTION

Cundy and Rollett's Egg

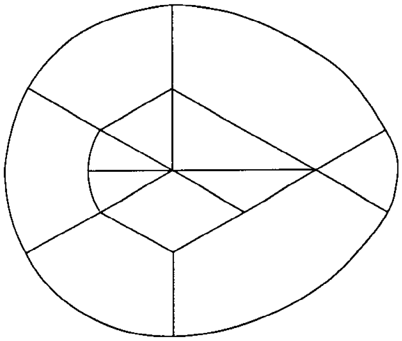

An OVAL dissected into pieces which are to used to create pictures. The resulting figures resemble those constructed out of TANGRAMS.

see also DISSECTION, EGG, OVAL, TANGRAM

References
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 19–21, 1989.
Dixon, R. *Mathographics.* New York: Dover, p. 11, 1991.

Cunningham Chain

A SEQUENCE of PRIMES $q_1 < q_2 < \ldots < q_k$ is a Cunningham chain of the first kind (second kind) of length k if $q_{i+1} = 2q_i + 1$ ($q_{i+1} = 2q_i - 1$) for $i = 1, \ldots, k-1$. Cunningham PRIMES of the first kind are SOPHIE GERMAIN PRIMES.

The two largest known Cunningham chains (of the first kind) of length three are $(384205437 \cdot 2^{4000} - 1,\ 384205437 \cdot 2^{4001} - 1,\ 384205437 \cdot 2^{4002} - 1)$ and $(651358155 \cdot 2^{3291} - 1,\ 651358155 \cdot 2^{3292} - 1,\ 651358155 \cdot 2^{3293} - 1)$, both discovered by W. Roonguthai in 1998.

see also PRIME ARITHMETIC PROGRESSION, PRIME CLUSTER

References
Guy, R. K. "Cunningham Chains." §A7 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 18–19, 1994.
Ribenboim, P. *The New Book of Prime Number Records.* New York: Springer-Verlag, p. 333, 1996.
Roonguthai, W. "Yves Gallot's Proth and Cunningham Chains." `http://ksc9.th.com/warut/cunningham.html`.

Cunningham Function

Sometimes also called the PEARSON-CUNNINGHAM FUNCTION. It can be expressed using WHITTAKER FUNCTIONS (Whittaker and Watson 1990, p. 353).

$$\omega_{n,m}(x) \equiv \frac{e^{\pi i(m/2-n)+x}}{\Gamma(1+n-\frac{1}{2}m)} U(\tfrac{1}{2}m-n, 1+m, x),$$

where U is a CONFLUENT HYPERGEOMETRIC FUNCTION OF THE SECOND KIND (Abramowitz and Stegun 1972, p. 510).

see also CONFLUENT HYPERGEOMETRIC FUNCTION OF THE SECOND KIND, WHITTAKER FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, 1972.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Cunningham Number

A BINOMIAL NUMBER of the form $C^{\pm}(b,n) \equiv b^n \pm 1$. Bases b^k which are themselves powers need not be considered since they correspond to $(b^k)^n \pm 1 = b^{kn} \pm 1$. PRIME NUMBERS of the form $C^{\pm}(b,n)$ are very rare.

A NECESSARY (but not SUFFICIENT) condition for $C^+(2,n) = 2^n + 1$ to be PRIME is that n be of the form $n = 2^m$. Numbers of the form $F_m = C^+(2,2^m) = 2^{2^m} + 1$ are called FERMAT NUMBERS, and the only known PRIMES occur for $C^+(2,1) = 3$, $C^+(2,2) = 5$, $C^+(2,4) = 17$, $C^+(2,8) = 257$, and $C^+(2,16) = 65537$ (i.e., $n = 0$, 1, 2, 3, 4). The only other PRIMES $C^+(b,n)$ for nontrivial $b \leq 11$ and $2 \leq n \leq 1000$ are $C^+(6,2) = 37$, $C^+(6,4) = 1297$, and $C^+(10,2) = 101$.

PRIMES of the form $C^-(b,n)$ are also very rare. The MERSENNE NUMBERS $M_n = C^-(2,n) = 2^n - 1$ are known to be prime only for 37 values, the first few of which are $n = 2$, 3, 5, 7, 13, 17, 19, ... (Sloane's A000043). There are no other PRIMES $C^-(b,n)$ for nontrivial $b \leq 20$ and $2 \leq n \leq 1000$.

In 1925, Cunningham and Woodall (1925) gathered together all that was known about the PRIMALITY and factorization of the numbers $C^{\pm}(b,n)$ and published a small book of tables. These tables collected from scattered sources the known prime factors for the bases 2 and 10 and also presented the authors' results of 30 years' work with these and other bases.

Since 1925, many people have worked on filling in these tables. D. H. Lehmer, a well-known mathematician who died in 1991, was for many years a leader of these efforts. Lehmer was a mathematician who was at the forefront of computing as modern electronic computers became a reality. He was also known as the inventor of some ingenious pre-electronic computing devices specifically designed for factoring numbers.

Updated factorizations were published in Brillhart *et al.* (1988). The current archive of Cunningham number factorizations for $b = 1, \ldots, \pm 12$ is kept on `ftp://sable.ox.ac.uk/pub/math/cunningham`. The tables have been extended by Brent and te Riele (1992) to $b = 13, \ldots, 100$ with $m < 255$ for $b < 30$ and $m < 100$ for $b \geq 30$. All numbers with exponent 58 and smaller, and all composites with ≤ 90 digits have now been factored.

see also BINOMIAL NUMBER, CULLEN NUMBER, FERMAT NUMBER, MERSENNE NUMBER, REPUNIT, RIESEL NUMBER, SIERPIŃSKI NUMBER OF THE FIRST KIND, WOODALL NUMBER

References

Brent, R. P. and te Riele, H. J. J. "Factorizations of $a^n \pm 1$, $13 \leq a < 100$." Report NM-R9212, *Centrum voor Wiskunde en Informatica.* Amsterdam, June 1992. The text is available electronically at `ftp://sable.ox.ac.uk/pub/math/factors/BMtR_13-99.dvi`, and the files at `BMtR_13-99`. Updates are given in `BMtR_13-99_update1` (94-09-01) and `BMtR_13-99_update2` (95-06-01).

Brillhart, J.; Lehmer, D. H.; Selfridge, J.; Wagstaff, S. S. Jr.; and Tuckerman, B. *Factorizations of* $b^n \pm 1$, $b = 2, 3, 5, 6, 7, 10, 11, 12$ *Up to High Powers, rev. ed.* Providence, RI: Amer. Math. Soc., 1988. Updates are available electronically from `ftp://sable.ox.ac.uk/pub/math/cunningham/`.

Cunningham, A. J. C. and Woodall, H. J. *Factorisation of* $y^n \mp 1$, $y = 2, 3, 5, 6, 7, 10, 11, 12$ *Up to High Powers (n).* London: Hodgson, 1925.

Mudge, M. "Not Numerology but Numeralogy!" *Personal Computer World,* 279–280, 1997.

Ribenboim, P. "Numbers $k \times 2^n \pm 1$." §5.7 in *The New Book of Prime Number Records.* New York: Springer-Verlag, pp. 355–360, 1996.

Sloane, N. J. A. Sequence A000043/M0672 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Cunningham Project

see CUNNINGHAM NUMBER

Cupola

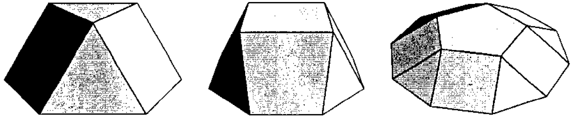

An n-gonal cupola Q_n (possible for only $n = 3$, 4, 5) is a POLYHEDRON having n TRIANGULAR and n SQUARE faces separating an $\{n\}$ and a $\{2n\}$ REGULAR POLYGON. The coordinates of the base VERTICES are

$$\left(R\cos\left[\frac{\pi(2k+1)}{2n}\right], R\sin\left[\frac{\pi(2k+1)}{2n}\right], 0\right), \qquad (1)$$

and the coordinates of the top VERTICES are

$$\left(r\cos\left[\frac{2k\pi}{n}\right], r\sin\left[\frac{2k\pi}{n}\right], z\right), \qquad (2)$$

where R and r are the CIRCUMRADII of the base and top

$$R = \tfrac{1}{2} a \csc\left(\frac{\pi}{2n}\right) \tag{3}$$

$$r = \tfrac{1}{2} a \csc\left(\frac{\pi}{n}\right), \tag{4}$$

and z is the height, obtained by letting $k = 0$ in the equations (1) and (2) to obtain the coordinates of neighboring bottom and top VERTICES,

$$\mathbf{b} = \begin{bmatrix} R\cos\left(\frac{\pi}{2n}\right) \\ R\sin\left(\frac{\pi}{2n}\right) \\ 0 \end{bmatrix} \tag{5}$$

$$\mathbf{t} = \begin{bmatrix} r \\ 0 \\ z \end{bmatrix}. \tag{6}$$

Since all side lengths are a,

$$|\mathbf{b} - \mathbf{t}|^2 = a^2. \tag{7}$$

Solving for z then gives

$$\left[R\cos\left(\frac{\pi}{2n}\right) - r\right]^2 + R^2 \sin^2\left(\frac{\pi}{2n}\right) + z^2 = a^2 \tag{8}$$

$$z^2 + R^2 + r^2 - 2rR\cos\left(\frac{\pi}{2n}\right) = a^2 \tag{9}$$

$$\begin{aligned} z &= \sqrt{a^2 - 2rR\cos\left(\frac{\pi}{2n}\right) - r^2 - R^2} \\ &= a\sqrt{1 - \tfrac{1}{4}\csc^2\left(\frac{\pi}{n}\right)}. \end{aligned} \tag{10}$$

see also BICUPOLA, ELONGATED CUPOLA, GYROELONGATED CUPOLA, PENTAGONAL CUPOLA, SQUARE CUPOLA, TRIANGULAR CUPOLA

References

Johnson, N. W. "Convex Polyhedra with Regular Faces." *Canad. J. Math.* **18**, 169–200, 1966.

Cupolarotunda

A CUPOLA adjoined to a ROTUNDA.

see also GYROCUPOLAROTUNDA, ORTHOCUPOLAROTUNDA

Curl

The curl of a TENSOR field is given by

$$(\nabla \times A)^\alpha = \epsilon^{\alpha\mu\nu} A_{\nu;\mu}, \tag{1}$$

where ϵ_{ijk} is the LEVI-CIVITA TENSOR and ";" is the COVARIANT DERIVATIVE. For a VECTOR FIELD, the curl is denoted

$$\operatorname{curl}(\mathbf{F}) \equiv \nabla \times \mathbf{F}, \tag{2}$$

and $\nabla \times \mathbf{F}$ is normal to the PLANE in which the "circulation" is MAXIMUM. Its magnitude is the limiting value of circulation per unit AREA,

$$(\nabla \times \mathbf{F}) \cdot \hat{\mathbf{n}} \equiv \lim_{A \to 0} \frac{\oint_C \mathbf{F} \cdot d\mathbf{s}}{A}. \tag{3}$$

Let

$$\mathbf{F} \equiv F_1 \hat{\mathbf{u}}_1 + F_2 \hat{\mathbf{u}}_2 + F_3 \hat{\mathbf{u}}_3 \tag{4}$$

and

$$h_i \equiv \left|\frac{\partial \mathbf{r}}{\partial u_i}\right|, \tag{5}$$

then

$$\begin{aligned} \nabla \times \mathbf{F} &\equiv \frac{1}{h_1 h_2 h_3} \begin{vmatrix} h_1 \hat{\mathbf{u}}_1 & h_2 \hat{\mathbf{u}}_2 & h_3 \hat{\mathbf{u}}_3 \\ \frac{\partial}{\partial u_1} & \frac{\partial}{\partial u_2} & \frac{\partial}{\partial u_3} \\ h_1 F_1 & h_2 F_2 & h_3 F_3 \end{vmatrix} \\ &= \frac{1}{h_2 h_3}\left[\frac{\partial}{\partial u_2}(h_3 F_3) - \frac{\partial}{\partial u_3}(h_2 F_2)\right] \hat{\mathbf{u}}_1 \\ &\quad + \frac{1}{h_1 h_3}\left[\frac{\partial}{\partial u_3}(h_1 F_1) - \frac{\partial}{\partial u_1}(h_3 F_3)\right] \hat{\mathbf{u}}_2 \\ &\quad + \frac{1}{h_1 h_2}\left[\frac{\partial}{\partial u_1}(h_2 F_2) - \frac{\partial}{\partial u_2}(h_1 F_1)\right] \hat{\mathbf{u}}_3. \end{aligned} \tag{6}$$

Special cases of the curl formulas above can be given for CURVILINEAR COORDINATES.

see also CURL THEOREM, DIVERGENCE, GRADIENT, VECTOR DERIVATIVE

References

Arfken, G. "Curl, $\nabla\times$." §1.8 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 42–47, 1985.

Curl Theorem

A special case of STOKES' THEOREM in which F is a VECTOR FIELD and M is an oriented, compact embedded 2-MANIFOLD with boundary in $\mathbb{R}^3$, given by

$$\int_S (\nabla \times \mathbf{F}) \cdot d\mathbf{a} = \int_{\partial S} \mathbf{F} \cdot d\mathbf{s}. \tag{1}$$

There are also alternate forms. If

$$\mathbf{F} \equiv \mathbf{c}F, \tag{2}$$

then

$$\int_S d\mathbf{a} \times \nabla F = \int_C F d\mathbf{s}. \tag{3}$$

and if

$$\mathbf{F} \equiv \mathbf{c} \times \mathbf{P}, \tag{4}$$

then

$$\int_S (d\mathbf{a} \times \nabla) \times \mathbf{P} = \int_C d\mathbf{s} \times \mathbf{P}. \tag{5}$$

see also CHANGE OF VARIABLES THEOREM, CURL, STOKES' THEOREM

References

Arfken, G. "Stokes's Theorem." §1.12 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 61–64, 1985.

Curlicue Fractal

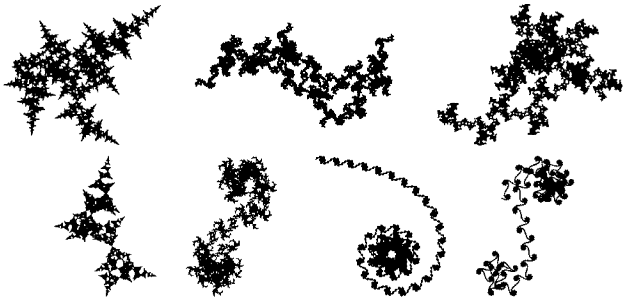

The curlicue fractal is a figure obtained by the following procedure. Let s be an IRRATIONAL NUMBER. Begin with a line segment of unit length, which makes an ANGLE $\phi_0 \equiv 0$ to the horizontal. Then define θ_n iteratively by

$$\theta_{n+1} = (\theta_n + 2\pi s) \pmod{2\pi},$$

with $\theta_0 = 0$. To the end of the previous line segment, draw a line segment of unit length which makes an angle

$$\phi_{n+1} = \theta_n + \phi_n \pmod{2\pi},$$

to the horizontal (Pickover 1995). The result is a FRACTAL, and the above figures correspond to the curlicue fractals with 10,000 points for the GOLDEN RATIO ϕ, $\ln 2$, e, $\sqrt{2}$, the EULER-MASCHERONI CONSTANT γ, π, and FEIGENBAUM CONSTANT δ.

The TEMPERATURE of these curves is given in the following table.

Constant	Temperature
golden ratio ϕ	46
$\ln 2$	51
e	58
$\sqrt{2}$	58
Euler-Mascheroni constant γ	63
π	90
Feigenbaum constant δ	92

References

Berry, M. and Goldberg, J. "Renormalization of Curlicues." *Nonlinearity* **1**, 1–26, 1988.

Moore, R. and van der Poorten, A. "On the Thermodynamics of Curves and Other Curlicues." *McQuarie Univ. Math. Rep.* 89-0031, April 1989.

Pickover, C. A. "The Fractal Golden Curlicue is Cool." Ch. 21 in *Keys to Infinity.* New York: W. H. Freeman, pp. 163–167, 1995.

Pickover, C. A. *Mazes for the Mind: Computers and the Unexpected.* New York: St. Martin's Press, 1993.

Sedgewick, R. *Algorithms.* Reading, MA: Addison-Wesley, 1988.

Stewart, I. *Another Fine Math You've Got Me Into....* New York: W. H. Freeman, 1992.

Current

A linear FUNCTIONAL on a smooth differential form.

see also FLAT NORM, INTEGRAL CURRENT, RECTIFIABLE CURRENT

Curtate Cycloid

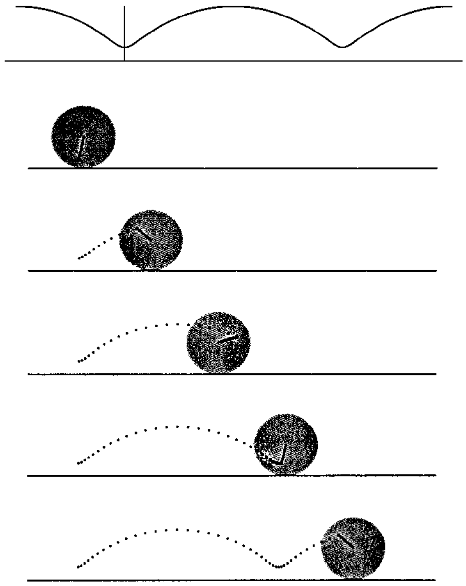

The path traced out by a fixed point at a RADIUS $b < a$, where a is the RADIUS of a rolling CIRCLE, sometimes also called a CONTRACTED CYCLOID.

$$x = a\phi - b\sin\phi \tag{1}$$
$$y = a - b\cos\phi. \tag{2}$$

The ARC LENGTH from $\phi = 0$ is

$$s = 2(a+b)E(u), \tag{3}$$

where

$$\sin(\tfrac{1}{2}\phi) = \operatorname{sn} u \tag{4}$$
$$k^2 = \frac{4ab}{(a+c)^2}, \tag{5}$$

and $E(u)$ is a complete ELLIPTIC INTEGRAL OF THE SECOND KIND and $\operatorname{sn} u$ is a JACOBI ELLIPTIC FUNCTION.

see also CYCLOID, PROLATE CYCLOID

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., 1989.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 46–50, 1991.

Curtate Cycloid Evolute

The EVOLUTE of the CURTATE CYCLOID

$$x = a\phi - b\sin\phi \tag{1}$$
$$y = a - b\cos\phi. \tag{2}$$

is given by

$$x = \frac{a[-2b\phi + 2a\phi\cos\phi - 2a\sin\phi + b\sin(2\phi)]}{2(a\cos\phi - b)} \tag{3}$$
$$y = \frac{a(a - b\cos\phi)^2}{b(a\cos\phi - b)}. \tag{4}$$

Curvature

In general, there are two important types of curvature: EXTRINSIC CURVATURE and INTRINSIC CURVATURE. The EXTRINSIC CURVATURE of curves in 2- and 3-space was the first type of curvature to be studied historically, culminating in the FRENET FORMULAS, which describe a SPACE CURVE entirely in terms of its "curvature," TORSION, and the initial starting point and direction.

After the curvature of 2- and 3-D curves was studied, attention turned to the curvature of surfaces in 3-space. The main curvatures which emerged from this scrutiny are the MEAN CURVATURE, GAUSSIAN CURVATURE, and the WEINGARTEN MAP. MEAN CURVATURE was the most important for applications at the time and was the most studied, but Gauss was the first to recognize the importance of the GAUSSIAN CURVATURE.

Because GAUSSIAN CURVATURE is "intrinsic," it is detectable to 2-dimensional "inhabitants" of the surface, whereas MEAN CURVATURE and the WEINGARTEN MAP are not detectable to someone who can't study the 3-dimensional space surrounding the surface on which he resides. The importance of GAUSSIAN CURVATURE to an inhabitant is that it controls the surface AREA of SPHERES around the inhabitant.

Riemann and many others generalized the concept of curvature to SECTIONAL CURVATURE, SCALAR CURVATURE, the RIEMANN TENSOR, RICCI CURVATURE, and a host of other INTRINSIC and EXTRINSIC CURVATURES. General curvatures no longer need to be numbers, and can take the form of a MAP, GROUP, GROUPOID, tensor field, etc.

The simplest form of curvature and that usually first encountered in CALCULUS is an EXTRINSIC CURVATURE. In 2-D, let a PLANE CURVE be given by CARTESIAN parametric equations $x = x(t)$ and $y = y(t)$. Then the curvature κ is defined by

$$\kappa \equiv \frac{d\phi}{ds} = \frac{\frac{d\phi}{dt}}{\frac{ds}{dt}} = \frac{\frac{d\phi}{dt}}{\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}} = \frac{\frac{d\phi}{dt}}{\sqrt{x'^2 + y'^2}}, \tag{1}$$

where ϕ is the POLAR ANGLE and s is the ARC LENGTH. As can readily be seen from the definition, curvature therefore has units of inverse distance. The $d\phi/dt$ derivative in the above equation can be eliminated by using the identity

$$\tan\phi = \frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{y'}{x'}, \tag{2}$$

so

$$\frac{d}{dt}(\tan\phi) = \sec^2\phi \frac{d\phi}{dt} = \frac{x'y'' - y'x''}{x'^2} \tag{3}$$

and

$$\begin{aligned}\frac{d\phi}{dt} &= \frac{1}{1 + \tan^2\phi} \frac{x'y'' - y'x''}{x'^2} \\ &= \frac{1}{\frac{y'^2}{x'^2}} \frac{x'y'' - y'x''}{x'^2} = \frac{x'y'' - y'x''}{x'^2 + y'^2}.\end{aligned} \tag{4}$$

Combining (2) and (4) gives

$$\kappa = \frac{x'y'' - y'x''}{(x'^2 + y'^2)^{3/2}}. \tag{5}$$

For a 2-D curve written in the form $y = f(x)$, the equation of curvature becomes

$$\kappa = \frac{\frac{d^2y}{dx^2}}{\left[1 + \left(\frac{dy}{dx}\right)^2\right]^{3/2}}. \tag{6}$$

If the 2-D curve is instead parameterized in POLAR COORDINATES, then

$$\kappa = \frac{r^2 + 2{r_\theta}^2 - r r_{\theta\theta}}{(r^2 + {r_\theta}^2)^{3/2}}, \tag{7}$$

where $r_\theta \equiv \partial r/\partial\theta$ (Gray 1993). In PEDAL COORDINATES, the curvature is given by

$$\kappa = \frac{1}{r}\frac{dp}{dr}. \tag{8}$$

The curvature for a 2-D curve given implicitly by $g(x, y) = 0$ is given by

$$\kappa = \frac{g_{xx}{g_y}^2 - 2g_{xy}g_x g_y + g_{yy}{g_x}^2}{({g_x}^2 + {g_y}^2)^{3/2}} \tag{9}$$

(Gray 1993).

Now consider a parameterized SPACE CURVE $\mathbf{r}(t)$ in 3-D for which the TANGENT VECTOR $\hat{\mathbf{T}}$ is defined as

$$\hat{\mathbf{T}} \equiv \frac{\frac{d\mathbf{r}}{dt}}{\left|\frac{d\mathbf{r}}{dt}\right|} = \frac{\frac{d\mathbf{r}}{dt}}{\frac{ds}{dt}}. \tag{10}$$

Therefore,

$$\frac{d\mathbf{r}}{dt} = \frac{ds}{dt}\hat{\mathbf{T}} \tag{11}$$

$$\frac{d^2\mathbf{r}}{dt^2} = \frac{d^2s}{dt^2}\hat{\mathbf{T}} + \frac{ds}{dt}\frac{d\hat{\mathbf{T}}}{dt} = \frac{ds^2}{dt^2}\hat{\mathbf{T}} + \kappa\hat{\mathbf{N}}\left(\frac{ds}{dt}\right)^2, \tag{12}$$

where $\hat{\mathbf{N}}$ is the NORMAL VECTOR. But

$$\begin{aligned}\frac{d\mathbf{r}}{dt} \times \frac{d^2\mathbf{r}}{dt^2} &= \frac{ds}{dt}\frac{d^2s}{dt^2}(\hat{\mathbf{T}} \times \hat{\mathbf{T}}) + \kappa\left(\frac{ds}{dt}\right)^3 (\hat{\mathbf{T}} \times \hat{\mathbf{N}}) \\ &= \kappa\left(\frac{ds}{dt}\right)^3 (\hat{\mathbf{T}} \times \hat{\mathbf{N}})\end{aligned} \tag{13}$$

$$\left|\frac{d\mathbf{r}}{dt}\times\frac{d^2\mathbf{r}}{dt^2}\right| = \kappa\left(\frac{ds}{dt}\right)^3 = \kappa\left|\frac{d\mathbf{r}}{dt}\right|^3, \tag{14}$$

so

$$\kappa = \left|\frac{d\hat{\mathbf{T}}}{ds}\right| = \frac{\left|\frac{d\mathbf{r}}{dt}\times\frac{d^2\mathbf{r}}{dt^2}\right|}{\left|\frac{d\mathbf{r}}{dt}\right|^3}. \tag{15}$$

The curvature of a 2-D curve is related to the Radius of Curvature of the curve's Osculating Circle. Consider a Circle specified parametrically by

$$x = a\cos t \tag{16}$$

$$y = a\sin t \tag{17}$$

which is tangent to the curve at a given point. The curvature is then

$$\kappa = \frac{x'y'' - y'x''}{(x'^2 + y'^2)^{3/2}} = \frac{a^2}{a^3} = \frac{1}{a}, \tag{18}$$

or one over the Radius of Curvature. The curvature of a Circle can also be repeated in vector notation. For the Circle with $0 \leq t < 2\pi$, the Arc Length is

$$\begin{aligned} s(t) &= \int_0^t \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\,dt \\ &= \int_0^t \sqrt{a^2\cos^2 t + a^2\sin^2 t}\,dt = at, \end{aligned} \tag{19}$$

so $t = s/a$ and the equations of the Circle can be rewritten as

$$x = a\cos\left(\frac{s}{a}\right) \tag{20}$$

$$y = a\sin\left(\frac{s}{a}\right). \tag{21}$$

The Position Vector is then given by

$$\mathbf{r}(s) = a\cos\left(\frac{s}{a}\right)\hat{\mathbf{x}} + a\sin\left(\frac{s}{a}\right)\hat{\mathbf{y}}, \tag{22}$$

and the Tangent Vector is

$$\hat{\mathbf{T}} = \frac{d\mathbf{r}}{ds} = -\sin\left(\frac{s}{a}\right)\hat{\mathbf{x}} + \cos\left(\frac{s}{a}\right)\hat{\mathbf{y}}, \tag{23}$$

so the curvature is related to the Radius of Curvature a by

$$\begin{aligned} \kappa &= \left|\frac{d\hat{\mathbf{T}}}{ds}\right| = \left|-\frac{1}{a}\cos\left(\frac{s}{a}\right)\hat{\mathbf{x}} - \frac{1}{a}\sin\left(\frac{s}{a}\right)\hat{\mathbf{y}}\right| \\ &= \sqrt{\frac{\cos^2\left(\frac{s}{a}\right) + \sin^2\left(\frac{s}{a}\right)}{a^2}} = \frac{1}{a}, \end{aligned} \tag{24}$$

as expected.

Four very important derivative relations in differential geometry related to the Frenet Formulas are

$$\dot{\mathbf{r}} = \mathbf{T} \tag{25}$$

$$\ddot{\mathbf{r}} = \kappa\mathbf{N} \tag{26}$$

$$\dddot{\mathbf{r}} = \dot{\kappa}\mathbf{N} + \kappa(\tau\mathbf{B} - \kappa\mathbf{T}) \tag{27}$$

$$[\dot{\mathbf{r}}, \ddot{\mathbf{r}}, \dddot{\mathbf{r}}] = \kappa^2\tau, \tag{28}$$

where $\mathbf{T}$ is the Tangent Vector, $\mathbf{N}$ is the Normal Vector, $\mathbf{B}$ is the Binormal Vector, and τ is the Torsion (Coxeter 1969, p. 322).

The curvature at a point on a surface takes on a variety of values as the Plane through the normal varies. As κ varies, it achieves a minimum and a maximum (which are in perpendicular directions) known as the Principal Curvatures. As shown in Coxeter (1969, pp. 352–353),

$$\kappa^2 - \sum b_i^i\kappa + \det(b_i^j) = 0 \tag{29}$$

$$\kappa^2 - 2H\kappa + K = 0, \tag{30}$$

where K is the Gaussian Curvature, H is the Mean Curvature, and det denotes the Determinant.

The curvature κ is sometimes called the First Curvature and the Torsion τ the Second Curvature. In addition, a Third Curvature (sometimes called Total Curvature)

$$\sqrt{ds_T{}^2 + ds_B{}^2} \tag{31}$$

is also defined. A signed version of the curvature of a Circle appearing in the Descartes Circle Theorem for the radius of the fourth of four mutually tangent circles is called the Bend.

see also Bend (Curvature), Curvature Center, Curvature Scalar, Extrinsic Curvature, First Curvature, Four-Vertex Theorem, Gaussian Curvature, Intrinsic Curvature, Lancret Equation, Line of Curvature, Mean Curvature, Normal Curvature, Principal Curvatures, Radius of Curvature, Ricci Curvature, Riemann Tensor, Second Curvature, Sectional Curvature, Soddy Circles, Third Curvature, Torsion (Differential Geometry), Weingarten Map

References

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, 1969.

Fischer, G. (Ed.). Plates 79–85 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 74–81, 1986.

Gray, A. "Curvature of Curves in the Plane," "Drawing Plane Curves with Assigned Curvature," and "Drawing Space Curves with Assigned Curvature." §1.5, 6.4, and 7.8 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 11–13, 68–69, 113–118, and 145–147, 1993.

Kreyszig, E. "Principal Normal, Curvature, Osculating Circle." §12 in *Differential Geometry.* New York: Dover, pp. 34–36, 1991.

Yates, R. C. "Curvature." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 60–64, 1952.

Curvature Center

The point on the POSITIVE RAY of the NORMAL VECTOR at a distance $\rho(s)$, where ρ is the RADIUS OF CURVATURE. It is given by

$$\mathbf{z} = \mathbf{x} + \rho\mathbf{N} = \mathbf{x} + \rho^2 \frac{\mathbf{T}}{ds}, \tag{1}$$

where $\mathbf{N}$ is the NORMAL VECTOR and $\mathbf{T}$ is the TANGENT VECTOR. It can be written in terms of $\mathbf{x}$ explicitly as

$$\mathbf{z} = \mathbf{x} + \frac{\mathbf{x}''(\mathbf{x}' \cdot \mathbf{x}')^2 - \mathbf{x}'(\mathbf{x}' \cdot \mathbf{x}')(\mathbf{x}' \cdot \mathbf{x}'')}{(\mathbf{x}' \cdot \mathbf{x}')(\mathbf{x}'' \cdot \mathbf{x}'') - (\mathbf{x}' \cdot \mathbf{x}'')^2}. \tag{2}$$

For a CURVE represented parametrically by $(f(t), g(t))$,

$$\alpha = f - \frac{(f'^2 - g'^2)g'}{f'g'' - f''g'} \tag{3}$$

$$\beta = g + \frac{(f'^2 - g'^2)f'}{f'g'' - f''g'}. \tag{4}$$

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, 1993.

Curvature Scalar

The curvature scalar is given by

$$R \equiv g^{\mu\kappa} R_{\mu\kappa},$$

where $g^{\mu\kappa}$ is the METRIC TENSOR and $R_{\mu\kappa}$ is the RICCI TENSOR.

see also CURVATURE, GAUSSIAN CURVATURE, MEAN CURVATURE, METRIC TENSOR, RADIUS OF CURVATURE, RICCI TENSOR, RIEMANN-CHRISTOFFEL TENSOR

Curvature Vector

$$\mathbf{K} \equiv \frac{d\mathbf{T}}{ds},$$

where $\mathbf{T}$ is the TANGENT VECTOR defined by

$$\mathbf{T} \equiv \frac{\frac{d\mathbf{x}}{ds}}{\left|\frac{d\mathbf{x}}{ds}\right|}.$$

Curve

A CONTINUOUS MAP from a 1-D SPACE to an n-D SPACE. Loosely speaking, the word "curve" is often used to mean the GRAPH of a 2- or 3-D curve. The simplest curves can be represented parametrically in n-D SPACE as

$$\begin{aligned} x_1 &= f_1(t) \\ x_2 &= f_2(t) \\ &\vdots \\ x_n &= f_n(t). \end{aligned}$$

Other simple curves can be simply defined only implicitly, i.e., in the form

$$f(x_1, x_2, \ldots) = 0.$$

see also ARCHIMEDEAN SPIRAL, ASTROID, ASYMPTOTIC CURVE, BASEBALL COVER, BATRACHION, BICORN, BIFOLIUM, BOW, BULLET NOSE, BUTTERFLY CURVE, CARDIOID, CASSINI OVALS, CATALAN'S TRISECTRIX, CATENARY, CAUSTIC, CAYLEY'S SEXTIC, CESÀRO EQUATION, CIRCLE, CIRCLE INVOLUTE, CISSOID, CISSOID OF DIOCLES, COCHLEOID, CONCHOID, CONCHOID OF NICOMEDES, CROSS CURVE, CRUCIFORM, CUBICAL PARABOLA, CURVE OF CONSTANT PRECESSION, CURVE OF CONSTANT WIDTH, CURTATE CYCLOID, CYCLOID, DELTA CURVE, DELTOID, DEVIL'S CURVE, DEVIL ON TWO STICKS, DUMBBELL CURVE, DÜRER'S CONCHOID, EIGHT CURVE, ELECTRIC MOTOR CURVE, ELLIPSE, ELLIPSE INVOLUTE, ELLIPTIC CURVE, ENVELOPE, EPICYCLOID, EQUIPOTENTIAL CURVE, EUDOXUS'S KAMPYLE, EVOLUTE, EXPONENTIAL RAMP, FERMAT CONIC, FOLIUM OF DESCARTES, FREETH'S NEPHROID, FREY CURVE, GAUSSIAN FUNCTION, GERONO LEMNISCATE, GLISSETTE, GUDERMANNIAN FUNCTION, GUTSCHOVEN'S CURVE, HIPPOPEDE, HORSE FETTER, HYPERBOLA, HYPERELLIPSE, HYPOCYCLOID, HYPOELLIPSE, INVOLUTE, ISOPTIC CURVE, KAPPA CURVE, KERATOID CUSP, KNOT CURVE, LAMÉ CURVE, LEMNISCATE, L'HOSPITAL'S CUBIC, LIMAÇON, LINKS CURVE, LISSAJOUS CURVE, LITUUS, LOGARITHMIC SPIRAL, MACLAURIN TRISECTRIX, MALTESE CROSS, MILL, NATURAL EQUATION, NEGATIVE PEDAL CURVE, NEPHROID, NIELSEN'S SPIRAL, ORTHOPTIC CURVE, PARABOLA, PEAR CURVE, PEAR-SHAPED CURVE, PEARLS OF SLUZE, PEDAL CURVE, PEG TOP, PIRIFORM, PLATEAU CURVES, POLICEMAN ON POINT DUTY CURVE, PROLATE CYCLOID, PURSUIT CURVE, QUADRATRIX OF HIPPIAS, RADIAL CURVE, RHODONEA, ROSE, ROULETTE, SEMICUBICAL PARABOLA, SERPENTINE CURVE, SICI SPIRAL, SIGMOID CURVE, SINUSOIDAL SPIRAL, SPACE CURVE, STROPHOID, SUPERELLIPSE, SWASTIKA, SWEEP SIGNAL, TALBOT'S CURVE, TEARDROP CURVE, TRACTRIX, TRIDENT, TRIDENT OF DESCARTES, TRIDENT OF NEWTON, TROCHOID, TSCHIRNHAUSEN CUBIC, VERSIERA, WATT'S CURVE, WHEWELL EQUATION, WITCH OF AGNESI

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 71–75, 1989.

"Geometry." *The New Encyclopædia Britannica, 15th ed.* **19**, pp. 946–951, 1990.

Gray, A. "Famous Plane Curves." Ch. 3 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 37–55, 1993.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, 1972.

Lee, X. "A Catalog of Special Plane Curves." `http://www.best.com/~xah/SpecialPlaneCurves_dir/specialPlaneCurves.html`.

Lockwood, E. H. *A Book of Curves.* Cambridge, England: Cambridge University Press, 1961.

MacTutor History of Mathematics Archive. `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Curves.html`.

Oakley, C. O. *Analytic Geometry.* New York: Barnes and Noble, 1957.

Shikin, E. V. *Handbook and Atlas of Curves.* Boca Raton, FL: CRC Press, 1995.

Smith, P. F.; Gale, A. S.; and Neelley, J. H. *New Analytic Geometry, Alternate Edition.* Boston, MA: Ginn and Company, 1938.

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, 1993.

Walker, R. J. *Algebraic Curves.* New York: Springer-Verlag, 1978.

Weisstein, E. W. "Plane Curves." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Curves.m`.

Yates, R. C. *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, 1947.

Yates, R. C. *The Trisection Problem.* Reston, VA: National Council of Teachers of Mathematics, 1971.

Zwillinger, D. (Ed.). "Algebraic Curves." §8.1 in *CRC Standard Mathematical Tables and Formulae, 3rd ed.* Boca Raton, FL: CRC Press, 1996. `http://www.geom.umn.edu/docs/reference/CRC-formulas/node33.html`.

Curve of Constant Breadth

see CURVE OF CONSTANT WIDTH

Curve of Constant Precession

A curve whose CENTRODE revolves about a fixed axis with constant ANGLE and SPEED when the curve is traversed with unit SPEED. The TANGENT INDICATRIX of a curve of constant precession is a SPHERICAL HELIX. An ARC LENGTH parameterization of a curve of constant precession with NATURAL EQUATIONS

$$\kappa(s) = -\omega\sin(\mu s) \tag{1}$$

$$\tau(s) = \omega\cos(\mu s) \tag{2}$$

is

$$x(s) = \frac{\alpha+\mu}{2\alpha}\frac{\sin[(\alpha-\mu)s]}{\alpha-\mu} - \frac{\alpha-\mu}{2\alpha}\frac{\sin[(\alpha+\mu)s]}{\alpha+\mu} \tag{3}$$

$$y(s) = -\frac{\alpha+\mu}{2\alpha}\frac{\cos[(\alpha-\mu)s]}{\alpha-\mu} + \frac{\alpha-\mu}{2\alpha}\frac{\cos[(\alpha+\mu)s]}{\alpha+\mu} \tag{4}$$

$$z(s) = \frac{\omega}{\mu\alpha}\sin(\mu s), \tag{5}$$

where

$$\alpha \equiv \sqrt{\omega^2+\mu^2} \tag{6}$$

and ω, and μ are constant. This curve lies on a circular one-sheeted HYPERBOLOID

$$x^2 + y^2 - \frac{\mu^2}{\omega^2}z^2 = \frac{4\mu^2}{\omega^4}. \tag{7}$$

The curve is closed IFF μ/α is RATIONAL.

References

Scofield, P. D. "Curves of Constant Precession." *Amer. Math. Monthly* **102**, 531–537, 1995.

Curve of Constant Slope

see GENERALIZED HELIX

Curve of Constant Width

Curves which, when rotated in a square, make contact with all four sides. The "width" of a closed convex curve is defined as the distance between parallel lines bounding it ("supporting lines"). Every curve of constant width is convex. Curves of constant width have the same "width" regardless of their orientation between the parallel lines. In fact, they also share the same PERIMETER (BARBIER'S THEOREM). Examples include the CIRCLE (with largest AREA), and REULEAUX TRIANGLE (with smallest AREA) but there are an infinite number. A curve of constant width can be used in a special drill chuck to cut square "HOLES."

A generalization gives solids of constant width. These do not have the same surface AREA for a given width, but their shadows are curves of constant width with the *same* width!

see also DELTA CURVE, KAKEYA NEEDLE PROBLEM, REULEAUX TRIANGLE

References

Bogomolny, A. "Shapes of Constant Width." `http://www.cut-the-knot.com/do_you_know/cwidth.html`.

Böhm, J. "Convex Bodies of Constant Width." Ch. 4 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 96–100, 1986.

Fischer, G. (Ed.). Plates 98–102 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 89 and 96, 1986.

Gardner, M. Ch. 18 in *The Unexpected Hanging and Other Mathematical Diversions.* Chicago, IL: Chicago University Press, 1991.

Goldberg, M. "Circular-Arc Rotors in Regular Polygons." *Amer. Math. Monthly* **55**, 393–402, 1948.

Kelly, P. *Convex Figures.* New York: Harcourt Brace, 1995.

Rademacher, H. and Toeplitz, O. *The Enjoyment of Mathematics: Selections from Mathematics for the Amateur.* Princeton, NJ: Princeton University Press, 1957.

Yaglom, I. M. and Boltyanski, V. G. *Convex Figures.* New York: Holt, Rinehart, and Winston, 1961.

Curvilinear Coordinates

A general METRIC $g_{\mu\nu}$ has a LINE ELEMENT

$$ds^2 = g_{\mu\nu}du^\mu du^\nu, \tag{1}$$

where EINSTEIN SUMMATION is being used. Curvilinear coordinates are defined as those with a diagonal METRIC so that

$$g_{\mu\nu} \equiv \delta^\mu_\nu {h_\mu}^2, \tag{2}$$

where δ^μ_ν is the KRONECKER DELTA. Curvilinear coordinates therefore have a simple LINE ELEMENT

$$ds^2 = \delta^\mu_\nu {h_\mu}^2 du^\mu du^\nu = {h_\mu}^2 du^{\mu 2}, \tag{3}$$

which is just the PYTHAGOREAN THEOREM, so the differential VECTOR is

$$d\mathbf{r} = h_\mu du_\mu \hat{\mathbf{u}}_\mu, \tag{4}$$

or

$$d\mathbf{r} = \frac{\partial \mathbf{r}}{\partial u_1}\,du_1 + \frac{\partial \mathbf{r}}{\partial u_2}\,du_2 + \frac{\partial \mathbf{r}}{\partial u_3}\,du_3, \tag{5}$$

where the SCALE FACTORS are

$$h_i \equiv \left|\frac{\partial \mathbf{r}}{\partial u_i}\right| \tag{6}$$

and

$$\hat{\mathbf{u}}_i \equiv \frac{\frac{\partial \mathbf{r}}{\partial u_i}}{\left|\frac{\partial \mathbf{r}}{\partial u_i}\right|} = \frac{1}{h_i}\frac{\partial \mathbf{r}}{\partial u_i}. \tag{7}$$

Equation (5) may therefore be re-expressed as

$$d\mathbf{r} = h_1 du_1 \hat{\mathbf{u}}_1 + h_2 du_2 \hat{\mathbf{u}}_2 + h_3 du_3 \hat{\mathbf{u}}_3. \tag{8}$$

The GRADIENT is

$$\operatorname{grad}(\phi) \equiv \nabla\phi = \frac{1}{h_1}\frac{\partial \phi}{\partial u_1}\hat{\mathbf{u}}_1 + \frac{1}{h_2}\frac{\partial \phi}{\partial u_2}\hat{\mathbf{u}}_2 + \frac{1}{h_3}\frac{\partial \phi}{\partial u_3}\hat{\mathbf{u}}_3, \tag{9}$$

the DIVERGENCE is

$$\begin{aligned}\operatorname{div}(F) \equiv \nabla\cdot\mathbf{F} \equiv \frac{1}{h_1h_2h_3}&\left[\frac{\partial}{\partial u_1}(h_2h_3F_1)\right.\\ &\left.+\frac{\partial}{\partial u_2}(h_3h_1F_2) + \frac{\partial}{\partial u_3}(h_1h_2F_3)\right],\end{aligned} \tag{10}$$

and the CURL is

$$\begin{aligned}\nabla\times\mathbf{F} &\equiv \frac{1}{h_1h_2h_3}\begin{vmatrix} h_1\hat{\mathbf{u}}_1 & h_2\hat{\mathbf{u}}_2 & h_3\hat{\mathbf{u}}_3 \\ \frac{\partial}{\partial u_1} & \frac{\partial}{\partial u_2} & \frac{\partial}{\partial u_3} \\ h_1F_1 & h_2F_2 & h_3F_3 \end{vmatrix}\\ &= \frac{1}{h_2h_3}\left[\frac{\partial}{\partial u_2}(h_3F_3) - \frac{\partial}{\partial u_3}(h_2F_2)\right]\hat{\mathbf{u}}_1\\ &\quad+\frac{1}{h_1h_3}\left[\frac{\partial}{\partial u_3}(h_1F_1) - \frac{\partial}{\partial u_1}(h_3F_3)\right]\hat{\mathbf{u}}_2\\ &\quad+\frac{1}{h_1h_2}\left[\frac{\partial}{\partial u_1}(h_2F_2) - \frac{\partial}{\partial u_2}(h_1F_1)\right]\hat{\mathbf{u}}_3.\end{aligned} \tag{11}$$

Orthogonal curvilinear coordinates satisfy the additional constraint that

$$\hat{\mathbf{u}}_i\cdot\hat{\mathbf{u}}_j = \delta_{ij}. \tag{12}$$

Therefore, the LINE ELEMENT is

$$ds^2 = d\mathbf{r}\cdot d\mathbf{r} = {h_1}^2{du_1}^2 + {h_2}^2{du_2}^2 + {h_3}^2{du_3}^2 \tag{13}$$

and the VOLUME ELEMENT is

$$\begin{aligned} dV &= |(h_1\hat{\mathbf{u}}_1\,du_1)\cdot(h_2\hat{\mathbf{u}}_2\,du_2)\times(h_3\hat{\mathbf{u}}_3\,du_3)|\\ &= h_1h_2h_3\,du_1\,du_2\,du_3\\ &= \left|\frac{\partial r}{\partial u_1}\cdot\frac{\partial r}{\partial u_2}\times\frac{\partial r}{\partial u_3}\right|du_1\,du_2\,du_3\\ &= \begin{vmatrix} \frac{\partial x}{\partial u_1} & \frac{\partial x}{\partial u_2} & \frac{\partial x}{\partial u_3}\\ \frac{\partial y}{\partial u_1} & \frac{\partial y}{\partial u_2} & \frac{\partial y}{\partial u_3}\\ \frac{\partial z}{\partial u_1} & \frac{\partial z}{\partial u_2} & \frac{\partial z}{\partial u_3} \end{vmatrix} du_1\,du_2\,du_3\\ &= \left|\frac{\partial(x,y,z)}{\partial(u_1,u_2,u_3)}\right| du_1\,du_2\,du_3, \end{aligned} \tag{14}$$

where the latter is the JACOBIAN.

Orthogonal curvilinear coordinate systems include BIPOLAR CYLINDRICAL COORDINATES, BISPHERICAL COORDINATES, CARTESIAN COORDINATES, CONFOCAL ELLIPSOIDAL COORDINATES, CONFOCAL PARABOLOIDAL COORDINATES, CONICAL COORDINATES, CYCLIDIC COORDINATES, CYLINDRICAL COORDINATES, ELLIPSOIDAL COORDINATES, ELLIPTIC CYLINDRICAL COORDINATES, OBLATE SPHEROIDAL COORDINATES, PARABOLIC COORDINATES, PARABOLIC CYLINDRICAL COORDINATES, PARABOLOIDAL COORDINATES, POLAR COORDINATES, PROLATE SPHEROIDAL COORDINATES, SPHERICAL COORDINATES, and TOROIDAL COORDINATES. These are degenerate cases of the CONFOCAL ELLIPSOIDAL COORDINATES.

see also CHANGE OF VARIABLES THEOREM, CURL, DIVERGENCE, GRADIENT, JACOBIAN, LAPLACIAN

References

Arfken, G. "Curvilinear Coordinates" and "Differential Vector Operators." §2.1 and 2.2 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 86–90 and 90–94, 1985.

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 1084–1088, 1980.

Morse, P. M. and Feshbach, H. "Curvilinear Coordinates" and "Table of Properties of Curvilinear Coordinates." §1.3 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 21–31 and 115–117, 1953.

Cushion

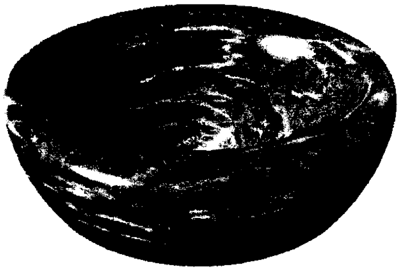

The QUARTIC SURFACE resembling a squashed round cushion on a barroom stool and given by the equation

$$\begin{aligned} &z^2x^2 - z^4 - 2zx^2 + 2z^3 + x^2 - z^2\\ &\quad -(x^2-z)^2 - y^4 - 2x^2y^2 - y^2z^2 + 2y^2z + y^2 = 0. \end{aligned}$$

see also QUARTIC SURFACE

References

Nordstrand, T. "Surfaces." `http://www.uib.no/people/nfytn/surfaces.htm`.

Cusp

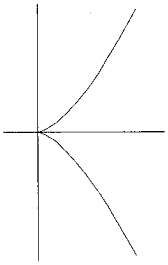

A function $f(x)$ has a cusp (also called a SPINODE) at a point x_0 if $f(x)$ is CONTINUOUS at x_0 and

$$\lim_{x\to x_0} f'(x) = \infty$$

from one side while

$$\lim_{x\to x_0} f'(x) = -\infty$$

from the other side, so the curve is CONTINUOUS but the DERIVATIVE is not. A cusp is a type of DOUBLE POINT. The above plot shows the curve $x^3 - y^2 = 0$, which has a cusp at the ORIGIN.

see also DOUBLE CUSP, DOUBLE POINT, ORDINARY DOUBLE POINT, RAMPHOID CUSP, SALIENT POINT

References

Walker, R. J. *Algebraic Curves.* New York: Springer-Verlag, pp. 57–58, 1978.

Cusp Catastrophe

A CATASTROPHE which can occur for two control factors and one behavior axis. The equation $y = x^{2/3}$ has a cusp catastrophe.

see also CATASTROPHE

References

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 28, 1993.

Cusp Form

A cusp form on $\Gamma_0(N)$, the group of INTEGER matrices with determinant 1 which are upper triangular mod N, is an ANALYTIC FUNCTION on the upper half-plane consisting of the COMPLEX NUMBERS with POSITIVE IMAGINARY PART. Weight n cusp forms satisfy

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^n f(z)$$

for all matrices

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \Gamma_0(N).$$

see also MODULAR FORM

Cusp Map

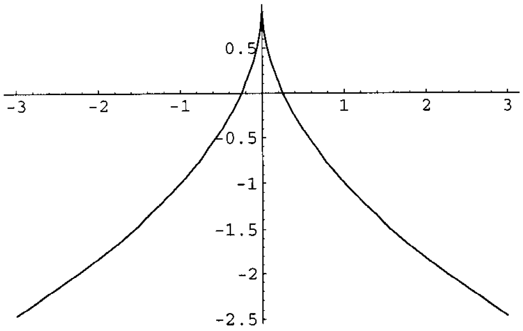

The function

$$f(x) = 1 - 2|x|^{1/2}$$

for $x \in [-1, 1]$. The INVARIANT DENSITY is

$$\rho(y) = \tfrac{1}{2}(1 - y).$$

References

Beck, C. and Schlögl, F. *Thermodynamics of Chaotic Systems.* Cambridge, England: Cambridge University Press, p. 195, 1995.

Cusp Point

see CUSP

Cut-Vertex

see ARTICULATION VERTEX

Cutting

see ARRANGEMENT, CAKE CUTTING, CIRCLE CUTTING, CYLINDER CUTTING, PANCAKE CUTTING, PIE CUTTING, SQUARE CUTTING, TORUS CUTTING

CW-Approximation Theorem

If X is any SPACE, then there is a CW-COMPLEX Y and a MAP $f : Y \to X$ inducing ISOMORPHISMS on all HOMOTOPY, HOMOLOGY, and COHOMOLOGY groups.

CW-Complex

A CW-complex is a homotopy-theoretic generalization of the notion of a SIMPLICIAL COMPLEX. A CW-complex is any SPACE X which can be built by starting off with a discrete collection of points called X^0, then attaching 1-D DISKS D^1 to X^0 along their boundaries S^0, writing X^1 for the object obtained by attaching the D^1s to X^0, then attaching 2-D DISKS D^2 to X^1 along their boundaries S^1, writing X^2 for the new SPACE, and so on, giving spaces X^n for every n. A CW-complex is any SPACE that has this sort of decomposition into SUBSPACES X^n built up in such a hierarchical fashion (so the X^ns must exhaust all of X). In particular, X^n may be built from X^{n-1} by attaching infinitely many n-DISKS, and the attaching MAPS $S^{n-1} \to X^{n-1}$ may be any continuous MAPS.

The main importance of CW-complexes is that, for the sake of HOMOTOPY, HOMOLOGY, and COHOMOLOGY groups, every SPACE is a CW-complex. This is called the CW-APPROXIMATION THEOREM. Another is WHITEHEAD'S THEOREM, which says that MAPS between CW-complexes that induce ISOMORPHISMS on all HOMOTOPY GROUPS are actually HOMOTOPY equivalences.

see also COHOMOLOGY, CW-APPROXIMATION THEOREM, HOMOLOGY GROUP, HOMOTOPY GROUP, SIMPLICIAL COMPLEX, SPACE, SUBSPACE, WHITEHEAD'S THEOREM

Cycle (Circle)

A CIRCLE with an arrow indicating a direction.

Cycle (Graph)

A subset of the EDGE-set of a graph that forms a CHAIN (GRAPH), the first node of which is also the last (also called a CIRCUIT).

see also CYCLIC GRAPH, HAMILTONIAN CYCLE, WALK

Cycle Graph

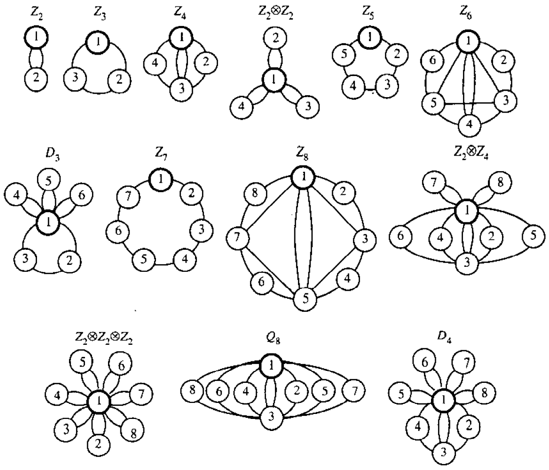

A cycle graph is a GRAPH which shows cycles of a GROUP as well as the connectivity between the cycles. Several examples are shown above. For Z_4, the group elements A_i satisfy ${A_i}^4 = 1$, where 1 is the IDENTITY ELEMENT, and two elements satisfy ${A_1}^2 = {A_3}^2 = 1$.

For a CYCLIC GROUP of COMPOSITE ORDER n (e.g., Z_4, Z_6, Z_8), the degenerate subcycles corresponding to factors dividing n are often not shown explicitly since their presence is implied.

see also CHARACTERISTIC FACTOR, CYCLIC GROUP

References

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 83–98, 1993.

Cycle (Map)

An n-cycle is a finite sequence of points $Y_0, \ldots, Y_{n-1}$ such that, under a MAP G,

$$\begin{aligned} Y_1 &= G(Y_0) \\ Y_2 &= G(Y_1) \\ Y_{n-1} &= G(Y_{n-2}) \\ Y_0 &= G(Y_{n-1}). \end{aligned}$$

In other words, it is a periodic trajectory which comes back to the same point after n iterations of the cycle. Every point Y_j of the cycle satisfies $Y_j = G^n(Y_j)$ and is therefore a FIXED POINT of the mapping G^n. A fixed point of G is simply a CYCLE of period 1.

Cycle (Permutation)

A SUBSET of a PERMUTATION whose elements trade places with one another. A cycle decomposition of a PERMUTATION can therefore be viewed as a CLASS of a PERMUTATION GROUP. For example, in the PERMUTATION GROUP $\{4, 2, 1, 3\}$, $\{1, 3, 4\}$ is a 3-cycle ($1 \to 3$, $3 \to 4$, and $4 \to 1$) and $\{2\}$ is a 1-cycle ($2 \to 2$). Every PERMUTATION GROUP on n symbols can be uniquely expressed as a product of disjoint cycles. The cyclic decomposition of a PERMUTATION can be computed in *Mathematica*® (Wolfram Research, Champaign, IL) with the function `ToCycles` and the PERMUTATION corresponding to a cyclic decomposition can be computed with `FromCycles`. According to Vardi (1991), the Mathematica code for `ToCycles` is one of the most obscure ever written.

To find the number $N(m, n)$ of m cycles in a PERMUTATION GROUP of order n, take

$$N(n, m) = (-1)^{n-m} S_1(n, m),$$

where S_1 is the STIRLING NUMBER OF THE FIRST KIND.

see also GOLOMB-DICKMAN CONSTANT, PERMUTATION, PERMUTATION GROUP, SUBSET

References

Skiena, S. *Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica.* Reading, MA: Addison-Wesley, p. 20, 1990.

Vardi, I. *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, p. 223, 1991.

Cyclic Graph

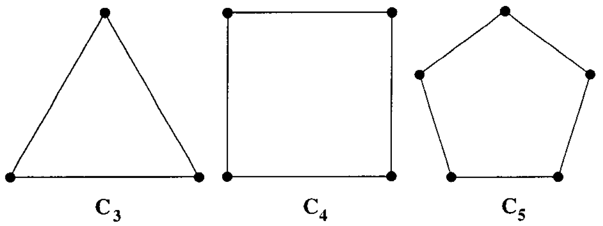

A GRAPH of n nodes and n edges such that node i is connected to the two adjacent nodes $i+1$ and $i-1$ (mod n), where the nodes are numbered $0, 1, \ldots, n-1$.

see also CYCLE (GRAPH), CYCLE GRAPH, STAR GRAPH, WHEEL GRAPH

Cyclic Group

A cyclic group Z_n of ORDER n is a GROUP defined by the element X (the GENERATOR) and its n POWERS up to

$$X^n = I,$$

where I is the IDENTITY ELEMENT. Cyclic groups are both ABELIAN and SIMPLE. There exists a unique cyclic group of every order $n \geq 2$, so cyclic groups of the same order are always isomorphic (Shanks 1993, p. 74), and all GROUPS of PRIME ORDER are cyclic.

Examples of cyclic groups include Z_2, Z_3, Z_4, and the MODULO MULTIPLICATION GROUPS M_m such that $m = 2$, 4, p^n, or $2p^n$, for p an ODD PRIME and $n \geq 1$ (Shanks 1993, p. 92). By computing the CHARACTERISTIC FACTORS, any ABELIAN GROUP can be expressed as a DIRECT PRODUCT of cyclic SUBGROUPS, for example, $Z_2 \otimes Z_4$ or $Z_2 \otimes Z_2 \otimes Z_2$.

see also ABELIAN GROUP, CHARACTERISTIC FACTOR, FINITE GROUP—Z_2, FINITE GROUP—Z_3, FINITE GROUP—Z_2, FINITE GROUP—Z_5, FINITE GROUP—Z_6, MODULO MULTIPLICATION GROUP, SIMPLE GROUP

References

Lomont, J. S. "Cyclic Groups." §3.10.A in *Applications of Finite Groups.* New York: Dover, p. 78, 1987.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, 1993.

Cyclic Hexagon

A hexagon (not necessarily regular) on whose VERTICES a CIRCLE may be CIRCUMSCRIBED. Let

$$\sigma_i \equiv \sum_{i,j,\ldots,n=1} {a_i}^2 {a_j}^2 \cdots {a_n}^2, \tag{1}$$

where the sum runs over all distinct permutations of the SQUARES of the six side lengths, so

$$\begin{aligned}
\sigma_1 &= {a_1}^2 + {a_2}^2 + {a_3}^2 + {a_4}^2 + {a_5}^2 + {a_6}^2 && (2)\\
\sigma_2 &= {a_1}^2{a_2}^2 + {a_1}^2{a_3}^2 + {a_1}^2{a_4}^2 + {a_1}^2{a_5}^2 + {a_1}^2{a_6}^2\\
&\quad + {a_2}^2{a_3}^2 + {a_2}^2{a_4}^2 + {a_2}^2{a_5}^2 + {a_2}^2{a_6}^2\\
&\quad + {a_3}^2{a_4}^2 + {a_3}^2{a_5}^2 + {a_3}^2{a_6}^2\\
&\quad + {a_4}^2{a_5}^2 + {a_4}^2{a_6}^2 + {a_5}^2{a_6}^2 && (3)\\
\sigma_3 &= {a_1}^2{a_2}^2{a_3}^2 + {a_1}^2{a_2}^2{a_4}^2 + {a_1}^2{a_2}^2{a_5}^2 + {a_1}^2{a_2}^2{a_6}^2\\
&\quad + {a_2}^2{a_3}^2{a_4}^2 + {a_2}^2{a_3}^2{a_5}^2 + {a_2}^2{a_3}^2{a_6}^2\\
&\quad + {a_3}^2{a_4}^2{a_5}^2 + {a_3}^2{a_4}^2{a_6}^2 + {a_4}^2{a_5}^2{a_6}^2 && (4)\\
\sigma_4 &= {a_1}^2{a_2}^2{a_3}^2{a_4}^2 + {a_1}^2{a_2}^2{a_3}^2{a_5}^2 + {a_1}^2{a_2}^2{a_3}^2{a_6}^2\\
&\quad + {a_1}^2{a_3}^2{a_4}^2{a_5}^2 + {a_1}^2{a_3}^2{a_4}^2{a_6}^2\\
&\quad + {a_1}^2{a_3}^2{a_5}^2{a_6}^2 + {a_1}^2{a_4}^2{a_5}^2{a_6}^2\\
&\quad + {a_2}^2{a_3}^2{a_4}^2{a_5}^2 + {a_2}^2{a_3}^2{a_4}^2{a_6}^2 + {a_2}^2{a_3}^2{a_5}^2{a_6}^2\\
&\quad + {a_2}^2{a_4}^2{a_5}^2{a_6}^2 + {a_3}^2{a_4}^2{a_5}^2{a_6}^2 && (5)\\
\sigma_5 &= {a_1}^2{a_2}^2{a_3}^2{a_4}^2{a_5}^2 + {a_1}^2{a_2}^2{a_3}^2{a_4}^2{a_6}^2\\
&\quad + {a_1}^2{a_2}^2{a_3}^2{a_5}^2{a_6}^2 + {a_1}^2{a_2}^2{a_4}^2{a_5}^2{a_6}^2\\
&\quad + {a_1}^2{a_3}^2{a_4}^2{a_5}^2{a_6}^2 + {a_2}^2{a_3}^2{a_4}^2{a_5}^2{a_6}^2 && (6)\\
\sigma_6 &= {a_1}^2{a_2}^2{a_3}^2{a_4}^2{a_5}^2{a_6}^2. && (7)
\end{aligned}$$

Then define

$$\begin{aligned}
t_2 &= u - 4\sigma_2 + {\sigma_1}^2 && (8)\\
t_3 &= 8\sigma_3 + \sigma_1 t_2 - 16\sqrt{\sigma_6} && (9)\\
t_4 &= {t_2}^2 - 64\sigma_4 + 64\sigma_1\sqrt{\sigma_6} && (10)\\
t_5 &= 128\sigma_5 + 32t_2\sqrt{\sigma_6} && (11)\\
u &= 16K^2. && (12)
\end{aligned}$$

The AREA of the hexagon then satisfies

$$u{t_4}^3 + {t_3}^2{t_4}^2 - 16{t_3}^3 t_5 - 18ut_3t_4t_5 - 27u^2{t_5}^2 = 0, \tag{13}$$

or this equation with $\sqrt{\sigma_6}$ replaced by $-\sqrt{\sigma_6}$, a seventh order POLYNOMIAL in u. This is $1/(4u^2)$ times the DISCRIMINANT of the CUBIC EQUATION

$$z^3 + 2t_3z^2 - ut_4z + 2y^2t_5. \tag{14}$$

see also CONCYCLIC, CYCLIC PENTAGON, CYCLIC POLYGON, FUHRMANN'S THEOREM

References

Robbins, D. P. "Areas of Polygons Inscribed in a Circle." *Discr. Comput. Geom.* **12**, 223–236, 1994.

Robbins, D. P. "Areas of Polygons Inscribed in a Circle." *Amer. Math. Monthly* **102**, 523–530, 1995.

Cyclic-Inscriptable Quadrilateral

see BICENTRIC QUADRILATERAL

Cyclic Number

A number having $n-1$ DIGITS which, when MULTIPLIED by 1, 2, 3, ..., $n-1$, produces the same digits in a different order. Cyclic numbers are generated by the UNIT FRACTIONS $1/n$ which have maximal period DECIMAL EXPANSIONS (which means n must be PRIME). The first few numbers which generate cyclic numbers are 7, 17, 19, 23, 29, 47, 59, 61, 97, ... (Sloane's A001913). A much larger generator is 17389.

It has been conjectured, but not yet proven, that an INFINITE number of cyclic numbers exist. In fact, the FRACTION of PRIMES which generate cyclic numbers seems to be approximately 3/8. See Yates (1973) for a table of PRIME period lengths for PRIMES $< 1{,}370{,}471$. When a cyclic number is multiplied by its generator, the result is a string of 9s. This is a special case of MIDY'S THEOREM.

$$\begin{aligned}
07 &= 0.142857\\
17 &= 0.0588235294117647\\
19 &= 0.052631578947368421\\
23 &= 0.0434782608695652173913\\
29 &= 0.0344827586206896551724137931\\
47 &= 0.0212765957446808510638297872340425531914\cdots\\
&\quad\cdots 4893617\\
59 &= 0.0169491525423728813559322033898305084745\cdots\\
&\quad\cdots 76271186440677966\,1\\
61 &= 0.0163934426229508196721311475409836065573\cdots
\end{aligned}$$

$$\cdots 377049180327868852459$$

$$97 = 0.01030927835051546391752577319587628865 9\cdots$$

$$\cdots 79381443298969072164948453608247422680412\cdots$$

$$\cdots 3711340206185567$$

see also DECIMAL EXPANSION, MIDY'S THEOREM

References

Gardner, M. Ch. 10 in *Mathematical Circus: More Puzzles, Games, Paradoxes and Other Mathematical Entertainments from Scientific American.* New York: Knopf, 1979.

Guttman, S. "On Cyclic Numbers." *Amer. Math. Monthly* **44**, 159–166, 1934.

Kraitchik, M. "Cyclic Numbers." §3.7 in *Mathematical Recreations.* New York: W. W. Norton, pp. 75–76, 1942.

Rao, K. S. "A Note on the Recurring Period of the Reciprocal of an Odd Number." *Amer. Math. Monthly* **62**, 484–487, 1955.

Sloane, N. J. A. Sequence A001913/M4353 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Yates, S. *Primes with Given Period Length.* Trondheim, Norway: Universitetsforlaget, 1973.

Cyclic Pentagon

A cyclic pentagon is a not necessarily regular PENTAGON on whose VERTICES a CIRCLE may be CIRCUMSCRIBED. Let

$$\sigma_i \equiv \sum_{i,j,\ldots,n=1} {a_i}^2 {a_j}^2 \cdots {a_n}^2, \tag{1}$$

where the SUM runs over all distinct PERMUTATIONS of the SQUARES of the 5 side lengths, so

$$\sigma_1 = {a_1}^2 + {a_2}^2 + {a_3}^2 + {a_4}^2 + {a_5}^2 \tag{2}$$

$$\begin{aligned}\sigma_2 = {} & {a_1}^2{a_2}^2 + {a_1}^2{a_3}^2 + {a_1}^2{a_4}^2 + {a_1}^2{a_5}^2 + {a_2}^2{a_3}^2 \\ & + {a_2}^2{a_4}^2 + {a_2}^2{a_5}^2 + {a_3}^2{a_4}^2 + {a_3}^2{a_5}^2 \\ & + {a_4}^2{a_5}^2\end{aligned} \tag{3}$$

$$\begin{aligned}\sigma_3 = {} & {a_1}^2{a_2}^2{a_3}^2 + {a_1}^2{a_2}^2{a_4}^2 + {a_1}^2{a_2}^2{a_5}^2 \\ & + {a_2}^2{a_3}^2{a_4}^2 + {a_2}^2{a_3}^2{a_5}^2 + {a_3}^2{a_4}^2{a_5}^2\end{aligned} \tag{4}$$

$$\begin{aligned}\sigma_4 = {} & {a_1}^2{a_2}^2{a_3}^2{a_4}^2 + {a_1}^2{a_2}^2{a_3}^2{a_5}^2 + {a_1}^2{a_3}^2{a_4}^2{a_5}^2 \\ & + {a_2}^2{a_3}^2{a_4}^2{a_5}^2\end{aligned} \tag{5}$$

$$\sigma_5 = {a_1}^2{a_2}^2{a_3}^2{a_4}^2{a_5}^2. \tag{6}$$

Then define

$$t_2 = u - 4\sigma_2 + {\sigma_1}^2 \tag{7}$$

$$t_3 = 8\sigma_3 + \sigma_1 t_2 \tag{8}$$

$$t_4 = -64\sigma_4 + {t_2}^2 \tag{9}$$

$$t_5 = 128\sigma_5 \tag{10}$$

$$u = 16K^2. \tag{11}$$

The AREA of the pentagon then satisfies

$$u{t_4}^3 + {t_3}^2{t_4}^2 - 16{t_3}^3 t_5 - 18ut_3t_4t_5 - 27u^2{t_5}^2 = 0, \tag{12}$$

a seventh order POLYNOMIAL in u. This is $1/(4u^2)$ times the DISCRIMINANT of the CUBIC EQUATION

$$z^3 + 2t_3z^2 - ut_4z + 2y^2t_5. \tag{13}$$

see also CONCYCLIC, CYCLIC HEXAGON, CYCLIC POLYGON

References

Robbins, D. P. "Areas of Polygons Inscribed in a Circle." *Discr. Comput. Geom.* **12**, 223–236, 1994.

Robbins, D. P. "Areas of Polygons Inscribed in a Circle." *Amer. Math. Monthly* **102**, 523–530, 1995.

Cyclic Permutation

A PERMUTATION which shifts all elements of a SET by a fixed offset, with the elements shifted off the end inserted back at the beginning. For a SET with elements a_0, a_1, $\ldots$, a_{n-1}, this can be written $a_i \to a_{i+k \pmod n}$ for a shift of k.

see also PERMUTATION

Cyclic Polygon

A cyclic polygon is a POLYGON with VERTICES upon which a CIRCLE can be CIRCUMSCRIBED. Since every TRIANGLE has a CIRCUMCIRCLE, every TRIANGLE is cyclic. It is conjectured that for a cyclic polygon of $2m+1$ sides, $16K^2$ (where K is the AREA) satisfies a MONIC POLYNOMIAL of degree Δ_m, where

$$\Delta_m \equiv \sum_{k=0}^{m-1} (m-k)\binom{2m+1}{k} \tag{1}$$

$$= \frac{1}{2}\left[(2m+1)\binom{2m}{m} - 2^{2m}\right] \tag{2}$$

(Robbins 1995). It is also conjectured that a cyclic polygon with $2m+2$ sides satisfies one of two POLYNOMIALS of degree Δ_m. The first few values of Δ_m are 1, 7, 38, 187, 874, ... (Sloane's A000531).

For TRIANGLES ($n = 3 = 2\cdot 1 + 1$), the POLYNOMIAL is HERON'S FORMULA, which may be written

$$16K^2 = 2a^2b^2 + 2a^2c^2 + 2b^2c^2 - a^4 - b^4 - c^4, \tag{3}$$

and which is of order $\Delta_1 = 1$ in $16K^2$. For a CYCLIC QUADRILATERAL, the POLYNOMIAL is BRAHMAGUPTA'S FORMULA, which may be written

$$\begin{aligned}16K^2 = {} & -a^4 + 2a^2b^2 - b^4 + 2a^2c^2 + 2b^2c^2 - c^4 \\ & + 8abcd + 2a^2d^2 + 2b^2d^2 + 2c^2d^2 - d^4,\end{aligned} \tag{4}$$

which is of order $\Delta_1 = 1$ in $16K^2$. Robbins (1995) gives the corresponding FORMULAS for the CYCLIC PENTAGON and CYCLIC HEXAGON.

see also CONCYCLIC, CYCLIC HEXAGON, CYCLIC PENTAGON, CYCLIC QUADRANGLE, CYCLIC QUADRILATERAL

References

Robbins, D. P. "Areas of Polygons Inscribed in a Circle." *Discr. Comput. Geom.* **12**, 223–236, 1994.

Robbins, D. P. "Areas of Polygons Inscribed in a Circle." *Amer. Math. Monthly* **102**, 523–530, 1995.

Sloane, N. J. A. Sequence A000531 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Cyclic Quadrangle

Let A_1, A_2, A_3, and A_4 be four POINTS on a CIRCLE, and H_1, H_2, H_3, H_4 the ORTHOCENTERS of TRIANGLES $\Delta A_2A_3A_4$, etc. If, from the eight POINTS, four with different subscripts are chosen such that three are from one set and the fourth from the other, these POINTS form an ORTHOCENTRIC SYSTEM. There are eight such systems, which are analogous to the six sets of ORTHOCENTRIC SYSTEMS obtained using the feet of the ANGLE BISECTORS, ORTHOCENTER, and VERTICES of a generic TRIANGLE.

On the other hand, if all the POINTS are chosen from one set, or two from each set, with all different subscripts, the four POINTS lie on a CIRCLE. There are four pairs of such CIRCLES, and eight POINTS lie by fours on eight equal CIRCLES.

The SIMSON LINE of A_4 with regard to TRIANGLE $\Delta A_1A_2A_3$ is the same as that of H_4 with regard to the TRIANGLE $\Delta H_1A_2A_3$.

see also ANGLE BISECTOR, CONCYCLIC, CYCLIC POLYGON, CYCLIC QUADRILATERAL, ORTHOCENTRIC SYSTEM

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 251–253, 1929.

Cyclic Quadrilateral

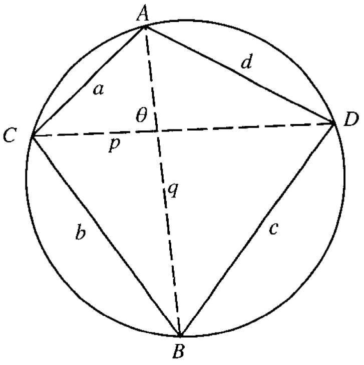

A QUADRILATERAL for which a CIRCLE can be circumscribed so that it touches each VERTEX. The AREA is then given by a special case of BRETSCHNEIDER'S FORMULA. Let the sides have lengths a, b, c, and d, let s be the SEMIPERIMETER

$$s \equiv \tfrac{1}{2}(a+b+c+d), \tag{1}$$

and let R be the CIRCUMRADIUS. Then

$$A = \sqrt{(s-a)(s-b)(s-c)(s-d)} \tag{2}$$

$$= \frac{\sqrt{(ac+bd)(ad+bc)(ab+cd)}}{4R}. \tag{3}$$

Solving for the CIRCUMRADIUS gives

$$R = \tfrac{1}{4}\sqrt{\frac{(ac+bd)(ad+bc)(ab+cd)}{(s-a)(s-b)(s-c)(s-d)}}. \tag{4}$$

The DIAGONALS of a cyclic quadrilateral have lengths

$$p = \sqrt{\frac{(ab+cd)(ac+bd)}{ad+bc}} \tag{5}$$

$$q = \sqrt{\frac{(ac+bd)(ad+bc)}{ab+cd}}, \tag{6}$$

so that $pq = ac+bd$. In general, there are three essentially distinct cyclic quadrilaterals (modulo ROTATION and REFLECTION) whose edges are permutations of the lengths a, b, c, and d. Of the six corresponding DIAGONAL lengths, three are distinct. In addition to p and q, there is therefore a "third" DIAGONAL which can be denoted r. It is given by the equation

$$r = \sqrt{\frac{(ad+bc)(ab+cd)}{ac+bd}}. \tag{7}$$

This allows the AREA formula to be written in the particularly beautiful and simple form

$$A = \frac{pqr}{4R}. \tag{8}$$

The DIAGONALS are sometimes also denoted p, q, and r.

The AREA of a cyclic quadrilateral is the MAXIMUM possible for any QUADRILATERAL with the given side lengths. Also, the opposite ANGLES of a cyclic quadrilateral sum to π RADIANS (Dunham 1990).

A cyclic quadrilateral with RATIONAL sides a, b, c, and d, DIAGONALS p and q, CIRCUMRADIUS R, and AREA A is given by $a = 25$, $b = 33$, $c = 39$, $d = 65$, $p = 60$, $q = 52$, $R = 65/2$, and $A = 1344$.

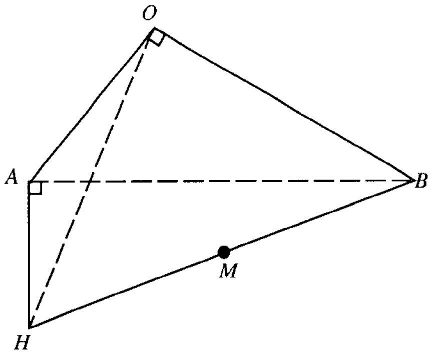

Let $AHBO$ be a QUADRILATERAL such that the angles $\angle HAB$ and $\angle HOB$ are RIGHT ANGLES, then $AHBO$ is a cyclic quadrilateral (Dunham 1990). This is a COROLLARY of the theorem that, in a RIGHT TRIANGLE, the MIDPOINT of the HYPOTENUSE is equidistant from the three VERTICES. Since M is the MIDPOINT of both RIGHT TRIANGLES ΔAHB and ΔBOH, it is equidistant from all four VERTICES, so a CIRCLE centered at M may be drawn through them. This theorem is one of the building blocks of Heron's derivation of HERON'S FORMULA.

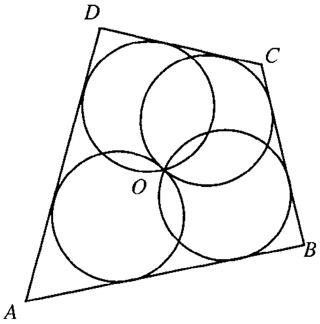

Place four equal CIRCLES so that they intersect in a point. The quadrilateral $ABCD$ is then a cyclic quadrilateral (Honsberger 1991). For a CONVEX cyclic quadrilateral Q, consider the set of CONVEX cyclic quadrilaterals $Q_{||}$ whose sides are PARALLEL to Q. Then the $Q_{||}$ of maximal AREA is the one whose DIAGONALS are PERPENDICULAR (Gürel 1996).

see also BRETSCHNEIDER'S FORMULA, CONCYCLIC, CYCLIC POLYGON, CYCLIC QUADRANGLE, EULER BRICK, HERON'S FORMULA, PTOLEMY'S THEOREM, QUADRILATERAL

References
Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 123, 1987.
Dunham, W. *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, p. 121, 1990.
Gürel, E. Solution to Problem 1472. "Maximal Area of Quadrilaterals." *Math. Mag.* **69**, 149, 1996.
Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., pp. 36–37, 1991.

Cyclic Redundancy Check

A sophisticated CHECKSUM (often abbreviated CRC), which is based on the algebra of polynomials over the integers (mod 2). It is substantially more reliable in detecting transmission errors, and is one common error-checking protocol used in modems.

see also CHECKSUM, ERROR-CORRECTING CODE

References
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Cyclic Redundancy and Other Checksums." Ch. 20.3 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 888–895, 1992.

Cyclid

see CYCLIDE

Cyclide

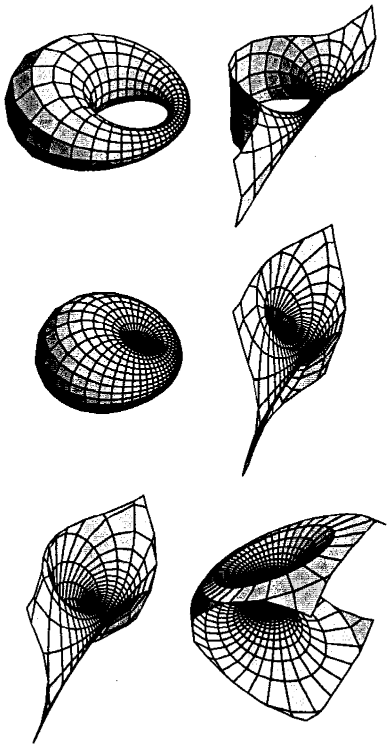

A pair of focal conics which are the envelopes of two one-parameter families of spheres, sometimes also called a CYCLID. The cyclide is a QUARTIC SURFACE, and the lines of curvature on a cyclide are all straight lines or circular arcs (Pinkall 1986). The STANDARD TORI and their inversions in a SPHERE S centered at a point $\mathbf{x}_0$ and of RADIUS r, given by

$$I(\mathbf{x}_0, r) = \mathbf{x}_0 + \frac{\mathbf{x} - \mathbf{x}_0 r^2}{|\mathbf{x} - \mathbf{x}_0|^2},$$

are both cyclides (Pinkall 1986). Illustrated above are RING CYCLIDES, HORN CYCLIDES, and SPINDLE CYCLIDES. The figures on the right correspond to $\mathbf{x}_0$ lying on the torus itself, and are called the PARABOLIC RING CYCLIDE, PARABOLIC HORN CYCLIDE, and PARABOLIC SPINDLE CYCLIDE, respectively.

see also CYCLIDIC COORDINATES, HORN CYCLIDE, PARABOLIC HORN CYCLIDE, PARABOLIC RING CYCLIDE, RING CYCLIDE, SPINDLE CYCLIDE, STANDARD TORI

References
Bierschneider-Jakobs, A. "Cyclides." `http://www.mi.uni-erlangen.de/~biersch/cyclides.html`.
Byerly, W. E. *An Elementary Treatise on Fourier's Series, and Spherical, Cylindrical, and Ellipsoidal Harmonics, with Applications to Problems in Mathematical Physics.* New York: Dover, p. 273, 1959.
Eisenhart, L. P. "Cyclides of Dupin." §133 in *A Treatise on the Differential Geometry of Curves and Surfaces.* New York: Dover, pp. 312–314, 1960.

Fischer, G. (Ed.). Plates 71–77 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 66–72, 1986.
Nordstrand, T. "Dupin Cyclide." `http://www.uib.no/people/nfytn/dupintxt.htm`.
Pinkall, U. "Cyclides of Dupin." §3.3 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 28–30, 1986.
Salmon, G. *Analytic Geometry of Three Dimensions.* New York: Chelsea, p. 527, 1979.

Cyclidic Coordinates

A general system of CURVILINEAR COORDINATES based on the CYCLIDE in which LAPLACE'S EQUATION is SEPARABLE.

References
Byerly, W. E. *An Elementary Treatise on Fourier's Series, and Spherical, Cylindrical, and Ellipsoidal Harmonics, with Applications to Problems in Mathematical Physics.* New York: Dover, p. 273, 1959.

Cycloid

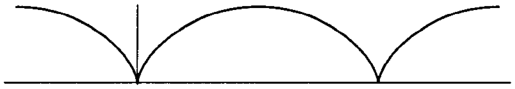

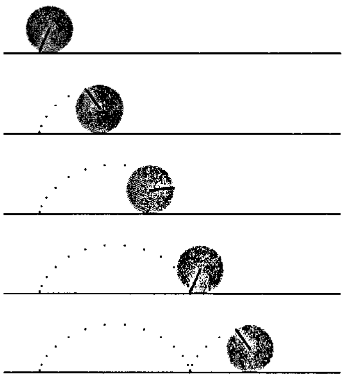

The cycloid is the locus of a point on the rim of a CIRCLE of RADIUS a rolling along a straight LINE. It was studied and named by Galileo in 1599. Galileo attempted to find the AREA by weighing pieces of metal cut into the shape of the cycloid. Torricelli, Fermat, and Descartes all found the AREA. The cycloid was also studied by Roberval in 1634, Wren in 1658, Huygens in 1673, and Johann Bernoulli in 1696. Roberval and Wren found the ARC LENGTH (MacTutor Archive). Gear teeth were also made out of cycloids, as first proposed by Desargues in the 1630s (Cundy and Rollett 1989).

In 1696, Johann Bernoulli challenged other mathematicians to find the curve which solves the BRACHISTOCHRONE PROBLEM, knowing the solution to be a cycloid. Leibniz, Newton, Jakob Bernoulli and L'Hospital all solved Bernoulli's challenge. The cycloid also solves the TAUTOCHRONE PROBLEM. Because of the frequency with which it provoked quarrels among mathematicians in the 17th century, the cycloid became known as the "Helen of Geometers" (Boyer 1968, p. 389).

The cycloid is the CATACAUSTIC of a CIRCLE for a RADIANT POINT on the circumference, as shown by Jakob and Johann Bernoulli in 1692. The CAUSTIC of the cycloid when the rays are parallel to the y-axis is a cycloid with twice as many arches. The RADIAL CURVE of a CYCLOID is a CIRCLE. The EVOLUTE and INVOLUTE of a cycloid are identical cycloids.

If the cycloid has a CUSP at the ORIGIN, its equation in CARTESIAN COORDINATES is

$$x = a\cos^{-1}\left(\frac{a-y}{a}\right) \mp \sqrt{2ay - y^2}. \tag{1}$$

In parametric form, this becomes

$$x = a(t - \sin t) \tag{2}$$

$$y = a(1 - \cos t). \tag{3}$$

If the cycloid is upside-down with a cusp at $(0, a)$, (2) and (3) become

$$x = 2a\sin^{-1}\left(\frac{y}{2a}\right) + \sqrt{2ay - y^2} \tag{4}$$

or

$$x = a(t + \sin t) \tag{5}$$

$$y = a(1 - \cos t) \tag{6}$$

(sign of $\sin t$ flipped for x).

The DERIVATIVES of the parametric representation (2) and (3) are

$$x' = a(1 - \cos t) \tag{7}$$

$$y' = a\sin t \tag{8}$$

$$\begin{aligned}\frac{dy}{dx} = \frac{y'}{x'} &= \frac{a\sin t}{a(1-\cos t)} = \frac{\sin t}{1-\cos t}\\ &= \frac{2\sin(\frac{1}{2}t)\cos(\frac{1}{2}t)}{2\sin^2(\frac{1}{2}t)} = \cot(\tfrac{1}{2}t).\end{aligned} \tag{9}$$

The squares of the derivatives are

$$x'^2 = a^2(1 - 2\cos t + \cos^2 t) \tag{10}$$

$$y'^2 = a^2\sin^2 t, \tag{11}$$

so the ARC LENGTH of a single cycle is

$$\begin{aligned}L &= \int ds = \int_0^{2\pi} \sqrt{x'^2 + y'^2}\,dt\\ &= a\int_0^{2\pi} \sqrt{(1 - 2\cos t + \cos^2 t) + \sin^2 t}\,dt\\ &= a\sqrt{2}\int_0^{2\pi} \sqrt{1-\cos t}\,dt = 2a\int_0^{2\pi} \sqrt{\frac{1-\cos t}{2}}\,dt\\ &= 2a\int_0^{2\pi} \left|\sin(\tfrac{1}{2}t)\right|\,dt.\end{aligned} \tag{12}$$

Now let $u \equiv t/2$ so $du = dt/2$. Then

$$L = 4a \int_0^\pi \sin u \, du = 4a[-\cos u]_0^\pi$$
$$= -4a[(-1) - 1] = 8a. \qquad (13)$$

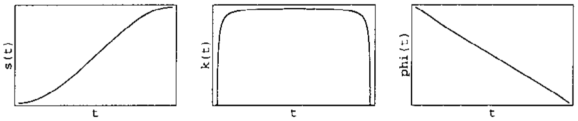

The ARC LENGTH, CURVATURE, and TANGENTIAL ANGLE are

$$s = 8a \sin^2(\tfrac{1}{4}t) \qquad (14)$$
$$\kappa = -\tfrac{1}{4}a \csc(\tfrac{1}{2}t) \qquad (15)$$
$$\phi = -\tfrac{1}{2}at. \qquad (16)$$

The AREA under a single cycle is

$$\begin{aligned} A &= \int_0^{2\pi} y\,dx = a^2 \int_0^{2\pi} (1-\cos\phi)(1-\cos\phi)\,d\phi \\ &= a^2 \int_0^{2\pi} (1-\cos\phi)^2\,d\phi \\ &= a^2 \int_0^{2\pi} (1 - 2\cos\phi + \cos^2\phi)\,d\phi \\ &= a^2 \int_0^{2\pi} \{1 - 2\cos\phi + \tfrac{1}{2}[1+\cos(2\phi)]\}\,d\phi \\ &= a^2 \int_0^{2\pi} [\tfrac{3}{2} - 2\cos\phi + \tfrac{1}{2}\cos(2\phi)]\,d\phi \\ &= a^2[\tfrac{3}{2}\phi - 2\sin\phi + \tfrac{1}{4}\sin(2\phi)]_0^{2\pi} \\ &= a^2 \tfrac{3}{2} 2\pi = 3\pi a^2. \end{aligned} \qquad (17)$$

The NORMAL is

$$\hat{\mathbf{T}} = \frac{1}{\sqrt{2-2\cos t}} \begin{bmatrix} 1-\cos t \\ \sin t \end{bmatrix}. \qquad (18)$$

see also CURTATE CYCLOID, CYCLIDE, CYCLOID EVOLUTE, CYCLOID INVOLUTE, EPICYCLOID, HYPOCYCLOID, PROLATE CYCLOID, TROCHOID

References

Bogomolny, A. "Cycloids." `http://www.cut-the-knot.com/pythagoras/cycloids.html`.

Boyer, C. B. *A History of Mathematics.* New York: Wiley, 1968.

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., 1989.

Gray, A. "Cycloids." §3.1 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 37–39, 1993.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 192 and 197, 1972.

Lee, X. "Cycloid." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Cycloid_dir/cycloid.html`.

Lockwood, E. H. "The Cycloid." Ch. 9 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 80–89, 1967.

MacTutor History of Mathematics Archive. "Cycloid." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Cycloid.html`.

Muterspaugh, J.; Driver, T.; and Dick, J. E. "The Cycloid and Tautochronism." `http://ezinfo.ucs.indiana.edu/~jedick/project/intro.html`.

Pappas, T. "The Cycloid—The Helen of Geometry." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 6–8, 1989.

Wagon, S. "Rolling Circles." Ch. 2 in *Mathematica in Action.* New York: W. H. Freeman, pp. 39–66, 1991.

Yates, R. C. "Cycloid." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 65–70, 1952.

Cycloid Evolute

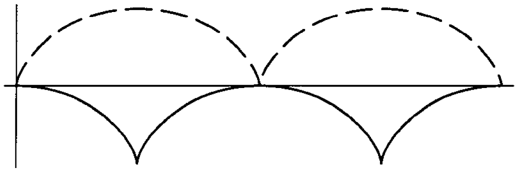

The EVOLUTE of the CYCLOID

$$x(t) = a(t - \sin t)$$
$$y(t) = a(1 - \cos t)$$

is given by

$$x(t) = a(t + \sin t)$$
$$y(t) = a(\cos t - 1).$$

As can be seen in the above figure, the EVOLUTE is simply a shifted copy of the original CYCLOID, so the CYCLOID is its own EVOLUTE.

Cycloid Involute

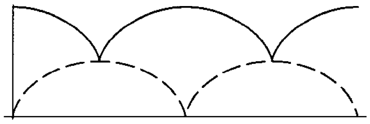

The INVOLUTE of the CYCLOID

$$x(t) = a(t - \sin t)$$
$$y(t) = a(1 - \cos t)$$

is given by

$$x(t) = a(t + \sin t)$$
$$y(t) = a(3 + \cos t).$$

As can be seen in the above figure, the INVOLUTE is simply a shifted copy of the original CYCLOID, so the CYCLOID is its own INVOLUTE!

Cycloid Radial Curve

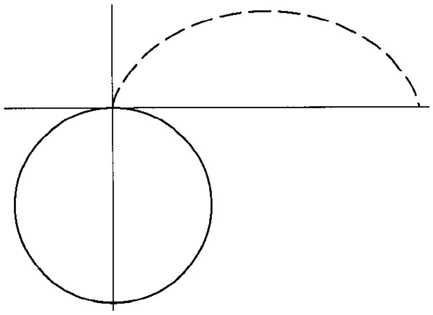

The RADIAL CURVE of the CYCLOID is the CIRCLE

$$\begin{aligned} x &= x_0 + 2a\sin\phi \\ y &= -2a + y_0 + 2a\cos\phi. \end{aligned}$$

Cyclomatic Number

see CIRCUIT RANK

Cyclotomic Equation

The equation

$$x^p = 1,$$

where solutions $\zeta_k = e^{2\pi ik/p}$ are the ROOTS OF UNITY sometimes called DE MOIVRE NUMBERS. Gauss showed that the cyclotomic equation can be reduced to solving a series of QUADRATIC EQUATIONS whenever p is a FERMAT PRIME. Wantzel (1836) subsequently showed that this condition is not only SUFFICIENT, but also NECESSARY. An "irreducible" cyclotomic equation is an expression of the form

$$\frac{x^p - 1}{x - 1} = x^{p-1} + x^{p-2} + \ldots + 1 = 0,$$

where p is PRIME. Its ROOTS z_i satisfy $|z_i| = 1$.

see also CYCLOTOMIC POLYNOMIAL, DE MOIVRE NUMBER, POLYGON, PRIMITIVE ROOT OF UNITY

References

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 99–100, 1996.

Scott, C. A. "The Binomial Equation $x^p - 1 = 0$." *Amer. J. Math.* **8**, 261–264, 1886.

Wantzel, P. L. "Recherches sur les moyens de reconnaître si un Problème de Géométrie peut se résoudre avec la règle et le compas." *J. Math. pures appliq.* **1**, 366–372, 1836.

Cyclotomic Factorization

$$z^p - y^p = (z - y)(z - \zeta y)\cdots(z - \zeta^{p-1}y),$$

where $\zeta \equiv e^{2\pi i/p}$ (a DE MOIVRE NUMBER) and p is a PRIME.

Cyclotomic Field

The smallest field containing $m \in \mathbb{Z} \geq 1$ with ζ a PRIME ROOT OF UNITY is denoted $\mathbb{R}_m(\zeta)$.

$$x^p + y^p = \prod_{k=1}^{p}(x + \zeta^k y).$$

Specific cases are

$$\begin{aligned} \mathbb{R}_3 &= \mathbb{Q}(\sqrt{-3}) \\ \mathbb{R}_4 &= \mathbb{Q}(\sqrt{-1}) \\ \mathbb{R}_6 &= \mathbb{Q}(\sqrt{-3}), \end{aligned}$$

where $\mathbb{Q}$ denotes a QUADRATIC FIELD.

Cyclotomic Integer

A number of the form

$$a_0 + a_1\zeta + \ldots + a_{p-1}\zeta^{p-1},$$

where

$$\zeta \equiv e^{2\pi i/p}$$

is a DE MOIVRE NUMBER and p is a PRIME number. Unique factorizations of cyclotomic INTEGERS fail for $p > 23$.

Cyclotomic Invariant

Let p be an ODD PRIME and F_n the CYCLOTOMIC FIELD of p^{n+1}th ROOTS of unity over the rational FIELD. Now let $p^{e(n)}$ be the POWER of p which divides the CLASS NUMBER h_n of F_n. Then there exist INTEGERS $\mu_p, \lambda_p \geq 0$ and ν_p such that

$$e(n) = \mu_p p^n + \lambda_p n + \nu_p$$

for all sufficiently large n. For REGULAR PRIMES, $\mu_p = \lambda_p = \nu_p = 0$.

References

Johnson, W. "Irregular Primes and Cyclotomic Invariants." *Math. Comput.* **29**, 113–120, 1975.

Cyclotomic Number

see DE MOIVRE NUMBER, SYLVESTER CYCLOTOMIC NUMBER

Cyclotomic Polynomial

A polynomial given by

$$\Phi_d(x) = \prod_{k=1}^{d}(x - \zeta_k), \tag{1}$$

where ζ_i are the primitive dth ROOTS OF UNITY in $\mathbb{C}$ given by $\zeta_k = e^{2\pi ik/d}$. The numbers ζ_k are sometimes called DE MOIVRE NUMBERS. $\Phi_d(x)$ is an irreducible

POLYNOMIAL in $\mathbb{Z}[x]$ with degree $\phi(d)$, where ϕ is the TOTIENT FUNCTION. For d PRIME,

$$\Phi_p = \sum_{k=0}^{p-1} x^k, \tag{2}$$

i.e., the coefficients are all 1. Φ_{105} has coefficients of -2 for x^7 and x^{41}, making it the first cyclotomic polynomial to have a coefficient other than ± 1 and 0. This is true because 105 is the first number to have three distinct ODD PRIME factors, i.e., $105 = 3 \cdot 5 \cdot 7$ (McClellan and Rader 1979, Schroeder 1997). Migotti (1883) showed that COEFFICIENTS of Φ_{pq} for p and q distinct PRIMES can be only 0, ± 1. Lam and Leung (1996) considered

$$\Phi_{pq} \equiv \sum_{k=0}^{pq-1} a_k x^k \tag{3}$$

for p, q PRIME. Write the TOTIENT FUNCTION as

$$\phi(pq) = (p-1)(q-1) = rp + sq \tag{4}$$

and let

$$0 \leq k \leq (p-1)(q-1), \tag{5}$$

then

1. $a_k = 1$ IFF $k = ip + jq$ for some $i \in [0, r]$ and $j \in [0, s]$,
2. $a_k = -1$ IFF $k + pq = ip + jq$ for $i \in [r+1, q-1]$ and $j \in [s+1, p-1]$,
3. otherwise $a_k = 0$.

The number of terms having $a_k = 1$ is $(r+1)(s+1)$, and the number of terms having $a_k = -1$ is $(p-s-1)(q-r-1)$. Furthermore, assume $q > p$, then the middle COEFFICIENT of Φ_{pq} is $(-1)^r$.

The LOGARITHM of the cyclotomic polynomial

$$\Phi_n(x) = \prod_{d|n} (1 - x^{n/d})^{\mu(d)} \tag{6}$$

is the MÖBIUS INVERSION FORMULA (Vardi 1991, p. 225).

The first few cyclotomic POLYNOMIALS are

$$\begin{aligned}
\Phi_1(x) &= x - 1 \\
\Phi_2(x) &= x + 1 \\
\Phi_3(x) &= x^2 + x + 1 \\
\Phi_4(x) &= x^2 + 1 \\
\Phi_5(x) &= x^4 + x^3 + x^2 + x + 1 \\
\Phi_6(x) &= x^2 - x + 1 \\
\Phi_7(x) &= x^6 + x^5 + x^4 + x^3 + x^2 + x + 1 \\
\Phi_8(x) &= x^4 + 1 \\
\Phi_9(x) &= x^6 + x^3 + 1 \\
\Phi_{10}(x) &= x^4 - x^3 + x^2 - x + 1.
\end{aligned}$$

The smallest values of n for which Φ_n has one or more coefficients ± 1, ± 2, ± 3, ... are 0, 105, 385, 1365, 1785, 2805, 3135, 6545, 6545, 10465, 10465, 10465, 10465, 10465, 11305, ... (Sloane's A013594).

The POLYNOMIAL $x^n - 1$ can be factored as

$$x^n - 1 = \prod_{d|n} \Phi_d(x), \tag{7}$$

where $\Phi_d(x)$ is a CYCLOTOMIC POLYNOMIAL. Furthermore,

$$x^n + 1 = \frac{x^{2n-1}}{x^n - 1} = \frac{\prod_{d|2n} \Phi_d(x)}{\prod_{d|n} \Phi_d(x)} = \prod_{d|m} \Phi_{2^t d}(x). \tag{8}$$

The COEFFICIENTS of the inverse of the cyclotomic POLYNOMIAL

$$\begin{aligned}
\frac{1}{1+x+x^2} &= 1 - x + x^3 - x^4 + x^6 - x^7 + x^9 - x^{10} + \ldots \\
&\equiv \sum_{n=0}^{\infty} c_n x^n
\end{aligned} \tag{9}$$

can also be computed from

$$c_n = 1 - 2\left\lfloor \tfrac{1}{3}(n+2) \right\rfloor + \left\lfloor \tfrac{1}{3}(n+1) \right\rfloor + \left\lfloor \tfrac{1}{3}n \right\rfloor, \tag{10}$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION.

see also AURIFEUILLEAN FACTORIZATION, MÖBIUS INVERSION FORMULA

References

Beiter, M. "The Midterm Coefficient of the Cyclotomic Polynomial $F_{pq}(x)$." *Amer. Math. Monthly* **71**, 769–770, 1964.

Beiter, M. "Magnitude of the Coefficients of the Cyclotomic Polynomial F_{pq}." *Amer. Math. Monthly* **75**, 370–372, 1968.

Bloom, D. M. "On the Coefficients of the Cyclotomic Polynomials." *Amer. Math. Monthly* **75**, 372–377, 1968.

Carlitz, L. "The Number of Terms in the Cyclotomic Polynomial $F_{pq}(x)$." *Amer. Math. Monthly* **73**, 979–981, 1966.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, 1996.

de Bruijn, N. G. "On the Factorization of Cyclic Groups." *Indag. Math.* **15**, 370–377, 1953.

Lam, T. Y. and Leung, K. H. "On the Cyclotomic Polynomial $\Phi_{pq}(X)$." *Amer. Math. Monthly* **103**, 562–564, 1996.

Lehmer, E. "On the Magnitude of Coefficients of the Cyclotomic Polynomials." *Bull. Amer. Math. Soc.* **42**, 389–392, 1936.

McClellan, J. H. and Rader, C. *Number Theory in Digital Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1979.

Migotti, A. "Zur Theorie der Kreisteilungsgleichung." *Sitzber. Math.-Naturwiss. Classe der Kaiser. Akad. der Wiss., Wien* **87**, 7–14, 1883.

Schroeder, M. R. *Number Theory in Science and Communication, with Applications in Cryptography, Physics, Digital Information, Computing, and Self-Similarity, 3rd ed.* New York: Springer-Verlag, p. 245, 1997.

Sloane, N. J. A. Sequence A013594 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Vardi, I. *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 8 and 224–225, 1991.

Cylinder

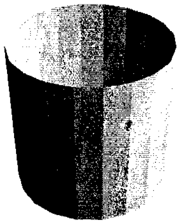

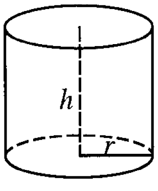

A cylinder is a solid of circular CROSS-SECTION in which the centers of the CIRCLES all lie on a single LINE. The cylinder was extensively studied by Archimedes in his 2-volume work *On the Sphere and Cylinder* in ca. 225 BC.

A cylinder is called a right cylinder if it is "straight" in the sense that its cross-sections lie directly on top of each other; otherwise, the cylinder is called oblique. The surface of a cylinder of height h and RADIUS r can be described parametrically by

$$x = r\cos\theta \tag{1}$$
$$y = r\sin\theta \tag{2}$$
$$z = z, \tag{3}$$

for $z \in [0,h]$ and $\theta \in [0,2\pi)$. These are the basis for CYLINDRICAL COORDINATES. The SURFACE AREA (of the sides) and VOLUME of the cylinder of height h and RADIUS r are

$$S = 2\pi rh \tag{4}$$
$$V = \pi r^2 h. \tag{5}$$

Therefore, if top and bottom caps are added, the volume-to-surface area ratio for a cylindrical container is

$$\frac{V}{S} = \frac{\pi r^2 h}{2\pi rh + 2\pi r^2} = \frac{1}{2}\left(\frac{1}{r}+\frac{1}{h}\right)^{-1}, \tag{6}$$

which is related to the HARMONIC MEAN of the radius r and height h.

see also CONE, CYLINDER-SPHERE INTERSECTION, CYLINDRICAL SEGMENT, ELLIPTIC CYLINDER, GENERALIZED CYLINDER, SPHERE, STEINMETZ SOLID, VIVIANI'S CURVE

References
Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 129, 1987.

Cylinder Cutting

The maximum number of pieces into which a cylinder can be divided by n oblique cuts is given by

$$f(n) = \binom{n+1}{3} + n + 1 = \tfrac{1}{6}(n+2)(n+3),$$

where $\binom{a}{b}$ is a BINOMIAL COEFFICIENT. This problem is sometimes also called CAKE CUTTING or PIE CUTTING. For $n = 1, 2, \ldots$ cuts, the maximum number of pieces is 2, 4, 8, 15, 26, 42, ... (Sloane's A000125).

see also CIRCLE CUTTING, HAM SANDWICH THEOREM, PANCAKE THEOREM, TORUS CUTTING

References
Bogomolny, A. "Can You Cut a Cake into 8 Pieces with Three Movements." `http://www.cut-the-knot.com/do_you_know/cake.html`.
Sloane, N. J. A. Sequence A000125/M1100 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Cylinder-Cylinder Intersection

see STEINMETZ SOLID

Cylinder Function

The cylinder function is defined as

$$C(x,y) \equiv \begin{cases} 1 & \text{for } \sqrt{x^2+y^2} \leq a \\ 0 & \text{for } \sqrt{x^2+y^2} > a. \end{cases} \tag{1}$$

The BESSEL FUNCTIONS are sometimes also called cylinder functions. To find the FOURIER TRANSFORM of the cylinder function, let

$$k_x = k\cos\alpha \tag{2}$$
$$k_y = k\sin\alpha \tag{3}$$

$$x = r\cos\theta \tag{4}$$
$$y = r\sin\theta. \tag{5}$$

Then

$$\begin{aligned} F(k,a) &= \mathcal{F}(C(x,y)) \\ &= \int_0^{2\pi}\int_0^a e^{i(k\cos\alpha r\cos\theta + k\sin\alpha r\sin\theta)}r\,dr\,d\theta \\ &= \int_0^{2\pi}\int_0^a e^{ikr\cos(\theta-\alpha)}r\,dr\,d\theta. \end{aligned} \tag{6}$$

Let $b = \theta - \alpha$, so $db = d\theta$. Then

$$\begin{aligned} F(k,a) &= \int_{-\alpha}^{2\pi-\alpha}\int_0^a e^{ikr\cos b}r\,dr\,d\theta \\ &= \int_0^{2\pi}\int_0^a e^{ikr\cos b}r\,dr\,d\theta \\ &= 2\pi\int_0^a J_0(kr)r\,dr, \end{aligned} \tag{7}$$

where J_0 is a zeroth order BESSEL FUNCTION OF THE FIRST KIND. Let $u \equiv kr$, so $du = k\,dr$, then

$$\begin{aligned} F(k,a) &= \frac{2\pi}{k^2}\int_0^{ka} J_0(u)u\,du = \frac{2\pi}{k^2}[uJ_1(u)]_0^{ka} \\ &= \frac{2\pi a}{k}J_1(ka) = 2\pi a^2\frac{J_1(ka)}{ka}. \end{aligned} \tag{8}$$

As defined by Watson (1966), a "cylinder function" is any function which satisfies the RECURRENCE RELATIONS

$$\mathcal{C}_{\nu-1}(z)+\mathcal{C}_{\nu+1}(z)=\frac{2\nu}{z}\mathcal{C}_\nu(z) \tag{9}$$

$$\mathcal{C}_{\nu-1}(z)-\mathcal{C}_{\nu+1}(z)=2\mathcal{C}_\nu'(z). \tag{10}$$

This class of functions can be expressed in terms of BESSEL FUNCTIONS.

see also BESSEL FUNCTION OF THE FIRST KIND, CYLINDER FUNCTION, CYLINDRICAL FUNCTION, HEMISPHERICAL FUNCTION

References

Watson, G. N. *A Treatise on the Theory of Bessel Functions, 2nd ed.* Cambridge, England: Cambridge University Press, 1966.

Cylinder-Sphere Intersection

see VIVIANI'S CURVE

Cylindrical Coordinates

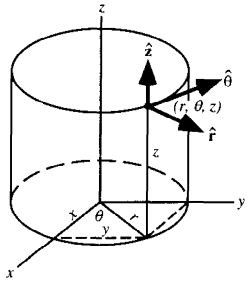

Cylindrical coordinates are a generalization of 2-D POLAR COORDINATES to 3-D by superposing a height (z) axis. Unfortunately, there are a number of different notations used for the other two coordinates. Either r or ρ is used to refer to the radial coordinate and either ϕ or θ to the azimuthal coordinates. Arfken (1985), for instance, uses (ρ,ϕ,z), while Beyer (1987) uses (r,θ,z). In this work, the NOTATION (r,θ,z) is used.

$$r=\sqrt{x^2+y^2} \tag{1}$$

$$\theta=\tan^{-1}\left(\frac{y}{x}\right) \tag{2}$$

$$z=z, \tag{3}$$

where $r\in[0,\infty)$, $\theta\in[0,2\pi)$, and $z\in(-\infty,\infty)$. In terms of x, y, and z

$$x=r\cos\theta \tag{4}$$

$$y=r\sin\theta \tag{5}$$

$$z=z. \tag{6}$$

Morse and Feshbach (1953) define the cylindrical coordinates by

$$x=\xi_1\xi_2 \tag{7}$$

$$y=\xi_1\sqrt{1-{\xi_2}^2} \tag{8}$$

$$z=\xi_3, \tag{9}$$

where $\xi_1=r$ and $\xi_2=\cos\theta$. The METRIC elements of the cylindrical coordinates are

$$g_{rr}=1 \tag{10}$$

$$g_{\theta\theta}=r^2 \tag{11}$$

$$g_{zz}=1, \tag{12}$$

so the SCALE FACTORS are

$$g_r=1 \tag{13}$$

$$g_\theta=r \tag{14}$$

$$g_z=1. \tag{15}$$

The LINE ELEMENT is

$$ds=dr\,\hat{\mathbf{r}}+r\,d\theta\,\hat{\boldsymbol{\theta}}+dz\,\hat{\mathbf{z}}, \tag{16}$$

and the VOLUME ELEMENT is

$$dV=r\,dr\,d\theta\,dz. \tag{17}$$

The JACOBIAN is

$$\left|\frac{\partial(x,y,z)}{\partial(r,\theta,z)}\right|=r. \tag{18}$$

A CARTESIAN VECTOR is given in CYLINDRICAL COORDINATES by

$$\mathbf{r}=\begin{bmatrix} r\cos\theta \\ r\sin\theta \\ z \end{bmatrix}. \tag{19}$$

To find the UNIT VECTORS,

$$\hat{\mathbf{r}}\equiv\frac{\frac{d\mathbf{r}}{dr}}{\left|\frac{d\mathbf{r}}{dr}\right|}=\begin{bmatrix} \cos\theta \\ \sin\theta \\ 0 \end{bmatrix} \tag{20}$$

$$\hat{\boldsymbol{\theta}}\equiv\frac{\frac{d\mathbf{r}}{d\theta}}{\left|\frac{d\mathbf{r}}{d\theta}\right|}=\begin{bmatrix} -\sin\theta \\ \cos\theta \\ 0 \end{bmatrix} \tag{21}$$

$$\hat{\mathbf{z}}\equiv\frac{\frac{d\mathbf{r}}{dz}}{\left|\frac{d\mathbf{r}}{dz}\right|}=\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{22}$$

Derivatives of unit VECTORS with respect to the coordinates are

$$\frac{\partial\hat{\mathbf{r}}}{\partial r}=\mathbf{0} \tag{23}$$

$$\frac{\partial\hat{\mathbf{r}}}{\partial\theta}=\begin{bmatrix} -\sin\theta \\ \cos\theta \\ 0 \end{bmatrix}=\hat{\boldsymbol{\theta}} \tag{24}$$

$$\frac{\partial\hat{\mathbf{r}}}{\partial z}=\mathbf{0} \tag{25}$$

$$\frac{\partial\hat{\boldsymbol{\theta}}}{\partial r}=\mathbf{0} \tag{26}$$

$$\frac{\partial\hat{\boldsymbol{\theta}}}{\partial\theta}=\begin{bmatrix} -\cos\theta \\ -\sin\theta \\ 0 \end{bmatrix}=-\hat{\mathbf{r}} \tag{27}$$

$$\frac{\partial\hat{\boldsymbol{\theta}}}{\partial z}=\mathbf{0} \tag{28}$$

$$\frac{\partial\hat{\mathbf{z}}}{\partial r}=\mathbf{0} \tag{29}$$

$$\frac{\partial\hat{\mathbf{z}}}{\partial\theta}=\mathbf{0} \tag{30}$$

$$\frac{\partial\hat{\mathbf{z}}}{\partial z}=\mathbf{0}. \tag{31}$$

The GRADIENT of a VECTOR FIELD in cylindrical coordinates is given by

$$\nabla \equiv \hat{\mathbf{r}}\frac{\partial}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial}{\partial \theta} + \hat{\mathbf{z}}\frac{\partial}{\partial z}, \tag{32}$$

so the GRADIENT components become

$$\nabla_r \hat{\mathbf{r}} = \mathbf{0} \tag{33}$$

$$\nabla_\theta \hat{\mathbf{r}} = \frac{1}{r}\hat{\boldsymbol{\theta}} \tag{34}$$

$$\nabla_z \hat{\mathbf{r}} = \mathbf{0} \tag{35}$$

$$\nabla_r \hat{\boldsymbol{\theta}} = \mathbf{0} \tag{36}$$

$$\nabla_\theta \hat{\boldsymbol{\theta}} = -\frac{1}{r}\hat{\mathbf{r}} \tag{37}$$

$$\nabla_z \hat{\boldsymbol{\theta}} = \mathbf{0} \tag{38}$$

$$\nabla_r \hat{\mathbf{z}} = \mathbf{0} \tag{39}$$

$$\nabla_\theta \hat{\mathbf{z}} = \mathbf{0} \tag{40}$$

$$\nabla_z \hat{\mathbf{z}} = \mathbf{0}. \tag{41}$$

Now, since the CONNECTION COEFFICIENTS are defined by

$$\Gamma^i_{jk} = \hat{\mathbf{x}}_i \cdot (\nabla_k \hat{\mathbf{x}}_j), \tag{42}$$

$$\Gamma^r = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -\frac{1}{r} & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{43}$$

$$\Gamma^\theta = \begin{bmatrix} 0 & \frac{1}{r} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{44}$$

$$\Gamma^z = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{45}$$

The COVARIANT DERIVATIVES, given by

$$A_{j;k} = \frac{1}{g^{kk}}\frac{\partial A_j}{\partial x_k} - \Gamma^i_{jk}A_i, \tag{46}$$

are

$$A_{r;r} = \frac{\partial A_r}{\partial r} - \Gamma^i_{rr}A_i = \frac{\partial A_r}{\partial r} \tag{47}$$

$$A_{r;\theta} = \frac{1}{r}\frac{\partial A_r}{\partial \theta} - \Gamma^i_{r\theta}A_i = \frac{1}{r}\frac{\partial A_\theta}{\partial r} - \Gamma^\theta_{r\theta}A_\theta$$

$$= \frac{1}{r}\frac{\partial A_r}{\partial \theta} - \frac{A_\theta}{r} \tag{48}$$

$$A_{r;z} = \frac{\partial A_r}{\partial z} - \Gamma^i_{rz}A_i = \frac{\partial A_r}{\partial z} \tag{49}$$

$$A_{\theta;r} = \frac{\partial A_\theta}{\partial r}\Gamma^i_{\theta r}A_i = \frac{\partial A_\theta}{\partial r} \tag{50}$$

$$A_{\theta;\theta} = \frac{1}{r}\frac{\partial A_\theta}{\partial \theta} - \Gamma^i_{\theta\theta}A_i = \frac{1}{r}\frac{\partial A_\theta}{\partial \theta} - \Gamma^r_{\theta\theta}A_r$$

$$= \frac{1}{r}\frac{\partial A_\theta}{\partial \theta} + \frac{A_r}{r} \tag{51}$$

$$A_{\theta;z} = \frac{\partial A_\theta}{\partial z} - \Gamma^i_{\theta z}A_i = \frac{\partial A_\theta}{\partial z} \tag{52}$$

$$A_{z;r} = \frac{\partial A_z}{\partial r} - \Gamma^i_{zr}A_i = \frac{\partial A_z}{\partial r} \tag{53}$$

$$A_{z;\theta} = \frac{1}{r}\frac{\partial A_z}{\partial \theta} - \Gamma^i_{z\theta}A_i = \frac{1}{r}\frac{\partial A_z}{\partial \theta} \tag{54}$$

$$A_{z;z} = \frac{\partial A_z}{\partial z} - \Gamma^i_{zz}A_i = \frac{\partial A_z}{\partial z}. \tag{55}$$

CROSS PRODUCTS of the coordinate axes are

$$\hat{\mathbf{r}} \times \hat{\mathbf{z}} = -\hat{\boldsymbol{\theta}} \tag{56}$$

$$\hat{\boldsymbol{\theta}} \times \hat{\mathbf{z}} = \hat{\mathbf{r}} \tag{57}$$

$$\hat{\mathbf{r}} \times \hat{\boldsymbol{\theta}} = \hat{\mathbf{z}}. \tag{58}$$

The COMMUTATION COEFFICIENTS are given by

$$c^\mu_{\alpha\beta}\vec{e}_\mu = [\vec{e}_\alpha, \vec{e}_\beta] = \nabla_\alpha \vec{e}_\beta - \nabla_\beta \vec{e}_\alpha, \tag{59}$$

But

$$[\hat{\mathbf{r}}, \hat{\mathbf{r}}] = [\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}] = [\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\phi}}] = \mathbf{0}, \tag{60}$$

so $c^\alpha_{rr} = c^\alpha_{\theta\theta} = c^\alpha_{\phi\phi} = 0$, where $\alpha = r, \theta, \phi$. Also

$$[\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}] = -[\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}] = \nabla_r \hat{\boldsymbol{\theta}} - \nabla_\theta \hat{\mathbf{r}} = 0 - \frac{1}{r}\hat{\boldsymbol{\theta}} = -\frac{1}{r}\hat{\boldsymbol{\theta}}, \tag{61}$$

so $c^\theta_{r\theta} = -c^\theta_{\theta r} = -\frac{1}{r}$, $c^r_{r\theta} = c^\phi_{r\theta} = 0$. Finally,

$$[\hat{\mathbf{r}}, \hat{\boldsymbol{\phi}}] = [\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}}] = 0. \tag{62}$$

Summarizing,

$$c^r = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{63}$$

$$c^\theta = \begin{bmatrix} 0 & -\frac{1}{r} & 0 \\ \frac{1}{r} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{64}$$

$$c^\phi = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{65}$$

Time DERIVATIVES of the VECTOR are

$$\dot{\mathbf{r}} = \begin{bmatrix} \cos\theta\,\dot{r} - r\sin\theta\,\dot{\theta} \\ \sin\theta\,\dot{r} + r\cos\theta\,\dot{\theta} \\ \dot{z} \end{bmatrix} = \dot{r}\,\hat{\mathbf{r}} + r\dot{\theta}\,\hat{\boldsymbol{\theta}} + \dot{z}\,\hat{\mathbf{z}} \tag{66}$$

$$\begin{aligned} \ddot{\mathbf{r}} &= \begin{bmatrix} -\sin\theta\,\dot{r}\dot{\theta} + \cos\theta\,\ddot{r} - \sin\theta\,\dot{r}\dot{\theta} - r\cos\theta\,\dot{\theta}^2 - r\sin\theta\,\ddot{\theta} \\ \cos\theta\,\dot{r}\dot{\theta} + \sin\theta\,\ddot{r} + \cos\theta\,\dot{r}\dot{\theta} - r\sin\theta\,\dot{\theta}^2 + r\cos\theta\,\ddot{\theta} \\ \ddot{z} \end{bmatrix} \\ &= \begin{bmatrix} -2\sin\theta\,\dot{r}\dot{\theta} + \cos\theta\,\ddot{r} - r\cos\theta\,\dot{\theta}^2 - r\sin\theta\,\ddot{\theta} \\ 2\cos\theta\,\dot{r}\dot{\theta} + \sin\theta\,\ddot{r} - r\sin\theta\,\dot{\theta}^2 + r\cos\theta\,\ddot{\theta} \\ \ddot{z} \end{bmatrix} \\ &= (\ddot{r} - r\dot{\theta}^2)\hat{\mathbf{r}} + (2\dot{r}\dot{\theta} + r\ddot{\theta})\hat{\boldsymbol{\theta}} + \ddot{z}\,\hat{\mathbf{z}}. \end{aligned} \tag{67}$$

SPEED is given by

$$v \equiv |\dot{\mathbf{r}}| = \sqrt{\dot{r}^2 + r^2\dot{\theta}^2 + \dot{z}^2}. \tag{68}$$

Time derivatives of the unit VECTORS are

$$\dot{\hat{\mathbf{r}}} = \begin{bmatrix} -\sin\theta\,\dot{\theta} \\ \cos\theta\,\dot{\theta} \\ 0 \end{bmatrix} = \dot{\theta}\,\hat{\boldsymbol{\theta}} \tag{69}$$

$$\dot{\hat{\boldsymbol{\theta}}} = \begin{bmatrix} -\cos\theta\,\dot{\theta} \\ -\sin\theta\,\dot{\theta} \\ 0 \end{bmatrix} = -\dot{\theta}\,\hat{\mathbf{r}} \tag{70}$$

$$\dot{\hat{\mathbf{z}}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0}. \tag{71}$$

CROSS PRODUCTS of the axes are

$$\hat{\mathbf{r}} \times \hat{\mathbf{z}} = -\hat{\boldsymbol{\theta}} \tag{72}$$

$$\hat{\boldsymbol{\theta}} \times \hat{\mathbf{z}} = \hat{\mathbf{r}} \tag{73}$$

$$\hat{\mathbf{r}} \times \hat{\boldsymbol{\theta}} = \hat{\mathbf{z}}. \tag{74}$$

The CONVECTIVE DERIVATIVE is

$$\frac{D\dot{\mathbf{r}}}{Dt} \equiv \left(\frac{\partial}{\partial t} + \dot{\mathbf{r}} \cdot \nabla\right) \dot{\mathbf{r}} = \frac{\partial \dot{\mathbf{r}}}{\partial t} + \dot{\mathbf{r}} \cdot \nabla \dot{\mathbf{r}}. \tag{75}$$

To rewrite this, use the identity

$$\nabla(\mathbf{A}\cdot\mathbf{B}) = \mathbf{A}\times(\nabla\times\mathbf{B})+\mathbf{B}\times(\nabla\times\mathbf{A})+(\mathbf{A}\cdot\nabla)\mathbf{B}+(\mathbf{B}\cdot\nabla)\mathbf{A} \tag{76}$$

and set $\mathbf{A} = \mathbf{B}$, to obtain

$$\nabla(\mathbf{A} \cdot \mathbf{A}) = 2\mathbf{A} \times (\nabla \times \mathbf{A}) + 2(\mathbf{A} \cdot \nabla)\mathbf{A}, \tag{77}$$

so

$$(\mathbf{A} \cdot \nabla)\mathbf{A} = \nabla(\tfrac{1}{2}\mathbf{A}^2) - \mathbf{A} \times (\nabla \times \mathbf{A}). \tag{78}$$

Then

$$\frac{D\dot{\mathbf{r}}}{Dt} = \ddot{\mathbf{r}}+\nabla(\tfrac{1}{2}\dot{\mathbf{r}}^2)-\dot{\mathbf{r}}\times(\nabla\times\dot{\mathbf{r}}) = \ddot{\mathbf{r}}+(\nabla\times\dot{\mathbf{r}})\times\dot{\mathbf{r}}+\nabla(\tfrac{1}{2}\dot{\mathbf{r}}^2). \tag{79}$$

The CURL in the above expression gives

$$\nabla \times \dot{\mathbf{r}} = \frac{1}{r}\frac{\partial}{\partial r}(r^2\dot{\theta})\hat{\mathbf{z}} = 2\dot{\theta}\hat{\mathbf{z}}, \tag{80}$$

so

$$\begin{aligned} -\dot{\mathbf{r}} \times (\nabla \times \dot{\mathbf{r}}) &= -2\dot{\theta}(\dot{r}\hat{\mathbf{r}} \times \hat{\mathbf{z}} + r\dot{\theta}\hat{\boldsymbol{\theta}} \times \hat{\mathbf{z}}) \\ &= -2\dot{\theta}(-\dot{r}\hat{\boldsymbol{\theta}} + r\dot{\theta}\hat{\mathbf{r}}) = 2\dot{r}\dot{\theta}\hat{\boldsymbol{\theta}} - 2r\dot{\theta}^2\hat{\mathbf{r}}. \end{aligned} \tag{81}$$

We expect the gradient term to vanish since SPEED does not depend on position. Check this using the identity $\nabla(f^2) = 2f\nabla f$,

$$\nabla(\tfrac{1}{2}\dot{\mathbf{r}}^2) = \tfrac{1}{2}\nabla(\dot{r}^2 + r^2\dot{\theta}^2 + \dot{z}^2) = \dot{r}\nabla\dot{r} + r\dot{\theta}\nabla(r\dot{\theta}) + \dot{z}\nabla\dot{z}. \tag{82}$$

Examining this term by term,

$$\dot{r}\nabla\dot{r} = \dot{r}\frac{\partial}{\partial t}\nabla r = \dot{r}\frac{\partial}{\partial t}\hat{\mathbf{r}} = \dot{r}\dot{\hat{\mathbf{r}}} = \dot{r}\dot{\theta}\hat{\boldsymbol{\theta}} \tag{83}$$

$$\begin{aligned} r\dot{\theta}\nabla(r\dot{\theta}) &= r\dot{\theta}\left[r\frac{\partial}{\partial t}\nabla\theta + \dot{\theta}\nabla r\right] = r\dot{\theta}\left[r\frac{\partial}{\partial t}\left(\frac{1}{r}\hat{\boldsymbol{\theta}}\right) + \dot{\theta}\hat{\mathbf{r}}\right] \\ &= r\dot{\theta}\left[r\left(-\frac{1}{r^2}\dot{r}\hat{\boldsymbol{\theta}} + \frac{1}{r}\dot{\hat{\boldsymbol{\theta}}}\right) + \dot{\theta}\hat{\mathbf{r}}\right] \\ &= -\dot{\theta}\dot{r}\hat{\boldsymbol{\theta}} + r\dot{\theta}(-\dot{\theta}\hat{\mathbf{r}}) + r\dot{\theta}^2\hat{\mathbf{r}} = -\dot{\theta}\dot{r}\hat{\boldsymbol{\theta}} \end{aligned} \tag{84}$$

$$\dot{z}\nabla\dot{z} = \dot{z}\frac{\partial}{\partial t}\nabla z = \dot{z}\frac{\partial}{\partial t}\hat{\mathbf{z}} = \dot{z}\dot{\hat{\mathbf{z}}} = \mathbf{0}, \tag{85}$$

so, as expected,

$$\nabla(\tfrac{1}{2}\dot{\mathbf{r}}^2) = \mathbf{0}. \tag{86}$$

We have already computed $\ddot{\mathbf{r}}$, so combining all three pieces gives

$$\begin{aligned} \frac{D\dot{\mathbf{r}}}{Dt} &= (\ddot{r} - r\dot{\theta}^2 - 2r\dot{\theta}^2)\hat{\mathbf{r}} + (2\dot{r}\dot{\theta} + 2\dot{r}\dot{\theta} + r\ddot{\theta})\hat{\boldsymbol{\theta}} + \ddot{z}\hat{\mathbf{z}} \\ &= (\ddot{r} - 3r\dot{\theta}^2)\hat{\mathbf{r}} + (4\dot{r}\dot{\theta} + r\ddot{\theta})\hat{\boldsymbol{\theta}} + \ddot{z}\hat{\mathbf{z}}. \end{aligned} \tag{87}$$

The DIVERGENCE is

$$\begin{aligned} \nabla \cdot A = A^r_{;r} &= A^r_{,r} + \left(\Gamma^r_{rr}A^t + \Gamma^r_{\theta r}A^\theta + \Gamma^r_{zr}A^z\right) + A^\theta_{,\theta} \\ &\quad + \left(\Gamma^\theta_{r\theta}A^r + \Gamma^\theta_{\theta\theta}A^\theta + \Gamma^\theta_{z\theta}A^z\right) \\ &\quad + A^z_{,z} + \left(\Gamma^z_{rz}A^r + \Gamma^z_{\theta z}A^\theta + \Gamma^z_{zz}A^z\right) \\ &= A^r_{,r} + A^\theta_{,\theta} + A^z_{,z} + (0+0+0) + \left(\frac{1}{r} + 0 + 0\right) \\ &\quad + (0+0+0) \\ &= \frac{1}{g_r}\frac{\partial}{\partial r}A^r + \frac{1}{g_\theta}\frac{\partial}{\partial \theta}A^\theta + \frac{1}{g_z}\frac{\partial}{\partial z}A^z + \frac{1}{r}A^r \\ &= \left(\frac{\partial}{\partial r} + \frac{1}{r}\right)A^r + \frac{1}{r}\frac{\partial}{\partial \theta}A^\theta + \frac{\partial}{\partial z}A^z, \end{aligned} \tag{88}$$

or, in VECTOR notation

$$\nabla \cdot \mathbf{F} = \frac{1}{r}\frac{\partial}{\partial r}(rF_r) + \frac{1}{r}\frac{\partial F_\theta}{\partial \theta} + \frac{\partial F_z}{\partial z}. \tag{89}$$

The CROSS PRODUCT is

$$\begin{aligned} \nabla \times \mathbf{F} &= \left(\frac{1}{r}\frac{\partial F_z}{\partial \theta} - \frac{\partial F_\theta}{\partial z}\right)\hat{\mathbf{r}} + \left(\frac{\partial F_r}{\partial z} - \frac{\partial F_z}{\partial r}\right)\hat{\boldsymbol{\theta}} \\ &\quad + \frac{1}{r}\left[\frac{\partial}{\partial r}(rF_\theta) - \frac{\partial F_r}{\partial \theta}\right]\hat{\mathbf{z}}, \end{aligned} \tag{90}$$

and the LAPLACIAN is

$$\begin{aligned} \nabla^2 f &\equiv \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial f}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} + \frac{\partial^2 f}{\partial z^2} \\ &= \frac{\partial^2 f}{\partial r^2} + \frac{1}{r}\frac{\partial f}{\partial r} + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} + \frac{\partial^2 f}{\partial z^2}. \end{aligned} \tag{91}$$

The vector LAPLACIAN is

$$\nabla^2\mathbf{v} = \begin{bmatrix} \frac{\partial^2 v_r}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 v_r}{\partial \phi^2} + \frac{\partial^2 v_r}{z^2} + \frac{1}{r}\frac{\partial v_r}{\partial r} - \frac{2}{r^2}\frac{\partial v_\phi}{\partial \phi} - \frac{v_r}{r^2} \\ \frac{\partial^2}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 v_\phi}{\partial \phi^2} + \frac{\partial^2 v_\phi}{\partial z^2} + \frac{1}{r}\frac{\partial v_\phi}{\partial r} + \frac{2}{r^2}\frac{\partial v_r}{\partial \phi} - \frac{v_\phi}{r^2} \\ \frac{\partial^2 v_z}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 v_z}{\partial \phi^2} + \frac{\partial^2 v_z}{\partial z^2} + \frac{1}{r}\frac{\partial v_z}{\partial r} \end{bmatrix}. \tag{92}$$

The HELMHOLTZ DIFFERENTIAL EQUATION is separable in cylindrical coordinates and has STÄCKEL DETERMINANT $S = 1$ (for r, θ, z) or $S = 1/(1 - {\xi_2}^2)$ (for Morse and Feshbach's ξ_1, ξ_2, ξ_3).

see also ELLIPTIC CYLINDRICAL COORDINATES, HELMHOLTZ DIFFERENTIAL EQUATION—CIRCULAR CYLINDRICAL COORDINATES, POLAR COORDINATES, SPHERICAL COORDINATES

References

Arfken, G. "Circular Cylindrical Coordinates." §2.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 95–101, 1985.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 212, 1987.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 657, 1953.

Cylindrical Equal-Area Projection

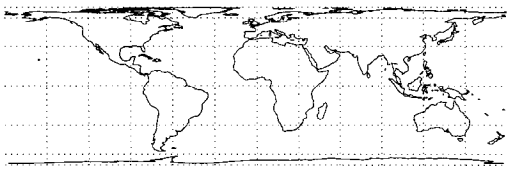

The MAP PROJECTION having transformation equations,

$$x = (\lambda - \lambda_0)\cos\phi_s \tag{1}$$

$$y = \frac{\sin\phi}{\cos\phi_s} \tag{2}$$

for the normal aspect, and inverse transformation equations

$$\phi = \sin^{-1}(y\cos\phi_s) \tag{3}$$

$$\lambda = \frac{x}{\cos\phi_s} + \lambda_0. \tag{4}$$

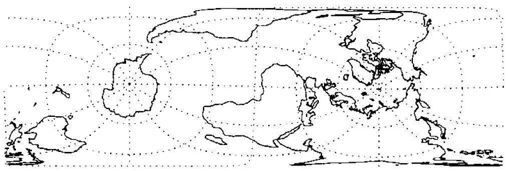

An oblique form of the cylindrical equal-area projection is given by the equations

$$\lambda_p = \tan^{-1}\left(\frac{\cos\phi_1\sin\phi_2\cos\lambda_1 - \sin\phi_1\cos\phi_2\cos\lambda_2}{\sin\phi_1\cos\phi_2\sin\lambda_2 - \cos\phi_1\sin\phi_2\sin\lambda_1}\right) \tag{5}$$

$$\phi_p = \tan^{-1}\left[-\frac{\cos(\lambda_p - \lambda_1)}{\tan\phi_1}\right], \tag{6}$$

and the inverse FORMULAS are

$$\phi = \sin^{-1}(y\sin\phi_p + \sqrt{1-y^2}\,\cos\phi_p\sin x) \tag{7}$$

$$\lambda = \lambda_0 + \tan^{-1}\left(\frac{\sqrt{1-y^2}\,\sin\phi_p\sin x - y\cos\phi_p}{\sqrt{1-y^2}\,\cos x}\right). \tag{8}$$

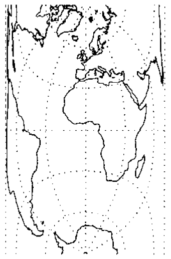

A transverse form of the cylindrical equal-area projection is given by the equations

$$x = \cos\phi\sin(\lambda - \lambda_0) \tag{9}$$

$$y = \tan^{-1}\left[\frac{\tan\phi}{\cos(\lambda - \lambda_0)}\right] - \phi_0, \tag{10}$$

and the inverse FORMULAS are

$$\phi = \sin^{-1}[\sqrt{1-x^2}\sin(y + \phi_0)] \tag{11}$$

$$\lambda = \lambda_0 + \tan^{-1}\left[\frac{x}{\sqrt{1-x^2}}\cos(y + \phi_0)\right]. \tag{12}$$

References

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 76–85, 1987.

Cylindrical Equidistant Projection

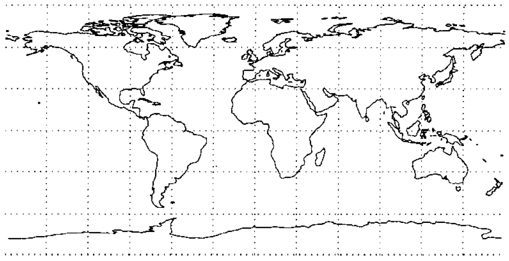

The MAP PROJECTION having transformation equations

$$x = (\lambda - \lambda_0)\cos\phi_1 \tag{1}$$

$$y = \phi, \tag{2}$$

and the inverse FORMULAS are

$$\phi = y \tag{3}$$

$$\lambda = \lambda_0 + \frac{x}{\cos\phi_1}. \tag{4}$$

References

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 90–91, 1987.

Cylindrical Function

$$R_m(x,y) \equiv \frac{J'_m(x)Y'_m(y) - J'_m(y)Y'_m(x)}{J_m(x)Y'_m(y) - J'_m(y)Y_m(x)}$$

$$S_m(x,y) \equiv \frac{J'_m(x)Y_m(y) - J_m(y)Y'_m(x)}{J_m(x)Y_m(y) - J_m(y)Y_m(x)}.$$

see also CYLINDER FUNCTION, HEMISPHERICAL FUNCTION

Cylindrical Harmonics

see Bessel Function of the First Kind

Cylindrical Hoof

see Cylindrical Wedge

Cylindrical Projection

see Behrmann Cylindrical Equal-Area Projection, Cylindrical Equal-Area Projection, Cylindrical Equidistant Projection, Gall's Stereographic Projection, Mercator Projection, Miller Cylindrical Projection, Peters Projection, Pseudocylindrical Projection

Cylindrical Segment

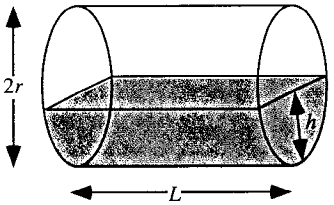

The solid portion of a Cylinder below a cutting Plane which is oriented Parallel to the Cylinder's axis of symmetry. For a Cylinder of Radius r and length L, the Volume of the cylindrical segment is given by multiplying the Area of a circular Segment of height h by L,

$$V = Lr^2\cos^{-1}\left(\frac{r-h}{r}\right) - (r-h)L\sqrt{2rh-h^2}.$$

see also Cylindrical Wedge, Sector, Segment, Spherical Segment

Cylindrical Wedge

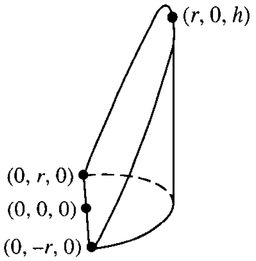

The solid cut from a Cylinder by a tilted Plane passing through a Diameter of the base. It is also called a Cylindrical Hoof. Let the height of the wedge be h and the radius of the Cylinder from which it is cut r. Then plugging the points $(0, -r, 0)$, $(0, r, 0)$, and $(r, 0, h)$ into the 3-point equation for a Plane gives the equation for the plane as

$$hx - rz = 0. \tag{1}$$

Combining with the equation of the Circle which describes the curved part remaining of the cylinder (and writing $t = x$) then gives the parametric equations of the "tongue" of the wedge as

$$x = t \tag{2}$$

$$y = \pm\sqrt{r^2 - t^2} \tag{3}$$

$$z = \frac{ht}{r} \tag{4}$$

for $t \in [0, r]$. To examine the form of the tongue, it needs to be rotated into a convenient plane. This can be accomplished by first rotating the plane of the curve by 90° about the x-Axis using the Rotation Matrix $\mathsf{R}_x(90°)$ and then by the Angle

$$\theta = \tan^{-1}\left(\frac{h}{r}\right) \tag{5}$$

above the z-Axis. The transformed plane now rests in the xz-plane and has parametric equations

$$x = \frac{t\sqrt{h^2 + r^2}}{r} \tag{6}$$

$$z = \pm\sqrt{r^2 - t^2} \tag{7}$$

and is shown below.

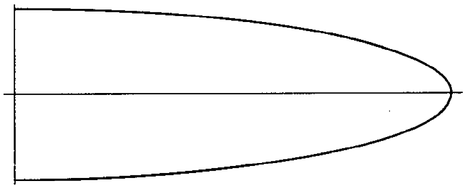

The length of the tongue (measured down its middle) is obtained by plugging $t = r$ into the above equation for x, which becomes

$$L = \sqrt{h^2 + r^2} \tag{8}$$

(and which follows immediately from the Pythagorean Theorem). The Volume of the wedge is given by

$$V = \tfrac{2}{3}r^2h. \tag{9}$$

see also Conical Wedge, Cylindrical Segment

Cylindroid

see Plücker's Conoid

D

d'Alembert's Equation

The ORDINARY DIFFERENTIAL EQUATION

$$y = xf(y') + g(y'),$$

where $y' \equiv dy/dx$ and f and g are given functions.

d'Alembert Ratio Test

see RATIO TEST

d'Alembert's Solution

A method of solving the 1-D WAVE EQUATION.

see also WAVE EQUATION

d'Alembert's Theorem

If three CIRCLES A, B, and C are taken in pairs, the external similarity points of the three pairs lie on a straight line. Similarly, the external similarity point of one pair and the two internal similarity points of the other two pairs lie upon a straight line, forming a similarity axis of the three CIRCLES.

References
Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, p. 155, 1965.

d'Alembertian Operator

Written in the NOTATION of PARTIAL DERIVATIVES,

$$\Box^2 \equiv \nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2},$$

where c is the speed of light. Writing in TENSOR notation

$$\Box^2\phi \equiv (g^{\lambda\kappa}\phi_{;\lambda})_{;\kappa} = g^{\lambda\kappa}\frac{\partial^2\phi}{\partial x^\lambda \partial x^\kappa} - \Gamma^\lambda \frac{\partial\phi}{\partial x^\lambda}.$$

see also HARMONIC COORDINATES

d-Analog

The d-analog of INFINITY FACTORIAL is given by

$$[\infty!]_d = \prod_{n=3}^{\infty}\left(1 - \frac{2^d}{n^d}\right).$$

This INFINITE PRODUCT can be evaluated in closed form for small POSITIVE integral $d \geq 2$.

see also q-ANALOG

References
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/infprd/infprd.html`.

D-Number

A NATURAL NUMBER $n > 3$ such that

$$n|(a^{n-2} - a)$$

whenever $(a, n) = 1$ (a and n are RELATIVELY PRIME) and $a \leq n$. There are an infinite number of such numbers, the first few being 9, 15, 21, 33, 39, 51, ... (Sloane's A033553).

see also KNÖDEL NUMBERS

References
Makowski, A. "Generalization of Morrow's *D*-Numbers." *Simon Stevin* **36**, 71, 1962/1963.
Sloane, N. J. A. Sequence A033553 in "An On-Line Version of the Encyclopedia of Integer Sequences."

D-Statistic

see KOLMOGOROV-SMIRNOV TEST

D-Triangle

Let the circles c_2 and c_3' used in the construction of the BROCARD POINTS which are tangent to A_2A_3 at A_2 and A_3, respectively, meet again at D_1. The points $D_1D_2D_3$ then define the D-triangle. The VERTICES of the D-triangle lie on the respective APOLLONIUS CIRCLES.

see also APOLLONIUS CIRCLES, BROCARD POINTS

References
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 284–285, 296 and 307, 1929.

Daisy

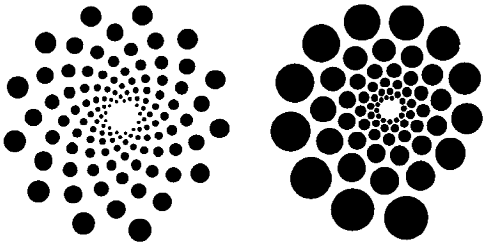

A figure resembling a daisy or sunflower in which copies of a geometric figure of increasing size are placed at regular intervals along a spiral. The resulting figure appears to have multiple spirals spreading out from the center.

see also PHYLLOTAXIS, SPIRAL, SWIRL, WHIRL

References
Dixon, R. "On Drawing a Daisy." §5.1 in *Mathographics.* New York: Dover, pp. 122–143, 1991.

Damped Exponential Cosine Integral

$$\int_0^\infty e^{-\omega T} \cos(\omega t)\, d\omega. \qquad (1)$$

Integrate by parts with

$$u \equiv e^{-\omega T} \qquad dv = \cos(\omega t)\, d\omega \qquad (2)$$

$$du \equiv -Te^{-\omega T}\, d\omega \qquad v = \frac{1}{t}\sin(\omega t), \qquad (3)$$

so

$$\int e^{-\omega T}\cos(\omega t)\, d\omega$$
$$= \frac{1}{t} e^{-\omega t}\sin(\omega t) + \frac{T}{t}\int e^{-\omega T}\sin(\omega t)\, d\omega. \qquad (4)$$

Now integrate

$$\int e^{-\omega T}\sin(\omega t)\, d\omega \qquad (5)$$

by parts. Let

$$v = e^{-\omega T} \qquad dv = \sin(\omega t)\, d\omega \qquad (6)$$

$$du = -Te^{-\omega T}\, d\omega \qquad v = -\frac{1}{t}\cos(\omega t), \qquad (7)$$

so

$$\int e^{-\omega t}\sin(\omega t)\, d\omega = -\frac{1}{t}\cos(\omega t) - \frac{T}{t}\int e^{-\omega T}\cos(\omega t)\, d\omega \qquad (8)$$

and

$$\int e^{\omega T}\cos(\omega t)\, d\omega = \frac{1}{t}e^{-\omega t}\sin(\omega t)$$
$$-\frac{T}{t^2}e^{-\omega t}\cos(\omega t) - \frac{T^2}{t^2}\int e^{-\omega T}\cos(\omega t)\, d\omega \qquad (9)$$

$$\left(1 + \frac{T^2}{t^2}\right)\int e^{-\omega T}\cos(\omega t)\, d\omega$$
$$= e^{-\omega T}\left[\frac{1}{t}\sin(\omega t) - \frac{T}{t^2}\cos(\omega t)\right] \qquad (10)$$

$$\frac{t^2 + T^2}{t^2}\int e^{-\omega T}\cos(\omega t)\, d\omega$$
$$= \frac{e^{-\omega t}}{t^2}[t\sin(\omega T) - T\cos(\omega t)] \qquad (11)$$

$$\int e^{-\omega T}\cos(\omega t)\, d\omega = \frac{e^{-\omega T}}{t^2 + T^2}[t\sin(\omega t) - T\cos(\omega T)]. \qquad (12)$$

Therefore,

$$\int_0^\infty e^{-\omega T}\cos(\omega t)\, d\omega = 0 + \frac{T}{t^2 + T^2} = \frac{T}{t^2 + T^2}. \qquad (13)$$

see also COSINE INTEGRAL, FOURIER TRANSFORM—LORENTZIAN FUNCTION, LORENTZIAN FUNCTION

Dandelin Spheres

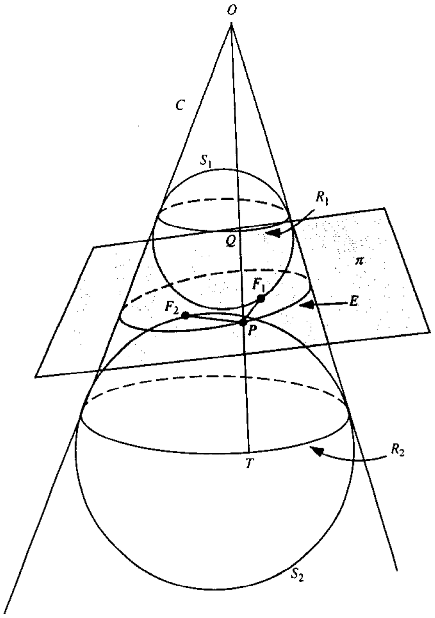

The inner and outer SPHERES TANGENT internally to a CONE and also to a PLANE intersecting the CONE are called Dandelin spheres.

The SPHERES can be used to show that the intersection of the PLANE with the CONE is an ELLIPSE. Let π be a PLANE intersecting a right circular CONE with vertex O in the curve E. Call the SPHERES TANGENT to the CONE and the PLANE S_1 and S_2, and the CIRCLES on which the CIRCLES are TANGENT to the CONE R_1 and R_2. Pick a line along the CONE which intersects R_1 at Q, E at P, and R_2 at T. Call the points on the PLANE where the CIRCLES are TANGENT F_1 and F_2. Because intersecting tangents have the same length,

$$F_1P = QP$$

$$F_2P = TP.$$

Therefore,

$$PF_1 + PF_2 = QP + PT = QT,$$

which is a constant independent of P, so E is an ELLIPSE with $a = QT/2$.

see also CONE, SPHERE

References

Honsberger, R. "Kepler's Conics." Ch. 9 in *Mathematical Plums* (Ed. R. Honsberger). Washington, DC: Math. Assoc. Amer., p. 170, 1979.

Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., pp. 40–44, 1991.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 80–81, 1990.

Ogilvy, C. S. *Excursions in Mathematics.* New York: Dover, pp. 68–69, 1994.

Danielson-Lanczos Lemma

The DISCRETE FOURIER TRANSFORM of length N (where N is EVEN) can be rewritten as the sum of two DISCRETE FOURIER TRANSFORMS, each of length $N/2$. One is formed from the EVEN numbered points; the other from the ODD numbered points. Denote the kth point of the DISCRETE FOURIER TRANSFORM by F_n. Then

$$\begin{aligned} F_n &= \sum_{k=0}^{N-1} f_k e^{-2\pi i n k/N} \\ &= \sum_{k=0}^{N/2-1} e^{-2\pi i k n/(N/2)} f_{2k} + W^n \sum_{k=0}^{N/2-1} e^{-2\pi i k n/(N/2)} f_{2k+1} \\ &= F_n^e + W_n F_n^o, \end{aligned}$$

where $W \equiv e^{-2\pi i/N}$ and $n = 0, \ldots, N$. This procedure can be applied recursively to break up the $N/2$ even and ODD points to their $N/4$ EVEN and ODD points. If N is a POWER of 2, this procedure breaks up the original transform into $\lg N$ transforms of length 1. Each transform of an individual point has $F_n^{eeo\cdots} = f_k$ for some k. By reversing the patterns of evens and odds, then letting $e = 0$ and $o = 1$, the value of k in BINARY is produced. This is the basis for the FAST FOURIER TRANSFORM.

see also DISCRETE FOURIER TRANSFORM, FAST FOURIER TRANSFORM, FOURIER TRANSFORM

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge, England: Cambridge University Press, pp. 407–411, 1989.

Darboux Integral

A variant of the RIEMANN INTEGRAL defined when the UPPER and LOWER INTEGRALS, taken as limits of the LOWER SUM

$$L(f;\phi;N) = \sum_{r=1}^{n} m(f;\delta_r) - \phi(x_{r-1})$$

and UPPER SUM

$$U(f;\phi;N) = \sum_{r=1}^{n} M(f;\delta_r) - \phi(x_{r-1}),$$

are equal. Here, $f(x)$ is a REAL FUNCTION, $\phi(x)$ is a monotonic increasing function with respect to which the sum is taken, $m(f;S)$ denotes the lower bound of $f(x)$ over the interval S, and $M(f;S)$ denotes the upper bound.

see also LOWER INTEGRAL, LOWER SUM, RIEMANN INTEGRAL, UPPER INTEGRAL, UPPER SUM

References

Kestelman, H. *Modern Theories of Integration, 2nd rev. ed.* New York: Dover, p. 250, 1960.

Darboux-Stieltjes Integral

see DARBOUX INTEGRAL

Darboux Vector

The rotation VECTOR of the TRIHEDRON of a curve with CURVATURE $\kappa \neq 0$ when a point moves along a curve with unit SPEED. It is given by

$$\mathbf{D} = \tau\mathbf{T} + \kappa\mathbf{B}, \tag{1}$$

where τ is the TORSION, $\mathbf{T}$ the TANGENT VECTOR, and $\mathbf{B}$ the BINORMAL VECTOR. The Darboux vector field satisfies

$$\dot{\mathbf{T}} = \mathbf{D} \times \mathbf{T} \tag{2}$$

$$\dot{\mathbf{N}} = \mathbf{D} \times \mathbf{N} \tag{3}$$

$$\dot{\mathbf{B}} = \mathbf{D} \times \mathbf{B}. \tag{4}$$

see also BINORMAL VECTOR, CURVATURE, TANGENT VECTOR, TORSION (DIFFERENTIAL GEOMETRY)

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 151, 1993.

Darling's Products

A generalization of the HYPERGEOMETRIC FUNCTION identity

$$\begin{aligned} &{}_2F_1(\alpha,\beta;\gamma;z)\,{}_2F_1(1-\alpha,1-\beta;2-\gamma;z) \\ &= {}_2F_1(\alpha+1-\gamma,\beta+1-\gamma;2-\gamma;z)\,{}_2F_1(\gamma-\alpha,\gamma-\beta;\gamma;z) \end{aligned} \tag{1}$$

to the GENERALIZED HYPERGEOMETRIC FUNCTION ${}_3F_2(a,b,c;d,e;x)$. Darling's products are

$$\begin{aligned} &{}_3F_2\begin{bmatrix} \alpha,\beta,\gamma;z \\ \delta,\epsilon \end{bmatrix} {}_3F_2\begin{bmatrix} 1-\alpha,1-\beta,1-\gamma;z \\ 2-\delta,2-\epsilon \end{bmatrix} \\ &\quad= \frac{\epsilon-1}{\epsilon-\delta}\,{}_3F_2\begin{bmatrix} \alpha+1-\delta,\beta+1-\delta,\gamma+1-\delta;z \\ 2-\delta,\epsilon+1-\delta \end{bmatrix} \\ &\qquad\times {}_3F_2\begin{bmatrix} \delta-\alpha,\delta-\beta,\delta-\gamma;z \\ \delta,\delta+1-\epsilon \end{bmatrix} \\ &\quad+\frac{\delta-1}{\delta-\epsilon}\,{}_3F_2\begin{bmatrix} \alpha+1-\epsilon,\beta+1-\epsilon,\gamma+1-\epsilon;z \\ 2-\epsilon,\delta+1-\epsilon \end{bmatrix} \\ &\qquad\times {}_3F_2\begin{bmatrix} \epsilon-\alpha,\epsilon-\beta,\epsilon-\gamma;z \\ \epsilon,\epsilon+1-\delta \end{bmatrix} \end{aligned} \tag{2}$$

and

$$\begin{aligned} &(1-z)^{\alpha+\beta+\gamma-\delta-\epsilon}\,{}_3F_2\begin{bmatrix} \alpha,\beta,\gamma;z \\ \delta,\epsilon \end{bmatrix} \\ &\quad= \frac{\epsilon-1}{\epsilon-\delta}\,{}_3F_2\begin{bmatrix} \delta-\alpha,\delta-\beta,\delta-\gamma;z \\ \delta,\delta+1-\epsilon \end{bmatrix} \\ &\qquad\times {}_3F_2\begin{bmatrix} \epsilon-\alpha,\epsilon-\beta,\epsilon-\gamma;z \\ \epsilon-1,\epsilon+1-\delta \end{bmatrix} \\ &\quad+\frac{\delta-1}{\delta-\epsilon}\,{}_3F_2\begin{bmatrix} \epsilon-\alpha,\epsilon-\beta,\epsilon-\gamma;z \\ \epsilon,\epsilon+1-\delta \end{bmatrix} \\ &\qquad\times {}_3F_2\begin{bmatrix} \delta-\alpha,\delta-\beta,\delta-\gamma;z \\ \delta-1,\delta+1-\epsilon \end{bmatrix}, \end{aligned} \tag{3}$$

which reduce to (1) when $\gamma = \epsilon \to \infty$.

References

Bailey, W. N. "Darling's Theorems of Products." §10.3 in *Generalised Hypergeometric Series.* Cambridge, England: Cambridge University Press, pp. 88–92, 1935.

Dart

see PENROSE TILES

Darwin-de Sitter Spheroid

A SURFACE OF REVOLUTION of the form

$$r(\phi) = a[1 - e\sin^2\phi - (\tfrac{3}{8}e^2 + k)\sin^2(2\phi)],$$

where k is a second-order correction to the figure of a rotating fluid.

see also OBLATE SPHEROID, PROLATE SPHEROID, SPHEROID

References

Zharkov, V. N. and Trubitsyn, V. P. *Physics of Planetary Interiors.* Tucson, AZ: Pachart Publ. House, 1978.

Darwin's Expansions

Series expansions of the PARABOLIC CYLINDER FUNCTION $U(a,x)$ and $W(a,x)$. The formulas can be found in Abramowitz and Stegun (1972).

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 689–690 and 694–695, 1972.

Data Structure

A formal structure for the organization of information. Examples of data structures include the LIST, QUEUE, STACK, and TREE.

Database

A database can be roughly defined as a structure consisting of

1. A collection of information (the data),
2. A collection of queries that can be submitted, and
3. A collection of algorithms by which the structure responds to queries, searches the data, and returns the results.

References

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, p. 48, 1996.

Daubechies Wavelet Filter

A WAVELET used for filtering signals. Daubechies (1988, p. 980) has tabulated the numerical values up to order $p = 10$.

References

Daubechies, I. "Orthonormal Bases of Compactly Supported Wavelets." *Comm. Pure Appl. Math.* **41**, 909–996, 1988.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Interpolation and Extrapolation." Ch. 3 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 584–586, 1992.

Davenport-Schinzel Sequence

Form a sequence from an ALPHABET of letters $[1, n]$ such that there are no consecutive letters and no alternating subsequences of length greater than d. Then the sequence is a Davenport-Schinzel sequence if it has maximal length $N_d(n)$. The value of $N_1(n)$ is the trivial sequence of 1s: 1, 1, 1, ... (Sloane's A000012). The values of $N_2(n)$ are the POSITIVE INTEGERS 1, 2, 3, 4, ... (Sloane's A000027). The values of $N_3(n)$ are the ODD INTEGERS 1, 3, 5, 7, ... (Sloane's A005408). The first nontrivial Davenport-Schinzel sequence $N_4(n)$ is given by 1, 4, 8, 12, 17, 22, 27, 32, ... (Sloane's A002004). Additional sequences are given by Guy (1994, p. 221) and Sloane.

References

Davenport, H. and Schinzel, A. "A Combinatorial Problem Connected with Differential Equations." *Amer. J. Math.* **87**, 684–690, 1965.

Guy, R. K. "Davenport-Schinzel Sequences." §E20 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 220–222, 1994.

Roselle, D. P. and Stanton, R. G. "Results of Davenport-Schinzel Sequences." In *Proc. Louisiana Conference on Combinatorics, Graph Theory, and Computing. Louisiana State University, Baton Rouge, March 1–5, 1970* (Ed. R. C. Mullin, K. B. Reid, and D. P. Roselle). Winnipeg, Manitoba: Utilitas Mathematica, pp. 249–267, 1960.

Sharir, M. and Agarwal, P. *Davenport-Schinzel Sequences and Their Geometric Applications.* New York: Cambridge University Press, 1995.

Sloane, N. J. A. Sequences A000012/M0003, A000027/M0472, A002004/M3328, and A005408/M2400 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Dawson's Integral

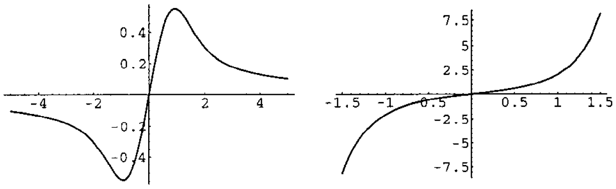

An INTEGRAL which arises in computation of the Voigt lineshape:

$$D(x) \equiv e^{-x^2} \int_0^x e^{y^2}\, dy. \tag{1}$$

It is sometimes generalized such that

$$D_\pm(x) \equiv e^{\mp x^2} \int_0^x e^{\pm y^2}\,dy, \tag{2}$$

giving

$$D_+(x) = \tfrac{1}{2}\sqrt{\pi}\,e^{-x^2}\operatorname{erfi}(x) \tag{3}$$

$$D_-(x) = \tfrac{1}{2}\sqrt{\pi}\,e^{x^2}\operatorname{erf}(x), \tag{4}$$

where $\operatorname{erf}(z)$ is the ERF function and $\operatorname{erfi}(z)$ is the imaginary error function ERFI. $D_+(x)$ is illustrated in the left figure above, and $D_-(x)$ in the right figure. D_+ has a maximum at $D'_+(x) = 0$, or

$$1 - \sqrt{\pi}\,e^{-x^2}x^2\operatorname{erfi}(x) = 0, \tag{5}$$

giving

$$D_+(0.9241388730) = 0.5410442246, \tag{6}$$

and an inflection at $D''_+(x) = 0$, or

$$-2x + \sqrt{\pi}\,e^{-x^2}(2x^2 - 1)\operatorname{erfi}(x) = 0, \tag{7}$$

giving

$$D_+(1.5019752683) = 0.4276866160. \tag{8}$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 298, 1972.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Dawson's Integrals." §6.10 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 252–254, 1992.

Spanier, J. and Oldham, K. B. "Dawson's Integral." Ch. 42 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 405–410, 1987.

Day of Week

see FRIDAY THE THIRTEENTH, WEEKDAY

de Bruijn Constant

Also called the COPSON-DE BRUIJN CONSTANT. It is defined by

$$\sum_{n=1}^{\infty} a_n \leq c \sum_{n=1}^{\infty} \sqrt{\frac{{a_n}^2 + {a_{n+1}}^2 + {a_{n+2}}^2 + \ldots}{n}}.$$

where

$$c = 1.0164957714\ldots.$$

References

Copson, E. T. "Note on Series of Positive Terms." *J. London Math. Soc.* **2**, 9–12, 1927.

Copson, E. T. "Note on Series of Positive Terms." *J. London Math. Soc.* **3**, 49–51, 1928.

de Bruijn, N. G. *Asymptotic Methods in Analysis.* New York: Dover, 1981.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/copson/copson.html`.

de Bruijn Diagram

see DE BRUIJN GRAPH

de Bruijn Graph

A graph whose nodes are sequences of symbols from some ALPHABET and whose edges indicate the sequences which might overlap.

References

Golomb, S. W. *Shift Register Sequences.* San Francisco, CA: Holden-Day, 1967.

Ralston, A. "de Bruijn Sequences—A Model Example of the Interaction of Discrete Mathematics and Computer Science." *Math. Mag.* **55**, 131–143, 1982.

de Bruijn-Newman Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let Ξ be the XI FUNCTION defined by

$$\Xi(iz) = \tfrac{1}{2}(z^2 - \tfrac{1}{4})\pi^{-z/2-\frac{1}{4}}\Gamma(\tfrac{1}{2}z + \tfrac{1}{4})\zeta(z + \tfrac{1}{2}). \tag{1}$$

$\Xi(z/2)/8$ can be viewed as the FOURIER TRANSFORM of the signal

$$\Phi(t) = \sum_{n=1}^{\infty}(2\pi^2 n^4 e^{9t} - 3\pi n^2 e^{5t})e^{-\pi n^2 e^{4t}} \tag{2}$$

for $t \in \mathbb{R} \geq 0$. Then denote the FOURIER TRANSFORM of $\Phi(t)e^{\lambda t^2}$ as $H(\lambda, z)$,

$$\mathcal{F}[\Phi(t)e^{\lambda t^2}] = H(\lambda, z). \tag{3}$$

de Bruijn (1950) proved that H has only REAL zeros for $\lambda \geq 1/2$. C. M. Newman (1976) proved that there exists a constant Λ such that H has only REAL zeros IFF $\lambda \geq \Lambda$. The best current lower bound (Csordas *et al.* 1993, 1994) is $\Lambda > -5.895 \times 10^{-9}$. The RIEMANN HYPOTHESIS is equivalent to the conjecture that $\Lambda \leq 0$.

References

Csordas, G.; Odlyzko, A.; Smith, W.; and Varga, R. S. "A New Lehmer Pair of Zeros and a New Lower Bound for the de Bruijn-Newman Constant." *Elec. Trans. Numer. Analysis* **1**, 104–111, 1993.

Csordas, G.; Smith, W.; and Varga, R. S. "Lehmer Pairs of Zeros, the de Bruijn-Newman Constant and the Riemann Hypothesis." *Constr. Approx.* **10**, 107–129, 1994.

de Bruijn, N. G. "The Roots of Trigonometric Integrals." *Duke Math. J.* **17**, 197–226, 1950.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/dbnwm/dbnwm.html`.

Newman, C. M. "Fourier Transforms with only Real Zeros." *Proc. Amer. Math. Soc.* **61**, 245–251, 1976.

de Bruijn Sequence

The shortest sequence such that every string of length n on the Alphabet a occurs as a contiguous subrange of the sequence described by a. Every de Bruijn sequence corresponds to an Eulerian Cycle on a "de Bruijn Graph." Surprisingly, it turns out that the lexicographic sequence of Lyndon Words of lengths Divisible by n gives the lexicographically smallest de Bruijn sequence (Ruskey).

References

Ruskey, F. "Information on Necklaces, Lyndon Words, de Bruijn Sequences." `http://sue.csc.uvic.ca/~cos/inf/neck/NecklaceInfo.html`.

de Bruijn's Theorem

A box can be packed with a Harmonic Brick $a \times ab \times abc$ Iff the box has dimensions $ap \times abq \times abcr$ for some natural numbers p, q, r (i.e., the box is a multiple of the brick).

see also Box-Packing Theorem, Conway Puzzle, de Bruijn's Theorem, Klarner's Theorem

References

Honsberger, R. *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., pp. 69–72, 1976.

de Jonquières Theorem

The total number of groups of a g_N^r consisting in a point of multiplicity k_1, one of multiplicity k_2, ..., one of multiplicity k_ρ, where

$$\sum k_i = N \tag{1}$$

$$\sum (k_i - 1) = r, \tag{2}$$

and where α_1 points have one multiplicity, α_2 another, etc., and

$$\Pi = k_1 k_2 \cdots k_\rho \tag{3}$$

is

$$\frac{\Pi p(p-1)\cdots(p-\rho)}{\alpha_1!\alpha_2!\cdots} \left[\frac{\Pi}{p-\rho} - \frac{\sum_i \frac{\partial \Pi}{\partial k_i}}{p-\rho+1} + \frac{\sum_{ij} \frac{\partial^2 \Pi}{\partial k_i \partial k_j}}{p-\rho+2} + \ldots\right]. \tag{4}$$

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 288, 1959.

de Jonquières Transformation

A transformation which is of the same type as its inverse. A de Jonquières transformation is always factorable.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, pp. 203–204, 1959.

de la Loubere's Method

A method for constructing Magic Squares of Odd order, also called the Siamese Method.

see also Magic Square

de Longchamps Point

The reflection of the Orthocenter about the Circumcenter. This point is also the Orthocenter of the Anticomplementary Triangle. It has Triangle Center Function

$$\alpha = \cos A - \cos B \cos C.$$

It lies on the Euler Line.

References

Altshiller-Court, N. "On the de Longchamps Circle of the Triangle." *Amer. Math. Monthly* **33**, 368–375, 1926.

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Vandeghen, A. "Soddy's Circles and the de Longchamps Point of a Triangle." *Amer. Math. Monthly* **71**, 176–179, 1964.

de Mere's Problem

The probability of getting at least one "6" in four rolls of a single 6-sided Die is

$$1 - \left(\tfrac{5}{6}\right)^4 = 0.518\ldots, \tag{1}$$

which is slightly higher than the probability of at least one double 6 in 24 throws,

$$1 - \left(\tfrac{35}{36}\right)^{24} = 0.491\ldots. \tag{2}$$

de Mere suspected that (1) was higher than (2). He posed the question to Pascal, who solved the problem and proved de Mere correct.

see also Dice

References

Kraitchik, M. "A Dice Problem." §6.2 in *Mathematical Recreations.* New York: W. W. Norton, pp. 118–119, 1942.

de Moivre's Identity

$$e^{i(n\theta)} = (e^{i\theta})^n. \tag{1}$$

From the Euler Formula it follows that

$$\cos(n\theta) + i\sin(n\theta) = (\cos\theta + i\sin\theta)^n. \tag{2}$$

A similar identity holds for the Hyperbolic Functions,

$$(\cosh z + \sinh z)^n = \cosh(nz) + \sinh(nz). \tag{3}$$

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 356–357, 1985.

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 96–100, 1996.

de Moivre Number

A solution $\zeta_k = e^{2\pi i k/d}$ to the CYCLOTOMIC EQUATION

$$x^d = 1.$$

The de Moivre numbers give the coordinates in the COMPLEX PLANE of the VERTICES of a regular POLYGON with d sides and unit RADIUS.

n	de Moivre Numbers
2	± 1
3	$1, \frac{1}{2}(-1 \pm i\sqrt{3})$
4	$\pm 1, \pm i$
5	$1, \frac{1}{4}\left(-1+\sqrt{5} \pm (1+\sqrt{5})\sqrt{\frac{5-5\sqrt{5}}{2}}\, i\right),$ $-\frac{1+\sqrt{5}}{4} \pm \frac{\sqrt{5-\sqrt{5}}}{2\sqrt{2}}\, i$
6	$\pm 1, \pm\frac{1}{2}(\pm 1 + i\sqrt{3})$

see also CYCLOTOMIC EQUATION, CYCLOTOMIC POLYNOMIAL, EUCLIDEAN NUMBER

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, 1996.

de Moivre-Laplace Theorem

The sum of those terms of the BINOMIAL SERIES of $(p+q)^s$ for which the number of successes x falls between d_1 and d_2 is approximately

$$Q \approx \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-t^2/2}\, dt, \tag{1}$$

where

$$t_1 \equiv \frac{d_1 - \frac{1}{2} - sp}{\sigma} \tag{2}$$

$$t_2 \equiv \frac{d_2 + \frac{1}{2}s - sp}{\sigma} \tag{3}$$

$$\sigma \equiv \sqrt{spq}. \tag{4}$$

Uspensky (1937) has shown that

$$Q = \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-t^2/2}\, dt + \frac{q-p}{6\sigma}\left[(1-t^2)\frac{1}{2\pi}e^{-t^2/2}\right]_{t_1}^{t_2} + \Omega, \tag{5}$$

where

$$|\Omega| < \frac{0.12 + 0.18|p-q|}{\sigma^2} + e^{-3\sigma/2} \tag{6}$$

for $\sigma \geq 5$.

A COROLLARY states that the probability that x successes in s trials will differ from the expected value sp by more than d is

$$P_\delta \approx 1 - 2\int_0^\delta \phi(t)\, dt, \tag{7}$$

where

$$\delta \equiv \frac{d + \frac{1}{2}}{\sigma}. \tag{8}$$

Uspensky (1937) showed that

$$Q_{\delta_1} \equiv P(|x - sp| \leq d) = 2\int_0^{\delta_1} \phi(t)\, dt + \frac{1 - \theta_1 - \theta_2}{\sigma}\phi(\delta_1) + \Omega_1, \tag{9}$$

where

$$\delta_1 \equiv \frac{d}{\delta} \tag{10}$$

$$\theta_1 \equiv (sq + d) - \lfloor sq + d \rfloor \tag{11}$$

$$\theta_2 \equiv (sp + d) - \lfloor sp + d \rfloor \tag{12}$$

and

$$|\Omega_1| < \frac{0.20 + 0.25|p-q|}{\sigma^2} + e^{-3\sigma/2}, \tag{13}$$

for $\sigma \geq 5$.

References

Uspensky, J. V. *Introduction to Mathematical Probability.* New York: McGraw-Hill, 1937.

de Moivre's Quintic

$$x^5 + ax^3 + \tfrac{1}{5}a^2x + b = 0.$$

see also QUINTIC EQUATION

de Morgan's and Bertrand's Test

see BERTRAND'S TEST

de Morgan's Duality Law

For every proposition involving logical addition and multiplication ("or" and "and"), there is a corresponding proposition in which the words "addition" and "multiplication" are interchanged.

de Morgan's Laws

Let $\cup$ represent "or", $\cap$ represent "and", and $'$ represent "not." Then, for two logical units E and F,

$$(E \cup F)' = E' \cap F'$$

$$(E \cap F)' = E' \cup F'.$$

de Polignac's Conjecture

Every EVEN NUMBER is the difference of two consecutive PRIMES in infinitely many ways. If true, taking the difference 2, this conjecture implies that there are infinitely many TWIN PRIMES (Ball and Coxeter 1987). The CONJECTURE has never been proven true or refuted.

see also EVEN NUMBER, TWIN PRIMES

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 64, 1987.
de Polignac, A. "Six propositions arithmologiques déduites de crible d'Ératosthène." *Nouv. Ann. Math.* **8**, 423–429, 1849.

de Rham Cohomology

de Rham cohomology is a formal set-up for the analytic problem: If you have a Differential k-Form ω on a Manifold M, is it the Exterior Derivative of another Differential k-Form ω'? Formally, if $\omega = d\omega'$ then $d\omega = 0$. This is more commonly stated as $d \circ d = 0$, meaning that if ω is to be the Exterior Derivative of a Differential k-Form, a Necessary condition that ω must satisfy is that its Exterior Derivative is zero.

de Rham cohomology gives a formalism that aims to answer the question, "Are all differential k-forms on a Manifold with zero Exterior Derivative the Exterior Derivatives of $(k+1)$-forms?" In particular, the kth de Rham cohomology vector space is defined to be the space of all k-forms with Exterior Derivative 0, modulo the space of all boundaries of $(k+1)$-forms. This is the trivial Vector Space Iff the answer to our question is yes.

The fundamental result about de Rham cohomology is that it is a topological invariant of the Manifold, namely: the kth de Rham cohomology Vector Space of a Manifold M is canonically isomorphic to the Alexander-Spanier Cohomology Vector Space $H^k(M;\mathbb{R})$ (also called cohomology with compact support). In the case that M is compact, Alexander-Spanier Cohomology is exactly singular cohomology.

see also Alexander-Spanier Cohomology, Change of Variables Theorem, Differential k-Form, Exterior Derivative, Vector Space

de Sluze Conchoid

see Conchoid of de Sluze

de Sluze Pearls

see Pearls of Sluze

Debye's Asymptotic Representation

An asymptotic expansion for a Hankel Function of the First Kind

$$\begin{aligned} H_\nu^{(1)}(x) &\sim \frac{1}{\sqrt{\pi}} \exp\{ix[\cos\alpha + (\alpha - \pi/2)\sin\alpha]\} \\ &\quad \times \left[\frac{e^{i\pi/4}}{X} + (\tfrac{1}{8} + \tfrac{5}{24}\tan^2\alpha)\frac{3e^{3\pi i/4}}{2X^3} \right. \\ &\quad \left. + (\tfrac{3}{128} + \tfrac{77}{576}\tan^\alpha + \tfrac{385}{3456}\tan^4\alpha)\frac{3\cdot e^{5\pi i/4}}{2^2X^5} + \ldots \right], \end{aligned}$$

where

$$\frac{\nu}{x} = \sin\alpha,$$

$$1 - \frac{\nu}{x} > \frac{3}{x}\nu^{1/2},$$

and

$$X \equiv \sqrt{-x\cos(\tfrac{1}{2}\alpha)}.$$

see also Hankel Function of the First Kind

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1475, 1980.

Debye Functions

$$\int_0^x \frac{t^n\,dt}{e^t - 1} = x^n \left[\frac{1}{n} - \frac{x}{2(n+1)} + \sum_{k=1}^\infty \frac{B_{2k}x^{2k}}{(2k+n)(2k!)} \right], \tag{1}$$

where $|x| < 2\pi$ and B_n are Bernoulli Numbers.

$$\begin{aligned} \int_x^\infty \frac{t^n\,dt}{e^t - 1} = \sum_{k=1}^\infty e^{-kx} &\left[\frac{x^n}{k} + \frac{nx^{n-1}}{k^2} \right. \\ &\left. + \frac{n(n-1)x^{n-2}}{k^3} + \ldots + \frac{n!}{k^{n+1}} \right], \end{aligned} \tag{2}$$

where $x > 0$. The sum of these two integrals is

$$\int_0^\infty \frac{t^n\,dt}{e^t - 1} = n!\zeta(n+1), \tag{3}$$

where $\zeta(z)$ is the Riemann Zeta Function.

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Debye Functions." §27.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 998, 1972.

Decagon

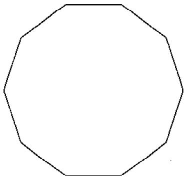

The constructible regular 10-sided Polygon with Schläfli Symbol $\{10\}$. The Inradius r, Circumradius R, and Area can be computed directly from the formulas for a general regular Polygon with side length s and $n = 10$ sides,

$$r = \tfrac{1}{2}s\cot\left(\frac{\pi}{10}\right) = \tfrac{1}{2}\sqrt{25 - 10\sqrt{5}}\,s \tag{1}$$

$$R = \tfrac{1}{2}s\csc\left(\frac{\pi}{10}\right) = \tfrac{1}{2}(1 + \sqrt{5})s = \phi s \tag{2}$$

$$A = \tfrac{1}{4}ns^2\cot\left(\frac{\pi}{10}\right) = \tfrac{5}{2}\sqrt{5 + 2\sqrt{5}}\,s^2. \tag{3}$$

Here, ϕ is the Golden Mean.

see also Decagram, Dodecagon, Trigonometry Values—$\pi/10$, Undecagon

References

Dixon, R. *Mathographics.* New York: Dover, p. 18, 1991.

Decagonal Number

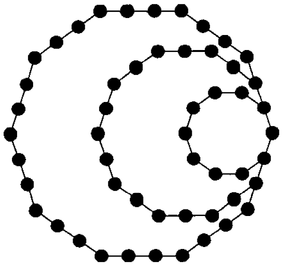

A FIGURATE NUMBER of the form $4n^2 - 3n$. The first few are 1, 10, 27, 52, 85, ... (Sloane's A001107). The GENERATING FUNCTION giving the decagonal numbers is

$$\frac{x(7x+1)}{(1-x)^3} = x + 10x^2 + 27x^3 + 52x^4 + \ldots.$$

References

Sloane, N. J. A. Sequence A001107/M4690 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Decagram

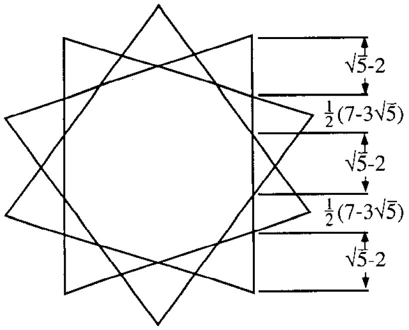

The STAR POLYGON $\left\{ \begin{smallmatrix} 10 \\ 3 \end{smallmatrix} \right\}$.

see also DECAGON, STAR POLYGON

Decic Surface

A SURFACE which can be represented implicitly by a POLYNOMIAL of degree 10 in x, y, and z. An example is the BARTH DECIC.

see also BARTH DECIC, CUBIC SURFACE, QUADRATIC SURFACE, QUARTIC SURFACE

Decidable

A "theory" in LOGIC is decidable if there is an ALGORITHM that will decide on input ϕ whether or not ϕ is a SENTENCE true of the FIELD of REAL NUMBERS $\mathbb{R}$.

see also CHURCH'S THESIS, GÖDEL'S COMPLETENESS THEOREM, GÖDEL'S INCOMPLETENESS THEOREM, KREISEL CONJECTURE, TARSKI'S THEOREM, UNDECIDABLE, UNIVERSAL STATEMENT

References

Kemeny, J. G. "Undecidable Problems of Elementary Number Theory." *Math. Ann.* **135**, 160–169, 1958.

Decillion

In the American system, 10^{33}.

see also LARGE NUMBER

Decimal

The base 10 notational system for representing REAL NUMBERS.

see also 10, BASE (NUMBER), BINARY, HEXADECIMAL, OCTAL

References

Pappas, T. "The Evolution of Base Ten." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 2–3, 1989.

Weisstein, E. W. "Bases." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Bases.m`.

Decimal Expansion

The decimal expansion of a number is its representation in base 10. For example, the decimal expansion of 25^2 is 625, of π is 3.14159..., and of 1/9 is 0.1111....

If $r \equiv p/q$ has a finite decimal expansion, then

$$\begin{aligned} r &= \frac{a_1}{10} + \frac{a_2}{10^2} + \ldots + \frac{a_n}{10^n} \\ &= \frac{a_1 10^{n-1} + a_2 10^{n-2} + \ldots + a_n}{10^n} \\ &= \frac{a_1 10^{n-1} + a_2 10^{n-2} + \ldots + a_n}{2^n 5^n}. \end{aligned} \tag{1}$$

FACTORING possible common multiples gives

$$r = \frac{p}{2^\alpha 5^\beta}, \tag{2}$$

where $p \not\equiv 0 \pmod{2, 5}$. Therefore, the numbers with finite decimal expansions are fractions of this form. The number of decimals is given by $\max(\alpha, \beta)$. Numbers which have a finite decimal expansion are called REGULAR NUMBERS.

Any NONREGULAR fraction m/n is periodic, and has a period $\lambda(n)$ independent of m, which is at most $n-1$ DIGITS long. If n is RELATIVELY PRIME to 10, then the period of m/n is a divisor of $\phi(n)$ and has at most $\phi(n)$ DIGITS, where ϕ is the TOTIENT FUNCTION. When a rational number m/n with $(m,n) = 1$ is expanded, the period begins after s terms and has length t, where s and t are the smallest numbers satisfying

$$10^2 \equiv 10^{s+t} \pmod{n}. \tag{3}$$

When $n \not\equiv 0 \pmod{2, 5}$, $s = 0$, and this becomes a purely periodic decimal with

$$10^t \equiv 1 \pmod{n}. \tag{4}$$

As an example, consider $n = 84$.

$$\begin{array}{llll} 10^0 \equiv 1 & 10^1 \equiv 10 & 10^2 \equiv 16 & 10^3 \equiv -8 \\ 10^4 \equiv 4 & 10^5 \equiv 40 & 10^6 \equiv -20 & 10^7 \equiv -32, \\ 10^8 \equiv 16 & & & \end{array}$$

so $s = 2$, $t = 6$. The decimal representation is $1/84 = 0.01\overline{190476}$. When the DENOMINATOR of a fraction m/n has the form $n = n_0 2^\alpha 5^\beta$ with $(n_0, 10) = 1$, then the period begins after $\max(\alpha, \beta)$ terms and the length of the period is the exponent to which 10 belongs (mod n_0), i.e., the number x such that $10^x \equiv 1 \pmod{n_0}$. If q is PRIME and $\lambda(q)$ is EVEN, then breaking the repeating DIGITS into two equal halves and adding gives all 9s. For example, $1/7 = 0.\overline{142857}$, and $142 + 857 = 999$. For $1/q$ with a PRIME DENOMINATOR other than 2 or 5, all cycles n/q have the same length (Conway and Guy 1996).

If n is a PRIME and 10 is a PRIMITIVE ROOT of n, then the period $\lambda(n)$ of the repeating decimal $1/n$ is given by

$$\lambda(n) = \phi(n), \tag{5}$$

where $\phi(n)$ is the TOTIENT FUNCTION. Furthermore, the decimal expansions for p/n, with $p = 1, 2, \ldots, n-1$ have periods of length $n-1$ and differ only by a cyclic permutation. Such numbers are called LONG PRIMES by Conway and Guy (1996). An equivalent definition is that

$$10^i \equiv 1 \pmod{n} \tag{6}$$

for $i = n - 1$ and no i less than this. In other words, a NECESSARY (but not SUFFICIENT) condition is that the number $9R_{n-1}$ (where R_n is a REPUNIT) is DIVISIBLE by n, which means that R_n is DIVISIBLE by n.

The first few numbers with maximal decimal expansions, called FULL REPTEND PRIMES, are 7, 17, 19, 23, 29, 47, 59, 61, 97, 109, 113, 131, 149, 167, ... (Sloane's A001913). The decimals corresponding to these are called CYCLIC NUMBERS. No general method is known for finding FULL REPTEND PRIMES. Artin conjectured that ARTIN'S CONSTANT $C = 0.3739558136\ldots$ is the fraction of PRIMES p for with $1/p$ has decimal maximal period (Conway and Guy 1996). D. Lehmer has generalized this conjecture to other bases, obtaining values which are small rational multiples of C.

To find DENOMINATORS with short periods, note that

$$\begin{aligned}
10^1 - 1 &= 3^2 \\
10^2 - 1 &= 3^2 \cdot 11 \\
10^3 - 1 &= 3^3 \cdot 37 \\
10^4 - 1 &= 3^2 \cdot 11 \cdot 101 \\
10^5 - 1 &= 3^2 \cdot 41 \cdot 271 \\
10^6 - 1 &= 3^3 \cdot 7 \cdot 11 \cdot 13 \cdot 37 \\
10^7 - 1 &= 3^2 \cdot 239 \cdot 4649 \\
10^8 - 1 &= 3^2 \cdot 11 \cdot 73 \cdot 101 \cdot 137 \\
10^9 - 1 &= 3^4 \cdot 37 \cdot 333667 \\
10^{10} - 1 &= 3^2 \cdot 11 \cdot 41 \cdot 271 \cdot 9091 \\
10^{11} - 1 &= 3^2 \cdot 21649 \cdot 513239 \\
10^{12} - 1 &= 3^3 \cdot 7 \cdot 11 \cdot 13 \cdot 37 \cdot 101 \cdot 9901.
\end{aligned}$$

The period of a fraction with DENOMINATOR equal to a PRIME FACTOR above is therefore the POWER of 10 in which the factor first appears. For example, 37 appears in the factorization of $10^3 - 1$ and $10^9 - 1$, so its period is 3. Multiplication of any FACTOR by a $2^\alpha 5^\beta$ still gives the same period as the FACTOR alone. A DENOMINATOR obtained by a multiplication of two FACTORS has a period equal to the first POWER of 10 in which both FACTORS appear. The following table gives the PRIMES having small periods (Sloane's A046106, A046107, and A046108; Ogilvy and Anderson 1988).

period	primes
1	3
2	11
3	37
4	101
5	41, 271
6	7, 13
7	239, 4649
8	73, 137
9	333667
10	9091
11	21649, 513239
12	9901
13	53, 79, 265371653
14	909091
15	31, 2906161
16	17, 5882353
17	2071723, 5363222357
18	19, 52579
19	1111111111111111111
20	3541, 27961

A table of the periods e of small PRIMES other than the special $p = 5$, for which the decimal expansion is not periodic, follows (Sloane's A002371).

p	e	p	e	p	e
3	1	31	15	67	33
7	6	37	3	71	35
11	2	41	5	73	8
13	6	43	21	79	13
17	16	47	46	83	41
19	18	53	13	89	44
23	22	59	58	97	96
29	28	61	60	101	4

Shanks (1873ab) computed the periods for all PRIMES up to 120,000 and published those up to 29,989.

see also FRACTION, MIDY'S THEOREM, REPEATING DECIMAL

References

Conway, J. H. and Guy, R. K. "Fractions Cycle into Decimals." In *The Book of Numbers.* New York: Springer-Verlag, pp. 157–163 and 166–171, 1996.

Das, R. C. "On Bose Numbers." *Amer. Math. Monthly* **56**, 87–89, 1949.

Dickson, L. E. *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, pp. 159–179, 1952.

Lehmer, D. H. "A Note on Primitive Roots." *Scripta Math.* **26**, 117–119, 1963.
Ogilvy, C. S. and Anderson, J. T. *Excursions in Number Theory.* New York: Dover, p. 60, 1988.
Rademacher, H. and Toeplitz, O. *The Enjoyment of Mathematics: Selections from Mathematics for the Amateur.* Princeton, NJ: Princeton University Press, pp. 147–163, 1957.
Rao, K. S. "A Note on the Recurring Period of the Reciprocal of an Odd Number." *Amer. Math. Monthly* **62**, 484–487, 1955.
Shanks, W. "On the Number of Figures in the Period of the Reciprocal of Every Prime Number Below 20,000." *Proc. Roy. Soc. London* **22**, 200, 1873a.
Shanks, W. "On the Number of Figures in the Period of the Reciprocal of Every Prime Number Between 20,000 and 30,000." *Proc. Roy. Soc. London* **22**, 384, 1873b.
Shiller, J. K. "A Theorem in the Decimal Representation of Rationals." *Amer. Math. Monthly* **66**, 797–798, 1959.
Sloane, N. J. A. Sequences A001913/M4353 and A002371/M4050 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Decimal Period

see DECIMAL EXPANSION

Decision Problem

Does there exist an ALGORITHM for deciding whether or not a specific mathematical assertion does or does not have a proof? The decision problem is also known as the ENTSCHEIDUNGSPROBLEM (which, not so coincidentally, is German for "decision problem"). Using the concept of the TURING MACHINE, Turing showed the answer to be NEGATIVE for elementary NUMBER THEORY. J. Robinson and Tarski showed the decision problem is undecidable for arbitrary FIELDS.

Decision Theory

A branch of GAME THEORY dealing with strategies to maximize the outcome of a given process in the face of uncertain conditions.

see also NEWCOMB'S PARADOX, OPERATIONS RESEARCH, PRISONER'S DILEMMA

Decomposition

A rewriting of a given quantity (e.g., a MATRIX) in terms of a combination of "simpler" quantities.

see also CHOLESKY DECOMPOSITION, CONNECTED SUM DECOMPOSITION, JACO-SHALEN-JOHANNSON TORUS DECOMPOSITION, LU DECOMPOSITION, QR DECOMPOSITION, SINGULAR VALUE DECOMPOSITION

Deconvolution

The inversion of a CONVOLUTION equation, i.e., the solution for f of an equation of the form

$$f * g = h + \epsilon,$$

given g and h, where ϵ is the NOISE and $*$ denotes the CONVOLUTION. Deconvolution is ill-posed and will usually not have a unique solution even in the absence of NOISE.

Linear deconvolution ALGORITHMS include INVERSE FILTERING and WIENER FILTERING. Nonlinear ALGORITHMS include the CLEAN ALGORITHM, MAXIMUM ENTROPY METHOD, and LUCY.

see also CLEAN ALGORITHM, CONVOLUTION, LUCY, MAXIMUM ENTROPY METHOD, WIENER FILTER

References

Cornwell, T. and Braun, R. "Deconvolution." Ch. 8 in *Synthesis Imaging in Radio Astronomy: Third NRAO Summer School, 1988* (Ed. R. A. Perley, F. R. Schwab, and A. H. Bridle). San Francisco, CA: Astronomical Society of the Pacific, pp. 167–183, 1989.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Convolution and Deconvolution Using the FFT." §13.1 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 531–537, 1992.

Decreasing Function

A function $f(x)$ decreases on an INTERVAL I if $f(b) < f(a)$ for all $b > a$, where $a, b \in I$. Conversely, a function $f(x)$ increases on an INTERVAL I if $f(b) > f(a)$ for all $b > a$ with $a, b \in I$.

If the DERIVATIVE $f'(x)$ of a CONTINUOUS FUNCTION $f(x)$ satisfies $f'(x) < 0$ on an OPEN INTERVAL (a, b), then $f(x)$ is decreasing on (a, b). However, a function may decrease on an interval without having a derivative defined at all points. For example, the function $-x^{1/3}$ is decreasing everywhere, including the origin $x = 0$, despite the fact that the DERIVATIVE is not defined at that point.

see also DERIVATIVE, INCREASING FUNCTION, NONDECREASING FUNCTION, NONINCREASING FUNCTION

Decreasing Sequence

A SEQUENCE $\{a_1, a_2, \ldots\}$ for which $a_1 \geq a_2 \geq \ldots$.

see also INCREASING SEQUENCE

Decreasing Series

A SERIES $s_1, s_2, \ldots$ for which $s_1 \geq s_2 \geq \ldots$.

Dedekind's Axiom

For every partition of all the points on a line into two nonempty SETS such that no point of either lies between two points of the other, there is a point of one SET which lies between every other point of that SET and every point of the other SET.

Dedekind Cut

A set partition of the RATIONAL NUMBERS into two nonempty subsets S_1 and S_2 such that all members of S_1 are less than those of S_2 and such that S_1 has no greatest member. REAL NUMBERS can be defined using either Dedekind cuts or CAUCHY SEQUENCES.

see also CANTOR-DEDEKIND AXIOM, CAUCHY SEQUENCE

References

Courant, R. and Robbins, H. "Alternative Methods of Defining Irrational Numbers. Dedekind Cuts." §2.2.6 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 71–72, 1996.

Dedekind Eta Function

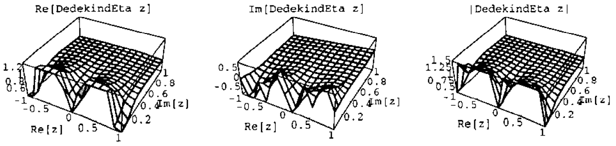

Let

$$q = e^{2\pi iz}, \tag{1}$$

then the Dedekind eta function is defined by

$$\eta(z) \equiv q^{1/24} \prod_{n=1}^{\infty} (1 - q^n), \tag{2}$$

which can be written as

$$\eta(z) = q^{1/24} \left\{ 1 + \sum_{n=1}^{\infty} (-1)^n [q^{n(3n-1)/2} + q^{n(3n+1)/2}] \right\} \tag{3}$$

(Weber 1902, pp. 85 and 112; Atkin and Morain 1993). η is a MODULAR FORM. Letting $\zeta_{24} = 2^{2\pi i/24}$ be a ROOT OF UNITY, $\eta(z)$ satisfies

$$\eta(z+1) = \zeta_{24}\eta(z) \tag{4}$$

$$\eta\left(-\frac{1}{z}\right) = -\sqrt{zi}\,\eta(z) \tag{5}$$

(Weber 1902, p. 113; Atkin and Morain 1993).

see also DIRICHLET ETA FUNCTION, THETA FUNCTION, WEBER FUNCTIONS

References

Atkin, A. O. L. and Morain, F. "Elliptic Curves and Primality Proving." *Math. Comput.* **61**, 29–68, 1993.

Weber, H. *Lehrbuch der Algebra, Vols. I–II.* New York: Chelsea, 1902.

Dedekind Function

$$\psi(n) = n \prod_{\substack{\text{distinct prime} \\ \text{factors } p \text{ of } n}} (1 + p^{-1}),$$

where the PRODUCT is over the distinct PRIME FACTORS of n. The first few values are 1, 3, 4, 6, 6, 12, 8, 12, 12, 18, ... (Sloane's A001615).

see also DEDEKIND ETA FUNCTION, EULER PRODUCT, TOTIENT FUNCTION

References

Cox, D. A. *Primes of the Form x^2+ny^2: Fermat, Class Field Theory and Complex Multiplication.* New York: Wiley, p. 228, 1997.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 96, 1994.

Sloane, N. J. A. Sequence A001615/M2315 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Dedekind's Problem

The determination of the number of monotone BOOLEAN FUNCTIONS of n variables is called Dedekind's problem.

Dedekind Ring

A abstract commutative RING in which every NONZERO IDEAL is a unique product of PRIME IDEALS.

Dedekind Sum

Given RELATIVELY PRIME INTEGERS p and q, the Dedekind sum is defined by

$$s(p,q) \equiv \sum_{i=1}^{q} \left(\left(\frac{i}{q}\right)\right) \left(\left(\frac{pi}{q}\right)\right), \tag{1}$$

where

$$((x)) \equiv \begin{cases} x - \lfloor x \rfloor - \frac{1}{2} & x \in \mathbb{Z} \\ 0 & x \notin \mathbb{Z}. \end{cases} \tag{2}$$

Dedekind sums obey 2-term

$$s(p,q) + s(q,p) = -\frac{1}{4} + \frac{1}{12}\left(\frac{p}{q} + \frac{q}{p} + \frac{1}{pq}\right), \tag{3}$$

and 3-term

$$s(bc',a)+s(ca',b)+s(ab',c) = -\frac{1}{4}+\frac{1}{12}\left(\frac{a}{bc} + \frac{b}{ca} + \frac{c}{ab}\right) \tag{4}$$

reciprocity laws, where a, b, c are pairwise COPRIME and

$$aa' \equiv 1 \pmod{b} \tag{5}$$

$$bb' \equiv 1 \pmod{c} \tag{6}$$

$$cc' \equiv 1 \pmod{a}. \tag{7}$$

Let p, q, u, $v \in \mathbb{N}$ with $(p,q) = (u,v) = 1$ (i.e., are pairwise RELATIVELY PRIME), then the Dedekind sums also satisfy

$$s(p,q) + s(u,v) = s(pu' - qv', pv + qu) - \tfrac{1}{4} + \frac{1}{12}\left(\frac{q}{vt} + \frac{v}{tq} + \frac{t}{qv}\right), \quad (8)$$

where $t = pv + qu$, and u', v' are any INTEGERS such that $uu' + vv' = 1$ (Pommersheim 1993).

References

Pommersheim, J. "Toric Varieties, Lattice Points, and Dedekind Sums." *Math. Ann.* **295**, 1–24, 1993.

Deducible

If q is logically deducible from p, this is written $p \vdash q$.

Deep Theorem

Qualitatively, a deep theorem is a theorem whose proof is long, complicated, difficult, or appears to involve branches of mathematics which are not obviously related to the theorem itself (Shanks 1993). Shanks (1993) cites the QUADRATIC RECIPROCITY THEOREM as an example of a deep theorem.

see also THEOREM

References

Shanks, D. "Is the Quadratic Reciprocity Law a Deep Theorem?" §2.25 in *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 64–66, 1993.

Defective Matrix

A MATRIX whose EIGENVECTORS are not COMPLETE.

Defective Number

see DEFICIENT NUMBER

Deficiency

The deficiency of a BINOMIAL COEFFICIENT $\binom{n+k}{k}$ with $k \leq n$ as the number of i for which $b_i = 1$, where

$$n + i = a_i b_i,$$

$1 \leq i \leq k$, the PRIME factors of b_i are $> k$, and $\prod a_i = k!$, where $k1$ is the FACTORIAL.

see also ABUNDANCE

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 84–85, 1994.

Deficient Number

Numbers which are not PERFECT and for which

$$s(N) \equiv \sigma(N) - N < N,$$

or equivalently

$$\sigma(n) < 2n,$$

where $\sigma(N)$ is the DIVISOR FUNCTION. Deficient numbers are sometimes called DEFECTIVE NUMBERS (Singh 1997). PRIMES, POWERS of PRIMES, and any divisors of a PERFECT or deficient number are all deficient. The first few deficient numbers are 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 19, 21, 22, 23, ... (Sloane's A002855).

see also ABUNDANT NUMBER, LEAST DEFICIENT NUMBER, PERFECT NUMBER

References

Dickson, L. E. *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, pp. 3–33, 1952.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 45, 1994.

Singh, S. *Fermat's Enigma: The Epic Quest to Solve the World's Greatest Mathematical Problem.* New York: Walker, p. 11, 1997.

Sloane, N. J. A. Sequence A002855/M0514 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Definable Set

An ANALYTIC, BOREL, or COANALYTIC SET.

Defined

If A and B are equal by definition (i.e., A is defined as B), then this is written symbolically as $A \equiv B$ or $A := B$.

Definite Integral

An INTEGRAL

$$\int_a^b f(x)\,dx$$

with upper and lower limits. The first FUNDAMENTAL THEOREM OF CALCULUS allows definite integrals to be computed in terms of INDEFINITE INTEGRALS, since if F is the INDEFINITE INTEGRAL for $f(x)$, then

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

see also CALCULUS, FUNDAMENTAL THEOREMS OF CALCULUS, INDEFINITE INTEGRAL, INTEGRAL

Degenerate

A limiting case in which a class of object changes its nature so as to belong to another, usually simpler, class. For example, the POINT is a degenerate case of the CIRCLE as the RADIUS approaches 0, and the CIRCLE is a degenerate form of an ELLIPSE as the ECCENTRICITY approaches 0. Another example is the two identical ROOTS of the second-order POLYNOMIAL $(x-1)^2$. Since the n ROOTS of an nth degree POLYNOMIAL are usually distinct, ROOTS which coincide are said to be degenerate. Degenerate cases often require special treatment in numerical and analytical solutions. For example, a simple search for both ROOTS of the above equation would find only a single one: 1

The word degenerate also has several very specific and technical meanings in different branches of mathematics.

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 513–514, 1985.

Degree

The word "degree" has many meanings in mathematics. The most common meaning is the unit of ANGLE measure defined such that an entire rotation is 360°. This unit harks back to the Babylonians, who used a base 60 number system. 360° likely arises from the Babylonian year, which was composed of 360 days (12 months of 30 days each). The degree is subdivided into 60 MINUTES per degree, and 60 SECONDS per MINUTE.

see also ARC MINUTE, ARC SECOND, DEGREE OF FREEDOM, DEGREE (MAP), DEGREE (POLYNOMIAL), DEGREE (VERTEX), INDEGREE, LOCAL DEGREE, OUTDEGREE

Degree (Algebraic Surface)

see ORDER (ALGEBRAIC SURFACE)

Degree of Freedom

The number of degrees of freedom in a problem, distribution, etc., is the number of parameters which may be independently varied.

see also LIKELIHOOD RATIO

Degree (Map)

Let $f : M \mapsto N$ be a MAP between two compact, connected, oriented n-D MANIFOLDS without boundary. Then f induces a HOMEOMORPHISM f_* from the HOMOLOGY GROUPS $H_n(M)$ to $H_n(N)$, both canonically isomorphic to the INTEGERS, and so f_* can be thought of as a HOMEOMORPHISM of the INTEGERS. The INTEGER $d(f)$ to which the number 1 gets sent is called the degree of the MAP f.

There is an easy way to compute $d(f)$ if the MANIFOLDS involved are smooth. Let $x \in \mathbb{N}$, and approximate f by a smooth map HOMOTOPIC to f such that x is a "regular value" of f (which exist and are everywhere by SARD'S THEOREM). By the IMPLICIT FUNCTION THEOREM, each point in $f^{-1}(x)$ has a NEIGHBORHOOD such that f restricted to it is a DIFFEOMORPHISM. If the DIFFEOMORPHISM is orientation preserving, assign it the number +1, and if it is orientation reversing, assign it the number −1. Add up all the numbers for all the points in $f^{-1}(x)$, and that is the $d(f)$, the degree of f. One reason why the degree of a map is important is because it is a HOMOTOPY invariant. A sharper result states that two self-maps of the n-sphere are homotopic IFF they have the same degree. This is equivalent to the result that the nth HOMOTOPY GROUP of the n-SPHERE is the set $\mathbb{Z}$ of INTEGERS. The ISOMORPHISM is given by taking the degree of any representation.

One important application of the degree concept is that homotopy classes of maps from n-spheres to n-spheres are classified by their degree (there is exactly one homotopy class of maps for every INTEGER n, and n is the degree of those maps).

Degree (Polynomial)

see ORDER (POLYNOMIAL)

Degree Sequence

Given an (undirected) GRAPH, a degree sequence is a monotonic nonincreasing sequence of the degrees of its VERTICES. A degree sequence is said to be k-connected if there exists some k-CONNECTED GRAPH corresponding to the degree sequence. For example, while the degree sequence $\{1, 2, 1\}$ is 1- but not 2-connected, $\{2, 2, 2\}$ is 2-connected. The number of degree sequences for $n = 1, 2, \ldots$ is given by 1, 2, 4, 11, 31, 102, ... (Sloane's A004251).

see also GRAPHICAL PARTITION

References

Ruskey, F. "Information on Degree Sequences." `http://sue.csc.uvic.ca/~cos/inf/nump/DegreeSequences.html`.

Ruskey, F.; Cohen, R.; Eades, P.; and Scott, A. "Alley CATs in Search of Good Homes." *Congres. Numer.* **102**, 97–110, 1994.

Sloane, N. J. A. Sequence A004251/M1250 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Degree (Vertex)

see VERTEX DEGREE

Dehn Invariant

An invariant defined using the angles of a 3-D POLYHEDRON. It remains constant under solid DISSECTION and reassembly. However, solids with the same volume can have different Dehn invariants. Two POLYHEDRA can be dissected into each other only if they have the same volume and the same Dehn invariant.

see also DISSECTION, EHRHART POLYNOMIAL

Dehn's Lemma

If you have an embedding of a 1-SPHERE in a 3-MANIFOLD which exists continuously over the 2-DISK, then it also extends over the DISK as an embedding. It was proposed by Dehn in 1910, but a correct proof was not obtained until the work of Papakyriakopoulos (1957ab).

References
Hempel, J. *3-Manifolds.* Princeton, NJ: Princeton University Press, 1976.
Papakyriakopoulos, C. D. "On Dehn's Lemma and the Asphericity of Knots." *Proc. Nat. Acad. Sci. USA* **43**, 169–172, 1957a.
Papakyriakopoulos, C. D. "On Dehn's Lemma and the Asphericity of Knots." *Ann. Math.* **66**, 1–26, 1957.
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 100–101, 1976.

Dehn Surgery

The operation of drilling a tubular NEIGHBORHOOD of a KNOT K in $\mathbb{S}^3$ and then gluing in a solid TORUS so that its meridian curve goes to a (p, q)-curve on the TORUS boundary of the KNOT exterior. Every compact connected 3-MANIFOLD comes from Dehn surgery on a LINK in $\mathbb{S}^3$.

see also KIRBY CALCULUS

References
Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, p. 260, 1994.

Del

see GRADIENT

Del Pezzo Surface

A SURFACE which is related to CAYLEY NUMBERS.

References
Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, p. 211, 1973.
Hunt, B. "Del Pezzo Surfaces." §4.1.4 in *The Geometry of Some Special Arithmetic Quotients.* New York: Springer-Verlag, pp. 128–129, 1996.

Delannoy Number

The Delannoy numbers are defined by

$$D(a,b) = D(a-1,b) + D(a,b-1) + D(a-1,b-1),$$

where $D(0,0) = 1$. They are the number of lattice paths from $(0,0)$ to (b,a) in which only east (1, 0), north (0, 1), and northeast (1, 1) steps are allowed (i.e, →, ↑, and ↗).

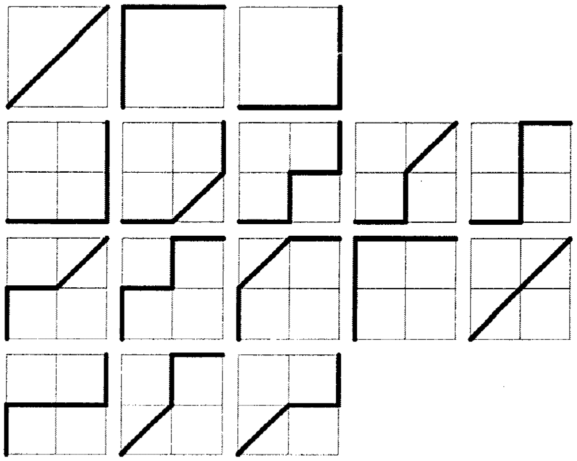

For $n \equiv a = b$, the Delannoy numbers are the number of "king walks"

$$D(n,n) = P_n(3),$$

where $P_n(x)$ is a LEGENDRE POLYNOMIAL (Moser 1955, Vardi 1991). Another expression is

$$D(n,n) = \sum_{k=0}^{n} \binom{n}{k} \binom{n+k}{k} = {}_2F_1(-n, n+1; 1, -1),$$

where $\binom{a}{b}$ is a BINOMIAL COEFFICIENT and ${}_2F_1(a,b;c;z)$ is a HYPERGEOMETRIC FUNCTION. The values of $D(n,n)$ for $n = 1, 2, \ldots$ are 3, 13, 63, 321, 1683, 8989, 48639, ... (Sloane's A001850).

The SCHRÖDER NUMBERS bear the same relation to the Delannoy numbers as the CATALAN NUMBERS do to the BINOMIAL COEFFICIENTS.

see also BINOMIAL COEFFICIENT, CATALAN NUMBER, MOTZKIN NUMBER, SCHRÖDER NUMBER

References
Sloane, N. J. A. Sequence A001850/M2942 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Delaunay Triangulation

The NERVE of the cells in a VORONOI DIAGRAM, which is the triangular of the CONVEX HULL of the points in the diagram. The Delaunay triangulation of a VORONOI DIAGRAM in $\mathbb{R}^2$ is the diagram's planar dual.

see also TRIANGULATION

Delian Constant

The number $2^{1/3}$ (the CUBE ROOT of 2) which is to be constructed in the CUBE DUPLICATION problem. This number is not a EUCLIDEAN NUMBER although it is an ALGEBRAIC of third degree.

References
Conway, J. H. and Guy, R. K. "Three Greek Problems." In *The Book of Numbers.* New York: Springer-Verlag, pp. 192–194, 1996.

Delian Problem

see CUBE DUPLICATION

Delta Amplitude

Given an AMPLITUDE ϕ and a MODULUS m in an ELLIPTIC INTEGRAL,

$$\Delta(\phi) \equiv \sqrt{1 - m \sin^2 \phi}.$$

see also AMPLITUDE, ELLIPTIC INTEGRAL, MODULUS (ELLIPTIC INTEGRAL)

Delta Curve

A curve which can be turned continuously inside an EQUILATERAL TRIANGLE. There are an infinite number of delta curves, but the simplest are the CIRCLE and lens-shaped Δ-biangle. All the Δ curves of height h have the same PERIMETER $2\pi h/3$. Also, at each position of a Δ curve turning in an EQUILATERAL TRIANGLE, the perpendiculars to the sides at the points of contact are CONCURRENT at the instantaneous center of rotation.

see also REULEAUX TRIANGLE

References

Honsberger, R. *Mathematical Gems I.* Washington, DC: Math. Assoc. Amer., pp. 56–59, 1973.

Delta Function

Defined as the limit of a class of DELTA SEQUENCES. Sometimes called the IMPULSE SYMBOL. The most commonly used (equivalent) definitions are

$$\delta(x) \equiv \lim_{n\to\infty} \frac{1}{2\pi} \frac{\sin[(n+\frac{1}{2})x]}{\sin(\frac{1}{2}x)} \tag{1}$$

(the so-called DIRICHLET KERNEL) and

$$\begin{aligned} \delta(x) &\equiv \lim_{n\to\infty} \frac{\sin(nx)}{\pi x} & (2)\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ikx}\,dk & (3)\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ikx}\,dk & (4)\\ &= \mathcal{F}[1], & (5) \end{aligned}$$

where $\mathcal{F}$ is the FOURIER TRANSFORM. Some identities include

$$\delta(x-a) = 0 \tag{6}$$

for $x \neq a$,

$$\int_{a-\epsilon}^{a+\epsilon} \delta(x-a)\,dx = 1, \tag{7}$$

where ϵ is any POSITIVE number, and

$$\int_{-\infty}^{\infty} f(x)\delta(x-a)\,dx = f(a) \tag{8}$$

$$\int_{-\infty}^{\infty} f(x)\delta'(x-a)dx = -f'(a) \tag{9}$$

$$x\int f(x)\delta(x-x_0)\,dx = x_0\int f(x)\delta(x-x_0)\,dx \tag{10}$$

$$\delta' * f = \int_{-\infty}^{\infty} \delta'(a-x)f(x)\,dx = f'(x) \tag{11}$$

$$\int_{-\infty}^{\infty} |\delta'(x)|\,dx = \infty \tag{12}$$

$$x^2\delta'(x) = 0 \tag{13}$$

$$\delta'(-x) = -\delta'(x) \tag{14}$$

$$x\delta'(x) = -\delta(x). \tag{15}$$

(15) can be established using INTEGRATION BY PARTS as follows:

$$\begin{aligned} \int f(x)x\delta'(x)\,dx &= -\int \delta(x)\frac{d}{dx}[xf(x)]\,dx\\ &= -\int \delta[f(x) + xf'(x)]\,dx\\ &= -\int f(x)\delta(x)\,dx. & (16) \end{aligned}$$

Additional identities are

$$\delta(ax) = \frac{1}{a}\delta(x) \tag{17}$$

$$\delta(x^2 - a^2) = \frac{1}{2a}[\delta(x+a) + \delta(x-a)] \tag{18}$$

$$\delta[g(x)] = \sum_i \frac{\delta(x-x_i)}{|g'(x_i)|}, \tag{19}$$

where the x_is are the ROOTS of g. For example, examine

$$\delta(x^2 + x - 2) = \delta[(x-1)(x+2)]. \tag{20}$$

Then $g'(x) = 2x+1$, so $g'(x_1) = g'(1) = 3$ and $g'(x_2) = g'(-2) = -3$, and we have

$$\delta(x^2 + x - 2) = \tfrac{1}{3}\delta(x-1) + \tfrac{1}{3}\delta(x+2). \tag{21}$$

A FOURIER SERIES expansion of $\delta(x-a)$ gives

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} \delta(x-a)\cos(nx)\,dx = \frac{1}{\pi}\cos(na) \tag{22}$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} \delta(x-a)\sin(nx)\,dx = \frac{1}{\pi}\sin(na), \tag{23}$$

so

$$\begin{aligned} \delta(x-a) &= \frac{1}{2\pi} + \frac{1}{\pi}\sum_{n=1}^{\infty}[\cos(na)\cos(nx) + \sin(na)\sin(nx)]\\ &= \frac{1}{2\pi} + \frac{1}{\pi}\sum_{n=1}^{\infty}\cos[n(x-a)]. & (24) \end{aligned}$$

The FOURIER TRANSFORM of the delta function is

$$\mathcal{F}[\delta(x-x_0)] = \int_{-\infty}^{\infty} e^{-2\pi ikx}\delta(x-x_0)\,dx = e^{-2\pi ikx_0}. \tag{25}$$

Delta functions can also be defined in 2-D, so that in 2-D CARTESIAN COORDINATES

$$\delta^2(x-x_0, y-y_0) = \delta(x-x_0)\delta(y-y_0), \tag{26}$$

and in 3-D, so that in 3-D CARTESIAN COORDINATES

$$\delta^3(x-x_0, y-y_0, z-z_0) = \delta(x-x_0)\delta(y-y_0)\delta(z-z_0), \tag{27}$$

in CYLINDRICAL COORDINATES

$$\delta^3(r,\theta,z) = \frac{\delta(r)\delta(z)}{\pi r}, \tag{28}$$

and in SPHERICAL COORDINATES,

$$\delta^3(r,\theta,\phi) = \frac{\delta(r)}{2\pi r^2}. \tag{29}$$

A series expansion in CYLINDRICAL COORDINATES gives

$$\begin{aligned}\delta^3(\mathbf{r}_1-\mathbf{r}_2) &= \frac{1}{r_1}\delta(r_1-r_2)\delta(\phi_1-\phi_2)\delta(z_1-z_2)\\ &= \frac{1}{r_1}\delta(r_1-r_2)\frac{1}{2\pi}\sum_{m=-\infty}^{\infty} e^{im(\phi_1-\phi_2)}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ik(z_1-z_2)}\,dk.\end{aligned} \tag{30}$$

The delta function also obeys the so-called SIFTING PROPERTY

$$\int f(\mathbf{y})\delta(\mathbf{x}-\mathbf{y})\,d\mathbf{y} = f(\mathbf{x}). \tag{31}$$

see also DELTA SEQUENCE, DOUBLET FUNCTION, FOURIER TRANSFORM—DELTA FUNCTION

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 481–485, 1985.

Spanier, J. and Oldham, K. B. "The Dirac Delta Function $\delta(x-a)$." Ch. 10 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 79–82, 1987.

Delta Sequence

A SEQUENCE of strongly peaked functions for which

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} \delta_n(x)f(x)\,dx = f(n) \tag{1}$$

so that in the limit as $n \to \infty$, the sequences become DELTA FUNCTIONS. Examples include

$$\begin{aligned}\delta_n(x) &= \begin{cases} 0 & x < -\frac{1}{2n}\\ n & -\frac{1}{2n} < x < \frac{1}{2n}\\ 0 & x > \frac{1}{2n}\end{cases} && (2)\\ &= \frac{n}{\sqrt{\pi}}e^{-n^2x^2} && (3)\\ &= \frac{n}{\pi}\operatorname{sinc}(ax) \equiv \frac{\sin(nx)}{\pi x} && (4)\\ &= \frac{1}{\pi x}\frac{e^{inx}-e^{-inx}}{2i} && (5)\\ &= \frac{1}{2\pi ix}\left[e^{ixt}\right]_{-n}^{n} && (6)\\ &= \frac{1}{2\pi}\int_{-n}^{n} e^{ixt}\,dt && (7)\\ &= \frac{1}{2\pi}\frac{\sin\left[\left(n+\frac{1}{2}\right)x\right]}{\sin\left(\frac{1}{2}x\right)}, && (8)\end{aligned}$$

where (8) is known as the DIRICHLET KERNEL.

Delta Variation

see VARIATION

Deltahedron

A semiregular POLYHEDRON whose faces are all EQUILATERAL TRIANGLES. There are an infinite number of deltahedra, but only eight convex ones (Freudenthal and van der Waerden 1947). They have 4, 6, 8, 10, 12, 14, 16, and 20 faces. These are summarized in the table below, and illustrated in the following figures.

n	Name
4	tetrahedron
6	triangular dipyramid
8	octahedron
10	pentagonal dipyramid
12	snub disphenoid
14	triaugmented triangular prism
16	gyroelongated square dipyramid
20	icosahedron

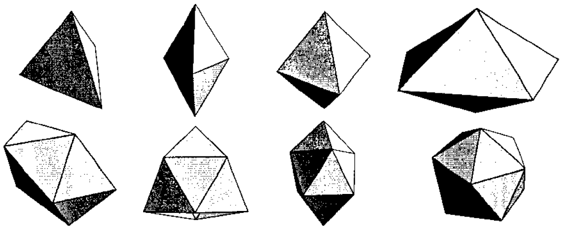

The STELLA OCTANGULA is a concave deltahedron with 24 sides:

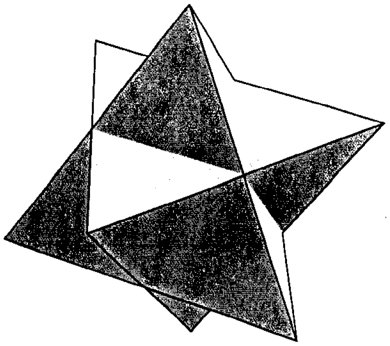

Another with 60 faces is a "caved in" DODECAHEDRON which is ICOSAHEDRON STELLATION I_{20}.

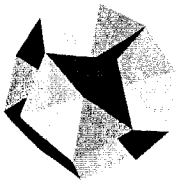

Cundy (1952) identifies 17 concave deltahedra with two kinds of VERTICES.

see also GYROELONGATED SQUARE DIPYRAMID, ICOSAHEDRON, OCTAHEDRON, PENTAGONAL DIPYRAMID, SNUB DISPHENOID TETRAHEDRON, TRIANGULAR DIPYRAMID, TRIAUGMENTED TRIANGULAR PRISM

References
Cundy, H. M. "Deltahedra." *Math. Gaz.* **36**, 263–266, 1952.
Freudenthal, H. and van der Waerden, B. L. "On an Assertion of Euclid." *Simon Stevin* **25**, 115–121, 1947.
Gardner, M. *Fractal Music, Hypercards, and More: Mathematical Recreations from Scientific American Magazine.* New York: W. H. Freeman, pp. 40, 53, and 58–60, 1992.
Pugh, A. *Polyhedra: A Visual Approach.* Berkeley, CA: University of California Press, pp. 35–36, 1976.

Deltoid

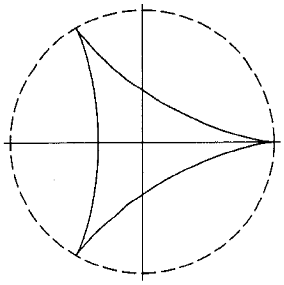

A 3-cusped HYPOCYCLOID, also called a TRICUSPOID, which has $n \equiv a/b = 3$ or $3/2$, where a is the RADIUS of the large fixed CIRCLE and b is the RADIUS of the small rolling CIRCLE. The deltoid was first considered by Euler in 1745 in connection with an optical problem. It was also investigated by Steiner in 1856 and is sometimes called STEINER'S HYPOCYCLOID (MacTutor Archive). The equation of the deltoid is obtained by setting $n = 3$ in the equation of the HYPOCYCLOID, yielding the parametric equations

$$x = [\tfrac{2}{3}\cos\phi - \tfrac{1}{3}\cos(2\phi)]a = 2b\cos\phi + b\cos(2\phi) \quad (1)$$

$$y = [\tfrac{2}{3}\sin\phi + \tfrac{1}{3}\sin(2\phi)]a = 2b\sin\phi - b\sin(2\phi). \quad (2)$$

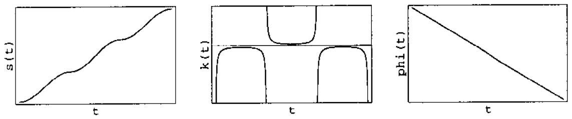

The ARC LENGTH, CURVATURE, and TANGENTIAL ANGLE are

$$s(t) = 4\int_0^t |\sin(\tfrac{3}{2}t')|\,dt' = \tfrac{16}{3}\sin^2(\tfrac{3}{4}t) \quad (3)$$

$$\kappa(t) = -\tfrac{1}{8}\csc(\tfrac{3}{2}t) \quad (4)$$

$$\phi(t) = -\tfrac{1}{2}t. \quad (5)$$

As usual, care must be taken in the evaluation of $s(t)$ for $t > 2\pi/3$. Since the form given above comes from an integral involving the ABSOLUTE VALUE of a function, it must be monotonic increasing. Each branch can be treated correctly by defining

$$n = \left\lfloor \frac{3t}{2\pi} \right\rfloor + 1, \quad (6)$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION, giving the formula

$$s(t) = (-1)^{1+[n \pmod 2]}\tfrac{16}{3}\sin^2(\tfrac{3}{4}t) + \tfrac{32}{3}\left\lfloor \tfrac{1}{2}n \right\rfloor. \quad (7)$$

The total ARC LENGTH is computed from the general HYPOCYCLOID equation

$$s_n = \frac{8a(n-1)}{n}. \quad (8)$$

With $n = 3$, this gives

$$s_3 = \tfrac{16}{3}a. \quad (9)$$

The AREA is given by

$$A_n = \frac{(n-1)(n-2)}{n^2}\pi a^2 \quad (10)$$

with $n = 3$,

$$A_3 = \tfrac{2}{9}\pi a^2. \quad (11)$$

The length of the tangent to the tricuspoid, measured between the two points P, Q in which it cuts the curve again, is constant and equal to $4a$. If you draw TANGENTS at P and Q, they are at RIGHT ANGLES.

References
Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 53, 1993.
Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 131–135, 1972.

Lee, X. "Deltoid." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Deltoid_dir/deltoid.html`.
Lockwood, E. H. "The Deltoid." Ch. 8 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 72–79, 1967.
Macbeth, A. M. "The Deltoid, I." *Eureka* **10**, 20–23, 1948.
Macbeth, A. M. "The Deltoid, II." *Eureka* **11**, 26–29, 1949.
Macbeth, A. M. "The Deltoid, III." *Eureka* **12**, 5–6, 1950.
MacTutor History of Mathematics Archive. "Tricuspoid." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Tricuspoid.html`.
Yates, R. C. "Deltoid." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 71–74, 1952.

Deltoid Caustic

The caustic of the DELTOID when the rays are PARALLEL in any direction is an ASTROID.

Deltoid Evolute

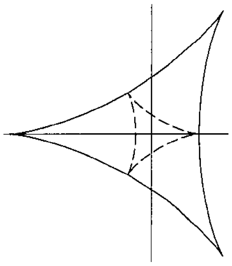

A HYPOCYCLOID EVOLUTE for $n = 3$ is another DELTOID scaled by a factor $n/(n-2) = 3/1 = 3$ and rotated $1/(2 \cdot 3) = 1/6$ of a turn.

Deltoid Involute

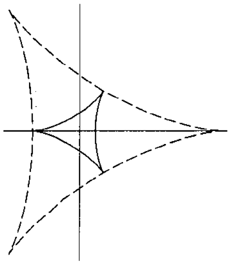

A HYPOCYCLOID INVOLUTE for $n = 3$ is another DELTOID scaled by a factor $(n-2)/n = 1/3$ and rotated $1/(2 \cdot 3) = 1/6$ of a turn.

Deltoid Pedal Curve

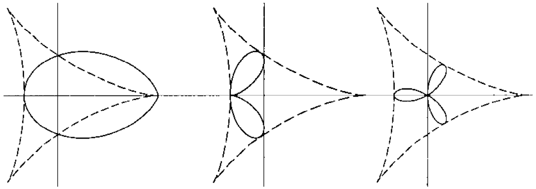

The PEDAL CURVE for a DELTOID with the PEDAL POINT at the CUSP is a FOLIUM. For the PEDAL POINT at the CUSP (NEGATIVE x-intercept), it is a BIFOLIUM. At the center, or anywhere on the inscribed EQUILATERAL TRIANGLE, it is a TRIFOLIUM.

Deltoid Radial Curve

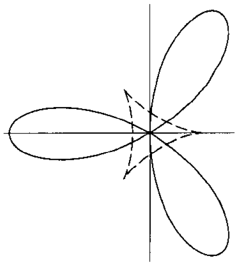

The TRIFOLIUM

$$x = x_0 + 4a\cos\phi - 4a\cos(2\phi)$$
$$y = y_0 + 4a\sin\phi + 4a\sin(2\phi).$$

Deltoidal Hexecontahedron

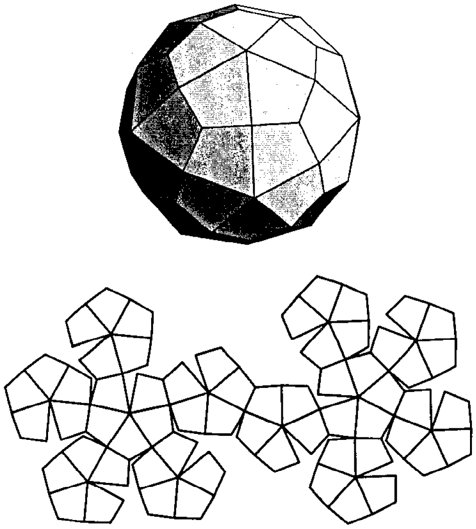

The DUAL POLYHEDRON of the RHOMBICOSIDODECAHEDRON.

Deltoidal Icositetrahedron

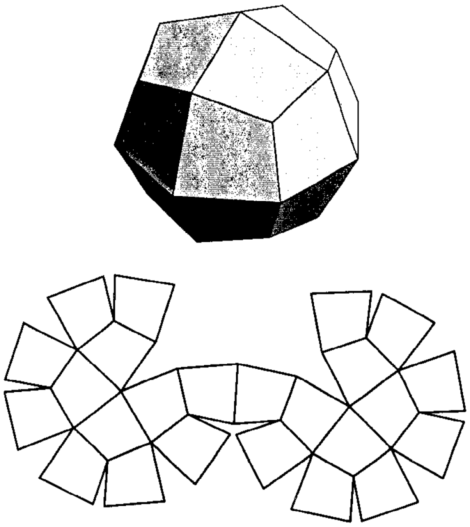

The DUAL POLYHEDRON of the SMALL RHOMBICUBOCTAHEDRON. It is also called the TRAPEZOIDAL ICOSITETRAHEDRON.

Demlo Number

The initially PALINDROMIC NUMBERS 1, 121, 12321, 1234321, 123454321, ... (Sloane's A002477). For the first through ninth terms, the sequence is given by the GENERATING FUNCTION

$$-\frac{10x+1}{(x-1)(10x-1)(100x-1)} = 1 + 121x + 12321x^2 + 1234321x^3 + \ldots$$

(Plouffe 1992, Sloane and Plouffe 1995). The definition of this sequence is slightly ambiguous from the tenth term on.

see also CONSECUTIVE NUMBER SEQUENCES, PALINDROMIC NUMBER

References

Kaprekar, D. R. "On Wonderful Demlo Numbers." *Math. Student* **6**, 68–70, 1938.
Plouffe, S. "Approximations de Séries Génératrices et quelques conjectures." Montréal, Canada: Université du Québec à Montréal, Mémoire de Maîtrise, UQAM, 1992.
Sloane, N. J. A. Sequence A002477/M5386 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Dendrite Fractal

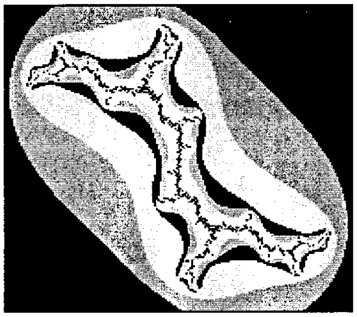

A JULIA SET with $c = i$.

Denjoy Integral

A type of INTEGRAL which is an extension of both the RIEMANN INTEGRAL and the LEBESGUE INTEGRAL. The original Denjoy integral is now called a Denjoy integral "in the restricted sense," and a more general type is now called a Denjoy integral "in the wider sense." The independently discovered PERON INTEGRAL turns out to be equivalent to the Denjoy integral "in the restricted sense."

see also INTEGRAL, LEBESGUE INTEGRAL, PERON INTEGRAL, RIEMANN INTEGRAL

References

Iyanaga, S. and Kawada, Y. (Eds.). "Denjoy Integrals." §103 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 337–340, 1980.
Kestelman, H. "General Denjoy Integral." §9.2 in *Modern Theories of Integration, 2nd rev. ed.* New York: Dover, pp. 217–227, 1960.

Denominator

The number q in a FRACTION p/q.

see also FRACTION, NUMERATOR, RATIO, RATIONAL NUMBER

Dense

A set A in a FIRST-COUNTABLE SPACE is dense in B if $B = A \cup L$, where L is the limit of sequences of elements of A. For example, the rational numbers are dense in the reals. In general, a SUBSET A of X is dense if its CLOSURE $\mathrm{cl}(A) = X$.

see also CLOSURE, DENSITY, DERIVED SET, PERFECT SET

Density

see DENSITY (POLYGON), DENSITY (SEQUENCE), NATURAL DENSITY

Density (Polygon)

The number q in a STAR POLYGON $\{\frac{p}{q}\}$.

see also STAR POLYGON

Density (Sequence)

Let a SEQUENCE $\{a_i\}_{i=1}^{\infty}$ be strictly increasing and composed of NONNEGATIVE INTEGERS. Call $A(x)$ the number of terms not exceeding x. Then the density is given by $\lim_{x\to\infty} A(x)/x$ if the LIMIT exists.

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 199, 1994.

Denumerable Set

A SET is denumerable if a prescription can be given for identifying its members one at a time. Such a set is said to have CARDINAL NUMBER $\aleph_0$. Examples of denumerable sets include ALGEBRAIC NUMBERS, INTEGERS, and RATIONAL NUMBERS. Once one denumerable set S is given, any other set which can be put into a ONE-TO-ONE correspondence with S is also denumerable. Examples of nondenumerable sets include the REAL, COMPLEX, IRRATIONAL, and TRANSCENDENTAL NUMBERS.

see also ALEPH-0, ALEPH-1, CANTOR DIAGONAL SLASH, CONTINUUM, HILBERT HOTEL

References

Courant, R. and Robbins, H. "The Denumerability of the Rational Number and the Non-Denumerability of the Continuum." §2.4.2 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 79–83, 1996.

Denumerably Infinite

see DENUMERABLE SET

Depth (Graph)

The depth $E(G)$ of a GRAPH G is the minimum number of PLANAR GRAPHS P_i needed such that the union $\cup_i P_i = G$.

see also PLANAR GRAPH

Depth (Size)

The depth of a box is the horizontal DISTANCE from front to back (usually not necessarily defined to be smaller than the WIDTH, the horizontal DISTANCE from side to side).

see also HEIGHT, WIDTH (SIZE)

Depth (Statistics)

The smallest RANK (either up or down) of a set of data.

References

Tukey, J. W. *Explanatory Data Analysis.* Reading, MA: Addison-Wesley, p. 30, 1977.

Depth (Tree)

The depth of a RESOLVING TREE is the number of levels of links, not including the top. The depth of the link is the minimal depth for any RESOLVING TREE of that link. The only links of length 0 are the trivial links. A KNOT of length 1 is always a trivial KNOT and links of depth one are always HOPF LINKS, possibly with a few additional trivial components (Bleiler and Scharlemann). The LINKS of depth two have also been classified (Thompson and Scharlemann).

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, p. 169, 1994.

Derangement

A PERMUTATION of n ordered objects in which none of the objects appears in its natural place. The function giving this quantity is the SUBFACTORIAL $!n$, defined by

$$!n \equiv n! \sum_{k=0}^{n} \frac{(-1)^k}{k!} \tag{1}$$

or

$$!n \equiv \left[\frac{n!}{e}\right], \tag{2}$$

where $k!$ is the usual FACTORIAL and $[x]$ is the NINT function. These are also called RENCONTRES NUMBERS (named after rencontres solitaire), or COMPLETE PERMUTATIONS, or derangements. The number of derangements $!n = d(n)$ of length n satisfy the RECURRENCE RELATIONS

$$d(n) = (n-1)[d(n-1) + d(n-2)] \tag{3}$$

and

$$d(n) = nd(n-1) + (-1)^n, \tag{4}$$

with $d(1) = 0$ and $d(2) = 1$. The first few are 0, 1, 2, 9, 44, 265, 1854, ... (Sloane's A000166). This sequence cannot be expressed as a fixed number of hypergeometric terms (Petkovšek *et al.* 1996, pp. 157–160).

see also MARRIED COUPLES PROBLEM, PERMUTATION, ROOT, SUBFACTORIAL

References

Aitken, A. C. *Determinants and Matrices.* Westport, CT: Greenwood Pub., p. 135, 1983.

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 46–47, 1987.

Coolidge, J. L. *An Introduction to Mathematical Probability.* Oxford, England: Oxford University Press, p. 24, 1925.

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 115–116, 1996.

de Montmort, P. R. *Essai d'analyse sur les jeux de hasard.* Paris, p. 132, 1713.

Dickau, R. M. "Derangements." `http://forum.swarthmore.edu/advanced/robertd/derangements.html`.

Durell, C. V. and Robson, A. *Advanced Algebra.* London, p. 459, 1937.

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, 1996.

Roberts, F. S. *Applied Combinatorics.* Englewood Cliffs, NJ: Prentice-Hall, 1984.

Ruskey, F. "Information on Derangements." `http://sue.csc.uvic.ca/~cos/inf/perm/Derangements.html`.

Sloane, N. J. A. Sequence A000166/M1937 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Stanley, R. P. *Enumerative Combinatorics, Vol. 1.* New York: Cambridge University Press, p. 67, 1986.

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 123, 1991.

Derivative

The derivative of a FUNCTION represents an infinitesimal change in the function with respect to whatever parameters it may have. The "simple" derivative of a function f with respect to x is denoted either $f'(x)$ or $\frac{df}{dx}$ (and often written in-line as df/dx). When derivatives are taken with respect to time, they are often denoted using Newton's FLUXION notation, $\frac{dx}{dt} = \dot{x}$. The derivative of a function $f(x)$ with respect to the variable x is defined as

$$f'(x) \equiv \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}. \tag{1}$$

Note that in order for the limit to exist, both $\lim_{h\to 0^+}$ and $\lim_{h\to 0^-}$ must exist and be equal, so the FUNCTION must be continuous. However, continuity is a NECESSARY but *not* SUFFICIENT condition for differentiability. Since some DISCONTINUOUS functions can be integrated, in a sense there are "more" functions which can be integrated than differentiated. In a letter to Stieltjes, Hermite wrote, "I recoil with dismay and horror at this lamentable plague of functions which do not have derivatives."

A 3-D generalization of the derivative to an arbitrary direction is known as the DIRECTIONAL DERIVATIVE.

In general, derivatives are mathematical objects which exist between smooth functions on manifolds. In this formalism, derivatives are usually assembled into "TANGENT MAPS."

Simple derivatives of some simple functions follow.

$$\frac{d}{dx}x^n = nx^{n-1} \tag{2}$$

$$\frac{d}{dx}\ln|x| = \frac{1}{x} \tag{3}$$

$$\frac{d}{dx}\sin x = \cos x \tag{4}$$

$$\frac{d}{dx}\cos x = -\sin x \tag{5}$$

$$\begin{aligned}\frac{d}{dx}\tan x &= \frac{d}{dx}\left(\frac{\sin x}{\cos x}\right) = \frac{\cos x\cos x - \sin x(-\sin x)}{\cos^2 x}\\ &= \frac{1}{\cos^2 x} = \sec^2 x\end{aligned} \tag{6}$$

$$\begin{aligned}\frac{d}{dx}\csc x &= \frac{d}{dx}(\sin x)^{-1} = -(\sin x)^{-2}\cos x = -\frac{\cos x}{\sin^2 x}\\ &= -\csc x\cot x\end{aligned} \tag{7}$$

$$\begin{aligned}\frac{d}{dx}\sec x &= \frac{d}{dx}(\cos x)^{-1} = -(\cos x)^{-2}(-\sin x) = \frac{\sin x}{\cos^2 x}\\ &= \sec x\tan x\end{aligned} \tag{8}$$

$$\begin{aligned}\frac{d}{dx}\cot x &= \frac{d}{dx}\left(\frac{\cos x}{\sin x}\right) = \frac{\sin x(-\sin x) - \cos x\cos x}{\cos^2 x}\\ &= -\frac{1}{\cos^2 x} = -\csc^2 x\end{aligned} \tag{9}$$

$$\frac{d}{dx}e^x = e^x \tag{10}$$

$$\begin{aligned}\frac{d}{dx}a^x &= \frac{d}{dx}e^{\ln a^x} = \frac{d}{dx}e^{x\ln a}\\ &= (\ln a)e^{x\ln a} = (\ln a)a^x\end{aligned} \tag{11}$$

$$\frac{d}{dx}\sin^{-1}x = \frac{1}{\sqrt{1-x^2}} \tag{12}$$

$$\frac{d}{dx}\cos^{-1}x = -\frac{1}{\sqrt{1-x^2}} \tag{13}$$

$$\frac{d}{dx}\tan^{-1}x = \frac{1}{1+x^2} \tag{14}$$

$$\frac{d}{dx}\cot^{-1}x = -\frac{1}{1+x^2} \tag{15}$$

$$\frac{d}{dx}\sec^{-1}x = \frac{1}{x\sqrt{x^2-1}} \tag{16}$$

$$\frac{d}{dx}\csc^{-1}x = -\frac{1}{x\sqrt{x^2-1}} \tag{17}$$

$$\frac{d}{dx}\sinh x = \cosh x \tag{18}$$

$$\frac{d}{dx}\cosh x = \sinh x \tag{19}$$

$$\frac{d}{dx}\tanh x = \operatorname{sech}^2 x \tag{20}$$

$$\frac{d}{dx}\coth x = -\operatorname{csch}^2 x \tag{21}$$

$$\frac{d}{dx}\operatorname{sech} x = -\operatorname{sech} x\tanh x \tag{22}$$

$$\frac{d}{dx}\operatorname{csch} x = -\operatorname{csch} x\coth x \tag{23}$$

$$\frac{d}{dx}\operatorname{sn} x = \operatorname{cn} x\operatorname{dn} x \tag{24}$$

$$\frac{d}{dx}\operatorname{cn} x = -\operatorname{sn} x\operatorname{dn} x \tag{25}$$

$$\frac{d}{dx}\operatorname{dn} x = -k^2\operatorname{sn} x\operatorname{cn} x. \tag{26}$$

Derivatives of sums are equal to the sum of derivatives so that

$$[f(x)+\ldots+h(x)]' = f'(x)+\ldots+h'(x). \tag{27}$$

In addition, if c is a constant,

$$\frac{d}{dx}[cf(x)] = cf'(x). \tag{28}$$

Furthermore,

$$\frac{d}{dx}[f(x)g(x)] = f(x)g'(x)+f'(x)g(x), \tag{29}$$

where f' denotes the DERIVATIVE of f with respect to x. This derivative rule can be applied iteratively to yield derivate rules for products of three or more functions, for example,

$$\begin{aligned}[fgh]' &= (fg)h' + (fg)'h = fgh' + (fg'+f'g)h\\ &= f'gh + fg'h + fgh'.\end{aligned} \tag{30}$$

Other rules involving derivatives include the CHAIN RULE, POWER RULE, PRODUCT RULE, and QUOTIENT RULE. Miscellaneous other derivative identities include

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} \tag{31}$$

$$\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}}. \tag{32}$$

If $F(x,y) = C$, where C is a constant, then

$$dF = \frac{\partial F}{\partial y}dy + \frac{\partial F}{\partial x}dx = 0, \tag{33}$$

so

$$\frac{dy}{dx} = -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}}. \tag{34}$$

A vector derivative of a vector function

$$\mathbf{X}(t) \equiv \begin{bmatrix} x_1(t)\\ x_2(t)\\ \vdots\\ x_k(t)\end{bmatrix} \tag{35}$$

can be defined by

$$\frac{d\mathbf{X}}{dt} = \begin{bmatrix} \frac{dx_1}{dt} \\ \frac{dx_2}{dt} \\ \vdots \\ \frac{dx_k}{dt}. \end{bmatrix} \tag{36}$$

see also BLANCMANGE FUNCTION, CARATHÉODORY DERIVATIVE, COMMA DERIVATIVE, CONVECTIVE DERIVATIVE, COVARIANT DERIVATIVE, DIRECTIONAL DERIVATIVE, EULER-LAGRANGE DERIVATIVE, FLUXION, FRACTIONAL CALCULUS, FRÉCHET DERIVATIVE, LAGRANGIAN DERIVATIVE, LIE DERIVATIVE, POWER RULE, SCHWARZIAN DERIVATIVE, SEMICOLON DERIVATIVE, WEIERSTRAß FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 11, 1972.

Anton, H. *Calculus with Analytic Geometry, 5th ed.* New York: Wiley, 1987.

Beyer, W. H. "Derivatives." *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 229–232, 1987.

Derivative Test

see FIRST DERIVATIVE TEST, SECOND DERIVATIVE TEST

Derived Set

The LIMIT POINTS of a SET P, denoted P'.

see also DENSE, LIMIT POINT, PERFECT SET

Dervish

A QUINTIC SURFACE having the maximum possible number of ORDINARY DOUBLE POINTS (31), which was constructed by W. Barth in 1994 (Endraß). The implicit equation of the surface is

$$\begin{aligned} &64(x-w)[x^4 - 4x^3w - 10x^2y^2 - 4x^2w^2 \\ &\quad +16xw^3 - 20xy^2w + 5y^4 + 16w^4 - 20y^2w^2] \\ &\quad -5\sqrt{5-\sqrt{5}}\,(2z - \sqrt{5-\sqrt{5}}\,w) \\ &\qquad\qquad \times[4(x^2+y^2+z^2) + (1+3\sqrt{5})w^2]^2, \end{aligned}$$

where w is a parameter (Endraß). The surface can also be described by the equation

$$aF + q = 0, \tag{1}$$

where

$$F = h_1h_2h_3h_4h_5, \tag{2}$$

$$h_1 = x - z \tag{3}$$

$$h_2 = \cos\left(\frac{2\pi}{5}\right)x - \sin\left(\frac{2\pi}{5}\right)y - z \tag{4}$$

$$h_3 = \cos\left(\frac{4\pi}{5}\right)x - \sin\left(\frac{4\pi}{5}\right)y - z \tag{5}$$

$$h_4 = \cos\left(\frac{6\pi}{5}\right)x - \sin\left(\frac{6\pi}{5}\right)y - z \tag{6}$$

$$h_5 = \cos\left(\frac{8\pi}{5}\right)x - \sin\left(\frac{8\pi}{5}\right)y - z \tag{7}$$

$$q = (1 - cz)(x^2 + y^2 - 1 + rz^2)^2, \tag{8}$$

and

$$r = \tfrac{1}{4}(1+\sqrt{5}) \tag{9}$$

$$a = -\frac{8}{5}\left(1 + \frac{1}{\sqrt{5}}\right)\sqrt{5-\sqrt{5}} \tag{10}$$

$$c = \tfrac{1}{2}\sqrt{5-\sqrt{5}} \tag{11}$$

(Nordstrand).

The dervish is invariant under the GROUP D_5 and contains exactly 15 lines. Five of these are the intersection of the surface with a D_5-invariant cone containing 16 nodes, five are the intersection of the surface with a D_5-invariant plane containing 10 nodes, and the last five are the intersection of the surface with a second D_5-invariant plane containing no nodes (Endraß).

References

Endraß, S. "Togliatti Surfaces." `http://www.mathematik.uni-mainz.de/AlgebraischeGeometrie/docs/Etogliatti.shtml`.

Endraß, S. "Flächen mit vielen Doppelpunkten." *DMV-Mitteilungen* **4**, 17–20, 4/1995.

Endraß, S. *Symmetrische Fläche mit vielen gewöhnlichen Doppelpunkten.* Ph.D. thesis. Erlangen, Germany, 1996.

Nordstrand, T. "Dervish." `http://www.uib.no/people/nfytn/dervtxt.htm`.

Desargues' Theorem

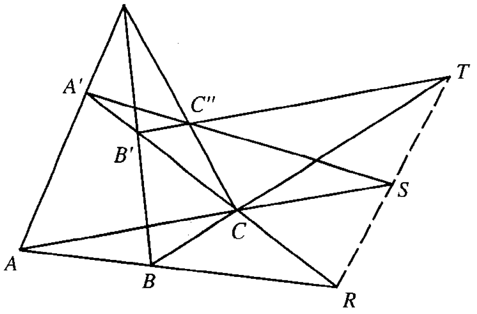

If the three straight LINES joining the corresponding VERTICES of two TRIANGLES ABC and $A'B'C'$ all meet in a point (the PERSPECTIVE CENTER), then the three intersections of pairs of corresponding sides lie on a straight LINE (the PERSPECTIVE AXIS). Equivalently, if two TRIANGLES are PERSPECTIVE from a POINT, they are PERSPECTIVE from a LINE.

Desargues' theorem is essentially its own dual according to the DUALITY PRINCIPLE of PROJECTIVE GEOMETRY.

see also DUALITY PRINCIPLE, PAPPUS'S HEXAGON THEOREM, PASCAL LINE, PASCAL'S THEOREM, PERSPECTIVE AXIS, PERSPECTIVE CENTER, PERSPECTIVE TRIANGLES

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 70–72, 1967.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 89–92, 1990.

Descartes Circle Theorem

A special case of APOLLONIUS' PROBLEM requiring the determination of a CIRCLE touching three mutually tangent CIRCLES (also called the KISSING CIRCLES PROBLEM). There are two solutions: a small circle surrounded by the three original CIRCLES, and a large circle surrounding the original three. Frederick Soddy gave the FORMULA for finding the RADIUS of the so-called inner and outer SODDY CIRCLES given the RADII of the other three. The relationship is

$$2({\kappa_1}^2 + {\kappa_2}^2 + {\kappa_3}^2 + {\kappa_4}^2) = (\kappa_1 + \kappa_2 + \kappa_3 + \kappa_4)^2,$$

where κ_i are the CURVATURES of the CIRCLES. Here, the NEGATIVE solution corresponds to the outer SODDY CIRCLE and the POSITIVE solution to the inner SODDY CIRCLE. This formula was known to Descartes and Viète (Boyer and Merzbach 1991, p. 159), but Soddy extended it to SPHERES. In n-D space, $n+2$ mutually touching n-SPHERES can always be found, and the relationship of their CURVATURES is

$$n\left(\sum_{i=1}^{n+2} {\kappa_i}^2\right) = \left(\sum_{i=1}^{n+2} \kappa_i\right)^2.$$

see also APOLLONIUS' PROBLEM, FOUR COINS PROBLEM, SODDY CIRCLES, SPHERE PACKING

References

Boyer, C. B. and Merzbach, U. C. *A History of Mathematics, 2nd ed.* New York: Wiley, 1991.

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 13–16, 1969.

Wilker, J. B. "Four Proofs of a Generalization of the Descartes Circle Theorem." *Amer. Math. Monthly* **76**, 278–282, 1969.

Descartes-Euler Polyhedral Formula

see POLYHEDRAL FORMULA

Descartes Folium

see FOLIUM OF DESCARTES

Descartes' Formula

see DESCARTES TOTAL ANGULAR DEFECT

Descartes Ovals

see CARTESIAN OVALS

Descartes' Sign Rule

A method of determining the maximum number of POSITIVE and NEGATIVE REAL ROOTS of a POLYNOMIAL.

For POSITIVE ROOTS, start with the SIGN of COEFFICIENT of the lowest (or highest) POWER. Count the number of SIGN changes n as you proceed from the lowest to the highest POWER (ignoring POWERS which do not appear). Then n is the *maximum* number of POSITIVE ROOTS. Furthermore, the number of allowable ROOTS is n, $n-2$, $n-4$, ... For example, consider the POLYNOMIAL

$$f(x) = x^7 + x^6 - x^4 - x^3 - x^2 + x - 1.$$

Since there are three SIGN changes, there are a maximum of three possible POSITIVE ROOTS.

For NEGATIVE ROOTS, starting with a POLYNOMIAL $f(x)$, write a new POLYNOMIAL $g(x)$ with the SIGNS of all ODD POWERS reversed, while leaving the SIGNS of the EVEN POWERS unchanged. Then proceed as before to count the number of SIGN changes n. Then n is the *maximum* number of NEGATIVE ROOTS. For example, consider the POLYNOMIAL

$$f(x) = x^7 + x^6 - x^4 - x^3 - x^2 + x - 1,$$

and compute the new POLYNOMIAL

$$g(x) = -x^7 + x^6 - x^4 + x^3 - x^2 - x - 1.$$

There are four SIGN changes, so there are a maximum of four NEGATIVE ROOTS.

see also BOUND, STURM FUNCTION

References

Anderson, B.; Jackson, J.; and Sitharam, M. "Descartes' Rule of Signs Revisited." *Amer. Math. Monthly* **105**, 447–451, 1998.

Hall, H. S. and Knight, S. R. *Higher Algebra: A Sequel to Elementary Algebra for Schools.* London: Macmillan, pp. 459–460, 1950.

Struik, D. J. (Ed.). *A Source Book in Mathematics 1200–1800.* Princeton, NJ: Princeton University Press, pp. 89–93, 1986.

Descartes Total Angular Defect

The total angular defect is the sum of the ANGULAR DEFECTS over all VERTICES of a POLYHEDRON, where the ANGULAR DEFECT δ at a given VERTEX is the difference between the sum of face angles and 2π. For any convex POLYHEDRON, the Descartes total angular defect is

$$\Delta = \sum_i \delta_i = 4\pi. \tag{1}$$

This is equivalent to the POLYHEDRAL FORMULA for a closed rectilinear surface, which satisfies

$$\Delta = 2\pi(V - E + F). \tag{2}$$

A POLYHEDRON with N_0 equivalent VERTICES is called a PLATONIC SOLID and can be assigned a SCHLÄFLI SYMBOL $\{p, q\}$. It then satisfies

$$N_0 = \frac{4\pi}{\delta} \tag{3}$$

and

$$\delta = 2\pi - q\left(1 - \frac{2}{p}\right)\pi, \tag{4}$$

so

$$N_0 = \frac{4p}{2p + 2q - pq}. \tag{5}$$

see also ANGULAR DEFECT, PLATONIC SOLID, POLYHEDRAL FORMULA, POLYHEDRON

Descriptive Set Theory

The study of DEFINABLE SETS and functions in POLISH SPACES.

References

Becker, H. and Kechris, A. S. *The Descriptive Set Theory of Polish Group Actions.* New York: Cambridge University Press, 1996.

Design

A formal description of the constraints on the possible configurations of an experiment which is subject to given conditions. A design is sometimes called an EXPERIMENTAL DESIGN.

see also BLOCK DESIGN, COMBINATORICS, DESIGN THEORY, HADAMARD DESIGN, HOWELL DESIGN, SPHERICAL DESIGN, SYMMETRIC BLOCK DESIGN, TRANSVERSAL DESIGN

Design Theory

The study of DESIGNS and, in particular, NECESSARY and SUFFICIENT conditions for the existence of a BLOCK DESIGN.

see also BRUCK-RYSER-CHOWLA THEOREM, FISHER'S BLOCK DESIGN INEQUALITY

References

Assmus, E. F. Jr. and Key, J. D. *Designs and Their Codes.* New York: Cambridge University Press, 1993.

Colbourn, C. J. and Dinitz, J. H. *CRC Handbook of Combinatorial Designs.* Boca Raton, FL: CRC Press, 1996.

Dinitz, J. H. and Stinson, D. R. (Eds.). "A Brief Introduction to Design Theory." Ch. 1 in *Contemporary Design Theory: A Collection of Surveys.* New York: Wiley, pp. 1–12, 1992.

Desmic Surface

Let Δ_1, Δ_2, and Δ_3 be tetrahedra in projective 3-space $\mathbb{P}^3$. Then the tetrahedra are said to be desmically related if there exist constants α, β, and γ such that

$$\alpha\Delta_1 + \beta\Delta_2 + \gamma\Delta_3 = 0.$$

A desmic surface is then defined as a QUARTIC SURFACE which can be written as

$$a\Delta_1 + b\Delta_2 + c\Delta_3 = 0$$

for desmically related tetrahedra Δ_1, Δ_2, and Δ_3. Desmic surfaces have 12 ORDINARY DOUBLE POINTS, which are the vertices of three tetrahedra in 3-space (Hunt).

see also QUARTIC SURFACE

References

Hunt, B. "Desmic Surfaces." §B.5.2 in *The Geometry of Some Special Arithmetic Quotients.* New York: Springer-Verlag, pp. 311–315, 1996.

Jessop, C. §13 in *Quartic Surfaces with Singular Points.* Cambridge, England: Cambridge University Press, 1916.

Destructive Dilemma

A formal argument in LOGIC in which it is stated that

1. $P \Rightarrow Q$ and $R \Rightarrow S$ (where $\Rightarrow$ means "IMPLIES"), and
2. Either not-Q or not-S is true, from which two statements it follows that either not-P or not-R is true.

see also CONSTRUCTIVE DILEMMA, DILEMMA

Determinant

Determinants are mathematical objects which are very useful in the analysis and solution of systems of linear equations. As shown in CRAMER'S RULE, a nonhomogeneous system of linear equations has a nontrivial solution IFF the determinant of the system's MATRIX is NONZERO (so that the MATRIX is nonsingular). A 2×2 determinant is defined to be

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} \equiv \begin{vmatrix} a & b \\ c & d \end{vmatrix} \equiv ad - bc. \tag{1}$$

A $k \times k$ determinant can be expanded by MINORS to obtain

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1k} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & a_{k3} & \cdots & a_{kk} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k2} & a_{k3} & \cdots & a_{kk} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k3} & \cdots & a_{kk} \end{vmatrix} + \ldots \pm a_{1k} \begin{vmatrix} a_{21} & a_{22} & \cdots & a_{2(k-1)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{k(k-1)} \end{vmatrix}. \quad (2)$$

A general determinant for a MATRIX A has a value

$$|\mathsf{A}| = \sum_i a_{ij} a^{ij}, \quad (3)$$

with no implied summation over i and where a^{ij} is the COFACTOR of a_{ij} defined by

$$a^{ij} \equiv (-1)^{i+j} C_{ij}. \quad (4)$$

Here, C is the $(n-1) \times (n-1)$ MATRIX formed by eliminating row i and column j from A, i.e., by DETERMINANT EXPANSION BY MINORS.

Given an $n \times n$ determinant, the additive inverse is

$$|-\mathsf{A}| = (-1)^n |\mathsf{A}|. \quad (5)$$

Determinants are also DISTRIBUTIVE, so

$$|\mathsf{AB}| = |\mathsf{A}|\,|\mathsf{B}|. \quad (6)$$

This means that the determinant of a MATRIX INVERSE can be found as follows:

$$|\mathsf{I}| = |\mathsf{AA}^{-1}| = |\mathsf{A}|\,|\mathsf{A}^{-1}| = 1, \quad (7)$$

where I is the IDENTITY MATRIX, so

$$|\mathsf{A}| = \frac{1}{|\mathsf{A}^{-1}|}. \quad (8)$$

Determinants are MULTILINEAR in rows and columns, since

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{vmatrix} = \begin{vmatrix} a_1 & 0 & 0 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{vmatrix} + \begin{vmatrix} 0 & a_2 & 0 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{vmatrix} \quad (9)$$

and

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ 0 & a_5 & a_6 \\ 0 & a_8 & a_9 \end{vmatrix} + \begin{vmatrix} 0 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ 0 & a_8 & a_9 \end{vmatrix} + \begin{vmatrix} 0 & a_2 & a_3 \\ 0 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{vmatrix}. \quad (10)$$

The determinant of the SIMILARITY TRANSFORMATION of a matrix is equal to the determinant of the original MATRIX

$$|\mathsf{BAB}^{-1}| = |\mathsf{B}|\,|\mathsf{A}|\,|\mathsf{B}^{-1}| = |\mathsf{B}|\,|\mathsf{A}|\frac{1}{|\mathsf{B}^{-1}|} = |\mathsf{A}|. \quad (11)$$

The determinant of a similarity transformation minus a multiple of the unit MATRIX is given by

$$\begin{aligned} |\mathsf{B}^{-1}\mathsf{AB} - \lambda\mathsf{I}| &= |\mathsf{B}^{-1}\mathsf{AB} - \mathsf{B}^{-1}\lambda\mathsf{IB}| = |\mathsf{B}^{-1}(\mathsf{A} - \lambda\mathsf{I})\mathsf{B}| \\ &= |\mathsf{B}^{-1}|\,|\mathsf{A} - \lambda\mathsf{I}|\,|\mathsf{B}| = |\mathsf{A} - \lambda\mathsf{I}|. \end{aligned} \quad (12)$$

The determinant of a MATRIX TRANSPOSE cquals the determinant of the original MATRIX,

$$|\mathsf{A}| = |\mathsf{A}^{\mathrm{T}}|, \quad (13)$$

and the determinant of a COMPLEX CONJUGATE is equal to the COMPLEX CONJUGATE of the determinant

$$|\mathsf{A}^*| = |\mathsf{A}|^*. \quad (14)$$

Let ϵ be a small number. Then

$$|\mathsf{I} + \epsilon\mathsf{A}| = 1 + \epsilon\,\mathrm{Tr}(\mathsf{A}) + \mathcal{O}(\epsilon^2), \quad (15)$$

where $\mathrm{Tr}(\mathsf{A})$ is the trace of A. The determinant takes on a particularly simple form for a TRIANGULAR MATRIX

$$\begin{vmatrix} a_{11} & a_{21} & \cdots & a_{k1} \\ 0 & a_{22} & \cdots & a_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{kk} \end{vmatrix} = \prod_{n=1}^{k} a_{nn}. \quad (16)$$

Important properties of the determinant include the following.

1. Switching two rows or columns changes the sign.
2. Scalars can be factored out from rows and columns.
3. Multiples of rows and columns can be added together without changing the determinant's value.
4. Scalar multiplication of a row by a constant c multiplies the determinant by c.
5. A determinant with a row or column of zeros has value 0.
6. Any determinant with two rows or columns equal has value 0.

Property 1 can be established by induction. For a 2 × 2 MATRIX, the determinant is

$$\begin{aligned}\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} &= a_1b_2 - b_1a_2 = -(b_1a_2 - a_1b_2) \\ &= -\begin{vmatrix} b_1 & a_1 \\ b_2 & a_2 \end{vmatrix}\end{aligned} \tag{17}$$

For a 3 × 3 MATRIX, the determinant is

$$\begin{aligned}&\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \\ &= a_1\begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - b_1\begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix} + c_1\begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix} \\ &= -\left(a_1\begin{vmatrix} c_2 & b_2 \\ c_3 & b_3 \end{vmatrix} + b_1\begin{vmatrix} c_2 & a_2 \\ c_3 & a_3 \end{vmatrix} - c_1\begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix}\right) \\ &= -\begin{vmatrix} a_1 & c_1 & b_1 \\ a_2 & c_2 & b_2 \\ a_3 & c_3 & b_3 \end{vmatrix} \\ &= -\left(-a_1\begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} + b_1\begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix} + c_1\begin{vmatrix} b_2 & a_2 \\ b_3 & a_3 \end{vmatrix}\right) \\ &= -\begin{vmatrix} b_1 & a_1 & c_1 \\ b_2 & a_2 & c_2 \\ b_3 & a_3 & c_3 \end{vmatrix} \\ &= -\left(-a_1\begin{vmatrix} c_2 & b_2 \\ c_3 & b_3 \end{vmatrix} - b_1\begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix} + c_1\begin{vmatrix} b_2 & a_2 \\ b_3 & a_3 \end{vmatrix}\right) \\ &= -\begin{vmatrix} c_1 & b_1 & a_1 \\ c_2 & b_2 & a_2 \\ c_3 & b_3 & a_3 \end{vmatrix}.\end{aligned} \tag{18}$$

Property 2 follows likewise. For 2 × 2 and 3 × 3 matrices,

$$\begin{vmatrix} ka_1 & b_1 \\ ka_2 & b_2 \end{vmatrix} = k(a_1b_2) - k(b_1a_2) = k\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \tag{19}$$

and

$$\begin{aligned}\begin{vmatrix} ka_1 & b_1 & c_1 \\ ka_2 & b_2 & c_2 \\ ka_3 & b_3 & c_3 \end{vmatrix} &= ka_1\begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - b_1\begin{vmatrix} ka_2 & c_2 \\ ka_3 & c_3 \end{vmatrix} \\ &+ c_1\begin{vmatrix} ka_2 & b_2 \\ ka_3 & b_3 \end{vmatrix} = k\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.\end{aligned} \tag{20}$$

Property 3 follows from the identity

$$\begin{aligned}\begin{vmatrix} a_1+kb_1 & b_1 & c_1 \\ a_2+kb_2 & b_2 & c_2 \\ a_3+kb_3 & b_3 & c_3 \end{vmatrix} &= (a_1+kb_1)\begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} \\ &- b_1\begin{vmatrix} a+kb_2 & c_2 \\ a_3+kb_3 & c_3 \end{vmatrix} + c_1\begin{vmatrix} a_2+kb_2 & b_2 \\ a_3+kb_3 & b_3 \end{vmatrix}.\end{aligned} \tag{21}$$

If a_{ij} is an $n \times n$ MATRIX with a_{ij} REAL NUMBERS, then $\det[a_{ij}]$ has the interpretation as the oriented n-dimensional CONTENT of the PARALLELEPIPED spanned by the column vectors $[a_{i,1}], \ldots, [a_{i,n}]$ in $\mathbb{R}^n$. Here, "oriented" means that, up to a change of + or − SIGN, the number is the n-dimensional CONTENT, but the SIGN depends on the "orientation" of the column vectors involved. If they agree with the standard orientation, there is a + SIGN; if not, there is a − SIGN. The PARALLELEPIPED spanned by the n-D vectors $\mathbf{v}_1$ through $\mathbf{v}_i$ is the collection of points

$$t_1\mathbf{v}_1 + \ldots + t_i\mathbf{v}_i, \tag{22}$$

where t_j is a REAL NUMBER in the CLOSED INTERVAL [0,1].

There are an infinite number of 3 × 3 determinants with no 0 or ±1 entries having unity determinant. One parametric family is

$$\begin{vmatrix} -8n^2-8n & 2n+1 & 4n \\ -4n^2-4n & n+1 & 2n+1 \\ -4n^2-4n-1 & n & 2n-1 \end{vmatrix}. \tag{23}$$

Specific examples having small entries include

$$\begin{vmatrix} 2 & 3 & 2 \\ 4 & 2 & 3 \\ 9 & 6 & 7 \end{vmatrix}, \begin{vmatrix} 2 & 3 & 5 \\ 3 & 2 & 3 \\ 9 & 5 & 7 \end{vmatrix}, \begin{vmatrix} 2 & 3 & 6 \\ 3 & 2 & 3 \\ 17 & 11 & 16 \end{vmatrix}, \ldots \tag{24}$$

(Guy 1989, 1994).

see also CIRCULANT DETERMINANT, COFACTOR, HESSIAN DETERMINANT, HYPERDETERMINANT, IMMANANT, JACOBIAN, KNOT DETERMINANT, MATRIX, MINOR, PERMANENT, VANDERMONDE DETERMINANT, WRONSKIAN

References

Arfken, G. "Determinants." §4.1 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 168–176, 1985.

Guy, R. K. "Unsolved Problems Come of Age." *Amer. Math. Monthly* **96**, 903–909, 1989.

Guy, R. K. "A Determinant of Value One." §F28 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 265–266, 1994.

Determinant (Binary Quadratic Form)

The determinant of a BINARY QUADRATIC FORM

$$Au^2 + 2Buv + Cv^2$$

is

$$D \equiv B^2 - AC.$$

It is equal to 1/4 of the corresponding DISCRIMINANT.

Determinant Expansion by Minors

Also known as LAPLACIAN DETERMINANT EXPANSION BY MINORS. Let $|\mathsf{M}|$ denote the DETERMINANT of a MATRIX M, then

$$|\mathsf{M}| = \sum_{i=1}^{k} (-1)^{i+j} a_i M_{ij},$$

where M_{ij} is called a MINOR,

$$|\mathsf{M}| = \sum_{i=1}^{k} a_i C_{ij},$$

where C_{ij} is called a COFACTOR.

see also COFACTOR, DETERMINANT

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 169–170, 1985.

Determinant (Knot)

see KNOT DETERMINANT

Determinant Theorem

Given a MATRIX m, the following are equivalent:

1. $|\mathsf{m}| \neq 0$.
2. The columns of m are linearly independent.
3. The rows of m are linearly independent.
4. $\text{Range}(\mathsf{m}) = \mathbb{R}^n$.
5. $\text{Null}(\mathsf{m}) = \{0\}$.
6. m has a MATRIX INVERSE.

see also DETERMINANT, MATRIX INVERSE, NULLSPACE, RANGE (IMAGE)

Developable Surface

A surface on which the GAUSSIAN CURVATURE K is everywhere 0.

see also BINORMAL DEVELOPABLE, NORMAL DEVELOPABLE, SYNCLASTIC, TANGENT DEVELOPABLE

Deviation

The DIFFERENCE of a quantity from some fixed value, usually the "correct" or "expected" one.

see ABSOLUTE DEVIATION, AVERAGE ABSOLUTE DEVIATION, DIFFERENCE, DISPERSION (STATISTICS), MEAN DEVIATION, SIGNED DEVIATION, STANDARD DEVIATION

Devil's Curve

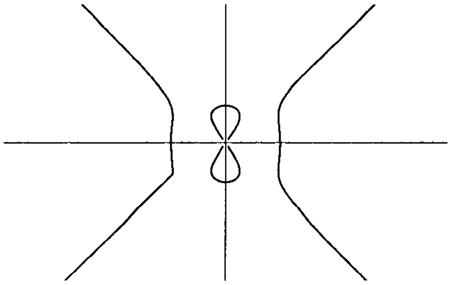

The devil's curve was studied by G. Cramer in 1750 and Lacroix in 1810 (MacTutor Archive). It appeared in *Nouvelles Annales* in 1858. The Cartesian equation is

$$y^4 - a^2y^2 = x^4 - b^2x^2, \tag{1}$$

equivalent to

$$y^2(y^2 - a^2) = x^2(x^2 - b^2), \tag{2}$$

the polar equation is

$$r^2(\sin^2\theta - \cos^2\theta) = a^2\sin^2\theta - b^2\cos^2\theta, \tag{3}$$

and the parametric equations are

$$x = \cos t\sqrt{\frac{a^2\sin^2 t - b^2\cos^2 t}{\sin^2 t - \cos^2 t}} \tag{4}$$

$$y = \sin t\sqrt{\frac{a^2\sin^2 t - b^2\cos^2 t}{\sin^2 t - \cos^2 t}}. \tag{5}$$

A special case of the Devil's curve is the so-called ELECTRIC MOTOR CURVE:

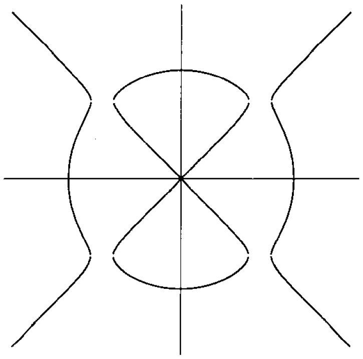

$$y^2(y^2 - 96) = x^2(x^2 - 100) \tag{6}$$

(Cundy and Rollett 1989).

see also ELECTRIC MOTOR CURVE

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 71, 1989.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 71, 1993.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 151–152, 1972.

MacTutor History of Mathematics Archive. "Devil's Curve." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Devils.html`.

Devil's Staircase

A plot of the WINDING NUMBER W resulting from MODE LOCKING as a function of Ω for the CIRCLE MAP with $K = 1$. At each value of Ω, the WINDING NUMBER is some RATIONAL NUMBER. The result is a monotonic increasing "staircase" for which the simplest RATIONAL NUMBERS have the largest steps. For $K = 1$, the MEASURE of quasiperiodic states (Ω IRRATIONAL) on the Ω-axis has become zero, and the measure of MODE-LOCKED state has become 1. The DIMENSION of the Devil's staircase $\approx 0.8700 \pm 3.7 \times 10^{-4}$.

see also CANTOR FUNCTION

References

Mandelbrot, B. B. *The Fractal Geometry of Nature.* New York: W. H. Freeman, 1983.
Ott, E. *Chaos in Dynamical Systems.* New York: Cambridge University Press, 1993.
Rasband, S. N. *Chaotic Dynamics of Nonlinear Systems.* New York: Wiley, p. 132, 1990.

Devil on Two Sticks

see DEVIL'S CURVE

Diabolical Cube

A 6-piece POLYCUBE DISSECTION of the 3×3 CUBE.

see also CUBE DISSECTION, SOMA CUBE

References

Gardner, M. "Polycubes." Ch. 3 in *Knotted Doughnuts and Other Mathematical Entertainments.* New York: W. H. Freeman, pp. 29–30, 1986.

Diabolical Square

see PANMAGIC SQUARE

Diabolo

A 2-POLYABOLO.

Diacaustic

The ENVELOPE of refracted rays for a given curve.

see also CATACAUSTIC, CAUSTIC

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, p. 60, 1972.

Diagonal Matrix

A diagonal matrix is a MATRIX A of the form

$$a_{ij} = c_i \delta_{ij}, \tag{1}$$

where δ is the KRONECKER DELTA, c_i are constants, and there is no summation over indices. The general diagonal matrix is therefore SQUARE and of the form

$$\begin{bmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_n \end{bmatrix}. \tag{2}$$

Given a MATRIX equation of the form

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}, \tag{3}$$

multiply through to obtain

$$\begin{bmatrix} a_{11}\lambda_1 & \cdots & a_{1n}\lambda_n \\ \vdots & \ddots & \vdots \\ a_{n1}\lambda_1 & \cdots & a_{nn}\lambda_n \end{bmatrix} = \begin{bmatrix} a_{11}\lambda_1 & \cdots & a_{1n}\lambda_1 \\ \vdots & \ddots & \vdots \\ a_{n1}\lambda_n & \cdots & a_{nn}\lambda_n \end{bmatrix}. \tag{4}$$

Since in general, $\lambda_i \neq \lambda_j$ for $i \neq j$, this can be true only if off-diagonal components vanish. Therefore, A must be diagonal.

Given a diagonal matrix T,

$$\mathsf{T}^n = \begin{bmatrix} t_1 & 0 & \cdots & 0 \\ 0 & t_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t_k \end{bmatrix}^n = \begin{bmatrix} {t_1}^n & 0 & \cdots & 0 \\ 0 & {t_2}^n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & {t_k}^n \end{bmatrix}. \tag{5}$$

see also MATRIX, TRIANGULAR MATRIX, TRIDIAGONAL MATRIX

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 181–184 and 217–229, 1985.

Diagonal Metric

A METRIC g_{ij} which is zero for $i \neq j$.

see also METRIC

Diagonal (Polygon)

A LINE SEGMENT connecting two nonadjacent VERTICES of a POLYGON. The number of ways a fixed convex n-gon can be divided into TRIANGLES by nonintersecting diagonals is C_{n-2} (with C_{n-3} diagonals), where C_n is a CATALAN NUMBER. This is EULER'S POLYGON DIVISION PROBLEM. Counting the number of regions determined by drawing the diagonals of a regular n-gon is a more difficult problem, as is determining the number of n-tuples of CONCURRENT diagonals (Beller *et al.* 1972, Item 2).

The number of regions which the diagonals of a CONVEX POLYGON divide its center if no three are concurrent in its interior is

$$N = \binom{n}{4} + \binom{n-1}{2} = \tfrac{1}{24}(n-1)(n-2)(n^2 - 3n + 12).$$

The first few values are 0, 0, 1, 4, 11, 25, 50, 91, 154, 246, ... (Sloane's A006522).

see also Catalan Number, Diagonal (Polyhedron), Euler's Polygon Division Problem, Polygon, Vertex (Polygon)

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Sloane, N. J. A. Sequence A006522/M3413 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Diagonal (Polyhedron)

A Line Segment connecting two nonadjacent sides of a Polyhedron. The only simple Polyhedron with no diagonals is the Tetrahedron. The only known Toroidal Polyhedron with no diagonals is the Császár Polyhedron.

see also Diagonal (Polygon), Euler Brick, Polyhedron, Space Diagonal

Diagonal Ramsey Number

A Ramsey Number of the form $R(k,k;2)$.

see also Ramsey Number

Diagonal Slash

see Cantor Diagonal Slash

Diagonal (Solidus)

see Solidus

Diagonalization

see Matrix Diagonalization

Diagonals Problem

see Euler Brick

Diagram

A schematic mathematical illustration showing the relationships between or properties of mathematical objects.

see also Alternating Knot Diagram, Argand Diagram, Coxeter-Dynkin Diagram, de Bruijn Diagram, Dynkin Diagram, Ferrers Diagram, Hasse Diagram, Heegaard Diagram, Knot Diagram, Link Diagram, Stem-and-Leaf Diagram, Venn Diagram, Voronoi Diagram, Young Diagram

Diameter

The diameter of a Circle is the Distance from a point on the Circle to point π Radians away. If r is the Radius, $d = 2r$.

see also Brocard Diameter, Circumference, Diameter (General), Diameter (Graph), Pi, Radius, Transfinite Diameter

Diameter (General)

The farthest Distance between two points on the boundary of a closed figure.

see also Borsuk's Conjecture

References

Eppstein, D. "Width, Diameter, and Geometric Inequalities." `http://www.ics.uci.edu/~eppstein/junkyard/diam.html`.

Diameter (Graph)

The length of the "longest shortest path" between two Vertices of a Graph. In other words, a graph's diameter is the largest number of vertices which must be traversed in order to travel from one vertex to another when paths which backtrack, detour, or loop are excluded from consideration.

Diamond

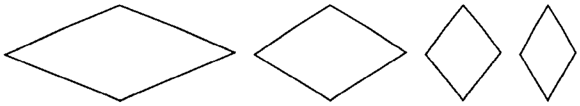

A convex Quadrilateral having sides of equal length and Perpendicular Planes of symmetry passing through opposite pairs of Vertices. The Lozenge is a special case of a diamond.

see also Kite, Lozenge, Parallelogram, Quadrilateral, Rhombus

Dice

A die (plural "dice") is a Solid with markings on each of its faces. The faces are usually all the same shape, making Platonic Solids and Archimedean Solid Duals the obvious choices. The die can be "rolled" by throwing it in the air and allowing it to come to rest on one of its faces. Dice are used in many games of chance as a way of picking Random Numbers on which to bet, and are used in board or roll-playing games to determine the number of spaces to move, results of a conflict, etc. A Coin can be viewed as a degenerate 2-sided case of a die.

The most common type of die is a six-sided Cube with the numbers 1–6 placed on the faces. The value of the roll is indicated by the number of "spots" showing on the top. For the six-sided die, opposite faces are arranged to always sum to seven. This gives two possible Mirror Image arrangements in which the numbers 1, 2, and 3 may be arranged in a clockwise or counterclockwise order about a corner. Commercial dice may, in fact, have either orientation. The illustrations below show 6-sided dice with counterclockwise and clockwise arrangements, respectively.

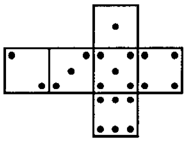

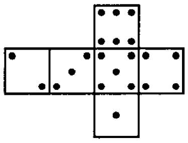

The CUBE has the nice property that there is an upward-pointing face opposite the bottom face from which the value of the "roll" can easily be read. This would not be true, for instance, for a TETRAHEDRAL die, which would have to be picked up and turned over to reveal the number underneath (although it could be determined by noting which number 1–4 was *not* visible on one of the upper three faces). The arrangement of spots ⁙ corresponding to a roll of 5 on a six-sided die is called the QUINCUNX. There are also special names for certain rolls of two six-sided dice: two 1s are called SNAKE EYES and two 6s are called BOXCARS.

Shapes of dice other than the usual 6-sided CUBE are commercially available from companies such as Dice & Games, Ltd.®

Diaconis and Keller (1989) show that there exist "fair" dice other than the usual PLATONIC SOLIDS and duals of the ARCHIMEDEAN SOLIDS, where a fair die is one for which its symmetry group acts transitively on its faces. However, they did not explicitly provide any examples.

The probability of obtaining p points (a roll of p) on n s-sided dice can be computed as follows. The number of ways in which p can be obtained is the COEFFICIENT of x^p in

$$f(x)=(x+x^2+\ldots+x^s)^n, \tag{1}$$

since each possible arrangement contributes one term. $f(x)$ can be written as a MULTINOMIAL SERIES

$$f(x)=x^n\left(\sum_{i=0}^{s-1}x^i\right)^n=x^n\left(\frac{1-x^s}{1-x}\right)^n, \tag{2}$$

so the desired number c is the COEFFICIENT of x^p in

$$x^n(1-x^s)^n(1-x)^{-n}. \tag{3}$$

Expanding,

$$x^n\sum_{k=0}^{n}(-1)^k\binom{n}{k}x^{sk}\sum_{l=0}^{\infty}\binom{n+l-1}{l}x^l, \tag{4}$$

so in order to get the COEFFICIENT of x^p, include all terms with

$$p=n+sk+l. \tag{5}$$

c is therefore

$$c=\sum_{k=0}^{n}(-1)^k\binom{n}{k}\binom{p-sk-1}{p-sk-n}. \tag{6}$$

But $p-sk-n>0$ only when $k<(p-n)/s$, so the other terms do not contribute. Furthermore,

$$\binom{p-sk-1}{p-sk-n}=\binom{p-sk-1}{n-1}, \tag{7}$$

so

$$c=\sum_{k=0}^{\lfloor (p-n)/s\rfloor}(-1)^k\binom{n}{k}\binom{p-sk-1}{n-1}, \tag{8}$$

where $\lfloor x\rfloor$ is the FLOOR FUNCTION, and

$$P(p,n,s)=\frac{1}{s^n}\sum_{k=0}^{\lfloor (p-n)/s\rfloor}(-1)^k\binom{n}{k}\binom{p-sk-1}{n-1}. \tag{9}$$

Consider now $s=6$. For $n=2$ six-sided dice,

$$k_{\max}\equiv\left\lfloor\frac{p-2}{6}\right\rfloor=\begin{cases}0 & \text{for } 2\leq p\leq 7\\ 1 & \text{for } 12\leq p\leq 8,\end{cases} \tag{10}$$

and

$$\begin{aligned}P(p,2,6)&=\frac{1}{6^2}\sum_{k=0}^{k_{\max}}(-1)^k\binom{2}{k}\binom{p-6k-1}{1}\\&=\frac{1}{6^2}\sum_{k=0}^{k_{\max}}(-1)^k\frac{2!}{k!(2-k)!}(p-6k-1)\\&=\frac{1}{36}\sum_{k=0}^{k_{\max}}(1-2k)(k+1)(p-6k-1)\\&=\frac{1}{36}\begin{cases}p-1 & \text{for } 2\leq p\leq 7\\ 13-p & \text{for } 8\leq p\leq 12\end{cases}\\&=\frac{6-|p-7|}{36}\quad\text{for } 2\leq p\leq 12.\end{aligned} \tag{11}$$

The most common roll is therefore seen to be a 7, with probability $6/36=1/6$, and the least common rolls are 2 and 12, both with probability $1/36$.

For $n=3$ six-sided dice,

$$k_{\max}=\left\lfloor\frac{p-3}{6}\right\rfloor=\begin{cases}0 & \text{for } 3\leq p\leq 8\\ 1 & \text{for } 9\leq p\leq 14\\ 2 & \text{for } 15\leq p\leq 18,\end{cases} \tag{12}$$

and

$$\begin{aligned}&P(p,3,6)\\&=\frac{1}{6^3}\sum_{k=0}^{k_{\max}}(-1)^k\binom{3}{k}\binom{p-6k-1}{2}\\&=\frac{1}{6^3}\sum_{k=0}^{k_{\max}}(-1)^k\frac{3!}{k!(3-k)!}\frac{(p-6k-1)(p-6k-2)}{2}\\&=\frac{1}{216}\begin{cases}\frac{(p-1)(p-2)}{2} & \text{for } 3\leq p\leq 8\\ \frac{(p-1)(p-2)}{2}-3\frac{(p-7)(p-8)}{2} & \text{for } 9\leq p\leq 14\\ \frac{(p-1)(p-2)}{2}-3\frac{(p-7)(p-8)}{2}+3\frac{(p-13)(p-14)}{2} & \text{for } 15\leq p\leq 18\end{cases}\\&=\frac{1}{216}\begin{cases}\frac{1}{2}(p-1)(p-2) & \text{for } 3\leq p\leq 8\\ -p^2+21p-83 & \text{for } 9\leq p\leq 14\\ \frac{1}{2}(19-p)(20-p) & \text{for } 15\leq p\leq 18.\end{cases}\end{aligned} \tag{13}$$

For three six-sided dice, the most common rolls are 10 and 11, both with probability $1/8$; and the least common rolls are 3 and 18, both with probability $1/216$.

For four six-sided dice, the most common roll is 14, with probability 73/648; and the least common rolls are 4 and 24, both with probability 1/1296.

In general, the likeliest roll p_L for n s-sided dice is given by

$$p_L(n,s) = \left\lfloor \tfrac{1}{2}n(s+1) \right\rfloor, \tag{14}$$

which can be written explicitly as

$$p_L(n,s) = \begin{cases} \tfrac{1}{2}n(s+1) & \text{for } n \text{ even} \\ \tfrac{1}{2}[n(s+1)-1] & \text{for } n \text{ odd, } s \text{ even} \\ \tfrac{1}{2}n(s+1) & \text{for } n \text{ odd, } s \text{ odd.} \end{cases} \tag{15}$$

For 6-sided dice, the likeliest rolls are given by

$$p_L(n,6) = \left\lfloor \tfrac{7}{2}n \right\rfloor = \begin{cases} \tfrac{7}{2}n & \text{for } n \text{ even} \\ \tfrac{1}{2}(7n-1) & \text{for } n \text{ odd, } s \text{ even} \\ \tfrac{7}{2}n & \text{for } n \text{ odd, } s \text{ odd,} \end{cases} \tag{16}$$

or 7, 10, 14, 17, 21, 24, 28, 31, 35, ... for $n = 2$, 3, ... (Sloane's A030123) dice. The probabilities corresponding to the most likely rolls can be computed by plugging $p = p_L$ into the general formula together with

$$k_L(n,s) = \begin{cases} \tfrac{1}{2}n & \text{for } n \text{ even} \\ \left\lfloor \frac{n(s-1)-1}{2s} \right\rfloor & \text{for } n \text{ odd, } s \text{ even} \\ \left\lfloor \frac{n(s-1)}{2s} \right\rfloor & \text{for } n \text{ odd, } s \text{ odd.} \end{cases} \tag{17}$$

Unfortunately, $P(p_L, n, s)$ does not have a simple closed-form expression in terms of s and n. However, the probabilities of obtaining the likeliest roll totals can be found explicitly for a particular s. For n 6-sided dice, the probabilities are 1/6, 1/8, 73/648, 65/648, 361/3888, 24017/279936, 7553/93312, ... for $n = 2$, 3, ...

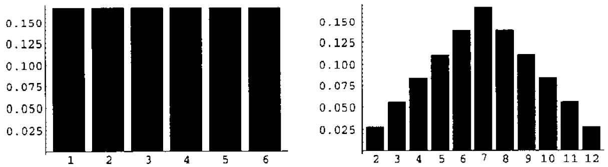

one die *two dice*

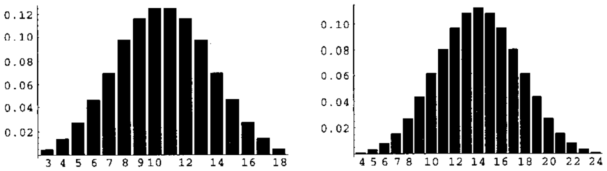

three dice *four dice*

The probabilities for obtaining a given total using n 6-sided dice are shown above for $n = 1$, 2, 3, and 4 dice. They can be seen to approach a GAUSSIAN DISTRIBUTION as the number of dice is increased.

see also BOXCARS, COIN TOSSING, CRAPS, DE MERE'S PROBLEM, EFRON'S DICE, POKER, QUINCUNX, SICHERMAN DICE, SNAKE EYES

References

Diaconis, P. and Keller, J. B. "Fair Dice." *Amer. Math. Monthly* **96**, 337–339, 1989.

Dice & Games, Ltd. "Dice & Games Hobby Games Accessories." `http://www.dice.co.uk/hob.htm`.

Gardner, M. "Dice." Ch. 18 in *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, pp. 251–262, 1978.

Robertson, L. C.; Shortt, R. M.; Landry, S. G. "Dice with Fair Sums." *Amer. Math. Monthly* **95**, 316–328, 1988.

Sloane, N. J. A. Sequence A030123 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Dichroic Polynomial

A POLYNOMIAL $Z_G(q,v)$ in two variables for abstract GRAPHS. A GRAPH with one VERTEX has $Z = q$. Adding a VERTEX not attached by any EDGES multiplies the Z by q. Picking a particular EDGE of a GRAPH G, the POLYNOMIAL for G is defined by adding the POLYNOMIAL of the GRAPH with that EDGE deleted to v times the POLYNOMIAL of the graph with that EDGE collapsed to a point. Setting $v = -1$ gives the number of distinct VERTEX colorings of the GRAPH. The dichroic POLYNOMIAL of a PLANAR GRAPH can be expressed as the SQUARE BRACKET POLYNOMIAL of the corresponding ALTERNATING LINK by

$$Z_G(q,v) = q^{N/2} B_{L(G)},$$

where N is the number of VERTICES in G. Dichroic POLYNOMIALS for some simple GRAPHS are

$$\begin{aligned} Z_{K_1} &= q \\ Z_{K_2} &= q^2 + vq \\ Z_{K_3} &= q^3 + 3vq^2 + 3v^2q + v^3q. \end{aligned}$$

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 231–235, 1994.

Dido's Problem

Find the figure bounded by a line which has the maximum AREA for a given PERIMETER. The solution is a SEMICIRCLE.

see also ISOPERIMETRIC PROBLEM, ISOVOLUME PROBLEM, PERIMETER, SEMICIRCLE

Diesis

The musical interval by which an octave exceeds three major thirds,

$$\frac{2}{(\frac{5}{4})^3} = \frac{2^7}{5^3} = \frac{128}{125} = 1.024.$$

Taking CONTINUED FRACTION CONVERGENTS of $\log(5/4)/\log(2)$ gives the increasing accurate approximations m/n of m octaves and n major thirds: 1/3, 9/28, 19/59, 47/146, 207/643, 1289/4004, ... (Sloane's A046103 and A046104). Other near equalities of m octaves and n major thirds having

$$R \equiv \frac{2^m}{(\frac{5}{4})^n} = \frac{2^{m+2n}}{5^n}$$

with $|R-1| < 0.02$ are given in the following table.

m	n	Ratio	m	n	Ratio
9	**28**	0.9903520314	104	323	1.012011267
10	31	1.01412048	113	351	1.002247414
18	56	0.9807971462	122	379	0.9925777621
19	**59**	1.004336278	123	382	1.016399628
28	87	0.9946464728	131	407	0.983001403
29	90	1.018517988	132	410	1.006593437
37	115	0.9850501549	141	438	0.9968818549
38	118	1.008691359	150	466	0.9872639701
47	**146**	0.9989595361	151	469	1.010958305
56	174	0.9893216059	160	497	1.001204611
57	177	1.013065324	169	525	0.9915450208
66	205	1.003291302	170	528	1.015342101
75	233	0.9936115791	178	553	0.9819786256
76	236	1.017458257	179	556	1.005546113
84	261	0.9840252458	188	584	0.9958446353
85	264	1.007641852	189	587	1.019744907
94	292	0.9979201548	197	612	0.9862367575
103	320	0.9882922525	198	615	1.00990644

see also COMMA OF DIDYMUS, COMMA OF PYTHAGORAS, SCHISMA

References

Sloane, N. J. A. Sequences A046103 and A046104 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Diffeomorphism

A diffeomorphism is a MAP between MANIFOLDS which is DIFFERENTIABLE and has a DIFFERENTIABLE inverse.

see also ANOSOV DIFFEOMORPHISM, AXIOM A DIFFEOMORPHISM, SYMPLECTIC DIFFEOMORPHISM, TANGENT MAP

Difference

The difference of two numbers n_1 and n_2 is $n_1 - n_2$, where the MINUS sign denotes SUBTRACTION.

see also BACKWARD DIFFERENCE, FINITE DIFFERENCE, FORWARD DIFFERENCE

Difference Equation

A difference equation is the discrete analogue of a DIFFERENTIAL EQUATION. A difference equation involves a FUNCTION with INTEGER-valued arguments $f(n)$ in a form like

$$f(n) - f(n-1) = g(n), \qquad (1)$$

where g is some FUNCTION. The above equation is the discrete analog of the first-order ORDINARY DIFFERENTIAL EQUATION

$$f'(x) = g(x). \qquad (2)$$

Examples of difference equations often arise in DYNAMICAL SYSTEMS. Examples include the iteration involved in the MANDELBROT and JULIA SET definitions,

$$f(n+1) = f(n)^2 + c, \qquad (3)$$

with c a constant, as well as the LOGISTIC EQUATION

$$f(n+1) = rf(n)[1 - f(n)], \qquad (4)$$

with r a constant.

see also FINITE DIFFERENCE, RECURRENCE RELATION

References

Batchelder, P. M. *An Introduction to Linear Difference Equations.* New York: Dover, 1967.

Bellman, R. E. and Cooke, K. L. *Differential-Difference Equations.* New York: Academic Press, 1963.

Beyer, W. H. "Finite Differences." *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 429–460, 1988.

Brand, L. *Differential and Difference Equations.* New York: Wiley, 1966.

Goldberg, S. *Introduction to Difference Equations, with Illustrative Examples from Economics, Psychology, and Sociology.* New York: Dover, 1986.

Levy, H. and Lessman, F. *Finite Difference Equations.* New York: Dover, 1992.

Richtmyer, R. D. and Morton, K. W. *Difference Methods for Initial-Value Problems, 2nd ed.* New York: Interscience Publishers, 1967.

Difference Operator

see BACKWARD DIFFERENCE, FORWARD DIFFERENCE

Difference Quotient

$$\Delta_h f(x) \equiv \frac{f(x+h) - f(x)}{h} = \frac{\Delta f}{h}.$$

It gives the slope of the SECANT LINE passing through $f(x)$ and $f(x+h)$. In the limit $n \to 0$, the difference quotient becomes the PARTIAL DERIVATIVE

$$\lim_{h\to 0} \Delta_{x(h)} f(x,y) = \frac{\partial f}{\partial x}.$$

Difference Set

Let G be a GROUP of ORDER h and D be a set of k elements of G. If the set of differences $d_i - d_j$ contains every NONZERO element of G exactly λ times, then D is a (h,k,λ)-difference set in G of ORDER $n = k - \lambda$. If $\lambda = 1$, the difference set is called planar. The quadratic residues in the GALOIS FIELD $GF(11)$ form a difference set. If there is a difference set of size k in a group G, then $2\binom{k}{2}$ must be a multiple of $|G| - 1$, where $\binom{k}{2}$ is a BINOMIAL COEFFICIENT.

see also BRUCK-RYSER-CHOWLA THEOREM, FIRST MULTIPLIER THEOREM, PRIME POWER CONJECTURE

References

Gordon, D. M. "The Prime Power Conjecture is True for $n < 2{,}000{,}000$." *Electronic J. Combinatorics* **1**, R6, 1–7, 1994. `http://www.combinatorics.org/Volume_1/volume1.html#R6`.

Difference of Successes

If x_1/n_1 and x_2/n_2 are the observed proportions from standard NORMALLY DISTRIBUTED samples with proportion of success θ, then the probability that

$$w \equiv \frac{x_1}{n_1} - \frac{x_2}{n_2} \tag{1}$$

will be as great as observed is

$$P_\delta = 1 - 2\int_0^{|\delta|} \phi(t)\,dt, \tag{2}$$

where

$$\delta \equiv \frac{w}{\sigma_w} \tag{3}$$

$$\sigma_w \equiv \sqrt{\hat\theta(1-\hat\theta)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \tag{4}$$

$$\hat\theta \equiv \frac{x_1+x_2}{n_1+n_2}. \tag{5}$$

Here, $\hat\theta$ is the UNBIASED ESTIMATOR. The SKEWNESS and KURTOSIS of this distribution are

$${\gamma_1}^2 = \frac{(n_1-n_2)^2}{n_1n_2(n_1+n_2)}\,\frac{1-4\hat\theta(1-\hat\theta)}{\hat\theta(1-\hat\theta)} \tag{6}$$

$$\gamma_2 = \frac{{n_1}^2 - n_1n_2 + {n_2}^2}{n_1n_2(n_1+n_2)}\,\frac{1-6\hat\theta(1-\hat\theta)}{\hat\theta(1-\hat\theta)}. \tag{7}$$

Difference Table

A table made by subtracting adjacent entries in a sequence, then repeating the process with those numbers.

see also FINITE DIFFERENCE, QUOTIENT-DIFFERENCE TABLE

Different

Two quantities are said to be different (or "unequal") if they are not EQUAL.

The term "different" also has a technical usage related to MODULES. Let a MODULE M in an INTEGRAL DOMAIN D_1 for $R(\sqrt{D})$ be expressed using a two-element basis as

$$M = [\xi_1, \xi_2],$$

where ξ_1 and ξ_2 are in D_1. Then the different of the MODULE is defined as

$$\Delta = \Delta(M) = \begin{vmatrix} \xi_1 & \xi_2 \\ \xi_1' & \xi_2' \end{vmatrix} = \xi_1\xi_2' - \xi_1'\xi_2.$$

The different $\Delta \neq 0$ IFF ξ_1 and ξ_2 are linearly independent. The DISCRIMINANT is defined as the square of the different.

see also DISCRIMINANT (MODULE), EQUAL, MODULE

References

Cohn, H. *Advanced Number Theory.* New York: Dover, pp. 72–73, 1980.

Different Prime Factors

see DISTINCT PRIME FACTORS

Differentiable

A FUNCTION is said to be differentiable at a point if its DERIVATIVE exists at that point. Let $z = x + iy$ and $f(z) = u(x,y)+iv(x,y)$ on some region G containing the point z_0. If $f(z)$ satisfies the CAUCHY-RIEMANN EQUATIONS and has continuous first PARTIAL DERIVATIVES at z_0, then $f'(z_0)$ exists and is given by

$$f'(z_0) = \lim_{z\to z_0} \frac{f(z)-f(z_0)}{z-z_0},$$

and the function is said to be COMPLEX DIFFERENTIABLE. Amazingly, there exist CONTINUOUS FUNCTIONS which are nowhere differentiable. Two examples are the BLANCMANGE FUNCTION and WEIERSTRAß FUNCTION.

see also BLANCMANGE FUNCTION, CAUCHY-RIEMANN EQUATIONS, COMPLEX DIFFERENTIABLE, CONTINUOUS FUNCTION, DERIVATIVE, PARTIAL DERIVATIVE, WEIERSTRAß FUNCTION

Differentiable Manifold

see SMOOTH MANIFOLD

Differential

A DIFFERENTIAL 1-FORM.

see also EXACT DIFFERENTIAL, INEXACT DIFFERENTIAL

Differential Calculus

That portion of "the" CALCULUS dealing with DERIVATIVES.

see also INTEGRAL CALCULUS

Differential Equation

An equation which involves the DERIVATIVES of a function as well as the function itself. If PARTIAL DERIVATIVES are involved, the equation is called a PARTIAL DIFFERENTIAL EQUATION; if only ordinary DERIVATIVES are present, the equation is called an ORDINARY DIFFERENTIAL EQUATION. Differential equations play an extremely important and useful role in applied math, engineering, and physics, and much mathematical and numerical machinery has been developed for the solution of differential equations.

see also INTEGRAL EQUATION, ORDINARY DIFFERENTIAL EQUATION, PARTIAL DIFFERENTIAL EQUATION

References

Arfken, G. "Differential Equations." Ch. 8 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 437–496, 1985.

Dormand, J. R. *Numerical Methods for Differential Equations: A Computational Approach.* Boca Raton, FL: CRC Press, 1996.

Differential Form

see Differential k-Form

Differential Geometry

Differential geometry is the study of Riemannian Manifolds. Differential geometry deals with metrical notions on Manifolds, while Differential Topology deals with those nonmetrical notions of Manifolds.

see also Differential Topology

References

Eisenhart, L. P. *A Treatise on the Differential Geometry of Curves and Surfaces.* New York: Dover, 1960.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, 1993.

Kreyszig, E. *Differential Geometry.* New York: Dover, 1991.

Lipschutz, M. M. *Theory and Problems of Differential Geometry.* New York: McGraw-Hill, 1969.

Spivak, M. *A Comprehensive Introduction to Differential Geometry, 2nd ed, 5 vols.* Berkeley, CA: Publish or Perish Press, 1979.

Struik, D. J. *Lectures on Classical Differential Geometry.* New York: Dover, 1988.

Weatherburn, C. E. *Differential Geometry of Three Dimensions, 2 vols.* Cambridge, England: Cambridge University Press, 1961.

Differential k-Form

A differential k-form is a Tensor of Rank k which is antisymmetric under exchange of any pair of indices. The number of algebraically independent components in n-D is $\binom{n}{p}$, where this is a Binomial Coefficient. In particular, a 1-form (often simply called a "differential") is a quantity

$$\omega^1 = b_1\,dx_1 + b_2\,dx_2, \tag{1}$$

where $b_1 = b_1(x_1, x_2)$ and $b_2 = b_2(x_1, x_2)$ are the components of a Covariant Tensor. Changing variables from $\mathbf{x}$ to $\mathbf{y}$ gives

$$\omega^1 = \sum_{i=1}^{2} b_i\,dx_i = \sum_{i=1}^{2}\sum_{j=1}^{2} b_i \frac{\partial x_i}{\partial y_j} = \sum_{j=1}^{2} \bar{b}_j\,dy_j, \tag{2}$$

where

$$\bar{b}_j \equiv \sum_{i=1}^{2} b_i \frac{\partial x_i}{\partial y_j}, \tag{3}$$

which is the covariant transformation law. 2-forms can be constructed from the Wedge Product of 1-forms. Let

$$\theta_1 \equiv b_1\,dx_1 + b_2\,dx_2 \tag{4}$$

$$\theta_2 \equiv c_1\,dx_1 + c_2\,dx_2, \tag{5}$$

then $\theta_1 \wedge \theta_2$ is a 2-form denoted ω^2. Changing variables $x_1(y_1, y_2)$ to $x_2(y_1, y_2)$ gives

$$dx_1 = \frac{\partial x_1}{\partial y_1} dy_1 + \frac{\partial x_1}{\partial y_2} dy_2 \tag{6}$$

$$dx_2 = \frac{\partial x_2}{\partial y_1} dy_1 + \frac{\partial x_2}{\partial y_2} dy_2, \tag{7}$$

so

$$\begin{aligned} dx_1 \wedge dx_2 &= \left(\frac{\partial x_1}{\partial y_1}\frac{\partial x_2}{\partial y_2} - \frac{\partial x_1}{\partial y_2}\frac{\partial x_2}{\partial y_1} \right) dy_1 \wedge dy_2 \\ &= \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} dy_1 \wedge dy_2. \end{aligned} \tag{8}$$

Similarly, a 4-form can be constructed from Wedge Products of two 2-forms or four 1-forms

$$\omega^4 = {\omega_1}^2 \wedge {\omega_2}^2 = ({\omega_1}^1 \wedge {\omega_2}^1) \wedge ({\omega_3}^1 \wedge {\omega_4}^1). \tag{9}$$

see also Angle Bracket, Bra, Exterior Derivative, Ket, One-Form, Symplectic Form, Wedge Product

References

Weintraub, S. H. *Differential Forms: A Complement to Vector Calculus.* San Diego, CA: Academic Press, 1996.

Differential Operator

The Operator representing the computation of a Derivative,

$$\tilde{D} \equiv \frac{d}{dx}.$$

The second derivative is then denoted $\tilde{D}^2$, the third $\tilde{D}^3$, etc. The Integral is denoted $\tilde{D}^{-1}$.

see also Convective Derivative, Derivative, Fractional Derivative, Gradient

Differential Structure

see Exotic R4, Exotic Sphere

Differential Topology

The motivating force of Topology, consisting of the study of smooth (differentiable) Manifolds. Differential topology deals with nonmetrical notions of Manifolds, while Differential Geometry deals with metrical notions of Manifolds.

see also Differential Geometry

References

Dieudonné, J. *A History of Algebraic and Differential Topology: 1900–1960.* Boston, MA: Birkhäuser, 1989.

Munkres, J. R. *Elementary Differential Topology.* Princeton, NJ: Princeton University Press, 1963.

Differentiation

The computation of a Derivative.

see also Calculus, Derivative, Integral, Integration

Digamma Function

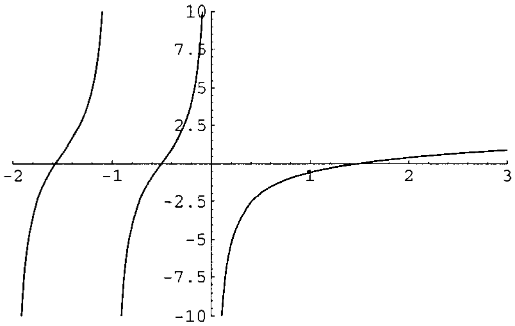

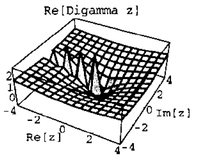

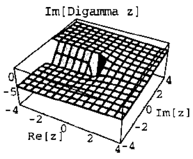

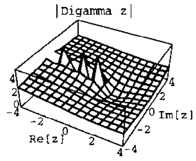

Two notations are used for the digamma function. The $\Psi(z)$ digamma function is defined by

$$\Psi(z) \equiv \frac{d}{dz} \ln \Gamma(z) = \frac{\Gamma'(z)}{\Gamma(z)}, \tag{1}$$

where Γ is the GAMMA FUNCTION, and is the function returned by the function `PolyGamma[z]` in *Mathematica®* (Wolfram Research, Champaign, IL). The F digamma function is defined by

$$F(z) \equiv \frac{d}{dz} \ln z! \tag{2}$$

and is equal to

$$F(z) = \Psi(z+1). \tag{3}$$

From a series expansion of the FACTORIAL function,

$$\begin{aligned} F(z) &= \frac{d}{dz} \lim_{n\to\infty} [\ln n! + z \ln n \\ &\quad - \ln(z+1) - \ln(z+2) - \ldots - \ln(z+n)] \qquad (4) \\ &= \lim_{n\to\infty} \left(\ln n - \frac{1}{z+1} - \frac{1}{z+2} - \ldots - \frac{1}{z+n} \right) \qquad (5) \\ &= -\gamma - \sum_{n=1}^{\infty} \left(\frac{1}{z+n} - \frac{1}{n} \right) \qquad (6) \\ &= -\gamma + \sum_{n=1}^{\infty} \frac{z}{n(n+z)} \qquad (7) \\ &= \ln z + \frac{1}{2z} - \sum_{n=1}^{\infty} \frac{B_{2n}}{2nz^{2n}}, \qquad (8) \end{aligned}$$

where γ is the EULER-MASCHERONI CONSTANT and B_{2n} are BERNOULLI NUMBERS.

The nth DERIVATIVE of $\Psi(z)$ is called the POLYGAMMA FUNCTION and is denoted $\psi_n(z)$. Since the digamma function is the zeroth derivative of $\Psi(z)$ (i.e., the function itself), it is also denoted $\psi_0(z)$.

The digamma function satisfies

$$\Psi(z) = \int_0^\infty \left(\frac{e^{-t}}{t} - \frac{e^{-zt}}{1-e^{-t}} \right) dt. \tag{9}$$

For integral $z \equiv n$,

$$\Psi(n) = -\gamma + \sum_{k=1}^{n-1} \frac{1}{k} = -\gamma + H_{n-1}, \tag{10}$$

where γ is the EULER-MASCHERONI CONSTANT and H_n is a HARMONIC NUMBER. Other identities include

$$\frac{d\Psi}{dz} = \sum_{n=0}^{\infty} \frac{1}{(z+n)^2} \tag{11}$$

$$\Psi(1-z) - \Psi(z) = \pi \cot(\pi z) \tag{12}$$

$$\Psi(z+1) = \Psi(z) + \frac{1}{z} \tag{13}$$

$$\Psi(2z) = \tfrac{1}{2}\Psi(z) + \tfrac{1}{2}\Psi(z+\tfrac{1}{2}) + \ln 2. \tag{14}$$

Special values are

$$\Psi(\tfrac{1}{2}) = -\gamma - 2\ln 2 \tag{15}$$

$$\Psi(1) = -\gamma. \tag{16}$$

At integral values,

$$\psi_0(n+1) = -\gamma + \sum_{k=1}^{n} \frac{1}{k}, \tag{17}$$

and at half-integral values,

$$\psi_0(\tfrac{1}{2} \pm n) = -\ln(4\gamma) + 2\sum_{k=1}^{n} \frac{1}{2k-1}. \tag{18}$$

At rational arguments, $\psi_0(p/q)$ is given by the explicit equation

$$\begin{aligned} \psi_0\left(\frac{p}{q}\right) &= -\gamma - \ln(2q) - \tfrac{1}{2}\pi \cot\left(\frac{p}{q}\pi\right) \\ &\quad + 2\sum_{k=1}^{\lceil q/2 \rceil - 1} \cos\left(\frac{2\pi pk}{q}\right) \ln\left[\sin\left(\frac{\pi k}{q}\right)\right] \qquad (19) \end{aligned}$$

for $0 < p < q$ (Knuth 1973). These give the special values

$$\psi_0(\tfrac{1}{2}) = -\gamma - 2\ln 2 \tag{20}$$

$$\psi_0(\tfrac{1}{3}) = \tfrac{1}{6}(-6\gamma - \pi\sqrt{3} - 9\ln 3) \tag{21}$$

$$\psi_0(\tfrac{2}{3}) = \tfrac{1}{6}(-6\gamma + \pi\sqrt{3} - 9\ln 3) \tag{22}$$

$$\psi_0(\tfrac{1}{4}) = -\gamma - \tfrac{1}{2}\pi - 3\ln 2 \tag{23}$$

$$\psi_0(\tfrac{3}{4}) = \tfrac{1}{2}(-2\gamma + \pi - 6\ln 2) \tag{24}$$

$$\psi_0(1) = -\gamma, \tag{25}$$

where γ is the EULER-MASCHERONI CONSTANT. Sums and differences of $\psi_1(r/s)$ for small integral r and s can be expressed in terms of CATALAN'S CONSTANT and π.

see also GAMMA FUNCTION, HARMONIC NUMBER, HURWITZ ZETA FUNCTION, POLYGAMMA FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Psi (Digamma) Function." §6.3 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 258–259, 1972.

Arfken, G. "Digamma and Polygamma Functions." §10.2 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 549–555, 1985.

Knuth, D. E. *The Art of Computer Programming, Vol. 1: Fundamental Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, p. 94, 1973.

Spanier, J. and Oldham, K. B. "The Digamma Function $\psi(x)$." Ch. 44 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 423–434, 1987.

Digimetic

A CRYPTARITHM in which DIGITS are used to represent other DIGITS.

Digit

The number of digits D in an INTEGER n is the number of numbers in some base (usually 10) required to represent it. The numbers 1 to 9 are therefore single digits, while the numbers 10 to 99 are double digits. Terms such as "double-digit inflation" are occasionally encountered, although this particular usage has thankfully not been needed in the U.S. for some time. The number of (base 10) digits in a number n can be calculated as

$$D = \lfloor \log_{10} n + 1 \rfloor,$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION.

see also 196-ALGORITHM, ADDITIVE PERSISTENCE, DIGITADITION, DIGITAL ROOT, FACTORION, FIGURES, LENGTH (NUMBER), MULTIPLICATIVE PERSISTENCE, NARCISSISTIC NUMBER, SCIENTIFIC NOTATION, SIGNIFICANT DIGITS, SMITH NUMBER

Digitadition

Start with an INTEGER n, known as the GENERATOR. Add the SUM of the GENERATOR's digits to the GENERATOR to obtain the digitadition n'. A number can have more than one GENERATOR. If a number has no GENERATOR, it is called a SELF NUMBER. The sum of all numbers in a digitadition series is given by the last term minus the first plus the sum of the DIGITS of the last.

If the digitadition process is performed on n' to yield *its* digitadition n'', on n'' to yield n''', etc., a single-digit number, known as the DIGITAL ROOT of n, is eventually obtained. The digital roots of the first few integers are 1, 2, 3, 4, 5, 6, 7, 8, 9, 1, 2, 3, 4, 5, 6, 7, 9, 1, ... (Sloane's A010888).

If the process is generalized so that the kth (instead of first) powers of the digits of a number are repeatedly added, a periodic sequence of numbers is eventually obtained for any given starting number n. If the original number n is equal to the sum of the kth powers of its digits, it is called a NARCISSISTIC NUMBER. If the original number is the smallest number in the eventually periodic sequence of numbers in the repeated k-digitaditions, it is called a RECURRING DIGITAL INVARIANT. Both NARCISSISTIC NUMBERS and RECURRING DIGITAL INVARIANTS are relatively rare.

The only possible periods for repeated 2-digitaditions are 1 and 8, and the periods of the first few positive integers are 1, 8, 8, 8, 8, 8, 1, 8, 8, 1, ... The possible periods p for n-digitaditions are summarized in the following table, together with digitaditions for the first few integers and the corresponding sequence numbers.

n	Sloane	*ps*	n-Digitaditions
2	031176	1, 8	1, 8, 8, 8, 8, 8, 1, 8, 8, ...
3	031178	1, 2, 3	1, 1, 1, 3, 1, 1, 1, 1, 1, ...
4	031182	1, 2, 7	1, 7, 7, 7, 7, 7, 7, 7, 7, ...
5	031186	1, 2, 4, 6, 10, 12, 22, 28	1, 12, 22, 4, 10, 22, 28, 10, 22, 1, ...
6	031195	1, 2, 3, 4, 10, 30	1, 10, 30, 30, 30, 10, 10, 10, 3, 1, 10, ...
7	031200	1, 2, 3, 6, 12, 14, 21, 27, 30, 56, 92	1, 92, 14, 30, 92, 56, 6, 92, 56, 1, 92, 27, ...
8	031211	1, 25, 154	1, 25, 154, 154, 154, 154, 25, 154, 154, 1, 25, ...
9	031212	1, 2, 3, 4, 8, 10, 19, 24, 28, 30, 80, 93	1, 30, 93, 1, 19, 80, 4, 30, 80, 1, 30, 93, 4, 10, ...
10	031212	1, 6, 7, 17, 81, 123	1, 30, 93, 1, 19, 80, 4, 30, 80, 1, 30, 93, 4, 10, ...

The numbers having period-1 2-digitaded sequences are also called HAPPY NUMBERS. The first few numbers having period p n-digitaditions are summarized in the following table, together with their sequence numbers.

n	p	Sloane	Members
2	1	007770	1, 7, 10, 13, 19, 23, 28, 31, 32, ...
2	8	031177	2, 3, 4, 5, 6, 8, 9, 11, 12, 14, 15, ...
3	1	031179	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, ...
3	2	031180	49, 94, 136, 163, 199, 244, 316, ...
3	3	031181	4, 13, 16, 22, 25, 28, 31, 40, 46, ...
4	1	031183	1, 10, 12, 17, 21, 46, 64, 71, 100, ...
4	2	031184	66, 127, 172, 217, 228, 271, 282, ...
4	7	031185	2, 3, 4, 5, 6, 7, 8, 9, 11, 13, 14, ...
5	1	031187	1, 10, 100, 145, 154, 247, 274, ...
5	2	031188	133, 139, 193, 199, 226, 262, ...
5	4	031189	4, 37, 40, 55, 73, 124, 142, ...
5	6	031190	16, 61, 106, 160, 601, 610, 778, ...
5	10	031191	5, 8, 17, 26, 35, 44, 47, 50, 53, ...
5	12	031192	2, 11, 14, 20, 23, 29, 32, 38, 41, ...
5	22	031193	3, 6, 9, 12, 15, 18, 21, 24, 27, ...
5	28	031194	7, 13, 19, 22, 25, 28, 31, 34, 43, ...
6	1	011557	1, 10, 100, 1000, 10000, 100000, ...
6	2	031357	3468, 3486, 3648, 3684, 3846, ...
6	3	031196	9, 13, 31, 37, 39, 49, 57, 73, 75, ...
6	4	031197	255, 466, 525, 552, 646, 664, ...
6	10	031198	2, 6, 7, 8, 11, 12, 14, 15, 17, 19, ...
6	30	031199	3, 4, 5, 16, 18, 22, 29, 30, 33, ...
7	1	031201	1, 10, 100, 1000, 1259, 1295, ...
7	2	031202	22, 202, 220, 256, 265, 526, 562, ...
7	3	031203	124, 142, 148, 184, 214, 241, 259, ...
7	6		7, 70, 700, 7000, 70000, 700000, ...
7	12	031204	17, 26, 47, 59, 62, 71, 74, 77, 89, ...
7	14	031205	3, 30, 111, 156, 165, 249, 294, ...
7	21	031206	19, 34, 43, 91, 109, 127, 172, 190, ...
7	27	031207	12, 18, 21, 24, 39, 42, 45, 54, 78, ...
7	30	031208	4, 13, 16, 25, 28, 31, 37, 40, 46, ...
7	56	031209	6, 9, 15, 27, 33, 36, 48, 51, 57, ...
7	92	031210	2, 5, 8, 11, 14, 20, 23, 29, 32, 35, ...
8	1		1, 10, 14, 17, 29, 37, 41, 71, 73, ...
8	25		2, 7, 11, 15, 16, 20, 23, 27, 32, ...
8	154		3, 4, 5, 6, 8, 9, 12, 13, 18, 19, ...
9	1		1, 4, 10, 40, 100, 400, 1000, 1111, ...
9	2		127, 172, 217, 235, 253, 271, 325, ...
9	3		444, 4044, 4404, 4440, 4558, ...
9	4		7, 13, 31, 67, 70, 76, 103, 130, ...
9	8		22, 28, 34, 37, 43, 55, 58, 73, 79, ...
9	10		14, 38, 41, 44, 83, 104, 128, 140, ...
9	19		5, 26, 50, 62, 89, 98, 155, 206, ...
9	24		16, 61, 106, 160, 337, 373, 445, ...
9	28		19, 25, 46, 49, 52, 64, 91, 94, ...
9	30		2, 8, 11, 17, 20, 23, 29, 32, 35, ...
9	80		6, 9, 15, 18, 24, 33, 42, 48, 51, ...
9	93		3, 12, 21, 27, 30, 36, 39, 45, 54, ...
10	1	011557	1, 10, 100, 1000, 10000, 100000, ...
10	6		266, 626, 662, 1159, 1195, 1519, ...
10	7		46, 58, 64, 85, 122, 123, 132, ...
10	17		2, 4, 5, 11, 13, 20, 31, 38, 40, ...
10	81		17, 18, 37, 71, 73, 81, 107, 108, ...
10	123		3, 6, 7, 8, 9, 12, 14, 15, 16, 19, ...

see also 196-ALGORITHM, ADDITIVE PERSISTENCE, DIGIT, DIGITAL ROOT, MULTIPLICATIVE PERSISTENCE, NARCISSISTIC NUMBER, RECURRING DIGITAL INVARIANT

Digital Root

Consider the process of taking a number, adding its DIGITS, then adding the DIGITS of numbers derived from it, etc., until the remaining number has only one DIGIT. The number of additions required to obtain a single DIGIT from a number n is called the ADDITIVE PERSISTENCE of n, and the DIGIT obtained is called the digital root of n.

For example, the sequence obtained from the starting number 9876 is (9876, 30, 3), so 9876 has an ADDITIVE PERSISTENCE of 2 and a digital root of 3. The digital roots of the first few integers are 1, 2, 3, 4, 5, 6, 7, 8, 9, 1, 2, 3, 4, 5, 6, 7, 9, 1, ... (Sloane's A010888). The digital root of an INTEGER n can therefore be computed without actually performing the iteration using the simple congruence formula

$$\begin{cases} n \pmod 9 & n \not\equiv 0 \pmod 9 \\ 9 & n \equiv 0 \pmod 9. \end{cases}$$

see also ADDITIVE PERSISTENCE, DIGITADITION, KAPREKAR NUMBER, MULTIPLICATIVE DIGITAL ROOT, MULTIPLICATIVE PERSISTENCE, NARCISSISTIC NUMBER, RECURRING DIGITAL INVARIANT, SELF NUMBER

References

Sloane, N. J. A. Sequences A010888 and A007612/M1114 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Digon

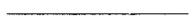

The DEGENERATE POLYGON (corresponding to a LINE SEGMENT) with SCHLÄFLI SYMBOL $\{2\}$.

see also LINE SEGMENT, POLYGON, TRIGONOMETRY VALUES—$\pi/2$

Digraph

see DIRECTED GRAPH

Dihedral Angle

The ANGLE between two PLANES. The dihedral angle between the planes

$$A_1x + B_1y + C_1z + D_1 = 0 \tag{1}$$

$$A_2x + B_2y + C_2z + D_2 = 0 \tag{2}$$

is

$$\cos\theta = \frac{A_1A_2 + B_1B_2 + C_1C_2}{\sqrt{A_1{}^2 + B_1{}^2 + C_1{}^2}\sqrt{A_2{}^2 + B_2{}^2 + C_2{}^2}}. \tag{3}$$

see also ANGLE, PLANE, VERTEX ANGLE

Dihedral Group

A GROUP of symmetries for an n-sided REGULAR POLYGON, denoted D_n. The ORDER of D_n is $2n$.

see also FINITE GROUP—D_3, FINITE GROUP—D_4

References

Arfken, G. "Dihedral Groups, D_n." *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 248, 1985.

Lomont, J. S. "Dihedral Groups." §3.10.B in *Applications of Finite Groups.* New York: Dover, pp. 78–80, 1987.

Dijkstra's Algorithm

An ALGORITHM for finding the shortest path between two VERTICES.

see also FLOYD'S ALGORITHM

Dijkstra Tree

The shortest path-spanning TREE from a VERTEX of a GRAPH.

Dilation

An AFFINE TRANSFORMATION in which the scale is reduced. A dilation is also known as a CONTRACTION or HOMOTHECY. Any dilation which is not a simple translation has a unique FIXED POINT. The opposite of a dilation is an EXPANSION.

see also AFFINE TRANSFORMATION, EXPANSION, HOMOTHECY

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 94–95, 1967.

Dilemma

Informally, a situation in which a decision must be made from several alternatives, none of which is obviously the optimal one. In formal LOGIC, a dilemma is a specific type of argument using two conditional statements which may take the form of a CONSTRUCTIVE DILEMMA or a DESTRUCTIVE DILEMMA.

see also CONSTRUCTIVE DILEMMA, DESTRUCTIVE DILEMMA, MONTY HALL PROBLEM, PARADOX, PRISONER'S DILEMMA

Dilogarithm

A special case of the POLYLOGARITHM $\operatorname{Li}_n(z)$ for $n = 2$. It is denoted $\operatorname{Li}_2(z)$, or sometimes $L_2(z)$, and is defined by the sum

$$\operatorname{Li}_2(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^2}$$

or the integral

$$\operatorname{Li}_2(z) \equiv \int_z^0 \frac{\ln(1-t)\,dt}{t}.$$

There are several remarkable identities involving the POLYLOGARITHM function.

see also ABEL'S FUNCTIONAL EQUATION, POLYLOGARITHM, SPENCE'S INTEGRAL

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Dilogarithm." §27.7 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 1004–1005, 1972.

Dilworth's Lemma

The WIDTH of a set P is equal to the minimum number of CHAINS needed to COVER P. Equivalently, if a set P of $ab+1$ elements is PARTIALLY ORDERED, then P contains a CHAIN of size $a+1$ or an ANTICHAIN of size $b+1$. Letting N be the CARDINALITY of P, W the WIDTH, and L the LENGTH, this last statement says $N \leq LW$. Dilworth's lemma is a generalization of the ERDŐS-SZEKERES THEOREM. RAMSEY'S THEOREM generalizes Dilworth's Lemma.

see also COMBINATORICS, ERDŐS-SZEKERES THEOREM, RAMSEY'S THEOREM

Dilworth's Theorem

see DILWORTH'S LEMMA

Dimension

The notion of dimension is important in mathematics because it gives a precise parameterization of the conceptual or visual complexity of any geometric object. In fact, the concept can even be applied to abstract objects which cannot be directly visualized. For example, the notion of time can be considered as one-dimensional, since it can be thought of as consisting of only "now," "before" and "after." Since "before" and "after," regardless of how far back or how far into the future they are, are extensions, time is like a line, a 1-dimensional object.

To see how lower and higher dimensions relate to each other, take any geometric object (like a POINT, LINE, CIRCLE, PLANE, etc.), and "drag" it in an opposing direction (drag a POINT to trace out a LINE, a LINE to trace out a box, a CIRCLE to trace out a CYLINDER, a DISK to a solid CYLINDER, etc.). The result is an object which is qualitatively "larger" than the previous object, "qualitative" in the sense that, regardless of how you drag the original object, you always trace out an object of the same "qualitative size." The POINT could be made into a straight LINE, a CIRCLE, a HELIX, or some other CURVE, but all of these objects are qualitatively of the same dimension. The notion of dimension was invented for the purpose of measuring this "qualitative" topological property.

Making things a bit more formal, finite collections of objects (e.g., points in space) are considered 0-dimensional. Objects that are "dragged" versions of 0-dimensional objects are then called 1-dimensional. Similarly, objects which are dragged 1-dimensional objects are 2-dimensional, and so on. Dimension is formalized in mathematics as the intrinsic dimension of a TOPOLOGICAL SPACE. This dimension is called the LEBESGUE COVERING DIMENSION (also known simply as the TOPOLOGICAL DIMENSION). The archetypal example is EUCLIDEAN n-space $\mathbb{R}^n$, which has topological dimension n. The basic ideas leading up to this result (including the DIMENSION INVARIANCE THEOREM, DOMAIN INVARIANCE THEOREM, and LEBESGUE COVERING DIMENSION) were developed by Poincaré, Brouwer, Lebesgue, Urysohn, and Menger.

There are several branchings and extensions of the notion of topological dimension. Implicit in the notion of the LEBESGUE COVERING DIMENSION is that dimension, in a sense, is a measure of how an object fills space. If it takes up a lot of room, it is higher dimensional, and if it takes up less room, it is lower dimensional. HAUSDORFF DIMENSION (also called FRACTAL DIMENSION) is a fine tuning of this definition that allows notions of objects with dimensions other than INTEGERS. FRACTALS are objects whose HAUSDORFF DIMENSION is different from their TOPOLOGICAL DIMENSION.

The concept of dimension is also used in ALGEBRA, primarily as the dimension of a VECTOR SPACE over a FIELD. This usage stems from the fact that VECTOR SPACES over the reals were the first VECTOR SPACES to be studied, and for them, their topological dimension can be calculated by purely algebraic means as the CARDINALITY of a maximal linearly independent subset. In particular, the dimension of a SUBSPACE of $\mathbb{R}^n$ is equal to the number of LINEARLY INDEPENDENT VECTORS needed to generate it (i.e., the number of VECTORS in its BASIS). Given a transformation A of $\mathbb{R}^n$,

$$\dim[\mathrm{Range}(A)] + \dim[\mathrm{Null}(A)] = \dim(\mathbb{R}^n).$$

see also CAPACITY DIMENSION, CODIMENSION, CORRELATION DIMENSION, EXTERIOR DIMENSION, FRACTAL DIMENSION, HAUSDORFF DIMENSION, HAUSDORFF-BESICOVITCH DIMENSION, KAPLAN-YORKE DIMENSION, KRULL DIMENSION, LEBESGUE COVERING DIMENSION, LEBESGUE DIMENSION, LYAPUNOV DIMENSION, POSET DIMENSION, q-DIMENSION, SIMILARITY DIMENSION, TOPOLOGICAL DIMENSION

References

Abbott, E. A. *Flatland: A Romance of Many Dimensions.* New York: Dover, 1992.

Hinton, C. H. *The Fourth Dimension.* Pomeroy, WA: Health Research, 1993.

Manning, H. *The Fourth Dimension Simply Explained.* Magnolia, MA: Peter Smith, 1990.

Manning, H. *Geometry of Four Dimensions.* New York: Dover, 1956.

Neville, E. H. *The Fourth Dimension.* Cambridge, England: Cambridge University Press, 1921.

Rucker, R. von Bitter. *The Fourth Dimension: A Guided Tour of the Higher Universes.* Boston, MA: Houghton Mifflin, 1984.

Sommerville, D. M. Y. *An Introduction to the Geometry of n Dimensions.* New York: Dover, 1958.

Dimension Axiom

One of the EILENBERG-STEENROD AXIOMS. Let X be a single point space. $H_n(X) = 0$ unless $n = 0$, in which case $H_0(X) = G$ where G are some GROUPS. The H_0 are called the COEFFICIENTS of the HOMOLOGY THEORY $H(\cdot)$.

see also EILENBERG-STEENROD AXIOMS, HOMOLOGY (TOPOLOGY)

Dimension Invariance Theorem

$\mathbb{R}^n$ is HOMEOMORPHIC to $\mathbb{R}^m$ IFF $n = m$. This theorem was first proved by Brouwer.

see also DOMAIN INVARIANCE THEOREM

Dimensionality Theorem

For a finite GROUP of h elements with an n_ith dimensional ith irreducible representation,

$$\sum_i {n_i}^2 = h.$$

Diminished Polyhedron

A UNIFORM POLYHEDRON with pieces removed.

Diminished Rhombicosidodecahedron

see JOHNSON SOLID

Dini Expansion

An expansion based on the ROOTS of

$$x^{-n}[xJ_n'(x) + HJ_n(x)] = 0,$$

where $J_n(x)$ is a BESSEL FUNCTION OF THE FIRST KIND, is called a Dini expansion.

see also BESSEL FUNCTION FOURIER EXPANSION

References

Bowman, F. *Introduction to Bessel Functions.* New York: Dover, p. 109, 1958.

Dini's Surface

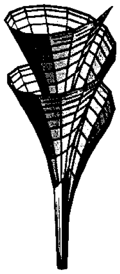

A surface of constant NEGATIVE CURVATURE obtained by twisting a PSEUDOSPHERE and given by the parametric equations

$$x = a\cos u \sin v \quad (1)$$

$$y = a\sin u \sin v \quad (2)$$

$$z = a\{\cos v + \ln[\tan(\tfrac{1}{2}v)]\} + bu. \quad (3)$$

The above figure corresponds to $a = 1$, $b = 0.2$, $u \in [0, 4\pi]$, and $v \in (0, 2]$.

see also PSEUDOSPHERE

References

Geometry Center. "Dini's Surface." `http://www.geom.umn.edu/zoo/diffgeom/surfspace/dini/`.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 494–495, 1993.

Nordstrand, T. "Dini's Surface." `http://www.uib.no/people/nfytn/dintxt.htm`.

Dini's Test

A test for the convergence of FOURIER SERIES. Let

$$\phi_x(t) \equiv f(x+t) + f(x-t) - 2f(x),$$

then if

$$\int_0^\pi \frac{|\phi_x(t)|\,dt}{t}$$

is FINITE, the FOURIER SERIES converges to $f(x)$ at x.

see also FOURIER SERIES

References

Sansone, G. *Orthogonal Functions, rev. English ed.* New York: Dover, pp. 65–68, 1991.

Dinitz Problem

Given any assignment of n-element sets to the n^2 locations of a square $n \times n$ array, is it always possible to find a PARTIAL LATIN SQUARE? The fact that such a PARTIAL LATIN SQUARE can always be found for a 2×2 array can be proven analytically, and techniques were developed which also proved the existence for 4×4 and 6×6 arrays. However, the general problem eluded solution until it was answered in the affirmative by Galvin in 1993 using results of Janssen (1993ab) and F. Maffray.

see also PARTIAL LATIN SQUARE

References

Chetwynd, A. and Häggkvist, R. "A Note on List-Colorings." *J. Graph Th.* **13**, 87–95, 1989.

Cipra, B. "Quite Easily Done." In *What's Happening in the Mathematical Sciences* **2**, pp. 41–46, 1994.

Erdős, P.; Rubin, A.; and Taylor, H. "Choosability in Graphs." *Congr. Numer.* **26**, 125–157, 1979.

Häggkvist, R. "Towards a Solution of the Dinitz Problem?" *Disc. Math.* **75**, 247–251, 1989.

Janssen, J. C. M. "The Dinitz Problem Solved for Rectangles." *Bull. Amer. Math. Soc.* **29**, 243–249, 1993a.

Janssen, J. C. M. *Even and Odd Latin Squares.* Ph.D. thesis. Lehigh University, 1993b.

Kahn, J. "Recent Results on Some Not-So-Recent Hypergraph Matching and Covering Problems." *Proceedings of the Conference on Extremal Problems for Finite Sets.* Visegràd, Hungary, 1991.

Kahn, J. "Coloring Nearly-Disjoint Hypergraphs with $n + o(n)$ Colors." *J. Combin. Th. Ser. A* **59**, 31–39, 1992.

Diocles's Cissoid

see CISSOID OF DIOCLES

Diophantine Equation

An equation in which only INTEGER solutions are allowed. HILBERT'S 10TH PROBLEM asked if a technique for solving a general Diophantine existed. A general method exists for the solution of first degree Diophantine equations. However, the impossibility of obtaining a general solution was proven by Julia Robinson and Martin Davis in 1970, following proof of the result that the equation $n = F_{2m}$ (where F_{2m} is a FIBONACCI NUMBER) is Diophantine by Yuri Matijasevič (Matijasevič 1970, Davis 1973, Davis and Hersh 1973, Matijasevič 1993).

No general method is known for quadratic or higher Diophantine equations. Jones and Matijasevič (1982) proved that no ALGORITHMS can exist to determine if an arbitrary Diophantine equation in nine variables has solutions. Ogilvy and Anderson (1988) give a number of Diophantine equations with known and unknown solutions.

D. Wilson has compiled a list of the smallest nth POWERS which are the sums of n *distinct* smaller nth POWERS. The first few are 3, 5, 6, 15, 12, 25, 40, ... (Sloane's A030052):

$$\begin{aligned}
3^1 &= 1^1 + 2^1 \\
5^2 &= 3^2 + 4^2 \\
6^3 &= 3^3 + 4^3 + 5^3 \\
15^4 &= 4^4 + 6^4 + 8^4 + 9^4 + 14^4 \\
12^5 &= 4^5 + 5^5 + 6^5 + 7^5 + 9^5 + 11^5 \\
25^6 &= 1^6 + 2^6 + 3^6 + 5^6 + 6^6 + 7^6 + 8^6 + 9^6 + 10^6 \\
&\quad + 12^6 + 13^6 + 15^6 + 16^6 + 17^6 + 18^6 + 23^6 \\
40^7 &= 1^7 + 3^7 + 5^7 + 9^7 + 12^7 + 14^7 + 16^7 + 17^7 \\
&\quad + 18^7 + 20^7 + 21^7 + 22^7 + 25^7 + 28^7 + 39^7 \\
84^8 &= 1^8 + 2^8 + 3^8 + 5^8 + 7^8 + 9^8 + 10^8 + 11^8 \\
&\quad + 12^8 + 13^8 + 14^8 + 15^8 + 16^8 + 17^8 + 18^8 \\
&\quad + 19^8 + 21^8 + 23^8 + 24^8 + 25^8 + 26^8 + 27^8 \\
&\quad + 29^8 + 32^8 + 33^8 + 35^8 + 37^8 + 38^8 + 39^8 \\
&\quad + 41^8 + 42^8 + 43^8 + 45^8 + 46^8 + 47^8 + 48^8 \\
&\quad + 49^8 + 51^8 + 52^8 + 53^8 + 57^8 + 58^8 + 59^8 \\
&\quad + 61^8 + 63^8 + 69^8 + 73^8 \\
47^9 &= 1^9 + 2^9 + 4^9 + 7^9 + 11^9 + 14^9 + 15^9 + 18^9 \\
&\quad + 26^9 + 27^9 + 30^9 + 31^9 + 32^9 + 33^9 \\
&\quad + 36^9 + 38^9 + 39^9 + 43^9 \\
63^{10} &= 1^{10} + 2^{10} + 4^{10} + 5^{10} + 6^{10} + 8^{10} + 12^{10} \\
&\quad + 15^{10} + 16^{10} + 17^{10} + 20^{10} + 21^{10} + 25^{10} \\
&\quad + 26^{10} + 27^{10} + 28^{10} + 30^{10} + 36^{10} + 37^{10} \\
&\quad + 38^{10} + 40^{10} + 51^{10} + 62^{10}.
\end{aligned}$$

see also ABC CONJECTURE, ARCHIMEDES' CATTLE PROBLEM, BACHET EQUATION, BRAHMAGUPTA'S

Problem, Cannonball Problem, Catalan's Problem, Diophantine Equation—Linear, Diophantine Equation—Quadratic, Diophantine Equation—Cubic, Diophantine Equation—Quartic, Diophantine Equation—5th Powers, Diophantine Equation—6th Powers, Diophantine Equation—7th Powers, Diophantine Equation—8th Powers, Diophantine Equation—9th Powers, Diophantine Equation—10th Powers, Diophantine Equation—nth Powers, Diophantus Property, Euler Brick, Euler Quartic Conjecture, Fermat's Last Theorem, Fermat Sum Theorem, Genus Theorem, Hurwitz Equation, Markov Number, Monkey and Coconut Problem, Multigrade Equation, p-adic Number, Pell Equation, Pythagorean Quadruple, Pythagorean Triple

References

Beiler, A. H. *Recreations in the Theory of Numbers: The Queen of Mathematics Entertains.* New York: Dover, 1966.

Carmichael, R. D. *The Theory of Numbers, and Diophantine Analysis.* New York: Dover, 1959.

Chen, S. "Equal Sums of Like Powers: On the Integer Solution of the Diophantine System." `http://www.nease.net/~chin/eslp/`.

Chen, S. "References." `http://www.nease.net/~chin/eslp/referenc.htm`

Davis, M. "Hilbert's Tenth Problem is Unsolvable." *Amer. Math. Monthly* **80**, 233–269, 1973.

Davis, M. and Hersh, R. "Hilbert's 10th Problem." *Sci. Amer.*, pp. 84–91, Nov. 1973.

Dörrie, H. "The Fermat-Gauss Impossibility Theorem." §21 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 96–104, 1965.

Guy, R. K. "Diophantine Equations." Ch. D in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 139–198, 1994.

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, 1979.

Hunter, J. A. H. and Madachy, J. S. "Diophantos and All That." Ch. 6 in *Mathematical Diversions.* New York: Dover, pp. 52–64, 1975.

Ireland, K. and Rosen, M. "Diophantine Equations." Ch. 17 in *A Classical Introduction to Modern Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 269–296, 1990.

Jones, J. P. and Matijasevič, Yu. V. "Exponential Diophantine Representation of Recursively Enumerable Sets." *Proceedings of the Herbrand Symposium, Marseilles, 1981.* Amsterdam, Netherlands: North-Holland, pp. 159–177, 1982.

Lang, S. *Introduction to Diophantine Approximations, 2nd ed.* New York: Springer-Verlag, 1995.

Matijasevič, Yu. V. "Solution to of the Tenth Problem of Hilbert." *Mat. Lapok* **21**, 83–87, 1970.

Matijasevič, Yu. V. *Hilbert's Tenth Problem.* Cambridge, MA: MIT Press, 1993.

Mordell, L. J. *Diophantine Equations.* New York: Academic Press, 1969.

Nagel, T. *Introduction to Number Theory.* New York: Wiley, 1951.

Ogilvy, C. S. and Anderson, J. T. "Diophantine Equations." Ch. 6 in *Excursions in Number Theory.* New York: Dover, pp. 65–83, 1988.

Sloane, N. J. A. Sequence A030052 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Diophantine Equation—5th Powers

The 2-1 fifth-order Diophantine equation

$$A^5 + B^5 = C^5 \tag{1}$$

is a special case of Fermat's Last Theorem with $n = 5$, and so has no solution. No solutions to the 2-2 equation

$$A^5 + B^5 = C^5 + D^5 \tag{2}$$

are known, despite the fact that sums up to 1.02×10^{26} have been checked (Guy 1994, p. 140), improving on the results on Lander *et al.* (1967), who checked up to 2.8×10^{14}. (In fact, no solutions are known for Powers of 6 or 7 either.)

No solutions to the 3-1 equation

$$A^5 + B^5 + C^5 = D^5 \tag{3}$$

are known (Lander *et al.* 1967), nor are any 3-2 solutions up to 8×10^{12} (Lander *et al.* 1967).

Parametric solutions are known for the 3-3 (Guy 1994, pp. 140 and 142). Swinnerton-Dyer (1952) gave two parametric solutions to the 3-3 equation but, forty years later, W. Gosper discovered that the second scheme has an unfixable bug. The smallest primitive 3-3 solutions are

$$24^5 + 28^5 + 67^5 = 3^5 + 54^5 + 62^5 \tag{4}$$
$$18^5 + 44^5 + 66^5 = 13^5 + 51^5 + 64^5 \tag{5}$$
$$21^5 + 43^5 + 76^5 = 8^5 + 62^5 + 68^5 \tag{6}$$
$$56^5 + 67^5 + 83^5 = 53^5 + 72^5 + 81^5 \tag{7}$$
$$49^5 + 75^5 + 107^5 = 39^5 + 92^5 + 100^5 \tag{8}$$

(Moessner 1939, Moessner 1948, Lander *et al.* 1967).

For 4 fifth Powers, we have the 4-1 equation

$$27^5 + 84^5 + 110^5 + 133^5 = 144^5 \tag{9}$$

(Lander and Parkin 1967, Lander *et al.* 1967), but it is not known if there is a parametric solution (Guy 1994, p. 140). Sastry's (1934) 5-1 solution gives some 4-2 solutions. The smallest primitive 4-2 solutions are

$$4^5 + 10^5 + 20^5 + 28^5 = 3^5 + 29^5 \tag{10}$$
$$5^5 + 13^5 + 25^5 + 37^5 = 12^5 + 38^5 \tag{11}$$
$$26^5 + 29^5 + 35^5 + 50^5 = 28^5 + 52^5 \tag{12}$$
$$5^5 + 25^5 + 62^5 + 63^5 = 61^5 + 64^5 \tag{13}$$
$$6^5 + 50^5 + 53^5 + 82^5 = 16^5 + 85^5 \tag{14}$$
$$56^5 + 63^5 + 72^5 + 86^5 = 31^5 + 96^5 \tag{15}$$
$$44^5 + 58^5 + 67^5 + 94^5 = 14^5 + 99^5 \tag{16}$$
$$11^5 + 13^5 + 37^5 + 99^5 = 63^5 + 97^5 \tag{17}$$
$$48^5 + 57^5 + 76^5 + 100^5 = 25^5 + 106^5 \tag{18}$$
$$58^5 + 76^5 + 79^5 + 102^5 = 54^5 + 111^5 \tag{19}$$

(Rao 1934, Moessner 1948, Lander *et al.* 1967).

A two-parameter solution to the 4-3 equation was given by Xeroudakes and Moessner (1958). Gloden (1949) also gave a parametric solution. The smallest solution is

$$1^5 + 8^5 + 14^5 + 27^5 = 3^5 + 22^5 + 25^5 \qquad (20)$$

(Rao 1934, Lander *et al.* 1967). Several parametric solutions to the 4-4 equation were found by Xeroudakes and Moessner (1958). The smallest 4-4 solution is

$$5^5 + 6^5 + 6^5 + 8^5 = 4^5 + 7^5 + 7^5 + 7^5 \qquad (21)$$

(Rao 1934, Lander *et al.* 1967). The first 4-4-4 equation is

$$\begin{aligned} 3^5 + 48^5 + 52^5 + 61^5 &= 13^5 + 36^5 + 51^5 + 64^5 \\ &= 18^5 + 36^5 + 44^5 + 66^5 \qquad (22) \end{aligned}$$

(Lander *et al.* 1967).

Sastry (1934) found a 2-parameter solution for 5-1 equations

$$\begin{aligned} &(75v^5 - u^5)^5 + (u^5 + 25v^5)^5 + (u^5 - 25v^5)^5 \\ &\quad +(10u^3v^2)^5 + (50uv^4)^5 = (u^5 + 75v^5)^5 \qquad (23) \end{aligned}$$

(quoted in Lander and Parkin 1967), and Lander and Parkin (1967) found the smallest numerical solutions. Lander *et al.* (1967) give a list of the smallest solutions, the first few being

$$19^5 + 43^5 + 46^5 + 47^5 + 67^5 = 72^5 \qquad (24)$$
$$21^5 + 23^5 + 37^5 + 79^5 + 84^5 = 94^5 \qquad (25)$$
$$7^5 + 43^5 + 57^5 + 80^5 + 100^5 = 107^5 \qquad (26)$$
$$8^5 + 120^5 + 191^5 + 259^5 + 347^5 = 365^5 \qquad (27)$$
$$79^5 + 202^5 + 258^5 + 261^5 + 395^5 = 415^5 \qquad (28)$$
$$4^5 + 26^5 + 139^5 + 296^5 + 412^5 = 427^5 \qquad (29)$$
$$31^5 + 105^5 + 139^5 + 314^5 + 416^5 = 435^5 \qquad (30)$$
$$54^5 + 91^5 + 101^5 + 404^5 + 430^5 = 480^5 \qquad (31)$$
$$19^5 + 201^5 + 347^5 + 388^5 + 448^5 = 503^5 \qquad (32)$$
$$159^5 + 172^5 + 200^5 + 356^5 + 513^5 = 530^5 \qquad (33)$$
$$218^5 + 276^5 + 385^5 + 409^5 + 495^5 = 553^5 \qquad (34)$$
$$2^5 + 298^5 + 351^5 + 474^5 + 500^5 = 575^5 \qquad (35)$$

(Lander and Parkin 1967, Lander *et al.* 1967).

The smallest primitive 5-2 solutions are

$$4^5 + 5^5 + 7^5 + 16^5 + 21^5 = 1^5 + 22^5 \qquad (36)$$
$$9^5 + 11^5 + 14^5 + 18^5 + 30^5 = 23^5 + 29^5 \qquad (37)$$
$$10^5 + 14^5 + 26^5 + 31^5 + 33^5 = 16^5 + 38^5 \qquad (38)$$
$$4^5 + 22^5 + 29^5 + 35^5 + 36^5 = 24^5 + 42^5 \qquad (39)$$
$$8^5 + 15^5 + 17^5 + 19^5 + 45^5 = 30^5 + 44^5 \qquad (40)$$
$$5^5 + 6^5 + 26^5 + 27^5 + 44^5 = 36^5 + 42^5 \qquad (41)$$

(Rao 1934, Lander *et al.* 1967).

The 6-1 equation has solutions

$$4^5 + 5^5 + 6^5 + 7^5 + 9^5 + 11^5 = 12^5 \qquad (42)$$
$$5^5 + 10^5 + 11^5 + 16^5 + 19^5 + 29^5 = 30^5 \qquad (43)$$
$$15^5 + 16^5 + 17^5 + 22^5 + 24^5 + 28^5 = 32^5 \qquad (44)$$
$$13^5 + 18^5 + 23^5 + 31^5 + 36^5 + 66^5 = 67^5 \qquad (45)$$
$$7^5 + 20^5 + 29^5 + 31^5 + 34^5 + 66^5 = 67^5 \qquad (46)$$
$$22^5 + 35^5 + 48^5 + 58^5 + 61^5 + 64^5 = 78^5 \qquad (47)$$
$$4^5 + 13^5 + 19^5 + 20^5 + 67^5 + 96^5 = 99^5 \qquad (48)$$
$$6^5 + 17^5 + 60^5 + 64^5 + 73^5 + 89^5 = 99^5 \qquad (49)$$

(Martin 1887, 1888, Lander and Parkin 1967, Lander *et al.* 1967).

The smallest 7-1 solution is

$$1^5 + 7^5 + 8^5 + 14^5 + 15^5 + 18^5 + 20^5 = 23^5 \qquad (50)$$

(Lander *et al.* 1967).

References

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, p. 95, 1994.

Gloden, A. "Über mehrgeradige Gleichungen." *Arch. Math.* **1**, 482–483, 1949.

Guy, R. K. "Sums of Like Powers. Euler's Conjecture." §D1 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 139–144, 1994.

Lander, L. J. and Parkin, T. R. "A Counterexample to Euler's Sum of Powers Conjecture." *Math. Comput.* **21**, 101–103, 1967.

Lander, L. J.; Parkin, T. R.; and Selfridge, J. L. "A Survey of Equal Sums of Like Powers." *Math. Comput.* **21**, 446–459, 1967.

Martin, A. "Methods of Finding *n*th-Power Numbers Whose Sum is an *n*th Power; With Examples." *Bull. Philos. Soc. Washington* **10**, 107–110, 1887.

Martin, A. *Smithsonian Misc. Coll.* **33**, 1888.

Martin, A. "About Fifth-Power Numbers whose Sum is a Fifth Power." *Math. Mag.* **2**, 201–208, 1896.

Moessner, A. "Einige numerische Identitäten." *Proc. Indian Acad. Sci. Sect. A* **10**, 296–306, 1939.

Moessner, A. "Alcune richerche di teoria dei numeri e problemi diofantei." *Bol. Soc. Mat. Mexicana* **2**, 36–39, 1948.

Rao, K. S. "On Sums of Fifth Powers." *J. London Math. Soc.* **9**, 170–171, 1934.

Sastry, S. "On Sums of Powers." *J. London Math. Soc.* **9**, 242–246, 1934.

Swinnerton-Dyer, H. P. F. "A Solution of $A^5 + B^5 + C^5 = D^5 + E^5 + F^5$." *Proc. Cambridge Phil. Soc.* **48**, 516–518, 1952.

Xeroudakes, G. and Moessner, A. "On Equal Sums of Like Powers." *Proc. Indian Acad. Sci. Sect. A* **48**, 245–255, 1958.

Diophantine Equation—6th Powers

The 2-1 equation

$$A^6 + B^6 = C^6 \qquad (1)$$

is a special case of FERMAT'S LAST THEOREM with $n = 6$, and so has no solution. Ekl (1996) has searched and found no solutions to the 2-2

$$A^6 + B^6 = C^6 + D^6 \tag{2}$$

with sums less than 7.25×10^{26}.

No solutions are known to the 3-1 or 3-2 equations. However, parametric solutions are known for the 3-3 equation

$$A^6 + B^6 + C^6 = D^6 + E^6 + F^6 \tag{3}$$

(Guy 1994, pp. 140 and 142). Known solutions are

$$3^6 + 19^6 + 22^6 = 10^6 + 15^6 + 23^6 \tag{4}$$
$$36^6 + 37^6 + 67^6 = 15^6 + 52^6 + 65^6 \tag{5}$$
$$33^6 + 47^6 + 74^6 = 23^6 + 54^6 + 73^6 \tag{6}$$
$$32^6 + 43^6 + 81^6 = 3^6 + 55^6 + 80^6 \tag{7}$$
$$37^6 + 50^6 + 81^6 = 11^6 + 65^6 + 78^6 \tag{8}$$
$$25^6 + 62^6 + 138^6 = 82^6 + 92^6 + 135^6 \tag{9}$$
$$51^6 + 113^6 + 136^6 = 40^6 + 125^6 + 129^6 \tag{10}$$
$$71^6 + 92^6 + 147^6 = 1^6 + 132^6 + 133^6 \tag{11}$$
$$111^6 + 121^6 + 230^6 = 26^6 + 169^6 + 225^6 \tag{12}$$
$$75^6 + 142^6 + 245^6 = 14^6 + 163^6 + 243^6 \tag{13}$$

(Rao 1934, Lander *et al.* 1967).

No solutions are known to the 4-1 or 4-2 equations. The smallest primitive 4-3 solutions are

$$41^6 + 58^6 + 73^6 = 15^6 + 32^6 + 65^6 + 70^6 \tag{14}$$
$$61^6 + 62^6 + 85^6 = 52^6 + 56^6 + 69^6 + 83^6 \tag{15}$$
$$61^6 + 74^6 + 85^6 = 26^6 + 56^6 + 71^6 + 87^6 \tag{16}$$
$$11^6 + 88^6 + 90^6 = 21^6 + 74^6 + 78^6 + 92^6 \tag{17}$$
$$26^6 + 83^6 + 95^6 = 23^6 + 24^6 + 28^6 + 101^6 \tag{18}$$

(Lander *et al.* 1967). Moessner (1947) gave three parametric solutions to the 4-4 equation. The smallest 4-4 solution is

$$2^6 + 2^6 + 9^6 + 9^6 = 3^6 + 5^6 + 6^6 + 10^6 \tag{19}$$

(Rao 1934, Lander *et al.* 1967). The smallest 4-4-4 solution is

$$\begin{aligned} 1^6 + 34^6 + 49^6 + 111^6 &= 7^6 + 43^6 + 69^6 + 110^6 \\ &= 18^6 + 25^6 + 77^6 + 109^6 \end{aligned} \tag{20}$$

(Lander *et al.* 1967).

No n-1 solutions are known for $n \leq 6$ (Lander *et al.* 1967). No solution to the 5-1 equation is known (Guy 1994, p. 140) or the 5-2 equation.

No solutions are known to the 6-1 or 6-2 equations.

The smallest 7-1 solution is

$$74^6 + 234^6 + 402^6 + 474^6 + 702^6 + 894^6 + 1077^6 = 1141^6 \tag{21}$$

(Lander *et al.* 1967). The smallest 7-2 solution is

$$18^6 + 22^6 + 36^6 + 58^6 + 69^6 + 78^6 + 78^6 = 56^6 + 91^6 \tag{22}$$

(Lander *et al.* 1967).

The smallest primitive 8-1 solutions are

$$\begin{aligned} 8^6 + 12^6 + 30^6 + 78^6 + 102^6 & \\ +138^6 + 165^6 + 246^6 &= 251^6 \end{aligned} \tag{23}$$
$$\begin{aligned} 48^6 + 111^6 + 156^6 + 186^6 + 188^6 & \\ +228^6 + 240^6 + 426^6 &= 431^6 \end{aligned} \tag{24}$$
$$\begin{aligned} 93^6 + 93^6 + 195^6 + 197^6 + 303^6 & \\ +303^6 + 303^6 + 411^6 &= 440^6 \end{aligned} \tag{25}$$
$$\begin{aligned} 219^6 + 255^6 + 261^6 + 267^6 + 289^6 & \\ +351^6 + 351^6 + 351^6 &= 440^6 \end{aligned} \tag{26}$$
$$\begin{aligned} 12^6 + 66^6 + 138^6 + 174^6 + 212^6 & \\ +288^6 + 306^6 + 441^6 &= 455^6 \end{aligned} \tag{27}$$
$$\begin{aligned} 12^6 + 48^6 + 222^6 + 236^6 + 333^6 & \\ +384^6 + 390^6 + 426^6 &= 493^6 \end{aligned} \tag{28}$$
$$\begin{aligned} 66^6 + 78^6 + 144^6 + 228^6 + 256^6 & \\ +288^6 + 435^6 + 444^6 &= 499^6 \end{aligned} \tag{29}$$
$$\begin{aligned} 16^6 + 24^6 + 60^6 + 156^6 + 204^6 & \\ +276^6 + 330^6 + 492^6 &= 502^6 \end{aligned} \tag{30}$$
$$\begin{aligned} 61^6 + 96^6 + 156^6 + 228^6 + 276^6 & \\ +318^6 + 354^6 + 534^6 &= 547^6 \end{aligned} \tag{31}$$
$$\begin{aligned} 170^6 + 177^6 + 276^6 + 312^6 + 312^6 & \\ +408^6 + 450^6 + 498^6 &= 559^6 \end{aligned} \tag{32}$$
$$\begin{aligned} 60^6 + 102^6 + 126^6 + 261^6 + 270^6 & \\ +338^6 + 354^6 + 570^6 &= 581^6 \end{aligned} \tag{33}$$
$$\begin{aligned} 57^6 + 146^6 + 150^6 + 360^6 + 390^6 & \\ +402^6 + 444^6 + 528^6 &= 583^6 \end{aligned} \tag{34}$$
$$\begin{aligned} 33^6 + 72^6 + 122^6 + 192^6 + 204^6 & \\ +390^6 + 534^6 + 534^6 &= 607^6 \end{aligned} \tag{35}$$
$$\begin{aligned} 12^6 + 90^6 + 114^6 + 114^6 + 273^6 & \\ +306^6 + 492^6 + 592^6 &= 623^6 \end{aligned} \tag{36}$$

(Lander *et al.* 1967). The smallest 8-2 solution is

$$8^6 + 10^6 + 12^6 + 15^6 + 24^6 + 30^6 + 33^6 + 36^6 = 35^6 + 37^6 \tag{37}$$

(Lander *et al.* 1967).

The smallest 9-1 solution is

$$1^6 + 17^6 + 19^6 + 22^6 + 31^6 + 37^6 + 37^6 + 41^6 + 49^6 = 54^6 \tag{38}$$

(Lander *et al.* 1967). The smallest 9-2 solution is

$$1^6+5^6+5^6+7^6+13^6+13^6+13^6+17^6+19^6=6^6+21^6 \quad (39)$$

(Lander *et al.* 1967).

The smallest 10-1 solution is

$$2^6+4^6+7^6+14^6+16^6+26^6+26^6+30^6+32^6+32^6=39^6 \quad (40)$$

(Lander *et al.* 1967). The smallest 10-2 solution is

$$1^6+1^6+1^6+4^6+4^6+7^6+9^6+11^6+11^6+11^6=12^6+12^6 \quad (41)$$

(Lander *et al.* 1967).

The smallest 11-1 solution is

$$2^6+5^6+5^6+5^6+7^6+7^6+9^6+9^6+10^6+14^6+17^6=18^6 \quad (42)$$

(Lander *et al.* 1967).

There is also at least one 16-1 identity,

$$\begin{aligned}1^6+2^6+4^6+5^6+6^6+7^6+9^6+12^6+13^6+15^6\\+16^6+18^6+20^6+21^6+22^6+23^6=28^6\end{aligned} \quad (43)$$

(Martin 1893). Moessner (1959) gave solutions for 16-1, 18-1, 20-1, and 23-1.

References

Ekl, R. L. "Equal Sums of Four Seventh Powers." *Math. Comput.* **65**, 1755–1756, 1996.

Guy, R. K. "Sums of Like Powers. Euler's Conjecture." §D1 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 139–144, 1994.

Lander, L. J.; Parkin, T. R.; and Selfridge, J. L. "A Survey of Equal Sums of Like Powers." *Math. Comput.* **21**, 446–459, 1967.

Martin, A. "On Powers of Numbers Whose Sum is the Same Power of Some Number." *Quart. J. Math.* **26**, 225–227, 1893.

Moessner, A. "On Equal Sums of Like Powers." *Math. Student* **15**, 83–88, 1947.

Moessner, A. "Einige zahlentheoretische Untersuchungen und diophantische Probleme." *Glasnik Mat.-Fiz. Astron. Drustvo Mat. Fiz. Hrvatske Ser. 2* **14**, 177–182, 1959.

Rao, S. K. "On Sums of Sixth Powers." *J. London Math. Soc.* **9**, 172–173, 1934.

Diophantine Equation—7th Powers

The 2-1 equation

$$A^7+B^7=C^7 \quad (1)$$

is a special case of FERMAT'S LAST THEOREM with $n=7$, and so has no solution. No solutions to the 2-2 equation

$$A^7+B^7=C^7+D^7 \quad (2)$$

are known

No solutions to the 3-1 or 3-2 equations are known, neither are solutions to the 3-3 equation

$$A^7+B^7+C^7=D^7+E^7+F^7 \quad (3)$$

(Ekl 1996).

No 4-1, 4-2, or 4-3 solutions are known. Guy (1994, p. 140) asked if a 4-4 equation exists for 7th POWERS. An affirmative answer was provided by (Ekl 1996),

$$149^7+123^7+14^7+10^7=146^7+129^7+90^7+15^7 \quad (4)$$

$$194^7+150^7+105^7+23^7=192^7+152^7+132^7+38^7. \quad (5)$$

A 4-5 solution is known.

No 5-1, 5-2, or 5-3 solutions are known. Numerical solutions to the 5-4 equation are given by Gloden (1948). The smallest 5-4 solution is

$$3^7+11^7+26^7+29^7+52^7=12^7+16^7+43^7+50^7 \quad (6)$$

(Lander *et al.* 1967). Gloden (1949) gives parametric solutions to the 5-5 equation. The first few 5-5 solutions are

$$\begin{aligned}&8^7+8^7+13^7+16^7+19^7\\&\qquad=2^7+12^7+15^7+17^7+18^7 \quad (7)\\&4^7+8^7+14^7+16^7+23^7\\&\qquad=7^7+7^7+9^7+20^7+22^7 \quad (8)\\&11^7+12^7+18^7+21^7+26^7\\&\qquad=9^7+10^7+22^7+23^7+24^7 \quad (9)\\&6^7+12^7+20^7+22^7+27^7\\&\qquad=10^7+13^7+13^7+25^7+26^7 \quad (10)\\&3^7+13^7+17^7+24^7+38^7\\&\qquad=14^7+26^7+32^7+32^7+33^7 \quad (11)\end{aligned}$$

(Lander *et al.* 1967).

No 6-1, 6-2, or 6-3 solutions are known. A parametric solution to the 6-6 equation was given by Sastry and Rai (1948). The smallest is

$$2^7+3^7+6^7+6^7+10^7+13^7=1^7+1^7+7^7+7^7+12^7+12^7 \quad (12)$$

(Lander *et al.* 1967).

There are no known solutions to the 7-1 equation (Guy 1994, p. 140). A 7^2-2 solution is

$$\begin{aligned}&2^7+26^7\\&=4^7+8^7+13^7+14^7+14^7+16^7+18^7+22^7+23^7+23^7\\&=7^7+7^7+9^7+13^7+14^7+18^7+20^7+22^7+22^7+23^7\end{aligned} \quad (13)$$

(Lander *et al.* 1967). The smallest 7-3 solution is

$$7^7+7^7+12^7+16^7+27^7+28^7+31^7 = 26^7+30^7+30^7 \quad (14)$$

(Lander *et al.* 1967).

The smallest 8-1 solution is

$$12^7+35^7+53^7+58^7+64^7+83^7+85^7+90^7 = 102^7 \quad (15)$$

(Lander *et al.* 1967). The smallest 8-2 solution is

$$5^7+6^7+7^7+15^7+15^7+20^7+28^7+31^7 = 10^7+33^7 \quad (16)$$

(Lander *et al.* 1967).

The smallest 9-1 solution is

$$6^7+14^7+20^7+22^7+27^7+33^7+41^7+50^7+59^7 = 62^7 \quad (17)$$

(Lander *et al.* 1967).

References

Ekl, R. L. "Equal Sums of Four Seventh Powers." *Math. Comput.* **65**, 1755–1756, 1996.

Gloden, A. "Zwei Parameterlösungen einer mehrgeradigen Gleichung." *Arch. Math.* **1**, 480–482, 1949.

Guy, R. K. "Sums of Like Powers. Euler's Conjecture." §D1 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 139–144, 1994.

Lander, L. J.; Parkin, T. R.; and Selfridge, J. L. "A Survey of Equal Sums of Like Powers." *Math. Comput.* **21**, 446–459, 1967.

Sastry, S. and Rai, T. "On Equal Sums of Like Powers." *Math. Student* **16**, 18–19, 1948.

Diophantine Equation—8th Powers

The 2-1 equation

$$A^8+B^8=C^8 \quad (1)$$

is a special case of FERMAT'S LAST THEOREM with $n = 8$, and so has no solution. No 2-2 solutions are known.

No 3-1, 3-2, or 3-3 solutions are known.

No 4-1, 4-2, 4-3, or 4-4 solutions are known.

No 5-1, 5-2, 5-3, or 5-4 solutions are known, but Letac (1942) found a solution to the 5-5 equation. The smallest 5-5 solution is

$$1^8+10^8+11^8+20^8+43^8 = 5^8+28^8+32^8+35^8+41^8 \quad (2)$$

(Lander *et al.* 1967).

No 6-1, 6-2, 6-3, or 6-4 solutions are known. Moessner and Gloden (1944) found solutions to the 6-6 equation. The smallest 6-6 solution is

$$3^8+6^8+8^8+10^8+15^8+23^8 = 5^8+9^8+9^8+12^8+20^8+22^8 \quad (3)$$

(Lander *et al.* 1967).

No 7-1, 7-2, or 7-3 solutions are known. The smallest 7-4 solution is

$$7^8+9^8+16^8+22^8+22^8+28^8+34^8 = 6^8+11^8+20^8+35^8 \quad (4)$$

(Lander *et al.* 1967). Moessner and Gloden (1944) found solutions to the 7-6 equation. Parametric solutions to the 7-7 equation were given by Moessner (1947) and Gloden (1948). The smallest 7-7 solution is

$$\begin{aligned}1^8+3^8+5^8+6^8+6^8+8^8+13^8 \\ = 4^8+7^8+9^8+9^8+10^8+11^8+12^8 \quad (5)\end{aligned}$$

(Lander *et al.* 1967).

No 8-1 or 8-2 solutions are known. The smallest 8-3 solution is

$$6^8+12^8+16^8+16^8+38^8+38^8+40^8+47^8 = 8^8+17^8+50^8 \quad (6)$$

(Lander *et al.* 1967). Sastry (1934) used the smallest 17-1 solution to give a parametric 8-8 solution. The smallest 8-8 solution is

$$\begin{aligned}1^8+3^8+7^8+7^8+7^8+10^8+10^8+12^8 \\ = 4^8+5^8+5^8+6^8+6^8+11^8+11^8+11^8 \quad (7)\end{aligned}$$

(Lander *et al.* 1967).

No solutions to the 9-1 equation is known. The smallest 9-2 solution is

$$2^8+7^8+8^8+16^8+17^8+20^8+20^8+24^8+24^8 = 11^8+27^8 \quad (8)$$

(Lander *et al.* 1967). Letac (1942) found solutions to the 9-9 equation.

No solutions to the 10-1 equation are known.

The smallest 11-1 solution is

$$\begin{aligned}14^8+18^8+2\cdot 44^8+66^8+70^8+92^8 \\ +93^8+96^8+106^8+112^8 = 125^8 \quad (9)\end{aligned}$$

(Lander *et al.* 1967).

The smallest 12-1 solution is

$$\begin{aligned}2\cdot 8^8+10^8+3\cdot 24^8+26^8+30^8 \\ +34^8+44^8+52^8+63^8 = 65^8 \quad (10)\end{aligned}$$

(Lander *et al.* 1967).

The general identity

$$\begin{aligned}(2^{8k+4}+1)^8 = (2^{8k+4}-1)^8+(2^{7k+4})^8 \\ +(2^{k+1})^8+7[(2^{5k+3})^8+(2^{3k+2})^8] \quad (11)\end{aligned}$$

gives a solution to the 17-1 equation (Lander *et al.* 1967).

References

Gloden, A. "Parametric Solutions of Two Multi-Degreed Equalities." *Amer. Math. Monthly* **55**, 86–88, 1948.

Lander, L. J.; Parkin, T. R.; and Selfridge, J. L. "A Survey of Equal Sums of Like Powers." *Math. Comput.* **21**, 446–459, 1967.

Letac, A. *Gazetta Mathematica* **48**, 68–69, 1942.

Moessner, A. "On Equal Sums of Like Powers." *Math. Student* **15**, 83–88, 1947.

Moessner, A. and Gloden, A. "Einige Zahlentheoretische Untersuchungen und Resultante." *Bull. Sci. École Polytech. de Timisoara* **11**, 196–219, 1944.

Sastry, S. "On Sums of Powers." *J. London Math. Soc.* **9**, 242–246, 1934.

Diophantine Equation—9th Powers

The 2-1 equation

$$A^9 + B^9 = C^9 \tag{1}$$

is a special case of FERMAT'S LAST THEOREM with $n = 9$, and so has no solution. There is no known 2-2 solution.

There are no known 3-1, 3-2, or 3-3 solutions.

There are no known 4-1, 4-2, 4-3, or 4-4 solutions.

There are no known 5-1, 5-2, 5-3, 5-4, or 5-5 solutions.

There are no known 6-1, 6-2, 6-3, 6-4, or 6-5 solutions. The smallest 6-6 solution is

$$1^9 + 13^9 + 13^9 + 14^9 + 18^9 + 23^9 = 5^9 + 9^9 + 10^9 + 15^9 + 21^9 + 22^9 \tag{2}$$

(Lander *et al.* 1967).

There are no known 7-1, 7-2, 7-3, 7-4, or 7-5 solutions.

There are no known 8-1, 8-2, 8-3, 8-4, or 8-5 solutions.

There are no known 9-1, 9-2, 9-3, 9-4, or 9-5 solutions.

There are no known 10-1, 10-2, or 10-3 solutions. The smallest 10-4 solution is

$$2^9 + 6^9 + 6^9 + 9^9 + 10^9 + 11^9 + 14^9 + 18^9 + 2 \cdot 19^9 = 5^9 + 12^9 + 16^9 + 21^9 \tag{3}$$

(Lander *et al.* 1967). No 10-5 solution is known. Moessner (1947) gives a parametric solution to the 10-10 equation.

There are no known 11-1 or 11-2 solutions. The smallest 11-3 solution is

$$2^9 + 3^9 + 6^9 + 7^9 + 9^9 + 9^9 + 19^9 + 19^9 + 21^9 + 25^9 + 29^9 = 13^9 + 16^9 + 30^9 \tag{4}$$

(Lander *et al.* 1967). The smallest 11-5 solution is

$$3^9 + 5^9 + 5^9 + 9^9 + 9^9 + 12^9 + 15^9 + 15^9 + 16^9 + 21^9 + 21^9 = 7^9 + 8^9 + 14^9 + 20^9 + 22^9 \tag{5}$$

(Lander *et al.* 1967). Palamá (1953) gave a solution to the 11-11 equation.

There is no known 12-1 solution. The smallest 12-2 solution is

$$4 \cdot 2^9 + 2 \cdot 3^9 + 4^9 + 7^9 + 16^9 + 17^9 + 2 \cdot 19^9 = 15^9 + 21^9 \tag{6}$$

(Lander *et al.* 1967).

There are no known 13-1 or 14-1 solutions. The smallest 15-1 solution is

$$2^9 + 2^9 + 4^9 + 6^9 + 6^9 + 7^9 + 9^9 + 9^9 + 10^9 + 15^9 + 18^9 + 21^9 + 21^9 + 23^9 + 23^9 = 26^9 \tag{7}$$

(Lander *et al.* 1967).

References

Lander, L. J.; Parkin, T. R.; and Selfridge, J. L. "A Survey of Equal Sums of Like Powers." *Math. Comput.* **21**, 446–459, 1967.

Moessner, A. "On Equal Sums of Like Powers." *Math. Student* **15**, 83–88, 1947.

Palamá, G. "Diophantine Systems of the Type $\sum_{i=1}^{p} a_i{}^k = \sum_{i=1}^{p} b_i{}^k$ ($k = 1, 2, \ldots, n, n+2, n+4, \ldots, n+2r$)." *Scripta Math.* **19**, 132–134, 1953.

Diophantine Equation—10th Powers

The 2-1 equation

$$A^{10} + B^{10} = C^{10} \tag{1}$$

is a special case of FERMAT'S LAST THEOREM with $n = 10$, and so has no solution. The smallest values for which n-1, n-2, etc., have solutions are 23, 19, 24, 23, 16, 27, and 7, corresponding to

$$5 \cdot 1^{10} + 2^{10} + 3^{10} + 6^{10} + 6 \cdot 7^{10} + 4 \cdot 9^{10} + 10^{10} + 2 \cdot 12^{10} + 13^{10} + 14^{10} = 15^{10} \tag{2}$$

$$5 \cdot 2^{10} + 5^{10} + 6^{10} + 10^{10} + 6 \cdot 11^{10} + 2 \cdot 12^{10} + 3 \cdot 15^{10} = 9^{10} + 17^{10} \tag{3}$$

$$1^{10} + 2^{10} + 3^{10} + 10 \cdot 4^{10} + 7^{10} + 7 \cdot 8^{10} + 10^{10} + 12^{10} + 16^{10} = 11^{10} + 2 \cdot 15^{10} \tag{4}$$

$$5 \cdot 1^{10} + 2 \cdot 2^{10} + 3 \cdot 3^{10} + 4^{10} + 4 \cdot 6^{10} + 3 \cdot 7^{10} + 8^{10} + 2 \cdot 10^{10} + 2 \cdot 14^{10} + 15^{10} = 3 \cdot 11^{10} + 16^{10} \tag{5}$$

$$4 \cdot 1^{10} + 2^{10} + 2 \cdot 4^{10} + 6^{10} + 2 \cdot 12^{10} + 5 \cdot 13^{10} + 15^{10} = 2 \cdot 3^{10} + 8^{10} + 14^{10} + 16^{10} \tag{6}$$

$$\begin{aligned}&1^{10}+4\cdot 3^{10}+2\cdot 4^{10}+2\cdot 5^{10}+7\cdot 6^{10}\\&+9\cdot 7^{10}+10^{10}+13^{10}=2\cdot 2^{10}+8^{10}+11^{10}+2\cdot 12^{10}\end{aligned}\tag{7}$$

$$\begin{aligned}&1^{10}+28^{10}+31^{10}+32^{10}+55^{10}+61^{10}+68^{10}\\&\quad=17^{10}+20^{10}+23^{10}+44^{10}+49^{10}+64^{10}+67^{10}\end{aligned}\tag{8}$$

(Lander *et al.* 1967).

References

Lander, L. J.; Parkin, T. R.; and Selfridge, J. L. "A Survey of Equal Sums of Like Powers." *Math. Comput.* **21**, 446–459, 1967.

Diophantine Equation—Cubic

The 2-1 equation

$$A^3+B^3=C^3\tag{1}$$

is a case of FERMAT'S LAST THEOREM with $n=3$. In fact, this particular case was known not to have any solutions long before the general validity of FERMAT'S LAST THEOREM was established. The 2-2 equation

$$A^3+B^3=C^3+D^3\tag{2}$$

has a known parametric solution (Dickson 1966, pp. 550–554; Guy 1994, p. 140), and 10 solutions with sum $<10^5$,

$$1729=1^3+12^3=9^3+10^3\tag{3}$$
$$4104=2^3+16^3=9^3+15^3\tag{4}$$
$$13832=2^3+24^3=18^3+20^3\tag{5}$$
$$20683=10^3+27^3=19^3+24^3\tag{6}$$
$$32832=4^3+32^3=18^3+30^3\tag{7}$$
$$39312=2^3+34^3=15^3+33^3\tag{8}$$
$$40033=9^3+34^3=16^3+33^3\tag{9}$$
$$46683=3^3+36^3=16^3+33^3\tag{10}$$
$$64232=17^3+39^3=26^3+36^3\tag{11}$$
$$65728=12^3+40^3=31^3+33^3\tag{12}$$

(Sloane's A001235; Moreau 1898). The first number (Madachy 1979, pp. 124 and 141) in this sequence, the so-called HARDY-RAMANUJAN NUMBER, is associated with a story told about Ramanujan by G. H. Hardy, but was known as early as 1657 (Berndt and Bhargava 1993). The smallest number representable in n ways as a sum of cubes is called the nth TAXICAB NUMBER.

Ramanujan gave a general solution to the 2-2 equation as

$$(\alpha+\lambda^2\gamma)^3+(\lambda\beta+\gamma)^3=(\lambda\alpha+\gamma)^3+(\beta+\lambda^2\gamma)^3\tag{13}$$

where

$$\alpha^2+\alpha\beta+\beta^2=3\lambda\gamma^2\tag{14}$$

(Berndt 1994, p. 107). Another form due to Ramanujan is

$$\begin{aligned}&(A^2+7AB-9B^2)^3+(2A^2-4AB+12B^2)^3\\&\qquad=(2A^2+10B^2)^3+(A^2-9AB-B^2)^3.\end{aligned}\tag{15}$$

Hardy and Wright (1979, Theorem 412) prove that there are numbers that are expressible as the sum of two cubes in n ways for any n (Guy 1994, pp. 140–141). The proof is constructive, providing a method for computing such numbers: given RATIONALS NUMBERS r and s, compute

$$t=\frac{r(r^3+2s^3)}{r^3-s^3}\tag{16}$$
$$u=\frac{s(2r^3+s^3)}{r^3-s^3}\tag{17}$$
$$v=\frac{t(t^3-2u^3)}{t^3+u^3}\tag{18}$$
$$w=\frac{u(2t^3-u^3)}{t^3+u^3}.\tag{19}$$

Then

$$r^3+s^3=t^3-u^3=v^3+w^3\tag{20}$$

The DENOMINATORS can now be cleared to produce an integer solution. If r/s is picked to be large enough, the v and w will be POSITIVE. If r/s is still larger, the v/w will be large enough for v and w to be used as the inputs to produce a third pair, etc. However, the resulting integers may be quite large, even for $n=2$. E.g., starting with $3^3+1^3=28$, the algorithm finds

$$28=\left(\tfrac{28340511}{21446828}\right)^3+\left(\tfrac{63284705}{21446828}\right)^3,\tag{21}$$

giving

$$28\cdot 21446828^3=(3\cdot 21446828)^3+21446828^3\tag{22}$$
$$=28340511^3+63284705^3.\tag{23}$$

The numbers representable in three ways as a sum of two cubes (a 2-2-2 equation) are

$$87539319=167^3+436^3=228^3+423^3=255^3+414^3\tag{24}$$
$$119824488=11^3+493^3=90^3+492^3=346^3+428^3\tag{25}$$
$$143604279=111^3+522^3=359^3+460^3=408^3+423^3\tag{26}$$
$$175959000=70^3+560^3=198^3+552^3=315^3+525^3\tag{27}$$
$$327763000=300^3+670^3=339^3+661^3=510^3+580^3\tag{28}$$

(Guy 1994, Sloane's A003825). Wilson (1997) found 32 numbers representable in four ways as the sum of two cubes (a 2-2-2-2 equation). The first is

$$\begin{aligned} 6963472309248 &= 2421^2 + 19083^2 = 5436^2 + 18948^2 \\ &= 102020^3 + 18072^2 = 13322^3 + 15530^3. \end{aligned} \tag{29}$$

The smallest known numbers so representable are 6963472309248, 12625136269928, 21131226514944, 26059452841000, ... (Sloane's A003826). Wilson also found six five-way sums,

$$\begin{aligned} 48988659276962496 &= 38787^3 + 365757^3 \\ &= 107839^3 + 362753^3 \\ &= 205292^3 + 342952^3 \\ &= 221424^3 + 336588^3 \\ &= 231518^3 + 331954^3 \end{aligned} \tag{30}$$

$$\begin{aligned} 490593422681271000 &= 48369^3 + 788631^3 \\ &= 233775^3 + 781785^3 \\ &= 285120^3 + 776070^3 \\ &= 543145^3 + 691295^3 \\ &= 579240^3 + 666630^3 \end{aligned} \tag{31}$$

$$\begin{aligned} 6355491080314102272 &= 103113^3 + 1852215^3 \\ &= 580488^3 + 1833120^3 \\ &= 788724^3 + 1803372^3 \\ &= 1150792^3 + 1690544^3 \\ &= 1462050^3 + 1478238^3 \end{aligned} \tag{32}$$

$$\begin{aligned} 27365551142421413376 &= 167751^3 + 3013305^3 \\ &= 265392^3 + 3012792^3 \\ &= 944376^3 + 2982240^3 \\ &= 1283148^3 + 2933844^3 \\ &= 1872184^3 + 2750288^3 \end{aligned} \tag{33}$$

$$\begin{aligned} 1199962860219870469632 &= 591543^3 + 10625865^3 \\ &= 935856^3 + 10624056^3 \\ &= 3330168^3 + 10516320^3 \\ &= 6601912^3 + 9698384^3 \\ &= 8387550^3 + 8480418^3 \end{aligned} \tag{34}$$

$$\begin{aligned} 111549833098123426841016 &= 1074073^3 + 48137999^3 \\ &= 8787870^3 + 48040356^3 \\ &= 13950972^3 + 47744382^3 \\ &= 24450192^3 + 45936462^3 \\ &= 33784478^3 + 41791204^3, \end{aligned} \tag{35}$$

and a single six-way sum

$$\begin{aligned} &8230545258248091551205888 \\ &= 11239317^3 + 201891435^3 \\ &= 17781264^3 + 201857064^3 \\ &= 63273192^3 + 199810080^3 \\ &= 85970916^3 + 196567548^3 \\ &= 125436328^3 + 184269296^3 \\ &= 159363450^3 + 161127942^3. \end{aligned} \tag{36}$$

The first rational solution to the 3-1 equation

$$A^3 + B^3 + C^3 = D^3 \tag{37}$$

was found by Euler and Vieta (Dickson 1966, pp. 550–554). Hardy and Wright (1979, pp. 199–201) give a solution which can be based on the identities

$$\begin{aligned} a^3(a^3 + b^3)^3 &= b^3(a^3 + b^3)^3 + a^3(a^3 - 2b^3)^3 \\ &\quad + b^3(2a^3 - b^3)^3 \end{aligned} \tag{38}$$

$$\begin{aligned} a^3(a^3 + 2b^3)^3 &= a^3(a^3 - b^3)^3 + b^3(a^3 - b^3)^3 \\ &\quad + b^3(2a^3 + b^3)^3. \end{aligned} \tag{39}$$

This is equivalent to the general 2-2 solution found by Ramanujan (Berndt 1994, pp. 54 and 107). The smallest integral solutions are

$$3^3 + 4^3 + 5^3 = 6^3 \tag{40}$$

$$1^3 + 6^3 + 8^3 = 9^3 \tag{41}$$

$$7^3 + 14^3 + 17^3 = 20^3 \tag{42}$$

$$11^3 + 15^3 + 27^3 = 29^3 \tag{43}$$

$$28^3 + 53^3 + 75^3 = 84^3 \tag{44}$$

$$26^3 + 55^3 + 78^3 = 87^3 \tag{45}$$

$$33^3 + 70^3 + 92^3 = 105^3 \tag{46}$$

(Beeler *et al.* 1972; Madachy 1979, pp. 124 and 141). Other general solutions have been found by Binet (1841) and Schwering (1902), although Ramanujan's formulation is the simplest. No general solution giving *all* POSITIVE integral solutions is known (Dickson 1966, pp. 550–561).

4-1 equations include

$$11^3 + 12^3 + 13^3 + 14^3 = 20^3 \tag{47}$$

$$5^3 + 7^3 + 9^3 + 10^3 = 13^3. \tag{48}$$

A solution to the 4-4 equation is

$$2^3 + 3^3 + 10^3 + 11^3 = 1^3 + 5^3 + 8^3 + 12^3 \tag{49}$$

(Madachy 1979, pp. 118 and 133).

5-1 equations

$$1^3 + 3^3 + 4^3 + 5^3 + 8^3 = 9^3 \tag{50}$$

$$3^3 + 4^3 + 5^3 + 8^3 + 10^3 = 12^3, \tag{51}$$

and a 6-1 equation is given by

$$1^3 + 5^3 + 6^3 + 7^3 + 8^3 + 10^3 = 13^3. \tag{52}$$

A 6-6 equation also exists:

$$1^3 + 2^3 + 4^3 + 8^3 + 9^3 + 12^3 = 3^3 + 5^3 + 6^3 + 7^3 + 10^3 + 11^3 \tag{53}$$

(Madachy 1979, p. 142).

Euler gave the general solution to

$$A^3+B^3=C^2 \tag{54}$$

as

$$A=3n^3+6n^2-n \tag{55}$$
$$B=-3n^3+6n^2+n \tag{56}$$
$$C=6n^2(3n^2+1). \tag{57}$$

see also CANNONBALL PROBLEM, HARDY-RAMANUJAN NUMBER, SUPER-3 NUMBER, TAXICAB NUMBER, TRIMORPHIC NUMBER

References
Beeler, M.; Gosper, R. W.; and Schroeppel, R. Item 58 in *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.
Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, 1994.
Berndt, B. C. and Bhargava, S. "Ramanujan—For Lowbrows." *Amer. Math. Monthly* **100**, 645–656, 1993.
Binet, J. P. M. "Note sur une question relative à la théorie des nombres." *C. R. Acad. Sci. (Paris)* **12**, 248–250, 1841.
Dickson, L. E. *History of the Theory of Numbers, Vol. 2: Diophantine Analysis.* New York: Chelsea, 1966.
Guy, R. K. "Sums of Like Powers. Euler's Conjecture." §D1 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 139–144, 1994.
Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, p. 68, 1959.
Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, 1979.
Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, 1979.
Moreau, C. "Plus petit nombre égal à la somme de deux cubes de deux façons." *L'Intermediaire Math.* **5**, 66, 1898.
Schwering, K. "Vereinfachte Lösungen des Eulerschen Aufgabe: $x^3+y^3+z^3+v^3=0$." *Arch. Math. Phys.* **2**, 280–284, 1902.
Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 157, 1993.
Sloane, N. J. A. Sequences A001235 and A003825 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Wilson, D. Personal communication, Apr. 17, 1997.

Diophantine Equation—Linear

A linear Diophantine equation (in two variables) is an equation of the general form

$$ax+by=c, \tag{1}$$

where solutions are sought with a, b, and c INTEGERS. Such equations can be solved completely, and the first known solution was constructed by Brahmagupta. Consider the equation

$$ax+by=1. \tag{2}$$

Now use a variation of the EUCLIDEAN ALGORITHM, letting $a=r_1$ and $b=r_2$

$$r_1=q_1r_2+r_3 \tag{3}$$
$$r_2=q_2r_3+r_4 \tag{4}$$
$$r_{n-3}=q_{n-3}r_{n-2}+r_{n-1} \tag{5}$$
$$r_{n-2}=q_{n-2}r_{n-1}+1. \tag{6}$$

Starting from the bottom gives

$$1=r_{n-2}-q_{n-2}r_{n-1} \tag{7}$$
$$r_{n-1}=r_{n-3}-q_{n-3}r_{n-2}, \tag{8}$$

so

$$\begin{aligned}1&=r_{n-2}-q_{n-2}(r_{n-3}-q_{n-3}r_{n-2})\\&=-q_{n-2}r_{n-3}+(1-q_{n-2}q_{n-3})r_{n-2}.\end{aligned} \tag{9}$$

Continue this procedure all the way back to the top.

Take as an example the equation

$$1027x+712y=1. \tag{10}$$

Proceed as follows.

$$\begin{array}{rl}
1027=712\cdot 1+315\;| & 1=-165\cdot 1027+238\cdot 712\;\uparrow\\
712=315\cdot 2+82\;| & 1=73\cdot 712-165\cdot 315\;|\\
315=82\cdot 3+69\;| & 1=-19\cdot 315+73\cdot 82\;|\\
82=69\cdot 1+13\;| & 1=16\cdot 82-19\cdot 69\;|\\
69=13\cdot 5+4\;| & 1=-3\cdot 69+16\cdot 13\;|\\
13=4\cdot 3+1\;\downarrow & 1=1\cdot 13-3\cdot 4\;|\\
 & 1=0\cdot 4+1\cdot 1,\;|
\end{array}$$

The solution is therefore $x=-165$, $y=238$. The above procedure can be simplified by noting that the two leftmost columns are offset by one entry and alternate signs, as they must since

$$1=-A_{i+1}r_i+A_ir_{i+1} \tag{11}$$
$$r_{i+1}=r_{i-1}-r_iq_{i-1} \tag{12}$$
$$1=A_ir_{i-1}-(A_iq_{i-1}+A_{i+1}), \tag{13}$$

so the COEFFICIENTS of r_{i-1} and r_{i+1} are the same and

$$A_{i-1}=-(A_iq_{i-1}+A_{i+1}). \tag{14}$$

Repeating the above example using this information therefore gives

$$\begin{array}{rl}
1027=712\cdot 1+315\;| & (-)\;\; 165\cdot 1+73=238\;\uparrow\\
712=315\cdot 2+82\;| & (+)\;\; 73\cdot 2+19=165\;|\\
315=82\cdot 3+69\;| & (-)\;\; 19\cdot 3+16=73\;|\\
82=69\cdot 1+13\;| & (+)\;\; 16\cdot 1+3=19\;|\\
69=13\cdot 5+4\;| & (-)\;\; 3\cdot 5+1=16\;|\\
13=4\cdot 3+1\;\downarrow & (+)\;\; 1\cdot 3+0=3\;|\\
 & (-)\;\; 0\cdot 1+1=1\;|
\end{array}$$

and we recover the above solution.

Call the solutions to

$$ax + by = 1 \tag{15}$$

x_0 and y_0. If the signs in front of ax or by are NEGATIVE, then solve the above equation and take the signs of the solutions from the following table:

equation	x	y
$ax + by = 1$	x_0	y_0
$ax - by = 1$	x_0	$-y_0$
$-ax + by = 1$	$-x_0$	y_0
$-ax - by = 1$	$-x_0$	$-y_0$

In fact, the solution to the equation

$$ax - by = 1 \tag{16}$$

is equivalent to finding the CONTINUED FRACTION for a/b, with a and b RELATIVELY PRIME (Olds 1963). If there are n terms in the fraction, take the $(n-1)$th convergent p_{n-1}/q_{n-1}. But

$$p_n q_{n-1} - p_{n-1} q_n = (-1)^n, \tag{17}$$

so one solution is $x_0 = (-1)^n q_{n-1}$, $y_0 = (-1)^n p_{n-1}$, with a general solution

$$x = x_0 + kb \tag{18}$$

$$y = y_0 + ka \tag{19}$$

with k an arbitrary INTEGER. The solution in terms of smallest POSITIVE INTEGERS is given by choosing an appropriate k.

Now consider the general first-order equation of the form

$$ax + by = c. \tag{20}$$

The GREATEST COMMON DIVISOR $d \equiv \mathrm{GCD}(a,b)$ can be divided through yielding

$$a'x + b'y = c', \tag{21}$$

where $a' \equiv a/d$, $b' \equiv b/d$, and $c' \equiv c/d$. If $d \nmid c$, then c' is not an INTEGER and the equation cannot have a solution in INTEGERS. A necessary and sufficient condition for the general first-order equation to have solutions in INTEGERS is therefore that $d|c$. If this is the case, then solve

$$a'x + b'y = 1 \tag{22}$$

and multiply the solutions by c', since

$$a'(c'x) + b'(c'y) = c'. \tag{23}$$

References

Courant, R. and Robbins, H. "Continued Fractions. Diophantine Equations." §2.4 in Supplement to Ch. 1 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 49–51, 1996.

Dickson, L. E. "Linear Diophantine Equations and Congruences." Ch. 2 in *History of the Theory of Numbers, Vol. 2: Diophantine Analysis.* New York: Chelsea, pp. 41–99, 1952.

Olds, C. D. Ch. 2 in *Continued Fractions.* New York: Random House, 1963.

Diophantine Equation—nth Powers

The 2-1 equation

$$A^n + B^n = C^n \tag{1}$$

is a special case of FERMAT'S LAST THEOREM and so has no solutions for $n \geq 3$. Lander *et al.* (1967) give a table showing the smallest n for which a solution to

$$x_1{}^k + x_2{}^k + \ldots + x_m{}^k = y_1{}^k + y_2{}^k + \ldots + y_n{}^k,$$

with $1 \leq m \leq n$ is known.

				k					
m	2	3	4	5	6	7	8	9	10
1	2	3	3	4	7	8	11	15	23
2	2	2	2	4	7	8	9	12	19
3				3	3	7	8	11	24
4						4	7	10	23
5						5	5	11	16
6								6	27
7									7

Take the results from the RAMANUJAN 6-10-8 IDENTITY that for $ad = bc$, with

$$\begin{aligned} F_{2m}(a,b,c,d) &= (a+b+c)^{2m} + (b+c+d)^{2m} \\ &-(c+d+a)^{2m} - (d+a+b)^{2m} + (a-d)^{2m} - (b-c)^{2m} \end{aligned} \tag{2}$$

and

$$\begin{aligned} f_{2m}(x,y) &= (1+x+y)^{2m} + (x+y+xy)^{2m} \\ &-(y+xy+1)^{2m} - (xy+1+x)^{2m} + (1-xy)^{2m} - (x-y)^{2m}, \end{aligned} \tag{3}$$

then

$$F_{2m}(a,b,c,d) = a^{2m} f_{2m}(x,y). \tag{4}$$

Using

$$f_2(x,y) = 0 \tag{5}$$

$$f_4(x,y) = 0 \tag{6}$$

now gives

$$\begin{aligned} &(a+b+c)^n + (b+c+d)^n + (a-d)^n \\ &\quad = (c+d+a)^n + (d+a+b)^n + (b-c)^n \end{aligned} \tag{7}$$

for $n = 2$ or 4.

see also RAMANUJAN 6-10-8 IDENTITY

References

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, p. 101, 1994.

Berndt, B. C. and Bhargava, S. "Ramanujan—For Lowbrows." *Amer. Math. Monthly* **100**, 644–656, 1993.

Dickson, L. E. *History of the Theory of Numbers, Vol. 2: Diophantine Analysis.* New York: Chelsea, pp. 653–657, 1966.

Gloden, A. *Mehrgradige Gleichungen.* Groningen, Netherlands: P. Noordhoff, 1944.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, 1994.

Lander, L. J.; Parkin, T. R.; and Selfridge, J. L. "A Survey of Equal Sums of Like Powers." *Math. Comput.* **21**, 446–459, 1967.

Reznick, B. "Sums of Even Powers of Real Linear Forms." *Mem. Amer. Math. Soc.* No. 463, **96**. Providence, RI: Amer. Math. Soc., 1992.

Diophantine Equation—Quadratic

An equation of the form

$$x^2 - Dy^2 = 1, \tag{1}$$

where D is an INTEGER is called a PELL EQUATION. Pell equations, as well as the analogous equation with a minus sign on the right, can be solved by finding the CONTINUED FRACTION for $\sqrt{D}$. (The trivial solution $x = 1$, $y = 0$ is ignored in all subsequent discussion.) Let p_n/q_n denote the nth CONVERGENT $[a_1, a_2, \ldots, a_n]$, then we are looking for a convergent which obeys the identity

$${p_n}^2 - D{q_n}^2 = (-1)^n, \tag{2}$$

which turns out to always be possible since the CONTINUED FRACTION of a QUADRATIC SURD always becomes periodic at some term a_{r+1}, where $a_{r+1} = 2a_1$, i.e.,

$$\sqrt{D} = [a_1, \overline{a_2, \ldots, a_r, 2a_1}\,]. \tag{3}$$

Writing $n = rk$ gives

$${p_{rk}}^2 - D{q_{rk}}^2 = (-1)^{rk}, \tag{4}$$

for k a POSITIVE INTEGER. If r is ODD, solutions to

$$x^2 - Dy^2 = \pm 1 \tag{5}$$

can be obtained if k is chosen to be EVEN or ODD, but if r is EVEN, there are no values of k which can make the exponent ODD.

If r is EVEN, then $(-1)^r$ is POSITIVE and the solution in terms of smallest INTEGERS is $x = p_r$ and $y = q_r$, where p_r/q_r is the rth CONVERGENT. If r is ODD, then $(-1)^r$ is NEGATIVE, but we can take $k = 2$ in this case, to obtain

$${p_{2r}}^2 - D{q_{2r}}^2 = 1, \tag{6}$$

so the solution in smallest INTEGERS is $x = p_{2r}$, $y = q_{2r}$. Summarizing,

$$(x, y) = \begin{cases} (p_r, q_r) & \text{for } r \text{ even} \\ (p_{2r}, p_{2r}) & \text{for } r \text{ odd.} \end{cases} \tag{7}$$

The more complicated equation

$$x^2 - Dy^2 = \pm c \tag{8}$$

can also be solved for certain values of c and D, but the procedure is more complicated (Chrystal 1961). However, if a single solution to the above equation is known, other solutions can be found. Let p and q be solutions to (8), and r and s solutions to the "unit" form". Then

$$(p^2 - Dq^2)(r^2 - Ds^2) = \pm c \tag{9}$$

$$(pr \pm Dqs)^2 - D(ps \pm qr)^2 = \pm c. \tag{10}$$

Call a Diophantine equation consisting of finding m POWERS equal to a sum of n equal POWERS an "$m - n$ equation." The 2-1 equation

$$A^2 = B^2 + C^2, \tag{11}$$

which corresponds to finding a PYTHAGOREAN TRIPLE (A, B, C) has a well-known general solution (Dickson 1966, pp. 165-170). To solve the equation, note that every PRIME of the form $4x + 1$ can be expressed as the sum of two RELATIVELY PRIME squares in exactly one way. To find in how many ways a general number m can be expressed as a sum of two squares, factor it as follows

$$m = 2^{a_0} {p_1}^{2a_1} \cdots {p_n}^{2a_n} {q_1}^{b_1} \cdots {q_r}^{b_r}, \tag{12}$$

where the ps are primes of the form $4x - 1$ and the qs are primes of the form $x + 1$. If the as are integral, then define

$$B \equiv (2b_1 + 1)(2b_2 + 1) \cdots (2b_r + 1) - 1. \tag{13}$$

Then m is a sum of two *unequal* squares in

$$N(m) = \begin{cases} 0 \\ \quad \text{for any } a_i \text{ half-integral} \\ \frac{1}{2}(b_1 + 1)(b_2 + 1) \cdots (b_r + 1) \\ \quad \text{for all } a_i \text{ integral}, B \text{ odd} \\ \frac{1}{2}(b_1 + 1)(b_2 + 1) \cdots (b_r + 1) - \frac{1}{2} \\ \quad \text{for all } a_i \text{ integral}, B \text{ even.} \end{cases} \tag{14}$$

If zero is counted as a square, both POSITIVE and NEGATIVE numbers are included, and the order of the two squares is distinguished, Jacobi showed that the number of ways a number can be written as the sum of two squares is four times the excess of the number of DIVISORS of the form $4x + 1$ over the number of DIVISORS of the form $4x - 1$.

A set of INTEGERS satisfying the 3-1 equation

$$A^2 + B^2 + C^2 = D^2 \tag{15}$$

is called a PYTHAGOREAN QUADRUPLE. Parametric solutions to the 2-2 equation

$$A^2 + B^2 = C^2 + D^2 \tag{16}$$

are known (Dickson 1966; Guy 1994, p. 140).

Solutions to an equation of the form

$$(A^2+B^2)(C^2+D^2)=E^2+F^2 \tag{17}$$

are given by the FIBONACCI IDENTITY

$$(a^2+b^2)(c^2+d^2)=(ac\pm bd)^2+(bc\mp ad)^2\equiv e^2+f^2. \tag{18}$$

Another similar identity is the EULER FOUR-SQUARE IDENTITY

$$\begin{aligned}(a_1{}^2+a_2{}^2)(b_1{}^2+b_2{}^2)(c_1{}^2+c_2{}^2)(d_1{}^2+d_2{}^2)\\ =e_1{}^2+e_2{}^2+e_3{}^2+e_4{}^2\end{aligned} \tag{19}$$

$$\begin{aligned}&(a_1{}^2+a_2{}^2+a_3{}^2+a_4{}^2)(b_1{}^2+b_2{}^2+b_3{}^2+b_4{}^2)\\ &=(a_1b_1-a_2b_2-a_3b_3-a_4b_4)^2\\ &\quad+(a_1b_2+a_2b_1+a_3b_4-a_4b_3)^2\\ &\quad+(a_1b_3-a_2b_4+a_3b_1+a_4b_2)^2\\ &\quad+(a_1b_4+a_2b_3-a_3b_2+a_4b_1)^2.\end{aligned} \tag{20}$$

Degen's eight-square identity holds for eight squares, but no other number, as proved by Cayley. The two-square identity underlies much of TRIGONOMETRY, the four-square identity some of QUATERNIONS, and the eight-square identity, the CAYLEY ALGEBRA (a noncommutative nonassociative algebra; Bell 1945).

RAMANUJAN'S SQUARE EQUATION

$$2^n-7=x^2 \tag{21}$$

has been proved to have only solutions $n=3$, 4, 5, 7, and 15 (Beeler *et al.* 1972, Item 31).

see also ALGEBRA, CANNONBALL PROBLEM, CONTINUED FRACTION, FERMAT DIFFERENCE EQUATION, LAGRANGE NUMBER (DIOPHANTINE EQUATION), PELL EQUATION, PYTHAGOREAN QUADRUPLE, PYTHAGOREAN TRIPLE, QUADRATIC RESIDUE

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Beiler, A. H. "The Pellian." Ch. 22 in *Recreations in the Theory of Numbers: The Queen of Mathematics Entertains.* New York: Dover, pp. 248–268, 1966.

Bell, E. T. *The Development of Mathematics, 2nd ed.* New York: McGraw-Hill, p. 159, 1945.

Chrystal, G. *Textbook of Algebra, 2 vols.* New York: Chelsea, 1961.

Degan, C. F. *Canon Pellianus.* Copenhagen, Denmark, 1817.

Dickson, L. E. "Number of Representations as a Sum of 5, 6, 7, or 8 Squares." Ch. 13 in *Studies in the Theory of Numbers.* Chicago, IL: University of Chicago Press, 1930.

Dickson, L. E. *History of the Theory of Numbers, Vol. 2: Diophantine Analysis.* New York: Chelsea, 1966.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, 1994.

Lam, T. Y. *The Algebraic Theory of Quadratic Forms.* Reading, MA: W. A. Benjamin, 1973.

Rajwade, A. R. *Squares.* Cambridge, England: Cambridge University Press, 1993.

Scharlau, W. *Quadratic and Hermitian Forms.* Berlin: Springer-Verlag, 1985.

Shapiro, D. B. "Products of Sums and Squares." *Expo. Math.* **2**, 235–261, 1984.

Smarandache, F. "Un metodo de resolucion de la ecuacion diofantica." *Gaz. Math.* **1**, 151–157, 1988.

Smarandache, F. " Method to Solve the Diophantine Equation $ax^2-by^2+c=0$." In *Collected Papers, Vol. 1.* Bucharest, Romania: Tempus, 1996.

Taussky, O. "Sums of Squares." *Amer. Math. Monthly* **77**, 805–830, 1970.

Whitford, E. E. *Pell Equation.* New York: Columbia University Press, 1912.

Diophantine Equation—Quartic

Call an equation involving quartics m-n if a sum of m quartics is equal to a sum of n fourth POWERS. The 2-1 equation

$$A^4+B^4=C^4 \tag{1}$$

is a case of FERMAT'S LAST THEOREM with $n=4$ and therefore has no solutions. In fact, the equations

$$A^4\pm B^4=C^2 \tag{2}$$

also have no solutions in INTEGERS.

Parametric solutions to the 2-2 equation

$$A^4+B^4=C^4+D^4 \tag{3}$$

are known (Euler 1802; Gérardin 1917; Guy 1994, pp. 140–141). A few specific solutions are

$$59^4+158^4=133^4+134^4=635{,}318{,}657 \tag{4}$$

$$7^4+239^4=157^4+227^4=3{,}262{,}811{,}042 \tag{5}$$

$$193^4+292^4=256^4+257^4=8{,}657{,}437{,}697 \tag{6}$$

$$298^4+497^4=271^4+502^4=68{,}899{,}596{,}497 \tag{7}$$

$$514^4+359^4=103^4+542^4=86{,}409{,}838{,}577 \tag{8}$$

$$222^4+631^4=503^4+558^4=160{,}961{,}094{,}577 \tag{9}$$

$$21^4+717^4=471^4+681^4=264{,}287{,}694{,}402 \tag{10}$$

$$76^4+1203^4=653^4+1176^4=2{,}094{,}447{,}251{,}857 \tag{11}$$

$$997^4+1342^4=878^4+1381^4=4{,}231{,}525{,}221{,}377 \tag{12}$$

$$27^4+2379^4=577^4+728^4=32{,}031{,}536{,}780{,}322 \tag{13}$$

(Sloane's A001235; Richmond 1920, Leech 1957), the smallest of which is due to Euler. Lander *et al.* (1967) give a list of 25 primitive 2-2 solutions. General (but incomplete) solutions are given by

$$x=a+b \tag{14}$$

$$y=c-d \tag{15}$$

$$u=a-b \tag{16}$$

$$v=c+d, \tag{17}$$

where

$$a = n(m^2+n^2)(-m^4+18m^2n^2-n^4) \tag{18}$$
$$b = 2m(m^6+10m^4n^2+m^2n^4+4n^6) \tag{19}$$
$$c = 2n(4m^6+m^4n^2+10m^2n^4+n^6) \tag{20}$$
$$d = m(m^2+n^2)(-m^4+18m^2n^2-n^4) \tag{21}$$

(Hardy and Wright 1979).

In 1772, Euler proposed that the 3-1 equation

$$A^4+B^4+C^4=D^4 \tag{22}$$

had no solutions in INTEGERS (Lander *et al.* 1967). This assertion is known as the EULER QUARTIC CONJECTURE. Ward (1948) showed there were no solutions for $D \leq 10{,}000$, which was subsequently improved to $D \leq 220{,}000$ by Lander *et al.* (1967). However, the EULER QUARTIC CONJECTURE was disproved in 1987 by Noam D. Elkies, who, using a geometric construction, found

$$2{,}682{,}440^4+15{,}365{,}639^4+18{,}796{,}760^4=20{,}615{,}673^4 \tag{23}$$

and showed that infinitely many solutions existed (Guy 1994, p. 140). In 1988, Roger Frye found

$$95{,}800^4+217{,}519^4+414{,}560^4=422{,}481^4 \tag{24}$$

and proved that there are no solutions in smaller INTEGERS (Guy 1994, p. 140). Another solution was found by Allan MacLeod in 1997,

$$638{,}523{,}249^4 = 630{,}662{,}624^4+275{,}156{,}240^4+219{,}076{,}465^4. \tag{25}$$

It is not known if there is a parametric solution.

In contrast, there are many solutions to the 3-1 equation

$$A^4+B^4+C^4=2D^4 \tag{26}$$

(see below).

Parametric solutions to the 3-2 equation

$$A^4+B^4=C^4+D^4+E^4 \tag{27}$$

are known (Gérardin 1910, Ferrari 1913). The smallest 3-2 solution is

$$3^4+5^4+8^4=7^4+7^4 \tag{28}$$

(Lander *et al.* 1967).

Ramanujan gave the 3-3 equations

$$2^4+4^4+7^4=3^4+6^4+6^4 \tag{29}$$
$$3^4+7^4+8^4=1^4+2^4+9^4 \tag{30}$$
$$6^4+9^4+12^4=2^4+2^4+13^4 \tag{31}$$

(Berndt 1994, p. 101). Similar examples can be found in Martin (1896). Parametric solutions were given by Gérardin (1911).

Ramanujan also gave the general expression

$$3^4+(2x^4-1)^4+(4x^5+x)^4 = (4x^4+1)^4+(6x^4-3)^4+(4x^5-5x)^4 \tag{32}$$

(Berndt 1994, p. 106). Dickson (1966, pp. 653–655) cites several FORMULAS giving solutions to the 3-3 equation, and Haldeman (1904) gives a general FORMULA.

The 4-1 equation

$$A^4+B^4+C^4+D^4=E^4 \tag{33}$$

has solutions

$$30^4+120^4+272^4+315^4=353^4 \tag{34}$$
$$240^4+340^4+430^4+599^4=651^4 \tag{35}$$
$$435^4+710^4+1384^4+2420^4=2487^4 \tag{36}$$
$$1130^4+1190^4+1432^4+2365^4=2501^4 \tag{37}$$
$$850^4+1010^4+1546^4+2745^4=2829^4 \tag{38}$$
$$2270^4+2345^4+2460^4+3152^4=3723^4 \tag{39}$$
$$350^4+1652^4+3230^4+3395^4=3973^4 \tag{40}$$
$$205^4+1060^4+2650^4+4094^4=4267^4 \tag{41}$$
$$1394^4+1750^4+3545^4+3670^4=4333^4 \tag{42}$$
$$699^4+700^4+2840^4+4250^4=4449^4 \tag{43}$$
$$380^4+1660^4+1880^4+4907^4=4949^4 \tag{44}$$
$$1000^4+1120^4+3233^4+5080^4=5281^4 \tag{45}$$
$$410^4+1412^4+3910^4+5055^4=5463^4 \tag{46}$$
$$955^4+1770^4+2634^4+5400^4=5491^4 \tag{47}$$
$$30^4+1680^4+3043^4+5400^4=5543^4 \tag{48}$$
$$1354^4+1810^4+4355^4+5150^4=5729^4 \tag{49}$$
$$542^4+2770^4+4280^4+5695^4=6167^4 \tag{50}$$
$$50^4+885^4+5000^4+5984^4=6609^4 \tag{51}$$
$$1490^4+3468^4+4790^4+6185^4=6801^4 \tag{52}$$
$$1390^4+2850^4+5365^4+6368^4=7101^4 \tag{53}$$
$$160^4+1345^4+2790^4+7166^4=7209^4 \tag{54}$$
$$800^4+3052^4+5440^4+6635^4=7339^4 \tag{55}$$
$$2230^4+3196^4+5620^4+6995^4=7703^4 \tag{56}$$

(Norrie 1911, Patterson 1942, Leech 1958, Brudno 1964, Lander *et al.* 1967), but it is not known if there is a parametric solution (Guy 1994, p. 139).

Ramanujan gave the 4-2 equation

$$3^4+9^4=5^4+5^4+6^4+6^4, \tag{57}$$

and the 4-3 identities

$$2^4 + 2^4 + 7^4 = 4^4 + 4^4 + 5^4 + 6^4 \tag{58}$$
$$3^4 + 9^4 + 14^4 = 7^4 + 8^4 + 10^4 + 13^4 \tag{59}$$
$$7^4 + 10^4 + 13^4 = 5^4 + 5^4 + 6^4 + 14^4 \tag{60}$$

(Berndt 1994, p. 101). Haldeman (1904) gives general FORMULAS for 4-2 and 4-3 equations.

There are an infinite number of solutions to the 5-1 equation

$$A^4 + B^4 + C^4 + D^4 + E^4 = F^4. \tag{61}$$

Some of the smallest are

$$2^4 + 2^4 + 3^4 + 4^2 + 4^2 = 5^4 \tag{62}$$
$$4^4 + 6^4 + 8^4 + 9^4 + 14^4 = 15^4 \tag{63}$$
$$4^4 + 21^4 + 22^4 + 26^4 + 28^4 = 35^4 \tag{64}$$
$$1^4 + 2^4 + 12^4 + 24^4 + 44^4 = 45^4 \tag{65}$$
$$1^4 + 8^4 + 12^4 + 32^4 + 64^4 = 65^4 \tag{66}$$
$$2^4 + 39^4 + 44^4 + 46^4 + 52^4 = 65^4 \tag{67}$$
$$22^4 + 52^4 + 57^4 + 74^4 + 76^4 = 95^4 \tag{68}$$
$$22^4 + 28^4 + 63^4 + 72^4 + 94^4 = 105^4 \tag{69}$$

(Berndt 1994). Berndt and Bhargava (1993) and Berndt (1994, pp. 94–96) give Ramanujan's solutions for arbitrary s, t, m, and n,

$$\begin{aligned}(8s^2 + 40st - 24t^2)^4 + (6s^2 - 44st - 18t^2)^4 \\ +(14s^2 - 4st - 42t^2)^4 + (9s^2 + 27t^2)^4 + (4s^2 + 12t^2)^4 \\ = (15s^2 + 45t^2)^4,\end{aligned} \tag{70}$$

and

$$\begin{aligned}(4m^2 - 12n^2)^4 + (3m^2 + 9n^2)^4 + (2m^2 - 12mn - 6n^2)^4 \\ +(4m^2+12n^2)^4+(2m^2+12mn-6n^2)^4 = (5m^2+15n^2)^4.\end{aligned} \tag{71}$$

These are also given by Dickson (1966, p. 649), and two general FORMULAS are given by Beiler (1966, p. 290). Other solutions are given by Fauquembergue (1898), Haldeman (1904), and Martin (1910).

Ramanujan gave

$$2(ab + ac + bc)^2 = a^4 + b^4 + c^4 \tag{72}$$

$$2(ab+ac+bc)^4 = a^4(b-c)^4+b^4(c-a)^4+c^4(a-b)^4 \tag{73}$$

$$\begin{aligned}2(ab + ac + bc)^6 = (a^2b + b^2c + c^2a)^4 \\ +(ab^2 + bc^2 + ca^2)^4 + (3abc)^4\end{aligned} \tag{74}$$

$$\begin{aligned}2(ab + ac + bc)^8 = (a^3 + 2abc)^4(b - c)^4 \\ +(b^3 + 2abc)^4(c - a)^4 + (c^3 + 2abc)^4(a - b)^4,\end{aligned} \tag{75}$$

where

$$a + b + c = 0 \tag{76}$$

(Berndt 1994, pp. 96–97). FORMULA (73) is equivalent to FERRARI'S IDENTITY

$$\begin{aligned}(a^2 + 2ac - 2bc - b^2)^4 + (b^2 - 2ab - 2ac - c^2)^4 \\ +(c^2+2ab+2bc-a^2)^4 = 2(a^2+b^2+c^2-ab+ac+bc)^4.\end{aligned} \tag{77}$$

BHARGAVA'S THEOREM is a general identity which gives the above equations as a special case, and may have been the route by which Ramanujan proceeded. Another identity due to Ramanujan is

$$\begin{aligned}(a + b + c)^4 + (b + c + d)^4 + (a - d)^4 \\ = (c + d + a)^4 + (d + a + b)^4 + (b - c)^4,\end{aligned} \tag{78}$$

where $a/b = c/d$, and 4 may also be replaced by 2 (Ramanujan 1957, Hirschhorn 1998).

V. Kyrtatas noticed that $a = 3$, $b = 7$, $c = 20$, $d = 25$, $e = 38$, and $f = 39$ satisfy

$$\frac{a^4 + b^4 + c^4}{d^4 + e^4 + f^4} = \frac{a + b + c}{d + e + f} \tag{79}$$

and asks if there are any other distinct integer solutions.

The first few numbers n which are a sum of two or more fourth POWERS ($m - 1$ equations) are 353, 651, 2487, 2501, 2829, ... (Sloane's A003294). The only number of the form

$$4x^4 + y^4 \tag{80}$$

which is PRIME is 5 (Baudran 1885, Le Lionnais 1983).

see also BHARGAVA'S THEOREM, FORD'S THEOREM

References

Barbette, E. *Les sommes de p-iémes puissances distinctes égales à une p-iéme puissance.* Doctoral Dissertation, Liege, Belgium. Paris: Gauthier-Villars, 1910.

Beiler, A. H. *Recreations in the Theory of Numbers: The Queen of Mathematics Entertains.* New York: Dover, 1966.

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, 1994.

Berndt, B. C. and Bhargava, S. "Ramanujan—For Lowbrows." *Am. Math. Monthly* **100**, 645–656, 1993.

Bhargava, S. "On a Family of Ramanujan's Formulas for Sums of Fourth Powers." *Ganita* **43**, 63–67, 1992.

Brudno, S. "A Further Example of $A^4+B^4+C^4+D^4 = E^4$." *Proc. Cambridge Phil. Soc.* **60**, 1027–1028, 1964.

Dickson, L. E. *History of the Theory of Numbers, Vol. 2: Diophantine Analysis.* New York: Chelsea, 1966.

Euler, L. *Nova Acta Acad. Petrop. as annos 1795–1796* **13**, 45, 1802.

Fauquembergue, E. *L'intermédiaire des Math.* **5**, 33, 1898.

Ferrari, F. *L'intermédiaire des Math.* **20**, 105–106, 1913.

Guy, R. K. "Sums of Like Powers. Euler's Conjecture" and "Some Quartic Equations." §D1 and D23 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 139–144 and 192–193, 1994.

Haldeman, C. B. "On Biquadrate Numbers." *Math. Mag.* **2**, 285–296, 1904.

Hardy, G. H. and Wright, E. M. §13.7 in *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, 1979.

Hirschhorn, M. D. "Two or Three Identities of Ramanujan." *Amer. Math. Monthly* **105**, 52–55, 1998.

Lander, L. J.; Parkin, T. R.; and Selfridge, J. L. "A Survey of Equal Sums of Like Powers." *Math. Comput.* **21**, 446–459, 1967.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 56, 1983.

Leech, J. "Some Solutions of Diophantine Equations." *Proc. Cambridge Phil. Soc.* **53**, 778–780, 1957.

Leech, J. "On $A^4+B^4+C^4+D^4=E^4$." *Proc. Cambridge Phil. Soc.* **54**, 554–555, 1958.

Martin, A. "About Biquadrate Numbers whose Sum is a Biquadrate." *Math. Mag.* **2**, 173–184, 1896.

Martin, A. "About Biquadrate Numbers whose Sum is a Biquadrate—II." *Math. Mag.* **2**, 325–352, 1904.

Norrie, R. *University of St. Andrews 500th Anniversary Memorial Volume.* Edinburgh, Scotland: pp. 87–89, 1911.

Patterson, J. O. "A Note on the Diophantine Problem of Finding Four Biquadrates whose Sum is a Biquadrate." *Bull. Amer. Math. Soc.* **48**, 736–737, 1942.

Ramanujan, S. *Notebooks.* New York: Springer-Verlag, pp. 385–386, 1987.

Richmond, H. W. "On Integers Which Satisfy the Equation $t^3 \pm x^3 \pm y^3 \pm z^3 = 0$." *Trans. Cambridge Phil. Soc.* **22**, 389–403, 1920.

Sloane, N. J. A. Sequences A001235 and A003294/M5446 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Ward, M. "Euler's Problem on Sums of Three Fourth Powers." *Duke Math. J.* **15**, 827–837, 1948.

Diophantine Quadruple

see DIOPHANTINE SET

Diophantine Set

A set S of POSITIVE integers is said to be Diophantine IFF there exists a POLYNOMIAL Q with integral coefficients in $m \geq 1$ indeterminates such that

$$S = \{Q(x_1,\ldots,x_m) \geq 1 : x_1 \geq 1,\ldots,x_m \geq 1\}.$$

It has been proved that the set of PRIME numbers is a Diophantine set.

References

Ribenboim, P. *The New Book of Prime Number Records.* New York: Springer-Verlag, pp. 189–192, 1995.

Diophantus Property

A set of POSITIVE INTEGERS $S=\{a_1,\ldots,a_m\}$ satisfies the Diophantus property $D(n)$ of order n if, for all $i,j = 1, \ldots, m$ with $i \neq j$,

$$a_i a_j + n = {b_{ij}}^2, \tag{1}$$

where n and b_{ij} are INTEGERS. The set S is called a Diophantine n-tuple. Fermat found the first $D(1)$ quadruple: $\{1,3,8,120\}$. General $D(1)$ quadruples are

$$\{F_{2n}, F_{2n+2}, F_{2n+4}, 4F_{2n+1}F_{2n+2}F_{2n+3},\} \tag{2}$$

where F_n are FIBONACCI NUMBERS, and

$$\{n, n+2, 4n+4, 4(n+1)(2n+1)(2n+3)\}. \tag{3}$$

The quadruplet

$$\{2F_{n-1}, 2F_{n+1}, 2{F_n}^3F_{n+1}F_{n+2}, \\ 2F_{n+1}F_{n+2}F_{n+3}(2{F_{n+1}}^2 - {F_n}^2)\} \tag{4}$$

is $D({F_n}^2)$ (Dujella 1996). Dujella (1993) showed there exist no Diophantine quadruples $D(4k+2)$.

References

Aleksandriiskii, D. *Arifmetika i kniga o mnogougol'nyh chislakh.* Moscow: Nauka, 1974.

Brown, E. "Sets in Which $xy+k$ is Always a Square." *Math. Comput.* **45**, 613–620, 1985.

Davenport, H. and Baker, A. "The Equations $3x^2-2=y^2$ and $8x^2-7=z^2$." *Quart. J. Math. (Oxford) Ser. 2* **20**, 129–137, 1969.

Dujella, A. "Generalization of a Problem of Diophantus." *Acta Arithm.* **65**, 15–27, 1993.

Dujella, A. "Diophantine Quadruples for Squares of Fibonacci and Lucas Numbers." *Portugaliae Math.* **52**, 305–318, 1995.

Dujella, A. "Generalized Fibonacci Numbers and the Problem of Diophantus." *Fib. Quart.* **34**, 164–175, 1996.

Hoggatt, V. E. Jr. and Bergum, G. E. "A Problem of Fermat and the Fibonacci Sequence." *Fib. Quart.* **15**, 323–330, 1977.

Jones, B. W. "A Variation of a Problem of Davenport and Diophantus." *Quart. J. Math. (Oxford) Ser. (2)* **27**, 349–353, 1976.

Diophantus' Riddle

"Diophantus' youth lasts 1/6 of his life. He grew a beard after 1/12 more of his life. After 1/7 more of his life, Diophantus married. Five years later, he had a son. The son lived exactly half as long as his father, and Diophantus died just four years after his son's death. All of this totals the years Diophantus lived."

Let D be the number of years Diophantus lived, and let S be the number of years his son lived. Then the above word problem gives the two equations

$$\begin{aligned} D &= \left(\tfrac{1}{6}+\tfrac{1}{12}+\tfrac{1}{7}\right)D+5+S+4 \\ S &= \tfrac{1}{2}D. \end{aligned}$$

Solving this simultaneously gives $S=42$ as the age of the son and $D=84$ as the age of Diophantus.

References

Pappas, T. "Diophantus' Riddle." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 123 and 232, 1989.

Dipyramid

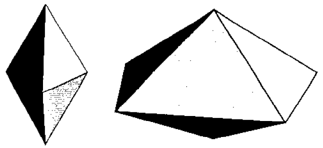

Two PYRAMIDS symmetrically placed base-to-base, also called a BIPYRAMID. They are the DUALS of the Archimedean PRISMS.

see also ELONGATED DIPYRAMID, PENTAGONAL DIPYRAMID, PRISM, PYRAMID, TRAPEZOHEDRON, TRIANGULAR DIPYRAMID, TRIGONAL DIPYRAMID

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 117, 1989.

Dirac Delta Function

see DELTA FUNCTION

Dirac Matrices

Define the 4×4 matrices

$$\sigma_i = \mathsf{I} \otimes \sigma_{i,\ \text{Pauli}} \tag{1}$$

$$\rho_i = \sigma_{i,\ \text{Pauli}} \otimes \mathsf{I}, \tag{2}$$

where $\sigma_{i,\ \text{Pauli}}$ are the PAULI MATRICES, I is the IDENTITY MATRIX, $i = 1, 2, 3$, and $\mathsf{A} \otimes \mathsf{B}$ is the matrix DIRECT PRODUCT. Explicitly,

$$\mathsf{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

$$\sigma_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{4}$$

$$\sigma_2 = \begin{bmatrix} 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{bmatrix} \tag{5}$$

$$\sigma_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \tag{6}$$

$$\rho_1 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{7}$$

$$\rho_2 = \begin{bmatrix} 0 & 0 & -i & 0 \\ 0 & 0 & 0 & -i \\ i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \end{bmatrix} \tag{8}$$

$$\rho_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \tag{9}$$

These matrices satisfy the anticommutation identities

$$\sigma_i\sigma_j + \sigma_j\sigma_i = 2\delta_{ij}\mathsf{I} \tag{10}$$

$$\rho_i\rho_j + \rho_j\rho_i = 2\delta_{ij}\mathsf{I}, \tag{11}$$

where δ_{ij} is the KRONECKER DELTA, the commutation identity

$$[\sigma_i, \rho_j] = \sigma_i\rho_j - \rho_j\sigma_i = 0, \tag{12}$$

and are cyclic under permutations of indices

$$\sigma_i\sigma_j = i\sigma_k \tag{13}$$

$$\rho_i\rho_j = i\rho_k. \tag{14}$$

A total of 16 Dirac matrices can be defined via

$$\mathsf{E}_{ij} = \rho_i\sigma_j \tag{15}$$

for $i, j = 0, 1, 2, 3$ and where $\sigma_0 = \rho_0 \equiv \mathsf{I}$. These matrix satisfy

1. $|\mathsf{E}_{ij}| = 1$, where $|\mathsf{A}|$ is the DETERMINANT,
2. $\mathsf{E}_{ij}^2 = \mathsf{I}$,
3. $\mathsf{E}_{ij} = \mathsf{E}_{ij}^\dagger$, making them Hermitian, and therefore unitary,
4. $\operatorname{tr}(\mathsf{E}_{ij}) = 0$, except $\operatorname{tr}(\mathsf{E}_{00}) = 4$,
5. Any two E_{ij} multiplied together yield a Dirac matrix to within a multiplicative factor of $-i$ or $\pm i$,
6. The E_{ij} are linearly independent,
7. The E_{ij} form a complete set, i.e., any 4×4 constant matrix may be written as

$$\mathsf{A} = \sum_{i,j=0}^{3} c_{ij}\mathsf{E}_{ij}, \tag{16}$$

where the c_{ij} are real or complex and are given by

$$c_{mn} = \tfrac{1}{4}\operatorname{tr}(\mathsf{A}\mathsf{E}_{mn}) \tag{17}$$

(Arfken 1985).

Dirac's original matrices were written α_i and were defined by

$$\alpha_i = \mathsf{E}_{1i} = \rho_1\sigma_i \tag{18}$$

$$\alpha_4 = \mathsf{E}_{30} = \rho_3, \tag{19}$$

for $i = 1, 2, 3$, giving

$$\alpha_1 = \mathsf{E}_{1i} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{20}$$

$$\alpha_2 = \mathsf{E}_{2i} = \begin{bmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{bmatrix} \tag{21}$$

$$\alpha_3 = \mathsf{E}_{3i} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} \tag{22}$$

$$\alpha_4 = \mathsf{E}_{30} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \tag{23}$$

The additional matrix

$$\alpha_5 = \mathsf{E}_{20} = \rho_2 = \begin{bmatrix} 0 & 0 & -i & 0 \\ 0 & 0 & 0 & -i \\ i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \end{bmatrix} \qquad (24)$$

is sometimes defined. Other sets of Dirac matrices are sometimes defined as

$$y_i = \mathsf{E}_{2i} \qquad (25)$$

$$y_4 = \mathsf{E}_{30} \qquad (26)$$

$$y_5 = -\mathsf{E}_{10} \qquad (27)$$

and

$$\delta_i = \mathsf{E}_{3i} \qquad (28)$$

for $i = 1, 2, 3$ (Arfken 1985) and

$$\gamma_i = \begin{bmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{bmatrix} \qquad (29)$$

$$\gamma_4 = \begin{bmatrix} \mathsf{I} & 0 \\ 2\mathsf{I} & -\mathsf{I} \end{bmatrix} \qquad (30)$$

for $i = 1, 2, 3$ (Goldstein 1980).

Any of the 15 Dirac matrices (excluding the identity matrix) commute with eight Dirac matrices and anticommute with the other eight. Let $\mathsf{M} \equiv \frac{1}{2}(1 + \mathsf{E}_{ij})$, then

$$\mathsf{M}^2 = \mathsf{M}. \qquad (31)$$

In addition

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \times \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = 2i\sigma. \qquad (32)$$

The products of α_i and y_i satisfy

$$\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5 = 1 \qquad (33)$$

$$y_1y_2y_3y_4y_5 = 1. \qquad (34)$$

The 16 Dirac matrices form six anticommuting sets of five matrices each:

1. α_1, α_2, α_3, α_4, α_5,
2. y_1, y_2, y_3, y_4, y_5,
3. δ_1, δ_2, δ_3, ρ_1, ρ_2,
4. α_1, y_1, δ_1, σ_2, σ_3,
5. α_2, y_2, δ_2, σ_1, σ_3,
6. α_3, y_3, δ_3, σ_1, σ_2.

see also PAULI MATRICES

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 211–213, 1985.

Goldstein, H. *Classical Mechanics, 2nd ed.* Reading, MA: Addison-Wesley, p. 580, 1980.

Dirac's Theorem

A GRAPH with $n \geq 3$ VERTICES in which each VERTEX has VALENCY $\geq n/2$ has a HAMILTONIAN CIRCUIT.

see also HAMILTONIAN CIRCUIT

Direct Product (Group)

The expression of a GROUP as a product of SUBGROUPS. The CHARACTERS of the representations of a direct product are equal to the products of the CHARACTERS of the representations based on the individual sets of functions. For R_1 and R_2,

$$\chi(R_1 \otimes R_2) = \chi(R_1)\chi(R_2).$$

The representation of a direct product Γ_{AB} will contain the totally symmetric representation only if the irreducible Γ_A equals the irreducible Γ_B.

Direct Product (Matrix)

Given two $n \times m$ MATRICES, their direct product $\mathsf{C} = \mathsf{A} \otimes \mathsf{B}$ is an $(mn) \times (nm)$ MATRIX with elements defined by

$$C_{\alpha\beta} = \mathsf{A}_{ij}\mathsf{B}_{kl}, \qquad (1)$$

where

$$\alpha \equiv n(i-1) + k \qquad (2)$$

$$\beta \equiv n(j-1) + l. \qquad (3)$$

For a 2×2 MATRIX,

$$\mathsf{A} \otimes \mathsf{B} = \begin{bmatrix} a_{11}\mathsf{B} & a_{12}\mathsf{B} \\ a_{21}\mathsf{B} & a_{22}\mathsf{B} \end{bmatrix} \qquad (4)$$

$$= \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{bmatrix}. \qquad (5)$$

Direct Product (Set)

The direct product of two sets A and B is defined to be the set of all points (a, b) where $a \in A$ and $b \in B$. The direct product is denoted $A \times B$ or $A \otimes B$ and is also called the CARTESIAN PRODUCT, since it originated in Descartes' formulation of analytic geometry. In the Cartesian view, points in the plane are specified by their vertical and horizontal coordinates, with points on a line being specified by just one coordinate. The main examples of direct products are EUCLIDEAN 3-space ($\mathbb{R} \otimes \mathbb{R} \otimes \mathbb{R}$, where $\mathbb{R}$ are the REAL NUMBERS), and the plane ($\mathbb{R} \times \mathbb{R}$).

Direct Product (Tensor)

For a first-RANK TENSOR (i.e., a VECTOR),

$$a_i' b'^j \equiv \frac{\partial x_k}{\partial x_i'} a_k \frac{\partial x_j'}{\partial x_l} b^l = \frac{\partial x_k}{\partial x_i'} \frac{\partial x_j'}{\partial x_l} (a_k b^l), \tag{1}$$

which is a second-RANK TENSOR. The CONTRACTION of a direct product of first-RANK TENSORS is the SCALAR

$$\operatorname{contr}(a_i' b'^j) = a_i' b'^i = a_k b^k. \tag{2}$$

For a second-RANK TENSOR,

$$A^i_j B_{kl} = C^{ikl}_j \tag{3}$$

$$C_j^{ikl\prime} = \frac{\partial x_i'}{\partial x_m} \frac{\partial x_n}{\partial x_j'} \frac{\partial x_k'}{\partial x_p} \frac{\partial x_l'}{\partial x_q} C_n^{mpq}. \tag{4}$$

For a general TENSOR, the direct product of two TENSORS is a TENSOR of RANK equal to the sum of the two initial RANKS. The direct product is ASSOCIATIVE, but not COMMUTATIVE.

References

Arfken, G. "Contraction, Direct Product." §3.2 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 124–126, 1985.

Direct Search Factorization

Direct search factorization is the simplest PRIME FACTORIZATION ALGORITHM. It consists of searching for factors of a number by systematically performing TRIAL DIVISIONS, usually using a sequence of increasing numbers. Multiples of small PRIMES are commonly excluded to reduce the number of trial DIVISORS, but just including them is sometimes faster than the time required to exclude them. This approach is very inefficient, and can be used only with fairly small numbers.

When using this method on a number n, only DIVISORS up to $\lfloor\sqrt{n}\rfloor$ (where $\lfloor x \rfloor$ is the FLOOR FUNCTION) need to be tested. This is true since if all INTEGERS less than this had been tried, then

$$\frac{n}{\lfloor\sqrt{n}\rfloor + 1} < \sqrt{n}. \tag{1}$$

In other words, all possible FACTORS have had their COFACTORS already tested. It is also true that, when the smallest PRIME FACTOR p of n is $> \sqrt[3]{n}$, then its COFACTOR m (such that $n = pm$) must be PRIME. To prove this, suppose that the smallest p is $> \sqrt[3]{n}$. If $m = ab$, then the smallest value a and b could assume is p. But then

$$n = pm = pab = p^3 > n, \tag{2}$$

which cannot be true. Therefore, m must be PRIME, so

$$n = p_1 p_2. \tag{3}$$

see also PRIME FACTORIZATION ALGORITHMS, TRIAL DIVISION

Direct Sum (Module)

The direct sum of two MODULES V and W over the same RING R is given by $V \otimes W$ with MODULE operations defined by

$$r \cdot (v, w) = (rv, rw)$$

$$(v, w) \oplus (y, z) = (v + y, w + z).$$

The direct sum of an arbitrary family of MODULES over the same RING is also defined. If J is the indexing set for the family of MODULES, then the direct sum is represented by the collection of functions with finite support from J to the union of all these MODULES such that the function sends $j \in J$ to an element in the MODULE indexed by j.

The dimension of a direct sum is the product of the dimensions of the quantities summed. The significant property of the direct sum is that it is the coproduct in the category of MODULES. This general definition gives as a consequence the definition of the direct sum of ABELIAN GROUPS (since they are MODULES over the INTEGERS) and the direct sum of VECTOR SPACES (since they are MODULES over a FIELD).

Directed Angle

The symbol $\measuredangle ABC$ denotes the directed angle from AB to BC, which is the signed angle through which AB must be rotated about B to coincide with BC. Four points $ABCD$ lie on a CIRCLE (i.e., are CONCYCLIC) IFF $\measuredangle ABC = \measuredangle ADC$. It is also true that

$$\measuredangle l_1 l_2 + \measuredangle l_2 l_1 = 0^\circ \text{ or } 180^\circ.$$

Three points A, B, and C are COLLINEAR IFF $\measuredangle ABC = 0$. For any four points, A, B, C, and D,

$$\measuredangle ABC + \measuredangle CDA = \measuredangle BAD + \measuredangle DCB.$$

see also ANGLE, COLLINEAR, CONCYCLIC, MIQUEL EQUATION

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 11–15, 1929.

Directed Graph

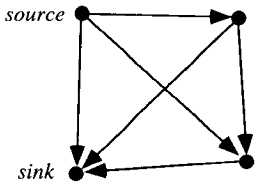

A GRAPH in which each EDGE is replaced by a directed EDGE, also called a DIGRAPH or REFLEXIVE GRAPH. A COMPLETE directed graph is called a TOURNAMENT. If G is an undirected connected GRAPH, then one can always direct the circuit EDGES of G and leave the SEPARATING EDGES undirected so that there is a directed path from any node to another. Such a GRAPH is said to be transitive if the adjacency relation is transitive. The number of directed graphs of n nodes for $n = 1$, 2, ... are 1, 1, 3, 16, 218, 9608, ... (Sloane's A000273).

see also ARBORESCENCE, CAYLEY GRAPH, INDEGREE, NETWORK, OUTDEGREE, SINK (DIRECTED GRAPH), SOURCE, TOURNAMENT

References

Sloane, N. J. A. Sequence A000273/M3032 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Direction Cosine

Let a be the ANGLE between $\mathbf{v}$ and $\mathbf{x}$, b the ANGLE between $\mathbf{v}$ and $\mathbf{y}$, and c the ANGLE between $\mathbf{v}$ and $\mathbf{z}$. Then the direction cosines are equivalent to the (x, y, z) coordinates of a UNIT VECTOR $\hat{\mathbf{v}}$,

$$\alpha \equiv \cos a \equiv \frac{\mathbf{v} \cdot \hat{\mathbf{x}}}{|\mathbf{v}|} \tag{1}$$

$$\beta \equiv \cos b \equiv \frac{\mathbf{v} \cdot \hat{\mathbf{y}}}{|\mathbf{v}|} \tag{2}$$

$$\gamma \equiv \cos c \equiv \frac{\mathbf{v} \cdot \hat{\mathbf{z}}}{|\mathbf{v}|}. \tag{3}$$

From these definitions, it follows that

$$\alpha^2 + \beta^2 + \gamma^2 = 1. \tag{4}$$

To find the JACOBIAN when performing integrals over direction cosines, use

$$\theta = \sin^{-1}\left(\sqrt{\alpha^2 + \beta^2}\right) \tag{5}$$

$$\phi = \tan^{-1}\left(\frac{\beta}{\alpha}\right) \tag{6}$$

$$\gamma = \sqrt{1 - \alpha^2 - \beta^2}. \tag{7}$$

The JACOBIAN is

$$\left|\frac{\partial(\theta, \phi)}{\partial(\alpha, \beta)}\right| = \begin{vmatrix} \frac{\partial\theta}{\partial\alpha} & \frac{\partial\theta}{\partial\beta} \\ \frac{\partial\phi}{\partial\alpha} & \frac{\partial\phi}{\partial\beta} \end{vmatrix}. \tag{8}$$

Using

$$\frac{d}{dx}(\sin^{-1} x) = \frac{1}{\sqrt{1 - x^2}} \tag{9}$$

$$\frac{d}{dx}(\tan^{-1} x) = \frac{1}{1 + x^2}, \tag{10}$$

$$\begin{aligned} \left|\frac{\partial(\theta, \phi)}{\partial(\alpha, \beta)}\right| &= \begin{vmatrix} \frac{\frac{1}{2}(\alpha^2+\beta^2)^{-1/2}2\alpha}{\sqrt{1-\alpha^2-\beta^2}} & \frac{\frac{1}{2}(\alpha^2+\beta^2)^{-1/2}2\beta}{\sqrt{1-\alpha^2-\beta^2}} \\ \frac{-\alpha^{-2}\beta}{1+\frac{\beta^2}{\alpha^2}} & \frac{\alpha^{-1}}{1+\frac{\beta^2}{\alpha^2}} \end{vmatrix} \\ &= \frac{1}{\sqrt{1 - \alpha^2 - \beta^2}} \frac{(\alpha^2 + \beta^2)^{-1/2}}{1 + \frac{\beta^2}{\alpha^2}} \left(1 + \frac{\beta^2}{\alpha^2}\right) \\ &= \frac{1}{\sqrt{(\alpha^2 + \beta^2)(1 - \alpha^2 - \beta^2)}}, \end{aligned} \tag{11}$$

so

$$\begin{aligned} d\Omega = \sin\theta\, d\phi\, d\theta &= \sqrt{\alpha^2 + \beta^2} \left|\frac{\partial(\theta, \phi)}{\partial(\alpha, \beta)}\right| d\alpha\, d\beta \\ &= \frac{d\alpha\, d\beta}{\sqrt{1 - \alpha^2 - \beta^2}} = \frac{d\alpha\, d\beta}{\gamma}. \end{aligned} \tag{12}$$

Direction cosines can also be defined between two sets of CARTESIAN COORDINATES,

$$\alpha_1 \equiv \hat{\mathbf{x}}' \cdot \hat{\mathbf{x}} \tag{13}$$

$$\alpha_2 \equiv \hat{\mathbf{x}}' \cdot \hat{\mathbf{y}} \tag{14}$$

$$\alpha_3 \equiv \hat{\mathbf{x}}' \cdot \hat{\mathbf{z}} \tag{15}$$

$$\beta_1 \equiv \hat{\mathbf{y}}' \cdot \hat{\mathbf{x}} \tag{16}$$

$$\beta_2 \equiv \hat{\mathbf{y}}' \cdot \hat{\mathbf{y}} \tag{17}$$

$$\beta_3 \equiv \hat{\mathbf{y}}' \cdot \hat{\mathbf{z}} \tag{18}$$

$$\gamma_1 \equiv \hat{\mathbf{z}}' \cdot \hat{\mathbf{x}} \tag{19}$$

$$\gamma_2 \equiv \hat{\mathbf{z}}' \cdot \hat{\mathbf{y}} \tag{20}$$

$$\gamma_3 \equiv \hat{\mathbf{z}}' \cdot \hat{\mathbf{z}}. \tag{21}$$

Projections of the unprimed coordinates onto the primed coordinates yield

$$\hat{\mathbf{x}}' = (\hat{\mathbf{x}}' \cdot \hat{\mathbf{x}})\hat{\mathbf{x}} + (\hat{\mathbf{x}}' \cdot \hat{\mathbf{y}})\hat{\mathbf{y}} + (\hat{\mathbf{x}}' \cdot \hat{\mathbf{z}})\hat{\mathbf{z}} = \alpha_1\hat{\mathbf{x}} + \alpha_2\hat{\mathbf{y}} + \alpha_3\hat{\mathbf{z}} \tag{22}$$

$$\hat{\mathbf{y}}' = (\hat{\mathbf{y}}' \cdot \hat{\mathbf{x}})\hat{\mathbf{x}} + (\hat{\mathbf{y}}' \cdot \hat{\mathbf{y}})\hat{\mathbf{y}} + (\hat{\mathbf{y}}' \cdot \hat{\mathbf{z}})\hat{\mathbf{z}} = \beta_1\hat{\mathbf{x}} + \beta_2\hat{\mathbf{y}} + \beta_3\hat{\mathbf{z}} \tag{23}$$

$$\hat{\mathbf{z}}' = (\hat{\mathbf{z}}' \cdot \hat{\mathbf{x}})\hat{\mathbf{x}} + (\hat{\mathbf{z}}' \cdot \hat{\mathbf{x}})\hat{\mathbf{y}} + (\hat{\mathbf{z}}' \cdot \hat{\mathbf{z}})\hat{\mathbf{z}} = \gamma_1\hat{\mathbf{x}} + \gamma_2\hat{\mathbf{y}} + \gamma_3\hat{\mathbf{z}}, \tag{24}$$

and

$$x' = \mathbf{r} \cdot \hat{\mathbf{x}}' = \alpha_1 x + \alpha_2 y + \alpha_3 z \tag{25}$$

$$y' = \mathbf{r} \cdot \hat{\mathbf{y}}' = \beta_1 x + \beta_2 y + \beta_3 z \tag{26}$$

$$z' = \mathbf{r} \cdot \hat{\mathbf{z}}' = \gamma_1 x + \gamma_2 y + \gamma_3 z. \tag{27}$$

Projections of the primed coordinates onto the unprimed coordinates yield

$$\begin{aligned} \hat{\mathbf{x}} &= (\hat{\mathbf{x}} \cdot \hat{\mathbf{x}}')\hat{\mathbf{x}}' + (\hat{\mathbf{x}} \cdot \hat{\mathbf{y}}')\hat{\mathbf{y}}' + (\hat{\mathbf{x}} \cdot \hat{\mathbf{z}}')\hat{\mathbf{z}}' \\ &= \alpha_1\hat{\mathbf{x}}' + \beta_1\hat{\mathbf{y}}' + \gamma_1\hat{\mathbf{z}}' \end{aligned} \tag{28}$$

$$\begin{aligned} \hat{\mathbf{y}} &= (\hat{\mathbf{y}} \cdot \hat{\mathbf{x}}')\hat{\mathbf{x}}' + (\hat{\mathbf{y}} \cdot \hat{\mathbf{y}}')\hat{\mathbf{y}}' + (\hat{\mathbf{y}} \cdot \hat{\mathbf{z}}')\hat{\mathbf{z}}' \\ &= \alpha_2\hat{\mathbf{x}}' + \beta_2\hat{\mathbf{y}}' + \gamma_2\hat{\mathbf{z}}' \end{aligned} \tag{29}$$

$$\begin{aligned} \hat{\mathbf{z}} &= (\hat{\mathbf{z}} \cdot \hat{\mathbf{x}}')\hat{\mathbf{x}}' + (\hat{\mathbf{z}} \cdot \hat{\mathbf{x}}')\hat{\mathbf{y}}' + (\hat{\mathbf{z}} \cdot \hat{\mathbf{z}}')\hat{\mathbf{z}}' \\ &= \alpha_3\hat{\mathbf{x}}' + \beta_3\hat{\mathbf{y}}' + \gamma_3\hat{\mathbf{z}}', \end{aligned} \tag{30}$$

and

$$x = \mathbf{r} \cdot \hat{\mathbf{x}} = \alpha_1 x + \beta_1 y + \gamma_1 z \tag{31}$$

$$y = \mathbf{r} \cdot \hat{\mathbf{y}} = \alpha_2 x + \beta_2 y + \gamma_2 z \tag{32}$$

$$z = \mathbf{r} \cdot \hat{\mathbf{z}} = \alpha_3 x + \beta_3 y + \gamma_3 z. \qquad (33)$$

Using the orthogonality of the coordinate system, it must be true that

$$\hat{\mathbf{x}} \cdot \hat{\mathbf{y}} = \hat{\mathbf{y}} \cdot \hat{\mathbf{z}} = \hat{\mathbf{z}} \cdot \hat{\mathbf{x}} = 0 \qquad (34)$$

$$\hat{\mathbf{x}} \cdot \hat{\mathbf{x}} = \hat{\mathbf{y}} \cdot \hat{\mathbf{y}} = \hat{\mathbf{z}} \cdot \hat{\mathbf{z}} = 1, \qquad (35)$$

giving the identities

$$\alpha_l \alpha_m + \beta_l \beta_m + \gamma_l \gamma_m = 0 \qquad (36)$$

for $l, m = 1, 2, 3$ and $l \neq m$, and

$$\alpha_l{}^2 + \beta_l{}^2 + \gamma_l{}^2 = 1 \qquad (37)$$

for $l = 1, 2, 3$. These two identities may be combined into the single identity

$$\alpha_l \alpha_m + \beta_l \beta_m + \gamma_l \gamma_m = \delta_{lm}, \qquad (38)$$

where δ_{lm} is the KRONECKER DELTA.

Directional Derivative

$$\nabla_{\mathbf{u}} f \equiv \nabla f \cdot \frac{\mathbf{u}}{|\mathbf{u}|} \propto \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{u}) - f(\mathbf{x})}{h}. \qquad (1)$$

$\nabla_{\mathbf{u}} f(x_0, y_0, z_0)$ is the rate at which the function $w = f(x, y, z)$ changes at (x_0, y_0, z_0) in the direction $\mathbf{u}$. Let $\mathbf{u}$ be a UNIT VECTOR in CARTESIAN COORDINATES, so

$$|\mathbf{u}| = \sqrt{u_x{}^2 + u_y{}^2 + u_z{}^2} = 1, \qquad (2)$$

then

$$\nabla_{\mathbf{u}} f = \frac{\partial f}{\partial x} u_x + \frac{\partial f}{\partial y} u_y + \frac{\partial f}{\partial z} u_z. \qquad (3)$$

The directional derivative is often written in the notation

$$\frac{d}{ds} \equiv \hat{\mathbf{s}} \cdot \nabla = s_x \frac{\partial}{\partial x} + s_y \frac{\partial}{\partial y} + s_z \frac{\partial}{\partial z}. \qquad (4)$$

Directly Similar

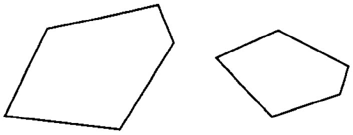

directly similar

Two figures are said to be SIMILAR when all corresponding ANGLES are equal, and are directly similar when all corresponding ANGLES are equal and described in the same rotational sense.

see also FUNDAMENTAL THEOREM OF DIRECTLY SIMILAR FIGURES, INVERSELY SIMILAR, SIMILAR

Director Curve

The curve $\mathbf{d}(u)$ in the RULED SURFACE parameterization

$$\mathbf{x}(u, v) = \mathbf{b}(u) + v\mathbf{d}(u).$$

see also DIRECTRIX (RULED SURFACE), RULED SURFACE, RULING

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 333, 1993.

Directrix (Conic Section)

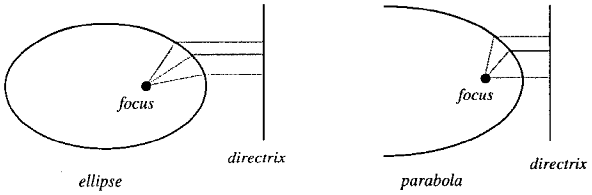

The LINE which, together with the point known as the FOCUS, serves to define a CONIC SECTION.

see also CONIC SECTION, ELLIPSE, FOCUS, HYPERBOLA, PARABOLA

References

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 115–116, 1969.

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 141–144, 1967.

Directrix (Graph)

A CYCLE.

Directrix (Ruled Surface)

The curve $\mathbf{b}(u)$ in the RULED SURFACE parameterization

$$\mathbf{x}(u, v) = \mathbf{b}(u) + v\mathbf{d}(u)$$

is called the directrix (or BASE CURVE).

see also DIRECTOR CURVE, RULED SURFACE

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 333, 1993.

Dirichlet Beta Function

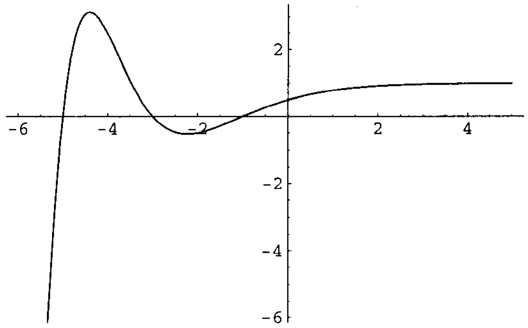

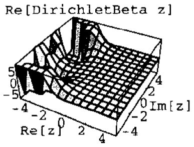

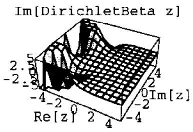

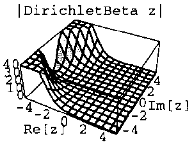

$$\beta(x) \equiv \sum_{n=0}^{\infty} (-1)^n (2n+1)^{-x} \qquad (1)$$

$$\beta(x) = 2^{-x}\Phi(-1, x, \tfrac{1}{2}), \qquad (2)$$

where Φ is the LERCH TRANSCENDENT. The beta function can be written in terms of the HURWITZ ZETA FUNCTION $\zeta(x,a)$ by

$$\beta(x) = \frac{1}{4^x}[\zeta(x, \tfrac{1}{4}) - \zeta(x, \tfrac{3}{4})]. \qquad (3)$$

The beta function can be evaluated directly for POSITIVE ODD x as

$$\beta(2k+1) = \frac{(-1)^k E_{2k}}{2(2k)!}(\tfrac{1}{2}\pi)^{2k+1}, \qquad (4)$$

where E_n is an EULER NUMBER. The beta function can be defined over the whole COMPLEX PLANE using ANALYTIC CONTINUATION,

$$\beta(1-z) = \left(\frac{2}{\pi}\right)^z \sin(\tfrac{1}{2}\pi z)\Gamma(z)\beta(z), \qquad (5)$$

where $\Gamma(z)$ is the GAMMA FUNCTION.

Particular values for β are

$$\beta(1) = \tfrac{1}{4}\pi \qquad (6)$$

$$\beta(2) \equiv K \qquad (7)$$

$$\beta(3) = \tfrac{1}{32}\pi^3, \qquad (8)$$

where K is CATALAN'S CONSTANT.

see also CATALAN'S CONSTANT, DIRICHLET ETA FUNCTION, DIRICHLET LAMBDA FUNCTION, HURWITZ ZETA FUNCTION, LERCH TRANSCENDENT, RIEMANN ZETA FUNCTION, ZETA FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 807–808, 1972.

Spanier, J. and Oldham, K. B. "The Zeta Numbers and Related Functions." Ch. 3 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 25–33, 1987.

Dirichlet Boundary Conditions

PARTIAL DIFFERENTIAL EQUATION BOUNDARY CONDITIONS which give the value of the function on a surface, e.g., $T = f(\mathbf{r}, t)$.

see also BOUNDARY CONDITIONS, CAUCHY BOUNDARY CONDITIONS

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 679, 1953.

Dirichlet's Box Principle

A.k.a. the PIGEONHOLE PRINCIPLE. Given n boxes and $m > n$ objects, at least one box must contain more than one object. This statement has important applications in number theory and was first stated by Dirichlet in 1834.

see also FUBINI PRINCIPLE

References

Chartrand, G. *Introductory Graph Theory.* New York: Dover, p. 38, 1985.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 161, 1993.

Dirichlet's Boxing-In Principle

see DIRICHLET'S BOX PRINCIPLE

Dirichlet Conditions

see DIRICHLET BOUNDARY CONDITIONS, DIRICHLET FOURIER SERIES CONDITIONS

Dirichlet Divisor Problem

Let $d(n) = \nu(n) = \sigma_0(n)$ be the number of DIVISORS of n (including n itself). For a PRIME p, $\nu(p) = 2$. In general,

$$\sum_{k=1}^{n} \nu(k) = n \ln n + (2\gamma - 1)n + \mathcal{O}(n^\theta),$$

where γ is the EULER-MASCHERONI CONSTANT. Dirichlet originally gave $\theta \approx 1/2$. As of 1988, this had been reduced to $\theta \approx 7/22$.

see also DIVISOR FUNCTION

Dirichlet Energy

Let h be a real-valued HARMONIC FUNCTION on a bounded DOMAIN Ω, then the Dirichlet energy is defined as $\int_\Omega |\nabla h|^2\, dx$, where ∇ is the GRADIENT.

see also ENERGY

Dirichlet Eta Function

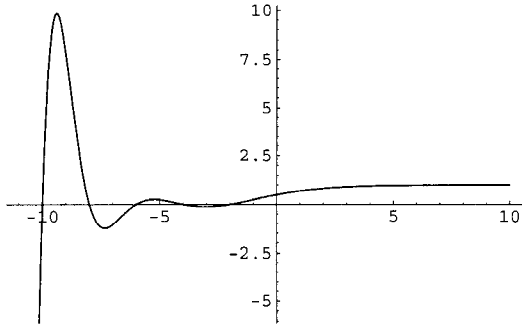

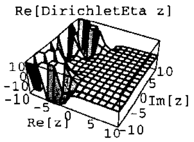

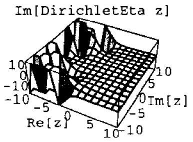

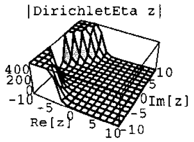

$$\eta(x) \equiv \sum_{n=1}^{\infty} (-1)^{n-1} n^{-x} = (1 - 2^{1-x})\zeta(x), \qquad (1)$$

where $n = 1, 2, \ldots$, and $\zeta(x)$ is the RIEMANN ZETA FUNCTION. Particular values are given in Abramowitz and Stegun (1972, p. 811). The eta function is related to the RIEMANN ZETA FUNCTION and DIRICHLET LAMBDA FUNCTION by

$$\frac{\zeta(\nu)}{2^\nu} = \frac{\lambda(\nu)}{2^\nu - 1} = \frac{\eta(\nu)}{2^\nu - 2} \qquad (2)$$

and

$$\zeta(\nu) + \eta(\nu) = 2\lambda(\nu) \qquad (3)$$

(Spanier and Oldham 1987). The value $\eta(1)$ may be computed by noting that the MACLAURIN SERIES for $\ln(1+x)$ for $-1 < x \leq 1$ is

$$\ln(1+x) = x - \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 - \tfrac{1}{4}x^4 + \ldots. \qquad (4)$$

Therefore,

$$\begin{aligned} \ln 2 = \ln(1+1) &= 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \ldots \\ &= \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} = \eta(1). \end{aligned} \qquad (5)$$

Values for EVEN INTEGERS are related to the analytical values of the RIEMANN ZETA FUNCTION. $\eta(0)$ is defined to be $\frac{1}{2}$.

$$\begin{aligned} \eta(0) &= \tfrac{1}{2} \\ \eta(1) &= \ln 2 \\ \eta(2) &= \frac{\pi^2}{12} \\ \eta(3) &= 0.90154\ldots \\ \eta(4) &= \frac{7\pi^4}{720}. \end{aligned}$$

see also DEDEKIND ETA FUNCTION, DIRICHLET BETA FUNCTION, DIRICHLET LAMBDA FUNCTION, RIEMANN ZETA FUNCTION, ZETA FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 807–808, 1972.

Spanier, J. and Oldham, K. B. "The Zeta Numbers and Related Functions." Ch. 3 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 25–33, 1987.

Dirichlet's Formula

If g is continuous and $\mu, \nu > 0$, then

$$\int_0^t (t-\xi)^{\mu-1}\,d\xi \int_0^\xi (\xi - x)^{\nu-1} g(\xi, x)\,dx = \int_0^t dx \int_x^t (t-\xi)^{\mu-1}(\xi - x)^{\nu-1} g(\xi, x)\,d\xi.$$

Dirichlet Fourier Series Conditions

A piecewise regular function which

1. Has a finite number of finite discontinuities and
2. Has a finite number of extrema

can be expanded in a FOURIER SERIES which converges to the function at continuous points and the mean of the POSITIVE and NEGATIVE limits at points of discontinuity.

see also FOURIER SERIES

Dirichlet Function

Let c and $d \neq c$ be REAL NUMBERS (usually taken as $c = 1$ and $d = 0$). The Dirichlet function is defined by

$$D(x) = \begin{cases} c & \text{for } x \text{ rational} \\ d & \text{for } x \text{ irrational.} \end{cases}$$

The function is CONTINUOUS at IRRATIONAL x and discontinuous at RATIONAL points. The function can be written analytically as

$$D(x) = \lim_{m,n\to\infty} \cos[(m!\pi x)^n].$$

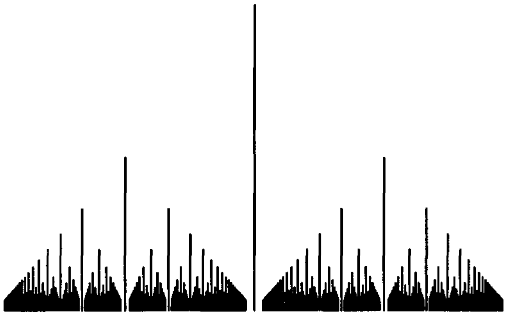

Because the Dirichlet function cannot be plotted without producing a solid blend of lines, a modified version can be defined as

$$D_M(x) = \begin{cases} 0 & \text{for } x \text{ rational} \\ b & \text{for } x = a/b \text{ with } a/b \text{ a reduced fraction} \end{cases}$$

(Dixon 1991), illustrated above.

see also CONTINUOUS FUNCTION, IRRATIONAL NUMBER, RATIONAL NUMBER

References

Dixon, R. *Mathographics.* New York: Dover, pp. 177 and 184–186, 1991.

Tall, D. "The Gradient of a Graph." *Math. Teaching* **111**, 48–52, 1985.

Dirichlet Integrals

There are several types of integrals which go under the name of a "Dirichlet integral." The integral

$$D[u] = \int_\Omega |\nabla u|^2\,dV \qquad (1)$$

appears in DIRICHLET'S PRINCIPLE.

The integral

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\frac{\sin[(n+\frac{1}{2})x]}{\sin(\frac{1}{2}x)}\,dx, \tag{2}$$

where the kernel is the DIRICHLET KERNEL, gives the nth partial sum of the FOURIER SERIES.

Another integral is denoted

$$\delta_k \equiv \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\sin\alpha_k\rho_k}{\rho_k}e^{i\rho_k\gamma_k}\,d\rho_k = \begin{cases} 0 & \text{for } |\gamma_k| > \alpha_k \\ 1 & \text{for } |\gamma_k| < \alpha_k \end{cases} \tag{3}$$

for $k = 1, \ldots, n$.

There are two types of Dirichlet integrals which are denoted using the letters C, D, I, and J. The type 1 Dirichlet integrals are denoted I, J, and IJ, and the type 2 Dirichlet integrals are denoted C, D, and CD.

The type 1 integrals are given by

$$\begin{aligned} I &\equiv \iint\cdots\int f(t_1+t_2+\ldots+t_n) \\ &\quad \times {t_1}^{\alpha_1-1}{t_2}^{\alpha_2-1}\cdots{t_n}^{\alpha_n-1}\,dt_1\,dt_2\,dt_n \\ &= \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_n)}{\Gamma\left(\sum_n \alpha_n\right)}\int_0^1 f(\tau)\tau^{\left(\sum_n \alpha\right)-1}\,d\tau, \end{aligned} \tag{4}$$

where $\Gamma(z)$ is the GAMMA FUNCTION. In the case $n = 2$,

$$I = \iint_T x^p y^q\,dx\,dy = \frac{p!q!}{(p+q+2)!} = \frac{B(p+1,q+1)}{p+q+2}, \tag{5}$$

where the integration is over the TRIANGLE T bounded by the x-axis, y-axis, and line $x + y = 1$ and $B(x, y)$ is the BETA FUNCTION.

The type 2 integrals are given for b-D vectors $\mathbf{a}$ and $\mathbf{r}$, and $0 \leq c \leq b$,

$$\begin{aligned} C_{\mathbf{a}}^{(b)}(\mathbf{r},m) &= \frac{\Gamma(m+R)}{\Gamma(m)\prod_{i=1}^b\Gamma(r_i)} \\ &\quad \times\int_0^{a_1}\cdots\int_0^{a_b}\frac{\prod_{i=1}^b {x_i}^{r_i-1}\,dx_i}{\left(1+\sum_{i=1}^b x_i\right)^{m+R}} \end{aligned} \tag{6}$$

$$\begin{aligned} D_{\mathbf{a}}^{(b)}(\mathbf{r},m) &= \frac{\Gamma(m+R)}{\Gamma(m)\prod_{i=1}^b\Gamma(r_i)} \\ &\quad \times\int_{a_1}^{\infty}\cdots\int_{a_k}^{\infty}\frac{\prod_{i=1}^b {x_i}^{r_i-1}\,dx_i}{\left(1+\sum_{i=1}^b x_i\right)^{m+R}} \end{aligned} \tag{7}$$

$$\begin{aligned} CD_{\mathbf{a}}^{(c,d-c)}(\mathbf{r},m) &= \frac{\Gamma(m+R)}{\Gamma(m)\prod_{i=1}^b\Gamma(r_i)} \\ &\quad \times\int_0^{a_c}\int_{a_{c+1}}^{\infty}\int_{a_b}^{\infty}\frac{\prod_{i=1}^b {x_i}^{r_i-1}\,dx_i}{\left(1+\sum_{i=1}^b x_i\right)^{m+R}}, \end{aligned} \tag{8}$$

where

$$R \equiv \sum_{i=1}^k r_i \tag{9}$$

$$a_i \equiv \frac{p_i}{1-\sum_{i=1}^k p_i}, \tag{10}$$

and p_i are the cell probabilities. For equal probabilities, $a_i = 1$. The Dirichlet D integral can be expanded as a MULTINOMIAL SERIES as

$$\begin{aligned} D_{\mathbf{a}}^{(b)}(\mathbf{r},m) &= \frac{1}{\left(1+\sum_{i=1}^b\right)^m} \\ &\quad \times\sum_{x_1<r_1}\cdots\sum_{x_b<r_b}\binom{m-1+\sum_{a=1}^b x_i}{m-1,x_1,\ldots,x_b} \\ &\quad \times\prod_{i=1} b\left(\frac{a_i}{1+\sum_{k=1}^b a_k}\right)^{x_i}. \end{aligned} \tag{11}$$

For small b, C and D can be expressed analytically either partially or fully for general arguments and $a_i = 1$.

$$C_1^{(1)}(r_2;r_1) = \frac{\Gamma(r_1+r_2)\,{}_2F_1(r_2,r_1+r_2;1+r_2;-1)}{r_2\Gamma(r_1)\Gamma(r_2)} \tag{12}$$

$$\begin{aligned} C_1^{(2)}(r_2,r_3;r_1) &= \frac{\Gamma(r_1+r_2+r_3)}{r_2\Gamma(r_1)\Gamma(r_2)\Gamma(r_3)} \\ &\quad \times\int_0^1 {}_2F_1\,y^{r_3-1}(1+y)^{-(r_1+r_2+r_3)}\,dy, \end{aligned} \tag{13}$$

where

$${}_2F_1 \equiv {}_2F_1(r_2,r_1+r_2+r_3;1+r_2,-(1+y)^{-1}) \tag{14}$$

is a HYPERGEOMETRIC FUNCTION.

$$D_1^{(1)}(r_2;r_1) = \frac{\Gamma(r_1+r_2)\,{}_2F_1(r_1,r_1+r_2;1+r_1;-1)}{r_1\Gamma(r_1)\Gamma(r_2)} \tag{15}$$

$$\begin{aligned} D_1^{(2)}(r_2,r_3;r_1) &= \frac{\Gamma(r_1+r_2+r_3)}{(r_1+r_3)\Gamma(r_1)\Gamma(r_2)\Gamma(r_3)} \\ &\quad \times\int_1^{\infty} {}_2F_1\,y^{r_3-1}\,dy, \end{aligned} \tag{16}$$

where

$${}_2F_1 \equiv {}_2F_1(r_1+r_3,r_1+r_2+r_3;1+r_1+r_3;-1-y). \tag{17}$$

References

Sobel, M.; Uppuluri, R. R.; and Frankowski, K. *Selected Tables in Mathematical Statistics, Vol. 4: Dirichlet Distribution—Type 1.* Providence, RI: Amer. Math. Soc., 1977.

Sobel, M.; Uppuluri, R. R.; and Frankowski, K. *Selected Tables in Mathematical Statistics, Vol. 9: Dirichlet Integrals of Type 2 and Their Applications.* Providence, RI: Amer. Math. Soc., 1985.

❀ Weisstein, E. W. "Dirichlet Integrals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/DirichletIntegrals.m`.

Dirichlet Kernel

The Dirichlet kernel D_n^M is obtained by integrating the CHARACTER $e^{i\langle\xi,x\rangle}$ over the BALL $|\xi| \leq M$,

$$D_n^M = -\frac{1}{2\pi r}\frac{d}{dr}D_{n-2}^M.$$

The Dirichlet kernel of a DELTA SEQUENCE is given by

$$\delta_n(x) \equiv \frac{1}{2\pi}\frac{\sin[(n+\frac{1}{2})x]}{\sin(\frac{1}{2}x)}.$$

The integral of this kernal is called the DIRICHLET INTEGRAL $D[u]$.

see also DELTA SEQUENCE, DIRICHLET INTEGRALS, DIRICHLET'S LEMMA

Dirichlet L-Series

Series of the form

$$L_k(s,\chi) \equiv \sum_{n=1}^{\infty} \chi_k(n) n^{-s}, \tag{1}$$

where the CHARACTER (NUMBER THEORY) $\chi_k(n)$ is an INTEGER function with period m. These series appear in number theory (they were used, for instance, to prove DIRICHLET'S THEOREM) and can be written as sums of LERCH TRANSCENDENTS with z a POWER of $e^{2\pi i/m}$. The DIRICHLET ETA FUNCTION

$$\eta(s) \equiv \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^s} = (1-2^{1-s})\zeta(s) \tag{2}$$

(for $s \neq 1$) and DIRICHLET BETA FUNCTION

$$L_{-4}(s) = \beta(s) \equiv \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^s} \tag{3}$$

and RIEMANN ZETA FUNCTION

$$L_{+1}(s) = \zeta(s) \tag{4}$$

are Dirichlet series (Borwein and Borwein 1987, p. 289). χ_k is called primitive if the CONDUCTOR $f(\chi) = k$. Otherwise, χ_k is imprimitive. A primitive L-series modulo k is then defined as one for which $\chi_k(n)$ is primitive. All imprimitive L-series can be expressed in terms of primitive L-series.

Let $P = 1$ or $P = \prod_{i=1}^{t} p_i$, where p_i are distinct ODD PRIMES. Then there are three possible types of primitive L-series with REAL COEFFICIENTS. The requirement of REAL COEFFICIENTS restricts the CHARACTER to $\chi_k(n) = \pm 1$ for all k and n. The three type are then

1. If $k = P$ (e.g., $k = 1, 3, 5, \ldots$) or $k = 4P$ (e.g., $k = 4, 12, 20$, dots), there is exactly one primitive L-series.
2. If $k = 8P$ (e.g., $k = 8, 24, \ldots$), there are two primitive L-series.
3. If $k = 2P, Pp_i$, or $2^\alpha P$ where $\alpha > 3$ (e.g., $k = 2, 6, 9, \ldots$), there are no primitive L-series

(Zucker and Robertson 1976). All primitive L-series are algebraically independent and divide into two types according to

$$\chi_k(k-1) = \pm 1. \tag{5}$$

Primitive L-series of these types are denoted $L_\pm$. For a primitive L-series with REAL CHARACTER (NUMBER THEORY), if $k = P$, then

$$L_k = \begin{cases} L_{-k} & \text{if } P \equiv 3 \pmod 4 \\ L_k & \text{if } P \equiv 1 \pmod 4. \end{cases} \tag{6}$$

If $k = 4P$, then

$$L_k = \begin{cases} L_{-k} & \text{if } P \equiv 1 \pmod 4 \\ L_k & \text{if } P \equiv 3 \pmod 4, \end{cases} \tag{7}$$

and if $k = 8P$, then there is a primitive function of each type (Zucker and Robertson 1976).

The first few primitive NEGATIVE L-series are L_{-3}, L_{-4}, L_{-7}, L_{-8}, L_{-11}, L_{-15}, L_{-19}, L_{-20}, L_{-23}, L_{-24}, L_{-31}, L_{-35}, L_{-39}, L_{-40}, L_{-43}, L_{-47}, L_{-51}, L_{-52}, L_{-55}, L_{-56}, L_{-59}, L_{-67}, L_{-68}, L_{-71}, L_{-79}, L_{-83}, L_{-84}, L_{-87}, L_{-88}, L_{-91}, L_{-95}, ... (Sloane's A003657), corresponding to the negated discriminants of imaginary quadratic fields. The first few primitive POSITIVE L-series are L_{+1}, L_{+5}, L_{+8}, L_{+12}, L_{+13}, L_{+17}, L_{+21}, L_{+24}, L_{+28}, L_{+29}, L_{+33}, L_{+37}, L_{+40}, L_{+41}, L_{+44}, L_{+53}, L_{+56}, L_{+57}, L_{+60}, L_{+61}, L_{+65}, L_{+69}, L_{+73}, L_{+76}, L_{+77}, L_{+85}, L_{+88}, L_{+89}, L_{+92}, L_{+93}, L_{+97}, ... (Sloane's A046113).

The KRONECKER SYMBOL is a REAL CHARACTER modulo k, and is in fact essentially the only type of REAL primitive CHARACTER (Ayoub 1963). Therefore,

$$L_{+d}(s) = \sum_{n=1}^{\infty} (d|n) n^{-s} \tag{8}$$

$$L_{-d}(s) = \sum_{n=1}^{\infty} (-d|n) n^{-s}, \tag{9}$$

where $(d|n)$ is the KRONECKER SYMBOL. The functional equations for $L_\pm$ are

$$L_{-k}(s) = 2^s \pi^{s-1} k^{-s+1/2} \Gamma(1-s) \cos(\tfrac{1}{2}s\pi) L_{-k}(1-s) \tag{10}$$

$$L_{+k}(s) = 2^s \pi^{s-1} k^{-s+1/2} \Gamma(1-s) \sin(\tfrac{1}{2}s\pi) L_{+k}(1-s). \tag{11}$$

For m a POSITIVE INTEGER

$$L_{+k}(-2m) = 0 \tag{12}$$

$$L_{-k}(1-2m) = 0 \tag{13}$$

$$L_{+k}(2m) = Rk^{-1/2}\pi^{2m} \tag{14}$$

$$L_{-k}(2m-1) = R'k^{-1/2}\pi^{2m-1} \tag{15}$$

$$L_{+k}(1-2m) = \frac{(-1)^m(2m-1)!R}{(2k)^{2m-1}} \tag{16}$$

$$L_{-k}(-2k) = \frac{(-1)^m R'(2m)!}{(2k)^{2m}}, \tag{17}$$

where R and R' are RATIONAL NUMBERS. $L_{+k}(1)$ can be expressed in terms of transcendentals by

$$L_d(1) = h(d)\kappa(d), \tag{18}$$

where $h(d)$ is the CLASS NUMBER and $\kappa(d)$ is the DIRICHLET STRUCTURE CONSTANT. Some specific values of primitive L-series are

$$\begin{aligned}
L_{-15}(1) &= \frac{2\pi}{\sqrt{15}}\\
L_{-11}(1) &= \frac{\pi}{\sqrt{11}}\\
L_{-8}(1) &= \frac{\pi}{2\sqrt{2}}\\
L_{-7}(1) &= \frac{\pi}{\sqrt{7}}\\
L_{-4}(1) &= \tfrac{1}{4}\pi\\
L_{-3}(1) &= \frac{\pi}{3\sqrt{3}}\\
L_{+5}(1) &= \frac{2}{\sqrt{5}}\ln\left(\frac{1+\sqrt{5}}{2}\right)\\
L_{+8}(1) &=\\
L_{+12}(1) &= \frac{\ln(2+\sqrt{3})}{\sqrt{3}}\\
L_{+13}(1) &= \frac{2}{\sqrt{13}}\ln\left(\frac{3+\sqrt{13}}{2}\right)\\
L_{+17}(1) &= \frac{2}{\sqrt{17}}\ln(4+\sqrt{17})\\
L_{+21}(1) &= \frac{2}{\sqrt{21}}\ln\left(\frac{5+\sqrt{21}}{2}\right)\\
L_{+24}(1) &= \frac{\ln(5+2\sqrt{6})}{\sqrt{6}}.
\end{aligned}$$

No general forms are known for $L_{-k}(2m)$ and $L_{+k}(2m-1)$ in terms of known transcendentals. For example,

$$L_{-4}(2) = \beta(2) \equiv K, \tag{19}$$

where K is defined as CATALAN'S CONSTANT.

see also DIRICHLET BETA FUNCTION, DIRICHLET ETA FUNCTION

References

Ayoub, R. G. *An Introduction to the Analytic Theory of Numbers.* Providence, RI: Amer. Math. Soc., 1963.

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, 1987.

Buell, D. A. "Small Class Numbers and Extreme Values of L-Functions of Quadratic Fields." *Math. Comput.* **139**, 786–796, 1977.

Ireland, K. and Rosen, M. "Dirichlet L-Functions." Ch. 16 in *A Classical Introduction to Modern Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 249–268, 1990.

Sloane, N. J. A. Sequences A046113 and A003657/M2332 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Weisstein, E. W. "Class Numbers." `http://www.astro.virginia.edu/~eww6n/math/notebooks/ClassNumbers.m.`

Zucker, I. J. and Robertson, M. M. "Some Properties of Dirichlet L-Series." *J. Phys. A: Math. Gen.* **9**, 1207–1214, 1976.

Dirichlet Lambda Function

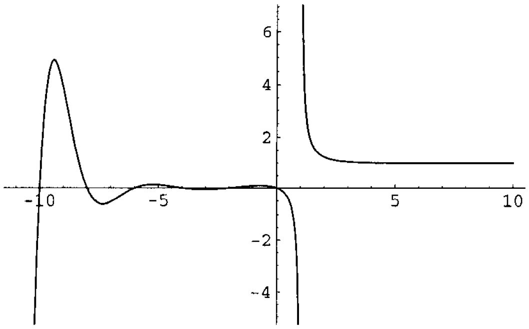

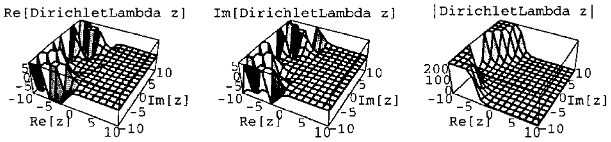

$$\lambda(x) \equiv \sum_{n=0}^{\infty}(2n+1)^{-x} = (1-2^{-x})\zeta(x) \tag{1}$$

for $x = 2, 3, \ldots$, where $\zeta(x)$ is the RIEMANN ZETA FUNCTION. The function is undefined at $x = 1$. It can be computed in closed form where $\zeta(x)$ can, that is for EVEN POSITIVE n. It is related to the RIEMANN ZETA FUNCTION and DIRICHLET ETA FUNCTION by

$$\frac{\zeta(\nu)}{2^\nu} = \frac{\lambda(\nu)}{2^\nu-1} = \frac{\eta(\nu)}{2^\nu-2} \tag{2}$$

and

$$\zeta(\nu)+\eta(\nu) = 2\lambda(\nu) \tag{3}$$

(Spanier and Oldham 1987). Special values of $\lambda(n)$ include

$$\lambda(2) = \frac{\pi^2}{8} \tag{4}$$

$$\lambda(4) = \frac{\pi^4}{96}. \tag{5}$$

see also DIRICHLET BETA FUNCTION, DIRICHLET ETA FUNCTION, RIEMANN ZETA FUNCTION, ZETA FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 807–808, 1972.

Spanier, J. and Oldham, K. B. "The Zeta Numbers and Related Functions." Ch. 3 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 25–33, 1987.

Dirichlet's Lemma

$$\int_0^\pi \frac{\sin[(n+\tfrac{1}{2})x]}{2\sin(\tfrac{1}{2}x)}\,dx = \tfrac{1}{2}\pi,$$

where the KERNEL is the DIRICHLET KERNEL.

References

Cohn, H. *Advanced Number Theory.* New York: Dover, p. 37, 1980.

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1101, 1979.

Dirichlet's Principle

Also known as THOMSON'S PRINCIPLE. There exists a function u that minimizes the functional

$$D[u] = \int_\Omega |\nabla u|^2\,dV$$

(called the DIRICHLET INTEGRAL) for $\Omega \subset \mathbb{R}^2$ or $\mathbb{R}^3$ among all the functions $u \in C^{(1)}(\Omega) \cap C^{(0)}(\bar{\Omega})$ which take on given values f on the boundary $\partial\Omega$ of Ω, and that function u satisfies $\nabla^2 = 0$ in Ω, $u|_{\partial\Omega} = f$, $u \in C^{(2)}(\Omega) \cap C^{(0)}(\bar{\Omega})$. Weierstraß showed that Dirichlet's argument contained a subtle fallacy. As a result, it can be claimed only that there exists a lower bound to which $D[u]$ comes arbitrarily close without being forced to actually reach it. Kneser, however, obtained a valid proof of Dirichlet's principle.

see also DIRICHLET'S BOX PRINCIPLE, DIRICHLET INTEGRALS

Dirichlet Region

see VORONOI POLYGON

Dirichlet Series

A sum $\sum a_n e^{\lambda_n z}$, where a_n and z are COMPLEX and λ_n is REAL and MONOTONIC increasing.

see also DIRICHLET L-SERIES

Dirichlet Structure Constant

$$\kappa(d) = \begin{cases} \frac{2\ln\eta(d)}{\sqrt{d}} & \text{for } d > 0 \\ \frac{2\pi}{w(d)\sqrt{|d|}} & \text{for } d < 0, \end{cases}$$

where $\eta(d)$ is the FUNDAMENTAL UNIT and $w(d)$ is the number of substitutions which leave the binary quadratic form unchanged

$$w(d) = \begin{cases} 6 & \text{for } d = -3 \\ 4 & \text{for } d = -4 \\ 2 & \text{otherwise.} \end{cases}$$

see also CLASS NUMBER, DIRICHLET L-SERIES

References

Weisstein, E. W. "Class Numbers." `http://www.astro.virginia.edu/~eww6n/math/notebooks/ClassNumbers.m`.

Dirichlet Tessellation

see VORONOI DIAGRAM

Dirichlet's Test

Let

$$\left|\sum_{n=1}^{p} a_n\right| < K,$$

where K is independent of p. Then if $f_n \geq f_{n+1} > 0$ and

$$\lim_{n\to\infty} f_n = 0,$$

it follows that

$$\sum_{n=1}^{\infty} a_n f_n$$

CONVERGES.

see also CONVERGENCE TESTS

Dirichlet's Theorem

Given an ARITHMETIC SERIES of terms $an+b$, for $n = 1$, 2, ..., the series contains an infinite number of PRIMES if a and b are RELATIVELY PRIME, i.e., $(a,b) = 1$. Dirichlet proved this theorem using DIRICHLET L-SERIES.

see also PRIME ARITHMETIC PROGRESSION, PRIME PATTERNS CONJECTURE, RELATIVELY PRIME, SIERPIŃSKI'S PRIME SEQUENCE THEOREM

References

Courant, R. and Robbins, H. "Primes in Arithmetical Progressions." §1.2b in Supplement to Ch. 1 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 26–27, 1996.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 22–23, 1993.

Dirty Beam

The FOURIER TRANSFORM of the (u, v) sampling distribution in synthesis imaging

$$b' = \mathcal{F}^{-1}(S), \tag{1}$$

also called the SYNTHESIZED BEAM. It is called a "beam" by way of analogy with the DIRTY MAP

$$\begin{aligned} I' &= \mathcal{F}^{-1}(VS) = \mathcal{F}^{-1}[V] * \mathcal{F}^{-1}[S] \\ &= I * \mathcal{F}^{-1}(S) \equiv I * b', \end{aligned} \tag{2}$$

where $*$ denotes CONVOLUTION. Here, I' is the intensity which would be observed for an extended source by an antenna with response pattern b_1,

$$I' = b_1(\theta'') * I(\theta''). \tag{3}$$

The dirty beam is often a complicated function. In order to avoid introducing any high spatial frequency features when CLEANing, an elliptical Gaussian is usually fit to the dirty beam, producing a CLEAN BEAM which is CONVOLVED with the final iteration.

see also CLEAN ALGORITHM, CLEAN MAP, DIRTY MAP

Dirty Map

From the van Cittert-Zernicke theorem, the relationship between observed visibility function $V(u, v)$ and source brightness $I(\xi, \eta)$ in synthesis imaging is given by

$$\begin{aligned} I(\xi, \eta) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} V(u, v) e^{2\pi i(\xi u + \eta v)}\, du\, dv \\ &= \mathcal{F}^{-1}[V(u, v)]. \end{aligned} \tag{1}$$

But the visibility function is sampled only at discrete points $S(u, v)$ (finite sampling), so only an approximation to I, called the "dirty map" and denoted I', is measured. It is given by

$$\begin{aligned} I'(\xi, \eta) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(u, v) V(u, v) e^{2\pi i(\xi u + \eta v)}\, du\, dv \\ &= \mathcal{F}^{-1}[VS], \end{aligned} \tag{2}$$

where $S(u, v)$ is the sampling function and $V(u, v)$ is the observed visibility function. Let $*$ denote CONVOLUTION and rearrange the CONVOLUTION THEOREM,

$$\mathcal{F}[f * g] = \mathcal{F}[f]\mathcal{F}[g] \tag{3}$$

into the form

$$\mathcal{F}[\mathcal{F}^{-1}[f] * \mathcal{F}^{-1}[g]] = fg, \tag{4}$$

from which it follows that

$$\mathcal{F}^{-1}[f] * \mathcal{F}^{-1}[g] = \mathcal{F}^{-1}[fg]. \tag{5}$$

Now note that

$$I = \mathcal{F}^{-1}[V] \tag{6}$$

is the CLEAN MAP, and define the "DIRTY BEAM" as the inverse FOURIER TRANSFORM of the sampling function,

$$b' \equiv \mathcal{F}^{-1}[S]. \tag{7}$$

The dirty map is then given by

$$I' = \mathcal{F}^{-1}[VS] = \mathcal{F}^{-1}[V] * \mathcal{F}^{-1}[S] = I * b'. \tag{8}$$

In order to deconvolve the desired CLEAN MAP I from the measured dirty map I' and the known DIRTY BEAM b', the CLEAN ALGORITHM is often used.

see also CLEAN ALGORITHM, CLEAN MAP, DIRTY BEAM

Disc

see DISK

Disconnected Form

A FORM which is the sum of two FORMS involving separate sets of variables.

Disconnectivity

Disconnectivities are mathematical entities which stand in the way of a SPACE being contractible (i.e., shrunk to a point, where the shrinking takes place inside the SPACE itself). When dealing with TOPOLOGICAL SPACES, a disconnectivity is interpreted as a "HOLE" in the space. Disconnectivities in SPACE are studied through the EXTENSION PROBLEM or the LIFTING PROBLEM.

see also EXTENSION PROBLEM, HOLE, LIFTING PROBLEM

Discontinuity

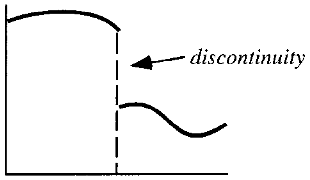

A point at which a mathematical object is DISCONTINUOUS.

Discontinuous

Not CONTINUOUS. A point at which a function is discontinuous is called a DISCONTINUITY, or sometimes a JUMP.

References

Yates, R. C. "Functions with Discontinuous Properties." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 100–107, 1952.

Discordant Permutation

see MARRIED COUPLES PROBLEM

Discrepancy Theorem

Let s_1, s_2, ... be an infinite series of real numbers lying between 0 and 1. Then corresponding to any arbitrarily large K, there exists a positive integer n and two subintervals of equal length such that the number of s_ν with $\nu = 1, 2, \ldots, n$ which lie in one of the subintervals differs from the number of such s_ν that lie in the other subinterval by more than K (van der Corput 1935ab, van Aardenne-Ehrenfest 1945, 1949, Roth 1954).

This statement can be refined as follows. Let N be a large integer and $s_1, s_2, \ldots, s_N$ be a sequence of N real numbers lying between 0 and 1. Then for any integer $1 \leq n \leq N$ and any real number α satisfying $0 < \alpha < 1$, let $D_n(\alpha)$ denote the number of s_ν with $\nu = 1, 2, \ldots, n$ that satisfy $0 \leq s_\nu < \alpha$. Then there exist n and α such that

$$|D_n(\alpha) - n\alpha| > c_1 \frac{\ln\ln N}{\ln\ln\ln N}$$

where c_1 is a positive constant.

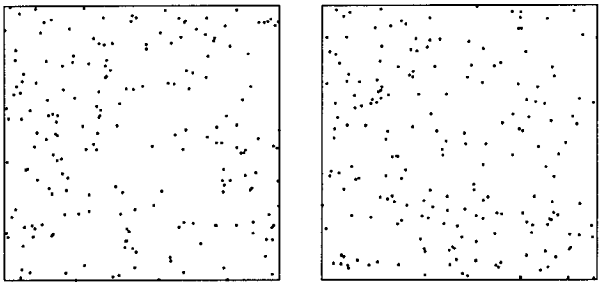

This result can be further strengthened, which is most easily done by reformulating the problem. Let $N > 1$ be an integer and $P_1, P_2, \ldots, P_N$ be N (not necessarily distinct) points in the square $0 \leq x \leq 1$, $0 \leq y \leq 1$. Then

$$\int_0^1 \int_0^1 [S(x,y) - Nxy]^2 \, dx \, dy > c_2 \ln N,$$

where c_2 is a positive constant and $S(u,v)$ is the number of points in the rectangle $0 \leq x < u$, $0 \leq y < v$ (Roth 1954). Therefore,

$$|S(x,y) - Nxy| > c_3 \sqrt{\ln N},$$

and the original result can be stated as the fact that there exist n and α such that

$$|D_n(\alpha) - n\alpha| > c_4 \sqrt{\ln N}\,.$$

The randomly distributed points shown in the above squares have $|S(x,y) - Nxy|^2 = 6.40$ and 9.11, respectively.

Similarly, the discrepancy of a set of N points in a unit d-Hypercube satisfies

$$|S(x,y) - Nxy| > c(\ln N)^{(d-1)/2}$$

(Roth 1954, 1976, 1979, 1980).

see also 18-Point Problem, Cube Point Picking

References

Berlekamp, E. R. and Graham, R. L. "Irregularities in the Distributions of Finite Sequences." *J. Number Th.* **2**, 152–161, 1970.

Roth, K. F. "On Irregularities of Distribution." *Mathematika* **1**, 73–79, 1954.

Roth, K. F. "On Irregularities of Distribution. II." *Comm. Pure Appl. Math.* **29**, 739–744, 1976.

Roth, K. F. "On Irregularities of Distribution. III." *Acta Arith.* **35**, 373–384, 1979.

Roth, K. F. "On Irregularities of Distribution. IV." *Acta Arith.* **37**, 67–75, 1980

van Aardenne-Ehrenfest, T. "Proof of the Impossibility of a Just Distribution of an Infinite Sequence Over an Interval." *Proc. Kon. Ned. Akad. Wetensch.* **48**, 3–8, 1945.

van Aardenne-Ehrenfest, T. *Proc. Kon. Ned. Akad. Wetensch.* **52**, 734–739, 1949.

van der Corput, J. G. *Proc. Kon. Ned. Akad. Wetensch.* **38**, 813–821, 1935a.

van der Corput, J. G. *Proc. Kon. Ned. Akad. Wetensch.* **38**, 1058–1066, 1935b.

Discrete Distribution

A Distribution whose variables can take on only discrete values. Abramowitz and Stegun (1972, p. 929) give a table of the parameters of most common discrete distributions.

see also Bernoulli Distribution, Binomial Distribution, Continuous Distribution, Distribution, Geometric Distribution, Hypergeometric Distribution, Negative Binomial Distribution, Poisson Distribution, Probability, Statistics, Uniform Distribution

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 927 and 929, 1972.

Discrete Fourier Transform

The Fourier Transform is defined as

$$f(\nu) = \mathcal{F}[f(t)] = \int_{-\infty}^{\infty} f(t) e^{-2\pi i \nu t} \, dt. \qquad (1)$$

Now consider generalization to the case of a discrete function, $f(t) \to f(t_k)$ by letting $f_k \equiv f(t_k)$, where $t_k \equiv k\Delta$, with $k = 0, \ldots, N-1$. Choose the frequency step such that

$$\nu_n = \frac{n}{N\Delta}, \qquad (2)$$

with $n = -N/2, \ldots, 0, \ldots, N/2$. There are $N+1$ values of n, so there is one relationship between the frequency components. Writing this out as per Press *et al.* (1989)

$$\mathcal{F}[f(t)] = \sum_{k=0}^{N-1} f_k e^{-2\pi i (n/N\Delta) k\Delta} \Delta = \Delta \sum_{k=0}^{N-1} f_k e^{-2\pi i n k/N}, \qquad (3)$$

and

$$F_n \equiv \sum_{k=0}^{N-1} f_k e^{-2\pi ink/N}. \tag{4}$$

The inverse transform is

$$f_k = \frac{1}{N} \sum_{n=0}^{N-1} F_n e^{2\pi ink/N}. \tag{5}$$

Note that $F_{-n} = F_{N-n}$, $n = 1, 2, \ldots$, so an alternate formulation is

$$\nu_n = \frac{n}{N\Delta}, \tag{6}$$

where the NEGATIVE frequencies $-\nu_c < \nu < 0$ have $N/2+1 \leq n \leq N-1$, POSITIVE frequencies $0 < \nu < \nu_c$ have $1 \leq n \leq N/2 - 1$, with zero frequency $n = 0$. $n = N/2$ corresponds to both $\nu = \nu_c$ and $\nu = -\nu_c$. The discrete Fourier transform can be computed using a FAST FOURIER TRANSFORM.

The discrete Fourier transform is a special case of the z-TRANSFORM.

see also FAST FOURIER TRANSFORM, FOURIER TRANSFORM, HARTLEY TRANSFORM, WINOGRAD TRANSFORM, z-TRANSFORM

References

Arfken, G. "Discrete Orthogonality—Discrete Fourier Transform." §14.6 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 787–792, 1985.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Fourier Transform of Discretely Sampled Data." §12.1 in *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge, England: Cambridge University Press, pp. 494–498, 1989.

Discrete Mathematics

The branch of mathematics dealing with objects which can assume only certain "discrete" values. Discrete objects can be characterized by INTEGERS (or RATIONAL NUMBERS), whereas continuous objects require REAL NUMBERS. The study of how discrete objects combine with one another and the probabilities of various outcomes is known as COMBINATORICS.

see also COMBINATORICS

References

Balakrishnan, V. K. *Introductory Discrete Mathematics.* New York: Dover, 1997.

Bobrow, L. S. and Arbib, M. A. *Discrete Mathematics: Applied Algebra for Computer and Information Science.* Philadelphia, PA: Saunders, 1974.

Dossey, J. A.; Otto, A. D.; Spence, L.; and Eynden, C. V. *Discrete Mathematics, 3rd ed.* Reading, MA: Addison-Wesley, 1997.

Skiena, S. S. *Implementing Discrete Mathematics.* Reading, MA: Addison-Wesley, 1990.

Discrete Set

A finite SET or an infinitely COUNTABLE SET of elements.

Discrete Uniform Distribution

see EQUALLY LIKELY OUTCOMES DISTRIBUTION

Discriminant

A discriminant is a quantity (usually invariant under certain classes of transformations) which characterizes certain properties of a quantity's ROOTS. The concept of the discriminant is used for BINARY QUADRATIC FORMS, ELLIPTIC CURVES, METRICS, MODULES, POLYNOMIALS, QUADRATIC CURVES, QUADRATIC FIELDS, QUADRATIC FORMS, and in the SECOND DERIVATIVE TEST.

Discriminant (Binary Quadratic Form)

The discriminant of a BINARY QUADRATIC FORM

$$au^2 + buv + cv^2$$

is defined by

$$d \equiv b^2 - 4ac.$$

It is equal to four times the corresponding DETERMINANT.

see also CLASS NUMBER

Discriminant (Elliptic Curve)

An ELLIPTIC CURVE is of the form

$$y^2 = x^3 + a_2x^2 + a_1x + a_0.$$

Let the ROOTS of y^2 be r_1, r_2, and r_3. The discriminant is then defined as

$$\Delta = k(r_1 - r_2)^2(r_1 - r_3)^2(r_2 - r_3)^2.$$

see also FREY CURVE, MINIMAL DISCRIMINANT

Discriminant (Metric)

Given a METRIC $g_{\alpha\beta}$, the discriminant is defined by

$$g \equiv \det(g_{\alpha\beta}) = \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} = g_{11}g_{22} - (g_{12})^2. \tag{1}$$

Let g be the discriminant and $\bar{g}$ the transformed discriminant, then

$$\bar{g} = D^2 g \tag{2}$$

$$g = \bar{D}^2 \bar{g}, \tag{3}$$

where

$$D \equiv \frac{\partial(u^1, u^2)}{\partial(\bar{u}^1, \bar{u}^2)} = \begin{vmatrix} \frac{\partial u^1}{\partial \bar{u}^1} & \frac{\partial u^1}{\partial \bar{u}^2} \\ \frac{\partial u^2}{\partial \bar{u}^1} & \frac{\partial u^2}{\partial \bar{u}^2} \end{vmatrix} \tag{4}$$

$$\bar{D} \equiv \frac{\partial(\bar{u}^1, \bar{u}^2)}{\partial(u^1, u^2)} = \begin{vmatrix} \frac{\partial \bar{u}^1}{\partial u^1} & \frac{\partial \bar{u}^1}{\partial u^2} \\ \frac{\partial \bar{u}^2}{\partial u^1} & \frac{\partial \bar{u}^2}{\partial u^2} \end{vmatrix}. \tag{5}$$

Discriminant (Module)

Let a MODULE M in an INTEGRAL DOMAIN D_1 for $R(\sqrt{D})$ be expressed using a two-element basis as

$$M = [\xi_1, \xi_2],$$

where ξ_1 and ξ_2 are in D_1. Then the DIFFERENT of the MODULE is defined as

$$\Delta = \Delta(M) = \begin{vmatrix} \xi_1 & \xi_2 \\ \xi_1' & \xi_2' \end{vmatrix} = \xi_1\xi_2' - \xi_1'\xi_2$$

and the discriminant is defined as the square of the DIFFERENT (Cohn 1980).

For IMAGINARY QUADRATIC FIELDS $\mathbb{Q}(\sqrt{n})$ (with $n < 0$), the discriminants are given in the following table.

-1	-2^2	-33	$-2^2 \cdot 3 \cdot 11$	-67	-67
-2	-2^3	-34	$-2^3 \cdot 17$	-69	$-2^2 \cdot 3 \cdot 23$
-3	-3	-35	$-5 \cdot 7$	-70	$-2^3 \cdot 5 \cdot 7$
-5	$-2^2 \cdot 5$	-37	$-2^2 \cdot 37$	-71	-71
-6	$-2^3 \cdot 3$	-39	$-3 \cdot 13$	-73	$-2^2 \cdot 73$
-7	-7	-41	$-2^2 \cdot 41$	-74	$-2^3 \cdot 37$
-10	$-2^3 \cdot 5$	-42	$-2^3 \cdot 3 \cdot 7$	-77	$-2^2 \cdot 7 \cdot 11$
-11	-11	-43	-43	-78	$-2^3 \cdot 3 \cdot 13$
-13	$-2^2 \cdot 13$	-46	$-2^3 \cdot 23$	-79	-79
-14	$-2^3 \cdot 7$	-47	-47	-82	$-2^3 \cdot 41$
-15	$-3 \cdot 5$	-51	$-3 \cdot 17$	-83	-83
-17	$-2^2 \cdot 17$	-53	$-2^2 \cdot 53$	-85	$-2^2 \cdot 5 \cdot 17$
-19	-19	-55	$-5 \cdot 11$	-86	$-2^3 \cdot 43$
-21	$-2^2 \cdot 3 \cdot 7$	-57	$-2^2 \cdot 3 \cdot 19$	-87	$-3 \cdot 29$
-22	$-2^3 \cdot 11$	-58	$-2^3 \cdot 29$	-89	$-2^2 \cdot 89$
-23	-23	-59	-59	-91	$-7 \cdot 13$
-26	$-2^3 \cdot 13$	-61	$-2^2 \cdot 61$	-93	$-2^2 \cdot 3 \cdot 31$
-29	$-2^2 \cdot 29$	-62	$-2^3 \cdot 31$	-94	$-2^3 \cdot 47$
-30	$-2^3 \cdot 3 \cdot 5$	-65	$-2^2 \cdot 5 \cdot 13$	-95	$-5 \cdot 19$
-31	-31	-66	$-2^3 \cdot 3 \cdot 11$	-97	$-2^2 \cdot 97$

The discriminants of REAL QUADRATIC FIELDS $\mathbb{Q}(\sqrt{n})$ ($n > 0$) are given in the following table.

2	2^3	34	$2^3 \cdot 17$	67	$67 \cdot 2^2$
3	$3 \cdot 2^2$	35	$7 \cdot 2^2 \cdot 5$	69	$3 \cdot 23$
5	5	37	37	70	$7 \cdot 2^3 \cdot 5$
6	$3 \cdot 2^3$	38	$19 \cdot 2^3$	71	$71 \cdot 2^2$
7	$7 \cdot 2^2$	39	$3 \cdot 2^2 \cdot 13$	73	73
10	$2^3 \cdot 5$	41	41	74	$2^3 \cdot 37$
11	$11 \cdot 2^2$	42	$3 \cdot 2^3 \cdot 7$	77	$7 \cdot 11$
13	13	43	$43 \cdot 2^2$	78	$3 \cdot 2^3 \cdot 13$
14	$7 \cdot 2^3$	46	$23 \cdot 2^3$	79	$79 \cdot 2^2$
15	$3 \cdot 2^2 \cdot 5$	47	$47 \cdot 2^2$	82	$2^3 \cdot 41$
17	17	51	$3 \cdot 2^2 \cdot 17$	83	$83 \cdot 2^2$
19	$19 \cdot 2^2$	53	53	85	$5 \cdot 17$
21	$3 \cdot 7$	55	$11 \cdot 2^2 \cdot 5$	86	$43 \cdot 2^3$
22	$11 \cdot 2^3$	57	$3 \cdot 19$	87	$3 \cdot 2^2 \cdot 13$
23	$23 \cdot 2^2$	58	$2^3 \cdot 29$	89	89
26	$2^3 \cdot 13$	59	$59 \cdot 2^2$	91	$7 \cdot 2^2 \cdot 13$
29	29	61	61	93	$3 \cdot 31$
30	$3 \cdot 2^3 \cdot 5$	62	$31 \cdot 2^3$	94	$47 \cdot 2^3$
31	$31 \cdot 2^2$	65	$5 \cdot 13$	95	$19 \cdot 2^2 \cdot 5$
33	$3 \cdot 11$	66	$3 \cdot 2^3 \cdot 11$	97	97

see also DIFFERENT, FUNDAMENTAL DISCRIMINANT, MODULE

References

Cohn, H. *Advanced Number Theory.* New York: Dover, pp. 72–73 and 261–274, 1980.

Discriminant (Polynomial)

The PRODUCT of the SQUARES of the differences of the POLYNOMIAL ROOTS x_i. For a POLYNOMIAL of degree n,

$$D_n \equiv \prod_{\substack{i,j \\ i<j}}^{n} (x_i - x_j)^2. \tag{1}$$

The discriminant of the QUADRATIC EQUATION

$$ax^2 + bx + c = 0 \tag{2}$$

is usually taken as

$$D = b^2 - 4ac. \tag{3}$$

However, using the general definition of the POLYNOMIAL DISCRIMINANT gives

$$D \equiv \prod_{i<j} (z_i - z_j)^2 = \frac{b^2 - 4ac}{a^2}, \tag{4}$$

where z_i are the ROOTS.

The discriminant of the CUBIC EQUATION

$$z^3 + a_2 z^2 + a_1 z + a_0 = 0 \tag{5}$$

is commonly defined as

$$D \equiv Q^3 + R^2, \tag{6}$$

where

$$Q \equiv \frac{3a_1 - a_2{}^2}{9} \tag{7}$$

$$R \equiv \frac{9a_2 a_1 - 27a_0 - 2a_2{}^3}{54}. \tag{8}$$

However, using the general definition of the polynomial discriminant for the standard form CUBIC EQUATION

$$z^3 + pz = q \tag{9}$$

gives

$$D \equiv \prod_{i<j} (z_i - z_j)^2 = P^2 = -4p^3 - 27q^2, \tag{10}$$

where z_i are the ROOTS and

$$P = (z_1 - z_2)(z_2 - z_3)(z_1 - z_3). \tag{11}$$

The discriminant of a QUARTIC EQUATION

$$x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = 0 \tag{12}$$

is

$$\begin{aligned}&-27a_1^4 + 18a_3a_2a_1^3 - 4a_3^3a_1^3 - 4a_2^3a_1^2 + a_3^2a_2^2a_1^2\\ &+a_0(144a_2a_1^2 - 6a_3^2a_1^2 - 80a_3a_2^2a_1 + 18a_3^3a_2a_1 + 16a_2^4\\ &-4a_3^2a_2^3) + a_0^2(-192a_3a_1 - 128a_2^2 + 144a_3^2a_2 - 27a_3^4)\\ &-256a_0^3\end{aligned} \tag{13}$$

(Beeler *et al.* 1972, Item 4).

see also RESULTANT

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Discriminant (Quadratic Curve)

Given a general QUADRATIC CURVE

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0, \tag{1}$$

the quantity X is known as the discriminant, where

$$X \equiv B^2 - 4AC, \tag{2}$$

and is invariant under ROTATION. Using the COEFFICIENTS from QUADRATIC EQUATIONS for a rotation by an angle θ,

$$\begin{aligned}A' &= \tfrac{1}{2}A[1+\cos(2\theta)] + \tfrac{1}{2}B\sin(2\theta) + \tfrac{1}{2}C[1-\cos(2\theta)]\\ &= \frac{A+C}{2} + \frac{B}{2}\sin(2\theta) + \frac{A-C}{2}\cos(2\theta) &(3)\\ B' &= G\cos\left(2\theta+\delta-\frac{\pi}{2}\right) = G\sin(2\theta+\delta) &(4)\\ C' &= \tfrac{1}{2}A[1-\cos(2\theta)] - \tfrac{1}{2}B\sin(2\theta) + \tfrac{1}{2})C[1+\cos(2\theta)]\\ &= \frac{A+C}{2} - \frac{B}{2}\sin(2\theta) + \frac{C-A}{2}\cos(2\theta). &(5)\end{aligned}$$

Now let

$$G \equiv \sqrt{B^2+(A-C)^2} \tag{6}$$

$$\delta \equiv \tan^{-1}\left(\frac{B}{C-A}\right) \tag{7}$$

$$\delta_2 \equiv \tan^{-1}\left(\frac{A-C}{B}\right) = -\cot^{-1}\left(\frac{B}{C-A}\right), \tag{8}$$

and use

$$\cot^{-1}(x) = \tfrac{1}{2}\pi - \tan^{-1}(x) \tag{9}$$

$$\delta_2 = \delta - \tfrac{1}{2}\pi \tag{10}$$

to rewrite the primed variables

$$A' = \frac{A+C}{2} + \tfrac{1}{2}G\cos(2\theta+\delta) \tag{11}$$

$$B' = B\cos(2\theta) + (C-A)\sin(2\theta) = G\cos(2\theta+\delta_2) \tag{12}$$

$$C' = \frac{A+C}{2} - \tfrac{1}{2}G\cos(2\theta+\delta). \tag{13}$$

From (11) and (13), it follows that

$$4A'C' = (A+C)^2 - G^2\cos(2\theta+\delta). \tag{14}$$

Combining with (12) yields, for an arbitrary θ

$$\begin{aligned}X &\equiv B'^2 - 4A'C'\\ &= G^2\sin^2(2\theta+\delta) + G^2\cos^2(2\theta+\delta) - (A+C)^2\\ &= G^2 - (A+C)^2 = B^2 + (A-C)^2 - (A+C)^2\\ &= B^2 - 4AC,\end{aligned} \tag{15}$$

which is therefore invariant under rotation. This invariant therefore provides a useful shortcut to determining the shape represented by a QUADRATIC CURVE. Choosing θ to make $B' = 0$ (see QUADRATIC EQUATION), the curve takes on the form

$$A'x^2 + C'y^2 + D'x + E'y + F = 0. \tag{16}$$

COMPLETING THE SQUARE and defining new variables gives

$$A'x'^2 + C'y'^2 = H. \tag{17}$$

Without loss of generality, take the sign of H to be positive. The discriminant is

$$X = B'^2 - 4A'C' = -4A'C'. \tag{18}$$

Now, if $-4A'C' < 0$, then A' and C' both have the same sign, and the equation has the general form of an ELLIPSE (if A' and B' are positive). If $-4A'C' > 0$, then A' and C' have opposite signs, and the equation has the general form of a HYPERBOLA. If $-4A'C' = 0$, then either A' or C' is zero, and the equation has the general form of a PARABOLA (if the NONZERO A' or C' is positive). Since the discriminant is invariant, these conclusions will also hold for an arbitrary choice of θ, so they also hold when $-4A'C'$ is replaced by the original $B^2 - 4AC$. The general result is

1. If $B^2-4AC < 0$, the equation represents an ELLIPSE, a CIRCLE (degenerate ELLIPSE), a POINT (degenerate CIRCLE), or has no graph.
2. If $B^2 - 4AC > 0$, the equation represents a HYPERBOLA or pair of intersecting lines (degenerate HYPERBOLA).
3. If $B^2 - 4AC = 0$, the equation represents a PARABOLA, a LINE (degenerate PARABOLA), a pair of PARALLEL lines (degenerate PARABOLA), or has no graph.

Discriminant (Quadratic Form)

see DISCRIMINANT (BINARY QUADRATIC FORM)

Discriminant (Second Derivative Test)

$$D \equiv f_{xx}f_{yy} - f_{xy}f_{yx} = f_{xx}f_{yy} - {f_{xy}}^2,$$

where f_{ij} are PARTIAL DERIVATIVES.

see also SECOND DERIVATIVE TEST

Disdyakis Dodecahedron

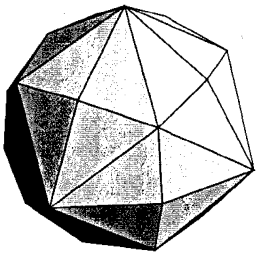

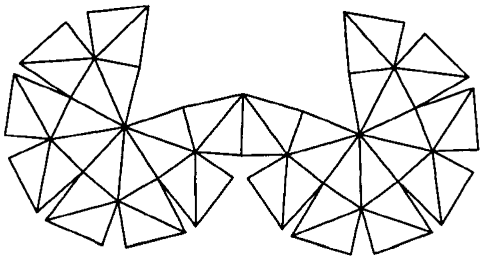

The DUAL POLYHEDRON of the ARCHIMEDEAN GREAT RHOMBICUBOCTAHEDRON, also called the HEXAKIS OCTAHEDRON.

see also GREAT DISDYAKIS DODECAHEDRON

Disdyakis Triacontahedron

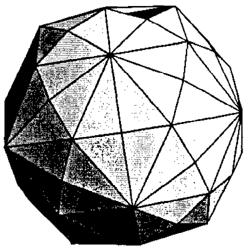

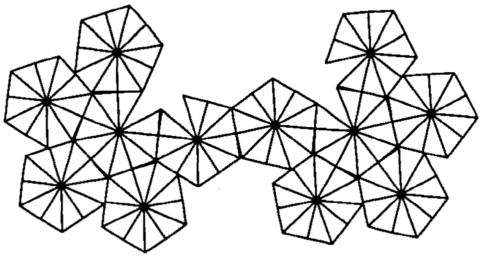

The DUAL POLYHEDRON of the ARCHIMEDEAN GREAT RHOMBICOSIDODECAHEDRON. It is also called the HEXAKIS ICOSAHEDRON.

Disjoint

see MUTUALLY EXCLUSIVE

Disjunction

A product of ORs, denoted

$$\bigvee_{k+1}^{n} A_k.$$

see also CONJUNCTION, OR

Disjunctive Game

see NIM-HEAP

Disk

An n-D disk (or DISC) of RADIUS r is the collection of points of distance $\leq r$ (CLOSED DISK) or $< r$ (OPEN DISK) from a fixed point in EUCLIDEAN n-space. A disk is the SHADOW of a BALL on a PLANE PERPENDICULAR to the BALL-RADIANT POINT line.

The n-disk for $n \geq 3$ is called a BALL, and the boundary of the n-disk is a $(n-1)$-HYPERSPHERE. The standard n-disk, denoted $\mathbb{D}^n$ (or $\mathbb{B}^n$), has its center at the ORIGIN and has RADIUS $r = 1$.

see also BALL, CLOSED DISK, DISK COVERING PROBLEM, FIVE DISKS PROBLEM, HYPERSPHERE, MERGELYAN-WESLER THEOREM, OPEN DISK, POLYDISK, SPHERE, UNIT DISK

Disk Covering Problem

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Given a UNIT DISK, find the smallest RADIUS $r(n)$ required for n equal disks to completely cover the UNIT DISK. For a symmetrical arrangement with $n = 5$ (the FIVE DISKS PROBLEM), $r(5) = \phi - 1 = 1/\phi = 0.6180340\ldots$, where ϕ is the GOLDEN RATIO. However, the radius can be reduced in the general disk covering problem where symmetry is not required. The first few such values are

$$\begin{aligned} r(1) &= 1 \\ r(2) &= 1 \\ r(3) &= \tfrac{1}{2}\sqrt{3} \\ r(4) &= \tfrac{1}{2}\sqrt{2} \\ r(5) &= 0.609382864\ldots \\ r(6) &= 0.555 \\ r(7) &= \tfrac{1}{2} \\ r(8) &= 0.437 \\ r(9) &= 0.422 \\ r(10) &= 0.398. \end{aligned}$$

Here, values for $n = 6, 8, 9, 10$ were obtained using computer experimentation by Zahn (1962). The value $r(5)$ is equal to $\cos(\theta+\phi/2)$, where θ and ϕ are solutions to

$$2\sin\theta - \sin(\theta + \tfrac{1}{2}\phi + \psi) - \sin(\psi - \theta - \tfrac{1}{2}\phi) = 0 \qquad (1)$$

$$2\sin\phi - \sin(\theta + \tfrac{1}{2}\phi + \chi) - \sin(\chi - \theta - \tfrac{1}{2}\phi) = 0 \qquad (2)$$

$$2\sin\theta + \sin(\chi + \theta) - \sin(\chi - \theta) - \sin(\psi + \phi) - \sin(\psi - \phi) - 2\sin(\psi - 2\theta) = 0 \qquad (3)$$

$$\cos(2\psi - \chi + \phi) - \cos(2\psi + \chi - \phi) - 2\cos\chi + \cos(2\psi + \chi - 2\theta) + \cos(2\psi - \chi - 2\theta) = 0 \qquad (4)$$

(Neville 1915). It is also given by $1/x$, where x is the largest real root of

$$a(y)x^6 - b(y)x^5 + c(y)x^4 - d(y)x^3 + e(y)x^2 - f(y)x + g(y) = 0 \qquad (5)$$

maximized over all y, subject to the constraints

$$\sqrt{2} < x < 2y + 1 \qquad (6)$$

$$-1 < y < 1, \qquad (7)$$

and with

$$a(y) = 80y^2 + 64y \qquad (8)$$

$$b(y) = 416y^3 + 384y^2 + 64y \qquad (9)$$

$$c(y) = 848y^4 + 928y^3 + 352y^2 + 32y \qquad (10)$$

$$d(y) = 768y^5 + 992y^4 + 736y^3 + 288y^2 + 96y$$

$$e(y) = 256y^6 + 384y^5 + 592y^4 + 480y^3 + 336y^2 + 96y + 16 \qquad (11)$$

$$f(y) = 128y^5 + 192y^4 + 256y^3 + 160y^2 + 96y + 32 \qquad (12)$$

$$g(y) = 64y^2 + 64y + 16 \qquad (13)$$

(Bezdek 1983, 1984).

Letting $N(\epsilon)$ be the smallest number of DISKS of RADIUS ϵ needed to cover a disk D, the limit of the ratio of the AREA of D to the AREA of the disks is given by

$$\lim_{\epsilon\to 0^+} \frac{1}{\epsilon^2 N(\epsilon)} = \frac{3\sqrt{3}}{2\pi} \qquad (14)$$

(Kershner 1939, Verblunsky 1949).

see also FIVE DISKS PROBLEM

References

Ball, W. W. R. and Coxeter, H. S. M. "The Five-Disc Problem." In *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 97–99, 1987.

Bezdek, K. "Über einige Kreisüberdeckungen." *Beiträge Algebra Geom.* **14**, 7–13, 1983.

Bezdek, K. "Über einige optimale Konfigurationen von Kreisen." *Ann. Univ. Sci. Budapest Eötvös Sect. Math.* **27**, 141–151, 1984.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/circle/circle.html`.

Kershner, R. "The Number of Circles Covering a Set." *Amer. J. Math.* **61**, 665–671, 1939.

Neville, E. H. "On the Solution of Numerical Functional Equations, Illustrated by an Account of a Popular Puzzle and of its Solution." *Proc. London Math. Soc.* **14**, 308–326, 1915.

Verblunsky, S. "On the Least Number of Unit Circles which Can Cover a Square." *J. London Math. Soc.* **24**, 164–170, 1949.

Zahn, C. T. "Black Box Maximization of Circular Coverage." *J. Res. Nat. Bur. Stand. B* **66**, 181–216, 1962.

Disk Lattice Points

see GAUSS'S CIRCLE PROBLEM

Dispersion Numbers

see MAGIC GEOMETRIC CONSTANTS

Dispersion Relation

Any pair of equations giving the REAL PART of a function as an integral of its IMAGINARY PART and the IMAGINARY PART as an integral of its REAL PART. Dispersion relationships imply causality in physics. Let

$$f(x_0) \equiv u(x_0) + iv(x_0), \qquad (1)$$

then

$$u(x_0) = \frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{v(x)\,dx}{x - x_0} \qquad (2)$$

$$v(x_0) = -\frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{u(x)\,dx}{x - x_0}, \qquad (3)$$

where PV denotes the CAUCHY PRINCIPAL VALUE and $u(x_0)$ and $v(x_0)$ are HILBERT TRANSFORMS of each other. If the COMPLEX function is symmetric such that $f(-x) = f^*(x)$, then

$$u(x_0) = \frac{2}{\pi} PV \int_0^{\infty} \frac{xv(x)\,dx}{x^2 - {x_0}^2} \qquad (4)$$

$$v(x_0) = -\frac{2}{\pi} PV \int_0^{\infty} \frac{xu(x)\,dx}{x^2 - {x_0}^2}. \qquad (5)$$

Dispersion (Sequence)

An array $B = b_{ij}$, $i, j \geq 1$ of POSITIVE INTEGERS is called a dispersion if

1. The first column of B is a strictly increasing sequence, and there exists a strictly increasing sequence $\{s_k\}$ such that
2. $b_{12} = s_1 \geq 2$,
3. The complement of the SET $\{b_{i1} : i \geq 1\}$ is the SET $\{s_k\}$,
4. $b_{ij} = s_{b_{i,j-1}}$ for all $j \geq 3$ for $i = 1$ and for all $g \geq 2$ for all $i \geq 2$.

If an array $B = b_{ij}$ is a dispersion, then it is an INTERSPERSION.

see also INTERSPERSION

References

Kimberling, C. "Interspersions and Dispersions." *Proc. Amer. Math. Soc.* **117**, 313–321, 1993.

Dispersion (Statistics)

$$(\Delta u)^2{}_i \equiv (u_i - \bar{u})^2.$$

see also ABSOLUTE DEVIATION, SIGNED DEVIATION, VARIANCE

Disphenocingulum

see JOHNSON SOLID

Disphenoid

A TETRAHEDRON with identical ISOSCELES or SCALENE faces.

Dissection

Any two rectilinear figures with equal AREA can be dissected into a finite number of pieces to form each other. This is the WALLACE-BOLYAI-GERWEIN THEOREM. For minimal dissections of a TRIANGLE, PENTAGON, and OCTAGON into a SQUARE, see Stewart (1987, pp. 169–170) and Ball and Coxeter (1987, pp. 89–91). The TRIANGLE to SQUARE dissection (HABERDASHER'S PROBLEM) is particularly interesting because it can be built from hinged pieces which can be folded and unfolded to yield the two shapes (Gardner 1961; Stewart 1987, p. 169; Pappas 1989).

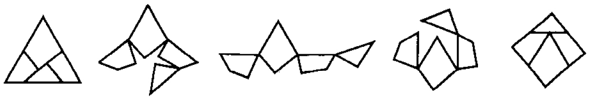

Laczkovich (1988) proved that the CIRCLE can be squared in a finite number of dissections ($\sim 10^{50}$). Furthermore, any shape whose boundary is composed of smoothly curving pieces can be dissected into a SQUARE.

The situation becomes considerably more difficult moving from 2-D to 3-D. In general, a POLYHEDRON cannot be dissected into other POLYHEDRA of a specified type. A CUBE *can* be dissected into n^3 CUBES, where n is any INTEGER. In 1900, Dehn proved that not every PRISM cannot be dissected into a TETRAHEDRON (Lenhard 1962, Ball and Coxeter 1987) The third of HILBERT'S PROBLEMS asks for the determination of two TETRAHEDRA which cannot be decomposed into congruent TETRAHEDRA directly or by adjoining congruent TETRAHEDRA. Max Dehn showed this could not be done in 1902, and W. F. Kagon obtained the same result independently in 1903. A quantity growing out of Dehn's work which can be used to analyze the possibility of performing a given solid dissection is the DEHN INVARIANT.

The table below is an updated version of the one given in Gardner (1991, p. 50). Many of the improvements are due to G. Theobald (Frederickson 1997). The minimum number of pieces known to dissect a regular n-gon (where n is a number in the first column) into a k-gon (where k is a number is the bottom row) is read off by the intersection of the corresponding row and column. In the table, $\{n\}$ denotes a regular n-gon, GR a GOLDEN RECTANGLE, GC a GREEK CROSS, LC a LATIN CROSS, MC a MALTESE CROSS, SW a SWASTIKA, $\{5/2\}$ a five-point star (solid PENTAGRAM), $\{6/2\}$ a six-point star (i.e., HEXAGRAM or solid STAR OF DAVID), and $\{8/3\}$ the solid OCTAGRAM.

	{3}	{4}	{5}	{6}	{7}	{8}	{9}	{10}	{12}
{4}	4								
{5}	6	6							
{6}	5	5	7						
{7}	8	7	9	8					
{8}	7	5	9	8	11				
{9}	8	9	12	11	14	13			
{10}	7	7	10	9	11	10	13		
{12}	8	6	10	6	11	10	14	12	
GR	4	3	6	5	7	6	9	6	7
GC	5	4	7	7	9	9	12	10	6
LC	5	5	8	6	8	8	11	10	7
MC		7		14					
SW		6		12					
{5/2}	7	7	9	9	11	10	14	6	12
{6/2}	5	5	8	6	9	8	11	9	9
{8/3}	8	8	9	9	12	6	13	12	12

	GR	GC	LC	MC	SW	{5/2}	{6/2}
GC	5						
LC	5	7					
MC		8					
SW		8	9				
{5/2}	7	12	10	10			
{6/2}	5	8	8			11	
{8/3}	7	10	11			13	10

The best-known dissections of one regular convex n-gon into another are shown for n = 3, 4, 5, 6, 7, 8, 9, 10, and 12 in the following illustrations due to Theobald.

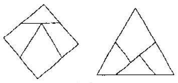
4-3

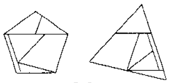
5-3

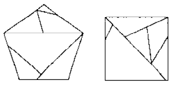
5-4

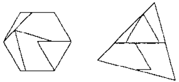
6-3

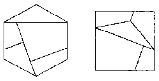
6-4

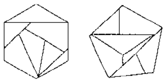
6-5

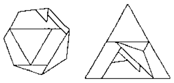
7-3

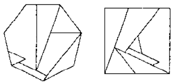
7-4

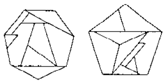
7-5

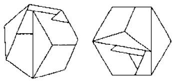
7-6

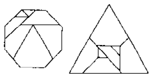
8-3

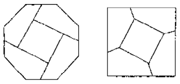
8-4

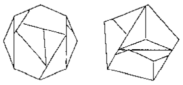
8-5

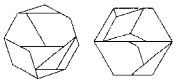
8-6

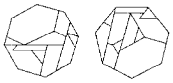
8-7

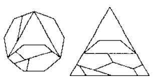
9-3

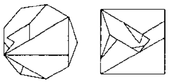
9-4

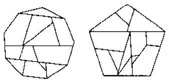
9-5

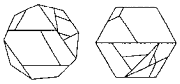
9-6

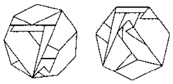
9-7

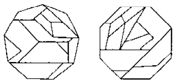
9-8

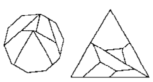
10-3

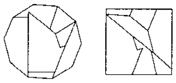
10-4

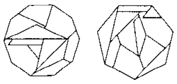
10-7

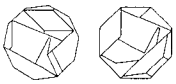
10-8

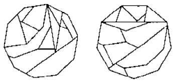
10-9

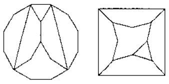
12-4

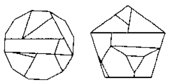
12-5

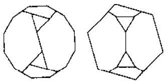
12-6

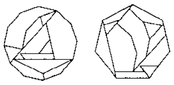
12-7

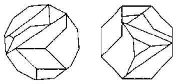
12-8

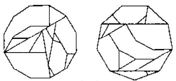
12-9

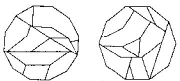
12-10

The best-known dissections of regular concave polygons are illustrated below for {5/2}, {6/2}, and {8/3} (Theobald).

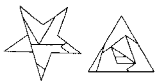
${}^{5}/_{2}$-3

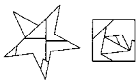
${}^{5}/_{2}$-4

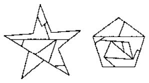
${}^{5}/_{2}$-5

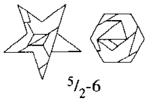
${}^{5}/_{2}$-6

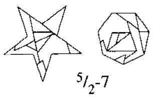
${}^{5}/_{2}$-7

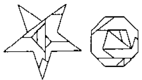
${}^{5}/_{2}$-8

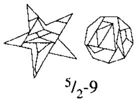
${}^{5}/_{2}$-9

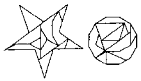
${}^{5}/_{2}$-12

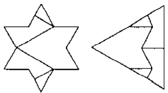
${}^{6}/_{2}$-3

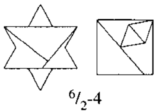
${}^{6}/_{2}$-4

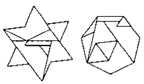
${}^{6}/_{2}$-7

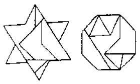
${}^{6}/_{2}$-8

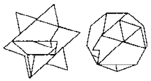
${}^{6}/_{2}$-9

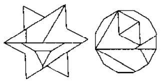
${}^{6}/_{2}$-12

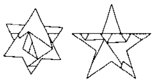
${}^{6}/_{2}$-${}^{5}/_{2}$

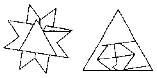
${}^{8}/_{3}$-3

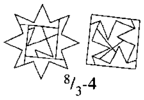
${}^{8}/_{3}$-4

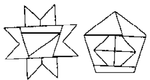
${}^{8}/_{3}$-5

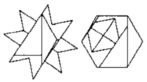
${}^{8}/_{3}$-6

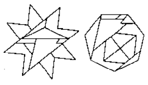
${}^{8}/_{3}$-7

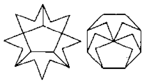
${}^{8}/_{3}$-8

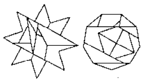
${}^{8}/_{3}$-9

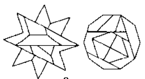
${}^{8}/_{3}$-10

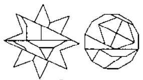
${}^{8}/_{3}$-12

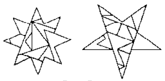
${}^{8}/_{3}$-${}^{5}/_{2}$

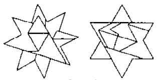
${}^{8}/_{3}$-${}^{6}/_{2}$

The best-known dissections of various crosses are illustrated below (Theobald).

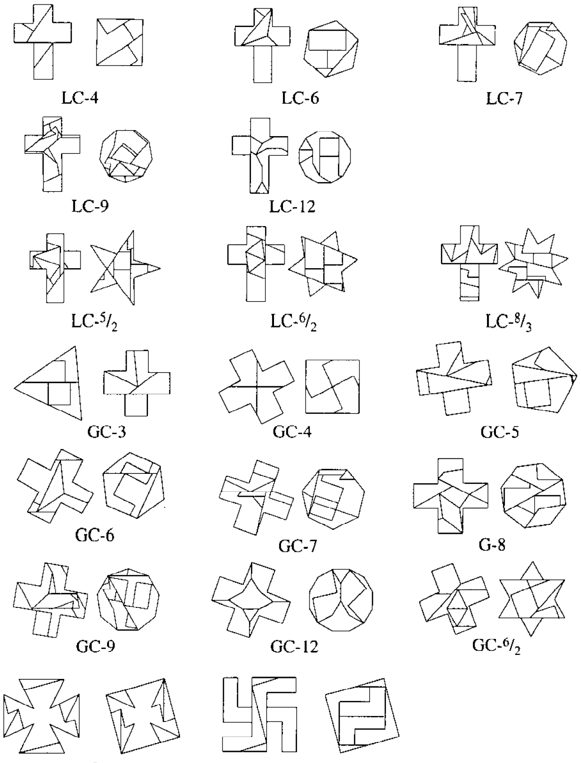

The best-known dissections of the GOLDEN RECTANGLE are illustrated below (Theobald).

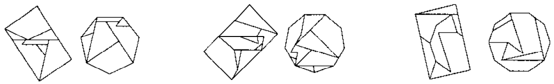

see also BANACH-TARSKI PARADOX, CUNDY AND ROLLETT'S EGG, DECAGON, DEHN INVARIANT, DIABOLICAL CUBE, DISSECTION PUZZLES, DODECAGON, EHRHART POLYNOMIAL, EQUIDECOMPOSABLE, EQUILATERAL TRIANGLE, GOLDEN RECTANGLE, HEPTAGON HEXAGON, HEXAGRAM, HILBERT'S PROBLEMS, LATIN CROSS, MALTESE CROSS, NONAGON, OCTAGON, OCTAGRAM, PENTAGON, PENTAGRAM, POLYHEDRON DISSECTION, PYTHAGOREAN SQUARE PUZZLE, PYTHAGOREAN THEOREM, REP-TILE, SOMA CUBE, SQUARE, STAR OF LAKSHMI, SWASTIKA, T-PUZZLE, TANGRAM, WALLACE-BOLYAI-GERWEIN THEOREM

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 87–94, 1987.

Coffin, S. T. *The Puzzling World of Polyhedral Dissections.* New York: Oxford University Press, 1990.

Cundy, H. and Rollett, A. Ch. 2 in *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., 1989.

Eppstein, D. "Dissection." `http://www.ics.uci.edu/~eppstein/junkyard/dissect.html`.

Eppstein, D. "Dissection Tiling." `http://www.ics.uci.edu/~eppstein/junkyard/distile`.

Eriksson, K. "Splitting a Polygon into Two Congruent Pieces." *Amer. Math. Monthly* **103**, 393–400, 1996.

Frederickson, G. *Dissections: Plane and Fancy.* New York: Cambridge University Press, 1997.

Gardner, M. *The Second Scientific American Book of Mathematical Puzzles & Diversions: A New Selection.* New York: Simon and Schuster, 1961.

Gardner, M. "Paper Cutting." Ch. 5 in *Martin Gardner's New Mathematical Diversions from Scientific American.* New York: Simon and Schuster, 1966.

Gardner, M. *The Unexpected Hanging and Other Mathematical Diversions.* Chicago, IL: Chicago University Press, 1991.

Hunter, J. A. H. and Madachy, J. S. *Mathematical Diversions.* New York: Dover, pp. 65–67, 1975.

Kraitchik, M. "Dissection of Plane Figures." §8.1 in *Mathematical Recreations.* New York: W. W. Norton, pp. 193–198, 1942.

Laczkovich, M. "Von Neumann's Paradox with Translation." *Fund. Math.* **131**, 1–12, 1988.

Lenhard, H.-C. "Über fünf neue Tetraeder, die einem Würfel äquivalent sind." *Elemente Math.* **17**, 108–109, 1962.

Lindgren, H. "Geometric Dissections." *Austral. Math. Teacher* **7**, 7–10, 1951.

Lindgren, H. "Geometric Dissections." *Austral. Math. Teacher* **9**, 17–21, 1953.

Lindgren, H. "Going One Better in Geometric Dissections." *Math. Gaz.* **45**, 94–97, 1961.

Lindgren, H. *Recreational Problems in Geometric Dissection and How to Solve Them.* New York: Dover, 1972.

Madachy, J. S. "Geometric Dissection." Ch. 1 in *Madachy's Mathematical Recreations.* New York: Dover, pp. 15–33, 1979.

Pappas, T. "A Triangle to a Square." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 9 and 230, 1989.

Stewart, I. *The Problems of Mathematics, 2nd ed.* Oxford, England: Oxford University Press, 1987.

Dissection Puzzles

A puzzle in which one object is to be converted to another by making a finite number of cuts and reassembling it. The cuts are often, but not always, restricted to straight lines. Sometimes, a given puzzle is precut and is to be re-assembled into two or more given shapes.

see also CUNDY AND ROLLETT'S EGG, PYTHAGOREAN SQUARE PUZZLE, T-PUZZLE, TANGRAM

Dissipative System

A system in which the phase space volume contracts along a trajectory. This means that the generalized DIVERGENCE is less than zero,

$$\frac{\partial f_i}{\partial x_i} < 0,$$

where EINSTEIN SUMMATION has been used.

Distance

Let $\gamma(t)$ be a smooth curve in a MANIFOLD M from x to y with $\gamma(0) = x$ and $\gamma(1) = y$. Then $\gamma'(t) \in T_{\gamma(t)}$, where

T_x is the TANGENT SPACE of M at x. The LENGTH of γ with respect to the Riemannian structure is given by

$$\int_0^1 ||\gamma'(t)||_{\gamma(t)}\,dt, \tag{1}$$

and the distance $d(x,y)$ between x and y is the shortest distance between x and y given by

$$d(x,y) = \inf_{\gamma:x \text{ to } y} \int ||\gamma'(t)||_{\gamma(t)}\,dt. \tag{2}$$

In order to specify the relative distances of $n > 1$ points in the plane, $1+2(n-2) = 2n-3$ coordinates are needed, since the first can always be taken as (0, 0) and the second as $(x, 0)$, which defines the x-AXIS. The remaining $n-2$ points need two coordinates each. However, the total number of distances is

$$\binom{n}{2} = \frac{n!}{2!(n-2)!} = \tfrac{1}{2}n(n-1), \tag{3}$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT. The distances between $n > 1$ points are therefore subject to m relationships, where

$$m \equiv \tfrac{1}{2}n(n-1) - (2n-3) = \tfrac{1}{2}(n-2)(n-3). \tag{4}$$

For $n = 1, 2, \ldots$, this gives 0, 0, 0, 1, 3, 6, 10, 15, 21, 28, ... (Sloane's A000217) relationships, and the number of relationships between n points is the TRIANGULAR NUMBER T_{n-3}.

Although there are no relationships for $n = 2$ and $n = 3$ points, for $n = 4$ (a QUADRILATERAL), there is one (Weinberg 1972):

$$\begin{aligned} 0 = {} & {d_{12}}^4 {d_{34}}^2 + {d_{13}}^4 {d_{24}}^2 + {d_{14}}^4 {d_{23}}^2 + {d_{23}}^4 {d_{14}}^2 \\ & + d_{24}^4 d_{13}^2 + d_{34}^4 d_{12}^2 \\ & + d_{12}^2 d_{23}^2 d_{31}^2 + d_{12}^2 d_{24}^2 d_{41}^2 + d_{13}^2 d_{34}^2 d_{41}^2 \\ & + d_{23}^2 d_{34}^2 d_{42}^2 - d_{12}^2 d_{23}^2 d_{34}^2 - d_{13}^2 d_{32}^2 d_{24}^2 \\ & - d_{12}^2 d_{24}^2 d_{43}^2 - d_{14}^2 d_{42}^2 d_{23}^2 - d_{13}^2 d_{34}^2 d_{42}^2 \\ & - d_{14}^2 d_{43}^2 d_{32}^2 - d_{23}^2 d_{31}^2 d_{14}^2 - d_{21}^2 d_{13}^2 d_{34}^2 \\ & - d_{24}^2 d_{41}^2 d_{13}^2 - d_{21}^2 d_{14}^2 d_{43}^2 - d_{31}^2 d_{12}^2 d_{24}^2 \\ & - d_{32}^2 d_{21}^2 d_{14}^2. \end{aligned} \tag{5}$$

This equation can be derived by writing

$$d_{ij} \equiv \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{6}$$

and eliminating x_i and y_j from the equations for d_{12}, d_{13}, d_{14}, d_{23}, d_{24}, and d_{34}.

see also ARC LENGTH, CUBE POINT PICKING, EXPANSIVE, LENGTH (CURVE), METRIC, PLANAR DISTANCE, POINT-LINE DISTANCE—2-D, POINT-LINE DISTANCE—3-D, POINT-PLANE DISTANCE, POINT-POINT DISTANCE—1-D, POINT-POINT DISTANCE—2-D, POINT-POINT DISTANCE—3-D, SPACE DISTANCE, SPHERE

References

Gray, A. "The Intuitive Idea of Distance on a Surface." §13.1 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 251–255, 1993.

Sloane, N. J. A. Sequence A000217/M2535 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Weinberg, S. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity.* New York: Wiley, p. 7, 1972.

Distinct Prime Factors

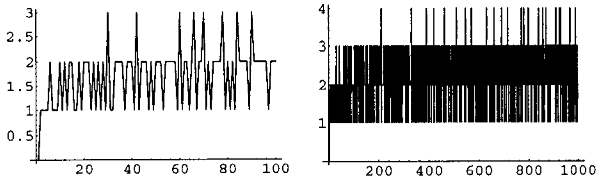

The number of distinct prime factors of a number n is denoted $\omega(n)$. The first few values for $n = 1, 2, \ldots$ are 0, 1, 1, 1, 1, 2, 1, 1, 1, 2, 1, 2, 1, 2, 2, 1, 1, 2, 1, 2, ... (Sloane's A001221). The first few values of the SUMMATORY FUNCTION

$$\sum_{k=2}^{n} \omega(k)$$

are 1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 14, 15, 17, 19, 20, 21, ... (Sloane's A013939), and the asymptotic value is

$$\sum_{k=2}^{n} \omega(k) = n \ln\ln n + B_1 n + o(n),$$

where B_1 is MERTENS CONSTANT. In addition,

$$\sum_{k=2}^{n} [\omega(k)]^2 = n(\ln\ln n)^2 + \mathcal{O}(n \ln\ln n).$$

see also DIVISOR FUNCTION, GREATEST PRIME FACTOR, HARDY-RAMANUJAN THEOREM, HETEROGENEOUS NUMBERS, LEAST PRIME FACTOR, MERTENS CONSTANT, PRIME FACTORS

References

Hardy, G. H. and Wright, E. M. "The Number of Prime Factors of n" and "The Normal Order of $\omega(n)$ and $\Omega(n)$." §22.10 and 22.11 in *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, pp. 354–358, 1979.

Sloane, N. J. A. Sequences A013939 and A001221/M0056 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Distribution

The distribution of a variable is a description of the relative numbers of times each possible outcome will occur in a number of trials. The function describing the distribution is called the Probability Function, and the function describing the probability that a given value or any value smaller than it will occur is called the Distribution Function.

Formally, a distribution can be defined as a normalized Measure, and the distribution of a Random Variable x is the Measure P_x on $\mathbb{S}'$ defined by setting

$$P_x(A') = P\{s \in S : x(s) \in A'\},$$

where $(S, \mathbb{S}, P)$ is a Probability Space, $(S, \mathbb{S})$ is a Measurable Space, and P a Measure on $\mathbb{S}$ with $P(S) = 1$.

see also Continuous Distribution, Discrete Distribution, Distribution Function, Measurable Space, Measure, Probability, Probability Density Function, Random Variable, Statistics

References

Doob, J. L. "The Development of Rigor in Mathematical Probability (1900–1950)." *Amer. Math. Monthly* **103**, 586–595, 1996.

Distribution Function

The distribution function $D(x)$, sometimes also called the Probability Distribution Function, describes the probability that a trial X takes on a value less than or equal to a number x. The distribution function is therefore related to a continuous Probability Density Function $P(x)$ by

$$D(x) = P(X \leq x) \equiv \int_{-\infty}^{x} P(x')\,dx', \tag{1}$$

so $P(x)$ (when it exists), is simply the derivative of the distribution function

$$P(x) = D'(x) = [P(x')]_{-\infty}^{x} = P(x) - P(-\infty). \tag{2}$$

Similarly, the distribution function is related to a discrete probability $P(x)$ by

$$D(x) = P(X \leq x) = \sum_{X \leq x} P(x). \tag{3}$$

In general, there exist distributions which are neither continuous nor discrete.

A Joint Distribution Function can be defined if outcomes are dependent on two parameters:

$$D(x,y) \equiv P(X \leq x, Y \leq y) \tag{4}$$

$$D_x(x) \equiv D(x, \infty) \tag{5}$$

$$D_y(y) \equiv D(\infty, y). \tag{6}$$

Similarly, a multiple distribution function can be defined if outcomes depend on n parameters:

$$D(a_1, \ldots, a_n) \equiv P(x_1 \leq a_1, \ldots, x_n \leq a_n). \tag{7}$$

Given a continuous $P(x)$, assume you wish to generate numbers distributed as $P(x)$ using a random number generator. If the random number generator yields a uniformly distributed value y_i in $[0,1]$ for each trial i, then compute

$$D(x) \equiv \int^{x} P(x')\,dx'. \tag{8}$$

The Formula connecting y_i with a variable distributed as $P(x)$ is then

$$x_i = D^{-1}(y_i), \tag{9}$$

where $D^{-1}(x)$ is the inverse function of $D(x)$. For example, if $P(x)$ were a Gaussian Distribution so that

$$D(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right], \tag{10}$$

then

$$x_i = \sigma\sqrt{2}\operatorname{erf}^{-1}(2y_i - 1) + \mu. \tag{11}$$

If $P(x) = Cx^n$ for $x \in (x_{\text{min}}, x_{\text{max}})$, then normalization gives

$$\int_{x_{\text{min}}}^{x_{\text{max}}} P(x)\,dx = C\frac{[x^{n+1}]_{x_{\text{min}}}^{x_{\text{max}}}}{n+1} = 1, \tag{12}$$

so

$$C = \frac{n+1}{{x_{\text{max}}}^{n+1} - {x_{\text{min}}}^{n+1}}. \tag{13}$$

Let y be a uniformly distributed variate on $[0,1]$. Then

$$\begin{aligned} D(x) &= \int_{x_{\text{min}}}^{x} P(x)\,dx = C\int_{x_{\text{min}}}^{x} x^n\,dx \\ &= \frac{C}{n+1}(x^{n+1} - {x_{\text{min}}}^{n+1}) \equiv y, \end{aligned} \tag{14}$$

and the variate given by

$$\begin{aligned} x &= \left(\frac{n+1}{C}y + {x_{\text{min}}}^{n+1}\right)^{1/(n+1)} \\ &= [({x_{\text{max}}}^{n+1} - {x_{\text{min}}}^{n+1})y + {x_{\text{min}}}^{n+1}]^{1/(n+1)} \end{aligned} \tag{15}$$

is distributed as $P(x)$.

A distribution with constant Variance of y for all values of x is known as a Homoscedastic distribution. The method of finding the value at which the distribution is a maximum is known as the Maximum Likelihood method.

see also Bernoulli Distribution, Beta Distribution, Binomial Distribution, Bivariate Distribution, Cauchy Distribution, Chi Distribution, Chi-Squared Distribution, Cornish-Fisher

ASYMPTOTIC EXPANSION, CORRELATION COEFFICIENT, DISTRIBUTION, DOUBLE EXPONENTIAL DISTRIBUTION, EQUALLY LIKELY OUTCOMES DISTRIBUTION, EXPONENTIAL DISTRIBUTION, EXTREME VALUE DISTRIBUTION, F-DISTRIBUTION, FERMI-DIRAC DISTRIBUTION, FISHER'S z-DISTRIBUTION, FISHER-TIPPETT DISTRIBUTION, GAMMA DISTRIBUTION, GAUSSIAN DISTRIBUTION, GEOMETRIC DISTRIBUTION, HALF-NORMAL DISTRIBUTION, HYPERGEOMETRIC DISTRIBUTION, JOINT DISTRIBUTION FUNCTION, LAPLACE DISTRIBUTION, LATTICE DISTRIBUTION, LÉVY DISTRIBUTION, LOGARITHMIC DISTRIBUTION, LOG-SERIES DISTRIBUTION, LOGISTIC DISTRIBUTION, LORENTZIAN DISTRIBUTION, MAXWELL DISTRIBUTION, NEGATIVE BINOMIAL DISTRIBUTION, NORMAL DISTRIBUTION, PARETO DISTRIBUTION, PASCAL DISTRIBUTION, PEARSON TYPE III DISTRIBUTION, POISSON DISTRIBUTION, PÓLYA DISTRIBUTION, RATIO DISTRIBUTION, RAYLEIGH DISTRIBUTION, RICE DISTRIBUTION, SNEDECOR'S F-DISTRIBUTION, STUDENT'S t-DISTRIBUTION, STUDENT'S z-DISTRIBUTION, UNIFORM DISTRIBUTION, WEIBULL DISTRIBUTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Probability Functions." Ch. 26 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 925–964, 1972.

Iyanaga, S. and Kawada, Y. (Eds.). "Distribution of Typical Random Variables." Appendix A, Table 22 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 1483–1486, 1980.

Distribution (Functional)

A functional distribution, also called a GENERALIZED FUNCTION, is a generalization of the concept of a function. Functional distributions are defined as continuous linear FUNCTIONALS over a SPACE of infinitely differentiable functions such that all continuous functions have SCHWARZIAN DERIVATIVES which are themselves distributions. The most commonly encountered functional distribution is the DELTA FUNCTION.

see also DELTA FUNCTION, GENERALIZED FUNCTION, SCHWARZIAN DERIVATIVE

References

Friedlander, F. G. *Introduction to the Theory of Distributions.* Cambridge, England: Cambridge University Press, 1982.

Gel'fand, I. M. and Shilov, G. E. *Generalized Functions, Vol. 1: Properties and Operations.* New York: Harcourt Brace, 1977.

Gel'fand, I. M. and Shilov, G. E. *Generalized Functions, Vol. 2: Spaces of Fundamental and Generalized Functions.* New York: Harcourt Brace, 1977.

Gel'fand, I. M. and Shilov, G. E. *Generalized Functions, Vol. 3: Theory of Differential Equations.* New York: Harcourt Brace, 1977.

Gel'fand, I. M. and Vilenkin, N. Ya. *Generalized Functions, Vol. 4: Applications of Harmonic Analysis.* New York: Harcourt Brace, 1977.

Gel'fand, I. M.; Graev, M. I.; and Vilenkin, N. Ya. *Generalized Functions, Vol. 5: Integral Geometry and Representation Theory.* New York: Harcourt Brace, 1977.

Griffel, D. H. *Applied Functional Analysis.* Englewood Cliffs, NJ: Prentice-Hall, 1984.

Halperin, I. and Schwartz, L. *Introduction to the Theory of Distributions, Based on the Lectures Given by Laurent Schwarz.* Toronto, Canada: University of Toronto Press, 1952.

Lighthill, M. J. *Introduction to Fourier Analysis and Generalised Functions.* Cambridge, England: Cambridge University Press, 1958.

Richards, I. and Young, H. *The Theory of Distributions: A Nontechnical Introduction.* New York: Cambridge University Press, 1995.

Rudin, W. *Functional Analysis, 2nd ed.* New York: McGraw-Hill, 1991.

Strichartz, R. *Fourier Transforms and Distribution Theory.* Boca Raton, FL: CRC Press, 1993.

Zemanian, A. H. *Distribution Theory and Transform Analysis: An Introduction to Generalized Functions, with Applications.* New York: Dover, 1987.

Distribution Parameter

The distribution parameter of a NONCYLINDRICAL RULED SURFACE parameterized by

$$\mathbf{x}(u,v) = \boldsymbol{\sigma}(u) + v\boldsymbol{\delta}(u), \tag{1}$$

where $\boldsymbol{\sigma}$ is the STRICTION CURVE and $\boldsymbol{\delta}$ the DIRECTOR CURVE, is the function p defined by

$$p = \frac{\det(\boldsymbol{\sigma}'\boldsymbol{\delta}\boldsymbol{\delta}')}{\boldsymbol{\delta}'\cdot\boldsymbol{\delta}'}. \tag{2}$$

The GAUSSIAN CURVATURE of a RULED SURFACE is given in terms of its distribution parameter by

$$K = -\frac{[p(u)]^2}{\{[p(u)]^2 + v^2\}^2}. \tag{3}$$

see also NONCYLINDRICAL RULED SURFACE, STRICTION CURVE

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 347–348, 1993.

Distribution (Statistical)

The set of probabilities for each possible event.

see DISTRIBUTION FUNCTION

Distributive

Elements of an ALGEBRA which obey the identity

$$A(B+C) = AB + AC$$

are said to be distributive over the operation $+$.

see also ASSOCIATIVE, COMMUTATIVE, TRANSITIVE

Distributive Lattice

A LATTICE which satisfies the identities

$$(x \wedge y) \vee (x \wedge z) = x \wedge (y \vee z)$$

$$(x \vee y) \wedge (x \vee z) = x \vee (y \wedge z)$$

is said to be distributive.

see also LATTICE, MODULAR LATTICE

References

Grätzer, G. *Lattice Theory: First Concepts and Distributive Lattices.* San Francisco, CA: W. H. Freeman, pp. 35–36, 1971.

Disymmetric

An object which is not superimposable on its MIRROR IMAGE is said to be disymmetric. All asymmetric objects are disymmetric, and an object with no IMPROPER ROTATION (rotoinversion) axis must also be disymmetric.

Ditrigonal Dodecadodecahedron

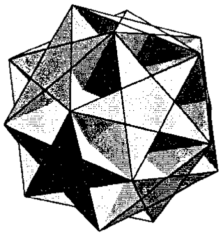

The UNIFORM POLYHEDRON U_{41}, also called the DITRIGONAL DODECAHEDRON, whose DUAL POLYHEDRON is the MEDIAL TRIAMBIC ICOSAHEDRON. It has WYTHOFF SYMBOL $3\,|\,\frac{5}{3}\,5$. Its faces are $12\{\frac{5}{2}\}+12\{5\}$. It is a FACETED version of the SMALL DITRIGONAL ICOSIDODECAHEDRON. The CIRCUMRADIUS for unit edge length is

$$R = \tfrac{1}{2}\sqrt{3}.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 123–124, 1989.

Ditrigonal Dodecahedron

see DITRIGONAL DODECADODECAHEDRON

Divergence

The divergence of a VECTOR FIELD $\mathbf{F}$ is given by

$$\operatorname{div}(\mathbf{F}) \equiv \nabla\cdot\mathbf{F} \equiv \lim_{V\to 0} \frac{\oint_S \mathbf{F}\cdot d\mathbf{a}}{V}. \tag{1}$$

Define

$$\mathbf{F} \equiv F_1\hat{\mathbf{u}}_1 + F_2\hat{\mathbf{u}}_2 + F_3\hat{\mathbf{u}}_3. \tag{2}$$

Then in arbitrary orthogonal CURVILINEAR COORDINATES,

$$\begin{aligned}\operatorname{div}(F) \equiv \nabla\cdot\mathbf{F} \equiv \frac{1}{h_1h_2h_3}\bigg[&\frac{\partial}{\partial u_1}(h_2h_3F_1)\\ &+\frac{\partial}{\partial u_2}(h_3h_1F_2) + \frac{\partial}{\partial u_3}(h_1h_2F_3)\bigg].\end{aligned} \tag{3}$$

If $\nabla\cdot\mathbf{F} = 0$, then the field is said to be a DIVERGENCELESS FIELD. For divergence in individual coordinate systems, see CURVILINEAR COORDINATES.

$$\nabla\cdot\frac{\mathsf{A}\mathbf{x}}{|\mathbf{x}|} = \frac{\operatorname{Tr}(\mathsf{A})}{|\mathbf{x}|} - \frac{\mathbf{x}^{\mathrm{T}}(\mathsf{A}\mathbf{x})}{|\mathbf{x}|^3}. \tag{4}$$

The divergence of a TENSOR A is

$$\nabla\cdot A \equiv A^\alpha_{;\alpha} = A^k_{,k} + \Gamma^k_{jk}A^j, \tag{5}$$

where ; is the COVARIANT DERIVATIVE. Expanding the terms gives

$$\begin{aligned}A^\alpha_{;\alpha} = A^\alpha_{,\alpha} &+ (\Gamma^\alpha_{\alpha\alpha}A^\alpha + \Gamma^\alpha_{\beta\alpha}A^\beta + \Gamma^\alpha_{\gamma\alpha}A^\gamma)\\ &+ A^\beta_{,\beta} + \left(\Gamma^\beta_{\alpha\beta}A^\alpha + \Gamma^\beta_{\beta\beta}A^\beta + \Gamma^\beta_{\gamma\beta}A^\gamma\right)\\ &+ A^\gamma_{,\gamma} + \left(\Gamma^\gamma_{\alpha\gamma}A^\alpha + \Gamma^\gamma_{\beta\gamma}A^\beta + \Gamma^\gamma_{\gamma\gamma}A^\gamma\right).\end{aligned} \tag{6}$$

see also CURL, CURL THEOREM, GRADIENT, GREEN'S THEOREM, DIVERGENCE THEOREM, VECTOR DERIVATIVE

References

Arfken, G. "Divergence, ∇·." §1.7 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 37–42, 1985.

Divergence Tests

If

$$\lim_{k\to\infty} u_k \neq 0,$$

then the series $\{u_n\}$ diverges.

see also CONVERGENCE TESTS, CONVERGENT SERIES, DINI'S TEST, SERIES

Divergence Theorem

A.k.a. GAUSS'S THEOREM. Let V be a region in space with boundary ∂V. Then

$$\int_V (\nabla\cdot\mathbf{F})\,dV = \int_{\partial V} \mathbf{F}\cdot d\mathbf{a}. \tag{1}$$

Let S be a region in the plane with boundary ∂S.

$$\int_S \nabla\cdot\mathbf{F}\,dA = \int_{\partial S} \mathbf{F}\cdot\mathbf{n}\,ds. \tag{2}$$

If the VECTOR FIELD $\mathbf{F}$ satisfies certain constraints, simplified forms can be used. If $\mathbf{F}(x,y,z) = v(x,y,z)\mathbf{c}$ where $\mathbf{c}$ is a constant vector $\neq \mathbf{0}$, then

$$\int_S \mathbf{F}\cdot d\mathbf{a} = \mathbf{c}\cdot\int_S v\,d\mathbf{a}. \tag{3}$$

But

$$\nabla \cdot (f\mathbf{v}) = (\nabla f) \cdot \mathbf{v} + f(\nabla \cdot \mathbf{v}), \tag{4}$$

so

$$\int_V \nabla \cdot (\mathbf{c}v)\, dV = \mathbf{c} \cdot \int_V (\nabla v + v\nabla \cdot \mathbf{c})\, dV = \mathbf{c} \cdot \int_V \nabla v\, dV \tag{5}$$

$$\mathbf{c} \cdot \left(\int_S v\, d\mathbf{a} - \int_V \nabla v\, dV \right) = 0. \tag{6}$$

But $\mathbf{c} \neq \mathbf{0}$, and $\mathbf{c} \cdot \mathbf{f}(v)$ must vary with v so that $\mathbf{c} \cdot \mathbf{f}(v)$ cannot always equal zero. Therefore,

$$\int_S v\, d\mathbf{a} = \int_V \nabla v\, dV. \tag{7}$$

If $\mathbf{F}(x, y, z) = \mathbf{c} \times P(x, y, z)$, where $\mathbf{c}$ is a constant vector $\neq \mathbf{0}$, then

$$\int_S d\mathbf{a} \times \mathbf{P} = \int_V \nabla \times \mathbf{P}\, dV. \tag{8}$$

see also CURL THEOREM, GRADIENT, GREEN'S THEOREM

References

Arfken, G. "Gauss's Theorem." §1.11 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 57–61, 1985.

Divergenceless Field

A divergenceless field, also called a SOLENOIDAL FIELD, is a FIELD for which $\nabla \cdot \mathbf{F} \equiv \mathbf{0}$. Therefore, there exists a $\mathbf{G}$ such that $\mathbf{F} = \nabla \times \mathbf{G}$. Furthermore, $\mathbf{F}$ can be written as

$$\mathbf{F} = \nabla \times (T\mathbf{r}) + \nabla^2(S\mathbf{r}) \equiv \mathbf{T} + \mathbf{S},$$

where

$$\mathbf{T} \equiv \nabla \times (T\mathbf{r}) = -\mathbf{r} \times (\nabla T)$$

$$\mathbf{S} \equiv \nabla^2(S\mathbf{r}) = \nabla \left[\frac{\partial}{\partial r}(rS) \right] - \mathbf{r}\nabla^2 S.$$

Following Lamb, $\mathbf{T}$ and $\mathbf{S}$ are called TOROIDAL FIELD and POLOIDAL FIELD.

see also BELTRAMI FIELD, IRROTATIONAL FIELD, POLOIDAL FIELD, SOLENOIDAL FIELD, TOROIDAL FIELD

Divergent Sequence

A divergent sequence is a SEQUENCE for which the LIMIT exists but is not CONVERGENT.

see also CONVERGENT SEQUENCE, DIVERGENT SERIES

Divergent Series

A SERIES which is not CONVERGENT. Series may diverge by marching off to infinity or by oscillating.

see also CONVERGENT SERIES, DIVERGENT SEQUENCE

References

Bromwich, T. J. I'a and MacRobert, T. M. *An Introduction to the Theory of Infinite Series, 3rd ed.* New York: Chelsea, 1991.

Diversity Condition

For any group of k men out of N, there must be at least k jobs for which they are collectively qualified.

Divide

To divide is to perform the operation of DIVISION, i.e., to see how many time a DIVISOR d goes into another number n. n divided by d is written n/d or $n \div d$. The result need not be an INTEGER, but if it is, some additional terminology is used. $d|n$ is read "d divides n" and means that d is a PROPER DIVISOR of n. In this case, n is said to be DIVISIBLE by d. Clearly, $1|n$ and $n|n$. By convention, $n|0$ for every n except 0 (Hardy and Wright 1979). The "divided" operation satisfies

$$b|a \text{ and } c|b \Rightarrow c|a$$

$$b|a \Rightarrow bc|ac$$

$$c|a \text{ and } c|b \Rightarrow c|(ma + nb).$$

$d' \nmid n$ is read "d' does not divide n" and means that d' is not a PROPER DIVISOR of n. $a^k || b$ means a^k divides b exactly.

see also CONGRUENCE, DIVISIBLE, DIVISION, DIVISOR

References

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, p. 1, 1979.

Divided Difference

The divided difference $f[x_1, x_2, \ldots, x_n]$ on n points x_1, x_2, ..., x_n of a function $f(x)$ is defined by $f[x_1] \equiv f(x_1)$ and

$$f[x_1, x_2, \ldots, x_n] = \frac{f[x_1, \ldots, x_n] - f[x_2, \ldots, x_n]}{x_1 - x_n} \tag{1}$$

for $n \geq 2$. The first few differences are

$$[x_0, x_1] = \frac{f_0 - f_1}{x_0 - x_1} \tag{2}$$

$$[x_0, x_1, x_2] = \frac{[x_0, x_1] - [x_1, x_2]}{x_0 - x_2} \tag{3}$$

$$[x_0, x_1, \ldots, x_n] = \frac{[x_0, \ldots, x_{n-1}] - [x_1, \ldots, x_n]}{x_0 - x_n}. \tag{4}$$

Defining

$$\pi_n(x) \equiv (x - x_0)(x - x_1) \cdots (x - x_n) \tag{5}$$

and taking the DERIVATIVE

$$\pi_n'(x_k) = (x_k - x_0) \cdots (x_k - x_{k-1})(x_k - x_{k+1}) \cdots (x_k - x_n) \tag{6}$$

gives the identity

$$[x_0, x_1, \ldots, x_n] = \sum_{k=0}^{n} \frac{f_k}{\pi_n'(x_k)}. \tag{7}$$

Consider the following question: does the property

$$f[x_1, x_2, \ldots, x_n] = h(x_1 + x_2 + \ldots + x_n) \qquad (8)$$

for $n \geq 2$ and $h(x)$ a given function guarantee that $f(x)$ is a POLYNOMIAL of degree $\leq n$? Aczél (1985) showed that the answer is "yes" for $n = 2$, and Bailey (1992) showed it to be true for $n = 3$ with differentiable $f(x)$. Schwaiger (1994) and Andersen (1996) subsequently showed the answer to be "yes" for all $n \geq 3$ with restrictions on $f(x)$ or $h(x)$.

see also NEWTON'S DIVIDED DIFFERENCE INTERPOLATION FORMULA, RECIPROCAL DIFFERENCE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 877–878, 1972.

Aczél, J. "A Mean Value Property of the Derivative of Quadratic Polynomials—Without Mean Values and Derivatives." *Math. Mag.* **58**, 42–45, 1985.

Andersen, K. M. "A Characterization of Polynomials." *Math. Mag.* **69**, 137–142, 1996.

Bailey, D. F. "A Mean-Value Property of Cubic Polynomials—Without Mean Values." *Math. Mag.* **65**, 123–124, 1992.

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 439–440, 1987.

Schwaiger, J. "On a Characterization of Polynomials by Divided Differences." *Aequationes Math.* **48**, 317–323, 1994.

Divine Proportion

see GOLDEN RATIO

Divisibility Tests

Write a decimal number a out digit by digit in the form $a_n \ldots a_3 a_2 a_1 a_0$. It is always true that $10^0 = 1 \equiv 1$ for any base.

2 $10^1 \equiv 0$, so $10^n \equiv 0$ for $n \geq 1$. Therefore, if the last digit a_0 is divisible by 2 (i.e., is EVEN), then so is a.

3 $10^1 \equiv 1$, $10^2 \equiv 1$, ..., $10^n \equiv 1$. Therefore, if $\sum_{i=1}^n a_i$ is divisible by 3, so is a.

4 $10^1 \equiv 2$, $10^2 \equiv 0$, $\ldots 10^n \equiv 0$. So if the last two digits are divisible by 4, more specifically if $r \equiv a_0 + 2a_1$ is, then so is a.

5 $10^1 \equiv 0$, so $10^n \equiv 0$ for $n \geq 1$. Therefore, if the last digit a_0 is divisible by 5 (i.e., is 5 or 0), then so is a_0.

6 $10^1 \equiv -2$, $10^2 \equiv -2$, so $10^n \equiv -2$. Therefore, if $r \equiv a_0 - 2\sum_{i=1}^n a_i$ is divisible by 6, so is a. If a is divisible by 3 and is EVEN, it is also divisible by 6.

7 $10^1 \equiv 3$, $10^2 \equiv 2$, $10^3 \equiv -1$, $10^4 \equiv -3$, $10^5 \equiv -2$, $10^6 \equiv 1$, and the sequence then repeats. Therefore, if $r \equiv (a_0 + 3a_1 + 2a_2 - a_3 - 3a_4 - 2a_5) + (a_6 + 3a_7 + \ldots) + \ldots$ is divisible by 7, so is a.

8 $10^1 \equiv 2$, $10^2 \equiv 4$, $10^3 \equiv 0$, ..., $10^n \equiv 0$. Therefore, if the last three digits are divisible by 8, more specifically if $r \equiv a_0 + 2a_1 + 4a_2$ is, then so is a.

9 $10^1 \equiv 1$, $10^2 \equiv 1$, ..., $10^3 \equiv 1$. Therefore, if $\sum_{i=1}^n a_i$ is divisible by 9, so is a.

10 $10^1 \equiv 0$, so if the last digit is 0, then a is divisible by 10.

11 $10^1 \equiv -1$, $10^2 \equiv 1$, $10^3 \equiv -1$, $10^4 \equiv 1$, ... Therefore, if $r \equiv a_0 - a_1 + a_2 - a_3 + \ldots$ is divisible by 11, then so is a.

12 $10^1 \equiv -2$, $10^2 \equiv 4$, $10^3 \equiv 4$, ... Therefore, if $r \equiv a_0 - 2a_1 + 4(a_2 + a_3 + \ldots)$ is divisible by 12, then so is a. Divisibility by 12 can also be checked by seeing if a is divisible by 3 and 4.

13 $10^1 \equiv -3$, $10^2 \equiv -4$, $10^3 \equiv -1$, $10^4 \equiv 3$, $10^5 \equiv 4$, $10^6 \equiv 1$, and the pattern repeats. Therefore, if $r \equiv (a_0 - 3a_1 - 4a_2 - a_3 + 3a_4 + 4a_5) + (a_6 - 3a_7 + \ldots) + \ldots$ is divisible by 13, so is a.

For additional tests for 13, see Gardner (1991).

References

Dickson, L. E. *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, pp. 337–346, 1952.

Gardner, M. Ch. 14 in *The Unexpected Hanging and Other Mathematical Diversions.* Chicago, IL: Chicago University Press, 1991.

Divisible

A number n is said to be divisible by d if d is a PROPER DIVISOR of n. The sum of any n consecutive INTEGERS is divisible by $n!$, where $n!$ is the FACTORIAL.

see also DIVIDE, DIVISOR, DIVISOR FUNCTION

References

Guy, R. K. "Divisibility." Ch. B in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 44–104, 1994.

Division

Taking the RATIO x/y of two numbers x and y, also written $x \div y$. Here, y is called the DIVISOR. The symbol "/" is called a SOLIDUS (or DIAGONAL), and the symbol "÷" is called the OBELUS. Division in which the fractional (remainder) is discarded is called INTEGER DIVISION, and is sometimes denoted using a backslash, \.

see also ADDITION, DIVIDE, INTEGER DIVISION, LONG DIVISION, MULTIPLICATION, OBELUS, ODDS, RATIO, SKELETON DIVISION, SOLIDUS, SUBTRACTION, TRIAL DIVISION

Division Algebra

A division algebra, also called a DIVISION RING or SKEW FIELD, is a RING in which every NONZERO element has a multiplicative inverse, but multiplication is not COMMUTATIVE. Explicitly, a division algebra is a set together with two BINARY OPERATORS $S(+, *)$ satisfying the following conditions:

1. Additive associativity: For all $a, b, c \in S$, $(a+b)+c = a+(b+c)$,
2. Additive commutativity: For all $a, b \in S$, $a+b = b+a$,
3. Additive identity: There exists an element $0 \in S$ such that for all $a \in S$, $0+a = a+0 = a$,
4. Additive inverse: For every $a \in S$ there exists a $-a \in S$ such that $a+(-a) = (-a)+a = 0$,
5. Multiplicative associativity: For all $a, b, c \in S$, $(a*b)*c = a*(b*c)$,
6. Multiplicative identity: There exists an element $1 \in S$ not equal to 0 such that for all $a \in S$, $1*a = a*1 = a$,
7. Multiplicative inverse: For every $a \in S$ not equal to 0, there exists $a^{-1} \in S$, $a*a^{-1} = a^{-1}*a = 1$,
8. Left and right distributivity: For all $a, b, c \in S$, $a*(b+c) = (a*b)+(a*c)$ and $(b+c)*a = (b*a)+(c*a)$.

Thus a division algebra $(S, +, *)$ is a UNIT RING for which $(S - \{0\}, *)$ is a GROUP. A division algebra must contain at least two elements. A COMMUTATIVE division algebra is called a FIELD.

In 1878 and 1880, Frobenius and Peirce proved that the only associative REAL division algebras are real numbers, COMPLEX NUMBERS, and QUATERNIONS. The CAYLEY ALGEBRA is the only NONASSOCIATIVE DIVISION ALGEBRA. Hurwitz (1898) proved that the ALGEBRAS of REAL NUMBERS, COMPLEX NUMBERS, QUATERNIONS, and CAYLEY NUMBERS are the only ones where multiplication by unit "vectors" is distance-preserving. Adams (1956) proved that n-D vectors form an ALGEBRA in which division (except by 0) is always possible only for $n = 1$, 2, 4, and 8.

see also CAYLEY NUMBER, FIELD, GROUP, NONASSOCIATIVE ALGEBRA, QUATERNION, UNIT RING

References

Dickson, L. E. *Algebras and Their Arithmetics.* Chicago, IL: University of Chicago Press, 1923.

Dixon, G. M. *Division Algebras: Octonions, Quaternions, Complex Numbers and the Algebraic Design of Physics.* Dordrecht, Netherlands: Kluwer, 1994.

Herstein, I. N. *Topics in Algebra, 2nd ed.* New York: Wiley, pp. 326–329, 1975.

Hurwitz, A. "Ueber die Composition der quadratischen Formen von beliebig vielen Variabeln." *Nachr. Gesell. Wiss. Göttingen, Math.-Phys. Klasse,* 309–316, 1898.

Kurosh, A. G. *General Algebra.* New York: Chelsea, pp. 221–243, 1963.

Petro, J. "Real Division Algebras of Dimension > 1 contain ℂ." *Amer. Math. Monthly* **94**, 445–449, 1987.

Division Lemma

When ac is DIVISIBLE by a number b that is RELATIVELY PRIME to a, then c must be DIVISIBLE by b.

Division Ring

see DIVISION ALGEBRA

Divisor

A divisor of a number N is a number d which DIVIDES N, also called a FACTOR. The total number of divisors for a given number N can be found as follows. Write a number in terms of its PRIME FACTORIZATION

$$N = {p_1}^{\alpha_1} {p_2}^{\alpha_2} \cdots {p_r}^{\alpha_r}. \tag{1}$$

For any divisor d of N, $N = dd'$ where

$$d = {p_1}^{\delta_1} {p_2}^{\delta_2} \cdots {p_r}^{\delta_r}, \tag{2}$$

so

$$d' = {p_1}^{\alpha_1-\delta_1} {p_2}^{\alpha_2-\delta_2} \cdots {p_r}^{\alpha_r-\delta_r}. \tag{3}$$

Now, $\delta_1 = 0, 1, \ldots, \alpha_1$, so there are $\alpha_1 + 1$ possible values. Similarly, for δ_n, there are $\alpha_n + 1$ possible values, so the total number of divisors $\nu(N)$ of N is given by

$$\nu(N) = \prod_{n=1}^{r} (\alpha_n + 1). \tag{4}$$

The function $\nu(N)$ is also sometimes denoted $d(N)$ or $\sigma_0(N)$. The product of divisors can be found by writing the number N in terms of all possible products

$$N = \begin{cases} d^{(1)} d'^{(1)} \\ \vdots \\ d^{(\nu)} d'^{(\nu)} \end{cases}, \tag{5}$$

so

$$\begin{aligned} N^{\nu(N)} &= [d^{(1)} \cdots d^{(\nu)}][d'^{(1)} d'^{(\nu)}] \\ &= \prod_{i=1}^{\nu} d_i \prod_{i=1}^{\nu} {d_i}' = \left(\prod d\right)^2, \end{aligned} \tag{6}$$

and

$$\prod d = N^{\nu(N)/2}. \tag{7}$$

The GEOMETRIC MEAN of divisors is

$$G \equiv \left(\prod d\right)^{\nu(N)/2} = [N^{\nu(n)/2}]^{1/\nu(N)} = \sqrt{N}. \tag{8}$$

The sum of the divisors can be found as follows. Let $N \equiv ab$ with $a \neq b$ and $(a, b) = 1$. For any divisor d of N, $d = a_i b_i$, where a_i is a divisor of a and b_i is a divisor of b. The divisors of a are 1, a_1, a_2, …, and a. The divisors of b are 1, b_1, b_2, …, b. The sums of the divisors are then

$$\sigma(a) = 1 + a_1 + a_2 + \ldots + a \tag{9}$$

$$\sigma(b) = 1 + b_1 + b_2 + \ldots + b. \tag{10}$$

For a given a_i,

$$a_i(1 + b_1 + b_2 + \ldots + b) = a_i \sigma(b). \tag{11}$$

Summing over all a_i,

$$(1 + a_1 + a_2 + \ldots + a)\sigma(b) = \sigma(a)\sigma(b), \qquad (12)$$

so $\sigma(N) = \sigma(ab) = \sigma(a)\sigma(b)$. Splitting a and b into prime factors,

$$\sigma(N) = \sigma({p_1}^{\alpha_1})\sigma({p_2}^{\alpha_2})\cdots\sigma({p_r}^{\alpha_r}). \qquad (13)$$

For a prime POWER ${p_i}^{\alpha_i}$, the divisors are 1, p_i, ${p_i}^2$, ..., ${p_i}^{\alpha_i}$, so

$$\sigma({p_i}^{\alpha_i}) = 1 + p_i + {p_i}^2 + \ldots + {p_i}^{\alpha_i} = \frac{{p_i}^{\alpha_i+1} - 1}{p_i - 1}. \qquad (14)$$

For N, therefore,

$$\sigma(N) = \prod_{i=1}^{r} \frac{{p_i}^{\alpha_i+1} - 1}{p_i - 1}. \qquad (15)$$

For the special case of N a PRIME, (15) simplifies to

$$\sigma(p) = \frac{p^2 - 1}{p - 1} = p + 1. \qquad (16)$$

For N a POWER of two, (15) simplifies to

$$\sigma(2^\alpha) = \frac{2^{\alpha+1} - 1}{2 - 1} = 2^{\alpha+1} - 1. \qquad (17)$$

The ARITHMETIC MEAN is

$$A(N) \equiv \frac{\sigma(N)}{\nu(N)}. \qquad (18)$$

The HARMONIC MEAN is

$$\frac{1}{H} \equiv \frac{1}{n}\left(\sum \frac{1}{d}\right). \qquad (19)$$

But $N = dd'$, so $\frac{1}{\nu(N)} = \frac{d'}{N}$ and

$$\sum \frac{1}{d} = \frac{1}{N}\sum d' = \frac{1}{N}\sum d = \frac{\sigma(N)}{N}, \qquad (20)$$

and we have

$$\frac{1}{H(N)} = \frac{1}{N(\nu)}\frac{\sigma(N)}{N} = \frac{A(N)}{N} \qquad (21)$$

$$N = A(N)H(N). \qquad (22)$$

Given three INTEGERS chosen at random, the probability that no common factor will divide them all is

$$[\zeta(3)]^{-1} \approx 1.202^{-1} = 0.832\ldots, \qquad (23)$$

where $\zeta(3)$ is APÉRY'S CONSTANT.

Let $f(n)$ be the number of elements in the greatest subset of $[1, n]$ such that none of its elements are divisible by two others. For n sufficiently large,

$$0.6725\ldots \le \frac{f(n)}{n} \le 0.673\ldots \qquad (24)$$

(Le Lionnais 1983, Lebensold 1976/1977).

see also ALIQUANT DIVISOR, ALIQUOT DIVISOR, ALIQUOT SEQUENCE, DIRICHLET DIVISOR PROBLEM, DIVISOR FUNCTION, *e*-DIVISOR, EXPONENTIAL DIVISOR, GREATEST COMMON DIVISOR, INFINARY DIVISOR, *k*-ARY DIVISOR, PERFECT NUMBER, PROPER DIVISOR, UNITARY DIVISOR

References

Guy, R. K. "Solutions of $d(n) = d(n+1)$." §B18 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 73–75, 1994.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 43, 1983.

Lebensold, K. "A Divisibility Problem." *Studies Appl. Math.* **56**, 291–294, 1976/1977.

Divisor Function

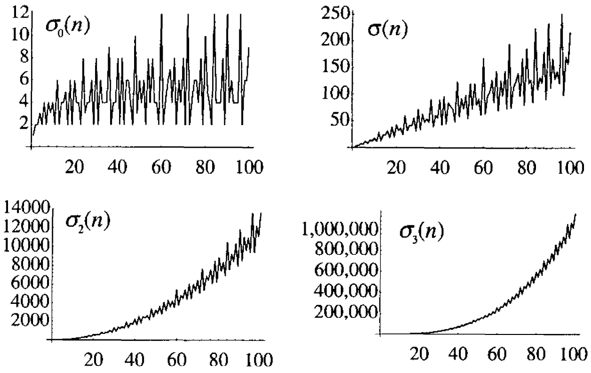

$\sigma_k(n)$ is defined as the sum of the kth POWERS of the DIVISORS of n. The function $\sigma_0(n)$ gives the total number of DIVISORS of n and is often denoted $d(n)$, $\nu(n)$, $\tau(n)$, or $\Omega(n)$ (Hardy and Wright 1979, pp. 354–355). The first few values of $\sigma_0(n)$ are 1, 2, 2, 3, 2, 4, 2, 4, 3, 4, 2, 6, ... (Sloane's A000005). The function $\sigma_1(n)$ is equal to the sum of DIVISORS of n and is often denoted $\sigma(n)$. The first few values of $\sigma(n)$ are 1, 3, 4, 7, 6, 12, 8, 15, 13, 18, ... (Sloane's A000203). The first few values of $\sigma_2(n)$ are 1, 5, 10, 21, 26, 50, 50, 85, 91, 130, ... (Sloane's A001157). The first few values of $\sigma_3(n)$ are 1, 9, 28, 73, 126, 252, 344, 585, 757, 1134, ... (Sloane's A001158).

The sum of the DIVISORS of n excluding n itself (i.e., the PROPER DIVISORS of n) is called the RESTRICTED DIVISOR FUNCTION and is denoted $s(n)$. The first few values are 0, 1, 1, 3, 1, 6, 1, 7, 4, 8, 1, 16, ... (Sloane's A001065).

As an illustrative example, consider the number 140, which has DIVISORS $d_i = 1, 2, 4, 5, 7, 10, 14, 20, 28, 35, 70$, and 140 (for a total of $N = 12$ of them). Therefore,

$$d(140) = N = 12 \tag{1}$$

$$\sigma(140) = \sum_i^N d_i = 336 \tag{2}$$

$$\sigma_2(140) = \sum_i^N {d_i}^2 = 27,300 \tag{3}$$

$$\sigma_3(140) = \sum_i^N {d_i}^3 = 3,164,112. \tag{4}$$

The $\sigma(n)$ function has the series expansion

$$\sigma(n) = \tfrac{1}{6}\pi^2 n \left[1 + \frac{(-1)^n}{2^2} + \frac{2\cos(\frac{2}{3}n\pi)}{3^2} + \frac{2\cos(\frac{1}{2}n\pi)}{4^2} + \frac{2[\cos(\frac{2}{5}n\pi) + \cos(\frac{4}{5}n\pi)]}{5^2} + \ldots\right] \tag{5}$$

(Hardy 1959). It also satisfies the INEQUALITY

$$\frac{\sigma(n)}{n \ln\ln n} \leq e^\gamma + \frac{2(1-\sqrt{2}) + \gamma - \ln(4\pi)}{\sqrt{\ln n}\,\ln\ln n} + \mathcal{O}\left(\frac{1}{\sqrt{\ln n}\,(\ln\ln n)^2}\right), \tag{6}$$

where γ is the EULER-MASCHERONI CONSTANT (Robin 1984, Erdős 1989).

Let a number n have PRIME factorization

$$n = \prod_{j=1}^r {p_j}^{\alpha_j}, \tag{7}$$

then

$$\sigma(n) = \prod_{j=1}^r \frac{{p_j}^{\alpha_j+1} - 1}{p_j - 1} \tag{8}$$

(Berndt 1985). GRONWALL'S THEOREM states that

$$\overline{\lim_{n\to\infty}} \frac{\sigma(n)}{n \ln\ln n} = e^\gamma, \tag{9}$$

where γ is the EULER-MASCHERONI CONSTANT.

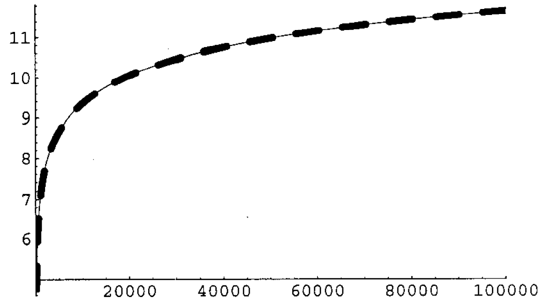

In general,

$$\sigma_k(n) \equiv \sum_{d|n} d^k. \tag{10}$$

In 1838, Dirichlet showed that the average number of DIVISORS of all numbers from 1 to n is asymptotic to

$$\frac{\sum_{i=1}^n \sigma_0(i)}{n} \sim \ln n + 2\gamma - 1 \tag{11}$$

(Conway and Guy 1996), as illustrated above, where the thin solid curve plots the actual values and the thick dashed curve plots the asymptotic function.

A curious identity derived using MODULAR FORM theory is given by

$$\sigma_7(n) = \sigma_3(n) + 120 \sum_{k=1}^{n-1} \sigma_3(k)\sigma_3(n-k). \tag{12}$$

The asymptotic SUMMATORY FUNCTION of $\sigma_0(n) = \Omega(n)$ is given by

$$\sum_{k=2}^n \Omega(k) = n \ln\ln n + B_2 + o(n), \tag{13}$$

where

$$B_2 = \gamma + \sum_{p \text{ prime}} \left[\ln(1 - p^{-1}) + \frac{1}{p-1}\right] \approx 1.034653 \tag{14}$$

(Hardy and Wright 1979, p. 355). This is related to the DIRICHLET DIVISOR PROBLEM. The SUMMATORY FUNCTIONS for σ_a with $a > 1$ are

$$\sum_{k=1}^n \sigma_a(k) = \frac{\zeta(a+1)}{a+1} n^{a+1} + \mathcal{O}(n^a). \tag{15}$$

For $a = 1$,

$$\sum_{k=1}^n \sigma_1(k) = \frac{\pi^2}{12} n^2 + \mathcal{O}(n \ln n). \tag{16}$$

The divisor function is ODD IFF n is a SQUARE NUMBER or twice a SQUARE NUMBER. The divisor function satisfies the CONGRUENCE

$$n\sigma(n) \equiv 2 \pmod{\phi(n)}, \tag{17}$$

for all PRIMES and no COMPOSITE NUMBERS with the exception of 4, 6, and 22 (Subbarao 1974). $\tau(n)$ is PRIME whenever $\sigma(n)$ is (Honsberger 1991). Factorizations of $\sigma(p^a)$ for PRIME p are given by Sorli.

see also DIRICHLET DIVISOR PROBLEM, DIVISOR, FACTOR, GREATEST PRIME FACTOR, GRONWALL'S THEOREM, LEAST PRIME FACTOR, MULTIPLY PERFECT

NUMBER, ORE'S CONJECTURE, PERFECT NUMBER, $r(n)$, RESTRICTED DIVISOR FUNCTION, SILVERMAN CONSTANT, TAU FUNCTION, TOTIENT FUNCTION, TOTIENT VALENCE FUNCTION, TWIN PEAKS

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Divisor Functions." §24.3.3 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 827, 1972.

Berndt, B. C. *Ramanujan's Notebooks: Part I.* New York: Springer-Verlag, p. 94, 1985.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 260–261, 1996.

Dickson, L. E. *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, pp. 279–325, 1952.

Dirichlet, G. L. "Sur l'usage des séries infinies dans la théorie des nombres." *J. reine angew. Math.* **18**, 259–274, 1838.

Erdős, P. "Ramanujan and I." In *Proceedings of the International Ramanujan Centenary Conference held at Anna University, Madras, Dec. 21, 1987.* (Ed. K. Alladi). New York: Springer-Verlag, pp. 1–20, 1989.

Guy, R. K. "Solutions of $m\sigma(m) = n\sigma(n)$," "Analogs with $d(n)$, $\sigma_k(n)$," "Solutions of $\sigma(n) = \sigma(n+1)$," and "Solutions of $\sigma(q) + \sigma(r) = \sigma(q+r)$." §B11, B12, B13 and B15 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 67–70, 1994.

Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, p. 141, 1959.

Hardy, G. H. and Weight, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Oxford University Press, pp. 354–355, 1979.

Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., pp. 250–251, 1991.

Robin, G. "Grandes valeurs de la fonction somme des diviseurs et hypothese de Riemann." *J. Math. Pures Appl.* **63**, 187–213, 1984.

Sloane, N. J. A. Sequences A000005/M0246, A000203/M2329, A001065/M2226, A001157/M3799, A001158/M4605 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sorli, R. "Factorization Tables." `http://www.maths.uts.edu.au/staff/ron/fact/fact.html`.

Subbarao, M. V. "On Two Congruences for Primality." *Pacific J. Math.* **52**, 261–268, 1974.

Divisor Theory

A generalization by Kronecker of Kummer's theory of PRIME IDEAL factors. A divisor on a full subcategory C of mod(A) is an additive mapping χ on C with values in a SEMIGROUP of IDEALS on A.

see also IDEAL, IDEAL NUMBER, PRIME IDEAL, SEMIGROUP

References

Edwards, H. M. *Divisor Theory.* Boston, MA: Birkhäuser, 1989.

Vasconcelos, W. V. *Divisor Theory in Module Categories.* Amsterdam, Netherlands: North-Holland, pp. 63–64, 1974.

Dixon's Factorization Method

In order to find INTEGERS x and y such that

$$x^2 \equiv y^2 \pmod{n} \tag{1}$$

(a modified form of FERMAT'S FACTORIZATION METHOD), in which case there is a 50% chance that $\mathrm{GCD}(n, x-y)$ is a FACTOR of n, choose a RANDOM INTEGER r_i, compute

$$g(r_i) \equiv {r_i}^2 \pmod{n}, \tag{2}$$

and try to factor $g(r_i)$. If $g(r_i)$ is not easily factorable (up to some small trial divisor d), try another r_i. In practice, the trial rs are usually taken to be $\lfloor\sqrt{n}\rfloor + k$, with $k = 1, 2, \ldots$, which allows the QUADRATIC SIEVE FACTORIZATION METHOD to be used. Continue finding and factoring $g(r_i)$s until $N \equiv \pi d$ are found, where π is the PRIME COUNTING FUNCTION. Now for each $g(r_i)$, write

$$g(r_i) = {p_{1i}}^{a_{1i}} {p_{2i}}^{a_{2i}} \cdots {p_{Ni}}^{a_{Ni}}, \tag{3}$$

and form the EXPONENT VECTOR

$$\mathbf{v}(r_i) = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{Ni} \end{bmatrix}. \tag{4}$$

Now, if a_{ki} are even for any k, then $g(r_i)$ is a SQUARE NUMBER and we have found a solution to (1). If not, look for a linear combination $\sum_i c_i \mathbf{v}(r_i)$ such that the elements are all even, i.e.,

$$c_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{N1} \end{bmatrix} + c_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{N2} \end{bmatrix} + \ldots + c_N \begin{bmatrix} a_{1N} \\ a_{2N} \\ \vdots \\ a_{NN} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \pmod{2} \tag{5}$$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \pmod{2}. \tag{6}$$

Since this must be solved only mod 2, the problem can be simplified by replacing the a_{ij}s with

$$b_{ij} = \begin{cases} 0 & \text{for } a_{ij} \text{ even} \\ 1 & \text{for } a_{ij} \text{ odd.} \end{cases} \tag{7}$$

GAUSSIAN ELIMINATION can then be used to solve

$$\mathbf{bc} = \mathbf{z} \tag{8}$$

for $\mathbf{c}$, where $\mathbf{z}$ is a VECTOR equal to $\mathbf{0}$ (mod 2). Once $\mathbf{c}$ is known, then we have

$$\prod_k g(r_k) \equiv \prod_k {r_k}^2 \pmod{n}, \tag{9}$$

where the products are taken over all k for which $c_k = 1$. Both sides are PERFECT SQUARES, so we have a 50% chance that this yields a nontrivial factor of n. If it does not, then we proceed to a different $\mathbf{z}$ and repeat the procedure. There is no guarantee that this method will yield a factor, but in practice it produces factors faster than any method using trial divisors. It is especially amenable to parallel processing, since each processor can work on a different value of r.

References
Bressoud, D. M. *Factorization and Prime Testing.* New York: Springer-Verlag, pp. 102–104, 1989.
Dixon, J. D. "Asymptotically Fast Factorization of Integers." *Math. Comput.* **36**, 255–260, 1981.
Lenstra, A. K. and Lenstra, H. W. Jr. "Algorithms in Number Theory." In *Handbook of Theoretical Computer Science, Volume A: Algorithms and Complexity* (Ed. J. van Leeuwen). New York: Elsevier, pp. 673–715, 1990.
Pomerance, C. "A Tale of Two Sieves." *Not. Amer. Math. Soc.* **43**, 1473–1485, 1996.

Dixon-Ferrar Formula

Let $J_\nu(z)$ be a BESSEL FUNCTION OF THE FIRST KIND, $Y_\nu(z)$ a BESSEL FUNCTION OF THE SECOND KIND, and $K_\nu(z)$ a MODIFIED BESSEL FUNCTION OF THE FIRST KIND. Also let $\Re[z] > 0$ and $|\Re[z]| < 1/2$. Then

$$J_\nu^2(z) + Y_\nu^2(z) = \frac{8\cos(\nu\pi)}{\pi^2} \int_0^\infty K_{2\nu}(2z\sinh t)\,dt.$$

see also NICHOLSON'S FORMULA, WATSON'S FORMULA

References
Gradshteyn, I. S. and Ryzhik, I. M. Eqn. 6.518 in *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 671, 1979.
Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1476, 1980.

Dixon's Random Squares Factorization Method

see DIXON'S FACTORIZATION METHOD

Dixon's Theorem

$$\begin{aligned} &{}_3F_2\left[\begin{matrix} n, -x, -y \\ x+n+1, y+n+1 \end{matrix}\right] \\ &= \Gamma(x+n+1)\Gamma(y+n+1)\Gamma(\tfrac{1}{2}n+1)\Gamma(x+y+\tfrac{1}{2}n+1) \\ &\times\Gamma(n+1)\Gamma(x+y+n+1)\Gamma(x+\tfrac{1}{2}n+1)\Gamma(y+\tfrac{1}{2}n+1), \end{aligned}$$

where ${}_3F_2(a,b,c;d,e;z)$ is a GENERALIZED HYPERGEOMETRIC FUNCTION and $\Gamma(z)$ is the GAMMA FUNCTION. It can be derived from the DOUGALL-RAMANUJAN IDENTITY. It can be written more symmetrically as

$${}_3F_2(a,b,c;d,e;1) = \frac{(\tfrac{1}{2}a)!(a-b)!(a-c)!(\tfrac{1}{2}a-b-c)!}{a!(\tfrac{1}{2}a-b)!(\tfrac{1}{2}a-c)!(a-b-c)!},$$

where $1 + a/2 - b - c$ has a positive REAL PART, $d = a - b + 1$, and $e = a - c + 1$. The identity can also be written as the beautiful symmetric sum

$$\sum_k (-1)^k \binom{a+b}{a+k}\binom{a+c}{c+k}\binom{b+c}{b+k} = \frac{(a+b+c)!}{a!b!c!}$$

(Petkovšek 1996).

see also DOUGALL-RAMANUJAN IDENTITY, GENERALIZED HYPERGEOMETRIC FUNCTION

References
Bailey, W. N. *Generalised Hypergeometric Series.* Cambridge, England: Cambridge University Press, 1935.
Cartier, P. and Foata, D. *Problèmes combinatoires de commutation et réarrangements.* New York: Springer-Verlag, 1969.
Knuth, D. E. *The Art of Computer Programming, Vol. 1: Fundamental Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, 1973.
Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, p. 43, 1996.
Zeilberger, D. and Bressoud, D. "A Proof of Andrew's q-Dyson Conjecture." *Disc. Math.* **54**, 201–224, 1985.

Dobiński's Formula

Gives the nth BELL NUMBER,

$$B_n = \frac{1}{e}\sum_{k=0}^\infty \frac{k^n}{k!}. \tag{1}$$

It can be derived by dividing the formula for a STIRLING NUMBER OF THE SECOND KIND by $m!$, yielding

$$\frac{m^n}{m!} = \sum_{k=1}^m \left\{ \begin{matrix} n \\ k \end{matrix} \right\} \frac{1}{(m-k)!}. \tag{2}$$

Then

$$\sum_{m=1}^\infty \frac{m^n}{m!}\lambda^m = \left(\sum_{k=1}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\} \lambda^k\right)\left(\sum_{k=0}^\infty \frac{\lambda^j}{j!}\right), \tag{3}$$

and

$$\sum_{k=1}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\} \lambda^k = e^{-\lambda}\sum_{m=1}^\infty \frac{m^n}{m!}\lambda^m. \tag{4}$$

Now setting $\lambda = 1$ gives the identity (Dobiński 1877; Rota 1964; Berge 1971, p. 44; Comtet 1974, p. 211; Roman 1984, p. 66; Lupas 1988; Wilf 1990, p. 106; Chen and Yeh 1994; Pitman 1997).

References
Berge, C. *Principles of Combinatorics.* New York: Academic Press, 1971.
Chen, B. and Yeh, Y.-N. "Some Explanations of Dobinski's Formula." *Studies Appl. Math.* **92**, 191–199, 1994.
Comtet, L. *Advanced Combinatorics.* Boston, MA: Reidel, 1974.
Dobiński, G. "Summierung der Reihe $\sum n^m/n!$ für $m = 1$, 2, 3, 4, 5, ..." *Grunert Archiv (Arch. Math. Phys.)* **61**, 333–336, 1877.

Foata, D. *La série génératrice exponentielle dans les problèmes d'énumération.* Vol. 54 of *Séminaire de Mathématiques supérieures.* Montréal, Canada: Presses de l'Université de Montréal, 1974.
Lupas, A. "Dobiński-Type Formula for Binomial Polynomials." *Stud. Univ. Babes-Bolyai Math.* **33**, 30–44, 1988.
Pitman, J. "Some Probabilistic Aspects of Set Partitions." *Amer. Math. Monthly* **104**, 201–209, 1997.
Roman, S. *The Umbral Calculus.* New York: Academic Press, 1984.
Rota, G.-C. "The Number of Partitions of a Set." *Amer. Math. Monthly* **71**, 498–504, 1964.
Wilf, H. *Generatingfunctionology, 2nd ed.* San Diego, CA: Academic Press, 1990.

Dodecadodecahedron

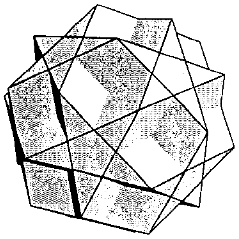

The UNIFORM POLYHEDRON U_{36} whose DUAL POLYHEDRON is the MEDIAL RHOMBIC TRIACONTAHEDRON. The solid is also called the GREAT DODECADODECAHEDRON, and its DUAL POLYHEDRON is also called the SMALL STELLATED TRIACONTAHEDRON. It can be obtained by TRUNCATING a GREAT DODECAHEDRON or FACETING a ICOSIDODECAHEDRON with PENTAGONS and covering remaining open spaces with PENTAGRAMS (Holden 1991, p. 103). A FACETED version is the GREAT DODECAHEMICOSAHEDRON. The dodecadodecahedron is an ARCHIMEDEAN SOLID STELLATION. The dodecadodecahedron has SCHLÄFLI SYMBOL $\{\frac{5}{2},5\}$ and WYTHOFF SYMBOL $2\,|\,\frac{5}{2}\,5$. Its faces are $12\{\frac{5}{2}\}+12\{5\}$, and its CIRCUMRADIUS for unit edge length is

$$R = 1.$$

References
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 123, 1989.
Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, 1991.
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 112, 1989.

Dodecagon

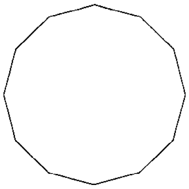

The constructible regular 12-sided POLYGON with SCHLÄFLI SYMBOL {12}. The INRADIUS r, CIRCUMRADIUS R, and AREA A can be computed directly from the formulas for a general regular POLYGON with side length s and $n = 12$ sides,

$$r = \tfrac{1}{2}s\cot\left(\frac{\pi}{12}\right) = \tfrac{1}{2}(2+\sqrt{3})s \qquad (1)$$

$$R = \tfrac{1}{2}s\cot\left(\frac{\pi}{12}\right) = \tfrac{1}{2}(\sqrt{2}+\sqrt{6})s \qquad (2)$$

$$A = \tfrac{1}{4}ns^2\cot\left(\frac{\pi}{12}\right) = 3(2+\sqrt{3})s^2. \qquad (3)$$

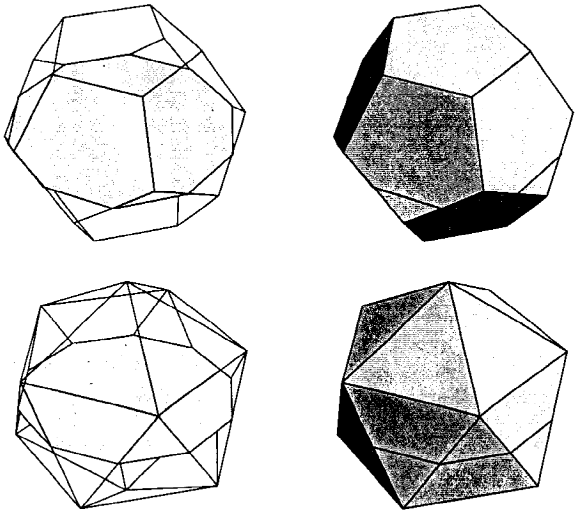

A PLANE PERPENDICULAR to a C_5 axis of a DODECAHEDRON or ICOSAHEDRON cuts the solid in a regular DECAGONAL CROSS-SECTION (Holden 1991, pp. 24–25).

The GREEK, LATIN, and MALTESE CROSSES are all irregular dodecagons.

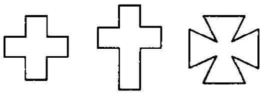

see also DECAGON, DODECAGRAM, DODECAHEDRON, GREEK CROSS, LATIN CROSS, MALTESE CROSS, TRIGONOMETRY VALUES—$\pi/12$, UNDECAGON

References
Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, 1991.

Dodecagram

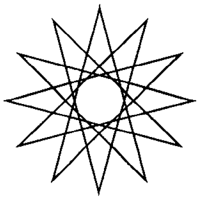

The STAR POLYGON $\left\{\frac{12}{5}\right\}$.

see also STAR POLYGON, TRIGONOMETRY VALUES—$\pi/12$

Dodecahedral Conjecture

In any unit SPHERE PACKING, the volume of any VORONOI CELL around any sphere is at least as large as a regular DODECAHEDRON of INRADIUS 1. If true, this would provide a bound on the densest possible sphere packing greater than any currently known. It would not, however, be sufficient to establish the KEPLER CONJECTURE.

Dodecahedral Graph

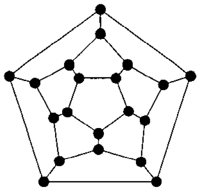

A POLYHEDRAL GRAPH.

see also CUBICAL GRAPH, ICOSAHEDRAL GRAPH, OCTAHEDRAL GRAPH, TETRAHEDRAL GRAPH

Dodecahedral Space

see POINCARÉ MANIFOLD

Dodecahedron

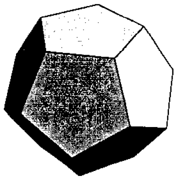

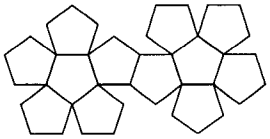

The regular dodecahedron is the PLATONIC SOLID (P_4) composed of 20 VERTICES, 30 EDGES, and 12 PENTAGONAL FACES. It is given by the symbol 12{5}, the SCHLÄFLISYMBOL {5,3}. It is also UNIFORM POLYHEDRON U_{23} and has WYTHOFF SYMBOL $3\,|\,2\,5$. The dodecahedron has the ICOSAHEDRAL GROUP I_h of symmetries.

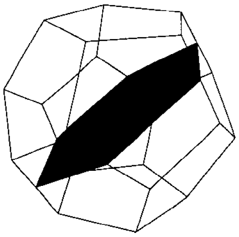

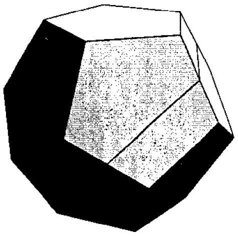

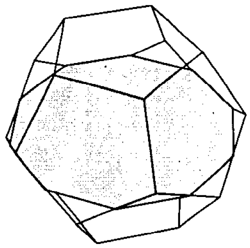

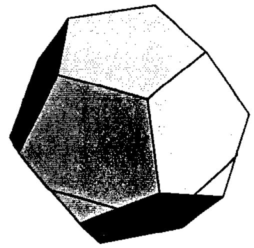

A PLANE PERPENDICULAR to a C_3 axis of a dodecahedron cuts the solid in a regular HEXAGONAL CROSS-SECTION (Holden 1991, p. 27). A PLANE PERPENDICULAR to a C_5 axis of a dodecahedron cuts the solid in a regular DECAGONAL CROSS-SECTION (Holden 1991, p. 24).

The DUAL POLYHEDRON of the dodecahedron is the ICOSAHEDRON.

When the dodecahedron with edge length $\sqrt{10-2\sqrt{5}}$ is oriented with two opposite faces parallel to the xy-PLANE, the vertices of the top and bottom faces lie at $z=\pm(\phi+1)$ and the other VERTICES lie at $z=\pm(\phi-1)$, where ϕ is the GOLDEN RATIO. The explicit coordinates are

$$\pm\left(2\cos(\tfrac{2}{5}\pi i), 2\sin(\tfrac{2}{5}\pi i), \phi+1\right) \qquad (1)$$

$$\pm\left(2\phi\cos(\tfrac{2}{5}\pi i), 2\phi\sin(\tfrac{2}{5}\pi i), \phi-1\right) \qquad (2)$$

with $i = 0, 1, \ldots, 4$, where ϕ is the GOLDEN RATIO. Explicitly, these coordinates are

$$\mathbf{x}_{10}^{\pm} = \pm(2, 0, \tfrac{1}{2}(3+\sqrt{5})) \qquad (3)$$

$$\mathbf{x}_{11}^{\pm} = \pm(\tfrac{1}{2}(\sqrt{5}-1), \tfrac{1}{2}\sqrt{10+2\sqrt{5}}, \tfrac{1}{2}(3+\sqrt{5})) \qquad (4)$$

$$\mathbf{x}_{12}^{\pm} = \pm(-\tfrac{1}{2}(1+\sqrt{5}), \tfrac{1}{2}\sqrt{10-2\sqrt{5}}, \tfrac{1}{2}(3+\sqrt{5})) \qquad (5)$$

$$\mathbf{x}_{13}^{\pm} = \pm(-\tfrac{1}{2}(1+\sqrt{5}), -\tfrac{1}{2}\sqrt{10-2\sqrt{5}}, \tfrac{1}{2}(3+\sqrt{5})) \qquad (6)$$

$$\mathbf{x}_{14}^{\pm} = \pm(\tfrac{1}{2}(\sqrt{5}-1), -\tfrac{1}{2}\sqrt{10+2\sqrt{5}}, \tfrac{1}{2}(3+\sqrt{5})) \qquad (7)$$

$$\mathbf{x}_{20}^{\pm} = \pm(1+\sqrt{5}, 0, \tfrac{1}{2}(\sqrt{5}-1)) \qquad (8)$$

$$\mathbf{x}_{21}^{\pm} = \pm(1, \sqrt{5+2\sqrt{5}}, \tfrac{1}{2}(\sqrt{5}-1)) \qquad (9)$$

$$\mathbf{x}_{22}^{\pm} = \pm(-\tfrac{1}{2}(3+\sqrt{5}), \tfrac{1}{2}\sqrt{10+2\sqrt{5}}, \tfrac{1}{2}(\sqrt{5}-1)) \qquad (10)$$

$$\mathbf{x}_{23}^{\pm} = \pm(-\tfrac{1}{2}(3+\sqrt{5}), -\tfrac{1}{2}\sqrt{10+2\sqrt{5}}, \tfrac{1}{2}(\sqrt{5}-1)) \qquad (11)$$

$$\mathbf{x}_{24}^{\pm} = \pm(1, -\sqrt{5+2\sqrt{5}}, \tfrac{1}{2}(\sqrt{5}-1)), \qquad (12)$$

where $\mathbf{x}_{1i}^{+}$ are the top vertices, $\mathbf{x}_{2i}^{+}$ are the vertices above the mid-plane, $\mathbf{x}_{2i}^{-}$ are the vertices below the mid-plane, and $\mathbf{x}_{2i}^{-}$ are the bottom vertices. The VERTICES of a dodecahedron can be given in a simple form for a dodecahedron of side length $a=\sqrt{5}-1$ by $(0, \pm\phi^{-1}, \pm\phi)$, $(\pm\phi, 0, \pm\phi^{-1})$, $(\pm\phi^{-1}, \pm\phi, 0)$, and $(\pm1, \pm1, \pm1)$.

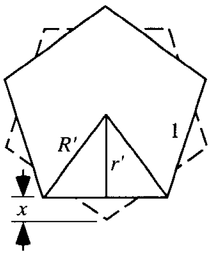

For a dodecahedron of unit edge length $a = 1$, the CIRCUMRADIUS R' and INRADIUS r' of a PENTAGONAL FACE are

$$R' = \tfrac{1}{10}\sqrt{50 + 10\sqrt{5}} \qquad (13)$$

$$r' = \tfrac{1}{10}\sqrt{25 + 10\sqrt{5}}. \qquad (14)$$

The SAGITTA x is then given by

$$x \equiv R' - r' = \tfrac{1}{10}\sqrt{125 - 10\sqrt{5}}. \qquad (15)$$

Now consider the following figure.

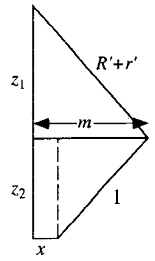

Using the PYTHAGOREAN THEOREM on the figure then gives

$$z_1{}^2 + m^2 = (R' + r)^2 \qquad (16)$$

$$z_2{}^2 + (m - x)^2 = 1 \qquad (17)$$

$$\left(\frac{z_1 + z_2}{2}\right)^2 + R'^2 = \left(\frac{z_1 - z_2}{2}\right)^2 + (m + r')^2. \qquad (18)$$

Equation (18) can be written

$$z_1 z_2 + r^2 = (m + r')^2. \qquad (19)$$

Solving (16), (17), and (19) simultaneously gives

$$m = r' = \tfrac{1}{10}\sqrt{25 + 10\sqrt{5}} \qquad (20)$$

$$z_1 = 2r' = \tfrac{1}{5}\sqrt{25 + 10\sqrt{5}} \qquad (21)$$

$$z_2 = R' = \tfrac{1}{10}\sqrt{50 + 10\sqrt{5}}. \qquad (22)$$

The INRADIUS of the dodecahedron is then given by

$$r = \tfrac{1}{2}(z_1 + z_2), \qquad (23)$$

so

$$\begin{aligned} r^2 &= \tfrac{1}{4}\left(\tfrac{1}{10}\sqrt{50 + 10\sqrt{5}} + \tfrac{1}{5}\sqrt{25 + 10\sqrt{5}}\right)^2 \\ &= \tfrac{1}{40}(25 + 11\sqrt{5}), \qquad (24) \end{aligned}$$

and

$$r = \sqrt{\frac{25 + 11\sqrt{5}}{40}} = \tfrac{1}{20}\sqrt{250 + 110\sqrt{5}} = 1.11351\ldots. \qquad (25)$$

Now,

$$\begin{aligned} R^2 &= R'^2 + r^2 = [\tfrac{1}{100}(50 + 10\sqrt{5}) + \tfrac{1}{400}(250 + 110\sqrt{5})] \\ &= \tfrac{3}{8}(3 + \sqrt{5}), \qquad (26) \end{aligned}$$

and the CIRCUMRADIUS is

$$R = a\sqrt{\tfrac{3}{8}(3 + \sqrt{5})} = \tfrac{1}{4}(\sqrt{15} + \sqrt{3}) = 1.40125\ldots. \qquad (27)$$

The INTERRADIUS is given by

$$\begin{aligned} \rho^2 &= r'^2 + r^2 = [\tfrac{1}{100}(25 + 10\sqrt{5}) + \tfrac{1}{400}(250 + 110\sqrt{5})] \\ &= \tfrac{1}{8}(7 + 3\sqrt{5}), \qquad (28) \end{aligned}$$

so

$$\rho = \tfrac{1}{4}(3 + \sqrt{5}) = 1.30901\ldots. \qquad (29)$$

The AREA of a single FACE is the AREA of a PENTAGON,

$$A = \tfrac{1}{4}\sqrt{25 + 10\sqrt{5}}. \qquad (30)$$

The VOLUME of the dodecahedron can be computed by summing the volume of the 12 constituent PENTAGONAL PYRAMIDS,

$$\begin{aligned} V &= 12(\tfrac{1}{3}Ar) \\ &= 12(\tfrac{1}{3})(\tfrac{1}{4}\sqrt{25 + 10\sqrt{5}})(\tfrac{1}{20}\sqrt{250 + 110\sqrt{5}}) \\ &= \tfrac{1}{20}(75 + 35\sqrt{5}) = \tfrac{1}{4}(15 + 7\sqrt{5}). \qquad (31) \end{aligned}$$

Apollonius showed that the VOLUME V and SURFACE AREA A of the dodecahedron and its DUAL the ICOSAHEDRON are related by

$$\frac{V_{\text{icosahedron}}}{V_{\text{dodecahedron}}} = \frac{A_{\text{icosahedron}}}{A_{\text{dodecahedron}}}. \qquad (32)$$

The HEXAGONAL SCALENOHEDRON is an irregular dodecahedron.

see also AUGMENTED DODECAHEDRON, AUGMENTED TRUNCATED DODECAHEDRON, DODECAGON, DODECAHEDRON-ICOSAHEDRON COMPOUND, ELONGATED DODECAHEDRON, GREAT DODECAHEDRON, GREAT STELLATED DODECAHEDRON, HYPERBOLIC DODECAHEDRON, ICOSAHEDRON, METABIAUGMENTED DODECAHEDRON, METABIAUGMENTED TRUNCATED DODECAHEDRON, PARABIAUGMENTED DODECAHEDRON, PARABIAUGMENTED TRUNCATED DODECAHEDRON, PYRITOHEDRON, RHOMBIC DODECAHEDRON, SMALL STELLATED DODECAHEDRON, TRIAUGMENTED DODECAHEDRON, TRIAUGMENTED TRUNCATED DODECAHEDRON, TRIGONAL DODECAHEDRON, TRIGONOMETRY VALUES—$\pi/5$ TRUNCATED DODECAHEDRON

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., 1989.

Davie, T. "The Dodecahedron." `http://www.dcs.st-and.ac.uk/~ad/mathrecs/polyhedra/dodecahedron.html`.

Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, 1991.

Dodecahedron-Icosahedron Compound

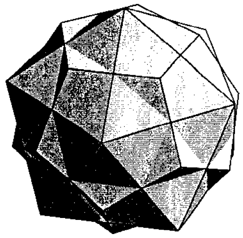

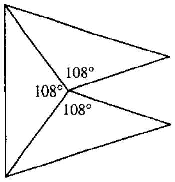

A POLYHEDRON COMPOUND of a DODECAHEDRON and ICOSAHEDRON which is most easily constructed by adding 20 triangular PYRAMIDS, constructed as above, to an ICOSAHEDRON. In the compound, the DODECAHEDRON and ICOSAHEDRON are rotated $\pi/5$ radians with respect to each other, and the ratio of the ICOSAHEDRON to DODECAHEDRON edges lengths are the GOLDEN RATIO ϕ.

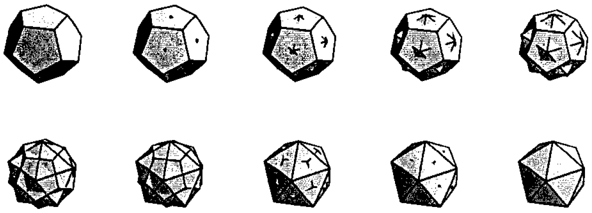

The above figure shows compounds composed of a DODECAHEDRON of unit edge length and ICOSAHEDRA having edge lengths varying from $\sqrt{5}/2$ (inscribed in the dodecahedron) to 2 (circumscribed about the dodecahedron).

The intersecting edges of the compound form the DIAGONALS of 30 RHOMBUSES comprising the TRIACONTAHEDRON, which is the the DUAL POLYHEDRON of the ICOSIDODECAHEDRON (Ball and Coxeter 1987). The dodecahedron-icosahedron is the first STELLATION of the ICOSIDODECAHEDRON.

see also DODECAHEDRON, ICOSAHEDRON, ICOSIDODECAHEDRON, POLYHEDRON COMPOUND

References
Cundy, H. and Rollett, A. *Mathematical Models, 2nd ed.* Stradbroke, England: Tarquin Pub., p. 131, 1989.
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 76, 1989.

Dodecahedron Stellations

The dodecahedron has three STELLATIONS: the GREAT DODECAHEDRON, GREAT STELLATED DODECAHEDRON, and SMALL STELLATED DODECAHEDRON. The only STELLATIONS of PLATONIC SOLIDS which are UNIFORM POLYHEDRA are these three and one ICOSAHEDRON STELLATION. Bulatov has produced 270 stellations of a deformed dodecahedron.

see also ICOSAHEDRON STELLATIONS, STELLATED POLYHEDRON, STELLATION

References
Bulatov, V.v "270 Stellations of Deformed Dodecahedron." http://www.physics.orst.edu/~bulatov/polyhedra/dodeca270/.

Dodecahedron 2-Compound

A compound of two dodecahedra with the symmetry of the CUBE arises by combining the two dodecahedra rotated 90° with respect to each other about a common C_2 axis (Holden 1991, p. 37).

see also POLYHEDRON COMPOUND

References
Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, 1991.

Domain

A connected OPEN SET. The term domain is also used to describe the set of values D for which a FUNCTION is defined. The set of values to which D is sent by the function (MAP) is then called the RANGE.

see also MAP, ONE-TO-ONE, ONTO, RANGE (IMAGE), REINHARDT DOMAIN

Domain Invariance Theorem

The Invariance of Domain Theorem is that if $f : A \to \mathbb{R}^n$ is a ONE-TO-ONE continuous MAP from A, a compact subset of $\mathbb{R}^n$, then the interior of A is mapped to the interior of $f(A)$.

see also DIMENSION INVARIANCE THEOREM

Dome

see BOHEMIAN DOME, GEODESIC DOME, HEMISPHERE, SPHERICAL CAP, TORISPHERICAL DOME, VAULT

Dominance

The dominance RELATION on a SET of points in EUCLIDEAN n-space is the INTERSECTION of the n coordinate-wise orderings. A point p dominates a point q provided that every coordinate of p is at least as large as the corresponding coordinate of q.

The dominance orders in $\mathbb{R}^n$ are precisely the POSETS of DIMENSION at most n.

see also PARTIALLY ORDERED SET, REALIZER

Domino

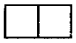

The unique 2-POLYOMINO consisting of two equal squares connected along a complete EDGE.

The FIBONACCI NUMBER F_{n+1} gives the number of ways for 2×1 dominoes to cover a $2 \times n$ CHECKERBOARD, as illustrated in the following diagrams (Dickau).

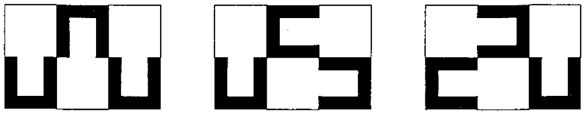

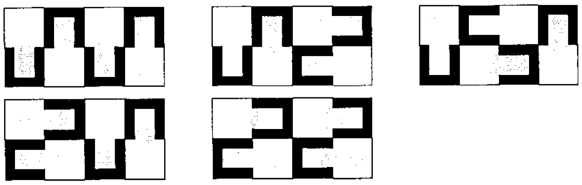

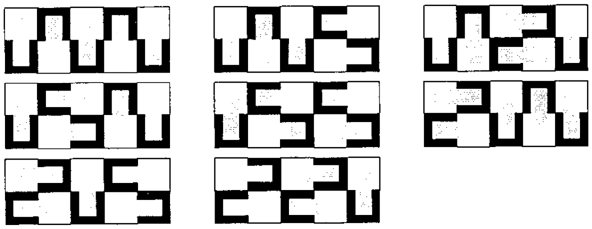

see also FIBONACCI NUMBER, GOMORY'S THEOREM, HEXOMINO, PENTOMINO, POLYOMINO, TETROMINO, TRIOMINO

References

Dickau, R. M. "Fibonacci Numbers." `http://www.prairienet.org/~pops/fibboard.html`.

Gardner, M. "Polyominoes." Ch. 13 in *The Scientific American Book of Mathematical Puzzles & Diversions.* New York: Simon and Schuster, pp. 124–140, 1959.

Kraitchik, M. "Dominoes." §12.1.22 in *Mathematical Recreations.* New York: W. W. Norton, pp. 298–302, 1942.

Lei, A. "Domino." `http://www.cs.ust.hk/~philipl/omino/domino.html`.

Madachy, J. S. "Domino Recreations." *Madachy's Mathematical Recreations.* New York: Dover, pp. 209–219, 1979.

Domino Problem

see WANG'S CONJECTURE

Donaldson Invariants

Distinguish between smooth MANIFOLDS in 4-D.

Donkin's Theorem

The product of three translations along the directed sides of a TRIANGLE through twice the lengths of these sides is the identity.

Donut

see TORUS

Doob's Theorem

A theorem proved by Doob (1942) which states that any random process which is both GAUSSIAN and MARKOV has the following forms for its correlation function, spectral density, and probability densities:

$$C_y(\tau) = {\sigma_y}^2 e^{-\tau/\tau_r}$$

$$G_y(f) = \frac{4{\tau_r}^{-1}{\sigma_y}^2}{(2\pi f)^2 + {\tau_r}^{-2}}$$

$$p_1(y) = \frac{1}{\sqrt{2\pi{\sigma_y}^2}} e^{-(y-\bar{y})^2/2{\sigma_y}^2}$$

$$p_2(y_1|y_2,\tau) = \frac{1}{\sqrt{2\pi(1-e^{-2\tau/\tau_r}){\sigma_y}^2}} \times \exp\left\{-\frac{[(y_2-\bar{y}) - e^{-\tau/\tau_r}(y_1-\bar{y})]^2}{2(1-e^{-2\tau/\tau_r}){\sigma_y}^2}\right\},$$

where $\bar{y}$ is the MEAN, σ_y the STANDARD DEVIATION, and τ_r the relaxation time.

References

Doob, J. L. "Topics in the Theory of Markov Chains." *Trans. Amer. Math. Soc.* **52**, 37–64, 1942.

Dot

The "dot" $\cdot$ has several meanings in mathematics, including MULTIPLICATION ($a \cdot b$ is pronounced "a times b"), computation of a DOT PRODUCT ($\mathbf{a}\cdot\mathbf{b}$ is pronounced "**a** dot **b**"), or computation of a time DERIVATIVE ($\dot{a}$ is pronounced "a dot").

see also DERIVATIVE, DOT PRODUCT, TIMES

Dot Product

The dot product can be defined by

$$\mathbf{X}\cdot\mathbf{Y} = |\mathbf{X}|\,|\mathbf{Y}|\cos\theta, \tag{1}$$

where θ is the angle between the vectors. It follows immediately that $\mathbf{X}\cdot\mathbf{Y} = 0$ if $\mathbf{X}$ is PERPENDICULAR to $\mathbf{Y}$. The dot product is also called the INNER PRODUCT and written $\langle a, b\rangle$. By writing

$$A_x = A\cos\theta_A \qquad B_x = B\cos\theta_B \tag{2}$$

$$A_y = A\sin\theta_A \qquad B_y = B\sin\theta_B, \tag{3}$$

it follows that (1) yields

$$\begin{aligned}\mathbf{A}\cdot\mathbf{B} &= AB\cos(\theta_A - \theta_B)\\ &= AB(\cos\theta_A\cos\theta_B + \sin\theta_A\sin\theta_B)\\ &= A\cos\theta_A B\cos\theta_B + A\sin\theta_A B\sin\theta_B\\ &= A_xB_x + A_yB_y.\end{aligned} \tag{4}$$

So, in general,

$$\mathbf{X}\cdot\mathbf{Y} = x_1y_1 + \ldots + x_ny_n. \tag{5}$$

The dot product is COMMUTATIVE

$$\mathbf{X}\cdot\mathbf{Y} = \mathbf{Y}\cdot\mathbf{X}, \tag{6}$$

ASSOCIATIVE

$$(r\mathbf{X})\cdot\mathbf{Y} = r(\mathbf{X}\cdot\mathbf{Y}), \tag{7}$$

and DISTRIBUTIVE

$$\mathbf{X}\cdot(\mathbf{Y}+\mathbf{Z}) = \mathbf{X}\cdot\mathbf{Y} + \mathbf{X}\cdot\mathbf{Z}. \tag{8}$$

The DERIVATIVE of a dot product of VECTORS is

$$\frac{d}{dt}[\mathbf{r}_1(t)\cdot\mathbf{r}_2(t)] = \mathbf{r}_1(t)\cdot\frac{d\mathbf{r}_2}{dt} + \frac{d\mathbf{r}_1}{dt}\cdot\mathbf{r}_2(t). \tag{9}$$

The dot product is invariant under rotations

$$\begin{aligned}\mathbf{A}' \cdot \mathbf{B}' = A'_i B'_i = a_{ij} A_j a_{ik} B_k = (a_{ij} a_{ik}) A_j B_k \\ = \delta_{jk} A_j B_k = A_j B_j = \mathbf{A} \cdot \mathbf{B}, \qquad (10)\end{aligned}$$

where EINSTEIN SUMMATION has been used.

The dot product is also defined for TENSORS A and B by

$$A \cdot B \equiv A^\alpha B_\alpha. \qquad (11)$$

see also CROSS PRODUCT, INNER PRODUCT, OUTER PRODUCT, WEDGE PRODUCT

References
Arfken, G. "Scalar or Dot Product." §1.3 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 13–18, 1985.

Douady's Rabbit Fractal

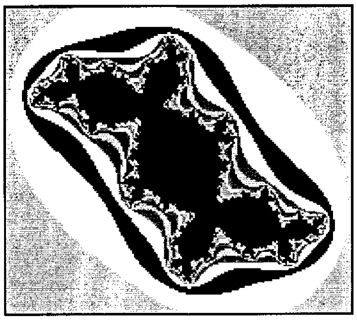

A JULIA SET with $c = -0.123 + 0.745i$, also known as the DRAGON FRACTAL.

see also SAN MARCO FRACTAL, SIEGEL DISK FRACTAL

References
Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, p. 176, 1991.

Double Bubble

The planar double bubble (three circular arcs meeting in two points at equal 120° ANGLES) has the minimum PERIMETER for enclosing two equal areas (Foisy 1993, Morgan 1995).

see also APPLE, BUBBLE, DOUBLE BUBBLE CONJECTURE, SPHERE-SPHERE INTERSECTION

References
Campbell, P. J. (Ed.). Reviews. *Math. Mag.* **68**, 321, 1995.
Foisy, J.; Alfaro, M.; Brock, J.; Hodges, N.; and Zimba, J. "The Standard Double Soap Bubble in $\mathbb{R}^2$ Uniquely Minimizes Perimeter." *Pacific J. Math.* **159**, 47–59, 1993.
Morgan, F. "The Double Bubble Conjecture." *FOCUS* **15**, 6–7, 1995.
Peterson, I. "Toil and Trouble over Double Bubbles." *Sci. News* **148**, 101, Aug. 12, 1995.

Double Bubble Conjecture

Two partial SPHERES with a separating boundary (which is planar for equal volumes) separate two volumes of air with less AREA than any other boundary. The planar case was proved true for equal volumes by J. Hass and R. Schlafy in 1995 by reducing the problem to a set of 200,260 integrals which they carried out on an ordinary PC.

see also DOUBLE BUBBLE

References
Haas, J. and Schlafy, R. "Double Bubbles Minimize." Preprint, 1995.

Double Contraction Relation

A TENSOR t is said to satisfy the double contraction relation when

$$t_{ij}^{m\,*} t_{ij}^n = \delta_{mn}.$$

This equation is satisfied by

$$\begin{aligned}\hat{t}^0 &= \frac{2\hat{\mathbf{z}}\hat{\mathbf{z}} - \hat{\mathbf{x}}\hat{\mathbf{x}} - \hat{\mathbf{y}}\hat{\mathbf{y}}}{\sqrt{6}} \\ \hat{t}^{\pm 1} &= \mp\tfrac{1}{2}(\hat{\mathbf{x}}\hat{\mathbf{z}} + \hat{\mathbf{z}}\hat{\mathbf{x}}) - \tfrac{1}{2}i(\hat{\mathbf{y}}\hat{\mathbf{z}} - \hat{\mathbf{z}}\hat{\mathbf{y}}) \\ \hat{t}^{\pm 2} &= \mp\tfrac{1}{2}(\hat{\mathbf{x}}\hat{\mathbf{x}} + \hat{\mathbf{y}}\hat{\mathbf{y}}) - \tfrac{1}{2}i(\hat{\mathbf{x}}\hat{\mathbf{y}} - \hat{\mathbf{y}}\hat{\mathbf{x}}),\end{aligned}$$

where the hat denotes zero trace, symmetric unit TENSORS. These TENSORS are used to define the SPHERICAL HARMONIC TENSOR.

see also SPHERICAL HARMONIC TENSOR, TENSOR

References
Arfken, G. "Alternating Series." *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 140, 1985.

Double Cusp

see DOUBLE POINT

Double Exponential Distribution

see FISHER-TIPPETT DISTRIBUTION, LAPLACE DISTRIBUTION

Double Exponential Integration

An excellent NUMERICAL INTEGRATION technique used by *Maple V R4*® (Waterloo Maple Inc.) for numerical computation of integrals.

see also INTEGRAL, INTEGRATION, NUMERICAL INTEGRATION

References
Davis, P. J. and Rabinowitz, P. *Methods of Numerical Integration, 2nd ed.* New York: Academic Press, p. 214, 1984.
Di Marco, G.; Favati, P.; Lotti, G.; and Romani, F. "Asymptotic Behaviour of Automatic Quadrature." *J. Complexity* **10**, 296–340, 1994.
Mori, M. *Developments in the Double Exponential Formula for Numerical Integration. Proceedings of the International Congress of Mathematicians, Kyoto 1990.* New York: Springer-Verlag, pp. 1585–1594, 1991.

Mori, M. and Ooura, T. "Double Exponential Formulas for Fourier Type Integrals with a Divergent Integrand." In *Contributions in Numerical Mathematics* (Ed. R. P. Agarwal). World Scientific Series in Applicable Analysis, Vol. 2, pp. 301–308, 1993.
Ooura, T. and Mori, M. "The Double Exponential Formula for Oscillatory Functions over the Half Infinite Interval." *J. Comput. Appl. Math.* **38**, 353–360, 1991.
Takahasi, H. and Mori, M. "Double Exponential Formulas for Numerical Integration." *Pub. RIMS Kyoto Univ.* **9**, 721–741, 1974.
Toda, H. and Ono, H. "Some Remarks for Efficient Usage of the Double Exponential Formulas." *Kokyuroku RIMS Kyoto Univ.* **339**, 74–109, 1978.

Double Factorial

The double factorial is a generalization of the usual FACTORIAL $n!$ defined by

$$n!! \equiv \begin{cases} n\cdot(n-2)\ldots 5\cdot 3\cdot 1 & n \text{ odd} \\ n\cdot(n-2)\ldots 6\cdot 4\cdot 2 & n \text{ even} \\ 1 & n=-1,0. \end{cases} \tag{1}$$

For $n = 0$, 1, 2, ..., the first few values are 1, 1, 2, 3, 8, 15, 48, 105, 384, ... (Sloane's A006882).

There are many identities relating double factorials to FACTORIALS. Since

$$\begin{aligned}&(2n+1)!!2^n n!\\ &= [(2n+1)(2n-1)\cdots 1][2n][2(n-1)][2(n-2)]\cdots 2(1)\\ &= [(2n+1)(2n-1)\cdots 1][2n(2n-2)(2n-4)\cdots 2]\\ &= (2n+1)(2n)(2n-1)(2n-2)(2n-3)(2n-4)\cdots 2(1)\\ &= (2n+1)!,\end{aligned} \tag{2}$$

it follows that $(2n+1)!! = \frac{(2n+1)!}{2^n n!}$. Since

$$\begin{aligned}(2n)!! &= (2n)(2n-2)(2n-4)\cdots 2\\ &= [2(n)][2(n-1)][2(n-2)]\cdots 2 = 2^n n!,\end{aligned} \tag{3}$$

it follows that $(2n)!! = 2^n n!$. Since

$$\begin{aligned}&(2n-1)!!2^n n!\\ &= [(2n-1)(2n-3)\cdots 1][2n][2(n-1)][2(n-2)]\cdots 2(1)\\ &= (2n-1)(2n-3)\cdots 1][2n(2n-2)(2n-4)\cdots 2]\\ &= 2n(2n-1)(2n-2)(2n-3)(2n-4)\cdots 2(1)\\ &= (2n)!,\end{aligned} \tag{4}$$

it follows that

$$(2n-1)!! = \frac{(2n)!}{2^n n!}. \tag{5}$$

Similarly, for $n = 0$, 1, ...,

$$(-2n-1)!! = \frac{(-1)^n}{(2n-1)!!} = \frac{(-1)n2^n n!}{(2n)!}. \tag{6}$$

For n ODD,

$$\begin{aligned}\frac{n!}{n!!} &= \frac{n(n-1)(n-2)\cdots(1)}{n(n-2)(n-4)\cdots(1)}\\ &= (n-1)(n-3)\cdots(1) = (n-1)!!.\end{aligned} \tag{7}$$

For n EVEN,

$$\begin{aligned}\frac{n!}{n!!} &= \frac{n(n-1)(n-2)\cdots(2)}{n(n-2)(n-4)\cdots(2)}\\ &= (n-1)(n-3)\cdots(2) = (n-1)!!.\end{aligned} \tag{8}$$

Therefore, for any n,

$$\frac{n!}{n!!} = (n-1)!! \tag{9}$$

$$n! = n!!(n-1)!!. \tag{10}$$

The FACTORIAL may be further generalized to the MULTIFACTORIAL

see also FACTORIAL, MULTIFACTORIAL

References
Sloane, N. J. A. Sequence A006882/M0876 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Double Folium

see BIFOLIUM

Double-Free Set

A SET of POSITIVE integers is double-free if, for any integer x, the SET $\{x, 2x\} \not\subset S$ (or equivalently, if $x \in S$ IMPLIES $2x \notin S$). Define

$$r(n) = \max\{S : S \subset \{1, 2, \ldots, n\} \text{ is double-free}\}.$$

Then an asymptotic formula is

$$r(n) \sim \tfrac{2}{3}n + \mathcal{O}(\ln n)$$

(Wang 1989).

see also TRIPLE-FREE SET

References
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/triple/triple.html`.
Wang, E. T. H. "On Double-Free Sets of Integers." *Ars Combin.* **28**, 97–100, 1989.

Double Gamma Function

see DIGAMMA FUNCTION

Double Point

A point traced out twice as a closed curve is traversed. The maximum number of double points for a nondegenerate QUARTIC CURVE is three. An ORDINARY DOUBLE POINT is called a NODE.

Arnold (1994) gives pictures of spherical and PLANE CURVES with up to five double points, as well as other curves.

see also BIPLANAR DOUBLE POINT, CONIC DOUBLE POINT, CRUNODE, CUSP, ELLIPTIC CONE POINT, GAUSS'S DOUBLE POINT THEOREM, NODE (ALGEBRAIC CURVE), ORDINARY DOUBLE POINT, QUADRUPLE POINT RATIONAL DOUBLE POINT, SPINODE, TACNODE, TRIPLE POINT, UNIPLANAR DOUBLE POINT

References

Aicardi, F. Appendix to "Plane Curves, Their Invariants, Perestroikas, and Classifications." In *Singularities & Bifurcations* (V. I. Arnold). Providence, RI: Amer. Math. Soc., pp. 80–91, 1994.

Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, pp. 12–13, 1986.

Double Sixes

Two sextuples of SKEW LINES on the general CUBIC SURFACE such that each line of one is SKEW to one LINE in the other set. Discovered by Schläfli.

see also BOXCARS, CUBIC SURFACE, SOLOMON'S SEAL LINES

References

Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, p. 11, 1986.

Double Sum

A nested sum over two variables. Identities involving double sums include the following:

$$\sum_{p=0}^{\infty}\sum_{q=0}^{p} a_{q,p-q} = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} a_{n,m} = \sum_{r=0}^{\infty}\sum_{s=0}^{\lfloor r/2\rfloor} a_{s,r-2s}, \qquad (1)$$

where

$$\lfloor r/2\rfloor = \begin{cases} \frac{1}{2}r & r \text{ even} \\ \frac{1}{2}(r-1) & r \text{ odd} \end{cases} \qquad (2)$$

is the FLOOR FUNCTION, and

$$\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j = n^2 \left\langle x^2 \right\rangle. \qquad (3)$$

Consider the sum

$$S(a,b,c;s) = \sum_{(m,n)\neq(0,0)} (am^2 + bmn + cn^2)^{-s} \qquad (4)$$

over binary QUADRATIC FORMS. If S can be decomposed into a linear sum of products of DIRICHLET L-SERIES, it is said to be solvable. The related sums

$$S_1(a,b,c;s) = \sum_{(m,n)\neq(0,0)} (-1)^m (am^2 + bmn + cn^2)^{-s} \qquad (5)$$

$$S_2(a,b,c;s) = \sum_{(m,n)\neq(0,0)} (-1)^n (am^2 + bmn + cn^2)^{-s} \qquad (6)$$

$$S_{1,2}(a,b,c;s) = \sum_{(m,n)\neq(0,0)} (-1)^{m+n} (am^2 + bmn + cn^2)^{-s} \qquad (7)$$

can also be defined, which gives rise to such impressive FORMULAS as

$$S_1(1,0,58;1) = -\frac{\pi \ln(27 + 5\sqrt{29}\,)}{\sqrt{58}}. \qquad (8)$$

A complete table of the principal solutions of all solvable $S(a,b,c;s)$ is given in Glasser and Zucker (1980, pp. 126–131).

see also EULER SUM

References

Glasser, M. L. and Zucker, I. J. "Lattice Sums in Theoretical Chemistry." *Theoretical Chemistry: Advances and Perspectives, Vol. 5.* New York: Academic Press, 1980.

Zucker, I. J. and Robertson, M. M. "A Systematic Approach to the Evaluation of $\sum_{(m,n\neq 0,0)}(am^2+bmn+cn^2)^{-s}$." *J. Phys. A: Math. Gen.* **9**, 1215–1225, 1976.

Doublet Function

$$y = \delta'(x-a),$$

where $\delta(x)$ is the DELTA FUNCTION.

see also DELTA FUNCTION

References

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 324, 1993.

Doubly Even Number

An even number N for which $N \equiv 0 \pmod 4$. The first few POSITIVE doubly even numbers are 4, 8, 12, 16, ... (Sloane's A008586).

see also EVEN FUNCTION, ODD NUMBER, SINGLY EVEN NUMBER

References

Sloane, N. J. A. Sequence A008586 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Doubly Magic Square

see BIMAGIC SQUARE

Dougall-Ramanujan Identity

Discovered by Ramanujan around 1910. From Hardy (1959, pp. 102–103),

$$\begin{aligned}\sum_{n=0}^{\infty}(-1)^n(s+2n)\frac{s^{(n)}}{1^{(n)}}\frac{(x+y+z+u+2s+1)^{(n)}}{(x+y+z+u-s)_{(n)}}\\ \times\prod_{x,y,z,u}\frac{x_{(n)}}{(x+s+1)^{(n)}}\\ =\frac{s}{\Gamma(s+1)\Gamma(x+y+z+u+s+1)}\\ \times\prod_{x,y,z,u}\frac{\Gamma(x+s+1)\Gamma(y+z+u+s+1)}{\Gamma(z+u+s+1)},\end{aligned}\tag{1}$$

where

$$a^{(n)}\equiv a(a+1)\cdots(a+n-1)\tag{2}$$

$$a_{(n)}\equiv a(a-1)\cdots(a-n+1)\tag{3}$$

(here, the POCHHAMMER SYMBOL has been written $a^{(n)}$). This can be rewritten as

$$\begin{aligned}{}_7F_6\left[\begin{matrix}s, 1+\tfrac{1}{2}s, -x-y, -z, -u, x-y+z+u+2s+1\\ \tfrac{1}{2}s, x+s+1, y+s+1, z+s+1, u+s+1,\\ -x-y-z-u-s\end{matrix};1\right]\\ =\frac{1}{\Gamma(s+1)\Gamma(x+y+z+u+s+1)}\\ \times\prod_{x,t,z,u}\frac{\Gamma(x+s+1)\Gamma(y+z+u+s+1)}{\Gamma(z+u+s+1)}.\end{aligned}\tag{4}$$

In a more symmetric form, if $n=2a_1+1=a_2+a_3+a_4+a_5$, $a_6=1+a_1/2$, $a_7=-n$, and $b_i=1+a_1-a_{i+1}$ for $i=1, 2, \ldots, 6$, then

$$\begin{aligned}{}_7F_6\left[\begin{matrix}a_1, a_2, a_3, a_4, a_5, a_6, a_7\\ b_1, b_2, b_3, b_4, b_5, b_6\end{matrix};1\right]\\ \frac{(a_1+1)_n(a_1-a_2-a_3+1)_n}{(a_1-a_2+1)_n(a_1-a_3+1)_n}\\ \times\frac{(a_1-a_2-a_4+1)_n(a_1-a_3-a_4+1)_n}{(a_1-a_4+1)_n(a_1-a_2-a_3-a_4+1)_n},\end{aligned}\tag{5}$$

where $(a)_n$ is the POCHHAMMER SYMBOL (Petkovšek *et al.* 1996).

The identity is a special case of JACKSON'S IDENTITY.

see also DIXON'S THEOREM, DOUGALL'S THEOREM, GENERALIZED HYPERGEOMETRIC FUNCTION, HYPERGEOMETRIC FUNCTION, JACKSON'S IDENTITY, SAALSCHÜTZ'S THEOREM

References

Dixon, A. C. "Summation of a Certain Series." *Proc. London Math. Soc.* **35**, 285–289, 1903.

Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, 1959.

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, pp. 43, 126–127, and 183–184, 1996.

Dougall's Theorem

$$\begin{aligned}{}_5F_4\left[\begin{matrix}\tfrac{1}{2}n+1, n, -x, -y, -z\\ \tfrac{1}{2}n, x+n+1, y+n+1, z+n+1\end{matrix}\right]=\\ \frac{\Gamma(x+n+1)\Gamma(y+n+1)\Gamma(z+n+1)\Gamma(x+y+z+n+1)}{\Gamma(n+1)\Gamma(x+y+n+1)\Gamma(y+z+n+1)\Gamma(x+z+n+1)},\end{aligned}$$

where ${}_5F_4(a,b,c,d,e;f,g,h,i;z)$ is a GENERALIZED HYPERGEOMETRIC FUNCTION and $\Gamma(z)$ is the GAMMA FUNCTION.

see also DOUGALL-RAMANUJAN IDENTITY, GENERALIZED HYPERGEOMETRIC FUNCTION

Doughnut

see TORUS

Douglas-Neumann Theorem

If the lines joining corresponding points of two directly similar figures are divided proportionally, then the LOCUS of the points of the division will be a figure directly similar to the given figures.

References

Eves, H. "Solution to Problem E521." *Amer. Math. Monthly* **50**, 64, 1943.

Musselman, J. R. "Problem E521." *Amer. Math. Monthly* **49**, 335, 1942.

Dovetailing Problem

see CUBE DOVETAILING PROBLEM

Dowker Notation

A simple way to describe a knot projection. The advantage of this notation is that it enables a KNOT DIAGRAM to be drawn quickly.

For an oriented ALTERNATING KNOT with n crossings, begin at an arbitrary crossing and label it 1. Now follow the undergoing strand to the next crossing, and denote it 2. Continue around the knot following the same strand until each crossing has been numbered twice. Each crossing will have one even number and one odd number, with the numbers running from 1 to $2n$.

Now write out the ODD NUMBERS 1, 3, ..., $2n-1$ in a row, and underneath write the even crossing number corresponding to each number. The Dowker NOTATION is this bottom row of numbers. When the sequence of even numbers can be broken into two permutations of consecutive sequences (such as $\{4,6,2\}$ $\{10,12,8\}$), the knot is composite and is not uniquely determined by the Dowker notation. Otherwise, the knot is prime and the NOTATION uniquely defines a single knot (for amphichiral knots) or corresponds to a single knot or its MIRROR IMAGE (for chiral knots).

For general nonalternating knots, the procedure is modified slightly by making the sign of the even numbers

POSITIVE if the crossing is on the top strand, and NEGATIVE if it is on the bottom strand.

These data are available only for knots, but not for links, from Berkeley's gopher site.

References
Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 35–40, 1994.
Dowker, C. H. and Thistlethwaite, M. B. "Classification of Knot Projections." *Topol. Appl.* **16**, 19–31, 1983.

Down Arrow Notation

An inverse of the up ARROW NOTATION defined by

$$e \downarrow n = \ln n$$
$$e \downarrow\downarrow n = \ln^* n$$
$$e \downarrow\downarrow\downarrow n = \ln^{**} n,$$

where $\ln^* n$ is the number of times the NATURAL LOGARITHM must be iterated to obtain a value $\leq e$.

see also ARROW NOTATION

References
Vardi, I. *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 12 and 231–232, 1991.

Dozen

12.

see also BAKER'S DOZEN, GROSS

Dragon Curve

Nonintersecting curves which can be iterated to yield more and more sinuosity. They can be constructed by taking a path around a set of dots, representing a left turn by 1 and a right turn by 0. The first-order curve is then denoted 1. For higher order curves, add a 1 to the end, then copy the string of digits preceding it to the end but switching its center digit. For example, the second-order curve is generated as follows: $(1)1 \to (1)1(0) \to 110$, and the third as: $(110)1 \to (110)1(100) \to 1101100$. Continuing gives 110110011100100... (Sloane's A014577). The OCTAL representation sequence is 1, 6, 154, 66344, ... (Sloane's A003460). The dragon curves of orders 1 to 9 are illustrated below.

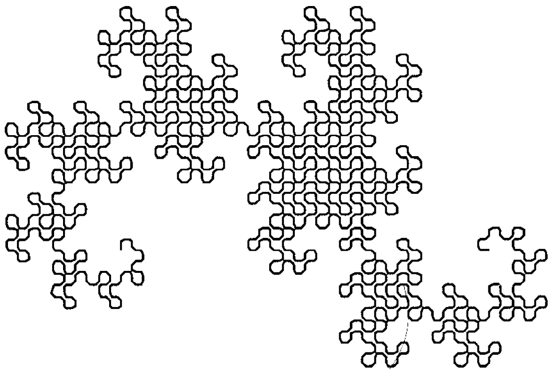

This procedure is equivalent to drawing a RIGHT ANGLE and subsequently replacing each RIGHT ANGLE with another smaller RIGHT ANGLE (Gardner 1978). In fact, the dragon curve can be written as a LINDENMAYER SYSTEM with initial string `"FX"`, STRING REWRITING rules `"X" -> "X+YF+"`, `"Y" -> "-FX-Y"`, and angle 90°.

see also LINDENMAYER SYSTEM, PEANO CURVE

References
Dickau, R. M. "Two-Dimensional L-Systems." `http://forum.swarthmore.edu/advanced/robertd/lsys2d.html`.
Dixon, R. *Mathographics.* New York: Dover, pp. 180–181, 1991.
Dubrovsky, V. "Nesting Puzzles, Part I: Moving Oriental Towers." *Quantum* **6**, 53–57 (Jan.) and 49–51 (Feb.), 1996.
Dubrovsky, V. "Nesting Puzzles, Part II: Chinese Rings Produce a Chinese Monster." *Quantum* **6**, 61–65 (Mar.) and 58–59 (Apr.), 1996.
Gardner, M. *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, pp. 207–209 and 215–220, 1978.
Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 48–53, 1991.
Peitgen, H.-O. and Saupe, D. (Eds.). *The Science of Fractal Images.* New York: Springer-Verlag, p. 284, 1988.
Sloane, N. J. A. Sequences A014577 and A003460/M4300 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Vasilyev, N. and Gutenmacher, V. "Dragon Curves." *Quantum* **6**, 5–10, 1995.

Dragon Fractal

see DOUADY'S RABBIT FRACTAL

Draughts

see CHECKERS

Drinfeld's Symmetric Space

A set of points which do not lie on any of a certain class of HYPERPLANES.

References
Teitelbaum, J. "The Geometry of p-adic Symmetric Spaces." *Not. Amer. Math. Soc.* **42**, 1120–1126, 1995.

Droz-Farny Circles

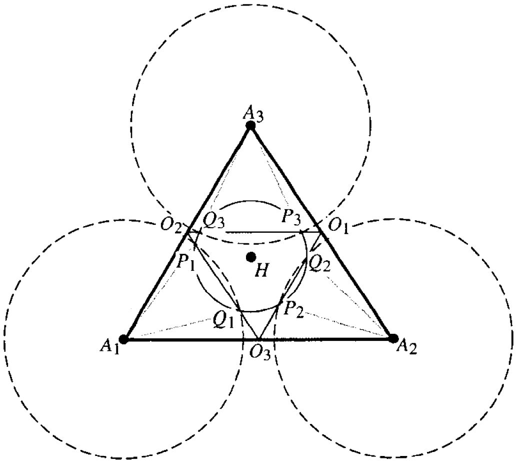

Draw a CIRCLE with center H which cuts the lines O_2O_3, O_3O_1, and O_1O_2 (where O_i are the MIDPOINTS) at P_1, Q_1; P_2, Q_2; and P_3, Q_3 respectively, then

$$\overline{A_1P_1} = \overline{A_2P_2} = \overline{A_3P_3} = \overline{A_1Q_1} = \overline{A_2Q_2} = \overline{A_3Q_3}.$$

Conversely, if equal CIRCLES are drawn about the VERTICES of a TRIANGLE, they cut the lines joining the MIDPOINTS of the corresponding sides in six points. These points lie on a CIRCLE whose center is the ORTHOCENTER. If r is the RADIUS of the equal CIRCLES centered on the vertices A_1, A_2, and A_3, and R_0 is the RADIUS of the CIRCLE about H, then

$$R_1{}^2 = 4R^2 + r^2 - \tfrac{1}{2}(a_1{}^2 + a_2{}^2 + a_3{}^2).$$

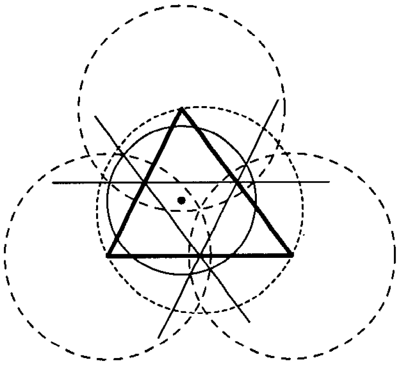

If the circles equal to the CIRCUMCIRCLE are drawn about the VERTICES of a triangle, they cut the lines joining midpoints of the adjacent sides in points of a CIRCLE R_2 with center H and RADIUS

$$R_2{}^2 = 5R^2 - \tfrac{1}{2}(a_1{}^2 + a_2{}^2 + a_3{}^2).$$

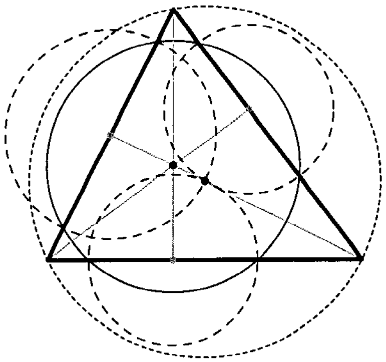

It is equivalent to the circle obtained by drawing circles with centers at the feet of the altitudes and passing through the CIRCUMCENTER. These circles cut the corresponding sides in six points on a circle R_2' whose center is H.

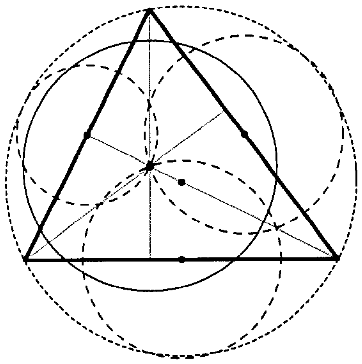

Furthermore, the circles about the midpoints of the sides and passing though H cut the sides in six points lying on another equivalent circle R_2'' whose center is O. In summary, the second Droz-Farny circle passes through 12 notable points, two on each of the sides and two on each of the lines joining midpoints of the sides.

References

Goormaghtigh, R. "Droz-Farny's Theorem." *Scripta Math.* **16**, 268–271, 1950.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 256–258, 1929.

Drum

see ISOSPECTRAL MANIFOLDS

Du Bois Raymond Constants

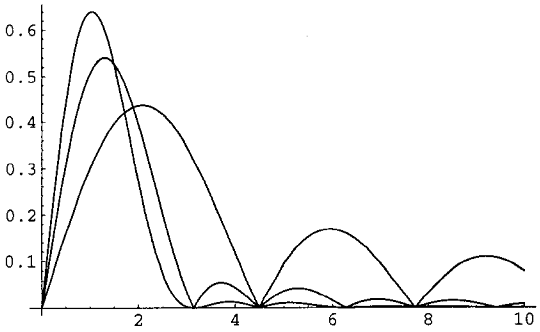

The constants C_n defined by

$$C_n \equiv \int_0^\infty \left| \frac{d}{dt} \left(\frac{\sin t}{t} \right)^n \right| dt - 1$$

which are difficult to compute numerically. The first few are

$$C_1 \approx 455$$
$$C_2 \approx 0.1945$$
$$C_3 \approx 0.028254$$
$$C_4 \approx 0.00524054.$$

Rather surprisingly, the second Du Bois Raymond constant is given analytically by

$$C_2 = \tfrac{1}{2}(e^2 - 7) = 0.1945280494\ldots$$

(Le Lionnais 1983).

References

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 23, 1983.

Plouffe, S. "Dubois-Raymond 2nd Constant." `http://lacim.uqam.ca/piDATA/dubois.txt`.

Dual Basis

Given a CONTRAVARIANT BASIS $\{\vec{e}_1, \ldots, \vec{e}_n\}$, its dual COVARIANT basis is given by

$$\vec{e}^{\,\alpha} \cdot \vec{e}_\beta = g(\vec{e}^{\,\alpha}, \vec{e}_\beta) = \delta^\alpha_\beta,$$

where g is the METRIC and δ^α_β is the mixed KRONECKER DELTA. In EUCLIDEAN SPACE with an ORTHONORMAL BASIS,

$$\vec{e}^{\,j} = \vec{e}_j,$$

so the BASIS and its dual are the same.

Dual Bivector

A dual BIVECTOR is defined by

$$\tilde{X}_{ab} \equiv \tfrac{1}{2}\epsilon_{abcd}X^{cd},$$

and a self-dual BIVECTOR by

$$X^*_{ab} \equiv X_{ab} + i\tilde{X}_{ab}.$$

Dual Graph

The dual graph G^* of a POLYHEDRAL GRAPH G has VERTICES each of which corresponds to a face of G and each of whose faces corresponds to a VERTEX of G. Two nodes in G^* are connected by an EDGE if the corresponding faces in G have a boundary EDGE in common.

Dual Map

see PULLBACK MAP

Dual Polyhedron

By the DUALITY PRINCIPLE, for every POLYHEDRON, there exists another POLYHEDRON in which faces and VERTICES occupy complementary locations. This POLYHEDRON is known as the dual, or RECIPROCAL. The dual polyhedron of a PLATONIC SOLID or ARCHIMEDEAN SOLID can be drawn by constructing EDGES tangent to the RECIPROCATING SPHERE (a.k.a. MIDSPHERE and INTERSPHERE) which are PERPENDICULAR to the original EDGES.

The dual of a general solid can be computed by connecting the midpoints of the sides surrounding each VERTEX, and constructing the corresponding tangent POLYGON. (The tangent polygon is the polygon which is tangent to the CIRCUMCIRCLE of the POLYGON produced by connecting the MIDPOINT on the sides surrounding the given VERTEX.) The process is illustrated below for the PLATONIC SOLIDS. The POLYHEDRON COMPOUNDS consisting of a POLYHEDRON and its dual are generally very attractive, and are also illustrated below for the PLATONIC SOLIDS.

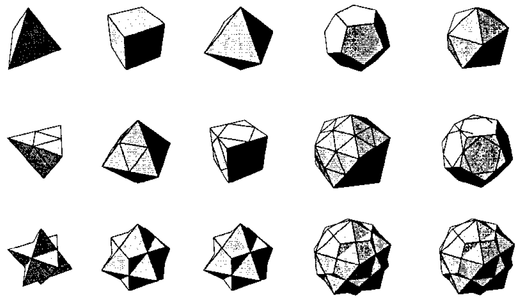

The ARCHIMEDEAN SOLIDS and their duals are illustrated below.

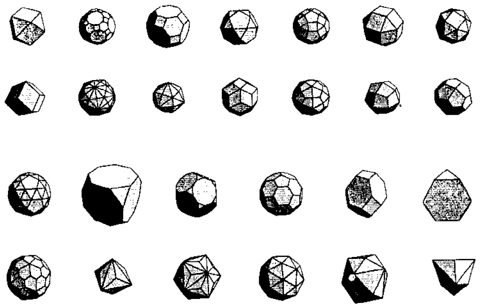

The following table gives a list of the duals of the PLATONIC SOLIDS and KEPLER-POINSOT SOLIDS together with the names of the POLYHEDRON-dual COMPOUNDS.

Polyhedron	Dual
Császár polyhedron	Szilassi polyhedron
cube	octahedron
cuboctahedron	rhombic dodecahedron
dodecahedron	icosahedron
great dodecahedron	small stellated dodec.
great icosahedron	great stellated dodec.
great stellated dodec.	great icosahedron
icosahedron	dodecahedron
octahedron	cube
small stellated dodec.	great dodecahedron
Szilassi polyhedron	Császár polyhedron
tetrahedron	tetrahedron

polyhedron compound	
cube	cube-octahedron compound
dodecahedron	dodec.-icosahedron compound
great dodecahedron	great dodecahedron-small stellated dodec. compound
great icosahedron	great icosahedron-great stellated dodec. compound
great stellated dodec.	great icosahedron-great stellated dodec. compound
icosahedron	dodec.-icosahedron compound
octahedron	cube-octahedron compound
small stellated dodec.	great dodec.-small stellated dodec. compound
tetrahedron	stella octangula

see also DUALITY PRINCIPLE, POLYHEDRON COMPOUND, RECIPROCATING SPHERE

References

Weisstein, E. W. "Polyhedron Duals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Duals.m`.

Wenninger, M. *Dual Models.* Cambridge, England: Cambridge University Press, 1983.

Dual Scalar

Given a third RANK TENSOR,

$$V_{ijk} \equiv \det[\mathbf{A} \quad \mathbf{B} \quad \mathbf{C}],$$

where det is the DETERMINANT, the dual scalar is defined as

$$V \equiv \frac{1}{3!}\epsilon_{ijk}V_{ijk},$$

where ϵ_{ijk} is the LEVI-CIVITA TENSOR.

see also DUAL TENSOR, LEVI-CIVITA TENSOR

Dual Solid

see DUAL POLYHEDRON

Dual Tensor

Given an antisymmetric second RANK TENSOR C_{ij}, a dual pseudotensor C_i is defined by

$$C_i \equiv \tfrac{1}{2}\epsilon_{ijk}C_{jk}, \qquad (1)$$

where

$$C_i \equiv \begin{bmatrix} C_{23} \\ C_{31} \\ C_{12} \end{bmatrix} \qquad (2)$$

$$C_{jk} \equiv \begin{bmatrix} 0 & C_{12} & -C_{31} \\ -C_{12} & 0 & C_{23} \\ C_{31} & -C_{23} & 0 \end{bmatrix}. \qquad (3)$$

see also DUAL SCALAR

References

Arfken, G. "Pseudotensors, Dual Tensors." §3.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 128–137, 1985.

Dual Voting

A term in SOCIAL CHOICE THEORY meaning each alternative receives equal weight for a single vote.

see also ANONYMOUS, MONOTONIC VOTING

Duality Principle

All the propositions in PROJECTIVE GEOMETRY occur in dual pairs which have the property that, starting from either proposition of a pair, the other can be immediately inferred by interchanging the parts played by the words "point" and "line." A similar duality exists for RECIPROCATION (Casey 1893).

see also BRIANCHON'S THEOREM, CONSERVATION OF NUMBER PRINCIPLE, DESARGUES' THEOREM, DUAL POLYHEDRON, PAPPUS'S HEXAGON THEOREM, PASCAL'S THEOREM, PERMANENCE OF MATHEMATICAL RELATIONS PRINCIPLE, PROJECTIVE GEOMETRY, RECIPROCATION

References

Casey, J. "Theory of Duality and Reciprocal Polars." Ch. 13 in *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions, with Numerous Examples, 2nd ed., rev. enl.* Dublin: Hodges, Figgis, & Co., pp. 382–392, 1893.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 107–110, 1990.

Duality Theorem

Dual pairs of LINEAR PROGRAMS are in "strong duality" if both are possible. The theorem was first conceived by John von Neumann. The first written proof was an Air Force report by George Dantzig, but credit is usually given to Tucker, Kuhn, and Gale.

see also LINEAR PROGRAMMING

Duffing Differential Equation

The most general forced form of the Duffing equation is

$$\ddot{x} + \delta\dot{x} + (\beta x^3 \pm {\omega_0}^2 x) = A\sin(\omega t + \phi). \qquad (1)$$

If there is no forcing, the right side vanishes, leaving

$$\ddot{x} + \delta\dot{x} + (\beta x^3 \pm {\omega_0}^2 x) = 0. \qquad (2)$$

If $\delta = 0$ and we take the plus sign,

$$\ddot{x} + {\omega_0}^2 x + \beta x^3 = 0. \qquad (3)$$

This equation can display chaotic behavior. For $\beta > 0$, the equation represents a "hard spring," and for $\beta < 0$, it represents a "soft spring." If $\beta < 0$, the phase portrait curves are closed. Returning to (1), take $\beta = 1$, $\omega_0 = 1$, $A = 0$, and use the minus sign. Then the equation is

$$\ddot{x} + \delta\dot{x} + (x^3 - x) = 0 \qquad (4)$$

(Ott 1993, p. 3). This can be written as a system of first-order ordinary differential equations by writing

$$\dot{x} = y, \qquad (5)$$

$$\dot{y} = x - x^3 - \delta y. \qquad (6)$$

The fixed points of these differential equations

$$\dot{x} = y = 0, \tag{7}$$

so $y = 0$, and

$$\dot{y} = x - x^3 - \delta y = x(1 - x^2) - 0 \tag{8}$$

giving $x = 0, \pm 1$. Differentiating,

$$\ddot{x} = \dot{y} = x - x^3 - \delta y \tag{9}$$

$$\ddot{y} = (1 - 3x^2)\dot{x} - \delta \dot{y} \tag{10}$$

$$\begin{bmatrix} \ddot{x} \\ \ddot{y} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 - 3x^2 & -\delta \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix}. \tag{11}$$

Examine the stability of the point (0,0):

$$\begin{vmatrix} 0 - \lambda & 1 \\ 1 & -\delta - \lambda \end{vmatrix} = \lambda(\lambda + \delta) - 1 = \lambda^2 + \lambda\delta - 1 = 0 \tag{12}$$

$$\lambda_\pm^{(0,0)} = \tfrac{1}{2}(-\delta \pm \sqrt{\delta^2 + 4}). \tag{13}$$

But $\delta^2 \geq 0$, so $\lambda_\pm^{(0,0)}$ is real. Since $\sqrt{\delta^2 + 4} > |\delta|$, there will always be one POSITIVE ROOT, so this fixed point is unstable. Now look at (±1, 0).

$$\begin{vmatrix} 0 - \lambda & 1 \\ -2 & -\delta - \lambda \end{vmatrix} = \lambda(\lambda + \delta) + 2 = \lambda^2 + \lambda\delta + 2 = 0 \tag{14}$$

$$\lambda_\pm^{(\pm 1,0)} = \tfrac{1}{2}(-\delta \pm \sqrt{\delta^2 - 8}). \tag{15}$$

For $\delta > 0$, $\Re[\lambda_\pm^{(\pm 1,0)}] < 0$, so the point is asymptotically stable. If $\delta = 0$, $\lambda_\pm^{(\pm 1,0)} = \pm i\sqrt{2}$, so the point is linearly stable. If $\delta \in (-2\sqrt{2}, 0)$, the radical gives an IMAGINARY PART and the REAL PART is > 0, so the point is unstable. If $\delta = -2\sqrt{2}$, $\lambda_\pm^{(\pm 1,0)} = \sqrt{2}$, which has a POSITIVE REAL ROOT, so the point is unstable. If $\delta < -2\sqrt{2}$, then $|\delta| < \sqrt{\delta^2 - 8}$, so both ROOTS are POSITIVE and the point is unstable. Summarizing,

$$\begin{cases} \text{asymptotically stable} & \delta > 0 \\ \text{linearly stable (superstable)} & \delta = 0 \\ \text{unstable} & \delta < 0. \end{cases} \tag{16}$$

Now specialize to the case $\delta = 0$, which can be integrated by quadratures. In this case, the equations become

$$\dot{x} = y \tag{17}$$

$$\dot{y} = x - x^3. \tag{18}$$

Differentiating (17) and plugging in (18) gives

$$\ddot{x} = \dot{y} = x - x^3. \tag{19}$$

Multiplying both sides by $\dot{x}$ gives

$$\ddot{x}\dot{x} - \dot{x}x + \dot{x}x^3 = 0 \tag{20}$$

$$\frac{d}{dt}(\tfrac{1}{2}\dot{x}^2 - \tfrac{1}{2}x^2 + \tfrac{1}{4}x^4) = 0, \tag{21}$$

so we have an invariant of motion h,

$$h \equiv \tfrac{1}{2}\dot{x}^2 - \tfrac{1}{2}x^2 + \tfrac{1}{4}x^4. \tag{22}$$

Solving for $\dot{x}^2$ gives

$$\dot{x}^2 = \left(\frac{dx}{dt}\right)^2 = 2h + x^2 - \tfrac{1}{2}x^4 \tag{23}$$

$$\frac{dx}{dt} = \sqrt{2h + x^2 + \tfrac{1}{2}x^2}, \tag{24}$$

so

$$t = \int dt = \int \frac{dx}{\sqrt{2h + x^2 + \frac{1}{2}x^2}}. \tag{25}$$

Note that the invariant of motion h satisfies

$$\dot{x} = \frac{\partial h}{\partial \dot{x}} = \frac{\partial h}{\partial y} \tag{26}$$

$$\frac{\partial h}{\partial x} = -x + x^3 = -\dot{y}, \tag{27}$$

so the equations of the Duffing oscillator are given by the HAMILTONIAN SYSTEM

$$\begin{cases} \dot{x} = \frac{\partial h}{\partial y} \\ \dot{y} = -\frac{\partial h}{\partial x}. \end{cases} \tag{28}$$

References

Ott, E. *Chaos in Dynamical Systems.* New York: Cambridge University Press, 1993.

Duhamel's Convolution Principle

Can be used to invert a LAPLACE TRANSFORM.

Dumbbell Curve

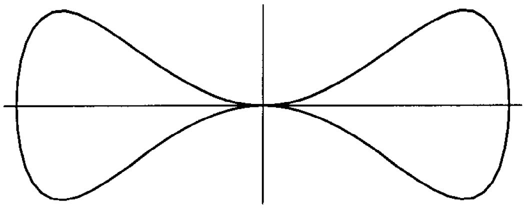

$$y^2 = a^2(x^4 - x^6).$$

see also BUTTERFLY CURVE, EIGHT CURVE, PIRIFORM

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 72, 1989.

Duodecillion

In the American system, 10^{39}.

see also LARGE NUMBER

Dupin's Cyclide

see CYCLIDE

Dupin's Indicatrix

A pair of conics obtained by expanding an equation in MONGE'S FORM $z = F(x,y)$ in a MACLAURIN SERIES

$$\begin{aligned} z &= z(0,0) + z_1 x + z_2 y \\ &\quad + \tfrac{1}{2}(z_{11}x^2 + 2z_{12}xy + z_{22}y^2) + \ldots \\ &= \tfrac{1}{2}(b_{11}x^2 + 2b_{12}xy + b_{22}y^2). \end{aligned}$$

This gives the equation

$$b_{11}x^2 + 2b_{12}xy + b_{22}y^2 = \pm 1.$$

Amazingly, the radius of the indicatrix in any direction is equal to the SQUARE ROOT of the RADIUS OF CURVATURE in that direction (Coxeter 1969).

References

Coxeter, H. S. M. "Dupin's Indicatrix" §19.8 in *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 363–365, 1969.

Dupin's Theorem

In three mutually orthogonal systems of the surfaces, the LINES OF CURVATURE on any surface in one of the systems are its intersections with the surfaces of the other two systems.

Duplication of the Cube

see CUBE DUPLICATION

Duplication Formula

see LEGENDRE DUPLICATION FORMULA

Durand's Rule

The NEWTON-COTES FORMULA

$$\int_{x_1}^{x_n} f(x)\,dx = h(\tfrac{2}{5}f_1 + \tfrac{11}{10}f_2 + f_3 + \ldots + f_{n-2} + \tfrac{11}{10}f_{n-1} + \tfrac{2}{5}f_n).$$

see also BODE'S RULE, HARDY'S RULE, NEWTON-COTES FORMULAS, SIMPSON'S 3/8 RULE, SIMPSON'S RULE, TRAPEZOIDAL RULE, WEDDLE'S RULE

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 127, 1987.

Dürer's Conchoid

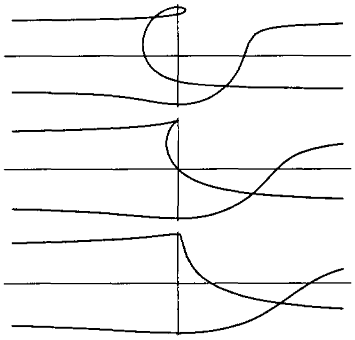

These curves appear in Dürer's work *Instruction in Measurement with Compasses and Straight Edge* (1525) and arose in investigations of perspective. Dürer constructed the curve by drawing lines QRP and $P'QR$ of length 16 units through $Q(q,0)$ and $R(r,0)$, where $q+r=13$. The locus of P and P' is the curve, although Dürer found only one of the two branches of the curve.

The ENVELOPE of the lines QRP and $P'QR$ is a PARABOLA, and the curve is therefore a GLISSETTE of a point on a line segment sliding between a PARABOLA and one of its TANGENTS.

Dürer called the curve "Muschellini," which means CONCHOID. However, it is not a true CONCHOID and so is sometimes called DÜRER'S SHELL CURVE. The Cartesian equation is

$$\begin{aligned} 2y^2(x^2+y^2) - 2by^2(x+y) + (b^2-3a^2)y^2 - a^2x^2 \\ + 2a^2b(x+y) + a^2(a^2-b^2) = 0. \end{aligned}$$

The above curves are for $(a,b) = (3,1)$, $(3,3)$, $(3,5)$. There are a number of interesting special cases. If $b=0$, the curve becomes two coincident straight lines $x=0$. For $a=0$, the curve becomes the line pair $x=b/2$, $x=-b/2$, together with the CIRCLE $x+y=b$. If $a=b/2$, the curve has a CUSP at $(-2a,a)$.

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 157–159, 1972.

Lockwood, E. H. *A Book of Curves.* Cambridge, England: Cambridge University Press, p. 163, 1967.

MacTutor History of Mathematics Archive. "Dürer's Shell Curves." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Durers.html`.

Dürer's Magic Square

16	3	2	13
5	10	11	8
9	6	7	12
4	15	14	1

Dürer's magic square is a MAGIC SQUARE with MAGIC CONSTANT 34 used in an engraving entitled *Melencolia I* by Albrecht Dürer (The British Museum). The engraving shows a disorganized jumble of scientific equipment lying unused while an intellectual sits absorbed in thought. Dürer's magic square is located in the upper left-hand corner of the engraving. The numbers 15 and 14 appear in the middle of the bottom row, indicating the date of the engraving, 1514.

References
Boyer, C. D. and Merzbach, U. C. *A History of Mathematics.* New York: Wiley, pp. 296–297, 1991.
Hunter, J. A. H. and Madachy, J. S. *Mathematical Diversions.* New York: Dover, p. 24, 1975.
Rivera, C. "Melancholia." `http://www.sci.net.mx/~crivera/melancholia.htm`.

Dürer's Shell Curve

see DÜRER'S CONCHOID

Durfee Polynomial

Let $F(n)$ be a family of PARTITIONS of n and let $F(n,d)$ denote the set of PARTITIONS in $F(n)$ with DURFEE SQUARE of size d. The Durfee polynomial of $F(n)$ is then defined as the polynomial

$$P_{F,n}=\sum|F(n,d)|y^d,$$

where $0\leq d\leq\sqrt{n}$.

see also DURFEE SQUARE, PARTITION

References
Canfield, E. R.; Corteel, S.; and Savage, C. D. "Durfee Polynomials." *Electronic J. Combinatorics* **5**, No. 1, R32, 1–21, 1998. `http://www.combinatorics.org/Volume_5/v5i1toc.html#R32`.

Durfee Square

The length of the largest-sized SQUARE contained within the FERRERS DIAGRAM of a PARTITION.

see also DURFEE POLYNOMIAL, FERRERS DIAGRAM, PARTITION

Dvoretzky's Theorem

Each centered convex body of sufficiently high dimension has an "almost spherical" k-dimensional central section.

Dyad

Dyads extend VECTORS to provide an alternative description to second RANK TENSORS. A dyad $D(\mathbf{A},\mathbf{B})$ of a pair of VECTORS $\mathbf{A}$ and $\mathbf{B}$ is defined by $D(\mathbf{A},\mathbf{B})\equiv\mathbf{AB}$. The DOT PRODUCT is defined by

$$\mathbf{A}\cdot\mathbf{BC}\equiv(\mathbf{A}\cdot\mathbf{B})\mathbf{C}$$

$$\mathbf{AB}\cdot\mathbf{C}\equiv\mathbf{A}(\mathbf{B}\cdot\mathbf{C}),$$

and the COLON PRODUCT by

$$\mathbf{AB}:\mathbf{CD}\equiv\mathbf{C}\cdot\mathbf{AB}\cdot\mathbf{D}=(\mathbf{A}\cdot\mathbf{C})(\mathbf{B}\cdot\mathbf{D}).$$

References
Morse, P. M. and Feshbach, H. "Dyadics and Other Vector Operators." §1.6 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 54–92, 1953.

Dyadic

A linear POLYNOMIAL of DYADS $\mathbf{AB}+\mathbf{CD}+\ldots$ consisting of nine components A_{ij} which transform as

$$(A_{ij})'=\sum_{m,n}\frac{h_mh_n}{h_i'h_j'}\frac{\partial x_m}{\partial x_i'}\frac{\partial x_n}{\partial x_j'}A_{mn}\tag{1}$$

$$=\sum_{m,n}\frac{h_i'h_j'}{h_mh_n}\frac{\partial x_i'}{\partial x_m}\frac{\partial x_j'}{\partial x_n}A_{mn}\tag{2}$$

$$=\sum_{m,n}\frac{h_i'h_n}{h_mh_j'}\frac{\partial x_i'}{\partial x_m}\frac{\partial x_n}{\partial x_j'}A_{mn}.\tag{3}$$

Dyadics are often represented by Gothic capital letters. The use of dyadics is nearly archaic since TENSORS perform the same function but are notationally simpler.

A unit dyadic is also called the IDEMFACTOR and is defined such that

$$\mathbf{I}\cdot\mathbf{A}\equiv\mathbf{A}.\tag{4}$$

In CARTESIAN COORDINATES,

$$\mathbf{I}=\hat{\mathbf{x}}\hat{\mathbf{x}}+\hat{\mathbf{y}}\hat{\mathbf{y}}+\hat{\mathbf{z}}\hat{\mathbf{z}},\tag{5}$$

and in SPHERICAL COORDINATES

$$\mathbf{I}=\nabla\mathbf{r}.\tag{6}$$

see also DYAD, TETRADIC

References
Arfken, G. "Dyadics." §3.5 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 137–140, 1985.
Morse, P. M. and Feshbach, H. "Dyadics and Other Vector Operators." §1.6 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 54–92, 1953.

Dyck's Theorem

see VON DYCK'S THEOREM

Dye's Theorem

For any two ergodic measure-preserving transformations on nonatomic PROBABILITY SPACES, there is an ISOMORPHISM between the two PROBABILITY SPACES carrying orbits onto orbits.

Dymaxion

Buckminster Fuller's term for the CUBOCTAHEDRON.

see also CUBOCTAHEDRON, MECON

Dynamical System

A means of describing how one state develops into another state over the course of time. Technically, a dynamical system is a smooth action of the reals or the INTEGERS on another object (usually a MANIFOLD). When the reals are acting, the system is called a continuous dynamical system, and when the INTEGERS are acting, the system is called a discrete dynamical system. If f is any CONTINUOUS FUNCTION, then the evolution of a variable x can be given by the formula

$$x_{n+1} = f(x_n). \qquad (1)$$

This equation can also be viewed as a difference equation

$$x_{n+1} - x_n = f(x_n) - x_n, \qquad (2)$$

so defining

$$g(x) \equiv f(x) - x \qquad (3)$$

gives

$$x_{n+1} - x_n = g(x_n) * 1, \qquad (4)$$

which can be read "as n changes by 1 unit, x changes by $g(x)$." This is the discrete analog of the DIFFERENTIAL EQUATION

$$x'(n) = g(x(n)). \qquad (5)$$

see also ANOSOV DIFFEOMORPHISM, ANOSOV FLOW, AXIOM A DIFFEOMORPHISM, AXIOM A FLOW, BIFURCATION THEORY, CHAOS, ERGODIC THEORY, GEODESIC FLOW

References
Aoki, N. and Hiraide, K. *Topological Theory of Dynamical Systems.* Amsterdam, Netherlands: North-Holland, 1994.
Golubitsky, M. *Introduction to Applied Nonlinear Dynamical Systems and Chaos.* New York: Springer-Verlag, 1997.
Guckenheimer, J. and Holmes, P. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields, 3rd ed.* New York: Springer-Verlag, 1997.
Lichtenberg, A. and Lieberman, M. *Regular and Stochastic Motion, 2nd ed.* New York: Springer-Verlag, 1994.
Ott, E. *Chaos in Dynamical Systems.* New York: Cambridge University Press, 1993.
Rasband, S. N. *Chaotic Dynamics of Nonlinear Systems.* New York: Wiley, 1990.
Strogatz, S. H. *Nonlinear Dynamics and Chaos, with Applications to Physics, Biology, Chemistry, and Engineering.* 1994.
Tabor, M. *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, 1989.

Dynkin Diagram

A diagram used to describe CHEVALLEY GROUPS.

see also COXETER-DYNKIN DIAGRAM

References
Jacobson, N. *Lie Algebras.* New York: Dover, p. 128, 1979.

E

e

The base of the NATURAL LOGARITHM, named in honor of Euler. It appears in many mathematical contexts involving LIMITS and DERIVATIVES, and can be defined by

$$e \equiv \lim_{x\to\infty} \left(1+\frac{1}{x}\right)^x, \tag{1}$$

or by the infinite sum

$$e = \sum_{k=0}^{\infty} \frac{1}{k!}. \tag{2}$$

The numerical value of e is

$$e = 2.718281828459045235360287471352662497757\ldots \tag{3}$$

(Sloane's A001113).

Euler proved that e is IRRATIONAL, and Liouville proved in 1844 that e does not satisfy any QUADRATIC EQUATION with integral COEFFICIENTS. Hermite proved e to be TRANSCENDENTAL in 1873. It is not known if $\pi + e$ or π/e is IRRATIONAL. However, it is known that $\pi + e$ and π/e do not satisfy any POLYNOMIAL equation of degree ≤ 8 with INTEGER COEFFICIENTS of average size 10^9 (Bailey 1988, Borwein *et al.* 1989).

The special case of the EULER FORMULA

$$e^{ix} = \cos x + i \sin x \tag{4}$$

with $x = \pi$ gives the beautiful identity

$$e^{i\pi} + 1 = 0, \tag{5}$$

an equation connecting the fundamental numbers i, PI, e, 1, and 0 (ZERO).

Some CONTINUED FRACTION representations of e include

$$e = 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{2}{3 + \cfrac{3}{\ddots}}}} \tag{6}$$

$$= [2, 1, 2, 1, 1, 4, 1, 1, 6, \ldots] \tag{7}$$

(Sloane's A003417) and

$$\frac{e-1}{e+1} = [2, 6, 10, 14, \ldots] \tag{8}$$

$$e - 1 = [1, 1, 2, 1, 1, 4, 1, 1, 6, \ldots] \tag{9}$$

$$\tfrac{1}{2}(e-1) = [0, 1, 6, 10, 14, \ldots] \tag{10}$$

$$\sqrt{e} = [1, 1, 1, 1, 5, 1, 1, 1, 9, 1, \ldots]. \tag{11}$$

The first few convergents of the CONTINUED FRACTION are 3, 8/3, 11/4, 19/7, 87/32, 106/39, 193/71, ... (Sloane's A007676 and A007677).

Using the RECURRENCE RELATION

$$a_n = n(a_{n-1} + 1) \tag{12}$$

with $a_1 = a^{-1}$, compute

$$\prod_{n=1}^{\infty} (1 + {a_n}^{-1}). \tag{13}$$

The result is e^a. Gosper gives the unusual equation connecting π and e,

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \cos\left(\frac{9}{n\pi + \sqrt{n^2\pi^2 - 9}}\right) = -\frac{\pi^2}{12e^3} = -0.040948222\ldots. \tag{14}$$

Rabinowitz and Wagon (1995) give an ALGORITHM for computing digits of e based on earlier DIGITS, but a much simpler SPIGOT ALGORITHM was found by Sales (1968). Around 1966, MIT hacker Eric Jensen wrote a very concise program (requiring less than a page of assembly language) that computed e by converting from factorial base to decimal.

Let $p(n)$ be the probability that a random ONE-TO-ONE function on the INTEGERS 1, ..., n has at least one FIXED POINT. Then

$$\lim_{n\to\infty} p(n) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k!} = 1 - \frac{1}{e} = 0.6321205588\ldots. \tag{15}$$

STIRLING'S FORMULA gives

$$\lim_{n\to\infty} \frac{(n!)^{1/n}}{n} = \frac{1}{e}. \tag{16}$$

Castellanos (1988) gives several curious approximations to e,

$$e \approx 2 + \frac{54^2 + 41^2}{80^2} \tag{17}$$

$$\approx (\pi^4 + \pi^5)^{1/6} \tag{18}$$

$$\approx \frac{271801}{99990} \tag{19}$$

$$\approx \left(150 - \frac{87^3 + 12^5}{83^3}\right)^{1/5} \tag{20}$$

$$\approx 4 - \frac{300^4 - 100^4 - 1291^2 + 9^2}{91^5} \tag{21}$$

$$\approx \left(1097 - \frac{55^5 + 311^3 - 11^3}{68^5}\right)^{1/7}, \tag{22}$$

which are good to 6, 7, 9, 10, 12, and 15 digits respectively.

Examples of *e* MNEMONICS (Gardner 1959, 1991) include:

> "By omnibus I traveled to Brooklyn" (6 digits).
>
> "To disrupt a playroom is commonly a practice of children" (10 digits).
>
> "It enables a numskull to memorize a quantity of numerals" (10 digits).
>
> "I'm forming a mnemonic to remember a function in analysis" (10 digits).
>
> "He repeats: I shouldn't be tippling, I shouldn't be toppling here!" (11 digits).
>
> "In showing a painting to probably a critical or venomous lady, anger dominates. O take guard, or she raves and shouts" (21 digits). Here, the word "O" stands for the number 0.

A much more extensive mnemonic giving 40 digits is

> "We present a mnemonic to memorize a constant so exciting that Euler exclaimed: '!' when first it was found, yes, loudly '!'. My students perhaps will compute *e*, use power or Taylor series, an easy summation formula, obvious, clear, elegant!"

(Barel 1995). In the latter, 0s are represented with "!". A list of *e* mnemonics in several languages is maintained by A. P. Hatzipolakis.

Scanning the decimal expansion of *e* until all *n*-digit numbers have occurred, the last appearing is 6, 12, 548, 1769, 92994, 513311, ... (Sloane's A032511). These end at positions 21, 372, 8092, 102128, 1061613, 12108841, ...

see also CARLEMAN'S INEQUALITY, COMPOUND INTEREST, DE MOIVRE'S IDENTITY, EULER FORMULA, EXPONENTIAL FUNCTION, HERMITE-LINDEMANN THEOREM, NATURAL LOGARITHM

References

Bailey, D. H. "Numerical Results on the Transcendence of Constants Involving π, *e*, and Euler's Constant." *Math. Comput.* **50**, 275–281, 1988.

Barel, Z. "A Mnemonic for *e*." *Math. Mag.* **68**, 253, 1995.

Borwein, J. M.; Borwein, P. B.; and Bailey, D. H. "Ramanujan, Modular Equations, and Approximations to Pi or How to Compute One Billion Digits of Pi." *Amer. Math. Monthly* **96**, 201–219, 1989.

Castellanos, D. "The Ubiquitous Pi. Part I." *Math. Mag.* **61**, 67–98, 1988.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 201 and 250–254, 1996.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/e/e.html`.

Gardner, M. "Memorizing Numbers." Ch. 11 in *The Scientific American Book of Mathematical Puzzles and Diversions.* New York: Simon and Schuster, pp. 103 and 109, 1959.

Gardner, M. Ch. 3 in *The Unexpected Hanging and Other Mathematical Diversions.* Chicago, IL: Chicago University Press, p. 40, 1991.

Hatzipolakis, A. P. "PiPhilology." `http://users.hol.gr/~xpolakis/piphil.html`.

Hermite, C. "Sur la fonction exponentielle." *C. R. Acad. Sci. Paris* **77**, 18–24, 74–79, and 226–233, 1873.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 47, 1983.

Maor, E. *e: The Story of a Number.* Princeton, NJ: Princeton University Press, 1994.

Minkus, J. "A Continued Fraction." Problem 10327. *Amer. Math. Monthly* **103**, 605–606, 1996.

Mitchell, U. G. and Strain, M. "The Number *e*." *Osiris* **1**, 476–496, 1936.

Olds, C. D. "The Simple Continued Fraction Expression of *e*." *Amer. Math. Monthly* **77**, 968–974, 1970.

Plouffe, S. "Plouffe's Inverter: Table of Current Records for the Computation of Constants." `http://lacim.uqam.ca/pi/records.html`.

Rabinowitz, S. and Wagon, S. "A Spigot Algorithm for the Digits of π." *Amer. Math. Monthly* **102**, 195–203, 1995.

Sales, A. H. J. "The Calculation of *e* to Many Significant Digits." *Computer J.* **11**, 229–230, 1968.

Sloane, N. J. A. Sequences A032511, A001113/M1727, A003417/M0088, A007676/M0869, and A007677/M2343 in "An On-Line Version of the Encyclopedia of Integer Sequences."

e-Divisor

d is called an *e*-divisor (or EXPONENTIAL DIVISOR) of

$$n = {p_1}^{a_1}{p_2}^{a_2}\cdots{p_r}^{a_r}$$

if $d|n$ and

$$d = {p_1}^{b_1}{p_2}^{b_2}\cdots{p_r}^{b_r}$$

where $b_j|a_j$ with $1 \leq j \leq r$.

see also *e*-PERFECT NUMBER

References

Guy, R. K. "Exponential-Perfect Numbers." §B17 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 73, 1994.

Straus, E. G. and Subbarao, M. V. "On Exponential Divisors." *Duke Math. J.* **41**, 465–471, 1974.

E_n-Function

The $\mathrm{E}_n(x)$ function is defined by the integral

$$\mathrm{E}_n(x) \equiv \int_1^\infty \frac{e^{-xt}\,dt}{t^n} \tag{1}$$

and is given by the *Mathematica*® (Wolfram Research, Champaign, IL) function `ExpIntegralE[n,x]`. Defining $t \equiv \eta^{-1}$ so that $dt = -\eta^{-2}d\eta$,

$$\mathrm{E}_n(x) = \int_0^1 e^{-x/\eta}\eta^{n-2}\,d\eta \tag{2}$$

$$\mathrm{E}_n(0) = \frac{1}{n-1}. \tag{3}$$

The function satisfies the RECURRENCE RELATIONS

$$\mathrm{E}'_n(x) = -\mathrm{E}_{n-1}(x) \tag{4}$$

$$n\,\mathrm{E}_{n+1}(x) = e^{-x} - x\,\mathrm{E}_n(x). \tag{5}$$

Equation (4) can be derived from

$$\mathrm{E}_n(x) = \int_1^\infty \frac{e^{-tx}}{t^n}\,dt \tag{6}$$

$$\begin{aligned} \mathrm{E}_n'(x) &= \frac{d}{dx}\int_1^\infty \frac{e^{-tx}}{t^n}\,dt = \int_1^\infty \frac{d}{dx}\left(\frac{e^{-tx}}{t^n}\right)dt \\ &= -\int_1^\infty t\frac{e^{-tx}}{t^n}\,dt \\ &= -\int_1^\infty \frac{e^{-tx}}{t^{n-1}}\,dt = -\mathrm{E}_{n-1}(x), \end{aligned} \tag{7}$$

and (5) using integrating by parts, letting

$$u = \frac{1}{t^n} \qquad dv = e^{-tx}\,dt \tag{8}$$

$$du = -\frac{n}{t^{n+1}}\,dt \qquad v = -\frac{e^{-tx}}{x}, \tag{9}$$

$$\begin{aligned} \mathrm{E}_n(x) &= \int u\,dv = uv - \int v\,du \\ &= -\frac{e^{-tx}}{xt^n} - \frac{n}{x}\int_1^\infty \frac{e^{-tx}\,dx}{t^{n+1}} \\ &= x\int_1^\infty \frac{e^{-tx}\,dt}{t^n} \\ &= -\left[\frac{1}{e^{tx}t^n}\right]_1^\infty - n\int_1^\infty \frac{e^{-tx}\,dx}{t^{n+1}} \\ &= x\,\mathrm{E}_n(x) = e^{-x} - n\,\mathrm{E}_{n+1}(x). \end{aligned} \tag{10}$$

Solving (10) for $n\,\mathrm{E}_n(x)$ gives (5). An asymptotic expansion gives

$$\begin{aligned} &(n-1)!\;\mathrm{E}_n(x) \\ &= (-x)^{n-1}\,\mathrm{E}_1(x) + e^{-x}\sum_{s=0}^{n} -2(n-s-2)!(-x)^s, \end{aligned} \tag{11}$$

so

$$\mathrm{E}_n(x) = \frac{e^{-x}}{x}\left[1 - \frac{n}{x} + \frac{n(n+1)}{x^2} + \ldots\right]. \tag{12}$$

The special case $n = 1$ gives

$$\mathrm{E}_1(x) \equiv -\operatorname{ei}(-x) = \int_1^\infty \frac{e^{-tx}\,dt}{t} = \int_x^\infty \frac{e^{-u}\,du}{u}, \tag{13}$$

where $\operatorname{ei}(x)$ is the EXPONENTIAL INTEGRAL, which is also equal to

$$\mathrm{E}_1(x) = -\gamma - \ln x - \sum_{n=1}^\infty \frac{(-1)^n x^n}{n!n}, \tag{14}$$

where γ is the EULER-MASCHERONI CONSTANT.

$$\mathrm{E}_1(0) = \infty \tag{15}$$

$$\mathrm{E}_1(ix) = -\operatorname{ci}(x) + i\operatorname{si}(x), \tag{16}$$

where $\operatorname{ci}(x)$ and $\operatorname{si}(x)$ are the COSINE INTEGRAL and SINE INTEGRAL.

see also COSINE INTEGRAL, E_t-FUNCTION, EXPONENTIAL INTEGRAL, GOMPERTZ CONSTANT, SINE INTEGRAL

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Exponential Integral and Related Functions." Ch. 5 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 227–233, 1972.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Exponential Integrals." §6.3 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 215–219, 1992.

Spanier, J. and Oldham, K. B. "The Exponential Integral Ei(x) and Related Functions." Ch. 37 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 351–360, 1987.

E_t-Function

A function which arises in FRACTIONAL CALCULUS.

$$E_t(\nu, a) = \frac{1}{\Gamma(\nu)}e^{at}\int_0^t x^{\nu-1}e^{-ax}\,dx = t^\nu e^{at}\gamma(\nu, at), \tag{1}$$

where γ is the incomplete GAMMA FUNCTION and Γ the complete GAMMA FUNCTION. The E_t function satisfies the RECURRENCE RELATION

$$E_t(\nu, a) = aE_t(\nu + 1, a) + \frac{t^\nu}{\Gamma(\nu + 1)}. \tag{2}$$

A special value is

$$E_t(0, a) = e^{at}. \tag{3}$$

see also E_n-FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Exponential Integral and Related Functions." Ch. 5 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 227–233, 1972.

e-Multiperfect Number

A number n is called a k e-perfect number if $\sigma_e(n) = kn$, where $\sigma_e(n)$ is the SUM of the e-DIVISORS of n.

see also e-DIVISOR, e-PERFECT NUMBER

References

Guy, R. K. "Exponential-Perfect Numbers." §B17 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 73, 1994.

e-Perfect Number

A number n is called an e-perfect number if $\sigma_e(n) = 2n$, where $\sigma_e(n)$ is the SUM of the e-DIVISORS of n. If m is SQUAREFREE, then $\sigma_e(m) = m$. As a result, if n is e-perfect and m is SQUAREFREE with $m \perp b$, then mn is e-perfect. There are no ODD e-perfect numbers.

see also e-DIVISOR

References

Guy, R. K. "Exponential-Perfect Numbers." §B17 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 73, 1994.

Subbarao, M. V. and Suryanarayan, D. "Exponential Perfect and Unitary Perfect Numbers." *Not. Amer. Math. Soc.* **18**, 798, 1971.

Ear

A PRINCIPAL VERTEX x_i of a SIMPLE POLYGON P is called an ear if the diagonal $[x_{i-1}, x_{i+1}]$ that bridges x_i lies entirely in P. Two ears x_i and x_j are said to overlap if

$$\operatorname{int}[x_{i-1}, x_i, x_{i+1}] \cap \operatorname{int}[x_{j-1}, x_j, x_{j+1}] = \varnothing.$$

The TWO-EARS THEOREM states that, except for TRIANGLES, every SIMPLE POLYGON has at least two nonoverlapping ears.

see also ANTHROPOMORPHIC POLYGON, MOUTH, TWO-EARS THEOREM

References

Meisters, G. H. "Polygons Have Ears." *Amer. Math. Monthly* **82**, 648–751, 1975.

Meisters, G. H. "Principal Vertices, Exposed Points, and Ears." *Amer. Math. Monthly* **87**, 284–285, 1980.

Toussaint, G. "Anthropomorphic Polygons." *Amer. Math. Monthly* **122**, 31–35, 1991.

Early Election Results

Let Jones and Smith be the only two contestants in an election that will end in a deadlock when all votes for Jones (J) and Smith (S) are counted. What is the EXPECTATION VALUE of $X_k \equiv |S - J|$ after k votes are counted? The solution is

$$\begin{aligned}\langle X_k\rangle &= \frac{2N\binom{N-1}{\lfloor k/2\rfloor}\binom{N-1}{\lfloor k/2\rfloor-1}}{\binom{2N}{k}}\\ &= \begin{cases} \frac{k(2N-k)}{2N}\binom{N}{k/2}^2\binom{2N}{k}^{-1} & \text{for } k \text{ even}\\ \frac{k(2N-k+1)}{2N}\binom{N}{(k-1)/2}^2\binom{2N}{k-1}^{-1} & \text{for } k \text{ odd.}\end{cases}\end{aligned}$$

References

Handelsman, M. B. Solution to Problem 10248. "Early Returns in a Tied Election." *Amer. Math. Monthly* **102**, 554–556, 1995.

Eban Number

The sequence of numbers whose names (in English) do not contain the letter "e" (i.e., "e" is "banned"). The first few eban numbers are 2, 4, 6, 30, 32, 34, 36, 40, 42, 44, 46, 50, 52, 54, 56, 60, 62, 64, 66, 2000, 2002, 2004, ... (Sloane's A006933); i.e., two, four, six, thirty, etc.

References

Sloane, N. J. A. Sequence A006933/M1030 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Eberhart's Conjecture

If q_n is the nth prime such that M_{q_n} is a MERSENNE PRIME, then

$$q_n \sim (3/2)^n.$$

It was modified by Wagstaff (1983) to yield

$$q_n \sim (2^{e^{-\gamma}})^n,$$

where γ is the EULER-MASCHERONI CONSTANT.

References

Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, pp. 332–333, 1989.

Wagstaff, S. S. "Divisors of Mersenne Numbers." *Math. Comput.* **40**, 385–397, 1983.

Eccentric

Not CONCENTRIC.

see also CONCENTRIC, CONCYCLIC

Eccentric Angle

The angle θ measured from the CENTER of an ELLIPSE to a point on the ELLIPSE.

see also ECCENTRICITY, ELLIPSE

Eccentric Anomaly

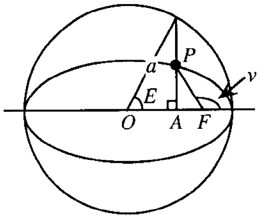

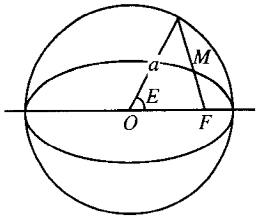

The ANGLE obtained by drawing the AUXILIARY CIRCLE of an ELLIPSE with center O and FOCUS F, and drawing a LINE PERPENDICULAR to the SEMIMAJOR AXIS and intersecting it at A. The ANGLE E is then defined as illustrated above. Then for an ELLIPSE with ECCENTRICITY e,

$$AF = OF - AO = ae - a\cos E. \qquad (1)$$

But the distance AF is also given in terms of the distance from the FOCUS $r = FP$ and the SUPPLEMENT of the ANGLE from the SEMIMAJOR AXIS v by

$$AF = r\cos(\pi - v) = -r\cos v. \qquad (2)$$

Equating these two expressions gives

$$r = \frac{a(\cos E - e)}{\cos v}, \tag{3}$$

which can be solved for $\cos v$ to obtain

$$\cos v = \frac{a(\cos E - e)}{r}. \tag{4}$$

To get E in terms of r, plug (4) into the equation of the ELLIPSE

$$r = \frac{a(1-e^2)}{1+e\cos v} \tag{5}$$

$$r(1+e\cos v) = a(1-e^2) \tag{6}$$

$$r\left(1 + \frac{ae\cos E}{r} - \frac{e^2}{r}\right) = r + ae\cos E - e^2 = a(1-e^2) \tag{7}$$

$$r = a(1-e^2) - ea\cos E + e^2 a = a(1 - e\cos E). \tag{8}$$

Differentiating gives

$$\dot{r} = ae\dot{E}\sin E. \tag{9}$$

The eccentric anomaly is a very useful concept in orbital mechanics, where it is related to the so-called mean anomaly M by KEPLER'S EQUATION

$$M = E - e\sin E. \tag{10}$$

M can also be interpreted as the AREA of the shaded region in the above figure (Finch).

see also ECCENTRICITY, ELLIPSE, KEPLER'S EQUATION

References

Danby, J. M. *Fundamentals of Celestial Mechanics, 2nd ed., rev. ed.* Richmond, VA: Willmann-Bell, 1988.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/lpc/lpc.html`.

Eccentricity

A quantity defined for a CONIC SECTION which can be given in terms of SEMIMAJOR and SEMIMINOR AXES for an ELLIPSE. For an ELLIPSE with SEMIMAJOR AXIS a and SEMIMINOR AXIS b,

$$e \equiv \sqrt{1 - \frac{b^2}{a^2}}.$$

The eccentricity can be interpreted as the fraction of the distance to the semimajor axis at which the FOCUS lies,

$$e = \frac{c}{a},$$

where c is the distance from the center of the CONIC SECTION to the FOCUS. The table below gives the type of CONIC SECTION corresponding to various ranges of eccentricity e.

e	Curve
$e = 0$	circle
$0 < e < 1$	ellipse
$e = 1$	parabola
$e > 1$	hyperbola

see also CIRCLE, CONIC SECTION, ECCENTRIC ANOMALY, ELLIPSE, FLATTENING, HYPERBOLA, OBLATENESS, PARABOLA, SEMIMAJOR AXIS, SEMIMINOR AXIS

Eccentricity (Graph)

The length of the longest shortest path from a VERTEX in a GRAPH.

see also DIAMETER (GRAPH)

Echidnahedron

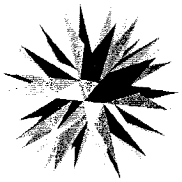

ICOSAHEDRON STELLATION #4.

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 65, 1971.

Eckardt Point

On the CLEBSCH DIAGONAL CUBIC, all 27 of the complex lines present on a general smooth CUBIC SURFACE are real. In addition, there are 10 points on the surface where three of the 27 lines meet. These points are called Eckardt points (Fischer 1986).

see also CLEBSCH DIAGONAL CUBIC, CUBIC SURFACE

References

Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, p. 11, 1986.

Eckert IV Projection

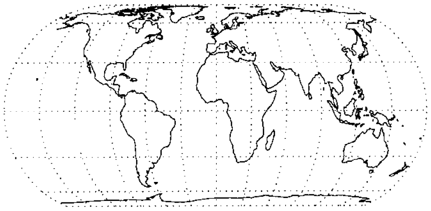

The equations are

$$x = \frac{2}{\sqrt{\pi(4+\pi)}}(\lambda - \lambda_0)(1+\cos\theta) \tag{1}$$

$$y = 2\sqrt{\frac{\pi}{4+\pi}}\sin\theta, \tag{2}$$

where θ is the solution to

$$\theta + \sin\theta\cos\theta + 2\sin\theta = (2 + \tfrac{1}{2}\pi)\sin\phi. \tag{3}$$

This can be solved iteratively using Newton's Method with $\theta_0 = \phi/2$ to obtain

$$\Delta\theta = -\frac{\theta + \sin\theta\cos\theta + 2\sin\theta - (2 - \frac{1}{2}\pi)\sin\phi}{2\cos\theta(1+\cos\theta)}. \tag{4}$$

The inverse Formulas are

$$\phi = \sin^{-1}\left(\frac{\theta + \sin\theta\cos\theta + 2\sin\theta}{2 + \frac{1}{2}\pi}\right) \tag{5}$$

$$\lambda = \lambda_0 + \frac{\pi\sqrt{4+\pi}\,x}{1+\cos\theta}, \tag{6}$$

where

$$\theta = \sin^{-1}\left(\frac{y}{2}\sqrt{\frac{4+\pi}{\pi}}\right). \tag{7}$$

References

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 253–258, 1987.

Eckert VI Projection

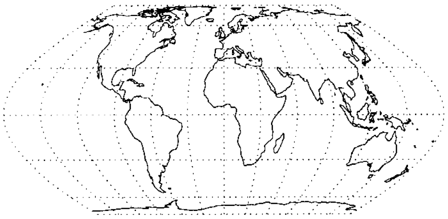

The equations are

$$x = \frac{(\lambda - \lambda_0)(1+\cos\theta)}{\sqrt{2+\pi}} \tag{1}$$

$$y = \frac{2\theta}{\sqrt{2+\pi}}, \tag{2}$$

where θ is the solution to

$$\theta + \sin\theta = (1 + \tfrac{1}{2}\pi)\sin\phi. \tag{3}$$

This can be solved iteratively using Newton's Method with $\theta_0 = \phi$ to obtain

$$\Delta\theta = -\frac{\theta + \sin\theta - (1 + \frac{1}{2}\pi)\sin\phi}{1+\cos\theta}. \tag{4}$$

The inverse Formulas are

$$\phi = \sin^{-1}\left(\frac{\theta + \sin\theta}{1 + \frac{1}{2}\pi}\right) \tag{5}$$

$$\lambda = \lambda_0 + \frac{\sqrt{2+\pi}\,x}{1+\cos\theta}, \tag{6}$$

where

$$\theta = \tfrac{1}{2}\sqrt{2+\pi}\,y. \tag{7}$$

References

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 253–258, 1987.

Economized Rational Approximation

A Padé Approximation perturbed with a Chebyshev Polynomial of the First Kind to reduce the leading Coefficient in the Error.

Eddington Number

$$136 \cdot 2^{256} \approx 1.575 \times 10^{79}.$$

According to Eddington, the exact number of protons in the universe, where 136 was the Reciprocal of the fine structure constant as best as it could be measured in his time.

see also Large Number

Edge-Coloring

An edge-coloring of a Graph G is a coloring of the edges of G such that adjacent edges (or the edges bounding different regions) receive different colors. Brelaz's Heuristic Algorithm can be used to find a good, but not necessarily minimal, edge-coloring.

see also Brelaz's Heuristic Algorithm, Chromatic Number, k-Coloring

References

Saaty, T. L. and Kainen, P. C. *The Four-Color Problem: Assaults and Conquest.* New York: Dover, p. 13, 1986.

Edge Connectivity

The minimum number of Edges whose deletion from a Graph disconnects it.

see also Vertex Connectivity

Edge (Graph)

For an undirected Graph, an unordered pair of nodes which specify the line connecting them. For a Directed Graph, the edge is an ordered pair of nodes.

see also Edge Number, Null Graph, Tait Coloring, Tait Cycle, Vertex (Graph)

Edge Number

The number of Edges in a Graph, denoted $|E|$.

see also Edge (Graph)

Edge (Polygon)

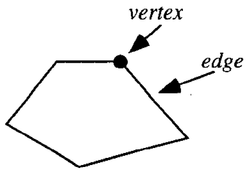

A Line Segment on the boundary of a Face, also called a Side.

see also Edge (Polyhedron), Vertex (Polygon)

Edge (Polyhedron)

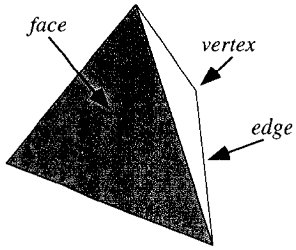

A LINE SEGMENT where two FACES of a POLYHEDRON meet, also called a SIDE.

see also EDGE (POLYGON), VERTEX (POLYHEDRON)

Edge (Polytope)

A 1-D LINE SEGMENT where two 2-D FACES of an n-D POLYTOPE meet, also called a SIDE.

see also EDGE (POLYGON), EDGE (POLYHEDRON)

Edgeworth Series

Approximate a distribution in terms of a NORMAL DISTRIBUTION. Let

$$\phi(t) \equiv \frac{1}{\sqrt{2\pi}} e^{-t^2/2},$$

then

$$f(t) = \phi(t) + \tfrac{1}{3!}\gamma_1 \phi^{(3)}(t) + \left[\frac{\gamma_2}{4!}\phi^{(4)}(t) + \frac{10{\gamma_1}^2}{6!}\phi^{(6)}(t)\right] + \ldots.$$

see also CORNISH-FISHER ASYMPTOTIC EXPANSION, GRAM-CHARLIER SERIES

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 935, 1972.

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, p. 108, 1951.

Edmonds' Map

A nonreflexible regular map of GENUS 7 with eight VERTICES, 28 EDGES, and eight HEPTAGONAL faces.

Efron's Dice

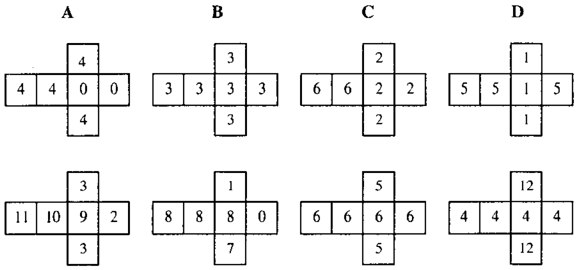

A set of four nontransitive DICE such that the probabilities of A winning against B, B against C, C against D, and D against A are all 2:1. A set in which ties may occur, in which case the DICE are rolled again, which gives ODDS of 11:6 is

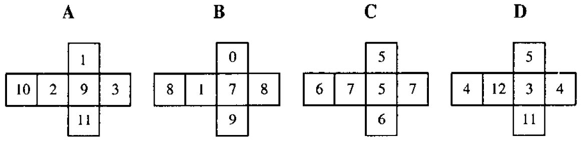

see also DICE, SICHERMAN DICE

References

Gardner, M. "Mathematical Games: The Paradox of the Nontransitive Dice and the Elusive Principle of Indifference." *Sci. Amer.* **223**, 110–114, Dec. 1970.

Honsberger, R. "Some Surprises in Probability." Ch. 5 in *Mathematical Plums* (Ed. R. Honsberger). Washington, DC: Math. Assoc. Amer., pp. 94–97, 1979.

Egg

An OVAL with one end more pointed than the other.

see also ELLIPSE, MOSS'S EGG, OVAL, OVOID, THOM'S EGGS

Egyptian Fraction

see UNIT FRACTION

Ehrhart Polynomial

Let Δ denote an integral convex POLYTOPE of DIMENSION n in a lattice M, and let $l_\Delta(k)$ denote the number of LATTICE POINTS in Δ dilated by a factor of the integer k,

$$l_\Delta(k) = \#(k\Delta \cap M) \tag{1}$$

for $k \in \mathbb{Z}^+$. Then l_Δ is a polynomial function in k of degree n with rational coefficients

$$l_\Delta(k) = a_n k^n + a_{n-1} k^{n-1} + \ldots + a_0 \tag{2}$$

called the Ehrhart polynomial (Ehrhart 1967, Pommersheim 1993). Specific coefficients have important geometric interpretations.

1. a_n is the CONTENT of Δ.
2. a_{n-1} is half the sum of the CONTENTS of the $(n-1)$-D faces of Δ.
3. $a_0 = 1$.

Let $S_2(\Delta)$ denote the sum of the lattice lengths of the edges of Δ, then the case $n = 2$ corresponds to PICK'S THEOREM,

$$l_\Delta(k) = \mathrm{Vol}(\Delta)k^2 + \tfrac{1}{2}S_2(\Delta) + 1. \tag{3}$$

Let $S_3(\Delta)$ denote the sum of the lattice volumes of the 2-D faces of Δ, then the case $n = 3$ gives

$$l_\Delta(k) = \mathrm{Vol}(\Delta)k^3 + \tfrac{1}{2}S_3(\Delta)k^2 + a_1 k + 1, \tag{4}$$

where a rather complicated expression is given by Pommersheim (1993), since a_1 can unfortunately *not* be interpreted in terms of the edges of Δ. The Ehrhart polynomial of the tetrahedron with vertices at (0, 0, 0), (a, 0, 0), (0, b, 0), (0, 0, c) is

$$\begin{aligned} l_\Delta(k) = {} & \tfrac{1}{6}abck^3 + \tfrac{1}{4}(ab+ac+bc+d)k^2 \\ & + \left[\frac{1}{12}\left(\frac{ac}{b}+\frac{bc}{a}+\frac{ab}{c}+\frac{d^2}{abc}\right)\right. \\ & + \tfrac{1}{4}(a+b+c+A+B+C) - As\left(\frac{bc}{d},\frac{aA}{d}\right) \\ & \left. - Bs\left(\frac{ac}{d},\frac{bB}{d}\right) - Cs\left(\frac{ab}{d},\frac{cC}{d}\right)\right]k+1, \end{aligned} \qquad (5)$$

where $s(x,y)$ is a DEDEKIND SUM, $A = \gcd(b,c)$, $B = \gcd(a,c)$, $C = \gcd(a,b)$ (here, gcd is the GREATEST COMMON DENOMINATOR), and $d = ABC$ (Pommersheim 1993).

see also DEHN INVARIANT, PICK'S THEOREM

References

Ehrhart, E. "Sur une problème de géométric diophantine linéaire." *J. Reine angew. Math.* **227**, 1–29, 1967.

MacDonald, I. G. "The Volume of a Lattice Polyhedron." *Proc. Camb. Phil. Soc.* **59**, 719–726, 1963.

McMullen, P. "Valuations and Euler-Type Relations on Certain Classes of Convex Polytopes." *Proc. London Math. Soc.* **35**, 113–135, 1977.

Pommersheim, J. "Toric Varieties, Lattices Points, and Dedekind Sums." *Math. Ann.* **295**, 1–24, 1993.

Reeve, J. E. "On the Volume of Lattice Polyhedra." *Proc. London Math. Soc.* **7**, 378–395, 1957.

Reeve, J. E. "A Further Note on the Volume of Lattice Polyhedra." *Proc. London Math. Soc.* **34**, 57–62, 1959.

Ei

see EXPONENTIAL INTEGRAL, E_n-FUNCTION

Eigenfunction

If $\tilde{L}$ is a linear OPERATOR on a FUNCTION SPACE, then f is an eigenfunction for $\tilde{L}$ and λ is the associated EIGENVALUE whenever $\tilde{L}f = \lambda f$.

see also EIGENVALUE, EIGENVECTOR

Eigenvalue

Let A be a linear transformation represented by a MATRIX A. If there is a VECTOR $\mathbf{X} \in \mathbb{R}^n \neq \mathbf{0}$ such that

$$\mathsf{A}\mathbf{X} = \lambda\mathbf{X} \qquad (1)$$

for some SCALAR λ, then λ is the eigenvalue of A with corresponding (right) EIGENVECTOR $\mathbf{X}$. Letting A be a $k \times k$ MATRIX,

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix} \qquad (2)$$

with eigenvalue λ, then the corresponding EIGENVECTORS satisfy

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}, \qquad (3)$$

which is equivalent to the homogeneous system

$$\begin{bmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk}-\lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \qquad (4)$$

Equation (4) can be written compactly as

$$(\mathsf{A} - \lambda\mathsf{I})\mathbf{X} = \mathbf{0}, \qquad (5)$$

where I is the IDENTITY MATRIX.

As shown in CRAMER'S RULE, a system of linear equations has nontrivial solutions only if the DETERMINANT vanishes, so we obtain the CHARACTERISTIC EQUATION

$$|\mathsf{A} - \lambda\mathsf{I}| = 0. \qquad (6)$$

If all k λs are different, then plugging these back in gives $k-1$ independent equations for the k components of each corresponding EIGENVECTOR. The EIGENVECTORS will then be orthogonal and the system is said to be nondegenerate. If the eigenvalues are n-fold DEGENERATE, then the system is said to be degenerate and the EIGENVECTORS are not linearly independent. In such cases, the additional constraint that the EIGENVECTORS be orthogonal,

$$\mathbf{X}_i \cdot \mathbf{X}_j = X_i X_j \delta_{ij}, \qquad (7)$$

where δ_{ij} is the KRONECKER DELTA, can be applied to yield n additional constraints, thus allowing solution for the EIGENVECTORS.

Assume A has nondegenerate eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and corresponding linearly independent EIGENVECTORS $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k$ which can be denoted

$$\begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1k} \end{bmatrix}, \begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2k} \end{bmatrix}, \ldots \begin{bmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kk} \end{bmatrix}. \qquad (8)$$

Define the matrices composed of eigenvectors

$$\mathsf{P} \equiv [\mathbf{X}_1 \quad \mathbf{X}_2 \quad \cdots \quad \mathbf{X}_k] = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{kk} \end{bmatrix} \qquad (9)$$

and eigenvalues

$$\mathsf{D} \equiv \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix}, \tag{10}$$

where D is a Diagonal Matrix. Then

$$\begin{aligned} \mathsf{AP} &= \mathsf{A}[\mathbf{X}_1 \quad \mathbf{X}_2 \quad \cdots \quad \mathbf{X}_k] \\ &= [\mathsf{A}\mathbf{X}_1 \quad \mathsf{A}\mathbf{X}_2 \quad \cdots \quad \mathsf{A}\mathbf{X}_k] \\ &= [\lambda_1\mathbf{X}_1 \quad \lambda_2\mathbf{X}_2 \quad \cdots \quad \lambda_k\mathbf{X}_k] \\ &= \begin{bmatrix} \lambda_1 x_{11} & \lambda_2 x_{21} & \cdots & \lambda_k x_{k1} \\ \lambda_1 x_{12} & \lambda_2 x_{22} & \cdots & \lambda_k x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 x_{1k} & \lambda_2 x_{2k} & \cdots & \lambda_k x_{kk} \end{bmatrix} \\ &= \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{kk} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix} \\ &= \mathsf{PD}, \end{aligned} \tag{11}$$

so

$$\mathsf{A} = \mathsf{PDP}^{-1}. \tag{12}$$

Furthermore,

$$\begin{aligned} \mathsf{A}^2 &= (\mathsf{PDP}^{-1})(\mathsf{PDP}^{-1}) = \mathsf{PD}(\mathsf{P}^{-1}\mathsf{P})\mathsf{DP}^{-1} \\ &= \mathsf{PD}^2\mathsf{P}^{-1}. \end{aligned} \tag{13}$$

By induction, it follows that for $n > 0$,

$$\mathsf{A}^n = \mathsf{PD}^n\mathsf{P}^{-1}. \tag{14}$$

The inverse of A is

$$\mathsf{A}^{-1} = (\mathsf{PDP}^{-1})^{-1} = \mathsf{PD}^{-1}\mathsf{P}^{-1}, \tag{15}$$

where the inverse of the Diagonal Matrix D is trivially given by

$$\mathsf{D}^{-1} = \frac{1}{k}\begin{bmatrix} {\lambda_1}^{-1} & 0 & \cdots & 0 \\ 0 & {\lambda_2}^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & {\lambda_k}^{-1} \end{bmatrix}. \tag{16}$$

Equation (14) therefore holds for both Positive and Negative n.

A further remarkable result involving the matrices P and D follows from the definition

$$\begin{aligned} e^{\mathsf{A}} &\equiv \sum_{n=0}^{\infty} \frac{\mathsf{A}^n}{n!} = \sum_{n=0}^{\infty} \frac{\mathsf{PD}^n\mathsf{P}^{-1}}{n!} \\ &= \mathsf{P}\left(\frac{\sum_{n=0}^{\infty}\mathsf{D}^n}{n!}\right)\mathsf{P}^{-1} = \mathsf{P}e^{\mathsf{D}}\mathsf{P}^{-1}. \end{aligned} \tag{17}$$

Since D is a Diagonal Matrix,

$$\begin{aligned} e^{\mathsf{D}} &= \sum_{n=0}^{\infty} \frac{\mathsf{D}^n}{n!} = \sum_{n=0}^{\infty} \frac{1}{n!} \begin{bmatrix} {\lambda_1}^n & 0 & \cdots & 0 \\ 0 & {\lambda_2}^n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & {\lambda_k}^n \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=0}^{\infty} {\lambda_1}^i & 0 & \cdots & 0 \\ 0 & \sum_{i=0}^{\infty} {\lambda_2}^i & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sum_{i=0}^{\infty} {\lambda_k}^i \end{bmatrix} \\ &= \begin{bmatrix} e^{\lambda_1} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\lambda_k} \end{bmatrix}, \end{aligned} \tag{18}$$

e^{D} can be found using

$$\mathsf{D}^n = \begin{bmatrix} {\lambda_1}^n & 0 & \cdots & 0 \\ 0 & {\lambda_2}^n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & {\lambda_k}^n \end{bmatrix}. \tag{19}$$

Assume we know the eigenvalue for

$$\mathsf{A}\mathbf{X} = \lambda\mathbf{X}. \tag{20}$$

Adding a constant times the Identity Matrix to A,

$$(\mathsf{A} + c\,\mathsf{I})\mathbf{X} = (\lambda + c)\mathbf{X} \equiv \lambda'\mathbf{X}, \tag{21}$$

so the new eigenvalues equal the old plus c. Multiplying A by a constant c

$$(c\,\mathsf{A})\mathbf{X} = c\,(\lambda\mathbf{X}) \equiv \lambda'\mathbf{X}, \tag{22}$$

so the new eigenvalues are the old multiplied by c.

Now consider a Similarity Transformation of A. Let $|\mathsf{A}|$ be the Determinant of A, then

$$\begin{aligned} |\mathsf{Z}^{-1}\mathsf{AZ} - \lambda\mathsf{I}| &= |\mathsf{Z}^{-1}(\mathsf{A} - \lambda\mathsf{I})\mathsf{Z}| \\ &= |\mathsf{Z}|\,|\mathsf{A} - \lambda\mathsf{I}|\,|\mathsf{Z}^{-1}| = |\mathsf{A} - \lambda\mathsf{I}|, \end{aligned} \tag{23}$$

so the eigenvalues are the same as for A.

see also Brauer's Theorem, Condition Number, Eigenfunction, Eigenvector, Frobenius Theorem, Geršgorin Circle Theorem, Lyapunov's First Theorem, Lyapunov's Second Theorem, Ostrowski's Theorem, Perron's Theorem, Perron-Frobenius Theorem, Poincaré Separation Theorem, Random Matrix, Schur's Inequalities, Sturmian Separation Theorem, Sylvester's Inertia Law, Wielandt's Theorem

References

Arfken, G. "Eigenvectors, Eigenvalues." §4.7 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 229–237, 1985.

Nash, J. C. "The Algebraic Eigenvalue Problem." Ch. 9 in *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation, 2nd ed.* Bristol, England: Adam Hilger, pp. 102–118, 1990.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Eigensystems." Ch. 11 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 449–489, 1992.

Eigenvector

A right eigenvector satisfies

$$\mathsf{A}\mathbf{X} = \lambda \mathbf{X}, \tag{1}$$

where $\mathbf{X}$ is a column VECTOR. The right EIGENVALUES therefore satisfy

$$|\mathsf{A} - \lambda \mathsf{I}| = 0. \tag{2}$$

A left eigenvector satisfies

$$\mathbf{X}\mathsf{A} = \lambda \mathbf{X}, \tag{3}$$

where $\mathbf{X}$ is a row VECTOR, so

$$(\mathbf{X}\mathsf{A})^{\mathrm{T}} = \lambda_L \mathbf{X}^{\mathrm{T}} \tag{4}$$

$$\mathsf{A}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}} = \lambda_L \mathbf{X}^{\mathrm{T}}, \tag{5}$$

where $\mathbf{X}^{\mathrm{T}}$ is the transpose of $\mathbf{X}$. The left EIGENVALUES satisfy

$$|\mathsf{A}^{\mathrm{T}} - \lambda_L \mathsf{I}| = |\mathsf{A}^{\mathrm{T}} - \lambda_L \mathsf{I}^{\mathrm{T}}| = |(\mathsf{A} - \lambda_L \mathsf{I})^{\mathrm{T}}| = |(\mathsf{A} - \lambda_L \mathsf{I})|, \tag{6}$$

(since $|\mathsf{A}| = |\mathsf{A}^{\mathrm{T}}|$) where $|\mathsf{A}|$ is the DETERMINANT of A. But this is the same equation satisfied by the right EIGENVALUES, so the left and right EIGENVALUES are the same. Let $\mathbf{X}_R$ be a MATRIX formed by the columns of the right eigenvectors and $\mathbf{X}_L$ be a MATRIX formed by the rows of the left eigenvectors. Let

$$\mathsf{D} \equiv \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}. \tag{7}$$

Then

$$\mathsf{A}\mathbf{X}_R = \mathbf{X}_R\mathsf{D} \qquad \mathbf{X}_L\mathsf{A} = \mathsf{D}\mathbf{X}_L \tag{8}$$

$$\mathbf{X}_L\mathsf{A}\mathbf{X}_R = \mathbf{X}_L\mathbf{X}_R\mathsf{D} \qquad \mathbf{X}_L\mathsf{A}\mathbf{X}_R = \mathsf{D}\mathbf{X}_L\mathbf{X}_R, \tag{9}$$

so

$$\mathbf{X}_L\mathbf{X}_R\mathsf{D} = \mathsf{D}\mathbf{X}_L\mathbf{X}_R. \tag{10}$$

But this equation is of the form $\mathsf{CD} = \mathsf{DC}$ where D is a DIAGONAL MATRIX, so it must be true that $\mathsf{C} \equiv \mathbf{X}_L\mathbf{X}_R$ is also diagonal. In particular, if A is a SYMMETRIC MATRIX, then the left and right eigenvectors are transposes of each other. If A is a SELF-ADJOINT MATRIX, then the left and right eigenvectors are conjugate HERMITIAN MATRICES.

Given a 3×3 MATRIX A with eigenvectors $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ and corresponding EIGENVALUES λ_1, λ_2, and λ_3, then an arbitrary VECTOR $\mathbf{y}$ can be written

$$\mathbf{y} = b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + b_3\mathbf{x}_3. \tag{11}$$

Applying the MATRIX A,

$$\begin{aligned} \mathsf{A}\mathbf{y} &= b_1\mathsf{A}\mathbf{x}_1 + b_2\mathsf{A}\mathbf{x}_2 + b_3\mathsf{A}\mathbf{x}_3 \\ &= \lambda_1\left(b_1\mathbf{x}_1 + \frac{\lambda_2}{\lambda_1}b_2\mathbf{x}_2 + \frac{\lambda_3}{\lambda_1}b_3\mathbf{x}_3\right), \end{aligned} \tag{12}$$

so

$$\mathsf{A}^n\mathbf{y} = {\lambda_1}^n\left[b_1\mathbf{x}_1 + \left(\frac{\lambda_2}{\lambda_1}\right)^n b_2\mathbf{x}_2 + \left(\frac{\lambda_3}{\lambda_1}\right)^n b_3\mathbf{x}_3\right]. \tag{13}$$

If $\lambda_1 > \lambda_2, \lambda_3$, it therefore follows that

$$\lim_{n\to\infty} \mathsf{A}^n\mathbf{y} = {\lambda_1}^n b_1\mathbf{x}_1, \tag{14}$$

so repeated application of the matrix to an arbitrary vector results in a vector proportional to the EIGENVECTOR having the largest EIGENVALUE.

see also EIGENFUNCTION, EIGENVALUE

References

Arfken, G. "Eigenvectors, Eigenvalues." §4.7 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 229–237, 1985.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Eigensystems." Ch. 11 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 449–489, 1992.

Eight Curve

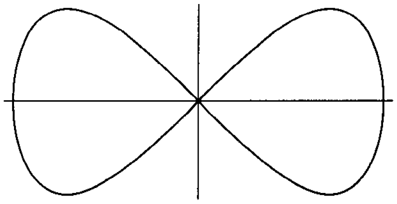

A curve also known as the GERONO LEMNISCATE. It is given by CARTESIAN COORDINATES

$$x^4 = a^2(x^2 - y^2), \tag{1}$$

POLAR COORDINATES,

$$r^2 = a^2 \sec^4\theta \cos(2\theta), \tag{2}$$

and parametric equations

$$x = a\sin t \tag{3}$$

$$y = a\sin t\cos t. \tag{4}$$

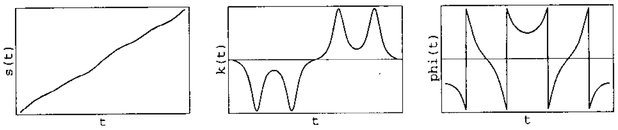

The CURVATURE and TANGENTIAL ANGLE are

$$\kappa(t) = -\frac{3\sin t + \sin(3t)}{2[\cos^2 t + \cos^2(2t)]^{3/2}} \tag{5}$$

$$\phi(t) = -\tan^{-1}[\cos t \sec(2t)]. \tag{6}$$

see also BUTTERFLY CURVE, DUMBBELL CURVE, EIGHT SURFACE, PIRIFORM

References
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 71, 1989.
Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 124–126, 1972.
Lee, X. "Lemniscate of Gerono." `http://www.best.com/~xah/SpecialPlaneCurves_dir/LemniscateOfGerono_dir/lemniscateOfGerono.html`.
MacTutor History of Mathematics Archive. "Eight Curve." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Eight.html`.

Eight-Point Circle Theorem

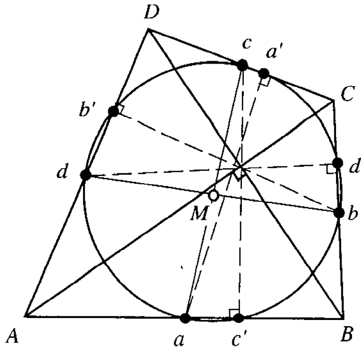

Let $ABCD$ be a QUADRILATERAL with PERPENDICULAR DIAGONALS. The MIDPOINTS of the sides (a, b, c, and d) determine a PARALLELOGRAM (the VARIGNON PARALLELOGRAM) with sides PARALLEL to the DIAGONALS. The eight-point circle passes through the four MIDPOINTS and the four feet of the PERPENDICULARS from the opposite sides a', b', c', and d'.

see also FEUERBACH'S THEOREM

References
Brand, L. "The Eight-Point Circle and the Nine-Point Circle." *Amer. Math. Monthly* **51**, 84–85, 1944.
Honsberger, R. *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., pp. 11–13, 1976.

Eight Surface

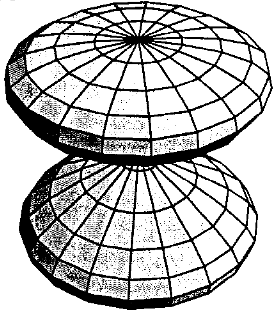

The SURFACE OF REVOLUTION given by the parametric equations

$$x(u,v) = \cos u \sin(2v) \tag{1}$$

$$y(u,v) = \sin u \sin(2v) \tag{2}$$

$$z(u,v) = \sin v \tag{3}$$

for $u \in [0, 2\pi)$ and $v \in [-\pi/2, \pi/2]$.

see also EIGHT CURVE

References
Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 209–210 and 224, 1993.

Eikonal Equation

$$\sum_{i=1}^{n}\left(\frac{\partial u}{\partial x_i}\right)^2 = 1.$$

Eilenberg-Mac Lane Space

For any ABELIAN GROUP G and any NATURAL NUMBER n, there is a unique SPACE (up to HOMOTOPY type) such that all HOMOTOPY GROUPS except for the nth are trivial (including the 0th HOMOTOPY GROUPS, meaning the SPACE is path-connected), and the nth HOMOTOPY GROUP is ISOMORPHIC to the GROUP G. In the case where $n = 1$, the GROUP G can be non-ABELIAN as well.

Eilenberg-Mac Lane spaces have many important applications. One of them is that every TOPOLOGICAL SPACE has the HOMOTOPY type of an iterated FIBRATION of Eilenberg-Mac Lane spaces (called a POSTNIKOV SYSTEM). In addition, there is a spectral sequence relating the COHOMOLOGY of Eilenberg-Mac Lane spaces to the HOMOTOPY GROUPS of SPHERES.

Eilenberg-Mac Lane-Steenrod-Milnor Axioms

see EILENBERG-STEENROD AXIOMS

Eilenberg-Steenrod Axioms

A family of FUNCTORS $H_n(\cdot)$ from the CATEGORY of pairs of TOPOLOGICAL SPACES and continuous maps, to the CATEGORY of ABELIAN GROUPS and group homomorphisms satisfies the Eilenberg-Steenrod axioms if the following conditions hold.

1. LONG EXACT SEQUENCE OF A PAIR AXIOM. For every pair (X, A), there is a natural long exact sequence

   $$\ldots \to H_n(A) \to H_n(X) \to H_n(X, A) \to H_{n-1}(A) \to \ldots,$$

   where the MAP $H_n(A) \to H_n(X)$ is induced by the INCLUSION MAP $A \to X$ and $H_n(X) \to H_n(X, A)$ is induced by the INCLUSION MAP $(X, \phi) \to (X, A)$. The MAP $H_n(X, A) \to H_{n-1}(A)$ is called the BOUNDARY MAP.
2. HOMOTOPY AXIOM. If $f : (X, A) \to (Y, B)$ is homotopic to $g : (X, A) \to (Y, B)$, then their INDUCED MAPS $f_* : H_n(X, A) \to H_n(Y, B)$ and $g_* : H_n(X, A) \to H_n(Y, B)$ are the same.
3. EXCISION AXIOM. If X is a SPACE with SUBSPACES A and U such that the CLOSURE of A is contained in the interior of U, then the INCLUSION MAP $(X\ U, A\ U) \to (X, A)$ induces an isomorphism $H_n(X\ U, A\ U) \to H_n(X, A)$.
4. DIMENSION AXIOM. Let X be a single point space. $H_n(X) = 0$ unless $n = 0$, in which case $H_0(X) = G$ where G are some GROUPS. The H_0 are called the COEFFICIENTS of the HOMOLOGY theory $H(\cdot)$.

These are the axioms for a generalized homology theory. For a cohomology theory, instead of requiring that $H(\cdot)$ be a FUNCTOR, it is required to be a co-functor (meaning the INDUCED MAP points in the opposite direction). With that modification, the axioms are essentially the same (except that all the induced maps point backwards).

see also ALEKSANDROV-ČECH COHOMOLOGY

Ein Function

$$\mathrm{Ein}(z) \equiv \int_0^z \frac{(1 - e^{-t})\,dt}{t} = \mathrm{E}_1(z) + \ln z + \gamma,$$

where γ is the EULER-MASCHERONI CONSTANT and E_1 is the E_n-FUNCTION with $n = 1$.

see also E_n-FUNCTION

Einstein Functions

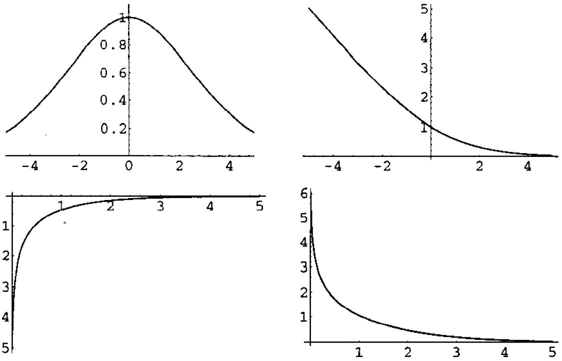

The functions $x^2e^x/(e^x - 1)^2$, $x/(e^x - 1)$, $\ln(1 - e^{-x})$, and $x/(e^x - 1) - \ln(1 - e^{-x})$.

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Debye Functions." §27.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 999–1000, 1972.

Einstein Summation

The implicit convention that repeated indices are summed over so that, for example,

$$a_i a_i \equiv \sum_i a_i a_i.$$

Eisenstein Integer

The numbers $a + b\omega$, where

$$\omega \equiv \tfrac{1}{2}(-1 + i\sqrt{3}\,)$$

is one of the ROOTS of $z^3 = 1$, the others being 1 and

$$\omega^2 = \tfrac{1}{2}(-1 - i\sqrt{3}\,).$$

Eisenstein integers are members of the QUADRATIC FIELD $\mathbb{Q}(\sqrt{-3}\,)$, and the COMPLEX NUMBERS $\mathbb{Z}[\omega]$. Every Eisenstein integer has a unique factorization. Specifically, any NONZERO Eisenstein integer is uniquely the product of POWERS of -1, ω, and the "positive" EISENSTEIN PRIMES (Conway and Guy 1996). Every Eisenstein integer is within a distance $|n|/\sqrt{3}$ of some multiple of a given Eisenstein integer n.

Dörrie (1965) uses the alternative notation

$$J \equiv \tfrac{1}{2}(1 + i\sqrt{3}\,) \tag{1}$$

$$O \equiv \tfrac{1}{2}(1 - i\sqrt{3}\,). \tag{2}$$

for $-\omega^2$ and $-\omega$, and calls numbers of the form $aJ+bO$ G-NUMBERS. O and J satisfy

$$J+O=1 \tag{3}$$

$$JO=1 \tag{4}$$

$$J^2+O=0 \tag{5}$$

$$O^2+J=0 \tag{6}$$

$$J^3=-1 \tag{7}$$

$$O^3=-1. \tag{8}$$

The sum, difference, and products of G numbers are also G numbers. The norm of a G number is

$$N(aJ+bO)=a^2+b^2-ab. \tag{9}$$

The analog of FERMAT'S THEOREM for Eisenstein integers is that a PRIME NUMBER p can be written in the form

$$a^2-ab+b^2=(a+b\omega)(a+b\omega^2)$$

IFF $3\nmid p+1$. These are precisely the PRIMES of the form $3m^2+n^2$ (Conway and Guy 1996).

see also EISENSTEIN PRIME, EISENSTEIN UNIT, GAUSSIAN INTEGER, INTEGER

References
Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 220–223, 1996.
Cox, D. A. §4A in *Primes of the Form x^2+ny^2: Fermat, Class Field Theory and Complex Multiplication.* New York: Wiley, 1989.
Dörrie, H. "The Fermat-Gauss Impossibility Theorem." §21 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 96–104, 1965.
Guy, R. K. "Gaussian Primes. Eisenstein-Jacobi Primes." §A16 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 33–36, 1994.
Riesel, H. Appendix 4 in *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, 1994.
Wagon, S. "Eisenstein Primes." *Mathematica in Action.* New York: W. H. Freeman, pp. 278–279, 1991.

Eisenstein-Jacobi Integer

see EISENSTEIN INTEGER

Eisenstein Prime

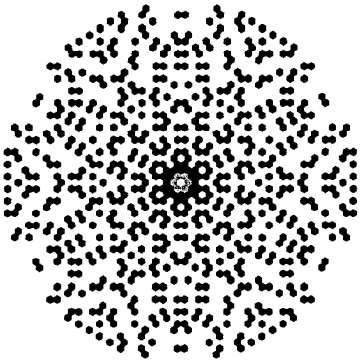

Let ω be the CUBE ROOT of unity $(-1+i\sqrt{3})/2$. Then the Eisenstein primes are

1. Ordinary PRIMES CONGRUENT to 2 (mod 3),
2. $1-\omega$ is prime in $\mathbb{Z}[\omega]$,
3. Any ordinary PRIME CONGRUENT to 1 (mod 3) factors as $\alpha\alpha^*$, where each of α and α^* are primes in $\mathbb{Z}[\omega]$ and α and α^* are not "associates" of each other (where associates are equivalent modulo multiplication by an EISENSTEIN UNIT).

References
Cox, D. A. §4A in *Primes of the Form x^2+ny^2: Fermat, Class Field Theory and Complex Multiplication.* New York: Wiley, 1989.
Guy, R. K. "Gaussian Primes. Eisenstein-Jacobi Primes." §A16 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 33–36, 1994.
Wagon, S. "Eisenstein Primes." *Mathematica in Action.* New York: W. H. Freeman, pp. 278–279, 1991.

Eisenstein Series

$$E_r(t)=\sum_{m,n}{}' \frac{1}{(mt+n)^{2r}},$$

where the sum Σ' excludes $m=n=0$, $\Im[t]>0$, and r is an INTEGER >2. The Eisenstein series satisfies the remarkable property

$$E_r\left(\frac{at+b}{ct+d}\right)=(ct+d)^{2r}E_r(t).$$

see also RAMANUJAN-EISENSTEIN SERIES

Eisenstein Unit

The Eisenstein units are the EISENSTEIN INTEGERS ±1, $\pm\omega$, $\pm\omega^2$, where

$$\omega=\tfrac{1}{2}(-1+i\sqrt{3})$$

$$\omega^2=\tfrac{1}{2}(-1-i\sqrt{3}).$$

see also EISENSTEIN INTEGER, EISENSTEIN PRIME

References
Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 220–223, 1996.

Elastica

The elastica formed by bent rods and considered in physics can be generalized to curves in a RIEMANNIAN MANIFOLD which are a CRITICAL POINT for

$$F^\lambda(\gamma)=\int_\gamma(\kappa^2+\lambda),$$

where κ is the GEODESIC CURVATURE of γ, λ is a REAL NUMBER, and γ is closed or satisfies some specified boundary condition. The curvature of an elastica must satisfy

$$0 = 2\kappa''(s) + \kappa^3(s) + 2\kappa(s)G(s) - \lambda\kappa(s),$$

where κ is the signed curvature of γ, $G(s)$ is the GAUSSIAN CURVATURE of the oriented Riemannian surface M along γ, κ'' is the second derivative of κ with respect to s, and λ is a constant.

References

Barros, M. and Garay, O. J. "Free Elastic Parallels in a Surface of Revolution." *Amer. Math. Monthly* **103**, 149–156, 1996.

Bryant, R. and Griffiths, P. "Reduction for Constrained Variational Problems and $\int(k^2/s)\,ds$." *Amer. J. Math.* **108**, 525–570, 1986.

Langer, J. and Singer, D. A. "Knotted Elastic Curves in R^3." *J. London Math. Soc.* **30**, 512–520, 1984.

Langer, J. and Singer, D. A. "The Total Squared of Closed Curves." *J. Diff. Geom.* **20**, 1–22, 1984.

Elation

A perspective COLLINEATION in which the center and axis are incident.

see also HOMOLOGY (GEOMETRY)

Elder's Theorem

A generalization of STANLEY'S THEOREM. It states that the total number of occurrences of an INTEGER k among all unordered PARTITIONS of n is equal to the number of occasions that a part occurs k or more times in a PARTITION, where a PARTITION which contains r parts that each occur k or more times contributes r to the sum in question.

see also STANLEY'S THEOREM

References

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer, pp. 8–9, 1985.

Election

see EARLY ELECTION RESULTS, VOTING

Electric Motor Curve

see DEVIL'S CURVE

Element

If x is a member of a set A, then x is said to be an element of A, written $x \in A$. If x is not an element of A, this is written $x \notin A$. The term element also refers to a particular member of a GROUP, or entry in a MATRIX.

Elementary Function

A function built up of compositions of the EXPONENTIAL FUNCTION and the TRIGONOMETRIC FUNCTIONS and their inverses by ADDITION, MULTIPLICATION, DIVISION, root extractions (the ELEMENTARY OPERATIONS) under repeated compositions. Not all functions are elementary. For example, the NORMAL DISTRIBUTION FUNCTION

$$\Phi(x) \equiv \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2}\,dt$$

is a notorious example of a nonelementary function. Nonelementary functions are called TRANSCENDENTAL FUNCTIONS.

see also ALGEBRAIC FUNCTION, ELEMENTARY OPERATION, ELEMENTARY SYMMETRIC FUNCTION, TRANSCENDENTAL FUNCTION

References

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 145, 1993.

Elementary Matrix

The elementary MATRICES are the PERMUTATION MATRIX p_{ij} and the SHEAR MATRIX e_{ij}^s.

Elementary Operation

One of the operations of ADDITION, SUBTRACTION, MULTIPLICATION, DIVISION, and root extraction.

see also ALGEBRAIC FUNCTION, ELEMENTARY FUNCTION

Elementary Symmetric Function

The elementary symmetric functions Π_n on n variables $\{x_1, \ldots, x_n\}$ are defined by

$$\Pi_1 = \sum_{1\leq i\leq n} x_i \qquad (1)$$

$$\Pi_2 = \sum_{1\leq i<j\leq n} x_i x_j \qquad (2)$$

$$\Pi_3 = \sum_{1\leq i<j<k\leq n} x_i x_j x_k \qquad (3)$$

$$\Pi_4 = \sum_{1\leq i<j<k<l\leq n} x_i x_j x_k x_l \qquad (4)$$

$$\vdots$$

$$\Pi_n = \prod_{i\leq i\leq n} x_i. \qquad (5)$$

Alternatively, Π_j can be defined as the coefficient of x^{n-j} in the GENERATING FUNCTION

$$\prod_{1\leq i\leq n} (x + x_i). \qquad (6)$$

The elementary symmetric functions satisfy the relationships

$$\sum_{i=1}^{n} {x_i}^2 = {\Pi_1}^2 - 2\Pi_2 \tag{7}$$

$$\sum_{i=1}^{n} {x_i}^3 = {\Pi_1}^3 - 3\Pi_1\Pi_2 + 3\Pi_3 \tag{8}$$

$$\sum_{i=1}^{n} {x_i}^4 = {\Pi_1}^4 - 4{\Pi_1}^2\Pi_2 + 2{\Pi_2}^2 + 4\Pi_1\Pi_3 - 4\Pi_4 \tag{9}$$

(Beeler *et al.* 1972, Item 6).

see also FUNDAMENTAL THEOREM OF SYMMETRIC FUNCTIONS, NEWTON'S RELATIONS, SYMMETRIC FUNCTION

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Elements

The classic treatise in geometry written by Euclid and used as a textbook for more than 1,000 years in western Europe. *The Elements*, which went through more than 2,000 editions and consisted of 465 propositions, are divided into 13 "books" (an archaic word for "chapters").

Book	Contents
1	triangles
2	rectangles
3	Circles
4	polygons
5	proportion
6	similarity
7–10	number theory
11	solid geometry
12	pyramids
13	platonic solids

The elements started with 23 definitions, five POSTULATES, and five "common notions," and systematically built the rest of plane and solid geometry upon this foundation. The five EUCLID'S POSTULATES are

1. It is possible to draw a straight LINE from any POINT to another POINT.
2. It is possible to produce a finite straight LINE continuously in a straight LINE.
3. It is possible to describe a CIRCLE with any CENTER and RADIUS.
4. All RIGHT ANGLES are equal to one another.
5. If a straight LINE falling on two straight LINES makes the interior ANGLES on the same side less than two RIGHT ANGLES, the straight LINES (if extended indefinitely) meet on the side on which the ANGLES which are less than two RIGHT ANGLES lie.

(Dunham 1990). Euclid's fifth postulate is known as the PARALLEL POSTULATE. After more than two millennia of study, this POSTULATE was found to be independent of the others. In fact, equally valid NON-EUCLIDEAN GEOMETRIES were found to be possible by changing the assumption of this POSTULATE. Unfortunately, Euclid's postulates were not rigorously complete and left a large number of gaps. Hilbert needed a total of 20 postulates to construct a logically complete geometry.

see also PARALLEL POSTULATE

References

Casey, J. *A Sequel to the First Six Books of the Elements of Euclid, 6th ed.* Dublin: Hodges, Figgis, & Co., 1892.

Dixon, R. *Mathographics.* New York: Dover, pp. 26–27, 1991.

Dunham, W. *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 30–83, 1990.

Heath, T. L. *The Thirteen Books of the Elements, 2nd ed., Vol. 1: Books I and II.* New York: Dover, 1956.

Heath, T. L. *The Thirteen Books of the Elements, 2nd ed., Vol. 2: Books III–IX.* New York: Dover, 1956.

Heath, T. L. *The Thirteen Books of the Elements, 2nd ed., Vol. 3: Books X–XIII.* New York: Dover, 1956.

Joyce, D. E. "Euclid's Elements." `http://aleph0.clarku.edu/~djoyce/java/elements/elements.html`

Elevator Paradox

A fact noticed by physicist G. Gamow when he had an office on the second floor and physicist M. Stern had an office on the sixth floor of a seven-story building (Gamow and Stern 1958, Gardner 1986). Gamow noticed that about 5/6 of the time, the first elevator to stop on his floor was going down, whereas about the same fraction of time, the first elevator to stop on the sixth floor was going up. This actually makes perfect sense, since 5 of the 6 floors 1, 3, 4, 5, 6, 7 are above the second, and 5 of the 6 floors 1, 2, 3, 4, 5, 7 are below the sixth. However, the situation takes some unexpected turns if more than one elevator is involved, as discussed by Gardner (1986).

References

Gamow, G. and Stern, M. *Puzzle Math.* New York: Viking, 1958.

Gardner, M. "Elevators." Ch. 10 in *Knotted Doughnuts and Other Mathematical Entertainments.* New York: W. H. Freeman, pp. 123–132, 1986.

Elkies Point

Given POSITIVE numbers s_a, s_b, and s_c, the Elkies point is the unique point Y in the interior of a TRIANGLE ΔABC such that the respective INRADII r_a, r_b, r_c of the TRIANGLES ΔBYC, ΔCYA, and ΔAYB satisfy $r_a : r_b : r_c = s_a : s_b : s_c$.

see also CONGRUENT INCIRCLES POINT, INRADIUS

References

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Kimberling, C. and Elkies, N. "Problem 1238 and Solution." *Math. Mag.* **60**, 116–117, 1987.

Ellipse

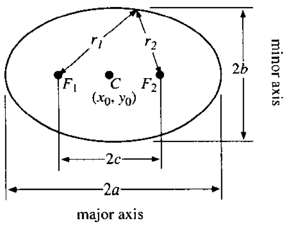

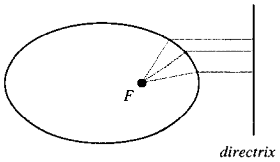

A curve which is the LOCUS of all points in the PLANE the SUM of whose distances r_1 and r_2 from two fixed points F_1 and F_2 (the FOCI) separated by a distance of $2c$ is a given POSITIVE constant $2a$ (left figure). This results in the two-center BIPOLAR COORDINATE equation

$$r_1 + r_2 = 2a, \qquad (1)$$

where a is the SEMIMAJOR AXIS and the ORIGIN of the coordinate system is at one of the FOCI. The ellipse can also be defined as the LOCUS of points whose distance from the FOCUS is proportional to the horizontal distance from a vertical line known as the DIRECTRIX (right figure).

The ellipse was first studied by Menaechmus, investigated by Euclid, and named by Apollonius. The FOCUS and DIRECTRIX of an ellipse were considered by Pappus. In 1602, Kepler believed that the orbit of Mars was OVAL; he later discovered that it was an ellipse with the Sun at one FOCUS. In fact, Kepler introduced the word "FOCUS" and published his discovery in 1609. In 1705 Halley showed that the comet which is now named after him moved in an elliptical orbit around the Sun (MacTutor Archive).

A ray passing through a FOCUS will pass through the other focus after a single bounce. Reflections not passing through a FOCUS will be tangent to a confocal HYPERBOLA or ELLIPSE, depending on whether the ray passes between the FOCI or not. Let an ellipse lie along the x-AXIS and find the equation of the figure (1) where r_1 and r_2 are at $(-c, 0)$ and $(c, 0)$. In CARTESIAN COORDINATES,

$$\sqrt{(x+c)^2 + y^2} + \sqrt{(x-c)^2 + y^2} = 2a. \qquad (2)$$

Bring the second term to the right side and square both sides,

$$(x+c)^2 + y^2 = 4a^2 - 4a\sqrt{(x-c)^2 + y^2} + (x-c)^2 + y^2. \quad (3)$$

Now solve for the SQUARE ROOT term and simplify

$$\begin{aligned}&\sqrt{(x-c)^2 + y^2}\\ &= -\frac{1}{4a}(x^2 + 2xc + c^2 + y^2 - 4a^2 - x^2 + 2xc - c^2 - y^2)\\ &= -\frac{1}{4a}(4xc - 4a^2) = a - \frac{c}{a}x. \qquad (4)\end{aligned}$$

Square one final time to clear the remaining SQUARE ROOT,

$$x^2 - 2xc + c^2 + y^2 = a^2 - 2cx + \frac{c^2}{a^2}x^2. \qquad (5)$$

Grouping the x terms then gives

$$x^2\frac{a^2 - c^2}{a^2} + y^2 = a^2 - c^2, \qquad (6)$$

which can be written in the simple form

$$\frac{x^2}{a^2} + \frac{y^2}{a^2 - c^2} = 1. \qquad (7)$$

Defining a new constant

$$b^2 \equiv a^2 - c^2 \qquad (8)$$

puts the equation in the particularly simple form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \qquad (9)$$

The parameter b is called the SEMIMINOR AXIS by analogy with the parameter a, which is called the SEMIMAJOR AXIS. The fact that b as defined above is actually the SEMIMINOR AXIS is easily shown by letting r_1 and r_2 be equal. Then two RIGHT TRIANGLES are produced, each with HYPOTENUSE a, base c, and height $b \equiv \sqrt{a^2 - c^2}$. Since the largest distance along the MINOR AXIS will be achieved at this point, b is indeed the SEMIMINOR AXIS.

If, instead of being centered at (0, 0), the CENTER of the ellipse is at (x_0, y_0), equation (9) becomes

$$\frac{(x - x_0)^2}{a^2} + \frac{(y - y_0)^2}{b^2} = 1. \qquad (10)$$

As can be seen from the CARTESIAN EQUATION for the ellipse, the curve can also be given by a simple parametric form analogous to that of a CIRCLE, but with the x and y coordinates having different scalings,

$$x = a\cos t \qquad (11)$$

$$y = b\sin t. \qquad (12)$$

The unit TANGENT VECTOR of the ellipse so parameterized is

$$x_T(t) = -\frac{a\sin t}{\sqrt{b^2\cos^2 t + a^2\sin^2 t}} \qquad (13)$$

$$y_T(t) = \frac{b\cos t}{\sqrt{b^2\cos^2 t + a^2\sin^2 t}}. \qquad (14)$$

A sequence of NORMAL and TANGENT VECTORS are plotted below for the ellipse.

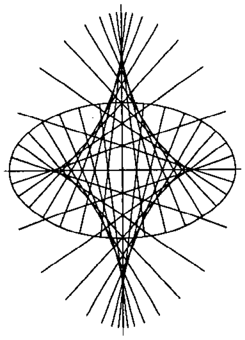

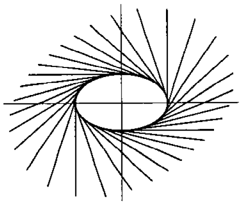

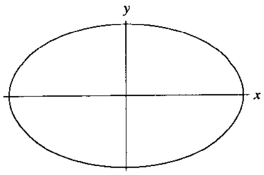

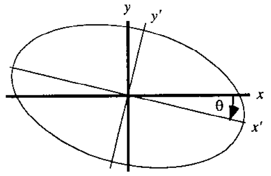

curve rotated by angle θ

For an ellipse centered at the ORIGIN but inclined at an arbitrary ANGLE θ to the x-AXIS, the parametric equations are

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} a\cos t \\ b\sin t \end{bmatrix} = \begin{bmatrix} a\cos\theta\cos t + b\sin\theta\sin t \\ -a\sin\theta\cos t + b\cos\theta\sin t \end{bmatrix}. \tag{15}$$

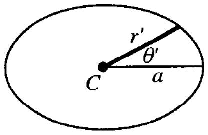

In POLAR COORDINATES, the ANGLE θ' measured from the *center* of the ellipse is called the ECCENTRIC ANGLE. Writing r' for the distance of a point from the ellipse center, the equation in POLAR COORDINATES is just given by the usual

$$x = r'\cos\theta' \tag{16}$$

$$y = r'\sin\theta'. \tag{17}$$

Here, the coordinates θ' and r' are written with primes to distinguish them from the more common polar coordinates for an ellipse which are centered on a *focus*. Plugging the polar equations into the Cartesian equation (9) and solving for r'^2 gives

$$r'^2 = \frac{b^2a^2}{b^2\cos^2\theta' + a^2\sin^2\theta'}. \tag{18}$$

Define a new constant $0 \leq e < 1$ called the ECCENTRICITY (where $e = 0$ is the case of a CIRCLE) to replace b

$$e \equiv \sqrt{1 - \frac{b^2}{a^2}}, \tag{19}$$

from which it also follows from (8) that

$$a^2e^2 = a^2 - b^2 \equiv c^2 \tag{20}$$

$$c = ae \tag{21}$$

$$b^2 = a^2(1 - e^2). \tag{22}$$

Therefore (18) can be written as

$$r'^2 = \frac{a^2(1 - e^2)}{1 - e^2\cos^2\theta'} \tag{23}$$

$$r' = a\sqrt{\frac{1 - e^2}{1 - e^2\cos^2\theta'}}. \tag{24}$$

If $e \ll 1$, then

$$r' = a\{1 - \tfrac{1}{2}e^2\sin^2\theta' - \tfrac{1}{16}e^4[5 + 3\cos(2\theta')]\sin^2\theta' + \ldots\}, \tag{25}$$

so

$$\frac{\Delta r'}{a} \equiv \frac{a - r'}{a} \approx \tfrac{1}{2}e^2\sin^2\theta'. \tag{26}$$

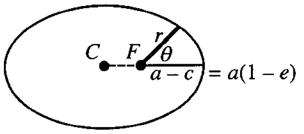

If r and θ are measured from a FOCUS instead of from the center, as they commonly are in orbital mechanics, then the equations of the ellipse are

$$x = c + r\cos\theta \tag{27}$$

$$y = r\sin\theta, \tag{28}$$

and (9) becomes

$$\frac{(c + r\cos\theta)^2}{a^2} + \frac{r^2\sin^2\theta}{b^2} = 1. \tag{29}$$

Clearing the DENOMINATORS gives

$$b^2(c^2 + 2cr\cos\theta + r^2\cos^2\theta) + a^2r^2\sin^2\theta = a^2b^2 \tag{30}$$

$$b^2c^2 + 2rcb^2\cos\theta + b^2r^2\cos^2\theta + a^2r^2 - a^2r^2\cos^2\theta = a^2b^2. \tag{31}$$

Plugging in (21) and (22) to re-express b and c in terms of a and e,

$$a^2(1 - e^2)a^2e^2 + 2aea^2(1 - e^2)r\cos\theta + a^2(1 - e^2)r^2\cos^2\theta + a^2r^2 - a^2r^2\cos^2\theta = a^2[a^2(1 - e^2)]. \tag{32}$$

Simplifying,

$$-r^2 + [er\cos\theta - a(1 - e^2)]^2 = 0 \tag{33}$$

$$r = \pm[er\cos\theta - a(1 - e^2)]. \tag{34}$$

The sign can be determined by requiring that r must be POSITIVE. When $e = 0$, (34) becomes $r = \pm(-a)$, but since a is always POSITIVE, we must take the NEGATIVE sign, so (34) becomes

$$r = a(1-e^2) - er\cos\theta \tag{35}$$

$$r(1+e\cos\theta) = a(1-e^2) \tag{36}$$

$$r = \frac{a(1-e^2)}{1+e\cos\theta}. \tag{37}$$

The distance from a FOCUS to a point with horizontal coordinate x is found from

$$\cos\theta = \frac{c+x}{r}. \tag{38}$$

Plugging this into (37) yields

$$r + e(c+x) = a(1-e^2) \tag{39}$$

$$r = a(1-e^2) - e(c+x). \tag{40}$$

Summarizing relationships among the parameters characterizing an ellipse,

$$b = a\sqrt{1-e^2} = \sqrt{a^2-c^2} \tag{41}$$

$$c = \sqrt{a^2-b^2} = ae \tag{42}$$

$$e = \sqrt{1-\frac{b^2}{a^2}} = \frac{c}{a}. \tag{43}$$

The ECCENTRICITY can therefore be interpreted as the position of the FOCUS as a fraction of the SEMIMAJOR AXIS.

In PEDAL COORDINATES with the PEDAL POINT at the FOCUS, the equation of the ellipse is

$$\frac{b^2}{p^2} = \frac{2a}{r} - 1. \tag{44}$$

To find the RADIUS OF CURVATURE, return to the parametric coordinates centered at the center of the ellipse and compute the first and second derivatives,

$$x' = -a\sin t \tag{45}$$

$$y' = b\cos t \tag{46}$$

$$x'' = -a\cos t \tag{47}$$

$$y'' = -b\sin t. \tag{48}$$

Therefore,

$$\begin{aligned} R &= \frac{(x'^2+y'^2)^{3/2}}{x'y''-x''y'} \\ &= \frac{(a^2\sin^2 t + b^2\cos^2 t)^{3/2}}{-a\sin t(-b\sin t) - (a\cos t)(b\cos t)} \\ &= \frac{(a^2\sin^2 t + b^2\cos^2 t)^{3/2}}{ab(\sin^2 t + \cos^2 t)} \\ &= \frac{(a^2\sin^2 t + b^2\cos^2 t)^{3/2}}{ab}. \end{aligned} \tag{49}$$

Similarly, the unit TANGENT VECTOR is given by

$$\hat{\mathbf{T}} = \begin{bmatrix} -a\sin t \\ b\cos t \end{bmatrix} \frac{1}{\sqrt{a^2\sin^2 t + b^2\cos^2 t}}. \tag{50}$$

The ARC LENGTH of the ellipse can be computed using

$$\begin{aligned} s &= \int \sqrt{x'^2+y'^2}\,dt = \int \sqrt{a^2\cos^2 t + b^2\sin^2 t}\,dt \\ &= a\int \sqrt{(1-\sin^2 t) + \frac{b^2}{a^2}\sin^2 t}\,dt \\ &= a\int \sqrt{1-\left(1-\frac{b^2}{a^2}\right)\sin^2 t}\,dt \\ &= a\int \sqrt{1-e^2\sin^2 t}\,dt = aE(t,e), \end{aligned} \tag{51}$$

where E is an incomplete ELLIPTIC INTEGRAL OF THE SECOND KIND. Again, note that t is a parameter which does not have a direct interpretation in terms of an ANGLE. However, the relationship between the polar angle from the ellipse center θ and the parameter t follows from

$$\theta = \tan^{-1}\left(\frac{y}{x}\right) = \tan^{-1}\left(\frac{b}{a}\tan t\right). \tag{52}$$

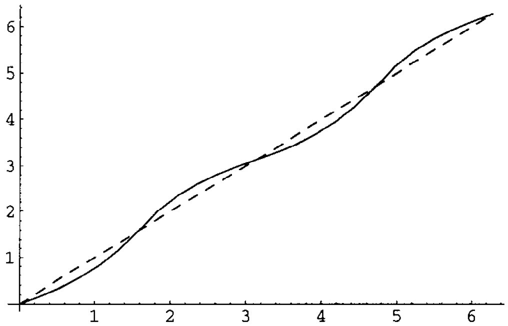

This function is illustrated above with θ shown as the solid curve and t as the dashed, with $b/a = 0.6$. Care must be taken to make sure that the correct branch of the INVERSE TANGENT function is used. As can be seen, θ weaves back and forth around t, with crossings occurring at multiples of $\pi/2$.

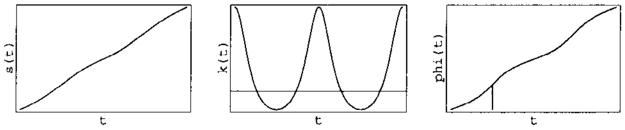

The CURVATURE and TANGENTIAL ANGLE of the ellipse are given by

$$\kappa = \frac{ab}{(b^2\cos^2 t + a^2\sin^2 t)^{3/2}} \tag{53}$$

$$\phi = -\tan^{-1}\left(\frac{b}{a}\cos t\right). \tag{54}$$

The entire Perimeter p of the ellipse is given by setting $t = 2\pi$ (corresponding to $\theta = 2\pi$), which is equivalent to four times the length of one of the ellipse's Quadrants,

$$p = aE(2\pi, e) = 4aE(\tfrac{1}{2}\pi, e) = 4aE(e), \tag{55}$$

where $E(e)$ is a complete Elliptic Integral of the Second Kind with Modulus k. The Perimeter can be computed numerically by the rapidly converging Gauss-Kummer Series

$$\begin{aligned} p &= \pi(a+b)\sum_{n=0}^{\infty}\binom{\frac{1}{2}}{n}^2 h^{2n} \\ &= \pi(a+b)(1 + \tfrac{1}{4}h^2 + \tfrac{1}{64}h^4 + \tfrac{1}{256}h^6 + \ldots), \end{aligned} \tag{56}$$

where

$$h \equiv \frac{a-b}{a+b} \tag{57}$$

and $\binom{n}{k}$ is a Binomial Coefficient. Approximations to the Perimeter include

$$\begin{aligned} p &\approx \pi\sqrt{2(a^2+b^2)} & (58)\\ &\approx \pi[3(a+b) - \sqrt{(a+3b)(3a+b)}] & (59)\\ &\approx \pi(a+b)\left(1 + \frac{3t}{10+\sqrt{4-3t}}\right), & (60) \end{aligned}$$

where the last two are due to Ramanujan (1913-14),

$$t \equiv \left(\frac{a-b}{a+b}\right)^2, \tag{61}$$

and (60) is accurate to within $\sim 3 \cdot 2^{-17}t^5$.

The maximum and minimum distances from the Focus are called the Apoapsis and Periapsis, and are given by

$$\begin{aligned} r_+ &= r_{\text{apoapsis}} = a(1+e) & (62)\\ r_- &= \text{periapsis} = a(1-e). & (63) \end{aligned}$$

The Area of an ellipse may be found by direct Integration

$$\begin{aligned} A &= \int_{-a}^{a}\int_{-b\sqrt{a^2-x^2}/a}^{b\sqrt{a^2-x^2}/a} dy\,dx = \int_{-a}^{a}\frac{2b}{a}\sqrt{a^2-x^2}\,dx \\ &= \frac{2b}{a}\left\{\frac{1}{2}\left[x\sqrt{a^2-x^2} + a^2\sin^{-1}\left(\frac{x}{|a|}\right)\right]\right\}_{x=-a}^{a} \\ &= ab[\sin^{-1}1 - \sin^{-1}(-1)] = ab\left[\frac{\pi}{2} - \left(\frac{\pi}{2}\right)\right] = \pi ab. \end{aligned} \tag{64}$$

The Area can also be computed more simply by making the change of coordinates $x' \equiv (b/a)x$ and $y' \equiv y$ from the elliptical region R to the new region R'. Then the equation becomes

$$\frac{1}{a^2}\left(\frac{a}{b}x'\right)^2 + \frac{y'^2}{b^2} = 1, \tag{65}$$

or $x'^2 + y'^2 = b^2$, so R' is a Circle of Radius b. Since

$$\frac{\partial x}{\partial x'} = \left(\frac{\partial x'}{\partial x}\right)^{-1} = \left(\frac{b}{a}\right)^{-1} = \frac{a}{b}, \tag{66}$$

the Jacobian is

$$\left|\tfrac{\partial(x,y)}{\partial(x',y')}\right| = \begin{vmatrix} \frac{\partial x}{\partial x'} & \frac{\partial y'}{\partial x'} \\ \frac{\partial x}{\partial y'} & \frac{\partial y}{\partial y'} \end{vmatrix} = \begin{vmatrix} \frac{a}{b} & 0 \\ 0 & 1 \end{vmatrix} = \frac{a}{b}. \tag{67}$$

The Area is therefore

$$\begin{aligned} \iint_R dx\,dy &= \iint_{R'}\left|\tfrac{\partial(x,y)}{\partial(x',y')}\right| dx'\,dy' \\ &= \frac{a}{b}\iint_{R'} dx'dy' = \frac{a}{b}(\pi b^2) = \pi ab, \end{aligned} \tag{68}$$

as before. The Area of an arbitrary ellipse given by the Quadratic Equation

$$ax^2 + bxy + cy^2 = 1 \tag{69}$$

is

$$A = \frac{2\pi}{\sqrt{4ac-b^2}}. \tag{70}$$

The Area of an Ellipse with semiaxes a and b with respect to a Pedal Point P is

$$A = \tfrac{1}{2}\pi(a^2 + b^2 + |OP|^2). \tag{71}$$

The ellipse Inscribed in a given Triangle and tangent at its Midpoints is called the Midpoint Ellipse. The Locus of the centers of the ellipses Inscribed in a Triangle is the interior of the Medial Triangle. Newton gave the solution to inscribing an ellipse in a convex Quadrilateral (Dörrie 1965, p. 217). The centers of the ellipses Inscribed in a Quadrilateral all lie on the straight line segment joining the Midpoints of the Diagonals (Chakerian 1979, pp. 136–139).

The Area of an ellipse with Barycentric Coordinates (α, β, γ) Inscribed in a Triangle of unit Area is

$$\Delta = \pi\sqrt{(1-2\alpha)(1-2\beta)(1-2\gamma)}. \tag{72}$$

(Chakerian 1979, pp. 142–145).

The Locus of the apex of a variable Cone containing an ellipse fixed in 3-space is a Hyperbola through the Foci of the ellipse. In addition, the Locus of the apex of a Cone containing that Hyperbola is the original ellipse. Furthermore, the Eccentricities of the ellipse and Hyperbola are reciprocals. The Locus of centers of a PAPPUS CHAIN of CIRCLES is an ellipse. Surprisingly, the locus of the end of a garage door mounted on rollers along a vertical track but extending beyond the track is a quadrant of an ellipse (the envelopes of positions is an ASTROID).

see also CIRCLE, CONIC SECTION, ECCENTRIC ANOMALY, ECCENTRICITY, ELLIPTIC CONE, ELLIPTIC CURVE, ELLIPTIC CYLINDER, HYPERBOLA, MIDPOINT ELLIPSE, PARABOLA, PARABOLOID, QUADRATIC CURVE, REFLECTION PROPERTY, SALMON'S THEOREM, STEINER'S ELLIPSE

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 126 and 198–199, 1987.

Casey, J. "The Ellipse." Ch. 6 in *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions, with Numerous Examples, 2nd ed., rev. enl.* Dublin: Hodges, Figgis, & Co., pp. 201–249, 1893.

Chakerian, G. D. "A Distorted View of Geometry." Ch. 7 in *Mathematical Plums* (Ed. R. Honsberger). Washington, DC: Math. Assoc. Amer., 1979.

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 75, 1996.

Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, 1965.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 72–78, 1972.

Lee, X. "Ellipse." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Ellipse_dir/ellipse.html`.

Lockwood, E. H. "The Ellipse." Ch. 2 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 13–24, 1967.

MacTutor History of Mathematics Archive. "Ellipse." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Ellipse.html`.

Ramanujan, S. "Modular Equations and Approximations to π." *Quart. J. Pure. Appl. Math.* **45**, 350–372, 1913–1914.

Ellipse Caustic Curve

For an ELLIPSE given by

$$x = r\cos t \tag{1}$$

$$y = \sin t \tag{2}$$

with light source at $(x, 0)$, the CAUSTIC is

$$x = \frac{N_x}{D_x} \tag{3}$$

$$y = \frac{N_y}{D_y}, \tag{4}$$

where

$$\begin{aligned} N_x &= 2rx(3-5r^2) + (-6r^2 + 6r^4 - 3x^2 + 9r^2x^2)\cos t \\ &\quad + 6rx(1-r^2)\cos(2t) \\ &\quad + (-2r^2 + 2r^4 - x^2 - r^2x^2)\cos(3t) \qquad (5) \\ D_x &= 2r(1+2r^2+4x^2) + 3x(1-5r^2)\cos t \\ &\quad + (6r+6r^3)\cos(2t) + x(1-r^2)\cos(3t) \qquad (6) \\ N_y &= 8r(-1+r^2-x^2)\sin^3 t \qquad (7) \\ D_y &= 2r(-1-r^2-4x^2) + 3(-x+5r^2)\cos t \\ &\quad + 6r(1-r^2)\cos(2t) + x(-1+r^2)\cos(3t). \qquad (8) \end{aligned}$$

At $(\infty, 0)$,

$$x = \frac{\cos t[-1+5r^2-\cos(2t)(1+r^2)]}{4r} \tag{9}$$

$$y = \sin^3 t. \tag{10}$$

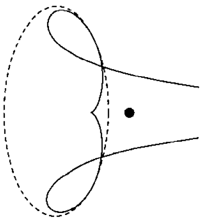

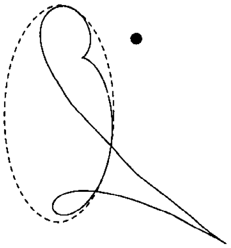

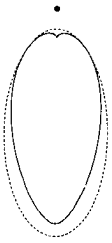

Ellipse Envelope

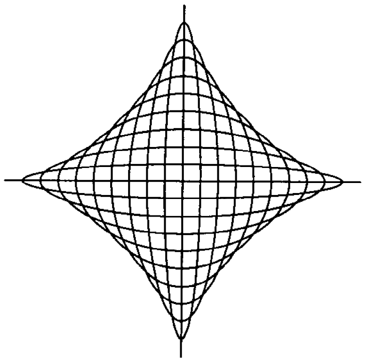

Consider the family of ELLIPSES

$$\frac{x^2}{c^2} + \frac{y^2}{(1-c)^2} - 1 = 0 \tag{1}$$

for $c \in [0, 1]$. The PARTIAL DERIVATIVE with respect to c is

$$-\frac{2x^2}{c^3} + \frac{2y^2}{(1-c)^3} = 0 \tag{2}$$

$$\frac{x^2}{c^3} - \frac{y^2}{(1-c)^3} = 0. \tag{3}$$

Combining (1) and (3) gives the set of equations

$$\begin{bmatrix} \frac{1}{c^2} & \frac{1}{(1-c)^2} \\ \frac{1}{c^3} & -\frac{1}{(1-c)^3} \end{bmatrix} \begin{bmatrix} x^2 \\ y^2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{4}$$

$$\begin{aligned} \begin{bmatrix} x^2 \\ y^2 \end{bmatrix} &= \frac{1}{\Delta} \begin{bmatrix} -\frac{1}{(1-c)^3} & -\frac{1}{(1-c)^2} \\ -\frac{1}{c^3} & \frac{1}{c^2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= \frac{1}{\Delta} \begin{bmatrix} -\frac{1}{(1-c)^3} \\ -\frac{1}{c^3} \end{bmatrix}, \qquad (5) \end{aligned}$$

where the DISCRIMINANT is

$$\Delta = -\frac{1}{c^2(1-c)^3} - \frac{1}{c^3(1-c)^2} = -\frac{1}{c^3(1-c)^3}, \tag{6}$$

so (5) becomes

$$\begin{bmatrix} x^2 \\ y^2 \end{bmatrix} = \begin{bmatrix} c^3 \\ (1-c)^3 \end{bmatrix}. \qquad (7)$$

Eliminating c then gives

$$x^{2/3} + y^{2/3} = 1, \qquad (8)$$

which is the equation of the ASTROID. If the curve is instead represented parametrically, then

$$x = c \cos t \qquad (9)$$

$$y = (1-c)\sin t. \qquad (10)$$

Solving

$$\begin{aligned} &\frac{\partial x}{\partial t}\frac{\partial y}{\partial c} - \frac{\partial x}{\partial c}\frac{\partial y}{\partial t} \\ &\quad = (-c\sin t)(-\sin t) - (\cos t)[(1-c)\cos t] \\ &\quad = c(\sin^2 t + \cos^2 t) - \cos^2 t = c - \cos^2 t = 0 \qquad (11) \end{aligned}$$

for c gives

$$c = \cos^2 t, \qquad (12)$$

so substituting this back into (9) and (10) gives

$$x = (\cos^2 t)\cos t = \cos^3 t \qquad (13)$$

$$y = (1-\cos^2 t)\sin t = \sin^3 t, \qquad (14)$$

the parametric equations of the ASTROID.

see also ASTROID, ELLIPSE, ENVELOPE

Ellipse Evolute

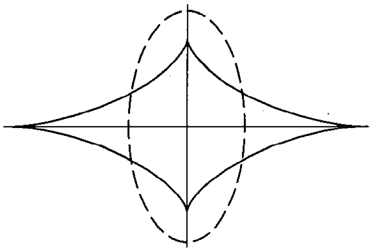

The EVOLUTE of an ELLIPSE is given by the parametric equations

$$x = \frac{a^2 - b^2}{a}\cos^3 t \qquad (1)$$

$$y = \frac{b^2 - a^2}{b}\sin^3 t, \qquad (2)$$

which can be combined and written

$$\begin{aligned} &(ax)^{2/3} + (by)^{2/3} \\ &\quad = [(a^2-b^2)\cos^3 t]^{2/3} + [(b^2-a^2)\sin^3 t]^{2/3} \\ &= (a^2-b^2)^{2/3}(\sin^2 t + \cos^2 t) = (a^2-b^2)^{2/3} = c^{4/3}, \qquad (3) \end{aligned}$$

which is a stretched ASTROID called the LAMÉ CURVE. From a point inside the EVOLUTE, four NORMALS can be drawn to the ellipse, but from a point outside, only two NORMALS can be drawn.

see also ASTROID, ELLIPSE, EVOLUTE, LAMÉ CURVE

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 77, 1993.

Ellipse Involute

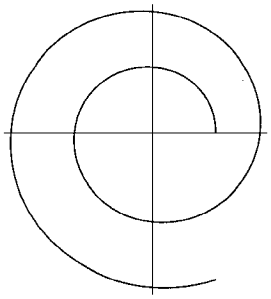

From ELLIPSE, the TANGENT VECTOR is

$$\mathbf{T} = \begin{bmatrix} -a\sin t \\ b\cos t \end{bmatrix}, \qquad (1)$$

and the ARC LENGTH is

$$s = a\int \sqrt{1 - e^2\sin^2 t}\,dt = aE(t,e), \qquad (2)$$

where $E(t,e)$ is an incomplete ELLIPTIC INTEGRAL OF THE SECOND KIND. Therefore,

$$\begin{aligned} \mathbf{r}_i = \mathbf{r} - s\hat{\mathbf{T}} &= \begin{bmatrix} a\cos t \\ b\sin t \end{bmatrix} - aeE(t,e)\begin{bmatrix} -a\sin t \\ b\cos t \end{bmatrix} \qquad (3) \\ &= \begin{bmatrix} a\{\cos t + aeE(t,e)\sin t\} \\ b\{\sin t - aeE(t,e)\cos t\}. \end{bmatrix} \qquad (4) \end{aligned}$$

Ellipse Pedal Curve

The pedal curve of an ellipse with a FOCUS as the PEDAL POINT is a CIRCLE.

Ellipsoid

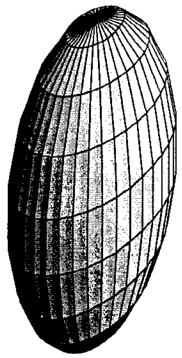

A QUADRATIC SURFACE which is given in CARTESIAN COORDINATES by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1, \qquad (1)$$

where the semi-axes are of lengths a, b, and c. In SPHERICAL COORDINATES, this becomes

$$\frac{r^2\cos^2\theta\sin^2\phi}{a^2} + \frac{r^2\sin^2\theta\sin^2\phi}{b^2} + \frac{r^2\cos^2\phi}{c^2} = 1. \qquad (2)$$

The parametric equations are

$$x = a\cos\theta\sin\phi \quad (3)$$

$$y = b\sin\theta\sin\phi \quad (4)$$

$$z = c\cos\phi. \quad (5)$$

The SURFACE AREA (Bowman 1961, pp. 31–32) is

$$S = 2\pi c^2 + \frac{2\pi b}{\sqrt{a^2-c^2}}[(a^2-c^2)E(\theta)+c^2\theta], \quad (6)$$

where $E(\theta)$ is a COMPLETE ELLIPTIC INTEGRAL OF THE SECOND KIND,

$$e_1{}^2 \equiv \frac{a^2-c^2}{a^2} \quad (7)$$

$$e_2{}^2 \equiv \frac{b^2-c^2}{b^2} \quad (8)$$

$$k \equiv \frac{e_2}{a_1}, \quad (9)$$

and θ is given by inverting the expression

$$e_1 = \mathrm{sn}(\theta, k), \quad (10)$$

where $\mathrm{sn}(\theta, k)$ is a JACOBI ELLIPTIC FUNCTION. The VOLUME of an ellipsoid is

$$V = \tfrac{4}{3}\pi abc. \quad (11)$$

If two axes are the same, the figure is called a SPHEROID (depending on whether $c < a$ or $c > a$, an OBLATE SPHEROID or PROLATE SPHEROID, respectively), and if all three are the same, it is a SPHERE.

A different parameterization of the ellipsoid is the so-called stereographic ellipsoid, given by the parametric equations

$$x(u,v) = \frac{a(1-u^2-v^2)}{1+u^2+v^2} \quad (12)$$

$$y(u,v) = \frac{2bu}{1+u^2+v^2} \quad (13)$$

$$z(u,v) = \frac{2cv}{1+u^2+v^2}. \quad (14)$$

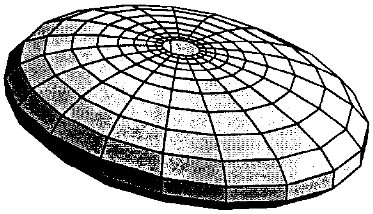

A third parameterization is the Mercator parameterization

$$x(u,v) = a\,\mathrm{sech}\,v\cos u \quad (15)$$

$$y(u,v) = b\,\mathrm{sech}\,v\sin u \quad (16)$$

$$z(u,v) = c\tanh v \quad (17)$$

(Gray 1993).

The SUPPORT FUNCTION of the ellipsoid is

$$h = \left(\frac{x^2}{a^4}+\frac{y^2}{b^4}+\frac{z^2}{c^4}\right)^{-1/2}, \quad (18)$$

and the GAUSSIAN CURVATURE is

$$K = \frac{h^4}{a^2b^2c^2} \quad (19)$$

(Gray 1993, p. 296).

see also CONVEX OPTIMIZATION THEORY, OBLATE SPHEROID, PROLATE SPHEROID, SPHERE, SPHEROID

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 131, 1987.

Bowman, F. *Introduction to Elliptic Functions, with Applications.* New York: Dover, 1961.

Fischer, G. (Ed.). Plate 65 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 60, 1986.

Gray, A. "The Ellipsoid" and "The Stereographic Ellipsoid." §11.2 and 11.3 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 215–217, and 296, 1993.

Ellipsoid Geodesic

An ELLIPSOID can be specified parametrically by

$$x = a\cos u\sin v \quad (1)$$

$$y = b\sin u\sin v \quad (2)$$

$$z = c\cos v. \quad (3)$$

The GEODESIC parameters are then

$$P = \sin^2 v(b^2\cos^2 u + a^2\sin^2 u) \quad (4)$$

$$Q = \tfrac{1}{4}(b^2-a^2)\sin(2u)\sin(2v) \quad (5)$$

$$R = \cos^2 v(a^2\cos^2 u + b^2\sin^2 u) + c^2\sin^2 v. \quad (6)$$

When the coordinates of a point are on the QUADRIC

$$\frac{x^2}{a}+\frac{y^2}{b}+\frac{z^2}{c} = 1 \quad (7)$$

and expressed in terms of the parameters p and q of the confocal quadrics passing through that point (in other words, having $a+p$, $b+p$, $c+p$, and $a+q$, $b+q$, $c+q$ for the squares of their semimajor axes), then the equation of a GEODESIC can be expressed in the form

$$\frac{q\,dq}{\sqrt{q(a+q)(b+q)(c+q)(\theta+q)}} \pm \frac{p\,dp}{\sqrt{p(a+p)(b+p)(c+p)(\theta+p)}} = 0, \quad (8)$$

with θ an arbitrary constant, and the ARC LENGTH element ds is given by

$$-2\frac{ds}{pq} = \frac{dq}{\sqrt{q(a+q)(b+q)(c+q)(\theta+q)}} \pm \frac{dp}{\sqrt{p(a+p)(b+p)(c+p)(\theta+p)}}, \qquad (9)$$

where upper and lower signs are taken together.

see also OBLATE SPHEROID GEODESIC, SPHERE GEODESIC

References

Eisenhart, L. P. *A Treatise on the Differential Geometry of Curves and Surfaces.* New York: Dover, pp. 236–241, 1960.

Forsyth, A. R. *Calculus of Variations.* New York: Dover, p. 447, 1960.

Ellipsoidal Calculus

Ellipsoidal calculus is a method for solving problems in control and estimation theory having unknown but bounded errors in terms of sets of approximating ellipsoidal-value functions. Ellipsoidal calculus has been especially useful in the study of LINEAR PROGRAMMING.

References

Kurzhanski, A. B. and Vályi, I. *Ellipsoidal Calculus for Estimation and Control.* Boston, MA: Birkhäuser, 1996.

Ellipsoidal Coordinates

see CONFOCAL ELLIPSOIDAL COORDINATES

Ellipsoidal Harmonic

see ELLIPSOIDAL HARMONIC OF THE FIRST KIND, ELLIPSOIDAL HARMONIC OF THE SECOND KIND

Ellipsoidal Harmonic of the First Kind

The first solution to LAMÉ'S DIFFERENTIAL EQUATION, denoted $E_n^m(x)$ for $m = 1, \ldots, 2n+1$. They are also called LAMÉ FUNCTIONS. The product of two ellipsoidal harmonics of the first kind is a SPHERICAL HARMONIC. Whittaker and Watson (1990, pp. 536–537) write

$$\Theta_p = \frac{x^2}{a^2+\theta_p} + \frac{y^2}{b^2+\theta_p} + \frac{z^2}{c^2+\theta_p} - 1 \qquad (1)$$

$$\Pi(\Theta) \equiv \Theta_1\Theta_2\cdots\Theta_m, \qquad (2)$$

and give various types of ellipsoidal harmonics and their highest degree terms as

1. $\Pi(\Theta) : 2m$
2. $x\Pi(\Theta), y\Pi(\Theta), z\Pi(\Theta) : 2m+1$
3. $yz\Pi(\Theta), zx\Pi(\Theta), xy\Pi(\Theta) : 2m+2$
4. $xyz\Pi(\Theta) : 2m+3$.

A Lamé function of degree n may be expressed as

$$(\theta+a^2)^{\kappa_1}(\theta+b^2)^{\kappa_2}(\theta+c^2)^{\kappa_3}\prod_{p=1}^{m}(\theta-\theta_p), \qquad (3)$$

where $\kappa_i = 0$ or $1/2$, θ_i are REAL and unequal to each other and to $-a^2$, $-b^2$, and $-c^2$, and

$$\tfrac{1}{2}n = m + \kappa_1 + \kappa_2 + \kappa_3. \qquad (4)$$

Byerly (1959) uses the RECURRENCE RELATIONS to explicitly compute some ellipsoidal harmonics, which he denotes by $K(x)$, $L(x)$, $M(x)$, and $N(x)$,

$$\begin{aligned}
K_0(x) &= 1\\
L_0(x) &= 0\\
M_0(x) &= 0\\
N_0(x) &= 0\\
K_1(x) &= x\\
L_1(x) &= \sqrt{x^2-b^2}\\
M_1(x) &= \sqrt{x^2-c^2}\\
N_1(x) &= 0\\
K_2^{p_1}(x) &= x^2 - \tfrac{1}{3}[b^2+c^2-\sqrt{(b^2+c^2)^2-3b^2c^2}\,]\\
K_2^{p_2}(x) &= x^2 - \tfrac{1}{3}[b^2+c^2+\sqrt{(b^2+c^2)^2-3b^2c^2}\,]\\
L_2(x) &= x\sqrt{x^2-b^2}\\
M_2(x) &= x\sqrt{x^2-c^2}\\
N_2(x) &= \sqrt{(x^2-b^2)(x^2-c^2)}\\
K_3^{p_1}(x) &= x^3 - \tfrac{1}{5}x[2(b^2+c^2)\\
&\quad - \sqrt{4(b^2+c^2)^2-15b^2c^2}\,]\\
K_3^{p_2}(x) &= x^3 - \tfrac{1}{5}x[2(b^2+c^2)\\
&\quad + \sqrt{4(b^2+c^2)^2-15b^2c^2}\,]\\
L_3^{q_1}(x) &= \sqrt{x^2-b^2}[x^2 - \tfrac{1}{5}(b^2+2c^2\\
&\quad - \sqrt{(b^2+2c^2)^2-5b^2c^2}\,)]\\
L_3^{q_2}(x) &= \sqrt{x^2-b^2}[x^2 - \tfrac{1}{5}(b^2+2c^2\\
&\quad + \sqrt{(b^2+2c^2)^2-5b^2c^2}\,)]\\
M_3^{q_1}(x) &= \sqrt{x^2-c^2}[x^2 - \tfrac{1}{5}(2b^2+c^2\\
&\quad - \sqrt{(2b^2+c^2)^2-5b^2c^2}\,)]\\
M_3^{q_2}(x) &= \sqrt{x^2-c^2}[x^2 - \tfrac{1}{5}(2b^2+c^2\\
&\quad + \sqrt{(2b^2+c^2)^2-5b^2c^2}\,)]\\
M_3^{q_3}(x) &= x\sqrt{(x^2-b^2)(x^2-c^2)}
\end{aligned}$$

see also ELLIPSOIDAL HARMONIC OF THE SECOND KIND, STIELTJES' THEOREM

References

Byerly, W. E. *An Elementary Treatise on Fourier's Series, and Spherical, Cylindrical, and Ellipsoidal Harmonics,*

with Applications to Problems in Mathematical Physics. New York: Dover, pp. 254–258, 1959.
Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Ellipsoidal Harmonic of the Second Kind

Given by

$$F_m^p(x) = (2m+1)E_m^p(x) \times \int_x^\infty \frac{dx}{(x^2-b^2)(x^2-c^2)[E_m^p(x)]^2}.$$

Elliptic Alpha Function

Elliptic alpha functions relate the complete Elliptic Integrals of the First $K(k_r)$ and Second Kinds $E(k_r)$ at Elliptic Integral Singular Values k_r according to

$$\alpha(r) = \frac{E'(k_r)}{K(k_r)} - \frac{\pi}{4[K(k_r)]^2} \tag{1}$$

$$= \frac{\pi}{4[K(k_r)]^2} + \sqrt{r} - \frac{E(k_r)\sqrt{r}}{K(k_r)} \tag{2}$$

$$= \frac{\pi^{-1} - 4\sqrt{r}\,q\frac{d\vartheta_4(q)}{dq}\frac{1}{\vartheta_4(q)}}{{\vartheta_3}^4(q)}, \tag{3}$$

where $\vartheta_3(q)$ is a Theta Function and

$$k_r = \lambda^*(r) \tag{4}$$

$$q = e^{-\pi\sqrt{r}}, \tag{5}$$

and $\lambda^*(r)$ is the Elliptic Lambda Function. The elliptic alpha function is related to the Elliptic Delta Function by

$$\alpha(r) = \tfrac{1}{2}[\sqrt{r} - \delta(r)]. \tag{6}$$

It satisfies

$$\alpha(4r) = (1+k_r)^2\alpha(r) - 2\sqrt{r}\,k_r, \tag{7}$$

and has the limit

$$\lim_{r\to\infty}\left[\alpha(r) - \frac{1}{\pi}\right] \approx 8\left(\sqrt{r} - \frac{1}{\pi}\right)e^{-\pi\sqrt{r}} \tag{8}$$

(Borwein *et al.* 1989). A few specific values (Borwein and Borwein 1987, p. 172) are

$$\alpha(1) = \tfrac{1}{2}$$
$$\alpha(2) = \sqrt{2} - 1$$
$$\alpha(3) = \tfrac{1}{2}(\sqrt{3} - 1)$$
$$\alpha(4) = 2(\sqrt{2} - 1)^2$$
$$\alpha(5) = \tfrac{1}{2}\left(\sqrt{5} - \sqrt{2\sqrt{5} - 2}\right)$$
$$\alpha(6) = 5\sqrt{6} + 6\sqrt{3} - 8\sqrt{2} - 11$$
$$\alpha(7) = \tfrac{1}{2}(\sqrt{7} - 2)$$
$$\alpha(8) = 2(10 + 7\sqrt{2})\left(1 - \sqrt{\sqrt{8} - 2}\right)^2$$
$$\alpha(9) = \tfrac{1}{2}[3 - 3^{3/4}\sqrt{2}(\sqrt{3} - 1)]$$
$$\alpha(10) = -103 + 72\sqrt{2} - 46\sqrt{5} + 33\sqrt{10}$$
$$\alpha(12) = 264 + 154\sqrt{3} - 188\sqrt{2} - 108\sqrt{6}$$
$$\alpha(13) = \tfrac{1}{2}\left(\sqrt{13} - \sqrt{74\sqrt{13} - 258}\right)$$
$$\alpha(15) = \tfrac{1}{2}(\sqrt{15} - \sqrt{5} - 1)$$
$$\alpha(16) = \frac{4(\sqrt{8} - 1)}{(2^{1/4} + 1)^4}$$
$$\alpha(18) = -3057 + 2163\sqrt{2} + 1764\sqrt{3} - 1248\sqrt{6}$$
$$\alpha(22) = -12479 - 8824\sqrt{2} + 3762\sqrt{11} + 2661\sqrt{22}$$
$$\alpha(25) = \tfrac{5}{2}[1 - 25^{1/4}(7 - 3\sqrt{5})]$$
$$\alpha(27) = 3[\tfrac{1}{2}(\sqrt{3} + 1) - 2^{1/3}]$$
$$\begin{aligned}\alpha(30) = \tfrac{1}{2}\{&\sqrt{30} - (2 + \sqrt{5})^2(3 + \sqrt{10})^2 \\ &\times \left(-6 - 5\sqrt{2} - 3\sqrt{5} - 2\sqrt{10} + \sqrt{6}\sqrt{57 + 40\sqrt{2}}\right) \\ &\times [56 + 38\sqrt{2} + \sqrt{30}(2 + \sqrt{5})(3 + \sqrt{10})]\}\end{aligned}$$
$$\alpha(37) = \tfrac{1}{2}\left[\sqrt{37} - (171 - 25\sqrt{37})\sqrt{\sqrt{37} - 6}\right]$$
$$\begin{aligned}\alpha(49) = {}&\tfrac{7}{2} \\ &- \sqrt{7[\sqrt{2}\,7^{3/4}(33011 + 12477\sqrt{7}) - 21(9567 + 3616\sqrt{7})]}\end{aligned}$$
$$\begin{aligned}\alpha(46) = \tfrac{1}{2}[&\sqrt{46} + \left(18 + 13\sqrt{2} + \sqrt{661 + 468\sqrt{2}}\right)^2 \\ &\times \left(18 + 13\sqrt{2} - 3\sqrt{2}\sqrt{147 + 104\sqrt{2}} + \sqrt{661 + 468\sqrt{2}}\right) \\ &\times \left(200 + 14\sqrt{2} + 26\sqrt{23} + 18\sqrt{46} + \sqrt{46}\sqrt{661 + 468\sqrt{2}}\right)]\end{aligned}$$
$$\begin{aligned}\alpha(58) &= [\tfrac{1}{2}(\sqrt{29} + 5)]^6(99\sqrt{29} - 444)(99\sqrt{2} - 70 - 13\sqrt{29}) \\ &= 3(-40768961 + 28828008\sqrt{2} - 7570606\sqrt{29} \\ &\quad + 5353227\sqrt{58})\end{aligned}$$
$$\alpha(64) = \frac{8[2(\sqrt{8} - 1) - (2^{1/4} - 1)^4]}{\left(\sqrt{\sqrt{2} + 1} + 2^{5/8}\right)^4}.$$

J. Borwein has written an Algorithm which uses lattice basis reduction to provide algebraic values for $\alpha(n)$.

see also Elliptic Integral of the First Kind, Elliptic Integral of the Second Kind, Elliptic Integral Singular Value, Elliptic Lambda Function

References

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, 1987.
Borwein, J. M.; Borwein, P. B.; and Bailey, D. H. "Ramanujan, Modular Equations, and Approximations to Pi, or How to Compute One Billion Digits of Pi." *Amer. Math. Monthly* **96**, 201–219, 1989.

Weisstein, E. W. "Elliptic Singular Values." `http://www.astro.virginia.edu/~eww6n/math/notebooks/EllipticSingular.m`.

Elliptic Cone

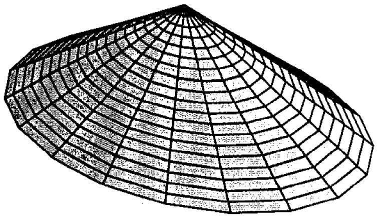

A CONE with ELLIPTICAL CROSS-SECTION. The parametric equations for an elliptic cone of height h, SEMIMAJOR AXIS a, and SEMIMINOR AXIS b are

$$\begin{aligned} x &= (h-z)a\cos\theta \\ y &= (h-z)b\sin\theta \\ z &= z, \end{aligned}$$

where $\theta \in [0, 2\pi)$ and $z \in [0, h]$.

see also CONE, ELLIPTIC CYLINDER, ELLIPTIC PARABOLOID, HYPERBOLIC PARABOLOID

References

Fischer, G. (Ed.). Plate 68 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 63, 1986.

Elliptic Cone Point

see ISOLATED SINGULARITY

Elliptic Curve

Informally, an elliptic curve is a type of CUBIC CURVE whose solutions are confined to a region of space which is topologically equivalent to a TORUS. Formally, an elliptic curve over a FIELD K is a nonsingular CUBIC CURVE in two variables, $f(X,Y) = 0$, with a K-rational point (which may be a point at infinity). The FIELD K is usually taken to be the COMPLEX NUMBERS $\mathbb{C}$, REALS $\mathbb{R}$, RATIONALS $\mathbb{Q}$, algebraic extensions of $\mathbb{Q}$, p-ADIC NUMBERS $\mathbb{Q}_p$, or a FINITE FIELD.

By an appropriate change of variables, a general elliptic curve over a FIELD of CHARACTERISTIC $\neq 2, 3$

$$\begin{aligned} Ax^3 + Bx^2y + Cxy^2 + Dy^3 + Ex^2 \\ +Fxy + Gy^2 + Hx + Iy + J = 0, \end{aligned} \tag{1}$$

where A, B, ..., are elements of K, can be written in the form

$$y^2 = x^3 + ax + b, \tag{2}$$

where the right side of (2) has no repeated factors. If K has CHARACTERISTIC three, then the best that can be done is to transform the curve into

$$y^2 = x^3 + ax^2 + bx + c \tag{3}$$

(the x^2 term cannot be eliminated). If K has CHARACTERISTIC two, then the situation is even worse. A general form into which an elliptic curve over any K can be transformed is called the WEIERSTRAß FORM, and is given by

$$y^2 + ay = x^3 + bx^2 + cxy + dx + e, \tag{4}$$

where a, b, c, d, and e are elements of K. Luckily, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ all have CHARACTERISTIC zero.

Whereas CONIC SECTIONS can be parameterized by the rational functions, elliptic curves cannot. The simplest parameterization functions are ELLIPTIC FUNCTIONS. ABELIAN VARIETIES can be viewed as generalizations of elliptic curves.

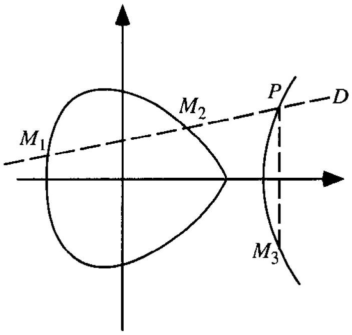

If the underlying FIELD of an elliptic curve is algebraically closed, then a straight line cuts an elliptic curve at three points (counting multiple roots at points of tangency). If two are known, it is possible to compute the third. If two of the intersection points are K-RATIONAL, then so is the third. Let (x_1, y_1) and (x_2, y_2) be two points on an elliptic curve E with DISCRIMINANT

$$\Delta_E = -16(4a^3 + 27b^2) \tag{5}$$

satisfying

$$\Delta_E \neq 0. \tag{6}$$

A related quantity known as the j-INVARIANT of E is defined as

$$j(E) \equiv \frac{2^8 3^3 a^3}{4a^3 + 27b^2}. \tag{7}$$

Now define

$$\lambda = \begin{cases} \frac{y_1 - y_2}{x_1 - x_2} & \text{for } x_1 \neq x_2 \\ \frac{3{x_1}^2 + a}{2y_1} & \text{for } x_1 = x_2. \end{cases} \tag{8}$$

Then the coordinates of the third point are

$$x_3 = \lambda^2 - x_1 - x_2 \tag{9}$$

$$y_3 = \lambda(x_3 - x_1) + y_1. \tag{10}$$

For elliptic curves over $\mathbb{Q}$, Mordell proved that there are a finite number of integral solutions. The MORDELL-WEIL THEOREM says that the GROUP of RATIONAL

Points of an elliptic curve over $\mathbb{Q}$ is finitely generated. Let the Roots of y^2 be r_1, r_2, and r_3. The discriminant is then

$$\Delta = k(r_1 - r_2)^2(r_1 - r_3)^2(r_2 - r_3)^2. \qquad (11)$$

The amazing Taniyama-Shimura Conjecture states that all rational elliptic curves are also modular. This fact is far from obvious, and despite the fact that the conjecture was proposed in 1955, it was not proved until 1995. Even so, Wiles' proof surprised most mathematicians, who had believed the conjecture unassailable. As a side benefit, Wiles' proof of the Taniyama-Shimura Conjecture also laid to rest the famous and thorny problem which had baffled mathematicians for hundreds of years, Fermat's Last Theorem.

Curves with small Conductors are listed in Swinnerton-Dyer (1975) and Cremona (1997). Methods for computing integral points (points with integral coordinates) are given in Gebel *et al.* and Stroeker and Tzanakis (1994).

see also Elliptic Curve Group Law, Fermat's Last Theorem, Frey Curve, j-Invariant, Minimal Discriminant, Mordell-Weil Theorem, Ochoa Curve, Ribet's Theorem, Siegel's Theorem, Swinnerton-Dyer Conjecture, Taniyama-Shimura Conjecture, Weierstraß Form

References

Atkin, A. O. L. and Morain, F. "Elliptic Curves and Primality Proving." *Math. Comput.* **61**, 29–68, 1993.

Cassels, J. W. S. *Lectures on Elliptic Curves.* New York: Cambridge University Press, 1991.

Cremona, J. E. *Algorithms for Modular Elliptic Curves, 2nd ed.* Cambridge, England: Cambridge University Press, 1997.

Cremona, J. E. "Elliptic Curve Data." `ftp://euclid.ex.ac.uk/pub/cremona/data/`.

Du Val, P. *Elliptic Functions and Elliptic Curves.* Cambridge: Cambridge University Press, 1973.

Gebel, J.; Pethő, A.; and Zimmer, H. G. "Computing Integral Points on Elliptic Curves." *Acta Arith.* **68**, 171–192, 1994.

Ireland, K. and Rosen, M. "Elliptic Curves." Ch. 18 in *A Classical Introduction to Modern Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 297–318, 1990.

Katz, N. M. and Mazur, B. *Arithmetic Moduli of Elliptic Curves.* Princeton, NJ: Princeton University Press, 1985.

Knapp, A. W. *Elliptic Curves.* Princeton, NJ: Princeton University Press, 1992.

Koblitz, N. *Introduction to Elliptic Curves and Modular Forms.* New York: Springer-Verlag, 1993.

Lang, S. *Elliptic Curves: Diophantine Analysis.* Berlin: Springer-Verlag, 1978.

Silverman, J. H. *The Arithmetic of Elliptic Curves.* New York: Springer-Verlag, 1986.

Silverman, J. H. *The Arithmetic of Elliptic Curves II.* New York: Springer-Verlag, 1994.

Silverman, J. H. and Tate, J. T. *Rational Points on Elliptic Curves.* New York: Springer-Verlag, 1992.

Stroeker, R. J. and Tzanakis, N. "Solving Elliptic Diophantine Equations by Estimating Linear Forms in Elliptic Logarithms." *Acta Arith.* **67**, 177–196, 1994.

Swinnerton-Dyer, H. P. F. "Correction to: 'On 1-adic Representations and Congruences for Coefficients of Modular Forms.'" In *Modular Functions of One Variable, Vol. 4, Proc. Internat. Summer School for Theoret. Phys., Univ. Antwerp, Antwerp, RUCA, July-Aug. 1972.* Berlin: Springer-Verlag, 1975.

Elliptic Curve Factorization Method

A factorization method, abbreviated ECM, which computes a large multiple of a point on a random Elliptic Curve modulo the number to be factored N. It tends to be faster than the Pollard ρ Factorization and Pollard $p - 1$ Factorization Method.

see also Atkin-Goldwasser-Kilian-Morain Certificate, Elliptic Curve Primality Proving, Elliptic Pseudoprime

References

Atkin, A. O. L. and Morain, F. "Finding Suitable Curves for the Elliptic Curve Method of Factorization." *Math. Comput.* **60**, 399–405, 1993.

Brent, R. P. "Some Integer Factorization Algorithms Using Elliptic Curves." *Austral. Comp. Sci. Comm.* **8**, 149–163, 1986.

Brent, R. P. "Parallel Algorithms for Integer Factorisation." In *Number Theory and Cryptography* (Ed. J. H. Loxton). New York: Cambridge University Press, 26–37, 1990. `ftp://nimbus.anu.edu.au/pub/Brent/115.dvi.Z`.

Brillhart, J.; Lehmer, D. H.; Selfridge, J.; Wagstaff, S. S. Jr.; and Tuckerman, B. *Factorizations of* $b^n \pm 1$, $b = 2, 3, 5, 6, 7, 10, 11, 12$ *Up to High Powers, rev. ed.* Providence, RI: Amer. Math. Soc., p. lxxxiii, 1988.

Eldershaw, C. and Brent, R. P. "Factorization of Large Integers on Some Vector and Parallel Computers." `ftp://nimbus.anu.edu.au/pub/Brent/156tr.dvi.Z`.

Lenstra, A. K. and Lenstra, H. W. Jr. "Algorithms in Number Theory." In *Handbook of Theoretical Computer Science, Volume A: Algorithms and Complexity* (Ed. J. van Leeuwen). Elsevier, pp. 673–715, 1990.

Lenstra, H. W. Jr. "Factoring Integers with Elliptic Curves." *Ann. Math.* **126**, 649–673, 1987.

Montgomery, P. L. "Speeding the Pollard and Elliptic Curve Methods of Factorization." *Math. Comput.* **48**, 243–264, 1987.

Elliptic Curve Group Law

The Group of an Elliptic Curve which has been transformed to the form

$$y^2 = x^3 + ax + b$$

is the set of K-Rational Points, including the single Point at Infinity. The group law (addition) is defined as follows: Take 2 K-Rational Points P and Q. Now 'draw' a straight line through them and compute the third point of intersection R (also a K-Rational Point). Then

$$P + Q + R = 0$$

gives the identity point at infinity. Now find the inverse of R, which can be done by setting $R = (a, b)$ giving $-R = (a, -b)$.

This remarkable result is only a special case of a more general procedure. Essentially, the reason is that this type of ELLIPTIC CURVE has a single point at infinity which is an inflection point (the line at infinity meets the curve at a single point at infinity, so it must be an intersection of multiplicity three).

Elliptic Curve Primality Proving

A class of algorithm, abbreviated ECPP, which provides certificates of primality using sophisticated results from the theory of ELLIPTIC CURVES. A detailed description and list of references are given by Atkin and Morain (1990, 1993).

Adleman and Huang (1987) designed an independent algorithm using elliptic curves of genus two.

see also ATKIN-GOLDWASSER-KILIAN-MORAIN CERTIFICATE, ELLIPTIC CURVE FACTORIZATION METHOD, ELLIPTIC PSEUDOPRIME

References

Adleman, L. M. and Huang, M. A. "Recognizing Primes in Random Polynomial Time." In *Proc. 19th STOC, New York City, May 25–27, 1986.* New York: ACM Press, pp. 462–469, 1987.

Atkin, A. O. L. Lecture notes of a conference, Boulder, CO, Aug. 1986.

Atkin, A. O. L. and Morain, F. "Elliptic Curves and Primality Proving." Res. Rep. 1256, INRIA, June 1990.

Atkin, A. O. L. and Morain, F. "Elliptic Curves and Primality Proving." *Math. Comput.* **61**, 29–68, 1993.

Bosma, W. "Primality Testing Using Elliptic Curves." Techn. Rep. 85–12, Math. Inst., Univ. Amsterdam, 1985.

Chudnovsky, D. V. and Chudnovsky, G. V. "Sequences of Numbers Generated by Addition in Formal Groups and New Primality and Factorization Tests." Res. Rep. RC 11262, IBM, Yorktown Heights, NY, 1985.

Cohen, H. *Cryptographie, factorisation et primalité: l'utilisation des courbes elliptiques.* Paris: C. R. J. Soc. Math. France, Jan. 1987.

Kaltofen, E.; Valente, R.; and Yui, N. "An Improved Las Vegas Primality Test." Res. Rep. 89–12, Rensselaer Polytechnic Inst., Troy, NY, May 1989.

Elliptic Cylinder

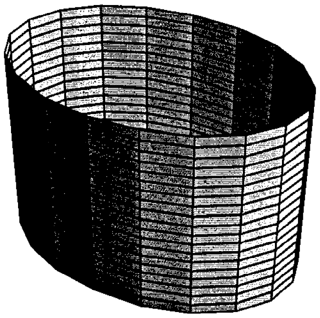

A CYLINDER with ELLIPTICAL CROSS-SECTION. The parametric equations for an elliptic cylinder of height h, SEMIMAJOR AXIS a, and SEMIMINOR AXIS b are

$$x = a\cos\theta$$
$$y = b\sin\theta$$
$$z = z,$$

where $\theta \in [0, 2\pi)$ and $z \in [0, h]$.

see also CONE, CYLINDER, ELLIPTIC CONE, ELLIPTIC PARABOLOID

Elliptic Cylindrical Coordinates

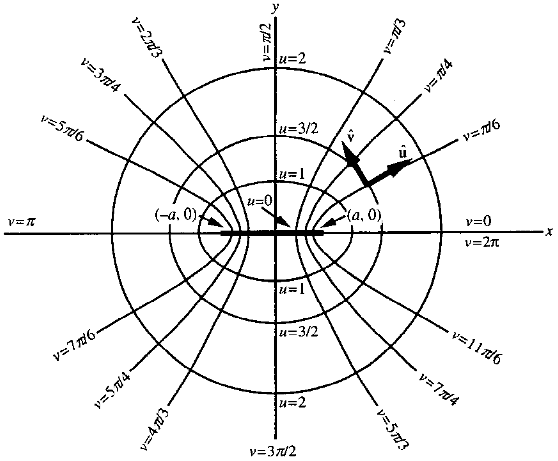

The v coordinates are the asymptotic angle of confocal PARABOLA segments symmetrical about the x axis. The u coordinates are confocal ELLIPSES centered on the origin.

$$x = a\cosh u\cos v \tag{1}$$
$$y = a\sinh u\sin v \tag{2}$$
$$z = z, \tag{3}$$

where $u \in [0, \infty)$, $v \in [0, 2\pi)$, and $z \in (-\infty, \infty)$. They are related to CARTESIAN COORDINATES by

$$\frac{x^2}{a^2\cosh^2 u} + \frac{y^2}{a^2\sinh^2 u} = 1 \tag{4}$$

$$\frac{x^2}{a^2\cos^2 v} - \frac{y^2}{a^2\sin^2 v} = 1. \tag{5}$$

The SCALE FACTORS are

$$h_1 = a\sqrt{\cosh^2 u\sin^2 v + \sinh^2 u\cos^2 v} \tag{6}$$
$$= a\sqrt{\frac{\cosh(2u) - \cos(2v)}{2}} \tag{7}$$
$$= a\sqrt{\sinh^2 u + \sin^2 v} \tag{8}$$
$$h_2 = a\sqrt{\sinh^2 u\sin^2 v + \sinh^2 u\cos^2 v} \tag{9}$$
$$= a\sqrt{\frac{\cosh(2u) - \cos(2v)}{2}} \tag{10}$$
$$= a\sqrt{\sinh^2 u + \sin^2 v} \tag{11}$$
$$h_3 = 1. \tag{12}$$

The LAPLACIAN is

$$\nabla^2 = \frac{1}{a^2(\sinh^2 u + \sin^2 v)}\left(\frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2}\right) + \frac{\partial^2}{\partial z^2}. \tag{13}$$

Let

$$q_1 = \cosh u \tag{14}$$

$$q_2 = \cos v \tag{15}$$

$$q_3 = z. \tag{16}$$

Then the new SCALE FACTORS are

$$h_{q_1} = a\sqrt{\frac{{q_1}^2 - {q_2}^2}{{q_1}^2 - 1}} \tag{17}$$

$$h_{q_2} = a\sqrt{\frac{{q_1}^2 - {q_2}^2}{1 - {q_1}^2}} \tag{18}$$

$$h_{q_3} = 1. \tag{19}$$

The HELMHOLTZ DIFFERENTIAL EQUATION is SEPARABLE.

see also CYLINDRICAL COORDINATES, HELMHOLTZ DIFFERENTIAL EQUATION—ELLIPTIC CYLINDRICAL COORDINATES

References

Arfken, G. "Elliptic Cylindrical Coordinates (u, v, z)." §2.7 in *Mathematical Methods for Physicists, 2nd ed.* Orlando, FL: Academic Press, pp. 95–97, 1970.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 657, 1953.

Elliptic Delta Function

$$\delta(r) = \sqrt{r} - 2\alpha(r),$$

where α is the ELLIPTIC ALPHA FUNCTION.

see also ELLIPTIC ALPHA FUNCTION, ELLIPTIC INTEGRAL SINGULAR VALUE

References

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, 1987.

Weisstein, E. W. "Elliptic Singular Values." `http://www.astro.virginia.edu/~eww6n/math/notebooks/EllipticSingular.m`.

Elliptic Exponential Function

The inverse of the ELLIPTIC LOGARITHM

$$\operatorname{eln}(x) \equiv \int_x^\infty \frac{dt}{\sqrt{t^3 + at^2 + bt}}.$$

It is doubly periodic in the COMPLEX PLANE.

Elliptic Fixed Point (Differential Equations)

A FIXED POINT for which the STABILITY MATRIX is purely IMAGINARY, $\lambda_\pm = \pm i\omega$ (for $\omega > 0$).

see also DIFFERENTIAL EQUATION, FIXED POINT, HYPERBOLIC FIXED POINT (DIFFERENTIAL EQUATIONS), PARABOLIC FIXED POINT, STABLE IMPROPER NODE, STABLE NODE, STABLE SPIRAL POINT, STABLE STAR, UNSTABLE IMPROPER NODE, UNSTABLE NODE, UNSTABLE SPIRAL POINT, UNSTABLE STAR

References

Tabor, M. "Classification of Fixed Points." §1.4.b in *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, pp. 22–25, 1989.

Elliptic Fixed Point (Map)

A FIXED POINT of a LINEAR TRANSFORMATION (MAP) for which the rescaled variables satisfy

$$(\delta - \alpha)^2 + 4\beta\gamma < 0.$$

see also HYPERBOLIC FIXED POINT (MAP), LINEAR TRANSFORMATION, PARABOLIC FIXED POINT

Elliptic Function

A doubly periodic function with periods $2\omega_1$ and $2\omega_2$ such that

$$f(z + 2\omega_1) = f(z + 2\omega_2) = f(z), \tag{1}$$

which is ANALYTIC and has no singularities except for POLES in the finite part of the COMPLEX PLANE. The ratio ω_1/ω_2 must not be purely real. If this ratio is real, the function reduces to a singly periodic function if it is rational and a constant if the ratio is irrational (Jacobi, 1835). ω_1 and ω_2 are labeled such that $\Im(\omega_2/\omega_1) > 0$. A "cell" of an elliptic function is defined as a parallelogram region in the COMPLEX PLANE in which the function is not multi-valued. Properties obeyed by elliptic functions include

1. The number of POLES in a cell is finite.
2. The number of ROOTS in a cell is finite.
3. The sum of RESIDUES in any cell is 0.
4. LIOUVILLE'S ELLIPTIC FUNCTION THEOREM: An elliptic function with no POLES in a cell is a constant.
5. The number of zeros of $f(z) - c$ (the "order") equals the number of POLES of $f(z)$.
6. The simplest elliptic function has order two, since a function of order one would have a simple irreducible POLE, which would need to have a NONZERO residue. By property (3), this is impossible.
7. Elliptic functions with a single POLE of order 2 with RESIDUE 0 are called WEIERSTRAß ELLIPTIC FUNCTIONS. Elliptic functions with two simple POLES having residues a_0 and $-a_0$ are called JACOBI ELLIPTIC FUNCTIONS.
8. Any elliptic function is expressible in terms of either WEIERSTRAß ELLIPTIC FUNCTION or JACOBI ELLIPTIC FUNCTIONS.
9. The sum of the AFFIXES of ROOTS equals the sum of the AFFIXES of the POLES.
10. An algebraic relationship exists between any two elliptic functions with the same periods.

The elliptic functions are inversions of the ELLIPTIC INTEGRALS. The two standard forms of these functions are known as JACOBI ELLIPTIC FUNCTIONS and WEIERSTRAß ELLIPTIC FUNCTIONS. JACOBI ELLIPTIC FUNCTIONS arise as solutions to differential equations of the form

$$\frac{d^2x}{dt^2} = A + Bx + Cx^2 + Dx^3, \tag{2}$$

and WEIERSTRAß ELLIPTIC FUNCTIONS arise as solutions to differential equations of the form

$$\frac{d^2x}{dt^2} = A + Bx + Cx^2. \qquad (3)$$

see also ELLIPTIC CURVE, ELLIPTIC INTEGRAL, JACOBI ELLIPTIC FUNCTIONS, LIOUVILLE'S ELLIPTIC FUNCTION THEOREM, MODULAR FORM, MODULAR FUNCTION, NEVILLE THETA FUNCTION, THETA FUNCTION, WEIERSTRAß ELLIPTIC FUNCTIONS

References

Akhiezer, N. I. *Elements of the Theory of Elliptic Functions.* Providence, RI: Amer. Math. Soc., 1990.

Bellman, R. E. *A Brief Introduction to Theta Functions.* New York: Holt, Rinehart and Winston, 1961.

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, 1987.

Bowman, F. *Introduction to Elliptic Functions, with Applications.* New York: Dover, 1961.

Byrd, P. F. and Friedman, M. D. *Handbook of Elliptic Integrals for Engineers and Scientists, 2nd ed., rev.* Berlin: Springer-Verlag, 1971.

Cayley, A. *An Elementary Treatise on Elliptic Functions, 2nd ed.* London: G. Bell, 1895.

Chandrasekharan, K. *Elliptic Functions.* Berlin: Springer-Verlag, 1985.

Du Val, P. *Elliptic Functions and Elliptic Curves.* Cambridge, England: Cambridge University Press, 1973.

Dutta, M. and Debnath, L. *Elements of the Theory of Elliptic and Associated Functions with Applications.* Calcutta, India: World Press, 1965.

Eagle, A. *The Elliptic Functions as They Should Be: An Account, with Applications, of the Functions in a New Canonical Form.* Cambridge, England: Galloway and Porter, 1958.

Greenhill, A. G. *The Applications of Elliptic Functions.* London: Macmillan, 1892.

Hancock, H. *Lectures on the Theory of Elliptic Functions.* New York: Wiley, 1910.

Jacobi, C. G. J. *Fundamentia Nova Theoriae Functionum Ellipticarum.* Regiomonti, Sumtibus fratrum Borntraeger, 1829.

King, L. V. *On the Direct Numerical Calculation of Elliptic Functions and Integrals.* Cambridge, England: Cambridge University Press, 1924.

Lang, S. *Elliptic Functions, 2nd ed.* New York: Springer-Verlag, 1987.

Lawden, D. F. *Elliptic Functions and Applications.* New York: Springer Verlag, 1989.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 427 and 433–434, 1953.

Murty, M. R. (Ed.). *Theta Functions.* Providence, RI: Amer. Math. Soc., 1993.

Neville, E. H. *Jacobian Elliptic Functions, 2nd ed.* Oxford, England: Clarendon Press, 1951.

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. "Elliptic Function Identities." §1.8 in *A=B.* Wellesley, MA: A. K. Peters, pp. 13–15, 1996.

Whittaker, E. T. and Watson, G. N. Chs. 20–22 in *A Course of Modern Analysis, 4th ed.* Cambridge, England: University Press, 1943.

Elliptic Functional

see COERCIVE FUNCTIONAL

Elliptic Geometry

A constant curvature NON-EUCLIDEAN GEOMETRY which replaces the PARALLEL POSTULATE with the statement "through any point in the plane, there exist no lines PARALLEL to a given line." Elliptic geometry is sometimes also called RIEMANNIAN GEOMETRY. It can be visualized as the surface of a SPHERE on which "lines" are taken as GREAT CIRCLES. In elliptic geometry, the sum of angles of a TRIANGLE is $> 180°$.

see also EUCLIDEAN GEOMETRY, HYPERBOLIC GEOMETRY, NON-EUCLIDEAN GEOMETRY

Elliptic Group Modulo p

$E(a,b)/p$ denotes the elliptic GROUP modulo p whose elements are 1 and ∞ together with the pairs of INTEGERS (x,y) with $0 \leq x, y < p$ satisfying

$$y^2 \equiv x^3 + ax + b \pmod{p} \qquad (1)$$

with a and b INTEGERS such that

$$4a^3 + 27b^2 \not\equiv 0 \pmod{p}. \qquad (2)$$

Given (x_1, y_1), define

$$(x_i, y_i) \equiv (x_1, y_1)^i \pmod{p}. \qquad (3)$$

The ORDER h of $E(a,b)/p$ is given by

$$h = 1 + \sum_{x=1}^{p} \left[\left(\frac{x^3 + ax + b}{p} \right) + 1 \right], \qquad (4)$$

where $(x^3 + ax + b/p)$ is the LEGENDRE SYMBOL, although this FORMULA quickly becomes impractical. However, it has been proven that

$$p + 1 - 2\sqrt{p} \leq h(E(a,b)/p) \leq p + 1 + 2\sqrt{p}. \qquad (5)$$

Furthermore, for p a PRIME > 3 and and INTEGER n in the above interval, there exists a and b such that

$$h(E(a,b)/p) = n, \qquad (6)$$

and the orders of elliptic GROUPS mod p are nearly uniformly distributed in the interval.

Elliptic Helicoid

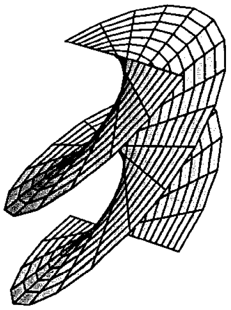

A generalization of the HELICOID to the parametric equations

$$x(u,v) = av\cos u$$
$$y(u,v) = bv\sin u$$
$$z(u,v) = cu.$$

see also HELICOID

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 264, 1993.

Elliptic Hyperboloid

The elliptic hyperboloid is the generalization of the HYPERBOLOID to three distinct semimajor axes. The elliptic hyperboloid of one sheet is a RULED SURFACE and has Cartesian equation

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1, \tag{1}$$

and parametric equations

$$x(u,v) = a\sqrt{1+u^2}\cos v \tag{2}$$
$$y(u,v) = b\sqrt{1+u^2}\sin v \tag{3}$$
$$z(u,v) = cu \tag{4}$$

for $v \in [0, 2\pi)$, or

$$x(u,v) = a(\cos u \mp v\sin u) \tag{5}$$
$$y(u,v) = b(\sin u \pm v\cos u) \tag{6}$$
$$z(u,v) = \pm cv, \tag{7}$$

or

$$x(u,v) = a\cosh v\cos u \tag{8}$$
$$y(u,v) = b\cosh v\sin u \tag{9}$$
$$z(u,v) = c\sinh v. \tag{10}$$

The two-sheeted elliptic hyperboloid oriented along the z-AXIS has Cartesian equation

$$\frac{x^2}{a^2}+\frac{y^2}{a^2}-\frac{z^2}{c^2}=-1, \tag{11}$$

and parametric equations

$$x = a\sinh u\cos v \tag{12}$$
$$y = b\sinh u\sin v \tag{13}$$
$$z = c \pm \cosh u. \tag{14}$$

The two-sheeted elliptic hyperboloid oriented along the x-AXIS has Cartesian equation

$$\frac{x^2}{a^2}-\frac{y^2}{a^2}-\frac{z^2}{c^2}=1 \tag{15}$$

and parametric equations

$$x = a\cosh u\cosh v \tag{16}$$
$$y = b\sinh u\cosh v \tag{17}$$
$$z = c\sinh v. \tag{18}$$

see also HYPERBOLOID, RULED SURFACE

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 296–297, 1993.

Elliptic Integral

An elliptic integral is an INTEGRAL of the form

$$\int \frac{A(x)+B(x)\sqrt{S(x)}}{C(x)+D(x)\sqrt{S(x)}}\,dx, \tag{1}$$

or

$$\int \frac{A(x)\,dx}{B(x)\sqrt{S(x)}}, \tag{2}$$

where A, B, C, and D are POLYNOMIALS in x and S is a POLYNOMIAL of degree 3 or 4. Another form is

$$\int R(w,x)\,dx, \tag{3}$$

where R is a RATIONAL FUNCTION of x and y, w^2 is a function of x CUBIC or QUADRATIC in x, $R(w,x)$ contains at least one ODD POWER of w, and w^2 has no repeated factors.

Elliptic integrals can be viewed as generalizations of the TRIGONOMETRIC FUNCTIONS and provide solutions to a wider class of problems. For instance, while the ARC LENGTH of a CIRCLE is given as a simple function of the parameter, computing the ARC LENGTH of an ELLIPSE requires an elliptic integral. Similarly, the position of a pendulum is given by a TRIGONOMETRIC FUNCTION as a function of time for small angle oscillations, but the full solution for arbitrarily large displacements requires the use of elliptic integrals. Many other problems in electromagnetism and gravitation are solved by elliptic integrals.

A very useful class of functions known as ELLIPTIC FUNCTIONS is obtained by inverting elliptic integrals (by analogy with the inverse trigonometric functions). ELLIPTIC FUNCTIONS (among which the JACOBI ELLIPTIC FUNCTIONS and WEIERSTRASS ELLIPTIC FUNCTION are the two most common forms) provide a powerful tool for analyzing many deep problems in NUMBER THEORY, as well as other areas of mathematics.

All elliptic integrals can be written in terms of three "standard" types. To see this, write

$$R(w,x) \equiv \frac{P(w,x)}{Q(w,x)} = \frac{wP(w,x)Q(-w,x)}{wQ(w,x)Q(-w,x)}. \tag{4}$$

But since $w^2 = f(x)$,

$$Q(w,x)Q(-w,x) \equiv Q_1(w,x) = Q_1(w,x), \tag{5}$$

then

$$\begin{aligned} wP(w,x)Q(-w,x) &= A + Bx + Cw + Dx^2 + Ewx \\ &\quad + Fw^2 + Gw^2x + Hw^3x \\ &= (A + Bx + Dx^2 + Fw^2 + Gw^2x) \\ &\quad + w(c + Ex + Hw^2x + \ldots) \\ &= P_1(x) + wP_2(x), \end{aligned} \tag{6}$$

so

$$R(w,x) = \frac{P_1(x) + wP_2(x)}{wQ_1(w)} = \frac{R_1(x)}{w} + R_2(x). \tag{7}$$

But any function $\int R_2(x)\,dx$ can be evaluated in terms of elementary functions, so the only portion that need be considered is

$$\frac{\int R_1(x)}{w}\,dx. \tag{8}$$

Now, any quartic can be expressed as $S_1 S_2$ where

$$S_1 \equiv a_1x^2 + 2b_1x + c_1 \tag{9}$$

$$S_2 \equiv a_2x^2 + 2b_2x + c_2. \tag{10}$$

The COEFFICIENTS here are real, since pairs of COMPLEX ROOTS are COMPLEX CONJUGATES

$$\begin{aligned} &[x - (R + Ii)][x - (R - Ii)] \\ &\quad = x^2 + x(-R + Ii - R - Ii) + (R^2 - I^2 i) \\ &\quad = x^2 - 2Rx + (R^2 + I^2). \end{aligned} \tag{11}$$

If all four ROOTS are real, they must be arranged so as not to interleave (Whittaker and Watson 1990, p. 514). Now define a quantity λ such that $S_1 + \lambda S_2$

$$(a_1 - \lambda a_2)x^2 - (2b_1 - 2b_2\lambda)x + (c_1 - \lambda c_2) \tag{12}$$

is a SQUARE NUMBER and

$$2\sqrt{(a_1 - \lambda a_2)(c_1 - \lambda_2)} = 2(b_1 - b_2\lambda) \tag{13}$$

$$(a_1 - \lambda a_2)(c_1 - \lambda c_2) - (b_1 - \lambda b_2)^2 = 0. \tag{14}$$

Call the ROOTS of this equation λ_1 and λ_2, then

$$\begin{aligned} S_1 - \lambda_1 S_2 &= \left[\sqrt{(a_1 - \lambda_1 a_2)x} + \sqrt{c_1 - \lambda c_2}\right]^2 \\ &= (a_1 - \lambda_1 a_2)\left(x + \sqrt{\frac{c_1 - \lambda_1 c_2}{a_1 - \lambda_1 a_2}}\right) \\ &\equiv (a_1 - \lambda_1 a_2)(x - \alpha)^2 \end{aligned} \tag{15}$$

$$\begin{aligned} S_1 - \lambda_2 S_2 &= \left[\sqrt{(a_1 - \lambda_1 a_2)x} + \sqrt{c_1 - \lambda c_2}\right]^2 \\ &= (a_1 - \lambda_1 a_2)\left(x + \sqrt{\frac{c_1 - \lambda_2 c_2}{a_1 - \lambda_2 a_2}}\right) \\ &\equiv (a_1 - \lambda_2 a_2)(x - \beta)^2. \end{aligned} \tag{16}$$

Taking (15)-(16) and $\lambda_2(1) - \lambda_1(2)$ gives

$$\begin{aligned} S_2(\lambda_2 - \lambda_1) &= (a_1 - \lambda_1 a_2)(x - \alpha)^2 \\ &\quad - (a_1 - \lambda_2 a_2)(x - \beta)^2 \end{aligned} \tag{17}$$

$$\begin{aligned} S_1(\lambda_2 - \lambda_1) &= \lambda_2(a_1 - \lambda_1 a_2)(x - \alpha)^2 \\ &\quad - \lambda_1(a_1 - \lambda_2 a_2)(x - \beta^2). \end{aligned} \tag{18}$$

Solving gives

$$\begin{aligned} S_1 &= \frac{a_1 - \lambda_1 a_2}{\lambda_2 - \lambda_1}(x - \alpha)^2 - \frac{a_1 - \lambda_2 a_2}{\lambda_2 - \lambda_1}(x - \beta)^2 \\ &\equiv A_1(x - \alpha)^2 + B_1(x - \beta)^2 \end{aligned} \tag{19}$$

$$\begin{aligned} S_2 &= \frac{\lambda_2(a_1 - \lambda_1 a_2)}{\lambda_2 - \lambda_1}(x - \alpha)^2 - \frac{\lambda_1(a_1 - \lambda_2 a_2)}{\lambda_2 - \lambda_1}(x - \beta)^2 \\ &\equiv A_2(x - \alpha)^2 + B_2(x - \beta)^2, \end{aligned} \tag{20}$$

so we have

$$\begin{aligned} w^2 &= S_1 S_2 \\ &= [A_1(x - \alpha)^2 + B_1(x - \beta)^2][A^2(x - \alpha)^2 + B^2(x - \beta)^2]. \end{aligned} \tag{21}$$

Now let

$$t \equiv \frac{x - \alpha}{x - \beta} \tag{22}$$

$$\begin{aligned} dy &= [(x - \beta)^{-1} - (x - \alpha)(x - \beta)^{-2}]\,dx \\ &= \frac{(x - \beta) - (x - \alpha)}{(x - \beta)^2}\,dx \\ &= \frac{\alpha - \beta}{(x - \beta)^2}\,dx, \end{aligned} \tag{23}$$

so

$$\begin{aligned} w^2 &= (x - \beta)^4\left[A_1\left(\frac{x - \alpha}{x - \beta}\right)^2 + B_1\right] \\ &\quad \times \left[A_2\left(\frac{x - \alpha}{x - \beta}\right) + B_2\right] \\ &= (x - \beta)^4(A_1t^2 + B_1)(A_2t^2 + B_2), \end{aligned} \tag{24}$$

and

$$w = (x - \beta)^2\sqrt{(A_1t^2 + B_1)(A_2t^2 + B_2)} \tag{25}$$

$$\begin{aligned} \frac{dx}{w} &= \left[\frac{(x - \beta)^2}{\alpha - \beta}\,dt\right]\frac{1}{(x - \beta)^2\sqrt{(A_1t^2 + B_1)(A_2t^2 + B_2)}} \\ &= \frac{dt}{(\alpha - \beta)\sqrt{(A_1t^2 + B_1)(A_2t^2 + B_2)}}. \end{aligned} \tag{26}$$

Now let

$$R_3(t) \equiv \frac{R_1(x)}{\alpha - \beta}, \tag{27}$$

so

$$\int \frac{R_1(x)\,dx}{w} = \int \frac{R_3(t)\,dt}{\sqrt{(A_1t^2 + B_1)(A_2t^2 + B_2)}}. \tag{28}$$

Rewriting the Even and Odd parts

$$R_3(t) + R_3(-t) \equiv 2R_4(t^2) \tag{29}$$

$$R_3(t) - R_3(-t) \equiv 2tR_5(t^2), \tag{30}$$

gives

$$R_3(t) \equiv \tfrac{1}{2}(R_{\text{even}} - R_{\text{odd}}) = R_4(t^2) + tR_5(t^2), \tag{31}$$

so we have

$$\int \frac{R_1(x)\,dx}{w} = \int \frac{R_4(t^2)\,dt}{\sqrt{(A_1t^2+B_1)(A_2t^2+B_2)}} + \int \frac{R_5(t^2)t\,dt}{\sqrt{(A_1t^2+B_1)(A_2t^2+B_2)}}. \tag{32}$$

Letting

$$u \equiv t^2 \tag{33}$$

$$du = 2t\,dt \tag{34}$$

reduces the second integral to

$$\frac{1}{2}\int \frac{R_5(u)\,du}{\sqrt{(A_1u+B_1)(A_2u+B_2)}}, \tag{35}$$

which can be evaluated using elementary functions. The first integral can then be reduced by Integration by Parts to one of the three Legendre elliptic integrals (also called Legendre-Jacobi Elliptic Integrals), known as incomplete elliptic integrals of the first, second, and third kind, denoted $F(\phi, k)$, $E(\phi, k)$, and $\Pi(n; \phi, k)$, respectively (von Kármán and Biot 1940, Whittaker and Watson 1990, p. 515). If $\phi = \pi/2$, then the integrals are called complete elliptic integrals and are denoted $K(k)$, $E(k)$, $\Pi(n; k)$.

Incomplete elliptic integrals are denoted using a Modulus k, Parameter $m \equiv k^2$, or Modular Angle $\alpha \equiv \sin^{-1} k$. An elliptic integral is written $I(\phi|m)$ when the Parameter is used, $I(\phi, k)$ when the Modulus is used, and $I(\phi\backslash\alpha)$ when the Modular Angle is used. Complete elliptic integrals are defined when $\phi = \pi/2$ and can be expressed using the expansion

$$(1 - k^2\sin^2\theta)^{-1/2} = \sum_{n=0}^{\infty} \frac{(2n-1)!!}{(2n)!!} k^{2n} \sin^{2n}\theta. \tag{36}$$

An elliptic integral in standard form

$$\int_a^x \frac{dx}{\sqrt{f(x)}}, \tag{37}$$

where

$$f(x) = a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0, \tag{38}$$

can be computed analytically (Whittaker and Watson 1990, p. 453) in terms of the Weierstraß Elliptic Function with invariants

$$g_2 = a_0a_4 - 4a_1a_3 + 3{a_2}^2 \tag{39}$$

$$g_3 = a_0a_2a_4 - 2a_1a_2a_3 - a_4{a_1}^2 - {a_3}^2a_0. \tag{40}$$

If $a \equiv x_0$ is a root of $f(x) = 0$, then the solution is

$$x = x_0 + \tfrac{1}{4}f'(x_0)[\wp(z; g_2, g_3) - \tfrac{1}{24}f''(x_0)]^{-1}. \tag{41}$$

For an arbitrary lower bound,

$$x = a + \frac{\sqrt{f(a)}\wp'(z) + \frac{1}{2}f'(a)[\wp(z) - \frac{1}{24}f''(a)] + \frac{1}{24}f(a)f'''(a)}{2[\wp(z) - \frac{1}{24}f''(a)]^2 - \frac{1}{48}f(a)f^{(a)}(a)}, \tag{42}$$

where $\wp(z) \equiv \wp(z; g_2, g_3)$ is a Weierstraß Elliptic Function.

A generalized elliptic integral can be defined by the function

$$T(a, b) \equiv \frac{2}{\pi}\int_0^{\pi/2} \frac{d\theta}{\sqrt{a^2\cos^2\theta + b^2\sin^2\theta}} \tag{43}$$

$$= \frac{2}{\pi}\int_0^{\pi/2}\int \frac{d\theta}{\cos\theta\sqrt{a^2 + b^2\tan^2\theta}} \tag{44}$$

(Borwein and Borwein 1987). Now let

$$t \equiv b\tan\theta \tag{45}$$

$$dt = b\sec^2\theta\,d\theta. \tag{46}$$

But

$$\sec\theta = \sqrt{1 + \tan^2\theta}, \tag{47}$$

so

$$dt = \frac{b}{\cos\theta}\sec\theta\,d\theta = \frac{b}{\cos\theta}\sqrt{1 + \tan^2\theta}\,d\theta = \frac{b}{\cos\theta}\sqrt{1 + \left(\frac{t}{b}\right)^2}\,d\theta = \frac{d\theta}{\cos\theta}\sqrt{b^2 + t^2}, \tag{48}$$

and

$$\frac{d\theta}{\cos\theta} = \frac{dt}{\sqrt{b^2 + t^2}}, \tag{49}$$

and the equation becomes

$$T(a, b) = \frac{2}{\pi}\int_0^\infty \frac{dt}{\sqrt{(a^2+t^2)(b^2+t^2)}} = \frac{1}{\pi}\int_{-\infty}^\infty \frac{dt}{\sqrt{(a^2+t^2)(b^2+t^2)}}. \tag{50}$$

Now we make the further substitution $u \equiv \frac{1}{2}(t - ab/t)$. The differential becomes

$$du = \tfrac{1}{2}(1 + ab/t^2)\,dt, \tag{51}$$

but $2u = t - ab/t$, so

$$2u/t = 1 - ab/t^2 \tag{52}$$

$$ab/t^2 = 1 - 2u/t \tag{53}$$

and

$$1 + ab/t^2 = 2 - 2u/t = 2(1 - u/t). \tag{54}$$

However, the left side is always positive, so

$$1 + ab/t^2 = 2 - 2u/t = 2|1 - u/t| \tag{55}$$

and the differential is

$$dt = \frac{du}{\left|1 - \frac{u}{t}\right|}. \tag{56}$$

We need to take some care with the limits of integration. Write (50) as

$$\int_{-\infty}^{\infty} f(t)\,dt = \int_{-\infty}^{0^-} f(t)\,dt + \int_{0+}^{\infty} f(t)\,dt. \tag{57}$$

Now change the limits to those appropriate for the u integration

$$\int_{-\infty}^{\infty} g(u)\,du + \int_{-\infty}^{\infty} g(u)\,du = 2\int_{-\infty}^{\infty} g(u)\,du, \tag{58}$$

so we have picked up a factor of 2 which must be included. Using this fact and plugging (56) in (50) therefore gives

$$T(a,b) = \frac{2}{\pi}\int_{-\infty}^{\infty} \frac{du}{\left|1 - \frac{u}{t}\right| \sqrt{a^2b^2 + (a^2+b^2)t^2 + t^4}}. \tag{59}$$

Now note that

$$u^2 = \frac{t^4 - 2abt^2 + a^2b^2}{4t^2} \tag{60}$$

$$4u^2t^2 = t^4 - 2abt^2 + 2abt^2 \tag{61}$$

$$a^2b^2 + t^4 = 4u^2t^2 + 2abt^2. \tag{62}$$

Plug (62) into (59) to obtain

$$\begin{aligned} T(a,b) &= \frac{2}{\pi}\int_{-\infty}^{\infty} \frac{du}{\left|1 - \frac{u}{t}\right| \sqrt{4u^2t^2 + 2abt^2 + (a^2+b^2)t^2}} \\ &= \frac{2}{\pi}\int_{-\infty}^{\infty} \frac{du}{|t-u|\sqrt{4u^2 + (a+b)^2}}. \end{aligned} \tag{63}$$

But

$$2ut = t^2 - ab \tag{64}$$

$$t^2 - 2ut - ab = 0 \tag{65}$$

$$t = \tfrac{1}{2}(2u \pm \sqrt{4u^2 + 4ab}) = u \pm \sqrt{u^2 + ab}, \tag{66}$$

so

$$t - u = \pm\sqrt{u^2 + ab}, \tag{67}$$

and (63) becomes

$$\begin{aligned} T(a,b) &= \frac{2}{\pi}\int_{-\infty}^{\infty} \frac{du}{\sqrt{[4u^2 + (a+b)^2](u^2 + ab)}} \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{du}{\sqrt{\left[u^2 + \left(\frac{a+b}{2}\right)^2\right](u^2 + ab)}}. \end{aligned} \tag{68}$$

We have therefore demonstrated that

$$T(a,b) = T(\tfrac{1}{2}(a+b), \sqrt{ab}). \tag{69}$$

We can thus iterate

$$a_{i+1} = \tfrac{1}{2}(a_i + b_i) \tag{70}$$

$$b_{i+1} = \sqrt{a_i b_i}, \tag{71}$$

as many times as we wish, without changing the value of the integral. But this iteration is the same as and therefore converges to the Arithmetic-Geometric Mean, so the iteration terminates at $a_i = b_i = M(a_0, b_0)$, and we have

$$\begin{aligned} T(a_0,b_0) &= T(M(a_0,b_0), M(a_0,b_0)) \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{dt}{M^2(a_0,b_0) + t^2} \\ &= \frac{1}{\pi M(a_0,b_0)}\left[\tan^{-1}\left(\frac{t}{M(a_0,b_0)}\right)\right]_{-\infty}^{\infty} \\ &= \frac{1}{\pi M(a_0,b_0)}\left[\frac{\pi}{2} - \left(\frac{\pi}{2}\right)\right] \\ &= \frac{1}{M(a_0,b_0)}. \end{aligned} \tag{72}$$

Complete elliptic integrals arise in finding the arc length of an Ellipse and the period of a pendulum. They also arise in a natural way from the theory of Theta Functions. Complete elliptic integrals can be computed using a procedure involving the Arithmetic-Geometric Mean. Note that

$$\begin{aligned} T(a,b) &\equiv \frac{2}{\pi}\int_0^{\pi/2} \frac{d\theta}{\sqrt{a^2\cos^2\theta + b^2\sin^2\theta}} \\ &= \frac{2}{\pi}\int_0^{\pi/2} \frac{d\theta}{a\sqrt{\cos^2\theta + \left(\frac{b}{a}\right)^2\sin^2\theta}} \\ &= \frac{2}{a\pi}\int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - \left(1 - \frac{b^2}{a^2}\right)^2\sin^2\theta}}. \end{aligned} \tag{73}$$

So we have

$$T(a,b) = \frac{2}{a\pi} K\left(1 - \frac{b^2}{a^2}\right) = \frac{1}{M(a,b)}, \tag{74}$$

where $K(k)$ is the complete ELLIPTIC INTEGRAL OF THE FIRST KIND. We are free to let $a \equiv a_0 \equiv 1$ and $b \equiv b_0 \equiv k'$, so

$$\frac{2}{\pi} K(\sqrt{1-k'^2}\,) = \frac{2}{\pi} K(k) = \frac{1}{M(1,k')}, \tag{75}$$

since $k \equiv \sqrt{1-k'^2}$, so

$$K(k) = \frac{\pi}{2M(1,k')}. \tag{76}$$

But the ARITHMETIC-GEOMETRIC MEAN is defined by

$$a_i = \tfrac{1}{2}(a_{i-1} + b_{i-1}) \tag{77}$$

$$b_i = \sqrt{a_{i-1} b_{i-1}} \tag{78}$$

$$c_i = \begin{cases} \tfrac{1}{2}(a_{i-1} - b_{i-1}) & i > 0 \\ \sqrt{{a_0}^2 - {b_0}^2} & i = 0, \end{cases} \tag{79}$$

where

$$c_{n-1} = \tfrac{1}{2} a_n - b_n = \frac{{c_n}^2}{4a_{n+1}} \leq \frac{{c_n}^2}{4M(a_0,b_0)}, \tag{80}$$

so we have

$$K(k) = \frac{\pi}{2a_N}, \tag{81}$$

where a_N is the value to which a_n converges. Similarly, taking instead $a_0' = 1$ and $b_0' = k$ gives

$$K'(k) = \frac{\pi}{2a_N'}. \tag{82}$$

Borwein and Borwein (1987) also show that defining

$$U(a,b) \equiv \frac{\pi}{2} \int_0^{\pi/2} \sqrt{a^2 \cos^2 + b^2 \sin^2 \theta}\, d\theta = aE'\left(\frac{b}{a}\right) \tag{83}$$

leads to

$$2U(a_{n+1}, b_{n+1}) - U(a_n, b_n) = a_n b_n T(a_n, b_n), \tag{84}$$

so

$$\frac{K(k) - E(k)}{K(k)} = \tfrac{1}{2}({c_0}^2 + 2{c_1}^2 + 2^2{c_2}^2 + \ldots + 2^n{c_n}^2) \tag{85}$$

for $a_0 \equiv 1$ and $b_0 \equiv k'$, and

$$\frac{K'(k) - E'(k)}{K'(k)} = \tfrac{1}{2}({c_0'}^2 + 2{c_1'}^2 + 2^2{c_2'}^2 + \ldots + 2^n{c_n'}^2). \tag{86}$$

The elliptic integrals satisfy a large number of identities. The complementary functions and moduli are defined by

$$K'(k) \equiv K(\sqrt{1-k^2}\,) = K(k'). \tag{87}$$

Use the identity of generalized elliptic integrals

$$T(a,b) = T(\tfrac{1}{2}(a+b), \sqrt{ab}\,) \tag{88}$$

to write

$$\begin{aligned} \frac{1}{a} K\left(\sqrt{1 - \frac{b^2}{a^2}}\right) &= \frac{2}{a+b} K\left(\sqrt{1 - \frac{4ab}{(a+b)^2}}\right) \\ &= \frac{2}{a+b} K\left(\sqrt{\frac{a^2 + b^2 - 2ab}{(a+b)^2}}\right) \\ &= \frac{2}{a+b} K\left(\frac{a-b}{a+b}\right) \end{aligned} \tag{89}$$

$$K\left(\sqrt{1 - \frac{b^2}{a^2}}\right) = \frac{2}{1 + \frac{b}{a}} K\left(\frac{1 - \frac{b}{a}}{1 + \frac{b}{a}}\right). \tag{90}$$

Define

$$k' \equiv \frac{b}{a}, \tag{91}$$

and use

$$k \equiv \sqrt{1 - k'^2}, \tag{92}$$

so

$$K(k) = \frac{2}{1+k'} K\left(\frac{1-k'}{1+k'}\right). \tag{93}$$

Now letting $l \equiv (1-k')/(1+k')$ gives

$$l(1+k') = 1 - k' \Rightarrow k'(l+1) = 1 - l \tag{94}$$

$$k' = \frac{1-l}{1+l} \tag{95}$$

$$\begin{aligned} k = \sqrt{1-k'^2} &= \sqrt{1 - \left(\frac{1-l}{1+l}\right)^2} \\ &= \sqrt{\frac{(1+l)^2 - (1-l)^2}{(1+l)^2}} = \frac{2\sqrt{l}}{1+l}, \end{aligned} \tag{96}$$

and

$$\begin{aligned} \tfrac{1}{2}(1+k') &= \frac{1}{2}\left(1 + \frac{1-l}{1+l}\right) = \frac{1}{2}\left[\frac{(1+l)+(1-l)}{1+l}\right] \\ &= \frac{1}{1+l}. \end{aligned} \tag{97}$$

Writing k instead of l,

$$K(k) = \frac{1}{k+1} K\left(\frac{2\sqrt{k}}{1+k}\right). \tag{98}$$

Similarly, from Borwein and Borwein (1987),

$$E(k) = \frac{1+k}{2} E\left(\frac{2\sqrt{k}}{1+k}\right) + \frac{k'^2}{2} K(k) \tag{99}$$

$$E(k) = (1+k')E\left(\frac{1-k'}{1+k'}\right) - k'K(k). \tag{100}$$

Expressions in terms of the complementary function can be derived from interchanging the moduli and their complements in (93), (98), (99), and (100).

$$\begin{aligned} K'(k) = K(k') &= \frac{2}{1+k} K\left(\frac{1-k}{1+k}\right) \\ &= \frac{2}{1+k} K'\left(\sqrt{1-\left(\frac{1-k}{1+k}\right)^2}\right) \\ &= \frac{2}{1+k} K'\left(\frac{2\sqrt{k}}{1+k}\right) \end{aligned} \tag{101}$$

$$K'(k) = \frac{1}{1+k'} K\left(\frac{2\sqrt{k'}}{1+k'}\right) = \frac{1}{1+k'} K'\left(\frac{1-k'}{1+k'}\right), \tag{102}$$

and

$$E'(k) = (1+k)E'\left(\frac{2\sqrt{k}}{1+k}\right) - kK'(k) \tag{103}$$

$$E'(k) = \left(\frac{1+k'}{2}\right) E'\left(\frac{1-k'}{1+k'}\right) + \frac{k^2}{2} K'(k). \tag{104}$$

Taking the ratios

$$\frac{K'(k)}{K(k)} = 2\frac{K'\left(\frac{2\sqrt{k}}{1+k}\right)}{K\left(\frac{2\sqrt{k}}{1+k}\right)} = \frac{1}{2}\frac{K'\left(\frac{1-k'}{1+k'}\right)}{K\left(\frac{1-k'}{1+k'}\right)} \tag{105}$$

gives the MODULAR EQUATION of degree 2. It is also true that

$$K(x) = \frac{4}{(1+\sqrt{x'})^2} K\left(\left[\frac{1-\sqrt[4]{1-x^4}}{1+\sqrt[4]{1-x^4}}\right]^2\right). \tag{106}$$

see also ABELIAN INTEGRAL, AMPLITUDE, ARGUMENT (ELLIPTIC INTEGRAL), CHARACTERISTIC (ELLIPTIC INTEGRAL), DELTA AMPLITUDE, ELLIPTIC FUNCTION, ELLIPTIC INTEGRAL OF THE FIRST KIND, ELLIPTIC INTEGRAL OF THE SECOND KIND, ELLIPTIC INTEGRAL OF THE THIRD KIND, ELLIPTIC INTEGRAL SINGULAR VALUE, HEUMAN LAMBDA FUNCTION, JACOBI ZETA FUNCTION, MODULAR ANGLE, MODULUS (ELLIPTIC INTEGRAL), NOME, PARAMETER

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Elliptic Integrals." Ch. 17 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 587–607, 1972.

Arfken, G. "Elliptic Integrals." §5.8 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 321–327, 1985.

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, 1987.

Hancock, H. *Elliptic Integrals.* New York: Wiley, 1917.

King, L. V. *The Direct Numerical Calculation of Elliptic Functions and Integrals.* London: Cambridge University Press, 1924.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Elliptic Integrals and Jacobi Elliptic Functions." §6.11 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 254–263, 1992.

Prudnikov, A. P.; Brychkov, Yu. A.; and Marichev, O. I. *Integrals and Series, Vol. 1: Elementary Functions.* New York: Gordon & Breach, 1986.

Timofeev, A. F. *Integration of Functions.* Moscow and Leningrad: GTTI, 1948.

von Kármán, T. and Biot, M. A. *Mathematical Methods in Engineering: An Introduction to the Mathematical Treatment of Engineering Problems.* New York: McGraw-Hill, p. 121, 1940.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Elliptic Integral of the First Kind

Let the MODULUS k satisfy $0 < k^2 < 1$. (This may also be written in terms of the PARAMETER $m \equiv k^2$ or MODULAR ANGLE $\alpha \equiv \sin^{-1} k$.) The incomplete elliptic integral of the first kind is then defined as

$$F(\phi, k) = \int_0^\phi \frac{d\theta}{\sqrt{1-k^2\sin^2\theta}}. \tag{1}$$

Let

$$t \equiv \sin\theta \tag{2}$$

$$dt = \cos\theta\, d\theta = \sqrt{1-t^2}\, d\theta \tag{3}$$

$$\begin{aligned} F(\phi, k) &= \int_0^{\sin\phi} \frac{1}{\sqrt{1-k^2t^2}} \frac{dt}{\sqrt{1-t^2}} \\ &= \int_0^{\sin\phi} \frac{dt}{\sqrt{(1-k^2t^2)(1-t^2)}}. \end{aligned} \tag{4}$$

Let

$$v \equiv \tan\theta \tag{5}$$

$$dv \equiv \sec^2\theta\, d\theta = (1+v^2)\, d\theta, \tag{6}$$

so the integral can also be written as

$$\begin{aligned} F(\phi, k) &= \int_0^{\tan\phi} \frac{1}{\sqrt{1-k^2\frac{v^2}{1+u^2}}} \frac{du}{1+v^2} \\ &= \int_0^{\tan\phi} \frac{dv}{\sqrt{1+v^2}\sqrt{(1+v^2)-k^2v^2}} \qquad (7) \\ &= \int_0^{\tan\phi} \frac{dv}{\sqrt{(1+v^2)(1+k'v^2)}}, \qquad (8) \end{aligned}$$

where $k'^2 \equiv 1 - k^2$ is the complementary MODULUS.

The integral

$$I = \frac{1}{\sqrt{2}} \int_0^{\theta_0} \frac{d\theta}{\sqrt{\cos\theta - \cos\theta_0}}, \tag{9}$$

which arises in computing the period of a pendulum, is also an elliptic integral of the first kind. Use

$$\cos\theta = 1 - 2\sin^2(\tfrac{1}{2}\theta) \tag{10}$$

$$\sin(\tfrac{1}{2}\theta) = \sqrt{\frac{1-\cos\theta}{2}} \tag{11}$$

to write

$$\begin{aligned} \sqrt{\cos\theta - \cos\theta_0} &= \sqrt{1 - 2\sin^2(\tfrac{1}{2}\theta) - \cos\theta_0} \\ &= \sqrt{1-\cos\theta_0}\sqrt{1 - \frac{2}{1-\cos\theta_0}\sin^2(\tfrac{1}{2}\theta)} \\ &= \sqrt{2}\sin(\tfrac{1}{2}\theta_0)\sqrt{1 - \csc^2(\tfrac{1}{2}\theta_0)\sin^2(\tfrac{1}{2}\theta)}, \end{aligned} \tag{12}$$

so

$$I = \frac{1}{2}\int_0^{\theta_0} \frac{d\theta}{\sin(\frac{1}{2}\theta_0)\sqrt{1-\csc^2(\frac{1}{2}\theta_0)\sin^2(\frac{1}{2}\theta)}}. \tag{13}$$

Now let

$$\sin(\tfrac{1}{2}\theta) = \sin(\tfrac{1}{2}\theta_0)\sin\phi, \tag{14}$$

so the angle θ is transformed to

$$\phi = \sin^{-1}\left(\frac{\sin\theta}{\theta_0}\right), \tag{15}$$

which ranges from 0 to $\pi/2$ as θ varies from 0 to θ_0. Taking the differential gives

$$\tfrac{1}{2}\cos(\tfrac{1}{2}\theta)\,d\theta = \sin(\tfrac{1}{2}\theta_0)\cos\phi\,d\phi, \tag{16}$$

or

$$\tfrac{1}{2}\sqrt{1-\sin^2(\tfrac{1}{2}\theta_0)\sin^2\phi}\,d\theta = \sin(\tfrac{1}{2}\theta_0)\cos\phi\,d\phi. \tag{17}$$

Plugging this in gives

$$\begin{aligned} I &= \int_0^{\pi/2} \frac{1}{\sqrt{1-\sin^2(\frac{1}{2}\theta_0)\sin^2\phi}} \frac{\sin(\frac{1}{2}\theta_0)\cos\phi\,d\phi}{\sin(\frac{1}{2}\theta_0)\sqrt{1-\sin^2\phi}} \\ &= \int_0^{\pi/2} \frac{d\phi}{\sqrt{1-\sin^2(\frac{1}{2}\theta_0)\sin^2\phi}} = K(\sin(\tfrac{1}{2}\theta_0)), \end{aligned} \tag{18}$$

so

$$I = \frac{1}{\sqrt{2}}\int_0^{\theta_0} \frac{d\theta}{\sqrt{\cos\theta - \cos\theta_0}} = K(\sin(\tfrac{1}{2}\theta_0)). \tag{19}$$

Making the slightly different substitution $\phi = \theta/2$, so $d\theta = 2\,d\phi$ leads to an equivalent, but more complicated expression involving an *incomplete* elliptic function of the first kind,

$$\begin{aligned} I &= 2\frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}}\csc(\tfrac{1}{2}\theta_0)\int_0^{\theta_0} \frac{d\phi}{\sqrt{1-\csc^2(\frac{1}{2}\theta_0)\sin^2\phi}} \\ &= \csc(\tfrac{1}{2}\theta_0)F(\tfrac{1}{2}\theta_0, \csc(\tfrac{1}{2}\theta_0)). \end{aligned} \tag{20}$$

Therefore, we have proven the identity

$$\csc x F(x, \csc x) = K(\sin x). \tag{21}$$

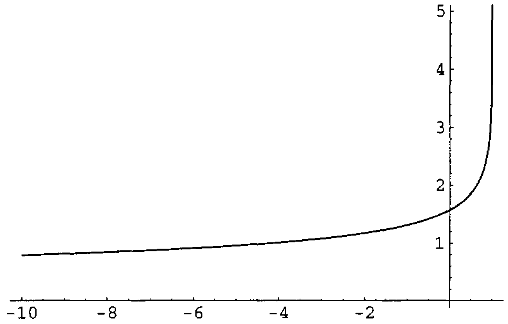

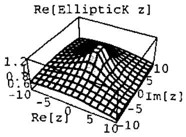

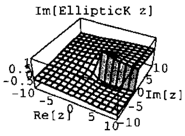

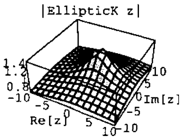

The complete elliptic integral of the first kind, illustrated above as a function of $m = k^2$, is defined by

$$\begin{aligned} K(k) &\equiv F(\tfrac{1}{2}\pi, k) && (22) \\ &= \sum_{n=0}^{\infty} \frac{(2n-1)!!}{(2n)!!} k^{2n} \int_0^{2\pi} \sin^{2n}\theta\,d\theta && (23) \\ &= \tfrac{1}{2}\pi{\vartheta_3}^2(q) && (24) \\ &= \sum_{n=0}^{\infty} \frac{(2n-1)!!}{(2n)!!} k^{2n} \frac{\pi}{2}\frac{(2n-1)!!}{(2n)!!} \\ &= \frac{\pi}{2}\sum_{n=0}^{\infty}\left[\frac{(2n-1)!!}{(2n)!!}\right]^2 k^{2n} && (25) \\ &= \tfrac{1}{2}\pi\,{}_2F_1(\tfrac{1}{2}, \tfrac{1}{2}, 1; k^2) && (26) \\ &= \frac{\pi}{2\sqrt{1-k^2}} P_{-1/2}\left(\frac{1+k^2}{1-k^2}\right), && (27) \end{aligned}$$

where

$$q = e^{-\pi K'(k)/K(k)} \tag{28}$$

is the NOME (for $|q| < 1$), ${}_2F_1(a,b;c;x)$ is the HYPERGEOMETRIC FUNCTION, and $P_n(x)$ is a LEGENDRE POLYNOMIAL. $K(k)$ satisfies the LEGENDRE RELATION

$$E(k)K'(k) + E'(k)K(k) - K(k)K'(k) = \tfrac{1}{2}\pi, \tag{29}$$

where $E(k)$ and $K(k)$ are complete elliptic integrals of the first and SECOND KINDS, and $E'(k)$ and $K'(k)$ are the complementary integrals. The modulus k is often suppressed for conciseness, so that $E(k)$ and $K(k)$ are often simply written E and K, respectively.

The DERIVATIVE of $K(k)$ is

$$\frac{dK}{dk} \equiv \int_0^1 \frac{dt}{\sqrt{(1-t^2)(1-k'^2t^2)}} = \frac{E(k)}{k(1-k^2)} - \frac{K(k)}{k} \tag{30}$$

$$\frac{d}{dk}\left(kk'^2\frac{dK}{dk}\right) = kK, \tag{31}$$

so

$$E = k(1-k^2)\left(\frac{dK}{dk}+\frac{K}{k}\right) = (1-k^2)\left(k\frac{dK}{dk}+k\right) \tag{32}$$

(Whittaker and Watson 1990, pp. 499 and 521).

see also AMPLITUDE, CHARACTERISTIC (ELLIPTIC INTEGRAL), ELLIPTIC INTEGRAL SINGULAR VALUE, GAUSS'S TRANSFORMATION, LANDEN'S TRANSFORMATION, LEGENDRE RELATION, MODULAR ANGLE, MODULUS (ELLIPTIC INTEGRAL), PARAMETER

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Elliptic Integrals." Ch. 17 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 587–607, 1972.

Spanier, J. and Oldham, K. B. "The Complete Elliptic Integrals $K(p)$ and $E(p)$" and "The Incomplete Elliptic Integrals $F(p;\phi)$ and $E(p;\phi)$." Chs. 61–62 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 609–633, 1987.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Elliptic Integral of the Second Kind

Let the MODULUS k satisfy $0 < k^2 < 1$. (This may also be written in terms of the PARAMETER $m \equiv k^2$ or MODULAR ANGLE $\alpha \equiv \sin^{-1} k$.) The incomplete elliptic integral of the second kind is then defined as

$$E(\phi,k) \equiv \int_0^\phi \sqrt{1-k^2\sin^2\theta}\,d\theta. \tag{1}$$

A generalization replacing $\sin\theta$ with $\sinh\theta$ gives

$$-iE(i\phi,-k) = \int_0^\phi \sqrt{1-k^2\sinh^2\theta}\,d\theta. \tag{2}$$

To place the elliptic integral of the second kind in a slightly different form, let

$$t \equiv \sin\theta \tag{3}$$

$$dt = \cos\theta\,d\theta = \sqrt{1-t^2}\,d\theta, \tag{4}$$

so the elliptic integral can also be written as

$$\begin{aligned} E(\phi,k) &= \int_0^{\sin\phi} \sqrt{1-k^2t^2}\,\frac{dt}{\sqrt{1-t^2}} \\ &= \int_0^{\sin\phi} \sqrt{\frac{1-k^2t^2}{1-t^2}}\,dt. \end{aligned} \tag{5}$$

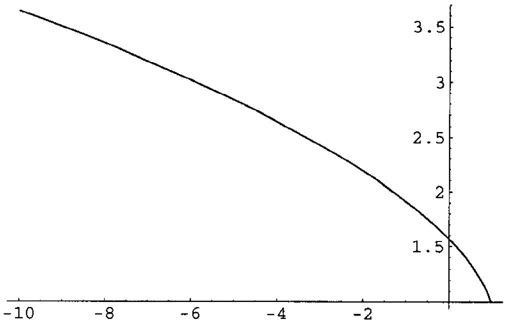

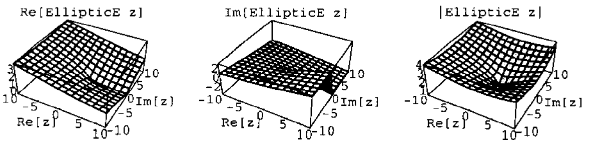

The complete elliptic integral of the second kind, illustrated above as a function of the PARAMETER m, is defined by

$$\begin{aligned} E(k) &\equiv E(\tfrac{1}{2}\pi,k) && (6)\\ &= \frac{\pi}{2}\left\{1-\sum_{n=1}^{\infty}\left[\frac{(2n-1)!!}{(2n)!!)}\right]^2 \frac{k^{2n}}{2n-1}\right\} && (7)\\ &= \tfrac{1}{2}\pi\,{}_2F_1(-\tfrac{1}{2},\tfrac{1}{2},1;k^2) && (8)\\ &= \int_0^K \mathrm{dn}^2\,u\,du, && (9) \end{aligned}$$

where ${}_2F_1(a,b;c;x)$ is the HYPERGEOMETRIC FUNCTION and $\mathrm{dn}\,u$ is a JACOBI ELLIPTIC FUNCTION. The complete elliptic integral of the second kind satisfies the LEGENDRE RELATION

$$E(k)K'(k) + E'(k)K(k) - K(k)K'(k) = \tfrac{1}{2}\pi, \tag{10}$$

where E and K are complete ELLIPTIC INTEGRALS OF THE FIRST and second kinds, and E' and K' are the complementary integrals. The DERIVATIVE is

$$\frac{dE}{dk} = \frac{E(k)-K(k)}{k} \tag{11}$$

(Whittaker and Watson 1990, p. 521). If k_r is a singular value (i.e.,

$$k_r = \lambda^*(r), \tag{12}$$

where λ^* is the ELLIPTIC LAMBDA FUNCTION), and $K(k_r)$ and the ELLIPTIC ALPHA FUNCTION $\alpha(r)$ are also known, then

$$E(k) = \frac{K(k)}{\sqrt{r}}\left[\frac{\pi}{3[K(k)]^2} - \alpha(r)\right] + K(k). \tag{13}$$

see also ELLIPTIC INTEGRAL OF THE FIRST KIND, ELLIPTIC INTEGRAL OF THE THIRD KIND, ELLIPTIC INTEGRAL SINGULAR VALUE

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Elliptic Integrals." Ch. 17 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 587–607, 1972.

Spanier, J. and Oldham, K. B. "The Complete Elliptic Integrals $K(p)$ and $E(p)$" and "The Incomplete Elliptic Integrals $F(p;\phi)$ and $E(p;\phi)$." Chs. 61 and 62 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 609–633, 1987.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Elliptic Integral of the Third Kind

Let $0 < k^2 < 1$. The incomplete elliptic integral of the third kind is then defined as

$$\Pi(n;\phi,k) = \int_0^\phi \frac{d\theta}{(1-n\sin^2\theta)\sqrt{1-k^2\sin^2\theta}} \tag{1}$$

$$= \int_0^{\sin\phi} \frac{dt}{(1-nt^2)\sqrt{(1-t^2)(1-k^2t^2)}}, \tag{2}$$

where n is a constant known as the CHARACTERISTIC.

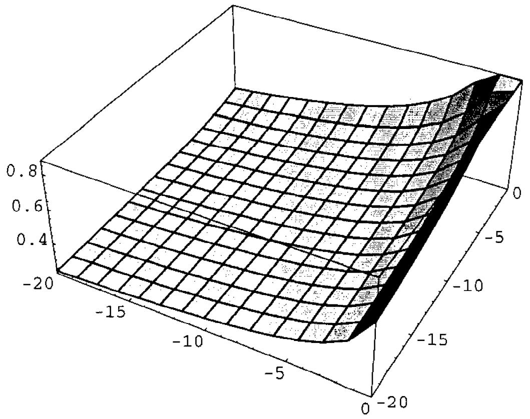

The complete elliptic integral of the second kind

$$\Pi(n|m) = \Pi(n; \tfrac{1}{2}\pi|m) \tag{3}$$

is illustrated above.

see also ELLIPTIC INTEGRAL OF THE FIRST KIND, ELLIPTIC INTEGRAL OF THE SECOND KIND, ELLIPTIC INTEGRAL SINGULAR VALUE

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Elliptic Integrals" and "Elliptic Integrals of the Third Kind." Ch. 17 and §17.7 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 587–607, 1972.

Elliptic Integral Singular Value

When the MODULUS k has a singular value, the complete elliptic integrals may be computed in analytic form in terms of GAMMA FUNCTIONS. Abel (quoted in Whittaker and Watson 1990, p. 525) proved that whenever

$$\frac{K'(k)}{K(k)} = \frac{a+b\sqrt{n}}{c+d\sqrt{n}}, \tag{1}$$

where a, b, c, d, and n are INTEGERS, $K(k)$ is a complete ELLIPTIC INTEGRAL OF THE FIRST KIND, and $K'(k) \equiv K(\sqrt{1-k^2})$ is the complementary complete ELLIPTIC INTEGRAL OF THE FIRST KIND, then the MODULUS k is the ROOT of an algebraic equation with INTEGER COEFFICIENTS.

A MODULUS k_r such that

$$\frac{K'(k_r)}{K(k_r)} = \sqrt{r}, \tag{2}$$

is called a singular value of the elliptic integral. The ELLIPTIC LAMBDA FUNCTION $\lambda^*(r)$ gives the value of k_r. Selberg and Chowla (1967) showed that $K(\lambda^*(r))$ and $E(\lambda^*(r))$ are expressible in terms of a finite number of GAMMA FUNCTIONS. The complete ELLIPTIC INTEGRALS OF THE SECOND KIND $E(k_r)$ and $E'(k_r)$ can be expressed in terms of $K(k_r)$ and $K'(k_r)$ with the aid of the ELLIPTIC ALPHA FUNCTION $\alpha(r)$.

The following table gives the values of $K(k_r)$ for small integral r in terms of GAMMA FUNCTIONS.

$$K(k_1) = \frac{\Gamma^2(\frac{1}{4})}{4\sqrt{\pi}}$$

$$K(k_2) = \frac{\sqrt{\sqrt{2}+1}\,\Gamma(\frac{1}{8})\Gamma(\frac{3}{8})}{2^{13/4}\sqrt{\pi}}$$

$$K(k_3) = \frac{3^{1/4}\Gamma^3(\frac{1}{3})}{2^{7/3}\pi}$$

$$K(k_4) = \frac{(\sqrt{2}+1)\Gamma^2(\frac{1}{4})}{2^{7/2}\sqrt{\pi}}$$

$$K(k_5) = (\sqrt{5}+2)^{1/4}\sqrt{\frac{\Gamma(\frac{1}{20})\Gamma(\frac{3}{20})\Gamma(\frac{7}{20})\Gamma(\frac{9}{20})}{160\pi}}$$

$$K(k_6) = \sqrt{(\sqrt{2}-1)(\sqrt{3}+\sqrt{2})(2+\sqrt{3})} \times \sqrt{\frac{\Gamma(\frac{1}{24})\Gamma(\frac{5}{24})\Gamma(\frac{7}{24})\Gamma(\frac{11}{24})}{384\pi}}$$

$$K(k_7) = \frac{\Gamma(\frac{1}{7})\Gamma(\frac{2}{7})\Gamma(\frac{4}{7})}{7^{1/4}\cdot 4\pi}$$

$$K(k_8) = \sqrt{\frac{2\sqrt{2}+\sqrt{1+5\sqrt{2}}}{4\sqrt{2}}}\,\frac{(\sqrt{2}+1)^{1/4}\Gamma(\frac{1}{8})\Gamma(\frac{3}{8})}{8\sqrt{\pi}}$$

$$K(k_9) = \frac{3^{1/4}\sqrt{2+\sqrt{3}}}{12\sqrt{\pi}\Gamma^2(\frac{1}{4})}$$

$$K(k_{10}) = \sqrt{(2+3\sqrt{2}+\sqrt{5})} \times \sqrt{\frac{\Gamma(\frac{1}{40})\Gamma(\frac{7}{40})\Gamma(\frac{9}{40})\Gamma(\frac{11}{40})\Gamma(\frac{13}{40})\Gamma(\frac{19}{40})\Gamma(\frac{23}{40})\Gamma(\frac{37}{40})}{2560\pi^3}}$$

$$K(k_{11}) = [2+(17+3\sqrt{33})^{1/3}-(3\sqrt{33}-17)^{1/3}]^2 \times \frac{\Gamma(\frac{1}{11})\Gamma(\frac{3}{11})\Gamma(\frac{4}{11})\Gamma(\frac{5}{11})\Gamma(\frac{9}{11})}{11^{1/4}144\pi^2}$$

$$K(k_{12}) = \frac{3^{1/4}(\sqrt{2}+1)(\sqrt{3}+\sqrt{2})\sqrt{2-\sqrt{3}}\Gamma^3(\frac{1}{3})}{2^{13/3}\pi}$$

$$K(k_{13}) = \frac{(18+5\sqrt{13})^{1/4}}{\sqrt{6656\pi^5}} \times \sqrt{\Gamma(\tfrac{1}{52})\Gamma(\tfrac{7}{52})\Gamma(\tfrac{9}{52})\Gamma(\tfrac{11}{52})\Gamma(\tfrac{15}{52})\Gamma(\tfrac{17}{52})} \times \sqrt{\Gamma(\tfrac{19}{52})\Gamma(\tfrac{25}{52})\Gamma(\tfrac{29}{52})\Gamma(\tfrac{31}{52})\Gamma(\tfrac{47}{52})\Gamma(\tfrac{49}{52})}$$

$$K(k_{15}) = \sqrt{\frac{(\sqrt{5}+1)\Gamma(\frac{1}{15})\Gamma(\frac{2}{15})\Gamma(\frac{4}{15})\Gamma(\frac{8}{15})}{240\pi}}$$

$$K(k_{16}) = \frac{(2^{1/4}+1)^2\Gamma^2(\frac{1}{4})}{2^{9/2}\sqrt{\pi}}$$

$$K(k_{17}) = C_1\left[\frac{\Gamma(\frac{1}{68})\Gamma(\frac{3}{68})\Gamma(\frac{7}{68})\Gamma(\frac{11}{68})\Gamma(\frac{13}{68})}{\Gamma(\frac{5}{68})\Gamma(\frac{15}{68})\Gamma(\frac{19}{68})\Gamma(\frac{29}{68})}\right]^{1/4} \times [\Gamma(\tfrac{21}{68})\Gamma(\tfrac{25}{68})\Gamma(\tfrac{27}{68})\Gamma(\tfrac{31}{68})\Gamma(\tfrac{33}{68})]^{1/4}$$

$$K(k_{25}) = \frac{\sqrt{5}+2}{20}\frac{\Gamma^2(\frac{1}{4})}{\sqrt{\pi}},$$

where $\Gamma(z)$ is the GAMMA FUNCTION and C_1 is an algebraic number (Borwein and Borwein 1987, p. 298).

Borwein and Zucker (1992) give amazing expressions for singular values of complete elliptic integrals in terms of CENTRAL BETA FUNCTIONS

$$\beta(p) \equiv B(p,p). \tag{3}$$

Furthermore, they show that $K(k_n)$ is *always* expressible in terms of these functions for $n \equiv 1, 2 \pmod 4$. In such cases, the Γ functions appearing in the expression are of the form $\Gamma(t/4n)$ where $1 \le t \le (2n-1)$ and $(t, 4n) = 1$. The terms in the numerator depend on the sign of the KRONECKER SYMBOL $\{t/4n\}$. Values for the first few n are

$$K(k_1) = 2^{-2}\beta(\tfrac{1}{4})$$

$$K(k_2) = 2^{-13/4}\beta(\tfrac{1}{8})$$

$$K(k_3) = 2^{-4/3}3^{-1/4}\beta(\tfrac{1}{3}) = 2^{-5/3}3^{-3/4}\beta(\tfrac{1}{6})$$

$$K(k_5) = 2^{-33/20}5^{-5/8}(11+5\sqrt{5})^{1/4}\sin(\tfrac{1}{20}\pi)\beta(\tfrac{1}{2}) = 2^{-29/20}5^{-3/8}(1+\sqrt{5})^{1/4}\sin(\tfrac{3}{20}\pi)\beta(\tfrac{3}{20})$$

$$K(k_6) = 2^{-47/12}3^{-3/4}(\sqrt{2}-1)(\sqrt{3}+1)\beta(\tfrac{1}{24}) = 2^{-43/12}3^{-1/4}(\sqrt{3}-1)\beta(\tfrac{5}{24})$$

$$K(k_7) = 2\cdot 7^{-3/4}\sin(\tfrac{1}{7}\pi)\sin(\tfrac{2}{7}\pi)B(\tfrac{1}{7},\tfrac{2}{7}) = 2^{-2/7}7^{-1/4}\frac{\beta(\frac{1}{7})\beta(\frac{2}{7})}{\beta(\frac{1}{14})}$$

$$K(k_{10}) = 2^{-61/20}5^{-1/4}(\sqrt{5}-2)^{1/2}(\sqrt{10}+3)\frac{\beta(\frac{1}{8})\beta(\frac{7}{40})}{\beta(\frac{1}{3}40)} = 2^{-15/4}5^{-3/4}(\sqrt{5}-2)^{1/2}\frac{\beta(\frac{1}{40})\beta(\frac{1}{9}40)}{\beta(\frac{3}{8})}$$

$$K(k_{11}) = R\cdot 2^{-7/11}\sin(\tfrac{1}{11}\pi)\sin(\tfrac{3}{11}\pi)B(\tfrac{1}{22},\tfrac{3}{22})$$

$$K(k_{13}) = 2^{-3}13^{-5/8}(5\sqrt{13}+18)^{1/4} \times [\tan(\tfrac{1}{52}\pi)\tan(\tfrac{3}{52}\pi)\tan(\tfrac{9}{52}\pi)]^{1/2}\frac{\beta(\frac{1}{52})\beta(\frac{9}{52})}{\beta(\frac{23}{52})}$$

$$K(k_{14}) = \sqrt{\sqrt{4\sqrt{2}+2}+\sqrt{2}+\sqrt{2\sqrt{2}-1}}\cdot 2^{-13/4}7^{-3/8} \times \left[\frac{\tan(\frac{5}{56}\pi)\tan(\frac{13}{56}\pi)}{\tan(\frac{11}{56}\pi)}\right]^{1/4}\sqrt{\frac{\beta(\frac{5}{56})\beta(\frac{13}{56})\beta(\frac{1}{8})}{\beta(\frac{11}{56})}}$$

$$K(k_{15}) = 2^{-1}3^{-3/4}5^{-7/12}B(\tfrac{1}{15},\tfrac{4}{15}) = \frac{2^{-2}3^{-3/4}5^{-3/4}(\sqrt{5}-1)\beta(\frac{1}{15})\beta(\frac{4}{15})}{\beta(\frac{1}{3})}$$

$$K(k_{17}) = C_2\left[\frac{\beta(\frac{1}{68})\beta(\frac{3}{68}(\beta(\frac{7}{68})\beta(\frac{9}{68})\beta(\frac{11}{68})\beta(\frac{13}{68})}{\beta(\frac{5}{68})\beta(\frac{15}{68})}\right]^{1/4},$$

where R is the REAL ROOT of

$$x^3 - 4x = 4 = 0 \tag{4}$$

and C_2 is an algebraic number (Borwein and Zucker 1992). Note that $K(k_{11})$ is the only value in the above list which cannot be expressed in terms of CENTRAL BETA FUNCTIONS.

Using the ELLIPTIC ALPHA FUNCTION, the ELLIPTIC INTEGRALS OF THE SECOND KIND can also be found from

$$E = \frac{\pi}{4\sqrt{r}K} + \left[1 - \frac{\alpha(r)}{\sqrt{r}}\right]K \tag{5}$$

$$E' = \frac{\pi}{4K} + \alpha(r)K, \tag{6}$$

and by definition,

$$K' = K\sqrt{n}. \tag{7}$$

see also CENTRAL BETA FUNCTION, ELLIPTIC ALPHA FUNCTION, ELLIPTIC DELTA FUNCTION, ELLIPTIC INTEGRAL OF THE FIRST KIND, ELLIPTIC INTEGRAL OF THE SECOND KIND, ELLIPTIC LAMBDA FUNCTION, GAMMA FUNCTION, MODULUS (ELLIPTIC INTEGRAL)

References

Abel, N. *J. für Math* 3, 184, 1881. Reprinted in Abel, N. H. *Oeuvres Completes* (Ed. L. Sylow and S. Lie). New York: Johnson Reprint Corp., p. 377, 1988.

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, pp. 139 and 298, 1987.

Borwein, J. M. and Zucker, I. J. "Elliptic Integral Evaluation of the Gamma Function at Rational Values of Small

Denominator." *IMA J. Numerical Analysis* **12**, 519–526, 1992.

Bowman, F. *Introduction to Elliptic Functions, with Applications.* New York: Dover, pp. 75, 95, and 98, 1961.

Glasser, M. L. and Wood, V. E. "A Closed Form Evaluation of the Elliptic Integral." *Math. Comput.* **22**, 535–536, 1971.

Selberg, A. and Chowla, S. "On Epstein's Zeta-Function." *J. Reine. Angew. Math.* **227**, 86–110, 1967.

Weisstein, E. W. "Elliptic Singular Values." `http://www.astro.virginia.edu/~eww6n/math/notebooks/EllipticSingular.m`.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, pp. 524–528, 1990.

Wrigge, S. "An Elliptic Integral Identity." *Math. Comput.* **27**, 837–840, 1973.

Zucker, I. J. "The Evaluation in Terms of Γ-Functions of the Periods of Elliptic Curves Admitting Complex Multiplication." *Math. Proc. Cambridge Phil. Soc.* **82**, 111–118, 1977.

Elliptic Integral Singular Value—k_1

The first SINGULAR VALUE k_1, corresponding to

$$K'(k_1) = K(k_1), \tag{1}$$

is given by

$$k_1 = \frac{1}{\sqrt{2}} \tag{2}$$

$$k_1' = \frac{1}{\sqrt{2}}. \tag{3}$$

As shown in LEMNISCATE FUNCTION,

$$\begin{aligned} K\left(\frac{1}{\sqrt{2}}\right) &\equiv \int_0^1 \frac{dt}{\sqrt{(1-t^2)\left(1-\frac{1}{2}t^2\right)}} \\ &= \sqrt{2}\int_0^1 \frac{dt}{\sqrt{1-t^4}}. \end{aligned} \tag{4}$$

Let

$$u \equiv t^4 \tag{5}$$

$$du = 4t^3\,dt = 4u^{3/4}\,dt \tag{6}$$

$$dt = \tfrac{1}{4}u^{-3/4}\,du, \tag{7}$$

then

$$\begin{aligned} K\left(\frac{1}{\sqrt{2}}\right) &= \frac{\sqrt{2}}{4}\int_0^1 u^{-3/4}(1-u)^{-1/2}\,du \\ &= \frac{\sqrt{2}}{4}B(\tfrac{1}{4},\tfrac{1}{2}) = \frac{\Gamma(\frac{1}{4})\Gamma(\frac{1}{2})}{\Gamma(\frac{3}{4})}\frac{\sqrt{2}}{4}, \end{aligned} \tag{8}$$

where $B(a,b)$ is the BETA FUNCTION and $\Gamma(z)$ is the GAMMA FUNCTION. Now use

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi} \tag{9}$$

and

$$\frac{1}{\Gamma(1-x)} = \frac{\sin(\pi x)}{\pi}\Gamma(x), \tag{10}$$

so

$$\frac{1}{\Gamma\left(\frac{3}{4}\right)} = \frac{1}{\Gamma\left(1-\frac{1}{4}\right)} = \frac{\sin\left(\frac{\pi}{4}\right)}{\pi}\Gamma(\tfrac{1}{4}) = \frac{1}{\pi\sqrt{2}}\Gamma(\tfrac{1}{4}). \tag{11}$$

Therefore,

$$K\left(\frac{1}{\sqrt{2}}\right) = \frac{\Gamma^2(\frac{1}{4})\sqrt{\pi}\sqrt{2}}{4\pi\sqrt{2}} = \frac{\Gamma^2(\frac{1}{4})}{4\sqrt{\pi}}. \tag{12}$$

Now consider

$$E\left(\frac{1}{\sqrt{2}}\right) \equiv \int_0^1 \sqrt{\frac{1-\frac{1}{2}t^2}{1-t^2}}\,dt. \tag{13}$$

Let

$$t^2 \equiv 1-u^2 \tag{14}$$

$$2t\,dt = -2u\,du \tag{15}$$

$$dt = -\frac{1}{t}u\,du = u(1-u^2)^{-1/2}\,du, \tag{16}$$

so

$$\begin{aligned} E\left(\frac{1}{\sqrt{2}}\right) &= \int_0^1 \sqrt{\frac{1-\frac{1}{2}(1-u^2)}{1-(1-u^2)}}\,u(1-u^2)^{-1/2}\,du \\ &= \int_0^1 \frac{\sqrt{\frac{1}{2}(1+u^2)}}{u}u(1-u^2)^{-1/2}\,du \\ &= \frac{1}{\sqrt{2}}\int_0^1 \sqrt{\frac{1+u^2}{1-u^2}}\,du. \end{aligned} \tag{17}$$

Now note that

$$\begin{aligned} \left(\frac{1}{\sqrt{1-u^4}}+\frac{u^2}{\sqrt{1-u^4}}\right)^2 &= \frac{(1+u^2)^2}{1-u^4} \\ &= \frac{(1+u^2)^2}{(1+u^2)(1-u^2)} = \frac{1+u^2}{1-u^2}, \end{aligned} \tag{18}$$

so

$$\begin{aligned} E\left(\frac{1}{\sqrt{2}}\right) &= \frac{1}{\sqrt{2}}\int_0^1 \sqrt{\frac{1+u^2}{1-u^2}}\,du \\ &= \frac{1}{\sqrt{2}}\int_0^1 \left(\frac{1}{\sqrt{1-u^4}}+\frac{u^2}{\sqrt{1-u^4}}\right)du \\ &= \frac{1}{2}K\left(\frac{1}{\sqrt{2}}\right)+\frac{1}{\sqrt{2}}\int_0^1 \frac{u^2\,du}{\sqrt{1-u^4}}. \end{aligned} \tag{19}$$

Now let

$$t \equiv u^4 \tag{20}$$

$$dt = 4u^3\,du, \tag{21}$$

so

$$\begin{aligned}\int_0^1 \frac{u^2\,du}{\sqrt{1-u^4}} &= \frac{1}{4}\int_0^1 t^{1/2}t^{-3/4}(1-t)^{-1/2}\,dt \\ &= \tfrac{1}{4}\int_0^1 t^{-1/4}(1-t)^{-1/2}\,dt \\ &= \tfrac{1}{4}B(\tfrac{3}{4},\tfrac{1}{2}) = \frac{\Gamma(\frac{3}{4})\Gamma(\frac{1}{2})}{4\Gamma(\frac{5}{4})}. \qquad (22)\end{aligned}$$

But

$$\begin{aligned}[\Gamma(\tfrac{5}{4})]^{-1} &= [\tfrac{1}{4}\Gamma(\tfrac{1}{4})]^{-1} \qquad (23)\\ \Gamma(\tfrac{3}{4}) &= \pi\sqrt{2}\,[\Gamma(\tfrac{1}{4})]^{-1} \qquad (24)\\ \Gamma(\tfrac{1}{2}) &= \sqrt{\pi}, \qquad (25)\end{aligned}$$

so

$$\int_0^1 \frac{u^2\,du}{\sqrt{1-u^4}} = \frac{1}{4}\frac{\pi\sqrt{2}\cdot 4\sqrt{\pi}}{\Gamma^2(\frac{1}{4})} = \frac{\sqrt{2}\pi^{3/2}}{\Gamma^2(\frac{1}{4})} \qquad (26)$$

$$\begin{aligned}E\left(\frac{1}{\sqrt{2}}\right) &= \tfrac{1}{2}K + \frac{\pi^{3/2}}{\Gamma^2(\frac{1}{4})} = \frac{\Gamma^2(\frac{1}{4})}{8\sqrt{\pi}} + \frac{\pi^{3/2}}{\Gamma^2(\frac{1}{4})}\\ &= \frac{1}{4}\sqrt{\frac{\pi}{2}}\left[\frac{\Gamma(\frac{1}{4})}{\Gamma(\frac{3}{4})} + \frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{5}{4})}\right]. \qquad (27)\end{aligned}$$

Summarizing (12) and (27) gives

$$\begin{aligned}K\left(\frac{1}{\sqrt{2}}\right) &= \frac{\Gamma^2(\frac{1}{4})}{4\sqrt{\pi}}\\ K'\left(\frac{1}{\sqrt{2}}\right) &= \frac{\Gamma^2(\frac{1}{4})}{4\sqrt{\pi}}\\ E\left(\frac{1}{\sqrt{2}}\right) &= \frac{\Gamma^2(\frac{1}{4})}{8\sqrt{\pi}} + \frac{\pi^{3/2}}{\Gamma^2(\frac{1}{4})}\\ E'\left(\frac{1}{\sqrt{2}}\right) &= \frac{\Gamma^2(\frac{1}{4})}{8\sqrt{\pi}} + \frac{\pi^{3/2}}{\Gamma^2(\frac{1}{4})}.\end{aligned}$$

Elliptic Integral Singular Value—k_2

The second SINGULAR VALUE k_2, corresponding to

$$K'(k_2) = \sqrt{2}\,K(k_2), \qquad (1)$$

is given by

$$\begin{aligned}k_2 &= \tan\left(\frac{\pi}{8}\right) = \sqrt{2}-1, \qquad (2)\\ k_2' &= \sqrt{2}\,(\sqrt{2}-1). \qquad (3)\end{aligned}$$

For this modulus,

$$E(\sqrt{2}-1) = \frac{1}{4}\sqrt{\frac{\pi}{4}}\left[\frac{\Gamma(\frac{1}{8})}{\Gamma(\frac{5}{8})} + \frac{\Gamma(\frac{5}{8})}{\Gamma(\frac{9}{8})}\right]. \qquad (4)$$

Elliptic Integral Singular Value—k_3

The third SINGULAR VALUE k_3, corresponding to

$$K'(k_3) = \sqrt{3}\,K(k_3), \qquad (1)$$

is given by

$$k_3 = \sin\left(\frac{\pi}{12}\right) = \tfrac{1}{4}(\sqrt{6}-\sqrt{2}). \qquad (2)$$

As shown by Legendre,

$$K(k_3) = \frac{\sqrt{\pi}}{2\cdot 3^{3/4}}\frac{\Gamma(\frac{1}{6})}{\Gamma(\frac{2}{3})} \qquad (3)$$

(Whittaker and Watson 1990, p. 525). In addition,

$$\begin{aligned}E(k_3) &= \frac{\pi}{4\sqrt{3}}\frac{1}{K} + \frac{\sqrt{3}+1}{2\sqrt{3}}K\\ &= \frac{1}{4}\left(\frac{\pi}{\sqrt{3}}\right)^{1/2}\left[\left(1+\frac{1}{\sqrt{3}}\right)\frac{\Gamma(\frac{1}{3})}{\Gamma(\frac{5}{6})} + \frac{2\Gamma(\frac{5}{6})}{\Gamma(\frac{1}{3})}\right], \qquad (4)\end{aligned}$$

and

$$E'(k_3) = \frac{\pi\sqrt{3}}{4}\frac{1}{K'(k_3)} + \frac{\sqrt{3}-1}{2\sqrt{3}}K'(k_3). \qquad (5)$$

Summarizing,

$$\begin{aligned}K[\tfrac{1}{4}(\sqrt{6}-\sqrt{2})] &= \frac{\sqrt{\pi}}{2\cdot 3^{3/4}}\frac{\Gamma(\frac{1}{6})}{\Gamma(\frac{2}{3})} \qquad (6)\\ K'[\tfrac{1}{4}(\sqrt{6}-\sqrt{2})] &= \sqrt{3}K = \frac{\sqrt{\pi}}{2\cdot 3^{1/4}}\frac{\Gamma(\frac{1}{6})}{\Gamma(\frac{2}{3})} \qquad (7)\\ E[\tfrac{1}{4}(\sqrt{6}-\sqrt{2})]&\\ &= \frac{1}{4}\left(\frac{\pi}{\sqrt{3}}\right)^{1/2}\left[\left(1+\frac{1}{\sqrt{3}}\right)\frac{\Gamma(\frac{1}{3})}{\Gamma(\frac{5}{6})} + \frac{2\Gamma(\frac{5}{6})}{\Gamma(\frac{1}{3})}\right] \qquad (8)\\ E'[\tfrac{1}{4}(\sqrt{6}-\sqrt{2})]&\\ &= \frac{\sqrt{\pi}}{2}\left[3^{3/4}\frac{\Gamma(\frac{2}{3})}{\Gamma(\frac{1}{6})} + \frac{\sqrt{3}-1}{2\cdot 3^{3/4}}\frac{\Gamma(\frac{1}{6})}{\Gamma(\frac{2}{3})}\right]. \qquad (9)\end{aligned}$$

(Whittaker and Watson 1990).

see also THETA FUNCTION

References

Ramanujan, S. "Modular Equations and Approximations to π." *Quart. J. Pure. Appl. Math.* **45**, 350–372, 1913–1914.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, pp. 525–527 and 535, 1990.

Elliptic Lambda Function

The λ GROUP is the SUBGROUP of the GAMMA GROUP with a and d ODD; b and c EVEN. The function

$$\lambda(t) \equiv \lambda(q) \equiv k^2(q) = \left[\frac{\vartheta_2(q)}{\vartheta_3(q)}\right]^4, \qquad (1)$$

where

$$q \equiv e^{i\pi t} \qquad (2)$$

is a λ-MODULAR FUNCTION and ϑ_i are THETA FUNCTIONS.

$\lambda^*(r)$ gives the value of the MODULUS k_r for which the complementary and normal complete ELLIPTIC INTEGRALS OF THE FIRST KIND are related by

$$\frac{K'(k_r)}{K(k_r)} = \sqrt{r}. \qquad (3)$$

It can be computed from

$$\lambda^*(r) \equiv k(q) = \frac{{\vartheta_2}^2(q)}{{\vartheta_3}^2(q)}, \qquad (4)$$

where

$$q \equiv e^{-\pi\sqrt{r}}, \qquad (5)$$

and ϑ_i is a THETA FUNCTION.

From the definition of the lambda function,

$$\lambda^*(r') = \lambda^*\left(\frac{1}{r}\right) = \lambda^{*\prime}(r). \qquad (6)$$

For all rational r, $K(\lambda^*(r))$ and $E(\lambda^*(r))$ are expressible in terms of a finite number of GAMMA FUNCTIONS (Selberg and Chowla 1967). $\lambda^*(r)$ is related to the RAMANUJAN g- AND G-FUNCTIONS by

$$\lambda^*(n) = \tfrac{1}{2}\left(\sqrt{1+G_n^{-12}} - \sqrt{1-G_n^{-12}}\right) \qquad (7)$$

$$\lambda^*(n) = g_n^6\left(\sqrt{g_n^{12}+g_n^{-12}} - g_n^6\right). \qquad (8)$$

Special values are

$$\lambda^*(\tfrac{2}{29}) = (13\sqrt{58}-99)(\sqrt{2}+1)^6$$
$$\lambda^*(\tfrac{2}{5}) = (\sqrt{10}-3)(\sqrt{2}+1)^2$$
$$\lambda^*(\tfrac{2}{3}) = (2-\sqrt{3})(\sqrt{2}+\sqrt{3})$$
$$\lambda^*(\tfrac{3}{4}) = (\sqrt{3}-\sqrt{2})^2(\sqrt{2}+1)^2$$
$$\lambda^*(1) = \frac{1}{\sqrt{2}}$$
$$\lambda^*(2) = \sqrt{2}-1$$
$$\lambda^*(3) = \tfrac{1}{4}\sqrt{2}(\sqrt{3}-1)$$
$$\lambda^*(4) = 3-2\sqrt{2}$$
$$\lambda^*(5) = \tfrac{1}{2}\left(\sqrt{\sqrt{5}-1} - \sqrt{3-\sqrt{5}}\right)$$
$$\lambda^*(6) = (2-\sqrt{3})(\sqrt{3}-\sqrt{2})$$
$$\lambda^*(7) = \tfrac{1}{8}\sqrt{2}(3-\sqrt{7})$$
$$\lambda^*(8) = \left(\sqrt{2}+1-\sqrt{2\sqrt{2}+2}\right)^2$$
$$\lambda^*(9) = \tfrac{1}{2}(\sqrt{2}-3^{1/4})(\sqrt{3}-1)$$
$$\lambda^*(10) = (\sqrt{10}-3)(\sqrt{2}-1)^2$$
$$\lambda^*(11) = \tfrac{1}{12}\sqrt{6}\left(\sqrt{1+2x_{11}-4{x_{11}}^{-1}}\right.$$
$$\left.-\sqrt{11+2x_{11}-4{x_{11}}^{-1}}\right)$$
$$\lambda^*(12) = (\sqrt{3}-\sqrt{2})^2(\sqrt{2}-1)^2$$
$$= 15-10\sqrt{2}+8\sqrt{3}-6\sqrt{6}$$
$$\lambda^*(13) = \tfrac{1}{2}\left(\sqrt{5\sqrt{13}-17} - \sqrt{19-5\sqrt{13}}\right)$$
$$\lambda^*(14) = -11-8\sqrt{2}-2(\sqrt{2}+2)\sqrt{5+4\sqrt{2}}$$
$$+\sqrt{11+8\sqrt{2}}(2+2\sqrt{2}+\sqrt{2}\sqrt{5+4\sqrt{2}})$$
$$\lambda^*(15) = \tfrac{1}{16}\sqrt{2}(3-\sqrt{5})(\sqrt{5}-\sqrt{3})(2-\sqrt{3})$$
$$\lambda^*(16) = \frac{(2^{1/4}-1)^2}{(2^{1/4}+1)^2}$$
$$\lambda^*(17) = \tfrac{1}{4}\sqrt{2}(\sqrt{42+10\sqrt{17}}$$
$$-13\sqrt{-3+\sqrt{17}}\sqrt{5+\sqrt{17}}$$
$$-3\sqrt{17}\sqrt{-3+\sqrt{17}}\sqrt{5+\sqrt{17}}$$
$$-\sqrt{-38-10\sqrt{17}+13\sqrt{-3+\sqrt{17}}}\sqrt{5+\sqrt{17}}$$
$$+3\sqrt{17}\sqrt{-3+\sqrt{17}}\sqrt{5+\sqrt{17}})$$
$$\lambda^*(18) = (\sqrt{2}-1)^3(2-\sqrt{3})^2$$
$$\lambda^*(22) = (3\sqrt{11}-7\sqrt{2})(10-3\sqrt{11})$$
$$\lambda^*(30) = (\sqrt{3}-\sqrt{2})^2(2-\sqrt{3})(\sqrt{6}-\sqrt{5})(4-\sqrt{15})$$
$$\lambda^*(34) = (\sqrt{2}-1)^2(3\sqrt{2}-\sqrt{17})$$
$$\times(\sqrt{297+72\sqrt{17}} - \sqrt{296+72\sqrt{17}})$$
$$\lambda^*(42) = (\sqrt{2}-1)^2(2-\sqrt{3})^2(\sqrt{7}-\sqrt{6})(8-3\sqrt{7})$$
$$\lambda^*(58) = (13\sqrt{58}-99)(\sqrt{2}-1)^6$$
$$\lambda^*(210) = (\sqrt{2}-1)^2(2-\sqrt{3})(\sqrt{7}-\sqrt{6})^2(8-3\sqrt{7})$$
$$\times(\sqrt{10}-3)^2(4-\sqrt{15})^2(\sqrt{15}-\sqrt{14})(6-\sqrt{35}),$$

where

$$x_{11} \equiv (17+3\sqrt{33})^{1/3}.$$

In addition,

$$\lambda^*(1') = \frac{1}{\sqrt{2}}$$
$$\lambda^*(2') = \sqrt{2\sqrt{2}-2}$$
$$\lambda^*(3') = \tfrac{1}{4}\sqrt{2}(\sqrt{3}+1)$$
$$\lambda^*(4') = 2^{1/4}(2\sqrt{2}-2)$$
$$\lambda^*(5') = \tfrac{1}{2}\left(\sqrt{\sqrt{5}-1} + \sqrt{3-\sqrt{5}}\right)$$
$$\lambda^*(7') = \tfrac{1}{8}\sqrt{2}\,(3+\sqrt{7})$$
$$\lambda^*(9') = \tfrac{1}{2}(\sqrt{2}+3^{1/4})(\sqrt{3}-1)$$
$$\lambda^*(12') = 2\sqrt{-208+147\sqrt{2}-120\sqrt{3}+85\sqrt{6}}.$$

see also ELLIPTIC ALPHA FUNCTION, ELLIPTIC INTEGRAL OF THE FIRST KIND, MODULUS (ELLIPTIC INTEGRAL), RAMANUJAN g- AND G-FUNCTIONS, THETA FUNCTION

References

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, pp. 139 and 298, 1987.

Bowman, F. *Introduction to Elliptic Functions, with Applications.* New York: Dover, pp. 75, 95, and 98, 1961.

Selberg, A. and Chowla, S. "On Epstein's Zeta-Function." *J. Reine. Angew. Math.* **227**, 86–110, 1967.

Watson, G. N. "Some Singular Moduli (1)." *Quart. J. Math.* **3**, 81–98, 1932.

Elliptic Logarithm

A generalization of integrals of the form

$$\int_\infty^x \frac{dt}{\sqrt{t^2+at}},$$

which can be expressed in terms of logarithmic and inverse trigonometric functions to

$$\operatorname{eln}(x) \equiv \int_x^\infty \frac{dt}{\sqrt{t^3+at^2+bt}}.$$

The inverse of the elliptic logarithm is the ELLIPTIC EXPONENTIAL FUNCTION.

Elliptic Modular Function

$$\varphi(z) = \left[\frac{\vartheta_2{}^4(0,z)}{\vartheta_3{}^4(0,z)}\right]^{1/8},$$

where ϑ is a THETA FUNCTION. A special case is

$$\varphi(-e^{-\pi\sqrt{3}}) = (4\sqrt{3}-7)^{1/8}.$$

see also MODULAR FUNCTION

Elliptic Paraboloid

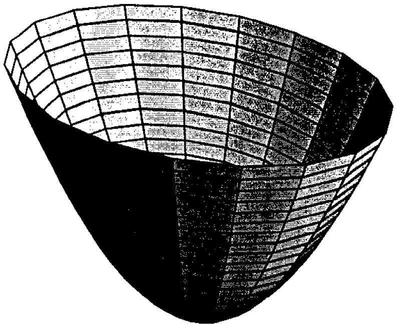

A QUADRATIC SURFACE which has ELLIPTICAL CROSS-SECTION. The elliptic paraboloid of height h, SEMIMAJOR AXIS a, and SEMIMINOR AXIS b can be specified parametrically by

$$x = a\sqrt{u}\cos v$$
$$y = b\sqrt{u}\sin v$$
$$z = u.$$

for $v \in [0, 2\pi)$ and $u \in [0, h]$.

see also ELLIPTIC CONE, ELLIPTIC CYLINDER, PARABOLOID

References

Fischer, G. (Ed.). Plate 66 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 61, 1986.

Elliptic Partial Differential Equation

A second-order PARTIAL DIFFERENTIAL EQUATION, i.e., one of the form

$$Au_{xx} + 2Bu_{xy} + Cu_{yy} + Du_x + Eu_y + F = 0, \tag{1}$$

is called elliptic if the MATRIX

$$\mathsf{Z} \equiv \begin{bmatrix} A & B \\ B & C \end{bmatrix} \tag{2}$$

is POSITIVE DEFINITE. LAPLACE'S EQUATION and POISSON'S EQUATION are examples of elliptic partial differential equations. For an elliptic partial differential equation, BOUNDARY CONDITIONS are used to give the constraint $u(x,y) = g(x,y)$ on $\partial\Omega$, where

$$u_{xx} + u_{yy} = f(u_x, u_y, u, x, y) \tag{3}$$

holds in Ω.

see also HYPERBOLIC PARTIAL DIFFERENTIAL EQUATION, PARABOLIC PARTIAL DIFFERENTIAL EQUATION, PARTIAL DIFFERENTIAL EQUATION

Elliptic Plane

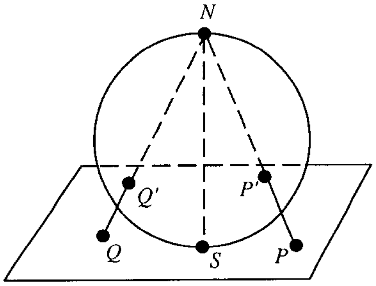

The REAL PROJECTIVE PLANE with elliptic METRIC where the distance between two points P and Q is defined as the RADIAN ANGLE between the projection of the points on the surface of a SPHERE (which is tangent to the plane at a point S) from the ANTIPODE N of the tangent point.

References

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, p. 94, 1969.

Elliptic Point

A point $\mathbf{p}$ on a Regular Surface $M \in \mathbb{R}^3$ is said to be elliptic if the Gaussian Curvature $K(\mathbf{p}) > 0$ or equivalently, the Principal Curvatures κ_1 and κ_2 have the same sign.

see also Anticlastic, Elliptic Fixed Point (Differential Equations), Elliptic Fixed Point (Map), Gaussian Curvature, Hyperbolic Point, Parabolic Point, Planar Point, Synclastic

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 280, 1993.

Elliptical Projection

see Mollweide Projection

Elliptic Pseudoprime

Let E be an Elliptic Curve defined over the Field of Rational Numbers $\mathbb{Q}(\sqrt{-d})$ having equation

$$y^2 = x^3 + ax + b$$

with a and b Integers. Let P be a point on E with integer coordinates and having infinite order in the additive group of rational points of E, and let n be a Composite Natural Number such that $(-d/n) = -1$, where $(-d/n)$ is the Jacobi Symbol. Then if

$$(n+1)P \equiv 0 \pmod{n},$$

n is called an elliptic pseudoprime for (E, P).

see also Atkin-Goldwasser-Kilian-Morain Certificate, Elliptic Curve Primality Proving, Strong Elliptic Pseudoprime

References

Balasubramanian, R. and Murty, M. R. "Elliptic Pseudoprimes. II." Submitted.

Gordon, D. M. "The Number of Elliptic Pseudoprimes." *Math. Comput.* **52**, 231–245, 1989.

Gordon, D. M. "Pseudoprimes on Elliptic Curves." In *Théorie des nombres* (Ed. J. M. DeKoninck and C. Levesque). Berlin: de Gruyter, pp. 290–305, 1989.

Miyamoto, I. and Murty, M. R. "Elliptic Pseudoprimes." *Math. Comput.* **53**, 415–430, 1989.

Ribenboim, P. *The New Book of Prime Number Records, 3rd ed.* New York: Springer-Verlag, pp. 132–134, 1996.

Elliptic Rotation

Leaves the Circle

$$x^2 + y^2 = 1$$

invariant.

$$x' = x\cos\theta - y\sin\theta$$
$$y' = x\sin\theta + y\sin\theta.$$

see also Equiaffinity

Elliptic Theta Function

see Neville Theta Function, Theta Function

Elliptic Torus

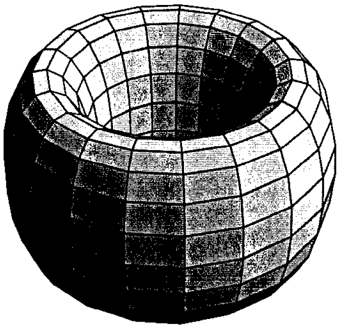

A generalization of the ring Torus produced by stretching or compressing in the z direction. It is given by the parametric equations

$$x(u,v) = (a + b\cos v)\cos u$$
$$y(u,v) = (a + b\cos v)\sin u$$
$$z(u,v) = c\sin v.$$

see also Torus

References

Gray, A. "Tori." §11.4 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 218–220, 1993.

Elliptic Umbilic Catastrophe

A Catastrophe which can occur for three control factors and two behavior axes.

see also Hyperbolic Umbilic Catastrophe

Ellipticity

Given a Spheroid with equatorial radius a and polar radius c,

$$e \equiv \begin{cases} \sqrt{\frac{a^2-c^2}{a^2}} & a > c \text{ (oblate spheroid)} \\ \sqrt{\frac{c^2-a^2}{a^2}}. & a < c \text{ (prolate spheroid)} \end{cases}$$

see also Flattening, Oblate Spheroid, Prolate Spheroid, Spheroid

Ellison–Mendès-France Constant

$$\sum_{n \le x} \frac{1}{n} \ln\left(\frac{x}{n}\right) = \tfrac{1}{2}(\ln x)^2 + \gamma \ln x + D + \mathcal{O}(x^{-1}),$$

where γ is the Euler-Mascheroni Constant, and

$$D = 2.723\ldots$$

is the Ellision-Mendès-France constant.

References

Ellison, W. J. and Mendès-France, M. *Les nombres premiers.* Paris: Hermann, 1975.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 47, 1983.

Elongated Cupola

A n-gonal CUPOLA adjoined to a $2n$-gonal PRISM.

see also ELONGATED PENTAGONAL CUPOLA, ELONGATED SQUARE CUPOLA, ELONGATED TRIANGULAR CUPOLA

Elongated Dipyramid

see also ELONGATED PENTAGONAL DIPYRAMID, ELONGATED SQUARE DIPYRAMID, ELONGATED TRIANGULAR DIPYRAMID

Elongated Dodecahedron

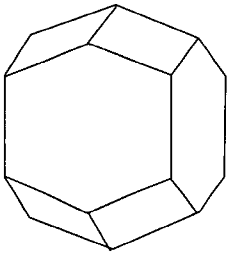

A SPACE-FILLING POLYHEDRON and PARALLELOHEDRON.

References

Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, pp. 29–30 and 257, 1973.

Elongated Gyrobicupola

see ELONGATED PENTAGONAL GYROBICUPOLA, ELONGATED SQUARE GYROBICUPOLA, ELONGATED TRIANGULAR GYROBICUPOLA

Elongated Gyrocupolarotunda

see ELONGATED PENTAGONAL GYROCUPOLAROTUNDA

Elongated Orthobicupola

see ELONGATED PENTAGONAL ORTHOBICUPOLA, ELONGATED TRIANGULAR ORTHOBICUPOLA

Elongated Orthobirotunda

see ELONGATED PENTAGONAL ORTHOBIROTUNDA

Elongated Orthocupolarotunda

see ELONGATED PENTAGONAL ORTHOCUPOLAROTUNDA

Elongated Pentagonal Cupola

see JOHNSON SOLID

Elongated Pentagonal Dipyramid

see JOHNSON SOLID

Elongated Pentagonal Gyrobicupola

see JOHNSON SOLID

Elongated Pentagonal Gyrobirotunda

see JOHNSON SOLID

Elongated Pentagonal Gyrocupolarotunda

see JOHNSON SOLID

Elongated Pentagonal Orthobicupola

see JOHNSON SOLID

Elongated Pentagonal Orthobirotunda

see JOHNSON SOLID

Elongated Pentagonal Orthocupolarotunda

see JOHNSON SOLID

Elongated Pentagonal Pyramid

see JOHNSON SOLID

Elongated Pentagonal Rotunda

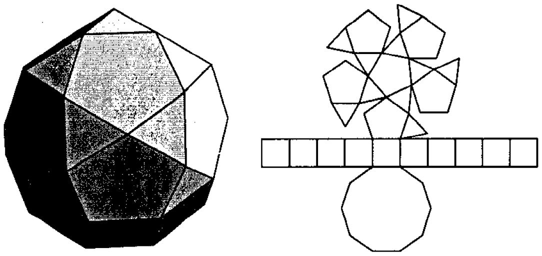

A PENTAGONAL ROTUNDA adjoined to a decagonal PRISM which is JOHNSON SOLID J_{21}.

Elongated Pyramid

An n-gonal PYRAMID adjoined to an n-gonal PRISM.

see also ELONGATED PENTAGONAL PYRAMID, ELONGATED SQUARE PYRAMID, ELONGATED TRIANGULAR PYRAMID, GYROELONGATED PYRAMID

Elongated Rotunda

see ELONGATED PENTAGONAL ROTUNDA

Elongated Square Cupola

see JOHNSON SOLID

Elongated Square Dipyramid

see JOHNSON SOLID

Elongated Square Gyrobicupola

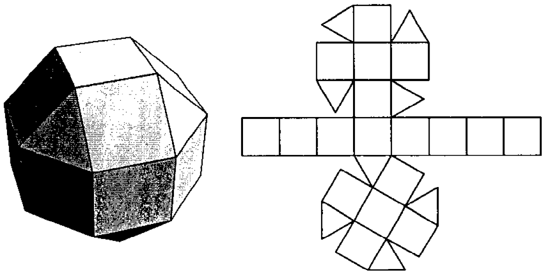

A nonuniform POLYHEDRON obtained by rotating the bottom third of a SMALL RHOMBICUBOCTAHEDRON (Ball and Coxeter 1987, p. 137). It is also called MILLER'S SOLID, the MILLER-AŠKINUZE SOLID, or the PSEUDORHOMBICUBOCTAHEDRON, and is JOHNSON SOLID J_{37}.

see also SMALL RHOMBICUBOCTAHEDRON

References

Aškinuze, V. G. "O čisle polupravil'nyh mnogogrannikov." *Math. Prosvešč.* **1**, 107–118, 1957.

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 137–138, 1987.

Cromwell, P. R. *Polyhedra.* New York: Cambridge University Press, pp. 91–92, 1997.

Elongated Square Pyramid

see JOHNSON SOLID

Elongated Triangular Cupola

see JOHNSON SOLID

Elongated Triangular Dipyramid

see JOHNSON SOLID

Elongated Triangular Gyrobicupola

see JOHNSON SOLID

Elongated Triangular Orthobicupola

see JOHNSON SOLID

Elongated Triangular Pyramid

see JOHNSON SOLID

Elsasser Function

$$E(y,u) \equiv \int_{-1/2}^{1/2} \exp\left[-\frac{2\pi y u \sinh(2\pi y)}{\cosh(2\pi y) - \cos(2\pi x)}\right] dx.$$

Embeddable Knot

A KNOT K is an n-embeddable knot if it can be placed on a GENUS n standard embedded surface without crossings, but K cannot be placed on any standardly embedded surface of lower GENUS without crossings. Any KNOT is an n-embeddable knot for some n. The FIGURE-OF-EIGHT KNOT is a 2-EMBEDDABLE KNOT. A knot with BRIDGE NUMBER b is an n-embeddable knot where $n \leq b$.

see also TUNNEL NUMBER

Embedding

see EXTRINSIC CURVATURE, HYPERBOLOID EMBEDDING, INJECTION, SPHERE EMBEDDING

Empty Set

The SET containing no elements, denoted $\varnothing$. Strangely, the empty set is both OPEN and CLOSED for any SET X and TOPOLOGY. A GROUPOID, SEMIGROUP, QUASIGROUP, RINGOID, and SEMIRING can be empty. A MONOID, GROUP, and RINGS must have at least one element, while DIVISION RINGS and FIELDS must have at least two elements.

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 266, 1996.

Enantiomer

Two objects which are MIRROR IMAGES of each other are called enantiomers. The term enantiomer is synonymous with ENANTIOMORPH.

see also AMPHICHIRAL KNOT, CHIRAL, DISYMMETRIC, HANDEDNESS, MIRROR IMAGE, REFLEXIBLE

References

Ball, W. W. R. and Coxeter, H. S. M. "Polyhedra." Ch. 5 in *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 130–161, 1987.

Enantiomorph

see ENANTIOMER

Encoding

An encoding is a way of representing a number or expression in terms of another (usually simpler) one. However, multiple expressions can also be encoded as a single expression, as in, for example,

$$(a,b) \equiv \tfrac{1}{2}[(a+b)^2 + 3a + b]$$

which encodes a and b uniquely as a single number.

a	b	(a,b)
0	0	0
0	1	1
1	0	2
0	2	3
1	2	4
2	0	5

see also CODE, CODING THEORY

Endogenous Variable

An economic variable which is independent of the relationships determining the equilibrium levels, but nonetheless affects the equilibrium.

see also EXOGENOUS VARIABLE

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 458, 1980.

Endomorphism

A SURJECTIVE MORPHISM from an object to itself. In ERGODIC THEORY, let X be a SET, F a SIGMA ALGEBRA on X and m a PROBABILITY MEASURE. A MAP $T : X \to X$ is called an endomorphism or MEASURE-PRESERVING TRANSFORMATION if

1. T is SURJECTIVE,
2. T is MEASURABLE,
3. $m(T^{-1}A) = m(A)$ for all $A \in F$.

An endomorphism is called ERGODIC if it is true that $T^{-1}A = A$ IMPLIES $m(A) = 0$ or 1, where $T^{-1}A = \{x \in X : T(x) \in A\}$.

see also MEASURABLE FUNCTION, MEASURE-PRESERVING TRANSFORMATION, MORPHISM, SIGMA ALGEBRA, SURJECTIVE

Endraß Octic

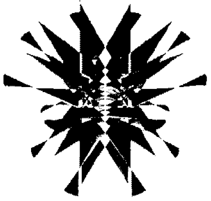

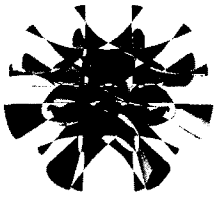

Endraß surfaces are a pair of OCTIC SURFACES which have 168 ORDINARY DOUBLE POINTS. This is the maximum number known to exist for an OCTIC SURFACE, although the rigorous upper bound is 174. The equations of the surfaces $X_8^{\pm}$ are

$$\begin{aligned}64(x^2 - w^2)(y^2 - w^2)[(x+y)^2 - 2w^2]\\ [(x-y)^2 - 2w^2] - \{-4(1 \pm \sqrt{2})(x^2+y^2)^2\\ +[8(2 \pm \sqrt{2})z^2 + 2(2 \pm 7\sqrt{2})w^2](x^2+y^2)\\ -16z^4 + 8(1 \mp 2\sqrt{2})z^2w^2 - (1+12\sqrt{2})w^4\}^2 = 0,\end{aligned}$$

where w is a parameter taken as $w = 1$ in the above plots. All ORDINARY DOUBLE POINTS of X_8^+ are real, while 24 of those in X_8^- are complex. The surfaces were discovered in a 5-D family of octics with 112 nodes, and are invariant under the GROUP $D_8 \otimes Z_2$.

see also OCTIC SURFACE

References

Endraß, S. "Octics with 168 Nodes." `http:// www . mathematik.uni-mainz.de/AlgebraischeGeometrie/docs /Eendrassoctic.shtml.`

Endraß, S. "Flächen mit vielen Doppelpunkten." *DMV-Mitteilungen* 4, 17–20, 4/1995.

Endraß, S. "A Proctive Surface of Degree Eight with 168 Nodes." *J. Algebraic Geom.* **6**, 325–334, 1997.

Energy

The term energy has an important physical meaning in physics and is an extremely useful concept. A much more abstract mathematical generalization is defined as follows. Let Ω be a SPACE with MEASURE $\mu \geq 0$ and let $\Phi(P, Q)$ be a real function on the PRODUCT SPACE $\Omega \times \Omega$. When

$$\begin{aligned}(\mu, nu) &= \iint \Phi(P,Q)\,d\mu(Q)\,d\nu(P)\\ &= \int \Phi(P,\mu)\,d\nu(P)\end{aligned}$$

exists for measures $\mu, \nu \geq 0$, (μ, ν) is called the MUTUAL ENERGY and (μ, μ) is called the ENERGY.

see also DIRICHLET ENERGY, MUTUAL ENERGY

References

Iyanaga, S. and Kawada, Y. (Eds.). "General Potential." §335.B in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1038, 1980.

Engel's Theorem

A finite-dimensional LIE ALGEBRA all of whose elements are ad-NILPOTENT is itself a NILPOTENT LIE ALGEBRA.

Enneacontagon

A 90-sided POLYGON.

Enneacontahedron

A ZONOHEDRON constructed from the 10 diameters of the DODECAHEDRON which has 90 faces, 30 of which are RHOMBS of one type and the other 60 of which are RHOMBS of another. The enneacontahedron somewhat resembles a figure of Sharp.

see also DODECAHEDRON, RHOMB, ZONOHEDRON

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 142–143, 1987.

Sharp, A. *Geometry Improv'd.* London, p. 87, 1717.

Enneadecagon

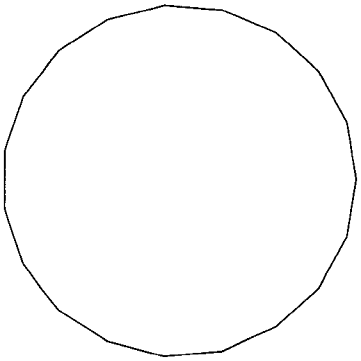

A 19-sided POLYGON, sometimes also called the ENNEAKAIDECAGON.

Enneagon

see NONAGON

Enneagonal Number

see NONAGONAL NUMBER

Enneakaidecagon

see ENNEADECAGON

Enneper's Surfaces

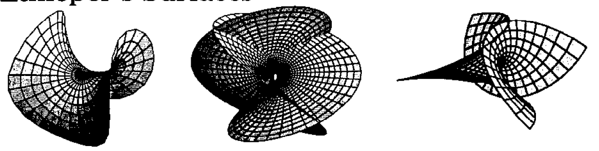

The Enneper surfaces are a three-parameter family of surfaces with constant curvature. In general, they are described by elliptic functions. However, special cases which can be specified parametrically using ELEMENTARY FUNCTION include the KUEN SURFACE, REMBS' SURFACES, and SIEVERT'S SURFACE. The surfaces shown above can be generated using the ENNEPER-WEIERSTRAß PARAMETERIZATION with

$$f(\zeta) = 1 \tag{1}$$

$$g(\zeta) = \zeta. \tag{2}$$

Letting $z = re^{i\phi}$ and taking the REAL PART give

$$x = \Re[re^{i\phi} - \tfrac{1}{3}r^3 e^{3i\phi}] \tag{3}$$

$$y = \Re[ire^{i\phi} + \tfrac{1}{3}ir^3 e^{3i\phi}] \tag{4}$$

$$z = \Re[r^2 e^{2i\phi}], \tag{5}$$

where $r \in [0, 1]$ and $\phi \in [-\pi, \pi)$. Letting $z = u + iv$ instead gives the figure on the right,

$$x = u - \tfrac{1}{3}u^3 + uv^2 \tag{6}$$

$$y = -v - u^2 v + \tfrac{1}{3}v^3 \tag{7}$$

$$z = u^2 - v^2 \tag{8}$$

(do Carmo 1986, Gray 1993, Nordstrand). This surface has a HOLE in its middle. Nordstrand gives the implicit form

$$\left(\frac{y^2 - x^2}{2z} + \tfrac{2}{9}z^2 + \tfrac{2}{3}\right)^3 - 6\left[\frac{(y^2 - x^2)}{4z} - \tfrac{1}{4}(x^2 + y^2 + \tfrac{8}{9}z^2) + \tfrac{2}{9}\right]^2 = 0. \tag{9}$$

References

Dickson, S. "Minimal Surfaces." *Mathematica J.* **1**, 38–40, 1990.

do Carmo, M. P. "Enneper's Surface." §3.5C in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, p. 43, 1986.

Enneper, A. "Analytisch-geometrische Untersuchungen." *Nachr. Königl. Gesell. Wissensch. Georg-Augustus-Univ. Göttingen* **12**, 258–277, 1868.

Fischer, G. (Ed.). Plate 92 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 88, 1986.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 265, 1993.

Maeder, R. *The Mathematica Programmer.* San Diego, CA: Academic Press, pp. 150–151, 1994.

Nordstrand, T. "Enneper's Minimal Surface." `http://www.uib.no/people/nfytn/enntxt.htm`.

Reckziegel, H. "Enneper's Surfaces." §3.4.4 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 37–39, 1986.

Wolfram Research "Mathematica Version 2.0 Graphics Gallery." `http://www.mathsource.com/cgi-bin/MathSource/Applications/Graphics/3D/0207-155`.

Enneper-Weierstraß Parameterization

Gives a parameterization of a MINIMAL SURFACE.

$$\Re \int \begin{bmatrix} f(1 - g^2) \\ if(1 + g^2) \\ 2fg \end{bmatrix} d\zeta.$$

see also MINIMAL SURFACE

References

Dickson, S. "Minimal Surfaces." *Mathematica J.* **1**, 38–40, 1990.

do Carmo, M. P. *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, p. 41, 1986.

Weierstraß, K. "Über die Flächen deren mittlere Krümmung überall gleich null ist." *Monatsber. Berliner Akad.*, 612–625, 1866.

Enormous Theorem

see CLASSIFICATION THEOREM

Enriques Surfaces

An Enriques surface X is a smooth compact complex surface having irregularity $q(X) = 0$ and nontrivial canonical sheaf K_X such that $K_X^2 = O_X$ (Endraß). Such surfaces cannot be embedded in projective 3-space, but there nonetheless exist transformations onto singular surfaces in projective 3-space. There exists a family of such transformed surfaces of degree six which passes through each edge of a TETRAHEDRON twice. A subfamily with tetrahedral symmetry is given by the two-parameter (r, c) family of surfaces

$$f_r x_0 x_1 x_2 x_3 + c({x_0}^2 {x_1}^2 {x_2}^2 + {x_0}^2 {x_1}^2 {x_3}^2 + {x_0}^2 {x_2}^2 {x_3}^2 + {x_1}^2 {x_2}^2 {x_3}^2 = 0$$

and the polynomial f_r is a sphere with radius r,

$$f_r = (3 - r)({x_0}^2 + {x_1}^2 + {x_2}^2 + {x_3}^2) - 2(1 + r)(x_0 x_1 + x_0 x_2 + x_0 x_3 + x_1 x_2 + x_1 x_3 + x_2 x_3)$$

(Endraß).

References
Angermüller, G. and Barth, W. "Elliptic Fibres on Enriques Surfaces." *Compos. Math.* **47**, 317–332, 1982.
Barth, W. and Peters, C. "Automorphisms of Enriques Surfaces." *Invent. Math.* **73**, 383–411, 1983.
Barth, W. P.; Peters, C. A.; and van de Ven, A. A. *Compact Complex Surfaces.* New York: Springer-Verlag, 1984.
Barth, W. "Lectures on K3- and Enriques Surfaces." In *Algebraic Geometry, Sitges (Barcelona) 1983, Proceedings of a Conference Held in Sitges (Barcelona), Spain, October 5–12, 1983* (Ed. E. Casas-Alvero, G. E. Welters, and S. Xambó-Descamps). New York: Springer-Verlag, pp. 21–57, 1983.
Endraß, S. "Enriques Surfaces." `http://www.mathematik.uni-mainz.de/Algebraische Geometrie/docs/enriques.shtml`.
Enriques, F. *Le superficie algebriche.* Bologna, Italy: Zanichelli, 1949.
Enriques, F. "Sulla classificazione." *Atti Accad. Naz. Lincei* **5**, 1914.
Hunt, B. *The Geometry of Some Special Arithmetic Quotients.* New York: Springer-Verlag, p. 317, 1996.

Entire Function

If a function is ANALYTIC on $\mathbb{C}^*$, where $\mathbb{C}^*$ denotes the extended COMPLEX PLANE, then it is said to be entire.

see also ANALYTIC FUNCTION, HOLOMORPHIC FUNCTION, MEROMORPHIC

Entringer Number

The Entringer numbers $E(n,k)$ are the number of PERMUTATIONS of $\{1,2,\ldots,n+1\}$, starting with $k+1$, which, after initially falling, alternately fall then rise. The Entringer numbers are given by

$$E(0,0) = 1$$
$$E(n,0) = 0$$

together with the RECURRENCE RELATION

$$E(n,k) = E(n,k+1) + E(n-1,n-k).$$

The numbers $E(n) = E(n,n)$ are the SECANT and TANGENT NUMBERS given by the MACLAURIN SERIES

$$\sec x + \tan x = A_0 + A_1 x + A_2\frac{x^2}{2!} + A_3\frac{x^3}{3!} + A_4\frac{x^4}{4!} + A_5\frac{x^5}{5!} + \ldots.$$

see also ALTERNATING PERMUTATION, BOUSTROPHEDON TRANSFORM, EULER ZIGZAG NUMBER, PERMUTATION, SECANT NUMBER, SEIDEL-ENTRINGER-ARNOLD TRIANGLE, TANGENT NUMBER, ZAG NUMBER, ZIG NUMBER

References
Entringer, R. C. "A Combinatorial Interpretation of the Euler and Bernoulli Numbers." *Nieuw. Arch. Wisk.* **14**, 241–246, 1966.
Millar, J.; Sloane, N. J. A.; and Young, N. E. "A New Operation on Sequences: The Boustrophedon Transform." *J. Combin. Th. Ser. A* **76**, 44–54, 1996.
Poupard, C. "De nouvelles significations enumeratives des nombres d'Entringer." *Disc. Math.* **38**, 265–271, 1982.

Entropy

In physics, the word entropy has important physical implications as the amount of "disorder" of a system. In mathematics, a more abstract definition is used. The (Shannon) entropy of a variable X is defined as

$$H(X) \equiv -\sum_x p(x)\ln[p(x)],$$

where $p(x)$ is the probability that X is in the state x, and $p\ln p$ is defined as 0 if $p=0$. The joint entropy of variables $X_1, \ldots, X_n$ is then defined by

$$H(X_1,\ldots,X_n) \equiv -\sum_{x_1}\cdots\sum_{x_n} p(x_1,\ldots,x_n)\ln[p(x_1,\ldots,x_n)].$$

see also KOLMOGOROV ENTROPY, KOLMOGOROV-SINAI ENTROPY, MAXIMUM ENTROPY METHOD, METRIC ENTROPY, ORNSTEIN'S THEOREM, REDUNDANCY, SHANNON ENTROPY, TOPOLOGICAL ENTROPY

References
Ott, E. "Entropies." §4.5 in *Chaos in Dynamical Systems.* New York: Cambridge University Press, pp. 138–144, 1993.

Entscheidungsproblem

see DECISION PROBLEM

Enumerative Geometry

Schubert's application of the CONSERVATION OF NUMBER PRINCIPLE.

see also CONSERVATION OF NUMBER PRINCIPLE, DUALITY PRINCIPLE, HILBERT'S PROBLEMS, PERMANENCE OF MATHEMATICAL RELATIONS PRINCIPLE

References
Bell, E. T. *The Development of Mathematics, 2nd ed.* New York: McGraw-Hill, p. 340, 1945.

Envelope

The envelope of a one-parameter family of curves given implicitly by

$$U(x,y,c) = 0, \tag{1}$$

or in parametric form by $(f(t,c), g(t,c))$, is a curve which touches every member of the family. For a curve represented by $(f(t,c), g(t,c))$, the envelope is found by solving

$$0 = \frac{\partial f}{\partial t}\frac{\partial g}{\partial c} - \frac{\partial f}{\partial c}\frac{\partial g}{\partial t}. \tag{2}$$

For a curve represented implicitly, the envelope is given by simultaneously solving

$$\frac{\partial U}{\partial c} = 0 \tag{3}$$

$$U(x,y,c) = 0. \tag{4}$$

see also ASTROID, CARDIOID, CATACAUSTIC, CAUSTIC, CAYLEYIAN CURVE, DÜRER'S CONCHOID, ELLIPSE ENVELOPE, ENVELOPE THEOREM, EVOLUTE, GLISSETTE, HEDGEHOG, KIEPERT'S PARABOLA, LINDELOF'S THEOREM, NEGATIVE PEDAL CURVE

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 33–34, 1972.

Lee, X. "Envelope." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Envelope_dir/envelope.html`.

Yates, R. C. "Envelopes." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 75–80, 1952.

Envelope Theorem

Relates EVOLUTES to single paths in the CALCULUS OF VARIATIONS. Proved in the general case by Darboux and Zermelo (1894) and Kneser (1898). It states: "When a single parameter family of external paths from a fixed point O has an ENVELOPE, the integral from the fixed point to any point A on the ENVELOPE equals the integral from the fixed point to any second point B on the ENVELOPE plus the integral along the envelope to the first point on the ENVELOPE, $J_{OA} = J_{OB} + J_{BA}$."

References

Kimball, W. S. *Calculus of Variations by Parallel Displacement.* London: Butterworth, p. 292, 1952.

Envyfree

An agreement in which all parties feel as if they have received the best deal.

Epicycloid

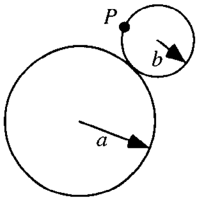

The path traced out by a point P on the EDGE of a CIRCLE of RADIUS b rolling on the outside of a CIRCLE of RADIUS a.

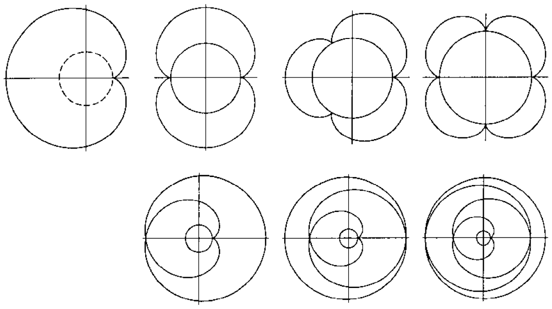

It is given by the equations

$$x = (a+b)\cos\phi - b\cos\left(\frac{a+b}{b}\phi\right) \tag{1}$$

$$y = (a+b)\sin\phi - b\sin\left(\frac{a+b}{b}\phi\right) \tag{2}$$

$$x^2 = (a+b)^2\cos^2\phi - 2b(a+b)\cos\phi\cos\left(\frac{a+b}{b}\phi\right) + b^2\cos^2\left(\frac{a+b}{b}\phi\right) \tag{3}$$

$$y^2 = (a+b)^2\sin^2\phi - 2b(a+b)\sin\phi\sin\left(\frac{a+b}{b}\phi\right) + b^2\sin^2\left(\frac{a+b}{b}\phi\right) \tag{4}$$

$$\begin{aligned} r^2 = x^2 + y^2 = (a+b)^2 + b^2 \\ - 2b(a+b)\left\{\cos\left[\left(\frac{a}{b}+1\right)\phi\right]\cos\phi \right. \\ \left. + \sin\left[\left(\frac{a}{b}+1\right)\phi\right]\sin\phi\right\}. \end{aligned} \tag{5}$$

But

$$\cos\alpha\cos\beta + \sin\alpha\sin\beta = \cos(\alpha-\beta), \tag{6}$$

so

$$\begin{aligned} r^2 &= (a+b)^2 + b^2 - 2b(a+b)\cos\left[\left(\frac{a}{b}+1\right)\phi - \phi\right] \\ &= (a+b)^2 + b^2 - 2b(a+b)\cos\left(\frac{a}{b}\phi\right). \end{aligned} \tag{7}$$

Note that ϕ is the parameter here, *not* the polar angle. The polar angle from the center is

$$\tan\theta = \frac{y}{x} = \frac{(a+b)\sin\phi - b\sin\left(\frac{a+b}{b}\phi\right)}{(a+b)\cos\phi - b\cos\left(\frac{a+b}{b}\phi\right)}. \tag{8}$$

To get n CUSPS in the epicycloid, $b = a/n$, because then n rotations of b bring the point on the edge back to its starting position.

$$\begin{aligned} r^2 &= a^2\left[\left(1+\frac{1}{n}\right)^2 + \left(\frac{1}{n}\right)^2 - 2\left(\frac{1}{n}\right)\left(1+\frac{1}{n}\right)\cos(n\phi)\right] \\ &= a^2\left[1 + \frac{2}{n} + \frac{1}{n^2} + \frac{1}{n^2} - \left(\frac{2}{n}\right)\left(\frac{n+1}{n}\right)\cos(n\phi)\right] \\ &= a^2\left[\frac{n^2+2n+2}{n^2} - \frac{2(n+1)}{n^2}\cos(n\phi)\right] \\ &= \frac{a^2}{n^2}\left[(n^2+2n+2) - 2(n+1)\cos(n\phi)\right], \end{aligned} \tag{9}$$

so

$$\begin{aligned} \tan\theta &= \frac{a\left(\frac{n+1}{n}\right)\sin\phi - \frac{a}{n}\sin[(n+1)\phi]}{a\left(\frac{n+1}{n}\right)\cos\phi - \frac{a}{n}\cos[(n+1)\phi]} \\ &= \frac{(n+1)\sin\phi - \sin[(n+1)\phi]}{(n+1)\cos\phi - \cos[(n+1)\phi]}. \end{aligned} \tag{10}$$

An epicycloid with one cusp is called a CARDIOID, one with two cusps is called a NEPHROID, and one with five cusps is called a RANUNCULOID.

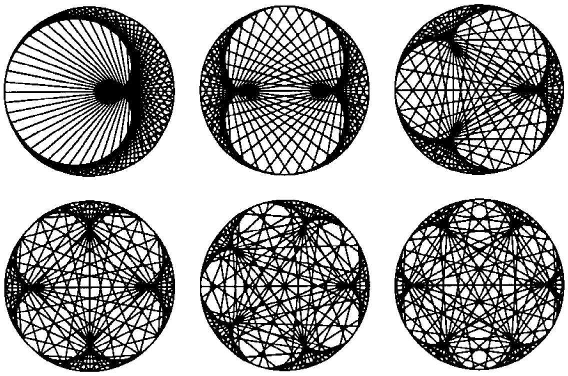

n-epicycloids can also be constructed by beginning with the DIAMETER of a CIRCLE, offsetting one end by a series of steps while at the same time offsetting the other end by steps n times as large. After traveling around the CIRCLE once, an n-cusped epicycloid is produced, as illustrated above (Madachy 1979).

Epicycloids have TORSION

$$\tau = 0 \tag{11}$$

and satisfy

$$\frac{s^2}{a^2} + \frac{\rho^2}{b^2} = 1, \tag{12}$$

where ρ is the RADIUS OF CURVATURE ($1/\kappa$).

see also CARDIOID, CYCLIDE, CYCLOID, EPICYCLOID—1-CUSPED, HYPOCYCLOID, NEPHROID, RANUNCULOID

References

Bogomolny, A. "Cycloids." `http://www.cut-the-knot.com/pythagoras/cycloids.html`.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 160–164 and 169, 1972.

Lee, X. "Epicycloid and Hypocycloid." `http://www.best.com/~xah/SpecialPlaneCurves_dir/EpiHypocycloid_dir/epiHypocycloid.html`.

MacTutor History of Mathematics Archive. "Epicycloid." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Epicycloid.html`.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 219–225, 1979.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 50–52, 1991.

Yates, R. C. "Epi- and Hypo-Cycloids." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 81–85, 1952.

Epicycloid—1-Cusped

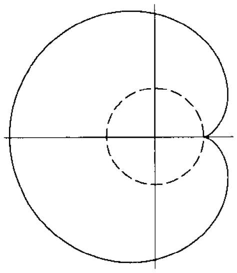

A 1-cusped epicycloid has $b = a$, so $n = 1$. The radius measured from the center of the large circle for a 1-cusped epicycloid is given by EPICYCLOID equation (9) with $n = 1$ so

$$\begin{aligned} r^2 &= \frac{a^2}{n^2}\left[(n^2+2n+2) - 2(n+1)\cos(n\phi)\right] \\ &= a^2[(1^2+2\cdot 1+2) - 2(1+1)\cos(1\cdot\phi)] \\ &= a^2(5-4\cos\phi) \end{aligned} \tag{1}$$

$$r = a\sqrt{5-4\cos\phi}, \tag{2}$$

and

$$\tan\theta = \frac{2\sin\phi - \sin(2\phi)}{2\cos\phi - \cos(2\phi)}. \tag{3}$$

The 1-cusped epicycloid is just an offset CARDIOID.

Epicycloid—2-Cusped

see NEPHROID

Epicycloid Evolute

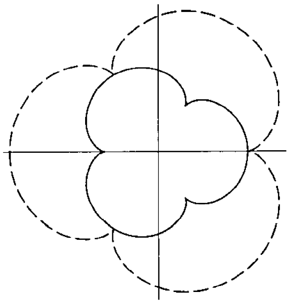

The EVOLUTE of the EPICYCLOID

$$\begin{aligned} x &= (a+b)\cos t - b\cos\left[\left(\frac{a+b}{b}\right)t\right] \\ y &= (a+b)\sin t - b\sin\left[\left(\frac{a+b}{b}\right)t\right] \end{aligned}$$

is another EPICYCLOID given by

$$\begin{aligned} x &= \frac{a}{a+2b}\left\{(a+b)\cos t + b\cos\left[\left(\frac{a+b}{b}\right)t\right]\right\} \\ y &= \frac{a}{a+2b}\left\{(a+b)\sin t + b\cos\left[\left(\frac{a+b}{b}\right)t\right]\right\}. \end{aligned}$$

Epicycloid Involute

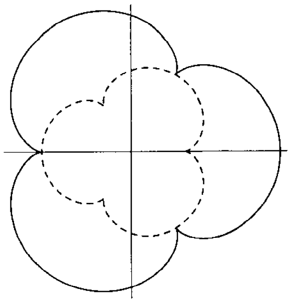

The INVOLUTE of the EPICYCLOID

$$x = (a+b)\cos t - b\cos\left[\left(\frac{a+b}{b}\right)t\right]$$
$$y = (a+b)\sin t - b\sin\left[\left(\frac{a+b}{b}\right)t\right]$$

is another EPICYCLOID given by

$$x = \frac{a+2b}{a}\left\{(a+b)\cos t + b\cos\left[\left(\frac{a+b}{b}\right)t\right]\right\}$$
$$y = \frac{a+2b}{a}\left\{(a+b)\sin t + b\cos\left[\left(\frac{a+b}{b}\right)t\right]\right\}.$$

Epicycloid Pedal Curve

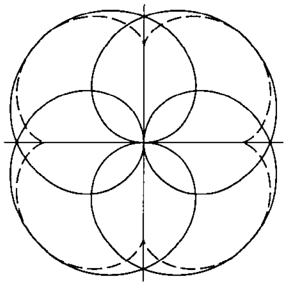

The PEDAL CURVE of an EPICYCLOID with PEDAL POINT at the center, shown for an epicycloid with four cusps, is not a ROSE as claimed by Lawrence (1972).

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, p. 204, 1972.

Epicycloid Radial Curve

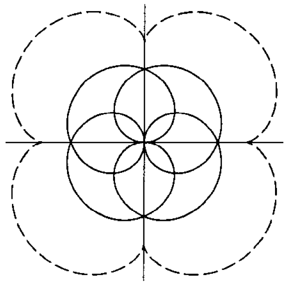

The RADIAL CURVE of an EPICYCLOID is shown above for an epicycloid with four cusps. It is not a ROSE, as claimed by Lawrence (1972).

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, p. 202, 1972.

Epimenides Paradox

A PARADOX, also called the LIAR'S PARADOX, attributed to the philosopher Epimenides in the sixth century BC. "All Cretans are liers... One of their own poets has said so." A sharper version of the paradox is the EUBULIDES PARADOX, "This statement is false."

see also EUBULIDES PARADOX, SOCRATES' PARADOX

References

Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, p. 17, 1989.

Epimorphism

A SURJECTIVE MORPHISM.

Epispiral

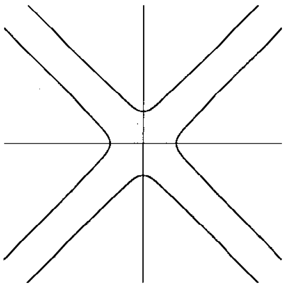

A plane curve with polar equation

$$r = \frac{a}{\cos(n\theta)}.$$

There are n sections if n is ODD and $2n$ if n is EVEN.

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 192–193, 1972.

Epispiral Inverse Curve

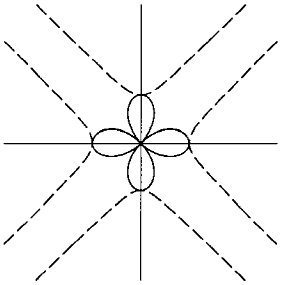

The INVERSE CURVE of the EPISPIRAL

$$r = a\sec(nt)$$

with INVERSION CENTER at the origin and inversion radius k is the ROSE

$$r = \frac{k\cos(nt)}{a}.$$

Epitrochoid

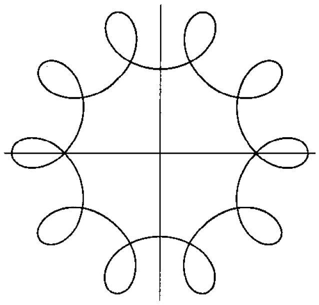

The ROULETTE traced by a point P attached to a CIRCLE of radius b rolling around the outside of a fixed

CIRCLE of radius a. These curves were studied by Dürer (1525), Desargues (1640), Huygens (1679), Leibniz, Newton (1686), L'Hospital (1690), Jakob Bernoulli (1690), la Hire (1694), Johann Bernoulli (1695), Daniel Bernoulli (1725), Euler (1745, 1781). An epitrochoid appears in Dürer's work *Instruction in Measurement with Compasses and Straight Edge* (1525). He called epitrochoids SPIDER LINES because the lines he used to construct the curves looked like a spider.

The parametric equations for an epitrochoid are

$$x = m\cos t - h\cos\left(\frac{m}{b}t\right)$$
$$y = m\sin t - h\sin\left(\frac{m}{b}t\right),$$

where $m \equiv a + b$ and h is the distance from P to the center of the rolling CIRCLE. Special cases include the LIMAÇON with $a = b$, the CIRCLE with $a = 0$, and the EPICYCLOID with $h = b$.

see also EPICYCLOID, HYPOTROCHOID, SPIROGRAPH

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 168–170, 1972.

Lee, X. "Epitrochoid." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Epitrochoid_dir/epitrochoid.html.`

Lee, X. "Epitrochoid and Hypotrochoid Movie Gallery." `http://www.best.com/~xah/SpecialPlaneCurves_dir/EpiHypoTMovieGallery_dir/epiHypoTMovieGallery.html.`

Epitrochoid Evolute

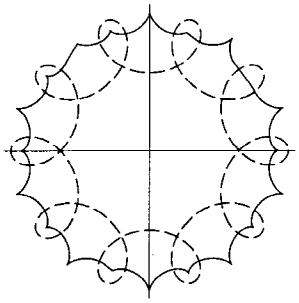

Epsilon

In mathematics, a small POSITIVE INFINITESIMAL quantity whose LIMIT is usually taken to be 0. The late mathematician P. Erdős also used the term "epsilons" to refer to children.

Epsilon-Neighborhood

see NEIGHBORHOOD

Epstein Zeta Function

$$Z\left|\begin{matrix}\mathbf{g}\\ \mathbf{h}\end{matrix}\right|(q;s)=\sum_{\mathbf{l}}\frac{e^{-2\pi i\mathbf{h}\cdot\mathbf{l}}}{[q(\mathbf{l}+\mathbf{g})]^{s/2}},$$

where $\mathbf{g}$ and $\mathbf{h}$ are arbitrary VECTORS, the SUM runs over a d-dimensional LATTICE, and $\mathbf{l} = -\mathbf{g}$ is omitted if $\mathbf{g}$ is a lattice VECTOR.

see also ZETA FUNCTION

References

Glasser, M. L. and Zucker, I. J. "Lattice Sums in Theoretical Chemistry." *Theoretical Chemistry: Advances and Perspectives, Vol. 5.* New York: Academic Press, pp. 69–70, 1980.

Shanks, D. "Calculation and Applications of Epstein Zeta Functions." *Math. Comput.* **29**, 271–287, 1975.

Equal

Two quantities are said to be equal if they are, in some well-defined sense, equivalent. Equality of quantities a and b is written $a = b$.

A symbol with three horizontal line segments ($\equiv$) resembling the equals sign is used to denote both equality by definition (e.g., $A \equiv B$ means A is DEFINED to be equal to B) and CONGRUENCE (e.g., $13 \equiv 12 \pmod 1$) means 13 divided by 12 leaves a REMAINDER of 1—a fact known to all readers of analog clocks).

see also CONGRUENCE, DEFINED, DIFFERENT, EQUAL BY DEFINITION, EQUALITY, EQUIVALENT, ISOMORPHISM

Equal by Definition

see DEFINED

Equal Detour Point

The center of an outer SODDY CIRCLE. It has TRIANGLE CENTER FUNCTION

$$\alpha = 1 + \frac{2\Delta}{a(b+c-a)} = \sec(\tfrac{1}{2}A)\cos(\tfrac{1}{2}B)\cos(\tfrac{1}{2}C) + 1.$$

Given a point Y not between A and B, a detour of length

$$|AY| + |YB| - |AB|$$

is made walking from A to B via Y, the point is of equal detour if the three detours from one side to another via Y are equal. If ABC has no ANGLE $> 2\sin^{-1}(4/5)$, then the point given by the above TRILINEAR COORDINATES is the unique equal detour point. Otherwise, the ISOPERIMETRIC POINT is also equal detour.

References

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Kimberling, C. "Isoperimetric Point and Equal Detour Point." `http://www.evansville.edu/~ck6/tcenters/recent/isoper.html.`

Veldkamp, G. R. "The Isoperimetric Point and the Point(s) of Equal Detour." *Amer. Math. Monthly* **92**, 546–558, 1985.

Equal Parallelians Point

The point of intersection of the three LINE SEGMENTS, each parallel to one side of a TRIANGLE and touching the other two, such that all three segments are of the same length. The TRILINEAR COORDINATES are

$$bc(ca+ab-bc) : ca(ab+bc-ca) : ab(bc+ca-ab).$$

References
Kimberling, C. "Equal Parallelians Point." `http://www.evansville.edu/~ck6/tcenters/recent/eqparal.html`.

Equality

A mathematical statement of the equivalence of two quantities. The equality "A is equal to B" is written $A = B$.

see also EQUAL, INEQUALITY

Equally Likely Outcomes Distribution

Let there be a set S with N elements, each of them having the same probability. Then

$$\begin{aligned} P(S) &= P\left(\bigcup_{i=1}^{N} E_i\right) = \sum_{i=1}^{N} P(E_i) \\ &= P(E_i)\sum_{i=1}^{N} 1 = NP(E_i). \end{aligned}$$

Using $P(S) \equiv 1$ gives

$$P(E_i) = \frac{1}{N}.$$

see also UNIFORM DISTRIBUTION

Equi-Brocard Center

The point Y for which the TRIANGLES BYC, CYA, and AYB have equal BROCARD ANGLES.

References
Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Equiaffinity

An AREA-preserving AFFINITY. Equiaffinities include the ELLIPTIC ROTATION, HYPERBOLIC ROTATION, HYPERBOLIC ROTATION (CROSSED), and PARABOLIC ROTATION.

Equiangular Spiral

see LOGARITHMIC SPIRAL

Equianharmonic Case

The case of the WEIERSTRAß ELLIPTIC FUNCTION with invariants $g_2 = 0$ and $g_3 = 1$.

see also LEMNISCATE CASE, PSEUDOLEMNISCATE CASE

References
Abramowitz, M. and Stegun, C. A. (Eds.). "Equianharmonic Case ($g_2 = 0$, $g_3 = 1$)." §18.13 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 652, 1972.

Equichordal Point

A point P for which all the CHORDS passing through P are of the same length. It satisfies

$$px + py = [\text{const}],$$

where p is the CHORD length. It is an open question whether a plane convex region can have two equichordal points.

see also EQUICHORDAL PROBLEM, EQUIPRODUCT POINT, EQUIRECIPROCAL POINT

Equichordal Problem

Is there a planar body bounded by a simple closed curve and star-shaped with respect to two interior points p and q whose point X-rays at p and q are both constant? Rychlik (1997) has answered the question in the negative.

see also EQUICHORDAL POINT

References
Rychlik, M. "The Equichordal Point Problem." *Elec. Res. Announcements Amer. Math. Soc.* **2**, 108–123, 1996.
Rychlik, M. "A Complete Solution to the Equichordal Problem of Fujiwara, Blaschke, Rothe, and Weitzenböck." *Invent. Math.* **129**, 141–212, 1997.

Equidecomposable

The ability of two plane or space regions to be DISSECTED into each other.

Equidistance Postulate

PARALLEL lines are everywhere equidistant. This POSTULATE is equivalent to the PARALLEL AXIOM.

References
Dunham, W. "Hippocrates' Quadrature of the Lune." Ch. 1 in *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, p. 54, 1990.

Equidistant Cylindrical Projection

see CYLINDRICAL EQUIDISTANT PROJECTION

Equidistributed Sequence

A sequence of REAL NUMBERS $\{x_n\}$ is equidistributed if the probability of finding x_n in any subinterval is proportional to the subinterval length.

see also WEYL'S CRITERION

References

Kuipers, L. and Niederreiter, H. *Uniform Distribution of Sequences.* New York: Wiley, 1974.

Pólya, G. and Szegő, G. *Problems and Theorems in Analysis I.* New York: Springer-Verlag, p. 88, 1972.

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 155–156, 1991.

Equilateral Hyperbola

see RECTANGULAR HYPERBOLA

Equilateral Triangle

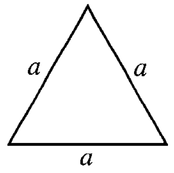

An equilateral triangle is a TRIANGLE with all three sides of equal length s. An equilateral triangle also has three equal 60° ANGLES.

An equilateral triangle can be constructed by TRISECTING all three ANGLES of any TRIANGLE (MORLEY'S THEOREM). NAPOLEON'S THEOREM states that if three equilateral triangles are drawn on the LEGS of any TRIANGLE (either all drawn inwards or outwards) and the centers of these triangles are connected, the result is another equilateral triangle.

Given the distances of a point from the three corners of an equilateral triangle, a, b, and c, the length of a side s is given by

$$3(a^4+b^4+c^4+s^4)=(a^2+b^2+c^2+s^2)^2 \qquad (1)$$

(Gardner 1977, pp. 56–57 and 63). There are infinitely many solutions for which a, b, and c are INTEGERS. In these cases, one of a, b, c, and s is DIVISIBLE by 3, one by 5, one by 7, and one by 8 (Guy 1994, p. 183).

The ALTITUDE h of an equilateral triangle is

$$h=\tfrac{1}{2}\sqrt{3}\,s, \qquad (2)$$

where s is the side length, so the AREA is

$$A=\tfrac{1}{2}sh=\tfrac{1}{4}\sqrt{3}\,s^2. \qquad (3)$$

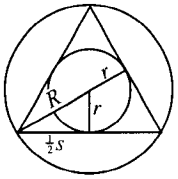

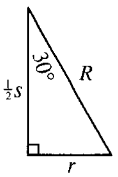

The INRADIUS r, CIRCUMRADIUS R, and AREA A can be computed directly from the formulas for a general regular POLYGON with side length s and $n=3$ sides,

$$r=\tfrac{1}{2}s\cot\left(\frac{\pi}{3}\right)=\tfrac{1}{2}s\tan\left(\frac{\pi}{6}\right)=\tfrac{1}{6}\sqrt{3}\,s \qquad (4)$$

$$R=\tfrac{1}{2}s\csc\left(\frac{\pi}{3}\right)=\tfrac{1}{2}s\sec\left(\frac{\pi}{6}\right)=\tfrac{1}{3}\sqrt{3}\,s \qquad (5)$$

$$A=\tfrac{1}{4}ns^2\cot\left(\frac{\pi}{3}\right)=\tfrac{1}{4}\sqrt{3}\,s^2. \qquad (6)$$

The AREAS of the INCIRCLE and CIRCUMCIRCLE are

$$A_r=\pi r^2=\tfrac{1}{12}\pi s^2 \qquad (7)$$

$$A_R=\pi R^2=\tfrac{1}{3}\pi s^2. \qquad (8)$$

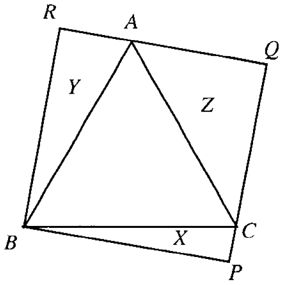

Let any RECTANGLE be circumscribed about an EQUILATERAL TRIANGLE. Then

$$X+Y=Z, \qquad (9)$$

where X, Y, and Z are the AREAS of the triangles in the figure (Honsberger 1985).

Begin with an arbitrary TRIANGLE and find the EXCENTRAL TRIANGLE. Then find the EXCENTRAL TRIANGLE of that triangle, and so on. Then the resulting triangle approaches an equilateral triangle. The only RATIONAL TRIANGLE is the equilateral triangle (Conway and Guy 1996). A POLYHEDRON composed of only equilateral triangles is known as a DELTAHEDRON.

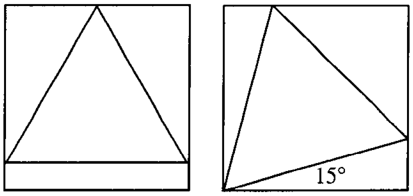

The largest equilateral triangle which can be inscribed in a UNIT SQUARE (left) has side length and area

$$s=1 \qquad (10)$$

$$A=\tfrac{1}{4}\sqrt{3}. \qquad (11)$$

The smallest equilateral triangle which can be inscribed (right) is oriented at an angle of 15° and has side length and area

$$s=\sec(15°)=\sqrt{6}-\sqrt{2} \qquad (12)$$

$$A=2\sqrt{3}-3 \qquad (13)$$

(Madachy 1979).

see also ACUTE TRIANGLE, DELTAHEDRON, EQUILIC QUADRILATERAL, FERMAT POINT, GYROELONGATED SQUARE DIPYRAMID, ICOSAHEDRON, ISOGONIC CENTERS, ISOSCELES TRIANGLE, MORLEY'S THEOREM, OCTAHEDRON, PENTAGONAL DIPYRAMID, RIGHT TRIANGLE, SCALENE TRIANGLE, SNUB DISPHENOID, TETRAHEDRON, TRIANGLE, TRIANGULAR DIPYRAMID, TRIAUGMENTED TRIANGULAR PRISM, VIVIANI'S THEOREM

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 121, 1987.
Conway, J. H. and Guy, R. K. "The Only Rational Triangle." In *The Book of Numbers.* New York: Springer-Verlag, pp. 201 and 228–239, 1996.
Dixon, R. *Mathographics.* New York: Dover, p. 33, 1991.
Gardner, M. *Mathematical Carnival: A New Round-Up of Tantalizers and Puzzles from Scientific American.* New York: Vintage Books, 1977.
Guy, R. K. "Rational Distances from the Corners of a Square." §D19 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 181–185, 1994.
Honsberger, R. "Equilateral Triangles." Ch. 3 in *Mathematical Gems I.* Washington, DC: Math. Assoc. Amer., 1973.
Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 19–21, 1985.
Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 115 and 129–131, 1979.

Equilibrium Point

An equilibrium point in GAME THEORY is a set of strategies $\{\hat{x}_1, \ldots, \hat{x}_n\}$ such that the ith payoff function $K_i(\mathbf{x})$ is larger or equal for any other ith strategy, i.e.,

$$K_i(\hat{x}_1, \ldots, \hat{x}_n) \geq K_i(\hat{x}_1, \ldots, \hat{x}_{i-1}, x_i, \hat{x}_{i+1}, \ldots, \hat{x}_n).$$

see NASH EQUILIBRIUM

Equilic Quadrilateral

A QUADRILATERAL in which a pair of opposite sides have the same length and are inclined at 60° to each other (or equivalently, satisfy $\langle A\rangle + \langle B\rangle = 120°$). Some interesting theorems hold for such quadrilaterals. Let $ABCD$ be an equilic quadrilateral with $AD = BC$ and $\langle A\rangle + \langle B\rangle = 120°$. Then

1. The MIDPOINTS P, Q, and R of the diagonals and the side CD always determine an EQUILATERAL TRIANGLE.
2. If EQUILATERAL TRIANGLE PCD is drawn outwardly on CD, then ΔPAB is also an EQUILATERAL TRIANGLE.
3. If EQUILATERAL TRIANGLES are drawn on AC, DC, and DB away from AB, then the three new VERTICES P, Q, and R are COLLINEAR.

See Honsberger (1985) for additional theorems.

References

Garfunkel, J. "The Equilic Quadrilateral." *Pi Mu Epsilon J.*, 317–329, Fall 1981.
Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 32–35, 1985.

Equinumerous

Let A and B be two classes of POSITIVE integers. Let $A(n)$ be the number of integers in A which are less than or equal to n, and let $B(n)$ be the number of integers in B which are less than or equal to n. Then if

$$A(n) \sim B(n),$$

A and B are said to be equinumerous.

The four classes of PRIMES $8k+1$, $8k+3$, $8k+5$, $8k+7$ are equinumerous. Similarly, since $8k+1$ and $8k+5$ are both of the form $4k+1$, and $8k+3$ and $8k+7$ are both of the form $4k+3$, $4k+1$ and $4k+3$ are also equinumerous.

see also BERTRAND'S POSTULATE, CHOQUET THEORY, PRIME COUNTING FUNCTION

References

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 21–22 and 31–32, 1993.

Equipollent

Two statements in LOGIC are said to be equipollent if they are deducible from each other. Two SETS with the same CARDINAL NUMBER are also said to be equipollent. The term EQUIPOTENT is sometimes used instead of equipollent.

Equipotent

see EQUIPOLLENT

Equipotential Curve

A curve in 2-D on which the value of a function $f(x, y)$ is a constant. Other synonymous terms are ISARITHM and ISOPLETH.

see also LEMNISCATE

Equiproduct Point

A point, such as interior points of a disk, such that

$$(px)(py) = [\text{const}],$$

where p is the CHORD length.

see also EQUICHORDAL POINT, EQUIRECIPROCAL POINT

Equireciprocal Point

A point, such as the FOCI of an ELLIPSE, which satisfies

$$\frac{1}{px} + \frac{1}{py} = [\text{const}],$$

where p is the CHORD length.

see also EQUICHORDAL POINT, EQUIPRODUCT POINT

Equirectangular Projection

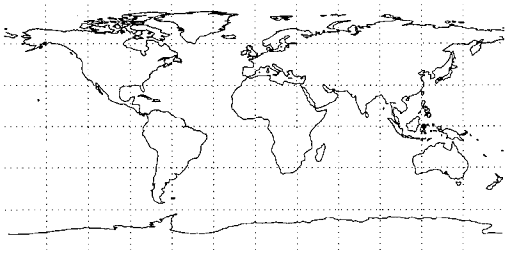

A MAP PROJECTION, also called a RECTANGULAR PROJECTION, in which the horizontal coordinate is the longitude and the vertical coordinate is the latitude.

Equiripple

A distribution of ERROR such that the ERROR remaining is always given approximately by the last term dropped.

Equitangential Curve

see TRACTRIX

Equivalence Class

An equivalence class is defined as a SUBSET of the form $\{x \in X : xRa\}$, where a is an element of X and the NOTATION "xRy" is used to mean that there is an EQUIVALENCE RELATION between x and y. It can be shown that any two equivalence classes are either equal or disjoint, hence the collection of equivalence classes forms a partition of X. For all $a, b \in X$, we have aRb IFF a and b belong to the same equivalence class.

A set of CLASS REPRESENTATIVES is a SUBSET of X which contains EXACTLY ONE element from each equivalence class.

For n a POSITIVE INTEGER, and a, b INTEGERS, consider the CONGRUENCE $a \equiv b \pmod{n}$, then the equivalence classes are the sets $\{\ldots, -2n, -n, 0, n, 2n, \ldots\}$, $\{\ldots, 1-2n, 1-n, 1, 1+n, 1+2n, \ldots\}$ etc. The standard CLASS REPRESENTATIVES are taken to be 0, 1, 2, ..., $n-1$.

see also CONGRUENCE, COSET

References

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 56–57, 1993.

Equivalence Problem

see METRIC EQUIVALENCE PROBLEM

Equivalence Relation

An equivalence relation on a set X is a SUBSET of $X \times X$, i.e., a collection R of ordered pairs of elements of X, satisfying certain properties. Write "xRy" to mean (x, y) is an element of R, and we say "x is related to y," then the properties are

1. Reflexive: aRa for all $a \in X$,
2. Symmetric: aRb IMPLIES bRa for all $a, b \in X$
3. Transitive: aRb and bRc imply aRc for all $a, b, c \in X$,

where these three properties are completely independent. Other notations are often used to indicate a relation, e.g., $a \equiv b$ or $a \sim b$.

see also EQUIVALENCE CLASS, TEICHMÜLLER SPACE

References

Stewart, I. and Tall, D. *The Foundations of Mathematics.* Oxford, England: Oxford University Press, 1977.

Equivalent

If $A \Rightarrow B$ and $B \Rightarrow A$ (i.e, $A \Rightarrow B \wedge B \Rightarrow A$, where $\Rightarrow$ denotes IMPLIES), then A and B are said to be equivalent, a relationship which is written symbolically as $A \Leftrightarrow B$ or $A \rightleftharpoons B$. However, if A and B are "equivalent by definition" (i.e., A is DEFINED to be B), this is written $A \equiv B$, a notation which conflicts with that for a CONGRUENCE.

see also DEFINED, IFF, IMPLIES

Equivalent Matrix

An $m \times n$ MATRIX A is said to be equivalent to another $m \times n$ MATRIX B IFF

$$\mathsf{B} = \mathsf{PAQ}$$

for P and Q any $m \times n$ and $n \times n$ MATRICES, respectively.

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1103, 1979.

Eratosthenes Sieve

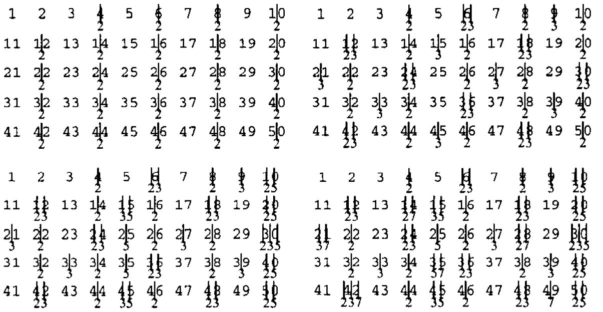
1 2 3 4 5 6 7 8 9 10
11 12 13 14 15 16 17 18 19 20
21 22 23 24 25 26 27 28 29 30
31 32 33 34 35 36 37 38 39 40
41 42 43 44 45 46 47 48 49 50
1 2 3 4 5 6 7 8 9 10
11 12 13 14 15 16 17 18 19 20
21 22 23 24 25 26 27 28 29 30
31 32 33 34 35 36 37 38 39 40
41 42 43 44 45 46 47 48 49 50
1 2 3 4 5 6 7 8 9 10
11 12 13 14 15 16 17 18 19 20
21 22 23 24 25 26 27 28 29 30
31 32 33 34 35 36 37 38 39 40
41 42 43 44 45 46 47 48 49 50
1 2 3 4 5 6 7 8 9 10
11 12 13 14 15 16 17 18 19 20
21 22 23 24 25 26 27 28 29 30
31 32 33 34 35 36 37 38 39 40
41 42 43 44 45 46 47 48 49 50

An ALGORITHM for making tables of PRIMES. Sequentially write down the INTEGERS from 2 to the highest number n you wish to include in the table. Cross out all numbers > 2 which are divisible by 2 (every second number). Find the smallest remaining number > 2. It is 3. So cross out all numbers > 3 which are divisible by 3 (every third number). Find the smallest remaining number > 3. It is 5. So cross out all numbers > 5 which are divisible by 5 (every fifth number).

Continue until you have crossed out all numbers divisible by $\lfloor \sqrt{n} \rfloor$, where $\lfloor x \rfloor$ is the FLOOR FUNCTION. The numbers remaining are PRIME. This procedure is illustrated in the above diagram which sieves up to 50, and therefore crosses out PRIMES up to $\lfloor\sqrt{50}\rfloor = 7$. If the procedure is then continued up to n, then the number of cross-outs gives the number of distinct PRIME factors of each number.

References
Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 127–130, 1996.
Ribenboim, P. *The New Book of Prime Number Records.* New York: Springer-Verlag, pp. 20–21, 1996.

Erdős-Anning Theorem

If an infinite number of points in the PLANE are all separated by INTEGER distances, then all the points lie on a straight LINE.

Erdős-Kac Theorem

A deeper result than the HARDY-RAMANUJAN THEOREM. Let $N(x, a, b)$ be the number of INTEGERS in $[3, x]$ such that inequality

$$a \leq \frac{\omega(n) - \ln\ln n}{\sqrt{\ln\ln n}} \leq b$$

holds, where $\omega(n)$ is the number of different PRIME factors of n. Then

$$\lim_{x\to\infty} N(x, a, b) = \frac{(x + o(x))}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\,dt.$$

The theorem is discussed in Kac (1959).

References
Kac, M. *Statistical Independence in Probability, Analysis and Number Theory.* New York: Wiley, 1959.
Riesel, H. "The Erdős-Kac Theorem." *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, pp. 158–159, 1994.

Erdős-Mordell Theorem

If O is any point inside a TRIANGLE ΔABC, and P, Q, and R are the feet of the perpendiculars from O upon the respective sides BC, CA, and AB, then

$$OA + OB + OC \geq 2(OP + OQ + OR).$$

Oppenheim (1961) and Mordell (1962) also showed that

$$OA \times OB \times OC \geq (OQ + OR)(OR + OP)(OP + OQ).$$

References
Bankoff, L. "An Elementary Proof of the Erdős-Mordell Theorem." *Amer. Math. Monthly* **65**, 521, 1958.
Brabant, H. "The Erdős-Mordell Inequality Again." *Nieuw Tijdschr. Wisk.* **46**, 87, 1958/1959.
Casey, J. *A Sequel to the First Six Books of the Elements of Euclid, 6th ed.* Dublin: Hodges, Figgis, & Co., p. 253, 1892.
Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, p. 9, 1969.
Erdős, P. "Problem 3740." *Amer. Math. Monthly* **42**, 396, 1935.
Fejes-Tóth, L. *Lagerungen in der Ebene auf der Kugel und im Raum.* Berlin: Springer, 1953.
Mordell, L. J. "On Geometric Problems of Erdős and Oppenheim." *Math. Gaz.* **46**, 213–215, 1962.
Mordell, L. J. and Barrow, D. F. "Solution to Problem 3740." *Amer. Math. Monthly* **44**, 252–254, 1937.
Oppenheim, A. "The Erdős Inequality and Other Inequalities for a Triangle." *Amer. Math. Monthly* **68**, 226–230 and 349, 1961.
Veldkamp, G. R. "The Erdős-Mordell Inequality." *Nieuw Tijdschr. Wisk.* **45**, 193–196, 1957/1958.

Erdős Number

An author's Erdős number is 1 if he has co-authored a paper with Erdős, 2 if he has co-authored a paper with someone who has co-authored a paper with Erdős, etc.

References
Grossman, J. and Ion, P. "The Erdős Number Project." `http://www.acs.oakland.edu/~grossman/erdoshp.html`.

Erdős Reciprocal Sum Constants

see A-SEQUENCE, $B2$-SEQUENCE, NONAVERAGING SEQUENCE

Erdős-Selfridge Function

The Erdős-Selfridge function $g(k)$ is defined as the least integer bigger than $k + 1$ such that all prime factors of $\binom{g(k)}{k}$ exceed k (Ecklund *et al.* 1974). The best lower bound known is

$$g(k) \geq \exp\left(c\frac{\ln^3 k}{\ln\ln k}^{1/2}\right)$$

(Granville and Ramare 1996). Scheidler and Williams (1992) tabulated $g(k)$ up to $k = 140$, and Lukes *et al.* (1997) tabulated $g(k)$ for $135 \leq k \leq 200$. The values for $n = 2, 3, \ldots$ are 4, 7, 7, 23, 62, 143, 44, 159, 46, 47, 174, 2239, ... (Sloane's A046105).

see also BINOMIAL COEFFICIENT, LEAST PRIME FACTOR

References
Ecklund, E. F. Jr.; Erdős, P.; and Selfridge, J. L. "A New Function Associated with the prime factors of $\binom{n}{k}$. *Math. Comput.* **28**, 647–649, 1974.
Erdős, P.; Lacampagne, C. B.; and Selfridge, J. L. "Estimates of the Least Prime Factor of a Binomial Coefficient." *Math. Comput.* **61**, 215–224, 1993.
Granville, A. and Ramare, O. "Explicit Bounds on Exponential Sums and the Scarcity of Squarefree Binomial Coefficients." *Mathematika* **43**, 73–107, 1996.
Lukes, R. F.; Scheidler, R.; and Williams, H. C. "Further Tabulation of the Erdős-Selfridge Function." *Math. Comput.* **66**, 1709–1717, 1997.
Scheidler, R. and Williams, H. C. "A Method of Tabulating the Number-Theoretic Function $g(k)$." *Math. Comput.* **59**, 251–257, 1992.
Sloane, N. J. A. Sequence A046105 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Erdős Squarefree Conjecture

The CENTRAL BINOMIAL COEFFICIENT $\binom{2n}{n}$ is never SQUAREFREE for $n > 4$. This was proved true for all sufficiently large n by SÁRKÖZY'S THEOREM. Goetgheluck (1988) proved the CONJECTURE true for $4 < n \leq 2^{42205184}$ and Vardi (1991) for $4 < n < 2^{774840978}$. The conjecture was proved true in its entirely by Granville and Ramare (1996).

see also CENTRAL BINOMIAL COEFFICIENT

References

Erdős, P. and Graham, R. L. *Old and New Problems and Results in Combinatorial Number Theory.* Geneva, Switzerland: L'Enseignement Mathématique Université de Genève, Vol. 28, p. 71, 1980.

Goetgheluck, P. "Prime Divisors of Binomial Coefficients." *Math. Comput.* **51**, 325–329, 1988.

Granville, A. and Ramare, O. "Explicit Bounds on Exponential Sums and the Scarcity of Squarefree Binomial Coefficients." *Mathematika* **43**, 73–107, 1996.

Sander, J. W. "On Prime Divisors of Binomial Coefficients." *Bull. London Math. Soc.* **24**, 140–142, 1992.

Sander, J. W. "A Story of Binomial Coefficients and Primes." *Amer. Math. Monthly* **102**, 802–807, 1995.

Sárközy, A. "On Divisors of Binomial Coefficients. I." *J. Number Th.* **20**, 70–80, 1985.

Vardi, I. "Applications to Binomial Coefficients." *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 25–28, 1991.

Erdős-Szekeres Theorem

Suppose a, $b \in \mathbb{N}$, $n = ab + 1$, and $x_1, \ldots, x_n$ is a sequence of n REAL NUMBERS. Then this sequence contains a MONOTONIC increasing (decreasing) subsequence of $a + 1$ terms or a MONOTONIC decreasing (increasing) subsequence of $b + 1$ terms. DILWORTH'S LEMMA is a generalization of this theorem.

see also COMBINATORICS

Erf

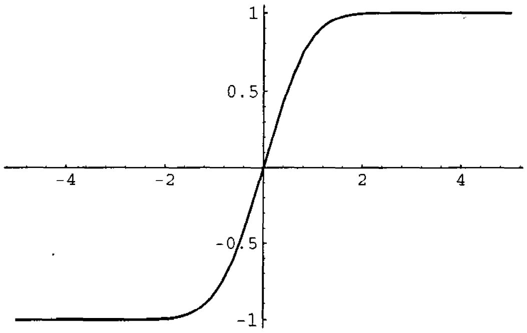

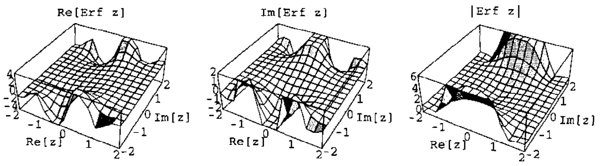

The "error function" encountered in integrating the GAUSSIAN DISTRIBUTION.

$$\begin{aligned} \operatorname{erf}(z) &\equiv \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\,dt & (1)\\ &= 1 - \operatorname{erfc}(z) & (2)\\ &= \sqrt{\pi}\,\gamma(\tfrac{1}{2}, z^2), & (3) \end{aligned}$$

where ERFC is the complementary error function and $\gamma(x, a)$ is the incomplete GAMMA FUNCTION. It can also be defined as a MACLAURIN SERIES

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{n!(2n+1)}. \qquad (4)$$

Erf has the values

$$\operatorname{erf}(0) = 0 \qquad (5)$$

$$\operatorname{erf}(\infty) = 1. \qquad (6)$$

It is an ODD FUNCTION

$$\operatorname{erf}(-z) = -\operatorname{erf}(z), \qquad (7)$$

and satisfies

$$\operatorname{erf}(z) + \operatorname{erfc}(z) = 1. \qquad (8)$$

Erf may be expressed in terms of a CONFLUENT HYPERGEOMETRIC FUNCTION OF THE FIRST KIND M as

$$\operatorname{erf}(z) = \frac{2z}{\sqrt{\pi}} M(\tfrac{1}{2}, \tfrac{3}{2}, -z^2) = \frac{2z}{\sqrt{\pi}} e^{-z^2} M(1, \tfrac{3}{2}, z^2). \quad (9)$$

Erf is bounded by

$$\frac{1}{x + \sqrt{x^2 + 2}} < e^{x^2} \int_x^\infty e^{-t^2}\,dt \leq \frac{1}{x + \sqrt{x^2 + \frac{4}{\pi}}}. \quad (10)$$

Its DERIVATIVE is

$$\frac{d^n}{dz^n} \operatorname{erf}(z) = (-1)^{n-1} \frac{2}{\sqrt{\pi}} H_n(z) e^{-z^2}, \qquad (11)$$

where H_n is a HERMITE POLYNOMIAL. The first DERIVATIVE is

$$\frac{d}{dz} \operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} e^{-z^2/2}, \qquad (12)$$

and the integral is

$$\int \operatorname{erf}(z)\,dz = z \operatorname{erf}(z) + \frac{e^{-z^2}}{\sqrt{\pi}}. \qquad (13)$$

For $x \ll 1$, erf may be computed from

$$\begin{aligned}
\operatorname{erf}(x) &= \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\,dt && (14)\\
&= \frac{2}{\sqrt{\pi}} \int_0^x \sum_{k=0}^{\infty} \frac{(-t^2)^k}{k!}\,dt\\
&= \frac{2}{\sqrt{\pi}} \int_0^x \sum_{k=0}^{\infty} \frac{(-1)^k t^{2k}}{k!}\,dt\\
&= \frac{2}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{x^{2k+1}(-1)^k}{k!(2k+1)} && (15)\\
&= \frac{2}{\sqrt{\pi}}(x - \tfrac{1}{3}x^3 + \tfrac{1}{10}x^5 - \tfrac{1}{42}x^7 + \tfrac{1}{216}x^9\\
&\quad - \tfrac{1}{1320}x^{11} + \ldots) && (16)\\
&= \frac{2}{\sqrt{\pi}} e^{-x^2} x \left[1 + \frac{2x^2}{1\cdot 3} + \frac{(2x)^2}{1\cdot 3\cdot 5} + \ldots\right] && (17)
\end{aligned}$$

(Acton 1990). For $x \gg 1$,

$$\begin{aligned}
\operatorname{erf}(x) &= \frac{2}{\sqrt{\pi}}\left(\int_0^\infty e^{-t^2}\,dt - \int_x^\infty e^{-t^2}\,dt\right)\\
&= 1 - \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2}\,dt. && (18)
\end{aligned}$$

Using INTEGRATION BY PARTS gives

$$\begin{aligned}
\int_x^\infty e^{-t^2}\,dt &= -\frac{1}{2}\int_x^\infty \frac{1}{x}\,d(e^{-x^2})\\
&= -\frac{1}{2}\left[\frac{e^{-t^2}}{t}\right]_x^\infty - \frac{1}{2}\int_x^\infty \frac{e^{-t^2}\,dt}{t^2}\\
&= \frac{e^{-x^2}}{2x} + \frac{1}{4}\int_x^\infty \frac{1}{t^3}\,d(e^{-t^2})\\
&= \frac{e^{-x^2}}{2x} - \frac{e^{-x^2}}{4x^3} - \ldots, && (19)
\end{aligned}$$

so

$$\operatorname{erf}(x) = 1 - \frac{e^{-x^2}}{\sqrt{\pi}\,x}\left(1 - \frac{1}{2x^2} - \ldots\right) \qquad (20)$$

and continuing the procedure gives the ASYMPTOTIC SERIES

$$\begin{aligned}
\operatorname{erf}(x) = 1 - \frac{e^{-x^2}}{\sqrt{\pi}}&(x^{-1} - \tfrac{1}{2}x^{-3}\\
&+ \tfrac{3}{4}x^{-5} - \tfrac{15}{8}x^{-7} + \tfrac{105}{16}x^{-9} + \ldots). && (21)
\end{aligned}$$

A COMPLEX generalization of erf is defined as

$$\begin{aligned}
w(z) &\equiv e^{-z^2}\operatorname{erfc}(-iz) && (22)\\
&= e^{-z^2}\left(1 + \frac{2i}{\sqrt{\pi}} + \frac{2i}{\sqrt{\pi}}\int_0^z e^{t^2}\,dt\right) && (23)\\
&= \frac{i}{\pi}\int_{-\infty}^{\infty} \frac{e^{-t^2}\,dt}{z-t} = \frac{2iz}{\pi}\int_0^\infty \frac{e^{-t^2}\,dt}{z^2-t^2}. && (24)
\end{aligned}$$

see also DAWSON'S INTEGRAL, ERFC, ERFI, GAUSSIAN INTEGRAL, NORMAL DISTRIBUTION FUNCTION, PROBABILITY INTEGRAL

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Error Function" and "Repeated Integrals of the Error Function." §7.1–7.2 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 297–300, 1972.

Acton, F. S. *Numerical Methods That Work, 2nd printing.* Washington, DC: Math. Assoc. Amer., p. 16, 1990.

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 568–569, 1985.

Spanier, J. and Oldham, K. B. "The Error Function $\operatorname{erf}(x)$ and Its Complement $\operatorname{erfc}(x)$." Ch. 40 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 385–393, 1987.

Erfc

The "complementary error function"

$$\begin{aligned}
\operatorname{erfc}(x) &\equiv \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}\,dt && (1)\\
&= 1 - \operatorname{erf}(x) && (2)\\
&= \sqrt{\pi}\,\gamma(\tfrac{1}{2}, z^2), && (3)
\end{aligned}$$

where γ is the incomplete GAMMA FUNCTION. It has the values

$$\operatorname{erfc}(0) = 1 \qquad (4)$$

$$\operatorname{erfc}(\infty) = 0 \qquad (5)$$

$$\operatorname{erfc}(-x) = 2 - \operatorname{erfc}(x) \qquad (6)$$

$$\int_0^\infty \operatorname{erfc}(x)\,dx = \frac{1}{\sqrt{\pi}} \qquad (7)$$

$$\int_0^\infty \operatorname{erfc}^2(x)\,dx = \frac{2-\sqrt{2}}{\sqrt{\pi}}. \qquad (8)$$

A generalization is obtained from the differential equation

$$\frac{d^2y}{dz^2} + 2z\frac{dy}{dz} - 2ny = 0. \qquad (9)$$

The general solution is then

$$y = A\operatorname{erfci}_n(z) + B\operatorname{erfci}_n(-z), \qquad (10)$$

where $\operatorname{erfci}_n(z)$ is the erfc integral. For integral $n \geq 1$,

$$\begin{aligned}
\operatorname{erfci}_n(z) &= \underbrace{\int \cdots \int}_{n} \operatorname{erfc}(z)\,dz && (11)\\
&= \frac{2}{\sqrt{\pi}}\int_0^\infty \frac{(t-z)^2}{n!} e^{-t^2}\,dt. && (12).
\end{aligned}$$

The definition can be extended to $n = -1$ and 0 using

$$\operatorname{erfci}_{-1}(z) = \frac{2}{\sqrt{\pi}} e^{-z^2} \tag{13}$$

$$\operatorname{erfci}_0(z) = \operatorname{erfc}(z). \tag{14}$$

see also ERF, ERFI

References
Abramowitz, M. and Stegun, C. A. (Eds.). "Repeated Integrals of the Error Function." §7.2 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 299–300, 1972.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Incomplete Gamma Function, Error Function, Chi-Square Probability Function, Cumulative Poisson Function." §6.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 209–214, 1992.
Spanier, J. and Oldham, K. B. "The Error Function erf(x) and Its Complement erfc(x)" and "The exp(x) and erfc($\sqrt{x}$) and Related Functions." Chs. 40 and 41 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 385–393 and 395–403, 1987.

Erfi

$$\operatorname{erfi}(z) \equiv -i \operatorname{erf}(iz).$$

see also ERF, ERFC

Ergodic Measure

An ENDOMORPHISM is called ergodic if it is true that $T^{-1}A = A$ IMPLIES $m(A) = 0$ or 1, where $T^{-1}A = \{x \in X : T(x) \in A\}$. Examples of ergodic endomorphisms include the MAP $X \to 2x$ mod 1 on the unit interval with LEBESGUE MEASURE, certain AUTOMORPHISMS of the TORUS, and "Bernoulli shifts" (and more generally "Markov shifts").

Given a MAP T and a SIGMA ALGEBRA, there may be many ergodic measures. If there is only one ergodic measure, then T is called uniquely ergodic. An example of a uniquely ergodic transformation is the MAP $x \mapsto x + a$ mod 1 on the unit interval when a is irrational. Here, the unique ergodic measure is LEBESGUE MEASURE.

Ergodic Theory

Ergodic theory can be described as the statistical and qualitative behavior of measurable group and semigroup actions on MEASURE SPACES. The GROUP is most commonly $\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}^+$, and $\mathbb{Z}$.

Ergodic theory had its origins in the work of Boltzmann in statistical mechanics. Its mathematical origins are due to von Neumann, Birkhoff, and Koopman in the 1930s. It has since grown to be a huge subject and has applications not only to statistical mechanics, but also to number theory, differential geometry, functional analysis, etc. There are also many internal problems (e.g., ergodic theory being applied to ergodic theory) which are interesting.

see also AMBROSE-KAKUTANI THEOREM, BIRKHOFF'S ERGODIC THEOREM, DYE'S THEOREM, DYNAMICAL SYSTEM, HOPF'S THEOREM, ORNSTEIN'S THEOREM

References
Billingsley, P. *Ergodic Theory and Information.* New York: Wiley, 1965.
Cornfeld, I.; Fomin, S.; and Sinai, Ya. G. *Ergodic Theory.* New York: Springer-Verlag, 1982.
Katok, A. and Hasselblatt, B. *An Introduction to the Modern Theory of Dynamical Systems.* Cambridge, England: Cambridge University Press, 1996.
Nadkarni, M. G. *Basic Ergodic Theory.* India: Hindustan Book Agency, 1995.
Parry, W. *Topics in Ergodic Theory.* Cambridge, England: Cambridge University Press, 1982.
Smorodinsky, M. *Ergodic Theory, Entropy.* Berlin: Springer-Verlag, 1971.
Walters, P. *Ergodic Theory: Introductory Lectures.* New York: Springer-Verlag, 1975.

Ergodic Transformation

A transformation which has only trivial invariant SUBSETS is said to be invariant.

Erlanger Program

A program initiated by F. Klein in an 1872 lecture to describe geometric structures in terms of their group AUTOMORPHISMS.

References
Klein, F. "Vergleichende Betrachtungen über neuere geometrische Forschungen." 1872.
Yaglom, I. M. *Felix Klein and Sophus Lie: Evolution of the Idea of Symmetry in the Nineteenth Century.* Boston, MA: Birkhäuser, 1988.

Ermakoff's Test

The series $\sum f(n)$ for a monotonic nonincreasing $f(x)$ is convergent if

$$\overline{\lim_{x \to \infty}} \frac{e^x f(e^x)}{f(x)} < 1$$

and divergent if

$$\underline{\lim_{x \to \infty}} \frac{e^x f(e^x)}{f(x)} > 1.$$

References
Bromwich, T. J. I'a and MacRobert, T. M. *An Introduction to the Theory of Infinite Series, 3rd ed.* New York: Chelsea, p. 43, 1991.

Error

The difference between a quantity and its estimated or measured quantity.

see also ABSOLUTE ERROR, PERCENTAGE ERROR, RELATIVE ERROR

Error-Correcting Code

An error-correcting code is an algorithm for expressing a sequence of numbers such that any errors which are introduced can be detected and corrected (within certain limitations) based on the remaining numbers. The study of error-correcting codes and the associated mathematics is known as CODING THEORY.

Error detection is much simpler than error correction, and one or more "check" digits are commonly embedded in credit card numbers in order to detect mistakes. Early space probes like Mariner used a type of error-correcting code called a block code, and more recent space probes use convolution codes. Error-correcting codes are also used in CD players, high speed modems, and cellular phones. Modems use error detection when they compute CHECKSUMS, which are sums of the digits in a given transmission modulo some number. The ISBN used to identify books also incorporates a check DIGIT.

A powerful check for 13 DIGIT numbers consists of the following. Write the number as a string of DIGITS $a_1a_2a_3\cdots a_{13}$. Take $a_1+a_3+\ldots+a_{13}$ and double. Now add the number of DIGITS in ODD positions which are > 4 to this number. Now add $a_2+a_4+\ldots+a_{12}$. The check number is then the number required to bring the last DIGIT to 0. This scheme detects all single DIGIT errors and all TRANSPOSITIONS of adject DIGITS except 0 and 9.

see also CHECKSUM, CODING THEORY, GALOIS FIELD, HADAMARD MATRIX, ISBN

References

Conway, J. H. and Sloane, N. J. A. "Error-Correcting Codes." §3.2 in *Sphere Packings, Lattices, and Groups, 2nd ed.* New York: Springer-Verlag, pp. 75–88, 1993.

Gallian, J. "How Computers Can Read and Correct ID Numbers." *Math Horizons*, pp. 14–15, Winter 1993.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 119–121, 1994.

MacWilliams, F. J. and Sloane, N. J. A. *The Theory of Error-Correcting Codes.* Amsterdam, Netherlands: North-Holland, 1977.

Error Curve

see GAUSSIAN FUNCTION

Error Function

see ERF, ERFC

Error Function Distribution

A NORMAL DISTRIBUTION with MEAN 0.

$$P(x)=\frac{h}{\sqrt{\pi}}e^{-h^2x^2}. \tag{1}$$

The CHARACTERISTIC FUNCTION is

$$\phi(t)=e^{-t^2/(4h^2)}. \tag{2}$$

The MEAN, VARIANCE, SKEWNESS, and KURTOSIS are

$$\mu=0 \tag{3}$$

$$\sigma^2=\frac{1}{2h^2} \tag{4}$$

$$\gamma_1=0 \tag{5}$$

$$\gamma_2=0. \tag{6}$$

The CUMULANTS are

$$\kappa_1=0 \tag{7}$$

$$\kappa_2=\frac{1}{2h^2} \tag{8}$$

$$\kappa_n=0 \tag{9}$$

for $n\geq 3$.

Error Propagation

Given a FORMULA $y=f(x)$ with an ABSOLUTE ERROR in x of dx, the ABSOLUTE ERROR is dy. The RELATIVE ERROR is dy/y. If $x=f(u,v)$, then

$$x_i-\bar{x}=(u_i-\bar{u})\frac{\partial x}{\partial u}+(v_i-\bar{v})\frac{\partial x}{\partial v}+\ldots, \tag{1}$$

so

$$\begin{aligned}{\sigma_x}^2&\equiv\frac{1}{N-1}\sum_i^N(x_i-\bar{x})^2\\&=\frac{1}{N-1}\sum_i^N\left[(u_i-\bar{u})^2\left(\frac{\partial x}{\partial u}\right)^2+(v_i-\bar{v})^2\left(\frac{\partial x}{\partial v}\right)^2\right.\\&\quad\left.+2(u_i-\bar{u})(v_i-\bar{v})\left(\frac{\partial x}{\partial u}\right)\left(\frac{\partial x}{\partial v}\right)+\ldots\right].\end{aligned} \tag{2}$$

The definitions of VARIANCE and COVARIANCE then give

$${\sigma_u}^2\equiv\frac{1}{N-1}\sum_{i=1}^N(u_i-\bar{u})^2 \tag{3}$$

$${\sigma_v}^2\equiv\frac{1}{N-1}\sum_{i=1}^N(v_i-\bar{v})^2 \tag{4}$$

$${\sigma_{uv}}^2\equiv\frac{1}{N-1}\sum_{i=1}^N(u_i-\bar{u})(v_i-\bar{v}), \tag{5}$$

so

$$\begin{aligned}{\sigma_x}^2={\sigma_u}^2\left(\frac{\partial x}{\partial u}\right)^2+{\sigma_v}^2\left(\frac{\partial x}{\partial v}\right)^2\\+2\sigma_{uv}\left(\frac{\partial x}{\partial u}\right)\left(\frac{\partial x}{\partial v}\right)+\ldots.\end{aligned} \tag{6}$$

If u and v are uncorrelated, then $\sigma_{uv}=0$ so

$${\sigma_x}^2={\sigma_u}^2\left(\frac{\partial x}{\partial u}\right)^2+{\sigma_v}^2. \tag{7}$$

Now consider addition of quantities with errors. For $x = au \pm bv$, $\partial x/\partial u = a$ and $\partial x/\partial v = \pm b$, so

$$\sigma_x{}^2 = a^2\sigma_u{}^2 + b^2\sigma_v{}^2 \pm 2ab\sigma_{uv}{}^2. \tag{8}$$

For division of quantities with $x = \pm au/v$, $\partial x/\partial u = \pm a/v$ and $\partial x/\partial v = \mp au/v^2$, so

$$\sigma_x{}^2 = \frac{a^2}{v^2}\sigma_u{}^2 + \frac{a^2u^2}{\sigma_v{}^4} - 2\frac{a}{v}\frac{au}{v^2}\sigma_{uv}{}^2. \tag{9}$$

$$\begin{aligned}\left(\frac{\sigma_x}{x}\right)^2 &= \frac{a^2}{v^2}\frac{v^2}{a^2u^2}\sigma_u{}^2 + \frac{a^2u^2}{v^4}\frac{v^2}{a^2u^2} - 2\left(\frac{a}{v}\right)\left(\frac{au}{v^2}\right)\sigma_{uv}{}^2 \\ &= \left(\frac{\sigma_u}{u}\right)^2 + \left(\frac{\sigma_v}{v}\right)^2 - 2\left(\frac{\sigma_{uv}}{u}\right)\left(\frac{\sigma_{uv}}{v}\right). \end{aligned} \tag{10}$$

For exponentiation of quantities with

$$x = a^{\pm bu} = (e^{\ln a})^{\pm bu} = e^{\pm b(\ln a)u}, \tag{11}$$

$$\frac{\partial x}{\partial u} = \pm b(\ln a)e^{\pm b\ln au} = \pm b(\ln a)x, \tag{12}$$

so

$$\sigma_x = \sigma_u b(\ln a)x \tag{13}$$

$$\frac{\sigma_x}{x} = b\ln a\sigma_u. \tag{14}$$

If $a = e$, then

$$\frac{\sigma_x}{x} = b\sigma_u. \tag{15}$$

For LOGARITHMS of quantities with $x = a\ln(\pm bu)$, $\partial x/\partial u = a(\pm b)/(\pm bu) = a/u$, so

$$\sigma_x{}^2 = \sigma_u{}^2\left(\frac{a^2}{u^2}\right) \tag{16}$$

$$\sigma_x = a\frac{\sigma_u}{u}. \tag{17}$$

For multiplication with $x = \pm auv$, $\partial x/\partial u = \pm av$ and $\partial x/\partial v = \pm au$, so

$$\sigma_x{}^2 = a^2v^2\sigma_u{}^2 + a^2u^2\sigma_v{}^2 + 2a^2uv\sigma_{uv}{}^2 \tag{18}$$

$$\begin{aligned}\left(\frac{\sigma_x}{x}\right)^2 &= \frac{a^2v^2}{a^2u^2v^2}\sigma_u{}^2 + \frac{a^2u^2}{a^2u^2v^2}\sigma_v{}^2 + \frac{2a^2uv}{a^2u^2v^2}\sigma_{uv}{}^2 \\ &= \left(\frac{\sigma_u}{u}\right)^2 + \left(\frac{\sigma_v}{v}\right)^2 + 2\left(\frac{\sigma_{uv}}{u}\right)\left(\frac{\sigma_{uv}}{v}\right). \end{aligned} \tag{19}$$

For POWERS, with $x = au^{\pm b}$, $\partial x/\partial u = \pm abu^{\pm b-1} = \pm bx/u$, so

$$\sigma_x{}^2 = \sigma_u{}^2\frac{b^2x^2}{u^2} \tag{20}$$

$$\frac{\sigma_x}{x} = b\frac{\sigma_u}{u}. \tag{21}$$

see also ABSOLUTE ERROR, PERCENTAGE ERROR, RELATIVE ERROR

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 14, 1972.

Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences.* New York: McGraw-Hill, pp. 58–64, 1969.

Escher's Map

$$f(z) \mapsto z^{(1+\cos\beta+i\sin\beta)/2}.$$

Escribed Circle

see EXCIRCLE

Essential Singularity

A SINGULARITY a for which $f(z)(z-a)^n$ is not DIFFERENTIABLE for any INTEGER $n > 0$.

see also PICARD'S THEOREM, WEIERSTRAß-CASORATI THEOREM

Estimate

An estimate is an educated guess for an unknown quantity or outcome based on known information. The making of estimates is an important part of statistics, since care is needed to provide as accurate an estimate as possible using as little input data as possible. Often, an estimate for the uncertainty ΔE of an estimate E can also be determined statistically. A rule that tells how to calculate an estimate based on the measurements contained in a sample is called an ESTIMATOR.

see also BIAS (ESTIMATOR), ERROR, ESTIMATOR

References

Iyanaga, S. and Kawada, Y. (Eds.). "Statistical Estimation and Statistical Hypothesis Testing." Appendix A, Table 23 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 1486–1489, 1980.

Estimator

An estimator is a rule that tells how to calculate an ESTIMATE based on the measurements contained in a sample. For example, the "sample MEAN" AVERAGE $\bar{x}$ is an estimator for the population MEAN μ.

The mean square error of an estimator $\tilde{\theta}$ is defined by

$$\mathrm{MSE} \equiv \left\langle(\tilde{\theta} - \theta)^2\right\rangle.$$

Let B be the BIAS, then

$$\begin{aligned}\mathrm{MSE} &= \left\langle[(\tilde{\theta} - \langle\tilde{\theta}\rangle) + B(\tilde{\theta})]^2\right\rangle \\ &= \left\langle(\tilde{\theta} - \langle\tilde{\theta}\rangle)^2\right\rangle + B^2(\tilde{\theta}) \equiv V(\tilde{\theta}) + B^2(\tilde{\theta}),\end{aligned}$$

where V is the estimator VARIANCE.

see also BIAS (ESTIMATOR), ERROR, ESTIMATE, *k*-STATISTIC

Eta Function

see DEDEKIND ETA FUNCTION, DIRICHLET ETA FUNCTION, THETA FUNCTION

Ethiopian Multiplication

see RUSSIAN MULTIPLICATION

Etruscan Venus Surface

A 3-D shadow of a 4-D KLEIN BOTTLE.

see also IDA SURFACE

References

Peterson, I. *Islands of Truth: A Mathematical Mystery Cruise.* New York: W. H. Freeman, pp. 42–44, 1990.

Eubulides Paradox

The PARADOX "This statement is false," stated in the fourth century BC. It is a sharper version of the EPIMENIDES PARADOX, "All Cretans are liers... One of their own poets has said so."

see also EPIMENIDES PARADOX, SOCRATES' PARADOX

References

Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, p. 17, 1989.

Euclid's Axioms

see EUCLID'S POSTULATES

Euclid's Elements

see ELEMENTS

Euclid's Fifth Postulate

see EUCLID'S POSTULATES

Euclid Number

The nth Euclid number is defined by

$$E_n \equiv 1 + \prod_{i=1}^{n} p_i,$$

where p_i is the ith PRIME. The first few E_n are 3, 7, 31, 211, 2311, 30031, 510511, 9699691, 223092871, 6469693231, ... (Sloane's A006862). The largest factor of E_n are 3, 7, 31, 211, 2311, 509, 277, 27953, ... (Sloane's A002585). The n of the first few PRIME Euclid numbers E_n are 1, 2, 3, 4, 5, 11, 75, 171, 172, 384, 457, 616, 643, ... (Sloane's A014545) up to a search limit of 700. It is not known if there are an INFINITE number of PRIME Euclid numbers (Guy 1994, Ribenboim 1996).

see also SMARANDACHE SEQUENCES

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, 1994.

Ribenboim, P. *The New Book of Prime Number Records.* New York: Springer-Verlag, 1996.

Sloane, N. J. A. Sequences A014544, A006862/M2698, and A002585/M2697 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 35–37, 1991.

Euclid's Postulates

1. A straight LINE SEGMENT can be drawn joining any two points.
2. Any straight LINE SEGMENT can be extended indefinitely in a straight LINE.
3. Given any straight LINE SEGMENT, a CIRCLE can be drawn having the segment as RADIUS and one endpoint as center.
4. All RIGHT ANGLES are congruent.
5. If two lines are drawn which intersect a third in such a way that the sum of the inner angles on one side is less than two RIGHT ANGLES, then the two lines inevitably must intersect each other on that side if extended far enough. This postulate is equivalent to what is known as the PARALLEL POSTULATE.

Euclid's fifth postulate cannot be proven as a theorem, although this was attempted by many people. Euclid himself used only the first four postulates ("ABSOLUTE GEOMETRY") for the first 28 propositions of the *Elements*, but was forced to invoke the PARALLEL POSTULATE on the 29th. In 1823, Janos Bolyai and Nicolai Lobachevsky independently realized that entirely self-consistent "NON-EUCLIDEAN GEOMETRIES" could be created in which the parallel postulate *did not hold.* (Gauss had also discovered but suppressed the existence of non-Euclidean geometries.)

see also ABSOLUTE GEOMETRY, CIRCLE, *Elements*, LINE SEGMENT, NON-EUCLIDEAN GEOMETRY, PARALLEL POSTULATE, PASCH'S THEOREM, RIGHT ANGLE

References

Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, pp. 88–92, 1989.

Euclid's Principle

see EUCLID'S THEOREMS

Euclid's Theorems

A theorem sometimes called "Euclid's First Theorem" or EUCLID'S PRINCIPLE states that if p is a PRIME and $p|ab$, then $p|a$ or $p|b$ (where $|$ means DIVIDES). A COROLLARY is that $p|a^n \Rightarrow p|a$ (Conway and Guy 1996). The FUNDAMENTAL THEOREM OF ARITHMETIC is another COROLLARY (Hardy and Wright 1979).

Euclid's Second Theorem states that the number of PRIMES is INFINITE. This theorem, also called the INFINITUDE OF PRIMES theorem, was proved by Euclid in Proposition IX.20 of the *Elements.* Ribenboim (1989) gives nine (and a half) proofs of this theorem. Euclid's elegant proof proceeds as follows. Given a finite sequence of consecutive PRIMES 2, 3, 5, ..., p, the number

$$N = 2 \cdot 3 \cdot 5 \cdots p + 1, \tag{1}$$

known as the ith EUCLID NUMBER when $p = p_i$ is the ith PRIME, is either a new PRIME or the product of PRIMES.

If N is a PRIME, then it must be greater than the previous PRIMES, since one plus the product of PRIMES must be greater than each PRIME composing the product. Now, if N is a product of PRIMES, then at least one of the PRIMES must be greater than p. This can be shown as follows. If N is COMPOSITE and not greater than p, then one of its factors (say F) must be one of the PRIMES in the sequence, 2, 3, 5, ..., p. It therefore DIVIDES the product $2 \cdot 3 \cdot 5 \cdots p$. However, since it is a factor of N, it also DIVIDES N. But a number which DIVIDES two numbers a and $b < a$ also DIVIDES their difference $a - b$, so F must also divide

$$N-(2{\cdot}3{\cdot}5\cdots p) = (2{\cdot}3{\cdot}5\cdots p+1)-(2{\cdot}3{\cdot}5\cdots p) = 1. \tag{2}$$

However, in order to divide 1, F must be 1, which is contrary to the assumption that it is a PRIME in the sequence 2, 3, 5, ... It therefore follows that if N is composite, it has at least one factor greater than p. Since N is either a PRIME greater than p or contains a factor greater than p, a PRIME larger than the largest in the finite sequence can always be found, so there are an infinite number of PRIMES. Hardy (1967) remarks that this proof is "as fresh and significant as when it was discovered" so that "two thousand years have not written a wrinkle" on it.

A similar argument shows that $p! \pm 1$ is PRIME, and

$$1 \cdot 3 \cdot 5 \cdot 7 \cdots p + 1 \tag{3}$$

must be either PRIME or be divisible by a PRIME $> p$. Kummer used a variation of this proof, which is also a proof by contradiction. It assumes that there exist only a finite number of PRIMES $N = p_1, p_2, \ldots, p_r$. Now consider $N - 1$. It must be a product of PRIMES, so it has a PRIME divisor p_i in common with N. Therefore, $p_i | N - (N-1) = 1$ which is nonsense, so we have proved the initial assumption is wrong by contradiction.

It is also true that there are runs of COMPOSITE NUMBERS which are arbitrarily long. This can be seen by defining

$$n \equiv j! = \prod_{i=1}^{j} i, \tag{4}$$

where $j!$ is a FACTORIAL. Then the $j - 1$ consecutive numbers $n + 2$, $n + 3$, ..., $n + j$ are COMPOSITE, since

$$n + 2 = (1 \cdot 2 \cdots j) + 2 = 2(1 \cdot 3 \cdot 4 \cdots n + 1) \tag{5}$$

$$n + 3 = (1 \cdot 2 \cdots j) + 3 = 3(1 \cdot 2 \cdot 4 \cdot 5 \cdots n + 1) \tag{6}$$

$$n + j = (1 \cdot 2 \cdots j) + j = j[1 \cdot 2 \cdots (j-1) + 1]. \tag{7}$$

Guy (1981, 1988) points out that while $p_1 p_2 \cdots p_n + 1$ is not necessarily PRIME, letting q be the next PRIME after $p_1 p_2 \cdots p_n + 1$, the number $q - p_1 p_2 \cdots p_n + 1$ is almost always a PRIME, although it has not been proven that this must *always* be the case.

see also DIVIDE, EUCLID NUMBER, PRIME NUMBER

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 60, 1987.

Conway, J. H. and Guy, R. K. "There are Always New Primes!" In *The Book of Numbers.* New York: Springer-Verlag, pp. 133–134, 1996.

Cosgrave, J. B. "A Remark on Euclid's Proof of the Infinitude of Primes." *Amer. Math. Monthly* **96**, 339–341, 1989.

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 22, 1996.

Dunham, W. "Great Theorem: The Infinitude of Primes." *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 73–75, 1990.

Guy, R. K. §A12 in *Unsolved Problems in Number Theory.* New York: Springer-Verlag, 1981.

Guy, R. K. "The Strong Law of Small Numbers." *Amer. Math. Monthly* **95**, 697–712, 1988.

Hardy, G. H. *A Mathematician's Apology.* Cambridge, England: Cambridge University Press, 1992.

Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, pp. 3–12, 1989.

Euclidean Algorithm

An ALGORITHM for finding the GREATEST COMMON DIVISOR of two numbers a and b, also called Euclid's algorithm. It is an example of a P-PROBLEM whose time complexity is bounded by a quadratic function of the length of the input values (Banach and Shallit). Let $a = bq + r$, then find a number u which DIVIDES both a and b (so that $a = su$ and $b = tu$), then u also DIVIDES r since

$$r = a - bq = su - qtu = (s - qt)u. \tag{1}$$

Similarly, find a number v which DIVIDES b and r (so that $b = s'v$ and $r = t'v$), then v DIVIDES a since

$$a = bq + r = s'vq + t'v = (s'q + t')v. \tag{2}$$

Therefore, every common DIVISOR of a and b is a common DIVISOR of b and r, so the procedure can be iterated as follows

$$a = bq_1 + r_1 \tag{3}$$

$$b = q_2 r_1 + r_2 \tag{4}$$

$$r_1 = q_3 r_2 + r_3 \tag{5}$$

$$r_{n-2} = q_n r_{n-1} + r_n \tag{6}$$

$$r_{n-1} = q_{n+1} r_n, \tag{7}$$

where r_n is $\mathrm{GCD}(a, b) \equiv (a, b)$. Lamé showed that the number of steps needed to arrive at the GREATEST COMMON DIVISOR for two numbers less than N is

$$\text{steps} \leq \frac{\log_{10} N}{\log_{10} \phi} + \frac{\log_{10} \sqrt{5}}{\log_{10} \phi} \tag{8}$$

where ϕ is the GOLDEN MEAN, or ≤ 5 times the number of digits in the smaller number. Numerically, Lamé's expression evaluates to

$$\text{steps} \leq 4.785 \log_{10} N + 1.6723. \tag{9}$$

As shown by LAMÉ'S THEOREM, the worst case occurs when the ALGORITHM is applied to two consecutive FIBONACCI NUMBERS. Heilbronn showed that the average number of steps is $12 \ln 2/\pi^2 \log_{10} n = 0.843 \log_{10} n$ for all pairs (n, b) with $b < n$. Kronecker showed that the shortest application of the ALGORITHM uses least absolute remainders. The QUOTIENTS obtained are distributed as shown in the following table (Wagon 1991).

Quotient	%
1	41.5
2	17.0
3	9.3

For details, see Uspensky and Heaslet (1939) or Knuth (1973). Let $T(m,n)$ be the number of divisions required to compute $\mathrm{GCD}(m,n)$ using the Euclidean algorithm, and define $T(m,0) = 0$ if $m \geq 0$. Then

$$T(m,n) = \begin{cases} 1 + T(n, m \bmod n) & \text{for } m \geq n \\ 1 + T(n,m) & \text{for } m < n. \end{cases} \tag{10}$$

Define the functions

$$T(n) = \frac{1}{n} \sum_{0 \leq m < n} T(m,n) \tag{11}$$

$$\tau(n) = \frac{1}{\phi(n)} \sum_{\substack{0 \leq m < n \\ \mathrm{GCD}(m,n)=1}} T(m,n) \tag{12}$$

$$A(N) = \frac{1}{N^2} \sum_{\substack{1 \leq m \leq N \\ 1 \leq n \leq N}} T(m,n), \tag{13}$$

where ϕ is the TOTIENT FUNCTION, $T(n)$ is the average number of divisions when n is fixed and m chosen at random, $\tau(n)$ is the average number of divisions when n is fixed and m is a random number coprime to n, and $A(N)$ is the average number of divisions when m and n are both chosen at random in $[1, N]$. Norton (1990) showed that

$$T(n) = \frac{12 \ln 2}{\pi^2} \left[\ln n - \sum_{d|n} \frac{\Lambda(d)}{d} \right] + C + \frac{1}{n} \sum_{d|n} \phi(d) \mathcal{O}(d^{-1/6+\epsilon}), \tag{14}$$

where Λ is the VON MANGOLDT FUNCTION and C is PORTER'S CONSTANT. Porter (1975) showed that

$$\tau(n) = \frac{12 \ln 2}{\pi^2} \ln n + C + \mathcal{O}(n^{-1/6} + \epsilon), \tag{15}$$

and Norton (1990) proved that

$$A(N) = \frac{12 \ln 2}{\pi^2} \left[\ln N - \tfrac{1}{2} + \frac{6}{\pi^2} \zeta'(2) \right] + C - \tfrac{1}{2} + \mathcal{O}(N^{-1/6+\epsilon}). \tag{16}$$

There exist 22 QUADRATIC FIELDS in which there is a Euclidean algorithm (Inkeri 1947).

see also FERGUSON-FORCADE ALGORITHM

References

Bach, E. and Shallit, J. *Algorithmic Number Theory, Vol. 1: Efficient Algorithms.* Cambridge, MA: MIT Press, 1996.

Courant, R. and Robbins, H. "The Euclidean Algorithm." §2.4 in Supplement to Ch. 1 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 42–51, 1996.

Dunham, W. *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 69–70, 1990.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/porter/porter.html`.

Inkeri, K. "Über den Euklidischen Algorithmus in quadratischen Zahlkörpern." *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys.* **1947**, 1–35, 1947.

Knuth, D. E. *The Art of Computer Programming, Vol. 1: Fundamental Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, 1973.

Knuth, D. E. *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, 1981.

Norton, G. H. "On the Asymptotic Analysis of the Euclidean Algorithm." *J. Symb. Comput.* **10**, 53–58, 1990.

Porter, J. W. "On a Theorem of Heilbronn." *Mathematika* **22**, 20–28, 1975.

Uspensky, J. V. and Heaslet, M. A. *Elementary Number Theory.* New York: McGraw-Hill, 1939.

Wagon, S. "The Ancient and Modern Euclidean Algorithm" and "The Extended Euclidean Algorithm." §8.1 and 8.2 in *Mathematica in Action.* New York: W. H. Freeman, pp. 247–252 and 252–256, 1991.

Euclidean Construction

see GEOMETRIC CONSTRUCTION

Euclidean Geometry

A GEOMETRY in which EUCLID'S FIFTH POSTULATE holds, sometimes also called PARABOLIC GEOMETRY. 2-D Euclidean geometry is called PLANE GEOMETRY, and 3-D Euclidean geometry is called SOLID GEOMETRY. Hilbert proved the CONSISTENCY of Euclidean geometry.

see also ELLIPTIC GEOMETRY, GEOMETRIC CONSTRUCTION, GEOMETRY, HYPERBOLIC GEOMETRY, NON-EUCLIDEAN GEOMETRY, PLANE GEOMETRY

References

Altshiller-Court, N. *College Geometry: A Second Course in Plane Geometry for Colleges and Normal Schools, 2nd ed., rev. enl.* New York: Barnes and Noble, 1952.

Casey, J. *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions with Numerous Examples, 2nd rev. enl. ed.* Dublin: Hodges, Figgis, & Co., 1893.

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., 1967
Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, 1969.
Gallatly, W. *The Modern Geometry of the Triangle, 2nd ed.* London: Hodgson, 1913.
Heath, T. L. *The Thirteen Books of the Elements, 2nd ed., Vol. 1: Books I and II.* New York: Dover, 1956.
Heath, T. L. *The Thirteen Books of the Elements, 2nd ed., Vol. 2: Books III–IX.* New York: Dover, 1956.
Heath, T. L. *The Thirteen Books of the Elements, 2nd ed., Vol. 3: Books X–XIII.* New York: Dover, 1956.
Honsberger, R. *Episodes in Nineteenth and Twentieth Century Euclidean Geometry.* Washington, DC: Math. Assoc. Amer., 1995.
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, 1929.
Johnson, R. A. *Advanced Euclidean Geometry.* New York: Dover, 1960.
Klee, V. "Some Unsolved Problems in Plane Geometry." *Math. Mag.* **52**, 131–145, 1979.
Klee, V. and Wagon, S. *Old and New Unsolved Problems in Plane Geometry and Number Theory, rev. ed.* Washington, DC: Math. Assoc. Amer., 1991.

Euclidean Group

The GROUP of ROTATIONS and TRANSLATIONS.

see also ROTATION, TRANSLATION

References

Lomont, J. S. *Applications of Finite Groups.* New York: Dover, 1987.

Euclidean Metric

The FUNCTION $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ that assigns to any two VECTORS $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ the number

$$\sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2},$$

and so gives the "standard" distance between any two VECTORS in $\mathbb{R}^n$.

Euclidean Motion

A Euclidean motion of $\mathbb{R}^n$ is an AFFINE TRANSFORMATION whose linear part is an ORTHOGONAL TRANSFORMATION.

see also RIGID MOTION

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 105, 1993.

Euclidean Norm

see L_2-NORM

Euclidean Number

A Euclidean number is a number which can be obtained by repeatedly solving the QUADRATIC EQUATION. Euclidean numbers, together with the RATIONAL NUMBERS, can be constructed using classical GEOMETRIC CONSTRUCTIONS. However, the cases for which the values of the TRIGONOMETRIC FUNCTIONS SINE, COSINE, TANGENT, etc., can be written in closed form involving square roots of REAL NUMBERS are much more restricted.

see also ALGEBRAIC INTEGER, ALGEBRAIC NUMBER, CONSTRUCTIBLE NUMBER, RADICAL INTEGER

References

Conway, J. H. and Guy, R. K. "Three Greek Problems." In *The Book of Numbers.* New York: Springer-Verlag, pp. 192–194, 1996.

Euclidean Plane

The 2-D EUCLIDEAN SPACE denoted $\mathbb{R}^2$.

see also COMPLEX PLANE, EUCLIDEAN SPACE

Euclidean Space

Euclidean n-space is the SPACE of all n-tuples of REAL NUMBERS, $(x_1, x_2, \ldots, x_n)$ and is denoted $\mathbb{R}^n$. $\mathbb{R}^n$ is a VECTOR SPACE and has LEBESGUE COVERING DIMENSION n. Elements of $\mathbb{R}^n$ are called n-VECTORS. $\mathbb{R}^1 = \mathbb{R}$ is the set of REAL NUMBERS (i.e., the REAL LINE), and $\mathbb{R}^2$ is called the EUCLIDEAN PLANE. In Euclidean space, COVARIANT and CONTRAVARIANT quantities are equivalent so $\vec{e}^j = \vec{e}_j$.

see also EUCLIDEAN PLANE, REAL LINE, VECTOR

References

Gray, A. "Euclidean Spaces." §1.1 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 2–4, 1993.

Eudoxus's Kampyle

see KAMPYLE OF EUDOXUS

Euler's $6n + 1$ Theorem

Every PRIME of the form $6n + 1$ can be written in the form $x^2 + 3y^2$.

Euler's Addition Theorem

Let $g(x) \equiv (1 - x^2)(1 - k^2x^2)$. Then

$$\int_0^a \frac{dx}{\sqrt{g(x)}} + \int_0^b \frac{dx}{\sqrt{g(x)}} = \int_0^c \frac{dx}{\sqrt{g(x)}},$$

where

$$c \equiv \frac{b\sqrt{g(a)} + a\sqrt{g(b)}}{\sqrt{1 - k^2a^2b^2}}.$$

Euler Angles

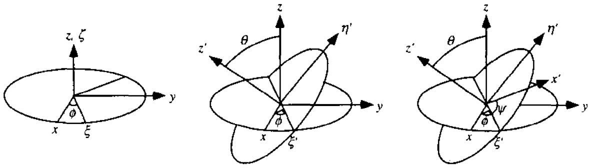

According to EULER'S ROTATION THEOREM, any ROTATION may be described using three ANGLES. If the ROTATIONS are written in terms of ROTATION MATRICES B, C, and D, then a general ROTATION A can be written as

$$\mathsf{A} = \mathsf{BCD}. \tag{1}$$

The three angles giving the three rotation matrices are called Euler angles. There are several conventions for Euler angles, depending on the axes about which the rotations are carried out. Write the MATRIX A as

$$\mathsf{A} \equiv \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \end{bmatrix}. \tag{2}$$

In the so-called "x-convention," illustrated above,

$$\mathsf{D} \equiv \begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

$$\mathsf{C} \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \tag{4}$$

$$\mathsf{B} \equiv \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{5}$$

so

$$\begin{aligned} a_{11} &= \cos\psi\cos\phi - \cos\theta\sin\phi\sin\psi \\ a_{12} &= \cos\psi\sin\phi + \cos\theta\cos\phi\sin\psi \\ a_{13} &= \sin\psi\sin\theta \\ a_{21} &= -\sin\psi\cos\phi - \cos\theta\sin\phi\cos\psi \\ a_{22} &= -\sin\psi\sin\phi + \cos\theta\cos\phi\cos\psi \\ a_{23} &= \cos\psi\sin\theta \\ a_{31} &= \sin\theta\sin\phi \\ a_{32} &= -\sin\theta\cos\phi \\ a_{33} &= \cos\theta \end{aligned}$$

To obtain the components of the ANGULAR VELOCITY $\boldsymbol{\omega}$ in the body axes, note that for a MATRIX

$$\mathsf{A} \equiv [\,\mathbf{A}_1 \quad \mathbf{A}_2 \quad \mathbf{A}_3\,], \tag{6}$$

it is true that

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} = \begin{bmatrix} a_{11}\omega_x + a_{12}\omega_y + a_{13}\omega_z \\ a_{21}\omega_x + a_{22}\omega_y + a_{23}\omega_z \\ a_{31}\omega_x + a_{32}\omega_y + a_{33}\omega_z \end{bmatrix} \tag{7}$$

$$= \mathbf{A}_1\omega_x + \mathbf{A}_2\omega_y + \mathbf{A}_3\omega_z. \tag{8}$$

Now, ω_z corresponds to rotation about the ϕ axis, so look at the ω_z component of $\mathsf{A}\boldsymbol{\omega}$,

$$\boldsymbol{\omega}_\phi = \mathbf{A}_1\omega_z = \begin{bmatrix} \sin\psi\sin\theta \\ \cos\psi\sin\theta \\ \cos\theta \end{bmatrix} \dot\phi. \tag{9}$$

The line of nodes corresponds to a rotation by θ about the ξ-axis, so look at the ω_ξ component of $\mathsf{B}\boldsymbol{\omega}$,

$$\boldsymbol{\omega}_\theta = \mathbf{B}_1\omega_\xi = \mathbf{B}_1\dot\theta = \begin{bmatrix} \cos\psi \\ -\sin\psi \\ 0 \end{bmatrix} \dot\theta. \tag{10}$$

Similarly, to find rotation by ψ about the remaining axis, look at the ω_ψ component of $\mathsf{B}\boldsymbol{\omega}$,

$$\boldsymbol{\omega}_\psi = \mathbf{B}_3\omega_\psi = \mathbf{B}_3\dot\psi = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \dot\psi. \tag{11}$$

Combining the pieces gives

$$\boldsymbol{\omega} = \begin{bmatrix} \sin\psi\sin\theta\dot\phi + \cos\psi\dot\theta \\ \cos\psi\sin\theta\dot\phi - \sin\psi \\ \cos\theta\dot\phi + \dot\psi. \end{bmatrix} \tag{12}$$

For more details, see Goldstein (1980, p. 176) and Landau and Lifschitz (1976, p. 111).

The x-convention Euler angles are given in terms of the CAYLEY-KLEIN PARAMETERS by

$$\begin{aligned} \phi &= -2i\ln\left(\pm\frac{\alpha^{1/2}\gamma^{1/4}}{\beta^{1/4}(1+\beta\gamma)^{1/4}}\right), \\ &\quad -2i\ln\left(\pm\frac{i\alpha^{1/2}\gamma^{1/4}}{\beta^{1/4}(1+\beta\gamma)^{1/4}}\right) \end{aligned} \tag{13}$$

$$\begin{aligned} \psi &= -2i\ln\left(\pm\frac{\alpha^{1/2}\beta^{1/4}}{\gamma^{1/4}(1+\beta\gamma)^{1/4}}\right), \\ &\quad -2i\ln\left(\pm\frac{i\alpha^{1/2}\beta^{1/4}}{\gamma^{1/4}(1+\beta\gamma)^{1/4}}\right) \end{aligned} \tag{14}$$

$$\theta = \pm 2\cos^{-1}(\pm\sqrt{1+\beta\gamma}\,). \tag{15}$$

In the "y-convention,"

$$\phi_x \equiv \phi_y + \tfrac{1}{2}\pi \tag{16}$$

$$\psi_x \equiv \psi_y - \tfrac{1}{2}\pi. \tag{17}$$

Therefore,

$$\sin\phi_x = \cos\phi_y \tag{18}$$

$$\cos\phi_x = -\sin\phi_y \tag{19}$$

$$\sin\psi_x = -\cos\psi_y \tag{20}$$

$$\cos\psi_x = \sin\psi_y \tag{21}$$

$$\mathsf{D} \equiv \begin{bmatrix} -\sin\phi & \cos\phi & 0 \\ -\cos\phi & -\sin\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{22}$$

$$\mathsf{C} \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \tag{23}$$

$$\mathsf{B} \equiv \begin{bmatrix} \sin\psi & -\cos\psi & 0 \\ \cos\psi & \sin\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{24}$$

and A is given by

$$\begin{aligned} a_{11} &= -\sin\psi\sin\phi + \cos\theta\cos\phi\cos\psi \\ a_{12} &= \sin\psi\cos\phi + \cos\theta\sin\phi\cos\psi \\ a_{13} &= -\cos\psi\sin\theta \\ a_{21} &= -\cos\psi\sin\phi - \cos\theta\cos\phi\sin\psi \\ a_{22} &= \cos\psi\cos\phi - \cos\theta\sin\phi\sin\psi \\ a_{23} &= \sin\psi\sin\theta \\ a_{31} &= \sin\theta\cos\phi \\ a_{32} &= \sin\theta\sin\phi \\ a_{33} &= \cos\theta. \end{aligned}$$

In the "xyz" (pitch-roll-yaw) convention, θ is pitch, ψ is roll, and ϕ is yaw.

$$\mathsf{D} \equiv \begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{25}$$

$$\mathsf{C} \equiv \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \tag{26}$$

$$\mathsf{B} \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & \sin\psi \\ 0 & -\sin\psi & \cos\psi \end{bmatrix} \tag{27}$$

and A is given by

$$\begin{aligned} a_{11} &= \cos\theta\cos\phi \\ a_{12} &= \cos\theta\sin\phi \\ a_{13} &= -\sin\theta \\ a_{21} &= \sin\psi\sin\theta\cos\phi - \cos\psi\sin\phi \\ a_{22} &= \sin\psi\sin\theta\sin\phi + \cos\psi\cos\phi \\ a_{23} &= \cos\theta\sin\psi \\ a_{31} &= \cos\psi\sin\theta\cos\phi + \sin\psi\sin\phi \\ a_{32} &= \cos\psi\sin\theta\sin\phi - \sin\psi\cos\phi \\ a_{33} &= \cos\theta\cos\psi. \end{aligned}$$

A set of parameters sometimes used instead of angles are the Euler Parameters e_0, e_1, e_2 and e_3, defined by

$$e_0 \equiv \cos\left(\frac{\phi}{2}\right) \tag{28}$$

$$\mathbf{e} \equiv \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} = \hat{\mathbf{n}}\sin\left(\frac{\phi}{2}\right). \tag{29}$$

Using Euler Parameters (which are Quaternions), an arbitrary Rotation Matrix can be described by

$$\begin{aligned} a_{11} &= {e_0}^2 + {e_1}^2 - {e_2}^2 - {e_3}^2 \\ a_{12} &= 2(e_1e_2 + e_0e_3) \\ a_{13} &= 2(e_1e_3 - e_0e_2) \\ a_{21} &= 2(e_1e_2 - e_0e_3) \\ a_{22} &= {e_0}^2 - {e_1}^2 + {e_2}^2 - {e_3}^2 \\ a_{23} &= 2(e_2e_3 + e_0e_1) \\ a_{31} &= 2(e_1e_3 + e_0e_2) \\ a_{32} &= 2(e_2e_3 - e_0e_1) \\ a_{33} &= {e_0}^2 - {e_1}^2 - {e_2}^2 + {e_3}^2 \end{aligned}$$

(Goldstein 1960, p. 153).

If the coordinates of two pairs of n points $\mathbf{x}_i$ and $\mathbf{x}'_i$ are known, one rotated with respect to the other, then the Euler rotation matrix can be obtained in a straightforward manner using Least Squares Fitting. Write the points as arrays of vectors, so

$$[\mathbf{x}'_1 \quad \cdots \quad \mathbf{x}'_n] = \mathsf{A}[\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n]. \tag{30}$$

Writing the arrays of vectors as matrices gives

$$\mathsf{X}' = \mathsf{A}\mathsf{X} \tag{31}$$

$$\mathsf{X}'\mathsf{X}^{\mathrm{T}} = \mathsf{A}\mathsf{X}\mathsf{X}^{\mathrm{T}}, \tag{32}$$

and solving for A gives

$$\mathsf{A} = \mathsf{X}'\mathsf{X}^{\mathrm{T}}(\mathsf{X}\mathsf{X}^{\mathrm{T}})^{-1}. \tag{33}$$

However, we want the angles θ, ϕ, and ψ, not their combinations contained in the Matrix A. Therefore, write the 3×3 Matrix

$$\mathsf{A} = \begin{bmatrix} f_1(\theta,\phi,\psi) & f_2(\theta,\phi,\psi) & f_3(\theta,\phi,\psi) \\ f_4(\theta,\phi,\psi) & f_5(\theta,\phi,\psi) & f_6(\theta,\phi,\psi) \\ f_7(\theta,\phi,\psi) & f_7(\theta,\phi,\psi) & f_9(\theta,\phi,\psi) \end{bmatrix} \tag{34}$$

as a 1×9 Vector

$$\mathbf{f} = \begin{bmatrix} f_1(\theta,\phi,\psi) \\ \vdots \\ f_9(\theta,\phi,\psi) \end{bmatrix}. \tag{35}$$

Now set up the matrices

$$\begin{bmatrix} \left.\frac{\partial f_1}{\partial\theta}\right|_{\theta_i,\phi_i,\psi_i} & \left.\frac{\partial f_1}{\partial\phi}\right|_{\theta_i,\phi_i,\psi_i} & \left.\frac{\partial f_1}{\partial\psi}\right|_{\theta_i,\phi_i,\psi_i} \\ \vdots & \vdots & \vdots \\ \left.\frac{\partial f_9}{\partial\theta}\right|_{\theta_i,\phi_i,\psi_i} & \left.\frac{\partial f_9}{\partial\phi}\right|_{\theta_i,\phi_i,\psi_i} & \left.\frac{\partial f_9}{\partial\psi}\right|_{\theta_i,\phi_i,\psi_i} \end{bmatrix} \begin{bmatrix} d\theta \\ d\phi \\ d\psi \end{bmatrix} = d\mathbf{f}. \tag{36}$$

Using Nonlinear Least Squares Fitting then gives solutions which converge to (θ, ϕ, ψ).

see also CAYLEY-KLEIN PARAMETERS, EULER PARAMETERS, EULER'S ROTATION THEOREM, INFINITESIMAL ROTATION, QUATERNION, ROTATION, ROTATION MATRIX

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 198–200, 1985.

Goldstein, H. "The Euler Angles" and "Euler Angles in Alternate Conventions." §4-4 and Appendix B in *Classical Mechanics, 2nd ed.* Reading, MA: Addison-Wesley, pp. 143–148 and 606–610, 1980.

Landau, L. D. and Lifschitz, E. M. *Mechanics, 3rd ed.* Oxford, England: Pergamon Press, 1976.

Euler-Bernoulli Triangle

see SEIDEL-ENTRINGER-ARNOLD TRIANGLE

Euler Brick

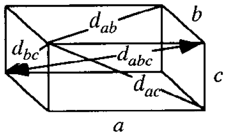

A RECTANGULAR PARALLELEPIPED ("BRICK") with integral edges $a > b > c$ and face diagonals d_{ij} given by

$$d_{ab} = \sqrt{a^2 + b^2} \quad (1)$$

$$d_{ac} = \sqrt{a^2 + c^2} \quad (2)$$

$$d_{bc} = \sqrt{b^2 + c^2}. \quad (3)$$

The problem is also called the BRICK, DIAGONALS PROBLEM, PERFECT BOX, PERFECT CUBOID, or RATIONAL CUBOID problem.

Euler found the smallest solution, which has sides $a = 240$, $b = 117$, and $c = 44$ and face DIAGONALS $d_{ab} = 267$, $d_{ac} = 244$, and $d_{bc} = 125$. Kraitchik gave 257 cuboids with the ODD edge less than 1 million (Guy 1994, p. 174). F. Helenius has compiled a list of the 5003 smallest (measured by the longest edge) Euler bricks. The first few are (240, 117, 44), (275, 252, 240), (693, 480, 140), (720, 132, 85), (792, 231, 160), ... (Sloane's A031173, A031174, and A031175). Parametric solutions for Euler bricks are also known.

No solution is known in which the oblique SPACE DIAGONAL

$$d_{abc} = \sqrt{a^2 + b^2 + c^2} \quad (4)$$

is also an INTEGER. If such a brick exists, the smallest side must be at least 1,281,000,000 (R. Rathbun 1996). Such a solution is equivalent to solving the DIOPHANTINE EQUATIONS

$$A^2 + B^2 = C^2 \quad (5)$$

$$A^2 + D^2 = E^2 \quad (6)$$

$$B^2 + D^2 = F^2 \quad (7)$$

$$B^2 + E^2 = G^2. \quad (8)$$

A solution with integral SPACE DIAGONAL and two out of three face diagonals is $a = 672$, $b = 153$, and $c = 104$, giving $d_{ab} = 3\sqrt{52777}$, $d_{ac} = 680$, $d_{bc} = 185$, and $d_{abc} = 697$. A solution giving integral space and face diagonals with only a single nonintegral EDGE is $a = 18720$, $b = \sqrt{211773121}$, and $c = 7800$, giving $d_{ab} = 23711$, $d_{ac} = 20280$, $d_{bc} = 16511$, and $d_{abc} = 24961$.

see also CUBOID, CYCLIC QUADRILATERAL, DIAGONAL (POLYHEDRON), PARALLELEPIPED, PYTHAGOREAN QUADRUPLE

References

Guy, R. K. "Is There a Perfect Cuboid? Four Squares whose Sums in Pairs are Square. Four Squares whose Differences are Square." §D18 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 173–181, 1994.

Helenius, F. First 1000 Primitive Euler Bricks. `notebooks/EulerBricks.dat`.

Leech, J. "The Rational Cuboid Revisited." *Amer. Math. Monthly* **84**, 518–533, 1977. Erratum in *Amer. Math. Monthly* **85**, 472, 1978.

Sloane, N. J. A. Sequences A031173, A031174, and A031175 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Rathbun, R. L. Personal communication, 1996.

Spohn, W. G. "On the Integral Cuboid." *Amer. Math. Monthly* **79**, 57–59, 1972.

Spohn, W. G. "On the Derived Cuboid." *Canad. Math. Bull.* **17**, 575–577, 1974.

Wells, D. G. *The Penguin Dictionary of Curious and Interesting Numbers.* London: Penguin, p. 127, 1986.

Euler Chain

A CHAIN (GRAPH) whose EDGES consist of all graph EDGES.

Euler Characteristic

Let a closed surface have GENUS g. Then the POLYHEDRAL FORMULA becomes the POINCARÉ FORMULA

$$\chi \equiv V - E + F = 2 - 2g, \quad (1)$$

where χ is the Euler characteristic, sometimes also known as the EULER-POINCARÉ CHARACTERISTIC. In terms of the INTEGRAL CURVATURE of the surface K,

$$\iint K\,da = 2\pi\chi. \quad (2)$$

The Euler characteristic is sometimes also called the EULER NUMBER. It can also be expressed as

$$\chi = p_0 - p_1 + p_2, \quad (3)$$

where p_i is the ith BETTI NUMBER of the space.

see also CHROMATIC NUMBER, MAP COLORING

Euler's Circle

see NINE-POINT CIRCLE

Euler's Conjecture

$$g(k) = 2^k + \left\lfloor \left(\tfrac{3}{2}\right)^k \right\rfloor - 2,$$

where $g(k)$ is the quantity appearing in WARING'S PROBLEM, and $\lfloor x \rfloor$ is the FLOOR FUNCTION.

see also WARING'S PROBLEM

Euler Constant

see e, EULER-MASCHERONI CONSTANT, MACLAURIN-CAUCHY THEOREM

Euler's Criterion

Let $p = 2m + 1$ be an ODD PRIME and a a POSITIVE INTEGER with $p \nmid a$. Then

$$a^m \equiv 1 \pmod{p} \tag{1}$$

IFF there exists an INTEGER t such that

$$p \equiv t^2 \pmod{p}. \tag{2}$$

In other words,

$$a^{(p-1)/2} \equiv \frac{a}{p} \pmod{p}, \tag{3}$$

where (a/p) is the LEGENDRE SYMBOL.

see also QUADRATIC RESIDUE

References

Rosen, K. H. Ch. 9 in *Elementary Number Theory and Its Applications, 3rd ed.* Reading, MA: Addison-Wesley, 1993.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 33–37, 1993.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, p. 293, 1991.

Euler Curvature Formula

$$\kappa = \kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta,$$

where κ is the normal CURVATURE in a direction making an ANGLE θ with the first principle direction.

Euler Differential Equation

The general nonhomogeneous equation is

$$x^2 \frac{d^2y}{dx^2} + \alpha x \frac{dy}{dx} + \beta y = S(x). \tag{1}$$

The homogeneous equation is

$$x^2 y'' + \alpha x y' + \beta y = 0 \tag{2}$$

$$y'' + \frac{\alpha}{x} y' + \frac{\beta}{x^2} y = 0. \tag{3}$$

Now attempt to convert the equation from

$$y'' + p(x)y' + q(x)y = 0 \tag{4}$$

to one with constant COEFFICIENTS

$$\frac{d^2y}{dz^2} + A\frac{dy}{dz} + By = 0 \tag{5}$$

by using the standard transformation for linear SECOND-ORDER ORDINARY DIFFERENTIAL EQUATIONS. Comparing (3) and (5), the functions $p(x)$ and $q(x)$ are

$$p(x) \equiv \frac{\alpha}{x} = \alpha x^{-1} \tag{6}$$

$$q(x) \equiv \frac{\beta}{x^2} = \beta x^{-2}. \tag{7}$$

Let $B \equiv \beta$ and define

$$\begin{aligned} z &\equiv B^{-1/2} \int \sqrt{q(x)}\,dx = \beta^{-1/2} \int \sqrt{\beta x^{-2}}\,dx \\ &= \int x^{-1}\,dx = \ln x. \end{aligned} \tag{8}$$

Then A is given by

$$\begin{aligned} A &\equiv \frac{q'(x) + 2p(x)q(x)}{2[q(x)]^{3/2}} B^{1/2} \\ &= \frac{-2\beta x^{-3} + 2(\alpha x^{-1})(\beta x^{-2})}{2(\beta x^{-2})^{3/2}} \beta^{1/2} \\ &= \alpha - 1, \end{aligned} \tag{9}$$

which is a constant. Therefore, the equation becomes a second-order ODE with constant COEFFICIENTS

$$\frac{d^2y}{dz^2} + (\alpha - 1)\frac{dy}{dz} + \beta y = 0. \tag{10}$$

Define

$$\begin{aligned} r_1 &\equiv \tfrac{1}{2}\left(-A + \sqrt{A^2 - 4B}\right) \\ &= \tfrac{1}{2}\left[1 - \alpha + \sqrt{(\alpha-1)^2 - 4\beta}\right] \end{aligned} \tag{11}$$

$$\begin{aligned} r_2 &\equiv \tfrac{1}{2}\left(-A - \sqrt{A^2 - 4B}\right) \\ &= \tfrac{1}{2}\left[1 - \alpha - \sqrt{(\alpha-1)^2 - 4\beta}\right] \end{aligned} \tag{12}$$

and

$$a \equiv \tfrac{1}{2}(1 - \alpha) \tag{13}$$

$$b \equiv \tfrac{1}{2}\sqrt{4\beta - (\alpha - 1)^2}. \tag{14}$$

The solutions are

$$y = \begin{cases} c_1 e^{r_1 z} + c_2 e^{r_2 z} & (\alpha-1)^2 > 4\beta \\ (c_1 + c_2 z)e^{az} & (\alpha-1)^2 = 4\beta \\ e^{az}[c_1\cos(bz) + c_2\sin(bz)] & (\alpha-1)^2 < 4\beta. \end{cases} \tag{15}$$

In terms of the original variable x,

$$y = \begin{cases} c_1|x|^{r_1} + c_2|x|^{r_2} & (\alpha-1)^2 > 4\beta \\ (c_1 + c_2\ln|x|)|x|^a & (\alpha-1)^2 = 4\beta \\ |x|^a[c_1\cos(b\ln|x|) + c_2\sin(b\ln|x|)] & (\alpha-1)^2 < 4\beta. \end{cases} \tag{16}$$

Euler's Displacement Theorem

The general displacement of a rigid body (or coordinate frame) with one point fixed is a ROTATION about some axis. Furthermore, a ROTATION may be described in any basis using three ANGLES.

see also EUCLIDEAN MOTION, EULER ANGLES, RIGID MOTION, ROTATION

Euler's Distribution Theorem

For signed distances,

$$\overline{AB}\cdot\overline{CD}+\overline{AC}\cdot\overline{DB}+\overline{AD}\cdot\overline{BC}=0,$$

since

$$(b-a)(d-c)+(c-a)(b-d)+(d-a)(c-b)=0.$$

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, p. 3, 1929.

Euler Equation

see also EULER DIFFERENTIAL EQUATION, EULER FORMULA, EULER-LAGRANGE DIFFERENTIAL EQUATION

Euler's Factorization Method

Works by expressing N as a QUADRATIC FORM in two different ways. Then

$$N=a^2+b^2=c^2+d^2, \tag{1}$$

so

$$a^2-c^2=d^2-b^2 \tag{2}$$

$$(a-c)(a+c)=(d-b)(d+b). \tag{3}$$

Let k be the GREATEST COMMON DIVISOR of $a-c$ and $d-b$ so

$$a-c=kl \tag{4}$$

$$d-b=km \tag{5}$$

$$(l,m)=1, \tag{6}$$

(where (l,m) denotes the GREATEST COMMON DIVISOR of l and m), and

$$l(a+c)=m(d+b). \tag{7}$$

But since $(l,m)=1$, $m|a+c$ and

$$a+c=mn, \tag{8}$$

which gives

$$b+d=ln, \tag{9}$$

so we have

$$\begin{aligned}[(\tfrac{1}{2}k)^2+(\tfrac{1}{2}n)^2](l^2+m^2)&=\tfrac{1}{4}(k^2+n^2)(l^2+m^2)\\&=\tfrac{1}{4}[(kn)^2+(kl)^2+(nm)^2+(nl)^2]\\&=\tfrac{1}{4}[(d-b)^2+(a-c)^2+(a+c)^2+(d+b)^2]\\&=\tfrac{1}{4}(2a^2+2b^2+2c^2+2d^2)\\&=\tfrac{1}{4}(2N+2N)=N.\end{aligned} \tag{10}$$

see also PRIME FACTORIZATION ALGORITHMS

Euler's Finite Difference Transformation

A transformation for the acceleration of the convergence of slowly converging ALTERNATING SERIES,

$$\sum_{k=0}^{\infty}(-1)^k a_k=\sum_{n=0}^{\infty}\frac{\Delta^k a_0}{2^{n+1}}.$$

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1163, 1980.

Euler Formula

The Euler formula states

$$e^{ix}=\cos x+i\sin x, \tag{1}$$

where i is the IMAGINARY NUMBER. Note that the EULER POLYHEDRAL FORMULA is sometimes also called the Euler formula, as is the EULER CURVATURE FORMULA. The equivalent expression

$$ix=\ln(\cos x+i\sin x) \tag{2}$$

had previously been published by Cotes (1714). The special case of the formula with $x=\pi$ gives the beautiful identity

$$e^{i\pi}+1=0, \tag{3}$$

an equation connecting the fundamental numbers i, PI, e, 1, and 0 (ZERO).

The Euler formula can be demonstrated using a series expansion

$$\begin{aligned}e^{ix}&=\sum_{n=0}^{\infty}\frac{(ix)^n}{n!}\\&=\sum_{n=0}^{\infty}\frac{(-1)^n x^{2n}}{(2n)!}+i\sum_{n=1}^{\infty}\frac{(-1)^{n-1}x^{2n-1}}{(2n-1)!}\\&=\cos x+i\sin x.\end{aligned} \tag{4}$$

It can also be proven using a COMPLEX integral. Let

$$z\equiv\cos\theta+i\sin\theta \tag{5}$$

$$dz=(-\sin\theta+i\cos\theta)\,d\theta=i(\cos\theta+i\sin\theta)\,d\theta=iz\,d\theta \tag{6}$$

$$\int\frac{dz}{z}=\int i\,d\theta \tag{7}$$

$$\ln z=i\theta, \tag{8}$$

so

$$z=e^{i\theta}\equiv\cos\theta+i\sin\theta. \tag{9}$$

see also DE MOIVRE'S IDENTITY, EULER POLYHEDRAL FORMULA

References

Castellanos, D. "The Ubiquitous Pi. Part I." *Math. Mag.* **61**, 67–98, 1988.

Conway, J. H. and Guy, R. K. "Euler's Wonderful Relation." *The Book of Numbers.* New York: Springer-Verlag, pp. 254-256, 1996.

Cotes, R. *Philosophical Transactions* **29**, 32, 1714.

Euler, L. *Miscellanea Berolinensia* **7**, 179, 1743.

Euler, L. *Introductio in Analysin Infinitorum, Vol. 1.* Lausanne, p. 104, 1748.

Euler Four-Square Identity

The amazing polynomial identity

$$\begin{aligned}(a_1{}^2+a_2{}^2+a_3{}^2+a_4{}^2)&(b_1{}^2+b_2{}^2+b_3{}^2+b_4{}^2)\\ &=(a_1b_1-a_2b_2-a_3b_3-a_4b_4)^2\\ &+(a_1b_2+a_2b_1+a_3b_4-a_4b_3)^2\\ &+(a_1b_3-a_2b_4+a_3b_1+a_4b_2)^2\\ &+(a_1b_4+a_2b_3-a_3b_2+a_4b_1)^2,\end{aligned}$$

communicated by Euler in a letter to Goldbach on April 15, 1705. The identity also follows from the fact that the norm of the product of two QUATERNIONS is the product of the norms (Conway and Guy 1996).

see also FIBONACCI IDENTITY, LAGRANGE'S FOUR-SQUARE THEOREM

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 232, 1996.

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, p. 8, 1996.

Euler's Graeco-Roman Squares Conjecture

Euler conjectured that there do not exist GRAECO-ROMAN SQUARES (now known as EULER SQUARES) of order $n = 4k + 2$ for $k = 1, 2, \ldots$. Such squares were found to exist in 1959, refuting the CONJECTURE.

see also EULER SQUARE, LATIN SQUARE

Euler Graph

A GRAPH containing an EULERIAN CIRCUIT. An undirected GRAPH is Eulerian IFF every VERTEX has EVEN DEGREE. A DIRECTED GRAPH is Eulerian IFF every VERTEX has equal INDEGREE and OUTDEGREE. A planar BIPARTITE GRAPH is DUAL to a planar Euler graph and vice versa. The number of Euler graphs with n nodes are 1, 1, 2, 3, 7, 16, 54, 243, ... (Sloane's A002854).

References

Sloane, N. J. A. Sequence A002854/M0846 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Euler's Homogeneous Function Theorem

Let $f(x, y)$ be a HOMOGENEOUS FUNCTION of order n so that

$$f(tx, ty) = t^n(x, y). \tag{1}$$

Then define $x' \equiv xt$ and $y' \equiv yt$. Then

$$\begin{aligned}nt^{n-1}f(x,y)&=\frac{\partial f}{\partial x'}\frac{\partial x'}{\partial t}+\frac{\partial f}{\partial y'}\frac{\partial y'}{\partial t}\\ &=x\frac{\partial f}{\partial x'}+y\frac{\partial f}{\partial y'}=x\frac{\partial f}{\partial(xt)}+y\frac{\partial f}{\partial(yt)}.\end{aligned} \tag{2}$$

Let $t = 1$, then

$$x\frac{\partial f}{\partial x}+y\frac{\partial f}{\partial y}=nf(x,y). \tag{3}$$

This can be generalized to an arbitrary number of variables

$$x_i\frac{\partial f}{\partial x_i}=nf(\mathbf{x}), \tag{4}$$

where EINSTEIN SUMMATION has been used.

Euler's Hypergeometric Transformations

$${}_2F_1(a,b;c;z)=\int_0^1\frac{t^{b-1}(1-t)^{c-b-1}}{(1-tz)^a}\,dt, \tag{1}$$

where ${}_2F_1(a, b; c; z)$ is a HYPERGEOMETRIC FUNCTION. The solution can be written using the Euler's transformations

$$t \to t \tag{2}$$

$$t \to 1-t \tag{3}$$

$$t \to (1-z-tz)^{-1} \tag{4}$$

$$t \to \frac{1-t}{1-tz} \tag{5}$$

in the equivalent forms

$$\begin{aligned}{}_2F_1(a,b;c;z)&=(1-z)^{-a}\,{}_2F_1(a,c-b;c;z/(z-1)) &(6)\\ &=(1-z)^{-b}\,{}_2F_1(c-a,b;c;z/(z-1)) &(7)\\ &=(1-z)^{c-a-b}\,{}_2F_1(c-a,c-b;c;z). &(8)\end{aligned}$$

see also HYPERGEOMETRIC FUNCTION

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 585–591, 1953.

Euler Identity

For $|z| < 1$,

$$\prod_{p=1}^{\infty}(1+z^p)=\prod_{q=1}^{\infty}(1-z^{2q-1})^{-1}.$$

see also JACOBI TRIPLE PRODUCT, q-SERIES

Euler's Idoneal Number

see IDONEAL NUMBER

Euler Integral

Euler integration was defined by Schanuel and subsequently explored by Rota, Chen, and Klain. The Euler integral of a FUNCTION $f : \mathbb{R} \to \mathbb{R}$ (assumed to be piecewise-constant with finitely many discontinuities) is the sum of

$$f(x) - \tfrac{1}{2}[f(x_+) + f(x_-)]$$

over the finitely many discontinuities of f. The n-D Euler integral can be defined for classes of functions $\mathbb{R}^n \to \mathbb{R}$. Euler integration is additive, so the Euler integral of $f + g$ equals the sum of the Euler integrals of f and g.

see also EULER MEASURE

Euler-Jacobi Pseudoprime

An Euler-Jacobi pseudoprime is a number n such that

$$2^{(n-1)/2} \equiv \frac{2}{n} \pmod{n}.$$

The first few are 561, 1105, 1729, 1905, 2047, 2465, ... (Sloane's A006971).

see also PSEUDOPRIME

References

Sloane, N. J. A. Sequence A006971/M5461 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Euler L-Function

A special case of the ARTIN L-FUNCTION for the POLYNOMIAL $x^2 + 1$. It is given by

$$L(s) = \prod_{p \text{ odd prime}} \frac{1}{1 - \chi^-(p)p^{-s}},$$

where

$$\begin{aligned} \chi^-(p) &\equiv \begin{cases} 1 & \text{for } p \equiv 1 \pmod 4 \\ -1 & \text{for } p \equiv 3 \pmod 4 \end{cases} \\ &= \left(\frac{-1}{p}\right), \end{aligned}$$

where $(-1/p)$ is a LEGENDRE SYMBOL.

References

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Euler-Lagrange Differential Equation

A fundamental equation of CALCULUS OF VARIATIONS which states that if J is defined by an INTEGRAL of the form

$$J = \int f(x, y, \dot{y})\, dx, \tag{1}$$

where

$$\dot{y} \equiv \frac{dy}{dt}, \tag{2}$$

then J has a STATIONARY VALUE if the Euler-Lagrange differential equation

$$\frac{\partial f}{\partial y} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{y}}\right) = 0 \tag{3}$$

is satisfied. If time DERIVATIVE NOTATION is replaced instead by space variable notation, the equation becomes

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y_x} = 0. \tag{4}$$

In many physical problems, f_x (the PARTIAL DERIVATIVE of f with respect to x) turns out to be 0, in which case a manipulation of the Euler-Lagrange differential equation reduces to the greatly simplified and partially integrated form known as the BELTRAMI IDENTITY,

$$f - y_x \frac{\partial f}{\partial y_x} = C. \tag{5}$$

For three independent variables (Arfken 1985, pp. 924–944), the equation generalizes to

$$\frac{\partial f}{\partial u} - \frac{\partial}{\partial x}\frac{\partial f}{\partial u_x} - \frac{\partial}{\partial y}\frac{\partial f}{\partial u_y} - \frac{\partial}{\partial z}\frac{\partial f}{\partial u_z} = 0. \tag{6}$$

Problems in the CALCULUS OF VARIATIONS often can be solved by solution of the appropriate Euler-Lagrange equation.

To derive the Euler-Lagrange differential equation, examine

$$\begin{aligned} \delta J \equiv \delta \int L(q, \dot{q}, t)\, dt &= \int \left(\frac{\partial L}{\partial q}\delta q + \frac{\partial L}{\partial \dot{q}}\delta \dot{q}\right) dt \\ &= \int \left[\frac{\partial L}{\partial q}\,\delta q + \frac{\partial L}{\partial \dot{q}}\frac{d(\delta q)}{dt}\right] dt, \end{aligned} \tag{7}$$

since $\delta \dot{q} = d(\delta q)/dt$. Now, integrate the second term by PARTS using

$$u = \frac{\partial L}{\partial \dot{q}} \qquad dv = d(\delta q) \tag{8}$$

$$du = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) dt \qquad v = \delta q, \tag{9}$$

so

$$\int \frac{\partial L}{\partial \dot{q}} \frac{d(\delta q)}{dt}\, dt = \int \frac{\partial L}{\partial \dot{q}}\, d(\delta q)$$
$$= \left[\frac{\partial L}{\partial \dot{q}} \delta q\right]_{t_1}^{t_2} - \int_{t_1}^{t_2} \left(\frac{d}{dt} \frac{\partial L}{\partial \dot{q}}\, dt\right) \delta q. \quad (10)$$

Combining (7) and (10) then gives

$$\delta J = \left[\frac{\partial L}{\partial \dot{q}} \delta q\right]_{t_1}^{t_2} + \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial q} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}}\right) \delta q\, dt. \quad (11)$$

But we are varying the path only, not the endpoints, so $\delta q(t_1) = \delta q(t_2) = 0$ and (11) becomes

$$\delta J = \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial q} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}}\right) \delta q\, dt. \quad (12)$$

We are finding the STATIONARY VALUES such that $\delta J = 0$. These must vanish for any small change δq, which gives from (12),

$$\frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) = 0. \quad (13)$$

This is the Euler-Lagrange differential equation.

The variation in J can also be written in terms of the parameter κ as

$$\delta J = \int [f(x, y + \kappa v, \dot{y} + \kappa \dot{v}) - f(x, y, \dot{y})]\, dt$$
$$= \kappa I_1 + \tfrac{1}{2}\kappa^2 I_2 + \tfrac{1}{6}\kappa^3 I_3 + \tfrac{1}{24}\kappa^4 I_4 + \ldots, \quad (14)$$

where

$$v = \delta y \quad (15)$$
$$\dot{v} = \delta \dot{y} \quad (16)$$

and the first, second, etc., variations are

$$I_1 = \int (v f_y + \dot{v} f_{\dot{y}})\, dt \quad (17)$$
$$I_2 = \int (v^2 f_{yy} + 2v\dot{v} f_{y\dot{y}} + \dot{v}^2 f_{\dot{y}\dot{y}})\, dt \quad (18)$$
$$I_3 = \int (v^3 f_{yyy} + 3v^2 \dot{v} f_{yy\dot{y}} + 3v\dot{v}^2 f_{y\dot{y}\dot{y}} + \dot{v}^3 f_{\dot{y}\dot{y}\dot{y}})\, dt \quad (19)$$
$$I_4 = \int (v^4 f_{yyyy} + 4v^3 \dot{v} f_{yyy\dot{y}} + 6v^2 \dot{v}^2 f_{yy\dot{y}\dot{y}} + 4v\dot{v}^3 f_{y\dot{y}\dot{y}\dot{y}} + \dot{v}^4 f_{\dot{y}\dot{y}\dot{y}\dot{y}})\, dt. \quad (20)$$

The second variation can be re-expressed using

$$\frac{d}{dt}(v^2 \lambda) = v^2 \dot{\lambda} + 2v\dot{v}\lambda, \quad (21)$$

so

$$I_2 + [v^2 \lambda]_2^1 = \int_1^2 [v^2 (f_{yy} + \dot{\lambda}) + 2v\dot{v}(f_{y\dot{y}} + \lambda) + \dot{v}^2 f_{\dot{y}\dot{y}}]\, dt. \quad (22)$$

But

$$[v^2 \lambda]_2^1 = 0. \quad (23)$$

Now choose λ such that

$$f_{\dot{y}\dot{y}}(f_{yy} + \dot{\lambda}) = (f_{y\dot{y}} + \lambda)^2 \quad (24)$$

and z such that

$$f_{y\dot{y}} + \lambda = -\frac{f_{\dot{y}\dot{y}}}{z} \frac{dz}{dt} \quad (25)$$

so that z satisfies

$$f_{\dot{y}\dot{y}} \ddot{z} + \dot{f}_{\dot{y}\dot{y}} \dot{z} - (f_{yy} - \dot{f}_{y\dot{y}}) z = 0. \quad (26)$$

It then follows that

$$I_2 = \int f_{\dot{y}\dot{y}} \left(\dot{v} + \frac{f_{y\dot{y}} + \lambda}{f_{\dot{y}\dot{y}} v}\right)^2 dt = \int f_{\dot{y}\dot{y}} \left(\dot{v} - \frac{v}{z} \frac{dz}{dt}\right)^2. \quad (27)$$

see also BELTRAMI IDENTITY, BRACHISTOCHRONE PROBLEM, CALCULUS OF VARIATIONS, EULER-LAGRANGE DERIVATIVE

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, 1985.

Forsyth, A. R. *Calculus of Variations.* New York: Dover, pp. 17–20 and 29, 1960.

Morse, P. M. and Feshbach, H. "The Variational Integral and the Euler Equations." §3.1 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 276–280, 1953.

Euler-Lagrange Derivative

The derivative

$$\frac{\delta L}{\delta q} \equiv \frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right)$$

appearing in the EULER-LAGRANGE DIFFERENTIAL EQUATION.

Euler Line

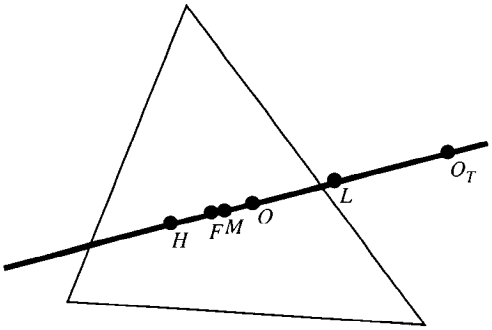

The line on which the ORTHOCENTER H, CENTROID M, CIRCUMCENTER O, DE LONGCHAMPS POINT L, NINE-POINT CENTER F, and the TANGENTIAL TRIANGLE CIRCUMCIRCLE O_T of a TRIANGLE lie. The INCENTER lies on the Euler line only if the TRIANGLE is an ISOSCELES TRIANGLE. The Euler line consists of all points with TRILINEAR COORDINATES $\alpha : \beta : \gamma$ which satisfy

$$\begin{vmatrix} \alpha & \beta & \gamma \\ \cos A & \cos B & \cos C \\ \cos B \cos C & \cos C \cos A & \cos A \cos B \end{vmatrix} = 0, \qquad (1)$$

which simplifies to

$$\alpha \cos A(\cos^2 B - \cos^2 C) + \beta \cos B(\cos^2 C - \cos^2 A) + \gamma \cos C(\cos^2 A - \cos^2 B) = 0. \qquad (2)$$

This can also be written

$$\alpha \sin(2A) \sin(B - C) + \beta \sin(2B) \sin(C - A) + \gamma \sin(2C) \sin(A - B) = 0. \qquad (3)$$

The Euler line may also be given parametrically by

$$P(\lambda) = O + \lambda H \qquad (4)$$

(Oldknow 1996).

λ	Center
-2	point at infinity
-1	de Longchamps point L
0	circumcenter O
1	centroid G
2	nine-point center F
∞	orthocenter H

The ORTHOCENTER is twice as far from the CENTROID as is the CIRCUMCENTER. The CIRCUMCENTER O, NINE-POINT CENTER F, CENTROID G, and ORTHOCENTER H form a HARMONIC RANGE.

The Euler line intersects the SODDY LINE in the DE LONGCHAMPS POINT, and the GERGONNE LINE in the EVANS POINT. The ISOTOMIC CONJUGATE of the Euler line is called JERABEK'S HYPERBOLA (Casey 1893, Vandeghen 1965).

see also CENTROID (TRIANGLE), CIRCUMCENTER, EVANS POINT, GERGONNE LINE, JERABEK'S HYPERBOLA, DE LONGCHAMPS POINT, NINE-POINT CENTER, ORTHOCENTER, SODDY LINE, TANGENTIAL TRIANGLE

References

Casey, J. *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions with Numerous Examples, 2nd rev. enl. ed.* Dublin: Hodges, Figgis, & Co., 1893.

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 18–20, 1967.

Dörrie, H. "Euler's Straight Line." §27 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 141–142, 1965.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 117–119, 1990.

Oldknow, A. "The Euler-Gergonne-Soddy Triangle of a Triangle." *Amer. Math. Monthly* **103**, 319–329, 1996.

Vandeghen, A. "Some Remarks on the Isogonal and Cevian Transforms. Alignments of Remarkable Points of a Triangle." *Amer. Math. Monthly* **72**, 1091–1094, 1965.

Euler-Lucas Pseudoprime

Let $U(P,Q)$ and $V(P,Q)$ be LUCAS SEQUENCES generated by P and Q, and define

$$D \equiv P^2 - 4Q.$$

Then

$$\begin{cases} U_{(n-(D/n))/2} \equiv 0 \pmod{n} & \text{when } (Q/n) = 1 \\ V_{(n-(D/n))/2} \equiv D \pmod{n} & \text{when } (Q/n) = -1, \end{cases}$$

where (Q/n) is the LEGENDRE SYMBOL. An ODD COMPOSITE NUMBER n such that $(n, QD) = 1$ (i.e., n and QD are RELATIVELY PRIME) is called an Euler-Lucas pseudoprime with parameters (P, Q).

see also PSEUDOPRIME, STRONG LUCAS PSEUDOPRIME

References

Ribenboim, P. "Euler-Lucas Pseudoprimes (elpsp(P,Q)) and Strong Lucas Pseudoprimes (slpsp(P,Q))." §2.X.C in *The New Book of Prime Number Records.* New York: Springer-Verlag, pp. 130–131, 1996.

Euler's Machin-Like Formula

The MACHIN-LIKE FORMULA

$$\tfrac{1}{4}\pi = \tan^{-1}(\tfrac{1}{2}) + \tan^{-1}(\tfrac{1}{3}).$$

The other 2-term MACHIN-LIKE FORMULAS are HERMANN'S FORMULA, HUTTON'S FORMULA, and MACHIN'S FORMULA.

see also INVERSE TANGENT

Euler-Maclaurin Integration Formulas

The first Euler-Maclaurin integration formula is

$$\int_0^1 f(x)\,dx = \tfrac{1}{2}[f(1) + f(0)] - \sum_{p=1}^{q} \frac{1}{(2p)!} B_{2p}[f^{(2p-1)}(1) - f^{(2p-1)}(0)] + \frac{1}{(2q)!} \int_0^1 f^{(2q)}(x) B_{2q}(x)\,dx, \qquad (1)$$

where B_n are BERNOULLI NUMBERS. SUMS may be converted to INTEGRALS by inverting the FORMULA to obtain

$$\sum_{m=1}^{n} f(m) = \int_1^n f(x)\,dx + \tfrac{1}{2}[f(1) + f(n)] + \frac{B_2}{2!}[f'(n) - f'(1)] + \ldots. \qquad (2)$$

For a more general case when $f(x)$ is tabulated at n values $f_1, f_2, \ldots, f_n$,

$$\int_{x_1}^{x_n} f(x)\,dx = h[\tfrac{1}{2}f_1 + f_2 + f_3 + \ldots + f_{n-1} + \tfrac{1}{2}f_n] - \sum_{k=1}^{\infty} \frac{B_{2k}h^{2k}}{(2k)!}[{f_n}^{(2k-1)} - {f_1}^{(2k-1)}]. \quad (3)$$

The Euler-Maclaurin formula is implemented in *Mathematica*® (Wolfram Research, Champaign, IL) as the function `NSum` with option `Method->Integrate`.

The second Euler-Maclaurin integration formula is used when $f(x)$ is tabulated at n values $f_{3/2}, f_{5/2}, \ldots, f_{n-1/2}$:

$$\int_{x_1}^{x_n} f(x)\,dx = h[f_{3/2} + f_{5/2} + f_{7/2} + \ldots + f_{n-3/2} + f_{n-1/2}] - \sum_{k=1}^{\infty} \frac{B_{2k}h^{2k}}{(2k)!}(1-2^{-2k+1})[{f_n}^{(2k-1)} - {f_1}^{(2k-1)}]. \quad (4)$$

see also SUM, WYNN'S EPSILON METHOD

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 16 and 806, 1972.

Arfken, G. "Bernoulli Numbers, Euler-Maclaurin Formula." §5.9 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 327–338, 1985.

Borwein, J. M.; Borwein, P. B.; and Dilcher, K. "Pi, Euler Numbers, and Asymptotic Expansions." *Amer. Math. Monthly* **96**, 681–687, 1989.

Vardi, I. "The Euler-Maclaurin Formula." §8.3 in *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 159–163, 1991.

Euler-Mascheroni Constant

The Euler-Mascheroni constant is denoted γ (or sometimes C) and has the numerical value

$$\gamma \approx 0.577215664901532860606512090082402431042\ldots \quad (1)$$

(Sloane's A001620). The CONTINUED FRACTION of the Euler-Mascheroni constant is [0, 1, 1, 2, 1, 2, 1, 4, 3, 13, 5, 1, 1, 8, 1, 2, 4, 1, 1, 40, ...] (Sloane's A002852). The first few CONVERGENTS are 1, 1/2, 3/5, 4/7, 11/19, 15/26, 71/123, 228/395, 3035/5258, 15403/26685, ... (Sloane's A046114 and A046115). The positions at which the digits 1, 2, ... first occur in the CONTINUED FRACTION are 2, 4, 9, 8, 11, 69, 24, 14, 139, 52, 22, ... (Sloane's A033149). The sequence of largest terms in the CONTINUED FRACTION is 1, 2, 4, 13, 40, 49, 65, 399, 2076, ... (Sloane's A033091), which occur at positions 2, 4, 8, 10, 20, 31, 34, 40, 529, ... (Sloane's A033092).

It is not known if this constant is IRRATIONAL, let alone TRANSCENDENTAL. However, Conway and Guy (1996) are "prepared to bet that it is transcendental," although they do not expect a proof to be achieved within their lifetimes.

The Euler-Mascheroni constant arises in many integrals

$$\gamma \equiv -\int_0^\infty e^{-x} \ln x\,dx \quad (2)$$

$$= \int_0^\infty \left(\frac{1}{1-e^{-x}} - \frac{1}{x}\right) e^{-x}\,dx \quad (3)$$

$$= \int_0^\infty \frac{1}{x}\left(\frac{1}{1+x} - e^{-x}\right) dx. \quad (4)$$

and sums

$$\gamma \equiv 1 + \sum_{k=2}^{\infty}\left[\frac{1}{k} + \ln\left(\frac{k-1}{k}\right)\right] \quad (5)$$

$$= \lim_{m\to\infty}\left(\sum_{n=1}^{m}\frac{1}{n} - \ln m\right) \quad (6)$$

$$= \sum_{n=2}^{\infty}(-1)^n\frac{\zeta(n)}{n} \quad (7)$$

$$= \ln\left(\frac{4}{\pi}\right) - \sum_{n=1}^{\infty}\frac{(-1)^n\zeta(n+1)}{2^n(n+1)} \quad (8)$$

$$= \lim_{n\to\infty}\left[\sum_{k=1}^{n}k^{-1} - \ln n - \frac{1}{2n} + \sum_{k=1}^{n}\frac{B_{2k}}{(2k)n^{2k}}\right], \quad (9)$$

where $\zeta(z)$ is the RIEMANN ZETA FUNCTION and B_n are the BERNOULLI NUMBERS. It is also given by the EULER PRODUCT

$$e^\gamma = \lim_{n\to\infty}\frac{1}{\ln n}\prod_{i=1}^{n}\frac{1}{1-\frac{1}{p_i}}, \quad (10)$$

where the product is over PRIMES p. Another connection with the PRIMES was provided by Dirichlet's 1838 proof that the average number of DIVISORS of all numbers from 1 to n is asymptotic to

$$\frac{\sum_{i=1}^{n}\sigma_0(i)}{n} \sim \ln n + 2\gamma - 1 \quad (11)$$

(Conway and Guy 1996). de la Vallée Poussin (1898) proved that, if a large number n is divided by all PRIMES $\leq n$, then the average amount by which the QUOTIENT is less than the next whole number is γ.

Infinite Products involving γ also arise from the G-Function with Positive Integer n. The cases $G(2)$ and $G(3)$ give

$$\prod_{n=1}^{\infty} e^{-1+1/(2n)}\left(1+\frac{1}{n}\right)^n = \frac{e^{1+\gamma/2}}{\sqrt{2\pi}} \tag{12}$$

$$\prod_{n=1}^{\infty} e^{-2+2/n}\left(1+\frac{2}{n}\right)^n = \frac{e^{3+2\gamma}}{2\pi}. \tag{13}$$

The Euler-Mascheroni constant is also given by the limits

$$\gamma = \lim_{s\to 1}\frac{\zeta(s)-1}{s-1} \tag{14}$$

$$= -\Gamma'(1) \tag{15}$$

$$= \lim_{x\to\infty}\left[x-\Gamma\left(\frac{1}{x}\right)\right] \tag{16}$$

(Le Lionnais 1983).

The difference between the nth convergent in (6) and γ is given by

$$\sum_{k=1}^{n}\frac{1}{k} - \ln n - \gamma = \int_n^{\infty}\frac{x-\lfloor x\rfloor}{x^2}\,dx, \tag{17}$$

where $\lfloor x\rfloor$ is the Floor Function, and satisfies the Inequality

$$\frac{1}{2(n+1)} < \sum_{k=1}^{n}\frac{1}{k} - \ln n - \gamma < \frac{1}{2n} \tag{18}$$

(Young 1991). A series with accelerated convergence is

$$\gamma = \tfrac{3}{2} - \ln 2 - \sum_{m=2}^{\infty}(-1)^m\frac{m-1}{m}[\zeta(m)-1] \tag{19}$$

(Flajolet and Vardi 1996). Another series is

$$\gamma = \sum_{n=1}^{\infty}(-1)^n\frac{\lfloor \lg n\rfloor}{n} \tag{20}$$

(Vacca 1910, Gerst 1969), where Lg is the Logarithm to base 2. The convergence of this series can be greatly improved using Euler's Convergence Improvement transformation to

$$\gamma = \sum_{k=1}^{\infty}2^{-(k+1)}\sum_{j=0}^{k-1}\frac{1}{\binom{2^{k-j}+j}{j}}, \tag{21}$$

where $\binom{a}{b}$ is a Binomial Coefficient (Beeler *et al.* 1972, Item 120, with $k-j$ replacing the undefined i). Bailey (1988) gives

$$\gamma = \frac{2^n}{e^{2^n}}\sum_{m=0}^{\infty}\frac{2^{mn}}{(m+1)!}\sum_{t=0}^{m}\frac{1}{t+1} - n\ln 2 + \mathcal{O}\left(\frac{1}{2^n e^{2^n}}\right), \tag{22}$$

which is an improvement over Sweeney (1963).

The symbol γ is sometimes also used for

$$\gamma' \equiv e^{\gamma} \approx 1.781072 \tag{23}$$

(Gradshteyn and Ryzhik 1979, p. xxvii).

Odena (1982–1983) gave the strange approximation

$$(0.11111111)^{1/4} = 0.577350\ldots, \tag{24}$$

and Castellanos (1988) gave

$$(\tfrac{7}{83})^{2/9} = 0.57721521\ldots \tag{25}$$

$$\left(\frac{520^6+22}{52^4}\right)^{1/6} = 0.5772156634\ldots \tag{26}$$

$$\left(\frac{80^3+92}{61^4}\right)^{1/6} = 0.57721566457\ldots \tag{27}$$

$$\frac{990^3-55^3-79^2-4^2}{70^5} = 0.5772156649015295\ldots. \tag{28}$$

No quadratically converging algorithm for computing γ is known (Bailey 1988). 7,000,000 digits of γ have been computed as of Feb. 1998 (Plouffe).

see also Euler Product, Mertens Theorem, Stieltjes Constants

References

Bailey, D. H. "Numerical Results on the Transcendence of Constants Involving π, e, and Euler's Constant." *Math. Comput.* **50**, 275–281, 1988.

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Brent, R. P. "Computation of the Regular Continued Fraction for Euler's Constant." *Math. Comput.* **31**, 771–777, 1977.

Brent, R. P. and McMillan, E. M. "Some New Algorithms for High-Precision Computation of Euler's Constant." *Math. Comput.* **34**, 305–312, 1980.

Castellanos, D. "The Ubiquitous Pi. Part I." *Math. Mag.* **61**, 67–98, 1988.

Conway, J. H. and Guy, R. K. "The Euler-Mascheroni Number." In *The Book of Numbers.* New York: Springer-Verlag, pp. 260–261, 1996.

de la Vallée Poussin, C.-J. Untitled communication. *Annales de la Soc. Sci. Bruxelles* **22**, 84–90, 1898.

DeTemple, D. W. "A Quicker Convergence to Euler's Constant." *Amer. Math. Monthly* **100**, 468–470, 1993.

Dirichlet, G. L. *J. für Math.* **18**, 273, 1838.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/euler/euler.html`.

Flajolet, P. and Vardi, I. "Zeta Function Expansions of Classical Constants." Unpublished manuscript, 1996. `http://pauillac.inria.fr/algo/flajolet/Publications/landau.ps`.

Gerst, I. "Some Series for Euler's Constant." *Amer. Math. Monthly* **76**, 273–275, 1969.

Glaisher, J. W. L. "On the History of Euler's Constant." *Messenger of Math.* **1**, 25–30, 1872.

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, 1979.

Knuth, D. E. "Euler's Constant to 1271 Places." *Math. Comput.* **16**, 275–281, 1962.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 28, 1983.

Plouffe, S. "Plouffe's Inverter: Table of Current Records for the Computation of Constants." `http://lacim.uqam.ca/pi/records.html`.

Sloane, N. J. A. Sequences A033091, A033092, A046114, A046115, A001620/M3755, and A002852/M0097 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sweeney, D. W. "On the Computation of Euler's Constant." *Math. Comput.* **17**, 170–178, 1963.

Vacca, G. "A New Series for the Eulerian Constant." *Quart. J. Pure Appl. Math.* **41**, 363–368, 1910.

Young, R. M. "Euler's Constant." *Math. Gaz.* **75**, 187–190, 1991.

Euler-Mascheroni Integrals

Define

$$I_n \equiv (-1)^n \int_0^\infty (\ln z)^n e^{-z}\,dz, \tag{1}$$

then

$$I_0 = \int_0^\infty e^{-z}\,dz = [-e^{-z}]_0^\infty = (0+1) = 1 \tag{2}$$

$$I_1 = -\int_0^\infty (\ln z)e^{-z}\,dz = \gamma \tag{3}$$

$$I_2 = \gamma^2 + \tfrac{1}{6}\pi^2 \tag{4}$$

$$I_3 = \gamma^3 + \tfrac{1}{2}\gamma\pi^2 + 2\zeta(3) \tag{5}$$

$$I_4 = \gamma^4 + \gamma^2\pi^2 - \tfrac{3}{20}\pi^4 + 8\gamma\zeta(3), \tag{6}$$

where γ is the EULER-MASCHERONI CONSTANT and $\zeta(3)$ is APÉRY'S CONSTANT.

Euler Measure

Define the Euler measure of a polyhedral set as the EULER INTEGRAL of its indicator function. It is easy to show by induction that the Euler measure of a closed bounded convex POLYHEDRON is always 1 (independent of dimension), while the Euler measure of a d-D relative-open bounded convex POLYHEDRON is $(-1)^d$.

Euler Number

The Euler numbers, also called the SECANT NUMBERS or ZIG NUMBERS, are defined for $|x| < \pi/2$ by

$$\operatorname{sech} x - 1 \equiv -\frac{E_1^* x^2}{2!} + \frac{E_2^* x^4}{4!} - \frac{E_3^* x^6}{6!} + \ldots \tag{1}$$

$$\sec x - 1 \equiv \frac{E_1^* x^2}{2!} + \frac{E_2^* x^4}{4!} + \frac{E_3^* x^6}{6!} + \ldots, \tag{2}$$

where sech is the HYPERBOLIC SECANT and sec is the SECANT. Euler numbers give the number of ODD ALTERNATING PERMUTATIONS and are related to GENOCCHI NUMBERS. The base e of the NATURAL LOGARITHM is sometimes known as Euler's number.

Some values of the Euler numbers are

$$\begin{aligned}
E_1^* &= 1\\
E_2^* &= 5\\
E_3^* &= 61\\
E_4^* &= 1{,}385\\
E_5^* &= 50{,}521\\
E_6^* &= 2{,}702{,}765\\
E_7^* &= 199{,}360{,}981\\
E_8^* &= 19{,}391{,}512{,}145\\
E_9^* &= 2{,}404{,}879{,}675{,}441\\
E_{10}^* &= 370{,}371{,}188{,}237{,}525\\
E_{11}^* &= 69{,}348{,}874{,}393{,}137{,}901\\
E_{12}^* &= 15{,}514{,}534{,}163{,}557{,}086{,}905
\end{aligned}$$

(Sloane's A000364). The first few PRIME Euler numbers E_n occur for $n = 2, 3, 19, 227, 255, \ldots$ (Sloane's A014547) up to a search limit of $n = 1415$.

The slightly different convention defined by

$$E_{2n} = (-1)^n E_n^* \tag{3}$$

$$E_{2n+1} = 0 \tag{4}$$

is frequently used. These are, for example, the Euler numbers computed by the *Mathematica*® (Wolfram Research, Champaign, IL) function `EulerE[n]`. This definition has the particularly simple series definition

$$\operatorname{sech} x - 1 \equiv \sum_{k=0}^{\infty} \frac{E_k x^k}{k!} \tag{5}$$

and is equivalent to

$$E_n = 2^n E_n(\tfrac{1}{2}), \tag{6}$$

where $E_n(x)$ is an EULER POLYNOMIAL.

To confuse matters further, the EULER CHARACTERISTIC is sometimes also called the "Euler number."

see also BERNOULLI NUMBER, EULERIAN NUMBER, EULER POLYNOMIAL, EULER ZIGZAG NUMBER, GENOCCHI NUMBER

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Bernoulli and Euler Polynomials and the Euler-Maclaurin Formula." §23.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 804–806, 1972.

Conway, J. H. and Guy, R. K. In *The Book of Numbers.* New York: Springer-Verlag, pp. 110–111, 1996.

Guy, R. K. "Euler Numbers." §B45 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 101, 1994.

Knuth, D. E. and Buckholtz, T. J. "Computation of Tangent, Euler, and Bernoulli Numbers." *Math. Comput.* **21**, 663–688, 1967.

Sloane, N. J. A. Sequences A014547 and A000364/M4019 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Spanier, J. and Oldham, K. B. "The Euler Numbers, E_n." Ch. 5 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 39–42, 1987.

Euler Parameters

The four parameters e_0, e_1, e_2, and e_3 describing a finite rotation about an arbitrary axis. The Euler parameters are defined by

$$e_0 \equiv \cos\left(\frac{\phi}{2}\right) \tag{1}$$

$$\mathbf{e} \equiv \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} = \hat{\mathbf{n}} \sin\left(\frac{\phi}{2}\right), \tag{2}$$

and are a QUATERNION in scalar-vector representation

$$(e_0, \mathbf{e}) = e_0 + e_1 x + e_2 j + e_3 k. \tag{3}$$

Because EULER'S ROTATION THEOREM states that an arbitrary rotation may be described by only three parameters, a relationship must exist between these four quantities

$${e_0}^2 + \mathbf{e}\cdot\mathbf{e} = {e_0}^2 + {e_1}^2 + {e_2}^2 + {e_3}^2 = 1 \tag{4}$$

(Goldstein 1980, p. 153). The rotation angle is then related to the Euler parameters by

$$\cos\phi = 2{e_0}^2 - 1 = {e_0}^2 - \mathbf{e}\cdot\mathbf{e} = {e_0}^2 - {e_1}^2 - {e_2}^2 - {e_3}^2 \tag{5}$$

$$\hat{\mathbf{n}} \sin\phi = 2\mathbf{e}e_0. \tag{6}$$

The Euler parameters may be given in terms of the EULER ANGLES by

$$e_0 = \cos[\tfrac{1}{2}(\phi + \psi)]\cos(\tfrac{1}{2}\theta) \tag{7}$$

$$e_1 = \sin[\tfrac{1}{2}(\phi - \psi)]\sin(\tfrac{1}{2}\theta) \tag{8}$$

$$e_2 = \cos[\tfrac{1}{2}(\phi - \psi)]\sin(\tfrac{1}{2}\theta) \tag{9}$$

$$e_3 = \sin[\tfrac{1}{2}(\phi + \psi)]\cos(\tfrac{1}{2}\theta) \tag{10}$$

(Goldstein 1980, p. 155).

Using the Euler parameters, the ROTATION FORMULA becomes

$$\mathbf{r}' = \mathbf{r}({e_0}^2 - {e_1}^2 - {e_2}^2 - {e_3}^2) + 2\mathbf{e}(\mathbf{e}\cdot\mathbf{r}) + (\mathbf{r}\times\hat{\mathbf{n}})\sin\phi, \tag{11}$$

and the ROTATION MATRIX becomes

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \mathsf{A} \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \tag{12}$$

where the elements of the matrix are

$$a_{ij} = \delta_{ij}({e_0}^2 - e_k e_k) + 2e_i e_j + 2\epsilon_{ijk} e_0 e_k. \tag{13}$$

Here, EINSTEIN SUMMATION has been used, δ_{ij} is the KRONECKER DELTA, and ϵ_{ijk} is the PERMUTATION SYMBOL. Written out explicitly, the matrix elements are

$$a_{11} = {e_0}^2 + {e_1}^2 - {e_2}^2 - {e_3}^2 \tag{14}$$

$$a_{12} = 2(e_1 e_2 + e_0 e_3) \tag{15}$$

$$a_{13} = 2(e_1 e_3 - e_0 e_2) \tag{16}$$

$$a_{21} = 2(e_1 e_2 - e_0 e_3) \tag{17}$$

$$a_{22} = {e_0}^2 - {e_1}^2 + {e_2}^2 - {e_3}^2 \tag{18}$$

$$a_{23} = 2(e_2 e_3 + e_0 e_1) \tag{19}$$

$$a_{31} = 2(e_1 e_3 + e_0 e_2) \tag{20}$$

$$a_{32} = 2(e_2 e_3 - e_0 e_1) \tag{21}$$

$$a_{33} = {e_0}^2 - {e_1}^2 - {e_2}^2 + {e_3}^2. \tag{22}$$

see also EULER ANGLES, QUATERNION, ROTATION MATRIX

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 198–200, 1985.

Goldstein, H. *Classical Mechanics, 2nd ed.* Reading, MA: Addison-Wesley, 1980.

Landau, L. D. and Lifschitz, E. M. *Mechanics, 3rd ed.* Oxford, England: Pergamon Press, 1976.

Euler's Pentagonal Number Theorem

$$\prod_{n=1}^{\infty}(1 - x^n) = \sum_{n=-\infty}^{\infty} (-1)^n x^{n(3n+1)/2}, \tag{1}$$

where $n(3n+1)/2$ are generalized PENTAGONAL NUMBERS. Related equalities are

$$\prod_{k=1}^{\infty}(1 - x^k t) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{n(n+1)/2} t^n}{\prod_{k=1}^{n}(1 - x^k)} \tag{2}$$

$$\prod_{k=1}^{\infty}(1 - x^k t)^{-1} = \sum_{n=0}^{\infty} \frac{t^n}{\prod_{k=1}^{n}(1 - x^k)}. \tag{3}$$

see also PARTITION FUNCTION P, PENTAGONAL NUMBER

Euler's Phi Function

see TOTIENT FUNCTION

Euler-Poincaré Characteristic

see EULER CHARACTERISTIC

Euler's Polygon Division Problem

The problem of finding in how many ways E_n a PLANE convex POLYGON of n sides can be divided into TRIANGLES by diagonals. Euler first proposed it to Christian Goldbach in 1751, and the solution is the CATALAN NUMBER $E_n = C_{n-2}$.

see also CATALAN NUMBER, CATALAN'S PROBLEM

References

Guy, R. K. "Dissecting a Polygon Into Triangles." *Bull. Malayan Math. Soc.* **5**, 57–60, 1958.

Euler Polyhedral Formula

see POLYHEDRAL FORMULA

Euler Polynomial

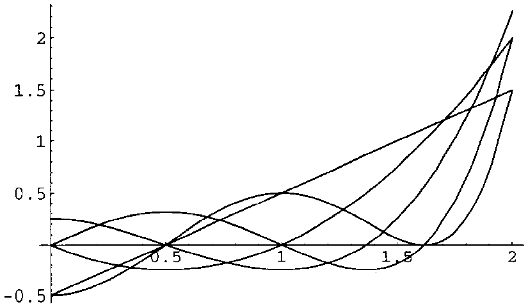

A POLYNOMIAL $E_n(x)$ given by the sum

$$\frac{2e^{xt}}{e^t+1} \equiv \sum_{n=0}^{\infty} E_n(x)\frac{t^n}{n!}. \tag{1}$$

Euler polynomials are related to the BERNOULLI NUMBERS by

$$E_{n-1}(x) = \frac{2^n}{n}\left[B_n\left(\frac{x+1}{2}\right) - B_n\left(\frac{x}{2}\right)\right] \tag{2}$$

$$= \frac{2}{n}\left[B_n(x) - 2^n B_n\left(\frac{x}{2}\right)\right] \tag{3}$$

$$E_{n-2}(x) = 2\binom{n}{2}^{-1}\sum_{k=0}^{n-2}\binom{n}{k}[(2^{n-k}-1)B_{n-k}B_k(x)], \tag{4}$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT. Setting $x = 1/2$ and normalizing by 2^n gives the EULER NUMBER

$$E_n = 2^n E_n(\tfrac{1}{2}). \tag{5}$$

Call $E'_n = E_n(0)$, then the first few terms are $-1/2$, 0, 1/4, $-1/2$, 0, 17/8, 0, 31/2, 0, ... The terms are the same but with the SIGNS reversed if $x = 1$. These values can be computed using the double sum

$$E_n(0) = 2^{-n}\sum_{j=1}^{n}\left[(-1)^{j+n+1}j^k\sum_{k=0}^{n-j}\binom{n+1}{k}\right]. \tag{6}$$

The BERNOULLI NUMBERS B_n for $n > 1$ can be expressed in terms of the E'_n by

$$B_n = -\frac{nE'_{n-1}}{2(2^n-1)}. \tag{7}$$

see also BERNOULLI POLYNOMIAL, EULER NUMBER, GENOCCHI NUMBER

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Bernoulli and Euler Polynomials and the Euler-Maclaurin Formula." §23.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 804–806, 1972.

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, 1979.

Spanier, J. and Oldham, K. B. "The Euler Polynomials $E_n(x)$." Ch. 20 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 175–181, 1987.

Euler Polynomial Identity

see EULER FOUR-SQUARE IDENTITY

Euler Power Conjecture

see EULER'S SUM OF POWERS CONJECTURE

Euler Product

For $\sigma > 1$,

$$\zeta(\sigma) \equiv \sum_{n=1}^{\infty}\frac{1}{n^\sigma} = \prod_p \frac{1}{1-\frac{1}{p^\sigma}},$$

where $\zeta(z)$ is the RIEMANN ZETA FUNCTION.

$$e^\gamma = \lim_{n\to\infty}\frac{1}{\ln n}\prod_{i=1}^{n}\frac{1}{1-\frac{1}{p_i}},$$

where the product is over PRIMES p, where γ is the EULER-MASCHERONI CONSTANT.

see also DEDEKIND FUNCTION

Euler Pseudoprime

Euler pseudoprimes to a base a are ODD COMPOSITE numbers such that $(a, n) = 1$ and the JACOBI SYMBOL satisfies

$$\left(\frac{a}{n}\right) \equiv a^{(n-1)/2} \pmod{n}.$$

No ODD COMPOSITE number is an Euler pseudoprime for all bases a RELATIVELY PRIME to it. This class includes some CARMICHAEL NUMBERS and all STRONG PSEUDOPRIMES to base a. An Euler pseudoprime is pseudoprime to at most 1/2 of all possible bases less than itself. The first few Euler pseudoprimes are 341, 561, 1105, 1729, 1905, 2047, ... (Sloane's A006970).

see also PSEUDOPRIME, STRONG PSEUDOPRIME

References

Guy, R. K. "Pseudoprimes. Euler Pseudoprimes. Strong Pseudoprimes." §A12 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 27–30, 1994.

Sloane, N. J. A. Sequence A006970/M5442 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Euler's Quadratic Residue Theorem

A number D that possesses no common divisor with a prime number p is either a QUADRATIC RESIDUE or nonresidue of p, depending whether $D^{(p-1)/2}$ is congruent mod p to ± 1.

Euler Quartic Conjecture

Euler conjectured that there are no POSITIVE INTEGER solutions to the quartic DIOPHANTINE EQUATION

$$A^4 + B^4 = C^4 + D^4.$$

This conjecture was disproved by N. D. Elkies in 1988, who found an infinite class of solutions.

see also DIOPHANTINE EQUATION—QUARTIC

References
Berndt, B. C. and Bhargava, S. "Ramanujan—For Lowbrows." *Amer. Math. Monthly* **100**, 644–656, 1993.
Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 139–140, 1994.
Lander, L. J.; Parkin, T. R.; and Selfridge, J. L. "A Survey of Equal Sums of Like Powers." *Math. Comput.* **21**, 446–459, 1967.
Ward, M. "Euler's Problem on Sums of Three Fourth Powers." *Duke Math. J.* **15**, 827–837, 1948.

Euler's Rotation Theorem

An arbitrary ROTATION may be described by only three parameters.

see also EULER ANGLES, EULER PARAMETERS, ROTATION MATRIX

Euler's Rule

The numbers $2^n pq$ and $2^n r$ are AMICABLE NUMBERS if the three INTEGERS

$$p \equiv 2^m(2^{n-m}+1) - 1$$
$$q \equiv 2^m(2^{n-m}+1) - 1$$
$$r \equiv 2^{n+m}(2^{n-m}+1)^2 - 1$$

are all PRIME numbers for some POSITIVE INTEGER m satisfying $1 \leq m \leq n-1$ (Dickson 1952, p. 42). However, there are exotic AMICABLE NUMBERS which do not satisfy Euler's rule, so it is a SUFFICIENT but not NECESSARY condition for amicability.

see also AMICABLE NUMBERS

References
Dickson, L. E. *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, 1952.

Euler's Series Transformation

Accelerates the rate of CONVERGENCE for an ALTERNATING SERIES

$$S = \sum_{s=0}^{\infty} (-1)^s u_s$$
$$= u_0 - u_1 + u_2 - \ldots - u_{n-1} + \sum_{s=0}^{\infty} \frac{(-1)^2}{2^{s+1}} [\Delta^s u_n] \qquad (1)$$

for n EVEN and Δ the FORWARD DIFFERENCE operator

$$\Delta^k u_n \equiv \sum_{m=0}^{k} (-1)^m \binom{k}{m} u_{n+k-m}, \qquad (2)$$

where $\binom{k}{m}$ are BINOMIAL COEFFICIENTS. The POSITIVE terms in the series can be converted to an ALTERNATING SERIES using

$$\sum_{r=1}^{\infty} v_r = \sum_{r=1}^{\infty} (-1)^{r-1} w_r, \qquad (3)$$

where

$$w_r \equiv v_r + 2v_{2r} + 4v_{4r} + 8v_{8r} + \ldots. \qquad (4)$$

see also ALTERNATING SERIES

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 16, 1972.

Euler's Spiral

see CORNU SPIRAL

Euler Square

A square ARRAY made by combining n objects of two types such that the first and second elements form LATIN SQUARES. Euler squares are also known as GRAECO-LATIN SQUARES, GRAECO-ROMAN SQUARES, or LATIN-GRAECO SQUARES. For many years, Euler squares were known to exist for $n = 3$, 4, and for every ODD n except $n = 3k$. EULER'S GRAECO-ROMAN SQUARES CONJECTURE maintained that there do not exist Euler squares of order $n = 4k + 2$ for $k = 1, 2, \ldots$. However, such squares were found to exist in 1959, refuting the CONJECTURE.

see also LATIN RECTANGLE, LATIN SQUARE, ROOM SQUARE

References
Beezer, R. "Graeco-Latin Squares." `http://buzzard.ups.edu/squares.html`.
Kraitchik, M. "Euler (Graeco-Latin) Squares." §7.12 in *Mathematical Recreations.* New York: W. W. Norton, pp. 179–182, 1942.

Euler Sum

In response to a letter from Goldbach, Euler considered DOUBLE SUMS of the form

$$s(m,n) = \sum_{k=1}^{\infty} \left(1 + \frac{1}{2} + \ldots + \frac{1}{k}\right)^m (k+1)^{-n} \qquad (1)$$
$$= \sum_{k=1}^{\infty} [\gamma + \psi_0(k+1)]^m (k+1)^{-n} \qquad (2)$$

with $m \geq 1$ and $n \geq 2$ and where γ is the EULER-MASCHERONI CONSTANT and $\Psi(x) = \psi_0(x)$ is the DIGAMMA FUNCTION. Euler found explicit formulas in terms of the RIEMANN ZETA FUNCTION for $s(1,n)$ with $n \geq 2$, and E. Au-Yeung numerically discovered

$$\sum_{k=1}^{\infty} \left(1+\frac{1}{2}+\ldots+\frac{1}{k}\right)^2 k^{-2} = \tfrac{17}{4}\zeta(4), \tag{3}$$

where $\zeta(z)$ is the RIEMANN ZETA FUNCTION, which was subsequently rigorously proven true (Borwein and Borwein 1995). Sums involving k^{-n} can be re-expressed in terms of sums the form $(k+1)^{-n}$ via

$$\begin{aligned}
&\sum_{k=1}^{\infty} \left(1+\frac{1}{2^m}+\ldots+\frac{1}{k^m}\right) k^{-n} \\
&\quad = \sum_{k=0}^{\infty} \left[1+\frac{2}{2^m}+\ldots+\frac{1}{(k+1)^m}\right] (k+1)^{-n} \\
&\quad = \sum_{k=1}^{\infty} \left(1+\frac{1}{2^m}+\ldots+\frac{1}{k^m}\right) (k+1)^{-n} + \sum_{k=1}^{\infty} k^{-(m+n)} \\
&\quad \equiv \sigma_h(m,n) + \zeta(m+n) \qquad (4)
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{k=1}^{\infty} \left(1+\frac{1}{2}+\ldots+\frac{1}{k}\right)^2 k^{-n} \\
&\qquad = s_h(2,n) + 2s_h(1,n+1) + \zeta(n+2), \qquad (5)
\end{aligned}$$

where σ_h is defined below.

Bailey *et al.* (1994) subsequently considered sums of the forms

$$s_h(m,n) = \sum_{k=1}^{\infty} \left(1+\frac{1}{2}+\ldots+\frac{1}{k}\right)^m (k+1)^{-n} \tag{6}$$

$$s_a(m,n) = \sum_{k=1}^{\infty} \left[1-\frac{1}{2}+\ldots+\frac{(-1)^{k+1}}{k}\right]^m (k+1)^{-n} \tag{7}$$

$$a_h(m,n) = \sum_{k=1}^{\infty} \left(1+\frac{1}{2}+\ldots+\frac{1}{k}\right)^m (-1)^{k+1}(k+1)^{-n} \tag{8}$$

$$a_a(m,n) = \sum_{k=1}^{\infty} \left(1-\frac{1}{2}+\ldots+\frac{(-1)^{k+1}}{k}\right)^m (-1)^{k+1}(k+1)^{-n} \tag{9}$$

$$\sigma_h(m,n) = \sum_{k=1}^{\infty} \left(1+\frac{1}{2^m}+\ldots+\frac{1}{k^m}\right) (k+1)^{-n} \tag{10}$$

$$\sigma_a(m,n) = \sum_{k=1}^{\infty} \left(1-\frac{1}{2^m}+\ldots+\frac{(-1)^{k+1}}{k^m}\right) (k+1)^{-n} \tag{11}$$

$$\alpha_h(m,n) = \sum_{k=1}^{\infty} \left(1+\frac{1}{2^m}+\ldots+\frac{1}{k^m}\right) (-1)^{k+1}(k+1)^{-n} \tag{12}$$

$$\alpha_a(m,n) = \sum_{k=1}^{\infty} \left(1-\frac{1}{2^m}+\ldots+\frac{(-1)^{k+1}}{k^m}\right) (-1)^{k+1}(k+1)^{-n}, \tag{13}$$

where s_h and s_a have the special forms

$$s_h = \sum_{k=1}^{\infty} [\gamma + \psi_0(n+1)]^m k + 1^{-n} \tag{14}$$

$$\begin{aligned}
a_a = {} & \sum_{k=1}^{\infty} \{\ln 2 + \tfrac{1}{2}(-1)^n \\
& \times[\psi_0(\tfrac{1}{2}n+\tfrac{1}{2}) - \psi_0(\tfrac{1}{2}n+1)]\}^m (k+1)^{-m}. \qquad (15)
\end{aligned}$$

Analytic single or double sums over $\zeta(z)$ can be constructed for

$$s_h(1,n) = \tfrac{1}{2}n\zeta(n+1) - \tfrac{1}{2}\sum_{k=1}^{n-2} \zeta(n-k)\zeta(k+1) \tag{16}$$

$$\begin{aligned}
s_h(2,n) = {} & \tfrac{1}{3}n(n+1)\zeta(n+2) + \zeta(2)\zeta(n) \\
& - \tfrac{1}{2}n \sum_{k=0}^{n-2} \zeta(n-k)\zeta(k+2) \\
& + \tfrac{1}{3}\sum_{k=2}^{n-2} \zeta(n-k) \sum_{j=1}^{k-1} \zeta(j+1)\zeta(k+1-j) + \sigma_h(2,n) \qquad (17)
\end{aligned}$$

$$\begin{aligned}
s_h(2,2n-1) = {} & \tfrac{1}{6}(2n^2-7n-3)\zeta(2n+1) + \zeta(2)\zeta(2n-1) \\
& - \tfrac{1}{2}\sum_{k=1}^{n-2} (2k-1)\zeta(2n-1-2k)\zeta(2k+2) \\
& + \tfrac{1}{3}\sum_{k=1}^{n-2} \zeta(2k+1) \sum_{j=1}^{n-2-k} \zeta(2j+1)\zeta(2n-1-2k-2j) \qquad (18)
\end{aligned}$$

$$\sigma_h(1,n) = s_h(1,n) \tag{19}$$

$$\begin{aligned}
\sigma_h(2,2n-1) = {} & -\tfrac{1}{2}(2n^2+n+1)\zeta(2n+1) + \zeta(2)\zeta(2n-1) \\
& + \sum_{k=1}^{n-1} 2k\zeta(k+1)\zeta(2n-2k) \qquad (20)
\end{aligned}$$

$$\begin{aligned}
\sigma_h(m \text{ even}, n \text{ odd}) = {} & \tfrac{1}{2}\left[\binom{m+n}{m} - 1\right] \zeta(m+n) + \zeta(m)\zeta(n) \\
& - \sum_{j=1}^{m+n} \left[\binom{2j-2}{m-1} + \binom{2j-2}{n-1}\right] \\
& \times \zeta(2j-1)\zeta(m+n-2j+1) \qquad (21)
\end{aligned}$$

$$\begin{aligned}
\sigma_h(m \text{ odd}, n \text{ even}) = {} & -\tfrac{1}{2}\left[\binom{m+n}{m} + 1\right] \zeta(m+n) \\
& + \sum_{k=1}^{m+n} \left[\binom{2j-2}{m-1} + \binom{2j-2}{n-1}\right] \\
& \times \zeta(2j-1)\zeta(m+n-2j+1), \qquad (22)
\end{aligned}$$

where $\binom{n}{m}$ is a BINOMIAL COEFFICIENT. Explicit formulas inferred using the PSLQ ALGORITHM include

$$s_h(2,2) = \tfrac{3}{2}\zeta(4) + \tfrac{1}{2}[\zeta(2)]^2 \tag{23}$$

$$= \tfrac{11}{360}\pi^4 \tag{24}$$

$$s_h(2,4) = \tfrac{2}{3}\zeta(6) - \tfrac{1}{3}\zeta(2)\zeta(4) + \tfrac{1}{3}[\zeta(2)]^3 - [\zeta(3)]^2 \tag{25}$$

$$= \tfrac{37}{22680}\pi^6 - [\zeta(3)]^2 \tag{26}$$

$$s_h(3,2) = \tfrac{15}{2}\zeta(5) + \zeta(2)\zeta(3) \tag{27}$$

$$s_h(3,3) = -\tfrac{33}{16}\zeta(6) + 2[\zeta(3)]^2 \tag{28}$$

$$s_h(3,4) = \tfrac{119}{16}\zeta(7) - \tfrac{33}{4}\zeta(3)\zeta(4) + 2\zeta(2)\zeta(5) \tag{29}$$

$$\begin{aligned} s_h(3,6) = {} & \tfrac{197}{24}\zeta(9) - \tfrac{33}{4}\zeta(4)\zeta(5) - \tfrac{37}{8}\zeta(3)\zeta(6) \\ & + [\zeta(3)]^3 + 3\zeta(2)\zeta(7) \end{aligned} \tag{30}$$

$$s_h(4,2) = \tfrac{859}{24}\zeta(6) + 3[\zeta(3)]^2 \tag{31}$$

$$s_h(4,3) = -\tfrac{109}{8}\zeta(7) + \tfrac{37}{2}\zeta(3)\zeta(4) - 5\zeta(2)\zeta(5) \tag{32}$$

$$\begin{aligned} s_h(4,5) = {} & -\tfrac{29}{2}\zeta(9) + \tfrac{37}{2}\zeta(4)\zeta(5) + \tfrac{33}{4}\zeta(3)\zeta(6) \\ & - \tfrac{8}{3}[\zeta(3)]^3 - 7\zeta(2)\zeta(7) \end{aligned} \tag{33}$$

$$s_h(5,2) = \tfrac{1855}{16}\zeta(7) + 33\zeta(3)\zeta(4) + \tfrac{57}{2}\zeta(2)\zeta(5) \tag{34}$$

$$\begin{aligned} s_h(5,4) = {} & \tfrac{890}{9}\zeta(9) + 66\zeta(4)\zeta(5) - \tfrac{4295}{24}\zeta(3)\zeta(6) - 5[\zeta(3)]^3 \\ & + \tfrac{265}{8}\zeta(2)\zeta(7) \end{aligned} \tag{35}$$

$$\begin{aligned} s_h(6,3) = {} & -\tfrac{3073}{12}\zeta(9) - 243\zeta(4)\zeta(5) + \tfrac{2097}{4}\zeta(3)\zeta(6) \\ & + \tfrac{67}{3}[\zeta(3)]^3 - \tfrac{651}{8}\zeta(2)\zeta(7) \end{aligned} \tag{36}$$

$$\begin{aligned} s_h(7,2) = {} & \tfrac{134701}{36}\zeta(9) + \tfrac{15697}{8}\zeta(4)\zeta(5) + \tfrac{29555}{24}\zeta(3)\zeta(6) \\ & + 56[\zeta(3)]^3 + \tfrac{3287}{4}\zeta(2)\zeta(7), \end{aligned} \tag{37}$$

$$s_a(2,2) = 6\,\mathrm{Li}_4(\tfrac{1}{2}) + \tfrac{1}{4}(\ln 2)^4 - \tfrac{29}{8}\zeta(4) + \tfrac{3}{2}\zeta(2)(\ln 2)^2 \tag{38}$$

$$\begin{aligned} s_a(2,3) = {} & 4\,\mathrm{Li}_5(\tfrac{1}{2}) - \tfrac{1}{30}(\ln 2)^5 - \tfrac{17}{32}\zeta(5) - \tfrac{11}{8}\zeta(4)\ln 2 \\ & + \tfrac{7}{4}\zeta(3)(\ln 2)^2 + \tfrac{1}{3}\zeta(2)(\ln 2)^3 - \tfrac{3}{4}\zeta(2)\zeta(3), \end{aligned} \tag{39}$$

$$\begin{aligned} s_a(3,2) = {} & -24\,\mathrm{Li}_5(\tfrac{1}{2}) + 6\ln 2\,\mathrm{Li}_4(\tfrac{1}{2}) + \tfrac{9}{20}(\ln 2)^5 + \tfrac{659}{32}\zeta(5) \\ & - \tfrac{285}{16}\zeta(4)\ln 2 + \tfrac{5}{2}\zeta(2)(\ln 2)^3 + \tfrac{1}{2}\zeta(2)\zeta(3), \end{aligned} \tag{40}$$

$$\begin{aligned} a_h(2,2) = {} & -2\,\mathrm{Li}_4(\tfrac{1}{2}) - \tfrac{1}{12}(\ln 2)^4 + \tfrac{99}{48}\zeta(4) - \tfrac{7}{4}\zeta(3)\ln 2 \\ & + \tfrac{1}{2}\zeta(2)(\ln 2)^2 \end{aligned} \tag{41}$$

$$\begin{aligned} a_h(2,3) = {} & -4\,\mathrm{Li}_5(\tfrac{1}{2}) - 4(\ln 2)\,\mathrm{Li}_4(\tfrac{1}{2}) - \tfrac{2}{15}(\ln 2)^5 + \tfrac{107}{32}\zeta(5) \\ & - \tfrac{7}{4}\zeta(3)(\ln 2)^2 + \tfrac{2}{3}\zeta(2)(\ln 2)^3 + \tfrac{3}{8}\zeta(2)\zeta(3) \end{aligned} \tag{42}$$

$$\begin{aligned} a_h(3,2) = {} & 6\,\mathrm{Li}_5(\tfrac{1}{2}) + 6(\ln 2)\,\mathrm{Li}_4(\tfrac{1}{2}) + \tfrac{1}{5}(\ln 2)^5 - \tfrac{33}{8}\zeta(5) \\ & + \tfrac{21}{8}\zeta(3)(\ln 2)^2 - \zeta(2)(\ln 2)^3 - \tfrac{15}{16}\zeta(2)\zeta(3), \end{aligned} \tag{43}$$

and

$$\begin{aligned} a_a(2,2) = {} & -4\,\mathrm{Li}_4(\tfrac{1}{2}) - \tfrac{1}{6}(\ln 2)^4 + \tfrac{37}{16}\zeta(4) + \tfrac{7}{4}\zeta(3)(\ln 2) \\ & - 2\zeta(2)(\ln 2)^2 \end{aligned} \tag{44}$$

$$\begin{aligned} a_a(2,3) = {} & 4(\ln 2)\,\mathrm{Li}_4(\tfrac{1}{2}) + \tfrac{1}{6}(\ln 2)^5 - \tfrac{79}{32}\zeta(5) \\ & + \tfrac{11}{8}\zeta(4)(\ln 2) - \zeta(2)(\ln 2)^3 + \tfrac{3}{8}\zeta(2)\zeta(3) \end{aligned} \tag{45}$$

$$\begin{aligned} a_a(3,2) = {} & 30\,\mathrm{Li}_5(\tfrac{1}{2}) - \tfrac{1}{4}(\ln 2)^5 - \tfrac{1813}{64}\zeta(5) \\ & + \tfrac{285}{16}\zeta(4)(\ln 2) + \tfrac{21}{8}\zeta(3)(\ln 2)^2 \\ & - \tfrac{7}{2}\zeta(2)(\ln 2)^3 + \tfrac{3}{4}\zeta(2)\zeta(3), \end{aligned} \tag{46}$$

where Li_n is a POLYLOGARITHM, and $\zeta(z)$ is the RIEMANN ZETA FUNCTION (Bailey and Plouffe). Of these, only $s_h(3,2)$, $s_h(3,3)$ and the identities for $s_a(m,n)$, $a_h(m,n)$ and $a_a(m,n)$ have been rigorously established.

References

Bailey, D. and Plouffe, S. "Recognizing Numerical Constants." `http://www.cecm.sfu.ca/organics/papers/bailey/`.

Bailey, D. H.; Borwein, J. M.; and Girgensohn, R. "Experimental Evaluation of Euler Sums." *Exper. Math.* **3**, 17–30, 1994.

Berndt, B. C. *Ramanujan's Notebooks: Part I.* New York: Springer-Verlag, 1985.

Borwein, D. and Borwein, J. M. "On an Intriguing Integral and Some Series Related to $\zeta(4)$." *Proc. Amer. Math. Soc.* **123**, 1191–1198, 1995.

Borwein, D.; Borwein, J. M.; and Girgensohn, R. "Explicit Evaluation of Euler Sums." *Proc. Edinburgh Math. Soc.* **38**, 277–294, 1995.

de Doelder, P. J. "On Some Series Containing $\Psi(x) - \Psi(y)$ and $(\Psi(x) - \Psi(y))^2$ for Certain Values of x and y." *J. Comp. Appl. Math.* **37**, 125–141, 1991.

Euler's Sum of Powers Conjecture

Euler conjectured that at least n nth POWERS are required for $n > 2$ to provide a sum that is itself an nth POWER. The conjecture was disproved by Lander and Parkin (1967) with the counterexample

$$27^5 + 84^5 + 110^5 + 133^5 = 144^5.$$

see also DIOPHANTINE EQUATION

References

Lander, L. J. and Parkin, T. R. "A Counterexample to Euler's Sum of Powers Conjecture." *Math. Comput.* **21**, 101–103, 1967.

Euler's Theorem

A generalization of FERMAT'S LITTLE THEOREM. Euler published a proof of the following more general theorem in 1736. Let $\phi(n)$ denote the TOTIENT FUNCTION. Then

$$a^{\phi(n)} \equiv 1 \pmod{n}$$

for all a RELATIVELY PRIME to n.

see also CHINESE HYPOTHESIS, EULER'S DISPLACEMENT THEOREM, EULER'S DISTRIBUTION THEOREM, FERMAT'S LITTLE THEOREM, TOTIENT FUNCTION

References

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 21 and 23–25, 1993.

Euler Totient Function

see TOTIENT FUNCTION

Euler's Totient Rule

The number of bases mod p in which $1/p$ has cycle length l is the same as the number of FRACTIONS $0/(p-1)$, $1/(p-1)$, …, $(p-2)/(p-1)$ which have least DENOMINATOR l.

see also TOTIENT FUNCTION

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 167–168, 1996.

Euler's Transform

A technique for SERIES CONVERGENCE IMPROVEMENT which takes a convergent alternating series

$$\sum_{k=0}^{\infty}(-1)^k a_k = a_0 - a_1 + a_2 - \ldots \tag{1}$$

into a series with more rapid convergence to the same value to

$$s = \sum_{k=0}^{\infty} \frac{(-1)^k \Delta^k a_0}{2^{k+1}}, \tag{2}$$

where the FORWARD DIFFERENCE is defined by

$$\Delta^k a_0 = \sum_{m=0}^{k} \equiv (-1)^m \binom{k}{m} a_{k-m} \tag{3}$$

(Abramowitz and Stegun 1972; Beeler *et al.* 1972, Item 120).

see also FORWARD DIFFERENCE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 16, 1972.

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Euler Transformation

see EULER'S FINITE DIFFERENCE TRANSFORMATION, EULER'S HYPERGEOMETRIC TRANSFORMATIONS, EULER'S TRANSFORM

Euler's Triangle

The triangle of numbers $A_{n,k}$ given by

$$A_{n,1} = A_{n,n} = 1$$

and the RECURRENCE RELATION

$$A_{n+1,k} = kA_{n,k} + (n+2-k)A_{n,k-1}$$

for $k \in [2,n]$, where $A_{n,k}$ are EULERIAN NUMBERS.

$$\begin{array}{c}
1 \\
1 \quad 1 \\
1 \quad 4 \quad 1 \\
1 \quad 11 \quad 11 \quad 1 \\
1 \quad 26 \quad 66 \quad 26 \quad 1 \\
1 \quad 57 \quad 302 \quad 302 \quad 57 \quad 1 \\
1 \; 120 \; 1191 \; 2416 \; 1191 \; 120 \; 1.
\end{array}$$

The numbers 1, 1, 1, 1, 4, 1, 1, 11, 11, 1, … are Sloane's A008292. Amazingly, the Z-TRANSFORMS of t^n

$$\frac{(z-1)^n}{T^n z} Z[t^n] = \frac{(1-z)^n}{T^n z} \lim_{x \to 0} \frac{\partial^n}{\partial x^n} \left(\frac{z}{z - e^{-xT}} \right)$$

are generators for Euler's triangle.

see also CLARK'S TRIANGLE, EULERIAN NUMBER, LEIBNIZ HARMONIC TRIANGLE, NUMBER TRIANGLE, PASCAL'S TRIANGLE, SEIDEL-ENTRINGER-ARNOLD TRIANGLE, Z-TRANSFORM

References

Sloane, N. J. A. Sequence A008292 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Euler Triangle Formula

Let O and I be the CIRCUMCENTER and INCENTER of a TRIANGLE with CIRCUMRADIUS R and INRADIUS r. Let d be the distance between O and I. Then

$$d^2 = R^2 - 2rR.$$

Euler Walk

see EULERIAN TRAIL

Euler Zigzag Number

The number of ALTERNATING PERMUTATIONS for n elements is sometimes called an Euler zigzag number. Denote the number of ALTERNATING PERMUTATIONS on n elements for which the first element is k by $E(n,k)$. Then $E(1,1)$ and

$$E(n,k) = \begin{cases} 0 & \text{for } k \geq n \text{ or } k < 1 \\ E(n,k+1) & \text{otherwise.} \\ \quad +E(n-1,n-k) & \end{cases}$$

see also ALTERNATING PERMUTATION, ENTRINGER NUMBER, SECANT NUMBER, TANGENT NUMBER

References

Ruskey, F. "Information of Alternating Permutations." `http://sue.csc.uvic.ca/~cos/inf/perm/Alternating.html`.

Sloane, N. J. A. Sequence A000111/M1492 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Eulerian Circuit

An EULERIAN TRAIL which starts and ends at the same VERTEX. In other words, it is a CYCLE which uses each EDGE exactly once. The term EULERIAN CYCLE is also used synonymously with Eulerian circuit. For technical reasons, Eulerian circuits are easier to study mathematically than are HAMILTONIAN CIRCUITS. As a generalization of the KÖNIGSBERG BRIDGE PROBLEM, Euler showed (without proof) that a CONNECTED GRAPH has an Eulerian circuit IFF it has no VERTICES of ODD DEGREE.

see also EULER GRAPH, HAMILTONIAN CIRCUIT

Eulerian Cycle

see EULERIAN CIRCUIT

Eulerian Integral of the First Kind

Legendre and Whittaker and Watson's (1990) term for the BETA FUNCTION.

References

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Eulerian Integral of the Second Kind

$$\begin{aligned}\Pi(z,n) &\equiv \int_0^n \left(1-\frac{t}{n}\right)^n t^{z-1}\,dt = n^z \int_0^1 (1-\tau)^n \tau^{z-1}\,d\tau \\ &= \frac{n!}{z(z+1)\cdots(z+n)} n^z.\end{aligned}$$

Eulerian Number

The number of PERMUTATIONS of length n with $k \leq n$ RUNS, denoted $\left\langle {n \atop k} \right\rangle$, $A_{n,k}$, or $A(n,k)$. The Eulerian numbers are given explicitly by the sum

$$\left\langle {n \atop k} \right\rangle = \sum_{j=0}^{k} (-1)^j \binom{n+1}{j} (k-j)^n. \qquad (1)$$

Making the definition

$$b_{n,1} = 1 \qquad (2)$$

$$b_{1,n} = 1 \qquad (3)$$

together with the RECURRENCE RELATION

$$b_{n,k} = n b_{n,k-1} + k b_{n-1,k} \qquad (4)$$

for $n > k$ then gives

$$\left\langle {n \atop k} \right\rangle = b_{k,n-k+1}. \qquad (5)$$

The arrangement of the numbers into a triangle gives EULER'S TRIANGLE, whose entries are 1, 1, 1, 1, 4, 1, 1, 11, 11, 1, ... (Sloane's A008292). Therefore, they represent a sort of generalization of the BINOMIAL COEFFICIENTS where the defining RECURRENCE RELATION weights the sum of neighbors by their row and column numbers, respectively.

The Eulerian numbers satisfy

$$\sum_{k=1}^{n} \left\langle {n \atop k} \right\rangle = n!. \qquad (6)$$

Eulerian numbers also arise in the surprising context of integrating the SINC FUNCTION, and also in sums of the form

$$\sum_{k=1}^{\infty} k^n r^k = \mathrm{Li}_{-n}(r) = \frac{r}{(1-r)^{n+1}} \sum_{i=1}^{n} \left\langle {n \atop i} \right\rangle r^{n-i}, \qquad (7)$$

where $\mathrm{Li}_m(z)$ is the POLYLOGARITHM function.

see also COMBINATION LOCK, EULER NUMBER, EULER'S TRIANGLE, EULER ZIGZAG NUMBER, POLYLOGARITHM, SINC FUNCTION, WORPITZKY'S IDENTITY, *Z*-TRANSFORM

References

Carlitz, L. "Eulerian Numbers and Polynomials." *Math. Mag.* **32**, 247–260, 1959.

Foata, D. and Schützenberger, M.-P. *Théorie Géométrique des Polynômes Eulériens.* Berlin: Springer-Verlag, 1970.

Kimber, A. C. "Eulerian Numbers." Supplement to *Encyclopedia of Statistical Sciences.* (Eds. S. Kotz, N. L. Johnson, and C. B. Read). New York: Wiley, pp. 59–60, 1989.

Salama, I. A. and Kupper, L. L. "A Geometric Interpretation for the Eulerian Numbers." *Amer. Math. Monthly* **93**, 51–52, 1986.

Sloane, N. J. A. Sequence A008292 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Eulerian Trail

A WALK on the EDGES of a GRAPH which uses each EDGE exactly once. A CONNECTED GRAPH has an Eulerian trail IFF it has at most two VERTICES of ODD DEGREE.

see also EULERIAN CIRCUIT

Eutactic Star

An orthogonal projection of a CROSS onto a 3-D SUBSPACE. It is said to be normalized if the CROSS vectors are all of unit length.

see also HADWIGER'S PRINCIPAL THEOREM

Evans Point

The intersection of the GERGONNE LINE and the EULER LINE. It does not appear to have a simple parametric representation.

References

Oldknow, A. "The Euler-Gergonne-Soddy Triangle of a Triangle." *Amer. Math. Monthly* **103**, 319–329, 1996.

Eve

see APPLE, ROOT, SNAKE, SNAKE EYES, SNAKE OIL METHOD, SNAKE POLYIAMOND

Even Function

A function $f(x)$ such that $f(x) = f(-x)$. An even function times an ODD FUNCTION is odd.

Even Number

An INTEGER of the form $N = 2n$, where n is an INTEGER. The even numbers are therefore $\ldots, -4, -2, 0, 2, 4, 6, 8, 10, \ldots$ (Sloane's A005843). Since the even numbers are integrally divisible by two, $N \equiv 0 \pmod 2$ for even N. An even number N for which $N \equiv 2 \pmod 4$ is called a SINGLY EVEN NUMBER, and an even number N for which $N \equiv 0 \pmod 4$ is called a DOUBLY EVEN NUMBER. An integer which is not even is called an ODD NUMBER. The GENERATING FUNCTION of the even numbers is

$$\frac{2x}{(x-1)^2} = 2x + 4x^2 + 6x^3 + 8x^4 + \ldots.$$

see also DOUBLY EVEN NUMBER, EVEN FUNCTION, ODD NUMBER, SINGLY EVEN NUMBER

References

Sloane, N. J. A. Sequence A005843/M0985 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Eventually Periodic

A PERIODIC SEQUENCE such as {1, 1, 1, 2, 1, 2, 1, 2, 1, 2, 1, 1, 2, 1, ...} which is periodic from some point onwards.

see also PERIODIC SEQUENCE

Everett's Formula

$$f_p = (1-p)f_0 + pf_1 + E_2\delta_0^2 + F_2\delta_1^2 + E_4\delta_0^4 + F_4\delta_1^4 + E_6\delta_0^6 + F_6\delta_1^6 + \ldots, \quad (1)$$

for $p \in [0, 1]$, where δ is the CENTRAL DIFFERENCE and

$$E_{2n} \equiv G_{2n} - G_{2n+1} \equiv B_{2n} - B_{2n+1} \quad (2)$$

$$F_{2n} \equiv G_{2n+1} \equiv B_{2n} + B_{2n+1}, \quad (3)$$

where G_k are the COEFFICIENTS from GAUSS'S BACKWARD FORMULA and GAUSS'S FORWARD FORMULA and B_k are the COEFFICIENTS from BESSEL'S FINITE DIFFERENCE FORMULA. The E_ks and F_ks also satisfy

$$E_{2n}(p) = F_{2n}(q) \quad (4)$$

$$F_{2n}(p) = E_{2n}(q), \quad (5)$$

for

$$q \equiv 1 - p. \quad (6)$$

see also BESSEL'S FINITE DIFFERENCE FORMULA

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 880–881, 1972.

Acton, F. S. *Numerical Methods That Work, 2nd printing.* Washington, DC: Math. Assoc. Amer., pp. 92–93, 1990.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 433, 1987.

Everett Interpolation

see EVERETT'S FORMULA

Eversion

A curve on the unit sphere S^2 is an eversion if it has no corners or cusps (but it may be self-intersecting). These properties are guaranteed by requiring that the curve's velocity never vanishes. A mapping $\boldsymbol{\sigma}: S^1 \to S^2$ forms an immersion of the CIRCLE into the SPHERE IFF, for all $\theta \in \mathbb{R}$,

$$\left|\frac{d}{d\theta}[\boldsymbol{\sigma}(e^{i\theta})]\right| > 0.$$

Smale (1958) showed it is possible to turn a SPHERE inside out (SPHERE EVERSION) using eversion.

see also SPHERE EVERSION

References

Smale, S. "A Classification of Immersions of the Two-Sphere." *Trans. Amer. Math. Soc.* **90**, 281–290, 1958.

Evolute

An evolute is the locus of centers of curvature (the envelope) of a plane curve's normals. The original curve is then said to be the INVOLUTE of its evolute. Given a plane curve represented parametrically by $(f(t), g(t))$, the equation of the evolute is given by

$$x = f - R\sin\tau \quad (1)$$

$$y = g + R\cos\tau, \quad (2)$$

where (x, y) are the coordinates of the running point, R is the RADIUS OF CURVATURE

$$R = \frac{(f'^2 + g'^2)^{3/2}}{f'g'' - f''g'}, \quad (3)$$

and τ is the angle between the unit TANGENT VECTOR

$$\hat{\mathbf{T}} = \frac{\mathbf{x}'}{|\mathbf{x}'|} = \frac{1}{\sqrt{f'^2 + g'^2}}\begin{bmatrix} f' \\ g' \end{bmatrix} \quad (4)$$

and the x-AXIS,

$$\cos\tau = \hat{\mathbf{T}} \cdot \hat{\mathbf{x}} \quad (5)$$

$$\sin\tau = \hat{\mathbf{T}} \cdot \hat{\mathbf{y}}. \quad (6)$$

Combining gives

$$x = f - \frac{(f'^2 + g'^2)g'}{f'g'' - f''g'} \tag{7}$$

$$y = g + \frac{(f'^2 + g'^2)g'}{f'g'' - f''g'}. \tag{8}$$

The definition of the evolute of a curve is independent of parameterization for any differentiable function (Gray 1993). If E is the evolute of a curve I, then I is said to be the INVOLUTE of E. The centers of the OSCULATING CIRCLES to a curve form the evolute to that curve (Gray 1993, p. 90).

The following table lists the evolutes of some common curves.

Curve	Evolute
astroid	astroid 2 times as large
cardioid	cardioid 1/3 as large
cayley's sextic	nephroid
circle	point (0, 0)
cycloid	equal cycloid
deltoid	deltoid 3 times as large
ellipse	Lamé curve
epicycloid	enlarged epicycloid
hypocycloid	similar hypocycloid
limaçon	circle catacaustic for a point source
logarithmic spiral	equal logarithmic spiral
nephroid	nephroid 1/2 as large
parabola	Neile's parabola
tractrix	catenary

see also INVOLUTE, OSCULATING CIRCLE

References
Cayley, A. "On Evolutes of Parallel Curves." *Quart. J. Pure Appl. Math.* **11**, 183–199, 1871.
Dixon, R. "String Drawings." Ch. 2 in *Mathographics.* New York: Dover, pp. 75–78, 1991.
Gray, A. "Evolutes." §5.1 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 76–80, 1993.
Jeffrey, H. M. "On the Evolutes of Cubic Curves." *Quart. J. Pure Appl. Math.* **11**, 78–81 and 145–155, 1871.
Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 40 and 202, 1972.
Lee, X. "Evolute." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Evolute_dir/evolute.html`.
Lockwood, E. H. "Evolutes and Involutes." Ch. 21 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 166–171, 1967.
Yates, R. C. "Evolutes." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 86–92, 1952.

Exact Covering System

A system of congruences a_i mod n_i with $1 \leq i \leq k$ is called a COVERING SYSTEM if every INTEGER y satisfies $y \equiv a_i \pmod{n}$ for at least one value of i. A covering system in which each integer is covered by just one congruence is called an exact covering system.

see also COVERING SYSTEM

References
Guy, R. K. "Exact Covering Systems." §F14 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 253–256, 1994.

Exact Differential

A differential of the form

$$df = P(x,y)\,dx + Q(x,y)\,dy \tag{1}$$

is exact (also called a TOTAL DIFFERENTIAL) if $\int df$ is path-independent. This will be true if

$$df = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy, \tag{2}$$

so P and Q must be of the form

$$P(x,y) = \frac{\partial f}{\partial x} \qquad Q(x,y) = \frac{\partial f}{\partial y}. \tag{3}$$

But

$$\frac{\partial P}{\partial y} = \frac{\partial^2 f}{\partial y \partial x} \tag{4}$$

$$\frac{\partial Q}{\partial x} = \frac{\partial^2 f}{\partial x \partial y}, \tag{5}$$

so

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}. \tag{6}$$

see also PFAFFIAN FORM, INEXACT DIFFERENTIAL

Exact Period

see LEAST PERIOD

Exact Trilinear Coordinates

The TRILINEAR COORDINATES $\alpha : \beta : \gamma$ of a point P relative to a TRIANGLE are PROPORTIONAL to the directed distances $a' : b' : c'$ from P to the side lines (i.e, $a' : b' : c' = k\alpha : b' = k\beta : k\gamma$). Letting k be the constant of proportionality,

$$k \equiv \frac{2\Delta}{a\alpha + b\beta + c\gamma},$$

where Δ is the AREA of ΔABC and a, b, and c are the lengths of its sides. When the trilinears are chosen so that $k = 1$, the coordinates are known as exact trilinear coordinates.

see also TRILINEAR COORDINATES

Exactly One

"Exactly one" means "one and only one," sometimes also referred to as "JUST ONE." J. H. Conway has also humorously suggested "onee" (one and only one) by analogy with IFF (if and only if), "twoo" (two and only two), and "threee" (three and only three). This refinement is sometimes needed in formal mathematical discourse because, for example, if you have two apples, you also have one apple, but you do not have *exactly one* apple.

In 2-valued LOGIC, exactly one is equivalent to the exclusive or operator XOR,

$$P(E) \text{ XOR } P(F) = P(E) + P(F) - 2P(E \cap F).$$

see also IFF, PRECISELY UNLESS, XOR

Exactly When

see IFF

Excenter

The center J_i of an EXCIRCLE. There are three excenters for a given TRIANGLE, denoted J_1, J_2, J_3. The INCENTER I and excenters J_i of a TRIANGLE are an ORTHOCENTRIC SYSTEM.

$$\overline{OI}^2 + \overline{OJ_1}^2 + \overline{OJ_2}^2 + \overline{OJ_3}^2 = 12R^2,$$

where O is the CIRCUMCENTER, J_i are the excenters, and R is the CIRCUMRADIUS (Johnson 1929, p. 190). Denote the MIDPOINTS of the original TRIANGLE M_1, M_2, and M_3. Then the lines J_1M_1, J_2M_2, and J_3M_3 intersect in a point known as the MITTENPUNKT.

see also CENTROID (ORTHOCENTRIC SYSTEM), EXCENTER-EXCENTER CIRCLE, EXCENTRAL TRIANGLE, EXCIRCLE, INCENTER, MITTENPUNKT

References

Dixon, R. *Mathographics.* New York: Dover, pp. 58–59, 1991.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, 1929.

Excenter-Excenter Circle

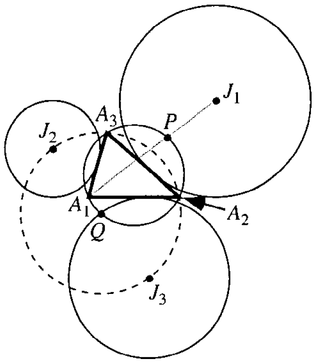

Given a TRIANGLE $\Delta A_1A_2A_3$, the points A_1, I, and J_1 lie on a line, where I is the INCENTER and J_1 is the EXCENTER corresponding to A_1. Furthermore, the circle with J_2J_3 as the diameter has Q as its center, where P is the intersection of A_1J_1 with the CIRCUMCIRCLE of $A_1A_2A_3$ and Q is the point opposite P on the CIRCUMCIRCLE. The circle with diameter J_2J_3 also passes through A_2 and A_3 and has radius

$$r = \tfrac{1}{2}a_1 \csc(\tfrac{1}{2}\alpha_1) = 2R\cos(\tfrac{1}{2}\alpha_1).$$

It arises because the points I, J_1, J_2, and J_3 form an ORTHOCENTRIC SYSTEM.

see also EXCENTER, INCENTER-EXCENTER CIRCLE, ORTHOCENTRIC SYSTEM

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 185–186, 1929.

Excentral Triangle

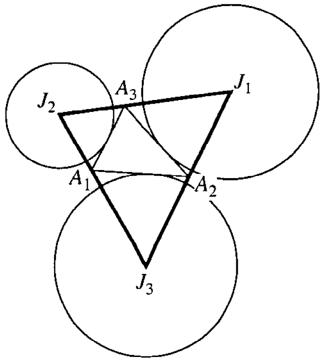

The TRIANGLE $J = \Delta J_1J_2J_3$ with VERTICES corresponding to the EXCENTERS of a given TRIANGLE A, also called the TRITANGENT TRIANGLE.

Beginning with an arbitrary TRIANGLE A, find the excentral triangle J. Then find the excentral triangle $J'$of that TRIANGLE, and so on. Then the resulting TRIANGLE $J^{(\infty)}$ approaches an EQUILATERAL TRIANGLE.

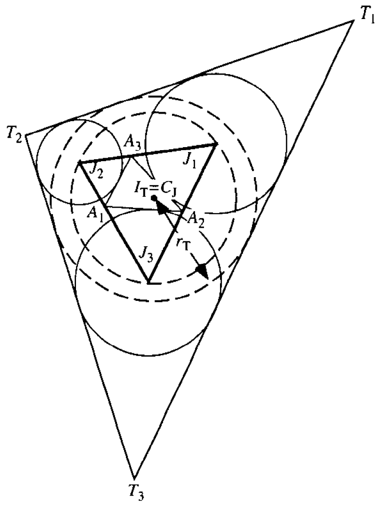

Call T the TRIANGLE tangent externally to the EXCIRCLES of A. Then the INCENTER I_T of K coincides with the CIRCUMCENTER C_J of TRIANGLE $\Delta J_1J_2J_3$, where J_i are the EXCENTERS of A. The INRADIUS r_T of the INCIRCLE of T is

$$r_T = 2R + r = \tfrac{1}{2}(r + r_1 + r_2 + r_3),$$

where R is the CIRCUMRADIUS of A, r is the INRADIUS, and r_i are the EXRADII (Johnson 1929, p. 192).

see also EXCENTER, EXCENTER-EXCENTER CIRCLE, EXCIRCLE, MITTENPUNKT

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, 1929.

Excess

see KURTOSIS

Excess Coefficient

see KURTOSIS

Excessive Number

see ABUNDANT NUMBER

Excircle

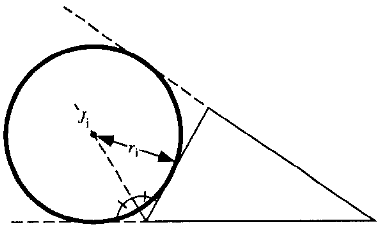

Given a TRIANGLE, extend two nonadjacent sides. The CIRCLE tangent to these two lines and to the other side of the TRIANGLE is called an ESCRIBED CIRCLE, or excircle. The CENTER J_i of the excircle is called the EXCENTER and lies on the external ANGLE BISECTOR of the opposite ANGLE. Every TRIANGLE has three excircles, and the TRILINEAR COORDINATES of the EXCENTERS are $-1 : 1 : 1$, $1 : -1 : 1$, and $1 : 1 : -1$. The RADIUS r_i of the excircle i is called its EXRADIUS.

Given a TRIANGLE with INRADIUS r, let h_i be the ALTITUDES of the excircles, and r_i their RADII (the EXRADII). Then

$$\frac{1}{h_1} + \frac{1}{h_2} + \frac{1}{h_3} = \frac{1}{r_1} + \frac{1}{r_2} + \frac{1}{r_3} = \frac{1}{r}$$

(Johnson 1929, p. 189).

see also EXCENTER, EXCENTER-EXCENTER CIRCLE, EXCENTRAL TRIANGLE, FEUERBACH'S THEOREM, NAGEL POINT, TRIANGLE TRANSFORMATION PRINCIPLE

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 11-13, 1967.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 176–177 and 182–194, 1929.

Excision Axiom

One of the EILENBERG-STEENROD AXIOMS which states that, if X is a SPACE with SUBSPACES A and U such that the CLOSURE of A is contained in the interior of U, then the INCLUSION MAP $(X\ U, A\ U) \to (X, A)$ induces an isomorphism $H_n(X\ U, A\ U) \to H_n(X, A)$.

Excluded Middle Law

A law in (2-valued) LOGIC which states there is no third alternative to TRUTH or FALSEHOOD. In other words, every statement must be either A or not-A. This fact no longer holds in THREE-VALUED LOGIC or FUZZY LOGIC.

Excludent

A method which can be used to solve any QUADRATIC CONGRUENCE. This technique relies on the fact that solving

$$x^2 \equiv b \pmod{p}$$

is equivalent to finding a value y such that

$$b + py = x^2.$$

Pick a few small moduli m. If y mod m does not make $b+py$ a quadratic residue of m, then this value of y may be excluded. Furthermore, values of $y > p/4$ are never necessary.

Excludent Factorization Method

Also known as the difference of squares. It was first used by Fermat and improved by Gauss. Gauss looked for INTEGERS x and y satisfying

$$y^2 \equiv x^2 - N \pmod{E}$$

for various moduli E. This allowed the exclusion of many potential factors. This method works best when factors are of approximately the same size, so it is sometimes better to attempt mN for some suitably chosen value of m.

see also PRIME FACTORIZATION ALGORITHMS

Exclusive Or

see XOR

Exeter Point

Define A' to be the point (other than the VERTEX A) where the MEDIAN through A meets the CIRCUMCIRCLE of ABC, and define B' and C' similarly. Then the Exeter point is the PERSPECTIVE CENTER of the TRIANGLE $A'B'C'$ and the TANGENTIAL TRIANGLE. It has TRIANGLE CENTER FUNCTION

$$\alpha = a(b^4 + c^4 - a^4).$$

References

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Kimberling, C. "Exeter Point." `http://www.evansville.edu/~ck6/tcenters/recent/exeter.html`.

Kimberling, C. and Lossers, O. P. "Problem 6557 and Solution." *Amer. Math. Monthly* **97**, 535–537, 1990.

Exhaustion Method

The method of exhaustion was a INTEGRAL-like limiting process used by Archimedes to compute the AREA and VOLUME of 2-D LAMINA and 3-D SOLIDS.

see also INTEGRAL, LIMIT

Existence

If at least one solution can be determined for a given problem, a solution to that problem is said to exist. Frequently, mathematicians seek to prove the existence of solutions and then investigate their UNIQUENESS.

see also EXISTS, UNIQUE

Existential Closure

A class of processes which attempt to round off a domain and simplify its theory by adjoining elements.

see also MODEL COMPLETION

References

Kenneth, M. "Domain Extension and the Philosophy of Mathematics." *J. Philos.* **86**, 553–562, 1989.

Exists

If there exists an A, this is written $\exists A$. Similarly, A does not exit is written $\nexists A$.

see also EXISTENCE, FOR ALL, QUANTIFIER

Exmedian

The line through the VERTEX of a TRIANGLE which is PARALLEL to the opposite side.

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, p. 176, 1929.

Exmedian Point

The point of intersection of two EXMEDIANS.

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, p. 176, 1929.

Exogenous Variable

An economic variable that is related to other economic variables and determines their equilibrium levels.

see also ENDOGENOUS VARIABLE

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 458, 1980.

Exotic $\mathbb{R}^4$

Donaldson (1983) showed there exists an exotic smooth DIFFERENTIAL STRUCTURE on $\mathbb{R}^4$. Donaldson's result has been extended to there being precisely a CONTINUUM of nondiffeomorphic DIFFERENTIAL STRUCTURES on $\mathbb{R}^4$.

see also EXOTIC SPHERE

References

Donaldson, S. K. "Self-Dual Connections and the Topology of Smooth 4-Manifold." *Bull. Amer. Math. Soc.* **8**, 81–83, 1983.

Monastyrsky, M. *Modern Mathematics in the Light of the Fields Medals.* Wellesley, MA: A. K. Peters, 1997.

Exotic Sphere

Milnor (1963) found more than one smooth structure on the 7-D HYPERSPHERE. Generalizations have subsequently been found in other dimensions. Using SURGERY theory, it is possible to relate the number of DIFFEOMORPHISM classes of exotic spheres to higher homotopy groups of spheres (Kosinski 1992). Kervaire and Milnor (1963) computed a list of the number $N(d)$ of distinct (up to DIFFEOMORPHISM) DIFFERENTIAL STRUCTURES on spheres indexed by the DIMENSION d of the sphere. For $d = 1, 2, \ldots$, assuming the POINCARÉ CONJECTURE, they are 1, 1, 1, ≥ 1, 1, 1, 28, 2, 8, 6, 992, 1, 3, 2, 16256, 2, 16, 16, ... (Sloane's A001676). The status of $d = 4$ is still unresolved: at least one exotic structure exists, but it is not known if others do as well.

The only exotic Euclidean spaces are a CONTINUUM of EXOTIC $\mathbb{R}^4$ structures.

see also EXOTIC $\mathbb{R}^4$, HYPERSPHERE

References

Kervaire, M. A. and Milnor, J. W. "Groups of Homotopy Spheres: I." *Ann. Math.* **77**, 504–537, 1963.

Kosinski, A. A. §X.6 in *Differential Manifolds.* Boston, MA: Academic Press, 1992.

Milnor, J. "Topological Manifolds and Smooth Manifolds." *Proc. Internat. Congr. Mathematicians (Stockholm, 1962)* Djursholm: Inst. Mittag-Leffler, pp. 132–138, 1963.

Milnor, J. W. and Stasheff, J. D. *Characteristic Classes.* Princeton, NJ: Princeton University Press, 1973.

Monastyrsky, M. *Modern Mathematics in the Light of the Fields Medals.* Wellesley, MA: A. K. Peters, 1997.

Novikov, S. P. (Ed.). *Topology I.* New York: Springer-Verlag, 1996.

Sloane, N. J. A. Sequence A001676/M5197 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Exp

see EXPONENTIAL FUNCTION

Expansion

An AFFINE TRANSFORMATION in which the scale is increased. It is the opposite of a DILATION (CONTRACTION).

see also DILATION

Expansive

Let ϕ be a MAP. Then ϕ is expansive if the DISTANCE $d(\phi^n x, \phi^n y) < \delta$ for all $n \in \mathbb{Z}$, then $x = y$. Equivalently, ϕ is expansive if the orbits of two points x and y are always very close.

Expectation Value

For one discrete variable,

$$\langle f(x) \rangle = \sum_x P(x). \tag{1}$$

For one continuous variable,

$$\langle f(x) \rangle = \int f(x) P(x)\, dx. \tag{2}$$

The expectation value satisfies

$$\langle ax + by \rangle = a\langle x \rangle + b\langle y \rangle \tag{3}$$

$$\langle a \rangle = a \tag{4}$$

$$\left\langle \sum x \right\rangle = \sum \langle x \rangle . \tag{5}$$

For multiple discrete variables

$$\langle f(x_1, \ldots, x_n) \rangle = \sum_{x_1, \ldots, x_n} P(x_1, \ldots, x_n). \tag{6}$$

For multiple continuous variables

$$\langle f(x_1, \ldots, x_n) \rangle = \int f(x_1, \ldots, x_n) P(x_1, \ldots, x_n)\, dx_1 \cdots dx_n. \tag{7}$$

The (multiple) expectation value satisfies

$$\begin{aligned} \langle (x - \mu_x)(y - \mu_y) \rangle &= \langle xy - \mu_x y - \mu_y x + \mu_x \mu_y \rangle \\ &= \langle xy \rangle - \mu_x \mu_y - \mu_y \mu_x + \mu_x \mu_y \\ &= \langle xy \rangle - \langle x \rangle \langle y \rangle , \end{aligned} \tag{8}$$

where μ_i is the MEAN for the variable i.

see also MEAN

Experimental Design

see DESIGN

Exploration Problem

see JEEP PROBLEM

Exponent

The POWER p in an expression a^p.

Exponent Laws

The laws governing the combination of EXPONENTS (POWERS) are

$$x^m \cdot x^n = x^{m+n} \tag{1}$$

$$\frac{x^m}{x^n} = x^{m-n} \tag{2}$$

$$(x^m)^n = x^{mn} \tag{3}$$

$$(xy)^m = x^m y^m \tag{4}$$

$$\left(\frac{x}{y}\right)^n = \frac{x^n}{y^n} \tag{5}$$

$$x^{-n} = \frac{1}{x^n} \tag{6}$$

$$\left(\frac{x}{y}\right)^{-n} = \left(\frac{y}{x}\right)^n , \tag{7}$$

where quantities in the DENOMINATOR are taken to be nonzero. Special cases include

$$x^1 = x \tag{8}$$

and

$$x^0 = 1 \tag{9}$$

for $x \neq 0$. The definition $0^0 = 1$ is sometimes used to simplify formulas, but it should be kept in mind that this equality is a definition and not a fundamental mathematical truth.

see also EXPONENT, POWER

Exponent Vector

Let p_i denote the ith PRIME, and write

$$m = \prod_i {p_i}^{v_i}.$$

Then the exponent vector is $\mathbf{v}(m) = (v_1, v_2, \ldots)$.

see also DIXON'S FACTORIZATION METHOD

References

Pomerance, C. "A Tale of Two Sieves." *Not. Amer. Math. Soc.* **43**, 1473–1485, 1996.

Exponential Digital Invariant

see NARCISSISTIC NUMBER

Exponential Distribution

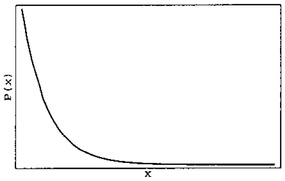

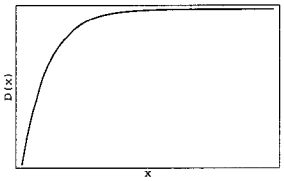

Given a POISSON DISTRIBUTION with rate of change λ, the distribution of waiting times between successive changes (with $k = 0$) is

$$\begin{aligned} D(x) &\equiv P(X \leq x) = 1 - P(X > x) \\ &= 1 - \frac{(\lambda x)^0 e^{-\lambda x}}{0!} = 1 - e^{-\lambda x} \quad (1) \\ P(x) &= D'(x) = \lambda e^{-\lambda x}, \quad (2) \end{aligned}$$

which is normalized since

$$\begin{aligned} \int_0^\infty P(x)\,dx &= \lambda \int_0^\infty e^{-\lambda x}\,dx \\ &= -[e^{-\lambda x}]_0^\infty = -(0-1) = 1. \quad (3) \end{aligned}$$

This is the only MEMORYLESS RANDOM DISTRIBUTION. Define the MEAN waiting time between successive changes as $\theta \equiv \lambda^{-1}$. Then

$$P(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & x \geq 0 \\ 0 & x < 0. \end{cases} \quad (4)$$

The MOMENT-GENERATING FUNCTION is

$$\begin{aligned} M(t) &= \int_0^\infty e^{tx} \left(\frac{1}{\theta}\right) e^{-x/\theta}\,dx = \frac{1}{\theta} \int_0^\infty e^{-(1-\theta t)x/\theta}\,dx \\ &= \left[\frac{e^{-(1-\theta t)x/\theta}}{1-\theta t}\right]_0^\infty = \frac{1}{1-\theta t} \quad (5) \\ M'(t) &= \frac{\theta}{(1-\theta t)^2} \quad (6) \\ M''(t) &= \frac{2\theta^2}{(1-\theta t)^3}, \quad (7) \end{aligned}$$

so

$$\begin{aligned} R(t) &\equiv \ln M(t) = -\ln(1-\theta t) \quad (8) \\ R'(t) &= \frac{\theta}{1-\theta t} \quad (9) \\ R''(t) &= \frac{\theta^2}{(1-\theta t)^2} \quad (10) \\ \mu &= R'(0) = \theta \quad (11) \\ \sigma^2 &= R''(0) = \theta^2. \quad (12) \end{aligned}$$

The SKEWNESS and KURTOSIS are given by

$$\begin{aligned} \gamma_1 &= 2 \quad (13) \\ \gamma_2 &= 6. \quad (14) \end{aligned}$$

The MEAN and VARIANCE can also be computed directly

$$\langle x \rangle \equiv \int_0^\infty P(x)\,dx = \frac{1}{s} \int_0^\infty x e^{-x/s}\,dx. \quad (15)$$

Use the integral

$$\int x e^{ax}\,dx = \frac{e^{ax}}{a^2}(ax - 1) \quad (16)$$

to obtain

$$\begin{aligned} \langle x \rangle &= \frac{1}{s} \left[\frac{e^{-x/s}}{\left(-\frac{1}{s}\right)^2} \left\{\left(-\frac{1}{s}\right) x - 1\right\}\right]_0^\infty \\ &= -s \left[e^{-x/s}\left(1 + \frac{x}{s}\right)\right]_0^\infty \\ &= -s(0-1) = s. \quad (17) \end{aligned}$$

Now, to find

$$\langle x^2 \rangle = \frac{1}{s} \int_0^\infty x^2 e^{-x/s}\,dx, \quad (18)$$

use the integral

$$\int x^2 e^{-x/s}\,dx = \frac{e^{ax}}{a^3}(2 - 2ax + a^2 x^2) \quad (19)$$

$$\begin{aligned} \langle x^2 \rangle &= \frac{1}{s} \left[\frac{e^{-x/s}}{\left(-\frac{1}{s}\right)^3} \left(2 + \frac{2}{s} x + \frac{1}{s^2} x^2\right)\right]_0^\infty \\ &= -s^2(0-2) = 2s^2, \quad (20) \end{aligned}$$

giving

$$\begin{aligned} \sigma^2 &\equiv \langle x^2 \rangle - \langle x \rangle^2 \\ &= 2s^2 - s^2 = s^2 \quad (21) \\ \sigma &\equiv \sqrt{\operatorname{var}(x)} = s. \quad (22) \end{aligned}$$

If a generalized exponential probability function is defined by

$$P_{(\alpha,\beta)}(x) = \frac{1}{\beta} e^{-(x-\alpha)/\beta}, \quad (23)$$

then the CHARACTERISTIC FUNCTION is

$$\phi(t) = \frac{e^{i\alpha t}}{1 - i\beta t}, \quad (24)$$

and the MEAN, VARIANCE, SKEWNESS, and KURTOSIS are

$$\begin{aligned} \mu &= \alpha + \beta \quad (25) \\ \sigma^2 &= \beta^2 \quad (26) \\ \gamma_1 &= 2 \quad (27) \\ \gamma_2 &= 6. \quad (28) \end{aligned}$$

see also DOUBLE EXPONENTIAL DISTRIBUTION

References

Balakrishnan, N. and Basu, A. P. *The Exponential Distribution: Theory, Methods, and Applications.* New York: Gordon and Breach, 1996.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 534–535, 1987.

Spiegel, M. R. *Theory and Problems of Probability and Statistics.* New York: McGraw-Hill, p. 119, 1992.

Exponential Divisor

see e-DIVISOR

Exponential Function

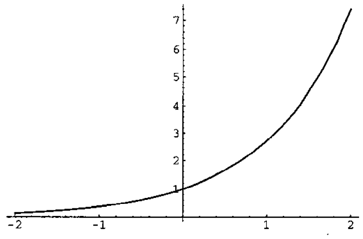

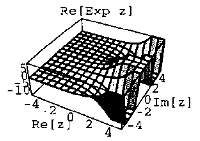

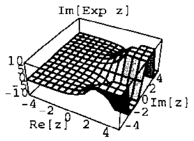

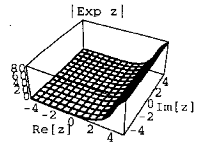

The exponential function is defined by

$$\exp(x) \equiv e^x, \tag{1}$$

where e is the constant 2.718.... It satisfies the identity

$$\exp(x+y) = \exp(x)\exp(y). \tag{2}$$

If $z \equiv x + iy$,

$$e^z = e^{x+iy} = e^x e^{iy} = e^x(\cos y + i\sin y). \tag{3}$$

If

$$a + bi = e^{x+iy}, \tag{4}$$

then

$$y = \tan^{-1}\left(\frac{b}{a}\right) \tag{5}$$

$$\begin{aligned} x &= \ln\left\{b\csc\left[\tan^{-1}\left(\frac{b}{a}\right)\right]\right\} \\ &= \ln\left\{a\sec\left[\tan^{-1}\left(\frac{b}{a}\right)\right]\right\}. \end{aligned} \tag{6}$$

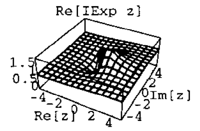

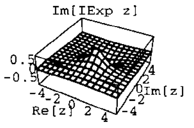

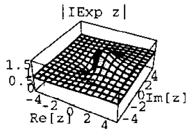

The above plot shows the function $e^{1/z}$.

see also EULER FORMULA, EXPONENTIAL RAMP, FOURIER TRANSFORM—EXPONENTIAL FUNCTION, SIGMOID FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Exponential Function." §4.2 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 69–71, 1972.

Fischer, G. (Ed.). Plates 127–128 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 124–125, 1986.

Spanier, J. and Oldham, K. B. "The Exponential Function $\exp(bx + c)$" and "Exponentials of Powers $\exp(-ax^\nu)$." Chs. 26–27 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 233–261, 1987.

Yates, R. C. "Exponential Curves." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 86–97, 1952.

Exponential Function (Truncated)

see EXPONENTIAL SUM FUNCTION

Exponential Inequality

For $c < 1$,

$$x^c < 1 + c(x-1).$$

For $c > 1$,

$$x^c > 1 + c(x-1).$$

Exponential Integral

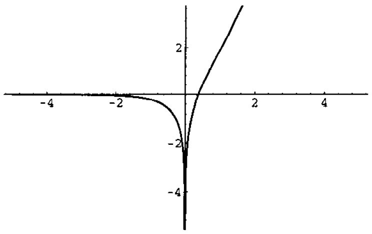

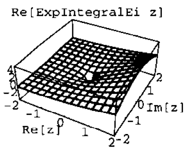

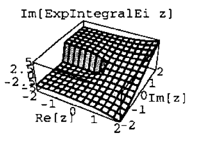

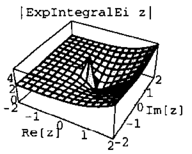

Let $\mathrm{E}_1(x)$ be the E_n-FUNCTION with $n = 1$,

$$\mathrm{E}_1(x) \equiv \int_1^\infty \frac{e^{-tx}\,dt}{t} = \int_x^\infty \frac{e^{-u}\,du}{u}. \tag{1}$$

Then define the exponential integral $\mathrm{ei}(x)$ by

$$\mathrm{E}_1(x) = -\,\mathrm{ei}(-x), \tag{2}$$

where the retention of the $-\,\mathrm{ei}(-x)$ NOTATION is a historical artifact. Then $\mathrm{ei}(x)$ is given by the integral

$$\mathrm{ei}(x) = -\int_{-x}^\infty \frac{e^{-t}\,dt}{t}. \tag{3}$$

This function is given by the *Mathematica*® (Wolfram Research, Champaign, IL) function `ExpIntegralEi[x]`. The exponential integral can also be written

$$\mathrm{ei}(ix) = \mathrm{ci}(x) + i\,\mathrm{si}(x), \tag{4}$$

where $\mathrm{ci}(x)$ and $\mathrm{si}(x)$ are COSINE and SINE INTEGRAL.

The real ROOT of the exponential integral occurs at 0.37250741078..., which is not known to be expressible in terms of other standard constants. The quantity $-e\,\mathrm{ei}(-1) = 0.596347362\ldots$ is known as the GOMPERTZ CONSTANT.

see also COSINE INTEGRAL, E_n-FUNCTION, GOMPERTZ CONSTANT, SINE INTEGRAL

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 566–568, 1985.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 434–435, 1953.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Exponential Integrals." §6.3 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 215–219, 1992.

Spanier, J. and Oldham, K. B. "The Exponential Integral Ei(x) and Related Functions." Ch. 37 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 351–360, 1987.

Exponential Map

On a LIE GROUP, exp is a MAP from the LIE ALGEBRA to its LIE GROUP. If you think of the LIE ALGEBRA as the TANGENT SPACE to the identity of the LIE GROUP, $\exp(v)$ is defined to be $h(1)$, where h is the unique LIE GROUP HOMEOMORPHISM from the REAL NUMBERS to the LIE GROUP such that its velocity at time 0 is v.

On a RIEMANNIAN MANIFOLD, exp is a MAP from the TANGENT BUNDLE of the MANIFOLD to the MANIFOLD, and $\exp(v)$ is defined to be $h(1)$, where h is the unique GEODESIC traveling through the base-point of v such that its velocity at time 0 is v.

The three notions of exp (exp from COMPLEX ANALYSIS, exp from LIE GROUPS, and exp from Riemannian geometry) are all linked together, the strongest link being between the LIE GROUPS and Riemannian geometry definition. If G is a compact LIE GROUP, it admits a left and right invariant RIEMANNIAN METRIC. With respect to that metric, the two exp maps agree on their common domain. In other words, one-parameter subgroups are geodesics. In the case of the MANIFOLD $\mathbb{S}^1$, the CIRCLE, if we think of the tangent space to 1 as being the IMAGINARY axis (y-AXIS) in the COMPLEX PLANE, then

$$\begin{aligned}\exp_{\text{Riemannian geometry}}(v) &= \exp_{\text{Lie Groups}}(v)\\ &= \exp_{\text{complex analysis}}(v),\end{aligned}$$

and so the three concepts of the exponential all agree in this case.

see also EXPONENTIAL FUNCTION

Exponential Matrix

see MATRIX EXPONENTIAL

Exponential Ramp

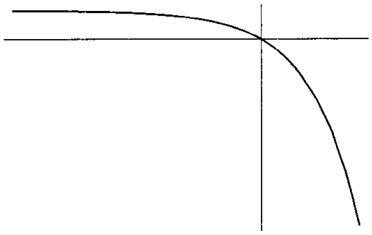

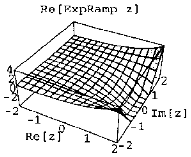

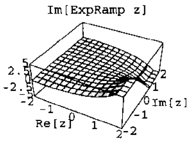

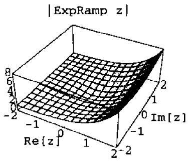

The curve

$$y = 1 - e^{ax}.$$

see also EXPONENTIAL FUNCTION, SIGMOID FUNCTION

References

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 158, 1993.

Exponential Sum Formulas

$$\begin{aligned}\sum_{n=0}^{N-1} e^{iNx} = \frac{1-e^{iNx}}{1-e^{ix}} &= \frac{-e^{iNx/2}\left(e^{-iNx/2}-e^{iNx/2}\right)}{-e^{ix/2}\left(e^{-ix/2}-e^{ix/2}\right)}\\ &= \frac{\sin(\frac{1}{2}Nx)}{\sin(\frac{1}{2}x)}e^{ix(N-1)/2},\end{aligned} \tag{1}$$

where

$$\sum_{n=0}^{N-1} r^n = \frac{1-r^n}{1-r} \tag{2}$$

has been used. Similarly,

$$\begin{aligned}\sum_{n=0}^{N-1} p^n e^{inx} = \frac{1-p^N e^{iNx}}{1-pe^{ix}} &= \frac{(1-p^N e^{iNx})(1-pe^{-ix})}{(1-pe^{ix})(1-pe^{-ix})}\\ &= \frac{1-p^N e^{iNx}-pe^{-ix}+p^{N+1}e^{ix(N-1)}}{1-p(e^{ix}+e^{-ix})+p^2}\\ &= \frac{p^{N+1}e^{ix(N-1)}-p^N e^{iNx}+1-pe^{-ix}}{1-2p\cos x+p^2}.\end{aligned} \tag{3}$$

This gives

$$\sum_{n=0}^{\infty} p^n e^{inx} = \lim_{N\to\infty}\sum_{n=0}^{N-1} p^n e^{inx} = \frac{1-pe^{-ix}}{1-2p\cos x+p^2}. \tag{4}$$

By looking at the REAL and IMAGINARY PARTS of these FORMULAS, sums involving sines and cosines can be obtained.

Exponential Sum Function

$$\mathrm{es}_n(x) \equiv \exp_n(x) \equiv \sum_{m=0}^{n} \frac{x^m}{m!}.$$

see also GAMMA FUNCTION

Exradius

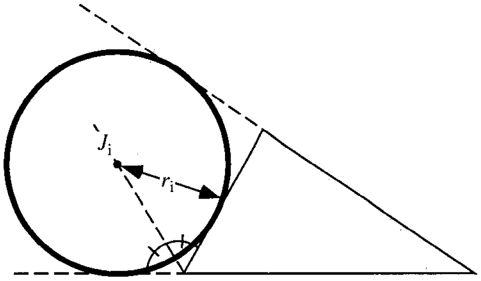

The RADIUS of an EXCIRCLE. Let a TRIANGLE have exradius r_a (sometimes denoted ρ_a), opposite side of length a, AREA Δ, and SEMIPERIMETER s. Then

$$\begin{aligned} {r_a}^2 &= \left(\frac{\Delta}{s-a}\right)^2 && (1)\\ &= \frac{s(s-c)(s-b)}{s-a} && (2)\\ &= 4R\sin(\tfrac{1}{2}\alpha_1)\cos(\tfrac{1}{2}\alpha_2)\cos(\tfrac{1}{2}\alpha_3) && (3) \end{aligned}$$

(Johnson 1929, p. 189) where R is the CIRCUMRADIUS. Let r be the INRADIUS, then

$$4R = r_a + r_b + r_c - r \tag{4}$$

$$\frac{1}{r_a} + \frac{1}{r_b} + \frac{1}{r_c} = \frac{1}{r} \tag{5}$$

$$r r_a r_b r_c = \Delta^2. \tag{6}$$

Some fascinating FORMULAS due to Feuerbach are

$$r_2 r_3 + r_3 r_1 + r_1 r_3 = s^2 \tag{7}$$

$$r(r_2 r_3 + r_3 r_1 + r_1 r_2) = s\Delta = r_1 r_2 r_3 \tag{8}$$

$$r(r_1 + r_2 + r_3) = a_2 a_3 + a_3 a_1 + a_1 a_2 - s^2 \tag{9}$$

$$r r_1 + r r_2 + r r_3 + r_1 r_2 + r_2 r_3 + r_3 r_1 = a_2 a_3 + a_3 a_1 + a_1 a_2 \tag{10}$$

$$r_2 r_3 + r_3 r_1 + r_1 r_2 - r r_1 - r r_2 - r r_3 - \tfrac{1}{2}({a_1}^2 + {a_2}^2 + {a_3}^2) \tag{11}$$

(Johnson 1929, pp. 190–191).

see also CIRCLE, CIRCUMRADIUS, EXCIRCLE, INRADIUS, RADIUS

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, 1929.

Mackay, J. S. "Formulas Connected with the Radii of the Incircle and Excircles of a Triangle." *Proc. Edinburgh Math. Soc.* **12**, 86–105.

Mackay, J. S. "Formulas Connected with the Radii of the Incircle and Excircles of a Triangle." *Proc. Edinburgh Math. Soc.* **13**, 103–104.

Exsecant

$$\operatorname{exsec} x \equiv \sec x - 1,$$

where $\sec x$ is the SECANT.

see also COVERSINE, HAVERSINE, SECANT, VERSINE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 78, 1972.

Extended Cycloid

see PROLATE CYCLOID

Extended Goldbach Conjecture

see GOLDBACH CONJECTURE

Extended Greatest Common Divisor

see GREATEST COMMON DIVISOR

Extended Mean-Value Theorem

Let the functions f and g be DIFFERENTIABLE on the OPEN INTERVAL (a,b) and CONTINUOUS on the CLOSED INTERVAL $[a,b]$. If $g'(x) \neq 0$ for any $x \in (a,b)$, then there is at least one point $c \in (a,b)$ such that

$$\frac{f'(c)}{g'(c)} = \frac{f(b)-f(a)}{g(b)-g(a)}.$$

see also MEAN-VALUE THEOREM

Extended Riemann Hypothesis

The first quadratic nonresidue mod p of a number is always less than $2(\ln p)^2$.

see also RIEMANN HYPOTHESIS

References

Bach, E. *Analytic Methods in the Analysis and Design of Number-Theoretic Algorithms.* Cambridge, MA: MIT Press, 1985.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, p. 295, 1991.

Extension

The definition of a SET by enumerating its members. An extensional definition can always be reduced to an INTENTIONAL one.

see also INTENSION

References

Russell, B. "Definition of Number." *Introduction to Mathematical Philosophy.* New York: Simon and Schuster, 1971.

Extension Problem

Given a SUBSPACE A of a SPACE X and a MAP from A to a SPACE Y, is it possible to extend that MAP to a MAP from X to Y?

see also LIFTING PROBLEM

Extensions Calculus

see EXTERIOR ALGEBRA

Extent

The RADIUS of the smallest CIRCLE centered at one of the points of an N-CLUSTER, which contains all the points in the N-CLUSTER.

see also N-CLUSTER

Exterior

That portion of a region lying "outside" a specified boundary.

see also INTERIOR

Exterior Algebra

The ALGEBRA of the EXTERIOR PRODUCT, also called an ALTERNATING ALGEBRA or GRASSMANN ALGEBRA. The study of exterior algebra is also called AUSDEHNUNGSLEHRE and EXTENSIONS CALCULUS. Exterior algebras are GRADED ALGEBRAS.

In particular, the exterior algebra of a VECTOR SPACE is the DIRECT SUM over k in the natural numbers of the VECTOR SPACES of alternating k-forms on that VECTOR SPACE. The product on this algebra is then the wedge product of forms. The exterior algebra for a VECTOR SPACE V is constructed by forming monomials u, $v \wedge w$, $x \wedge y \wedge z$, etc., where u, v, w, x, y, and z are vectors in V and $\wedge$ is asymmetric multiplication. The sums formed from linear combinations of the MONOMIALS are the elements of an exterior algebra.

References

Forder, H. G. *The Calculus of Extension.* Cambridge, England: Cambridge University Press, 1941.

Lounesto, P. "Counterexamples to Theorems Published and Proved in Recent Literature on Clifford Algebras, Spinors, Spin Groups, and the Exterior Algebra." `http://www.hit.fi/~lounesto/counterexamples.htm`.

Exterior Angle Bisector

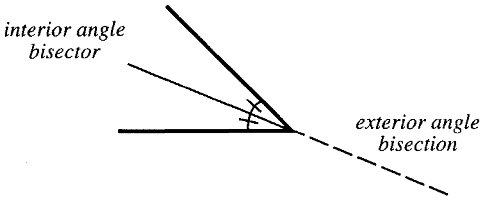

The exterior bisector of an ANGLE is the LINE or LINE SEGMENT which cuts it into two equal ANGLES on the opposite "side" as the ANGLE.

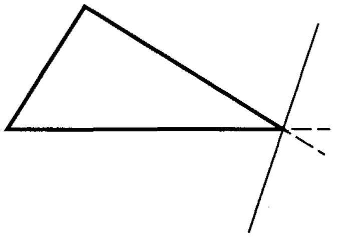

For a TRIANGLE, the exterior angle bisector bisects the SUPPLEMENTARY ANGLE at a given VERTEX. It also divides the opposite side externally in the ratio of adjacent sides.

see also ANGLE BISECTOR, ISODYNAMIC POINTS

Exterior Angle Theorem

In any TRIANGLE, if one of the sides is extended, the exterior angle is greater than both the interior and opposite angles.

References

Dunham, W. *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, p. 41, 1990.

Exterior Derivative

Consider a DIFFERENTIAL k-FORM

$$\omega^1 = b_1\,dx_1 + b_2\,dx_2. \tag{1}$$

Then its exterior derivative is

$$d\omega^1 = db_1 \wedge dx_1 + db_2 \wedge dx_2, \tag{2}$$

where $\wedge$ is the WEDGE PRODUCT. Similarly, consider

$$\omega^1 = b_1(x_1, x_2)\,dx_1 + b_2(x_1, x_2)\,dx_2. \tag{3}$$

Then

$$\begin{aligned} d\omega^1 &= db_1 \wedge dx_1 + db_2 \wedge dx_2 \\ &= \left(\frac{\partial b_1}{\partial x_1}dx_1 + \frac{\partial b_1}{\partial x_2}dx_2\right) \wedge dx_1 \\ &\quad + \left(\frac{\partial b_2}{\partial x_1}dx_1 + \frac{\partial b_2}{\partial x_2}dx_2\right) \wedge dx_2. \end{aligned} \tag{4}$$

Denote the exterior derivative by

$$Dt \equiv \frac{\partial}{\partial x} \wedge t. \tag{5}$$

Then for a 0-form t,

$$(Dt)_\mu \equiv \frac{\partial t}{\partial x^\mu}, \tag{6}$$

for a 1-form t,

$$(Dt)_{\mu\nu} \equiv \frac{1}{2}\left(\frac{\partial t_\nu}{\partial x^\mu} - \frac{\partial t_\mu}{\partial x^\nu}\right), \tag{7}$$

and for a 2-form t,

$$(Dt)_{ijk} \equiv \tfrac{1}{3}\epsilon_{ijk}\left(\frac{\partial t_{23}}{\partial x^1} + \frac{\partial t_{31}}{\partial x^2} + \frac{\partial t_{12}}{\partial x^3}\right), \tag{8}$$

where ϵ_{ijk} is the PERMUTATION TENSOR.

The second exterior derivative is

$$D^2 t = \frac{\partial}{\partial x} \wedge \left(\frac{\partial}{\partial x} \wedge t\right) = \left(\frac{\partial}{\partial x} \wedge \frac{\partial}{\partial x}\right) \wedge t = 0, \tag{9}$$

which is known as POINCARÉ'S LEMMA.

see also DIFFERENTIAL k-FORM, POINCARÉ'S LEMMA, WEDGE PRODUCT

Exterior Dimension

A type of DIMENSION which can be used to characterize FAT FRACTALS.

see also FAT FRACTAL

References

Grebogi, C.; McDonald, S. W.; Ott, E.; and Yorke, J. A. "Exterior Dimension of Fat Fractals." *Phys. Let. A* **110**, 1–4, 1985.

Grebogi, C.; McDonald, S. W.; Ott, E.; and Yorke, J. A. Erratum to "Exterior Dimension of Fat Fractals." *Phys. Let. A* **113**, 495, 1986.

Ott, E. *Chaos in Dynamical Systems.* New York: Cambridge University Press, p. 98, 1993.

Exterior Product

see WEDGE PRODUCT

Exterior Snowflake

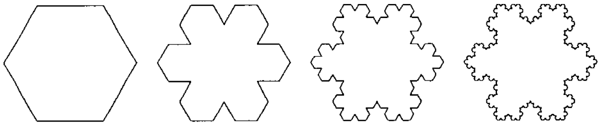

A FRACTAL.

see also FLOWSNAKE FRACTAL, KOCH ANTISNOWFLAKE, KOCH SNOWFLAKE, PENTAFLAKE

References

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 193–195, 1991.

Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m.`

Extra Strong Lucas Pseudoprime

Given the LUCAS SEQUENCE $U_n(b,-1)$ and $V_n(b,-1)$, define $\Delta = b^2 - 4$. Then an extra strong Lucas pseudoprime to the base b is a COMPOSITE NUMBER $n = 2^r s + (\Delta/n)$, where s is ODD and $(n, 2\Delta) = 1$ such that either $U_s \equiv 0 \pmod{n}$ and $V_s \equiv \pm 2 \pmod{n}$, or $V_{2^t s} \equiv 0 \pmod{n}$ for some t with $0 \leq t < r - 1$. An extra strong Lucas pseudoprime is a STRONG LUCAS PSEUDOPRIME with parameters $(b, -1)$. COMPOSITE n are extra strong pseudoprimes for at most 1/8 of possible bases (Grantham 1997).

see also LUCAS PSEUDOPRIME, STRONG LUCAS PSEUDOPRIME

References

Grantham, J. "Frobenius Pseudoprimes." `http://www.clark.net/pub/grantham/pseudo/pseudo.ps`

Grantham, J. "A Frobenius Probable Prime Test with High Confidence." 1997. `http://www.clark.net/pub/grantham/pseudo/pseudo2.ps`

Jones, J. P. and Mo, Z. "A New Primality Test Using Lucas Sequences." Preprint.

Extrapolation

see RICHARDSON EXTRAPOLATION

Extremal Coloring

see EXTREMAL GRAPH

Extremal Graph

A two-coloring of a COMPLETE GRAPH K_n of n nodes which contains exactly the number $N \equiv (R+B)_{\min}$ of MONOCHROMATIC FORCED TRIANGLES and no more (i.e., a minimum of $R+B$ where R and B are the numbers of red and blue TRIANGLES). Goodman (1959) showed that for an extremal graph,

$$N(n) = \begin{cases} \frac{1}{3}m(m-1)(m-2) & \text{for } n = 2m \\ \frac{1}{3}2m(m-1)(4m+1) & \text{for } n = 4m+1 \\ \frac{1}{3}2m(m+1)(4m-1) & \text{for } n = 4m+3. \end{cases}$$

This is sometimes known as GOODMAN'S FORMULA. Schwenk (1972) rewrote it in the form

$$N(n) = \binom{n}{3} - \left\lfloor \tfrac{1}{2} n \left\lfloor \tfrac{1}{4}(n-1)^2 \right\rfloor \right\rfloor,$$

sometimes known as SCHWENK'S FORMULA, where $\lfloor x \rfloor$ is the FLOOR FUNCTION. The first few values of $N(n)$ for $n = 1, 2, \ldots$ are 0, 0, 0, 0, 0, 2, 4, 8, 12, 20, 28, 40, 52, 70, 88, ... (Sloane's A014557).

see also BICHROMATIC GRAPH, BLUE-EMPTY GRAPH, GOODMAN'S FORMULA, MONOCHROMATIC FORCED TRIANGLE, SCHWENK'S FORMULA

References

Goodman, A. W. "On Sets of Acquaintances and Strangers at Any Party." *Amer. Math. Monthly* **66**, 778–783, 1959.

Schwenk, A. J. "Acquaintance Party Problem." *Amer. Math. Monthly* **79**, 1113–1117, 1972.

Sloane, N. J. A. Sequence A014557 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Extremals

A field of extremals is a plane region which is SIMPLY CONNECTED by a one-parameter family of extremals. The concept was invented by Weierstraß.

Extreme and Mean Ratio

see GOLDEN MEAN

Extreme Value Distribution

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let M_n denote the "extreme" (i.e., largest) ORDER STATISTIC $X^{(n)}$ for a distribution of n elements X_i taken from a continuous UNIFORM DISTRIBUTION. Then the distribution of the M_n is

$$P(M_n < x) = \begin{cases} 0 & \text{if } x < 0 \\ x^n & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1, \end{cases} \tag{1}$$

and the MEAN and VARIANCE are

$$\mu = \frac{n}{n+1} \tag{2}$$

$$\sigma^2 = \frac{n}{(n+1)^2(n+2)}. \tag{3}$$

If X_i are taken from a STANDARD NORMAL DISTRIBUTION, then its cumulative distribution is

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt = \tfrac{1}{2} + \Phi(x), \qquad (4)$$

where $\Phi(x)$ is the NORMAL DISTRIBUTION FUNCTION. The probability distribution of M_n is then

$$P(M_n < x) = [F(x)]^n = \frac{n}{\sqrt{2\pi}} \int_{-\infty}^{x} [F(t)]^{n-1} e^{-t^2/2}\, dt. \qquad (5)$$

The MEAN $\mu(n)$ and VARIANCE $\sigma^2(n)$ are expressible in closed form for small n,

$$\mu(1) = 0 \qquad (6)$$

$$\mu(2) = \frac{1}{\sqrt{\pi}} \qquad (7)$$

$$\mu(3) = \frac{3}{2\sqrt{\pi}} \qquad (8)$$

$$\mu(4) = \frac{3}{2\sqrt{\pi}} \left[1 + \frac{2}{\pi} \sin^{-1}(\tfrac{1}{3})\right] \qquad (9)$$

$$\mu(5) = \frac{5}{4\sqrt{\pi}} \left[1 + \frac{6}{\pi} \sin^{-1}(\tfrac{1}{3})\right] \qquad (10)$$

and

$$\sigma^2(1) = 1 \qquad (11)$$

$$\sigma^2(2) = 1 - \frac{1}{\pi} \qquad (12)$$

$$\sigma^2(3) = \frac{4\pi - 9 + 2\sqrt{3}}{4\pi} \qquad (13)$$

$$\sigma^2(4) = 1 + \frac{\sqrt{3}}{\pi} - [\mu(4)]^2 \qquad (14)$$

$$\sigma^2(5) = 1 + \frac{5\sqrt{3}}{4\pi} + \frac{5\sqrt{3}}{2\pi^2} \sin^{-1}(\tfrac{1}{4}) - [\mu(5)]^2. \qquad (15)$$

No exact expression is known for $\mu(6)$ or $\sigma^2(6)$, but there is an equation connecting them

$$[\mu(6)]^2 + \sigma^2(6) = 1 + \frac{5\sqrt{3}}{4\pi} + \frac{15\sqrt{3}}{2\pi^2} \sin^{-1}(\tfrac{1}{4}). \qquad (16)$$

An analog to the CENTRAL LIMIT THEOREM states that the asymptotic normalized distribution of M_n satisfies one of the three distributions

$$P(y) = \exp(-e^{-y}) \qquad (17)$$

$$P(y) = \begin{cases} 0 & \text{if } y \leq 0 \\ \exp(-y^{-a}) & \end{cases} \qquad (18)$$

$$P(y) = \begin{cases} \exp[-(-y)^a] & \text{if } y \leq 0 \\ 1 & \text{if } y > 0, \end{cases} \qquad (19)$$

also known as GUMBEL, Fréchet, and WEIBULL DISTRIBUTIONS, respectively.

see also FISHER-TIPPETT DISTRIBUTION, ORDER STATISTIC

References
Balakrishnan, N. and Cohen, A. C. *Order Statistics and Inference.* New York: Academic Press, 1991.
David, H. A. *Order Statistics, 2nd ed.* New York: Wiley, 1981.
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/extval/extval.html`.
Gibbons, J. D. *Nonparametric Statistical Inference.* New York: McGraw-Hill, 1971.

Extreme Value Theorem

If a function f is continuous on a closed interval $[a, b]$, then f has both a MAXIMUM and a MINIMUM on $[a, b]$. If f has an extreme value on an open interval (a, b), then the extreme value occurs at a CRITICAL POINT. This theorem is sometimes also called the WEIERSTRAß EXTREME VALUE THEOREM.

Extremum

A MAXIMUM or MINIMUM. An extremum may be LOCAL (a.k.a. a RELATIVE EXTREMUM; an extremum in a given region which is not the overall MAXIMUM or MINIMUM) or GLOBAL. Functions with many extrema can be very difficult to GRAPH. Notorious examples include the functions $\cos(1/x)$ and $\sin(1/x)$ near $x = 0$

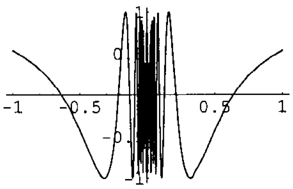

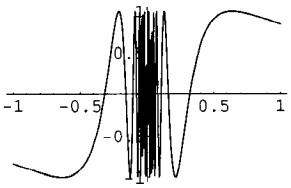

and $\sin(e^{2x+9})$ near 0 and 1.

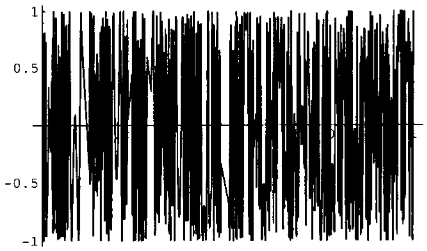

The latter has

$$\left\lfloor \frac{e^{11}}{\pi} - \frac{1}{2} \right\rfloor - \left\lceil \frac{e^9}{\pi} - \frac{1}{2} \right\rceil + 1 = 19058 - 2579 + 1 = 16480$$

extrema in the CLOSED INTERVAL [0,1] (Mulcahy 1996).

see also GLOBAL EXTREMUM, GLOBAL MAXIMUM, GLOBAL MINIMUM, KUHN-TUCKER THEOREM, LAGRANGE MULTIPLIER, LOCAL EXTREMUM, LOCAL MAXIMUM, LOCAL MINIMUM, MAXIMUM, MINIMUM

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 14, 1972.
Mulcahy, C. "Plotting and Scheming with Wavelets." *Math. Mag.* **69**, 323–343, 1996.
Tikhomirov, V. M. *Stories About Maxima and Minima.* Providence, RI: Amer. Math. Soc., 1991.

Extremum Test

Consider a function $f(x)$ in 1-D. If $f(x)$ has a relative extremum at x_0, then either $f'(x_0) = 0$ or f is not DIFFERENTIABLE at x_0. Either the first or second DERIVATIVE tests may be used to locate relative extrema of the first kind.

A NECESSARY condition for $f(x)$ to have a MINIMUM (MAXIMUM) at x_0 is

$$f'(x_0) = 0,$$

and

$$f''(x_0) \geq 0 \qquad (f''(x_0) \leq 0).$$

A SUFFICIENT condition is $f'(x_0) = 0$ and $f''(x_0) > 0$ ($f''(x_0) < 0$). Let $f'(x_0) = 0$, $f''(x_0) = 0$, ..., $f^{(n)}(x_0) = 0$, but $f^{(n+1)}(x_0) \neq 0$. Then $f(x)$ has a RELATIVE MAXIMUM at x_0 if n is ODD and $f^{(n+1)}(x_0) < 0$, and $f(x)$ has a RELATIVE MINIMUM at x_0 if n is ODD and $f^{(n+1)}(x_0) > 0$. There is a SADDLE POINT at x_0 if n is EVEN.

see also EXTREMUM, FIRST DERIVATIVE TEST, RELATIVE MAXIMUM, RELATIVE MINIMUM, SADDLE POINT (FUNCTION), SECOND DERIVATIVE TEST

Extrinsic Curvature

A curvature of a SUBMANIFOLD of a MANIFOLD which depends on its particular EMBEDDING. Examples of extrinsic curvature include the CURVATURE and TORSION of curves in 3-space, or the mean curvature of surfaces in 3-space.

see also CURVATURE, INTRINSIC CURVATURE, MEAN CURVATURE

F

F-Distribution

Arises in the testing of whether two observed samples have the same Variance. Let ${\chi_m}^2$ and ${\chi_n}^2$ be independent variates distributed as Chi-Squared with m and n Degrees of Freedom. Define a statistic $F_{n,m}$ as the ratio of the dispersions of the two distributions

$$F_{n,m} \equiv \frac{{\chi_n}^2/n}{{\chi_m}^2/m}. \tag{1}$$

This statistic then has an F-distribution with probability function and cumulative distribution

$$\begin{aligned} F_{n,m}(x) &= \frac{\Gamma(\frac{n+m}{2})n^{n/2}m^{m/2}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}\frac{x^{n/2-1}}{(m+nx)^{(n+m)/2}} && (2)\\ &= \frac{m^{m/2}n^{n/2}x^{n/2-1}}{(m+nx)^{(n+m)/2}B(\frac{1}{2}n,\frac{1}{2}m)} && (3)\\ &= I(1;\tfrac{1}{2}m;\tfrac{1}{2}n) - I\left(\frac{m}{m+nx};\tfrac{1}{2}m;\tfrac{1}{2}n\right), && (4)\end{aligned}$$

where $\Gamma(z)$ is the Gamma Function, $B(a,b)$ is the Beta Function, and $I(a,b;x)$ is the Regularized Beta Function. The Mean, Variance, Skewness and Kurtosis are

$$\mu = \frac{m}{m-2} \tag{5}$$

$$\sigma^2 = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)} \tag{6}$$

$$\gamma_1 = \frac{2(m+2n-2)}{m-6}\sqrt{\frac{2(m-4)}{n(m+n-2)}} \tag{7}$$

$$\begin{aligned}\gamma_2 &= \frac{12(-16+20m-8m^2+m^3+44n)}{n(m-6)(m-8)(n+m-2)}\\ &\quad + \frac{12(-32mn+5m^2n-22n^2+5mn^2)}{n(m-6)(m-8)(n+m-2)}. && (8)\end{aligned}$$

The probability that F would be as large as it is if the first distribution has a smaller variance than the second is denoted $Q(F_{n,m})$.

The noncentral F-distribution is given by

$$\begin{aligned}P(x) &= e^{-\lambda/2+(\lambda n_1 x)/[2(n_2+n_1x)]}\\ &\quad \times {n_1}^{n_1/2}{n_2}^{n_2/2}x^{n_1/2-1}(n_2+n_1x)^{-(n_1+n_2)/2}\\ &\quad \times \frac{\Gamma(\frac{1}{2}n_1)\Gamma(1+\frac{1}{2}n_2)L_{n_2/2}^{n_1/2-1}\left(-\frac{\lambda n_1 x}{2(n_2+n_1x)}\right)}{B(\frac{1}{2}n_1,\frac{1}{2}n_2)\Gamma[\frac{1}{2}(n_1+n_2)]}, && (9)\end{aligned}$$

where $\Gamma(z)$ is the Gamma Function, $B(\alpha,\beta)$ is the Beta Function, and $L_m^n(z)$ is an associated Laguerre Polynomial.

see also Beta Function, Gamma Function, Regularized Beta Function, Snedecor's F-Distribution

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 946–949, 1972.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Incomplete Beta Function, Student's Distribution, F-Distribution, Cumulative Binomial Distribution." §6.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 219–223, 1992.

Spiegel, M. R. *Theory and Problems of Probability and Statistics.* New York: McGraw-Hill, pp. 117–118, 1992.

F-Polynomial

see Kauffman Polynomial F

F-Ratio

The Ratio of two independent estimates of the Variance of a Normal Distribution.

see also F-Distribution, Normal Distribution, Variance

F-Ratio Distribution

see F-Distribution

Fabry Imbedding

A representation of a Planar Graph as a planar straight line graph such that no two Edges cross.

Face

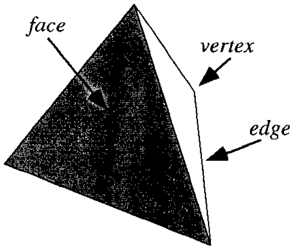

The intersection of an n-D Polytope with a tangent Hyperplane. 0-D faces are known as Vertices (nodes), 1-D faces as Edges, $(n-2)$-D faces as Ridges, and $(n-1)$-D faces as Facets.

see also Edge (Polyhedron), Facet, Polytope, Ridge, Vertex (Polyhedron)

Facet

An $(n-1)$-D Face of an n-D Polytope. A procedure for generating facets is known as Faceting.

Faceting

Using a set of corners of a SOLID that lie in a plane to form the VERTICES of a new POLYGON is called faceting. Such POLYGONS may outline new FACES that join to enclose a new SOLID, even if the sides of the POLYGONS do not fall along EDGES of the original SOLID.

References

Holden, A. *Shapes, Space, and Symmetry.* New York: Columbia University Press, p. 94, 1971.

Factor

A factor is a portion of a quantity, usually an INTEGER or POLYNOMIAL. The determination of factors is called FACTORIZATION (or sometimes "FACTORING"). It is usually desired to break factors down into the smallest possible pieces so that no factor is itself factorable. For INTEGERS, the determination of factors is called PRIME FACTORIZATION. For large quantities, the determination of all factors is usually very difficult except in exceptional circumstances.

see also DIVISOR, FACTORIZATION, GREATEST PRIME FACTOR, LEAST PRIME FACTOR, PRIME FACTORIZATION ALGORITHMS

Factor Base

The primes with LEGENDRE SYMBOL $(n/p) = 1$ (less than $N = \pi(d)$ for trial divisor d) which need be considered when using the QUADRATIC SIEVE FACTORIZATION METHOD.

see also DIXON'S FACTORIZATION METHOD

References

Morrison, M. A. and Brillhart, J. "A Method of Factoring and the Factorization of F_7." *Math. Comput.* **29**, 183–205, 1975.

Factor (Graph)

A 1-factor of a GRAPH with n VERTICES is a set of $n/2$ separate EDGES which collectively contain all n of the VERTICES of G among their endpoints.

Factor Group

see QUOTIENT GROUP

Factor Level

A grouping of statistics.

Factor Ring

see QUOTIENT RING

Factor Space

see QUOTIENT SPACE

Factorial

The factorial $n!$ is defined for a POSITIVE INTEGER n as

$$n! \equiv \begin{cases} n\cdot(n-1)\cdots 2\cdot 1 & n=1,2,\ldots \\ 1 & n=0. \end{cases} \tag{1}$$

The first few factorials for n = 0, 1, 2, ... are 1, 1, 2, 6, 24, 120, ... (Sloane's A000142). An older NOTATION for the factorial is $\underline{\lfloor n}$ (Dudeney 1970, Gardner 1978, Conway and Guy 1996).

As n grows large, factorials begin acquiring tails of trailing ZEROS. To calculate the number of trailing ZEROS for $n!$, use

$$Z = \sum_{k=1}^{k_{\max}} \left\lfloor \frac{n}{5^k} \right\rfloor, \tag{2}$$

where

$$k_{\max} \equiv \left\lfloor \frac{\ln n}{\ln 5} \right\rfloor \tag{3}$$

and $\lfloor x \rfloor$ is the FLOOR FUNCTION (Gardner 1978, p. 63; Ogilvy and Anderson 1988, pp. 112–114). For n = 1, 2, ..., the number of trailing zeros are 0, 0, 0, 0, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, ... (Sloane's A027868). This is a special application of the general result that the POWER of a PRIME p dividing $n!$ is

$$\epsilon_p(n) = \sum_{r\geq 0} \left\lfloor \frac{n}{p^r} \right\rfloor \tag{4}$$

(Graham *et al.* 1994, Vardi 1991). Stated another way, the exact POWER of a PRIME p which divides $n!$ is

$$\frac{n - \text{sum of digits of the base-}p\text{ representation of }n}{p-1}. \tag{5}$$

By noting that

$$n! = \Gamma(n+1), \tag{6}$$

where $\Gamma(n)$ is the GAMMA FUNCTION for INTEGERS n, the definition can be generalized to COMPLEX values

$$z! \equiv \Gamma(z+1) \equiv \int_0^\infty e^{-t} t^z \, dt. \tag{7}$$

This defines $z!$ for all COMPLEX values of z, except when z is a NEGATIVE INTEGER, in which case $z! = \infty$. Using the identities for GAMMA FUNCTIONS, the values of $(\frac{1}{2}n)!$ (half integral values) can be written explicitly

$$(-\tfrac{1}{2})! = \sqrt{\pi} \tag{8}$$

$$(\tfrac{1}{2})! = \tfrac{1}{2}\sqrt{\pi} \tag{9}$$

$$(n-\tfrac{1}{2})! = \frac{\sqrt{\pi}}{2^n}(2n-1)!! \tag{10}$$

$$(n+\tfrac{1}{2})! = \frac{\sqrt{\pi}}{2^{n+1}}(2n+1)!!, \tag{11}$$

where $n!!$ is a DOUBLE FACTORIAL.

For INTEGERS s and n with $s < n$,

$$\frac{(s-n)!}{(2s-2n)!} = \frac{(-1)^{n-s}(2n-2s)!}{(n-s)!}. \tag{12}$$

The LOGARITHM of $z!$ is frequently encountered

$$\ln(z!) = \frac{1}{2}\ln\left[\frac{\pi z}{\sin(\pi z)}\right] - \gamma - \sum_{n=1}^{\infty}\frac{\zeta(2n+1)}{2n+1}z^{2n+1} \tag{13}$$

$$= \frac{1}{2}\ln\left[\frac{\pi z}{\sin(\pi z)}\right] - \frac{1}{2}\ln\left(\frac{1+z}{1-z}\right) + (1-\gamma)z - \sum_{n=1}^{\infty}[\zeta(2n+1)-1]\frac{z^{2n+1}}{2n+1} \tag{14}$$

$$= \ln\left[\lim_{n\to\infty}\frac{n!}{(z+1)(z+2)\cdots(z+n)}n^z\right] \tag{15}$$

$$= \lim_{n\to\infty}[\ln(n!) + z\ln n - \ln(z+1) - \ln(z+2) - \ldots - \ln(z+n)] \tag{16}$$

$$= \sum_{n=1}^{\infty}\frac{z^n}{n!}F_{n-1}(0) \tag{17}$$

$$= -\gamma z + \sum_{n=2}^{\infty}(-1)^n\frac{z^n}{n}\zeta(n) \tag{18}$$

$$= -\ln(1+z) + z(1-\gamma) + \sum_{n=2}^{\infty}(-1)^n[\zeta(n)-1]\frac{z^n}{n}, \tag{19}$$

where γ is the EULER-MASCHERONI CONSTANT, ζ is the RIEMANN ZETA FUNCTION, and F_n is the POLYGAMMA FUNCTION. The factorial can be expanded in a series

$$z! = \sqrt{2\pi}\,z^{z+1/2}e^{-z}\left(1 + \tfrac{1}{12}z^{-1} + \tfrac{1}{288}z^{-2} - \tfrac{139}{51840}z^{-3} + \ldots\right). \tag{20}$$

STIRLING'S SERIES gives the series expansion for $\ln(z!)$,

$$\ln(z!) = \tfrac{1}{2}\ln(2\pi) + \left(z+\tfrac{1}{2}\right)\ln z - z + \frac{B_2}{2z} + \ldots + \frac{B_{2n}}{2n(2n-1)z^{2n-1}} + \ldots$$

$$= \tfrac{1}{2}\ln(2\pi) + \left(z+\tfrac{1}{2}\right)\ln z - z + \tfrac{1}{12}z^{-1} - \tfrac{1}{360}z^{-3} + \tfrac{1}{1260}z^{-5} - \ldots, \tag{21}$$

where B_n is a BERNOULLI NUMBER.

Identities satisfied by sums of factorials include

$$\sum_{k=0}^{\infty}\frac{1}{k!} = e = 2.718281828\ldots \tag{22}$$

$$\sum_{k=0}^{\infty}\frac{(-1)^k}{k!} = e^{-1} = 0.3678794412\ldots \tag{23}$$

$$\sum_{k=0}^{\infty}\frac{1}{(k!)^2} = I_0(2) = 2.279585302\ldots \tag{24}$$

$$\sum_{k=0}^{\infty}\frac{(-1)^k}{(k!)^2} = J_0(2) = 0.2238907791\ldots \tag{25}$$

$$\sum_{k=0}^{\infty}\frac{1}{(2k)!} = \cosh 1 = 1.543080635\ldots \tag{26}$$

$$\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k)!} = \cos 1 = 0.5403023059\ldots \tag{27}$$

$$\sum_{k=0}^{\infty}\frac{1}{(2k+1)!} = \sinh 1 = 1.175201194\ldots \tag{28}$$

$$\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)!} = \sin 1 = 0.8414709848\ldots \tag{29}$$

(Spanier and Oldham 1987), where I_0 is a MODIFIED BESSEL FUNCTION OF THE FIRST KIND, J_0 is a BESSEL FUNCTION OF THE FIRST KIND, cosh is the HYPERBOLIC COSINE, cos is the COSINE, sinh is the HYPERBOLIC SINE, and sin is the SINE.

Let h be the exponent of the greatest POWER of a PRIME p dividing $n!$. Then

$$h = \sum_{\substack{i=1\\ p^i\leq n}}\left\lfloor\frac{n}{p^i}\right\rfloor. \tag{30}$$

Let g be the number of 1s in the BINARY representation of n. Then

$$g + h = n \tag{31}$$

(Honsberger 1976). In general, as discovered by Legendre in 1808, the POWER m of the PRIME p dividing $n!$ is given by

$$m = \sum_{k=0}^{\infty}\left\lfloor\frac{n}{p^k}\right\rfloor = \frac{n-(n_0+n_1+\ldots+n_N)}{p-1}, \tag{32}$$

where the INTEGERS $n_1, \ldots, n_N$ are the digits of n in base p (Ribenboim 1989).

The sum-of-factorials function is defined by

$$\Sigma(n) \equiv \sum_{k=1}^{n}k!$$

$$= \frac{-e + \mathrm{ei}(1) + \pi i + \mathrm{E}_{2n+1}(-1)\Gamma(n+2)}{e}, \tag{33}$$

$$= \frac{-e + \mathrm{ei}(1) + \Re[\mathrm{E}_{2n+1}(-1)]\Gamma(n+2)}{e}, \tag{34}$$

where $\text{ei}(1) \approx 1.89512$ is the EXPONENTIAL INTEGRAL, E_n is the E_n-FUNCTION, and i is the IMAGINARY NUMBER. The first few values are 1, 3, 9, 33, 153, 873, 5913, 46233, 409113, ... (Sloane's A007489). $\Sigma(n)$ cannot be written as a hypergeometric term plus a constant (Petkovšek *et al.* 1996). However the sum

$$\Sigma'(n) \equiv \sum_{k=1}^{n} kk! = (n+1)! - 1 \tag{35}$$

has a simple form, with the first few values being 1, 5, 23, 119, 719, 5039, ... (Sloane's A033312).

The numbers $n! + 1$ are prime for $n = 1$, 2, 3, 11, 27, 37, 41, 73, 77, 116, 154, ... (Sloane's A002981), and the numbers $n! - 1$ are prime for $n = 3$, 4, 6, 7, 12, 14, 30, 32, 33, 38, 94, 166, ... (Sloane's A002982). In general, the power-product sequences (Mudge 1997) are given by $S_k^{\pm}(n) = (n!)^k \pm 1$. The first few terms of $S_2^+(n)$ are 2, 5, 37, 577, 14401, 518401, ... (Sloane's A020549), and $S_2^+(n)$ is PRIME for $n = 1$, 2, 3, 4, 5, 9, 10, 11, 13, 24, 65, 76, ... (Sloane's A046029). The first few terms of $S_2^-(n)$ are 0, 3, 35, 575, 14399, 518399, ... (Sloane's A046030), but $S_2^-(n)$ is PRIME for only $n = 2$ since $S_2^-(n) = (n!)^2 - 1 = (n!+1)(n!-1)$ for $n > 2$. The first few terms of $S_3^-(n)$ are 0, 7, 215, 13823, 1727999, ..., and the first few terms of $S_3^+(n)$ are 2, 9, 217, 13825, 1728001, ... (Sloane's A19514).

There are only four INTEGERS equal to the sum of the factorials of their digits. Such numbers are called FACTORIONS. While no factorial is a SQUARE NUMBER, D. Hoey listed sums $< 10^{12}$ of distinct factorials which give SQUARE NUMBERS, and J. McCranie gave the one additional sum less than $21! = 5.1 \times 10^{19}$:

$$\begin{aligned}
0! + 1! + 2! &= 2^2 \\
1! + 2! + 3! &= 3^2 \\
1! + 4! &= 5^2 \\
1! + 5! &= 11^2 \\
4! + 5! &= 12^2 \\
1! + 2! + 3! + 6! &= 27^2 \\
1! + 5! + 6! &= 29^2 \\
1! + 7! &= 71^2 \\
4! + 5! + 7! &= 72^2 \\
1! + 2! + 3! + 7! + 8! &= 213^2 \\
1! + 4! + 5! + 6! + 7! + 8! &= 215^2 \\
1! + 2! + 3! + 6! + 9! &= 603^2 \\
1! + 4! + 8! + 9! &= 635^2 \\
1! + 2! + 3! + 6! + 7! + 8! + 10! &= 1917^2
\end{aligned}$$

$$\begin{aligned}
&1! + 2! + 3! + 7! + 8! + 9! + 10! + 11! + 12! \\
&\qquad + 13! + 14! + 15! = 1183893^2
\end{aligned}$$

(Sloane's A014597). The first few values for which the alternating SUM

$$\sum_{i=1}^{n} (-1)^{n-i} i! \tag{36}$$

is PRIME are 3, 4, 5, 6, 7, 8, 41, 59, 61, 105, 160, ... (Sloane's A014615, Guy 1994, p. 100). The only known factorials which are products of factorial in an ARITHMETIC SEQUENCE are

$$\begin{aligned}
0!1! &= 1! \\
1!2! &= 2! \\
0!1!2! &= 2! \\
6!7! &= 10! \\
1!3!5! &= 6! \\
1!3!5!7! &= 10!
\end{aligned}$$

(Madachy 1979).

There are no identities of the form

$$n! = a_1!a_2! \cdots a_r! \tag{37}$$

for $r \geq 2$ with $a_i \geq a_j \geq 2$ for $i < j$ for $n \leq 18160$ except

$$9! = 7!3!3!2! \tag{38}$$

$$10! = 7!6! = 7!5!3! \tag{39}$$

$$16! = 14!5!2! \tag{40}$$

(Guy 1994, p. 80).

There are three numbers less than 200,000 for which

$$(n-1)! + 1 \equiv 0 \pmod{n^2}, \tag{41}$$

namely 5, 13, and 563 (Le Lionnais 1983). BROWN NUMBERS are pairs (m, n) of INTEGERS satisfying the condition of BROCARD'S PROBLEM, i.e., such that

$$n! + 1 = m^2. \tag{42}$$

Only three such numbers are known: (5, 4), (11, 5), (71, 7). Erdős conjectured that these are the only three such pairs (Guy 1994, p. 193).

see also ALLADI-GRINSTEAD CONSTANT, BROCARD'S PROBLEM, BROWN NUMBERS, DOUBLE FACTORIAL, FACTORIAL PRIME, FACTORION, GAMMA FUNCTION, HYPERFACTORIAL, MULTIFACTORIAL, POCHHAMMER SYMBOL, PRIMORIAL, ROMAN FACTORIAL, STIRLING'S SERIES, SUBFACTORIAL, SUPERFACTORIAL

References

Conway, J. H. and Guy, R. K. "Factorial Numbers." In *The Book of Numbers.* New York: Springer-Verlag, pp. 65–66, 1996.

Dudeney, H. E. *Amusements in Mathematics.* New York: Dover, p. 96, 1970.

Gardner, M. "Factorial Oddities." Ch. 4 in *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, pp. 50–65, 1978.

Graham, R. L.; Knuth, D. E.; and Patashnik, O. "Factorial Factors." §4.4 in *Concrete Mathematics: A Foundation for Computer Science.* Reading, MA: Addison-Wesley, pp. 111—115, 1990.

Guy, R. K. "Equal Products of Factorials," "Alternating Sums of Factorials," and "Equations Involving Factorial n." §B23, B43, and D25 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 80, 100, and 193–194, 1994.

Honsberger, R. *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., p. 2, 1976.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 56, 1983.

Leyland, P. `ftp://sable.ox.ac.uk/pub/math/factors/factorial-.Z` and `ftp://sable.ox.ac.uk/pub/math/factors/factorial+.Z`.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, p. 174, 1979.

Mudge, M. "Not Numerology but Numeralogy!" *Personal Computer World,* 279–280, 1997.

Ogilvy, C. S. and Anderson, J. T. *Excursions in Number Theory.* New York: Dover, 1988.

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, p. 86, 1996.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Gamma Function, Beta Function, Factorials, Binomial Coefficients." §6.1 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 206–209, 1992.

Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, pp. 22–24, 1989.

Sloane, N. J. A. Sequences A014615, A014597, A033312, A020549, A000142/M1675, and A007489/M2818 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Spanier, J. and Oldham, K. B. "The Factorial Function $n!$ and Its Reciprocal." Ch. 2 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 19–33, 1987.

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 67, 1991.

Factorial Moment

$$\nu_{(r)} \equiv \sum_x x^{(r)} f(x),$$

where

$$x^{(r)} \equiv x(x-1)\cdots(x-r+1).$$

Factorial Number

see FACTORIAL

Factorial Prime

A PRIME of the form $n! \pm 1$. $n!+1$ is PRIME for 1, 2, 3, 11, 27, 37, 41, 73, 77, 116, 154, 320, 340, 399, 427, 872, 1477, ... (Sloane's A002981) up to a search limit 4850. $n!-1$ is PRIME for 3, 4, 6, 7, 12, 14, 30, 32, 33, 38, 94, 116, 324, 379, 469, 546, 974, 1963, 3507, 3610, ... (Sloane's A002982) up to a search limit of 4850.

References

Borning, A. "Some Results for $k!+1$ and $2\cdot 3\cdot 5\cdot p+1$." *Math. Comput.* **26**, 567–570, 1972.

Buhler, J. P.; Crandall, R. E.; and Penk, M. A. "Primes of the Form $M!+1$ and $2\cdot3\cdot5\cdots p+1$." *Math. Comput.* **38**, 639–643, 1982.

Caldwell, C. K. "On the Primality of $N!\pm1$ and $2\cdot3\cdot5\cdots p\pm 1$." *Math. Comput.* **64**, 889–890, 1995.

Dubner, H. "Factorial and Primorial Primes." *J. Rec. Math.* **19**, 197–203, 1987.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 7, 1994.

Sloane, N. J. A. Sequences A002981/M0908 and A002982/M2321 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Temper, M. "On the Primality of $k!+1$ and $\cdot 3\cdot 5\cdots p+1$." *Math. Comput.* **34**, 303–304, 1980.

Factorial Sum

Sums with unity NUMERATOR and FACTORIALS in the DENOMINATOR which can be expressed analytically include

$$\sum_{i=1}^{n} \frac{1}{(n+i-k)!(n-i)!} = \frac{{}_2F_1(1,-n;1+n-k;-1)-1}{\Gamma(1+n)\Gamma(1+n-k)} \tag{1}$$

$$\sum_{i=1}^{n} \frac{1}{(n+i-1)!(n-i)!} = \frac{n\sqrt{\pi}}{2\Gamma(\frac{1}{2}+n)\Gamma(1+n)} \tag{2}$$

$$\sum_{i=1}^{n} \frac{1}{(n+i)!(n-i)!} = \frac{\sqrt{\pi}}{2\Gamma(\frac{1}{2}+n)\Gamma(1+n)} - \frac{1}{2\Gamma^2(1+n)} \tag{3}$$

$$\sum_{i=1}^{n} \frac{1}{(n+i+1)!(n-i)!} = \frac{\sqrt{\pi}}{2\Gamma(\frac{3}{2}+n)\Gamma(1+n)} - \frac{1}{2\Gamma(1+n)\Gamma(2+n)}, \tag{4}$$

where ${}_2F_1(a,b;c;z)$ is a HYPERGEOMETRIC FUNCTION and $\Gamma(z)$ is a GAMMA FUNCTION.

Sums with i in the NUMERATOR having analytic solutions include

$$\sum_{i=1}^{n} \frac{i}{(n+i-k)!(n-i)!} = \frac{n\,{}_2F_1(2,1-n;2-k+n;-1)}{(1-k+n)\Gamma(1+n)\Gamma(1-k+n)} \tag{5}$$

$$\sum_{i=1}^{n} \frac{i}{(n+i-1)!(n-i)!} = \frac{1}{2\Gamma(n)}\left[\frac{\sqrt{\pi}}{2\Gamma(\frac{1}{2}+n)} + \frac{n}{\Gamma(1+n)}\right] \tag{6}$$

$$\sum_{i=1}^{n} \frac{i}{(n+i)!(n-i)!} = \frac{n}{2\Gamma^2(1+n)} \tag{7}$$

$$\sum_{i=1}^{n} \frac{i}{(n+i+1)!(n-i)!} = \frac{1}{2\Gamma(1+n)} \left[\frac{1}{\Gamma(2+n)} - \frac{(n^2+3n+2)\sqrt{\pi}}{2\Gamma(\frac{3}{2}+n)} \right]. \quad (8)$$

A sum with i^2 in the NUMERATOR is

$$\begin{aligned} \sum_{i=1}^{n} \frac{i^2}{(n+i-k)!(n-i)!} &= \frac{n}{(1-k+n)(2-k+n)\Gamma(1+n)\Gamma(1-k+n)} \\ &\times[(2-k+n)\,{}_2F_1(2,1-n;2-k+n;-1) \\ &\quad +2(n-1)\,{}_2F_1(3,2-n;3-k+n;-1)], \end{aligned} \quad (9)$$

where ${}_2F_1(a,b;c;z)$ is the HYPERGEOMETRIC FUNCTION.

Sums of factorial POWERS include

$$\sum_{n=0}^{\infty} \frac{(n!)^2}{(2n)!} = \frac{4}{3} + \frac{2\pi}{9\sqrt{3}} \quad (10)$$

$$\sum_{n=0}^{\infty} \frac{(n!)^3}{(3n)!} = \int_0^1 [P(t) + Q(t)\cos^{-1} R(t)]\,dt, \quad (11)$$

where

$$P(t) = \frac{2(8+7t^2-7t^3)}{(4-t^2+t^3)^2} \quad (12)$$

$$Q(t) = \frac{4t(1-t)(5+t^2-t^3)}{(4-t^2+t^3)^2\sqrt{(1-t)(4-t^2+t^3)}} \quad (13)$$

$$R(t) = 1 - \tfrac{1}{2}(t^2-t^3) \quad (14)$$

(Beeler *et al.* 1972, Item 116).

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Factoring

see FACTORIZATION

Factorion

A factorion is an INTEGER which is equal to the sum of FACTORIALS of its digits. There are exactly four such numbers:

$$1 = 1! \quad (1)$$

$$2 = 2! \quad (2)$$

$$145 = 1! + 4! + 5! \quad (3)$$

$$40{,}585 = 4! + 0! + 5! + 8! + 5! \quad (4)$$

(Gardner 1978, Madachy 1979, Pickover 1995). The factorion of an n-digit number cannot exceed $n \cdot 9!$ digits.

see also FACTORIAL

References

Gardner, M. "Factorial Oddities." Ch. 4 in *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, pp. 61 and 64, 1978.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, p. 167, 1979.

Pickover, C. A. "The Loneliness of the Factorions." Ch. 22 in *Keys to Infinity.* New York: W. H. Freeman, pp. 169–171 and 319–320, 1995.

Factorization

The finding of FACTORS (DIVISORS) of a given INTEGER, POLYNOMIAL, etc. Factorization is also called FACTORING.

see also FACTOR, PRIME FACTORIZATION ALGORITHMS

Fagnano's Point

The point of coincidence of P and P' in FAGNANO'S PROBLEM.

Fagnano's Problem

In a given ACUTE-angled TRIANGLE ΔABC, INSCRIBE another TRIANGLE whose PERIMETER is as small as possible. The answer is the PEDAL TRIANGLE of ΔABC.

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 88–89, 1967.

Fagnano's Theorem

If $P(x,y)$ and $P(x',y')$ are two points on an ELLIPSE

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad (1)$$

with ECCENTRIC ANGLES ϕ and ϕ' such that

$$\tan\phi\tan\phi' = \frac{b}{a} \quad (2)$$

and $A = P(a,0)$ and $B = P(0,b)$. Then

$$\operatorname{arc} BP + \operatorname{arc} BP' = \frac{e^2 xx'}{a}. \quad (3)$$

This follows from the identity

$$E(u,k) + E(v,k) - E(k) = k^2 \operatorname{sn}(u,k)\operatorname{sn}(v,k), \quad (4)$$

where $E(u,k)$ is an incomplete ELLIPTIC INTEGRAL OF THE SECOND KIND, $E(k)$ is a complete ELLIPTIC INTEGRAL OF THE SECOND KIND, and $\operatorname{sn}(v,k)$ is a JACOBI ELLIPTIC FUNCTION. If P and P' coincide, the point where they coincide is called FAGNANO'S POINT.

Fair Game

A GAME which is not biased toward any player.

see also GAME, MARTINGALE

Fairy Chess

A variation of CHESS involving a change in the form of the board, the rules of play, or the pieces used. For example, the normal rules of chess can be used but with a cylindrical or MÖBIUS STRIP connection of the edges.

see also CHESS

References
Kraitchik, M. "Fairy Chess." §12.2 in *Mathematical Recreations.* New York: W. W. Norton, pp. 276–279, 1942.

Fallacy

A fallacy is an incorrect result arrived at by apparently correct, though actually specious reasoning. The most common example of a mathematical fallacy is the "proof" that $1 = 2$ as follows. Let $a = b$, then

$$ab = a^2$$

$$ab - b^2 = a^2 - b^2$$

$$b(a - b) = (a + b)(a - b)$$

$$b = a + b$$

$$b = 2b$$

$$1 = 2.$$

The incorrect step is division by $a - b$ (equal to 0), which is invalid. Ball and Coxeter (1987) give other such examples in the areas of both arithmetic and geometry.

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 41–45 and 76–84, 1987.
Pappas, T. "Geometric Fallacy & the Fibonacci Sequence." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 191, 1989.

False

A statement which is rigorously not TRUE. Regular two-valued LOGIC allows statements to be only TRUE or false, but FUZZY LOGIC treats "truth" as a continuum which can have a value between 0 and 1.

see also ALETHIC, FUZZY LOGIC, LOGIC, TRUE, TRUTH TABLE, UNDECIDABLE

False Position Method

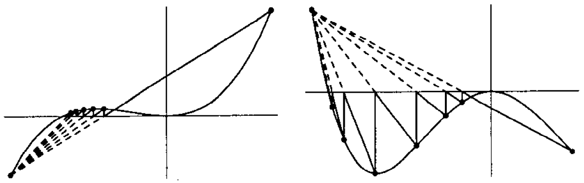

An ALGORITHM for finding ROOTS which uses the point where the linear approximation crosses the axis as the next iteration and keeps the same initial point for each iteration. Using the two-point form of the line

$$y - y_1 = \frac{f(x_{n-1}) - f(x_1)}{x_{n-1} - x_1}(x_n - x_1)$$

with $y = 0$, using $y_1 = f(x_1)$, and solving for x_n therefore gives the iteration

$$x_n = x_1 - \frac{x_{n-1} - x_1}{f(x_{n-1}) - f(x_1)} f(x_1).$$

see also BRENT'S METHOD, RIDDERS' METHOD, SECANT METHOD

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 18, 1972.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Secant Method, False Position Method, and Ridders' Method." §9.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 347–352, 1992.

Faltung (Form)

Let A and B be bilinear forms

$$A = A(x, y) = \sum\sum a_{ij} x_i y_i$$
$$B = B(x, y) = \sum\sum b_{ij} x_i y_i$$

and suppose that A and B are bounded in $[p, p']$ with bounds M and N. Then

$$F = F(A, B) = \sum\sum f_{ij} x_i y_j,$$

where the series

$$f_{ij} = \sum_k a_{ik} b_{kj}$$

is absolutely convergent, is called the faltung of A and B. F is bounded in $[p, p']$, and its bound does not exceed MN.

References
Hardy, G. H.; Littlewood, J. E.; and Pólya, G. *Inequalities, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 210–211, 1988.

Faltung (Function)

see CONVOLUTION

Fan

A SPREAD in which each node has a FINITE number of children.

see also SPREAD (TREE)

Fano's Axiom

The three diagonal points of a COMPLETE QUADRILATERAL are never COLLINEAR.

Fano Plane

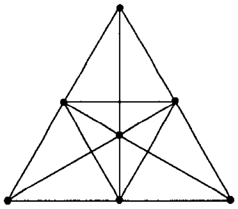

The 2-D PROJECTIVE PLANE over $GF(2)$ ("of order two"), illustrated above. It is a BLOCK DESIGN with $\nu = 7$, $k = 3$, $\lambda = 1$, $r = 3$, and $b = 7$, and is also the STEINER TRIPLE SYSTEM $S(7)$.

The Fano plane also solves the TRANSYLVANIA LOTTERY, which picks three numbers from the INTEGERS 1–14. Using two Fano planes we can guarantee matching two by playing just 14 times as follows. Label the VERTICES of one Fano plane by the INTEGERS 1–7, the other plane by the INTEGERS 8–14. The 14 tickets to play are the 14 lines of the two planes. Then if (a, b, c) is the winning ticket, at least two of a, b, c are either in the interval [1, 7] or [8, 14]. These two numbers are on exactly one line of the corresponding plane, so one of our tickets matches them.

The Lehmers (1974) found an application of the Fano plane for factoring INTEGERS via QUADRATIC FORMS. Here, the triples of forms used form the lines of the PROJECTIVE GEOMETRY on seven points, whose planes are Fano configurations corresponding to pairs of residue classes mod 24 (Lehmer and Lehmer 1974, Guy 1975, Shanks 1985). The group of AUTOMORPHISMS (incidence-preserving BIJECTIONS) of the Fano plane is the SIMPLE GROUP of ORDER 168 (Klein 1870).

see also DESIGN, PROJECTIVE PLANE, STEINER TRIPLE SYSTEM, TRANSYLVANIA LOTTERY

References

Guy, R. "How to Factor a Number." *Proc. Fifth Manitoba Conf. on Numerical Math.*, 49–89, 1975.

Lehmer, D. H. and Lehmer, E. "A New Factorization Technique Using Quadratic Forms." *Math. Comput.* **28**, 625–635, 1974.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 3rd ed.* New York: Chelsea, pp. 202 and 238, 1985.

Far Out

A word used by Tukey to describe data points which are outside the outer FENCES.

References

Tukey, J. W. *Explanatory Data Analysis.* Reading, MA: Addison-Wesley, p. 44, 1977.

Far-Out Point

For a TRIANGLE with side lengths a, b, and c, the far-out point has TRIANGLE CENTER FUNCTION

$$\alpha = a(b^4 + c^4 - a^4 - b^2c^2).$$

As $a : b : c$ approaches $1 : 1 : 1$, this point moves out along the EULER LINE to infinity.

References

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Kimberling, C.; Lyness, R. C.; and Veldkamp, G. R. "Problem 1195 and Solution." *Crux Math.* **14**, 177–179, 1988.

Farey Sequence

The Farey sequence F_n for any POSITIVE INTEGER n is the set of irreducible RATIONAL NUMBERS a/b with $0 \leq a \leq b \leq n$ and $(a, b) = 1$ arranged in increasing order.

$$F_1 = \{\tfrac{0}{1}, \tfrac{1}{1}\} \tag{1}$$

$$F_2 = \{\tfrac{0}{1}, \tfrac{1}{2}, \tfrac{1}{1}\} \tag{2}$$

$$F_3 = \{\tfrac{0}{1}, \tfrac{1}{3}, \tfrac{1}{2}, \tfrac{2}{3}, \tfrac{1}{1}\} \tag{3}$$

$$F_4 = \{\tfrac{0}{1}, \tfrac{1}{4}, \tfrac{1}{3}, \tfrac{1}{2}, \tfrac{2}{3}, \tfrac{3}{4}, \tfrac{1}{1}\} \tag{4}$$

$$F_5 = \{\tfrac{0}{1}, \tfrac{1}{5}, \tfrac{1}{4}, \tfrac{1}{3}, \tfrac{2}{5}, \tfrac{1}{2}, \tfrac{3}{5}, \tfrac{2}{3}, \tfrac{3}{4}, \tfrac{4}{5}, \tfrac{1}{1}\}. \tag{5}$$

There is always an ODD number of terms, and the middle term is always 1/2. Let p/q, p'/q', and p''/q'' be three successive terms in a Farey series. Then

$$qp' - pq' = 1 \tag{6}$$

$$\frac{p'}{q'} = \frac{p + p''}{q + q''}. \tag{7}$$

These two statements are actually equivalent.

The number of terms $N(n)$ in the Farey sequence for the INTEGER n is

$$N(n) = 1 + \sum_{k=1}^{n} \phi(k) = 1 + \Phi(n), \tag{8}$$

where $\phi(k)$ is the TOTIENT FUNCTION and $\Phi(n)$ is the SUMMATORY FUNCTION of $\phi(k)$, giving 2, 3, 5, 7, 11, 13, 19, ... (Sloane's A005728). The asymptotic limit for the function $N(n)$ is

$$N(n) \sim \frac{3n^2}{\pi^2} = 0.3039635509n^2 \tag{9}$$

(Vardi 1991, p. 155). For a method of computing a successive sequence from an existing one of n terms, insert the MEDIANT fraction $(a + b)/(c + d)$ between terms a/c and b/d when $c + d \leq n$ (Hardy and Wright 1979, pp. 25–26; Conway and Guy 1996).

FORD CIRCLES provide a method of visualizing the Farey sequence. The Farey sequence F_n defines a subtree of the STERN-BROCOT TREE obtained by pruning unwanted branches (Graham *et al.* 1994).

see also FORD CIRCLE, MEDIANT, RANK (SEQUENCE), STERN-BROCOT TREE

References

Beiler, A. H. "Farey Tails." Ch. 16 in *Recreations in the Theory of Numbers: The Queen of Mathematics Entertains.* New York: Dover, 1966.

Conway, J. H. and Guy, R. K. "Farey Fractions and Ford Circles." *The Book of Numbers.* New York: Springer-Verlag, pp. 152–154 and 156, 1996.

Dickson, L. E. *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, pp. 155–158, 1952.

Farey, J. "On a Curious Property of Vulgar Fractions." *London, Edinburgh and Dublin Phil. Mag.* **47**, 385, 1816.

Graham, R. L.; Knuth, D. E.; and Patashnik, O. *Concrete Mathematics: A Foundation for Computer Science, 2nd ed.* Reading, MA: Addison-Wesley, pp. 118–119, 1994.

Guy, R. K. "Mahler's Generalization of Farey Series." §F27 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 263–265, 1994.

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, 1979.

Sloane, N. J. A. Sequences A005728/M0661, A006842/M0041, and A006843/M0081 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sylvester, J. J. "On the Number of Fractions Contained in Any Farey Series of Which the Limiting Number is Given." *London, Edinburgh and Dublin Phil. Mag. (5th Series)* **15**, 251, 1883.

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 155, 1991.

Weisstein, E. W. "Plane Geometry." `http://www.astro.virginia.edu/~eww6n/math/notebooks/PlaneGeometry.m`.

Farey Series

see FAREY SEQUENCE

Farkas's Lemma

The INEQUALITY $\langle f_0, x \rangle \leq 0$ follows from

$$\langle f_1, x \rangle \leq 0, \ldots, \langle f_n, x \rangle \leq 0$$

IFF there exist NONNEGATIVE numbers $\lambda_1, \ldots, \lambda_n$ with

$$\sum_{k=1}^{n} \lambda_k f_k = f_0.$$

This LEMMA is used in the proof of the KUHN-TUCKER THEOREM.

see also KUHN-TUCKER THEOREM, LAGRANGE MULTIPLIER

Faro Shuffle

see RIFFLE SHUFFLE

Fast Fibonacci Transform

For a general second-order recurrence equation

$$f_{n+1} = x f_n + y f_{n-1}, \tag{1}$$

define a multiplication rule on ordered pairs by

$$(A, B)(C, D) = (AD + BC + xAC, BD + yAC). \tag{2}$$

The inverse is then given by

$$(A, B)^{-1} = \frac{(-A, xA + B)}{B^2 + xAB - yA^2}, \tag{3}$$

and we have the identity

$$(f_1, y f_0)(1, 0)^n = (f_{n+1}, y f_n) \tag{4}$$

(Beeler *et al.* 1972, Item 12).

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Fast Fourier Transform

The fast Fourier transform (FFT) is a DISCRETE FOURIER TRANSFORM ALGORITHM which reduces the number of computations needed for N points from $2N^2$ to $2N \lg N$, where LG is the base-2 LOGARITHM. If the function to be transformed is not harmonically related to the sampling frequency, the response of an FFT looks like a SINC FUNCTION (although the integrated POWER is still correct). ALIASING (LEAKAGE) can be reduced by APODIZATION using a TAPERING FUNCTION. However, ALIASING reduction is at the expense of broadening the spectral response.

FFTs were first discussed by Cooley and Tukey (1965), although Gauss had actually described the critical factorization step as early as 1805 (Gergkand 1969, Strang 1993). A DISCRETE FOURIER TRANSFORM can be computed using an FFT by means of the DANIELSON-LANCZOS LEMMA if the number of points N is a POWER of two. If the number of points N is not a POWER of two, a transform can be performed on sets of points corresponding to the prime factors of N which is slightly degraded in speed. An efficient real Fourier transform algorithm or a fast HARTLEY TRANSFORM (Bracewell 1965) gives a further increase in speed by approximately a factor of two. Base-4 and base-8 fast Fourier transforms use optimized code, and can be 20–30% faster than base-2 fast Fourier transforms. PRIME factorization is slow when the factors are large, but discrete Fourier transforms can be made fast for N = 2, 3, 4, 5, 7, 8, 11, 13, and 16 using the WINOGRAD TRANSFORM ALGORITHM (Press *et al.* 1992, pp. 412–413, Arndt).

Fast Fourier transform algorithms generally fall into two classes: decimation in time, and decimation in frequency. The Cooley-Tukey FFT ALGORITHM first rearranges the input elements in bit-reversed order, then builds the output transform (decimation in time). The basic idea is to break up a transform of length N into two transforms of length $N/2$ using the identity

$$\sum_{n=0}^{N-1} a_n e^{-2\pi ink/N} = \sum_{n=0}^{N/2-1} a_{2n} e^{-2\pi i(2n)k/N} + \sum_{n=0}^{N/2-1} a_{2n+1} e^{-2\pi i(2n+1)k/N}$$

$$= \sum_{n=0}^{N/2-1} a_n^{\text{even}} e^{-2\pi ink/(N/2)} + e^{-2\pi ik/N} \sum_{n=0}^{N/2-1} a_n^{\text{odd}} e^{-2\pi ink/(N/2)},$$

sometimes called the DANIELSON-LANCZOS LEMMA. The easiest way to visualize this procedure is perhaps via the FOURIER MATRIX.

The Sande-Tukey ALGORITHM (Stoer and Burlisch 1980) first transforms, then rearranges the output values (decimation in frequency).

see also DANIELSON-LANCZOS LEMMA, DISCRETE FOURIER TRANSFORM, FOURIER MATRIX, FOURIER TRANSFORM, HARTLEY TRANSFORM, NUMBER THEORETIC TRANSFORM, WINOGRAD TRANSFORM

References

Arndt, J. "FFT Code and Related Stuff." `http://www.jjj.de/fxt/`.
Bell Laboratories. "Netlib FFTPack." `http://netlib.bell-labs.com/netlib/fftpack/`.
Blahut, R. E. *Fast Algorithms for Digital Signal Processing.* New York: Addison-Wesley, 1984.
Bracewell, R. *The Fourier Transform and Its Applications.* New York: McGraw-Hill, 1965.
Brigham, E. O. *The Fast Fourier Transform and Applications.* Englewood Cliffs, NJ: Prentice Hall, 1988.
Cooley, J. W. and Tukey, O. W. "An Algorithm for the Machine Calculation of Complex Fourier Series." *Math. Comput.* **19**, 297–301, 1965.
Duhamel, P. and Vetterli, M. "Fast Fourier Transforms: A Tutorial Review." *Signal Processing* **19**, 259–299, 1990.
Gergkand, G. D. "A Guided Tour of the Fast Fourier Transform." *IEEE Spectrum,* pp. 41–52, July 1969.
Lipson, J. D. *Elements of Algebra and Algebraic Computing.* Reading, MA: Addison-Wesley, 1981.
Nussbaumer, H. J. *Fast Fourier Transform and Convolution Algorithms, 2nd ed.* New York: Springer-Verlag, 1982.
Papoulis, A. *The Fourier Integral and its Applications.* New York: McGraw-Hill, 1962.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Fast Fourier Transform." Ch. 12 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 490–529, 1992.
Stoer, J. and Burlisch, R. *Introduction to Numerical Analysis.* New York: Springer-Verlag, 1980.
Strang, G. "Wavelet Transforms Versus Fourier Transforms." *Bull. Amer. Math. Soc.* **28**, 288–305, 1993.
Van Loan, C. *Computational Frameworks for the Fast Fourier Transform.* Philadelphia, PA: SIAM, 1992.
Walker, J. S. *Fast Fourier Transform, 2nd ed.* Boca Raton, FL: CRC Press, 1996.

Fat Fractal

A CANTOR SET with LEBESGUE MEASURE greater than 0.

see also CANTOR SET, EXTERIOR DERIVATIVE, FRACTAL, LEBESGUE MEASURE

References

Ott, E. "Fat Fractals." §3.9 in *Chaos in Dynamical Systems.* New York: Cambridge University Press, pp. 97–100, 1993.

Fatou Dust

see FATOU SET

Fatou's Lemma

If a SEQUENCE $\{f_n\}$ of NONNEGATIVE measurable functions is defined on a measurable set E, then

$$\int_E \liminf_{n\to\infty} f_n \, d\mu \leq \liminf_{n\to\infty} \int_E f_n \, d\mu.$$

References

Zeidler, E. *Applied Functional Analysis: Applications to Mathematical Physics.* New York: Springer-Verlag, 1995.

Fatou Set

A set consisting of the complementary set of complex numbers to a JULIA SET.

see also JULIA SET

References

Schroeder, M. *Fractals, Chaos, Power Laws.* New York: W. H. Freeman, p. 39, 1991.

Fatou's Theorems

Let $f(\theta)$ be LEBESGUE INTEGRABLE and let

$$f(r,\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \frac{1-r^2}{1-2r\cos(t-\theta)+r^2} \, dt \qquad (1)$$

be the corresponding POISSON INTEGRAL. Then ALMOST EVERYWHERE in $-\pi \leq \theta \leq \pi$,

$$\lim_{r\to 0^-} f(r,\theta) = f(\theta). \qquad (2)$$

Let

$$F(z) = c_0 + c_1 z + c_2 z^2 + \ldots + c_n z^n + \ldots \qquad (3)$$

be regular for $|z| < 1$, and let the integral

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |F(re^{i\theta})|^2 \, d\theta \qquad (4)$$

be bounded for $r < 1$. This condition is equivalent to the convergence of

$$|c_0|^2 + |c_1|^2 + \ldots + |c_n|^2 + \ldots. \qquad (5)$$

Then almost everywhere in $-\pi \leq \theta \leq \pi$,

$$\lim_{r\to 0^-} F(re^{i\theta}) = F(e^{i\theta}). \tag{6}$$

Furthermore, $F(e^{i\theta})$ is measurable, $|F(e^{i\theta})|^2$ is LEBESGUE INTEGRABLE, and the FOURIER SERIES of $F(e^{i\theta})$ is given by writing $z = e^{i\theta}$.

References

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., p. 274, 1975.

Faulhaber's Formula

In a 1631 edition of *Academiae Algebrae*, J. Faulhaber published the general formula for the SUM of pth POWERS of the first n POSITIVE INTEGERS,

$$\sum_{k=1}^{n} k^p = \frac{1}{p+1}\sum_{i=1}^{p+1}(-1)^{\delta_{ip}}\binom{p+1}{i}B_{p+1-i}n^i, \tag{1}$$

where δ_{ip} is the KRONECKER DELTA, $\binom{n}{i}$ is a BINOMIAL COEFFICIENT, and B_i is the ith BERNOULLI NUMBER. Computing the sums for $p = 1, \ldots, 10$ gives

$$\sum_{k=1}^{n} k = \tfrac{1}{2}(n^2+n) \tag{2}$$

$$\sum_{k=1}^{n} k^2 = \tfrac{1}{6}(2n^3+3n^2+n) \tag{3}$$

$$\sum_{k=1}^{n} k^3 = \tfrac{1}{4}(n^4+2n^3+n^2) \tag{4}$$

$$\sum_{k=1}^{n} k^4 = \tfrac{1}{30}(6n^5+15n^4+10n^3-n) \tag{5}$$

$$\sum_{k=1}^{n} k^5 = \tfrac{1}{12}(2n^6+6n^5+5n^4-n^2) \tag{6}$$

$$\sum_{k=1}^{n} k^6 = \tfrac{1}{42}(6n^7+21n^6+21n^5-7n^3+n) \tag{7}$$

$$\sum_{k=1}^{n} k^7 = \tfrac{1}{24}(3n^8+12n^7+14n^6-7n^4+2n^2) \tag{8}$$

$$\begin{aligned}\sum_{k=1}^{n} k^8 = {} & \tfrac{1}{90}(10n^9+45n^8+60n^7-42n^5 \\ & +20n^3-3n)\end{aligned} \tag{9}$$

$$\begin{aligned}\sum_{k=1}^{n} k^9 = {} & \tfrac{1}{20}(2n^{10}+10n^9+15n^8-14n^6 \\ & +10n^4-3n^2)\end{aligned} \tag{10}$$

$$\begin{aligned}\sum_{k=1}^{n} k^{10} = {} & \tfrac{1}{66}(6n^{11}+33n^{10}+55n^9-66n^7 \\ & +66n^5-33n^3+5n).\end{aligned} \tag{11}$$

see also POWER, SUM

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 106, 1996.

Favard Constants

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let $T_n(x)$ be an arbitrary trigonometric POLYNOMIAL

$$T_n(x) = \tfrac{1}{2}a_0 + \left\{\sum_{k=1}^{n}[a_k\cos(kx) + b_k\sin(kx)]\right\},$$

where the COEFFICIENTS are real. Let the rth derivative of $T_n(x)$ be bounded in $[-1,1]$, then there exists a POLYNOMIAL $T_n(x)$ for which

$$|f(x) - T_n(x)| \leq \frac{K_r}{(n+1)^r},$$

for all x, where K_r is the rth Favard constant, which is the smallest constant possible.

$$K_r = \frac{4}{\pi}\sum_{k=0}^{\infty}\left[\frac{(-1)^k}{2k+1}\right]^{r+1}.$$

These can be expressed by

$$K_r = \begin{cases} \frac{4}{\pi}\lambda(r+1) & \text{for } r \text{ odd} \\ \frac{4}{\pi}\beta(r+1) & \text{for } r \text{ even,} \end{cases}$$

where λ is the DIRICHLET LAMBDA FUNCTION and β is the DIRICHLET BETA FUNCTION. Explicitly,

$$\begin{aligned} K_0 &= 1 \\ K_1 &= \tfrac{1}{2}\pi \\ K_2 &= \tfrac{1}{8}\pi^2 \\ K_3 &= \tfrac{1}{24}\pi^3. \end{aligned}$$

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/favard/favard.html`.

Kolmogorov, A. N. "Zur Grössenordnung des Restgliedes Fourierscher reihen differenzierbarer Funktionen." *Ann. Math.* **36**, 521–526, 1935.

Zygmund, A. G. *Trigonometric Series, Vols. 1–2, 2nd ed.* New York: Cambridge University Press, 1959.

Feigenbaum Constant

A universal constant for functions approaching CHAOS via period doubling. It was discovered by Feigenbaum in 1975 and demonstrated rigorously by Lanford (1982) and Collet and Eckmann (1979, 1980). The Feigenbaum constant δ characterizes the geometric approach of the bifurcation parameter to its limiting value. Let μ_k be the point at which a period 2^k cycle becomes unstable.

Denote the converged value by μ_∞. Assuming geometric convergence, the difference between this value and μ_k is denoted

$$\lim_{k\to\infty} \mu_\infty - \mu_k = \frac{\Gamma}{\delta^k}, \tag{1}$$

where Γ is a constant and δ is a constant > 1. Solving for δ gives

$$\delta \equiv \lim_{n\to\infty} \frac{\mu_{n+1} - \mu_n}{\mu_{n+2} - \mu_{n+1}} \tag{2}$$

(Rasband 1990, p. 23). For the LOGISTIC EQUATION,

$$\delta = 4.669216091\ldots \tag{3}$$

$$\Gamma = 2.637\ldots \tag{4}$$

$$\mu_\infty = 3.5699456\ldots. \tag{5}$$

Amazingly, the Feigenbaum constant $\delta \approx 4.669$ is "universal" (i.e., the same) for all 1-D MAPS $f(x)$ if $f(x)$ has a single locally quadratic MAXIMUM. More specifically, the Feigenbaum constant is universal for 1-D MAPS if the SCHWARZIAN DERIVATIVE

$$D_{\text{Schwarzian}} \equiv \frac{f'''(x)}{f'(x)} - \frac{3}{2}\left[\frac{f''(x)}{f'(x)}\right]^2 \tag{6}$$

is NEGATIVE in the bounded interval (Tabor 1989, p. 220). Examples of maps which are universal include the HÉNON MAP, LOGISTIC MAP, LORENZ SYSTEM, Navier-Stokes truncations, and sine map $x_{n+1} = a\sin(\pi x_n)$. The value of the Feigenbaum constant can be computed explicitly using functional group renormalization theory. The universal constant also occurs in phase transitions in physics and, curiously, is very nearly equal to

$$\pi + \tan^{-1}(e^\pi) = 4.669201932\ldots. \tag{7}$$

The CIRCLE MAP is *not* universal, and has a Feigenbaum constant of $\delta \approx 2.833$. For an AREA-PRESERVING 2-D MAP with

$$x_{n+1} = f(x_n, y_n) \tag{8}$$

$$y_{n+1} = g(x_n, y_n), \tag{9}$$

the Feigenbaum constant is $\delta = 0.7210978\ldots$ (Tabor 1989, p. 225). For a function of the form

$$f(x) = 1 - a|x|^n \tag{10}$$

with a and n constant and n an INTEGER, the Feigenbaum constant for various n is given in the following table (Briggs 1991, Briggs *et al.* 1991), which updates the values in Tabor (1989, p. 225).

n	δ
2	5.9679
4	7.2846
6	8.3494
8	9.2962

An additional constant α, defined as the separation of adjacent elements of PERIOD DOUBLED ATTRACTORS from one double to the next, has a value

$$\lim_{n\to\infty} \frac{d_n}{d_{n+1}} \equiv -\alpha = -2.502907875\ldots \tag{11}$$

for "universal" maps (Rasband 1990, p. 37). This value may be approximated from functional group renormalization theory to the zeroth order by

$$1 - \alpha^{-1} = \frac{1 - \alpha^{-2}}{[1 - \alpha^{-2}(1 - \alpha^{-1})]^2}, \tag{12}$$

which, when the QUINTIC EQUATION is numerically solved, gives $\alpha = -2.48634\ldots$, only 0.7% off from the actual value (Feigenbaum 1988).

see also ATTRACTOR, BIFURCATION, FEIGENBAUM FUNCTION, LINEAR STABILITY, LOGISTIC MAP, PERIOD DOUBLING

References

Briggs, K. "A Precise Calculation of the Feigenbaum Constants." *Math. Comput.* **57**, 435–439, 1991.

Briggs, K.; Quispel, G.; and Thompson, C. "Feigenvalues for Mandelsets." *J. Phys. A: Math. Gen.* **24** 3363–3368, 1991.

Briggs, K.; Quispel, G.; and Thompson, C. "Feigenvalues for Mandelsets." `http://epidem13.plantsci.cam.ac.uk/~kbriggs/`.

Collett, P. and Eckmann, J.-P. "Properties of Continuous Maps of the Interval to Itself." *Mathematical Problems in Theoretical Physics* (Ed. K. Osterwalder). New York: Springer-Verlag, 1979.

Collett, P. and Eckmann, J.-P. *Iterated Maps on the Interval as Dynamical Systems.* Boston, MA: Birkhäuser, 1980.

Eckmann, J.-P. and Wittwer, P. *Computer Methods and Borel Summability Applied to Feigenbaum's Equations.* New York: Springer-Verlag, 1985.

Feigenbaum, M. J. "Presentation Functions, Fixed Points, and a Theory of Scaling Function Dynamics." *J. Stat. Phys.* **52**, 527–569, 1988.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/fgnbaum/fgnbaum.html`.

Finch, S. "Generalized Feigenbaum Constants." `http://www.mathsoft.com/asolve/constant/fgnbaum/general.html`.

Lanford, O. E. "A Computer-Assisted Proof of the Feigenbaum Conjectures." *Bull. Amer. Math. Soc.* **6**, 427–434, 1982.

Rasband, S. N. *Chaotic Dynamics of Nonlinear Systems.* New York: Wiley, 1990.

Stephenson, J. W. and Wang, Y. "Numerical Solution of Feigenbaum's Equation." *Appl. Math. Notes* **15**, 68–78, 1990.

Stephenson, J. W. and Wang, Y. "Relationships Between the Solutions of Feigenbaum's Equations." *Appl. Math. Let.* **4**, 37–39, 1991.

Tabor, M. *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, 1989.

Feigenbaum Function

Consider an arbitrary 1-D MAP

$$x_{n+1} = F(x_n) \tag{1}$$

at the onset of CHAOS. After a suitable rescaling, the Feigenbaum function

$$g(x) = \lim_{n\to\infty} \frac{1}{F^{(2^n)}(0)} F^{(2^n)}(xF^{(2^n)}(0)) \tag{2}$$

is obtained. This function satisfies

$$g(g(x)) = -\frac{1}{\alpha} g(\alpha x), \tag{3}$$

with $\alpha = 2.50290\ldots$, a quantity related to the FEIGENBAUM CONSTANT.

see also BIFURCATION, CHAOS, FEIGENBAUM CONSTANT

References

Grassberger, P. and Procaccia, I. "Measuring the Strangeness of Strange Attractors." *Physica D* **9**, 189–208, 1983.

Feit-Thompson Conjecture

Concerns PRIMES p and q for which $p^q - 1$ and $q^p - 1$ have a common factor. The only (p, q) pair with both values less than 400,000 is (17, 3313), with a common factor 112,643.

References

Wells, D. G. *The Penguin Dictionary of Curious and Interesting Numbers.* London: Penguin, p. 17, 1986.

Feit-Thompson Theorem

Every FINITE SIMPLE GROUP (which is not CYCLIC) has EVEN ORDER, and the ORDER of every FINITE SIMPLE noncommutative group is DOUBLY EVEN, i.e., divisible by 4 (Feit and Thompson 1963).

see also BURNSIDE PROBLEM, FINITE GROUP, ORDER (GROUP), SIMPLE GROUP

References

Feit, W. and Thompson, J. G. "Solvability of Groups of Odd Order." *Pacific J. Math.* **13**, 775–1029, 1963.

Fejes Tóth's Integral

$$\frac{1}{2\pi(n+1)} \int_{-\pi}^{\pi} f(x) \left\{ \frac{\sin[\frac{1}{2}(n+1)x]}{\sin(\frac{1}{2}x)} \right\}^2 dx$$

gives the nth CESÀRO MEAN of the FOURIER SERIES of $f(x)$.

References

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., p. 12, 1975.

Fejes Tóth's Problem

How can n points be distributed on a UNIT SPHERE such that they maximize the minimum distance between any pair of points? In 1943, Fejes Tóth proved that for N points, there always exist two points whose distance d is

$$d \leq \sqrt{4 - \csc^2\left[\frac{\pi N}{6(N-2)}\right]},$$

and that the limit is exact for $N = 3$, 4, 6, and 12.

For two points, the points should be at opposite ends of a DIAMETER. For four points, they should be placed at the VERTICES of an inscribed TETRAHEDRON. There is no best solution for five points since the distance cannot be reduced below that for six points. For six points, they should be placed at the VERTICES of an inscribed OCTAHEDRON. For seven points, the best solution is four equilateral spherical triangles with angles of 80°. For eight points, the best dispersal is *not* the VERTICES of the inscribed CUBE, but of a square ANTIPRISM with equal EDGES. The solution for nine points is eight equilateral spherical triangles with angles of $\cos^{-1}(1/4)$. For 12 points, the solution is an inscribed ICOSAHEDRON.

The general problem has not been solved.

see also THOMSON PROBLEM

References

Ogilvy, C. S. *Excursions in Mathematics.* New York: Dover, p. 99, 1994.

Ogilvy, C. S. Solved by L. Moser. "Minimal Configuration of Five Points on a Sphere." Problem E946. *Amer. Math. Monthly* **58**, 592, 1951.

Schütte, K. and van der Waerden, B. L. "Auf welcher Kügel haben 5, 6, 7, 8 oder 9 Pünkte mit Mindestabstand Eins Platz?" *Math. Ann.* **123**, 96–124, 1951.

Whyte, L. L. "Unique Arrangement of Points on a Sphere." *Amer. Math. Monthly* **59**, 606–611, 1952.

Feller's Coin-Tossing Constants

see COIN TOSSING

Feller-Lévy Condition

Given a sequence of independent random variates X_1, X_2, ..., if ${\sigma_k}^2 = \mathrm{var}(X_k)$ and

$${\rho_n}^2 \equiv \max_{k\leq n}\left(\frac{{\sigma_k}^2}{{s_n}^2}\right),$$

then

$$\lim_{n\to\infty} {\rho_n}^2 = 0.$$

This means that if the LINDEBERG CONDITION holds for the sequence of variates $X_1, \ldots$, then the VARIANCE of an individual term in the sum S_n of X_k is asymptotically negligible. For such sequences, the LINDEBERG CONDITION is NECESSARY as well as SUFFICIENT for the LINDEBERG-FELLER CENTRAL LIMIT THEOREM to hold.

References

Zabell, S. L. "Alan Turing and the Central Limit Theorem." *Amer. Math. Monthly* **102**, 483–494, 1995.

Fence

Values one STEP outside the HINGES are called inner fences, and values two steps outside the HINGES are called outer fences. Tukey calls values outside the outer fences FAR OUT.

see also ADJACENT VALUE

References

Tukey, J. W. *Explanatory Data Analysis.* Reading, MA: Addison-Wesley, p. 44, 1977.

Fence Poset

A PARTIAL ORDER defined by $(i-1, j)$, $(i+1, j)$ for ODD i.

see also PARTIAL ORDER

References

Ruskey, F. "Information on Ideals of Partially Ordered Sets." `http://sue.csc.uvic.ca/~cos/inf/pose/Ideals.html`.

Ferguson-Forcade Algorithm

A practical algorithm for determining if there exist integers a_i for given real numbers x_i such that

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0,$$

or else establish bounds within which no such INTEGER RELATION can exist (Ferguson and Forcade 1979). A nonrecursive variant of the original algorithm was subsequently devised by Ferguson (1987). The Ferguson-Forcade algorithm has shown that there are no algebraic equations of degree ≤ 8 with integer coefficients having Euclidean norms below certain bounds for e/π, $e+\pi$, $\ln\pi$, γ, e^γ, γ/e, γ/π, and $\ln\gamma$, where e is the base for the NATURAL LOGARITHM, π is PI, and γ is the EULER-MASCHERONI CONSTANT (Bailey 1988).

Constant	Bound
e/π	6.1030×10^{14}
$e+\pi$	2.2753×10^{18}
$\ln\pi$	8.7697×10^{9}
γ	3.5739×10^{9}
e^γ	1.6176×10^{17}
γ/e	1.8440×10^{11}
γ/π	6.5403×10^{9}
$\ln\gamma$	2.6881×10^{10}

see also CONSTANT PROBLEM, EUCLIDEAN ALGORITHM, INTEGER RELATION, PSLQ ALGORITHM

References

Bailey, D. H. "Numerical Results on the Transcendence of Constants Involving π, e, and Euler's Constant." *Math. Comput.* **50**, 275–281, 1988.

Ferguson, H. R. P. "A Short Proof of the Existence of Vector Euclidean Algorithms." *Proc. Amer. Math. Soc.* **97**, 8–10, 1986.

Ferguson, H. R. P. "A Non-Inductive GL(n, Z) Algorithm that Constructs Linear Relations for n Z-Linearly Dependent Real Numbers." *J. Algorithms* **8**, 131–145, 1987.

Ferguson, H. R. P. and Forcade, R. W. "Generalization of the Euclidean Algorithm for Real Numbers to All Dimensions Higher than Two." *Bull. Amer. Math. Soc.* **1**, 912–914, 1979.

Fermat $4n+1$ Theorem

Every PRIME of the form $4n+1$ is a sum of two SQUARE NUMBERS in one unique way (up to the order of SUMMANDS). The theorem was stated by Fermat, but the first published proof was by Euler.

see also SIERPIŃSKI'S PRIME SEQUENCE THEOREM, SQUARE NUMBER

References

Hardy, G. H. and Wright, E. M. "Some Notation." Th. 251 in *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, 1979.

Fermat's Algorithm

see FERMAT'S FACTORIZATION METHOD

Fermat Compositeness Test

Uses FERMAT'S LITTLE THEOREM

Fermat's Congruence

see FERMAT'S LITTLE THEOREM

Fermat Conic

A PLANE CURVE of the form $y = x^n$. For $n > 0$, the curve is a generalized PARABOLA; for $n < 0$ it is a generalized HYPERBOLA.

see also CONIC SECTION, HYPERBOLA, PARABOLA

Fermat's Conjecture

see FERMAT'S LAST THEOREM

Fermat Difference Equation

see PELL EQUATION

Fermat Diophantine Equation

see FERMAT DIFFERENCE EQUATION

Fermat Equation

The DIOPHANTINE EQUATION

$$x^n + y^n = z^n.$$

The assertion that this equation has no nontrivial solutions for $n > 2$ is called FERMAT'S LAST THEOREM.

see also FERMAT'S LAST THEOREM

Fermat-Euler Theorem

see FERMAT'S LITTLE THEOREM

Fermat's Factorization Method

Given a number n, look for INTEGERS x and y such that $n = x^2 - y^2$. Then

$$n = (x-y)(x+y) \tag{1}$$

and n is factored. Any ODD NUMBER can be represented in this form since then $n = ab$, a and b are ODD, and

$$a = x + y \tag{2}$$

$$b = x - y. \tag{3}$$

Adding and subtracting,

$$a + b = 2x \tag{4}$$

$$a - b = 2y, \tag{5}$$

so solving for x and y gives

$$x = \tfrac{1}{2}(a+b) \tag{6}$$

$$y = \tfrac{1}{2}(a-b). \tag{7}$$

Therefore,

$$x^2 - y^2 = \tfrac{1}{4}[(a+b)^2 - (a-b)^2] = ab. \tag{8}$$

As the first trial for x, try $x_1 \lceil\sqrt{n}\rceil$, where $\lceil x\rceil$ is the CEILING FUNCTION. Then check if

$$\Delta x_1 = {x_1}^2 - n \tag{9}$$

is a SQUARE NUMBER. There are only 22 combinations of the last two digits which a SQUARE NUMBER can assume, so most combinations can be eliminated. If Δx_1 is not a SQUARE NUMBER, then try

$$x_2 = x_1 + 1, \tag{10}$$

so

$$\begin{aligned}\Delta x_2 &= {x_2}^2 - n = (x_1+1)^2 - n = {x_1}^2 + 2x_1 + 1 - n \\ &= \Delta x_1 + 2x_1 + 1. \end{aligned} \tag{11}$$

Continue with

$$\begin{aligned}\Delta x_3 &= {x_3}^2 - n = (x_2+1)^2 - n = {x_2}^2 + 2x_2 + 1 - n \\ &= \Delta x_2 + 2x_2 + 1 = \Delta x_2 + 2x_1 + 3, \end{aligned} \tag{12}$$

so subsequent differences are obtained simply by adding two.

Maurice Kraitchik sped up the ALGORITHM by looking for x and y satisfying

$$x^2 \equiv y^2 \pmod{n}, \tag{13}$$

i.e., $n|(x^2 - y^2)$. This congruence has uninteresting solutions $x \equiv \pm y \pmod{n}$ and interesting solutions $x \not\equiv \pm y \pmod{n}$. It turns out that if n is ODD and DIVISIBLE by at least two different PRIMES, then at least half of the solutions to $x^2 \equiv y^2 \pmod{n}$ with xy CO-PRIME to n are interesting. For such solutions, $(n, x-y)$ is neither n nor 1 and is therefore a nontrivial factor of n (Pomerance 1996). This ALGORITHM can be used to prove primality, but is not practical. In 1931, Lehmer and Powers discovered how to search for such pairs using CONTINUED FRACTIONS. This method was improved by Morrison and Brillhart (1975) into the CONTINUED FRACTION FACTORIZATION ALGORITHM, which was the fastest ALGORITHM in use before the QUADRATIC SIEVE FACTORIZATION METHOD was developed.

see also PRIME FACTORIZATION ALGORITHMS, SMOOTH NUMBER

References

Lehmer, D. H. and Powers, R. E. "On Factoring Large Numbers." *Bull. Amer. Math. Soc.* **37**, 770–776, 1931.

Morrison, M. A. and Brillhart, J. "A Method of Factoring and the Factorization of F_7." *Math. Comput.* **29**, 183–205, 1975.

Pomerance, C. "A Tale of Two Sieves." *Not. Amer. Math. Soc.* **43**, 1473–1485, 1996.

Fermat's Last Theorem

A theorem first proposed by Fermat in the form of a note scribbled in the margin of his copy of the ancient Greek text *Arithmetica* by Diophantus. The scribbled note was discovered posthumously, and the original is now lost. However, a copy was preserved in a book published by Fermat's son. In the note, Fermat claimed to have discovered a proof that the DIOPHANTINE EQUATION $x^n + y^n = z^n$ has no INTEGER solutions for $n > 2$.

The full text of Fermat's statement, written in Latin, reads "Cubum autem in duos cubos, aut quadratoquadratum in duos quadratoquadratos & generaliter nullam in infinitum ultra quadratum potestatem in duos eiusdem nominis fas est diuidere cuius rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet." In translation, "It is impossible for a cube to be the sum of two cubes, a fourth power to be the sum of two fourth powers, or in general for any number that is a power greater than the second to be the sum of two like powers. I have discovered a truly marvelous demonstration of this proposition that this margin is too narrow to contain."

As a result of Fermat's marginal note, the proposition that the DIOPHANTINE EQUATION

$$x^n + y^n = z^n, \tag{1}$$

where x, y, z, and n are INTEGERS, has no NONZERO solutions for $n > 2$ has come to be known as Fermat's Last Theorem. It was called a "THEOREM" on the strength of Fermat's statement, despite the fact that no other mathematician was able to prove it for hundreds of years.

Note that the restriction $n > 2$ is obviously necessary since there are a number of elementary formulas for generating an infinite number of PYTHAGOREAN TRIPLES (x, y, z) satisfying the equation for $n = 2$,

$$x^2 + y^2 = z^2. \tag{2}$$

A first attempt to solve the equation can be made by attempting to factor the equation, giving

$$(z^{n/2} + y^{n/2})(z^{n/2} - y^{n/2}) = x^n. \tag{3}$$

Since the product is an exact POWER,

$$\begin{cases} z^{n/2} + y^{n/2} = 2^{n-1}p^n \\ z^{n/2} - y^{n/2} = 2q^n \end{cases} \text{or} \begin{cases} z^{n/2} + y^{n/2} = 2p^n \\ z^{n/2} - y^{n/2} = 2^{n-1}q^n. \end{cases} \tag{4}$$

Solving for y and z gives

$$\begin{cases} z^{n/2} = 2^{n-2}p^n + q^n \\ y^{n/2} = 2^{n-2}p^n - q^n \end{cases} \text{or} \begin{cases} z^{n/2} = p^n + 2^{n-2}q^n \\ y^{n/2} = p^n - 2^{n-2}q^n, \end{cases} \tag{5}$$

which give

$$\begin{cases} z = (2^{n-2}p^n + q^n)^{2/n} \\ y = (2^{n-2}p^n - q^n)^{2/n} \end{cases} \text{or} \begin{cases} z = (p^n + 2^{n-2}q^n)^{2/n} \\ y = (p^n - 2^{n-2}q^n)^{2/n}. \end{cases} \tag{6}$$

However, since solutions to these equations in RATIONAL NUMBERS are no easier to find than solutions to the original equation, this approach unfortunately does not provide any additional insight.

It is sufficient to prove Fermat's Last Theorem by considering PRIME POWERS only, since the arguments can otherwise be written

$$(x^m)^p + (y^m)^p = (z^m)^p, \tag{7}$$

so redefining the arguments gives

$$x^p + y^p = z^p. \tag{8}$$

The so-called "first case" of the theorem is for exponents which are RELATIVELY PRIME to x, y, and z ($p \nmid x, y, z$) and was considered by Wieferich. Sophie Germain proved the first case of Fermat's Last Theorem for any ODD PRIME p when $2p + 1$ is also a PRIME. Legendre subsequently proved that if p is a PRIME such that $4p + 1$, $8p + 1$, $10p + 1$, $14p + 1$, or $16p + 1$ is also a PRIME, then the first case of Fermat's Last Theorem holds for p. This established Fermat's Last Theorem for $p < 100$. In 1849, Kummer proved it for all REGULAR PRIMES and COMPOSITE NUMBERS of which they are factors (Vandiver 1929, Ball and Coxeter 1987).

Kummer's attack led to the theory of IDEALS, and Vandiver developed VANDIVER'S CRITERIA for deciding if a given IRREGULAR PRIME satisfies the theorem. Genocchi (1852) proved that the first case is true for p if $(p, p - 3)$ is not an IRREGULAR PAIR. In 1858, Kummer showed that the first case is true if either $(p, p - 3)$ or $(p, p - 5)$ is an IRREGULAR PAIR, which was subsequently extended to include $(p, p - 7)$ and $(p, p - 9)$ by Mirimanoff (1905). Wieferich (1909) proved that if the equation is solved in integers RELATIVELY PRIME to an ODD PRIME p, then

$$2^{p-1} \equiv 1 \pmod{p^2}. \tag{9}$$

(Ball and Coxeter 1987). Such numbers are called WIEFERICH PRIMES. Mirimanoff (1909) subsequently showed that

$$3^{p-1} \equiv 1 \pmod{p^2} \tag{10}$$

must also hold for solutions RELATIVELY PRIME to an ODD PRIME p, which excludes the first two WIEFERICH PRIMES 1093 and 3511. Vandiver (1914) showed

$$5^{p-1} \equiv 1 \pmod{p^2}, \tag{11}$$

and Frobenius extended this to

$$11^{p-1}, 17^{p-1} \equiv 1 \pmod{p^2}. \tag{12}$$

It has also been shown that if p were a PRIME of the form $6x - 1$, then

$$7^{p-1}, 13^{p-1}, 19^{p-1} \equiv 1 \pmod{p^2}, \tag{13}$$

which raised the smallest possible p in the "first case" to 253,747,889 by 1941 (Rosser 1941). Granville and Monagan (1988) showed if there exists a PRIME p satisfying Fermat's Last Theorem, then

$$q^{p-1} \equiv 1 \pmod{p^2} \tag{14}$$

for $q = 5, 7, 11, \ldots, 71$. This establishes that the first case is true for all PRIME exponents up to 714,591,416,091,398 (Vardi 1991).

The "second case" of Fermat's Last Theorem (for $p | x, y, z$) proved harder than the first case.

Euler proved the general case of the theorem for $n = 3$, Fermat $n = 4$, Dirichlet and Lagrange $n = 5$. In 1832, Dirichlet established the case $n = 14$. The $n = 7$ case was proved by Lamé (1839), using the identity

$$\begin{aligned} &(X + Y + Z)^7 - (X^7 + Y^7 + Z^7) \\ &\quad = 7(X + Y)(X + Z)(Y + Z) \\ &\qquad \times [(X^2 + Y^2 + Z^2 + XY + XZ + YZ)^2 \\ &\qquad\qquad + XYZ(X + Y + Z)]. \end{aligned} \tag{15}$$

Although some errors were present in this proof, these were subsequently fixed by Lebesgue (1840). Much additional progress was made over the next 150 years, but no completely general result had been obtained. Buoyed by false confidence after his proof that PI is TRANSCENDENTAL, the mathematician Lindemann proceeded to publish several proofs of Fermat's Last Theorem, all of them invalid (Bell 1937, pp. 464–465). A prize of 100,000 German marks (known as the Wolfskel Prize) was also offered for the first valid proof (Ball and Coxeter 1987, p. 72).

A recent false alarm for a general proof was raised by Y. Miyaoka (Cipra 1988) whose proof, however, turned out to be flawed. Other attempted proofs among both professional and amateur mathematicians are discussed by vos Savant (1993), although vos Savant erroneously claims that work on the problem by Wiles (discussed below) is invalid. By the time 1993 rolled around, the general case of Fermat's Last Theorem had been shown to be true for all exponents up to 4×10^6 (Cipra 1993). However, given that a proof of Fermat's Last Theorem requires truth for *all* exponents, proof for any finite number of exponents does not constitute any significant progress towards a proof of the general theorem (although the fact that no counterexamples were found for this many cases is highly suggestive).

In 1993, a bombshell was dropped. In that year, the general theorem was partially proven by Andrew Wiles (Cipra 1993, Stewart 1993) by proving the SEMISTABLE case of the TANIYAMA-SHIMURA CONJECTURE. Unfortunately, several holes were discovered in the proof shortly thereafter when Wiles' approach via the TANIYAMA-SHIMURA CONJECTURE became hung up on properties of the SELMER GROUP using a tool called an "Euler system." However, the difficulty was circumvented by Wiles and R. Taylor in late 1994 (Cipra 1994, 1995ab) and published in Taylor and Wiles (1995) and Wiles (1995). Wiles' proof succeeds by (1) replacing ELLIPTIC CURVES with Galois representations, (2) reducing the problem to a CLASS NUMBER FORMULA, (3) proving that FORMULA, and (4) tying up loose ends that arise because the formalisms fail in the simplest degenerate cases (Cipra 1995a).

The proof of Fermat's Last Theorem marks the end of a mathematical era. Since virtually all of the tools which were eventually brought to bear on the problem had yet to be invented in the time of Fermat, it is interesting to speculate about whether he actually was in possession of an elementary proof of the theorem. Judging by the temerity with which the problem resisted attack for so long, Fermat's alleged proof seems likely to have been illusionary.

see also ABC CONJECTURE, BOGOMOLOV-MIYAOKA-YAU INEQUALITY, MORDELL CONJECTURE, PYTHAGOREAN TRIPLE, RIBET'S THEOREM, SELMER GROUP, SOPHIE GERMAIN PRIME, SZPIRO'S CONJECTURE, TANIYAMA-SHIMURA CONJECTURE, VOJTA'S CONJECTURE, WARING FORMULA

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 69–73, 1987.
Beiler, A. H. "The Stone Wall." Ch. 24 in *Recreations in the Theory of Numbers: The Queen of Mathematics Entertains.* New York: Dover, 1966.
Bell, E. T. *Men of Mathematics.* New York: Simon and Schuster, 1937.
Bell, E. T. *The Last Problem.* New York: Simon and Schuster, 1961.
Cipra, B. A. "Fermat Theorem Proved." *Science* **239**, 1373, 1988.
Cipra, B. A. "Mathematics—Fermat's Last Theorem Finally Yields." *Science* **261**, 32–33, 1993.
Cipra, B. A. "Is the Fix in on Fermat's Last Theorem?" *Science* **266**, 725, 1994.
Cipra, B. A. "Fermat's Theorem—At Last." *What's Happening in the Mathematical Sciences, 1995–1996, Vol. 3.* Providence, RI: Amer. Math. Soc., pp. 2–14, 1996.
Cipra, B. A. "Princeton Mathematician Looks Back on Fermat Proof." *Science* **268**, 1133–1134, 1995b.
Courant, R. and Robbins, H. "Pythagorean Numbers and Fermat's Last Theorem." §2.3 in Supplement to Ch. 1 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 40–42, 1996.
Cox, D. A. "Introduction to Fermat's Last Theorem." *Amer. Math. Monthly* **101**, 3–14, 1994.
Dickson, L. E. "Fermat's Last Theorem, $ax^r + by^s = cz^t$, and the Congruence $x^n + y^n \equiv z^n \pmod{p}$." Ch. 26 in *History of the Theory of Numbers, Vol. 2: Diophantine Analysis.* New York: Chelsea, pp. 731–776, 1952.
Edwards, H. M. *Fermat's Last Theorem: A Genetic Introduction to Algebraic Number Theory.* New York: Springer-Verlag, 1977.
Edwards, H. M. "Fermat's Last Theorem." *Sci. Amer.*, Oct. 1978.
Granville, A. "Review of BBC's Horizon Program, 'Fermat's Last Theorem'." *Not. Amer. Math. Soc.* **44**, 26–28, 1997.
Granville, A. and Monagan, M. B. "The First Case of Fermat's Last Theorem is True for All Prime Exponents up to 714,591,416,091,389." *Trans. Amer. Math. Soc.* **306**, 329–359, 1988.
Guy, R. K. "The Fermat Problem." §D2 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 144–146, 1994.
Hanson, A. "Fermat Project." `http://www.cica.indiana.edu/projects/Fermat/`.
Kolata, G. "Andrew Wiles: A Math Whiz Battles 350-Year-Old Puzzle." *New York Times*, June 29, 1993.
Lynch, J. "Fermat's Last Theorem." BBC Horizon television documentary. `http://www.bbc.co.uk/horizon/fermat.shtml`.
Lynch, J. (Producer and Writer). "The Proof." NOVA television episode. 52 mins. Broadcast by the U. S. Public Broadcasting System on Oct. 28, 1997.
Mirimanoff, D. "Sur le dernier théorème de Fermat et le critérium de wiefer." *Enseiggnement Math.* **11**, 455–459, 1909.
Mordell, L. J. *Fermat's Last Theorem.* New York: Chelsea, 1956.
Murty, V. K. (Ed.). *Fermat's Last Theorem: Proceedings of the Fields Institute for Research in Mathematical Sciences on Fermat's Last Theorem, Held 1993–1994 Toronto, Ontario, Canada.* Providence, RI: Amer. Math. Soc., 1995.
Osserman, R. (Ed.). *Fermat's Last Theorem. The Theorem and Its Proof: An Exploration of Issues and Ideas.* 98 min. videotape and 56 pp. book. 1994.
Ribenboim, P. *Lectures on Fermat's Last Theorem.* New York: Springer-Verlag, 1979.

Ribet, K. A. and Hayes, B. "Fermat's Last Theorem and Modern Arithmetic." *Amer. Sci.* **82**, 144–156, March/April 1994.

Ribet, K. A. and Hayes, B. Correction to "Fermat's Last Theorem and Modern Arithmetic." *Amer. Sci.* **82**, 205, May/June 1994.

Rosser, B. "On the First Case of Fermat's Last Theorem." *Bull. Amer. Math. Soc.* **45**, 636–640, 1939.

Rosser, B. "A New Lower Bound for the Exponent in the First Case of Fermat's Last Theorem." *Bull. Amer. Math. Soc.* **46**, 299–304, 1940.

Rosser, B. "An Additional Criterion for the First Case of Fermat's Last Theorem." *Bull. Amer. Math. Soc.* **47**, 109–110, 1941.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 144–149, 1993.

Singh, S. *Fermat's Enigma: The Quest to Solve the World's Greatest Mathematical Problem.* New York: Walker & Co., 1997.

Stewart, I. "Fermat's Last Time-Trip." *Sci. Amer.* **269**, 112–115, 1993.

Taylor, R. and Wiles, A. "Ring-Theoretic Properties of Certain Hecke Algebras." *Ann. Math.* **141**, 553–572, 1995.

van der Poorten, A. *Notes on Fermat's Last Theorem.* New York: Wiley, 1996.

Vandiver, H. S. "On Fermat's Last Theorem." *Trans. Amer. Math. Soc.* **31**, 613–642, 1929.

Vandiver, H. S. *Fermat's Last Theorem and Related Topics in Number Theory.* Ann Arbor, MI: 1935.

Vandiver, H. S. "Fermat's Last Theorem: Its History and the Nature of the Known Results Concerning It." *Amer. Math. Monthly,* **53**, 555–578, 1946.

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 59–61, 1991.

vos Savant, M. *The World's Most Famous Math Problem.* New York: St. Martin's Press, 1993.

Wieferich, A. "Zum letzten Fermat'schen Theorem." *J. reine angew. Math.* **136**, 293–302, 1909.

Wiles, A. "Modular Elliptic-Curves and Fermat's Last Theorem." *Ann. Math.* **141**, 443–551, 1995.

Fermat's Lesser Theorem

see FERMAT'S LITTLE THEOREM

Fermat's Little Theorem

If p is a PRIME number and a a NATURAL NUMBER, then

$$a^p \equiv a \pmod{p}. \tag{1}$$

Furthermore, if $p \nmid a$ (p does not divide a), then there exists some smallest exponent d such that

$$a^d - 1 \equiv 0 \pmod{p} \tag{2}$$

and d divides $p-1$. Hence,

$$a^{p-1} - 1 \equiv 0 \pmod{p}. \tag{3}$$

This is a generalization of the CHINESE HYPOTHESIS and a special case of EULER'S THEOREM. It is sometimes called FERMAT'S PRIMALITY TEST and is a NECESSARY but not SUFFICIENT test for primality. Although it was presumably proved (but suppressed) by Fermat, the first proof was published by Euler in 1749.

The theorem is easily proved using mathematical INDUCTION. Suppose $p | a^p - a$. Then examine

$$(a+1)^p - (a+1). \tag{4}$$

From the BINOMIAL THEOREM,

$$(a+1)^p = a^p + \binom{p}{1} a^{p-1} + \binom{p}{2} a^{p-2} + \ldots + \binom{p}{p-1} a + 1. \tag{5}$$

Rewriting,

$$(a+1)^p - a^p - 1 = \binom{p}{1} a^{p-1} + \binom{p}{2} a^{p-2} + \ldots + \binom{p}{p-1} a. \tag{6}$$

But p divides the right side, so it also divides the left side. Combining with the induction hypothesis gives that p divides the sum

$$[(a+1)^p - a^p - 1] + (a^p - a) = (a+1)^p - (a+1), \tag{7}$$

as assumed, so the hypothesis is true for any a. The theorem is sometimes called FERMAT'S SIMPLE THEOREM. WILSON'S THEOREM follows as a COROLLARY of Fermat's Little Theorem.

Fermat's little theorem shows that, if p is PRIME, there does not exists a base $a < p$ with $(a, p) = 1$ such that $a^{p-1} - 1$ possesses a nonzero residue modulo p. If such base a exists, p is therefore guaranteed to be composite. However, the lack of a nonzero residue in Fermat's little theorem does *not* guarantee that p is PRIME. The property of unambiguously certifying composite numbers while passing some PRIMES make Fermat's little theorem a COMPOSITENESS TEST which is sometimes called the FERMAT COMPOSITENESS TEST. COMPOSITE NUMBERS known as FERMAT PSEUDOPRIMES (or sometimes simply "PSEUDOPRIMES") have zero residue for some as and so are not identified as composite. Worse still, there exist numbers known as CARMICHAEL NUMBERS (the smallest of which is 561) which give zero residue for *any* choice of the base a RELATIVELY PRIME to p. However, FERMAT'S LITTLE THEOREM CONVERSE provides a criterion for certifying the primality of a number.

A number satisfying Fermat's little theorem for some nontrivial base and which is not known to be composite is called a PROBABLE PRIME. A table of the smallest PSEUDOPRIMES P for the first 100 bases a follows (Sloane's A007535).

a	P	a	P	a	P	a	P	a	P
2	341	22	69	42	205	62	63	82	91
3	91	23	33	43	77	63	341	83	105
4	15	24	25	44	45	64	65	84	85
5	124	25	28	45	76	65	133	85	129
6	35	26	27	46	133	66	91	86	87
7	25	27	65	47	65	67	85	87	91
8	9	28	87	48	49	68	69	88	91
9	28	29	35	49	66	69	85	89	99
10	33	30	49	50	51	70	169	90	91
11	15	31	49	51	65	71	105	91	115
12	65	32	33	52	85	72	85	92	93
13	21	33	85	53	65	73	111	93	301
14	15	34	35	54	55	74	75	94	95
15	341	35	51	55	63	75	91	95	141
16	51	36	91	56	57	76	77	96	133
17	45	37	45	57	65	77	95	97	105
18	25	38	39	58	95	78	341	98	99
19	45	39	95	59	87	79	91	99	145
20	21	40	91	60	341	80	81	100	259
21	55	41	105	61	91	81	85		

see also BINOMIAL THEOREM, CARMICHAEL NUMBER, CHINESE HYPOTHESIS, COMPOSITE NUMBER, COMPOSITENESS TEST, EULER'S THEOREM, FERMAT'S LITTLE THEOREM CONVERSE, FERMAT PSEUDOPRIME, MODULO MULTIPLICATION GROUP, PRATT CERTIFICATE, PRIMALITY TEST, PRIME NUMBER, PSEUDOPRIME, RELATIVELY PRIME, TOTIENT FUNCTION, WIEFERICH PRIME, WILSON'S THEOREM, WITNESS

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 61, 1987.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 141–142, 1996.

Courant, R. and Robbins, H. "Fermat's Theorem." §2.2 in Supplement to Ch. 1 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 37–38, 1996.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 20, 1993.

Sloane, N. J. A. Sequence A007535/M5440 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Fermat's Little Theorem Converse

The converse of FERMAT'S LITTLE THEOREM is also known as LEHMER'S THEOREM. It states that, if an INTEGER x is PRIME to m and $x^{m-1} \equiv 1 \pmod{m}$ and there is no INTEGER $e < m-1$ for which $x^e \equiv 1 \pmod{m}$, then m is PRIME. Here, x is called a WITNESS to the primality of m. This theorem is the basis for the PRATT PRIMALITY CERTIFICATE.

see also FERMAT'S LITTLE THEOREM, PRATT CERTIFICATE, PRIMALITY CERTIFICATE, WITNESS

References

Riesel, H. *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, p. 96, 1994.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 278–279, 1991.

Fermat-Lucas Number

A number of the form $2^n + 1$ obtained by setting $x = 1$ in a FERMAT-LUCAS POLYNOMIAL. The first few are 3, 5, 9, 17, 33, ... (Sloane's A000051).

see also FERMAT NUMBER (LUCAS)

References

Sloane, N. J. A. Sequence A000051/M0717 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Fermat Number

A BINOMIAL NUMBER of the form $F_n = 2^{2^n} + 1$. The first few for $n = 0, 1, 2, \ldots$ are 3, 5, 17, 257, 65537, 4294967297, ... (Sloane's A000215). The number of DIGITS for a Fermat number is

$$\begin{aligned} D(n) &= \lfloor [\log(2^{2^n} + 1)] + 1 \rfloor \approx \lfloor \log(2^{2^n}) + 1 \rfloor \\ &= \lfloor 2^n \log 2 + 1 \rfloor. \end{aligned} \tag{1}$$

Being a Fermat number is the NECESSARY (but not SUFFICIENT) form a number

$$N_n \equiv 2^n + 1 \tag{2}$$

must have in order to be PRIME. This can be seen by noting that if $N_n = 2^n + 1$ is to be PRIME, then n cannot have any ODD factors b or else N_n would be a factorable number of the form

$$\begin{aligned} 2^n + 1 &= (2^a)^b + 1 \\ &= (2^a + 1)[2^{a(b-1)} - 2^{a(b-2)} + 2^{a(b-3)} - \ldots + 1]. \end{aligned} \tag{3}$$

Therefore, for a PRIME N_n, n must be a POWER of 2.

Fermat conjectured in 1650 that every Fermat number is PRIME, but only COMPOSITE Fermat numbers F_n are known for $n \geq 5$. Eisenstein (1844) proposed as a problem the proof that there are an infinite number of Fermat primes (Ribenboim 1996, p. 88), but this has not yet been achieved. An anonymous writer proposed that numbers of the form $2^2 + 1$, $2^{2^2} + 1$, $2^{2^{2^2}} + 1$ were PRIME. However, this conjecture was refuted when Selfridge (1953) showed that

$$F_{16} = 2^{2^{16}} + 1 = 2^{2^{2^{2^2}}} + 1 \tag{4}$$

is COMPOSITE (Ribenboim 1996, p. 88). Numbers of the form $a^{2^n} + b^{2^n}$ are called generalized Fermat numbers (Ribenboim 1996, pp. 359–360).

Fermat numbers satisfy the RECURRENCE RELATION

$$F_m = F_0 F_1 \cdots F_{m-1} + 2. \tag{5}$$

F_n can be shown to be PRIME iff it satisfies PÉPIN'S TEST

$$3^{(F_n - 1)/2} \equiv -1 \pmod{F_n}. \tag{6}$$

PÉPIN'S THEOREM

$$3^{2^{2^n-1}} \equiv -1 \pmod{F_n} \tag{7}$$

is also NECESSARY and SUFFICIENT.

In 1770, Euler showed that any FACTOR of F_n must have the form

$$2^{n+1}K+1, \tag{8}$$

where K is a POSITIVE INTEGER. In 1878, Lucas increased the exponent of 2 by one, showing that FACTORS of Fermat numbers must be of the form

$$2^{n+2}L+1. \tag{9}$$

If

$$F = p_1 p_2 \cdots p_r \tag{10}$$

is the factored part of $F_n = FC$ (where C is the cofactor to be tested for primality), compute

$$A \equiv 3^{F_n-1} \pmod{F_n} \tag{11}$$

$$B \equiv 3^{F-1} \pmod{F_n} \tag{12}$$

$$R \equiv A - B \pmod{C}. \tag{13}$$

Then if $R \equiv 0$, the cofactor is a PROBABLE PRIME to the base 3^F; otherwise C is COMPOSITE.

In order for a POLYGON to be circumscribed about a CIRCLE (i.e., a CONSTRUCTIBLE POLYGON), it must have a number of sides N given by

$$N = 2^k F_0 \cdots F_n, \tag{14}$$

where the F_n are *distinct* Fermat primes. This is equivalent to the statement that the trigonometric functions $\sin(k\pi/N)$, $\cos(k\pi/N)$, etc., can be computed in terms of finite numbers of additions, multiplications, and square root extractions iff N is of the above form. The only known Fermat PRIMES are

$$F_0 = 3$$
$$F_1 = 5$$
$$F_2 = 17$$
$$F_3 = 257$$
$$F_4 = 65537$$

and it seems unlikely that any more exist.

Factoring Fermat numbers is extremely difficult as a result of their large size. In fact, only F_5 to F_{11} have been complete factored, as summarized in the following table. Written out explicitly, the complete factorizations are

$$\begin{aligned}
F_5 &= 641 \cdot 6700417 \\
F_6 &= 274177 \cdot 67280421310721 \\
F_7 &= 59649589127497217 \cdot 5704689200685129054721 \\
F_8 &= 1238926361552897 \cdot 93461639715357977769163 \cdots \\
&\quad \cdots 558199606896584051237541638188580280321 \\
F_9 &= 2424833 \cdot 7455602825647884208337395736200 4 \cdots \\
&\quad \cdots 54918783366342657 \cdot P99 \\
F_{10} &= 45592577 \cdot 6487031809 \cdot 4659775785220018 5 \cdots \\
&\quad \cdots 4326456074307677819289 7 \cdot P252 \\
F_{11} &= 319489 \cdot 974849 \cdot 167988556341760475137 \\
&\quad \cdot 3560841906445833920513 \cdot P564.
\end{aligned}$$

Here, the final large PRIME is not explicitly given since it can be computed by dividing F_n by the other given factors.

F_n	Digits	Facts.	Digits	Reference
5	10	2	3, 7	Euler 1732
6	20	2	6, 14	Landry 1880
7	39	2	7, 22	Morrison and Brillhart 1975
8	78	2	16, 62	Brent and Pollard 1981
9	155	3	7, 49, 99	Manasse and Lenstra (In Cipra 1993)
10	309	4	8, 10, 40, 252	Brent 1995
11	617	5	6, 6, 21, 22, 564	Brent 1988

Tables of known factors of Fermat numbers are given by Keller (1983), Brillhart *et al.* (1988), Young and Buell (1988), Riesel (1994), and Pomerance (1996). Young and Buell (1988) discovered that F_{20} is COMPOSITE, and Crandall *et al.* (1995) that F_{22} is COMPOSITE. A current list of the known factors of Fermat numbers is maintained by Keller, and reproduced in the form of a *Mathematica*® notebook by Weisstein. In these tables, since all factors are of the form $k2^n+1$, the known factors are expressed in the concise form (k, n). The number of factors for Fermat numbers F_n for $n = 0, 1, 2, \ldots$ are 1, 1, 1, 1, 1, 2, 2, 2, 2, 3, 4, 5, ...

see also CULLEN NUMBER, PÉPIN'S TEST, PÉPIN'S THEOREM, POCKLINGTON'S THEOREM, POLYGON, PROTH'S THEOREM, SELFRIDGE-HURWITZ RESIDUE, WOODALL NUMBER

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 68–69 and 94–95, 1987.

Brent, R. P. "Factorization of the Eighth Fermat Number." *Amer. Math. Soc. Abstracts* **1**, 565, 1980.

Brent, R. P. "Factorisation of F10." `http://cslab.anu.edu.au/~rpb/F10.html`.

Brent, R. P "Factorization of the Tenth and Eleventh Fermat Numbers." Submitted to *Math. Comput.* `ftp://nimbus.anu.edu.au/pub/Brent/rpb161tr.dvi.Z`.

Brent, R. P. and Pollard, J. M. "Factorization of the Eighth Fermat Number." *Math. Comput.* **36**, 627–630, 1981.

Brillhart, J.; Lehmer, D. H.; Selfridge, J.; Wagstaff, S. S. Jr.; and Tuckerman, B. *Factorizations of* $b^n \pm 1$, $b = 2, 3, 5, 6, 7, 10, 11, 12$ *Up to High Powers, rev. ed.* Providence, RI: Amer. Math. Soc., pp. lxxxvii and 2–3 of Update 2.2, 1988.

Cipra, B. "Big Number Breakdown." *Science* **248**, 1608, 1990.

Conway, J. H. and Guy, R. K. "Fermat's Numbers." In *The Book of Numbers.* New York: Springer-Verlag, pp. 137–141, 1996.

Cormack, G. V. and Williams, H. C. "Some Very Large Primes of the Form $k \cdot 2^m + 1$." *Math. Comput.* **35**, 1419–1421, 1980.

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 25–26 and 119, 1996.

Crandall, R.; Doenias, J.; Norrie, C.; and Young, J. "The Twenty-Second Fermat Number is Composite." *Math. Comput.* **64**, 863–868, 1995.

Dickson, L. E. "Fermat Numbers $F_n = 2^{2^n} + 1$." Ch. 15 in *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, pp. 375–380, 1952.

Dixon, R. *Mathographics.* New York: Dover, p. 53, 1991.

Euler, L. "Observationes de theoremate quodam Fermatiano aliisque ad numeros primos spectantibus." *Acad. Sci. Petropol.* **6**, 103–107, ad annos 1732–33 (1738). In *Leonhardi Euleri Opera Omnia,* Ser. I, Vol. II. Leipzig: Teubner, pp. 1–5, 1915.

Gostin, G. B. "A Factor of F_{17}." *Math. Comput.* **35**, 975–976, 1980.

Gostin, G. B. "New Factors of Fermat Numbers." *Math. Comput.* **64**, 393–395, 1995.

Gostin, G. B. and McLaughlin, P. B. Jr. "Six New Factors of Fermat Numbers." *Math. Comput.* **38**, 645–649, 1982.

Guy, R. K. "Mersenne Primes. Repunits. Fermat Numbers. Primes of Shape $k \cdot 2^n + 2$." §A3 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 8–13, 1994.

Hallyburton, J. C. Jr. and Brillhart, J. "Two New Factors of Fermat Numbers." *Math. Comput.* **29**, 109–112, 1975.

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, pp. 14–15, 1979.

Keller, W. "Factor of Fermat Numbers and Large Primes of the Form $k \cdot 2^n + 1$." *Math. Comput.* **41**, 661–673, 1983.

Keller, W. "Factors of Fermat Numbers and Large Primes of the Form $k \cdot 2^n + 1$, II." In prep.

Keller, W. "Prime Factors $k \cdot 2^n + 1$ of Fermat Numbers F_m and Complete Factoring Status." `http://ballingerr.xray.ufl.edu/proths/fermat.html`.

Kraitchik, M. "Fermat Numbers." §3.6 in *Mathematical Recreations.* New York: W. W. Norton, pp. 73–75, 1942.

Landry, F. "Note sur la décomposition du nombre $2^{64} + 1$ (Extrait)." *C. R. Acad. Sci. Paris,* **91**, 138, 1880.

Lenstra, A. K.; Lenstra, H. W. Jr.; Manasse, M. S.; and Pollard, J. M. "The Factorization of the Ninth Fermat Number." *Math. Comput.* **61**, 319–349, 1993.

Morrison, M. A. and Brillhart, J. "A Method of Factoring and the Factorization of F_7." *Math. Comput.* **29**, 183–205, 1975.

Pomerance, C. "A Tale of Two Sieves." *Not. Amer. Math. Soc.* **43**, 1473–1485, 1996.

Ribenboim, P. "Fermat Numbers" and "Numbers $k \times 2^n \pm 1$." §2.6 and 5.7 in *The New Book of Prime Number Records.* New York: Springer-Verlag, pp. 83–90 and 355–360, 1996.

Riesel, H. *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Basel: Birkhäuser, pp. 384–388, 1994.

Robinson, R. M. "A Report on Primes of the Form $k \cdot 2^n + 1$ and on Factors of Fermat Numbers." *Proc. Amer. Math. Soc.* **9**, 673–681, 1958.

Selfridge, J. L. "Factors of Fermat Numbers." *Math. Comput.* **7**, 274–275, 1953.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 13 and 78–80, 1993.

Sloane, N. J. A. Sequence A000215/M2503 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Weisstein, E. W. "Fermat Numbers." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fermat.m`.

Wrathall, C. P. "New Factors of Fermat Numbers." *Math. Comput.* **18**, 324–325, 1964.

Young, J. and Buell, D. A. "The Twentieth Fermat Number is Composite." *Math. Comput.* **50**, 261–263, 1988.

Fermat Number (Lucas)

A number of the form $2^n - 1$ obtained by setting $x = 1$ in a FERMAT POLYNOMIAL is called a MERSENNE NUMBER.

see also FERMAT-LUCAS NUMBER, MERSENNE NUMBER

Fermat Point

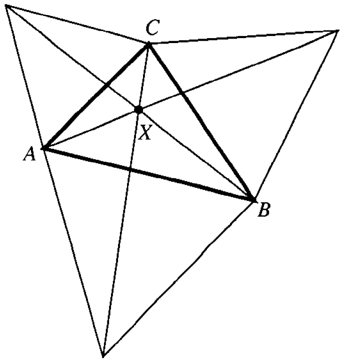

Also known as the first ISOGONIC CENTER and the TORRICELLI POINT. In a given ACUTE TRIANGLE ΔABC, the Fermat point is the point X which minimizes the sum of distances from A, B, and C,

$$|AX| + |BX| + |CX|. \tag{1}$$

This problem is called FERMAT'S PROBLEM or STEINER'S PROBLEM (Courant and Robbins 1941) and was proposed by Fermat to Torricelli. Torricelli's solution was published by his pupil Viviani in 1659 (Johnson 1929).

If all ANGLES of the TRIANGLE are less than 120° ($2\pi/3$), then the Fermat point is the interior point X from which each side subtends an ANGLE of 120°, i.e.,

$$\angle BXC = \angle CXA = \angle AXB = 120^\circ. \tag{2}$$

The Fermat point can also be constructed by drawing EQUILATERAL TRIANGLES on the outside of the given TRIANGLE and connecting opposite VERTICES. The three diagonals in the figure then intersect in the Fermat point. The TRIANGLE CENTER FUNCTION of the Fermat point is

$$\begin{aligned}\alpha &= \csc(A + \tfrac{1}{3}\pi) && (3)\\ &= bc[c^2a^2 + (c^2 + a^2 - b^2)^2][a^2b^2 - (a^2 + b^2 - c^2)^2]\\ &\quad\times [4\Delta - \sqrt{3}(b^2 + c^2 - a^2)]. && (4)\end{aligned}$$

The ANTIPEDAL TRIANGLE is EQUILATERAL and has AREA

$$\Delta' = 2\Delta\left[1 + \cot\omega\cot\left(\frac{\pi}{3}\right)\right], \qquad (5)$$

where ω is the BROCARD ANGLE.

Given three POSITIVE REAL NUMBERS l, m, n, the "generalized" Fermat point is the point P of a given ACUTE TRIANGLE ΔABC such that

$$l \cdot PA + m \cdot PB + n \cdot PC \qquad (6)$$

is a minimum (Greenberg and Robertello 1965, van de Lindt 1966, Tong and Chua 1995)

see also ISOGONIC CENTERS

References

Courant, R. and Robbins, H. *What is Mathematics?, 2nd ed.* Oxford, England: Oxford University Press, 1941.

Gallatly, W. *The Modern Geometry of the Triangle, 2nd ed.* London: Hodgson, p. 107, 1913.

Greenberg, I. and Robertello, R. A. "The Three Factory Problem." *Math. Mag.* **38**, 67–72, 1965.

Honsberger, R. *Mathematical Gems I.* Washington, DC: Math. Assoc. Amer., pp. 24–34, 1973.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 221–222, 1929.

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, p. 174, 1994.

Kimberling, C. "Fermat Point." `http://www.evansville.edu/~ck6/tcenters/class/fermat.html`.

Mowaffaq, H. "An Advanced Calculus Approach to Finding the Fermat Point." *Math. Mag.* **67**, 29–34, 1994.

Pottage, J. *Geometrical Investigations.* Reading, MA: Addison-Wesley, 1983.

Spain, P. G. "The Fermat Point of a Triangle." *Math. Mag.* **69**, 131–133, 1996.

Tong, J. and Chua, Y. S. "The Generalized Fermat's Point." *Math. Mag.* **68**, 214–215, 1995.

van de Lindt, W. J. "A Geometrical Solution of the Three Factory Problem." *Math. Mag.* **39**, 162–165, 1966.

Fermat's Polygonal Number Theorem

In 1638, Fermat proposed that every POSITIVE INTEGER is a sum of *at most* three TRIANGULAR NUMBERS, four SQUARE NUMBERS, five PENTAGONAL NUMBERS, and n n-POLYGONAL NUMBERS. Fermat claimed to have a proof of this result, although Fermat's proof has never been found. Gauss proved the triangular case, and noted the event in his diary on July 10, 1796, with the notation

$$**\,E\Upsilon RHKA \qquad num = \Delta + \Delta + \Delta.$$

This case is equivalent to the statement that every number of the form $8m + 3$ is a sum of three ODD SQUARES (Duke 1997). More specifically, a number is a sum of three SQUARES IFF it is not of the form $4^b(8m + 7)$ for $b \geq 0$, as first proved by Legendre in 1798.

Euler was unable to prove the square case of Fermat's theorem, but he left partial results which were subsequently used by Lagrange. The square case was finally proved by Jacobi and independently by Lagrange in 1772. It is therefore sometimes known as LAGRANGE'S FOUR-SQUARE THEOREM. In 1813, Cauchy proved the proposition in its entirety.

see also FIFTEEN THEOREM, VINOGRADOV'S THEOREM, LAGRANGE'S FOUR-SQUARE THEOREM, WARING'S PROBLEM

References

Cassels, J. W. S. *Rational Quadratic Forms.* New York: Academic Press, 1978.

Conway, J. H.; Guy, R. K.; Schneeberger, W. A.; and Sloane, N. J. A. "The Primary Pretenders." *Acta Arith.* **78**, 307–313, 1997.

Duke, W. "Some Old Problems and New Results about Quadratic Forms." *Not. Amer. Math. Soc.* **44**, 190–196, 1997.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 143–144, 1993.

Smith, D. E. *A Source Book in Mathematics.* New York: Dover, p. 91, 1984.

Fermat Polynomial

The POLYNOMIALS obtained by setting $p(x) = 3x$ and $q(x) = -2$ in the LUCAS POLYNOMIAL SEQUENCES. The first few Fermat polynomials are

$$\begin{aligned}\mathcal{F}_1(x) &= 1\\ \mathcal{F}_2(x) &= 3x\\ \mathcal{F}_3(x) &= 9x^2 - 2\\ \mathcal{F}_4(x) &= 27x^3 - 12x\\ \mathcal{F}_5(x) &= 81x^4 - 54x^2 + 4,\end{aligned}$$

and the first few Fermat-Lucas polynomials are

$$\begin{aligned}f_1(x) &= 3x\\ f_2(x) &= 9x^2 - 4\\ f_3(x) &= 27x^3 - 18x\\ f_4(x) &= 81x^4 - 72x^2 + 8\\ f_5(x) &= 243x^5 - 270x^3 + 60x.\end{aligned}$$

Fermat and Fermat-Lucas POLYNOMIALS satisfy

$$\mathcal{F}_n(1) = \mathcal{F}_n$$

$$f_n(1) = f_n$$

where $\mathcal{F}_n$ are FERMAT NUMBERS and f_n are FERMAT-LUCAS NUMBERS.

Fermat's Primality Test

see FERMAT'S LITTLE THEOREM

Fermat Prime

A FERMAT NUMBER $F_n = 2^{2^n} + 1$ which is PRIME.

see also CONSTRUCTIBLE POLYGON, FERMAT NUMBER

Fermat's Problem

In a given ACUTE TRIANGLE ΔABC, locate a point whose distances from A, B, and C have the smallest possible sum. The solution is the point from which each side subtends an angle of 120°, known as the FERMAT POINT.

see also ACUTE TRIANGLE, FERMAT POINT

Fermat Pseudoprime

A Fermat pseudoprime to a base a, written psp(a), is a COMPOSITE NUMBER n such that $a^{n-1} \equiv 1 \pmod{n}$ (i.e., it satisfies FERMAT'S LITTLE THEOREM, sometimes with the requirement that n must be ODD; Pomerance *et al.* 1980). psp(2)s are called POULET NUMBERS or, less commonly, SARRUS NUMBERS or FERMATIANS (Shanks 1993). The first few EVEN psp(2)s (including the PRIME 2 as a pseudoprime) are 2, 161038, 215326, ... (Sloane's A006935).

If base 3 is used in addition to base 2 to weed out potential COMPOSITE NUMBERS, only 4709 COMPOSITE NUMBERS remain $< 25 \times 10^9$. Adding base 5 leaves 2552, and base 7 leaves only 1770 COMPOSITE NUMBERS.

see also FERMAT'S LITTLE THEOREM, POULET NUMBER, PSEUDOPRIME

References

Pomerance, C.; Selfridge, J. L.; and Wagstaff, S. S. "The Pseudoprimes to $25 \cdot 10^9$." *Math. Comput.* **35**, 1003–1026, 1980. Available electronically from `ftp://sable.ox.ac.uk/pub/math/primes/ps2.Z`.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 115, 1993.

Sloane, N. J. A. Sequence A006935/M2190 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Fermat Quotient

The Fermat quotient for a number a and a PRIME base p is defined as

$$q_p(a) \equiv \frac{a^{p-1}-1}{p}. \tag{1}$$

If $p \nmid ab$, then

$$q_p(ab) = q_p(a) + q_p(b) \tag{2}$$

$$q_p(p \pm 1) = \mp 1 \tag{3}$$

$$q_p(2) = \frac{1}{p}\left(1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots - \frac{1}{p-1}\right), \tag{4}$$

all (mod p). The quantity $q_p(2) = (2^{p-1} - 1)/p$ is known to be SQUARE for only two PRIMES: the so-called WIEFERICH PRIMES 1093 and 3511 (Lehmer 1981, Crandall 1986).

see also WIEFERICH PRIME

References

Crandall, R. *Projects in Scientific Computation.* New York: Springer-Verlag, 1986.

Lehmer, D. H. "On Fermat's Quotient, Base Two." *Math. Comput.* **36**, 289–290, 1981.

Fermat's Right Triangle Theorem

The AREA of a RATIONAL RIGHT TRIANGLE cannot be a SQUARE NUMBER. This statement is equivalent to "a CONGRUUM cannot be a SQUARE NUMBER."

see also CONGRUUM, RATIONAL TRIANGLE, RIGHT TRIANGLE, SQUARE NUMBER

Fermat's Sigma Problem

Solve

$$\sigma(x^3) = y^2$$

and

$$\sigma(x^2) = y^3,$$

where σ is the DIVISOR FUNCTION.

see also WALLIS'S PROBLEM

Fermat's Simple Theorem

see FERMAT'S LITTLE THEOREM

Fermat's Spiral

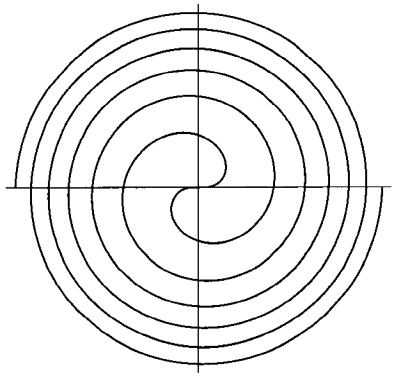

An ARCHIMEDEAN SPIRAL with $m = 2$ having polar equation

$$r = a\theta^{1/2},$$

discussed by Fermat in 1636 (MacTutor Archive). It is also known as the PARABOLIC SPIRAL. For any given POSITIVE value of θ, there are two corresponding values of r of opposite signs. The resulting spiral is therefore symmetrical about the line $y = -x$. The CURVATURE is

$$\kappa(\theta) = \frac{\frac{3a^2}{4\theta} + a^2\theta}{\left(\frac{a^2}{4\theta} + a^2\theta\right)^{3/2}}.$$

References

Dixon, R. *Mathographics.* New York: Dover, p. 121, 1991.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 69–70, 1993.
Lee, X. "Equiangular Spiral." `http://www.best.com/~xah/SpecialPlaneCurves_dir/EquiangularSpiral_dir/equiangularSpiral.html`.
Lockwood, E. H. *A Book of Curves.* Cambridge, England: Cambridge University Press, p. 175, 1967.
MacTutor History of Mathematics Archive. "Fermat's Spiral." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Fermats.html`.
Wells, D. *The Penguin Dictionary of Curious and Interesting Geometry.* Middlesex, England: Penguin Books, 1991.

Fermat Spiral Inverse Curve

The INVERSE CURVE of FERMAT'S SPIRAL with the origin taken as the INVERSION CENTER is the LITUUS.

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 186–187, 1972.

Fermat Sum Theorem

The only whole number solution to the DIOPHANTINE EQUATION

$$y^3 = x^2 + 2$$

is $y = 3$, $x = \pm 5$. This theorem was offered as a problem by Fermat, who suppressed his own proof.

Fermat's Theorem

A PRIME p can be represented in an essentially unique manner in the form $x^2 + y^2$ for integral x and y IFF $p \equiv 1 \pmod 4$ or $p = 2$. It can be restated by letting

$$Q(x, y) \equiv x^2 + y^2,$$

then all RELATIVELY PRIME solutions (x, y) to the problem of representing $Q(x, y) = m$ for m any INTEGER are achieved by means of successive applications of the GENUS THEOREM and COMPOSITION THEOREM. There is an analog of this theorem for EISENSTEIN INTEGERS.

see also EISENSTEIN INTEGER, SQUARE NUMBER

References

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 142–143, 1993.

Fermat's Two-Square Theorem

see FERMAT'S THEOREM

Fermatian

see POULET NUMBER

Fermi-Dirac Distribution

A distribution which arises in the study of half-integral spin particles in physics,

$$P(k) = \frac{k^s}{e^{k-\mu} + 1}.$$

Its integral is

$$\int_0^\infty \frac{k^s\,dk}{e^{k-\mu} + 1} = e^\mu \Gamma(s+1) \Phi(-e^\mu, s+1, 1),$$

where Φ is the LERCH TRANSCENDENT.

Fern

see BARNSLEY'S FERN

Ferrari's Identity

$$(a^2 + 2ac - 2bc - b^2)^4 + (b^2 - 2ab - 2ac - c^2)^4$$
$$+ (c^2 + 2ab + 2bc - a^2)^4 = 2(a^2 + b^2 + c^2 - ab + ac + bc)^4.$$

see also DIOPHANTINE EQUATION—QUARTIC

References

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 96–97, 1994.

Ferrers Diagram

see YOUNG DIAGRAM

Ferrers' Function

An alternative name for an associated LEGENDRE POLYNOMIAL.

see also LEGENDRE POLYNOMIAL

References

Sansone, G. *Orthogonal Functions, rev. English ed.* New York: Dover, p. 246, 1991.

Ferrier's Prime

According to Hardy and Wright (1979), the largest PRIME found before the days of electronic computers is the 44-digit number

$$F \equiv \tfrac{1}{17}(2^{148} + 1)$$
$$= 20988936657440586486151264256610222593863921,$$

which was found using only a mechanical calculator.

References

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, pp. 16–22, 1979.

Feuerbach Circle

see NINE-POINT CIRCLE

Feuerbach's Conic Theorem

The LOCUS of the centers of all CONICS through the VERTICES and ORTHOCENTER of a TRIANGLE (which are RECTANGULAR HYPERBOLAS when not degenerate), is a CIRCLE through the MIDPOINTS of the sides, the points half way from the ORTHOCENTER to the VERTICES, and the feet of the ALTITUDE.

see also ALTITUDE, CONIC SECTION, FEUERBACH'S THEOREM, KIEPERT'S HYPERBOLA, MIDPOINT, ORTHOCENTER, RECTANGULAR HYPERBOLA

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 198, 1959.

Feuerbach Point

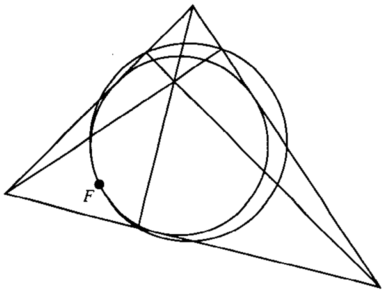

The point F at which the INCIRCLE and NINE-POINT CIRCLE are tangent. It has TRIANGLE CENTER FUNCTION

$$\alpha = 1 - \cos(B - C).$$

see also FEUERBACH'S THEOREM

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, p. 200, 1929.

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Salmon, G. *Conic Sections, 6th ed.* New York: Chelsea, p. 127, 1954.

Feuerbach's Theorem

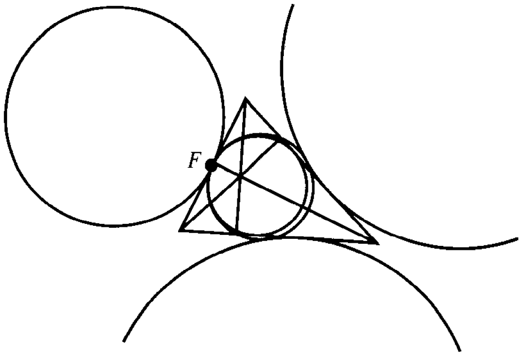

1. The CIRCLE which passes through the feet of the PERPENDICULARS dropped from the VERTICES of any TRIANGLE on the sides opposite them passes also through the MIDPOINTS of these sides as well as through the MIDPOINT of the segments which join the VERTICES to the point of intersection of the PERPENDICULAR (a NINE-POINT CIRCLE).
2. The NINE-POINT CIRCLE of any TRIANGLE is TANGENT internally to the INCIRCLE and TANGENT externally to the three EXCIRCLES.

see also EXCIRCLE, FEUERBACH POINT, INCIRCLE, MIDPOINT, NINE-POINT CIRCLE, PERPENDICULAR, TANGENT

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 117–119, 1967.

Dixon, R. *Mathographics.* New York: Dover, p. 59, 1991.

Feynman Point

The sequence of six 9s which begins at the 762th decimal place of PI,

$$\pi = 3.14159\ldots 134\underbrace{999999}837\ldots.$$

see also PI

FFT

see FAST FOURIER TRANSFORM

Fiber

A quantity F corresponding to a FIBER BUNDLE, where the FIBER BUNDLE is a MAP $f : E \to B$, with E the TOTAL SPACE of the FIBER BUNDLE and B the BASE SPACE of the FIBER BUNDLE.

see also FIBER BUNDLE, WHITNEY SUM

Fiber Bundle

A fiber bundle (also called simply a BUNDLE) with FIBER F is a MAP $f : E \to B$ where E is called the TOTAL SPACE of the fiber bundle and B the BASE SPACE of the fiber bundle. The main condition for the MAP to be a fiber bundle is that every point in the BASE SPACE $b \in B$ has a NEIGHBORHOOD U such that $f^{-1}(U)$ is HOMEOMORPHIC to $U \times F$ in a special way. Namely, if

$$h : f^{-1}(U) \to U \times F$$

is the HOMEOMORPHISM, then

$$\mathrm{proj}_U \circ h = f_{|f^{-1}(U)|},$$

where the MAP proj_U means projection onto the U component. The homeomorphisms h which "commute with projection" are called local TRIVIALIZATIONS for the fiber bundle f. In other words, E looks like the product $B \times F$ (at least locally), except that the fibers $f^{-1}(x)$ for $x \in B$ may be a bit "twisted."

Examples of fiber bundles include any product $B \times F \to B$ (which is a bundle over B with FIBER F), the MÖBIUS STRIP (which is a fiber bundle over the CIRCLE with FIBER given by the unit interval [0,1]; i.e, the BASE SPACE is the CIRCLE), and $\mathbb{S}^3$ (which is a bundle over $\mathbb{S}^2$ with fiber $\mathbb{S}^1$). A special class of fiber bundle is the VECTOR BUNDLE, in which the FIBER is a VECTOR SPACE.

see also BUNDLE, FIBER SPACE, FIBRATION

Fiber Space

A fiber space, depending on context, means either a FIBER BUNDLE or a FIBRATION.

see also FIBER BUNDLE, FIBRATION

Fibonacci Dual Theorem

Let F_n be the nth FIBONACCI NUMBER. Then the sequence $\{F_n\}_{n=2}^{\infty} = \{1, 2, 3, 5, 8, \ldots\}$ is COMPLETE, even if one is restricted to subsequences in which no two consecutive terms are both passed over (until the desired total is reached; Brown 1965, Honsberger 1985).

see also COMPLETE SEQUENCE, FIBONACCI NUMBER.

References

Brown, J. L. Jr. "A New Characterization of the Fibonacci Numbers." *Fib. Quart.* **3**, 1–8, 1965.

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., p. 130, 1985.

Fibonacci Hyperbolic Cosine

Let

$$\psi \equiv 1+\phi = \tfrac{1}{2}(3+\sqrt{5}\,) \approx 2.618034 \qquad (1)$$

where ϕ is the GOLDEN RATIO, and

$$\alpha = \ln\phi \approx 0.4812118. \qquad (2)$$

Then define

$$\begin{aligned} \mathrm{cFh}(x) &\equiv \frac{\psi^{x+1/2}+\psi^{-(x+1/2)}}{\sqrt{5}} && (3)\\ &= \frac{\phi^{(2x+1)}+\phi^{-(2x+1)}}{\sqrt{5}} && (4)\\ &= \frac{2}{\sqrt{5}}\cosh[(2x+1)\alpha]. && (5) \end{aligned}$$

This function satisfies

$$\mathrm{cFh}(-x) = \mathrm{cFh}(x-1). \qquad (6)$$

For $n \in \mathbb{Z}$, $\mathrm{cFh}(n) = F_{2n+1}$ where F_n is a FIBONACCI NUMBER.

References

Trzaska, Z. W. "On Fibonacci Hyperbolic Trigonometry and Modified Numerical Triangles." *Fib. Quart.* **34**, 129–138, 1996.

Fibonacci Hyperbolic Cotangent

$$\mathrm{ctFh}(x) \equiv \frac{\mathrm{cFh}(x)}{\mathrm{sFh}(x)},$$

where $\mathrm{cFh}(x)$ is the FIBONACCI HYPERBOLIC COSINE and $\mathrm{sFh}(x)$ is the FIBONACCI HYPERBOLIC SINE.

References

Trzaska, Z. W. "On Fibonacci Hyperbolic Trigonometry and Modified Numerical Triangles." *Fib. Quart.* **34**, 129–138, 1996.

Fibonacci Hyperbolic Sine

Let

$$\psi \equiv 1+\phi = \tfrac{1}{2}(3+\sqrt{5}\,) \approx 2.618034 \qquad (1)$$

where ϕ is the GOLDEN RATIO, and

$$\alpha = \ln\phi \approx 0.4812118. \qquad (2)$$

Then define

$$\begin{aligned} \mathrm{sFh}(x) &\equiv \frac{\psi^{x}-\psi^{-x}}{\sqrt{5}} && (3)\\ &= \frac{\phi^{2x}-\phi^{-2x}}{\sqrt{5}} && (4)\\ &= \frac{2}{\sqrt{5}}\sinh[2x\alpha]. && (5) \end{aligned}$$

For $n \in \mathbb{Z}$, $\mathrm{sFh}(n) = F_{2n}$ where F_n is a FIBONACCI NUMBER. The function satisfies

$$\mathrm{sFh}(-x) = -\,\mathrm{sFh}(x). \qquad (6)$$

References

Trzaska, Z. W. "On Fibonacci Hyperbolic Trigonometry and Modified Numerical Triangles." *Fib. Quart.* **34**, 129–138, 1996.

Fibonacci Hyperbolic Tangent

$$\mathrm{tFh}(x) \equiv \frac{\mathrm{sFh}(x)}{\mathrm{cFh}(x)},$$

where $\mathrm{sFh}(x)$ is the FIBONACCI HYPERBOLIC SINE and $\mathrm{cFh}(x)$ is the FIBONACCI HYPERBOLIC COSINE.

References

Trzaska, Z. W. "On Fibonacci Hyperbolic Trigonometry and Modified Numerical Triangles." *Fib. Quart.* **34**, 129–138, 1996.

Fibonacci Identity

Since

$$|(a+ib)(c+id)| = |a+ib|\,|c+di| \qquad (1)$$

$$|(ac-bd)+i(bc+ad)| = \sqrt{a^2+b^2}\sqrt{c^2+d^2}, \qquad (2)$$

it follows that

$$(a^2+b^2)(c^2+d^2) = (ac-bd)^2+(bc+ad)^2 \equiv e^2+f^2. \qquad (3)$$

This identity implies the 2-D CAUCHY-SCHWARZ SUM INEQUALITY.

see also CAUCHY-SCHWARZ SUM INEQUALITY, EULER FOUR-SQUARE IDENTITY

References

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, p. 9, 1996.

Fibonacci Matrix

A SQUARE MATRIX related to the FIBONACCI NUMBERS. The simplest is the FIBONACCI Q-MATRIX.

Fibonacci n-Step Number

An n-step Fibonacci sequence is given by defining $F_k = 0$ for $k \leq 0$, $F_1 = F_2 = 1$, $F_3 = 2$, and

$$F_k = \sum_{i=1}^{k} F_{n-i} \tag{1}$$

for $k > 3$. The case $n = 1$ corresponds to the degenerate 1, 1, 2, 2, 2, 2 ..., $n = 2$ to the usual FIBONACCI NUMBERS 1, 1, 2, 3, 5, 8, ... (Sloane's A000045), $n = 3$ to the TRIBONACCI NUMBERS 1, 1, 2, 4, 7, 13, 24, 44, 81, ... (Sloane's A000073), $n = 4$ to the TETRANACCI NUMBERS 1, 1, 2, 4, 8, 15, 29, 56, 108, ... (Sloane's A000078), etc.

The limit $\lim_{k\to\infty} F_k/F_{k-1}$ is given by solving

$$x^n(2-x) = 1 \tag{2}$$

for x and taking the REAL ROOT $x > 1$. If $n = 2$, the equation reduces to

$$x^2(2-x) = 1 \tag{3}$$

$$x^3 - 2x^2 + 1 = (x-1)(x^2 - x - 1) = 0, \tag{4}$$

giving solutions

$$x = 1, \tfrac{1}{2}(1 \pm \sqrt{5}). \tag{5}$$

The ratio is therefore

$$x = \tfrac{1}{2}(1 + \sqrt{5}) = \phi = 1.618\ldots, \tag{6}$$

which is the GOLDEN RATIO, as expected. Solutions for $n = 1$, 2, ... are given numerically by 1, 1.61803, 1.83929, 1.92756, 1.96595, ..., approaching 2 as $n \to \infty$.

see also FIBONACCI NUMBER, TRIBONACCI NUMBER

References

Sloane, N. J. A. Sequences A000045/M0692, A000073/M1074, and A000078/M1108 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Fibonacci Number

The sequence of numbers defined by the U_n in the LUCAS SEQUENCE. They are companions to the LUCAS NUMBERS and satisfy the same RECURRENCE RELATION,

$$F_n \equiv F_{n-2} + F_{n-1} \tag{1}$$

for $n = 3$, 4, ..., with $F_1 = F_2 = 1$. The first few Fibonacci numbers are 1, 1, 2, 3, 5, 8, 13, 21, ... (Sloane's A000045). The Fibonacci numbers give the number of pairs of rabbits n months after a single pair begins breeding (and newly born bunnies are assumed to begin breeding when they are two months old).

The ratios of alternate Fibonacci numbers are given by the convergents to ϕ^{-2}, where ϕ is the GOLDEN RATIO, and are said to measure the fraction of a turn between successive leaves on the stalk of a plant (PHYLLOTAXIS): 1/2 for elm and linden, 1/3 for beech and hazel, 2/5 for oak and apple, 3/8 for poplar and rose, 5/13 for willow and almond, etc. (Coxeter 1969, Ball and Coxeter 1987). The Fibonacci numbers are sometimes called PINE CONE NUMBERS (Pappas 1989, p. 224)

Another RECURRENCE RELATION for the Fibonacci numbers is

$$F_{n+1} = \left\lfloor \frac{F_n(1+\sqrt{5})+1}{2} \right\rfloor = \left\lfloor \phi F_n + \tfrac{1}{2} \right\rfloor, \tag{2}$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION and ϕ is the GOLDEN RATIO. This expression follows from the more general RECURRENCE RELATION that

$$\begin{vmatrix} F_n & F_{n+1} & \cdots & F_{n+k} \\ F_{n+k+1} & F_{n+k+2} & \cdots & F_{n+2k} \\ \vdots & \vdots & \ddots & \vdots \\ F_{n+k(k-1)+1} & F_{n+k(k-1)+2} & \cdots & F_{n+k^2} \end{vmatrix} = 0. \tag{3}$$

The GENERATING FUNCTION for the Fibonacci numbers is

$$g(x) = \sum_{n=0}^{\infty} F_n x^n = \frac{1}{1-x-x^2}. \tag{4}$$

Yuri Matijasevič (1970) proved that the equation $n = F_{2m}$ is a DIOPHANTINE EQUATION. This led to the proof of the impossibility of the tenth of HILBERT'S PROBLEMS (does there exist a general method for solving DIOPHANTINE EQUATIONS?) by Julia Robinson and Martin Davis in 1970.

The Fibonacci number F_{n+1} gives the number of ways for 2×1 DOMINOES to cover a $2 \times n$ CHECKERBOARD, as illustrated in the following diagrams (Dickau).

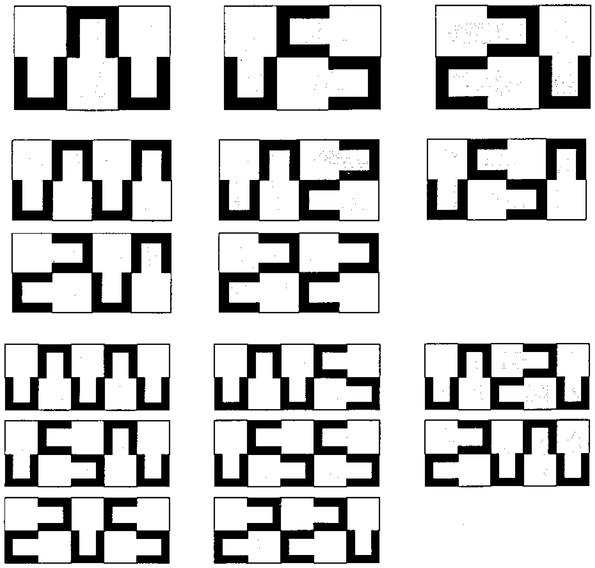

The number of ways of picking a SET (including the EMPTY SET) from the numbers 1, 2, ..., n without picking two consecutive numbers is F_{n+2}. The number of ways of picking a set (including the EMPTY SET) from the numbers 1, 2, ..., n without picking two consecutive numbers (where 1 and n are now consecutive) is $L_n = F_{n+1} + F_{n-1}$, where L_n is a LUCAS NUMBER. The probability of not getting two heads in a row in n tosses of a COIN is $F_{n+2}/2^n$ (Honsberger 1985, pp. 120–122). Fibonacci numbers are also related to the number of ways in which n COIN TOSSES can be made such that there are not three consecutive heads or tails. The number of ideals of an n-element FENCE POSET is the Fibonacci number F_n.

Sum identities are

$$\sum_{k=1}^{n} F_k = F_{n+2} - 1. \tag{5}$$

$$F_1 + F_3 + F_5 + \ldots + F_{2k+1} = F_{2k+2} \tag{6}$$

$$1 + F_2 + F_4 + F_6 + \ldots + F_{2k} = F_{2k+1} \tag{7}$$

$$\sum_{k=1}^{n} F_k^2 = F_n F_{n+1} \tag{8}$$

$$F_{2n} = F_{n+1}^2 - F_{n-1}^2 \tag{9}$$

$$F_{3n} = F_{n+1}^3 + F_n^3 + F_{n-1}^3. \tag{10}$$

Additional RECURRENCE RELATIONS are CASSINI'S IDENTITY

$$F_{n-1}F_{n+1} - F_n^2 = (-1)^n \tag{11}$$

and the relations

$$F_{2n+1} = 1 + F_2 + F_4 + \ldots + F_{2n} \tag{12}$$

$$F_{n+1}^2 = 4F_nF_{n-1} + F_{n-2}^2 \tag{13}$$

(Brousseau 1972),

$$F_{n+m} = F_{n-1}F_m + F_nF_{m+1} \tag{14}$$

$$F_{(k+1)n} = F_{n-1}F_{kn} + F_nF_{kn+1} \tag{15}$$

(Honsberger 1985, p. 107),

$$F_n = F_lF_{n-l+1} + F_{l-1}F_{n-l}, \tag{16}$$

so if $l = n - l + 1$, then $2l = n + 1$ and $l = (n + 1)/2$

$$F_n = F_{(n+1)/2}^2 + F_{(n-1)/2}^2. \tag{17}$$

Letting $k \equiv (n-1)/2$,

$$F_{2k+1} = F_{k+1}^2 + F_k^2 \tag{18}$$

$$F_{n+2}^2 - F_{n+1}^2 = F_nF_{n+3} \tag{19}$$

$$F_n^2 = F_{n-1}^2 + 3F_{n-2}^2 + 2F_{n-2}F_{n-3}. \tag{20}$$

Sum FORMULAS for F_n include

$$F_n = \frac{1}{2^{n-1}}\left[\binom{n}{1} + 5\binom{n}{3} + 5^2\binom{n}{5} + \ldots\right] \tag{21}$$

$$F_{n+1} = \binom{n}{0} + \binom{n-1}{1} + \binom{n-2}{2} + \ldots. \tag{22}$$

Cesàro derived the FORMULAS

$$\sum_{k=0}^{n} \binom{n}{k} F_k = F_{2n} \tag{23}$$

$$\sum_{k=0}^{n} \binom{n}{k} 2^k F_k = F_{3n} \tag{24}$$

(Honsberger 1985, pp. 109–110). Additional identities can be found throughout the *Fibonacci Quarterly* journal. A list of 47 generalized identities are given by Halton (1965).

In terms of the LUCAS NUMBER L_n,

$$F_{2n} = F_nL_n \tag{25}$$

$$F_{2n}(L_{2n}^2 - 1) = F_{6n} \tag{26}$$

$$F_{m+p} + (-1)^{p+1}F_{m-p} = F_pL_m \tag{27}$$

$$\sum_{k=a+1}^{a+4n} F_k = F_{a+4n+2} - F_{a+2} = F_{2n}L_{a+2n+2} \tag{28}$$

(Honsberger 1985, pp. 111–113). A remarkable identity is

$$\exp(L_1x + \tfrac{1}{2}L_2x^2 + \tfrac{1}{3}L_3x^3 + \ldots) = F_1 + F_2x + F_3x^2 + \ldots \tag{29}$$

(Honsberger 1985, pp. 118-119). It is also true that

$$5F_n^2 = L_n^2 - 4(-1)^n \tag{30}$$

and

$$\frac{L_n^2 - (-1)^aL_{n+a}^2}{F_n^2 - (-1)^aF_{n+a}^2} = 5 \tag{31}$$

for a ODD, and

$$\frac{L_n^2 + L_{n+a}^2 - 8(-1)^n}{F_n^2 + F_{n+a}^2} = 5 \tag{32}$$

for a EVEN (Freitag 1996).

The equation (1) is a LINEAR RECURRENCE SEQUENCE

$$x_n = Ax_{x-1} + Bx_{n-2} \qquad n \geq 3, \tag{33}$$

so the closed form for F_n is given by

$$F_n = \frac{\alpha^n - \beta^n}{\alpha - \beta}, \tag{34}$$

where α and β are the roots of $x^2 = Ax + B$. Here, $A = B = 1$, so the equation becomes

$$x^2 - x - 1 = 0, \tag{35}$$

which has ROOTS

$$x = \tfrac{1}{2}(1 \pm \sqrt{5}\,). \tag{36}$$

The closed form is therefore given by

$$F_n = \frac{(1+\sqrt{5}\,)^n - (1-\sqrt{5}\,)^n}{2^n\sqrt{5}}. \tag{37}$$

This is known as BINET'S FORMULA. Another closed form is

$$F_n = \left[\frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n\right] = \left[\frac{\phi^n}{\sqrt{5}}\right], \tag{38}$$

where $[x]$ is the NINT function.

From (1), the RATIO of consecutive terms is

$$\begin{aligned} \frac{F_n}{F_{n-1}} &= 1 + \frac{F_{n-2}}{F_{n-1}} = 1 + \frac{1}{\frac{F_{n-1}}{F_{n-2}}} \\ &= 1 + \frac{1}{1 + \frac{1}{\frac{F_{n-3}}{F_{n-2}}}} = \left[1, 1, \ldots, \frac{F_2}{F_1}\right] \\ &= [\underbrace{1, 1, \ldots, 1}_{n-1}], \end{aligned} \tag{39}$$

which is just the first few terms of the CONTINUED FRACTION for the GOLDEN RATIO ϕ. Therefore,

$$\lim_{n\to\infty} \frac{F_n}{F_{n-1}} = \phi. \tag{40}$$

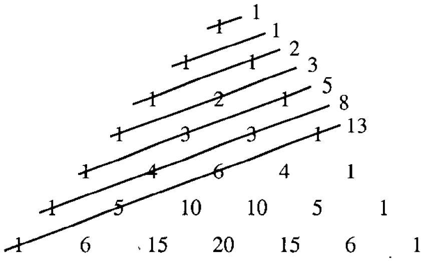

The "SHALLOW DIAGONALS" of PASCAL'S TRIANGLE sum to Fibonacci numbers (Pappas 1989),

$$\begin{aligned} \sum_{k=1}^{n} \binom{k}{n-k} &= \frac{(-1)^n {}_3F_2(1, 2, 1-n; \frac{1}{2}(3-n), 2-\frac{1}{2}n; -\frac{1}{4})}{\pi(2-3n+n^2)} \\ &= F_{n+1}, \end{aligned} \tag{41}$$

where ${}_3F_2(a, b, c; d, e; z)$ is a GENERALIZED HYPERGEOMETRIC FUNCTION.

The sequence of final digits in Fibonacci numbers repeats in cycles of 60. The last two digits repeat in 300, the last three in 1500, the last four in 15,000, etc.

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{F_n F_{n+2}} = 2 - \sqrt{5} \tag{42}$$

(Clark 1995). A very curious addition of the Fibonacci numbers is the following addition tree,

$$\begin{array}{l} 0 \\ \;\,1 \\ \;\;\,1 \\ \;\;\;\,2 \\ \;\;\;\;\,3 \\ \;\;\;\;\;\,5 \\ \;\;\;\;\;\;\,8 \\ \;\;\;\;\;\;\,13 \\ \;\;\;\;\;\;\;\,21 \\ \;\;\;\;\;\;\;\;\,34 \\ \;\;\;\;\;\;\;\;\;\,55 \\ \;\;\;\;\;\;\;\;\;\;\,89 \\ \;\;\;\;\;\;\;\;\;\;\;\;\;\ddots \\ \hline 0112359550561\ldots \end{array}$$

which is equal to the fractional digits of 1/89,

$$\sum_{n=0}^{\infty} \frac{F_n}{10^{n+1}} = \frac{1}{89}. \tag{43}$$

For $n \geq 3$, $F_n | F_m$ IFF $n | m$. $L_n | L_m$ IFF n divides into m an EVEN number of times. $(F_m, F_n) = F_{(m,n)}$ (Michael 1964; Honsberger 1985, pp. 131-132). No ODD Fibonacci number is divisible by 17 (Honsberger 1985, pp. 132 and 242). No Fibonacci number > 8 is ever of the form $p - 1$ or $p + 1$ where p is a PRIME number (Honsberger 1985, p. 133).

Consider the sum

$$s_k = \sum_{n=2}^{k} \frac{1}{F_{n-1}F_{n+1}} = \sum_{n=2}^{k} \left(\frac{1}{F_{n-1}F_n} - \frac{1}{F_n F_{n+1}}\right). \tag{44}$$

This is a TELESCOPING SUM, so

$$s_k = 1 - \frac{1}{F_{k+1}F_{k+2}}, \tag{45}$$

thus

$$S \equiv \lim_{k\to\infty} s_k = 1 \tag{46}$$

(Honsberger 1985, pp. 134–135). Using BINET'S FORMULA, it also follows that

$$\frac{F_{n+r}}{F_n} = \frac{\alpha^{n+r} - \beta^{n+r}}{\alpha^n - \beta^n} = \frac{\alpha^{n+r}}{\alpha^n} \frac{1 - \left(\frac{\beta}{\alpha}\right)^{n+r}}{1 - \left(\frac{\beta}{\alpha}\right)^n}, \tag{47}$$

where

$$\alpha = \tfrac{1}{2}(1 + \sqrt{5}\,) \tag{48}$$

$$\beta = \tfrac{1}{2}(1 - \sqrt{5}\,) \tag{49}$$

so

$$\frac{F_{n+r}}{F_n} = \alpha^r. \tag{50}$$

$$S' = \sum_{n=1}^{\infty} \frac{F_n}{F_{n+1}F_{n+2}} = 1 \tag{51}$$

(Honsberger 1985, pp. 138 and 242–243). The Millin Series has sum

$$S'' \equiv \sum_{n=0}^{\infty} \frac{1}{F_{2^n}} = \tfrac{1}{2}(7 - \sqrt{5}\,) \tag{52}$$

(Honsberger 1985, pp. 135–137).

The Fibonacci numbers are Complete. In fact, dropping one number still leaves a Complete Sequence, although dropping two numbers does not (Honsberger 1985, pp. 123 and 126). Dropping two terms from the Fibonacci numbers produces a sequence which is not even Weakly Complete (Honsberger 1985, p. 128). However, the sequence

$$F_n' \equiv F_n - (-1)^n \tag{53}$$

is Weakly Complete, even with any finite subsequence deleted (Graham 1964). $\{{F_n}^2\}$ is not Complete, but $\{{F_n}^2\} + \{{F_n}^2\}$ are. 2^{N-1} copies of $\{{F_n}^N\}$ are Complete.

For a discussion of Square Fibonacci numbers, see Cohn (1964), who proved that the only Square Number Fibonacci numbers are 1 and $F_{12} = 144$ (Cohn 1964, Guy 1994). Ming (1989) proved that the only Triangular Fibonacci numbers are 1, 3, 21, and 55. The Fibonacci and Lucas Numbers have no common terms except 1 and 3. The only Cubic Fibonacci numbers are 1 and 8.

$$(F_nF_{n+3}, 2F_{n+1}F_{n+2}, F_{2n+3} = {F_{n+1}}^2 + {F_{n+2}}^2) \tag{54}$$

is a Pythagorean Triple.

$$F_{4n}{}^2 + 8F_{2n}(F_{2n} + F_{6n}) = (3F_{4n})^2 \tag{55}$$

is always a Square Number (Honsberger 1985, p. 243).

In 1975, James P. Jones showed that the Fibonacci numbers are the Positive Integer values of the Polynomial

$$P(x,y) = -y^5 + 2y^4x + y^3x^2 - 2y^2x^3 - y(x^4 - 2) \tag{56}$$

for Gaussian Integers x and y (Le Lionnais 1983). If n and k are two Positive Integers, then between n^k and n^{k+1}, there can never occur more than n Fibonacci numbers (Honsberger 1985, pp. 104–105).

Every F_n that is Prime has a Prime n, but the converse is not necessarily true. The first few Prime Fibonacci numbers are for n = 3, 4, 5, 7, 11, 13, 17, 23, 29, 43, 47, 83, 131, 137, 359, 431, 433, 449, 509, 569, 571, ... (Sloane's A001605; Dubner and Keller 1998). Gardner's statement that F_{531} is prime is incorrect, especially since 531 is not even Prime (Gardner 1979, p. 161). It is not known if there are an Infinite number of Fibonacci primes.

The Fibonacci numbers F_n, are Squareful for n = 6, 12, 18, 24, 25, 30, 36, 42, 48, 50, 54, 56, 60, 66, ..., 300, 306, 312, 324, 325, 330, 336, ... (Sloane's A037917) and Squarefree for n = 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 13, ... (Sloane's A037918). The largest known Squareful Fibonacci number is F_{336}, and no Squareful Fibonacci numbers F_p are known with p Prime.

see also Cassini's Identity, Fast Fibonacci Transform, Fibonacci Dual Theorem, Fibonacci n-Step Number, Fibonacci Q-Matrix, Generalized Fibonacci Number, Inverse Tangent, Linear Recurrence Sequence, Lucas Sequence, Near Noble Number, Pell Sequence, Rabbit Constant, Stolarsky Array, Tetranacci Number, Tribonacci Number, Wythoff Array, Zeckendorf Representation, Zeckendorf's Theorem

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 56–57, 1987.

Basin, S. L. and Hoggatt, V. E. Jr. "A Primer on the Fibonacci Sequence." *Fib. Quart.* **1**, 1963.

Basin, S. L. and Hoggatt, V. E. Jr. "A Primer on the Fibonacci Sequence—Part II." *Fib. Quart.* **1**, 61–68, 1963.

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, pp. 94–101, 1987.

Brillhart, J.; Montgomery, P. L.; and Silverman, R. D. "Tables of Fibonacci and Lucas Factorizations." *Math. Comput.* **50**, 251–260 and S1–S15, 1988.

Brook, M. "Fibonacci Formulas." *Fib. Quart.* **1**, 60, 1963.

Brousseau, A. "Fibonacci Numbers and Geometry." *Fib. Quart.* **10**, 303–318, 1972.

Clark, D. Solution to Problem 10262. *Amer. Math. Monthly* **102**, 467, 1995.

Cohn, J. H. E. "On Square Fibonacci Numbers." *J. London Math. Soc.* **39**, 537–541, 1964.

Conway, J. H. and Guy, R. K. "Fibonacci Numbers." In *The Book of Numbers.* New York: Springer-Verlag, pp. 111–113, 1996.

Coxeter, H. S. M. "The Golden Section and Phyllotaxis." Ch. 11 in *Introduction to Geometry, 2nd ed.* New York: Wiley, 1969.

Dickau, R. M. "Fibonacci Numbers." `http://www.prairienet.org/~pops/fibboard.html`.

Dubner, H. and Keller, W. "New Fibonacci and Lucas Primes." *Math. Comput.* 1998.

Freitag, H. Solution to Problem B-772. "An Integral Ratio." *Fib. Quart.* **34**, 82, 1996.

Gardner, M. *Mathematical Circus: More Puzzles, Games, Paradoxes and Other Mathematical Entertainments from Scientific American.* New York: Knopf, 1979.

Graham, R. "A Property of Fibonacci Numbers." *Fib. Quart.* **2**, 1–10, 1964.

Guy, R. K. "Fibonacci Numbers of Various Shapes." §D26 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 194–195, 1994.

Halton, J. H. "On a General Fibonacci Identity." *Fib. Quart.* **3**, 31–43, 1965.

Hoggatt, V. E. Jr. *The Fibonacci and Lucas Numbers.* Boston, MA: Houghton Mifflin, 1969.

Hoggatt, V. E. Jr. and Ruggles, I. D. "A Primer on the Fibonacci Sequence—Part III." *Fib. Quart.* **1**, 61–65, 1963.

Hoggatt, V. E. Jr. and Ruggles, I. D. "A Primer on the Fibonacci Sequence—Part IV." *Fib. Quart.* **1**, 65–71, 1963.

Hoggatt, V. E. Jr. and Ruggles, I. D. "A Primer on the Fibonacci Sequence—Part V." *Fib. Quart.* **2**, 59–66, 1964.

Hoggatt, V. E. Jr.; Cox, N.; and Bicknell, M. "A Primer for the Fibonacci Numbers: Part XII." *Fib. Quart.* **11**, 317-331, 1973.

Honsberger, R. "A Second Look at the Fibonacci and Lucas Numbers." Ch. 8 in *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., 1985.

Knott, R. "Fibonacci Numbers and the Golden Section." `http://www.mcs.surrey.ac.uk/Personal/R.Knott/Fibonacci/fib.html`.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 146, 1983.

Leyland, P. `ftp://sable.ox.ac.uk/pub/math/factors/fibonacci.Z`.

Matijasevič, Yu. V. "Solution to of the Tenth Problem of Hilbert." *Mat. Lapok* **21**, 83–87, 1970.

Matijasevich, Yu. V. *Hilbert's Tenth Problem.* Cambridge, MA: MIT Press, 1993.

Michael, G. "A New Proof for an Old Property." *Fib. Quart.* **2**, 57–58, 1964.

Ming, L. "On Triangular Fibonacci Numbers." *Fib. Quart.* **27**, 98–108, 1989.

Ogilvy, C. S. and Anderson, J. T. "Fibonacci Numbers." Ch. 11 in *Excursions in Number Theory.* New York: Dover, pp. 133–144, 1988.

Pappas, T. "Fibonacci Sequence," "Pascal's Triangle, the Fibonacci Sequence & Binomial Formula," "The Fibonacci Trick," and "The Fibonacci Sequence & Nature." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 28–29, 40–41, 51, 106, and 222–225, 1989.

Schroeder, M. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise.* New York: W. H. Freeman, pp. 49–57, 1991.

Sloane, N. J. A. Sequences A037917, A037918, A000045/M0692, and A001605/M2309 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Vorob'ev, N. N. *Fibonacci Numbers.* New York: Blaisdell Publishing Co., 1961.

Fibonacci Polynomial

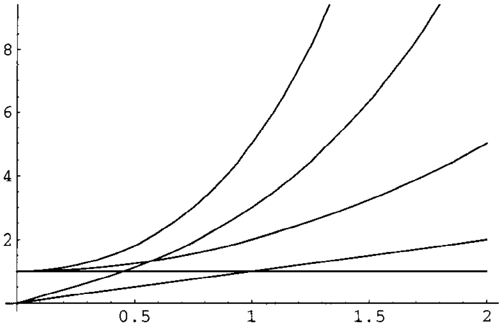

The W POLYNOMIALS obtained by setting $p(x) = x$ and $q(x) = 1$ in the LUCAS POLYNOMIAL SEQUENCE. (The corresponding w POLYNOMIALS are called LUCAS POLYNOMIALS.) The Fibonacci polynomials are defined by the RECURRENCE RELATION

$$F_{n+1}(x) = xF_n(x) + F_{n-1}(x), \tag{1}$$

with $F_1(x) = 1$ and $F_2(x) = x$. They are also given by the explicit sum formula

$$F_n(x) = \sum_{j=0}^{\lfloor (n-1)/2 \rfloor} \binom{n-j-1}{j} x^{n-2j-1}, \tag{2}$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION and $\binom{n}{m}$ is a BINOMIAL COEFFICIENT. The first few Fibonacci polynomials are

$$\begin{aligned} F_1(x) &= 1 \\ F_2(x) &= x \\ F_3(x) &= x^2 + 1 \\ F_4(x) &= x^3 + 2x \\ F_5(x) &= x^4 + 3x^2 + 1. \end{aligned}$$

The Fibonacci polynomials are normalized so that

$$F_n(1) = F_n, \tag{3}$$

where the F_ns are FIBONACCI NUMBERS.

The Fibonacci polynomials are related to the MORGAN-VOYCE POLYNOMIALS by

$$F_{2n+1}(x) = b_n(x^2) \tag{4}$$

$$F_{2n+n2}(x) = xB_n(x^2) \tag{5}$$

(Swamy 1968).

see also BRAHMAGUPTA POLYNOMIAL, FIBONACCI NUMBER, MORGAN-VOYCE POLYNOMIAL

References

Swamy, M. N. S. "Further Properties of Morgan-Voyce Polynomials." *Fib. Quart.* **6**, 167–175, 1968.

Fibonacci Pseudoprime

Consider a LUCAS SEQUENCE with $P > 0$ and $Q = \pm 1$. A Fibonacci pseudoprime is a COMPOSITE NUMBER n such that

$$V_n \equiv P \pmod{n}.$$

There exist no EVEN Fibonacci pseudoprimes with parameters $P = 1$ and $Q = -1$ (Di Porto 1993) or $P = Q = 1$ (André-Jeannin 1996). André-Jeannin (1996) also proved that if $(P, Q) \neq (1, -1)$ and $(P, Q) \neq (1, 1)$, then there exists at least one EVEN Fibonacci pseudoprime with parameters P and Q.

see also PSEUDOPRIME

References

André-Jeannin, R. "On the Existence of Even Fibonacci Pseudoprimes with Parameters P and Q." *Fib. Quart.* **34**, 75–78, 1996.

Di Porto, A. "Nonexistence of Even Fibonacci Pseudoprimes of the First Kind." *Fib. Quart.* **31**, 173–177, 1993.

Ribenboim, P. "Fibonacci Pseudoprimes." §2.X.A in *The New Book of Prime Number Records, 3rd ed.* New York: Springer-Verlag, pp. 127–129, 1996.

Fibonacci Q-Matrix

A Fibonacci Matrix of the form

$$\mathsf{M} = \begin{bmatrix} m & 1 \\ 1 & 0 \end{bmatrix}. \qquad (1)$$

If U and V are defined as Binet Forms

$$U_n = mU_{n-1} + U_{n-2} \quad (U_0 = 0, U_1 = 1) \qquad (2)$$

$$V_n = mV_{n-1} + V_{n-2} \quad (V_0 = 2, V_1 = m), \qquad (3)$$

then

$$\mathsf{M} = \begin{bmatrix} U_{n+1} & U_n \\ U_n & U_{n-1} \end{bmatrix} \qquad (4)$$

$$\mathsf{M}^{-1} = \mathsf{M} - m\mathsf{I} = \begin{bmatrix} 0 & 1 \\ 1 & -m \end{bmatrix}. \qquad (5)$$

Defining

$$\mathsf{Q} \equiv \begin{bmatrix} F_2 & F_1 \\ F_1 & F_0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \qquad (6)$$

then

$$\mathsf{Q}^n = \begin{bmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{bmatrix} \qquad (7)$$

(Honsberger 1985, pp. 106–107).

see also Binet Forms, Fibonacci Number

References

Honsberger, R. "A Second Look at the Fibonacci and Lucas Numbers." Ch. 8 in *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., 1985.

Fibonacci Sequence

see Fibonacci Number

Fibration

If $f : E \to B$ is a Fiber Bundle with B a Paracompact Topological Space, then f satisfies the Homotopy Lifting Property with respect to all Topological Spaces. In other words, if $g : [0,1] \times X \to B$ is a Homotopy from g_0 to g_1, and if g_0' is a Lift of the Map g_0 with respect to f, then g has a Lift to a Map g' with respect to f. Therefore, if you have a Homotopy of a Map into B, and if the beginning of it has a Lift, then that Lift can be extended to a Lift of the Homotopy itself.

A fibration is a Map between Topological Spaces $f : E \to B$ such that it satisfies the Homotopy Lifting Property.

see also Fiber Bundle, Fiber Space

Field

A field is any set of elements which satisfies the Field Axioms for both addition and multiplication and is a commutative Division Algebra. An archaic word for a field is Rational Domain. A field with a finite number of members is known as a Finite Field or Galois Field.

Because the identity condition must be different for addition and multiplication, every field must have at least two elements. Examples include the Complex Numbers ($\mathbb{C}$), Rational Numbers ($\mathbb{Q}$), and Real Numbers ($\mathbb{R}$), but *not* the Integers ($\mathbb{Z}$), which form a Ring. It has been proven by Hilbert and Weierstraß that all generalizations of the field concept to triplets of elements are equivalent to the field of Complex Numbers.

see also Adjunction, Algebraic Number Field, Coefficient Field, Cyclotomic Field, Field Axioms, Field Extension, Function Field, Galois Field, Mac Lane's Theorem, Module, Number Field, Quadratic Field, Ring, Skew Field, Vector Field

Field Axioms

The field axioms are generally written in additive and multiplicative pairs.

Name	Addition	Multiplication
Commutivity	$a + b = b + a$	$ab = ba$
Associativity	$(a + b) + c = a + (b + c)$	$(ab)c = a(bc)$
Distributivity	$a(b + c) = ab + ac$	$(a + b)c = ac + bc$
Identity	$a + 0 = a = 0 + a$	$a \cdot 1 = a = 1 \cdot a$
Inverses	$a + (-a) = 0 = (-a) + a$	$aa^{-1} = 1 = a^{-1}a$ if $a \neq 0$

see also Algebra, Field

Field Extension

A Field L is said to be a field extension of field K if K is a Subfield of L. This is denoted L/K (note that this Notation conflicts with that of a Quotient Group). The Complex Numbers are a field extension of the Real Numbers, and the Real Numbers are a field extension of the Rational Numbers.

see also Field

Fields Medal

The mathematical equivalent of the Nobel Prize (there is no Nobel Prize in mathematics) which is awarded by the International Mathematical Union every four years to one or more outstanding researchers, usually under 40 years of age. The first Fields Medal was awarded in 1936.

see also Burnside Problem, Mathematics Prizes, Poincaré Conjecture, Roth's Theorem, Tau Conjecture

References

MacTutor History of Mathematics Archives. "The Fields Medal." `http://www-groups.dcs.st-and.ac.uk/~history/Societies/FieldsMedal.html`.

Monastyrsky, M. *Modern Mathematics in the Light of the Fields Medals.* Wellesley, MA: A. K. Peters, 1997.

Fifteen Theorem

A theorem due to Conway *et al.* (1997) which states that, if a POSITIVE definite QUADRATIC FORM with integral matrix entries represents all natural numbers up to 15, then it represents all natural numbers. This theorem contains LAGRANGE'S FOUR-SQUARE THEOREM, since every number up to 15 is the sum of at most four SQUARES.

see also INTEGER-MATRIX FORM, LAGRANGE'S FOUR-SQUARE THEOREM, QUADRATIC FORM

References

Conway, J. H.; Guy, R. K.; Schneeberger, W. A.; and Sloane, N. J. A. "The Primary Pretenders." *Acta Arith.* **78**, 307–313, 1997.

Duke, W. "Some Old Problems and New Results about Quadratic Forms." *Not. Amer. Math. Soc.* **44**, 190–196, 1997.

Figurate Number

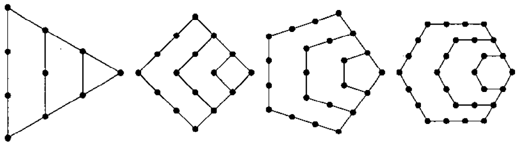

A number which can be represented by a regular geometrical arrangement of equally spaced points. If the arrangement forms a REGULAR POLYGON, the number is called a POLYGONAL NUMBER. The polygonal numbers illustrated above are called triangular, square, pentagonal, and hexagon numbers, respectively. Figurate numbers can also form other shapes such as centered polygons, L-shapes, 3-dimensional solids, etc. The following table lists the most common types of figurate numbers.

Name	Formula
biquadratic	n^4
centered cube	$(2n-1)(n^2-n+1)$
centered pentagonal	$\frac{1}{2}(5n^2-5n+2)$
centered square	$n^2+(n-1)^2$
centered triangular	$\frac{1}{2}(3n^2-3n+2)$
cubic	n^3
decagonal	$4n^2-3n$
gnomic	$2n-1$
heptagonal	$\frac{1}{2}n(5n-3)$
heptagonal pyramidal	$\frac{1}{6}n(n+1)(5n-2)$
hex	$3n^2-3n+1$
hexagonal	$n(2n-1)$
hexagonal pyramidal	$\frac{1}{6}n(n+1)(4n-1)$
octagonal	$n(3n-2)$
octahedral	$\frac{1}{3}n(2n^2+1)$
pentagonal	$\frac{1}{2}n(3n-1)$
pentagonal pyramidal	$\frac{1}{2}n^2(n+1)$
pentatope	$\frac{1}{24}n(n+1)(n+2)(n+3)$
pronic number	$n(n+1)$
rhombic dodecahedral	$(2n-1)(2n^2-2n+1)$
square	n^2
stella octangula	$n(2n^2-1)$
tetrahedral	$\frac{1}{6}n(n+1)(n+2)$
triangular	$\frac{1}{2}n(n+1)$
truncated octahedral	$16n^3-33n^2+24n-6$
truncated tetrahedral	$\frac{1}{6}n(23n^2-27n+10)$

An n-D FIGURATE NUMBER can be defined by

$$f_{m,s}^{r} = \frac{(rs+m-s)(r+m-2)}{m!(r-1)!}.$$

see also BIQUADRATIC NUMBER, CENTERED CUBE NUMBER, CENTERED PENTAGONAL NUMBER, CENTERED POLYGONAL NUMBER, CENTERED SQUARE NUMBER, CENTERED TRIANGULAR NUMBER, CUBIC NUMBER, DECAGONAL NUMBER, FIGURATE NUMBER TRIANGLE, GNOMIC NUMBER, HEPTAGONAL NUMBER, HEPTAGONAL PYRAMIDAL NUMBER, HEX NUMBER, HEX PYRAMIDAL NUMBER, HEXAGONAL NUMBER, HEXAGONAL PYRAMIDAL NUMBER, NEXUS NUMBER, OCTAGONAL NUMBER, OCTAHEDRAL NUMBER, PENTAGONAL NUMBER, PENTAGONAL PYRAMIDAL NUMBER, PENTATOPE NUMBER, POLYGONAL NUMBER, PRONIC NUMBER, PYRAMIDAL NUMBER, RHOMBIC DODECAHEDRAL NUMBER, SQUARE NUMBER, STELLA OCTANGULA NUMBER, TETRAHEDRAL NUMBER, TRIANGULAR NUMBER, TRUNCATED OCTAHEDRAL NUMBER, TRUNCATED TETRAHEDRAL NUMBER

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 30–62, 1996.

Dickson, L. E. "Polygonal, Pyramidal, and Figurate Numbers." Ch. 1 in *History of the Theory of Numbers, Vol. 2: Diophantine Analysis.* New York: Chelsea, pp. 1–39, 1952.

Goodwin, P. "A Polyhedral Sequence of Two." *Math. Gaz.* **69**, 191–197, 1985.
Guy, R. K. "Figurate Numbers." §D3 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 147–150, 1994.
Kraitchik, M. "Figurate Numbers." §3.4 in *Mathematical Recreations.* New York: W. W. Norton, pp. 66–69, 1942.

Figurate Number Triangle

A PASCAL'S TRIANGLE written in a square grid and padded with zeroes, as written by Jakob Bernoulli (Smith 1984). The figurate number triangle therefore has entries

$$a_{ij} = \binom{i}{j},$$

where i is the row number, j the column number, and $\binom{i}{j}$ a BINOMIAL COEFFICIENT. Written out explicitly (beginning each row with $j = 0$),

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 2 & 1 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 3 & 3 & 1 & 0 & 0 & 0 & \cdots \\ 1 & 4 & 6 & 4 & 1 & 0 & 0 & \cdots \\ 1 & 5 & 10 & 10 & 5 & 1 & 0 & \cdots \\ 1 & 6 & 15 & 20 & 15 & 6 & 1 & \cdots \\ 1 & 7 & 21 & 35 & 35 & 21 & 7 & \ddots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Then we have the sum identities

$$\sum_{j=0}^{i} a_{ij} = 2^i$$

$$\sum_{j=1}^{i} a_{ij} = 2^i - 1$$

$$\sum_{i=0}^{n} a_{ij} = a_{(n+1),(j+1)} = \frac{n+1}{j+1} a_{nj}.$$

see also BINOMIAL COEFFICIENT, FIGURATE NUMBER, PASCAL'S TRIANGLE

References

Smith, D. E. *A Source Book in Mathematics.* New York: Dover, p. 86, 1984.

Figure Eight Knot

see FIGURE-OF-EIGHT KNOT

Figure Eight Surface

see EIGHT SURFACE

Figure-of-Eight Knot

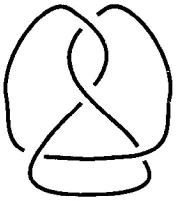

The knot 04_{001}, which is the unique PRIME KNOT of four crossings, and which is a 2-EMBEDDABLE KNOT. It is AMPHICHIRAL. It is also known as the FLEMISH KNOT and SAVOY KNOT, and it has BRAID WORD $\sigma_1\sigma_2{}^{-1}\sigma_1\sigma_2{}^{-1}$.

References

Owen, P. *Knots.* Philadelphia, PA: Courage, p. 16, 1993.

Figures

A number x is said to have "n figures" if it takes n DIGITS to express it. The number of figures is therefore equal to one more than the POWER of 10 in the SCIENTIFIC NOTATION representation of the number. The word is most frequently used in reference to monetary amounts, e.g., a "six-figure salary" would fall in the range of $100,000 to $999,999.

see also DIGIT, SCIENTIFIC NOTATION, SIGNIFICANT FIGURES

Filon's Integration Formula

A formula for NUMERICAL INTEGRATION,

$$\begin{aligned} \int_{x_0}^{x_n} & f(x)\cos(tx)\,dx \\ & = h\{\alpha(th)[f_{2n}\sin(tx_{2n}) - f_0\sin(tx_0)] \\ & \quad + \beta(th)C_{2n} + \gamma(th)C_{2n-1} + \tfrac{2}{45}th^4 S'_{2n-1}\} - R_n, \end{aligned} \tag{1}$$

where

$$C_{2n} = \sum_{i=0}^{n} f_{2i}\cos(tx_{2i}) - \tfrac{1}{2}[f_{2n}\cos(tx_{2n}) + f_0\cos(tx_0)] \tag{2}$$

$$C_{2n-1} = \sum_{i=1}^{n} f_{2i-1}\cos(tx_{2i-1}) \tag{3}$$

$$S'_{2n-1} = \sum_{i=1}^{n} f_{2i-1}^{(3)}\sin(tx_{2i-1}) \tag{4}$$

$$\alpha(\theta) = \frac{1}{\theta} + \frac{\sin(2\theta)}{2\theta^2} - \frac{2\sin^2\theta}{\theta^3} \tag{5}$$

$$\beta(\theta) = 2\left[\frac{1+\cos^2\theta}{\theta^2} - \frac{\sin(2\theta)}{\theta^3}\right] \tag{6}$$

$$\gamma(\theta) = 4\left(\frac{\sin\theta}{\theta^3} - \frac{\cos\theta}{\theta^2}\right), \tag{7}$$

and the remainder term is

$$R_n = \tfrac{1}{90}nh^5 f^{(4)}(\xi) + \mathcal{O}(th^7). \tag{8}$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 890–891, 1972.

Tukey, J. W. In *On Numerical Approximation: Proceedings of a Symposium Conducted by the Mathematics Research Center, United States Army, at the University of Wisconsin, Madison, April 21–23, 1958* (Ed. R. E. Langer). Madison, WI: University of Wisconsin Press, p. 400, 1959.

Filter

Formally, a filter is defined in terms of a SET X and a SET Φ of SUBSETS of X. Then Φ is called a filter if

1. $X \in \Phi$,
2. the EMPTY SET $\varnothing \notin \Phi$,
3. $A \subset B \subset X$ and $A \in \Phi$ IMPLIES $B \in \Phi$,
4. and $A, B \in \Phi$ IMPLIES $A \cup B \in \Phi$.

Informally, a filter is a function or procedure which removes unwanted parts of a signal. The concept of filtering and filter functions is particularly useful in engineering. One particularly elegant method of filtering FOURIER TRANSFORMS a signal into frequency space, performs the filtering operation there, then transforms back into the original space (Press *et al.* 1992).

see also SAVITZKY-GOLAY FILTER, WIENER FILTER

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Digital Filtering in the Time Domain." §13.5 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 551–556, 1992.

Fine's Equation

$$\prod_{n=1} \frac{(1-q^{2n})(1-q^{3n})(1-q^{8n})(1-q^{12n})}{(1-q^n)(1-q^{24n})} = 1 + \sum_{N=1} E_{1,5,7,11}(N;24)q^N,$$

where E is the sum of the DIVISORS of N CONGRUENT to 1, 5, 7, and 11 (mod 24) minus the sum of DIVISORS of N CONGRUENT to -1, -5, -7, and -11 (mod 24).

see also q-SERIES

Finite

A SET which contains a NONNEGATIVE integral number of elements is said to be finite. A SET which is not finite is said to be INFINITE. A finite or COUNTABLY INFINITE SET is said to be COUNTABLE. While the meaning of the term "finite" is fairly clear in common usage, precise definitions of FINITE and INFINITE are needed in technical mathematics and especially in SET THEORY.

see also COUNTABLE SET, COUNTABLY INFINITE SET, INFINITE, SET THEORY, UNCOUNTABLY INFINITE SET

Finite Difference

The finite difference is the discrete analog of the DERIVATIVE. The finite FORWARD DIFFERENCE of a function f_p is defined as

$$\Delta f_p \equiv f_{p+1} - f_p, \tag{1}$$

and the finite BACKWARD DIFFERENCE as

$$\nabla f_p \equiv f_p - f_{p-1}. \tag{2}$$

If the values are tabulated at spacings h, then the notation

$$f_p \equiv f(x_0 + ph) \equiv f(x) \tag{3}$$

is used. The kth FORWARD DIFFERENCE would then be written as $\Delta^k f_p$, and similarly, the kth BACKWARD DIFFERENCE as $\nabla^k f_p$.

However, when f_p is viewed as a discretization of the continuous function $f(x)$, then the finite difference is sometimes written

$$\Delta f(x) \equiv f(x+\tfrac{1}{2}) - f(x-\tfrac{1}{2}) = 2\,\mathrm{I_I}(x) * f(x), \tag{4}$$

where $*$ denotes CONVOLUTION and $\mathrm{I_I}(x)$ is the odd IMPULSE PAIR. The finite difference operator can therefore be written

$$\tilde{\Delta} = 2\,\mathrm{I_I}\,*. \tag{5}$$

An nth POWER has a constant nth finite difference. For example, take $n = 3$ and make a DIFFERENCE TABLE,

$$\begin{array}{cccccc}
x & x^3 & & & & \\
 & & \Delta & & & \\
1 & 1 & & \Delta^2 & & \\
 & & 7 & & \Delta^3 & \\
2 & 8 & & 12 & & \Delta^4 \\
 & & 19 & & 6 & \\
3 & 27 & & 18 & & 0 \\
 & & 37 & & 6 & \\
4 & 64 & & 24 & & \\
 & & 61 & & & \\
5 & 125 & & & &
\end{array}. \tag{6}$$

The Δ^3 column is the constant 6.

Finite difference formulas can be very useful for extrapolating a finite amount of data in an attempt to find the general term. Specifically, if a function $f(n)$ is known at only a few discrete values $n = 0, 1, 2, \ldots$ and it is desired to determine the analytical form of f, the following procedure can be used if f is assumed to be a POLYNOMIAL function. Denote the nth value in the SEQUENCE of interest by a_n. Then define b_n as the FORWARD DIFFERENCE $\Delta_n \equiv a_{n+1} - a_n$, c_n as the second FORWARD DIFFERENCE $\Delta_n^2 \equiv b_{n+1} - b_n$, etc., constructing a table as follows

$$\begin{array}{ccccc}
a_0 \equiv f(0) & a_1 \equiv f(1) & a_2 \equiv f(2) & \ldots & a_p \equiv f(p) \\
b_0 \equiv a_1 - a_0 & b_1 \equiv a_2 - a_1 & \ldots & b_{p-1} \equiv a_p - a_{p-1} & \\
 & c_0 \equiv b_2 - b_1 & \ldots & \ldots & \\
 & & \ddots & &
\end{array} \tag{7}$$

Continue computing d_0, e_0, etc., until a 0 value is obtained. Then the POLYNOMIAL function giving the values a_n is given by

$$f(n) = \sum_{k=0}^{p} \alpha_k \binom{n}{k} \tag{8}$$

$$= a_0 + b_0 n + \frac{c_0 n(n-1)}{2} + \frac{d_0 n(n-1)(n-2)}{2\cdot 3} + \ldots . \tag{9}$$

When the notation $\Delta_0 \equiv a_0$, $\Delta_0^2 \equiv b_0$, etc., is used, this beautiful equation is called NEWTON'S FORWARD DIFFERENCE FORMULA. To see a particular example, consider a SEQUENCE with first few values of 1, 19, 143, 607, 1789, 4211, and 8539. The difference table is then given by

$$\begin{array}{c}
1 \quad 19 \quad 143 \quad 607 \quad 1789 \quad 4211 \quad 8539 \\
18 \quad 124 \quad 464 \quad 1182 \quad 2422 \quad 4328 \\
106 \quad 340 \quad 718 \quad 1240 \quad 1906 \\
234 \quad 378 \quad 522 \quad 666 \\
144 \quad 144 \quad 144 \\
0 \quad 0
\end{array}$$

Reading off the first number in each row gives $a_0 = 1$, $b_0 = 18$, $c_0 = 106$, $d_0 = 234$, $e_0 = 144$. Plugging these in gives the equation

$$f(n) = 1 + 18n + 53n(n-1) + 39n(n-1)(n-2) + 6n(n-1)(n-2)(n-3), \tag{10}$$

which simplifies to $f(n) = 6n^4 + 3n^3 + 2n^2 + 7n + 1$, and indeed fits the original data exactly!

Beyer (1987) gives formulas for the derivatives

$$h^n \frac{d^n f(x_0 + ph)}{dx^n} \equiv h^n \frac{d^n f_p}{dx^n} \equiv \frac{d^n f_p}{dp^n} \tag{11}$$

(Beyer 1987, pp. 449–451) and integrals

$$\int_{x_0}^{x_n} f(x)\,dx = h \int_0^n f_p\,dp \tag{12}$$

(Beyer 1987, pp. 455-456) of finite differences.

Finite differences lead to DIFFERENCE EQUATIONS, finite analogs of DIFFERENTIAL EQUATIONS. In fact, UMBRAL CALCULUS displays many elegant analogs of well-known identities for continuous functions. Common finite difference schemes for PARTIAL DIFFERENTIAL EQUATIONS include the so-called Crank-Nicholson, Du Fort-Frankel, and Laasonen methods.

see also BACKWARD DIFFERENCE, BESSEL'S FINITE DIFFERENCE FORMULA, DIFFERENCE EQUATION, DIFFERENCE TABLE, EVERETT'S FORMULA, FORWARD DIFFERENCE, GAUSS'S BACKWARD FORMULA, GAUSS'S FORWARD FORMULA, INTERPOLATION, JACKSON'S DIFFERENCE FAN, NEWTON'S BACKWARD DIFFERENCE FORMULA, NEWTON-COTES FORMULAS, NEWTON'S DIVIDED DIFFERENCE INTERPOLATION FORMULA, NEWTON'S FORWARD DIFFERENCE FORMULA, QUOTIENT-DIFFERENCE TABLE, STEFFENSON'S FORMULA, STIRLING'S FINITE DIFFERENCE FORMULA, UMBRAL CALCULUS

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Differences." §25.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 877–878, 1972.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 429–515, 1987.

Boole, G. and Moulton, J. F. *A Treatise on the Calculus of Finite Differences, 2nd rev. ed.* New York: Dover, 1960.

Conway, J. H. and Guy, R. K. "Newton's Useful Little Formula." In *The Book of Numbers.* New York: Springer-Verlag, pp. 81–83, 1996.

Iyanaga, S. and Kawada, Y. (Eds.). "Interpolation." Appendix A, Table 21 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 1482–1483, 1980.

Jordan, K. *Calculus of Finite Differences, 2nd ed.* New York: Chelsea, 1950.

Levy, H. and Lessman, F. *Finite Difference Equations.* New York: Dover, 1992.

Milne-Thomson, L. M. *The Calculus of Finite Differences.* London: Macmillan, 1951.

Richardson, C. H. *An Introduction to the Calculus of Finite Differences.* New York: Van Nostrand, 1954.

Spiegel, M. *Calculus of Finite Differences and Differential Equations.* New York: McGraw-Hill, 1971.

Finite Field

A finite field is a FIELD with a finite ORDER (number of elements), also called a GALOIS FIELD. The order of a finite field is always a PRIME or a POWER of a PRIME (Birkhoff and Mac Lane 1965). For each PRIME POWER, there exists exactly one (up to an ISOMORPHISM) finite field GF(p^n), often written as $\mathbb{F}_{p^n}$ in current usage. GF(p) is called the PRIME FIELD of order p, and is the FIELD of RESIDUE CLASSES modulo p, where the p elements are denoted 0, 1, ..., $p-1$. $a = b$ in GF(p) means the same as $a \equiv b \pmod{p}$. Note, however, that $2 \times 2 \equiv 0 \pmod 4$ in the RING of residues modulo 4, so 2 has no reciprocal, and the RING of residues modulo 4 is distinct from the finite field with four elements. Finite fields are therefore denoted GF(p^n), instead of GF($p_1 \cdots p_n$) for clarity.

The finite field GF(2) consists of elements 0 and 1 which satisfy the following addition and multiplications tables.

+	0	1
0	0	1
1	1	0

×	0	1
0	0	0
1	0	1

If a subset S of the elements of a finite field F satisfies the above AXIOMS with the same operators of F, then S

is called a SUBFIELD. Finite fields are used extensively in the study of ERROR-CORRECTING CODES.

When $n > 1$, $\mathrm{GF}(p^n)$ can be represented as the FIELD of EQUIVALENCE CLASSES of POLYNOMIALS whose COEFFICIENTS belong to $\mathrm{GF}(p)$. Any IRREDUCIBLE POLYNOMIAL of degree n yields the same FIELD up to an ISOMORPHISM. For example, for $\mathrm{GF}(2^3)$, the modulus can be taken as $x^3+x^2+1=0$, x^3+x+1, or any other IRREDUCIBLE POLYNOMIAL of degree 3. Using the modulus x^3+x+1, the elements of $\mathrm{GF}(2^3)$—written 0, x^0, x^1, …—can be represented as POLYNOMIALS with degree less than 3. For instance,

$$\begin{aligned}
&x^3 \equiv -x-1 \equiv x+1\\
&x^4 \equiv x(x^3) \equiv x(x+1) \equiv x^2+x\\
&x^5 \equiv x(x^2+x) \equiv x^3+x^2 \equiv x^2-x-1 \equiv x^2+x+1\\
&x^6 \equiv x(x^2+x+1) \equiv x^3+x^2+x \equiv x^2-1 \equiv x^2+1\\
&x^7 \equiv x(x^2+1) \equiv x^3+x \equiv -1 \equiv 1 \equiv x_0.
\end{aligned}$$

Now consider the following table which contains several different representations of the elements of a finite field. The columns are the power, polynomial representation, triples of polynomial representation COEFFICIENTS (the vector representation), and the binary INTEGER corresponding to the vector representation (the regular representation).

	Representation		
Power	Polynomial	Vector	Regular
0	0	(000)	0
x^0	1	(001)	1
x^1	x	(010)	2
x^2	x^2	(100)	4
x^3	$x+1$	(011)	3
x^4	x^2+x	(110)	6
x^5	x^2+x+1	(111)	7
x^6	x^2+1	(101)	5

The set of POLYNOMIALS in the second column is closed under ADDITION and MULTIPLICATION modulo x^3+x+1, and these operations on the set satisfy the AXIOMS of finite field. This particular finite field is said to be an extension field of degree 3 of GF(2), written $\mathrm{GF}(2^3)$, and the field GF(2) is called the base field of $\mathrm{GF}(2^3)$. If an IRREDUCIBLE POLYNOMIAL generates all elements in this way, it is called a PRIMITIVE IRREDUCIBLE POLYNOMIAL. For any PRIME or PRIME POWER q and any POSITIVE INTEGER n, there exists a PRIMITIVE IRREDUCIBLE POLYNOMIAL of degree n over $\mathrm{GF}(q)$.

For any element c of $\mathrm{GF}(q)$, $c^q = c$, and for any NONZERO element d of $\mathrm{GF}(q)$, $d^{q-1}=1$. There is a smallest POSITIVE INTEGER n satisfying the sum condition $n \cdot 1 = 0$ in $\mathrm{GF}(q)$, which is called the characteristic of the finite field $\mathrm{GF}(q)$. The characteristic is a PRIME NUMBER for every finite field, and it is true that

$$(x+y)^p = x^p + y^p$$

over a finite field with characteristic p.

see also FIELD, HADAMARD MATRIX, RING, SUBFIELD

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 73–75, 1987.

Birkhoff, G. and Mac Lane, S. *A Survey of Modern Algebra, 3rd ed.* New York: Macmillan, p. 413, 1965.

Dickson, L. E. *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, p. viii, 1952.

Finite Game

A GAME in which each player has a finite number of moves and a finite number of of choices at each move.

see also GAME, ZERO-SUM GAME

References

Dresher, M. *The Mathematics of Games of Strategy: Theory and Applications.* New York: Dover, p. 2, 1981.

Finite Group

A GROUP of finite ORDER. Examples of finite groups are the MODULO MULTIPLICATION GROUPS and the POINT GROUPS. The CLASSIFICATION THEOREM of finite SIMPLE GROUPS states that the finite SIMPLE GROUPS can be classified completely into one of five types.

There is no known FORMULA to give the number of possible finite groups as a function of the ORDER h. It is possible, however, to determine the number of ABELIAN GROUPS using the KRONECKER DECOMPOSITION THEOREM, and there is at least one ABELIAN GROUP for every finite order h.

The following table gives the numbers and names of the first few groups of ORDER h. In the table, N_{NA} denotes the number of non-Abelian groups, N_A denotes the number of ABELIAN GROUPS, and N the total number of groups. In addition, Z_n denotes an CYCLIC GROUP of ORDER n, A_n an ALTERNATING GROUP, D_n a DIHEDRAL GROUP, Q_8 the group of the QUATERNIONS, T the cubic group, and $\otimes$ a DIRECT PRODUCT.

h	Name	N_{NA}	N_A	N
1	$\langle e \rangle$	1	0	1
2	Z_2	1	0	1
3	Z_3	1	0	1
4	$Z_2 \otimes Z_2, Z_4$	2	0	2
5	Z_5	1	0	1
6	Z_6, D_3	1	1	2
7	Z_7	1	0	1
8	$Z_2 \otimes Z_2 \otimes Z_2, Z_2 \otimes Z_4, Z_8, Q_8, D_4$	3	2	5
9	$Z_3 \otimes Z_3, Z_9$	2	0	2
10	Z_{10}, D_5	1	1	2
11	Z_{11}	1	0	1
12	$Z_2 \otimes Z_6, Z_{12}, A_4, D_6, T$	2	3	5
13	Z_{13}	1	0	1
14	Z_{14}, D_7	1	1	2
15	Z_{15}	1	0	1

Miller (1930) gave the number of groups for orders 1–100, including an erroneous 297 as the number of groups of ORDER 64. Senior and Lunn (1934, 1935) subsequently completed the list up to 215, but omitted 128 and 192. The number of groups of ORDER 64 was corrected in Hall and Senior (1964). James *et al.* (1990) found 2328 groups in 115 ISOCLINISM families of ORDER 128, correcting previous work, and O'Brien (1991) found the number of groups of ORDER 256. The number of groups is known for orders up to 1000, with the possible exception of 512 and 768. Besche and Eick (1998) have determined the number of finite groups of orders less than 1000 which are not powers of 2 or 3. These numbers appear in the *Magma*® database. The numbers of nonisomorphic finite groups N of each ORDER h for the first few hundred orders are given in the following table (Sloane's A000001—the very first sequence).

The number of ABELIAN GROUPS of ORDER h is denoted N_A (Sloane's A000688). The smallest order for which there exist $n = 1, 2, \ldots$ nonisomorphic groups are 1, 4, 75, 28, 8, 42, ... (Sloane's A046057). The incrementally largest number of nonisomorphic finite groups are 1, 2, 5, 14, 15, 51, 52, 267, 2328, ... (Sloane's A046058), which occur for orders 1, 4, 8, 16, 24, 32, 48, 64, 128, ... (Sloane's A046059).

h	N	N_A	h	N	N_A	h	N	N_A	h	N	N_A
1	1	1	51	1	1	101	1	1	151	1	1
2	1	1	52	5	2	102	4	1	152	12	3
3	1	1	53	1	1	103	1	1	153	2	2
4	2	2	54	15	3	104	14	3	154	4	1
5	1	1	55	2	1	105	2	1	155	2	1
6	2	1	56	13	3	106	2	1	156	18	2
7	1	1	57	2	1	107	1	1	157	1	1
8	5	3	58	2	1	108	45	6	158	2	1
9	2	2	59	1	1	109	1	1	159	1	1
10	2	1	60	13	2	110	6	1	160	238	7
11	1	1	61	1	1	111	2	1	161	1	1
12	5	2	62	2	1	112	43	5	162	55	5
13	1	1	63	4	2	113	1	1	163	1	1
14	2	1	64	267	11	114	6	1	164	5	2
15	1	1	65	1	1	115	1	1	165	2	1
16	14	5	66	4	1	116	5	2	166	2	1
17	1	1	67	1	1	117	4	2	167	1	1
18	5	2	68	5	2	118	2	1	168	57	3
19	1	1	69	1	1	119	1	1	169	2	2
20	5	2	70	4	1	120	47	3	170	4	1
21	2	1	71	1	1	121	2	2	171	5	2
22	2	1	72	50	6	122	2	1	172	4	2
23	1	1	73	1	1	123	1	1	173	1	1
24	15	3	74	2	1	124	4	2	174	4	1
25	2	2	75	3	2	125	5	3	175	2	2
26	2	1	76	4	2	126	16	2	176	42	5
27	5	3	77	1	1	127	1	1	177	1	1
28	4	2	78	6	1	128	2328	15	178	2	1
29	1	1	79	1	1	129	2	1	179	1	1
30	4	1	80	52	5	130	4	1	180	37	4
31	1	1	81	15	5	131	1	1	181	1	1
32	51	7	82	2	1	132	10	2	182	4	1
33	1	1	83	1	1	133	1	1	183	2	1
34	2	1	84	15	2	134	2	1	184	12	3
35	1	1	85	1	1	135	5	3	185	1	1
36	14	4	86	2	1	136	15	3	186	6	1
37	1	1	87	1	1	137	1	1	187	1	1
38	2	1	88	12	3	138	4	1	188	4	2
39	2	1	89	1	1	139	1	1	189	13	3
40	14	3	90	10	2	140	11	2	190	4	1
41	1	1	91	1	1	141	1	1	191	1	1
42	6	1	92	4	2	142	2	1	192	1543	11
43	1	1	93	2	1	143	1	1	193	1	1
44	4	2	94	2	1	144	197	10	194	2	1
45	2	2	95	1	1	145	1	1	195	2	1
46	2	1	96	230	7	146	2	1	196	17	4
47	1	1	97	1	1	147	6	2	197	1	1
48	52	5	98	5	2	148	5	2	198	10	2
49	2	2	99	2	2	149	1	1	199	1	1
50	2	2	100	16	4	150	13	2	200	52	6

h	N	N_A	h	N	N_A	h	N	N_A	h	N	N_A
201	2	1	251	1	1	301	2	1	351	14	3
202	2	1	252	46	4	302	2	1	352	195	7
203	2	1	253	2	1	303	1	1	353	1	1
204	12	2	254	2	1	304	42	5	354	4	1
205	2	1	255	1	1	305	2	1	355	2	1
206	2	1	256	56092	22	306	10	2	356	5	2
207	2	2	257	1	1	307	1	1	357	2	1
208	51	5	258	6	1	308	9	2	358	2	1
209	1	1	259	1	1	309	2	1	359	1	1
210	12	1	260	15	2	310	6	1	360	162	6
211	1	1	261	2	2	311	1	1	361	2	2
212	5	2	262	2	1	312	61	3	362	2	1
213	1	1	263	1	1	313	1	1	363	3	2
214	2	1	264	39	3	314	2	1	364	11	2
215	1	1	265	1	1	315	4	2	365	1	1
216	177	9	266	4	1	316	4	2	366	6	1
217	1	1	267	1	1	317	1	1	367	1	1
218	2	1	268	4	2	318	4	1	368	42	5
219	2	1	269	1	1	319	1	1	369	2	2
220	15	2	270	30	3	320	1640	11	370	4	1
221	1	1	271	1	1	321	1	1	371	1	1
222	6	1	272	54	5	322	4	1	372	15	2
223	1	1	273	5	1	323	1	1	373	1	1
224	197	7	274	2	1	324	176	10	374	4	1
225	6	4	275	4	2	325	2	2	375	7	3
226	2	1	276	10	2	326	2	1	376	12	3
227	1	1	277	1	1	327	2	1	377	1	1
228	15	2	278	2	1	328	15	3	378	60	3
229	1	1	279	4	2	329	1	1	379	1	1
230	4	1	280	40	3	330	12	1	380	11	2
231	2	1	281	1	1	331	1	1	381	2	1
232	14	3	282	4	1	332	4	2	382	2	1
233	1	1	283	1	1	333	5	2	383	1	1
234	16	2	284	4	2	334	2	1	384	20169	15
235	1	1	285	2	1	335	1	1	385	2	1
236	4	2	286	4	1	336	228	5	386	2	1
237	2	1	287	1	1	337	1	1	387	4	2
238	4	1	288	1045	14	338	5	2	388	5	2
239	1	1	289	2	2	339	1	1	389	1	1
240	208	5	290	4	1	340	15	2	390	12	1
241	1	1	291	2	1	341	1	1	391	1	1
242	5	2	292	5	2	342	18	2	392	44	6
243	67	7	293	1	1	343	5	3	393	1	1
244	5	2	294	23	2	344	12	3	394	2	1
245	2	2	295	1	1	345	1	1	395	1	1
246	4	1	296	14	3	346	2	1	396	30	4
247	1	1	297	5	3	347	1	1	397	1	1
248	12	3	298	2	1	348	12	2	398	2	1
249	1	1	299	1	1	349	1	1	399	5	1
250	15	3	300	49	4	350	10	2	400	221	10

see also Abelian Group, Abel's Theorem, Abhyankar's Conjecture, Alternating Group, Burnside's Lemma, Burnside Problem, Chevalley Groups, Classification Theorem, Composition Series, Dihedral Group, Group, Jordan-Hölder Theorem, Kronecker Decomposition Theorem, Lie Group, Lie-Type Group, Linear Group, Modulo Multiplication Group, Order (Group), Orthogonal Group, p-Group, Point Groups, Simple Group, Sporadic Group, Symmetric Group, Symplectic Group, Twisted Chevalley Groups, Unitary Group

References

Arfken, G. "Discrete Groups." §4.9 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 243–251, 1985.

Artin, E. "The Order of the Classical Simple Groups." *Comm. Pure Appl. Math.* **8**, 455–472, 1955.

Aschbacher, M. *Finite Group Theory.* Cambridge, England: Cambridge University Press, 1994.

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 73–75, 1987.

Besche and Eick. "Construction of Finite Groups." To Appear in *J. Symb. Comput.*

Besche and Eick. "The Groups of Order at Most 1000." To Appear in *J. Symb. Comput.*

Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, 1985.

Hall, M. Jr. and Senior, J. K. *The Groups of Order* $2^n (n \leq 6)$. New York: Macmillan, 1964.

James, R.; Newman, M. F.; and O'Brien, E. A. "The Groups of Order 128." *J. Algebra* **129**, 136–158, 1990.

Miller, G. A. "Determination of All the Groups of Order 64." *Amer. J. Math.* **52**, 617–634, 1930.

O'Brien, E. A. "The Groups of Order 256." *J. Algebra* **143**, 219–235, 1991.

O'Brien, E. A. and Short, M. W. "Bibliography on Classification of Finite Groups." Manuscript, Australian National University, 1988.

Senior, J. K. and Lunn, A. C. "Determination of the Groups of Orders 101–161, Omitting Order 128." *Amer. J. Math.* **56**, 328–338, 1934.

Senior, J. K. and Lunn, A. C. "Determination of the Groups of Orders 162–215, Omitting Order 192." *Amer. J. Math.* **57**, 254–260, 1935.

Simon, B. *Representations of Finite and Compact Groups.* Providence, RI: Amer. Math. Soc., 1996.

Sloane, N. J. A. Sequences A000001/M0098 and A000688/M0064 in "An On-Line Version of the Encyclopedia of Integer Sequences."

University of Sydney Computational Algebra Group. "The Magma Computational Algebra for Algebra, Number Theory and Geometry." `http://www.maths.usyd.edu.au:8000/u/magma/`.

✻ Weisstein, E. W. "Groups." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Groups.m`.

Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas`.

Finite Group—D_3

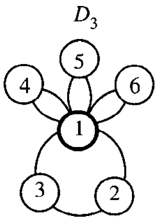

The Dihedral Group D_3 is one of the two groups of Order 6. It the non-Abelian group of smallest Order. Examples of D_3 include the Point Groups known as C_{3h}, C_{3v}, S_3, D_3, the symmetry group of the Equilateral Triangle, and the group of permutation of three objects. Its elements A_i satisfy $A_i^3 = 1$, and four of its elements satisfy $A_i^2 = 1$, where 1 is the Identity Element. The Cycle Graph is shown above, and the Multiplication Table is given below.

D_3	1	A	B	C	D	E
1	1	A	B	C	D	E
A	A	1	D	E	B	C
B	B	E	1	D	C	A
C	C	D	E	1	A	B
D	D	C	A	B	E	1
E	E	B	C	A	1	D

The Conjugacy Classes are $\{1\}$, $\{A, B, C\}$

$$A^{-1}AA = A \tag{1}$$

$$B^{-1}AB = C \tag{2}$$

$$C^{-1}AC = B \tag{3}$$

$$D^{-1}AD = C \tag{4}$$

$$E^{-1}AE = B, \tag{5}$$

and $\{D, E\}$,

$$DA^{-1}D = E \tag{6}$$

$$B^{-1}\mathcal{D}B = D. \tag{7}$$

A reducible 2-D representation using Real Matrices can be found by performing the spatial rotations corresponding to the symmetry elements of C_{3v}. Take the z-Axis along the C_3 axis.

$$I = R_z(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{8}$$

$$A = R_z(\tfrac{2}{3}\pi) = \begin{bmatrix} \cos(\frac{2}{3}\pi) & \sin(\frac{2}{3}\pi) \\ -\sin(\frac{2}{3}\pi) & \cos(\frac{2}{3}\pi) \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{bmatrix} \tag{9}$$

$$B = R_z(\tfrac{4}{3}\pi) = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2}\sqrt{3} \\ -\frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{bmatrix} \tag{10}$$

$$C = R_C(\pi) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \tag{11}$$

$$D = R_D(\pi) = CB = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2}\sqrt{3} \\ -\frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{bmatrix} \tag{12}$$

$$E = R_E(\pi) = CA = \begin{bmatrix} \frac{1}{2} & \frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{bmatrix}. \tag{13}$$

To find the irreducible representation, note that there are three Conjugacy Classes. Rule 5 requires that there be three irreducible representations satisfying

$$h = l_1^2 + l_2^2 + l_3^2 = 6, \tag{14}$$

so it must be true that

$$l_1 = l_2 = 1, l_3 = 2. \tag{15}$$

By rule 6, we can let the first representation have all 1s.

$\mathcal{D}_3$	1	A	B	C	D	E
Γ_1	1	1	1	1	1	1

To find representation orthogonal to the totally symmetric representation, we must have three +1 and three −1 Characters. We can also add the constraint that the components of the Identity Element 1 be positive. The three Conjugacy Classes have 1, 2, and 3 elements. Since we need a total of three +1s and we have required that a +1 occur for the Conjugacy Class of Order 1, the remaining +1s must be used for the elements of the Conjugacy Class of Order 2, i.e., A and B.

D_3	1	A	B	C	D	E
Γ_1	1	1	1	1	1	1
Γ_2	1	1	1	−1	−1	−1

Using the rule 1, we see that

$$1^2 + 1^2 + {\chi_3}^2(1) = 6, \tag{16}$$

so the final representation for 1 has Character 2. Orthogonality with the first two representations (rule 3) then yields the following constraints:

$$1 \cdot 1 \cdot 2 + 1 \cdot 2 \cdot \chi_2 + 1 \cdot 3 \cdot \chi_3 = 2 + 2\chi_2 + 3\chi_3 = 0 \tag{17}$$

$$1 \cdot 1 \cdot 2 + 1 \cdot 2 \cdot \chi_2 + (-1) \cdot 3 \cdot \chi_3 = 2 + 2\chi_2 - 3\chi_3 = 0. \tag{18}$$

Solving these simultaneous equations by adding and subtracting (18) from (17), we obtain $\chi_2 = -1$, $\chi_3 = 0$. The full Character Table is then

D_3	1	A	B	C	D	E
Γ_1	1	1	1	1	1	1
Γ_2	1	1	1	−1	−1	−1
Γ_3	2	−1	−1	0	0	0

Since there are only three Conjugacy Classes, this table is conventionally written simply as

D_3	1	$A = B$	$C = D = E$
Γ_1	1	1	1
Γ_2	1	1	−1
Γ_3	2	−1	0

Writing the irreducible representations in matrix form then yields

$$1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{19}$$

$$A = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2}\sqrt{3} & 0 & 0 \\ \frac{1}{2}\sqrt{3} & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{20}$$

$$B = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2}\sqrt{3} & 0 & 0 \\ -\frac{1}{2}\sqrt{3} & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{21}$$

$$C = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \tag{22}$$

$$D = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2}\sqrt{3} & 0 & 0 \\ -\frac{1}{2}\sqrt{3} & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \tag{23}$$

$$E = \begin{bmatrix} \frac{1}{2} & \frac{1}{2}\sqrt{3} & 0 & 0 \\ \frac{1}{2}\sqrt{3} & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \tag{24}$$

see also DIHEDRAL GROUP, FINITE GROUP—D_4, FINITE GROUP—Z_6

Finite Group—D_4

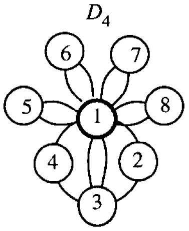

The DIHEDRAL GROUP D_4 is one of the two non-Abelian groups of the five groups total of ORDER 8. It is sometimes called the octic group. Examples of D_4 include the symmetry group of the SQUARE. The CYCLE GRAPH is shown above.

see also DIHEDRAL GROUP, FINITE GROUP—D_3, FINITE GROUP—Z_8, FINITE GROUP—$Z_2 \otimes Z_2 \otimes Z_2$, FINITE GROUP—$Z_2 \otimes Z_4$, FINITE GROUP—$Z_8$,

Finite Group—$\langle e \rangle$

The unique (and trivial) group of ORDER 1 is denoted $\langle e \rangle$. It is (trivially) ABELIAN and CYCLIC. Examples include the POINT GROUP C_1 and the integers modulo 1 under addition.

$\langle e \rangle$	1
1	1

The only class is $\{1\}$.

Finite Group—Q_8

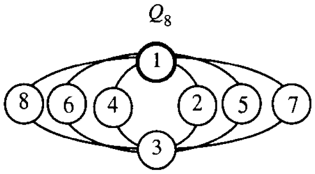

One of the three Abelian groups of the five groups total of ORDER 8. The group Q_8 has the MULTIPLICATION TABLE of $\pm 1, i, j, k$, where 1, i, j, and k are the QUATERNIONS. The CYCLE GRAPH is shown above.

see also FINITE GROUP—D_4, FINITE GROUP—$Z_2 \otimes Z_2 \otimes Z_2$, FINITE GROUP—$Z_2 \otimes Z_4$, FINITE GROUP—$Z_8$, QUATERNION

Finite Group—Z_2

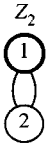

The unique group of ORDER 2. Z_2 is both ABELIAN and CYCLIC. Examples include the POINT GROUPS C_s, C_i, and C_2, the integers modulo 2 under addition, and the MODULO MULTIPLICATION GROUPS M_3, M_4, and M_6. The elements A_i satisfy ${A_i}^2 = 1$, where 1 is the IDENTITY ELEMENT. The CYCLE GRAPH is shown above, and the MULTIPLICATION TABLE is given below.

Z_2	1	A
1	1	A
A	A	1

The CONJUGACY CLASSES are $\{1\}$ and $\{A\}$. The irreducible representation for the C_2 group is $\{1, -1\}$.

Finite Group—$Z_2 \otimes Z_2$

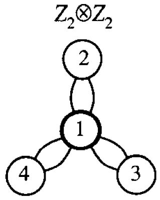

One of the two groups of ORDER 4. The name of this group derives from the fact that it is a DIRECT PRODUCT of two Z_2 SUBGROUPS. Like the group Z_4, $Z_2 \otimes Z_2$ is an ABELIAN GROUP. Unlike Z_4, however, it is not CYCLIC. In addition to satisfying ${A_i}^4 = 1$ for each element A_i, it also satisfies ${A_i}^2 = 1$, where 1 is the IDENTITY ELEMENT. Examples of the $Z_2 \otimes Z_2$ group include the VIERGRUPPE, POINT GROUPS D_2, C_{2h}, and C_{2v}, and the MODULO MULTIPLICATION GROUPS M_8 and M_{12}. That M_8, the RESIDUE CLASSES prime to 8 given by {1, 3, 5, 7}, are a group of type $Z_2 \otimes Z_2$ can be shown by verifying that

$$1^2 = 1 \quad 3^2 = 9 \equiv 1 \quad 5^2 = 25 \equiv 1 \quad 7^2 = 49 \equiv 1 \pmod{8} \tag{1}$$

and

$$3 \cdot 5 = 15 \equiv 7 \quad 3 \cdot 7 = 21 \equiv 5 \quad 5 \cdot 7 = 35 \equiv 3 \pmod{8}. \tag{2}$$

$Z_2 \otimes Z_2$ is therefore a MODULO MULTIPLICATION GROUP.

The CYCLE GRAPH is shown above, and the multiplication table for the $Z_2 \otimes Z_2$ group is given below.

$Z_2 \otimes Z_2$	1	A	B	C
1	1	A	B	C
A	A	1	C	B
B	B	C	1	A
C	C	B	A	1

The CONJUGACY CLASSES are $\{1\}$, $\{A\}$,

$$A^{-1}AA = A \tag{3}$$

$$B^{-1}AB = A \tag{4}$$

$$C^{-1}AC = A, \tag{5}$$

$\{B\}$,

$$A^{-1}BA = B \tag{6}$$

$$C^{-1}BC = B, \tag{7}$$

and $\{C\}$.

Now explicitly consider the elements of the C_{2v} POINT GROUP.

C_{2v}	E	C_2	σ_v	σ_v
E	E	C_2	σ_v	σ_v'
C_2	C_2	E	σ_v'	σ_v
σ_v	σ_v	σ_v'	E	C_2
σ_v'	σ_v'	σ_v	C_2	E

In terms of the VIERGRUPPE elements

V	I	V_1	V_2	V_3
I	V_1	V_2	V_3	V_4
V_1	V_1	I	V_3	V_2
V_2	V_2	V_3	I	V_1
V_3	V_3	V_2	V_1	I

A reducible representation using 2-D REAL MATRICES is

$$1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{8}$$

$$A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \tag{9}$$

$$B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{10}$$

$$C = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}. \tag{11}$$

Another reducible representation using 3-D REAL MATRICES can be obtained from the symmetry elements of the D_2 group (1, $C_2(z)$, $C_2(y)$, and $C_2(x)$) or C_{2v} group (1, C_2, σ_v, and σ_v'). Place the C_2 axis along the z-axis, σ_v in the x-y plane, and σ_v' in the y-z plane.

$$1 = E = E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{12}$$

$$A = R_x(\pi) = \sigma_v = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{13}$$

$$C = R_z(\pi) = C_2 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{14}$$

$$B = R_y(\pi) = \sigma_v' = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{15}$$

In order to find the irreducible representations, note that the traces are given by $\chi(1) = 3, \chi(C_2) = -1$, and $\chi(\sigma_v) = \chi(\sigma_v') = 1$. Therefore, there are at least three distinct CONJUGACY CLASSES. However, we see from the MULTIPLICATION TABLE that there are actually four CONJUGACY CLASSES, so group rule 5 requires that there must be four irreducible representations. By rule 1, we are looking for POSITIVE INTEGERS which satisfy

$$l_1{}^2 + l_2{}^2 + l_3{}^2 + l_4{}^2 = 4. \tag{16}$$

The only combination which will work is

$$l_1 = l_2 = l_3 = l_4 = 1, \tag{17}$$

so there are four one-dimensional representations. Rule 2 requires that the sum of the squares equal the ORDER $h = 4$, so each 1-D representation must have CHARACTER ±1. Rule 6 requires that a totally symmetric representation always exists, so we are free to start off with the first representation having all 1s. We then use orthogonality (rule 3) to build up the other representations. The simplest solution is then given by

C_{2v}	1	C_2	σ_v	σ_v'
Γ_1	1	1	1	1
Γ_2	1	-1	-1	1
Γ_3	1	-1	1	-1
Γ_4	1	1	-1	-1

These can be put into a more familiar form by switching Γ_1 and Γ_3, giving the CHARACTER TABLE

C_{2v}	1	C_2	σ_v	σ_v'
Γ_3	1	-1	1	-1
Γ_2	1	-1	-1	1
Γ_1	1	1	1	1
Γ_4	1	1	-1	-1

The matrices corresponding to this representation are now

$$1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{18}$$

$$C_2 = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{19}$$

$$\sigma_v = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \tag{20}$$

$$\sigma_v' = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}, \tag{21}$$

which consist of the previous representation with an additional component. These matrices are now orthogonal, and the order equals the matrix dimension. As before, $\chi(\sigma_v) = \chi(\sigma_v')$.

see also FINITE GROUP—Z_4

Finite Group—$Z_2 \otimes Z_2 \otimes Z_2$

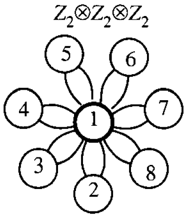

One of the three Abelian groups of the five groups total of ORDER 8. Examples include the MODULO MULTIPLICATION GROUP M_{24}. The elements A_i of this group satisfy $A_i{}^2 = 1$, where 1 is the IDENTITY ELEMENT. The CYCLE GRAPH is shown above.

see also FINITE GROUP—D_4, FINITE GROUP—Q_8, FINITE GROUP—$Z_2 \otimes Z_4$, FINITE GROUP—Z_8

Finite Group—$Z_2 \otimes Z_4$

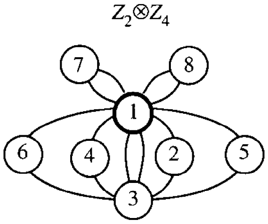

One of the three Abelian groups of the five groups total of ORDER 8. Examples include the MODULO MULTIPLICATION GROUPS M_{15}, M_{16}, M_{20}, and M_{30}. The elements A_i of this group satisfy $A_i{}^4 = 1$, where 1 is the IDENTITY ELEMENT, and four of the elements satisfy $A_i{}^2 = 1$. The CYCLE GRAPH is shown above.

see also FINITE GROUP—D_4, FINITE GROUP—Q_8, FINITE GROUP—$Z_2 \otimes Z_2 \otimes Z_2$, FINITE GROUP—$Z_8$

Finite Group—Z_3

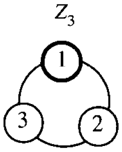

The unique group of ORDER 3. It is both ABELIAN and CYCLIC. Examples include the POINT GROUPS C_3 and D_3 and the integer modulo 3. The elements A_i of the group satisfy $A_i{}^3 = 1$ where 1 is the IDENTITY ELEMENT. The CYCLE GRAPH is shown above, and the MULTIPLICATION TABLE is given below.

Z_3	1	A	B
1	1	A	B
A	A	B	1
B	B	1	A

The CONJUGACY CLASSES are $\{1\}$, $\{A\}$,

$$A^{-1}AA = A$$
$$B^{-1}AB = A,$$

and $\{B\}$,

$$A^{-1}BA = B$$
$$B^{-1}BB = B.$$

The irreducible representation (CHARACTER TABLE) is therefore

Γ	1	A	B
Γ_1	1	1	1
Γ_2	1	1	-1
Γ_3	1	-1	1

Finite Group—Z_4

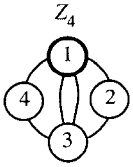

One of the two groups of ORDER 4. Like $Z_2 \otimes Z_2$, it is ABELIAN, but unlike $Z_2 \otimes Z_2$, it is a CYCLIC. Examples include the POINT GROUPS C_4 and S_4 and the MODULO MULTIPLICATION GROUPS M_5 and M_{10}. Elements A_i of the group satisfy $A_i{}^4 = 1$, where 1 is the IDENTITY ELEMENT, and two of the elements satisfy $A_i{}^2 = 1$.

The CYCLE GRAPH is shown above. The MULTIPLICATION TABLE for this group may be written in three equivalent ways—denoted here by $Z_4^{(1)}$, $Z_4^{(2)}$, and $Z_4^{(3)}$—by permuting the symbols used for the group elements.

$Z_4^{(1)}$	1	A	B	C
1	1	A	B	C
A	A	B	C	1
B	B	C	1	A
C	C	1	A	B

The MULTIPLICATION TABLE for $Z_4^{(2)}$ is obtained from $Z_4^{(1)}$ by interchanging A and B.

$Z_4^{(2)}$	1	A	B	C
1	1	A	B	C
A	A	1	C	B
B	B	C	A	1
C	C	B	1	A

The MULTIPLICATION TABLE for $Z_4^{(3)}$ is obtained from $Z_4^{(1)}$ by interchanging A and C.

$Z_4^{(3)}$	1	A	B	C
1	1	A	B	C
A	A	C	1	B
B	B	1	C	A
C	C	B	A	1

The CONJUGACY CLASSES of Z_4 are $\{1\}$, $\{A\}$,

$$A^{-1}AA = A \tag{1}$$

$$B^{-1}AB = A \tag{2}$$

$$C^{-1}AC = A, \tag{3}$$

$\{B\}$,

$$A^{-1}BA = B \tag{4}$$

$$B^{-1}BB = B \tag{5}$$

$$C^{-1}BC = B, \tag{6}$$

and $\{C\}$.

The group may be given a reducible representation using COMPLEX NUMBERS

$$1 = 1 \tag{7}$$

$$A = i \tag{8}$$

$$B = -1 \tag{9}$$

$$C = -i, \tag{10}$$

or REAL MATRICES

$$1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{11}$$

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{12}$$

$$B = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \tag{13}$$

$$C = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{14}$$

see also FINITE GROUP—$Z_2 \otimes Z_2$

Finite Group—Z_5

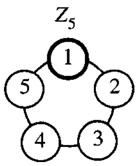

The unique GROUP of ORDER 5, which is ABELIAN. Examples include the POINT GROUP C_5 and the integers mod 5 under addition. The elements A_i satisfy $A_i{}^5 = 1$, where 1 is the IDENTITY ELEMENT. The CYCLE GRAPH is shown above, and the MULTIPLICATION TABLE is illustrated below.

Z_5	1	A	B	C	D
1	1	A	B	C	D
A	A	B	C	D	1
B	B	C	D	1	A
C	C	D	1	A	B
D	D	1	A	B	C

The CONJUGACY CLASSES are $\{1\}$, $\{A\}$, $\{B\}$, $\{C\}$, and $\{D\}$.

Finite Group—Z_6

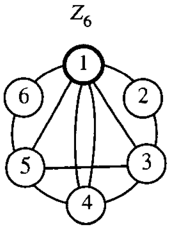

One of the two groups of ORDER 6 which, unlike D_3, is ABELIAN. It is also a CYCLIC. It is isomorphic to $Z_2 \otimes Z_3$. Examples include the POINT GROUPS C_6 and S_6, the integers modulo 6 under addition, and the MODULO MULTIPLICATION GROUPS M_7, M_9, and M_{14}. The elements A_i of the group satisfy $A_i{}^6 = 1$, where 1 is the IDENTITY ELEMENT, three elements satisfy $A_i{}^3 = 1$, and two elements satisfy $A_i{}^2 = 1$. The CYCLE GRAPH is shown above, and the MULTIPLICATION TABLE is given below.

Z_6	1	A	B	C	D	E
1	1	A	B	C	D	E
A	A	1	E	D	B	C
B	B	E	1	A	C	D
C	C	D	A	1	E	B
D	D	B	C	E	1	A
E	E	C	D	B	A	1

The CONJUGACY CLASSES are $\{1\}$, $\{A\}$, $\{B\}$, $\{C\}$, $\{D\}$, and $\{E\}$.

see also FINITE GROUP—D_3

Finite Group—Z_7

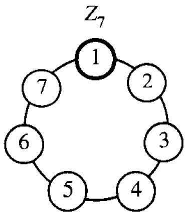

The unique GROUP of ORDER 7. It is ABELIAN and CYCLIC. Examples include the POINT GROUP C_7 and the integers modulo 7 under addition. The elements A_i of the group satisfy ${A_i}^7 = 1$, where 1 is the IDENTITY ELEMENT. The CYCLE GRAPH is shown above.

Z_7	1	A	B	C	D	E	F
1	1	A	B	C	D	E	F
A	A	B	C	D	E	F	1
B	B	C	D	E	F	1	A
C	C	D	E	F	1	A	B
D	D	E	F	1	A	B	C
E	E	F	1	A	B	C	D
F	F	1	A	B	C	D	E

The CONJUGACY CLASSES are $\{1\}$, $\{A\}$, $\{B\}$, $\{C\}$, $\{D\}$, $\{E\}$, and $\{F\}$.

Finite Group—Z_8

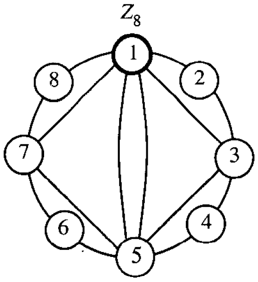

One of the three Abelian groups of the five groups total of ORDER 8. An example is the residue classes modulo 17 which QUADRATIC RESIDUES, i.e., {1, 2, 4, 8, 9, 13, 15, 16} under multiplication modulo 17. The elements A_i satisfy ${A_i}^8 = 1$, four of them satisfy ${A_i}^4 = 1$, and two satisfy ${A_i}^2 = 1$. The CYCLE GRAPH is shown above.

see also FINITE GROUP—D_4, FINITE GROUP—Q_8, FINITE GROUP—$Z_2 \otimes Z_4$, FINITE GROUP—$Z_2 \otimes Z_2 \otimes Z_2$

Finite Mathematics

The branch of mathematics which does not involve infinite sets, limits, or continuity.

see also COMBINATORICS, DISCRETE MATHEMATICS

References

Hildebrand, F. H. and Johnson, C. G. *Finite Mathematics.* Boston, MA: Prindle, Weber, and Schmidt, 1970.

Kemeny, J. G.; Snell, J. L.; and Thompson, G. L. *Introduction to Finite Mathematics, 3rd ed.* Englewood Cliffs, NJ: Prentice-Hall, 1974.

Marcus, M. *A Survey of Finite Mathematics.* New York: Dover, 1993.

Finite Simple Group

see SIMPLE GROUP

Finite Simple Group Classification Theorem

see CLASSIFICATION THEOREM

Finite-to-One Factor

A MAP $\psi : M \to M$, where M is a MANIFOLD, is a finite-to-one factor of a MAP $\Psi : X \to X$ if there exists a continuous ONTO MAP $\pi : X \to M$ such that $\psi \circ \pi = \pi \circ \Psi$ and $\pi^{-1}(x) \subset X$ is finite for each $x \in M$.

Finsler Geometry

The geometry of FINSLER SPACE.

Finsler Manifold

see FINSLER SPACE

Finsler Metric

A continuous real function $L(x, y)$ defined on the TANGENT BUNDLE $T(M)$ of an n-D DIFFERENTIABLE MANIFOLD M is said to be a Finsler metric if

1. $L(x, y)$ is DIFFERENTIABLE at $x \neq y$,
2. $L(x, \lambda y) = |\lambda| L(x, y)$ for any element $(x, y) \in T(M)$ and any REAL NUMBER λ,
3. Denoting the METRIC

$$g_{ij}(x, y) = \frac{1}{2} \frac{\partial^2 [L(x, y)]^2}{\partial y^i \partial y^j},$$

then g_{ij} is a POSITIVE DEFINITE MATRIX.

A DIFFERENTIABLE MANIFOLD M with a Finsler metric is called a FINSLER SPACE.

see also DIFFERENTIABLE MANIFOLD, FINSLER SPACE, TANGENT BUNDLE

References

Iyanaga, S. and Kawada, Y. (Eds.). "Finsler Spaces." §161 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 540–542, 1980.

Finsler Space

A general space based on the LINE ELEMENT

$$ds = F(x^1, \ldots, x^n; dx^1, \ldots, dx^n),$$

with $F(x, y) > 0$ for $y \neq 0$ a function on the TANGENT BUNDLE $T(M)$, and homogeneous of degree 1 in y. Formally, a Finsler space is a DIFFERENTIABLE MANIFOLD possessing a FINSLER METRIC. Finsler geometry is RIEMANNIAN GEOMETRY *without* the restriction that the LINE ELEMENT be quadratic of the form

$$F^2 = g_{ij}(x)\, dx^i\, dx^j.$$

A compact boundaryless Finsler space is locally Minkowskian IFF it has 0 "flag curvature."

see also FINSLER METRIC, HODGE'S THEOREM, RIEMANNIAN GEOMETRY, TANGENT BUNDLE

References
Akbar-Zadeh, H. "Sur les espaces de Finsler à courbures sectionnelles constantes." *Acad. Roy. Belg. Bull. Cl. Sci.* **74**, 281–322, 1988.
Bao, D.; Chern, S.-S.; and Shen, Z. (Eds.). *Finsler Geometry.* Providence, RI: Amer. Math. Soc., 1996.
Chern, S.-S. "Finsler Geometry is Just Riemannian Geometry without the Quadratic Restriction." *Not. Amer. Math. Soc.* **43**, 959–963, 1996.
Iyanaga, S. and Kawada, Y. (Eds.). "Finsler Spaces." §161 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 540–542, 1980.

Finsler-Hadwiger Theorem

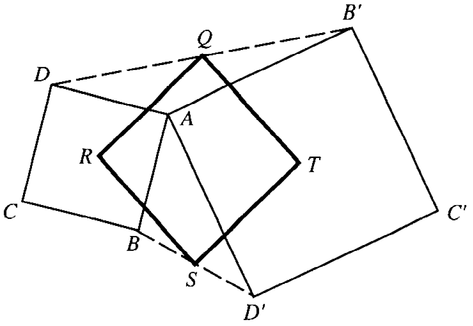

Let the SQUARES $\square ABCD$ and $\square AB'C'D'$ share a common VERTEX A. The midpoints Q and S of the segments $B'D$ and BD' together with the centers of the original squares R and T then form another square $\square QRST$. This theorem is a special case of the FUNDAMENTAL THEOREM OF DIRECTLY SIMILAR FIGURES (Detemple and Harold 1996).

see also FUNDAMENTAL THEOREM OF DIRECTLY SIMILAR FIGURES, SQUARE

References
Detemple, D. and Harold, S. "A Round-Up of Square Problems." *Math. Mag.* **69**, 15–27, 1996.
Finsler, P. and Hadwiger, H. "Einige Relationen im Dreieck." *Comment. Helv.* **10**, 316–326, 1937.
Fisher, J. C.; Ruoff, D.; and Shileto, J. "Polygons and Polynomials." In *The Geometric Vein: The Coxeter Festschrift.* New York: Springer-Verlag, 321–333, 1981.

First-Countable Space

A TOPOLOGICAL SPACE in which every point has a countable BASE for its neighborhood system.

First Curvature

see CURVATURE

First Derivative Test

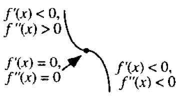
$f'(x) < 0$, $f''(x) > 0$
$f'(x) = 0$, $f''(x) = 0$
$f'(x) < 0$, $f''(x) < 0$

stationary point

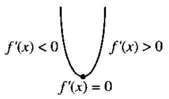

mininum

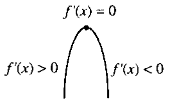

maximum

Suppose $f(x)$ is CONTINUOUS at a STATIONARY POINT x_0.

1. If $f'(x) > 0$ on an OPEN INTERVAL extending left from x_0 and $f'(x) < 0$ on an OPEN INTERVAL extending right from x_0, then f has a RELATIVE MAXIMUM (possibly a GLOBAL MAXIMUM) at x_0.
2. If $f'(x) < 0$ on an OPEN INTERVAL extending left from x_0 and $f'(x) > 0$ on an OPEN INTERVAL extending right from x_0, then f has a RELATIVE MINIMUM (possibly a GLOBAL MINIMUM) at x_0.
3. If $f'(x)$ has the same sign on an OPEN INTERVAL extending left from x_0 and on an OPEN INTERVAL extending right from x_0, then f does not have a RELATIVE EXTREMUM at x_0.

see also EXTREMUM, GLOBAL MAXIMUM, GLOBAL MINIMUM, INFLECTION POINT, MAXIMUM, MINIMUM, RELATIVE EXTREMUM, RELATIVE MAXIMUM, RELATIVE MINIMUM, SECOND DERIVATIVE TEST, STATIONARY POINT

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 14, 1972.

First Digit Law

see BENFORD'S LAW

First Digit Phenomenon

see BENFORD'S LAW

First Multiplier Theorem

Let D be a planar Abelian DIFFERENCE SET and t be any DIVISOR of n. Then t is a numerical multiplier of D, where a multiplier is defined as an automorphism α of G which takes D to a translation $g + D$ of itself for some $g \in G$. If α is of the form $\alpha : x \to tx$ for $t \in \mathbb{Z}$ relatively prime to the order of G, then α is called a numerical multiplier.

References
Gordon, D. M. "The Prime Power Conjecture is True for $n < 2,000,000$." *Electronic J. Combinatorics* **1**, R6, 1–7, 1994. `http://www.combinatorics.org/Volume_1/volume1.html#R6`.

Fischer's Baby Monster Group

see BABY MONSTER GROUP

Fischer Groups

The SPORADIC GROUPS Fi_{22}, Fi_{23}, and Fi'_{24}. These groups were discovered during the investigation of 3-TRANSPOSITION GROUPS.

see also SPORADIC GROUP

References
Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas/F22.html`, `F23.html`, and `F24.html`.

Fish Bladder

see LENS

Fisher-Behrens Problem

The determination of a test for the equality of MEANS for two NORMAL DISTRIBUTIONS with different VARIANCES given samples from each. There exists an exact test which, however, does not give a unique answer because it does not use all the data. There also exist approximate tests which do not use all the data.

see also NORMAL DISTRIBUTION

References

Fisher, R. A. "The Fiducial Argument in Statistical Inference." *Ann. Eugenics* **6**, 391–398, 1935.

Kenney, J. F. and Keeping, E. S. "The Behrens-Fisher Test." §9.8 in *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, pp. 257–260 and 261–264, 1951.

Sukhatme, P. V. "On Fisher and Behrens' Test of Significance of the Difference in Means of Two Normal Samples." *Sankhya* **4**, 39, 1938.

Fisher's Block Design Inequality

A balanced incomplete BLOCK DESIGN (v, k, λ, r, b) exists only or $b \geq v$ (or, equivalently, $r \geq k$).

see also BRUCK-RYSER-CHOWLA THEOREM

References

Dinitz, J. H. and Stinson, D. R. "A Brief Introduction to Design Theory." Ch. 1 in *Contemporary Design Theory: A Collection of Surveys* (Ed. J. H. Dinitz and D. R. Stinson). New York: Wiley, pp. 1–12, 1992.

Fisher's Estimator Inequality

Given T an UNBIASED ESTIMATOR of θ so that $\langle T\rangle = \theta$. Then

$$\operatorname{var}(T) \geq \frac{1}{N\int_{-\infty}^{\infty}\left[\frac{\partial(\ln f)}{\partial\theta}\right]^2 f\,dx},$$

where var is the VARIANCE.

Fisher's Exact Test

A STATISTICAL TEST used to determine if there are nonrandom associations between two CATEGORICAL VARIABLES. Let there exist two such variables X and Y, with m and n observed states, respectively. Now form an $n \times m$ MATRIX in which the entries a_{ij} represent the number of observations in which $x = i$ and $y = j$. Calculate the row and column sums R_i and C_j, respectively, and the total sum

$$N = \sum_i R_i = \sum_j C_j$$

of the MATRIX. Then calculate the conditional LIKELIHOOD (P-VALUE) of getting the actual matrix given the particular row and column sums, given by

$$P_{\text{crit}} = \frac{(R_1!R_2!\cdots R_m!)(C_1!C_2!\cdots C_n!)}{N!\prod_{i,j} a_{ij}!}$$

(which is a HYPERGEOMETRIC DISTRIBUTION). Now find all possible MATRICES of NONNEGATIVE INTEGERS consistent with the row and column sums R_i and C_j. For each one, calculate the associated P-VALUE using (0) (where the sum of these probabilities must be 1). Then the P-VALUE of the test is given by the sum of all P-VALUES which are $\leq P_{\text{crit}}$.

The test is most commonly applied to a 2×2 MATRICES, and is computationally unwieldy for large m or n.

As an example application of the test, let X be a journal, say either *Mathematics Magazine* or *Science*, and let Y be the number of articles on the topics of mathematics and biology appearing in a given issue of one of these journals. If *Mathematics Magazine* has five articles on math and one on biology, and *Science* has none on math and four on biology, then the relevant matrix would be

	Math. Mag.	*Science*	
math	5	0	$R_1 = 5$
biology	1	4	$R_2 = 5$
	$C_1 = 6$	$C_2 = 4$	$N = 10.$

Computing P_{crit} gives

$$P_{\text{crit}} = \frac{5!^2 6!4!}{10!(5!0!1!4!)} = 0.0238,$$

and the other possible matrices and their Ps are

$$\begin{bmatrix}4 & 1\\ 2 & 3\end{bmatrix} \quad P = 0.2381$$

$$\begin{bmatrix}3 & 2\\ 3 & 2\end{bmatrix} \quad P = 0.4762$$

$$\begin{bmatrix}2 & 3\\ 4 & 1\end{bmatrix} \quad P = 0.2381$$

$$\begin{bmatrix}1 & 4\\ 5 & 0\end{bmatrix} \quad P = 0.0238,$$

which indeed sum to 1, as required. The sum of P-values less than or equal to $P_{\text{crit}} = 0.0238$ is then 0.0476 which, because it is less than 0.05, is SIGNIFICANT. Therefore, in this case, there would be a statistically significant association between the journal and type of article appearing.

Fisher Index

The statistical INDEX

$$P_B \equiv \sqrt{P_L P_P},$$

where P_L is LASPEYRES' INDEX and P_P is PAASCHE'S INDEX.

see also INDEX

References

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, p. 66, 1962.

Fisher Kurtosis

$$\gamma_2 \equiv b_2 \equiv \frac{\mu_4}{{\mu_2}^2} - 3 = \frac{\mu_4}{\sigma^4} - 3,$$

where μ_i is the ith MOMENT about the MEAN and $\sigma = \sqrt{\mu_2}$ is the STANDARD DEVIATION.

see also FISHER SKEWNESS, KURTOSIS, PEARSON KURTOSIS

Fisher Sign Test

A robust nonparametric test which is an alternative to the PAIRED t-TEST. This test makes the basic assumption that there is information only in the signs of the differences between paired observations, not in their sizes. Take the paired observations, calculate the differences, and count the number of +s n_+ and −s n_-, where

$$N \equiv n_+ + n_-$$

is the sample size. Calculate the BINOMIAL COEFFICIENT

$$B \equiv \binom{N}{n_+}.$$

Then $B/2^N$ gives the probability of getting exactly this many +s and −s if POSITIVE and NEGATIVE values are equally likely. Finally, to obtain the P-VALUE for the test, sum all the COEFFICIENTS that are $\leq B$ and divide by 2^N.

see also HYPOTHESIS TESTING

Fisher Skewness

$$\gamma_1 = \frac{\mu_3}{{\mu_2}^{3/2}} = \frac{\mu_3}{\sigma^3},$$

where μ_i is the i MOMENT about the MEAN, and $\sigma = \sqrt{\mu_2}$ is the STANDARD DEVIATION.

see also FISHER KURTOSIS, MOMENT, SKEWNESS, STANDARD DEVIATION

Fisher's Theorem

Let A be a sum of squares of n independent normal standardized variates x_i, and suppose $A = B + C$ where B is a quadratic form in the x_i, distributed as CHI-SQUARED with h DEGREES OF FREEDOM. Then C is distributed as χ^2 with $n - h$ DEGREES OF FREEDOM and is independent of B. The converse of this theorem is known as COCHRAN'S THEOREM.

see also CHI-SQUARED DISTRIBUTION, COCHRAN'S THEOREM

Fisher-Tippett Distribution

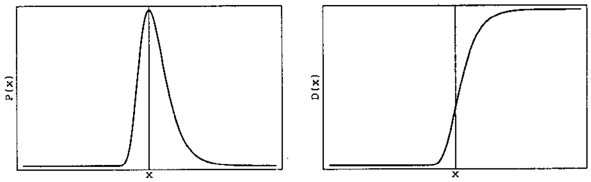

Also called the EXTREME VALUE DISTRIBUTION and LOG-WEIBULL DISTRIBUTION. It is the limiting distribution for the smallest or largest values in a large sample drawn from a variety of distributions.

$$P(x) = \frac{e^{(a-x)/b - e^{(a-x)/b}}}{b} \tag{1}$$

$$D(x) = e^{-e^{(a-x)/b}}. \tag{2}$$

These can be computed directly be defining

$$z \equiv \exp\left(\frac{a-x}{b}\right) \tag{3}$$

$$x = a - b\ln z \tag{4}$$

$$dz = -\frac{1}{b}\exp\left(\frac{a-x}{b}\right)dx. \tag{5}$$

Then the MOMENTS are

$$\begin{aligned}\mu_n &\equiv \int_{-\infty}^{\infty} x^n P(x)\,dx \\ &= \frac{1}{b}\int_{-\infty}^{\infty} x^n \exp\left(\frac{a}{x-b}\right)\exp[-e^{(a-x)/b}]\,dx \\ &= -\int_{\infty}^{0}(a-b\ln z)^n e^{-z}\,dz \\ &= \int_0^{\infty}(a-b\ln z)^n e^{-z}\,dz \\ &= \sum_{k=0}^{n}\binom{n}{k}(-1)^k a^{n-k} b^k \int_0^{\infty}(\ln z)^k e^{-z}\,dz \\ &= \sum_{k=0}^{n}\binom{n}{k} a^{n-k} b^k I(k), \end{aligned} \tag{6}$$

where $I(k)$ are EULER-MASCHERONI INTEGRALS. Plugging in the EULER-MASCHERONI INTEGRALS $I(k)$ gives

$$\mu_0 = 1 \tag{7}$$

$$\mu_1 = a + b\gamma \tag{8}$$

$$\mu_2 = a^2 + 2ab\gamma + b^2(\gamma^2 + \tfrac{1}{6}\pi^2) \tag{9}$$

$$\begin{aligned}\mu_3 &= a^3 + 3a^2b\gamma + 3ab^2(\gamma^2 + \tfrac{1}{6}\pi^2) \\ &\quad + b^3[\gamma^3 + \tfrac{1}{2}\gamma\pi^2 + 2\zeta(3)]\end{aligned} \tag{10}$$

$$\begin{aligned}\mu_4 &= a^4 + 4a^3b\gamma + 6a^2b^2(\gamma^2 + \tfrac{1}{6}\pi^2) \\ &\quad + 4ab^3[\gamma^3 + \tfrac{1}{2}\gamma\pi^2 + 2\zeta(3)] \\ &\quad + b^4[\gamma^4 + \gamma^2\pi^2 + \tfrac{3}{20}\pi^4 + 8\gamma\zeta(3)],\end{aligned} \tag{11}$$

where γ is the EULER-MASCHERONI CONSTANT and $\zeta(3)$ is APÉRY'S CONSTANT. The MEAN, VARIANCE, SKEWNESS, and KURTOSIS are therefore

$$\mu = a + b\gamma \tag{12}$$

$$\sigma^2 = \mu_2 - {\mu_1}^2 = \tfrac{1}{6}\pi^2 b^2 \tag{13}$$

$$\begin{aligned}\gamma_1 &= \frac{\mu_3}{\sigma^3} \\ &= \frac{6\sqrt{6}}{b^3\pi^3}\{a^3 + 3a^2 b\gamma + 3ab^2(\gamma^2 + \tfrac{1}{6}\pi^2) \\ &\quad + b^3[\gamma^3 + \tfrac{1}{2}\gamma\pi^2 + 2\zeta(3)]\}\end{aligned} \tag{14}$$

$$\begin{aligned}\gamma_2 &= \frac{\mu_4}{\sigma^4} - 3 \\ &= \frac{36}{b^4\pi^4}\{a^4 + 4a^3 b\gamma + a^2 b^2(6\gamma^2 + \pi^2) \\ &\quad + 4ab^3[\gamma^3 + \tfrac{1}{2}\gamma\pi^2 + 2\zeta(3)] \\ &\quad + b^4[\gamma^4 + \gamma^2\pi^2 + \tfrac{3}{20}\pi^4 + 8\gamma\zeta(3)]\}.\end{aligned} \tag{15}$$

The CHARACTERISTIC FUNCTION is

$$\phi(t) = \Gamma(1 - i\beta t)e^{i\alpha t}, \tag{16}$$

where $\Gamma(z)$ is the GAMMA FUNCTION. The special case of the Fisher-Tippett distribution with $a = 0$, $b = 1$ is called GUMBEL'S DISTRIBUTION.

see also EULER-MASCHERONI INTEGRALS, GUMBEL'S DISTRIBUTION

Fisher's z-Distribution

$$g(z) = \frac{2{n_1}^{n_1/2}{n_2}^{n_2/2}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \frac{e^{n_1 z}}{(n_1 e^{2z} + n_2)^{(n_1+n_1)/2}} \tag{1}$$

(Kenney and Keeping 1951). This general distribution includes the CHI-SQUARED DISTRIBUTION and STUDENT'S t-DISTRIBUTION as special cases. Let u^2 and v^2 be INDEPENDENT UNBIASED ESTIMATORS of the VARIANCE of a NORMALLY DISTRIBUTED variate. Define

$$z \equiv \ln\left(\frac{u}{v}\right) = \tfrac{1}{2}\ln\left(\frac{u^2}{v^2}\right). \tag{2}$$

Then let

$$F \equiv \frac{u^2}{v^2} = \frac{\frac{N{s_1}^2}{n_1}}{\frac{N{s_2}^2}{n_2}} \tag{3}$$

so that $n_1 F/n_2$ is a ratio of CHI-SQUARED variates

$$\frac{n_1 F}{n_2} = \frac{\chi^2(n_1)}{\chi^2(n_2)}, \tag{4}$$

which makes it a ratio of GAMMA DISTRIBUTION variates, which is itself a BETA PRIME DISTRIBUTION variate,

$$\frac{\gamma\left(\frac{n_1}{2}\right)}{\gamma\left(\frac{n_2}{2}\right)} = \beta'\left(\tfrac{n_1}{2}, \tfrac{n_2}{2}\right), \tag{5}$$

giving

$$f(F) = \frac{\left(\frac{n_1 F}{n_2}\right)^{n_1/2-1}\left(1 + \frac{n_1 F}{n_2}\right)^{-(n_1+n_2)/2}\frac{n_1}{n_2}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)}. \tag{6}$$

The MEAN is

$$\langle F\rangle = \frac{n_2}{n_2 - 2}, \tag{7}$$

and the MODE is

$$\frac{n_2}{n_2 + 2}\frac{n_1 - 2}{n_1}. \tag{8}$$

see also BETA DISTRIBUTION, BETA PRIME DISTRIBUTION, CHI-SQUARED DISTRIBUTION, GAMMA DISTRIBUTION, NORMAL DISTRIBUTION, STUDENT'S t-DISTRIBUTION

References

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, pp. 180–181, 1951.

Fisher's z'-Transformation

Let r be the CORRELATION COEFFICIENT. Then defining

$$z' \equiv \tanh^{-1} r \tag{1}$$

$$\zeta \equiv \tanh^{-1} \rho, \tag{2}$$

gives

$$\sigma_{z'} = (N - 3)^{-1/2} \tag{3}$$

$$\operatorname{var}(z') = \frac{1}{n} + \frac{4 - \rho^2}{2n^2} + \ldots \tag{4}$$

$$\gamma_1 = \frac{\rho\left|\rho^2 - \frac{9}{16}\right|}{n^{3/2}} \tag{5}$$

$$\gamma_2 = \frac{32 - 3\rho^4}{16N}, \tag{6}$$

where $n \equiv N - 1$.

see also CORRELATION COEFFICIENT

Fitting Subgroup

The unique smallest NORMAL NILPOTENT SUBGROUP of H, denoted $F(H)$. The generalized fitting subgroup is defined by $F^*(H) = F(H)E(H)$, where $E(H)$ is the commuting product of all components of H, and F is the fitting subgroup of H.

Five Cubes

see CUBE 5-COMPOUND

Five Disks Problem

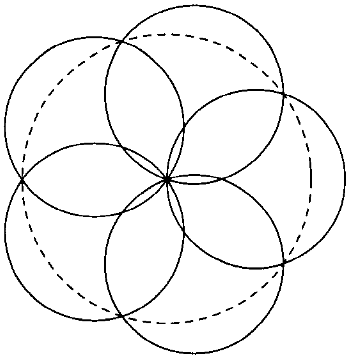

Given five *equal* DISKS placed *symmetrically* about a given center, what is the smallest RADIUS r for which the RADIUS of the circular AREA covered by the five disks is 1? The answer is $r = \phi - 1 = 1/\phi = 0.6180340\ldots$, where ϕ is the GOLDEN RATIO, and the centers c_i of the disks $i = 1, \ldots, 5$ are located at

$$c_i = \begin{bmatrix} \frac{1}{\phi}\cos\left(\frac{2\pi i}{5}\right) \\ \frac{1}{\phi}\sin\left(\frac{2\pi i}{5}\right) \end{bmatrix}.$$

The GOLDEN RATIO enters here through its connection with the regular PENTAGON. If the requirement that the disks be symmetrically placed is dropped (the general DISK COVERING PROBLEM), then the RADIUS for $n = 5$ disks can be reduced slightly to 0.609383... (Neville 1915).

see also ARC, DISK COVERING PROBLEM, FLOWER OF LIFE, SEED OF LIFE

References
Ball, W. W. R. and Coxeter, H. S. M. "The Five-Disc Problem." In *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 97–99, 1987.
Neville, E. H. "On the Solution of Numerical Functional Equations, Illustrated by an Account of a Popular Puzzle and of its Solution." *Proc. London Math. Soc.* **14**, 308–326, 1915.

Five Tetrahedra Compound

see TETRAHEDRON 5-COMPOUND

Fixed

When referring to a planar object, "fixed" means that the object is regarded as fixed in the plane so that it may not be picked up and flipped. As a result, MIRROR IMAGES are not necessarily equivalent for fixed objects.

see also FREE, MIRROR IMAGE

Fixed Element

see FIXED POINT (MAP)

Fixed Point

A point which does not change upon application of a MAP, system of DIFFERENTIAL EQUATIONS, etc.

see also FIXED POINT (DIFFERENTIAL EQUATIONS), FIXED POINT (MAP), FIXED POINT THEOREM

References
Shashkin, Yu. A. *Fixed Points.* Providence, RI: Amer. Math. Soc., 1991.

Fixed Point (Differential Equations)

Points of an AUTONOMOUS system of ordinary differential equations at which

$$\begin{cases} \frac{dx_1}{dt} = f_1(x_1, \ldots, x_n) = 0 \\ \vdots \\ \frac{dx_n}{dt} = f_n(x_1, \ldots, x_n) = 0. \end{cases}$$

If a variable is slightly displaced from a FIXED POINT, it may (1) move back to the fixed point ("asymptotically stable" or "superstable"), (2) move away ("unstable"), or (3) move in a neighborhood of the fixed point but not approach it ("stable" but not "asymptotically stable"). Fixed points are also called CRITICAL POINTS or EQUILIBRIUM POINTS. If a variable starts at a point that is not a CRITICAL POINT, it cannot reach a critical point in a finite amount of time. Also, a trajectory passing through at least one point that is not a CRITICAL POINT cannot cross itself unless it is a CLOSED CURVE, in which case it corresponds to a periodic solution.

A fixed point can be classified into one of several classes using LINEAR STABILITY analysis and the resulting STABILITY MATRIX.

see also ELLIPTIC FIXED POINT (DIFFERENTIAL EQUATIONS), HYPERBOLIC FIXED POINT (DIFFERENTIAL EQUATIONS), STABLE IMPROPER NODE, STABLE NODE, STABLE SPIRAL POINT, STABLE STAR, UNSTABLE IMPROPER NODE, UNSTABLE NODE, UNSTABLE SPIRAL POINT, UNSTABLE STAR

Fixed Point (Map)

A point x^* which is mapped to itself under a MAP G, so that $x^* = G(x^*)$. Such points are sometimes also called INVARIANT POINTS, or FIXED ELEMENTS (Woods 1961). Stable fixed points are called elliptical. Unstable fixed points, corresponding to an intersection of a stable and unstable invariant MANIFOLD, are called HYPERBOLIC (or SADDLE). Points may also be called asymptotically stable (a.k.a. superstable).

see also CRITICAL POINT, INVOLUNTARY

References
Shashkin, Yu. A. *Fixed Points.* Providence, RI: Amer. Math. Soc., 1991.
Woods, F. S. *Higher Geometry: An Introduction to Advanced Methods in Analytic Geometry.* New York: Dover, p. 14, 1961.

Fixed Point Theorem

If g is a continuous function $g(x) \in [a,b]$ FOR ALL $x \in [a,b]$, then g has a FIXED POINT in $[a,b]$. This can be proven by noting that

$$g(a) \geq a \qquad g(b) \leq b$$

$$g(a) - a \geq 0 \qquad g(b) - b \leq 0.$$

Since g is continuous, the INTERMEDIATE VALUE THEOREM guarantees that there exists a $c \in [a,b]$ such that

$$g(c) - c = 0,$$

so there must exist a c such that

$$g(c) = c,$$

so there must exist a FIXED POINT $\in [a,b]$.

see also BANACH FIXED POINT THEOREM, BROUWER FIXED POINT THEOREM, KAKUTANI'S FIXED POINT THEOREM, LEFSHETZ FIXED POINT FORMULA, LEFSHETZ TRACE FORMULA, POINCARÉ-BIRKHOFF FIXED POINT THEOREM, SCHAUDER FIXED POINT THEOREM

Fixed Point (Transformation)

see FIXED POINT (MAP)

Flag

A collection of FACES of an n-D POLYTOPE or simplicial COMPLEX, one of each DIMENSION $0, 1, \ldots, n-1$, which all have a common nonempty INTERSECTION. In normal 3-D, the flag consists of a half-plane, its bounding RAY, and the RAY's endpoint.

Flag Manifold

For any SEQUENCE of INTEGERS $0 < n_1 < \ldots < n_k$, there is a flag manifold of type $(n_1, \ldots, n_k)$ which is the collection of ordered pairs of vector SUBSPACES of $\mathbb{R}^{n_k}$ $(V_1, \ldots, V_k)$ with $\dim(V_i) = n_i$ and V_i a SUBSPACE of V_{i+1}. There are also COMPLEX flag manifolds with COMPLEX subspaces of $\mathbb{C}^{n_k}$ instead of REAL SUBSPACES of a REAL n_k-space. These flag manifolds admit the structure of MANIFOLDS in a natural way and are used in the theory of LIE GROUPS.

see also GRASSMANN MANIFOLD

References

Lu, J.-H. and Weinstein, A. "Poisson Lie Groups, Dressing Transformations, and the Bruhat Decomposition." *J. Diff. Geom.* **31**, 501–526, 1990.

Flat

A set in $\mathbb{R}^d$ formed by translating an affine subspace or by the intersection of a set of HYPERPLANES.

Flat Norm

The flat norm on a CURRENT is defined by

$$\mathcal{F}(S) = \int \{\text{Area } T + \text{vol } R : S - T = \partial R\},$$

where ∂R is the boundary of R.

see also COMPACTNESS THEOREM, CURRENT

References

Morgan, F. "What Is a Surface?" *Amer. Math. Monthly* **103**, 369–376, 1996.

Flat Space Theorem

If it is possible to transform a coordinate system to a form where the metric elements $g_{\mu\nu}$ are constants independent of x^μ, then the space is flat.

Flat Surface

A REGULAR SURFACE and special class of MINIMAL SURFACE for the GAUSSIAN CURVATURE vanishes everywhere. A TANGENT DEVELOPABLE, GENERALIZED CONE, and GENERALIZED CYLINDER are all flat surfaces.

see also MINIMAL SURFACE

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 280, 1993.

Flattening

The flattening of a SPHEROID (also called OBLATENESS) is denoted ε or f. It is defined as

$$\varepsilon \equiv \begin{cases} \frac{a-c}{a} = 1 - \frac{c}{a} & \text{oblate} \\ \frac{c-a}{a} = \frac{c}{a} - 1 & \text{prolate,} \end{cases}$$

where c is the polar RADIUS and a is the equatorial RADIUS.

see also ECCENTRICITY, ELLIPSOID, OBLATE SPHEROID, PROLATE SPHEROID, SPHEROID

Flemish Knot

see FIGURE-OF-EIGHT KNOT

Fletcher Point

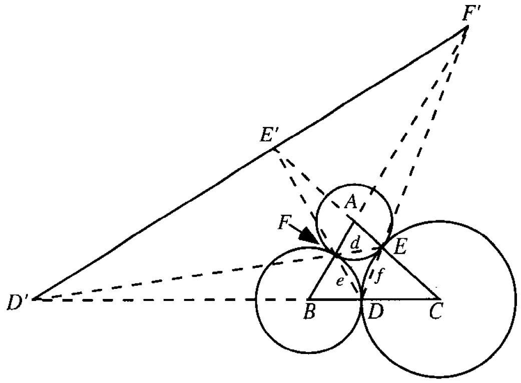

The intersection of the GERGONNE LINE and the SODDY LINE. It has TRILINEAR COORDINATES given by

$$Fl = I - \frac{1}{3}\left(\frac{1}{d} + \frac{1}{e} + \frac{1}{f}\right) Ge,$$

where I is the INCENTER, Ge the GERGONNE POINT, and d, e, and f are the lengths of the sides of the CONTACT TRIANGLE ΔDEF.

see also CONTACT TRIANGLE, GERGONNE LINE, GERGONNE POINT, SODDY LINE

References

Oldknow, A. "The Euler-Gergonne-Soddy Triangle of a Triangle." *Amer. Math. Monthly* **103**, 319–329, 1996.

Flexible Polyhedron

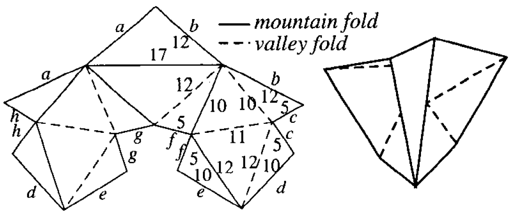

The RIGIDITY THEOREM states that if the faces of a *convex* POLYHEDRON are made of metal plates and the EDGES are replaced by hinges, the POLYHEDRON would be RIGID. The theorem was stated by Cauchy (1813), although a mistake in this paper went unnoticed for more than 50 years. Concave polyhedra need not be RIGID, and such nonrigid polyhedra are called flexible polyhedra. Connelly (1978) found the first example of a reflexible polyhedron, consisting of 18 triangular faces. A flexible polyhedron with only 14 triangular faces and 9 vertices (shown above), believed to be the simplest possible composed of only triangles, was subsequently found by Steffen (Mackenzie 1998). There also exists a six-vertex eight-face flexible polyhedron (Wunderlich and Schwabe 1986, Cromwell 1997).

Connelly *et al.* (1997) proved that a flexible polyhedron must keep its VOLUME constant (Mackenzie 1998).

see also POLYHEDRON, QUADRICORN, RIGID, RIGIDITY THEOREM

References

Cauchy, A. L. "Sur les polygons et le polyhéders." *XVIe Cahier* **IX**, 87–89, 1813.

Connelly, R. "A Flexible Sphere." *Math. Intel.* **1**, 130–131, 1978.

Connelly, R.; Sabitov, I.; and Walz, A. "The Bellows Conjecture." *Contrib. Algebra Geom.* **38**, 1–10, 1997.

Cromwell, P. R. *Polyhedra.* New York: Cambridge University Press, 1997.

Mackenzie, D. "Polyhedra Can Bend But Not Breathe." *Science* **279**, 1637, 1998.

Wunderlich, W. and Schwabe, C. "Eine Familie von geschlossen gleichflachigen Polyhedern, die fast beweglich sind." *Elem. Math.* **41**, 88–98, 1986.

Flexagon

An object created by FOLDING a piece of paper along certain lines to form loops. The number of states possible in an n-FLEXAGON is a CATALAN NUMBER. By manipulating the folds, it is possible to hide and reveal different faces.

see also FLEXATUBE, FOLDING, HEXAFLEXAGON, TETRAFLEXAGON

References

Crampin, J. "On Note 2449." *Math. Gazette* **41**, 55–56, 1957.

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 205–207, 1989.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 62–84, 1979.

Gardner, M. "Hexaflexagons." Ch. 1 in *The Scientific American Book of Mathematical Puzzles & Diversions.* New York: Simon and Schuster, 1959.

Gardner, M. Ch. 2 in *The Second Scientific American Book of Mathematical Puzzles & Diversions: A New Selection.* New York: Simon and Schuster, pp. 24–31, 1961.

Maunsell, F. G. "The Flexagon and the Hexaflexagon." *Math. Gazette* **38**, 213–214, 1954.

Oakley, C. O. and Wisner, R. J. "Flexagons." *Amer. Math. Monthly* **64**, 143–154, 1957.

Wheeler, R. F. "The Flexagon Family." *Math. Gaz.* **42**, 1–6, 1958.

Flexatube

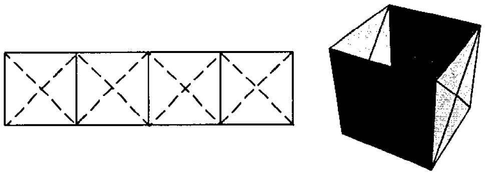

A FLEXAGON-like structure created by connecting the ends of a strip of four squares after folding along 45° diagonals. Using a number of folding movements, it is possible to flip the flexatube inside out so that the faces originally facing inward face outward. Gardner (1961) illustrated one possible solution, and Steinhaus (1983) gives a second.

see also FLEXAGON, HEXAFLEXAGON, TETRAFLEXAGON

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 205, 1989.

Gardner, M. *The Second Scientific American Book of Mathematical Puzzles & Diversions: A New Selection.* New York: Simon and Schuster, pp. 29–31, 1961.

Steinhaus, H. *Mathematical Snapshots, 3rd American ed.* New York: Oxford University Press, pp. 177–181 and 190, 1983.

Flip Bifurcation

Let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a one-parameter family of C^3 maps satisfying

$$f(0,0) = 0$$

$$\left[\frac{\partial f}{\partial x}\right]_{\mu=0,x=0} = -1$$

$$\left[\frac{\partial^2 f}{\partial x^2}\right]_{\mu=0,x=0} < 0$$

$$\left[\frac{\partial^3 f}{\partial x^3}\right]_{\mu=0,x=0} < 0.$$

Then there are intervals $(\mu_1, 0)$, $(0, \mu_2)$, and $\epsilon > 0$ such that

1. If $\mu \in (0, \mu_2)$, then $f_\mu(x)$ has one unstable fixed point and one stable orbit of period two for $x \in (-\epsilon, \epsilon)$, and
2. If $\mu \in (\mu_1, 0)$, then $f_\mu(x)$ has a single stable fixed point for $x \in (-\epsilon, \epsilon)$.

This type of BIFURCATION is known as a flip bifurcation. An example of an equation displaying a flip bifurcation is

$$f(x) = \mu - x - x^2.$$

see also BIFURCATION

References
Rasband, S. N. *Chaotic Dynamics of Nonlinear Systems.* New York: Wiley, pp. 27–30, 1990.

Floor Function

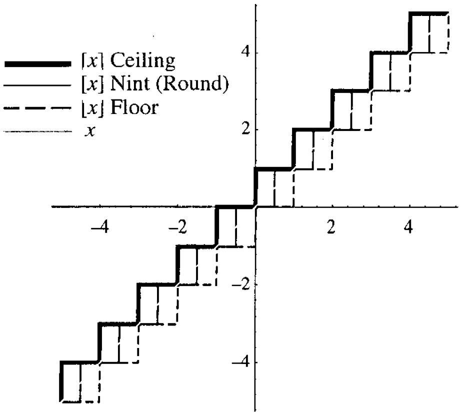

The function $\lfloor x \rfloor$ is the largest INTEGER $\leq x$, shown as the dashed curve in the above plot, and also called the GREATEST INTEGER FUNCTION. In many computer languages, the floor function is called the INTEGER PART function and is denoted `int(x)`. The name and symbol for the floor function were coined by K. E. Iverson (Graham *et al.* 1990).

Unfortunately, in many older and current works (e.g., Shanks 1993, Ribenboim 1996), the symbol $[x]$ is used instead of $\lfloor x \rfloor$. Because of the elegant symmetry of the floor function and CEILING FUNCTION symbols $\lfloor x \rfloor$ and $\lceil x \rceil$, and because $[x]$ is such a useful symbol when interpreted as an IVERSON BRACKET, the use of $[x]$ to denote the floor function should be deprecated. In this work, the symbol $[x]$ is used to denote the nearest integer NINT function since it naturally falls between the $\lfloor x \rfloor$ and $\lceil x \rceil$ symbols.

see also CEILING FUNCTION, FRACTIONAL PART, INT, IVERSON BRACKET, NINT

References
Graham, R. L.; Knuth, D. E.; and Patashnik, O. "Integer Functions." Ch. 3 in *Concrete Mathematics: A Foundation for Computer Science.* Reading, MA: Addison-Wesley, pp. 67–101, 1990.
Iverson, K. E. *A Programming Language.* New York: Wiley, p. 12, 1962.
Ribenboim, P. *The New Book of Prime Number Records.* New York: Springer-Verlag, pp. 180–182, 1996.
Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 14, 1993.
Spanier, J. and Oldham, K. B. "The Integer-Value Int(x) and Fractional-Value frac(x) Functions." Ch. 9 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 71–78, 1987.

Floquet Analysis

Given a system of periodic ORDINARY DIFFERENTIAL EQUATIONS of the form

$$\frac{d}{dt}\begin{bmatrix} x \\ y \\ v_x \\ v_y \end{bmatrix} = -\begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ \Phi_{xx} & \Phi_{yy} & 0 & 0 \\ \Phi_{xy} & \Phi_{yy} & 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ v_x \\ v_y \end{bmatrix}, \qquad (1)$$

the solution can be written as a linear combination of functions of the form

$$\begin{bmatrix} x(t) \\ y(t) \\ v_x(t) \\ v_y(t) \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \\ v_{x0} \\ v_{y0} \end{bmatrix} e^{\mu t} P_\mu(t), \qquad (2)$$

where $P_\mu(t)$ is a function periodic with the same period T as the equations themselves. Given an ORDINARY DIFFERENTIAL EQUATION of the form

$$\ddot{x} + g(t)x = 0, \qquad (3)$$

where $g(t)$ is periodic with period T, the ODE has a pair of independent solutions given by the REAL and IMAGINARY PARTS of

$$x(t) = w(t)e^{i\psi(t)} \qquad (4)$$

$$\dot{x} = (\dot{w} + iw\dot{\psi})e^{i\psi} \qquad (5)$$

$$\begin{aligned} \ddot{x} &= [\ddot{w} + i\dot{w}\dot{\psi} + i(\dot{w}\dot{\psi} + w\ddot{\psi} + iw\dot{\psi}^2)]e^{i\psi} \\ &= [(\ddot{w} - w\dot{\psi}^2) + i(2\dot{w}\dot{\psi} + w\ddot{\psi})]e^{i\psi}. \qquad (6) \end{aligned}$$

Plugging these into (3) gives

$$\ddot{w} + 2i\dot{w}\dot{\psi} + w(g + i\ddot{\psi} - \dot{\psi}^2) = 0, \qquad (7)$$

so the REAL and IMAGINARY PARTS are

$$\ddot{w} + w(g - \dot{\psi}^2) = 0 \qquad (8)$$

$$2\dot{w}\dot{\psi} + w\ddot{\psi} = 0. \qquad (9)$$

From (9),

$$\begin{aligned} \frac{2\dot{w}}{w} + \frac{\ddot{\psi}}{\dot{\psi}} &= 2\frac{d}{dt}(\ln w) + \frac{d}{dt}[\ln(\dot{\psi})] \\ &= \frac{d}{dt}\ln(\dot{\psi}w^2) = 0. \qquad (10) \end{aligned}$$

Integrating gives

$$\dot\psi = \frac{C}{w^2}, \tag{11}$$

where C is a constant which must equal 1, so ψ is given by

$$\psi = \int_{t_0}^{t} \frac{dt}{w^2}. \tag{12}$$

The REAL solution is then

$$x(t) = w(t)\cos[\psi(t)], \tag{13}$$

so

$$\begin{aligned}\dot x &= \dot w\cos\psi - w\dot\psi\sin\psi = \dot w\frac{x}{w} - w\dot\psi\sin\psi \\ &= \dot w\frac{x}{w} - w\frac{1}{w^2}\sin\psi = \dot w\frac{x}{w} - \frac{1}{w}\sin\psi \end{aligned} \tag{14}$$

and

$$\begin{aligned} 1 &= \cos^2\psi + \sin^2\psi = x^2w^{-2} + \left[w\left(\dot w\frac{x}{w} - \dot x\right)\right]^2 \\ &= x^2w^{-2} + (\dot w x - w\dot x)^2 \equiv I(x,\dot x,t), \end{aligned} \tag{15}$$

which is an integral of motion. Therefore, although $w(t)$ is not explicitly known, an integral I always exists. Plugging (10) into (8) gives

$$\ddot w + g(t)w - \frac{1}{w^3} = 0, \tag{16}$$

which, however, is not any easier to solve than (3).

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 727, 1972.

Binney, J. and Tremaine, S. *Galactic Dynamics.* Princeton, NJ: Princeton University Press, p. 175, 1987.

Lichtenberg, A. and Lieberman, M. *Regular and Stochastic Motion.* New York: Springer-Verlag, p. 32, 1983.

Margenau, H. and Murphy, G. M. *The Mathematics of Physics and Chemistry, 2 vols.* Princeton, NJ: Van Nostrand, 1956–64.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 556–557, 1953.

Floquet's Theorem

see FLOQUET ANALYSIS

Flow

An ACTION with $G = \mathbb{R}$. Flows are generated by VECTOR FIELDS and vice versa.

see also ACTION, AMBROSE-KAKUTANI THEOREM, ANOSOV FLOW, AXIOM A FLOW, CASCADE, GEODESIC FLOW, SEMIFLOW

Flow Line

A flow line for a map on a VECTOR FIELD $\mathbf{F}$ is a path $\sigma(t)$ such that $\sigma'(t) = \mathbf{F}(\sigma(t))$.

Flower

see DAISY, FLOWER OF LIFE, ROSE

Flower of Life

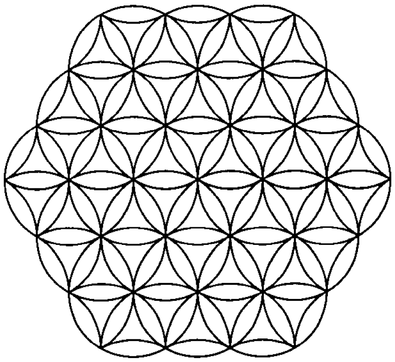

One of the beautiful arrangements of CIRCLES found at the Temple of Osiris at Abydos, Egypt (Rawles 1997). The CIRCLES are placed with six-fold symmetry, forming a mesmerizing pattern of CIRCLES and LENSES.

see also FIVE DISKS PROBLEM, REULEAUX TRIANGLE, SEED OF LIFE, VENN DIAGRAM

References

Rawles, B. *Sacred Geometry Design Sourcebook: Universal Dimensional Patterns.* Nevada City, CA: Elysian Pub., p. 15, 1997.

Wein, J. "The Flower of Life." `http://www2.cruzio.com/~flower`.

Weisstein, E. W. "Flower of Life." `http://www.astro.virginia.edu/~eww6n/math/notebooks/FlowerOfLife.m`.

Flowsnake

see PEANO-GOSPER CURVE

Flowsnake Fractal

see GOSPER ISLAND

Floyd's Algorithm

An algorithm for finding the shortest path between two VERTICES.

see also DIJKSTRA'S ALGORITHM

Fluent

Newton's term for a variable in his method of FLUXIONS (differential calculus).

Fluxion

The term for DERIVATIVE in Newton's CALCULUS.

Flype

A 180° rotation of a TANGLE.

see also FLYPING CONJECTURE, TANGLE

Flyping Conjecture

Also called the TAIT FLYPING CONJECTURE. Given two reduced alternating projections of the same knot, they are equivalent on the SPHERE IFF they are related by a series of FLYPES. It was proved by Menasco and Thistlethwaite (1991). It allows all possible REDUCED alternating projections of a given ALTERNATING KNOT to be drawn.

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 164–165, 1994.

Menasco, W. and Thistlethwaite, M. "The Tait Flyping Conjecture." *Bull. Amer. Math. Soc.* **25**, 403–412, 1991.

Stewart, I. *The Problems of Mathematics, 2nd ed.* Oxford, England: Oxford University Press, pp. 284–285, 1987.

Focus

A point related to the construction and properties of CONIC SECTIONS.

see also ELLIPSE, ELLIPSOID, HYPERBOLA, HYPERBOLOID, PARABOLA, PARABOLOID, REFLECTION PROPERTY

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 141–144, 1967.

Fold Bifurcation

Let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a one-parameter family of C^2 MAP satisfying

$$\begin{gathered} f(0,0) = 0 \\ \left[\frac{\partial f}{\partial x}\right]_{\mu=0,x=0} = 1 \\ \left[\frac{\partial^2 f}{\partial x^2}\right]_{\mu=0,x=0} > 0 \\ \left[\frac{\partial f}{\partial \mu}\right]_{\mu=0,x=0} > 0, \end{gathered}$$

then there exist intervals $(\mu_1, 0)$, $(0, \mu_2)$ and $\epsilon > 0$ such that

1. If $\mu \in (\mu_1, 0)$, then $f_\mu(x)$ has two fixed points in $(-\epsilon, \epsilon)$ with the positive one being unstable and the negative one stable, and
2. If $\mu \in (0, \mu_2)$, then $f_\mu(x)$ has no fixed points in $(-\epsilon, \epsilon)$.

This type of BIFURCATION is known as a fold bifurcation, sometimes also called a SADDLE-NODE BIFURCATION or TANGENT BIFURCATION. An example of an equation displaying a fold bifurcation is

$$\dot{x} = \mu - x^2$$

(Guckenheimer and Holmes 1997, p. 145).

see also BIFURCATION

References

Guckenheimer, J. and Holmes, P. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields, 3rd ed.* New York: Springer-Verlag, pp. 145–149, 1997.

Rasband, S. N. *Chaotic Dynamics of Nonlinear Systems.* New York: Wiley, pp. 27–28, 1990.

Fold Catastrophe

A CATASTROPHE which can occur for one control factor and one behavior axis.

Folding

The points accessible from c by a single fold which leaves $a_1, \ldots, a_n$ fixed are exactly those points interior to or on the boundary of the intersection of the CIRCLES through c with centers at a_i, for $i = 1, \ldots, n$. Given any three points in the plane a, b, and c, there is an EQUILATERAL TRIANGLE with VERTICES x, y, and z for which a, b, and c are the images of x, y, and z under a single fold. Given any four points in the plane a, b, c, and d, there is some SQUARE with VERTICES x, y, z, and w for which a, b, c, and d are the images of x, y, z, and w under a sequence of at most three folds. Also, any four collinear points are the images of the VERTICES of a suitable SQUARE under at most two folds. Every five (six) points are the images of the VERTICES of suitable regular PENTAGON (HEXAGON) under at most five (six) folds. The least number of folds required for $n \geq 4$ is not known, but some bounds are. In particular, every set of n points is the image of a suitable REGULAR n-gon under at most $F(n)$ folds, where

$$F(n) \leq \begin{cases} \frac{1}{2}(3n-2) & \text{for } n \text{ even} \\ \frac{1}{2}(3n-3) & \text{for } n \text{ odd.} \end{cases}$$

The first few values are 0, 2, 3, 5, 6, 8, 9, 11, 12, 14, 15, 17, 18, 20, 21, ... (Sloane's A007494).

see also FLEXAGON, MAP FOLDING, ORIGAMI

References

Sabinin, P. and Stone, M. G. "Transforming n-gons by Folding the Plane." *Amer. Math. Monthly* **102**, 620–627, 1995.

Sloane, N. J. A. Sequence A007494 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Foliation

Let M^n be an n-MANIFOLD and let $\mathsf{F} = \{F_\alpha\}$ denote a PARTITION of M into DISJOINT path-connected SUBSETS. Then F is called a foliation of M of codimension c (with $0 < c < n$) if there EXISTS a COVER of M by OPEN SETS U, each equipped with a HOMEOMORPHISM $h : U \to \mathbb{R}^n$ or $h : U \to \mathbb{R}^n_+$ which throws each nonempty component of $F_\alpha \cap U$ onto a parallel translation of the standard HYPERPLANE $\mathbb{R}^{n-c}$ in $\mathbb{R}^n$. Each F_α is then called a LEAF and is not necessarily closed or compact.

see also LEAF (FOLIATION), REEB FOLIATION

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 284, 1976.

Folium

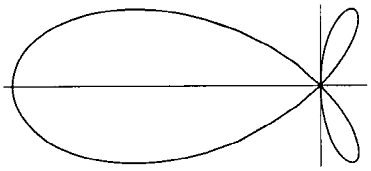

The word "folium" means leaf-shaped. The polar equation is

$$r = \cos\theta(4a\sin^2\theta - b).$$

If $b \geq 4a$, it is a single folium. If $b = 0$, it is a BIFOLIUM. If $0 < b < 4a$, it is a TRIFOLIUM. The simple folium is the PEDAL CURVE of the DELTOID where the PEDAL POINT is one of the CUSPS.

see also BIFOLIUM, FOLIUM OF DESCARTES, KEPLER'S FOLIUM, QUADRIFOLIUM, ROSE, TRIFOLIUM

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 152–153, 1972.

MacTutor History of Mathematics Archive. "Folium." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Folium.html`.

Folium of Descartes

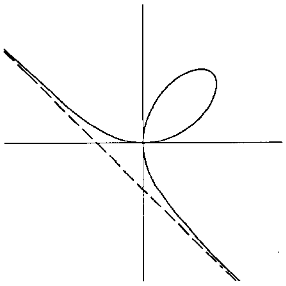

A plane curve proposed by Descartes to challenge Fermat's extremum-finding techniques. In parametric form,

$$x = \frac{3at}{1+t^3} \tag{1}$$

$$y = \frac{3at^2}{1+t^3}. \tag{2}$$

The curve has a discontinuity at $t = -1$. The left wing is generated as t runs from -1 to 0, the loop as t runs from 0 to ∞, and the right wing as t runs from $-\infty$ to -1.

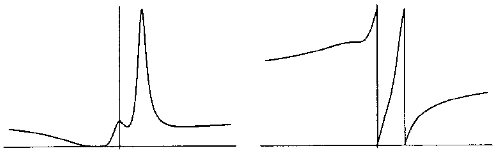

The CURVATURE and TANGENTIAL ANGLE of the folium of Descartes, illustrated above, are

$$\kappa(t) = \frac{2(1+t^3)^4}{3(1+4t^2-4t^3-4t^5+4t^6+t^8)^{3/2}} \tag{3}$$

$$\phi(t) = \frac{1}{2}\left[\pi + \tan^{-1}\left(\frac{1-2t^3}{t^4-2t}\right) - \tan^{-1}\left(\frac{2t^3-1}{t^4-2t}\right)\right]. \tag{4}$$

Converting the parametric equations to POLAR COORDINATES gives

$$r^2 = \frac{(3at)^2(1+t^2)}{(1+t^3)^2} \tag{5}$$

$$\theta = \tan^{-1}\left(\frac{y}{x}\right) = \tan^{-1} t, \tag{6}$$

so

$$d\theta = \frac{dt}{1+t^2}. \tag{7}$$

The AREA enclosed by the curve is

$$\begin{aligned} A &= \tfrac{1}{2}\int r^2\,d\theta = \tfrac{1}{2}\int_0^\infty \frac{(3at)^2(1+t^2)}{(1+t^3)^2}\,\frac{dt}{1+t^2} \\ &= \tfrac{3}{2}a^2\int_0^\infty \frac{3t^2\,dt}{(1+t^3)^2}. \end{aligned} \tag{8}$$

Now let $u \equiv 1+t^3$ so $du = 3t^2\,dt$

$$A = \tfrac{3}{2}a^2\int_1^\infty \frac{du}{u^2} = \tfrac{3}{2}a^2\left[-\frac{1}{u}\right]_1^\infty = \tfrac{3}{2}a^2(-0+1) = \tfrac{3}{2}a^2. \tag{9}$$

In CARTESIAN COORDINATES,

$$x^3 + y^3 = \frac{(3at)^3(1+t^3)}{(1+t^3)^3} = \frac{(3at)^3}{(1+t^3)^2} = 3axy \tag{10}$$

(MacTutor Archive). The equation of the ASYMPTOTE is

$$y = -a - x. \tag{11}$$

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 59–62, 1993.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 106–109, 1972.

MacTutor History of Mathematics Archive. "Folium of Descartes." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Foliumd.html`.

Stroeker, R. J. "Brocard Points, Circulant Matrices, and Descartes' Folium." *Math. Mag.* **61**, 172–187, 1988.

Yates, R. C. "Folium of Descartes." In *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 98–99, 1952.

Follows

see SUCCEEDS

Fontené Theorems

1. If the sides of the PEDAL TRIANGLE of a point P meet the corresponding sides of a TRIANGLE $\Delta O_1O_2O_3$ at X_1, X_2, and X_3, respectively, then P_1X_1, P_2X_2, P_3X_3 meet at a point L common to the CIRCLES $O_1O_2O_3$ and $P_1P_2P_3$. In other words, L is one of the intersections of the NINE-POINT CIRCLE of $A_1A_2A_3$ and the PEDAL CIRCLE of P.
2. If a point moves on a fixed line through the CIRCUMCENTER, then its PEDAL CIRCLE passes through a fixed point on the NINE-POINT CIRCLE.

3. The PEDAL CIRCLE of a point is tangent to the NINE-POINT CIRCLE IFF the point and its ISOGONAL CONJUGATE lie on a LINE through the ORTHOCENTER. FEUERBACH'S THEOREM is a special case of this theorem.

see also CIRCUMCENTER, FEUERBACH'S THEOREM, ISOGONAL CONJUGATE, NINE-POINT CIRCLE, ORTHOCENTER, PEDAL CIRCLE

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 245–247, 1929.

Foot

see PERPENDICULAR FOOT

For All

If a proposition P is true for all B, this is written $P \forall B$.

see also ALMOST ALL, EXISTS, QUANTIFIER

Forcing

A technique in SET THEORY invented by P. Cohen (1963, 1964, 1966) and used to prove that the AXIOM OF CHOICE and CONTINUUM HYPOTHESIS are independent of one another in ZERMELO-FRAENKEL SET THEORY.

see also AXIOM OF CHOICE, CONTINUUM HYPOTHESIS, SET THEORY, ZERMELO-FRAENKEL SET THEORY

References

Cohen, P. J. "The Independence of the Continuum Hypothesis." *Proc. Nat. Acad. Sci. U. S. A.* **50**, 1143–1148, 1963.

Cohen, P. J. "The Independence of the Continuum Hypothesis. II." *Proc. Nat. Acad. Sci. U. S. A.* **51**, 105–110, 1964.

Cohen, P. J. *Set Theory and the Continuum Hypothesis.* New York: W. A. Benjamin, 1966.

Ford Circle

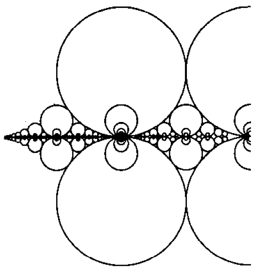

Pick any two INTEGERS h and k, then the CIRCLE of RADIUS $1/(2k^2)$ centered at $(h/k, 1/(2k^2))$ is known as a Ford circle. No matter what and how many hs and ks are picked, none of the Ford circles intersect (and all are tangent to the x-AXIS). This can be seen by examining the squared distance between the centers of the circles with (h,k) and (h',k'),

$$d^2 = \left(\frac{h'}{k'} - \frac{h}{k}\right)^2 + \left(\frac{1}{2k'^2} - \frac{1}{2k^2}\right). \tag{1}$$

Let s be the sum of the radii

$$s = r_1 + r_2 = \frac{1}{2k^2} + \frac{1}{2k'^2}, \tag{2}$$

then

$$d^2 - s^2 = \frac{(h'k - hk')^2 - 1}{k^2 k'^2}. \tag{3}$$

But $(h'k - k'h)^2 \geq 1$, so $d^2 - s^2 \geq 0$ and the distance between circle centers is $\geq$ the sum of the CIRCLE RADII, with equality (and therefore tangency) IFF $|h'k - k'h| = 1$. Ford circles are related to the FAREY SEQUENCE (Conway and Guy 1996).

see also ADJACENT FRACTION, FAREY SEQUENCE, STERN-BROCOT TREE

References

Conway, J. H. and Guy, R. K. "Farey Fractions and Ford Circles." *The Book of Numbers.* New York: Springer-Verlag, pp. 152–154, 1996.

Ford, L. R. "Fractions." *Amer. Math. Monthly* **45**, 586–601, 1938.

Pickover, C. A. "Fractal Milkshakes and Infinite Archery." Ch. 14 in *Keys to Infinity.* New York: W. H. Freeman, pp. 117–125, 1995.

Rademacher, H. *Higher Mathematics from an Elementary Point of View.* Boston, MA: Birkhäuser, 1983.

Ford's Theorem

Let a, b, and k be INTEGERS with $k \geq 1$. For $j = 0$, 1, 2, let

$$S_j \equiv \sum_{\substack{i=0 \\ i \equiv j \pmod{3}}} (-1)^j \binom{k}{i} a^{k-i} b^i.$$

Then

$$2(a^2 + ab + b^2)^{2k} = (S_0 - S_1)^4 + (S_1 - S_2)^4 + (S_2 - S_0)^4.$$

see also BHARGAVA'S THEOREM, DIOPHANTINE EQUATION—QUARTIC

References

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 100–101, 1994.

Forest

A GRAPH without any CIRCUITS (CYCLES), which therefore consists only of TREES. A forest with k components and n nodes has $n - k$ EDGES.

Fork

see TREE

Form

see CANONICAL FORM, CUSP FORM, DIFFERENTIAL k-FORM, FORM (GEOMETRIC), FORM (POLYNOMIAL), MODULAR FORM, NORMAL FORM, PFAFFIAN FORM, QUADRATIC FORM

Form (Geometric)

A 1-D geometric object such as a Pencil or Range.

Form (Polynomial)

A Homogeneous Polynomial in two or more variables.

see also Disconnected Form, k-Form

Formal Logic

see Symbolic Logic

Formosa Theorem

see Chinese Remainder Theorem

Formula

A mathematical equation or a formal logical expression. The correct Latin plural form of formula is "formulae," although the less pretentious-sounding "formulas" is used more commonly.

see also Archimedes' Recurrence Formula, Bayes' Formula, Benson's Formula, Bessel's Finite Difference Formula, Bessel's Interpolation Formula, Bessel's Statistical Formula, Binet's Formula, Binomial Formula, Brahmagupta's Formula, Brent-Salamin Formula, Bretschneider's Formula, Brioschi Formula, Calderón's Formula, Cardano's Formula, Cauchy's Formula, Cauchy's Cosine Integral Formula, Cauchy Integral Formula, Chasles-Cayley-Brill Formula, Chebyshev Approximation Formula, Christoffel-Darboux Formula, Christoffel Formula, Clausen Formula, Clenshaw Recurrence Formula, Descartes-Euler Polyhedral Formula, Descartes' Formula, Dirichlet's Formula, Dixon-Ferrar Formula, Dobiński's Formula, Duplication Formula, Enneper-Weierstraß Parameterization, Euler Curvature Formula, Euler Formula, Euler-Maclaurin Integration Formulas, Euler Polyhedral Formula, Euler Triangle Formula, Everett's Formula, Exponential Sum Formulas, Faulhaber's Formula, Frenet Formulas, Gauss's Backward Formula, Gauss-Bonnet Formula, Gauss's Formula, Gauss's Forward Formula, Gauss Multiplication Formula, Gauss-Salamin Formula, Girard's Spherical Excess Formula, Goodman's Formula, Gregory's Formula, Grenz-Formel, Grinberg Formula, Halley's Irrational Formula, Halley's Rational Formula, Hansen-Bessel Formula, Heron's Formula, Hook Length Formula, Jacobi Elliptic Functions, Jensen's Formula, Jonah Formula, Kac Formula, Kneser-Sommerfeld Formula, Kummer's Formulas, Laisant's Recurrence Formula, Landen's Formula, Lefshetz Fixed Point Formula, Lefshetz Trace Formula, Legendre Duplication Formula, Legendre's Formula, Lehmer's Formula, Lichnerowicz Formula, Lichnerowicz-Weitzenbock Formula, Lobachevsky's Formula, Logarithmic Binomial Formula, Ludwig's Inversion Formula, Machin's Formula, Machin-Like Formulas, Mehler's Bessel Function Formula, Mehler's Hermite Polynomial Formula, Meissel's Formula, Mensuration Formula, Möbius Inversion Formula, Morley's Formula, Newton's Backward Difference Formula, Newton-Cotes Formulas, Newton's Forward Difference Formula, Nicholson's Formula, Pascal's Formula, Pick's Formula, Poincaré Formula, Poisson's Bessel Function Formula, Poisson's Harmonic Function Formula, Poisson Sum Formula, Polyhedral Formula, Prosthaphaeresis Formulas, Quadratic Formula, Quadrature Formulas, Rayleigh's Formulas, Riemann's Formula, Rodrigues Formula, Rotation Formula, Schläfli's Formula, Schröter's Formula, Schwenk's Formula, Segner's Recurrence Formula, Serret-Frenet Formulas, Sherman-Morrison Formula, Sommerfeld's Formula, Sonine-Schafheitlin Formula, Steffenson's Formula, Stirling's Finite Difference Formula, Stirling's Formula, Strassen Formulas, Thiele's Interpolation Formula, Wallis Formula, Watson's Formula, Watson-Nicholson Formula, Weber's Formula, Weber-Sonine Formula, Weyrich's Formula, Woodbury Formula

References

Carr, G. S. *Formulas and Theorems in Pure Mathematics.* New York: Chelsea, 1970.

Spiegel, M. R. *Mathematical Handbook of Formulas and Tables.* New York: McGraw-Hill, 1968.

Tallarida, R. J. *Pocket Book of Integrals and Mathematical Formulas, 3rd ed.* Boca Raton, FL: CRC Press, 1992.

Fortunate Prime

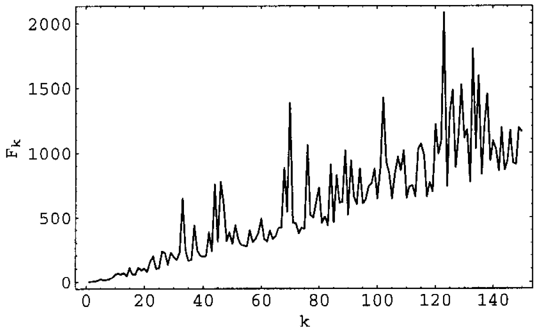

Let

$$X_k \equiv 1 + p_k\#,$$

where p_k is the kth Prime and $p\#$ is the Primorial, and let q_k be the Next Prime (i.e., the smallest Prime greater than X_k),

$$q_k = p_{1+\pi(X_k)} = p_{1+\pi(1+p_k\#)},$$

where $\pi(n)$ is the PRIME COUNTING FUNCTION. Then R. F. Fortune conjectured that $F_k \equiv q_k - X_k + 1$ is PRIME for all k. The first values of F_k are 3, 5, 7, 13, 23, 17, 19, 23, ... (Sloane's A005235), and all known values of F_k are indeed PRIME (Guy 1994). The indices of these primes are 2, 3, 4, 6, 9, 7, 8, 9, 12, 18, ... In numerical order with duplicates removed, the Fortunate primes are 3, 5, 7, 13, 17, 19, 23, 37, 47, 59, 61, 67, 71, 79, 89, ... (Sloane's A046066).

see also ANDRICA'S CONJECTURE, PRIMORIAL

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 7, 1994.

Sloane, N. J. A. Sequences A046066 and A005235/M2418 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Forward Difference

The forward difference is a FINITE DIFFERENCE defined by

$$\Delta f_p \equiv f_{p+1} - f_p. \tag{1}$$

Higher order differences are obtained by repeated operations of the forward difference operator, so

$$\begin{aligned}\Delta^2 f_p &= {\Delta_p}^2 = \Delta(\Delta_p) = \Delta(f_{p+1} - f_p)\\ &= \Delta_{p+1} - \Delta_p = f_{p+2} - 2f_{p+1} + f_p.\end{aligned} \tag{2}$$

In general,

$$\Delta_p^k \equiv \Delta^k f_p \equiv \sum_{m=0}^{k} (-1)^m \binom{k}{m} f_{p+k-m}, \tag{3}$$

where $\binom{k}{m}$ is a BINOMIAL COEFFICIENT.

NEWTON'S FORWARD DIFFERENCE FORMULA expresses f_p as the sum of the nth forward differences

$$f_p = f_0 + p\Delta_0 + \tfrac{1}{2!}p(p+1)\Delta_0^2 + \tfrac{1}{3!}p(p+1)(p+2)\Delta_0^3 + \ldots \tag{4}$$

where Δ_0^n is the first nth difference computed from the difference table.

see also BACKWARD DIFFERENCE, CENTRAL DIFFERENCE, DIFFERENCE EQUATION, DIVIDED DIFFERENCE, RECIPROCAL DIFFERENCE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 877, 1972.

Fountain

An (n, k) fountain is an arrangement of n coins in rows such that exactly k coins are in the bottom row and each coin in the $(i + 1)$st row touches exactly two in the ith row.

References

Berndt, B. C. *Ramanujan's Notebooks, Part III.* New York: Springer-Verlag, p. 79, 1985.

Four Coins Problem

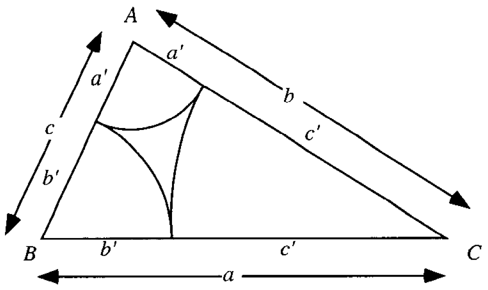

Given three coins of possibly different sizes which are arranged so that each is tangent to the other two, find the coin which is tangent to the other three coins. The solution is the inner SODDY CIRCLE.

see also APOLLONIUS CIRCLES, APOLLONIUS' PROBLEM, ARBELOS, BEND (CURVATURE), CIRCUMCIRCLE, COIN, DESCARTES CIRCLE THEOREM, HART'S THEOREM, PAPPUS CHAIN, SODDY CIRCLES, SPHERE PACKING, STEINER CHAIN

References

Oldknow, A. "The Euler-Gergonne-Soddy Triangle of a Triangle." *Amer. Math. Monthly* **103**, 319–329, 1996.

Four-Color Theorem

The four-color theorem states that any map in a PLANE can be colored using four-colors in such a way that regions sharing a common boundary (other than a single point) do not share the same color. This problem is sometimes also called GUTHRIE'S PROBLEM after F. Guthrie, who first conjectured the theorem in 1853. The CONJECTURE was then communicated to de Morgan and thence into the general community. In 1878, Cayley wrote the first paper on the conjecture.

Fallacious proofs were given independently by Kempe (1879) and Tait (1880). Kempe's proof was accepted for a decade until Heawood showed an error using a map with 18 faces (although a map with nine faces suffices to show the fallacy). The HEAWOOD CONJECTURE provided a very general result for map coloring, showing that in a GENUS 0 SPACE (i.e., either the SPHERE or PLANE), six colors suffice. This number can easily be reduced to five, but reducing the number of colors all the way to four proved very difficult.

Finally, Appel and Haken (1977) announced a computer-assisted proof that four colors were SUFFICIENT. However, because part of the proof consisted of an exhaustive analysis of many discrete cases by a computer, some mathematicians do not accept it. However, no flaws have yet been found, so the proof appears valid. A potentially independent proof has recently been constructed by N. Robertson, D. P. Sanders, P. D. Seymour, and R. Thomas.

Martin Gardner (1975) played an April Fool's joke by (incorrectly) claiming that the map of 110 regions illustrated below requires five colors and constitutes a counterexample to the four-color theorem.

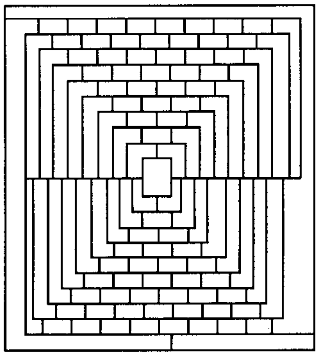

see also CHROMATIC NUMBER, HEAWOOD CONJECTURE, MAP COLORING, SIX-COLOR THEOREM

References

Appel, K. and Haken, W. "Every Planar Map is Four-Colorable, I and II." *Illinois J. Math.* **21**, 429–567, 1977.

Appel, K. and Haken, W. "The Solution of the Four-Color Map Problem." *Sci. Amer.* **237**, 108–121, 1977.

Appel, K. and Haken, W. *Every Planar Map is Four-Colorable.* Providence, RI: Amer. Math. Soc., 1989.

Barnette, D. *Map Coloring, Polyhedra, and the Four-Color Problem.* Providence, RI: Math. Assoc. Amer., 1983.

Birkhoff, G. D. "The Reducibility of Maps." *Amer. Math. J.* **35**, 114–128, 1913.

Chartrand, G. "The Four Color Problem." §9.3 in *Introductory Graph Theory.* New York: Dover, pp. 209–215, 1985.

Coxeter, H. S. M. "The Four-Color Map Problem, 1840–1890." *Math. Teach.*, Apr. 1959.

Franklin, P. *The Four-Color Problem.* New York: Scripta Mathematica, Yeshiva College, 1941.

Gardner, M. "Mathematical Games: The Celebrated Four-Color Map Problem of Topology." *Sci. Amer.* **203**, 218–222, Sep. 1960.

Gardner, M. "The Four-Color Map Theorem." Ch. 10 in *Martin Gardner's New Mathematical Diversions from Scientific American.* New York: Simon and Schuster, pp. 113–123, 1966.

Gardner, M. "Mathematical Games: Six Sensational Discoveries that Somehow or Another have Escaped Public Attention." *Sci. Amer.* **232**, 127–131, Apr. 1975.

Gardner, M. "Mathematical Games: On Tessellating the Plane with Convex Polygons." *Sci. Amer.* **232**, 112–117, Jul. 1975.

Kempe, A. B. "On the Geographical Problem of Four-Colors." *Amer. J. Math.* **2**, 193–200, 1879.

Kraitchik, M. §8.4.2 in *Mathematical Recreations.* New York: W. W. Norton, p. 211, 1942.

Ore, Ø. *The Four-Color Problem.* New York: Academic Press, 1967.

Pappas, T. "The Four-Color Map Problem: Topology Turns the Tables on Map Coloring." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 152–153, 1989.

Robertson, N.; Sanders, D. P.; and Thomas, R. "The Four-Color Theorem." `http://www.math.gatech.edu/~thomas/FC/fourcolor.html`.

Saaty, T. L. and Kainen, P. C. *The Four-Color Problem: Assaults and Conquest.* New York: Dover, 1986.

Tait, P. G. "Note on a Theorem in Geometry of Position." *Trans. Roy. Soc. Edinburgh* **29**, 657–660, 1880.

Four Travelers Problem

Let four LINES in a PLANE represent four roads in GENERAL POSITION, and let one traveler T_i be walking along each road at a constant (but not necessarily equal to any other traveler's) speed. Say that two travelers T_i and T_j have "met" if they were simultaneously at the intersection of their two roads. Then if T_1 has met all other three travelers (T_2, T_3, and T_4) and T_2, in addition to meeting T_1, has met T_3 and T_4, then T_3 and T_4 have also met!

References

Bogomolny, A. "Four Travellers Problem." `http://www.cut-the-knot.com/gproblems.html`.

Four-Vector

A four-element vector

$$a^\mu = \begin{bmatrix} a^0 \\ a^1 \\ a^2 \\ a^3 \end{bmatrix}, \tag{1}$$

which transforms under a LORENTZ TRANSFORMATION like the POSITION FOUR-VECTOR. This means it obeys

$$a'^\mu = \Lambda^\mu_\nu a^\nu \tag{2}$$

$$a_\mu \cdot b_\mu \equiv a_\mu b^\mu \tag{3}$$

$$a_\mu \cdot b^\mu = a'_\mu b'_\mu, \tag{4}$$

where Λ^μ_μ is the LORENTZ TENSOR. Multiplication of two four-vectors with the METRIC $g_{\mu\nu}$ gives products of the form

$$g_{\mu\nu} x^\mu x^\nu = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2. \tag{5}$$

In the case of the POSITION FOUR-VECTOR, $x^0 = ct$ (where c is the speed of light) and this product is an invariant known as the spacetime interval.

see also GRADIENT FOUR-VECTOR, LORENTZ TRANSFORMATION, POSITION FOUR-VECTOR, QUATERNION

References

Morse, P. M. and Feshbach, H. "The Lorentz Transformation, Four-Vectors, Spinors." §1.7 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 93–107, 1953.

Four-Vertex Theorem

A closed embedded smooth PLANE CURVE has at least four vertices, where a vertex is defined as an extremum of CURVATURE.

see also CURVATURE

References

Tabachnikov, S. "The Four-Vertex Theorem Revisited—Two Variations on the Old Theme." *Amer. Math. Monthly* **102**, 912–916, 1995.

Fourier-Bessel Series

see BESSEL FUNCTION FOURIER EXPANSION, SCHLÖMILCH'S SERIES

Fourier-Bessel Transform

see HANKEL TRANSFORM

Fourier Cosine Series

If $f(x)$ is an EVEN FUNCTION, then $b_n = 0$ and the FOURIER SERIES collapses to

$$f(x) = \tfrac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos(nx), \tag{1}$$

where

$$\begin{aligned} a_0 &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,dx = \frac{2}{\pi}\int_0^{\pi} f(x)\,dx && (2)\\ a_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)\,dx \\ &= \frac{2}{\pi}\int_0^{\pi} f(x)\cos(nx)\,dx, && (3)\end{aligned}$$

where the last equality is true because

$$f(x)\cos(nx) = f(-x)\cos(-nx). \tag{4}$$

Letting the range go to L,

$$a_0 = \frac{1}{L}\int_0^L f(x)\,dx \tag{5}$$

$$a_n = \frac{2}{L}\int_0^L f(x)\cos\left(\frac{n\pi x}{L}\right)dx. \tag{6}$$

see also EVEN FUNCTION, FOURIER COSINE TRANSFORM, FOURIER SERIES, FOURIER SINE SERIES

Fourier Cosine Transform

The Fourier cosine transform is the REAL PART of the full complex FOURIER TRANSFORM,

$$\mathcal{F}\cos[f(x)] = \Re[\mathcal{F}[f(x)]].$$

see also FOURIER SINE TRANSFORM, FOURIER TRANSFORM

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "FFT of Real Functions, Sine and Cosine Transforms." §12.3 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 504–515, 1992.

Fourier Integral

see FOURIER TRANSFORM

Fourier Matrix

The $n \times n$ SQUARE MATRIX F_n with entries given by

$$F_{jk} = e^{2\pi ijk/n} \tag{1}$$

for $j, k = 1, 2, \ldots, n$, and normalized by $1/\sqrt{n}$ to make it a UNITARY. The Fourier matrix F_2 is given by

$$\mathsf{F}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & i^2 \end{bmatrix}, \tag{2}$$

and the F_4 matrix by

$$\begin{aligned}\mathsf{F}_4 &= \frac{1}{\sqrt{4}}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} 1 & & 1 & \\ & 1 & & i \\ 1 & & -1 & \\ & 1 & & -i \end{bmatrix}\begin{bmatrix} 1 & 1 & & \\ 1 & i^2 & & \\ & & 1 & 1 \\ & & 1 & i^2 \end{bmatrix}\begin{bmatrix} 1 & & & \\ & & 1 & \\ & 1 & & \\ & & & 1 \end{bmatrix}.\end{aligned} \tag{3}$$

In general,

$$\mathsf{F}_{2n} = \begin{bmatrix} \mathsf{I}_n & \mathsf{D}_n \\ \mathsf{I}_n & -\mathsf{D}_n \end{bmatrix}\begin{bmatrix} \mathsf{F}_n & \\ & \mathsf{F}_n \end{bmatrix}\begin{bmatrix} \text{even-odd} \\ \text{shuffle} \end{bmatrix}, \tag{4}$$

with

$$\begin{aligned}\begin{bmatrix} \mathsf{F}_n & \\ & \mathsf{F}_n \end{bmatrix} &= \begin{bmatrix} \mathsf{I}_{n/2} & \mathsf{D}_{n/2} & & \\ \mathsf{I}_{n/2} & -\mathsf{D}_{n/2} & & \\ & & \mathsf{I}_{n/2} & \mathsf{D}_{n/2} \\ & & \mathsf{I}_{n/2} & -\mathsf{D}_{n/2} \end{bmatrix} \\ &\times \begin{bmatrix} \mathsf{F}_{n/2} & & & \\ & \mathsf{F}_{n/2} & & \\ & & \mathsf{F}_{n/2} & \\ & & & \mathsf{F}_{n/2} \end{bmatrix}\begin{bmatrix} \text{even-odd} \\ 0, 2 \pmod 4 \\ \text{even-odd} \\ 1, 3 \pmod 4 \end{bmatrix},\end{aligned} \tag{5}$$

where I_n is the $n \times n$ IDENTITY MATRIX. Note that the factorization (which is the basis of the FAST FOURIER TRANSFORM) has two copies of F_2 in the center factor MATRIX.

see also FAST FOURIER TRANSFORM, FOURIER TRANSFORM

References

Strang, G. "Wavelet Transforms Versus Fourier Transforms." *Bull. Amer. Math. Soc.* **28**, 288–305, 1993.

Fourier-Mellin Integral

The inverse of the LAPLACE TRANSFORM

$$F(t) = \mathcal{L}^{-1}[f(s)] = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{st} f(s)\,ds$$

$$f(s) = \mathcal{L}[F(t)] = \int_0^{\infty} F(t)e^{-st}\,dt.$$

see also BROMWICH INTEGRAL, LAPLACE TRANSFORM

Fourier Series

Fourier series are expansions of PERIODIC FUNCTIONS $f(x)$ in terms of an infinite sum of SINES and COSINES

$$f(x) = \sum_{n=0}^{\infty} a_n \cos(nx) + \sum_{n=0}^{\infty} b_n \sin(nx). \tag{1}$$

Fourier series make use of the ORTHOGONALITY relationships of the SINE and COSINE functions, which can be used to calculate the coefficients a_n and b_n in the sum. The computation and study of Fourier series is known as HARMONIC ANALYSIS.

To compute a Fourier series, use the integral identities

$$\int_{-\pi}^{\pi} \sin(mx)\sin(nx)\,dx = \pi\delta_{mn} \qquad \text{for } n, m \neq 0 \tag{2}$$

$$\int_{-\pi}^{\pi} \cos(mx)\cos(nx)\,dx = \pi\delta_{mn} \qquad \text{for } n, m \neq 0 \tag{3}$$

$$\int_{-\pi}^{\pi} \sin(mx)\cos(nx)\,dx = 0 \tag{4}$$

$$\int_{-\pi}^{\pi} \sin(mx)\,dx = 0 \tag{5}$$

$$\int_{-\pi}^{\pi} \cos(mx)\,dx = 0, \tag{6}$$

where δ_{mn} is the KRONECKER DELTA. Now, expand your function $f(x)$ as an infinite series of the form

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} a'_n \cos(nx) + \sum_{n=0}^{\infty} b_n \sin(nx) \\ &= \tfrac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx), \end{aligned} \tag{7}$$

where we have relabeled the $a_0 = 2a'_0$ term for future convenience but left $a_n = a'_n$. Assume the function is periodic in the interval $[-\pi, \pi]$. Now use the orthogonality conditions to obtain

$$\begin{aligned} &\int_{-\pi}^{\pi} f(x)\,dx \\ &= \int_{-\pi}^{\pi} \left[\sum_{n=1}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx) + \tfrac{1}{2}a_0\right] dx \\ &= \sum_{n=1}^{\infty} \int_{-\pi}^{\pi} [a_n \cos(nx) + b_n \sin(nx)]\,dx + \tfrac{1}{2}a_0 \int_{-\pi}^{\pi} dx \\ &= \sum_{n=1}^{\infty} (0+0) + \pi a_0 = \pi a_0 \end{aligned} \tag{8}$$

and

$$\begin{aligned} &\int_{-\pi}^{\pi} f(x)\sin(mx)\,dx \\ &= \int_{-\pi}^{\pi} \left[\sum_{n=1}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx) + \tfrac{1}{2}a_0\right] \times \sin(mx)\,dx \\ &= \sum_{n=1}^{\infty} \int_{-\pi}^{\pi} [a_n \cos(nx)\sin(mx) + b_n \sin(nx)\sin(mx)]\,dx \\ &\quad + \tfrac{1}{2}a_0 \int_{-\pi}^{\pi} \sin(mx)\,dx \\ &= \sum_{n=1}^{\infty} (0 + b_n\pi\delta_{mn}) + 0 = \pi b_n, \end{aligned} \tag{9}$$

so

$$\begin{aligned} \int_{-\pi}^{\pi} f(x)\cos(mx)\,dx &= \int_{-\pi}^{\pi} \left[\sum_{n=1}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx) + \tfrac{1}{2}a_0\right] \cos(mx)\,dx \\ &= \sum_{n=1}^{\infty} \int_{-\pi}^{\pi} [a_n \cos(nx)\cos(mx) \\ &\quad + b_n \sin(nx)\cos(mx)]\,dx + \tfrac{1}{2}a_0 \int_{-\pi}^{\pi} \cos(mx)\,dx \\ &= \sum_{n=1}^{\infty} (a_n\pi\delta_{mn} + 0) + 0 = \pi a_n. \end{aligned} \tag{10}$$

Plugging back into the original series then gives

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\,dx \tag{11}$$

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\cos(nx)\,dx \tag{12}$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\sin(nx)\,dx \tag{13}$$

for $n = 1, 2, 3, \ldots$. The series expansion converges to the function $\bar{f}$ (equal to the original function at points of continuity or to the average of the two limits at points of discontinuity)

$$\bar{f} \equiv \begin{cases} \frac{1}{2}\left[\lim_{x \to x_0^-} f(x) + \lim_{x \to x_0^+} f(x)\right] \\ \quad \text{for } -\pi < x_0 < \pi \\ \frac{1}{2}\left[\lim_{x \to -\pi^+} f(x) + \lim_{x \to \pi_-} f(x)\right] \\ \quad \text{for } x_0 = -\pi, \pi \end{cases} \tag{14}$$

if the function satisfies the DIRICHLET CONDITIONS.

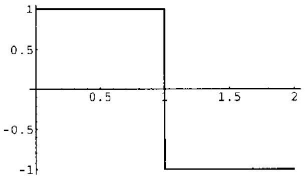

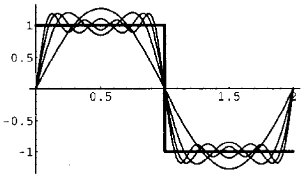

Near points of discontinuity, a "ringing" known as the GIBBS PHENOMENON, illustrated below, occurs. For a function $f(x)$ periodic on an interval $[-L, L]$, use a change of variables to transform the interval of integration to $[-1, 1]$. Let

$$x \equiv \frac{\pi x'}{L} \tag{15}$$

$$dx = \frac{\pi dx'}{L}. \tag{16}$$

Solving for x', $x' = Lx/\pi$. Plugging this in gives

$$f(x') = \tfrac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos\left(\frac{n\pi x'}{L}\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi x'}{L}\right) \tag{17}$$

$$\begin{cases} a_0 = \frac{1}{L}\int_{-L}^{L} f(x')\,dx' \\ a_n = \frac{1}{L}\int_{-L}^{L} f(x')\cos\left(\frac{n\pi x'}{L}\right)\,dx' \\ b_n = \frac{1}{L}\int_{-L}^{L} f(x')\sin\left(\frac{n\pi x'}{L}\right)\,dx' \end{cases}. \tag{18}$$

If a function is EVEN so that $f(x) = f(-x)$, then $f(x)\sin(nx)$ is ODD. (This follows since $\sin(nx)$ is ODD and an EVEN FUNCTION times an ODD FUNCTION is an ODD FUNCTION.) Therefore, $b_n = 0$ for all n. Similarly, if a function is ODD so that $f(x) = f(-x)$, then $f(x)\cos(nx)$ is ODD. (This follows since $\cos(nx)$ is EVEN and an EVEN FUNCTION times an ODD FUNCTION is an ODD FUNCTION.) Therefore, $a_n = 0$ for all n.

Because the SINES and COSINES form a COMPLETE ORTHOGONAL BASIS, the SUPERPOSITION PRINCIPLE holds, and the Fourier series of a linear combination of two functions is the same as the linear combination of the corresponding two series. The COEFFICIENTS for Fourier series expansions for a few common functions are given in Beyer (1987, pp. 411–412) and Byerly (1959, p. 51).

The notion of a Fourier series can also be extended to COMPLEX COEFFICIENTS. Consider a real-valued function $f(x)$. Write

$$f(x) = \sum_{n=-\infty}^{\infty} A_n e^{inx}. \tag{19}$$

Now examine

$$\begin{aligned} \int_{-\pi}^{\pi} f(x)e^{-imx}\,dx &= \int_{-\pi}^{\pi}\left(\sum_{n=-\infty}^{\infty} A_n e^{inx}\right)e^{-imx}\,dx \\ &= \sum_{n=-\infty}^{\infty} A_n \int_{-\pi}^{\pi} e^{i(n-m)x}\,dx \\ &= \sum_{n=-\infty}^{\infty} A_n \int_{-\pi}^{\pi} \{\cos[(n-m)x] + i\sin[(n-m)x]\}\,dx \\ &= \sum_{m=-\infty}^{\infty} A_n 2\pi\delta_{mn} = 2\pi A_m, \end{aligned} \tag{20}$$

so

$$A_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}\,dx. \tag{21}$$

The COEFFICIENTS can be expressed in terms of those in the FOURIER SERIES

$$\begin{aligned} A_n &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)[\cos(nx) - i\sin(nx)]\,dx \\ &= \begin{cases} \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)[\cos(nx) + i\sin(nx)]\,dx & n < 0 \\ \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,dx & n = 0 \\ \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)[\cos(nx) - i\sin(nx)]\,dx & n > 0 \end{cases} \\ &= \begin{cases} \frac{1}{2}(a_n + ib_n) & n < 0 \\ \frac{1}{2}a_0 & n = 0 \\ \frac{1}{2}(a_n - ib_n) & n > 0. \end{cases} \end{aligned} \tag{22}$$

For a function periodic in $[-L, L]$, these become

$$f(x) = \sum_{n=-\infty}^{\infty} A_n e^{i(2\pi nx/L)} \tag{23}$$

$$A_n = \frac{1}{L}\int_{-L/2}^{L/2} f(x)e^{-i(2\pi nx/L)}\,dx. \tag{24}$$

These equations are the basis for the extremely important FOURIER TRANSFORM, which is obtained by transforming A_n from a discrete variable to a continuous one as the length $L \to \infty$.

see also DIRICHLET FOURIER SERIES CONDITIONS, FOURIER COSINE SERIES, FOURIER SINE SERIES, FOURIER TRANSFORM, GIBBS PHENOMENON, LEBESGUE CONSTANTS (FOURIER SERIES), LEGENDRE SERIES, RIESZ-FISCHER THEOREM, SCHLÖMILCH'S SERIES

References

Arfken, G. "Fourier Series." Ch. 14 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 760–793, 1985.
Beyer, W. H. (Ed.). *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, 1987.
Brown, J. W. and Churchill, R. V. *Fourier Series and Boundary Value Problems, 5th ed.* New York: McGraw-Hill, 1993.
Byerly, W. E. *An Elementary Treatise on Fourier's Series, and Spherical, Cylindrical, and Ellipsoidal Harmonics,*

with Applications to Problems in Mathematical Physics. New York: Dover, 1959.

Carslaw, H. S. *Introduction to the Theory of Fourier's Series and Integrals, 3rd ed., rev. and enl.* New York: Dover, 1950.

Davis, H. F. *Fourier Series and Orthogonal Functions.* New York: Dover, 1963.

Dym, H. and McKean, H. P. *Fourier Series and Integrals.* New York: Academic Press, 1972.

Folland, G. B. *Fourier Analysis and Its Applications.* Pacific Grove, CA: Brooks/Cole, 1992.

Groemer, H. *Geometric Applications of Fourier Series and Spherical Harmonics.* New York: Cambridge University Press, 1996.

Körner, T. W. *Fourier Analysis.* Cambridge, England: Cambridge University Press, 1988.

Körner, T. W. *Exercises for Fourier Analysis.* New York: Cambridge University Press, 1993.

Lighthill, M. J. *Introduction to Fourier Analysis and Generalised Functions.* Cambridge, England: Cambridge University Press, 1958.

Morrison, N. *Introduction to Fourier Analysis.* New York: Wiley, 1994.

Sansone, G. "Expansions in Fourier Series." Ch. 2 in *Orthogonal Functions, rev. English ed.* New York: Dover, pp. 39–168, 1991.

Fourier Series—Power Series

For $f(x) = x^k$ on the INTERVAL $[-L, L)$ and periodic with period $2L$, the FOURIER SERIES is given by

$$\begin{aligned} a_n &= \frac{1}{L}\int_{-L}^{L} x^k \cos\left(\frac{n\pi x}{L}\right)dx \\ &= \frac{2L^k}{1+k}\,{}_1F_2\left(\begin{matrix} 1+\frac{1}{2}k \\ \frac{1}{2} \quad \frac{1}{2}(3+k)\end{matrix}; -\tfrac{1}{4}\pi^2 n^2\right) \\ b_n &= \frac{1}{L}\int_{-L}^{L} x^k \sin\left(\frac{n\pi x}{L}\right)dx \\ &= \frac{2n\pi L^k}{2+k}\,{}_1F_2\left(\begin{matrix} 1+\frac{1}{2}k \\ \frac{3}{2} \quad 2+\frac{1}{2}k\end{matrix}; -\tfrac{1}{4}\pi^2 n^2\right), \end{aligned}$$

where ${}_1F_2(a; b, c; x)$ is a generalized HYPERGEOMETRIC FUNCTION.

Fourier Series—Right Triangle

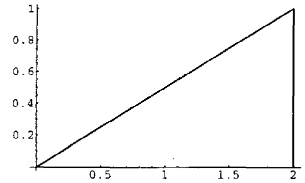

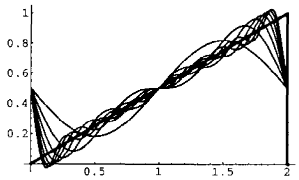

Consider a string of length $2L$ plucked at the right end, then

$$\begin{aligned} a_0 &= \frac{1}{L}\int_0^{2L} \frac{x}{2L}\,dx = \frac{1}{2L^2}\left[\tfrac{1}{2}x^2\right]_0^L = \frac{1}{4L^2}(2L)^2 = 1 \\ a_n &= \frac{1}{L}\int_0^{2L} \frac{x}{2L}\cos\left(\frac{n\pi x}{L}\right)dx \\ &= \frac{[2n\pi\cos(n\pi) - \sin(n\pi)]\sin(n\pi)}{n^2\pi^2} = 0 \\ b_n &= \frac{1}{L}\int_0^{2L} \frac{x}{2L}\sin\left(\frac{n\pi x}{L}\right)dx \\ &= \frac{-2n\pi\cos(2n\pi) + \sin(2n\pi)}{2n^2\pi^2} = -\frac{1}{n\pi}. \end{aligned}$$

The Fourier series is therefore

$$f(x) = \tfrac{1}{2} - \frac{1}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin\left(\frac{n\pi x}{L}\right).$$

see also FOURIER SERIES

Fourier Series—Square Wave

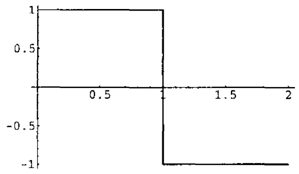

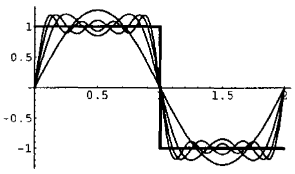

Consider a square wave of length $2L$. Since the function is ODD, $a_0 = a_n = 0$, and

$$\begin{aligned} b_n &= \frac{2}{L}\int_0^{L} \sin\left(\frac{n\pi x}{L}\right)dx \\ &= \frac{4}{n\pi}\sin^2(\tfrac{1}{2}n\pi) = \frac{4}{n\pi}\begin{cases} 0 & n \text{ even} \\ 1 & n \text{ odd.}\end{cases} \end{aligned}$$

The Fourier series is therefore

$$f(x) = \frac{4}{\pi}\sum_{n=1,3,5,\ldots}^{\infty}\frac{1}{n}\sin\left(\frac{n\pi x}{L}\right).$$

see also FOURIER SERIES, SQUARE WAVE

Fourier Series—Triangle

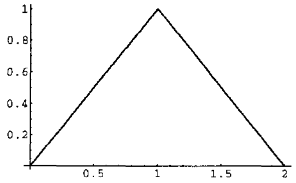

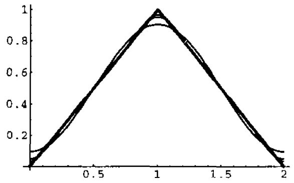

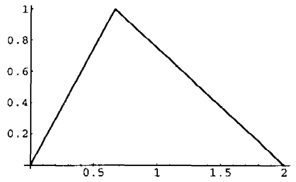

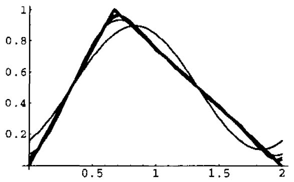

Let a string of length $2L$ have a y-displacement of unity when it is pinned an x-distance which is $(1/m)$th of the way along the string. The displacement as a function of x is then

$$f_m(x) = \begin{cases} \frac{mx}{2L} & 0 \le x \le \frac{2L}{m} \\ \frac{m}{1-m}\left(\frac{x}{2L} - 1\right) & \frac{2L}{m} \le x \le 2L. \end{cases}$$

The COEFFICIENTS are therefore

$$\begin{aligned} a_0 &= \frac{1}{L}\left[\int_0^{2L/m} \frac{nx}{2L}\,dx + \int_{2L/m}^{2L} \frac{n}{1-n}\left(\frac{x}{2L}-1\right)dx\right] \\ &= 1 \\ a_n &= \frac{m\left[1-m-\cos(2\pi n)+m\cos\left(\frac{2n\pi}{m}\right)\right]}{2(m-1)n^2\pi^2} \\ &= \frac{m^2\left[\cos\left(\frac{2n\pi}{m}\right)-1\right]}{2(m-1)n^2\pi^2} \\ b_n &= \frac{m\left[m\sin\left(\frac{2\pi n}{m}\right)-\sin(2\pi n)\right]}{2(m-1)n^2\pi^2} \\ &= \frac{m^2\sin\left(\frac{2\pi n}{m}\right)}{2(m-1)n^2\pi^2}. \end{aligned}$$

The Fourier series is therefore

$$\begin{aligned} f_m(x) = \tfrac{1}{2} &+ \frac{m^2}{2(m-1)\pi^2} \\ &\times \sum_{n=1}^{\infty}\left\{\frac{1}{n^2}\left[\cos\left(\frac{2n\pi}{m}\right)-1\right]\cos\left(\frac{n\pi x}{L}\right)\right. \\ &\qquad \left. + \frac{\sin\left(\frac{2\pi n}{m}\right)}{n^2}\sin\left(\frac{n\pi x}{L}\right)\right\}. \end{aligned}$$

If $m = 2$, then a_n and b_n simplify to

$$\begin{aligned} a_n &= -\frac{4}{n^2\pi^2}\sin^2\left(\tfrac{1}{2}n\pi\right) = -\frac{4}{n^2\pi^2}\begin{cases} 0 & n = 0,\ 2,\ \ldots \\ 1 & n = 1,\ 3,\ \ldots \end{cases} \\ b_n &= 0, \end{aligned}$$

giving

$$f_2(x) = \tfrac{1}{2} - \frac{4}{\pi^2}\sum_{n=1,3,5,\ldots}^{\infty} \frac{1}{n^2}\cos\left(\frac{n\pi x}{L}\right).$$

see also FOURIER SERIES

Fourier Series—Triangle Wave

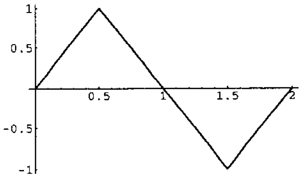

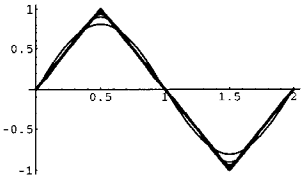

Consider a triangle wave of length $2L$. Since the function is ODD, $a_0 = a_n = 0$, and

$$\begin{aligned} b_n &= \frac{2}{L}\left\{\int_0^{L/2} \frac{x}{L/2}\sin\left(\frac{n\pi x}{L}\right)dx\right. \\ &\qquad \left. + \int_{L/2}^{L}\left[1-\frac{2}{L}\left(x-\tfrac{1}{2}L\right)\right]\sin\left(\frac{n\pi x}{L}\right)dx\right\}dx \\ &= \frac{32}{\pi^2 n^2}\cos\left(\tfrac{1}{4}n\pi\right)\sin^3\left(\tfrac{1}{4}n\pi\right) \\ &= \frac{32}{\pi^2 n^2}\begin{cases} 0 & n = 0,\ 4,\ \ldots \\ \frac{1}{4} & n = 1,\ 5,\ \ldots \\ 0 & n = 2,\ 6,\ \ldots \\ -\frac{1}{4} & n = 3,\ 7,\ \ldots \end{cases} \\ &= \frac{8}{\pi^2 n^2}\begin{cases} (-1)^{(n-1)/2} & \text{for } n \text{ odd} \\ 0 & \text{for } n \text{ even.} \end{cases} \end{aligned}$$

The Fourier series is therefore

$$f(x) = \frac{8}{\pi^2}\sum_{n=1,3,5,\ldots}^{\infty} \frac{(-1)^{(n-1)/2}}{n^2}\sin\left(\frac{n\pi x}{L}\right).$$

see also FOURIER SERIES

Fourier Sine Series

If $f(x)$ is an ODD FUNCTION, then $a_n = 0$ and the FOURIER SERIES collapses to

$$f(x) = \sum_{n=1}^{\infty} b_n \sin(nx), \tag{1}$$

where

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)\,dx = \frac{2}{\pi}\int_0^{\pi} f(x)\sin(nx)\,dx \tag{2}$$

for $n = 1,\ 2,\ 3,\ \ldots$. The last EQUALITY is true because

$$\begin{aligned} f(x)\sin(nx) &= [-f(-x)][-\sin(-nx)] \\ &= f(-x)\sin(-nx). \end{aligned} \tag{3}$$

Letting the range go to L,

$$b_n = \frac{2}{L}\int_0^L f(x)\sin\left(\frac{n\pi x}{L}\right)dx. \tag{4}$$

see also FOURIER COSINE SERIES, FOURIER SERIES, FOURIER SINE TRANSFORM

Fourier Sine Transform

The Fourier sine transform is the IMAGINARY PART of the full complex FOURIER TRANSFORM,

$$\mathcal{F}\sin[f(x)] = \Im[\mathcal{F}[f(x)]].$$

see also FOURIER COSINE TRANSFORM, FOURIER TRANSFORM

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "FFT of Real Functions, Sine and Cosine Transforms." §12.3 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 504–515, 1992.

Fourier-Stieltjes Transform

Let $f(x)$ be a positive definite, measurable function on the INTERVAL $(-\infty, \infty)$. Then there exists a monotone increasing, real-valued bounded function $\alpha(t)$ such that

$$f(x) = \int_{-\infty}^{\infty} e^{itx}\,d\alpha(t)$$

for "ALMOST ALL" x. If $\alpha(t)$ is nondecreasing and bounded and $f(x)$ is defined as above, then $f(x)$ is called the Fourier-Stieltjes transform of $\alpha(t)$, and is both continuous and positive definite.

see also FOURIER TRANSFORM, LAPLACE TRANSFORM

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 618, 1980.

Fourier Transform

The Fourier transform is a generalization of the Complex Fourier Series in the limit as $L \to \infty$. Replace the discrete A_n with the continuous $F(k)\,dk$ while letting $n/L \to k$. Then change the sum to an Integral, and the equations become

$$f(x) = \int_{-\infty}^{\infty} F(k) e^{2\pi ikx}\,dk \tag{1}$$

$$F(k) = \int_{-\infty}^{\infty} f(x) e^{-2\pi ikx}\,dx. \tag{2}$$

Here,

$$F(k) = \mathcal{F}[f(x)] = \int_{-\infty}^{\infty} f(x) e^{-2\pi ikx}\,dx \tag{3}$$

is called the *forward* $(-i)$ Fourier transform, and

$$f(x) = \mathcal{F}^{-1}[F(k)] = \int_{-\infty}^{\infty} F(k) e^{2\pi ikx}\,dk \tag{4}$$

is called the *inverse* $(+i)$ Fourier transform. Some authors (especially physicists) prefer to write the transform in terms of angular frequency $\omega \equiv 2\pi\nu$ instead of the oscillation frequency ν. However, this destroys the symmetry, resulting in the transform pair

$$H(\nu) = \mathcal{F}[h(t)] = \int_{-\infty}^{\infty} h(t) e^{-i\omega t}\,dt \tag{5}$$

$$h(t) = \mathcal{F}^{-1}[H(\nu)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} H(\nu) e^{i\omega t}\,d\omega. \tag{6}$$

In general, the Fourier transform pair may be defined using two arbitrary constants A and B as

$$F(\omega) = A\int_{-\infty}^{\infty} f(t) e^{Bi\omega t}\,dt \tag{7}$$

$$f(t) = \frac{B}{2\pi A}\int_{-\infty}^{\infty} F(\omega) e^{-Bi\omega t}\,d\omega. \tag{8}$$

The *Mathematica*® program (Wolfram Research, Champaign, IL) calls A the `$FourierOverallConstant` and B the `$FourierFrequencyConstant`, and defines $A = B = 1$ by default. Morse and Feshbach (1953) use $B = 1$ and $A = 1/\sqrt{2\pi}$. In this work, following Bracewell (1965, pp. 6–7), $A = 1$ and $B = -2\pi$ unless otherwise stated.

Since any function can be split up into Even and Odd portions $E(x)$ and $O(x)$,

$$f(x) = \tfrac{1}{2}[f(x)+f(-x)]+\tfrac{1}{2}[f(x)-f(-x)] = E(x)+O(x), \tag{9}$$

a Fourier transform can always be expressed in terms of the Fourier Cosine Transform and Fourier Sine Transform as

$$\mathcal{F}[f(x)] = \int_{-\infty}^{\infty} E(x)\cos(2\pi kx)\,dx - i\int_{-\infty}^{\infty} O(x)\sin(2\pi kx)\,dx. \tag{10}$$

A function $f(x)$ has a forward and inverse Fourier transform such that

$$f(x) = \begin{cases} \int_{-\infty}^{\infty} e^{2\pi ikx}\left[\int_{-\infty}^{\infty} f(x)e^{-2\pi ikx}\,dx\right]dk \\ \quad \text{for } f(x) \text{ continuous at } x \\ \frac{1}{2}[f(x_+)+f(x_-)] \\ \quad \text{for } f(x) \text{ discontinuous at } x, \end{cases} \tag{11}$$

provided that

1. $\int_{-\infty}^{\infty} |f(x)|\,dx$ exists.
2. Any discontinuities are finite.
3. The function has bounded variation. A Sufficient weaker condition is fulfillment of the Lipschitz Condition.

The smoother a function (i.e., the larger the number of continuous Derivatives), the more compact its Fourier transform.

The Fourier transform is linear, since if $f(x)$ and $g(x)$ have Fourier Transforms $F(k)$ and $G(k)$, then

$$\begin{aligned}\int [af(x)+bg(x)]e^{-2\pi ikx}\,dx &= a\int_{-\infty}^{\infty} f(x)e^{-2\pi ikx}\,dx + b\int_{-\infty}^{\infty} g(x)e^{-2\pi ikx}\,dx \\ &= F(k)+G(k).\end{aligned} \tag{12}$$

Therefore,

$$\mathcal{F}[af(x)+bg(x)] = a\mathcal{F}[f(x)]+b\mathcal{F}[g(x)] = aF(k)+bG(k). \tag{13}$$

The Fourier transform is also symmetric since $F(k) = \mathcal{F}[f(x)]$ implies $F(-k) = \mathcal{F}[f(x)]$.

Let $f * g$ denote the Convolution, then the transforms of convolutions of functions have particularly nice transforms,

$$\mathcal{F}[f * g] = \mathcal{F}[f]\mathcal{F}[g] \tag{14}$$

$$\mathcal{F}[fg] = \mathcal{F}[f] * \mathcal{F}[g] \tag{15}$$

$$\mathcal{F}[\mathcal{F}(f)+\mathcal{F}(g)] = f * g \tag{16}$$

$$\mathcal{F}[\mathcal{F}(f) * \mathcal{F}(g)] = fg. \tag{17}$$

The first of these is derived as follows:

$$\begin{aligned}\mathcal{F}[f*g] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-2\pi ikx} f(x')g(x-x')\,dx'\,dx\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [e^{-2\pi ikx'} f(x')\,dx']\\ &\quad\times [e^{-2\pi ik(x-x')} g(x-x')\,dx]\\ &= \left[\int_{-\infty}^{\infty} e^{-2\pi ikx'} f(x')\,dx'\right]\\ &\quad\times \left[\int_{-\infty}^{\infty} e^{-2\pi ikx''} g(x'')\,dx''\right]\\ &= \mathcal{F}[f]\mathcal{F}[g], \end{aligned} \qquad (18)$$

where $x'' \equiv x - x'$.

There is also a somewhat surprising and extremely important relationship between the AUTOCORRELATION and the Fourier transform known as the WIENER-KHINTCHINE THEOREM. Let $\mathcal{F}[f(x)] = F(k)$, and F^* denote the COMPLEX CONJUGATE of F, then the FOURIER TRANSFORM of the ABSOLUTE SQUARE of $F(k)$ is given by

$$\mathcal{F}[|F(k)|^2] = \int_{-\infty}^{\infty} f^*(\tau)f(\tau+x)\,d\tau. \qquad (19)$$

The Fourier transform of a DERIVATIVE $f'(x)$ of a function $f(x)$ is simply related to the transform of the function $f(x)$ itself. Consider

$$\mathcal{F}[f'(x)] = \int_{-\infty}^{\infty} f'(x)e^{-2\pi ikx}\,dx. \qquad (20)$$

Now use INTEGRATION BY PARTS

$$\int v\,du = [uv] - \int u\,dv \qquad (21)$$

with

$$du = f'(x)\,dx \qquad v = e^{-2\pi ikx} \qquad (22)$$

$$u = f(x) \qquad dv = -2\pi ike^{-2\pi ikx}\,dx, \qquad (23)$$

then

$$\mathcal{F}[f'(x)] = [f(x)e^{-2\pi ikx}]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(x)(-2\pi ike^{-2\pi ikx}\,dx). \qquad (24)$$

The first term consists of an oscillating function times $f(x)$. But if the function is bounded so that

$$\lim_{x\to\pm\infty} f(x) = 0 \qquad (25)$$

(as any physically significant signal must be), then the term vanishes, leaving

$$\mathcal{F}[f'(x)] = 2\pi ik\int_{-\infty}^{\infty} f(x)e^{-2\pi ikx}\,dx = 2\pi ik\mathcal{F}[f(x)]. \qquad (26)$$

This process can be iterated for the nth DERIVATIVE to yield

$$\mathcal{F}[f^{(n)}(x)] = (2\pi ik)^n\mathcal{F}[f(x)]. \qquad (27)$$

The important MODULATION THEOREM of Fourier transforms allows $\mathcal{F}[\cos(2\pi k_0x)f(x)]$ to be expressed in terms of $\mathcal{F}[f(x)] = F(k)$ as follows,

$$\begin{aligned}\mathcal{F}[\cos(2\pi k_0x)f(x)] &\equiv \int_{-\infty}^{\infty} f(x)\cos(2\pi k_0x)e^{-2\pi ikx}\,dx\\ &= \tfrac{1}{2}\int_{-\infty}^{\infty} f(x)e^{2\pi ik_0x}e^{-2\pi ikx}\,dx\\ &\quad+\tfrac{1}{2}\int_{-\infty}^{\infty} f(x)e^{-2\pi ik_0x}e^{-2\pi ikx}\,dx\\ &= \tfrac{1}{2}\int_{-\infty}^{\infty} f(x)e^{-2\pi i(k-k_0)x}\,dx\\ &\quad+\tfrac{1}{2}\int_{-\infty}^{\infty} f(x)e^{-2\pi i(k+k_0)x}\,dx\\ &= \tfrac{1}{2}[F(k-k_0)+F(k+k_0)]. \end{aligned} \qquad (28)$$

Since the DERIVATIVE of the FOURIER TRANSFORM is given by

$$F'(k) \equiv \frac{d}{dk}\mathcal{F}[f(x)] = \int_{-\infty}^{\infty} (-2\pi ix)f(x)e^{-2\pi ikx}\,dx, \qquad (29)$$

it follows that

$$F'(0) = -2\pi i\int_{-\infty}^{\infty} xf(x)\,dx. \qquad (30)$$

Iterating gives the general FORMULA

$$\mu_n \equiv \int_{-\infty}^{\infty} x^n f(x)\,dx = \frac{F^{(n)}(0)}{(-2\pi i)^n}. \qquad (31)$$

The VARIANCE of a FOURIER TRANSFORM is

$$\sigma_f{}^2 = \left\langle (xf - \langle xf\rangle)^2\right\rangle, \qquad (32)$$

and it is true that

$$\sigma_{f+g} = \sigma_f + \sigma_g. \qquad (33)$$

If $f(x)$ has the FOURIER TRANSFORM $F(k)$, then the Fourier transform has the shift property

$$\begin{aligned}&\int_{-\infty}^{\infty} f(x-x_0)e^{-2\pi ikx}\,dx\\ &\quad= \int_{-\infty}^{\infty} f(x-x_0)e^{-2\pi i(x-x_0)k}e^{-2\pi i(kx_0)}\,d(x-x_0)\\ &\quad= e^{-2\pi ikx_0}F(k), \end{aligned} \qquad (34)$$

so $f(x-x_0)$ has the FOURIER TRANSFORM

$$\mathcal{F}[f(x-x_0)] = e^{-2\pi i k x_0} F(k). \tag{35}$$

If $f(x)$ has a FOURIER TRANSFORM $F(k)$, then the Fourier transform obeys a similarity theorem.

$$\begin{aligned}\int_{-\infty}^{\infty} & f(ax)e^{-2\pi i k x}\,dx \\ &= \frac{1}{|a|}\int_{-\infty}^{\infty} f(ax)e^{-2\pi i(ax)(k/a)}\,d(ax) \\ &= \frac{1}{|a|}F\left(\frac{k}{a}\right),\end{aligned} \tag{36}$$

so $f(ax)$ has the FOURIER TRANSFORM $|a|^{-1}F\left(\frac{k}{a}\right)$.

The "equivalent width" of a Fourier transform is

$$w_e \equiv \frac{\int_{-\infty}^{\infty} f(x)\,dx}{f(0)} = \frac{F(0)}{\int_{-\infty}^{\infty} F(k)\,dk}. \tag{37}$$

The "autocorrelation width" is

$$w_a \equiv \frac{\int_{-\infty}^{\infty} f\star f^*\,dx}{[f\star f^*]_0} = \frac{\int_{-\infty}^{\infty} f\,dx \int_{-\infty}^{\infty} f^*\,dx}{\int_{-\infty}^{\infty} f f^*\,dx}, \tag{38}$$

where $f\star g$ denotes the CROSS-CORRELATION of f and g.

Any operation on $f(x)$ which leaves its AREA unchanged leaves $F(0)$ unchanged, since

$$\int_{-\infty}^{\infty} f(x)\,dx = \mathcal{F}[f(0)] = F(0). \tag{39}$$

In 2-D, the Fourier transform becomes

$$F(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(k_x,k_y)e^{-2\pi i(k_x x + k_y y)}\,dk_x\,dk_y \tag{40}$$

$$f(k_x,k_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x,y)e^{2\pi i(k_x x + k_y y)}\,dx\,dy. \tag{41}$$

Similarly, the n-D Fourier transform can be defined for $\mathbf{k}$, $\mathbf{x}\in\mathbb{R}^n$ by

$$F(\mathbf{x}) = \underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{n} f(\mathbf{k})e^{-2\pi i\mathbf{k}\cdot\mathbf{x}}\,d^n\mathbf{k} \tag{42}$$

$$f(\mathbf{k}) = \underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{n} F(\mathbf{x})e^{2\pi i\mathbf{k}\cdot\mathbf{x}}\,d^n\mathbf{x}. \tag{43}$$

see also AUTOCORRELATION, CONVOLUTION, DISCRETE FOURIER TRANSFORM, FAST FOURIER TRANSFORM, FOURIER SERIES, FOURIER-STIELTJES TRANSFORM, HANKEL TRANSFORM, HARTLEY TRANSFORM, INTEGRAL TRANSFORM, LAPLACE TRANSFORM, STRUCTURE FACTOR, WINOGRAD TRANSFORM

References

Arfken, G. "Development of the Fourier Integral," "Fourier Transforms—Inversion Theorem," and "Fourier Transform of Derivatives." §15.2–15.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 794–810, 1985.

Blackman, R. B. and Tukey, J. W. *The Measurement of Power Spectra, From the Point of View of Communications Engineering.* New York: Dover, 1959.

Bracewell, R. *The Fourier Transform and Its Applications.* New York: McGraw-Hill, 1965.

Brigham, E. O. *The Fast Fourier Transform and Applications.* Englewood Cliffs, NJ: Prentice Hall, 1988.

James, J. F. *A Student's Guide to Fourier Transforms with Applications in Physics and Engineering.* New York: Cambridge University Press, 1995.

Körner, T. W. *Fourier Analysis.* Cambridge, England: Cambridge University Press, 1988.

Morrison, N. *Introduction to Fourier Analysis.* New York: Wiley, 1994.

Morse, P. M. and Feshbach, H. "Fourier Transforms." §4.8 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 453–471, 1953.

Papoulis, A. *The Fourier Integral and Its Applications.* New York: McGraw-Hill, 1962.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge, England: Cambridge University Press, 1989.

Sansone, G. "The Fourier Transform." §2.13 in *Orthogonal Functions, rev. English ed.* New York: Dover, pp. 158–168, 1991.

Sneddon, I. N. *Fourier Transforms.* New York: Dover, 1995.

Sogge, C. D. *Fourier Integrals in Classical Analysis.* New York: Cambridge University Press, 1993.

Spiegel, M. R. *Theory and Problems of Fourier Analysis with Applications to Boundary Value Problems.* New York: McGraw-Hill, 1974.

Strichartz, R. *Fourier Transforms and Distribution Theory.* Boca Raton, FL: CRC Press, 1993.

Titchmarsh, E. C. *Introduction to the Theory of Fourier Integrals, 3rd ed.* Oxford, England: Clarendon Press, 1948.

Tolstov, G. P. *Fourier Series.* New York: Dover, 1976.

Walker, J. S. *Fast Fourier Transforms, 2nd ed.* Boca Raton, FL: CRC Press, 1996.

Fourier Transform—1

The FOURIER TRANSFORM of the CONSTANT FUNCTION $f(x)=1$ is given by

$$\mathcal{F}[1] = \int_{-\infty}^{\infty} e^{-2\pi i k x}\,dx = \delta(k),$$

according to the definition of the DELTA FUNCTION.

see also DELTA FUNCTION

Fourier Transform—$1/x$

The FOURIER TRANSFORM of the function $1/x$ is given by

$$\begin{aligned}\mathcal{F}\left(-\frac{1}{\pi x}\right) &= -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{e^{-2\pi ikx}}{x}\,dx\\ &= PV\int_{-\infty}^{\infty}\frac{\cos(2\pi kx)-i\sin(2\pi kx)}{x}\,dx\\ &= \begin{cases}-\frac{2i}{\pi}\int_0^\infty \frac{\sin(2\pi kx)}{x}\,dx & \text{for } k<0\\ \frac{2i}{\pi}\int_0^\infty \frac{\sin(2\pi kx)}{x}\,dx & \text{for } k>0\end{cases}\\ &= \begin{cases}-i & \text{for } k<0\\ i & \text{for } k>0,\end{cases}\end{aligned} \qquad (1)$$

which can also be written as the single equation

$$\mathcal{F}\left(-\frac{1}{\pi x}\right) = i[1-2H(-k)], \qquad (2)$$

where $H(x)$ is the HEAVISIDE STEP FUNCTION. The integrals follow from the identity

$$\begin{aligned}\int_0^\infty \frac{\sin(2\pi kx)}{x}\,dx &= \int_0^\infty \frac{\sin(2\pi kx)}{2\pi kx}\,d(2\pi kx)\\ &= \int_0^\infty \operatorname{sinc} z\,dz = \tfrac{1}{2}\pi.\end{aligned} \qquad (3)$$

Fourier Transform—Cosine

$$\begin{aligned}\mathcal{F}[\cos(2\pi k_0x)] &= \int_{-\infty}^{\infty} e^{-2\pi ikx}\left(\frac{e^{2\pi ik_0x}+e^{-2\pi ik_0x}}{2}\right)dx\\ &= \tfrac{1}{2}\int_{-\infty}^{\infty}[e^{-2\pi i(k-k_0)x}+e^{-2\pi i(k+k_0)x}]\,dx\\ &= \tfrac{1}{2}[\delta(k-k_0)+\delta(k+k_0)],\end{aligned}$$

where $\delta(x)$ is the DELTA FUNCTION.

see also COSINE, FOURIER TRANSFORM—SINE

Fourier Transform—Delta Function

The FOURIER TRANSFORM of the DELTA FUNCTION is given by

$$\mathcal{F}[\delta(x-x_0)] = \int_{-\infty}^{\infty}\delta(x-x_0)e^{-2\pi ikx}\,dx = e^{-2\pi ikx_0}.$$

see also DELTA FUNCTION

Fourier Transform—Exponential Function

The FOURIER TRANSFORM of $e^{-k_0|x|}$ is given by

$$\begin{aligned}\mathcal{F}[e^{-k_0|x|}] &= \int_{-\infty}^{\infty} e^{-k_0|x|}e^{-2\pi ikx}\,dx\\ &= \int_{-\infty}^{0} e^{-2\pi ikx}e^{2\pi xk_0}\,dx + \int_0^\infty e^{-2\pi ikx}e^{-2\pi k_0x}\,dx\\ &= \int_{-\infty}^{0}[\cos(2\pi kx)-i\sin(2\pi kx)]e^{2\pi k_0x}\,dx\\ &\quad + \int_0^\infty [\cos(2\pi kx)-i\sin(2\pi kx)]e^{-2\pi k_0x}\,dx.\end{aligned} \qquad (1)$$

Now let $u \equiv -x$ so $du = -dx$, then

$$\begin{aligned}\mathcal{F}[e^{-k_0|x|}] &= \int_0^\infty [\cos(2\pi ku)+i\sin(2\pi ku)]e^{-2\pi k_0u}\,du]\\ &\quad + \int_0^\infty [\cos(2\pi ku)-i\sin(2\pi ku)]e^{-2\pi k_0u}\,du]\\ &= 2\int_0^\infty \cos(2\pi ku)e^{-2\pi k_0u}\,du,\end{aligned} \qquad (2)$$

which, from the DAMPED EXPONENTIAL COSINE INTEGRAL, gives

$$\mathcal{F}[e^{-2\pi k_0|x|}] = \frac{1}{\pi}\frac{k_0}{k^2+{k_0}^2}, \qquad (3)$$

which is a LORENTZIAN FUNCTION.

see also DAMPED EXPONENTIAL COSINE INTEGRAL, EXPONENTIAL FUNCTION, LORENTZIAN FUNCTION

Fourier Transform—Gaussian

The FOURIER TRANSFORM of a GAUSSIAN FUNCTION $f(x) \equiv e^{-ax^2}$ is given by

$$\begin{aligned}F(k) &= \int_{-\infty}^{\infty} e^{-ax^2}e^{ikx}\,dx\\ &= \int_{-\infty}^{\infty} e^{-ax^2}[\cos(kx)+i\sin(kx)]\,dx\\ &= \int_{-\infty}^{\infty} e^{-ax^2}\cos(kx)\,dx + i\int_{-\infty}^{\infty} e^{-ax^2}\sin(kx)\,dx.\end{aligned}$$

The second integrand is EVEN, so integration over a symmetrical range gives 0. The value of the first integral is given by Abramowitz and Stegun (1972, p. 302, equation 7.4.6)

$$F(k) = \sqrt{\frac{\pi}{a}}\,e^{-k^2/4a},$$

so a GAUSSIAN transforms to a GAUSSIAN.

see also GAUSSIAN FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, 1972.

Fourier Transform—Heaviside Step Function

$$\mathcal{F}[H(x)] = \int_{-\infty}^{\infty} e^{-2\pi ikx} H(x)\, dx = \frac{1}{2}\left[\delta(k) - \frac{i}{\pi k}\right],$$

where $H(x)$ is the Heaviside Step Function and $\delta(k)$ is the Delta Function.

see also Heaviside Step Function

Fourier Transform—Lorentzian Function

$$\mathcal{F}\left[\frac{1}{\pi}\frac{\frac{1}{2}\Gamma}{(x-x_0)^2 + (\frac{1}{2}\Gamma)^2}\right] = e^{-2\pi ikx_0 - \Gamma\pi|k|}.$$

see also Lorentzian Function

Fourier Transform—Ramp Function

Let $R(x)$ be the Ramp Function, then the Fourier Transform of $R(x)$ is given by

$$\mathcal{F}[R(x)] = \int_{-\infty}^{\infty} e^{-2\pi ikx} R(x)\, dx = \pi i\delta'(2\pi k) - \frac{1}{4\pi^2 k^2},$$

where $\delta'(x)$ is the Derivative of the Delta Function.

see also Ramp Function

Fourier Transform—Rectangle Function

Let $\Pi(x)$ be the Rectangle Function, then the Fourier Transform is

$$\mathcal{F}[\Pi(x)] = \operatorname{sinc}(\pi k),$$

where $\operatorname{sinc}(x)$ is the Sinc Function.

see also Rectangle Function, Sinc Function

Fourier Transform—Sine

$$\begin{aligned}\mathcal{F}[\sin(2\pi k_0 x)] &= \int_{-\infty}^{\infty} e^{-2\pi ik_0x}\left(\frac{e^{2\pi i\nu_0 t} - e^{-2\pi ik_0x}}{2i}\right) dx\\ &= \tfrac{1}{2}i\int_{-\infty}^{\infty}[-e^{-2\pi i(k-k_0)x} + e^{-2\pi i(k+k_0)x}]\, dt\\ &= \tfrac{1}{2}i[\delta(k+k_0) - \delta(k-k_0)],\end{aligned}$$

where $\delta(x)$ is the Delta Function.

see also Fourier Transform—Cosine, Sine

Fox's *H*-Function

A very general function defined by

$$\begin{aligned}H(z) &= \mathbf{H}_{p,q}^{m,n}\left[z\left|\begin{matrix}(a_1,\alpha_1),\ldots,(a_p,\alpha_p)\\(b_1,\beta_1),\ldots,(b_p,\beta_p)\end{matrix}\right.\right]\\ &= \frac{1}{2\pi i}\int_C \frac{\prod_{j=1}^{m}\Gamma(b_j-\beta_i s)\prod_{j=1}^{n}\Gamma(1-a_j+\alpha_j s)}{\prod_{j=m+1}^{q}\Gamma(1-b_j+\beta_j s)\prod_{j=n+1}^{qp}\Gamma(a_j-\alpha_j s)} z^s\, ds,\end{aligned}$$

where $0 \le m \le q$, $0 \le n \le p$, $\alpha_j, \beta_j > 0$, and a_j, b_j are Complex Numbers such that the pole of $\Gamma(b_j - \beta_j s)$ for $j = 1, 2, \ldots, m$ coincides with any Pole of $\Gamma(1 - a_j + \alpha_j s)$ for $j = 1, 2, \ldots, n$. In addition C, is a Contour in the complex s-plane from $\omega - i\infty$ to $\omega + i\infty$ such that $(b_j + k)/\beta_j$ and $(a_j - 1 - k)/\alpha_j$ lie to the right and left of C, respectively.

see also MacRobert's *E*-Function, Meijer's *G*-Function

References

Carter, B. D. and Springer, M. D. "The Distribution of Products, Quotients, and Powers of Independent *H*-Functions." *SIAM J. Appl. Math.* **33**, 542–558, 1977.

Fox, C. "The *G* and *H*-Functions as Symmetrical Fourier Kernels." *Trans. Amer. Math. Soc.* **98**, 395–429, 1961.

Frac

see Fractional Part

Fractal

An object or quantity which displays Self-Similarity, in a somewhat technical sense, on all scales. The object need not exhibit *exactly* the same structure at all scales, but the same "type" of structures must appear on all scales. A plot of the quantity on a log-log graph versus scale then gives a straight line, whose slope is said to be the Fractal Dimension. The prototypical example for a fractal is the length of a coastline measured with different length Rulers. The shorter the Ruler, the longer the length measured, a Paradox known as the Coastline Paradox.

see also Backtracking, Barnsley's Fern, Box Fractal, Butterfly Fractal, Cactus Fractal, Cantor Set, Cantor Square Fractal, Carotid-Kundalini Fractal, Cesàro Fractal, Chaos Game, Circles-and-Squares Fractal, Coastline Paradox, Dragon Curve, Fat Fractal, Fatou Set, Flowsnake Fractal, Fractal Dimension, H-Fractal, Hénon Map, Iterated Function System, Julia Fractal, Kaplan-Yorke Map, Koch Antisnowflake, Koch Snowflake, Lévy Fractal, Lévy Tapestry, Lindenmayer System, Mandelbrot Set, Mandelbrot Tree, Menger Sponge, Minkowski Sausage, Mira Fractal, Newton's Method, Pentaflake, Pythagoras Tree, Rabinovich-Fabrikant Equation, San Marco Fractal, Sierpiński Carpet, Sierpiński Curve, Sierpiński Sieve, Star Fractal, Zaslavskii Map

References

Barnsley, M. F. and Rising, H. *Fractals Everywhere, 2nd ed.* Boston, MA: Academic Press, 1993.

Bogomolny, A. "Fractal Curves and Dimension." `http://www.cut-the-knot.com/do_you_know/dimension.html`.

Brandt, C.; Graf, S.; and Zähle, M. (Eds.). *Fractal Geometry and Stochastics.* Boston, MA: Birkhäuser, 1995.

Bunde, A. and Havlin, S. (Eds.). *Fractals and Disordered Systems, 2nd ed.* New York: Springer-Verlag, 1996.

Bunde, A. and Havlin, S. (Eds.). *Fractals in Science.* New York: Springer-Verlag, 1994.

Devaney, R. L. *Complex Dynamical Systems: The Mathematics Behind the Mandelbrot and Julia Sets.* Providence, RI: Amer. Math. Soc., 1994.

Devaney, R. L. and Keen, L. *Chaos and Fractals: The Mathematics Behind the Computer Graphics.* Providence, RI: Amer. Math. Soc., 1989.

Edgar, G. A. *Classics on Fractals.* Reading, MA: Addison-Wesley, 1994.

Eppstein, D. "Fractals." `http://www.ics.uci.edu/~eppstein/junkyard/fractal.html`.

Falconer, K. J. *The Geometry of Fractal Sets, 1st pbk. ed., with corr.* Cambridge, England Cambridge University Press, 1986.

Feder, J. *Fractals.* New York: Plenum Press, 1988.

Giffin, N. "The Spanky Fractal Database." `http://spanky.triumf.ca/www/welcome1.html`.

Hastings, H. M. and Sugihara, G. *Fractals: A User's Guide for the Natural Sciences.* New York: Oxford University Press, 1994.

Kaye, B. H. *A Random Walk Through Fractal Dimensions, 2nd ed.* New York: Wiley, 1994.

Lauwerier, H. A. *Fractals: Endlessly Repeated Geometrical Figures.* Princeton, NJ: Princeton University Press, 1991.

Mandelbrot, B. B. *Fractals: Form, Chance, & Dimension.* San Francisco, CA: W. H. Freeman, 1977.

Mandelbrot, B. B. *The Fractal Geometry of Nature.* New York: W. H. Freeman, 1983.

Massopust, P. R. *Fractal Functions, Fractal Surfaces, and Wavelets.* San Diego, CA: Academic Press, 1994.

Pappas, T. "Fractals—Real or Imaginary." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 78–79, 1989.

Peitgen, H.-O.; Jürgens, H.; and Saupe, D. *Chaos and Fractals: New Frontiers of Science.* New York: Springer-Verlag, 1992.

Peitgen, H.-O. and Richter, D. H. *The Beauty of Fractals: Images of Complex Dynamical Systems.* New York: Springer-Verlag, 1986.

Peitgen, H.-O. and Saupe, D. (Eds.). *The Science of Fractal Images.* New York: Springer-Verlag, 1988.

Pickover, C. A. (Ed.). *The Pattern Book: Fractals, Art, and Nature.* World Scientific, 1995.

Pickover, C. A. (Ed.). *Fractal Horizons: The Future Use of Fractals.* New York: St. Martin's Press, 1996.

Rietman, E. *Exploring the Geometry of Nature: Computer Modeling of Chaos, Fractals, Cellular Automata, and Neural Networks.* New York: McGraw-Hill, 1989.

Russ, J. C. *Fractal Surfaces.* New York: Plenum, 1994.

Schroeder, M. *Fractals, Chaos, Power Law: Minutes from an Infinite Paradise.* New York: W. H. Freeman, 1991.

Sprott, J. C. "Sprott's Fractal Gallery." `http://sprott.physics.wisc.edu/fractals.htm`.

Stauffer, D. and Stanley, H. E. *From Newton to Mandelbrot, 2nd ed.* New York: Springer-Verlag, 1995.

Stevens, R. T. *Fractal Programming in C.* New York: Henry Holt, 1989.

Takayasu, H. *Fractals in the Physical Sciences.* Manchester, England: Manchester University Press, 1990.

Taylor, M. C. "sci.fractals FAQ." `http://www.mta.ca/~mctaylor/sci.fractals-faq`.

Tricot, C. *Curves and Fractal Dimension.* New York: Springer-Verlag, 1995.

Triumf Mac Fractal Programs. `http://spanky.triumf.ca/pub/fractals/programs/MAC/`.

Vicsek, T. *Fractal Growth Phenomena, 2nd ed.* Singapore: World Scientific, 1992.

Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m`.

Yamaguti, M.; Hata, M.; and Kigami, J. *Mathematics of Fractals.* Providence, RI: Amer. Math. Soc., 1997.

Fractal Dimension

The term "fractal dimension" is sometimes used to refer to what is more commonly called the CAPACITY DIMENSION (which is, roughly speaking, the exponent D in the expression $n(\epsilon) = \epsilon^{-D}$, where $n(\epsilon)$ is the minimum number of OPEN SETS of diameter ϵ needed to cover the set). However, it can more generally refer to any of the dimensions commonly used to characterize fractals (e.g., CAPACITY DIMENSION, CORRELATION DIMENSION, INFORMATION DIMENSION, LYAPUNOV DIMENSION, MINKOWSKI-BOULIGAND DIMENSION).

see also BOX COUNTING DIMENSION, CAPACITY DIMENSION, CORRELATION DIMENSION, FRACTAL DIMENSION, HAUSDORFF DIMENSION, INFORMATION DIMENSION, LYAPUNOV DIMENSION, MINKOWSKI-BOULIGAND DIMENSION, POINTWISE DIMENSION, q-DIMENSION

References

Rasband, S. N. "Fractal Dimension." Ch. 4 in *Chaotic Dynamics of Nonlinear Systems.* New York: Wiley, pp. 71–83, 1990.

Fractal Land

see CAROTID-KUNDALINI FRACTAL

Fractal Process

A 1-D MAP whose increments are distributed according to a NORMAL DISTRIBUTION. Let $y(t-\Delta t)$ and $y(t+\Delta t)$ be values, then their correlation is given by the BROWN FUNCTION

$$r = 2^{2H-1} - 1.$$

When $H = 1/2$, $r = 0$ and the fractal process corresponds to 1-D Brownian motion. If $H > 1/2$, then $r > 0$ and the process is called a PERSISTENT PROCESS. If $H < 1/2$, then $r < 0$ and the process is called an ANTIPERSISTENT PROCESS.

see also ANTIPERSISTENT PROCESS, PERSISTENT PROCESS

References

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, 1993.

Fractal Sequence

Given an INFINITIVE SEQUENCE $\{x_n\}$ with associated array $a(i,j)$, then $\{x_n\}$ is said to be a fractal sequence

1. If $i+1 = x_n$, then there exists $m < n$ such that $i = x_m$,
2. If $h < i$, then, for every j, there is exactly one k such that $a(i,j) < a(h,k) < a(i,j+1)$.

(As i and j range through N, the array $A = a(i,j)$, called the associative array of x, ranges through all of N.) An example of a fractal sequence is 1, 1, 1, 1, 2, 1, 2, 1, 3, 2, 1, 3, 2, 1, 3, ...

If $\{x_n\}$ is a fractal sequence, then the associated array is an INTERSPERSION. If x is a fractal sequence, then the UPPER-TRIMMED SUBSEQUENCE is given by $\lambda(x) = x$, and the LOWER-TRIMMED SUBSEQUENCE $V(x)$ is another fractal sequence. The SIGNATURE of an IRRATIONAL NUMBER is a fractal sequence.

see also INFINITIVE SEQUENCE

References

Kimberling, C. "Fractal Sequences and Interspersions." *Ars Combin.* **45**, 157–168, 1997.

Fractal Valley

see CAROTID-KUNDALINI FUNCTION

Fraction

A RATIONAL NUMBER expressed in the form a/b, where a is called the NUMERATOR and b is called the DENOMINATOR. A PROPER FRACTION is a fraction such that $a/b < 1$, and a LOWEST TERMS FRACTION is a fraction with common terms canceled out of the NUMERATOR and DENOMINATOR.

The Egyptians expressed their fractions as sums (and differences) of UNIT FRACTIONS. Conway and Guy (1999) give a table of Roman NOTATION for fractions, in which multiples of 1/12 (the UNCIA) were given separate names.

see also ADJACENT FRACTION, ANOMALOUS CANCELLATION, CONTINUED FRACTION, DENOMINATOR, EGYPTIAN FRACTION, FAREY SEQUENCE, GOLDEN RULE, HALF, LOWEST TERMS FRACTION, MEDIANT, NUMERATOR, PROPER FRACTION, PYTHAGOREAN FRACTION, QUARTER, RATIONAL NUMBER, UNIT FRACTION

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 22–23, 1996.

Courant, R. and Robbins, H. "Decimal Fractions. Infinite Decimals." §2.2.2 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 61–63, 1996.

Fractional Calculus

Denote the nth DERIVATIVE D^n and the n-fold INTEGRAL D^{-n}. Then

$$D^{-1}f(t) = \int_0^t f(\xi)\,d\xi. \tag{1}$$

Now, if

$$D^{-n}f(t) = \frac{1}{(n-1)!}\int_0^t (t-\xi)^{n-1}f(\xi)\,d\xi \tag{2}$$

is true for n, then

$$\begin{aligned} D^{-(n+1)}f(t) &= D^{-1}\left[\frac{1}{(n-1)!}\int_0^t (t-\xi)^{n-1}f(\xi)\,d\xi\right] \\ &= \int_0^t \left[\frac{1}{(n-1)!}\int_0^x (x-\xi)^{n-1}f(\xi)\,d\xi\right]dx. \end{aligned} \tag{3}$$

Interchanging the order of integration gives

$$D^{-(n+1)}f(t) = \frac{1}{n!}\int_0^t (t-\xi)^n f(\xi)\,d\xi. \tag{4}$$

But (2) is true for $n = 1$, so it is also true for all n by INDUCTION. The fractional integral of $f(t)$ can then be defined by

$$D^{-\nu}f(t) = \frac{1}{\Gamma(\nu)}\int_0^t (t-\xi)^{\nu-1}f(\xi)\,d\xi, \tag{5}$$

where $\Gamma(\nu)$ is the GAMMA FUNCTION.

The fractional integral can only be given in terms of elementary functions for a small number of functions. For example,

$$D^{-\nu}t^{-\lambda} = \frac{\Gamma(\lambda+1)}{\Gamma(\lambda+\nu+1)}t^{\lambda+\nu} \quad \text{for } \lambda > -1, \nu > 0 \tag{6}$$

$$D^{-\nu}e^{at} = \frac{1}{\Gamma(\nu)}e^{at}\int_0^t x^{\nu-1}e^{-ax}\,dx \equiv E_t(\nu,a), \tag{7}$$

where $E_t(\nu,a)$ is the E_t-FUNCTION. The fractional derivative of f (if it exists) can be defined by

$$D^{\mu}f(t) = D^m[D^{-(m-\mu)}f(t)]. \tag{8}$$

An example is

$$\begin{aligned} D^{\mu}t^{\lambda} &= \frac{\Gamma(\lambda+1)}{\Gamma(\lambda+m-\mu+1)} \\ &= \frac{\Gamma(\lambda+1)}{\Gamma(\lambda-\mu+1)}t^{\lambda-\mu} \quad \text{for } \lambda > -1, \mu > 0 \end{aligned} \tag{9}$$

$$D^{\rho}E_t(\nu,a) = E_t(\nu-\rho,a) \qquad \text{for } \nu > 0, \rho \neq 0. \tag{10}$$

It is always true that, for $\mu, \nu > 0$,

$$D^{-\mu}D^{-\nu}f(t) = D^{-(\mu+\nu)}, \qquad (11)$$

but *not* always true that

$$D^{\mu}D^{\nu} = D^{\mu+\nu}. \qquad (12)$$

see also DERIVATIVE, INTEGRAL

References

Love, E. R. "Fractional Derivatives of Imaginary Order." *J. London Math. Soc.* **3**, 241–259, 1971.

McBride, A. C. *Fractional Calculus.* New York: Halsted Press, 1986.

Miller, K. S. "Derivatives of Noninteger Order." *Math. Mag.* **68**, 183–192, 1995.

Nishimoto, K. *Fractional Calculus.* New Haven, CT: University of New Haven Press, 1989.

Spanier, J. and Oldhan, K. B. *The Fractional Calculus.* New York: Academic Press, 1974.

Fractional Derivative

see FRACTIONAL CALCULUS

Fractional Differential Equation

The solution to the differential equation

$$[D^{2v} + aD^{v} + bD^{0}]y(t) = 0$$

is

$$y(t) = \begin{cases} e_\alpha(t) - e_\beta(t) \\ \quad \text{for } \alpha \neq \beta \\ te^{\alpha t}, \sum_{k=-(q-1)}^{q-1} \alpha^k (q - |k|) D^{1-(k+1)v}(te^{\alpha^q t}) \\ \quad \text{for } \alpha = \beta \neq 0 \\ \frac{t^{2\nu-1}}{\Gamma(2v)} \\ \quad \text{for } \alpha = \beta = 0, \end{cases}$$

where

$$q = \frac{1}{v}$$

$$e_\beta(t) = \sum_{k=0}^{q-1} \beta^{q-k-1} E_t(-kv, \beta^q),$$

$E_t(a,x)$ is the E_t-FUNCTION, and $\Gamma(n)$ is the GAMMA FUNCTION.

References

Miller, K. S. "Derivatives of Noninteger Order." *Math. Mag.* **68**, 183–192, 1995.

Fractional Fourier Transform

A z-TRANSFORM with

$$z \equiv e^{2\pi i\alpha/N}$$

for $\alpha \neq \pm 1$. This transform can be used to detect frequencies which are not INTEGER multiples of the lowest DISCRETE FOURIER TRANSFORM frequency.

see also z-TRANSFORM

References

Graham, R. L.; Knuth, D. E.; and Patashnik, O. *Concrete Mathematics, 2nd ed.* Reading, MA: Addison-Wesley, 1994.

Fractional Integral

see FRACTIONAL CALCULUS

Fractional Part

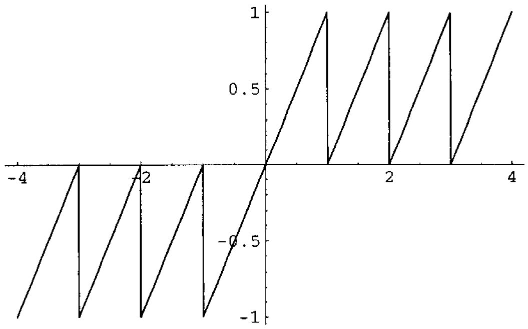

The function giving the fractional (nonintegral) part of a number and defined as

$$\operatorname{frac}(x) \equiv \begin{cases} x - \lfloor x \rfloor & x \geq 0 \\ x - \lfloor x \rfloor - 1 & x \leq 0, \end{cases}$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION.

see also CEILING FUNCTION, FLOOR FUNCTION, NINT, ROUND, TRUNCATE, WHOLE NUMBER

References

Spanier, J. and Oldham, K. B. "The Integer-Value Int(x) and Fractional-Value frac(x) Functions." Ch. 9 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 71–78, 1987.

Fractran

Fractran is an algorithm applied to a given list f_1, f_2, ..., f_k of FRACTIONS. Given a starting INTEGER N, the Fractran algorithm proceeds by repeatedly multiplying the integer at a given stage by the first element f_i given an integer PRODUCT. The algorithm terminates when there is no such f_i.

The list

$$\frac{17}{91}, \frac{78}{85}, \frac{19}{51}, \frac{23}{38}, \frac{29}{33}, \frac{77}{29}, \frac{95}{23}, \frac{77}{19}, \frac{1}{17}, \frac{11}{13}, \frac{13}{11}, \frac{15}{2}, \frac{1}{7}, \frac{55}{1}$$

with starting integer $N = 2$ generates a sequence 2, 15, 825, 725, 1925, 2275, 425, 390, 330, 290, 770, ... Conway (1987) showed that the only other powers of 2 which occur are those with PRIME exponent: 2^2, 2^3, 2^5, 2^7, ...

References

Conway, J. H. "Unpredictable Iterations." In *Proc. Number Theory Conf., Boulder, CO,* pp. 49–52, 1972.

Conway, J. H. "Fractran: A Simple Universal Programming Language for Arithmetic." Ch. 2 in *Open Problems in Communication and Computation* (Ed. T. M. Cover and B. Gopinath). New York: Springer-Verlag, pp. 4–26, 1987.

Framework

Consider a finite collection of points $p = (p_1, \ldots, p_n)$, $p_i \in \mathbb{R}^d$ EUCLIDEAN SPACE (known as a CONFIGURATION) and a graph G whose VERTICES correspond to pairs of points that are constrained to stay the same distance apart. Then the graph G together with the configuration p, denoted $G(p)$, is called a framework.

see also BAR (EDGE), CONFIGURATION, RIGID

Franklin Magic Square

52	61	4	13	20	29	36	45
14	3	62	51	46	35	30	19
53	60	5	12	21	28	37	44
11	6	59	54	43	38	27	22
55	58	7	10	23	26	39	42
9	8	57	56	41	40	25	24
50	63	2	15	18	31	34	47
16	1	64	49	48	33	32	17

Benjamin Franklin constructed the above 8×8 PANMAGIC SQUARE having MAGIC CONSTANT 260. Any half-row or half-column in this square totals 130, and the four corners plus the middle total 260. In addition, bent diagonals (such as 52-3-5-54-10-57-63-16) also total 260 (Madachy 1979, p. 87).

see also MAGIC SQUARE, PANMAGIC SQUARE

References

Madachy, J. S. "Magic and Antimagic Squares." Ch. 4 in *Madachy's Mathematical Recreations.* New York: Dover, pp. 103–113, 1979.

Pappas, T. "The Magic Square of Benjamin Franklin." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 97, 1989.

Fransén-Robinson Constant

$$F \equiv \int_0^\infty \frac{dx}{\Gamma(x)} = 2.8077702420\ldots,$$

where $\Gamma(x)$ is the GAMMA FUNCTION. The above plots show the functions $\Gamma(x)$ and $1/\Gamma(x)$.

see also GAMMA FUNCTION

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/fran/fran.html`.

Fransén, A. "Accurate Determination of the Inverse Gamma Integral." *BIT* **19**, 137–138, 1979.

Fransén, A. "Addendum and Corrigendum to 'High-Precision Values of the Gamma Function and of Some Related Coefficients.'" *Math. Comput.* **37**, 233–235, 1981.

Fransén, A. and Wrigge, S. "High-Precision Values of the Gamma Function and of Some Related Coefficients." *Math. Comput.* **34**, 553–566, 1980.

Plouffe, S. "Fransen-Robinson Constant." `http://lacim.uqam.ca/piDATA/fransen.txt`.

Fréchet Bounds

Any bivariate distribution function with marginal distribution functions F and G satisfies

$$\max\{F(x)+G(y)-1,0\} \leq H(x,y) \leq \min\{F(x),G(y)\}.$$

Fréchet Derivative

A function f is Fréchet differentiable at a if

$$\lim_{x \to a} \frac{f(x)-f(a)}{x-a}$$

exists. This is equivalent to the statement that ϕ has a removable DISCONTINUITY at a, where

$$\phi(x) \equiv \frac{f(x)-f(a)}{x-a}.$$

Every function which is Fréchet differentiable is also Carathéodory differentiable.

see also CARATHÉODORY DERIVATIVE, DERIVATIVE

Fréchet Space

A complete metrizable SPACE, sometimes also with the restriction that the space be locally convex.

Fredholm Integral Equation of the First Kind

An INTEGRAL EQUATION of the form

$$f(x) = \int_{-\infty}^{\infty} K(x,t)\phi(t)\,dt$$

$$\phi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{F(\omega)}{K(\omega)} e^{-i\omega x}\,d\omega.$$

see also FREDHOLM INTEGRAL EQUATION OF THE SECOND KIND, INTEGRAL EQUATION, VOLTERRA INTEGRAL EQUATION OF THE FIRST KIND, VOLTERRA INTEGRAL EQUATION OF THE SECOND KIND

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 865, 1985.

Fredholm Integral Equation of the Second Kind

An INTEGRAL EQUATION of the form

$$\phi(x) = f(x) + \lambda \int_{-\infty}^{\infty} K(x,t)\phi(t)\,dt$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{F(t)e^{-ixt}\,dt}{1-\sqrt{2\pi}\,\lambda K(t)}.$$

see also FREDHOLM INTEGRAL EQUATION OF THE FIRST KIND, INTEGRAL EQUATION, NEUMANN SERIES (INTEGRAL EQUATION), VOLTERRA INTEGRAL

EQUATION OF THE FIRST KIND, VOLTERRA INTEGRAL EQUATION OF THE SECOND KIND

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 865, 1985.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Fredholm Equations of the Second Kind." §18.1 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 782–785, 1992.

Free

When referring to a planar object, "free" means that the object is regarded as capable of being picked up out of the plane and flipped over. As a result, MIRROR IMAGES are equivalent for free objects.

A free abstract mathematical object is generated by n elements in a "free manner," i.e., such that the n elements satisfy no nontrivial relations among themselves. To make this more formal, an algebraic GADGET X is freely generated by a SUBSET G if, for any function $f : G \to Y$ where Y is any other algebraic GADGET, there exists a unique HOMOMORPHISM (which has different meanings depending on what kind of GADGETS you're dealing with) $g : X \to Y$ such that g restricted to G is f.

If the algebraic GADGETS are VECTOR SPACES, then G freely generates X IFF G is a BASIS for X. If the algebraic GADGETS are ABELIAN GROUPS, then G freely generates X IFF X is a DIRECT SUM of the INTEGERS, with G consisting of the standard BASIS.

see also FIXED, GADGET, MIRROR IMAGE, RANK

Free Group

The generators of a group G are defined to be the smallest subset of group elements such that all other elements of G can be obtained from them and their inverses. A GROUP is a free group if no relation exists between its generators (other than the relationship between an element and its inverse required as one of the defining properties of a group). For example, the additive group of whole numbers is free with a single generator, 1.

see also FREE SEMIGROUP

Free Semigroup

A SEMIGROUP with a noncommutative product in which no PRODUCT can ever be expressed more simply in terms of other ELEMENTS.

see also FREE GROUP, SEMIGROUP

Free Variable

An occurrence of a variable in a LOGIC FORMULA which is not inside the scope of a QUANTIFIER.

see also BOUND, SENTENCE

Freemish Crate

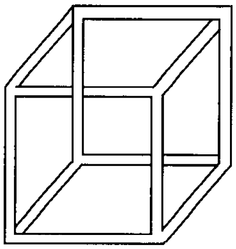

An IMPOSSIBLE FIGURE box which can be drawn but not built.

References

Fineman, M. *The Nature of Visual Illusion.* New York: Dover, p. 120–122, 1996.
Jablan, S. "Are Impossible Figures Possible?" `http://members.tripod.com/~modularity/kulpa.htm`.
Pappas, T. "The Impossible Tribar." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 13, 1989.

Freeth's Nephroid

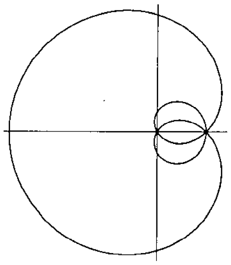

A STROPHOID of a CIRCLE with the POLE O at the CENTER of the CIRCLE and the fixed point P on the CIRCUMFERENCE of the CIRCLE. In a paper published by the London Mathematical Society in 1879, T. J. Freeth described it and various other STROPHOIDS (MacTutor Archive). If the line through P PARALLEL to the y-AXIS cuts the NEPHROID at A, then ANGLE AOP is $3\pi/7$, so this curve can be used to construct a regular HEPTAGON. The POLAR equation is

$$r = a[1 + 2\sin(\tfrac{1}{2}\theta)].$$

see also STROPHOID

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 175 and 177–178, 1972.
MacTutor History of Mathematics Archive. "Freeth's Nephroid." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Freeths.html`.

Freiman's Constant

The end of the last gap in the LAGRANGE SPECTRUM, given by

$$F \equiv \frac{2221564096 + 293748\sqrt{462}}{491993569} = 4.5278295661\ldots.$$

REAL NUMBERS greater than F are members of the MARKOV SPECTRUM.

see also LAGRANGE SPECTRUM, MARKOV SPECTRUM

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 188–189, 1996.

French Curve

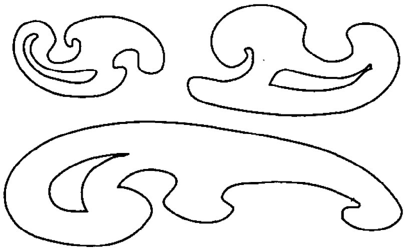

French curves are plastic (or wooden) templates having an edge composed of several different curves. French curves are used in drafting (or were before computer-aided design) to draw smooth curves of almost any desired curvature in mechanical drawings. Several typical French curves are illustrated above.

see also CORNU SPIRAL

Frenet Formulas

Also known as the SERRET-FRENET FORMULAS

$$\begin{bmatrix}\dot{\mathbf{T}}\\ \dot{\mathbf{N}}\\ \dot{\mathbf{B}}\end{bmatrix}=\begin{bmatrix}0 & \kappa & 0\\ -\kappa & 0 & \tau\\ 0 & -\tau & 0\end{bmatrix}\begin{bmatrix}\mathbf{T}\\ \mathbf{N}\\ \mathbf{B}\end{bmatrix},$$

where $\mathbf{T}$ is the unit TANGENT VECTOR, $\mathbf{N}$ is the unit NORMAL VECTOR, $\mathbf{B}$ is the unit BINORMAL VECTOR, τ is the TORSION, κ is the CURVATURE, and $\dot{\mathbf{x}}$ denotes $d\mathbf{x}/ds$.

see also CENTRODE, FUNDAMENTAL THEOREM OF SPACE CURVES, NATURAL EQUATION

References

Frenet, F. "Sur les courbes à double courbure." Thèse. Toulouse, 1847. Abstract in *J. de Math.* **17**, 1852.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 126, 1993.

Kreyszig, E. "Formulae of Frenet." §15 in *Differential Geometry.* New York: Dover, p. 40–43, 1991.

Serret, J. A. "Sur quelques formules relatives à la théorie des courbes à double courbure." *J. de Math.* **16**, 1851.

Frequency Curve

see GAUSSIAN FUNCTION

Fresnel's Elasticity Surface

A QUARTIC SURFACE given by

$$r=\sqrt{a^2x^2+b^2y^2+c^2z^2},$$

where

$$r^2\equiv x'^2+y'^2+z'^2,$$

also known as FRESNEL'S WAVE SURFACE. It was introduced by Fresnel in his studies of crystal optics.

References

Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, p. 16, 1986.

Fischer, G. (Ed.). Plates 38–39 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 38–39, 1986.

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 304, 1993.

Fresnel Integrals

In physics, the Fresnel integrals are most often defined by

$$\begin{aligned}C(u)+iS(u)&\equiv\int_0^u e^{i\pi x^2/2}\,dx\\ &=\int_0^u\cos(\tfrac{1}{2}\pi x^2)\,dx+i\int_0^u\sin(\tfrac{1}{2}\pi x^2)\,dx,\end{aligned}\qquad(1)$$

so

$$C(u)\equiv\int_0^u\cos(\tfrac{1}{2}\pi x^2)\,dx\qquad(2)$$

$$S(u)\equiv\int_0^u\sin(\tfrac{1}{2}\pi x^2)\,dx.\qquad(3)$$

They satisfy

$$C(\pm\infty)=-\tfrac{1}{2}\qquad(4)$$

$$S(\pm\infty)=\tfrac{1}{2}.\qquad(5)$$

Related functions are defined as

$$C_1(z)\equiv\sqrt{\frac{2}{\pi}}\int_0^x\cos t^2\,dt\qquad(6)$$

$$S_1(z)\equiv\sqrt{\frac{2}{\pi}}\int_0^x\sin t^2\,dt\qquad(7)$$

$$C_2(z)\equiv\frac{1}{\sqrt{2\pi}}\int\frac{\cos t}{\sqrt{t}}\,dt\qquad(8)$$

$$S_2(z)\equiv\frac{1}{\sqrt{2\pi}}\int\frac{\sin t}{\sqrt{t}}\,dt.\qquad(9)$$

An asymptotic expansion for $x\gg 1$ gives

$$C(u)\approx\frac{1}{2}+\frac{1}{\pi u}\sin(\tfrac{1}{2}\pi u^2)\qquad(10)$$

$$S(u)\approx\frac{1}{2}-\frac{1}{\pi u}\cos(\tfrac{1}{2}\pi u^2).\qquad(11)$$

Therefore, as $u\to\infty$, $C(u)=1/2$ and $S(u)=1/2$. The Fresnel integrals are sometimes alternatively defined as

$$x(t)=\int_0^t\cos(v^2)\,dv\qquad(12)$$

$$y(t)=\int_0^t\sin(v^2)\,dv.\qquad(13)$$

Letting $x \equiv v^2$ so $dx = 2v\,dv = 2\sqrt{x}\,dv$, and $dv = x^{-1/2}\,dx/2$

$$x(t) = \tfrac{1}{2}\int_0^{\sqrt{t}} x^{-1/2}\cos x\,dx \tag{14}$$

$$y(t) = \tfrac{1}{2}\int_0^{\sqrt{t}} x^{-1/2}\sin x\,dx. \tag{15}$$

In this form, they have a particularly simple expansion in terms of SPHERICAL BESSEL FUNCTIONS OF THE FIRST KIND. Using

$$j_0(x) = \frac{\sin x}{x} \tag{16}$$

$$n_1(x) = -j_{-1}(x) = -\frac{\cos x}{x}, \tag{17}$$

where $n_1(x)$ is a SPHERICAL BESSEL FUNCTION OF THE SECOND KIND

$$\begin{aligned} x(t^2) &= -\tfrac{1}{2}\int_0^t n_1(x)x^{1/2}\,dx \\ &= \tfrac{1}{2}\int_0^t j_{-1}(x)x^{1/2}\,dx = x^{1/2}\sum_{n=0}^{\infty} j_{2n}(x) \end{aligned} \tag{18}$$

$$\begin{aligned} y(t^2) &= \tfrac{1}{2}\int_0^t j_0(x)x^{1/2}\,dx \\ &= x^{1/2}\sum_{n=0}^{\infty} j_{2n+1}(x). \end{aligned} \tag{19}$$

see also CORNU SPIRAL

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Fresnel Integrals." §7.3 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 300–302, 1972.

Leonard, I. E. "More on Fresnel Integrals." *Amer. Math. Monthly* **95**, 431–433, 1988.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Fresnel Integrals, Cosine and Sine Integrals." §6.79 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 248–252, 1992.

Spanier, J. and Oldham, K. B. "The Fresnel Integrals $S(x)$ and $C(x)$." Ch. 39 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 373–383, 1987.

Fresnel's Wave Surface

see FRESNEL'S ELASTICITY SURFACE

Frey Curve

Let $a^p + b^p = c^p$ be a solution to FERMAT'S LAST THEOREM. Then the corresponding Frey curve is

$$y^2 = x(x - a^p)(x + b^p). \tag{1}$$

Frey showed that such curves cannot be MODULAR, so if the TANIYAMA-SHIMURA CONJECTURE were true, Frey curves couldn't exist and FERMAT'S LAST THEOREM would follow with b EVEN and $a \equiv -1 \pmod 4$. Frey curves are SEMISTABLE. Invariants include the DISCRIMINANT

$$(a^p - 0)^2(-b^p - 0)[a^p - (-b)^p]^2 = a^{2p}b^{2p}c^{2p}. \tag{2}$$

The MINIMAL DISCRIMINANT is

$$\Delta = 2^{-8}a^{2p}b^{2p}c^{2p}, \tag{3}$$

the CONDUCTOR is

$$N = \prod_{l|abc} l, \tag{4}$$

and the j-INVARIANT is

$$j = \frac{2^8(a^{2p} + b^{2p} + a^p b^p)^3}{a^{2p}b^{2p}c^{2p}} = \frac{2^8(c^{2p} - b^p c^p)^3}{(abc)^{2p}}. \tag{5}$$

see also ELLIPTIC CURVE, FERMAT'S LAST THEOREM, TANIYAMA-SHIMURA CONJECTURE

References

Cox, D. A. "Introduction to Fermat's Last Theorem." *Amer. Math. Monthly* **101**, 3–14, 1994.

Gouvêa, F. Q. "A Marvelous Proof." *Amer. Math. Monthly* **101**, 203–222, 1994.

Frey Elliptic Curve

see FREY CURVE

Friday the Thirteenth

The Gregorian calendar follows a pattern of leap years which repeats every 400 years. There are 4,800 months in 400 years, so the 13th of the month occurs 4,800 times in this interval. The number of times the 13th occurs on each weekday is given in the table below. As shown by Brown (1933), the thirteenth of the month is slightly more likely to be on a Friday than on any other day.

Day	Number of 13s	Fraction
Sunday	687	14.31%
Monday	685	14.27%
Tuesday	685	14.27%
Wednesday	687	14.31%
Thursday	684	14.25%
Friday	688	14.33%
Saturday	684	14.25%

see also 13, WEEKDAY

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 27, 1987.

Brown, B. H. "Solution to Problem E36." *Amer. Math. Monthly* **40**, 607, 1933.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 14–15, 1992.

Friend

A friend of a number n is another number m such that (m, n) is a FRIENDLY PAIR.

see also FRIENDLY PAIR, SOLITARY NUMBER

References

Anderson, C. W. and Hickerson, D. Problem 6020. "Friendly Integers." *Amer. Math. Monthly* **84**, 65–66, 1977.

Friendly Giant Group

see MONSTER GROUP

Friendly Pair

Define

$$\Sigma(n) \equiv \frac{\sigma(n)}{n},$$

where $\sigma(n)$ is the DIVISOR FUNCTION. Then a PAIR of distinct numbers (k, m) is a friendly pair (and k is said to be a FRIEND of m) if

$$\Sigma(k) = \Sigma(m).$$

For example, 4320 and 4680 are a friendly pair, since $\sigma(4320) = 15120$, $\sigma(4680) = 16380$, and

$$\begin{aligned} \Sigma(4320) &\equiv \tfrac{15120}{4320} = \tfrac{7}{2} \\ \Sigma(4680) &\equiv \tfrac{16380}{4680} = \tfrac{7}{2}. \end{aligned}$$

Numbers which do not have FRIENDS are called SOLITARY NUMBERS. SOLITARY NUMBERS satisfy $(\sigma(n), n) = 1$, where (a, b) is the GREATEST COMMON DIVISOR of a and b.

see also ALIQUOT SEQUENCE, FRIEND, SOLITARY NUMBER

References

Anderson, C. W. and Hickerson, D. Problem 6020. "Friendly Integers." *Amer. Math. Monthly* **84**, 65–66, 1977.

Frieze Pattern

$$\begin{array}{ccc} & b & \\ a & & d \\ & c & \end{array}$$

An arrangement of numbers at the intersection of two sets of perpendicular diagonals such that $a+d = b+c+1$ (for an additive frieze pattern) or $ad = bc + 1$ (for a multiplicative frieze pattern) in each diamond.

References

Conway, J. H. and Coxeter, H. S. M. "Triangulated Polygons and Frieze Patterns." *Math. Gaz.* **57**, 87–94, 1973.

Conway, J. H. and Guy, R. K. In *The Book of Numbers.* New York: Springer-Verlag, pp. 74–76 and 96–97, 1996.

Frobenius-König Theorem

The PERMANENT of an $n \times n$ MATRIX with all entries either 0 or 1 is 0 IFF the MATRIX contains an $r \times s$ submatrix of 0s with $r + s = n + 1$. This result follows from the KÖNIG-EGEVÁRY THEOREM.

see also KÖNIG-EGEVÁRY THEOREM, PERMANENT

Frobenius Map

A map $x \mapsto x^p$ where p is a PRIME.

Frobenius Method

If x_0 is an ordinary point of the ORDINARY DIFFERENTIAL EQUATION, expand y in a TAYLOR SERIES about x_0, letting

$$y = \sum_{n=0}^{\infty} a_n x^n. \tag{1}$$

Plug y back into the ODE and group the COEFFICIENTS by POWER. Now, obtain a RECURRENCE RELATION for the nth term, and write the TAYLOR SERIES in terms of the a_ns. Expansions for the first few derivatives are

$$y = \sum_{n=0}^{\infty} a_n x^n \tag{2}$$

$$y' = \sum_{n=1}^{\infty} n a_n x^{n-1} = \sum_{n=0}^{\infty} (n+1) a_{n+1} x^n \tag{3}$$

$$y'' = \sum_{n=2}^{\infty} n(n-1) a_n x^{n-2} = \sum_{n=0}^{\infty} (n+2)(n+1) a_{n+2} x^n. \tag{4}$$

If x_0 is a regular singular point of the ORDINARY DIFFERENTIAL EQUATION,

$$P(x)y'' + Q(x)y' + R(x)y = 0, \tag{5}$$

solutions may be found by the Frobenius method or by expansion in a LAURENT SERIES. In the Frobenius method, assume a solution of the form

$$y = x^k \sum_{n=0}^{\infty} a_n x^n, \tag{6}$$

so that

$$y = x^k \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} a_n x^{n+k} \tag{7}$$

$$y' = \sum_{n=0}^{\infty} a_n (n+k) x^{k+n-1} \tag{8}$$

$$y'' = \sum_{n=0}^{\infty} a_n (n+k)(n+k-1) x^{k+n-2}. \tag{9}$$

Now, plug y back into the ODE and group the COEFFICIENTS by POWER to obtain a recursion FORMULA for the a_nth term, and then write the TAYLOR SERIES in terms of the a_ns. Equating the a_0 term to 0 will produce the so-called INDICIAL EQUATION, which will give the allowed values of k in the TAYLOR SERIES.

FUCHS'S THEOREM guarantees that at least one POWER series solution will be obtained when applying the Frobenius method if the expansion point is an ordinary, or regular, SINGULAR POINT. For a regular SINGULAR POINT, a LAURENT SERIES expansion can also be used. Expand y in a LAURENT SERIES, letting

$$y = c_{-n}x^{-n} + \ldots + c_{-1}x^{-1} + c_0 + c_1x + \ldots + c_nx^n + \ldots. \tag{10}$$

Plug y back into the ODE and group the COEFFICIENTS by POWER. Now, obtain a recurrence FORMULA for the c_nth term, and write the TAYLOR EXPANSION in terms of the c_ns.

see also FUCHS'S THEOREM, ORDINARY DIFFERENTIAL EQUATION

References

Arfken, G. "Series Solutions—Frobenius' Method." §8.5 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 454–467, 1985.

Frobenius-Peron Equation

$$\rho_{n+1}(x) = \int \rho_n(y)\delta[x - M(y)]\,dy,$$

where $\delta(x)$ is a DELTA FUNCTION, $M(x)$ is a map, and ρ is the NATURAL DENSITY.

References

Ott, E. *Chaos in Dynamical Systems.* New York: Cambridge University Press, p. 51, 1993.

Frobenius Pseudoprime

Let $f(x)$ be a MONIC POLYNOMIAL of degree d with discriminant Δ. Then an ODD INTEGER n with $(n, f(0)\Delta) = 1$ is called a Frobenius pseudoprime with respect to $f(x)$ if it passes a certain algorithm given by Grantham (1996). A Frobenius pseudoprime with respect to a POLYNOMIAL $f(x) \in \mathbb{Z}[x]$ is then a composite Frobenius probably prime with respect to the POLYNOMIAL $x - a$.

While 323 is the first LUCAS PSEUDOPRIME with respect to the Fibonacci polynomial $x^2 - x - 1$, the first Frobenius pseudoprime is 5777. If $f(x) = x^3 - rx^2 + sx - 1$, then any Frobenius pseudoprime n with respect to $f(x)$ is also a PERRIN PSEUDOPRIME. Grantham (1997) gives a test based on Frobenius pseudoprimes which is passed by COMPOSITE NUMBERS with probability at most 1/7710.

see also PERRIN PSEUDOPRIME, PSEUDOPRIME, STRONG FROBENIUS PSEUDOPRIME

References

Grantham, J. "Frobenius Pseudoprimes." 1996. `http://www.clark.net/pub/grantham/pseudo/pseudo.ps`

Grantham, J. "A Frobenius Probable Prime Test with High Confidence." 1997. `http://www.clark.net/pub/grantham/pseudo/pseudo2.ps`

Grantham, J. "Pseudoprimes/Probable Primes." `http://www.clark.net/pub/grantham/pseudo`.

Frobenius Theorem

Let $\mathsf{A} = a_{ij}$ be a MATRIX with POSITIVE COEFFICIENTS so that $a_{ij} > 0$ for all $i, j = 1, 2, \ldots, n$, then A has a POSITIVE EIGENVALUE λ_0, and all its EIGENVALUES lie on the CLOSED DISK

$$|z| \leq \lambda_0.$$

see also CLOSED DISK, OSTROWSKI'S THEOREM

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1121, 1979.

Frobenius Triangle Identities

Let $C_{L,M}$ be a PADÉ APPROXIMANT. Then

$$C_{(L+1)/M}S_{(L-1)/M} - C_{L/(M+1)}S_{L/(M+1)} = C_{L/M}S_{L/M} \tag{1}$$

$$C_{L/(M+1)}S_{(L+1)/M} - C_{(L+1)/M}S_{L/(M+1)} = C_{(L+1)/(M+1)}xS_{L/M} \tag{2}$$

$$C_{(L+1)/M}S_{L/M} - C_{L/M}S_{(L+1)/M} = C_{(L+1)/(M+1)}xS_{L/(M-1)} \tag{3}$$

$$C_{L/(M+1)}S_{L/M} - C_{L/M}S_{L/(M+1)} = C_{(L+1)/(M+1)}xS_{(L-1)/M}, \tag{4}$$

where

$$S_{L/M} = G(x)P_L(x) + H(x)Q_M(x) \tag{5}$$

and C is the C-DETERMINANT.

see also C-DETERMINANT, PADÉ APPROXIMANT

References

Baker, G. A. Jr. *Essentials of Padé Approximants in Theoretical Physics.* New York: Academic Press, p. 31, 1975.

Frontier

see BOUNDARY

Frullani's Integral

If $f'(x)$ is continuous and the integral converges,

$$\int_0^\infty \frac{f(ax) - f(bx)}{x}\,dx = [f(0) - f(\infty)]\ln\left(\frac{b}{a}\right).$$

References

Spiegel, M. R. *Mathematical Handbook of Formulas and Tables.* New York: McGraw-Hill, 1968.

Frustum

The portion of a solid which lies between two PARALLEL PLANES cutting the solid. Degenerate cases are obtained for finite solids by cutting with a single PLANE only.

see also CONICAL FRUSTUM, PYRAMIDAL FRUSTUM, SPHERICAL SEGMENT

Fubini Principle

If the average number of envelopes per pigeonhole is a, then some pigeonhole will have at least a envelopes. Similarly, there must be a pigeonhole with at most a envelopes.

see also PIGEONHOLE PRINCIPLE

Fuchsian System

A system of linear differential equations

$$\frac{dy}{dz} = A(z)y,$$

with $A(z)$ an ANALYTIC $n \times n$ MATRIX, for which the MATRIX $A(z)$ is ANALYTIC in $\bar{\mathbb{C}}\backslash\{a_1, \ldots, a_N\}$ and has a POLE of order 1 at a_j for $j = 1, \ldots, N$. A system is Fuchsian IFF there exist $n \times n$ matrices $B_1, \ldots, B_N$ with entries in $\mathbb{Z}$ such that

$$A(z) = \sum_{j=1}^{N} \frac{B_j}{z - a_j}$$

$$\sum_{j=1}^{N} B_j = v.$$

Fuchs's Theorem

At least one POWER SERIES solution will be obtained when applying the FROBENIUS METHOD if the expansion point is an ordinary, or regular, SINGULAR POINT. The number of ROOTS is given by the ROOTS of the INDICIAL EQUATION.

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 462–463, 1985.

Fuhrmann Circle

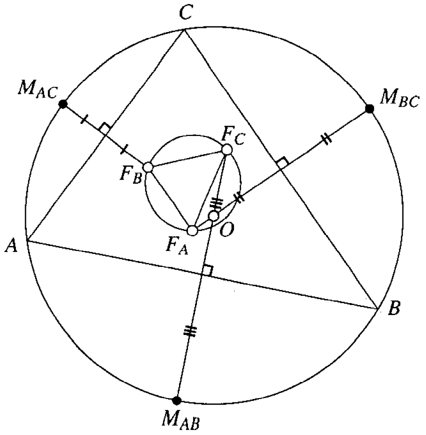

The CIRCUMCIRCLE of the FUHRMANN TRIANGLE.

see also FUHRMANN TRIANGLE, MID-ARC POINTS

References

Fuhrmann, W. *Synthetische Beweise Planimetrischer Sätze.* Berlin, p. 107, 1890.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 228–229, 1929.

Fuhrmann's Theorem

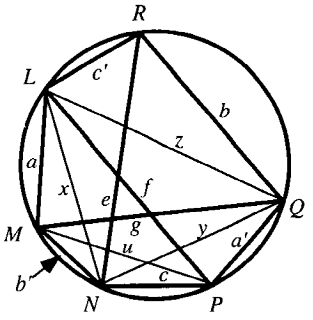

Let the opposite sides of a convex CYCLIC HEXAGON be a, a', b, b', c, and c', and let the DIAGONALS e, f, and g be so chosen that a, a', and e have no common VERTEX (and likewise for b, b', and f), then

$$efg = aa'e + bb'f + cc'g + abc + a'b'c'.$$

This is an extension of PTOLEMY'S THEOREM to the HEXAGON.

see also CYCLIC HEXAGON, HEXAGON, PTOLEMY'S THEOREM

References

Fuhrmann, W. *Synthetische Beweise Planimetrischer Sätze.* Berlin, p. 61, 1890.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 65–66, 1929.

Fuhrmann Triangle

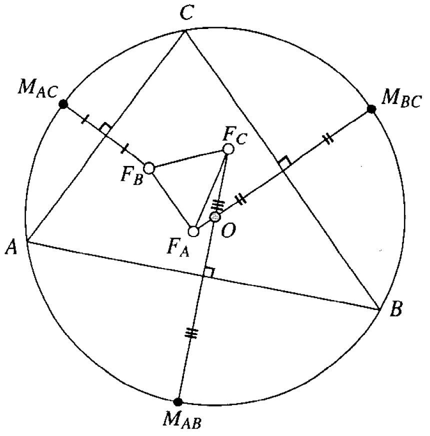

The Fuhrmann triangle of a TRIANGLE ΔABC is the TRIANGLE $\Delta F_C F_B F_A$ formed by reflecting the MID-ARC POINTS M_{AB}, M_{AC}, M_{BC} about the lines AB, AC, and BC. The CIRCUMCIRCLE of the Fuhrmann triangle is called the FUHRMANN CIRCLE, and the lines $F_A M_{BC}$, $F_B M_{AC}$, and $F_C M_{AB}$ CONCUR at the CIRCUMCENTER O.

see also FUHRMANN CIRCLE, MID-ARC POINTS

References

Fuhrmann, W. *Synthetische Beweise Planimetrischer Sätze.* Berlin, p. 107, 1890.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 228–229, 1929.

Full Reptend Prime

A PRIME p for which $1/p$ has a maximal period DECIMAL EXPANSION of $p-1$ DIGITS. The first few numbers with maximal decimal expansions are 7, 17, 19, 23, 29, 47, 59, 61, 97, ... (Sloane's A001913).

References

Sloane, N. J. A. Sequence A001913/M4353 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Full Width at Half Maximum

The full width at half maximum (FWHM) is a parameter commonly used to describe the width of a "bump" on a curve or function. It is given by the distance between points on the curve at which the function reaches half its maximum value. The following table gives the analytic and numerical full widths for several common curves.

Function	Formula	FWHM
Bartlett	$1-\frac{\lvert x\rvert}{a}$	a
Blackman		$0.810957a$
Connes	$\left(1-\frac{x^2}{a^2}\right)$	$\sqrt{4-2\sqrt{2}}\,a$
Cosine	$\cos\left(\frac{\pi x}{2a}\right)$	$\frac{4}{3}a$
Gaussian	$e^{-x^2/(2\sigma^2)}$	$2\sqrt{2\ln 2}\,\sigma$
Hamming		$1.05543a$
Hanning		a
Lorentzian	$\frac{\frac{1}{2}\Gamma}{x^2+(\frac{1}{2}\Gamma)^2}$	Γ
Welch	$1-\frac{x^2}{a^2}$	$\sqrt{2}\,a$

see also APODIZATION FUNCTION, MAXIMUM

Fuller Dome

see GEODESIC DOME

Function

A way of associating unique objects to every point in a given SET. A function from A to B is an object f such that for every $a \in A$, there is a unique object $f(a) \in B$. Examples of functions include $\sin x$, x, x^2, etc. The term MAP is synonymous with function.

Poincaré remarked with regard to the proliferation of pathological functions, "Formerly, when one invented a new function, it was to further some practical purpose; today one invents them in order to make incorrect the reasoning of our fathers, and nothing more will ever be accomplished by these inventions."

see also ABELIAN FUNCTION, ABSOLUTE VALUE, ACKERMANN FUNCTION, AIRY FUNCTIONS, ALGEBRAIC FUNCTION, ALGEBROIDAL FUNCTION, ALPHA FUNCTION, ANDREW'S SINE, ANGER FUNCTION, APODIZATION FUNCTION, APPARATUS FUNCTION, ARGUMENT (FUNCTION), ARTIN L-FUNCTION, AUTOMORPHIC FUNCTION, BACHELIER FUNCTION, BARNES G-FUNCTION, BARTLETT FUNCTION, BASSET FUNCTION, BATEMAN FUNCTION, BEI, BER, BERNOULLI FUNCTION, BESSEL FUNCTION OF THE FIRST KIND, BESSEL FUNCTION OF THE SECOND KIND, BESSEL FUNCTION OF THE THIRD KIND, BETA FUNCTION, BETA FUNCTION (EXPONENTIAL), BINOMIAL COEFFICIENT, BLACKMAN FUNCTION, BLANCMANGE FUNCTION, BOOLEAN FUNCTION, BOURGET FUNCTION, BOXCAR FUNCTION, BROWN FUNCTION, CAL, CANTOR FUNCTION, CARMICHAEL FUNCTION, CAROTID-KUNDALINI FUNCTION, CEILING FUNCTION, CENTER FUNCTION, CENTRAL BETA FUNCTION, CHARACTERISTIC FUNCTION, CHEBYSHEV FUNCTION, CIRCULAR FUNCTIONS, CLAUSEN FUNCTION, COMB FUNCTION, COMPLETE FUNCTIONS, COMPLEX CONJUGATE, COMPUTABLE FUNCTION, CONCAVE FUNCTION, CONFLUENT HYPERGEOMETRIC FUNCTION, CONFLUENT HYPERGEOMETRIC FUNCTION OF THE FIRST KIND, CONFLUENT HYPERGEOMETRIC FUNCTION OF THE SECOND KIND, CONFLUENT HYPERGEOMETRIC LIMIT FUNCTION, CONICAL FUNCTION, CONNES FUNCTION, CONSTANT FUNCTION, CONTIGUOUS FUNCTION, CONTINUOUS FUNCTION, CONVEX FUNCTION, COPULA, COSECANT, COSINE, COSINE APODIZATION FUNCTION, COTANGENT, COULOMB WAVE FUNCTION, COVERSINE, CUBE ROOT, CUBED, CUMULANT-GENERATING FUNCTION, CUMULATIVE DISTRIBUTION FUNCTION, CUNNINGHAM FUNCTION, CYLINDER FUNCTION, CYLINDRICAL FUNCTION, DEBYE FUNCTIONS, DECREASING FUNCTION, DEDEKIND ETA FUNCTION, DEDEKIND FUNCTION, DELTA FUNCTION, DIGAMMA FUNCTION, DILOGARITHM, DIRAC DELTA FUNCTION, DIRICHLET BETA FUNCTION, DIRICHLET ETA FUNCTION, DIRICHLET FUNCTION, DIRICHLET LAMBDA FUNCTION, DISTRIBUTION FUNCTION, DIVISOR FUNCTION, DOUBLE GAMMA FUNCTION, DOUBLET FUNCTION, E_n-FUNCTION, E_t-FUNCTION, EIGENFUNCTION, EIN FUNCTION, EINSTEIN FUNCTIONS, ELEMENTARY FUNCTION, ELLIPTIC ALPHA FUNCTION, ELLIPTIC DELTA FUNCTION, ELLIPTIC EXPONENTIAL FUNCTION, ELLIPTIC FUNCTION, ELLIPTIC FUNCTIONAL, ELLIPTIC LAMBDA FUNCTION, ELLIPTIC MODULAR FUNCTION, ELLIPTIC THETA FUNCTION, ELSASSER FUNCTION, ENTIRE FUNCTION, EPSTEIN ZETA FUNCTION, ERDŐS-SELFRIDGE FUNCTION, ERF, ERROR FUNCTION, EXPONENTIAL RAMP, EULER L-FUNCTION, EVEN FUNCTION, EXPONENTIAL FUNCTION, EXPONENTIAL FUNCTION (TRUNCATED), EXPONENTIAL SUM FUNCTION, EXSECANT, FLOOR FUNCTION, FOX'S H-FUNCTION,

Function Space, G-Function, Gamma Function, Gate Function, Gaussian Function, Gegenbauer Function, Generalized Function, Generalized Hyperbolic Functions, Generalized Hypergeometric Function, Generating Function, Gordon Function, Green's Function, Growth Function, Gudermannian Function, H-Function, Haar Function, Hamming Function, Hankel Function, Hankel Function of the First Kind, Hankel Function of the Second Kind, Hann Function, Hanning Function, Harmonic Function, Haversine, Heaviside Step Function, Hecke L-Function, Hemicylindrical Function, Hemispherical Function, Heuman Lambda Function, Hh Function, Hilbert Function, Holonomic Function, Homogeneous Function, Hurwitz Zeta Function, Hyperbolic Cosecant, Hyperbolic Cosine, Hyperbolic Cotangent, Hyperbolic Functions, Hyperbolic Secant, Hyperbolic Sine, Hyperbolic Tangent, Hyperelliptic Function, Hypergeometric Function, Identity Function, Implicit Function, Implicit Function Theorem, Incomplete Gamma Function, Increasing Function, Infinite Product, Instrument Function, Int, Inverse Cosecant, Inverse Cosine, Inverse Cotangent, Inverse Function, Inverse Hyperbolic Functions, Inverse Secant, Inverse Sine, Inverse Tangent, j-Function, Jacobi Elliptic Functions, Jacobi Function of the First Kind, Jacobi Function of the Second Kind, Jacobi Theta Function, Jacobi Zeta Function, Jinc Function, Joint Probability Density Function, Jonquière's Function, K-Function, Kei, Kelvin Functions, Ker, Koebe Function, L-Function, Lambda Function, Lambda Hypergeometric Function, Lambert's W-Function, Lamé Function, Legendre Function of the First Kind, Legendre Function of the Second Kind, Lemniscate Function, Lemniscate Function, Length Distribution Function, Lerch Transcendent, Lévy Function, Linearly Dependent Functions, Liouville Function, Lipschitz Function, Logarithm, Logarithmically Convex Function, Logit Transformation, Lommel Function, Lyapunov Function, MacRobert's E-Function, Mangoldt Function, Mathieu Function, Measurable Function, Meijer's G-Function, Meromorphic, Mertens Function, Mertz Apodization Function, Mittag-Leffler Function, Möbius Function, Möbius Periodic Function, Mock Theta Function, Modified Bessel Function of the First Kind, Modified Bessel Function of the Second Kind, Modified Spherical Bessel Function, Modified Struve Function, Modular Function, Modular Gamma Function, Modular Lambda Function, Moment-Generating Function, Monogenic Function, Monotonic Function, Mu Function, Multiplicative Function, Multivalued Function, Multivariate Function, Neumann Function, Nint, Nu Function, Null Function, Numeric Function, Oblate Spheroidal Wave Function, Odd Function, Omega Function, One-Way Function, Parabolic Cylinder Function, Partition Function P, Partition Function Q, Parzen Apodization Function, Pearson-Cunningham Function, Pearson's Function, Periodic Function, Planck's Radiation Function, Plurisubharmonic Function, Pochhammer Symbol, Poincaré-Fuchs-Klein Automorphic Function, Poisson-Charlier Function, Polygamma Function, Polygenic Function, Polylogarithm, Positive Definite Function, Potential Function, Power, Prime Counting Function, Prime Difference Function, Probability Density Function, Probability Distribution Function, Prolate Spheroidal Wave Function, Psi Function, Pulse Function, q-Beta Function, Q-Function, q-Gamma Function, Quasiperiodic Function, Rademacher Function, Ramanujan Function, Ramanujan g- and G- Functions, Ramanujan Theta Functions, Ramp Function, Rational Function, Real Function, Rectangle Function, Regular Function, Regularized Gamma Function, Restricted Divisor Function, Riemann Function, Riemann-Mangoldt Function, Riemann-Siegel Functions, Riemann Theta Function, Riemann Zeta Function, Ring Function, Sal, Sampling Function, Scalar Function, Schlomilch's Function, Secant, Sequency Function, Sgn, Shah Function, Siegel Modular Function, Sigma Function, Sigmoid Function, Sign, Sinc Function, Sine, Smarandache Function, Spence's Function, Spherical Bessel Function of the First Kind, Spherical Bessel Function of the Second Kind, Spherical Hankel Function of the First Kind, Spherical Hankel Function of the Second Kind, Spherical Harmonic, Spheroidal Wavefunction, Sprague-Grundy Function, Square Root, Squared, Step Function, Struve Function, Sturm Function, Summatory Function, Symmetric Function, TAK Function, Tangent, Tapering Function, Tau Function, Tetrachoric Function, Theta Function, Toroidal Function, Toronto Function, Total Function, Totient Function, Totient Valence Function, Transcendental Function, Transfer Function, Trapdoor Function, Triangle Center Function, Triangle Function, Tricomi Function, Trigonometric Functions, Uniform Apodization Function, Univalent Function, Vector Function, Versine, von Mangoldt Function, W-Function, Walsh Function, Weber Functions, Weierstraß Elliptic Function, Weierstraß Function, Weierstraß Sigma Function, Weierstraß Zeta Function, Weighting Function, Welch Apodization

FUNCTION, WHITTAKER FUNCTION, WIENER FUNCTION, WINDOW FUNCTION, XI FUNCTION, ZETA FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Miscellaneous Functions." Ch. 27 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 997–1010, 1972.

Arfken, G. "Special Functions." Ch. 13 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 712–759, 1985.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Special Functions." Ch. 6 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 205–265, 1992.

Function Field

see ALGEBRAIC FUNCTION FIELD

Function Space

$f(I)$ is the collection of all real-valued continuous functions defined on some interval I. $f^{(n)}(I)$ is the collection of all functions $\in f(I)$ with continuous nth DERIVATIVES. A function space is a TOPOLOGICAL VECTOR SPACE whose "points" are functions.

see also FUNCTIONAL, FUNCTIONAL ANALYSIS, OPERATOR

Functional

A mapping between FUNCTION SPACES if the range is on the REAL LINE or in the COMPLEX PLANE.

see also COERCIVE FUNCTIONAL, CURRENT, ELLIPTIC FUNCTIONAL, GENERALIZED FUNCTION, LAX-MILGRAM THEOREM, OPERATOR, RIESZ REPRESENTATION THEOREM

Functional Analysis

A branch of mathematics concerned with infinite dimensional spaces (mainly FUNCTION SPACES) and mappings between them. The SPACES may be of different, and possibly INFINITE, DIMENSIONS. These mappings are called OPERATORS or, if the range is on the REAL line or in the COMPLEX PLANE, FUNCTIONALS.

see also FUNCTIONAL, OPERATOR

References

Balakrishnan, A. V. *Applied Functional Analysis, 2nd ed.* New York: Springer-Verlag, 1981.

Berezansky, Y. M.; Us, G. F.; and Sheftel, Z. G. *Functional Analysis, Vol. 1.* Boston, MA: Birkhäuser, 1996.

Berezansky, Y. M.; Us, G. F.; and Sheftel, Z. G. *Functional Analysis, Vol. 2.* Boston, MA: Birkhäuser, 1996.

Birkhoff, G. and Kreyszig, E. "The Establishment of Functional Analysis." *Historia Math.* **11**, 258–321, 1984.

Hutson, V. and Pym, J. S. *Applications of Functional Analysis and Operator Theory.* New York: Academic Press, 1980.

Kreyszig, E. *Introductory Functional Analysis with Applications.* New York: Wiley, 1989.

Yoshida, K. *Functional Analysis and Its Applications.* New York: Springer-Verlag, 1971.

Zeidler, E. *Nonlinear Functional Analysis and Its Applications.* New York: Springer-Verlag, 1989.

Zeidler, E. *Applied Functional Analysis: Applications to Mathematical Physics.* New York: Springer-Verlag, 1995.

Functional Calculus

An early name for CALCULUS OF VARIATIONS.

Functional Derivative

A generalization of the concept of the DERIVATIVE to GENERALIZED FUNCTIONS.

Functor

A function between CATEGORIES which maps objects to objects and MORPHISMS to MORPHISMS. Functors exist in both covariant and contravariant types.

see also CATEGORY, EILENBERG-STEENROD AXIOMS, MORPHISM, SCHUR FUNCTOR

Fundamental Class

The canonical generator of the nonvanishing HOMOLOGY GROUP on a TOPOLOGICAL MANIFOLD.

see also CHERN NUMBER, PONTRYAGIN NUMBER, STIEFEL-WHITNEY NUMBER

Fundamental Continuity Theorem

Given two POLYNOMIALS of the same order in one variable where the first p COEFFICIENTS (but not the first $p-1$) are 0 and the COEFFICIENTS of the second approach the corresponding COEFFICIENTS of the first as limits, then the second POLYNOMIAL will have exactly p roots that increase indefinitely. Furthermore, exactly k ROOTS of the second will approach each ROOT of multiplicity k of the first as a limit.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 4, 1959.

Fundamental Discriminant

$-D$ is a fundamental discriminant if D is a POSITIVE INTEGER which is not DIVISIBLE by any square of an ODD PRIME and which satisfies $D \equiv 3 \pmod 4$ or $D \equiv 4, 8 \pmod{16}$.

see also DISCRIMINANT

References

Atkin, A. O. L. and Morain, F. "Elliptic Curves and Primality Proving." *Math. Comput.* **61**, 29–68, 1993.

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, p. 294, 1987.

Cohn, H. *Advanced Number Theory.* New York: Dover, 1980.

Dickson, L. E. *History of the Theory of Numbers, Vols. 1–3.* New York: Chelsea, 1952.

Fundamental Forms

There are three types of so-called fundamental forms. The most important are the first and second (since the third can be expressed in terms of these). The fundamental forms are extremely important and useful in determining the metric properties of a surface, such as LINE ELEMENT, AREA ELEMENT, NORMAL CURVATURE, GAUSSIAN CURVATURE, and MEAN CURVATURE. Let M be a REGULAR SURFACE with $\mathbf{v_p}, \mathbf{w_p}$ points on the TANGENT SPACE $M_\mathbf{p}$ of M. Then the first fundamental form is the INNER PRODUCT of tangent vectors,

$$\mathbf{I}(\mathbf{v_p}, \mathbf{w_p}) = \mathbf{v_p} \cdot \mathbf{w_p}. \tag{1}$$

For $M \in \mathbb{R}^3$, the second fundamental form is the symmetric bilinear form on the TANGENT SPACE $M_\mathbf{p}$,

$$\mathbf{II}(\mathbf{v_p}, \mathbf{w_p}) = S(\mathbf{v_p}) \cdot \mathbf{w_p}, \tag{2}$$

where S is the SHAPE OPERATOR. The third fundamental form is given by

$$\mathbf{III}(\mathbf{v_p}, \mathbf{w_p}) = S(\mathbf{v_p}) \cdot S(\mathbf{w_p}). \tag{3}$$

The first and second fundamental forms satisfy

$$\mathbf{I}(a\mathbf{x}_u + b\mathbf{x}_v, a\mathbf{x}_u + b\mathbf{x}_v) = Ea^2 + 2Fab + Gb^2 \tag{4}$$

$$\mathbf{II}(a\mathbf{x}_u + b\mathbf{x}_v, a\mathbf{x}_u + b\mathbf{x}_v) = ea^2 + 2fab + gb^2, \tag{5}$$

and so their ratio is simply the NORMAL CURVATURE

$$\kappa(\mathbf{v_p}) = \frac{\mathbf{II}(\mathbf{v_p})}{\mathbf{I}(\mathbf{v_p})} \tag{6}$$

for any nonzero TANGENT VECTOR. The third fundamental form is given in terms of the first and second forms by

$$\mathbf{III} - 2H\mathbf{II} + K\mathbf{I} = 0, \tag{7}$$

where H is the MEAN CURVATURE and K is the GAUSSIAN CURVATURE.

The first fundamental form (or LINE ELEMENT) is given explicitly by the RIEMANNIAN METRIC

$$ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2. \tag{8}$$

It determines the ARC LENGTH of a curve on a surface. The coefficients are given by

$$E = \mathbf{x}_{uu} = \left|\frac{\partial \mathbf{x}}{\partial u}\right|^2 \tag{9}$$

$$F = \mathbf{x}_{uv} = \frac{\partial \mathbf{x}}{\partial u} \cdot \frac{\partial \mathbf{x}}{\partial v} \tag{10}$$

$$G = \mathbf{x}_{vv} = \left|\frac{\partial \mathbf{x}}{\partial v}\right|^2. \tag{11}$$

The coefficients are also denoted $g_{uu} = E$, $g_{uv} = F$, and $g_{vv} = G$. In CURVILINEAR COORDINATES (where $F = 0$), the quantities

$$h_u \equiv \sqrt{g_{uu}} = \sqrt{E} \tag{12}$$

$$h_v \equiv \sqrt{g_{vv}} = \sqrt{G} \tag{13}$$

are called SCALE FACTORS.

The second fundamental form is given explicitly by

$$e\,du^2 + 2f\,du\,dv + g\,dv^2 \tag{14}$$

where

$$e = \sum_i X_i \frac{\partial^2 x_i}{\partial u^2} \tag{15}$$

$$f = \sum_i X_i \frac{\partial^2 x_i}{\partial u \partial v} \tag{16}$$

$$g = \sum_i X_i \frac{\partial^2 x_i}{\partial v^2}, \tag{17}$$

and X_i are the DIRECTION COSINES of the surface normal. The second fundamental form can also be written

$$e = -\mathbf{N}_u \cdot \mathbf{x}_u = \mathbf{N} \cdot \mathbf{x}_{uu} \tag{18}$$

$$f = -\mathbf{N}_v \cdot \mathbf{x}_u = \mathbf{N} \cdot \mathbf{x}_{uv} = \mathbf{N}_{vu} \cdot \mathbf{x}_{vu}$$

$$= -\mathbf{N}_u \cdot \mathbf{x}_v \tag{19}$$

$$g = -\mathbf{N}_v \cdot \mathbf{x}_v = \mathbf{N} \cdot \mathbf{x}_{vv}, \tag{20}$$

where $\mathbf{N}$ is the NORMAL VECTOR, or

$$e = \frac{\det(\mathbf{x}_{uu}\mathbf{x}_u\mathbf{x}_v)}{\sqrt{EG - F^2}} \tag{21}$$

$$f = \frac{\det(\mathbf{x}_{uv}\mathbf{x}_u\mathbf{x}_v)}{\sqrt{EG - F^2}} \tag{22}$$

$$g = \frac{\det(\mathbf{x}_{vv}\mathbf{x}_u\mathbf{x}_v)}{\sqrt{EG - F^2}}. \tag{23}$$

see also ARC LENGTH, AREA ELEMENT, GAUSSIAN CURVATURE, GEODESIC, KÄHLER MANIFOLD, LINE OF CURVATURE, LINE ELEMENT, MEAN CURVATURE, NORMAL CURVATURE, RIEMANNIAN METRIC, SCALE FACTOR, WEINGARTEN EQUATIONS

References

Gray, A. "The Three Fundamental Forms." §14.6 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 251–255, 259–260, 275–276, and 282–291, 1993.

Fundamental Group

The fundamental group of a CONNECTED SET S is the QUOTIENT GROUP of the GROUP of all paths with initial and final points at a given point P and the SUBGROUP of all paths HOMOTOPIC to the degenerate path consisting of the point P.

The fundamental group of the CIRCLE is the INFINITE CYCLIC GROUP. Two fundamental groups having different points P are ISOMORPHIC. If the fundamental group consists only of the identity element, then the set S is simply connected.

see also MILNOR'S THEOREM

Fundamental Homology Class

see also FUNDAMENTAL CLASS

Fundamental Lemma of Calculus of Variations

If

$$\int_a^b M(x)h(x)\,dx = 0$$

$\forall\ h(x)$ with CONTINUOUS second PARTIAL DERIVATIVES, then

$$M(x) = 0$$

on the OPEN INTERVAL (a, b).

Fundamental System

A set of ALGEBRAIC INVARIANTS for a QUANTIC such that any invariant of the QUANTIC is expressible as a POLYNOMIAL in members of the set. In 1868, Gordan proved the existence of finite fundamental systems of algebraic invariants and covariants for any binary QUANTIC. In 1890, Hilbert (1890) proved the HILBERT BASIS THEOREM, which is a finiteness theorem for the related concept of SYZYGIES.

see also HILBERT BASIS THEOREM, SYZYGY

References

Hilbert, D. "Über die Theorie der algebraischen Formen." *Math. Ann.* **36**, 473–534, 1890.

Fundamental Theorem of Algebra

Every POLYNOMIAL equation having COMPLEX COEFFICIENTS and degree ≥ 1 has at least one COMPLEX ROOT. This theorem was first proven by Gauss. It is equivalent to the statement that a POLYNOMIAL $P(z)$ of degree n has n values of z (some of them possibly degenerate) for which $P(z) = 0$. An example of a POLYNOMIAL with a single ROOT of multiplicity > 1 is $z^2 - 2z + 1 = (z-1)(z-1)$, which has $z = 1$ as a ROOT of multiplicity 2.

see also DEGENERATE, POLYNOMIAL

References

Courant, R. and Robbins, H. "The Fundamental Theorem of Algebra." §2.5.4 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 101–103, 1996.

Fundamental Theorem of Arithmetic

Any POSITIVE INTEGER can be represented in exactly one way as a PRODUCT of PRIMES. The theorem is also called the UNIQUE FACTORIZATION THEOREM. The fundamental theorem of algebra is a COROLLARY of the first of EUCLID'S THEOREMS (Hardy and Wright 1979).

see also EUCLID'S THEOREMS, INTEGER, PRIME NUMBER

References

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 23, 1996.

Hardy, G. H. and Wright, E. M. "Statement of the Fundamental Theorem of Arithmetic," "Proof of the Fundamental Theorem of Arithmetic," and "Another Proof of teh Fundamental Theorem of Arithmetic." §1.3, 2.10 and 2.11 in *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, pp. 3 and 21, 1979.

Fundamental Theorems of Calculus

The first fundamental theorem of calculus states that, if f is CONTINUOUS on the CLOSED INTERVAL $[a, b]$ and F is the ANTIDERIVATIVE (INDEFINITE INTEGRAL) of f on $[a, b]$, then

$$\int_a^b f(x)\,dx = F(b) - F(a). \tag{1}$$

The second fundamental theorem of calculus lets f be CONTINUOUS on an OPEN INTERVAL I and lets a be any point in I. If F is defined by

$$F(x) = \int_a^x f(t)\,dt, \tag{2}$$

then

$$F'(x) = f(x) \tag{3}$$

at each point in I.

The complex fundamental theorem of calculus states that if $f(z)$ has a CONTINUOUS ANTIDERIVATIVE $F(z)$ in a region R containing a parameterized curve $\gamma : z = z(t)$ for $\alpha \leq t \leq \beta$, then

$$\int_\gamma f(z)\,dz = F(z(\beta)) - F(z(\alpha)). \tag{4}$$

see also CALCULUS, DEFINITE INTEGRAL, INDEFINITE INTEGRAL, INTEGRAL

Fundamental Theorem of Curves

The CURVATURE and TORSION functions along a SPACE CURVE determine it up to an orientation-preserving ISOMETRY.

Fundamental Theorem of Directly Similar Figures

Let F_0 and F_1 denote two directly similar figures in the plane, where $P_1 \in F_1$ corresponds to $P_0 \in F_0$ under the given similarity. Let $r \in (0,1)$, and define $F_r = \{(1-r)P_0 + rP_1 : P_0 \in F_0\}$. Then F_r is also directly similar to F_0.

see also FINSLER-HADWIGER THEOREM

References

Detemple, D. and Harold, S. "A Round-Up of Square Problems." *Math. Mag.* **69**, 15–27, 1996.

Eves, H. Solution to Problem E521. *Amer. Math. Monthly* **50**, 64, 1943.

Fundamental Theorem of Gaussian Quadrature

The ABSCISSAS of the N point GAUSSIAN QUADRATURE FORMULA are precisely the ROOTS of the ORTHOGONAL POLYNOMIAL for the same INTERVAL and WEIGHTING FUNCTION.

see also GAUSSIAN QUADRATURE

Fundamental Theorem of Genera

$$2^{\omega(d)-1}|h(-d)|,$$

where $\omega(d)$ is the genus of forms and $h(-d)$ is the CLASS NUMBER of an IMAGINARY QUADRATIC FIELD.

References

Arno, S.; Robinson, M. L.; and Wheeler, F. S. "Imaginary Quadratic Fields with Small Odd Class Number." `http://www.math.uiuc.edu/Algebraic-Number-Theory/0009/`.

Cohn, H. *Advanced Number Theory.* New York: Dover, p. 224, 1980.

Gauss, C. F. *Disquisitiones Arithmeticae.* New Haven, CT: Yale University Press, 1966.

Fundamental Theorem of Plane Curves

Two unit-speed plane curves which have the same CURVATURE differ only by a EUCLIDEAN MOTION.

see also FUNDAMENTAL THEOREM OF SPACE CURVES

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 103 and 110–111, 1993.

Fundamental Theorem of Projective Geometry

A PROJECTIVITY is determined when three points of one RANGE and the corresponding three points of the other are given.

see also PROJECTIVE GEOMETRY

Fundamental Theorem of Space Curves

If two single-valued continuous functions $\kappa(s)$ (CURVATURE) and $\tau(s)$ (TORSION) are given for $s > 0$, then there exists EXACTLY ONE SPACE CURVE, determined except for orientation and position in space (i.e., up to a EUCLIDEAN MOTION), where s is the ARC LENGTH, κ is the CURVATURE, and τ is the TORSION.

see also ARC LENGTH, CURVATURE, EUCLIDEAN MOTION, FUNDAMENTAL THEOREM OF PLANE CURVES, TORSION (DIFFERENTIAL GEOMETRY)

References

Gray, A. "The Fundamental Theorem of Space Curves." §7.7 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 123 and 142–145, 1993.

Struik, D. J. *Lectures on Classical Differential Geometry.* New York: Dover, p. 29, 1988.

Fundamental Theorem of Symmetric Functions

Any symmetric polynomial (respectively, symmetric rational function) can be expressed as a POLYNOMIAL (respectively, RATIONAL FUNCTION) in the ELEMENTARY SYMMETRIC FUNCTIONS on those variables.

see also ELEMENTARY SYMMETRIC FUNCTION

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 2, 1959.

Herstein, I. N. *Noncommutative Rings.* Washington, DC: Math. Assoc. Amer., 1968.

Fundamental Unit

In a real QUADRATIC FIELD, there exists a special UNIT η known as the fundamental unit such that all units ρ are given by $\rho = \pm\eta^m$, for $m = 0, \pm1, \pm2, \ldots$. The following table gives the fundamental units for the first few real quadratic fields.

d	$\eta(d)$	d	$\eta(d)$
2	$1+\sqrt{2}$	51	$50+7\sqrt{51}$
3	$2+\sqrt{3}$	53	$\frac{1}{2}(7+\sqrt{53})$
5	$\frac{1}{2}(1+\sqrt{5})$	55	$89+12\sqrt{55}$
6	$5+2\sqrt{6}$	57	$151+20\sqrt{57}$
7	$8+3\sqrt{7}$	58	$99+13\sqrt{58}$
10	$3+\sqrt{10}$	59	$530+69\sqrt{59}$
11	$10+3\sqrt{11}$	61	$\frac{1}{2}(39+5\sqrt{61})$
13	$\frac{1}{2}(3+\sqrt{13})$	62	$63+8\sqrt{62}$
14	$15+4\sqrt{14}$	65	$8+\sqrt{65}$
15	$4+\sqrt{15}$	66	$65+8\sqrt{66}$
17	$4+\sqrt{17}$	67	$48842+5967\sqrt{67}$
19	$170+39\sqrt{19}$	69	$\frac{1}{2}(25+3\sqrt{69})$
21	$\frac{1}{2}(5+\sqrt{21})$	70	$251+30\sqrt{70}$
22	$197+42\sqrt{22}$	71	$3480+413\sqrt{71}$
23	$24+5\sqrt{23}$	73	$1068+125\sqrt{73}$
26	$5+\sqrt{26}$	74	$43+5\sqrt{74}$
29	$\frac{1}{2}(5+\sqrt{29})$	77	$\frac{1}{2}(9+\sqrt{77})$
30	$11+2\sqrt{30}$	78	$53+6\sqrt{78}$
31	$1520+273\sqrt{31}$	79	$80+9\sqrt{79}$
33	$5+4\sqrt{33}$	82	$9+\sqrt{82}$
34	$35+6\sqrt{34}$	83	$82+9\sqrt{83}$
35	$6+\sqrt{35}$	85	$\frac{1}{2}(9+\sqrt{85})$
37	$6+\sqrt{37}$	86	$10405+1122\sqrt{86}$
38	$37+6\sqrt{38}$	87	$28+3\sqrt{87}$
39	$25+4\sqrt{39}$	89	$501+54\sqrt{89}$
41	$32+5\sqrt{41}$	91	$1574+165\sqrt{91}$
42	$13+2\sqrt{42}$	93	$\frac{1}{2}(29+3\sqrt{93})$
43	$3482+531\sqrt{43}$	95	$39+4\sqrt{95}$
46	$24335+3588\sqrt{46}$	97	$5604+569\sqrt{97}$
47	$48+7\sqrt{47}$		

see also QUADRATIC FIELD, UNIT

References

Cohn, H. "Fundamental Units" and "Construction of Fundamental Units." §6.4 and 6.5 in *Advanced Number Theory.* New York: Dover, pp. 98–102, and 261–274, 1980.

Weisstein, E. W. "Class Numbers." `http://www.astro.virginia.edu/~eww6n/math/notebooks/ClassNumbers.m`.

Funnel

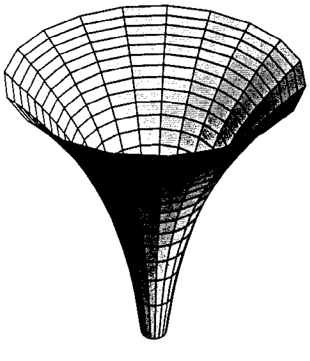

The funnel surface is a REGULAR SURFACE defined by the Cartesian equation

$$z = \tfrac{1}{2}\ln(x^2+y^2) \qquad (1)$$

and the parametric equations

$$x(r,\theta) = r\cos\theta \qquad (2)$$

$$y(r,\theta) = r\sin\theta \qquad (3)$$

$$z(r,\theta) = \ln r. \qquad (4)$$

see also GABRIEL'S HORN, PSEUDOSPHERE, SINCLAIR'S SOAP FILM PROBLEM

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 325–327, 1993.

Fuss's Problem

see BICENTRIC POLYGON

Futile Game

A GAME which permits a draw ("tie") when played properly by both players.

Fuzzy Logic

An extension of two-valued LOGIC such that statements need not be TRUE or FALSE, but may have a degree of truth between 0 and 1. Such a system can be extremely useful in designing control logic for real-world systems such as elevators.

see also ALETHIC, FALSE, LOGIC, TRUE

References

McNeill, D. and Freiberger, P. *Fuzzy Logic.* New York: Simon & Schuster, 1993.

Nguyen, H. T. and Walker, E. A. *A First Course in Fuzzy Logic.* Boca Raton, FL: CRC Press, 1996.

Yager, R. R. and Zadeh, L. A. (Eds.) *An Introduction to Fuzzy Logic Applications in Intelligent Systems.* Boston, MA: Kluwer, 1992.

Zadeh, L. and Kacprzyk, J. (Eds.). *Fuzzy Logic for the Management of Uncertainty.* New York: Wiley, 1992.

FWHM

see FULL WIDTH AT HALF MAXIMUM

G

g-Function

see RAMANUJAN *g*- AND *G*-FUNCTIONS

G-Function

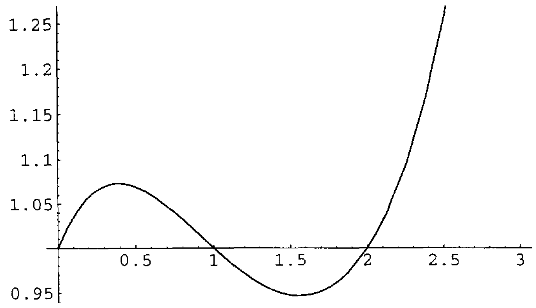

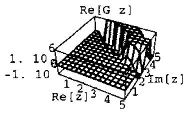

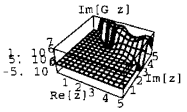

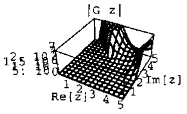

Defined in Whittaker and Watson (1990, p. 264) and also called the BARNES *G*-FUNCTION.

$$G(z+1) \equiv (2\pi)^{z/2} e^{-[z(z+1)+\gamma z^2]/2} \times \prod_{n=1}^{\infty} \left[\left(1+\frac{z}{n}\right)^n e^{-z+z^2/(2n)} \right], \quad (1)$$

where γ is the EULER-MASCHERONI CONSTANT. This is an ANALYTIC CONTINUATION of the *G* function defined in the construction of the GLAISHER-KINKELIN CONSTANT

$$G(n+1) \equiv \frac{(n!)^n}{K(n+1)}, \quad (2)$$

which has the special values

$$G(n) = \begin{cases} 0 & \text{if } n = 0, -1, -2, \ldots \\ 1 & \text{if } n = 1 \\ 0!1!2! \cdots (n-2)! & \text{if } n = 2, 3, 4, \ldots \end{cases} \quad (3)$$

for INTEGER n. This function is what Sloane and Plouffe (1995) call the SUPERFACTORIAL, and the first few values for $n = 1, 2, \ldots$ are 1, 1, 1, 2, 12, 288, 34560, 24883200, 125411328000, 5056584744960000, ... (Sloane's A000178).

The *G*-function is the reciprocal of the DOUBLE GAMMA FUNCTION. It satisfies

$$G(z+1) = \Gamma(z)G(z) \quad (4)$$

$$\frac{(n!)^n}{G(n+1)} = 1^1 \cdot 2^2 \cdot 3^3 \cdots n^n \quad (5)$$

$$\frac{G'(z+1)}{G(z+1)} = \tfrac{1}{2}\ln(2\pi) - \tfrac{1}{2} - z + z\frac{\Gamma'(z)}{\Gamma(z)} \quad (6)$$

$$\ln\left[\frac{G(1-z)}{G(1+z)}\right] = \int_0^z \pi z \cot(\pi z)\, dz - z\ln(2\pi) \quad (7)$$

and has the special values

$$G(\tfrac{1}{2}) = A^{-3/2} \pi^{-1/4} e^{1/8} 2^{1/24} \quad (8)$$

$$G(1) = 1, \quad (9)$$

where

$$A = \exp\left[-\frac{\zeta'(2)}{2\pi^2} + \frac{\ln(2\pi)}{12} + \frac{\gamma}{2}\right] = 1.28242713\ldots. \quad (10)$$

The *G*-function can arise in spectral functions in mathematical physics (Voros 1987).

An unrelated pair of functions are denoted g_n and G_n and are known as RAMANUJAN *g*- AND *G*-FUNCTIONS.

see also EULER-MASCHERONI CONSTANT, GLAISHER-KINKELIN CONSTANT, *K*-FUNCTION, MEIJER'S *G*-FUNCTION, RAMANUJAN *g*- AND *G*-FUNCTIONS, SUPERFACTORIAL

References

Barnes, E. W. "The Theory of the *G*-Function." *Quart. J. Pure Appl. Math.* **31**, 264–314, 1900.

Glaisher, J. W. L. "On a Numerical Continued Product." *Messenger Math.* **6**, 71–76, 1877.

Glaisher, J. W. L. "On the Product $1^1 2^2 3^3 \cdots n^n$." *Messenger Math.* **7**, 43–47, 1878.

Glaisher, J. W. L. "On Certain Numerical Products." *Messenger Math.* **23**, 145–175, 1893.

Glaisher, J. W. L. "On the Constant which Occurs in the Formula for $1^1 2^2 3^3 \cdots n^n$." *Messenger Math.* **24**, 1–16, 1894.

Kinkelin. "Über eine mit der Gammafunktion verwandte Transcendente und deren Anwendung auf die Integralrechnung." *J. Reine Angew. Math.* **57**, 122–158, 1860.

Sloane, N. J. A. Sequence A000178/M2049 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Voros, A. "Spectral Functions, Special Functions and the Selberg Zeta Function." *Commun. Math. Phys.* **110**, 439–465, 1987.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

G-Number

see EISENSTEIN INTEGER

G-Space

A *G*-space is a special type of HAUSDORFF SPACE. Consider a point x and a HOMEOMORPHISM of an open NEIGHBORHOOD V of x onto an OPEN SET of $\mathbb{R}^n$. Then a space is a *G*-space if, for any two such NEIGHBORHOODS V' and V'', the images of $V' \cup V''$ under the different HOMEOMORPHISMS are ISOMETRIC. If $n = 2$, the HOMEOMORPHISMS need only be conformal (but not necessarily orientation-preserving).

see also GREEN SPACE

Gabriel's Horn

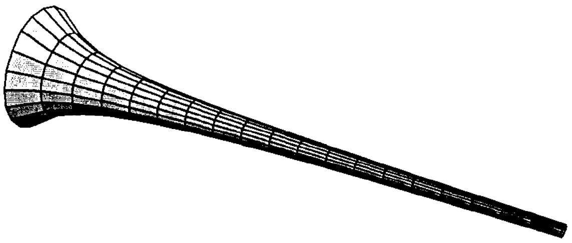

The SURFACE OF REVOLUTION of the function $y = 1/x$ about the x-axis for $x \geq 1$. It has FINITE VOLUME

$$\begin{aligned} V &= \int_1^\infty \pi y^2\,dx = \pi \int_1^\infty \frac{dx}{x^2} \\ &= \pi \left[-\frac{1}{x}\right]_1^\infty = \pi[0-(-1)] = \pi, \end{aligned}$$

but INFINITE SURFACE AREA, since

$$\begin{aligned} S &= \int_1^\infty 2\pi y\sqrt{1+y'^2}\,dx \\ &> 2\pi \int_1^\infty y\,dx = 2\pi \int_1^\infty \frac{dx}{x} = 2\pi[\ln x]_1^\infty \\ &= 2\pi[\ln\infty - 0] = \infty. \end{aligned}$$

This leads to the paradoxical consequence that while Gabriel's horn can be filled up with π cubic units of paint, an INFINITE number of square units of paint are needed to cover its surface!

see also FUNNEL, PSEUDOSPHERE

Gabriel's Staircase

The SUM

$$\sum_{k=1}^\infty kr^k = \frac{r}{(1-r)^2},$$

valid for $0 < r < 1$.

Gadget

A term of endearment used by ALGEBRAIC TOPOLOGISTS when talking about their favorite power tools such as ABELIAN GROUPS, BUNDLES, HOMOLOGY GROUPS, HOMOTOPY GROUPS, k-THEORY, MORSE THEORY, OBSTRUCTIONS, stable homotopy theory, VECTOR SPACES, etc.

see also ABELIAN GROUP, ALGEBRAIC TOPOLOGY, BUNDLE, FREE, HOMOLOGY GROUP, HOMOTOPY GROUP, k-THEORY, OBSTRUCTION, MORSE THEORY, VECTOR SPACE

Gale-Ryser Theorem

Let p and q be PARTITIONS of a POSITIVE INTEGER, then there exists a (0,1)-matrix A such that $c(A) = p$, $r(A) = q$ IFF q is dominated by p^*.

References

Brualdi, R. and Ryser, H. J. §6.2.4 in *Combinatorial Matrix Theory.* New York: Cambridge University Press, 1991.

Krause, M. "A Simple Proof of the Gale-Ryser Theorem." *Amer. Math. Monthly* **103**, 335–337, 1996.

Robinson, G. §1.4 in *The Representation Theory of the Symmetric Group.* Toronto, Canada: University of Toronto Press, 1961.

Ryser, H. J. "The Class $\mathcal{A}(\mathbf{R}, \mathbf{S})$." *Combinatorial Mathematics.* Buffalo, NY: Math. Assoc. Amer., pp. 61–65, 1963.

Galilean Transformation

A transformation from one reference frame to another moving with a constant VELOCITY v with respect to the first for classical motion. However, special relativity shows that the transformation must be modified to the LORENTZ TRANSFORMATION for relativistic motion. The forward Galilean transformation is

$$\begin{bmatrix} t' \\ x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -v & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix},$$

and the inverse transformation is

$$\begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ v & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} t' \\ x' \\ y' \\ z' \end{bmatrix}.$$

see also LORENTZ TRANSFORMATION

Gall's Stereographic Projection

A CYLINDRICAL PROJECTION which projects the equator onto a tangent cylinder which intersects the globe at ± 45°. The transformation equations are

$$\begin{aligned} x &= \lambda \\ y &= \tan(\tfrac{1}{2}\phi), \end{aligned}$$

where λ is the LONGITUDE and ϕ the LATITUDE.

see also STEREOGRAPHIC PROJECTION

References

Dana, P. H. "Map Projections." `http://www.utexas.edu/depts/grg/gcraft/notes/mapproj/mapproj.html`.

Gallows

Schroeder (1991) calls the CEILING FUNCTION symbols $\lceil$ and $\rceil$ the "gallows" because of their similarity in appearance to the structure used for hangings.

see also CEILING FUNCTION

References

Schroeder, M. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise.* New York: W. H. Freeman, p. 57, 1991.

Gallucci's Theorem

If three SKEW LINES all meet three other SKEW LINES, any TRANSVERSAL to the first set of three meets any TRANSVERSAL to the second set of three.

see also SKEW LINES, TRANSVERSAL LINE

Galois Extension Field

The splitting FIELD for a separable POLYNOMIAL over a FINITE FIELD K, where L is a FIELD EXTENSION of K.

Galois Field

see FINITE FIELD

Galois Group

Let L be a FIELD EXTENSION of K, denoted L/K, and let G be the set of AUTOMORPHISMS of L/K, that is, the set of AUTOMORPHISMS σ of L such that $\sigma(x) = x$ for every $x \in K$, so that K is fixed. Then G is a GROUP of transformations of L, called the Galois group of L/K.

The Galois group of $(\mathbb{C}/\mathbb{R})$ consists of the IDENTITY ELEMENT and COMPLEX CONJUGATION. These functions both take a given REAL to the same real.

see also ABHYANKAR'S CONJECTURE, FINITE GROUP, GROUP

References

Jacobson, N. *Basic Algebra I, 2nd ed.* New York: W. H. Freeman, p. 234, 1985.

Galois Imaginary

A mathematical object invented to solve irreducible CONGRUENCES of the form

$$F(x) \equiv 0 \pmod{p},$$

where p is PRIME.

Galois's Theorem

An algebraic equation is algebraically solvable IFF its GROUP is SOLVABLE. In order that an irreducible equation of PRIME degree be solvable by radicals, it is NECESSARY and SUFFICIENT that all its ROOTS be rational functions of two ROOTS.

see also ABEL'S IMPOSSIBILITY THEOREM, SOLVABLE GROUP

Galois Theory

If there exists a ONE-TO-ONE correspondence between two SUBGROUPS and SUBFIELDS such that

$$\begin{aligned} G(E(G')) &= G' \\ E(G(E')) &= E', \end{aligned}$$

then E is said to have a Galois theory.

Galoisian

An algebraic extension E of F for which every IRREDUCIBLE POLYNOMIAL in F which has a single ROOT in E has *all* its ROOTS in E is said to be Galoisian. Galoisian extensions are also called algebraically normal.

Gambler's Ruin

Let two players each have a finite number of pennies (say, n_1 for player one and n_2 for player two). Now, flip one of the pennies (from either player), with each player having 50% probability of winning, and give the penny to the winner. If the process is repeated indefinitely, the probability that one or the other player will *eventually* lose all his pennies is unity. However, the chances that the individual players will be rendered penniless are

$$\begin{aligned} P_1 &= \frac{n_1}{n_1 + n_2} \\ P_2 &= \frac{n_2}{n_1 + n_2}. \end{aligned}$$

see also COIN TOSSING, MARTINGALE, SAINT PETERSBURG PARADOX

References

Cover, T. M. "Gambler's Ruin: A Random Walk on the Simplex." §5.4 in In *Open Problems in Communications and Computation.* (Ed. T. M. Cover and B. Gopinath). New York: Springer-Verlag, p. 155, 1987.

Hajek, B. "Gambler's Ruin: A Random Walk on the Simplex." §6.3 in In *Open Problems in Communications and Computation.* (Ed. T. M. Cover and B. Gopinath). New York: Springer-Verlag, pp. 204–207, 1987.

Kraitchik, M. "The Gambler's Ruin." §6.20 in *Mathematical Recreations.* New York: W. W. Norton, p. 140, 1942.

Game

A game is defined as a conflict involving gains and losses between two or more opponents who follow formal rules. The study of games belongs to a branch of mathematics known as GAME THEORY.

see also GAME THEORY

Game Expectation

Let the elements in a PAYOFF MATRIX be denoted a_{ij}, where the is are player A's STRATEGIES and the js are player B's STRATEGIES. Player A can get at least

$$\min_{j \leq n} a_{ij} \tag{1}$$

for STRATEGY i. Player B can force player A to get no more than $\max_{j \leq m} a_{ij}$ for a STRATEGY j. The best STRATEGY for player A is therefore

$$\min_{i \leq m} \min_{j \leq n} a_{ij}, \tag{2}$$

and the best STRATEGY for player B is

$$\min_{j \leq n} \max_{i \leq m} a_{ij}. \tag{3}$$

In general,

$$\min_{i\le m}\min_{j\le n} a_{ij} \le \min_{j\le n}\max_{i\le m} a_{ij}. \tag{4}$$

Equality holds only if a SADDLE POINT is present, in which case the quantity is called the VALUE of the game.

see also GAME, PAYOFF MATRIX, SADDLE POINT (GAME), STRATEGY, VALUE

Game of Life

see LIFE

Game Matrix

see PAYOFF MATRIX

Game Theory

A branch of MATHEMATICS and LOGIC which deals with the analysis of GAMES (i.e., situations in which parties are involved in situations where their interests conflict). In addition to the mathematical elegance and complete "solution" which is possible for simple games, the principles of game theory also find applications to complicated games such as cards, checkers, and chess, as well as real-world problems as diverse as economics, property division, politics, and warfare.

see also BOREL DETERMINACY THEOREM, CATEGORICAL GAME, CHECKERS, CHESS, DECISION THEORY, EQUILIBRIUM POINT, FINITE GAME, FUTILE GAME, GAME EXPECTATION, GO, HI-Q, IMPARTIAL GAME, MEX, MINIMAX THEOREM, MIXED STRATEGY, NASH EQUILIBRIUM, NASH'S THEOREM, NIM, NIM-VALUE, PARTISAN GAME, PAYOFF MATRIX, PEG SOLITAIRE, PERFECT INFORMATION, SADDLE POINT (GAME), SAFE, SPRAGUE-GRUNDY FUNCTION, STRATEGY, TACTIX, TIT-FOR-TAT, UNSAFE, VALUE, WYTHOFF'S GAME, ZERO-SUM GAME

References

Berlekamp, E. R.; Conway, J. H; and Guy, R. K. *Winning Ways, For Your Mathematical Plays, Vol. 1: Games in General.* London: Academic Press, 1982.

Berlekamp, E. R.; Conway, J. H; and Guy, R. K. *Winning Ways, For Your Mathematical Plays, Vol. 2: Games in Particular.* London: Academic Press, 1982.

Dresher, M. *The Mathematics of Games of Strategy: Theory and Applications.* New York: Dover, 1981.

Eppstein, D. "Combinatorial Game Theory." `http://www.ics.uci.edu/~eppstein/cgt/`.

Gardner, M. "Game Theory." Ch. 3 in *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, 1978.

Karlin, S. *Mathematical Methods and Theory in Games, Programming, and Economics, 2 Vols. Vol. 1: Matrix Games, Programming, and Mathematical Economics. Vol. 2: The Theory of Infinite Games.* New York: Dover, 1992.

Kuhn, H. W. (Ed.). *Classics in Game Theory.* Princeton, NJ: Princeton University Press, 1997.

McKinsey, J. C. C. *Introduction to the Theory of Games.* New York: McGraw-Hill, 1952.

Neumann, J. von and Morgenstern, O. *Theory of Games and Economic Behavior, 3rd ed.* New York: Wiley, 1964.

Packel, E. *The Mathematics of Games and Gambling.* Washington, DC: Math. Assoc. Amer., 1981.

Straffin, P. D. Jr. *Game Theory and Strategy.* Washington, DC: Math. Assoc. Amer., 1993.

Vajda, S. *Mathematical Games and How to Play Them.* New York: Routledge, 1992.

Walker, P. "An Outline of the History of Game Theory." `http://william-king.www.drexel.edu/top/class/histf.html`.

Williams, J. D. *The Compleat Strategyst, Being a Primer on the Theory of Games of Strategy.* New York: Dover, 1986.

Gamma Distribution

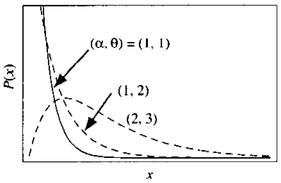

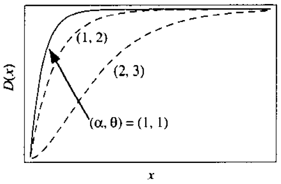

A general type of statistical DISTRIBUTION which is related to the BETA DISTRIBUTION and arises naturally in processes for which the waiting times between POISSON DISTRIBUTED events are relevant. Gamma distributions have two free parameters, labeled α and θ, a few of which are illustrated above.

Given a POISSON DISTRIBUTION with a rate of change λ, the DISTRIBUTION FUNCTION $D(x)$ giving the waiting times until the hth change is

$$\begin{aligned} D(x) &= P(X \le x) = 1 - P(X > x) \\ &= 1 - \sum_{k=0}^{h-1} \frac{(\lambda x)^k e^{-\lambda x}}{k!} \\ &= 1 - e^{-\lambda x} \sum_{k=0}^{h-1} \frac{(\lambda x)^k}{k!} \end{aligned} \tag{1}$$

for $x \ge 0$. The probability function $P(x)$ is then obtained by differentiating $D(x)$,

$$\begin{aligned} P(x) &= D'(x) \\ &= \lambda e^{-\lambda x} \sum_{k=0}^{h-1} \frac{(\lambda x)^k}{k!} - e^{-\lambda x} \sum_{k=0}^{h-1} \frac{k(\lambda x)^{k-1}\lambda}{k!} \\ &= \lambda e^{-\lambda x} + \lambda e^{-\lambda x} \sum_{k=1}^{h-1} \frac{(\lambda x)^k}{k!} - e^{-\lambda x} \sum_{k=1}^{h-1} \frac{k(\lambda x)^{k-1}\lambda}{k!} \\ &= \lambda e^{-\lambda x} - \lambda e^{-\lambda x} \sum_{k=1}^{h-1} \left[\frac{k(\lambda x)^{k-1}}{k!} - \frac{(\lambda x)^k}{k!} \right] \\ &= \lambda e^{-\lambda x} \left\{ 1 - \sum_{k=1}^{h-1} \left[\frac{(\lambda x)^{k-1}}{(k-1)!} - \frac{(\lambda x)^k}{k!} \right] \right\} \\ &= \lambda e^{-\lambda x} \left\{ 1 - \left[1 - \frac{(\lambda x)^{h-1}}{(h-1)!} \right] \right\} = \frac{\lambda(\lambda x)^{h-1}}{(h-1)!} e^{-\lambda x}. \end{aligned} \tag{2}$$

Now let $\alpha \equiv h$ and define $\theta \equiv 1/\lambda$ to be the time between changes. Then the above equation can be written

$$P(x) = \begin{cases} \frac{x^{\alpha-1}e^{-x/\theta}}{\Gamma(\alpha)\theta^\alpha} & 0 \le x < \infty \\ 0 & x < 0. \end{cases} \tag{3}$$

The CHARACTERISTIC FUNCTION describing this distribution is

$$\phi(t) = (1 - it)^{-p}, \tag{4}$$

and the MOMENT-GENERATING FUNCTION is

$$\begin{aligned} M(t) &= \int_0^\infty \frac{e^{tx}x^{\alpha-1}e^{-x/\theta}\,dx}{\Gamma(\alpha)\theta^\alpha} \\ &= \int_0^\infty \frac{x^{\alpha-1}e^{-(1-\theta t)x/\theta}\,dx}{\Gamma(\alpha)\theta^\alpha}. \end{aligned} \tag{5}$$

In order to find the MOMENTS of the distribution, let

$$y \equiv \frac{(1-\theta t)x}{\theta} \tag{6}$$

$$dy = \frac{1-\theta t}{\theta}\,dx, \tag{7}$$

so

$$\begin{aligned} M(t) &= \int_0^\infty \left(\frac{\theta y}{1-\theta t}\right)^{\alpha-1} \frac{e^{-y}}{\Gamma(\alpha)\theta^\alpha}\frac{\theta\,dy}{1-\theta t} \\ &= \frac{1}{(1-\theta t)^\alpha\Gamma(\alpha)}\int_0^\infty y^{\alpha-1}e^{-y}\,dy \\ &= \frac{1}{(1-\theta t)^\alpha}, \end{aligned} \tag{8}$$

and the logarithmic Moment-Generating function is

$$R(t) \equiv \ln M(t) = -\alpha\ln(1-\theta t) \tag{9}$$

$$R'(t) = \frac{\alpha\theta}{1-\theta t} \tag{10}$$

$$R''(t) = \frac{\alpha\theta^2}{(1-\theta t)^2}. \tag{11}$$

The MEAN, VARIANCE, SKEWNESS, and KURTOSIS are then

$$\mu = R'(0) = \alpha\theta \tag{12}$$

$$\sigma^2 = R''(0) = \alpha\theta^2 \tag{13}$$

$$\gamma_1 = \frac{2}{\sqrt{\alpha}} \tag{14}$$

$$\gamma_2 = \frac{6}{\alpha}. \tag{15}$$

The gamma distribution is closely related to other statistical distributions. If $X_1, X_2, \ldots, X_n$ are independent random variates with a gamma distribution having parameters (α_1,θ), (α_2,θ), $\ldots$, (α_n,θ), then $\sum_{i=1}^n X_i$ is distributed as gamma with parameters

$$\alpha = \sum_{i=1}^n \alpha_i \tag{16}$$

$$\theta = \theta. \tag{17}$$

Also, if X_1 and X_2 are independent random variates with a gamma distribution having parameters (α_1,θ) and (α_2,θ), then $X_1/(X_1+X_2)$ is a BETA DISTRIBUTION variate with parameters (α_1,α_2). Both can be derived as follows.

$$P(x,y) = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)}e^{x_1+x_2}x_1^{\alpha_1-1}x_2^{\alpha_2-1}. \tag{18}$$

Let

$$u = x_1 + x_2 \qquad x_1 = uv \tag{19}$$

$$v = \frac{x_1}{x_1+x_2} \qquad x_2 = u(1-v), \tag{20}$$

then the JACOBIAN is

$$J\left(\frac{x_1,x_2}{u,v}\right) = \begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix} = -u, \tag{21}$$

so

$$g(u,v)\,du\,dv = f(x,y)\,dx\,dy = f(x,y)u\,du\,dv. \tag{22}$$

$$\begin{aligned} g(u,v) &= \frac{u}{\Gamma(\alpha_1)\Gamma(\alpha_2)}e^{-u}(uv)^{\alpha_1-1}u^{\alpha_2-1}(1-v)^{\alpha_2-1} \\ &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)}e^{-u}u^{\alpha_1+\alpha_2-1}v^{\alpha_1-1}(1-v)^{\alpha_2-1}. \end{aligned} \tag{23}$$

The sum $X_1 + X_2$ therefore has the distribution

$$f(u) = f(x_1+x_2) = \int_0^1 g(u,v)\,dv = \frac{e^{-u}u^{\alpha_1+\alpha_2-1}}{\Gamma(\alpha_1+\alpha_2)}, \tag{24}$$

which is a gamma distribution, and the ratio $X_1/(X_1+X_2)$ has the distribution

$$\begin{aligned} h(v) = h\left(\frac{x_1}{x_1+x_2}\right) &= \int_0^\infty g(u,v)\,du \\ &= \frac{v^{\alpha_1-1}(1-v)^{\alpha_2-1}}{B(\alpha_1,\alpha_2)}, \end{aligned} \tag{25}$$

where B is the BETA FUNCTION, which is a BETA DISTRIBUTION.

If X and Y are gamma variates with parameters α_1 and α_2, the X/Y is a variate with a BETA PRIME DISTRIBUTION with parameters α_1 and α_2. Let

$$u = x + y \qquad v = \frac{x}{y}, \tag{26}$$

then the JACOBIAN is

$$J\left(\frac{u,v}{x,y}\right) = \begin{vmatrix} 1 & 1 \\ \frac{1}{y} & -\frac{x}{y^2} \end{vmatrix} = -\frac{x+y}{y^2} = -\frac{(1+v)^2}{u}, \quad (27)$$

so

$$dx\,dy = \frac{u}{(1+v)^2}\,du\,dv \quad (28)$$

$$\begin{aligned} g(u,v) &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-u} \left(\frac{uv}{1+v}\right)^{\alpha_1-1} \\ &\qquad \left(\frac{u}{1+v}\right)^{\alpha_2-1} \frac{u}{(1+v)^2} \\ &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-u} u^{\alpha_1+\alpha_2-1} v^{\alpha_2-1} (1+v)^{-\alpha_1-\alpha_2}. \end{aligned} \quad (29)$$

The ratio X/Y therefore has the distribution

$$h(v) = \int_0^\infty g(u,v)\,du = \frac{v^{\alpha_1-1}(1+v)^{-\alpha_1-\alpha_2}}{B(\alpha_1,\alpha_2)}, \quad (30)$$

which is a BETA PRIME DISTRIBUTION with parameters (α_1, α_2).

The "standard form" of the gamma distribution is given by letting $y \equiv x/\theta$, so $dy = dx/\theta$ and

$$\begin{aligned} P(y)\,dy &= \frac{x^{\alpha-1}e^{-x/\theta}}{\Gamma(\alpha)\theta^\alpha}\,dx = \frac{(\theta y)^{\alpha-1}e^{-y}}{\Gamma(\alpha)\theta^\alpha}\,(\theta\,dy) \\ &= \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)}\,dy, \end{aligned} \quad (31)$$

so the MOMENTS about 0 are

$$\nu_r = \frac{1}{\Gamma(\alpha)} \int_0^\infty e^{-x} x^{\alpha-1+r}\,dx = \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)} = (\alpha)_r, \quad (32)$$

where $(\alpha)_r$ is the POCHHAMMER SYMBOL. The MOMENTS about $\mu = \mu_1$ are then

$$\mu_1 = \alpha \quad (33)$$

$$\mu_2 = \alpha \quad (34)$$

$$\mu_3 = 2\alpha \quad (35)$$

$$\mu_4 = 3\alpha^2 + 6\alpha. \quad (36)$$

The MOMENT-GENERATING FUNCTION is

$$M(t) = \frac{1}{(1-t)^\alpha}, \quad (37)$$

and the CUMULANT-GENERATING FUNCTION is

$$K(t) = \alpha \ln(1-t) = \alpha\left(t + \tfrac{1}{2}t^2 + \tfrac{1}{3}t^3 + \ldots\right), \quad (38)$$

so the CUMULANTS are

$$\kappa_r = \alpha\Gamma(r). \quad (39)$$

If x is a NORMAL variate with MEAN μ and STANDARD DEVIATION σ, then

$$y \equiv \frac{(x-\mu)^2}{2\sigma^2} \quad (40)$$

is a standard gamma variate with parameter $\alpha = 1/2$.

see also BETA DISTRIBUTION, CHI-SQUARED DISTRIBUTION

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 534, 1987.

Gamma Function

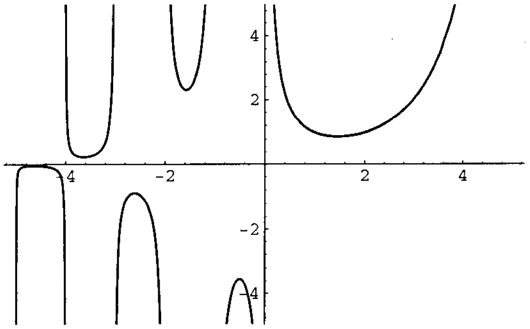

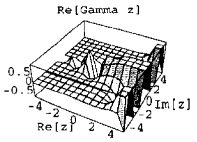

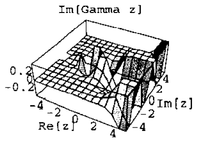

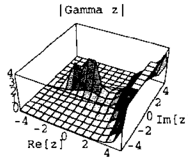

The complete gamma function is defined to be an extension of the FACTORIAL to COMPLEX and REAL NUMBER arguments. It is ANALYTIC everywhere except at $z = 0$, -1, -2, ... It can be defined as a DEFINITE INTEGRAL for $\Re[z] > 0$ (Euler's integral form)

$$\Gamma(z) \equiv \int_0^\infty t^{z-1}e^{-t}\,dt \quad (1)$$

$$= 2\int_0^\infty e^{-t^2} t^{2z-1}\,dt, \quad (2)$$

or

$$\Gamma(z) \equiv \int_0^1 \left[\ln\left(\frac{1}{t}\right)\right]^{z-1} dt. \quad (3)$$

INTEGRATING (1) by parts for a REAL argument, it can be seen that

$$\begin{aligned} \Gamma(x) &= \int_0^\infty t^{x-1}e^{-t}\,dt \\ &= [-t^{x-1}e^{-t}]_0^\infty + \int_0^\infty (x-1)t^{x-2}e^{-t}\,dt \\ &= (x-1)\int_0^\infty t^{x-2}e^{-t}\,dt = (x-1)\Gamma(x-1). \end{aligned} \quad (4)$$

If x is an INTEGER $n = 1, 2, 3, \ldots$ then

$$\begin{aligned}\Gamma(n) &= (n-1)\Gamma(n-1) = (n-1)(n-2)\Gamma(n-2)\\ &= (n-1)(n-2)\cdots 1 = (n-1)!,\end{aligned} \tag{5}$$

so the gamma function reduces to the FACTORIAL for a POSITIVE INTEGER argument.

BINET'S FORMULA is

$$\ln\Gamma(a) = (a-\tfrac{1}{2})\ln a - a + \tfrac{1}{2}\ln(2\pi) + 2\int_0^\infty \frac{\tan\left(\frac{z}{a}\right)}{e^{2\pi z}-1}\,dz \tag{6}$$

for $\Re[a] > 0$ (Whittaker and Watson 1990, p. 251). The gamma function can also be defined by an INFINITE PRODUCT form (Weierstraß Form)

$$\Gamma(z) \equiv \left[ze^{\gamma z}\prod_{r=1}^{\infty}\left(1+\frac{z}{r}\right)e^{-z/r}\right]^{-1}, \tag{7}$$

where γ is the EULER-MASCHERONI CONSTANT. This can be written

$$\Gamma(z) = \frac{1}{z}\exp\left[\sum_{k=1}^{\infty}\frac{(-1)^k s_k}{k}x^k\right], \tag{8}$$

where

$$s_1 \equiv \gamma \tag{9}$$

$$s_k \equiv \zeta(k) \tag{10}$$

for $k \geq 2$, where ζ is the RIEMANN ZETA FUNCTION (Finch). Taking the logarithm of both sides of (7),

$$-\ln[\Gamma(z)] = \ln z + \gamma z + \sum_{n=1}^{\infty}\left[\ln\left(1+\frac{z}{n}\right) - \frac{z}{n}\right]. \tag{11}$$

Differentiating,

$$\begin{aligned}-\frac{\Gamma'(z)}{\Gamma(z)} &= \frac{1}{z} + \gamma + \sum_{n=1}^{\infty}\left(\frac{\frac{1}{n}}{1+\frac{z}{n}} - \frac{1}{n}\right)\\ &= \frac{1}{z} + \gamma + \sum_{n=1}^{\infty}\left(\frac{1}{n+z} - \frac{1}{n}\right)\end{aligned} \tag{12}$$

$$\Gamma'(z) = -\Gamma(z)\left[\frac{1}{z} + \gamma + \sum_{n=1}^{\infty}\left(\frac{1}{n+z} - \frac{1}{n}\right)\right] \tag{13}$$

$$\equiv \Gamma(z)\Psi(z) = \Gamma(z)\psi_0(z) \tag{14}$$

$$\begin{aligned}\Gamma'(1) &= -\Gamma(1) - \{1 + \gamma + [(\tfrac{1}{2} - 1) + (\tfrac{1}{3} - \tfrac{1}{2})\\ &\quad + \ldots + \left(\frac{1}{n+1} - \frac{1}{n}\right) + \ldots]\}\\ &= -(1 + \gamma - 1) = -\gamma\end{aligned} \tag{15}$$

$$\begin{aligned}\Gamma'(n) &= -\Gamma(n)\left\{\frac{1}{n} + \gamma + \left[\left(\frac{1}{1+n} - 1\right) + \left(\frac{1}{2+n} - \frac{1}{2}\right)\right.\right.\\ &\quad \left.\left. + \left(\frac{1}{3+n} - \frac{1}{3}\right) + \ldots\right]\right\}\\ &= -(n-1)!\left(\frac{1}{n} + \gamma - \sum_{k=1}^{n}\frac{1}{k}\right),\end{aligned} \tag{16}$$

where $\Psi(z)$ is the DIGAMMA FUNCTION and $\psi_0(z)$ is the POLYGAMMA FUNCTION. nth derivatives are given in terms of the POLYGAMMA FUNCTIONS $\psi_n, \psi_{n-1}, \ldots, \psi_0$.

The minimum value x_0 of $\Gamma(x)$ for REAL POSITIVE $x = x_0$ is achieved when

$$\Gamma'(x_0) = \Gamma(x_0)\psi_0(x_0) = 0 \tag{17}$$

$$\psi_0(x_0) = 0, \tag{18}$$

This can be solved numerically to give $x_0 = 1.46163\ldots$ (Sloane's A030169), which has CONTINUED FRACTION [1, 2, 6, 63, 135, 1, 1, 1, 1, 4, 1, 38, ...] (Sloane's A030170). At x_0, $\Gamma(x_0)$ achieves the value 0.8856031944... (Sloane's A030171), which has CONTINUED FRACTION [0, 1, 7, 1, 2, 1, 6, 1, 1, ...] (Sloane's A030172).

The Euler limit form is

$$\begin{aligned}\frac{1}{\Gamma(z)} &= z\left[\lim_{m\to\infty} e^{(1+1/2+\ldots+1/m-\ln m)z}\right]\\ &\quad \times\left[\lim_{m\to\infty}\prod_{n=1}^{m}\left\{\left(1+\frac{z}{n}\right)e^{-z/n}\right\}\right]\\ &= \frac{1}{z}\prod_{n=1}^{\infty}\left[\left(1+\frac{1}{n}\right)^z\left(1+\frac{z}{n}\right)^{-1}\right],\end{aligned} \tag{19}$$

so

$$\Gamma(z) \equiv \lim_{n\to\infty}\frac{1\cdot 2\cdot 3\cdots n}{z(z+1)(z+2)\cdots(z+n)}n^z. \tag{20}$$

The LANCZOS APPROXIMATION for $z > 0$ is

$$\begin{aligned}\Gamma(z+1) &= (z+\gamma+\tfrac{1}{2})^{z+1/2}e^{z+\gamma+1/2}\sqrt{2\pi}\\ &\quad \times\left[c_0 + \frac{c_1}{z+1} + \frac{c_2}{z+2} + \ldots + \frac{c_n}{z+n} + \epsilon\right].\end{aligned} \tag{21}$$

The complete gamma function $\Gamma(x)$ can be generalized to the incomplete gamma function $\Gamma(x, a)$ such that $\Gamma(x) = \Gamma(x, 0)$. The gamma function satisfies the recurrence relations

$$\Gamma(1+z) = z\Gamma(z) \tag{22}$$

$$\Gamma(1-z) = -z\Gamma(-z). \tag{23}$$

Additional identities are

$$\Gamma(x)\Gamma(-x) = -\frac{\pi}{x\sin(\pi x)} \tag{24}$$

$$\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin(\pi x)} \tag{25}$$

$$\begin{aligned}\ln[\Gamma(x+iy+1)] &= \ln(x^2+y^2) + i\tan^{-1}\left(\frac{y}{x}\right)\\ &\quad + \ln[\Gamma(x+iy)]\end{aligned} \tag{26}$$

$$|(ix)!|^2 = \frac{\pi x}{\sinh(\pi x)} \tag{27}$$

$$|(n+ix)!| = \sqrt{\frac{\pi x}{\sinh(\pi x)}}\prod_{s=1}^{n}\sqrt{s^2+x^2}. \tag{28}$$

For integral arguments, the first few values are 1, 1, 2, 6, 24, 120, 720, 5040, 40320, 362880, ... (Sloane's A000142). For half integral arguments,

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi} \quad (29)$$

$$\Gamma(\tfrac{3}{2}) = \tfrac{1}{2}\sqrt{\pi} \quad (30)$$

$$\Gamma(\tfrac{5}{2}) = \tfrac{3}{4}\sqrt{\pi}. \quad (31)$$

In general, for m a Positive Integer $m = 1, 2, \ldots$

$$\begin{aligned}\Gamma(\tfrac{1}{2}+m) &= \frac{1\cdot 3\cdot 5\cdots(2m-1)}{2^m}\sqrt{\pi} \\ &= \frac{(2m-1)!!}{2^m}\sqrt{\pi} \quad (32)\end{aligned}$$

$$\begin{aligned}\Gamma(\tfrac{1}{2}-m) &= \frac{(-1)^m 2^m}{1\cdot 3\cdot 5\cdots(2m-1)}\sqrt{\pi} \\ &= \frac{(-1)^m 2^m}{(2m-1)!!}\sqrt{\pi}. \quad (33)\end{aligned}$$

For $\Re[x] = -\frac{1}{2}$,

$$|(-\tfrac{1}{2}+iy)!|^2 = \frac{\pi}{\cosh(\pi y)}. \quad (34)$$

Gamma functions of argument $2z$ can be expressed using the Legendre Duplication Formula

$$\Gamma(2z) = (2\pi)^{-1/2}2^{2z-1/2}\Gamma(z)\Gamma(z+\tfrac{1}{2}). \quad (35)$$

Gamma functions of argument $3z$ can be expressed using a triplication Formula

$$\Gamma(3z) = (2\pi)^{-1}3^{3z-1/2}\Gamma(z)\Gamma(z+\tfrac{1}{3})\Gamma(z+\tfrac{2}{3}). \quad (36)$$

The general result is the Gauss Multiplication Formula

$$\Gamma(z)\Gamma(z+\tfrac{1}{n})\cdots\Gamma(z+\tfrac{n+1}{n}) = (2\pi)^{(n-1)/2}n^{1/2-nz}\Gamma(nz). \quad (37)$$

The gamma function is also related to the Riemann Zeta Function ζ by

$$\Gamma\left(\frac{s}{2}\right)\pi^{-2/2}\zeta(s) = \Gamma\left(\frac{1-s}{2}\right)\pi^{-(1-s)/2}\zeta(1-s). \quad (38)$$

Borwein and Zucker (1992) give a variety of identities relating gamma functions to square roots and Elliptic Integral Singular Values k_n, i.e., Moduli k_n such that

$$\frac{K'(k_n)}{K(k_n)} = \sqrt{n}, \quad (39)$$

where $K(k)$ is a complete Elliptic Integral of the First Kind and $K'(k) = K(k') = K(\sqrt{1-k^2})$ is the complementary integral.

$$\Gamma(\tfrac{1}{3}) = 2^{7/9}3^{-1/12}\pi^{1/3}[K(k_3)]^{1/3} \quad (40)$$

$$\Gamma(\tfrac{1}{4}) = 2\pi^{1/4}[K(k_1)]^{1/2} \quad (41)$$

$$\Gamma(\tfrac{1}{6}) = 2^{-1/3}3^{1/2}\pi^{-1/2}[\Gamma(\tfrac{1}{3})]^2 \quad (42)$$

$$\Gamma(\tfrac{1}{8})\Gamma(\tfrac{3}{8}) = (\sqrt{2}-1)^{1/2}2^{13/4}\pi^{1/2}K(k_2) \quad (43)$$

$$\frac{\Gamma(\frac{1}{8})}{\Gamma(\frac{3}{8})} = 2(\sqrt{2}+1)^{1/2}\pi^{-1/4}[K(k_1)]^{1/2} \quad (44)$$

$$\Gamma(\tfrac{1}{12}) = 2^{-1/4}3^{3/8}(\sqrt{3}+1)^{1/2}\pi^{-1/2}\Gamma(\tfrac{1}{4})\Gamma(\tfrac{1}{3}) \quad (45)$$

$$\Gamma(\tfrac{5}{12}) = 2^{1/4}3^{-1/8}(\sqrt{3}-1)^{1/2}\pi^{1/2}\frac{\Gamma(\frac{1}{4})}{\Gamma(\frac{1}{3})} \quad (46)$$

$$\frac{\Gamma(\frac{1}{24})\Gamma(\frac{11}{24})}{\Gamma(\frac{5}{24})\Gamma(\frac{7}{24})} = \sqrt{3}\sqrt{2+\sqrt{3}} \quad (47)$$

$$\frac{\Gamma(\frac{1}{24})\Gamma(\frac{5}{24})}{\Gamma(\frac{7}{24})\Gamma(\frac{11}{24})} = 4\cdot 3^{1/4}(\sqrt{3}+\sqrt{2})\pi^{-1/2}K(k_1) \quad (48)$$

$$\frac{\Gamma(\frac{1}{24})\Gamma(\frac{7}{24})}{\Gamma(\frac{5}{24})\Gamma(\frac{11}{24})} = 2^{25/18}3^{1/3}(\sqrt{2}+1)\pi^{-1/3}[K(k_3)]^{2/3} \quad (49)$$

$$\begin{aligned}&\Gamma(\tfrac{1}{24})\Gamma(\tfrac{5}{25})\Gamma(\tfrac{7}{24})\Gamma(\tfrac{11}{24}) \\ &\quad = 384(\sqrt{2}+1)(\sqrt{3}-\sqrt{2})(2-\sqrt{3})\pi[K(k_6)]^2 \quad (50)\end{aligned}$$

$$\Gamma(\tfrac{1}{10}) = 2^{-7/10}5^{1/4}(\sqrt{5}+1)^{1/2}\pi^{-1/2}\Gamma(\tfrac{1}{5})\Gamma(\tfrac{2}{5}) \quad (51)$$

$$\Gamma(\tfrac{3}{10}) = 2^{-3/5}(\sqrt{5}-1)\pi^{1/2}\frac{\Gamma(\frac{1}{5})}{\Gamma(\frac{2}{5})} \quad (52)$$

$$\frac{\Gamma(\frac{1}{15})\Gamma(\frac{4}{15})\Gamma(\frac{7}{15})}{\Gamma(\frac{2}{15})} = 2\cdot 3^{1/2}5^{1/6}\sin(\tfrac{2}{15}\pi)[\Gamma(\tfrac{1}{3})]^2 \quad (53)$$

$$\begin{aligned}&\frac{\Gamma(\frac{1}{15})\Gamma(\frac{2}{15})\Gamma(\frac{7}{15})}{\Gamma(\frac{4}{15})} \\ &\quad = 2^2\cdot 3^{2/5}\sin(\tfrac{1}{5}\pi)\sin(\tfrac{4}{15}\pi)[\Gamma(\tfrac{1}{5})]^2 \quad (54)\end{aligned}$$

$$\begin{aligned}&\frac{\Gamma(\frac{2}{15})\Gamma(\frac{4}{15})\Gamma(\frac{7}{15})}{\Gamma(\frac{1}{15})} \\ &\quad = \frac{2^{-3/2}3^{-1/5}5^{1/4}(\sqrt{5}-1)^{1/2}[\Gamma(\frac{2}{5})]^2}{\sin(\frac{4}{15}\pi)} \quad (55)\end{aligned}$$

$$\frac{\Gamma(\frac{1}{15})\Gamma(\frac{2}{15})\Gamma(\frac{4}{15})}{\Gamma(\frac{7}{15})} = 60(\sqrt{5}-1)\sin(\tfrac{7}{15}\pi)[K(k_{15})]^2 \quad (56)$$

$$\frac{\Gamma(\frac{1}{20})\Gamma(\frac{9}{20})}{\Gamma(\frac{3}{20})\Gamma(\frac{7}{20})} = 2^{-1}5^{1/4}(\sqrt{5}+1) \quad (57)$$

$$\begin{aligned}&\frac{\Gamma(\frac{1}{20})\Gamma(\frac{3}{20})}{\Gamma(\frac{7}{20})\Gamma(\frac{9}{20})} \\ &\quad = 2^{4/5}(10-2\sqrt{5})^{1/2}\pi^{-1}\sin(\tfrac{7}{20}\pi)\sin(\tfrac{9}{20}\pi)[\Gamma(\tfrac{1}{5})]^2 \quad (58)\end{aligned}$$

$$\frac{\Gamma(\frac{1}{20})\Gamma(\frac{7}{20})}{\Gamma(\frac{3}{20})\Gamma(\frac{9}{20})}$$

$$= 2^{3/5}(10+2\sqrt{5}\,)^{1/2}\pi^{-1}\sin(\tfrac{3}{20}\pi)\sin(\tfrac{9}{20}\pi)[\Gamma(\tfrac{2}{5})]^2 \quad (59)$$

$$\Gamma(\tfrac{1}{20})\Gamma(\tfrac{3}{20})\Gamma(\tfrac{7}{20})\Gamma(\tfrac{9}{20}) = 160(\sqrt{5}-2)^{1/2}\pi[K(k_5)]^2. \quad (60)$$

A few curious identities include

$$\prod_{n=1}^{8}\Gamma(\tfrac{1}{3}n) = \frac{640}{3^6}\left(\frac{\pi}{\sqrt{3}}\right)^3 \quad (61)$$

$$\frac{[\Gamma(\frac{1}{4})]^4}{16\pi^2} = \frac{3^2}{3^2-1}\,\frac{5^2-1}{5^2}\,\frac{7^2}{7^2-1}\cdots \quad (62)$$

$$\frac{\Gamma'(1)}{\Gamma(1)} - \frac{\Gamma'(\frac{1}{2})}{\Gamma(\frac{1}{2})} = 2\ln 2 \quad (63)$$

(Magnus and Oberhettinger 1949, p. 1). Ramanujan also gave a number of fascinating identities:

$$\frac{\Gamma^2(n+1)}{\Gamma(n+xi+1)\Gamma(n-xi+1)} = \prod_{k=1}^{\infty}\left[1+\frac{x^2}{(n+k)^2}\right] \quad (64)$$

$$\phi(m,n)\phi(n,m) = \frac{\Gamma^3(m+1)\Gamma^3(n+1)}{\Gamma(2m+n+1)\Gamma(2n+m+1)} \times\frac{\cosh[\pi(m+n)\sqrt{3}\,]-\cos[\pi(m-n)]}{2\pi^2(m^2+mn+n^2)}, \quad (65)$$

where

$$\phi(m,n) \equiv \prod_{k=1}^{\infty}\left[1+\left(\frac{m+n}{k+m}\right)^3\right], \quad (66)$$

$$\prod_{k=1}^{\infty}\left[1+\left(\frac{n}{k}\right)^3\right]\prod_{k=1}^{\infty}\left[1+3\left(\frac{n}{n+2k}\right)^2\right] = \frac{\Gamma(\frac{1}{2}n)}{\Gamma[\frac{1}{2}(n+1)]}\,\frac{\cosh(\pi n\sqrt{3}\,)-\cos(\pi n)}{2^{n+2}\pi^{3/2}n} \quad (67)$$

(Berndt 1994).

The following ASYMPTOTIC SERIES is occasionally useful in probability theory (e.g., the 1-D RANDOM WALK):

$$\frac{\Gamma(J+\frac{1}{2})}{\Gamma(J)} = \sqrt{J}\left(1-\frac{1}{8J}+\frac{1}{128J^2} +\frac{5}{1024J^3}-\frac{21}{32768J^4}+\ldots\right) \quad (68)$$

(Graham *et al.* 1994). This series also gives a nice asymptotic generalization of STIRLING NUMBERS OF THE FIRST KIND to fractional values.

It has long been known that $\Gamma(\frac{1}{4})\pi^{-1/4}$ is TRANSCENDENTAL (Davis 1959), as is $\Gamma(\frac{1}{3})$ (Le Lionnais 1983), and Chudnovsky has apparently recently proved that $\Gamma(\frac{1}{4})$ is itself TRANSCENDENTAL.

The upper incomplete gamma function is given by

$$\Gamma(a,x) \equiv \int_x^{\infty} t^{a-1}e^{-t}\,dt = 1-\gamma(a,x), \quad (69)$$

where γ is the lower incomplete gamma function. For a an INTEGER n

$$\Gamma(n,x) = (n-1)!e^{-x}\sum_{s=0}^{n-1}\frac{x^s}{s!} = (n-1)!e^{-x}\,\mathrm{es}_{n-1}(x), \quad (70)$$

where es is the EXPONENTIAL SUM FUNCTION. The lower incomplete gamma function is given by

$$\begin{aligned}\gamma(a,x) \equiv \Gamma(a)-\Gamma(a,x) &= \int_0^x e^{-t}t^{a-1}\,dt \\ &= a^{-1}x^a e^{-x}{}_1F_1(1;1+a;x) \\ &= a^{-1}x^a{}_1F_1(a;1+a;-x), \end{aligned} \quad (71)$$

where ${}_1F_1(a;b;x)$ is the CONFLUENT HYPERGEOMETRIC FUNCTION OF THE FIRST KIND. For a an INTEGER n,

$$\begin{aligned}\gamma(n,x) &= (n-1)!\left(1-e^{-x}\sum_{s=0}^{n-1}\frac{x^s}{s!}\right) \\ &= (n-1)!\,[1-\mathrm{es}_{n-1}(x)]. \end{aligned} \quad (72)$$

The function $\Gamma(a,z)$ is denoted `Gamma[a,z]` and the function $\gamma(a,z)$ is denoted `Gamma[a,0,z]` in *Mathematica*® (Wolfram Research, Champaign, IL).

see also DIGAMMA FUNCTION, DOUBLE GAMMA FUNCTION, FRANSÉN-ROBINSON CONSTANT, *G*-FUNCTION, GAUSS MULTIPLICATION FORMULA, LAMBDA FUNCTION, LEGENDRE DUPLICATION FORMULA, MU FUNCTION, NU FUNCTION, PEARSON'S FUNCTION, POLYGAMMA FUNCTION, REGULARIZED GAMMA FUNCTION, STIRLING'S SERIES

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Gamma (Factorial) Function" and "Incomplete Gamma Function." §6.1 and 6.5 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 255–258 and 260–263, 1972.

Arfken, G. "The Gamma Function (Factorial Function)." Ch. 10 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 339–341 and 539–572, 1985.

Artin, E. *The Gamma Function.* New York: Holt, Rinehart, and Winston, 1964.

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 334–342, 1994.

Borwein, J. M. and Zucker, I. J. "Elliptic Integral Evaluation of the Gamma Function at Rational Values of Small Denominator." *IMA J. Numerical Analysis* **12**, 519–526, 1992.

Davis, H. T. *Tables of the Higher Mathematical Functions.* Bloomington, IN: Principia Press, 1933.

Davis, P. J. "Leonhard Euler's Integral: A Historical Profile of the Gamma Function." *Amer. Math. Monthly* **66**, 849–869, 1959.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/fran/fran.html`.
Graham, R. L.; Knuth, D. E.; and Patashnik, O. Answer to problem 9.60 in *Concrete Mathematics: A Foundation for Computer Science*. Reading, MA: Addison-Wesley, 1994.
Le Lionnais, F. *Les nombres remarquables*. Paris: Hermann, p. 46, 1983.
Magnus, W. and Oberhettinger, F. *Formulas and Theorems for the Special Functions of Mathematical Physics*. New York: Chelsea, 1949.
Nielsen, H. *Die Gammafunktion*. New York: Chelsea, 1965.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Gamma Function, Beta Function, Factorials, Binomial Coefficients" and "Incomplete Gamma Function, Error Function, Chi-Square Probability Function, Cumulative Poisson Function." §6.1 and 6.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 206–209 and 209–214, 1992.
Sloane, N. J. A. Sequences A030169, A030170, A030171, A030172, and A000142/M1675 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Spanier, J. and Oldham, K. B. "The Gamma Function $\Gamma(x)$" and "The Incomplete Gamma $\gamma(\nu;x)$ and Related Functions." Chs. 43 and 45 in *An Atlas of Functions*. Washington, DC: Hemisphere, pp. 411–421 and 435–443, 1987.
Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Gamma Group

The gamma group Γ is the set of all transformations w of the form

$$w(t) = \frac{at+b}{ct+d},$$

where a, b, c, and d are Integers and $ad - bc = 1$.

see also Klein's Absolute Invariant, Lambda Group, Theta Function

References

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity*. New York: Wiley, pp. 127–132, 1987.

Gamma-Modular

see Modular Gamma Function

Gamma Statistic

$$\gamma_r \equiv \frac{\kappa_r}{\sigma^{r+2}},$$

where κ_r are Cumulants and σ is the Standard Deviation.

see also Kurtosis, Skewness

Garage Door

see Astroid

Gårding's Inequality

Gives a lower bound for the inner product (Lu, u), where L is a linear elliptic Real differential operator of order m, and u has compact support.

References

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Garman-Kohlhagen Formula

$$V_t = e^{-y\tau} S_t N(d_1) - e^{-r\tau} KN(d_2),$$

where N is the cumulative Normal Distribution and

$$d_1, d_2 = \frac{\log\left(\frac{S_t}{K}\right) + (r - y \pm \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}.$$

If $y = 0$, this is the standard form of the Black-Scholes formula.

see also Black-Scholes Theory

References

Garman, M. B. and Kohlhagen, S. W. "Foreign Currency Option Values." *J. International Money and Finance* **2**, 231–237, 1983.
Price, J. F. "Optional Mathematics is Not Optional." *Not. Amer. Math. Soc.* **43**, 964–971, 1996.

Gate Function

Bracewell's term for the Rectangle Function.

References

Bracewell, R. *The Fourier Transform and Its Applications*. New York: McGraw-Hill, 1965.

Gauche Conic

see Skew Conic

Gaullist Cross

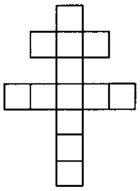

A Cross also called the Cross of Lorraine or Patriarchal Cross.

see also Cross, Dissection

Gauss's Backward Formula

$$f_p = f_0 + p\delta_{-1/2} + G_2^*\delta_0^2 + G_3\delta_{-1/2}^3 + G_4^*\delta_0^4 + G_5\delta_{-1/2}^5 + \ldots,$$

for $p \in [0,1]$, where δ is the Central Difference and

$$G_{2n}^* = \binom{p+n}{2n}$$

$$G_{2n+1} = \binom{p+n}{2n+1},$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT.

see also CENTRAL DIFFERENCE, GAUSS'S FORWARD FORMULA

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 433, 1987.

Gauss-Bodenmiller Theorem

The CIRCLES on the DIAGONALS of a COMPLETE QUADRILATERAL as DIAMETERS are COAXAL. Furthermore, the ORTHOCENTERS of the four TRIANGLES of a COMPLETE QUADRILATERAL are COLLINEAR on the RADICAL AXIS of the COAXAL CIRCLES.

see also COAXAL CIRCLES, COLLINEAR, COMPLETE QUADRILATERAL, DIAGONAL (POLYGON), ORTHOCENTER, RADICAL AXIS

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, p. 172, 1929.

Gauss-Bolyai-Lobachevsky Space

A non-Euclidean space with constant NEGATIVE GAUSSIAN CURVATURE.

see also LOBACHEVSKY-BOLYAI-GAUSS GEOMETRY, NON-EUCLIDEAN GEOMETRY

Gauss-Bonnet Formula

The Gauss-Bonnet formula has several formulations. The simplest one expresses the total GAUSSIAN CURVATURE of an embedded triangle in terms of the total GEODESIC CURVATURE of the boundary and the JUMP ANGLES at the corners.

More specifically, if M is any 2-D RIEMANNIAN MANIFOLD (like a surface in 3-space) and if T is an embedded triangle, then the Gauss-Bonnet formula states that the integral over the whole triangle of the GAUSSIAN CURVATURE with respect to AREA is given by 2π minus the sum of the JUMP ANGLES minus the integral of the GEODESIC CURVATURE over the whole of the boundary of the triangle (with respect to ARC LENGTH),

$$\iint_T K\,dA = 2\pi - \sum \alpha_i - \int_{\partial T} \kappa_g\,ds, \tag{1}$$

where K is the GAUSSIAN CURVATURE, dA is the AREA measure, the α_is are the JUMP ANGLES of ∂T, and κ_g is the GEODESIC CURVATURE of ∂T, with ds the ARC LENGTH measure.

The next most common formulation of the Gauss-Bonnet formula is that for any compact, boundaryless 2-D RIEMANNIAN MANIFOLD, the integral of the GAUSSIAN CURVATURE over the entire MANIFOLD with respect to AREA is 2π times the EULER CHARACTERISTIC of the MANIFOLD,

$$\iint_M K\,dA = 2\pi\chi(M). \tag{2}$$

This is somewhat surprising because the total GAUSSIAN CURVATURE is differential-geometric in character, but the EULER CHARACTERISTIC is topological in character and does not depend on differential geometry at all. So if you distort the surface and change the curvature at any location, regardless of how you do it, the same total curvature is maintained.

Another way of looking at the Gauss-Bonnet theorem for surfaces in 3-space is that the GAUSS MAP of the surface has DEGREE given by half the EULER CHARACTERISTIC of the surface

$$\iint_M K\,dA = 2\pi\chi(M) - \sum \alpha_i - \int_{\partial M} \kappa_g\,ds, \tag{3}$$

which works only for ORIENTABLE SURFACES. This makes the Gauss-Bonnet theorem a simple consequence of the POINCARE-HOPF INDEX THEOREM, which is a nice way of looking at things if you're a topologist, but not so nice for a differential geometer. This proof can be found in Guillemin and Pollack (1974). Millman and Parker (1977) give a standard differential-geometric proof of the Gauss-Bonnet theorem, and Singer and Thorpe (1996) give a GAUSS'S THEOREMA EGREGIUM-inspired proof which is entirely intrinsic, without any reference to the ambient EUCLIDEAN SPACE.

A general Gauss-Bonnet formula that takes into account both formulas can also be given. For any compact 2-D RIEMANNIAN MANIFOLD with corners, the integral of the GAUSSIAN CURVATURE over the 2-MANIFOLD with respect to AREA is 2π times the EULER CHARACTERISTIC of the MANIFOLD minus the sum of the JUMP ANGLES and the total GEODESIC CURVATURE of the boundary.

References

Chavel, I. *Riemannian Geometry: A Modern Introduction.* New York: Cambridge University Press, 1994.

Guillemin, V. and Pollack, A. *Differential Topology.* Englewood Cliffs, NJ: Prentice-Hall, 1974.

Millman, R. S. and Parker, G. D. *Elements of Differential Geometry.* Prentice-Hall, 1977.

Reckziegel, H. In *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, p. 31, 1986.

Singer, I. M. and Thorpe, J. A. *Lecture Notes on Elementary Topology and Geometry.* New York: Springer-Verlag, 1996.

Gauss-Bonnet Theorem

see GAUSS-BONNET FORMULA

Gauss's Circle Problem

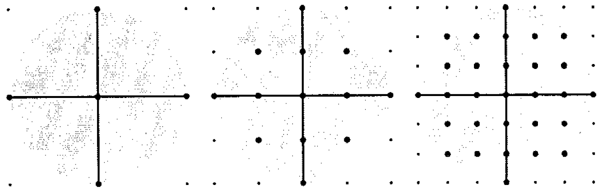

Count the number of LATTICE POINTS $N(r)$ inside the boundary of a CIRCLE of RADIUS r with center at the origin. The exact solution is given by the SUM

$$N(r) = 1 + 4\lfloor r\rfloor + 4\sum_{i=1}^{\lfloor r\rfloor}\left\lfloor\sqrt{r^2-i^2}\right\rfloor. \tag{1}$$

The first few values are 1, 5, 13, 29, 49, 81, 113, 149, ... (Sloane's A000328).

Gauss showed that

$$N(r) = \pi r^2 + E(r), \tag{2}$$

where

$$|E(r)| \leq 2\sqrt{2}\,\pi r. \tag{3}$$

Writing $|E(r)| \leq Cr^\theta$, the best bounds on θ are $1/2 < \theta \leq 46/73 \approx 0.630137$ (Huxley 1990). The problem has also been extended to CONICS and higher dimensions. The limit 1/2 was obtained by Hardy and Landau (1915), and the limit 46/73 improves previous values of $24/37 \approx 0.64864$ (Cheng 1963) and $34/53 \approx 0.64150$ (Vinogradov), and $7/11 \approx 0.63636$.

see also CIRCLE LATTICE POINTS

References

Cheng, J. R. "The Lattice Points in a Circle." *Sci. Sinica* **12**, 633–649, 1963.

Cilleruello, J. "The Distribution of Lattice Points on Circles." *J. Number Th.* **43**, 198–202, 1993.

Guy, R. K. "Gauß's Lattice Point Problem." §F1 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 240–2417, 1994.

Huxley, M. N. "Exponential Sums and Lattice Points." *Proc. London Math. Soc.* **60**, 471–502, 1990.

Huxley, M. N. "Corrigenda: 'Exponential Sums and Lattice Points'." *Proc. London Math. Soc.* **66**, 70, 1993.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 24, 1983.

Sloane, N. J. A. Sequence A000328/M3829 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Weisstein, E. W. "Circle Lattice Points." `http://www.astro.virginia.edu/~eww6n/math/notebooks/CircleLatticePoints.m`.

Gauss's Class Number Conjecture

In his monumental treatise *Disquisitiones Arithmeticae,* Gauss conjectured that the CLASS NUMBER $h(-d)$ of an IMAGINARY quadratic field with DISCRIMINANT $-d$ tends to infinity with d. A proof was finally given by Heilbronn (1934), and Siegel (1936) showed that for any $\epsilon > 0$, there exists a constant $c_\epsilon > 0$ such that

$$h(-d) > c_\epsilon d^{1/2-\epsilon}$$

as $d \to \infty$. However, these results were not effective in actually determining the values for a given m of a complete list of fundamental discriminants $-d$ such that $h(-d) = m$, a problem known as GAUSS'S CLASS NUMBER PROBLEM.

Goldfeld (1976) showed that if there exists a "Weil curve" whose associated DIRICHLET L-SERIES has a zero of at least third order at $s = 1$, then for any $\epsilon > 0$, there exists an effectively computable constant c_ϵ such that

$$h(-d) > c_\epsilon(\ln d)^{1-\epsilon}.$$

Gross and Zaiger (1983) showed that certain curves must satisfy the condition of Goldfeld, and Goldfeld's proof was simplified by Oesterlé (1985).

see also CLASS NUMBER, GAUSS'S CLASS NUMBER PROBLEM, HEEGNER NUMBER

References

Arno, S.; Robinson, M. L.; and Wheeler, F. S. "Imaginary Quadratic Fields with Small Odd Class Number." `http://www.math.uiuc.edu/Algebraic-Number-Theory/0009/`.

Böcherer, S. "Das Gauß'sche Klassenzahlproblem." *Mitt.. Math. Ges. Hamburg* **11**, 565–589, 1988.

Gauss, C. F. *Disquisitiones Arithmeticae.* New Haven, CT: Yale University Press, 1966.

Goldfeld, D. M. "The Class Number of Quadratic Fields and the Conjectures of Birch and Swinnerton-Dyer." *Ann. Scuola Norm. Sup. Pisa* **3**, 623–663, 1976.

Gross, B. and Zaiger, D. "Points de Heegner et derivées de fonctions L." *C. R. Acad. Sci. Paris* **297**, 85–87, 1983.

Heilbronn, H. "On the Class Number in Imaginary Quadratic Fields." *Quart. J. Math. Oxford Ser.* **25**, 150–160, 1934.

Oesterlé, J. "Nombres de classes des corps quadratiques imaginaires." *Astérique* **121-122**, 309–323, 1985.

Siegel, C. L. "Über die Klassenzahl quadratischer Zahlkörper." *Acta. Arith.* **1**, 83–86, 1936.

Gauss's Class Number Problem

For a given m, determine a complete list of fundamental DISCRIMINANTS $-d$ such that the CLASS NUMBER is given by $h(-d) = m$. Heegner (1952) gave a solution for $m = 1$, but it was not completely accepted due to a number of apparent gaps. However, subsequent examination of Heegner's proof show it to be "essentially" correct (Conway and Guy 1996). Conway and Guy (1996) therefore call the nine values of $n(-d)$ having $h(-d) = 1$ where $-d$ is the DISCRIMINANT corresponding to a QUADRATIC FIELD $a + b\sqrt{-n}$ ($n = -1, -2, -3, -7, -11, -19, -43, -67$, and -163; Sloane's A003173) the HEEGNER NUMBERS. The HEEGNER NUMBERS have a number of fascinating properties.

Stark (1967) and Baker (1966) gave independent proofs of the fact that only nine such numbers exist; both proofs were accepted. Baker (1971) and Stark (1975) subsequently and independently solved the generalized class number problem completely for $m = 2$. Oesterlé (1985) solved the case $m = 3$, and Arno (1992) solved the case $m = 4$. Wagner (1996) solve the cases $n = 5$, 6, and 7. Arno *et al.* (1993) solved the problem for ODD m satisfying $5 \leq m \leq 23$.

see also Class Number, Gauss's Class Number Conjecture, Heegner Number

References

Arno, S. "The Imaginary Quadratic Fields of Class Number 4." *Acta Arith.* **40**, 321–334, 1992.

Arno, S.; Robinson, M. L.; and Wheeler, F. S. "Imaginary Quadratic Fields with Small Odd Class Number." Dec. 1993. `http://www.math.uiuc.edu/Algebraic-Number-Theory/0009/`.

Baker, A. "Linear Forms in the Logarithms of Algebraic Numbers. I." *Mathematika* **13**, 204–216, 1966.

Baker, A. "Imaginary Quadratic Fields with Class Number 2." *Ann. Math.* **94**, 139–152, 1971.

Conway, J. H. and Guy, R. K. "The Nine Magic Discriminants." In *The Book of Numbers.* New York: Springer-Verlag, pp. 224–226, 1996.

Goldfeld, D. M. "Gauss' Class Number Problem for Imaginary Quadratic Fields." *Bull. Amer. Math. Soc.* **13**, 23–37, 1985.

Heegner, K. "Diophantische Analysis und Modulfunktionen." *Math. Z.* **56**, 227–253, 1952.

Heilbronn, H. A. and Linfoot, E. H. "On the Imaginary Quadratic Corpora of Class-Number One." *Quart. J. Math. (Oxford)* **5**, 293–301, 1934.

Lehmer, D. H. "On Imaginary Quadratic Fields whose Class Number is Unity." *Bull. Amer. Math. Soc.* **39**, 360, 1933.

Montgomery, H. and Weinberger, P. "Notes on Small Class Numbers." *Acta. Arith.* **24**, 529–542, 1974.

Oesterlé, J. "Nombres de classes des corps quadratiques imaginaires." *Astérique* **121–122**, 309–323, 1985.

Oesterlé, J. "Le problème de Gauss sur le nombre de classes." *Enseign Math.* **34**, 43–67, 1988.

Serre, J.-P. $\Delta = b^2 - 4ac$." *Math. Medley* **13**, 1–10, 1985.

Shanks, D. "On Gauss's Class Number Problems." *Math. Comput.* **23**, 151–163, 1969.

Sloane, N. J. A. Sequence A003173/M0827 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Stark, H. M. "A Complete Determination of the Complex Quadratic Fields of Class Number One." *Michigan Math. J.* **14**, 1–27, 1967.

Stark, H. M. "On Complex Quadratic Fields with Class Number Two." *Math. Comput.* **29**, 289–302, 1975.

Wagner, C. "Class Number 5, 6, and 7." *Math. Comput.* **65**, 785–800, 1996.

Gauss's Constant

The Reciprocal of the Arithmetic-Geometric Mean of 1 and $\sqrt{2}$,

$$\begin{aligned} \frac{1}{M(1,\sqrt{2})} &= \frac{2}{\pi}\int_0^1 \frac{1}{\sqrt{1-x^4}}\,dx && (1) \\ &= \frac{2}{\pi}\int_0^{\pi/2} \frac{d\theta}{\sqrt{1+\sin^2\theta}} && (2) \\ &= \frac{\sqrt{2}}{\pi}K\left(\frac{1}{\sqrt{2}}\right) \\ &= \frac{1}{(2\pi)^{3/2}}[\Gamma(\tfrac{1}{4})]^2 && (3) \\ &= 0.83462684167\ldots, && (4) \end{aligned}$$

where $K(k)$ is the complete Elliptic Integral of the First Kind and $\Gamma(z)$ is the Gamma Function.

see also Arithmetic-Geometric Mean, Gauss-Kuzmin-Wirsing Constant

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/gauss/gauss.html`.

Gauss's Criterion

Let p be an Odd Prime and b a Positive Integer not divisible by p. Then for each Positive Odd Integer $2k-1 < p$, let r_i be

$$r_k \equiv (2k-1)b \pmod{p}$$

with $0 < r_k < p$, and let t be the number of Even r_is. Then

$$(b/p) = (-1)^t,$$

where (b/p) is the Legendre Symbol.

References

Shanks, D. "Gauss's Criterion." §1.17 in *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 38–40, 1993.

Gauss's Double Point Theorem

If a sequence of Double Points is passed as a Closed Curve is traversed, each Double Point appears once in an Even place and once in an Odd place.

References

Rademacher, H. and Toeplitz, O. *The Enjoyment of Mathematics: Selections from Mathematics for the Amateur.* Princeton, NJ: Princeton University Press, pp. 61–66, 1957.

Gauss Equations

If $\mathbf{x}$ is a regular patch on a Regular Surface in $\mathbb{R}^3$ with normal $\hat{\mathbf{N}}$, then

$$\begin{aligned} \mathbf{x}_{uu} &= \Gamma_{11}^1\mathbf{x}_u + \Gamma_{11}^2\mathbf{x}_v + e\hat{\mathbf{N}} && (1) \\ \mathbf{x}_{uv} &= \Gamma_{12}^1\mathbf{x}_u + \Gamma_{12}^2\mathbf{x}_v + f\hat{\mathbf{N}} && (2) \\ \mathbf{x}_{vv} &= \Gamma_{22}^1\mathbf{x}_u + \Gamma_{22}^2\mathbf{x}_v + g\hat{\mathbf{N}}, && (3) \end{aligned}$$

where e, f, and g are coefficients of the second Fundamental Form and Γ_{ij}^k are Christoffel Symbols of the Second Kind.

see also Christoffel Symbol of the Second Kind, Fundamental Forms, Mainardi-Codazzi Equations

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 398–400, 1993.

Gauss's Equation (Radius Derivatives)

Expresses the second derivatives of $\mathbf{r}$ in terms of the Christoffel Symbol of the Second Kind.

$$\mathbf{r}_{ij} = \Gamma_{ij}^k\mathbf{r}_k + (\mathbf{r}_{ij}\cdot\mathbf{n})\mathbf{n}.$$

Gauss's Formula

$$4\frac{x^p - y^p}{x-y} = R^2(x,y) - (-1)^{(p-1)/2} p S^2(x,y),$$

where R and S are HOMOGENEOUS POLYNOMIALS in x and y with integral COEFFICIENTS.

see also AURIFEUILLEAN FACTORIZATION, GAUSS'S BACKWARD FORMULA, GAUSS'S FORWARD FORMULA

References

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 105, 1993.

Gauss's Formulas

Let a SPHERICAL TRIANGLE have sides a, b, and c with A, B, and C the corresponding opposite angles. Then

$$\frac{\sin[\frac{1}{2}(a-b)]}{\sin(\frac{1}{2}c)} = \frac{\sin[\frac{1}{2}(A-B)]}{\cos(\frac{1}{2}C)} \tag{1}$$

$$\frac{\sin[\frac{1}{2}(a+b)]}{\sin(\frac{1}{2}c)} = \frac{\cos[\frac{1}{2}(A-B)]}{\sin(\frac{1}{2}C)} \tag{2}$$

$$\frac{\cos[\frac{1}{2}(a-b)]}{\cos(\frac{1}{2}c)} = \frac{\sin[\frac{1}{2}(A+B)]}{\cos(\frac{1}{2}C)} \tag{3}$$

$$\frac{\cos[\frac{1}{2}(a+b)]}{\cos(\frac{1}{2}c)} = \frac{\cos[\frac{1}{2}(A+B)]}{\sin(\frac{1}{2}C)}. \tag{4}$$

see also SPHERICAL TRIGONOMETRY

Gauss's Forward Formula

$$f_p = f_0 + p\delta_{1/2} + G_2\delta_0^2 + G_3\delta_{1/2}^3 + G_4\delta_0^4 + G_5\delta_{1/2}^5 + \ldots,$$

for $p \in [0,1]$, where δ is the CENTRAL DIFFERENCE and

$$G_{2n} = \binom{p+n-1}{2n}$$

$$G_{2n+1} = \binom{p+n}{2n+1},$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT.

see also CENTRAL DIFFERENCE, GAUSS'S BACKWARD FORMULA

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 433, 1987.

Gauss's Harmonic Function Theorem

If a function ϕ is HARMONIC in a SPHERE, then the value of ϕ at the center of the SPHERE is the ARITHMETIC MEAN of its value on the surface.

Gauss's Hypergeometric Theorem

$${}_2F_1(a,b;c;1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}$$

for $\Re[c-a-b] > 0$, where ${}_2F_1(a,b;c;x)$ is a HYPERGEOMETRIC FUNCTION. If a is a NEGATIVE INTEGER $-n$, this becomes

$${}_2F_1(-n,b;c;1) = \frac{(c-b)_n}{(c)_n},$$

which is known as the VANDERMONDE THEOREM.

see also GENERALIZED HYPERGEOMETRIC FUNCTION, HYPERGEOMETRIC FUNCTION

References

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, pp. 42 and 126, 1996.

Gauss's Inequality

If a distribution has a single MODE at μ_0, then

$$P(|x-\mu_0| \geq \lambda\tau) \leq \frac{4}{9\lambda^2},$$

where

$$\tau^2 \equiv \sigma^2 + (\mu - \mu_0)^2.$$

Gauss's Interpolation Formula

$$f(x) \approx t_n(x) = \sum_{k=0}^{2n} f_k \zeta_k(x),$$

where $t_n(x)$ is a trigonometric POLYNOMIAL of degree n such that $t_n(x_k) = f_k$ for $k = 0, \ldots, 2n$, and

$$\zeta_k(x) = \frac{\sin[\frac{1}{2}(x-x_0)]\cdots\sin[\frac{1}{2}(x-x_{k-1})]}{\sin[\frac{1}{2}(x_k-x_0)]\cdots\sin[\frac{1}{2}(x_k-x_{k-1})]} \frac{\sin[\frac{1}{2}(x-x_{k+1})]\cdots\sin[\frac{1}{2}(x-x_{2n})]}{\sin[\frac{1}{2}(x_k-x_{k+1})]\cdots\sin[\frac{1}{2}(x_k-x_{2n})]}.$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 881, 1972.

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 442–443, 1987.

Gauss-Jacobi Mechanical Quadrature

If $x_1 < x_2 < \ldots < x_n$ denote the zeros of $p_n(x)$, there exist REAL NUMBERS $\lambda_1, \lambda_2, \ldots, \lambda_n$ such that

$$\int_a^b \rho(x)\,d\alpha(x) = \lambda_1\rho(x_1) + \lambda_2\rho(x_2) + \ldots + \lambda_n\rho(x_n),$$

for an arbitrary POLYNOMIAL of order $2n-1$ and the $\lambda'_n s$ are called CHRISTOFFEL NUMBERS. The distribution $d\alpha(x)$ and the INTEGER n uniquely determine these numbers λ_ν.

References

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., p. 47, 1975.

Gauss-Jordan Elimination

A method for finding a MATRIX INVERSE. To apply Gauss-Jordan elimination, operate on a MATRIX

$$[\mathsf{A}\quad \mathsf{I}] \equiv \begin{bmatrix} a_{11} & \cdots & a_{1n} & 1 & 0 & \cdots & 0 \\ a_{21} & \cdots & a_{2n} & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} & 0 & 0 & \cdots & 1 \end{bmatrix},$$

where I is the IDENTITY MATRIX, to obtain a MATRIX of the form

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & b_{11} & \cdots & b_{1n} \\ 0 & 1 & \cdots & 0 & b_{21} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & b_{n1} & \cdots & b_{nn} \end{bmatrix}.$$

The MATRIX

$$\mathsf{B} \equiv \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ b_{21} & \cdots & b_{2n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nn} \end{bmatrix}$$

is then the MATRIX INVERSE of A. The procedure is numerically unstable unless PIVOTING (exchanging rows and columns as appropriate) is used. Picking the largest available element as the pivot is usually a good choice.

see also GAUSSIAN ELIMINATION, LU DECOMPOSITION, MATRIX EQUATION

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Gauss-Jordan Elimination" and "Gaussian Elimination with Backsubstitution." §2.1 and 2.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 27–32 and 33–34, 1992.

Gauss-Kummer Series

$$ {}_2F_1(-\tfrac{1}{2}, -\tfrac{1}{2}; 1; h^2) = \sum_{n=0}^{\infty} \binom{\frac{1}{2}}{n}^2 h^{2n} = 1 + \tfrac{1}{4}h^2 + \tfrac{1}{64}h^4 + \tfrac{1}{256}h^6 + \ldots, $$

where ${}_2F_1(a, b; c; x)$ is a HYPERGEOMETRIC FUNCTION. This can be derived using KUMMER'S QUADRATIC TRANSFORMATION.

Gauss-Kuzmin-Wirsing Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let x_0 be a random number from $[0, 1]$ written as a simple CONTINUED FRACTION

$$x_0 = 0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ldots}}}. \tag{1}$$

Define

$$x_n = 0 + \cfrac{1}{a_{n+1} + \cfrac{1}{a_{n+2} + \cfrac{1}{a_{n+3} + \ldots}}} = \frac{1}{x_{n-1}} - \left\lfloor \frac{1}{x_{n-1}} \right\rfloor. \tag{2}$$

Gauss (1800) showed that if $F(n, x)$ is the probability that $x_n < x$, then

$$\lim_{n \to \infty} F(n, x) = \frac{\ln(1 + x)}{\ln 2}. \tag{3}$$

Kuzmin (1928) published the first proof, which was subsequently improved by Lévy (1929). Wirsing (1974) showed, among other results, that

$$\lim_{n \to \infty} \frac{F(n, x) - \frac{\ln(1+x)}{\ln 2}}{(-\lambda)^n} = \Psi(x), \tag{4}$$

where $\lambda = 0.3036630029\ldots$ and $\Psi(x)$ is an analytic function with $\Psi(0) = \Psi(1) = 0$. This constant is connected to the efficiency of the EUCLIDEAN ALGORITHM (Knuth 1981).

References

Babenko, K. I. "On a Problem of Gauss." *Soviet Math. Dokl.* **19**, 136–140, 1978.

Daudé, H.; Flajolet, P.; and Vallée, B. "An Average-Case Analysis of the Gaussian Algorithm for Lattice Reduction." Submitted.

Durner, A. "On a Theorem of Gauss-Kuzmin-Lévy." *Arch. Math.* **58**, 251–256, 1992.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/kuzmin/kuzmin.html`.

Flajolet, P. and Vallée, B. "On the Gauss-Kuzmin-Wirsing Constant." Unpublished memo. 1995. `http://pauillac.inria.fr/algo/flajolet/Publications/gauss-kuzmin.ps`.

Knuth, D. E. *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, 1981.

MacLeod, A. J. "High-Accuracy Numerical Values of the Gauss-Kuzmin Continued Fraction Problem." *Computers Math. Appl.* **26**, 37–44, 1993.

Wirsing, E. "On the Theorem of Gauss-Kuzmin-Lévy and a Frobenius-Type Theorem for Function Spaces." *Acta Arith.* **24**, 507–528, 1974.

Gauss-Laguerre Quadrature

see LAGUERRE-GAUSS QUADRATURE

Gauss's Lemma

Let the multiples $m, 2m, \ldots, [(p-1)/2]m$ of an INTEGER such that $p \nmid m$ be taken. If there are an EVEN number r of least POSITIVE RESIDUES mod p of these numbers $> p/2$, then m is a QUADRATIC RESIDUE of p. If r is ODD, m is a QUADRATIC NONRESIDUE. Gauss's lemma can therefore be stated as $(m|p) = (-1)^r$, where $(m|p)$ is the LEGENDRE SYMBOL. It was proved by Gauss as a step along the way to the QUADRATIC RECIPROCITY THEOREM.

see also QUADRATIC RECIPROCITY THEOREM

Gauss's Machin-Like Formula

The MACHIN-LIKE FORMULA

$$\tfrac{1}{4}\pi = 12\cot^{-1} 18 + 8\cot^{-1} 57 - 5\cot^{-1} 239.$$

Gauss-Manin Connection

A connection defined on a smooth ALGEBRAIC VARIETY defined over the COMPLEX NUMBERS.

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 81, 1980.

Gauss Map

The Gauss map is a function from an ORIENTABLE SURFACE in EUCLIDEAN SPACE to a SPHERE. It associates to every point on the surface its oriented NORMAL VECTOR. For surfaces in 3-space, the Gauss map of the surface has DEGREE given by half the EULER CHARACTERISTIC of the surface

$$\iint_M K\,dA = 2\pi\chi(M) - \sum \alpha_i - \int_{\partial M} \kappa_g\,ds,$$

which works only for ORIENTABLE SURFACES.

see also CURVATURE, NIRENBERG'S CONJECTURE, PATCH

References

Gray, A. "The Local Gauss Map" and "The Gauss Map via Mathematica." §10.3 and §15.3 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 193–194 and 310–316, 1993.

Gauss's Mean-Value Theorem

Let $f(z)$ be an ANALYTIC FUNCTION in $|z - a| < R$. Then

$$f(z) = \frac{1}{2\pi}\int_0^{2\pi} f(z + re^{i\theta})\,d\theta$$

for $0 < r < R$.

Gauss Measure

The standard Gauss measure of a finite dimensional REAL HILBERT SPACE H with norm $||\cdot||_H$ has the BOREL MEASURE

$$\mu_H\,(dh) = (\sqrt{2\pi}\,)^{-\dim(H)} \exp(\tfrac{1}{2}||h||_H^2)\lambda_H\,(dh),$$

where λ_H is the LEBESGUE MEASURE on H.

Gauss Multiplication Formula

$$\begin{aligned}(2n\pi)^{(n-1)/2} & n^{1/2-nz}\Gamma(nz) \\ &= \Gamma(z)\Gamma(z+\tfrac{1}{n})\Gamma(z+\tfrac{2}{n})\cdots\Gamma(z+\tfrac{n-1}{n}) \\ &= \prod_{k=0}^{n-1}\Gamma\left(z+\frac{k}{n}\right),\end{aligned}$$

where $\Gamma(z)$ is the GAMMA FUNCTION.

see also GAMMA FUNCTION, LEGENDRE DUPLICATION FORMULA, POLYGAMMA FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 256, 1972.

Gauss Plane

see COMPLEX PLANE

Gauss's Polynomial Theorem

If a POLYNOMIAL

$$f(x) = x^N + C_1x^{N-1} + C_2x^{N-2} + \ldots + C_N$$

with integral COEFFICIENTS is divisible into a product of two POLYNOMIALS $f = \psi\phi$

$$\begin{aligned}\psi &= x^m + \alpha_1x^{m-1} + \ldots + \alpha_m \\ \phi &= x^n + \beta_1x^{n-1} + \ldots + \beta_n,\end{aligned}$$

then the COEFFICIENTS of this POLYNOMIAL are INTEGERS.

see also ABEL'S IRREDUCIBILITY THEOREM, ABEL'S LEMMA, KRONECKER'S POLYNOMIAL THEOREM, POLYNOMIAL, SCHOENEMANN'S THEOREM

References

Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, p. 119, 1965.

Gauss's Reciprocity Theorem

see QUADRATIC RECIPROCITY THEOREM

Gauss-Salamin Formula

see BRENT-SALAMIN FORMULA

Gauss's Test

If $u_n > 0$ and given $B(n)$ a bounded function of n as $n \to \infty$, express the ratio of successive terms as

$$\frac{u_n}{u_{n+1}} = 1 + \frac{h}{n} + \frac{B(n)}{n^2}.$$

The SERIES converges for $h > 1$ and diverges for $h \le 1$.

see also CONVERGENCE TESTS

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 287–288, 1985.

Gauss's Theorem

see DIVERGENCE THEOREM

Gauss's Theorema Egregium

Gauss's theorema egregium states that the GAUSSIAN CURVATURE of a surface embedded in 3-space may be understood intrinsically to that surface. "Residents" of the surface may observe the GAUSSIAN CURVATURE of the surface without ever venturing into full 3-dimensional space; they can observe the curvature of the surface they live in without even knowing about the 3-dimensional space in which they are embedded.

In particular, GAUSSIAN CURVATURE can be measured by checking how closely the ARC LENGTH of small RADIUS CIRCLES correspond to what they should be in EUCLIDEAN SPACE, $2\pi r$. If the ARC LENGTH of CIRCLES tends to be smaller than what is expected in EUCLIDEAN SPACE, then the space is positively curved; if larger, negatively; if the same, 0 GAUSSIAN CURVATURE.

Gauss (effectively) expressed the theorema egregium by saying that the GAUSSIAN CURVATURE at a point is given by $-R(v,w)v,w$, where R is the RIEMANN TENSOR, and v and w are an orthonormal basis for the TANGENT SPACE.

see also CHRISTOFFEL SYMBOL OF THE SECOND KIND, GAUSS EQUATIONS, GAUSSIAN CURVATURE

References

Gray, A. "Gauss's Theorema Egregium." §20.2 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 395–397, 1993.

Reckziegel, H. In *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 31–32, 1986.

Gauss's Transformation

If

$$(1+x\sin^2\alpha)\sin\beta=(1+x)\sin\alpha,$$

then

$$(1+x)\int_0^\alpha \frac{d\phi}{\sqrt{1-x^2\sin^2\phi}}=\int_0^\beta \frac{d\phi}{\sqrt{1-\frac{4x}{(1+x)^2}\sin^2\phi}}.$$

see also ELLIPTIC INTEGRAL OF THE FIRST KIND, LANDEN'S TRANSFORMATION

Gaussian Approximation Algorithm

see ARITHMETIC-GEOMETRIC MEAN

Gaussian Bivariate Distribution

The Gaussian bivariate distribution is given by

$$P(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{z}{2(1-\rho^2)}\right], \quad (1)$$

where

$$z\equiv\frac{(x_1-\mu_1)^2}{{\sigma_1}^2}-\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{{\sigma_2}^2}, \quad (2)$$

and

$$\rho\equiv\operatorname{cov}(x_1,x_2)=\frac{\langle x_1x_2\rangle-\langle x_1\rangle\langle x_2\rangle}{\sigma_1\sigma_2} \quad (3)$$

is the COVARIANCE. Let X_1 and X_2 be normally and independently distributed variates with MEAN 0 and VARIANCE 1. Then define

$$Y_1\equiv\mu_1+\sigma_{11}X_1+\sigma_{12}X_2 \quad (4)$$

$$Y_2\equiv\mu_2+\sigma_{21}X_1+\sigma_{22}X_2. \quad (5)$$

These new variates are normally distributed with MEAN μ_1 and μ_2, VARIANCE

$${\sigma_1}^2\equiv{\sigma_{11}}^2+{\sigma_{12}}^2 \quad (6)$$

$${\sigma_2}^2\equiv{\sigma_{21}}^2+{\sigma_{22}}^2, \quad (7)$$

and COVARIANCE

$$V_{12}\equiv\sigma_{11}\sigma_{21}+\sigma_{12}\sigma_{22}. \quad (8)$$

The COVARIANCE matrix is

$$V_{ij}=\begin{bmatrix}{\sigma_1}^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & {\sigma_2}^2\end{bmatrix}, \quad (9)$$

where

$$\rho\equiv\frac{V_{12}}{\sigma_1\sigma_2}=\frac{\sigma_{11}\sigma_{21}+\sigma_{12}\sigma_{22}}{\sigma_1\sigma_2}. \quad (10)$$

The joint probability density function for x_1 and x_2 is

$$f(x_1,x_2)\,dx_1\,dx_2=\frac{1}{2\pi}e^{-({x_1}^2+{x_2}^2)/2}\,dx_1\,dx_2. \quad (11)$$

However, from (4) and (5) we have

$$\begin{bmatrix}y_1-\mu_1\\ y_2-\mu_2\end{bmatrix}=\begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}. \quad (12)$$

Now, if

$$\begin{vmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{vmatrix}\neq 0, \quad (13)$$

then this can be inverted to give

$$\begin{aligned}\begin{bmatrix}x_1\\ x_2\end{bmatrix}&=\begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{bmatrix}^{-1}\begin{bmatrix}y_1-\mu_1\\ y_2-\mu_2\end{bmatrix}\\ &=\frac{1}{\sigma_{11}\sigma_{22}-\sigma_{12}\sigma_{21}}\begin{bmatrix}\sigma_{22} & -\sigma_{12}\\ -\sigma_{21} & \sigma_{11}\end{bmatrix}\begin{bmatrix}y_1-\mu_1\\ y_2-\mu_2\end{bmatrix}.\end{aligned} \quad (14)$$

Therefore,

$$\begin{aligned}{x_1}^2+{x_2}^2&=\frac{[\sigma_{22}(y_1-\mu_1)-\sigma_{12}(y_2-\mu_2)]^2}{(\sigma_{11}\sigma_{22}-\sigma_{12}\sigma_{21})^2}\\ &\quad+\frac{[-\sigma_{21}(y_1-\mu_1)+\sigma_{11}(y_2-\mu_2)]^2}{(\sigma_{11}\sigma_{22}-\sigma_{12}\sigma_{21})^2}.\end{aligned} \quad (15)$$

Expanding the Numerator gives

$$\begin{aligned}&{\sigma_{22}}^2(y_1-\mu_1)^2-2\sigma_{12}\sigma_{22}(y_1-\mu_1)(y_2-\mu_2)\\&\quad+{\sigma_{12}}^2(y_2-\mu_2)^2+{\sigma_{21}}^2(y_1-\mu_1)^2\\&\quad-2\sigma_{11}\sigma_{21}(y_1-\mu_1)(y_2-\mu_2)+{\sigma_{11}}^2(y_2-\mu_2)^2,\end{aligned} \tag{16}$$

so

$$\begin{aligned}&({x_1}^2+{x_2}^2)(\sigma_{11}\sigma_{22}-\sigma_{12}\sigma_{21})^2\\&=(y_1-\mu_1)^2({\sigma_{21}}^2+{\sigma_{22}}^2)\\&\quad-2(y_1-\mu_1)(y_2-\mu_2)(\sigma_{11}\sigma_{21}+\sigma_{12}\sigma_{22})\\&\quad+(y_2-\mu_2)^2({\sigma_{11}}^2+{\sigma_{12}}^2)\\&={\sigma_2}^2(y_1-\mu_1)^2-2(y_1-\mu_1)(y_2-\mu_2)(\rho\sigma_1\sigma_2)\\&\quad+{\sigma_1}^2(y_2-\mu_2)^2\\&={\sigma_1}^2{\sigma_2}^2\left[\frac{(y_1-\mu_1)^2}{{\sigma_1}^2}-\frac{2\rho(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(y_2-\mu_2)^2}{{\sigma_2}^2}\right].\end{aligned} \tag{17}$$

But

$$\begin{aligned}\frac{1}{1-\rho^2}&=\frac{1}{1-\frac{{V_{12}}^2}{{\sigma_1}^2{\sigma_2}^2}}=\frac{{\sigma_1}^2{\sigma_2}^2}{{\sigma_1}^2{\sigma_2}^2-{V_{12}}^2}\\&=\frac{{\sigma_1}^2{\sigma_2}^2}{({\sigma_{11}}^2+{\sigma_{12}}^2)({\sigma_{21}}^2+{\sigma_{22}}^2)-(\sigma_{11}\sigma_{21}+\sigma_{12}\sigma_{22})^2}.\end{aligned} \tag{18}$$

The Denominator is

$$\begin{aligned}&{\sigma_{11}}^2{\sigma_{21}}^2+{\sigma_{11}}^2{\sigma_{22}}^2+{\sigma_{12}}^2{\sigma_{21}}^2+{\sigma_{12}}^2{\sigma_{22}}^2-{\sigma_{11}}^2{\sigma_{21}}^2\\&-2\sigma_{11}\sigma_{12}\sigma_{21}\sigma_{22}-{\sigma_{12}}^2{\sigma_{22}}^2=(\sigma_{11}\sigma_{22}-\sigma_{12}\sigma_{21})^2,\end{aligned} \tag{19}$$

so

$$\frac{1}{1-\rho^2}=\frac{{\sigma_1}^2{\sigma_2}^2}{(\sigma_{11}\sigma_{22}-\sigma_{12}\sigma_{21})^2} \tag{20}$$

and

$$\begin{aligned}&{x_1}^2+{x_2}^2=\frac{1}{1-\rho^2}\\&\times\left[\frac{(y_1-\mu_1)^2}{{\sigma_1}^2}-\frac{2\rho(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(y_2-\mu_2)^2}{{\sigma_2}^2}\right].\end{aligned} \tag{21}$$

Solving for x_1 and x_2 and defining

$$\rho'\equiv\frac{\sigma_1\sigma_2\sqrt{1-\rho^2}}{\sigma_{11}\sigma_{22}-\sigma_{12}\sigma_{21}} \tag{22}$$

gives

$$x_1=\frac{\sigma_{22}(y_1-\mu_1)-\sigma_{12}(y_2-\mu_2)}{\rho'} \tag{23}$$

$$x_2=\frac{-\sigma_{21}(y_1-\mu_1)+\sigma_{11}(y_2-\mu_2)}{\rho'}. \tag{24}$$

The Jacobian is

$$\begin{aligned}J\left(\frac{x_1,x_2}{y_1,y_2}\right)&=\begin{vmatrix}\frac{\partial x_1}{\partial y_1}&\frac{\partial x_1}{\partial y_2}\\\frac{\partial x_2}{\partial y_1}&\frac{\partial x_2}{\partial y_2}\end{vmatrix}=\begin{vmatrix}\frac{\sigma_{22}}{\rho'}&-\frac{\sigma_{12}}{\rho'}\\-\frac{\sigma_{21}}{\rho'}&\frac{\sigma_{11}}{\rho'}\end{vmatrix}\\&=\frac{1}{\rho'^2}(\sigma_{11}\sigma_{22}-\sigma_{12}\sigma_{21})\\&=\frac{1}{\rho'}=\frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2}}.\end{aligned} \tag{25}$$

Therefore,

$$dx_1\,dx_2=\frac{dy_1\,dy_2}{\sigma_1\sigma_2\sqrt{1-\rho^2}} \tag{26}$$

and

$$\begin{aligned}&\frac{1}{2\pi}e^{-({x_1}^2+{x_2}^2)/2}\,dx_1\,dx_2\\&\quad=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-v/2}\,dy_1\,dy_2,\end{aligned} \tag{27}$$

where

$$\begin{aligned}&v\equiv\frac{1}{1-\rho^2}\\&\times\left[\frac{(y_1-\mu_1)^2}{{\sigma_1}^2}-\frac{2\rho(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(y_2-\mu_2)^2}{{\sigma_2}^2}\right].\end{aligned} \tag{28}$$

Now, if

$$\begin{vmatrix}\sigma_{11}&\sigma_{12}\\\sigma_{21}&\sigma_{22}\end{vmatrix}=0, \tag{29}$$

then

$$\sigma_{11}\sigma_{12}=\sigma_{12}\sigma_{21} \tag{30}$$

$$y_1=\mu_1+\sigma_{11}x_1+\sigma_{12}x_2 \tag{31}$$

$$\begin{aligned}y_2&=\mu_2+\frac{\sigma_{12}\sigma_{21}}{\sigma_{11}}x_2=\mu_2+\frac{\sigma_{11}\sigma_{21}x_1+\sigma_{12}\sigma_{21}x_2}{\sigma_{11}}\\&=\mu_2+\frac{\sigma_{21}}{\sigma_{11}}(\sigma_{11}x_1+\sigma_{12}x_2),\end{aligned} \tag{32}$$

so

$$y_1=\mu_1+x_3 \tag{33}$$

$$y_2=\mu_2+\frac{\sigma_{21}}{\sigma_{11}}x_3, \tag{34}$$

where

$$x_3=y_1-\mu_1=\frac{\sigma_{11}}{\sigma_{21}}(y_2-\mu_2). \tag{35}$$

The Characteristic Function is given by

$$\begin{aligned}\phi(t_1,t_2)&\equiv\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{i(t_1x_1+t_2x_2)}P(x_1,x_2)\,dx_1\,dx_2\\&=N\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{i(t_1x_1+t_2x_2)}\exp\left[-\frac{z}{2(1-\rho^2)}\right]dx_1\,dx_2,\end{aligned} \tag{36}$$

where

$$z \equiv \left[\frac{(x_1-\mu_1)^2}{{\sigma_1}^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{{\sigma_2}^2}\right] \tag{37}$$

and

$$N \equiv \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}. \tag{38}$$

Now let

$$u \equiv x_1 - \mu_1 \tag{39}$$

$$w \equiv x_2 - \mu_2. \tag{40}$$

Then

$$\phi(t_1,t_2) = N' \int_{-\infty}^{\infty} \left(e^{it_2 w} \exp\left[-\frac{1}{2(1-\rho^2)}\frac{w^2}{{\sigma_2}^2}\right]\right) \times \int_{-\infty}^{\infty} e^v e^{t_1 u}\,du\,dw, \tag{41}$$

where

$$v \equiv -\frac{1}{2(1-\rho^2)}\frac{1}{{\sigma_1}^2}\left[u^2 - \frac{2\rho\sigma_1 w}{\sigma_2}u\right]$$

$$N' \equiv \frac{e^{i(t_1\mu_1+t_2\mu_2)}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}. \tag{42}$$

COMPLETE THE SQUARE in the inner integral

$$\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\frac{1}{{\sigma_1}^2}\left[u^2 - \frac{2\rho\sigma_1 w}{\sigma_2}u\right]\right\} e^{t_1 u}\,du$$
$$= \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2{\sigma_1}^2(1-\rho^2)}\left[u - \frac{\rho_1\sigma_1 w}{\sigma^2}\right]^2\right\} \times \left\{\frac{1}{2{\sigma_1}^2(1-\rho^2)}\left(\frac{\rho_1\sigma_1 w}{\sigma_2}\right)^2\right\} e^{it_1 u}\,du. \tag{43}$$

Rearranging to bring the exponential depending on w outside the inner integral, letting

$$v \equiv u - \rho\frac{\sigma_1 w}{\sigma_2}, \tag{44}$$

and writing

$$e^{it_1 u} = \cos(t_1 u) + i\sin(t_1 u) \tag{45}$$

gives

$$\phi(t_1,t_2) = N' \int_{-\infty}^{\infty} e^{it_2 w} \exp\left[-\frac{1}{2{\sigma_2}^2(1-\rho^2)}w^2\right] \times \exp\left[\frac{\rho^2}{2{\sigma_2}^2(1-\rho^2)}w^2\right] \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2{\sigma_2}^2(1-\rho^2)}v^2\right] \times \left\{\cos\left[t_1\left(v + \frac{\rho\sigma_1 w}{\sigma_2}\right)\right] + i\sin\left[t_1\left(v + \frac{\rho\sigma_1 w}{\sigma_2}\right)\right]\right\} dv\,dw. \tag{46}$$

Expanding the term in braces gives

$$\left[\cos(t_1 v)\cos\left(\frac{\rho\sigma_1 w t_1}{\sigma_2}\right) - \sin(t_1 v)\sin\left(\frac{\rho\sigma_1 w}{\sigma_2 t_1}\right)\right] + i\left[\sin(t_1 v)\cos\left(\frac{\rho\sigma_1 w}{\sigma_2 t_1}\right) + \cos(t_1 v)\sin\left(\frac{\rho\sigma_1 w t_1}{\sigma_2}\right)\right]$$
$$= \left[\cos\left(\frac{\rho\sigma_1 w t_1}{\sigma_2}\right) + i\sin\left(\frac{\rho\sigma_1 w t_1}{\sigma_2}\right)\right] [\cos(t_1 v) + i\sin(t_1 v)]$$
$$= \exp\left(\frac{i\rho\sigma_1 w}{\sigma_2}t_1\right)[\cos(t_1 v) + i\sin(t_1 v)]. \tag{47}$$

But $e^{-ax^2}\sin(bx)$ is ODD, so the integral over the sine term vanishes, and we are left with

$$\phi(t_1,t_2) = N' \int_{-\infty}^{\infty} e^{it_2 w} \exp\left[-\frac{w^2}{2{\sigma_2}^2}\right] \times \exp\left[\frac{\rho^2 w^2}{2{\sigma_2}^2(1-\rho^2)}\right] \exp\left[\frac{i\rho\sigma_1 w t_1}{\sigma_2}\right] dw \times \int_{-\infty}^{\infty} \exp\left[-\frac{v^2}{2{\sigma_1}^2(1-\rho^2)}\right]\cos(t_1 v)\,dv$$
$$= N' \int_{-\infty}^{\infty} \exp\left[iw\left(t_2 + t_1\left(\rho\frac{\sigma_1}{\sigma_2}\right)\right)\right] \exp\left[-\frac{w^2}{2{\sigma_2}^2}\right] dw \times \int_{-\infty}^{\infty} \exp\left[-\frac{v^2}{2{\sigma_1}^2(1-\rho^2)}\right]\cos(t_1 v)\,dv. \tag{48}$$

Now evaluate the GAUSSIAN INTEGRAL

$$\int_{-\infty}^{\infty} e^{ikx}e^{-ax^2}\,dx = \int_{-\infty}^{\infty} e^{-ax^2}\cos(kx)\,dx = \sqrt{\frac{\pi}{a}}e^{-k^2/4a} \tag{49}$$

to obtain the explicit form of the CHARACTERISTIC FUNCTION,

$$\phi(t_1,t_2) = \frac{e^{i(t_1\mu_1+t_2\mu_2)}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \left\{\sigma_2\sqrt{2\pi}\exp\left[-\frac{1}{4}\left(t_2 + \rho\frac{\sigma_1}{\sigma_2}t_1\right)^2 2{\sigma_2}^2\right]\right\} \times \left\{\sigma_1\sqrt{2\pi(1-\rho^2)}\exp\left[-\tfrac{1}{4}{t_1}^2 2{\sigma_1}^2(1-\rho^2)\right]\right\}$$
$$= e^{i(t_1\mu_1+t_2\mu_2)}\exp\{-\tfrac{1}{2}[{t_2}^2{\sigma_2}^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + \rho^2{\sigma_1}^2{t_1}^2 + (1-\rho^2){\sigma_1}^2{t_1}^2]\}$$
$$= \exp[i(t_1\mu_1 + t_2\mu_2) - \tfrac{1}{2}({\sigma_1}^2{t_1}^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + {\sigma_1}^2{t_1}^2)]. \tag{50}$$

Let z_1 and z_2 be two independent Gaussian variables with MEANS $\mu_i = 0$ and ${\sigma_i}^2 = 1$ for $i = 1, 2$. Then the variables a_1 and a_2 defined below are Gaussian bivariates with unit VARIANCE and CROSS-CORRELATION COEFFICIENT ρ:

$$a_1 = \sqrt{\frac{1+\rho}{2}}\,z_1 + \sqrt{\frac{1-\rho}{2}}\,z_2 \tag{51}$$

$$a_2 = \sqrt{\frac{1+\rho}{2}}\, z_1 - \sqrt{\frac{1-\rho}{2}}\, z_2. \tag{52}$$

The conditional distribution is

$$P(x_2|x_1) = \frac{1}{\sigma_2\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{(x^2-\mu'^2)^2}{2\sigma_2'^{\,2}}\right], \tag{53}$$

where

$$\mu'_2 \equiv \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1) \tag{54}$$

$$\sigma'_2 \equiv \sigma_2\sqrt{1-\rho_2}\,. \tag{55}$$

The marginal probability density is

$$\begin{aligned} P(x_2) &= \int_{-\infty}^{\infty} P(x_1, x_2)\, dx_1 \\ &= \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left[-\frac{(x_2-\mu_2)^2}{2{\sigma_2}^2}\right]. \end{aligned} \tag{56}$$

see also BOX-MULLER TRANSFORMATION, GAUSSIAN DISTRIBUTION, MCMOHAN'S THEOREM, NORMAL DISTRIBUTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 936–937, 1972.

Spiegel, M. R. *Theory and Problems of Probability and Statistics.* New York: McGraw-Hill, p. 118, 1992.

Gaussian Brackets

Published by Gauss in *Disquisitiones Arithmeticae.* They are defined as follows.

$$[\,] = 1 \tag{1}$$

$$[a_1] = a_1 \tag{2}$$

$$[a_1, a_2] = [a_1]a_2 + [\,] \tag{3}$$

$$\begin{aligned} [a_1, a_2, \ldots, a_n] = [a_1, a_2, \ldots, a_{n-1}]a_n \\ +[a_1, a_2, \ldots, a_{n-2}]. \end{aligned} \tag{4}$$

Gaussian brackets are useful for treating CONTINUED FRACTIONS because

$$\cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ldots + \cfrac{1}{a_n}}}} = \frac{[a_2, a_n]}{[a_1, a_n]}. \tag{5}$$

The NOTATION $[x]$ conflicts with that of GAUSSIAN POLYNOMIALS and the NINT function.

References

Herzberger, M. *Modern Geometrical Optics.* New York: Interscience Publishers, pp. 457–462, 1958.

Gaussian Coefficient

see q-BINOMIAL COEFFICIENT

Gaussian Coordinate System

A coordinate system which has a METRIC satisfying $g_{ii} = -1$ and $\partial g_{ij}/\partial x_j = 0$.

Gaussian Curvature

An intrinsic property of a space independent of the coordinate system used to describe it. The Gaussian curvature of a REGULAR SURFACE in $\mathbb{R}^3$ at a point $\mathbf{p}$ is formally defined as

$$K(\mathbf{p}) = \tfrac{1}{2}\det(S(\mathbf{p})), \tag{1}$$

where S is the SHAPE OPERATOR and det denotes the DETERMINANT.

If $\mathbf{x} : U \to \mathbb{R}^3$ is a REGULAR PATCH, then the Gaussian curvature is given by

$$K = \frac{eg - f^2}{EG - F^2}, \tag{2}$$

where E, F, and G are coefficients of the first FUNDAMENTAL FORM and e, f, and g are coefficients of the second FUNDAMENTAL FORM (Gray 1993, p. 282). The Gaussian curvature can be given entirely in terms of the first FUNDAMENTAL FORM

$$ds^2 = E\, du^2 + 2F\, du\, dv + G\, dv^2 \tag{3}$$

and the DISCRIMINANT

$$g \equiv EG - F^2 \tag{4}$$

by

$$K = \frac{1}{\sqrt{g}}\left[\frac{\partial}{\partial v}\left(\frac{\sqrt{g}}{E}\Gamma^2_{11}\right) - \frac{\partial}{\partial u}\left(\frac{\sqrt{g}}{E}\Gamma^2_{12}\right)\right], \tag{5}$$

where Γ^k_{ij} are the CONNECTION COEFFICIENTS. Equivalently,

$$\begin{aligned} K = \frac{1}{g^2}\begin{vmatrix} E & F & \frac{\partial F}{\partial v} - \frac{1}{2}\frac{\partial G}{\partial u} \\ F & G & \frac{1}{2}\frac{\partial G}{\partial v} \\ \frac{1}{2}\frac{\partial E}{\partial u} & k_{23} & k_{33} \end{vmatrix} \\ - \frac{1}{g^2}\begin{vmatrix} E & F & \frac{1}{2}\frac{\partial E}{\partial v} \\ F & G & \frac{1}{2}\frac{\partial G}{\partial u} \\ \frac{1}{2}\frac{\partial E}{\partial v} & \frac{1}{2}\frac{\partial G}{\partial v} & 0 \end{vmatrix}, \end{aligned} \tag{6}$$

where

$$k_{23} \equiv \frac{\partial F}{\partial u} - \frac{1}{2}\frac{\partial E}{\partial v} \tag{7}$$

$$k_{33} \equiv -\frac{1}{2}\frac{\partial^2 E}{\partial v^2} + \frac{\partial^2 F}{\partial u \partial v} - \frac{1}{2}\frac{\partial^2 G}{\partial u^2}. \tag{8}$$

Writing this out,

$$\begin{aligned}K = {} & \frac{1}{2g}\left[2\frac{\partial^2 F}{\partial u \partial v} - \frac{\partial^2 E}{\partial v^2} - \frac{\partial^2 G}{\partial u^2}\right] \\ & - \frac{G}{4g^2}\left[\frac{\partial E}{\partial u}\left(2\frac{\partial F}{\partial v} - \frac{\partial G}{\partial u}\right) - \left(\frac{\partial E}{\partial v}\right)^2\right] \\ & + \frac{F}{4g_2}\left[\frac{\partial E}{\partial u}\frac{\partial G}{\partial v} - 2\frac{\partial E}{\partial v}\frac{\partial G}{\partial u}\right. \\ & \left. + \left(2\frac{\partial F}{\partial u} - \frac{\partial E}{\partial v}\right)\left(2\frac{\partial F}{\partial v} - \frac{\partial G}{\partial u}\right)\right] \\ & - \frac{E}{4g^2}\left[\frac{\partial G}{\partial v}\left(2\frac{\partial F}{\partial u} - \frac{\partial E}{\partial v}\right) - \left(\frac{\partial G}{\partial u}\right)^2\right]. \end{aligned} \tag{9}$$

The Gaussian curvature is also given by

$$K = \frac{\det(\mathbf{x}_{uu}\mathbf{x}_u\mathbf{x}_v)\det(\mathbf{x}_{vv}\mathbf{x}_u\mathbf{x}_v) - [\det(\mathbf{x}_{uv}\mathbf{x}_u\mathbf{x}_v)]^2}{[|\mathbf{x}_u|^2|\mathbf{x}_v|^2 - (\mathbf{x}_u \cdot \mathbf{x}_v)^2]^2} \tag{10}$$

(Gray 1993, p. 285), as well as

$$K = \frac{[\hat{\mathbf{N}}\,\hat{\mathbf{N}}_1\,\hat{\mathbf{N}}_2]}{\sqrt{g}} = \frac{\epsilon^{ij}[\hat{\mathbf{N}}\,\hat{\mathbf{T}}\,\hat{\mathbf{T}}_i]_j}{\sqrt{g}}, \tag{11}$$

where ϵ^{ij} is the LEVI-CIVITA SYMBOL, $\hat{\mathbf{N}}$ is the unit NORMAL VECTOR and $\hat{\mathbf{T}}$ is the unit TANGENT VECTOR. The Gaussian curvature is also given by

$$K = -\frac{R}{2} = \kappa_1\kappa_2 = \frac{1}{R_1R_2}, \tag{12}$$

where R is the CURVATURE SCALAR, κ_1 and κ_2 the PRINCIPAL CURVATURES, and R_1 and R_2 the PRINCIPAL RADII OF CURVATURE. For a MONGE PATCH with $z = h(u,v)$,

$$K = \frac{h_{uu}h_{vv} - {h_{uv}}^2}{(1 + {h_u}^2 + {h_v}^2)^2}. \tag{13}$$

The Gaussian curvature K and MEAN CURVATURE H satisfy

$$H^2 \geq K, \tag{14}$$

with equality only at UMBILIC POINTS, since

$$H^2 - K^2 = \tfrac{1}{4}(\kappa_1 - \kappa_2)^2. \tag{15}$$

If $\mathbf{p}$ is a point on a REGULAR SURFACE $M \subset \mathbb{R}^3$ and $\mathbf{v_p}$ and $\mathbf{w_p}$ are tangent vectors to M at $\mathbf{p}$, then the Gaussian curvature of M at $\mathbf{p}$ is related to the SHAPE OPERATOR S by

$$S(\mathbf{v_p}) \times S(\mathbf{w_p}) = K(\mathbf{p})\mathbf{v_p} \times \mathbf{w_p}. \tag{16}$$

Let $\mathbf{Z}$ be a nonvanishing VECTOR FIELD on M which is everywhere PERPENDICULAR to M, and let V and W be VECTOR FIELDS tangent to M such that $V \times W = \mathbf{Z}$, then

$$K = \frac{\mathbf{Z} \cdot (D_V\mathbf{Z} \times D_W\mathbf{Z})}{2|\mathbf{Z}|^4} \tag{17}$$

(Gray 1993, pp. 291–292).

For a SPHERE, the Gaussian quadrature is $K = 1/a^2$. For EUCLIDEAN SPACE, the Gaussian quadrature is $K = 0$. For GAUSS-BOLYAI-LOBACHEVSKY SPACE, the Gaussian quadrature is $K = -1/a^2$. A FLAT SURFACE is a REGULAR SURFACE and special class of MINIMAL SURFACE on which Gaussian curvature vanishes everywhere.

A point $\mathbf{p}$ on a REGULAR SURFACE $M \in \mathbb{R}^3$ is classified based on the sign of $K(\mathbf{p})$ as given in the following table (Gray 1993, p. 280), where S is the SHAPE OPERATOR.

Sign	Point
$K(\mathbf{p}) > 0$	elliptic point
$K(\mathbf{p}) < 0$	hyperbolic point
$K(\mathbf{p}) = 0$ but $S(\mathbf{p}) \neq 0$	parabolic point
$K(\mathbf{p}) = 0$ and $S(\mathbf{p}) = 0$	planar point

A surface on which the Gaussian curvature K is everywhere POSITIVE is called SYNCLASTIC, while a surface on which K is everywhere NEGATIVE is called ANTICLASTIC. Surfaces with constant Gaussian curvature include the CONE, CYLINDER, KUEN SURFACE, PLANE, PSEUDOSPHERE, and SPHERE. Of these, the CONE and CYLINDER are the only FLAT SURFACES OF REVOLUTION.

see also ANTICLASTIC, BRIOSCHI FORMULA, DEVELOPABLE SURFACE, ELLIPTIC POINT, FLAT SURFACE, HYPERBOLIC POINT, INTEGRAL CURVATURE, MEAN CURVATURE, METRIC TENSOR, MINIMAL SURFACE, PARABOLIC POINT, PLANAR POINT, SYNCLASTIC, UMBILIC POINT

References

Geometry Center. "Gaussian Curvature." `http://www.geom.umn.edu/zoo/diffgeom/surfspace/concepts/curvatures/gauss-curv.html`.

Gray, A. "The Gaussian and Mean Curvatures" and "Surfaces of Constant Gaussian Curvature." §14.5 and Ch. 19 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 279–285 and 375–387, 1993.

Gaussian Differential Equation

see HYPERGEOMETRIC DIFFERENTIAL EQUATION

Gaussian Distribution

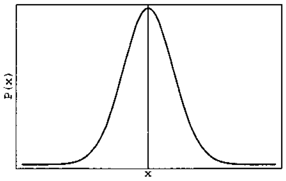

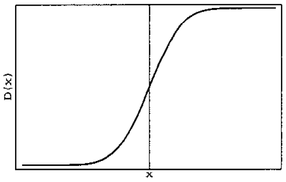

The Gaussian probability distribution with MEAN μ and STANDARD DEVIATION σ is a GAUSSIAN FUNCTION of the form

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \tag{1}$$

where $P(x)\,dx$ gives the probability that a variate with a Gaussian distribution takes on a value in the range $[x, x+dx]$. This distribution is also called the NORMAL DISTRIBUTION or, because of its curved flaring shape, the BELL CURVE. The distribution $P(x)$ is properly normalized for $x \in (-\infty, \infty)$ since

$$\int_{-\infty}^{\infty} P(x)\,dx = 1. \tag{2}$$

The cumulative DISTRIBUTION FUNCTION, which gives the probability that a variate will assume a value $\leq x$, is then

$$D(x) \equiv \int_{-\infty}^{x} P(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(x-\mu)^2/2\sigma^2}\,dx. \tag{3}$$

Gaussian distributions have many convenient properties, so random variates with unknown distributions are often assumed to be Gaussian, especially in physics and astronomy. Although this can be a dangerous assumption, it is often a good approximation due to a surprising result known as the CENTRAL LIMIT THEOREM. This theorem proves that the MEAN of any set of variates with any distribution having a finite MEAN and VARIANCE tends to the Gaussian distribution. Many common attributes such as test scores, height, etc., follow roughly Gaussian distributions, with few members at the high and low ends and many in the middle.

Making the transformation

$$z \equiv \frac{x-\mu}{\sigma} \tag{4}$$

so that $dz = dz/\sigma$ gives a variate with unit VARIANCE and 0 MEAN

$$P(x)\,dx = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz, \tag{5}$$

known as a standard NORMAL DISTRIBUTION. So defined, z is known as a z-SCORE).

The NORMAL DISTRIBUTION FUNCTION gives the probability that a standard normal variate assumes a value in the interval $[0, z]$,

$$\Phi(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-z^2/2}\,dz = \tfrac{1}{2}\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right). \tag{6}$$

Here, ERF is a function sometimes called the error function. Neither $\Phi(z)$ nor ERF can be expressed in terms of finite additions, subtractions, multiplications, and root extractions, and so both must be either computed numerically or otherwise approximated. The value of a for which $P(x)$ falls within the interval $[-a, a]$ with a given probability P is called the P CONFIDENCE INTERVAL.

The Gaussian distribution is also a special case of the CHI-SQUARED DISTRIBUTION, since substituting

$$z \equiv \frac{(x-\mu)^2}{\sigma^2} \tag{7}$$

so that

$$dz = \tfrac{1}{2}\frac{2(x-\mu)}{\sigma}\,dx = \sqrt{\frac{z}{\sigma}}\,dx \tag{8}$$

(where an extra factor of 1/2 has been added to dz since z runs from 0 to ∞ instead of from $-\infty$ to ∞), gives

$$\begin{aligned} P(x)\,dx &= \frac{1}{\sqrt{2\pi}} e^{-(z/\sigma)/2} \left(\frac{z}{\sigma}\right)^{-1/2} d\left(\frac{z}{\sigma}\right)\,dz \\ &= \frac{1}{2^{1/2}\Gamma(\frac{1}{2})} e^{-(z/\sigma)/2} \left(\frac{z}{\sigma}\right)^{-1/2} d\left(\frac{z}{\sigma}\right)\,dz, \end{aligned} \tag{9}$$

which is a CHI-SQUARED DISTRIBUTION in z/σ with $r = 1$ (i.e., a GAMMA DISTRIBUTION with $\alpha = 1/2$ and $\theta = 2$).

Cramer showed in 1936 that if X and Y are INDEPENDENT variates and $X + Y$ has a Gaussian distribution, then both X and Y must be Gaussian (CRAMER'S THEOREM).

The ratio X/Y of independent Gaussian-distributed variates with zero MEAN is distributed with a CAUCHY DISTRIBUTION. This can be seen as follows. Let X and Y both have MEAN 0 and standard deviations of σ_x and σ_y, respectively, then the joint probability density function is the GAUSSIAN BIVARIATE DISTRIBUTION with $\rho = 0$,

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-[x^2/(2{\sigma_x}^2)+y^2/(2{\sigma_y}^2)]}. \tag{10}$$

From RATIO DISTRIBUTION, the distribution of $U = Y/X$ is

$$\begin{aligned} P(u) &= \int_{-\infty}^{\infty} |x| f(x, ux)\,dx \\ &= \frac{1}{2\pi\sigma_x\sigma_y} \int_{-\infty}^{\infty} |x| e^{-[x^2/(2{\sigma_x}^2)+u^2x^2/(2{\sigma_y}^2))]} \\ &= \frac{1}{\pi\sigma_x\sigma_y} \int_0^{\infty} x \exp\left[-x^2\left(\frac{1}{2{\sigma_x}^2} + \frac{u^2}{2{\sigma_y}^2}\right)\right]\,dx. \end{aligned} \tag{11}$$

But

$$\int_0^{\infty} x e^{-ax^2}\,dx = \left[-\frac{1}{2a} e^{-ax^2}\right]_0^{\infty} = \frac{1}{2a}[0-(-1)] = \frac{1}{2a}, \tag{12}$$

so

$$P(u) = \frac{1}{\pi\sigma_x\sigma_y}\frac{1}{2\left(\frac{1}{2\sigma_x^2}+\frac{u^2}{2\sigma_y^2}\right)} = \frac{1}{\pi}\frac{\sigma_x\sigma_y}{u^2\sigma_x^2+\sigma_y^2}$$

$$= \frac{1}{\pi}\frac{\frac{\sigma_y}{\sigma_x}}{u^2+\left(\frac{\sigma_y}{\sigma_x}\right)^2}, \qquad (13)$$

which is a CAUCHY DISTRIBUTION with MEAN $\mu = 0$ and full width

$$\Gamma = \frac{2\sigma_y}{\sigma_x}. \qquad (14)$$

The CHARACTERISTIC FUNCTION for the Gaussian distribution is

$$\phi(t) = e^{imt-\sigma^2t^2/2}, \qquad (15)$$

and the MOMENT-GENERATING FUNCTION is

$$M(t) = \langle e^{tx}\rangle = \int_{-\infty}^{\infty}\frac{e^{tx}}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}\,dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2\sigma^2}\left[x^2-2(\mu+\sigma^2t)x+\mu^2\right]\right\}dx. \qquad (16)$$

COMPLETING THE SQUARE in the exponent,

$$\frac{1}{2\sigma^2}[x^2-2(\mu+\sigma^2t)x+\mu^2]$$

$$= \frac{1}{2\sigma^2}\left\{[x-(\mu+\sigma^2t)]^2+[\mu^2-(\mu+\sigma^2t)^2]\right\}. \qquad (17)$$

Let

$$y \equiv x-(\mu+\sigma^2t) \qquad (18)$$

$$dy = dx \qquad (19)$$

$$a \equiv \frac{1}{2\sigma^2}. \qquad (20)$$

The integral then becomes

$$M(t) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left[-ay^2+\frac{2\mu\sigma^2t+\sigma^4t^2}{2\sigma^2}\right]dy$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp[-ay^2+\mu t+\tfrac{1}{2}\sigma^2t^2]\,dy$$

$$= \frac{1}{\sigma\sqrt{2\pi}}e^{\mu t+\sigma^2t^2/2}\int_{-\infty}^{\infty}e^{-ay^2}\,dy$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\sqrt{\frac{\pi}{a}}\,e^{\mu t+\sigma^2t^2/2}$$

$$= \frac{\sqrt{2\sigma^2\pi}}{\sigma\sqrt{2\pi}}\,e^{\mu t+\sigma^2t^2/2} = e^{\mu t+\sigma^2t^2/2}, \qquad (21)$$

so

$$M'(t) = (\mu+\sigma^2t)e^{\mu t+\sigma^2t^2/2} \qquad (22)$$

$$M''(t) = \sigma^2e^{\mu t+\sigma^2t^2/2}+e^{\mu t+\sigma^2t^2/2}(\mu+t\sigma^2)^2, \qquad (23)$$

and

$$\mu = M'(0) = \mu \qquad (24)$$

$$\sigma^2 = M''(0)-[M'(0)]^2$$

$$= (\sigma^2+\mu^2)-\mu^2 = \sigma^2. \qquad (25)$$

These can also be computed using

$$R(t) = \ln[M(t)] = \mu t+\tfrac{1}{2}\sigma^2t^2 \qquad (26)$$

$$R'(t) = \mu+\sigma^2t \qquad (27)$$

$$R''(t) = \sigma^2, \qquad (28)$$

yielding, as before,

$$\mu = R'(0) = \mu \qquad (29)$$

$$\sigma^2 = R''(0) = \sigma^2. \qquad (30)$$

The moments can also be computed directly by computing the MOMENTS about the origin $\mu'_n \equiv \langle x^n\rangle$,

$$\mu'_n = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}x^ne^{-(x-\mu)^2/2\sigma^2}\,dx. \qquad (31)$$

Now let

$$u \equiv \frac{x-\mu}{\sqrt{2}\sigma} \qquad (32)$$

$$du = \frac{dx}{\sqrt{2}\sigma} \qquad (33)$$

$$x = \sigma u\sqrt{2}+\mu, \qquad (34)$$

giving

$$\mu'_n = \frac{\sqrt{2}\,\sigma}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}x^ne^{-u^2}\,du = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}x^ne^{-u^2}\,du, \qquad (35)$$

so

$$\mu'_0 = 1 \qquad (36)$$

$$\mu'_1 = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}xe^{-u^2}\,du$$

$$= \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}(\sqrt{2}\,\sigma u+\mu)e^{-u^2}\,du$$

$$= [\sqrt{2}\sigma H_1(1)+\mu H_0(1)] = (0+\mu) = \mu \qquad (37)$$

$$\mu'_2 = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}x^2e^{-u^2}\,du$$

$$= \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}(2\sigma^2u^2+2\sqrt{2}\,\sigma\mu u+\mu^2)e^{-u^2}\,du$$

$$= [2\sigma^2H_2(1)+2\sqrt{2}\,\sigma\mu H_1(1)+\mu^2H_0(1)]$$

$$= (2\sigma^2\tfrac{1}{2}+0+\mu^2) = \mu^2+\sigma^2 \qquad (38)$$

$$\mu'_3 = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}x^3e^{-u^2}\,du$$

$$\begin{aligned}
&= \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}(2\sqrt{2}\,\sigma^3u^3+6\mu\sigma^2u^2\\
&\quad +3\sqrt{2}\,\mu^2\sigma u+\mu^3)e^{-u^2}\,du\\
&=[2\sqrt{2}\,\sigma^3H_3(1)+6\mu\sigma^2H_2(1)\\
&\quad +3\sqrt{2}\,\mu^2\sigma H_1(1)+\mu^3H_0(1)]\\
&=(0+6\mu^2\sigma^2\tfrac{1}{2}+0+\mu^3)=\mu(\mu^2+3\sigma^2) \qquad (39)
\end{aligned}$$

$$\begin{aligned}
\mu_4' &= \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}x^3e^{-u^2}\,du\\
&= \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}(4\sigma^4u^4+8\sqrt{2}\,\mu\sigma^3u^3\\
&\quad +12\mu^2\sigma^2u^2+4\sqrt{2}\,\mu^3\sigma u+\mu^4)e^{-u^2}\,du\\
&=[4\sigma^4H_4(1)+8\sqrt{2}\,\mu\sigma^3H_3(1)+12\mu^2\sigma^2H_2(1)\\
&\quad +4\sqrt{2}\,\mu^3\sigma H_1(1)+\mu^4H_0(1)]\\
&=(4\sigma^4\tfrac{3}{4}+0+12\mu^2\sigma^2\tfrac{1}{2}+0+\mu^4)\\
&=\mu^4+6\mu^2\sigma^2+3\sigma^4, \qquad (40)
\end{aligned}$$

where $H_n(a)$ are GAUSSIAN INTEGRALS.

Now find the MOMENTS about the MEAN,

$$\mu_1 \equiv 0 \qquad (41)$$

$$\mu_2 \equiv \mu_2' - (\mu_1')^2 = (\mu^2+\sigma^2)-\mu^2=\sigma^2 \qquad (42)$$

$$\begin{aligned}
\mu_3 &\equiv \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3\\
&= \mu(\mu^2+3\sigma^2)-3(\sigma^2+\mu^2)\mu+2\mu^3=0 \qquad (43)
\end{aligned}$$

$$\begin{aligned}
\mu_4 &\equiv \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4\\
&= (\mu^4+6\mu^2\sigma^2+3\sigma^4)-4(\mu^3+3\mu\sigma^2)\mu\\
&\quad +6(\mu^2+\sigma^2)\mu^2-3\mu^4\\
&= 3\sigma^4, \qquad (44)
\end{aligned}$$

so the VARIANCE, STANDARD DEVIATION, SKEWNESS, and KURTOSIS are given by

$$\operatorname{var}(x) \equiv \mu_2 = \sigma^2 \qquad (45)$$

$$\operatorname{stdv}(x) \equiv \sqrt{\operatorname{var}(x)} = \sigma \qquad (46)$$

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = 0 \qquad (47)$$

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{3\sigma^4}{\sigma^4} - 3 = 0. \qquad (48)$$

The VARIANCE of the SAMPLE VARIANCE s^2 for a sample taken from a population with a Gaussian distribution is

$$\begin{aligned}
\operatorname{var}(s^2) &= \frac{(N-1)[(N-1)\mu_4' - (N-3){\mu_2'}^2}{N^3}\\
&= \frac{(N-1)}{N^3}[(N-1)(\mu^4+6\mu^2\sigma^2+3\sigma^4)\\
&\quad -(N-3)(\mu^2+\sigma^2)^2]\\
&= \frac{2(N-1)(\mu^4+2\mu^2N\sigma^2+N\sigma^4)}{N^3}. \qquad (49)
\end{aligned}$$

If $\mu = 0$, this expression simplifies to

$$\operatorname{var}(s^2) = \frac{2(N-1)N\sigma^4}{N^3} = \frac{2\sigma^4(N-1)}{N^2}, \qquad (50)$$

and the STANDARD ERROR is

$$[\text{standard error}] = \frac{\sqrt{2(N-1)}}{N}. \qquad (51)$$

The CUMULANT-GENERATING FUNCTION for a Gaussian distribution is

$$K(h) = \ln(e^{\nu_1 h}e^{\sigma^2h^2/2}) = \nu_1 h + \tfrac{1}{2}\sigma^2h^2, \qquad (52)$$

so

$$\kappa_1 = \nu_1 \qquad (53)$$

$$\kappa_2 = \sigma^2 \qquad (54)$$

$$\kappa_r = 0 \quad \text{for } r>2. \qquad (55)$$

For Gaussian variates, $\kappa_r = 0$ for $r > 2$, so the variance of k-STATISTIC k_3 is

$$\begin{aligned}
\operatorname{var}(k_3) &= \frac{\kappa_6}{N} + \frac{9\kappa_2\kappa_4}{N-1} + \frac{9{\kappa_3}^2}{N-1} + \frac{6{\kappa_2}^3}{N(N-1)(N-2)}\\
&= \frac{6{\kappa_2}^3}{N(N-1)(N-2)}. \qquad (56)
\end{aligned}$$

Also,

$$\operatorname{var}(k_4) = \frac{24{k_2}^4N(N-1)^2}{(N-3)(N-2)(N+3)(N+5)} \qquad (57)$$

$$\operatorname{var}(g_1) = \frac{6N(N-1)}{(N-2)(N+1)(N+3)} \qquad (58)$$

$$\operatorname{var}(g_2) = \frac{24N(N-1)^2}{(N-3)(N-2)(N+3)(N+5)}, \qquad (59)$$

where

$$g_1 \equiv \frac{k_3}{{k_2}^{3/2}} \qquad (60)$$

$$g_2 \equiv \frac{k_4}{{k_2}^2}. \qquad (61)$$

If $P(x)$ is a Gaussian distribution, then

$$D(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right], \qquad (62)$$

so variates x_i with a Gaussian distribution can be generated from variates y_i having a UNIFORM DISTRIBUTION in (0,1) via

$$x_i = \sigma\sqrt{2}\,\operatorname{erf}^{-1}(2y_i-1)+\mu. \qquad (63)$$

However, a simpler way to obtain numbers with a Gaussian distribution is to use the BOX-MULLER TRANSFORMATION.

The Gaussian distribution is an approximation to the BINOMIAL DISTRIBUTION in the limit of large numbers,

$$P(n_1) = \frac{1}{\sqrt{2\pi Npq}} \exp\left[-\frac{(n_1 - Np)^2}{2Npq}\right], \tag{64}$$

where n_1 is the number of steps in the POSITIVE direction, N is the number of trials ($N \equiv n_1 + n_2$), and p and q are the probabilities of a step in the POSITIVE direction and NEGATIVE direction ($q \equiv 1 - p$).

The differential equation having a Gaussian distribution as its solution is

$$\frac{dy}{dx} = \frac{y(\mu - x)}{\sigma^2}, \tag{65}$$

since

$$\frac{dy}{y} = \frac{\mu - x}{\sigma^2}\,dx \tag{66}$$

$$\ln\left(\frac{y}{y_0}\right) = -\frac{1}{2\sigma^2}(\mu - x)^2 \tag{67}$$

$$y = y_0 e^{-(x-\mu)^2/2\sigma^2} \tag{68}$$

This equation has been generalized to yield more complicated distributions which are named using the so-called PEARSON SYSTEM.

see also BINOMIAL DISTRIBUTION, CENTRAL LIMIT THEOREM, ERF, GAUSSIAN BIVARIATE DISTRIBUTION, LOGIT TRANSFORMATION, NORMAL DISTRIBUTION, NORMAL DISTRIBUTION FUNCTION, PEARSON SYSTEM, RATIO DISTRIBUTION, z-SCORE

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 533–534, 1987.

Kraitchik, M. "The Error Curve." §6.4 in *Mathematical Recreations.* New York: W. W. Norton, pp. 121–123, 1942.

Spiegel, M. R. *Theory and Problems of Probability and Statistics.* New York: McGraw-Hill, p. 109–111, 1992.

Gaussian Distribution—Linear Combination of Variates

If x is NORMALLY DISTRIBUTED with MEAN μ and VARIANCE σ^2, then a linear function of x,

$$y = ax + b, \tag{1}$$

is also NORMALLY DISTRIBUTED. The new distribution has MEAN $a\mu + b$ and VARIANCE $a^2\sigma^2$, as can be derived using the MOMENT-GENERATING FUNCTION

$$\begin{aligned} M(t) &= \left\langle e^{t(ax+b)}\right\rangle = e^{tb}\left\langle e^{atx}\right\rangle = e^{tb}e^{\mu at+\sigma^2(at)^2/2} \\ &= e^{tb+\mu at+\sigma^2a^2t^2/2} = e^{(b+a\mu)t+a^2\sigma^2t^2/2}, \end{aligned} \tag{2}$$

which is of the standard form with

$$\mu' = b + a\mu \tag{3}$$

$$\sigma'^2 = a^2\sigma^2. \tag{4}$$

For a weighted sum of independent variables

$$y \equiv \sum_{i=1}^{n} a_i x_i, \tag{5}$$

the expectation is given by

$$\begin{aligned} M(t) &= \left\langle e^{yt}\right\rangle = \left\langle \exp\left(t\sum_{i=1}^{n} a_i x_i\right)\right\rangle \\ &= \left\langle e^{a_1tx_1}e^{a_2tx_2}\cdots e^{a_ntx_n}\right\rangle \\ &= \prod_{i=1}^{n}\left\langle e^{a_itx_i}\right\rangle = \prod_{i=1}^{n}\exp(a_i\mu_it + \tfrac{1}{2}a_i{}^2\sigma_i{}^2t^2). \end{aligned} \tag{6}$$

Setting this equal to

$$\exp(\mu t + \tfrac{1}{2}\sigma^2t^2) \tag{7}$$

gives

$$\mu \equiv \sum_{i=1}^{n} a_i\mu_i \tag{8}$$

$$\sigma^2 \equiv \sum_{i=1}^{n} a_i{}^2\sigma_i{}^2. \tag{9}$$

Therefore, the MEAN and VARIANCE of the weighted sums of n RANDOM VARIABLES are their weighted sums.

If x_i are INDEPENDENT and NORMALLY DISTRIBUTED with MEAN 0 and VARIANCE σ^2, define

$$y_i \equiv \sum_j c_{ij}x_j, \tag{10}$$

where c obeys the ORTHOGONALITY CONDITION

$$c_{ik}c_{jk} = \delta_{ij}, \tag{11}$$

with δ the KRONECKER DELTA. Then y_i are also independent and normally distributed with MEAN 0 and VARIANCE σ^2.

Gaussian Elimination

A method for solving MATRIX EQUATIONS of the form

$$\mathsf{Ax} = \mathsf{b}. \tag{1}$$

Starting with the system of equations

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}, \tag{2}$$

compose the augmented MATRIX equation

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1k} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2k} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} & b_k \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}. \tag{3}$$

Then, perform MATRIX operations to put the augmented MATRIX into the form

$$\left[\begin{array}{cccc|c} a'_{11} & a'_{12} & \cdots & a'_{1k} & b'_1 \\ 0 & a'_{22} & \cdots & a'_{2k} & b'_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & a'_{kk} & b'_k \end{array}\right] \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_k \end{bmatrix}. \tag{4}$$

Solve for a'_{kk}, then substitute back in to obtain solutions for $n = 1, 2, \ldots, k-1$,

$$x_i = \frac{1}{a'_{ii}} \left(b'_i - \sum_{j=i+1}^{k} a'_{ij} x'_j \right). \tag{5}$$

see also GAUSS-JORDAN ELIMINATION, LU DECOMPOSITION, MATRIX EQUATION, SQUARE ROOT METHOD

Gaussian Function

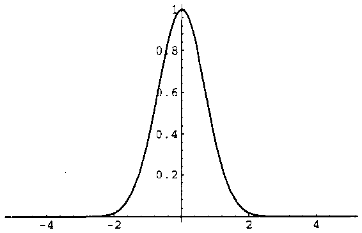

In 1-D, the Gaussian function is the function from the GAUSSIAN DISTRIBUTION,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \tag{1}$$

sometimes also called the FREQUENCY CURVE. The FULL WIDTH AT HALF MAXIMUM (FWHM) for a Gaussian is found by finding the half-maximum points x_0. The constant scaling factor can be ignored, so we must solve

$$e^{-(x_0-\mu)^2/2\sigma^2} = \tfrac{1}{2} f(x_{\max}) \tag{2}$$

But $f(x_{\max})$ occurs at $x_{\max} = \mu$, so

$$e^{-(x_0-\mu)^2/2\sigma^2} = \tfrac{1}{2} f(\mu) = \tfrac{1}{2}. \tag{3}$$

Solving,

$$e^{-(x_0-\mu)^2/2\sigma^2} = 2^{-1} \tag{4}$$

$$-\frac{(x_0-\mu)^2}{2\sigma^2} = -\ln 2 \tag{5}$$

$$(x_0-\mu)^2 = 2\sigma^2 \ln 2 \tag{6}$$

$$x_0 = \pm\sigma\sqrt{2\ln 2} + \mu. \tag{7}$$

The FULL WIDTH AT HALF MAXIMUM is therefore given by

$$\text{FWHM} \equiv x_+ - x_- = 2\sqrt{2\ln 2}\,\sigma \approx 2.3548\sigma. \tag{8}$$

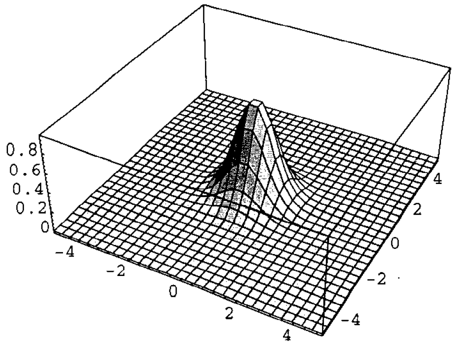

In 2-D, the circular Gaussian function is the distribution function for uncorrelated variables x and y having a GAUSSIAN BIVARIATE DISTRIBUTION and equal STANDARD DEVIATION $\sigma = \sigma_x = \sigma_y$,

$$f(x,y) = \frac{1}{2\pi\sigma^2} e^{-[(x-\mu_x)^2+(y-\mu_y)^2]/2\sigma^2}. \tag{9}$$

The corresponding elliptical Gaussian function corresponding to $\sigma_x \neq \sigma_y$ is given by

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-[(x-\mu_x)^2/2\sigma_x{}^2+[(y-\mu_y)^2/2\sigma_y{}^2]}. \tag{10}$$

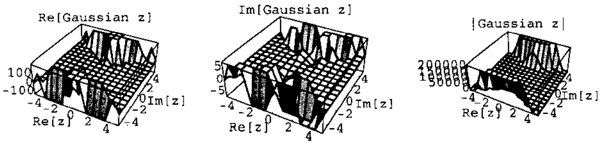

The above plots show the real and imaginary parts of $(2\pi)^{-1/2} e^{-z^2}$ together with the complex absolute value $|(2\pi)^{-1/2} e^{-z^2}|$.

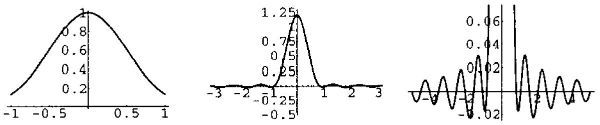

The Gaussian function can also be used as an APODIZATION FUNCTION, shown above with the corresponding INSTRUMENT FUNCTION.

The HYPERGEOMETRIC FUNCTION is also sometimes known as the Gaussian function.

see also ERF, ERFC, FOURIER TRANSFORM—GAUSSIAN, GAUSSIAN BIVARIATE DISTRIBUTION, GAUSSIAN DISTRIBUTION, NORMAL DISTRIBUTION

References

MacTutor History of Mathematics Archive. "Frequency Curve." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Frequency.html`.

Gaussian Hypergeometric Series

see HYPERGEOMETRIC FUNCTION

Gaussian Integer

A COMPLEX NUMBER $a+bi$ where a and b are INTEGERS. The Gaussian integers are members of the QUADRATIC FIELD $\mathbb{Q}(\sqrt{-1})$. The sum, difference, and product of two Gaussian integers are Gaussian integers, but $a+bi|c+di$ only if there is an $e+fi$ such that

$$(a+bi)(e+fi) = (ae-bf)+(af+be)i = c+di.$$

Gaussian INTEGERS can be uniquely factored in terms of other Gaussian INTEGERS up to POWERS of i and rearrangements.

The norm of a Gaussian integer is defined by

$$n(x+iy) = x^2+y^2.$$

GAUSSIAN PRIMES are Gaussian integers $a+ib$ for which $n(a+ib) = a^2+b^2$ is PRIME and a a PRIME INTEGER a such that $a \equiv 3 \pmod 4$.

1. If $2|n(x+iy)$, then $1+i$ and $1-i|x+iy$. These factors are equivalent since $-i(i-1) = i+1$. For example, $2 = (1+i)(1-i)$ is not a Gaussian prime.
2. If $n(x+iy) \equiv 3 \pmod 4)\,|n(x+iy)$, then $n(a+ib)|x+iy$.
3. If $n(x+iy) \equiv 1 \pmod 4)\,|n(x+iy)$, then $a+ib$ or $b+ia|x+iy$. If both do, then $n(a+ib)|x+iy$.

The Gaussian primes with $|a|, |b| \leq 5$ are given by $-5-4i$, $-5-2i$, $-5+2i$, $-5+4i$, $-4-5i$, $-4-i$, $-4+i$, $-4+5i$, $-3-2i$, -3, $-3+2i$, $-2-5i$, $-2-3i$, $-2-i$, $-2+i$, $-2+3i$, $-2+5i$, $-1-4i$, $-1-2i$, $-1-i$, $-1+i$, $-1+2i$, $-1+4i$, $-3i$, $3i$, $1-4i$, $1-2i$, $1-i$, $1+i$, $1+2i$, $1+4i$, $2-5i$, $2-3i$, $2-i$, $2+i$, $2+3i$, $2+5i$, $3-2i$, 3, $3+2i$, $4-5i$, $4-i$, $4+i$, $4+5i$, $5-4i$, $5-2i$, $5+2i$, $5+4i$.

Every Gaussian integer is within $|n|/\sqrt{2}$ of a multiple of a Gaussian integer n.

see also COMPLEX NUMBER, EISENSTEIN INTEGER, GAUSSIAN PRIME, INTEGER, OCTONION

References

Conway, J. H. and Guy, R. K. "Gauss's Whole Numbers." In *The Book of Numbers.* New York: Springer-Verlag, pp. 217–223, 1996.

Shanks, D. "Gaussian Integers and Two Applications." §50 in *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 149–151, 1993.

Gaussian Integral

The Gaussian integral, also called the PROBABILITY INTEGRAL, is the integral of the 1-D Gaussian over $(-\infty, \infty)$. It can be computed using the trick of combining two 1-D Gaussians

$$\begin{aligned}\int_{-\infty}^{\infty} e^{-x^2}\,dx &= \sqrt{\left(\int_{-\infty}^{\infty} e^{-y^2}\,dy\right)\left(\int_{-\infty}^{\infty} e^{-x^2}\,dx\right)} \\ &= \sqrt{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\,dy\,dx} \qquad (1)\end{aligned}$$

and switching to POLAR COORDINATES,

$$\begin{aligned}\int_{-\infty}^{\infty} e^{-x^2}\,dx &= \sqrt{\int_0^{2\pi}\int_0^{\infty} e^{-r^2} r\,dr\,d\theta} \\ &= \sqrt{2\pi\left[-\tfrac{1}{2}e^{-r^2} r\right]_0^{\infty}} = \sqrt{\pi}. \qquad (2)\end{aligned}$$

However, a simple proof can also be given which does not require transformation to POLAR COORDINATES (Nicholas and Yates 1950).

The integral from 0 to a finite upper limit a can be given by the CONTINUED FRACTION

$$\int_0^a e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}\,\frac{1}{a+}\,\frac{2}{2a+}\,\frac{3}{a+}\,\frac{4}{2a+\ldots}. \qquad (3)$$

The general class of integrals of the form

$$I_n(a) \equiv \int_0^{\infty} e^{-ax^2} x^n\,dx \qquad (4)$$

can be solved analytically by setting

$$x \equiv a^{-1/2} y \qquad (5)$$

$$dx = a^{-1/2}\,dy \qquad (6)$$

$$y^2 = ax^2. \qquad (7)$$

Then

$$\begin{aligned}I_n(a) &= a^{-1/2}\int_0^{\infty} e^{-y^2}(a^{-1/2})^n\,dy \\ &= a^{-(1+n)/2}\int_0^{\infty} e^{-y^2} y^n\,dy. \qquad (8)\end{aligned}$$

For $n = 0$, this is just the usual Gaussian integral, so

$$I_0(a) = \frac{\sqrt{\pi}}{2}a^{-1/2} = \frac{1}{2}\sqrt{\frac{\pi}{a}}. \qquad (9)$$

For $n = 1$, the integrand is integrable by quadrature,

$$I_1(a) = a^{-1}\int_0^{\infty} e^{-y^2} y\,dy = a^{-1}[-\tfrac{1}{2}e^{-y^2}]_0^{\infty} = \tfrac{1}{2}a^{-1}. \qquad (10)$$

To compute $I_n(a)$ for $n > 1$, use the identity

$$\begin{aligned} -\frac{\partial}{\partial a} I_{n-2}(a) &= -\frac{\partial}{\partial a} \int_0^\infty e^{-ax^2} x^{n-2}\, dx \\ &= -\int_0^\infty -x^2 e^{-ax^2} x^{n-2}\, dx \\ &= \int_0^\infty e^{-ax^2} x^n\, dx = I_n(a). \end{aligned} \quad (11)$$

For $n = 2s$ EVEN,

$$\begin{aligned} I_n(a) &= \left(-\frac{\partial}{\partial a}\right) I_{n-2}(a) = \left(-\frac{\partial}{\partial a}\right)^2 I_{n-4} \\ &= \ldots = \left(-\frac{\partial}{\partial a}\right)^{n/2} I_0(a) \\ &= \frac{\partial^{n/2}}{\partial a^{n/2}} I_0(a) = \frac{\sqrt{\pi}}{2} \frac{\partial^{n/2}}{\partial a^{n/2}} a^{-1/2}, \end{aligned} \quad (12)$$

so

$$\int_0^\infty x^{2s} e^{-ax^2}\, dx = \frac{(s-\frac{1}{2})!}{2a^{s+1/2}} = \frac{(2s-1)!!}{2^{s+1} a^s} \sqrt{\frac{\pi}{a}}. \quad (13)$$

If $n = 2s+1$ is ODD, then

$$\begin{aligned} I_n(a) &= \left(-\frac{\partial}{\partial a}\right) I_{n-2}(a) = \left(-\frac{\partial}{\partial a}\right)^2 I_{n-4}(a) \\ &= \ldots = \left(-\frac{\partial}{\partial a}\right)^{(n-1)/2} I_1(a) \\ &= \frac{\partial^{(n-1)/2}}{\partial a^{(n-1)/2}} I_1(a) = \frac{1}{2} \frac{\partial^{(n-1)/2}}{\partial a^{(n-1)/2}} a^{-1}, \end{aligned} \quad (14)$$

so

$$\int_0^\infty x^{2s+1} e^{-ax^2}\, dx = \frac{s!}{2a^{s+1}}. \quad (15)$$

The solution is therefore

$$\int_0^\infty e^{-ax^2} x^n\, dx = \begin{cases} \frac{(n-1)!!}{2^{n/2+1} a^{n/2}} \sqrt{\frac{\pi}{a}} & \text{for } n \text{ even} \\ \frac{[(n+1)/2]!}{2a^{(n+1)/2}} & \text{for } n \text{ odd.} \end{cases} \quad (16)$$

The first few values are therefore

$$I_0(a) = \frac{1}{2}\sqrt{\frac{\pi}{a}} \quad (17)$$

$$I_1(a) = \frac{1}{2a} \quad (18)$$

$$I_2(a) = \frac{1}{4a}\sqrt{\frac{\pi}{a}} \quad (19)$$

$$I_3(a) = \frac{1}{2a^2} \quad (20)$$

$$I_4(a) = \frac{3}{8a^2}\sqrt{\frac{\pi}{a}} \quad (21)$$

$$I_5(a) = \frac{1}{a^3} \quad (22)$$

$$I_6(a) = \frac{15}{16a^3}\sqrt{\frac{\pi}{a}}. \quad (23)$$

A related, often useful integral is

$$H_n(a) \equiv \frac{1}{\sqrt{\pi}} \int_{-\infty}^\infty e^{-ax^2} x^n\, dx, \quad (24)$$

which is simply given by

$$H_n = \begin{cases} \frac{2I_n(a)}{\sqrt{\pi}} & \text{for } n \text{ even} \\ 0 & \text{for } n \text{ odd.} \end{cases} \quad (25)$$

References

Nicholas, C. B. and Yates, R. C. "The Probability Integral." *Amer. Math. Monthly* **57**, 412–413, 1950.

Gaussian Integral (Linking Number)

see LINKING NUMBER

Gaussian Joint Variable Theorem

Also called the MULTIVARIATE THEOREM. Given an EVEN number of variates from a NORMAL DISTRIBUTION with MEANS all 0,

$$\langle x_1 x_2 \rangle = \langle x_1 \rangle \langle x_2 \rangle, \quad (1)$$

$$\begin{aligned} &\langle x_1 x_2 x_3 x_4 \rangle \\ &\quad = \langle x_1 x_2 \rangle \langle x_3 x_4 \rangle + \langle x_1 x_3 \rangle \langle x_2 x_4 \rangle + \langle x_1 x_4 \rangle \langle x_2 x_3 \rangle, \end{aligned} \quad (2)$$

etc. Given an ODD number of variates,

$$\langle x_1 \rangle = 0, \quad (3)$$

$$\langle x_1 x_2 x_3 \rangle = 0, \quad (4)$$

etc.

Gaussian Mountain Range

see CAROTID-KUNDALINI FUNCTION

Gaussian Multivariate Distribution

see also GAUSSIAN BIVARIATE DISTRIBUTION, JOINT THEOREM, MULTIVARIATE THEOREM

Gaussian Polynomial

Defined by

$$[l] \equiv \frac{1-q^l}{1-q} \quad (1)$$

for integral l, and

$$\begin{bmatrix} n \\ k \end{bmatrix} \equiv \begin{cases} \prod_{l=1}^k \frac{[n-l+1]}{[l]} & \text{for } 0 \leq k \leq n \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Unfortunately, the NOTATION conflicts with that of GAUSSIAN BRACKETS and the NEAREST INTEGER

FUNCTION. Gaussian POLYNOMIALS satisfy the identities

$$\frac{\begin{bmatrix} n+1 \\ k+1 \end{bmatrix}}{\begin{bmatrix} n \\ k+1 \end{bmatrix}} = \frac{1-q^{n+1}}{1-q^{n-k}} \tag{3}$$

$$\frac{\begin{bmatrix} n+1 \\ k+1 \end{bmatrix}}{\begin{bmatrix} n+1 \\ k \end{bmatrix}} = \frac{1-q^{n-k+1}}{1-q^{k+1}}. \tag{4}$$

For $q = 1$, the Gaussian polynomial turns into the BINOMIAL COEFFICIENT.

see also BINOMIAL COEFFICIENT, GAUSSIAN COEFFICIENT, q-SERIES

Gaussian Prime

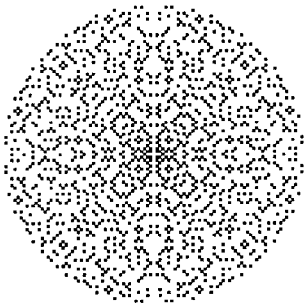

Gaussian primes are GAUSSIAN INTEGERS $a + ib$ for which $n(a + ib) = a^2 + b^2$ is PRIME and a a PRIME INTEGER a such that $a \equiv 3 \pmod 4$. The above plot of the COMPLEX PLANE shows the Gaussian primes as filled squares.

see also EISENSTEIN INTEGER, GAUSSIAN INTEGER

References

Guy, R. K. "Gaussian Primes. Eisenstein-Jacobi Primes." §A16 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 33–36, 1994.

Wagon, S. "Gaussian Primes." §9.4 in *Mathematica in Action.* New York: W. H. Freeman, pp. 298–303, 1991.

Gaussian Quadrature

Seeks to obtain the best numerical estimate of an integral by picking optimal ABSCISSAS x_i at which to evaluate the function $f(x)$. The FUNDAMENTAL THEOREM OF GAUSSIAN QUADRATURE states that the optimal ABSCISSAS of the m-point GAUSSIAN QUADRATURE FORMULAS are precisely the roots of the orthogonal POLYNOMIAL for the same interval and WEIGHTING FUNCTION. Gaussian quadrature is optimal because it fits all POLYNOMIALS up to degree $2m$ exactly. Slightly less optimal fits are obtained from RADAU QUADRATURE and LAGUERRE QUADRATURE.

$W(x)$	Interval	x_i Are Roots Of
1	$(-1,1)$	$P_n(x)$
e^{-t}	$(0,\infty)$	$L_n(x)$
e^{-t^2}	$(-\infty,\infty)$	$H_n(x)$
$(1-t^2)^{-1/2}$	$(-1,1)$	$T_n(x)$
$(1-t^2)^{1/2}$	$(-1,1)$	$U_n(x)$
$x^{1/2}$	$(0,1)$	$x^{-1/2}P_{2n+1}(\sqrt{x})$
$x^{-1/2}$	$(0,1)$	$P_n(\sqrt{x})$

To determine the weights corresponding to the Gaussian ABSCISSAS, compute a LAGRANGE INTERPOLATING POLYNOMIAL for $f(x)$ by letting

$$\pi(x) \equiv \prod_{j=1}^{m} (x - x_j) \tag{1}$$

(where Chandrasekhar 1967 uses F instead of π), so

$$\pi'(x_j) = \left[\frac{d\pi}{dx}\right]_{x=x_j} = \prod_{\substack{i=1\\ i\neq j}}^{m} (x_j - x_i). \tag{2}$$

Then fitting a LAGRANGE INTERPOLATING POLYNOMIAL through the m points gives

$$\phi(x) = \sum_{j=1}^{m} \frac{\pi(x)}{(x - x_j)\pi'(x_j)} f(x_j) \tag{3}$$

for arbitrary points x_i. We are therefore looking for a set of points x_j and weights w_j such that for a WEIGHTING FUNCTION $W(x)$,

$$\begin{aligned} \int_a^b \phi(x)W(x)\,dx &= \int_a^b \sum_{j=1}^{m} \frac{\pi(x)W(x)}{(x - x_j)\pi'(x_j)}\,dx\, f(x_j) \\ &\equiv \sum_{j=1}^{m} w_j f(x_j), \end{aligned} \tag{4}$$

with WEIGHT

$$w_j = \frac{1}{\pi'(x_j)} \int_a^b \frac{\pi(x)W(x)}{x - x_j}\,dx. \tag{5}$$

The weights w_j are sometimes also called the CHRISTOFFEL NUMBER (Chandrasekhar 1967). For orthogonal POLYNOMIALS $\phi_j(x)$ with j=1, …, n,

$$\phi_j(x) = A_j \pi(x) \tag{6}$$

(Hildebrand 1956, p. 322), where A_n is the COEFFICIENT of x^n in $\phi_n(x)$, then

$$\begin{aligned} w_j &= \frac{1}{\phi_n'(x_j)} \int_a^b W(x) \frac{\phi(x)}{x - x_j}\,dx \\ &= -\frac{A_{n+1}\gamma_n}{A_n \phi_n'(x_j)\phi_{n+1}(x)}, \end{aligned} \tag{7}$$

where

$$\gamma_m \equiv \int [\phi_m(x)]^2 W(x)\,dx. \tag{8}$$

Using the relationship

$$\phi_{n+1}(x_i) = -\frac{A_{n+1}A_{n-1}}{{A_n}^2}\frac{\gamma_n}{\gamma_{n-1}}\phi_{n-1}(x_i) \tag{9}$$

(Hildebrand 1956, p. 323) gives

$$w_j = \frac{A_n}{A_{n-1}}\frac{\gamma_{n-1}}{\phi_n'(x_j)\phi_{n-1}(x_j)}. \tag{10}$$

(Note that Press *et al.* 1992 omit the factor A_n/A_{n-1}.) In Gaussian quadrature, the weights are all POSITIVE. The error is given by

$$E_n = \frac{f^{(2n)}(\xi)}{(2n)!}\int_a^b W(x)[\pi(x)]^2\,dx = \frac{\gamma_n}{{A_n}^2}\frac{f^{(2n)}(\xi)}{(2n)!}, \tag{11}$$

where $a < \xi < b$ (Hildebrand 1956, pp. 320–321).

Other curious identities are

$$\sum_{k=0}^{m}\frac{[\phi_k(x)]^2}{\gamma_k} = \frac{A_m}{A_{m+1}\gamma_m}[\phi_{m+1}'(x)\phi_m(x) - \phi_m'(x)\phi_{m+1}(x)] \tag{12}$$

and

$$\sum_{k=0}^{m}\frac{[\phi_k(x)]^2}{\gamma_k} = -\frac{A_m\phi_m'(x_i)\phi_{m+1}(x_i)}{A_{m+1}\gamma_m} = \frac{1}{w_i} \tag{13}$$

(Hildebrand 1956, p. 323).

In the NOTATION of Szegő (1975), let $x_{1n} < \ldots < x_{nn}$ be an ordered set of points in $[a,b]$, and let $\lambda_{1n}, \ldots, \lambda_{nn}$ be a set of REAL NUMBERS. If $f(x)$ is an arbitrary function on the CLOSED INTERVAL $[a,b]$, write the MECHANICAL QUADRATURE as

$$Q_n(f) = \sum_{\nu=1}^{n}\lambda_{\nu n}f(x_{\nu n}). \tag{14}$$

Here $x_{\nu n}$ are the ABSCISSAS and $\lambda_{\nu n}$ are the COTES NUMBERS.

see also CHEBYSHEV QUADRATURE, CHEBYSHEV-GAUSS QUADRATURE, CHEBYSHEV-RADAU QUADRATURE, FUNDAMENTAL THEOREM OF GAUSSIAN QUADRATURE, HERMITE-GAUSS QUADRATURE, JACOBI-GAUSS QUADRATURE, LAGUERRE-GAUSS QUADRATURE, LEGENDRE-GAUSS QUADRATURE, LOBATTO QUADRATURE, MEHLER QUADRATURE, RADAU QUADRATURE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 887–888, 1972.

Acton, F. S. *Numerical Methods That Work, 2nd printing.* Washington, DC: Math. Assoc. Amer., p. 103, 1990.

Arfken, G. "Appendix 2: Gaussian Quadrature." *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 968–974, 1985.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 461, 1987.

Chandrasekhar, S. *An Introduction to the Study of Stellar Structure.* New York: Dover, 1967.

Hildebrand, F. B. *Introduction to Numerical Analysis.* New York: McGraw-Hill, pp. 319–323, 1956.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Gaussian Quadratures and Orthogonal Polynomials." §4.5 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 140–155, 1992.

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., pp. 37–48 and 340–349, 1975.

Whittaker, E. T. and Robinson, G. *The Calculus of Observations: A Treatise on Numerical Mathematics, 4th ed.* New York: Dover, pp. 152–163, 1967.

Gaussian Sum

$$S(p,q) \equiv \sum_{r=0}^{q-1} e^{-\pi i r^2 p/q}, \tag{1}$$

where p and q are RELATIVELY PRIME INTEGERS. If $(n,n') = 1$, then

$$S(m,nn') = S(mn',n)S(mn,n'). \tag{2}$$

Gauss showed

$$\sum_{r=0}^{q-1} e^{2\pi i r^2/q} = \frac{1-i^q}{1-i}\sqrt{q} \tag{3}$$

for ODD q. A more general result was obtained by Schaar. For p and q of opposite PARITY (i.e., one is EVEN and the other is ODD), SCHAAR'S IDENTITY states

$$\frac{1}{\sqrt{q}}\sum_{r=0}^{q-1} e^{-\pi i r^2 p/q} = \frac{e^{-\pi i/4}}{\sqrt{p}}\sum_{r=0}^{p-1} e^{\pi i r^2 q/p}. \tag{4}$$

Such sums are important in the theory of QUADRATIC RESIDUES.

see also KLOOSTERMAN'S SUM, SCHAAR'S IDENTITY, SINGULAR SERIES

References

Evans, R. and Berndt, B. "The Determination of Gauss Sums." *Bull. Amer. Math. Soc.* **5**, 107–129, 1981.

Katz, N. M. *Gauss Sums, Kloosterman Sums, and Monodromy Groups.* Princeton, NJ: Princeton University Press, 1987.

Riesel, H. *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, pp. 132–134, 1994.

Gear Graph

A WHEEL GRAPH with a VERTEX added between each pair of adjacent VERTICES.

Gegenbauer Function

see ULTRASPHERICAL FUNCTION

Gegenbauer Polynomial

see ULTRASPHERICAL POLYNOMIAL

Gelfond-Schneider Constant

The number $2^{\sqrt{2}} = 2.66514414\ldots$ which is known to be TRANSCENDENTAL by GELFOND'S THEOREM.

References
Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 107, 1996.

Gelfond-Schneider Theorem

see GELFOND'S THEOREM

Gelfond's Theorem

Also called the GELFOND-SCHNEIDER THEOREM. a^b is TRANSCENDENTAL if

1. a is ALGEBRAIC $\neq 0$, 1 and
2. b is ALGEBRAIC and IRRATIONAL.

This provides the solution to the seventh of HILBERT'S PROBLEMS.

see also ALGEBRAIC NUMBER, HILBERT'S PROBLEMS, IRRATIONAL NUMBER, TRANSCENDENTAL NUMBER

References
Baker, A. *Transcendental Number Theory.* London: Cambridge University Press, 1990.
Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 107, 1996.

Genaille Rods

Numbered rods which can be used to perform multiplication.

see also NAPIER'S BONES

References
Gardner, M. "Napier's Bones." Ch. 7 in *Knotted Doughnuts and Other Mathematical Entertainments.* New York: W. H. Freeman, 1986.

Genera

see FUNDAMENTAL THEOREM OF GENERA

General Linear Group

The general linear group $GL_n(q)$ is the set of $n \times n$ MATRICES with entries in the FIELD $\mathbb{F}_q$ which have NONZERO DETERMINANT.

see also LANGLANDS RECIPROCITY, PROJECTIVE GENERAL LINEAR GROUP, PROJECTIVE SPECIAL LINEAR GROUP, SPECIAL LINEAR GROUP

References
Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. "The Groups $GL_n(q)$, $SL_n(q)$, $PGL_n(q)$, and $PSL_n(q) = L_n(q)$." §2.1 in *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, p. x, 1985.

General Orthogonal Group

The general orthogonal group $GO_n(q,F)$ is the SUBGROUP of all elements of the PROJECTIVE GENERAL LINEAR GROUP that fix the particular nonsingular QUADRATIC FORM F. The determinant of such an element is ±1.

see also PROJECTIVE GENERAL LINEAR GROUP

References
Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. "The Groups $GO_n(q)$, $SO_n(q)$, $PGO_n(q)$, and $PSO_n(q)$, and $O_n(q)$." §2.4 in *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, pp. xi–xii, 1985.

General Position

An arrangement of points with no three COLLINEAR, or of lines with no three concurrent.

see also ORDINARY LINE, NEAR-PENCIL

References
Guy, R. K. "Unsolved Problems Come of Age." *Amer. Math. Monthly* **96**, 903–909, 1989.

General Prismatoid

A solid such that the AREA A_y of any section parallel to and a distance y from a fixed PLANE can be expressed as

$$A_y = ay^3 + by^2 + cy + d.$$

The volume of such a solid is the same as for a PRISMATOID,

$$V = \tfrac{1}{6}h(A_1 + 4M + A_2).$$

Examples include the CONE, CYLINDER, PRISMATOID, SPHERE, and SPHEROID.

see also PRISMATOID, PRISMOID

References
Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 132, 1987.

General Unitary Group

The general unitary group $GU_n(q)$ is the SUBGROUP of all elements of the GENERAL LINEAR GROUP $GL(q^2)$ that fix a given nonsingular Hermitian form. This is equivalent, in the canonical case, to the definition of GU_n as the group of UNITARY MATRICES.

References
Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. "The Groups $GU_n(q)$, $SU_n(q)$, $PGU_n(q)$, and $PSU_n(q) = U_n(q)$." §2.2 in *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, p. x, 1985.

Generalized Cone

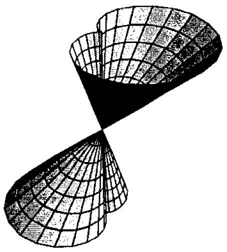

A RULED SURFACE is called a generalized cone if it can be parameterized by $\mathbf{x}(u,v) = \mathbf{p} + v\mathbf{y}(u)$, where $\mathbf{p}$ is a fixed point which can be regarded as the vertex of the cone. A generalized cone is a REGULAR SURFACE wherever $v\mathbf{y}\times\mathbf{y}' \neq \mathbf{0}$. The above surface is a generalized cylinder over a CARDIOID. A generalized cone is a FLAT SURFACE.

see also CONE

References
Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 341–342, 1993.

Generalized Cylinder

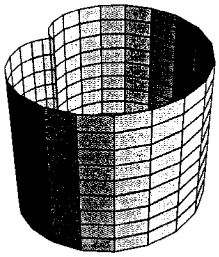

A RULED SURFACE is called a generalized cylinder if it can be parameterized by $\mathbf{x}(u,v) = v\mathbf{p} + \mathbf{y}(u)$, where $\mathbf{p}$ is a fixed point. A generalized cylinder is a REGULAR SURFACE wherever $\mathbf{y}' \times \mathbf{p} \neq \mathbf{0}$. The above surface is a generalized cylinder over a CARDIOID. A generalized cylinder is a FLAT SURFACE.

see also CYLINDER

References
Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 341–342, 1993.

Generalized Fibonacci Number

A generalization of the FIBONACCI NUMBERS defined by $1 = G_1 = G_2 = \ldots = G_{c-1}$ and the RECURRENCE RELATION

$$G_n = G_{n-1} + G_{n-c}. \qquad (1)$$

These are the sums of elements on successive diagonals of a left-justified PASCAL'S TRIANGLE beginning in the left-most column and moving in steps of $c-1$ up and 1 right. The case $c = 2$ equals the usual FIBONACCI NUMBER. These numbers satisfy the identities

$$G_1 + G_2 + G_3 + \ldots + G_n = G_{n+3} - 1 \qquad (2)$$

$$G_3 + G_6 + G_9 + \ldots + G_{3k} = G_{3k+1} - 1 \qquad (3)$$

$$G_1 + G_4 + G_7 + \ldots + G_{3k+1} = G_{3k+2} \qquad (4)$$

$$G_2 + G_5 + G_8 + \ldots + G_{3k+2} = G_{3k+3} \qquad (5)$$

(Bicknell-Johnson and Spears 1996). For the special case $c = 3$,

$$G_{n+w} = G_{w-2}G_n + G_{w-3}G_{n+1} + G_{w-1}G_{n+2}. \qquad (6)$$

Bicknell-Johnson and Spears (1996) give many further identities.

Horadam (1965) defined the generalized Fibonacci numbers $\{w_n\}$ as $w_n = w_n(a,b;p,q)$, where a, b, p, and q are INTEGERS, $w_0 = a$, $w_1 = b$, and $w_n = pw_{n-1} - qw_{n-2}$ for $n \geq 2$. They satisfy the identities

$$w_n w_{n+2r} - eq^n U_r = w_{n+r}{}^2 \qquad (7)$$

$$4w_n w_{n+1}{}^2 w_{n+2} + (wq^n)^2 = (w_n w_{n+2} + w_{n+1}{}^2)^2 \qquad (8)$$

$$\begin{aligned} &w_n w_{n+1} w_{n+3} w_{n+4} \\ &\quad = w_{n+2}{}^4 + eq^n(p^2+q)w_{n+2}{}^2 + e^2q^{2n+1}p^2 \qquad (9) \end{aligned}$$

$$\begin{aligned} &4w_n w_{n+1} w_{n+2} w_{n+4} w_{n+5} w_{n+6} \\ &\quad + e^2q^{2n}(w_nU_4U_5 - w_{n+1}U_2U_6 - w_nU_1U_8)^2 \\ &\quad = (w_{n+1}w_{n+2}w_{n+6} + w_nw_{n+4}w_{n+5})^2, \qquad (10) \end{aligned}$$

where

$$e \equiv pab - qa^2 - b^2 \qquad (11)$$

$$U_n \equiv w_n(0,1;p,q). \qquad (12)$$

The final above result is due to Morgado (1987) and is called the MORGADO IDENTITY.

Another generalization of the Fibonacci numbers is denoted x_n. Given x_1 and x_2, define the generalized Fibonacci number by $x_n \equiv x_{n-2} + x_{n-1}$ for $n \geq 3$,

$$\sum_{i=1}^{n} x_n = x_{n+2} - x_2 \qquad (13)$$

$$\sum_{i=1}^{10} x_n = 11x_7 \tag{14}$$

$$x_n{}^2 - x_{n-1}x_{n+2} = (-1)^n(x_2{}^2 - x_1{}^2 - x_1x_2), \tag{15}$$

where the plus and minus signs alternate.

see also FIBONACCI NUMBER

References

Bicknell, M. "A Primer for the Fibonacci Numbers, Part VIII: Sequences of Sums from Pascal's Triangle." *Fib. Quart.* **9**, 74–81, 1971.

Bicknell-Johnson, M. and Spears, C. P. "Classes of Identities for the Generalized Fibonacci Numbers $G_n = G_{n-1}+G_{n-c}$ for Matrices with Constant Valued Determinants." *Fib. Quart.* **34**, 121–128, 1996.

Dujella, A. "Generalized Fibonacci Numbers and the Problem of Diophantus." *Fib. Quart.* **34**, 164–175, 1996.

Horadam, A. F. "Generating Functions for Powers of a Certain Generalized Sequence of Numbers." *Duke Math. J.* **32**, 437–446, 1965.

Horadam, A. F. "Generalization of a Result of Morgado." *Portugaliae Math.* **44**, 131–136, 1987.

Horadam, A. F. and Shannon, A. G. "Generalization of Identities of Catalan and Others." *Portugaliae Math.* **44**, 137–148, 1987.

Morgado, J. "Note on Some Results of A. F. Horadam and A. G. Shannon Concerning a Catalan's Identity on Fibonacci Numbers." *Portugaliae Math.* **44**, 243–252, 1987.

Generalized Function

The class of all regular sequences of PARTICULARLY WELL-BEHAVED FUNCTIONS equivalent to a given regular sequence (sometimes also called a DISTRIBUTION or FUNCTIONAL). A generalized function $p(x)$ has the properties

$$\int_{-\infty}^{\infty} p'(x)f(x)\,dx = -\int_{-\infty}^{\infty} p(x)f'(x)\,dx$$

$$\int_{-\infty}^{\infty} p^{(n)}f(x)\,dx = (-1)^n\int_{-\infty}^{\infty} p(x)f^{(n)}(x)\,dx.$$

The DELTA FUNCTION is a generalized function.

see also DELTA FUNCTION

Generalized Helicoid

The SURFACE generated by a twisted curve C when rotated about a fixed axis A and, at the same time, displaced PARALLEL to A so that the velocity of displacement is always proportional to the ANGULAR VELOCITY of ROTATION.

see also GENERALIZED HELIX, HELICOID, HELIX

References

do Carmo, M. P.; Fischer, G.; Pinkall, U.; and Reckziegel, H. "General Helicoids." §3.4.3 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 36–37, 1986.

Fischer, G. (Ed.). Plate 89 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 85, 1986.

Kreyszig, E. *Differential Geometry.* New York: Dover, p. 88, 1991.

Generalized Helix

The GEODESICS on a general cylinder generated by lines PARALLEL to a line l with which the TANGENT makes a constant ANGLE.

see also HELIX

Generalized Hyperbolic Functions

In 1757, V. Riccati first recorded the generalizations of the HYPERBOLIC FUNCTIONS defined by

$$F^{\alpha}_{n,r}(x) \equiv C\sum_{k=0}^{\infty}\frac{\alpha^k}{(nk+r)!}x^{nk+r}, \tag{1}$$

for $r = 0, \ldots, n-1$, where α is COMPLEX, and where the normalization is taken so that

$$F^{\alpha}_{n,0}(0) = 1. \tag{2}$$

This is called the α-hyperbolic function of order n of the kth kind. The functions $F^{\alpha}_{n,r}$ satisfy

$$F^{\alpha}_{n,r}(x) = (\sqrt[n]{\alpha})^{-r}(\sqrt[n]{\alpha}\,x) \tag{3}$$

and

$$f^{(k)}(x) = \alpha f(x), \tag{4}$$

where

$$f^{(k)}(0) = \begin{cases} 0 & k \neq r,\ 0 \leq k \leq n-1, \\ 1 & k = r. \end{cases} \tag{5}$$

In addition,

$$\frac{d}{dx}F^{\alpha}_{n,r}(x) = \begin{cases} F^{\alpha}_{n,r-1}(x) & \text{for } 0 < r \leq n-1 \\ \alpha F^{\alpha}_{n,n-1}(x) & \text{for } r = 0. \end{cases} \tag{6}$$

The functions give a generalized EULER FORMULA

$$e^{\sqrt[n]{\alpha}} = \sum_{r=0}^{n-1}(\sqrt[n]{\alpha})^r F^{\alpha}_{n,r}(x). \tag{7}$$

Since there are n nth roots of α, this gives a system of n linear equations. Solving for $F^{\alpha}_{n,r}$ gives

$$F^{\alpha}_{n,r}(x) = \frac{1}{n}(\sqrt[n]{\alpha})^{-r}\sum_{k=0}^{n-1}\omega_n{}^{-rk}\exp(\omega_n{}^k\sqrt[n]{\alpha}\,x), \tag{8}$$

where

$$\omega_n = \exp\left(\frac{2\pi i}{n}\right) \tag{9}$$

is a PRIMITIVE ROOT OF UNITY.

The LAPLACE TRANSFORM is

$$\int_0^{\infty} e^{-st}F^{\alpha}_{n,r}(at)\,dt = \frac{s^{n-r-1}a^r}{s^n + \alpha a_n}. \tag{10}$$

The generalized hyperbolic function is also related to the Mittag-Leffler Function $E_\gamma(x)$ by

$$F_{n,0}^1(x) = E_n(x^n). \tag{11}$$

The values $n = 1$ and $n = 2$ give the exponential and circular/hyperbolic functions (depending on the sign of α), respectively.

$$F_{1,0}^\alpha(x) = e^{\alpha x} \tag{12}$$

$$F_{2,0}^\alpha(x) = \cosh(\sqrt{\alpha}\,x) \tag{13}$$

$$F_{2,1}^\alpha(x) = \frac{\sinh(\sqrt{\alpha}\,x)}{\sqrt{\alpha}}. \tag{14}$$

For $\alpha = 1$, the first few functions are

$$\begin{aligned}
F_{1,0}^1(x) &= e^x \\
F_{2,0}^1(x) &= \cosh x \\
F_{2,1}^1(x) &= \sinh x \\
F_{3,0}^1(x) &= \tfrac{1}{3}[e^x + 2e^{-x/2}\cos(\tfrac{1}{2}\sqrt{3}\,x)] \\
F_{3,1}^1(x) &= \tfrac{1}{3}[e^x + 2e^{-x/2}\cos(\tfrac{1}{2}\sqrt{3}\,x + \tfrac{1}{3}\pi)] \\
F_{3,2}^1(x) &= \tfrac{1}{3}[e^x + 2e^{-x/2}\cos(\tfrac{1}{2}\sqrt{3}\,x - \tfrac{1}{3}\pi)] \\
F_{4,0}^1(x) &= \tfrac{1}{2}(\cosh x + \cos x) \\
F_{4,1}^1(x) &= \tfrac{1}{2}(\sinh x + \sin x) \\
F_{4,2}^1(x) &= \tfrac{1}{2}(\cosh x - \cos x) \\
F_{4,3}^1(x) &= \tfrac{1}{2}(\sinh x - \sin x).
\end{aligned}$$

see also Hyperbolic Functions, Mittag-Leffler Function

References

Kaufman, H. "A Biographical Note on the Higher Sine Functions." *Scripta Math.* **28**, 29–36, 1967.

Muldoon, M. E. and Ungar, A. A. "Beyond Sin and Cos." *Math. Mag.* **69**, 3–14, 1996.

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B*. Wellesley, MA: A. K. Peters, 1996.

Ungar, A. "Generalized Hyperbolic Functions." *Amer. Math. Monthly* **89**, 688–691, 1982.

Ungar, A. "Higher Order Alpha-Hyperbolic Functions." *Indian J. Pure. Appl. Math.* **15**, 301–304, 1984.

Generalized Hypergeometric Function

The generalized hypergeometric function is given by a Hypergeometric Series, i.e., a series for which the ratio of successive terms can be written

$$\frac{a_{k+1}}{a_k} = \frac{P(k)}{Q(k)} = \frac{(k+a_1)(k+a_2)\cdots(k+a_p)}{(k+b_1)(k+b_2)\cdots(k+b_q)(k+1)}x. \tag{1}$$

(The factor of $k+1$ in the Denominator is present for historical reasons of notation.) The resulting generalized hypergeometric function is written

$$\sum_{k=0} a_k x^k = {}_pF_q\left[\begin{matrix} a_1, a_2, \cdots, a_p \\ b_1, b_2, \cdots, b_q \end{matrix}; x\right] \tag{2}$$

$$= \sum_{k=0}^{\infty} \frac{(a_1)_k (a_2)_k \cdots (a_p)_k}{(b_1)_k b(b_2)_k \cdots (b_q)_k} \frac{x^k}{k!}, \tag{3}$$

where $(a)_k$ is the Pochhammer Symbol or Rising Factorial

$$(a)_k \equiv \frac{\Gamma(a+k)}{\Gamma(a)} = a(a+1)\cdots(a+k-1). \tag{4}$$

If the argument $x = 1$, then the function is abbreviated

$${}_pF_q\left[\begin{matrix} a_1, a_2 \ldots, a_p \\ b_1, b_2, \ldots, b_q \end{matrix}\right] \equiv {}_pF_q\left[\begin{matrix} a_1, a_2 \ldots, a_p \\ b_1, b_2, \ldots, b_q \end{matrix}; x\right]. \tag{5}$$

${}_2F_1(a, b; c; z)$ is "the" Hypergeometric Function, and ${}_1F_1(a; b; z) \equiv M(z)$ is the Confluent Hypergeometric Function. A function of the form ${}_0F_1(; b; z)$ is called a Confluent Hypergeometric Limit Function.

The generalized hypergeometric function

$${}_{p+1}F_p\left[\begin{matrix} a_1, a_2, \ldots, a_{p+1} \\ b_1, b_2, \ldots, b_p \end{matrix}; z\right] \tag{6}$$

is a solution to the Differential Equation

$$\begin{aligned}
&[\vartheta(\vartheta + b - 1)\cdots(\vartheta + b_p - 1) \\
&\qquad -z(\vartheta + a_1)(\vartheta + a_2)\cdots(\vartheta + a_{p+1})]y = 0,
\end{aligned} \tag{7}$$

where

$$\vartheta = z\frac{d}{dz}. \tag{8}$$

The other linearly independent solution is

$$z^{1-b_1}\,{}_{p+1}F_p\left[\begin{matrix} 1 + a_1 - b_1, 1 - a_2 - b_2, \\ \ldots, 1 + a_{p+1} - b_1 \\ 2 - b_1, 1 - b_2 - b_1, \ldots, \\ 1 - b_p - b_1 \end{matrix}; z\right]. \tag{9}$$

A generalized hypergeometric equation is termed "well posed" if

$$1 + a_1 = b_1 + a_2 = \ldots = b_p + a_{p+1}. \tag{10}$$

Many sums can be written as generalized hypergeometric functions by inspection of the ratios of consecutive terms in the generating Hypergeometric Series. For example, for

$$f(n) \equiv \sum_k (-1)^k \binom{2n}{k}^2, \tag{11}$$

the ratio of successive terms is

$$\frac{a_{k+1}}{a_k} = \frac{(-1)^{k+1}\binom{2n}{k+1}^2}{(-1)^k\binom{2n}{k}^2} = -\frac{(k-2n)^2}{(k+1)^2}, \tag{12}$$

yielding

$$f(n) = {}_2F_1\left[\begin{matrix}-2n,-2n\\1\end{matrix};-1\right] = {}_2F_1(-2n,-2n;1;-1) \tag{13}$$

(Petkovšek 1996, pp. 44–45).

Gosper (1978) discovered a slew unusual hypergeometric function identities, many of which were subsequently proven by Gessel and Stanton (1982). An important generalization of Gosper's technique, called ZEILBERGER'S ALGORITHM, in turn led to the powerful machinery of the WILF-ZEILBERGER PAIR (Zeilberger 1990).

Special hypergeometric identities include GAUSS'S HYPERGEOMETRIC THEOREM

$${}_2F_1(a,b;c;1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} \tag{14}$$

for $\Re[c-a-b] > 0$, KUMMER'S FORMULA

$${}_2F_1(a,b;c;-1) = \frac{\Gamma(\frac{1}{2}b+1)\Gamma(b-a+1)}{\Gamma(b+1)\Gamma(\frac{1}{2}b-a+1)}, \tag{15}$$

where $a-b+c=1$ and b is a positive integer, SAALSCHÜTZ'S THEOREM

$${}_3F_2(a,b,c;d,e;1) = \frac{(d-a)_{|c|}(d-b)_{|c|}}{d_{|c|}(d-a-b)_{|c|}} \tag{16}$$

for $d+e=a+b+c+1$ with c a negative integer and $(a)_n$ the POCHHAMMER SYMBOL, DIXON'S THEOREM

$${}_3F_2(a,b,c;d,e;1) = \frac{(\frac{1}{2}a)!(a-b)!(a-c)!(\frac{1}{2}a-b-c)!}{a!(\frac{1}{2}a-b)!(\frac{1}{2}a-c)!(a-b-c)!}, \tag{17}$$

where $1+a/2-b-c$ has a positive REAL PART, $d = a-b+1$, and $e = a-c+1$, the CLAUSEN FORMULA

$${}_4F_3\left[\begin{matrix}a & b & c & d\\ e & f & g & \end{matrix};1\right] = \frac{(2a)_{|d|}(a+b)_{|d|}(2b)_{|d|}}{(2a+2b)_{|d|}a_{|d|}b_{|d|}}, \tag{18}$$

for $a+b+c-d=1/2$, $e=a+b+1/2$, $a+f=d+1=b+g$, d a nonpositive integer, and the DOUGALL-RAMANUJAN IDENTITY

$$\begin{aligned}{}_7F_6&\left[\begin{matrix}a_1,a_2,a_3,a_4,a_5,a_6,a_7\\ b_1,b_2,b_3,b_4,b_5,b_6\end{matrix};1\right]\\ &= \frac{(a_1+1)_n(a_1-a_2-a_3+1)_n}{(a_1-a_2+1)_n(a_1-a_3+1)_n}\\ &\quad\times\frac{(a_1-a_2-a_4+1)_n(a_1-a_3-a_4+1)_n}{(a_1-a_4+1)_n(a_1-a_2-a_3-a_4+1)_n},\end{aligned} \tag{19}$$

where $n = 2a_1+1 = a_2+a_3+a_4+a_5$, $a_6 = 1+a_1/2$, $a_7 = -n$, and $b_i = 1+a_1-a_{i+1}$ for $i = 1, 2, \ldots, 6$. For all these identities, $(a)_n$ is the POCHHAMMER SYMBOL.

Gessel (1994) found a slew of new identities using WILF-ZEILBERGER PAIRS, including the following:

$${}_5F_4\left[\begin{matrix}-a-b,n+1,n+c+1,2n-a-b+1,n+\frac{1}{2}(3-a-b)\\ n-a-b-c+1,n-a-b+1,2n+2,n+\frac{1}{2}(1-a-b)\end{matrix};1\right] = 0 \tag{20}$$

$${}_3F_2\left[\begin{matrix}-3n,\frac{2}{3}-c,3n+2\\ \frac{3}{2},1-3c\end{matrix};\frac{3}{4}\right] = \frac{(c+\frac{2}{3})_n(\frac{1}{3})_n}{(1-c)_n(\frac{4}{3})_n} \tag{21}$$

$${}_3F_2\left[\begin{matrix}-3b,-\frac{3}{2}n,\frac{1}{2}(1-3n)\\ -3n,\frac{2}{3}-b-n\end{matrix};\frac{4}{3}\right] = \frac{(\frac{1}{3}-b)_n}{(\frac{1}{3}+b)_n} \tag{22}$$

$${}_4F_3\left[\begin{matrix}\frac{3}{2}+\frac{1}{5}n,\frac{2}{3},-n,2n+2\\ n+\frac{11}{6},\frac{4}{3},\frac{1}{5}n+\frac{1}{2}\end{matrix};\frac{2}{27}\right] = \frac{(\frac{5}{2})_n(\frac{11}{6})_n}{(\frac{3}{2})_n(\frac{7}{2})_n} \tag{23}$$

(Petkovšek *et al.* 1996, pp. 135–137).

see also CARLSON'S THEOREM, CLAUSEN FORMULA, CONFLUENT HYPERGEOMETRIC FUNCTION, CONFLUENT HYPERGEOMETRIC LIMIT FUNCTION, DIXON'S THEOREM, DOUGALL-RAMANUJAN IDENTITY, DOUGALL'S THEOREM, GOSPER'S ALGORITHM, HEINE HYPERGEOMETRIC SERIES, HYPERGEOMETRIC FUNCTION, HYPERGEOMETRIC IDENTITY, HYPERGEOMETRIC SERIES, JACKSON'S IDENTITY, KUMMER'S THEOREM, RAMANUJAN'S HYPERGEOMETRIC IDENTITY, SAALSCHÜTZ'S THEOREM, SAALSCHÜTZIAN, SISTER CELINE'S METHOD, THOMAE'S THEOREM, WATSON'S THEOREM, WHIPPLE'S TRANSFORMATION, WILF-ZEILBERGER PAIR, ZEILBERGER'S ALGORITHM

References

Bailey, W. N. *Generalised Hypergeometric Series.* Cambridge, England: Cambridge University Press, 1935.

Dwork, B. *Generalized Hypergeometric Functions.* Oxford, England: Clarendon Press, 1990.

Exton, H. *Multiple Hypergeometric Functions and Applications.* New York: Wiley, 1976.

Gessel, I. "Finding Identities with the WZ Method." *Theoret. Comput. Sci.* To appear.

Gessel, I. and Stanton, D. "Strange Evaluations of Hypergeometric Series." *SIAM J. Math. Anal.* **13**, 295–308, 1982.

Gosper, R. W. "Decision Procedures for Indefinite Hypergeometric Summation." *Proc. Nat. Acad. Sci. USA* **75**, 40–42, 1978.

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, 1996.

Saxena, R. K. and Mathai, A. M. *Generalized Hypergeometric Functions with Applications in Statistics and Physical Sciences.* New York: Springer-Verlag, 1973.

Slater, L. J. *Generalized Hypergeometric Functions.* Cambridge, England: Cambridge University Press, 1966.

Zeilberger, D. "A Fast Algorithm for Proving Terminating Hypergeometric Series Identities." *Discrete Math.* **80**, 207–211, 1990.

Generalized Matrix Inverse

see MOORE-PENROSE GENERALIZED MATRIX INVERSE

Generalized Mean

A generalized version of the MEAN

$$m(t) \equiv \left(\frac{1}{n} \sum_{k=1}^{n} {a_k}^t \right)^{1/t} \tag{1}$$

with parameter t which gives the GEOMETRIC MEAN, ARITHMETIC MEAN, and HARMONIC MEAN as special cases:

$$\lim_{t \to 0} m(t) = G \tag{2}$$

$$m(1) = A \tag{3}$$

$$m(-1) = H. \tag{4}$$

see also MEAN

Generalized Remainder Method

An algorithm for computing a UNIT FRACTION.

Generating Function

A POWER SERIES

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

whose COEFFICIENTS give the SEQUENCE $\{a_0, a_1, \ldots\}$. The *Mathematica*® (Wolfram Research, Champaign, IL) function `DiscreteMath'RSolve'PowerSum` gives the generating function of a given expression, and `ExponentialPowerSum` gives the exponential generating function.

Generating functions for the first few powers are

$$\begin{array}{lll}
1: & \frac{x}{1-x} & = x + x^2 + x^3 + \ldots \\
n: & \frac{x}{(1-x)^2} & = x + 2x^2 + 3x^3 + 4x^4 + \ldots \\
n^2: & \frac{x(x+1)}{(1-x)^3} & = x + 4x^2 + 9x^3 + 16x^4 + \ldots \\
n^3: & \frac{x(x^2+4x+1)}{(x-1)^4} & = x + 8x^2 + 27x^3 + \ldots \\
n^4: & \frac{x(x+1)(x^2+10x+1)}{(x-1)^5} & = x + 16x^2 + 81x^3 + \ldots.
\end{array}$$

see also MOMENT-GENERATING FUNCTION, RECURRENCE RELATION

References

Wilf, H. S. *Generatingfunctionology, 2nd ed.* New York: Academic Press, 1990.

Generation

In population studies, the direct offspring of a reference population (roughly) constitutes a single generation. For a CELLULAR AUTOMATON, the fundamental unit of time during which the rules of reproduction are applied once is called a generation.

Generator (Digitadition)

An INTEGER used to generate a DIGITADITION. A number can have more than one generator. If a number has no generator, it is called a SELF NUMBER.

Generator (Group)

An element of a CYCLIC GROUP, the POWERS of which generate the entire GROUP.

References

Arfken, G. "Generators." §4.11 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 261–267, 1985.

Genetic Algorithm

An adaptive ALGORITHM involving search and optimization first used by John Holland. Holland created an electronic organism as a binary string ("chromosome"), and then used genetic and evolutionary principles of fitness-proportionate selection for reproduction (including random crossover and mutation) to search enormous solution spaces efficiently. So-called genetic programming languages apply the same principles, using an expression tree instead of a bit string as the "chromosome."

see also CELLULAR AUTOMATON

Genocchi Number

A number given by the GENERATING FUNCTION

$$\frac{2t}{e^t + 1} = \sum_{n=1}^{\infty} G_n \frac{t^n}{n!}.$$

It satisfies $G_1 = 1$, $G_3 = G_5 = G_7 = \ldots$, and even coefficients are given by

$$\begin{aligned} G_{2n} &= 2(1 - 2^{2n}) B_{2n} \\ &= 2n E_{2n-1}(0), \end{aligned}$$

where B_n is a BERNOULLI NUMBER and $E_n(x)$ is an EULER POLYNOMIAL. The first few Genocchi numbers for n EVEN are −1, 1, −3, 17, −155, 2073, ... (Sloane's A001469).

see also BERNOULLI NUMBER, EULER POLYNOMIAL

References

Comtet, L. *Advanced Combinatorics: The Art of Finite and Infinite Expansions, rev. enl. ed.*ordrecht, Netherlands: Reidel, p. 49, 1974.

Kreweras, G. "An Additive Generation for the Genocchi Numbers and Two of its Enumerative Meanings." *Bull. Inst. Combin. Appl.* **20**, 99–103, 1997.

Kreweras, G. "Sur les permutations comptées par les nombres de Genocchi de 1-ière et 2-ième espèce." *Europ. J. Comb.* **18**, 49–58, 1997.

Sloane, N. J. A. Sequence A001469/M3041 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Gentle Diagonal

see PASCAL'S TRIANGLE

Gentle Giant Group

see MONSTER GROUP

Genus (Curve)

One of the PLÜCKER CHARACTERISTICS, defined by

$$p \equiv \tfrac{1}{2}(n-1)(n-2)-(\delta+\kappa) = \tfrac{1}{2}(m-1)(m-2)-(\tau+\iota),$$

where m is the class, n the order, δ the number of nodes, κ the number of CUSPS, ι the number of stationary tangents (INFLECTION POINTS), and τ the number of BITANGENTS.

see also RIEMANN CURVE THEOREM

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 100, 1959.

Genus (Knot)

The least genus of any SEIFERT SURFACE for a given KNOT. The UNKNOT is the only KNOT with genus 0.

Genus (Surface)

A topologically invariant property of a surface defined as the largest number of nonintersecting simple closed curves that can be drawn on the surface without separating it. Roughly speaking, it is the number of HOLES in a surface.

see also EULER CHARACTERISTIC

Genus Theorem

A DIOPHANTINE EQUATION

$$x^2+y^2=p$$

can be solved for p a PRIME IFF $p \equiv 1 \pmod 4$ or $p=2$. The representation is unique except for changes of sign or rearrangements of x and y.

see also COMPOSITION THEOREM, FERMAT'S THEOREM

Geocentric Latitude

An AUXILIARY LATITUDE given by

$$\phi_g = \tan^{-1}[(1-e^2)\tan\phi].$$

The series expansion is

$$\phi_g = \phi - e_2\sin(2\phi) + \tfrac{1}{2}{e_2}^2\sin(4\phi) + \tfrac{1}{3}{e_2}^3\sin(6\phi) + \ldots,$$

where

$$e_2 \equiv \frac{e^2}{2-e^2}.$$

see also LATITUDE

References

Adams, O. S. "Latitude Developments Connected with Geodesy and Cartography with Tables, Including a Table for Lambert Equal-Area Meridional Projections." Spec. Pub. No. 67. U. S. Coast and Geodetic Survey, 1921.

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 17–18, 1987.

Geodesic

Given two points on a surface, the geodesic is defined as the shortest path on the surface connecting them. Geodesics have many interesting properties. The NORMAL VECTOR to any point of a GEODESIC arc lies along the normal to a surface at that point (Weinstock 1974, p. 65).

Furthermore, no matter how badly a SPHERE is distorted, there exist an infinite number of closed geodesics on it. This general result, demonstrated in the early 1990s, extended earlier work by Birkhoff, who proved in 1917 that there exists at least one closed geodesic on a distorted sphere, and Lyusternik and Schirelmann, who proved in 1923 that there exist at least three closed geodesics on such a sphere (Cipra 1993).

For a surface $g(x,y,z)=0$, the geodesic can be found by minimizing the ARC LENGTH

$$L \equiv \int ds = \int \sqrt{dx^2+dy^2+dz^2}. \tag{1}$$

But

$$dx = \frac{\partial x}{\partial u}du + \frac{\partial x}{\partial v}dv \tag{2}$$

$$dx^2 = \left(\frac{\partial x}{\partial u}\right)^2 du^2 + 2\frac{\partial x}{\partial u}\frac{\partial x}{\partial v}du\,dv + \left(\frac{\partial x}{\partial v}\right)^2 dv^2, \tag{3}$$

and similarly for dy^2 and dz^2. Plugging in,

$$\begin{aligned} L = \int \Bigg\{ & \left[\left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2 + \left(\frac{\partial z}{\partial u}\right)^2\right] du^2 \\ & + 2\left[\frac{\partial x}{\partial u}\frac{\partial x}{\partial v} + \frac{\partial y}{\partial u}\frac{\partial y}{\partial v} + \frac{\partial z}{\partial u}\frac{\partial z}{\partial v}\right] du\,dv \\ & + \left[\left(\frac{\partial x}{\partial v}\right)^2 + \left(\frac{\partial y}{\partial v}\right)^2 + \left(\frac{\partial z}{\partial v}\right)^2\right] dv^2 \Bigg\}^{1/2}. \end{aligned} \tag{4}$$

This can be rewritten as

$$L = \int \sqrt{P + 2Qv' + Rv'^2}\,du \tag{5}$$

$$= \int \sqrt{Pu'^2 + 2Qu' + R}\,dv, \tag{6}$$

where

$$v' \equiv \frac{dv}{du} \tag{7}$$

$$u' \equiv \frac{du}{dv} \tag{8}$$

and

$$P \equiv \left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2 + \left(\frac{\partial z}{\partial u}\right)^2 \tag{9}$$

$$Q \equiv \frac{\partial x}{\partial u}\frac{\partial x}{\partial v} + \frac{\partial y}{\partial u}\frac{\partial y}{\partial v} + \frac{\partial z}{\partial u}\frac{\partial z}{\partial v} \tag{10}$$

$$R \equiv \left(\frac{\partial x}{\partial v}\right)^2 + \left(\frac{\partial y}{\partial v}\right)^2 + \left(\frac{\partial z}{\partial v}\right)^2. \tag{11}$$

Taking derivatives,

$$\frac{\partial L}{\partial v} = \tfrac{1}{2}(P + 2Qv' + Rv'^2)^{-1/2} \times \left(\frac{\partial P}{\partial v} + 2\frac{\partial Q}{\partial v}v' + \frac{\partial R}{\partial v}v'^2\right) \qquad (12)$$

$$\frac{\partial L}{\partial v'} = \tfrac{1}{2}(P + 2Qv' + Rv'^2)^{-1/2}(2Q + 2Rv'), \qquad (13)$$

so the Euler-Lagrange Differential Equation then gives

$$\frac{\frac{\partial P}{\partial v} + 2v'\frac{\partial Q}{\partial v} + v'^2\frac{\partial R}{\partial v}}{2\sqrt{P + 2Qv' + Rv'^2}} - \frac{d}{du}\left(\frac{Q + Rv'}{\sqrt{P + 2Qv' + Rv'^2}}\right) = 0. \qquad (14)$$

In the special case when P, Q, and R are explicit functions of u only,

$$\frac{Q + Rv'}{\sqrt{P + 2Qv' + Rv'^2}} = c_1 \qquad (15)$$

$$\frac{Q^2 + 2QRv' + R^2v'^2}{P + 2Qv' + Rv'^2} = {c_1}^2 \qquad (16)$$

$$v'^2R(R - {c_1}^2) + 2v'Q(R - {c_1}^2) + (Q^2 - P{c_1}^2) = 0 \qquad (17)$$

$$v' = \frac{1}{2R(R - {c_1}^2)}[2Q({c_1}^2 - R) \pm\sqrt{4Q^2(R - {c_1}^2)^2 - 4R(R - {c_1}^2)(Q^2 - P{c_1}^2)}\,]. \qquad (18)$$

Now, if P and R are explicit functions of u only *and* $Q = 0$,

$$v' = \frac{\sqrt{4R(R - {c_1}^2)P{c_1}^2}}{2R(R - {c_1}^2)} = c_1\sqrt{\frac{P}{R(R - {c_1}^2)}}, \qquad (19)$$

so

$$v = c_1 \int \sqrt{\frac{P}{R(R - {c_1}^2)}}\,du. \qquad (20)$$

In the case $Q = 0$ where P and R are explicit functions of v only, then

$$\frac{\frac{\partial P}{\partial v} + v'^2\frac{\partial R}{\partial v}}{2\sqrt{P + Rv'^2}} - \frac{d}{du}\left(\frac{Rv'}{\sqrt{P + Rv'^2}}\right) = 0, \qquad (21)$$

so

$$\frac{\partial P}{\partial v} + v'^2\frac{\partial R}{\partial v} - 2\sqrt{P + Rv'^2}R \times \left[\frac{v''}{\sqrt{P + Rv'^2}} + (-\tfrac{1}{2})\frac{v'(2Rv'v'')}{(P + Rv'^2)^{3/2}}\right] = 0 \qquad (22)$$

$$\frac{\partial P}{\partial v} + v'^2\frac{\partial R}{\partial v} - 2Rv'' + \frac{2R^2v'^2v''}{P + Rv'^2} = 0 \qquad (23)$$

$$\frac{Rv'^2}{\sqrt{P + Rv'^2}} - \sqrt{P + Rv'^2} = c_1 \qquad (24)$$

$$Rv'^2 - (P + Rv'^2) = c_1\sqrt{P + Rv'^2} \qquad (25)$$

$$\left(-\frac{P}{c_1}\right)^2 = P + Rv'^2 \qquad (26)$$

$$\frac{P^2 - {c_1}^2P}{R{c_1}^2} = v'^2, \qquad (27)$$

and

$$u = c_1 \int \sqrt{\frac{R}{P^2 - {c_1}^2P}}\,dv. \qquad (28)$$

For a surface of revolution in which $y = g(x)$ is rotated about the x-axis so that the equation of the surface is

$$y^2 + z^2 = g^2(x), \qquad (29)$$

the surface can be parameterized by

$$x = u \qquad (30)$$

$$y = g(u)\cos v \qquad (31)$$

$$z = g(u)\sin v. \qquad (32)$$

The equation of the geodesics is then

$$v = c_1 \int \frac{\sqrt{1 + [g'(u)]^2}\,du}{g(u)\sqrt{[g(u)]^2 - {c_1}^2}}. \qquad (33)$$

see also Ellipsoid Geodesic, Geodesic Curvature, Geodesic Dome, Geodesic Equation, Geodesic Triangle, Great Circle, Harmonic Map, Oblate Spheroid Geodesic, Paraboloid Geodesic

References

Cipra, B. *What's Happening in the Mathematical Sciences, Vol. 1.* Providence, RI: Amer. Math. Soc., pp. 27, 1993.

Weinstock, R. *Calculus of Variations, with Applications to Physics and Engineering.* New York: Dover, pp. 26–28 and 45–46, 1974.

Geodesic Curvature

For a unit speed curve on a surface, the length of the surface-tangential component of acceleration is the geodesic curvature κ_g. Curves with $\kappa_g = 0$ are called Geodesics. For a curve parameterized as $\boldsymbol{\alpha}(t) = \mathbf{x}(u(t), v(t))$, the geodesic curvature is given by

$$\kappa_g = \sqrt{EG - F^2}[-\Gamma^2_{11}u'^3 + \Gamma^1_{22}v'^3 - (2\Gamma^2_{12} - \Gamma^1_{11})u'^2v' + (2\Gamma^1_{12} - \Gamma^2_{22})u'v'^2 + u''v' - v''u'],$$

where E, F, and G are coefficients of the first Fundamental Form and Γ^k_{ij} are Christoffel Symbols of the Second Kind.

see also Geodesic

References

Gray, A. "Geodesic Curvature." §20.5 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 402–407, 1993.

Geodesic Dome

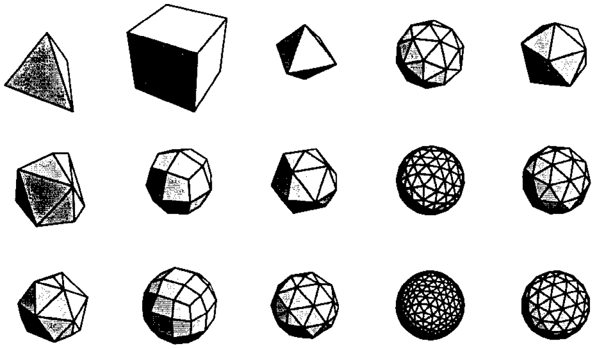

A TRIANGULATION of a PLATONIC SOLID or other POLYHEDRON to produce a close approximation to a SPHERE. The nth order geodesation operation replaces each polygon of the polyhedron by the projection onto the CIRCUMSPHERE of the order n regular tessellation of that polygon. The above figure shows geodesations of orders 1 to 3 (from top to bottom) of the TETRAHEDRON, CUBE, OCTAHEDRON, DODECAHEDRON, and ICOSAHEDRON (from left to right).

R. Buckminster Fuller designed the first geodesic dome (i.e., geodesation of a HEMISPHERE). Fuller's dome was constructed from an ICOSAHEDRON by adding ISOSCELES TRIANGLES about each VERTEX and slightly repositioning the VERTICES. In such domes, neither the VERTICES nor the centers of faces necessarily lie at exactly the same distances from the center. However, these conditions are approximately satisfied.

In the geodesic domes discussed by Kniffen (1994), the sum of VERTEX angles is chosen to be a constant. Given a PLATONIC SOLID, let $e' \equiv 2e/v$ be the number of EDGES meeting at a VERTEX and n be the number of EDGES of the constituent POLYGON. Call the angle of the old VERTEX point A and the angle of the new VERTEX point F. Then

$$A = B \qquad (1)$$

$$2e'A = nF \qquad (2)$$

$$2A + F = 180°. \qquad (3)$$

Solving for A gives

$$2A + \frac{2e'}{n}A = 2A\left(1 + \frac{e'}{n}\right) = 180° \qquad (4)$$

$$A = 90°\frac{n}{e'+n}, \qquad (5)$$

and

$$F = \frac{2e'}{n}A = 180°\frac{e'}{e'+n}. \qquad (6)$$

The VERTEX sum is

$$\Sigma = nF = 180°\frac{e'n}{e'+n}. \qquad (7)$$

Solid	f	v	e'	n	A	F	Σ
tetrahedron			3	3	45°	90°	270°
cube	24	14	3	4	$51\frac{3}{7}$ °	$81\frac{3}{7}$ °	$308\frac{4}{7}$ °
octahedron			4	3	$38\frac{4}{7}$ °	$108\frac{4}{7}$ °	$308\frac{4}{7}$ °
dodecahedron	60	32	3	5	$56\frac{1}{4}$ °	$71\frac{1}{4}$ °	$337\frac{1}{2}$ °
icosahedron			5	3	$33\frac{3}{4}$ °	$118\frac{3}{4}$ °	$337\frac{1}{2}$ °

see also TRIANGULAR SYMMETRY GROUP

References

Kenner, H. *Geodesic Math and How to Use It.* Berkeley, CA: University of California Press, 1976.

Kniffen, D. "Geodesic Domes for Amateur Astronomers." *Sky and Telescope,* pp. 90–94, Oct. 1994.

Pappas, T. "Geodesic Dome of Leonardo da Vinci." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 81, 1989.

Geodesic Equation

$$d\tau^2 = -\eta_{\alpha\beta}\, d\xi^\alpha\, d\xi^\beta,$$

or

$$\frac{d^2\xi^\alpha}{d\tau^2} = 0.$$

see also GEODESIC

Geodesic Flow

A type of FLOW technically defined in terms of the TANGENT BUNDLE of a MANIFOLD.

see also DYNAMICAL SYSTEM

Geodesic Triangle

A TRIANGLE formed by the arcs of three GEODESICS on a smooth surface.

see also INTEGRAL CURVATURE

Geodetic Latitude

see LATITUDE

Geographic Latitude

see LATITUDE

Geometric Construction

In antiquity, geometric constructions of figures and lengths were restricted to use of only a STRAIGHTEDGE and COMPASS. Although the term "RULER" is sometimes used instead of "STRAIGHTEDGE," no markings which could be used to make measurements were allowed according to the Greek prescription. Furthermore, the "COMPASS" could not even be used to mark off distances by setting it and then "walking" it along, so the COMPASS had to be considered to automatically collapse when not in the process of drawing a CIRCLE.

Because of the prominent place Greek geometric constructions held in Euclid's *Elements*, these constructions are sometimes also known as Euclidean Constructions. Such constructions lay at the heart of the Geometric Problems of Antiquity of Circle Squaring, Cube Duplication, and Trisection of an Angle. The Greeks were unable to solve these problems, but it was not until hundreds of years later that the problems were proved to be actually impossible under the limitations imposed.

Simple algebraic operations such as $a+b$, $a-b$, ra (for r a Rational Number), a/b, ab, and $\sqrt{x}$ can be performed using geometric constructions (Courant and Robbins 1996). Other more complicated constructions, such as the solution of Apollonius' Problem and the construction of Inverse Points can also accomplished.

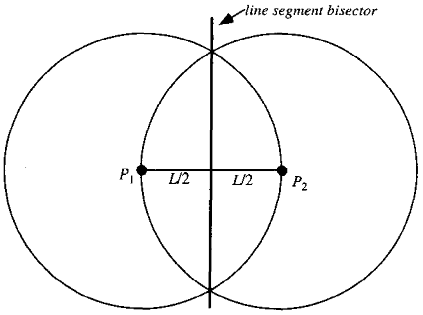

One of the simplest geometric constructions is the construction of a Bisector of a Line Segment, illustrated above.

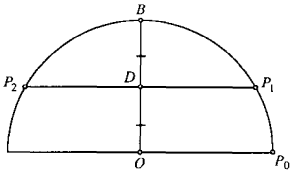

Equilateral Triangle

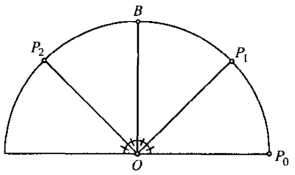

Square

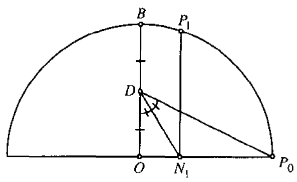

Pentagon

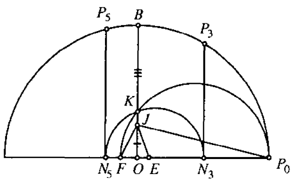

17-gon

The Greeks were very adept at constructing Polygons, but it took the genius of Gauss to mathematically determine which constructions were possible and which were not. As a result, Gauss determined that a series of Polygons (the smallest of which has 17 sides; the Heptadecagon) had constructions unknown to the Greeks. Gauss showed that the Constructible Polygons (several of which are illustrated above) were closely related to numbers called the Fermat Primes.

Wernick (1982) gave a list of 139 sets of three located points from which a Triangle was to be constructed. Of Wernick's original list of 139 problems, 20 had not yet been solved as of 1996 (Meyers 1996).

It is possible to construct Rational Numbers and Euclidean Numbers using a Straightedge and Compass construction. In general, the term for a number which can be constructed using a Compass and Straightedge is a Constructible Number. Some Irrational Numbers, but *no* Transcendental Numbers, can be constructed.

It turns out that all constructions possible with a Compass and Straightedge can be done with a Compass alone, as long as a line is considered constructed when its two endpoints are located. The reverse is also true, since Jacob Steiner showed that all constructions possible with Straightedge and Compass can be done using only a straightedge, as long as a fixed Circle and its center (or two intersecting Circles without their centers, or three nonintersecting Circles) have been drawn beforehand. Such a construction is known as a Steiner Construction.

Geometrography is a quantitative measure of the simplicity of a geometric construction. It reduces geometric constructions to five types of operations, and seeks to reduce the total number of operations (called the "Simplicity") needed to effect a geometric construction.

Dixon (1991, pp. 34–51) gives approximate constructions for some figures (the Heptagon and Nonagon) and lengths (Pi) which cannot be rigorously constructed. Ramanujan (1913–14) and Olds (1963) give geometric constructions for $355/113 \approx \pi$. Gardner (1966, pp. 92–93) gives a geometric construction for

$$3+\tfrac{16}{113}=3.1415929\ldots\approx\pi.$$

Constructions for π are approximate (but inexact) forms of Circle Squaring.

see also Circle Squaring, Compass, Constructible Number, Constructible Polygon, Cube Duplication, Elements, Fermat Prime, Geometric Problems of Antiquity, Geometrography, Mascheroni Construction, Napoleon's Problem, Neusis Construction, Plane Geometry, Polygon, Poncelet-Steiner Theorem, Rectification, Simplicity, Steiner Construction, Straightedge, Trisection

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 96–97, 1987.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 191–202, 1996.

Courant, R. and Robbins, H. "Geometric Constructions. The Algebra of Number Fields." Ch. 3 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 117–164, 1996.

Dantzig, T. *Number, The Language of Science.* New York: Macmillan, p. 316, 1954.

Dixon, R. *Mathographics.* New York: Dover, 1991.

Eppstein, D. "Geometric Models." `http://www.ics.uci.edu/~eppstein/junkyard/model.html`.
Gardner, M. "The Transcendental Number Pi." Ch. 8 in *Martin Gardner's New Mathematical Diversions from Scientific American.* New York: Simon and Schuster, 1966.
Gardner, M. "Mascheroni Constructions." Ch. 17 in *Mathematical Circus: More Puzzles, Games, Paradoxes and Other Mathematical Entertainments from Scientific American.* New York: Knopf, pp. 216–231, 1979.
Herterich, K. *Die Konstruktion von Dreiecken.* Stuttgart: Ernst Klett Verlag, 1986.
Krötenheerdt, O. "Zur Theorie der Dreieckskonstruktionen." *Wissenschaftliche Zeitschrift der Martin-Luther-Univ. Halle-Wittenberg, Math. Naturw. Reihe* **15**, 677–700, 1966.
Meyers, L. F. "Update on William Wernick's 'Triangle Constructions with Three Located Points.'" *Math. Mag.* **69**, 46–49, 1996.
Olds, C. D. *Continued Fractions.* New York: Random House, pp. 59–60, 1963.
Petersen, J. "Methods and Theories for the Solution of Problems of Geometrical Constructions." Reprinted in *String Figures and Other Monographs.* New York: Chelsea, 1960.
Plouffe, S.. "The Computation of Certain Numbers Using a Ruler and Compass." Dec. 12, 1997. `http://www.research.att.com/~njas/sequences/JIS/compass.html`.
Posamentier, A. S. and Wernick, W. *Advanced Geometric Constructions.* Palo Alto, CA: Dale Seymour Pub., 1988.
Ramanujan, S. "Modular Equations and Approximations to π." *Quart. J. Pure. Appl. Math.* **45**, 350–372, 1913–1914.
Wernick, W. "Triangle Constructions with Three Located Points." *Math. Mag.* **55**, 227–230, 1982.

Geometric Distribution

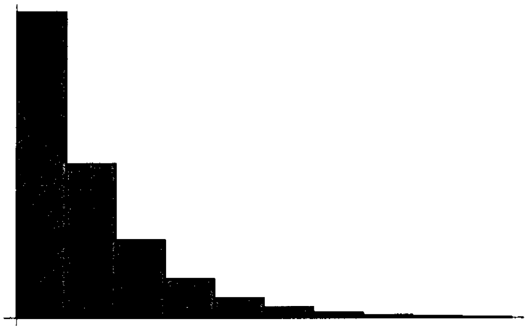

A distribution such that

$$P(n) = q^{n-1}p = p(1-p)^{n-1}, \tag{1}$$

where $q \equiv 1-p$ and for $n = 1, 2, \ldots$. The distribution is normalized since

$$\sum_{n=1}^{\infty} P(n) = \sum_{n=1}^{\infty} q^{n-1}p = p\sum_{n=0}^{\infty} q^n = \frac{p}{1-q} = \frac{p}{p} = 1. \tag{2}$$

The MOMENT-GENERATING FUNCTION is

$$\phi(t) = p[1-(1-p)e^{it}]^{-1}, \tag{3}$$

or

$$\begin{aligned} M(t) = \langle e^{tn} \rangle &= \sum_{n=1}^{\infty} e^{tn}pq^{n-1} = p\sum_{n=0}^{\infty} e^{t(n+1)}q^n \\ &= pe^t \sum_{n=0}^{\infty} (e^t q)^n = \frac{pe^t}{1-e^t q} \end{aligned} \tag{4}$$

$$\begin{aligned} M'(t) &= p\left[\frac{(1-e^t q)e^t - e^t(-e^t q)}{(1-e^t q)^2}\right] \\ &= \frac{p(e^t - qe^{2t} + qe^{2t})}{(1-e^t q)^2} = \frac{pe^t}{(1-e^t q)^2} \end{aligned} \tag{5}$$

$$\begin{aligned} M''(t) &= p\frac{(1-e^t q)^2 e^t - e^t 2(1-e^t q)(-e^t q)}{(1-e^t q)^4} \\ &= p\frac{(1-2e^t q + e^{2t}q^2)e^t + 2qe^{2t}(1-e^t q)}{(1-e^t q)^4} \\ &= p\frac{e^t - 2e^{2t}q + e^{3t}q^2 + 2qe^{2t} - 2q^2e^{3t}}{(1-e^t q)^4} \\ &= p\frac{e^t - q^2e^{3t}}{(1-e^t q)^4} = \frac{pe^t(1-q^2e^{2t})}{(1-e^t q)^4} \\ &= \frac{pe^t(1+qe^t)}{(1-e^t q)^3} \end{aligned} \tag{6}$$

$$M'''(t) = \frac{pe^t[1+4e^t(1-p)+e^{2t}(1-p)^2]}{(1-e^t+e^t p)^4}. \tag{7}$$

Therefore,

$$M'(0) = \mu_1' = \mu = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p} \tag{8}$$

$$M''(0) = \mu_2' = \frac{p(1+q)}{(1+q)^3} = \frac{p(2-p)}{p^3} = \frac{2-p}{p^2} \tag{9}$$

$$M'''(0) = \mu_3' = \frac{(6-6p+p^2)}{p^3} \tag{10}$$

$$M^{(4)}(0) = \mu_4' = \frac{(p-2)(-p^2+12p-12)}{p^4}, \tag{11}$$

and

$$\begin{aligned} \mu_2 \equiv \mu_2' - (\mu_1')^2 &= \left(\frac{2}{p^2} - \frac{1}{p}\right) - \frac{1}{p^2} = \frac{1-p}{p^2} \\ &= \frac{q}{p^2} \end{aligned} \tag{12}$$

$$\begin{aligned} \mu_3 &\equiv \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3 \\ &= \frac{6-6p+p^2}{p^3} - 3\frac{2-p}{p^2}\frac{1}{p} + 2\left(\frac{1}{p}\right)^3 \\ &= \frac{6-6p+p^2-3(2-p)+2}{p^3} \\ &= \frac{(p-1)(p-2)}{p^3} \end{aligned} \tag{13}$$

$$\begin{aligned} \mu_4 &\equiv \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4 \\ &= \frac{(p-2)(-p^2+12p-12)}{p^4} - 4\frac{6-6p+p^2}{p^3}\frac{1}{p} \\ &\quad + 6\frac{2-p}{p^2}\left(\frac{1}{p}\right)^2 - 3\left(\frac{1}{p}\right)^4 \\ &= \frac{(p-1)(-p^2+9p-9)}{p^4}, \end{aligned} \tag{14}$$

so the MEAN, VARIANCE, SKEWNESS, and KURTOSIS are given by

$$\mu \equiv \mu_1' = \frac{1}{p} \tag{15}$$

$$\sigma^2 = \mu_2 = \frac{q}{p^2} \tag{16}$$

$$\begin{aligned}\gamma_1 &= \frac{\mu_3}{\sigma^3} = \frac{(p-1)(p-2)}{p^3}\left(\frac{p^2}{1-p}\right)^{3/2} \\ &= \frac{(p-1)(p-2)}{(1-p)\sqrt{1-p}} = \frac{2-p}{\sqrt{1-p}} = \frac{2-p}{\sqrt{q}}\end{aligned} \tag{17}$$

$$\begin{aligned}\gamma_2 &= \frac{\mu_4}{\sigma^4} - 3 = \frac{(p-1)(-p^2+9p-9)}{p^4\frac{(1-p)^2}{p^4}} - 3 \\ &= \frac{-9+9p-p^2}{(p-1)} - 3 \\ &= \frac{p^2-6p+6}{1-p}.\end{aligned} \tag{18}$$

In fact, the moments of the distribution are given analytically in terms of the POLYLOGARITHM function,

$$\mu_k' \equiv \sum_{n=1}^{\infty} P(n)n^k = \sum_{n=1}^{\infty} p(1-p)^{n-1}n^k = \frac{p\,\mathrm{Li}_{-k}(1-p)}{1-p}. \tag{19}$$

For the case $p = 1/2$ (corresponding to the distribution of the number of COIN TOSSES needed to win in the SAINT PETERSBURG PARADOX) this formula immediately gives

$$\mu_1' = 2 \tag{20}$$

$$\mu_2' = 6 \tag{21}$$

$$\mu_3' = 26 \tag{22}$$

$$\mu_4' = 150, \tag{23}$$

so the MEAN, VARIANCE, SKEWNESS, and KURTOSIS in this case are

$$\mu = 2 \tag{24}$$

$$\sigma^2 = 2 \tag{25}$$

$$\gamma_1 = \tfrac{3}{2}\sqrt{2} \tag{26}$$

$$\gamma_2 = \tfrac{13}{2}. \tag{27}$$

The first CUMULANT of the geometric distribution is

$$\kappa_1 = \frac{1-p}{p}, \tag{28}$$

and subsequent CUMULANTS are given by the RECURRENCE RELATION

$$\kappa_{r+1} = (1-p)\frac{d\kappa_r}{dp}. \tag{29}$$

see also SAINT PETERSBURG PARADOX

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 531–532, 1987.

Spiegel, M. R. *Theory and Problems of Probability and Statistics.* New York: McGraw-Hill, p. 118, 1992.

Geometric Mean

$$G \equiv \left(\prod_{i=1}^{n} a_i\right)^{1/n}.$$

Hoehn and Niven (1985) show that

$$G(a_1+c, a_2+c, \ldots, a_n+c) > c + G(a_1, a_2, \ldots, a_n)$$

for any POSITIVE constant c.

see also ARITHMETIC MEAN, ARITHMETIC-GEOMETRIC MEAN, CARLEMAN'S INEQUALITY, HARMONIC MEAN, MEAN, ROOT-MEAN-SQUARE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 10, 1972.

Hoehn, L. and Niven, I. "Averages on the Move." *Math. Mag.* **58**, 151–156, 1985.

Geometric Mean Index

The statistical INDEX

$$P_G \equiv \left[\prod \left(\frac{p_n}{p_0}\right)^{v_0}\right]^{1/\Sigma v_0},$$

where p_n is the price per unit in period n, q_n is the quantity produced in period n, and $v_n \equiv p_n q_n$ the value of the n units.

see also INDEX

References

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, p. 69, 1962.

Geometric Probability Constants

see CUBE POINT PICKING, CUBE TRIANGLE PICKING

Geometric Problems of Antiquity

The Greek problems of antiquity were a set of geometric problems whose solution was sought using only COMPASS and STRAIGHTEDGE:

1. CIRCLE SQUARING.
2. CUBE DUPLICATION.
3. TRISECTION of an ANGLE.

Only in modern times, more than 2,000 years after they were formulated, were all three ancient problems proved insoluble using only COMPASS and STRAIGHTEDGE.

Another ancient geometric problem not proved impossible until 1997 is ALHAZEN'S BILLIARD PROBLEM. As Ogilvy (1990) points out, constructing the general REGULAR POLYHEDRON was really a "fourth" unsolved problem of antiquity.

see also ALHAZEN'S BILLIARD PROBLEM, CIRCLE SQUARING, COMPASS, CONSTRUCTIBLE NUMBER, CONSTRUCTIBLE POLYGON, CUBE DUPLICATION, GEOMETRIC CONSTRUCTION, REGULAR POLYHEDRON, STRAIGHTEDGE, TRISECTION

References

Conway, J. H. and Guy, R. K. "Three Greek Problems." In *The Book of Numbers.* New York: Springer-Verlag, pp. 190–191, 1996.

Courant, R. and Robbins, H. "The Unsolvability of the Three Greek Problems." §3.3 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 117–118 and 134–140, 1996.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 135–138, 1990.

Pappas, T. "The Impossible Trio." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 130–132, 1989.

Jones, A.; Morris, S.; and Pearson, K. *Abstract Algebra and Famous Impossibilities.* New York: Springer-Verlag, 1991.

Geometric Progression

see GEOMETRIC SEQUENCE

Geometric Sequence

A geometric sequence is a SEQUENCE $\{a_k\}$, $k = 1, 2, \ldots$, such that each term is given by a multiple r of the previous one. Another equivalent definition is that a sequence is geometric IFF it has a zero BIAS. If the multiplier is r, then the kth term is given by

$$a_k = ra_{k-1} = r^2 a_{k-2} = a_0 r^k.$$

Without loss of generality, take $a_0 = 1$, giving

$$a_k = r^k.$$

Geometric Series

A geometric series $\sum_k a_k$ is a series for which the ratio of each two consecutive terms a_{k+1}/a_k is a constant function of the summation index k, say r. Then the terms a_k are of the form $a_k = a_0 r^k$, so $a_{k+1}/a_k = r$. If $\{a_k\}$, with $k = 1, 2, \ldots$, is a GEOMETRIC SEQUENCE with multiplier $-1 < r < 1$ and $a_0 = 1$, then the geometric series

$$S_n = \sum_{k=0}^{n} a_k = \sum_{k=0}^{n} r^k \tag{1}$$

is given by

$$S_n \equiv \sum_{k=0}^{n} r^k = 1 + r + r^2 + \ldots + r^n, \tag{2}$$

so

$$rS_n = r + r^2 + r^3 + \ldots + r^{n+1}. \tag{3}$$

Subtracting

$$\begin{aligned}(1-r)S_n &= (1 + r + r^2 + \ldots + r^n) \\ &\quad - (r + r^2 + r^3 + \ldots + r^{n+1}) \\ &= 1 - r^{n+1},\end{aligned} \tag{4}$$

so

$$S_n \equiv \sum_{k=0}^{n} r^k = \frac{1 - r^{n+1}}{1 - r}. \tag{5}$$

As $n \to \infty$, then

$$S \equiv S_\infty = \sum_{k=0}^{\infty} r^k = \frac{1}{1-r}. \tag{6}$$

see also ARITHMETIC SERIES, GABRIEL'S STAIRCASE, HARMONIC SERIES, HYPERGEOMETRIC SERIES, WHEAT AND CHESSBOARD PROBLEM

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 10, 1972.

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 278–279, 1985.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 8, 1987.

Courant, R. and Robbins, H. "The Geometric Progression." §1.2.3 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 13–14, 1996.

Pappas, T. "Perimeter, Area & the Infinite Series." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 134–135, 1989.

Geometrization Conjecture

see THURSTON'S GEOMETRIZATION CONJECTURE

Geometrography

A quantitative measure of the simplicity of a GEOMETRIC CONSTRUCTION which reduces geometric constructions to five steps. It was devised by É. Lemoine.

- S_1 Place a STRAIGHTEDGE'S EDGE through a given POINT,
- S_2 Draw a straight LINE,
- C_1 Place a POINT of a COMPASS on a given POINT,
- C_2 Place a POINT of a COMPASS on an indeterminate POINT on a LINE,
- C_3 Draw a CIRCLE.

Geometrography seeks to reduce the number of operations (called the "SIMPLICITY") needed to effect a construction. If the number of the above operations are denoted m_1, m_2, n_1, n_2, and n_3, respectively, then the SIMPLICITY is $m_1 + m_2 + n_1 + n_2 + n_3$ and the symbol is $m_1S_1 + m_2S_2 + n_1C_1 + n_2C_2 + n_3C_3$. It is apparently an unsolved problem to determine if a given GEOMETRIC CONSTRUCTION is of the smallest possible simplicity.

see also SIMPLICITY

References

De Temple, D. W. "Carlyle Circles and the Lemoine Simplicity of Polygonal Constructions." *Amer. Math. Monthly* **98**, 97–108, 1991.

Eves, H. *An Introduction to the History of Mathematics, 6th ed.* New York: Holt, Rinehart, and Winston, 1990.

Geometry

Geometry is the study of figures in a SPACE of a given number of dimensions and of a given type. The most common types of geometry are PLANE GEOMETRY (dealing with objects like the LINE, CIRCLE, TRIANGLE, and POLYGON), SOLID GEOMETRY (dealing with objects like the LINE, SPHERE, and POLYHEDRON), and SPHERICAL GEOMETRY (dealing with objects like the SPHERICAL TRIANGLE and SPHERICAL POLYGON).

Historically, the study of geometry proceeds from a small number of accepted truths (AXIOMS or POSTULATES), then builds up true statements using a systematic and rigorous step-by-step PROOF. However, there is much more to geometry than this relatively dry textbook approach, as evidenced by some of the beautiful and unexpected results of PROJECTIVE GEOMETRY (not to mention Schubert's powerful but questionable ENUMERATIVE GEOMETRY).

Formally, a geometry is defined as a complete locally homogeneous RIEMANNIAN METRIC. In $\mathbb{R}^2$, the possible geometries are Euclidean planar, hyperbolic planar, and elliptic planar. In $\mathbb{R}^3$, the possible geometries include Euclidean, hyperbolic, and elliptic, but also include five other types.

see also ABSOLUTE GEOMETRY, AFFINE GEOMETRY, COORDINATE GEOMETRY, DIFFERENTIAL GEOMETRY, ENUMERATIVE GEOMETRY, FINSLER GEOMETRY, INVERSIVE GEOMETRY, MINKOWSKI GEOMETRY, NIL GEOMETRY, NON-EUCLIDEAN GEOMETRY, ORDERED GEOMETRY, PLANE GEOMETRY, PROJECTIVE GEOMETRY, SOL GEOMETRY, SOLID GEOMETRY, SPHERICAL GEOMETRY, THURSTON'S GEOMETRIZATION CONJECTURE

References

Altshiller-Court, N. *College Geometry: A Second Course in Plane Geometry for Colleges and Normal Schools, 2nd ed., rev. enl.* New York: Barnes and Noble, 1952.

Brown, K. S. "Geometry." `http://www.seanet.com/~ksbrown/igeometr.htm`.

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, 1969.

Croft, H. T.; Falconer, K. J.; and Guy, R. K. *Unsolved Problems in Geometry.* New York: Springer-Verlag, 1994.

Eppstein, D. "Geometry Junkyard." `http://www.ics.uci.edu/~eppstein/junkyard/`.

Eppstein, D. "Many-Dimensional Geometry." `http://www.ics.uci.edu/~eppstein/junkyard/highdim.html`.

Eppstein, D. "Planar Geometry." `http://www.ics.uci.edu/~eppstein/junkyard/2d.html`.

Eppstein, D. "Three-Dimensional Geometry." `http://www.ics.uci.edu/~eppstein/junkyard/3d.html`.

Eves, H. W. *A Survey of Geometry, rev. ed.* Boston, MA: Allyn and Bacon, 1972.

Geometry Center. `http://www.geom.umn.edu`.

Ghyka, M. C. *The Geometry of Art and Life, 2nd ed.* New York: Dover, 1977.

Hilbert, D. *The Foundations of Geometry, 2nd ed.* Chicago, IL: The Open Court Publishing Co., 1921.

Johnson, R. A. *Advanced Euclidean Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* New York: Dover, 1960.

King, J. and Schattschneider, D. (Eds.). *Geometry Turned On: Dynamic Software in Learning, Teaching and Research.* Washington, DC: Math. Assoc. Amer., 1997.

Klein, F. *Famous Problems of Elementary Geometry and Other Monographs.*082840108X New York: Dover, 1956.

Melzak, Z. A. *Invitation to Geometry.* New York: Wiley, 1983.

Moise, E. E. *Elementary Geometry from an Advanced Standpoint, 3rd ed.* Reading, MA: Addison-Wesley, 1990.

Ogilvy, C. S. "Some Unsolved Problems of Modern Geometry." Ch. 11 in *Excursions in Geometry.* New York: Dover, pp. 143–153, 1990.

Simon, M. *Über die Entwicklung der Elementargeometrie im XIX Jahrhundert.* Berlin, pp. 97–105, 1906.

Woods, F. S. *Higher Geometry: An Introduction to Advanced Methods in Analytic Geometry.* New York: Dover, 1961.

Gergonne Line

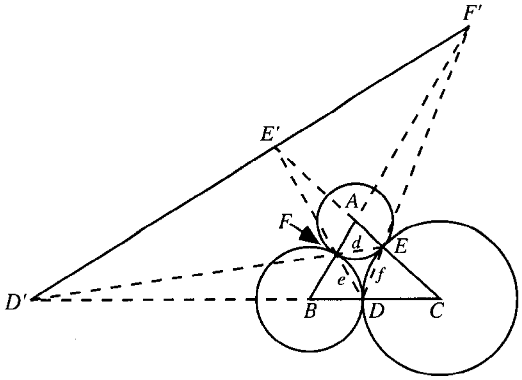

The perspective line for the CONTACT TRIANGLE ΔDEF and its TANGENTIAL TRIANGLE ΔABC. It is determined by the NOBBS POINTS D', E', and F'. In addition to the NOBBS POINTS, the FLETCHER POINT and EVANS POINT also lie on the Gergonne line where it intersects the SODDY LINE and EULER LINE, respectively. The D and D' coordinates are given by

$$D = B + \frac{f}{e}C$$
$$D' = B - \frac{f}{e}C,$$

so $BDCD'$ form a HARMONIC RANGE. The equation of the Gergonne line is

$$\frac{\alpha}{d} + \frac{\beta}{e} + \frac{\gamma}{f} = 0.$$

see also CONTACT TRIANGLE, EULER LINE, EVANS POINT, FLETCHER POINT, NOBBS POINTS, SODDY LINE, TANGENTIAL TRIANGLE

References

Oldknow, A. "The Euler-Gergonne-Soddy Triangle of a Triangle." *Amer. Math. Monthly* **103**, 319–329, 1996.

Gergonne Point

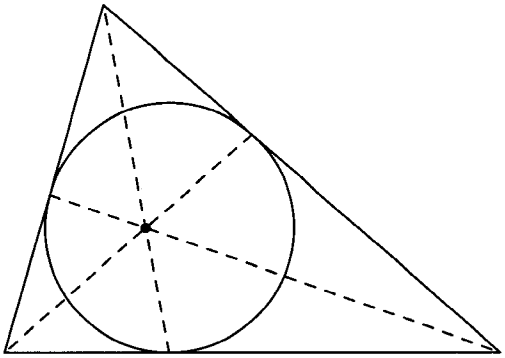

The common point of the CONCURRENT lines from the TANGENT points of a TRIANGLE'S INCIRCLE to the opposite VERTICES. It has TRIANGLE CENTER FUNCTION

$$\alpha = [a(b+c-a)]^{-1} = \tfrac{1}{2}\sec^2 A.$$

It is the ISOTOMIC CONJUGATE POINT of the NAGEL POINT. The CONTACT TRIANGLE and TANGENTIAL TRIANGLE are perspective from the Gergonne point.

see also GERGONNE LINE

References

Altshiller-Court, N. *College Geometry: A Second Course in Plane Geometry for Colleges and Normal Schools, 2nd ed.* New York: Barnes and Noble, pp. 160–164, 1952.

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* New York: Random House, pp. 11–13, 1967.

Eves, H. W. *A Survey of Geometry, rev. ed.* Boston, MA: Allyn and Bacon, p. 83, 1972.

Gallatly, W. *The Modern Geometry of the Triangle, 2nd ed.* London: Hodgson, p. 22, 1913.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 184 and 216, 1929.

Kimberling, C. "Gergonne Point." `http://www.evansville.edu/~ck6/tcenters/class/gergonne.html`.

Germain Primes

see SOPHIE GERMAIN PRIME

Gerono Lemniscate

see EIGHT CURVE

Geršgorin Circle Theorem

Gives a region in the COMPLEX PLANE containing all the EIGENVALUES of a COMPLEX SQUARE MATRIX. Let

$$|x_k| = \max\{|x_i| : 1 \leq i \leq n\} > 0 \tag{1}$$

and define

$$R_i = \sum_{\substack{i=1 \\ j\neq i}}^{n} |a_{ij}|. \tag{2}$$

Then each EIGENVALUE of the MATRIX A of order n is in at least one of the disks

$$\{z : |z - a_{ii}| \leq R_i\}. \tag{3}$$

The theorem can be made stronger as follows. Let r be an INTEGER with $1 \leq r \leq n$, then each EIGENVALUE of A is either in one of the disks Γ_1

$$\{z : |z - a_{jj}| \leq S_j^{(r-1)}\}, \tag{4}$$

or in one of the regions

$$\left\{z : \sum_{i=1}^{r} |z - a_{ii}| \leq \sum_{i=1}^{r} R_i\right\}, \tag{5}$$

where $S_j^{(r-1)}$ is the sum of magnitudes of the $r-1$ largest off-diagonal elements in column j.

References

Buraldi, R. A. and Mellendorf, S. "Regions in the Complex Plane Containing the Eigenvalues of a Matrix." *Amer. Math. Monthly* **101**, 975–985, 1994.

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 1120–1121, 1979.

Taussky-Todd, O. "A Recurring Theorem on Determinants." *Amer. Math. Monthly* **56**, 672–676, 1949.

Ghost

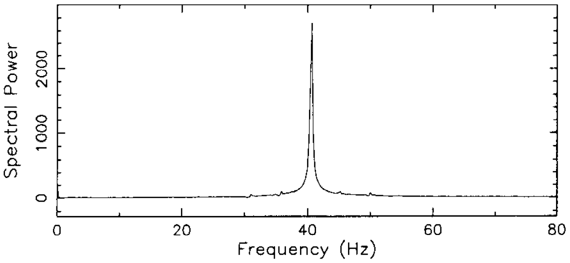

If the sampling of an interferogram is modulated at a definite frequency instead of being uniformly sampled, spurious spectral features called "ghosts" are produced (Brault 1985). Periodic ruling or sampling errors introduce a modulation superposed on top of the expected fringe pattern due to uniform stage translation. Because modulation is a multiplicative process, spurious features are generated in spectral space at the sum and difference of the true fringe and ghost fringe frequencies, thus throwing power out of its spectral band.

Ghosts are copies of the actual spectrum, but appear at reduced strength. The above shows the power spectrum for a pure sinusoidal signal sampled by translating a Fourier transform spectrometer mirror at constant speed. The small blips on either side of the main peaks are ghosts.

In order for a ghost to appear, the process producing it must exist for most of the interferogram. However, if the ruling errors are not truly sinusoidal but vary across the length of the screw, a longer travel path can reduce their effect.

see also JITTER

References

Brault, J. W. "Fourier Transform Spectroscopy." In *High Resolution in Astronomy: 15th Advanced Course of the Swiss Society of Astronomy and Astrophysics* (Ed. A. Benz, M. Huber, and M. Mayor). Geneva Observatory, Sauverny, Switzerland, 1985.

Gibbs Constant

see WILBRAHAM-GIBBS CONSTANT

Gibbs Effect

see GIBBS PHENOMENON

Gibbs Phenomenon

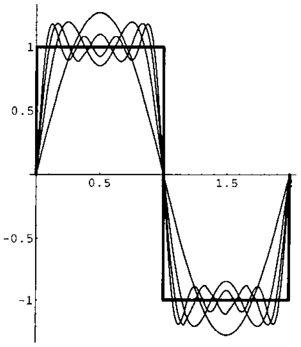

An overshoot of FOURIER SERIES and other EIGENFUNCTION series occurring at simple DISCONTINUITIES. It can be removed with the LANCZOS σ FACTOR.

see also FOURIER SERIES

References

Arfken, G. "Gibbs Phenomenon." §14.5 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 783–787, 1985.

Foster, J. and Richards, F. B. "The Gibbs Phenomenon for Piecewise-Linear Approximation." *Amer. Math. Monthly* **98**, 47–49, 1991.

Gibbs, J. W. "Fourier Series." *Nature* **59**, 200 and 606, 1899.

Hewitt, E. and Hewitt, R. "The Gibbs-Wilbraham Phenomenon: An Episode in Fourier Analysis." *Arch. Hist. Exact Sci.* **21**, 129–160, 1980.

Sansone, G. "Gibbs' Phenomenon." §2.10 in *Orthogonal Functions, rev. English ed.* New York: Dover, pp. 141–148, 1991.

Gigantic Prime

A PRIME with 10,000 or more decimal digits. As of Nov. 15, 1995, 127 were known.

see also TITANIC PRIME

References

Caldwell, C. "The Ten Largest Known Primes." `http://www.utm.edu/research/primes/largest.html#largest`.

Gilbrat's Distribution

A CONTINUOUS DISTRIBUTION in which the LOGARITHM of a variable x has a NORMAL DISTRIBUTION,

$$P(x)=\frac{1}{\sqrt{2\pi}}e^{-(\ln x)^2/2}.$$

It is a special case of the LOG NORMAL DISTRIBUTION

$$P(x)=\frac{1}{S\sqrt{2\pi}}e^{-(\ln x-M)^2/2S^2}.$$

with $S=1$ and $M=0$.

see also LOG NORMAL DISTRIBUTION

Gilbreath's Conjecture

Let the DIFFERENCE of successive PRIMES be defined by $d_n\equiv p_{n+1}-p_n$, and d_n^k by

$$d_n^k\equiv\begin{cases}d_n & \text{for } k=1\\ |d_{n+1}^{k-1}-d_n^{k-1}| & \text{for } k>1.\end{cases}$$

N. L. Gilbreath claimed that $d_1^k=1$ for all k (Guy 1994). It has been verified for $k<63419$ and all PRIMES up to $\pi(10^{13})$, where π is the PRIME COUNTING FUNCTION.

References

Guy, R. K. "Gilbreath's Conjecture." §A10 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 25–26, 1994.

Gill's Method

A formula for numerical solution of differential equations,

$$y_{n+1}=y_n+\tfrac{1}{6}[k_1+(2-\sqrt{2})k_2 +(2+\sqrt{2})k_3+k_4)+\mathcal{O}(h^5),$$

where

$$\begin{aligned}k_1&=hf(x_n,y_n)\\ k_2&=hf(x_n+\tfrac{1}{2}h,y_n+\tfrac{1}{2}k_1)\\ k_3&=hf(x_n+\tfrac{1}{2}h,y_n+\tfrac{1}{2}(-1+\sqrt{2})k_1+(1-\tfrac{1}{2}\sqrt{2})k_2)\\ k_4&=hf(x_n+h,y_n-\tfrac{1}{2}\sqrt{2}\,k_2+(1+\tfrac{1}{2}\sqrt{2})k_3).\end{aligned}$$

see also ADAMS' METHOD, MILNE'S METHOD, PREDICTOR-CORRECTOR METHODS, RUNGE-KUTTA METHOD

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 896, 1972.

Gingerbreadman Map

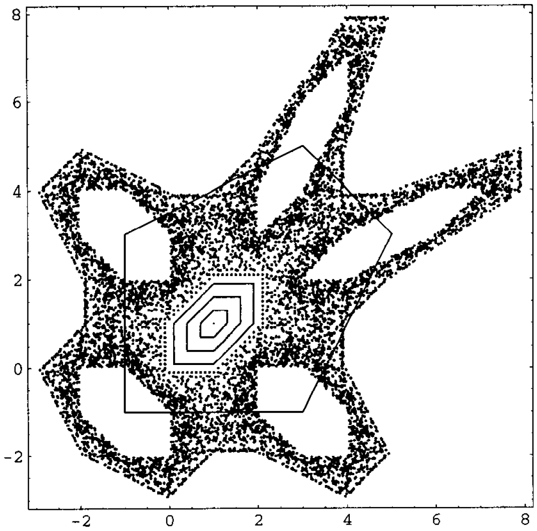

A 2-D piecewise linear MAP defined by

$$x_{n+1} = 1 - y_n + |x_n|$$
$$y_{n+1} = x_n.$$

The map is chaotic in the filled region above and stable in the six hexagonal regions. Each point in the interior hexagon defined by the vertices (0, 0), (1, 0), (2, 1), (2, 2), (1, 2), and (0, 1) has an orbit with period six (except the point (1, 1), which has period 1). Orbits in the other five hexagonal regions circulate from one to the other. There is a unique orbit of period five, with all others having period 30. The points having orbits of period five are (−1, 3), (−1, −1), (3, −1), (5, 3), and (3, 5), indicated in the above figure by the black line. However, there are infinitely many distinct periodic orbits which have an arbitrarily long period.

References

Devaney, R. L. "A Piecewise Linear Model for the Zones of Instability of an Area Preserving Map." *Physica D* **10**, 387–393, 1984.

Peitgen, H.-O. and Saupe, D. (Eds.). "A Chaotic Gingerbreadman." §3.2.3 in *The Science of Fractal Images.* New York: Springer-Verlag, pp. 149–150, 1988.

Girard's Spherical Excess Formula

Let a SPHERICAL TRIANGLE Δ have angles A, B, and C. Then the SPHERICAL EXCESS is given by

$$\Delta = A + B + C - \pi.$$

see also ANGULAR DEFECT, L'HUILIER'S THEOREM, SPHERICAL EXCESS

References

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 94–95, 1969.

Girko's Circular Law

Let λ be EIGENVALUES of a set of RANDOM $n \times n$ MATRICES. Then $\lambda/\sqrt{n}$ is uniformly distributed on the DISK.

References

Girko, V. L. *Theory of Random Determinants.* Boston, MA: Kluwer, 1990.

Girth

The length of the shortest CYCLE in a GRAPH.

Girth	Example
3	tetrahedron
4	cube, $K_{3,3}$
5	Petersen graph

Giuga's Conjecture

If $n > 1$ and

$$n | 1^{n-1} + 2^{n-1} + \ldots + (n-1)^{n-1} + 1,$$

is n necessarily a PRIME? In other words, defining

$$s_n \equiv \sum_{k=1}^{n-1} k^{n-1},$$

does there exist a COMPOSITE n such that $s_n \equiv -1 \pmod{n}$? It is known that $s_n \equiv -1 \pmod{n}$ IFF for each prime divisor p of n, $(p-1)|(n/p-1)$ and $p|(n/p-1)$ (Giuga 1950, Borwein *et al.* 1996); therefore, any counterexample must be SQUAREFREE. A composite INTEGER n satisfies $s_n \equiv -1 \pmod{n}$ IFF it is both a CARMICHAEL NUMBER and a GIUGA NUMBER. Giuga showed that there are no exceptions to the conjecture up to 10^{1000}. This was later improved to 10^{1700} (Bedocchi 1985) and 10^{13800} (Borwein *et al.* 1996).

see also ARGOH'S CONJECTURE

References

Bedocchi, E. "The $\mathbb{Z}(\sqrt{14})$ Ring and the Euclidean Algorithm." *Manuscripta Math.* **53**, 199–216, 1985.

Borwein, D.; Borwein, J. M.; Borwein, P. B.; and Girgensohn, R. "Giuga's Conjecture on Primality." *Amer. Math. Monthly* **103**, 40–50, 1996.

Giuga, G. "Su una presumibile propertietà caratteristica dei numeri primi." *Ist. Lombardo Sci. Lett. Rend. A* **83**, 511–528, 1950.

Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, pp. 20–21, 1989.

Giuga Number

Any COMPOSITE NUMBER n with $p|(n/p-1)$ for all PRIME DIVISORS p of n. n is a Giuga number IFF

$$\sum_{k=1}^{n-1} k^{\phi(n)} \equiv -1 \pmod{n}$$

where ϕ is the TOTIENT FUNCTION and IFF

$$\sum_{p|n} \frac{1}{p} - \prod_{p|n} \frac{1}{p} \in \mathbb{N}.$$

n is a Giuga number IFF

$$nB_{\phi(n)} \equiv -1 \pmod{n},$$

where B_k is a BERNOULLI NUMBER and ϕ is the TOTIENT FUNCTION. Every counterexample to Giuga's conjecture is a contradiction to ARGOH'S CONJECTURE and vice versa. The smallest known Giuga numbers are 30 (3 factors), 858, 1722 (4 factors), 66198 (5 factors), 2214408306, 24423128562 (6 factors), 432749205173838, 14737133470010574, 550843391309130318 (7 factors),

$$24419700098249971508786634 6,\quad 554079914617070801288578559178$$

(8 factors), ... (Sloane's A007850).

It is not known if there are an infinite number of Giuga numbers. All the above numbers have sum minus product equal to 1, and any Giuga number of higher order must have at least 59 factors. The smallest ODD Giuga number must have at least nine PRIME factors.

see also ARGOH'S CONJECTURE, BERNOULLI NUMBER, TOTIENT FUNCTION

References

Borwein, D.; Borwein, J. M.; Borwein, P. B.; and Girgensohn, R. "Giuga's Conjecture on Primality." *Amer. Math. Monthly* **103**, 40–50, 1996.

Sloane, N. J. A. Sequence A007850 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Giuga Sequence

A finite, increasing sequence of INTEGERS $\{n_1, \ldots, n_m\}$ such that

$$\sum_{i=1}^{m} \frac{1}{n_i} - \prod_{i=1}^{m} \frac{1}{n_i} \in \mathbb{N}.$$

A sequence is a Giuga sequence IFF it satisfies

$$n_i | (n_1 \cdots n_{i-1} \cdot n_{i+1} \cdot n_m - 1)$$

for $i = 1, \ldots, m$. There are no Giuga sequences of length 2, one of length 3 ({2, 3, 5}), two of length 4 ({2, 3, 7, 41} and {2, 3, 11, 13}), 3 of length 5 ({2, 3, 7, 43, 1805}, {2, 3, 7, 83, 85}, and {2, 3, 11, 17, 59}), 17 of length 6, 27 of length 7, and hundreds of length 8. There are infinitely many Giuga sequences. It is possible to generate longer Giuga sequences from shorter ones satisfying certain properties.

see also CARMICHAEL SEQUENCE

References

Borwein, D.; Borwein, J. M.; Borwein, P. B.; and Girgensohn, R. "Giuga's Conjecture on Primality." *Amer. Math. Monthly* **103**, 40–50, 1996.

Glaisher-Kinkelin Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Define

$$K(n+1) \equiv 0^0 1^1 2^2 3^3 \cdots n^n \tag{1}$$

$$G(n+1) \equiv \frac{(n!)^n}{K(n+1)} = \begin{cases} 1 & \text{if } n = 0 \\ 0!1!2!\cdots(n-1)! & \text{if } n > 0. \end{cases} \tag{2}$$

where G is the G-FUNCTION and K is the K-FUNCTION. Then

$$\lim_{n\to\infty} \frac{K(n+1)}{n^{n^2/2+n/2+1/2} e^{-n^2/4}} = A \tag{3}$$

$$\lim_{n\to\infty} \frac{G(n+1)}{n^{n^2/2-1/12}(2\pi)^{n/2} e^{-3n^2/4}} = \frac{e^{1/12}}{A}, \tag{4}$$

where

$$A = \exp\left[-\frac{\zeta'(2)}{2\pi^2} + \frac{\ln(2\pi)}{12} + \frac{\gamma}{2}\right] = 1.28242713\ldots, \tag{5}$$

where $\zeta(z)$ is the RIEMANN ZETA FUNCTION, π is PI, and γ is the EULER-MASCHERONI CONSTANT (Kinkelin 1860, Glaisher 1877, 1878, 1893, 1894). Glaisher (1877) also obtained

$$A = 2^{7/36}\pi^{-1/6} \exp\left\{\frac{1}{3} + \frac{2}{3}\int_0^{1/2} \ln[\Gamma(x+1)]\,dx\right\}. \tag{6}$$

Glaisher (1894) showed that

$$1^{1/1} 2^{1/4} 3^{1/9} 4^{1/16} 5^{1/25} \cdots = \left(\frac{A^{12}}{2\pi e^{\gamma}}\right)^{\pi^2/6} \tag{7}$$

$$1^{1/1} 3^{1/9} 5^{1/25} 7^{1/49} 9^{1/81} \cdots = \left(\frac{A^{12}}{2^{4/3}\pi e^{\gamma}}\right)^{\pi^2/8} \tag{8}$$

$$\frac{1^{1/1} 5^{1/125} 9^{1/729} \cdots}{3^{1/27} 7^{1/343} 11^{1/1331} \cdots} = \left(\frac{A}{2^{5/32}\pi^{1/32} e^{3/32+\gamma/48+s/4}}\right)^{\pi^3}, \tag{9}$$

where

$$s \equiv \frac{\zeta(3)}{3\cdot 4\cdot 5}\frac{1}{4^3} + \frac{\zeta(5)}{5\cdot 6\cdot 7}\frac{1}{4^5} + \frac{\zeta(7)}{7\cdot 8\cdot 9}\frac{1}{4^7} + \ldots. \tag{10}$$

see also G-FUNCTION, HYPERFACTORIAL, K-FUNCTION

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/glshkn/glshkn.html`.

Glaisher, J. W. L. "On a Numerical Continued Product." *Messenger Math.* **6**, 71–76, 1877.

Glaisher, J. W. L. "On the Product $1^1 2^2 3^3 \cdots n^n$." *Messenger Math.* **7**, 43–47, 1878.

Glaisher, J. W. L. "On Certain Numerical Products." *Messenger Math.* **23**, 145–175, 1893.

Glaisher, J. W. L. "On the Constant which Occurs in the Formula for $1^1 2^2 3^3 \cdots n^n$." *Messenger Math.* **24**, 1–16, 1894.

Kinkelin. "Über eine mit der Gammafunktion verwandte Transcendente und deren Anwendung auf die Integralrechnung." *J. Reine Angew. Math.* **57**, 122–158, 1860.

Glide

A product of a REFLECTION in a line and TRANSLATION along the same line.

see also REFLECTION, TRANSLATION

Glissette

The locus of a point P (or the envelope of a line) fixed in relation to a curve C which slides between fixed curves. For example, if C is a line segment and P a point on the line segment, then P describes an ELLIPSE when C slides so as to touch two ORTHOGONAL straight LINES. The glissette of the LINE SEGMENT C itself is, in this case, an ASTROID.

see also ROULETTE

References

Besant, W. H. *Notes on Roulettes and Glissettes, 2nd enl. ed.* Cambridge, England: Deighton, Bell & Co., 1890.

Lockwood, E. H. "Glissettes." Ch. 20 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 160–165, 1967.

Yates, R. C. "Glissettes." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 108–112, 1952.

Global $C(G;T)$ Theorem

If a SYLOW 2-SUBGROUP T of G lies in a unique maximal 2-local P of G, then P is a "strongly embedded" SUBGROUP of G, and G is known.

Global Extremum

A GLOBAL MINIMUM or GLOBAL MAXIMUM.

see also LOCAL EXTREMUM

Global Maximum

The largest overall value of a set, function, etc., over its entire range.

see also GLOBAL MINIMUM, LOCAL MAXIMUM, MAXIMUM

Global Minimum

The smallest overall value of a set, function, etc., over its entire range.

see also GLOBAL MAXIMUM, KUHN-TUCKER THEOREM, LOCAL MINIMUM, MINIMUM

Globe

A SPHERE which acts as a model of a spherical (or ellipsoidal) celestial body, especially the Earth, and on which the outlines of continents, oceans, etc. are drawn.

see also LATITUDE, LONGITUDE, SPHERE

Glove Problem

Let there be m doctors and $n \leq m$ patients, and let all mn possible combinations of examinations of patients by doctors take place. Then what is the minimum number of surgical gloves needed $G(m,n)$ so that no doctor must wear a glove contaminated by a patient and no patient is exposed to a glove worn by another doctor? In this problem, the gloves can be turned inside out and even placed on top of one another if necessary, but no "decontamination" of gloves is permitted. The optimal solution is

$$g(m,n) = \begin{cases} 2 & m = n = 2 \\ \frac{1}{2}(m+1) & n = 1,\, m = 2k+1 \\ \left\lceil \frac{1}{2}(m) + \frac{2}{3}n \right\rceil & \text{otherwise,} \end{cases}$$

where $\lceil x \rceil$ is the CEILING FUNCTION (Vardi 1991). The case $m = n = 2$ is straightforward since two gloves have a total of four surfaces, which is the number needed for $mn = 4$ examinations.

References

Gardner, M. *Aha! Aha! Insight.* New York: Scientific American, 1978.

Gardner, M. *Science Fiction Puzzle Tales.* New York: Crown, pp. 5, 67, and 104–150, 1981.

Hajnal, A. and Lovász, L. "An Algorithm to Prevent the Propagation of Certain Diseases at Minimum Cost." §10.1 in *Interfaces Between Computer Science and Operations Research* (Ed. J. K. Lenstra, A. H. G. Rinnooy Kan, and P. van Emde Boas). Amsterdam: Matematisch Centrum, 1978.

Orlitzky, A. and Shepp, L. "On Curbing Virus Propagation." Exercise 10.2 in Technical Memo. Bell Labs, 1989.

Vardi, I. "The Condom Problem." Ch. 10 in *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, p. 203–222, 1991.

Glue Vector

A VECTOR specifying how layers are stacked in a LAMINATED LATTICE.

Gnomic Number

A FIGURATE NUMBER of the form $g_n = 2n - 1$ which are the areas of square gnomons, obtained by removing a SQUARE of side $n-1$ from a SQUARE of side n,

$$g_n = n^2 - (n-1)^2 = 2n - 1.$$

The gnomic numbers are therefore equivalent to the ODD NUMBERS, and the first few are 1, 3, 5, 7, 9, 11, ... (Sloane's A005408). The GENERATING FUNCTION for the gnomic numbers is

$$\frac{x(1+x)}{(x-1)^2} = x + 3x^2 + 5x^3 + 7x^4 + \ldots.$$

see also ODD NUMBER

References

Sloane, N. J. A. Sequence A005408/M2400 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Gnomic Projection

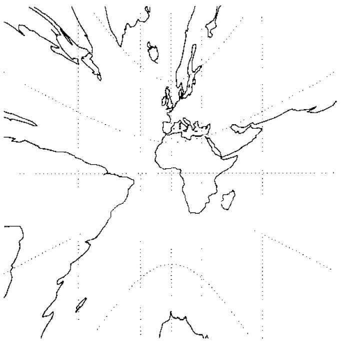

A nonconformal projection from a SPHERE's center in which ORTHODROMES are straight LINES.

$$x = \frac{\cos\phi \sin(\lambda - \lambda_0)}{\cos c} \quad (1)$$

$$y = \frac{\cos\phi_1 \sin\phi - \sin\phi_1 \cos\phi \cos(\lambda - \lambda_0)}{\cos c}, \quad (2)$$

where

$$\cos c = \sin\phi_1 \sin\phi + \cos\phi_1 \cos\phi \cos(\lambda - \lambda_0). \quad (3)$$

The inverse FORMULAS are

$$\phi = \sin^{-1}(\cos c \sin\phi_1 + y \sin c \cos c \cos\phi_1) \quad (4)$$

$$\lambda = \lambda_0 + \tan^{-1}\left(\frac{x}{\cos\phi_1 - y \sin\phi_1}\right). \quad (5)$$

References
Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 150–153, 1967.
Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 164–168, 1987.

Gnomon

A shape which, when added to a figure, yields another figure SIMILAR to the original.

References
Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 123, 1993.

Gnomon Magic Square

A 3×3 array of numbers in which the elements in each 2×2 corner have the same sum.

see also MAGIC SQUARE

Go

There are estimated to be about 4.63×10^{170} possible positions on a 19×19 board (Flammenkamp). The number of n-move Go games are 1, 362, 130683, 47046242, ... (Sloane's A007565).

References
Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Item 96, Feb. 1972.
Flammenkamp, A. "A Short, Concise Ruleset of Go." `http://www.minet.uni-jena.de/~achim/gorules.html`.
Kraitchik, M. "Go." §12.4 in *Mathematical Recreations.* New York: W. W. Norton, pp. 279–280, 1942.
Sloane, N. J. A. Sequence A007565/M5447 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Göbel's Sequence

Consider the RECURRENCE RELATION

$$x_n = \frac{1 + {x_0}^2 + {x_1}^2 + \ldots + {x_{n-1}}^2}{n}, \quad (1)$$

with $x_0 = 1$. The first few iterates of x_n are 1, 2, 3, 5, 10, 28, 154, ... (Sloane's A003504). The terms grow extremely rapidly, but are given by the asymptotic formula

$$x_n \approx (n^2 + 2n - 1 + 4n^{-1} - 21n^{-2} + 137n^{-3} - \ldots)C^{2^n}, \quad (2)$$

where

$$C = 1.0478314475764112295599094627431375545 9\ldots \quad (3)$$

(Zagier). It is more convenient to work with the transformed sequence

$$s_n = 2 + {x_1}^2 + {x_2}^2 + \ldots + {x_{n-1}}^2 = nx_n, \quad (4)$$

which gives the new recurrence

$$s_{n+1} = s_n + \frac{{s_n}^2}{n^2} \quad (5)$$

with initial condition $s_1 = 2$. Now, s_{n+1} will be nonintegral IFF $n \nmid s_n$. The smallest p for which $s_p \not\equiv 0 \pmod{p}$ therefore gives the smallest nonintegral s_{p+1}. In addition, since $p \nmid s_p$, $x_p = s_p/p$ is also the smallest nonintegral x_p.

For example, we have the sequences $\{s_n \pmod{k}\}_{n=1}^{k}$:

$$2, 6 \equiv 2, \tfrac{5}{4} \equiv 0, 0, 0 \pmod{5} \quad (6)$$

$$2, 6, 15 \equiv 1, \tfrac{5}{4} \equiv 0, 0, 0, 0 \pmod{7} \quad (7)$$

$$2, 6, 15 \equiv 4, \tfrac{52}{9} \equiv 7, \tfrac{161}{16} \equiv 8, \tfrac{264}{5} \equiv 0, 0, \ldots, 0 \pmod{11}. \quad (8)$$

Testing values of k shows that the first nonintegral x_n is x_{43}. Note that a direct verification of this fact is impossible since

$$x_{43} \approx 5.4093 \times 10^{178485291567} \quad (9)$$

(calculated using the asymptotic formula) is much too large to be computed and stored explicitly.

A sequence even more striking for remaining integral over many terms is the 3-Göbel sequence

$$x_n = \frac{1 + {x_0}^3 + {x_1}^3 + \ldots + {x_{n-1}}^3}{n}. \qquad (10)$$

The first few terms of this sequence are 1, 2, 5, 45, 22815, ... (Sloane's A005166).

The Göbel sequences can be generalized to k powers by

$$x_n = \frac{1 + {x_0}^k + {x_1}^k + \ldots + {x_{n-1}}^k}{n}. \qquad (11)$$

see also SOMOS SEQUENCE

References
Guy, R. K. "The Strong Law of Small Numbers." *Amer. Math. Monthly* **95**, 697–712, 1988.
Guy, R. K. "A Recursion of Göbel." §E15 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 214–215, 1994.
Sloane, N. J. A. Sequences A003504/M0728 and A005166/M1551 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Zaiger, D. "Solution: Day 5, Problem 3." `http://www-groups.dcs.st-and.ac.uk/~john/Zagier/Solution5.3.html`.

Goblet Illusion

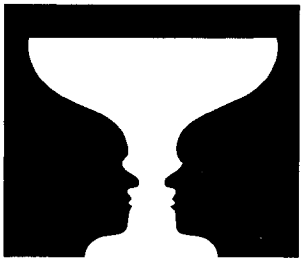

An ILLUSION in which the eye alternately sees two black faces, or a white goblet.

References
Fineman, M. *The Nature of Visual Illusion.* New York: Dover, pp. 111 and 115, 1996.
Rubin, E. *Synoplevede Figurer.* Copenhagen, Denmark: Gyldendalske, 1915.
What's Up with Kids Magazine. "Reversible Goblet." `http://wuwk.spurtek.com/COI_reversible_goblet.htm`.

Gödel's Completeness Theorem

If T is a set of AXIOMS in a first-order language, and a statement p holds for any structure M satisfying T, then p can be formally deduced from T in some appropriately defined fashion.

see also GÖDEL'S INCOMPLETENESS THEOREM

Gödel's Incompleteness Theorem

Informally, Gödel's incompleteness theorem states that all consistent axiomatic formulations of NUMBER THEORY include undecidable propositions (Hofstadter 1989). This is is sometimes called Gödel's first incompleteness theorem, and answers in the negative HILBERT'S PROBLEM asking whether mathematics is "complete" (in the sense that every statement in the language of NUMBER THEORY can be either proved or disproved). Formally, Gödel's theorem states, "To every ω-consistent recursive class κ of FORMULAS, there correspond recursive class-signs r such that neither (v Gen r) nor Neg(v Gen r) belongs to Flg(κ), where v is the FREE VARIABLE of r" (Gödel 1931).

A statement sometimes known as Gödel's second incompleteness theorem states that if NUMBER THEORY is consistent, then a proof of this fact does not exist using the methods of first-order PREDICATE CALCULUS. Stated more colloquially, any formal system that is interesting enough to formulate its own consistency can prove its own consistency IFF it is inconsistent.

Gerhard Gentzen showed that the consistency and completeness of arithmetic can be proved if "transfinite" induction is used. However, this approach does not allow proof of the consistency of all mathematics.

see also GÖDEL'S COMPLETENESS THEOREM, HILBERT'S PROBLEMS, KREISEL CONJECTURE, NATURAL INDEPENDENCE PHENOMENON, NUMBER THEORY, RICHARDSON'S THEOREM, UNDECIDABLE

References
Barrow, J. D. *Pi in the Sky: Counting, Thinking, and Being.* Oxford, England: Clarendon Press, p. 121, 1992.
Gödel, K. "Über Formal Unentscheidbare Sätze der *Principia Mathematica* und Verwandter Systeme, I." *Monatshefte für Math. u. Physik* **38**, 173–198, 1931.
Gödel, K. *On Formally Undecidable Propositions of Principia Mathematica and Related Systems.* New York: Dover, 1992.
Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, p. 17, 1989.
Kolata, G. "Does Gödel's Theorem Matter to Mathematics?" *Science* **218**, 779–780, 1982.
Smullyan, R. M. *Gödel's Incompleteness Theorems.* New York: Oxford University Press, 1992.
Whitehead, A. N. and Russell, B. *Principia Mathematica.* New York: Cambridge University Press, 1927.

Gödel Number

A Gödel number is a unique number associated to a statement about arithmetic. It is formed as the PRODUCT of successive PRIMES raised to the POWER of the number corresponding to the individual symbols that comprise the sentence. For example, the statement $(\exists x)(x = sy)$ that reads "there EXISTS an x such that x is the immediate successor of y" is coded

$$(2^8)(3^4)(5^{13})(7^9)(11^8)(13^{13})(17^5)(19^7)(23^{16})(29^9),$$

where the numbers in the set (8, 4, 13, 9, 8, 13, 5, 7, 16, 9) correspond to the symbols that make up $(\exists x)(x = sy)$.

see also GÖDEL'S INCOMPLETENESS THEOREM

References

Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, p. 18, 1989.

Goldbach Conjecture

Goldbach's original conjecture, written in a 1742 letter to Euler, states that every INTEGER > 5 is the SUM of three PRIMES. As re-expressed by Euler, an equivalent of this CONJECTURE (called the "strong" Goldbach conjecture) asserts that all POSITIVE EVEN INTEGERS ≥ 4 can be expressed as the SUM of two PRIMES. Schnirelmann (1931) proved that every EVEN number can be written as the sum of not more than 300,000 PRIMES (Dunham 1990), which seems a rather far cry from a proof for *four* PRIMES! The strong Goldbach conjecture has been shown to be true up to 4×10^{11} by Sinisalo (1993). Pogorzelski (1977) claimed to have proven the Goldbach conjecture, but his proof is not generally accepted (Shanks 1993).

The conjecture that all ODD numbers ≥ 9 are the SUM of three ODD PRIMES is called the "weak" Goldbach conjecture. Vinogradov proved that all ODD INTEGERS starting at some sufficiently large value are the SUM of three PRIMES (Guy 1994). The original "sufficiently large" $N \geq 3^{3^{15}} = e^{e^{16.573}}$ was subsequently reduced to $e^{e^{11.503}}$ by Chen and Wang (1989). Chen (1973, 1978) also showed that all sufficiently large EVEN NUMBERS are the sum of a PRIME and the PRODUCT of at most two PRIMES (Guy 1994, Courant and Robbins 1996).

It has been shown that if the weak Goldbach conjecture is false, then there are only a FINITE number of exceptions.

Other variants of the Goldbach conjecture include the statements that every EVEN number ≥ 6 is the SUM of two ODD PRIMES, and every INTEGER > 17 the sum of exactly three distinct PRIMES. Let $R(n)$ be the number of representations of an EVEN INTEGER n as the sum of two PRIMES. Then the "extended" Goldbach conjecture states that

$$R(n) \sim 2\Pi_2 \prod_{\substack{k=2\\ p_k \mid n}} \frac{p_k - 1}{p_k - 2} \int_2^x \frac{dx}{(\ln x)^2},$$

where Π_2 is the TWIN PRIMES CONSTANT (Halberstam and Richert 1974).

If the Goldbach conjecture is true, then for every number m, there are PRIMES p and q such that

$$\phi(p) + \phi(q) = 2m,$$

where $\phi(x)$ is the TOTIENT FUNCTION (Guy 1994, p. 105).

Vinogradov (1937ab, 1954) proved that every sufficiently large ODD NUMBER is the sum of three PRIMES, and Estermann (1938) proves that almost all EVEN NUMBERS are the sums of two PRIMES.

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 64, 1987.
Chen, J.-R. "On the Representation of a Large Even Number as the Sum of a Prime and the Product of at Most Two Primes." *Sci. Sinica* **16**, 157–176, 1973.
Chen, J.-R. "On the Representation of a Large Even Number as the Sum of a Prime and the Product of at Most Two Primes, II." *Sci. Sinica* **21**, 421–430, 1978.
Chen, J.-R. and Wang, T.-Z. "On the Goldbach Problem." *Acta Math. Sinica* **32**, 702–718, 1989.
Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 30–31, 1996.
Devlin, K. *Mathematics: The New Golden Age.* London: Penguin Books, 1988.
Dunham, W. *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, p. 83, 1990.
Estermann, T. "On Goldbach's Problem: Proof that Almost All Even Positive Integers are Sums of Two Primes." *Proc. London Math. Soc. Ser. 2* **44**, 307–314, 1938.
Guy, R. K. "Goldbach's Conjecture." §C1 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 105–107, 1994.
Hardy, G. H. and Littlewood, J. E. "Some Problems of Partitio Numerorum (V): A Further Contribution to the Study of Goldbach's Problem." *Proc. London Math. Soc. Ser. 2* **22**, 46–56, 1924.
Halberstam, H. and Richert, H.-E. *Sieve Methods.* New York: Academic Press, 1974.
Pogorzelski, H. A. "Goldbach Conjecture." *J. Reine Angew. Math.* **292**, 1–12, 1977.
Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 30–31 and 222, 1985.
Sinisalo, M. K. "Checking the Goldbach Conjecture up to $4 \cdot 10^{11}$." *Math. Comput.* **61**, 931–934, 1993.
Vinogradov, I. M. "Representation of an Odd Number as a Sum of Three Primes." *Comtes rendus (Doklady) de l'Académie des Sciences de l'U.R.S.S.* **15**, 169–172, 1937a.
Vinogradov, I. "Some Theorems Concerning the Theory of Primes." *Recueil Math.* **2**, 179–195, 1937b.
Vinogradov, I. M. *The Method of Trigonometrical Sums in the Theory of Numbers.* London: Interscience, p. 67, 1954.
Yuan, W. *Goldbach Conjecture.* Singapore: World Scientific, 1984.

Golden Mean

see GOLDEN RATIO

Golden Ratio

A number often encountered when taking the ratios of distances in simple geometric figures such as the DECAGON and DODECAGON. It is denoted ϕ, or sometimes τ (which is an abbreviation of the Greek "tome," meaning "to cut"). ϕ is also known as the DIVINE PROPORTION, GOLDEN MEAN, and GOLDEN SECTION and is a PISOT-VIJAYARAGHAVAN CONSTANT. It has surprising connections with CONTINUED FRACTIONS and the

EUCLIDEAN ALGORITHM for computing the GREATEST COMMON DIVISOR of two INTEGERS.

Given a RECTANGLE having sides in the ratio $1 : \phi$, ϕ is defined such that partitioning the original RECTANGLE into a SQUARE and new RECTANGLE results in a new RECTANGLE having sides with a ratio $1 : \phi$. Such a RECTANGLE is called a GOLDEN RECTANGLE, and successive points dividing a GOLDEN RECTANGLE into SQUARES lie on a LOGARITHMIC SPIRAL. This figure is known as a WHIRLING SQUARE.

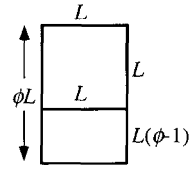

This means that

$$\frac{1}{\phi - 1} = \phi \qquad (1)$$

$$\phi^2 - \phi - 1 = 0. \qquad (2)$$

So, by the QUADRATIC EQUATION,

$$\phi = \tfrac{1}{2}(1 \pm \sqrt{1+4}) = \tfrac{1}{2}(1 + \sqrt{5}) \qquad (3)$$

$$= 1.6180339887498948482045868343656381177720\ldots \qquad (4)$$

(Sloane's A001622).

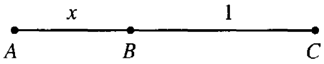

A geometric definition can be given in terms of the above figure. Let the ratio $x \equiv AB/BC$. The NUMERATOR and DENOMINATOR can then be taken as $\overline{AB} = x$ and $\overline{BC} = 1$ without loss of generality. Now define the position of B by

$$\frac{BC}{AB} = \frac{AB}{AC}. \qquad (5)$$

Plugging in gives

$$\frac{1}{x} = \frac{x}{1+x}, \qquad (6)$$

or

$$x^2 - x - 1 = 0, \qquad (7)$$

which can be solved using the QUADRATIC EQUATION to obtain

$$\phi \equiv x = \frac{1 \pm \sqrt{1^2 - (-4)}}{2} = \tfrac{1}{2}(1 + \sqrt{5}). \qquad (8)$$

ϕ is the "most" IRRATIONAL number because it has a CONTINUED FRACTION representation

$$\phi = [1, 1, 1, \ldots] \qquad (9)$$

(Sloane's A000012). Another infinite representation in terms of a CONTINUED SQUARE ROOT is

$$\phi = \sqrt{1 + \sqrt{1 + \sqrt{1 + \sqrt{1 + \ldots}}}}. \qquad (10)$$

Ramanujan gave the curious CONTINUED FRACTION identities

$$\frac{1}{(\sqrt{\phi\sqrt{5}})e^{2\pi/5}} = 1 + \cfrac{e^{-2\pi}}{1 + \cfrac{e^{-4\pi}}{1 + \cfrac{e^{-6\pi}}{1 + \cfrac{e^{-8\pi}}{1 + \cfrac{e^{-10\pi}}{1 + \ldots}}}}} \qquad (11)$$

$$\frac{1}{\left\{\frac{\sqrt{5}}{1+\left[5^{3/4}(\phi-1)^{5/2}-1\right]-\phi}\right\} e^{2\pi/\sqrt{5}}} = 1 + \cfrac{e^{-2\pi\sqrt{5}}}{1 + \cfrac{e^{-4\pi\sqrt{5}}}{1 + \cfrac{e^{-6\pi\sqrt{5}}}{1 + \cfrac{e^{-8\pi\sqrt{5}}}{1 + \cfrac{e^{-10\pi\sqrt{5}}}{1 + \ldots}}}}} \qquad (12)$$

(Ramanathan 1984).

The legs of a GOLDEN TRIANGLE are in a golden ratio to its base. In fact, this was the method used by Pythagoras to construct ϕ. Euclid used the following construction.

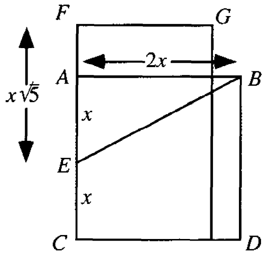

Draw the SQUARE $\square ABDC$, call E the MIDPOINT of AC, so that $AE = EC \equiv x$. Now draw the segment BE, which has length

$$x\sqrt{2^2 + 1^2} = x\sqrt{5}, \qquad (13)$$

and construct EF with this length. Now construct $FG = EF$, then

$$\phi = \frac{FC}{CD} = \frac{EF + CE}{CD} = \frac{x(\sqrt{5}+1)}{2x} = \tfrac{1}{2}(\sqrt{5}+1). \qquad (14)$$

The ratio of the CIRCUMRADIUS to the length of the side of a DECAGON is also ϕ,

$$\frac{R}{s} = \tfrac{1}{2}\csc\left(\frac{\pi}{10}\right) = \tfrac{1}{2}(1+\sqrt{5}\,) = \phi. \qquad (15)$$

Similarly, the legs of a GOLDEN TRIANGLE (an ISOSCELES TRIANGLE with a VERTEX ANGLE of 36°) are in a GOLDEN RATIO to the base. Bisecting a GAULLIST CROSS also gives a golden ratio (Gardner 1961, p. 102).

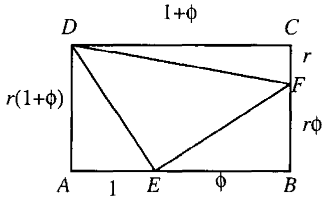

In the figure above, three TRIANGLES can be INSCRIBED in the RECTANGLE $\square ABCD$ of arbitrary aspect ratio $1 : r$ such that the three RIGHT TRIANGLES have equal areas by dividing AB and BC in the golden ratio. Then

$$K_{\triangle ADE} = \tfrac{1}{2}\cdot r(1+\phi)\cdot 1 = \tfrac{1}{2}r\phi^2 \qquad (16)$$

$$K_{\triangle BEF} = \tfrac{1}{2}\cdot r\phi\cdot\phi = \tfrac{1}{2}r\phi^2 \qquad (17)$$

$$K_{\triangle CDF} = \tfrac{1}{2}(1+\phi)\cdot r = \tfrac{1}{2}r\phi^2, \qquad (18)$$

which are all equal.

The golden ratio also satisfies the RECURRENCE RELATION

$$\phi^{n+1} = \phi^{n-1} + \phi^n. \qquad (19)$$

Taking $n = 0$ gives

$$\phi = \phi^{-1} + 1 \qquad (20)$$

$$\phi^2 = 1 + \phi. \qquad (21)$$

But this is the definition equation for ϕ (when the root with the plus sign is used). Squaring gives

$$\begin{aligned}\phi^2 &= \tfrac{1}{4}(5+2\sqrt{5}+1) = \tfrac{1}{4}(6+2\sqrt{5}\,) = \tfrac{1}{2}(3+\sqrt{5}\,)\\ &= \tfrac{1}{2}(\sqrt{5}+1)+1 = \phi^0+\phi^1 \qquad (22)\\ \phi^3 &= (\phi^0+\phi^1)\phi^1 = \phi^0\phi^1 + (\phi^1)^2 = \phi^1+\phi^2, \qquad (23)\end{aligned}$$

and so on.

For the difference equations

$$\begin{cases} x_0 = 1 \\ x_n = 1 + \dfrac{1}{x_{n-1}} & \text{for } n = 1, 2, 3, \end{cases} \qquad (24)$$

ϕ is also given by

$$\phi = \lim_{n\to\infty} x_n. \qquad (25)$$

In addition,

$$\phi = \lim_{n\to\infty}\frac{F_n}{F_{n-1}}, \qquad (26)$$

where F_n is the nth FIBONACCI NUMBER.

The SUBSTITUTION MAP

$$0 \to 01 \qquad (27)$$

$$1 \to 0 \qquad (28)$$

gives

$$0 \to 01 \to 010 \to 01001 \to \ldots, \qquad (29)$$

giving rise to the sequence

$$0100101001001010010100100101\ldots \qquad (30)$$

(Sloane's A003849). Here, the zeros occur at positions 1, 3, 4, 6, 8, 9, 11, 12, ... (Sloane's A000201), and the ones occur at positions 2, 5, 7, 10, 13, 15, 18, ... (Sloane's A001950). These are complementary BEATTY SEQUENCES generated by $\lfloor n\phi\rfloor$ and $\lfloor n\phi^2\rfloor$. The sequence also has many connections with the FIBONACCI NUMBERS.

Salem showed that the set of PISOT-VIJAYARAGHAVAN CONSTANTS is closed, with ϕ the smallest accumulation point of the set (Le Lionnais 1983).

see also BERAHA CONSTANTS, DECAGON, FIVE DISKS PROBLEM, GOLDEN RATIO CONJUGATE, GOLDEN TRIANGLE, ICOSIDODECAHEDRON, NOBLE NUMBER, PENTAGON, PENTAGRAM, PHI NUMBER SYSTEM, SECANT METHOD

References

Boyer, C. B. *History of Mathematics.* New York: Wiley, p. 56, 1968.

Coxeter, H. S. M. "The Golden Section, Phyllotaxis, and Wythoff's Game." *Scripta Mathematica* **19**, 135–143, 1953.

Dixon, R. *Mathographics.* New York: Dover, pp. 30–31 and 50, 1991.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/cntfrc/cntfrc.html`.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/gold/gold.html`.

Gardner, M. "Phi: The Golden Ratio." Ch. 8 in *The Second Scientific American Book of Mathematical Puzzles & Diversions, A New Selection.* New York: Simon and Schuster, 1961.

Gardner, M. "Notes on a Fringe-Watcher: The Cult of the Golden Ratio." *Skeptical Inquirer* **18**, 243–247, 1994.

Herz-Fischler, R. *A Mathematical History of the Golden Number.* New York: Dover, 1998.

Huntley, H. E. *The Divine Proportion.* New York: Dover, 1970.

Knott, R. "Fibonacci Numbers and the Golden Section." `http://www.mcs.surrey.ac.uk/Personal/R.Knott/Fibonacci/fib.html`.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 40, 1983.

Markowsky, G. "Misconceptions About the Golden Ratio." *College Math. J.* **23**, 2–19, 1992.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 122–134, 1990.

Pappas, T. "Anatomy & the Golden Section." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 32–33, 1989.
Ramanathan, K. G. "On Ramanujan's Continued Fraction." *Acta. Arith.* **43**, 209–226, 1984.
Sloane, N. J. A. Sequences A003849, A000012/M0003, A000201/M2322, A001622/M4046, and A001950/M1332 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Golden Ratio Conjugate

The quantity

$$\phi_C \equiv \frac{1}{\phi} = \phi - 1 = \frac{\sqrt{5}-1}{2} \approx 0.6180339885, \qquad (1)$$

where ϕ is the GOLDEN RATIO. The golden ratio conjugate is sometimes also called the SILVER RATIO. A quantity similar to the FEIGENBAUM CONSTANT can be found for the nth CONTINUED FRACTION representation

$$[a_0, a_1, a_2, \ldots]. \qquad (2)$$

Taking the limit of

$$\delta_n \equiv \frac{\sigma_n - \sigma_{n-1}}{\sigma_n - \sigma_{n+1}} \qquad (3)$$

gives

$$\delta \equiv \lim_{n\to\infty} = 1 + \phi = 2 + \phi_C. \qquad (4)$$

see also GOLDEN RATIO, SILVER RATIO

Golden Rectangle

Given a RECTANGLE having sides in the ratio $1 : \phi$, the GOLDEN RATIO ϕ is defined such that partitioning the original RECTANGLE into a SQUARE and new RECTANGLE results in a new RECTANGLE having sides with a ratio $1 : \phi$. Such a RECTANGLE is called a golden rectangle, and successive points dividing a golden rectangle into SQUARES lie on a LOGARITHMIC SPIRAL.

see also GOLDEN RATIO, LOGARITHMIC SPIRAL, RECTANGLE

References

Pappas, T. "The Golden Rectangle." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 102–106, 1989.

Golden Rule

The mathematical golden rule states that, for any FRACTION, both NUMERATOR and DENOMINATOR may be multiplied by the same number without changing the fraction's value.

see also DENOMINATOR, FRACTION, NUMERATOR

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 151, 1996.

Golden Section

see GOLDEN RATIO

Golden Theorem

see QUADRATIC RECIPROCITY THEOREM

Golden Triangle

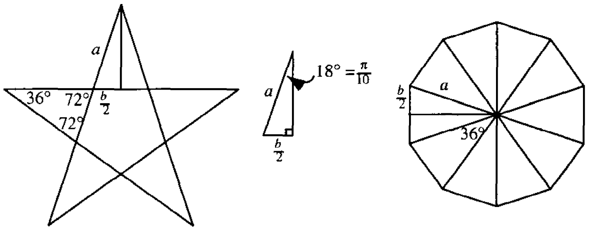

An ISOSCELES TRIANGLE with VERTEX angles 36°. Such TRIANGLES occur in the PENTAGRAM and DECAGON. The legs are in a GOLDEN RATIO to the base. For such a TRIANGLE,

$$\sin(18^\circ) = \sin(\tfrac{1}{10}\pi) = \frac{\frac{1}{2}b}{l} \qquad (1)$$

$$b = 2a\sin(\tfrac{1}{10}\pi) = 2a\frac{\sqrt{5}-1}{4} = \tfrac{1}{2}a(\sqrt{5}-1) \qquad (2)$$

$$b + l = \tfrac{1}{2}a(\sqrt{5}+1) \qquad (3)$$

$$\frac{b+a}{a} = \frac{\sqrt{5}+1}{2} = \phi. \qquad (4)$$

see also DECAGON, GOLDEN RATIO, ISOSCELES TRIANGLE, PENTAGRAM

References

Pappas, T. "The Pentagon, the Pentagram & the Golden Triangle." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 188–189, 1989.

Goldschmidt Solution

The discontinuous solution of the SURFACE OF REVOLUTION AREA minimization problem for surfaces connecting two CIRCLES. When the CIRCLES are sufficiently far apart, the usual CATENOID is no longer stable and the surface will break and form two surfaces with the CIRCLES as boundaries.

see also CALCULUS OF VARIATIONS, SURFACE OF REVOLUTION

Golomb Constant

see GOLOMB-DICKMAN CONSTANT

Golomb-Dickman Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let Π be a PERMUTATION of n elements, and let α_i be the number of CYCLES of length i in this PERMUTATION. Picking Π at RANDOM gives

$$\left\langle \sum_{j=1}^{\infty} \alpha_j \right\rangle = \sum_{i=1}^{n} \frac{1}{i} = \ln n + \gamma + \mathcal{O}\left(\frac{1}{n}\right) \tag{1}$$

$$\operatorname{var}\left(\sum_{j=1}^{\infty} \alpha_j\right) = \sum_{i=1}^{n} \frac{i-1}{i^2} = \ln n + \gamma - \tfrac{1}{6}\pi^2 + \mathcal{O}\left(\frac{1}{n}\right) \tag{2}$$

$$\lim_{n\to\infty} P(\alpha_1 = 0) = \frac{1}{e} \tag{3}$$

(Shepp and Lloyd 1966, Wilf 1990). Goncharov (1942) showed that

$$\lim_{n\to\infty} P(\alpha_j = k) = \frac{1}{k!} e^{-1/j} j^{-k}, \tag{4}$$

which is a POISSON DISTRIBUTION, and

$$\lim_{n\to\infty} P\left[\left(\sum_{j=1}^{\infty} \alpha_j - \ln n\right) (\ln n)^{-1/2} \leq x\right] = \Phi(x), \tag{5}$$

which is a NORMAL DISTRIBUTION, γ is the EULER-MASCHERONI CONSTANT, and Φ is the NORMAL DISTRIBUTION FUNCTION. Let

$$M(\alpha) \equiv \max_{1\leq j<\infty} \alpha_j \tag{6}$$

$$m(\alpha) \equiv \min_{1\leq j<\infty} \alpha_j. \tag{7}$$

Golomb (1959) derived

$$\lambda \equiv \lim_{n\to\infty} \frac{\langle M(\alpha)\rangle}{n} = 0.6243299885\ldots, \tag{8}$$

which is known as the GOLOMB CONSTANT or Golomb-Dickman constant. Knuth (1981) asked for the constants b and c such that

$$\lim_{n\to\infty} n^b[\langle M(\alpha)\rangle - \lambda n - \tfrac{1}{2}\lambda] = c, \tag{9}$$

and Gourdon (1996) showed that

$$\langle M(\alpha)\rangle = \lambda(n + \tfrac{1}{2}) - \frac{e^\gamma}{24n} + \frac{\frac{1}{48}e^\gamma - \frac{1}{8}(-1)^n}{n^2} + \frac{\frac{17}{3840}e^\gamma + \frac{1}{8}(-1)^n + \frac{1}{6}j^{1+2n} + \frac{1}{6}j^{2+n}}{n^3}, \tag{10}$$

where

$$j \equiv e^{2\pi i/3}. \tag{11}$$

λ can be expressed in terms of the function $f(x)$ defined by $f(x) = 1$ for $1 \leq x \leq 2$ and

$$\frac{df}{dx} = -\frac{f(x-1)}{x-1} \tag{12}$$

for $x > 2$, by

$$\lambda = \int_1^\infty \frac{f(x)}{x^2}\,dx. \tag{13}$$

Shepp and Lloyd (1966) derived

$$\begin{aligned}\lambda &= \int_0^\infty \exp\left(-x - \int_x^\infty \frac{e^{-y}}{y}\,dy\right)\\ &= \int_0^1 \exp\left(\int_0^x \frac{dy}{\ln y}\right) dx.\end{aligned} \tag{14}$$

Mitchell (1968) computed λ to 53 decimal places.

Surprisingly enough, there is a connection between λ and PRIME FACTORIZATION (Knuth and Pardo 1976, Knuth 1981, pp. 367-368, 395, and 611). Dickman (1930) investigated the probability $P(x,n)$ that the largest PRIME FACTOR p of a random INTEGER between 1 and n satisfies $p < n^x$ for $x \in (0,1)$. He found that

$$F(x) \equiv \lim_{n\to\infty} P(x,n) = \begin{cases} 1 & \text{if } x \geq 1 \\ \int_0^x F\left(\frac{t}{1-t}\right)\frac{dt}{t} & \text{if } 0 \leq x < 1. \end{cases} \tag{15}$$

Dickman then found the average value of x such that $p = n^x$, obtaining

$$\begin{aligned}\mu \equiv \lim_{n\to\infty} \langle x\rangle &= \lim_{n\to\infty}\left\langle\frac{\ln p}{\ln n}\right\rangle = \int_0^1 x\frac{dF}{dx}\,dx\\ &= \int_0^1 F\left(\frac{1}{1-t}\right) dt = 0.62432999,\end{aligned} \tag{16}$$

which is λ.

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/golomb/golomb.html`.

Gourdon, X. 1996. `http://www.mathsoft.com/asolve/constant/golomb/gourdon.html`.

Knuth, D. E. *The Art of Computer Programming, Vol. 1: Fundamental Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, 1973.

Knuth, D. E. *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, 1981.

Knuth, D. E. and Pardo, L. T. "Analysis of a Simple Factorization Algorithm." *Theor. Comput. Sci.* **3**, 321–348, 1976.

Mitchell, W. C. "An Evaluation of Golomb's Constant." *Math. Comput.* **22**, 411–415, 1968.

Purdom, P. W. and Williams, J. H. "Cycle Length in a Random Function." *Trans. Amer. Math. Soc.* **133**, 547–551, 1968.

Shepp, L. A. and Lloyd, S. P. "Ordered Cycle Lengths in Random Permutation." *Trans. Amer. Math. Soc.* **121**, 350–557, 1966.

Wilf, H. S. *Generatingfunctionology, 2nd ed.* New York: Academic Press, 1993.

Golomb Ruler

A Golomb ruler is a set of NONNEGATIVE integers such that all pairwise POSITIVE differences are distinct. The optimum Golomb ruler with n marks is the Golomb ruler having the smallest possible maximum element ("length"). The set (0, 1, 3, 7) is an order four Golomb ruler since its differences are ($1 = 1 - 0$, $2 = 3 - 1$, $3 = 3 - 0$, $4 = 7 - 3$, $6 = 7 - 1$, $7 = 7 - 0$), all of which are distinct. However, the optimum 4-mark Golomb ruler is (0, 1, 4, 6), which measures the distances (1, 2, 3, 4, 5, 6) (and is therefore also a PERFECT RULER).

The lengths of the optimal n-mark Golomb rulers for $n = 2, 3, 4, \ldots$ are 1, 3, 6, 11, 17, 25, 34, ... (Sloane's A003022, Vanderschel and Garry). The lengths of the optimal n-mark Golomb rulers are not known for $n \geq 20$.

see also PERFECT DIFFERENCE SET, PERFECT RULER, RULER, TAYLOR'S CONDITION, WEIGHINGS

References

Atkinson, M. D.; Santoro, N.; and Urrutia, J. "Integer Sets with Distinct Sums and Differences and Carrier Frequency Assignments for Nonlinear Repeaters." *IEEE Trans. Comm.* **34**, 614–617, 1986.

Colbourn, C. J. and Dinitz, J. H. (Eds.) *CRC Handbook of Combinatorial Designs.* Boca Raton, FL: CRC Press, p. 315, 1996.

Guy, R. K. "Modular Difference Sets and Error Correcting Codes." §C10 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 118–121, 1994.

Lam, A. W. and D. V. Sarwate, D. V. "On Optimum Time Hopping Patterns." *IEEE Trans. Comm.* **36**, 380–382, 1988.

Robinson, J. P. and Bernstein, A. J. "A Class of Binary Recurrent Codes with Limited Error Propagation." *IEEE Trans. Inform. Th. 13*, 106–113, 1967.

Sloane, N. J. A. Sequence A003022/M2540 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Vanderschel, D. and Garry, M. "In Search of the Optimal 20 & 21 Mark Golomb Rulers." `http://members.aol.com/golomb20/`.

Golygon

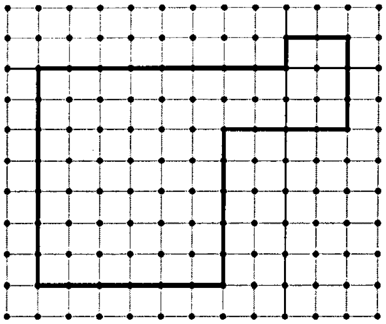

A PLANE path on a set of equally spaced LATTICE POINTS, starting at the ORIGIN, where the first step is one unit to the north or south, the second step is two units to the east or west, the third is three units to the north or south, etc., and continuing until the ORIGIN is again reached. No crossing or backtracking is allowed. The simplest golygon is (0, 0), (0, 1), (2, 1), (2, −2), (−2, −2), (−2, −7), (−8, −7), (−8, 0), (0, 0).

A golygon can be formed if there exists an EVEN INTEGER n such that

$$\pm 1 \pm 3 \pm \ldots \pm (n-1) = 0 \tag{1}$$

$$\pm 2 \pm 4 \pm \ldots \pm n = 0 \tag{2}$$

(Vardi 1991). Gardner proved that all golygons are of the form $n = 8k$. The number of golygons of length n (EVEN), with each initial direction counted separately, is the PRODUCT of the COEFFICIENT of $x^{n^2/8}$ in

$$(1+x)(1+x^3)\cdots(1+x^{n-1}), \tag{3}$$

with the COEFFICIENT of $x^{n(n/2+1)/8}$ in

$$(1+x)(1+x^2)\cdots(1+x^{n/2}). \tag{4}$$

The number of golygons $N(n)$ of length $8n$ for the first few n are 4, 112, 8432, 909288, ... (Sloane's A006718) and is asymptotic to

$$N(n) \sim \frac{3 \cdot 2^{8n-4}}{\pi n^2(4n+1)} \tag{5}$$

(Sallows *et al.* 1991, Vardi 1991).

see also LATTICE PATH

References

Dudeney, A. K. "An Odd Journey Along Even Roads Leads to Home in Golygon City." *Sci. Amer.* **263**, 118–121, July 1990.

Sallows, L. C. F. "New Pathways in Serial Isogons." *Math. Intell.* **14**, 55–67, 1992.

Sallows, L.; Gardner, M.; Guy, R. K.; and Knuth, D. "Serial Isogons of 90 Degrees." *Math Mag.* **64**, 315–324, 1991.

Sloane, N. J. A. Sequence A006718/M3707 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Vardi, I. "American Science." §5.3 in *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 90–96, 1991.

Gomory's Theorem

Regardless of where one white and one black square are deleted from an ordinary 8×8 CHESSBOARD, the reduced board can always be covered exactly with 31 DOMINOES (of dimension 2 × 1).

see also CHESSBOARD

Gompertz Constant

$$G \equiv \int_0^\infty \frac{e^{-u}}{1+u}\,du = -e\,\mathrm{ei}(-1) = 0.596347362\ldots,$$

where $\mathrm{ei}(x)$ is the EXPONENTIAL INTEGRAL. Stieltjes showed it has the CONTINUED FRACTION representation

$$G = \frac{1}{2-}\,\frac{1^2}{4-}\,\frac{2^2}{6-}\,\frac{3^2}{8-}\cdots.$$

see also EXPONENTIAL INTEGRAL

References

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 29, 1983.

Gompertz Curve

The function defined by

$$y = ab^{q^x}.$$

It is used in actuarial science for specifying a simplified mortality law. Using $s(x)$ as the probability that a newborn will achieve age x, the Gompertz law (1825) is

$$s(x) = \exp[-m(c^x - 1)],$$

for $c > 1$, $x \geq 0$.

see also LIFE EXPECTANCY, LOGISTIC GROWTH CURVE, MAKEHAM CURVE, POPULATION GROWTH

References

Bowers, N. L. Jr.; Gerber, H. U.; Hickman, J. C.; Jones, D. A.; and Nesbitt, C. J. *Actuarial Mathematics.* Itasca, IL: Society of Actuaries, p. 71, 1997.

Gompertz, B. "On the Nature of the Function Expressive of the Law of Human Mortality." *Phil. Trans. Roy. Soc. London,* 1825.

Gonal Number

see POLYGONAL NUMBER

Good Path

see p-GOOD PATH

Good Prime

A PRIME p_n is called "good" if

$${p_n}^2 > p_{n-i}p_{n+i}$$

for all $1 \leq i \leq n-1$ (there is a typo in Guy 1994 in which the is are replaced by 1s). There are infinitely many good primes, and the first few are 5, 11, 17, 29, 37, 41, 53, ... (Sloane's A028388).

see also ANDRICA'S CONJECTURE, PÓLYA CONJECTURE

References

Guy, R. K. "'Good' Primes and the Prime Number Graph." §A14 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 32–33, 1994.

Sloane, N. J. A. Sequence A028388 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Goodman's Formula

A two-coloring of a COMPLETE GRAPH K_n of n nodes which contains exactly the number of MONOCHROMATIC FORCED TRIANGLES and no more (i.e., a minimum of $R+B$ where R and B are the number of red and blue TRIANGLES) is called an EXTREMAL GRAPH. Goodman (1959) showed that for an extremal graph,

$$R+B = \begin{cases} \frac{1}{3}m(m-1)(m-2) & \text{for } n = 2m \\ \frac{2}{3}m(m-1)(4m+1) & \text{for } n = 4m+1 \\ \frac{2}{3}m(m+1)(4m-1) & \text{for } n = 4m+3. \end{cases}$$

Schwenk (1972) rewrote the equation in the form

$$R+B = \binom{n}{3} - \left\lfloor \tfrac{1}{2}n \left\lfloor \tfrac{1}{4}(n-1)^2 \right\rfloor \right\rfloor,$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT and $\lfloor x \rfloor$ is the FLOOR FUNCTION.

see also BLUE-EMPTY GRAPH, EXTREMAL GRAPH, MONOCHROMATIC FORCED TRIANGLE

References

Goodman, A. W. "On Sets of Acquaintances and Strangers at Any Party." *Amer. Math. Monthly* **66**, 778–783, 1959.

Schwenk, A. J. "Acquaintance Party Problem." *Amer. Math. Monthly* **79**, 1113–1117, 1972.

Goodstein Sequence

Given a HEREDITARY REPRESENTATION of a number n in BASE, let $B[b](n)$ be the NONNEGATIVE INTEGER which results if we syntactically replace each b by $b+1$ (i.e., $B[b]$ is a base change operator that 'bumps the base' from b up to $b+1$). The HEREDITARY REPRESENTATION of 266 in base 2 is

$$\begin{aligned} 266 &= 2^8 + 2^3 + 2 \\ &= 2^{2^{2+1}} + 2^{2+1} + 2, \end{aligned}$$

so bumping the base from 2 to 3 yields

$$B[2](266) = 3^{3^{3+1}} + 3^{3+1} + 3.$$

Now repeatedly bump the base and subtract 1,

$$\begin{aligned} G_0(266) &= 266 = 2^{2^{2+1}} + 2^{2+1} + 2 \\ G_1(266) &= B[2](266) - 1 = 3^{3^{3+1}} + 3^{3+1} + 2 \\ G_2(266) &= B[3](G_1) - 1 = 4^{4^{4+1}} + 4^{4+1} + 1 \\ G_3(266) &= B[4](G_2) - 1 = 5^{5^{5+1}} + 5^{5+1} \\ G_4(266) &= B[5](G_3) - 1 = 6^{6^{6+1}} + 6^{6+1} - 1 \\ &= 6^{6^{6+1}} + 5 \cdot 6^6 + 5 \cdot 6^5 + \ldots + 5 \cdot 6 + 5 \\ G_5(266) &= B[6](G_4) - 1 \\ &= 7^{7^{7+1}} + 5 \cdot 7^7 + 5 \cdot 7^5 + \ldots + 5 \cdot 7 + 4, \end{aligned}$$

etc. Starting this procedure at an INTEGER n gives the Goodstein sequence $\{G_k(n)\}$. Amazingly, despite the apparent rapid increase in the terms of the sequence, GOODSTEIN'S THEOREM states that $G_k(n)$ is 0 for any n and any sufficiently large k.

see also GOODSTEIN'S THEOREM, HEREDITARY REPRESENTATION

Goodstein's Theorem

For all n, there exists a k such that the kth term of the GOODSTEIN SEQUENCE $G_k(n) = 0$. In other words, every GOODSTEIN SEQUENCE converges to 0.

The secret underlying Goodstein's theorem is that the HEREDITARY REPRESENTATION of n in base b mimics an ordinal notation for ordinals less than some number. For such ordinals, the base bumping operation leaves the ordinal fixed whereas the subtraction of one decreases the ordinal. But these ordinals are well-ordered, and this allows us to conclude that a Goodstein sequence eventually converges to zero.

Goodstein's theorem cannot be proved in PEANO ARITHMETIC (i.e., formal NUMBER THEORY).

see also NATURAL INDEPENDENCE PHENOMENON, PEANO ARITHMETIC

Googol

A LARGE NUMBER equal to 10^{100}, or

$$\begin{gathered}10000000000000000000000000\\0000000000000000000000000\\0000000000000000000000000\\0000000000000000000000000.\end{gathered}$$

see also GOOGOLPLEX, LARGE NUMBER

References

Kasner, E. and Newman, J. R. *Mathematics and the Imagination.* Redmond, WA: Tempus Books, pp. 20–27, 1989.

Pappas, T. "Googol & Googolplex." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 76, 1989.

Googolplex

A LARGE NUMBER equal to $10^{10^{100}}$.

see also GOOGOL, LARGE NUMBER

References

Kasner, E. and Newman, J. R. *Mathematics and the Imagination.* Redmond, WA: Tempus Books, pp. 23–27, 1989.

Pappas, T. "Googol & Googolplex." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 76, 1989.

Gordon Function

Another name for the CONFLUENT HYPERGEOMETRIC FUNCTION OF THE SECOND KIND, defined by

$$G(a|c|z) = e^{i\pi a}\frac{\Gamma(c)}{\Gamma(a)}\left\{\frac{\Gamma(1-c)}{\Gamma(1-a)}\left[e^{-\pi c}+\frac{\sin[\pi(a-c)]}{\sin(\pi a)}\right]\right.$$
$$\left.\times {}_1F_1(a;c;z) - 2\frac{\Gamma(c-1)}{\Gamma(c-a)}z^{1-c}{}_1F_1(a-c+1;2-c;z)\right\},$$

where $\Gamma(x)$ is the GAMMA FUNCTION and ${}_1F_1(a;b;z)$ is the CONFLUENT HYPERGEOMETRIC FUNCTION OF THE FIRST KIND.

see also CONFLUENT HYPERGEOMETRIC FUNCTION OF THE SECOND KIND

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 671–672, 1953.

Gorenstein Ring

An algebraic RING which appears in treatments of duality in ALGEBRAIC GEOMETRY. Let A be a local ARTINIAN RING with $m \subset A$ its maximal IDEAL. Then A is a Gorenstein ring if the ANNIHILATOR of m has DIMENSION 1 as a VECTOR SPACE over $K = A/m$.

see also CAYLEY-BACHARACH THEOREM

References

Eisenbud, D.; Green, M.; and Harris, J. "Cayley-Bacharach Theorems and Conjectures." *Bull. Amer. Math. Soc.* **33**, 295–324, 1996.

Gosper's Algorithm

An ALGORITHM for finding closed form HYPERGEOMETRIC IDENTITIES The algorithm treats sums whose successive terms have ratios which are RATIONAL FUNCTIONS. Not only does it decide conclusively whether there exists a hypergeometric sequence z_n such that

$$t_n = z_{n+1} - z_n,$$

but actually produces z_n if it exists. If not, it produces $\sum_{k=0}^{n-1} t_n$. An outline of the algorithm follows (Petkovšek 1996):

1. For the ratio $r(n) = t_{n+1}/t_n$ which is a RATIONAL FUNCTION of n.
2. Write
$$r(n) = \frac{a(n)}{b(n)}\frac{c(n+1)}{c(n)},$$
where $a(n)$, $b(n)$, and $c(n)$ are polynomials satisfying
$$\mathrm{GCD}(a(n), b(n+h)) = 1$$
for all nonnegative integers h.
3. Find a nonzero polynomial solution $x(n)$ of
$$a(n)x(n+1) - b(n-1)x(n) = c(n),$$
if one exists.
4. Return $b(n-1)x(n)/c(n)t_n$ and stop.

Petkovšek *et al.* (1996) describe the algorithm as "one of the landmarks in the history of computerization of the problem of closed form summation." Gosper's algorithm is vital in the operation of ZEILBERGER'S ALGORITHM and the machinery of WILF-ZEILBERGER PAIRS.

see also HYPERGEOMETRIC IDENTITY, SISTER CELINE'S METHOD, WILF-ZEILBERGER PAIR, ZEILBERGER'S ALGORITHM

References

Gessel, I. and Stanton, D. "Strange Evaluations of Hypergeometric Series." *SIAM J. Math. Anal.* **13**, 295–308, 1982.

Gosper, R. W. "Decision Procedure for Indefinite Hypergeometric Summation." *Proc. Nat. Acad. Sci. USA* **75**, 40–42, 1978.

Graham, R. L.; Knuth, D. E.; and Patashnik, O. *Concrete Mathematics: A Foundation for Computer Science, 2nd ed.* Reading, MA: Addison-Wesley, 1994.

Lafron, J. C. "Summation in Finite Terms." In *Computer Algebra Symbolic and Algebraic Computation, 2nd ed.* (Ed. B. Buchberger, G. E. Collins, and R. Loos). New York: Springer-Verlag, 1983.
Paule, P. and Schorn, M. "A Mathematica Version of Zeilberger's Algorithm for Proving Binomial Coefficient Identities." *J. Symb. Comput.* **20**, 673–698, 1995.
Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. "Gosper's Algorithm." Ch. 5 in *A=B*. Wellesley, MA: A. K. Peters, pp. 73–99, 1996.
Zeilberger, D. "The Method of Creative Telescoping." *J. Symb. Comput.* **11**, 195–204, 1991.

Gosper Island

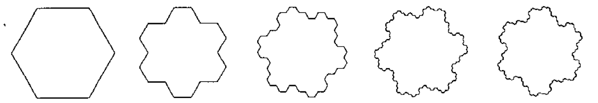

A modification of the KOCH SNOWFLAKE which has FRACTAL DIMENSION

$$D = \frac{2\ln 3}{\ln 7} = 1.12915\ldots.$$

The term "Gosper island" was used by Mandelbrot (1977) because this curve bounds the space filled by the PEANO-GOSPER CURVE; Gosper and Gardner use the term FLOWSNAKE FRACTAL instead. Gosper islands can TILE the PLANE.

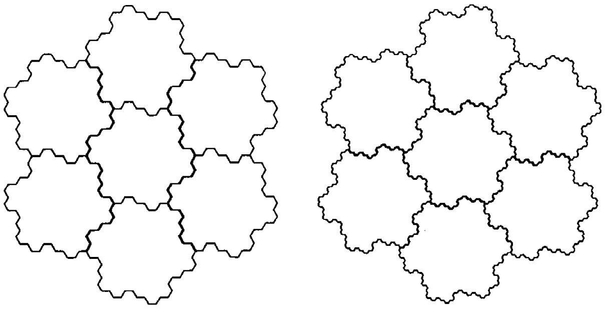

see also KOCH SNOWFLAKE, PEANO-GOSPER CURVE

References
Mandelbrot, B. B. *Fractals: Form, Chance, & Dimension.* San Francisco, CA: W. H. Freeman, Plate 46, 1977.

Gosper's Method

see GOSPER'S ALGORITHM

Graceful Graph

A LABELLED GRAPH which can be "gracefully numbered" is called a graceful graph. Label the nodes with distinct NONNEGATIVE INTEGERS. Then label the EDGES with the absolute differences between node values. If the EDGE numbers then run from 1 to e, the graph is gracefully numbered. In order for a graph to be graceful, it must be without loops or multiple EDGES.

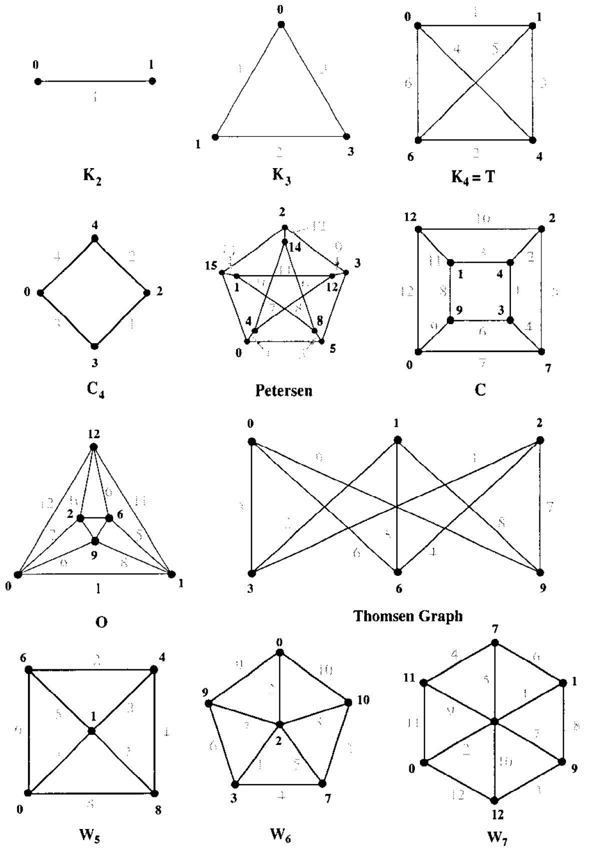

Golomb showed that the number of EDGES connecting the EVEN-numbered and ODD-numbered sets of nodes is $\lfloor (e+1)/2 \rfloor$, where e is the number of EDGES. In addition, if the nodes of a graph are all of EVEN ORDER, then the graph is graceful only if $\lfloor (e+1)/2 \rfloor$ is EVEN. The only ungraceful simple graphs with ≤ 5 nodes are shown below.

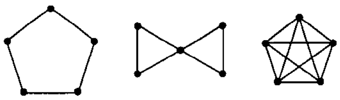

There are exactly $e!$ graceful graphs with e EDGES (Sheppard 1976), where $e!/2$ of these correspond to different labelings of the same graph. Golomb (1974) showed that all complete bipartite graphs are graceful. CATERPILLAR GRAPHS; COMPLETE GRAPHS K_2, K_3, $K_4 = W_4 = T$ (and only these; Golomb 1974); CYCLIC GRAPHS C_n when $n \equiv 0$ or 3 (mod 4), when the number of consecutive chords $k = 2$, 3, or $n-3$ (Koh and Punim 1982), or when they contain a P_k chord (Delorme *et al.* 1980, Koh and Yap 1985, Punnim and Pabhapote 1987); GEAR GRAPHS; PATH GRAPHS; the PETERSEN GRAPH; POLYHEDRAL GRAPHS $T = K_4 = W_4$, C, O, D, and I (Gardner 1983); STAR GRAPHS; the THOMSEN GRAPH (Gardner 1983); and WHEEL GRAPHS (Frucht 1988) are all graceful.

Some graceful graphs have only one numbering, but others have more than one. It is conjectured that all trees are graceful (Bondy and Murty 1976), but this has only been proved for trees with ≤ 16 Vertices. It has also been conjectured that all unicyclic graphs are graceful.

An excellent on-line resource is Brundage (`http://www.math.washington.edu/~brundage/oldgraceful/`).

see also Harmonious Graph, Labelled Graph

References

Abraham, J. and Kotzig, A. "All 2-Regular Graphs Consisting of 4-Cycles are Graceful." *Disc. Math.* **135**, 1–24, 1994.

Abraham, J. and Kotzig, A. "Extensions of Graceful Valuations of 2-Regular Graphs Consisting of 4-Gons." *Ars Combin.* **32**, 257–262, 1991.

Bloom, G. S. and Golomb, S. W. "Applications of Numbered Unidirected Graphs." *Proc. IEEE* **65**, 562–570, 1977.

Bolian, L. and Xiankun, Z. "On Harmonious Labellings of Graphs." *Ars Combin.* **36**, 315–326, 1993.

Brualdi, R. A. and McDougal, K. F. "Semibandwidth of Bipartite Graphs and Matrices." *Ars Combin.* **30**, 275–287, 1990.

Brundage, M. "Graceful Graphs." `http://www.math.washington.edu/~brundage/oldgraceful/`.

Cahit, I. "Are All Complete Binary Trees Graceful?" *Amer. Math. Monthly* **83**, 35–37, 1976.

Delorme, C.; Maheo, M.; Thuillier, H.; Koh, K. M.; and Teo, H. K. "Cycles with a Chord are Graceful." *J. Graph Theory* **4**, 409–415, 1980.

Frucht, R. W. and Gallian, J. A. "Labelling Prisms." *Ars Combin.* **26**, 69–82, 1988.

Gallian, J. A. "A Survey: Recent Results, Conjectures, and Open Problems in Labelling Graphs." *J. Graph Th.* **13**, 491–504, 1989.

Gallian, J. A. "Open Problems in Grid Labeling." *Amer. Math. Monthly* **97**, 133–135, 1990.

Gallian, J. A. "A Guide to the Graph Labelling Zoo." *Disc. Appl. Math.* **49**, 213–229, 1994.

Gallian, J. A.; Prout, J.; and Winters, S. "Graceful and Harmonious Labellings of Prism Related Graphs." *Ars Combin.* **34**, 213–222, 1992.

Gardner, M. "Golomb's Graceful Graphs." Ch. 15 in *Wheels, Life, and Other Mathematical Amusements.* New York: W. H. Freeman, pp. 152–165, 1983.

Golomb, S. W. "The Largest Graceful Subgraph of the Complete Graph." *Amer. Math. Monthly* **81**, 499–501, 1974.

Guy, R. "Monthly Research Problems, 1969–75." *Amer. Math. Monthly* **82**, 995–1004, 1975.

Guy, R. "Monthly Research Problems, 1969–1979." *Amer. Math. Monthly* **86**, 847–852, 1979.

Guy, R. K. "The Corresponding Modular Covering Problem. Harmonious Labelling of Graphs." §C13 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 127–128, 1994.

Huang, J. H. and Skiena, S. "Gracefully Labelling Prisms." *Ars Combin.* **38**, 225–242, 1994.

Koh, K. M. and Punnim, N. "On Graceful Graphs: Cycles with 3-Consecutive Chords." *Bull. Malaysian Math. Soc.* **5**, 49–64, 1982.

Jungreis, D. S. and Reid, M. "Labelling Grids." *Ars Combin.* **34**, 167–182, 1992.

Koh, K. M. and Yap, K. Y. "Graceful Numberings of Cycles with a P_3-Chord." *Bull. Inst. Math. Acad. Sinica* **13**, 41–48, 1985.

Morris, P. A. "On Graceful Trees." `http://www.math.washington.edu/~brundage/math/graceful/source/on_graceful_trees.ps`.

Moulton, D. "Graceful Labellings of Triangular Snakes." *Ars Combin.* **28**, 3–13, 1989.

Murty, U. S. R. and Bondy, J. A. *Graph Theory with Applications.* New York: North Holland, p. 248, 1976.

Punnim, N. and Pabhapote, N. "On Graceful Graphs: Cycles with a P_k-Chord, $k \geq 4$." *Ars Combin. A* **23**, 225–228, 1987.

Rosa, A. "On Certain Valuations of the Vertices of a Graph." In *Theory of Graphs, International Symposium, Rome, July 1966.* New York: Gordon and Breach, pp. 349–355, 1967.

Sheppard, D. A. "The Factorial Representation of Balanced Labelled Graphs." *Discr. Math.* **15**, 379–388, 1976.

Sierksma, G. and Hoogeveen, H. "Seven Criteria for Integer Sequences Being Graphic." *J. Graph Th.* **15**, 223–231, 1991.

Slater, P. J. "Note on k-Graceful, Locally Finite Graphs." *J. Combin. Th. Ser. B* **35**, 319–322, 1983.

Snevily, H. S. "New Families of Graphs That Have α-Labellings." Preprint.

Snevily, H. S. "Remarks on the Graceful Tree Conjecture." Preprint.

Xie, L. T. and Liu, G. Z. "A Survey of the Problem of Graceful Trees." *Qufu Shiyuan Xuebao* **1**, 8–15, 1984.

Graded Algebra

If A is a graded module and there Exists a degree-preserving linear map $\phi : A \otimes A \to A$, then (A, ϕ) is called a graded algebra.

References

Jacobson, N. *Lie Algebras.* New York: Dover, p. 163, 1979.

Gradian

A unit of angular measure in which the angle of an entire Circle is 400 gradians. A Right Angle is therefore 100 gradians.

see also Degree, Radian

Gradient

The gradient is a Vector operator denoted ∇ and sometimes also called Del or Nabla. It most often is applied to a real function of three variables $f(u_1, u_2, u_3)$, and may be denoted

$$\nabla f \equiv \mathrm{grad}(f). \tag{1}$$

For general Curvilinear Coordinates, the gradient is given by

$$\nabla \phi = \frac{1}{h_1}\frac{\partial \phi}{\partial u_1}\hat{\mathbf{u}}_1 + \frac{1}{h_2}\frac{\partial \phi}{\partial u_2}\hat{\mathbf{u}}_2 + \frac{1}{h_3}\frac{\partial \phi}{\partial u_3}\hat{\mathbf{u}}_3, \tag{2}$$

which simplifies to

$$\nabla \phi(x, y, z) = \frac{\partial \phi}{\partial x}\hat{\mathbf{x}} + \frac{\partial \phi}{\partial y}\hat{\mathbf{y}} + \frac{\partial \phi}{\partial z}\hat{\mathbf{z}} \tag{3}$$

in Cartesian Coordinates.

The direction of ∇f is the orientation in which the Directional Derivative has the largest value and $|\nabla f|$ is the value of that Directional Derivative. Furthermore, if $\nabla f \neq 0$, then the gradient is Perpendicular to the Level Curve through (x_0, y_0) if $z = f(x, y)$ and Perpendicular to the level surface through (x_0, y_0, z_0) if $F(x, y, z) = 0$.

In TENSOR notation, let

$$ds^2 = g_\mu \, dx_\mu{}^2 \tag{4}$$

be the LINE ELEMENT in principal form. Then

$$\nabla_{\vec{e}_\alpha} \vec{e}_\beta = \nabla_\alpha \vec{e}_\beta = \frac{1}{\sqrt{g_\alpha}} \frac{\partial}{\partial x_\alpha} \vec{e}_\beta. \tag{5}$$

For a MATRIX A,

$$\nabla |\mathsf{Ax}| = \frac{(\mathsf{Ax})^{\mathrm{T}} \mathsf{A}}{|\mathsf{Ax}|}. \tag{6}$$

For expressions giving the gradient in particular coordinate systems, see CURVILINEAR COORDINATES.

see also CONVECTIVE DERIVATIVE, CURL, DIVERGENCE, LAPLACIAN, VECTOR DERIVATIVE

References

Arfken, G. "Gradient, ∇" and "Successive Applications of ∇." §1.6 and 1.9 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 33–37 and 47–51, 1985.

Gradient Four-Vector

The 4-dimensional version of the GRADIENT, encountered frequently in general relativity and special relativity, is

$$\nabla_\mu = \begin{bmatrix} \frac{1}{c} \frac{\partial}{\partial t} \\ \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{bmatrix},$$

which can be written

$$(\nabla^\mu)^2 \equiv \Box^2,$$

where $\Box^2$ is the D'ALEMBERTIAN OPERATOR.

see also D'ALEMBERTIAN OPERATOR, GRADIENT, TENSOR, VECTOR

References

Morse, P. M. and Feshbach, H. "The Differential Operator ∇." §1.4 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 31–44, 1953.

Gradient Theorem

$$\int_b^a (\nabla f) \cdot d\mathbf{s} = f(b) - f(a),$$

where ∇ is the GRADIENT, and the integral is a LINE INTEGRAL. It is this relationship which makes the definition of a scalar potential function f so useful in gravitation and electromagnetism as a concise way to encode information about a VECTOR FIELD.

see also DIVERGENCE THEOREM, GREEN'S THEOREM, LINE INTEGRAL

Graeco-Latin Square

see EULER SQUARE

Graeco-Roman Square

see EULER SQUARE

Graeffe's Method

A ROOT-finding method which proceeds by multiplying a POLYNOMIAL $f(x)$ by $f(-x)$ and noting that

$$f(x) = (x - a_1)(x - a_2) \cdots (x - a_n) \tag{1}$$

$$f(-x) = (-1)^n (x + a_1)(x + a_2) \cdots (x + a_n) \tag{2}$$

so the result is

$$f(x) f(-x) = (-1)^n (x^2 - a_1{}^2)(x^2 - a_2{}^2) \cdots (x^2 - a_n{}^2). \tag{3}$$

Repeat ν times, then write this in the form

$$y^n + b_1 y^{n-1} + \ldots + b_n = 0 \tag{4}$$

where $y \equiv x^{2\nu}$. Since the coefficients are given by NEWTON'S RELATIONS

$$b_1 = -(y_1 + y_2 + \ldots + y_n) \tag{5}$$

$$b_2 = (y_1 y_2 + y_1 y_3 + \ldots + y_{n-1} y_n) \tag{6}$$

$$b_n = (-1)^n y_1 y_2 \cdots y_n, \tag{7}$$

and since the squaring procedure has separated the roots, the first term is larger than rest. Therefore,

$$b_1 \approx -y_1 \tag{8}$$

$$b_2 \approx y_1 y_2 \tag{9}$$

$$b_n \approx (-1)^n y_1 y_2 \cdots y_n, \tag{10}$$

giving

$$y_1 \approx -b_1 \tag{11}$$

$$y_2 \approx -\frac{b_2}{b_1} \tag{12}$$

$$y_n \approx -\frac{b_n}{b_{n-1}}. \tag{13}$$

Solving for the original roots gives

$$a_1 \approx \sqrt[2\nu]{-b_1} \tag{14}$$

$$a_2 \approx \sqrt[2\nu]{-\frac{b_2}{b_1}} \tag{15}$$

$$a_n \approx \sqrt[2\nu]{-\frac{b_n}{b_{n-1}}}. \tag{16}$$

This method works especially well if all roots are real.

References

von Kármán, T. and Biot, M. A. "Squaring the Roots (Graeffe's Method)." §5.8.c in *Mathematical Methods in Engineering: An Introduction to the Mathematical Treatment of Engineering Problems.* New York: McGraw-Hill, pp. 194–196, 1940.

Graham's Biggest Little Hexagon

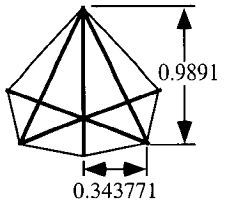

The largest possible (not necessarily regular) HEXAGON for which no two of the corners are more than unit distance apart. In the above figure, the heavy lines are all of unit length. The AREA of the hexagon is $A = 0.674981\ldots$, where A is a ROOT of

$$\begin{aligned}&4096A^{10} - 8192A^9 - 3008A^8 - 30,848A^7 + 21,056A^6\\ &\quad +146,496A^5 - 221,360A^4 + 1232A^3 + 144,464A^2\\ &\qquad -78,488A + 11,993 = 0.\end{aligned}$$

see also CALABI'S TRIANGLE

References
Conway, J. H. and Guy, R. K. "Graham's Biggest Little Hexagon." In *The Book of Numbers.* New York: Springer-Verlag, pp. 206–207, 1996.
Graham, R. L. "The Largest Small Hexagon." *J. Combin. Th. Ser. A* **18**, 165–170, 1975.

Graham's Number

The smallest dimension of a HYPERCUBE such that if the lines joining all pairs of corners are two-colored, a PLANAR COMPLETE GRAPH K_4 of one color will be forced. That an answer exists was proved by R. L. Graham and B. L. Rothschild. The actual answer is believed to be 6, but the best bound proved is

$$64\left\{\begin{array}{c}\underbrace{3\uparrow\uparrow\uparrow\uparrow 3}\\ \underbrace{3\uparrow 3}\\ \vdots\\ \underbrace{\qquad}\\ 3\uparrow 3,\end{array}\right.$$

where $\uparrow$ is stacked ARROW NOTATION. It is less than $3 \to 3 \to 3 \to 3$, where CHAINED ARROW NOTATION has been used.

see also ARROW NOTATION, CHAINED ARROW NOTATION, SKEWES NUMBER

References
Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 61–62, 1996.
Gardner, M. "Mathematical Games." *Sci. Amer.* **237**, 18–28, Nov. 1977.

Gram-Charlier Series

Approximates a distribution in terms of a NORMAL DISTRIBUTION. Let

$$\phi(t) \equiv \frac{1}{\sqrt{2\pi}} e^{-t^2/2},$$

then

$$f(t) = \phi(t) + \tfrac{1}{6}\gamma_1\phi^{(3)}(t) + \tfrac{1}{24}\gamma_2\phi^{(4)}(t) + \ldots.$$

see also EDGEWORTH SERIES

References
Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, pp. 107–108, 1951.

Gram Determinant

The DETERMINANT

$$G(f_1, f_2, \ldots, f_n) = \begin{vmatrix} \int f_1^2\,dt & \int f_1 f_2\,dt & \cdots & \int f_1 f_n\,dt\\ \int f_2 f_1\,dt & \int f_2^2\,dt & \cdots & \int f_2 f_n\,dt\\ \vdots & \vdots & \ddots & \vdots\\ \int f_1 f_n\,dt & \int f_1 f_n\,dt & \cdots & \int f_n^2\,dt \end{vmatrix}.$$

see also GRAM-SCHMIDT ORTHONORMALIZATION, WRONSKIAN

References
Sansone, G. *Orthogonal Functions, rev. English ed.* New York: Dover, p. 2, 1991.

Gram's Inequality

Let $f_1(x), \ldots, f_n(x)$ be REAL INTEGRABLE FUNCTIONS over the CLOSED INTERVAL $[a, b]$, then the DETERMINANT of their integrals satisfies

$$\begin{vmatrix} \int_a^b f_1{}^2(x)\,dx & \int_a^b f_1(x)f_2(x)\,dx & \cdots & \int_a^b f_1(x)f_n(x)\,dx\\ \int_a^b f_2(x)f_1(x)\,dx & \int_a^b f_2{}^2(x)\,dx & \cdots & \int_a^b f_2(x)f_n(x)\,dx\\ \vdots & \vdots & \ddots & \vdots\\ \int_a^b f_n(x)f_1(x)\,dx & \int_a^b f_n(x)f_2(x)\,dx & \cdots & \int_a^b f_n(x)f_n(x)\,dx \end{vmatrix} \geq 0.$$

see also GRAM-SCHMIDT ORTHONORMALIZATION

References
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1100, 1979.

Gram Matrix

Given m points with n-D vector coordinates $\mathbf{v}_i$, let M be the $n \times m$ matrix whose jth column consists of the coordinates of the vector $\mathbf{v}_j$, with $j = 1, \ldots, m$. Then define the $m \times m$ Gram matrix of dot products $a_{ij} = \mathbf{v}_i \cdot \mathbf{v}_j$ as

$$\mathsf{A} = \mathsf{M}^{\mathrm{T}}\mathsf{M},$$

where A^{T} denotes the TRANSPOSE. The Gram matrix determines the vectors $\mathbf{v}_i$ up to ISOMETRY.

Gram-Schmidt Orthonormalization

A procedure which takes a nonorthogonal set of LINEARLY INDEPENDENT functions and constructs an ORTHOGONAL BASIS over an arbitrary interval with respect to an arbitrary WEIGHTING FUNCTION $w(x)$. Given an original set of linearly independent functions $\{u_n\}$, let $\{\psi_n\}$ denote the orthogonalized (but not normalized) functions and $\{\phi_n\}$ the orthonormalized functions.

$$\psi_0(x) \equiv u_1(x) \tag{1}$$

$$\phi_0(x) \equiv \frac{\psi_0(x)}{\sqrt{\int \psi_0{}^2(x) w(x)\, dx}}. \tag{2}$$

Take

$$\psi_1(x) = u_1(x) + a_{10}\phi_0(x), \tag{3}$$

where we require

$$\int \psi_1 \phi_0 w\, dx = \int u_1 \phi_0 w\, dx + a_{10} \int \phi_0{}^2 w\, dx = 0. \tag{4}$$

By definition,

$$\int \phi_0{}^2 w\, dx = 1, \tag{5}$$

so

$$a_{10} = -\int u_1 \phi_0 w\, dx. \tag{6}$$

The first orthogonalized function is therefore

$$\psi_1 = u_1(x) - \left[\int u_1 \phi_0 w\, dx\right] \phi_0, \tag{7}$$

and the corresponding normalized function is

$$\phi_1 = \frac{\psi_1(x)}{\sqrt{\int \psi_1{}^2 w\, dx}}. \tag{8}$$

By mathematical induction, it follows that

$$\phi_i(x) = \frac{\psi_i(x)}{\sqrt{\int \psi_i{}^2 w\, dx}}, \tag{9}$$

where

$$\psi_i(x) = u_i + a_{i0}\phi_0 + a_{i1}\phi_1 \ldots + a_{i,i-1}\phi_{i-1} \tag{10}$$

and

$$a_{ij} \equiv -\int u_i \phi_j w\, dx. \tag{11}$$

If the functions are normalized to N_j instead of 1, then

$$\int_a^b [\phi_j(x)]^2 w\, dx = N_j{}^2 \tag{12}$$

$$\phi_i(x) = N_i \frac{\psi_i(x)}{\sqrt{\int \psi_i{}^2 w\, dx}} \tag{13}$$

$$a_{ij} = -\frac{\int u_i \phi_j w\, dx}{N_j{}^2}. \tag{14}$$

ORTHOGONAL POLYNOMIALS are especially easy to generate using GRAM-SCHMIDT ORTHONORMALIZATION. Use the notation

$$\langle x_i | x_j \rangle \equiv \langle x_i | w | x_j \rangle \equiv \int_a^b x_i(x) x_j(x) w(x)\, dx, \tag{15}$$

where $w(x)$ is a WEIGHTING FUNCTION, and define the first few POLYNOMIALS,

$$p_0(x) \equiv 1 \tag{16}$$

$$p_1(x) = \left[x - \frac{\langle x p_0 | p_0 \rangle}{\langle p_0 | p_0 \rangle}\right] p_0. \tag{17}$$

As defined, p_0 and p_1 are ORTHOGONAL POLYNOMIALS, as can be seen from

$$\begin{aligned} \langle p_0 | p_1 \rangle &= \left\langle \left[x - \frac{\langle x p_0 | p_0 \rangle}{\langle p_0 | p_0 \rangle}\right] p_0 \right\rangle \\ &= \langle x p_0 \rangle - \frac{\langle x p_0 | p_0 \rangle}{\langle p_0 | p_0 \rangle} \langle p_0 \rangle \\ &= \langle x p_0 \rangle - \langle x p_0 \rangle = 0. \end{aligned} \tag{18}$$

Now use the RECURRENCE RELATION

$$p_{i+1}(x) = \left[x - \frac{\langle x p_i | p_i \rangle}{\langle p_i | p_i \rangle}\right] p_i - \left[\frac{\langle p_i | p_i \rangle}{\langle p_{i-1} | p_{i-1} \rangle}\right] p_{i-1} \tag{19}$$

to construct all higher order POLYNOMIALS.

To verify that this procedure does indeed produce ORTHOGONAL POLYNOMIALS, examine

$$\begin{aligned} \langle p_{i+1} | p_i \rangle &= \left\langle \left[x - \frac{\langle x p_i | p_i \rangle}{\langle p_i | p_i \rangle}\right] p_i \middle| p_i \right\rangle \\ &\quad - \left\langle \frac{\langle p_i | p_i \rangle}{\langle p_{i-1} | p_{i-1} \rangle} p_{i-1} \middle| p_i \right\rangle \\ &= \langle x p_i | p_i \rangle - \frac{\langle x p_i | p_i \rangle}{\langle p_i | p_i \rangle} \langle p_i | p_i \rangle \\ &\quad - \frac{\langle p_i | p_i \rangle}{\langle p_{i-1} | p_{i-1} \rangle} \langle p_{i-1} | p_i \rangle \\ &= -\frac{\langle p_i | p_i \rangle}{\langle p_{i-1} | p_{i-1} \rangle} \langle p_{i-1} | p_i \rangle \\ &= -\frac{\langle p_i | p_i \rangle}{\langle p_{i-1} | p_{i-1} \rangle} \left[-\frac{\langle p_{i-1} | p_{j-1} \rangle}{\langle p_{j-2} | p_{j-2} \rangle} \langle p_{j-2} | p_{j-1} \rangle\right] \\ &= \ldots = (-1)^j \frac{\langle p_j | p_j \rangle}{\langle p_0 | p_0 \rangle} \langle p_0 | p_1 \rangle = 0, \end{aligned} \tag{20}$$

since $\langle p_0 | p_1 \rangle = 0$. Therefore, all the POLYNOMIALS $p_i(x)$ are orthogonal. Many common ORTHOGONAL POLYNOMIALS of mathematical physics can be generated in this manner. However, the process is numerically unstable (Golub and van Loan 1989).

see also GRAM DETERMINANT, GRAM'S INEQUALITY, ORTHOGONAL POLYNOMIALS

References
Arfken, G. "Gram-Schmidt Orthogonalization." §9.3 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 516–520, 1985.
Golub, G. H. and van Loan, C. F. *Matrix Computations, 3rd ed.* Baltimore, MD: Johns Hopkins, 1989.

Gram Series

$$R(x) = 1 + \sum_{k=1}^{\infty} \frac{(\ln x)^k}{k k! \zeta(k+1)},$$

where ζ is the RIEMANN ZETA FUNCTION. This approximation to the PRIME COUNTING FUNCTION is 10 times better than $\mathrm{Li}(x)$ for $x < 10^9$ but has been proven to be worse infinitely often by Littlewood (Ingham 1990). An equivalent formulation due to Ramanujan is

$$G(x) \equiv \frac{4}{\pi} \sum_{k=1}^{\infty} \frac{(-1)^{k-1} k}{B_{2k}(2k-1)} \left(\frac{\ln x}{2\pi}\right)^{2k-1} \sim \pi(x)$$

(Berndt 1994), where B_{2k} is a BERNOULLI NUMBER. The integral analog, also found by Ramanujan, is

$$J(x) \equiv \int_0^{\infty} \frac{(\ln x)^t \, dt}{t \Gamma(t+1) \zeta(t+1)} \sim \pi(x)$$

(Berndt 1994).

References
Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 124–129, 1994.
Gram, J. P. "Undersøgelser angaaende Maengden af Primtal under en given Graeense." *K. Videnskab. Selsk. Skr.* **2**, 183–308, 1884.
Ingham, A. E. Ch. 5 in *The Distribution of Prime Numbers.* New York: Cambridge, 1990.
Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 74, 1991.

Granny Knot

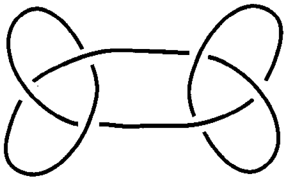

A COMPOSITE KNOT of seven crossings consisting of a KNOT SUM of TREFOILS. The granny knot has the same ALEXANDER POLYNOMIAL $(x^2 - x + 1)^2$ as the SQUARE KNOT.

Graph (Function)

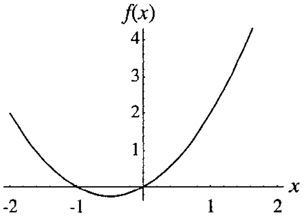

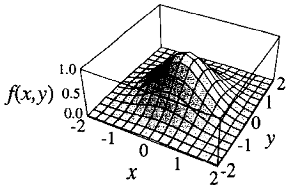

Technically, the graph of a function is its RANGE (a.k.a. image). Informally, given a FUNCTION $f(x_1, \ldots, x_n)$ defined on a DOMAIN U, the graph of f is defined as a CURVE or SURFACE showing the values taken by f over U (or some portion of U),

$$\operatorname{graph} f(x) \equiv \{(x, F(x)) \in \mathbb{R}^2 : x \in U\}$$

$$\operatorname{graph} f(x_1, \ldots, x_n) \equiv \{(x_1, \ldots, x_n, f(x_1, \ldots, x_n)) \in \mathbb{R}^{n+1} : (x_1, \ldots, x_n) \in U\}.$$

A graph is sometimes also called a PLOT.

Good routines for plotting graphs use adaptive algorithms which plot more points in regions where the function varies most rapidly (Wagon 1991, Math Works 1992, Heck 1993, Wickham-Jones 1994).

see also CURVE, EXTREMUM, GRAPH (GRAPH THEORY), HISTOGRAM, MAXIMUM, MINIMUM

References
Cleveland, W. S. *The Elements of Graphing Data, rev. ed.* Summit, NJ: Hobart, 1994.
Heck, A. *Introduction to Maple, 2nd ed.* New York: Springer-Verlag, pp. 303–304, 1993.
Math Works. *Matlab Reference Guide.* Natick, MA: The Math Works, p. 216, 1992.
Tufte, E. R. *The Visual Display of Quantitative Information.* Cheshire, CN: Graphics Press, 1983.
Tufte, E. R. *Envisioning Information.* Cheshire, CN: Graphics Press, 1990.
Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 24–25, 1991.
Wickham-Jones, T. *Computer Graphics with Mathematica.* Santa Clara, CA: TELOS, pp. 579–584, 1994.
Yates, R. C. "Sketching." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 188–205, 1952.

Graph (Graph Theory)

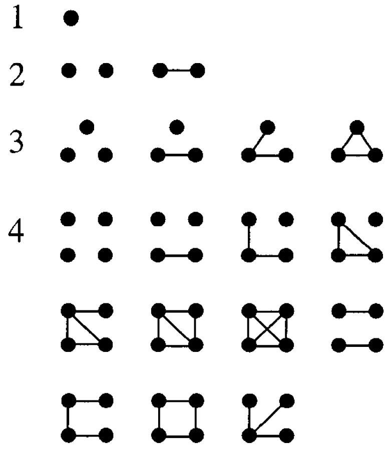

A mathematical object composed of points known as VERTICES or NODES and lines connecting some (possibly empty) SUBSET of them, known as EDGES. The study of graphs is known as GRAPH THEORY. Graphs are 1-D COMPLEXES, and there are always an EVEN number of ODD NODES in a graph. The number of nonisomorphic graphs with v NODES is given by the PÓLYA ENUMERATION THEOREM. The first few values for $n = 1, 2, \ldots,$ are 1, 2, 4, 11, 34, 156, 1044, ... (Sloane's A000088; see above figure).

Graph sums, differences, powers, and products can be defined, as can graph eigenvalues.

Before applying PÓLYA ENUMERATION THEOREM, define the quantity

$$h_{\mathbf{j}} = \frac{p!}{\prod_{i=1}^{p} i^{j_i} j_i!}, \tag{1}$$

where $p!$ is the FACTORIAL of p, and the related polynomial

$$Z_p(S) = \sum_i h_{\mathbf{j}_i} \prod_{k=1}^{p} f_k{}^{(\mathbf{j}_i)_k}, \tag{2}$$

where the $\mathbf{j}_i = (j_1, \ldots, j_p)_i$ are all of the p-VECTORS satisfying

$$j_1 + 2j_2 + 3j_3 + \ldots + pj_p = p. \tag{3}$$

For example, for $p = 3$, the three possible values of $\mathbf{j}$ are

$$\mathbf{j}_1 = (3,0,0), \text{ since } (1\cdot 3) + (2\cdot 0) + (3\cdot 0) = 3,$$

$$\text{giving } h_{\mathbf{j}_1} = \frac{3!}{(1^3 3!)(2^0 0!)(3^0 0!)} = 1 \tag{4}$$

$$\mathbf{j}_2 = (1,1,0), \text{ since } (1\cdot 1) + (2\cdot 1) + (3\cdot 0) = 3,$$

$$\text{giving } h_{\mathbf{j}_2} = \frac{3!}{(1^1 1!)(2^1 1!)(3^0 0!)} = 3, \tag{5}$$

$$\mathbf{j}_3 = (0,0,1), \text{ since } (1\cdot 0) + (2\cdot 0) + (3\cdot 1) = 3$$

$$\text{giving } h_{\mathbf{j}_3} = \frac{3!}{(1^0 0!)(2^0 0!)(3^1 1!)} = 2. \tag{6}$$

Therefore,

$$Z_3(S) = f_1{}^3 + 3f_1f_2 + 2f_3. \tag{7}$$

For small p, the first few values of $Z_p(S)$ are given by

$$Z_2(S) = f_1{}^2 + f_2 \tag{8}$$

$$Z_3(S) = f_1{}^3 + 3f_1f_2 + 2f_3 \tag{9}$$

$$Z_4(S) = f_1{}^4 + 6f_1{}^2f_2 + 3f_2{}^2 + 8f_1f_3 + 6f_4 \tag{10}$$

$$\begin{aligned} Z_5(S) = {} & f_1{}^5 + 10f_1{}^3f_2 + 15f_1f_2{}^2 + 20f_1{}^2f_3 \\ & + 20f_2f_3 + 30f_1f_4 + 24f_5 \end{aligned} \tag{11}$$

$$\begin{aligned} Z_6(S) = {} & f_1{}^6 + 15f_1{}^4f_2 + 45f_1{}^2f_2{}^2 + 15f_2{}^3 \\ & + 40f_1{}^3f_3 + 120f_1f_2f_3 + 40f_3{}^2 \\ & + 90f_1{}^2f_4 + 90f_2f_4 + 144f_1f_5 + 120f_6 \end{aligned} \tag{12}$$

$$\begin{aligned} Z_7(S) = {} & f_1{}^7 + 21f_1{}^5f_2 + 105f_1{}^3f_2{}^2 + 105f_1f_2{}^3 \\ & + 70f_1{}^4f_3 + 420f_1{}^2f_2f_3 + 210f_2{}^2f_3 \\ & + 280f_1f_3{}^2 + 210f_1{}^3f_4 + 630f_1f_2f_4 \\ & + 420f_3f_4 + 504f_1{}^2f_5 + 504f_2f_5 \\ & + 840f_1f_6 + 720f_7. \end{aligned} \tag{13}$$

Application of the PÓLYA ENUMERATION THEOREM then gives the formula

$$\begin{aligned} Z(R) = {} & \frac{1}{p!} \sum_{(j)} h_{\mathbf{j}} \prod_{n=0}^{\lfloor (p-1)/2 \rfloor} g_{2n+1}{}^{nj_{2n+1} + (2n+1)\binom{j_{2n+1}}{2}} \\ & \times \prod_{n=1}^{\lfloor p/2 \rfloor} [(g_n g_{2n})^{n-1}]^{j_{2n}} g_{2n}{}^{2n\binom{j_{2n}}{2}} \\ & \times \prod_{q=1}^{p} \prod_{r=q+1}^{p} g_{\mathrm{LCM}(q,r)}{}^{j_q j_r \mathrm{GCD}(q,r)}, \end{aligned} \tag{14}$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION, $\binom{n}{m}$ is a BINOMIAL COEFFICIENT, LCM is the LEAST COMMON MULTIPLE, GCD is the GREATEST COMMON DIVISOR, and the SUM (j) is over all $\mathbf{j}_i$ satisfying the sum identity described above. The first few generating functions $Z_p(R)$ are

$$Z_2(R) = 2g_1 \tag{15}$$

$$Z_3(R) = g_1{}^3 + 3g_1g_2 + 2g_3 \tag{16}$$

$$Z_4(R) = g_1{}^6 + 9g_1{}^2g_2{}^2 + 8g_3{}^2 + 6g_2g_4 \tag{17}$$

$$\begin{aligned} Z_5(R) = {} & g_1{}^{10} + 10g_1{}^4g_2{}^3 + 15g_1{}^2g_2{}^4 + 20g_1g_3{}^3 \\ & + 30g_2g_4{}^2 + 24g_5{}^2 + 20g_1g_3g_6 \end{aligned} \tag{18}$$

$$\begin{aligned} Z_6(R) = {} & g_1{}^{15} + 15g_1{}^7g_2{}^4 + 60g_1{}^3g_2{}^6 + 40g_1{}^3g_3{}^4 \\ & + 40g_3{}^5 + 180g_1g_2g_4{}^3 + 144g_5{}^3 \\ & + 120g_1g_2g_3{}^2g_6 + 120g_3g_6{}^2 \end{aligned}$$

$$Z_7(R) = g_1{}^{21} + 21g_1{}^{11}g_2{}^5 + 105g_1{}^5g_2{}^8 \tag{19}$$

$$\begin{aligned} & + 105g_1{}^3g_2{}^9 + 70g_1{}^6g_3{}^5 + 280g_3{}^7 \\ & + 210g_1{}^3g_2g_4{}^4 + 630g_1g_2{}^2g_4{}^4 \\ & + 504g_1g_5{}^4 + 420g_1{}^2g_2{}^2g_3{}^3g_6 \\ & + 210g_1{}^2g_2{}^2g_3g_6{}^2 + 840g_3g_6{}^3 + 720g_7{}^3 \\ & + 504g_1g_5{}^2g_{10} + 420g_2g_3g_4g_{12}. \end{aligned} \tag{20}$$

Letting $g_i = 1 + x^i$ then gives a POLYNOMIAL $S_i(x)$, which is a GENERATING FUNCTION for (i.e., the terms of x^i give) the number of graphs with i EDGES. The total number of graphs having i edges is $S_i(1)$. The first few $S_i(x)$ are

$$S_2 = 1 + x \tag{21}$$

$$S_3 = 1 + x + x^2 + x^3 \tag{22}$$

$$S_4 = 1 + x + 2x^2 + 3x^3 + 2x^4 + x^5 + x^6 \tag{23}$$

$$\begin{aligned} S_5 = {} & 1 + x + 2x^2 + 4x^3 + 6x^4 + 6x^5 + 6x^6 \\ & + 4x^7 + 2x^8 + x^9 + x^{10} \end{aligned} \tag{24}$$

$$\begin{aligned} S_6 = {} & 1 + x + 2x^2 + 5x^3 + 9x^4 + 15x^5 \\ & + 21x^6 + 24x^7 + 24x^8 + 21x^9 \\ & + 15x^{10} + 9x^{11} + 5x^{12} \\ & + 2x^{13} + x^{14} + x^{15} \end{aligned} \tag{25}$$

$$\begin{aligned} S_7 = {} & 1 + x + 2x^2 + 5x^3 + 10x^4 + 21x^5 \\ & + 21x^6 + 24x^7 + 41x^6 + 65x^7 + 97x^8 \end{aligned}$$

$$+131x^9+148x^{10}+148x^{11}$$
$$+131x^{12}+97x^{13}+65x^{14}+41x^{15}$$
$$+21x^{16}+10x^{17}+5x^{18}+2x^{19}+x^{20}+x^{21}, \tag{26}$$

giving the number of graphs with n nodes as 1, 2, 4, 11, 34, 156, 1044, ... (Sloane's A000088). King and Palmer (cited in Read 1981) have calculated S_n up to $n = 24$, for which

$$S_{24} = 195,704,906,302,078,447,922,174,862,416,\cdots$$
$$\cdots 726,256,004,122,075,267,063,365,754,368. \tag{27}$$

see also BIPARTITE GRAPH, CATERPILLAR GRAPH, CAYLEY GRAPH, CIRCULANT GRAPH, COCKTAIL PARTY GRAPH, COMPARABILITY GRAPH, COMPLEMENT GRAPH, COMPLETE GRAPH, CONE GRAPH, CONNECTED GRAPH, COXETER GRAPH, CUBICAL GRAPH, DE BRUIJN GRAPH, DIGRAPH, DIRECTED GRAPH, DODECAHEDRAL GRAPH, EULER GRAPH, EXTREMAL GRAPH, GEAR GRAPH, GRACEFUL GRAPH, GRAPH THEORY, HANOI GRAPH, HARARY GRAPH, HARMONIOUS GRAPH, HOFFMAN-SINGLETON GRAPH, ICOSAHEDRAL GRAPH, INTERVAL GRAPH, ISOMORPHIC GRAPHS, LABELLED GRAPH, LADDER GRAPH, LATTICE GRAPH, MATCHSTICK GRAPH, MINOR GRAPH, MOORE GRAPH, NULL GRAPH, OCTAHEDRAL GRAPH, PATH GRAPH, PETERSEN GRAPHS, PLANAR GRAPH, RANDOM GRAPH, REGULAR GRAPH, SEQUENTIAL GRAPH, SIMPLE GRAPH, STAR GRAPH, SUBGRAPH, SUPERGRAPH, SUPERREGULAR GRAPH, SYLVESTER GRAPH, TETRAHEDRAL GRAPH, THOMASSEN GRAPH, TOURNAMENT, TRIANGULAR GRAPH, TURAN GRAPH, TUTTE'S GRAPH, UNIVERSAL GRAPH, UTILITY GRAPH, WEB GRAPH, WHEEL GRAPH

References

Bogomolny, A. "Graph Puzzles." `http://www.cut-the-knot.com/do_you_know/graphs2.html`.

Fujii, J. N. *Puzzles and Graphs.* Washington, DC: National Council of Teachers, 1966.

Harary, F. "The Number of Linear, Directed, Rooted, and Connected Graphs." *Trans. Amer. Math. Soc.* **78**, 445–463, 1955.

Pappas, T. "Networks." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 126–127, 1989.

Read, R. "The Graph Theorists Who Count—and What They Count." In *The Mathematical Gardner* (Ed. D. Klarner). Boston, MA: Prindle, Weber, and Schmidt, pp. 326–345, 1981.

Sloane, N. J. A. Sequences A000088/M1253 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sloane, N. J. A. and Plouffe, S. Extended entry in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.

Graph Theory

The mathematical study of the properties of the formal mathematical structures called GRAPHS.

see also ADJACENCY MATRIX, ADJACENCY RELATION, ARTICULATION VERTEX, BLUE-EMPTY COLORING, BRIDGE (GRAPH), CHROMATIC NUMBER, CHROMATIC POLYNOMIAL, CIRCUIT RANK, CROSSING NUMBER (GRAPH), CYCLE (GRAPH), CYCLOMATIC NUMBER, DEGREE, DIAMETER (GRAPH), DIJKSTRA'S ALGORITHM, ECCENTRICITY, EDGE-COLORING, EDGE CONNECTIVITY, EULERIAN CIRCUIT, EULERIAN TRAIL, FACTOR (GRAPH), FLOYD'S ALGORITHM, GIRTH, GRAPH TWO-COLORING, GROUP THEORY, HAMILTONIAN CIRCUIT, HASSE DIAGRAM, HUB, INDEGREE, INTEGRAL DRAWING, ISTHMUS, JOIN (GRAPH), LOCAL DEGREE, MONOCHROMATIC FORCED TRIANGLE, OUTDEGREE, PARTY PROBLEM, PÓLYA ENUMERATION THEOREM, PÓLYA POLYNOMIAL, RADIUS (GRAPH), RAMSEY NUMBER, RE-ENTRANT CIRCUIT, SEPARATING EDGE, TAIT COLORING, TAIT CYCLE, TRAVELING SALESMAN PROBLEM, TREE, TUTTE'S THEOREM, UNICURSAL CIRCUIT, VALENCY, VERTEX COLORING, WALK

References

Berge, C. *The Theory of Graphs.* New York: Wiley, 1962.

Bogomolny, A. "Graphs." `http://www.cut-the-knot.com/do_you_know/graphs.html`.

Bollobás, B. *Graph Theory: An Introductory Course.* New York: Springer-Verlag, 1979.

Chartrand, G. *Introductory Graph Theory.* New York: Dover, 1985.

Foulds, L. R. *Graph Theory Applications.* New York: Springer-Verlag, 1992.

Chung, F. and Graham, R. *Erdős on Graphs: His Legacy of Unsolved Problems.* New York: A. K. Peters, 1998.

Grossman, I. and Magnus, W. *Groups and Their Graphs.* Washington, DC: Math. Assoc. Amer., 1965.

Harary, F. *Graph Theory.* Reading, MA: Addison-Wesley, 1994.

Hartsfield, N. and Ringel, G. *Pearls in Graph Theory: A Comprehensive Introduction, 2nd ed.* San Diego, CA: Academic Press, 1994.

Ore, Ø. *Graphs and Their Uses.* New York: Random House, 1963.

Ruskey, F. "Information on (Unlabelled) Graphs." `http://sue.csc.uvic.ca/~cos/inf/grap/GraphInfo.html`.

Saaty, T. L. and Kainen, P. C. *The Four-Color Problem: Assaults and Conquest.* New York: Dover, 1986.

Skiena, S. S. *Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica.* Redwood City, CA: Addison-Wesley, 1988.

Trudeau, R. J. *Introduction to Graph Theory.* New York: Dover, 1994.

Graph Two-Coloring

Assignment of each EDGE of a GRAPH to one of two color classes ("red" or "green").

see also BLUE-EMPTY GRAPH, MONOCHROMATIC FORCED TRIANGLE

Graphical Partition

A graphical partition of order n is the DEGREE SEQUENCE of a GRAPH with $n/2$ EDGES and no isolated VERTICES. For $n = 2, 4, 6, \ldots$, the number of graphical partitions is 1, 2, 5, 9, 17, ... (Sloane's A000569).

References

Barnes, T. M. and Savage, C. D. "A Recurrence for Counting Graphical Partitions." *Electronic J. Combinatorics* **2**, R11, 1–10, 1995. `http://www.combinatorics.org/Volume_2/volume2.html#R11`.

Barnes, T. M. and Savage, C. D. "Efficient Generation of Graphical Partitions." Submitted.

Ruskey, F. "Information on Graphical Partitions." `http://sue . csc . uvic . ca / ~ cos / inf / nump / GraphicalPartition.html`.

Sloane, N. J. A. Sequence A000569 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Grassmann Algebra

see EXTERIOR ALGEBRA

Grassmann Coordinates

An $(m+1)$-D SUBSPACE W of an $(n+1)$-D VECTOR SPACE V can be specified by an $(m+1) \times (n+1)$ MATRIX whose rows are the coordinates of a BASIS of W. The set of all $\binom{n+1}{m+1}$ $(m+1) \times (m+1)$ MINORS of this MATRIX are then called the Grassmann coordinates of w (where $\binom{a}{b}$ is a BINOMIAL COEFFICIENT).

see also CHOW COORDINATES

References

Wilson, W. S.; Chern, S. S.; Abhyankar, S. S.; Lang, S.; and Igusa, J.-I. "Wei-Liang Chow." *Not. Amer. Math. Soc.* **43**, 1117–1124, 1996.

Grassmann Manifold

A special case of a FLAG MANIFOLD. A Grassmann manifold is a certain collection of vector SUBSPACES of a VECTOR SPACE. In particular, $G_{n,k}$ is the Grassmann manifold of k-dimensional subspaces of the VECTOR SPACE $\mathbb{R}^n$. It has a natural MANIFOLD structure as an orbit-space of the STIEFEL MANIFOLD $V_{n,k}$ of orthonormal k-frames in $\mathbb{R}^n$. One of the main things about Grassmann manifolds is that they are classifying spaces for VECTOR BUNDLES.

Gray Code

An encoding of numbers so that adjacent numbers have a single DIGIT differing by 1. A BINARY Gray code with n DIGITS corresponds to a HAMILTONIAN PATH on an n-D HYPERCUBE (including direction reversals). The term Gray code is often used to refer to a "reflected" code, or more specifically still, the binary reflected Gray code.

To convert a BINARY number $d_1\,d_2 \cdots d_{n-1}\,d_n$ to its corresponding binary reflected Gray code, start at the right with the digit d_n (the nth, or last, DIGIT). If the d_{n-1} is 1, replace d_n by $1 - d_n$; otherwise, leave it unchanged. Then proceed to d_{n-1}. Continue up to the first DIGIT d_1, which is kept the same since d_0 is assumed to be a 0. The resulting number $g_1\,g_2 \cdots g_{n-1}\,g_n$ is the reflected binary Gray code.

To convert a binary reflected Gray code $g_1\,g_2 \cdots g_{n-1}\,g_n$ to a BINARY number, start again with the nth digit, and compute

$$\Sigma_n \equiv \sum_{i=1}^{n-1} g_i \pmod 2.$$

If Σ_n is 1, replace g_n by $1 - g_n$; otherwise, leave it the unchanged. Next compute

$$\Sigma_{n-1} \equiv \sum_{i=1}^{n-2} g_i \pmod 2,$$

and so on. The resulting number $d_1\,d_2 \cdots d_{n-1}\,d_n$ is the BINARY number corresponding to the initial binary reflected Gray code.

The code is called reflected because it can be generated in the following manner. Take the Gray code 0, 1. Write it forwards, then backwards: 0, 1, 1, 0. Then append 0s to the first half and 1s to the second half: 00, 01, 11, 10. Continuing, write 00, 01, 11, 10, 10, 11, 01, 00 to obtain: 000, 001, 011, 010, 110, 111, 101, 100, ... (Sloane's A014550). Each iteration therefore doubles the number of codes. The Gray codes corresponding to the first few nonnegative integers are given in the following table.

0	0	20	11110	40	111100
1	1	21	11111	41	111101
2	11	22	11101	42	111111
3	10	23	11100	43	111110
4	110	24	10100	44	111010
5	111	25	10101	45	111011
6	101	26	10111	46	111001
7	100	27	10110	47	111000
8	1100	28	10010	48	101000
9	1101	29	10011	49	101001
10	1111	30	10001	50	101011
11	1110	31	10000	51	101010
12	1010	32	110000	52	101110
13	1011	33	110001	53	101111
14	1001	34	110011	54	101101
15	1000	35	110010	55	101100
16	11000	36	110110	56	100100
17	11001	37	110111	57	100101
18	11011	38	110101	58	100111
19	11010	39	110100	59	100110

The binary reflected Gray code is closely related to the solution of the TOWERS OF HANOI as well as the BAGUENAUDIER.

see also BAGUENAUDIER, BINARY, HILBERT CURVE, MORSE-THUE SEQUENCE, RYSER FORMULA, TOWERS OF HANOI

References

Gardner, M. "The Binary Gray Code." Ch. 2 in *Knotted Doughnuts and Other Mathematical Entertainments.* New York: W. H. Freeman, 1986.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Gray Codes." §20.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 886–888, 1992.

Sloane, N. J. A. Sequence A014550 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Vardi, I. *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 111–112 and 246, 1991.

Great Circle

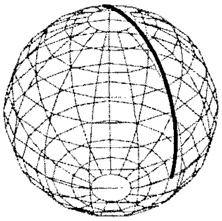

The shortest path between two points on a SPHERE, also known as an ORTHODROME. To find the great circle (GEODESIC) distance between two points located at LATITUDE δ and LONGITUDE λ of (δ_1, λ_1) and (δ_2, λ_2) on a SPHERE of RADIUS a, convert SPHERICAL COORDINATES to CARTESIAN COORDINATES using

$$\mathbf{r}_i = a \begin{bmatrix} \cos\lambda_i \cos\delta_i \\ \sin\lambda_i \cos\delta_i \\ \sin\delta_i \end{bmatrix}. \qquad (1)$$

(Note that the LATITUDE δ is related to the COLATITUDE ϕ of SPHERICAL COORDINATES by $\delta = 90^\circ - \phi$, so the conversion to CARTESIAN COORDINATES replaces $\sin\phi$ and $\cos\phi$ by $\cos\delta$ and $\sin\delta$, respectively.) Now find the ANGLE α between $\mathbf{r}_1$ and $\mathbf{r}_2$ using the DOT PRODUCT,

$$\begin{aligned} \cos\alpha &= \hat{\mathbf{r}}_1 \cdot \hat{\mathbf{r}}_2 \\ &= \cos\delta_1 \cos\delta_2 (\sin\lambda_1 \sin\lambda_2 + \cos\lambda_1 \cos\lambda_2) \\ &\quad + \sin\delta_1 \sin\delta_2 \\ &= \cos\delta_1 \cos\delta_2 \cos(\lambda_1 - \lambda_2) + \sin\delta_1 \sin\delta_2. \qquad (2) \end{aligned}$$

The great circle distance is then

$$d = a \cos^{-1}[\cos\delta_1 \cos\delta_2 \cos(\lambda_1 - \lambda_2) + \sin\delta_1 \sin\delta_2]. \quad (3)$$

For the Earth, the *equatorial* RADIUS is $a \approx 6378$ km, or 3963 (statute) miles. Unfortunately, the FLATTENING of the Earth cannot be taken into account in this simple derivation, since the problem is considerable more complicated for a SPHEROID or ELLIPSOID (each of which has a RADIUS which is a function of LATITUDE).

The equation of the great circle can be explicitly computed using the GEODESIC formalism. Writing

$$u = \lambda \qquad (4)$$

$$v = \delta = \tfrac{1}{2}\pi - \phi \qquad (5)$$

gives the P, Q, and R parameters of the GEODESIC (which are just combinations of the PARTIAL DERIVATIVES) as

$$P \equiv \left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2 + \left(\frac{\partial z}{\partial u}\right)^2 = a^2 \sin^2 v \qquad (6)$$

$$Q \equiv \frac{\partial x}{\partial u}\frac{\partial x}{\partial v} + \frac{\partial y}{\partial u}\frac{\partial y}{\partial v} + \frac{\partial z}{\partial u}\frac{\partial z}{\partial v} = 0 \qquad (7)$$

$$R \equiv \left(\frac{\partial x}{\partial v}\right)^2 + \left(\frac{\partial y}{\partial v}\right)^2 + \left(\frac{\partial z}{\partial v}\right)^2 = a^2. \qquad (8)$$

The GEODESIC differential equation then becomes

$$\cos v \sin^4 v + 2\cos v \sin^2 v v'^2 + \cos v v'^4 - \sin v v'' = 0. \quad (9)$$

However, because this is a special case of $Q = 0$ with P and R explicit functions of v only, the GEODESIC solution takes on the special form

$$\begin{aligned} v &= c_1 \int \sqrt{\frac{R}{P^2 - c_1{}^2 P}}\, dv = c_1 \int \frac{dv}{a^2 \sin^4 v - c_1{}^2 \sin^2 v} \\ &= \int \frac{dv}{\sin v \sqrt{\left(\frac{a}{c_1}\right)^2 \sin^2 v - 1}} \\ &= -\tan^{-1}\left[\frac{\cos v}{\sqrt{\left(\frac{a}{c_1}\right)^2 - 1}}\right] + c_2 \qquad (10) \end{aligned}$$

(Gradshteyn and Ryzhik 1979, p. 174, eqn. 2.599.6), which can be rewritten as

$$v = -\sin^{-1}\left(\frac{\cot v}{\sqrt{\left(\frac{a}{c_1}\right)^2 - 1}}\right) + c_2. \qquad (11)$$

It therefore follows that

$$(\sin c_2) a \sin v \cos u - (\cos c_2) a \sin v \sin u - \frac{a \cos v}{\sqrt{\left(\frac{a}{c_1}\right)^2 - 1}} = 0. \quad (12)$$

This equation can be written in terms of the CARTESIAN COORDINATES as

$$x \sin c_2 - y \cos c_2 - \frac{z}{\sqrt{\left(\frac{a}{c_1}\right)^2 - 1}} = 0, \qquad (13)$$

which is simply a PLANE passing through the center of the SPHERE and the two points on the surface of the SPHERE.

see also Geodesic, Great Sphere, Loxodrome, Mikusiński's Problem, Orthodrome, Point-Point Distance—2-D, Sphere

References
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, 1979.
Weinstock, R. *Calculus of Variations, with Applications to Physics and Engineering.* New York: Dover, pp. 26–28 and 62–63, 1974.

Great Cubicuboctahedron

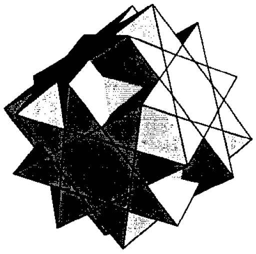

The Uniform Polyhedron U_{14} whose Dual Polyhedron is the Great Hexacronic Icositetrahedron. It has Wythoff Symbol $3\,4\,|\,\frac{4}{3}$. Its faces are $8\{3\}+6\{4\}+6\{\frac{8}{3}\}$. It is a Faceted version of the Cube. The Circumradius of a great cubicuboctahedron with unit edge length is

$$R = \tfrac{1}{2}\sqrt{5-2\sqrt{2}}.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 118–119, 1989.

Great Deltoidal Hexecontahedron

The Dual of the Great Rhombicosidodecahedron (Uniform).

Great Deltoidal Icositetrahedron

The Dual of the Great Rhombicuboctahedron (Uniform).

Great Dirhombicosidodecacron

The Dual of the Great Dirhombicosidodecahedron.

Great Dirhombicosidodecahedron

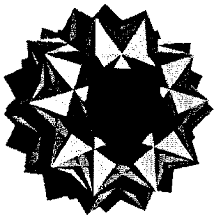

The Uniform Polyhedron U_{75} whose Dual is the Great Dirhombicosidodecacron. This Polyhedron is exceptional because it cannot be derived from Schwarz Triangles and because it is the only Uniform Polyhedron with more than six Polygons surrounding each Vertex (four Squares alternating with two Triangles and two Pentagrams). It has Wythoff Symbol $|\,\frac{3}{2}\,\frac{5}{3}\,3\,\frac{5}{2}$. Its faces are $40\{3\} + 60\{4\} + 24\{\frac{5}{2}\}$, and its Circumradius for unit edge length is

$$R = \tfrac{1}{2}\sqrt{2}.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 200–203, 1989.

Great Disdyakis Dodecahedron

The Dual of the Great Truncated Cuboctahedron.

Great Disdyakis Triacontahedron

The Dual of the Great Truncated Icosidodecahedron.

Great Ditrigonal Dodecacronic Hexecontahedron

The Dual of the Great Ditrigonal Dodecicosidodecahedron.

Great Ditrigonal Dodecicosidodecahedron

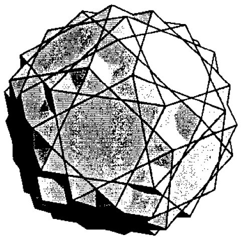

The Uniform Polyhedron U_{42} whose Dual is the Great Ditrigonal Dodecacronic Hexecontahedron. It has Wythoff Symbol $3\,5\,|\,\frac{5}{3}$. Its faces are $20\{3\}+12\{5\}+12\{\frac{10}{3}\}$, and its Circumradius for unit edge length is

$$R = \tfrac{1}{4}\sqrt{34-6\sqrt{5}}.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 125, 1989.

Great Ditrigonal Icosidodecahedron

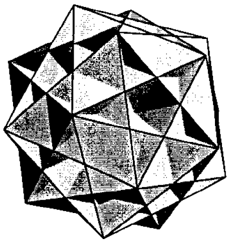

The UNIFORM POLYHEDRON U_{47} whose DUAL is the GREAT TRIAMBIC ICOSAHEDRON. It has WYTHOFF SYMBOL $\frac{3}{2}\,|\,3\,5$. Its faces are $20\{3\} + 12\{5\}$, and its CIRCUMRADIUS for unit edge length is

$$R = \tfrac{1}{2}\sqrt{3}.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 135–136, 1989.

Great Dodecacronic Hexecontahedron

The DUAL of the GREAT DODECICOSIDODECAHEDRON.

Great Dodecadodecahedron

see DODECADODECAHEDRON

Great Dodecahedron

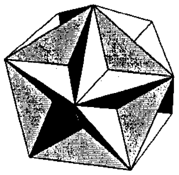

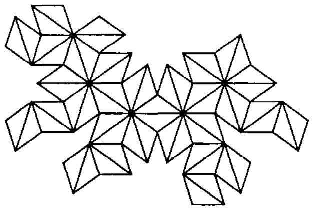

The UNIFORM POLYHEDRON U_{35} which is the DUAL of the SMALL STELLATED DODECAHEDRON and one of the KEPLER-POINSOT SOLIDS. Its faces are $12\{5\}$. Its SCHLÄFLI SYMBOL is $\{5, \frac{5}{2}\}$, and its WYTHOFF SYMBOL is $\frac{5}{2}\,|\,2\,5$. Its faces are $12\{5\}$. Its CIRCUMRADIUS for unit edge length is

$$R = \tfrac{1}{2}5^{1/4}\phi^{1/2}a = \tfrac{1}{4}5^{1/4}\sqrt{2(1+\sqrt{5})},$$

where ϕ is the GOLDEN RATIO.

see also GREAT ICOSAHEDRON, GREAT STELLATED DODECAHEDRON, KEPLER-POINSOT SOLID, SMALL STELLATED DODECAHEDRON

References
Fischer, G. (Ed.). Plate 105 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 104, 1986.

Great Dodecahedron-Small Stellated Dodecahedron Compound

A POLYHEDRON COMPOUND in which the GREAT DODECAHEDRON is interior to the SMALL STELLATED DODECAHEDRON.

see also POLYHEDRON COMPOUND

Great Dodecahemicosacron

The DUAL of the GREAT DODECAHEMICOSAHEDRON.

Great Dodecahemicosahedron

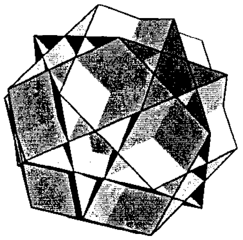

The UNIFORM POLYHEDRON U_{65} whose DUAL is the GREAT DODECAHEMICOSACRON. It has WYTHOFF SYMBOL $\frac{5}{3}\,\frac{5}{2}\,|\,\frac{5}{3}$. Its faces are $12\{\frac{5}{2}\} + 6\{\frac{10}{3}\}$. It is a FACETED DODECADODECAHEDRON. The CIRCUMRADIUS for unit edge length is

$$R = \tfrac{1}{2}\sqrt{3}.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 106–107, 1989.

Great Dodecahemidodecacron

The DUAL of the GREAT DODECAHEMIDODECAHEDRON.

Great Dodecahemidodecahedron

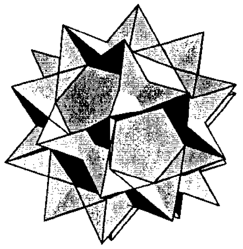

The UNIFORM POLYHEDRON U_{70} whose DUAL is the GREAT DODECAHEMIDODECACRON. It has WYTHOFF SYMBOL $\frac{5}{3}\,\frac{5}{2}\,|\,\frac{5}{3}$. Its faces are $12\{\frac{5}{2}\} + 6\{\frac{10}{3}\}$. Its CIRCUMRADIUS for unit edge length is

$$R = \phi^{-1},$$

where ϕ is the GOLDEN RATIO.

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 165, 1989.

Great Dodecicosacron

The DUAL of the GREAT DODECICOSAHEDRON.

Great Dodecicosahedron

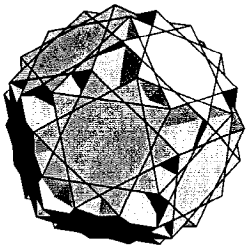

The UNIFORM POLYHEDRON U_{63} whose DUAL is the GREAT DODECICOSACRON. It has WYTHOFF SYMBOL $3\,\frac{5}{3}\mid \begin{array}{c}\frac{3}{2}\\ \frac{5}{2}\end{array}\Big|$. Its faces are $20\{6\}+12\{\frac{10}{3}\}$. Its CIRCUMRADIUS for unit edge length is

$$R=\tfrac{1}{4}\sqrt{34-6\sqrt{5}}.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 156–157, 1989.

Great Dodecicosidodecahedron

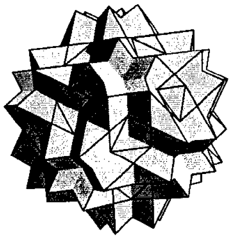

The UNIFORM POLYHEDRON U_{61} whose DUAL is the GREAT DODECACRONIC HEXECONTAHEDRON. Its WYTHOFF SYMBOL is $2\,\frac{5}{2}\mid 3$. Its faces are $20\{6\}+12\{\frac{5}{2}\}$, and its CIRCUMRADIUS for unit edge length is

$$R=\tfrac{1}{4}\sqrt{58-18\sqrt{5}}.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 148, 1989.

Great Hexacronic Icositetrahedron

The DUAL of the GREAT CUBICUBOCTAHEDRON.

Great Hexagonal Hexecontahedron

The DUAL of the GREAT SNUB DODECICOSIDODECAHEDRON.

Great Icosacronic Hexecontahedron

The DUAL of the GREAT ICOSICOSIDODECAHEDRON.

Great Icosahedron

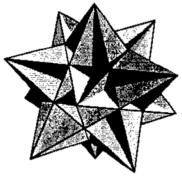

One of the KEPLER-POINSOT SOLIDS whose DUAL is the GREAT STELLATED DODECAHEDRON. Its faces are $20\{3\}$. It is also UNIFORM POLYHEDRON U_{53} and has WYTHOFF SYMBOL $3\,\frac{5}{2}\mid\frac{5}{3}$. Its faces are $20\{3\}+12\{\frac{5}{2}\}+12\{\frac{10}{3}\}$. Its CIRCUMRADIUS for unit edge length is

$$R=\tfrac{1}{2}\sqrt{11-4\sqrt{5}}.$$

see also GREAT DODECAHEDRON, GREAT ICOSAHEDRON, GREAT STELLATED DODECAHEDRON, KEPLER-POINSOT SOLID, SMALL STELLATED DODECAHEDRON, TRUNCATED GREAT ICOSAHEDRON

References

Fischer, G. (Ed.). Plate 106 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 105, 1986.
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 154, 1989.

Great Icosahedron-Great Stellated Dodecahedron Compound

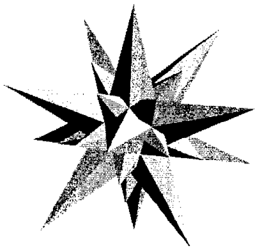

A POLYHEDRON COMPOUND most easily constructed by adding the VERTICES of a GREAT ICOSAHEDRON to a GREAT STELLATED DODECAHEDRON.

see also POLYHEDRON COMPOUND

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 132–133, 1989.

Great Icosicosidodecahedron

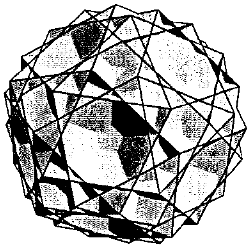

The UNIFORM POLYHEDRON U_{48} whose DUAL is the GREAT ICOSACRONIC HEXECONTAHEDRON. It has WYTHOFF SYMBOL $\frac{3}{2}\,5\,|\,3$. Its faces are $20\{3\}+20\{6\}+12\{5\}$. Its CIRCUMRADIUS for unit edge length is

$$R=\tfrac{1}{4}\sqrt{34-6\sqrt{5}}\,.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 137–139, 1989.

Great Icosidodecahedron

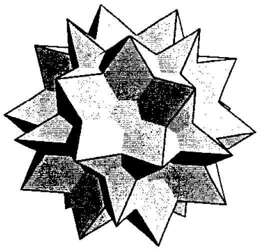

A UNIFORM POLYHEDRON U_{54} whose DUAL is the GREAT RHOMBIC TRIACONTAHEDRON (also called the GREAT STELLATED TRIACONTAHEDRON). It is a STELLATED ARCHIMEDEAN SOLID. It has SCHLÄFLI SYMBOL $\left\{\begin{matrix}3\\ \frac{5}{2}\end{matrix}\right\}$ and WYTHOFF SYMBOL $2\,|\,3\,\frac{5}{2}$. Its faces are $20\{3\}+12\{\frac{5}{2}\}$. Its CIRCUMRADIUS for unit edge length is

$$R=\phi^{-1},$$

where ϕ is the GOLDEN RATIO.

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 124, 1989.
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 147, 1989.

Great Icosihemidodecacron

The DUAL of the GREAT ICOSIHEMIDODECAHEDRON.

Great Icosihemidodecahedron

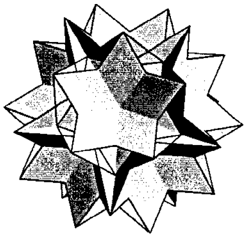

The UNIFORM POLYHEDRON U_{71} whose DUAL is the GREAT ICOSIHEMIDODECACRON. It has WYTHOFF SYMBOL $\frac{3}{2}\,3\,|\,\frac{5}{3}$. Its faces are $20\{3\}+6\{\frac{10}{3}\}$. For unit edge length, its CIRCUMRADIUS is

$$R=\phi^{-1},$$

where ϕ is the GOLDEN RATIO.

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 164, 1989.

Great Inverted Pentagonal Hexecontahedron

The DUAL of the GREAT INVERTED SNUB ICOSIDODECAHEDRON.

Great Inverted Retrosnub Icosidodecahedron

see GREAT RETROSNUB ICOSIDODECAHEDRON

Great Inverted Snub Icosidodecahedron

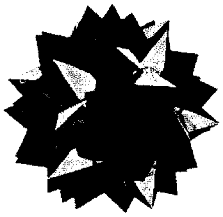

The UNIFORM POLYHEDRON U_{69} whose DUAL is the GREAT INVERTED PENTAGONAL HEXECONTAHEDRON. It has WYTHOFF SYMBOL $|\,2\,3\,\frac{5}{2}$. Its faces are $80\{3\}+12\{\frac{5}{2}\}$. For unit edge length, it has CIRCUMRADIUS

$$\begin{aligned}R&=\frac{1}{2}\sqrt{\frac{8\cdot 2^{2/3}-16x+2^{1/3}x^2}{8\cdot 2^{2/3}-10x+2^{1/3}x^2}}\\&=0.816080674799923,\end{aligned}$$

where

$$x\equiv\left(49-27\sqrt{5}+3\sqrt{6}\sqrt{93-49\sqrt{5}}\right)^{1/3}.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 179, 1989.

Great Pentagonal Hexecontahedron

The DUAL of the GREAT SNUB ICOSIDODECAHEDRON.

Great Pentagrammic Hexecontahedron

The DUAL of the GREAT RETROSNUB ICOSIDODECAHEDRON.

Great Pentakis Dodecahedron

The DUAL of the SMALL STELLATED TRUNCATED DODECAHEDRON.

Great Quasitruncated Icosidodecahedron

see GREAT TRUNCATED ICOSIDODECAHEDRON

Great Retrosnub Icosidodecahedron

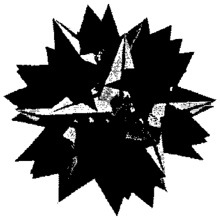

The Uniform Polyhedron U_{74}, also called the Great Inverted Retrosnub Icosidodecahedron, whose Dual is the Great Pentagrammic Hexecontahedron. It has Wythoff Symbol $|\,2\,\frac{3}{2}\,\frac{5}{3}$. Its faces are $80\{3\}+12\{\frac{5}{2}\}$. For unit edge length, it has Circumradius

$$R=\tfrac{1}{2}\sqrt{\frac{2-x}{1-x}}\approx 0.5800015,$$

where x is the smaller Negative root of

$$x^3+2x^2-\phi^{-2}=0,$$

with ϕ the Golden Mean.

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 189–193, 1989.

Great Rhombic Triacontahedron

A Zonohedron which is the Dual of the Great Icosidodecahedron. It is also called the Great Stellated Triacontahedron.

References
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 126, 1989.

Great Rhombicosidodecahedron (Archimedean)

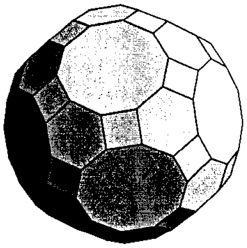

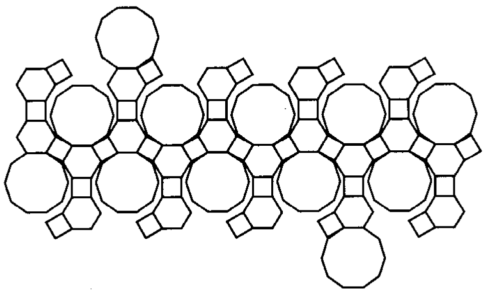

An Archimedean Solid also known as the Rhombitruncated Icosidodecahedron. It is sometimes improperly called the Truncated Icosidodecahedron, a name which is inappropriate since Truncation would yield Rectangular instead of Square. The great rhombicosidodecahedron is also Uniform Polyhedron U_{28}. Its Dual is the Disdyakis Triacontahedron, also called the Hexakis Icosahedron. It has Schläfli Symbol $t\left\{\begin{smallmatrix}3\\5\end{smallmatrix}\right\}$ and Wythoff Symbol $2\,3\,5\,|$. The Inradius, Midradius, and Circumradius for $a=1$ are

$$r=\tfrac{1}{241}(105+6\sqrt{5})\sqrt{31+12\sqrt{5}}\approx 3.73665$$

$$\rho=\tfrac{1}{2}\sqrt{30+12\sqrt{5}}\approx 3.76938$$

$$R=\tfrac{1}{2}\sqrt{31+12\sqrt{5}}\approx 3.80239.$$

see also Small Rhombicosidodecahedron

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 137, 1987.

Great Rhombicosidodecahedron (Uniform)

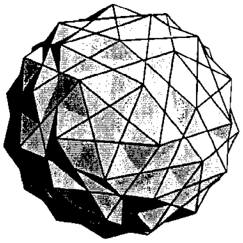

The Uniform Polyhedron U_{67}, also called the Quasirhombicosidodecahedron, whose Dual is the Great Deltoidal Hexecontahedron. It has Schläfli Symbol $r'\left\{\begin{matrix}3\\ \frac{5}{2}\end{matrix}\right\}$. It has Wythoff Symbol $3\,\frac{5}{3}\,|\,2$. Its faces are $20\{3\}+30\{4\}+12\{\frac{5}{2}\}$. For unit edge length, its Circumradius is

$$R=\tfrac{1}{2}\sqrt{11-4\sqrt{5}}.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 162–163, 1989.

Great Rhombicuboctahedron (Archimedean)

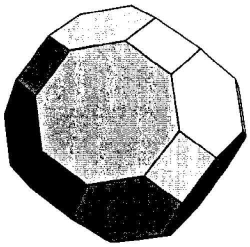

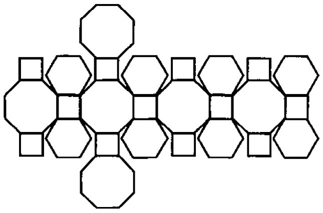

An ARCHIMEDEAN SOLID sometimes (improperly) called the TRUNCATED CUBOCTAHEDRON and also called the RHOMBITRUNCATED CUBOCTAHEDRON. Its DUAL is the DISDYAKIS DODECAHEDRON, also called the HEXAKIS OCTAHEDRON. It has SCHLÄFLI SYMBOL $t\left\{\begin{smallmatrix}3\\4\end{smallmatrix}\right\}$. It is also UNIFORM POLYHEDRON U_{11} and has WYTHOFF SYMBOL $2\,3\,4\,|$. Its faces are $8\{6\}+12\{4\}+6\{8\}$. The SMALL CUBICUBOCTAHEDRON is a FACETED version. The INRADIUS, MIDRADIUS, and CIRCUMRADIUS for unit edge length are

$$r = \tfrac{3}{97}(14+\sqrt{2})\sqrt{13+6\sqrt{2}} \approx 2.20974$$

$$\rho = \tfrac{1}{2}\sqrt{12+6\sqrt{2}} \approx 2.26303$$

$$R = \tfrac{1}{2}\sqrt{13+6\sqrt{2}} \approx 2.31761.$$

Additional quantities are

$$t = \tan(\tfrac{1}{8}\pi) = \sqrt{2} - 1$$

$$l = 2t = 2(\sqrt{2} - 1)$$

$$h = 1 + l\sin(\tfrac{1}{4}\pi) = 3 - \sqrt{2}.$$

see also SMALL RHOMBICUBOCTAHEDRON, GREAT TRUNCATED CUBOCTAHEDRON

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 138, 1987.

Great Rhombicuboctahedron (Uniform)

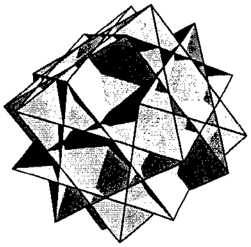

The UNIFORM POLYHEDRON U_{17}, also known as the QUASIRHOMBICUBOCTAHEDRON, whose DUAL is the GREAT DELTOIDAL ICOSITETRAHEDRON. It has SCHLÄFLI SYMBOL $r'\left\{\begin{smallmatrix}3\\4\end{smallmatrix}\right\}$ and WYTHOFF SYMBOL $\tfrac{3}{2}\,4\,|\,2$. Its faces are $8\{3\}+20\{4\}$. Its CIRCUMRADIUS for unit edge length is

$$R = \tfrac{1}{2}\sqrt{5-2\sqrt{2}}.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 132–133, 1989.

Great Rhombidodecacron

The DUAL of the GREAT RHOMBIDODECAHEDRON.

Great Rhombidodecahedron

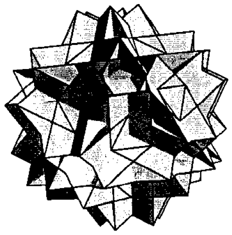

The UNIFORM POLYHEDRON U_{73} whose DUAL is the Great Rhombidodecacron. It WYTHOFF SYMBOL $2\,\tfrac{5}{3}\,|\,\begin{array}{c}\frac{3}{2}\\ \frac{5}{4}\end{array}\Big|$. Its faces are $30\{4\}+12\{\tfrac{10}{3}\}$. Its CIRCUMRADIUS for unit edge length is

$$R = \tfrac{1}{2}\sqrt{11-4\sqrt{5}}.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 168–170, 1989.

Great Rhombihexacron

The DUAL of the GREAT RHOMBIHEXAHEDRON.

Great Rhombihexahedron

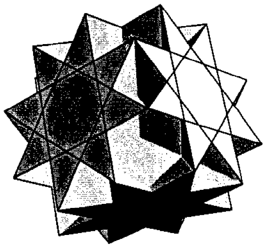

The UNIFORM POLYHEDRON U_{21} whose DUAL is the GREAT RHOMBIHEXACRON. It has WYTHOFF SYMBOL $2\,\tfrac{4}{3}\,\begin{array}{c}\frac{3}{2}\\ \frac{4}{2}\end{array}\Big|$. Its faces are $12\{4\}+6\{\tfrac{5}{3}\}$. Its CIRCUMRADIUS for unit edge length is

$$R = \tfrac{1}{2}\sqrt{5-2\sqrt{2}}.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 159–160, 1989.

Great Snub Dodecicosidodecahedron

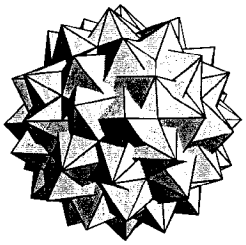

The UNIFORM POLYHEDRON U_{64} whose DUAL is the GREAT HEXAGONAL HEXECONTAHEDRON. It has WYTHOFF SYMBOL $|\,3\,\frac{5}{3}\,\frac{5}{2}$. Its faces are $80\{3\}+24\{\frac{5}{2}\}$. Its CIRCUMRADIUS for unit edge length is

$$R = \tfrac{1}{2}\sqrt{2}.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 183–185, 1989.

Great Snub Icosidodecahedron

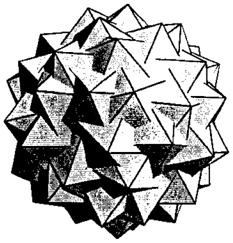

The UNIFORM POLYHEDRON U_{57} whose DUAL is the GREAT PENTAGONAL HEXECONTAHEDRON. It has WYTHOFF SYMBOL $|\,2\,3\,\frac{5}{3}$. Its faces are $80\{3\}+12\{\frac{5}{2}\}$. For unit edge length, it has CIRCUMRADIUS

$$R = \tfrac{1}{2}\sqrt{\frac{2-x}{1-x}} \approx 0.6450202,$$

where x is the most NEGATIVE ROOT of

$$x^3 + 2x^2 - \phi^{-2} = 0,$$

with ϕ the GOLDEN RATIO.

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 186–188, 1989.

Great Sphere

The great sphere on the surface of a HYPERSPHERE is the 3-D analog of the GREAT CIRCLE on the surface of a SPHERE. Let $2h$ be the number of reflecting SPHERES, and let great spheres divide a HYPERSPHERE into g 4-D TETRAHEDRA. Then for the POLYTOPE with SCHLÄFLI SYMBOL $\{p,q,r\}$,

$$\frac{64h}{g} = 12 - p - 2q - r + \frac{4}{p} + \frac{4}{r}.$$

see also GREAT CIRCLE

Great Stellapentakis Dodecahedron

The DUAL of the GREAT TRUNCATED ICOSAHEDRON.

Great Stellated Dodecahedron

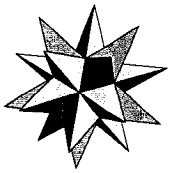

One of the KEPLER-POINSOT SOLIDS whose DUAL is the GREAT ICOSAHEDRON. Its SCHLÄFLI SYMBOL is $\{\frac{5}{2},3\}$. It is also UNIFORM POLYHEDRON U_{52} and has WYTHOFF SYMBOL $3\,|\,2\,\frac{5}{2}$. Its faces are $12\{\frac{5}{2}\}$. Its CIRCUMRADIUS for unit edge length is

$$R = \tfrac{1}{2}\sqrt{3}\phi^{-1} = \tfrac{1}{4}\sqrt{3}(\sqrt{5}-1).$$

The easiest way to construct it is to make 12 TRIANGULAR PYRAMIDS

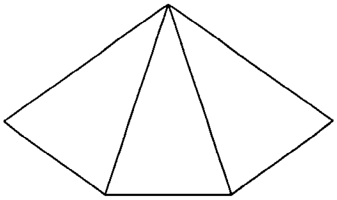

with side length $\phi = (1+\sqrt{5})/2$ (the GOLDEN RATIO) times the base and attach them to the sides of an ICOSAHEDRON.

see also GREAT DODECAHEDRON, GREAT ICOSAHEDRON, GREAT STELLATED TRUNCATED DODECAHEDRON, KEPLER-POINSOT SOLID, SMALL STELLATED DODECAHEDRON

References

Fischer, G. (Ed.). Plate 104 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 103, 1986.

Great Stellated Triacontahedron

see GREAT RHOMBIC TRIACONTAHEDRON

Great Stellated Truncated Dodecahedron

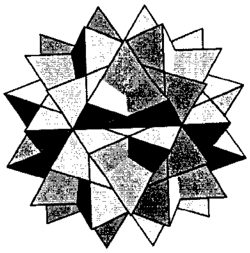

The UNIFORM POLYHEDRON U_{66}, also called the QUASITRUNCATED GREAT STELLATED DODECAHEDRON, whose DUAL is the GREAT TRIAKIS ICOSAHEDRON. It has SCHLÄFLI SYMBOL t'$\{\frac{5}{2},3\}$ and WYTHOFF SYMBOL

$2\,3\,|\,\frac{5}{3}$. Its faces are $20\{3\}+12\{\frac{10}{3}\}$. Its Circumradius for unit edge length is

$$R=\tfrac{1}{4}\sqrt{74-30\sqrt{5}}.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 161, 1989.

Great Triakis Icosahedron

The Dual of the Great Stellated Truncated Dodecahedron.

Great Triakis Octahedron

The Dual of the Stellated Truncated Hexahedron.

see also Small Triakis Octahedron

Great Triambic Icosahedron

The Dual of the Great Ditrigonal Icosidodecahedron.

Great Truncated Cuboctahedron

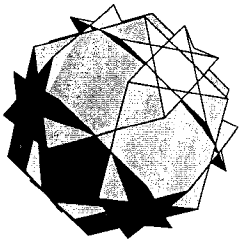

The Uniform Polyhedron U_{20}, also called the Quasitruncated Cuboctahedron, whose Dual is the Great Disdyakis Dodecahedron. It has Schläfli Symbol $t'\{\frac{3}{4}\}$ and Wythoff Symbol $2\,3\,\frac{4}{3}\,|$. Its Circumradius for unit edge length is

$$R=\tfrac{1}{2}\sqrt{13-6\sqrt{2}}.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 145–146, 1989.

Great Truncated Icosahedron

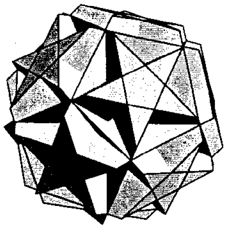

The Uniform Polyhedron U_{55}, also called the Truncated Great Icosahedron, whose Dual is the Great Stellapentakis Dodecahedron. It has Schläfli Symbol $t\{3,\frac{5}{2}\}$ and Wythoff Symbol $2\,\frac{5}{2}\,|\,3$. Its faces are $20\{6\}+12\{\frac{5}{2}\}$. Its Circumradius for unit edge length is

$$R=\tfrac{1}{4}\sqrt{58-18\sqrt{5}}.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 148, 1989.

Great Truncated Icosidodecahedron

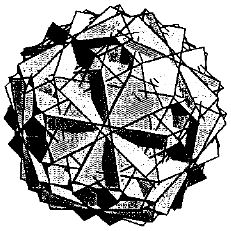

The Uniform Polyhedron U_{68}, also called the Great Quasitruncated Icosidodecahedron, whose Dual is the Great Disdyakis Triacontahedron. It has Schläfli Symbol $t'\left\{\begin{matrix}3\\ \frac{5}{2}\end{matrix}\right\}$ and Wythoff Symbol $2\,3\,\frac{5}{3}\,|$. Its faces are $20\{6\}+30\{4\}+12\{\frac{10}{3}\}$. Its Circumradius for unit edge length is

$$R=\tfrac{1}{2}\sqrt{31-12\sqrt{5}}.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 166–167, 1989.

Greater

A quantity a is said to be greater than b if a is larger than b, written $a>b$. If a is greater than or Equal to b, the relationship is written $a\geq b$. If a is Much Greater than b, this is written $a\gg b$. Statements involving greater than and Less than symbols are called Inequalities.

see also Equal, Greater Than/Less Than Symbol, Inequality, Less, Much Greater

Greater Than/Less Than Symbol

When applied to a system possessing a length R at which solutions in a variable r change character (such as the gravitational field of a sphere as r runs from the interior to the exterior), the symbols

$$r_{>}\equiv\max(r,R)$$

$$r_{<}\equiv\min(r,R)$$

are sometimes used.

see also Equal, Greater, Less

Greatest Common Denominator

see GREATEST COMMON DIVISOR

Greatest Common Divisor

The greatest common divisor of a and b GCD(a,b), sometimes written (a,b), is the largest DIVISOR common to a and b. Symbolically, let

$$a \equiv \prod_i {p_i}^{\alpha_i} \tag{1}$$

$$b \equiv \prod_i {p_i}^{\beta_i} \tag{2}$$

Then the greatest common divisor is given by

$$(a,b) = \prod_i {p_i}^{\min(\alpha_i,\beta_i)}, \tag{3}$$

where min denotes the MINIMUM. The GCD is DISTRIBUTIVE

$$(ma, mb) = m(a,b) \tag{4}$$

$$(ma, mb, mc) = m(a,b,c), \tag{5}$$

and ASSOCIATIVE

$$(a,b,c) = ((a,b),c) = (a,(b,c)) \tag{6}$$

$$(ab, cd) = (a,c)(b,d)\left(\frac{a}{(a,c)}, \frac{d}{(b,d)}\right)\left(\frac{c}{(a,c)}, \frac{b}{(b,d)}\right). \tag{7}$$

If $a = a_1(a,b)$ and $b = b_1(a,b)$, then

$$(a,b) = (a_1(a,b), b_1(a,b)) = (a,b)(a_1,b_1), \tag{8}$$

so $(a_1,b_1) = 1$ and a_1 and b_1 are said to be RELATIVELY PRIME. The GCD is also IDEMPOTENT

$$(a,a) = a, \tag{9}$$

COMMUTATIVE

$$(a,b) = (b,a), \tag{10}$$

and satisfies the ABSORPTION LAW

$$[a,(a,b)] = a. \tag{11}$$

The probability that two INTEGERS picked at random are RELATIVELY PRIME is $[\zeta(2)]^{-1} = 6/\pi^2$, where $\zeta(z)$ is the RIEMANN ZETA FUNCTION. Polezzi (1997) observed that $(m,n) = k$, where k is the number of LATTICE POINTS in the PLANE on the straight LINE connecting the VECTORS (0, 0) and (m,n) (excluding (m,n) itself). This observation is intimately connected with the probability of obtaining RELATIVELY PRIME integers, and also with the geometric interpretation of a REDUCED FRACTION y/x as a string through a LATTICE of points with ends at (1,0) and (x,y). The pegs it presses against (x_i,y_i) give alternate CONVERGENTS y_i/x_i of the CONTINUED FRACTION for y/x, while the other CONVERGENTS are obtained from the pegs it presses against with the initial end at (0, 1).

Knuth showed that

$$(2^p - 1, q^q - 1) = 2^{(p,q)} - 1 \tag{12}$$

for p, q PRIME.

The extended greatest common divisor of two INTEGERS m and n can be defined as the greatest common divisor of m and n which also satisfies the constraint $g = rm + sn$ for r and s given INTEGERS. It is used in solving LINEAR DIOPHANTINE EQUATIONS.

see also BEZOUT NUMBERS, EUCLIDEAN ALGORITHM, LEAST PRIME FACTOR

References

Polezzi, M. "A Geometrical Method for Finding an Explicit Formula for the Greatest Common Divisor." *Amer. Math. Monthly* **104**, 445–446, 1997.

Greatest Common Divisor Theorem

Given m and n, it is possible to choose c and d such that $cm + dn$ is a common factor of m and n.

Greatest Common Factor

see GREATEST COMMON DIVISOR

Greatest Integer Function

see FLOOR FUNCTION

Greatest Lower Bound

see INFIMUM, LEAST UPPER BOUND

Greatest Prime Factor

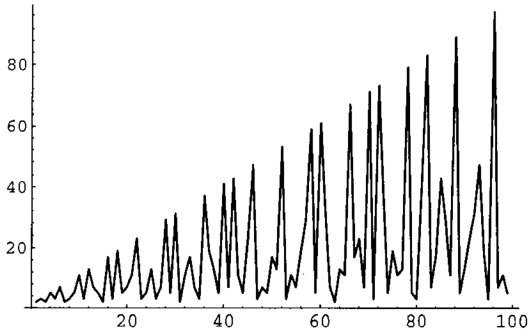

For an INTEGER $n \geq 2$, let gpf(x) denote the greatest prime factor of n, i.e., the number p_k in the factorization

$$n = {p_1}^{a_1} \cdots {p_k}^{a_k},$$

with $p_i < p_j$ for $i < j$. For $n = 2, 3, \ldots$, the first few are 2, 3, 2, 5, 3, 7, 2, 3, 5, 11, 3, 13, 7, 5, ... (Sloane's A006530). The greatest *multiple* prime factors for SQUAREFUL integers are 2, 2, 3, 2, 2, 3, 2, 2, 5, 3, 2, 2, 3, ... (Sloane's A046028).

see also DISTINCT PRIME FACTORS, FACTOR, LEAST COMMON MULTIPLE, LEAST PRIME FACTOR, MANGOLDT FUNCTION, PRIME FACTORS, TWIN PEAKS

References

Erdős, P. and Pomerance, C. "On the Largest Prime Factors of n and $n+1$." *Aequationes Math.* **17**, 211-321, 1978.

Guy, R. K. "The Largest Prime Factor of n." §B46 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 101, 1994.

Heath-Brown, D. R. "The Largest Prime Factor of the Integers in an Interval." *Sci. China Ser. A* **39**, 449–476, 1996.

Mahler, K. "On the Greatest Prime Factor of $ax^m + by^n$." *Nieuw Arch. Wiskunde* 1, 113–122, 1953.

Sloane, N. J. A. Sequence A006530/M0428 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Grebe Point

see LEMOINE POINT

Greedy Algorithm

An algorithm used to recursively construct a SET of objects from the smallest possible constituent parts.

Given a SET of k INTEGERS (a_1, a_2, ..., a_k) with $a_1 < a_2 < \ldots < a_k$, a greedy algorithm can be used to find a VECTOR of coefficients (c_1, c_2, ..., c_k) such that

$$\sum_{i=1}^{k} c_i a_i = \mathbf{c} \cdot \mathbf{a} = n, \tag{1}$$

where $\mathbf{c} \cdot \mathbf{a}$ is the DOT PRODUCT, for some given INTEGER n. This can be accomplished by letting $c_i = 0$ for $i = 1$, ..., $k-1$ and setting

$$c_k = \left\lfloor \frac{n}{a_k} \right\rfloor. \tag{2}$$

Now define the difference between the representation and n as

$$\Delta \equiv n - \mathbf{c} \cdot \mathbf{a}. \tag{3}$$

If $\Delta = 0$ at any step, a representation has been found. Otherwise, decrement the NONZERO a_i term with least i, set all $a_j = 0$ for $j < i$, and build up the remaining terms from

$$c_j = \left\lfloor \frac{\Delta_j}{a_k} \right\rfloor \tag{4}$$

for $j = i-1$, ..., 1 until $\Delta = 0$ or all possibilities have been exhausted.

For example, MCNUGGET NUMBERS are numbers which are representable using only $(a_1, a_2, a_3) = (6, 9, 20)$. Taking $n = 62$ and applying the algorithm iteratively gives the sequence (0, 0, 3), (0, 2, 2), (2, 1, 2), (3, 0, 2), (1, 4, 1), at which point $\Delta = 0$. 62 is therefore a MCNUGGET NUMBER with

$$62 = (1 \cdot 6) + (4 \cdot 9) + (1 \cdot 20). \tag{5}$$

If *any* INTEGER n can be represented with $c_i = 0$ or 1 using a sequence (a_1, a_2, ...), then this sequence is called a COMPLETE SEQUENCE.

A greedy algorithm can also be used to break down arbitrary fractions into UNIT FRACTIONS in a finite number of steps. For a FRACTION a/b, find the least INTEGER x_1 such that $1/x_1 \leq a/b$, i.e.,

$$x_1 = \frac{\lceil b \rceil}{a}, \tag{6}$$

where $\lceil x \rceil$ is the CEILING FUNCTION. Then find the least INTEGER x_2 such that $1/x_2 \leq a/b - 1/x_1$. Iterate until there is no remainder. The ALGORITHM gives two or fewer terms for $1/n$ and $2/n$, three or fewer terms for $3/n$, and four or fewer for $4/n$.

Paul Erdős and E. G. Strays have conjectured that the DIOPHANTINE EQUATION

$$\frac{4}{n} = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} \tag{7}$$

always can be solved, and W. Sierpiński conjectured that

$$\frac{5}{n} = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} \tag{8}$$

can be solved.

see also COMPLETE SEQUENCE, INTEGER RELATION, LEVINE-O'SULLIVAN GREEDY ALGORITHM, MCNUGGET NUMBER, REVERSE GREEDY ALGORITHM, SQUARE NUMBER, SYLVESTER'S SEQUENCE, UNIT FRACTION

References

Greek Cross

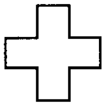

An irregular DODECAHEDRON CROSS in the shape of a PLUS SIGN.

see also CROSS, DISSECTION, DODECAHEDRON, LATIN CROSS, PLUS SIGN, SAINT ANDREW'S CROSS

Greek Problems

see GEOMETRIC PROBLEMS OF ANTIQUITY

Green's Function

Let

$$\tilde{L} = \tilde{D}^n + a_{n-1}(t)\tilde{D}^{n-1} + \ldots + a_1(t)\tilde{D} + a_0(t) \tag{1}$$

be a differential OPERATOR in 1-D, with $a_i(t)$ CONTINUOUS for $i = 0, 1, \ldots, n-1$ on the interval I, and assume we wish to find the solution $y(t)$ to the equation

$$\tilde{L}y(t) = h(t), \tag{2}$$

where $h(t)$ is a given CONTINUOUS on I. To solve equation (2), we look for a function $g : C^n(I) \mapsto C(I)$ such that $\tilde{L}(g(h)) = h$, where

$$y(t) = g(h(t)). \tag{3}$$

This is a CONVOLUTION equation of the form

$$y = g * h, \tag{4}$$

so the solution is

$$y(t) = \int_{t_0}^{t} g(t - x)h(x)\,dx, \tag{5}$$

where the function $g(t)$ is called the Green's function for $\tilde{L}$ on I.

Now, note that if we take $h(t) = \delta(t)$, then

$$y(t) = \int_{t_0}^{t} g(t - x)\delta(x)\,dx = g(t), \tag{6}$$

so the Green's function can be defined by

$$\tilde{L}g(t) = \delta(t). \tag{7}$$

However, the Green's function can be uniquely determined only if some initial or boundary conditions are given.

For an arbitrary linear differential operator $\tilde{L}$ in 3-D, the Green's function $G(\mathbf{r},\mathbf{r}')$ is defined by analogy with the 1-D case by

$$\tilde{L}G(\mathbf{r},\mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'). \tag{8}$$

The solution to $\tilde{L}\phi = f$ is then

$$\phi(\mathbf{r}) = \int G(\mathbf{r},\mathbf{r}')f(\mathbf{r}')\,d^3\mathbf{r}'. \tag{9}$$

Explicit expressions for $G(\mathbf{r},\mathbf{r}')$ can often be found in terms of a basis of given eigenfunctions $\phi_n(\mathbf{r}_1)$ by expanding the Green's function

$$G(\mathbf{r}_1,\mathbf{r}_2) = \sum_{n=0}^{\infty} a_n(\mathbf{r}_2)\phi_n(\mathbf{r}_1) \tag{10}$$

and DELTA FUNCTION,

$$\delta^3(\mathbf{r}_1 - \mathbf{r}_2) = \sum_{n=0}^{\infty} b_n\phi_n(\mathbf{r}_1). \tag{11}$$

Multiplying both sides by $\phi_m(\mathbf{r}_2)$ and integrating over $\mathbf{r}_1$ space,

$$\int \phi_m(\mathbf{r}_2)\delta^3(\mathbf{r}_1-\mathbf{r}_2)\,d^3\mathbf{r}_1 = \sum_{n=0}^{\infty} b_n \int \phi_m(\mathbf{r}_2)\phi_n(\mathbf{r}_1)\,d^3\mathbf{r}_1 \tag{12}$$

$$\phi_m(\mathbf{r}_2) = \sum_{n=0}^{\infty} b_n\delta_{nm} = b_m, \tag{13}$$

so

$$\delta^3(\mathbf{r}_1 - \mathbf{r}_2) = \sum_{n=0}^{\infty} \phi_n(\mathbf{r}_1)\phi_n(\mathbf{r}_2). \tag{14}$$

By plugging in the differential operator, solving for the a_ns, and substituting into G, the original nonhomogeneous equation then can be solved.

References

Arfken, G. "Nonhomogeneous Equation—Green's Function," "Green's Functions—One Dimension," and "Green's Functions—Two and Three Dimensions." §8.7 and §16.5–16.6 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 480–491 and 897–924, 1985.

Green's Function—Helmholtz Differential Equation

The inhomogeneous HELMHOLTZ DIFFERENTIAL EQUATION is

$$\nabla^2\psi(\mathbf{r}) + k^2\psi(\mathbf{r}) = \rho(\mathbf{r}), \tag{1}$$

where the Helmholtz operator is defined as $\tilde{L} \equiv \nabla^2 + k^2$. The Green's function is then defined by

$$(\nabla^2 + k^2)G(\mathbf{r}_1,\mathbf{r}_2) = \delta^3(\mathbf{r}_1 - \mathbf{r}_2). \tag{2}$$

Define the basis functions ϕ_n as the solutions to the homogeneous HELMHOLTZ DIFFERENTIAL EQUATION

$$\nabla^2\phi_n(\mathbf{r}) + {k_n}^2\phi_n(\mathbf{r}) = 0. \tag{3}$$

The Green's function can then be expanded in terms of the ϕ_ns,

$$G(\mathbf{r}_1,\mathbf{r}_2) = \sum_{n=0}^{\infty} a_n(\mathbf{r}_2)\phi_n(\mathbf{r}_1), \tag{4}$$

and the DELTA FUNCTION as

$$\delta^3(\mathbf{r}_1 - \mathbf{r}_2) = \sum_{n=0}^{\infty} \phi_n(\mathbf{r}_1)\phi_n(\mathbf{r}_2). \tag{5}$$

Plugging (4) and (5) into (2) gives

$$\nabla^2\left[\sum_{n=0}^{\infty} a_n(\mathbf{r}_2)\phi_n(\mathbf{r}_1)\right] + k^2\sum_{n=0}^{\infty} a_n(\mathbf{r}_2)\phi_n(\mathbf{r}_1) = \sum_{n=0}^{\infty} \phi_n(\mathbf{r}_1)\phi_n(\mathbf{r}_2). \tag{6}$$

Using (3) gives

$$-\sum_{n=0}^{\infty} a_n(\mathbf{r}_2){k_n}^2\phi_n(\mathbf{r}) + k^2\sum_{n=0}^{\infty} a_n(\mathbf{r}_2)\phi_n(\mathbf{r}_1) = \sum_{n=0}^{\infty} \phi_n(\mathbf{r}_1)\phi_n(\mathbf{r}_2) \tag{7}$$

$$\sum_{n=0}^{\infty} a_n(\mathbf{r}_2)\phi_n(\mathbf{r}_1)(k^2-{k_n}^2) = \sum_{n=0}^{\infty} \phi_n(\mathbf{r}_1)\phi_n(\mathbf{r}_2). \quad (8)$$

This equation must hold true for each n, so

$$a_n(\mathbf{r}_2)\phi_n(\mathbf{r}_1)(k^2-{k_n}^2) = \phi_n(\mathbf{r}_1)\phi_n(\mathbf{r}_2) \quad (9)$$

$$a_n(\mathbf{r}_2) = \frac{\phi_n(\mathbf{r}_2)}{k^2-{k_n}^2}, \quad (10)$$

and (4) can be written

$$G(\mathbf{r}_1,\mathbf{r}_2) = \sum_{n=0}^{\infty} \frac{\phi_n(\mathbf{r}_1)\phi_n(\mathbf{r}_2)}{k^2-{k_n}^2}. \quad (11)$$

The general solution to (1) is therefore

$$\begin{aligned}\psi(\mathbf{r}_1) &= \int G(\mathbf{r}_1,\mathbf{r}_2)\rho(\mathbf{r}_2)\,d^3\mathbf{r}_2 \\ &= \sum_{n=0}^{\infty} \int \frac{\phi_n(\mathbf{r}_1)\phi_n(\mathbf{r}_2)\rho(\mathbf{r}_2)}{k^2-{k_n}^2}\,d^3\mathbf{r}_2. \quad (12)\end{aligned}$$

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 529–530, 1985.

Green's Function—Poisson's Equation

Poisson's Equation equation is

$$\nabla^2\phi = 4\pi\rho, \quad (1)$$

where ϕ is often called a potential function and ρ a density function, so the differential operator in this case is $\tilde{L} = \nabla^2$. As usual, we are looking for a Green's function $G(\mathbf{r}_1,\mathbf{r}_2)$ such that

$$\nabla^2 G(\mathbf{r}_1,\mathbf{r}_2) = \delta^3(\mathbf{r}_1-\mathbf{r}_2). \quad (2)$$

But from Laplacian,

$$\nabla^2\left(\frac{1}{|\mathbf{r}-\mathbf{r}'|}\right) = -4\pi\delta^3(\mathbf{r}-\mathbf{r}'), \quad (3)$$

so

$$G(\mathbf{r},\mathbf{r}') = -\frac{1}{4\pi|\mathbf{r}-\mathbf{r}'|}, \quad (4)$$

and the solution is

$$\phi(\mathbf{r}) = \int G(\mathbf{r},\mathbf{r}')[4\pi\rho(\mathbf{r}')]\,d^3\mathbf{r}' = -\int \frac{\rho(\mathbf{r}')\,d^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}. \quad (5)$$

Expanding $G(\mathbf{r}_1,\mathbf{r}_2)$ in the Spherical Harmonics Y_l^m gives

$$\begin{aligned}&G(\mathbf{r}_1,\mathbf{r}_2) \\ &= \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \frac{1}{2l+1}\frac{r_<^l}{r_>^{l+1}} Y_l^m(\theta_1,\phi_1){Y_l^m}^*(\theta_2,\phi_2), \quad (6)\end{aligned}$$

where $r_<$ and $r_>$ are Greater Than/Less Than Symbols. This expression simplifies to

$$G(\mathbf{r}_1,\mathbf{r}_2) = \frac{1}{4\pi}\sum_{l=0}^{\infty}\frac{r_<^l}{r_>^{l+1}}P_l(\cos\gamma), \quad (7)$$

where P_l are Legendre Polynomials, and $\cos\gamma \equiv \mathbf{r}_1\cdot\mathbf{r}_2$. Equations (6) and (7) give the addition theorem for Legendre Polynomials.

In Cylindrical Coordinates, the Green's function is much more complicated,

$$\begin{aligned}&G(\mathbf{r}_1,\mathbf{r}_2) = \frac{1}{2\pi^2}\sum_{m=-\infty}^{\infty} \\ &\int_0^{\infty} I_m(k\rho_<)K_m(k\rho_>)e^{im(\phi_1-\phi_2)}\cos[k(z_1-z_2)]\,dk, \\ &\hfill (8)\end{aligned}$$

where $I_m(x)$ and $K_m(x)$ are Modified Bessel Functions of the First and Second Kinds (Arfken 1985).

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 485–486, 905, and 912, 1985.

Green's Identities

Green's identities are a set of three vector derivative/integral identities which can be derived starting with the vector derivative identities

$$\nabla\cdot(\psi\nabla\phi) = \psi\nabla^2\phi + (\nabla\psi)\cdot(\nabla\phi) \quad (1)$$

and

$$\nabla\cdot(\phi\nabla\psi) = \phi\nabla^2\psi + (\nabla\phi)\cdot(\nabla\psi), \quad (2)$$

where $\nabla\cdot$ is the Divergence, ∇ is the Gradient, ∇^2 is the Laplacian, and $\mathbf{a}\cdot\mathbf{b}$ is the Dot Product. From the Divergence Theorem,

$$\int_V (\nabla\cdot\mathbf{F})\,dV = \int_S \mathbf{F}\cdot d\mathbf{a}. \quad (3)$$

Plugging (2) into (3),

$$\int_S \phi(\nabla\psi)\cdot d\mathbf{a} = \int_V [\phi\nabla^2\psi + (\nabla\phi)\cdot(\nabla\psi)]\,dV. \quad (4)$$

This is Green's first identity.

Subtracting (2) from (1),

$$\nabla\cdot(\phi\nabla\psi - \psi\nabla\phi) = \phi\nabla^2\psi - \psi\nabla^2\phi. \quad (5)$$

Therefore,

$$\int_V (\phi\nabla^2\psi - \psi\nabla^2\phi)\,dV = \int_S (\phi\nabla\psi - \psi\nabla\phi)\cdot d\mathbf{a}. \quad (6)$$

This is Green's second identity.

Let u have continuous first PARTIAL DERIVATIVES and be HARMONIC inside the region of integration. Then Green's third identity is

$$u(x,y) = \frac{1}{2\pi} \oint_C \left[\ln\left(\frac{1}{r}\right) \frac{\partial u}{\partial n} - u \frac{\partial}{\partial n} \ln\left(\frac{1}{r}\right) \right] ds \qquad (7)$$

(Kaplan 1991, p. 361).

References

Kaplan, W. *Advanced Calculus, 4th ed.* Reading, MA: Addison-Wesley, 1991.

Greene's Method

A method for predicting the onset of widespread CHAOS. It is based on the hypothesis that the dissolution of an invariant torus can be associated with the sudden change from stability to instability of nearly closed orbits (Tabor 1989, p. 163).

see also OVERLAPPING RESONANCE METHOD

References

Tabor, M. *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, 1989.

Green Space

A G-SPACE provides local notions of harmonic, hyperharmonic, and superharmonic functions. When there exists a nonconstant superharmonic function greater than 0, it is a called a Green space. Examples are $\mathbb{R}^n$ (for $n \geq 3$) and any bounded domain of $\mathbb{R}^n$.

Green's Theorem

Green's theorem is a vector identity which is equivalent to the CURL THEOREM in the PLANE. Over a region D in the plane with boundary ∂D,

$$\int_{\partial D} f(x,y)\,dx + g(x,y)\,dy = \iint_D \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right) dx\,dy$$

$$\int_{\partial D} \mathbf{F} \cdot d\mathbf{s} = \iint_D (\nabla \times \mathbf{F}) \cdot \mathbf{k}\,dA.$$

If the region D is on the left when traveling around ∂D, then AREA of D can be computed using

$$A = \tfrac{1}{2} \int_{\partial D} x\,dy - y\,dx.$$

see also CURL THEOREM, DIVERGENCE THEOREM

References

Arfken, G. "Gauss's Theorem." §1.11 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 57–61, 1985.

Gregory's Formula

A series FORMULA for PI found by Gregory and Leibniz,

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} + \ldots.$$

It converges very slowly, but its convergence can be accelerated using certain transformations, in particular

$$\pi = \sum_{k=1}^{\infty} \frac{3^k - 1}{4^k} \zeta(k+1),$$

where $\zeta(z)$ is the RIEMANN ZETA FUNCTION (Vardi 1991).

see also MACHIN'S FORMULA, MACHIN-LIKE FORMULAS, PI

References

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 157–158, 1991.

Gregory Number

A number

$$t_x = \tan^{-1}(\tfrac{1}{x}) = \cot^{-1} x,$$

where x is an INTEGER or RATIONAL NUMBER, $\tan^{-1} x$ is the INVERSE TANGENT, and $\cot^{-1} x$ is the INVERSE COTANGENT. Gregory numbers arise in the determination of MACHIN-LIKE FORMULAS. Every Gregory number t_x can be expressed uniquely as a sum of t_ns where the ns are STØRMER NUMBERS.

References

Conway, J. H. and Guy, R. K. "Gregory's Numbers" In *The Book of Numbers.* New York: Springer-Verlag, pp. 241–242, 1996.

Grelling's Paradox

A semantic PARADOX, also called the HETEROLOGICAL PARADOX, which arises by defining "heterological" to mean "a word which does not describe itself." The word "heterological" is therefore heterological IFF it is not.

see also RUSSELL'S PARADOX

References

Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, pp. 20–21, 1989.

Grenz-Formel

An equation derived by Kronecker:

$$\sum{}' (x^2 + y^2 + dz^2)^{-s} = 4\zeta(s)\eta(s) + \frac{2\pi}{s-1} \frac{\zeta(2s-2)}{d^{s-1}}$$

$$+ \frac{2\pi^s}{\Gamma(s)} d^{(1-s)/2} \sum_{n=1}^{\infty} n^{(s-1)/2} \sum_{u^2|n} \frac{r\left(\frac{n}{u^2}\right)}{u^{2s-2}} \int_0^{\infty} e^{\pi\sqrt{nd}\,(y+y^{-1})} y^{s-2}\,dy,$$

where

$$r(n) = 4 \sum_{d|n} \sin(\tfrac{1}{2}\pi d),$$

$\zeta(z)$, is the RIEMANN ZETA FUNCTION, $\eta(z)$ is the DIRICHLET ETA FUNCTION, $\Gamma(z)$ is the GAMMA FUNCTION, and the primed sum omits infinite terms (Selberg and Chowla 1967).

References

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, pp. 296–297, 1987.

Selberg, A. and Chowla, S. "On Epstein's Zeta-Function." *J. Reine. Angew. Math.* **227**, 86–110, 1967.

Griffiths Points

"The" Griffiths point is the fixed point in GRIFFITHS' THEOREM. Given four points on a CIRCLE and a line through the center of the CIRCLE, the four corresponding Griffiths points are COLLINEAR (Tabov 1995).

The points

$$Gr = I + 4Ge$$

$$Gr' = I - 4Ge,$$

are known as the first and second Griffiths points, where I is the INCENTER and Ge is the GERGONNE POINT.

see also GERGONNE POINT, GRIFFITHS' THEOREM, INCENTER, OLDKNOW POINTS, RIGBY POINTS

References

Oldknow, A. "The Euler-Gergonne-Soddy Triangle of a Triangle." *Amer. Math. Monthly* **103**, 319–329, 1996.

Tabov, J. "Four Collinear Griffiths Points." *Math. Mag.* **68**, 61–64, 1995.

Griffiths' Theorem

When a point P moves along a line through the CIRCUMCENTER of a given TRIANGLE Δ, the CIRCUMCIRCLE of the PEDAL TRIANGLE of P with respect to Δ passes through a fixed point (the GRIFFITHS POINT) on the NINE-POINT CIRCLE of Δ.

see also CIRCUMCENTER, GRIFFITHS POINTS, NINE-POINT CIRCLE, PEDAL TRIANGLE

Grimm's Conjecture

Grimm conjectures that if $n+1$, $n+2$, ..., $n+k$ are all COMPOSITE NUMBERS, then there are distinct PRIMES p_{i_j} such that $p_{i_j} | (n+j)$ for $1 \leq j \leq k$.

References

Guy, R. K. "Grimm's Conjecture." §B32 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 86, 1994.

Grinberg Formula

A formula satisfied by all HAMILTONIAN CIRCUITS with n nodes. Let f_j be the number of regions inside the circuit with j sides, and let g_j be the number of regions outside the circuit with j sides. If there are d interior diagonals, then there must be $d+1$ regions

$$[\#\text{ regions in interior}] = d+1 = f_2 + f_3 + \ldots + f_n. \quad (1)$$

Any region with j sides is bounded by j EDGES, so such regions contribute jf_j to the total. However, this counts each diagonal twice (and each EDGE only once). Therefore,

$$2f_2 + 3f_3 + \ldots + nf_n = 2d + n. \quad (2)$$

Take (2) $- 2 \times$ (1),

$$f_3 + 2f_4 + 3f_5 + \ldots + (n-2)f_n = n - 2. \quad (3)$$

Similarly,

$$g_3 + 2g_4 + \ldots + (n-2)g_n = n - 2, \quad (4)$$

so

$$(f_3 - g_3) + 2(f_4 - g_4) + 3(f_5 - g_5) + \ldots + (n-2)(f_n - g_n) = 0. \quad (5)$$

Gröbner Basis

A Gröbner basis for a system of POLYNOMIAL equations is an equivalence system that possesses useful properties. It is very roughly analogous to computing an ORTHONORMAL BASIS from a set of BASIS VECTORS and can be described roughly as a combination of GAUSSIAN ELIMINATION (for linear systems) and the EUCLIDEAN ALGORITHM (for UNIVARIATE POLYNOMIALS over a FIELD).

Gröbner bases are useful in the construction of symbolic algebra algorithms. The algorithm for computing Gröbner bases is known as BUCHBERGER'S ALGORITHM.

see also BUCHBERGER'S ALGORITHM, COMMUTATIVE ALGEBRA

References

Adams, W. W. and Loustaunau, P. *An Introduction to Gröbner Bases.* Providence, RI: Amer. Math. Soc., 1994.

Becker, T. and Weispfennig, V. *Gröbner Bases: A Computational Approach to Commutative Algebra.* New York: Springer-Verlag, 1993.

Cox, D.; Little, J.; and O'Shea, D. *Ideals, Varieties, and Algorithms: An Introduction to Algebraic Geometry and Commutative Algebra, 2nd ed.* New York: Springer-Verlag, 1996.

Eisenbud, D. *Commutative Algebra with a View toward Algebraic Geometry.* New York: Springer-Verlag, 1995.

Mishra, B. *Algorithmic Algebra.* New York: Springer-Verlag, 1993.

Groemer Packing

A honeycomb-like packing that forms HEXAGONS.

see also GROEMER THEOREM

References

Stewart, I. "A Bundling Fool Beats the Wrap." *Sci. Amer.* **268**, 142–144, 1993.

Groemer Theorem

Given n CIRCLES and a PERIMETER p, the total AREA of the CONVEX HULL is

$$A_{\text{Convex Hull}} = 2\sqrt{3}(n-1) + p(1 - \tfrac{1}{2}\sqrt{3}\,) + \pi(\sqrt{3} - 1).$$

Furthermore, the actual AREA equals this value IFF the packing is a GROEMER PACKING. The theorem was proved in 1960 by Helmut Groemer.

see also CONVEX HULL

Gronwall's Theorem

Let $\sigma(n)$ be the DIVISOR FUNCTION. Then

$$\overline{\lim_{n\to\infty}}\, \frac{\sigma(n)}{n \ln \ln n} = e^{\gamma},$$

where γ is the EULER-MASCHERONI CONSTANT. Ramanujan independently discovered a less precise version of this theorem (Berndt 1994). Robin (1984) showed that the validity of the inequality

$$\sigma(n) < e^{\gamma} n \ln \ln n$$

for $n \geq 5041$ is equivalent to the RIEMANN HYPOTHESIS.

References

Berndt, B. C. *Ramanujan's Notebooks: Part I.* New York: Springer-Verlag, p. 94, 1985.

Gronwall, T. H. "Some Asymptotic Expressions in the Theory of Numbers." *Trans. Amer. Math. Soc.* **37**, 113–122, 1913.

Nicholas, J.-L. "On Highly Composite Numbers." In *Ramanujan Revisited: Proceedings of the Centenary Conference* (Ed. G. E. Andrews, B. C. Berndt, and R. A. Rankin). Boston, MA: Academic Press, pp. 215–244, 1988.

Robin, G. "Grandes Valeurs de la foction somme des diviseurs et hypothèse de Riemann." *J. Math. Pures Appl.* **63**, 187–213, 1984.

Gross

A DOZEN DOZEN, or the SQUARE NUMBER 144.

see also 12, DOZEN

Grossencharacter

In the original formulation, a quantity associated with ideal class groups. According to Chevalley's formulation, a Grossencharacter is a MULTIPLICATIVE CHARACTER of the group of ADÈLES that is trivial on the diagonally embedded $k^{\times}$, where k is a number FIELD.

References

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Grossman's Constant

Define the sequence $a_0 = 1$, $a_1 = x$, and

$$a_{n+2} = \frac{a_n}{1 + a_{n+1}}$$

for $n \geq 0$. Janssen and Tjaden (1987) showed that this sequence converges for exactly one value of x, $x = 0.73733830336929\ldots$, confirming Grossman's conjecture.

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/grssmn/grssmn.html.`

Janssen, A. J. E. M. and Tjaden, D. L. A. Solution to Problem 86-2. *Math. Intel.* **9**, 40–43, 1987.

Grothendieck's Majorant

The best known majorant of Grothendieck's constant. Let A be an $n \times n$ REAL SQUARE MATRIX such that

$$\left| \sum_{1 \leq i,j \leq n} a_{ij} x_i y_j \right| \tag{1}$$

in which x_i and y_j have REAL ABSOLUTE VALUES $<$ 1. Grothendieck has shown there exists a number K_G independent of A and n satisfying

$$\left| \sum_{1 \leq i,j \leq n} a_{ij} \langle x_i, y_j \rangle \right| \tag{2}$$

in which the vectors x_i and y_j have a norm < 1 in HILBERT SPACE. The Grothendieck constant is the smallest REAL NUMBER for which this inequality has been proven. Krivine (1977) showed that

$$1.676\ldots \leq K_G \leq 1.782\ldots, \tag{3}$$

and has postulated that

$$K_G \equiv \frac{\pi}{2\ln(1+\sqrt{2}\,)} = 1.7822139\ldots. \tag{4}$$

It is related to KHINTCHINE'S CONSTANT.

References

Krivine, J. L. "Sur las constante de Grothendieck." *C. R. A. S.* **284**, 8, 1977.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 42, 1983.

Grothendieck's Theorem

Let E and F be paired spaces with S a family of absolutely convex bounded sets of F such that the sets of S generate F and, if $B_1, B_2 \in S$, then there exists a $B_3 \in S$ such that $B_3 \supset B_1$ and $B_3 \supset B_2$. Then E_S is complete IFF algebraic linear functional $f(y)$ of F that is weakly continuous on every $B \in S$ is expressed as $f(y) = \langle x, y \rangle$ for some $x \in E$. When E_S is not complete, the space of all linear functionals satisfying this condition gives the completion $\hat{E}_S$ of E_S.

see also MACKEY'S THEOREM

References

Iyanaga, S. and Kawada, Y. (Eds.). "Grothendieck's Theorem." §407L in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1274, 1980.

Ground Set

A PARTIALLY ORDERED SET is defined as an ordered pair $P = (X, \leq)$. Here, X is called the GROUND SET of P and $\leq$ is the PARTIAL ORDER of P.

see also PARTIAL ORDER, PARTIALLY ORDERED SET

Group

A group G is defined as a finite or infinite set of OPERANDS (called "elements") $A, B, C, \ldots$ that may be combined or "multiplied" via a BINARY OPERATOR to form well-defined products and which furthermore satisfy the following conditions:

1. Closure: If A and B are two elements in G, then the product AB is also in G.
2. Associativity: The defined multiplication is associative, i.e., for all $A, B, C \in G$, $(AB)C = A(BC)$.
3. Identity: There is an IDENTITY ELEMENT I (a.k.a. 1, E, or e) such that $IA = AI = A$ for every element $A \in G$.
4. Inverse: There must be an inverse or reciprocal of each element. Therefore, the set must contain an element $B = A^{-1}$ such that $AA^{-1} = A^{-1}A = I$ for each element of G.

A group is therefore a MONOID for which every element is invertible. A group must contain at least one element.

The study of groups is known as GROUP THEORY. If there are a finite number of elements, the group is called a FINITE GROUP and the number of elements is called the ORDER of the group.

Since each element A, B, C, ..., X, and Y is a member of the GROUP, GROUP property 1 requires that the product

$$D \equiv ABC \cdots XY \tag{1}$$

must also be a member. Now apply D to $Y^{-1}X^{-1} \cdots C^{-1}B^{-1}A^{-1}$,

$$\begin{aligned} &D(Y^{-1}X^{-1} \cdots C^{-1}B^{-1}A^{-1}) \\ &\quad = (ABC \cdots XY)(Y^{-1}X^{-1} \cdots C^{-1}B^{-1}A^{-1}). \end{aligned} \tag{2}$$

But

$$\begin{aligned} &ABC \cdots XYY^{-1}X^{-1} \cdots C^{-1}B^{-1}A^{-1} \\ &\quad = ABC \cdots XIX^{-1} \cdots C^{-1}B^{-1}A^{-1} \\ &\quad = ABC \cdots C^{-1}B^{-1}A^{-1} = \ldots = AA^{-1} = I, \end{aligned} \tag{3}$$

so

$$I = D(Y^{-1}X^{-1} \cdots C^{-1}B^{-1}A^{-1}), \tag{4}$$

which means that

$$D^{-1} = Y^{-1}X^{-1} \cdots C^{-1}B^{-1}A^{-1} \tag{5}$$

and

$$(ABC \cdots XY)^{-1} = Y^{-1}X^{-1} \cdots C^{-1}B^{-1}A^{-1}. \tag{6}$$

An IRREDUCIBLE REPRESENTATION of a group is a representation for which there exists no UNITARY TRANSFORMATION which will transform the representation MATRIX into block diagonal form. The IRREDUCIBLE REPRESENTATION has some remarkable properties. Let the ORDER of a GROUP be h, and the dimension of the ith representation (the order of each constituent matrix) be l_i (a POSITIVE INTEGER). Let any operation be denoted R, and let the mth row and nth column of the matrix corresponding to a matrix R in the ith IRREDUCIBLE REPRESENTATION be $\Gamma_i(R)_{mn}$. The following properties can be derived from the GROUP ORTHOGONALITY THEOREM,

$$\sum_R \Gamma_i(R)_{mn} \Gamma_j(R)_{m'n'}{}^* = \frac{h}{\sqrt{l_i l_j}} \delta_{ij} \delta_{mm'} \delta_{nn'}. \tag{7}$$

1. The DIMENSIONALITY THEOREM:

   $$h = \sum_i l_i{}^2 = l_1{}^2 + l_2{}^2 + l_3{}^2 + \ldots = \sum_i \chi_i{}^2(I), \tag{8}$$

   where each l_i must be a POSITIVE INTEGER and χ is the CHARACTER (trace) of the representation.
2. The sum of the squares of the CHARACTERS in any IRREDUCIBLE REPRESENTATION i equals h,

   $$h = \sum_R \chi_i{}^2(R). \tag{9}$$

3. ORTHOGONALITY of different representations

   $$\sum_R \chi_i(R) \chi_j(R) = 0 \quad \text{for } i \neq j. \tag{10}$$

4. In a given representation, reducible or irreducible, the CHARACTERS of all MATRICES belonging to operations in the same class are identical (but differ from those in other representations).
5. The number of IRREDUCIBLE REPRESENTATIONS of a GROUP is equal to the number of CONJUGACY CLASSES in the GROUP. This number is the dimension of the Γ MATRIX (although some may have zero elements).
6. A one-dimensional representation with all 1s (totally symmetric) will always exist for any GROUP.
7. A 1-D representation for a GROUP with elements expressed as MATRICES can be found by taking the CHARACTERS of the MATRICES.
8. The number a_i of IRREDUCIBLE REPRESENTATIONS χ_i present in a reducible representation c is given by

   $$a_i = \frac{1}{h} \sum_R \chi(R) \chi_i(R), \tag{11}$$

where h is the Order of the Group and the sum must be taken over all elements in each class. Written explicitly,

$$a_i = \frac{1}{h}\sum_R \chi(R)\chi_i{}'(R)n_R, \tag{12}$$

where $\chi_i{}'$ is the Character of a single entry in the Character Table and n_R is the number of elements in the corresponding Conjugacy Class.

see also Abelian Group, Adéle Group, Affine Group, Alternating Group, Artinian Group, Aschbacher's Component Theorem, B_p-Theorem, Baby Monster Group, Betti Group, Bimonster, Bordism Group, Braid Group, Brauer Group, Burnside Problem, Center (Group), Centralizer, Character (Group), Character (Multiplicative), Chevalley Groups, Classical Groups, Cobordism Group, Cohomotopy Group, Component, Conjugacy Class, Coset, Conway Groups, Coxeter Group, Cyclic Group, Dihedral Group, Dimensionality Theorem, Dynkin Diagram, Elliptic Group Modulo p, Engel's Theorem, Euclidean Group, Feit-Thompson Theorem, Finite Group, Fischer Groups, Fischer's Baby Monster Group, Fundamental Group, General Linear Group, General Orthogonal Group, General Unitary Group, Global $C(G;T)$ Theorem, Groupoid, Group Orthogonality Theorem, Hall-Janko Group, Hamiltonian Group, Harada-Norton Group, Heisenberg Group, Held Group, Hermann-Mauguin Symbol, Higman-Sims Group, Homeomorphic Group, Hypergroup, Icosahedral Group, Irreducible Representation, Isomorphic Groups, Janko Groups, Jordan-Hölder Theorem, Kleinian Group, Kummer Group, $L_{p'}$-Balance Theorem, Lagrange's Group Theorem, Local Group Theory, Linear Group, Lyons Group, Mathieu Groups, Matrix Group, McLaughlin Group, Möbius Group, Modular Group, Modulo Multiplication Group, Monodromy Group, Monoid, Monster Group, Mulliken Symbols, Néron-Severi Group, Nilpotent Group, Noncommutative Group, Normal Subgroup, Normalizer, O'Nan Group, Octahedral Group, Order (Group), Orthogonal Group, Orthogonal Rotation Group, Outer Automorphism Group, p-Group, p'-Group, p-Layer, Point Groups, Positive Definite Function, Prime Group, Projective General Linear Group, Projective General Orthogonal Group, Projective General Unitary Group, Projective Special Linear Group, Projective Special Orthogonal Group, Projective Special Unitary Group, Projective Symplectic Group, Pseudogroup, Quasigroup, Quasisimple Group, Quasithin Theorem, Quasi-Unipotent Group, Representation, Residue Class, Rubik's Cube, Rudvalis Group, Schönflies Symbol, Schur Multiplier, Semisimple, Signalizer Functor Theorem, Selmer Group, Semigroup, Simple Group, Solvable Group, Space Groups, Special Linear Group, Special Orthogonal Group, Special Unitary Group, Sporadic Group, Stochastic Group, Strongly Embedded Theorem, Subgroup, Subnormal, Support, Suzuki Group, Symmetric Group, Symplectic Group, Tetrahedral Group, Thompson Group, Tightly Embedded, Tits Group, Triangular Symmetry Group, Twisted Chevalley Groups, Unimodular Group, Unipotent, Unitary Group, Viergruppe, von Dyck's Theorem

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 237–276, 1985.

Farmer, D. *Groups and Symmetry.* Providence, RI: Amer. Math. Soc., 1995.

Weisstein, E. W. "Groups." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Groups.m`.

Weyl, H. *The Classical Groups: Their Invariants and Representations.* Princeton, NJ: Princeton University Press, 1997.

Wybourne, B. G. *Classical Groups for Physicists.* New York: Wiley, 1974.

Group Convolution

The convolution of two Complex-valued functions on a Group G is defined as

$$(a*b)(g) = \sum_{k\in G} a(k)b(k^{-1}g)$$

where the Support (set which is not zero) of each function is finite.

References

Weinstein, A. "Groupoids: Unifying Internal and External Symmetry." *Not. Amer. Math. Soc.* **43**, 744–752, 1996.

Group Orthogonality Theorem

Let Γ be a representation for a Group of Order h, then

$$\sum_R \Gamma_i(R)_{mn}\Gamma_j(R)_{m'n'}{}^* = \frac{h}{\sqrt{l_i l_j}}\delta_{ij}\delta_{mm'}\delta_{nn'}.$$

The proof is nontrivial and may be found in Eyring *et al.* (1944).

References

Eyring, H.; Walker, J.; and Kimball, G. E. *Quantum Chemistry.* New York: Wiley, p. 371, 1944.

Group Ring

The set of sums $\sum_x a_x x$ ranging over a multiplicative Group and a_i are elements of a Field with all but a finite number of $a_i = 0$.

Group Theory

The study of Groups. Gauss developed but did not publish parts of the mathematics of group theory, but Galois is generally considered to have been the first to develop the theory. Group theory is a powerful formal method for analyzing abstract and physical systems in which Symmetry is present and has surprising importance in physics, especially quantum mechanics.

see also Finite Group, Group, Plethysm, Symmetry

References

Arfken, G. "Introduction to Group Theory." §4.8 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 237–276, 1985.

Burnside, W. *Theory of Groups of Finite Order, 2nd ed.* New York: Dover, 1955.

Burrow, M. *Representation Theory of Finite Groups.* New York: Dover, 1993.

Carmichael, R. D. *Introduction to the Theory of Groups of Finite Order.* New York: Dover, 1956.

Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, 1985.

Cotton, F. A. *Chemical Applications of Group Theory, 3rd ed.* New York: Wiley, 1990.

Dixon, J. D. *Problems in Group Theory.* New York: Dover, 1973.

Grossman, I. and Magnus, W. *Groups and Their Graphs.* Washington, DC: Math. Assoc. Amer., 1965.

Hamermesh, M. *Group Theory and Its Application to Physical Problems.* New York: Dover, 1989.

Lomont, J. S. *Applications of Finite Groups.* New York: Dover, 1987.

Magnus, W.; Karrass, A.; and Solitar, D. *Combinatorial Group Theory: Presentations of Groups in Terms of Generators and Relations.* New York: Dover, 1976.

Robinson, D. J. S. *A Course in the Theory of Groups, 2nd ed.* New York: Springer-Verlag, 1995.

Rose, J. S. *A Course on Group Theory.* New York: Dover, 1994.

Rotman, J. J. *An Introduction to the Theory of Groups, 4th ed.* New York: Springer-Verlag, 1995.

Groupoid

There are at least two definitions of "groupoid" currently in use.

The first type of groupoid is an algebraic structure on a Set with a Binary Operator. The only restriction on the operator is Closure (i.e., applying the Binary Operator to two elements of a given set S returns a value which is itself a member of S). Associativity, commutativity, etc., are not required (Rosenfeld 1968, pp. 88–103). A groupoid can be empty. The numbers of nonisomorphic groupoids of this type having n elements are 1, 1, 10, 3330, 178981952, ... (Sloane's A001329), and the numbers of nonisomorphic and nonantiisimorphic groupoids are 1, 7, 1734, 89521056, ... (Sloane's A001424). An associative groupoid is called a Semigroup.

The second type of groupoid is an algebraic structure first defined by Brandt (1926) and also known as a Virtual Group. A groupoid with base B is a set G with mappings α and β from G onto B and a partially defined binary operation $(g,h) \mapsto gh$, satisfying the following four conditions:

1. gh is defined only when $\beta(G) = \alpha(h)$ for certain maps α and β from G onto $\mathbb{R}^2$ with $\alpha : (x,\gamma,y) \mapsto x$ and $\beta : (x,\gamma,y) \mapsto y$.
2. Associativity: If either $(gh)k$ or $g(hk)$ is defined, then so is the other and $(gh)k = g(hk)$.
3. For each g in G, there are left and right Identity Elements λ_g and ρ_g such that $\lambda_g g = g = g\rho_g$.
4. Each g in G has an inverse g^{-1} for which $gg^{-1} = \lambda_g$ and $g^{-1}g = \rho_g$

(Weinstein 1996). A groupoid is a small Category with every morphism invertible.

see also Binary Operator, Inverse Semigroup, Lie Algebroid, Lie Groupoid, Monoid, Quasigroup, Semigroup, Topological Groupoid

References

Brandt, W. "Über eine Verallgemeinerung des Gruppengriffes." *Math. Ann.* **96**, 360–366, 1926.

Brown, R. "From Groups to Groupoids: A Brief Survey." *Bull. London Math. Soc.* **19**, 113–134, 1987.

Brown, R. *Topology: A Geometric Account of General Topology, Homotopy Types, and the Fundamental Groupoid.* New York: Halsted Press, 1988.

Higgins, P. J. *Notes on Categories and Groupoids.* London: Van Nostrand Reinhold, 1971.

Ramsay, A.; Chiaramonte, R.; and Woo, L. "Groupoid Home Page." `http://amath-www.colorado.edu:80/math/researchgroups/groupoids/groupoids.shtml`.

Rosenfeld, A. *An Introduction to Algebraic Structures.* New York: Holden-Day, 1968.

Sloane, N. J. A. Sequences A001329/M4760 and A001424/M4465 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Weinstein, A. "Groupoids: Unifying Internal and External Symmetry." *Not. Amer. Math. Soc.* **43**, 744–752, 1996.

Growth

A general term which refers to an increase (or decrease in the case of the oxymoron "Negative growth") in a given quantity.

see also Growth Function, Growth Spiral

Growth Function

see Block Growth

Growth Spiral

see Logarithmic Spiral

Grundy's Game

A special case of NIM played by the following rules. Given a heap of size n, two players alternately select a heap and divide it into two unequal heaps. A player loses when he cannot make a legal move because all heaps have size 1 or 2. Flammenkamp gives a table of the extremal SPRAGUE-GRUNDY VALUES for this game. The first few values of Grundy's game are 0, 0, 0, 1, 0, 2, 1, 0, 2, ... (Sloane's A002188).

References

Flammenkamp, A. "Sprague-Grundy Values of Grundy's Game." `http://www.minet.uni-jena.de/~achim/grundy.html`.

Sloane, N. J. A. Sequence A002188/M0044 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Grundy-Sprague Number

see NIM-VALUE

Gudermannian Function

Denoted either $\gamma(x)$ or $\operatorname{gd}(x)$.

$$\operatorname{gd}(x) \equiv \tan^{-1}(\sinh x) = 2\tan^{-1}(e^x) - \tfrac{1}{2}\pi \qquad (1)$$

$$\operatorname{gd}^{-1}(x) = \ln[\tan(\tfrac{1}{4}\pi + \tfrac{1}{2}x)] = \ln(\sec x + \tan x). \qquad (2)$$

The derivatives are given by

$$\frac{d}{dx}\operatorname{gd}(x) = \operatorname{sech} x \qquad (3)$$

$$\frac{d}{dx}\operatorname{gd}^{-1}(x) = \sec x. \qquad (4)$$

Guldinus Theorem

see PAPPUS'S CENTROID THEOREM

Gumbel's Distribution

A special case of the FISHER-TIPPETT DISTRIBUTION with $a = 0$, $b = 1$. The MEAN, VARIANCE, SKEWNESS, and KURTOSIS are

$$\begin{aligned} \mu &= \gamma \\ \sigma^2 &= \tfrac{1}{6}\pi^2 \\ \gamma_1 &= \frac{12\sqrt{6}\,\zeta(3)}{\pi^3} \\ \gamma_2 &= \tfrac{12}{5}. \end{aligned}$$

where γ is the EULER-MASCHERONI CONSTANT, and $\zeta(3)$ is APÉRY'S CONSTANT.

see also FISHER-TIPPETT DISTRIBUTION

Guthrie's Problem

The problem of deciding if four-colors are sufficient to color any map on a plane or SPHERE.

see also FOUR-COLOR THEOREM

Gutschoven's Curve

see KAPPA CURVE

Guy's Conjecture

Guy's conjecture, which has not yet been proven or disproven, states that the CROSSING NUMBER for a COMPLETE GRAPH of order n is

$$\frac{1}{4}\left\lfloor\frac{n}{2}\right\rfloor\left\lfloor\frac{n-1}{2}\right\rfloor\left\lfloor\frac{n-2}{2}\right\rfloor\left\lfloor\frac{n-3}{2}\right\rfloor,$$

where $\lfloor x\rfloor$ is the FLOOR FUNCTION, which can be rewritten

$$\begin{cases} \frac{1}{64}n(n-2)^2(n-4) & \text{for } n \text{ even} \\ \frac{1}{64}(n-1)^2(n-3)^2 & \text{for } n \text{ odd.} \end{cases}$$

The first few values are 0, 0, 0, 0, 1, 3, 9, 18, 36, 60, ... (Sloane's A000241).

see also CROSSING NUMBER (GRAPH)

References

Sloane, N. J. A. Sequence A000241/M2772 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Gyrate Bidiminished Rhombicosidodecahedron

see JOHNSON SOLID

Gyrate Rhombicosidodecahedron

see JOHNSON SOLID

Gyrobicupola

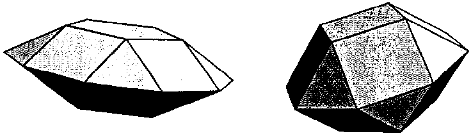

A BICUPOLA in which the bases are in opposite orientations.

see also BICUPOLA, PENTAGONAL GYROBICUPOLA, SQUARE GYROBICUPOLA

Gyrobifastigium

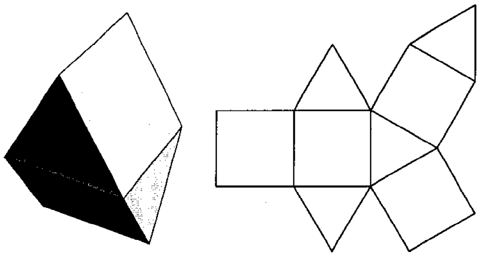

JOHNSON SOLID J_{26}, consisting of two joined triangular PRISMS.

Gyrobirotunda

A BIROTUNDA in which the bases are in opposite orientations.

Gyrocupolarotunda

A CUPOLAROTUNDA in which the bases are in opposite orientations.

see also ORTHOCUPOLAROTUNDA

Gyroelongated Cupola

A n-gonal CUPOLA adjoined to a $2n$-gonal ANTIPRISM.

see also GYROELONGATED PENTAGONAL CUPOLA, GYROELONGATED SQUARE CUPOLA, GYROELONGATED TRIANGULAR CUPOLA

Gyroelongated Dipyramid

see GYROELONGATED PYRAMID, GYROELONGATED SQUARE DIPYRAMID

Gyroelongated Pentagonal Bicupola

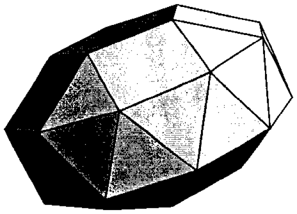

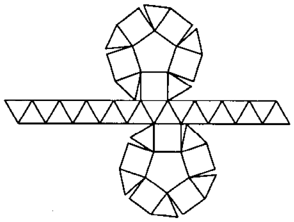

JOHNSON SOLID J_{46}, which consists of a PENTAGONAL ROTUNDA adjoined to a decagonal ANTIPRISM.

Gyroelongated Pentagonal Birotunda

see JOHNSON SOLID

Gyroelongated Pentagonal Cupola

see JOHNSON SOLID

Gyroelongated Pentagonal Cupolarotunda

see JOHNSON SOLID

Gyroelongated Pentagonal Pyramid

see JOHNSON SOLID

Gyroelongated Pentagonal Rotunda

see JOHNSON SOLID

Gyroelongated Pyramid

An n-gonal pyramid adjoined to an n-gonal ANTIPRISM.

see also ELONGATED PYRAMID, GYROELONGATED DIPYRAMID, GYROELONGATED PENTAGONAL PYRAMID, GYROELONGATED SQUARE DIPYRAMID, GYROELONGATED SQUARE PYRAMID

Gyroelongated Rotunda

see GYROELONGATED PENTAGONAL ROTUNDA

Gyroelongated Square Bicupola

see JOHNSON SOLID

Gyroelongated Square Cupola

see JOHNSON SOLID

Gyroelongated Square Dipyramid

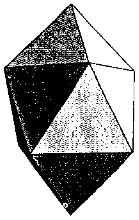

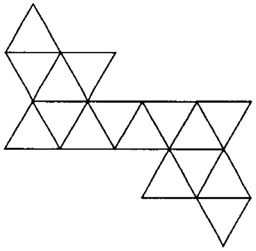

One of the eight convex DELTAHEDRA. It consists of two oppositely faced SQUARE PYRAMIDS rotated 45° to each other and separated by a ribbon of eight side-to-side TRIANGLES. It is JOHNSON SOLID J_{17}.

Call the coordinates of the upper PYRAMID bases $(\pm 1, \pm 1, h_1)$ and of the lower $(\pm\sqrt{2}, 0, -h_1)$ and $(0, \pm\sqrt{2}, -h_1)$. Call the PYRAMID apexes $(0, 0, \pm(h_1 + h_2))$. Consider the points $(1, 1, 0)$ and $(0, 0, h_1 + h_2)$. The height of the PYRAMID is then given by

$$\sqrt{1^2 + 1^2 + {h_2}^2} = \sqrt{2 + {h_2}^2} = 2 \qquad (1)$$

$$h_2 = \sqrt{2}. \qquad (2)$$

Now consider the points $(1, 1, h_1)$ and $(\sqrt{2}, 0, -h_1)$. The height of the base is given by

$$\begin{aligned}(1 - \sqrt{2})^2 + 1^2 + (2h_1)^2 &= 1 - 2\sqrt{2} + 2 + 1 + 4{h_1}^2 \\ &= 4 - 2\sqrt{2} + 4{h_1}^2 = 2^2 = 4 \qquad (3)\end{aligned}$$

$$4{h_1}^2 = 2\sqrt{2} \qquad (4)$$

$${h_1}^2 = \frac{\sqrt{2}}{2} = \frac{1}{\sqrt{2}} = 2^{-1/2}, \qquad (5)$$

so

$$h_1 = 2^{-1/4} \qquad (6)$$

$$h_2 = 2^{1/2}. \qquad (7)$$

Gyroelongated Square Pyramid

see JOHNSON SOLID

Gyroelongated Triangular Bicupola

see JOHNSON SOLID

Gyroelongated Triangular Cupola

see JOHNSON SOLID

H

h-Cobordism

An *h*-cobordism is a COBORDISM W between two MANIFOLDS M_1 and M_2 such that W is SIMPLY CONNECTED and the inclusion maps $M_1 \to W$ and $M_2 \to W$ are HOMOTOPY equivalences.

h-Cobordism Theorem

If W is a SIMPLY CONNECTED, COMPACT MANIFOLD with a boundary that has two components, M_1 and M_2, such that inclusion of each is a HOMOTOPY equivalence, then W is DIFFEOMORPHIC to the product $M_1 \times [0,1]$ for $\dim(M_1) \geq 5$. In other words, if M and M' are two simply connected MANIFOLDS of DIMENSION ≥ 5 and there exists an *h*-COBORDISM W between them, then W is a product $M \times I$ and M is DIFFEOMORPHIC to M'.

The proof of the *h*-cobordism theorem can be accomplished using SURGERY. A particular case of the *h*-cobordism theorem is the POINCARÉ CONJECTURE in dimension $n \geq 5$. Smale proved this theorem in 1961.

see also DIFFEOMORPHISM, POINCARÉ CONJECTURE, SURGERY

References

Smale, S. "Generalized Poincaré's Conjecture in Dimensions Greater than Four." *Ann. Math.* **74**, 391–406, 1961.

H-Fractal

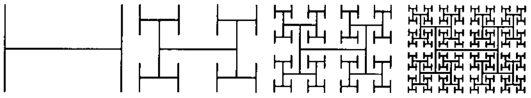

The FRACTAL illustrated above.

References

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 1–2, 1991.

Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m`.

H-Function

see FOX'S *H*-FUNCTION

H-Spread

The difference $H_2 - H_1$, where H_1 and H_2 are HINGES. It is the same as the INTERQUARTILE RANGE for $N = 5$, 9, 13, ... points.

see also HINGE, INTERQUARTILE RANGE, STEP

References

Tukey, J. W. *Explanatory Data Analysis.* Reading, MA: Addison-Wesley, p. 44, 1977.

H-Transform

A 2-D generalization of the HAAR TRANSFORM which is used for the compression of astronomical images. The algorithm consists of dividing the $2^N \times 2^N$ image into blocks of 2×2 pixels, calling the pixels in the block a_{00}, a_{10}, a_{01}, and a_{11}. For each block, compute the four coefficients

$$\begin{aligned}
h_0 &\equiv \tfrac{1}{2}(a_{11} + a_{10} + a_{01} + a_{00}) \\
h_x &\equiv \tfrac{1}{2}(a_{11} + a_{10} - a_{01} - a_{00}) \\
h_y &\equiv \tfrac{1}{2}(a_{11} - a_{10} + a_{01} - a_{00}) \\
h_c &\equiv \tfrac{1}{2}(a_{11} - a_{10} - a_{01} + a_{00}).
\end{aligned}$$

Construct a $2^{N-1} \times 2^{N-1}$ image from the h_0 values, and repeat until only one h_0 value remains. The H-transform can be performed in place and requires about $16N^2/3$ additions for an $N \times N$ image.

see also HAAR TRANSFORM

References

Capaccioli, M.; Held, E. V.; Lorenz, H.; Richter, G. M.; and Ziener, R. "Application of an Adaptive Filtering Technique to Surface Photometry of Galaxies. I. The Method Tested on NGC 3379." *Astron. Nachr.* **309**, 69–80, 1988.

Fritze, K.; Lange, M.; Möstle, G.; Oleak, H.; and Richter, G. M. "A Scanning Microphotometer with an On-Line Data Reduction for Large Field Schmidt Plates." *Astron. Nachr.* **298**, 189–196, 1977.

Richter, G. M. "The Evaluation of Astronomical Photographs with the Automatic Area Photometer." *Astron. Nachr.* **299**, 283–303, 1978.

White, R. L.; Postman, M.; and Lattanzi, M. G. "Compression of the Guide Star Digitised Schmidt Plates." In *Digitised Optical Sky Surveys: Proceedings of the Conference on "Digitised Optical Sky Surveys" held in Edinburgh, Scotland, 18–21 June 1991* (Ed. H. T. MacGillivray and E. B. Thompson). Dordrecht, Netherlands: Kluwer, pp. 167–175, 1992.

Haar Function

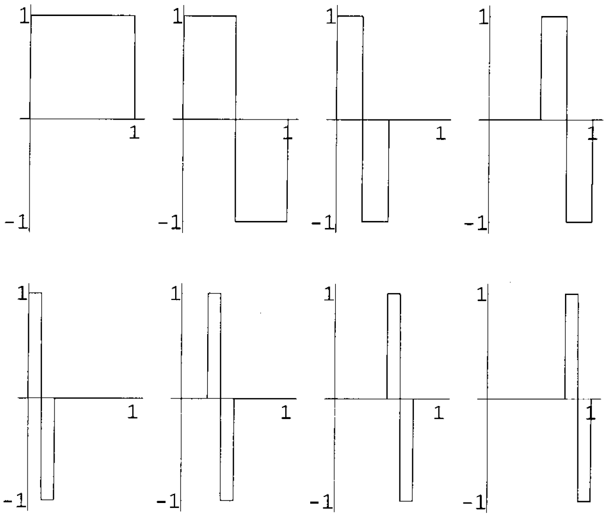

Define

$$\psi(x) \equiv \begin{cases} 1 & 0 \le x \le \frac{1}{2} \\ -1 & \frac{1}{2} \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

and

$$\psi_{jk}(x) \equiv \psi(2^j x - k), \tag{2}$$

where the FUNCTIONS plotted above are

$$\begin{aligned} \psi_{00} &= \psi(x) \\ \psi_{10} &= \psi(2x) \\ \psi_{11} &= \psi(2x-1) \\ \psi_{20} &= \psi(4x) \\ \psi_{21} &= \psi(4x-1) \\ \psi_{21} &= \psi(4x-2) \\ \psi_{21} &= \psi(4x-3). \end{aligned}$$

Then a FUNCTION $f(x)$ can be written as a series expansion by

$$f(x) = c_0 + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} c_{jk}\psi_{jk}(x). \tag{3}$$

The FUNCTIONS ψ_{jk} and ψ are all ORTHOGONAL in $[0,1]$, with

$$\int_0^1 \phi(x)\phi_{jk}(x)\,dx = 0 \tag{4}$$

$$\int_0^1 \phi_{jk}(x)\phi_{lm}(x)\,dx = 0. \tag{5}$$

These functions can be used to define WAVELETS. Let a FUNCTION be defined on n intervals, with n a POWER of 2. Then an arbitrary function can be considered as an n-VECTOR $\mathbf{f}$, and the COEFFICIENTS in the expansion $\mathbf{b}$ can be determined by solving the MATRIX equation

$$\mathbf{f} = \mathsf{W}_n \mathbf{b} \tag{6}$$

for $\mathbf{b}$, where W is the MATRIX of ψ basis functions. For example,

$$\begin{aligned} \mathsf{W}_4 &= \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & 0 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & & \\ 1 & -1 & & \\ & & 1 & 1 \\ & & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & & 1 & \\ & 1 & & \\ & & & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & & \\ 1 & -1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}. \end{aligned} \tag{7}$$

The WAVELET MATRIX can be computed in $\mathcal{O}(n)$ steps, compared to $\mathcal{O}(n \lg n)$ for the FOURIER MATRIX.

see also WAVELET, WAVELET TRANSFORM

References

Haar, A. "Zur Theorie der orthogonalen Funktionensysteme." *Math. Ann.* **69**, 331–371, 1910.

Strang, G. "Wavelet Transforms Versus Fourier Transforms." *Bull. Amer. Math. Soc.* **28**, 288–305, 1993.

Haar Integral

The INTEGRAL associated with the HAAR MEASURE.

see also HAAR MEASURE

Haar Measure

Any locally compact Hausdorff topological group has a unique (up to scalars) NONZERO left invariant measure which is finite on compact sets. If the group is Abelian or compact, then this measure is also right invariant and is known as the Haar measure.

Haar Transform

A 1-D transform which makes use of the HAAR FUNCTIONS.

see H-TRANSFORM, HAAR FUNCTION

References

Haar, A. "Zur Theorie der orthogonalen Funktionensysteme." *Math. Ann.* **69**, 331–371, 1910.

Haberdasher's Problem

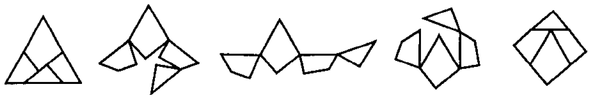

With four cuts, DISSECT an EQUILATERAL TRIANGLE into a SQUARE. First proposed by Dudeney (1907) and discussed in Gardner (1961, p. 34) and Stewart (1987, p. 169). The solution can be hinged so that the three pieces collapse into either the TRIANGLE or the SQUARE.

see also DISSECTION

References

Gardner, M. *The Second Scientific American Book of Mathematical Puzzles & Diversions: A New Selection.* New York: Simon and Schuster, 1961.

Stewart, I. *The Problems of Mathematics, 2nd ed.* Oxford, England: Oxford University Press, 1987.

Hadamard Design

A SYMMETRIC BLOCK DESIGN $(4n+3, n+1, n)$ which is equivalent to a HADAMARD MATRIX of order $4n+4$. It is conjectured that Hadamard designs exist from all integers $n > 0$, but this has not yet been proven. This elusive proof (or disproof) remains one of the most important unsolved problems in COMBINATORICS.

References

Dinitz, J. H. and Stinson, D. R. "A Brief Introduction to Design Theory." Ch. 1 in *Contemporary Design Theory: A Collection of Surveys* (Ed. J. H. Dinitz and D. R. Stinson). New York: Wiley, pp. 1–12, 1992.

Hadamard's Inequality

Let $\mathsf{A} = a_{ii}$ be an arbitrary $n \times n$ nonsingular MATRIX with REAL elements and DETERMINANT $|\mathsf{A}|$, then

$$|\mathsf{A}|^2 \le \prod_{i=1}^{n} \left(\sum_{k=1}^{n} a_{ik}{}^2 \right).$$

see also HADAMARD'S THEOREM

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1110, 1979.

Hadamard Matrix

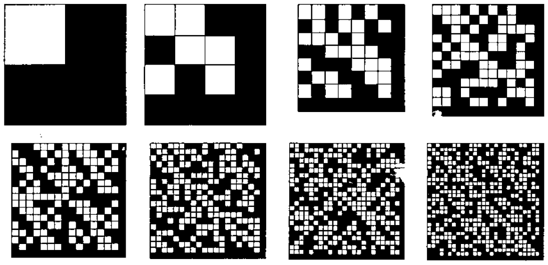

A class of SQUARE MATRIX invented by Sylvester (1867) under the name of ANALLAGMATIC PAVEMENT. A Hadamard matrix is a SQUARE MATRIX containing only 1s and −1s such that when any two columns or rows are placed side by side, HALF the adjacent cells are the same SIGN and half the other (excepting from the count an L-shaped "half-frame" bordering the matrix on two sides which is composed entirely of 1s). When viewed as pavements, cells with 1s are colored black and those with −1s are colored white. Therefore, the $n \times n$ Hadamard matrix H_n must have $n(n-1)/2$ white squares (−1s) and $n(n+1)/2$ black squares (1s).

This is equivalent to the definition

$$\mathsf{H}_n \mathsf{H}_n{}^{\mathrm{T}} = n \mathsf{I}_n, \tag{1}$$

where I_n is the $n \times n$ IDENTITY MATRIX. A Hadamard matrix of order $4n+4$ corresponds to a HADAMARD DESIGN ($4n+3$, $2n+1$, n).

PALEY'S THEOREM guarantees that there always exists a Hadamard matrix H_n when n is divisible by 4 and of the form $2^e(q^m+1)$, where p is an ODD PRIME. In such cases, the MATRICES can be constructed using a PALEY CONSTRUCTION. The PALEY CLASS k is undefined for the following values of $m < 1000$: 92, 116, 156, 172, 184, 188, 232, 236, 260, 268, 292, 324, 356, 372, 376, 404, 412, 428, 436, 452, 472, 476, 508, 520, 532, 536, 584, 596, 604, 612, 652, 668, 712, 716, 732, 756, 764, 772, 808, 836, 852, 856, 872, 876, 892, 904, 932, 940, 944, 952, 956, 964, 980, 988, 996.

Sawade (1985) constructed H_{268}. It is conjectured (and verified up to $n < 428$) that H_n exists for all n DIVISIBLE by 4 (van Lint and Wilson 1993). However, the proof of this CONJECTURE remains an important problem in CODING THEORY. The number of Hadamard matrices of order $4n$ are 1, 1, 1, 5, 3, 60, 487, ... (Sloane's A007299).

If H_n and H_m are known, then H_{nm} can be obtained by replacing all 1s in H_m by H_n and all −1s by $-\mathsf{H}_n$. For $n \leq 100$, Hadamard matrices with $n = 12$, 20, 28, 36, 44, 52, 60, 68, 76, 84, 92, and 100 cannot be built up from lower order Hadamard matrices.

$$\mathsf{H}_2 = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \tag{2}$$

$$\begin{aligned} \mathsf{H}_4 = \begin{bmatrix} \mathsf{H}_2 & \mathsf{H}_2 \\ -\mathsf{H}_2 & \mathsf{H}_2 \end{bmatrix} &= \begin{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\ -\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix}. \end{aligned} \tag{3}$$

H_8 can be similarly generated from H_4. Hadamard matrices can also be expressed in terms of the WALSH FUNCTIONS Cal and Sal

$$H_8 = \begin{bmatrix} \mathrm{Cal}(0,t) \\ \mathrm{Sal}(4,t) \\ \mathrm{Sal}(2,t) \\ \mathrm{Cal}(2,t) \\ \mathrm{Sal}(1,t) \\ \mathrm{Cal}(3,t) \\ \mathrm{Cal}(1,t) \\ \mathrm{Sal}(3,t) \end{bmatrix}. \tag{4}$$

Hadamard matrices can be used to make ERROR-CORRECTING CODES.

see also HADAMARD DESIGN, PALEY CONSTRUCTION, PALEY'S THEOREM, WALSH FUNCTION

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 107–109 and 274, 1987.

Beth, T.; Jungnickel, D.; and Lenz, H. *Design Theory.* New York: Cambridge University Press, 1986.

Colbourn, C. J. and Dinitz, J. H. (Eds.) "Hadamard Matrices and Designs." Ch. 24 in *CRC Handbook of Combinatorial Designs.* Boca Raton, FL: CRC Press, pp. 370–377, 1996.

Geramita, A. V. *Orthogonal Designs: Quadratic Forms and Hadamard Matrices.* New York: Marcel Dekker, 1979.

Golomb, S. W. and Baumert, L. D. "The Search for Hadamard Matrices." *Amer. Math. Monthly* **70**, 12–17, 1963.

Hall, M. Jr. *Combinatorial Theory, 2nd ed.* New York: Wiley, p. 207, 1986.

Hedayat, A. and Wallis, W. D. "Hadamard Matrices and Their Applications." *Ann. Stat.* **6**, 1184–1238, 1978.

Kimura, H. "Classification of Hadamard Matrices of Order 28." *Disc. Math.* **133**, 171–180, 1994.

Kimura, H. "Classification of Hadamard Matrices of Order 28 with Hall Sets." *Disc. Math.* **128**, 257–269, 1994.

Kitis, L. "Paley's Construction of Hadamard Matrices." `http://www.mathsource.com/cgi-bin/MathSource/Applications/Mathematics/0205-760`.

Ogilvie, G. A. "Solution to Problem 2511." *Math. Questions and Solutions* **10**, 74–76, 1868.

Paley, R. E. A. C. "On Orthogonal Matrices." *J. Math. Phys.* **12**, 311–320, 1933.

Ryser, H. J. *Combinatorial Mathematics.* Buffalo, NY: Math. Assoc. Amer., pp. 104–122, 1963.

Sawade, K. "A Hadamard Matrix of Order-268." *Graphs Combinatorics* **1**, 185–187, 1985.
Seberry, J. and Yamada, M. "Hadamard Matrices, Sequences, and Block Designs." Ch. 11 in *Contemporary Design Theory: A Collection of Surveys* (Eds. J. H. Dinitz and D. R. Stinson). New York: Wiley, pp. 431–560, 1992.
Sloane, N. J. A. Sequence A007299/M3736 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Spence, E. "Classification of Hadamard Matrices of Order 24 and 28." *Disc. Math* **140**, 185–243, 1995.
Sylvester, J. J. "Thoughts on Orthogonal Matrices, Simultaneous Sign-Successions, and Tessellated Pavements in Two or More Colours, with Applications to Newton's Rule, Ornamental Tile-Work, and the Theory of Numbers." *Phil. Mag.* **34**, 461–475, 1867.
Sylvester, J. J. "Problem 2511." *Math. Questions and Solutions* **10**, 74, 1868.
van Lint, J. H. and Wilson, R. M. *A Course in Combinatorics.* New York: Cambridge University Press, 1993.

Hadamard's Theorem

Let $|\mathsf{A}|$ be an $n \times n$ DETERMINANT with COMPLEX (or REAL) elements a_{ij}, then $|\mathsf{A}| \neq 0$ if

$$|a_{ii}| > \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|.$$

see also HADAMARD'S INEQUALITY

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1110, 1979.

Hadamard Transform

A FAST FOURIER TRANSFORM-like ALGORITHM which produces a hologram of an image.

Hadamard-Vallée Poussin Constants

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

The sum of RECIPROCALS of PRIMES diverges, but

$$\lim_{n\to\infty}\left[\sum_{k=1}^{\pi(n)}\frac{1}{p_k} - \ln(\ln n)\right] = \gamma + \sum_{k=1}^{\infty}\left[\ln\left(1-\frac{1}{p_k}\right)+\frac{1}{p_k}\right] \equiv C_1 = 0.2614972128\ldots, \qquad (1)$$

where $\pi(n)$ is the PRIME COUNTING FUNCTION and γ is the EULER-MASCHERONI CONSTANT (Le Lionnais 1983). Hardy and Wright (1985) show that, if $\omega(n)$ is the number of distinct PRIME factors of n, then

$$\lim_{n\to\infty}\left[\frac{1}{n}\sum_{k=1}^{n}\omega(k) - \ln(\ln n)\right] = C_1. \qquad (2)$$

Furthermore, if $\Omega(n)$ is the total number of PRIME factors of n, then

$$\lim_{n\to\infty}\left[\frac{1}{n}\sum_{k=1}^{n}\Omega(k) - \ln(\ln n)\right] = C_1 + \sum_{k=1}^{\infty}\frac{1}{p_k(p_k-1)} = 1.0346538819\ldots. \qquad (3)$$

Similarly,

$$\lim_{n\to\infty}\left(\sum_{k=1}^{\pi(n)}\frac{\ln p_k}{p_k} - \ln n\right) = -\gamma - \sum_{j=2}^{\infty}\sum_{k=1}^{\infty}\frac{\ln p_k}{{p_k}^j} \equiv -C_2 = -1.3325822757\ldots. \qquad (4)$$

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/hdmrd/hdmrd.html`.
Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, 1985.
Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 24, 1983.
Rosser, J. B. and Schoenfeld, L. "Approximate Formulas for Some Functions of Prime Numbers." *Ill. J. Math.* **6**, 64–94, 1962.

Hadwiger's Principal Theorem

The VECTORS $\pm\mathbf{a}_1, \ldots, \pm\mathbf{a}_n$ in a 3-space form a normalized EUTACTIC STAR IFF $T\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in the 3-space.

Hadwiger Problem

What is the largest number of subcubes (not necessarily different) into which a CUBE cannot be divided by plane cuts? The answer is 47.

see also CUBE DISSECTION

Hafner-Sarnak-McCurley Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Given two randomly chosen INTEGER $n \times n$ matrices, what is the probability $D(n)$ that the corresponding determinants are coprime? Hafner *et al.* (1993) showed that

$$D(n) = \prod_{p_k}\left\{1 - \left[1 - \prod_{j=1}^{n}(1-{p_k}^{-j})\right]^2\right\}, \qquad (1)$$

where the product is over PRIMES. The case $D(1)$ is just the probability that two random INTEGERS are coprime,

$$D(1) = \frac{6}{\pi^2} = 0.6079271019\ldots. \qquad (2)$$

Vardi (1991) computed the limit

$$\sigma \equiv \lim_{n\to\infty} D(n) = 0.3532363719\ldots. \qquad (3)$$

The speed of convergence is roughly $\sim 0.57^n$ (Flajolet and Vardi 1996).

References
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/hafner/hafner.html`.
Flajolet, P. and Vardi, I. "Zeta Function Expansions of Classical Constants." Unpublished manuscript. 1996. `http://pauillac.inria.fr/algo/flajolet/Publications/landau.ps`.
Hafner, J. L.; Sarnak, P.; and McCurley, K. "Relatively Prime Values of Polynomials." In *Contemporary Mathematics* Vol. 143 (Ed. M. Knopp and M. Seingorn). Providence, RI: Amer. Math. Soc., 1993.
Vardi, I. *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, 1991.

Hahn-Banach Theorem

A linear FUNCTIONAL defined on a SUBSPACE of a VECTOR SPACE V and which is dominated by a sublinear function defined on V has a linear extension which is also dominated by the sublinear function.

References
Zeidler, E. *Applied Functional Analysis: Applications to Mathematical Physics.* New York: Springer-Verlag, 1995.

Hailstone Number

Sequences of INTEGERS generated in the COLLATZ PROBLEM. For example, for a starting number of 7, the sequence is 7, 22, 11, 34, 17, 52, 26, 13, 40, 20, 10, 5, 16, 8, 4, 2, 1, 4, 2, 1, ... Such sequences are called hailstone sequences because the values typically rise and fall, somewhat analogously to a hailstone inside a cloud.

While a hailstone eventually becomes so heavy that it falls to ground, every starting INTEGER ever tested has produced a hailstone sequence that eventually drops down to the number 1 and then "bounces" into the small loop 4, 2, 1, ...

see also COLLATZ PROBLEM

References
Schwartzman, S. *The Words of Mathematics: An Etymological Dictionary of Mathematical Terms Used in English.* Washington, DC: Math. Assoc. Amer., 1994.

Hairy Ball Theorem

There does not exist an everywhere NONZERO VECTOR FIELD on the 2-SPHERE $\mathbb{S}^2$. This implies that somewhere on the surface of the Earth, there is a point with zero horizontal wind velocity.

Half

The UNIT FRACTION 1/2.

see also QUARTER, SQUARE ROOT, UNIT FRACTION

Half-Closed Interval

An INTERVAL in which one endpoint is included but not the other. A half-closed interval is denoted $[a, b)$ or $(a, b]$ and is also called a HALF-OPEN INTERVAL.

see also CLOSED INTERVAL, OPEN INTERVAL

Half-Normal Distribution

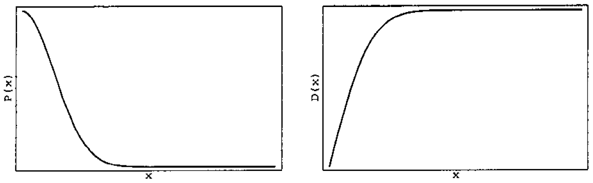

A NORMAL DISTRIBUTION with MEAN 0 and STANDARD DEVIATION $1/\theta$ limited to the domain $[0, \infty)$.

$$P(x) = \frac{2\theta}{\pi} e^{-x^2\theta^2/\pi} \qquad (1)$$

$$D(x) = \operatorname{erf}\left(\frac{tx}{\sqrt{\pi}}\right). \qquad (2)$$

The MOMENTS are

$$\mu_1 = \frac{1}{t} \qquad (3)$$

$$\mu_2 = \frac{\pi}{2t^2} \qquad (4)$$

$$\mu_3 = \frac{\pi}{t^3} \qquad (5)$$

$$\mu_4 = \frac{3\pi^2}{4t^4}, \qquad (6)$$

so the MEAN, VARIANCE, SKEWNESS, and KURTOSIS are

$$\mu = \frac{1}{\theta} \qquad (7)$$

$$\sigma^2 \equiv \mu_2 - {\mu_1}^2 = \frac{\pi - 2}{2t^2} \qquad (8)$$

$$\gamma_1 = 2\sqrt{\frac{2}{\pi}} \qquad (9)$$

$$\gamma_2 = 0. \qquad (10)$$

see also NORMAL DISTRIBUTION

Half-Open Interval

see HALF-CLOSED INTERVAL

Hall-Janko Group

The SPORADIC GROUP HJ, also denoted J_2.

see also JANKO GROUPS

Halley's Irrational Formula

A ROOT-finding ALGORITHM which makes use of a third-order TAYLOR SERIES

$$f(x)=f(x_n)+f'(x_n)(x-x_n)+\tfrac{1}{2}f''(x_n)(x-x_n)^2+\ldots. \qquad (1)$$

A ROOT of $f(x)$ satisfies $f(x)=0$, so

$$0\approx f(x_n)+f'(x_n)(x_{n+1}-x_n)+\tfrac{1}{2}f''(x_n)(x_{n+1}-x_n)^2. \qquad (2)$$

Using the QUADRATIC EQUATION then gives

$$x_{n+1}=x_n+\frac{-f'(x_n)\pm\sqrt{[f'(x_n)]^2-2f(x_n)f''(x_n)}}{f''(x_n)}. \qquad (3)$$

Picking the plus sign gives the iteration function

$$C_f(x)=x-\frac{1-\sqrt{1-\frac{2f(x)f''(x)}{[f'(x)]^2}}}{\frac{f''(x)}{f'(x)}}. \qquad (4)$$

This equation can be used as a starting point for deriving HALLEY'S METHOD.

If the alternate form of the QUADRATIC EQUATION is used instead in solving (2), the iteration function becomes instead

$$C_f(x)=x-\frac{2f(x)}{f'(x)\pm\sqrt{[f'(x)]^2-2f(x)f''(x)}}. \qquad (5)$$

This form can also be derived by setting $n=2$ in LAGUERRE'S METHOD. Numerically, the SIGN in the DENOMINATOR is chosen to maximize its ABSOLUTE VALUE. Note that in the above equation, if $f''(x)=0$, then NEWTON'S METHOD is recovered. This form of Halley's irrational formula has cubic convergence, and is usually found to be substantially more stable than NEWTON'S METHOD. However, it does run into difficulty when both $f(x)$ and $f'(x)$ or $f'(x)$ and $f''(x)$ are simultaneously near zero.

see also HALLEY'S METHOD, LAGUERRE'S METHOD, NEWTON'S METHOD

References

Qiu, H. "A Robust Examination of the Newton-Raphson Method with Strong Global Convergence Properties." Master's Thesis. University of Central Florida, 1993.

Scavo, T. R. and Thoo, J. B. "On the Geometry of Halley's Method." *Amer. Math. Monthly* **102**, 417–426, 1995.

Halley's Method

Also known as the TANGENT HYPERBOLAS METHOD or HALLEY'S RATIONAL FORMULA. As in HALLEY'S IRRATIONAL FORMULA, take the second-order TAYLOR POLYNOMIAL

$$f(x)=f(x_n)+f'(x_n)(x-x_n)+\tfrac{1}{2}f''(x_n)(x-x_n)^2+\ldots. \qquad (1)$$

A ROOT of $f(x)$ satisfies $f(x)=0$, so

$$0\approx f(x_n)+f'(x_n)(x_{n+1}-x_n)+\tfrac{1}{2}f''(x_n)(x_{n+1}-x_n)^2. \qquad (2)$$

Now write

$$0=f(x_n)+(x_{n+1}-x_n)[f'(x_n)+\tfrac{1}{2}f''(x_n)(x_{n+1}-x_n)], \qquad (3)$$

giving

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)+\frac{1}{2}f''(x_n)(x_{n+1}-x_n)}. \qquad (4)$$

Using the result from NEWTON'S METHOD,

$$x_{n+1}-x_n=-\frac{f(x_n)}{f'(x_n)}, \qquad (5)$$

gives

$$x_{n+1}=x_n-\frac{2f(x_n)f'(x_n)}{2[f'(x_n)]^2-f(x_n)f''(x_n)}, \qquad (6)$$

so the iteration function is

$$H_f(x)=x-\frac{2f(x)f'(x)}{2[f'(x)]^2-f(x)f''(x)}. \qquad (7)$$

This satisfies $H_f'(\alpha)=H_f''(\alpha)=0$ where α is a ROOT, so it is third order for simple zeros. Curiously, the third derivative

$$H_f'''(\alpha)=-\left\{\frac{f'''(\alpha)}{f'(\alpha)}-\frac{3}{2}\left[\frac{f''(\alpha)}{f'(\alpha)}\right]^2\right\} \qquad (8)$$

is the SCHWARZIAN DERIVATIVE. Halley's method may also be derived by applying NEWTON'S METHOD to $ff'^{-1/2}$. It may also be derived by using an OSCULATING CURVE of the form

$$y(x)=\frac{(x-x_n)+c}{a(x-x_n)+b}. \qquad (9)$$

Taking derivatives,

$$f(x_n)=\frac{c}{b} \qquad (10)$$

$$f'(x_n)=\frac{b-ac}{b^2} \qquad (11)$$

$$f''(x_n)=\frac{2a(ac-b)}{b^3}, \qquad (12)$$

which has solutions

$$a=-\frac{f''(x_n)}{2[f'(x_n)]^2-f(x_n)f''(x_n)} \qquad (13)$$

$$b=\frac{2f'(x_n)}{2[f'(x_n)]^2-f(x_n)f''(x_n)} \qquad (14)$$

$$c=\frac{2f(x_n)f'(x_n)}{2[f'(x_n)]^2-f(x_n)f''(x_n)}, \qquad (15)$$

so at a Root, $y(x_{n+1}) = 0$ and

$$x_{n+1} = x_n - c, \tag{16}$$

which is Halley's method.

see also Halley's Irrational Formula, Laguerre's Method, Newton's Method

References
Scavo, T. R. and Thoo, J. B. "On the Geometry of Halley's Method." *Amer. Math. Monthly* **102**, 417–426, 1995.

Halley's Rational Formula

see Halley's Method

Halphen Constant

see One-Ninth Constant

Halphen's Transformation

A curve and its polar reciprocal with regard to the fixed Conic have the same Halphen transformation.

References
Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, pp. 346–347, 1959.

Halting Problem

The determination of whether a Turing Machine will come to a halt given a particular input program. This problem is formally Undecidable, as first proved by Turing.

see also Busy Beaver, Chaitin's Constant, Turing Machine, Undecidable

References
Chaitin, G. J. "Computing the Busy Beaver Function." §4.4 in *Open Problems in Communication and Computation* (Ed. T. M. Cover and B. Gopinath). New York: Springer-Verlag, pp. 108–112, 1987.
Davis, M. "What It a Computation." In *Mathematics Today: Twelve Informal Essays* (Ed. L. A. Steen). New York: Springer-Verlag, pp. 241–267, 1978.
Penrose, R. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics.* Oxford, England: Oxford University Press, pp. 63–66, 1989.

Ham Sandwich Theorem

The volumes of any n n-D solids can always be simultaneously bisected by a $(n-1)$-D Hyperplane. Proving the theorem for $n = 2$ (where it is known as the Pancake Theorem) is simple and can be found in Courant and Robbins (1978). The theorem was proved for $n > 3$ by Stone and Tukey (1942).

see also Pancake Theorem

References
Chinn, W. G. and Steenrod, N. E. *First Concepts of Topology.* Washington, DC: Math. Assoc. Amer., 1966.
Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods.* Oxford, England: Oxford University Press, 1978.
Davis, P. J. and Hersh, R. *The Mathematical Experience.* Boston, MA: Houghton Mifflin, pp. 274–284, 1981.
Hunter, J. A. H. and Madachy, J. S. *Mathematical Diversions.* New York: Dover, pp. 67–69, 1975.
Stone, A. H. and Tukey, J. W. "Generalized 'Sandwich' Theorems." *Duke Math. J.* **9**, 356–359, 1942.

Hamilton's Equations

The equations defined by

$$\dot{q} = \frac{\partial H}{\partial p} \tag{1}$$

$$\dot{p} = -\frac{\partial H}{\partial q}, \tag{2}$$

where $\dot{x} \equiv dx/dt$ and H is the so-called Hamiltonian, are called Hamilton's equations. These equations frequently arise in problems of celestial mechanics. Another formulation related to Hamilton's equation is

$$p = \frac{\partial L}{\partial \dot{q}}, \tag{3}$$

where L is the so-called Lagrangian.

References
Morse, P. M. and Feshbach, H. "Hamilton's Principle and Classical Dynamics." §3.2 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 280–301, 1953.

Hamilton's Rules

The rules for the Multiplication of Quaternions.

see also Quaternion

Hamiltonian Circuit

A closed loop through a Graph that visits each node exactly once and ends adjacent to the initial point. The Hamiltonian circuit is named after Sir William Rowan Hamilton, who devised a puzzle in which such a path along the Edges of an Icosahedron was sought (the Icosian Game).

All Platonic Solids have a Hamiltonian circuit, as do planar 4-connected graphs. However, no foolproof method is known for determining whether a given general Graph has a Hamiltonian circuit. The number of Hamiltonian circuits on an n-Hypercube is 2, 8, 96, 43008, ... (Sloane's A006069, Gardner 1986, pp. 23–24).

see also Chvátal's Theorem, Dirac's Theorem, Euler Graph, Grinberg Formula, Hamiltonian Graph, Hamiltonian Path, Icosian Game, Kozyrev-Grinberg Theory, Ore's Theorem, Pósa's Theorem, Smith's Network Theorem

References
Chartrand, G. *Introductory Graph Theory.* New York: Dover, p. 68, 1985.
Gardner, M. "The Binary Gray Code." In *Knotted Doughnuts and Other Mathematical Entertainments.* New York: W. H. Freeman, pp. 23–24, 1986.
Sloane, N. J. A. Sequence A006069/M1903 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Hamiltonian Cycle

see Hamiltonian Circuit

Hamiltonian Graph

A Graph possessing a Hamiltonian Circuit.

see also Hamiltonian Circuit, Hamiltonian Path

References

Chartrand, G. *Introductory Graph Theory.* New York: Dover, p. 68, 1985.

Chartrand, G.; Kapoor, S. F.; and Kronk, H. V. "The Many Facets of Hamiltonian Graphs." *Math. Student* **41**, 327–336, 1973.

Hamiltonian Group

A non-Abelian Group all of whose Subgroups are self-conjugate.

References

Carmichael, R. D. "Hamiltonian Groups." §31 in *Introduction to the Theory of Groups of Finite Order.* New York: Dover, p. 113–116, 1956.

Hamiltonian Map

Consider a 1-D Hamiltonian Map of the form

$$H(p,q) = \tfrac{1}{2}p^2 + V(q), \tag{1}$$

which satisfies Hamilton's Equations

$$\dot q = \frac{\partial H}{\partial p} \tag{2}$$

$$\dot p = -\frac{\partial H}{\partial q}. \tag{3}$$

Now, write

$$\dot q_i = \frac{(q_{i+1} - q_i)}{\Delta t}, \tag{4}$$

where

$$q_i = q(t) \tag{5}$$

$$q_{i+1} = q(t + \Delta t). \tag{6}$$

Then the equations of motion become

$$q_{i+1} = q_i + p_i \Delta t \tag{7}$$

$$p_{i+1} = p_i - \Delta t \left(\frac{\partial V}{\partial q_i}\right)_{q=q_i}. \tag{8}$$

Note that equations (7) and (8) are not Area-Preserving, since

$$\frac{\partial(q_{i+1}, p_{i+1})}{\partial(q_i, p_i)} = \begin{vmatrix} 1 & -\Delta t \frac{\partial^2 V}{\partial {q_i}^2} \\ \Delta t & 1 \end{vmatrix} = 1 + (\Delta t)^2 \frac{\partial^2 V}{\partial {q_i}^2} \neq 1. \tag{9}$$

However, if we take instead of (7) and (8),

$$q_{i+1} = q_i + p_i \Delta t \tag{10}$$

$$p_{i+1} = p_i - \Delta t \left(\frac{\partial V}{\partial q_i}\right)_{q=q_{i+1}} \tag{11}$$

$$\begin{aligned} \frac{\partial(q_{i+1}, p_{i+1})}{\partial(q_i, p_i)} &= \begin{vmatrix} 1 & -\Delta t \frac{\partial}{\partial q_i}\left(\frac{\partial V}{\partial q}\right)_{q=q_{i+1}} \\ \Delta t & 1 \end{vmatrix} \\ &= 1 + (\Delta t)^2 \frac{\partial^2 V}{\partial {q_i}^2} = 1, \end{aligned} \tag{12}$$

which is Area-Preserving.

Hamiltonian Path

A loop through a Graph that visits each node exactly once but does *not* end adjacent to the initial point. The number of Hamiltonian paths on an n-Hypercube is 0, 0, 48, 48384, ... (Sloane's A006070, Gardner 1986, pp. 23–24).

see also Hamiltonian Circuit, Hamiltonian Graph

References

Gardner, M. "The Binary Gray Code." In *Knotted Doughnuts and Other Mathematical Entertainments.* New York: W. H. Freeman, pp. 23–24, 1986.

Sloane, N. J. A. Sequence A006070/M5295 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Hamiltonian System

A system of variables which can be written in the form of Hamilton's Equations.

Hammer-Aitoff Equal-Area Projection

A Map Projection whose inverse is defined using the intermediate variable

$$z \equiv \sqrt{1 - (\tfrac{1}{4}x)^2 - (\tfrac{1}{2}y)^2}.$$

Then the longitude and latitude are given by

$$\lambda = 2\tan^{-1}\left(\frac{zx}{2(2z^2 - 1)}\right)$$

$$\phi = \sin^{-1}(yz).$$

Hamming Function

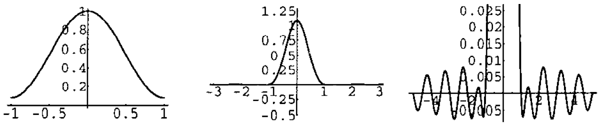

An Apodization Function chosen to minimize the height of the highest sidelobe. The Hamming function is given by

$$A(x) = 0.54 + 0.46\cos\left(\frac{\pi x}{a}\right), \tag{1}$$

Its Full Width at Half Maximum is $1.05543a$. The corresponding Instrument Function is

$$I(k) = \frac{a(1.08 - 0.64a^2k^2)\operatorname{sinc}(2\pi a k)}{1 - 4a^2k^2}. \tag{2}$$

This APODIZATION FUNCTION is close to the one produced by the requirement that the APPARATUS FUNCTION goes to 0 at $ka = 5/4$. From APODIZATION FUNCTION, a general symmetric apodization function $A(x)$ can be written as a FOURIER SERIES

$$A(x) = a_0 + 2\sum_{n=1}^{\infty} a_n \cos\left(\frac{n\pi x}{b}\right), \tag{3}$$

where the COEFFICIENTS satisfy

$$a_0 + 2\sum_{n=1}^{\infty} a_n = 1. \tag{4}$$

The corresponding apparatus function is

$$I(t) = 2b\{a_0 \operatorname{sinc}(2\pi kb) + \sum_{n=1}^{\infty}[\operatorname{sinc}(2\pi kb + n\pi) + \operatorname{sinc}(2\pi kb - n\pi)]\}. \tag{5}$$

To obtain an APODIZATION FUNCTION with zero at $ka = 3/4$, use

$$a_0 + 2a_1 = 1, \tag{6}$$

so

$$a_0 \operatorname{sinc}(\tfrac{5}{2}\pi) + a_1[\operatorname{sinc}(\tfrac{7}{2}\pi) + \operatorname{sinc}(\tfrac{3}{2}\pi) = 0 \tag{7}$$

$$(1-2a_1)\frac{2}{5\pi} - a_1\left(\frac{2}{7\pi} + \frac{2}{3\pi}\right) = (1-2a_1)\tfrac{1}{5} - a_1(\tfrac{1}{7}+\tfrac{1}{3}) = 0 \tag{8}$$

$$a_1(\tfrac{1}{7} + \tfrac{1}{3} + \tfrac{2}{5}) = \tfrac{1}{5} \tag{9}$$

$$\begin{aligned} a_1 &= \frac{\frac{1}{5}}{\frac{2}{5} + \frac{1}{7} + \frac{1}{3}} = \frac{7\cdot 3}{2\cdot 3\cdot 7 + 3\cdot 5 + 5\cdot 7} \\ &= \tfrac{21}{92} \approx 0.2283 \end{aligned} \tag{10}$$

$$\begin{aligned} a_0 &= 1 - 2a_1 = \frac{92 - 2\cdot 21}{92} = \frac{92-42}{92} \\ &= \tfrac{50}{92} = \tfrac{25}{46} \approx 0.5435. \end{aligned} \tag{11}$$

The FWHM is 1.81522, the peak is 1.08, the peak NEGATIVE and POSITIVE sidelobes (in units of the peak) are −0.00689132 and 0.00734934, respectively.

see also APODIZATION FUNCTION, HANNING FUNCTION, INSTRUMENT FUNCTION

References
Blackman, R. B. and Tukey, J. W. "Particular Pairs of Windows." In *The Measurement of Power Spectra, From the Point of View of Communications Engineering.* New York: Dover, pp. 98–99, 1959.

Handedness

Objects which are identical except for a mirror reflection are said to display handedness and to be CHIRAL.

see also AMPHICHIRAL, CHIRAL, ENANTIOMER, MIRROR IMAGE

Handkerchief Surface

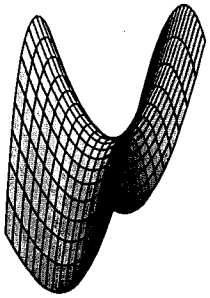

A surface given by the parametric equations

$$\begin{aligned} x(u,v) &= u \\ y(u,v) &= v \\ z(u,v) &= \tfrac{1}{3}u^3 + uv^2 + 2(u^2 - v^2). \end{aligned}$$

References
Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 628, 1993.

Handle

Handles are to MANIFOLDS as CELLS are to CW-COMPLEXES. If M is a MANIFOLD together with a $(k-1)$-SPHERE $\mathbb{S}^{k-1}$ embedded in its boundary with a trivial TUBULAR NEIGHBORHOOD, we attach a k-handle to M by gluing the tubular NEIGHBORHOOD of the $(k-1)$-SPHERE $\mathbb{S}^{k-1}$ to the TUBULAR NEIGHBORHOOD of the standard $(k-1)$-SPHERE $\mathbb{S}^{k-1}$ in the $\dim(M)$-dimensional DISK.

In this way, attaching a k-handle is essentially just the process of attaching a fattened-up k-DISK to M along the $(k-1)$-SPHERE $\mathbb{S}^{k-1}$. The embedded DISK in this new MANIFOLD is called the k-handle in the UNION of M and the handle.

see also HANDLEBODY, SURGERY, TUBULAR NEIGHBORHOOD

Handlebody

A handlebody of type (n, k) is an n-D MANIFOLD that is attained from the standard n-DISK by attaching only k-D HANDLES.

see also HANDLE, HEEGAARD SPLITTING, SURGERY

References
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 46, 1976.

Hankel Function

A COMPLEX function which is a linear combination of BESSEL FUNCTIONS OF THE FIRST and SECOND KINDS.

see also HANKEL FUNCTION OF THE FIRST KIND, HANKEL FUNCTION OF THE SECOND KIND, SPHERICAL HANKEL FUNCTION OF THE FIRST KIND, SPHERICAL HANKEL FUNCTION OF THE SECOND KIND

References

Arfken, G. "Hankel Functions." §11.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 604–610, 1985.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 623–624, 1953.

Hankel Function of the First Kind

$$H_n^{(1)}(z) \equiv J_n(z) + iY_n(z),$$

where $J_n(z)$ is a Bessel Function of the First Kind and $Y_n(z)$ is a Bessel Function of the Second Kind. Hankel functions of the first kind can be represented as a Contour Integral using

$$H_n^{(1)}(z) = \frac{1}{i\pi} \int_{0\text{ [upper half plane]}}^{\infty} \frac{e^{(z/2)(t-1/t)}}{t^{n+1}}\,dt.$$

see also Debye's Asymptotic Representation, Watson-Nicholson Formula, Weyrich's Formula

References

Arfken, G. "Hankel Functions." §11.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 604–610, 1985.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 623–624, 1953.

Hankel Function of the Second Kind

$$H_n^{(2)}(z) \equiv J_n(z) - iY_n(z),$$

where $J_n(z)$ is a Bessel Function of the First Kind and $Y_n(z)$ is a Bessel Function of the Second Kind. Hankel functions of the second kind can be represented as a Contour Integral using

$$H_n^{(2)}(z) = \frac{1}{i\pi} \int_{-\infty\text{ [lower half plane]}}^{0} \frac{e^{(z/2)(t-1/t)}}{t^{n+1}}\,dt.$$

see also Watson-Nicholson Formula

References

Arfken, G. "Hankel Functions." §11.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 604–610, 1985.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 623–624, 1953.

Hankel's Integral

$$J_m(x) = \frac{x^m}{2^{m-1}\sqrt{\pi}\,\Gamma\left(m+\frac{1}{2}\right)} \times \int_0^1 \cos(xt)(1-t^2)^{m-1/2}\,dt,$$

where $J_m(x)$ is a Bessel Function of the First Kind and $\Gamma(z)$ is the Gamma Function. Hankel's integral can be derived from Sonine's Integral.

see also Poisson Integral, Sonine's Integral

Hankel Matrix

A Matrix with identical values for each element in a given diagonal. Define H_n to be the Hankel matrix with leading column made up of the Integers $1, \ldots, n$, then

$$\mathsf{H}_2 = \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix}$$

$$\mathsf{H}_3 = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 0 \\ 3 & 0 & 0 \end{bmatrix}.$$

Hankel Transform

Equivalent to a 2-D Fourier Transform with a radially symmetric Kernel, and also called the Fourier-Bessel Transform.

$$g(u,v) = \mathcal{F}[f(r)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(r)e^{-2\pi i(ux+vy)}\,dx\,dy. \tag{1}$$

Let

$$x + iy = re^{i\theta} \tag{2}$$

$$u + iv = qe^{i\phi} \tag{3}$$

so that

$$x = r\cos\theta \tag{4}$$

$$y = r\sin\theta \tag{5}$$

$$r = \sqrt{x^2+y^2} \tag{6}$$

$$u = q\cos\phi \tag{7}$$

$$v = q\sin\phi \tag{8}$$

$$q = \sqrt{u^2+v^2}. \tag{9}$$

Then

$$\begin{aligned} g(q) &= \int_0^\infty \int_0^{2\pi} f(r)e^{-2\pi irq(\cos\phi\cos\theta+\sin\phi\sin\theta)} r\,dr\,d\theta \\ &= \int_0^\infty \int_0^{2\pi} f(r)e^{-2\pi irq\cos(\theta-\phi)} r\,dr\,d\theta \\ &= \int_0^\infty \int_{-\phi}^{2\pi-\phi} f(r)e^{-2\pi irq\cos\theta} r\,dr\,d\theta \\ &= \int_0^\infty \int_0^{2\pi} f(r)e^{-2\pi irq\cos\theta} r\,dr\,d\theta \\ &= \int_0^\infty f(r)\left[\int_0^{2\pi} e^{-2\pi irq\cos\theta}\,d\theta\right] r\,dr \\ &= 2\pi \int_0^\infty f(r)J_0(2\pi qr)r\,dr, \end{aligned} \tag{10}$$

where $J_0(z)$ is a zeroth order BESSEL FUNCTION OF THE FIRST KIND. Therefore, the Hankel transform pairs are

$$g(k) = \int_0^\infty f(x) J_0(kx) x\, dx \tag{11}$$

$$f(x) = \int_0^\infty g(k) J_0(kx) k\, dk. \tag{12}$$

see also BESSEL FUNCTION OF THE FIRST KIND, FOURIER TRANSFORM, LAPLACE TRANSFORM

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 795, 1985.

Bracewell, R. *The Fourier Transform and Its Applications.* New York: McGraw-Hill, pp. 244–250, 1965.

Hann Function

see HANNING FUNCTION

Hanning Function

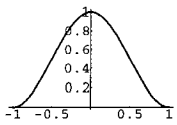

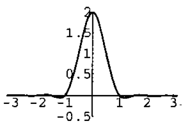

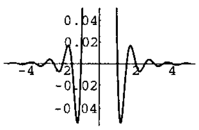

An APODIZATION FUNCTION, also called the HANN FUNCTION, frequently used to reduce ALIASING in FOURIER TRANSFORMS. The illustrations above show the Hanning function, its INSTRUMENT FUNCTION, and a blowup of the INSTRUMENT FUNCTION sidelobes. The Hanning function is given by

$$f(x) = \cos^2\left(\frac{\pi x}{2a}\right) = \tfrac{1}{2} - \tfrac{1}{2}\cos\left(\frac{\pi x}{a}\right). \tag{1}$$

The INSTRUMENT FUNCTION for Hanning apodization can also be written

$$a[\operatorname{sinc}(2\pi k a) + \tfrac{1}{2}\operatorname{sinc}(2\pi k a - \pi) + \tfrac{1}{2}\operatorname{sinc}(2\pi k a + \pi)]. \tag{2}$$

Its FULL WIDTH AT HALF MAXIMUM is a. It has APPARATUS FUNCTION

$$\begin{aligned} A(x) &= \int_{-a}^{a} \left[\tfrac{1}{2} - \tfrac{1}{2}\cos\left(\frac{\pi x}{a}\right)\right] e^{-2\pi i k x}\, dx \\ &= \tfrac{1}{2}\int_{-a}^{a} e^{-2\pi i k x}\, dx - \tfrac{1}{2}\int_{-a}^{a} e^{-2\pi i k x}\, dx \\ &\equiv \tfrac{1}{2}(A_1 + A_2). \end{aligned} \tag{3}$$

The first integral is

$$I_1 = \int_{-a}^{a} e^{-2\pi i k x}\, dx = \frac{\sin(2\pi k a)}{\pi k} = 2a \operatorname{sinc}(2\pi k a). \tag{4}$$

The second integral can be rewritten

$$\begin{aligned} I_2 &= \int_{-a}^{0} \cos\left(\frac{\pi x}{a}\right) e^{-2\pi i k x}\, dx \\ &\quad + \int_0^a \cos\left(\frac{\pi x}{a}\right) e^{-2\pi i k x}\, dx \\ &= \int_0^a \cos\left(\frac{\pi x}{a}\right) (e^{2\pi i k x} + e^{-2\pi i k x})\, dx \\ &= 2\int_0^a \cos\left(\frac{\pi x}{a}\right) \cos(2\pi k x)\, dx \\ &= 2\left\{\frac{\sin\left(\frac{\pi}{a} - 2\pi k\right)x}{2\left(\frac{\pi}{a} - 2\pi k\right)} + \frac{\sin\left(\frac{\pi}{a} + 2\pi k\right)x}{2\left(\frac{\pi}{a} + 2\pi k\right)}\right\}_0^a \\ &= a\left[\frac{\sin(\pi - 2\pi k a)}{\pi - 2\pi k a} + \frac{\sin(\pi + 2\pi k a)}{\pi + 2\pi k a}\right] \\ &= \frac{a}{\pi}\left[\frac{\sin(2\pi k a)}{1 - 2ka} - \frac{\sin(2\pi k a)}{1 + 2ka}\right] \\ &= a[\operatorname{sinc}(\pi - 2\pi k a) + \operatorname{sinc}(\pi + 2\pi k a)]. \end{aligned} \tag{5}$$

Combining (4) and (5) gives

$$\begin{aligned} A(x) = a[&\operatorname{sinc}(2\pi k a) + \tfrac{1}{2}\operatorname{sinc}(\pi - 2\pi k a) \\ &+ \tfrac{1}{2}\operatorname{sinc}(\pi + 2\pi k a)]. \end{aligned} \tag{6}$$

To find the extrema, define $x \equiv 2\pi k a$ and rewrite (6) as

$$A(x) = a[\sin x + \tfrac{1}{2}\operatorname{sinc}(x - \pi) + \tfrac{1}{2}\operatorname{sinc}(x + \pi)]. \tag{7}$$

Then solve

$$\begin{aligned} \frac{dA}{dx} &= \frac{\pi^2(-x^3\cos x + 3x^2\sin x + \pi^2 x\cos x - \pi^2\sin x)}{x^2(\pi^2 - x^2)^2} \\ &= 0 \end{aligned} \tag{8}$$

to find the extrema. The roots are $x = 7.42023$ and 10.7061, giving a peak NEGATIVE sidelobe of -0.026708 and a peak POSITIVE sidelobe (in units of a) of 0.00843441. The peak in units of a is 1, and the full-width at half maximum is given by setting (7) equal to 1/2 and solving for x, yielding

$$x_{1/2} = 2\pi k_{1/2} a = \pi. \tag{9}$$

Therefore, with $L \equiv 2a$, the FULL WIDTH AT HALF MAXIMUM is

$$\text{FWHM} = 2k_{1/2} = \frac{1}{a} = \frac{2}{L}. \tag{10}$$

see also APODIZATION FUNCTION, HAMMING FUNCTION

Hanoi Graph

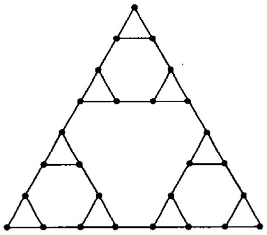

A GRAPH H_n arising in conjunction with the TOWERS OF HANOI problem. The above figure is the Hanoi graph H_3.

see also TOWERS OF HANOI

Hanoi Towers

see TOWERS OF HANOI

Hansen-Bessel Formula

$$\begin{aligned} J_n(z) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{iz\cos t}e^{in(t-\pi/2)}\,dt \\ &= \frac{i^{-n}}{\pi}\int_0^{\pi} e^{iz\cos t}cos(nt)\,dt \\ &= \frac{1}{\pi}\int_0^{\pi}\cos(z\sin t - nt)\,dt \end{aligned}$$

for $n = 0, 1, 2, \ldots$, where $J_n(z)$ is a BESSEL FUNCTION OF THE FIRST KIND.

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1472, 1980.

Hansen Chain

An ADDITION CHAIN for which there is a SUBSET H of members such that each member of the chain uses the largest element of H which is less than the member.

see also ADDITION CHAIN, BRAUER CHAIN, HANSEN NUMBER

References

Guy, R. K. "Addition Chains. Brauer Chains. Hansen Chains." §C6 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 111–113, 1994.

Hansen Number

A number n for which a shortest chain exists which is a HANSEN CHAIN is called a Hansen number.

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 111–112, 1994.

Hansen's Problem

A SURVEYING PROBLEM: from the position of two known but inaccessible points A and B, determine the position of two unknown accessible points P and P' by bearings from A, B, P' to P and A, B, P to P'.

see also SURVEYING PROBLEMS

References

Dörrie, H. "Annex to a Survey." §40 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 193–197, 1965.

Happy Number

Let the sum of the SQUARES of the DIGITS of a POSITIVE INTEGER s_0 be represented by s_1. In a similar way, let the sum of the SQUARES of the DIGITS of s_1 be represented by s_2, and so on. If some $s_i = 1$ for $i \geq 1$, then the original INTEGER s_0 is said to be happy.

Once it is known whether a number is happy (or not), then any number in the sequence s_1, s_2, s_3, ... will also be happy (or not). A number which is not happy is called UNHAPPY. Unhappy numbers have EVENTUALLY PERIODIC sequences of s_i 4, 16, 37, 58, 89, 145, 42, 20, 4, ... which do not reach 1.

Any PERMUTATION of the DIGITS of an UNHAPPY or happy number must also be unhappy or happy. This follows from the fact that ADDITION is COMMUTATIVE. The first few happy numbers are 1, 7, 10, 13, 19, 23, 28, 31, 32, 44, 49, 68, 70, 79, 82, 86, 91, 94, 97, 100, ... (Sloane's A007770). These are also the numbers whose 2-RECURRING DIGITAL INVARIANT sequences have period 1.

see also KAPREKAR NUMBER, RECURRING DIGITAL INVARIANT , UNHAPPY NUMBER

References

Dudeney, H. E. Problem 143 in *536 Puzzles & Curious Problems.* New York: Scribner, pp. 43 and 258–259, 1967.

Guy, R. K. "Happy Numbers." §E34 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 234–235, 1994.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 163–165, 1979.

Schwartzman, S. *The Words of Mathematics: An Etymological Dictionary of Mathematical Terms Used in English.* Washington, DC: Math. Assoc. Amer., 1994.

Sloane, N. J. A. Sequence A007770 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Harada-Norton Group

The SPORADIC GROUP *HN.*

References

Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas/HN.html`.

Harary Graph

The smallest k-connected GRAPH with n VERTICES.

Hard Hexagon Entropy Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

A constant related to the HARD SQUARE ENTROPY CONSTANT. This constant is given by

$$\kappa_h \equiv \lim_{N\to\infty} [G(N)]^{1/N} = 1.395485972\ldots, \qquad (1)$$

where $G(N)$ is the number of configurations of nonattacking KINGS on an $n \times n$ chessboard with regular hexagonal cells, where $N \equiv n^2$. Amazingly, κ_h is algebraic and given by

$$\kappa_h \equiv \kappa_1\kappa_2\kappa_3\kappa_4, \qquad (2)$$

where

$$\kappa_1 \equiv 4^{-1}3^{5/4}11^{-5/12}c^{-2} \qquad (3)$$

$$\kappa_2 \equiv [1-\sqrt{1-c}+\sqrt{2+c+2\sqrt{1+c+c^2}}\,]^2 \qquad (4)$$

$$\kappa_3 \equiv [-1-\sqrt{1-c}+\sqrt{2+c+2\sqrt{1+c+c^2}}\,]^2 \qquad (5)$$

$$\kappa_4 \equiv [\sqrt{1-a}+\sqrt{2+a+2\sqrt{1+a+a^2}}\,]^{-1/2} \qquad (6)$$

$$a \equiv -\tfrac{124}{363}11^{1/3} \qquad (7)$$

$$b \equiv \tfrac{2501}{11979}33^{1/2} \qquad (8)$$

$$c \equiv \{\tfrac{1}{4}+\tfrac{3}{8}a[(b+1)^{1/3}-(b-1)^{1/3}]\}^{1/3}. \qquad (9)$$

(Baxter 1980, Joyce 1988).

References

Baxter, R. J. "Hard Hexagons: Exact Solution." *J. Physics A* **13**, 1023–1030, 1980.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/square/square.html`.

Joyce, G. S. "On the Hard Hexagon Model and the Theory of Modular Functions." *Phil. Trans. Royal Soc. London A* **325**, 643–702, 1988.

Plouffe, S. "Hard Hexagons Constant." `http://lacim.uqam.ca/piDATA/hardhex.html`.

Hard Square Entropy Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let $F(n^2)$ be the number of binary $n\times n$ MATRICES with no adjacent 1s (in either columns or rows). Define $N \equiv n^2$, then the hard square entropy constant is defined by

$$\kappa \equiv \lim_{N\to\infty} [F(N)]^{1/N} = 1.503048082\ldots.$$

The quantity $\ln\kappa$ arises in statistical physics (Baxter *et al.* 1980, Pearce and Seaton 1988), and is known as the entropy per site of hard squares. A related constant known as the HARD HEXAGON ENTROPY CONSTANT can also be defined.

References

Baxter, R. J.; Enting, I. G.; and Tsang, S. K. "Hard-Square Lattice Gas." *J. Statist. Phys.* **22**, 465–489, 1980.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/square/square.html`.

Pearce, P. A. and Seaton, K. A. "A Classical Theory of Hard Squares." *J. Statist. Phys.* **53**, 1061–1072, 1988.

Hardy's Inequality

Let $\{a_n\}$ be a NONNEGATIVE SEQUENCE and $f(x)$ a NONNEGATIVE integrable FUNCTION. Define

$$A_n = \sum_{k=1}^{n} a_k \qquad (1)$$

$$B_n = \sum_{k=n}^{\infty} a_k \qquad (2)$$

and

$$F(x) = \int_0^x f(t)\,dt \qquad (3)$$

$$G(x) = \int_x^\infty f(t)\,dt, \qquad (4)$$

and take $p > 1$. For sums,

$$\sum_{n=1}^{\infty}\left(\frac{A_n}{n}\right)^p < \left(\frac{p}{p-1}\right)^p \sum_{n=1}^{\infty}(a_n)^p \qquad (5)$$

(unless all $a_n = 0$), and for integrals,

$$\int_0^\infty \left[\frac{F(x)}{x}\right]^p dx < \left(\frac{p}{p-1}\right)^p \int_0^\infty [f(x)]^p\,dx \qquad (6)$$

(unless f is identically 0).

References

Hardy, G. H.; Littlewood, J. E.; and Pólya, G. *Inequalities, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 239–243, 1988.

Mitrinovic, D. S.; Pecaric, J. E.; and Fink, A. M. *Inequalities Involving Functions and Their Integrals and Derivatives.* New York: Kluwer, 1991.

Opic, B. and Kufner, A. *Hardy-Type Inequalities.* Essex, England: Longman, 1990.

Hardy-Littlewood Conjectures

The first Hardy-Littlewood conjecture is called the k-TUPLE CONJECTURE. It states that the asymptotic number of PRIME CONSTELLATIONS can be computed explicitly.

The second Hardy-Littlewood conjecture states that

$$\pi(x+y) - \pi(x) \leq \pi(y)$$

for all x and y, where $\pi(x)$ is the PRIME COUNTING FUNCTION. Although it is not obvious, Richards (1974) proved that this conjecture is incompatible with the first Hardy-Littlewood conjecture.

see also PRIME CONSTELLATION, PRIME COUNTING FUNCTION

References

Richards, I. "On the Incompatibility of Two Conjectures Concerning Primes." *Bull. Amer. Math. Soc.* **80**, 419–438, 1974.

Riesel, H. *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, pp. 61–62 and 68–69, 1994.

Hardy-Littlewood Constants

see PRIME CONSTELLATION

Hardy-Littlewood Tauberian Theorem

Let $a_n \geq 0$ and suppose

$$\sum_{n=1}^{\infty} a_n e^{-an} \sim \frac{1}{a}$$

as $a \to 0^+$. Then

$$\sum_{n \leq x} a_n \sim x$$

as $x \to \infty$.

see also TAUBERIAN THEOREM

References

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 118–119, 1994.

Hardy-Littlewood k-Tuple Conjecture

see PRIME PATTERNS CONJECTURE

Hardy-Ramanujan Number

The smallest nontrivial TAXICAB NUMBER, i.e., the smallest number representable in two ways as a sum of two CUBES. It is given by

$$1729 = 1^3 + 12^3 = 9^3 + 10^3.$$

The number derives its name from the following story G. H. Hardy told about Ramanujan. "Once, in the taxi from London, Hardy noticed its number, 1729. He must have thought about it a little because he entered the room where Ramanujan lay in bed and, with scarcely a hello, blurted out his disappointment with it. It was, he declared, 'rather a dull number,' adding that he hoped that wasn't a bad omen. 'No, Hardy,' said Ramanujan, 'it is a very interesting number. It is the smallest number expressible as the sum of two [POSITIVE] cubes in two different ways'" (Hofstadter 1989, Kanigel 1991, Snow 1993).

see also DIOPHANTINE EQUATION—CUBIC, TAXICAB NUMBER

References

Guy, R. K. "Sums of Like Powers. Euler's Conjecture." §D1 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 139–144, 1994.

Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, p. 68, 1959.

Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, p. 564, 1989.

Kanigel, R. *The Man Who Knew Infinity: A Life of the Genius Ramanujan.* New York: Washington Square Press, p. 312, 1991.

Snow, C. P. Foreword to Hardy, G. H. *A Mathematician's Apology, reprinted with a foreword by C. P. Snow.* New York: Cambridge University Press, p. 37, 1993.

Hardy-Ramanujan Theorem

Let $\omega(n)$ be the number of DISTINCT PRIME FACTORS of n. If $\Psi(x)$ tends steadily to infinity with x, then

$$\ln\ln x - \Psi(x)\sqrt{\ln\ln x} < \omega(n) < \ln\ln x + \Psi(x)\sqrt{\ln\ln x}$$

for ALMOST ALL numbers $n < x$. "ALMOST ALL" means here the frequency of those INTEGERS n in the interval $1 \leq n \leq x$ for which

$$|\omega(n) - \ln\ln x| > \Psi(x)\sqrt{\ln\ln x}$$

approaches 0 as $x \to \infty$.

see also DISTINCT PRIME FACTORS, ERDŐS-KAC THEOREM

Hardy's Rule

Let the values of a function $f(x)$ be tabulated at intervals equally spaced by h about x_0, so that $f_{-3} = f(x_0 - 3h)$, $f_{-2} = f(x_0 - 2h)$, etc. Then Hardy's rule gives the approximation to the integral of $f(x)$ as

$$\int_{x_0-3h}^{x_0+3h} f(x)\,dx = \tfrac{1}{100}h(28f_{-3} + 162f_{-2} + 22f_0 + 162f_2 + 28f_3) + \tfrac{9}{1400}h^7[2f^{(4)}(\xi_2) - h^2 f^{(8)}(\xi_1)],$$

where the final term gives the error, with $\xi_1, \xi_2 \in [x_0 - 3h, x_0 + 3h]$.

see also BODE'S RULE, DURAND'S RULE, NEWTON-COTES FORMULAS, SIMPSON'S 3/8 RULE, SIMPSON'S RULE, TRAPEZOIDAL RULE, WEDDLE'S RULE

Harmonic Addition Theorem

To convert an equation of the form

$$f(\theta) = a\cos\theta + b\sin\theta \qquad (1)$$

to the form

$$f(\theta) = c\cos(\theta + \delta), \qquad (2)$$

expand (2) using the trigonometric addition formulas to obtain

$$f(\theta) = c\cos\theta\cos\delta - c\sin\theta\sin\delta. \qquad (3)$$

Now equate the COEFFICIENTS of (1) and (3)

$$a = c\cos\delta \qquad (4)$$

$$b = -c\sin\delta, \qquad (5)$$

so

$$\tan\delta = -\frac{b}{a} \qquad (6)$$

$$a^2 + b^2 = c^2, \qquad (7)$$

and we have

$$\delta = \tan^{-1}\left(-\frac{b}{a}\right) \qquad (8)$$

$$c = \sqrt{a^2 + b^2}. \qquad (9)$$

Given two general sinusoidal functions with frequency ω:

$$\psi_1 = A_1 \sin(\omega t + \delta_1) \tag{10}$$

$$\psi_2 = A_2 \sin(\omega t + \delta_2), \tag{11}$$

their sum ψ can be expressed as a sinusoidal function with frequency ω

$$\begin{aligned}\psi \equiv \psi_1 + \psi_2 &= A_1[\sin(\omega t)\cos\delta_1 + \sin\delta_1 \cos(\omega t)] \\ &\quad + A_2[\sin(\omega t)\cos\delta_2 + \sin\delta_2\cos(\omega t)] \\ &= [A_1\cos\delta_1 + A_2\cos\delta_2]\sin(\omega t) \\ &\quad + [A_1\sin\delta_1 + A_2\sin\delta_2]\cos(\omega t).\end{aligned} \tag{12}$$

Now, define

$$A\cos\delta \equiv A_1\cos\delta_1 + A_2\cos\delta_2 \tag{13}$$

$$A\sin\delta \equiv A_1\sin\delta_1 + A_2\sin\delta_2. \tag{14}$$

Then (12) becomes

$$A\cos\delta\sin(\omega t) + A\sin\delta\cos(\omega t) = A\sin(\omega t + \delta). \tag{15}$$

Square and add (13) and (14)

$$A_2 = {A_1}^2 + {A_2}^2 + 2A_1A_2\cos(\delta_2 - \delta_1). \tag{16}$$

Also, divide (14) by (13)

$$\tan\delta = \frac{A_1\sin\delta_1 + A_2\sin\delta_2}{A_1\cos\delta_1 + A_2\cos\delta_2}, \tag{17}$$

so

$$\psi = A\sin(\omega t + \delta), \tag{18}$$

where A and δ are defined by (16) and (17).

This procedure can be generalized to a sum of n harmonic waves, giving

$$\psi = \sum_{i=1}^{n} A_i \cos(\omega t + \delta_i) = A\cos(\omega t + \delta), \tag{19}$$

where

$$A^2 \equiv \sum_{i=1}^{n} {A_i}^2 + 2\sum_{j>i}^{n}\sum_{i=1}^{n} A_iA_j\cos(\delta_i - \delta_j) \tag{20}$$

and

$$\tan\delta = -\frac{\sum_{i=1}^{n} A_i\sin\delta_i}{\sum_{i=1}^{n} A_i\cos\delta_i}. \tag{21}$$

Harmonic Analysis

see also FOURIER SERIES

Harmonic Brick

A right-angled PARALLELEPIPED with dimensions $a \times ab \times abc$, where a, b, and c are INTEGERS.

see also BRICK, DE BRUIJN'S THEOREM, EULER BRICK

Harmonic Conjugate Function

The harmonic conjugate to a given function $u(x,y)$ is a function $v(x,y)$ such that

$$f(x,y) = u(x,y) + v(x,y)$$

is COMPLEX DIFFERENTIABLE (i.e., satisfies the CAUCHY-RIEMANN EQUATIONS). It is given by

$$v(z) = \int ux\,dy - uy\,dx.$$

Harmonic Conjugate Points

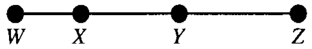

Given COLLINEAR points W, X, Y, and Z, Y and Z are harmonic conjugates with respect to W and X if

$$\frac{|WY|}{|YX|} = \frac{|WZ|}{|ZX|}.$$

The distances between such points are said to be in HARMONIC RATIO, and the LINE SEGMENT depicted above is called a HARMONIC SEGMENT.

Harmonic conjugate points are also defined for a TRIANGLE. If W and X have TRILINEAR COORDINATES $\alpha : \beta : \gamma$ and $\alpha' : \beta' : \gamma'$, then the TRILINEAR COORDINATES of the harmonic conjugates are

$$Y = \alpha + \alpha' : \beta + \beta' : \gamma + \gamma'$$

$$Z = \alpha - \alpha' : \beta - \beta' : \gamma - \gamma'$$

(Kimberling 1994).

see also HARMONIC RANGE, HARMONIC RATIO

References

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 13–14, 1990.

Phillips, A. W. and Fisher, I. *Elements of Geometry.* New York: American Book Co., 1896.

Wells, D. *The Penguin Dictionary of Curious and Interesting Geometry.* New York: Viking Penguin, p. 92, 1992.

Harmonic Coordinates

Satisfy the condition

$$\Gamma^\lambda \equiv g^{\mu\nu}{\Gamma_{\mu\nu}}^\lambda = 0, \tag{1}$$

or equivalently,

$$\frac{\partial}{\partial x^\kappa}\left(\sqrt{g}\, g^{\lambda\kappa}\right) = 0. \tag{2}$$

It is always possible to choose such a system. Using the D'ALEMBERTIAN OPERATOR,

$$\Box^2\phi \equiv (g^{\lambda\kappa}\phi_{;\lambda})_{;\kappa} = g^{\lambda\kappa}\frac{\partial^2\phi}{\partial x^\lambda \partial x^\kappa} - \Gamma^\lambda \frac{\partial\phi}{\partial x^\lambda}. \qquad (3)$$

But since $\Gamma^\lambda \equiv 0$ for harmonic coordinates,

$$\Box^2 x^\mu = 0. \qquad (4)$$

Harmonic Decomposition

A POLYNOMIAL function in the elements of $\mathbf{x}$ can be uniquely decomposed into a sum of harmonic POLYNOMIALS times POWERS of $|\mathbf{x}|$.

Harmonic Divisor Number

A number n for which the HARMONIC MEAN of the DIVISORS of n, i.e., $nd(n)/\sigma(n)$, is an INTEGER, where $d(n)$ is the number of POSITIVE integral DIVISORS of n and $\sigma(n)$ is the DIVISOR FUNCTION. For example, the divisors of $n = 140$ are 1, 2, 4, 5, 7, 10, 14, 20, 28, 35, 70, and 140, giving

$$\begin{aligned} d(140) &= 12 \\ \sigma(140) &= 336 \\ \frac{140d(140)}{\sigma(140)} &= \frac{140 \cdot 12}{335} = 5, \end{aligned}$$

so 140 is a harmonic divisor number. Harmonic divisor numbers are also called ORE NUMBERS. Garcia (1954) gives the 45 harmonic divisor numbers less than 10^7. The first few are 1, 6, 140, 270, 672, 1638, ... (Sloane's A007340).

For distinct PRIMES p and q, harmonic divisor numbers are equivalent to EVEN PERFECT NUMBERS for numbers of the form $p^r q$. Mills (1972) proved that if there exists an ODD POSITIVE harmonic divisor number n, then n has a prime-POWER factor greater than 10^7.

Another type of number called "harmonic" is the HARMONIC NUMBER.

see also DIVISOR FUNCTION, HARMONIC NUMBER

References

Edgar, H. M. W. "Harmonic Numbers." *Amer. Math. Monthly* **99**, 783–789, 1992.

Garcia, M. "On Numbers with Integral Harmonic Mean." *Amer. Math. Monthly* **61**, 89–96, 1954.

Guy, R. K. "Almost Perfect, Quasi-Perfect, Pseudoperfect, Harmonic, Weird, Multiperfect and Hyperperfect Numbers." §B2 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 45–53, 1994.

Mills, W. H. "On a Conjecture of Ore." *Proceedings of the 1972 Number Theory Conference.* University of Colorado, Boulder, pp. 142–146, 1972.

Ore, Ø. "On the Averages of the Divisors of a Number." *Amer. Math. Monthly* **55**, 615–619, 1948.

Pomerance, C. "On a Problem of Ore: Harmonic Numbers." Unpublished manuscript, 1973.

Sloane, N. J. A. Sequences A007340/M4299 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sloane, N. J. A. and Plouffe, S. Extended entry in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.

Zachariou, A. and Zachariou, E. "Perfect, Semi-Perfect and Ore Numbers." *Bull. Soc. Math. Grèce (New Ser.)* **13**, 12–22, 1972.

Harmonic Equation

see LAPLACE'S EQUATION

Harmonic Function

Any real-valued function $u(x, y)$ with continuous second PARTIAL DERIVATIVES which satisfies LAPLACE'S EQUATION

$$\nabla^2 u(x, y) = 0 \qquad (1)$$

is called a harmonic function. Harmonic functions are called POTENTIAL FUNCTIONS in physics and engineering. Potential functions are extremely useful, for example, in electromagnetism, where they reduce the study of a 3-component VECTOR FIELD to a 1-component SCALAR FUNCTION. A scalar harmonic function is called a SCALAR POTENTIAL, and a vector harmonic function is called a VECTOR POTENTIAL.

To find a class of such functions in the PLANE, write the LAPLACE'S EQUATION in POLAR COORDINATES

$$u_{rr} + \frac{1}{r}u_r + \frac{1}{r^2}u_{\theta\theta} = 0, \qquad (2)$$

and consider only radial solutions

$$u_{rr} + \frac{1}{r}u_r = 0. \qquad (3)$$

This is integrable by quadrature, so define $v \equiv du/dr$,

$$\frac{dv}{dr} + \frac{1}{r}v = 0 \qquad (4)$$

$$\frac{dv}{v} = -\frac{dr}{r} \qquad (5)$$

$$\ln\left(\frac{v}{A}\right) = -\ln r \qquad (6)$$

$$\frac{v}{A} = \frac{1}{r} \qquad (7)$$

$$v = \frac{du}{dr} = \frac{A}{r} \qquad (8)$$

$$du = A\frac{dr}{r}, \qquad (9)$$

so the solution is

$$u = A \ln r. \qquad (10)$$

Ignoring the trivial additive and multiplicative constants, the general pure radial solution then becomes

$$u = \ln[(x-a)^2+(y-b)^2]^{1/2} = \tfrac{1}{2}\ln\left[(x-a)^2 + (y-b)^2\right]. \qquad (11)$$

Other solutions may be obtained by differentiation, such as

$$u = \frac{x-a}{(x-a)^2+(y-b)^2} \tag{12}$$

$$v = \frac{y-b}{(x-a)^2+(y-b)^2}, \tag{13}$$

$$u = e^x \sin y \tag{14}$$

$$v = e^x \cos y, \tag{15}$$

and

$$\tan^{-1}\left(\frac{y-b}{x-a}\right). \tag{16}$$

Harmonic functions containing azimuthal dependence include

$$u = r^n \cos(n\theta) \tag{17}$$

$$v = r^n \sin(n\theta). \tag{18}$$

The POISSON KERNEL

$$u(r,R,\theta,\phi) = \frac{R^2-r^2}{R^2-2rR\cos(\theta-\phi)+r^2} \tag{19}$$

is another harmonic function.

see also SCALAR POTENTIAL, VECTOR POTENTIAL

References

Ash, J. M. (Ed.) *Studies in Harmonic Analysis.* Washington, DC: Math. Assoc. Amer., 1976.

Axler, S.; Pourdon, P.; and Ramey, W. *Harmonic Function Theory.* Springer-Verlag, 1992.

Benedetto, J. J. *Harmonic Analysis and Applications.* Boca Raton, FL: CRC Press, 1996.

Cohn, H. *Conformal Mapping on Riemann Surfaces.* New York: Dover, 1980.

Harmonic-Geometric Mean

Let

$$\alpha_{n+1} = \frac{2\alpha_n\beta_n}{\alpha_n+\beta_n}$$

$$\beta_{n+1} = \sqrt{\alpha_n\beta_n},$$

then

$$H(\alpha_0,\beta_0) \equiv \lim_{n\to\infty} a_n = \frac{1}{M({\alpha_0}^{-1},{\beta_0}^{-1})},$$

where M is the ARITHMETIC-GEOMETRIC MEAN.

see also ARITHMETIC MEAN, ARITHMETIC-GEOMETRIC MEAN, GEOMETRIC MEAN, HARMONIC MEAN

Harmonic Homology

A PERSPECTIVE COLLINEATION with center O and axis o not incident is called a HOMOLOGY. A HOMOLOGY is said to be harmonic if the points A and A' on a line through O are harmonic conjugates with respect to O and $o \cdot a$. Every PERSPECTIVE COLLINEATION of period two is a harmonic homology.

see also HOMOLOGY (GEOMETRY), PERSPECTIVE COLLINEATION

References

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, p. 248, 1969.

Harmonic Logarithm

For all INTEGERS n and NONNEGATIVE INTEGERS t, the harmonic logarithms $\lambda_n^{(t)}(x)$ of order t and degree n are defined as the unique functions satisfying

1. $\lambda_0^{(t)}(x) = (\ln x)^t$,
2. $\lambda_n^{(t)}(x)$ has no constant term except $\lambda_0^{(0)}(x) = 1$,
3. $\frac{d}{dx}\lambda_n^{(t)}(x) = \lfloor n\rceil\, \lambda_{n-1}^{(t)}(x)$,

where the "ROMAN SYMBOL" $\lfloor n\rceil$ is defined by

$$\lfloor n\rceil \equiv \begin{cases} n & \text{for } n \neq 0 \\ 1 & \text{for } n = 0 \end{cases} \tag{1}$$

(Roman 1992). This gives the special cases

$$\lambda_n^{(0)}(x) = \begin{cases} x^n & \text{for } n \geq 0 \\ 0 & \text{for } n < 0 \end{cases} \tag{2}$$

$$\lambda_n^{(1)}(x) = \begin{cases} x^n(\ln x - H_n) & \text{for } n \geq 0 \\ x^n & \text{for } n < 0, \end{cases} \tag{3}$$

where H_n is a HARMONIC NUMBER

$$H_n \equiv \sum_{k=1}^{n} \frac{1}{k}. \tag{4}$$

The harmonic logarithm has the INTEGRAL

$$\int \lambda_n^{(1)}(x)\,dx = \frac{1}{\lfloor n+1\rceil}\lambda_{n+1}^{(1)}(x). \tag{5}$$

The harmonic logarithm can be written

$$\lambda_n^{(t)}(x) = \lfloor n\rceil!\tilde{D}^{-n}(\ln x)^t, \tag{6}$$

where $\tilde{D}$ is the DIFFERENTIAL OPERATOR, (so $\tilde{D}^{-n}$ is the nth INTEGRAL). Rearranging gives

$$\tilde{D}^k\lambda_n^{(t)}(x) = \left\lfloor \frac{\lfloor n\rceil!}{\lfloor n-k} \right\rceil !\lambda_{n-k}^{(t)}(x). \tag{7}$$

This formulation gives an analog of the BINOMIAL THEOREM called the LOGARITHMIC BINOMIAL FORMULA. Another expression for the harmonic logarithm is

$$\lambda_n^{(t)}(x) = x^n \sum_{j=0}^{t} (-1)^j (t)_j c_n^{(j)} (\ln x)^{t-j}, \tag{8}$$

where $(t)_j = t(t-1)\cdots(t-j+1)$ is a POCHHAMMER SYMBOL and $c_n^{(j)}$ is a two-index HARMONIC NUMBER (Roman 1992).

see also LOGARITHM, ROMAN FACTORIAL

References

Loeb, D. and Rota, G.-C. "Formal Power Series of Logarithmic Type." *Advances Math.* **75**, 1–118, 1989.

Roman, S. "The Logarithmic Binomial Formula." *Amer. Math. Monthly* **99**, 641–648, 1992.

Harmonic Map

A harmonic map between RIEMANNIAN MANIFOLDS can be viewed as a generalization of a GEODESIC when the domain DIMENSION is one, or of a HARMONIC FUNCTION when the range is a EUCLIDEAN SPACE.

see also BOCHNER IDENTITY, EUCLIDEAN SPACE, GEODESIC, HARMONIC FUNCTION, RIEMANNIAN MANIFOLD

References

Burstal, F.; Lemaire, L.; and Rawnsley, J. "Harmonic Maps Bibliography." `http://www.bath.ac.uk/~masfeb/harmonic.html`.

Eels, J. and Lemaire, L. "A Report on Harmonic Maps." *Bull. London Math. Soc.* **10**, 1–68, 1978.

Eels, J. and Lemaire, L. "Another Report on Harmonic Maps." *Bull. London Math. Soc.* **20**, 385–524, 1988.

Harmonic Mean

The harmonic mean $H(x_1, \ldots, x_n)$ of n points x_i (where $i = 1, \ldots, n$) is

$$\frac{1}{H} \equiv \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}. \qquad (1)$$

The special case of $n = 2$ is therefore

$$\frac{1}{H} = \frac{1}{2}\left(\frac{1}{x_1}+\frac{1}{x_2},\right) \qquad (2)$$

or

$$\frac{1}{H} = \frac{x_1+x_2}{2x_1x_2}. \qquad (3)$$

The VOLUME-to-SURFACE AREA ratio for a cylindrical container with height h and radius r and the MEAN CURVATURE of a general surface are related to the harmonic mean.

Hoehn and Niven (1985) show that

$$H(a_1+c, a_2+c, \ldots, a_n+c) > c + H(a_1, a_2, \ldots, a_n) \qquad (4)$$

for any POSITIVE constant c.

see also ARITHMETIC MEAN, ARITHMETIC-GEOMETRIC MEAN, GEOMETRIC MEAN, HARMONIC-GEOMETRIC MEAN, ROOT-MEAN-SQUARE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 10, 1972.

Hoehn, L. and Niven, I. "Averages on the Move." *Math. Mag.* **58**, 151–156, 1985.

Harmonic Mean Index

The statistical INDEX

$$P_H \equiv \frac{\sum p_0 q_0}{\sum \frac{{p_0}^2 q_0}{p_n}},$$

where p_n is the price per unit in period n, q_n is the quantity produced in period n, and $v_n \equiv p_n q_n$ the value of the n units.

see also INDEX

References

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, p. 69, 1962.

Harmonic Number

A number of the form

$$H_n = \sum_{k=1}^{n}\frac{1}{k}. \qquad (1)$$

This can be expressed analytically as

$$H_n = \gamma + \psi_0(n+1), \qquad (2)$$

where γ is the EULER-MASCHERONI CONSTANT and $\Psi(x) = \psi_0(x)$ is the DIGAMMA FUNCTION. The number formed by taking alternate signs in the sum also has an analytic solution

$$\begin{aligned} H'_n &= \sum_{k=1}^{n}\frac{(-1)^{k+1}}{k} && (3)\\ &= \ln 2 + \tfrac{1}{2}(-1)^n[\psi_0(\tfrac{1}{2}n+\tfrac{1}{2}) - \psi_0(\tfrac{1}{2}n+1)]. && (4)\end{aligned}$$

The first few harmonic numbers H_n are 1, 3/2, 11/6, 25/12, 137/60, ... (Sloane's A001008 and A002805). The HARMONIC NUMBER H_n is never an INTEGER (except for H_1), which can be proved by using the strong triangle inequality to show that the 2-ADIC VALUE of H_n is greater than 1 for $n > 1$. The harmonic numbers have ODD NUMERATORS and EVEN DENOMINATORS. The nth harmonic number is given asymptotically by

$$H_n \sim \ln n + \gamma + \frac{1}{2n}, \qquad (5)$$

where γ is the EULER-MASCHERONI CONSTANT (Conway and Guy 1996). Gosper gave the interesting identity

$$\sum_{i=0}^{\infty}\frac{z^i H_i}{i!} = -e^z\sum_{k=1}^{\infty}\frac{(-z)^k}{kk!} = e^z[\ln z + \Gamma(0,z) + \gamma], \qquad (6)$$

where $\Gamma(0,z)$ is the incomplete GAMMA FUNCTION and γ is the EULER-MASCHERONI CONSTANT. Borwein and Borwein (1995) show that

$$\sum_{n=1}^{\infty} \frac{{H_n}^2}{(n+1)^2} = \tfrac{11}{4}\zeta(4) = \tfrac{11}{360}\pi^4 \tag{7}$$

$$\sum_{n=1}^{\infty} \frac{{H_n}^2}{n^2} = \tfrac{17}{4}\zeta(4) = \tfrac{17}{360}\pi^4 \tag{8}$$

$$\sum_{n=1}^{\infty} \frac{H_n}{n^3} = \tfrac{5}{4}\zeta(4) = \tfrac{1}{72}\pi^4, \tag{9}$$

where $\zeta(z)$ is the RIEMANN ZETA FUNCTION. The first of these had been previously derived by de Doelder (1991), and the last by Euler (1775). These identities are corollaries of the identity

$$\frac{1}{\pi}\int_0^{\pi} x^2\{\ln[2\cos(\tfrac{1}{2}x)]\}^2\,dx = \tfrac{11}{2}\zeta(4) = \tfrac{11}{180}\pi^4 \tag{10}$$

(Borwein and Borwein 1995). Additional identities due to Euler are

$$\sum_{n=1}^{\infty} \frac{H_n}{n^2} = 2\zeta(3) \tag{11}$$

$$2\sum_{n=1}^{\infty} \frac{H_n}{n^m} = (m+2)\zeta(m+1) - \sum_{n=1}^{m-2} \zeta(m-n)\zeta(n+1) \tag{12}$$

for $m = 2, 3, \ldots$ (Borwein and Borwein 1995), where $\zeta(3)$ is APÉRY'S CONSTANT. These sums are related to so-called EULER SUMS.

Conway and Guy (1996) define the second harmonic number by

$$H_n^{(2)} \equiv \sum_{i=1}^{n} H_i = (n+1)(H_{n+1}-1) = (n+1)(H_{n+1}-H_1), \tag{13}$$

the third harmonic number by

$$H_n^{(3)} \equiv \sum_{i=1}^{n} H_i^{(2)} = \binom{n+2}{2}(H_{n+2} - H_2), \tag{14}$$

and the nth harmonic number by

$$H_n^{(k)} = \binom{n+k-1}{k-1}(H_{n+k-1} - H_{k-1}). \tag{15}$$

A slightly different definition of a two-index harmonic number $c_n^{(j)}$ is given by Roman (1992) in connection with the HARMONIC LOGARITHM. Roman (1992) defines this by

$$c_n^{(0)} = \begin{cases} 1 & \text{for } n \geq 0 \\ 0 & \text{for } n < 0 \end{cases} \tag{16}$$

$$c_0^{(j)} = \begin{cases} 1 & \text{for } j = 0 \\ 0 & \text{for } j \neq 0 \end{cases} \tag{17}$$

plus the recurrence relation

$$cn_n^{(j)} = c_n^{(j-1)} + nc_{n-1}^{(j)}. \tag{18}$$

For general $n > 0$ and $j > 0$, this is equivalent to

$$c_n^{(j)} = \sum_{i=1}^{n} \frac{1}{i} c_i^{(j-1)}, \tag{19}$$

and for $n > 0$, it simplifies to

$$c_n^{(j)} = \sum_{i=1}^{n} \binom{n}{i} (-1)^{i-1} i^{-j}. \tag{20}$$

For $n < 0$, the harmonic number can be written

$$c_n^{(j)} = (-1)^j \lfloor n \rceil ! s(-n, j), \tag{21}$$

where $\lfloor n \rceil !$ is the ROMAN FACTORIAL and s is a STIRLING NUMBER OF THE FIRST KIND.

A separate type of number sometimes also called a "harmonic number" is a HARMONIC DIVISOR NUMBER (or ORE NUMBER).

see also APÉRY'S CONSTANT, EULER SUM, HARMONIC LOGARITHM, HARMONIC SERIES, ORE NUMBER

References

Borwein, D. and Borwein, J. M. "On an Intriguing Integral and Some Series Related to $\zeta(4)$." *Proc. Amer. Math. Soc.* **123**, 1191–1198, 1995.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 143 and 258–259, 1996.

de Doelder, P. J. "On Some Series Containing $\Psi(x) - \Psi(y)$ and $(\Psi(x) - \Psi(y))^2$ for Certain Values of x and y." *J. Comp. Appl. Math.* **37**, 125–141, 1991.

Roman, S. "The Logarithmic Binomial Formula." *Amer. Math. Monthly* **99**, 641–648, 1992.

Sloane, N. J. A. Sequences A001008/M2885 and A002805/M1589 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Harmonic Progression

see Harmonic Series

Harmonic Range

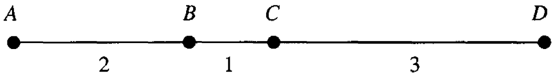

A set of four COLLINEAR points A, B, C, and D arranged such that

$$AB : BC = 2 : 1$$

$$AD : DC = 6 : 3.$$

Hardy (1967) uses the term HARMONIC SYSTEM OF POINTS to refer to a harmonic range.

see also EULER LINE, GERGONNE LINE, HARMONIC CONJUGATE POINTS, SODDY LINE

References

Hardy, G. H. *A Course of Pure Mathematics, 10th ed.* Cambridge, England: Cambridge University Press, pp. 99 and 106, 1967.

Harmonic Ratio

see HARMONIC CONJUGATE POINTS

Harmonic Segment

see HARMONIC CONJUGATE POINTS

Harmonic Series

The SUM

$$\sum_{k=1}^{\infty} \frac{1}{k} \tag{1}$$

is called the harmonic series. It can be shown to DIVERGE using the INTEGRAL TEST by comparison with the function $1/x$. The divergence, however, is very slow. In fact, the sum

$$\sum_{p} \frac{1}{p} \tag{2}$$

taken over all PRIMES also diverges. The generalization of the harmonic series

$$\zeta(n) \equiv \sum_{k=1}^{\infty} \frac{1}{k^n} \tag{3}$$

is known as the RIEMANN ZETA FUNCTION.

The sum of the first few terms of the harmonic series is given analytically by the nth HARMONIC NUMBER

$$H_n = \sum_{j=1}^{n} \frac{1}{j} = \gamma + \psi_0(n+1), \tag{4}$$

where γ is the EULER-MASCHERONI CONSTANT and $\Psi(x) = \psi_0(x)$ is the DIGAMMA FUNCTION. The number of terms needed to exceed 1, 2, 3, ... are 1, 4, 11, 31, 83, 227, 616, 1674, 4550, 12367, 33617, 91380, 248397, ... (Sloane's A004080). Using the analytic form shows that after 2.5×10^8 terms, the sum is still less than 20. Furthermore, to achieve a sum greater than 100, more than 1.509×10^{43} terms are needed!

Progressions of the form

$$\frac{1}{a_1}, \frac{1}{a_1+d}, \frac{1}{a_1+2d}, \ldots \tag{5}$$

are also sometimes called harmonic series (Beyer 1987).

The modified harmonic series, given by the sum

$$T = \sum_{k=1}^{\infty} \frac{1}{p_k}, \tag{6}$$

where p_k is the kth PRIME, diverges.

see also ARITHMETIC SERIES, BERNOULLI'S PARADOX, BOOK STACKING PROBLEM, EULER SUM, ZIPF'S LAW

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 279–280, 1985.

Beyer, W. H. (Ed.). *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 8, 1987.

Boas, R. P. and Wrench, J. W. "Partial Sums of the Harmonic Series." *Amer. Math. Monthly* **78**, 864–870, 1971.

Honsberger, R. "An Intriguing Series." Ch. 10 in *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., pp. 98–103, 1976.

Sloane, N. J. A. Sequence A004080 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Harmonic System of Points

see HARMONIC RANGE

Harmonious Graph

A connected LABELLED GRAPH with n EDGES in which all VERTICES can be labeled with distinct INTEGERS (mod n) so that the sums of the PAIRS of numbers at the ends of each EDGE are also distinct (mod n). The LADDER GRAPH, FAN, WHEEL GRAPH, PETERSEN GRAPH, TETRAHEDRAL GRAPH, DODECAHEDRAL GRAPH, and ICOSAHEDRAL GRAPH are all harmonious (Graham and Sloane 1980).

see also GRACEFUL GRAPH, LABELLED GRAPH, POSTAGE STAMP PROBLEM, SEQUENTIAL GRAPH

References

Gallian, J. A. "Open Problems in Grid Labeling." *Amer. Math. Monthly* **97**, 133–135, 1990.

Gardner, M. *Wheels, Life, and other Mathematical Amusements.* New York: W. H. Freeman, p. 164, 1983.

Graham, R. L. and Sloane, N. "On Additive Bases and Harmonious Graphs." *SIAM J. Algebraic Discrete Math.* **1**, 382–404, 1980.

Guy, R. K. "The Corresponding Modular Covering Problem. Harmonious Labelling of Graphs." §C13 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 127–128, 1994.

Harnack's Inequality

Let $D = D(z_0, R)$ be an OPEN DISK, and let u be a HARMONIC FUNCTION on D such that $u(z) \geq 0$ for all $z \in D$. Then for all $z \in D$, we have

$$0 \leq u(z) \leq \left(\frac{R}{R - |z - z_0|}\right)^2 u(z_0).$$

see also LIOUVILLE'S CONFORMALITY THEOREM

References

Flanigan, F. J. "Harnack's Inequality." §2.5.1 in *Complex Variables: Harmonic and Analytic Functions.* New York: Dover, pp. 88–90, 1983.

Harnack's Theorems

Harnack's first theorem states that a real irreducible curve of order n cannot have more than

$$\tfrac{1}{2}(n-1)(n-2)-\sum s_i(s_i-1)+1$$

circuits (Coolidge 1959, p. 57).

Harnack's second theorem states that there exists a curve of every order with the maximum number of circuits compatible with that order and with a certain number of double points, provided that number is not permissible for a curve of lower order (Coolidge 1959, p. 61).

References
Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, 1959.

Harshad Number

A Positive Integer which is Divisible by the sum of its Digits, also called a Niven Number (Kennedy *et al.* 1980). The first few are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 18, 20, 21, 24, ... (Sloane's A005349). Grundman (1994) proved that there is no sequence of more than 20 consecutive Harshad numbers, and found the smallest sequence of 20 consecutive Harshad numbers, each member of which has 44,363,342,786 digits.

Grundman (1994) defined an n-Harshad (or n-Niven) number to be a Positive Integer which is Divisible by the sum of its digits in base $n \geq 2$. Cai (1996) showed that for $n = 2$ or 3, there exists an infinite family of sequences of consecutive n-Harshad numbers of length $2n$.

References
Cai, T. "On 2-Niven Numbers and 3-Niven Numbers." *Fib. Quart.* **34**, 118–120, 1996.
Cooper, C. N. and Kennedy, R. E. "Chebyshev's Inequality and Natural Density." *Amer. Math. Monthly* **96**, 118–124, 1989.
Cooper, C. N. and Kennedy, R. "On Consecutive Niven Numbers." *Fib. Quart.* **21**, 146–151, 1993.
Grundman, H. G. "Sequences of Consecutive n-Niven Numbers." *Fib. Quart.* **32**, 174–175, 1994.
Kennedy, R.; Goodman, R.; and Best, C. "Mathematical Discovery and Niven Numbers." *MATYC J.* **14**, 21–25, 1980.
Sloane, N. J. A. Sequence A005349/M0481 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Vardi, I. "Niven Numbers." §2.3 in *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 19 and 28–31, 1991.

Hart's Inversor

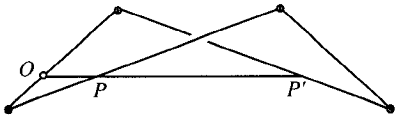

A linkage which draws the inverse of a given curve. It can also convert circular to linear motion. The rods satisfy $AB = CD$ and $BC = DA$, and O, P, and P' remain Collinear. Coxeter (1969, p. 428) shows that if $AO = \mu AB$, then

$$OP \times OP' = \mu(1-\mu)(AD^2 - AB^2).$$

see also Peaucellier Inversor

References
Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods.* Oxford, England: Oxford University Press, p. 157, 1978.
Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 82–83, 1969.
Rademacher, H. and Toeplitz, O. *The Enjoyment of Mathematics: Selections from Mathematics for the Amateur.* Princeton, NJ: Princeton University Press, pp. 124–129, 1957.

Hart's Theorem

Any one of the eight Apollonius Circles of three given Circles is Tangent to a Circle C, as are the other three Apollonius Circles having (1) like contact with two of the given Circles and (2) unlike contact with the third.

see also Apollonius Circles

References
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 127–128, 1929.

Hartley Transform

An Integral Transform which shares some features with the Fourier Transform, but which (in the discrete case), multiplies the Kernel by

$$\cos\left(\frac{2\pi kn}{N}\right)-\sin\left(\frac{2\pi kn}{N}\right) \tag{1}$$

instead of

$$e^{-2\pi ikn/N}=\cos\left(\frac{2\pi kn}{N}\right)-i\sin\left(\frac{2\pi kn}{N}\right). \tag{2}$$

The Hartley transform produces Real output for a Real input, and is its own inverse. It therefore can have computational advantages over the Discrete Fourier Transform, although analytic expressions are usually more complicated for the Hartley transform.

The discrete version of the Hartley transform can be written explicitly as

$$\begin{aligned}\mathcal{H}[a] &\equiv \frac{1}{\sqrt{N}}\sum_{n=0}^{N-1} a_n\left[\cos\left(\frac{2\pi kn}{N}\right)-\sin\left(\frac{2\pi kn}{N}\right)\right] && (3)\\ &= \Re\mathcal{F}[a]-\Im\mathcal{F}[a], && (4)\end{aligned}$$

where $\mathcal{F}$ denotes the Fourier Transform. The Hartley transform obeys the Convolution property

$$\mathcal{H}[a*b]_k=\tfrac{1}{2}(A_kB_k-\bar{A}_k\bar{B}_k+A_k\bar{B}_k+\bar{A}_kB_k), \tag{5}$$

where

$$\bar{a}_0 \equiv a_0 \tag{6}$$
$$\bar{a}_{n/2} \equiv a_{n/2} \tag{7}$$
$$\bar{a}_k \equiv a_{n-k} \tag{8}$$

(Arndt). Like the FAST FOURIER TRANSFORM, there is a "fast" version of the Hartley transform. A decimation in time algorithm makes use of

$$\mathcal{H}_n^{\text{left}}[a] = \mathcal{H}_{n/2}[a^{\text{even}}] + \mathcal{X}\mathcal{H}_{n/2}[a^{\text{odd}}] \tag{9}$$
$$\mathcal{H}_n^{\text{right}}[a] = \mathcal{H}_{n/2}[a^{\text{even}}] - \mathcal{X}\mathcal{H}_{n/2}[a^{\text{odd}}], \tag{10}$$

where $\mathcal{X}$ denotes the sequence with elements

$$a_n \cos\left(\frac{\pi n}{N}\right) - \bar{a}_n \sin\left(\frac{\pi n}{N}\right). \tag{11}$$

A decimation in frequency algorithm makes use of

$$\mathcal{H}_n^{\text{even}}[a] = \mathcal{H}_{n/2}[a^{\text{left}} + a^{\text{right}}], \tag{12}$$
$$\mathcal{H}_n^{\text{odd}}[a] = \mathcal{H}_{n/2}[\mathcal{X}(a^{\text{left}} - a^{\text{right}})]. \tag{13}$$

The DISCRETE FOURIER TRANSFORM

$$A_k \equiv \mathcal{F}[a] = \sum_{n=0}^{N-1} e^{-2\pi i k n/N} a_n \tag{14}$$

can be written

$$\begin{bmatrix} A_k \\ A_{-k} \end{bmatrix} = \sum_{n=0}^{N-1} \underbrace{\begin{bmatrix} e^{-2\pi i k n/N} & 0 \\ 0 & e^{2\pi i k n/N} \end{bmatrix}}_{\mathsf{F}} \begin{bmatrix} a_n \\ a_n \end{bmatrix} \tag{15}$$

$$= \sum_{n=0}^{N-1} \underbrace{\frac{1}{2}\begin{bmatrix} 1-i & 1+i \\ 1+i & 1-i \end{bmatrix}}_{\mathsf{T}^{-1}} \underbrace{\begin{bmatrix} \cos\left(\frac{2\pi k n}{N}\right) & \sin\left(\frac{2\pi k n}{N}\right) \\ -\sin\left(\frac{2\pi k n}{N}\right) & \cos\left(\frac{2\pi k n}{N}\right) \end{bmatrix}}_{\mathsf{H}}$$
$$\times \underbrace{\frac{1}{2}\begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix}}_{\mathsf{T}} \begin{bmatrix} a_n \\ a_n \end{bmatrix}, \tag{16}$$

so

$$\mathsf{F} = \mathsf{T}^{-1}\mathsf{H}\mathsf{T}. \tag{17}$$

see also DISCRETE FOURIER TRANSFORM, FAST FOURIER TRANSFORM, FOURIER TRANSFORM

References

Arndt, J. "The Hartley Transform (HT)." Ch. 2 in "Remarks on FFT Algorithms." `http://www.jjj.de/fxt/`.

Bracewell, R. N. *The Fourier Transform and Its Applications.* New York: McGraw-Hill, 1965.

Bracewell, R. N. *The Hartley Transform.* New York: Oxford University Press, 1986.

HashLife

A LIFE ALGORITHM that achieves remarkable speed by storing subpatterns in a hash table, and using them to skip forward, sometimes thousands of generations at a time. HashLife takes tremendous amounts of memory and can't show patterns at every step, but can quickly calculate the outcome of a pattern that takes millions of generations to complete.

References

Hensel, A. "A Brief Illustrated Glossary of Terms in Conway's Game of Life." `http://www.cs.jhu.edu/~callahan/glossary.html`.

Hasse's Algorithm

see COLLATZ PROBLEM

Hasse's Conjecture

Define the ZETA FUNCTION of a VARIETY over a NUMBER FIELD by taking the product over all PRIME IDEALS of the ZETA FUNCTIONS of this VARIETY reduced modulo the PRIMES. Hasse conjectured that this product has a MEROMORPHIC continuation over the whole plane and a functional equation.

References

Lang, S. "Some History of the Shimura-Taniyama Conjecture." *Not. Amer. Math. Soc.* **42**, 1301–1307, 1995.

Hasse-Davenport Relation

Let F be a FINITE FIELD with q elements, and let F_s be a FIELD containing F such that $[F_s : F] = s$. Let χ be a nontrivial MULTIPLICATIVE CHARACTER of F and $\chi' = \chi \circ N_{F_s/F}$ a character of F_s. Then

$$(-g(\chi))^s = -g(\chi'),$$

where $g(x)$ is a GAUSSIAN SUM.

see also GAUSSIAN SUM, MULTIPLICATIVE CHARACTER

References

Ireland, K. and Rosen, M. "A Proof of the Hasse-Davenport Relation." §11.4 in *A Classical Introduction to Modern Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 162–165, 1990.

Hasse Diagram

A graphical rendering of a PARTIALLY ORDERED SET displayed via the COVER relation of the PARTIALLY ORDERED SET with an implied upward orientation. A point is drawn for each element of the POSET, and line segments are drawn between these points according to the following two rules:

1. If $x < y$ in the poset, then the point corresponding to x appears lower in the drawing than the point corresponding to y.
2. The line segment between the points corresponding to any two elements x and y of the poset is included in the drawing IFF x covers y or y covers x.

Hasse diagrams are also called UPWARD DRAWINGS.

Hasse-Minkowski Theorem

Two nonsingular forms are equivalent over the rationals IFF they have the same DETERMINANT and the same p-SIGNATURES for all p.

Hasse Principle

A collection of equations satisfies the Hasse principle if, whenever one of the equations has solutions in $\mathbb{R}$ and all the $\mathbb{Q}_p$, then the equations have solutions in the RATIONALS $\mathbb{Q}$. Examples include the set of equations

$$ax^2 + bxy + cy^2 = 0$$

with a, b, and c INTEGERS, and the set of equations

$$x^2 + y^2 = a$$

for a rational. The trivial solution $x = y = 0$ is usually not taken into account when deciding if a collection of homogeneous equations satisfies the Hasse principle. The Hasse principle is sometimes called the LOCAL-GLOBAL PRINCIPLE.

see also LOCAL FIELD

Hasse's Resolution Modulus Theorem

The JACOBI SYMBOL $(a/y) = \chi(y)$ as a CHARACTER can be extended to the KRONECKER SYMBOL $(f(a)/y) = \chi^*(y)$ so that $\chi^*(y) = \chi(y)$ whenever $\chi(y) \neq 0$. When y is RELATIVELY PRIME to $f(a)$, then $\chi^*(y) \neq 0$, and for NONZERO values $\chi^*(y_1) = \chi^*(y_2)$ IFF $y_1 \equiv y_2 \text{ mod}^+ f(a)$. In addition, $|f(a)|$ is the minimum value for which the latter congruence property holds in any extension symbol for $\chi(y)$.

see also CHARACTER (NUMBER THEORY), JACOBI SYMBOL, KRONECKER SYMBOL

References
Cohn, H. *Advanced Number Theory.* New York: Dover, pp. 35–36, 1980.

Hat

The hat is a caret-shaped symbol most commonly used to denote a UNIT VECTOR ($\hat{\mathbf{v}}$) or an ESTIMATOR ($\hat{x}$).

see also ESTIMATOR, UNIT VECTOR

Haupt-Exponent

The smallest exponent e for which $b^e \equiv 1 \pmod{p}$, where b and p are given numbers, is the haupt-exponent of $b \pmod{p}$. The number of bases having a haupt-exponent e is $\phi(e)$, where $\phi(e)$ is the TOTIENT FUNCTION. Cunningham (1922) published the haupt-exponents for primes to 25409 and bases 2, 3, 5, 6, 7, 10, 11, and 12.

see also COMPLETE RESIDUE SYSTEM, RESIDUE INDEX

References
Cunningham, A. *Haupt-Exponents, Residue Indices, Primitive Roots.* London: F. Hodgson, 1922.

Hausdorff Axioms

Describe subsets of elements x in a NEIGHBORHOOD SET E of x. The NEIGHBORHOOD is assumed to satisfy:

1. There corresponds to each point x at least one NEIGHBORHOOD $U(x)$, and each NEIGHBORHOOD $U(x)$ contains the point x.
2. If $U(x)$ and $V(x)$ are two NEIGHBORHOODS of the same point x, there must exist a NEIGHBORHOOD $W(x)$ that is a subset of both.
3. If the point y lies in $U(x)$, there must exist a NEIGHBORHOOD $U(y)$ that is a SUBSET of $U(x)$.
4. For two different points x and y, there are two corresponding NEIGHBORHOODS $U(x)$ and $U(y)$ with no points in common.

Hausdorff-Besicovitch Dimension

see CAPACITY DIMENSION

Hausdorff Dimension

Let A be a SUBSET of a METRIC SPACE X. Then the Hausdorff dimension $D(A)$ of A is the INFIMUM of $d \geq 0$ such that the d-dimensional HAUSDORFF MEASURE of A is 0. Note that this need not be an INTEGER.

In many cases, the Hausdorff dimension correctly describes the correction term for a resonator with FRACTAL PERIMETER in Lorentz's conjecture. However, in general, the proper dimension to use turns out to be the MINKOWSKI-BOULIGAND DIMENSION (Schroeder 1991).

see also CAPACITY DIMENSION, FRACTAL DIMENSION, MINKOWSKI-BOULIGAND DIMENSION

References
Federer, H. *Geometric Measure Theory.* New York: Springer-Verlag, 1969.
Hausdorff, F. "Dimension und äußeres Maß." *Math. Ann.* **79**, 157–179, 1919.
Ott, E. "Appendix: Hausdorff Dimension." *Chaos in Dynamical Systems.* New York: Cambridge University Press, pp. 100–103, 1993.
Schroeder, M. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise.* New York: W. H. Freeman, pp. 41–45, 1991.

Hausdorff Measure

Let X be a METRIC SPACE, A be a SUBSET of X, and d a number ≥ 0. The d-dimensional Hausdorff measure of A, $H^d(A)$, is the INFIMUM of POSITIVE numbers y such that for every $r > 0$, A can be covered by a countable family of closed sets, each of diameter less than r, such that the sum of the dth POWERS of their diameters is less than y. Note that $H^d(A)$ may be infinite, and d need not be an INTEGER.

References
Federer, H. *Geometric Measure Theory.* New York: Springer-Verlag, 1969.
Ott, E. *Chaos in Dynamical Systems.* Cambridge, England: Cambridge University Press, p. 103, 1993.

Hausdorff Paradox

For $n \geq 3$, there exist no additive finite and invariant measures for the group of displacements in $\mathbb{R}^n$.

References

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 49, 1983.

Hausdorff Space

A TOPOLOGICAL SPACE in which any two points have disjoint NEIGHBORHOODS.

Haversine

$$\operatorname{hav}(z) \equiv \tfrac{1}{2}\operatorname{vers}(z) = \tfrac{1}{2}(1-\cos z),$$

where $\operatorname{vers}(z)$ is the VERSINE and cos is the COSINE. Using a trigonometric identity, the haversine is equal to

$$\operatorname{hav}(z) = \sin^2(\tfrac{1}{2}z).$$

see also COSINE, COVERSINE, EXSECANT, VERSINE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 78, 1972.

Heads Minus Tails Distribution

A fair COIN is tossed $2n$ times. Let $D \equiv |H-T|$ be the absolute difference in the number of heads and tails obtained. Then the probability distribution is given by

$$P(D=2k) = \begin{cases} (\frac{1}{2})^{2n}\binom{2n}{n} & k=0 \\ 2(\frac{1}{2})^{2n}\binom{2n}{n+k} & k=1,\,2,\,\ldots, \end{cases}$$

where $P(D=2k-1)=0$. The most probable value of D is $D=2$, and the expectation value is

$$\langle D\rangle = \frac{n\binom{2n}{n}}{2^{2n-1}}.$$

see also BERNOULLI DISTRIBUTION, COIN, COIN TOSSING

References

Handelsman, M. B. Solution to Problem 436, "Distributing 'Heads' Minus 'Tails.'" *College Math. J.* **22**, 444–446, 1991.

Heap

A SET of N members forms a heap if it satisfies $a_{\lfloor j/2\rfloor} \geq a_j$ for $1 \leq \lfloor j/2\rfloor < j \leq N$, where $\lfloor x\rfloor$ is the FLOOR FUNCTION.

see also HEAPSORT

Heapsort

An $N\lg N$ SORTING ALGORITHM which is not quite as fast as QUICKSORT. It is a "sort-in-place" algorithm and requires no auxiliary storage, which makes it particularly concise and elegant to implement.

see also QUICKSORT, SORTING

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Heapsort." §8.3 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 327–329, 1992.

Heart Surface

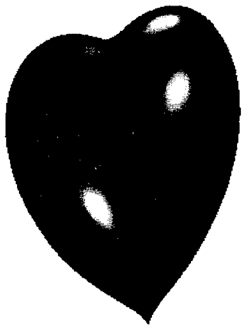

A heart-shaped surface given by the SEXTIC EQUATION

$$(2x^2+2y^2+z^2-1)^3 - \tfrac{1}{10}x^2z^3 - y^2z^3 = 0.$$

see also BONNE PROJECTION, PIRIFORM

References

Nordstrand, T. "Heart." `http://www.uib.no/people/nfytn/hearttxt.htm`.

Heat Conduction Equation

A diffusion equation of the form

$$\frac{\partial T}{\partial t} = \kappa\nabla^2 T. \tag{1}$$

Physically, the equation commonly arises in situations where κ is the thermal diffusivity and T the temperature.

The 1-D heat conduction equation is

$$\frac{\partial T}{\partial t} = \kappa\frac{\partial^2 T}{\partial x^2}. \tag{2}$$

This can be solved by SEPARATION OF VARIABLES using

$$T(x,t) = X(x)T(t). \tag{3}$$

Then

$$X\frac{dT}{dt} = \kappa T\frac{d^2X}{dx^2}. \tag{4}$$

Dividing both sides by κXT gives

$$\frac{1}{\kappa T}\frac{dT}{dt} = \frac{1}{X}\frac{d^2X}{dx^2} = -\frac{1}{\lambda^2}, \tag{5}$$

where each side must be equal to a constant. Anticipating the exponential solution in T, we have picked a negative separation constant so that the solution remains finite at all times and λ has units of length. The T solution is

$$T(t) = Ae^{-\kappa t/\lambda^2}, \tag{6}$$

and the X solution is

$$X(x) = C\cos\left(\frac{x}{\lambda}\right) + D\sin\left(\frac{x}{\lambda}\right). \tag{7}$$

The general solution is then

$$\begin{aligned} T(x,t) &= T(t)X(x) \\ &= Ae^{-\kappa t/\lambda^2}\left[C\cos\left(\frac{x}{\lambda}\right) + D\sin\left(\frac{x}{\lambda}\right)\right] \\ &= e^{-\kappa t/\lambda^2}\left[D\cos\left(\frac{x}{\lambda}\right) + E\sin\left(\frac{x}{\lambda}\right)\right]. \end{aligned} \tag{8}$$

If we are given the boundary conditions

$$T(0,t) = 0 \tag{9}$$

and

$$T(L,t) = 0, \tag{10}$$

then applying (9) to (8) gives

$$D\cos\left(\frac{x}{\lambda}\right) = 0 \Rightarrow D = 0, \tag{11}$$

and applying (10) to (8) gives

$$E\sin\left(\frac{L}{\lambda}\right) = 0 \Rightarrow \frac{L}{\lambda} = n\pi \Rightarrow \lambda = \frac{L}{n\pi}, \tag{12}$$

so (8) becomes

$$T_n(x,t) = E_n e^{-\kappa(n\pi/L)^2 t}\sin\left(\frac{n\pi x}{L}\right). \tag{13}$$

Since the general solution can have any n,

$$T(x,t) = \sum_{n=1}^{\infty} c_n \sin\left(\frac{n\pi x}{L}\right) e^{-\kappa(n\pi/L)^2 t}. \tag{14}$$

Now, if we are given an initial condition $T(x,0)$, we have

$$T(x,0) = \sum_{n=1}^{\infty} c_n \sin\left(\frac{n\pi x}{L}\right). \tag{15}$$

Multiplying both sides by $\sin(m\pi x/L)$ and integrating from 0 to L gives

$$\int_0^L \sin\left(\frac{m\pi x}{L}\right) T(x,0)\,dx = \int_0^L \sum_{n=1}^{\infty} c_n \sin\left(\frac{m\pi x}{L}\right)\sin\left(\frac{n\pi x}{L}\right) dx. \tag{16}$$

Using the ORTHOGONALITY of $\sin(nx)$ and $\sin(mx)$,

$$\begin{aligned} \sum_{n=1}^{\infty} c_n \int_0^L \sin\left(\frac{n\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right) dx &= \sum_{n=1}^{\infty} \tfrac{1}{2}\pi\delta_{mn}c_n \\ = \tfrac{1}{2}\pi c_m &= \int_0^L \sin\left(\frac{m\pi x}{L}\right) T(x,0)\,dx, \end{aligned} \tag{17}$$

so

$$c_n = \frac{2}{\pi}\int_0^L \sin\left(\frac{m\pi x}{L}\right) T(x,0)\,dx. \tag{18}$$

If the boundary conditions are replaced by the requirement that the derivative of the temperature be zero at the edges, then (9) and (10) are replaced by

$$\left.\frac{\partial T}{\partial x}\right|_{(0,t)} = 0 \tag{19}$$

$$\left.\frac{\partial T}{\partial x}\right|_{(L,t)} = 0. \tag{20}$$

Following the same procedure as before, a similar answer is found, but with sine replaced by cosine:

$$T(x,t) = \sum_{n=1}^{\infty} c_n \cos\left(\frac{n\pi x}{L}\right) e^{-\kappa(n\pi/L)^2 t}, \tag{21}$$

where

$$c_n = \frac{2}{\pi}\int_0^L \cos\left(\frac{m\pi x}{L}\right) \left.\frac{\partial T(x,0)}{\partial x}\right|_{t=0} dx. \tag{22}$$

Heat Conduction Equation—Disk

To solve the HEAT CONDUCTION EQUATION on a 2-D disk of radius $R = 1$, try to separate the equation using

$$T(r,\theta,t) = R(r)\Theta(\theta)T(t). \tag{1}$$

Writing the θ and r terms of the LAPLACIAN in SPHERICAL COORDINATES gives

$$\nabla^2 = \frac{d^2R}{dr^2} + \frac{2}{r}\frac{dR}{dr} + \frac{1}{r^2}\frac{d^2\Theta}{d\theta^2}, \tag{2}$$

so the HEAT CONDUCTION EQUATION becomes

$$\frac{R\Theta}{\kappa}\frac{d^2T}{dt^2} = \frac{d^2R}{dr^2}\Theta T + \frac{2}{r}\frac{dR}{dr}\Theta T + \frac{1}{r^2}\frac{d^2\Theta}{d\theta^2}RT. \tag{3}$$

Multiplying through by $r^2/R\Theta T$ gives

$$\frac{r^2}{\kappa T}\frac{d^2T}{dt^2} = \frac{r^2}{R}\frac{d^2R}{dr^2} + \frac{2r}{R}\frac{dR}{dr} + \frac{d^2\Theta}{d\theta^2}\frac{1}{\Theta}. \tag{4}$$

The θ term can be separated.

$$\frac{d^2\Theta}{d\theta^2}\frac{1}{\Theta} = -n(n+1), \tag{5}$$

which has a solution

$$\Theta(\theta) = A\cos\left[\sqrt{n(n+1)}\,\theta\right] + B\sin\left[\sqrt{n(n+1)}\,\theta\right]. \tag{6}$$

The remaining portion becomes

$$\frac{r^2}{\kappa T}\frac{d^2T}{dt^2} = \frac{r^2}{R}\frac{d^2R}{dr^2} + \frac{2r}{R}\frac{dR}{dr} - n(n+1). \tag{7}$$

Dividing by r^2 gives

$$\frac{1}{\kappa T}\frac{d^2T}{dt^2} = \frac{1}{R}\frac{d^2R}{dr^2} + \frac{2}{rR}\frac{dR}{dr} - \frac{n(n+1)}{r^2} = -\frac{1}{\lambda^2}, \tag{8}$$

where a NEGATIVE separation constant has been chosen so that the t portion remains finite

$$T(t) = Ce^{-\kappa t/\lambda^2}. \tag{9}$$

The radial portion then becomes

$$\frac{1}{R}\frac{d^2R}{dr^2} + \frac{2}{rR}\frac{dR}{dr} - \frac{n(n+1)}{r^2} + \frac{1}{\lambda^2} = 0 \tag{10}$$

$$r^2\frac{d^2R}{dr^2} + 2r\frac{dR}{dr} + \left[\frac{r^2}{\lambda^2} - n(n+1)\right]R = 0, \tag{11}$$

which is the SPHERICAL BESSEL DIFFERENTIAL EQUATION. If the initial temperature is $T(r,0) = 0$ and the boundary condition is $T(1,t) = 1$, the solution is

$$T(r,t) = 1 - 2\sum_{n=1}^{\infty}\frac{J_0(\alpha_n r)}{\alpha_n J_1(\alpha_n)}e^{\alpha_n{}^2 t}, \tag{12}$$

where α_n is the nth POSITIVE zero of the BESSEL FUNCTION OF THE FIRST KIND J_0.

Heaviside Calculus

A method of solving differential equations using FOURIER TRANSFORMS and LAPLACE TRANSFORMS.

see also FOURIER TRANSFORM, LAPLACE TRANSFORM

Heaviside Step Function

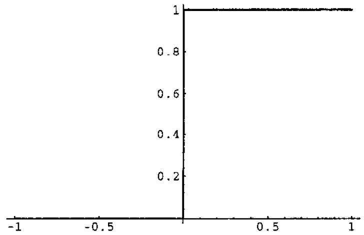

A discontinuous "step" function, also called the UNIT STEP, and defined by

$$H(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & x = 0 \\ 1 & x > 0. \end{cases} \tag{1}$$

It is related to the BOXCAR FUNCTION. The DERIVATIVE is given by

$$\frac{d}{dx}H(x) = \delta(x), \tag{2}$$

where $\delta(x)$ is the DELTA FUNCTION, and the step function is related to the RAMP FUNCTION $R(x)$ by

$$\frac{d}{dx}R(x) = -H(x). \tag{3}$$

Bracewell (1965) gives many identities, some of which include the following. Letting $*$ denote the CONVOLUTION,

$$H(x) * f(x) = \int_{-\infty}^{x} f(x')\,dx' \tag{4}$$

$$H(T) * H(T) = \int_{-\infty}^{\infty} H(u)H(T-u)\,du \tag{5}$$

$$\begin{aligned} &= H(0)\int_{0}^{\infty} H(T-u)\,du \\ &= H(0)H(T)\int_{0}^{T} du = TH(T). \end{aligned} \tag{6}$$

Additional identities are

$$H(x)H(y) = \begin{cases} H(x) & x > y \\ H(y) & x < y \end{cases} \tag{7}$$

$$\begin{aligned} H(ax+b) &= H\left(x+\frac{b}{a}\right)H(a) + H\left(-x-\frac{b}{a}\right)H(-a) \\ &= \begin{cases} H\left(x+\frac{b}{a}\right) & a > 0 \\ H\left(-x-\frac{b}{a}\right) & a < 0. \end{cases} \end{aligned} \tag{8}$$

The step function obeys the integral identities

$$\int_{-a}^{b} H(u-u_0)f(u)\,du = H(u_0)\int_{u_0}^{b} f(u)\,du \tag{9}$$

$$\int_{-a}^{b} H(u_1-u)f(u)\,du = H(u_1)\int_{-a}^{u_1} f(u)\,du \tag{10}$$

$$\int_{-a}^{b} H(u-u_0)H(u_1-u)f(u)\,du = H(u_0)H(u_1)\int_{u_0}^{u_1} f(u)\,du. \qquad (11)$$

The Heaviside step function can be defined by the following limits,

$$\begin{aligned} H(x) &= \lim_{t\to 0}\left[\tfrac{1}{2}+\frac{1}{\pi}\tan^{-1}\left(-\frac{x}{t}\right)\right] && (12)\\ &= \tfrac{1}{2}\lim_{t\to 0}\operatorname{erfc}(-\tfrac{x}{t}) && (13)\\ &= \frac{1}{\sqrt{\pi}}\lim_{t\to 0}\int_{-s}^{\infty} t^{-1}e^{-u^2/t}\,du && (14)\\ &= \tfrac{1}{2}+\tfrac{1}{\pi}\lim_{t\to 0}\operatorname{si}(\tfrac{\pi x}{t}) && (15)\\ &= \lim_{t\to 0}\int_{-\infty}^{x} t^{-1}\operatorname{sinc}\left(\frac{u}{t}\right)du && (16)\\ &= \lim_{t\to 0}\begin{cases} \tfrac{1}{2}(1-e^{-x/t}) & x>0\\ -\tfrac{1}{2}(1-e^{-x/t}) & x<0 \end{cases} && (17)\\ &= \lim_{t\to 0}\int_{-\infty}^{x} t^{-1}\Lambda\left(\frac{x-\frac{1}{2}t}{t}\right)dx, && (18) \end{aligned}$$

where Λ is the one-argument Triangle Function and $\operatorname{si}(x)$ is the Sine Integral.

The Fourier Transform of the Heaviside step function is given by

$$\mathcal{F}[H(x)] = \int_{-\infty}^{\infty} e^{-2\pi ikx}H(x)\,dx = \frac{1}{2}\left[\delta(k)-\frac{i}{\pi k}\right], \qquad (19)$$

where $\delta(k)$ is the Delta Function.

see also Boxcar Function, Delta Function, Fourier Transform—Heaviside Step Function, Ramp Function, Ramp Function, Rectangle Function, Square Wave

References

Bracewell, R. *The Fourier Transform and Its Applications.* New York: McGraw-Hill, 1965.

Spanier, J. and Oldham, K. B. "The Unit-Step $u(x-a)$ and Related Functions." Ch. 8 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 63–69, 1987.

Heawood Conjecture

The bound for the number of colors which are Sufficient for Map Coloring on a surface of Genus g,

$$\chi(g) = \left\lfloor \tfrac{1}{2}(7+\sqrt{48g+1})\right\rfloor$$

is the best possible, where $\lfloor x\rfloor$ is the Floor Function. $\chi(g)$ is called the Chromatic Number, and the first few values for $g = 0, 1, \ldots$ are 4, 7, 8, 9, 10, 11, 12, 12, 13, 13, 14, ... (Sloane's A000934).

The fact that $\chi(g)$ is also Necessary was proved by Ringel and Youngs (1968) with two exceptions: the Sphere (Plane), and the Klein Bottle (for which the Heawood Formula gives seven, but the correct bound is six). When the Four-Color Theorem was proved in 1976, the Klein Bottle was left as the only exception. The four most difficult cases to prove were $g = 59$, 83, 158, and 257.

see also Chromatic Number, Four-Color Theorem, Map Coloring, Six-Color Theorem, Torus Coloring

References

Ringel, G. *Map Color Theorem.* New York: Springer-Verlag, 1974.

Ringel, G. and Youngs, J. W. T. "Solution of the Heawood Map-Coloring Problem." *Proc. Nat. Acad. Sci. USA* **60**, 438–445, 1968.

Sloane, N. J. A. Sequence A000934/M3292 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Wagon, S. "Map Coloring on a Torus." §7.5 in *Mathematica in Action.* New York: W. H. Freeman, pp. 232–237, 1991.

Hebesphenomegacorona

see Johnson Solid

Hecke Algebra

An associative Ring, also called a Hecke Ring, which has a technical definition in terms of commensurable Subgroups.

Hecke L-Function

A generalization of the Euler L-Function associated with a Grossencharacter.

References

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Hecke Operator

A family of operators on each Space of Modular Forms. Hecke operators Commute with each other.

Hecke Ring

see Hecke Algebra

Hectogon

A 100-sided Polygon.

Hedgehog

An envelope parameterized by its Gauss Map. The parametric equations for a hedgehog are

$$\begin{aligned} x &= p(\theta)\cos\theta + p'(\theta)\sin\theta\\ y &= p(\theta)\sin\theta + p'(\theta)\cos\theta. \end{aligned}$$

A plane convex hedgehog has at least four Vertices where the Curvature has a stationary value. A plane convex hedgehog of constant width has at least six VERTICES (Martinez-Maure 1996).

References
Langevin, R.; Levitt, G.; and Rosenberg, H. "Hérissons et Multihérissons (Enveloppes paramétrées par leu application de Gauss." Warsaw: Singularities, 245–253, 1985. Banach Center Pub. 20, PWN Warsaw, 1988.
Martinez-Maure, Y. "A Note on the Tennis Ball Theorem." *Amer. Math. Monthly* **103**, 338–340, 1996.

Heegaard Diagram

A diagram expressing how the gluing operation that connects the HANDLEBODIES involved in a HEEGAARD SPLITTING proceeds, usually by showing how the meridians of the HANDLEBODY are mapped.

see also HANDLEBODY, HEEGAARD SPLITTING

References
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 239, 1976.

Heegaard Splitting

A Heegaard splitting of a connected orientable 3-MANIFOLD M is any way of expressing M as the UNION of two (3,1)-HANDLEBODIES along their boundaries. The boundary of such a (3,1)-HANDLEBODY is an orientable SURFACE of some GENUS, which determines the number of HANDLES in the (3,1)-HANDLEBODIES. Therefore, the HANDLEBODIES involved in a Heegaard splitting are the same, but they may be glued together in a strange way along their boundary. A diagram showing how the gluing is done is known as a HEEGAARD DIAGRAM.

References
Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, p. 255, 1994.

Heegner Number

The values of $-d$ for which QUADRATIC FIELDS $\mathbb{Q}(\sqrt{-d})$ are uniquely factorable into factors of the form $a + b\sqrt{-d}$. Here, a and b are half-integers, except for $d = 1$ and 2, in which case they are INTEGERS. The Heegner numbers therefore correspond to DISCRIMINANTS $-d$ which have CLASS NUMBER $h(-d)$ equal to 1, except for Heegner numbers -1 and -2, which correspond to $d = -4$ and -8, respectively.

The determination of these numbers is called GAUSS'S CLASS NUMBER PROBLEM, and it is now known that there are only nine Heegner numbers: -1, -2, -3, -7, -11, -19, -43, -67, and -163 (Sloane's A003173), corresponding to discriminants -4, -8, -3, -7, -11, -19, -43, -67, and -163, respectively.

Heilbronn and Linfoot (1934) showed that if a larger d existed, it must be $> 10^9$. Heegner (1952) published a proof that only nine such numbers exist, but his proof was not accepted as complete at the time. Subsequent examination of Heegner's proof show it to be "essentially" correct (Conway and Guy 1996).

The Heegner numbers have a number of fascinating connections with amazing results in PRIME NUMBER theory. In particular, the j-FUNCTION provides stunning connections between e, π, and the ALGEBRAIC INTEGERS. They also explain why Euler's PRIME-GENERATING POLYNOMIAL $n^2 - n + 41$ is so surprisingly good at producing PRIMES.

see also CLASS NUMBER, DISCRIMINANT (BINARY QUADRATIC FORM), GAUSS'S CLASS NUMBER PROBLEM, j-Function, PRIME-GENERATING POLYNOMIAL, QUADRATIC FIELD

References
Conway, J. H. and Guy, R. K. "The Nine Magic Discriminants." In *The Book of Numbers.* New York: Springer-Verlag, pp. 224–226, 1996.
Heegner, K. "Diophantische Analysis und Modulfunktionen." *Math. Z.* **56**, 227–253, 1952.
Heilbronn, H. A. and Linfoot, E. H. "On the Imaginary Quadratic Corpora of Class-Number One." *Quart. J. Math. (Oxford)* **5**, 293–301, 1934.
Sloane, N. J. A. Sequence A003173/M0827 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Heesch Number

The Heesch number of a closed plane figure is the maximum number of times that figure can be completely surrounded by copies of itself. The determination of the maximum possible (finite) Heesch number is known as HEESCH'S PROBLEM. The Heesch number of a TRIANGLE, QUADRILATERAL, regular HEXAGON, or any other shape that can TILE or TESSELLATE the plane, is infinity. Conversely, any shape with infinite Heesch number must tile the plane (Eppstein). The largest known (finite) Heesch number is 3, and corresponds to a tile invented by R. Ammann (Senechal 1995).

References
Eppstein, D. "Heesch's Problem." `http://www.ics.uci.edu/~eppstein/junkyard/heesch/`.
Fontaine, A. "An Infinite Number of Plane Figures with Heesch Number Two." *J. Comb. Th. A* **57**, 151–156, 1991.
Senechal, M. *Quasicrystals and Geometry.* New York: Cambridge University Press, 1995.

Heesch's Problem

How many times can a shape be completely surrounded by copies of itself without being able to TILE the entire plane, i.e., what is the maximum (finite) HEESCH NUMBER?

References
Eppstein, D. "Heesch's Problem." `http://www.ics.uci.edu/~eppstein/junkyard/heesch/`.

Height

The vertical length of an object from top to bottom.

see also LENGTH (SIZE), WIDTH (SIZE)

Heilbronn Triangle Problem

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Given any arrangement of n points within a UNIT SQUARE, let H_n be the smallest value for which there is at least one TRIANGLE formed from three of the points with AREA $\leq H_n$. The first few values are

$$\begin{aligned}
H_3 &= \tfrac{1}{2}\\
H_4 &= \tfrac{1}{2}\\
H_5 &= \tfrac{1}{9}\sqrt{3}\\
H_6 &= \tfrac{1}{8}\\
H_7 &\geq \tfrac{1}{12}\\
H_8 &\geq \tfrac{1}{4}(2-\sqrt{3})\\
H_9 &\geq \tfrac{1}{21}\\
H_{10} &\geq \tfrac{1}{32}(3\sqrt{17}-11)\\
H_{11} &\geq \tfrac{1}{27}\\
H_{12} &\geq \tfrac{1}{33}\\
H_{13} &\geq 0.030\\
H_{14} &\geq 0.022\\
H_{15} &\geq 0.020\\
H_{16} &\geq 0.0175.
\end{aligned}$$

Komlós *et al.* (1981, 1982) have shown that there are constants c such that

$$\frac{c\ln n}{n^2} \leq H_n \leq \frac{C}{n^{8/7}-\epsilon},$$

for any $\epsilon > 0$ and all sufficiently large n.

Using an EQUILATERAL TRIANGLE of unit AREA instead gives the constants

$$\begin{aligned}
h_3 &= 1\\
h_4 &= \tfrac{1}{3}\\
h_5 &= 3-2\sqrt{2}\\
h_6 &= \tfrac{1}{8}.
\end{aligned}$$

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/hlb/hlb.html`.

Goldberg, M. "Maximizing the Smallest Triangle Made by N Points in a Square." *Math. Mag.* **45**, 135–144, 1972.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 242–244, 1994.

Komloś, J.; Pintz, J.; and Szemerédi, E. "On Heilbronn's Triangle Problem." *J. London Math. Soc.* **24**, 385–396, 1981.

Komloś, J.; Pintz, J.; and Szemerédi, E. "A Lower Bound for Heilbronn's Triangle Problem." *J. London Math. Soc.* **25**, 13–24, 1982.

Roth, K. F. "Developments in Heilbronn's Triangle Problem." *Adv. Math.* **22**, 364–385, 1976.

Heine-Borel Theorem

If a CLOSED SET of points on a line can be covered by a set of intervals so that every point of the set is an interior point of at least one of the intervals, then there exist a finite number of intervals with the covering property.

Heine Hypergeometric Series

$${}_r\phi_s\begin{bmatrix}\alpha_1,\alpha_2,\ldots,\alpha_r\\ \beta_1,\ldots,\beta_s\end{bmatrix};z\Big] \equiv \sum_{n=0}^{\infty}\frac{(\alpha_1;q)_n(\alpha_2;q)_n\cdots(\alpha_r;q)_n}{(q;q)_n(\beta_1;q)_n\cdots(\beta_s;q)_n}z^n, \quad (1)$$

where

$$(a;q)_n = (1-a)(1-aq)(1-aq^2)\cdots(1-aq^{n-1}), \quad (2)$$

$$(a;q)_0 = 1. \quad (3)$$

In particular,

$${}_2\psi_1(a,b;c;q,z) = \sum_{n=0}^{\infty}\frac{(a;q)_n(b;q)_n z^n}{(q;q)_n(c;q)_n} \quad (4)$$

(Andrews 1986, p. 10). Heine proved the transformation formula

$${}_2\phi_1(a,b;c;q,z) = \frac{(b;q)_\infty(az;q)_\infty}{(c;q)_\infty(z;q)_\infty}{}_2\phi_1(c/b,a;az;q,b), \quad (5)$$

and Rogers (1893) obtained the formulas

$$\begin{aligned}
&{}_2\phi_1(a,b;c;q,z)\\
&\quad = \frac{(c/b;q)_\infty(bz;q)_\infty}{(z;q)_\infty(c;q)_\infty}{}_2\phi_1(b,abz/c;bz;q,c/b) \quad (6)\\
&{}_2\phi_1(a,b,c;q,z)\\
&\quad = (abz/c;q)_\infty(z;q)_\infty{}_2\phi_1(c/a,c/b;c;q,abz/c) \quad (7)
\end{aligned}$$

(Andrews 1986, pp. 10–11).

see also q-SERIES

References

Andrews, G. E. *q-Series: Their Development and Application in Analysis, Number Theory, Combinatorics, Physics, and Computer Algebra.* Providence, RI: Amer. Math. Soc., p. 10, 1986.

Heine, E. "Über die Reihe $1+\frac{(q^\alpha-1)(q^\beta-1)}{(q-1)(q^\gamma-1)}x + \frac{(q^\alpha-1)(q^{\alpha+1}-1)(q^\beta-1)(q^{\beta+1}-1)}{(q-1)(q^2-1)(q^\gamma-1)(q^{\gamma+1}-1)}x^2+\ldots$." *J. reine angew. Math.* **32**, 210–212, 1846.

Heine, E. "Untersuchungen über die Reihe $1+\frac{(1-q^\alpha)(1-q^\beta)}{(1-q)(1-q^\gamma)}\cdot x+\frac{(1-q^\alpha)(1-q^{\alpha+1})(1-q^\beta)(1-q^{\beta+1})}{(1-q)(1-q^2)(1-q^\gamma)(1-q^{\gamma+1})}\cdot x^2+\ldots$." *J. reine angew. Math.* **34**, 285–328, 1847.

Heine, E. *Theorie der Kugelfunctionen und der verwandten Functionen, Vol. 1.* Berlin: Reimer, 1878.

Rogers, L. J. "On a Three-Fold Symmetry in the Elements of Heine's Series." *Proc. London Math. Soc.* **24**, 171–179, 1893.

Heisenberg Group

The Heisenberg group H^n in n COMPLEX variables is the GROUP of all (z,t) with $z \in \mathbb{C}^n$ and $t \in \mathbb{R}$ having multiplication

$$(w,t)(z,t') = (w+z, t+t'+\Im[w^{\mathrm{T}} z])$$

where w^{T} is the conjugate transpose. The Heisenberg group is ISOMORPHIC to the group of MATRICES

$$\begin{bmatrix} 1 & z^{\mathrm{T}} & \frac{1}{2}|z|^2 + it \\ 0 & 1 & z \\ 0 & 0 & 1 \end{bmatrix},$$

and satisfies

$$(z,t)^{-1} = (-z,-t).$$

Every finite-dimensional unitary representation is trivial on Z and therefore factors to a REPRESENTATION of the quotient $\mathbb{C}^n$.

see also NIL GEOMETRY

References
Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Heisenberg Space

The boundary of COMPLEX HYPERBOLIC 2-SPACE.

see also HYPERBOLIC SPACE

Held Group

The SPORADIC GROUP *He.*

References
Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas/He.html`.

Helen of Geometers

see CYCLOID

Helicoid

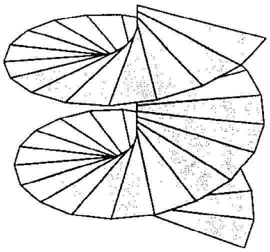

The MINIMAL SURFACE having a HELIX as its boundary. It is the only RULED MINIMAL SURFACE other than the PLANE (Catalan 1842, do Carmo 1986). For many years, the helicoid remained the only known example of a complete embedded MINIMAL SURFACE of finite topology with infinite CURVATURE. However, in 1992 a second example, known as HOFFMAN'S MINIMAL SURFACE and consisting of a helicoid with a HOLE, was discovered (*Sci. News* 1992).

The equation of a helicoid in CYLINDRICAL COORDINATES is

$$z = c\theta. \tag{1}$$

In CARTESIAN COORDINATES, it is

$$\frac{y}{x} = \tan\left(\frac{z}{c}\right). \tag{2}$$

It can be given in parametric form by

$$x = u\cos v \tag{3}$$

$$y = u\sin v \tag{4}$$

$$z = cu, \tag{5}$$

which has an obvious generalization to the ELLIPTIC HELICOID. The differentials are

$$dx = \cos v\,du - u\sin v\,dv \tag{6}$$

$$dy = \sin v\,du + u\cos v\,dv \tag{7}$$

$$dz = 2cu\,dy, \tag{8}$$

so the LINE ELEMENT on the surface is

$$\begin{aligned} ds^2 &= dx^2 + dy^2 + dz^2 \\ &= \cos^2 v\,du^2 - 2u\sin v\cos v\,du\,dv + u^2\sin^2 v\,dv^2 \\ &\quad + \sin^2 v\,du^2 + 2u\sin v\cos v\,du\,dv + u^2\cos^2 v\,dv^2 \\ &\quad + 4c^2u^2\,du^2 \\ &= (1+4c^2u^2)\,du^2 + u^2\,dv^2, \end{aligned} \tag{9}$$

and the METRIC components are

$$g_{uu} = 1 + 4c^2u^2 \tag{10}$$

$$g_{uv} = 0 \tag{11}$$

$$g_{vv} = u^2. \tag{12}$$

From GAUSS'S THEOREMA EGREGIUM, the GAUSSIAN CURVATURE is then

$$K = \frac{4c^2}{(1+4c^2u^2)^2}. \tag{13}$$

The MEAN CURVATURE is

$$H = 0, \tag{14}$$

and the equation for the LINES OF CURVATURE is

$$u = \pm c\sinh(v-k). \tag{15}$$

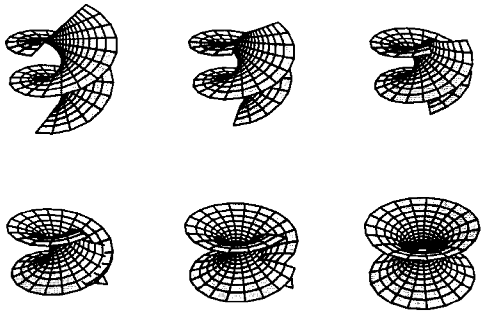

The helicoid can be continuously deformed into a CATENOID by the transformation

$$x(u,v)=\cos\alpha\sinh v\sin u+\sin\alpha\cosh v\cos u \quad (16)$$
$$y(u,v)=-\cos\alpha\sinh v\cos u+\sin\alpha\cosh v\sin u \quad (17)$$
$$z(u,v)=u\cos\alpha+v\sin\alpha, \quad (18)$$

where $\alpha=0$ corresponds to a helicoid and $\alpha=\pi/2$ to a CATENOID.

If a twisted curve C (i.e., one with TORSION $\tau\neq 0$) rotates about a fixed axis A and, at the same time, is displaced parallel to A such that the speed of displacement is always proportional to the angular velocity of rotation, then C generates a GENERALIZED HELICOID.

see also CALCULUS OF VARIATIONS, CATENOID, ELLIPTIC HELICOID, GENERALIZED HELICOID, HELIX, HOFFMAN'S MINIMAL SURFACE, MINIMAL SURFACE

References

Catalan E. "Sur les surfaces régléess dont l'aire est un minimum." *J. Math. Pure Appl.* **7**, 203–211, 1842.

do Carmo, M. P. "The Helicoid." §3.5B in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 44–45, 1986.

Fischer, G. (Ed.). Plate 91 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 87, 1986.

Geometry Center. "The Helicoid." `http://www.geom.umn.edu/zoo/diffgeom/surfspace/helicoid/`.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 264, 1993.

Kreyszig, E. *Differential Geometry.* New York: Dover, p. 88, 1991.

Meusnier, J. B. "Mémoire sur la courbure des surfaces." *Mém. des savans étrangers* **10** (lu 1776), 477–510, 1785.

Peterson, I. "Three Bites in a Doughnut." *Sci. News* **127**, 168, Mar. 16, 1985.

"Putting a Handle on a Minimal Helicoid." *Sci. News* **142**, 276, Oct. 24, 1992.

Wolfram, S. *The Mathematica Book, 3rd ed.* Champaign, IL: Wolfram Media, p. 164, 1996.

Helix

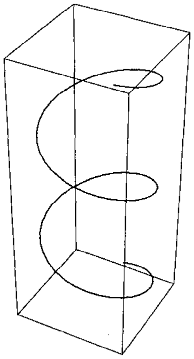

A helix is also called a CURVE OF CONSTANT SLOPE. It can be defined as a curve for which the TANGENT makes a constant ANGLE with a fixed line. The helix is a SPACE CURVE with parametric equations

$$x=r\cos t \quad (1)$$
$$y=r\sin t \quad (2)$$
$$z=ct, \quad (3)$$

where c is a constant. The CURVATURE of the helix is given by

$$\kappa=\frac{r}{r^2+c^2}, \quad (4)$$

and the LOCUS of the centers of CURVATURE of a helix is another helix. The ARC LENGTH is given by

$$s=\int\sqrt{x'^2+y'^2+z'^2}\,dt=\sqrt{r^2+c^2}\,t. \quad (5)$$

The TORSION of a helix is given by

$$\begin{aligned}\tau&=\frac{1}{r^2(r^2+c^2)}\begin{vmatrix}-r\sin t & -r\cos t & r\sin t\\ r\cos t & -r\sin t & -r\cos t\\ c & 0 & 0\end{vmatrix}\\ &=\frac{c}{r^2+c^2}, \end{aligned} \quad (6)$$

so

$$\frac{\kappa}{\tau}=\frac{\frac{r}{r^2+c^2}}{\frac{c}{r^2+c^2}}=\frac{r}{c}, \quad (7)$$

which is a constant. In fact, LANCRET'S THEOREM states that a NECESSARY and SUFFICIENT condition for a curve to be a helix is that the ratio of CURVATURE to TORSION be constant. The OSCULATING PLANE of the helix is given by

$$\begin{vmatrix}z_1-r\cos t & z_2-r\sin t & z_3-ct\\ -r\sin t & r\cos t & c\\ -r\cos t & -r\sin t & 0\end{vmatrix}=0 \quad (8)$$

$$z_1c\sin t-z_2c\cos t+(z_3-ct)r=0. \quad (9)$$

The MINIMAL SURFACE of a helix is a HELICOID.

see also GENERALIZED HELIX, HELICOID, SPHERICAL HELIX

References

Geometry Center. "The Helix." `http://www.geom.umn.edu/zoo/diffgeom/surfspace/helicoid/helix.html`.

Gray, A. "The Helix and Its Generalizations." §7.5 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 138–140, 1993.

Isenberg, C. Plate 4.11 in *The Science of Soap Films and Soap Bubbles.* New York: Dover, 1992.

Pappas, T. "The Helix—Mathematics & Genetics." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 166–168, 1989.

Wolfram, S. *The Mathematica Book, 3rd ed.* Champaign, IL: Wolfram Media, p. 163, 1996.

Helly Number

Given a Euclidean n-space,

$$H_n \equiv n+1.$$

see also EUCLIDEAN SPACE, HELLY'S THEOREM

Helly's Theorem

If F is a family of more than n bounded closed convex sets in Euclidean n-space $\mathbb{R}^n$, and if every H_n (where H_n is the HELLY NUMBER) members of F have at least one point in common, then all the members of F have at least one point in common.

see also CARATHÉODORY'S FUNDAMENTAL THEOREM, HELLY NUMBER

Helmholtz Differential Equation

A PARTIAL DIFFERENTIAL EQUATION which can be written in a SCALAR version

$$\nabla^2\psi + k^2\psi = 0, \tag{1}$$

or VECTOR form,

$$\nabla^2\mathbf{A} + k^2\mathbf{A} = 0, \tag{2}$$

where ∇^2 is the LAPLACIAN. When $k = 0$, the Helmholtz differential equation reduces to LAPLACE'S EQUATION. When $k^2 < 0$, the equation becomes the space part of the diffusion equation.

The Helmholtz differential equation can be solved by SEPARATION OF VARIABLES in only 11 coordinate systems, 10 of which (with the exception of CONFOCAL PARABOLOIDAL COORDINATES) are particular cases of the CONFOCAL ELLIPSOIDAL system: CARTESIAN, CONFOCAL ELLIPSOIDAL, CONFOCAL PARABOLOIDAL, CONICAL, CYLINDRICAL, ELLIPTIC CYLINDRICAL, OBLATE SPHEROIDAL, PARABOLOIDAL, PARABOLIC CYLINDRICAL, PROLATE SPHEROIDAL, and SPHERICAL COORDINATES (Eisenhart 1934). LAPLACE'S EQUATION (the Helmholtz differential equation with $k = 0$) is separable in the two additional BISPHERICAL COORDINATES and TOROIDAL COORDINATES.

If Helmholtz's equation is separable in a 3-D coordinate system, then Morse and Feshbach (1953, pp. 509–510) show that

$$\frac{h_1h_2h_3}{{h_n}^2} = f_n(u_n)g_n(u_i,u_j), \tag{3}$$

where $i \neq j \neq n$. The LAPLACIAN is therefore of the form

$$\begin{aligned}\nabla^2 = \frac{1}{h_1h_2h_3}\bigg\{ & g_1(u_2,u_3)\frac{\partial}{\partial u_1}\left[f_1(u_1)\frac{\partial}{\partial u_1}\right] \\ & +g_2(u_1,u_3)\frac{\partial}{\partial u_2}\left[f_2(u_2)\frac{\partial}{\partial u_2}\right] \\ & +g_3(u_1,u_2)\frac{\partial}{\partial u_3}\left[f_3(u_3)\frac{\partial}{\partial u_3}\right]\bigg\},\end{aligned} \tag{4}$$

which simplifies to

$$\begin{aligned}\nabla^2 = & \frac{1}{{h_1}^2 f_1}\frac{\partial}{\partial u_1}\left[f_1(u_1)\frac{\partial}{\partial u_1}\right] \\ & +\frac{1}{{h_2}^2 f_2}\frac{\partial}{\partial u_2}\left[f_2(u_2)\frac{\partial}{\partial u_2}\right] \\ & +\frac{1}{{h_3}^2 f_3}\frac{\partial}{\partial u_3}\left[f_3(u_3)\frac{\partial}{\partial u_3}\right].\end{aligned} \tag{5}$$

Such a coordinate system obeys the ROBERTSON CONDITION, which means that the STÄCKEL DETERMINANT is of the form

$$S = \frac{h_1h_2h_3}{f_1(u_1)f_2(u_2)f_3(u_3)}. \tag{6}$$

Coordinate System	Variables	Solution Functions
Cartesian	$X(x)Y(y)Z(z)$	exponential, circular, hyperbolic
circular cylindrical	$R(r)\Theta(\theta)Z(z)$	Bessel, exponential, circular
conical		ellipsoidal harmonics, power
ellipsoidal	$\Lambda(\lambda)M(\mu)N(\nu)$	ellipsoidal harmonics
elliptic cylindrical	$U(u)V(v)Z(z)$	Mathieu, circular
oblate spheroidal	$\Lambda(\lambda)M(\mu)N(\nu)$	Legendre, circular
parabolic		Bessel, circular
parabolic cylindrical		Parabolic cylinder, Bessel, circular
paraboloidal	$U(u)V(v)\Theta(\theta)$	Baer functions, circular
prolate spheroidal	$\Lambda(\lambda)M(\mu)N(\nu)$	Legendre, circular
spherical	$R(r)\Theta(\theta)\Phi(\phi)$	Legendre, power, circular

see also LAPLACE'S EQUATION, POISSON'S EQUATION, SEPARATION OF VARIABLES, SPHERICAL BESSEL DIFFERENTIAL EQUATION

References

Eisenhart, L. P. "Separable Systems in Euclidean 3-Space." *Physical Review* **45**, 427–428, 1934.

Eisenhart, L. P. "Separable Systems of Stäckel." *Ann. Math.* **35**, 284–305, 1934.

Eisenhart, L. P. "Potentials for Which Schroedinger Equations Are Separable." *Phys. Rev.* **74**, 87–89, 1948.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 125–126 and 509–510, 1953.

Helmholtz Differential Equation—Bipolar Coordinates

In BIPOLAR COORDINATES, the HELMHOLTZ DIFFERENTIAL EQUATION is not separable, but LAPLACE'S EQUATION is.

see also LAPLACE'S EQUATION—BIPOLAR COORDINATES

Helmholtz Differential Equation—Cartesian Coordinates

In 2-D CARTESIAN COORDINATES, attempt SEPARATION OF VARIABLES by writing

$$F(x,y) = X(x)Y(y), \tag{1}$$

then the HELMHOLTZ DIFFERENTIAL EQUATION becomes

$$\frac{d^2X}{dx^2}Y+\frac{d^2Y}{dy^2}X+k^2XY=0. \tag{2}$$

Dividing both sides by XY gives

$$\frac{1}{X}\frac{d^2X}{dx^2}+\frac{1}{Y}\frac{d^2Y}{dy^2}+k^2=0. \tag{3}$$

This leads to the two coupled ordinary differential equations with a separation constant m^2,

$$\frac{1}{X}\frac{d^2X}{dx^2}=m^2 \tag{4}$$

$$\frac{1}{Y}\frac{d^2Y}{dy^2}=-(m^2+k^2), \tag{5}$$

where X and Y could be interchanged depending on the boundary conditions. These have solutions

$$X=A_me^{mx}+B_me^{-mx} \tag{6}$$

$$\begin{aligned} Y &= C_me^{i\sqrt{m^2+k^2}\,y}+D_me^{-i\sqrt{m^2+k^2}\,y} \\ &= E_m\sin(\sqrt{m^2+k^2}\,y)+F_m\cos(\sqrt{m^2+k^2}\,y). \end{aligned} \tag{7}$$

The general solution is then

$$\begin{aligned} F(x,y) &= \sum_{m=1}^{\infty}(A_me^{mx}+B_me^{-mx}) \\ &\quad \times[E_m\sin(\sqrt{m^2+k^2}\,y)+F_m\cos(\sqrt{m^2+k^2}\,y)]. \end{aligned} \tag{8}$$

In 3-D CARTESIAN COORDINATES, attempt SEPARATION OF VARIABLES by writing

$$F(x,y,z)=X(x)Y(y)Z(z), \tag{9}$$

then the HELMHOLTZ DIFFERENTIAL EQUATION becomes

$$\frac{d^2X}{dx^2}YZ+\frac{d^2Y}{dy^2}XZ+\frac{d^2Z}{dz^2}XY+k^2XY=0. \tag{10}$$

Dividing both sides by XYZ gives

$$\frac{1}{X}\frac{d^2X}{dx^2}+\frac{1}{Y}\frac{d^2Y}{dy^2}+\frac{1}{Z}\frac{d^2Z}{dz^2}+k^2=0. \tag{11}$$

This leads to the three coupled differential equations

$$\frac{1}{X}\frac{d^2X}{dx^2}=l^2 \tag{12}$$

$$\frac{1}{Y}\frac{d^2Y}{dy^2}=m^2 \tag{13}$$

$$\frac{1}{Z}\frac{d^2Z}{dz^2}=-(k^2+l^2+m^2), \tag{14}$$

where X, Y, and Z could be permuted depending on boundary conditions. The general solution is therefore

$$\begin{aligned} &F(x,y,z) \\ &\quad =\sum_{l=1}^{\infty}\sum_{m=1}^{\infty}(A_le^{lx}+B_le^{-lx})(C_me^{my}+D_me^{-my}) \\ &\quad \times(E_{lm}e^{-i\sqrt{k^2+l^2+m^2}\,z}+F_{lm}e^{i\sqrt{k^2+l^2+m^2}\,z}). \end{aligned} \tag{15}$$

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 501–502, 513–514 and 656, 1953.

Helmholtz Differential Equation—Circular Cylindrical Coordinates

In CYLINDRICAL COORDINATES, the SCALE FACTORS are $h_r=1$, $h_\theta=r$, $h_z=1$ and the separation functions are $f_1(r)=r$, $f_2(\theta)=1$, $f_3(z)=1$, so the STÄCKEL DETERMINANT is 1. Attempt SEPARATION OF VARIABLES by writing

$$F(r,\theta,z)=R(r)\Theta(\theta)Z(z), \tag{1}$$

then the HELMHOLTZ DIFFERENTIAL EQUATION becomes

$$\frac{d^2R}{dr^2}\Theta Z+\frac{1}{r}\frac{dR}{dr}\Theta Z+\frac{1}{r^2}\frac{d^2\Theta}{d\theta^2}RZ+\frac{d^2Z}{dz^2}R\Theta+k^2R\Theta Z=0. \tag{2}$$

Now divide by $R\Theta Z$,

$$\left(\frac{r^2}{R}\frac{d^2R}{dr^2}+\frac{r}{R}\frac{dR}{dr}\right)+\frac{d^2\Theta}{d\theta^2}\frac{1}{\Theta}+\frac{d^2Z}{dz^2}\frac{r^2}{Z}+k^2=0, \tag{3}$$

so the equation has been separated. Since the solution must be periodic in Θ from the definition of the circular cylindrical coordinate system, the solution to the second part of (3) must have a NEGATIVE separation constant

$$\frac{d^2\Theta}{d\theta^2}\frac{1}{\Theta}=-(k^2+m^2), \tag{4}$$

which has a solution

$$\Theta(\theta)=C_me^{-i\sqrt{k^2+m^2}\,\theta}+D_me^{i\sqrt{k^2+m^2}\,\theta}. \tag{5}$$

Plugging (5) back into (3) gives

$$\frac{r^2}{R}\frac{d^2R}{dr^2}+\frac{r}{R}\frac{dR}{dr}-m^2+\frac{d^2Z}{dz^2}\frac{r^2}{Z}=0 \tag{6}$$

$$\frac{1}{R}\frac{d^2R}{dr^2}+\frac{1}{rR}\frac{dR}{dr}-\frac{m^2}{r^2}+\frac{d^2Z}{dz^2}\frac{1}{Z}=0. \tag{7}$$

The solution to the second part of (7) must not be sinusoidal at ±∞ for a physical solution, so the differential equation has a POSITIVE separation constant

$$\frac{d^2Z}{dz^2}\frac{1}{Z}=n^2, \tag{8}$$

and the solution is

$$Z(z) = E_n e^{-nz} + F_n e^{nz}. \tag{9}$$

Plugging (9) back into (7) and multiplying through by R yields

$$\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr} + \left(n^2 - \frac{m^2}{r^2}\right) R = 0 \tag{10}$$

$$\frac{1}{n^2}\frac{d^2R}{dr^2} + \frac{1}{(nr)}\frac{1}{n}\frac{dR}{dr} + \left[1 - \frac{m^2}{(nr)^2}\right] R = 0 \tag{11}$$

$$\frac{d^2R}{d(nr)^2} + \frac{1}{(nr)}\frac{dR}{d(nr)} + \left[1 - \frac{m^2}{(nr)^2}\right] R = 0. \tag{12}$$

This is the BESSEL DIFFERENTIAL EQUATION, which has a solution

$$R(r) = A_{mn} J_m(nr) + B_{mn} Y_m(nr), \tag{13}$$

where $J_n(x)$ and $Y_n(x)$ are BESSEL FUNCTIONS OF THE FIRST and SECOND KINDS, respectively. The general solution is therefore

$$\begin{aligned} F(r,\theta,z) &\\ = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}&[A_{mn}J_m(nr) + B_{mn}Y_m(nr)] \\ \times(C_m e^{-i\sqrt{k^2+m^2}\,\theta} &+ D_m e^{i\sqrt{k^2+m^2}\,\theta})(E_n e^{-nz} + F_n e^{nz}). \end{aligned} \tag{14}$$

Actually, the HELMHOLTZ DIFFERENTIAL EQUATION is separable for general k of the form

$$k^2(r,\theta,z) = f(r) + \frac{g(\theta)}{r^2} + h(z) + k'^2. \tag{15}$$

see also CYLINDRICAL COORDINATES, HELMHOLTZ DIFFERENTIAL EQUATION

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 514 and 656–657, 1953.

Helmholtz Differential Equation—Confocal Ellipsoidal Coordinates

Using the NOTATION of Byerly (1959, pp. 252–253), LAPLACE'S EQUATION can be reduced to

$$\nabla^2 F = (\mu^2-\nu^2)\frac{\partial^2 F}{\partial\alpha^2} + (\lambda^2-\nu^2)\frac{\partial^2 F}{\partial\beta^2} + (\lambda^2-\mu^2)\frac{\partial^2 F}{\partial\gamma^2} = 0, \tag{1}$$

where

$$\begin{aligned} \alpha &= c\int_c^\lambda \frac{d\lambda}{\sqrt{(\lambda^2-b^2)(\lambda^2-c^2)}} \\ &= F\Bigl[\frac{b}{c},\frac{\pi}{2}\Bigr) - F\Bigl(\frac{b}{c},\sin^{-1}\Bigl(\frac{c}{\lambda}\Bigr)\Bigr)\Bigr] \end{aligned} \tag{2}$$

$$\begin{aligned} \beta &= c\int_b^\mu \frac{d\mu}{\sqrt{(c^2-\mu^2)(\mu^2-b^2)}} \\ &= F\left[\sqrt{1-b^2-c^2}, \sin^{-1}\left(\sqrt{\frac{1-\frac{b^2}{\mu^2}}{1-\frac{b^2}{c^2}}}\right)\right] \end{aligned} \tag{3}$$

$$\begin{aligned} \gamma &= c\int_0^\nu \frac{d\nu}{\sqrt{(b^2-\nu^2)(c^2-\nu^2)}} \\ &= F\left(\frac{b}{c},\sin^{-1}\left(\frac{\nu}{b}\right)\right). \end{aligned} \tag{4}$$

In terms of α, β, and γ,

$$\lambda = c\,\mathrm{dc}\left(\alpha,\frac{b}{c}\right) \tag{5}$$

$$\mu = b\,\mathrm{nd}\left(\beta,\sqrt{1-\frac{b^2}{c^2}}\right) \tag{6}$$

$$\nu = b\,\mathrm{sn}\left(\gamma,\frac{b}{c}\right). \tag{7}$$

Equation (1) is not separable using a function of the form

$$F = L(\alpha)M(\beta)N(\gamma), \tag{8}$$

but it is if we let

$$\frac{1}{L}\frac{d^2L}{d\alpha^2} = \sum a_k\lambda^k \tag{9}$$

$$\frac{1}{M}\frac{d^2M}{d\beta^2} = \sum b_k\mu^k \tag{10}$$

$$\frac{1}{N}\frac{d^2N}{d\gamma^2} = \sum c_k\nu^k. \tag{11}$$

These give

$$a_0 = -b_0 = c_0 \tag{12}$$

$$a_2 = -b_2 = c_2, \tag{13}$$

and all others terms vanish. Therefore (1) can be broken up into the equations

$$\frac{d^2L}{d\alpha^2} = (a_0 + a_2\lambda^2)L \tag{14}$$

$$\frac{d^2M}{d\beta^2} = -(a_0 + a_2\mu^2)M \tag{15}$$

$$\frac{d^2N}{d\gamma^2} = (a_0 + a_2\nu^2)N. \tag{16}$$

For future convenience, now write

$$a_0 = -(b^2 + c^2)p \tag{17}$$

$$a_2 = m(m+1), \tag{18}$$

then

$$\frac{d^2L}{d\alpha^2} - [m(m+1)\lambda^2 - (b^2+c^2)p]L = 0 \tag{19}$$

$$\frac{d^2M}{d\beta^2} + [m(m+1)\mu^2 - (b^2+c^2)p]M = 0 \tag{20}$$

$$\frac{d^2N}{d\gamma^2} - [m(m+1)\nu^2 - (b^2+c^2)p]N = 0. \tag{21}$$

Now replace α, β, and γ to obtain

$$(\lambda^2-b^2)(\lambda^2-c^2)\frac{d^2L}{d\lambda^2} + \lambda(\lambda^2-b^2+\lambda^2-c^2)\frac{dL}{d\lambda} - [m(m+1)\lambda^2 - (b^2+c^2)p]L = 0 \tag{22}$$

$$(\mu^2-b^2)(\mu^2-c^2)\frac{d^2M}{d\mu^2} + \mu(\mu^2-b^2+\mu^2-c^2)\frac{dM}{d\mu} - [m(m+1)\mu^2 - (b^2+c^2)p]M = 0 \tag{23}$$

$$(\nu^2-b^2)(\nu^2-c^2)\frac{d^2N}{d\nu^2} + \nu(\nu^2-b^2+\nu^2-c^2)\frac{dN}{d\nu} - [m(m+1)\nu^2 - (b^2+c^2)p]N = 0. \tag{24}$$

Each of these is a LAMÉ'S DIFFERENTIAL EQUATION, whose solution is called an ELLIPSOIDAL HARMONIC. Writing

$$L(\lambda) = E_m^p(\lambda) \tag{25}$$

$$M(\lambda) = E_m^p(\mu) \tag{26}$$

$$N(\lambda) = E_m^p(\nu) \tag{27}$$

gives the solution to (1) as a product of ELLIPSOIDAL HARMONICS $E_m^p(x)$.

$$F = E_m^p(\lambda)E_m^p(\mu)E_m^p(\nu). \tag{28}$$

References

Arfken, G. "Confocal Ellipsoidal Coordinates (ξ_1, ξ_2, ξ_3)." §2.15 in *Mathematical Methods for Physicists, 2nd ed.* Orlando, FL: Academic Press, pp. 117–118, 1970.

Byerly, W. E. *An Elementary Treatise on Fourier's Series, and Spherical, Cylindrical, and Ellipsoidal Harmonics, with Applications to Problems in Mathematical Physics.* New York: Dover, pp. 251–258, 1959.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 663, 1953.

Helmholtz Differential Equation—Confocal Paraboloidal Coordinates

As shown by Morse and Feshbach (1953), the HELMHOLTZ DIFFERENTIAL EQUATION is separable in CONFOCAL PARABOLOIDAL COORDINATES.

see also CONFOCAL PARABOLOIDAL COORDINATES

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 664, 1953.

Helmholtz Differential Equation—Conical Coordinates

In CONICAL COORDINATES, LAPLACE'S EQUATION can be written

$$\frac{\partial^2 V}{\partial\alpha^2} + \frac{\partial^2 V}{\partial\beta^2} + (\mu^2-\nu^2)\frac{\partial}{\partial\lambda}\left(\lambda^2\frac{\partial V}{\partial\lambda}\right) = 0, \tag{1}$$

where

$$\alpha = \int_a^\mu \frac{d\mu}{\sqrt{(\mu^2-a^2)(b^2-\mu^2)}} \tag{2}$$

$$\beta = \int_0^\nu \frac{d\nu}{\sqrt{(a^2-\nu^2)(b^2-\nu^2)}} \tag{3}$$

(Byerly 1959). Letting

$$V = U(u)R(r) \tag{4}$$

breaks (1) into the two equations,

$$\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) = m(m+1)R \tag{5}$$

$$\frac{\partial^2 U}{\partial\alpha^2} + \frac{\partial^2 U}{\partial\beta^2} + m(m+1)(\mu^2-\nu^2)U = 0. \tag{6}$$

Solving these gives

$$R(r) = Ar^m + Br^{-m-1} \tag{7}$$

$$U(u) = E_m^p(\mu)E_m^p(\nu), \tag{8}$$

where E_m^p are ELLIPSOIDAL HARMONICS. The regular solution is therefore

$$V = Ar^m E_m^p(\mu)E_m^p(\nu). \tag{9}$$

However, because of the cylindrical symmetry, the solution $E_m^p(\mu)E_m^p(\nu)$ is an mth degree SPHERICAL HARMONIC.

References

Arfken, G. "Conical Coordinates (ξ_1, ξ_2, ξ_3)." §2.16 in *Mathematical Methods for Physicists, 2nd ed.* Orlando, FL: Academic Press, pp. 118–119, 1970.

Byerly, W. E. *An Elementary Treatise on Fourier's Series, and Spherical, Cylindrical, and Ellipsoidal Harmonics, with Applications to Problems in Mathematical Physics.* New York: Dover, p. 263, 1959.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 514 and 659, 1953.

Helmholtz Differential Equation—Elliptic Cylindrical Coordinates

In ELLIPTIC CYLINDRICAL COORDINATES, the SCALE FACTORS are $h_u = h_v = a\sqrt{\sinh^2 u + \sin^2 v}$, $h_z = 1$, and the separation functions are $f_1(u) = f_2(v) = f_3(z) = 1$, giving a STÄCKEL DETERMINANT of $S = a^2(\sin^2 v + \sinh^2 u)$. The Helmholtz differential equation is

$$\frac{1}{a^2(\sinh^2 u + \sin^2 v)}\left(\frac{\partial^2 F}{\partial u^2} + \frac{\partial^2 F}{\partial v^2}\right) + \frac{\partial^2 F}{\partial z^2} + k^2 = 0. \tag{1}$$

Attempt SEPARATION OF VARIABLES by writing

$$F(u, v, z) = U(u)V(v)Z(z), \tag{2}$$

then the HELMHOLTZ DIFFERENTIAL EQUATION becomes

$$\frac{Z}{\sinh^2 u + \sin^2 v}\left(V\frac{d^2U}{du^2} + U\frac{d^2V}{dv^2}\right) + UV\frac{d^2Z}{dz^2} + k^2UVZ = 0. \tag{3}$$

Now divide by UVZ to give

$$\frac{1}{\sinh^2 u + \sin^2 v}\left(\frac{1}{U}\frac{d^2U}{du^2} + \frac{1}{V}\frac{d^2V}{dv^2}\right) + \frac{1}{Z}\frac{d^2Z}{dz^2} + k^2 = 0. \tag{4}$$

Separating the Z part,

$$\frac{1}{Z}\frac{d^2Z}{dz^2} = -(k^2 + m^2) \tag{5}$$

$$\frac{1}{\sinh^2 u + \sin^2 v}\left(\frac{1}{U}\frac{d^2U}{du^2} + \frac{1}{V}\frac{d^2V}{dv^2}\right) = m^2, \tag{6}$$

so

$$\frac{d^2Z}{dz^2} = -(k^2 + m^2)Z, \tag{7}$$

which has the solution

$$Z(z) = A\cos(\sqrt{k^2 + m^2}\, z) + B\sin(\sqrt{k^2 + m^2}\, z). \tag{8}$$

Rewriting (6) gives

$$\left(\frac{1}{U}\frac{d^2U}{du^2} - m^2\sinh^2 u\right) + \left(\frac{1}{V}\frac{d^2V}{dv^2} - m^2\sin^2 v\right) = 0, \tag{9}$$

which can be separated into

$$\frac{1}{U}\frac{d^2U}{du^2} - m^2\sinh^2 u = c \tag{10}$$

$$c + \frac{1}{V}\frac{d^2V}{dv^2} - m^2\sin^2 v = 0, \tag{11}$$

so

$$\frac{d^2U}{du^2} - (c + m^2\sinh^2 u)U = 0 \tag{12}$$

$$\frac{d^2V}{dv^2} + (c - m^2\sin^2 v)V = 0. \tag{13}$$

Now use

$$\sinh^2 u = \tfrac{1}{2}[1 - \cosh(2u)] \tag{14}$$

$$\sin^2 v = \tfrac{1}{2}[1 - \cos(2v)] \tag{15}$$

to obtain

$$\frac{d^2U}{du^2} - \{c + \tfrac{1}{2}m^2[1 - \cosh(2u)]\}U = 0 \tag{16}$$

$$\frac{d^2V}{dv^2} + \{c + \tfrac{1}{2}m^2[1 - \cos(2v)]\}V = 0. \tag{17}$$

Regrouping gives

$$\frac{d^2U}{du^2} - [(c + \tfrac{1}{2}m^2) - \tfrac{1}{4}m^2 2\cosh(2u)]U = 0 \tag{18}$$

$$\frac{d^2V}{dv^2} + [(c + \tfrac{1}{2}m^2) - \tfrac{1}{4}m^2 2\cos(2v)]V = 0. \tag{19}$$

Let $b \equiv \frac{1}{2}m^2 + c$ and $q \equiv \frac{1}{4}m^2$, then these become

$$\frac{d^2U}{du^2} - [b - 2q\cosh(2u)]U = 0 \tag{20}$$

$$\frac{d^2V}{dv^2} + [b - 2q\cos(2v)]V = 0. \tag{21}$$

Here, (21) is the MATHIEU DIFFERENTIAL EQUATION and (20) is the modified MATHIEU DIFFERENTIAL EQUATION. These solutions are known as MATHIEU FUNCTIONS.

see also ELLIPTIC CYLINDRICAL COORDINATES, MATHIEU DIFFERENTIAL EQUATION, MATHIEU FUNCTION

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 514 and 657, 1953.

Helmholtz Differential Equation—Oblate Spheroidal Coordinates

As shown by Morse and Feshbach (1953) and Arfken (1970), the HELMHOLTZ DIFFERENTIAL EQUATION is separable in OBLATE SPHEROIDAL COORDINATES.

References

Arfken, G. "Oblate Spheroidal Coordinates (u, v, φ)." §2.11 in *Mathematical Methods for Physicists, 2nd ed.* Orlando, FL: Academic Press, pp. 107–109, 1970.
Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 662, 1953.

Helmholtz Differential Equation—Parabolic Coordinates

The SCALE FACTORS are $h_u = h_v = \sqrt{u^2 + v^2}$, $h_\theta = uv$ and the separation functions are $f_1(u) = u$, $f_2(v) = v$, $f_3(\theta) = 1$, given a STÄCKEL DETERMINANT of $S = u^2 + v^2$. The LAPLACIAN is

$$\frac{1}{u^2 + v^2}\left(\frac{1}{u}\frac{\partial F}{\partial u} + \frac{\partial^2 F}{\partial u^2} + \frac{1}{v}\frac{\partial F}{\partial v} + \frac{\partial^2 F}{\partial v^2}\right) + \frac{1}{u^2 v^2}\frac{\partial^2 F}{\partial \theta^2} + k^2 = 0. \tag{1}$$

Attempt SEPARATION OF VARIABLES by writing

$$F(u,v,z) \equiv U(u)V(v)\Theta(\theta), \tag{2}$$

then the HELMHOLTZ DIFFERENTIAL EQUATION becomes

$$\frac{1}{u^2+v^2}\left[V\Theta\left(\frac{1}{u}\frac{dU}{du}+\frac{d^2U}{du^2}\right)+U\Theta\left(\frac{1}{v}\frac{dV}{dv}+\frac{d^2V}{dv^2}\right)\right]+k^2UV\Theta=0. \tag{3}$$

Now divide by $UV\Theta$,

$$\frac{u^2v^2}{u^2+v^2}\left[\frac{1}{U}\left(\frac{1}{u}\frac{dU}{du}+\frac{d^2U}{du^2}\right)+\frac{1}{V}\left(\frac{1}{v}\frac{dV}{dv}+\frac{d^2V}{dv^2}\right)\right]+\frac{1}{\Theta}\frac{d^2\Theta}{d\theta^2}+k^2=0. \tag{4}$$

Separating the Θ part,

$$\frac{1}{\Theta}\frac{d^2\Theta}{f\theta^2}=-(k^2+m^2) \tag{5}$$

$$\frac{u^2v^2}{u^2+v^2}\left[\frac{1}{U}\left(\frac{1}{u}\frac{dU}{du}+\frac{d^2U}{du^2}\right)+\frac{1}{V}\left(\frac{1}{v}\frac{dV}{dv}+\frac{d^2V}{dv^2}\right)\right]=k^2, \tag{6}$$

so

$$\frac{d^2\Theta}{d\theta^2}=-(k^2+m^2)\Theta, \tag{7}$$

which has solution

$$\Theta(\theta)=A\cos(\sqrt{k^2+m^2}\,\theta)+B\sin(\sqrt{k^2+m^2}\,\theta), \tag{8}$$

and

$$\left[\frac{1}{U}\left(\frac{1}{u}\frac{dU}{du}+\frac{d^2U}{du^2}\right)\right]+\left[\frac{1}{V}\left(\frac{1}{v}\frac{dV}{dv}+\frac{d^2V}{dv^2}\right)\right]-k^2\frac{u^2+v^2}{u^2v^2}=0 \tag{9}$$

$$\left[\frac{1}{U}\left(\frac{1}{u}\frac{dU}{du}+\frac{d^2U}{du^2}\right)-\frac{k^2}{u^2}\right]+\left[\frac{1}{V}\left(\frac{1}{v}\frac{dV}{dv}+\frac{d^2V}{dv^2}\right)-\frac{k^2}{v^2}\right]=0. \tag{10}$$

This can be separated

$$\frac{1}{U}\left(\frac{1}{u}\frac{dU}{du}+\frac{d^2U}{du^2}\right)-\frac{k^2}{u^2}=c \tag{11}$$

$$\frac{1}{V}\left(\frac{1}{v}\frac{dV}{dv}+\frac{d^2V}{dv^2}\right)-\frac{k^2}{v^2}=-c, \tag{12}$$

so

$$u^2\frac{d^2U}{du^2}+\frac{dU}{du}-(c+k^2)U=0 \tag{13}$$

$$v^2\frac{d^2V}{dv^2}+\frac{dV}{dv}+(c-k^2)V=0. \tag{14}$$

References

Arfken, G. "Parabolic Coordinates (ξ,η,ϕ)." §2.12 in *Mathematical Methods for Physicists, 2nd ed.* Orlando, FL: Academic Press, pp. 109–111, 1970.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 514–515 and 660, 1953.

Helmholtz Differential Equation—Parabolic Cylindrical Coordinates

In PARABOLIC CYLINDRICAL COORDINATES, the SCALE FACTORS are $h_u = h_v = \sqrt{u^2+v^2}$, $h_z = 1$ and the separation functions are $f_1(u) = f_2(v) = f_3(z) = 1$, giving STÄCKEL DETERMINANT of $S = u^2 + v^2$. The HELMHOLTZ DIFFERENTIAL EQUATION is

$$\frac{1}{u^2+v^2}\left(\frac{\partial^2F}{\partial u^2}+\frac{\partial^2F}{\partial v^2}\right)+\frac{\partial^2F}{\partial z^2}+k^2=0. \tag{1}$$

Attempt SEPARATION OF VARIABLES by writing

$$F(u,v,z) \equiv U(u)V(v)Z(z), \tag{2}$$

then the HELMHOLTZ DIFFERENTIAL EQUATION becomes

$$\frac{1}{u^2+v^2}\left(VZ\frac{d^2U}{du^2}+UZ\frac{d^2V}{dv^2}\right)+UV\frac{d^2Z}{dz^2}+k^2UVZ=0. \tag{3}$$

Divide by UVZ,

$$\frac{1}{u^2+v^2}\left(\frac{1}{U}\frac{d^2U}{du^2}+\frac{1}{V}\frac{d^2V}{dv^2}\right)+\frac{1}{Z}\frac{d^2Z}{dz^2}+k^2=0. \tag{4}$$

Separating the Z part,

$$\frac{1}{Z}\frac{d^2Z}{dz^2}=-(k^2+m^2) \tag{5}$$

$$\frac{1}{u^2+v^2}\left(\frac{1}{U}\frac{d^2U}{du^2}+\frac{1}{V}\frac{d^2V}{dv^2}\right)-k^2=0 \tag{6}$$

$$\frac{1}{U}\frac{d^2U}{du^2}+\frac{1}{V}\frac{d^2V}{dv^2}-k^2(u^2+v^2)=0, \tag{7}$$

so

$$\frac{d^2Z}{dz^2}=-(k^2+m^2)Z, \tag{8}$$

which has solution

$$Z(z)=A\cos(\sqrt{k^2+m^2}\,z)+B\sin(\sqrt{k^2+m^2}\,z), \tag{9}$$

and

$$\left(\frac{1}{U}\frac{d^2U}{du^2} - k^2u^2\right) + \left(\frac{1}{V}\frac{d^2V}{dv^2} - k^2v^2\right) = 0. \quad (10)$$

This can be separated

$$\frac{1}{U}\frac{d^2U}{du^2} - k^2u^2 = c \quad (11)$$

$$\frac{1}{V}\frac{d^2V}{dv^2} - k^2v^2 = -c, \quad (12)$$

so

$$\frac{d^2U}{du^2} - (c + k^2u^2)U = 0 \quad (13)$$

$$\frac{d^2V}{dv^2} + (c - k^2v^2)V = 0. \quad (14)$$

These are the WEBER DIFFERENTIAL EQUATIONS, and the solutions are known as PARABOLIC CYLINDER FUNCTIONS.

see also PARABOLIC CYLINDER FUNCTION, PARABOLIC CYLINDRICAL COORDINATES, WEBER DIFFERENTIAL EQUATIONS

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 515 and 658, 1953.

Helmholtz Differential Equation—Polar Coordinates

In 2-D POLAR COORDINATES, attempt SEPARATION OF VARIABLES by writing

$$F(r,\theta) = R(r)\Theta(\theta), \quad (1)$$

then the HELMHOLTZ DIFFERENTIAL EQUATION becomes

$$\frac{d^2R}{dr^2}\Theta + \frac{1}{r}\frac{dR}{dr}\Theta + \frac{1}{r^2}\frac{d^2\Theta}{d\theta^2}R + k^2R\Theta = 0. \quad (2)$$

Divide both sides by $R\Theta$

$$\left(\frac{r^2}{R}\frac{d^2R}{dr^2} + \frac{r}{R}\frac{dR}{dr}\right) + \left(\frac{1}{\Theta}\frac{d^2\Theta}{d\theta^2} + k^2\right) = 0. \quad (3)$$

The solution to the second part of (3) must be periodic, so the differential equation is

$$\frac{d^2\Theta}{d\theta^2}\frac{1}{\Theta} = -(k^2 + m^2), \quad (4)$$

which has solutions

$$\Theta(\theta) = c_1 e^{i\sqrt{k^2+m^2}\,\theta} + c_2 e^{-i\sqrt{k^2+m^2}\,\theta}$$
$$= c_3 \sin(\sqrt{k^2+m^2}\,\theta) + c_4 \cos(\sqrt{k^2+m^2}\,\theta). \quad (5)$$

Plug (4) back into (3)

$$r^2R'' + rR' - m^2R = 0. \quad (6)$$

This is an EULER DIFFERENTIAL EQUATION with $\alpha \equiv 1$ and $\beta \equiv -m^2$. The roots are $r = \pm m$. So for $m = 0$, $r = 0$ and the solution is

$$R(r) = c_1 + c_2 \ln r. \quad (7)$$

But since $\ln r$ blows up at $r = 0$, the only possible physical solution is $R(r) = c_1$. When $m > 0$, $r = \pm m$, so

$$R(r) = c_1 r^m + c_2 r^{-m}. \quad (8)$$

But since r^{-m} blows up at $r = 0$, the only possible physical solution is $R_m(r) = c_1 r^m$. The solution for R is then

$$R_m(r) = c_m r^m \quad (9)$$

for $m = 0, 1, \ldots$ and the general solution is

$$F(r,\theta) = \sum_{m=0}^{\infty} [a_m r^m \sin(\sqrt{k^2+m^2}\,\theta)$$
$$+ b_m r^m \cos(\sqrt{k^2+m^2}\,\theta)]. \quad (10)$$

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 502–504, 1953.

Helmholtz Differential Equation—Prolate Spheroidal Coordinates

As shown by Morse and Feshbach (1953) and Arfken (1970), the HELMHOLTZ DIFFERENTIAL EQUATION is separable in PROLATE SPHEROIDAL COORDINATES.

References

Arfken, G. "Prolate Spheroidal Coordinates (u, v, φ)." §2.10 in *Mathematical Methods for Physicists, 2nd ed.* Orlando, FL: Academic Press, pp. 103–107, 1970.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 661, 1953.

Helmholtz Differential Equation—Spherical Coordinates

In SPHERICAL COORDINATES, the SCALE FACTORS are $h_r = 1$, $h_\theta = r\sin\phi$, $h_\phi = r$, and the separation functions are $f_1(r) = r^2$, $f_2(\theta) = 1$, $f_3(\phi) = \sin\phi$, giving a STÄCKEL DETERMINANT of $S = 1$. The LAPLACIAN is

$$\nabla^2 \equiv \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin^2\phi}\frac{\partial^2}{\partial\theta^2}$$
$$+ \frac{1}{r^2\sin\phi}\frac{\partial}{\partial\phi}\left(\sin\phi\frac{\partial}{\partial\phi}\right). \quad (1)$$

To solve the HELMHOLTZ DIFFERENTIAL EQUATION in SPHERICAL COORDINATES, attempt SEPARATION OF VARIABLES by writing

$$F(r,\theta,\phi) = R(r)\Theta(\theta)\Phi(\phi). \quad (2)$$

Then the HELMHOLTZ DIFFERENTIAL EQUATION becomes

$$\frac{d^2R}{dr^2}\Phi\Theta + \frac{2}{r}\frac{dR}{dr}\Phi\Theta + \frac{1}{r^2\sin^2\phi}\frac{d^2\Theta}{d\theta^2}\Phi R + \frac{\cos\phi}{r^2\sin\phi}\frac{d\Phi}{d\phi}\Theta R + \frac{1}{r^2}\frac{d^2\Phi}{d\phi^2}\Theta R = 0. \quad (3)$$

Now divide by $R\Theta\Phi$,

$$\frac{r^2\sin^2\phi}{\Phi R\Theta}\Phi\Theta\frac{d^2R}{dr^2} + \frac{2}{r}\frac{r^2\sin^2\phi}{\Phi R\Theta}\Phi\Theta\frac{dR}{dr} + \frac{1}{r^2\sin^2\phi}\frac{r^2\sin^2\phi}{\Phi R\Theta}\Phi R\frac{d^2\Theta}{d\theta^2} + \frac{\cos\phi}{r^2\sin\phi}\frac{r^2\sin^2\phi}{\Phi\Theta R}\frac{d\Phi}{d\phi}\Theta R + \frac{1}{r^2}\frac{r^2\sin^2\phi}{\Phi R\Theta}\frac{d^2\Phi}{d\phi^2}\Theta R = 0 \quad (4)$$

$$\left(\frac{r^2\sin^2\phi}{R}\frac{d^2R}{dr^2} + \frac{2r\sin^2\phi}{R}\frac{dR}{dr}\right) + \left(\frac{1}{\Theta}\frac{d^2\Theta}{d\theta^2}\right) + \left(\frac{\cos\phi\sin\phi}{\Phi}\frac{d\Phi}{d\phi} + \frac{\sin^2\phi}{\Phi}\frac{d^2\Phi}{d\phi^2}\right) = 0. \quad (5)$$

The solution to the second part of (5) must be sinusoidal, so the differential equation is

$$\frac{d^2\Theta}{d\theta^2}\frac{1}{\Theta} = -m^2, \quad (6)$$

which has solutions which may be defined either as a COMPLEX function with $m = -\infty, \ldots, \infty$

$$\Theta(\theta) = A_m e^{im\theta}, \quad (7)$$

or as a sum of REAL sine and cosine functions with $m = -\infty, \ldots, \infty$

$$\Theta(\theta) = S_m\sin(m\theta) + C_m\cos(m\theta). \quad (8)$$

Plugging (6) back into (7),

$$\frac{r^2}{R}\frac{d^2R}{dr^2} + \frac{2r}{R}\frac{dR}{dr} - \frac{1}{\sin^2\phi}\left(m^2 + \frac{\cos\phi\sin\phi}{\Phi}\right)\frac{d\Phi}{d\phi} + \frac{\sin^2\phi}{\Phi}\frac{d^2\Phi}{d\phi^2} = 0. \quad (9)$$

The radial part must be equal to a constant

$$\frac{r^2}{R}\frac{d^2R}{dr^2} + \frac{2r}{R}\frac{dR}{dr} = l(l+1) \quad (10)$$

$$r^2\frac{d^2R}{dr^2} + 2r\frac{dR}{dr} = l(l+1)R. \quad (11)$$

But this is the EULER DIFFERENTIAL EQUATION, so we try a series solution of the form

$$R = \sum_{n=0}^{\infty} a_n r^{n+c}. \quad (12)$$

Then

$$r^2\sum_{n=0}^{\infty}(n+c)(n+c-1)a_n r^{n+c-2} + 2r\sum_{n=0}^{\infty}(n+c)a_n r^{n+c-1} - l(l+1)\sum_{n=0}^{\infty}a_n r^{n+c} = 0 \quad (13)$$

$$\sum_{n=0}^{\infty}(n+c)(n+c-1)a_n r^{n+c} + 2\sum_{n=0}^{\infty}(n+c)a_n r^{n+c} - l(l+1)\sum_{n=0}^{\infty}a_n r^{n+c} = 0 \quad (14)$$

$$\sum_{n=0}^{\infty}[(n+c)(n+c+1) - l(l+1)]a_n r^{n+c} = 0. \quad (15)$$

This must hold true for all POWERS of r. For the r^c term (with $n = 0$),

$$c(c+1) = l(l+1), \quad (16)$$

which is true only if $c = l, -l-1$ and all other terms vanish. So $a_n = 0$ for $n \neq l, -l-1$. Therefore, the solution of the R component is given by

$$R_l(r) = A_l r^l + B_l r^{-l-1}. \quad (17)$$

Plugging (17) back into (9),

$$l(l+1) - \frac{m^2}{\sin^2\phi} + \frac{\cos\phi}{\sin\phi}\frac{1}{\Phi}\frac{d\Phi}{d\phi} + \frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2} = 0 \quad (18)$$

$$\Phi'' + \frac{\cos\phi}{\sin\phi}\Phi' + \left[l(l+1) - \frac{m^2}{\sin^2\phi}\right]\Phi = 0, \quad (19)$$

which is the associated LEGENDRE DIFFERENTIAL EQUATION for $x = \cos\phi$ and $m = 0, \ldots, l$. The general COMPLEX solution is therefore

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(A_l r^l + B_l r^{-l-1})P_l^m(\cos\phi)e^{-im\theta} \equiv \sum_{l=0}^{\infty}\sum_{m=-1}^{l}(A_l r^l + B_l r^{-l-1})Y_l^m(\theta,\phi), \quad (20)$$

where

$$Y_l^m(\theta,\phi) \equiv P_l^m(\cos\phi)e^{-im\theta} \quad (21)$$

are the (COMPLEX) SPHERICAL HARMONICS. The general REAL solution is

$$\sum_{l=0}^{\infty}\sum_{m=0}^{l}(A_l r^l + B_l r^{-l-1})P_l^m(\cos\phi) \times [S_m\sin(m\theta) + C_m\cos(m\theta)]. \quad (22)$$

Some of the normalization constants of P_l^m can be absorbed by S_m and C_m, so this equation may appear in the form

$$\sum_{l=0}^{\infty}\sum_{m=0}^{l}(A_l r^l + B_l r^{-l-1})P_l^m(\cos\phi) \times[S_l^m \sin(m\theta) + C_l^m \cos(m\theta)]$$

$$\equiv \sum_{l=0}^{\infty}\sum_{m=0}^{l}(A_l r^l + B_l r^{-l-1}) \times[S_l^m Y_l^{m(o)}(\theta,\phi) + C_l^m Y_l^{m(e)}(\theta,\phi)], \quad (23)$$

where

$$Y_l^{m(o)}(\theta,\phi) \equiv P_l^m(\cos\theta)\sin(m\theta) \quad (24)$$

$$Y_l^{m(e)}(\theta,\phi) \equiv P_l^m(\cos\theta)\cos(m\theta) \quad (25)$$

are the EVEN and ODD (real) SPHERICAL HARMONICS. If azimuthal symmetry is present, then $\Theta(\theta)$ is constant and the solution of the Φ component is a LEGENDRE POLYNOMIAL $P_l(\cos\phi)$. The general solution is then

$$F(r,\phi) = \sum_{l=0}^{\infty}(A_l r^l + B_l r^{-l-1})P_l(\cos\phi). \quad (26)$$

Actually, the equation is separable under the more general condition that k^2 is of the form

$$k^2(r,\theta,\phi) = f(r) + \frac{g(\theta)}{r^2} + \frac{h(\phi)}{r^2\sin\theta} + k'^2. \quad (27)$$

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 514 and 658, 1953.

Helmholtz Differential Equation—Spherical Surface

On the surface of a SPHERE, attempt SEPARATION OF VARIABLES in SPHERICAL COORDINATES by writing

$$F(\theta,\phi) = \Theta(\theta)\Phi(\phi), \quad (1)$$

then the HELMHOLTZ DIFFERENTIAL EQUATION becomes

$$\frac{1}{\sin^2\phi}\frac{d^2\Theta}{d\theta^2}\Phi + \frac{\cos\phi}{\sin\phi}\frac{d\Phi}{d\phi}\Theta + \frac{d^2\Phi}{d\phi^2}\Theta + k^2\Theta\Phi = 0. \quad (2)$$

Dividing both sides by $\Phi\Theta$,

$$\left(\frac{\cos\phi\sin\phi}{\Phi}\frac{d\Phi}{d\phi} + \frac{\sin^2\phi}{\Phi}\frac{d^2\Phi}{d\phi^2}\right) + \left(\frac{1}{\Theta}\frac{d^2\Theta}{d\theta^2} + k^2\right) = 0, \quad (3)$$

which can now be separated by writing

$$\frac{d^2\Theta}{d\theta^2}\frac{1}{\Theta} = -(k^2 + m^2). \quad (4)$$

The solution to this equation must be periodic, so m must be an INTEGER. The solution may then be defined either as a COMPLEX function

$$\Theta(\theta) = A_m e^{i\sqrt{k^2+m^2}\,\theta} + B_m e^{-i\sqrt{k^2+m^2}\,\theta} \quad (5)$$

for $m = -\infty, \ldots, \infty$, or as a sum of REAL sine and cosine functions

$$\Theta(\theta) = S_m \sin(\sqrt{k^2+m^2}\,\theta) + C_m \cos(\sqrt{k^2+m^2}\,\theta) \quad (6)$$

for $m = 0, \ldots, \infty$. Plugging (4) into (3) gives

$$\frac{\cos\phi\sin\phi}{\Phi}\frac{d\Phi}{d\phi} + \frac{\sin^2\phi}{\Phi}\frac{d^2\Phi}{d\phi^2} + m^2 = 0 \quad (7)$$

$$\Phi'' + \frac{\cos\phi}{\sin\phi}\Phi' + \frac{m^2}{\sin^2\phi}\Phi = 0, \quad (8)$$

which is the LEGENDRE DIFFERENTIAL EQUATION for $x = \cos\phi$ with

$$m^2 \equiv l(l+1), \quad (9)$$

giving

$$l^2 + l - m^2 = 0 \quad (10)$$

$$l = \tfrac{1}{2}(-1 \pm \sqrt{1+4m^2}). \quad (11)$$

Solutions are therefore LEGENDRE POLYNOMIALS with a COMPLEX index. The general COMPLEX solution is then

$$F(\theta,\phi) = \sum_{m=-\infty}^{\infty} P_l(\cos\phi)(A_m e^{im\theta} + B_m e^{-im\theta}), \quad (12)$$

and the general REAL solution is

$$F(\theta,\phi) = \sum_{m=0}^{\infty} P_l(\cos\phi)[S_m \sin(m\theta) + C_m \cos(m\theta)]. \quad (13)$$

Note that these solutions depend on only a single variable m. However, on the surface of a sphere, it is usual to express solutions in terms of the SPHERICAL HARMONICS derived for the 3-D spherical case, which depend on the two variables l and m.

Helmholtz Differential Equation—Toroidal Coordinates

The HELMHOLTZ DIFFERENTIAL EQUATION is not separable.

see LAPLACE'S EQUATION—TOROIDAL COORDINATES

Helmholtz's Theorem

Any VECTOR FIELD **v** satisfying

$$[\nabla \cdot \mathbf{v}]_\infty = 0 \qquad (1)$$

$$[\nabla \times \mathbf{v}]_\infty = 0 \qquad (2)$$

may be written as the sum of an IRROTATIONAL part and a SOLENOIDAL part,

$$\mathbf{v} = -\nabla\phi + \nabla \times \mathbf{A}, \qquad (3)$$

where for a VECTOR FIELD F,

$$\phi = -\int_V \frac{\nabla \cdot \mathbf{F}}{4\pi|\mathbf{r}' - \mathbf{r}|}\, d^3\mathbf{r}' \qquad (4)$$

$$\mathbf{A} = \int_V \frac{\nabla \times \mathbf{F}}{4\pi|\mathbf{r}' - \mathbf{r}|} d^3\mathbf{r}'. \qquad (5)$$

see also IRROTATIONAL FIELD, SOLENOIDAL FIELD, VECTOR FIELD

References

Arfken, G. "Helmholtz's Theorem." §1.15 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 78–84, 1985.

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 1084, 1980.

Helson-Szegő Measure

An absolutely continuous measure on ∂D whose density has the form $\exp(x + \bar{y})$, where x and y are real-valued functions in L^∞, $||y||_\infty < \pi/2$, exp is the EXPONENTIAL FUNCTION, and $||y||$ is the NORM.

Hemicylindrical Function

A function $S_n(z)$ which satisfies the RECURRENCE RELATION

$$S_{n-1}(z) - S_{n+1}(z) = 2S_n'(z)$$

together with

$$S_1(z) = -S_0'(z)$$

is called a hemicylindrical function.

References

Sonine, N. "Recherches sur les fonctions cylindriques et le développement des fonctions continues en séries." *Math. Ann.* **16**, 1–9 and 71–80, 1880.

Watson, G. N. "Hemi-Cylindrical Functions." §10.8 in *A Treatise on the Theory of Bessel Functions, 2nd ed.* Cambridge, England: Cambridge University Press, p. 353, 1966.

Hemisphere

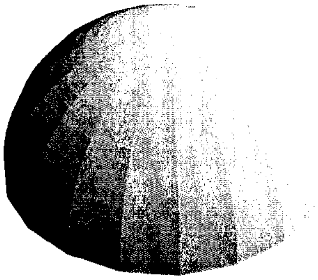

Half of a SPHERE cut by a PLANE passing through its CENTER. A hemisphere of RADIUS r can be given by the usual SPHERICAL COORDINATES

$$x = r\cos\theta\sin\phi \qquad (1)$$

$$y = r\sin\theta\sin\phi \qquad (2)$$

$$z = r\cos\phi, \qquad (3)$$

where $\theta \in [0, 2\pi)$ and $\phi \in [0, \pi/2]$. All CROSS-SECTIONS passing through the z-axis are SEMICIRCLES.

The VOLUME of the hemisphere is

$$V = \pi \int_0^r (r^2 - z^2)\, dz = \tfrac{2}{3}\pi r^3. \qquad (4)$$

The weighted mean of z over the hemisphere is

$$\langle z \rangle = \pi \int_0^r z(r^2 - z^2)\, dz = \tfrac{1}{4}\pi r^2. \qquad (5)$$

The CENTROID is then given by

$$\bar{z} = \frac{\langle z \rangle}{V} = \tfrac{3}{8} r \qquad (6)$$

(Beyer 1987).

see also SEMICIRCLE, SPHERE

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 133, 1987.

Hemispherical Function

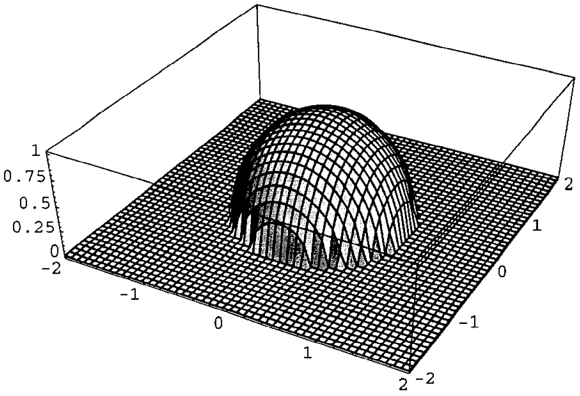

The hemisphere function is defined as

$$H(x,y) = \begin{cases} \sqrt{a-x^2-y^2} & \text{for } \sqrt{x^2+y^2} \leq a \\ 0 & \text{for } \sqrt{x^2+y^2} > a. \end{cases}$$

Watson (1966) defines a hemispherical function as a function S which satisfies the RECURRENCE RELATIONS

$$S_{n-1}(z) - S_{n+1}(z) = 2S_n{}'(z)$$

with

$$S_1(z) = -S_0'(z).$$

see also CYLINDER FUNCTION, CYLINDRICAL FUNCTION

References

Watson, G. N. *A Treatise on the Theory of Bessel Functions, 2nd ed.* Cambridge, England: Cambridge University Press, p. 353, 1966.

Hempel's Paradox

A purple cow is a confirming instance of the hypothesis that all crows are black.

References

Carnap, R. *Logical Foundations of Probability.* Chicago, IL: University of Chicago Press, pp. 224 and 469, 1950.

Gardner, M. *The Scientific American Book of Mathematical Puzzles & Diversions.* New York: Simon and Schuster, pp. 52–54, 1959.

Goodman, N. Ch. 3 in *Fact, Fiction, and Forecast.* Cambridge, MA: Harvard University Press, 1955.

Hempel, C. G. "A Purely Syntactical Definition of Confirmation." *J. Symb. Logic* **8**, 122–143, 1943.

Hempel, C. G. "Studies in Logic and Confirmation." *Mind* **54**, 1–26, 1945.

Hempel, C. G. "Studies in Logic and Confirmation. II." *Mind* **54**, 97–121, 1945.

Hempel, C. G. "A Note on the Paradoxes of Confirmation." *Mind* **55**, 1946.

Hosiasson-Lindenbaum, J. "On Confirmation." *J. Symb. Logic* **5**, 133–148, 1940.

Whiteley, C. H. "Hempel's Paradoxes of Confirmation." *Mind* **55**, 156–158, 1945.

Hendecagon

see UNDECAGON

Henneberg's Minimal Surface

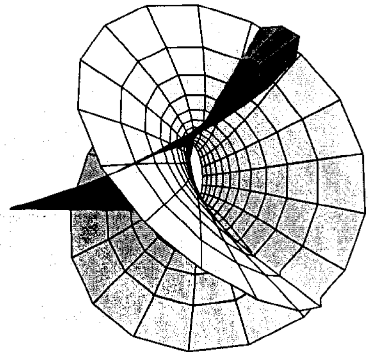

A double algebraic surface of 15th order and fifth class which can be given by parametric equations

$$x(u,v) = 2\sinh u \cos v - \tfrac{2}{3}\sinh(3u)\cos(3v) \quad (1)$$

$$y(u,v) = 2\sinh u \sin v - \tfrac{2}{3}\sinh(3u)\sin(3v) \quad (2)$$

$$z(u,v) = 2\cosh(2u)\cos(2v). \quad (3)$$

It can also be obtained from the ENNEPER-WEIERSTRAß PARAMETERIZATION with

$$f = 2 - 2z^{-4} \quad (4)$$

$$g = z. \quad (5)$$

see also MINIMAL SURFACE

References

Eisenhart, L. P. *A Treatise on the Differential Geometry of Curves and Surfaces.* New York: Dover, p. 267, 1960.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 446–448, 1993.

Nitsche, J. C. C. *Introduction to Minimal Surfaces.* Cambridge, England: Cambridge University Press, p. 144, 1989.

Hénon Attractor

see HÉNON MAP

Hénon-Heiles Equation

A nonlinear nonintegrable HAMILTONIAN SYSTEM with

$$\ddot{x} = -\frac{\partial V}{\partial x} \quad (1)$$

$$\ddot{y} = -\frac{\partial V}{\partial y}, \quad (2)$$

where

$$V(x,y) = \tfrac{1}{2}(x^2 + y^2 + 2x^2y - \tfrac{2}{3}y^3) \quad (3)$$

$$V(r,\theta) = \tfrac{1}{2}r^2 + \frac{1}{3}r^3 \sin(3\theta). \quad (4)$$

The energy is

$$E = V(x,y) + \tfrac{1}{2}(\dot{x}^2 + \dot{y}^2). \quad (5)$$

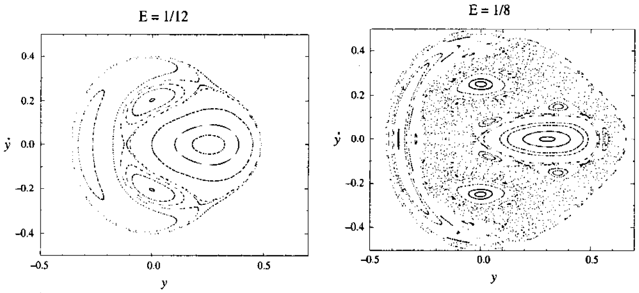

The above plots are SURFACES OF SECTION for $E = 1/12$ and $E = 1/8$. The Hamiltonian for a generalized Hénon-Heiles potential is

$$H = \tfrac{1}{2}(p_x{}^2 + p_y{}^2 + Ax^2 + By^2) + Dx^2y - \tfrac{1}{3}Cy^3. \quad (6)$$

The equations of motion are integrable only for

1. $D/C = 0$,
2. $D/C = -1, A/B = 1$,
3. $D/C = -1/6$, and
4. $D/C = -1/16, A/B = 1/6$.

References

Gleick, J. *Chaos: Making a New Science.* New York: Penguin Books, pp. 144–153, 1988.

Hénon, M. and Heiles, C. "The Applicability of the Third Integral of Motion: Some Numerical Experiments." *Astron. J.* **69**, 73–79, 1964.

Hénon Map

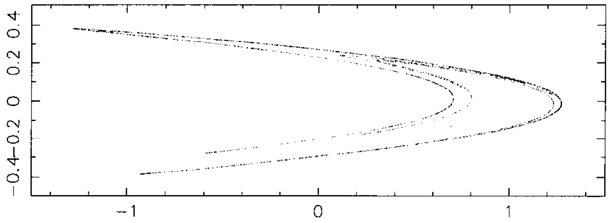

A quadratic 2-D Map given by the equations

$$x_{n+1} = 1 - \alpha {x_n}^2 + y_n \tag{1}$$

$$y_{n+1} = \beta x_n \tag{2}$$

or

$$x_{n+1} = x_n \cos\alpha - (y_n - {x_n}^2)\sin\alpha \tag{3}$$

$$y_{n+1} = x_n \sin\alpha + (y_n - {x_n}^2)\cos\alpha. \tag{4}$$

The above map is for $\alpha = 1.4$ and $\beta = 0.3$. The Hénon map has Correlation Exponent 1.25 ± 0.02 (Grassberger and Procaccia 1983) and Capacity Dimension 1.261 ± 0.003 (Russell *et al.* 1980). Hitzl and Zele (1985) give conditions for the existence of periods 1 to 6.

see also Bogdanov Map, Lozi Map, Quadratic Map

References

Dickau, R. M. "The Hénon Attractor." `http://forum.swarthmore.edu/advanced/robertd/henon.html`.

Gleick, J. *Chaos: Making a New Science.* New York: Penguin Books, pp. 144–153, 1988.

Grassberger, P. and Procaccia, I. "Measuring the Strangeness of Strange Attractors." *Physica D* **9**, 189–208, 1983.

Hitzl, D. H. and Zele, F. "An Exploration of the Hénon Quadratic Map." *Physica D* **14**, 305–326, 1985.

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 128–133, 1991.

Peitgen, H.-O. and Saupe, D. (Eds.). "A Chaotic Set in the Plane." §3.2.2 in *The Science of Fractal Images.* New York: Springer-Verlag, pp. 146–148, 1988.

Russell, D. A.; Hanson, J. D.; and Ott, E. "Dimension of Strange Attractors." *Phys. Rev. Let.* **45**, 1175–1178, 1980.

Hensel's Lemma

An important result in Valuation Theory which gives information on finding roots of Polynomials. Hensel's lemma is formally stated as follow. Let $(K, |\cdot|)$ be a complete non-Archimedean valuated field, and let R be the corresponding Valuation Ring. Let $f(x)$ be a Polynomial whose Coefficients are in R and suppose a_0 satisfies

$$|f(a_0)| < |f'(a_0)|^2, \tag{1}$$

where f' is the (formal) Derivative of f. Then there exists a unique element $a \in R$ such that $f(a) = 0$ and

$$|a - a_0| \leq \left|\frac{f(a_0)}{f'(a_0)}\right|. \tag{2}$$

Less formally, if $f(x)$ is a Polynomial with "Integer" Coefficients and $f(a_0)$ is "small" compared to $f'(a_0)$, then the equation $f(x) = 0$ has a solution "near" a_0. In addition, there are no other solutions near a_0, although there may be other solutions. The proof of the Lemma is based around the Newton-Raphson method and relies on the non-Archimedean nature of the valuation.

Consider the following example in which Hensel's lemma is used to determine that the equation $x^2 = -1$ is solvable in the 5-adic numbers $\mathbb{Q}_5$ (and so we can embed the Gaussian Integers inside $\mathbb{Q}_5$ in a nice way). Let K be the 5-adic numbers $\mathbb{Q}_5$, let $f(x) = x^2 + 1$, and let $a_0 = 2$. Then we have $f(2) = 5$ and $f'(2) = 4$, so

$$|f(2)|_5 = \tfrac{1}{5} < |f'(2)|_5^2 = 1, \tag{3}$$

and the condition is satisfied. Hensel's lemma then tells us that there is a 5-adic number a such that $a^2 + 1 = 0$ and

$$|a - 2|_5 <= |\tfrac{5}{4}|_5 = \tfrac{1}{5}. \tag{4}$$

Similarly, there is a 5-adic number b such that $b^2 + 1 = 0$ and

$$|b - 3|_5 <= |\tfrac{10}{7}|_5 = \tfrac{1}{5}. \tag{5}$$

Therefore, we have found both the square roots of -1 in $\mathbb{Q}_5$. It is possible to find the roots of any Polynomial using this technique.

Henstock-Kurzweil Integral

see HK Integral

Heptacontagon

A 70-sided Polygon.

Heptadecagon

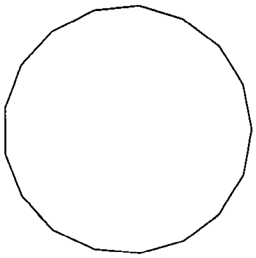

The Regular Polygon of 17 sides is called the Heptadecagon, or sometimes the Heptakaidecagon. Gauss proved in 1796 (when he was 19 years old) that the heptadecagon is Constructible with a Compass and Straightedge. Gauss's proof appears in his monumental work *Disquisitiones Arithmeticae.* The proof relies on the property of irreducible Polynomial equations that Roots composed of a finite number of Square Root extractions only exist when the order of the equation is a product of the form $2^a3^bF_c \cdot F_d \cdots F_e$, where the F_n are distinct Primes of the form

$$F_n = 2^{2^n} + 1,$$

known as Fermat Primes. Constructions for the regular Triangle (3^1), Square (2^2), Pentagon ($2^{2^1}+1$), Hexagon ($2^1 3^1$), etc., had been given by Euclid, but constructions based on the Fermat Primes ≥ 17 were unknown to the ancients. The first explicit construction of a heptadecagon was given by Erchinger in about 1800.

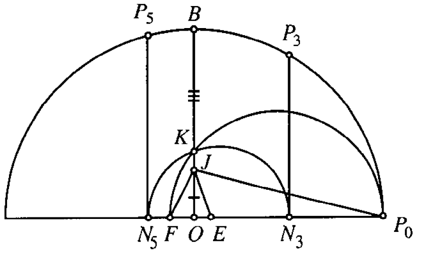

17-gon

The following elegant construction for the heptadecagon (Yates 1949, Coxeter 1969, Stewart 1977, Wells 1992) was first given by Richmond (1893).

1. Given an arbitrary point O, draw a Circle centered on O and a Diameter drawn through O.
2. Call the right end of the Diameter dividing the Circle into a Semicircle P_0.
3. Construct the Diameter Perpendicular to the original Diameter by finding the Perpendicular Bisector OB.
4. Find J a Quarter the way up OB.
5. Join JP_0 and find E so that $\angle OJE$ is a Quarter of $\angle OJP_0$.
6. Find F so that $\angle EJF$ is 45°.
7. Construct the Semicircle with Diameter FP_0.
8. This Semicircle cuts OB at K.
9. Draw a Semicircle with center E and Radius EK.
10. This cuts the extension of OP_0 at N_3.
11. Construct a line Perpendicular to OP_0 through N_3.
12. This line meets the original Semicircle at P_3.
13. You now have points P_0 and P_3 of a heptadecagon.
14. Use P_0 and P_3 to get the remaining 15 points of the heptadecagon around the original Circle by constructing P_0, P_3, P_6, P_9, P_{12}, P_{15}, P_1, P_4, P_7, P_{10}, P_{13}, P_{16}, P_2, P_5, P_8, P_{11}, and P_{14}.
15. Connect the adjacent points P_i.

This construction, when suitably streamlined, has Simplicity 53. The construction of Smith (1920) has a greater Simplicity of 58. Another construction due to Tietze (1965) and reproduced in Hall (1970) has a Simplicity of 50. However, neither Tietze (1965) nor Hall (1970) provides a proof that this construction is correct. Both Richmond's and Tietze's constructions require extensive calculations to prove their validity. De Temple (1991) gives an elegant construction involving the Carlyle Circles which has Geometrography symbol $8S_1 + 4S_2 + 22C_1 + 11C_3$ and Simplicity 45. The construction problem has now been automated to some extent (Bishop 1978).

see also 257-gon, 65537-gon, Compass, Constructible Polygon, Fermat Number, Fermat Prime, Regular Polygon, Straightedge, Trigonometry Values—$\pi/17$

References

Archibald, R. C. "The History of the Construction of the Regular Polygon of Seventeen Sides." *Bull. Amer. Math. Soc.* **22**, 239–246, 1916.

Archibald, R. C. "Gauss and the Regular Polygon of Seventeen Sides." *Amer. Math. Monthly* **27**, 323–326, 1920.

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 95–96, 1987.

Bishop, W. "How to Construct a Regular Polygon." *Amer. Math. Monthly* **85**, 186–188, 1978.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 201 and 229–230, 1996.

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 26–28, 1969.

De Temple, D. W. "Carlyle Circles and the Lemoine Simplicity of Polygonal Constructions." *Amer. Math. Monthly* **98**, 97–108, 1991.

Dixon, R. "Gauss Extends Euclid." §1.4 in *Mathographics.* New York: Dover, pp. 52–54, 1991.

Gauss, C. F. §365 and 366 in *Disquisitiones Arithmeticae.* Leipzig, Germany, 1801. New Haven, CT: Yale University Press, 1965.

Hall, T. *Carl Friedrich Gauss: A Biography.* Cambridge, MA: MIT Press, 1970.

Klein, F. *Famous Problems of Elementary Geometry and Other Monographs.* New York: Chelsea, 1956.

Ore, Ø. *Number Theory and Its History.* New York: Dover, 1988.

Rademacher, H. *Lectures on Elementary Number Theory.* New York: Blaisdell, 1964.

Richmond, H. W. "A Construction for a Regular Polygon of Seventeen Sides." *Quart. J. Pure Appl. Math.* **26**, 206–207, 1893.
Smith, L. L. "A Construction of the Regular Polygon of Seventeen Sides." *Amer. Math. Monthly* **27**, 322–323, 1920.
Stewart, I. "Gauss." *Sci. Amer.* **237**, 122–131, 1977.
Tietze, H. *Famous Problems of Mathematics.* New York: Graylock Press, 1965.
Wells, D. *The Penguin Dictionary of Curious and Interesting Geometry.* New York: Viking Penguin, 1992.
Yates, R. C. *Geometrical Tools.* St. Louis, MO: Educational Publishers, 1949.

Heptagon

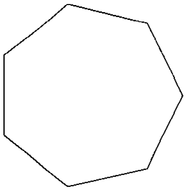

The unconstructible regular seven-sided POLYGON, illustrated above, has SCHLÄFLI SYMBOL {7}.

Although the regular heptagon is not a CONSTRUCTIBLE POLYGON, Dixon (1991) gives several close approximations. While the ANGLE subtended by a side is $360°/7 \approx 51.428571°$, Dixon gives constructions containing angles of $2\sin^{-1}(\sqrt{3}/4) \approx 51.317812°$, $\tan^{-1}(5/4) \approx 51.340191°$, and $30° + \sin^{-1}((\sqrt{3}-1)/2) \approx 51.470701°$.

Madachy (1979) illustrates how to construct a heptagon by folding and knotting a strip of paper.

see also EDMONDS' MAP, TRIGONOMETRY VALUES—$\pi/7$

References
Courant, R. and Robbins, H. "The Regular Heptagon." §3.3.4 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 138–139, 1996.
Dixon, R. *Mathographics.* New York: Dover, pp. 35–40, 1991.
Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 59–61, 1979.

Heptagonal Number

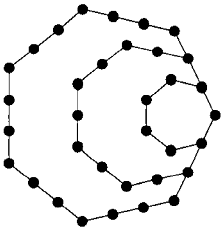

A FIGURATE NUMBER of the form $n(5n-3)/2$. The first few are 1, 7, 18, 34, 55, 81, 112, ... (Sloane's A000566). The GENERATING FUNCTION for the heptagonal numbers is

$$\frac{x(4x+1)}{(1-x)^3} = x + 7x^2 + 18x^3 + 34x^4 + \ldots.$$

References
Sloane, N. J. A. Sequence A000566/M4358 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Heptagonal Pyramidal Number

A PYRAMIDAL NUMBER of the form $n(n+1)(5n-2)/6$, The first few are 1, 8, 26, 60, 115, ... (Sloane's A002413). The GENERATING FUNCTION for the heptagonal pyramidal numbers is

$$\frac{x(4x+1)}{(x-1)^4} = x + 8x^2 + 26x^3 + 60x^4 + \ldots.$$

References
Sloane, N. J. A. Sequence A002413/M4498 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Heptahedron

The regular heptahedron is a one-sided surface made from four TRIANGLES and three QUADRILATERALS. It is topologically equivalent to the ROMAN SURFACE (Wells 1991). While all of the faces are regular and vertices equivalent, the heptahedron is self-intersecting and is therefore not considered an ARCHIMEDEAN SOLID. There are three semiregular heptahedra: the pentagonal and pentagrammic PRISMS, and a FACETED OCTAHEDRON (Holden 1991).

References
Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, p. 95, 1991.
Wells, D. *The Penguin Dictionary of Curious and Interesting Geometry.* New York: Viking Penguin, p. 98, 1992.

Heptakaidecagon

see HEPTADECAGON

Heptaparallelohedron

see CUBOCTAHEDRON

Heptomino

The heptominoes are the 7-POLYOMINOES. There are 108 different heptominoes.

see also HERSCHEL, PI HEPTOMINO, POLYOMINO

Herbrand's Theorem

Let an ideal class be in $\mathcal{A}$ if it contains an IDEAL whose lth power is PRINCIPAL. Let i be an ODD INTEGER $1 \leq i \leq l$ and define j by $i + j = 1$. Then $\mathcal{A}_1 = \langle e \rangle$. If $i \geq 3$ and $l \nmid B_j$, then $\mathcal{A}_i = \langle e \rangle$.

References
Ireland, K. and Rosen, M. "Herbrand's Theorem." §15.3 in *A Classical Introduction to Modern Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 241–248, 1990.

Hereditary Representation

The representation of a number as a sum of powers of a BASE b, followed by expression of each of the exponents as a sum of powers of b, etc., until the process stops. For example, the hereditary representation of 266 in base 2 is

$$\begin{aligned}266 &= 2^8 + 2^3 + 2\\ &= 2^{2^{2+1}} + 2^{2+1} + 2.\end{aligned}$$

see also GOODSTEIN SEQUENCE

Heredity

A property of a SPACE which is also true of each of its SUBSPACES. Being "COUNTABLE" is hereditary, but having a given GENUS is not.

Hermann's Formula

The MACHIN-LIKE FORMULA

$$\tfrac{1}{4}\pi = 2\tan^{-1}(\tfrac{1}{2}) - \tan^{-1}(\tfrac{1}{7}).$$

The other 2-term MACHIN-LIKE FORMULAS are EULER'S MACHIN-LIKE FORMULA, HUTTON'S FORMULA, and MACHIN'S FORMULA.

Hermann Grid Illusion

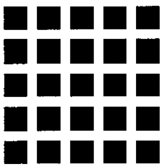

A regular 2-D arrangement of squares separated by vertical and horizontal "canals." Looking at the grid produces the illusion of gray spots in the white AREA between square VERTICES. The illusion was noted by Hermann (1870) while reading a book on sound by J. Tyndall.

References

Fineman, M. *The Nature of Visual Illusion.* New York: Dover, pp. 139–140, 1996.

Hermann-Hering Illusion

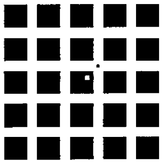

The illusion in view by staring at the small black dot for a half minute or so, then switching to the white dot. The black squares appear stationary when staring at the white dot, but a fainter grid of moving squares also appears to be present.

Hermann-Mauguin Symbol

A symbol used to represent the point and space groups (e.g., $2/m\bar{3}$). Some symbols have abbreviated form. The equivalence between Hermann-Mauguin symbols ("crystallographic symbol") and SCHÖNFLIES SYMBOLS for the POINT GROUPS is given by Cotton (1990).

see also POINT GROUPS

References

Cotton, F. A. *Chemical Applications of Group Theory, 3rd ed.* New York: Wiley, p. 379, 1990.

Hermit Point

see ISOLATED POINT

Hermite Constants

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

The Hermite constant is defined for DIMENSION n as the value

$$\gamma_n = \frac{\sup_f \min_{x_i} f(x_1, x_2, \ldots, x_n)}{[\text{discriminant}(f)]^{1/n}}$$

(Le Lionnais 1983). In other words, they are given by

$$\gamma_n = 4\left(\frac{\delta_n}{V_n}\right)^{2/n},$$

where δ_n is the maximum *lattice* PACKING DENSITY for HYPERSPHERE PACKING and V_n is the CONTENT of the n-HYPERSPHERE. The first few values of $(\gamma_n)^n$ are 1, 4/3, 2, 4, 8, 64/3, 64, 256, ... Values for larger n are not known.

For sufficiently large n,

$$\frac{1}{2\pi e} \leq \frac{\gamma_n}{n} \leq \frac{1.744\ldots}{2\pi e}.$$

see also HYPERSPHERE PACKING, KISSING NUMBER, SPHERE PACKING

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/hermit/hermit.html`.

Conway, J. H. and Sloane, N. J. A. *Sphere Packings, Lattices, and Groups, 2nd ed.* New York: Springer-Verlag, p. 20, 1993.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 38, 1983.

Hermite Differential Equation

$$\frac{d^2y}{dx^2} - 2x\frac{dy}{dx} + \lambda y = 0. \tag{1}$$

This differential equation has an irregular singularity at ∞. It can be solved using the series method

$$\sum_{n=0}^{\infty}(n+2)(n+1)a_{n+2}x^n - \sum_{n=1}^{\infty} 2na_nx^n + \sum_{n=0}^{\infty} \lambda a_nx^n = 0 \tag{2}$$

$$(2a_2+\lambda a_4)+\sum_{n=1}^{\infty}[(n+2)(n+1)a_{n+2}-2na_n+\lambda a_n]x^n = 0. \tag{3}$$

Therefore,

$$a_2 = -\frac{\lambda a_0}{2} \tag{4}$$

and

$$a_{n+2} = \frac{2n-\lambda}{(n+2)(n+1)}a_n \tag{5}$$

for $n = 1, 2, \ldots$. Since (4) is just a special case of (5),

$$a_{n+2} = \frac{2n-\lambda}{(n+2)(n+1)}a_n \tag{6}$$

for $n = 0, 1, \ldots$. The linearly independent solutions are then

$$y_1 = a_0\left[1-\frac{\lambda}{2!}x^2-\frac{(4-\lambda)\lambda}{4!}x^4 -\frac{(8-\lambda)(4-\lambda)\lambda}{6!}x^6-\ldots\right] \tag{7}$$

$$y_2 = a_1\left[x+\frac{(2-\lambda)}{3!}x^3+\frac{(6-\lambda)(2-\lambda)}{5!}x^5+\ldots\right]. \tag{8}$$

If $\lambda \equiv 4n = 0, 4, 8, \ldots$, then y_1 terminates with the POWER x^λ, and y_1 (normalized so that the COEFFICIENT of x^n is 2^n) is the regular solution to the equation, known as the HERMITE POLYNOMIAL. If $\lambda \equiv 4n+2 = 2, 6, 10, \ldots$, then y_2 terminates with the POWER x^λ, and y_2 (normalized so that the COEFFICIENT of x^n is 2^n) is the regular solution to the equation, known as the HERMITE POLYNOMIAL.

If $\lambda = 0$, then Hermite's differential equation becomes

$$y''-2xy' = 0, \tag{9}$$

which is of the form $P_2(x)y''+P_1(x)y' = 0$ and so has solution

$$\begin{aligned} y &= c_1\int\frac{dx}{\exp\left(\int\frac{P_1}{P_2}\,dx\right)}+c_2 \\ &= c_1\int\frac{dx}{\exp\int -2x\,dx}+c_2 \\ &= c_1\int\frac{dx}{e^{-x^2}}+c_2 = c_1\int e^{x^2}\,dx+c_2. \end{aligned} \tag{10}$$

Hermite-Gauss Quadrature

Also called HERMITE QUADRATURE. A GAUSSIAN QUADRATURE over the interval $(-\infty,\infty)$ with WEIGHTING FUNCTION $W(x) = e^{-x^2}$. The ABSCISSAS for quadrature order n are given by the roots of the HERMITE POLYNOMIALS $H_n(x)$, which occur symmetrically about 0. The WEIGHTS are

$$w_i = -\frac{A_{n+1}\gamma_n}{A_nH_n'(x_i)H_{n+1}(x_i)} = \frac{A_n}{A_{n-1}}\frac{\gamma_{n-1}}{H_{n-1}(x_i)H_n'(x_i)}, \tag{1}$$

where A_n is the COEFFICIENT of x^n in $H_n(x)$. For HERMITE POLYNOMIALS,

$$A_n = 2^n, \tag{2}$$

so

$$\frac{A_{n+1}}{A_n} = 2. \tag{3}$$

Additionally,

$$\gamma_n = \sqrt{\pi}\,2^n n!, \tag{4}$$

so

$$\begin{aligned} w_i &= -\frac{2^{n+1}n!\sqrt{\pi}}{H_{n+1}(x_i)H_n'(x_i)} \\ &= \frac{2^n(n-1)!\sqrt{\pi}}{H_{n-1}(x_i)H_n'(x_i)}. \end{aligned} \tag{5}$$

Using the RECURRENCE RELATION

$$H_n'(x) = 2nH_{n-1}(x) = 2xH_n(x)-H_{n+1}(x) \tag{6}$$

yields

$$H_n'(x_i) = 2nH_{n-1}(x_i) = -H_{n+1}(x_i) \tag{7}$$

and gives

$$w_i = \frac{2^{n+1}n!\sqrt{\pi}}{[H_n'(x_i)]^2} = \frac{2^{n+1}n!\sqrt{\pi}}{[H_{n+1}(x_i)]^2}. \tag{8}$$

The error term is

$$E = \frac{n!\sqrt{\pi}}{2^n(2n)!}f^{(2n)}(\xi). \tag{9}$$

Beyer (1987) gives a table of ABSCISSAS and weights up to n=12.

n	x_i	w_i
2	±0.707107	0.886227
3	0	1.18164
	±1.22474	0.295409
4	±0.524648	0.804914
	±1.65068	0.0813128
5	0	0.945309
	±0.958572	0.393619
	±2.02018	0.0199532

The ABSCISSAS and weights can be computed analytically for small n.

n	x_i	w_i
2	$\pm\frac{1}{2}\sqrt{2}$	$\frac{1}{2}\sqrt{\pi}$
3	0	$\frac{2}{3}\sqrt{\pi}$
	$\pm\frac{1}{2}\sqrt{6}$	$\frac{1}{6}\sqrt{\pi}$
4	$\pm\sqrt{\frac{3-\sqrt{6}}{2}}$	$\frac{\sqrt{\pi}}{4(3-\sqrt{6})}$
	$\pm\sqrt{\frac{3+\sqrt{6}}{2}}$	$\frac{\sqrt{\pi}}{4(3+\sqrt{6})}$

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 464, 1987.

Hildebrand, F. B. *Introduction to Numerical Analysis.* New York: McGraw-Hill, pp. 327–330, 1956.

Hermite Interpolation

see HERMITE'S INTERPOLATING FUNDAMENTAL POLYNOMIAL

Hermite's Interpolating Fundamental Polynomial

Let $l(x)$ be an nth degree POLYNOMIAL with zeros at $x_1, \ldots, x_m$. Then the fundamental POLYNOMIALS are

$$h_\nu^{(1)}(x) = \left[1 - \frac{l''(x_\nu)}{l'(x_\nu)}\right][l_\nu(x)]^2 \tag{1}$$

and

$$h_\nu^{(2)}(x) = (x - x_\nu)[l_\nu(x)]^2. \tag{2}$$

They have the properties

$$h_\nu^{(1)}(x_\mu) = \delta_{\nu\mu} \tag{3}$$

$$h^{(1)'}{}_\nu(x_\mu) = 0 \tag{4}$$

$$h^{(2)}(x_\mu) = 0 \tag{5}$$

$$h^{(2)'}(x_\mu) = \delta_{\nu\mu}. \tag{6}$$

Now let $f_1, \ldots, f_n$ and $f_1', \ldots, f_\nu'$ be values. Then the expansion

$$W_n(x) = \sum_{\nu=1}^n f_\nu h_\nu^{(1)}(x) + \sum_{\nu=1}^n f_\nu' h^{(2)}(x) \tag{7}$$

gives the unique HERMITE'S INTERPOLATING FUNDAMENTAL POLYNOMIAL for which

$$W_n(x_\nu) = f_\nu \tag{8}$$

$$W_n'(x_\nu) = f_\nu'. \tag{9}$$

If $f_\nu' = 0$, these are called STEP POLYNOMIALS. The fundamental POLYNOMIALS satisfy

$$h_1(x) + \ldots + h_n(x) = 1 \tag{10}$$

and

$$\sum_{\nu=1}^n x_\nu h_\nu^{(1)}(x) + \sum_{\nu=1}^n h_\nu^{(2)}(x) = x. \tag{11}$$

Also,

$$\int_a^b h_\nu^{(1)}(x)\,d\alpha(x) = \lambda_\nu \tag{12}$$

$$\int_a^b h_\nu^{(1)}(x)\,d\alpha(x) = 0 \tag{13}$$

$$\int_a^b x h_\nu'(x)\,d\alpha(x) = 0 \tag{14}$$

$$\int_a^b h_\nu^{(2)}(x)\,d\alpha(x) = 0 \tag{15}$$

$$\int_a^b h^{(2)'}{}_\nu\,d\alpha(x) = \lambda_\nu \tag{16}$$

$$\int_a^b x h^{(2)'}{}_\nu(x)\,dx = \lambda_\nu x_\nu, \tag{17}$$

for $\nu = 1, \ldots, n$.

References

Hildebrand, F. B. *Introduction to Numerical Analysis.* New York: McGraw-Hill, pp. 314–319, 1956.

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., pp. 330–332, 1975.

Hermite-Lindemann Theorem

The expression

$$A_1 e^{\alpha_1} + A_2 e^{\alpha_2} + A_3 e^{\alpha_3} + \ldots,$$

in which the COEFFICIENTS A_i differ from zero and in which the exponents α_i are ALGEBRAIC NUMBERS differing from each other, cannot equal zero.

see also ALGEBRAIC NUMBER, CONSTANT PROBLEM, INTEGER RELATION, LINDEMANN-WEIERSTRAß THEOREM

References

Dörrie, H. "The Hermite-Lindemann Transcendence Theorem." §26 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 128–137, 1965.

Hermite Polynomial

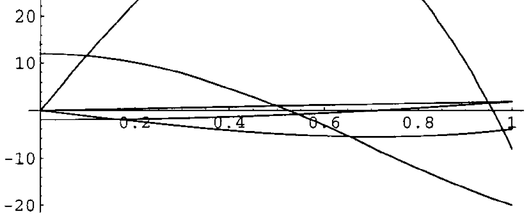

A set of ORTHOGONAL POLYNOMIALS. The Hermite polynomials $H_n(x)$ are illustrated above for $x \in [0,1]$ and $n = 1, 2, \ldots, 5$.

The GENERATING FUNCTION for Hermite polynomials is

$$\exp(2xt - t^2) \equiv \sum_{n=0}^{\infty} \frac{H_n(x)t^n}{n!}. \tag{1}$$

Using a TAYLOR SERIES shows that,

$$\begin{aligned} H_n(x) &= \left[\left(\frac{\partial}{\partial t}\right)^n \exp(2xt - t^2)\right]_{t=0} \\ &= \left[e^{x^2}\left(\frac{\partial}{\partial t}\right)^n e^{-(x-t)^2}\right]_{t=0}. \end{aligned} \tag{2}$$

Since $\partial f(x-t)/\partial t = -\partial f(x-t)/\partial x$,

$$\begin{aligned} H_n(x) &= (-1)^n e^{x^2}\left[\left(\frac{\partial}{\partial x}\right)^n e^{-(x-t)^2}\right]_{t=0} \\ &= (-1)^n e^{x^2}\frac{d^n}{dx^n}e^{-x^2}. \end{aligned} \tag{3}$$

Now define operators

$$\tilde{O}_1 \equiv -e^{x^2}\frac{d}{dx}e^{-x^2} \tag{4}$$

$$\tilde{O}_2 \equiv e^{x^2/2}\left(x - \frac{d}{dx}\right)e^{-x^2/2}. \tag{5}$$

It follows that

$$\tilde{O}_1 f = -e^{x^2}\frac{d}{dx}[fe^{-x^2}] = 2xf - \frac{df}{dx} \tag{6}$$

$$\begin{aligned} \tilde{O}_2 f &= e^{x^2/2}\left(x - \frac{d}{dx}\right)[fe^{-x^2/2}] \\ &= xf + xf - \frac{df}{dx} = 2xf - \frac{df}{dx}, \end{aligned} \tag{7}$$

so

$$\tilde{O}_1 = \tilde{O}_2, \tag{8}$$

and

$$-e^{x^2}\frac{d}{dx}e^{-x^2} = e^{x^2/2}\left(x - \frac{d}{dx}\right)e^{-x^2/2}, \tag{9}$$

which means the following definitions are equivalent:

$$\exp(2xt - t^2) \equiv \sum_{n=0}^{\infty}\frac{H_n(x)t^n}{n!} \tag{10}$$

$$H_n(x) \equiv (-1)^n e^{x^2}\frac{d^n}{dx^n}e^{-x^2} \tag{11}$$

$$H_n(x) \equiv e^{x^2/2}\left(x - \frac{d}{dx}\right)ne^{-x^2/2}. \tag{12}$$

The Hermite POLYNOMIALS are related to the derivative of the ERROR FUNCTION by

$$H_n(z) = (-1)^2\frac{\sqrt{\pi}}{2}e^{z^2}\frac{d^{n+1}}{dz^{n+1}}\operatorname{erf}(z). \tag{13}$$

They have a contour integral representation

$$H_n(x) = \frac{n!}{2\pi i}\int e^{-t^2+2tx}t^{-n-1}\,dt. \tag{14}$$

They are orthogonal in the range $(-\infty, \infty)$ with respect to the WEIGHTING FUNCTION e^{-x^2}

$$\int_{-\infty}^{\infty} H_n(x)H_m(x)e^{-x^2}\,dx = \delta_{mn}2^n n!\sqrt{\pi}. \tag{15}$$

Define the associated functions

$$u_n(x) \equiv \sqrt{\frac{a}{\pi^{1/2}n!2^n}}\,H_n(ax)e^{-a^2x^2/2}. \tag{16}$$

These obey the orthogonality conditions

$$\int_{-\infty}^{\infty} u_n(x)\frac{du_m}{dx}\,dx = \begin{cases} a\sqrt{\frac{n+1}{2}} & m = n+1 \\ -a\sqrt{\frac{n}{2}} & m = n-1 \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

$$\int_{-\infty}^{\infty} u_m(x)u_n(x)\,dx = \delta_{mn} \tag{18}$$

$$\int_{-\infty}^{\infty} u_m(x)xu_n(x)\,dx = \begin{cases} \frac{1}{a}\sqrt{\frac{n+!}{2}} & m = n+1 \\ \frac{1}{a}\sqrt{\frac{n}{2}} & m = n-1 \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

$$\int_{-\infty}^{\infty} u_m(x)x^2u_n(x)\,dx = \begin{cases} \frac{2n+1}{2a^2} & m = n \\ \frac{\sqrt{(n+1)(n+2)}}{2a^2} & m = n+2 \\ 0 & m \neq n \neq n \pm 2 \end{cases} \tag{20}$$

$$\int_{-\infty}^{\infty} e^{-x^2}H_\alpha H_\beta H_\gamma\,dx = \sqrt{\pi}\frac{2^s\alpha!\beta!\gamma!}{(s-\alpha)!(s-\beta)!(s-\gamma)!}, \tag{21}$$

if $\alpha + \beta + \gamma = 2s$ is EVEN and $s \geq \alpha$, $s \geq \beta$, and $s \geq \gamma$. Otherwise, the last integral is 0 (Szegő 1975, p. 390).

They also satisfy the RECURRENCE RELATIONS

$$H_{n+1} = 2xH_n(x) - 2nH_{n-1}(x) \tag{22}$$

$$H_n'(x) = 2nH_{n-1}(x). \tag{23}$$

The DISCRIMINANT is

$$D_n = 2^{3n(n-1)/2}\prod_{\nu=1}^{n}\nu^\nu \tag{24}$$

(Szegő 1975, p. 143).

An interesting identity is

$$\sum_{\nu=0}^{n}\binom{n}{\nu}H_\nu(x)H_{n-\nu}(y) = 2^{n/2}H_n[2^{-1/2}(x+y)]. \tag{25}$$

The first few POLYNOMIALS are

$$\begin{aligned}
H_0(x) &= 1\\
H_1(x) &= 2x\\
H_2(x) &= 4x^2 - 2\\
H_3(x) &= 8x^3 - 12x\\
H_4(x) &= 16x^4 - 48x^2 + 12\\
H_5(x) &= 32x^5 - 160x^3 + 120x\\
H_6(x) &= 64x^6 - 480x^4 + 720x^2 - 120\\
H_7(x) &= 128x^7 - 1344x^5 + 3360x^3 - 1680x\\
H_8(x) &= 256x^8 - 3594x^6 + 13440x^4 - 13440x^2\\
&\quad + 160\\
H_9(x) &= 512x^9 - 9216x^7 + 48384x^5 - 80640x^3\\
&\quad + 30240x\\
H_{10}(x) &= 1024x^{10} - 23040x^8 + 161280x^6 - 403200x^4\\
&\quad + 302400x^2 - 30240.
\end{aligned}$$

A class of generalized Hermite POLYNOMIALS $\gamma_n^m(x)$ satisfying

$$e^{mxt-t^m} = \sum_{n=0}^{\infty} \gamma_n^m(x)t^n \tag{26}$$

was studied by Subramanyan (1990). A class of related POLYNOMIALS defined by

$$h_{n,m} = \gamma_n^m\left(\frac{2x}{m}\right) \tag{27}$$

and with GENERATING FUNCTION

$$e^{2xt-t^m} = \sum_{n=0}^{\infty} h_{n,m}(x)t^n \tag{28}$$

was studied by Djordjević (1996). They satisfy

$$H_n(x) = n!h_{n,2}(x). \tag{29}$$

A modified version of the HERMITE POLYNOMIAL is sometimes defined by

$$\mathrm{He}_n(x) \equiv H_n\left(\frac{x}{\sqrt{2}}\right). \tag{30}$$

see also MEHLER'S HERMITE POLYNOMIAL FORMULA, WEBER FUNCTIONS

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Orthogonal Polynomials." Ch. 22 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 771–802, 1972.

Arfken, G. "Hermite Functions." §13.1 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 712–721, 1985.

Chebyshev, P. L. "Sur le développement des fonctions à une seule variable." *Bull. ph.-math., Acad. Imp. Sc. St. Pétersbourg* **1**, 193–200, 1859.

Chebyshev, P. L. *Oeuvres, Vol. 1.* New York: Chelsea, pp. 49–508, 1987.

Djordjević, G. "On Some Properties of Generalized Hermite Polynomials." *Fib. Quart.* **34**, 2–6, 1996.

Hermite, C. "Sur un nouveau développement en série de fonctions." *Compt. Rend. Acad. Sci. Paris* **58**, 93–100 and 266–273, 1864. Reprinted in Hermite, C. *Oeuvres complètes, Vol. 2.* Paris, pp. 293–308, 1908.

Hermite, C. *Oeuvres complètes, Vol. 3.* Paris, p. 432, 1912.

Iyanaga, S. and Kawada, Y. (Eds.). "Hermite Polynomials." Appendix A, Table 20.IV in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 1479–1480, 1980.

Sansone, G. "Expansions in Laguerre and Hermite Series." Ch. 4 in *Orthogonal Functions, rev. English ed.* New York: Dover, pp. 295–385, 1991.

Spanier, J. and Oldham, K. B. "The Hermite Polynomials $H_n(x)$." Ch. 24 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 217–223, 1987.

Subramanyan, P. R. "Springs of the Hermite Polynomials." *Fib. Quart.* **28**, 156–161, 1990.

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., 1975.

Hermite Quadrature

see HERMITE-GAUSS QUADRATURE

Hermite's Theorem

e is TRANSCENDENTAL.

Hermitian Form

A combination of variables x and y given by

$$axx^* + bxy^* + b^*x^*y + cyy^*,$$

where x^* and y^* are COMPLEX CONJUGATES.

Hermitian Matrix

If a MATRIX is SELF-ADJOINT, it is said to be a Hermitian matrix. Therefore, a Hermitian MATRIX is defined as one for which

$$\mathsf{A} = \mathsf{A}^\dagger, \tag{1}$$

where † denotes the ADJOINT MATRIX. Hermitian MATRICES have REAL EIGENVALUES with ORTHOGONAL EIGENVECTORS. For REAL MATRICES, Hermitian is the same as symmetrical. Any MATRIX C which is not Hermitian can be expressed as the sum of two Hermitian matrices

$$\mathsf{C} = \tfrac{1}{2}(\mathsf{C} + \mathsf{C}^\dagger) + \tfrac{1}{2}(\mathsf{C} - \mathsf{C}^\dagger). \tag{2}$$

Let U be a UNITARY MATRIX and A be a Hermitian matrix. Then the ADJOINT MATRIX of a SIMILARITY TRANSFORMATION is

$$\begin{aligned}
(\mathsf{UAU}^{-1})^\dagger &= [(\mathsf{UA})(\mathsf{U}^{-1})]^\dagger = (\mathsf{U}^{-1})^\dagger(\mathsf{UA})^\dagger\\
&= (\mathsf{U}^\dagger)^\dagger(\mathsf{A}^\dagger\mathsf{U}^\dagger) = \mathsf{UAU}^\dagger = \mathsf{UAU}^{-1}.
\end{aligned} \tag{3}$$

The specific matrix

$$\mathsf{H}(x,y,z)=\begin{bmatrix} z & x+iy \\ x-iy & -z \end{bmatrix}=x\mathsf{P}_1+y\mathsf{P}_2+z\mathsf{P}_3, \quad (4)$$

where P_i are PAULI SPIN MATRICES, is sometimes called "the" Hermitian matrix.

see also ADJOINT MATRIX, HERMITIAN OPERATOR, PAULI SPIN MATRICES

References

Arfken, G. "Hermitian Matrices, Unitary Matrices." §4.5 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 209–217, 1985.

Hermitian Operator

A Hermitian OPERATOR $\tilde{L}$ is one which satisfies

$$\int_a^b v^*\tilde{L}u\,dx=\int_a^b u\tilde{L}v^*\,dx. \quad (1)$$

As shown in STURM-LIOUVILLE THEORY, if $\tilde{L}$ is SELF-ADJOINT and satisfies the boundary conditions

$$[v^*pu']_{x=a}=[v^*pu']_{x=b}, \quad (2)$$

then it is automatically Hermitian. Hermitian operators have REAL EIGENVALUES, ORTHOGONAL EIGENFUNCTIONS, and the corresponding EIGENFUNCTIONS form a COMPLETE set when $\tilde{L}$ is second-order and linear. In order to prove that EIGENVALUES must be REAL and EIGENFUNCTIONS ORTHOGONAL, consider

$$\tilde{L}u_i+\lambda_i w u_i=0. \quad (3)$$

Assume there is a second EIGENVALUE λ_j such that

$$\tilde{L}u_j+\lambda_j w u_j=0 \quad (4)$$

$$\tilde{L}u_j{}^*+\lambda_j{}^* w u_j{}^*=0. \quad (5)$$

Now multiply (3) by $u_j{}^*$ and (5) by u_i

$$u_j{}^*\tilde{L}u_i+u_j{}^*\lambda_i w u_i=0 \quad (6)$$

$$u_i\tilde{L}u_j{}^*+u_i\lambda_j{}^* w u_j{}^*=0 \quad (7)$$

$$u_j{}^*\tilde{L}u_i-u_i\tilde{L}u_j{}^*=(\lambda_j{}^*-\lambda_i)wu_iu_j{}^*. \quad (8)$$

Now integrate

$$\int_a^b u_j{}^*\tilde{L}u_i-\int_a^b u_i\tilde{L}u_j{}^*=(\lambda_j{}^*-\lambda_i)\int_a^b wu_iu_j{}^*. \quad (9)$$

But because $\tilde{L}$ is Hermitian, the left side vanishes.

$$(\lambda_j{}^*-\lambda_i)\int_a^b wu_iu_j{}^*=0. \quad (10)$$

If EIGENVALUES λ_i and λ_j are not degenerate, then $\int_a^b wu_iu_j{}^*=0$, so the EIGENFUNCTIONS are ORTHOGONAL. If the EIGENVALUES are degenerate, the EIGENFUNCTIONS are not necessarily orthogonal. Now take $i=j$.

$$(\lambda_i{}^*-\lambda_i)\int_a^b wu_iu_i{}^*=0. \quad (11)$$

The integral cannot vanish unless $u_i=0$, so we have $\lambda_i{}^*=\lambda_i$ and the EIGENVALUES are real.

For a Hermitian operator $\tilde{O}$,

$$\langle\phi|\tilde{O}\psi\rangle=\langle\phi|\tilde{O}\psi\rangle^*=\langle\tilde{O}\phi|\psi\rangle. \quad (12)$$

In integral notation,

$$\int(\tilde{A}\phi)^*\psi\,dx=\int\phi^*\tilde{A}\psi\,dx. \quad (13)$$

Given Hermitian operators $\tilde{A}$ and $\tilde{B}$,

$$\langle\phi|\tilde{A}\tilde{B}\psi\rangle=\langle\tilde{A}\phi|\tilde{B}\psi\rangle=\langle\tilde{B}\tilde{A}\phi|\psi\rangle=\langle\phi|\tilde{B}\tilde{A}\psi\rangle^*. \quad (14)$$

Because, for a Hermitian operator $\tilde{A}$ with EIGENVALUE a,

$$\langle\psi|\tilde{A}\psi\rangle=\langle\tilde{A}\psi|\psi\rangle \quad (15)$$

$$a\langle\psi|\psi\rangle=a^*\langle\psi|\psi\rangle. \quad (16)$$

Therefore, either $\langle\psi|\psi\rangle=0$ or $a=a^*$. But $\langle\psi|\psi\rangle=0$ IFF $\psi=0$, so

$$\langle\psi|\psi\rangle\neq 0, \quad (17)$$

for a nontrivial EIGENFUNCTION. This means that $a=a^*$, namely that Hermitian operators produce REAL expectation values. Every observable must therefore have a corresponding Hermitian operator. Furthermore,

$$\langle\psi_n|\tilde{A}\psi_m\rangle=\langle\tilde{A}\psi_n|\psi_m\rangle \quad (18)$$

$$a_m\langle\psi_n|\psi_m\rangle=a_n{}^*\langle\psi_n|\psi_m\rangle=a_n\langle\psi_n|\psi_m\rangle, \quad (19)$$

since $a_n=a_n{}^*$. Then

$$(a_m-a_n)\langle\psi_n|\psi_m\rangle=0 \quad (20)$$

For $a_m\neq a_n$ (i.e., $\psi_n\neq\psi_m$),

$$\langle\psi_n|\psi_m\rangle=0. \quad (21)$$

For $a_m=a_n$ (i.e., $\psi_n=\psi_m$),

$$\langle\psi_n|\psi_m\rangle=\langle\psi_n|\psi_n\rangle\equiv 1. \quad (22)$$

Therefore,

$$\langle\psi_n|\psi_m\rangle=\delta_{nm}, \quad (23)$$

so the basis of EIGENFUNCTIONS corresponding to a Hermitian operator are ORTHONORMAL. Given two Hermitian operators $\tilde{A}$ and $\tilde{B}$,

$$(\tilde{A}\tilde{B})^\dagger = \tilde{B}^\dagger \tilde{A}^\dagger = \tilde{B}\tilde{A} = \tilde{A}\tilde{B} + [\tilde{B}, \tilde{A}], \tag{24}$$

the operator $\tilde{A}\tilde{B}$ equals $(\tilde{A}\tilde{B})^\dagger$, and is therefore Hermitian, only if

$$[\tilde{B}, \tilde{A}] = 0. \tag{25}$$

Given an arbitrary operator $\tilde{A}$,

$$\begin{aligned} \langle \psi_1 | (\tilde{A} + \tilde{A}^\dagger)\psi_2 \rangle &= \langle (\tilde{A}^\dagger + \tilde{A})\psi_1 | \psi_2 \rangle \\ &= \langle (\tilde{A} + \tilde{A}^\dagger)\psi_1 | \psi_2 \rangle , \end{aligned} \tag{26}$$

so $\tilde{A} + \tilde{A}^\dagger$ is Hermitian.

$$\begin{aligned} \langle \psi_1 | i(\tilde{A} - \tilde{A}^\dagger)\psi_2 \rangle &= \langle -i(\tilde{A}^\dagger - \tilde{A})\psi_1 | \psi_2 \rangle \\ &= \langle i(\tilde{A} - \tilde{A}^\dagger)\psi_1 | \psi_2 \rangle , \end{aligned} \tag{27}$$

so $i(\tilde{A} - \tilde{A}^\dagger)$ is Hermitian. Similarly,

$$\langle \psi_1 | (\tilde{A}\tilde{A}^\dagger)\psi_2 \rangle = \langle \tilde{A}^\dagger \psi_1 | \tilde{A} \dagger \psi_2 \rangle = \langle (\tilde{A}\tilde{A}^\dagger)\psi_1 | \psi_2 \rangle , \tag{28}$$

so $\tilde{A}\tilde{A}^\dagger$ is Hermitian.

Define the Hermitian conjugate operator $\tilde{A}^\dagger$ by

$$\langle \tilde{A}\psi | \psi \rangle \equiv \langle \psi | \tilde{A}^\dagger \psi \rangle . \tag{29}$$

For a Hermitian operator, $\tilde{A} = \tilde{A}^\dagger$. Furthermore, given two Hermitian operators $\tilde{A}$ and $\tilde{B}$,

$$\begin{aligned} \langle \psi_2 | (\tilde{A}\tilde{B})^\dagger \psi_1 \rangle = \langle (\tilde{A}\tilde{B})\psi_2 | \psi_1 \rangle &= \langle \tilde{B}\psi_2 | \tilde{A}^\dagger \psi_1 \rangle \\ = \langle \psi_2 | \tilde{B}^\dagger \tilde{A}^\dagger \psi_1 \rangle , \end{aligned} \tag{30}$$

so

$$(\tilde{A}\tilde{B})^\dagger = \tilde{B}^\dagger \tilde{A}^\dagger. \tag{31}$$

By further iterations, this can be generalized to

$$(\tilde{A}\tilde{B} \cdots \tilde{Z})^\dagger = \tilde{Z}^\dagger \cdots \tilde{B}^\dagger \tilde{A}^\dagger. \tag{32}$$

see also ADJOINT OPERATOR, HERMITIAN MATRIX, SELF-ADJOINT OPERATOR, STURM-LIOUVILLE THEORY

References

Arfken, G. "Hermitian (Self-Adjoint) Operators." §9.2 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 504–506 and 510–516, 1985.

Heron's Formula

Gives the AREA of a TRIANGLE in terms of the lengths of the sides a, b, and c and the SEMIPERIMETER

$$s \equiv \tfrac{1}{2}(a + b + c). \tag{1}$$

Heron's formula then states

$$\Delta = \sqrt{s(s-a)(s-b)(s-c)}. \tag{2}$$

Expressing the side lengths a, b, and c in terms of the radii a', b', and c' of the mutually tangent circles centered on the TRIANGLE vertices (which define the SODDY CIRCLES),

$$a = b' + c' \tag{3}$$

$$b = a' + c' \tag{4}$$

$$c = a' + b', \tag{5}$$

gives the particularly pretty form

$$\Delta = \sqrt{a'b'c'(a' + b' + c')} \tag{6}$$

The proof of this fact was discovered by Heron (ca. 100 BC–100 AD), although it was already known to Archimedes prior to 212 BC (Kline 1972). Heron's proof (Dunham 1990) is ingenious but extremely convoluted, bringing together a sequence of apparently unrelated geometric identities and relying on the properties of CYCLIC QUADRILATERALS and RIGHT TRIANGLES.

Heron's proof can be found in Proposition 1.8 of his work *Metrica.* This manuscript had been lost for centuries until a fragment was discovered in 1894 and a complete copy in 1896 (Dunham 1990, p. 118). More recently, writings of the Arab scholar Abu'l Raihan Muhammed al-Biruni have credited the formula to Heron's predecessor Archimedes (Dunham 1990, p. 127).

A much more accessible algebraic proof proceeds from the LAW OF COSINES,

$$\cos A = \frac{b^2 + c^2 - a^2}{2bc}. \tag{7}$$

Then

$$\sin A = \frac{\sqrt{-a^4 - b^4 - c^4 + 2b^2c^2 + 2c^2a^2 + 2a^2b^2}}{2bc}, \tag{8}$$

giving

$$\Delta = \tfrac{1}{2}bc \sin A \tag{9}$$

$$= \tfrac{1}{4}\sqrt{-a^4 - b^4 - c^4 + 2b^2c^2 + 2c^2a^2 + 2a^2b^2} \tag{10}$$

$$= \tfrac{1}{4}[(a+b+c)(-a+b+c)(a-b+c)(a+b-c)]^{1/2} \tag{11}$$

$$= \sqrt{s(s-a)(s-b)(s-c)} \tag{12}$$

(Coxeter 1969). Heron's formula contains the PYTHAGOREAN THEOREM.

see also BRAHMAGUPTA'S FORMULA, BRETSCHNEIDER'S FORMULA, HERONIAN TETRAHEDRON, HERONIAN TRIANGLE, SODDY CIRCLES, SSS THEOREM, TRIANGLE

References

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, p. 12, 1969.

Dunham, W. "Heron's Formula for Triangular Area." Ch. 5 in *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 113–132, 1990.

Kline, M. *Mathematical Thought from Ancient to Modern Times.* New York: Oxford University Press, 1972.

Pappas, T. "Heron's Theorem." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 62, 1989.

Heron Triangle

see HERONIAN TRIANGLE

Heronian Tetrahedron

A TETRAHEDRON with RATIONAL sides, FACE AREAS, and VOLUME. The smallest examples have pairs of opposite sides (148, 195, 203), (533, 875, 888), (1183, 1479, 1804), (2175, 2296, 2431), (1825, 2748, 2873), (2180, 2639, 3111), (1887, 5215, 5512), (6409, 6625, 8484), and (8619, 10136, 11275).

see also HERON'S FORMULA, HERONIAN TRIANGLE

References

Guy, R. K. "Simplexes with Rational Contents." §D22 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 190–192, 1994.

Heronian Triangle

A TRIANGLE with RATIONAL side lengths and RATIONAL AREA. Brahmagupta gave a parametric solution for *integer* Heronian triangles (the three side lengths and area can be multiplied by their LEAST COMMON MULTIPLE to make them all INTEGERS): side lengths $c(a^2+b^2)$, $b(a^2+c^2)$, and $(b+c)(a^2-bc)$, giving SEMIPERIMETER

$$s = a^2(b+c)$$

and AREA

$$\Delta = abc(a+b)(a^2-bc).$$

The first few integer Hernonian triangles, sorted by increasing maximal side lengths, are (3, 4, 5), (6, 8, 10), (5, 12, 13), (9, 12, 15), (4, 13, 15), (13, 14, 15), (9, 10, 17), ... (Sloane's A046128, A046129, and A046130), having areas 6, 24, 30, 54, 24, 84, 36, ... (Sloane's A046131).

Schubert (1905) claimed that Heronian triangles with two rational MEDIANS do not exist (Dickson 1952). This was shown to be incorrect by Buchholz and Rathbun (1997), who discovered six such triangles.

see also HERON'S FORMULA, MEDIAN (TRIANGLE), PYTHAGOREAN TRIPLE, TRIANGLE

References

Buchholz, R. H. *On Triangles with Rational Altitudes, Angle Bisectors or Medians.* Doctoral Dissertation. Newcastle, England: Newcastle University, 1989.

Buchholz, R. H. and Rathbun, R. L. "An Infinite Set of Heron Triangles with Two Rational Medians." *Amer. Math. Monthly* **104**, 107–115, 1997.

Dickson, L. E. *History of the Theory of Numbers, Vol. 2: Diophantine Analysis.* New York: Chelsea, pp. 199 and 208, 1952.

Guy, R. K. "Simplexes with Rational Contents." §D22 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 190–192, 1994.

Kraitchik, M. "Heronian Triangles." §4.13 in *Mathematical Recreations.* New York: W. W. Norton, pp. 104–108, 1942.

Schubert, H. "Die Ganzzahligkeit in der algebraischen Geometrie." In *Festgabe 48 Versammlung d. Philologen und Schulmänner zu Hamburg.* Leipzig, Germany, pp. 1–16, 1905.

Wells, D. G. *The Penguin Dictionary of Curious and Interesting Puzzles.* London: Penguin Books, p. 34, 1992.

Herschel

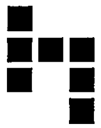

A HEPTOMINO shaped like the astronomical symbol for Uranus (which was discovered by William Herschel).

Herschfeld's Convergence Theorem

For real, NONNEGATIVE terms x_n and REAL p with $0 < p < 1$, the expression

$$\lim_{k\to\infty} x_0 + (x_1 + (x_2 + (\ldots + (x_k)^p)^p)^p)^p$$

converges IFF $(x_n)^{p^n}$ is bounded.

see also CONTINUED SQUARE ROOT

References

Herschfeld, A. "On Infinite Radicals." *Amer. Math. Monthly* **42**, 419–429, 1935.

Jones, D. J. "Continued Powers and a Sufficient Condition for Their Convergence." *Math. Mag.* **68**, 387–392, 1995.

Hesse's Theorem

If two pairs of opposite VERTICES of a COMPLETE QUADRILATERAL are pairs of CONJUGATE POINTS, then the third pair of opposite VERTICES is likewise a pair of CONJUGATE POINTS.

Hessenberg Matrix

A matrix of the form

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1(n-1)} & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2(n-1)} & a_{2n} \\ 0 & a_{32} & a_{33} & \cdots & a_{3(n-1)} & a_{3n} \\ 0 & 0 & a_{43} & \cdots & a_{4(n-1)} & a_{4n} \\ 0 & 0 & 0 & \cdots & a_{5(n-1)} & a_{5n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & a_{(n-1)(n-1)} & a_{(n-1)n} \\ 0 & 0 & 0 & 0 & a_{n(n-1)} & a_{nn} \end{bmatrix}.$$

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Reduction of a General Matrix to Hessenberg Form." §11.5 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 476–480, 1992.

Hessian Covariant

$$H \equiv |aa'a''|a_{x^{n-2}}a'_{x^{n-2}}a''_{x^{n-2}} = 0.$$

The nonsingular inflections of a curve are its nonsingular intersections with the Hessian.

References
Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, pp. 79, 95–98, and 151–161, 1959.

Hessian Determinant

The DETERMINANT

$$Hf(x,y) \equiv \begin{vmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{vmatrix}$$

appearing in the SECOND DERIVATIVE TEST as $D \equiv Hf(x,y)$.

see also SECOND DERIVATIVE TEST

References
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 1112–1113, 1979.

Heteroclinic Point

If intersecting stable and unstable MANIFOLDS (SEPARATRICES) emanate from FIXED POINTS of different families, they are called heteroclinic points.

see also HOMOCLINIC POINT

Heterogeneous Numbers

Two numbers are heterogeneous if their PRIME factors are distinct.

see also DISTINCT PRIME FACTORS, HOMOGENEOUS NUMBERS

References
Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 146, 1983.

Heterological Paradox

see GRELLING'S PARADOX

Heteroscedastic

A set of STATISTICAL DISTRIBUTIONS having different VARIANCES.

see also HOMOSCEDASTIC, VARIANCE

Heterosquare

9	8	7
2	1	6
3	4	5

1	2	3	4
5	6	7	8
9	10	11	12
13	14	16	15

A heterosquare is an $n \times n$ ARRAY of the integers from 1 to n^2 such that the rows, columns, and diagonals have different sums. (By contrast, in a MAGIC SQUARE, they have the *same* sum.) There are no heterosquares of order two, but heterosquares of every ODD order exist. They can be constructed by placing consecutive INTEGERS in a SPIRAL pattern (Fults 1974, Madachy 1979).

An ANTIMAGIC SQUARE is a special case of a heterosquare for which the sums of rows, columns, and main diagonals form a SEQUENCE of consecutive integers.

see also ANTIMAGIC SQUARE, MAGIC SQUARE, TALISMAN SQUARE

References
Duncan, D. "Problem 86." *Math. Mag.* **24**, 166, 1951.
Fults, J. L. *Magic Squares.* Chicago, IL: Open Court, 1974.
Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 101–103, 1979.
Weisstein, E. W. "Magic Squares." `http://www.astro.virginia.edu/~eww6n/math/notebooks/MagicSquares.m`.

Heuman Lambda Function

$$\Lambda_0(\phi|m) \equiv \frac{F(\phi|1-m)}{K(1-m)} + \frac{2}{\pi}K(m)Z(\phi|1-m),$$

where ϕ is the AMPLITUDE, m is the PARAMETER, Z is the JACOBI ZETA FUNCTION, and $F(\phi|m')$ and $K(m)$ are incomplete and complete ELLIPTIC INTEGRALS OF THE FIRST KIND.

see also ELLIPTIC INTEGRAL OF THE FIRST KIND, JACOBI ZETA FUNCTION

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 595, 1972.

Heun's Differential Equation

$$\frac{d^2w}{dx^2} + \left(\frac{\gamma}{x} + \frac{\delta}{x-1} + \frac{\epsilon}{x-a}\right)\frac{dw}{dx} + \frac{\alpha\beta x - q}{x(x-1)(x-a)}w = 0,$$

where

$$\alpha + \beta - \gamma - \delta - \epsilon + 1 = 0.$$

References
Erdelyi, A.; Magnus, W.; Oberhettinger, F.; and Tricomi, F. G. *Higher Transcendental Functions, Vol. 3.* Krieger, pp. 57–62, 1981.
Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, p. 576, 1990.

Heuristic

(1) Based on or involving trial and error. (2) Convincing without being rigorous.

Hex Game

A two-player GAME. There is a winning strategy for the first player if there is an even number of cells on each side; otherwise, there is a winning strategy for the second player.

References

Gardner, M. Ch. 8 in *The Scientific American Book of Mathematical Puzzles & Diversions.* New York, NY: Simon and Schuster, 1959.

Hex Number

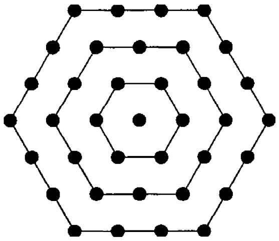

The CENTERED HEXAGONAL NUMBER given by

$$H_n = 1 + 6T_n = 2H_{n-1} - H_{n-2} + 6 = 3n^2 - 3n + 1,$$

where T_n is the nth TRIANGULAR NUMBER. The first few hex numbers are 1, 7, 19, 37, 61, 91, 127, 169, ... (Sloane's A003215). The GENERATING FUNCTION of the hex numbers is

$$\frac{x(x^2+4x+1)}{(1-x)^3} = x + 7x^2 + 19x^3 + 37x^4 + \ldots.$$

The first TRIANGULAR hex numbers are 1 and 91, and the first few SQUARE ones are 1, 169, 32761, 6355441, ... (Sloane's A006051). SQUARE hex numbers are obtained by solving the DIOPHANTINE EQUATION

$$3x^2 + 1 = y^2.$$

The only hex number which is SQUARE and TRIANGULAR is 1. There are no CUBIC hex numbers.

see also MAGIC HEXAGON, CENTERED SQUARE NUMBER, STAR NUMBER, TALISMAN HEXAGON

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 41, 1996.

Gardner, M. "Hexes and Stars." Ch. 2 in *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, 1988.

Hindin, H. "Stars, Hexes, Triangular Numbers, and Pythagorean Triples." *J. Recr. Math.* **16**, 191–193, 1983–1984.

Sloane, N. J. A. Sequences A003215/M4362 and A006051/M5409 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Hex (Polyhex)

see POLYHEX

Hex Pyramidal Number

A FIGURATE NUMBER which is equal to the CUBIC NUMBER n^3. The first few are 1, 8, 27, 64, ... (Sloane's A000578).

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 42–44, 1996.

Sloane, N. J. A. Sequence A000578/M4499 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Hexa

see POLYHEX

Hexabolo

A 6-POLYABOLO.

Hexacontagon

A 60-sided POLYGON.

Hexacronic Icositetrahedron

see GREAT HEXACRONIC ICOSITETRAHEDRON, SMALL HEXACRONIC ICOSITETRAHEDRON

Hexad

A SET of six.

see also MONAD, QUARTET, QUINTET, TETRAD, TRIAD

Hexadecagon

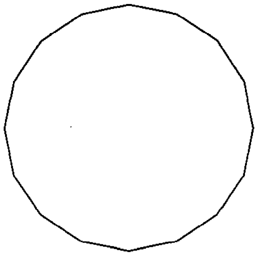

A 16-sided POLYGON, sometimes also called a HEXAKAIDECAGON.

see also POLYGON, REGULAR POLYGON, TRIGONOMETRY VALUES—$\pi/16$

Hexadecimal

The base 16 notational system for representing REAL NUMBERS. The digits used to represent numbers using hexadecimal NOTATION are 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, and F.

see also BASE (NUMBER), BINARY, DECIMAL, METADROME, OCTAL, QUATERNARY, TERNARY, VIGESIMAL

References

Weisstein, E. W. "Bases." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Bases.m.`

Hexaflexagon

A FLEXAGON made by folding a strip into adjacent EQUILATERAL TRIANGLES. The number of states possible in a hexaflexagon is the CATALAN NUMBER $C_4 = 42$.

see also FLEXAGON, FLEXATUBE, TETRAFLEXAGON

References
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 205–207, 1989.
Gardner, M. Ch. 1 in *The Scientific American Book of Mathematical Puzzles & Diversions.* New York: Simon and Schuster, 1959.
Gardner, M. Ch. 2 in *The Second Scientific American Book of Mathematical Puzzles & Diversions: A New Selection.* New York: Simon and Schuster, 1961.
Maunsell, F. G. "The Flexagon and the Hexaflexagon." *Math. Gazette* **38**, 213–214, 1954.
Wheeler, R. F. "The Flexagon Family." *Math. Gaz.* **42**, 1–6, 1958.

Hexagon

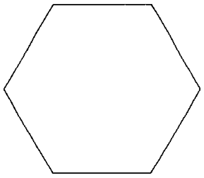

A six-sided POLYGON. In proposition IV.15, Euclid showed how to inscribe a regular hexagon in a CIRCLE. The INRADIUS r, CIRCUMRADIUS R, and AREA A can be computed directly from the formulas for a general regular POLYGON with side length s and $n = 6$ sides,

$$r = \tfrac{1}{2}s\cot\left(\frac{\pi}{6}\right) = \tfrac{1}{2}\sqrt{3}\,s \tag{1}$$

$$R = \tfrac{1}{2}s\csc\left(\frac{\pi}{6}\right) = s \tag{2}$$

$$A = \tfrac{1}{4}ns^2\cot\left(\frac{\pi}{6}\right) = \tfrac{3}{2}\sqrt{3}\,s^2. \tag{3}$$

Therefore, for a regular hexagon,

$$\frac{R}{r} = \sec\left(\frac{\pi}{6}\right) = \frac{2}{\sqrt{3}}, \tag{4}$$

so

$$\frac{A_R}{A_r} = \left(\frac{R}{r}\right)^2 = \frac{4}{3}. \tag{5}$$

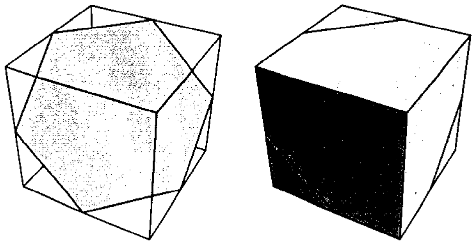

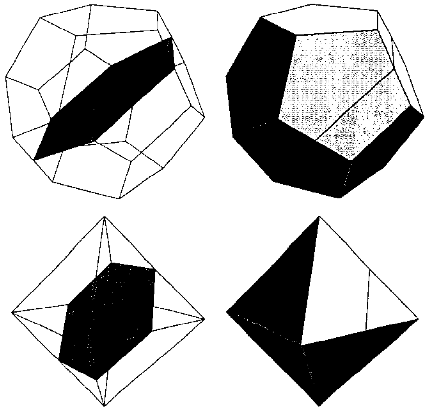

A PLANE PERPENDICULAR to a C_3 axis of a CUBE, DODECAHEDRON, or ICOSAHEDRON cuts the solid in a regular HEXAGONAL CROSS-SECTION (Holden 1991, pp. 22–23 and 27). For the CUBE, the PLANE passes through the MIDPOINTS of opposite sides (Steinhaus 1983, p. 170; Cundy and Rollett 1989, p. 157; Holden 1991, pp. 22–23). Since there are four such axes for the CUBE and OCTAHEDRON, there are four possibly hexagonal cross-sections. Since there are four such axes in each case, there are also four possibly hexagonal cross-sections.

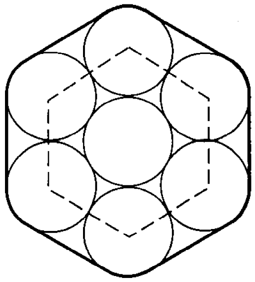

Take seven CIRCLES and close-pack them together in a hexagonal arrangement. The PERIMETER obtained by wrapping a band around the CIRCLE then consists of six straight segments of length d (where d is the DIAMETER) and 6 arcs with total length 1/6 of a CIRCLE. The PERIMETER is therefore

$$p = (12 + 2\pi)r = 2(6 + \pi)r. \tag{6}$$

see also CUBE, CYCLIC HEXAGON, DISSECTION, DODECAHEDRON, GRAHAM'S BIGGEST LITTLE HEXAGON, HEXAGON POLYIAMOND, HEXAGRAM, MAGIC HEXAGON, OCTAHEDRON, PAPPUS'S HEXAGON THEOREM, PASCAL'S THEOREM, TALISMAN HEXAGON

References
Cundy, H. and Rollett, A. "Hexagonal Section of a Cube." §3.15.1 in *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 157, 1989.
Dixon, R. *Mathographics.* New York: Dover, p. 16, 1991.
Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, 1991.
Pappas, T. "Hexagons in Nature." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 74–75, 1989.
Steinhaus, H. *Mathematical Snapshots, 3rd American ed.* New York: Oxford University Press, 1983.

Hexagon Polyiamond

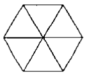

A 6-POLYIAMOND.

see also HEXAGON

References

Golomb, S. W. *Polyominoes: Puzzles, Patterns, Problems, and Packings, 2nd ed.* Princeton, NJ: Princeton University Press, p. 92, 1994.

Hexagonal Number

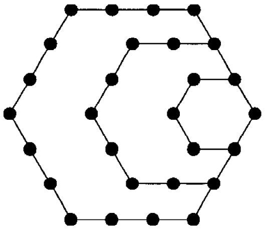

A FIGURATE NUMBER and 6-POLYGONAL NUMBER of the form $n(2n-1)$. The first few are 1, 6, 15, 28, 45, ... (Sloane's A000384). The GENERATING FUNCTION of the hexagonal numbers

$$\frac{x(3x+1)}{(1-x)^3} = x + 6x^2 + 15x^3 + 28x^4 + \ldots.$$

Every hexagonal number is a TRIANGULAR NUMBER since

$$r(2r-1) = \tfrac{1}{2}(2r-1)[(2r-1)+1].$$

In 1830, Legendre (1979) proved that every number larger than 1791 is a sum of four hexagonal numbers, and Duke and Schulze-Pillot (1990) improved this to three hexagonal numbers for every sufficiently large integer. The numbers 11 and 26 can only be represented as a sum using the maximum possible of six hexagonal numbers:

$$\begin{aligned} 11 &= 1+1+1+1+1+6 \\ 26 &= 1+1+6+6+6+6. \end{aligned}$$

see also FIGURATE NUMBER, HEX NUMBER, TRIANGULAR NUMBER

References

Duke, W. and Schulze-Pillot, R. "Representations of Integers by Positive Ternary Quadratic Forms and Equidistribution of Lattice Points on Ellipsoids." *Invent. Math.* **99**, 49–57, 1990.

Guy, R. K. "Sums of Squares." §C20 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 136–138, 1994.

Legendre, A.-M. *Théorie des nombres, 4th ed., 2 vols.* Paris: A. Blanchard, 1979.

Sloane, N. J. A. Sequence A000384/M4108 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Hexagonal Pyramidal Number

A PYRAMIDAL NUMBER of the form $n(n+1)(4n-1)/6$, The first few are 1, 7, 22, 50, 95, ... (Sloane's A002412). The GENERATING FUNCTION of the hexagonal pyramidal numbers is

$$\frac{x(3x+1)}{(x-1)^4} = x + 7x^2 + 22x^3 + 50x^4 + \ldots.$$

References

Sloane, N. J. A. Sequence A002412/M4374 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Hexagonal Scalenohedron

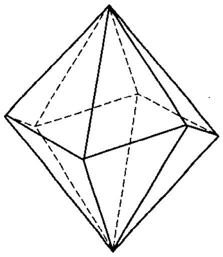

An irregular DODECAHEDRON which is also a TRAPEZOHEDRON.

see also DODECAHEDRON, TRAPEZOHEDRON

References

Cotton, F. A. *Chemical Applications of Group Theory, 3rd ed.* New York: Wiley, p. 63, 1990.

Hexagonal Tiling

see TILING

Hexagram

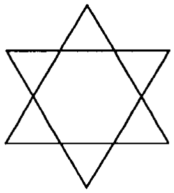

The STAR POLYGON $\{6,2\}$, also known as the STAR OF DAVID.

see also DISSECTION, PENTAGRAM, SOLOMON'S SEAL KNOT, STAR FIGURE, STAR OF LAKSHMI

Hexagrammum Mysticum Theorem

see PASCAL'S THEOREM

Hexahedron

A hexahedron is a six-sided POLYHEDRON. The regular hexahedron is the CUBE, although there are seven topologically different CONVEX hexahedra (Guy 1994).

see also CUBE

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 189, 1994.

Hexahemioctahedron

The DUAL POLYHEDRON of the CUBOHEMIOCTAHEDRON.

Hexakaidecagon

see HEXADECAGON

Hexakis Icosahedron

see DISDYAKIS TRIACONTAHEDRON

Hexakis Octahedron

see DISDYAKIS DODECAHEDRON

Hexlet

Also called SODDY'S HEXLET. Consider three mutually tangent SPHERES A, B, and C. Then construct a chain of SPHERES tangent to each of A, B, and C threading and interlocking with the $A-B-C$ ring. Surprisingly, every chain closes into a "necklace" after six SPHERES regardless of where the first SPHERE is placed. This is a special case of KOLLROS' THEOREM. The centers of a Soddy hexlet always lie on an ELLIPSE (Ogilvy 1990, p. 63).

see also COXETER'S LOXODROMIC SEQUENCE OF TANGENT CIRCLES, KOLLROS' THEOREM, STEINER CHAIN

References
Coxeter, H. S. M. "Interlocking Rings of Spheres." *Scripta Math.* **18**, 113–121, 1952.
Gosset, T. "The Hexlet." *Nature* **139**, 251–252, 1937.
Honsberger, R. *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., pp. 49–50, 1976.
Morley, F. "The Hexlet." *Nature* **139**, 72–73, 1937.
Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 60–72, 1990.
Soddy, F. "The Bowl of Integers and the Hexlet." *Nature* **139**, 77–79, 1937.
Soddy, F. "The Hexlet." *Nature* **139**, 154 and 252, 1937.

HexLife

An alternative LIFE game similar to Conway's, which is played on a hexagonal grid. No set of rules has yet emerged as uniquely interesting.

see also HIGHLIFE

References
Hensel, A. "A Brief Illustrated Glossary of Terms in Conway's Game of Life." `http://www.cs.jhu.edu/~callahan/glossary.html`.

Hexomino

One of the 35 6-POLYOMINOES.

References
Pappas, T. "Triangular, Square & Pentagonal Numbers." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 214, 1989.

Heyting Algebra

An ALGEBRA which is a special case of a LOGOS.

see also LOGOS, TOPOS

Hh Function

Let

$$Z(x) \equiv \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad (1)$$

$$Q(x) \equiv \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\,dt, \qquad (2)$$

where Z and Q are closely related to the NORMAL DISTRIBUTION FUNCTION, then

$$\mathrm{Hh}_{-n}(x) = (-1)^{n-1}\sqrt{2\pi}\, z^{(n-1)}(x) \qquad (3)$$

$$\mathrm{Hh}_n(x) = \frac{(-1)^n}{n!} \mathrm{Hh}_{-1}(x) \frac{d^n}{dx^n}\left[\frac{Q(x)}{Z(x)}\right]. \qquad (4)$$

see also NORMAL DISTRIBUTION FUNCTION, TETRACHORIC FUNCTION

Hi-Q

A triangular version of PEG SOLITAIRE with 15 holes and 14 pegs. Numbering hole 1 at the apex of the triangle and thereafter from left to right on the next lower row, etc., the following table gives possible ending holes for a single peg removed (Beeler *et al.* 1972, Item 76). Because of symmetry, only the first five pegs need be considered. Also because of symmetry, removing peg 2 is equivalent to removing peg 3 and flipping the board horizontally.

remove	possible ending pegs
1	1, 7 = 10, 13
2	2, 6, 11, 14
4	3 = 12, 4, 9, 15
5	13

References
Beeler, M.; Gosper, R. W.; and Schroeppel, R. Item 75 in *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Higher Arithmetic

An archaic term for NUMBER THEORY.

Highest Weight Theorem

A theorem proved by É. Cartan (1913) which classifies the irreducible representations of COMPLEX semisimple LIE ALGEBRAS.

References
Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

HighLife

An alternate set of LIFE rules similar to Conway's, but with the additional rule that six neighbors generate a birth. Most of the interest in this variant is due to the presence of a so-called replicator.

see also HEXLIFE, LIFE

References
Hensel, A. "A Brief Illustrated Glossary of Terms in Conway's Game of Life." `http://www.cs.jhu.edu/~callahan/glossary.html`.

Highly Abundant Number

see HIGHLY COMPOSITE NUMBER

Highly Composite Number

A COMPOSITE NUMBER (also called a SUPERABUNDANT NUMBER) is a number n which has more FACTORS than any other number less than n. In other words, $\sigma(n)/n$ exceeds $\sigma(k)/k$ for all $k < n$, where $\sigma(n)$ is the DIVISOR FUNCTION. They were called highly composite numbers by Ramanujan, who found the first 100 or so, and superabundant by Alaoglu and Erdős (1944).

There are an infinite numbers of highly composite numbers, and the first few are 2, 4, 6, 12, 24, 36, 48, 60, 120, 180, 240, 360, 720, 840, 1260, 1680, 2520, 5040, ... (Sloane's A002182). Ramanujan (1915) listed 102 up to 6746328388800 (but omitted 293, 318, 625, 600, and 29331862500). Robin (1983) gives the first 5000 highly composite numbers, and a comprehensive survey is given by Nicholas (1988).

If

$$N = 2^{a_2}3^{a_3}\cdots p^{a_p} \tag{1}$$

is the prime decomposition of a highly composite number, then

1. The PRIMES 2, 3, ..., p form a string of consecutive PRIMES,
2. The exponents are nonincreasing, so $a_2 \geq a_3 \geq \ldots \geq a_p$, and
3. The final exponent a_p is always 1, except for the two cases $N = 4 = 2^2$ and $N = 36 = 2^2 \cdot 3^2$, where it is 2.

Let $Q(x)$ be the number of highly composite numbers $\leq x$. Ramanujan (1915) showed that

$$\lim_{x\to\infty} \frac{Q(x)}{\ln x} = \infty. \tag{2}$$

Erdős (1944) showed that there exists a constant $c_1 > 0$ such that

$$Q(x) \geq (\ln x)^{1+c_1} \tag{3}$$

Nicholas proved that there exists a constant $c_2 > 0$ such that

$$Q(x) \ll (\ln x)^{c_2}. \tag{4}$$

see also ABUNDANT NUMBER

References

Alaoglu, L. and Erdős, P. "On Highly Composite and Similar Numbers." *Trans. Amer. Math. Soc.* **56**, 448–469, 1944.

Andree, R. V. "Ramanujan's Highly Composite Numbers." *Abacus* **3**, 61–62, 1986.

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, p. 53, 1994.

Dickson, L. E. *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, p. 323, 1952.

Flammenkamp, A. `http://www.minet.uni-jena.de/~achim/highly.html`.

Honsberger, R. *Mathematical Gems I.* Washington, DC: Math. Assoc. Amer., p. 112, 1973.

Honsberger, R. "An Introduction to Ramanujan's Highly Composite Numbers." Ch. 14 in *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 193–207, 1985.

Kanigel, R. *The Man Who Knew Infinity: A Life of the Genius Ramanujan.* New York: Washington Square Press, p. 232, 1991.

Nicholas, J.-L. "On Highly Composite Numbers." In *Ramanujan Revisited: Proceedings of the Centenary Conference* (Ed. G. E. Andrews, B. C. Berndt, and R. A. Rankin). Boston, MA: Academic Press, pp. 215–244, 1988.

Ramanujan, S. "Highly Composite Numbers." *Proc. London Math. Soc.* **14**, 347–409, 1915.

Ramanujan, S. *Collected Papers.* New York: Chelsea, 1962.

Robin, G. "Méthodes d'optimalisation pour un problème de théories des nombres." *RAIRO Inform. Théor.* **17**, 239–247, 1983.

Sloane, N. J. A. Sequence A002182/M1025 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Wells, D. *The Penguin Dictionary of Curious and Interesting Numbers.* New York: Penguin Books, p. 128, 1986.

Higman-Sims Group

The SPORADIC GROUP *HS*.

References

Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas/HS.html`.

Hilbert's Axioms

The 21 assumptions which underlie the GEOMETRY published in Hilbert's classic text *Grundlagen der Geometrie.* The eight INCIDENCE AXIOMS concern collinearity and intersection and include the first of EUCLID'S POSTULATES. The four ORDERING AXIOMS concern the arrangement of points, the five CONGRUENCE AXIOMS concern geometric equivalence, and the three CONTINUITY AXIOMS concern continuity. There is also a single parallel axiom equivalent to Euclid's PARALLEL POSTULATE.

see also CONGRUENCE AXIOMS, CONTINUITY AXIOMS, INCIDENCE AXIOMS, ORDERING AXIOMS, PARALLEL POSTULATE

References

Hilbert, D. *The Foundations of Geometry, 2nd ed.* Chicago, IL: Open Court, 1980.

Iyanaga, S. and Kawada, Y. (Eds.). "Hilbert's System of Axioms." §163B in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 544–545, 1980.

Hilbert Basis Theorem

If R is a NOETHERIAN RING, then $S = R[X]$ is also a NOETHERIAN RING.

see also ALGEBRAIC VARIETY, FUNDAMENTAL SYSTEM, SYZYGY

References

Hilbert, D. "Über die Theorie der algebraischen Formen." *Math. Ann.* **36**, 473–534, 1890.

Hilbert's Constants

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Extend HILBERT'S INEQUALITY by letting $p, q > 1$ and

$$\frac{1}{p}+\frac{1}{q}\geq 1, \tag{1}$$

so that

$$0<\lambda=2-\frac{1}{p}-\frac{1}{q}\leq 1. \tag{2}$$

Levin (1937) and Stečkin (1949) showed that

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\frac{a_m b_n}{(m+n)^{\lambda}}\leq\left\{\pi\csc\left[\frac{\pi(q-1)}{\lambda q}\right]\right\}^{\lambda} \times\left(\int_0^{\infty}[f(x)]^p\,dx\right)^{1/p}\left(\int_0^{\infty}[g(x)]^q\,dx\right)^{1/q} \tag{3}$$

and

$$\int_0^{\infty}\int_0^{\infty}\frac{f(x)g(y)}{(x+y)^{\lambda}}\,dx\,dy<\pi\csc\left[\frac{\pi(q-1)}{p}\right]^{\lambda} \times\left(\int_0^{\infty}[f(x)]^p\,dx\right)^{1/p}\left(\int_0^{\infty}[g(x)]^q\,dx\right)^{1/q}. \tag{4}$$

Mitrinovic *et al.* (1991) indicate that this constant is the best possible.

see also HILBERT'S INEQUALITY

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/hilbert/hilbert.html`.

Mitrinovic, D. S.; Pecaric, J. E.; and Fink, A. M. *Inequalities Involving Functions and Their Integrals and Derivatives.* Dordrecht, Netherlands: Kluwer, 1991.

Stečkin, S. B. "On the Degree of Best Approximation to Continuous Functions." *Dokl. Akad. Nauk SSSR* **65**, 135–137, 1949.

Hilbert Curve

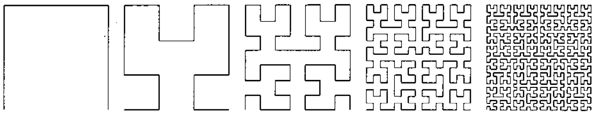

A LINDENMAYER SYSTEM invented by Hilbert (1891) whose limit is a PLANE-FILLING CURVE which fills a square. Traversing the VERTICES of an n-D HYPERCUBE in GRAY CODE order produces a generator for the n-D Hilbert curve (Goetz). The Hilbert curve can be simply encoded with initial string `"L"`, STRING REWRITING rules `"L" -> "+RF-LFL-FR+"`, `"R"->"-LF+RFR+FL-"`, and angle 90° (Peitgen and Saupe 1988, p. 278).

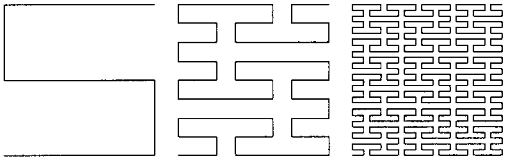

A related curve is the Hilbert II curve, shown above (Peitgen and Saupe 1988, p. 284). It is also a LINDENMAYER SYSTEM and the curve can be encoded with initial string `"X"`, STRING REWRITING rules `"X" -> "XFYFX+F+YFXFY-F-XFYFX"`, `"Y" -> "YFXFY-F-XFYFX+F+YFXFY"`, and angle 90°.

see also LINDENMAYER SYSTEM, PEANO CURVE, PLANE-FILLING CURVE, SIERPIŃSKI CURVE, SPACE-FILLING CURVE

References

Bogomolny, A. "Plane Filling Curves." `http://www.cut-the-knot.com/do_you_know/hilbert.html`.

Dickau, R. M. "Two-Dimensional L-Systems." `http://forum.swarthmore.edu/advanced/robertd/lsys2d.html`.

Dickau, R. M. "Three-Dimensional L-Systems." `http://forum.swarthmore.edu/advanced/robertd/lsys3d.html`.

Goetz, P. "Phil's Good Enough Complexity Dictionary." `http://www.cs.buffalo.edu/~goetz/dict.html`.

Hilbert, D. "Über die stetige Abbildung einer Linie auf ein Flachenstück." *Math. Ann.* **38**, 459–460, 1891.

Peitgen, H.-O. and Saupe, D. (Eds.). *The Science of Fractal Images.* New York: Springer-Verlag, pp. 278 and 284, 1988.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 198–206, 1991.

Hilbert Function

Let $\Gamma=\{p_1,\ldots,p_m\}\subset\mathbb{P}^2$ be a collection of m distinct points. Then the number of conditions imposed by Γ on forms of degree d is called the Hilbert function h_Γ of Γ. If curves X_1 and X_2 of degrees d and e meet in a collection Γ of $d\cdot e$ points, then for any k, the number $h_\Gamma(k)$ of conditions imposed by Γ on forms of degree k is independent of X_1 and X_2 and is given by

$$h_\Gamma(k)=\binom{k+2}{2}-\binom{k-d+2}{2} -\binom{k-e+2}{2}+\binom{k-d-e+2}{2},$$

where the BINOMIAL COEFFICIENT $\binom{a}{2}$ is taken as 0 if $a<2$ (Cayley 1843).

References

Eisenbud, D.; Green, M.; and Harris, J. "Cayley-Bacharach Theorems and Conjectures." *Bull. Amer. Math. Soc.* **33**, 295–324, 1996.

Hilbert Hotel

Let a hotel have a DENUMERABLE set of rooms numbered 1, 2, 3, ... Then any finite number n of guests can be accommodated without evicting the current guests by moving the current guests from room i to room $i+n$. Furthermore, a DENUMERABLE number of guests can be similarly accommodated by moving the existing guests from i to $2i$, freeing up a DENUMERABLE number of rooms $2i-1$.

References

Lauwerier, H. "Hilbert Hotel." In *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, p. 22, 1991.

Pappas, T. "Hotel Infinity." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 37, 1989.

Hilbert's Inequality

Given a POSITIVE SEQUENCE $\{a_n\}$,

$$\sqrt{\sum_{j=-\infty}^{\infty}\left|\sum_{\substack{n=-\infty\\ n\neq j}}^{\infty}\frac{a_n}{j-n}\right|^2}\leq\pi\sqrt{\sum_{n=-\infty}^{\infty}|a_n|^2},$$

where the a_ns are REAL and "square summable."

Another INEQUALITY known as Hilbert's applies to NONNEGATIVE sequences $\{a_n\}$ and $\{b_n\}$,

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\frac{a_m b_m}{m+n} < \pi\csc\left(\frac{\pi}{p}\right)\left(\sum_{m=1}^{\infty}{a_m}^p\right)^{1/p}\left(\sum_{n=1}^{\infty}{b_n}^q\right)^{1/q}$$

unless all a_n or all b_n are 0. If $f(x)$ and $g(x)$ are NONNEGATIVE integrable functions, then the integral form is

$$\int_0^\infty\int_0^\infty\frac{f(x)g(y)}{x+y}\,dx\,dy<\pi\csc\left(\frac{\pi}{p}\right) \times\left(\int_0^\infty[f(x)]^p\,dx\right)^{1/p}\left(\int_0^\infty[g(x)]^q\,dx\right)^{1/q}.$$

The constant $\pi\csc(\pi/P)$ is the best possible, in the sense that counterexamples can be constructed for any smaller value.

References

Hardy, G. H.; Littlewood, J. E.; and Pólya, G. *Inequalities, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 308–309, 1988.

Hilbert Matrix

A MATRIX H with elements

$$H_{ij}\equiv(i+j-1)^{-1}$$

for $i, j = 1, 2, \ldots, n$. Although the MATRIX INVERSE is given analytically by

$$(H^{-1})_{ij}=\frac{(-1)^{i+j}}{i+j-1}\frac{(n+i-1)!(n+j-1)!}{[(i-1)!(j-1)!]^2(n-i)!(n-j)!},$$

Hilbert matrices are difficult to invert numerically. The DETERMINANTS for the first few values of H_n are given in the following table.

n	$\det(\mathsf{H}_n)$
1	1
2	8.33333×10^{-2}
3	4.62963×10^{-4}
4	1.65344×10^{-7}
5	3.74930×10^{-12}
6	5.36730×10^{-18}

Hilbert's Nullstellansatz

Let K be an algebraically closed field and let I be an IDEAL in $K(x)$, where $x=(x_1,x_2,\ldots,x_n)$ is a finite set of indeterminates. Let $p\in K(x)$ be such that for any $(c_1,\ldots,c_n)$ in K^n, if every element of I vanishes when evaluated if we set each $(x_i=c_i)$, then p also vanishes. Then p^j lies in I for some j. Colloquially, the theory of algebraically closed fields is a complete model.

Hilbert Number

see GELFOND-SCHNEIDER CONSTANT

Hilbert Polynomial

Let Γ be an ALGEBRAIC CURVE in a projective space of DIMENSION n, and let p be the PRIME IDEAL defining Γ, and let $\chi(p,m)$ be the number of linearly independent forms of degree m modulo p. For large m, $\chi(p,m)$ is a POLYNOMIAL known as the Hilbert polynomial.

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 36, 1980.

Hilbert's Problems

A set of (originally) unsolved problems in mathematics proposed by Hilbert. Of the 23 total, ten were presented at the Second International Congress in Paris in 1900. These problems were designed to serve as examples for the kinds of problems whose solutions would lead to the furthering of disciplines in mathematics.

1a. Is there a transfinite number between that of a DENUMERABLE SET and the numbers of the CONTINUUM? This question was answered by Gödel and Cohen to the effect that the answer depends on the particular version of SET THEORY assumed.

1b. Can the CONTINUUM of numbers be considered a WELL-ORDERED SET? This question is related to Zermelo's AXIOM OF CHOICE. In 1963, the AXIOM OF CHOICE was demonstrated to be independent of all other AXIOMS in SET THEORY, so there appears to be no universally valid solution to this question either.

2. Can it be proven that the AXIOMS of logic are consistent? GÖDEL'S INCOMPLETENESS THEOREM indicated that the answer is "no," in the sense

that any formal system interesting enough to formulate its own consistency can prove its own consistency IFF it is inconsistent.

3. Give two TETRAHEDRA which cannot be decomposed into congruent TETRAHEDRA directly or by adjoining congruent TETRAHEDRA. Max Dehn showed this could not be done in 1902. W. F. Kagon obtained the same result independently in 1903.
4. Find GEOMETRIES whose AXIOMS are closest to those of EUCLIDEAN GEOMETRY if the ORDERING and INCIDENCE AXIOMS are retained, the CONGRUENCE AXIOMS weakened, and the equivalent of the PARALLEL POSTULATE omitted. This problem was solved by G. Hamel.
5. Can the assumption of differentiability for functions defining a continuous transformation GROUP be avoided? (This is a generalization of the CAUCHY FUNCTIONAL EQUATION.) Solved by John von Neumann in 1930 for bicompact groups. Also solved for the ABELIAN case, and for the solvable case in 1952 with complementary results by Montgomery and Zipin (subsequently combined by Yamabe in 1953). Andrew Glean showed in 1952 that the answer is also "yes" for all locally bicompact groups.
6. Can physics be axiomized?
7. Let $\alpha \neq 1 \neq 0$ be ALGEBRAIC and β IRRATIONAL. Is α^β then TRANSCENDENTAL? Proved true in 1934 by Aleksander Gelfond (GELFOND'S THEOREM; Courant and Robins 1996).
8. Prove the RIEMANN HYPOTHESIS. The CONJECTURE has still been neither proved nor disproved.
9. Construct generalizations of the RECIPROCITY THEOREM of NUMBER THEORY.
10. Does there exist a universal algorithm for solving DIOPHANTINE EQUATIONS? The impossibility of obtaining a general solution was proven by Julia Robinson and Martin Davis in 1970, following proof of the result that the equation $n = F_{2m}$ (where F_{2m} is a FIBONACCI NUMBER) is Diophantine by Yuri Matijasevich (Matijasevič 1970, Davis 1973, Davis and Hersh 1973, Matijasevič 1993).
11. Extend the results obtained for quadratic fields to arbitrary INTEGER algebraic fields.
12. Extend a theorem of Kronecker to arbitrary algebraic fields by explicitly constructing Hilbert class fields using special values. This calls for the construction of HOLOMORPHIC FUNCTIONS in several variables which have properties analogous to the exponential function and elliptic modular functions (Holtzapfel 1995).
13. Show the impossibility of solving the general seventh degree equation by functions of two variables.
14. Show the finiteness of systems of relatively integral functions.
15. Justify Schubert's ENUMERATIVE GEOMETRY (Bell 1945).
16. Develop a topology of REAL algebraic curves and surfaces. The SHIMURA-TANIYAMA CONJECTURE postulates just this connection. See Ilyashenko and Yakovenko (1995) and Gudkov and Utkin (1978).
17. Find a representation of definite form by SQUARES.
18. Build spaces with congruent POLYHEDRA.
19. Analyze the analytic character of solutions to variational problems.
20. Solve general BOUNDARY VALUE PROBLEMS.
21. Solve differential equations given a MONODROMY GROUP. More technically, prove that there always exists a FUCHSIAN SYSTEM with given singularities and a given MONODROMY GROUP. Several special cases had been solved, but a NEGATIVE solution was found in 1989 by B. Bolibruch (Anasov and Bolibruch 1994).
22. Uniformization.
23. Extend the methods of CALCULUS OF VARIATIONS.

References

Anasov, D. V. and Bolibruch, A. A. *The Riemann-Hilbert Problem.* Braunschweig, Germany: Vieweg, 1994.

Bell, E. T. *The Development of Mathematics, 2nd ed.* New York: McGraw-Hill, p. 340, 1945.

Borowski, E. J. and Borwein, J. M. (Eds.). "Hilbert Problems." Appendix 3 in *The Harper Collins Dictionary of Mathematics.* New York: Harper-Collins, p. 659, 1991.

Boyer, C. and Merzbach, U. "The Hilbert Problems." *History of Mathematics, 2nd ed.* New York: Wiley, pp. 610–614, 1991.

Browder, Felix E. (Ed.). *Mathematical Developments Arising from Hilbert Problems.* Providence, RI: Amer. Math. Soc., 1976.

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 107, 1996.

Davis, M. "Hilbert's Tenth Problem is Unsolvable." *Amer. Math. Monthly* **80**, 233–269, 1973.

Davis, M. and Hersh, R. "Hilbert's 10th Problem." *Sci. Amer.*, pp. 84–91, Nov. 1973.

Gudkov, D. and Utkin, G. A. *Nine Papers on Hilbert's 16th Problem.* Providence, RI: Amer. Math. Soc., 1978.

Holtzapfel, R.-P. *The Ball and Some Hilbert Problems.* Boston, MA: Birkhäuser, 1995.

Ilyashenko, Yu. and Yakovenko, S. (Eds.). *Concerning the Hilbert 16th Problem.* Providence, RI: Amer. Math. Soc., 1995.

Matijasevič, Yu. V. "Solution to of the Tenth Problem of Hilbert." *Mat. Lapok* **21**, 83–87, 1970.

Matijasevich, Yu. V. *Hilbert's Tenth Problem.* Cambridge, MA: MIT Press, 1993.

Hilbert-Schmidt Norm

The Hilbert-Schmidt norm of a MATRIX A is defined as

$$|\mathsf{A}|_2 \equiv \sqrt{\sum_{i,j} a_{ij}}.$$

Hilbert-Schmidt Theory

The study of linear integral equations of the Fredholm type with symmetric kernels

$$K(x,t) = K(t,x).$$

References
Arfken, G. "Hilbert-Schmidt Theory." §16.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 890–897, 1985.

Hilbert Space

A Hilbert space is VECTOR SPACE H with an INNER PRODUCT $\langle f,g\rangle$ such that the NORM defined by

$$|f| = \sqrt{\langle f,f\rangle}$$

turns H into a COMPLETE METRIC SPACE. If the INNER PRODUCT does not so define a NORM, it is instead known as an INNER PRODUCT SPACE.

Examples of FINITE-dimensional Hilbert spaces include

1. The REAL NUMBERS $\mathbb{R}^n$ with $\langle v,u\rangle$ the vector DOT PRODUCT of v and u.
2. The COMPLEX NUMBERS $\mathbb{C}^n$ with $\langle v,u\rangle$ the vector DOT PRODUCT of v and the COMPLEX CONJUGATE of u.

An example of an INFINITE-dimensional Hilbert space is L^2, the SET of all FUNCTIONS $f:\mathbb{R}\to\mathbb{R}$ such that the INTEGRAL of f^2 over the whole REAL LINE is FINITE. In this case, the INNER PRODUCT is

$$\langle f,g\rangle = \int f(x)g(x)\,dx.$$

A Hilbert space is always a BANACH SPACE, but the converse need not hold.

see also BANACH SPACE, L_2-NORM, L_2-SPACE, LIOUVILLE SPACE, PARALLELOGRAM LAW, VECTOR SPACE

References
Sansone, G. "Elementary Notions of Hilbert Space." §1.3 in *Orthogonal Functions, rev. English ed.* New York: Dover, pp. 5–10, 1991.
Stone, M. H. *Linear Transformations in Hilbert Space and Their Applications Analysis.* Providence, RI: Amer. Math. Soc., 1932.

Hilbert's Theorem

Every MODULAR SYSTEM has a MODULAR SYSTEM BASIS consisting of a finite number of POLYNOMIALS. Stated another way, for every order n there exists a nonsingular curve with the maximum number of circuits and the maximum number for any one nest.

References
Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 61, 1959.

Hilbert Transform

$$g(y) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{f(x)\,dx}{x-y}$$
$$f(x) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{g(y)\,dy}{y-x}.$$

see also TITCHMARSH THEOREM

References
Bracewell, R. *The Fourier Transform and Its Applications.* New York: McGraw-Hill, pp. 267–272, 1965.

Hill Determinant

A DETERMINANT which arises in the solution of the second-order ORDINARY DIFFERENTIAL EQUATION

$$x^2\frac{d^2\psi}{dx^2} + x\frac{d\psi}{dx} + \left(\tfrac{1}{4}h^2x^2 + \tfrac{1}{2}h^2 - b + \frac{h^2}{4x^2}\right)\psi = 0. \quad (1)$$

Writing the solution as a POWER SERIES

$$\psi = \sum_{n=-\infty}^{\infty} a_n x^{s+2n} \quad (2)$$

gives a RECURRENCE RELATION

$$h^2a_{n+1} + [2h^2 - 4b + 16(n+\tfrac{1}{2}s)^2]a_n + h^2a_{n-1} = 0. \quad (3)$$

The value of s can be computed using the Hill determinant

$$\Delta(s) = \begin{vmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & ⋰ \\ \cdots & \frac{(\sigma+2)-\alpha^2}{4-\alpha^2} & \frac{\beta^2}{4-\alpha^2} & 0 & 0 & \cdots \\ \cdots & 0 & -\frac{\beta^2}{\alpha^2} & -\frac{\sigma^2-\alpha^2}{\alpha^2} & -\frac{\beta^2}{\alpha^2} & \cdots \\ \cdots & 0 & 0 & -\frac{\beta^2}{1-\alpha^2} & \frac{(\sigma-1)^2-\alpha^2}{1-\alpha^2} & \cdots \\ ⋰ & \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}, \quad (4)$$

where

$$\sigma = \tfrac{1}{2}s \quad (5)$$
$$\alpha^2 = \tfrac{1}{4}b - \tfrac{1}{8}h^2 \quad (6)$$
$$\beta = \tfrac{1}{4}h, \quad (7)$$

and σ is the variable to solve for. The determinant can be given explicitly by the amazing formula

$$\Delta(s) = \Delta(0) - \frac{\sin^2(\pi s/2)}{\sin^2(\frac{1}{2}\pi\sqrt{b - \frac{1}{2}h^2})}, \qquad (8)$$

where

$$\Delta(0) = \begin{vmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \iddots \\ \cdots & 1 & \frac{h^2}{144+2h^2-4b} & 0 & 0 & \cdots \\ \cdots & \frac{h^2}{64+2h^2-4b} & 1 & \frac{h^2}{64+2h^2-4b} & 0 & \cdots \\ \cdots & 0 & \frac{h^2}{16+2h^2-4b} & 1 & \frac{h^2}{16+2h^2-4b} & \cdots \\ \cdots & 0 & 0 & \frac{h^2}{2h^2-4b} & 1 & \cdots \\ \cdots & 0 & 0 & 0 & \frac{h^2}{16+2h^2-4b} & \cdots \\ \iddots & \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}, \qquad (9)$$

leading to the implicit equation for s,

$$\sin^2(\tfrac{1}{2}\pi s) = \Delta(0)\sin^2\left(\tfrac{1}{2}\pi\sqrt{b - \tfrac{1}{2}h^2}\right). \qquad (10)$$

see also HILL'S DIFFERENTIAL EQUATION

References
Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 555–562, 1953.

Hill's Differential Equation

$$\frac{d^2x}{dt^2} = \phi(t)x,$$

where ϕ is periodic. It can be written as

$$\frac{d^2y}{dx^2} + \left[\theta_0 + 2\sum_{n=1}^{\infty}\theta_n\cos(2nz)\right] = 0,$$

where θ_n are known constants. A solution can be given by taking the "DETERMINANT" of an infinite MATRIX.

see also HILL DETERMINANT

Hillam's Theorem

If $f : [a,b] \to [a,b]$ (where $[a,b]$ denotes the CLOSED INTERVAL from a to b on the REAL LINE) satisfies a LIPSCHITZ CONDITION with constant K, i.e., if

$$|f(x) - f(y)| \leq K|x - y|$$

for all $x, y \in [a,b]$, then the iteration scheme

$$x_{n+1} = (1-\lambda)x_n + \lambda f(x_n),$$

where $\lambda = 1/(K+1)$, converges to a FIXED POINT of f.

References
Falkowski, B.-J. "On the Convergence of Hillam's Iteration Scheme." *Math. Mag.* **69**, 299–303, 1996.
Geist, R.; Reynolds, R.; and Suggs, D. "A Markovian Framework for Digital Halftoning." *ACM Trans. Graphics* **12**, 136–159, 1993.
Hillam, B. P. "A Generalization of Krasnoselski's Theorem on the Real Line." *Math. Mag.* **48**, 167–168, 1975.
Krasnoselski, M. A. "Two Remarks on the Method of Successive Approximations." *Uspehi Math. Nauk (N. S.)* **10**, 123–127, 1955.

Hindu Check

see CASTING OUT NINES

Hinge

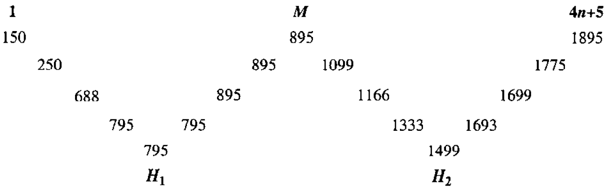

The upper and lower hinges are descriptive statistics of a set of N data values, where N is of the form $N = 4n+5$ with $n = 0, 1, 2, \ldots$. The hinges are obtained by ordering the data in increasing order $a_1, \ldots, a_N$, and writing them out in the shape of a "w" as illustrated above. The values at the bottom legs are called the hinges H_1 and H_2 (and the central peak is the MEDIAN). In this ordering,

$$\begin{aligned} H_1 &= a_{n+2} = a_{(N+3)/4} \\ M &= a_{2n+3} = a_{(N+1)/2} \\ H_2 &= a_{3n+4} = a_{(3N+1)/4}. \end{aligned}$$

For N of the form $4n+5$, the hinges are identical to the QUARTILES. The difference $H_2 - H_1$ is called the H-SPREAD.

see also H-SPREAD, HABERDASHER'S PROBLEM, MEDIAN (STATISTICS), ORDER STATISTIC, QUARTILE, TRIMEAN

References
Tukey, J. W. *Explanatory Data Analysis.* Reading, MA: Addison-Wesley, pp. 32–34, 1977.

Hippias' Quadratrix

see QUADRATRIX OF HIPPIAS

Hippopede

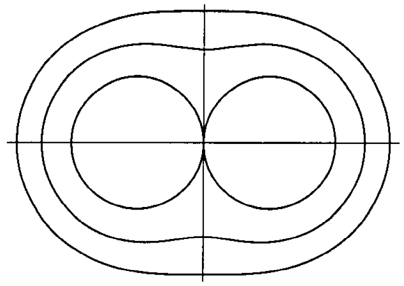

A curve also known as a HORSE FETTER and given by the polar equation

$$r^2 = 4b(a - b\sin^2\theta).$$

References
Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 144–146, 1972.

Histogram

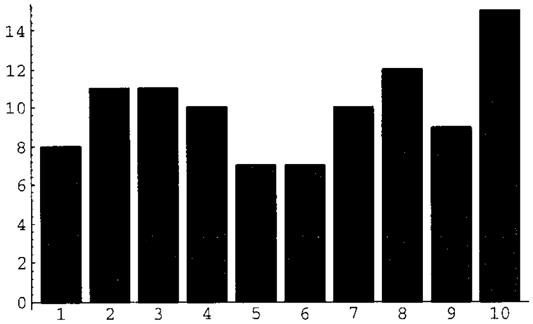

The grouping of data into bins (spaced apart by the so-called CLASS INTERVAL) plotting the number of members in each bin versus the bin number. The above histogram shows the number of variates in bins with CLASS INTERVAL 1 for a sample of 100 real variates with a UNIFORM DISTRIBUTION from 0 and 10. Therefore, bin 1 gives the number of variates in the range 0–1, bin 2 gives the number of variates in the range 1–2, etc.

see also OGIVE

Hitch

A KNOT that secures a rope to a post, ring, another rope, etc., but does not keep its shape by itself.

see also CLOVE HITCH, KNOT, LINK, LOOP (KNOT)

References
Owen, P. *Knots.* Philadelphia, PA: Courage, p. 17, 1993.

Hitting Set

Let S be a collection S of subsets of a finite set X. The smallest subset Y of X that meets every member of S is called the hitting set or VERTEX COVER. Finding the hitting set is an NP-COMPLETE PROBLEM.

Hjelmslev's Theorem

When all the points P on one line are related by an ISOMETRY to all points P' on another, the MIDPOINTS of the segments PP' are either distinct and collinear or coincident.

HJLS Algorithm

An algorithm for finding INTEGER RELATIONS whose running time is bounded by a polynomial in the number of real variables. It is much faster than other algorithms such as the FERGUSON-FORCADE ALGORITHM, LLL ALGORITHM, and PSOS ALGORITHM.

Unfortunately, it is numerically unstable and therefore requires extremely high precision. The cause of this instability is not known (Ferguson and Bailey 1992), but is believed to derive from its reliance on GRAM-SCHMIDT ORTHONORMALIZATION, which is know to be numerically unstable (Golub and van Loan 1989).

see also FERGUSON-FORCADE ALGORITHM, INTEGER RELATION, LLL ALGORITHM, PSLQ ALGORITHM, PSOS ALGORITHM

References
Ferguson, H. R. P. and Bailey, D. H. "A Polynomial Time, Numerically Stable Integer Relation Algorithm." RNR Techn. Rept. RNR-91-032, Jul. 14, 1992.
Golub, G. H. and van Loan, C. F. *Matrix Computations, 3rd ed.* Baltimore, MD: Johns Hopkins, 1996.
Hastad, J.; Just, B.; Lagarias, J. C.; and Schnorr, C. P. "Polynomial Time Algorithms for Finding Integer Relations Among Real Numbers." *SIAM J. Comput.* **18**, 859–881, 1988.

HK Integral

Named after Henstock and Kurzweil. Every LEBESGUE INTEGRABLE function is HK integrable with the same value.

References
Shenitzer, A. and Steprans, J. "The Evolution of Integration." *Amer. Math. Monthly* **101**, 66–72, 1994.

Hodge Star

On an oriented n-D RIEMANNIAN MANIFOLD, the Hodge star is a linear FUNCTION which converts alternating DIFFERENTIAL k-FORMS to alternating $(n-k)$-forms. If w is an alternating k-FORM, its Hodge star is given by

$$w(v_1, \ldots, v_k) = ({}^*w)(v_{k+1}, \ldots, v_n)$$

when $v_1, \ldots, v_n$ is an oriented orthonormal basis.

see also STOKES' THEOREM

Hodge's Theorem

On a compact oriented FINSLER MANIFOLD without boundary, every COHOMOLOGY class has a UNIQUE harmonic representative. The DIMENSION of the SPACE of all harmonic forms of degree p is the pth BETTI NUMBER of the MANIFOLD.

see also BETTI NUMBER, COHOMOLOGY, DIMENSION, FINSLER MANIFOLD

References
Chern, S.-S. "Finsler Geometry is Just Riemannian Geometry without the Quadratic Restriction." *Not. Amer. Math. Soc.* **43**, 959–963, 1996.

Hoehn's Theorem

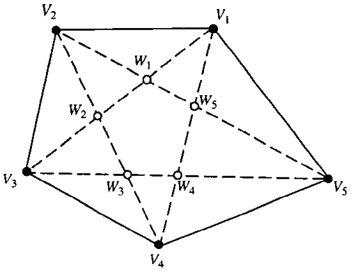

A geometric theorem related to the PENTAGRAM and also called the PRATT-KASAPI THEOREM.

$$\frac{|V_1W_1|}{|W_2V_3|}\frac{|V_2W_2|}{|W_3V_4|}\frac{|V_3W_3|}{|W_4V_5|}\frac{|V_4W_4|}{|W_5V_1|}\frac{|V_5W_5|}{|W_1V_2|}=1$$

$$\frac{|V_1W_2|}{|W_1V_3|}\frac{|V_2W_3|}{|W_2V_4|}\frac{|V_3W_4|}{|W_3V_5|}\frac{|V_4W_5|}{|W_4V_1|}\frac{|V_5W_1|}{|W_5V_2|}=1.$$

In general, it is also true that

$$\frac{|V_iW_i|}{|W_{i+1}V_{i+2}|}=\frac{|V_iV_{i+1}V_{i+4}|}{|V_iV_{i+1}V_{i+2}V_{i+4}|}\frac{|V_iV_{i+1}V_{i+2}V_{i+3}|}{|V_{i+2}V_{i+3}V_{i+1}|}.$$

This type of identity was generalized to other figures in the plane and their duals by Pinkernell (1996).

References
Chou, S. C. *Mechanical Geometry Theorem Proving.* Dordrecht, Netherlands: Reidel, 1987.
Grünbaum, B. and Shepard, G. C. "Ceva, Menelaus, and the Area Principle." *Math. Mag.* **68**, 254–268, 1995.
Hoehn, L. "A Menelaus-Type Theorem for the Pentagram." *Math. Mag.* **68**, 254–268, 1995.
Pinkernell, G. M. "Identities on Point-Line Figures in the Euclidean Plane." *Math. Mag.* **69**, 377–383, 1996.

Hoffman's Minimal Surface

A minimal embedded surface discovered in 1992 consisting of a helicoid with a HOLE and HANDLE (*Science News* 1992). It has the same topology as a punctured sphere with a handle, and is only the second complete embedded minimal surface of finite topology and infinite total curvature discovered (the HELICOID being the first).

A three-ended minimal surface GENUS 1 is sometimes also called Hoffman's minimal surface (Peterson 1988).

see also HELICOID

References
Peterson, I. *Mathematical Tourist: Snapshots of Modern Mathematics.* New York: W. H. Freeman, pp. 57–59, 1988.
"Putting a Handle on a Minimal Helicoid." *Sci. News* **142**, 276, Oct. 24, 1992.

Hoffman-Singleton Graph

The only GRAPH of DIAMETER 2, GIRTH 5, and VALENCY 7. It contains many copies of the PETERSEN GRAPH.

References
Hoffman, A. J. and Singleton, R. R. "On Moore Graphs of Diameter Two and Three." *IBM J. Res. Develop.* **4**, 497–504, 1960.

Hofstadter-Conway $10,000 Sequence

The INTEGER SEQUENCE defined by the RECURRENCE RELATION

$$a(n)=a(a(n-1))+a(n-a(n-1)),$$

with $a(1)=a(2)=1$. The first few values are 1, 1, 2, 2, 3, 4, 4, 4, 5, 6, ... (Sloane's A004001). Plotting $a(n)/n$ against n gives the BATRACHION plotted below. Conway (1988) showed that $\lim_{n\to\infty} a(n)/n = 1/2$ and offered a prize of $10,000 to the discoverer of a value of n for which $|a(i)/i - 1/2| < 1/20$ for $i > n$. The prize was subsequently claimed by Mallows, after adjustment to Conway's "intended" prize of $1,000 (Schroeder 1991), who found $n = 1489$.

$a(n)/n$ takes a value of 1/2 for n of the form 2^k with $k = 1, 2, \ldots$. Pickover (1996) gives a table of analogous values of n corresponding to different values of $|a(n)/n - 1/2| < e$.

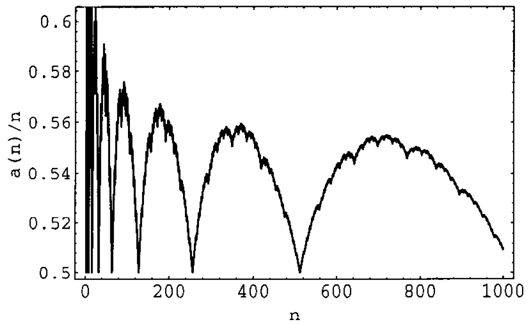

see also BLANCMANGE FUNCTION, HOFSTADTER'S Q-SEQUENCE, MALLOW'S SEQUENCE

References
Conolly, B. W. "Meta-Fibonacci Sequences." In *Fibonacci and Lucas Numbers, and the Golden Section* (Ed. S. Vajda). New York: Halstead Press, pp. 127–138, 1989.
Conway, J. "Some Crazy Sequences." Lecture at AT&T Bell Labs, July 15, 1988.
Guy, R. K. "Three Sequences of Hofstadter." §E31 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 231–232, 1994.
Kubo, T. and Vakil, R. "On Conway's Recursive Sequence." *Disc. Math.* **152**, 225–252, 1996.
Mallows, C. "Conway's Challenging Sequence." *Amer. Math. Monthly* **98**, 5–20, 1991.
Pickover, C. A. "The Drums of Ulupu." In *Mazes for the Mind: Computers and the Unexpected.* New York: St. Martin's Press, 1993.
Pickover, C. A. "The Crying of Fractal Batrachion 1,489." Ch. 25 in *Keys to Infinity.* New York: W. H. Freeman, pp. 183–191, 1995.
Schroeder, M. "John Horton Conway's 'Death Bet.'" *Fractals, Chaos, Power Laws.* New York: W. H. Freeman, pp. 57–59, 1991.
Sloane, N. J. A. Sequence A004001/M0276 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Hofstadter Figure-Figure Sequence

Define $F(1) = 1$ and $S(1) = 2$ and write

$$F(n) = F(n-1) + S(n-1),$$

where the sequence $\{S(n)\}$ consists of those integers not already contained in $\{F(n)\}$. For example, $F(2) = F(1) + S(1) = 3$, so the next term of $S(n)$ is $S(2) = 4$, giving $F(3) = F(2) + S(2) = 7$. The next integer is 5, so $S(3) = 5$ and $F(4) = F(3) + S(3) = 12$. Continuing in this manner gives the "figure" sequence $F(n)$ as 1, 3, 7, 12, 18, 26, 35, 45, 56, ... (Sloane's A005228) and the "space" sequence as 2, 4, 5, 6, 8, 9, 10, 11, 13, 14, ... (Sloane's A030124).

References

Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, p. 73, 1989.

Sloane, N. J. A. Sequences A030124 and A005288/M2629 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Hofstadter G-Sequence

The sequence defined by $G(0) = 0$ and

$$G(n) = n - G(G(n-1)).$$

The first few terms are 1, 1, 2, 3, 3, 4, 4, 5, 6, 6, 7, 8, 8, 9, 9, ... (Sloane's A005206).

References

Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, p. 137, 1989.

Sloane, N. J. A. Sequence A005206/M0436 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Hofstadter H-Sequence

The sequence defined by $H(0) = 0$ and

$$H(n) = n - H(H(H(n-1))).$$

The first few terms are 1, 1, 2, 3, 4, 4, 5, 5, 6, 7, 7, 8, 9, 10, 10, 11, 12, 13, 13, 14, ... (Sloane's A005374).

References

Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, p. 137, 1989.

Sloane, N. J. A. Sequence A005374/M0449 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Hofstadter Male-Female Sequences

The pair of sequences defined by $F(0) = 1$, $M(0) = 0$, and

$$F(n) = n - M(F(n-1))$$
$$M(n) = n - F(M(n-1)).$$

The first few terms of the "male" sequence $M(n)$ are 0, 1, 2, 2, 3, 4, 4, 5, 6, 6, 7, 7, 8, 9, 9, ... (Sloane's A005379), and the first few terms of the "female" sequence $F(n)$ are 1, 2, 2, 3, 3, 4, 5, 5, 6, 6, 7, 8, 8, 9, 9, ... (Sloane's A005378).

References

Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, p. 137, 1989.

Sloane, N. J. A. Sequences A005378/M0263 and A005379/M0278 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Hofstadter Point

The r-HOFSTADTER TRIANGLE of a given TRIANGLE ΔABC is perspective to ΔABC, and the PERSPECTIVE CENTER is called the Hofstadter point. The TRIANGLE CENTER FUNCTION is

$$\alpha = \frac{\sin(rA)}{\sin(r - rA)}.$$

As $r \to 0$, the TRIANGLE CENTER FUNCTION approaches

$$\alpha = \frac{A}{a},$$

and as $r \to 1$, the TRIANGLE CENTER FUNCTION approaches

$$\alpha = \frac{a}{A}.$$

see also HOFSTADTER TRIANGLE

References

Kimberling, C. "Hofstadter Points." *Nieuw Arch. Wiskunder* **12**, 109–114, 1994.

Kimberling, C. "Major Centers of Triangles." *Amer. Math. Monthly* **104**, 431–438, 1997.

Kimberling, C. "Hofstadter Points." `http://www.evansville.edu/~ck6/tcenters/recent/hofstad.html`.

Hofstadter's Q-Sequence

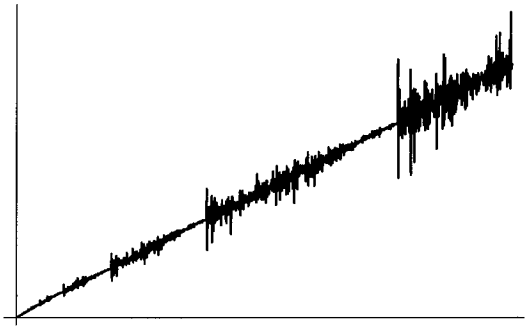

The INTEGER SEQUENCE given by

$$Q(n) = Q(n - Q(n-1)) + Q(n - Q(n-2)),$$

with $Q(1) = Q(2) = 1$. The first few values are 1, 1, 2, 3, 3, 4, 5, 5, 6, 6, ... (Sloane's A005185; illustrated above). These numbers are sometimes called Q-NUMBERS.

see also HOFSTADTER-CONWAY \$10,000 SEQUENCE, MALLOW'S SEQUENCE

References

Conolly, B. W. "Meta-Fibonacci Sequences." In *Fibonacci and Lucas Numbers, and the Golden Section* (Ed. S. Vajda). New York: Halstead Press, pp. 127–138, 1989.

Guy, R. "Some Suspiciously Simple Sequences." *Amer. Math. Monthly* **93**, 186–191, 1986.

Hofstadter, D. R. *Gödel, Escher Bach: An Eternal Golden Braid.* New York: Vintage Books, pp. 137–138, 1980.

Pickover, C. A. "The Crying of Fractal Batrachion 1,489." Ch. 25 in *Keys to Infinity.* New York: W. H. Freeman, pp. 183–191, 1995.

Sloane, N. J. A. Sequence A005185/M0438 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Hofstadter Sequences

Let $b_1 = 1$ and $b_2 = 2$ and for $n \geq 3$, let b_n be the least INTEGER $> b_{n-1}$ which can be expressed as the SUM of two or more consecutive terms. The resulting sequence is 1, 2, 3, 5, 6, 8, 10, 11, 14, 16, ... (Sloane's A005243). Let $c_1 = 2$ and $c_2 = 3$, form all possible expressions of the form $c_i c_j - 1$ for $1 \leq i \leq j \leq n$, and append them. The resulting sequence is 2, 3, 5, 9, 14, 16, 17, 18, ... (Sloane's A05244).

see also HOFSTADTER-CONWAY $10,000 SEQUENCE, HOFSTADTER'S Q-SEQUENCE

References

Guy, R. K. "Three Sequences of Hofstadter." §E31 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 231–232, 1994.

Sloane, N. J. A. Sequences A005243/M0623 and A00524/M0705 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Hofstadter Triangle

For a NONZERO REAL NUMBER r and a TRIANGLE ΔABC, swing LINE SEGMENT BC about the vertex B towards vertex A through an ANGLE rB. Call the line along the rotated segment L. Construct a second line L' by rotating LINE SEGMENT BC about vertex C through an ANGLE rC. Now denote the point of intersection of L and L' by $A(r)$. Similarly, construct $B(r)$ and $C(r)$. The TRIANGLE having these points as vertices is called the Hofstadter r-triangle. Kimberling (1994) showed that the Hofstadter triangle is perspective to ΔABC, and calls PERSPECTIVE CENTER the HOFSTADTER POINT.

see also HOFSTADTER POINT

References

Kimberling, C. "Hofstadter Points." *Nieuw Arch. Wiskunde* **12**, 109–114, 1994.

Kimberling, C. "Hofstadter Points." `http://www.evansville.edu/~ck6/tcenters/recent/hofstad.html`.

Hölder Condition

A function $\phi(t)$ satisfies the Hölder condition on two points t_1 and t_2 on an arc L when

$$|\phi(t_2) - \phi(t_1)| \leq A|t_2 - t_1|^{\mu},$$

with A and μ POSITIVE REAL constants.

Hölder Integral Inequality

If

$$\frac{1}{p} + \frac{1}{q} = 1$$

with p, $q > 1$, then

$$\int_a^b |f(x)g(x)|\,dx \leq \left[\int_a^b |f(x)|^p\,dx\right]^{1/p} \left[\int_a^b |g(x)|^q\,dx\right]^{1/q},$$

with equality when

$$|g(x)| = c|f(x)|^{p-1}.$$

If $p = q = 2$, this inequality becomes SCHWARZ'S INEQUALITY.

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 11, 1972.

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1099, 1993.

Hölder, O. "Über einen Mittelwertsatz." *Göttingen Nachr.*, 44, 1889.

Riesz, F. "Untersuchungen über Systeme integrierbarer Funktionen." *Math. Ann.* **69**, 456, 1910.

Riesz, F. "Su alcune disuguaglianze." *Boll. Un. Mat. It.* **7**, 77–79, 1928.

Sansone, G. *Orthogonal Functions, rev. English ed.* New York: Dover, pp. 32–33, 1991.

Hölder Sum Inequality

If

$$\frac{1}{p} + \frac{1}{q} = 1$$

with p, $q > 1$, then

$$\sum_{k=1}^{n} |a_k b_k| \leq \left(\sum_{k=1}^{n} |a_k|^p\right)^{1/p} \left(\sum_{k=1}^{n} |b_k|^q\right)^{1/q},$$

with equality when $|b_k| = c|a_k|^{p-1}$. If $p = q = 2$, this becomes the CAUCHY INEQUALITY.

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 11, 1972.

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1092, 1979.

Hardy, G. H.; Littlewood, J. E.; and Pólya, G. *Inequalities, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 10–15, 1988.

Hole

A hole in a mathematical object is a TOPOLOGICAL structure which prevents the object from being continuously shrunk to a point. When dealing with TOPOLOGICAL SPACES, a DISCONNECTIVITY is interpreted as a hole in the space. Examples of holes are things like the hole in the "center" of a SPHERE or a CIRCLE and the hole produced in EUCLIDEAN SPACE cutting a KNOT out from it.

Singular HOMOLOGY GROUPS form a MEASURE of the hole structure of a SPACE, but they are one particular measure and they don't always pick up everything. HOMOTOPY GROUPS of a SPACE are another measure of holes in a SPACE, as well as BORDISM GROUPS, k-THEORY, COHOMOTOPY GROUPS, and so on.

There are many ways to measure holes in a space. Some holes are picked up by HOMOTOPY GROUPS that are not picked up by HOMOLOGY GROUPS, and some holes are picked up by HOMOLOGY GROUPS that are not picked up by HOMOTOPY GROUPS. (For example, in the TORUS, HOMOTOPY GROUPS "miss" the two-dimensional hole that is given by the TORUS itself, but the second HOMOLOGY GROUP picks that hole up.) In addition, HOMOLOGY GROUPS don't pick up the varying hole structures of the complement of KNOTS in 3-space, but the first HOMOTOPY GROUP (the fundamental group) does.

see also BRANCH CUT, BRANCH POINT, CORK PLUG, CROSS-CAP, GENUS (SURFACE), SINGULAR POINT (FUNCTION), SPHERICAL RING, TORUS

Holomorphic Function

A synonym for ANALYTIC FUNCTION.

see also ANALYTIC FUNCTION, HOMEOMORPHIC

Holonomic Constant

A limiting value of a HOLONOMIC FUNCTION near a SINGULAR POINT. Holonomic constants include APÉRY'S CONSTANT, CATALAN'S CONSTANT, PÓLYA'S RANDOM WALK CONSTANTS for $d > 2$, and PI.

Holonomic Function

A solution of a linear homogeneous ORDINARY DIFFERENTIAL EQUATION with POLYNOMIAL COEFFICIENTS.

see also HOLONOMIC CONSTANT

References

Zeilberger, D. "A Holonomic Systems Approach to Special Function Identities." *J. Comput. Appl. Math.* **32**, 321–348, 1990.

Holonomy

A general concept in CATEGORY THEORY involving the globalization of topological or differential structures.

see also MONODROMY

Home Plate

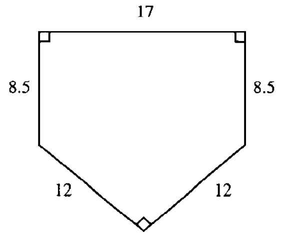

Home plate in the game of BASEBALL is an irregular PENTAGON. However, the Little League rulebook's specification of the shape of home plate (Kreutzer and Kerley 1990), illustrated above, is not physically realizable, since it requires the existence of a (12, 12, 17) RIGHT TRIANGLE, whereas

$$12^2 + 12^2 = 288 \neq 289 = 17^2$$

(Bradley 1996).

see also BASEBALL COVER

References

Bradley, M. J. "Building Home Plate: Field of Dreams or Reality?" *Math. Mag.* **69**, 44–45, 1996.
Kreutzer, P. and Kerley, T. *Little League's Official How-to-Play Baseball Book.* New York: Doubleday, 1990.

Homeoid

A shell bounded by two similar ELLIPSOIDS having a constant ratio of axes. Given a CHORD passing through a homeoid, the distance between inner and outer intersections is equal on both sides. Since a spherical shell is a symmetric case of a homeoid, this theorem is also true for spherical shells (CONCENTRIC CIRCLES in the PLANE), for which it is easily proved by symmetry arguments.

see also CHORD, ELLIPSOID

Homeomorphic

There are two possible definitions:

1. Possessing similarity of form,
2. Continuous, ONE-TO-ONE, ONTO, and having a continuous inverse.

The most common meaning is possessing intrinsic topological equivalence. Two objects are homeomorphic if they can be deformed into each other by a continuous, invertible mapping. Homeomorphism ignores the space in which surfaces are embedded, so the deformation can be completed in a higher dimensional space than the surface was originally embedded. MIRROR IMAGES are homeomorphic, as are MÖBIUS BANDS with an EVEN number of half twists, and MÖBIUS BANDS with an ODD number of twists.

In CATEGORY THEORY terms, homeomorphisms are ISOMORPHISMS in the CATEGORY of TOPOLOGICAL SPACES and continuous maps.

see also HOMOMORPHIC, POLISH SPACE

Homeomorphic Group

If the ELEMENTS of two GROUPS are n to 1 and the correspondences satisfy the same GROUP multiplication table, the GROUPS are said to be homeomorphic.

see also ISOMORPHIC GROUPS

Homeomorphic Type

The following three pieces of information completely determine the homeomorphic type of the surface (Massey 1967):

1. Orientability,
2. Number of boundary components,
3. EULER CHARACTERISTIC.

see also ALGEBRAIC TOPOLOGY, EULER CHARACTERISTIC

References

Massey, W. S. *Algebraic Topology: An Introduction.* New York: Springer-Verlag, 1996.

Homeomorphism

see HOMEOMORPHIC, HOMEOMORPHIC GROUP, HOMEOMORPHIC TYPE, TOPOLOGICALLY CONJUGATE

HOMFLY Polynomial

A 2-variable oriented KNOT POLYNOMIAL $P_L(a,z)$ motivated by the JONES POLYNOMIAL (Freyd *et al.* 1985). Its name is an acronym for the last names of its co-discoverers: Hoste, Ocneanu, Millett, Freyd, Lickorish, and Yetter (Freyd *et al.* 1985). Independent work related to the HOMFLY polynomial was also carried out by Prztycki and Traczyk (1987). HOMFLY polynomial is defined by the SKEIN RELATIONSHIP

$$a^{-1}P_{L_+}(a,z) - aP_{L_-}(a,z) = zP_{L_0}(a,z) \qquad (1)$$

(Doll and Hoste 1991), where v is sometimes written instead of a (Kanenobu and Sumi 1993) or, with a slightly different relationship, as

$$\alpha P_{L_+}(\alpha,z) - \alpha^{-1}P_{L_-}(\alpha,z) = zP_{L_0}(\alpha,z) \qquad (2)$$

(Kauffman 1991). It is also defined as $P_L(\ell,m)$ in terms of SKEIN RELATIONSHIP

$$\ell P_{L_+} + \ell^{-1}P_{L_-} + mP_{L_0} = 0 \qquad (3)$$

(Lickorish and Millett 1988). It can be regarded as a nonhomogeneous POLYNOMIAL in two variables or a homogeneous POLYNOMIAL in three variables. In three variables the SKEIN RELATIONSHIP is written

$$xP_{L_+}(x,y,z) + yP_{L_-}(x,y,z) + zP_{L_0}(x,y,z) = 0. \qquad (4)$$

It is normalized so that $P_{\text{unknot}} = 1$. Also, for n unlinked unknotted components,

$$P_L(x,y,z) = \left(-\frac{x+y}{z}\right)^{n-1}. \qquad (5)$$

This POLYNOMIAL usually detects CHIRALITY but does not detect the distinct ENANTIOMERS of the KNOTS 09_{042}, 10_{048}, 10_{071}, 10_{091}, 10_{104}, and 10_{125} (Jones 1987). The HOMFLY polynomial of an oriented KNOT is the same if the orientation is reversed. It is a generalization of the JONES POLYNOMIAL $V(t)$, satisfying

$$V(t) = P(a = t, z = t^{1/2} - t^{-1/2}) \qquad (6)$$

$$V(t) = P(\ell = it^{-1}, m = i(t^{-1/2} - t^{1/2})). \qquad (7)$$

It is also a generalization of the ALEXANDER POLYNOMIAL $\nabla(z)$, satisfying

$$\Delta(z) = P(a = 1, z = t^{1/2} - t^{-1/2}). \qquad (8)$$

The HOMFLY POLYNOMIAL of the MIRROR IMAGE K^* of a KNOT K is given by

$$P_{K^*}(\ell,m) = P_K(\ell^{-1},m), \qquad (9)$$

so P usually but not always detects CHIRALITY.

A split union of two links (i.e., bringing two links together without intertwining them) has HOMFLY polynomial

$$P(L_1 \cup L_2) = -(\ell + \ell^{-1})m^{-1}P(L_1)P(L_2). \qquad (10)$$

Also, the composition of two links

$$P(L_1 \# L_2) = P(L_1)P(L_2), \qquad (11)$$

so the POLYNOMIAL of a COMPOSITE KNOT factors into POLYNOMIALS of its constituent knots (Adams 1994).

MUTANTS have the same HOMFLY polynomials. In fact, there are infinitely many distinct KNOTS with the same HOMFLY POLYNOMIAL (Kanenobu 1986). Examples include (05_{001}, 10_{132}), (08_{008}, 10_{129}) (08_{016}, 10_{156}), and (10_{025}, 10_{056}) (Jones 1987). Incidentally, these also have the same JONES POLYNOMIAL.

M. B. Thistlethwaite has tabulated the HOMFLY polynomial for KNOTS up to 13 crossings.

see also ALEXANDER POLYNOMIAL, JONES POLYNOMIAL, KNOT POLYNOMIAL

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 171–172, 1994.

Doll, H. and Hoste, J. "A Tabulation of Oriented Links." *Math. Comput.* **57**, 747–761, 1991.

Freyd, P.; Yetter, D.; Hoste, J.; Lickorish, W. B. R.; Millett, K.; and Oceanu, A. "A New Polynomial Invariant of Knots and Links." *Bull. Amer. Math. Soc.* **12**, 239–246, 1985.

Jones, V. "Hecke Algebra Representations of Braid Groups and Link Polynomials." *Ann. Math.* **126**, 335–388, 1987.
Kanenobu, T. "Infinitely Many Knots with the Same Polynomial." *Proc. Amer. Math. Soc.* **97**, 158–161, 1986.
Kanenobu, T. and Sumi, T. "Polynomial Invariants of 2-Bridge Knots through 22 Crossings." *Math. Comput.* **60**, 771–778 and S17–S28, 1993.
Kauffman, L. H. *Knots and Physics.* Singapore: World Scientific, p. 52, 1991.
Lickorish, W. B. R. and Millett, B. R. "The New Polynomial Invariants of Knots and Links." *Math. Mag.* **61**, 1–23, 1988.
Morton, H. R. and Short, H. B. "Calculating the 2-Variable Polynomial for Knots Presented as Closed Braids." *J. Algorithms* **11**, 117–131, 1990.
Przytycki, J. and Traczyk, P. "Conway Algebras and Skein Equivalence of Links." *Proc. Amer. Math. Soc.* **100**, 744–748, 1987.
Stoimenow, A. "Jones Polynomials." `http://www.informatik.hu-berlin.de/~stoimeno/ptab/j10.html`.
Weisstein, E. W. "Knots and Links." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Knots.m`.

Homoclinic Point

A point where a stable and an unstable separatrix (invariant manifold) from the same fixed point or same family intersect. Therefore, the limits

$$\lim_{k\to\infty} f^k(X)$$

and

$$\lim_{k\to -\infty} f^k(X)$$

exist and are equal.

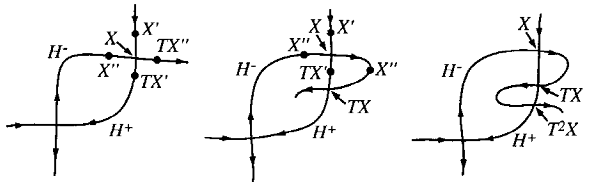

Refer to the above figure. Let X be the point of intersection, with X_1 ahead of X on one MANIFOLD and X_2 ahead of X of the other. The mapping of each of these points TX_1 and TX_2 must be ahead of the mapping of X, TX. The only way this can happen is if the MANIFOLD loops back and crosses itself at a new homoclinic point. Another loop must be formed, with T^2X another homoclinic point. Since T^2X is closer to the hyperbolic point than TX, the distance between T^2X and TX is less than that between X and TX. Area preservation requires the AREA to remain the same, so each new curve (which is closer than the previous one) must extend further. In effect, the loops become longer and thinner. The network of curves leading to a dense AREA of homoclinic points is known as a homoclinic tangle or tendril. Homoclinic points appear where CHAOTIC regions touch in a hyperbolic FIXED POINT.

A small DISK centered near a homoclinic point includes infinitely many periodic points of different periods. Poincaré showed that if there is a single homoclinic point, there are an infinite number. More specifically, there are infinitely many homoclinic points in each small disk (Nusse and Yorke 1996).

see also HETEROCLINIC POINT

References

Nusse, H. E. and Yorke, J. A. "Basins of Attraction." *Science* **271**, 1376–1380, 1996.
Tabor, M. *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, p. 145, 1989.

Homogeneous Coordinates

see TRILINEAR COORDINATES

Homogeneous Function

A function which satisfies

$$f(tx,ty) = t^n f(x,y)$$

for a fixed n. MEANS, the WEIERSTRAß ELLIPTIC FUNCTION, and TRIANGLE CENTER FUNCTIONS are homogeneous functions. A transformation of the variables of a TENSOR changes the TENSOR into another whose components are linear homogeneous functions of the components of the original TENSOR.

see also EULER'S HOMOGENEOUS FUNCTION THEOREM

Homogeneous Numbers

Two numbers are homogeneous if they have identical PRIME FACTORS. An example of a homogeneous pair is (6, 36), both of which share PRIME FACTORS 2 and 3:

$$6 = 2\cdot 3$$
$$36 = 2^2\cdot 3^2.$$

see also HETEROGENEOUS NUMBERS, PRIME FACTORS, PRIME NUMBER

References

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 146, 1983.

Homogeneous Polynomial

A multivariate polynomial (i.e., a POLYNOMIAL in more than one variable) with all terms having the same degree. For example, $x^3+xyz+y^2z+z^3$ is a homogeneous polynomial of degree three.

see also POLYNOMIAL

Homographic

see MÖBIUS TRANSFORMATION

Homography

A CIRCLE-preserving transformation composed of an EVEN number of inversions.

see also ANTIHOMOGRAPHY

Homological Algebra

An abstract ALGEBRA concerned with results valid for many different kinds of SPACES.

References
Hilton, P. and Stammbach, U. *A Course in Homological Algebra, 2nd ed.* New York: Springer-Verlag, 1997.
Weibel, C. A. *An Introduction to Homological Algebra.* New York: Cambridge University Press, 1994.

Homologous Points

The extremities of PARALLEL RADII of two CIRCLES are called homologous with respect to the SIMILITUDE CENTER collinear with them.

see also ANTIHOMOLOGOUS POINTS

References
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, p. 19, 1929.

Homolographic Equal Area Projection

see MOLLWEIDE PROJECTION

Homology (Geometry)

A PERSPECTIVE COLLINEATION in which the center and axis are not incident.

see also ELATION, HARMONIC HOMOLOGY, PERSPECTIVE COLLINEATION

Homology Group

The term "homology group" usually means a singular homology group, which is an ABELIAN GROUP which partially counts the number of HOLES in a TOPOLOGICAL SPACE. In particular, singular homology groups form a MEASURE of the HOLE structure of a SPACE, but they are one particular measure and they don't always pick up everything.

In addition, there are "generalized homology groups" which are *not* singular homology groups.

Homology (Topology)

Historically, the term "homology" was first used in a topological sense by Poincaré. To him, it meant pretty much what is now called a COBORDISM, meaning that a homology was thought of as a relation between MANIFOLDS mapped into a MANIFOLD. Such MANIFOLDS form a homology when they form the boundary of a higher-dimensional MANIFOLD inside the MANIFOLD in question.

To simplify the definition of homology, Poincaré simplified the spaces he dealt with. He assumed that all the spaces he dealt with had a triangulation (i.e., they were "SIMPLICIAL COMPLEXES"). Then instead of talking about general "objects" in these spaces, he restricted himself to subcomplexes, i.e., objects in the space made up only on the simplices in the TRIANGULATION of the space. Eventually, Poincaré's version of homology was dispensed with and replaced by the more general SINGULAR HOMOLOGY. SINGULAR HOMOLOGY is the concept mathematicians mean when they say "homology."

In modern usage, however, the word homology is used to mean HOMOLOGY GROUP. For example, if someone says "X did Y by computing the homology of Z," they mean "X did Y by computing the HOMOLOGY GROUPS of Z." But sometimes homology is used more loosely in the context of a "homology in a SPACE," which corresponds to *singular* homology groups.

Singular homology groups of a SPACE measure the extent to which there are finite (compact) boundaryless GADGETS in that SPACE, such that these GADGETS are not the boundary of other finite (compact) GADGETS in that SPACE.

A generalized homology or cohomology theory must satisfy all of the EILENBERG-STEENROD AXIOMS with the exception of the DIMENSION AXIOM.

see also COHOMOLOGY, DIMENSION AXIOM, EILENBERG-STEENROD AXIOMS, GADGET, HOMOLOGICAL ALGEBRA, HOMOLOGY GROUP, SIMPLICIAL COMPLEX, SIMPLICIAL HOMOLOGY, SINGULAR HOMOLOGY

Homomorphic

Related to one another by a HOMOMORPHISM.

Homomorphism

A term used in CATEGORY THEORY to mean a general MORPHISM.

see also HOMEOMORPHISM, MORPHISM

Homoscedastic

A set of STATISTICAL DISTRIBUTIONS having the same VARIANCE.

see also HETEROSCEDASTIC

Homothecy

see DILATION

Homothetic

Two figures are homothetic if they are related by a DILATION (a dilation is also known as a HOMOTHECY). This means that they lie in the same plane and corresponding sides are PARALLEL; such figures have connectors of corresponding points which are CONCURRENT at a point known as the HOMOTHETIC CENTER. The HOMOTHETIC CENTER divides each connector in the same ratio k, known as the SIMILITUDE RATIO. For figures which are similar but do not have PARALLEL sides, a SIMILITUDE CENTER exists.

see also DILATION, HOMOTHETIC CENTER, PERSPECTIVE, SIMILITUDE RATIO

References
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, 1929.

Homothetic Center

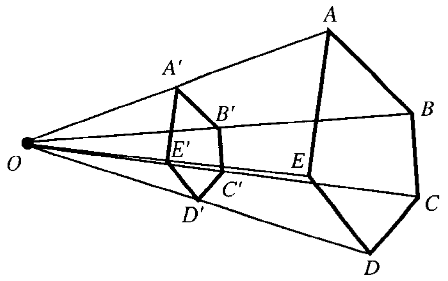

The meeting point of lines that connect corresponding points from HOMOTHETIC figures. In the above figure, O is the homothetic center of the HOMOTHETIC figures $ABCDE$ and $A'B'C'D'E'$. For figures which are similar but do not have PARALLEL sides, a SIMILITUDE CENTER exists (Johnson 1929, pp. 16–20).

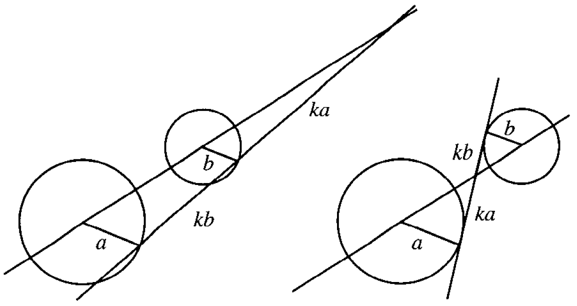

Given two nonconcentric CIRCLES, draw RADII PARALLEL and in the same direction. Then the line joining the extremities of the RADII passes through a fixed point on the line of centers which divides that line externally in the ratio of RADII. This point is called the external homothetic center, or external center of similitude (Johnson 1929, pp. 19–20 and 41).

If RADII are drawn PARALLEL but instead in opposite directions, the extremities of the RADII pass through a fixed point on the line of centers which divides that line internally in the ratio of RADII (Johnson 1929, pp. 19–20 and 41). This point is called the internal homothetic center, or internal center of similitude (Johnson 1929, pp. 19–20 and 41).

The position of the homothetic centers for two circles of radii r_i, centers (x_i, y_i), and segment angle θ are given by solving the simultaneous equations

$$y - y_2 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_2)$$
$$y - y_2^{\pm} = \frac{y_2^{\pm} - y_1^{\pm}}{x_2^{\pm} - x_1^{\pm}}(x - x_2^{\pm})$$

for (x, y), where

$$x_i^{\pm} \equiv x_i + (-1)^i r_i \cos\theta$$
$$y_i^{\pm} \equiv y_i + (-1)^i r_i \sin\theta,$$

and the plus signs give the external homothetic center, while the minus signs give the internal homothetic center.

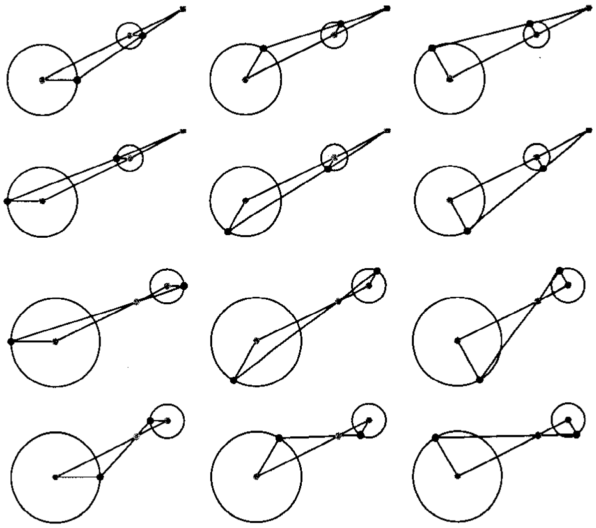

As the above diagrams show, as the angles of the parallel segments are varied, the positions of the homothetic centers remain the same. This fact provides a (slotted) LINKAGE for converting circular motion with one radius to circular motion with another.

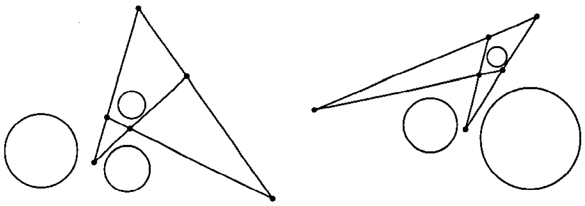

The six homothetic centers of three circles lie three by three on four lines (Johnson 1929, p. 120), which "enclose" the smallest circle.

The homothetic center of triangles is the PERSPECTIVE CENTER of HOMOTHETIC TRIANGLES. It is also called the SIMILITUDE CENTER (Johnson 1929, pp. 16–17).

see also APOLLONIUS' PROBLEM, PERSPECTIVE, SIMILITUDE CENTER

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, 1929.

✽ Weisstein, E. W. "Plane Geometry." `http://www.astro.virginia.edu/~eww6n/math/notebooks/PlaneGeometry.m`.

Homothetic Position

Two similar figures with PARALLEL homologous LINES and connectors of HOMOLOGOUS POINTS CONCURRENT at the HOMOTHETIC CENTER are said to be in homothetic position. If two SIMILAR figures are in the same plane but the corresponding sides are not PARALLEL, there exists a self-HOMOLOGOUS POINT which occupies the same homologous position with respect to the two figures.

Homothetic Triangles

Nonconcurrent TRIANGLES with PARALLEL sides are always HOMOTHETIC. Homothetic triangles are always PERSPECTIVE TRIANGLES. Their PERSPECTIVE CENTER is called their HOMOTHETIC CENTER.

Homotopy

A continuous transformation from one FUNCTION to another. A homotopy between two functions f and g from a SPACE X to a SPACE Y is a continuous MAP G from $X \in [0,1] \mapsto Y$ such that $G(x,0) = f(x)$ and $G(x,1) = g(x)$. Another way of saying this is that a homotopy is a path in the mapping SPACE Map(X,Y) from the first FUNCTION to the second.

see also h-COBORDISM

Homotopy Axiom

One of the EILENBERG-STEENROD AXIOMS which states that, if $f:(X,A) \to (Y,B)$ is homotopic to $g:(X,A) \to (Y,B)$, then their INDUCED MAPS $f_*: H_n(X,A) \to H_n(Y,B)$ and $g_*: H_n(X,A) \to H_n(Y,B)$ are the same.

Homotopy Group

A GROUP related to the HOMOTOPY classes of MAPS from SPHERES $\mathbb{S}^n$ into a SPACE X.

see also COHOMOTOPY GROUP

Homotopy Theory

The branch of ALGEBRAIC TOPOLOGY which deals with HOMOTOPY GROUPS.

References
Aubry, M. *Homotopy Theory and Models.* Boston, MA: Birkhäuser, 1995.

Honeycomb

A TESSELLATION in n-D, for $n \geq 3$. The only regular honeycomb in 3-D is $\{4,3,4\}$, which consists of eight cubes meeting at each VERTEX. The only quasiregular honeycomb (with regular cells and semiregular VERTEX FIGURES) has each VERTEX surrounded by eight TETRAHEDRA and six OCTAHEDRA and is denoted $\left\{\begin{matrix}3\\3,4\end{matrix}\right\}$.

There are many semiregular honeycombs, such as $\left\{\begin{matrix}3,3\\4\end{matrix}\right\}$, in which each VERTEX consists of two OCTAHEDRA $\{3,4\}$ and four CUBOCTAHEDRA $\left\{\begin{matrix}3\\4\end{matrix}\right\}$.

see also SPONGE, TESSELLATION

References
Bulatov, V. "Infinite Regular Polyhedra." `http://www.physics.orst.edu/~bulatov/polyhedra/infinite/`.

Hoof

see CYLINDRICAL WEDGE

Hook

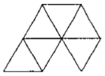

A 6-POLYIAMOND.

References
Golomb, S. W. *Polyominoes: Puzzles, Patterns, Problems, and Packings, 2nd ed.* Princeton, NJ: Princeton University Press, p. 92, 1994.

Hook Length Formula

A FORMULA for the number of YOUNG TABLEAUX associated with a given YOUNG DIAGRAM. In each box, write the sum of one plus the number of boxes horizontally to the right and vertically below the box (the "hook length"). The number of tableaux is then $n!$ divided by the product of all "hook lengths". The `Combinatorica`NumberOfTableaux` function in *Mathematica*® implements the hook length formula.

see also YOUNG DIAGRAM, YOUNG TABLEAU

References
Jones, V. "Hecke Algebra Representations of Braid Groups and Link Polynomials." *Ann. Math.* **126**, 335–388, 1987.
Skiena, S. *Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica.* Reading, MA: Addison-Wesley, 1990.

Hopf Algebra

Let a graded module A have a multiplication ϕ and a co-multiplication ψ. Then if ϕ and ψ have the unity of k as unity and $\psi:(A,\phi) \to (A,\phi) \otimes (A,\phi)$ is an algebra homomorphism, then (A,ϕ,ψ) is called a Hopf algebra.

Hopf Bifurcation

The BIFURCATION of a FIXED POINT to a LIMIT CYCLE (Tabor 1989).

References
Guckenheimer, J. and Holmes, P. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields, 3rd ed.* New York: Springer-Verlag, pp. 150–154, 1997.
Marsden, J. and McCracken, M. *Hopf Bifurcation and Its Applications.* New York: Springer-Verlag, 1976.
Tabor, M. *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, p. 197, 1989.

Hopf Circle

see HOPF MAP

Hopf Link

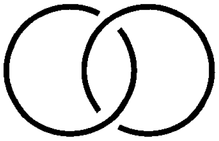

The LINK 2_1^2 which has JONES POLYNOMIAL

$$V(t) = -t - t^{-1}$$

and HOMFLY POLYNOMIAL

$$P(z,\alpha) = z^{-1}(\alpha^{-1} - \alpha^{-3}) + z\alpha^{-1}.$$

It has BRAID WORD $\sigma_1{}^2$.

Hopf Map

The first example discovered of a MAP from a higher-dimensional SPHERE to a lower-dimensional SPHERE which is not null-HOMOTOPIC. Its discovery was a shock to the mathematical community, since it was believed at the time that all such maps were null-HOMOTOPIC, by analogy with HOMOLOGY GROUPS. The Hopf map takes points (X_1, X_2, X_3, X_4) on a 3-sphere to points on a 2-sphere (x_1, x_2, x_3)

$$\begin{aligned} x_1 &= 2(X_1X_2 + X_3X_4) \\ x_2 &= 2(X_1X_4 - X_2X_3) \\ x_3 &= (X_1{}^2 + X_3{}^2) - (X_2{}^2 + X_4{}^2). \end{aligned}$$

Every point on the two SPHERES corresponds to a CIRCLE called the HOPF CIRCLE on the 3-SPHERE.

Hopf's Theorem

A NECESSARY and SUFFICIENT condition for a MEASURE which is quasi-invariant under a transformation to be equivalent to an invariant PROBABILITY MEASURE is that the transformation cannot (in a measure theoretic sense) compress the SPACE.

Horizontal

Oriented in position PERPENDICULAR to up-down, and therefore PARALLEL to a flat surface.

see also VERTICAL

Horizontal-Vertical Illusion

see VERTICAL-HORIZONTAL ILLUSION

Horn Angle

The configuration formed by two curves starting at a point, called the VERTEX V, in a common direction. They are concrete illustrations of non-Archimedean geometries.

References

Kasner, E. "The Recent Theory of the Horn Angle." *Scripta Math* **11**, 263–267, 1945.

Horn Cyclide

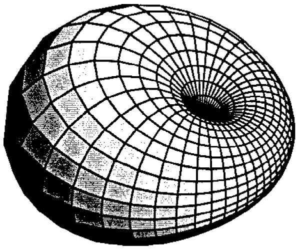

The inversion of a HORN TORUS. If the inversion center lies on the torus, then the horn cyclide degenerates to a PARABOLIC HORN CYCLIDE.

see also CYCLIDE, HORN TORUS, PARABOLIC CYCLIDE, RING CYCLIDE, SPINDLE CYCLIDE, TORUS

Horn Torus

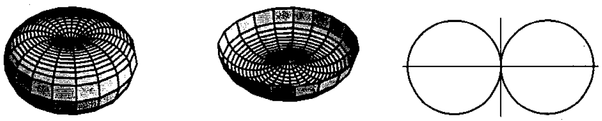

One of the three STANDARD TORI given by the parametric equations

$$x = (c + a\cos v)\cos u \tag{1}$$
$$y = (c + a\cos v)\sin u \tag{2}$$
$$z = a\sin v \tag{3}$$

with $a = c$. The inversion of a horn torus is a HORN CYCLIDE (or PARABOLIC HORN CYCLIDE). The above left figure shows a horn torus, the middle a cutaway, and the right figure shows a CROSS-SECTION of the horn torus through the xz-plane.

see also CYCLIDE, HORN CYCLIDE, RING TORUS, SPINDLE TORUS, STANDARD TORI, TORUS

References

Gray, A. "Tori." §11.4 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 218–220, 1993.

Pinkall, U. "Cyclides of Dupin." §3.3 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 28–30, 1986.

Horned Sphere

see ALEXANDER'S HORNED SPHERE, ANTOINE'S HORNED SPHERE

Horner's Method

Let

$$P(x) = a_nx^n + \ldots + a_0 \tag{1}$$

and $b_n \equiv a_n$. If we then define

$$b_k \equiv a_k + b_{k-1}x_0 \tag{2}$$

for $k = n-1$, $n-2$, …, 0, we obtain $b_0 = P(x_0)$. It therefore follows that

$$P(x) = (x - x_0)Q(x) + b_0, \tag{3}$$

where

$$Q(x) \equiv b_nx^{n-1} + b_{n-1}x^{n-2} + \ldots + b_2x + b_1. \tag{4}$$

In addition,

$$P'(x) = Q(x) + (x - x_0)Q'(x) \tag{5}$$
$$P'(x_0) = Q(x_0). \tag{6}$$

Horner's Rule

A rule for POLYNOMIAL computation which both reduces the number of necessary multiplications and results in less numerical instability due to potential subtraction of one large number from another. The rule simply factors out POWERS of x, giving

$$a_nx^n+a_{n-1}x^{n-1}+\ldots+a_0=((a_nx+a_{n-1})x+\ldots)x+a_0.$$

References
Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 9, 1991.

Horocycle

The LOCUS of a point which is derived from a fixed point Q by continuous parallel displacement.

References
Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, p. 300, 1969.

Horse Fetter

see HIPPOPEDE

Horseshoe Map

see SMALE HORSESHOE MAP

Hough Transform

A technique used to detect boundaries in digital images.

Householder's Method

A ROOT-finding algorithm based on the iteration formula

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)}\left\{1-\frac{[f(x_n)]^2f''(x_n)}{2[f'(x_n)]^2}\right\}.$$

This method, like NEWTON'S METHOD, has poor convergence properties near any point where the DERIVATIVE $f'(x)=0$.

see also NEWTON'S METHOD

References
Householder, A. S. *The Numerical Treatment of a Single Nonlinear Equation.* New York: McGraw-Hill, 1970.

Howell Design

Let S be a set of $n+1$ symbols, then a Howell design $H(s,2n)$ on symbol set S is an $s\times s$ array H such that

1. Every cell of H is either empty or contains an unordered pair of symbols from S,
2. Every symbol of S occurs once in each row and column of H, and
3. Every unordered pair of symbols occurs in at most one cell of H.

References
Colbourn, C. J. and Dinitz, J. H. (Eds.) "Howell Designs." Ch. 26 in *CRC Handbook of Combinatorial Designs.* Boca Raton, FL: CRC Press, pp. 381–385, 1996.

Hub

The central point in a WHEEL GRAPH W_n. The hub has DEGREE $n-1$.

see also WHEEL GRAPH

Huffman Coding

A lossless data compression algorithm which uses a small number of bits to encode common characters. Huffman coding approximates the probability for each character as a POWER of 1/2 to avoid complications associated with using a nonintegral number of bits to encode characters using their actual probabilities.

References
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Huffman Coding and Compression of Data." Ch. 20.4 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 896–901, 1992.

Hull

see AFFINE HULL, CONVEX HULL

Humbert's Theorem

The NECESSARY and SUFFICIENT condition that an algebraic curve has an algebraic INVOLUTE is that the ARC LENGTH is a two-valued algebraic function of the coordinates of the extremities. Furthermore, this function is a ROOT of a QUADRATIC EQUATION whose COEFFICIENTS are rational functions of x and y.

References
Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 195, 1959.

Hundkurve

see TRACTRIX

Hundred

$100=10^2$. Madachy (1979) gives a number of algebraic equations using the digits 1 to 9 which evaluate to 100, such as

$$\begin{aligned}(7-5)^2+96+8-4-3-1&=100\\ 3^2+91+7+8-6-5-4&=100\\ \sqrt{9}-6+72-(1)(3!)-8+45&=100\\ 123-45-67+89&=100,\end{aligned}$$

and so on.

see also 10, BILLION, HUNDRED, LARGE NUMBER, MILLION, THOUSAND

References
Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 156–159, 1979.

Hunt's Surface

An ALGEBRAIC SURFACE given by the implicit equation

$$4(x^2+y^2+z^2-13)^3+27(3x^2+y^2-4z^2-12)^2=0.$$

References

Hunt, B. "Algebraic Surfaces." `http://www.mathematik.uni-kl.de/~wwwagag/Galerie.html`.

Nordstrand, T. "Hunt's Surface." `http://www.uib.no/people/nfytn/hunttxt.htm`.

Huntington Equation

An equation proposed by Huntington (1933) as part of his definition of a BOOLEAN ALGEBRA,

$$n(n(x)+y)+n(n(x)+n(y))=x.$$

see also ROBBINS ALGEBRA, ROBBINS EQUATION

References

Huntington, E. V. "New Sets of Independent Postulates for the Algebra of Logic, with Special Reference to Whitehead and Russell's *Principia Mathematica*." *Trans. Amer. Math. Soc.* **35**, 274–304, 1933.

Huntington, E. V. "Boolean Algebra. A Correction." *Trans. Amer. Math. Soc.* **35**, 557–558, 1933.

Hurwitz Equation

The DIOPHANTINE EQUATION

$${x_1}^2+{x_2}^2+\ldots+{x_n}^2=ax_1x_2\cdots x_n$$

which has no INTEGER solutions for $a>n$.

see also LAGRANGE NUMBER (DIOPHANTINE EQUATION)

References

Guy, R. K. "Markoff Numbers." §D12 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 166–168, 1994.

Hurwitz's Irrational Number Theorem

As Lagrange showed, any IRRATIONAL NUMBER α has an infinity of rational approximations p/q which satisfy

$$\left|\alpha-\frac{p}{q}\right|<\frac{1}{\sqrt{5}\,q^2}. \qquad (1)$$

Similarly, if $\alpha\neq\frac{1}{2}(1+\sqrt{5}\,)$,

$$\left|\alpha-\frac{p}{q}\right|<\frac{1}{\sqrt{8}\,q^2}, \qquad (2)$$

and if $\alpha\neq\frac{1}{2}(1+\sqrt{5}\,)\neq\sqrt{2}$,

$$\left|\alpha-\frac{p}{q}\right|<\frac{5}{\sqrt{221}}\frac{1}{q^2}. \qquad (3)$$

In general, even tighter bounds of the form

$$\left|\alpha-\frac{p}{q}\right|<\frac{1}{L_nq^2} \qquad (4)$$

can be obtained for the best rational approximation possible for an arbitrary irrational number α, where the L_n are called LAGRANGE NUMBERS and get steadily larger for each "bad" set of irrational numbers which is excluded.

see also HURWITZ'S IRRATIONAL NUMBER THEOREM, LIOUVILLE'S RATIONAL APPROXIMATION THEOREM, LIOUVILLE-ROTH CONSTANT, MARKOV NUMBER, ROTH'S THEOREM, SEGRE'S THEOREM, THUE-SIEGEL-ROTH THEOREM

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 40, 1987.

Chandrasekharan, K. *An Introduction to Analytic Number Theory.* Berlin: Springer-Verlag, p. 23, 1968.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 187–189, 1996.

Hurwitz Number

A number with a continued fraction whose terms are the values of one or more POLYNOMIALS evaluated on consecutive INTEGERS and then interleaved. This property is preserved by MÖBIUS TRANSFORMATIONS (Beeler *et al.* 1972, p. 44).

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Hurwitz Polynomial

A POLYNOMIAL with REAL POSITIVE COEFFICIENTS and ROOTS which are either NEGATIVE or pairwise conjugate with NEGATIVE REAL PARTS.

Hurwitz-Radon Theorem

Determined the possible values of r and n for which there is an IDENTITY of the form

$$(x_1{}^2 + \ldots + x_r{}^2)(y_1{}^2 + \ldots + y_r{}^2) = z_1{}^2 + \ldots + z_n{}^2.$$

Hurwitz's Root Theorem

Let $\{f(x)\}$ be a SEQUENCE of ANALYTIC FUNCTIONS REGULAR in a region G, and let this sequence be UNIFORMLY CONVERGENT in every CLOSED SUBSET of G. If the ANALYTIC FUNCTION

$$\lim_{n\to\infty} f_n(x) = f(x)$$

does not vanish identically, then if $x = a$ is a zero of $f(x)$ of order k, a NEIGHBORHOOD $|x-a| < \delta$ of $x = a$ and a number N exist such that if $n > N$, $f_n(x)$ has exactly k zeros in $|x-a| < \delta$.

References

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., p. 22, 1975.

Hurwitz Zeta Function

A generalization of the RIEMANN ZETA FUNCTION with a FORMULA

$$\zeta(s,a) \equiv \sum_{k=0}^{\infty} \frac{1}{(k+a)^s}, \tag{1}$$

where any term with $k + a = 0$ is excluded. The Hurwitz zeta function can also be given by the functional equation

$$\zeta\left(s, \frac{p}{q}\right) = 2\Gamma(1-s)(2\pi q)^{s-1} \sum_{n=1}^{q} \sin\left(\frac{\pi s}{2} + \frac{2\pi n p}{q}\right) \zeta\left(1-s, \frac{n}{q}\right) \tag{2}$$

(Apostol 1976, Miller and Adamchik), or the integral

$$\zeta(s,a) = \tfrac{1}{2}a^{-3} + \frac{a^{1-s}}{s-1} + 2\int_0^\infty (a^2+y^2)^{-s/2} \left\{\sin\left[s\tan^{-1}\left(\frac{y}{a}\right)\right]\right\} \frac{dy}{e^{2\pi y}-1}. \tag{3}$$

If $\Re[z] < 0$, then

$$\zeta(z,a) = \frac{2\Gamma(1-z)}{(2\pi)^{1-z}} \left[\sin\left(\frac{\pi z}{2}\right) \sum_{n=1}^{\infty} \frac{\cos(2\pi a n)}{n^{1-z}} + \cos\left(\frac{\pi z}{2}\right) \sum_{n=1}^{\infty} \frac{\sin(2\pi a n)}{n^{1-z}}\right]. \tag{4}$$

The Hurwitz zeta function satisfies

$$\zeta(0,a) = \tfrac{1}{2} - a \tag{5}$$

$$\frac{d}{ds}\zeta(0,a) = \ln[\Gamma(a)] - \tfrac{1}{2}\ln(2\pi) \tag{6}$$

$$\frac{d}{ds}\zeta(0,0) = \tfrac{1}{2}\ln(2\pi), \tag{7}$$

where $\Gamma(z)$ is the GAMMA FUNCTION. The POLYGAMMA FUNCTION $\psi_m(z)$ can be expressed in terms of the Hurwitz zeta function by

$$\psi_m(z) = (-1)^{m+1} m!\,\zeta(1+m,z). \tag{8}$$

For POSITIVE integers k, p, and $q > p$,

$$\begin{aligned}\zeta'\left(-2k+1, \frac{p}{q}\right) &= \frac{[\psi(2k) - \ln(2\pi q)]B_{2k}(p/q)}{2k} - \frac{[\psi(2k) - \ln(2\pi)]B_{2k}}{q^{2k}2k} \\ &\quad + \frac{(-1)^{k+1}\pi}{(2\pi q)^{2k}} \sum_{n=1}^{q-1} \sin\left(\frac{2\pi p n}{q}\right) \psi_{(2k-1)}\left(\frac{n}{q}\right) \\ &\quad + \frac{(-1)^{k+1}2(2k-1)!}{(2\pi q)^{2k}} \sum_{n=1}^{q-1} \cos\left(\frac{2\pi p n}{q}\right) \zeta'\left(2k, \frac{n}{q}\right) \\ &\quad + \frac{\zeta'(-2k+1)}{q^{2k}},\end{aligned} \tag{9}$$

where B_n is a BERNOULLI NUMBER, $B_n(x)$ a BERNOULLI POLYNOMIAL, $\psi_n(z)$ is a POLYGAMMA FUNCTION, and $\zeta(z)$ is a RIEMANN ZETA FUNCTION (Miller and Adamchik). Miller and Adamchik also give the closed-form expressions

$$\zeta'(-2k+1, \tfrac{1}{2}) = -\frac{B_{2k}\ln 2}{4^k k} - \frac{(2^{2k-1}-1)\zeta'(-2k+1)}{2^{2k-1}} \tag{10}$$

$$\begin{aligned}\zeta'\left(-2k+1, \begin{smallmatrix}\frac{1}{3}\\ \frac{2}{3}\end{smallmatrix}\right) &= \mp\frac{(9^k-1)B_{2k}\pi}{\sqrt{3}(3^{2k-1}-1)8k} - \frac{B_{2k}\ln 3}{(3^{2k-1})4k} \\ &\quad \mp \frac{(-1)^k\psi_{2k-1}(\frac{1}{3})}{2\sqrt{3}(6\pi)^{2k-1}} - \frac{(3^{2k-1}-1)\zeta'(-2k+1)}{2(3^{2k-1})}\end{aligned} \tag{11}$$

$$\begin{aligned}\zeta'\left(-2k+1, \begin{smallmatrix}\frac{1}{4}\\ \frac{3}{4}\end{smallmatrix}\right) &= \mp\frac{(4^k+1)B_{2k}\pi}{4^{k+1}k} + \frac{(4^{k-1}-1)B_{2k}\ln 2}{2^{3k-1}k} \\ &\quad \mp \frac{(-1)^k\psi_{2k-1}(\frac{1}{4})}{4(8\pi)^{2k-1}} - \frac{(2^{2k-1}-1)\zeta'(-2k+1)}{2^{4k-1}}\end{aligned} \tag{12}$$

$$\begin{aligned}\zeta'\left(-2k+1, \begin{smallmatrix}\frac{1}{6}\\ \frac{5}{6}\end{smallmatrix}\right) &= \mp\frac{(9^k-1)(2^{2k-1}+1)B_{2k}\pi}{\sqrt{3}(6^{2k-1})8k} \\ &\quad + \frac{B_{2k}(3^{2k-1}-1)\ln 2}{(6^{2k-1})4k} + \frac{B_{2k}(2^{2k-1}-1)\ln 3}{(6^{2k-1})4k} \\ &\quad \mp \frac{(-1)^k(2^{2k-1}+1)\psi_{2k-1}(\frac{1}{3})}{2\sqrt{3}\,(12\pi)^{2k-1}} \\ &\quad + \frac{(2^{2k-1}-1)(3^{2k-1}-1)\zeta'(-2k+1)}{2(6^{2k-1})}.\end{aligned} \tag{13}$$

see also KHINTCHINE'S CONSTANT, POLYGAMMA FUNCTION, PSI FUNCTION, RIEMANN ZETA FUNCTION, ZETA FUNCTION

References

Apostol, T. M. *Introduction to Analytic Number Theory.* New York: Springer-Verlag, 1995.

Elizalde, E.; Odintsov, A. D.; and Romeo, A. *Zeta Regularization Techniques with Applications.* River Edge, NJ: World Scientific, 1994.

Knopfmacher, J. "Generalised Euler Constants." *Proc. Edinburgh Math. Soc.* **21**, 25–32, 1978.

Magnus, W. and Oberhettinger, F. *Formulas and Theorems for the Special Functions of Mathematical Physics, 3rd ed.* New York: Springer-Verlag, 1966.

Miller, J. and Adamchik, V. "Derivatives of the Hurwitz Zeta Function for Rational Arguments." Submitted to *J. Symb. Comput.* `http://www.wolfram.com/~victor/articles/hurwitz.html`.

Spanier, J. and Oldham, K. B. "The Hurwitz Function $\zeta(\nu;u)$." Ch. 62 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 653–664, 1987.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, pp. 268–269, 1950.

Hutton's Formula

The MACHIN-LIKE FORMULA

$$\tfrac{1}{4}\pi = 2\tan^{-1}(\tfrac{1}{3}) + \tan^{-1}(\tfrac{1}{7}).$$

The other two-term MACHIN-LIKE FORMULAS are EULER'S MACHIN-LIKE FORMULA, HERMANN'S FORMULA, and MACHIN'S FORMULA.

Hutton's Method

see LAMBERT'S METHOD

Hyperbola

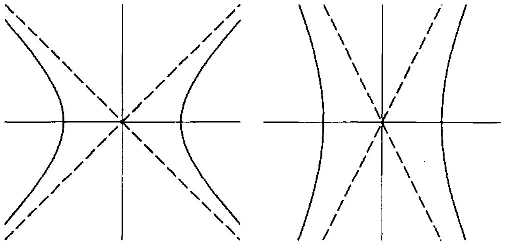

In general, a hyperbola is defined as the LOCUS of all points in the PLANE the difference of whose distance from two fixed points (the FOCI F_1 and F_2) separated by a distance $2c$, where

$$c \equiv \sqrt{a^2+b^2}, \tag{1}$$

is a given POSITIVE constant. By analogy with the definition of the ELLIPSE, the equation for a hyperbola with SEMIMAJOR AXIS a parallel to the x-AXIS and SEMIMINOR AXIS b parallel to the y-AXIS is given by

$$\frac{(x-x_0)^2}{a^2} - \frac{(y-y_0)^2}{b^2} = 1. \tag{2}$$

Unlike the ELLIPSE, no points of the hyperbola actually lie on the SEMIMINOR AXIS, but rather the ratio b/a determined the vertical scaling of the hyperbola. The ECCENTRICITY of the hyperbola is defined as

$$e \equiv \frac{c}{a} = \sqrt{1+\frac{b^2}{a^2}}. \tag{3}$$

In the standard equation of the hyperbola, the center is located at (x_0, y_0), the FOCI are at $(x_0 \pm c, y_0)$, and the vertices are at $(x_0 \pm a, y_0)$. The so-called ASYMPTOTES (shown as the dashed lines in the above figures) can be found by substituting 0 for the 1 on the right side of the general equation (2),

$$y = \pm\frac{b}{a}(x-x_0) + y_0, \tag{4}$$

and therefore have SLOPES $\pm b/a$.

The special case $a = b$ (the left diagram above) is known as a RIGHT HYPERBOLA because the ASYMPTOTES are PERPENDICULAR.

In POLAR COORDINATES, the equation of a hyperbola centered at the ORIGIN (i.e., with $x_0 = y_0 = 0$) is

$$r^2 = \frac{a^2b^2}{b^2\cos^2\theta - a^2\sin^2\theta}. \tag{5}$$

In POLAR COORDINATES centered at a FOCUS,

$$r = \frac{a(e^2-1)}{1-e\cos\theta}. \tag{6}$$

The two-center BIPOLAR COORDINATES equation with origin at a FOCUS is

$$r_1 - r_2 = \pm 2a. \tag{7}$$

The parametric equations for the hyperbola are

$$x = \pm a\cosh t \tag{8}$$

$$y = b\sinh t. \tag{9}$$

The CURVATURE and TANGENTIAL ANGLE are

$$\kappa(t) = -[\cosh(2t)]^{-3/2} \tag{10}$$

$$\phi(t) = -\tan^{-1}(\tanh t). \tag{11}$$

The special case of the RIGHT HYPERBOLA was first studied by Menaechmus. Euclid and Aristaeus wrote about the general hyperbola, but only studied one branch of it. The hyperbola was given its present name by Apollonius, who was the first to study both branches. The FOCUS and DIRECTRIX were considered by Pappus (MacTutor Archive). The hyperbola is the shape of an orbit of a body on an escape trajectory (i.e., a body with positive energy), such as some comets, about a fixed mass, such as the sun.

The Locus of the apex of a variable Cone containing an Ellipse fixed in 3-space is a hyperbola through the Foci of the Ellipse. In addition, the Locus of the apex of a Cone containing that hyperbola is the original Ellipse. Furthermore, the Eccentricities of the Ellipse and hyperbola are reciprocals.

see also Conic Section, Ellipse, Hyperboloid, Jerabek's Hyperbola, Kiepert's Hyperbola, Parabola, Quadratic Curve, Rectangular Hyperbola, Reflection Property, Right Hyperbola

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 199–200, 1987.

Casey, J. "The Hyperbola." Ch. 7 in *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions, with Numerous Examples, 2nd ed., rev. enl.* Dublin: Hodges, Figgis, & Co., pp. 250–284, 1893.

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 75–76, 1996.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 79–82, 1972.

Lee, X. "Hyperbola." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Hyperbola_dir/hyperbola.html`.

Lockwood, E. H. "The Hyperbola." Ch. 3 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 24–33, 1967.

MacTutor History of Mathematics Archive. "Hyperbola." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Hyperbola.html`.

Hyperbola Evolute

The Evolute of a Rectangular Hyperbola is the Lamé Curve

$$(ax)^{2/3} - (by)^{2/3} = (a+b)^{2/3}.$$

From a point between the two branches of the Evolute, two Normals can be drawn to the Hyperbola. However, from a point beyond the Evolute, four Normals can be drawn.

Hyperbola Inverse Curve

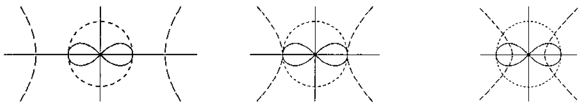

For a Hyperbola with $a = b$ with Inversion Center at the center, the Inverse Curve

$$x = \frac{2k\cos t}{a[3-\cos(2t)]} \tag{1}$$

$$y = \frac{k\sin(2t)}{a[3-\cos(2t)]} \tag{2}$$

is a Lemniscate.

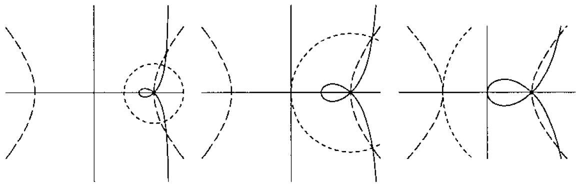

For an Inversion Center at the Vertex, the Inverse Curve

$$x = a + \frac{4k\cos t\sin^2(\tfrac{1}{2}t)}{a[5-4\cos t+\cos(2t)-2\sin(2t)]} \tag{3}$$

$$y = a + \frac{k(\tan t - 1)}{a[(\sec t - 1)^2 + (\tan t - 1)^2]} \tag{4}$$

is a Right Strophoid.

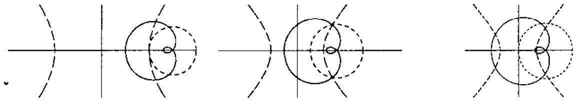

For an Inversion Center at the Focus, the Inverse Curve

$$x = ae = \frac{k\cos t(1-e\cos t)}{a(\cos t - e)^2} \tag{5}$$

$$y = \frac{\sqrt{e^2-1}\,k\sin(2t)}{2a(\cos t - e)^2} \tag{6}$$

is a Limaçon, where e is the Eccentricity.

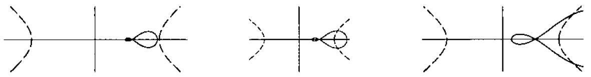

For a Hyperbola with $a = \sqrt{3}\,b$ and Inversion Center at the Vertex, the Inverse Curve

$$x = b + \frac{2k\cos t(\sqrt{3}-\cos t)}{b[9-4\sqrt{3}\cos t+\cos(2t)-2\sin(2t)]} \tag{7}$$

$$y = b + \frac{k(\tan t - 1)}{b[(\sqrt{3}\sec t - 1)^2 + (\tan t - 1)^2]} \tag{8}$$

is a Maclaurin Trisectrix.

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, p. 203, 1972.

Hyperbola Pedal Curve

The Pedal Curve of a Hyperbola with the Pedal Point at the Focus is a Circle. The Pedal Curve of a Rectangular Hyperbola with Pedal Point at the center is a Lemniscate.

Hyperbolic Automorphism

see Anosov Automorphism

Hyperbolic Cosecant

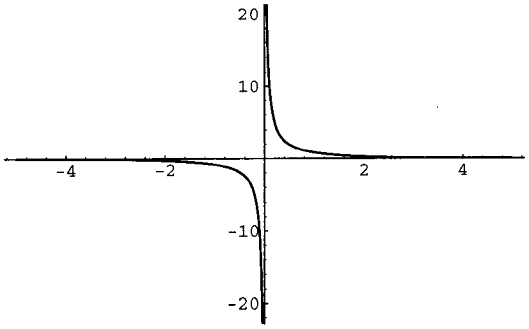

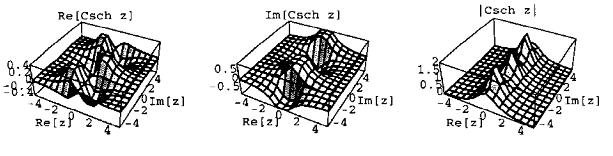

The hyperbolic cosecant is defined as

$$\operatorname{csch} x \equiv \frac{1}{\sinh x} = \frac{2}{e^x - e^{-x}}.$$

see also BERNOULLI NUMBER, BIPOLAR COORDINATES, BIPOLAR CYLINDRICAL COORDINATES, COSECANT, HELMHOLTZ DIFFERENTIAL EQUATION—TOROIDAL COORDINATES, HYPERBOLIC SINE, POINSOT'S SPIRALS, SURFACE OF REVOLUTION, TOROIDAL FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Hyperbolic Functions." §4.5 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 83–86, 1972.

Spanier, J. and Oldham, K. B. "The Hyperbolic Secant sech(x) and Cosecant csch(x) Functions." Ch. 29 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 273–278, 1987.

Hyperbolic Cosine

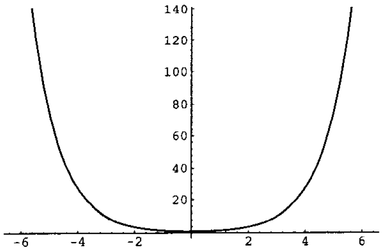

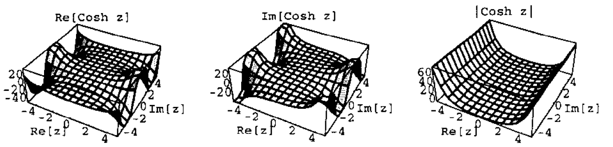

The hyperbolic cosine is defined as

$$\cosh x \equiv \tfrac{1}{2}(e^x + e^{-x}).$$

This function describes the shape of a hanging cable, known as the CATENARY.

see also BIPOLAR COORDINATES, BIPOLAR CYLINDRICAL COORDINATES, BISPHERICAL COORDINATES, CATENARY, CATENOID, CHI, CONICAL FUNCTION, CORRELATION COEFFICIENT—GAUSSIAN BIVARIATE DISTRIBUTION, COSINE, CUBIC EQUATION, DE MOIVRE'S IDENTITY, ELLIPTIC CYLINDRICAL COORDINATES, ELSASSER FUNCTION, FIBONACCI HYPERBOLIC COSINE, FIBONACCI HYPERBOLIC SINE, HYPERBOLIC GEOMETRY, HYPERBOLIC LEMNISCATE FUNCTION, HYPERBOLIC SINE, HYPERBOLIC SECANT, HYPERBOLIC TANGENT, INVERSIVE DISTANCE, LAPLACE'S EQUATION—BIPOLAR COORDINATES, LAPLACE'S EQUATION—BISPHERICAL COORDINATES, LAPLACE'S EQUATION—TOROIDAL COORDINATES, LEMNISCATE FUNCTION, LORENTZ GROUP, MATHIEU DIFFERENTIAL EQUATION, MEHLER'S BESSEL FUNCTION FORMULA, MERCATOR PROJECTION, MODIFIED BESSEL FUNCTION OF THE FIRST KIND, OBLATE SPHEROIDAL COORDINATES, PROLATE SPHEROIDAL COORDINATES, PSEUDOSPHERE, RAMANUJAN COS/COSH IDENTITY, SINE-GORDON EQUATION, SURFACE OF REVOLUTION, TOROIDAL COORDINATES

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Hyperbolic Functions." §4.5 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 83–86, 1972.

Spanier, J. and Oldham, K. B. "The Hyperbolic Sine sinh(x) and Cosine cosh(x) Functions." Ch. 28 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 263–271, 1987.

Hyperbolic Cotangent

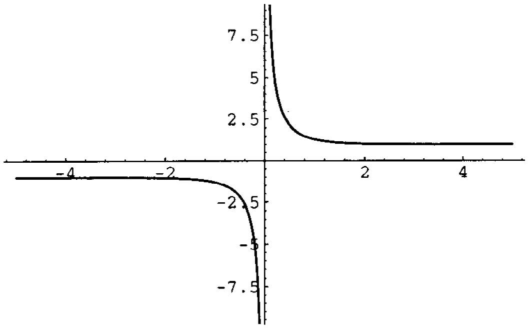

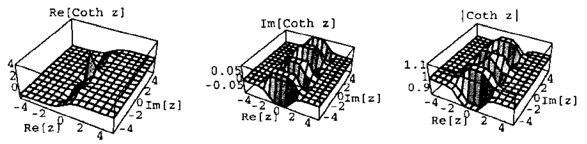

The hyperbolic cotangent is defined as

$$\coth x \equiv \frac{e^x + e^{-x}}{e^x - e^{-x}} = \frac{e^{2x} + 1}{e^{2x} - 1}.$$

Its LAURENT SERIES is

$$\coth x = \frac{1}{x} + \tfrac{1}{3}x - \tfrac{1}{45}x^3 + \ldots .$$

see also BERNOULLI NUMBER, BIPOLAR COORDINATES, BIPOLAR CYLINDRICAL COORDINATES, COTANGENT, FIBONACCI HYPERBOLIC COTANGENT, HYPERBOLIC TANGENT, LAPLACE'S EQUATION—TOROIDAL COORDINATES, LEBESGUE CONSTANTS (FOURIER SERIES), PROLATE SPHEROIDAL COORDINATES, SURFACE OF REVOLUTION, TOROIDAL COORDINATES, TOROIDAL FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Hyperbolic Functions." §4.5 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 83–86, 1972.

Spanier, J. and Oldham, K. B. "The Hyperbolic Tangent tanh(x) and Cotangent coth(x) Functions." Ch. 30 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 279–284, 1987.

Hyperbolic Cube

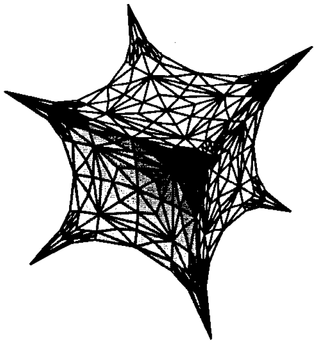

A hyperbolic version of the Euclidean CUBE.

see also HYPERBOLIC DODECAHEDRON, HYPERBOLIC OCTAHEDRON, HYPERBOLIC TETRAHEDRON

References

Rivin, I. "Hyperbolic Polyhedron Graphics." `http://www.mathsource.com/cgi-bin/MathSource/Applications/Graphics/3D/0201-788`.

Hyperbolic Cylinder

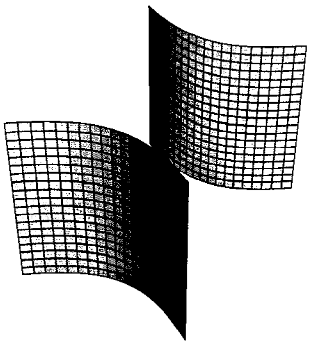

A QUADRATIC SURFACE given by the equation

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = -1.$$

see also ELLIPTIC PARABOLOID, PARABOLOID

Hyperbolic Dodecahedron

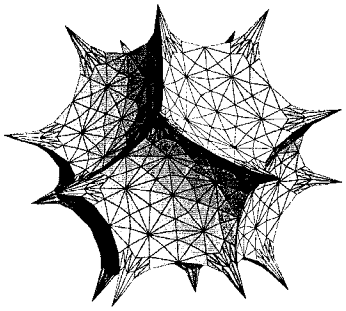

A hyperbolic version of the Euclidean DODECAHEDRON.

see also HYPERBOLIC CUBE, HYPERBOLIC OCTAHEDRON, HYPERBOLIC TETRAHEDRON

References

Rivin, I. "Hyperbolic Polyhedron Graphics." `http://www.mathsource.com/cgi-bin/MathSource/Applications/Graphics/3D/0201-788`.

Hyperbolic Fixed Point (Differential Equations)

A FIXED POINT for which the STABILITY MATRIX has EIGENVALUES $\lambda_1 < 0 < \lambda_2$, also called a SADDLE POINT.

see also ELLIPTIC FIXED POINT (DIFFERENTIAL EQUATIONS), FIXED POINT, STABLE IMPROPER NODE, STABLE SPIRAL POINT, STABLE STAR, UNSTABLE IMPROPER NODE, UNSTABLE NODE, UNSTABLE SPIRAL POINT, UNSTABLE STAR

References

Tabor, M. "Classification of Fixed Points." §1.4.b in *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, pp. 22–25, 1989.

Hyperbolic Fixed Point (Map)

A FIXED POINT of a LINEAR TRANSFORMATION (MAP) for which the rescaled variables satisfy

$$(\delta - \alpha)^2 + 4\beta\gamma > 0.$$

see also ELLIPTIC FIXED POINT (MAP), LINEAR TRANSFORMATION, PARABOLIC FIXED POINT

Hyperbolic Functions

The hyperbolic functions sinh, cosh, tanh, csch, sech, coth (Hyperbolic Sine, Hyperbolic Cosine, etc.) share many properties with the corresponding Circular Functions. The hyperbolic functions arise in many problems of mathematics and mathematical physics in which integrals involving $\sqrt{1+x^2}$ arise (whereas the Circular Functions involve $\sqrt{1-x^2}$).

For instance, the Hyperbolic Sine arises in the gravitational potential of a cylinder and the calculation of the Roche limit. The Hyperbolic Cosine function is the shape of a hanging cable (the so-called Catenary). The Hyperbolic Tangent arises in the calculation of magnetic moment and rapidity of special relativity. All three appear in the Schwarzschild metric using external isotropic Kruskal coordinates in general relativity. The Hyperbolic Secant arises in the profile of a laminar jet. The Hyperbolic Cotangent arises in the Langevin function for magnetic polarization.

The hyperbolic functions are defined by

$$\sinh z \equiv \frac{e^z - e^{-z}}{2} = -\sinh(-z) \tag{1}$$

$$\cosh z \equiv \frac{e^z + e^{-z}}{2} = \cosh(-z) \tag{2}$$

$$\tanh z \equiv \frac{e^z - e^{-z}}{e^z + e^{-z}} = \frac{e^{2z} - 1}{e^{2z} + 1} \tag{3}$$

$$\operatorname{csch} z \equiv \frac{2}{e^z - e^{-z}} \tag{4}$$

$$\operatorname{sech} z \equiv \frac{2}{e^z + e^{-z}} \tag{5}$$

$$\coth z \equiv \frac{e^z + e^{-z}}{e^z - e^{-z}} = \frac{e^{2z} + 1}{e^{2z} - 1}. \tag{6}$$

For purely Imaginary arguments,

$$\sinh(iz) = i \sin z \tag{7}$$

$$\cosh(iz) = \cos z. \tag{8}$$

The hyperbolic functions satisfy many identities anomalous to the trigonometric identities (which can be inferred using Osborne's Rule) such as

$$\cosh^2 x - \sinh^2 x = 1 \tag{9}$$

$$\cosh x + \sinh x = e^x \tag{10}$$

$$\cosh x - \sinh x = e^{-x}. \tag{11}$$

See also Beyer (1987, p. 168). Some half-angle Formulas are

$$\tanh\left(\frac{z}{2}\right) = \frac{\sinh x + i \sin y}{\cosh x + \cos y} \tag{12}$$

$$\coth\left(\frac{z}{2}\right) = \frac{\sinh x - i \sin y}{\cosh x - \cos y}. \tag{13}$$

Some double-angle Formulas are

$$\sinh(2x) = 2 \sinh x \cosh x \tag{14}$$

$$\cosh(2x) = 2\cosh^2 x - 1 = 1 + 2\sinh^2 x. \tag{15}$$

Identities for Complex arguments include

$$\sinh(x+iy) = \sinh x \cos y + i \cosh x \sin y \tag{16}$$

$$\cosh(x+iy) = \cosh x \cos y + i \sinh x \sin y. \tag{17}$$

The Absolute Squares for Complex arguments are

$$|\sinh(z)|^2 = \sinh^2 x + \sin^2 y \tag{18}$$

$$|\cosh(z)|^2 = \sinh^2 x + \cos^2 y. \tag{19}$$

Integrals involving hyperbolic functions include

$$\begin{aligned} \int \frac{dx}{x\sqrt{a+bx}} &= \ln\left|\frac{\sqrt{a+bx}-\sqrt{a}}{\sqrt{a+bx}+\sqrt{a}}\right| \\ &= \ln\left|\frac{(\sqrt{a+bx}-\sqrt{a})^2}{(a+bx)-a}\right| \\ &= \ln\left|\frac{(a+bx)-2\sqrt{a(a+bx)}+a}{bx}\right|. \end{aligned} \tag{20}$$

If $b > 0$, then

$$\begin{aligned} \int \frac{dx}{x\sqrt{a+bx}} &= \ln\left|\frac{2a+bx-2\sqrt{a(a+bx)}}{bx}\right| \\ &= \ln\left|\left(\frac{2a}{bx}+1\right)-2\sqrt{\frac{a}{bx}\left(\frac{a}{bx}+1\right)}\right|. \end{aligned} \tag{21}$$

Let $z \equiv 2a/bx + 1$, and $a/bx = (z-1)/2$ and

$$\begin{aligned} \int \frac{dx}{x\sqrt{a+bx}} &= \ln\left[z - 2\sqrt{\tfrac{1}{2}(z-1)\tfrac{1}{2}(z+1)}\right] \\ &= \ln\left[z - \sqrt{(z-1)(z+1)}\right] \\ &= \ln\left(z - \sqrt{z^2-1}\right) = \cosh^{-1}(z) \\ &= \cosh^{-1}\left(1 + \frac{2a}{bx}\right) \\ &= 2\tanh\left(-\sqrt{\frac{a}{a+bx}}\right). \end{aligned} \tag{22}$$

see also Hyperbolic Cosecant, Hyperbolic Cosine, Hyperbolic Cotangent, Generalized Hyperbolic Functions, Hyperbolic Inverse Functions, Hyperbolic Secant, Hyperbolic Sine, Hyperbolic Tangent, Hyperbolic Inverse Functions, Osborne's Rule

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Hyperbolic Functions." §4.5 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 83–86, 1972.

Beyer, W. H. "Hyperbolic Function." *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 168–186, 1987.

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 126–131, 1967.

Yates, R. C. "Hyperbolic Functions." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 113–118, 1952.

Hyperbolic Geometry

A NON-EUCLIDEAN GEOMETRY, also called LOBACHEVSKY-BOLYAI-GAUSS GEOMETRY, having constant SECTIONAL CURVATURE −1. This GEOMETRY satisfies all of EUCLID'S POSTULATES *except* the PARALLEL POSTULATE, which is modified to read: For any infinite straight LINE L and any POINT P not on it, there are *many other* infinitely extending straight LINES that pass through P and which do not intersect L.

In hyperbolic geometry, the sum of ANGLES of a TRIANGLE is less than 180°, and TRIANGLES with the same angles have the same areas. Furthermore, not all TRIANGLES have the same ANGLE sum (c.f. the AAA THEOREM for TRIANGLES in Euclidean 2-space). The best-known example of a hyperbolic space are SPHERES in Lorentzian 4-space. The POINCARÉ HYPERBOLIC DISK is a hyperbolic 2-space. Hyperbolic geometry is well understood in 2-D, but not in 3-D.

Geometric models of hyperbolic geometry include the KLEIN-BELTRAMI MODEL, which consists of an OPEN DISK in the Euclidean plane whose open chords correspond to hyperbolic lines. A 2-D model is the POINCARÉ HYPERBOLIC DISK. Felix Klein constructed an analytic hyperbolic geometry in 1870 in which a POINT is represented by a pair of REAL NUMBERS (x_1, x_2) with

$$x_1{}^2 + x_2{}^2 < 1$$

(i.e., points of an OPEN DISK in the COMPLEX PLANE) and the distance between two points is given by

$$d(x, X) = a\cosh^{-1}\left[\frac{1 - x_1X_1 - x_2X_2}{\sqrt{1 - x_1{}^2 - x_2{}^2}\,\sqrt{1 - X_1{}^2 - X_2{}^2}}\right].$$

The geometry generated by this formula satisfies all of EUCLID'S POSTULATES except the fifth. The METRIC of this geometry is given by the CAYLEY-KLEIN-HILBERT METRIC,

$$g_{11} = \frac{a^2(1 - x_2{}^2)}{(1 - x_1{}^2 - x_2{}^2)^2}$$

$$g_{12} = \frac{a^2x_1x_2}{(1 - x_1{}^2 - x_2{}^2)^2}$$

$$g_{22} = \frac{a^2(1 - x_1{}^2)}{(1 - x_1{}^2 - x_2{}^2)^2}.$$

Hilbert extended the definition to general bounded sets in a EUCLIDEAN SPACE.

see also ELLIPTIC GEOMETRY, EUCLIDEAN GEOMETRY, HYPERBOLIC METRIC, KLEIN-BELTRAMI MODEL, NON-EUCLIDEAN GEOMETRY, SCHWARZ-PICK LEMMA

References

Dunham, W. *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 57–60, 1990.

Eppstein, D. "Hyperbolic Geometry." `http://www.ics.uci.edu/~eppstein/junkyard/hyper.html`.

Stillwell, J. *Sources of Hyperbolic Geometry.* Providence, RI: Amer. Math. Soc., 1996.

Hyperbolic Inverse Functions

$$\sinh^{-1}\left(\frac{a}{b}\right) = \ln\left(a + \sqrt{a^2 + b^2}\right) \quad (1)$$

$$\cosh^{-1} z = \ln\left(z \pm \sqrt{z^2 - 1}\right) \quad (2)$$

$$\tanh^{-1}\left(\frac{a}{b}\right) = \tfrac{1}{2}\ln\left(\frac{b + a}{b - a}\right) \quad (3)$$

$$\operatorname{csch}^{-1} z = \ln\left(1 \pm \sqrt{1 + z^2}\right) \quad (4)$$

$$\operatorname{sech}^{-1} z = \ln\left(\frac{1 \pm \sqrt{1 - z^2}}{z}\right) \quad (5)$$

$$\coth^{-1} z = \tfrac{1}{2}\ln\left(\frac{z + 1}{z - 1}\right). \quad (6)$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Hyperbolic Functions." §4.6 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 86–89, 1972.

Hyperbolic Knot

A hyperbolic knot is a KNOT that has a complement that can be given a metric of constant curvature −1. The only KNOTS which are not hyperbolic are TORUS KNOTS and SATELLITE KNOTS (including COMPOSITE KNOTS), as proved by Thurston in 1978. Therefore, all but six of the PRIME KNOTS with 10 or fewer crossings are hyperbolic. The exceptions with nine or fewer crossings are 03_{001} (the(3, 2)-TORUS KNOT), 05_{001}, 07_{001}, 08_{019} (the (4, 3)-TORUS KNOT), and 09_{001}.

Almost all hyperbolic knots can be distinguished by their hyperbolic volumes (exceptions being 05_{002} and a certain 12-crossing knot; see Adams 1994, p. 124). It has been conjectured that the smallest hyperbolic volume is 2.0298..., that of the FIGURE-OF-EIGHT KNOT.

MUTANT KNOTS have the same hyperbolic knot volume.

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 119–127, 1994.

Adams, C.; Hildebrand, M.; and Weeks, J. "Hyperbolic Invariants of Knots and Links." *Trans. Amer. Math. Soc.* **326**, 1–56, 1991.

Weisstein, E. W. "Knots and Links." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Knots.m`.

Hyperbolic Lemniscate Function

By analogy with the LEMNISCATE FUNCTIONS, hyperbolic lemniscate functions can also be defined

$$\operatorname{arcsinhlemn} x \equiv \int_0^x (1+t^4)^{1/2}\,dt \tag{1}$$

$$\operatorname{arccoshlemn} x \equiv \int_x^1 (1+t^4)^{1/2}\,dt. \tag{2}$$

Let $0 \leq \theta \leq \pi/2$ and $0 \leq v \leq 1$, and write

$$\frac{\theta\mu}{2} = \int_0^v \frac{dt}{\sqrt{1+t^2}}, \tag{3}$$

where μ is the constant obtained by setting $\theta = \pi/2$ and $v = 1$. Then

$$\mu = \frac{2}{\pi} K\left(\frac{1}{\sqrt{2}}\right), \tag{4}$$

where $K(k)$ is a complete ELLIPTIC INTEGRAL OF THE FIRST KIND, and Ramanujan showed

$$2\tan^{-1} v = \theta + \sum_{n=1}^{\infty} \frac{\sin(2n\theta)}{n\cosh(n\pi)}, \tag{5}$$

$$\tfrac{1}{8}\pi - \tfrac{1}{2}\tan^{-1}(v^2) = \sum_{n=0}^{\infty} \frac{(-1)^n \cos[(2n+1)\theta]}{(2n+1)\cosh[\frac{1}{2}(2n+1)\pi]} \tag{6}$$

and

$$\ln\left(\frac{1+v}{1-v}\right) = \ln[\tan(\tfrac{1}{4}\pi + \tfrac{1}{2}\theta)] + 4\sum_{n=0}^{\infty} \frac{(-1)^n \sin[(2n+1)\theta]}{(2n+1)[e^{(2n+1)\pi} - 1]} \tag{7}$$

(Berndt 1994).

see also LEMNISCATE FUNCTION

References

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 255–258, 1994.

Hyperbolic Map

A linear MAP $\mathbb{R}^n$ is hyperbolic if none of its EIGENVALUES have modulus 1. This means that $\mathbb{R}^n$ can be written as a direct sum of two A-invariant SUBSPACES E^s and E^u (where s stands for stable and u for unstable). This means that there exist constants $C > 0$ and $0 < \lambda < 1$ such that

$$||A^n v|| \leq C\lambda^n ||v|| \quad \text{if } v \in E^s$$

$$||A^{-n} v|| \leq C\lambda^n ||v|| \quad \text{if } v \in E^u$$

for $n = 0, 1, \ldots$.

see also PESIN THEORY

Hyperbolic Metric

The METRIC for the POINCARÉ HYPERBOLIC DISK, a model for HYPERBOLIC GEOMETRY. The hyperbolic metric is invariant under conformal maps of the disk onto itself.

see also HYPERBOLIC GEOMETRY, POINCARÉ HYPERBOLIC DISK

References

Bear, H. S. "Part Metric and Hyperbolic Metric." *Amer. Math. Monthly* **98**, 109–123, 1991.

Hyperbolic Octahedron

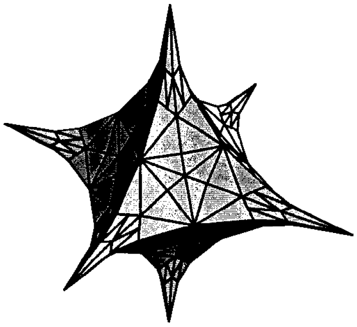

A hyperbolic version of the Euclidean OCTAHEDRON, which is a special case of the ASTROIDAL ELLIPSOID with $a = b = c = 1$. It is given by the parametric equations

$$x = (\cos u \cos v)^3$$
$$y = (\sin u \cos v)^3$$
$$z = \sin^3 v$$

for $u \in [-\pi/2, \pi/2]$ and $v \in [-\pi, \pi]$.

see also ASTROIDAL ELLIPSOID, HYPERBOLIC CUBE, HYPERBOLIC DODECAHEDRON, HYPERBOLIC TETRAHEDRON

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 305–306, 1993.

Nordstrand, T. "Astroidal Ellipsoid." `http://www.uib.no/people/nfytn/asttxt.htm`.

Rivin, I. "Hyperbolic Polyhedron Graphics." `http://www.mathsource.com/cgi-bin/MathSource/Applications/Graphics/3D/0201-788`.

Hyperbolic Paraboloid

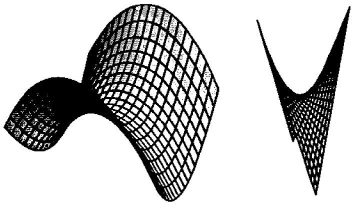

A QUADRATIC SURFACE given by the Cartesian equation

$$z = \frac{y^2}{b^2} - \frac{x^2}{a^2} \tag{1}$$

(left figure). This form has parametric equations

$$x(u,v) = a(u+v) \tag{2}$$
$$y(u,v) = \pm bv \tag{3}$$
$$z(u,v) = u^2 + 2uv \tag{4}$$

(Gray 1993, p. 336). An alternative form is

$$z = xy \tag{5}$$

(right figure; Fischer 1986), which has parametric equations

$$x(u,v) = u \tag{6}$$
$$y(u,v) = v \tag{7}$$
$$z(u,v) = uv. \tag{8}$$

see also ELLIPTIC PARABOLOID, PARABOLOID, RULED SURFACE

References

Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, pp. 3–4, 1986.

Fischer, G. (Ed.). Plates 7–9 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 8–10, 1986.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 211–212 and 336, 1993.

Meyer, W. "Spezielle algebraische Flächen." *Encylopädie der Math. Wiss. III*, **22B**, 1439–1779.

Salmon, G. *Analytic Geometry of Three Dimensions.* New York: Chelsea, 1979.

Hyperbolic Partial Differential Equation

A PARTIAL DIFFERENTIAL EQUATION of second-order, i.e., one of the form

$$Au_{xx} + 2Bu_{xy} + Cu_{yy} + Du_x + Eu_y + F = 0, \tag{1}$$

is called hyperbolic if the MATRIX

$$\mathsf{Z} \equiv \begin{bmatrix} A & B \\ B & C \end{bmatrix} \tag{2}$$

satisfies $\det(\mathsf{Z}) < 0$. The WAVE EQUATION is an example of a hyperbolic partial differential equation. Initial-boundary conditions are used to give

$$u(x,y,t) = g(x,y,t) \quad \text{for } x \in \partial\Omega, t > 0 \tag{3}$$
$$u(x,y,0) = v_0(x,y) \quad \text{in } \Omega \tag{4}$$
$$u_t(x,y,0) = v_1(x,y) \quad \text{in } \Omega, \tag{5}$$

where

$$u_{xy} = f(u_x, u_t, x, y) \tag{6}$$

holds in Ω.

see also ELLIPTIC PARTIAL DIFFERENTIAL EQUATION, PARABOLIC PARTIAL DIFFERENTIAL EQUATION, PARTIAL DIFFERENTIAL EQUATION

Hyperbolic Plane

In the hyperbolic plane $\mathbb{H}^2$, a pair of LINES can be PARALLEL (diverging from one another in one direction and intersecting at an IDEAL POINT at infinity in the other), can intersect, or can be HYPERPARALLEL (diverge from each other in both directions).

see also EUCLIDEAN PLANE, RIGID MOTION

Hyperbolic Point

A point $\mathbf{p}$ on a REGULAR SURFACE $M \in \mathbb{R}^3$ is said to be hyperbolic if the GAUSSIAN CURVATURE $K(\mathbf{p}) < 0$ or equivalently, the PRINCIPAL CURVATURES κ_1 and κ_2, have opposite signs.

see also ANTICLASTIC, ELLIPTIC POINT, GAUSSIAN CURVATURE, HYPERBOLIC FIXED POINT (DIFFERENTIAL EQUATIONS), HYPERBOLIC FIXED POINT (MAP), PARABOLIC POINT, PLANAR POINT, SYNCLASTIC

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 280, 1993.

Hyperbolic Polyhedron

A POLYHEDRON in a HYPERBOLIC GEOMETRY.

see HYPERBOLIC CUBE, HYPERBOLIC DODECAHEDRON, HYPERBOLIC OCTAHEDRON, HYPERBOLIC TETRAHEDRON

Hyperbolic Rotation

Also known as the LORENTZ TRANSFORMATION or PROCRUSTIAN STRETCH. Leaves each branch of the HYPERBOLA $x'y' = xy$ invariant and transforms CIRCLES into ELLIPSES with the same AREA.

$$x' = \mu^{-1}x$$
$$y' = \mu y.$$

Hyperbolic Rotation (Crossed)

Exchanges branches of the HYPERBOLA $x'y' = xy$.

$$x' = \mu^{-1}x$$
$$y' = -\mu y.$$

Hyperbolic Secant

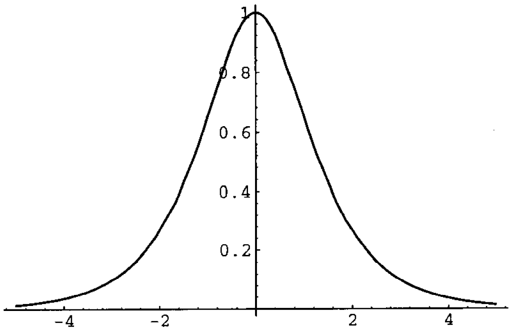

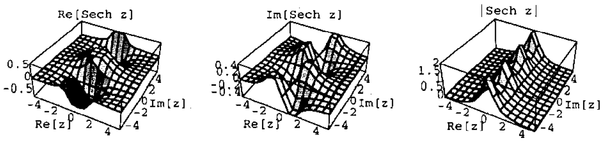

The hyperbolic secant is defined as

$$\operatorname{sech} x \equiv \frac{1}{\cosh x} = \frac{2}{e^{x} + e^{-x}}.$$

It has a MAXIMUM at $x = 0$ and inflection points at $x = \pm \operatorname{sech}^{-1}(1/\sqrt{2}) \approx 0.881374$.

see also BENSON'S FORMULA, CATENARY, CATENOID, EULER NUMBER, HYPERBOLIC COSINE, OBLATE SPHEROIDAL COORDINATES, PSEUDOSPHERE, SECANT, SURFACE OF REVOLUTION, TRACTRIX, TRACTROID

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Hyperbolic Functions." §4.5 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 83–86, 1972.

Spanier, J. and Oldham, K. B. "The Hyperbolic Secant sech(x) and Cosecant csch(x) Functions." Ch. 29 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 273–278, 1987.

Hyperbolic Sine

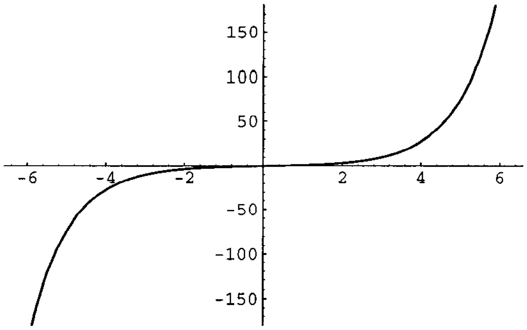

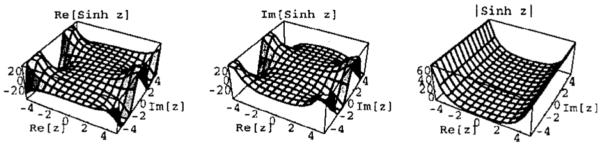

The hyperbolic sine is defined as

$$\sinh x \equiv \tfrac{1}{2}(e^{x} - e^{-x}).$$

see also BETA FUNCTION (EXPONENTIAL), BIPOLAR COORDINATES, BIPOLAR CYLINDRICAL COORDINATES, BISPHERICAL COORDINATES, CATENARY, CATENOID, CONICAL FUNCTION, CUBIC EQUATION, DE MOIVRE'S IDENTITY, DIXON-FERRAR FORMULA, ELLIPTIC CYLINDRICAL COORDINATES, ELSASSER FUNCTION, FIBONACCI HYPERBOLIC COSINE, FIBONACCI HYPERBOLIC SINE, GUDERMANNIAN FUNCTION, HELICOID, HELMHOLTZ DIFFERENTIAL EQUATION—ELLIPTIC CYLINDRICAL COORDINATES, HYPERBOLIC COSECANT, LAPLACE'S EQUATION—BISPHERICAL COORDINATES, LAPLACE'S EQUATION—TOROIDAL COORDINATES, LEBESGUE CONSTANTS (FOURIER SERIES), LORENTZ GROUP, MERCATOR PROJECTION, MILLER CYLINDRICAL PROJECTION, MODIFIED BESSEL FUNCTION OF THE SECOND KIND, MODIFIED SPHERICAL BESSEL FUNCTION, MODIFIED STRUVE FUNCTION, NICHOLSON'S FORMULA, OBLATE SPHEROIDAL COORDINATES, PARABOLA INVOLUTE, PARTITION FUNCTION P, POINSOT'S SPIRALS, PROLATE SPHEROIDAL COORDINATES, RAMANUJAN'S TAU FUNCTION, SCHLÄFLI'S FORMULA, SHI, SINE, SINE-GORDON EQUATION, SURFACE OF REVOLUTION, TOROIDAL COORDINATES, TOROIDAL FUNCTION, TRACTRIX, WATSON'S FORMULA

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Hyperbolic Functions." §4.5 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 83–86, 1972.

Spanier, J. and Oldham, K. B. "The Hyperbolic Sine sinh(x) and Cosine cosh(x) Functions." Ch. 28 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 263–271, 1987.

Hyperbolic Space

see HYPERBOLIC GEOMETRY

Hyperbolic Spiral

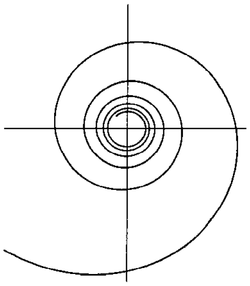

An ARCHIMEDEAN SPIRAL with POLAR equation

$$r = \frac{a}{\theta}.$$

The hyperbolic spiral originated with Pierre Varignon in 1704 and was studied by Johann Bernoulli between 1710 and 1713, as well as by Cotes in 1722 (MacTutor Archive).

see also ARCHIMEDEAN SPIRAL, SPIRAL

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 69–70, 1993.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 186 and 188, 1972.

Lockwood, E. H. *A Book of Curves.* Cambridge, England: Cambridge University Press, p. 175, 1967.

MacTutor History of Mathematics Archive. "Hyperbolic Spiral." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Hyperbolic.html`.

Hyperbolic Spiral Inverse Curve

Taking the pole as the INVERSION CENTER, the HYPERBOLIC SPIRAL inverts to ARCHIMEDES' SPIRAL

$$r = a\theta.$$

Hyperbolic Spiral Roulette

The ROULETTE of the pole of a HYPERBOLIC SPIRAL rolling on a straight line is a TRACTRIX.

Hyperbolic Substitution

A substitution which can be used to transform integrals involving square roots into a more tractable form.

Form	Substitution
$\sqrt{x^2+a^2}$	$x = a\sinh u$
$\sqrt{x^2-a^2}$	$x = a\cosh u$

see also TRIGONOMETRIC SUBSTITUTION

Hyperbolic Tangent

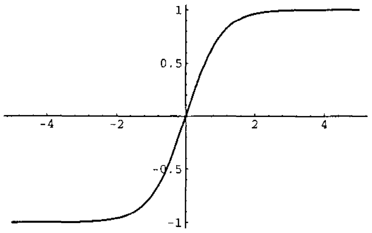

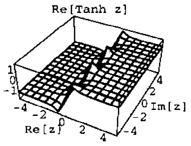

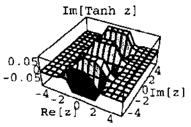

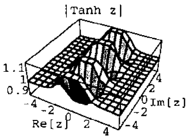

By way of analogy with the usual TANGENT

$$\tan x \equiv \frac{\sin x}{\cos x},$$

the hyperbolic tangent is defined as

$$\tanh x \equiv \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x}-1}{e^{2x}+1},$$

where $\sinh x$ is the HYPERBOLIC SINE and $\cosh x$ is the HYPERBOLIC COSINE. The hyperbolic tangent can be written using a CONTINUED FRACTION as

$$\tanh x = \cfrac{x}{1 + \cfrac{x^2}{3 + \cfrac{x^2}{5 + \ldots}}}.$$

see also BERNOULLI NUMBER, CATENARY, CORRELATION COEFFICIENT—GAUSSIAN BIVARIATE DISTRIBUTION, FIBONACCI HYPERBOLIC TANGENT, FISHER'S z'-TRANSFORMATION, HYPERBOLIC COTANGENT, LORENTZ GROUP, MERCATOR PROJECTION, OBLATE SPHEROIDAL COORDINATES, PSEUDOSPHERE, SURFACE OF REVOLUTION, TANGENT, TRACTRIX, TRACTROID

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Hyperbolic Functions." §4.5 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 83–86, 1972.

Spanier, J. and Oldham, K. B. "The Hyperbolic Tangent $\tanh(x)$ and Cotangent $\coth(x)$ Functions." Ch. 30 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 279–284, 1987.

Hyperbolic Tetrahedron

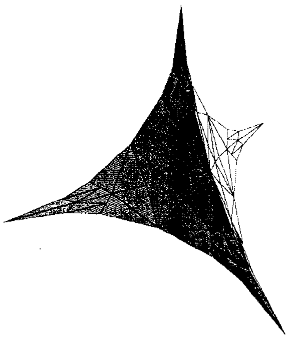

A hyperbolic version of the Euclidean TETRAHEDRON.

see also HYPERBOLIC CUBE, HYPERBOLIC DODECAHEDRON, HYPERBOLIC OCTAHEDRON

References

Rivin, I. "Hyperbolic Polyhedron Graphics." `http://www.mathsource.com/cgi-bin/MathSource/Applications/Graphics/3D/0201-788`.

Hyperbolic Umbilic Catastrophe

A CATASTROPHE which can occur for three control factors and two behavior axes.

see also ELLIPTIC UMBILIC CATASTROPHE

Hyperboloid

A QUADRATIC SURFACE which may be one- or two-sheeted.

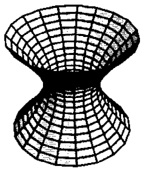

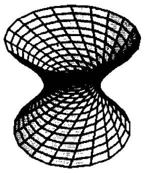

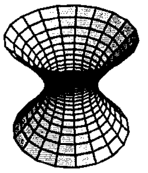

The one-sheeted circular hyperboloid is a doubly RULED SURFACE. When oriented along the z-AXIS, the one-sheeted circular hyperboloid has CARTESIAN COORDINATES equation

$$\frac{x^2}{a^2} + \frac{y^2}{a^2} - \frac{z^2}{c^2} = 1, \tag{1}$$

and parametric equation

$$x = a\sqrt{1+u^2}\cos v \qquad (2)$$

$$y = a\sqrt{1+u^2}\sin v \qquad (3)$$

$$z = cu \qquad (4)$$

for $v \in [0, 2\pi)$ (left figure). Other parameterizations include

$$x(u,v) = a(\cos u \mp v \sin u) \qquad (5)$$

$$y(u,v) = a(\sin u \pm v \cos u) \qquad (6)$$

$$z(u,v) = \pm cv, \qquad (7)$$

(middle figure), or

$$x(u,v) = a\cosh v \cos u \qquad (8)$$

$$y(u,v) = a\cosh v \sin u \qquad (9)$$

$$z(u,v) = c\sinh v \qquad (10)$$

(right figure). An obvious generalization gives the one-sheeted ELLIPTIC HYPERBOLOID.

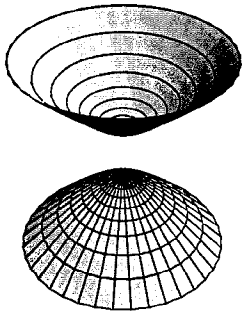

A two-sheeted circular hyperboloid oriented along the z-AXIS has CARTESIAN COORDINATES equation

$$\frac{x^2}{a^2} + \frac{y^2}{a^2} - \frac{z^2}{c^2} = -1. \qquad (11)$$

The parametric equations are

$$x = a\sinh u \cos v \qquad (12)$$

$$y = a\sinh u \sin v \qquad (13)$$

$$z = \pm c\cosh u \qquad (14)$$

for $v \in [0, 2\pi)$. Note that the plus and minus signs in z correspond to the upper and lower sheets. The two-sheeted circular hyperboloid oriented along the x-AXIS has Cartesian equation

$$\frac{x^2}{a^2} - \frac{y^2}{a^2} - \frac{z^2}{c^2} = 1 \qquad (15)$$

and parametric equations

$$x = \pm a\cosh u \cosh v \qquad (16)$$

$$y = a\sinh u \cosh v \qquad (17)$$

$$z = c\sinh v \qquad (18)$$

(Gray 1993, p. 313). Again, an obvious generalization gives the two-sheeted ELLIPTIC HYPERBOLOID.

The SUPPORT FUNCTION of the hyperboloid of one sheet

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1 \qquad (19)$$

is

$$h = \left(\frac{x^2}{a^4} + \frac{y^2}{b^4} + \frac{z^2}{c^4}\right)^{-1/2}, \qquad (20)$$

and the GAUSSIAN CURVATURE is

$$K = -\frac{h^4}{a^2b^2c^2}. \qquad (21)$$

The SUPPORT FUNCTION of the hyperboloid of two sheets

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1 \qquad (22)$$

is

$$h = \left(\frac{x^2}{a^4} - \frac{y^2}{b^4} + \frac{z^2}{c^4}\right)^{-1/2}, \qquad (23)$$

and the GAUSSIAN CURVATURE is

$$K = \frac{h^4}{a^2b^2c^2} \qquad (24)$$

(Gray 1993, pp. 296–297).

see also CATENOID, ELLIPSOID, ELLIPTIC HYPERBOLOID, HYPERBOLOID EMBEDDING, PARABOLOID, RULED SURFACE

References

Fischer, G. (Ed.). Plates 67 and 69 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 62 and 64, 1986.

Gray, A. "The Hyperboloid of Revolution." §18.5 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 296–297, 311–314, and 369–370, 1993.

Hyperboloid Embedding

A 4-HYPERBOLOID has NEGATIVE CURVATURE, with

$$R^2 = x^2 + y^2 + z^2 - w^2 \qquad (1)$$

$$2x\frac{dx}{dw} + 2y\frac{dy}{dw} + 2z\frac{dz}{dw} - 2w = 0. \qquad (2)$$

Since

$$\mathbf{r} \equiv x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}, \qquad (3)$$

$$dw = \frac{x\,dx + y\,dy + z\,dz}{w} = \frac{\mathbf{r}\cdot d\mathbf{r}}{\sqrt{r^2 - R^2}}. \qquad (4)$$

To stay on the surface of the HYPERBOLOID,

$$\begin{aligned} ds^2 &= dx^2 + dy^2 + dz^2 - dw^2 \\ &= dx^2 + dy^2 + dz^2 - \frac{r^2\,dr^2}{r^2 - R^2} \\ &= dr^2 + r^2 d\Omega^2 + \frac{dr^2}{1 - \frac{R^2}{r^2}}. \end{aligned} \tag{5}$$

Hypercomplex Number

A number having properties departing from those of the REAL and COMPLEX NUMBERS. The most common examples are BIQUATERNIONS, EXTERIOR ALGEBRAS, GROUP algebras, MATRICES, OCTONIONS, and QUATERNIONS.

References

van der Waerden, B. L. *A History of Algebra from al-Khwarizmi to Emmy Noether.* New York: Springer-Verlag, pp. 177–217, 1985.

Hypercube

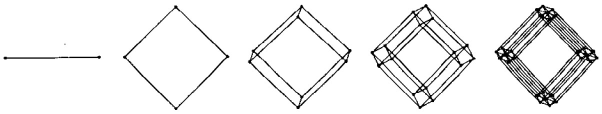

The generalization of a 3-CUBE to n-D, also called a MEASURE POLYTOPE. It is a regular POLYTOPE with mutually PERPENDICULAR sides, and is therefore an ORTHOTOPE. It is denoted γ_n and has SCHLÄFLI SYMBOL $\{4, \underbrace{3,3}_{n-2}\}$. The number of k-cubes contained in an n-cube can be found from the COEFFICIENTS of $(2k+1)^n$.

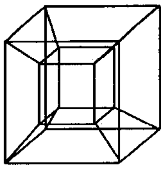

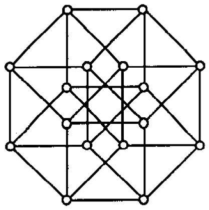

The 1-hypercube is a LINE SEGMENT, the 2-hypercube is the SQUARE, and the 3-hypercube is the CUBE. The hypercube in $\mathbb{R}^4$, called a TESSERACT, has the SCHLÄFLI SYMBOL $\{4,3,3\}$ and VERTICES $(\pm1, \pm1, \pm1, \pm1)$. The above figures show two visualizations of the TESSERACT. The figure on the left is a projection of the TESSERACT in 3-space (Gardner 1977), and the figure on the right is the GRAPH of the TESSERACT symmetrically projected into the PLANE (Coxeter 1973). A TESSERACT has 16 VERTICES, 32 EDGES, four SQUARES, and eight CUBES.

see also CROSS POLYTOPE, CUBE, HYPERSPHERE, ORTHOTOPE, PARALLELEPIPED, POLYTOPE, SIMPLEX, TESSERACT

References

Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, p. 123, 1973.

Gardner, M. "Hypercubes." Ch. 4 in *Mathematical Carnival: A New Round-Up of Tantalizers and Puzzles from Scientific American.* New York: Vintage Books, 1977.

Geometry Center. "The Tesseract (or Hypercube)." `http://www.geom.umn.edu/docs/outreach/4-cube/`.

Pappas, T. "How Many Dimensions are There?" *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 204–205, 1989.

Hyperdeterminant

A technically defined extension of the ordinary DETERMINANT to "higher dimensional" HYPERMATRICES. Cayley (1845) originally coined the term, but subsequently used it to refer to an ALGEBRAIC INVARIANT of a multilinear form. The hyperdeterminant of the $2\times2\times2$ HYPERMATRIX $A = a_{ijk}$ (for $i, j, k = 0, 1$) is given by

$$\begin{aligned} \det(A) = {} & ({a_{000}}^2 {a_{111}}^2 + {a_{001}}^2 {a_{110}}^2 + {a_{010}}^2 {a_{101}}^2 + {a_{011}}^2 {a_{100}}^2) \\ & - 2(a_{000}a_{001}a_{110}a_{111} + a_{000}a_{010}a_{101}a_{111} + a_{000}a_{011}a_{100}a_{111} \\ & + a_{001}a_{010}a_{101}a_{110} + a_{001}a_{011}a_{110}a_{100} + a_{010}a_{011}a_{101}a_{100}) \\ & + 4(a_{000}a_{011}a_{101}a_{110} + a_{001}a_{010}a_{100}a_{111}). \end{aligned}$$

The above hyperdeterminant vanishes IFF the following system of equations in six unknowns has a nontrivial solution,

$$\begin{aligned} a_{000}x_0y_0 + a_{010}x_0y_1 + a_{100}x_1y_0 + a_{110}x_1y_1 &= 0 \\ a_{001}x_0y_0 + a_{011}x_0y_1 + a_{101}x_1y_0 + a_{111}x_1y_1 &= 0 \\ a_{000}x_0z_0 + a_{001}x_0z_1 + a_{100}x_1z_0 + a_{101}x_1z_1 &= 0 \\ a_{010}x_0z_0 + a_{011}x_0z_1 + a_{110}x_1z_0 + a_{111}x_1z_1 &= 0 \\ a_{000}y_0z_0 + a_{001}y_0z_1 + a_{010}y_1z_0 + a_{011}y_1z_1 &= 0 \\ a_{100}y_0z_0 + a_{101}y_0z_1 + a_{110}y_1z_0 + a_{111}y_1z_1 &= 0. \end{aligned}$$

see also DETERMINANT, HYPERMATRIX

References

Cayley, A. "On the Theory of Linear Transformations." *Cambridge Math. J.* **4**, 193–209, 1845.

Gel'fand, I. M.; Kapranov, M. M.; and Zelevinsky, A. V. "Hyperdeterminants." *Adv. Math.* **96**, 226–263, 1992.

Schläfli, L. "Über die Resultante eine Systemes mehrerer algebraischer Gleichungen." *Denkschr. Kaiserl. Akad. Wiss., Math.-Naturwiss. Klasse* **4**, 1852.

Hyperellipse

$$y^{n/m} + c\left|\frac{x}{a}\right|^{n/m} - c = 0,$$

with $n/m > 2$. If $n/m < 2$, the curve is a HYPOELLIPSE.

see also ELLIPSE, HYPOELLIPSE, SUPERELLIPSE

References

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 82, 1993.

Hyperelliptic Function

see ABELIAN FUNCTION

Hyperelliptic Integral

see ABELIAN INTEGRAL

Hyperfactorial

The function defined by

$$H(n) \equiv K(n+1) \equiv 1^1 2^2 3^3 \cdots n^n,$$

where K is the K-FUNCTION and the first few values for $n = 1, 2, \ldots$ are 1, 4, 108, 27648, 86400000, 4031078400000, 3319766398771200000, ... (Sloane's A002109), and these numbers are called hyperfactorials by Sloane and Plouffe (1995).

see also G-FUNCTION, GLAISHER-KINKELIN CONSTANT, K-FUNCTION

References

Sloane, N. J. A. Sequence A002109/M3706 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Hypergeometric Differential Equation

$$x(x-1)\frac{d^2y}{dx^2} + [(1+\alpha+\beta)x - \gamma]\frac{dy}{dx} + \alpha\beta y = 0.$$

It has REGULAR SINGULAR POINTS at 0, 1, and ∞. Every ORDINARY DIFFERENTIAL EQUATION of second-order with at most three REGULAR SINGULAR POINTS can be transformed into the hypergeometric differential equation.

see also CONFLUENT HYPERGEOMETRIC DIFFERENTIAL EQUATION, CONFLUENT HYPERGEOMETRIC FUNCTION, HYPERGEOMETRIC FUNCTION

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 542–543, 1953.

Hypergeometric Distribution

Let there be n ways for a successful and m ways for an unsuccessful trial out of a total of $n+m$ possibilities. Take N samples and let x_i equal 1 if selection i is successful and 0 if it is not. Let x be the total number of successful selections,

$$x \equiv \sum_{i=1}^{N} x_i. \tag{1}$$

The probability of i successful selections is then

$$\begin{aligned} P(x=i) &= \\ &\frac{[\#\text{ ways for } i \text{ successes}][\#\text{ ways for } N-i \text{ unsuccesses}]}{[\text{total number of ways to select}]} \\ &= \frac{\binom{n}{i}\binom{m}{N-i}}{\binom{n+m}{N}} = \frac{\frac{n!}{i!(n-i!)}\frac{m!}{(m+i-N)!(N-i)!}}{\frac{(n+m)!}{N!(N-n-m)!}} \\ &= \frac{n!m!N!(N-m-n)!}{i!(n-i)!(m+i-N)!(N-i)!(n+m)!}. \end{aligned} \tag{2}$$

The ith selection has an equal likelihood of being in any trial, so the fraction of acceptable selections p is

$$p \equiv \frac{n}{n+m} \tag{3}$$

$$P(x_i = 1) = \frac{n}{n+m} \equiv p. \tag{4}$$

The expectation value of x is

$$\begin{aligned} \mu \equiv \langle x \rangle &= \left\langle \sum_{i=1}^{N} x_i \right\rangle = \sum_{i=1}^{N} \langle x_i \rangle \\ &= \sum_{i=1}^{N} \frac{n}{n+m} = \frac{nN}{n+m} = Np. \end{aligned} \tag{5}$$

The VARIANCE is

$$\operatorname{var}(x) \equiv \sum_{i=1}^{N} \operatorname{var}(x_i) + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \operatorname{cov}(x_i, x_j). \tag{6}$$

Since x_i is a BERNOULLI variable,

$$\begin{aligned} \operatorname{var}(x_i) = p(1-p) &= \frac{n}{n+m}\left(1 - \frac{n}{n+m}\right) \\ &= \frac{n}{n+m}\left(1 - \frac{n}{n+m}\right) \\ &= \frac{n}{n+m}\left(\frac{n+m-n}{n+m}\right) = \frac{nm}{(n+m)^2}, \end{aligned} \tag{7}$$

so

$$\sum_{i=1}^{N} \operatorname{var}(x_i) = \frac{Nnm}{(n+m)^2}. \tag{8}$$

For $i < j$, the COVARIANCE is

$$\operatorname{cov}(x_i, x_j) = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle. \tag{9}$$

The probability that both i and j are successful for $i \neq j$ is

$$\begin{aligned} P(x_i = 1, x_j = 1) &= P(x_i = 1)P(x_j = 1 | x_i = 1) \\ &= \frac{n}{n+m}\frac{n-1}{n+m-1} \\ &= \frac{n(n-1)}{(n+m)(n+m-1)}. \end{aligned} \tag{10}$$

But since x_i and x_j are random BERNOULLI variables (each 0 or 1), their product is also a BERNOULLI variable. In order for $x_i x_j$ to be 1, both x_i and x_j must be 1,

$$\begin{aligned} \langle x_i x_j \rangle = P(x_i x_j = 1) &= P(x_i = 1, x_j = 1) \\ &= \frac{n}{n+m}\frac{n-1}{n+m-1} \\ &= \frac{n(n-1)}{(n+m)(n+m-1)}. \end{aligned} \tag{11}$$

Combining (11) with

$$\langle x_i\rangle\,\langle x_j\rangle = \frac{n}{n+m}\frac{n}{n+m} = \frac{n^2}{(n+m)^2}, \tag{12}$$

gives

$$\begin{aligned}\operatorname{cov}(x_i, x_j) &= \frac{(n+m)(n^2-n) - n^2(n+m-1)}{(n+m)^2(n+m-1)}\\ &= \frac{n^3+mn^2-n^2-mn-n^3-n^2m+n^2}{(n+m)^2(n+m-1)}\\ &= -\frac{mn}{(n+m)^2(n+m-1)}.\end{aligned} \tag{13}$$

There are a total of N^2 terms in a double summation over N. However, $i = j$ for N of these, so there are a total of $N^2 - N = N(N-1)$ terms in the COVARIANCE summation

$$\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} \operatorname{cov}(x_i, x_j) = -\frac{N(N-1)mn}{(n+m)^2(n+m-1)}. \tag{14}$$

Combining equations (6), (8), (11), and (14) gives the VARIANCE

$$\begin{aligned}\operatorname{var}(x) &= \frac{Nmn}{(n+m)^2} - \frac{N(N-1)mn}{(n+m)^2(n+m-1)}\\ &= \frac{Nmn}{(m+n)^2}\left(1 - \frac{N-1}{n+m-1}\right)\\ &= \frac{Nmn}{(n+m)^2}\left(\frac{N+m-1-N+1}{n+m-1}\right)\\ &= \frac{Nmn(n+m-N)}{(n+m)^2(n+m-1)},\end{aligned} \tag{15}$$

so the final result is

$$\langle x\rangle = Np \tag{16}$$

and, since

$$1 - p = \frac{m}{n+m} \tag{17}$$

and

$$np(1-p) = \frac{mn}{(n+m)^2}, \tag{18}$$

we have

$$\begin{aligned}\sigma^2 = \operatorname{var}(x) &= Np(1-p)\left(1 - \frac{N-1}{n+m-1}\right)\\ &= \frac{mnN(m+n-N)}{(m+n)^2(m+n-1)}.\end{aligned} \tag{19}$$

The SKEWNESS is

$$\begin{aligned}\gamma_1 &= \frac{q-p}{\sqrt{npq}}\sqrt{\frac{N-1}{N-m}}\left(\frac{N-2n}{N-2}\right)\\ &= \frac{(m-n)(m+n-2N)}{m+n-2}\sqrt{\frac{m+n-1}{mnN(m+n-N)}},\end{aligned} \tag{20}$$

and the KURTOSIS

$$\gamma_2 = \frac{F(m,n,N)}{mnN(-3+m+n)(-2+m+n)(-m-n+N)}, \tag{21}$$

where

$$\begin{aligned}F(m,n,N) = {} & m^3 - m^5 + 3m^2n - 6m^3n + m^4n + 3mn^2\\ & - 12m^2n^2 + 8m^3n^2 + n^3 - 6mn^3 + 8m^2n^3\\ & + mn^4 - n^5 - 6m^3N + 6m^4N + 18m^2nN\\ & - 6m^3nN + 18mn^2N - 24m^2n^2N - 6n^3N\\ & - 6mn^3N + 6n^4N + 6m^2N^2 - 6m^3N^2\\ & - 24mnN^2 + 12m^2nN^2 + 6n^2N^2\\ & + 12mn^2N^2 - 6n^3N^2.\end{aligned} \tag{22}$$

The GENERATING FUNCTION is

$$\phi(t) = \frac{\binom{m}{N}}{\binom{n+m}{N}}\,{}_2F_1(-N, -n; m-N+1; e^{it}), \tag{23}$$

where ${}_2F_1(a, b; c; z)$ is the HYPERGEOMETRIC FUNCTION.

If the hypergeometric distribution is written

$$h_n(x,s) = \frac{\binom{np}{x}\binom{nq}{s-x}}{\binom{n}{s}}, \tag{24}$$

then

$$\sum_{x=0}^{s} h_n(x,s)u^x = A\,{}_2F_1(-s, -np; nq-s+1; u). \tag{25}$$

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 532–533, 1987.

Spiegel, M. R. *Theory and Problems of Probability and Statistics.* New York: McGraw-Hill, pp. 113–114, 1992.

Hypergeometric Function

A GENERALIZED HYPERGEOMETRIC FUNCTION ${}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; x)$ is a function which can be defined in the form of a HYPERGEOMETRIC SERIES, i.e., a series for which the ratio of successive terms can be written

$$\frac{a_{k+1}}{a_k} = \frac{P(k)}{Q(k)} = \frac{(k+a_1)(k+a_2)\cdots(k+a_p)}{(k+b_1)(k+b_2)\cdots(k+b_q)(k+1)}x. \tag{1}$$

(The factor of $k+1$ in the DENOMINATOR is present for historical reasons of notation.) The function ${}_2F_1(a, b; c; x)$ corresponding to $p = 2$, $q = 1$ is the first hypergeometric function to be studied (and, in general, arises the most frequently in physical problems), and so is frequently known as "the" hypergeometric equation.

To confuse matters even more, the term "hypergeometric function" is less commonly used to mean CLOSED FORM.

The hypergeometric functions are solutions to the HYPERGEOMETRIC DIFFERENTIAL EQUATION, which has a REGULAR SINGULAR POINT at the ORIGIN. To derive the hypergeometric function based on the HYPERGEOMETRIC DIFFERENTIAL EQUATION, plug

$$y = \sum_{n=0}^{\infty} A_n z^n \tag{2}$$

$$y' = \sum_{n=0}^{\infty} nA_n z^{n-1} \tag{3}$$

$$y'' = \sum_{n=0}^{\infty} n(n-1)A_n z^{n-2} \tag{4}$$

into

$$z(1-z)y'' + [c-(a+b+1)a]y' - aby = 0 \tag{5}$$

to obtain

$$\begin{aligned}\sum_{n=0}^{\infty} n(n-1)A_n z^{n-1} - \sum_{n=0}^{\infty} n(n-1)A_n z^n \\ +c\sum_{n=0}^{\infty} nA_n z^{n-1} + (a+b+1)\sum_{n=0}^{\infty} nA_n z^n \\ -ab\sum_{n=0}^{\infty} A_n z^n = 0\end{aligned} \tag{6}$$

$$\begin{aligned}\sum_{n=2}^{\infty} n(n-1)A_n z^{n-1} - \sum_{n=0}^{\infty} n(n-1)A_n z^n \\ +c\sum_{n=1}^{\infty} nA_n z^{n-1} - (a+b+1)\sum_{n=1}^{\infty} nA_n z^n \\ -ab\sum_{n=0}^{\infty} A_n z^n = 0\end{aligned} \tag{7}$$

$$\begin{aligned}\sum_{n=0}^{\infty} (n+1)nA_{n+1} z^n - \sum_{n=0}^{\infty} n(n-1)A_n z^n \\ +c\sum_{n=0}^{\infty} (n+1)A_{n+1} z^n - (a+b+1)\sum_{n=0}^{\infty} nA_n z^n \\ -ab\sum_{n=0}^{\infty} A_n z^n = 0\end{aligned} \tag{8}$$

$$\begin{aligned}\sum_{n=0}^{\infty} [n(n+1)A_{n+1} - n(n-1)A_n + c(n+1)A_{n-1} \\ -(a+b+1)nA_n - abA_n]z^n = 0\end{aligned} \tag{9}$$

$$\begin{aligned}\sum_{n=0}^{\infty} \{(n+1)(n+c)A_{n+1} \\ -[n(n-1+a+b+1)+ab]A_n\}\, z^n = 0\end{aligned} \tag{10}$$

$$\begin{aligned}\sum_{n=0}^{\infty} \{(n+1)(n+c)A_{n+1} \\ -[n^2+(a+b)n+ab]A_n\}z^n = 0,\end{aligned} \tag{11}$$

so

$$A_{n+1} = \frac{(n+a)(n+b)}{(n+1)(n+c)} A_n \tag{12}$$

and

$$y = A_0 \left[1 + \frac{ab}{1!c}z + \frac{a(a+1)b(b+1)}{2!c(c+1)}z^2 + \ldots\right]. \tag{13}$$

This is the regular solution and is denoted

$$\begin{aligned}{}_2F_1(a,b;c;z) &= 1 + \frac{ab}{1!c}z + \frac{a(a+1)b(b+1)}{2!c(c+1)}z^2 + \ldots \\ &= \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!},\end{aligned} \tag{14}$$

where $(a)_n$ are POCHHAMMER SYMBOLS. The hypergeometric series is convergent for REAL $-1 < z < 1$, and for $z = \pm 1$ if $c > a+b$. The complete solution to the HYPERGEOMETRIC DIFFERENTIAL EQUATION is

$$y = A\,{}_2F_1(a,b;c;z) + Bz^{1-c}\,{}_2F_1(a+1-c, b+1-c; 2-c; z). \tag{15}$$

Derivatives are given by

$$\frac{d\,{}_2F_1(a,b;c;z)}{dz} = \frac{ab}{c}\,{}_2F_1(a+1,b+1;c+1;z) \tag{16}$$

$$\frac{d^2\,{}_2F_1(a,b;c;z)}{dz^2} = \frac{a(a+1)b(b+1)}{c(c+1)} \times {}_2F_1(a+2,b+2;c+2;z) \tag{17}$$

(Magnus and Oberhettinger 1949, p. 8). An integral giving the hypergeometric function is

$${}_2F_1(a,b;c;z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-tz)^a}\,dt \tag{18}$$

as shown by Euler in 1748.

A hypergeometric function can be written using EULER'S HYPERGEOMETRIC TRANSFORMATIONS

$$t \to t \tag{19}$$

$$t \to 1-t \tag{20}$$

$$t \to (1-z-tz)^{-1} \tag{21}$$

$$t \to \frac{1-t}{1-tz} \tag{22}$$

in any one of four equivalent forms

$$ {}_2F_1(a,b;c;z) = (1-z)^{-a}\,{}_2F_1(a,c-b;c;z/(z-1)) \tag{23}$$

$$ = (1-z)^{-b}\,{}_2F_1(c-a,b;c;z/(z-1)) \tag{24}$$

$$ = (1-z)^{c-a-b}\,{}_2F_1(c-a,c-b;c;z). \tag{25}$$

It can also be written as a linear combination

$$\begin{aligned} {}_2F_1(a,b;c;z) &= \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}{}_2F_1(a,b;a+b+1-c;1-z) \\ &\quad + \frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)}(1-z)^{c-a-b} \\ &\quad \times {}_2F_1(c-a,c-b;1+c-a-b;1-z). \end{aligned} \tag{26}$$

Kummer found all six solutions (not necessarily regular at the origin) to the HYPERGEOMETRIC DIFFERENTIAL EQUATION,

$$\begin{aligned} u_1(x) &= {}_2F_1(a,b;c;z) \\ u_2(x) &= {}_2F_1(a,b;a+b+1-c;1-z) \\ u_3(x) &= z^{-a}\,{}_2F_1(a,a+1-c;a+1-b;1/z) \\ u_4(x) &= z^{-b}\,{}_2F_1(b+1-c,b;b+1-a;1/z) \\ u_5(x) &= z^{1-c}\,{}_2F_1(b+1-c,a+1-c;2-c;z) \\ u_6(x) &= (1-z)^{c-a-b}\,{}_2F_1(c-a,c-b;c+1-a-b;1-z). \end{aligned}$$

Applying EULER'S HYPERGEOMETRIC TRANSFORMATIONS to the Kummer solutions then gives all 24 possible forms which are solutions to the HYPERGEOMETRIC DIFFERENTIAL EQUATION

$$\begin{aligned} u_1^{(1)}(x) &= {}_2F_1(a,b;c;z) \\ u_1^{(2)}(x) &= (1-z)^{-a}\,{}_2F_1(a,c-b;c;z/(z-1)) \\ u_1^{(3)}(x) &= (1-z)^{-b}\,{}_2F_1(c-a,b;c;z/(z-1)) \\ u_1^{(4)}(x) &= (1-z)^{c-a-b}\,{}_2F_1(c-a,c-b;c;z) \\ u_2^{(1)}(x) &= {}_2F_1(a,b;a+b+1-c;1-z) \\ u_2^{(2)}(x) &= z^{-a}\,{}_2F_1(a,a+1-c;a+b+1-c;1-1/z) \\ u_2^{(3)}(x) &= z^{-b}\,{}_2F_1(b+1-c,b;a+b+1-c;1-1/z) \\ u_2^{(4)}(x) &= z^{1-c}\,{}_2F_1(b+1-c,a+1-c;a+b+1-c;1-z) \\ u_3^{(1)}(x) &= z^{-a}\,{}_2F_1(a,a+1-c;a+1-b;1/z) \\ u_3^{(2)}(x) &= z^{-a}(1-1/1z)^{-a}\,{}_2F_1(a,c-b;a+1-b;1/(1-z)) \\ u_3^{(3)}(x) &= z^{-a}(1-1/z)^{c-a-1} \\ &\quad \times {}_2F_1(1-b,a+1-c;a+1-b;1/(1-z)) \\ u_3^{(4)}(x) &= z^{-a}(1-1/z)^{c-a-b}\,{}_2F_1(1-b,c-b;a+1-b;1/z) \\ u_4^{(1)}(x) &= z^{-a}\,{}_2F_1(b+1-c,b;b+1-a;1/z) \\ u_4^{(2)}(x) &= z^{-b}(1-1/z)^{c-b-1} \\ &\quad \times {}_2F_1(b_1-c,1-a;b+1-a;1/(1-z)) \\ u_4^{(3)}(x) &= z^{-b}(1-1/z)^{-b}\,{}_2F_1(c-a,b;b+1-a;1/(1-z)) \\ u_4^{(4)}(x) &= z^{-b}(1-1/z)^{c-a-b}\,{}_2F_1(c-a,1-a;b+1-a;1/z) \\ u_5^{(1)}(x) &= z^{1-c}\,{}_2F_1(b+1-c,a+1-c;2-c;z) \\ u_5^{(2)}(x) &= z^{1-c}(1-z)^{c-b-1}\,{}_2F_1(b+1-c,1-a;2-c;z/(z-1)) \\ u_5^{(2)}(x) &= z^{1-c}(1-z)^{c-a-1}\,{}_2F_1(1-b,a+1-c;2-c;z/(z-1)) \\ u_5^{(4)}(x) &= z^{1-c}(1-z)^{c-a-b}\,{}_2F_1(1-b,1-a;2-c;z) \\ u_6^{(1)}(x) &= (1-z)^{c-a-b}\,{}_2F_1(c-a,c-b;c+1-a-b;1-z) \\ u_6^{(2)}(x) &= z^{a-c}(1-z)^{c-a-b} \\ &\quad \times {}_2F_1(c-a,1-a;c+1-a-b;1-1/z) \\ u_6^{(2)}(x) &= z^{b-c}(1-z)^{c-a-b} \\ &\quad \times {}_2F_1(1-b,c-b;c+1-a-b;1-1/z) \\ u_6^{(4)}(x) &= z^{c-a-b}(1-z)^{c-a-b} \\ &\quad \times {}_2F_1(1-b,1-a;c+1-a-b;1-z). \end{aligned}$$

Goursat (1881) gives many hypergeometric transformation FORMULAS, including several cubic transformation FORMULAS.

Many functions of mathematical physics can be expressed as special cases of the hypergeometric functions. For example,

$$ {}_2F_1(-l,l+1,1;(1-z)/2) = P_l(z), \tag{27}$$

where $P_l(z)$ is a LEGENDRE POLYNOMIAL.

$$ (1+z)^n = {}_2F_1(-n,b;b;-z) \tag{28}$$

$$ \ln(1+z) = z\,{}_2F_1(1,1;2;-z) \tag{29}$$

Complete ELLIPTIC INTEGRALS and the RIEMANN P-SERIES can also be expressed in terms of ${}_2F_1(a,b;c;z)$. Special values include

$$\begin{aligned} {}_2F_1(a,b;a-b+1;-1) \\ = 2^{-a}\sqrt{\pi}\frac{\Gamma(1+a+b)}{\Gamma(1+\frac{1}{2}a-b)\Gamma(\frac{1}{2}+\frac{1}{2}a)} \end{aligned} \tag{30}$$

$$ {}_2F_1(1,-a;a;-1) = \frac{\sqrt{\pi}}{2}\frac{\Gamma(a)}{\Gamma(a+\frac{1}{2})}+1 \tag{31}$$

$$ {}_2F_1(a,b;c;\tfrac{1}{2}) = 2^a\,{}_2F_1(a,c-b;c;-1) \tag{32}$$

$$ {}_2F_1(a,b;\tfrac{1}{2}(a+b+1);\tfrac{1}{2}) = \frac{\Gamma(\frac{1}{2})\Gamma[(\frac{1}{2}(1+a+b)]}{\Gamma[\frac{1}{2}(1+a)]\Gamma[\frac{1}{2}(1+b)]} \tag{33}$$

$$ {}_2F_1(a,1-a;c;\tfrac{1}{2}) = \frac{\Gamma(\frac{1}{2}c)\Gamma[\frac{1}{2}(c+1)]}{\Gamma[\frac{1}{2}(a+c)]\Gamma[\frac{1}{2}(1+c-a)]} \tag{34}$$

$$ {}_2F_1(a,b;c;1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}. \tag{35}$$

KUMMER'S FIRST FORMULA gives

$$\begin{aligned} {}_2F_1(\tfrac{1}{2}+m-k,-n;2m+1;1) \\ = \frac{\Gamma(2m+1)\Gamma(m+\frac{1}{2}+k+n)}{\Gamma(m+\frac{1}{2}+k)\Gamma(2m+1+n)}, \end{aligned} \tag{36}$$

where $m \neq -1/2, -1, -3/2, \ldots$. Many additional identities are given by Abramowitz and Stegun (1972, p. 557).

Hypergeometric functions can be generalized to GENERALIZED HYPERGEOMETRIC FUNCTIONS

$$ {}_nF_m(a_1,\ldots,a_n;b_1,\ldots,b_m;z). \qquad (37)$$

A function of the form ${}_1F_1(a;b;z)$ is called a CONFLUENT HYPERGEOMETRIC FUNCTION, and a function of the form ${}_0F_1(;b;z)$ is called a CONFLUENT HYPERGEOMETRIC LIMIT FUNCTION.

see also APPELL HYPERGEOMETRIC FUNCTION, BARNES' LEMMA, BRADLEY'S THEOREM, CAYLEY'S HYPERGEOMETRIC FUNCTION THEOREM, CLAUSEN FORMULA, CLOSED FORM, CONFLUENT HYPERGEOMETRIC FUNCTION, CONFLUENT HYPERGEOMETRIC LIMIT FUNCTION, CONTIGUOUS FUNCTION, DARLING'S PRODUCTS, GENERALIZED HYPERGEOMETRIC FUNCTION, GOSPER'S ALGORITHM, HYPERGEOMETRIC IDENTITY, HYPERGEOMETRIC SERIES, JACOBI POLYNOMIAL, KUMMER'S FORMULAS, KUMMER'S QUADRATIC TRANSFORMATION, KUMMER'S RELATION, ORR'S THEOREM, RAMANUJAN'S HYPERGEOMETRIC IDENTITY, SAALSCHÜTZIAN, SISTER CELINE'S METHOD, ZEILBERGER'S ALGORITHM

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Hypergeometric Functions." Ch. 15 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 555–566, 1972.

Arfken, G. "Hypergeometric Functions." §13.5 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 748–752, 1985.

Fine, N. J. *Basic Hypergeometric Series and Applications.* Providence, RI: Amer. Math. Soc., 1988.

Gasper, G. and Rahman, M. *Basic Hypergeometric Series.* Cambridge, England: Cambridge University Press, 1990.

Gauss, C. F. "Disquisitiones Generales Circa Seriem Infinitam $\left[\frac{\alpha\beta}{1\cdot\gamma}\right]x+\left[\frac{\alpha(\alpha+1)\beta(\beta+1)}{1\cdot2\cdot\gamma(\gamma+1)}\right]x^2 + \left[\frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{1\cdot2\cdot3\cdot\gamma(\gamma+1)(\gamma+2)}\right]x^3+$ etc. Pars Prior." *Commentationes Societiones Regiae Scientiarum Gottingensis Recentiores, Vol. II.* 1813.

Gessel, I. and Stanton, D. "Strange Evaluations of Hypergeometric Series." *SIAM J. Math. Anal.* **13**, 295–308, 1982.

Gosper, R. W. "Decision Procedures for Indefinite Hypergeometric Summation." *Proc. Nat. Acad. Sci. USA* **75**, 40–42, 1978.

Goursat, M. E. "Sur l'équation différentielle linéaire qui admet pour intégrale la série hypergéométrique." *Ann. Sci. École Norm. Super. Sup.* **10**, S3–S142, 1881.

Iyanaga, S. and Kawada, Y. (Eds.). "Hypergeometric Functions and Spherical Functions." Appendix A, Table 18 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 1460–1468, 1980.

Kummer, E. E. "Über die Hypergeometrische Reihe." *J. für die Reine Angew. Mathematik* **15**, 39–83 and 127–172, 1837.

Magnus, W. and Oberhettinger, F. *Formulas and Theorems for the Special Functions of Mathematical Physics.* New York: Chelsea, 1949.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 541–547, 1953.

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, 1996.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Hypergeometric Functions." §6.12 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 263–265, 1992.

Seaborn, J. B. *Hypergeometric Functions and Their Applications.* New York: Springer-Verlag, 1991.

Snow, C. *Hypergeometric and Legendre Functions with Applications to Integral Equations of Potential Theory.* Washington, DC: U. S. Government Printing Office, 1952.

Spanier, J. and Oldham, K. B. "The Gauss Function $F(a,b;c;x)$." Ch. 60 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 599–607, 1987.

Hypergeometric Identity

A relation expressing a sum potentially involving BINOMIAL COEFFICIENTS, FACTORIALS, RATIONAL FUNCTIONS, and power functions in terms of a simple result. Thanks to results by Fasenmyer, Gosper, Zeilberger, Wilf, and Petkovšek, the problem of determining whether a given hypergeometric sum is expressible in simple closed form and, if so, finding the form, is now (subject to a mild restriction) completely solved. The algorithm which does so has been implemented in several computer algebra packages and is called ZEILBERGER'S ALGORITHM.

see also GENERALIZED HYPERGEOMETRIC FUNCTION, GOSPER'S ALGORITHM, HYPERGEOMETRIC SERIES, SISTER CELINE'S METHOD, WILF-ZEILBERGER PAIR, ZEILBERGER'S ALGORITHM

References

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, p. 18, 1996.

Hypergeometric Polynomial

see JACOBI POLYNOMIAL

Hypergeometric Series

A hypergeometric series $\sum_k a_k$ is a series for which $a_0 = 1$ and the ratio of consecutive terms is a RATIONAL FUNCTION of the summation index k, i.e., one for which

$$\frac{a_{k+1}}{a_k}=\frac{P(k)}{Q(k)},$$

with $P(k)$ and $Q(k)$ POLYNOMIALS. The functions generated by hypergeometric series are called HYPERGEOMETRIC FUNCTIONS or, more generally, GENERALIZED HYPERGEOMETRIC FUNCTIONS. If the polynomials are completely factored, the ratio of successive terms can be written

$$\frac{a_{k+1}}{a_k}=\frac{P(k)}{Q(k)}=\frac{(k+a_1)(k+a_2)\cdot(k+a_p)}{(k+b_1)(k+b_2)\cdots(k+b_q)(k+1)}x,$$

where the factor of $k+1$ in the DENOMINATOR is present for historical reasons of notation, and the resulting GENERALIZED HYPERGEOMETRIC FUNCTION is written

$${}_pF_q\begin{bmatrix}a_1 & a_2 & \cdots & a_p\\ b_1 & b_2 & \cdots & b_q\end{bmatrix};x\Bigg]=\sum_{k=0}a_kx^k.$$

If $p = 2$ and $q = 1$, the function becomes a traditional HYPERGEOMETRIC FUNCTION ${}_2F_1(a, b; c; x)$.

Many sums can be written as GENERALIZED HYPERGEOMETRIC FUNCTIONS by inspections of the ratios of consecutive terms in the generating hypergeometric series.

see also GENERALIZED HYPERGEOMETRIC FUNCTION, GEOMETRIC SERIES, HYPERGEOMETRIC FUNCTION, HYPERGEOMETRIC IDENTITY

References
Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. "Hypergeometric Series," "How to Identify a Series as Hypergeometric," and "Software That Identifies Hypergeometric Series." §3.2–3.4 in *A=B.* Wellesley, MA: A. K. Peters, pp. 34–42, 1996.

Hypergroup

A MEASURE ALGEBRA which has many properties associated with the convolution MEASURE ALGEBRA of a GROUP, but no algebraic structure is assumed for the underlying SPACE.

References
Bloom, W. R.; and Heyer, H. *The Harmonic Analysis of Probability Measures on Hypergroups.* Berlin: de Gruyter, 1995.
Jewett, R. I. "Spaces with an Abstract Convolution of Measures." *Adv. Math.* **18**, 1–101, 1975.

Hypermatrix

A generalization of the MATRIX to an $n_1 \times n_2 \times \cdots$ array of numbers.

see also HYPERDETERMINANT

References
Gel'fand, I. M.; Kapranov, M. M.; and Zelevinsky, A. V. "Hyperdeterminants." *Adv. Math.* **96**, 226–263, 1992.

Hyperparallel

Two lines in HYPERBOLIC GEOMETRY which diverge from each other in both directions.

see also ANTIPARALLEL, IDEAL POINT, PARALLEL

Hyperperfect Number

A number n is called k-hyperperfect if

$$n = 1 + k \sum_i d_i,$$

where the summation is over the PROPER DIVISORS with $1 < d_i < n$, giving

$$k\sigma(n) = (k + 1)n + k + 1,$$

where $\sigma(n)$ is the DIVISOR FUNCTION. The first few hyperperfect numbers are 21, 301, 325, 697, 1333, ... (Sloane's A007592). 2-hyperperfect numbers include 21, 2133, 19521, 176661, ... (Sloane's A007593), and the first 3-hyperperfect number is 325.

References
Guy, R. K. "Almost Perfect, Quasi-Perfect, Pseudoperfect, Harmonic, Weird, Multiperfect and Hyperperfect Numbers." §B2 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 45–53, 1994.
Sloane, N. J. A. Sequences A007592/M5113 and A007593/M5121 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Hyperplane

Let a_1, a_2, ..., a_n be SCALARS not all equal to 0. Then the SET S consisting of all VECTORS

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

in $\mathbb{R}^n$ such that

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0$$

is a SUBSPACE of $\mathbb{R}^n$ called a hyperplane. More generally, a hyperplane is any co-dimension 1 vector SUBSPACE of a VECTOR SPACE. Equivalently, a hyperplane V in a VECTOR SPACE W is any SUBSPACE such that W/V is 1-dimensional. Equivalently, a hyperplane is the KERNEL of any NONZERO linear MAP from the VECTOR SPACE to the underlying FIELD.

Hyperreal Number

Hyperreal numbers are an extension of the REAL NUMBERS to include certain classes of infinite and infinitesimal numbers. A hyperreal number is said to be finite IFF $|x| < n$ for some INTEGER n. x is said to be infinitesimal IFF $|x| < 1/n$ for all INTEGERS n.

see also AX-KOCHEN ISOMORPHISM THEOREM, NONSTANDARD ANALYSIS

References
Apps, P. "The Hyperreal Line." `http://www.math.wisc.edu/~apps/line.html`.
Keisler, H. J. "The Hyperreal Line." In *Real Numbers, Generalizations of the Reals, and Theories of Continua* (Ed. P. Ehrlich). Norwell, MA: Kluwer, 1994.

Hyperspace

A SPACE having DIMENSION $n > 3$.

Hypersphere

The n-hypersphere (often simply called the n-sphere) is a generalization of the CIRCLE ($n = 2$) and SPHERE ($n = 3$) to dimensions $n \geq 4$. It is therefore defined as the set of n-tuples of points (x_1, x_2, ..., x_n) such that

$$x_1{}^2 + x_2{}^2 + \ldots + x_n{}^2 = R^2, \tag{1}$$

where R is the RADIUS of the hypersphere. The CONTENT (i.e., n-D VOLUME) of an n-hypersphere of RADIUS R is given by

$$V_n = \int_0^R S_n r^{n-1}\,dr = \frac{S_n R^n}{n}, \qquad (2)$$

where S_n is the hyper-SURFACE AREA of an n-sphere of unit radius. But, for a unit hypersphere, it must be true that

$$\begin{aligned} S_n \int_0^\infty e^{-r^2} r^{n-1}\,dr &= \underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{n} e^{-({x_1}^2+\ldots+{x_n}^2)}\,dx_1\cdots dx_m \\ &= \left(\int_{-\infty}^{\infty} e^{-x^2}\,dx\right)^n. \end{aligned} \qquad (3)$$

But the GAMMA FUNCTION can be defined by

$$\Gamma(m) = 2\int_0^\infty e^{-r^2} r^{2m-1}\,dr, \qquad (4)$$

so

$$\tfrac{1}{2} S_n \Gamma(\tfrac{1}{2}n) = [\Gamma(\tfrac{1}{2})]^n = (\pi^{1/2})^n \qquad (5)$$

$$S_n = \frac{2\pi^{n/2}}{\Gamma(\frac{1}{2}n)}. \qquad (6)$$

This gives the RECURRENCE RELATION

$$S_{n+2} = \frac{2\pi S_n}{n}. \qquad (7)$$

Using $\Gamma(n+1) = n\Gamma(n)$ then gives

$$V_n = \frac{S_n R^n}{n} = \frac{\pi^{n/2} R^n}{(\frac{1}{2}n)\Gamma(\frac{1}{2}n)} = \frac{\pi^{n/2} R^n}{\Gamma(1+\frac{1}{2}n)} \qquad (8)$$

(Conway and Sloane 1993).

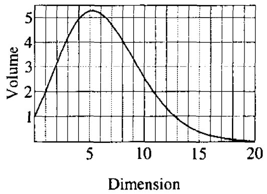

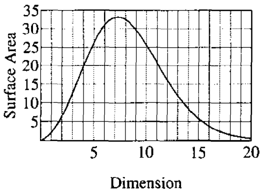

Strangely enough, the hyper-SURFACE AREA and CONTENT reach MAXIMA and then decrease towards 0 as n increases. The point of MAXIMAL hyper-SURFACE AREA satisfies

$$\frac{dS_n}{dn} = \frac{\pi^{n/2}[\ln\pi - \psi_0(\frac{1}{2}n)]}{\Gamma(\frac{1}{2}n)} = 0, \qquad (9)$$

where $\psi_0(x) \equiv \Psi(x)$ is the DIGAMMA FUNCTION. The point of MAXIMAL CONTENT satisfies

$$\frac{dV_n}{dn} = \frac{\pi^{n/2}[\ln\pi - \psi_0(1+\frac{1}{2}n)]}{2\Gamma(1+\frac{1}{2}n)} = 0. \qquad (10)$$

Neither can be solved analytically for n, but the numerical solutions are $n = 7.25695\ldots$ for hyper-SURFACE AREA and $n = 5.25695\ldots$ for CONTENT. As a result, the 7-D and 5-D hyperspheres have MAXIMAL hyper-SURFACE AREA and CONTENT, respectively (Le Lionnais 1983).

n	V_n	$V_n/V_{n-\text{cube}}$	S_n
0	1	1	0
1	2	1	2
2	π	$\frac{1}{4}\pi$	2π
3	$\frac{4}{3}\pi$	$\frac{1}{6}\pi$	4π
4	$\frac{1}{2}\pi^2$	$\frac{1}{32}\pi^2$	$2\pi^2$
5	$\frac{8}{15}\pi^2$	$\frac{1}{60}\pi^2$	$\frac{8}{3}\pi^2$
6	$\frac{1}{6}\pi^3$	$\frac{1}{384}\pi^3$	π^3
7	$\frac{16}{105}\pi^3$	$\frac{1}{840}\pi^3$	$\frac{16}{15}\pi^3$
8	$\frac{1}{24}\pi^4$	$\frac{1}{6144}\pi^4$	$\frac{1}{3}\pi^4$
9	$\frac{32}{945}\pi^4$	$\frac{1}{15120}\pi^4$	$\frac{32}{105}\pi^4$
10	$\frac{1}{120}\pi^5$	$\frac{1}{122880}\pi^5$	$\frac{1}{12}\pi^5$

In 4-D, the generalization of SPHERICAL COORDINATES is defined by

$$x_1 = R\sin\psi\sin\phi\cos\theta \qquad (11)$$

$$x_2 = R\sin\psi\sin\phi\sin\theta \qquad (12)$$

$$x_3 = R\sin\psi\cos\phi \qquad (13)$$

$$x_4 = R\cos\psi. \qquad (14)$$

The equation for a 4-sphere is

$${x_1}^2 + {x_2}^2 + {x_3}^2 + {x_4}^2 = R^2, \qquad (15)$$

and the LINE ELEMENT is

$$ds^2 = R^2[d\psi^2 + \sin^2\psi(d\phi^2 + \sin^2\phi\,d\theta^2)]. \qquad (16)$$

By defining $r \equiv R\sin\psi$, the LINE ELEMENT can be rewritten

$$ds^2 = \frac{dr^2}{\left(1-\frac{r^2}{R^2}\right)} + r^2(d\phi^2 + \sin^2\phi\,d\theta^2). \qquad (17)$$

The hyper-SURFACE AREA is therefore given by

$$\begin{aligned} S_4 &= \int_0^\pi R\,d\psi \int_0^\pi R\sin\psi\,d\phi \int_0^{2\pi} R\sin\psi\sin\phi\,d\theta \\ &= 2\pi^2 R^3. \end{aligned} \qquad (18)$$

see also CIRCLE, HYPERCUBE, HYPERSPHERE PACKING, MAZUR'S THEOREM, SPHERE, TESSERACT

References

Conway, J. H. and Sloane, N. J. A. *Sphere Packings, Lattices, and Groups, 2nd ed.* New York: Springer-Verlag, p. 9, 1993.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 58, 1983.

Peterson, I. *The Mathematical Tourist: Snapshots of Modern Mathematics.* New York: W. H. Freeman, pp. 96–101, 1988.

Hypersphere Packing

Draw unit n-spheres in an n-D space centered at all ± 1 coordinates. Then place an additional HYPERSPHERE at the origin tangent to the other HYPERSPHERES. Then the central HYPERSPHERE is contained with the HYPERSPHERE with VERTICES at the center of the other spheres for n between 2 and 8. However, for $n = 9$, the central HYPERSPHERE just touches the bounding HYPERSPHERE, and for $n > 9$, the HYPERSPHERE is partially outside the hypercube. This can be seen by finding the distance from the origin to the center of one of the HYPERSPHERES

$$\underbrace{\sqrt{(\pm1)^2+\ldots+(\pm1)^2}}_{n}=\sqrt{n}.$$

The radius of the central sphere is therefore $\sqrt{n}-1$. The distance from the origin to the center of the bounding hypercube is always 2 (two radii), so the center HYPERSPHERE is tangent when $\sqrt{n}-1=2$, or $n=9$, and outside for $n>9$.

The analog of face-centered cubic packing is the densest lattice in 4- and 5-D. In 8-D, the densest lattice packing is made up of two copies of face-centered cubic. In 6- and 7-D, the densest lattice packings are cross-sections of the 8-D case. In 24-D, the densest packing appears to be the LEECH LATTICE. For high dimensions ($\sim$ 1000-D), the densest known packings are nonlattice. The densest lattice packings in n-D have been rigorously proved to have PACKING DENSITY 1, $\pi/(2\sqrt{3})$, $\pi/(3\sqrt{2})$, $\pi^2/16$, $\pi^2/(15\sqrt{2})$, $\pi^3/(48\sqrt{3})$, $\pi^3/105$, and $\pi^4/384$ (Finch).

The largest number of unit CIRCLES which can touch another is six. For SPHERES, the maximum number is 12. Newton considered this question long before a proof was published in 1874. The maximum number of hyperspheres that can touch another in n-D is the so-called KISSING NUMBER.

see also KISSING NUMBER, LEECH LATTICE, SPHERE PACKING

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/hermit/hermit.html`.

Gardner, M. *Martin Gardner's New Mathematical Diversions from Scientific American.* New York: Simon and Schuster, pp. 89–90, 1966.

Hypervolume

see CONTENT

Hypocycloid

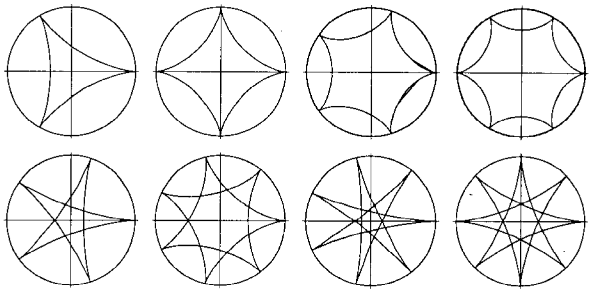

The curve produced by a small CIRCLE of RADIUS b rolling around the inside of a large CIRCLE of RADIUS $a > b$. A hypocycloid is a HYPOTROCHOID with $h = b$. To derive the equations of the hypocycloid, call the ANGLE by which a point on the small CIRCLE rotates about its center ϑ, and the ANGLE from the center of the large CIRCLE to that of the small CIRCLE ϕ. Then

$$(a-b)\phi = b\vartheta, \tag{1}$$

so

$$\vartheta = \frac{a-b}{b}\phi. \tag{2}$$

Call $\rho \equiv a - 2b$. If $x(0) = \rho$, then the first point is at minimum radius, and the Cartesian parametric equations of the hypocycloid are

$$\begin{aligned} x &= (a-b)\cos\phi - b\cos\vartheta \\ &= (a-b)\cos\phi - b\cos\left(\frac{a-b}{b}\phi\right) \qquad (3)\\ y &= (a-b)\sin\phi + b\sin\vartheta \\ &= (a-b)\sin\phi + b\sin\left(\frac{a-b}{b}\phi\right). \qquad (4)\end{aligned}$$

If $x(0) = a$ instead so the first point is at maximum radius (on the CIRCLE), then the equations of the hypocycloid are

$$x = (a-b)\cos\phi + b\cos\left(\frac{a-b}{b}\phi\right) \tag{5}$$

$$y = (a-b)\sin\phi - b\sin\left(\frac{a-b}{b}\phi\right). \tag{6}$$

An n-cusped non-self-intersecting hypocycloid has $a/b = n$. A 2-cusped hypocycloid is a LINE SEGMENT, as can be seen by setting $a = b$ in equations (3) and (4) and noting that the equations simplify to

$$x = a\sin\phi \tag{7}$$

$$y = 0. \tag{8}$$

A 3-cusped hypocycloid is called a DELTOID or TRICUSPOID, and a 4-cusped hypocycloid is called an ASTROID. If a/b is rational, the curve closes on itself and has b cusps. If a/b is IRRATIONAL, the curve never closes and fills the entire interior of the CIRCLE.

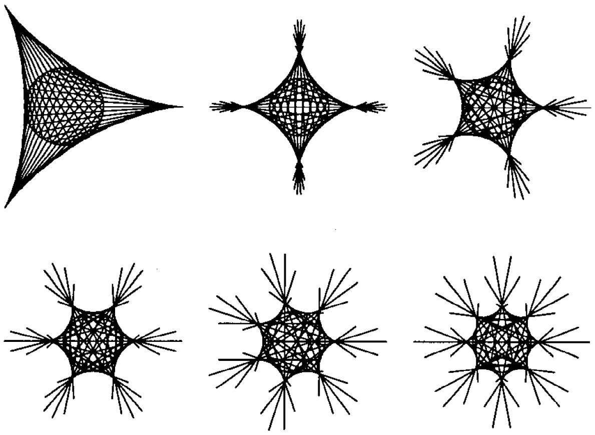

n-hypocycloids can also be constructed by beginning with the DIAMETER of a CIRCLE, offsetting one end by a series of steps while at the same time offsetting the other end by steps n times as large in the opposite direction and extending beyond the edge of the CIRCLE. After traveling around the CIRCLE once, an n-cusped hypocycloid is produced, as illustrated above (Madachy 1979).

Let r be the radial distance from a fixed point. For RADIUS OF TORSION ρ and ARC LENGTH s, a hypocycloid can given by the equation

$$s^2 + \rho^2 = 16r^2 \tag{9}$$

(Kreyszig 1991, pp. 63–64). A hypocycloid also satisfies

$$\sin^2\psi = \frac{\rho^2}{a^2-\rho^2}\frac{a^2-r^2}{r^2}, \tag{10}$$

where

$$r\frac{dr}{d\theta} = \tan\psi \tag{11}$$

and ψ is the ANGLE between the RADIUS VECTOR and the TANGENT to the curve.

The ARC LENGTH of the hypocycloid can be computed as follows

$$\begin{aligned} x' &= -(a-b)\sin\phi - (a-b)\sin\left(\frac{a-b}{b}\phi\right) \\ &= (a-b)\left[\sin\phi + \sin\left(\frac{a-b}{b}\phi\right)\right] \end{aligned} \tag{12}$$

$$\begin{aligned} y' &= (a-b)\cos\phi - (a-b)\cos\left(\frac{a-b}{a}\phi\right) \\ &= (a-b)\left[\cos\phi - \cos\left(\frac{a-b}{b}\phi\right)\right] \end{aligned} \tag{13}$$

$$\begin{aligned} x'^2 + y'^2 &= (a-b)^2\left[\sin^2\phi + 2\sin\phi\sin\left(\frac{a-b}{b}\phi\right)\right. \\ &\quad + \sin^2\left(\frac{a-b}{b}\phi\right) + \cos^2\phi - 2\cos\phi\cos\left(\frac{a-b}{b}\phi\right) \\ &\quad \left. + \cos^2\left(\frac{a-b}{b}\phi\right)\right] \\ &= (a-b)^2\left\{2 + 2\left[\sin\phi\sin\left(\frac{a-b}{a}\phi\right)\right.\right. \\ &\quad \left.\left. - \cos\phi\cos\left(\frac{a-b}{b}\phi\right)\right]\right\} \\ &= 2(a-b)^2\left[1 - \cos\left(\phi + \frac{a-b}{b}\phi\right)\right] \\ &= 4(a-b)^2\tfrac{1}{2}\left[1-\cos\left(\frac{a}{b}\phi\right)\right] = 4(a-b)^2\sin^2\left(\frac{a\phi}{2b}\right), \end{aligned} \tag{14}$$

so

$$ds = \sqrt{x'^2 + y'^2}\,d\phi = 2(a-b)\sin\left(\frac{a\phi}{2b}\right)d\phi \tag{15}$$

for $\phi \leq (b/2a)\pi$. Integrating,

$$\begin{aligned} s(\phi) &= \int_0^\phi ds = 2(a-b)\left[-\frac{2b}{a}\cos\left(\frac{a\phi}{2b}\right)\right]_0^\phi \\ &= \frac{4b(a-b)}{a}\left[-\cos\left(\frac{a}{2b}\phi\right) + 1\right] \\ &= \frac{8b(a-b)}{a}\sin^2\left(\frac{a}{4b}\phi\right). \end{aligned} \tag{16}$$

The length of a single cusp is then

$$s\left(2\pi\frac{b}{a}\right) = \frac{8b(a-b)}{a}\sin^2\left(\frac{\pi}{2}\right) = \frac{8b(a-b)}{a}. \tag{17}$$

If $n \equiv a/b$ is rational, then the curve closes on itself without intersecting after n cusps. For $n \equiv a/b$ and with $x(0) = a$, the equations of the hypocycloid become

$$x = \frac{1}{n}[(n-1)\cos\phi - \cos[(n-1)\phi]a, \tag{18}$$

$$y = \frac{1}{n}[(n-1)\sin\phi + \sin[(n-1)\phi]a, \tag{19}$$

and

$$s_n = n\frac{8b(bn-b)}{nb} = 8b(n-1) = \frac{8a(n-1)}{n}. \tag{20}$$

Compute

$$\begin{aligned} xy' - yx' &= \left[(a-b)\cos\phi + b\cos\left(\frac{a-b}{a}\phi\right)\right](b-a) \\ &\quad \times \left[\sin\phi + \sin\left(\frac{a-b}{b}\phi\right)\right] \\ &\quad - \left[(a-b)\sin\phi - b\sin\left(\frac{a-b}{b}\phi\right)\right](a-b) \\ &\quad \times \left[\cos\phi - \cos\left(\frac{a-b}{b}\phi\right)\right] \\ &= 2(a^2 - 3ab + 2b^2)\sin^2\left(\frac{a\phi}{2b}\right). \end{aligned} \tag{21}$$

The AREA of one cusp is then

$$\begin{aligned} A &= \tfrac{1}{2}\int_0^{2\pi b/a} (xy' - yx')\,d\phi \\ &= (a^2 - 3ab + 2b^2)\left[\frac{at - b\sin\left(\frac{at}{b}\right)}{2a}\right]_a^{2\pi b/a} \\ &= (a^2 - 3ab + 2b^2)\left[\frac{a\left(2\pi\frac{b}{a}\right)}{2a}\right] \\ &= \frac{b(a^2 - 3ab + 2b^2)}{a}\pi. \end{aligned} \tag{22}$$

If $n = a/b$ is rational, then after n cusps,

$$\begin{aligned} A_n &= n\pi\frac{b(a^2 - 3ab + 2b^2)}{a} = n\pi\frac{\frac{a}{n}\left(a^2 - 3a\frac{a}{n} + 2\frac{a^2}{n^2}\right)}{a} \\ &= \frac{n^2 - 3n + 2}{n^2}\pi a^2 = \frac{(n-1)(n-2)}{n^2}\pi a^2. \end{aligned} \tag{23}$$

The equation of the hypocycloid can be put in a form which is useful in the solution of CALCULUS OF VARIATIONS problems with radial symmetry. Consider the case $x(0) = \rho$, then

$$\begin{aligned} r^2 &= x^2 + y^2 \\ &= \Big[(a-b)^2\cos^2\phi - 2(a-b)b\cos\phi\cos\left(\frac{a-b}{b}\phi\right) \\ &\quad + b^2\cos^2\left(\frac{a-b}{b}\phi\right) \\ &\quad + (a-b)^2\sin^2\phi + 2(a-b)b\sin\phi\sin\left(\frac{a-b}{b}\phi\right) \\ &\quad + b^2\sin^2\left(\frac{a-b}{b}\phi\right)\Big] \\ &= \Big\{(a-b)^2 + b^2 - 2(a-b)b \\ &\quad \times \left[\cos\phi\cos\left(\frac{a-b}{b}\phi\right) - \sin\phi\sin\left(\frac{a-b}{b}\phi\right)\right]\Big\} \\ &= (a-b)^2 + b^2 - 2(a-b)b\cos\left(\frac{a}{b}\phi\right). \end{aligned} \tag{24}$$

But $\rho = a - 2b$, so $b = (a-\rho)/2$, which gives

$$\begin{aligned} (a-b)^2 + b^2 &= [a - \tfrac{1}{2}(a-\rho)]^2 + [\tfrac{1}{2}(a-\rho)]^2 \\ &= [\tfrac{1}{2}(a+\rho)]^2 + [\tfrac{1}{2}(a-\rho)]^2 \\ &= \tfrac{1}{4}(a^2 + 2a\rho + \rho^2 + a^2 - 2a\rho + \rho^2) \\ &= \tfrac{1}{2}(a^2 + \rho^2) \end{aligned} \tag{25}$$

$$\begin{aligned} 2(a-b)b &= 2[a - \tfrac{1}{2}(a-\rho)]\tfrac{1}{2}(a-\rho) \\ &= \tfrac{1}{2}(a+\rho)(a-\rho) = \tfrac{1}{2}(a^2 - \rho^2). \end{aligned} \tag{26}$$

Now let

$$2\Omega t \equiv \frac{a}{b}\phi, \tag{27}$$

so

$$\phi = \frac{a-\rho}{a}\Omega t \tag{28}$$

$$\frac{\phi}{a-\rho} = \frac{\Omega t}{a}, \tag{29}$$

then

$$\begin{aligned} r^2 &= \tfrac{1}{2}(a^2 + \rho^2) - \tfrac{1}{2}(a^2 - \rho^2)\cos\left(\frac{a}{b}\phi\right) \\ &= \tfrac{1}{2}(a^2 + \rho^2) - \tfrac{1}{2}(a^2 - \rho^2)\cos(2\Omega t). \end{aligned} \tag{30}$$

The POLAR ANGLE is

$$\tan\theta \equiv \frac{y}{x} = \frac{(a-b)\sin\phi + b\sin\left(\frac{a-b}{a}\phi\right)}{(a-b)\cos\phi - b\cos\left(\frac{a-b}{a}\phi\right)}. \tag{31}$$

But

$$b = \tfrac{1}{2}(a-\rho) \tag{32}$$

$$a - b = \tfrac{1}{2}(a+\rho) \tag{33}$$

$$\frac{a-b}{b} = \frac{a+\rho}{a-\rho}, \tag{34}$$

so

$$\begin{aligned} \tan\theta &= \frac{\frac{1}{2}(a+\rho)\sin\phi + \frac{1}{2}(a-\rho)\sin\left(\frac{a+\rho}{a-\rho}\phi\right)}{\frac{1}{2}(a+\rho)\cos\phi - \frac{1}{2}(a-\rho)\cos\left(\frac{a+\rho}{a-\rho}\phi\right)} \\ &= \frac{(a+\rho)\sin\left(\frac{a-\rho}{a}\Omega t\right) + (a-\rho)\sin\left(\frac{a+\rho}{a}\Omega t\right)}{(a+\rho)\cos\left(\frac{a-\rho}{a}\Omega t\right) - (a-\rho)\cos\left(\frac{a+\rho}{a}\Omega t\right)} \\ &= \frac{\begin{array}{c} a\left[\sin\left(\frac{a-\rho}{a}\Omega t\right) + \sin\left(\frac{a+\rho}{a}\Omega t\right)\right] \\ +\rho\left[\sin\left(\frac{a-\rho}{a}\Omega t\right) - \sin\left(\frac{a+\rho}{a}\Omega t\right)\right] \end{array}}{\begin{array}{c} a\left[\cos\left(\frac{a-\rho}{a}\Omega t\right) - \cos\left(\frac{a+\rho}{a}\Omega t\right)\right] \\ +\rho\left[\cos\left(\frac{a-\rho}{a}\Omega t\right) + \cos\left(\frac{a+\rho}{a}\Omega t\right)\right] \end{array}} \\ &= \frac{2a\sin(\Omega t)\cos\left(\frac{\rho}{a}\Omega t\right) - 2\rho\cos(\Omega t)\sin\left(\frac{\rho}{a}\Omega t\right)}{2a\sin(\Omega t)\sin\left(\frac{\rho}{a}\Omega t\right) + 2\rho\cos(\Omega t)\cos\left(\frac{\rho}{a}\Omega t\right)} \\ &= \frac{a\tan(\Omega t) - \rho\tan\left(\frac{\rho}{a}\Omega t\right)}{a\tan(\Omega t)\tan\left(\frac{\rho}{a}\Omega t\right) + \rho}. \end{aligned} \tag{35}$$

Computing

$$\begin{aligned} \tan\left(\theta + \frac{\rho}{a}\Omega t\right) &= \frac{\begin{array}{r} \left[a\tan(\Omega t) - \rho\tan\left(\frac{\rho}{a}\Omega t\right) + \tan\left(\frac{\rho}{a}\Omega t\right)\right] \\ \times\left[a\tan(\Omega t)\tan\left(\frac{\rho}{a}\Omega t\right) + \rho\right] \end{array}}{\begin{array}{l} \left[a\tan(\Omega t)\tan\left(\frac{\rho}{a}\Omega t\right) + \rho\right] \\ - \left[a\tan(\Omega t) - \rho\tan\left(\frac{\rho}{a}\Omega t\right)\right]\tan\left(\frac{\rho}{a}\Omega t\right) \end{array}} \\ &= \frac{a\tan(\Omega t)\left[1 + \tan^2\left(\frac{\rho}{a}\Omega t\right)\right]}{\rho\left[1 + \tan^2\left(\frac{\rho}{a}\Omega t\right)\right]} \\ &= \frac{a}{\rho}\tan(\Omega t), \end{aligned} \tag{36}$$

then gives

$$\theta = \tan^{-1}\left[\frac{a}{\rho}\tan(\Omega t)\right] - \frac{\rho}{a}\Omega t. \tag{37}$$

Finally, plugging back in gives

$$\begin{aligned}\theta &= \tan^{-1}\left[\frac{a}{\rho}\tan\left(\frac{a}{a-\rho}\phi\right)\right] - \frac{\rho}{a}\frac{a}{a-\rho}\phi \\ &= \tan^{-1}\left[\frac{a}{\rho}\tan\left(\frac{a}{a-\rho}\phi\right)\right] - \frac{\rho}{a-\rho}\phi. \qquad (38)\end{aligned}$$

This form is useful in the solution of the SPHERE WITH TUNNEL problem, which is the generalization of the BRACHISTOCHRONE PROBLEM, to find the shape of a tunnel drilled through a SPHERE (with gravity varying according to Gauss's law for gravitation) such that the travel time between two points on the surface of the SPHERE under the force of gravity is minimized.

see also CYCLOID, EPICYCLOID

References

Bogomolny, A. "Cycloids." `http://www.cut-the-knot.com/pythagoras/cycloids.html`.

Kreyszig, E. *Differential Geometry.* New York: Dover, 1991.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 171–173, 1972.

Lee, X. "Epicycloid and Hypocycloid." `http://www.best.com/~xah/SpecialPlaneCurves_dir/EpiHypocycloid_dir/epiHypocycloid.html`.

MacTutor History of Mathematics Archive. "Hypocycloid." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Hypocycloid.html`.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 225–231, 1979.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 50–52, 1991.

Yates, R. C. "Epi- and Hypo-Cycloids." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 81–85, 1952.

Hypocycloid—3-Cusped

see DELTOID

Hypocycloid—4-Cusped

see ASTROID

Hypocycloid Evolute

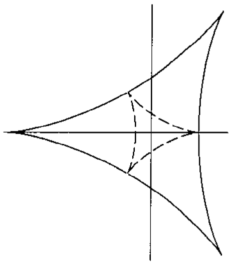

For $x(0) = a$,

$$\begin{aligned}x &= \frac{a}{a-2b}\left[(a-b)\cos\phi - b\cos\left(\frac{a-b}{b}\phi\right)\right] \\ y &= \frac{a}{a-2b}\left[(a-b)\sin\phi + b\sin\left(\frac{a-b}{b}\phi\right)\right].\end{aligned}$$

If $a/b = n$, then

$$\begin{aligned}x &= \frac{1}{n-2}[(n-1)\cos\phi - \cos[(n-1)\phi]a \\ y &= \frac{1}{n-2}[(n-1)\sin\phi + \sin[(n-1)\phi]a.\end{aligned}$$

This is just the original HYPOCYCLOID scaled by the factor $(n-2)/n$ and rotated by $1/(2n)$ of a turn.

Hypocycloid Involute

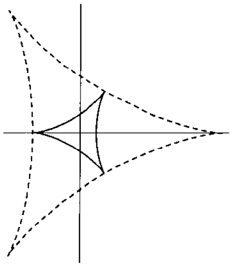

The HYPOCYCLOID

$$\begin{aligned}x &= \frac{a}{a-2b}\left[(a-b)\cos\phi - b\cos\left(\frac{a-b}{b}\phi\right)\right] \\ y &= \frac{a}{a-2b}\left[(a-b)\sin\phi + b\sin\left(\frac{a-b}{b}\phi\right)\right]\end{aligned}$$

has INVOLUTE

$$\begin{aligned}x &= \frac{a-2b}{a}\left[(a-b)\cos\phi + b\cos\left(\frac{a-b}{b}\phi\right)\right] \\ y &= \frac{a-2b}{a}\left[(a-b)\sin\phi - b\sin\left(\frac{a-b}{b}\phi\right)\right],\end{aligned}$$

which is another HYPOCYCLOID.

Hypocycloid Pedal Curve

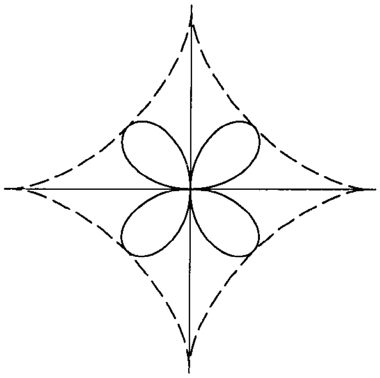

The PEDAL CURVE for a PEDAL POINT at the center is a ROSE.

Hypoellipse

$$y^{n/m} + c\left|\frac{x}{a}\right|^{n/m} - c = 0,$$

with $n/m < 2$. If $n/m > 2$, the curve is a HYPERELLIPSE.

see also ELLIPSE, HYPERELLIPSE, SUPERELLIPSE

References

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 82, 1993.

Hypotenuse

The longest LEG of a RIGHT TRIANGLE (which is the side opposite the RIGHT ANGLE).

Hypothesis

A proposition that is consistent with known data, but has been neither verified nor shown to be false. It is synonymous with CONJECTURE.

see also BOURGET'S HYPOTHESIS, CHINESE HYPOTHESIS, CONTINUUM HYPOTHESIS, HYPOTHESIS TESTING, NESTED HYPOTHESIS, NULL HYPOTHESIS, POSTULATE, RAMANUJAN'S HYPOTHESIS, RIEMANN HYPOTHESIS, SCHINZEL'S HYPOTHESIS, SOUSLIN'S HYPOTHESIS

Hypothesis Testing

The use of statistics to determine the probability that a given hypothesis is true.

see also BONFERRONI CORRECTION, ESTIMATE, FISHER SIGN TEST, PAIRED t-TEST, STATISTICAL TEST, TYPE I ERROR, TYPE II ERROR, WILCOXON SIGNED RANK TEST

References

Hoel, P. G.; Port, S. C.; and Stone, C. J. "Testing Hypotheses." Ch. 3 in *Introduction to Statistical Theory.* New York: Houghton Mifflin, pp. 52–110, 1971.

Iyanaga, S. and Kawada, Y. (Eds.). "Statistical Estimation and Statistical Hypothesis Testing." Appendix A, Table 23 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 1486–1489, 1980.

Shaffer, J. P. "Multiple Hypothesis Testing." *Ann. Rev. Psych.* **46**, 561–584, 1995.

Hypotrochoid

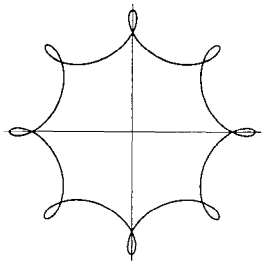

The ROULETTE traced by a point P attached to a CIRCLE of radius b rolling around the inside of a fixed CIRCLE of radius a. The parametric equations for a hypotrochoid are

$$x = n\cos t + h\cos\left(\frac{n}{b}t\right) \tag{1}$$

$$y = n\sin t - h\sin\left(\frac{n}{b}t\right), \tag{2}$$

where $n \equiv a - b$ and h is the distance from P to the center of the rolling CIRCLE. Special cases include the HYPOCYCLOID with $h = b$, the ELLIPSE with $a = 2b$, and the ROSE with

$$a = \frac{2nh}{n+1} \tag{3}$$

$$b = \frac{(n-1)h}{n+1}. \tag{4}$$

see also EPITROCHOID, HYPOCYCLOID, SPIROGRAPH

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 165–168, 1972.

Lee, X. "Hypotrochoid." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Hypotrochoid_dir/hypotrochoid.html`.

Lee, X. "Epitrochoid and Hypotrochoid Movie Gallery." `http://www.best.com/~xah/SpecialPlaneCurves_dir/EpiHypoTMovieGallery_dir/epiHypoTMovieGallery.html`.

MacTutor History of Mathematics Archive. "Hypotrochoid." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Hypotrochoid.html`.

Hypotrochoid Evolute

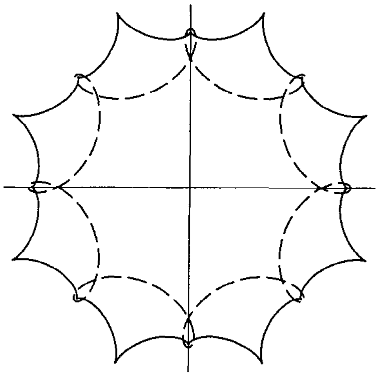

The EVOLUTE of the HYPOTROCHOID is illustrated above.

Hyzer's Illusion

see FREEMISH CRATE

I

i

The IMAGINARY NUMBER i is defined as $i \equiv \sqrt{-1}$. However, for some reason engineers and physicists prefer the symbol j to i. Numbers of the form $z = x + iy$ where x and y are REAL NUMBERS are called COMPLEX NUMBERS, and when z is used to denote a COMPLEX NUMBER, it is sometimes (in older texts) called an "AFFIX."

The SQUARE ROOT of i is

$$\sqrt{i} = \pm\frac{i+1}{\sqrt{2}}, \tag{1}$$

since

$$\left[\frac{1}{\sqrt{2}}(i+1)\right]^2 = \tfrac{1}{2}(i^2 + 2i + 1) = i. \tag{2}$$

This can be immediately derived from the EULER FORMULA with $x = \pi/2$,

$$i = e^{i\pi/2} \tag{3}$$

$$\sqrt{i} = \sqrt{e^{i\pi/2}} = e^{i\pi/4} = \cos(\tfrac{1}{4}\pi) + i\sin(\tfrac{1}{4}\pi) = \frac{1+i}{\sqrt{2}}. \tag{4}$$

The PRINCIPAL VALUE of i^i is

$$i^i = (e^{i\pi/2})^i = e^{i^2\pi/2} = e^{-\pi/2} = 0.207879\ldots. \tag{5}$$

see also COMPLEX NUMBER, IMAGINARY IDENTITY, IMAGINARY NUMBER, REAL NUMBER, SURREAL NUMBER

References

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 89, 1996.

$\mathbb{I}$

see $\mathbb{Z}$

I-Signature

see SIGNATURE (RECURRENCE RELATION)

Iamond

see POLYIAMOND

Ice Fractal

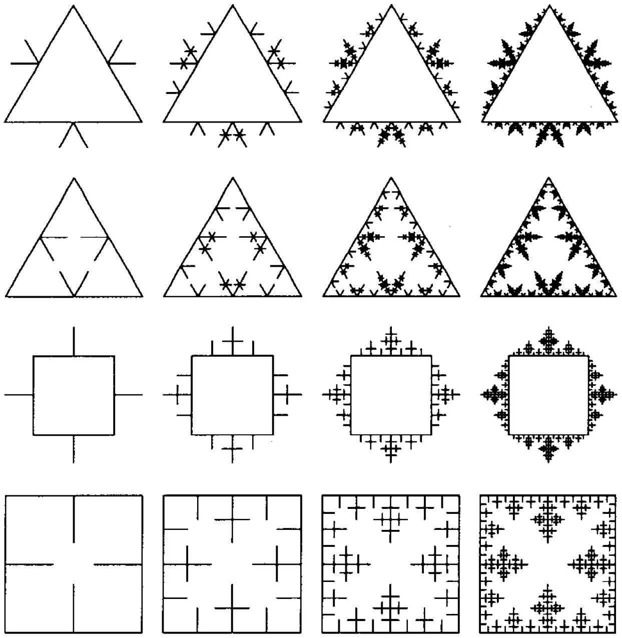

A FRACTAL (square, triangle, etc.) based on a simple generating motif. The above plots show the ice triangle, antitriangle, square, and antisquare. The base curves and motifs for the fractals illustrated above are shown below.

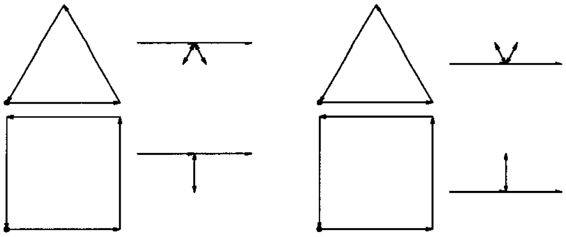

see also FRACTAL

References

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, p. 44, 1991.

Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m`.

Icosagon

A 20-sided POLYGON. The SWASTIKA is an irregular icosagon.

see also SWASTIKA

Icosahedral Equation

Hunt (1996) gives the "dehomogenized" icosahedral equation as

$$[(z^{20}+1) - 228(z^{15} - z^5) + 494z^{10})^3 + 1728uz^5(z^{10} + 11z^5 - 1)^5 = 0.$$

Other forms include

$$I(u,v,Z)=u^5v^5(u^{10}+11u^5v^5-v^{10})^5\\-[u^{30}+v^{30}-10005(u^{20}v^{10}+u^{10}v^{20})\\+522(u^25v^5-u^5v^25)]^2Z=0$$

and

$$I(z,1,Z)=z^5(-1+11z^5+z^{10})^5\\-[1+z^{30}-10005(z^{10}+z^{20})+522(-z^5+z^{25})]^2Z=0.$$

References

Hunt, B. *The Geometry of Some Special Arithmetic Quotients.* New York: Springer-Verlag, p. 146, 1996.

Icosahedral Graph

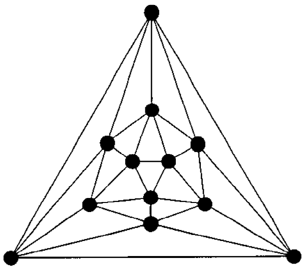

A Polyhedral Graph.

see also Cubical Graph, Dodecahedral Graph, Octahedral Graph, Tetrahedral Graph

Icosahedral Group

The Group I_h of symmetries of the Icosahedron and Dodecahedron. The icosahedral group consists of the symmetry operations E, $12C_5$, $12C_5^2$, $20C_3$, $15C_2$, i, $12S_{10}$, $12S_{10}^3$, $20S_6$, and 15σ (Cotton 1990).

see also Dodecahedron, Icosahedron, Octahedral Group, Tetrahedral Group

References

Cotton, F. A. *Chemical Applications of Group Theory, 3rd ed.* New York: Wiley, p. 48–50, 1990.

Lomont, J. S. "Icosahedral Group." §3.10.E in *Applications of Finite Groups.* New York: Dover, p. 82, 1987.

Icosahedron

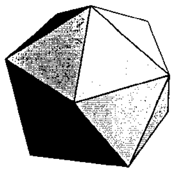

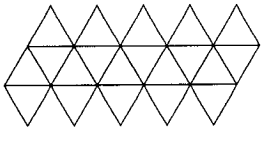

A Platonic Solid (P_5) with 12 Vertices, 30 Edges, and 20 equivalent Equilateral Triangle faces $20\{3\}$. It is described by the Schläfli Symbol $\{3,5\}$. It is also Uniform Polyhedron U_{22} and has Wythoff Symbol $5\,|\,2\,3$. The icosahedron has the Icosahedral Group I_h of symmetries.

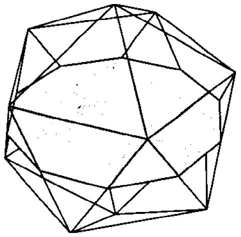

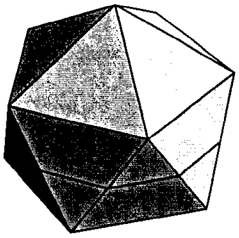

A plane Perpendicular to a C_5 axis of an icosahedron cuts the solid in a regular Decagonal Cross-Section (Holden 1991, pp. 24–25).

A construction for an icosahedron with side length $a=\sqrt{50-10\sqrt{5}}/5$ places the end vertices at $(0,0,\pm1)$ and the central vertices around two staggered Circles of Radii $\frac{2}{5}\sqrt{5}$ and heights $\pm\frac{1}{5}\sqrt{5}$, giving coordinates

$$\pm\left(\tfrac{2}{5}\sqrt{5}\cos(\tfrac{2}{5}i\pi),\tfrac{2}{5}\sqrt{5}\sin(\tfrac{2}{5}i\pi),\tfrac{1}{5}\sqrt{5}\right)\tag{1}$$

for $i=0, 1, \ldots, 4$, where all the plus signs or minus signs are taken together. Explicitly, these coordinates are

$$\mathbf{x}_0^\pm=\pm(\tfrac{2}{5}\sqrt{5},0,\tfrac{1}{5}\sqrt{5})\tag{2}$$

$$\mathbf{x}_1^\pm=\pm(\tfrac{1}{10}(5-\sqrt{5}),\tfrac{1}{10}\sqrt{50+10\sqrt{5}},\tfrac{1}{5}\sqrt{5})\tag{3}$$

$$\mathbf{x}_2^\pm=\pm(-\tfrac{1}{10}(\sqrt{5}+5),\tfrac{1}{10}\sqrt{50-10\sqrt{5}},\tfrac{1}{5}\sqrt{5})\tag{4}$$

$$\mathbf{x}_3^\pm=\pm(-\tfrac{1}{10}(\sqrt{5}-5),-\tfrac{1}{10}\sqrt{50-10\sqrt{5}},\tfrac{1}{5}\sqrt{5})\tag{5}$$

$$\mathbf{x}_4^\pm=\pm(\tfrac{1}{10}(5-\sqrt{5}),-\tfrac{1}{10}\sqrt{50+10\sqrt{5}},\tfrac{1}{5}\sqrt{5}).\tag{6}$$

By a suitable rotation, the Vertices of an icosahedron of side length 2 can also be placed at $(0,\pm\phi,\pm1)$, $(\pm1,0,\pm\phi)$, and $(\pm\phi,\pm1,0)$, where ϕ is the Golden Ratio. These points divide the Edges of an Octahedron into segments with lengths in the ratio $\phi:1$.

The Dual Polyhedron of the icosahedron is the Dodecahedron. There are 59 distinct icosahedra when each Triangle is colored differently (Coxeter 1969).

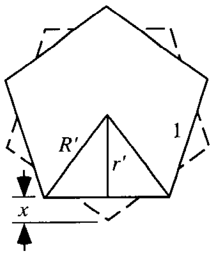

To derive the Volume of an icosahedron having edge length a, consider the orientation so that two Vertices are oriented on top and bottom. The vertical distance between the top and bottom Pentagonal Dipyramids is then given by

$$z=\sqrt{\ell^2-x^2},\tag{7}$$

where

$$\ell=\tfrac{1}{2}\sqrt{3}\,a\tag{8}$$

is the height of an Isosceles Triangle, and the Sagitta $x \equiv R' - r'$ of the pentagon is

$$x = \tfrac{1}{2} a \tfrac{1}{10} \sqrt{25 - 10\sqrt{5}}\, a, \tag{9}$$

giving

$$x^2 = \tfrac{1}{20} \sqrt{5 - 2\sqrt{5}}\, a^2. \tag{10}$$

Plugging (8) and (10) into (7) gives

$$\begin{aligned} z &= a\sqrt{\tfrac{3}{4} - \tfrac{1}{20}(5 - 2\sqrt{5})} = a\sqrt{\frac{15 - (5 - 2\sqrt{5})}{20}} \\ &= a\sqrt{\frac{10 + 2\sqrt{5}}{20}} = \tfrac{1}{2} a \sqrt{\frac{10 + 2\sqrt{5}}{5}} \\ &= \tfrac{1}{10} \sqrt{50 + 10\sqrt{5}}\, a, \end{aligned} \tag{11}$$

which is identical to the radius of a Pentagon of side a. The Circumradius is then

$$R = h + \tfrac{1}{2} z, \tag{12}$$

where

$$h = \tfrac{1}{10} \sqrt{50 - 10\sqrt{5}}\, a \tag{13}$$

is the height of a Pentagonal Dipyramid. Therefore,

$$\begin{aligned} R^2 &= (h + \tfrac{1}{2} z)^2 \\ &= (\tfrac{1}{10} \sqrt{50 - 10\sqrt{5}} + \tfrac{1}{20} \sqrt{50 + 10\sqrt{5}}\,)^2 a^2 \\ &= \left(\frac{5}{8} - \frac{3}{8\sqrt{5}} + \frac{\sqrt{20}}{10}\right) a^2 = \tfrac{1}{8}(5 + \sqrt{5}\,)a. \end{aligned} \tag{14}$$

Taking the square root gives the Circumradius

$$R = \sqrt{\tfrac{1}{8}(5 + \sqrt{5})}\, a = \tfrac{1}{4} \sqrt{10 + 2\sqrt{5}}\, a \approx 0.95105a. \tag{15}$$

The Inradius is

$$r = \tfrac{1}{12}(3\sqrt{3} + \sqrt{15}\,)a \approx 0.75576a. \tag{16}$$

The square of the Interradius is

$$\begin{aligned} \rho^2 &= (\tfrac{1}{2} z)^2 + {x_l}^2 \\ &= [(\tfrac{1}{4})(\tfrac{1}{100})(50 + 10\sqrt{5}\,) + \tfrac{1}{100}(25 + 10\sqrt{5}\,)]a^2 \\ &= \tfrac{1}{8}(3 + \sqrt{5})a^2, \end{aligned} \tag{17}$$

so

$$\rho = \sqrt{\tfrac{1}{8}(3 + \sqrt{5}\,)}\, a = \tfrac{1}{4}(1 + \sqrt{5}\,)a \approx 0.80901a. \tag{18}$$

The Area of one face is the Area of an Equilateral Triangle

$$A = \tfrac{1}{4} a^2 \sqrt{3}. \tag{19}$$

The volume can be computed by taking 20 pyramids of height r

$$\begin{aligned} V &= 20[(\tfrac{1}{3} A)r] = 20 \tfrac{1}{3} \tfrac{1}{4} \sqrt{3}\, a^2 \tfrac{1}{12}(3\sqrt{3} + \sqrt{15}\,)a \\ &= \tfrac{5}{12}(3 + \sqrt{5}\,)a^3. \end{aligned} \tag{20}$$

Apollonius showed that

$$\frac{V_{\text{icosahedron}}}{V_{\text{dodecahedron}}} = \frac{A_{\text{icosahedron}}}{A_{\text{dodecahedron}}}, \tag{21}$$

where V is the volume and A the surface area.

see also Augmented Tridiminished Icosahedron, Decagon, Dodecahedron, Great Icosahedron, Icosahedron Stellations, Metabidiminished Icosahedron, Tridiminished Icosahedron, Trigonometry Values—$\pi/5$

References

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, 1969.

Davie, T. "The Icosahedron." `http://www.dcs.st-and.ac.uk/~ad/mathrecs/polyhedra/icosahedron.html`.

Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, 1991.

Klein, F. *Lectures on the Icosahedron.* New York: Dover, 1956.

Pappas, T. "The Icosahedron & the Golden Rectangle." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 115, 1989.

Icosahedron Stellations

Applying the Stellation process to the Icosahedron gives

$$20 + 30 + 60 + 20 + 60 + 120 + 12 + 30 + 60 + 60$$

cells of ten different shapes and sizes in addition to the Icosahedron itself. After application of five restrictions due to J. C. P. Miller to define which forms should be considered distinct, 59 stellations are found to be possible. Miller's restrictions are

1. The faces must lie in the twenty bounding planes of the icosahedron.
2. The parts of the faces in the twenty planes must be congruent, but those parts lying in one place may be disconnected.
3. The parts lying in one plane must have threefold rotational symmetry with or without reflections.
4. All parts must be accessible, i.e., lie on the outside of the solid.
5. Compounds are excluded that can be divided into two sets, each of which has the full symmetry of the whole.

Of these, 32 have full icosahedral symmetry and 27 are Enantiomeric forms. Four are Polyhedron Compounds, one is a Kepler-Poinsot Solid, and one is the Dual Polyhedron of an Archimedean Solid.

The only STELLATIONS of PLATONIC SOLIDS which are UNIFORM POLYHEDRA are the three DODECAHEDRON STELLATIONS the GREAT ICOSAHEDRON (stellation # 11).

n	name
1	icosahedron
2	triakisicosahedron
3	octahedron 5-compound
4	echidnahedron
11	great icosahedron
18	tetrahedron 10-compound
20	deltahedron-60
36	tetrahedron 5-compound

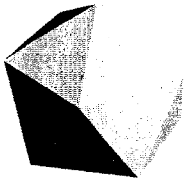
01

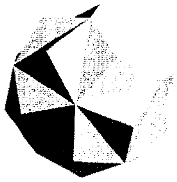
02

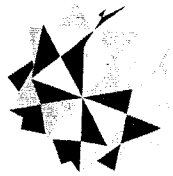
03

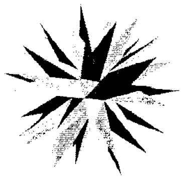
04

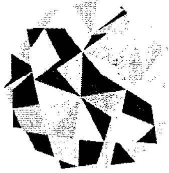
05

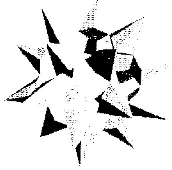
06

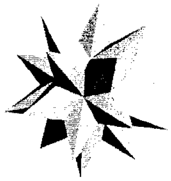
07

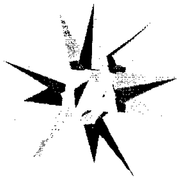
08

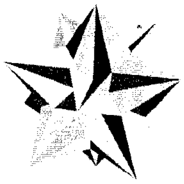
09

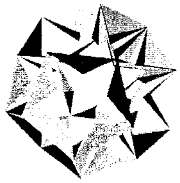
10

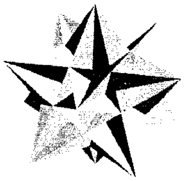
11

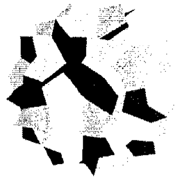
12

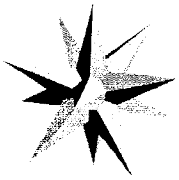
13

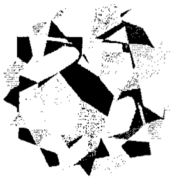
14

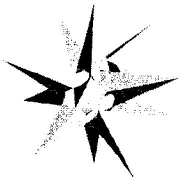
15

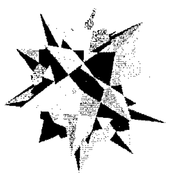
16

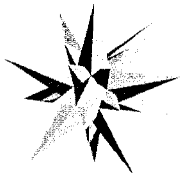
17

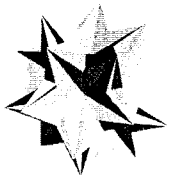
18

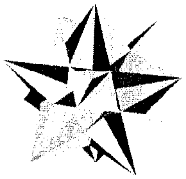
19

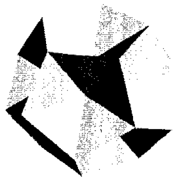
20

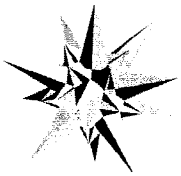
21

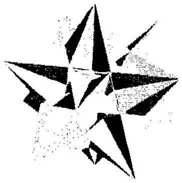
22

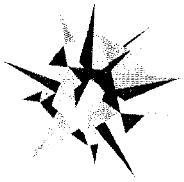
23

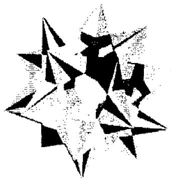
24

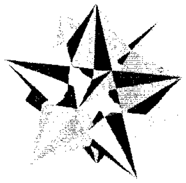
25

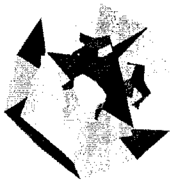
26

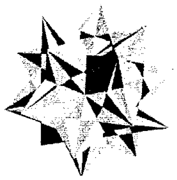
27

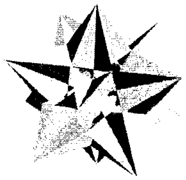
28

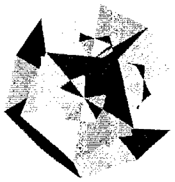
29

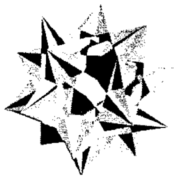
30

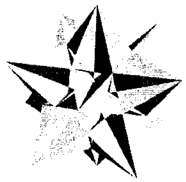
31

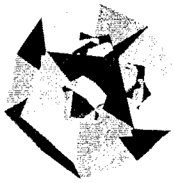
32

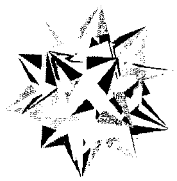
33

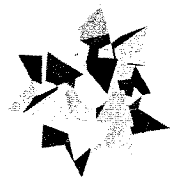
34

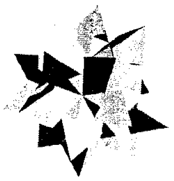
35

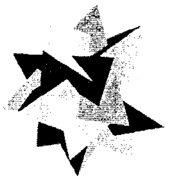
36

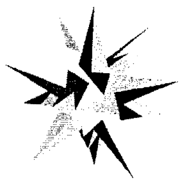
37

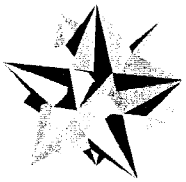
38

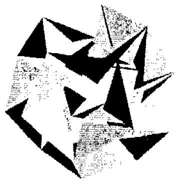
39

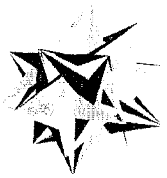
40

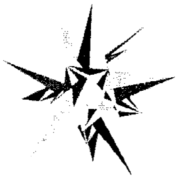
41

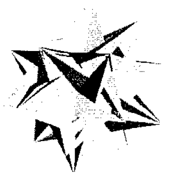
42

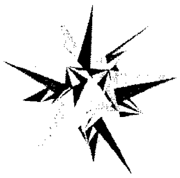
43

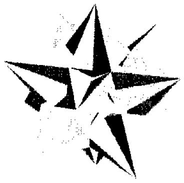
44

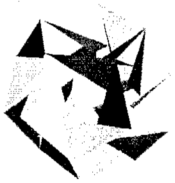
45

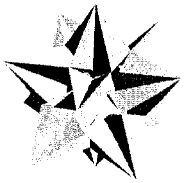
46

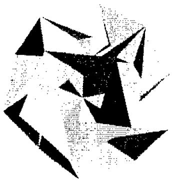
47

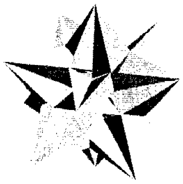
48

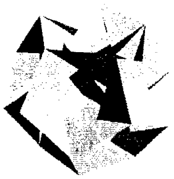
49

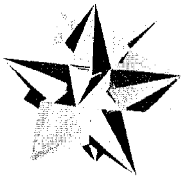
50

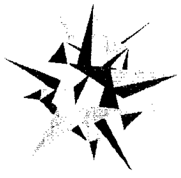
51

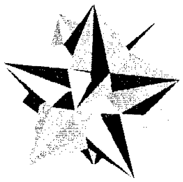
52

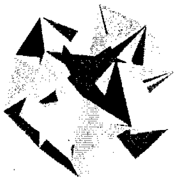
53

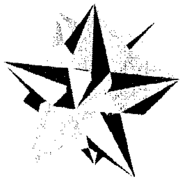
54

55

56

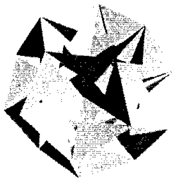
57

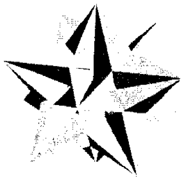
58

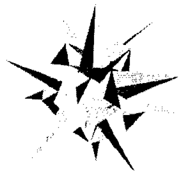
59

see also ARCHIMEDEAN SOLID STELLATION, DODECAHEDRON STELLATIONS, STELLATION

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 146–147, 1987.

Bulatov, V. "Stellations of Icosahedron." `http://www.physics.orst.edu/~bulatov/polyhedra/icosahedron/`.

Coxeter, H. S. M. *The Fifty-Nine Icosahedra.* New York: Springer-Verlag, 1982.

Hart, G. W. "59 Stellations of the Icosahedron." `http://www.li.net/~george/virtual-polyhedra/stellations-icosahedron-index.html`.

Maeder, R. E. `Icosahedra.m` notebook. `http://www.inf.ethz.ch/department/TI/rm/programs.html`.

Maeder, R. E. "The Stellated Icosahedra." *Mathematica in Education* **3**, 1994. `ftp://ftp.inf.ethz.ch/doc/papers/ti/scs/icosahedra94.ps.gz`.

Maeder, R. E. "Stellated Icosahedra." `http://www.mathconsult.ch/showroom/icosahedra/`.

Wang, P. "Polyhedra." `http://www.ugcs.caltech.edu/~peterw/portfolio/polyhedra/`.

Wenninger, M. J. *Polyhedron Models.* New York: Cambridge University Press, pp. 41–65, 1989.

Wheeler, A. H. "Certain Forms of the Icosahedron and a Method for Deriving and Designating Higher Polyhedra." *Proc. Internat. Math. Congress* **1**, 701–708, 1924.

Icosian Game

The problem of finding a HAMILTONIAN CIRCUIT along the edges of an ICOSAHEDRON, i.e., a path such that every vertex is visited a single time, no edge is visited twice, and the ending point is the same as the starting point.

see also HAMILTONIAN CIRCUIT, ICOSAHEDRON

References

Herschel, A. S. "Sir Wm. Hamilton's Icosian Game." *Quart. J. Pure Applied Math.* **5**, 305, 1862.

Icosidodecadodecahedron

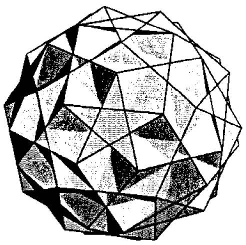

The UNIFORM POLYHEDRON U_{44} whose DUAL POLYHEDRON is the MEDIAL ICOSACRONIC HEXECONTAHEDRON. It has WYTHOFF SYMBOL $\frac{5}{3}\,5\,|\,3$. Its faces are $20\{6\}+12\{\frac{5}{2}\}+12\{5\}$. Its CIRCUMRADIUS for unit edge length is

$$R=\tfrac{1}{2}\sqrt{7}.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 128–129, 1989.

Icosidodecahedron

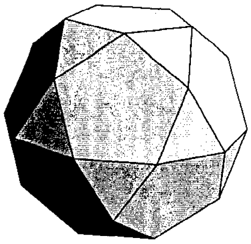

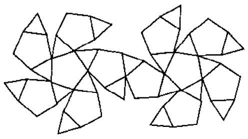

An ARCHIMEDEAN SOLID whose DUAL POLYHEDRON is the RHOMBIC TRIACONTAHEDRON. It is one of the two convex QUASIREGULAR POLYHEDRA and has SCHLÄFLI SYMBOL $\left\{\begin{matrix}3\\5\end{matrix}\right\}$. It is also UNIFORM POLYHEDRON U_{24} and has WYTHOFF SYMBOL $2\,|\,3\,5$. Its faces are $20\{3\}+12\{5\}$. The VERTICES of an icosidodecahedron of EDGE length $2\phi^{-1}$ are $(\pm2,0,0)$, $(0,\pm2,0)$, $(0,0,\pm2)$, $(\pm1,\pm\phi^{-1},\pm1)$, $(\pm1,\pm\phi,\pm\phi^{-1})$, $(\pm\phi^{-1},\pm1,\pm\phi)$. The 30 VERTICES of an OCTAHEDRON 5-COMPOUND form an icosidodecahedron (Ball and Coxeter 1987). FACETED versions include the SMALL ICOSIHEMIDODECAHEDRON and SMALL DODECAHEMIDODECAHEDRON.

The faces of the icosidodecahedron consist of 20 triangles and 12 pentagons. Furthermore, its 60 edges are bisected perpendicularly by those of the reciprocal RHOMBIC TRIACONTAHEDRON (Ball and Coxeter 1987).

The INRADIUS, MIDRADIUS, and CIRCUMRADIUS for unit edge length are

$$\begin{aligned} r &= \tfrac{1}{8}(5+3\sqrt{5}\,) \approx 1.46353 \\ \rho &= \tfrac{1}{2}\sqrt{5+2\sqrt{5}} \approx 1.53884 \\ R &= \tfrac{1}{2}(1+\sqrt{5}\,) = \phi \approx 1.61803. \end{aligned}$$

see also ARCHIMEDEAN SOLID, GREAT ICOSIDODECAHEDRON, QUASIREGULAR POLYHEDRON, SMALL ICOSIHEMIDODECAHEDRON, SMALL DODECAHEMIDODECAHEDRON

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 137, 1987.
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 73, 1989.

Icosidodecahedron Stellation

The first stellation is a DODECAHEDRON-ICOSAHEDRON COMPOUND.

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 73–96, 1989.

Icosidodecatruncated Icosidodecahedron

see ICOSITRUNCATED DODECADODECAHEDRON

Icositruncated Dodecadodecahedron

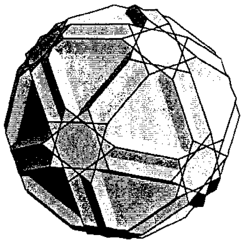

The UNIFORM POLYHEDRON U_{45} also called the ICOSIDODECATRUNCATED ICOSIDODECAHEDRON whose DUAL POLYHEDRON is the TRIDYAKIS ICOSAHEDRON. It has WYTHOFF SYMBOL $3\,\frac{5}{3}\,5\,|$. Its faces are $20\{6\}+12\{10\}+12\{\frac{10}{3}\}$. Its CIRCUMRADIUS for unit edge length is

$$R = 2.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 130–131, 1989.

Ida Surface

A 3-D shadow of a 4-D KLEIN BOTTLE.

see also KLEIN BOTTLE

References
Peterson, I. *Islands of Truth: A Mathematical Mystery Cruise.* New York: W. H. Freeman, pp. 44–45, 1990.

Ideal

A subset I of elements in a RING R which forms an additive GROUP and has the property that, whenever x belongs to R and y belongs to I, then xy and yx belong to I. For example, the set of EVEN INTEGERS is an ideal in the RING of INTEGERS. Given an ideal I, it is possible to define a FACTOR RING R/I.

An ideal may be viewed as a lattice and specified as the finite list of algebraic integers that form a basis for the lattice. Any two bases for the same lattice are equivalent. Ideals have multiplication, and this is basically the Kronecker product of the two bases.

For any ideal I, there is an ideal I_i such that

$$II_i = z,$$

where z is a PRINCIPAL IDEAL, (i.e., an ideal of rank 1). Moreover there is a finite list of ideals I_i such that this equation may be satisfied for every I. The size of this list is known as the CLASS NUMBER. In effect, the above relation imposes an EQUIVALENCE RELATION on ideals, and the number of ideals modulo this relation is the class number. When the CLASS NUMBER is 1, the corresponding number RING has unique factorization and, in a sense, the class number is a measure of the failure of unique factorization in the original number ring.

Dedekind (1871) showed that every NONZERO ideal in the domain of INTEGERS of a FIELD is a unique product of PRIME IDEALS.

see also CLASS NUMBER, DIVISOR THEORY, IDEAL NUMBER, MAXIMAL IDEAL, PRIME IDEAL, PRINCIPAL IDEAL

References

Malgrange, B. *Ideals of Differentiable Functions.* London: Oxford University Press, 1966.

Ideal Number

A type of number involving the ROOTS OF UNITY which was developed by Kummer while trying to solve FERMAT'S LAST THEOREM. Although factorization over the INTEGERS is unique (the FUNDAMENTAL THEOREM OF ALGEBRA), factorization is *not* unique over the COMPLEX NUMBERS. Over the ideal numbers, however, factorization in terms of the COMPLEX NUMBERS becomes unique. Ideal numbers were so powerful that they were generalized by Dedekind into the more abstract IDEALS in general RINGS which are a key part of modern abstract ALGEBRA.

see also DIVISOR THEORY, FERMAT'S LAST THEOREM, IDEAL

Ideal (Partial Order)

An ideal I of a PARTIAL ORDER P is a subset of the elements of P which satisfy the property that if $y \in I$ and $x < y$, then $x \in I$. For k disjoint chains in which the ith chain contains n_i elements, there are $(1+n_1)(1+n_2)\cdots(1+n_k)$ ideals. The number of ideals of a n-element FENCE POSET is the FIBONACCI NUMBER F_n.

References

Ruskey, F. "Information on Ideals of Partially Ordered Sets." `http://sue.csc.uvic.ca/~cos/inf/pose/Ideals.html`.

Steiner, G. "An Algorithm to Generate the Ideals of a Partial Order." *Operat. Res. Let.* **5**, 317–320, 1986.

Ideal Point

A type of POINT AT INFINITY in which parallel lines in the HYPERBOLIC PLANE intersect at infinity in one direction, while diverging from one another in the other.

see also HYPERPARALLEL

Idele

The multiplicative subgroup of all elements in the product of the multiplicative groups $k_\nu^\times$ whose absolute value is 1 at all but finitely many ν, where k is a number FIELD and ν a PLACE.

see also ADÉLE

References

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Idemfactor

see DYADIC

Idempotent

An OPERATOR $\tilde{A}$ such that $\tilde{A}^2 = \tilde{A}$ or an element of an ALGEBRA x such that $x^2 = x$.

see also AUTOMORPHIC NUMBER, BOOLEAN ALGEBRA, GROUP, SEMIGROUP

Identity

An identity is a mathematical relationship equating one quantity to another (which may initially appear to be different).

see also ABEL'S IDENTITY, ANDREWS-SCHUR IDENTITY, BAC-CAB IDENTITY, BEAUZAMY AND DÉGOT'S IDENTITY, BELTRAMI IDENTITY, BIANCHI IDENTITIES, BOCHNER IDENTITY, BRAHMAGUPTA IDENTITY, CASSINI'S IDENTITY, CAUCHY-LAGRANGE IDENTITY, CHRISTOFFEL-DARBOUX IDENTITY, CHU-VANDERMONDE IDENTITY, DE MOIVRE'S IDENTITY, DOUGALL-RAMANUJAN IDENTITY, EULER FOUR-SQUARE IDENTITY, EULER IDENTITY, EULER POLYNOMIAL IDENTITY, FERRARI'S IDENTITY, FIBONACCI IDENTITY, FROBENIUS TRIANGLE IDENTITIES, GREEN'S IDENTITIES, HYPERGEOMETRIC IDENTITY, IMAGINARY IDENTITY, JACKSON'S IDENTITY, JACOBI IDENTITIES, JACOBI'S DETERMINANT IDENTITY, LAGRANGE'S IDENTITY, LE CAM'S IDENTITY, LEIBNIZ IDENTITY, LIOUVILLE POLYNOMIAL IDENTITY, MATRIX POLYNOMIAL IDENTITY, MORGADO IDENTITY, NEWTON'S IDENTITIES, QUINTUPLE PRODUCT IDENTITY, RAMANUJAN 6-10-8 IDENTITY, RAMANUJAN COS/COSH IDENTITY, RAMANUJAN'S IDENTITY, RAMANUJAN'S SUM IDENTITY, REZNIK'S IDENTITY, ROGERS-RAMANUJAN IDENTITIES, SCHAAR'S IDENTITY, STREHL IDENTITY, SYLVESTER'S DETERMINANT IDENTITY, TRINOMIAL IDENTITY, VISIBLE POINT VECTOR IDENTITY, WATSON QUINTUPLE PRODUCT IDENTITY, WORPITZKY'S IDENTITY

Identity Element

The identity element I (also denoted E, e, or 1) of a GROUP or related mathematical structure S is the unique elements such that $IA = AI = I$ for every element $A \in S$. The symbol "E" derives from the German word for unity, "Einheit."

see also BINARY OPERATOR, GROUP, INVOLUTION (GROUP), MONOID

Identity Function

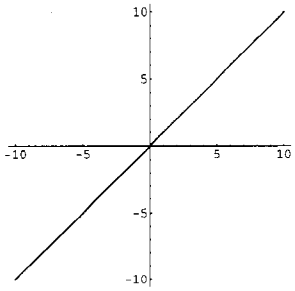

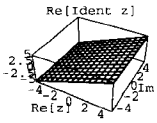

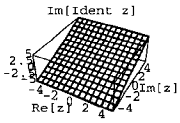

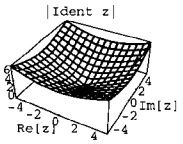

The function $f(x) = x$ which assigns every Real Number x to the same Real Number x. It is identical to the Identity Map.

Identity Map

The Map which assigns every Real Number to the same Real Number $\mathrm{id}_{\mathbb{R}}$. It is identical to the Identity Function.

Identity Matrix

The identity matrix is defined as the Matrix I (or I) such that

$$\mathsf{I}(\mathbf{X}) \equiv \mathbf{X}$$

for all Vectors $\mathbf{X}$. The identity matrix is

$$I_{ij} = \delta_{ij}$$

for $i, j = 1, 2, \ldots, n$, where δ_{ij} is the Kronecker Delta. Written explicitly,

$$\mathsf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Identity Operator

The Operator $\tilde{I}$ which takes a Real Number to the same Real Number $\tilde{I}r = r$.

see also Identity Function, Identity Map

Idoneal Number

A Positive value of D for which the fact that a number is a Monomorph (i.e., the number is expressible in only one way as $x^2 + Dy^2$ or $x^2 - Dy^2$ where x^2 is Relatively Prime to Dy^2) guarantees it to be a Prime, Power of a Prime, or twice one of these. The numbers are also called Euler's Idoneal Numbers, or Suitable Numbers.

The 65 idoneal numbers found by Gauss and Euler and conjectured to be the *only* such numbers (Shanks 1969) are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 15, 16, 18, 21, 22, 24, 25, 28, 30, 33, 37, 40, 42, 45, 48, 57, 58, 60, 70, 72, 78, 85, 88, 93, 102, 105, 112, 120, 130, 133, 165, 168, 177, 190, 210, 232, 240, 253, 273, 280, 312, 330, 345, 357, 385, 408, 462, 520, 760, 840, 1320, 1365, and 1848 (Sloane's A000926).

References

Shanks, D. "On Gauss's Class Number Problems." *Math. Comput.* **23**, 151–163, 1969.

Sloane, N. J. A. Sequence A000926/M0476 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Iff

If and only if (i.e., Necessary and Sufficient). The terms "Just If" or "Exactly When" are sometimes used instead. A iff B is written symbolically as $A \leftrightarrow B$. A iff B is also equivalent to $A \Rightarrow B$, together with $B \Rightarrow A$, where the symbol $\Rightarrow$ denotes "Implies."

J. H. Conway believes that the word originated with P. Halmos and was transmitted through Kelley (1975). Halmos has stated, "To the best of my knowledge, I DID invent the silly thing, but I wouldn't swear to it in a court of law. So there—give me credit for it anyway" (Asimov 1997).

see also Equivalent, Exactly One, Implies, Necessary, Sufficient

References

Asimov, D. "Iff." math-fun@cs.arizona.edu posting, Sept. 19, 1997.

Kelley, J. L. *General Topology.* New York: Springer-Verlag, 1975.

Ill-Conditioned

A system is ill-conditioned if the Condition Number is too large (and singular if it is Infinite).

see also Condition Number

Illumination Problem

In the early 1950s, Ernst Straus asked

1. Is every Polygonal region illuminable from every point in the region?
2. Is every Polygonal region illuminable from at least one point in the region?

Here, illuminable means that there is a path from every point to every other by repeated reflections. Tokarsky (1995) showed that unilluminable rooms exist in the plane and 3-D, but question (2) remains open. The smallest known counterexample to (1) in the Plane has 26 sides.

see also Art Gallery Theorem

References

Klee, V. "Is Every Polygonal Region Illuminable from Some Point?" *Math. Mag.* **52**, 180, 1969.

Tokarsky, G. W. "Polygonal Rooms Not Illuminable from Every Point." *Amer. Math. Monthly* **102**, 867–879, 1995.

Illusion

An object or drawing which appears to have properties which are physically impossible, deceptive, or counterintuitive.

see also Benham's Wheel, Freemish Crate, Goblet Illusion, Hermann Grid Illusion, Hermann-Hering Illusion, Hyzer's Illusion, Impossible Figure, Irradiation Illusion, Kanizsa Triangle, Müller-Lyer Illusion, Necker Cube, Orbison's Illusion, Parallelogram Illusion, Penrose

Stairway, Poggendorff Illusion, Ponzo's Illusion, Rabbit-Duck Illusion, Tribar, Vertical-Horizontal Illusion, Young Girl-Old Woman Illusion, Zollner's Illusion

References
Ausbourne, B. "A Sensory Adventure." `http://www.lainet.com/illusions/`.
Ausbourne, B. "Optical Illusions: A Collection." `http://www.lainet.com/~ausbourn/`.
Ernst, B. *Optical Illusions.* New York: Taschen, 1996.
Fineman, M. *The Nature of Visual Illusion.* New York: Dover, 1996.
Gardner, M. "Optical Illusions." Ch. 1 in *Mathematical Circus: More Puzzles, Games, Paradoxes and Other Mathematical Entertainments from Scientific American.* New York: Knopf, 1979.
Gregory, R. L. *Eye and Brain, 5th ed.* Princeton, NJ: Princeton University Press, 1997.
"Illusions: Central Station." `http://www.heureka.fi/i/Illusions_ctrl_station.html.en`.
Landrigad, D. "Gallery of Illusions." `http://valley.uml.edu/psychology/illusion.html`.
Luckiesh, M. *Visual Illusions: Their Causes, Characteristics, and Applications.* New York: Dover, 1965.
Pappas, T. "History of Optical Illusions." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 172–173, 1989.
Tolansky, S. *Optical Illusions.* New York: Pergamon Press, 1964.

Image

see Range (Image)

Imaginary Identity

see i

Imaginary Number

A Complex Number which has zero Real Part, so that it can be written as a Real Number multiplied by the "imaginary unit" i (equal to $\sqrt{-1}$).

see also Complex Number, Galois Imaginary, Gaussian Integer, i, Real Number

References
Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 211–216, 1996.

Imaginary Part

The imaginary part $\Im$ of a Complex Number $z = x+iy$ is the Real Number multiplying i, so $\Im[x+iy] = y$. In terms of z itself,

$$\Im[z] = \frac{z - z^*}{2i},$$

where z^* is the Complex Conjugate of z.

see also Absolute Square, Complex Conjugate, Real Part

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 16, 1972.

Imaginary Point

A pair of values x and y one or both of which is Complex.

References
Woods, F. S. *Higher Geometry: An Introduction to Advanced Methods in Analytic Geometry.* New York: Dover, p. 2, 1961.

Imaginary Quadratic Field

A Quadratic Field $\mathbb{Q}(\sqrt{D})$ with $D < 0$.

see also Quadratic Field

Immanant

For an $n \times n$ matrix, let S denote any permutation $e_1, e_2, \ldots, e_n$ of the set of numbers $1, 2, \ldots, n$, and let $\chi^{(\lambda)}(S)$ be the character of the symmetric group corresponding to the partition (λ). Then the immanant $|a_{mn}|^{(\lambda)}$ is defined as

$$|a_{mn}|^{(\lambda)} = \sum \chi^{(\lambda)}(S) P_S$$

where the summation is over the $n!$ permutations of the Symmetric Group and

$$P_s = a_{1e_1} a_{2e_2} \cdots a_{ne_n}.$$

see also Determinant, Permanent

References
Littlewood, D. E. and Richardson, A. R. "Group Characters and Algebra." *Philos. Trans. Roy. Soc. London A* **233**, 99–141, 1934.
Littlewood, D. E. and Richardson, A. R. "Immanants of Some Special Matrices." *Quart. J. Math. (Oxford)* **5**, 269–282, 1934.
Wybourne, B. G. "Immanants of Matrices." §2.19 in *Symmetry Principles and Atomic Spectroscopy.* New York: Wiley, pp. 12–13, 1970.

Immersed Minimal Surface

see Enneper's Surfaces

Immersion

A special nonsingular Map from one Manifold to another such that at every point in the domain of the map, the Derivative is an injective linear map. This is equivalent to saying that every point in the Domain has a Neighborhood such that, up to Diffeomorphisms of the Tangent Space, the map looks like the inclusion map from a lower-dimensional Euclidean Space to a higher-dimensional Euclidean Space.

see also Boy Surface, Eversion, Smale-Hirsch Theorem

References
Boy, W. "Über die Curvatura integra und die Topologie geschlossener Flächen." *Math. Ann* **57**, 151–184, 1903.
Pinkall, U. "Models of the Real Projective Plane." Ch. 6 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 63–67, 1986.

Impartial Game

A GAME in which the possible moves are the same for each player in any position. All positions in all impartial GAMES form an additive ABELIAN GROUP. For impartial games in which the last player wins (normal form games), the nim-value of the sum of two GAMES is the nim-sum of their nim-values. If the last player loses, the GAME is said to be in misère form and the analysis is much more difficult.

see also FAIR GAME, GAME, PARTISAN GAME

Implicit Function

A function which is not defined explicitly, but rather is defined in terms of an algebraic relationship (which can not, in general, be "solved" for the function in question). For example, the ECCENTRIC ANOMALY E of a body orbiting on an ELLIPSE with ECCENTRICITY e is defined implicitly in terms of the mean anomaly M by KEPLER'S EQUATION

$$M = E - e\sin E.$$

Implicit Function Theorem

Given

$$F_1(x,y,z,u,v,w) = 0$$

$$F_2(x,y,z,u,v,w) = 0$$

$$F_3(x,y,z,u,v,w) = 0,$$

if the JACOBIAN

$$JF(u,v,w) = \frac{\partial(F_1,F_2,F_3)}{\partial(u,v,w)} \neq 0,$$

then u, v, and w can be solved for in terms of x, y, and z and PARTIAL DERIVATIVES of u, v, w with respect to x, y, and z can be found by differentiating implicitly.

More generally, let A be an OPEN SET in $\mathbb{R}^{n+k}$ and let $f : A \to \mathbb{R}^n$ be a C^r FUNCTION. Write f in the form $f(x,y)$, where x and y are elements of $\mathbb{R}^k$ and $\mathbb{R}^n$. Suppose that (a, b) is a point in A such that $f(a,b) = 0$ and the DETERMINANT of the $n \times n$ MATRIX whose elements are the DERIVATIVES of the n component FUNCTIONS of f with respect to the n variables, written as y, evaluated at (a,b), is not equal to zero. The latter may be rewritten as

$$\operatorname{rank}(Df(a,b)) = n.$$

Then there exists a NEIGHBORHOOD B of a in $\mathbb{R}^k$ and a unique C^r FUNCTION $g : B \to \mathbb{R}^n$ such that $g(a) = b$ and $f(x,g(x)) = 0$ for all $x \in B$.

see also CHANGE OF VARIABLES THEOREM, JACOBIAN

References

Munkres, J. R. *Analysis on Manifolds.* Reading, MA: Addison-Wesley, 1991.

Implies

The symbol $\Rightarrow$ means "implies" in the mathematical sense. Let A be true. If this implies that B is also true, then the statement is written symbolically as $A \Rightarrow B$, or sometimes $A \subset B$. If $A \Rightarrow B$ and $B \Rightarrow A$ (i.e, $A \Rightarrow B \wedge B \Rightarrow A$), then A and B are said to be EQUIVALENT, a relationship which is written symbolically as $A \Leftrightarrow B$ or $A \rightleftharpoons B$.

see also EQUIVALENT

Impossible Figure

A class of ILLUSION in which an object which is physically unrealizable is apparently depicted.

see also FREEMISH CRATE, HOME PLATE, ILLUSION, NECKER CUBE, PENROSE STAIRWAY, TRIBAR

References

Cowan, T. M. "The Theory of Braids and the Analysis of Impossible Figures." *J. Math. Psych.* **11**, 190–212, 1974.

Cowan, T. M. "Supplementary Report: Braids, Side Segments, and Impossible Figures." *J. Math. Psych.* **16**, 254–260, 1977.

Cowan, T. M. "Organizing the Properties of Impossible Figures." *Perception* **6**, 41–56, 1977.

Cowan, T. M. and Pringle, R. "An Investigation of the Cues Responsible for Figure Impossibility." *J. Exper. Psych. Human Perception Performance* **4**, 112–120, 1978.

Ernst, B. *Adventures with Impossible Figures.* Stradbroke, England: Tarquin, 1987.

Harris, W. F. "Perceptual Singularities in Impossible Pictures Represent Screw Dislocations." *South African J. Sci.* **69**, 10–13, 1973.

Fineman, M. *The Nature of Visual Illusion.* New York: Dover, pp. 119–122, 1996.

Jablan, S. "Impossible Figures." `http://members.tripod.com/~modularity/impos.htm` and "Are Impossible Figures Possible?" `http://members.tripod.com/~modularity/kulpa.htm`.

Kulpa, Z. "Are Impossible Figures Possible?" *Signal Processing* **5**, 201–220, 1983.

Kulpa, Z. "Putting Order in the Impossible." *Perception* **16**, 201–214, 1987.

Sugihara, K. "Classification of Impossible Objects." *Perception* **11**, 65–74, 1982.

Terouanne, E. "Impossible Figures and Interpretations of Polyhedral Figures." *J. Math. Psych.* **27**, 370–405, 1983.

Terouanne, E. "On a Class of 'Impossible' Figures: A New Language for a New Analysis." *J. Math. Psych.* **22**, 24–47, 1983.

Thro, E. B. "Distinguishing Two Classes of Impossible Objects." *Perception* **12**, 733–751, 1983.

Wilson, R. "Stamp Corner: Impossible Figures." *Math. Intell.* **13**, 80, 1991.

Impredicative

Definitions about a SET which depend on the entire SET.

Improper Integral

An INTEGRAL which has either or both limits INFINITE or which has an INTEGRAND which approaches INFINITY at one or more points in the range of integration.

see also DEFINITE INTEGRAL, INDEFINITE INTEGRAL, INTEGRAL, PROPER INTEGRAL

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Improper Integrals." §4.4 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 135–140, 1992.

Improper Node

A FIXED POINT for which the STABILITY MATRIX has equal nonzero EIGENVECTORS.

see also STABLE IMPROPER NODE, UNSTABLE IMPROPER NODE

Improper Rotation

The SYMMETRY OPERATION corresponding to a a ROTATION followed by an INVERSION OPERATION, also called a ROTOINVERSION. This operation is denoted $\bar{n}$ for an improper rotation by $360°/n$, so the CRYSTALLOGRAPHY RESTRICTION gives only $\bar{1}$, $\bar{2}$, $\bar{3}$, $\bar{4}$, $\bar{6}$ for crystals. The MIRROR PLANE symmetry operation is $(x, y, z) \to (x, y, -z)$, etc., which is equivalent to $\bar{2}$.

Impulse Pair

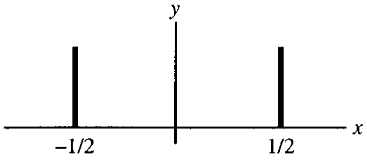

The even impulse pair is the FOURIER TRANSFORM of $\cos(\pi k)$,

$$\mathrm{II}(x) \equiv \tfrac{1}{2}\delta(x + \tfrac{1}{2}) + \tfrac{1}{2}\delta(x - \tfrac{1}{2}). \tag{1}$$

It satisfies

$$\mathrm{II}(x) * f(x) = \tfrac{1}{2}f(x + \tfrac{1}{2}) + \tfrac{1}{2}f(x - \tfrac{1}{2}), \tag{2}$$

where $*$ denotes CONVOLUTION, and

$$\int_{-\infty}^{\infty} \mathrm{II}(x)\,dx = 1. \tag{3}$$

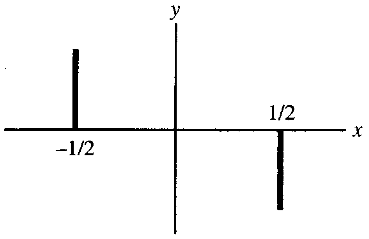

The odd impulse pair is the FOURIER TRANSFORM of $i\sin(\pi s)$,

$$\mathrm{I_I}(x) \equiv \tfrac{1}{2}\delta(x + \tfrac{1}{2}) - \tfrac{1}{2}\delta(x - \tfrac{1}{2}). \tag{4}$$

Impulse Symbol

Bracewell's term for the DELTA FUNCTION.

see also IMPULSE PAIR

References

Bracewell, R. *The Fourier Transform and Its Applications.* New York: McGraw-Hill, 1965.

In-and-Out Curve

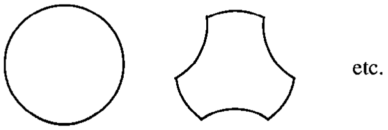

A curve created by starting with a circle, dividing it into six arcs, and flipping three alternating arcs. The process is then repeated an infinite number of times.

Inaccessible Cardinal

An inaccessible cardinal is a CARDINAL NUMBER which cannot be expressed in terms of a smaller number of smaller cardinals.

Inaccessible Cardinals Axiom

see also LEBESGUE MEASURABILITY PROBLEM

Inadmissible

A word or string which is not ADMISSIBLE.

Incenter

The center I of a TRIANGLE'S INCIRCLE. It can be found as the intersection of ANGLE BISECTORS, and it is the interior point for which distances to the sidelines are equal. Its TRILINEAR COORDINATES are 1:1:1. The distance between the incenter and CIRCUMCENTER is $\sqrt{R(R-2r)}$.

The incenter lies on the EULER LINE only for an ISOSCELES TRIANGLE. It does, however, lie on the SODDY LINE. For an EQUILATERAL TRIANGLE, the CIRCUMCENTER O, CENTROID G, NINE-POINT CENTER F, ORTHOCENTER H, and DE LONGCHAMPS POINT Z all coincide with I.

The incenter and EXCENTERS of a TRIANGLE are an ORTHOCENTRIC SYSTEM. The POWER of the incenter with respect to the CIRCUMCIRCLE is

$$p = \frac{a_1 a_2 a_3}{a_1 + a_2 + a_3}$$

(Johnson 1929, p. 190). If the incenters of the TRIANGLES $\Delta A_1H_2H_3$, $\Delta A_2H_3A_1$, and $\Delta A_3H_1H_2$ are X_1, X_2, and X_3, then X_2X_3 is equal and parallel to I_2I_3, where H_i are the FEET of the ALTITUDES and I_i are the incenters of the TRIANGLES. Furthermore, X_1, X_2, X_3, are the reflections of I with respect to the sides of the TRIANGLE $\Delta I_1I_2I_3$ (Johnson 1929, p. 193).

If four points are on a Circle (i.e., they are Concyclic), the incenters of the four Triangles form a Rectangle whose sides are parallel to the lines connecting the middle points of opposite arcs. Furthermore, the connectors pass through the center of the Rectangle (Fuhrmann 1890, p. 50; Johnson 1929, pp. 254–255). More generally, the 16 incenters and excenters of the Triangles whose Vertices are four points on a Circle, are the intersections of two sets of four Parallel lines which are mutually Perpendicular (Johnson 1929, p. 255).

see also Centroid (Orthocentric System), Circumcenter, Excenter, Gergonne Point, Incircle, Inradius, Orthocenter

References

Carr, G. S. *Formulas and Theorems in Pure Mathematics, 2nd ed.* New York: Chelsea, p. 622, 1970.

Dixon, R. *Mathographics.* New York: Dover, p. 58, 1991.

Fuhrmann, W. *Synthetische Beweise Planimetrischer Sätze.* Berlin, 1890.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 182–194, 1929.

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Kimberling, C. "Incenter." `http://www.evansville.edu/~ck6/tcenters/class/incenter.html`.

Incenter-Excenter Circle

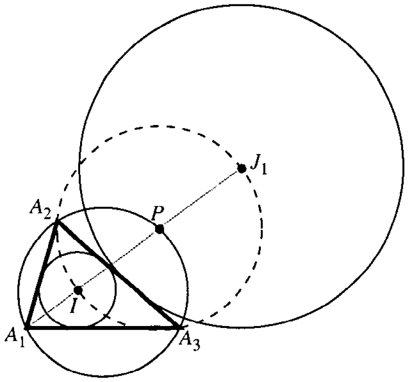

Given a triangle $\Delta A_1A_2A_3$, the points A_1, I, and J_1 lie on a line, where I is the Incenter and J_1 is the Excenter corresponding to A_1. Furthermore, the Circle with IJ_1 as the Diameter has P as its center, where P is the intersection of A_1J_1 with the Circumcircle of $\Delta A_1A_2A_3$, and passes through A_2 and A_3. This Circle has Radius

$$r = \tfrac{1}{2}a_1 \sec(\tfrac{1}{2}\alpha_1) = 2R\sin(\tfrac{1}{2}\alpha_1).$$

It arises because $IJ_1J_2J_3$ forms an Orthocentric System.

see also Circumcircle, Excenter, Excenter-Excenter Circle, Incenter, Orthocentric System

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, p. 185, 1929.

Incidence Axioms

The eight of Hilbert's Axioms which concern collinearity and intersection; they include the first four of Euclid's Postulates.

see also Absolute Geometry, Congruence Axioms, Continuity Axioms, Euclid's Postulates, Hilbert's Axioms, Ordering Axioms, Parallel Postulate

References

Hilbert, D. *The Foundations of Geometry, 2nd ed.* Chicago, IL: Open Court, 1980.

Iyanaga, S. and Kawada, Y. (Eds.). "Hilbert's System of Axioms." §163B in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 544–545, 1980.

Incidence Matrix

For a k-D Polytope Π_k, the incidence matrix is defined by

$$\eta^k_{ij} = \begin{cases} 1 & \text{if } \Pi^i_{k-1} \text{ belongs to } \Pi^j_k \\ 0 & \text{if } \Pi^i_{k-1} \text{ does not belong to } \Pi^j_k. \end{cases}$$

The ith row shows which Π_ks surround Π^i_{k-1}, and the jth column shows which Π_{k-1}s bound Π^j_k. Incidence matrices are also used to specify Projective Planes. The incidence matrices for a Tetrahedron $ABCD$ are

η^0	1	A	B	C
1	1	1	1	1

η^1	AD	BD	CD	BC	AC	AB
A	1	0	0	0	1	1
B	0	1	0	1	0	1
C	0	0	1	1	1	0
D	1	1	1	0	0	0

η^2	BCD	ACD	ABD	ABC
AD	0	1	1	0
BD	1	0	1	0
CD	1	1	0	0
BC	1	0	0	1
AC	0	1	0	1
AB	0	0	1	1

η^3	$ABCD$
BCD	1
ACD	1
ABD	1
ABC	1

see also Adjacency Matrix, k-Chain, k-Circuit

Incident

Two objects which touch each other are said to be incident.

see also Incidence Matrix

Incircle

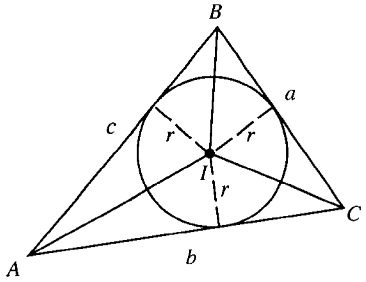

The INSCRIBED CIRCLE of a TRIANGLE ΔABC. The center I is called the INCENTER and the RADIUS r the INRADIUS. The points of intersection of the incircle with T are the VERTICES of the PEDAL TRIANGLE of T with the INCENTER as the PEDAL POINT (c.f. TANGENTIAL TRIANGLE). This TRIANGLE is called the CONTACT TRIANGLE.

The AREA K of the TRIANGLE ΔABC is given by

$$\begin{aligned}K &= \Delta AIC + \Delta CIB + \Delta AIB \\ &= \tfrac{1}{2}br + \tfrac{1}{2}ar + \tfrac{1}{2}cr = \tfrac{1}{2}(a+b+c)r = sr,\end{aligned}$$

where s is the SEMIPERIMETER.

Using the incircle of a TRIANGLE as the INVERSION CENTER, the sides of the TRIANGLE and its CIRCUMCIRCLE are carried into four equal CIRCLES (Honsberger 1976, p. 21). Pedoe (1995, p. xiv) gives a GEOMETRIC CONSTRUCTION for the incircle.

see also CIRCUMCIRCLE, CONGRUENT INCIRCLES POINT, CONTACT TRIANGLE, INRADIUS, TRIANGLE TRANSFORMATION PRINCIPLE

References
Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 11–13, 1967.
Honsberger, R. *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., 1976.
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 182–194, 1929.
Pedoe, D. *Circles: A Mathematical View, rev. ed.* Washington, DC: Math. Assoc. Amer., 1995.

Inclusion-Exclusion Principle

If A_1, ..., A_k are finite sets, then

$$\left|\bigcup_{i=1}^{k} A_i\right| = \sum_{i=1}^{k}(-1)^{i+1}\xi_i,$$

where ξ_i is the sum of the CARDINALITIES of the intersections of the sets taken i at a time.

Inclusion Map

Given a SUBSET B of a SET A, the INJECTION $f : B \to A$ defined by $f(b) = b$ for all $b \in B$ is called the inclusion map.

see also LONG EXACT SEQUENCE OF A PAIR AXIOM

Incommensurate

Two lengths are called incommensurate or incommensurable if their ratio cannot be expressed as a ratio of whole numbers. IRRATIONAL NUMBERS and TRANSCENDENTAL NUMBERS are incommensurate with the integers.

see also IRRATIONAL NUMBER, PYTHAGORAS'S CONSTANT, TRANSCENDENTAL NUMBER

Incomplete Gamma Function

see GAMMA FUNCTION

Incompleteness

A formal theory is said to be incomplete if it contains fewer theorems than would be possible while still retaining CONSISTENCY.

see also CONSISTENCY, GÖDEL'S INCOMPLETENESS THEOREM

References
Chaitin, G. J. "G. J. Chaitin's Home Page." `http://www.cs.auckland.ac.nz/CDMTCS/chaitin`.

Increasing Function

A function $f(x)$ increases on an INTERVAL I if $f(b) > f(a)$ for all $b > a$, where $a, b \in I$. Conversely, a function $f(x)$ decreases on an INTERVAL I if $f(b) < f(a)$ for all $b > a$ with $a, b \in I$.

If the DERIVATIVE $f'(x)$ of a CONTINUOUS FUNCTION $f(x)$ satisfies $f'(x) > 0$ on an OPEN INTERVAL (a, b), then $f(x)$ is increasing on (a, b). However, a function may increase on an interval without having a derivative defined at all points. For example, the function $x^{1/3}$ is increasing everywhere, including the origin $x = 0$, despite the fact that the DERIVATIVE is not defined at that point.

see also DECREASING FUNCTION, DERIVATIVE, NONDECREASING FUNCTION, NONINCREASING FUNCTION

Increasing Sequence

For a SEQUENCE $\{a_n\}$, if $a_{n+1} - a_n > 0$ for $n \geq x$, then a is increasing for $n \geq x$. Conversely, if $a_{n+1} - a_n < 0$ for $n \geq x$, then a is DECREASING for $n \geq x$. If $a_{n+1}/a_n > 1$ for all $n \geq x$, then a is increasing for $n \geq x$. Conversely, if $a_{n+1}/a_n < 1$ for all $n \geq x$, then a is decreasing for $n \geq x$.

Indefinite Integral

An INTEGRAL

$$\int f(x)\,dx$$

without upper and lower limits, also called an ANTIDERIVATIVE. The first FUNDAMENTAL THEOREM OF CALCULUS allows DEFINITE INTEGRALS to be computed in terms of indefinite integrals. If F is the indefinite integral for $f(x)$, then

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

see also ANTIDERIVATIVE, CALCULUS, DEFINITE INTEGRAL, FUNDAMENTAL THEOREMS OF CALCULUS, INTEGRAL

Indefinite Quadratic Form

A QUADRATIC FORM $Q(\mathbf{x})$ is indefinite if it is less than 0 for some values and greater than 0 for others. The QUADRATIC FORM, written in the form $(\mathbf{x}, \mathsf{A}\mathbf{x})$, is indefinite if EIGENVALUES of the MATRIX A are of both signs.

see also POSITIVE DEFINITE QUADRATIC FORM, POSITIVE SEMIDEFINITE QUADRATIC FORM

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1106, 1979.

Indegree

The number of inward directed EDGES from a given VERTEX in a DIRECTED GRAPH.

see also LOCAL DEGREE, OUTDEGREE

Independence Axiom

A rational choice between two alternatives should depend only on how they differ.

Independence Complement Theorem

If sets E and F are INDEPENDENT, then so are E and F', where F' is the complement of F (i.e., the set of all possible outcomes not contained in F). Let $\cup$ denote "or" and $\cap$ denote "and." Then

$$\begin{aligned} P(E) &= P(EF \cup EF') && (1)\\ &= P(EF) + P(EF') - P(EF \cap EF'), && (2) \end{aligned}$$

where AB is an abbreviation for $A \cap B$. But E and F are independent, so

$$P(EF) = P(E)P(F). \qquad (3)$$

Also, since F and F' are complements, they contain no common elements, which means that

$$P(EF \cap EF') = 0 \qquad (4)$$

for any E. Plugging (4) and (3) into (2) then gives

$$P(E) = P(E)P(F) + P(EF'). \qquad (5)$$

Rearranging,

$$P(EF') = P(E)[1 - P(F)] = P(E)P(F'), \qquad (6)$$

Q.E.D.

see also INDEPENDENT STATISTICS

Independence Number

The number

$$\alpha(G) = \max(|U| : U \subset V \text{ independent})$$

for a GRAPH G. The independence number of the DE BRUIJN GRAPH of order n is given by 1, 2, 3, 7, 13, 28, ... (Sloane's A006946).

References

Sloane, N. J. A. Sequence A006946/M0834 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Independent Equations

see LINEARLY INDEPENDENT

Independent Sequence

see STRONGLY INDEPENDENT, WEAKLY INDEPENDENT

Independent Statistics

Two variates A and B are statistically independent IFF the CONDITIONAL PROBABILITY $P(A|B)$ of A given B satisfies

$$P(A|B) = P(A), \qquad (1)$$

in which case the probability of A and B is just

$$P(AB) = P(A \cap B) = P(A)P(B). \qquad (2)$$

Similarly, n events $A_1, A_2, \ldots, A_n$ are independent IFF

$$P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i). \qquad (3)$$

Statistically independent variables are always UNCORRELATED, but the converse is not necessarily true.

see also BAYES' FORMULA, CONDITIONAL PROBABILITY, INDEPENDENCE COMPLEMENT THEOREM, UNCORRELATED

Independent Vertices

A set of VERTICES A of a GRAPH with EDGES V is independent if it contains no EDGES.

see also INDEPENDENCE NUMBER

Indeterminate Problems

see DIOPHANTINE EQUATION—LINEAR

Index

A statistic which assigns a single number to several individual statistics in order to quantify trends. The best-known index in the United States is the consumer price index, which gives a sort of "average" value for inflation based on the price changes for a group of selected products.

Let p_n be the price per unit in period n, q_n be the quantity produced in period n, and $v_n \equiv p_n q_n$ be the value of the n units. Let q_a be the estimated relative importance of a product. There are several types of indices defined, among them those listed in the following table.

Index	Abbr.	Formula
Bowley index	P_B	$\frac{1}{2}(P_L + P_P)$
Fisher index	P_F	$\sqrt{P_L P_P}$
Geometric mean index	P_G	$\left[\prod \left(\frac{p_n}{p_0}\right)^{v_0}\right]^{1/\sum v_0}$
Harmonic mean index	P_H	$\frac{\sum p_0 q_0}{\sum \frac{{p_0}^2 q_0}{p_n}}$
Laspeyres's index	P_L	$\frac{\sum p_n q_0}{\sum p_0 q_0}$
Marshall-Edgeworth index	P_{ME}	$\frac{\sum p_n (q_0 + q_n)}{\sum (v_0 + v_n)}$
Mitchell index	P_M	$\frac{\sum p_n q_a}{\sum p_0 q_a}$
Paasche's index	P_P	$\frac{\sum p_n q_n}{\sum p_0 q_n}$
Walsh index	P_W	$\frac{\sum \sqrt{q_0 q_n}\, p_n}{\sum \sqrt{q_0 q_n}\, p_0}$

see also Bowley Index, Fisher Index, Geometric Mean Index, Harmonic Mean Index, Laspeyres' Index, Marshall-Edgeworth Index, Mitchell Index, Paasche's Index, Residue Index, Walsh Index

References

Fisher, I. *The Making of Index Numbers: A Study of Their Varieties, Tests and Reliability, 3rd ed.* New York: Augustus M. Kelly, 1967.

Kenney, J. F. and Keeping, E. S. "Index Numbers." Ch. 5 in *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, pp. 64–74, 1962.

Mudgett, B. D. *Index Numbers.* New York: Wiley, 1951.

Index Set

A Stochastic Process is a family of Random Variables $\{x(t, \bullet), t \in \mathcal{J}\}$ from some Probability Space $(S, \mathbb{S}, P)$ into a State Space $(S', \mathbb{S}')$, where $\mathcal{J}$ is the index set of the process.

References

Doob, J. L. "The Development of Rigor in Mathematical Probability (1900–1950)." *Amer. Math. Monthly* **103**, 586–595, 1996.

Index Theory

A branch of Topology dealing with topological invariants of Manifolds.

References

Roe, J. *Index Theory, Coarse Geometry, and Topology of Manifolds.* Providence, RI: Amer. Math. Soc., 1996.

Upmeier, H. *Toeplitz Operators and Index Theory in Several Complex Variables.* Boston, MA: Birkhäuser, 1996.

Indicatrix

A spherical image of a curve. The most common indicatrix is Dupin's Indicatrix.

see also Dupin's Indicatrix

Indicial Equation

The Recurrence Relation obtained during application of the Frobenius Method of solving a second-order ordinary differential equation. The indicial equation (also called the Characteristic Equation) is obtained by noting that, by definition, the lowest order term x^k (that corresponding to $n = 0$) must have a Coefficient of zero. For an example of the construction of an indicial equation, see Bessel Differential Equation.

1. If the two Roots are equal, only one solution can be obtained.
2. If the two Roots differ by a noninteger, two solutions can be obtained.
3. If the two Roots differ by an Integer, the larger will yield a solution. The smaller may or may not.

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 532–534, 1953.

Indifference Principle

see Insufficient Reason Principle

Induced Map

If $f : (X, A) \to (Y, B)$ is homotopic to $g : (X, A) \to (Y, B)$, then $f_* : H_n(X, A) \to H_n(Y, B)$ and $g_* : H_n(X, A) \to H_n(Y, B)$ are said to be the induced maps.

see also Eilenberg-Steenrod Axioms

Induced Norm

see Natural Norm

Induction

The use of the Induction Principle in a Proof. Induction used in mathematics is often called Mathematical Induction.

References

Buck, R. C. "Mathematical Induction and Recursive Definitions." *Amer. Math. Monthly* **70**, 128–135, 1963.

Induction Axiom

The fifth of Peano's Axioms, which states: If a Set S of numbers contains zero and also the successor of every number in S, then every number is in S.

see also Peano's Axioms

Induction Principle

The truth of an INFINITE sequence of propositions P_i for $i = 1, \ldots, \infty$ is established if (1) P_1 is true, and (2) P_k IMPLIES P_{k+1} for all k.

References

Courant, R. and Robbins, H. "The Principle of Mathematical Induction" and "Further Remarks on Mathematical Induction." §1.2.1 and 1.7 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 9–11 and 18–20, 1996.

Inequality

A mathematical statement that one quantity is greater than or less than another. "a is less than b" is denoted $a < b$, and "a is greater than b" is denoted $a > b$. "a is less than or equal to b" is denoted $a \leq b$, and "a is greater than or equal to b" is denoted $a \geq b$. The symbols $a \ll b$ and $a \gg b$ are used to denote "a is much less than b" and "a is much greater than b," respectively.

Solutions to the inequality $|x - a| < b$ consist of the set $\{x : -b < x - a < b\}$, or equivalently $\{x : a - b < x < a + b\}$. Solutions to the inequality $|x - a| > b$ consist of the set $\{x : x - a > b\} \cup \{x : x - a < -b\}$. If a and b are both POSITIVE or both NEGATIVE and $a < b$, then $1/a > 1/b$.

see also ABC CONJECTURE, ARITHMETIC-LOGARITHMIC-GEOMETRIC MEAN INEQUALITY, BERNOULLI INEQUALITY, BERNSTEIN'S INEQUALITY, BERRY-OSSEEN INEQUALITY, BIENAYMÉ-CHEBYSHEV INEQUALITY, BISHOP'S INEQUALITY, BOGOMOLOV-MIYAOKA-YAU INEQUALITY, BOMBIERI'S INEQUALITY, BONFERRONI'S INEQUALITY, BOOLE'S INEQUALITY, CARLEMAN'S INEQUALITY, CAUCHY INEQUALITY, CHEBYSHEV INEQUALITY, CHI INEQUALITY, COPSON'S INEQUALITY, ERDŐS-MORDELL THEOREM, EXPONENTIAL INEQUALITY, FISHER'S BLOCK DESIGN INEQUALITY, FISHER'S ESTIMATOR INEQUALITY, GÅRDING'S INEQUALITY, GAUSS'S INEQUALITY, GRAM'S INEQUALITY, HADAMARD'S INEQUALITY, HARDY'S INEQUALITY, HARNACK'S INEQUALITY, HÖLDER INTEGRAL INEQUALITY, HÖLDER'S SUM INEQUALITY, ISOPERIMETRIC INEQUALITY, JARNICK'S INEQUALITY, JENSEN'S INEQUALITY, JORDAN'S INEQUALITY, KANTROVICH INEQUALITY, MARKOV'S INEQUALITY, MINKOWSKI INTEGRAL INEQUALITY, MINKOWSKI SUM INEQUALITY, MORSE INEQUALITIES, NAPIER'S INEQUALITY, NOSARZEWSKA'S INEQUALITY, OSTROWSKI'S INEQUALITY, PTOLEMY INEQUALITY, ROBBIN'S INEQUALITY, SCHRÖDER-BERNSTEIN THEOREM, SCHUR'S INEQUALITIES, SCHWARZ'S INEQUALITY, SQUARE ROOT INEQUALITY, STEFFENSEN'S INEQUALITY, STOLARSKY'S INEQUALITY, STRONG SUBADDITIVITY INEQUALITY, TRIANGLE INEQUALITY, TURÁN'S INEQUALITIES, WEIERSTRAß PRODUCT INEQUALITY, WIRTINGER'S INEQUALITY, YOUNG INEQUALITY

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 16, 1972.
Beckenbach, E. F. and Bellman, Richard E. *An Introduction to Inequalities.* New York: Random House, 1961.
Beckenbach, E. F. and Bellman, Richard E. *Inequalities, 2nd rev. print.* Berlin: Springer-Verlag, 1965.
Hardy, G. H.; Littlewood, J. E.; and Pólya, G. *Inequalities, 2nd ed.* Cambridge, England: Cambridge University Press, 1952.
Kazarinoff, N. D. *Geometric Inequalities.* New York: Random House, 1961.
Mitrinovic, D. S. *Analytic Inequalities.* New York: Springer-Verlag, 1970.
Mitrinovic, D. S.; Pecaric, J. E.; and Fink, A. M. *Classical & New Inequalities in Analysis.* Dordrecht, Netherlands: Kluwer, 1993.
Mitrinovic, D. S.; Pecaric, J. E.; Fink, A. M. *Inequalities Involving Functions & Their Integrals & Derivatives.* Dordrecht, Netherlands: Kluwer, 1991.
Mitrinovic, D. S.; Pecaric, J. E.; and Volenec, V. *Recent Advances in Geometric Inequalities.* Dordrecht, Netherlands: Kluwer, 1989.

Inexact Differential

An infinitesimal which is not the differential of an actual function and which cannot be expressed as

$$dz = \left(\frac{\partial z}{\partial x}\right)_y dx + \left(\frac{\partial z}{\partial y}\right)_z dy,$$

the way an EXACT DIFFERENTIAL can. Inexact differentials are denoted with a bar through the d. The most common example of an inexact differential is the change in heat $đQ$ encountered in thermodynamics.

see also EXACT DIFFERENTIAL, PFAFFIAN FORM

References

Zemansky, M. W. *Heat and Thermodynamics, 5th ed.* New York: McGraw-Hill, p. 38, 1968.

Inf

see INFIMUM, INFIMUM LIMIT

Infimum

The greatest lower bound of a set. It is denoted

$$\inf_S .$$

see also INFIMUM LIMIT, SUPREMUM

Infimum Limit

The limit infimum of a set is the greatest lower bound of the CLOSURE of a set. It is denoted

$$\liminf_S .$$

see also INFIMUM, SUPREMUM

Infinary Divisor

p^x is an infinary divisor of p^y (with $y > 0$) if $p^x|_{y-1}p^y$. This generalizes the concept of the k-ARY DIVISOR.

see also INFINARY PERFECT NUMBER, k-ARY DIVISOR

References

Cohen, G. L. "On an Integer's Infinary Divisors." *Math. Comput.* **54**, 395–411, 1990.

Cohen, G. and Hagis, P. "Arithmetic Functions Associated with the Infinary Divisors of an Integer." *Internat. J. Math. Math. Sci.* **16**, 373–383, 1993.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 54, 1994.

Infinary Multiperfect Number

Let $\sigma_\infty(n)$ be the SUM of the INFINARY DIVISORS of a number n. An infinary k-multiperfect number is a number n such that $\sigma_\infty(n) = kn$. Cohen (1990) found 13 infinary 3-multiperfects, seven 4-multiperfects, and two 5-multiperfects.

see also INFINARY PERFECT NUMBER

References

Cohen, G. L. "On an Integer's Infinary Divisors." *Math. Comput.* **54**, 395–411, 1990.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 54, 1994.

Infinary Perfect Number

Let $\sigma_\infty(n)$ be the SUM of the INFINARY DIVISORS of a number n. An infinary perfect number is a number n such that $\sigma_\infty(n) = 2n$. Cohen (1990) found 14 such numbers. The first few are 6, 60, 90, 36720, ... (Sloane's A007257).

see also INFINARY MULTIPERFECT NUMBER

References

Cohen, G. L. "On an Integer's Infinary Divisors." *Math. Comput.* **54**, 395–411, 1990.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 54, 1994.

Sloane, N. J. A. Sequence A007257/M4267 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Infinite

Greater than any assignable quantity of the sort in question. In mathematics, the concept of the infinite is made more precise through the notion of an INFINITE SET.

see also COUNTABLE SET, COUNTABLY INFINITE SET, FINITE, INFINITE SET, INFINITESIMAL, INFINITY

Infinite Product

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

A PRODUCT involving an INFINITE number of terms. Such products can converge. In fact, for POSITIVE a_n, the PRODUCT $\prod_{n=1}^\infty a_n$ converges to a NONZERO number IFF $\sum_{n=1}^\infty \ln a_n$ converges.

Infinite products can be used to define the COSINE

$$\cos x = \prod_{n=1}^\infty \left[1 - \frac{4x^2}{\pi^2(2n-1)^2}\right], \tag{1}$$

GAMMA FUNCTION

$$\Gamma(z) = \left[ze^{\gamma z}\prod_{r=1}^\infty \left(1+\frac{z}{r}\right)e^{-z/r}\right]^{-1}, \tag{2}$$

SINE, and SINC FUNCTION. They also appear in the POLYGON CIRCUMSCRIBING CONSTANT

$$K = \prod_{n=3}^\infty \frac{1}{\cos\left(\frac{\pi}{n}\right)}. \tag{3}$$

An interesting infinite product formula due to Euler which relates π and the nth PRIME p_n is

$$\pi = \frac{2}{\prod_{i=n}^\infty \left[1 + \frac{\sin(\frac{1}{2}\pi p_n)}{p_n}\right]} \tag{4}$$

$$= \frac{2}{\prod_{i=n}^\infty \left[1 + \frac{(-1)^{(p_n-1)/2}}{p_n}\right]} \tag{5}$$

(Blatner 1997).

The product

$$\prod_{n=1}^\infty \left(1+\frac{1}{n^p}\right) \tag{6}$$

has closed form expressions for small POSITIVE integral $p \geq 2$,

$$\prod_{n=1}^\infty \left(1+\frac{1}{n^2}\right) = \frac{\sinh\pi}{\pi} \tag{7}$$

$$\prod_{n=1}^\infty \left(1+\frac{1}{n^3}\right) = \frac{1}{\pi}\cosh(\tfrac{1}{2}\pi\sqrt{3}) \tag{8}$$

$$\prod_{n=1}^\infty \left(1+\frac{1}{n^4}\right) = \frac{\cosh(\pi\sqrt{2}) - \cos(\pi\sqrt{2})}{2\pi^2} \tag{9}$$

$$\prod_{n=1}^\infty \left(1+\frac{1}{n^5}\right) = \left|\Gamma[\exp(\tfrac{2}{5}\pi i)]\Gamma[\exp(\tfrac{6}{5}\pi i)]\right|^{-2}. \tag{10}$$

The d-ANALOG expression

$$[\infty!]_d = \prod_{n=3}^\infty \left(1 - \frac{2^d}{n^d}\right) \tag{11}$$

also has closed form expressions,

$$\prod_{n=3}^{\infty}\left(1-\frac{4}{n^2}\right)=\frac{1}{6} \tag{12}$$

$$\prod_{n=3}^{\infty}\left(1-\frac{8}{n^3}\right)=\frac{\sinh(\pi\sqrt{3}\,)}{42\pi\sqrt{3}} \tag{13}$$

$$\prod_{n=3}^{\infty}\left(1-\frac{16}{n^4}\right)=\frac{\sinh(2\pi)}{120\pi} \tag{14}$$

$$\prod_{n=3}^{\infty}\left(1-\frac{32}{n^5}\right)=\left|\Gamma[\exp(\tfrac{1}{5}\pi i)]\Gamma[2\exp(\tfrac{7}{5}\pi i)]\right|^{-2}. \tag{15}$$

see also COSINE, DIRICHLET ETA FUNCTION, EULER IDENTITY, GAMMA FUNCTION, ITERATED EXPONENTIAL CONSTANTS, POLYGON CIRCUMSCRIBING CONSTANT, POLYGON INSCRIBING CONSTANT, *Q*-FUNCTION, SINE

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 75, 1972.
Arfken, G. "Infinite Products." §5.11 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 346–351, 1985.
Blatner, D. *The Joy of Pi.* New York: Walker, p. 119, 1997.
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/infprd/infprd.html`.
Hansen, E. R. *A Table of Series and Products.* Englewood Cliffs, NJ: Prentice-Hall, 1975.
Whittaker, E. T. and Watson, G. N. §7.5 and 7.6 in *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Infinite Series

A SERIES with an INFINITE number of terms.

see also SERIES

Infinite Set

A SET of S elements is said to be infinite if the elements of a PROPER SUBSET S' can be put into ONE-TO-ONE correspondence with the elements of S. An infinite set whose elements can be put into a ONE-TO-ONE correspondence with the set of INTEGERS is said to be COUNTABLY INFINITE; otherwise, it is called UNCOUNTABLY INFINITE.

see also ALEPH-0, ALEPH-1, CARDINAL NUMBER, COUNTABLY INFINITE SET, CONTINUUM, FINITE, INFINITE, INFINITY, ORDINAL NUMBER, TRANSFINITE NUMBER, UNCOUNTABLY INFINITE SET

References
Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 77, 1996.

Infinitesimal

A quantity which yields 0 after the application of some LIMITING process. The understanding of infinitesimals was a major roadblock to the acceptance of CALCULUS and its placement on a firm mathematical foundation.

see also INFINITE, INFINITY, NONSTANDARD ANALYSIS

Infinitesimal Analysis

An archaic term for CALCULUS.

Infinitesimal Matrix Change

Let B, A, and e be square matrices with e small, and define

$$\mathsf{B}\equiv\mathsf{A}(\mathsf{I}+\mathsf{e}), \tag{1}$$

where I is the IDENTITY MATRIX. Then the inverse of B is approximately

$$\mathsf{B}^{-1}=(\mathsf{I}-\mathsf{e})\mathsf{A}^{-1}. \tag{2}$$

This can be seen by multiplying

$$\begin{aligned}\mathsf{B}\mathsf{B}^{-1}&=(\mathsf{A}+\mathsf{A}\mathsf{e})(\mathsf{A}^{-1}-\mathsf{e}\mathsf{A}^{-1})\\&=\mathsf{A}\mathsf{A}^{-1}-\mathsf{A}\mathsf{e}\mathsf{A}^{-1}+\mathsf{A}\mathsf{e}\mathsf{A}^{-1}-\mathsf{A}\mathsf{e}^2\mathsf{A}^{-1}\\&=\mathsf{I}-\mathsf{A}\mathsf{e}^2\mathsf{A}^{-1}\approx 1.\end{aligned} \tag{3}$$

Note that if we instead let $\mathsf{B}'\equiv\mathsf{A}+\mathsf{e}$, and look for an inverse of the form $\mathsf{B}'^{-1}=\mathsf{A}^{-1}+\mathsf{C}$, we obtain

$$\begin{aligned}\mathsf{B}'\mathsf{B}'^{-1}&=(\mathsf{A}+\mathsf{e})(\mathsf{A}^{-1}+\mathsf{C})=\mathsf{A}\mathsf{A}^{-1}+\mathsf{A}\mathsf{C}+\mathsf{e}\mathsf{A}^{-1}+\mathsf{e}\mathsf{C}\\&=\mathsf{I}+\mathsf{A}\mathsf{C}+\mathsf{e}(\mathsf{C}+\mathsf{A}^{-1})\equiv\mathsf{I}.\end{aligned} \tag{4}$$

In order to eliminate the e term, we require $\mathsf{C}=-\mathsf{A}^{-1}$. However, then $\mathsf{A}\mathsf{C}=-\mathsf{I}$, so $\mathsf{B}\mathsf{B}^{-1}=0$ so there can be no inverse of this form.

The exact inverse of B can be found as follows.

$$\mathsf{B}=\mathsf{A}(\mathsf{I}+\mathsf{e})=\mathsf{A}(\mathsf{I}+\mathsf{A}^{-1}\mathsf{e}), \tag{5}$$

so

$$\mathsf{B}^{-1}=[\mathsf{A}(\mathsf{I}+\mathsf{A}^{-1}\mathsf{e})]^{-1}. \tag{6}$$

Using a general MATRIX INVERSE identity then gives

$$\mathsf{B}^{-1}=(\mathsf{I}+\mathsf{A}^{-1}\mathsf{e})^{-1}\mathsf{A}^{-1}. \tag{7}$$

Infinitesimal Rotation

An infinitesimal transformation of a VECTOR $\mathbf{r}$ is given by

$$\mathbf{r}'=(\mathsf{I}+\mathsf{e})\mathbf{r}, \tag{1}$$

where the MATRIX e is infinitesimal and I is the IDENTITY MATRIX. (Note that the infinitesimal transformation may not correspond to an inversion, since inversion is a discontinuous process.) The COMMUTATIVITY of infinitesimal transformations e_1 and e_2 is established by the equivalence of

$$(\mathsf{I}+\mathsf{e}_1)(\mathsf{I}+\mathsf{e}_2) = \mathsf{I}^2+\mathsf{e}_1\mathsf{I}+\mathsf{I}\mathsf{e}_2+\mathsf{e}_1\mathsf{e}_2 \approx \mathsf{I}+\mathsf{e}_1+\mathsf{e}_2 \quad (2)$$

$$(\mathsf{I}+\mathsf{e}_2)(\mathsf{I}+\mathsf{e}_1) = \mathsf{I}^2+\mathsf{e}_2\mathsf{I}+\mathsf{I}\mathsf{e}_1+\mathsf{e}_2\mathsf{e}_1 \approx \mathsf{I}+\mathsf{e}_2+\mathsf{e}_1. \quad (3)$$

Now let

$$\mathsf{A} \equiv \mathsf{I}+\mathsf{e}. \quad (4)$$

The inverse A^{-1} is then $\mathsf{I}-\mathsf{e}$, since

$$\mathsf{A}\mathsf{A}^{-1} = (\mathsf{I}+\mathsf{e})(\mathsf{I}-\mathsf{e}) = \mathsf{I}^2-\mathsf{e}^2 \approx \mathsf{I}. \quad (5)$$

Since we are defining our infinitesimal transformation to be a rotation, ORTHOGONALITY of ROTATION MATRICES requires that

$$\mathsf{A}^{\mathrm{T}} = \mathsf{A}^{-1}, \quad (6)$$

but

$$\mathsf{A}^{-1} = \mathsf{I}-\mathsf{e} \quad (7)$$

$$(\mathsf{I}+\mathsf{e})^{\mathrm{T}} = \mathsf{I}^{\mathrm{T}}+\mathsf{e}^{\mathrm{T}} = \mathsf{I}+\mathsf{e}^{\mathrm{T}}, \quad (8)$$

so $\mathsf{e} = -\mathsf{e}^{\mathrm{T}}$ and the infinitesimal rotation is ANTISYMMETRIC. It must therefore have a MATRIX of the form

$$\mathsf{e} = \begin{bmatrix} 0 & d\Omega_3 & -d\Omega_2 \\ -d\Omega_3 & 0 & d\Omega_1 \\ d\Omega_2 & -d\Omega_1 & 0 \end{bmatrix}. \quad (9)$$

The differential change in a vector $\mathbf{r}$ upon application of the ROTATION MATRIX is then

$$d\mathbf{r} \equiv \mathbf{r}' - \mathbf{r} = (\mathsf{I}+\mathsf{e})\mathbf{r} - \mathbf{r} = \mathsf{e}\mathbf{r}. \quad (10)$$

Writing in MATRIX form,

$$\begin{aligned} d\mathbf{r} &= \begin{bmatrix} x \\ y \\ z \end{bmatrix} \begin{bmatrix} 0 & d\Omega_3 & -d\Omega_2 \\ -d\Omega_3 & 0 & d\Omega_1 \\ d\Omega_2 & -d\Omega_1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} y\,d\Omega_3 - z\,d\Omega_2 \\ z\,d\Omega_1 - x\,d\Omega_3 \\ x\,d\Omega_2 - y\,d\Omega_1 \end{bmatrix} \quad (11) \\ &= (y\,d\Omega_3 - z\,d\Omega_2)\hat{\mathbf{x}} + (z\,d\Omega_1 - x\,d\Omega_3)\hat{\mathbf{y}} \\ &\quad + (x\,d\Omega_2 - y\,d\Omega_1)\hat{\mathbf{z}} \\ &= \mathbf{r} \times d\mathbf{\Omega}. \quad (12) \end{aligned}$$

Therefore,

$$\left(\frac{d\mathbf{r}}{dt}\right)_{\text{rotation, body}} = \mathbf{r} \times \frac{d\mathbf{\Omega}}{dt} = \mathbf{r} \times \boldsymbol{\omega}, \quad (13)$$

where

$$\boldsymbol{\omega} \equiv \frac{d\mathbf{\Omega}}{dt} = \hat{\mathbf{n}}\frac{d\phi}{dt}. \quad (14)$$

The total rotation observed in the stationary frame will be a sum of the rotational velocity and the velocity in the rotating frame. However, note that an observer in the stationary frame will see a velocity opposite in direction to that of the observer in the frame of the rotating body, so

$$\left(\frac{d\mathbf{r}}{dt}\right)_{\text{space}} = \left(\frac{d\mathbf{r}}{dt}\right)_{\text{body}} + \boldsymbol{\omega} \times \mathbf{r}. \quad (15)$$

This can be written as an operator equation, known as the ROTATION OPERATOR, defined as

$$\left(\frac{d}{dt}\right)_{\text{space}} = \left(\frac{d}{dt}\right)_{\text{body}} + \boldsymbol{\omega} \times . \quad (16)$$

see also ACCELERATION, EULER ANGLES, ROTATION, ROTATION MATRIX, ROTATION OPERATOR

Infinitive Sequence

A sequence $\{x_n\}$ is called an infinitive sequence if, for every i, $x_n = i$ for infinitely many n. Write $a(i,j)$ for the jth index n for which $x_n = i$. Then as i and j range through N, the array $A = a(i,j)$, called the associative array of x, ranges through all of N.

see also FRACTAL SEQUENCE

References

Kimberling, C. "Fractal Sequences and Interspersions." *Ars Combin.* **45**, 157–168, 1997.

Infinitude of Primes

see EUCLID'S THEOREMS

Infinity

An unbounded number greater than every REAL NUMBER, most often denoted as ∞. The symbol ∞ had been used as an alternative to M (1,000) in ROMAN NUMERALS until 1655, when John Wallis suggested it be used instead for infinity.

Infinity is a very tricky concept to work with, as evidenced by some of the counterintuitive results which follow from Georg Cantor's treatment of INFINITE SETS. Informally, $1/\infty = 0$, a statement which can be made rigorous using the LIMIT concept,

$$\lim_{x\to\infty} \frac{1}{x} = 0.$$

Similarly,

$$\lim_{x\to 0^+} \frac{1}{x} = \infty,$$

where the notation 0^+ indicates that the LIMIT is taken from the POSITIVE side of the REAL LINE.

see also ALEPH, ALEPH-0, ALEPH-1, CARDINAL NUMBER, CONTINUUM, CONTINUUM HYPOTHESIS, HILBERT HOTEL, INFINITE, INFINITE SET, INFINITESIMAL, LINE AT INFINITY, L'HOSPITAL'S RULE, POINT AT INFINITY, TRANSFINITE NUMBER, UNCOUNTABLY INFINITE SET, ZERO

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 19, 1996.

Courant, R. and Robbins, H. "The Mathematical Analysis of Infinity." §2.4 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 77–88, 1996.

Hardy, G. H. *Orders of Infinity, the 'infinitarcalcul' of Paul Du Bois-Reymond, 2nd ed.* Cambridge, England: Cambridge University Press, 1924.

Lavine, S. *Understanding the Infinite.* Cambridge, MA: Harvard University Press, 1994.

Maor, E. *To Infinity and Beyond: A Cultural History of the Infinite.* Boston, MA: Birkhäuser, 1987.

Moore, A. W. *The Infinite.* New York: Routledge, 1991.

Morris, R. *Achilles in the Quantum Universe: The Definitive History of Infinity.* New York: Henry Holt, 1997.

Péter, R. *Playing with Infinity.* New York: Dover, 1976.

Smail, L. L. *Elements of the Theory of Infinite Processes.* New York: McGraw-Hill, 1923.

Vilenskin, N. Ya. *In Search of Infinity.* Boston, MA: Birkhäuser, 1995.

Wilson, A. M. *The Infinite in the Finite.* New York: Oxford University Press, 1996.

Zippin, L. *Uses of Infinity.* New York: Random House, 1962.

Inflection Point

A point on a curve at which the SIGN of the CURVATURE (i.e., the concavity) changes. The FIRST DERIVATIVE TEST can sometimes distinguish inflection points from EXTREMA for DIFFERENTIABLE functions $f(x)$.

see also CURVATURE, DIFFERENTIABLE, EXTREMUM, FIRST DERIVATIVE TEST, STATIONARY POINT

Information Dimension

Define the "information function" to be

$$I \equiv -\sum_{i=1}^{N} P_i(\epsilon)\ln[P_i(\epsilon)], \tag{1}$$

where $P_i(\epsilon)$ is the NATURAL MEASURE, or probability that element i is populated, normalized such that

$$\sum_{i=1}^{N} P_i(\epsilon) = 1. \tag{2}$$

The information dimension is then defined by

$$\begin{aligned} d_{\mathrm{inf}} &\equiv -\lim_{\epsilon\to 0^+} \frac{I}{\ln(\epsilon)} \\ &= \lim_{\epsilon\to 0^+} \sum_{i=1}^{N} \frac{P_i(\epsilon)\ln[P_i(\epsilon)]}{\ln(\epsilon)}. \end{aligned} \tag{3}$$

If every element is equally likely to be visited, then $P_i(\epsilon)$ is independent of i, and

$$\sum_{i=1}^{N} P_i(\epsilon) = NP_i(\epsilon) = 1, \tag{4}$$

so

$$P_i(\epsilon) = \frac{1}{N}, \tag{5}$$

and

$$\begin{aligned} d_{\mathrm{inf}} &= \lim_{\epsilon\to 0^+} \frac{\sum\limits_{i=1}^{N} \frac{1}{N}\ln\left(\frac{1}{N}\right)}{\ln\epsilon} \\ &= \lim_{\epsilon\to 0^+} \frac{\ln(N^{-1})}{\ln\epsilon} = -\lim_{\epsilon\to 0^+} \frac{\ln N}{\ln(\epsilon)} = d_{\mathrm{cap}}, \end{aligned} \tag{6}$$

where d_{cap} is the CAPACITY DIMENSION.

see also CORRELATION EXPONENT

References

Farmer, J. D. "Chaotic Attractors of an Infinite-dimensional Dynamical System." *Physica D* 4, 366–393, 1982.

Nayfeh, A. H. and Balachandran, B. *Applied Nonlinear Dynamics: Analytical, Computational, and Experimental Methods.* New York: Wiley, pp. 545–547, 1995.

Information Entropy

see ENTROPY

Information Theory

The branch of mathematics dealing with the efficient and accurate storage, transmission, and representation of information.

see also CODING THEORY, ENTROPY

References

Goldman, S. *Information Theory.* New York: Dover, 1953.

Lee, Y. W. *Statistical Theory of Communication.* New York: Wiley, 1960.

Pierce, J. R. *An Introduction to Information Theory.* New York: Dover, 1980.

Reza, F. M. *An Introduction to Information Theory.* New York: Dover, 1994.

Singh, J. *Great Ideas in Information Theory, Language and Cybernetics.* New York: Dover, 1966.

Zayed, A. I. *Advances in Shannon's Sampling Theory.* Boca Raton, FL: CRC Press, 1993.

Initial Value Problem

An initial value problem is a problem that has its conditions specified at some time $t = t_0$. Usually, the problem is an ORDINARY DIFFERENTIAL EQUATION or a PARTIAL DIFFERENTIAL EQUATION. For example,

$$\begin{cases} \frac{\partial^2 u}{\partial t^2} - \nabla^2 u = f & \text{in } \Omega \\ u = u_0 & t = t_0 \\ u = u_1 & \text{on } \partial\Omega, \end{cases}$$

where $\partial\Omega$ denotes the boundary of Ω, is an initial value problem.

see also BOUNDARY CONDITIONS, BOUNDARY VALUE PROBLEM, PARTIAL DIFFERENTIAL EQUATION

References

Eriksson, K.; Estep, D.; Hansbo, P.; and Johnson, C. *Computational Differential Equations.* Lund, Sweden: Studentlitteratur, 1996.

Injection

see ONE-TO-ONE

Injective

A MAP is injective when it is ONE-TO-ONE, i.e., f is injective when $x \neq y$ IMPLIES $f(x) \neq f(y)$.

see also ONE-TO-ONE, SURJECTIVE

Injective Patch

An injective patch is a PATCH such that $\mathbf{x}(u_1, v_1) = \mathbf{x}(u_2, v_2)$ implies that $u_1 = u_2$ and $v_1 = v_2$. An example of a PATCH which is injective but not REGULAR is the function defined by (u^3, v^3, uv) for $u, v \in (-1, 1)$. However, if $\mathbf{x} : U \to \mathbb{R}^n$ is an injective regular patch, then $\mathbf{x}$ maps U diffeomorphically onto $\mathbf{x}(U)$.

see also PATCH, REGULAR PATCH

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 187, 1993.

Inner Automorphism Group

A particular type of AUTOMORPHISM GROUP which exists only for GROUPS. For a GROUP G, the inner automorphism group is defined by

$$\mathrm{Inn}(G) = \{\sigma_a : a \in G\} \subset \mathrm{Aut}(G)$$

where σ_a is an AUTOMORPHISM of G defined by

$$\sigma_a(x) = axa^{-1}.$$

see also AUTOMORPHISM, AUTOMORPHISM GROUP

Inner Product

see DOT PRODUCT

Inner Product Space

An inner product space is a VECTOR SPACE which has an INNER PRODUCT. If the INNER PRODUCT defines a NORM, then the inner product space is called a HILBERT SPACE.

see also HILBERT SPACE, INNER PRODUCT, NORM

Inradius

The radius of a TRIANGLE'S INCIRCLE or of a POLYHEDRON'S INSPHERE, denoted r. For a TRIANGLE,

$$r = \frac{1}{2}\sqrt{\frac{(b+c-a)(c+a-b)(a+b-c)}{a+b+c}} = \frac{\Delta}{s} \quad (1)$$

$$= 4R\sin(\tfrac{1}{2}\alpha_1)\sin(\tfrac{1}{2}\alpha_2)\sin(\tfrac{1}{2}\alpha_3), \quad (2)$$

where Δ is the AREA of the TRIANGLE, a, b, and c are the side lengths, s is the SEMIPERIMETER, and R is the CIRCUMRADIUS (Johnson 1929, p. 189).

Equation (1) can be derived easily using TRILINEAR COORDINATES. Since the INCENTER is equally spaced from all three sides, its trilinear coordinates are 1:1:1, and its exact trilinear coordinates are $r : r : r$. The ratio k of the exact trilinears to the homogeneous coordinates is given by

$$k = \frac{2\Delta}{a+b+c} = \frac{\Delta}{s}. \quad (3)$$

But since $k = r$ in this case,

$$r = k = \frac{\Delta}{s}, \quad (4)$$

Q. E. D.

Other equations involving the inradius include

$$Rr = \frac{abc}{4s} \quad (5)$$

$$\Delta^2 = rr_1r_2r_3 \quad (6)$$

$$\cos A + \cos B + \cos C = 1 + \frac{r}{R} \quad (7)$$

$$r = 2R\cos A\cos B\cos C \quad (8)$$

$$a^2 + b^2 + c^2 = 4rR + 8R^2, \quad (9)$$

where r_i are the EXRADII (Johnson 1929, pp. 189–191).

As shown in RIGHT TRIANGLE, the inradius of a RIGHT TRIANGLE of integral side lengths x, y, and z is also integral, and is given by

$$r = \frac{xy}{x+y+z}, \quad (10)$$

where z is the HYPOTENUSE. Let d be the distance between inradius r and CIRCUMRADIUS R, $d = \overline{rR}$. Then

$$R^2 - d^2 = 2Rr \quad (11)$$

$$\frac{1}{R-d} + \frac{1}{R+d} = \frac{1}{r} \quad (12)$$

(Mackay 1886–87). These and many other identities are given in Johnson (1929, pp. 186–190).

Expressing the MIDRADIUS ρ and CIRCUMRADIUS R in terms of the midradius gives

$$r = \frac{\rho^2}{\sqrt{\rho^2 + \frac{1}{4}a^2}} \quad (13)$$

$$r = \frac{R^2 - \frac{1}{4}a^2}{R} \quad (14)$$

for an ARCHIMEDEAN SOLID.

see also CARNOT'S THEOREM, CIRCUMRADIUS, MIDRADIUS

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, 1929.

Mackay, J. S. "Historical Notes on a Geometrical Theorem and its Developments [18th Century]." *Proc. Edinburgh Math. Soc.* **5**, 62–78, 1886–1887.

Mackay, J. S. "Formulas Connected with the Radii of the Incircle and Excircles of a Triangle." *Proc. Edinburgh Math. Soc.* **12**, 86–105.

Mackay, J. S. "Formulas Connected with the Radii of the Incircle and Excircles of a Triangle." *Proc. Edinburgh Math. Soc.* **13**, 103–104.

Inscribed

A geometric figure which touches only the sides (or interior) of another figure.

see also Circumscribed, Incenter, Incircle, Inradius

Inscribed Angle

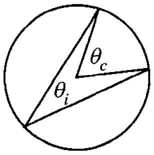

The Angle with Vertex on a Circle's Circumference formed by two points on a Circle's Circumference. For Angles with the same endpoints,

$$\theta_c = 2\theta_i,$$

where θ_c is the Central Angle.

see also Central Angle

References

Pedoe, D. *Circles: A Mathematical View, rev. ed.* Washington, DC: Math. Assoc. Amer., pp. xxi–xxii, 1995.

Inside-Outside Theorem

Let $P(z)$ and $Q(z)$ be Polynomials with Complex arguments and $\deg(Q) \geq \deg(P+2)$. Then

$$\int_\gamma \frac{P(z)}{Q(z)}\,dz = \begin{cases} 2\pi i \sum_{\text{inside }\gamma} \operatorname{Res} f(z) & \text{inside } \gamma \\ -2\pi i \sum_{\text{outside }\gamma} \operatorname{Res} f(z) & \text{outside } \gamma, \end{cases}$$

where Res are the Residues.

Insphere

A Sphere Inscribed in a given solid.

see also Circumsphere, Midsphere

Instrument Function

The finite Fourier Cosine Transform of an Apodization Function, also known as an Apparatus Function. The instrument function $I(k)$ corresponding to a given Apodization Function $A(x)$ is then given by

$$I(k) = \int_{-a}^{a} \cos(2\pi kx) A(x)\,dx.$$

see also Apodization Function, Fourier Cosine Transform

Insufficient Reason Principle

A principle also called the Indifference Principle which was first enunciated by Johann Bernoulli. The insufficient reason principle states that, if we are ignorant of the ways an event can occur and therefore have no reason to believe that one way will occur preferentially to another, it will occur equally likely in any way.

Int

see Integer Part

Integer

One of the numbers $\ldots, -2, -1, 0, 1, 2, \ldots$. The Set of Integers forms a Ring which is denoted $\mathbb{Z}$. A given Integer n may be Negative ($n \in \mathbb{Z}^-$), Nonnegative ($n \in \mathbb{Z}^*$), Zero ($n = 0$), or Positive ($n \in \mathbb{Z}^+ = \mathbb{N}$). The Ring $\mathbb{Z}$ has Cardinality of $\aleph_0$. The Generating Function for the Positive Integers is

$$f(x) = \frac{1}{(1-x)^2} = x + 2x^2 + 3x^3 + 4x^4 + \ldots.$$

There are several symbols used to perform operations having to do with conversion between Real Numbers and integers. The symbol $\lfloor x \rfloor$ ("Floor x") means "the largest integer not greater than x," i.e., `int(x)` in computer parlance. The symbol $[x]$ means "the nearest integer to x" (Nint), i.e., `nint(x)` in computer parlance. The symbol $\lceil x \rceil$ ("Ceiling x") means "the smallest integer not smaller x," or `-int(-x)`, where `int(x)` is the Integer Part of x.

see also Algebraic Integer, Almost Integer, Complex Number, Counting Number, Cyclotomic Integer, Eisenstein Integer, Gaussian Integer, $\mathbb{N}$, Natural Number, Negative, Positive, Radical Integer, Real Number, Whole Number, $\mathbb{Z}$, $\mathbb{Z}^-$, $\mathbb{Z}^+$, $\mathbb{Z}^*$, Zero

Integer Division

Division in which the fractional part (remainder) is discarded is called integer division and is sometimes denoted $\backslash$. Integer division can be defined as $a\backslash b \equiv \lfloor a/b \rfloor$, where "/" denotes normal division and $\lfloor x \rfloor$ is the Floor Function. For example,

$$\begin{aligned} 10/3 &= 3 + 1/3 \\ 10\backslash 3 &= 3. \end{aligned}$$

Integer Factorization

see Prime Factorization

Integer-Matrix Form

Let $Q(x) \equiv Q(\mathbf{x}) = Q(x_1, x_2, \ldots, x_n)$ be an integer-valued n-ary Quadratic Form, i.e., a Polynomial with integer Coefficients which satisfies $Q(x) > 0$ for Real $x \neq 0$. Then $Q(x)$ can be represented by

$$Q(x) = \mathbf{x}^{\mathrm{T}} \mathsf{A} \mathbf{x},$$

where

$$\mathsf{A} = \frac{1}{2} \frac{\partial^2 Q(x)}{\partial x_i \partial x_j}$$

is a POSITIVE symmetric matrix (Duke 1997). If A has POSITIVE entries, then $Q(x)$ is called an integer matrix form. Conway *et al.* (1997) have proven that, if a POSITIVE integer matrix quadratic form represents each of 1, 2, 3, 5, 6, 7, 10, 14, and 15, then it represents all POSITIVE INTEGERS.

see also FIFTEEN THEOREM

References
Conway, J. H.; Guy, R. K.; Schneeberger, W. A.; and Sloane, N. J. A. "The Primary Pretenders." *Acta Arith.* **78**, 307–313, 1997.
Duke, W. "Some Old Problems and New Results about Quadratic Forms." *Not. Amer. Math. Soc.* **44**, 190–196, 1997.

Integer Module

see ABELIAN GROUP

Integer Part

The function int(x) gives the INTEGER PART of x. In many computer languages, the function is denoted `int(x)`, but in mathematics, it is usually called the FLOOR FUNCTION and denoted $\lfloor x \rfloor$.

see also CEILING FUNCTION, FLOOR FUNCTION, NINT

Integer Relation

A set of REAL NUMBERS $x_1, \ldots, x_n$ is said to possess an integer relation if there exist integers a_i such that

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0,$$

with not all $a_i = 0$. An interesting example of such a relation is the 17-VECTOR $(1, x, x^2, \ldots, x^{16})$ with $x = 3^{1/4} - 2^{1/4}$, which has an integer relation (1, 0, 0, 0, −3860, 0, 0, 0, −666, 0, 0, 0, −20, 0, 0, 0, 1), i.e.,

$$1 - 3860x^4 - 666x^8 - 20x^{12} + x^{16} = 0.$$

This is a special case of finding the polynomial of degree $n = rs$ satisfied by $x = 3^{1/r} - 2^{1/s}$.

Algorithms for finding integer relations include the FERGUSON-FORCADE ALGORITHM, HJLS ALGORITHM, LLL ALGORITHM, PSLQ ALGORITHM, PSOS ALGORITHM, and the algorithm of Lagarias and Odlyzko (1985). Perhaps the simplest (and unfortunately most inefficient) such algorithm is the GREEDY ALGORITHM. Plouffe's "Inverse Symbolic Calculator" site includes a huge 54 million database of REAL NUMBERS which are algebraically related to fundamental mathematical constants.

see also CONSTANT PROBLEM, FERGUSON-FORCADE ALGORITHM, GREEDY ALGORITHM, HERMITE-LINDEMANN THEOREM, HJLS ALGORITHM, LATTICE REDUCTION, LLL ALGORITHM, PSLQ ALGORITHM, PSOS ALGORITHM, REAL NUMBER, LINDEMANN-WEIERSTRAß THEOREM

References
Bailey, D. and Plouffe, S. "Recognizing Numerical Constants." `http://www.cecm.sfu.ca/organics/papers/bailey`.
Lagarias, J. C. and Odlyzko, A. M. "Solving Low-Density Subset Sum Problems." *J. ACM* **32**, 229–246, 1985.
Plouffe, S. "Inverse Symbolic Calculator." `http://www.cecm.sfu.ca/projects/ISC/`.

Integer Sequence

A SEQUENCE whose terms are INTEGERS. The most complete printed references for such sequences are Sloane (1973) and its update, Sloane and Plouffe (1995). Sloane also maintains the sequences from both works together with many additional sequences in an on-line listing. In this listing, sequences are identified by a unique 6-DIGIT A-number. Sequences appearing in Sloane and Plouffe (1995) are ordered lexicographically and identified with a 4-DIGIT M-number, and those appearing in Sloane (1973) are identified with a 4-DIGIT N-number.

Sloane's huge (and enjoyable) database is accessible by either e-mail or web browser. To look up sequences by e-mail, send a message to either `sequences@research.att.com` or `superseeker@research.att.com` containing lines of the form `lookup 5 14 42 132 ....` To use the browser version, point to `http://www.research.att.com/~njas/sequences/eisonline.html`.

see also ARONSON'S SEQUENCE, COMBINATORICS, CONSECUTIVE NUMBER SEQUENCES, CONWAY SEQUENCE, EBAN NUMBER, HOFSTADTER-CONWAY $10,000 SEQUENCE, HOFSTADTER'S Q-SEQUENCE, LEVINE-O'SULLIVAN SEQUENCE, LOOK AND SAY SEQUENCE, MALLOW'S SEQUENCE, MIAN-CHOWLA SEQUENCE, MORSE-THUE SEQUENCE, NEWMAN-CONWAY SEQUENCE, NUMBER, PADOVAN SEQUENCE, PERRIN SEQUENCE, RATS SEQUENCE, SEQUENCE, SMARANDACHE SEQUENCES

References
Aho, A. V. and Sloane, N. J. A. "Some Doubly Exponential Sequences." *Fib. Quart.* **11**, 429–437, 1973.
Bernstein, M. and Sloane, N. J. A. "Some Canonical Sequences of Integers." *Linear Algebra and Its Applications* **226–228**, 57–72, 1995.
Erdős, P.; Sárközy, E.; and Szemerédi, E. "On Divisibility Properties of Sequences of Integers." In *Number Theory, Colloq. Math. Soc. János Bolyai, Vol. 2.* Amsterdam, Netherlands: North-Holland, pp. 35–49, 1970.
Guy, R. K. "Sequences of Integers." Ch. E in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 199–239, 1994.
Krattenthaler, C. "RATE: A Mathematica Guessing Machine." `http://radon.mat.univie.ac.at/People/kratt/rate/rate.html`.
Ostman, H. *Additive Zahlentheorie I, II.* Heidelberg, Germany: Springer-Verlag, 1956.
Pomerance, C. and Sárközy, A. "Combinatorial Number Theory." In *Handbook of Combinatorics* (Ed. R. Graham, M. Grötschel, and L. Lovász). Amsterdam, Netherlands: North-Holland, 1994.
Ruskey, F. "The (Combinatorial) Object Server." `http://sue.csc.uvic.ca/~cos`.
Sloane, N. J. A. *A Handbook of Integer Sequences.* Boston, MA: Academic Press, 1973.

Sloane, N. J. A. "Find the Next Term." *J. Recr. Math.* **7**, 146, 1974.

Sloane, N. J. A. "An On-Line Version of the Encyclopedia of Integer Sequences." *Elec. J. Combin.* **1**, F1 1–5, 1994. `http://www.combinatorics.org/Volume_1/volume1.html#F1`.

Sloane, N. J. A. "Some Important Integer Sequences." In *CRC Standard Mathematical Tables and Formulae* (Ed. D. Zwillinger). Boca Raton, FL: CRC Press, 1995.

Sloane, N. J. A. "An On-Line Version of the Encyclopedia of Integer Sequences." `http://www.research.att.com/~njas/sequences/eisonline.html`.

Sloane, N. J. A. and Plouffe, S. *The Encyclopedia of Integer Sequences.* San Diego, CA: Academic Press, 1995.

Stöhr, A. "Gelöste und ungelöste Fragen über Basen der natürlichen Zahlenreihe I, II." *J. reine angew. Math.* **194**, 40–65 and 111–140, 1955.

Turán, P. (Ed.). *Number Theory and Analysis: A Collection of Papers in Honor of Edmund Landau (1877–1938).* New York: Plenum Press, 1969.

Weisstein, E. W. "Integer Sequences." `http://www.astro.virginia.edu/~eww6n/math/notebooks/IntegerSequences.m`.

Integrable

A function for which the INTEGRAL can be computed is said to be integrable.

see also DIFFERENTIABLE, INTEGRAL, INTEGRATION

Integral

An integral is a mathematical object which can be interpreted as an AREA or a generalization of AREA. Integrals, together with DERIVATIVES, are the fundamental objects of CALCULUS. Other words for integral include ANTIDERIVATIVE and PRIMITIVE. The RIEMANN INTEGRAL is the simplest integral definition and the only one usually encountered in elementary CALCULUS. The RIEMANN INTEGRAL of the function $f(x)$ over x from a to b is written

$$\int_a^b f(x)\,dx. \tag{1}$$

Every definition of an integral is based on a particular MEASURE. For instance, the RIEMANN INTEGRAL is based on JORDAN MEASURE, and the LEBESGUE INTEGRAL is based on LEBESGUE MEASURE. The process of computing an integral is called INTEGRATION (a more archaic term for INTEGRATION is QUADRATURE), and the approximate computation of an integral is termed NUMERICAL INTEGRATION.

There are two classes of (Riemann) integrals: DEFINITE INTEGRALS

$$\int_a^b f(x)\,dx, \tag{2}$$

which have upper and lower limits, and INDEFINITE INTEGRALS, which are written without limits. The first FUNDAMENTAL THEOREM OF CALCULUS allows DEFINITE INTEGRALS to be computed in terms of INDEFINITE INTEGRALS, since if F is the INDEFINITE INTEGRAL for $f(x)$, then

$$\int_a^b f(x)\,dx = F(b) - F(a). \tag{3}$$

Wolfram Research (`http://www.integrals.com`) maintains a web site which will integrate many common (and not so common) functions. However, it cannot solve some simple integrals such as

$$\int \left[\frac{d}{dx}(x\sqrt{\sin x})\right] dx = \int \left(\frac{x\cos x}{2\sqrt{\sin x}} + \sqrt{\sin x}\right) dx \tag{4}$$

$$\int \left[\frac{d}{dx}L_2(x\ln x)\right] dx = -\int \left[\frac{(\ln x + 1)\ln(1 - x\ln x)}{x\ln x}\right] dx, \tag{5}$$

where L_2 is the DILOGARITHM. Furthermore, it gives an incorrect answer of $\pi^{1-2\sqrt{3}}/(\sqrt{3}\cdot 4^{\sqrt{3}})$ to

$$I(\sqrt{3}) = \int_0^{\pi/2} \frac{dx}{1 + (\tan x)^{\sqrt{3}}} = \tfrac{1}{2}\pi. \tag{6}$$

This integral and, in fact, the generalized integral for arbitrary a

$$I(a) = \int_0^{\pi/2} \frac{dx}{1 + (\tan x)^a}, \tag{7}$$

have a "trick" solution which takes advantage of the trigonometric identity

$$\tan(\tfrac{1}{2}\pi - x) = \cot x. \tag{8}$$

Letting $z \equiv (\tan x)^a$,

$$\begin{aligned} I(a) &= \int_0^{\pi/4} \frac{dx}{1+z} + \int_{\pi/4}^{\pi/2} \frac{dx}{1+1} \\ &= \int_0^{\pi/4} \frac{dx}{1+z} + \int_0^{\pi/4} \frac{dx}{1+\frac{1}{z}} \\ &= \int_0^{\pi/4} \left(\frac{1}{1+z} + \frac{1}{1+\frac{1}{z}}\right) = \int_0^{\pi/4} dx \\ &= \tfrac{1}{4}\pi. \end{aligned} \tag{9}$$

Here is a list of common INDEFINITE INTEGRALS:

$$\int x^r\,dx = \frac{x^{r+1}}{r+1} + C \tag{10}$$

$$\int \frac{dx}{x} = \ln|x| + C \tag{11}$$

$$\int a^x\,dx = \frac{a^x}{\ln a} + C \tag{12}$$

$$\int \sin x\,dx = -\cos x + C \tag{13}$$

$$\int \cos x\,dx = \sin x + C \tag{14}$$

$$\int \tan x\,dx = \ln|\sec x| + C \tag{15}$$

$$\int \csc x\,dx = \ln|\csc x - \cot x| + C \tag{16}$$

$$= \ln\left[\tan(\tfrac{1}{2}x)\right] + C \tag{17}$$

$$= \frac{1}{2}\ln\left(\frac{1-\cos x}{1+\cos x}\right) + C \tag{18}$$

$$\int \sec x\,dx = \ln|\sec x + \tan x| + C$$

$$= \operatorname{gd}^{-1}(x) + C \tag{19}$$

$$\int \cot x\,dx = \ln|\sin x| + C \tag{20}$$

$$\int \sec^2 x\,dx = \tan x + C \tag{21}$$

$$\int \csc^2 x\,dx = -\cot x + C \tag{22}$$

$$\int \sec x\,\tan x\,dx = \sec x + C \tag{23}$$

$$\int \cos^{-1} x\,dx = x\cos^{-1} x - \sqrt{1-x^2} + C \tag{24}$$

$$\int \sin^{-1} x\,dx = x\sin^{-1} x + \sqrt{1-x^2} + C \tag{25}$$

$$\int \tan^{-1} x\,dx = x\tan^{-1} x - \tfrac{1}{2}\ln(1+x^2) + C \tag{26}$$

$$\int \frac{dx}{\sqrt{a^2-x^2}} = \sin^{-1}\left(\frac{x}{a}\right) + C \tag{27}$$

$$\int \frac{dx}{\sqrt{a^2-x^2}} = \cos^{-1}\left(\frac{x}{a}\right) + C \tag{28}$$

$$\int \frac{dx}{a^2+x^2} = \frac{1}{a}\tan^{-1}\left(\frac{x}{a}\right) + C \tag{29}$$

$$\int \frac{dx}{a^2+x^2} = -\frac{1}{a}\cot^{-1}\left(\frac{x}{a}\right) + C \tag{30}$$

$$\int \frac{dx}{x\sqrt{x^2-a^2}} = \frac{1}{a}\sec^{-1}\left(\frac{x}{a}\right) + C \tag{31}$$

$$\int \frac{dx}{x\sqrt{x^2-a^2}} = -\frac{1}{a}\csc^{-1}\left(\frac{x}{a}\right) + C \tag{32}$$

$$\int \sin^2(ax)\,dx = \frac{x}{2} - \frac{1}{4a}\sin(2ax) + C \tag{33}$$

$$\int \operatorname{sn} u\,du = k^{-1}\ln(\operatorname{dn} u - k\operatorname{cn} u) + C \tag{34}$$

$$\int \operatorname{sn}^2 u\,du = \frac{u - E(u)}{k^2} + C \tag{35}$$

$$\int \operatorname{cn} u\,du = k^{-1}\sin^{-1}(k\operatorname{sn} u) + C \tag{36}$$

$$\int \operatorname{dn} u\,du = \sin^{-1}(\operatorname{sn} u) + C, \tag{37}$$

where $\sin x$ is the SINE; $\cos x$ is the COSINE; $\tan x$ is the TANGENT; $\csc x$ is the COSECANT; $\sec x$ is the SECANT; $\cot x$ is the COTANGENT; $\cos^{-1} x$ is the INVERSE COSINE; $\sin^{-1} x$ is the INVERSE SINE; $\tan^{-1}$ is the INVERSE TANGENT; $\operatorname{sn} u$, $\operatorname{cn} u$, and $\operatorname{dn} u$ are JACOBI ELLIPTIC FUNCTIONS; $E(u)$ is a complete ELLIPTIC INTEGRAL OF THE SECOND KIND; and $\operatorname{gd}(x)$ is the GUDERMANNIAN FUNCTION.

To derive (15), let $u \equiv \cos x$, so $du = -\sin x\,dx$ and

$$\begin{aligned}\int \tan x &= \int \frac{\sin x}{\cos x}\,dx = -\int \frac{du}{u}\\ &= -\ln|u| + C = -\ln|\cos x| + C\\ &= \ln|\cos x|^{-1} + C = \ln|\sec x| + C.\end{aligned} \tag{38}$$

To derive (18), let $u \equiv \csc x - \cot x$, so $du = (-\csc x\cot x + \csc^2 x)\,dx$ and

$$\begin{aligned}\int \csc x\,dx &= \int \csc x\frac{\csc x - \cot x}{\csc x - \cot x}\,dx\\ &= \int \frac{\csc^2 x + \cot x\csc x}{\csc x + \cot x}\,dx\\ &= \int \frac{du}{u} = \ln|u| + C\\ &= \ln|\csc x - \cot x| + C.\end{aligned} \tag{39}$$

To derive (19), let

$$u \equiv \sec x + \tan x, \tag{40}$$

so

$$du = (\sec x\tan x + \sec^2 x)\,dx \tag{41}$$

and

$$\begin{aligned}\int \sec x\,dx &= \int \sec x\frac{\sec x + \tan x}{\sec x + \tan x}\,dx\\ &= \int \frac{\sec^2 x + \sec x + \tan x}{\sec x + \tan x}\,dx\\ &= \int \frac{du}{u} = \ln|u| + C\\ &= \ln|\sec x + \tan x| + C.\end{aligned} \tag{42}$$

To derive (20), let $u \equiv \sin x$, so $du = \cos x\,dx$ and

$$\begin{aligned}\int \cot x\,dx &= \int \frac{\cos x}{\sin x}\,dx = \int \frac{du}{u} \\ &= \ln|u| + C = \ln|\sin x| + C. \qquad (43)\end{aligned}$$

Differentiating integrals leads to some useful and powerful identities, for instance

$$\frac{d}{dx}\int_a^x f(x)\,dx = f(x), \qquad (44)$$

which is the first FUNDAMENTAL THEOREM OF CALCULUS.

$$\frac{d}{dx}\int_x^b f(x)\,dx = -f(x) \qquad (45)$$

$$\frac{d}{dx}\int_a^b f(x,t)\,dt = \int_a^b \frac{\partial}{\partial x} f(x,t)\,dt \qquad (46)$$

$$\frac{d}{dx}\int_a^x f(x,t)\,dt = f(x,t) + \int_a^x \frac{\partial}{\partial x} f(x,t)\,dt. \qquad (47)$$

If $f(x,t)$ is singular or INFINITE, then

$$\begin{aligned}&\frac{d}{dx}\int_a^x f(x,t)\,dx \\ &= \frac{1}{x-a}\int_a^x \left[(x-a)\frac{\partial f}{\partial x} + (t-a)\frac{\partial f}{\partial t} + f\right] dt. \qquad (48)\end{aligned}$$

The LEIBNIZ IDENTITY is

$$\begin{aligned}\frac{d}{dx}\int_{u(x)}^{v(x)} f(x,t)\,dt &= v'(x) f(x, v(x)) - u' f(x, u(x)) \\ &\quad + \int_{u(x)}^{v(x)} \frac{\partial}{\partial x} f(x,t)\,dt. \qquad (49)\end{aligned}$$

Other integral identities include

$$\int_a^x \int_a^x f(t)\,dt\,dx = \int_a^x (x-t) f(t)\,dt \qquad (50)$$

$$\begin{aligned}&\int_0^x dt_n \int_0^{t_n} dt_{n-1} \cdots \int_0^{t_3} dt_2 \int_0^{t_2} f(t_1)\,dt_1 \\ &\qquad = \frac{1}{(n-1)!}\int_0^x (x-t)^{n-1} f(t)\,dt \qquad (51)\end{aligned}$$

$$\frac{\partial}{\partial x_k}(x_j J_k) = \delta_{jk} J_k + x_j \frac{\partial}{\partial x_k} J_k = \mathbf{J} + \mathbf{r}\nabla\cdot\mathbf{J} \qquad (52)$$

$$\begin{aligned}\int_V \mathbf{J}\,d^3\mathbf{r} &= \int_V \frac{\partial}{\partial x_k}(x_i J_k) - \int_V \mathbf{r}\nabla\cdot\mathbf{J}\,d^3\mathbf{r} \\ &= -\int_V \mathbf{r}\nabla\cdot\mathbf{J}\,d^3\mathbf{r}. \qquad (53)\end{aligned}$$

Integrals of the form

$$\int_a^b f(x)\,dx \qquad (54)$$

with one INFINITE LIMIT and the other NONZERO may be expressed as finite integrals over transformed functions. If $f(x)$ decreases at least as fast as $1/x^2$, then let

$$t \equiv \frac{1}{x} \qquad (55)$$

$$dt = -\frac{dx}{x^2} \qquad (56)$$

$$dx = -x^2\,dt = -\frac{dt}{t^2}, \qquad (57)$$

and

$$\int_a^b f(x)\,dx = -\int_{1/a}^{1/b} \frac{1}{t^2} f\left(\frac{1}{t}\right) dt = \int_{1/b}^{1/a} \frac{1}{t^2} f\left(\frac{1}{t}\right) dt. \qquad (58)$$

If $f(x)$ diverges as $(x-a)^\gamma$ for $\gamma \in [0, 1]$, let

$$x \equiv t^{1/(1-\gamma)} + a \qquad (59)$$

$$\begin{aligned}dx &= \frac{1}{1-\gamma} t^{(1/1-\gamma)-1}\,dt = \frac{1}{1-\gamma} t^{[1-(1-\gamma)]/(1-\gamma)}\,dt \\ &= \frac{1}{\gamma-1} t^{\gamma/(1-\gamma)}\,dt \qquad (60)\end{aligned}$$

$$t = (x-a)^{1-\gamma}, \qquad (61)$$

and

$$\begin{aligned}\int_a^b f(x)\,dx &= \frac{1}{1-\gamma} \\ &= \int_0^{(b-a)^{1-\gamma}} t^{\gamma/(1-\gamma)} f(t^{1/(1-\gamma)} + a)\,dt. \qquad (62)\end{aligned}$$

If $f(x)$ diverges as $(x+b)^\gamma$ for $\gamma \in [0, 1]$, let

$$x \equiv b - t^{1/(1-\gamma)} \qquad (63)$$

$$dx = -\frac{1}{\gamma-1} t^{\gamma/(1-\gamma)}\,dt \qquad (64)$$

$$t = (b-x)^{1-\gamma}, \qquad (65)$$

and

$$\begin{aligned}\int_a^b f(x)\,dx &= \frac{1}{1-\gamma} \\ &= \int_0^{(b-a)^{1-\gamma}} t^{\gamma/(1-\gamma)} f(b - t^{1/(1-\gamma)})\,dt. \qquad (66)\end{aligned}$$

If the integral diverges exponentially, then let

$$t \equiv e^{-x} \qquad (67)$$

$$dt = -e^{-x}\,dx \qquad (68)$$

$$x = -\ln t, \qquad (69)$$

and

$$\int_a^\infty f(x)\,dx = \int_0^{e^{-a}} f(-\ln t)\frac{dt}{t}. \tag{70}$$

Integrals with rational exponents can often be solved by making the substitution $u = x^{1/n}$, where n is the Least Common Multiple of the Denominator of the exponents.

Integration rules include

$$\int_a^a f(x)\,dx = 0 \tag{71}$$

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx. \tag{72}$$

For $c \in (a,b)$,

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx. \tag{73}$$

If g' is continuous on $[a,b]$ and f is continuous and has an antiderivative on an Interval containing the values of $g(x)$ for $a \leq x \leq b$, then

$$\int_a^b f(g(x))g'(x)\,dx = \int_{g(a)}^{g(b)} f(u)\,du. \tag{74}$$

Liouville showed that the integrals

$$\int e^{-x^2}\,dx \quad \int \frac{e^x}{x}\,dx \quad \int \frac{\sin x}{x}\,dx \quad \int \frac{dx}{\ln x} \tag{75}$$

cannot be expressed as terms of a finite number of elementary functions. Other irreducibles include

$$\int x^x\,dx \quad \int x^{-x}\,dx \quad \int \sqrt{\sin x}\,dx. \tag{76}$$

Chebyshev proved that if U, V, and W are Rational Numbers, then

$$\int x^U (A + Bx^V)^W\,dx \tag{77}$$

is integrable in terms of elementary functions Iff $(U+1)/V$, W, or $W + (U+1)/V$ is an Integer (Ritt 1948, Shanks 1993).

There are a wide range of methods available for Numerical Integration. A good source for such techniques is Press *et al.* (1992). The most straightforward numerical integration technique uses the Newton-Cotes Formulas (also called Quadrature Formulas), which approximate a function tabulated at a sequence of regularly spaced Intervals by various degree Polynomials. If the endpoints are tabulated, then the 2- and 3-point formulas are called the Trapezoidal Rule and Simpson's Rule, respectively. The 5-point formula is called Bode's Rule. A generalization of the Trapezoidal Rule is Romberg Integration, which can yield accurate results for many fewer function evaluations.

If the analytic form of a function is known (instead of its values merely being tabulated at a fixed number of points), the best numerical method of integration is called Gaussian Quadrature. By picking the optimal Abscissas at which to compute the function, Gaussian quadrature produces the most accurate approximations possible. However, given the speed of modern computers, the additional complication of the Gaussian Quadrature formalism often makes it less desirable than the brute-force method of simply repeatedly calculating twice as many points on a regular grid until convergence is obtained. An excellent reference for Gaussian Quadrature is Hildebrand (1956).

see also A-Integrable, Abelian Integral, Calculus, Chebyshev-Gauss Quadrature, Chebyshev Quadrature, Darboux Integral, Definite Integral, Denjoy Integral, Derivative, Double Exponential Integration, Euler Integral, Fundamental Theorem of Gaussian Quadrature, Gauss-Jacobi Mechanical Quadrature, Gaussian Quadrature, Haar Integral, Hermite-Gauss Quadrature, Hermite Quadrature, HK Integral, Indefinite Integral, Integration, Jacobi-Gauss Quadrature, Jacobi Quadrature, Laguerre-Gauss Quadrature, Laguerre Quadrature, Lebesgue Integral, Lebesgue-Stieltjes Integral, Legendre-Gauss Quadrature, Legendre Quadrature, Lobatto Quadrature, Mechanical Quadrature, Mehler Quadrature, Newton-Cotes Formulas, Numerical Integration, Peron Integral, Quadrature, Radau Quadrature, Recursive Monotone Stable Quadrature, Riemann-Stieltjes Integral, Romberg Integration, Riemann Integral, Stieltjes Integral

References

Beyer, W. H. "Integrals." *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 233–296, 1987.

Bronstein, M. *Symbolic Integration I: Transcendental Functions.* New York: Springer-Verlag, 1996.

Gordon, R. A. *The Integrals of Lebesgue, Denjoy, Perron, and Henstock.* Providence, RI: Amer. Math. Soc., 1994.

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, 1993.

Hildebrand, F. B. *Introduction to Numerical Analysis.* New York: McGraw-Hill, pp. 319–323, 1956.

Piessens, R.; de Doncker, E.; Uberhuber, C. W.; and Kahaner, D. K. *QUADPACK: A Subroutine Package for Automatic Integration.* New York: Springer-Verlag, 1983.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Integration of Functions." Ch. 4 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing,* *2nd ed.* Cambridge, England: Cambridge University Press, pp. 123–158, 1992.

Ritt, J. F. *Integration in Finite Terms.* New York: Columbia University Press, p. 37, 1948.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 145, 1993.

Wolfram Research. "The Integrator." `http://www.integrals.com`

Integral Brick

see EULER BRICK

Integral Calculus

That portion of "the" CALCULUS dealing with INTEGRALS.

see also CALCULUS, DIFFERENTIAL CALCULUS, INTEGRAL

Integral Cuboid

see EULER BRICK

Integral Current

A RECTIFIABLE CURRENT whose boundary is also a RECTIFIABLE CURRENT.

Integral Curvature

Given a GEODESIC TRIANGLE (a triangle formed by the arcs of three GEODESICS on a smooth surface),

$$\int_{ABC} K\,da = A + B + C - \pi.$$

Given the EULER CHARACTERISTIC χ,

$$\iint K\,da = 2\pi\chi,$$

so the integral curvature of a closed surface is not altered by a topological transformation.

see also GAUSS-BONNET FORMULA, GEODESIC TRIANGLE

Integral Domain

A RING that is COMMUTATIVE under multiplication, has a unit element, and has no divisors of 0. The INTEGERS form an integral domain.

see also FIELD, RING

Integral Drawing

A GRAPH drawn such that the EDGES have only INTEGER lengths. It is conjectured that every PLANAR GRAPH has an integral drawing.

References

Harborth, H. and Möller, M. "Minimum Integral Drawings of the Platonic Graphs." *Math. Mag.* **67**, 355–358, 1994.

Integral Equation

If the limits are fixed, an integral equation is called a Fredholm integral equation. If one limit is variable, it is called a Volterra integral equation. If the unknown function is only under the integral sign, the equation is said to be of the "first kind." If the function is both inside and outside, the equation is called of the "second kind." A Fredholm equation of the first kind is of the form

$$f(x) = \int_a^b K(x,t)\phi(t)\,dt. \tag{1}$$

A Fredholm equation of the second kind is of the form

$$\phi(x) = f(x) + \lambda \int_a^b K(x,t)\phi(t)\,dt. \tag{2}$$

A Volterra equation of the first kind is of the form

$$f(x) = \int_a^x K(x,t)\phi(t)\,dt. \tag{3}$$

A Volterra equation of the second kind is of the form

$$\phi(x) = f(x) + \int_a^x K(x,t)\phi(t)\,dt, \tag{4}$$

where the functions $K(x,t)$ are known as KERNELS. Integral equations may be solved directly if they are SEPARABLE. Otherwise, a NEUMANN SERIES must be used.

A KERNEL is separable if

$$K(x,t) = \lambda \sum_{j=1}^n M_j(x)N_j(t). \tag{5}$$

This condition is satisfied by all POLYNOMIALS and many TRANSCENDENTAL FUNCTIONS. A FREDHOLM INTEGRAL EQUATION OF THE SECOND KIND with separable KERNEL may be solved as follows:

$$\begin{aligned}\phi(x) &= f(x) + \int_a^b K(x,t)\phi(t)\,dt\\ &= f(x) + \lambda \sum_{j=1}^n M_j(x) \int_a^b N_j(t)\phi(t)\,dt\\ &= f(x) + \lambda \sum_{j=1}^n c_j M_j(x),\end{aligned} \tag{6}$$

where

$$c_j \equiv \int_a^b N_j(t)\phi(t)\,dt. \tag{7}$$

Now multiply both sides of (7) by $N_i(x)$ and integrate over dx.

$$\begin{aligned}&\int_a^b \phi(x)N_i(x)\,dx\\ &= \int_a^b f(x)N_i(x)\,dx + \lambda \sum_{j=1}^n c_j \int_a^b M_j(x)N_i(x)\,dx.\end{aligned} \tag{8}$$

By (7), the first term is just c_i. Now define

$$b_i \equiv \int_a^b N_i(x) f(x)\,dx \tag{9}$$

$$a_{ij} = \int_a^b N_i(x) M_j(x)\,dx, \tag{10}$$

so (8) becomes

$$c_i = b_i + \lambda \sum_{j=1}^{n} a_{ij} c_j. \tag{11}$$

Writing this in matrix form,

$$\mathbf{C} = \mathbf{B} + \lambda \mathsf{A}\mathbf{C}, \tag{12}$$

so

$$(\mathsf{I} - \lambda\mathsf{A})\mathbf{C} = \mathbf{B} \tag{13}$$

$$\mathbf{C} = (\mathsf{I} - \lambda\mathsf{A})^{-1}\mathbf{B}. \tag{14}$$

see also FREDHOLM INTEGRAL EQUATION OF THE FIRST KIND, FREDHOLM INTEGRAL EQUATION OF THE SECOND KIND, VOLTERRA INTEGRAL EQUATION OF THE FIRST KIND, VOLTERRA INTEGRAL EQUATION OF THE SECOND KIND

References
Corduneanu, C. *Integral Equations and Applications.* Cambridge, England: Cambridge University Press, 1991.
Davis, H. T. *Introduction to Nonlinear Differential and Integral Equations.* New York: Dover, 1962.
Kondo, J. *Integral Equations.* Oxford, England: Clarendon Press, 1992.
Lovitt, W. V. *Linear Integral Equations.* New York: Dover, 1950.
Mikhlin, S. G. *Integral Equations and Their Applications to Certain Problems in Mechanics, Mathematical Physics and Technology, 2nd rev. ed.* New York: Macmillan, 1964.
Mikhlin, S. G. *Linear Integral Equations.* New York: Gordon & Breach, 1961.
Pipkin, A. C. *A Course on Integral Equations.* New York: Springer-Verlag, 1991.
Porter, D. and Stirling, D. S. G. *Integral Equations: A Practical Treatment, from Spectral Theory to Applications.* Cambridge, England: Cambridge University Press, 1990.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Integral Equations and Inverse Theory." Ch. 18 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 779–817, 1992.
Tricomi, F. G. *Integral Equations.* New York: Dover, 1957.

Integral of Motion

A function of the coordinates which is constant along a trajectory in PHASE SPACE. The number of DEGREES OF FREEDOM of a DYNAMICAL SYSTEM such as the DUFFING DIFFERENTIAL EQUATION can be decreased by one if an integral of motion can be found. In general, it is very difficult to discover integrals of motion.

Integral Sign

The symbol $\int$ used to denote an INTEGRAL $\int f(x)\,dx$. The symbol was chosen to be a stylized script "S" to stand for "summation."

see also INTEGRAL

Integral Test

Let $\sum u_k$ be a series with POSITIVE terms and let $f(x)$ be the function that results when k is replaced by x in the FORMULA for u_k. If f is decreasing and continuous for $x \geq 1$ and

$$\lim_{x\to\infty} f(x) = 0,$$

then

$$\sum_{k=1}^{\infty} u_k$$

and

$$\int_t^{\infty} f(x)\,dx$$

both converge or diverge, where $1 \leq t < \infty$. The test is also called the CAUCHY INTEGRAL TEST or MACLAURIN INTEGRAL TEST.

see also CONVERGENCE TESTS

References
Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 283–284, 1985.

Integral Transform

A general integral transform is defined by

$$g(\alpha) = \int_a^b f(t) K(\alpha, t)\,dt,$$

where $K(\alpha, t)$ is called the KERNEL of the transform.

see also FOURIER TRANSFORM, FOURIER-STIELTJES TRANSFORM, H-TRANSFORM, HADAMARD TRANSFORM, HANKEL TRANSFORM, HARTLEY TRANSFORM, HOUGH TRANSFORM, OPERATIONAL MATHEMATICS, RADON TRANSFORM, WAVELET TRANSFORM, Z-TRANSFORM

References
Arfken, G. "Integral Transforms." Ch. 16 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 794–864, 1985.
Carslaw, H. S. and Jaeger, J. C. *Operational Methods in Applied Mathematics.*
Davies, B. *Integral Transforms and Their Applications, 2nd ed.* New York: Springer-Verlag, 1985.
Poularikas, A. D. (Ed.). *The Transforms and Applications Handbook.* Boca Raton, FL: CRC Press, 1995.
Zayed, A. I. *Handbook of Function and Generalized Function Transformations.* Boca Raton, FL: CRC Press, 1996.

Integrand

The quantity being INTEGRATED, also called the KERNEL. For example, in $\int f(x)\,dx$, $f(x)$ is the integrand.

see also INTEGRAL, INTEGRATION

Integrating Factor

A FUNCTION by which an ORDINARY DIFFERENTIAL EQUATION is multiplied in order to make it integrable.

see also ORDINARY DIFFERENTIAL EQUATION

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 526–529, 1953.

Integration

The process of computing or obtaining an INTEGRAL. A more archaic term for integration is QUADRATURE.

see also CONTOUR INTEGRATION, INTEGRAL, INTEGRATION BY PARTS, MEASURE THEORY, NUMERICAL INTEGRATION

Integration Lattice

A discrete subset of $\mathbb{R}^s$ which is closed under addition and subtraction and which contains $\mathbb{Z}^s$ as a SUBSET.

see also LATTICE

References

Sloan, I. H. and Joe, S. *Lattice Methods for Multiple Integration.* New York: Oxford University Press, 1994.

Integration by Parts

A first-order (single) integration by parts uses

$$d(uv) = u\,dv + v\,du \tag{1}$$

$$\int d(uv) = uv = \int u\,dv + \int v\,du, \tag{2}$$

so

$$\int u\,dv = uv - \int v\,du \tag{3}$$

and

$$\int_a^b u\,dv = [uv]_a^b - \int_{f(a)}^{f(b)} v\,du. \tag{4}$$

Now apply this procedure n times to $\int f^{(n)}(x)g(x)\,dx$.

$$u = g(x) \qquad dv = f^{(n)}(x)\,dx \tag{5}$$

$$du = g'(x)\,dx \qquad v = f^{(n-1)}(x). \tag{6}$$

Therefore,

$$\int f^{(n)}g(x)\,dx = g(x)f^{(n-1)}(x) - \int f^{(n-1)}(x)g'(x)\,dx. \tag{7}$$

But

$$\int f^{(n-1)}(x)g'(x)\,dx = g'(x)f^{(n-2)}(x) - \int f^{(n-2)}(x)g''(x)\,dx \tag{8}$$

$$\int f^{(n-2)}(x)g''(x)\,dx = g''(x)f^{(n-3)}(x) - \int f^{(n-3)}(x)g^{(3)}(x)\,dx, \tag{9}$$

so

$$\int f^{(n)}(x)g(x)\,dx = g(x)f^{(n-1)}(x) - g'(x)f^{(n-2)}(x) + g''(x)f^{(n-3)}(x) - \ldots + (-1)^n \int f(x)g^{(n)}(x)\,dx. \tag{10}$$

Now consider this in the slightly different form $\int f(x)g(x)\,dx$. Integrate by parts a first time

$$u = f(x) \qquad dv = g(x)\,dx \tag{11}$$

$$du = f'(x)\,dx \qquad v = \int g(x)\,dx, \tag{12}$$

so

$$\int f(x)g(x)\,dx = f(x)\int g(x)\,dx - \int \left[\int g(x)\,dx\right] f'(x)\,dx. \tag{13}$$

Now integrate by parts a second time,

$$u = f'(x) \qquad dv = \int g(x)(dx)^2 \tag{14}$$

$$du = f''(x)\,dx \qquad v = \iint g(x)(dx)^2, \tag{15}$$

so

$$\int f(x)g(x)\,dx = f(x)\int g(x)\,dx - f'(x)\iint g(x)(dx)^2 + \int \left[\iint g(x)(dx)^2\right] f''(x)\,dx. \tag{16}$$

Repeating a third time,

$$\int f(x)g(x)\,dx = f(x)\int g(x)\,dx - f'(x)\iint g(x)(dx)^2 + f''(x)\iiint g(x)(dx)^3 - \int \left[\iiint g(x)(dx)^3\right] f'''(x)\,dx. \tag{17}$$

Therefore, after n applications,

$$\begin{aligned}\int f(x)g(x)\,dx &= f(x)\int g(x)\,dx - f'(x)\iint g(x)(dx)^2 \\ &\quad + f''(x)\iiint g(x)(dx)^3 - \ldots \\ &\quad + (-1)^{n+1} f^{(n)}(x) \underbrace{\int\cdots\int}_{n+1} g(x)(dx)^{n+1} \\ &\quad + (-1)^n \int \left[\underbrace{\int\cdots\int}_{n+1} g(x)(dx)^{n+1}\right] f^{(n+1)}(x)\,dx. \end{aligned} \tag{18}$$

If $f^{(n+1)}(x) = 0$ (e.g., for an nth degree POLYNOMIAL), the last term is 0, so the sum terminates after n terms and

$$\begin{aligned}\int f(x)g(x)\,dx &= f(x)\int g(x)\,dx \\ &\quad - f'(x)\iint g(x)(dx)^2 + f''(x)\iiint g(x)(dx)^3 - \ldots \\ &\quad + (-1)^{n+1} f^{(n)}(x) \underbrace{\int\cdots\int}_{n+1} g(x)(dx)^{n+1}. \end{aligned} \tag{19}$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 12, 1972.

Intension

A definition of a SET by mentioning a defining property.

see also EXTENSION

References

Russell, B. "Definition of Number." *Introduction to Mathematical Philosophy.* New York: Simon and Schuster, 1971.

Interchange Graph

see LINE GRAPH

Interest

Interest is a fee (or payment) made for the borrowing (or lending) of money. The two most common types of interest are SIMPLE INTEREST, for which interest is paid only on the initial PRINCIPAL, and COMPOUND INTEREST, for which interest earned can be re-invested to generate further interest.

see also COMPOUND INTEREST, CONVERSION PERIOD, RULE OF 72, SIMPLE INTEREST

References

Kellison, S. G. *Theory of Interest, 2nd ed.* Burr Ridge, IL: Richard D. Irwin, 1991.

Interior

That portion of a region lying "inside" a specified boundary. For example, the interior of the SPHERE is a BALL.

see also EXTERIOR

Interior Angle Bisector

see ANGLE BISECTOR

Intermediate Value Theorem

If f is continuous on a CLOSED INTERVAL $[a, b]$ and c is any number between $f(a)$ and $f(b)$ inclusive, there is at least one number x in the CLOSED INTERVAL such that $f(x) = c$.

see also WEIERSTRAß INTERMEDIATE VALUE THEOREM

Internal Bisectors Problem

see STEINER-LEHMUS THEOREM

Internal Knot

One of the knots $t_{p+1}, \ldots, t_{m-p-1}$ of a B-SPLINE with control points $\mathbf{P}_0, \ldots, \mathbf{P}_n$ and KNOT VECTOR

$$\mathbf{T} = \{t_0, t_1, \ldots, t_m\},$$

where

$$p \equiv m - n - 1.$$

see also B-SPLINE, KNOT VECTOR

Interpolation

The computation of points or values between ones that are known or tabulated using the surrounding points or values.

see also AITKEN INTERPOLATION, BESSEL'S INTERPOLATION FORMULA, EVERETT INTERPOLATION, EXTRAPOLATION, FINITE DIFFERENCE, GAUSS'S INTERPOLATION FORMULA, HERMITE INTERPOLATION, LAGRANGE INTERPOLATING POLYNOMIAL, NEWTON-COTES FORMULAS, NEWTON'S DIVIDED DIFFERENCE INTERPOLATION FORMULA, OSCULATING INTERPOLATION, THIELE'S INTERPOLATION FORMULA

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Interpolation." §25.2 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 878–882, 1972.

Iyanaga, S. and Kawada, Y. (Eds.). "Interpolation." Appendix A, Table 21 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 1482–1483, 1980.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Interpolation and Extrapolation." Ch. 3 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 99–122, 1992.

Interquartile Range

Divide a set of data into two groups (high and low) of equal size at the MEDIAN if there is an EVEN number of data points, or two groups consisting of points on either side of the MEDIAN itself plus the MEDIAN if there is an ODD number of data points. Find the MEDIANS of the low and high groups, denoting these first and third quartiles by Q_1 and Q_3. The interquartile range is then defined by

$$\text{IQR} \equiv Q_3 - Q_1.$$

see also H-SPREAD, HINGE, MEDIAN (STATISTICS)

Interradius

see MIDRADIUS

Intersection

The intersection of two sets A and B is the set of elements common to A and B. This is written $A \cap B$, and is pronounced "A intersection B" or "A cap B." The intersection of sets A_1 through A_n is written $\bigcap_{i=1}^{n} A_i$. The intersection of lines AB and CD is written $AB \cap CD$.

see also AND, UNION

Interspersion

An ARRAY $A = a_{ij}$, $i, j \geq 1$ of POSITIVE INTEGERS is called an interspersion if

1. The rows of A comprise a PARTITION of the POSITIVE INTEGERS,
2. Every row of A is an increasing sequence,
3. Every column of A is a (possibly FINITE) increasing sequence,
4. If (u_j) and (v_j) are distinct rows of A and if p and q are any indices for which $u_p < v_q < u_{p+1}$, then $u_{p+1} < v_{q+1} < u_{p+2}$.

If an array $A = a_{ij}$ is an interspersion, then it is a DISPERSION. If an array $A = a(i, j)$ is an interspersion, then the sequence $\{x_n\}$ given by $\{x_n = i : n = (i, j)\}$ for some j is a FRACTAL SEQUENCE. Examples of interspersion are the STOLARSKY ARRAY and WYTHOFF ARRAY.

see also DISPERSION (SEQUENCE), FRACTAL SEQUENCE, STOLARSKY ARRAY

References

Kimberling, C. "Interspersions and Dispersions." *Proc. Amer. Math. Soc.* **117**, 313–321, 1993.

Kimberling, C. "Fractal Sequences and Interspersions." *Ars Combin.* **45**, 157–168, 1997.

Intersphere

see MIDSPHERE

Interval

A collection of points on a LINE SEGMENT. If the endpoints a and b are FINITE and are included, the interval is called CLOSED and is denoted $[a, b]$. If one of the endpoints is $\pm\infty$, then the interval still contains all of its LIMIT POINTS, so $[a, \infty)$ and $(-\infty, b]$ are also closed intervals. If the endpoints are not included, the interval is called OPEN and denoted (a, b). If one endpoint is included but not the other, the interval is denoted $[a, b)$ or $(a, b]$ and is called a HALF-CLOSED (or HALF-OPEN) interval.

see also CLOSED INTERVAL, HALF-CLOSED INTERVAL, LIMIT POINT, OPEN INTERVAL, PENCIL

Interval Graph

A GRAPH $G = (V, E)$ is an interval graph if it captures the INTERSECTION RELATION for some set of INTERVALS on the REAL LINE. Formally, P is an interval graph provided that one can assign to each $v \in V$ an interval I_v such that $I_u \cap I_v$ is nonempty precisely when $uv \in E$.

see also COMPARABILITY GRAPH

References

Booth, K. S. and Lueker, G. S. "Testing for the Consecutive Ones Property, Interval Graphs, and Graph Planarity using PQ-Tree Algorithms." *J. Comput. System Sci.* **13**, 335–379, 1976.

Fishburn, P. C. *Interval Orders and Interval Graphs: A Study of Partially Ordered Sets.* New York: Wiley, 1985.

Gilmore, P. C. and Hoffman, A. J. "A Characterization of Comparability Graphs and of Interval Graphs." *Canad. J. Math.* **16**, 539–548, 1964.

Lekkerkerker, C. G. and Boland, J. C. "Representation of a Finite Graph by a Set of Intervals on the Real Line." *Fund. Math.* **51**, 45–64, 1962.

Interval Order

A POSET $P = (X, \leq)$ is an interval order if it is ISOMORPHIC to some set of INTERVALS on the REAL LINE ordered by left-to-right precedence. Formally, P is an interval order provided that one can assign to each $x \in X$ an INTERVAL $[x_L, x_R]$ such that $x_R < y_L$ in the REAL NUMBERS IFF $x < y$ in P.

see also PARTIALLY ORDERED SET

References

Fishburn, P. C. *Interval Orders and Interval Graphs: A Study of Partially Ordered Sets.* New York: Wiley, 1985.

Wiener, N. "A Contribution to the Theory of Relative Position." *Proc. Cambridge Philos. Soc.* **17**, 441–449, 1914.

Intrinsic Curvature

A CURVATURE such as GAUSSIAN CURVATURE which is detectable to the "inhabitants" of a surface and not just outside observers. An EXTRINSIC CURVATURE, on the other hand, is not detectable to someone who can't study the 3-dimensional space surrounding the surface on which he resides.

see also CURVATURE, EXTRINSIC CURVATURE, GAUSSIAN CURVATURE

Intrinsic Equation

An equation which specifies a CURVE in terms of intrinsic properties such as ARC LENGTH, RADIUS OF CURVATURE, and TANGENTIAL ANGLE instead of with reference to artificial coordinate axes. Intrinsic equations are also called NATURAL EQUATIONS.

see also CESÀRO EQUATION, NATURAL EQUATION, WHEWELL EQUATION

References

Yates, R. C. "Intrinsic Equations." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 123–126, 1952.

Intrinsically Linked

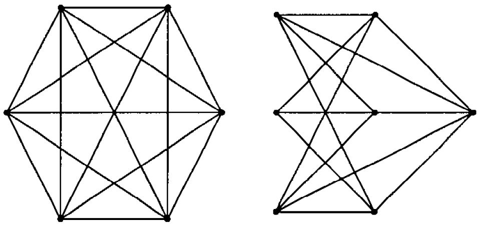

A GRAPH is intrinsically linked if any embedding of it in 3-D contains a nontrivial LINK. A GRAPH is intrinsically linked IFF it contains one of the seven PETERSEN GRAPHS (Robertson *et al.* 1993).

The COMPLETE GRAPH K_6 (left) is intrinsically linked because it contains at least two linked TRIANGLES. The COMPLETE k-PARTITE GRAPH $K_{3,3,1}$ (right) is also intrinsically linked.

see also COMPLETE GRAPH, COMPLETE k-PARTITE GRAPH, PETERSEN GRAPHS

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 217–221, 1994.

Robertson, N.; Seymour, P. D.; and Thomas, R. "Linkless Embeddings of Graphs in 3-Space." *Bull. Amer. Math. Soc.* **28**, 84–89, 1993.

Invariant

A quantity which remains unchanged under certain classes of transformations. Invariants are extremely useful for classifying mathematical objects because they usually reflect intrinsic properties of the object of study.

see ADIABATIC INVARIANT, ALEXANDER INVARIANT, ALGEBRAIC INVARIANT, ARF INVARIANT, INTEGRAL OF MOTION

References

Hunt, B. "Invariants." Appendix B.1 in *The Geometry of Some Special Arithmetic Quotients.* New York: Springer-Verlag, pp. 282–290, 1996.

Invariant Density

see NATURAL INVARIANT

Invariant (Elliptic Function)

The invariants of a WEIERSTRAß ELLIPTIC FUNCTION are defined by

$$g_2 \equiv 60\Sigma'\Omega_{mn}{}^{-4}$$
$$g_3 \equiv 140\Sigma'\Omega_{mn}{}^{-6}.$$

Here,

$$\Omega_{mn} \equiv 2m\omega_1 - 2n\omega_2,$$

where ω_1 and ω_2 are the periods of the ELLIPTIC FUNCTION.

Invariant Manifold

When stable and unstable invariant MANIFOLDS intersect, they do so in a HYPERBOLIC FIXED POINT (SADDLE POINT). The invariant MANIFOLDS are then called SEPARATRICES. A HYPERBOLIC FIXED POINT is characterized by two ingoing stable MANIFOLDS and two outgoing unstable MANIFOLDS. In integrable systems, incoming W^s and outgoing W^u MANIFOLDS all join up smoothly.

A stable invariant MANIFOLD W^s of a FIXED POINT Y^* is the set of all points Y_0 such that the trajectory passing through Y_0 tends to Y^* as $j \to \infty$.

An unstable invariant MANIFOLD W^u of a FIXED POINT Y^* is the set of all points Y_0 such that the trajectory passing through Y_0 tends to Y^* as $j \to -\infty$.

see also HOMOCLINIC POINT

Invariant Point

see FIXED POINT (TRANSFORMATION)

Invariant Subgroup

see NORMAL SUBGROUP

Inverse Cosecant

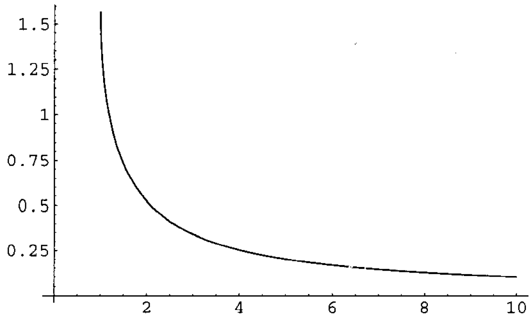

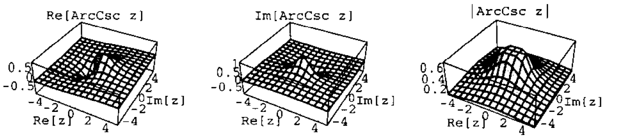

The function $\csc^{-1} x$, also denoted $\mathrm{arccsc}(x)$, where $\csc x$ is the COSECANT and the SUPERSCRIPT -1 denotes an

INVERSE FUNCTION, *not* the multiplicative inverse. The inverse cosecant satisfies

$$\csc^{-1} x = \sec^{-1}\left(\frac{x}{\sqrt{x^2-1}}\right) \tag{1}$$

for POSITIVE or NEGATIVE x, and

$$\csc^{-1} x = \pi + \csc^{-1}(-x) \tag{2}$$

for $x \geq 0$. The inverse cosecant is given in terms of other inverse trigonometric functions by

$$\csc^{-1} = \cos^{-1}\left(\frac{\sqrt{x^2-1}}{x}\right) \tag{3}$$

$$= \cot^{-1}(\sqrt{x^2-1}) \tag{4}$$

$$= \tfrac{1}{2}\pi - \sec^{-1} x = -\tfrac{1}{2}\pi - \sec^{-1}(-x) \tag{5}$$

$$= \sin^{-1}\left(\frac{1}{x}\right) \tag{6}$$

for $x \geq 0$.

see also COSECANT INVERSE SINE, SINE

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 142–143, 1987.

Inverse Cosine

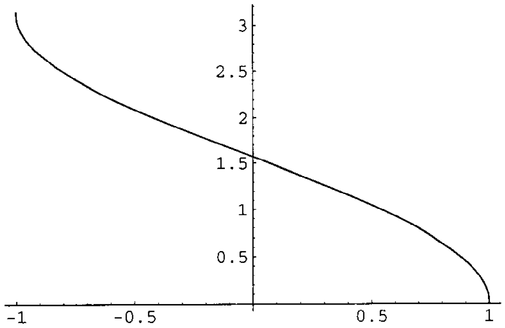

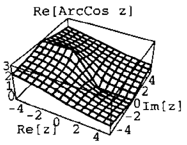

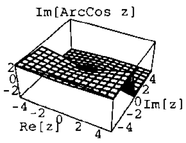

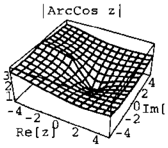

The function $\cos^{-1} x$, also denoted $\arccos(x)$, where $\cos x$ is the COSINE and the superscript -1 denotes an INVERSE FUNCTION, *not* the multiplicative inverse. The MACLAURIN SERIES for the inverse cosine range $-1 < x < 1$ is

$$\cos^{-1} x = \tfrac{1}{2}\pi - x - \tfrac{1}{6}x^3 - \tfrac{3}{40}x^5 - \tfrac{5x}{112}x^7 - \tfrac{35}{1152}x^9 - \ldots. \tag{1}$$

The inverse cosine satisfies

$$\cos^{-1} x = \pi - \cos^{-1}(-x) \tag{2}$$

for POSITIVE and NEGATIVE x, and

$$\cos^{-1} = \tfrac{1}{2}\pi - \cos^{-1}(\sqrt{1-x^2}) \tag{3}$$

for $x \geq 0$. The inverse cosine is given in terms of other inverse trigonometric functions by

$$\cos^{-1} x = \cot^{-1}\left(\frac{x}{\sqrt{1-x^2}}\right) \tag{4}$$

$$= \tfrac{1}{2}\pi + \sin^{-1}(-x) = \tfrac{1}{2}\pi - \sin^{-1} x \tag{5}$$

$$= \tfrac{1}{2}\pi - \tan^{-1}\left(\frac{x}{\sqrt{1-x^2}}\right) \tag{6}$$

for POSITIVE or NEGATIVE x, and

$$\cos^{-1} x = \csc^{-1}\left(\frac{1}{\sqrt{1-x^2}}\right) \tag{7}$$

$$= \sec^{-1}\left(\frac{1}{x}\right) \tag{8}$$

$$= \sin^{-1}(\sqrt{1-x^2}) \tag{9}$$

$$= \tan^{-1}\left(\frac{\sqrt{1-x^2}}{x}\right) \tag{10}$$

for $x \geq 0$.

see also COSINE, INVERSE SECANT

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Inverse Circular Functions." §4.4 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 79–83, 1972.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 142–143, 1987.

Inverse Cotangent

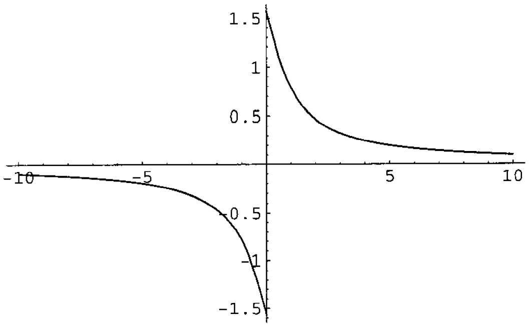

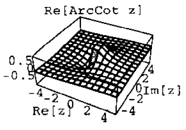

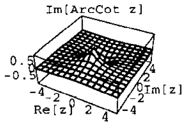

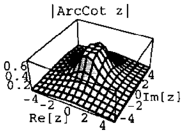

The function $\cot^{-1} x$, also denoted $\operatorname{arccot}(x)$, where $\cot x$ is the COTANGENT and the superscript -1 denotes an INVERSE FUNCTION and *not* the multiplicative inverse. The MACLAURIN SERIES is given by

$$\cot^{-1} x = \tfrac{1}{2}\pi - x + \tfrac{1}{3}x^3 - \tfrac{1}{5}x^5 + \tfrac{1}{7}x^7 - \tfrac{1}{9}x^9 + \ldots, \tag{1}$$

and POWER SERIES by

$$\cot^{-1} x = x^{-1} - \tfrac{1}{3}x^{-3} + \tfrac{1}{5}x^{-5} - \tfrac{1}{7}x^{-7} + \tfrac{1}{9}x^{-9} + \ldots. \quad (2)$$

Euler derived the INFINITE series

$$\cot^{-1} x = x\left[\frac{1}{x^2+1} + \frac{2}{3(x^2+1)^2} + \frac{2\cdot 4}{3\cdot 5(x^2+1)^3} + \ldots\right] \quad (3)$$

(Wetherfield 1996).

The inverse cotangent satisfies

$$\cot^{-1} x = \pi - \cot^{-1}(-x) \quad (4)$$

for POSITIVE and NEGATIVE x, and

$$\cot^{-1} = \tfrac{1}{2}\pi - \cot^{-1}\left(\frac{1}{x}\right) \quad (5)$$

for $x \geq 0$. The inverse cotangent is given in terms of other inverse trigonometric functions by

$$\begin{aligned}\cot^{-1} x &= \cos^{-1}\left(\frac{x}{\sqrt{x^2+1}}\right) && (6)\\ &= \tfrac{1}{2}\pi - \sin^{-1}\left(\frac{x}{\sqrt{x^2+1}}\right) && (7)\\ &= \tfrac{1}{2}\pi + \tan^{-1}(-x) = \tfrac{1}{2}\pi - \tan^{-1} x && (8)\end{aligned}$$

for POSITIVE or NEGATIVE x, and

$$\begin{aligned}\cot^{-1} x &= \csc^{-1}(\sqrt{x^2+1}) && (9)\\ &= \sec^{-1}\left(\frac{\sqrt{x^2+1}}{x}\right) && (10)\\ &= \sin^{-1}\left(\frac{1}{\sqrt{x^2+1}}\right) && (11)\\ &= \tan^{-1}\left(\frac{1}{x}\right) && (12)\end{aligned}$$

for $x \geq 0$.

A number

$$t_x = \cot^{-1} x, \quad (13)$$

where x is an INTEGER or RATIONAL NUMBER, is sometimes called a GREGORY NUMBER. Lehmer (1938a) showed that $\cot^{-1}(a/b)$ can be expressed as a finite sum of inverse cotangents of INTEGER arguments

$$\cot^{-1}\left(\frac{a}{b}\right) = \sum_{i=1}^{k}(-1)^{i-1}\cot^{-1} n_i, \quad (14)$$

where

$$n_i = \left\lfloor \frac{a_i}{b_i} \right\rfloor, \quad (15)$$

with $\lfloor x \rfloor$ the FLOOR FUNCTION, and

$$\begin{aligned}a_{i+1} &= a_i n + i + b_i && (16)\\ b_{i+1} &= a_i - n_i b_i, && (17)\end{aligned}$$

with $a_0 = a$ and $b_0 = b$, and where the recurrence is continued until $b_{k+1} = 0$. If an INVERSE TANGENT sum is written as

$$\tan^{-1} n = \sum_{k=1} f_k \tan^{-1} n_k + f \tan^{-1}, \quad (18)$$

then equation (14) becomes

$$\cot^{-1} n = \sum_{k=1} f_k \cot^{-1} n_k + c\cot^{-1} 1, \quad (19)$$

where

$$c = 2 - f - 2\sum_{k=1} f_r. \quad (20)$$

Inverse cotangent sums can be used to generate MACHIN-LIKE FORMULAS.

An interesting inverse cotangent identity attributed to Charles Dodgson (Lewis Carroll) by Lehmer (1938b; Bromwich 1965, Castellanos 1988ab) is

$$\cot^{-1}(p+r) + \tan^{-1}(p+q) = \tan^{-1} p, \quad (21)$$

where

$$1 + p^2 = qr. \quad (22)$$

Other inverse cotangent identities include

$$2\cot^{-1}(2x) - \cot^{-1} x = \cot^{-1}(4x^3 + 3x) \quad (23)$$

$$3\cot^{-1}(3x) - \cot^{-1} x = \cot^{-1}\left(\frac{27x^4 + 18x^2 - 1}{8x}\right), \quad (24)$$

as well as many others (Bennett 1926, Lehmer 1938b).

see also COTANGENT, INVERSE TANGENT, MACHIN'S FORMULA, MACHIN-LIKE FORMULAS, TANGENT

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Inverse Circular Functions." §4.4 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 79–83, 1972.

Bennett, A. A. "The Four Term Diophantine Arccotangent Relation." *Ann. Math.* **27**, 21–24, 1926.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 142–143, 1987.

Bromwich, T. J. I. and MacRobert, T. M. *An Introduction to the Theory of Infinite Series, 3rd ed.* New York: Chelsea, 1991.

Castellanos, D. "The Ubiquitous Pi. Part I." *Math. Mag.* **61**, 67–98, 1988a.

Castellanos, D. "The Ubiquitous Pi. Part II." *Math. Mag.* **61**, 148–163, 1988b.

Lehmer, D. H. "A Cotangent Analogue of Continued Fractions." *Duke Math. J.* **4**, 323–340, 1938a.

Lehmer, D. H. "On Arccotangent Relations for π." *Amer. Math. Monthly* **45**, 657–664, 1938b.

Weisstein, E. W. "Arccotangent Series." `http://www.astro.virginia.edu/~eww6n/math/notebooks/CotSeries.m`.

Wetherfield, M. "The Enhancement of Machin's Formula by Todd's Process." *Math. Gaz.*, 333–344, July 1996.

Inverse Curve

Given a Circle C with Center O and Radius k, then two points P and Q are inverse with respect to C if $OP \cdot OQ = k^2$. If P describes a curve C_1, then Q describes a curve C_2 called the inverse of C_1 with respect to the circle C (with Inversion Center O). If the Polar equation of C is $r(\theta)$, then the inverse curve has polar equation

$$r = \frac{k^2}{r(\theta)}.$$

If $O = (x_0, y_0)$ and $P = (f(t), g(t))$, then the inverse has equations

$$x = x_0 + \frac{k^2(f - x_0])}{(f - x_0)^2 + (g - y_0)^2}$$

$$y = y_0 + \frac{k^2(g - y_0)}{(f - x_0)^2 + (g - y_0)^2}.$$

Curve	Inversion Center	Inverse Curve
Archimedean spiral	origin	Archimedean spiral
cardioid	cusp	parabola
circle	any pt.	another circle
cissoid of Diocles	cusp	parabola
cochleoid	origin	quadratrix of Hippias
epispiral	origin	Rose
Fermat's spiral	origin	lituus
hyperbola	center	lemniscate
hyperbola	vertex	right strophoid
hyperbola with $a = \sqrt{3}$	vertex	Maclaurin trisectrix
lemniscate	center	hyperbola
lituus	origin	Fermat spiral
logarithmic spiral	origin	logarithmic spiral
Maclaurin trisectrix	focus	Tschirnhausen's cubic
parabola	focus	cardioid
parabola	vertex	cissoid of Diocles
quadratrix of Hippias		cochleoid
right strophoid	origin	the same right strophoid
sinusoidal spiral	origin	sinusoidal spiral inverse curve
Tschirnhausen cubic		sinusoidal spiral

see also Inversion, Inversion Center, Inversion Circle

References

Lee, X. "Inversion." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Inversion_dir/inversion.html`.

Lee, X. "Inversion Gallery." `http://www.best.com/~xah/SpecialPlaneCurves_dir/InversionGallery_dir/inversionGallery.html`.

Yates, R. C. "Inversion." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 127–134, 1952.

Inverse Filter

A linear Deconvolution Algorithm.

Inverse Function

Given a Function $f(x)$, its inverse $f^{-1}(x)$ is defined by $f(f^{-1}(x)) \equiv x$. Therefore, $f(x)$ and $f^{-1}(x)$ are reflections about the line $y = x$.

Inverse Hyperbolic Cosecant

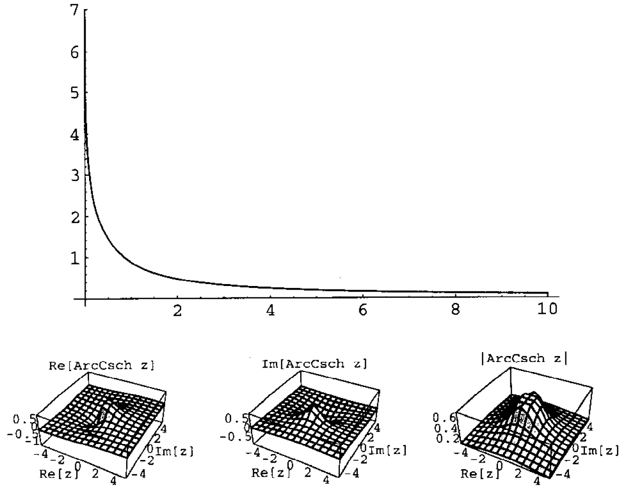

The Inverse Function of the Hyperbolic Cosecant, denoted $\operatorname{csch}^{-1} x$.

see also Hyperbolic Cosecant

Inverse Hyperbolic Cosine

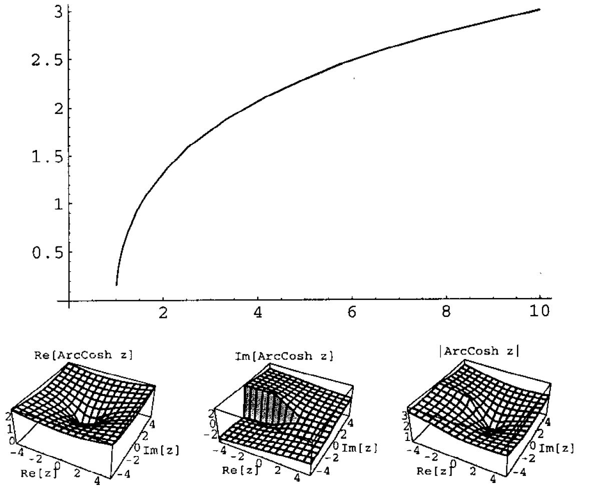

The Inverse Function of the Hyperbolic Cosine, denoted $\cosh^{-1} x$.

see also Hyperbolic Cosine

Inverse Hyperbolic Cotangent

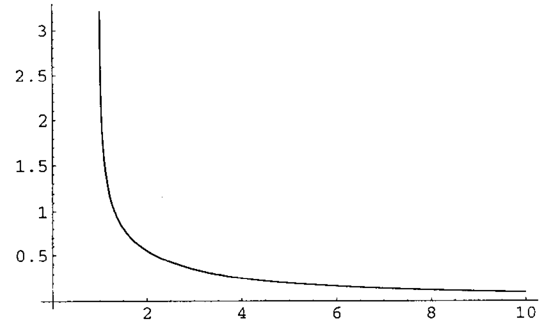

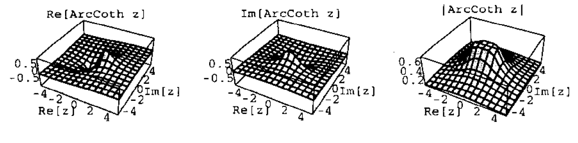

The INVERSE FUNCTION of the HYPERBOLIC COTANGENT, denoted $\coth^{-1} x$.

see also HYPERBOLIC COTANGENT

Inverse Hyperbolic Functions

The INVERSE of the HYPERBOLIC FUNCTIONS, denoted $\cosh^{-1} x$, $\coth^{-1} x$, $\operatorname{csch}^{-1} x$, $\operatorname{sech}^{-1} x$, $\sinh^{-1} x$, and $\tanh^{-1} x$.

see also HYPERBOLIC FUNCTIONS

References

Spanier, J. and Oldham, K. B. "The Inverse Hyperbolic Functions." Ch. 31 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 285–293, 1987.

Inverse Hyperbolic Secant

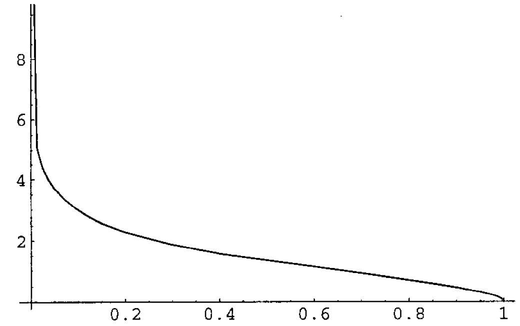

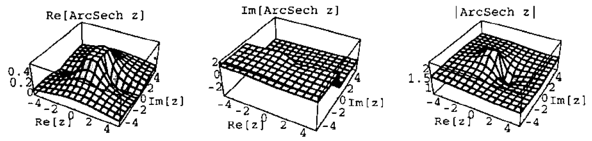

The INVERSE FUNCTION of the HYPERBOLIC SECANT, denoted $\operatorname{sech}^{-1} x$.

see also HYPERBOLIC SECANT

Inverse Hyperbolic Sine

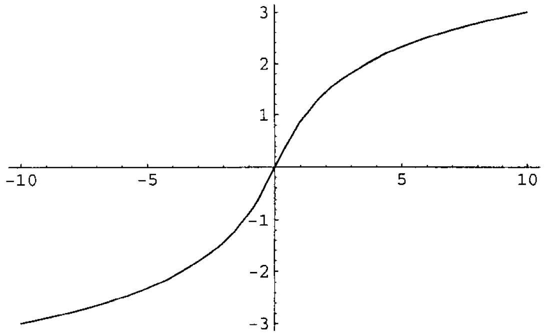

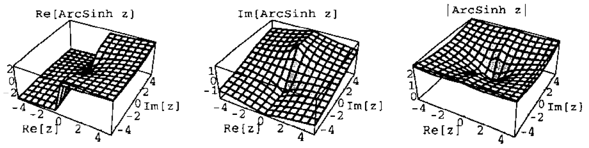

The INVERSE FUNCTION of the HYPERBOLIC SINE, denoted $\sinh^{-1} x$.

see also HYPERBOLIC SINE

Inverse Hyperbolic Tangent

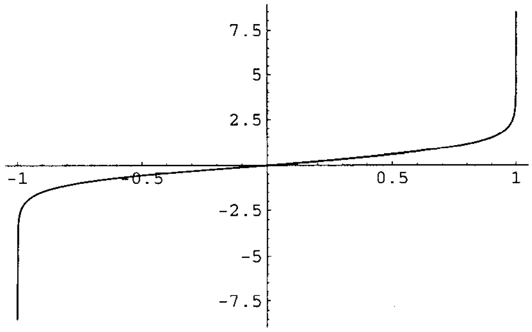

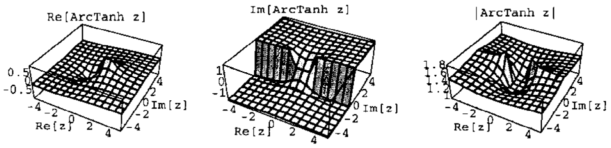

The INVERSE FUNCTION of the HYPERBOLIC TANGENT, denoted $\tanh^{-1} x$.

see also HYPERBOLIC TANGENT

Inverse Matrix

see also MATRIX INVERSE

Inverse Points

Points which are transformed into each other through INVERSION about a given INVERSION CIRCLE. The point P' which is the inverse point of a given point P with respect to an INVERSION CIRCLE C may be constructed geometrically using a COMPASS only (Courant and Robbins 1996).

see also GEOMETRIC CONSTRUCTION, INVERSION, POLAR, POLE (GEOMETRY)

References

Courant, R. and Robbins, H. "Geometrical Construction of Inverse Points." §3.4.3 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 144–145, 1996.

Inverse Quadratic Interpolation

The use of three prior points in a ROOT-finding ALGORITHM to estimate the zero crossing.

Inverse Scattering Method

A method which can be used to solve the initial value problem for certain classes of nonlinear PARTIAL DIFFERENTIAL EQUATIONS. The method reduces the initial value problem to a linear INTEGRAL EQUATION in which time appears only implicitly. However, the solutions $u(x,t)$ and various of their derivatives must approach zero as $x \to \pm\infty$ (Infeld and Rowlands 1990).

see also ABLOWITZ-RAMANI-SEGUR CONJECTURE, BÄCKLUND TRANSFORMATION

References

Infeld, E. and Rowlands, G. "Inverse Scattering Method." §7.4 in *Nonlinear Waves, Solitons, and Chaos.* Cambridge, England: Cambridge University Press, pp. 192–196, 1990.

Miura, R. M. (Ed.) *Bäcklund Transformations, the Inverse Scattering Method, Solitons, and Their Applications.* New York: Springer-Verlag, 1974.

Inverse Secant

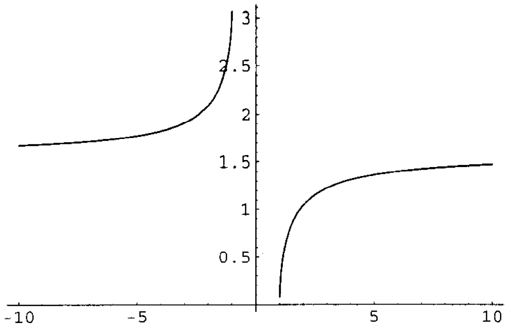

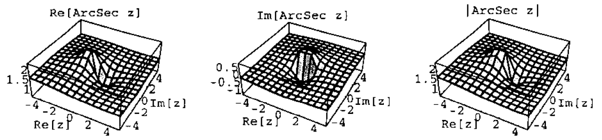

The function $\sec^{-1} x$, where $\sec x$ is the SECANT and the superscript -1 denotes the INVERSE FUNCTION, *not* the multiplicative inverse. The inverse secant satisfies

$$\sec^{-1} x = \csc^{-1}\left(\frac{x}{\sqrt{x^2-1}}\right) \tag{1}$$

for POSITIVE or NEGATIVE x, and

$$\sec^{-1} x = \pi + \sec^{-1}(-x) \tag{2}$$

for $x \geq 0$. The inverse secant is given in terms of other inverse trigonometric functions by

$$\begin{aligned}
\sec^{-1} x &= \cos^{-1}\left(\frac{1}{x}\right) && (3)\\
&= \cot^{-1}\left(\frac{1}{\sqrt{x^2-1}}\right) && (4)\\
&= \tfrac{1}{2}\pi - \csc^{-1} x = -\tfrac{1}{2}\pi - \csc^{-1}(-x) && (5)\\
&= \sin^{-1}\left(\frac{\sqrt{x^2-1}}{x}\right) && (6)\\
&= \tan^{-1}(\sqrt{x^2-1}) && (7)
\end{aligned}$$

for $x \geq 0$.

see also INVERSE COSECANT, SECANT

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 141–143, 1987.

Inverse Semigroup

The abstract counterpart of a PSEUDOGROUP formed by certain subsets of a GROUPOID which admit a MULTIPLICATION.

References

Weinstein, A. "Groupoids: Unifying Internal and External Symmetry." *Not. Amer. Math. Soc.* **43**, 744–752, 1996.

Inverse Sine

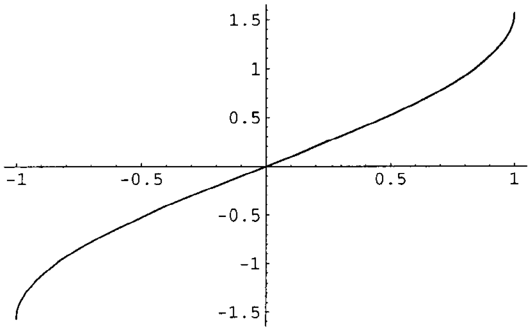

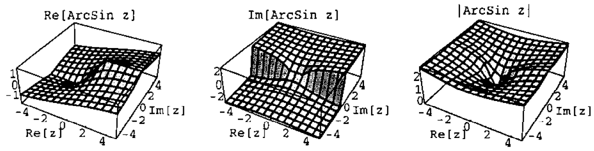

The function $\sin^{-1} x$, where $\sin x$ is the SINE and the superscript -1 denotes the INVERSE FUNCTION, *not* the multiplicative inverse. The inverse sine satisfies

$$\sin^{-1} x = -\sin^{-1}(-x) \tag{1}$$

for POSITIVE and NEGATIVE x, and

$$\sin^{-1} = \tfrac{1}{2}\pi - \sin^{-1}(\sqrt{1-x^2}) \tag{2}$$

for $x \geq 0$. The inverse sine is given in terms of other inverse trigonometric functions by

$$\begin{aligned}\sin^{-1} x &= \cos^{-1}(-x) - \tfrac{1}{2}\pi = \tfrac{1}{2}\pi - \cos^{-1} x &(3)\\ &= \tfrac{1}{2}\pi - \cot^{-1}\left(\frac{x}{\sqrt{1-x^2}}\right) &(4)\\ &= \tan^{-1}\left(\frac{x}{\sqrt{1-x^2}}\right) &(5)\end{aligned}$$

for POSITIVE or NEGATIVE x, and

$$\begin{aligned}\sin^{-1} x &= \cos^{-1}(\sqrt{1-x^2}\,) &(6)\\ &= \cot^{-1}\left(\frac{\sqrt{1-x^2}}{x}\right) &(7)\\ &= \csc^{-1}\left(\frac{1}{x}\right) &(8)\\ &= \sec^{-1}\left(\frac{1}{\sqrt{1-x^2}}\right) &(9)\end{aligned}$$

for $x \geq 0$.

see also INVERSE COSINE, SINE

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Inverse Circular Functions." §4.4 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 79–83, 1972.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 142–143, 1987.

Inverse Tangent

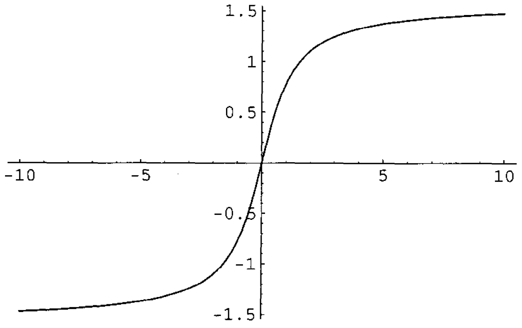

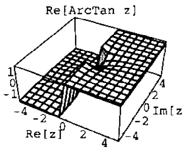

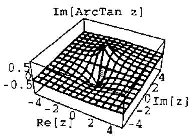

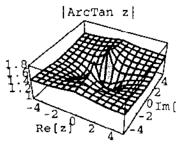

The inverse tangent is also called the arctangent and is denoted either $\tan^{-1} x$ or $\arctan x$. It has the MACLAURIN SERIES

$$\tan^{-1} x = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{2n+1} = x - \tfrac{1}{3}x^3 + \tfrac{1}{5}x^5 - \tfrac{1}{7}x^7 + \ldots. \qquad (1)$$

A more rapidly converging form due to Euler is given by

$$\tan^{-1} x = \sum_{n=0}^{\infty} \frac{2^{2n}(n!)^2}{(2n+1)!} \frac{x^{2n+1}}{(+x^2)^{n+1}} \qquad (2)$$

(Castellanos 1988). The inverse tangent satisfies

$$\tan^{-1} x = -\tan^{-1}(-x) \qquad (3)$$

for POSITIVE and NEGATIVE x, and

$$\tan^{-1} = \tfrac{1}{2}\pi - \tan^{-1}\left(\frac{1}{x}\right) \qquad (4)$$

for $x \geq 0$. The inverse tangent is given in terms of other inverse trigonometric functions by

$$\begin{aligned}\tan^{-1} x &= \tfrac{1}{2}\pi - \cos^{-1}\left(\frac{x}{\sqrt{x^2+1}}\right) &(5)\\ &= \cot^{-1}(-x) - \tfrac{1}{2}\pi = \tfrac{1}{2}\pi - \cot^{-1} x &(6)\\ &= \sin^{-1}\left(\frac{x}{\sqrt{x^2+1}}\right) &(7)\end{aligned}$$

for POSITIVE or NEGATIVE x, and

$$\begin{aligned}\tan^{-1} x &= \cos^{-1}\left(\frac{1}{\sqrt{x^2+1}}\right) &(8)\\ &= \cot^{-1}\left(\frac{1}{x}\right) &(9)\\ &= \csc^{-1}\left(\frac{\sqrt{x^2+1}}{x}\right) &(10)\\ &= \sec^{-1}(\sqrt{x^2+1}\,) &(11)\end{aligned}$$

for $x \geq 0$.

In terms of the HYPERGEOMETRIC FUNCTION,

$$\begin{aligned}\tan^{-1} x &= x\,{}_2F_1(1, \tfrac{1}{2}; \tfrac{3}{2}; -x^2) &(12)\\ &= \frac{x}{1+x^2}\,{}_2F_1\left(1, 1; \tfrac{3}{2}; \frac{x^2}{1+x^2}\right) &(13)\end{aligned}$$

(Castellanos 1988). Castellanos (1986, 1988) also gives some curious formulas in terms of the FIBONACCI NUMBERS,

$$\begin{aligned}\tan^{-1} x &= \sum_{n=0}^{\infty} \frac{(-1)^n F_{2n+1} t^{2n+1}}{5^n(2n+1)} &(14)\\ &= 5\sum_{n=0}^{\infty} \frac{(-1)^n {F_{2n+1}}^2}{(2n+1)(u+\sqrt{u^2+1}\,)^{2n+1}} &(15)\\ &= \sum_{n=0}^{\infty} \frac{(-1)^n 5^{n+2} {F_{2n+1}}^3}{(2n+1)(v+\sqrt{v^2+5}\,)^{2n+1}}, &(16)\end{aligned}$$

where

$$t \equiv \frac{2x}{1+\sqrt{\frac{4x^2}{5}}} \tag{17}$$

$$u \equiv \frac{5}{4x}\left(1+\sqrt{1+\tfrac{24}{25}x^2}\right), \tag{18}$$

and v is the largest POSITIVE ROOT of

$$8xv^4 - 100v^3 - 450xv^2 + 875v + 625x = 0. \tag{19}$$

The inverse tangent satisfies the addition FORMULA

$$\tan^{-1}x + \tan^{-1}y = \tan^{-1}\left(\frac{x+y}{1-xy}\right) \tag{20}$$

as well as the more complicated FORMULAS

$$\tan^{-1}\left(\frac{1}{a-b}\right) = \tan^{-1}\left(\frac{1}{a}\right) + \tan^{-1}\left(\frac{b}{a^2-ab+1}\right) \tag{21}$$

$$\tan^{-1}\left(\frac{1}{a}\right) = 2\tan^{-1}\left(\frac{1}{2a}\right) - \tan^{-1}\left(\frac{1}{4a^3+3a}\right) \tag{22}$$

$$\tan^{-1}\left(\frac{1}{p}\right) = \tan^{-1}\left(\frac{1}{p+q}\right) + \tan^{-1}\left(\frac{q}{p^2+pq+1}\right), \tag{23}$$

the latter of which was known to Euler. The inverse tangent FORMULAS are connected with many interesting approximations to PI

$$\tan^{-1}(1+x) = \tfrac{1}{4}\pi + \tfrac{1}{2}x - \tfrac{1}{4}x^2 + \tfrac{1}{12}x^3 + \tfrac{1}{40}x^5 + \tfrac{1}{48}x^6 + \tfrac{1}{112}x^7 + \ldots. \tag{24}$$

Euler gave

$$\tan^{-1}x = \frac{y}{x}\left(\frac{2}{3}y + \frac{2\cdot 4}{3\cdot 5}y^2 + \frac{2\cdot 4\cdot 6}{3\cdot 5\cdot 7}y^3 + \ldots\right), \tag{25}$$

where

$$y \equiv \frac{x^2}{1+x^2}. \tag{26}$$

The inverse tangent has CONTINUED FRACTION representations

$$\tan^{-1}x = \cfrac{x}{1+\cfrac{x^2}{3+\cfrac{4x^2}{5+\cfrac{9x^2}{7+\cfrac{16x^2}{9+\ldots}}}}} \tag{27}$$

$$= \cfrac{x}{1+\cfrac{x^2}{3-x^2+\cfrac{9x^2}{5-3x^2+\cfrac{25x^2}{7-5x^2+\ldots}}}}. \tag{28}$$

To find $\tan^{-1}x$ numerically, the following ARITHMETIC-GEOMETRIC MEAN-like ALGORITHM can be used. Let

$$a_0 = (1+x^2)^{-1/2} \tag{29}$$

$$b_0 = 1. \tag{30}$$

Then compute

$$a_{i+1} = \tfrac{1}{2}(a_i + b_i) \tag{31}$$

$$b_{i+1} = \sqrt{a_{i+1}b_i}, \tag{32}$$

and the inverse tangent is given by

$$\tan^{-1}x = \lim_{n\to\infty}\frac{x}{\sqrt{1+x^2}\,a_n} \tag{33}$$

(Acton 1990).

An inverse tangent $\tan^{-1}n$ with integral n is called reducible if it is expressible as a finite sum of the form

$$\tan^{-1}n = \sum_{k=1} f_k \tan^{-1}n_k, \tag{34}$$

where f_k are POSITIVE or NEGATIVE INTEGERS and n_i are iINTEGERS $< n$. $\tan^{-1}m$ is reducible IFF all the PRIME factors of $1+m^2$ occur among the PRIME factors of $1+n^2$ for $n = 1, \ldots, m-1$. A second NECESSARY and SUFFICIENT condition is that the largest PRIME factor of $1+m^2$ is less than $2m$. Equivalent to the second condition is the statement that every GREGORY NUMBER $t_x = \cot^{-1}x$ can be uniquely expressed as a sum in terms of t_ms for which m is a STØRMER NUMBER (Conway and Guy 1996). To find this decomposition, write

$$\arg(1+in) = \arg\prod_{k=1}(1+n_k i)^{f_k}, \tag{35}$$

so the ratio

$$r = \frac{\prod_{k=1}(1+n_k i)^{f_k}}{1+in} \tag{36}$$

is a RATIONAL NUMBER. Equation (36) can also be written

$$r^2(1+n^2) = \prod_{k=1}(1+{n_k}^2)^{f_k}. \tag{37}$$

Writing (34) in the form

$$\tan^{-1}n = \sum_{k=1} f_k \tan^{-1}n_k + f\tan^{-1}1 \tag{38}$$

allows a direct conversion to a corresponding INVERSE COTANGENT FORMULA

$$\cot^{-1}n = \sum_{k=1} f_k \cot^{-1}n_k + c\cot^{-1}1, \tag{39}$$

where

$$c = 2 - f - 2\sum_{k=1} f_r. \tag{40}$$

Todd (1949) gives a table of decompositions of $\tan^{-1} n$ for $n \leq 342$. Conway and Guy (1996) give a similar table in terms of STØRMER NUMBERS.

Arndt and Gosper give the remarkable inverse tangent identity

$$\sin\left(\sum_{k=1}^{2n+1} \tan^{-1} a_k\right) = \frac{(-1)^n}{2n+1} \frac{\sum_{k=1}^{2n+1} \prod_{j=1}^{2n+1} \left[a_j - \tan\left(\frac{\pi(j-k)}{2n+1}\right)\right]}{\sqrt{\prod_{j=1}^{2n+1} (a_j{}^2 + 1)}}. \quad (41)$$

see also INVERSE COTANGENT, TANGENT

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Inverse Circular Functions." §4.4 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 79–83, 1972.

Acton, F. S. "The Arctangent." In *Numerical Methods that Work, upd. and rev.* Washington, DC: Math. Assoc. Amer., pp. 6–10, 1990.

Arndt, J. "Completely Useless Formulas." `http://www.jjj.de/hfloat/hfloatpage.html#formulas`.

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Item 137, Feb. 1972.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 142–143, 1987.

Castellanos, D. "Rapidly Converging Expansions with Fibonacci Coefficients." *Fib. Quart.* **24**, 70–82, 1986.

Castellanos, D. "The Ubiquitous Pi. Part I." *Math. Mag.* **61**, 67–98, 1988.

Conway, J. H. and Guy, R. K. "Størmer's Numbers." *The Book of Numbers.* New York: Springer-Verlag, pp. 245–248, 1996.

Todd, J. "A Problem on Arc Tangent Relations." *Amer. Math. Monthly* **56**, 517–528, 1949.

Inverse Trigonometric Functions

INVERSE FUNCTIONS of the TRIGONOMETRIC FUNCTIONS written $\cos^{-1} x$, $\cot^{-1} x$, $\csc^{-1} x$, $\sec^{-1} x$, $\sin^{-1} x$, and $\tan^{-1} x$.

see also INVERSE COSECANT, INVERSE COSINE, INVERSE COTANGENT, INVERSE SECANT, INVERSE SINE, INVERSE TANGENT

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Inverse Circular Functions." §4.4 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 79–83, 1972.

Spanier, J. and Oldham, K. B. "Inverse Trigonometric Functions." Ch. 35 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 331–341, 1987.

Inversely Similar

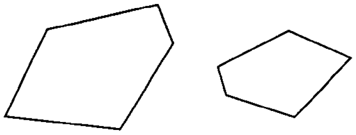

inversely similar

Two figures are said to be SIMILAR when all corresponding ANGLES are equal, and are inversely similar when all corresponding ANGLES are equal and described in the opposite rotational sense.

see also DIRECTLY SIMILAR, SIMILAR

Inversion

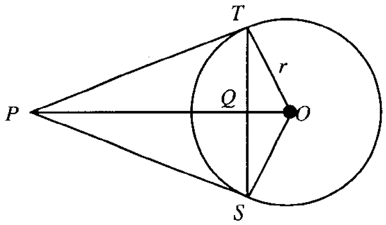

Inversion is the process of transforming points to their INVERSE POINTS. This sort of inversion was first systematically investigated by Jakob Steiner. Two points are said to be inverses with respect to an INVERSION CIRCLE with INVERSION CENTER $O = (x_0, y_0)$ and INVERSION RADIUS k if PT and PS are line segments symmetric about OP and tangent to the CIRCLE, and Q is the intersection of OP and ST. The curve to which a given curve is transformed under inversion is called its INVERSE CURVE.

Note that a point on the CIRCUMFERENCE of the INVERSION CIRCLE is its own inverse point. The inverse points obey

$$\frac{OP}{k} = \frac{k}{OQ}, \quad (1)$$

or

$$k^2 = OP \times OQ, \quad (2)$$

where k^2 is called the POWER. The equation for the inverse of the point (x, y) relative to the INVERSION CIRCLE with INVERSION CENTER (x_0, y_0) and inversion radius k is therefore

$$x' = x_0 + \frac{k^2(x - x_0)}{(x - x_0)^2 + (y - y_0)^2} \quad (3)$$

$$y' = y_0 + \frac{k^2(y - y_0)}{(x - x_0)^2 + (y - y_0)^2}. \quad (4)$$

In vector form,

$$\mathbf{x}' = \mathbf{x}_0 + \frac{k^2(\mathbf{x} - \mathbf{x}_0)}{|\mathbf{x} - \mathbf{x}_0|^2}. \quad (5)$$

Any ANGLE inverts to an opposite ANGLE.

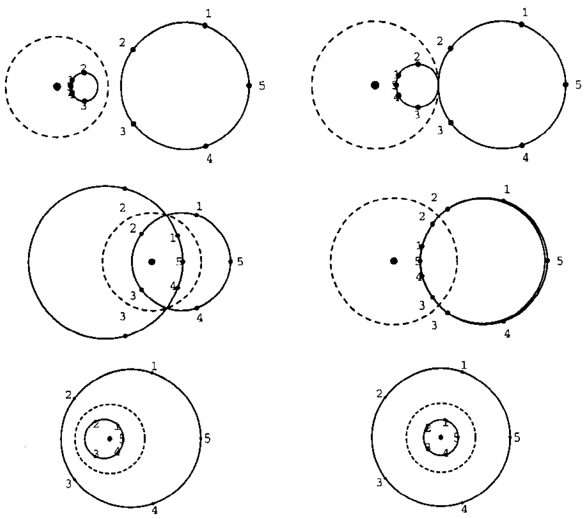

Treating LINES as CIRCLES of INFINITE RADIUS, all CIRCLES invert to CIRCLES. Furthermore, any two nonintersecting circles can be inverted into concentric circles by taking the INVERSION CENTER at one of the two limiting points (Coxeter 1969), and ORTHOGONAL CIRCLES invert to ORTHOGONAL CIRCLES (Coxeter 1969).

The inverse of a CIRCLE of RADIUS a with CENTER (x, y) with respect to an inversion circle with INVERSION CENTER $(0, 0)$ and INVERSION RADIUS k is another CIRCLE with CENTER $(x', y') = (sx, sy)$ and RADIUS $r' = |s|a$, where

$$s \equiv \frac{k^2}{x^2 + y^2 - a^2}. \tag{6}$$

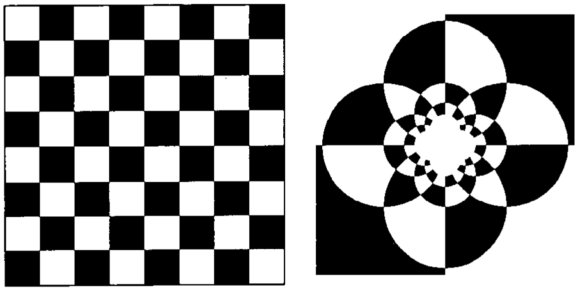

The above plot shows a checkerboard centered at (0, 0) and its inverse about a small circle also centered at (0, 0) (Dixon 1991).

see also ARBELOS, HEXLET, INVERSE CURVE, INVERSION CIRCLE, INVERSION OPERATION, INVERSION RADIUS, INVERSIVE DISTANCE, INVERSIVE GEOMETRY, MIDCIRCLE, PAPPUS CHAIN, PEAUCELLIER INVERSOR, POLAR, POLE (GEOMETRY), POWER (CIRCLE), RADICAL LINE, STEINER CHAIN, STEINER'S PORISM

References

Courant, R. and Robbins, H. "Geometrical Transformations. Inversion." §3.4 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 140–146, 1996.

Coxeter, H. S. M. "Inversion in a Circle" and "Inversion of Lines and Circles." §6.1 and 6.3 in *Introduction to Geometry, 2nd ed.* New York: Wiley, p. 77–83, 1969.

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 108–114, 1967.

Dixon, R. "Inverse Points and Mid-Circles." §1.6 in *Mathographics.* New York: Dover, pp. 62–73, 1991.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 43–57, 1929.

Lockwood, E. H. "Inversion." Ch. 23 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 176–181, 1967.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 25–31, 1990.

Weisstein, E. W. "Plane Geometry." `http://www.astro.virginia.edu/~eww6n/math/notebooks/PlaneGeometry.m`.

Inversion Center

The point that INVERSION OF A CURVE is performed with respect to.

see also INVERSE POINTS, INVERSION CIRCLE, INVERSION RADIUS, INVERSIVE DISTANCE, POLAR, POLE (GEOMETRY), POWER (CIRCLE)

Inversion Circle

The CIRCLE with respect to which a INVERSE CURVE is computed or relative to which INVERSE POINTS are computed.

see also INVERSE POINTS, INVERSION CENTER, INVERSION RADIUS, INVERSIVE DISTANCE, MIDCIRCLE, POLAR, POLE (GEOMETRY), POWER (CIRCLE)

Inversion Operation

The SYMMETRY OPERATION $(x, y, z) \to (-x, -y, -z)$. When used in conjunction with a ROTATION, it becomes an IMPROPER ROTATION.

Inversion Radius

The RADIUS used in performing an INVERSION with respect to an INVERSION CIRCLE.

see also INVERSE POINTS, INVERSION CENTER, INVERSION CIRCLE, INVERSIVE DISTANCE, POLAR, POLE (GEOMETRY), POWER (CIRCLE)

Inversive Distance

The inversive distance is the NATURAL LOGARITHM of the ratio of two concentric circles into which the given circles can be inverted. Let c be the distance between the centers of two nonintersecting CIRCLES of RADII a and $b < a$. Then the inversive distance is

$$\delta = \cosh^{-1}\left|\frac{a^2 + b^2 - c^2}{2ab}\right|$$

(Coxeter and Greitzer 1967).

The inversive distance between the SODDY CIRCLES is given by

$$\delta = 2\cosh^{-1} 2,$$

and the CIRCUMCIRCLE and INCIRCLE of a TRIANGLE with CIRCUMRADIUS R and INRADIUS r are at inversive distance

$$\delta = 2\sinh^{-1}\left(\frac{1}{2}\sqrt{\frac{r}{R}}\right)$$

(Coxeter and Greitzer 1967, pp. 130–131).

References
Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 123–124 and 127–131, 1967.

Inversive Geometry

The GEOMETRY resulting from the application of the INVERSION operation. It can be especially powerful for solving apparently difficult problems such as STEINER'S PORISM and APOLLONIUS' PROBLEM.

see also HEXLET, INVERSE CURVE, INVERSION, PEAUCELLIER INVERSOR, POLAR, POLE (GEOMETRY), POWER (CIRCLE), RADICAL LINE

References
Ogilvy, C. S. "Inversive Geometry" and "Applications of Inversive Geometry." Chs. 3—4 in *Excursions in Geometry.* New York: Dover, pp. 24–55, 1990.

Inverted Funnel

see also FUNNEL, SINCLAIR'S SOAP FILM PROBLEM

Inverted Snub Dodecadodecahedron

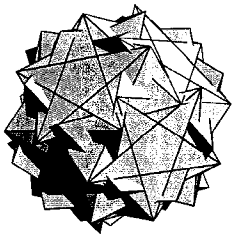

The UNIFORM POLYHEDRON U_{60} whose DUAL POLYHEDRON is the MEDIAL INVERTED PENTAGONAL HEXECONTAHEDRON. It has WYTHOFF SYMBOL $|\,2\,\frac{5}{3}\,5$. Its faces are $12\{\frac{5}{3}\}+60\{3\}+12\{5\}$. It has CIRCUMRADIUS for unit edge length of

$$R \approx 0.8516302.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 180–182, 1989.

Invertible Knot

A knot which can be deformed into itself but with the orientation reversed. The simplest noninvertible knot is 08_{017}. No general technique is known for determining if a KNOT is invertible. Burde and Zieschang (1985) give a tabulation from which it is possible to extract the invertible knots up to 10 crossings.

see also AMPHICHIRAL KNOT

References
Burde, G. and Zieschang, H. *Knots.* Berlin: de Gruyter, 1985.

Involuntary

A LINEAR TRANSFORMATION of period two. Since a LINEAR TRANSFORMATION has the form,

$$\lambda' = \frac{\alpha\lambda+\beta}{\gamma\lambda+\delta}, \tag{1}$$

applying the transformation a second time gives

$$\lambda'' = \frac{\alpha\lambda'+\beta}{\gamma\lambda'+\delta} = \frac{(\alpha^2+\beta\gamma)\lambda+\beta(\alpha+\delta)}{(\alpha+\delta)\gamma\lambda+\beta\gamma+\delta^2}. \tag{2}$$

For an involuntary, $\lambda'' = \lambda$, so

$$\gamma(\alpha+\delta)\lambda^2 + (\delta^2-\alpha^2)\lambda - (\alpha+\delta)\beta = 0. \tag{3}$$

Since each COEFFICIENT must vanish separately,

$$\alpha\gamma + \gamma\delta = 0 \tag{4}$$
$$\delta^2 - \alpha^2 = 0 \tag{5}$$
$$\alpha\beta + \beta\delta = 0. \tag{6}$$

The first equation gives $\delta = \pm\alpha$. Taking $\delta = \alpha$ would require $\gamma = \beta = 0$, giving $\lambda = \lambda'$, the identity transformation. Taking $\delta = -\alpha$ gives $\delta = -\alpha$, so

$$\lambda' = \frac{\alpha\lambda+\beta}{\gamma\lambda-\alpha}, \tag{7}$$

the general form of an INVOLUTION.

see also CROSS-RATIO, INVOLUTION (LINE)

References
Woods, F. S. *Higher Geometry: An Introduction to Advanced Methods in Analytic Geometry.* New York: Dover, pp. 14–15, 1961.

Involute

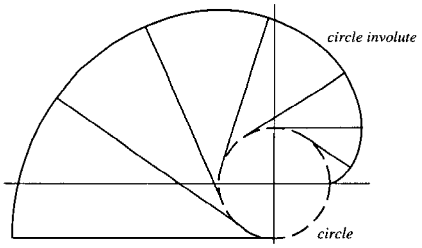

Attach a string to a point on a curve. Extend the string so that it is tangent to the curve at the point of attachment. Then wind the string up, keeping it always taut. The LOCUS of points traced out by the end of the string is the involute of the original curve, and the original curve is called the EVOLUTE of its involute. Although a curve has a unique EVOLUTE, it has infinitely many involutes corresponding to different choices of initial point. An involute can also be thought of as any curve ORTHOGONAL to all the TANGENTS to a given curve.

The equation of the involute is

$$\mathbf{r}_i = \mathbf{r} - s\hat{\mathbf{T}}, \tag{1}$$

where $\hat{\mathbf{T}}$ is the TANGENT VECTOR

$$\hat{\mathbf{T}} = \frac{\frac{d\mathbf{r}}{dt}}{\left|\frac{d\mathbf{r}}{dt}\right|} \tag{2}$$

and s is the ARC LENGTH

$$s = \int ds = \int \frac{ds}{dt}\,dt = \int \frac{\sqrt{ds^2}}{dt}\,dt = \int \sqrt{f'^2 + g'^2}\,dt. \tag{3}$$

This can be written for a parametrically represented function $(f(t), g(t))$ as

$$x(t) = f - \frac{sf'}{\sqrt{f'^2 + g'^2}} \tag{4}$$

$$y(t) = g - \frac{sg'}{\sqrt{f'^2 + g'^2}}. \tag{5}$$

Curve	Involute
astroid	astroid 1/2 as large
cardioid	cardioid 3 times as large
catenary	tractrix
circle catacaustic for a point source	limaçon
circle	circle involute (a spiral)
cycloid	equal cycloid
deltoid	deltoid 1/3 as large
ellipse	ellipse involute
epicycloid	reduced epicycloid
hypocycloid	similar hypocycloid
logarithmic spiral	equal logarithmic spiral
Neile's parabola	parabola
nephroid	Cayley's sextic
nephroid	nephroid 2 times as large

see also EVOLUTE, HUMBERT'S THEOREM

References

Cundy, H. and Rollett, A. "Roulettes and Involutes." §2.6 in *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 46–55, 1989.

Dixon, R. "String Drawings." Ch. 2 in *Mathographics.* New York: Dover, pp. 75–78, 1991.

Gray, A. "Involutes." §5.4 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 81–85, 1993.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 40–42 and 202, 1972.

Lee, X. "Involute." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Involute_dir/involute.html`.

Lockwood, E. H. "Evolutes and Involutes." Ch. 21 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 166–171, 1967.

Pappas, T. "The Involute." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 187, 1989.

Yates, R. C. "Involutes." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 135–137, 1952.

Involution (Group)

An element of order 2 in a GROUP (i.e., an element A of a GROUP such that $A^2 = I$, where I is the IDENTITY ELEMENT).

see also GROUP, IDENTITY ELEMENT

Involution (Line)

Pairs of points of a line, the product of whose distances from a FIXED POINT is a given constant. This is more concisely defined as a PROJECTIVITY of period two.

see also INVOLUNTARY

Involution (Operator)

An OPERATOR of period 2, i.e., an OPERATOR $*$ which satisfies $((a)^*)^* = a$.

Involution (Set)

An involution of a SET S is a PERMUTATION of S which does not contain any cycles of length > 2. The PERMUTATION MATRICES of an involution are SYMMETRIC. The number of involutions $I(n)$ of a SET containing the first n integers is given by the RECURRENCE RELATION

$$I(n) = I(n-1) + (n-1)I(n-2).$$

For $n = 1, 2, \ldots$, the first few values of $I(n)$ are 1, 2, 4, 10, 26, 76, ... (Sloane's A000085). The number of involutions on n symbols cannot be expressed as a fixed number of hypergeometric terms (Petkovšek *et al.* 1996, p. 160).

see also PERMUTATION

References

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, 1996.

Ruskey, F. "Information on Involutions." `http://sue.csc.uvic.ca/~cos/inf/perm/Involutions.html`.

Sloane, N. J. A. Sequence A00085/M1221 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Involution (Transformation)

A TRANSFORMATION of period 2.

Irradiation Illusion

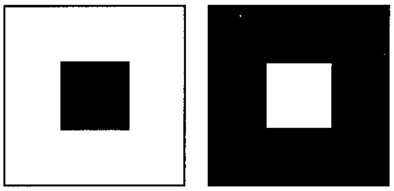

The ILLUSION shown above which was discovered by Helmholtz in the 19th century. Despite the fact that the two above figures are identical in size, the white hole looks bigger than the black one in this ILLUSION.

References

Pappas, T. "Irradiation Optical Illusion." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 199, 1989.

Irrational Number

A number which cannot be expressed as a FRACTION p/q for any INTEGERS p and q. Every TRANSCENDENTAL NUMBER is irrational. Numbers of the form $n^{1/m}$ are irrational unless n is the mth POWER of an INTEGER.

Numbers of the form $\log_n m$, where log is the LOGARITHM, are irrational if m and n are INTEGERS, one of which has a PRIME factor which the other lacks. e^r is irrational for rational $r \neq 0$. The irrationality of e was proven by Lambert in 1761; for the general case, see Hardy and Wright (1979, p. 46). π^n is irrational for POSITIVE integral n. The irrationality of π was proven by Lambert in 1760; for the general case, see Hardy and Wright (1979, p. 47). APÉRY'S CONSTANT $\zeta(3)$ (where $\zeta(z)$ is the RIEMANN ZETA FUNCTION) was proved irrational by Apéry (Apéry 1979, van der Poorten 1979).

From GELFOND'S THEOREM, a number of the form a^b is TRANSCENDENTAL (and therefore irrational) if a is ALGEBRAIC $\neq 0$, 1 and b is irrational and ALGEBRAIC. This establishes the irrationality of e^π (since $(-1)^{-i} = (e^{i\pi})^{-i} = e^\pi$), $2^{\sqrt{2}}$, and $e\pi$. Nesterenko (1996) proved that $\pi + e^\pi$ is irrational. In fact, he proved that π, e^π and $\Gamma(1/4)$ are algebraically independent, but it was not previously known that $\pi + e^\pi$ was irrational.

Given a POLYNOMIAL equation

$$x^m + c_{m-1}x^{m-1} + \ldots + c_0, \qquad (1)$$

where c_i are INTEGERS, the roots x_i are either integral or irrational. If $\cos(2\theta)$ is irrational, then so are $\cos\theta$, $\sin\theta$, and $\tan\theta$.

Irrationality has not yet been established for 2^e, π^e, $\pi^{\sqrt{2}}$, or γ (where γ is the EULER-MASCHERONI CONSTANT).

QUADRATIC SURDS are irrational numbers which have periodic CONTINUED FRACTIONS.

HURWITZ'S IRRATIONAL NUMBER THEOREM gives bounds of the form

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{L_n q^2} \qquad (2)$$

for the best rational approximation possible for an arbitrary irrational number α, where the L_n are called LAGRANGE NUMBERS and get steadily larger for each "bad" set of irrational numbers which is excluded.

The SERIES

$$\sum_{n=1}^\infty \frac{\sigma_k(n)}{n!}, \qquad (3)$$

where $\sigma_k(n)$ is the DIVISOR FUNCTION, is irrational for $k = 1$ and 2. The series

$$\sum_{n=1}^\infty \frac{1}{2^n - 1} = \sum_{n=1}^\infty \frac{d(n)}{2^n}, \qquad (4)$$

where $d(n)$ is the number of divisors of n, is also irrational, as are

$$\sum_{n=1}^\infty \frac{1}{q^n + r} \qquad \text{for} \qquad \sum_{n=1}^\infty \frac{(-1)^n}{q^n + r} \qquad (5)$$

for q an INTEGER other than p, ±1, and r a RATIONAL NUMBER other than 0 or $-q^n$ (Guy 1994).

see also ALGEBRAIC INTEGER, ALGEBRAIC NUMBER, ALMOST INTEGER, DIRICHLET FUNCTION, FERGUSON-FORCADE ALGORITHM, GELFOND'S THEOREM, HURWITZ'S IRRATIONAL NUMBER THEOREM, NEAR NOBLE NUMBER, NOBLE NUMBER, PYTHAGORAS'S THEOREM, QUADRATIC IRRATIONAL NUMBER, RATIONAL NUMBER, SEGRE'S THEOREM, TRANSCENDENTAL NUMBER

References
Apéry, R. "Irrationalité de $\zeta(2)$ et $\zeta(3)$." *Astérisque* **61**, 11–13, 1979.
Courant, R. and Robbins, H. "Incommensurable Segments, Irrational Numbers, and the Concept of Limit." §2.2 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 58-61, 1996.
Guy, R. K. "Some Irrational Series." §B14 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 69, 1994.
Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, 1979.
Manning, H. P. *Irrational Numbers and Their Representation by Sequences and Series.* New York: Wiley, 1906.
Nesterenko, Yu. "Modular Functions and Transcendence Problems." *C. R. Acad. Sci. Paris Sér. I Math.* **322**, 909–914, 1996.
Nesterenko, Yu. V. "Modular Functions and Transcendence Questions." *Mat. Sb.* **187**, 65–96, 1996.
Niven, I. M. *Irrational Numbers.* New York: Wiley, 1956.
Niven, I. M. *Numbers: Rational and Irrational.* New York: Random House, 1961.
Pappas, T. "Irrational Numbers & the Pythagoras Theorem." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 98–99, 1989.
van der Poorten, A. "A Proof that Euler Missed... Apéry's Proof of the Irrationality of $\zeta(3)$." *Math. Intel.* **1**, 196–203, 1979.

Irrationality Measure

see LIOUVILLE-ROTH CONSTANT

Irrationality Sequence

A sequence of POSITIVE INTEGERS $\{a_n\}$ such that $\sum 1/(a_n b_n)$ is IRRATIONAL for all integer sequences $\{b_n\}$. Erdős showed that $\{2^{2^n}\}$ is an irrationality sequence.

References
Guy, R. K. "Irrationality Sequence." §E24 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 225, 1994.

Irreducible Matrix

A SQUARE MATRIX which is not REDUCIBLE is said to be irreducible.

Irreducible Polynomial

A POLYNOMIAL or polynomial equation is said to be irreducible if it cannot be factored into polynomials of lower degree over the same FIELD.

The number of binary irreducible polynomials of degree n is equal to the number of n-bead fixed NECKLACES of two colors: 1, 2, 3, 4, 6, 8, 14, 20, 36, ... (Sloane's A000031), the first few of which are given in the following table.

n	Polynomials
1	x
2	x, $x+1$
3	x, x^2+x+1, $x+1$
4	x, x^3+x+1, x^3+x^2+1, $x+1$

see also FIELD, GALOIS FIELD, NECKLACE, POLYNOMIAL, PRIMITIVE IRREDUCIBLE POLYNOMIAL

References

Sloane, N. J. A. Sequences A000031/M0564 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sloane, N. J. A. and Plouffe, S. Extended entry in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.

Irreducible Representation

An irreducible representation of a GROUP is a representation for which there exists no UNITARY TRANSFORMATION which will transform the representation MATRIX into block diagonal form. The irreducible representation has a number of remarkable properties.

see also GROUP, ITÔ'S THEOREM, UNITARY TRANSFORMATION

Irreducible Semiperfect Number

see PRIMITIVE PSEUDOPERFECT NUMBER

Irreducible Tensor

Given a general second RANK TENSOR A_{ij} and a METRIC g_{ij}, define

$$\theta \equiv A_{ij}g^{ij} = A_i^i \tag{1}$$

$$\omega^i \equiv \epsilon^{ijk}A_{jk} \tag{2}$$

$$\sigma_{ij} \equiv \tfrac{1}{2}(A_{ij}+A_{ji}) - \tfrac{1}{3}g_{ij}A_k^k, \tag{3}$$

where δ_{ij} is the KRONECKER DELTA and ϵ^{ijk} is the LEVI-CIVITA SYMBOL. Then

$$\begin{aligned}\sigma_{ij} &+ \tfrac{1}{3}\theta g_{ij} + \tfrac{1}{2}\epsilon_{ijk}\omega^k \\ &= [\tfrac{1}{2}(A_{ij}+A_{ji}) - \tfrac{1}{3}g_{ij}A_k^k] + \tfrac{1}{3}A_k^k g_{ij} + \tfrac{1}{2}\epsilon_{ijk}[\epsilon^{\lambda\mu k}A_{\lambda\mu}] \\ &= \tfrac{1}{2}(A_{ij}+A_{ji}) + \tfrac{1}{2}(\delta_i^\lambda\delta_j^\mu - \delta_i^\mu\delta_j^\lambda)A_{\lambda\mu} \\ &= \tfrac{1}{2}(A_{ij}+A_{ji}) + \tfrac{1}{2}(A_{ij}-A_{ji}) = A_{ij}, \end{aligned} \tag{4}$$

where θ, ω^i, and σ_{ij} are TENSORS of RANK 0, 1, and 2.

see also TENSOR

Irredundant Ramsey Number

Let G_1, G_2, ..., G_t be a t-EDGE coloring of the COMPLETE GRAPH K_n, where for each $i = 1, 2, \ldots, t$, G_i is the spanning SUBGRAPH of K_n consisting of all EDGES colored with the ith color. The irredundant Ramsey number $s(q_1, \ldots, q_t)$ is the smallest INTEGER n such that for any t-EDGE coloring of K_n, the COMPLEMENT GRAPH $\overline{G_i}$ has an irredundant set of size q_i for at least one $i = 1, \ldots, t$. Irredundant Ramsey numbers were introduced by Brewster *et al.* (1989) and satisfy

$$s(q_1, \ldots, q_t) \leq R(q_1, \ldots, q_t).$$

For a summary, see Mynhardt (1992).

s	Bounds	Reference
$s(3,3)$	6	Brewster *et al.* 1989
$s(3,4)$	8	Brewster *et al.* 1989
$s(3,5)$	12	Brewster *et al.* 1989
$s(3,6)$	15	Brewster *et al.* 1990
$s(3,7)$	18	Chen and Rousseau 1995, Cockayne *et al.* 1991
$s(4,4)$	13	Cockayne *et al.* 1992
$s(3,3,3)$	13	Cockayne and Mynhardt 1994

References

Brewster, R. C.; Cockayne, E. J.; and Mynhardt, C. M. "Irredundant Ramsey Numbers for Graphs." *J. Graph Theory* **13**, 283–290, 1989.

Brewster, R. C.; Cockayne, E. J.; and Mynhardt, C. M. "The Irredundant Ramsey Number $s(3,6)$." *Quaest. Math.* **13**, 141–157, 1990.

Chen, G. and Rousseau, C. C. "The Irredundant Ramsey Number $s(3,7)$." *J. Graph. Th.* **19**, 263–270, 1995.

Cockayne, E. J.; Exoo, G.; Hattingh, J. H.; and Mynhardt, C. M. "The Irredundant Ramsey Number $s(4,4)$." *Util. Math.* **41**, 119–128, 1992.

Cockayne, E. J.; Hattingh, J. H.; and Mynhardt, C. M. "The Irredundant Ramsey Number $s(3,7)$." *Util. Math.* **39**, 145–160, 1991.

Cockayne, E. J. and Mynhardt, C. M. "The Irredundant Ramsey Number $s(3,3,3) = 13$." *J. Graph. Th.* **18**, 595–604, 1994.

Hattingh, J. H. "On Irredundant Ramsey Numbers for Graphs." *J. Graph Th.* **14**, 437–441, 1990.

Mynhardt, C. M. "Irredundant Ramsey Numbers for Graphs: A Survey." *Congres. Numer.* **86**, 65–79, 1992.

Irreflexive

A RELATION R on a SET S is irreflexive provided that no element is related to itself; in other words, xRx for no x in S.

see also RELATION

Irregular Pair

If p divides the NUMERATOR of the BERNOULLI NUMBER B_{2k} for $0 < 2k < p-1$, then $(p, 2k)$ is called an irregular pair. For $p < 30000$, the irregular pairs of various forms are $p = 16843$ for $(p, p-3)$, $p = 37$ for $(p, p-5)$, none for $(p, p-7)$, and $p = 67,877$ for $(p, p-9)$.

see also BERNOULLI NUMBER, IRREGULAR PRIME

References

Johnson, W. "Irregular Primes and Cyclotomic Invariants." *Math. Comput.* **29**, 113–120, 1975.

Irregular Prime

PRIMES for which Kummer's theorem on the unsolvability of FERMAT'S LAST THEOREM does not apply. An irregular prime p divides the NUMERATOR of one of the BERNOULLI NUMBERS B_{10}, B_{12}, ..., B_{2p-2}, as shown by Kummer in 1850. The FERMAT EQUATION has no solutions for REGULAR PRIMES.

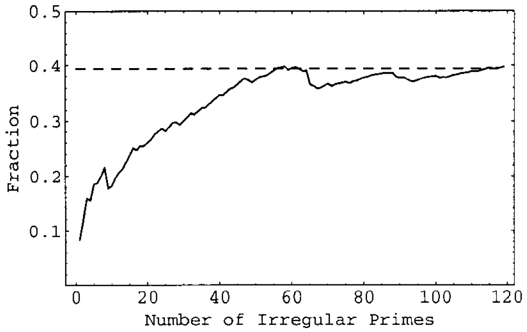

An INFINITE number of irregular PRIMES exist, as proven in 1915 by Jensen. The first few irregular primes are 37, 59, 67, 101, 103, 131, 149, 157, ... (Sloane's A000928). Of the 283,145 PRIMES less than 4×10^6, 111,597 (or 39.41%) are regular. The conjectured FRACTION is $1 - e^{-1/2} \approx 39.35\%$ (Ribenboim 1996, p. 415).

see also BERNOULLI NUMBER, FERMAT'S LAST THEOREM, IRREGULAR PAIR, REGULAR PRIME

References

Buhler, J.; Crandall, R.; Ernvall, R.; and Metsänkylä, T. "Irregular Primes and Cyclotomic Invariants to Four Million." *Math. Comput.* **60**, 151–153, 1993.

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, p. 202, 1979.

Johnson, W. "Irregular Primes and Cyclotomic Invariants." *Math. Comput.* **29**, 113–120, 1975.

Ribenboim, P. *The New Book of Prime Number Records.* New York: Springer-Verlag, pp. 325–329 and 414–425, 1996.

Sloane, N. J. A. Sequence A000928/M5260 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Stewart, C. L. "A Note on the Fermat Equation." *Mathematika* **24**, 130–132, 1977.

Irregular Singularity

Consider a second-order ORDINARY DIFFERENTIAL EQUATION

$$y'' + P(x)y' + Q(x)y = 0.$$

If $P(x)$ and $Q(x)$ remain FINITE at $x = x_0$, then x_0 is called an ORDINARY POINT. If either $P(x)$ or $Q(x)$ diverges as $x \to x_0$, then x_0 is called a singular point. If $P(x)$ diverges more quickly than $1/(x - x_0)$, so $(x - x_0)P(x)$ approaches INFINITY as $x \to x_0$, or $Q(x)$ diverges more quickly than $1/(x - x_0)^2 Q$ so that $(x - x_0)^2 Q(x)$ goes to INFINITY as $x \to x_0$, then x_0 is called an IRREGULAR SINGULARITY (or ESSENTIAL SINGULARITY).

see also ORDINARY POINT, REGULAR SINGULAR POINT, SINGULAR POINT (DIFFERENTIAL EQUATION)

References

Arfken, G. "Singular Points." §8.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 451–453 and 461–463, 1985.

Irrotational Field

A VECTOR FIELD $\mathbf{v}$ for which the CURL vanishes,

$$\nabla \times \mathbf{v} = \mathbf{0}.$$

see also BELTRAMI FIELD, CONSERVATIVE FIELD, SOLENOIDAL FIELD, VECTOR FIELD

Isarithm

see EQUIPOTENTIAL CURVE

ISBN

Publisher	Digits
Addison-Wesley	0201
Amer. Math. Soc.	0821
Cambridge University Press	0521
CRC Press	0849
Dover	0486
McGraw-Hill	0070
Oxford University Press	0198
Springer-Verlag	0387
Wiley	0471

The International Standard Book Number (ISBN) is a 10-digit CODE which is used to identify a book uniquely. The first four digits specify the publisher, the next five digits the book, and the last digit d_{10} is a check digit which may be in the range 0–9 or X (where X equals 10). The check digit is computed from the equation

$$10d_1 + 9d_2 + 8d_3 + \ldots + 2d_9 + d_{10} \equiv 0 \pmod{11}.$$

For example, the number for this book is 0-8493-9640-9, and

$$\begin{aligned}&10 \cdot 0 + 9 \cdot 8 + 8 \cdot 4 + 7 \cdot 9 + 6 \cdot 3 + 5 \cdot 9\\&+4 \cdot 6 + 3 \cdot 4 + 2 \cdot 0 + 1 \cdot 9 = 275 = 25 \cdot 11 \equiv 0 \pmod{11},\end{aligned}$$

as required.

see also CODE

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 894, 1992.

Island

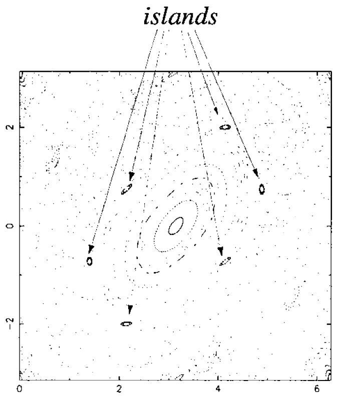

If an integrable QUASIPERIODIC system is slightly perturbed so that it becomes nonintegrable, only a finite number of n-CYCLES remain as a result of MODE LOCKING. One will be elliptical and one will be hyperbolic.

Surrounding the ELLIPTIC FIXED POINT is a region of stable ORBITS which circle it, as illustrated above in the STANDARD MAP with $K = 1.5$. As the map is iteratively applied, the island is mapped to a similar structure surrounding the next point of the elliptic cycle. The map thus has a chain of islands, with the FIXED POINT alternating between ELLIPTIC (at the center of the islands) and HYPERBOLIC (between islands). Because the unperturbed system goes through an INFINITY of rational values, the perturbed system must have an INFINITE number of island chains.

see also MODE LOCKING, ORBIT (MAP), QUASIPERIODIC FUNCTION

References

Kármán, T. von and Biot, M. A. *Mathematical Methods in Engineering: An Introduction to the Mathematical Treatment of Engineering Problems.* New York: McGraw-Hill, pp. 3 and 7, 1940.

Isobaric Polynomial

A POLYNOMIAL in which the sum of SUBSCRIPTS is the same in each term.

see also HOMOGENEOUS POLYNOMIAL

Isochronous Curve

see SEMICUBICAL PARABOLA, TAUTOCHRONE PROBLEM

Isoclinal

see ISOCLINE

Isocline

A graphical method of solving an ORDINARY DIFFERENTIAL EQUATION of the form

$$\frac{dy}{dx} = f(x,y)$$

by plotting a series of curves $f(x,y) = [\text{const}]$, then drawing a curve PERPENDICULAR to each curve such that it satisfies the initial condition. This curve is the solution to the ORDINARY DIFFERENTIAL EQUATION.

Isoclinic Groups

Two GROUPS G and H are said to be isoclinic if there are isomorphisms $G/Z(G) \to H/Z(H)$ and $G' \to H'$, where $Z(G)$ is the CENTER of the group, which identify the two commutator maps.

References

Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. "Isoclinism." §6.7 in *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, pp. xxiii–xxiv, 1985.

Isodynamic Points

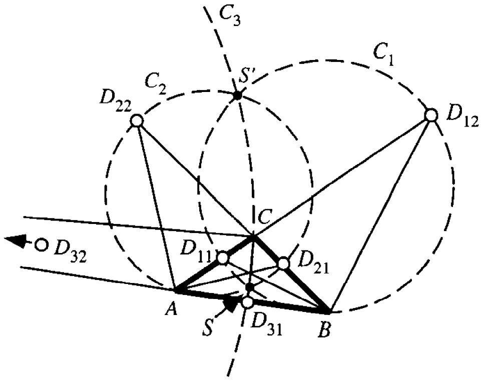

The first and second isodynamic points of a TRIANGLE ΔABC can be constructed by drawing the triangle's ANGLE BISECTORS and EXTERIOR ANGLE BISECTORS. Each pair of bisectors intersects a side of the triangle (or its extension) in two points D_{i1} and D_{i2}, for $i = 1$, 2, 3. The three CIRCLES having $D_{11}D_{12}$, $D_{21}D_{22}$, and $D_{31}D_{32}$ as DIAMETERS are the APOLLONIUS CIRCLES C_1, C_2, and C_3. The points S and S' in which the three APOLLONIUS CIRCLES intersect are the first and second isodynamic points, respectively.

S and S' have TRIANGLE CENTER FUNCTIONS

$$\alpha = \sin(A \pm \tfrac{1}{3}\pi),$$

respectively. The ANTIPEDAL TRIANGLES of both points are EQUILATERAL and have AREAS

$$\Delta' = 2\Delta[\cot\omega\cot(\tfrac{1}{3}\pi)],$$

where ω is the BROCARD ANGLE.

The isodynamic points are ISOGONAL CONJUGATES of the ISOGONIC CENTERS. They lie on the BROCARD AXIS. The distances from either isodynamic point to the VERTICES are inversely proportional to the sides. The PEDAL TRIANGLE of either isodynamic point is an EQUILATERAL TRIANGLE. An INVERSION with either isodynamic point as the INVERSION CENTER transforms the triangle into an EQUILATERAL TRIANGLE.

The CIRCLE which passes through both the isodynamic points and the CENTROID of a TRIANGLE is known as the PARRY CIRCLE.

see also APOLLONIUS CIRCLES, BROCARD AXIS, CENTROID (TRIANGLE), ISOGONIC CENTERS, PARRY CIRCLE

References
Gallatly, W. *The Modern Geometry of the Triangle, 2nd ed.* London: Hodgson, p. 106, 1913.
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 295–297, 1929.
Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Isoenergetic Nondegeneracy

The condition for isoenergetic nondegeneracy for a Hamiltonian

$$H = H_0(\mathbf{I}) + \epsilon H_1(\mathbf{I}, \boldsymbol{\theta})$$

is

$$\begin{vmatrix} \frac{\partial^2 H_0}{\partial I_i \partial I_j} & \frac{\partial H_0}{\partial I_i} \\ \frac{\partial H_0}{\partial I_j} & 0 \end{vmatrix} \neq 0,$$

which guarantees the EXISTENCE on every energy level surface of a set of invariant tori whose complement has a small MEASURE.

References
Tabor, M. *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, pp. 113–114, 1989.

Isogonal Conjugate

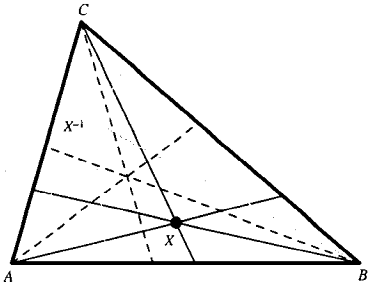

The isogonal conjugate X^{-1} of a point X in the plane of the TRIANGLE ΔABC is constructed by reflecting the lines AX, BX, and CX about the ANGLE BISECTORS at A, B, and C. The three reflected lines CONCUR at the isogonal conjugate. The TRILINEAR COORDINATES of the isogonal conjugate of the point with coordinates

$$\alpha : \beta : \gamma$$

are

$$\alpha^{-1} : \beta^{-1} : \gamma^{-1}.$$

Isogonal conjugation maps the interior of a TRIANGLE onto itself. This mapping transforms lines onto CONIC SECTIONS that CIRCUMSCRIBE the TRIANGLE. The type of CONIC SECTION is determined by whether the line d meets the CIRCUMCIRCLE C',

1. If d does not intersect C', the isogonal transform is an ELLIPSE;
2. If d is tangent to C', the transform is a PARABOLA;
3. If d cuts C', the transform is a HYPERBOLA, which is a RECTANGULAR HYPERBOLA if the line passes through the CIRCUMCENTER

(Casey 1893, Vandeghen 1965).

The isogonal conjugate of a point on the CIRCUMCIRCLE is a POINT AT INFINITY (and conversely). The sides of the PEDAL TRIANGLE of a point are PERPENDICULAR to the connectors of the corresponding VERTICES with the isogonal conjugate. The isogonal conjugate of a set of points is the LOCUS of their isogonal conjugate points.

The product of ISOTOMIC and isogonal conjugation is a COLLINEATION which transforms the sides of a TRIANGLE to themselves (Vandeghen 1965).

see also ANTIPEDAL TRIANGLE, COLLINEATION, ISOGONAL LINE, ISOTOMIC CONJUGATE POINT, LINE AT INFINITY, SYMMEDIAN LINE

References
Casey, J. *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions with Numerous Examples, 2nd rev. enl. ed.* Dublin: Hodges, Figgis, & Co., 1893.
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 153–158, 1929.
Vandeghen, A. "Some Remarks on the Isogonal and Cevian Transforms. Alignments of Remarkable Points of a Triangle." *Amer. Math. Monthly* **72**, 1091–1094, 1965.

Isogonal Line

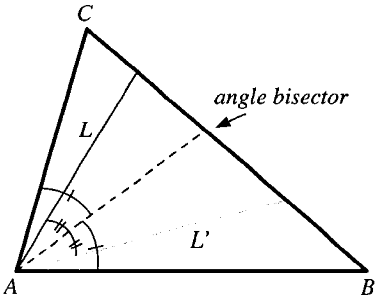

The line L' through a TRIANGLE VERTEX obtained by reflecting an initial line L (also through a VERTEX) about the ANGLE BISECTOR. If three lines from the VERTICES of a TRIANGLE ΔABC are CONCURRENT at $X = L_1 L_2 L_3$, then their isogonal lines are also CONCURRENT, and the point of concurrence $X' = L_1' L_2' L_3'$ is called the ISOGONAL CONJUGATE point.

see also ISOGONAL CONJUGATE

References
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 153–157, 1929.

Isogonic Centers

The first isogonic center F_1 of a TRIANGLE is the FERMAT POINT. The second isogonic center F_2 is constructed analogously with the first isogonic center except that for F_2, the EQUILATERAL TRIANGLES are constructed on the same side of the opposite VERTEX. The second isogonic center has TRIANGLE CENTER FUNCTION

$$\alpha = \csc(A - \tfrac{1}{3}\pi).$$

Its ANTIPEDAL TRIANGLE is EQUILATERAL and has AREA

$$2\Delta = [-1 + \cot\omega \cot(\tfrac{1}{3}\pi)],$$

where ω is the BROCARD ANGLE.

The first and second isogonic centers are ISOGONAL CONJUGATES of the ISODYNAMIC POINTS.

see also BROCARD ANGLE, EQUILATERAL TRIANGLE, FERMAT POINT, ISODYNAMIC POINTS, ISOGONAL CONJUGATE

References

Gallatly, W. *The Modern Geometry of the Triangle, 2nd ed.* London: Hodgson, p. 107, 1913.

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Isograph

The substitution of $re^{i\theta}$ for z in a POLYNOMIAL $p(z)$. $p(z)$ is then plotted as a function of θ for a given r in the COMPLEX PLANE. By varying r so that the curve passes through the ORIGIN, it is possible to determine a value for one ROOT of the POLYNOMIAL.

Isohedral Tiling

Let $S(T)$ be the group of symmetries which map a MONOHEDRAL TILING T onto itself. The TRANSITIVITY CLASS of a given tile T is then the collection of all tiles to which T can be mapped by one of the symmetries of $S(T)$. If T has k TRANSITIVITY CLASSES, then T is said to be k-isohedral. Berglund (1993) gives examples of k-isohedral tilings for $k = 1$, 2, and 4.

see also ANISOHEDRAL TILING

References

Berglund, J. "Is There a k-Anisohedral Tile for $k \geq 5$?" *Amer. Math. Monthly* **100**, 585-588, 1993.

Grünbaum, B. and Shephard, G. C. "The 81 Types of Isohedral Tilings of the Plane." *Math. Proc. Cambridge Philos. Soc.* **82**, 177-196, 1977.

Isohedron

A convex POLYHEDRON with symmetries acting transitively on its faces. Every isohedron has an EVEN number of faces (Grünbaum 1960).

References

Grünbaum, B. "On Polyhedra in E^3 Having All Faces Congruent." *Bull. Research Council Israel* **8F**, 215–218, 1960.

Grünbaum, B. and Shepard, G. C. "Spherical Tilings with Transitivity Properties." In *The Geometric Vein: The Coxeter Festschrift* (Ed. C. Davis, B. Grünbaum, and F. Shenk). New York: Springer-Verlag, 1982.

Isolated Point

A point on a curve, also known as an ACNODE or HERMIT POINT, which has no other points in its NEIGHBORHOOD.

Isolated Singularity

An isolated singularity is a SINGULARITY for which there exists a (small) REAL NUMBER ϵ such that there are no other SINGULARITIES within a NEIGHBORHOOD of radius ϵ centered about the SINGULARITY.

The types of isolated singularities possible for CUBIC SURFACES have been classified (Schläfli 1864, Cayley 1869, Bruce and Wall 1979) and are summarized in the following table from Fischer (1986).

Double Pt. Name	Symbol	Normal Form	Coxeter Diagram
conic	C_2	$x^2 + y^2 + z^2$	A_1
biplanar	B_3	$x^2 + y^2 + z^3$	A_2
biplanar	B_4	$x^2 + y^2 + z^4$	A_3
biplanar	B_5	$x^2 + y^2 + z^5$	A_4
biplanar	B_6	$x^2 + y^2 + z^6$	A_5
uniplanar	U_6	$x^2 + z(y^2 + z^2)$	D_4
uniplanar	U_7	$x^2 + z(y^2 + z^3)$	D_5
uniplanar	U_8	$x^2 + y^3 + z^4$	E_6
elliptic cone pt.	—	$xy^2 - 4z^3$ $-g_2x^2y + g_3x^3$	$\tilde{E}_6$

see also CUBIC SURFACE, RATIONAL DOUBLE POINT, SINGULARITY

References

Bruce, J. and Wall, C. T. C. "On the Classification of Cubic Surfaces." *J. London Math. Soc.* **19**, 245–256, 1979.

Cayley, A. "A Memoir on Cubic Surfaces." *Phil. Trans. Roy. Soc.* **159**, 231–326, 1869.

Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, pp. 12–13, 1986.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 380–381, 1953.

Schläfli, L. "On the Distribution of Surfaces of Third Order into Species." *Phil. Trans. Roy. Soc.* **153**, 193–247, 1864.

Isolating Integral

An integral of motion which restricts the PHASE SPACE available to a DYNAMICAL SYSTEM.

Isometry

A BIJECTIVE MAP between two METRIC SPACES that preserves distances, i.e.,

$$d(f(x), f(y)) = d(x, y),$$

where f is the MAP and $d(a, b)$ is the DISTANCE function.

An isometry of the PLANE is a linear transformation which preserves length. Isometries include ROTATION, TRANSLATION, REFLECTION, GLIDES, and the IDENTITY MAP. If an isometry has more than one FIXED

POINT, it must be either the identity transformation or a reflection. Every isometry of period two (two applications of the transformation preserving lengths in the original configuration) is either a reflection or a half turn rotation. Every isometry in the plane is the product of at most three reflections (at most two if there is a FIXED POINT). Every finite group of isometries has at least one FIXED POINT.

see also DISTANCE, EUCLIDEAN MOTION, HJELMSLEV'S THEOREM, LENGTH (CURVE), REFLECTION, ROTATION, TRANSLATION

References
Gray, A. "Isometries of Surfaces." §13.2 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 255–258, 1993.

Isometric Latitude

An AUXILIARY LATITUDE which is directly proportional to the spacing of parallels of LATITUDE from the equator on an ellipsoidal MERCATOR PROJECTION. It is defined by

$$\psi = \ln\left|\tan(\tfrac{1}{4}\pi + \tfrac{1}{2}\phi)\left(\frac{1 - e\sin\phi}{1 + e\sin\phi}\right)^{e/2}\right|, \tag{1}$$

where the symbol τ is sometimes used instead of ψ. The isometric latitude is related to the CONFORMAL LATITUDE by

$$\psi = \ln\tan(\tfrac{1}{4}\pi + \tfrac{1}{2}\chi). \tag{2}$$

The inverse is found by iterating

$$\phi = 2\tan^{-1}\left[\exp(\psi)\left(\frac{1 + e\sin\phi}{1 - e\sin\phi}\right)^{e/2}\right] - \tfrac{1}{2}\pi, \tag{3}$$

with the first trial as

$$\phi_0 = 2\tan^{-1}(e^\psi) - \tfrac{1}{2}\pi. \tag{4}$$

see also LATITUDE

References
Adams, O. S. "Latitude Developments Connected with Geodesy and Cartography with Tables, Including a Table for Lambert Equal-Area Meridional Projections." Spec. Pub. No. 67. U. S. Coast and Geodetic Survey, 1921.
Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, p. 15, 1987.

Isomorphic Graphs

Two GRAPHS which contain the same number of VERTICES connected in the same way are said to be isomorphic. Formally, two graphs G and H with VERTICES $V_n = \{1, 2, \ldots, n\}$ are said to be isomorphic if there is a PERMUTATION p of V_n such that $\{u, v\}$ is in the set of EDGES $E(G)$ IFF $\{p(u), p(v)\}$ is in the set of EDGES $E(H)$.

References
Chartrand, G. "Isomorphic Graphs." §2.2 in *Introductory Graph Theory.* New York: Dover, pp. 32–40, 1985.

Isomorphic Groups

Two GROUPS are isomorphic if the correspondence between them is ONE-TO-ONE and the "multiplication" table is preserved. For example, the POINT GROUPS C_2 and D_1 are isomorphic GROUPS, written $C_2 \cong D_1$ or $C_2 \rightleftharpoons D_1$ (Shanks 1993). Note that the symbol $\cong$ is also used to denote geometric CONGRUENCE.

References
Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, 1993.

Isomorphic Posets

Two POSETS are said to be isomorphic if their "structures" are entirely analogous. Formally, POSETS $P = (X, \leq)$ and $Q = (X', \leq')$ are isomorphic if there is a BIJECTION f from X to X' such that $x \leq x'$ precisely when $f(x) \leq' f(x')$.

Isomorphism

Isomorphism is a very general concept which appears in several areas of mathematics. Formally, an isomorphism is BIJECTIVE MORPHISM. Informally, an isomorphism is a map which preserves sets and relations among elements.

A space isomorphism is a VECTOR SPACE in which addition and scalar multiplication are preserved. An isomorphism of a TOPOLOGICAL SPACE is called a HOMEOMORPHISM.

Two groups G_1 and G_2 with binary operators $+$ and $\times$ are isomorphic if there exists a map $f : G_1 \mapsto G_2$ which satisfies

$$f(x + y) = f(x) \times f(y).$$

An isomorphism preserves the identities and inverses of a GROUP. A GROUP which is isomorphic to itself is called an AUTOMORPHISM.

see also AUTOMORPHISM, AX-KOCHEN ISOMORPHISM THEOREM, HOMEOMORPHISM, MORPHISM

Isoperimetric Inequality

Let a PLANE figure have AREA A and PERIMETER p. Let the CIRCLE of PERIMETER p have RADIUS r. Then

$$\frac{4\pi A}{p^2} \leq 1,$$

where the quantity on the left is known as the ISOPERIMETRIC QUOTIENT.

Isoperimetric Point

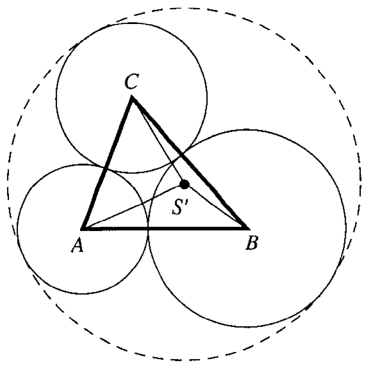

The point S' which makes the PERIMETERS of the TRIANGLES $\Delta BS'C$, $\Delta CS'A$, and $\Delta AS'B$ equal. The isoperimetric point exists IFF the largest ANGLE of the triangle satisfies

$$\max(A,B,C) < 2\sin^{-1}(\tfrac{4}{5}) \approx 1.85459 \text{ rad} \approx 106.26^\circ,$$

or equivalently

$$a+b+c > 4R+r,$$

where a, b, and c are the side lengths of ΔABC, r is the INRADIUS, and R is the CIRCUMRADIUS. The isoperimetric point is also the center of the outer SODDY CIRCLE of ΔABC and has TRIANGLE CENTER FUNCTION

$$\alpha = 1 - \frac{2\Delta}{a(b+c-a)} = \sec(\tfrac{1}{2}A)\cos(\tfrac{1}{2}B)\cos(\tfrac{1}{2}C) - 1.$$

see also EQUAL DETOUR POINT, PERIMETER, SODDY CIRCLES

References

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Kimberling, C. "Isoperimetric Point and Equal Detour Point." http://www.evansville.edu/~ck6/tcenters/recent/isoper.html.

Kimberling, C. and Wagner, R. W. "Problem E 3020 and Solution." *Amer. Math. Monthly* **93**, 650–652, 1986.

Veldkamp, G. R. "The Isoperimetric Point and the Point(s) of Equal Detour." *Amer. Math. Monthly* **92**, 546–558, 1985.

Isoperimetric Problem

Find a closed plane curve of a given length which encloses the greatest AREA. The solution is a CIRCLE. If the class of curves to be considered is limited to smooth curves, the isoperimetric problem can be stated symbolically as follows: find an arc with parametric equations $x = x(t)$, $y = y(t)$ for $t \in [t_1, t_2]$ such that $x(t_1) = x(t_2)$, $y(t_1) = y(t_2)$ (where no further intersections occur) constrained by

$$l = \int_{t_1}^{t_2} \sqrt{x'^2 + y'^2}\,dt$$

such that

$$A = \tfrac{1}{2}\int_{t_1}^{t_2} (xy' - x'y)\,dt$$

is a MAXIMUM.

see also DIDO'S PROBLEM, ISOVOLUME PROBLEM

References

Bogomolny, A. "Isoperimetric Theorem and Inequality." http://www.cut-the-knot.com/do_you_know/isoperimetric.html.

Isenberg, C. Appendix V in *The Science of Soap Films and Soap Bubbles.* New York: Dover, 1992.

Isoperimetric Quotient

A quantity defined in the ISOPERIMETRIC INEQUALITY

$$Q \equiv \frac{4\pi A}{p^2}.$$

see also ISOPERIMETRIC INEQUALITY

Isoperimetric Theorem

Of all convex n-gons of a given PERIMETER, the one which maximizes AREA is the regular n-gon.

see also ISOPERIMETRIC INEQUALITY, ISOPERIMETRIC PROBLEM

Isopleth

see EQUIPOTENTIAL CURVE

Isoptic Curve

For a given curve C, consider the locus of the point P from where the TANGENTS from P to C meet at a fixed given ANGLE. This is called an isoptic curve of the given curve.

Curve	Isoptic
cycloid	curtate or prolate cycloid
epicycloid	epitrochoid
hypocycloid	hypotrochoid
parabola	hyperbola
sinusoidal spiral	sinusoidal spiral

see also ORTHOPTIC CURVE

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 58–59 and 206, 1972.

Yates, R. C. "Isoptic Curves." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 138–140, 1952.

Isosceles Tetrahedron

A nonregular TETRAHEDRON in which each pair of opposite EDGES are equal such that all triangular faces are congruent. A TETRAHEDRON is isosceles IFF the sum of the face angles at each VERTEX is 180°, and IFF its INSPHERE and CIRCUMSPHERE are concentric.

The only way for all the faces of a TETRAHEDRON to have the same PERIMETER or to have the same AREA is for them to be fully congruent, in which case the tetrahedron is isosceles.

see also CIRCUMSPHERE, INSPHERE, ISOSCELES TRIANGLE, TETRAHEDRON

References

Brown, B. H. "Theorem of Bang. Isosceles Tetrahedra." *Amer. Math. Monthly* **33**, 224–226, 1926.

Honsberger, R. "A Theorem of Bang and the Isosceles Tetrahedron." Ch. 9 in *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., pp. 90–97, 1976.

Isosceles Triangle

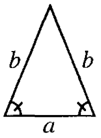

A TRIANGLE with two equal sides (and two equal ANGLES). The name derives from the Greek *iso* (same) and *skelos* (LEG). The height of the above isosceles triangle can be found from the PYTHAGOREAN THEOREM as

$$h = \sqrt{b^2 - \tfrac{1}{4}a^2}. \qquad (1)$$

The AREA is therefore given by

$$A = \tfrac{1}{2}ah = \tfrac{1}{2}a\sqrt{b^2 - \tfrac{1}{4}a^2}. \qquad (2)$$

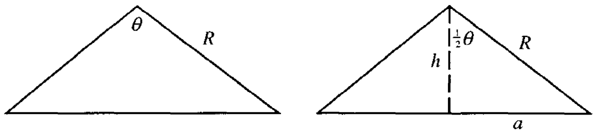

There is a surprisingly simple relationship between the AREA and VERTEX ANGLE θ. As shown in the above diagram, simple TRIGONOMETRY gives

$$h = R\cos(\tfrac{1}{2}\theta) \qquad (3)$$

$$a = R\sin(\tfrac{1}{2}\theta), \qquad (4)$$

so the AREA is

$$A = \tfrac{1}{2}(2a)h = ah = R^2\cos(\tfrac{1}{2}\theta)\sin(\tfrac{1}{2}\theta) = \tfrac{1}{2}R^2\sin\theta. \qquad (5)$$

No set of $n > 6$ points in the PLANE can determine only ISOSCELES TRIANGLES.

see also ACUTE TRIANGLE, EQUILATERAL TRIANGLE, INTERNAL BISECTORS PROBLEM, ISOSCELES TETRAHEDRON, ISOSCELIZER, OBTUSE TRIANGLE, POINT PICKING, PONS ASINORUM, RIGHT TRIANGLE, SCALENE TRIANGLE, STEINER-LEHMUS THEOREM

Isoscelizer

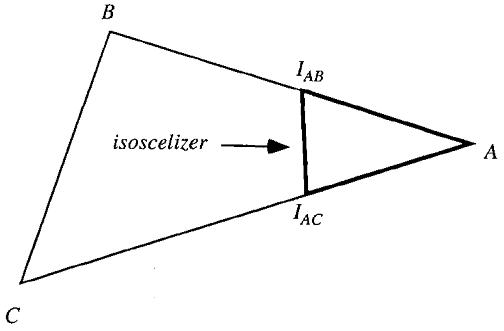

An isoscelizer of an ANGLE A in a TRIANGLE ΔABC is a LINE SEGMENT $I_{AB}I_{AC}$ where I_{AB} lies on AB and I_{AC} on AC such that $\Delta AI_{AB}I_{AC}$ is an ISOSCELES TRIANGLE.

see also CONGRUENT ISOSCELIZERS POINT, ISOSCELES TRIANGLE, YFF CENTER OF CONGRUENCE

Isospectral Manifolds

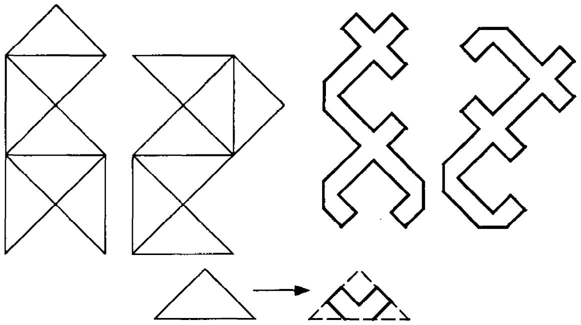

DRUMS that sound the same, i.e., have the same eigenfrequency spectrum. Two drums with differing AREA, PERIMETER, or GENUS can always be distinguished. However, Kac (1966) asked if it was possible to construct differently shaped drums which have the same eigenfrequency spectrum. This question was answered in the affirmative by Gordon *et al.* (1992). Two such isospectral manifolds are shown in the right figure above (Cipra 1992).

References

Chapman, S. J. "Drums That Sound the Same." *Amer. Math. Monthly* **102**, 124–138, 1995.

Cipra, B. "You Can't Hear the Shape of a Drum." *Science* **255**, 1642–1643, 1992.

Gordon, C.; Webb, D.; and Wolpert, S. "Isospectral Plane Domains and Surfaces via Riemannian Orbifolds." *Invent. Math.* **110**, 1–22, 1992.

Gordon, C.; Webb, D.; and Wolpert, S. "You Cannot Hear the Shape of a Drum." *Bull. Amer. Math. Soc.* **27**, 134–138, 1992.

Kac, M. "Can One Hear the Shape of a Drum?" *Amer. Math. Monthly* **73**, 1–23, 1966.

Isothermal Parameterization

A parameterization is isothermal if, for $\zeta \equiv u + iv$ and

$$\phi_k(\zeta) = \frac{\partial x_k}{\partial u} - i\frac{\partial x_k}{\partial v},$$

the identity

$$\phi_1{}^2(\zeta) + \phi_2{}^2(\zeta) + \phi_3{}^2(\zeta) = 0$$

holds.

see also Minimal Surface, Temperature

Isotomic Conjugate Point

The point of concurrence Q of the Isotomic Lines relative to a point P. The isotomic conjugate $\alpha' : \beta' : \gamma'$ of a point with Trilinear Coordinates $\alpha : \beta : \gamma$ is

$$(a^2\alpha)^{-1} : (b^2\beta)^{-1} : (c^2\gamma)^{-1}. \tag{1}$$

The isotomic conjugate of a Line d having trilinear equation

$$l\alpha + m\beta + n\gamma \tag{2}$$

is a Conic Section circumscribed on the Triangle ΔABC (Casey 1893, Vandeghen 1965). The isotomic conjugate of the Line at Infinity having trilinear equation

$$a\alpha + b\beta + c\gamma = 0 \tag{3}$$

is Steiner's Ellipse

$$\frac{\beta'\gamma'}{a} + \frac{\gamma'\alpha'}{b} + \frac{\alpha'\beta'}{c} = 0 \tag{4}$$

(Vandeghen 1965). The type of Conic Section to which d is transformed is determined by whether the line d meets Steiner's Ellipse E.

1. If d does not intersect E, the isotomic transform is an Ellipse.
2. If d is tangent to E, the transform is a Parabola.
3. If d cuts E, the transform is a Hyperbola, which is a Rectangular Hyperbola if the line passes through the isotomic conjugate of the Orthocenter

(Casey 1893, Vandeghen 1965).

There are four points which are isotomically self-conjugate: the Centroid M and each of the points of intersection of lines through the Vertices Parallel to the opposite sides. The isotomic conjugate of the Euler Line is called Jerabek's Hyperbola (Casey 1893, Vandeghen 1965).

Vandeghen (1965) calls the transformation taking points to their isotomic conjugate points the Cevian Transform. The product of isotomic and Isogonal is a Collineation which transforms the sides of a Triangle to themselves (Vandeghen 1965).

see also Cevian Transform, Gergonne Point, Isogonal Conjugate, Jerabek's Hyperbola, Nagel Point, Steiner's Ellipse

References

Casey, J. *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions with Numerous Examples, 2nd rev. enl. ed.* Dublin: Hodges, Figgis, & Co., 1893.

Eddy, R. H. and Fritsch, R. "The Conics of Ludwig Kiepert: A Comprehensive Lesson in the Geometry of the Triangle." *Math. Mag.* **67**, 188–205, 1994.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 157–159, 1929.

Vandeghen, A. "Some Remarks on the Isogonal and Cevian Transforms. Alignments of Remarkable Points of a Triangle." *Amer. Math. Monthly* **72**, 1091–1094, 1965.

Isotomic Lines

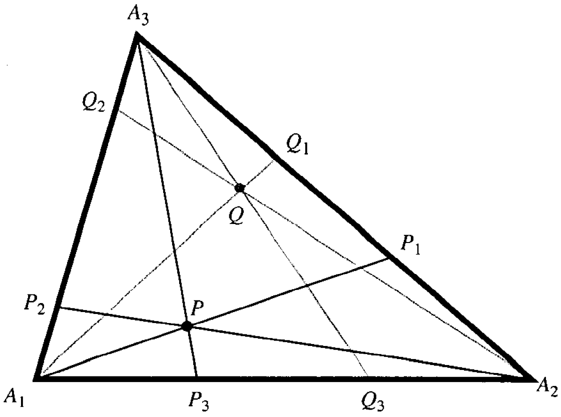

Given a point P in the interior of a Triangle $\Delta A_1A_2A_3$, draw the Cevians through P from each Vertex which meet the opposite sides at P_1, P_2, and P_3. Now, mark off point Q_1 along side A_2A_3 such that $A_3P_1 = A_2Q_1$, etc., i.e., so that Q_i and P_i are equidistance from the Midpoint of A_jA_k. The lines A_1Q_1, A_2Q_2, and A_3Q_3 then coincide in a point Q known as the Isotomic Conjugate Point.

see also Cevian, Isotomic Conjugate Point, Midpoint

Isotone Map

A Map which is monotone increasing and therefore order-preserving.

Isotope

To rearrange without cutting or pasting.

Isotopy

A Homotopy from one embedding of a Manifold M in N to another such that at every time, it is an embedding. The notion of isotopy is category independent, so notions of topological, piecewise-linear, smooth, isotopy (and so on) exist. When no explicit mention is made, "isotopy" usually means "smooth isotopy."

see also Ambient Isotopy, Regular Isotopy

Isotropic Tensor

A TENSOR which has the same components in all rotated coordinate systems.

rank	isotropic tensors
0	all
1	none
2	Kronecker delta
3	1
4	3

Isovolume Problem

Find the surface enclosing the maximum volume per unit surface AREA $I \equiv V/S$. The solution is a SPHERE, which has

$$I_{\text{sphere}} = \frac{\frac{4}{3}\pi r^3}{4\pi r^2} = \tfrac{1}{3}r.$$

see also DIDO'S PROBLEM, ISOPERIMETRIC PROBLEM

References

Bogomolny, A. "Isoperimetric Theorem and Inequality." `http://www.cut-the-knot.com/do_you_know/isoperimetric.html`.

Isenberg, C. Appendix VI in *The Science of Soap Films and Soap Bubbles.* New York: Dover, 1992.

Isthmus

see BRIDGE (GRAPH)

Iterated Exponential Constants

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Euler (Le Lionnais 1983) and Eisenstein (1844) showed that the function $h(x) = x^{x^{x^{\cdot^{\cdot^{\cdot}}}}}$, where x^{x^x} is an abbreviation for $x^{(x^x)}$, converges only for $e^{-e} \leq x \leq e^{1/e}$, that is, $0.0659\ldots \leq x \leq 1.44466\ldots$. The value it converges to is the inverse of $x^{1/x}$, which has a closed form expression in terms of LAMBERT'S W-FUNCTION,

$$h(z) = \frac{W(-\ln z)}{-\ln z} \tag{1}$$

(Corless *et al.*). Knoebel (1981) gives

$$h(z) = 1 + \ln x + \frac{3^2(\ln z)^2}{3!} + \frac{4^3(\ln z)^3}{4!} + \ldots \tag{2}$$

(Vardi 1991). A CONTINUED FRACTION due to Khovanskii (1963) is

$$x^{1/x} = 1 + \cfrac{2(x-1)}{x^2+1-\cfrac{(x^2-1)(x-1)^2}{3x(x+1)-\cfrac{(4x^2-1)(x-1)^2}{5x(x+1)-\cfrac{(9x^2-1)(x-1)^2}{7x(x+1)-\ldots}}}}. \tag{3}$$

The function $g(x) = x^{(1/x)^{(1/x)^{\cdot^{\cdot^{\cdot}}}}}$ converges only for $e^{-1/e} \leq x \leq e^e$, that is, $0.692\ldots \leq x \leq 15.154\ldots$. The value it converges to is the inverse of x^x.

Some interesting related integrals are

$$\int_0^1 x^x\,dx = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^n} = 0.7834305107\ldots \tag{4}$$

$$\int_0^1 x^{-x}\,dx = \sum_{n=1}^{\infty} \frac{1}{n^n} = 1.2912859971\ldots \tag{5}$$

(Spiegel 1968, Abramowitz and Stegun 1972).

see also LAMBERT'S W-FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, 1972.

Baker, I. N. and Rippon, P. J. "A Note on Complex Iteration." *Amer. Math. Monthly* **92**, 501–504, 1985.

Barrows, D. F. "Infinite Exponentials." *Amer. Math. Monthly* **43**, 150–160, 1936.

Corless, R. M.; Gonnet, G. H.; Hare, D. E. G.; and Jeffrey, D. J. "On Lambert's W Function." `ftp://watdragon.uwaterloo.ca/cs-archive/CS-93-03/W.ps.Z`.

Creutz, M. and Sternheimer, R. M. "On the Convergence of Iterated Exponentiation, Part I." *Fib. Quart.* **18**, 341–347, 1980.

Creutz, M. and Sternheimer, R. M. "On the Convergence of Iterated Exponentiation, Part II." *Fib. Quart.* **19**, 326–335, 1981.

de Villiers, J. M. and Robinson, P. N. "The Interval of Convergence and Limiting Functions of a Hyperpower Sequence." *Amer. Math. Monthly* **93**, 13–23, 1986.

Eisenstein, G. "Entwicklung von $\alpha^{\alpha^{\alpha^{\cdot^{\cdot^{\cdot}}}}}$." *J. Reine angew. Math.* **28**, 49–52, 1844.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/itrexp/itrexp.html`.

Khovanskii, A. N. *The Application of Continued Fractions and Their Generalizations to Problems in Approximation Theory.* Groningen, Netherlands: P. Noordhoff, 1963.

Knoebel, R. A. "Exponentials Reiterated." *Amer. Math. Monthly* **88**, 235–252, 1981.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, pp. 22 and 39, 1983.

Mauerer, H. "Über die Funktion $x^{x^{x^{\cdot^{\cdot^{\cdot}}}}}$ für ganzzahliges Argument (Abundanzen)." *Mitt. Math. Gesell. Hamburg* **4**, 33–50, 1901.

Spiegel, M. R. *Mathematical Handbook of Formulas and Tables.* New York: McGraw-Hill, 1968.

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 12, 1991.

Iterated Function System

A finite set of contraction maps w_i for $i = 1, 2, \ldots, N$, each with a contractivity factor $s < 1$, which map a compact METRIC SPACE onto itself. It is the basis for FRACTAL image compression techniques.

see also BARNSLEY'S FERN, SELF-SIMILARITY

References

Barnsley, M. F. "Fractal Image Compression." *Not. Amer. Math. Soc.* **43**, 657–662, 1996.

Barnsley, M. *Fractals Everywhere, 2nd ed.* Boston, MA: Academic Press, 1993.

Barnsley, M. F. and Demko, S. G. "Iterated Function Systems and the Global Construction of Fractals." *Proc. Roy. Soc. London, Ser. A* **399**, 243–275, 1985.

Barnsley, M. F. and Hurd, L. P. *Fractal Image Compression.* Wellesley, MA: A. K. Peters, 1993.

Diaconis, P. M. and Shashahani, M. "Products of Random Matrices and Computer Image Generation." *Contemp. Math.* **50**, 173–182, 1986.

Fisher, Y. *Fractal Image Compression.* New York: Springer-Verlag, 1995.

Hutchinson, J. "Fractals and Self-Similarity." *Indiana Univ. J. Math.* **30**, 713–747, 1981.

Wagon, S. "Iterated Function Systems." §5.2 in *Mathematica in Action.* New York: W. H. Freeman, pp. 149–156, 1991.

Iterated Radical

see NESTED RADICAL

Iteration Sequence

A SEQUENCE $\{a_j\}$ of POSITIVE INTEGERS is called an iteration sequence if there EXISTS a strictly increasing sequence $\{s_k\}$ of POSITIVE INTEGERS such that $a_1 = s_1 \geq 2$ and $a_j = s_{a_{j-1}}$ for $j = 2, 3, \ldots$. A NECESSARY and SUFFICIENT condition for $\{a_j\}$ to be an iteration sequence is

$$a_j \geq 2a_{j-1} - a_{j-2}$$

for all $j \geq 3$.

References

Kimberling, C. "Interspersions and Dispersions." *Proc. Amer. Math. Soc.* **117**, 313–321, 1993.

Itô's Lemma

$$V_t - V_0 = \int_0^t f_x(S_u, T-u)\,dS_u - \int_0^t f_\tau(S_u, T-u)\,du + \tfrac{1}{2}\sigma^2 \int_0^t S_u{}^2 f_{xx}(S_u, T-u)\,du,$$

where $V_t = f(S_t, \tau)$ for $0 \leq \tau \equiv T - t \leq T$, and $f \in C^{2,1}((0,\infty) \times [0,T])$.

References

Price, J. F. "Optional Mathematics is Not Optional." *Not. Amer. Math. Soc.* **43**, 964–971, 1996.

Itô's Theorem

The dimension d of any IRREDUCIBLE REPRESENTATION of a GROUP G must be a DIVISOR of the index of each maximal normal Abelian SUBGROUP of G.

see also ABELIAN GROUP, IRREDUCIBLE REPRESENTATION, SUBGROUP

References

Lomont, J. S. *Applications of Finite Groups.* New York: Dover, p. 55, 1993.

Iverson Bracket

Let S be a mathematical statement, then the Iverson bracket is defined by

$$[S] \equiv \begin{cases} 0 & \text{if } S \text{ is true} \\ 1 & \text{if } S \text{ is false.} \end{cases}$$

This notation conflicts with the brackets sometimes used to denote the FLOOR FUNCTION. For this reason, and because of the elegant symmetry of the FLOOR FUNCTION and CEILING FUNCTION symbols $\lfloor x \rfloor$ and $\lceil x \rceil$, the use of $[x]$ to denote the FLOOR FUNCTION should be deprecated.

see also CEILING FUNCTION, FLOOR FUNCTION

References

Graham, R. L.; Knuth, D. E.; and Patashnik, O. *Concrete Mathematics: A Foundation for Computer Science.* Reading, MA: Addison-Wesley, p. 24, 1990.

Iverson, K. E. *A Programming Language.* New York: Wiley, p. 11, 1962.

Iwasawa's Theorem

Every finite-dimensional LIE ALGEBRA of characteristic $p \neq 0$ has a faithful finite-dimensional representation.

References

Jacobson, N. *Lie Algebras.* New York: Dover, pp. 204–205, 1979.

J

j

The symbol used by engineers and some physicists to denote i, the IMAGINARY NUMBER $\sqrt{-1}$.

j-Conductor

see FREY CURVE

j-Function

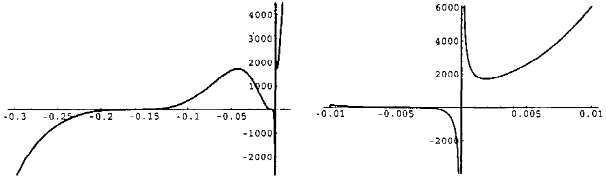

The j-function is defined as

$$j(q) \equiv 1728J(\sqrt{q}), \tag{1}$$

where

$$J(q) \equiv \frac{4}{27}\frac{[1-\lambda(q)+\lambda^2(q)]^3}{\lambda^2(q)[1-\lambda(q)]^2} \tag{2}$$

is KLEIN'S ABSOLUTE INVARIANT, $\lambda(q)$ the ELLIPTIC LAMBDA FUNCTION

$$\lambda(q) \equiv k^2(q) = \left[\frac{\vartheta_2(q)}{\vartheta_3(q)}\right]^4, \tag{3}$$

and ϑ_i a THETA FUNCTION. This function can also be specified in terms of the WEBER FUNCTIONS f, f_1, f_2, γ_2, and γ_3 as

$$\begin{aligned} j(z) &= \frac{[f^{24}(z)-16]^3}{f^{24}(z)} && (4)\\ &= \frac{[{f_1}^{24}(z)+16]^3}{{f_1}^{24}(z)} && (5)\\ &= \frac{[{f_2}^{24}(z)+16]^3}{{f_2}^{24}(z)} && (6)\\ &= {\gamma_2}^3(z) && (7)\\ &= {\gamma_3}^2(z)+1728 && (8)\end{aligned}$$

(Weber 1902, p. 179; Atkin and Morain 1993).

The j-function is MEROMORPHIC function on the upper half of the COMPLEX PLANE which is invariant with respect to the SPECIAL LINEAR GROUP $SL(2,Z)$. It has a FOURIER SERIES

$$j(q) = \sum_{n=-\infty}^{\infty} a_n q^n, \tag{9}$$

for the NOME

$$q \equiv e^{2\pi i t} \tag{10}$$

with $\Im[t] > 0$. The coefficients in the expansion of the j-function satisfy:

1. $a_n = 0$ for $n < -1$ and $a_{-1} = 1$,
2. all a_ns are INTEGERS with fairly limited growth with respect to n, and
3. $j(q)$ is an ALGEBRAIC NUMBER, sometimes a RATIONAL NUMBER, and sometimes even an INTEGER at certain very special values of q (or t).

The latter result is the end result of the massive and beautiful theory of COMPLEX multiplication and the first step of Kronecker's so-called "JUGENDTRAUM."

Then all of the COEFFICIENTS in LAURENT SERIES

$$j(q) = \frac{1}{q} + 744 + 196884q + 21494760q^2$$
$$+864299970q^3 + 20245856256q^4 + 333202640600q^5 + \ldots \tag{11}$$

(Sloane's A000521) are POSITIVE INTEGERS (Rankin 1977). Let d be a POSITIVE SQUAREFREE INTEGER, and define

$$t \equiv \begin{cases} i\sqrt{d} & \text{for } d \equiv 1 \text{ or } 2 \pmod 4 \\ \frac{1}{2}(1+i\sqrt{d}) & \text{for } d \equiv 3 \pmod 4. \end{cases} \tag{12}$$

Then the NOME is

$$\begin{aligned} q \equiv e^{i\pi\tau} &= \begin{cases} e^{2\pi i(i\sqrt{d})} \\ e^{2\pi i(1+i\sqrt{d})/2} \end{cases} \\ &= \begin{cases} e^{-2\pi\sqrt{d}} & \text{for } d \equiv 1 \text{ or } 2 \pmod 4 \\ -e^{-\pi\sqrt{d}} & \text{for } d \equiv 3 \pmod 4. \end{cases} \end{aligned} \tag{13}$$

It then turns out that $j(q)$ is an ALGEBRAIC INTEGER of degree $h(-d)$, where $h(-d)$ is the CLASS NUMBER of the DISCRIMINANT $-d$ of the QUADRATIC FIELD $\mathbb{Q}(\sqrt{n})$ (Silverman 1986). The first term in the LAURENT SERIES is then $q^{-1} = e^{-2\pi\sqrt{n}}$ or $-e^{-\pi\sqrt{n}}$, and all the later terms are POWERS of q^{-1}, which are small numbers. The larger n, the faster the series converges. If $h(-d) = 1$, then $j(q)$ is a ALGEBRAIC INTEGER of degree 1, i.e., just a plain INTEGER. Furthermore, the INTEGER is a perfect CUBE.

The numbers whose LAURENT SERIES give INTEGERS are those with CLASS NUMBER 1. But these are precisely the HEEGNER NUMBERS -1, -2, -3, -7, -11, -19, -43, -67, -163. The greater (in ABSOLUTE VALUE) the HEEGNER NUMBER d, the closer to an INTEGER is the expression $e^{\pi\sqrt{-n}}$, since the initial term in $j(q)$ is the largest and subsequent terms are the smallest. The best approximations with $h(-d) = 1$ are therefore

$$\begin{aligned} e^{\pi\sqrt{43}} &\approx 960^3 + 744 - 2.2 \times 10^{-4} && (14)\\ e^{\pi\sqrt{67}} &\approx 5280^3 + 744 - 1.3 \times 10^{-6} && (15)\\ e^{\pi\sqrt{163}} &\approx 640320^3 + 744 - 7.5 \times 10^{-13}. && (16)\end{aligned}$$

The exact values of $j(q)$ corresponding to the HEEGNER NUMBERS are

$$j(-e^{-\pi}) = 12^3 \tag{17}$$
$$j(e^{-2\pi\sqrt{2}}) = 20^3 \tag{18}$$
$$j(-e^{-\pi\sqrt{3}}) = 0^3 \tag{19}$$
$$j(-e^{-\pi\sqrt{7}}) = -15^3 \tag{20}$$
$$j(-e^{-\pi\sqrt{11}}) = -32^3 \tag{21}$$
$$j(-e^{-\pi\sqrt{19}}) = -96^3 \tag{22}$$
$$j(-e^{-\pi\sqrt{43}}) = -960^3 \tag{23}$$
$$j(-e^{-\pi\sqrt{67}}) = -5280^3 \tag{24}$$
$$j(-e^{-\pi\sqrt{163}}) = -640320^3. \tag{25}$$

(The number 5280 is particularly interesting since it is also the number of feet in a mile.) The ALMOST INTEGER generated by the last of these, $e^{\pi\sqrt{163}}$ (corresponding to the field $\mathbb{Q}(\sqrt{-163})$ and the IMAGINARY quadratic field of maximal discriminant), is known as the RAMANUJAN CONSTANT.

$e^{\pi\sqrt{22}}$, $e^{\pi\sqrt{37}}$, and $e^{\pi\sqrt{58}}$ are also ALMOST INTEGERS. These correspond to binary quadratic forms with discriminants -88, -148, and -232, all of which have CLASS NUMBER two and were noted by Ramanujan (Berndt 1994).

It turns out that the j-function also is important in the CLASSIFICATION THEOREM for finite simple groups, and that the factors of the orders of the SPORADIC GROUPS, including the celebrated MONSTER GROUP, are also related.

see also ALMOST INTEGER, KLEIN'S ABSOLUTE INVARIANT, WEBER FUNCTIONS

References

Atkin, A. O. L. and Morain, F. "Elliptic Curves and Primality Proving." *Math. Comput.* **61**, 29–68, 1993.
Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 90–91, 1994.
Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, pp. 117–118, 1987.
Cohn, H. *Introduction to the Construction of Class Fields.* New York: Dover, p. 73, 1994.
Conway, J. H. and Guy, R. K. "The Nine Magic Discriminants." In *The Book of Numbers.* New York: Springer-Verlag, pp. 224–226, 1996.
Morain, F. "Implementation of the Atkin-Goldwasser-Kilian Primality Testing Algorithm." Rapport de Récherche 911, INRIA, Oct. 1988.
Rankin, R. A. *Modular Forms.* New York: Wiley, 1985.
Rankin, R. A. *Modular Forms and Functions.* Cambridge, England: Cambridge University Press, p. 199, 1977.
Serre, J. P. *Cours d'arithmétique.* Paris: Presses Universitaires de France, 1970.
Silverman, J. H. *The Arithmetic of Elliptic Curves.* New York: Springer-Verlag, p. 339, 1986.
Sloane, N. J. A. Sequence A000521/M5477 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Weber, H. *Lehrbuch der Algebra, Vols. I–II.* New York: Chelsea, 1979.
Weisstein, E. W. "j-Function." `http://www.astro.virginia.edu/~eww6n/math/notebooks/jFunction.m`.

j-Invariant

An invariant of an ELLIPTIC CURVE closely related to the DISCRIMINANT and defined by

$$j(E) \equiv \frac{2^8 3^3 a^3}{4a^3 + 27b^2}.$$

The determination of j as an ALGEBRAIC INTEGER in the QUADRATIC FIELD $\mathbb{Q}(j)$ is discussed by Greenhill (1891), Weber (1902), Berwick (1928), Watson (1938), Gross and Zaiger (1985), and Dorman (1988). The norm of j in $\mathbb{Q}(j)$ is the CUBE of an INTEGER in $\mathbb{Z}$.

see also DISCRIMINANT (ELLIPTIC CURVE), ELLIPTIC CURVE, FREY CURVE

References

Berwick, W. E. H. "Modular Invariants Expressible in Terms of Quadratic and Cubic Irrationalities." *Proc. London Math. Soc.* **28**, 53–69, 1928.
Dorman, D. R. "Special Values of the Elliptic Modular Function and Factorization Formulae." *J. reine angew. Math.* **383**, 207–220, 1988.
Greenhill, A. G. "Table of Complex Multiplication Moduli." *Proc. London Math. Soc.* **21**, 403–422, 1891.
Gross, B. H. and Zaiger, D. B. "On Singular Moduli." *J. reine angew. Math.* **355**, 191–220, 1985.
Watson, G. N. "Ramanujans Vermutung über Zerfällungsanzahlen." *J. reine angew. Math.* **179**, 97–128, 1938.
Weber, H. *Lehrbuch der Algebra, Vols. I–II.* New York: Chelsea, 1979.

Jackson's Difference Fan

If, after constructing a DIFFERENCE TABLE, no clear pattern emerges, turn the paper through an ANGLE of 60° and compute a new table. If necessary, repeat the process. Each ROTATION reduces POWERS by 1, so the sequence $\{k^n\}$ multiplied by any POLYNOMIAL in n is reduced to 0s by a k-fold difference fan.

References

Conway, J. H. and Guy, R. K. "Jackson's Difference Fans." In *The Book of Numbers.* New York: Springer-Verlag, pp. 84–85, 1996.

Jackson's Identity

A q-SERIES identity involving

$$\frac{(aq)_q^m (aqde)_q^m (adec)_q^m (aqcd)_q^m}{(aqc)_q^m (aqd)_q^m (aqe)_q^m (aqcde)_q^m},$$

where

$$(a)_q^n = (1-a)(1-aq)\cdots(1-aq^{n-1}).$$

see also q-SERIES

References

Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, pp. 109–110, 1959.
Jackson, F. H. "Summation of q-Hypergeometric Series." *Messenger Math.* **47**, 101–112, 1917.

Jackson's Theorem

Jackson's theorem is a statement about the error $E_n(f)$ of the best uniform approximation to a REAL FUNCTION $f(x)$ on $[-1,1]$ by REAL POLYNOMIALS of degree at most n. Let $f(x)$ be of bounded variation in $[-1,1]$ and let M' and V' denote the least upper bound of $|f(x)|$ and the total variation of $f(x)$ in $[-1,1]$, respectively. Given the function

$$F(x) = F(-1) + \int_{-1}^{x} f(x)\,dx, \tag{1}$$

then the coefficients

$$a_n = \tfrac{1}{2}(2n+1)\int_{-1}^{1} F(x)P_n(x)\,dx \tag{2}$$

of its LEGENDRE SERIES, where $P_n(x)$ is a LEGENDRE POLYNOMIAL, satisfy the inequalities

$$|a_n| < \begin{cases} \frac{6}{\sqrt{\pi}}(M'+V')n^{-3/2} & \text{for } n \geq 1 \\ \frac{4}{\sqrt{\pi}}(M'+V')n^{-3/2} & \text{for } n \geq 2. \end{cases} \tag{3}$$

Moreover, the LEGENDRE SERIES of $F(x)$ converges uniformly and absolutely to $F(x)$ in $[-1,1]$.

Bernstein strengthened Jackson's theorem to

$$2nE_{2n}(\alpha) \leq \frac{4n}{\pi(2n+1)} < \frac{2}{\pi} = 0.6366. \tag{4}$$

A specific application of Jackson's theorem shows that if

$$\alpha(x) = |x|, \tag{5}$$

then

$$E_n(\alpha) \leq \frac{6}{n}. \tag{6}$$

see also LEGENDRE SERIES, PICONE'S THEOREM

References

Cheney, E. W. *Introduction to Approximation Theory.* New York: McGraw-Hill, 1966.

Jackson, D. *The Theory of Approximation.* New York: Amer. Math. Soc., p. 76, 1930.

Rivlin, T. J. *An Introduction to the Approximation of Functions.* New York: Dover, 1981.

Sansone, G. *Orthogonal Functions, rev. English ed.* New York: Dover, pp. 205–208, 1991.

Jaco-Shalen-Johannson Torus Decomposition

Irreducible orientable COMPACT 3-MANIFOLDS have a canonical (up to ISOTOPY) minimal collection of disjointly EMBEDDED incompressible TORI such that each component of the 3-MANIFOLD removed by the TORI is either "atoroidal" or "Seifert-fibered."

Jacobi Algorithm

A method which can be used to solve a TRIDIAGONAL MATRIX equation with largest absolute values in each row and column dominated by the diagonal element. Each diagonal element is solved for, and an approximate value plugged in. The process is then iterated until it converges. This algorithm is a stripped-down version of the JACOBI METHOD of matrix diagonalization.

see also JACOBI METHOD, TRIDIAGONAL MATRIX

References

Acton, F. S. *Numerical Methods That Work, 2nd printing.* Washington, DC: Math. Assoc. Amer., pp. 161–163, 1990.

Jacobi-Anger Expansion

$$e^{iz\cos\theta} = \sum_{n=-\infty}^{\infty} i^n J_n(z)e^{in\theta},$$

where $J_n(z)$ is a BESSEL FUNCTION OF THE FIRST KIND. The identity can also be written

$$e^{iz\cos\theta} = J_0(z) + 2\sum_{n=1}^{\infty} i^n J_n(z)\cos(n\theta).$$

This expansion represents an expansion of plane waves into a series of cylindrical waves.

see also BESSEL FUNCTION OF THE FIRST KIND

Jacobi's Curvature Theorem

The principal normal indicatrix of a closed space curve with nonvanishing curvature bisects the AREA of the unit sphere if it is embedded.

Jacobi's Determinant Identity

Let

$$\mathsf{A} = \begin{bmatrix} \mathsf{B} & \mathsf{D} \\ \mathsf{E} & \mathsf{C} \end{bmatrix} \tag{1}$$

$$\mathsf{A}^{-1} = \begin{bmatrix} \mathsf{W} & \mathsf{X} \\ \mathsf{Y} & \mathsf{Z} \end{bmatrix}, \tag{2}$$

where B and W are $k \times k$ MATRICES. Then

$$(\det \mathsf{Z})(\det \mathsf{A}) = \det \mathsf{B}. \tag{3}$$

The proof follows from equating determinants on the two sides of the block matrices

$$\begin{bmatrix} \mathsf{B} & \mathsf{D} \\ \mathsf{E} & \mathsf{C} \end{bmatrix} \begin{bmatrix} \mathsf{I} & \mathsf{X} \\ \mathsf{0} & \mathsf{Z} \end{bmatrix} = \begin{bmatrix} \mathsf{B} & \mathsf{0} \\ \mathsf{E} & \mathsf{I} \end{bmatrix}, \tag{4}$$

where I is the IDENTITY MATRIX and $\mathsf{0}$ is the zero matrix.

References

Gantmacher, F. R. *The Theory of Matrices, Vol. 1.* New York: Chelsea, p. 21, 1960.

Horn, R. A. and Johnson, C. R. *Matrix Analysis.* Cambridge, England: Cambridge University Press, p. 21, 1985.

Jacobi Differential Equation

$$(1-x^2)y''+[\beta-\alpha-(\alpha+\beta+2)x]y'+n(n+\alpha+\beta+1)y=0 \qquad (1)$$

or

$$\frac{d}{dx}[(1-x)^{\alpha+1}(1+x)^{\beta+1}y'] + n(n+\alpha+\beta+1)(1-x)^{\alpha}(1+x)^{\beta}y=0. \qquad (2)$$

The solutions are JACOBI POLYNOMIALS. They can be transformed to

$$\frac{d^2u}{dx^2}+\left[\frac{1}{4}\frac{1-\alpha^2}{(1-x)^2}+\frac{1}{4}\frac{1-\beta^2}{(1+x)^2} + \frac{n(n+\alpha+\beta+1)+\frac{1}{2}(\alpha+1)(\beta+1)}{1-x^2}\right]u=0, \qquad (3)$$

where

$$u=u(x)=(1-x)^{(\alpha+1)/2}(1+x)^{(\beta+1)/2}P_n^{(\alpha,\beta)}(x), \qquad (4)$$

and

$$\frac{d^2u}{d\theta^2}+\left[\frac{\frac{1}{4}-\alpha^2}{4\sin^2(\frac{1}{2}\theta)}+\frac{\frac{1}{4}-\beta^2}{4\cos^2(\frac{1}{2}\theta)} + \left(n+\frac{\alpha+\beta+1}{2}\right)^2\right]u=0, \qquad (5)$$

where

$$u=u(\theta)=\sin^{\alpha+1/2}(\tfrac{1}{2}\theta)\cos^{\beta+1/2}(\tfrac{1}{2}\theta)P_n^{(\alpha,\beta)}(\cos\theta). \qquad (6)$$

Jacobi Differential Equation (Calculus of Variations)

$$\frac{d}{dx}\Omega_{\eta'}-\Omega_{\eta}=\frac{d}{dx}(f_{y'y}\eta+f_{y'y}\eta')-(f_{yy}\eta+f_{yy'}\eta')=0,$$

where

$$\Omega(x,\eta,\eta')\equiv\tfrac{1}{2}(f_{yy}\eta^2+2f_{yy'}\eta\eta'+f_{y'y}\eta'^2).$$

This equations arises in the CALCULUS OF VARIATIONS.

References

Bliss, G. A. *Calculus of Variations.* Chicago, IL: Open Court, pp. 162–163, 1925.

Jacobi Elliptic Functions

The Jacobi elliptic functions are standard forms of ELLIPTIC FUNCTIONS. The three basic functions are denoted $\mathrm{cn}(u,k)$, $\mathrm{dn}(u,k)$, and $\mathrm{sn}(u,k)$, where k is known as the MODULUS. In terms of THETA FUNCTIONS,

$$\mathrm{sn}(u,k)=\frac{\vartheta_3}{\vartheta_4}\frac{\vartheta_1(u\vartheta_3{}^{-2})}{\vartheta_4(u\vartheta_3{}^{-2})} \qquad (1)$$

$$\mathrm{cn}(u,k)=\frac{\vartheta_4}{\vartheta_2}\frac{\vartheta_2(u\vartheta_3{}^{-2})}{\vartheta_4(u\vartheta_3{}^{-2})} \qquad (2)$$

$$\mathrm{dn}(u,k)=\frac{\vartheta_4}{\vartheta_3}\frac{\vartheta_3(u\vartheta_3{}^{-2})}{\vartheta_4(u\vartheta_3{}^{-2})} \qquad (3)$$

(Whittaker and Watson 1990, p. 492), where $\vartheta_i\equiv\vartheta_i(0)$ (Whittaker and Watson 1990, p. 464). Ratios of Jacobi elliptic functions are denoted by combining the first letter of the NUMERATOR elliptic function with the first of the DENOMINATOR elliptic function. The multiplicative inverses of the elliptic functions are denoted by reversing the order of the two letters. These combinations give a total of 12 functions: cd, cn, cs, dc, dn, ds, nc, nd, ns, sc, sd, and sn. The AMPLITUDE ϕ is defined in terms of $\mathrm{sn}\,u$ by

$$y=\sin\phi=\mathrm{sn}(u,k). \qquad (4)$$

The k argument is often suppressed for brevity so, for example, $\mathrm{sn}(u,k)$ can be written $\mathrm{sn}\,u$.

The Jacobi elliptic functions are periodic in $K(k)$ and $K'(k)$ as

$$\mathrm{sn}(u+2mK+2niK',k)=(-1)^m\,\mathrm{sn}(u,k) \qquad (5)$$

$$\mathrm{cn}(u+2mK+2niK',k)=(-1)^{m+n}\,\mathrm{cn}(u,k) \qquad (6)$$

$$\mathrm{dn}(u+2mK+2niK',k)=(-1)^{n}\,\mathrm{dn}(u,k), \qquad (7)$$

where $K(k)$ is the complete ELLIPTIC INTEGRAL OF THE FIRST KIND, $K'(k)\equiv K(k')$, and $k'\equiv\sqrt{1-k^2}$ (Whittaker and Watson 1990, p. 503).

The $\mathrm{cn}\,x$, $\mathrm{dn}\,x$, and $\mathrm{sn}\,x$ functions may also be defined as solutions to the differential equations

$$\frac{d^2y}{dx^2}=-(1+k^2)y+2k^2y^3 \qquad (8)$$

$$\frac{d^2y}{dx^2}=-(1-2k^2)y-2k^2y^3 \qquad (9)$$

$$\frac{d^2y}{dx^2}=(2-k^2)y-2y^3. \qquad (10)$$

The standard Jacobi elliptic functions satisfy the identities

$$\mathrm{sn}^2\,u+\mathrm{cn}^2\,u=1 \qquad (11)$$

$$k^2\,\mathrm{sn}^2\,u+\mathrm{dn}^2\,u=1 \qquad (12)$$

$$k^2\,\mathrm{cn}^2\,u+k'^2=\mathrm{dn}^2\,u \qquad (13)$$

$$\mathrm{cn}^2\,u+k'^2\,\mathrm{sn}^2\,u=\mathrm{dn}^2\,u. \qquad (14)$$

Special values are

$$\operatorname{cn}(0) = 1 \tag{15}$$

$$\operatorname{dn}(0) = 1 \tag{16}$$

$$\operatorname{cn}(K) = 0 \tag{17}$$

$$\operatorname{dn}(K) = k' \equiv \sqrt{1-k^2}, \tag{18}$$

$$\operatorname{sn}(K) = 1, \tag{19}$$

where $K = K(k)$ is a complete ELLIPTIC INTEGRAL OF THE FIRST KIND and k' is the complementary MODULUS (Whittaker and Watson 1990, pp. 498–499).

In terms of integrals,

$$u = \int_0^{\operatorname{sn} u} (1-t^2)^{1-/2}(1-k^2t^2)^{-1/2}\,dt \tag{20}$$

$$= \int_{\operatorname{ns} u}^{\infty} (t^2-1)^{-1/2}(t^2-l^2)^{-1/2}\,dt \tag{21}$$

$$= \int_{\operatorname{cn} u}^{1} (1-t^2)^{-1/2}(k'^2+k^2t^2)^{-1/2}\,dt \tag{22}$$

$$= \int_1^{\operatorname{nc} u} (t^2-1)^{-1/2}(k'^2t^2+k^2)^{-1/2}\,dt \tag{23}$$

$$= \int_{\operatorname{dn} u}^{1} (1-t^2)^{-1/2}(t^2-k'^2)^{-1/2}\,dt \tag{24}$$

$$= \int_1^{\operatorname{nd} u} (t^2-1)^{-1/2}(1-k'^2t^2)^{-1/2}\,dt \tag{25}$$

$$= \int_0^{\mathrm{U}} (1+t^2)^{-1/2}(1+k'^2t^2)^{-1/2}\,dt \tag{26}$$

$$= \int_{\operatorname{cs} u}^{\infty} (t^2+1)^{-1/2}(t^2+k'^2)^{-1/2}\,dt \tag{27}$$

$$= \int_0^{\operatorname{sd} u} (1-k'^2t^2)^{-1/2}(1+k^2t^2)^{-1/2}\,dt \tag{28}$$

$$= \int_{\operatorname{ds} u}^{\infty} (t^2-k'^2)^{-1/2}(t^2+k^2)^{-1/2}\,dt \tag{29}$$

$$= \int_1^{\operatorname{cd} u} (1-t^2)^{-1/2}(1-k^2t^2)^{-1/2}\,dt \tag{30}$$

$$= \int_{\operatorname{dc} u}^{1} (t^2-1)^{-1/2}(t^2-k^2)^{-1/2}\,dt \tag{31}$$

(Whittaker and Watson 1990, p. 494).

Jacobi elliptic functions addition formulas include

$$\operatorname{sn}(u+v) = \frac{\operatorname{sn} u \operatorname{cn} v \operatorname{dn} v + \operatorname{sn} v \operatorname{cn} u \operatorname{dn} u}{1-k^2\operatorname{sn}^2 u \operatorname{sn}^2 v} \tag{32}$$

$$\operatorname{cn}(u+v) = \frac{\operatorname{cn} u \operatorname{cn} v - \operatorname{sn} u \operatorname{sn} v \operatorname{dn} u \operatorname{dn} v}{1-k^2\operatorname{sn}^2 u \operatorname{sn}^2 v} \tag{33}$$

$$\operatorname{dn}(u+v) = \frac{\operatorname{dn} u \operatorname{dn} v - k^2 \operatorname{sn} u \operatorname{sn} v \operatorname{cn} u \operatorname{cn} v}{1-k^2\operatorname{sn}^2 u \operatorname{sn}^2 v}. \tag{34}$$

Extended to integral periods,

$$\operatorname{sn}(u+K) = \frac{\operatorname{cn} u}{\operatorname{dn} u} \tag{35}$$

$$\operatorname{cn}(u+K) = -\frac{k' \operatorname{sn} u}{\operatorname{dn} u} \tag{36}$$

$$\operatorname{dn}(u+K) = \frac{k'}{\operatorname{dn} u} \tag{37}$$

$$\operatorname{sn}(u+2K) = -\operatorname{sn} u \tag{38}$$

$$\operatorname{cn}(u+2K) = -\operatorname{cn} u \tag{39}$$

$$\operatorname{dn}(u+2K) = \operatorname{dn} u \tag{40}$$

For COMPLEX arguments,

$$\operatorname{sn}(u+iv) = \frac{\operatorname{sn}(u,k)\operatorname{dn}(v,k')}{1-\operatorname{dn}^2(u,k)\operatorname{sn}^2(v,k')} + \frac{i\operatorname{cn}(u,k)\operatorname{dn}(u,k)\operatorname{sn}(v,k')\operatorname{cn}(v,k')}{1-\operatorname{dn}^2(u,k)\operatorname{sn}^2(v,k')} \tag{41}$$

$$\operatorname{cn}(u+iv) = \frac{\operatorname{cn}(u,k)\operatorname{cn}(v,k')}{1-\operatorname{dn}^2(u,k)\operatorname{sn}^2(v,k')} - \frac{i\operatorname{sn}(u,k)\operatorname{dn}(u,k)\operatorname{sn}(v,k')\operatorname{dn}(v,k')}{1-\operatorname{dn}^2(u,k)\operatorname{sn}^2(v,k')} \tag{42}$$

$$\operatorname{dn}(u+iv) = \frac{\operatorname{dn}(u,k)\operatorname{cn}(v,k')\operatorname{dn}(v,k')}{1-\operatorname{dn}^2(u,k)\operatorname{sn}^2(v,k')} - \frac{ik^2\operatorname{sn}(u,k)\operatorname{cn}(u,k)\operatorname{sn}(v,k')}{1-\operatorname{dn}^2(u,k)\operatorname{sn}^2(v,k')}. \tag{43}$$

DERIVATIVES of the Jacobi elliptic functions include

$$\frac{d\operatorname{sn} u}{du} = \operatorname{cn} u \operatorname{dn} u \tag{44}$$

$$\frac{d\operatorname{cn} u}{du} = \operatorname{sn} u \operatorname{dn} u \tag{45}$$

$$\frac{d\operatorname{dn} u}{du} = -k^2 \operatorname{sn} u \operatorname{cn} u. \tag{46}$$

Double-period formulas involving the Jacobi elliptic functions include

$$\operatorname{sn}(2u) = \frac{2\operatorname{sn} u \operatorname{cn} u \operatorname{dn} u}{1-k^2\operatorname{sn}^4 u} \tag{47}$$

$$\operatorname{cn}(2u) = \frac{1-2\operatorname{sn}^2 u + k^2\operatorname{sn}^4 u}{1-k^2\operatorname{sn}^4 u} \tag{48}$$

$$\operatorname{dn}(2u) = \frac{1-2k^2\operatorname{sn}^2 u + k^2\operatorname{sn}^4 u}{1-k^2\operatorname{sn}^4 u}. \tag{49}$$

Half-period formulas involving the Jacobi elliptic functions include

$$\operatorname{sn}(\tfrac{1}{2}K) = \frac{1}{\sqrt{1+k'}} \tag{50}$$

$$\operatorname{cn}(\tfrac{1}{2}K) = \sqrt{\frac{k'}{1+k'}} \tag{51}$$

$$\operatorname{dn}(\tfrac{1}{2}K) = \sqrt{k'}. \tag{52}$$

Squared formulas include

$$\operatorname{sn}^2 u = \frac{1-\operatorname{cn}(2u)}{1+\operatorname{dn}(2u)} \tag{53}$$

$$\operatorname{cn}^2 u = \frac{\operatorname{dn}(2u)+\operatorname{cn}(2u)}{1+\operatorname{dn}(2u)} \tag{54}$$

$$\operatorname{dn}^2 u = \frac{\operatorname{dn}(2u)+\operatorname{cn}(2u)}{1+\operatorname{cn}(2u)}. \tag{55}$$

see also AMPLITUDE, ELLIPTIC FUNCTION, JACOBI'S IMAGINARY TRANSFORMATION, THETA FUNCTION, WEIERSTRASS ELLIPTIC FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Jacobian Elliptic Functions and Theta Functions." Ch. 16 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 567–581, 1972.

Bellman, R. E. *A Brief Introduction to Theta Functions.* New York: Holt, Rinehart and Winston, 1961.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 433, 1953.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Elliptic Integrals and Jacobi Elliptic Functions." §6.11 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 254–263, 1992.

Spanier, J. and Oldham, K. B. "The Jacobian Elliptic Functions." Ch. 63 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 635–652, 1987.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Jacobi Function of the First Kind

see JACOBI POLYNOMIAL

Jacobi Function of the Second Kind

$$Q_n^{(\alpha,\beta)}(x) = 2^{-n-1}(x-1)^{-\alpha}(x+1)^{-\beta} \times \int_{-1}^{1} (1-t)^{n+\alpha}(1+t)^{n+\beta}(x-t)^{-n-1}\,dt.$$

In the exceptional case $n=0$, $\alpha+\beta+1=0$, a nonconstant solution is given by

$$Q^{(\alpha)}(x) = \ln(x+1) + \pi^{-1}\sin(\pi\alpha)(x-1)^{-\alpha}(x+1)^{-\beta} \times \int_{-1}^{1} \frac{(1-t)^{\alpha}(1+t)^{\beta}}{x-t}\ln(1+t)\,dt.$$

References

Szegő, G. "Jacobi Polynomials." Ch. 4 in *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., pp. 73–79, 1975.

Jacobi-Gauss Quadrature

Also called JACOBI QUADRATURE or MEHLER QUADRATURE. A GAUSSIAN QUADRATURE over the interval $[-1,1]$ with WEIGHTING FUNCTION $W(x)=(1-x)^{\alpha}(1+x)^{\beta}$. The ABSCISSAS for quadrature order n are given by the roots of the JACOBI POLYNOMIALS $P_n^{(\alpha,\beta)}(x)$. The weights are

$$\begin{aligned} w_i &= -\frac{A_{n+1}\gamma_n}{A_n P_n^{(\alpha,\beta)\prime}(x_i)P_{n+1}^{(\alpha,\beta)}(x_i)} \\ &= \frac{A_n}{A_{n-1}}\frac{\gamma_{n-1}}{P_{n-1}^{(\alpha,\beta)}(x_i)P_n^{(\alpha,\beta)\prime}(x_i)}, \end{aligned} \tag{1}$$

where A_n is the COEFFICIENT of x^n in $P_n^{(\alpha,\beta)}(x)$. For JACOBI POLYNOMIALS,

$$A_n = \frac{\Gamma(2n+\alpha+\beta+1)}{2^n n!\,\Gamma(n+\alpha+\beta+1)}, \tag{2}$$

where $\Gamma(z)$ is a GAMMA FUNCTION. Additionally,

$$\gamma_n = \frac{1}{2^{2n}(n!)^2}\frac{2^{2n+\alpha+\beta+1}n!}{2n+\alpha+\beta+1} \times \frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{\Gamma(n+\alpha+\beta+1)}, \tag{3}$$

so

$$w_i = \frac{2n+\alpha+\beta+2}{n+\alpha+\beta+1}\frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{\Gamma(n+\alpha+\beta+1)} \times \frac{2^{2n+\alpha+\beta+1}n!}{V_n'(x_i)V_{n+1}(x_i)} \tag{4}$$

$$= \frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{\Gamma(n+\alpha+\beta+1)}\frac{2^{2n+\alpha+\beta+1}n!}{(1-{x_i}^2)[V_n'(x_i)]^2}, \tag{5}$$

where

$$V_m \equiv P_n^{(\alpha,\beta)}(x)\frac{2^n n!}{(-1)^n}. \tag{6}$$

The error term is

$$E_n = \frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)\Gamma(n+\alpha+\beta+1)}{(2n+\alpha+\beta+1)[\Gamma(2n+\alpha+\beta+1)]^2} \times \frac{2^{2n+\alpha+\beta+1}n!}{(2n)!}f^{(2n)}(\xi) \tag{7}$$

(Hildebrand 1959).

References

Hildebrand, F. B. *Introduction to Numerical Analysis.* New York: McGraw-Hill, pp. 331–334, 1956.

Jacobi Identities

"The" Jacobi identity is a relationship

$$[A,[B,C]]+[B,[C,A]]+[C,[A,B]]=0, \qquad (1),$$

between three elements A, B, and C, where $[A,B]$ is the COMMUTATOR. The elements of a LIE GROUP satisfy this identity.

Relationships between the Q-FUNCTIONS Q_i are also known as Jacobi identities:

$$Q_1Q_2Q_3=1, \qquad (2)$$

equivalent to the JACOBI TRIPLE PRODUCT (Borwein and Borwein 1987, p. 65) and

$$Q_2{}^8=16qQ_1{}^8+Q_3{}^8, \qquad (3)$$

where

$$q\equiv e^{-\pi K'(k)/K(k)}, \qquad (4)$$

$K=K(k)$ is the complete ELLIPTIC INTEGRAL OF THE FIRST KIND, and $K'(k)=K(k')=K(\sqrt{1-k^2})$. Using WEBER FUNCTIONS

$$f_1=q^{-1/24}Q_3 \qquad (5)$$

$$f_2=2^{1/2}q^{1/12}Q_1 \qquad (6)$$

$$f=q^{-1/24}Q_2, \qquad (7)$$

(5) and (6) become

$$f_1f_2f=\sqrt{2} \qquad (8)$$

$$f^8=f_1{}^8+f_2{}^8 \qquad (9)$$

(Borwein and Borwein 1987, p. 69).

see also COMMUTATOR, JACOBI TRIPLE PRODUCT, Q-FUNCTION, WEBER FUNCTIONS

References

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, 1987.

Jacobi's Imaginary Transformation

For JACOBI ELLIPTIC FUNCTIONS $\operatorname{sn}u$, $\operatorname{cn}u$, and $\operatorname{dn}u$,

$$\operatorname{sn}(iu,k)=i\frac{\operatorname{sn}(u,k')}{\operatorname{cn}(u,k')}$$

$$\operatorname{cn}(iu,k)=\frac{1}{\operatorname{cn}(u,k')}$$

$$\operatorname{dn}(iu,k)=\frac{\operatorname{dn}(u,k')}{\operatorname{cn}(u,k')}$$

(Abramowitz and Stegun 1972, Whittaker and Watson 1990).

see also JACOBI ELLIPTIC FUNCTIONS

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 592 and 595, 1972.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, p. 505, 1990.

Jacobi Matrix

see JACOBI ROTATION MATRIX, JACOBIAN

Jacobi Method

A method of diagonalizing MATRICES using JACOBI ROTATION MATRICES. It consists of a sequence of ORTHOGONAL SIMILARITY TRANSFORMATIONS, each of which eliminates one off-diagonal element.

see also JACOBI ALGORITHM, JACOBI ROTATION MATRIX

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Jacobi Transformation of a Symmetric Matrix." §11.1 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 456–462, 1992.

Jacobi Polynomial

Also known as the HYPERGEOMETRIC POLYNOMIALS, they occur in the study of ROTATION GROUPS and in the solution to the equations of motion of the symmetric top. They are solutions to the JACOBI DIFFERENTIAL EQUATION. Plugging

$$y=\sum_{\nu=0}^{\infty}a_\nu(x-1)^\nu \qquad (1)$$

into the differential equation gives the RECURRENCE RELATION

$$[\gamma-\nu(\nu+\alpha+\beta+1)]a_\nu-2(\nu+1)(\nu+\alpha+1)a_{\nu+1}=0 \quad (2)$$

for $\nu=0, 1, \ldots,$ where

$$\gamma\equiv n(n+\alpha+\beta+1). \qquad (3)$$

Solving the RECURRENCE RELATION gives

$$P_n^{(\alpha,\beta)}(x)=\frac{(-1)^n}{2^n n!}(1-x)^{-\alpha}(1+x)^{-\beta} \times\frac{d^n}{dx^n}[(1-x)^{\alpha+n}(1+x)^{\beta+n}] \quad (4)$$

for $\alpha,\beta>-1$. They form a complete orthogonal system in the interval $[-1,1]$ with respect to the weighting function

$$w_n(x)=(1-x)^\alpha(1+x)^\beta, \qquad (5)$$

and are normalized according to

$$P_n^{(\alpha,\beta)}(1)=\binom{n+\alpha}{n}, \qquad (6)$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT. Jacobi polynomials can also be written

$$P_n^{(\alpha,\beta)}=\frac{\Gamma(2n+\alpha+\beta+1)}{n!\Gamma(n+\alpha+\beta+1)} \times G_n(\alpha+\beta+1,\beta+1,\tfrac{1}{2}(x+1)), \qquad (7)$$

where $\Gamma(z)$ is the GAMMA FUNCTION and

$$G_n(p,q,x) \equiv \frac{n!\Gamma(n+p)}{\Gamma(2n+p)} P_n^{(p-q,q-1)}(2x-1). \quad (8)$$

Jacobi polynomials are ORTHOGONAL satisfying

$$\int_{-1}^{1} P_m^{(\alpha,\beta)} P_n^{(\alpha,\beta)} (1-x)^\alpha (1+x)^\beta \, dx$$
$$= \frac{2^{\alpha+\beta+1}}{2n+\alpha+\beta+1} \frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{n!\Gamma(n+\alpha+\beta+1)} \delta_{mn}. \quad (9)$$

The COEFFICIENT of the term x^n in $P_n^{(\alpha,beta)}(x)$ is given by

$$A_n = \frac{\Gamma(2n+\alpha+\beta+1)}{2^n n!\Gamma(n+\alpha+\beta+1)}. \quad (10)$$

They satisfy the RECURRENCE RELATION

$$2(n+1)(n+\alpha+\beta+1)(2n+\alpha+\beta)P_{n+1}^{(\alpha,\beta)}(x)$$
$$= [(2n+\alpha+\beta+1)(\alpha^2-\beta^2) + (2n+\alpha+\beta)_3 x] P_n^{(\alpha,\beta)}(x)$$
$$-2(n+\alpha)(n+\beta)(2n+\alpha+\beta+2)P_{n-1}^{(\alpha,\beta)}(x), \quad (11)$$

where $(m)_n$ is the RISING FACTORIAL

$$(m)_n \equiv m(m+1)\cdots(m+n-1) = \frac{(m+n-1)!}{(m-1)!}. \quad (12)$$

The DERIVATIVE is given by

$$\frac{d}{dx}[P_n^{(\alpha,\beta)}(x)] = \tfrac{1}{2}(n+\alpha+\beta+1)P_{n-1}^{(\alpha+1,\beta+1)}(x). \quad (13)$$

The ORTHOGONAL POLYNOMIALS with WEIGHTING FUNCTION $(b-x)^\alpha (x-a)^\beta$ on the CLOSED INTERVAL $[a,b]$ can be expressed in the form

$$[\text{const.}] P_n^{(\alpha,\beta)} \left(2\frac{x-a}{b-a} - 1\right) \quad (14)$$

(Szegő 1975, p. 58).

Special cases with $\alpha = \beta$ are

$$P_{2\nu}^{(\alpha,\alpha)}(x) = \frac{\Gamma(2\nu+\alpha+1)\Gamma(\nu+1)}{\Gamma(\nu+\alpha+1)\Gamma(2\nu+1)} P_\nu^{(\alpha,-1/2)}(2x^2-1)$$
$$= (-1)^\nu \frac{\Gamma(2\nu+\alpha+1)\Gamma(\nu+1)}{\Gamma(\nu+\alpha+1)\Gamma(2\nu+1)} P_\nu^{(-1/2,\alpha)}(1-2x^2) \quad (15)$$

$$P_{2\nu+1}^{(\alpha,\alpha)}(x) = \frac{\Gamma(2\nu+\alpha+2)\Gamma(\nu+1)}{\Gamma(\nu+\alpha+1)\Gamma(2\nu+2)} x P_\nu^{(\alpha,1/2)}(2x^2-1)$$
$$= (-1)^\nu \frac{\Gamma(2\nu+\alpha+2)\Gamma(\nu+1)}{\Gamma(\nu+\alpha+1)\Gamma(2\nu+2)} x P_\nu^{(1/2,\alpha)}(1-2x^2). \quad (16)$$

Further identities are

$$P_n^{(\alpha+1,\beta)}(x) = \frac{2}{2n+\alpha+\beta+2}$$
$$\times \frac{(n+\alpha+1)P_n^{(\alpha,\beta)} - (n+1)P_{n+1}^{(\alpha,\beta)}(x)}{1-x} \quad (17)$$

$$P_n^{(\alpha,\beta+1)}(x) = \frac{2}{2n+\alpha+\beta+2}$$
$$\frac{(n+\beta+1)P_n^{(\alpha,\beta)}(x) + (n+1)P_{n+1}^{\alpha,\beta)}(x)}{1+x} \quad (18)$$

$$\sum_{\nu=0}^{n} \frac{2\nu+\alpha+\beta+1}{2^{\alpha+\beta+1}} \frac{\Gamma(\nu+1)\Gamma(\nu+\alpha+\beta+1)}{\Gamma(\nu+\alpha+1)\Gamma(\nu+\beta+1)}$$
$$\times P_\nu^{(\alpha,\beta)}(x) Q_\nu^{(\alpha,\beta)}(y)$$
$$= \frac{1}{2} \frac{(y-1)^{-\alpha}(y+1)^{-\beta}}{y-x} + \frac{2^{-\alpha-\beta}}{2n+\alpha+\beta+2}$$
$$\times \frac{\Gamma(n+2)\Gamma(n+\alpha+\beta+2)}{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}$$
$$\frac{P_{n+1}^{(\alpha,\beta)}(x)Q_n^{(\alpha,\beta)}(y) - P_n^{(\alpha,\beta)}(x)Q_{n+1}^{\alpha,\beta}(y)}{x-y} \quad (19)$$

(Szegő 1975, p. 79).

The KERNEL POLYNOMIAL is

$$K_n^{(\alpha,\beta)}(x,y) = \frac{2^{-\alpha-\beta}}{2n+\alpha+\beta+2}$$
$$\times \frac{\Gamma(n+2)\Gamma(n+\alpha+\beta+2)}{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}$$
$$\times \frac{P_{n+1}^{(\alpha,\beta)}(x)P_n^{(\alpha,\beta)}(y) - P_n^{(\alpha,\beta)}(x)P_{n+1}^{(\alpha,\beta)}(y)}{x-y} \quad (20)$$

(Szegő 1975, p. 71).

The DISCRIMINANT is

$$D_n^{(\alpha,\beta)} = 2^{-n(n-1)} \prod_{\nu=1}^{n} \nu^{\nu-2n+2}(\nu+\alpha)^{\nu-1}(\nu+\beta)^{\nu-1}$$
$$\times (n+\nu+\alpha+\beta)^{n-\nu} \quad (21)$$

(Szegő 1975, p. 143).

For $\alpha = \beta = 0$, $P_n^{(0,0)}(x)$ reduces to a LEGENDRE POLYNOMIAL. The GEGENBAUER POLYNOMIAL

$$G_n(p,q,x) = \frac{n!\Gamma(n+p)}{\Gamma(2n+p)} P_n^{(p-q,q-1)}(2x-1) \quad (22)$$

and CHEBYSHEV POLYNOMIAL OF THE FIRST KIND can also be viewed as special cases of the Jacobi POLYNOMIALS. In terms of the HYPERGEOMETRIC FUNCTION,

$$P_n^{(\alpha,\beta)}(x) = \binom{n+\alpha}{n}$$
$$\times {}_2F_1(-n, n+\alpha+\beta; \alpha+1; \tfrac{1}{2}(1-x)) \quad (23)$$

$$P_n^{(\alpha,\beta)}(x) = \binom{n+\alpha}{n} \left(\frac{x+1}{2}\right)^2$$
$$\times {}_2F_1\left(-n, -n-\beta; \alpha+1; \frac{x-1}{x+1}\right). \quad (24)$$

Let N_1 be the number of zeros in $x \in (-1,1)$, N_2 the number of zeros in $x \in (-\infty,-1)$, and N_3 the number of zeros in $x \in (1,\infty)$. Define Klein's symbol

$$E(u) = \begin{cases} 0 & \text{if } u \leq 0 \\ \lfloor u \rfloor & \text{if } u \text{ positive and nonintegral} \\ u-1 & \text{if } u = 1,\, 2,\, \ldots, \end{cases} \tag{25}$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION, and

$$X(\alpha,\beta) = E[\tfrac{1}{2}(|2n+\alpha+\beta+1| - |\alpha| - |\beta| + 1)] \tag{26}$$

$$Y(\alpha,\beta) = E[\tfrac{1}{2}(-|2n+\alpha+\beta+1| + |\alpha| - |\beta| + 1)] \tag{27}$$

$$Z(\alpha,\beta) = E[\tfrac{1}{2}(-|2n+\alpha+\beta+1| - |\alpha| + |\beta| + 1)]. \tag{28}$$

If the cases $\alpha = -1,\, -2,\, \ldots,\, -n$, $\beta = -1,\, -2,\, \ldots,\, -n$, and $n+\alpha+\beta = -1,\, -2,\, \ldots,\, -n$ are excluded, then the number of zeros of $P_n^{(\alpha,\beta)}$ in the respective intervals are

$$N_1(\alpha,\beta) = \begin{cases} 2\left\lfloor \frac{1}{2}(X+1) \right\rfloor & \text{for } (-1)^n \binom{n+\alpha}{n}\binom{n+\beta}{n} > 0 \\ 2\left\lfloor \frac{1}{2}X \right\rfloor + 1 & \text{for } (-1)^n \binom{n+\alpha}{n}\binom{n+\beta}{n} < 0 \end{cases} \tag{29}$$

$$N_2(\alpha,\beta) = \begin{cases} 2\left\lfloor \frac{1}{2}(Y+1) \right\rfloor & \text{for } \binom{2n+\alpha+\beta}{n}\binom{n+\beta}{n} > 0 \\ 2\left\lfloor \frac{1}{2}Y \right\rfloor + 1 & \text{for } \binom{2n+\alpha+\beta}{n}\binom{n+\beta}{n} < 0 \end{cases} \tag{30}$$

$$N_3(\alpha,\beta) = \begin{cases} 2\left\lfloor \frac{1}{2}(Z+1) \right\rfloor & \text{for } \binom{2n+\alpha+\beta}{n}\binom{n+\alpha}{n} > 0 \\ 2\left\lfloor \frac{1}{2}Z \right\rfloor + 1 & \text{for } \binom{2n+\alpha+\beta}{n}\binom{n+\alpha}{n} < 0 \end{cases} \tag{31}$$

(Szegő 1975, pp. 144–146).

The first few POLYNOMIALS are

$$\begin{aligned} P_0^{(\alpha,\beta)}(x) &= 1 \\ P_1^{(\alpha,\beta)}(x) &= \tfrac{1}{2}[2(\alpha+1) + (\alpha+\beta+2)(x-1)] \\ P_2^{(\alpha,\beta)}(x) &= \tfrac{1}{8}[4(\alpha+1)_2 + 4(\alpha+\beta+3)(\alpha+2)(x-1) \\ &\quad + (\alpha+\beta+2)_2(x-1)^2], \end{aligned}$$

where $(m)_n$ is a RISING FACTORIAL. See Abramowitz and Stegun (1972, pp. 782–793) and Szegő (1975, Ch. 4) for additional identities.

see also CHEBYSHEV POLYNOMIAL OF THE FIRST KIND, GEGENBAUER POLYNOMIAL, JACOBI FUNCTION OF THE SECOND KIND, RISING FACTORIAL, ZERNIKE POLYNOMIAL

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Orthogonal Polynomials." Ch. 22 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 771–802, 1972.

Iyanaga, S. and Kawada, Y. (Eds.). "Jacobi Polynomials." Appendix A, Table 20.V in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1480, 1980.

Szegő, G. "Jacobi Polynomials." Ch. 4 in *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., 1975.

Jacobi Quadrature

see JACOBI-GAUSS QUADRATURE

Jacobi Rotation Matrix

A MATRIX used in the JACOBI TRANSFORMATION method of diagonalizing MATRICES. It contains $\cos\phi$ in p rows and columns and $\sin\phi$ in q rows and columns,

$$P_{pq} \equiv \begin{bmatrix} 1 & & & & & & 0 \\ & \ddots & & \vdots & & \iddots & \\ & \cos\phi & \cdots & 0 & \cdots & \sin\phi & \\ \cdots & 0 & \cdots & 1 & \cdots & 0 & \cdots \\ & \sin\phi & \cdots & 0 & \cdots & \cos\phi & \\ & \iddots & & \vdots & & \ddots & \\ 0 & & & & & & 1 \end{bmatrix}.$$

see also JACOBI TRANSFORMATION

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Jacobi Transformation of a Symmetric Matrix." §11.1 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 456–462, 1992.

Jacobi Symbol

The product of LEGENDRE SYMBOLS (n/p_i) for each of the PRIME factors p_i such that $m = \prod_i p_i$, denoted (n/m). When m is a PRIME, the Jacobi symbol reduces to the LEGENDRE SYMBOL. The Jacobi symbol satisfies the same rules as the LEGENDRE SYMBOL

$$(n/m)(n/m') = (n/(mm')) \tag{1}$$

$$(n/m)(n'/m) = ((nn')/m) \tag{2}$$

$$(n^2/m) = (n/m^2) = 1 \qquad \text{if } (m,n) = 1 \tag{3}$$

$$(n/m) = (n'/m) \qquad \text{if } n \equiv n' \pmod{m} \tag{4}$$

$$(-1/m) = (-1)^{(m-1)/2} = \begin{cases} 1 & \text{for } m \equiv 1 \pmod{4} \\ -1 & \text{for } m \equiv -1 \pmod{4} \end{cases} \tag{5}$$

$$(2/m) = (-1)^{(m^2-1)/8} = \begin{cases} 1 & \text{for } m \equiv \pm 1 \pmod{8} \\ -1 & \text{for } m \equiv \pm 3 \pmod{8} \end{cases} \tag{6}$$

$$(n/m) = \begin{cases} (m/n) & \text{for } m \text{ or } n \equiv 1 \pmod{4} \\ -(m/n) & \text{for } m, n \equiv 3 \pmod{4}. \end{cases} \tag{7}$$

Written another way, for m and n RELATIVELY PRIME ODD INTEGERS with $n \geq 3$,

$$(m/n) = (-1)^{(m-1)(n-1)/4}(n/m). \tag{8}$$

The Jacobi symbol is not defined if $m \leq 0$ or m is EVEN.

Bach and Shallit (1996) show how to compute the Jacobi symbol in terms of the SIMPLE CONTINUED FRACTION of a RATIONAL NUMBER a/b.

see also KRONECKER SYMBOL

References

Bach, E. and Shallit, J. *Algorithmic Number Theory, Vol. 1: Efficient Algorithms.* Cambridge, MA: MIT Press, pp. 343–344, 1996.

Guy, R. K. "Quadratic Residues. Schur's Conjecture." §F5 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 244–245, 1994.

Riesel, H. "Jacobi's Symbol." *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, pp. 281–284, 1994.

Jacobi Tensor

$$J^{\mu}_{\nu\alpha\beta} = J^{\mu}_{\nu\beta\alpha} \equiv \tfrac{1}{2}(R^{\mu}_{\alpha\nu\beta} + R^{\mu}_{\beta\nu\alpha}),$$

where R is the RIEMANN TENSOR.

see also RIEMANN TENSOR

Jacobi's Theorem

Let M_r be an r-rowed MINOR of the nth order DETERMINANT $|\mathsf{A}|$ associated with an $n \times n$ MATRIX $\mathsf{A} = a_{ij}$ in which the rows $i_1, i_2, \ldots, i_r$ are represented with columns $k_1, k_2, \ldots, k_r$. Define the complementary minor to M_r as the $(n-k)$-rowed MINOR obtained from $|\mathsf{A}|$ by deleting all the rows and columns associated with M_r and the signed complementary minor $M^{(r)}$ to M_r to be

$$\begin{aligned} M^{(r)} = {} & (-1)^{i_1+i_2+\ldots+i_r+k_1+k_2+\ldots+k_r} \\ & \times [\text{complementary minor to} M_r]. \end{aligned}$$

Let the MATRIX of cofactors be given by

$$\Delta = \begin{vmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{vmatrix},$$

with M_r and M'_r the corresponding r-rowed minors of $|\mathsf{A}|$ and Δ, then it is true that

$$M'_r = |\mathsf{A}|^{r-1} M^{(r)}.$$

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 1109–1100, 1979.

Jacobi Theta Function

see THETA FUNCTION

Jacobi Transformation

see JACOBI METHOD

Jacobi Triple Product

The Jacobi triple product is the beautiful identity

$$\prod_{n=1}^{\infty}(1-x^{2n})(1+x^{2n-1}z^2)\left(1+\frac{x^{2n-1}}{z^2}\right) = \sum_{m=-\infty}^{\infty} x^{m^2} z^{2m}. \quad (1)$$

In terms of the Q-FUNCTIONS, (1) is written

$$Q_1 Q_2 Q_3 = 1, \quad (2)$$

which is one of the two JACOBI IDENTITIES. For the special case of $z = 1$, (1) becomes

$$\begin{aligned} \varphi(x) \equiv G(1) &= \prod_{n=1}^{\infty}(1+x^{2n-1})^2(1-x^{2n}) \\ &= \sum_{m=-\infty}^{\infty} x^{m^2} = 1 + 2\sum_{m=0}^{\infty} x^{m^2}, \end{aligned} \quad (3)$$

where $\varphi(x)$ is the one-variable RAMANUJAN THETA FUNCTION.

To prove the identity, define the function

$$\begin{aligned} F(z) &\equiv \prod_{n=1}^{\infty}(1+x^{2n-1}z^2)\left(1+\frac{x^{2n-1}}{z^2}\right) \\ &= (1+xz^2)\left(1+\frac{x}{z^2}\right)(1+x^3z^2)\left(1+\frac{x^3}{z^2}\right) \\ &\quad \times (1+x^5z^2)\left(1+\frac{x^5}{z^2}\right)\cdots. \end{aligned} \quad (4)$$

Then

$$\begin{aligned} F(xz) = {} & (1+x^3z^2)\left(1+\frac{1}{xz^2}\right)(1+x^5z^2)\left(1+\frac{x}{z^2}\right) \\ & (1+x^7z^2)\left(1+\frac{x^3}{z^2}\right)\cdots. \end{aligned} \quad (5)$$

Taking (5) ÷ (4),

$$\begin{aligned} \frac{F(xz)}{F(z)} &= \left(1+\frac{1}{xz^2}\right)\left(\frac{1}{1+xz^2}\right) \\ &= \frac{xz^2+1}{xz^2}\frac{1}{1+xz^2} = \frac{1}{xz^2}, \end{aligned} \quad (6)$$

which yields the fundamental relation

$$xz^2 F(xz) = F(z). \quad (7)$$

Now define

$$G(z) \equiv F(z)\prod_{n=1}^{\infty}(1-x^{2n}) \quad (8)$$

$$G(xz) = F(xz)\prod_{n=1}^{\infty}(1-x^{2n}). \tag{9}$$

Using (7), (9) becomes

$$G(xz) = \frac{F(z)}{xz^2}\prod_{n=1}^{\infty}(1-x^{2n}) = \frac{G(z)}{xz^2}, \tag{10}$$

so

$$G(z) = xz^2 G(xz). \tag{11}$$

Expand G in a LAURENT SERIES. Since G is an EVEN FUNCTION, the LAURENT SERIES contains only even terms.

$$G(z) = \sum_{m=-\infty}^{\infty} a_m z^{2m}. \tag{12}$$

Equation (11) then requires that

$$\begin{aligned}\sum_{m=-\infty}^{\infty} a_m z^{2m} &= xz^2 \sum_{m=-\infty}^{\infty} a_m (xz)^{2m}\\ &= \sum_{m=-\infty}^{\infty} a_m x^{2m+1} z^{2m+2}. \end{aligned}\tag{13}$$

This can be re-indexed with $m' \equiv m-1$ on the left side of (13)

$$\sum_{m=-\infty}^{\infty} a_m z^{2m} = \sum_{m=-\infty}^{\infty} a_m x^{2m-1} z^{2m}, \tag{14}$$

which provides a RECURRENCE RELATION

$$a_m = a_{m-1} x^{2m-1}, \tag{15}$$

so

$$a_1 = a_0 x \tag{16}$$

$$a_2 = a_1 x^3 = a_0 x^{3+1} = a_0 x^4 = a_0 x^{2^2} \tag{17}$$

$$a_3 = a_2 x^5 = a_0 x^{5+4} = a_0 x^9 = a_0 x^{3^2}. \tag{18}$$

The exponent grows greater by $(2m-1)$ for each increase in m of 1. It is given by

$$\sum_{n=1}^{m}(2m-1) = 2\frac{m(m+1)}{2} - m = m^2. \tag{19}$$

Therefore,

$$a_m = a_0 x^{m^2}. \tag{20}$$

This means that

$$G(z) = a_0 \sum_{m=-\infty}^{\infty} x^{m^2} z^{2m}. \tag{21}$$

The COEFFICIENT a_0 must be determined by going back to (4) and (8) and letting $z=1$. Then

$$\begin{aligned}F(1) &= \prod_{n=1}^{\infty}(1+x^{2n-1})(1+x^{2n-1})\\ &= \prod_{n=1}^{\infty}(1+x^{2n-1})^2 \end{aligned}\tag{22}$$

$$\begin{aligned}G(1) &= F(1)\prod_{n=1}^{\infty}(1-x^{2n})\\ &= \prod_{n=1}^{\infty}(1+x^{2n-1})^2 \prod_{n=1}^{\infty}(1-x^{2n})\\ &= \prod_{n=1}^{\infty}(1+x^{2n-1})^2(1-x^{2n}), \end{aligned}\tag{23}$$

since multiplication is ASSOCIATIVE. It is clear from this expression that the a_0 term must be 1, because all other terms will contain higher POWERS of x. Therefore,

$$a_0 = 1, \tag{24}$$

so we have the Jacobi triple product,

$$\begin{aligned}G(z) &= \prod_{n=1}^{\infty}(1-x^{2n})(1+x^{2n-1}z^2)\left(1+\frac{x^{2n-1}}{z^2}\right)\\ &= \sum_{m=-\infty}^{\infty} x^{m^2} z^{2m}. \end{aligned}\tag{25}$$

see also EULER IDENTITY, JACOBI IDENTITIES, Q-FUNCTION, QUINTUPLE PRODUCT IDENTITY, RAMANUJAN PSI SUM, RAMANUJAN THETA FUNCTIONS, SCHRÖTER'S FORMULA, THETA FUNCTION

References

Andrews, G. E. *q-Series: Their Development and Application in Analysis, Number Theory, Combinatorics, Physics, and Computer Algebra.* Providence, RI: Amer. Math. Soc., pp. 63–64, 1986.

Borwein, J. M. and Borwein, P. B. "Jacobi's Triple Product and Some Number Theoretic Applications." Ch. 3 in *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, pp. 62–101, 1987.

Jacobi, C. G. J. *Fundamentia Nova Theoriae Functionum Ellipticarum.* Regiomonti, Sumtibus fratrum Borntraeger, p. 90, 1829.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, p. 470, 1990.

Jacobi Zeta Function

Denoted $\mathrm{zn}(u,k)$ or $Z(u)$.

$$Z(\phi|m) \equiv E(\phi|m) - \frac{E(m)F(\phi|m)}{K(m)},$$

where ϕ is the AMPLITUDE, m is the PARAMETER, and F and K are ELLIPTIC INTEGRALS OF THE FIRST KIND, and E is an ELLIPTIC INTEGRAL OF THE SECOND KIND. See Gradshteyn and Ryzhik (1980, p. xxxi) for expressions in terms of THETA FUNCTIONS.

see also ZETA FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 595, 1972.

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, 1979.

Jacobian

Given a set $\mathbf{y} = \mathbf{f}(\mathbf{x})$ of n equations in n variables x_1, ..., x_n, written explicitly as

$$\mathbf{y} \equiv \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix}, \tag{1}$$

or more explicitly as

$$\begin{cases} y_1 = f_1(x_1, \ldots, x_n) \\ \vdots \\ y_n = f_n(x_1, \ldots, x_n), \end{cases} \tag{2}$$

the Jacobian matrix, sometimes simply called "the Jacobian" (Simon and Blume 1994) is defined by

$$\mathsf{J}(x_1, x_2, x_3) = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_n} \end{bmatrix}. \tag{3}$$

The DETERMINANT of J is the JACOBIAN DETERMINANT (confusingly, often called "the Jacobian" as well) and is denoted

$$J = \left| \frac{\partial(y_1, \ldots, y_n)}{\partial(x_1, \ldots, x_n)} \right|. \tag{4}$$

Taking the differential

$$d\mathbf{y} = \mathbf{y}_{\mathbf{x}}\, d\mathbf{x} \tag{5}$$

shows that J is the DETERMINANT of the MATRIX $\mathbf{y}_{\mathbf{x}}$, and therefore gives the ratios of n-D volumes (CONTENTS) in y and x,

$$dy_1 \cdots dy_n = \left| \frac{\partial(y_1, \ldots, y_n)}{\partial(x_1, \ldots, x_n)} \right| dx_1 \cdots dx_n. \tag{6}$$

The concept of the Jacobian can also be applied to n functions in more than n variables. For example, considering $f(u, v, w)$ and $g(u, v, w)$, the Jacobians

$$\frac{\partial(f, g)}{\partial(u, v)} = \begin{vmatrix} f_u & f_v \\ g_u & g_v \end{vmatrix} \tag{7}$$

$$\frac{\partial(f, g)}{\partial(u, w)} = \begin{vmatrix} f_u & f_w \\ g_u & g_w \end{vmatrix} \tag{8}$$

can be defined (Kaplan 1984, p. 99).

For the case of $n = 3$ variables, the Jacobian takes the special form

$$Jf(x_1, x_2, x_3) \equiv \left| \frac{\partial \mathbf{y}}{\partial x_1} \cdot \frac{\partial \mathbf{y}}{\partial x_2} \times \frac{\partial \mathbf{y}}{\partial x_3} \right|, \tag{9}$$

where $\mathbf{a} \cdot \mathbf{b}$ is the DOT PRODUCT and $\mathbf{b} \times \mathbf{c}$ is the CROSS PRODUCT, which can be expanded to give

$$\left| \frac{\partial(y_1, y_2, y_3)}{\partial(x_1, x_2, x_3)} \right| = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \frac{\partial y_1}{\partial x_3} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \frac{\partial y_2}{\partial x_3} \\ \frac{\partial y_3}{\partial x_1} & \frac{\partial y_3}{\partial x_2} & \frac{\partial y_3}{\partial x_3} \end{vmatrix}. \tag{10}$$

see also CHANGE OF VARIABLES THEOREM, CURVILINEAR COORDINATES, IMPLICIT FUNCTION THEOREM

References

Kaplan, W. *Advanced Calculus, 3rd ed.* Reading, MA: Addison-Wesley, pp. 98–99, 123, and 238–245, 1984.

Simon, C. P. and Blume, L. E. *Mathematics for Economists.* New York: W. W. Norton, 1994.

Jacobian Conjecture

If $\det[F'(x)] = 1$ for a POLYNOMIAL mapping F (where det is the DETERMINANT), then F is BIJECTIVE with POLYNOMIAL inverse.

Jacobian Curve

The Jacobian of a linear net of curves of order n is a curve of order $3(n - 1)$. It passes through all points common to all curves of the net. It is the LOCUS of points where the curves of the net touch one another and of singular points of the curve.

see also CAYLEYIAN CURVE, HESSIAN COVARIANT, STEINERIAN CURVE

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 149, 1959.

Jacobian Determinant

see JACOBIAN

Jacobian Group

The Jacobian group of a 1-D linear series is given by intersections of the base curve with the JACOBIAN CURVE of itself and two curves cutting the series.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 283, 1959.

Jacobsthal-Lucas Number

see JACOBSTHAL NUMBER

Jacobsthal-Lucas Polynomial

see JACOBSTHAL POLYNOMIAL

Jacobsthal Number

The Jacobsthal numbers are the numbers obtained by the U_ns in the LUCAS SEQUENCE with $P = 1$ and $Q = -2$, corresponding to $a = 2$ and $b = -1$. They and the Jacobsthal-Lucas numbers (the V_ns) satisfy the RECURRENCE RELATION

$$J_n = J_{n-1} + 2J_{n-2}. \qquad (1)$$

The Jacobsthal numbers satisfy $J_0 = 0$ and $J_1 = 1$ and are 0, 1, 1, 3, 5, 11, 21, 43, 85, 171, 341, ... (Sloane's A001045). The Jacobsthal-Lucas numbers satisfy $j_0 = 2$ and $j_1 = 1$ and are 2, 1, 5, 7, 17, 31, 65, 127, 257, 511, 1025, ... (Sloane's A014551). The properties of these numbers are summarized in Horadam (1996). They are given by the closed form expressions

$$J_n = \sum_{r=0}^{\lfloor (n-1)/2 \rfloor} \binom{n-1-r}{r} 2^r \qquad (2)$$

$$j_n = \sum_{r=0}^{\lfloor n/2 \rfloor} \frac{n}{n-r} \binom{n-r}{r} 2^r, \qquad (3)$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION and $\binom{n}{k}$ is a BINOMIAL COEFFICIENT. The Binet forms are

$$J_n = \tfrac{1}{3}(a^n - b^n) = \tfrac{1}{3}[2^n - (-1)^n] \qquad (4)$$

$$j_n = a^n + b^n = 2^n + (-1)^n. \qquad (5)$$

The GENERATING FUNCTIONS are

$$\sum_{i=1}^{\infty} J_i x^{i-1} = (1 - x - 2x^2)^{-1} \qquad (6)$$

$$\sum_{i=1}^{\infty} j_i x^{i-1} = (1 + 4x)(1 - x - 2x^2)^{-1}. \qquad (7)$$

The Simson FORMULAS are

$$J_{n+1}J_{n-1} - J_n{}^2 = (-1)^n 2^{n-1} \qquad (8)$$

$$j_{n+1}j_{n-1} - j_n{}^2 = 9(-1)^{n-1}2^{n-1} = -9(J_{n+1}J_{n-1} - J_n{}^2). \qquad (9)$$

Summation FORMULAS include

$$\sum_{i=2}^{n} J_i = \tfrac{1}{2}(J_{n+2} - 3) \qquad (10)$$

$$\sum_{i=1}^{n} j_i = \tfrac{1}{2}(j_{n+2} - 5). \qquad (11)$$

Interrelationships are

$$j_n J_n = J_{2n} \qquad (12)$$

$$j_n = J_{n+1} + 2J_{n-1} \qquad (13)$$

$$9J_n = j_{n+1} + 2j_{n-1} \qquad (14)$$

$$j_{n+1} + j_n = 3(J_{n+1} + J_n) = 3 \cdot 2^n \qquad (15)$$

$$j_{n+1} - j_n = 3(J_{n+1} - J_n) + 4(-1)^{n+1} = 2^n + 2(-1)^{n+1} \qquad (16)$$

$$j_{n+1} - 2j_n = 3(2J_n - J_{n+1}) = 3(-1)^{n+1} \qquad (17)$$

$$2j_{n+1} + j_{n-1} = 3(2J_{n+1} + J_{n-1}) + 6(-1)^{n+1} \qquad (18)$$

$$j_{n+r} + j_{n-r} = 3(J_{n+r} + J_{n-r}) + 4(-1)^{n-r} \qquad (19)$$

$$= 2^{n-r}(2^{2r} + 1) + 2(-1)^{n-r} \qquad (20)$$

$$j_{n+r} - j_{n-r} = 3(J_{n+r} - J_{n-r}) = 2^{n-r}(2^{2r} - 1) \qquad (21)$$

$$j_n = 3J_n + 2(-1)^n \qquad (22)$$

$$3J_n + j_n = 2^{n+1} \qquad (23)$$

$$J_n + j_n = 2J_{n+1} \qquad (24)$$

$$j_{n+2}j_{n-2} - j_n{}^2 = -9(J_{n+2}J_{n-2} - J_n)^2 = 9(-1)^n 2^{n-2} \qquad (25)$$

$$J_m j_n + J_n j_m = 2J_{m+n} \qquad (26)$$

$$j_m j_n + 9J_m J_n = 2j_{m+n} \qquad (27)$$

$$j_n{}^2 + 9J_n{}^2 = 2j_{2n} \qquad (28)$$

$$J_m j_n - J_n j_m = (-1)^n 2^{n+1} J_{m-n} \qquad (29)$$

$$j_m j_n - 9J_m J_n = (-1)^n 2^{n+1} j_{m-n} \qquad (30)$$

$$j_n{}^2 - 9J_n{}^2 = (-1)^n 2^{n+2} \qquad (31)$$

(Horadam 1996).

References

Horadam, A. F. "Jacobsthal and Pell Curves." *Fib. Quart.* **26**, 79–83, 1988.

Horadam, A. F. "Jacobsthal Representation Numbers." *Fib. Quart.* **34**, 40–54, 1996.

Sloane, N. J. A. Sequences A014551 and A001045/M2482 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Jacobsthal Polynomial

The Jacobsthal polynomials are the POLYNOMIALS obtained by setting $p(x) = 1$ and $q(x) = 2x$ in the LUCAS POLYNOMIAL SEQUENCE. The first few Jacobsthal polynomials are

$$\begin{aligned} J_1(x) &= 1 \\ J_2(x) &= 1 \\ J_3(x) &= 1 + 2x \\ J_4(x) &= 1 + 4x \\ J_5(x) &= 4x^2 + 6x + 1, \end{aligned}$$

and the first few Jacobsthal-Lucas polynomials are

$$
\begin{aligned}
j_1(x) &= 1\\
j_2(x) &= 4x+1\\
j_3(x) &= 6x+1\\
j_4(x) &= 8x^2+8x+1\\
j_5(x) &= 20x^2+10x+1.
\end{aligned}
$$

Jacobsthal and Jacobsthal-Lucas polynomials satisfy

$$J_n(1) = J_n$$

$$j_n(1) = j_n$$

where J_n is a Jacobsthal Number and j_n is a Jacobsthal-Lucas Number.

Janko Groups

The Sporadic Groups J_1, J_2, J_3 and J_4. The Janko group J_2 is also known as the Hall-Janko Group.

see also Sporadic Group

References

Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas/J1.html`, `J2.html`, `J3.html`, and `J4.html`.

Japanese Triangulation Theorem

Let a convex Cyclic Polygon be Triangulated in any manner, and draw the Incircle to each Triangle so constructed. Then the sum of the Inradii is a constant independent of the Triangulation chosen. This theorem can be proved using Carnot's Theorem. It is also true that if the sum of Inradii does not depend on the Triangulation of a Polygon, then the Polygon is Cyclic.

see also Carnot's Theorem, Cyclic Polygon, Incircle, Inradius, Triangulation

References

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 24–26, 1985.

Lambert, T. "The Delaunay Triangulation Maximizes the Mean Inradius." *Proc. Sixth Canadian Conf. Comput. Geometry.* Saskatoon, Saskatchewan, Canada, pp. 201–206, August 1994.

Jarnick's Inequality

Given a Convex plane region with Area A and Perimeter p, then

$$|N - A| < p,$$

where N is the number of enclosed Lattice Points.

see also Lattice Point, Nosarzewska's Inequality

Jeep Problem

Maximize the distance a jeep can penetrate into the desert using a given quantity of fuel. The jeep is allowed to go forward, unload some fuel, and then return to its base using the fuel remaining in its tank. At its base, it may refuel and set out again. When it reaches fuel it has previously stored, it may then use it to partially fill its tank. This problem is also called the Exploration Problem (Ball and Coxeter 1987).

Given $n + f$ (with $0 \leq f < 1$) drums of fuel at the edge of the desert and a jeep capable of holding one drum (and storing fuel in containers along the way), the maximum one-way distance which can be traveled (assuming the jeep travels one unit of distance per drum of fuel expended) is

$$
\begin{aligned}
d &= \frac{f}{2n+1} + \sum_{i=1}^{n} \frac{1}{2i-1}\\
&= \frac{f}{2n+1} + \tfrac{1}{2}[\gamma + 2\ln 2 + \psi_0(\tfrac{1}{2}+n)],
\end{aligned}
$$

where γ is the Euler-Mascheroni Constant and $\psi_n(z)$ the Polygamma Function.

For example, the farthest a jeep with $n = 1$ drum can travel is obviously 1 unit. However, with $n = 2$ drums of gas, the maximum distance is achieved by filling up the jeep's tank with the first drum, traveling 1/3 of a unit, storing 1/3 of a drum of fuel there, and then returning to base with the remaining 1/3 of a tank. At the base, the tank is filled with the second drum. The jeep then travels 1/3 of a unit (expending 1/3 of a drum of fuel), refills the tank using the 1/3 of a drum of fuel stored there, and continues an additional 1 unit of distance on a full tank, giving a total distance of 4/3. The solutions for $n = 1, 2, \ldots$ drums are 1, 4/3, 23/15, 176/105, 563/315, ..., which can also be written as $a(n)/b(n)$, where

$$
\begin{aligned}
a(n) &= \left(\frac{1}{1} + \frac{1}{3} + \ldots + \frac{1}{2n-1}\right) \mathrm{LCM}(1,3,5,\ldots,2n-1)\\
b(n) &= \mathrm{LCM}(1,3,5,\ldots,2n-1)
\end{aligned}
$$

(Sloane's A025550 and A025547).

see also Harmonic Number

References

Alway, G. C. "Crossing the Desert." *Math. Gaz.* **41**, 209, 1957.

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 32, 1987.

Bellman, R. Exercises 54–55 *Dynamic Programming.* Princeton, NJ: Princeton University Press, p. 103, 1955.

Fine, N. J. "The Jeep Problem." *Amer. Math. Monthly* **54**, 24–31, 1947.

Gale, D. "The Jeep Once More or Jeeper by the Dozen." *Amer. Math. Monthly* **77**, 493–501, 1970.

Gardner, M. *The Second Scientific American Book of Mathematical Puzzles & Diversions: A New Selection.* New York: Simon and Schuster, pp. 152 and 157–159, 1961.

Haurath, A.; Jackson, B.; Mitchem, J.; and Schmeichel, E. "Gale's Round-Trip Jeep Problem." *Amer. Math. Monthly* **102**, 299–309, 1995.
Helmer, O. "A Problem in Logistics: The Jeep Problem." Project Rand Report No. Ra 15015, Dec. 1947.
Phipps, C. G. "The Jeep Problem, A More General Solution." *Amer. Math. Monthly* **54**, 458–462, 1947.

Jenkins-Traub Method

A complicated POLYNOMIAL ROOT-finding algorithm which is used in the *IMSL*® (IMSL, Houston, TX) library and which Press *et al.* (1992) describe as "practically a standard in black-box POLYNOMIAL ROOT-finders."

References
IMSL, Inc. *IMSL Math/Library User's Manual.* Houston, TX: IMSL, Inc.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 369, 1992.
Ralston, A. and Rabinowitz, P. §8.9–8.13 in *A First Course in Numerical Analysis, 2nd ed.* New York: McGraw-Hill, 1978.

Jensen's Formula

$$\int_0^{2\pi} \ln|z+e^{i\theta}|\,d\theta = 2\pi \ln^+|z|,$$

where

$$\ln^+ \equiv \max(0, \ln x)$$

and $\ln x$ is the NATURAL LOGARITHM.

Jensen's Inequality

For a REAL CONTINUOUS CONCAVE FUNCTION

$$\frac{\sum f(x_i)}{n} \leq f\left(\frac{\sum x_i}{n}\right)$$

if f is concave down,

$$\frac{\sum f(x_i)}{n} \geq f\left(\frac{\sum x_i}{n}\right)$$

if f is concave up, and

$$\frac{\sum f(x_i)}{n} = f\left(\frac{\sum x_i}{n}\right)$$

IFF $x_1 = x_2 = \ldots = x_n$. A special case is

$$\sqrt[n]{x_1 x_2 \cdots x_n} \leq \frac{x_1 + x_2 + \ldots + x_n}{n},$$

with equality IFF $x_1 = x_2 = \ldots = x_n$.

see also CONCAVE FUNCTION, MAHLER'S MEASURE

References
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1101, 1979.

Jensen Polynomial

Let $f(x)$ be a real ENTIRE FUNCTION of the form

$$f(x) = \sum_{k=0}^{\infty} \gamma_k \frac{x^k}{k!},$$

where the γ_ks are POSITIVE and satisfy TURÁN'S INEQUALITIES

$$\gamma_k{}^2 - \gamma_{k-1}\gamma_{k+1} \geq 0$$

for $k = 1, 2, \ldots$. The Jensen polynomial $g(t)$ associated with $f(x)$ is then given by

$$g_n(t) = \sum_{k=0}^{n} \binom{n}{k} \gamma_k t^k,$$

where $\binom{a}{b}$ is a BINOMIAL COEFFICIENT.

References
Csordas, G.; Varga, R. S.; and Vincze, I. "Jensen Polynomials with Applications to the Riemann ζ-Function." *J. Math. Anal. Appl.* **153**, 112–135, 1990.

Jerabek's Hyperbola

The ISOGONAL CONJUGATE of the EULER LINE. It passes through the the vertices of a TRIANGLE, the ORTHOCENTER, CIRCUMCENTER, the LEMOINE POINT, and the ISOGONAL CONJUGATE points of the NINE-POINT CENTER and DE LONGCHAMPS POINT.

see also CIRCUMCENTER, DE LONGCHAMPS POINT, EULER LINE, ISOGONAL CONJUGATE, LEMOINE POINT, NINE-POINT CENTER, ORTHOCENTER

References
Casey, J. *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions with Numerous Examples, 2nd rev. enl. ed.* Dublin: Hodges, Figgis, & Co., 1893.
Pinkernell, G. M. "Cubic Curves in the Triangle Plane." *J. Geom.* **55**, 141–161, 1996.
Vandeghen, A. "Some Remarks on the Isogonal and Cevian Transforms. Alignments of Remarkable Points of a Triangle." *Amer. Math. Monthly* **72**, 1091–1094, 1965.

Jerk

The jerk $\mathbf{j}$ is defined as the time DERIVATIVE of the VECTOR ACCELERATION $\mathbf{a}$,

$$\mathbf{j} \equiv \frac{d\mathbf{a}}{dt}.$$

see also ACCELERATION, VELOCITY

Jinc Function

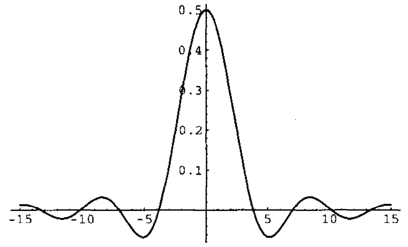

The jinc function is defined as

$$\operatorname{jinc}(x) \equiv \frac{J_1(x)}{x},$$

where $J_1(x)$ is a Bessel Function of the First Kind, and satisfies $\lim_{x\to 0} \operatorname{jinc}(x) = 1/2$. The Derivative of the jinc function is given by

$$\operatorname{jinc}'(x) = -\frac{J_2(x)}{x}.$$

The function is sometimes normalized by multiplying by a factor of 2 so that $\operatorname{jinc}(0) = 1$ (Siegman 1986, p. 729).

see also Bessel Function of the First Kind, Sinc Function

References
Siegman, A. E. *Lasers.* Sausalito, CA: University Science Books, 1986.

Jitter

A Sampling phenomenon produced when a waveform is not sampled uniformly at an interval t each time, but rather at a series of slightly shifted intervals $t+\Delta t_i$ such that the average $\langle \Delta t_i \rangle = 0$.

see also Ghost, Sampling

Joachimsthal's Equation

Using Clebsch-Aronhold Notation,

$$\xi_1^n a_y^n + \xi_1^{n-1}\xi_2 a_y^{n-1} a_x + \tfrac{1}{2}n(n-1)\xi_1^{n-2}\xi_2^2 a_y^{n-2} a_x^2 + \ldots + n\xi_1\xi_2^{n-1} a_y a_x^{n-1} + \xi_2^n a_x^n = 0.$$

References
Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 89, 1959.

Johnson Circle

The Circumcircle in Johnson's Theorem.

see also Johnson's Theorem

Johnson's Equation

The Partial Differential Equation

$$\frac{\partial}{\partial x}\left(u_t + uu_x + \tfrac{1}{2}u_{xxx} + \frac{u}{2t}\right) + \frac{3\alpha^2}{2t^2}u_{yy} = 0$$

which arises in the study of water waves.

References
Infeld, E. and Rowlands, G. *Nonlinear Waves, Solitons, and Chaos.* Cambridge, England: Cambridge University Press, p. 223, 1990.

Johnson Solid

The Johnson solids are the Convex Polyhedra having regular faces (with the exception of the completely regular Platonic Solids, the "Semiregular" Archimedean Solids, and the two infinite families of Prisms and Antiprisms). There are 28 simple (i.e., cannot be dissected into two other regular-faced polyhedra by a plane) regular-faced polyhedra in addition to the Prisms and Antiprisms (Zalgaller 1969), and Johnson (1966) proposed and Zalgaller (1969) proved that there exist exactly 92 Johnson solids in all.

A database of solids and Vertex Nets of these solids is maintained on the Bell Laboratories Netlib server, but a few errors exist in several entries. A concatenated and corrected version of the files is given by Weisstein, together with *Mathematica®* (Wolfram Research, Champaign, IL) code to display the solids and nets. The following table summarizes the names of the Johnson solids and gives their images and nets.

1. Square pyramid

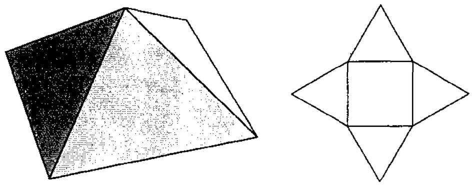

2. Pentagonal pyramid

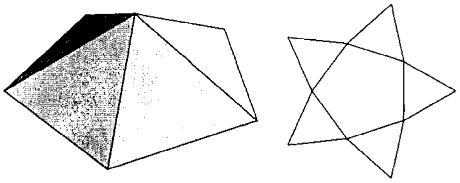

3. Triangular cupola

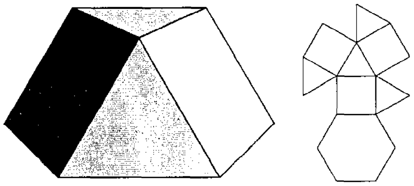

4. Square cupola

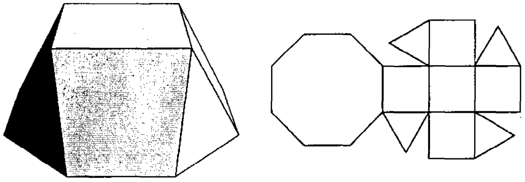

5. Pentagonal cupola

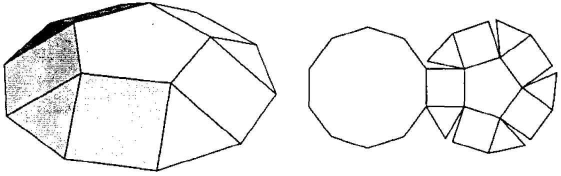

6. Pentagonal rotunda

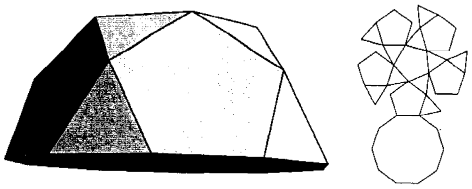

7. Elongated triangular pyramid

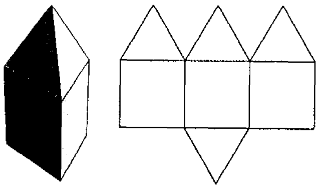

8. Elongated square pyramid

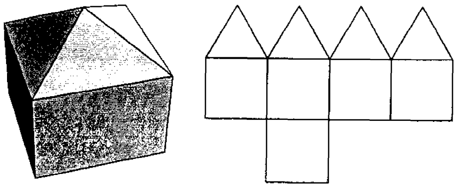

9. Elongated pentagonal pyramid

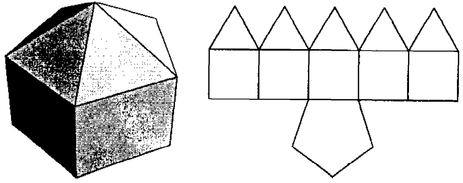

10. Gyroelongated square pyramid

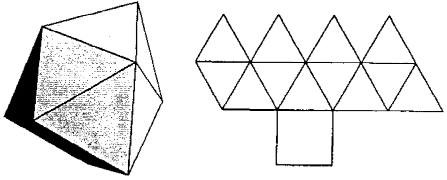

11. Gyroelongated pentagonal pyramid

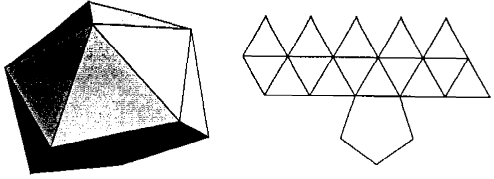

12. Triangular dipyramid

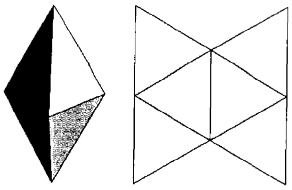

13. Pentagonal dipyramid

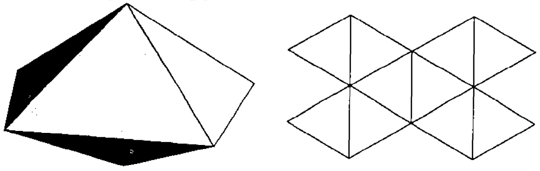

14. Elongated triangular dipyramid

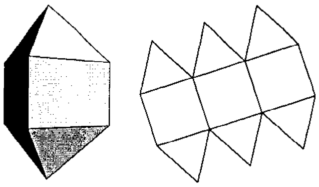

15. Elongated square dipyramid

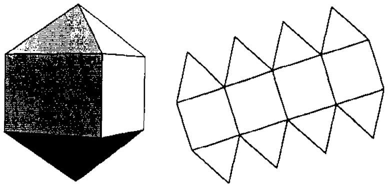

16. Elongated pentagonal dipyramid

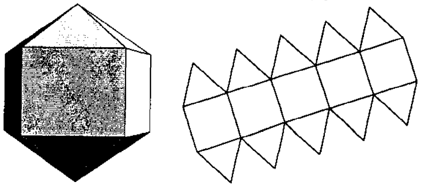

17. Gyroelongated square dipyramid

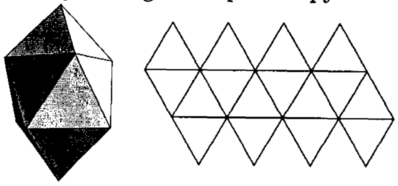

18. Elongated triangular cupola

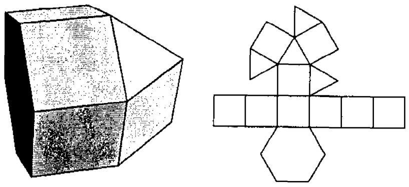

19. Elongated square cupola

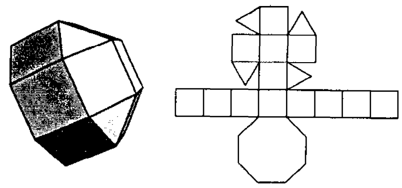

20. Elongated pentagonal cupola

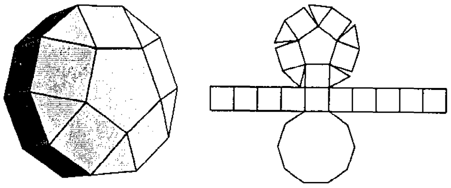

21. Elongated pentagonal rotunda

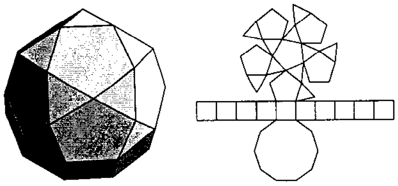

22. Gyroelongated triangular cupola

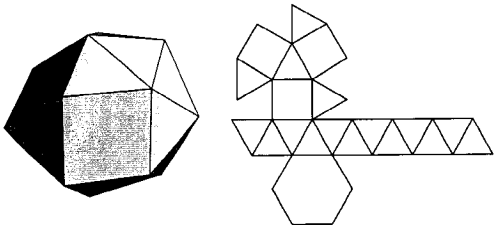

23. Gyroelongated square cupola

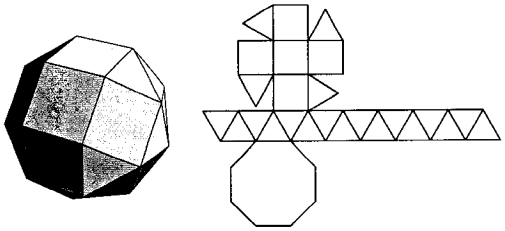

24. Gyroelongated pentagonal cupola

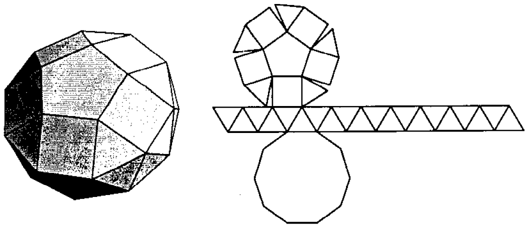

25. Gyroelongated pentagonal rotunda

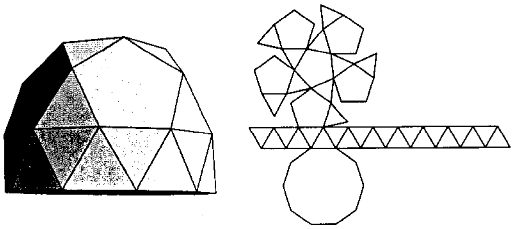

26. Gyrobifastigium

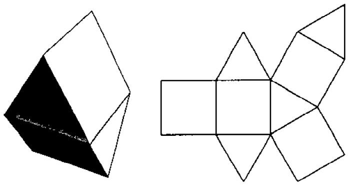

27. Triangular orthobicupola

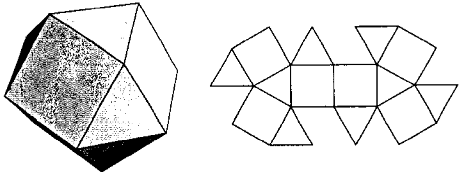

28. Square orthobicupola

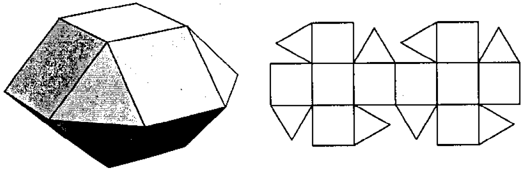

29. Square gyrobicupola

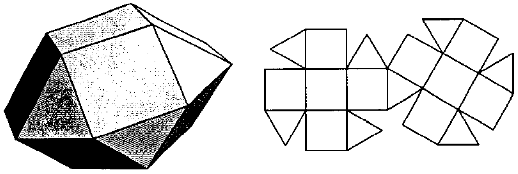

30. Pentagonal orthobicupola

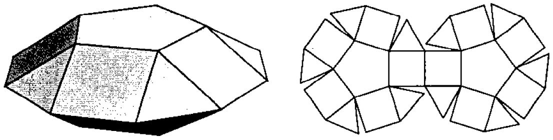

31. Pentagonal gyrobicupola

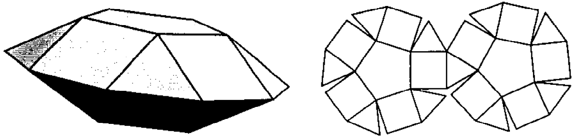

32. Pentagonal orthocupolarontunda

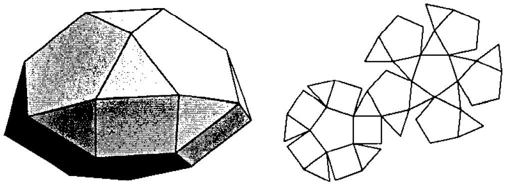

33. Pentagonal gyrocupolarotunda

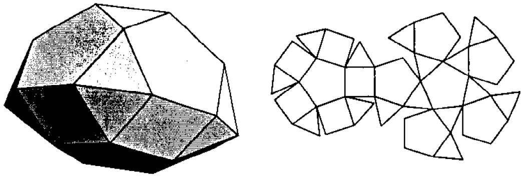

34. Pentagonal orthobirotunda

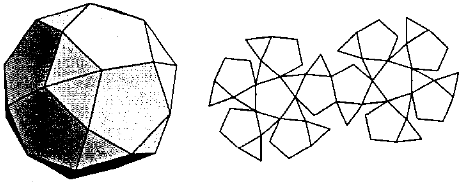

35. Elongated triangular orthobicupola

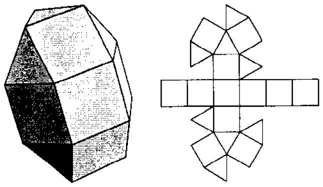

36. Elongated triangular gyrobicupola

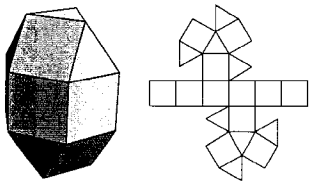

37. Elongated square gyrobicupola

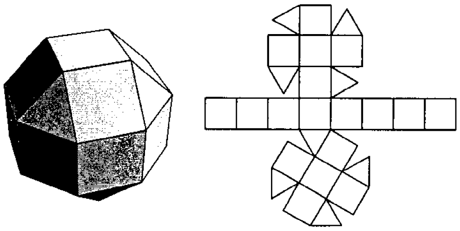

38. Elongated pentagonal orthobicupola

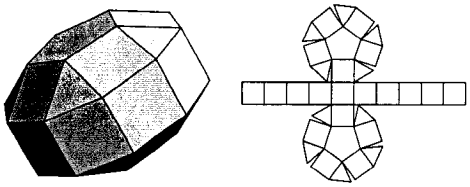

39. Elongated pentagonal gyrobicupola

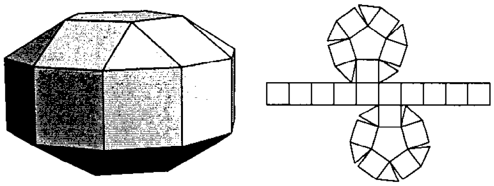

40. Elongated pentagonal orthocupolarotunda

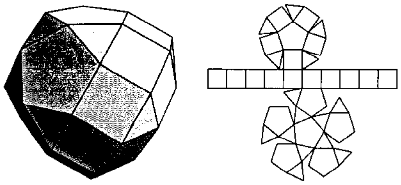

41. Elongated pentagonal gyrocupolarotunda

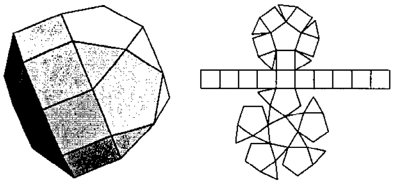

42. Elongated pentagonal orthobirotunda

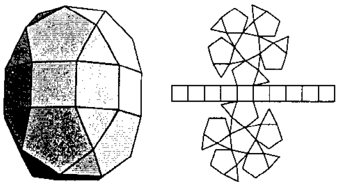

43. Elongated pentagonal gyrobirotunda

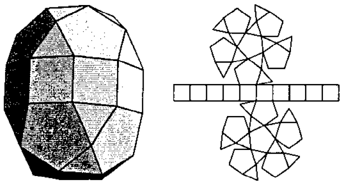

44. Gyroelongated triangular bicupola

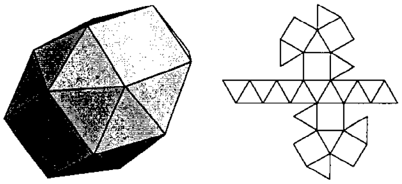

45. Gyroelongated square bicupola

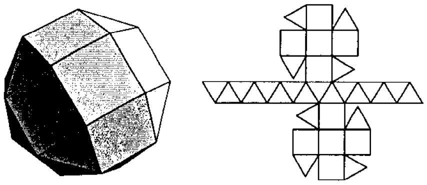

46. Gyroelongated pentagonal bicupola

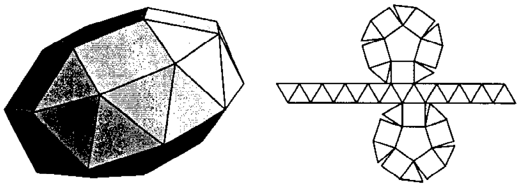

47. Gyroelongated pentagonal cupolarotunda

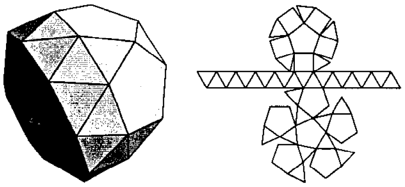

48. Gyroelongated pentagonal birotunda

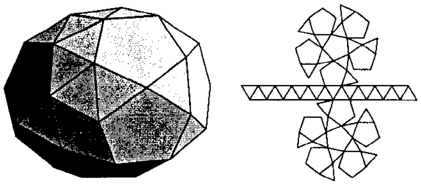

49. Augmented triangular prism

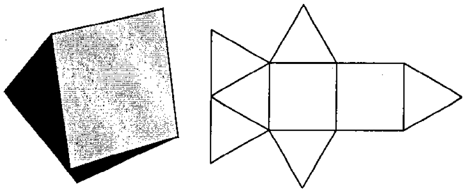

50. Biaugmented triangular prism

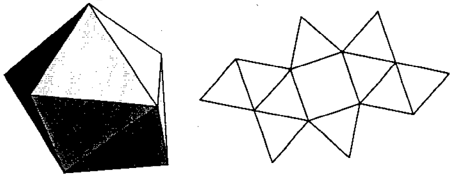

51. Triaugmented triangular prism

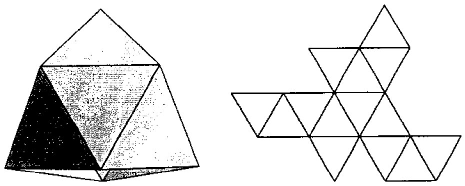

52. Augmented pentagonal prism

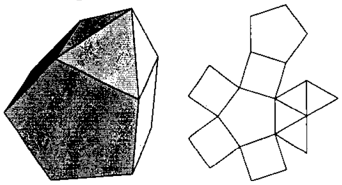

53. Biaugmented pentagonal prism

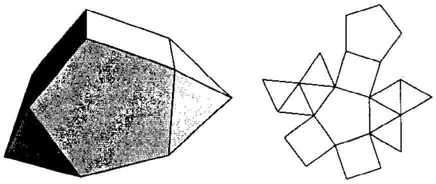

54. Augmented hexagonal prism

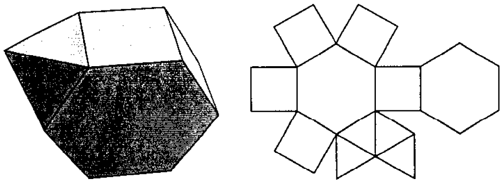

55. Parabiaugmented hexagonal prism

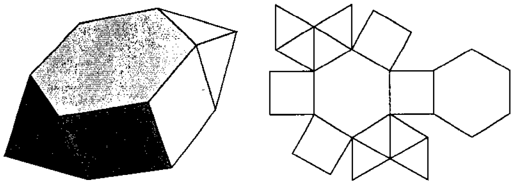

56. Metabiaugmented hexagonal prism

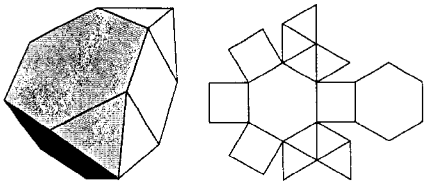

57. Triaugmented hexagonal prism

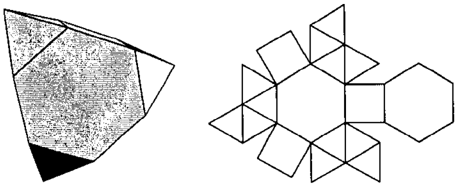

58. Augmented dodecahedron

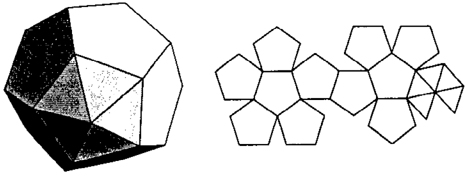

59. Parabiaugmented dodecahedron

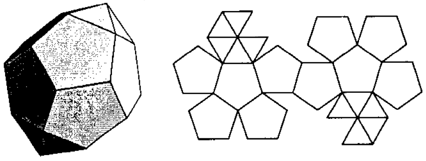

60. Metabiaugmented dodecahedron

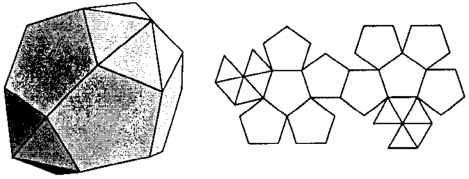

61. Triaugmented dodecahedron

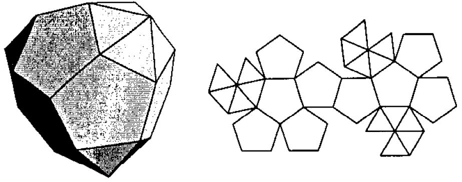

62. Metabidiminished icosahedron

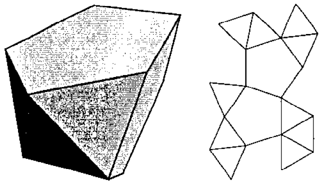

63. Tridiminished icosahedron

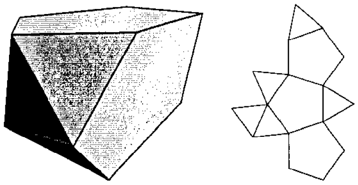

64. Augmented tridiminished icosahedron

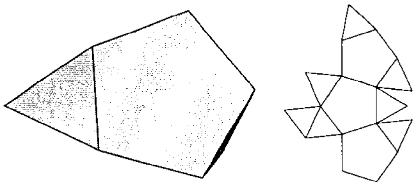

65. Augmented truncated tetrahedron

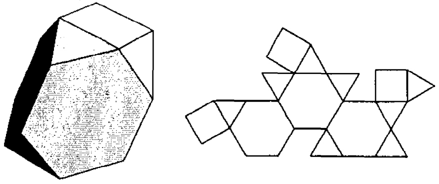

66. Augmented truncated cube

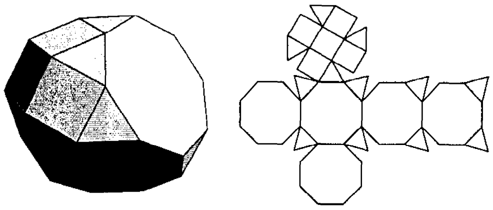

67. Biaugmented truncated cube

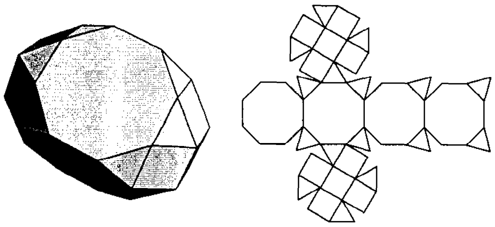

68. Augmented truncated dodecahedron

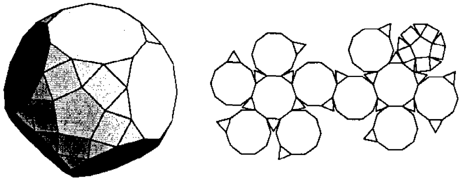

69. Parabiaugmented truncated dodecahedron

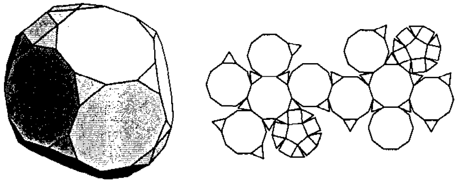

70. Metabiaugmented truncated dodecahedron

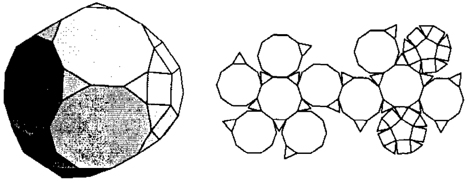

71. Triaugmented truncated dodecahedron

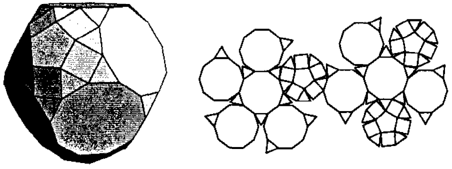

72. Gyrate rhombicosidodecahedron

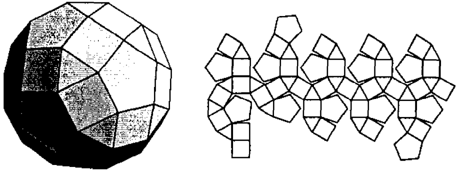

73. Parabigyrate rhombicosidodecahedron

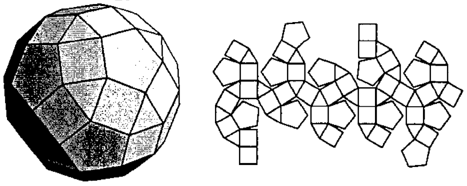

74. Metabigyrate rhombicosidodecahedron

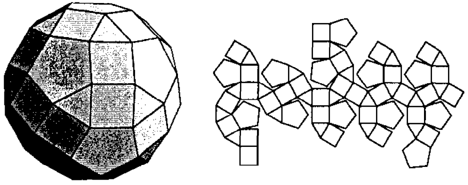

75. Trigyrate rhombicosidodecahedron

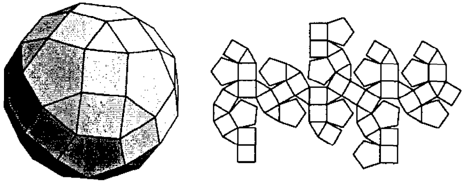

76. Diminished rhombicosidodecahedron

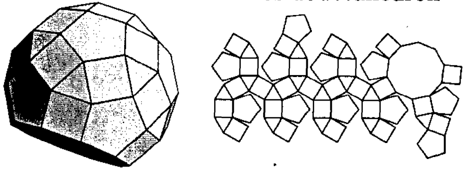

77. Paragyrate diminished rhombicosidodecahedron

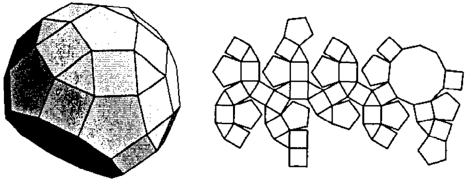

78. Metagyrate diminished rhombicosidodecahedron

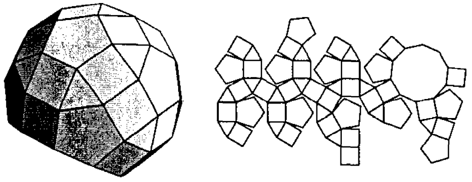

79. Bigyrate diminished rhombicosidodecahedron

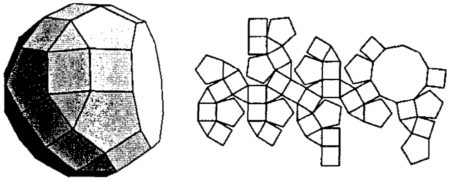

80. Parabidiminished rhombicosidodecahedron

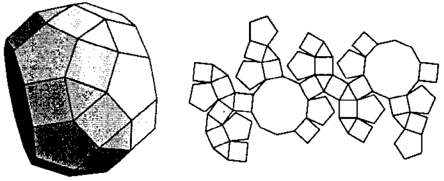

81. Metabidiminished rhombicosidodecahedron

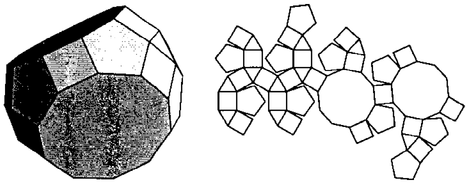

82. Gyrate bidiminished rhombicosidodecahedron

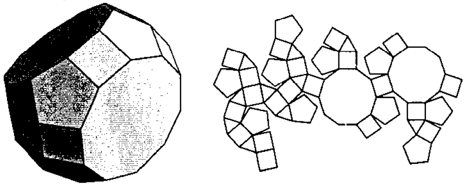

83. Tridiminished rhombicosidodecahedron

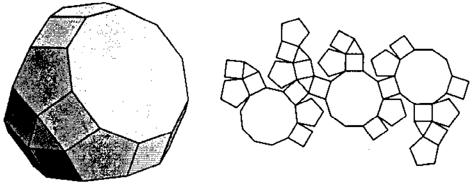

84. Snub disphenoid

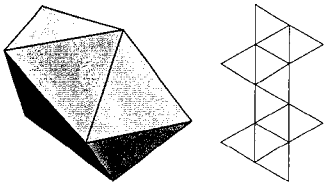

85. Snub square antiprism

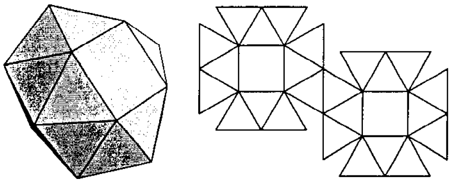

86. Sphenocorona

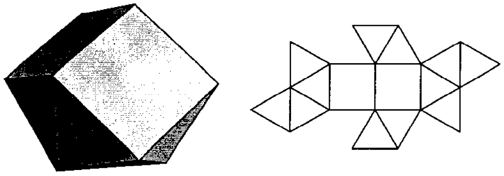

87. Augmented sphenocorona

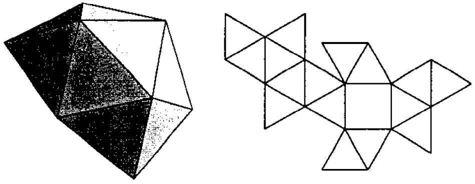

88. Sphenomegacorona

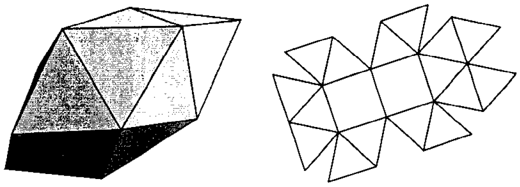

89. Hebesphenomegacorona

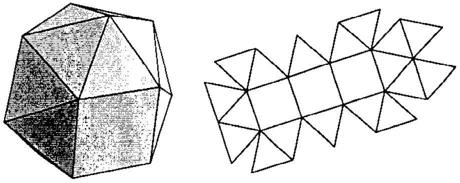

90. Disphenocingulum

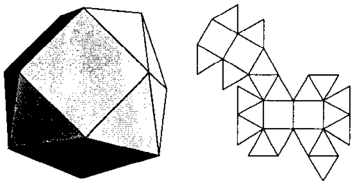

91. Bilunabirotunda

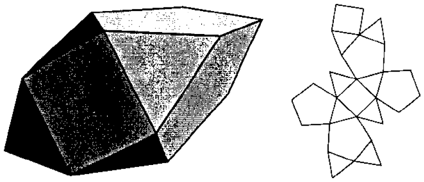

92. Triangular hebesphenorotunda

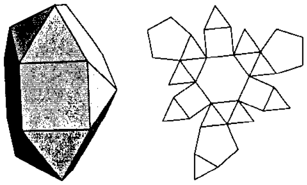

The number of constituent n-gons ($\{n\}$) for each Johnson solid are given in the following table.

J_n	{3}	{4}	{5}	{6}	{8}	{10}
1	4	1				
2	5		1			
3	4	3		1		
4	4	5			1	
5	5	5	1			1
6	10		6			1
7	4	3				
8	4	5				
9	5	5	1			
10	12	1				
11	15		1			
12	6					
13	10					
14	6	3				
15	8	4				
16	10	5				
17	16					
18	4	9		1		
19	4	13			1	
20	5	15	1			1
21	10	10	6			1
22	16	3		1		
23	20	5			1	
24	25	5	1			1
25	30		6			1
26	4	4				
27	8	6				
28	8	10				
29	8	10				
30	10	10	2			
31	10	10	2			
32	15	5	7			
33	15	5	7			
34	20		12			
35	8	12				
36	8	12				
37	8	18				
38	10	20	2			
39	10	20	2			
40	15	15	7			
41	15	15	7			
42	20	10	12			
43	20	10	12			
44	20	6				
45	24	10				
46	30	10	2			
47	35	5	7			
48	40		12			
49	6	2				
50	10	1				
51	14					
52	4	4	2			
53	8	3	2			
54	4	5		2		
55	8	4		2		
56	8	4		2		
57	12	3		2		
58	5		11			
59	10		10			
60	10		10			
61	15		9			
62	10		2			
63	5		3			
64	7		3			
65	8	3		3		
66	12	5			5	
67	16	10			4	
68	25	5	1			11
69	30	10	2			10
70	30	10	2			10
71	35	15	3			9
72	20	30	12			
73	20	30	12			
74	20	30	12			
75	20	30	12			
76	15	25	11			1
77	15	25	11			1
78	15	25	11			1
79	15	25	11			1
80	10	20	10			2
81	10	20	10			2
82	10	20	10			2
83	5	15	9			3
84	12					
85	24	2				
86	12	2				
87	16	1				
88	16	2				
89	18	3				
90	20	4				
91	8	2	4			
92	13	3	3	1		

see also ANTIPRISM, ARCHIMEDEAN SOLID, CONVEX POLYHEDRON, KEPLER-POINSOT SOLID, POLYHEDRON, PLATONIC SOLID, PRISM, UNIFORM POLYHEDRON

References

Bell Laboratories. `http://netlib.bell-labs.com/netlib/polyhedra/`.

Bulatov, V. "Johnson Solids." `http://www.physics.orst.edu/~bulatov/polyhedra/johnson/`.

Cromwell, P. R. *Polyhedra.* New York: Cambridge University Press, pp. 86–92, 1997.

Hart, G. W. "NetLib Polyhedra DataBase." `http://www.li.net/~george/virtual-polyhedra/netlib-info.html`.

Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, 1991.

Hume, A. *Exact Descriptions of Regular and Semi-Regular Polyhedra and Their Duals.* Computer Science Technical Report #130. Murray Hill, NJ: AT&T Bell Laboratories, 1986.

Johnson, N. W. "Convex Polyhedra with Regular Faces." *Canad. J. Math.* **18**, 169–200, 1966.

Pugh, A. "Further Convex Polyhedra with Regular Faces." Ch. 3 in *Polyhedra: A Visual Approach.* Berkeley, CA: University of California Press, pp. 28–35, 1976.

Weisstein, E. W. "Johnson Solids." `http://www.astro.virginia.edu/~eww6n/math/notebooks/JohnsonSolids.m`.

Weisstein, E. W. "Johnson Solid Netlib Database." `http://www.astro.virginia.edu/~eww6n/math/notebooks/JohnsonSolids.dat`.

Zalgaller, V. *Convex Polyhedra with Regular Faces.* New York: Consultants Bureau, 1969.

Johnson's Theorem

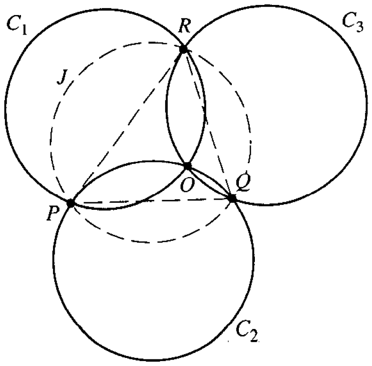

Let three equal CIRCLES with centers C_1, C_2, and C_3 intersect in a single point O and intersect pairwise in the points P, Q, and R. Then the CIRCUMCIRCLE J of ΔPQR (the so-called JOHNSON CIRCLE) is congruent to the original three.

see also CIRCUMCIRCLE, JOHNSON CIRCLE

References

Emch, A. "Remarks on the Foregoing Circle Theorem." *Amer. Math. Monthly* **23**, 162–164, 1916.

Honsberger, R. *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., pp. 18–21, 1976.

Johnson, R. "A Circle Theorem." *Amer. Math. Monthly* **23**, 161–162, 1916.

Join (Graph)

Let x and y be distinct nodes of G which are not joined by an EDGE. Then the graph G/xy which is formed by adding the EDGE (x,y) to G is called a join of G.

Join (Spaces)

Let X and Y be TOPOLOGICAL SPACES. Then their join is the factor space

$$X*Y=(X\times Y\times I)/\sim,$$

where $\sim$ is the EQUIVALENCE RELATION

$$(x,y,t)\sim(x',y',t')\Leftrightarrow\begin{cases}t=t'=0\text{ and }x=x'\\ \quad\text{or}\\ t=t'=1\text{ and }y=y.\end{cases}$$

see also CONE (SPACE), SUSPENSION

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 6, 1976.

Joint Distribution Function

A joint distribution function is a DISTRIBUTION FUNCTION in two variables defined by

$$D(x,y)\equiv P(X\leq x,Y\leq y) \tag{1}$$

$$D_x(x)\equiv D(x,\infty) \tag{2}$$

$$D_y(y)\equiv D(\infty,y) \tag{3}$$

so that the joint probability function

$$P[(x,y)\in C)]=\iint_{(x,y)\in C}P(x,y)\,dx\,dy \tag{4}$$

$$P(x\in A,y\in B)=\int_B\int_A P(x,y)\,dx\,dy \tag{5}$$

$$\begin{aligned}P(x,y)&=P\{x\in(-\infty,x],y\in(-\infty,y]\}\\&=\int_{-\infty}^{b}\int_{-\infty}^{a}P(x,y)\,dx\,dy\end{aligned} \tag{6}$$

$$\begin{aligned}&P(a\leq x\leq a+da,b\leq y\leq b+db)\\&\quad=\int_b^{b+db}\int_a^{a+da}P(x,y)\,dx\,dy\approx P(a,b)\,da\,db.\end{aligned} \tag{7}$$

A multiple distribution function is of the form

$$D(a_1,\ldots,a_n)\equiv P(x_1\leq a_1,\ldots,x_n\leq a_n). \tag{8}$$

see also DISTRIBUTION FUNCTION

Joint Probability Density Function

see JOINT DISTRIBUTION FUNCTION

Joint Theorem

see GAUSSIAN JOINT VARIABLE THEOREM

Jonah Formula

A formula for the generalized Catalan Number ${}_pd_{qi}$. The general formula is

$$\binom{n-q}{k-1} = \sum_{i=1}^{k} {}_pd_{qi}\binom{n-pi}{k-i},$$

where $\binom{n}{k}$ is a Binomial Coefficient, although Jonah's original formula corresponded to $p = 2$, $q = 0$ (Hilton and Pederson 1991).

References

Hilton, P. and Pederson, J. "Catalan Numbers, Their Generalization, and Their Uses." *Math. Intel.* **13**, 64–75, 1991.

Jones Polynomial

The second Knot Polynomial discovered. Unlike the first-discovered Alexander Polynomial, the Jones polynomial can sometimes distinguish handedness (as can its more powerful generalization, the HOMFLY Polynomial). Jones polynomials are Laurent Polynomials in t assigned to an $\mathbb{R}^3$ Knot. The Jones polynomials are denoted $V_L(t)$ for Links, $V_K(t)$ for Knots, and normalized so that

$$V_{\text{unknot}}(t) = 1. \qquad (1)$$

For example, the Jones polynomial of the Trefoil Knot is given by

$$V_{\text{trefoil}}(t) = t + t^3 - t^4. \qquad (2)$$

If a Link has an Odd number of components, then V_L is a Laurent Polynomial over the Integers; if the number of components is Even, $V_L(t)$ is $t^{1/2}$ times a Laurent Polynomial. The Jones polynomial of a Knot Sum $L_1\#L_2$ satisfies

$$V_{L_1\#L_2} = (V_{L_1})(V_{L_2}). \qquad (3)$$

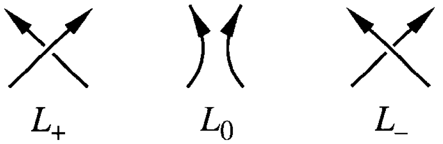

The Skein Relationship for under- and overcrossings is

$$t^{-1}V_{L_+} - tV_{L_-} = (t^{1/2} - t^{-1/2})V_{L_0}. \qquad (4)$$

Combined with the link sum relationship, this allows Jones polynomials to be built up from simple knots and links to more complicated ones.

Some interesting identities from Jones (1985) follow. For any Link L,

$$V_L(-1) = \Delta_L(-1), \qquad (5)$$

where Δ_L is the Alexander Polynomial, and

$$V_L(1) = (-2)^{p-1}, \qquad (6)$$

where p is the number of components of L. For any Knot K,

$$V_K(e^{2\pi i/3}) = 1 \qquad (7)$$

and

$$\frac{d}{dt}V_K(1) = 0. \qquad (8)$$

Let K^* denote the Mirror Image of a Knot K. Then

$$V_{K^*}(t) = V_K(t^{-1}). \qquad (9)$$

For example, the right-hand and left-hand Trefoil Knots have polynomials

$$V_{\text{trefoil}}(t) = t + t^3 - t^4 \qquad (10)$$

$$V_{\text{trefoil}^*}(t) = t^{-1} + t^{-3} - t^{-4}. \qquad (11)$$

Jones defined a simplified trace invariant for knots by

$$W_K(t) = \frac{1 - V_K(t)}{(1-t^3)(1-t)}. \qquad (12)$$

The Arf Invariant of W_K is given by

$$\text{Arf}(K) = W_K(i) \qquad (13)$$

(Jones 1985), where i is $\sqrt{-1}$. A table of the W polynomials is given by Jones (1985) for knots of up to eight crossings, and by Jones (1987) for knots of up to 10 crossings. (Note that in these papers, an additional polynomial which Jones calls V is also tabulated, but it is not the conventionally defined Jones polynomial.)

Jones polynomials were subsequently generalized to the two-variable HOMFLY Polynomials, the relationship being

$$V(t) = P(a = t, x = t^{1/2} - t^{-1/2}) \qquad (14)$$

$$V(t) = P(\ell = it, m = i(t^{-1/2} - t^{1/2})). \qquad (15)$$

They are related to the Kauffman Polynomial F by

$$V(t) = F(-t^{-3/4}, t^{-1/4} + t^{1/4}). \qquad (16)$$

Jones (1987) gives a table of Braid Words and W polynomials for knots up to 10 crossings. Jones polynomials for Knots up to nine crossings are given in Adams (1994) and for oriented links up to nine crossings by Doll and Hoste (1991). All Prime Knots with 10 or fewer crossings have distinct Jones polynomials. It is not known if there is a nontrivial knot with Jones polynomial 1. The Jones polynomial of an (m,n)-Torus Knot is

$$\frac{t^{(m-1)(n-1)/2}(1 - t^{m+1} - t^{n+1} + t^{m+n})}{1-t^2}. \qquad (17)$$

Let k be one component of an oriented LINK L. Now form a new oriented LINK L^* by reversing the orientation of k. Then

$$V_{L^*} = t^{-3\lambda} V(L),$$

where V is the Jones polynomial and λ is the LINKING NUMBER of k and $L - k$. No such result is known for HOMFLY POLYNOMIALS (Lickorish and Millett 1988).

Birman and Lin (1993) showed that substituting the POWER SERIES for e^x as the variable in the Jones polynomial yields a POWER SERIES whose COEFFICIENTS are VASSILIEV POLYNOMIALS.

Let L be an oriented connected LINK projection of n crossings, then

$$n \geq \text{span } V(L), \tag{18}$$

with equality if L is ALTERNATING and has no REMOVABLE CROSSING (Lickorish and Millett 1988).

There exist distinct KNOTS with the same Jones polynomial. Examples include (05_{001}, 10_{132}), (08_{008}, 10_{129}), (08_{016}, 10_{156}), (10_{025}, 10_{056}), (10_{022}, 10_{035}), (10_{041}, 10_{094}), (10_{043}, 10_{091}), (10_{059}, 10_{106}), (10_{060}, 10_{083}), (10_{071}, 10_{104}), (10_{073}, 10_{086}), (10_{081}, 10_{109}), and (10_{137}, 10_{155}) (Jones 1987). Incidentally, the first four of these also have the same HOMFLY POLYNOMIAL.

Witten (1989) gave a heuristic definition in terms of a topological quantum field theory, and Sawin (1996) showed that the "quantum group" $U_q(sl_2)$ gives rise to the Jones polynomial.

see also ALEXANDER POLYNOMIAL, HOMFLY POLYNOMIAL, KAUFFMAN POLYNOMIAL F, KNOT, LINK, VASSILIEV POLYNOMIAL

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, 1994.

Birman, J. S. and Lin, X.-S. "Knot Polynomials and Vassiliev's Invariants." *Invent. Math.* **111**, 225–270, 1993.

Doll, H. and Hoste, J. "A Tabulation of Oriented Links." *Math. Comput.* **57**, 747–761, 1991.

Jones, V. "A Polynomial Invariant for Knots via von Neumann Algebras." *Bull. Am. Math. Soc.* **12**, 103–111, 1985.

Jones, V. "Hecke Algebra Representations of Braid Groups and Link Polynomials." *Ann. Math.* **126**, 335–388, 1987.

Lickorish, W. B. R. and Millett, B. R. "The New Polynomial Invariants of Knots and Links." *Math. Mag.* **61**, 1–23, 1988.

Murasugi, K. "Jones Polynomials and Classical Conjectures in Knot Theory." *Topology* **26**, 297–307, 1987.

Praslov, V. V. and Sossinsky, A. B. *Knots, Links, Braids and 3-Manifolds: An Introduction to the New Invariants in Low-Dimensional Topology.* Providence, RI: Amer. Math. Soc., 1996.

Sawin, S. "Links, Quantum Groups, and TQFTS." *Bull. Amer. Math. Soc.* **33**, 413–445, 1996.

Stoimenow, A. "Jones Polynomials." `http://www.informatik.hu-berlin.de/~stoimeno/ptab/j10.html`.

Thistlethwaite, M. "A Spanning Tree Expansion for the Jones Polynomial." *Topology* **26**, 297–309, 1987.

Weisstein, E. W. "Knots and Links." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Knots.m`.

Witten, E. "Quantum Field Theory and the Jones Polynomial." *Comm. Math. Phys.* **121**, 351–399, 1989.

Jonquière's Function

see POLYGAMMA FUNCTION

Jordan Algebra

A nonassociative algebra with the product of elements A and B defined by the ANTICOMMUTATOR $\{A, B\} = AB + BA$.

see also ANTICOMMUTATOR

Jordan Curve

A Jordan curve is a plane curve which is topologically equivalent to (a HOMEOMORPHIC image of) the UNIT CIRCLE.

It is not known if every Jordan curve contains all four VERTICES of some SQUARE, but it has been proven true for "sufficiently smooth" curves and closed convex curves (Schnirelmann). For every TRIANGLE T and Jordan curve J, J has an INSCRIBED TRIANGLE similar to T.

see also JORDAN CURVE THEOREM, UNIT CIRCLE

Jordan Curve Theorem

If J is a simple closed curve in $\mathbb{R}^2$, then $\mathbb{R}^2 - J$ has two components (an "inside" and "outside"), with J the BOUNDARY of each.

see also JORDAN CURVE, SCHÖNFLIES THEOREM

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 9, 1976.

Jordan Decomposition Theorem

Let $V \neq (0)$ be a finite dimensional VECTOR SPACE over the COMPLEX NUMBERS, and let A be a linear operator on V. Then V can be expressed as a DIRECT SUM of cyclic subspaces.

References

Gohberg, I. and Goldberg, S. "A Simple Proof of the Jordan Decomposition Theorem for Matrices." *Amer. Math. Monthly* **103**, 157–159, 1996.

Jordan-Hölder Theorem

The composition quotient groups belonging to two COMPOSITION SERIES of a FINITE GROUP G are, apart from their sequence, ISOMORPHIC in pairs. In other words, if

$$I \subset H_s \subset \ldots \subset H_2 \subset H_1 \subset G$$

is one COMPOSITION SERIES and

$$I \subset K_t \subset \ldots \subset K_2 \subset K_1 \subset G$$

is another, then $t = s$, and corresponding to any composition quotient group K_j/K_{j+1}, there is a composition quotient group H_i/H_{i+1} such that

$$\frac{K_j}{K_{j+1}} = \frac{H_i}{H_{i+1}}.$$

This theorem was proven in 1869–1889.

see also COMPOSITION SERIES, FINITE GROUP, ISOMORPHIC GROUPS

References
Lomont, J. S. *Applications of Finite Groups.* New York: Dover, p. 26, 1993.

Jordan's Inequality

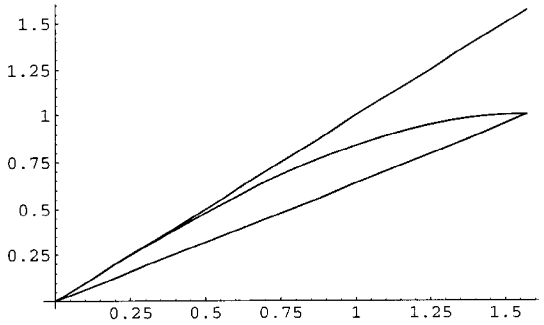

For $0 \leq x \leq \pi/2$,

$$\frac{2}{\pi}x \leq \sin x \leq x.$$

References
Yuefeng, F. "Jordan's Inequality." *Math. Mag.* **69**, 126, 1996.

Jordan's Lemma

Jordan's lemma shows the value of the INTEGRAL

$$I \equiv \int_{-\infty}^{\infty} f(x)e^{iax}\,dx \tag{1}$$

along the REAL AXIS is 0 for "nice" functions which satisfy $\lim_{R\to\infty}|f(Re^{i\theta})| = 0$. This is established using a CONTOUR INTEGRAL I_R which satisfies

$$\lim_{R\to\infty}|I_R| \leq \frac{\pi}{a}\lim_{R\to\infty}\epsilon = 0. \tag{2}$$

To derive the lemma, write

$$x \equiv Re^{i\theta} = R(\cos\theta + i\sin\theta) \tag{3}$$
$$dx = iRe^{i\theta}\,d\theta \tag{4}$$

and define the CONTOUR INTEGRAL

$$I_R = \int_0^{\pi} f(Re^{i\theta})e^{iaR\cos\theta - aR\sin\theta}iRe^{i\theta}\,d\theta \tag{5}$$

Then

$$\begin{aligned}|I_R| &= R\int_0^{\pi}|f(Re^{i\theta})|\,|e^{iaR\cos\theta}|\,|e^{-aR\sin\theta}|\,|i|\,|e^{i\theta}|\,d\theta\\ &= R\int_0^{\pi}|f(Re^{i\theta})|e^{-aR\sin\theta}\,d\theta\\ &= 2R\int_0^{\pi/2}|f(Re^{i\theta})|e^{-aR\sin\theta}\,d\theta. \end{aligned}\tag{6}$$

Now, if $\lim_{R\to\infty}|f(Re^{i\theta})| = 0$, choose an ϵ such that $|f(Re^{i\theta})| \leq \epsilon$, so

$$|I_R| \leq 2R\epsilon\int_0^{\pi/2} e^{-aR\sin\theta}\,d\theta. \tag{7}$$

But, for $\theta \in [0, \pi/2]$,

$$\frac{2}{\pi}\theta \leq \sin\theta, \tag{8}$$

so

$$\begin{aligned}|I_R| &\leq 2R\epsilon\int_0^{\pi/2} e^{-2aR\theta/\pi}\,d\theta\\ &= 2\epsilon R\frac{1-e^{-aR}}{\frac{2aR}{\pi}} = \frac{\pi\epsilon}{a}(1-e^{-aR}). \end{aligned}\tag{9}$$

As long as $\lim_{R\to\infty}|f(z)| = 0$, Jordan's lemma

$$\lim_{R\to\infty}|I_R| \leq \frac{\pi}{a}\lim_{R\to\infty}\epsilon = 0 \tag{10}$$

then follows.

see also CONTOUR INTEGRATION

References
Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 406–408, 1985.

Jordan Measure

Let the set M correspond to a bounded, NONNEGATIVE function f on an interval $0 \leq f(x) \leq c$ for $x \in [a, b]$. The Jordan measure, when it exists, is the common value of the outer and inner Jordan measures of M.

The outer Jordan measure is the greatest lower bound of the areas of the covering of M, consisting of finite unions of RECTANGLES. The inner Jordan measure of M is the difference between the AREA $c(a-b)$ of the RECTANGLE S with base $[a, b]$ and height c, and the outer measure of the complement of M in S.

References
Shenitzer, A. and Steprans, J. "The Evolution of Integration." *Amer. Math. Monthly* **101**, 66–72, 1994.

Jordan Polygon

see SIMPLE POLYGON

Josephus Problem

Given a group of n men arranged in a CIRCLE under the edict that every mth man will be executed going around the CIRCLE until only one remains, find the position $L(n,m)$ in which you should stand in order to be the last survivor (Ball and Coxeter 1987). The original problem consisted of a CIRCLE of 41 men with every third man killed ($n = 41$, $m = 3$). In order for the lives of the last two men to be spared, they must be placed at positions 31 (last) and 16 (second-to-last).

The following array gives the original position of the last survivor out of a group of $n = 1, 2, \ldots$, if every mth man is killed:

1									
2	1								
3	3	2							
4	1	1	2						
5	3	4	1	2					
6	5	1	5	1	4				
7	7	4	2	6	3	5			
8	1	7	6	3	1	4	4		
9	3	1	1	8	7	2	3	8	
10	5	4	5	3	3	9	1	7	8

(Sloane's A032434). The survivor for $m = 2$ can be given analytically by

$$L(n,2) = 1 + 2n - 2^{1+\lfloor \lg n \rfloor},$$

where $\lfloor n \rfloor$ is the FLOOR FUNCTION and LG is the LOGARITHM to base 2. The first few solutions are therefore 1, 1, 3, 1, 3, 5, 7, 1, 3, 5, 7, 9, 11, 13, 15, 1, ... (Sloane's A006257).

Mott-Smith (1954) discusses a card game called "Out and Under" in which cards at the top of a deck are alternately discarded and placed at the bottom. This is a Josephus problem with parameter $m = 2$, and Mott-Smith hints at the above closed-form solution.

The original position of the second-to-last survivor is given in the following table for $n = 2, 3, \ldots$:

1	1								
2	1	1							
3	1	1	2						
4	3	2	1	2					
5	1	1	5	1	4				
6	3	1	2	1	3	4			
7	1	4	6	3	1	3	4		
8	3	1	1	2	7	1	3	7	
9	5	4	5	3	3	8	1	6	4

(Sloane's A032435).

Another version of the problem considers a CIRCLE of two groups (say, "A" and "B") of 15 men each, with every ninth man cast overboard. To save all the members of the "A" group, the men must be placed at positions 1, 2, 3, 4, 10, 11, 13, 14, 15, 17, 20, 21, 25, 28, 29, giving the ordering

$$AAAABBBBBAABAAABABBAABBBABBAAB$$

which can be remembered with the aid of the MNEMONIC "From numbers' aid and art, never will fame depart." Consider the vowels only, assign $a = 1$, $e = 2$, $i = 3$, $o = 4$, $u = 5$, and alternately add a number of letters corresponding to a vowel value, so 4A (o), 5B (u), 2A (e), etc. (Ball and Coxeter 1987).

If every tenth man is instead thrown overboard, the men from the "A" group must be placed in positions 1, 2, 4, 5, 6, 12, 13, 16, 17, 18, 19, 21, 25, 28, 29, giving the sequence

$$AABAAABBBBBAABBAAAABABBBABBAAB$$

which can be constructed using the MNEMONIC "Rex paphi cum gente bona dat signa serena" (Ball and Coxeter 1987).

see also KIRKMAN'S SCHOOLGIRL PROBLEM, NECKLACE

References

Bachet, C. G. Problem 23 in *Problèmes plaisans et délectables, 2nd ed.* p. 174, 1624.

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 32–36, 1987.

Kraitchik, M. "Josephus' Problem." §3.13 in *Mathematical Recreations.* New York: W. W. Norton, pp. 93–94, 1942.

Mott-Smith, G. *Mathematical Puzzles for Beginners and Enthusiasts.* New York: Dover, 1954.

Sloane, N. J. A. Sequence A006257/M2216 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Jug

see THREE JUG PROBLEM

Jugendtraum

Kronecker proved that all the Galois extensions of the RATIONALS $\mathbb{Q}$ with ABELIAN Galois groups are SUBFIELDS of cyclotomic fields $Q(\mu_n)$, where μ_n is the group of nth ROOTS OF UNITY. He then sought to find a similar function whose division values would generate the Abelian extensions of an arbitrary NUMBER FIELD. He discovered that the j-FUNCTION works for IMAGINARY quadratic number fields K, but the completion of this problem, known as Kronecker's Jugendtraum ("dream of youth"), for other fields remains one of the great unsolved problems in NUMBER THEORY.

see also j-FUNCTION

References

Shimura, G. *Introduction to the Arithmetic Theory of Automorphic Functions.* Princeton, NJ: Princeton University Press, 1981.

Juggling

The throwing and catching of multiple objects such that at least one is always in the air. Some aspects of juggling turn out to be quite mathematical. The best examples are the two-handed asynchronous juggling sequences known as "SITESWAPS."

see also SITESWAP

References

Buhler, J.; Eisenbud, D.; Graham, R.; and Wright, C. "Juggling Drops and Descents." *Amer. Math. Monthly* **101**, 507–519, 1994.

Donahue, B. "Jugglers Now Juggle Numbers to Compute New Tricks for Ancient Art." *New York Times,* pp. B5 and B10, Apr. 16, 1996.

Juggling Information Service. "Siteswaps." `http://www.juggling.org/help/siteswap`.

Julia Fractal

see JULIA SET

Julia Set

Let $R(z)$ be a rational function

$$R(z) \equiv \frac{P(z)}{Q(z)}, \qquad (1)$$

where $z \in \mathbb{C}^*$, $\mathbb{C}^*$ is the RIEMANN SPHERE $\mathbb{C}\cup\{\infty\}$, and P and Q are POLYNOMIALS without common divisors. The "filled-in" Julia set J_R is the set of points z which do not approach infinity after $R(z)$ is repeatedly applied. The true Julia set is the boundary of the filled-in set (the set of "exceptional points"). There are two types of Julia sets: connected sets and CANTOR SETS.

For a Julia set J_c with $c \ll 1$, the CAPACITY DIMENSION is

$$d_{\text{capacity}} = 1 + \frac{|c|^2}{4\ln 2} + \mathcal{O}(|c|^3). \qquad (2)$$

For small c, J_c is also a JORDAN CURVE, although its points are not COMPUTABLE.

Quadratic Julia sets are generated by the quadratic mapping

$$z_{n+1} = {z_n}^2 + c \qquad (3)$$

for fixed c. The special case $c = -0.123 + 0.745i$ is called DOUADY'S RABBIT FRACTAL, $c = -0.75$ is called the SAN MARCO FRACTAL, and $c = -0.391 - 0.587i$ is the SIEGEL DISK FRACTAL. For every c, this transformation generates a FRACTAL. It is a CONFORMAL TRANSFORMATION, so angles are preserved. Let J be the JULIA SET, then $x' \mapsto x$ leaves J invariant. If a point P is on J, then all its iterations are on J. The transformation has a two-valued inverse. If $b = 0$ and y is started at 0, then the map is equivalent to the LOGISTIC MAP. The set of all points for which J is connected is known as the MANDELBROT SET.

see also DENDRITE FRACTAL, DOUADY'S RABBIT FRACTAL, FATOU SET, MANDELBROT SET, NEWTON'S METHOD, SAN MARCO FRACTAL, SIEGEL DISK FRACTAL

References

Dickau, R. M. "Julia Sets." `http://forum.swarthmore.edu/advanced/robertd/julias.html`.

Dickau, R. M. "Another Method for Calculating Julia Sets." `http://forum.swarthmore.edu/advanced/robertd/inversejulia.html`.

Douady, A. "Julia Sets and the Mandelbrot Set." In *The Beauty of Fractals: Images of Complex Dynamical Systems* (Ed. H.-O. Peitgen and D. H. Richter). Berlin: Springer-Verlag, p. 161, 1986.

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 124–126, 138–148, and 177–179, 1991.

Peitgen, H.-O. and Saupe, D. (Eds.). "The Julia Set," "Julia Sets as Basin Boundaries," "Other Julia Sets," and "Exploring Julia Sets." §3.3.2 to 3.3.5 in *The Science of Fractal Images.* New York: Springer-Verlag, pp. 152–163, 1988.

Schroeder, M. *Fractals, Chaos, Power Laws.* New York: W. H. Freeman, p. 39, 1991.

Wagon, S. "Julia Sets." §5.4 in *Mathematica in Action.* New York: W. H. Freeman, pp. 163–178, 1991.

Jump

A point of DISCONTINUITY.

see also DISCONTINUITY, JUMP ANGLE, JUMPING CHAMPION

Jump Angle

A GEODESIC TRIANGLE with oriented boundary yields a curve which is piecewise DIFFERENTIABLE. Furthermore, the TANGENT VECTOR varies continuously at all but the three corner points, where it changes suddenly. The angular difference of the tangent vectors at these corner points are called the jump angles.

see also ANGULAR DEFECT, GAUSS-BONNET FORMULA

Jumping Champion

An integer n is called a JUMPING CHAMPION if n is the most frequently occurring difference between consecutive primes $n \leq N$ for some N (Odlyzko *et al.*). This term was coined by J. H. Conway in 1993. There are occasionally several jumping champions in a range. Odlyzko *et al.* give a table of jumping champions for $n \leq 1000$, consisting mainly of 2, 4, and 6. 6 is the jumping champion up to about $n \approx 1.74 \times 10^{35}$, at which point 30 dominates. At $n \approx 10^{425}$, 210 becomes champion, and subsequent PRIMORIALS are conjectured to take over at larger and larger n. Erdős and Straus (1980) proved that the jumping champions tend to infinity under the assumption of a quantitative form of the k-tuples conjecture.

see also PRIME DIFFERENCE FUNCTION, PRIME GAPS, PRIME NUMBER, PRIMORIAL

References

Erdős, P.; and Straus, E. G. "Remarks on the Differences Between Consecutive Primes." *Elem. Math.* **35**, 115–118, 1980.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, 1994.
Nelson, H. "Problem 654." *J. Recr. Math.* **11**, 231, 1978–1979.
Odlyzko, A.; Rubinstein, M.; and Wolf, M. "Jumping Champions." `http://www.research.att.com/~amo/doc/recent.html`.

Jung's Theorem

Every finite set of points with SPAN d has an enclosing CIRCLE with RADIUS no greater than $\sqrt{3}\,d/3$.

References
Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 28, 1983.
Rademacher, H. and Toeplitz, O. *The Enjoyment of Mathematics: Selections from Mathematics for the Amateur.* Princeton, NJ: Princeton University Press, pp. 103–110, 1957.

Just If

see IFF

Just One

see EXACTLY ONE

K

k-ary Divisor

Let a DIVISOR d of n be called a 1-ary divisor if $d \perp n/d$. Then d is called a k-ary divisor of n, written $d|_k n$, if the GREATEST COMMON $(k-1)$-ary divisor of d and (n/d) is 1.

In this notation, $d|n$ is written $d|_0 n$, and $d||n$ is written $d|_1 n$. p^x is an INFINARY DIVISOR of p^y (with $y > 0$) if $p^x|_{y-1}p^y$.

see also DIVISOR, GREATEST COMMON DIVISOR, INFINARY DIVISOR

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 54, 1994.

k-Chain

Any sum of a selection of Π_ks, where Π_k denotes a k-D POLYTOPE.

see also k-CIRCUIT

k-Circuit

A k-CHAIN whose bounding $(k-1)$-CHAIN vanishes.

k-Coloring

A k-coloring of a GRAPH G is an assignment of one of k possible colors to each vertex of G such that no two adjacent vertices receive the same color.

see also COLORING, EDGE-COLORING

References

Saaty, T. L. and Kainen, P. C. *The Four-Color Problem: Assaults and Conquest.* New York: Dover, p. 13, 1986.

k-Form

see DIFFERENTIAL k-FORM

K-Function

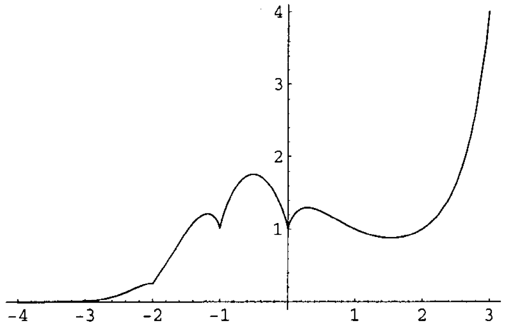

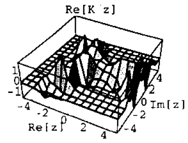

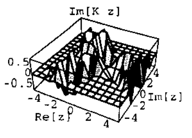

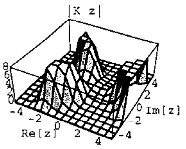

An extension of the K-function

$$K(n+1) \equiv 0^0 1^1 2^2 3^3 \cdots n^n \tag{1}$$

defined by

$$K(z) = \frac{[\Gamma(z)]^z}{G(z)}. \tag{2}$$

Here, $G(z)$ is the G-FUNCTION defined by

$$G(n+1) \equiv \frac{(n!)^n}{K(n+1)} = \begin{cases} 1 & \text{if } n = 0 \\ 0!1!2!\cdots(n-1)! & \text{if } n > 0. \end{cases} \tag{3}$$

The K-function is given by the integral

$$K(z) = (2\pi)^{-(z-1)/2} \exp\left[\binom{z}{2} + \int_0^{z-1} \ln(t!)\,dt\right] \tag{4}$$

and the closed-form expression

$$K(z) = \exp[\zeta'(-1,z) - \zeta'(-1)], \tag{5}$$

where $\zeta(z)$ is the RIEMANN ZETA FUNCTION, $\zeta'(z)$ its DERIVATIVE, $\zeta(a,z)$ is the HURWITZ ZETA FUNCTION, and

$$\zeta'(a,z) \equiv \left[\frac{d\zeta(s,z)}{ds}\right]_{s=a}. \tag{6}$$

$K(z)$ also has a STIRLING-like series

$$\begin{aligned} K(z+1) &= (2^{1/3}\pi_1 z)^{1/12} z^{\binom{z+1}{2}} \\ &\times \exp\left(\tfrac{1}{4}z^2 + \tfrac{1}{12} - \frac{B_4}{2\cdot 3\cdot 4z^2} - \frac{B_6}{4\cdot 5\cdot 6z^4} - \ldots\right), \end{aligned} \tag{7}$$

where

$$\pi_1 \equiv [K(\tfrac{1}{2})]^8 \tag{8}$$

$$= e^{-(\ln 2)/3 - 12\zeta'(-1)} \tag{9}$$

$$= 2^{2/3}\pi e^{\gamma - 1 - \zeta'(2)/\zeta(2)}, \tag{10}$$

and γ is the EULER-MASCHERONI CONSTANT (Gosper).

The first few values of $K(n)$ for $n = 2, 3, \ldots$ are 1, 4, 108, 27648, 86400000, 4031078400000, ... (Sloane's A002109). These numbers are called HYPERFACTORIALS by Sloane and Plouffe (1995).

see also G-FUNCTION, GLAISHER-KINKELIN CONSTANT, HYPERFACTORIAL, STIRLING'S SERIES

References

Sloane, N. J. A. Sequence A002109/M3706 in "An On-Line Version of the Encyclopedia of Integer Sequences."

K-Graph

The GRAPH obtained by dividing a set of VERTICES $\{1, \ldots, n\}$ into $k-1$ pairwise disjoint subsets with VERTICES of degree $n_1, \ldots, n_{k-1}$, satisfying

$$n = n_1 + \ldots + n_{k-1},$$

and with two VERTICES joined IFF they lie in distinct VERTEX sets. Such GRAPHS are denoted $K_{n_1,\ldots,n_k}$.

see also BIPARTITE GRAPH, COMPLETE GRAPH, COMPLETE k-PARTITE GRAPH, k-PARTITE GRAPH

k-Matrix

A k-matrix is a kind of CUBE ROOT of the IDENTITY MATRIX defined by

$$\mathsf{k} = \begin{bmatrix} 0 & 0 & -i \\ i & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

It satisfies

$$\mathsf{k}^3 = \mathsf{I},$$

where I is the IDENTITY MATRIX.

see also CUBE ROOT, QUATERNION

k-Partite Graph

A k-partite graph is a GRAPH whose VERTICES can be partitioned into k disjoint sets so that no two vertices within the same set are adjacent.

see also COMPLETE k-PARTITE GRAPH, K-GRAPH

References

Saaty, T. L. and Kainen, P. C. *The Four-Color Problem: Assaults and Conquest.* New York: Dover, p. 12, 1986.

k-Statistic

An UNBIASED ESTIMATOR of the CUMULANTS κ_i of a DISTRIBUTION. The expectation values of the k-statistics are therefore given by the corresponding CUMULANTS

$$\langle k_1 \rangle = \kappa_1 \qquad (1)$$

$$\langle k_2 \rangle = \kappa_2 \qquad (2)$$

$$\langle k_3 \rangle = \kappa_3 \qquad (3)$$

$$\langle k_4 \rangle = \kappa_4 \qquad (4)$$

(Kenney and Keeping 1951, p. 189). For a sample of size, N, the first few k-statistics are given by

$$k_1 = m_1 \qquad (5)$$

$$k_2 = \frac{N}{N-1} m_2 \qquad (6)$$

$$k_3 = \frac{N^2}{(N-1)(N-2)} m_3 \qquad (7)$$

$$k_4 = \frac{N^2[(N+1)m_4 - 3(N-1)m_2{}^2]}{(N-1)(N-2)(N-3)}, \qquad (8)$$

where m_1 is the sample MEAN, m_2 is the sample VARIANCE, and m_i is the sample ith MOMENT about the MEAN (Kenney and Keeping 1951, pp. 109–110, 163–165, and 189; Kenney and Keeping 1962). These statistics are obtained from inverting the relationships

$$\langle m_1 \rangle = \mu \qquad (9)$$

$$\langle m_2 \rangle = \frac{N-1}{N} \mu_2 \qquad (10)$$

$$\langle m_2{}^2 \rangle = \frac{(N-1)[(N-1)\mu_4 + (N^2-2N+3)\mu_2{}^2]}{N^3} \qquad (11)$$

$$\langle m_3 \rangle = \frac{(N-1)(N-2)}{N^2} \mu_3 \qquad (12)$$

$$\langle m_4 \rangle = \frac{(N-1)[(N^2-3N+3)\mu_4 + 3(2N-3)\mu_2{}^2]}{N^3}. \qquad (13)$$

The first moment (sample MEAN) is

$$m_1 \equiv \langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i, \qquad (14)$$

and the expectation value is

$$\langle m_1 \rangle = \left\langle \frac{1}{N} \sum_{i=1}^{N} x_i \right\rangle = \mu. \qquad (15)$$

The second MOMENT (sample STANDARD DEVIATION) is

$$\begin{aligned} m_2 &\equiv \langle (x-\mu)^2 \rangle = \langle x^2 \rangle - 2\mu \langle x \rangle + \mu^2 = \langle x^2 \rangle - \mu^2 \\ &= \frac{1}{N} \sum_{i=1}^{N} x_i{}^2 - \left(\frac{1}{N} \sum_{i=1}^{N} x_i \right)^2 \\ &= \frac{1}{N} \sum_{i=1}^{N} x_i{}^2 - \frac{1}{N^2} \left(\sum_{i=1}^{N} x_i{}^2 + \sum_{\substack{i,j=1 \\ i \neq j}}^{N} x_i x_j \right) \\ &= \frac{N-1}{N^2} \sum_{i=1}^{N} x_i{}^2 - \frac{1}{N^2} \sum_{\substack{i,j=1 \\ i \neq j}}^{N} x_i x_j, \end{aligned} \qquad (16)$$

and the expectation value is

$$\begin{aligned} \langle m_2 \rangle &= \frac{N-1}{N} \left\langle \frac{1}{N} \sum_{i=1}^{N} x_i{}^2 \right\rangle - \frac{1}{N^2} \left\langle \sum_{\substack{i,j=1 \\ i \neq j}}^{N} x_i x_j \right\rangle \\ &= \frac{N-1}{N} \mu_2' - \frac{N(N-1)}{N^2} \mu^2, \end{aligned} \qquad (17)$$

since there are $N(N-1)$ terms $x_i x_j$, using

$$\langle x_i x_j \rangle = \langle x_i \rangle \langle x_j \rangle = \langle x_i \rangle^2, \qquad (18)$$

and where μ_2' is the MOMENT about 0. Using the identity

$$\mu_2' = \mu_2 + \mu^2 \tag{19}$$

to convert to the MOMENT μ_2 about the MEAN and simplifying then gives

$$\langle m_2 \rangle = \frac{N-1}{N}\mu_2 \tag{20}$$

The factor $(N-1)/N$ is known as BESSEL'S CORRECTION.

The third MOMENT is

$$\begin{aligned} m_3 &\equiv \left\langle (x-\mu)^3 \right\rangle = \left\langle x^3 - 3\mu x^2 + 3\mu^2 x - \mu^3 \right\rangle \\ &= \left\langle x^3 \right\rangle - 3\mu \left\langle x^2 \right\rangle + 3\mu^2 \left\langle x \right\rangle - \mu^3 \\ &= \left\langle x^3 \right\rangle - 3\mu \left\langle x^2 \right\rangle + 3\mu^3 - \mu^3 \\ &= \left\langle x^3 \right\rangle - 3\mu \left\langle x^2 \right\rangle + 2\mu^3 \\ &= \frac{1}{N}\sum {x_i}^3 - 3\left(\frac{1}{N}\sum x_i\right)\left(\frac{1}{N}\sum {x_j}^2\right) \\ &\quad + 2\left(\frac{1}{N}\sum x_i\right)^3 \\ &= \frac{1}{N}\sum {x_i}^3 - \frac{3}{N^2}\left(\sum x_i\right)\left(\sum {x_j}^2\right) \\ &\quad + \frac{2}{N^3}\left(\sum x_i\right)^3. \end{aligned} \tag{21}$$

Now use the identities

$$\left(\sum {x_i}^2\right)\left(\sum x_j\right) = \sum {x_i}^3 + \sum {x_i}^2 x_j \tag{22}$$

$$\left(\sum x_i\right)^3 = \sum {x_i}^3 + 3\sum {x_i}^2 x_j + 6\sum x_i x_j x_k, \tag{23}$$

where it is understood that sums over products of variables exclude equal indices. Plugging in

$$\begin{aligned} m_3 &= \left(\frac{1}{N} - \frac{3}{N^2} + \frac{2}{N^3}\right)\sum {x_i}^3 \\ &+ \left(-\frac{3}{N^2} + 3\cdot\frac{2}{N^3}\right)\sum {x_i}^2 x_j + 6\cdot\frac{2}{N^3}\sum x_i x_j x_k. \end{aligned} \tag{24}$$

The expectation value is then given by

$$\begin{aligned} \langle m_3 \rangle &= \left(\frac{1}{N} - \frac{3}{N^2} + \frac{2}{N^3}\right) N\mu_3' \\ &+ \left(-\frac{3}{N^2} + \frac{6}{N^3}\right) N(N-1)\mu_2'\mu + \frac{12}{N^3}\tfrac{1}{6}N(N-1)(N-2)\mu^3 \end{aligned} \tag{25}$$

where μ_2' is the MOMENT about 0. Plugging in the identities

$$\mu_2' = \mu_2 + \mu^2 \tag{26}$$

$$\mu_3' = \mu_3 + 3\mu_2\mu + \mu^3 \tag{27}$$

and simplifying then gives

$$\langle m_3 \rangle = \frac{(N-1)(N-2)}{N^2}\mu_3 \tag{28}$$

(Kenney and Keeping 1951, p. 189).

The fourth MOMENT is

$$\begin{aligned} m_4 &= \left\langle (x-\mu)^4 \right\rangle = \left\langle x^4 - 4x^3\mu + 6x^2\mu^2 - 4x\mu^3 + \mu^4 \right\rangle \\ &= \left\langle x^4 \right\rangle - 4\mu \left\langle x^3 \right\rangle + 6\mu^2 \left\langle x^2 \right\rangle - 3\mu^4 \\ &= \frac{1}{N}\sum {x_i}^4 - \frac{4}{N^2}\left(\sum x_i\right)\left(\sum {x_j}^3\right) \\ &\quad + \frac{6}{N^3}\left(\sum x_i\right)^2\left(\sum {x_j}^2\right) - \frac{3}{N^4}\left(\sum x_i\right)^4 \end{aligned} \tag{29}$$

Now use the identities

$$\left(\sum x_i\right)\left(\sum {x_j}^3\right) = \sum {x_i}^4 + \sum {x_i}^3 x_j \tag{30}$$

$$\begin{aligned} \left(\sum x_i\right)^2\left(\sum {x_j}^2\right) &= \sum {x_i}^4 + 2\sum {x_i}^3 x_j \\ &\quad + 2\sum {x_i}^2 {x_j}^2 + 2\sum {x_i}^2 x_j x_k \end{aligned} \tag{31}$$

$$\begin{aligned} \left(\sum x_i\right)^4 &= \sum {x_i}^4 + 4\sum {x_i}^3 x_j + 6\sum {x_i}^2 {x_j}^2 \\ &\quad + 12\sum_{x_i}^{2} x_j x_k + 24\sum x_i x_j x_k x_l. \end{aligned} \tag{32}$$

Plugging in,

$$\begin{aligned} m_4 &= \left(\frac{1}{N} - \frac{4}{N^2} + \frac{6}{N^3} - \frac{3}{N^4}\right)\sum {x_i}^4 \\ &+ \left(-\frac{4}{N^2} + 2\cdot\frac{6}{N^3} - 4\cdot\frac{3}{N^4}\right)\sum {x_i}^3 x_j \\ &+ \left(2\cdot\frac{6}{N^3} - 6\cdot\frac{3}{N^4}\right)\sum {x_i}^2 {x_j}^2 \\ &+ \left(2\cdot\frac{6}{N^3} - 12\cdot\frac{3}{N^4}\right)\sum {x_i}^2 x_j x_k \\ &- 24\cdot\frac{3}{N^4}\sum x_i x_j x_k x_l. \end{aligned} \tag{33}$$

The expectation value is then given by

$$\begin{aligned} \langle m_4 \rangle &= \left(\frac{1}{N} - \frac{4}{N^2} + \frac{6}{N^3} - \frac{3}{N^4}\right) N\mu_4' \\ &+ \left(-\frac{4}{N^2} + \frac{12}{N^3} - \frac{12}{N^4}\right) N(N-1)\mu_3'\mu \\ &+ \left(\frac{12}{N^3} - \frac{18}{N^4}\right) \tfrac{1}{2}N(N-1){\mu_2'}^2 \\ &+ \left(\frac{18}{N^3} - \frac{36}{N^4}\right) \tfrac{1}{2}N(N-1)(N-2)\mu_2'\mu^2 \\ &- \frac{72}{N^4}\tfrac{1}{24}N(N-1)(N-2)(N-3)\mu^4, \end{aligned} \tag{34}$$

where μ_i' are MOMENTS about 0. Using the identities

$$\mu_2' = \mu_2 + \mu^2 \tag{35}$$

$$\mu_3' = \mu_3 + 3\mu_2\mu + \mu^3 \tag{36}$$

$$\mu_4' = \mu_4 + 4\mu_3\mu + 6\mu_2\mu^2 + \mu^4 \tag{37}$$

and simplifying gives

$$\langle m_4\rangle = \frac{(N-1)[(N^2-3N+3)\mu_4+3(2N-3){\mu_2}^2]}{N^3} \tag{38}$$

(Kenney and Keeping 1951, p. 189).

The square of the second moment is

$$\begin{aligned} {m_2}^2 &= (\langle x^2\rangle - \mu^2)^2 = \langle x^2\rangle^2 - 2\mu^2\langle x^2\rangle + \mu^4 \\ &= \left(\frac{1}{N}\sum {x_i}^2\right)^2 - 2\left(\frac{1}{N}\sum x_i\right)^2\left(\frac{1}{N}\sum {x_i}^2\right) \\ &\quad + \left(\frac{1}{N}\sum x_i\right)^4 \\ &= \frac{1}{N^2}\left(\sum {x_i}^2\right)^2 - \frac{2}{N^3}\left(\sum x_i\right)^2\left(\sum {x_j}^2\right) \\ &\quad + \frac{1}{N^4}\left(\sum x_i\right)^4. \end{aligned} \tag{39}$$

Now use the identities

$$\left(\sum {x_i}^2\right)^2 = \sum {x_i}^4 + 2\sum {x_i}^2{x_j}^2 \tag{40}$$

$$\begin{aligned} \left(\sum x_i\right)^2\left(\sum {x_j}^2\right) &= \sum {x_i}^4 + 2\sum {x_i}^2{x_j}^2 \\ &\quad +2\sum {x_i}^3x_j + 2\sum {x_i}^2x_jx_k \end{aligned} \tag{41}$$

$$\begin{aligned} \left(\sum x_i\right)^4 &= \sum {x_i}^4 + 6\sum {x_i}^2{x_j}^2 \\ &\quad +4\sum {x_i}^3x_j + 12\sum_{x_i}^{2} x_jx_k + 24\sum x_ix_jx_kx_l. \end{aligned} \tag{42}$$

Plugging in,

$$\begin{aligned} {m_2}^2 &= \left(\frac{1}{N^2} - \frac{2}{N^3} + \frac{1}{N^4}\right)\sum {x_i}^4 \\ &\quad + \left(2\cdot\frac{1}{N^2} - 2\cdot\frac{2}{N^3} + 6\cdot\frac{1}{N^4}\right)\sum {x_i}^2{x_j}^2 \\ &\quad + \left(-2\cdot\frac{2}{N^3} + 4\cdot\frac{1}{N^4}\right)\sum {x_i}^3x_j \\ &\quad + \left(-2\cdot\frac{2}{N^3} + 12\cdot\frac{1}{N^4}\right)\sum {x_i}^2x_jx_k \\ &\quad + \frac{24}{N^4}\sum x_ix_jx_kx_l \end{aligned} \tag{43}$$

The expectation value is then given by

$$\begin{aligned} \langle {m_2}^2\rangle &= \left(\frac{1}{N^2} - \frac{2}{N^3} + \frac{1}{N^4}\right)N\mu_4' \\ &\quad + \left(\frac{2}{N^2} - \frac{4}{N^3} + \frac{6}{N^4}\right)\tfrac{1}{2}N(N-1){\mu_2'}^2 \\ &\quad + \left(-\frac{4}{N^3} + \frac{4}{N^4}\right)N(N-1)\mu_3'\mu \\ &\quad + \left(-\frac{4}{N^3} + \frac{12}{N^4}\right)\tfrac{1}{2}N(N-1)(N-2)\mu_2'\mu^2 \\ &\quad + \frac{24}{N^4}\tfrac{1}{24}N(N-1)(N-2)(N-3)\mu^4 \end{aligned} \tag{44}$$

where μ_i' are MOMENTS about 0. Using the identities

$$\mu_2' = \mu_2 + \mu^2 \tag{45}$$

$$\mu_3' = \mu_3 + 3\mu_2\mu + \mu^3 \tag{46}$$

$$\mu_4' = \mu_4 + 4\mu_3\mu + 6\mu_2\mu^2 + \mu^4 \tag{47}$$

and simplifying gives

$$\langle {m_2}^2\rangle = \frac{(N-1)[(N-1)\mu_4 + (N^2-2N+3){\mu_2}^2]}{N^3} \tag{48}$$

(Kenney and Keeping 1951, p. 189).

The VARIANCE of k_2 is given by

$$\operatorname{var}(k_2) = \frac{\kappa_4}{N} + \frac{2}{(N-1){\kappa_2}^2}, \tag{49}$$

so an unbiased estimator of $\operatorname{var}(k_2)$ is given by

$$\widehat{\operatorname{var}}(k_2) = \frac{2{k_2}^2N + (N-1)k_4}{N(N+1)} \tag{50}$$

(Kenney and Keeping 1951, p. 189). The VARIANCE of k_3 can be expressed in terms of CUMULANTS by

$$\operatorname{var}(k_3) = \frac{\kappa_6}{N} + \frac{9\kappa_2\kappa_4}{N-1} + \frac{9{\kappa_3}^2}{N-1} + \frac{6{\kappa_2}^3}{N(N-1)(N-2)}. \tag{51}$$

An UNBIASED ESTIMATOR for $\operatorname{var}(k_3)$ is

$$\widehat{\operatorname{var}}(k_3) = \frac{6{k_2}^2N(N-1)}{(N-2)(N+1)(N+3)} \tag{52}$$

(Kenney and Keeping 1951, p. 190).

Now consider a finite population. Let a sample of N be taken from a population of M. Then UNBIASED ESTIMATORS M_2 for the population MEAN μ, M_2 for the population VARIANCE μ_2, G_1 for the population SKEWNESS γ_1, and G_2 for the population KURTOSIS γ_2 are

$$M_1 = \mu \tag{53}$$

$$M_2 = \frac{M-N}{N(M-1)}\mu_2 \tag{54}$$

$$G_1 = \frac{M-2N}{M-2}\sqrt{\frac{M-1}{N(M-N)}}\,\gamma_1 \tag{55}$$

$$\begin{aligned} G_2 &= \frac{(M-1)(M^2-6MN+M+6N^2)\gamma_2}{N(M-2)(M-3)(M-N)} \\ &\quad - \frac{6M(MN+M-N^2-1)}{N(M-2)(M-3)(M-N)} \end{aligned} \tag{56}$$

(Church 1926, p. 357; Carver 1930; Irwin and Kendall 1944; Kenney and Keeping 1951, p. 143), where γ_1 is the sample SKEWNESS and γ_2 is the sample KURTOSIS.

see also GAUSSIAN DISTRIBUTION, KURTOSIS, MEAN, MOMENT, SKEWNESS, VARIANCE

References

Carver, H. C. (Ed.). "Fundamentals of the Theory of Sampling." *Ann. Math. Stat.* **1**, 101–121, 1930.

Church, A. E. R. "On the Means and Squared Standard-Deviations of Small Samples from Any Population." *Biometrika* **18**, 321–394, 1926.

Irwin, J. O. and Kendall, M. G. "Sampling Moments of Moments for a Finite Population." *Ann. Eugenics* **12**, 138–142, 1944.

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, 1951.

Kenney, J. F. and Keeping, E. S. "The *k*-Statistics." §7.9 in *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, pp. 99–100, 1962.

k-Subset

A *k*-subset is a Subset containing exactly *k* elements.

see also Subset

k-Theory

A branch of mathematics which brings together ideas from algebraic geometry, Linear Algebra, and Number Theory. In general, there are two main types of *k*-theory: topological and algebraic.

Topological *k*-theory is the "true" *k*-theory in the sense that it came first. Topological *k*-theory has to do with Vector Bundles over Topological Spaces. Elements of a *k*-theory are Stable Equivalence classes of Vector Bundles over a Topological Space. You can put a Ring structure on the collection of Stably Equivalent bundles by defining Addition through the Whitney Sum, and Multiplication through the Tensor Product of Vector Bundles. This defines "the reduced real topological *k*-theory of a space."

"The reduced *k*-theory of a space" refers to the same construction, but instead of Real Vector Bundles, Complex Vector Bundles are used. Topological *k*-theory is significant because it forms a generalized Cohomology theory, and it leads to a solution to the vector fields on spheres problem, as well as to an understanding of the *J*-homeomorphism of Homotopy Theory.

Algebraic *k*-theory is somewhat more involved. Swan (1962) noticed that there is a correspondence between the Category of suitably nice Topological Spaces (something like regular Hausdorff Spaces) and C^*-Algebras. The idea is to associate to every Space the C^*-Algebra of Continuous Maps from that Space to the Reals.

A Vector Bundle over a Space has sections, and these sections can be multiplied by Continuous Functions to the Reals. Under Swan's correspondence, Vector Bundles correspond to modules over the C^*-Algebra of Continuous Functions, the Modules being the modules of sections of the Vector Bundle. This study of Modules over C^*-Algebra is the starting point of algebraic *k*-theory.

The Quillen-Lichtenbaum Conjecture connects algebraic *k*-theory to Étale cohomology.

see also C^*-Algebra

References

Srinivas, V. *Algebraic k-Theory, 2nd ed.* Boston, MA: Birkhäuser, 1995.

Swan, R. G. "Vector Bundles and Projective Modules." *Trans. Amer. Math. Soc.* **105**, 264–277, 1962.

k-Tuple Conjecture

The first of the Hardy-Littlewood Conjectures. The *k*-tuple conjecture states that the asymptotic number of Prime Constellations can be computed explicitly. In particular, unless there is a trivial divisibility condition that stops p, $p+a_1$, …, $p+a_k$ from consisting of Primes infinitely often, then such Prime Constellations will occur with an asymptotic density which is computable in terms of a_1, …, a_k. Let $0 < m_1 < m_2 < \ldots < m_k$, then the *k*-tuple conjecture predicts that the number of Primes $p \leq x$ such that $p+2m_1$, $p+2m_2$, …, $p+2m_k$ are all Prime is

$$P(x; m_1, m_2, \ldots, m_k) \sim C(m_1, m_2, \ldots, m_k) \int_2^x \frac{dt}{\ln^{k+1} t}, \tag{1}$$

where

$$C(m_1, m_2, \ldots, m_k) = 2^k \prod_q \frac{1 - \frac{w(q; m_1, m_2, \ldots, m_k)}{q}}{\left(1 - \frac{1}{q}\right)^{k+1}}, \tag{2}$$

the product is over Odd Primes q, and

$$w(q; m_1, m_2, \ldots, m_k) \tag{3}$$

denotes the number of distinct residues of 0, m_1, …, m_k (mod q) (Halberstam and Richert 1974, Odlyzko). If $k = 1$, then this becomes

$$C(m) = 2 \prod_q \frac{q(q-2)}{(q-1)^2} \prod_{q|m} \frac{q-1}{q-2}. \tag{4}$$

This conjecture is generally believed to be true, but has not been proven (Odlyzko *et al.*). The following special case of the conjecture is sometimes known as the Prime Patterns Conjecture. Let S be a Finite set of Integers. Then it is conjectured that there exist infinitely many k for which $\{k + s : s \in S\}$ are all Prime Iff S does not include all the Residues of any Prime. The Twin Prime Conjecture is a special case of the prime patterns conjecture with $S = \{0, 2\}$. This conjecture also implies that there are arbitrarily long Arithmetic Progressions of Primes.

see also Arithmetic Progression, Dirichlet's Theorem, Hardy-Littlewood Conjectures, *k*-Tuple Conjecture, Prime Arithmetic Progression, Prime Constellation, Prime Quadruplet, PRIME PATTERNS CONJECTURE, TWIN PRIME CONJECTURE, TWIN PRIMES

References

Brent, R. P. "The Distribution of Small Gaps Between Successive Primes." *Math. Comput.* **28**, 315–324, 1974.

Brent, R. P. "Irregularities in the Distribution of Primes and Twin Primes." *Math. Comput.* **29**, 43–56, 1975.

Halberstam, E. and Richert, H.-E. *Sieve Methods.* New York: Academic Press, 1974.

Hardy, G. H. and Littlewood, J. E. "Some Problems of 'Partitio Numerorum.' III. On the Expression of a Number as a Sum of Primes." *Acta Math.* **44**, 1–70, 1922.

Odlyzko, A.; Rubinstein, M.; and Wolf, M. "Jumping Champions."

Riesel, H. *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, pp. 66–68, 1994.

Kabon Triangles

The largest number $N(n)$ of nonoverlapping TRIANGLES which can be produced by n straight LINE SEGMENTS. The first few terms are 1, 2, 5, 7, 11, 15, 21, ... (Sloane's A006066).

References

Sloane, N. J. A. Sequence A006066/M1334 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Kac Formula

The expected number of REAL zeros E_n of a RANDOM POLYNOMIAL of degree n is

$$E_n = \frac{1}{\pi}\int_{-\infty}^{\infty}\sqrt{\frac{1}{(t^2-1)^2}-\frac{(n+1)^2t^{2n}}{(t^{2n+2}-1)^2}}\,dt \qquad (1)$$

$$= \frac{4}{\pi}\int_0^1\sqrt{\frac{1}{(1-t^2)^2}-\frac{(n+1)^2t^{2n}}{(1-t^{2n+2})^2}}\,dt. \qquad (2)$$

As $n \to \infty$,

$$E_n = \frac{2}{\pi}\ln n + C_1 + \frac{2}{\pi n} + \mathcal{O}(n^{-2}), \qquad (3)$$

where

$$C_1 = \frac{2}{\pi}\left[\ln 2 + \int_0^\infty \left(\sqrt{\frac{1}{x^2}-\frac{4e^{-2x}}{(1-e^{-2x})^2}} - \frac{1}{x+1}\right)dx\right] = 0.6257358072\ldots. \qquad (4)$$

The initial term was derived by Kac (1943).

References

Edelman, A. and Kostlan, E. "How Many Zeros of a Random Polynomial are Real?" *Bull. Amer. Math. Soc.* **32**, 1–37, 1995.

Kac, M. "On the Average Number of Real Roots of a Random Algebraic Equation." *Bull. Amer. Math. Soc.* **49**, 314–320, 1943.

Kac, M. "A Correction to 'On the Average Number of Real Roots of a Random Algebraic Equation'." *Bull. Amer. Math. Soc.* **49**, 938, 1943.

Kac Matrix

The $(n+1)\times(n+1)$ TRIDIAGONAL MATRIX (also called the CLEMENT MATRIX) defined by

$$S_n = \begin{bmatrix} 0 & n & 0 & 0 & \cdots & 0 \\ 1 & 0 & n-1 & 0 & \cdots & 0 \\ 0 & 2 & 0 & n-2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & n-1 & 0 & 1 \\ 0 & 0 & 0 & 0 & n & 0 \end{bmatrix}.$$

The EIGENVALUES are $2k-n$ for $k = 0, 1, \ldots, n$.

Kähler Manifold

A manifold for which the EXTERIOR DERIVATIVE of the FUNDAMENTAL FORM Ω associated with the given Hermitian metric vanishes, so $d\Omega = 0$.

References

Amorós, J. *Fundamental Groups of Compact Kähler Manifolds.* Providence, RI: Amer. Math. Soc., 1996.

Iyanaga, S. and Kawada, Y. (Eds.). "Kähler Manifolds." §232 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 732–734, 1980.

Kakeya Needle Problem

What is the plane figure of least AREA in which a line segment of width 1 can be freely rotated (where translation of the segment is also allowed)? Besicovitch (1928) proved that there is *no* MINIMUM AREA. This can be seen by rotating a line segment inside a DELTOID, star-shaped 5-oid, star-shaped 7-oid, etc. When the figure is restricted to be convex, Cunningham and Schoenberg (1965) found there is still *no* minimum AREA. However, the smallest *simple convex* domain in which one can put a segment of length 1 which will coincide with itself when rotated by 180° is

$$\tfrac{1}{24}(5-2\sqrt{2})\pi = 0.284258\ldots$$

(Le Lionnais 1983).

see also CURVE OF CONSTANT WIDTH, LEBESGUE MINIMAL PROBLEM, REULEAUX POLYGON, REULEAUX TRIANGLE

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 99–101, 1987.

Besicovitch, A. S. "On Kakeya's Problem and a Similar One." *Math. Z.* **27**, 312–320, 1928.

Besicovitch, A. S. "The Kakeya Problem." *Amer. Math. Monthly* **70**, 697–706, 1963.

Cunningham, F. Jr. and Schoenberg, I. J. "On the Kakeya Constant." *Canad. J. Math.* **17**, 946–956, 1965.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 24, 1983.

Ogilvy, C. S. *A Calculus Notebook.* Boston: Prindle, Weber, & Schmidt, 1968.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 147–153, 1990.

Pál, J. "Ein Minimumproblem für Ovale." *Math. Ann.* **88**, 311–319, 1921.
Plouffe, S. "Kakeya Constant." `http://lacim.uqam.ca/piDATA/kakeya.txt`.
Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 50–52, 1991.

Kakutani's Fixed Point Theorem

Every correspondence that maps a compact convex subset of a locally convex space into itself with a closed graph and convex nonempty images has a fixed point.

see also FIXED POINT THEOREM

Kakutani's Problem

see COLLATZ PROBLEM

Kalman Filter

An ALGORITHM in CONTROL THEORY introduced by R. Kalman in 1960 and refined by Kalman and R. Bucy. It is an ALGORITHM which makes optimal use of imprecise data on a linear (or nearly linear) system with Gaussian errors to continuously update the best estimate of the system's current state.

see also WIENER FILTER

References
Chui, C. K. and Chen, G. *Kalman Filtering: With Real-Time Applications, 2nd ed.* Berlin: Springer-Verlag, 1991.
Grewal, M. S. *Kalman FIltering: Theory & Practice.* Englewood Cliffs, NJ: Prentice-Hall, 1993.

KAM Theorem

see KOLMOGOROV-ARNOLD-MOSER THEOREM

Kampyle of Eudoxus

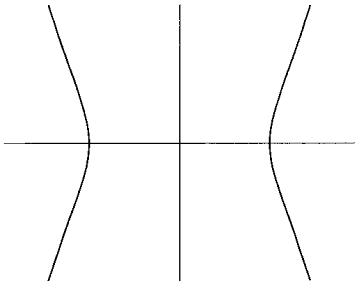

A curve studied by Eudoxus in relation to the classical problem of CUBE DUPLICATION. It is given by the polar equation

$$r\cos^2\theta = a,$$

and the parametric equations

$$x = a\sec t$$
$$y = a\tan t\sec t$$

with $t \in [-\pi/2, \pi/2]$.

References
Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 141–143, 1972.
MacTutor History of Mathematics Archive. "Kampyle of Eudoxus." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Kampyle.html`.

Kanizsa Triangle

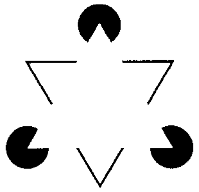

An optical ILLUSION, illustrated above, in which the eye perceives a white upright EQUILATERAL TRIANGLE where none is actually drawn.

see also ILLUSION

References
Bradley, D. R. and Petry, H. M. "Organizational Determinants of Subjective Contour." *Amer. J. Psychology* **90**, 253–262, 1977.
Fineman, M. *The Nature of Visual Illusion.* New York: Dover, pp. 26, 137, and 156, 1996.

Kantrovich Inequality

Suppose $x_1 < x_2 < \ldots < x_n$ are given POSITIVE numbers. Let $\lambda_1, \ldots, \lambda_n \geq 0$ and $\sum \lambda_j = 1$. Then

$$\left(\sum \lambda_j x_j\right)\left(\sum \lambda_j {x_j}^{-1}\right) \leq A^2 G^{-2},$$

where

$$A = \tfrac{1}{2}(x_1 + x_n)$$
$$G = \sqrt{x_1 x_n}$$

are the ARITHMETIC and GEOMETRIC MEAN, respectively, of the first and last numbers.

References
Pták, V. "The Kantrovich Inequality." *Amer. Math. Monthly* **102**, 820–821, 1995.

Kaplan-Yorke Conjecture

There are several versions of the Kaplan-Yorke conjecture, with many of the higher dimensional ones remaining unsettled. The original Kaplan-Yorke conjecture (Kaplan and Yorke 1979) proposed that, for a two-dimensional mapping, the CAPACITY DIMENSION D equals the KAPLAN-YORKE DIMENSION D_{KY},

$$D = D_{KY} = d_{\text{Lya}} = 1 + \frac{\sigma_1}{\sigma_2},$$

where σ_1 and σ_2 are the LYAPUNOV CHARACTERISTIC EXPONENTS. This was subsequently proven to be true in 1982. A later conjecture held that the KAPLAN-YORKE DIMENSION is generically equal to a probabilistic dimension which appears to be identical to the INFORMATION DIMENSION (Frederickson *et al.* 1983). This conjecture is partially verified by Ledrappier (1981). For invertible 2-D maps, $\nu = \sigma = D$, where ν is the CORRELATION EXPONENT, σ is the INFORMATION DIMENSION, and D is the CAPACITY DIMENSION (Young 1984).

see also CAPACITY DIMENSION, KAPLAN-YORKE DIMENSION, LYAPUNOV CHARACTERISTIC EXPONENT, LYAPUNOV DIMENSION

References

Chen, Z. M. "A Note on Kaplan-Yorke-Type Estimates on the Fractal Dimension of Chaotic Attractors." *Chaos, Solitons, and Fractals* **3**, 575–582, 1994.

Frederickson, P.; Kaplan, J. L.; Yorke, E. D.; and Yorke, J. A. "The Liapunov Dimension of Strange Attractors." *J. Diff. Eq.* **49**, 185–207, 1983.

Kaplan, J. L. and Yorke, J. A. In *Functional Differential Equations and Approximations of Fixed Points* (Ed. H.-O. Peitgen and H.-O. Walther). Berlin: Springer-Verlag, p. 204, 1979.

Ledrappier, F. "Some Relations Between Dimension and Lyapunov Exponents." *Commun. Math. Phys.* **81**, 229–238, 1981.

Worzbusekros, A. "Remark on a Conjecture of Kaplan and Yorke." *Proc. Amer. Math. Soc.* **85**, 381–382, 1982.

Young, L. S. "Dimension, Entropy, and Lyapunov Exponents in Differentiable Dynamical Systems." *Phys. A* **124**, 639–645, 1984

Kaplan-Yorke Dimension

$$D_{\text{KY}} \equiv j + \frac{\sigma_1 + \ldots + \sigma_j}{|\sigma_{j+1}|},$$

where $\sigma_1 \leq \sigma_n$ are LYAPUNOV CHARACTERISTIC EXPONENTS and j is the largest INTEGER for which

$$\lambda_1 + \ldots + \lambda_j \geq 0.$$

If $\nu = \sigma = D$, where ν is the CORRELATION EXPONENT, σ the INFORMATION DIMENSION, and D the HAUSDORFF DIMENSION, then

$$D \leq D_{\text{KY}}$$

(Grassberger and Procaccia 1983).

References

Grassberger, P. and Procaccia, I. "Measuring the Strangeness of Strange Attractors." *Physica D* **9**, 189–208, 1983.

Kaplan-Yorke Map

$$\begin{aligned} x_{n+1} &= 2x_n \\ y_{n+1} &= \alpha y_n + \cos(4\pi x_n), \end{aligned}$$

where x_n, y_n are computed mod 1. (Kaplan and Yorke 1979). The Kaplan-Yorke map with $\alpha = 0.2$ has CORRELATION EXPONENT 1.42 ± 0.02 (Grassberger Procaccia 1983) and CAPACITY DIMENSION 1.43 (Russell *et al.* 1980).

References

Grassberger, P. and Procaccia, I. "Measuring the Strangeness of Strange Attractors." *Physica D* **9**, 189–208, 1983.

Kaplan, J. L. and Yorke, J. A. In *Functional Differential Equations and Approximations of Fixed Points* (Ed. H.-O. Peitgen and H.-O. Walther). Berlin: Springer-Verlag, p. 204, 1979.

Russell, D. A.; Hanson, J. D.; and Ott, E. "Dimension of Strange Attractors." *Phys. Rev. Let.* **45**, 1175–1178, 1980.

Kappa Curve

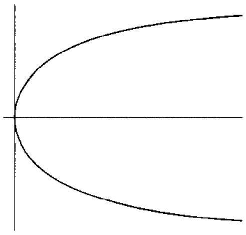

A curve also known as GUTSCHOVEN'S CURVE which was first studied by G. van Gutschoven around 1662 (MacTutor Archive). It was also studied by Newton and, some years later, by Johann Bernoulli. It is given by the Cartesian equation

$$(x^2 + y^2)y^2 = a^2x^2, \tag{1}$$

by the polar equation

$$r = a\cot\theta, \tag{2}$$

and the parametric equations

$$x = a\cos t\cot t \tag{3}$$

$$y = a\cos t. \tag{4}$$

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 136 and 139–141, 1972.

MacTutor History of Mathematics Archive. "Kappa Curve." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Kappa.html`.

Kaprekar Number

Consider an n-digit number k. Square it and add the right n digits to the left n or $n-1$ digits. If the resultant sum is k, then k is called a Kaprekar number. The first few are 1, 9, 45, 55, 99, 297, 703, ... (Sloane's A006886).

$$9^2 = 81 \qquad 8 + 1 = 9$$

$$297^2 = 88{,}209 \qquad 88 + 209 = 297.$$

see also DIGITAL ROOT, DIGITADITION, HAPPY NUMBER, KAPREKAR ROUTINE, NARCISSISTIC NUMBER, RECURRING DIGITAL INVARIANT

References

Sloane, N. J. A. Sequence A006886/M4625 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Kaprekar Routine

A routine discovered in 1949 by D. R. Kaprekar for 4-digit numbers, but which can be generalized to k-digit numbers. To apply the Kaprekar routine to a number n, arrange the digits in descending (n') and ascending (n'') order. Now compute $K(n) \equiv n' - n''$ and iterate. The algorithm reaches 0 (a degenerate case), a constant, or a cycle, depending on the number of digits in k and the value of n.

For a 3-digit number n in base 10, the Kaprekar routine reaches the number 495 in at most six iterations. In base r, there is a unique number $((r-2)/2, r-1, r/2)_r$ to which n converges in at most $(r+2)/2$ iterations Iff r is Even. For any 4-digit number n in base-10, the routine terminates on the number 6174 after seven or fewer steps (where it enters the 1-cycle $K(6174) = 6174$).

2. 0, 0, 9, 21, {(45), (49)}, ...,
3. 0, 0, (32, 52), 184, (320, 580, 484), ...,
4. 0, 30, {201, (126, 138)}, (570, 765), {(2550), (3369), (3873)}, ...,
5. 8, (48, 72), 392, (1992, 2616, 2856, 2232), (7488, 10712, 9992, 13736, 11432), ...,
6. 0, 105, (430, 890, 920, 675, 860, 705), {5600, (4305, 5180)}, {(27195), (33860), (42925), (16840, 42745, 35510)}, ...,
7. 0, (144, 192), (1068, 1752, 1836), (9936, 15072, 13680, 13008, 10608), (55500, 89112, 91800, 72012, 91212, 77388), ...,
8. 21, 252, {(1589, 3178, 2723), (1022, 3122, 3290, 2044, 2212)}, {(17892, 20475), (21483, 25578, 26586, 21987)}, ...,
9. (16, 48), (320, 400), {(2256, 5312, 3856), (3712, 5168, 5456)}, {41520, (34960, 40080, 55360, 49520, 42240)}, ...,
10. 0, 495, 6174, {(53955, 59994), (61974, 82962, 75933, 63954), (62964, 71973, 83952, 74943)}, ...,

see also 196-Algorithm, Kaprekar Number, RATS Sequence

References

Eldridge, K. E. and Sagong, S. "The Determination of Kaprekar Convergence and Loop Convergence of All 3-Digit Numbers." *Amer. Math. Monthly* **95**, 105–112, 1988.

Kaprekar, D. R. "An Interesting Property of the Number 6174." *Scripta Math.* **15**, 244-245, 1955.

Trigg, C. W. "All Three-Digit Integers Lead to..." *The Math. Teacher*, **67**, 41–45, 1974.

Young, A. L. "A Variation on the 2-digit Kaprekar Routine." *Fibonacci Quart.* **31**, 138–145, 1993.

Kaps-Rentrop Methods

A generalization of the Runge-Kutta Method for solution of Ordinary Differential Equations, also called Rosenbrock Methods.

see also Runge-Kutta Method

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 730–735, 1992.

Kapteyn Series

A series of the form

$$\sum_{n=0}^{\infty} \alpha_n J_{\nu+n}[(\nu+n)z],$$

where $J_n(z)$ is a Bessel Function of the First Kind. Examples include Kapteyn's original series

$$\frac{1}{1-z} = 1 + 2\sum_{n=1}^{\infty} J_n(nz)$$

and

$$\frac{z^2}{2(1-z^2)} = \sum_{n=1}^{\infty} J_{2n}(2nz).$$

see also Bessel Function of the First Kind, Neumann Series (Bessel Function)

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1473, 1980.

Karatsuba Multiplication

It is possible to perform Multiplication of Large Numbers in (many) fewer operations than the usual brute-force technique of "long multiplication." As discovered by Karatsuba and Ofman (1962), Multiplication of two n-Digit numbers can be done with a Bit Complexity of less than n^2 using identities of the form

$$\begin{aligned}(a + b\cdot 10^n)(c + d\cdot 10^n)\\ = ac + [(a+b)(c+d) - ac - bd]10^n + bd\cdot 10^{2n}. \quad (1)\end{aligned}$$

Proceeding recursively then gives Bit Complexity $\mathcal{O}(n^{\lg 3})$, where $\lg 3 = 1.58\ldots < 2$ (Borwein *et al.* 1989). The best known bound is $\mathcal{O}(n \lg n \lg\lg n)$ steps for $n \gg 1$ (Schönhage and Strassen 1971, Knuth 1981). However, this Algorithm is difficult to implement, but a procedure based on the Fast Fourier Transform is straightforward to implement and gives Bit Complexity $\mathcal{O}((\lg n)^{2+\epsilon} n)$ (Brigham 1974, Borodin and Munro 1975, Knuth 1981, Borwein *et al.* 1989).

As a concrete example, consider Multiplication of two numbers each just two "digits" long in base w,

$$N_1 = a_0 + a_1 w \quad (2)$$

$$N_2 = b_0 + b_1 w, \quad (3)$$

then their Product is

$$\begin{aligned}P &\equiv N_1 N_2\\ &= a_0 b_0 + (a_0 b_1 + a_1 b_0)w + a_1 b_1 w^2\\ &= p_0 + p_1 w + p_2 w^2. \quad (4)\end{aligned}$$

Instead of evaluating products of individual digits, now write

$$q_0 = a_0 b_0 \tag{5}$$

$$q_1 = (a_0 + a_1)(b_0 + b_1) \tag{6}$$

$$q_2 = a_1 b_1. \tag{7}$$

The key term is q_1, which can be expanded, regrouped, and written in terms of the p_j as

$$q_1 = p_1 + p_0 + p_2. \tag{8}$$

However, since $p_0 = q_0$, and $p_2 = q_2$, it immediately follows that

$$p_0 = q_0 \tag{9}$$

$$p_1 = q_1 - q_0 - q_2 \tag{10}$$

$$p_2 = q_2, \tag{11}$$

so the three "digits" of p have been evaluated using three multiplications rather than four. The technique can be generalized to multidigit numbers, with the trade-off being that more additions and subtractions are required.

Now consider four-"digit" numbers

$$N_1 = a_0 + a_1 w + a_2 w^2 + a_3 w^3, \tag{12}$$

which can be written as a two-"digit" number represented in the base w^2,

$$N_1 = (a_0 + a_1 w) + (a_2 + a_3 w) * w^2. \tag{13}$$

The "digits" in the new base are now

$$a_0' = a_0 + a_1 w \tag{14}$$

$$a_1' = a_2 + a_3 w, \tag{15}$$

and the Karatsuba algorithm can be applied to N_1 and N_2 in this form. Therefore, the Karatsuba algorithm is not restricted to multiplying two-digit numbers, but more generally expresses the multiplication of two numbers in terms of multiplications of numbers of half the size. The asymptotic speed the algorithm obtains by recursive application to the smaller required subproducts is $\mathcal{O}(n^{\lg 3})$ (Knuth 1981).

When this technique is recursively applied to multidigit numbers, a point is reached in the recursion when the overhead of additions and subtractions makes it more efficient to use the usual $\mathcal{O}(n^2)$ MULTIPLICATION algorithm to evaluate the partial products. The most efficient overall method therefore relies on a combination of Karatsuba and conventional multiplication.

see also COMPLEX MULTIPLICATION, MULTIPLICATION, STRASSEN FORMULAS

References

Borodin, A. and Munro, I. *The Computational Complexity of Algebraic and Numeric Problems.* New York: American Elsevier, 1975.

Borwein, J. M.; Borwein, P. B.; and Bailey, D. H. "Ramanujan, Modular Equations, and Approximations to Pi, or How to Compute One Billion Digits of Pi." *Amer. Math. Monthly* **96**, 201–219, 1989.

Brigham, E. O. *The Fast Fourier Transform.* Englewood Cliffs, NJ: Prentice-Hall, 1974.

Brigham, E. O. *Fast Fourier Transform and Applications.* Englewood Cliffs, NJ: Prentice-Hall, 1988.

Cook, S. A. *On the Minimum Computation Time of Functions.* Ph.D. Thesis. Cambridge, MA: Harvard University, pp. 51–77, 1966.

Hollerbach, U. "Fast Multiplication & Division of Very Large Numbers." `sci.math.research` posting, Jan. 23, 1996.

Karatsuba, A. and Ofman, Yu. "Multiplication of Many-Digital Numbers by Automatic Computers." *Doklady Akad. Nauk SSSR* **145**, 293–294, 1962. Translation in *Physics–Doklady* **7**, 595–596, 1963.

Knuth, D. E. *The Art of Computing, Vol. 2: Seminumerical Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, pp. 278–286, 1981.

Schönhage, A. and Strassen, V. "Schnelle Multiplikation Grosser Zahlen." *Computing* **7**, 281–292, 1971.

Toom, A. L. "The Complexity of a Scheme of Functional Elements Simulating the Multiplication of Integers." *Dokl. Akad. Nauk SSSR* **150**, 496–498, 1963. English translation in *Soviet Mathematics* **3**, 714–716, 1963.

Zuras, D. "More on Squaring and Multiplying Large Integers." *IEEE Trans. Comput.* **43**, 899–908, 1994.

Katona's Problem

Find the minimum number $f(n)$ of subsets in a SEPARATING FAMILY for a SET of n elements, where a SEPARATING FAMILY is a SET of SUBSETS in which each pair of adjacent elements is found separated, each in one of two disjoint subsets. For example, the 26 letters of the alphabet can be separated by a family of nine:

$$\begin{matrix} (abcdefghi) & (jklmnopqr) & (stuvwxyz) \\ (abcjklstu) & (defmnovwx) & (ghipqryz) \\ (adgjmpsvy) & (behknqtwz) & (cfilorux) \end{matrix}.$$

The problem was posed by Katona (1973) and solved by C. Mao-Cheng in 1982,

$$f(n) = \min\left\{2p + 3\left\lceil \log_3\left(\frac{n}{2^p}\right)\right\rceil : p = 0, 1, 2\right\},$$

where $\lceil x \rceil$ is the CEILING FUNCTION. $f(n)$ is nondecreasing, and the values for $n = 1, 2, \ldots$ are 0, 2, 3, 4, 5, 5, 6, 6, 6, 7, ... (Sloane's A07600). The values at which $f(n)$ increases are 1, 2, 3, 4, 5, 7, 10, 13, 19, 28, 37, ... (Sloane's A007601), so $f(26) = 9$, as illustrated in the preceding example.

see also SEPARATING FAMILY

References

Honsberger, R. "Cai Mao-Cheng's Solution to Katona's Problem on Families of Separating Subsets." Ch. 18 in *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 224–239, 1985.

Katona, G. O. H. "Combinatorial Search Problem." In *A Survey of Combinatorial Theory* (Ed. J. N. Srivasta *et al.*). Amsterdam, Netherlands: North-Holland, pp. 285–308, 1973.

Sloane, N. J. A. Sequences A007600/M0456 and A007601/M0525 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Kauffman Polynomial F

A semi-oriented 2-variable Knot Polynomial defined by

$$F_L(a,z) = a^{-w(L)}\langle|L|\rangle, \qquad (1)$$

where L is an oriented Link Diagram, $w(L)$ is the Writhe of L, $|L|$ is the unoriented diagram corresponding to L, and $\langle L\rangle$ is the Bracket Polynomial. It was developed by Kauffman by extending the BLM/Ho Polynomial Q to two variables, and satisfies

$$F(1,x) = Q(x). \qquad (2)$$

The Kauffman Polynomial is a generalization of the Jones Polynomial $V(t)$ since it satisfies

$$V(t) = F(-t^{-3/4}, t^{-1/4} + t^{1/4}), \qquad (3)$$

but its relationship to the HOMFLY Polynomial is not well understood. In general, it has more terms than the HOMFLY Polynomial, and is therefore more powerful for discriminating Knots. It is a semi-oriented Polynomial because changing the orientation only changes F by a Power of a. In particular, suppose L^* is obtained from L by reversing the orientation of component k, then

$$F_{L^*} = a^{4\lambda}F_L, \qquad (4)$$

where λ is the Linking Number of k with $L-k$ (Lickorish and Millett 1988). F is unchanged by Mutation.

$$F_{L_1+F_{L_2}} = F(L_1)F(L_2) \qquad (5)$$

$$F_{L_1\cup L_2} = [(a^{-1}+a)x^{-1}-1]F_{L_1}F_{L_2}. \qquad (6)$$

M. B. Thistlethwaite has tabulated the Kauffman 2-variable Polynomial for Knots up to 13 crossings.

References

Lickorish, W. B. R. and Millett, B. R. "The New Polynomial Invariants of Knots and Links." *Math. Mag.* **61**, 1–23, 1988.

Stoimenow, A. "Kauffman Polynomials." `http://www.informatik.hu-berlin.de/~stoimeno/ptab/k10.html`.

Weisstein, E. W. "Knots and Links." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Knots.m`.

Kauffman Polynomial X

A 1-variable Knot Polynomial denoted X or $\mathcal{L}$.

$$\mathcal{L}_L(A) \equiv (-A^3)^{-w(L)}\langle L\rangle, \qquad (1)$$

where $\langle L\rangle$ is the Bracket Polynomial and $w(L)$ is the Writhe of L. This Polynomial is invariant under Ambient Isotopy, and relates Mirror Images by

$$\mathcal{L}_{L^*} = \mathcal{L}_L(A^{-1}). \qquad (2)$$

It is identical to the Jones Polynomial with the change of variable

$$\mathcal{L}(t^{-1/4}) = V(t). \qquad (3)$$

The X Polynomial of the Mirror Image K^* is the same as for K but with A replaced by A^{-1}.

References

Kauffman, L. H. *Knots and Physics.* Singapore: World Scientific, p. 33, 1991.

Kei

The Imaginary Part of

$$e^{-\nu\pi i/2}K_\nu(xe^{\pi i/4}) = \mathrm{ker}_\nu(x) + i\,\mathrm{kei}_\nu(x).$$

see also Bei, Ber, Ker, Kelvin Functions

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Kelvin Functions." §9.9 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 379–381, 1972.

Keith Number

A Keith number is an n-digit Integer N such that if a Fibonacci-like sequence (in which each term in the sequence is the sum of the n previous terms) is formed with the first n terms taken as the decimal digits of the number N, then N itself occurs as a term in the sequence. For example, 197 is a Keith number since it generates the sequence 1, 9, 7, 17, 33, 57, 107, 197, ... (Keith). Keith numbers are also called Repfigit Numbers.

There is no known general technique for finding Keith numbers except by exhaustive search. Keith numbers are much rarer than the Primes, with only 52 Keith numbers with < 15 digits: 14, 19, 28, 47, 61, 75, 197, 742, 1104, 1537, 2208, 2580, 3684, 4788, 7385, 7647, 7909, ... (Sloane's A007629). In addition, three 15-digit Keith numbers are known (Keith 1994). It is not known if there are an Infinite number of Keith numbers.

References

Esche, H. A. "Non-Decimal Replicating Fibonacci Digits." *J. Recr. Math.* **26**, 193–194, 1994.

Heleen, B. "Finding Repfigits—A New Approach." *J. Recr. Math.* **26**, 184–187, 1994.

Keith, M. "Repfigit Numbers." *J. Recr. Math.* **19**, 41–42, 1987.

Keith, M. "All Repfigit Numbers Less than 100 Billion (10^{11})." *J. Recr. Math.* **26**, 181–184, 1994.

Keith, M. "Keith Numbers." `http://users.aol.com/s6sj7gt/mikekeit.htm`.

Robinson, N. M. "All Known Replicating Fibonacci Digits Less than One Thousand Billion (10^{12})." *J. Recr. Math.* **26**, 188–191, 1994.

Shirriff, K. "Computing Replicating Fibonacci Digits." *J. Recr. Math.* **26**, 191–193, 1994.

Sloane, N. J. A. Sequence A007629/M4922 in "An On-Line Version of the Encyclopedia of Integer Sequences."

"Table: Repfigit Numbers (Base 10*) Less than 10^{15}." *J. Recr. Math.* **26**, 195, 1994.

Keller's Conjecture

Keller conjectured that tiling an n-D space with n-D HYPERCUBES of equal size yields an arrangement in which at least two hypercubes have an entire $(n-1)$-D "side" in common. The CONJECTURE has been proven true for $n = 1$ to 6, but disproven for $n \geq 10$.

References

Cipra, B. "If You Can't See It, Don't Believe It." *Science* **259**, 26–27, 1993.

Cipra, B. *What's Happening in the Mathematical Sciences, Vol. 1.* Providence, RI: Amer. Math. Soc., pp. 24, 1993.

Kelvin's Conjecture

What space-filling arrangement of similar polyhedral cells of equal volume has minimal surface AREA? Kelvin proposed the 14-sided TRUNCATED OCTAHEDRON. Wearie and Phelan (1994) discovered another 14-sided POLYHEDRON that has 3% less SURFACE AREA.

References

Gray, J. "Parsimonious Polyhedra." *Nature* **367**, 598–599, 1994.

Wearie, D. and Phelan, R. "A Counter-Example to Kelvin's Conjecture on Minimal Surfaces." *Philos. Mag. Let.* **69**, 107–110, 1994.

Kelvin Functions

Kelvin defined the Kelvin functions BEI and BER according to

$$J_\nu(xe^{3\pi i/4}) = \operatorname{ber}_\nu(x) + i\operatorname{bei}_\nu(x), \qquad (1)$$

where $J_\nu(s)$ is a BESSEL FUNCTION OF THE FIRST KIND, and the functions KEI and KER by

$$e^{-\nu\pi i/2}K_\nu(xe^{\pi i/4}) = \ker_\nu(x) + i\operatorname{kei}_\nu(x), \qquad (2)$$

where $K_\nu(x)$ is a MODIFIED BESSEL FUNCTION OF THE SECOND KIND. For the special case $\nu = 0$,

$$J_0(i\sqrt{i}\,x) = J_0(\tfrac{1}{2}\sqrt{2}(i-1)x) \equiv \operatorname{ber}(x) + i\operatorname{bei}(x). \qquad (3)$$

see also BEI, BER, KEI, KER

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Kelvin Functions." §9.9 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 379–381, 1972.

Spanier, J. and Oldham, K. B. "The Kelvin Functions." Ch. 55 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 543–554, 1987.

Kelvin Transformation

The transformation

$$v(x_1',\ldots,x_n') = \left(\frac{a}{r'}\right)^{n-2} u\left(\frac{a^2x_1'}{r'^2},\ldots,\frac{a^2x_n'}{r'^2}\right),$$

where

$$r'^2 = x_1'^2 + \ldots + x_n'^2.$$

If $u(x_1,\ldots,x_n)$ is a HARMONIC FUNCTION on a DOMAIN D of $\mathbb{R}^n$ (with $n \geq 3$), then $v(x_1',\ldots,x_n')$ is HARMONIC on D'.

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 623, 1980.

Kempe Linkage

A double rhomboid LINKAGE which gives rectilinear motion from circular without an inversion.

References

Rademacher, H. and Toeplitz, O. *The Enjoyment of Mathematics: Selections from Mathematics for the Amateur.* Princeton, NJ: Princeton University Press, pp. 126–127, 1957.

Kepler Conjecture

In 1611, Kepler proposed that close packing (cubic or hexagonal) is the densest possible SPHERE PACKING (has the greatest η), and this assertion is known as the Kepler conjecture. Finding the densest (not necessarily periodic) packing of spheres is known as the KEPLER PROBLEM.

A putative proof of the Kepler conjecture was put forward by W.-Y. Hsiang (Hsiang 1992, Cipra 1993), but was subsequently determined to be flawed (Conway *et al.* 1994, Hales 1994). According to J. H. Conway, nobody who has read Hsiang's proof has any doubts about its validity: it is nonsense.

see also DODECAHEDRAL CONJECTURE, KEPLER PROBLEM

References

Cipra, B. "Gaps in a Sphere Packing Proof?" *Science* **259**, 895, 1993.

Conway, J. H.; Hales, T. C.; Muder, D. J.; and Sloane, N. J. A. "On the Kepler Conjecture." *Math. Intel.* **16**, 5, Spring 1994.

Eppstein, D. "Sphere Packing and Kissing Numbers." `http://www.ics.uci.edu/~eppstein/junkyard/spherepack.html`.

Hales, T. C. "The Sphere Packing Problem." *J. Comput. Appl. Math.* **44**, 41–76, 1992.

Hales, T. C. "Remarks on the Density of Sphere Packings in 3 Dimensions." *Combinatori* **13**, 181–197, 1993.

Hales, T. C. "The Status of the Kepler Conjecture." *Math. Intel.* **16**, 47–58, Summer 1994.

Hales, T. C. 'The Kepler Conjecture." `http://www.math.lsa.umich.edu/~hales/kepler.html`.

Hsiang, W.-Y. "On Soap Bubbles and Isoperimetric Regions in Noncompact Symmetrical Spaces. 1." *Tôhoku Math. J.* **44**, 151–175, 1992.

Hsiang, W.-Y. "A Rejoinder to Hales's Article." *Math. Intel.* **17**, 35–42, Winter 1995.

Kepler's Equation

Let M be the mean anomaly and E the ECCENTRIC ANOMALY of a body orbiting on an ELLIPSE with ECCENTRICITY e, then

$$M = E - e \sin E. \tag{1}$$

For M not a multiple of π, Kepler's equation has a unique solution, but is a TRANSCENDENTAL EQUATION and so cannot be inverted and solved directly for E given an arbitrary M. However, many algorithms have been derived for solving the equation as a result of its importance in celestial mechanics.

Writing a E as a POWER SERIES in e gives

$$E = M + \sum_{n=1}^{\infty} a_n e^n, \tag{2}$$

where the coefficients are given by the LAGRANGE INVERSION THEOREM as

$$a_n = \frac{1}{2^{n-1} n!} \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k \binom{n}{k} (n-2k)^{n-1} \sin[(n-2k)M] \tag{3}$$

(Wintner 1941, Moulton 1970, Henrici 1974, Finch). Surprisingly, this series diverges for

$$e > 0.6627434193\ldots, \tag{4}$$

a value known as the LAPLACE LIMIT. In fact, E converges as a GEOMETRIC SERIES with ratio

$$r = \frac{e}{1+\sqrt{1+e^2}} \exp(\sqrt{1+e^2}\,) \tag{5}$$

(Finch).

There is also a series solution in BESSEL FUNCTIONS OF THE FIRST KIND,

$$E = M + \sum_{n=1}^{\infty} \frac{2}{n} J_n(ne) \sin(nM). \tag{6}$$

This series converges for all $e < 1$ like a GEOMETRIC SERIES with ratio

$$r = \frac{e}{1+\sqrt{1-e^2}} \exp(\sqrt{1-e^2}\,). \tag{7}$$

The equation can also be solved by letting ψ be the ANGLE between the planet's motion and the direction PERPENDICULAR to the RADIUS VECTOR. Then

$$\tan \psi = \frac{e \sin E}{\sqrt{1-e^2}}. \tag{8}$$

Alternatively, we can define e in terms of an intermediate variable ϕ

$$e \equiv \sin \phi, \tag{9}$$

then

$$\sin[\tfrac{1}{2}(v - E)] = \sqrt{\frac{r}{p}} \sin(\tfrac{1}{2}\phi) \sin v \tag{10}$$

$$\sin[\tfrac{1}{2}(v + E)] = \sqrt{\frac{r}{p}} \cos(\tfrac{1}{2}\phi) \sin v. \tag{11}$$

Iterative methods such as the simple

$$E_{i+1} = M + e \sin E_i \tag{12}$$

with $E_0 = 0$ work well, as does NEWTON'S METHOD,

$$E_{i+1} = E_i + \frac{M + e \sin E_i - E_i}{1 - e \cos E_i}. \tag{13}$$

In solving Kepler's equation, Stieltjes required the solution to

$$e^x(x-1) = e^{-x}(x+1), \tag{14}$$

which is 1.1996678640257734... (Goursat 1959, Le Lionnais 1983).

see also ECCENTRIC ANOMALY

References

Danby, J. M. *Fundamentals of Celestial Mechanics, 2nd ed., rev. ed.* Richmond, VA: Willmann-Bell, 1988.

Dörrie, H. "The Kepler Equation." §81 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 330–334, 1965.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/lpc/lpc.html`.

Goldstein, H. *Classical Mechanics, 2nd ed.* Reading, MA: Addison-Wesley, pp. 101–102 and 123–124, 1980.

Goursat, E. *A Course in Mathematical Analysis, Vol. 2.* New York: Dover, p. 120, 1959.

Henrici, P. *Applied and Computational Complex Analysis, Vol. 1: Power Series–Integration–Conformal Mapping–Location of Zeros.* New York: Wiley, 1974.

Ioakimids, N. I. and Papadakis, K. E. "A New Simple Method for the Analytical Solution of Kepler's Equation." *Celest. Mech.* **35**, 305–316, 1985.

Ioakimids, N. I. and Papadakis, K. E. "A New Class of Quite Elementary Closed-Form Integrals Formulae for Roots of Nonlinear Systems." *Appl. Math. Comput.* **29**, 185–196, 1989.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 36, 1983.

Marion, J. B. and Thornton, S. T. "Kepler's Equations." §7.8 in *Classical Dynamics of Particles & Systems, 3rd ed.* San Diego, CA: Harcourt Brace Jovanovich, pp. 261–266, 1988.

Moulton, F. R. *An Introduction to Celestial Mechanics, 2nd rev. ed.* New York: Dover, pp. 159–169, 1970.

Siewert, C. E. and Burniston, E. E. "An Exact Analytical Solution of Kepler's Equation." *Celest. Mech.* **6**, 294–304, 1972.

Wintner, A. *The Analytic Foundations of Celestial Mechanics.* Princeton, NJ: Princeton University Press, 1941.

Kepler's Folium

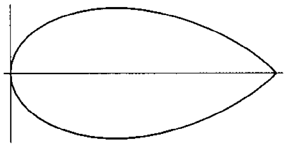

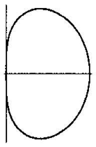

The curve with implicit equation

$$[(x-b)^2+y^2][x(x-b)+y^2]-4a(x-b)y^2.$$

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 71–72, 1993.

Kepler-Poinsot Solid

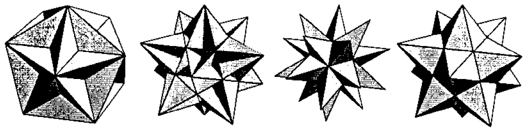

The Kepler-Poinsot solids are the four regular CONCAVE POLYHEDRA with intersecting facial planes. They are composed of regular CONCAVE POLYGONS and were unknown to the ancients. Kepler discovered two of them about 1619. These two were subsequently rediscovered by Poinsot, who also discovered the other two, in 1809. As shown by Cauchy, they are stellated forms of the DODECAHEDRON and ICOSAHEDRON.

The Kepler-Poinsot solids, illustrated above, are known as the GREAT DODECAHEDRON, GREAT ICOSAHEDRON, GREAT STELLATED DODECAHEDRON, and SMALL STELLATED DODECAHEDRON. Cauchy (1813) proved that these four exhaust all possibilities for regular star polyhedra (Ball and Coxeter 1987).

A table listing these solids, their DUALS, and COMPOUNDS is given below.

Polyhedron	Dual
great dodecahedron	small stellated dodec.
great Icosahedron	great stellated dodec.
great stellated dodec.	great icosahedron
small stellated dodec.	great dodecahedron

Polyhedron	Compound
great dodecahedron	great dodecahedron-small stellated dodec.
great icosahedron	great icosahedron-great stellated dodec.
great stellated dodec.	great icosahedron-great stellated dodec.
small stellated dodec.	great dodecahedron-small stellated dodec.

The polyhedra $\{\frac{5}{2},5\}$ and $\{5,\frac{5}{2}\}$ fail to satisfy the POLYHEDRAL FORMULA

$$V-E+F=2,$$

where V is the number of faces, E the number of edges, and F the number of faces, despite the fact that formula holds for all ordinary polyhedra (Ball and Coxeter 1987). This unexpected result led none less than Schläfli (1860) to conclude that they could not exist.

In 4-D, there are 10 Kepler-Poinsot solids, and in n-D with $n \geq 5$, there are none. In 4-D, nine of the solids have the same VERTICES as $\{3,3,5\}$, and the tenth has the same as $\{5,3,3\}$. Their SCHLÄFLI SYMBOLS are $\{\frac{5}{2},5,3\}$, $\{3,5,\frac{5}{2}\}$, $\{5,\frac{5}{2},5\}$, $\{\frac{5}{2},3,5\}$, $\{5,3,\frac{5}{2}\}$, $\{\frac{5}{2},5,\frac{5}{2}\}$, $\{5,\frac{5}{2},3\}$, $\{3,\frac{5}{2},5\}$, $\{\frac{5}{2},3,3\}$, and $\{3,3,\frac{5}{2}\}$.

Coxeter *et al.* (1954) have investigated star "Archimedean" polyhedra.

see also ARCHIMEDEAN SOLID, DELTAHEDRON, JOHNSON SOLID, PLATONIC SOLID, POLYHEDRON COMPOUND, UNIFORM POLYHEDRON

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 144–146, 1987.

Cauchy, A. L. "Recherches sur les polyèdres." *J. de l'École Polytechnique* **9**, 68–86, 1813.

Coxeter, H. S. M.; Longuet-Higgins, M. S.; and Miller, J. C. P. "Uniform Polyhedra." *Phil. Trans. Roy. Soc. London Ser. A* **246**, 401–450, 1954.

Pappas, T. "The Kepler-Poinsot Solids." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 113, 1989.

Quaisser, E. "Regular Star-Polyhedra." Ch. 5 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 56–62, 1986.

Schläfli. *Quart. J. Math.* **3**, 66–67, 1860.

Kepler Problem

Finding the densest not necessarily periodic SPHERE PACKING.

see also KEPLER CONJECTURE, SPHERE PACKING

Kepler Solid

see KEPLER-POINSOT SOLID

Ker

The REAL PART of

$$e^{-\nu\pi i/2}K_\nu(xe^{\pi i/4}) = \mathrm{ker}_\nu(x) + i\,\mathrm{kei}_\nu(x),$$

where $K_\nu(x)$ is a MODIFIED BESSEL FUNCTION OF THE SECOND KIND.

see also BEI, BER, KEI, KELVIN FUNCTIONS

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Kelvin Functions." §9.9 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 379–381, 1972.

Keratoid Cusp

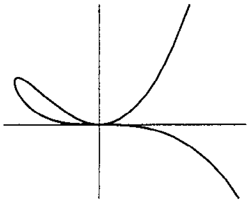

The PLANE CURVE given by the Cartesian equation

$$y^2 = x^2 y + x^5.$$

References
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 72, 1989.

Kernel (Integral)

The function $K(\alpha, t)$ in an INTEGRAL or INTEGRAL TRANSFORM

$$g(\alpha) = \int_a^b f(t) K(\alpha, t)\, dt.$$

see also BERGMAN KERNEL, POISSON KERNEL

Kernel (Linear Algebra)

see NULLSPACE

Kernel Polynomial

The function

$$K_n(x_0, x) = \overline{K_n(x, x_0)} = K_n(\bar{x}, \bar{x}_0)$$

which is useful in the study of many POLYNOMIALS.

References
Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., 1975.

Kervaire's Characterization Theorem

Let G be a GROUP, then there exists a piecewise linear KNOT K^{n-2} in $\mathbb{S}^n$ for $n \geq 5$ with $G = \pi_1(\mathbb{S}^n - K)$ IFF G satisfies

1. G is finitely presentable,
2. The Abelianization of G is infinite cyclic,
3. The normal closure of some single element is all of G,
4. $H_2(G) = 0$; the second homology of the group is trivial.

References
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 350–351, 1976.

Ket

A CONTRAVARIANT VECTOR, denoted $|\psi\rangle$. The ket is DUAL to the COVARIANT BRA 1-VECTOR $\langle\psi|$. Taken together, the BRA and ket form an ANGLE BRACKET (bra+ket = bracket) $\langle\psi|\psi\rangle$. The ket is commonly encountered in quantum mechanics.

see also ANGLE BRACKET, BRA, BRACKET PRODUCT, CONTRAVARIANT VECTOR, COVARIANT VECTOR, DIFFERENTIAL k-FORM, ONE-FORM

Khinchin Constant

see KHINTCHINE'S CONSTANT

Khintchine's Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

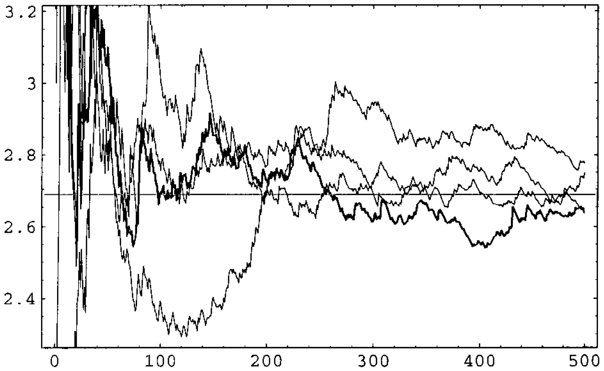

Let

$$x = [q_0, q_1, \ldots] = q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \cfrac{1}{q_3 + \ldots}}} \tag{1}$$

be the SIMPLE CONTINUED FRACTION of a REAL NUMBER x, where the numbers q_i are called PARTIAL QUOTIENTS. Khintchine (1934) considered the limit of the GEOMETRIC MEAN

$$G_n(x) = (q_1 q_2 \cdots q_n)^{1/n} \tag{2}$$

as $n \to \infty$. Amazingly enough, this limit is a constant *independent* of x—except if x belongs to a set of MEASURE 0–given by

$$K = 2.685452001\ldots \tag{3}$$

(Sloane's A002210), as proved in Kac (1959). The values $G_n(x)$ are plotted above for $n = 1$ to 500 and $x = \pi$, $1/\pi$, $\sin 1$, the EULER-MASCHERONI CONSTANT γ, and the COPELAND-ERDŐS CONSTANT. REAL NUMBERS x for which $\lim_{n\to\infty} G_n(x) \neq K$ include $x = e$, $\sqrt{2}$, $\sqrt{3}$, and the GOLDEN RATIO ϕ, all of which have periodic PARTIAL QUOTIENTS, plotted below.

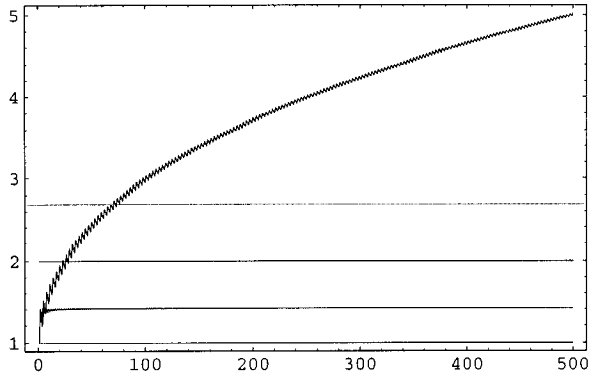

The CONTINUED FRACTION for K is [2, 1, 2, 5, 1, 1, 2, 1, 1, ...] (Sloane's A002211). It is not known if K is IRRATIONAL, let alone TRANSCENDENTAL. Bailey *et al.* (1995) have computed K to 7350 DIGITS.

Explicit expressions for K include

$$K = \prod_{n=1}^{\infty} \left[1 + \frac{1}{n(n+2)}\right]^{\ln n/\ln 2} \tag{4}$$

$$\ln 2 \ln K = \tfrac{1}{12}\pi^2 + \tfrac{1}{2}(\ln 2)^2 + \int_0^\pi \frac{\ln(\theta|\cot\theta|)\,d\theta}{\theta} \tag{5}$$

$$\ln K = \frac{1}{\ln 2} \sum_{m=1}^{\infty} \frac{h_{m-1}}{m}[\zeta(2m) - 1], \tag{6}$$

where $\zeta(z)$ is the RIEMANN ZETA FUNCTION and

$$h_m = \sum_{j=1}^{m} \frac{(-1)^{j-1}}{j} \tag{7}$$

(Shanks and Wrench 1959). Gosper gave

$$\ln K = \frac{1}{\ln 2} \sum_{j=2}^{\infty} \frac{(-1)^j(2-2^j)\zeta'(j)}{j}, \tag{8}$$

where $\zeta'(z)$ is the DERIVATIVE of the RIEMANN ZETA FUNCTION. An extremely rapidly converging sum also due to Gosper is

$$\begin{aligned} \ln K = \frac{1}{\ln 2} \sum_{k=0}^{\infty} \Bigg\{ & -\ln(k+1)[\ln(k+3) \\ & -2\ln(k+2) + \ln(k+1)] \\ & -\frac{(-1)^k(2-2^{k+2})}{k+2}\left[\frac{\ln(k+1)}{(k+1)^{k+2}} - \zeta'(k+2, k+2)\right] \\ & + \ln(k+1)\left[\sum_{s=1}^{k+2} \frac{(-1)^s(2-2^s)}{(k+1)^s s}\right]\Bigg\}, \end{aligned} \tag{9}$$

where $\zeta(s, a)$ is the HURWITZ ZETA FUNCTION.

Khintchine's constant is also given by the integral

$$\ln 2 \ln(\tfrac{1}{2}K) = \int_0^1 \frac{1}{x(1+x)} \ln\left[\frac{\pi x(1-x^2)}{\sin(\pi x)}\right] dx. \tag{10}$$

If P_n/Q_n is the nth CONVERGENT of the CONTINUED FRACTION of x, then

$$\lim_{n\to\infty} (Q_n)^{1/n} = \lim_{n\to\infty} \left(\frac{P_n}{x}\right)^{1/n} = e^{\pi^2/(12\ln 2)} \approx 3.27582 \tag{11}$$

for almost all REAL x (Lévy 1936, Finch). This number is sometimes called the LÉVY CONSTANT, and the argument of the exponential is sometimes called the KHINTCHINE-LÉVY CONSTANT.

Define the following quantity in terms of the kth partial quotient q_k,

$$M(s, n, x) = \left(\frac{1}{n}\sum_{k=1}^{n} {q_k}^s\right)^{1/s}. \tag{12}$$

Then

$$\lim_{n\to\infty} M(1, n, x) = \infty \tag{13}$$

for almost all real x (Khintchine, Knuth 1981, Finch), and

$$M(1, n, x) \sim \mathcal{O}(\ln n). \tag{14}$$

Furthermore, for $s < 1$, the limiting value

$$\lim_{n\to\infty} M(s, n, x) = K(s) \tag{15}$$

exists and is a constant $K(s)$ with probability 1 (Rockett and Szüsz 1992, Khintchine 1997).

see also CONTINUED FRACTION, CONVERGENT, KHINTCHINE-LÉVY CONSTANT, LÉVY CONSTANT, PARTIAL QUOTIENT, SIMPLE CONTINUED FRACTION

References
Bailey, D. H.; Borwein, J. M.; and Crandall, R. E. "On the Khintchine Constant." *Math. Comput.* **66**, 417–431, 1997.
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/khntchn/khntchn.html`.
Kac, M. *Statistical Independence and Probability, Analysis and Number Theory.* Providence, RI: Math. Assoc. Amer., 1959.
Khinchin, A. Ya. *Continued Fractions.* New York: Dover, 1997.
Knuth, D. E. Exercise 24 in *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, p. 604, 1981.
Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 46, 1983.
Lehmer, D. H. "Note on an Absolute Constant of Khintchine." *Amer. Math. Monthly* **46**, 148–152, 1939.
Phillipp, W. "Some Metrical Theorems in Number Theory." *Pacific J. Math.* **20**, 109–127, 1967.
Plouffe, S. "Plouffe's Inverter: Table of Current Records for the Computation of Constants." `http://lacim.uqam.ca/pi/records.html`.

Rockett, A. M. and Szüsz, P. *Continued Fractions.* Singapore: World Scientific, 1992.
Shanks, D. and Wrench, J. W. "Khintchine's Constant." *Amer. Math. Monthly* **66**, 148–152, 1959.
Sloane, N. J. A. Sequences A002210/M1564 and A002211/M0118 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Vardi, I. "Khinchin's Constant." §8.4 in *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 163–171, 1991.
Wrench, J. W. "Further Evaluation of Khintchine's Constant." *Math. Comput.* **14**, 370–371, 1960.

Khintchine-Lévy Constant

A constant related to KHINTCHINE'S CONSTANT defined by

$$KL \equiv \frac{\pi^2}{12\ln 2} = 1.1865691104\ldots.$$

see also KHINTCHINE'S CONSTANT, LÉVY CONSTANT

References

Plouffe, S. "Khintchine-Levy Constant." `http://lacim.uqam.ca/piDATA/klevy.txt`.

Khovanski's Theorem

If $f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$ are exponential polynomials, then $\{x \in \mathbb{R}^n : f_1(x) = \cdots f_n(x) = 0\}$ has finitely many connected components.

References

Marker, D. "Model Theory and Exponentiation." *Not. Amer. Math. Soc.* **43**, 753–759, 1996.

Kiepert's Conics

see KIEPERT'S HYPERBOLA, KIEPERT'S PARABOLA

Kiepert's Hyperbola

A curve which is related to the solution of LEMOINE'S PROBLEM and its generalization to ISOSCELES TRIANGLES constructed on the sides of a given TRIANGLE. The VERTICES of the constructed TRIANGLES are

$$A' = -\sin\phi : \sin(C+\phi) : \sin(B+\phi) \qquad (1)$$

$$B' = \sin(\mathbf{C}+\phi) : -\sin\phi : \sin(A+\phi) \qquad (2)$$

$$C' = \sin(B+\phi) : \sin(A+\phi) : -\sin\phi, \qquad (3)$$

where ϕ is the base ANGLE of the ISOSCELES TRIANGLE. Kiepert showed that the lines connecting the VERTICES of the given TRIANGLE and the corresponding peaks of the ISOSCELES TRIANGLES CONCUR. The TRILINEAR COORDINATES of the point of concurrence are

$$\sin(B+\phi)\sin(C+\phi) : \sin(C+\phi)\sin(A+\phi) : \sin(A+\phi)\sin(B+\phi). \qquad (4)$$

The locus of this point as the base ANGLE varies is given by the curve

$$\frac{\sin(B-C)}{\alpha} + \frac{\sin(C-A)}{\beta} + \frac{\sin(A-B)}{\gamma} = \frac{bc(c^2-c^2)}{\alpha} + \frac{ca(c^2-a^2)}{\beta} + \frac{ab(a^2-b^2)}{\gamma} = 0. \qquad (5)$$

Writing the TRILINEAR COORDINATES as

$$\alpha_i = d_i s_i, \qquad (6)$$

where d_i is the distance to the side opposite α_i of length s_i and using the POINT-LINE DISTANCE FORMULA with (x_0, y_0) written as (x, y),

$$d_i = \frac{|(y_{i+2}-y_{i+1})(x-x_{i+1})}{s_i} - \frac{(x_{i+2}-x_{i+1})(y-y_{i+1})|}{s_i}, \qquad (7)$$

where $y_4 \equiv y_1$ and $y_5 \equiv y_2$ gives the FORMULA

$$\sum_{i=1}^{3} s_{i+1}s_{i+2}(s_{i+1}^2 - s_{i+2}^2) \times \frac{s_i}{(y_{i+2}-y_{i+1})(x-x_{i+1}) - (x_{i+2}-x_{i+1})(y-y_{i+1})} = 0 \qquad (8)$$

$$\sum_{i=1}^{3} \frac{(s_{i+1}^2 - s_{i+2}^2)}{(y_{i+2}-y_{i+1})(x-x_{i+1}) - (x_{i+2}-x_{i+1})(y-y_{i+1})} = 0. \qquad (9)$$

Bringing this equation over a common DENOMINATOR then gives a quadratic in x and y, which is a CONIC SECTION (in fact, a HYPERBOLA). The curve can also be written as $\csc(A+t) : \csc(B+t) : \csc(C+t)$, as t varies over $[-\pi/4, \pi/4]$.

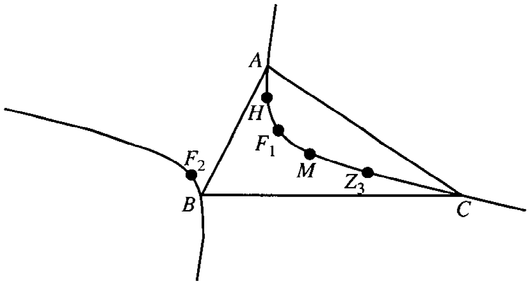

Kiepert's hyperbola passes through the triangle's CENTROID M ($\phi = 0$), ORTHOCENTER H ($\phi = \pi/2$), VERTICES A ($\phi = -\alpha$ if $\alpha \leq \pi/2$ and $\phi = \pi - \alpha$ if $\alpha > \pi/2$), B ($\phi = -\beta$), C ($\phi = -\gamma$), FERMAT POINT F_1 ($\phi = \pi/3$), second ISOGONIC CENTER F_2 ($\phi = -\pi/3$), ISOGONAL CONJUGATE of the BROCARD MIDPOINT ($\phi = \omega$), and BROCARD'S THIRD POINT Z_3 ($\phi = \omega$), where ω is the BROCARD ANGLE (Eddy and Fritsch 1994, p. 193).

The ASYMPTOTES of Kiepert's hyperbola are the SIMSON LINES of the intersections of the BROCARD AXIS with the CIRCUMCIRCLE. Kiepert's hyperbola is a RECTANGULAR HYPERBOLA. In fact, all nondegenerate conics through the VERTICES and ORTHOCENTER of a TRIANGLE are RECTANGULAR HYPERBOLAS the centers of which lie halfway between the Isogonic Centers and on the Nine-Point Circle. The Locus of centers of these Hyperbolas is the Nine-Point Circle.

The Isogonal Conjugate curve of Kiepert's hyperbola is the Brocard Axis. The center of the Incircle of the Triangle constructed from the Midpoints of the sides of a given Triangle lies on Kiepert's hyperbola of the original Triangle.

see also Brocard Angle, Brocard Axis, Brocard Points, Centroid (Triangle), Circumcircle, Isogonal Conjugate, Isogonic Centers, Isosceles Triangle, Lemoine's Problem, Nine-Point Circle, Orthocenter, Simson Line

References

Casey, J. *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions with Numerous Examples, 2nd rev. enl. ed.* Dublin: Hodges, Figgis, & Co., 1893.

Eddy, R. H. and Fritsch, R. "The Conics of Ludwig Kiepert: A Comprehensive Lesson in the Geometry of the Triangle." *Math. Mag.* **67**, 188–205, 1994.

Kelly, P. J. and Merriell, D. "Concentric Polygons." *Amer. Math. Monthly* **71**, 37–41, 1964.

Mineuer, A. "Sur les asymptotes de l'hyperbole de Kiepert." *Mathesis* **49**, 30–33, 1935.

Rigby, J. F. "A Concentrated Dose of Old-Fashioned Geometry." *Math. Gaz.* **57**, 296–298, 1953.

Vandeghen, A. "Some Remarks on the Isogonal and Cevian Transforms. Alignments of Remarkable Points of a Triangle." *Amer. Math. Monthly* **72**, 1091–1094, 1965.

Kiepert's Parabola

Let three similar Isosceles Triangles $\Delta A'BC$, $\Delta AB'C$, and $\Delta ABC'$ be constructed on the sides of a Triangle ΔABC. Then the Envelope of the axis of the Triangles ΔABC and $\Delta A'B'C'$ is Kiepert's parabola, given by

$$\frac{\sin A(\sin^2 B - \sin^2 C)}{u} + \frac{\sin B(\sin^2 C - \sin^2 A)}{v} + \frac{\sin C(\sin^2 A - \sin^2 B)}{w} = 0 \quad (1)$$

$$\frac{a(b^2 - c^2)}{u} + \frac{b(c^2 - a^2)}{v} + \frac{c(a^2 - b^2)}{w} = 0, \quad (2)$$

where $[u, v, w]$ are the Trilinear Coordinates for a line tangent to the parabola. It is tangent to the sides of the Triangle, the line at infinity, and the Lemoine Line. The Focus has Triangle Center Function

$$\alpha = \csc(B - C). \quad (3)$$

The Euler Line of a triangle is the Directrix of Kiepert's parabola. In fact, the Directrices of all parabolas inscribed in a Triangle pass through the Orthocenter. The Brianchon Point for Kiepert's parabola is the Steiner Point.

see also Brianchon Point, Envelope, Euler Line, Isosceles Triangle, Lemoine Line, Steiner Points

Kieroid

Let the center B of a Circle of Radius a move along a line BA. Let O be a fixed point located a distance c away from AB. Draw a Secant Line through O and D, the Midpoint of the chord cut from the line DE (which is parallel to AB) and a distance b away. Then the Locus of the points of intersection of OD and the Circle P_1 and P_2 is called a kieroid.

Special Case	Curve
$b = 0$	conchoid of Nicomedes
$b = a$	cissoid plus asymptote
$b = a = -c$	strophoid plus asymptote

References

Yates, R. C. "Kieroid." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 141–142, 1952.

Killing's Equation

The equation defining Killing Vectors.

$$\mathcal{L}_X g_{ab} = X_{a;b} + X_{b;a} = 2X_{(a;b)} = 0,$$

where $\mathcal{L}$ is the Lie Derivative.

see also Killing Vectors

Killing Vectors

If any set of points is displaced by $X^i dx_i$ where all distance relationships are unchanged (i.e., there is an Isometry), then the Vector field is called a Killing vector.

$$g_{ab} = \frac{\partial x'^c}{\partial x^a}\frac{\partial x'^d}{\partial x^b} g_{cd}(x'), \quad (1)$$

so let

$$x'^a = x^a + \epsilon x^a$$

$$\frac{\partial x'^a}{\partial x^b} = \delta^a_b + \epsilon x^a{}_{,b} \quad (2)$$

$$\begin{aligned} g_{ab}(x) &= (\delta^c_a + \epsilon x^c{}_{,a})\left(\delta^d_b + \epsilon x^d{}_{,b}\right) g_{cd}(x^e + \epsilon X^e) \\ &= (\delta^c_a + \epsilon x^c{}_{,a})\left(\delta^d_b + \epsilon x^d{}_{,b}\right) [g_{cd}(x) + \epsilon X^e g_{cd}(x)_{,e} + \ldots] \\ &= g_{ab}(x) + \epsilon[g_{ad}X^d{}_{,b} + g_{bd}X^d{}_{,a} + X^e g_{ab,e}] + \mathcal{O}(\epsilon^2) \\ &= \mathcal{L}_X g_{ab}, \end{aligned} \quad (3)$$

where $\mathcal{L}$ is the Lie Derivative. An ordinary derivative can be replaced with a covariant derivative in a Lie Derivative, so we can take as the definition

$$g_{ab;c} = 0 \quad (4)$$

$$g_{ab}g^{bc} = \delta^c_a, \quad (5)$$

which gives Killing's Equation

$$\mathcal{L}_X g_{ab} = X_{a;b} + X_{b;a} = 2X_{(a;b)} = 0. \quad (6)$$

A Killing vector X^b satisfies

$$g^{bc}X_{c;ab} - R_{ab}X^b = 0 \tag{7}$$

$$X_{a;bc} = R_{abcd}X^d \tag{8}$$

$$X^{a;b}{}_{;b} + R^a_c X^c = 0, \tag{9}$$

where R_{ab} is the RICCI TENSOR and R_{abcd} is the RIEMANN TENSOR.

A 2-sphere with METRIC

$$ds^2 = d\theta^2 + \sin^2\theta \, d\phi^2 \tag{10}$$

has three Killing vectors, given by the angular momentum operators

$$\tilde{L}_x = -\cos\phi \frac{\partial}{\partial\theta} + \cot\theta \sin\phi \frac{\partial}{\partial\phi} \tag{11}$$

$$\tilde{L}_y = \sin\phi \frac{\partial}{\partial\theta} + \cot\theta \cos\phi \frac{\partial}{\partial\phi} \tag{12}$$

$$\tilde{L}_z = \frac{\partial}{\partial\phi}. \tag{13}$$

The Killing vectors in Euclidean 3-space are

$$x^1 = \frac{\partial}{\partial x} \tag{14}$$

$$x^2 = \frac{\partial}{\partial y} \tag{15}$$

$$x^3 = \frac{\partial}{\partial z} \tag{16}$$

$$x^4 = y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y} \tag{17}$$

$$x^5 = z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z} \tag{18}$$

$$x^6 = x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}. \tag{19}$$

In MINKOWSKI SPACE, there are 10 Killing vectors

$$X_i^\mu = a_i{}^\mu \quad \text{for } i = 1, 2, 3, 4 \tag{20}$$

$$X_k^0 = 0 \tag{21}$$

$$X_k^l = \epsilon^{lkm} x_m \quad \text{for } k = 1, 2, 3 \tag{22}$$

$$X_\mu^k = \delta_\mu{}^{[0_x k]} \quad \text{for } k = 1, 2, 3. \tag{23}$$

The first group is TRANSLATION, the second ROTATION, and the final corresponds to a "boost."

Kimberling Sequence

A sequence generated by beginning with the POSITIVE integers, then iteratively applying the following algorithm:

1. In iteration i, discard the ith element,
2. Alternately write the $i+k$ and $i-k$th elements until $k = i$,
3. Write the remaining elements in order.

The first few iterations are therefore

$\boxed{1}$	2	3	4	5	6	7	8	9	10	11
2	$\boxed{3}$	4	5	6	7	8	9	10	11	12
4	2	$\boxed{5}$	6	7	8	9	10	11	12	13.
6	2	7	$\boxed{4}$	8	9	10	11	12	13	14
8	7	9	2	$\boxed{10}$	6	11	12	13	14	15

The diagonal elements form the sequence 1, 3, 5, 4, 10, 7, 15, ... (Sloane's A007063).

References

Guy, R. K. "The Kimberling Shuffle." §E35 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 235–236, 1994.

Kimberling, C. "Problem 1615." *Crux Math.* **17**, 44, 1991.

Sloane, N. J. A. Sequence A007063/M2387 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Kimberling Shuffle

see also KIMBERLING SEQUENCE

Kings Problem

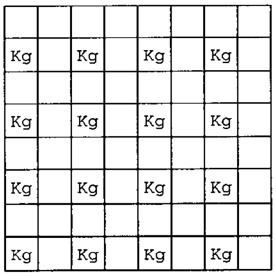

The problem of determining how many nonattacking kings can be placed on an $n \times n$ CHESSBOARD. For $n = 8$, the solution is 16, as illustrated above (Madachy 1979). In general, the solutions are

$$K(n) = \begin{cases} \frac{1}{4}n^2 & n \text{ even} \\ \frac{1}{4}(n+1)^2 & n \text{ odd} \end{cases} \tag{1}$$

(Madachy 1979), giving the sequence of doubled squares 1, 1, 4, 4, 9, 9, 16, 16, ... (Sloane's A008794). This sequence has GENERATING FUNCTION

$$\frac{1+x^2}{(1-x^2)^2(1-x)} = 1 + x + 4x^2 + 4x^3 + 9x^4 + 9x^5 + \ldots. \tag{2}$$

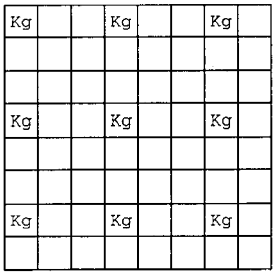

The minimum number of kings needed to attack or occupy all squares on an 8×8 CHESSBOARD is nine, illustrated above (Madachy 1979).

see also BISHOPS PROBLEM, CHESS, HARD HEXAGON ENTROPY CONSTANT, KNIGHTS PROBLEM, QUEENS PROBLEM, ROOKS PROBLEM

References
Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, p. 39, 1979.

King Walk

see DELANNOY NUMBER

Kinney's Set

A set of plane MEASURE 0 that contains a CIRCLE of every RADIUS.

References
Falconer, K. J. *The Geometry of Fractal Sets.* New York: Cambridge University Press, 1985.
Fejzić, H. "On Thin Sets of Circles." *Amer. Math. Monthly* **103**, 582–585, 1996.
Kinney, J. R. "A Thin Set of Circles." *Amer. Math. Monthly* **75**, 1077–1081, 1968.

Kinoshita-Terasaka Knot

The KNOT with BRAID WORD

$$\sigma_1{}^3\sigma_3{}^2\sigma_2\sigma_3{}^{-1}\sigma_1{}^{-2}\sigma_2\sigma_1{}^{-1}\sigma_3{}^{-1}\sigma_2{}^{-1}.$$

Its JONES POLYNOMIAL is

$$t^{-4}(-1+2t-2t^2+2t^3+t^6-2t^7+2t^8-2t^9+t^{10}),$$

the *same* as for CONWAY'S KNOT. It has the same ALEXANDER POLYNOMIAL as the UNKNOT.

References
Kinoshita, S. and Terasaka, H. "On Unions of Knots." *Osaka Math. J.* **9**, 131–153, 1959.

Kinoshita-Terasaka Mutants

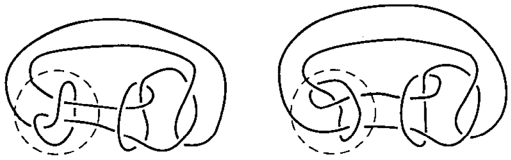

References
Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 49–50, 1994.

Kirby Calculus

The manipulation of DEHN SURGERY descriptions by a certain set of operations.

References
Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, p. 263, 1994.

Kirby's List

A list of problems in low-dimensional TOPOLOGY maintained by R. C. Kirby. The list currently runs about 380 pages.

References
Kirby, R. "Problems in Low-Dimensional Topology." `http://math.berkeley.edu/~kirby/`.

Kirkman's Schoolgirl Problem

In a boarding school there are fifteen schoolgirls who always take their daily walks in rows of threes. How can it be arranged so that each schoolgirl walks in the same row with every other schoolgirl exactly once a week? Solution of this problem is equivalent to constructing a KIRKMAN TRIPLE SYSTEM of order $n = 2$. The following table gives one of the 7 distinct (up to permutations of letters) solutions to the problem.

Sun	Mon	Tue	Wed	Thu	Fri	Sat
ABC	ADE	AFG	AHI	AJK	ALM	ANO
DHL	BIK	BHJ	BEG	CDF	BEF	BDG
EJN	CMO	CLN	BMN	CLO	CIJ	CHK
FIO	FHN	DIM	DJO	EHM	DKN	EIL
GKM	GJL	EKO	FKL	GIN	GHO	FJM

(The table of Dörrie 1965 contains a misprint in which the $a_1 = B$ and $a_2 = C$ entries for Wednesday and Thursday are written simply as a.)

see also JOSEPHUS PROBLEM, KIRKMAN TRIPLE SYSTEM, STEINER TRIPLE SYSTEM

References
Abel, R. J. R. and Furino, S. C. "Kirkman Triple Systems." §I.6.3 in *The CRC Handbook of Combinatorial Designs* (Ed. C. J. Colbourn and J. H. Dinitz). Boca Raton, FL: CRC Press, pp. 88–89, 1996.
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 287–289, 1987.
Dörrie, H. §5 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 14–18, 1965.
Frost, A. "General Solution and Extension of the Problem of the 15 Schoolgirls." *Quart. J. Pure Applied Math.* **11**, 1871.
Kirkman, T. P. "On a Problem in COmbinatorics." *Cambridge and Dublin Math. J.* **2**, 191–204, 1847.
Kirkman, T. P. *Lady's and Gentleman's Diary.* 1850.
Kraitchik, M. §9.3.1 in *Mathematical Recreations.* New York: W. W. Norton, pp. 226–227, 1942.
Peirce, B. "Cyclic Solutions of the School-Girl Puzzle." *Astron. J.* **6**, 169–174, 1859–1861.
Ryser, H. J. *Combinatorial Mathematics.* Buffalo, NY: Math. Assoc. Amer., pp. 101–102, 1963.

Kirkman Triple System

A Kirkman triple system of order $v = 6n + 3$ is a STEINER TRIPLE SYSTEM with parallelism (Ball and Coxeter 1987), i.e., one with the following additional stipulation: the set of $b = (2n+1)(3n+1)$ triples is partitioned into $3n+1$ components such that each component is a $(2n+1)$-subset of triples and each of the v elements appears exactly once in each component. The

STEINER TRIPLE SYSTEMS of order 3 and 9 are Kirkman triple systems with $n = 0$ and 1. Solution to KIRKMAN'S SCHOOLGIRL PROBLEM requires construction of a Kirkman triple system of order $n = 2$.

Ray-Chaudhuri and Wilson (1971) showed that there exists at least one Kirkman triple system for every NONNEGATIVE order n. Earlier editions of Ball and Coxeter (1987) gave constructions of Kirkman triple systems with $9 \leq v \leq 99$. For $n = 1$, there is a single unique (up to an isomorphism) solution, while there are 7 different systems for $n = 2$ (Mulder 1917, Cole 1922, Ball and Coxeter 1987).

see also STEINER TRIPLE SYSTEM

References

Abel, R. J. R. and Furino, S. C. "Kirkman Triple Systems." §I.6.3 in *The CRC Handbook of Combinatorial Designs* (Ed. C. J. Colbourn and J. H. Dinitz). Boca Raton, FL: CRC Press, pp. 88–89, 1996.

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 287–289, 1987.

Cole, F. N. *Bull. Amer. Math. Soc.* **28**, 435–437, 1922.

Kirkman, T. P. *Cambridge and Dublin Math. J.* **2**, 191–204, 1947.

Lindner, C. C. and Rodger, C. A. *Design Theory.* Boca Raton, FL: CRC Press, 1997.

Mulder, P. *Kirkman-Systemen.* Groningen Dissertation. Leiden, Netherlands, 1917.

Ray-Chaudhuri, D. K. and Wilson, R. M. "Solution of Kirkman's Schoolgirl Problem." *Combinatorics, Proc. Sympos. Pure Math., Univ. California, Los Angeles, Calif., 1968* **19**, 187–203, 1971.

Ryser, H. J. *Combinatorial Mathematics.* Buffalo, NY: Math. Assoc. Amer., pp. 101–102, 1963.

Kiss Surface

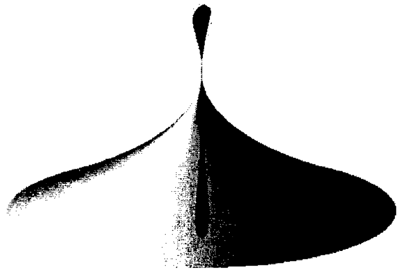

The QUINTIC SURFACE given by the equation

$$\tfrac{1}{2}x^5 + \tfrac{1}{2}x^4 - (y^2 + z^2) = 0.$$

References

Nordstrand, T. "Surfaces." `http://www.uib.no/people/nfytn/surfaces.htm`.

Kissing Circles Problem

see DESCARTES CIRCLE THEOREM, SODDY CIRCLES

Kissing Number

The number of equivalent HYPERSPHERES in n-D which can touch an equivalent HYPERSPHERE without any intersections, also sometimes called the NEWTON NUMBER, CONTACT NUMBER, COORDINATION NUMBER, or LIGANCY. Newton correctly believed that the kissing number in 3-D was 12, but the first proofs were not produced until the 19th century (Conway and Sloane 1993, p. 21) by Bender (1874), Hoppe (1874), and Günther (1875). More concise proofs were published by Schütte and van der Waerden (1953) and Leech (1956). Exact values for *lattice packings* are known for $n = 1$ to 9 and $n = 24$ (Conway and Sloane 1992, Sloane and Nebe). Odlyzko and Sloane (1979) found the exact value for 24-D.

The following table gives the largest known kissing numbers in DIMENSION D for lattice (L) and nonlattice (NL) packings (if a nonlattice packing with higher number exists). In nonlattice packings, the kissing number may vary from sphere to sphere, so the largest value is given below (Conway and Sloane 1993, p. 15). An more extensive and up-to-date tabulation is maintained by Sloane and Nebe.

D	L	NL	D	L	NL
1	2		13	≥ 918	$\geq 1,130$
2	6		14	$\geq 1,422$	$\geq 1,582$
3	12		15	$\geq 2,340$	
4	24		16	$\geq 4,320$	
5	40		17	$\geq 5,346$	
6	72		18	$\geq 7,398$	
7	126		19	$\geq 10,668$	
8	240		20	$\geq 17,400$	
9	272	≥ 306	21	$\geq 27,720$	
10	≥ 336	≥ 500	22	$\geq 49,896$	
11	≥ 438	≥ 582	23	$\geq 93,150$	
12	≥ 756	≥ 840	24	196,560	

The lattices having maximal packing numbers in 12- and 24-D have special names: the COXETER-TODD LATTICE and LEECH LATTICE, respectively. The general form of the lower bound of n-D lattice densities given by

$$\eta \geq \frac{\zeta(n)}{2^{n-1}},$$

where $\zeta(n)$ is the RIEMANN ZETA FUNCTION, is known as the MINKOWSKI-HLAWKA THEOREM.

see also COXETER-TODD LATTICE, HERMITE CONSTANTS, HYPERSPHERE PACKING, LEECH LATTICE, MINKOWSKI-HLAWKA THEOREM

References

Bender, C. "Bestimmung der grössten Anzahl gleich Kugeln, welche sich auf eine Kugel von demselben Radius, wie die übrigen, auflegen lassen." *Archiv Math. Physik (Grunert)* **56**, 302–306, 1874.

Conway, J. H. and Sloane, N. J. A. "The Kissing Number Problem" and "Bounds on Kissing Numbers." §1.2 and Ch. 13 in *Sphere Packings, Lattices, and Groups, 2nd ed.* New York: Springer-Verlag, pp. 21–24 and 337–339, 1993.

Edel, Y.; Rains, E. M.; Sloane, N. J. A. "On Kissing Numbers in Dimensions 32 to 128." *Electronic J. Combinatorics* **5**, No. 1, R22, 1–5, 1998. `http://www.combinatorics.org/Volume_5/v5i1toc.html`.
Günther, S. "Ein stereometrisches Problem." *Archiv Math. Physik* **57**, 209–215, 1875.
Hoppe, R. "Bemerkung der Redaction." *Archiv Math. Physik. (Grunert)* **56**, 307–312, 1874.
Kuperberg, G. "Average Kissing Numbers for Sphere Packings." Preprint.
Kuperberg, G. and Schramm, O. "Average Kissing Numbers for Non-Congruent Sphere Packings." *Math. Res. Let.* **1**, 339–344, 1994.
Leech, J. "The Problem of Thirteen Spheres." *Math. Gaz.* **40**, 22–23, 1956.
Odlyzko, A. M. and Sloane, N. J. A. "New Bounds on the Number of Unit Spheres that Can Touch a Unit Sphere in *n* Dimensions." *J. Combin. Th. A* **26**, 210–214, 1979.
Schütte, K. and van der Waerden, B. L. "Das Problem der dreizehn Kugeln." *Math. Ann.* **125**, 325–334, 1953.
Sloane, N. J. A. Sequence A001116/M1585 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Sloane, N. J. A. and Nebe, G. "Table of Highest Kissing Numbers Presently Known." `http://www.research.att.com/~njas/lattices/kiss.html`.
Stewart, I. *The Problems of Mathematics, 2nd ed.* Oxford, England: Oxford University Press, pp. 82–84, 1987.

Kite

see DIAMOND, LOZENGE, PARALLELOGRAM, PENROSE TILES, QUADRILATERAL, RHOMBUS

Klarner-Rado Sequence

The thinnest sequence which contains 1, and whenever it contains x, also contains $2x$, $3x+2$, and $6x+3$: 1, 2, 4, 5, 8, 9, 10, 14, 15, 16, 17, ... (Sloane's A005658).

References

Guy, R. K. "Klarner-Rado Sequences." §E36 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 237, 1994.
Klarner, D. A. and Rado, R. "Linear Combinations of Sets of Consecutive Integers." *Amer. Math. Monthly* **80**, 985–989, 1973.
Sloane, N. J. A. Sequence A005658/M0969 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Klarner's Theorem

An $a \times b$ RECTANGLE can be packed with $1 \times n$ strips IFF $n|a$ or $n|b$.

see also BOX-PACKING THEOREM, CONWAY PUZZLE, DE BRUIJN'S THEOREM, SLOTHOUBER-GRAATSMA PUZZLE

References

Honsberger, R. *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., p. 88, 1976.

Klein's Absolute Invariant

$$J(q) \equiv \frac{4}{27}\frac{[1-\lambda(q)+\lambda^2(q)]^3}{\lambda^2(q)[1-\lambda(q)]^2} = \frac{[E_4(q)]^3}{[E_4(q)]^3-[E_6(q)^2]}$$

(Cohn 1994), where $q \equiv e^{i\pi t}$ is the NOME, $\lambda(q)$ is the ELLIPTIC LAMBDA FUNCTION

$$\lambda(q) \equiv k^2(q) = \left[\frac{\vartheta_2(q)}{\vartheta_3(q)}\right]^4,$$

$\vartheta_i(q)$ is a THETA FUNCTION, and the $E_i(q)$ are RAMANUJAN-EISENSTEIN SERIES. $J(t)$ is GAMMA-MODULAR.

see also ELLIPTIC LAMBDA FUNCTION, j-FUNCTION, PI, RAMANUJAN-EISENSTEIN SERIES, THETA FUNCTION

References

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, pp. 115 and 179, 1987.
Cohn, H. *Introduction to the Construction of Class Fields.* New York: Dover, p. 73, 1994.
Weisstein, E. W. "j-Function." `http://www.astro.virginia.edu/~eww6n/math/notebooks/jFunction.m`.

Klein-Beltrami Model

The Klein-Beltrami model of HYPERBOLIC GEOMETRY consists of an OPEN DISK in the Euclidean plane whose open chords correspond to hyperbolic lines. Two lines l and m are then considered parallel if their chords fail to intersect and are PERPENDICULAR under the following conditions,

1. If at least one of l and m is a diameter of the DISK, they are hyperbolically perpendicular IFF they are perpendicular in the Euclidean sense.
2. If neither is a diameter, l is perpendicular to m IFF the Euclidean line extending l passes through the pole of m (defined as the point of intersection of the tangents to the disk at the "endpoints" of m).

There is an isomorphism between the POINCARÉ HYPERBOLIC DISK model and the Klein-Beltrami model. Consider a Klein disk in Euclidean 3-space with a SPHERE of the same radius seated atop it, tangent at the ORIGIN. If we now project chords on the disk orthogonally upward onto the SPHERE's lower HEMISPHERE, they become arcs of CIRCLES orthogonal to the equator. If we then stereographically project the SPHERE's lower HEMISPHERE back onto the plane of the Klein disk from the north pole, the equator will map onto a disk somewhat larger than the Klein disk, and the chords of the original Klein disk will now be arcs of CIRCLES orthogonal to this larger disk. That is, they will be Poincaré lines. Now we can say that two Klein lines or angles are congruent iff their corresponding Poincaré lines and angles under this isomorphism are congruent in the sense of the Poincaré model.

see also HYPERBOLIC GEOMETRY, POINCARÉ HYPERBOLIC DISK

Klein Bottle

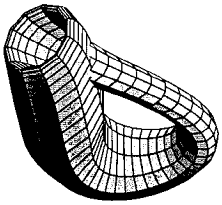

A closed NONORIENTABLE SURFACE of GENUS one having no inside or outside. It can be physically realized only in 4-D (since it must pass through itself without the presence of a HOLE). Its TOPOLOGY is equivalent to a pair of CROSS-CAPS with coinciding boundaries. It can be cut in half along its length to make two MÖBIUS STRIPS.

The above picture is an IMMERSION of the Klein bottle in $\mathbb{R}^3$ (3-space). There is also another possible IMMERSION called the "figure-8" IMMERSION (Geometry Center).

The equation for the usual IMMERSION is given by the implicit equation

$$(x^2+y^2+z^2+2y-1)[(x^2+y^2+z^2-2y-1)^2-8z^2] \\ +16xz(x^2+y^2+z^2-2y-1)=0 \quad (1)$$

(Stewart 1991). Nordstrand gives the parametric form

$$x=\cos u[\cos(\tfrac{1}{2}u)(\sqrt{2}+\cos v)+\sin(\tfrac{1}{2}u)\sin v\cos v] \quad (2)$$

$$y=\sin u[\cos(\tfrac{1}{2}u)(\sqrt{2}+\cos v)+\sin(\tfrac{1}{2}u)\sin v\cos v] \quad (3)$$

$$z=-\sin(\tfrac{1}{2}u)(\sqrt{2}+\cos v)+\cos(\tfrac{1}{2}u)\sin v\cos v. \quad (4)$$

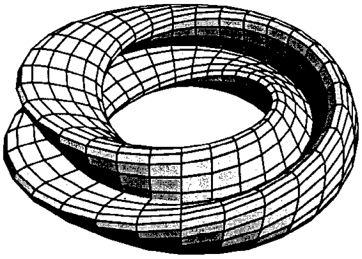

The "figure-8" form of the Klein bottle is obtained by rotating a figure eight about an axis while placing a twist in it, and is given by parametric equations

$$x(u,v)=[a+\cos(\tfrac{1}{2}u)\sin(v)-\sin(\tfrac{1}{2}u)\sin(2v)]\cos(u) \quad (5)$$

$$y(u,v)=[a+\cos(\tfrac{1}{2}u)\sin(v)-\sin(\tfrac{1}{2}u)\sin(2v)]\sin(u) \quad (6)$$

$$z(u,v)=\sin(\tfrac{1}{2}u)\sin(v)+\cos(\tfrac{1}{2}u)\sin(2v) \quad (7)$$

for $u \in [0,2\pi)$, $v \in [0,2\pi)$, and $a>2$ (Gray 1993).

The image of the CROSS-CAP map of a TORUS centered at the ORIGIN is a Klein bottle (Gray 1993, p. 249).

Any set of regions on the Klein bottle can be colored using ss colors only (Franklin 1934, Saaty 1986).

see also CROSS-CAP, ETRUSCAN VENUS SURFACE, IDA SURFACE, MAP COLORING MÖBIUS STRIP

References

Dickson, S. "Klein Bottle Graphic." `http://www.mathsource.com/cgi-bin/MathSource/Applications/Graphics/3D/0201-801`.

Franklin, P. "A Six Colour Problem." *J. Math. Phys.* **13**, 363–369, 1934.

Geometry Center. "The Klein Bottle." `http://www.geom.umn.edu/zoo/toptype/klein/`.

Geometry Center. "The Klein Bottle in Four-Space." `http://www.geom.umn.edu/~banchoff/Klein4D/Klein4D.html`.

Gray, A. "The Klein Bottle." §12.4 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 239–240, 1993.

Nordstrand, T. "The Famed Klein Bottle." `http://www.uib.no/people/nfytn/kleintxt.htm`.

Pappas, T. "The Moebius Strip & the Klein Bottle." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 44–46, 1989.

Saaty, T. L. and Kainen, P. C. *The Four-Color Problem: Assaults and Conquest.* New York: Dover, p. 45, 1986.

Stewart, I. *Game, Set and Math.* New York: Viking Penguin, 1991.

Wang, P. "Renderings." `http://www.ugcs.caltech.edu/~peterw/portfolio/renderings/`.

Klein's Equation

If a REAL curve has no singularities except nodes and CUSPS, BITANGENTS, and INFLECTION POINTS, then

$$n+2\tau_2'+\iota'=m+2\delta_2'+\kappa',$$

where n is the order, τ' is the number of conjugate tangents, ι' is the number of REAL inflections, m is the class, δ' is the number of REAL conjugate points, and κ' is the number of REAL CUSPS. This is also called KLEIN'S THEOREM.

see also PLÜCKER'S EQUATION

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 114, 1959.

Klein Four-Group

see VIERGRUPPE

Klein-Gordon Equation

$$\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2}=\frac{\partial^2\psi}{\partial x^2}-\mu^2\psi.$$

see also SINE-GORDON EQUATION, WAVE EQUATION

Klein Quartic

The 3-holed TORUS.

Klein's Theorem

see KLEIN'S EQUATION

Kleinian Group

A finitely generated discontinuous group of linear fractional transformation acting on a domain in the COMPLEX PLANE.

References
Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 425, 1980.
Kra, I. *Automorphic Forms and Kleinian Groups.* Reading, MA: W. A. Benjamin, 1972.

Kloosterman's Sum

$$S(u,v,n) \equiv \sum_n \exp\left[\frac{2\pi i(uh+v\bar{h})}{n}\right], \qquad (1)$$

where h runs through a complete set of residues RELATIVELY PRIME to n, and $\bar{h}$ is defined by

$$h\bar{h} \equiv 1 \pmod{n}. \qquad (2)$$

If $(n,n')=1$ (if n and n' are RELATIVELY PRIME), then

$$S(u,v,n)S(u,v',n') = S(u, vn'^2 + v'n^2, nn'). \qquad (3)$$

Kloosterman's sum essentially solves the problem introduced by Ramanujan of representing sufficiently large numbers by QUADRATIC FORMS $ax_1{}^2 + bx_2{}^2 + cx_3{}^2 + dx_4{}^2$. Weil improved on Kloosterman's estimate for Ramanujan's problem with the best possible estimate

$$|S(u,u,n)| \leq 2\sqrt{n} \qquad (4)$$

(Duke 1997).

see also GAUSSIAN SUM

References
Duke, W. "Some Old Problems and New Results about Quadratic Forms." *Not. Amer. Math. Soc.* **44**, 190–196, 1997.
Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, p. 56, 1979.
Katz, N. M. *Gauss Sums, Kloosterman Sums, and Monodromy Groups.* Princeton, NJ: Princeton University Press, 1987.
Kloosterman, H. D. "On the Representation of Numbers in the Form $ax^2 + by^2 + cz^2 + dt^2$." *Acta Math.* **49**, 407–464, 1926.
Ramanujan, S. "On the Expression of a Number in the Form $ax^2 + by^2 + cz^2 + du^2$." *Collected Papers.* New York: Chelsea, 1962.

Knapsack Problem

Given a SUM and a set of WEIGHTS, find the WEIGHTS which were used to generate the SUM. The values of the weights are then encrypted in the sum. The system relies on the existence of a class of knapsack problems which can be solved trivially (those in which the weights are separated such that they can be "peeled off" one at a time using a GREEDY-like algorithm), and transformations which convert the trivial problem to a difficult one and vice versa. Modular multiplication is used as the TRAPDOOR FUNCTION. The simple knapsack system was broken by Shamir in 1982, the Graham-Shamir system by Adleman, and the iterated knapsack by Ernie Brickell in 1984.

References
Coppersmith, D. "Knapsack Used in Factoring." §4.6 in *Open Problems in Communication and Computation* (Ed. T. M. Cover and B. Gopinath). New York: Springer-Verlag, pp. 117–119, 1987.
Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 163–166, 1985.

Kneser-Sommerfeld Formula

Let J_ν be a BESSEL FUNCTION OF THE FIRST KIND, N_ν a NEUMANN FUNCTION, and $j_{\nu,n}$the zeros of $z^{-\nu}J_\nu(z)$ in order of ascending REAL PART. Then for $0 < x < X < 1$ and $\Re[z] > 0$,

$$\frac{\pi J_\nu(xz)}{4J_\nu(z)}[J_\nu(z)N_\nu(Xz) - N_\nu(z)J_\nu(Xz)] = \sum_{n=1}^{\infty} \frac{J_\nu(j_{\nu,n}x)J_\nu(j_{\nu,n}X)}{(z^2 - j_{\nu,n}{}^2)J'_{\nu,n}{}^2(j_{\nu,n})}.$$

References
Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1474, 1980.

Knights Problem

Kt		Kt		Kt		Kt	
	Kt		Kt		Kt		Kt
Kt		Kt		Kt		Kt	
	Kt		Kt		Kt		Kt
Kt		Kt		Kt		Kt	
	Kt		Kt		Kt		Kt
Kt		Kt		Kt		Kt	
	Kt		Kt		Kt		Kt

The problem of determining how many nonattacking knights $K(n)$ can be placed on an $n \times n$ CHESSBOARD. For $n = 8$, the solution is 32 (illustrated above). In general, the solutions are

$$K(n) = \begin{cases} \frac{1}{2}n^2 & n > 2 \text{ even} \\ \frac{1}{2}(n^2+1) & n > 1 \text{ odd,} \end{cases}$$

giving the sequence 1, 4, 5, 8, 13, 18, 25, ... (Sloane's A030978, Dudeney 1970, p. 96; Madachy 1979).

					Kt		
	Kt	Kt		Kt	Kt		
		Kt					
					Kt		
		Kt	Kt		Kt	Kt	
		Kt					

The minimal number of knights needed to occupy or attack every square on an $n \times n$ CHESSBOARD is given by 1, 4, 4, 4, 5, 8, 10, ... (Sloane's A006075). The number of such solutions are given by 1, 1, 2, 3, 8, 22, 3, ... (Sloane's A006076).

see also BISHOPS PROBLEM, CHESS, KINGS PROBLEM, KNIGHT'S TOUR, QUEENS PROBLEM, ROOKS PROBLEM

References

Dudeney, H. E. "The Knight-Guards." §319 in *Amusements in Mathematics.* New York: Dover, p. 95, 1970.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 38–39, 1979.

Moser, L. "King Paths on a Chessboard." *Math. Gaz.* **39**, 54, 1955.

Sloane, N. J. A. Sequences A030978, A006076/M0884, and A006075/M3224 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sloane, N. J. A. and Plouffe, S. Extended entry for M3224 in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.

Vardi, I. *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 196–197, 1991.

Wilf, H. S. "The Problem of Kings." *Electronic J. Combinatorics* **2**, 3, 1–7, 1995. `http://www.combinatorics.org/Volume_2/volume2.html#3`.

Knights of the Round Table

see NECKLACE

Knight's Tour

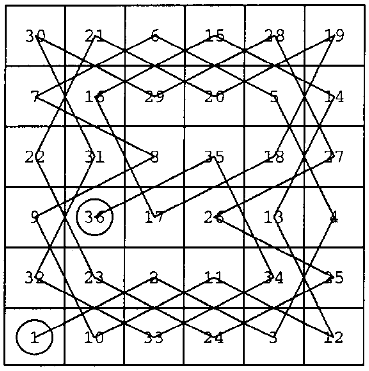

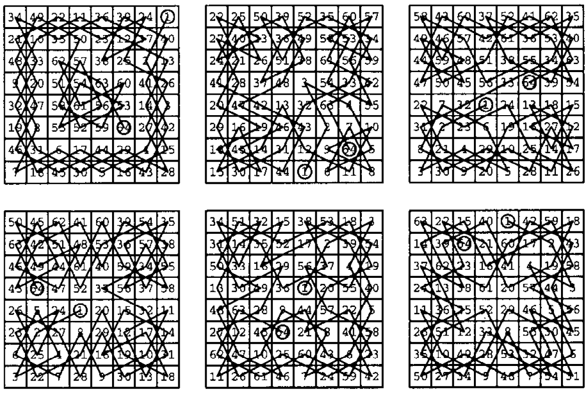

A knight's tour of a CHESSBOARD (or any other grid) is a sequence of moves by a knight CHESS piece (which may only make moves which simultaneously shift one square along one axis and two along the other) such that each square of the board is visited exactly once (i.e., a HAMILTONIAN CIRCUIT). If the final position is a knight's move away from the first position, the tour is called re-entrant. The first figure above shows a knight's tour on a 6×6 CHESSBOARD. The second set of figures shows six knight's tours on an 8×8 CHESSBOARD, all but the first of which are re-entrant. The final tour has the additional property that it is a SEMIMAGIC SQUARE with row and column sums of 260 and main diagonal sums of 348 and 168.

Löbbing and Wegener (1996) computed the number of cycles covering the directed knight's graph for an 8×8 CHESSBOARD. They obtained α^2, where $\alpha = 2,849,759,680$, i.e., 8,121,130,233,753,702,400. They also computed the number of undirected tours, obtaining an incorrect answer 33,439,123,484,294 (which is not divisible by 4 as it must be), and so are currently redoing the calculation.

The following results are given by Kraitchik (1942). The number of possible tours on a $4k \times 4k$ board for $k = 3$, 4, ... are 8, 0, 82, 744, 6378, 31088, 189688, 1213112, ... (Kraitchik 1942, p. 263). There are 14 tours on the 3×7 rectangle, two of which are symmetrical. There are 376 tours on the 3×8 rectangle, none of which is closed. There are 16 symmetric tours on the 3×9 rectangle and 8 closed tours on the 3×10 rectangle. There are 58 symmetric tours on the 3×11 rectangle and 28 closed tours on the 3×12 rectangle. There are five doubly symmetric tours on the 6×6 square. There are 1728 tours on the 5×5 square, 8 of which are symmetric. The longest "uncrossed" knight's tours on an $n \times n$ board for $n = 3$, 4, ... are 2, 5, 10, 17, 24, 35, ... (Sloane's A003192).

see also CHESS, KINGS PROBLEM, KNIGHTS PROBLEM, MAGIC TOUR, QUEENS PROBLEM, TOUR

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 175–186, 1987.

Chartrand, G. "The Knight's Tour." §6.2 in *Introductory Graph Theory.* New York: Dover, pp. 133–135, 1985.

Gardner, M. "Knights of the Square Table." Ch. 14 in *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, pp. 188–202, 1978.

Guy, R. K. "The n Queens Problem." §C18 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 133–135, 1994.

Kraitchik, M. "The Problem of the Knights." Ch. 11 in *Mathematical Recreations.* New York: W. W. Norton, pp. 257–266, 1942.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 87–89, 1979.

Ruskey, F. "Information on the n Knight's Tour Problem." `http://sue.csc.uvic.ca/~cos/inf/misc/Knight.html`.

Sloane, N. J. A. Sequences A003192/M1369 and A006075/M3224 in "An On-Line Version of the Encyclopedia of Integer Sequences."

van der Linde, A. *Geschichte und Literatur des Schachspiels, Vol. 2.* Berlin, pp. 101–111, 1874.

Volpicelli, P. "Soluzione completa e generale, mediante la geometria di situazione, del problema relativo alle corse del cavallo sopra qualunque scacchiere." *Atti della Reale Accad. dei Lincei* **25**, 87–162, 1872.

Wegener, I. and Löbbing, M. "The Number of Knight's Tours Equals 33,439,123,484,294—Counting with Binary Decision Diagrams." *Electronic J. Combinatorics* **3**, R5, 1–4, 1996. `http://www.combinatorics.org/Volume_3/volume3.html#R5`.

Knödel Numbers

For every $k \geq 1$, let C_k be the set of COMPOSITE numbers $n > k$ such that if $1 < a < n$, $\text{GCD}(a, n) = 1$ (where GCD is the GREATEST COMMON DIVISOR), then $a^{n-k} \equiv 1 \pmod{n}$. C_1 is the set of CARMICHAEL NUMBERS. Makowski (1962/1963) proved that there are infinitely many members of C_k for $k \geq 2$.

see also CARMICHAEL NUMBER, D-NUMBER, GREATEST COMMON DIVISOR

References

Makowski, A. "Generalization of Morrow's D-Numbers." *Simon Stevin* **36**, 71, 1962/1963.

Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, p. 101, 1989.

Knot

A knot is defined as a closed, non-self-intersecting curve embedded in 3-D. A knot is a single component LINK. Klein proved that knots cannot exist in an EVEN-numbered dimensional space ≥ 4. It has since been shown that a knot cannot exist in *any* dimension ≥ 4. Two distinct knots cannot have the same KNOT COMPLEMENT (Gordon and Luecke 1989), but two LINKS can! (Adams 1994, p. 261). The KNOT SUM of any number of knots cannot be the UNKNOT unless each knot in the sum is the UNKNOT.

Knots can be cataloged based on the minimum number of crossings present. Knots are usually further broken down into PRIME KNOTS. Knot theory was given its first impetus when Lord Kelvin proposed a theory that atoms were vortex loops, with different chemical elements consisting of different knotted configurations (Thompson 1867). P. G. Tait then cataloged possible knots by trial and error.

Thistlethwaite has used DOWKER NOTATION to enumerate the number of PRIME KNOTS of up to 13 crossings, and ALTERNATING KNOTS up to 14 crossings. In this compilation, MIRROR IMAGES are counted as a single knot type. The number of distinct PRIME KNOTS $N(n)$ for knots from $n = 3$ to 13 crossings are 1, 1, 2, 3, 7, 21, 49, 165, 552, 2176, 9988 (Sloane's A002863). Combining PRIME KNOTS gives one additional type of knot each for knots six and seven crossings.

Let $C(n)$ be the number of distinct PRIME KNOTS of n crossings, *counting* CHIRAL *versions of the same knot separately.* Then

$$\tfrac{1}{3}(2^{n-2} - 1) \leq N(n) \lesssim e^n$$

(Ernst and Summers 1987). Welsh has shown that the number of knots is bounded by an exponential in n.

A pictorial enumeration of PRIME KNOTS of up to 10 crossings appears in Rolfsen (1976, Appendix C). Note, however, that in this table, the PERKO PAIR 10_{161} and 10_{162} are actually identical, and the uppermost crossing in 10_{144} should be changed (Jones 1987). The kth knot having n crossings in this (arbitrary) ordering of knots is given the symbol n_k. Another possible representation for knots uses the BRAID GROUP. A knot with $n + 1$ crossings is a member of the BRAID GROUP n. There is no general method known for deciding whether two given knots are equivalent or interlocked. There is no general ALGORITHM to determine if a tangled curve is a knot. Haken (1961) has given an ALGORITHM, but it is too complex to apply to even simple cases.

If a knot is AMPHICHIRAL, the "amphichirality" is $A = 1$, otherwise $A = 0$ (Jones 1987). ARF INVARIANTS are designated a. BRAID WORDS are denoted b (Jones 1987). CONWAY'S KNOT NOTATION C for knots up to 10 crossings is given by Rolfsen (1976). Hyperbolic volumes are given (Adams, Hildebrand, and Weeks 1991; Adams 1994). The BRAID INDEX i is given by Jones (1987). ALEXANDER POLYNOMIALS Δ are given in Rolfsen (1976), but with the POLYNOMIALS for 10_{083} and 10_{086} reversed (Jones 1987). The ALEXANDER POLYNOMIALS are normalized according to Conway, and given in abbreviated form $[a_1, a_2, \ldots$ for $a_1 + a_2(x^{-1} + x) + \ldots$.

The JONES POLYNOMIALS W for knots of up to 10 crossings are given by Jones (1987), and the JONES POLYNOMIALS V can be either computed from these, or taken from Adams (1994) for knots of up to 9 crossings (although most POLYNOMIALS are associated with the wrong knot in the first printing). The JONES POLYNOMIALS are listed in the abbreviated form $\{n\}\ a_0\ a_1 \ldots$ for $t^{-n}(a_0 + a_1 t + \ldots)$, and correspond either to the knot depicted by Rolfsen or its MIRROR IMAGE, whichever has the lower POWER of t^{-1}. The HOMFLY POLYNOMIAL $P(\ell, m)$ and KAUFFMAN POLYNOMIAL $F(a, x)$ are given in Lickorish and Millett (1988) for knots of up to 7 crossings.

M. B. Thistlethwaite has tabulated the HOMFLY POLYNOMIAL and KAUFFMAN POLYNOMIAL F for KNOTS of up to 13 crossings.

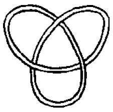
3_1

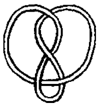
4_1

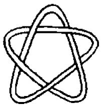
5_1

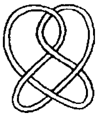
5_2

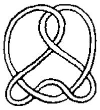
6_1

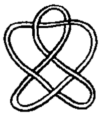
6_2

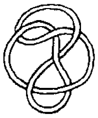
6_3

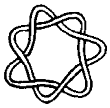
7_1

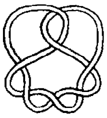
7_2

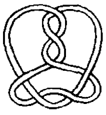
7_3

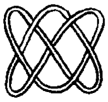
7_4

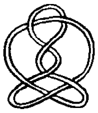
7_5

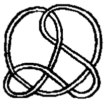
7_6

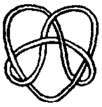
7_7

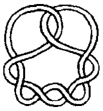
8_1

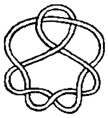
8_2

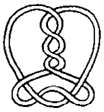
8_3

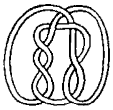
8_4

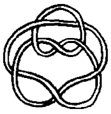
8_5

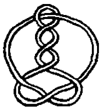
8_6

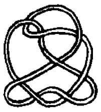
8_7

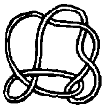
8_8

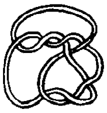
8_9

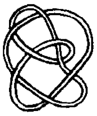
8_{10}

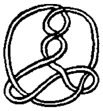
8_{11}

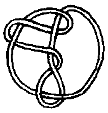
8_{12}

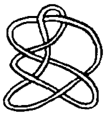
8_{13}

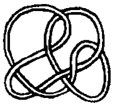
8_{14}

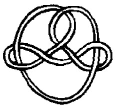
8_{15}

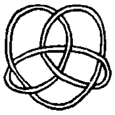
8_{16}

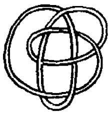
8_{17}

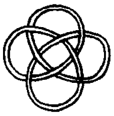
8_{18}

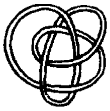
8_{19}

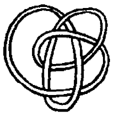
8_{20}

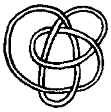
8_{21}

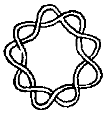
9_1

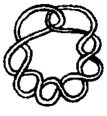
9_2

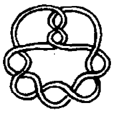
9_3

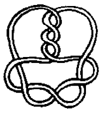
9_4

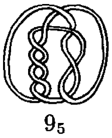
9_5

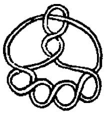
9_6

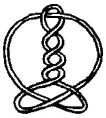
9_7

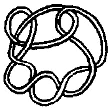
9_8

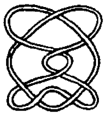
9_9

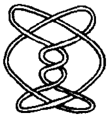
9_{10}

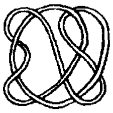
9_{11}

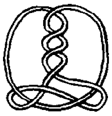
9_{12}

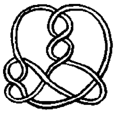
9_{13}

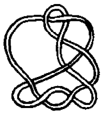
9_{14}

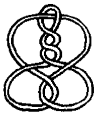
9_{15}

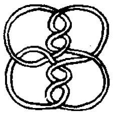
9_{16}

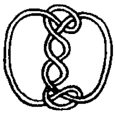
9_{17}

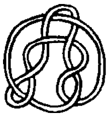
9_{18}

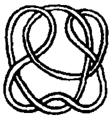
9_{19}

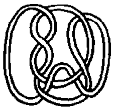
9_{20}

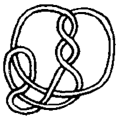
9_{21}

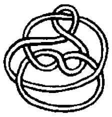
9_{22}

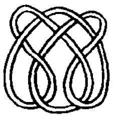
9_{23}

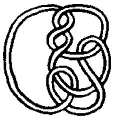
9_{24}

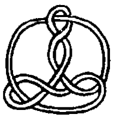
9_{25}

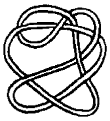
9_{26}

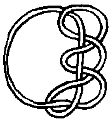
9_{27}

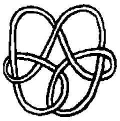
9_{28}

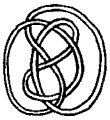
9_{29}

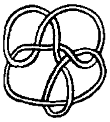
9_{30}

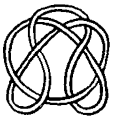
9_{31}

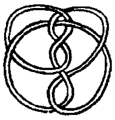
9_{32}

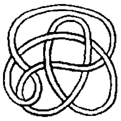
9_{33}

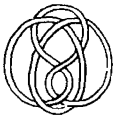
9_{34}

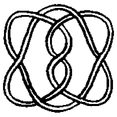
9_{35}

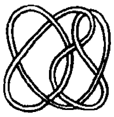
9_{36}

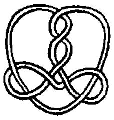
9_{37}

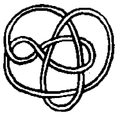
9_{38}

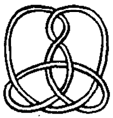
9_{39}

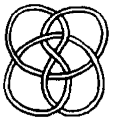
9_{40}

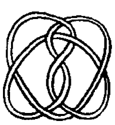
9_{41}

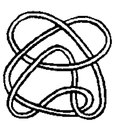
9_{42}

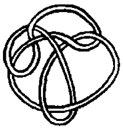
9_{43}

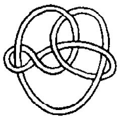
9_{44}

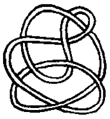
9_{45}

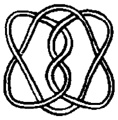
9_{46}

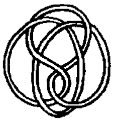
9_{47}

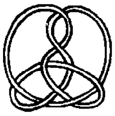
9_{48}

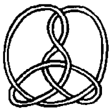
9_{49}

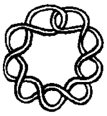
10_{1}

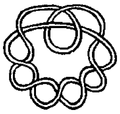
10_{2}

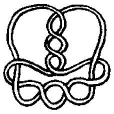
10_{3}

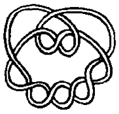
10_{4}

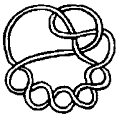
10_{5}

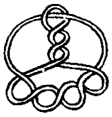
10_{6}

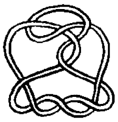
10_{7}

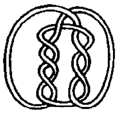
10_{8}

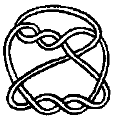
10_{9}

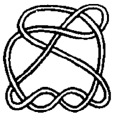
10_{10}

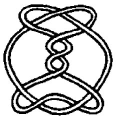
10_{11}

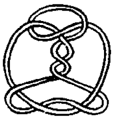
10_{12}

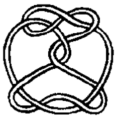
10_{13}

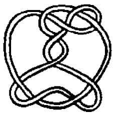
10_{14}

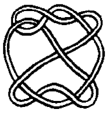
10_{15}

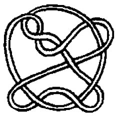
10_{16}

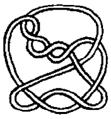
10_{17}

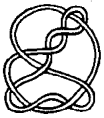
10_{18}

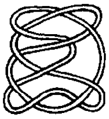
10_{19}

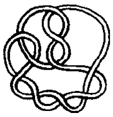
10_{20}

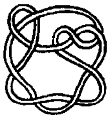
10_{21}

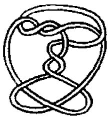
10_{22}

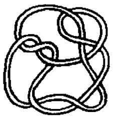
10_{23}

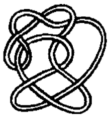
10_{24}

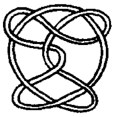
10_{25}

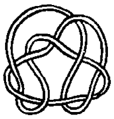
10_{26}

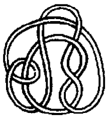
10_{27}

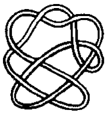
10_{28}

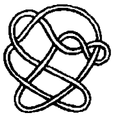
10_{29}

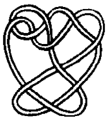
10_{30}

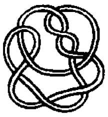
10_{31}

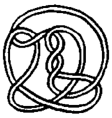
10_{32}

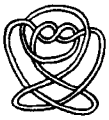
10_{33}

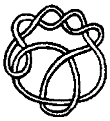
10_{34}

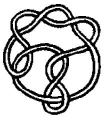
10_{35}

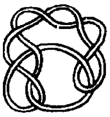
10_{36}

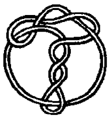
10_{37}

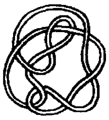
10_{38}

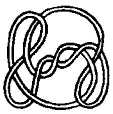
10_{39}

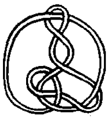
10_{40}

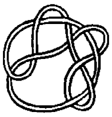
10_{41}

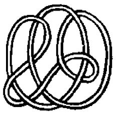
10_{42}

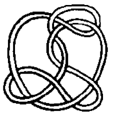
10_{43}

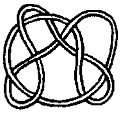
10_{44}

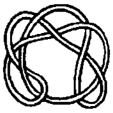
10_{45}

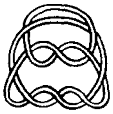
10_{46}

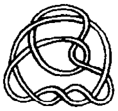
10_{47}

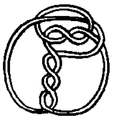
10_{48}

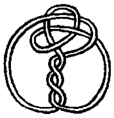
10_{49}

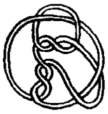
10_{50}

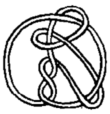
10_{51}

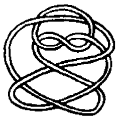
10_{52}

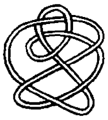
10_{53}

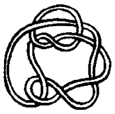
10_{54}

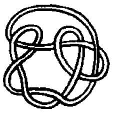
10_{55}

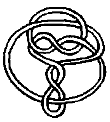
10_{56}

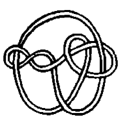
10_{57}

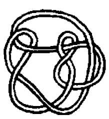
10_{58}

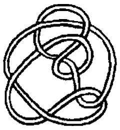
10_{59}

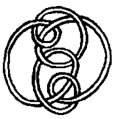
10_{60}

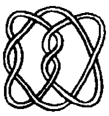
10_{61}

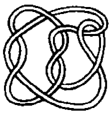
10_{62}

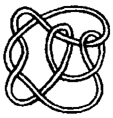
10_{63}

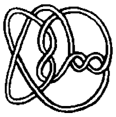
10_{64}

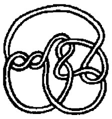
10_{65}

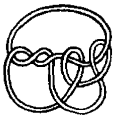
10_{66}

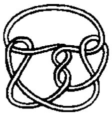
10_{67}

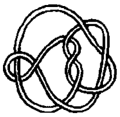
10_{68}

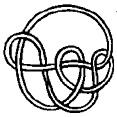
10_{69}

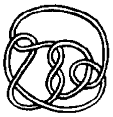
10_{70}

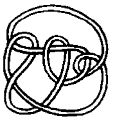
10_{71}

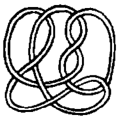
10_{72}

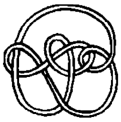
10_{73}

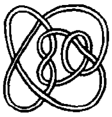
10_{74}

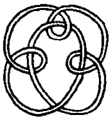
10_{75}

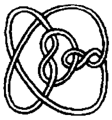
10_{76}

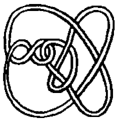
10_{77}

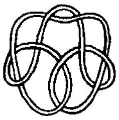
10_{78}

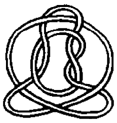
10_{79}

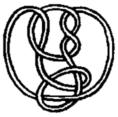
10_{80}

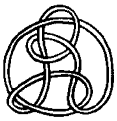
10_{81}

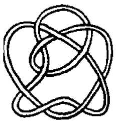
10_{82}

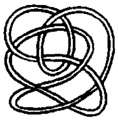
10_{83}

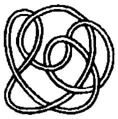
10_{84}

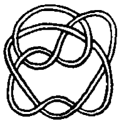
10_{85}

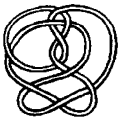
10_{86}

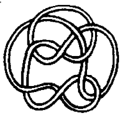
10_{87}

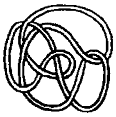
10_{88}

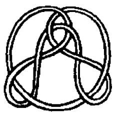
10_{89}

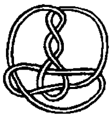
10_{90}

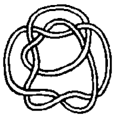
10_{91}

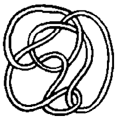
10_{92}

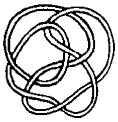
10_{93}

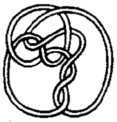
10_{94}

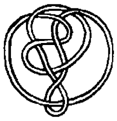
10_{95}

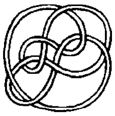
10_{96}

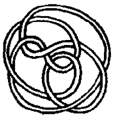
10_{97}

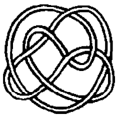
10_{98}

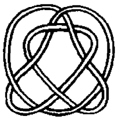
10_{99}

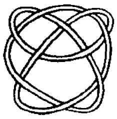
10_{100}

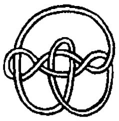
10_{101}

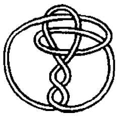
10_{102}

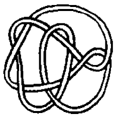
10_{103}

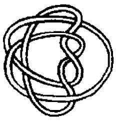
10_{104}

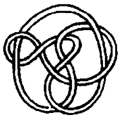
10_{105}

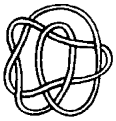
10_{106}

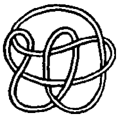
10_{107}

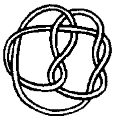
10_{108}

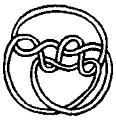
10_{109}

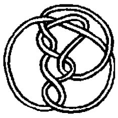
10_{110}

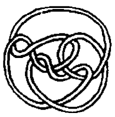
10_{111}

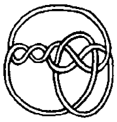
10_{112}

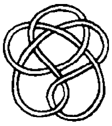
10_{113}

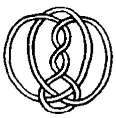
10_{114}

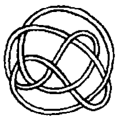
10_{115}

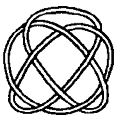
10_{116}

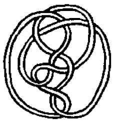
10_{117}

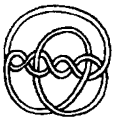
10_{118}

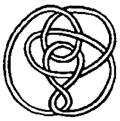
10_{119}

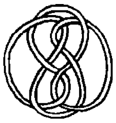
10_{120}

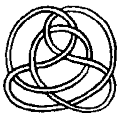
10_{121}

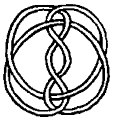
10_{122}

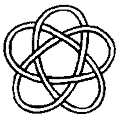
10_{123}

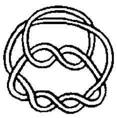
10_{124}

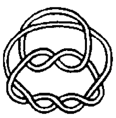
10_{125}

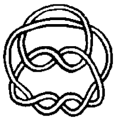
10_{126}

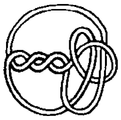
10_{127}

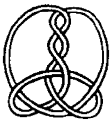
10_{128}

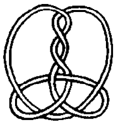
10_{129}

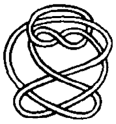
10_{130}

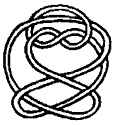
10_{131}

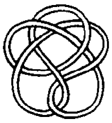
10_{132}

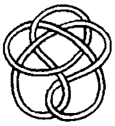
10_{133}

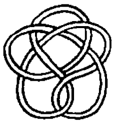
10_{134}

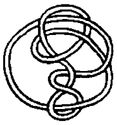
10_{135}

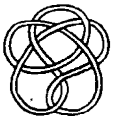
10_{136}

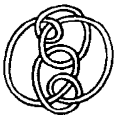
10_{137}

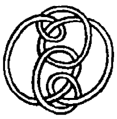
10_{138}

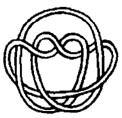
10_{139}

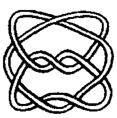
10_{140}

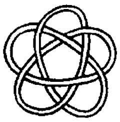
10_{141}

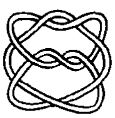
10_{142}

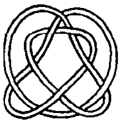
10_{143}

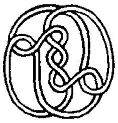
10_{144}

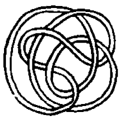
10_{145}

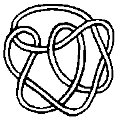
10_{146}

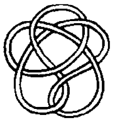
10_{147}

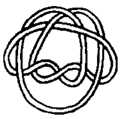
10_{148}

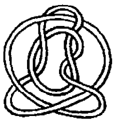
10_{149}

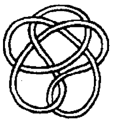
10_{150}

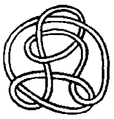
10_{151}

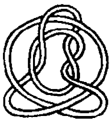
10_{152}

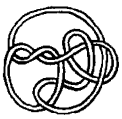
10_{153}

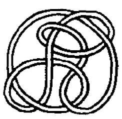
10_{154}

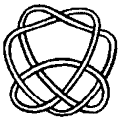
10_{155}

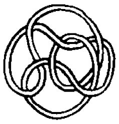
10_{156}

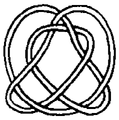
10_{157}

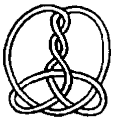
10_{158}

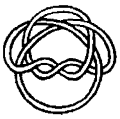
10_{159}

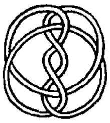
10_{160}

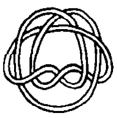
10_{161}

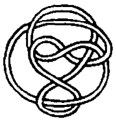
10_{162}

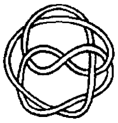
10_{163}

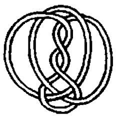
10_{164}

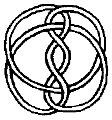
10_{165}

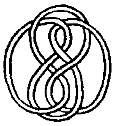
10_{166}

see also ALEXANDER POLYNOMIAL, ALEXANDER'S HORNED SPHERE, AMBIENT ISOTOPY, AMPHICHIRAL, ANTOINE'S NECKLACE, BEND (KNOT), BENNEQUIN'S CONJECTURE, BORROMEAN RINGS, BRAID GROUP, BRUNNIAN LINK, BURAU REPRESENTATION, CHEFALO KNOT, CLOVE HITCH, COLORABLE, CONWAY'S KNOT, CROOKEDNESS, DEHN'S LEMMA, DOWKER NOTATION, FIGURE-OF-EIGHT KNOT, GRANNY KNOT, HITCH, INVERTIBLE KNOT, JONES POLYNOMIAL, KINOSHITA-TERASAKA KNOT, KNOT POLYNOMIAL, KNOT SUM, LINKING NUMBER, LOOP (KNOT), MARKOV'S THEOREM, MENASCO'S THEOREM, MILNOR'S CONJECTURE, NASTY KNOT, PRETZEL KNOT, PRIME KNOT, REIDEMEISTER MOVES, RIBBON KNOT, RUNNING KNOT, SCHÖNFLIES THEOREM, SHORTENING, SIGNATURE (KNOT), SKEIN RELATIONSHIP, SLICE KNOT, SLIP KNOT, SMITH CONJECTURE, SOLOMON'S SEAL KNOT, SPAN (LINK), SPLITTING, SQUARE KNOT, STEVEDORE'S KNOT, STICK NUMBER, STOPPER KNOT, TAIT'S KNOT CONJECTURES, TAME KNOT, TANGLE, TORSION NUMBER, TREFOIL KNOT, UNKNOT, UNKNOTTING NUMBER, VASSILIEV POLYNOMIAL, WHITEHEAD LINK

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 280–286, 1994.

Adams, C.; Hildebrand, M.; and Weeks, J. "Hyperbolic Invariants of Knots and Links." *Trans. Amer. Math. Soc.* **1**, 1–56, 1991.

Anderson, J. "The Knotting Dictionary of Kännet." `http://www.netg.se/~jan/knopar/english/index.htm`.

Ashley, C. W. *The Ashley Book of Knots.* New York: McGraw-Hill, 1996.

Bogomolny, A. "Knots...." `http://www.cut-the-knot.com/do_you_know/knots.html`.

Conway, J. H. "An Enumeration of Knots and Links." In *Computational Problems in Abstract Algebra* (Ed. J. Leech). Oxford, England: Pergamon Press, pp. 329–358, 1970.

Eppstein, D. "Knot Theory." `http://www.ics.uci.edu/~eppstein/junkyard/knot.html`.

Eppstein, D. "Knot Theory." `http://www.ics.uci.edu/~eppstein/junkyard/knot/`.

Erdener, K.; Candy, C.; and Wu, D. "Verification and Extension of Topological Knot Tables." `ftp://chs.cusd.claremont.edu/pub/knot/FinalReport.sit.hqx`.

Ernst, C. and Sumner, D. W. "The Growth of the Number of Prime Knots." *Proc. Cambridge Phil. Soc.* **102**, 303–315, 1987.

Gordon, C. and Luecke, J. "Knots are Determined by their Complements." *J. Amer. Math. Soc.* **2**, 371–415, 1989.

Haken, W. "Theorie der Normalflachen." *Acta Math.* **105**, 245–375, 1961.

Kauffman, L. *Knots and Applications.* River Edge, NJ: World Scientific, 1995.

Kauffman, L. *Knots and Physics.* Teaneck, NJ: World Scientific, 1991.

Lickorish, W. B. R. and Millett, B. R. "The New Polynomial Invariants of Knots and Links." *Math. Mag.* **61**, 1–23, 1988.

Livingston, C. *Knot Theory.* Washington, DC: Math. Assoc. Amer., 1993.

Praslov, V. V. and Sossinsky, A. B. *Knots, Links, Braids and 3-Manifolds: An Introduction to the New Invariants in Low-Dimensional Topology.* Providence, RI: Amer. Math. Soc., 1996.

Rolfsen, D. "Table of Knots and Links." Appendix C in *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 280–287, 1976.

"Ropers Knots Page." `http://huizen.dds.nl/~erpprs/kne/kroot.htm`.

Sloane, N. J. A. Sequences A002863/M0851 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sloane, N. J. A. and Plouffe, S. Extended entry in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.

Stoimenow, A. "Polynomials of Knots with Up to 10 Crossings." Rev. March 16, 1998. `http://www.informatik.hu-berlin.de/~stoimeno/poly.ps`.

Suber, O. "Knots on the Web." `http://www.earlham.edu/suber/knotlink.htm`.

Tait, P. G. "On Knots I, II, and III." *Scientific Papers, Vol. 1.* Cambridge: University Press, pp. 273–347, 1898.

Thistlethwaite, M. B. "Knot Tabulations and Related Topics." In *Aspects of Topology in Memory of Hugh Dowker 1912–1982* (Ed. I. M. James and E. H. Kronheimer). Cambridge, England: Cambridge University Press, pp. 2–76, 1985.

Thistlethwaite, M. B. `ftp://chs.cusd.claremont.edu/pub/knot/Thistlethwaite_Tables/`.

Thompson, W. T. "On Vortex Atoms." *Philos. Mag.* **34**, 15–24, 1867.

Weisstein, E. W. "Knots." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Knots.m`.

Knot Complement

Two distinct knots cannot have the same Knot Complement (Gordon and Luecke 1989).

References

Cipra, B. "To Have and Have Knot: When are Two Knots Alike?" *Science* **241**, 1291–1292, 1988.

Gordon, C. and Luecke, J. "Knots are Determined by their Complements." *J. Amer. Math. Soc.* **2**, 371–415, 1989.

Knot Curve

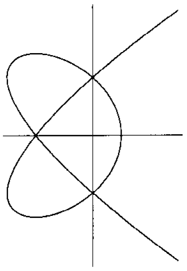

$$(x^2-1)^2 = y^2(3+2y).$$

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 72, 1989.

Knot Determinant

The determinant of a knot is $|\Delta(-1)|$, where $\Delta(z)$ is the Alexander Polynomial.

Knot Diagram

A picture of a projection of a Knot onto a Plane. Usually, only double points are allowed (no more than two points are allowed to be superposed), and the double or crossing points must be "genuine crossings" which transverse in the plane. This means that double points must look like the below diagram on the left, and not the one on the right.

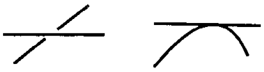

Also, it is usually demanded that a knot diagram contain the information if the crossings are overcrossings or undercrossings so that the original knot can be reconstructed. Here is a knot diagram of the Trefoil Knot,

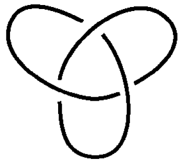

Knot Polynomials can be computed from knot diagrams. Such Polynomials often (but not always) allow the knots corresponding to given diagrams to be uniquely identified.

Knot Exterior

The COMPLEMENT of an open solid TORUS knotted at the KNOT. The removed open solid TORUS is called a tubular NEIGHBORHOOD.

Knot Linking

In general, it is possible to link two n-D HYPERSPHERES in $(n+2)$-D space in an infinite number of inequivalent ways. In dimensions greater than $n+2$ in the piecewise linear category, it is true that these spheres are themselves unknotted. However, they may still form nontrivial links. In this way, they are something like higher dimensional analogs of two 1-spheres in 3-D. The following table gives the number of nontrivial ways that two n-D HYPERSPHERES can be linked in k-D.

D of spheres	D of space	Distinct Linkings
23	40	239
31	48	959
102	181	3
102	182	10438319
102	183	3

Two 10-D HYPERSPHERES link up in 12, 13, 14, 15, and 16-D, then unlink in 17-D, link up again in 18, 19, 20, and 21-D. The proof of these results consists of an "easy part" (Zeeman 1962) and a "hard part" (Ravenel 1986). The hard part is related to the calculation of the (stable and unstable) HOMOTOPY GROUPS of SPHERES.

References

Bing, R. H. *The Geometric Topology of 3-Manifolds.* Providence, RI: Amer. Math. Soc., 1983.

Ravenel, D. *Complex Cobordism and Stable Homotopy Groups of Spheres.* New York: Academic Press, 1986.

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 7, 1976.

Zeeman. "Isotopies and Knots in Manifolds." In *Topology of 3-Manifolds and Related Topics* (Ed. M. K. Fort). Englewood Cliffs, NJ: Prentice-Hall, 1962.

Knot Polynomial

A knot invariant in the form of a POLYNOMIAL such as the ALEXANDER POLYNOMIAL, BLM/HO POLYNOMIAL, BRACKET POLYNOMIAL, CONWAY POLYNOMIAL, JONES POLYNOMIAL, KAUFFMAN POLYNOMIAL F, KAUFFMAN POLYNOMIAL X, and VASSILIEV POLYNOMIAL.

References

Lickorish, W. B. R. and Millett, K. C. "The New Polynomial Invariants of Knots and Links." *Math. Mag.* **61**, 3–23, 1988.

Knot Problem

The problem of deciding if two KNOTS in 3-space are equivalent such that one can be continuously deformed into another.

Knot Shadow

A LINK DIAGRAM which does not specify whether crossings are under- or overcrossings.

Knot Sum

Two oriented knots (or links) can be summed by placing them side by side and joining them by straight bars so that orientation is preserved in the sum. This operation is denoted #, so the knot sum of knots K_1 and K_2 is written

$$K_1 \# K_2 = K_2 \# K_1.$$

see also CONNECTED SUM

Knot Theory

The mathematical study of KNOTS. Knot theory considers questions such as the following:

1. Given a tangled loop of string, is it really knotted or can it, with enough ingenuity and/or luck, be untangled without having to cut it?
2. More generally, given two tangled loops of string, when are they deformable into each other?
3. Is there an effective algorithm (or any algorithm to speak of) to make these determinations?

Although there has been almost explosive growth in the number of important results proved since the discovery of the JONES POLYNOMIAL, there are still many "knotty" problems and conjectures whose answers remain unknown.

see also KNOT, LINK

Knot Vector

see B-SPLINE

Koch Antisnowflake

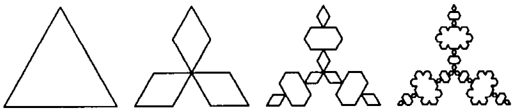

A FRACTAL derived from the KOCH SNOWFLAKE. The base curve and motif for the fractal are illustrated below.

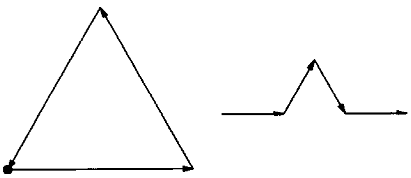

The AREA after the nth iteration is

$$A_n = A_{n-1} - \frac{1}{3}\frac{\ell_{n-1}}{a}\frac{\Delta}{3^n},$$

where Δ is the area of the original EQUILATERAL TRIANGLE, so from the derivation for the KOCH SNOWFLAKE,

$$A \equiv \lim_{n\to\infty} A_n = (1-\tfrac{3}{5})\Delta = \tfrac{2}{5}\Delta.$$

see also EXTERIOR SNOWFLAKE, FLOWSNAKE FRACTAL, KOCH SNOWFLAKE, PENTAFLAKE, SIERPIŃSKI CURVE

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 66–67, 1989.

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 36–37, 1991.

Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m`.

Koch Island

see KOCH SNOWFLAKE

Koch Snowflake

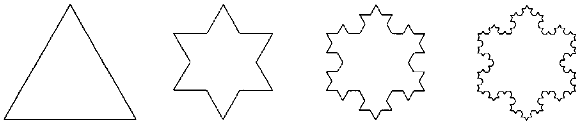

A FRACTAL, also known as the KOCH ISLAND, which was first described by Helge von Koch in 1904. It is built by starting with an EQUILATERAL TRIANGLE, removing the inner third of each side, building another EQUILATERAL TRIANGLE at the location where the side was removed, and then repeating the process indefinitely. The Koch snowflake can be simply encoded as a LINDENMAYER SYSTEM with initial string `"F--F--F"`, STRING REWRITING rule `"F" -> "F+F--F+F"`, and angle 60°. The zeroth through third iterations of the construction are shown above. The fractal can also be constructed using a base curve and motif, illustrated below.

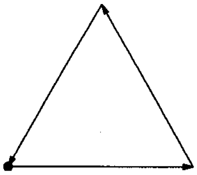

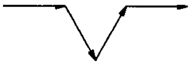

Let N_n be the number of sides, L_n be the length of a single side, ℓ_n be the length of the PERIMETER, and A_n the snowflake's AREA after the nth iteration. Further, denote the AREA of the initial $n = 0$ TRIANGLE Δ, and the length of an initial $n = 0$ side 1. Then

$$N_n = 3 \cdot 4^n \tag{1}$$

$$L_n = (\tfrac{1}{3})^n = 3^{-n} \tag{2}$$

$$\ell_n \equiv N_n L_n = 3(\tfrac{4}{3})^n \tag{3}$$

$$\begin{aligned} A_n &= A_{n-1} + \tfrac{1}{4} N_n {L_n}^2 \Delta = A_{n-1} + \frac{3 \cdot 4^n}{4} \left(\frac{1}{3}\right)^{2n} \Delta \\ &= A_{n-1} + \frac{3 \cdot 4^{n-1}}{9^n} \Delta = A_{n-1} + \frac{3 \cdot 4^{4-1}}{9 \cdot 9^{n-1}} \Delta \\ &= A_{n-1} + \tfrac{1}{3}(\tfrac{4}{9})^{n-1} \Delta. \end{aligned} \tag{4}$$

The CAPACITY DIMENSION is then

$$\begin{aligned} d_{\text{cap}} &= -\lim_{n \to \infty} \frac{\ln N_n}{\ln L_n} = -\lim_{n \to \infty} \frac{\ln(3 \cdot 4^n)}{\ln(3^{-n})} \\ &= \lim_{n \to \infty} \frac{\ln 3 + n \ln 4}{n \ln 3} \\ &= \frac{\ln 4}{\ln 3} = \frac{2 \ln 2}{\ln 3} = 1.261859507\ldots. \end{aligned} \tag{5}$$

Now compute the AREA explicitly,

$$A_0 = \Delta \tag{6}$$

$$A_1 = A_0 + \frac{1}{3}\left(\frac{4}{9}\right)^0 \Delta = \Delta \left\{1 + \frac{1}{3}\left(\tfrac{4}{9}\right)^0\right\} \tag{7}$$

$$A_2 = A_1 + \frac{1}{3}\left(\frac{4}{9}\right)^1 \Delta = \Delta \left\{1 + \frac{1}{3}\left[\left(\tfrac{4}{9}\right)^0 + \left(\tfrac{4}{9}\right)^1\right]\right\} \tag{8}$$

$$A_n = \left[1 + \frac{1}{3}\sum_{k=0}^{n} \left(\tfrac{4}{9}\right)^k\right] \Delta, \tag{9}$$

so as $n \to \infty$,

$$\begin{aligned} A \equiv A_\infty &= \left[1 + \frac{1}{3}\sum_{k=1}^{\infty} \left(\tfrac{4}{9}\right)^k\right] = \left(1 + \tfrac{1}{3}\frac{1}{1 - \frac{4}{9}}\right) \Delta \\ &= \tfrac{8}{5}\Delta. \end{aligned} \tag{10}$$

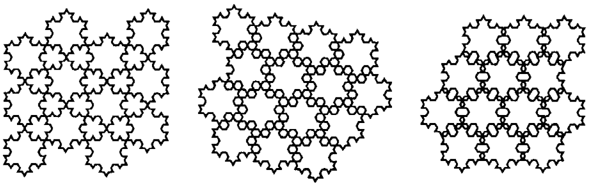

Some beautiful TILINGS, a few examples of which are illustrated above, can be made with iterations toward Koch snowflakes.

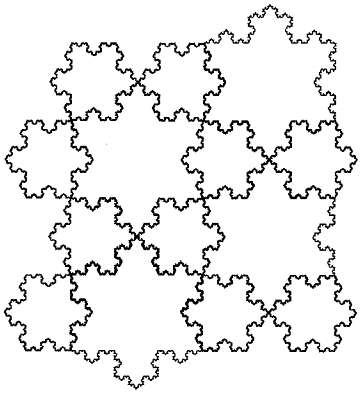

In addition, two sizes of Koch snowflakes in AREA ratio 1:3 TILE the PLANE, as shown above (Gosper).

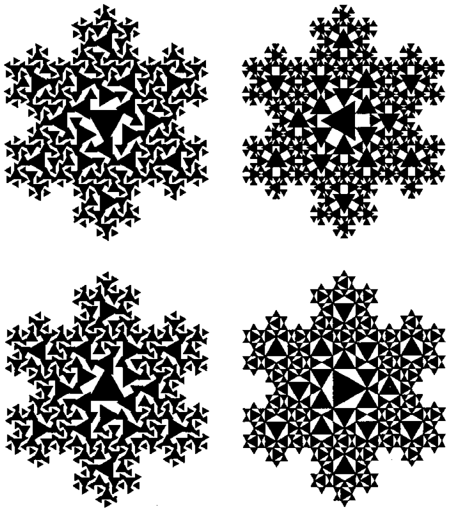

Another beautiful modification of the Koch snowflake involves inscribing the constituent triangles with filled-in triangles, possibly rotated at some angle. Some sample results are illustrated above for 3 and 4 iterations.

see also Cesàro Fractal, Exterior Snowflake, Gosper Island, Koch Antisnowflake, Peano-Gosper Curve, Pentaflake, Sierpiński Sieve

References
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 65–66, 1989.
Dickau, R. M. "Two-Dimensional L-Systems." `http://forum.swarthmore.edu/advanced/robertd/lsys2d.html`.
Dixon, R. *Mathographics.* New York: Dover, pp. 175–177 and 179, 1991.
Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 28–29 and 32–36, 1991.
Pappas, T. "The Snowflake Curve." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 78 and 160–161, 1989.
Peitgen, H.-O.; Jürgens, H.; and Saupe, D. *Chaos and Fractals: New Frontiers of Science.* New York: Springer-Verlag, 1992.
Peitgen, H.-O. and Saupe, D. (Eds.). "The von Koch Snowflake Curve Revisited." §C.2 in *The Science of Fractal Images.* New York: Springer-Verlag, pp. 275–279, 1988.
Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 185–195, 1991.
Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m`.

Kochansky's Approximation

The approximation for Pi,

$$\pi \approx \sqrt{\frac{40}{3} - \sqrt{12}} = 3.141533\ldots.$$

Koebe's Constant

A Constant equal to one Quarter, 1/4.

see also Quarter

References
Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 24, 1983.

Koebe Function

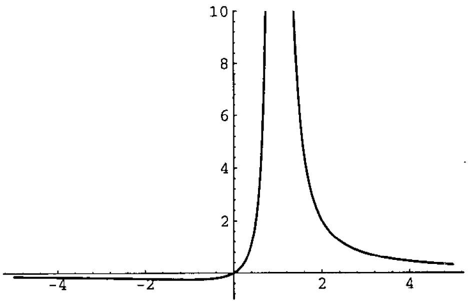

The function

$$f(z) \equiv \frac{z}{(1-z)^2}.$$

It has a Minimum at $z = -1$, where

$$f'(z) = -\frac{1+z}{(z-1)^3} = 0,$$

and an Inflection Point at $z = -2$, where

$$f''(z) = \frac{2(2+z)}{(z-1)^4} = 0.$$

References
Stewart, I. *From Here to Infinity: A Guide to Today's Mathematics.* Oxford, England: Oxford University Press, pp. 164–165, 1996.

Kollros' Theorem

For every ring containing p Spheres, there exists a ring of q Spheres, each touching each of the p Spheres, where

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{3}.$$

The Hexlet is a special case with $p = 3$.

see also Hexlet, Sphere

References
Honsberger, R. *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., p. 50, 1976.

Kolmogorov-Arnold-Moser Theorem

A theorem outlined in 1954 by Kolmogorov which was subsequently proved in the 1960s by Arnold and Moser (Tabor 1989, p. 105). It gives conditions under which Chaos is restricted in extent. Moser's 1962 proof was valid for Twist Maps

$$\theta' = \theta + 2\pi f(I) + g(\theta, I) \tag{1}$$
$$I' = I + f(\theta, I). \tag{2}$$

In 1963, Arnold produced a proof for Hamiltonian systems

$$H = H_0(\mathbf{I}) + \epsilon H_1(\mathbf{I}). \tag{3}$$

The original theorem required perturbations $\epsilon \sim 10^{-48}$, although this has since been significantly increased. Arnold's proof required C^∞, and Moser's original proof required C^{333}. Subsequently, Moser's version has been reduced to C^6, then $C^{2+\epsilon}$, although counterexamples are known for C^2. Conditions for applicability of the KAM theorem are:

1. small perturbations,
2. smooth perturbations, and
3. sufficiently irrational WINDING NUMBER.

Moser considered an integrable Hamiltonian function H_0 with a TORUS T_0 and set of frequencies ω having an incommensurate frequency vector $\boldsymbol{\omega}^*$ (i.e., $\boldsymbol{\omega}\cdot\mathbf{k} \neq 0$ for all INTEGERS k_i). Let H_0 be perturbed by some periodic function H_1. The KAM theorem states that, if H_1 is small enough, then for almost every ω^* there exists an invariant TORUS $T(\omega^*)$ of the perturbed system such that $T(\omega^*)$ is "close to" $T_0(\omega^*)$. Moreover, the TORI $T(\omega^*)$ form a set of POSITIVE measures whose complement has a measure which tends to zero as $|H_1| \to 0$. A useful paraphrase of the KAM theorem is, "For sufficiently small perturbation, almost all TORI (excluding those with rational frequency vectors) are preserved." The theorem thus explicitly excludes TORI with rationally related frequencies, that is, $n-1$ conditions of the form

$$\boldsymbol{\omega}\cdot\mathbf{k} = 0. \tag{4}$$

These TORI are destroyed by the perturbation. For a system with two DEGREES OF FREEDOM, the condition of closed orbits is

$$\sigma = \frac{\omega_1}{\omega_2} = \frac{r}{s}. \tag{5}$$

For a QUASIPERIODIC ORBIT, σ is IRRATIONAL. KAM shows that the preserved TORI satisfy the irrationality condition

$$\left|\frac{\omega_1}{\omega_2} - \frac{r}{s}\right| > \frac{K(\epsilon)}{s^{2.5}} \tag{6}$$

for all r and s, although not much is known about $K(\epsilon)$.

The KAM theorem broke the deadlock of the small divisor problem in classical perturbation theory, and provides the starting point for an understanding of the appearance of CHAOS. For a HAMILTONIAN SYSTEM, the ISOENERGETIC NONDEGENERACY condition

$$\left|\frac{\partial^2 H_0}{\partial I_j \partial I_j}\right| \neq 0 \tag{7}$$

guarantees preservation of most invariant TORI under small perturbations $\epsilon \ll 1$. The Arnold version states that

$$\left|\sum_{k=1}^n m_k\omega_k\right| > K(\epsilon)\left(\sum_{k=1}^n |m_k|\right)^{-n-1} \tag{8}$$

for all $m_k \in \mathbb{Z}$. This condition is less restrictive than Moser's, so fewer points are excluded.

see also CHAOS, HAMILTONIAN SYSTEM, QUASIPERIODIC FUNCTION, TORUS

References

Tabor, M. *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, 1989.

Kolmogorov Complexity

The complexity of a pattern parameterized as the shortest ALGORITHM required to reproduce it. Also known as ALGORITHMIC COMPLEXITY.

References

Goetz, P. "Phil's Good Enough Complexity Dictionary." `http://www.cs.buffalo.edu/~goetz/dict.html`.

Kolmogorov Constant

The exponent 5/3 in the spectrum of homogeneous turbulence, $k^{-5/3}$.

References

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 41, 1983.

Kolmogorov Entropy

Also known as METRIC ENTROPY. Divide PHASE SPACE into D-dimensional HYPERCUBES of CONTENT ϵ^D. Let $P_{i_0,\ldots,i_n}$ be the probability that a trajectory is in HYPERCUBE i_0 at $t = 0$, i_1 at $t = T$, i_2 at $t = 2T$, etc. Then define

$$K_n = h_K = -\sum_{i_0,\ldots,i_n} P_{i_0,\ldots,i_n} \ln P_{i_0,\ldots,i_n}, \tag{1}$$

where $K_{N+1} - K_N$ is the information needed to predict which HYPERCUBE the trajectory will be in at $(n+1)T$ given trajectories up to nT. The Kolmogorov entropy is then defined by

$$K \equiv \lim_{T\to 0}\lim_{\epsilon\to 0^+}\lim_{N\to\infty}\frac{1}{NT}\sum_{n=0}^{N-1}(K_{n+1} - K_n). \tag{2}$$

The Kolmogorov entropy is related to LYAPUNOV CHARACTERISTIC EXPONENTS by

$$h_K = \int_P \sum_{\sigma_i>0} \sigma_i \, d\mu. \tag{3}$$

see also HYPERCUBE, LYAPUNOV CHARACTERISTIC EXPONENT

References

Ott, E. *Chaos in Dynamical Systems.* New York: Cambridge University Press, p. 138, 1993.

Schuster, H. G. *Deterministic Chaos: An Introduction, 3rd ed.* New York: Wiley, p. 112, 1995.

Kolmogorov-Sinai Entropy

see KOLMOGOROV ENTROPY, METRIC ENTROPY

Kolmogorov-Smirnov Test

A goodness-of-fit test for any Distribution. The test relies on the fact that the value of the sample cumulative density function is asymptotically normally distributed.

To apply the Kolmogorov-Smirnov test, calculate the cumulative frequency (normalized by the sample size) of the observations as a function of class. Then calculate the cumulative frequency for a true distribution (most commonly, the Normal Distribution). Find the greatest discrepancy between the observed and expected cumulative frequencies, which is called the "D-Statistic." Compare this against the critical D-Statistic for that sample size. If the calculated D-Statistic is greater than the critical one, then reject the Null Hypothesis that the distribution is of the expected form. The test is an R-Estimate.

see also Anderson-Darling Statistic, D-Statistic, Kuiper Statistic, Normal Distribution, R-Estimate

References

Boes, D. C.; Graybill, F. A.; and Mood, A. M. *Introduction to the Theory of Statistics, 3rd ed.* New York: McGraw-Hill, 1974.

Knuth, D. E. §3.3.1B in *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, pp. 45–52, 1981.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Kolmogorov-Smirnov Test." In *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 617–620, 1992.

König-Egeváry Theorem

A theorem on Bipartite Graphs.

see also Bipartite Graph, Frobenius-König Theorem

König's Theorem

If an Analytic Function has a single simple Pole at the Radius of Convergence of its Power Series, then the ratio of the coefficients of its Power Series converges to that Pole.

see also Pole

References

König, J. "Über eine Eigenschaft der Potenzreihen." *Math. Ann.* **23**, 447–449, 1884.

Königsberg Bridge Problem

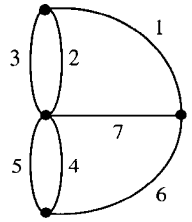

The Königsberg bridges cannot all be traversed in a single trip without doubling back. This problem was solved by Euler, and represented the beginning of Graph Theory.

see also Eulerian Circuit, Graph Theory

References

Bogomolny, A. "Graphs." `http://www.cut-the-knot.com/do_you_know/graphs.html`.

Chartrand, G. "The Königsberg Bridge Problem: An Introduction to Eulerian Graphs." §3.1 in *Introductory Graph Theory.* New York: Dover, pp. 51–66, 1985.

Kraitchik, M. §8.4.1 in *Mathematical Recreations.* New York: W. W. Norton, pp. 209–211, 1942.

Newman, J. "Leonhard Euler and the Königsberg Bridges." *Sci. Amer.* **189**, 66–70, 1953.

Pappas, T. "Königsberg Bridge Problem & Topology." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 124–125, 1989.

Korselt's Criterion

n Divides $a^n - a$ for all Integers a Iff n is Squarefree and $(p-1)|n/p-1$ for all Prime Divisors p of n. Carmichael Numbers satisfy this Criterion.

References

Borwein, D.; Borwein, J. M.; Borwein, P. B.; and Girgensohn, R. "Giuga's Conjecture on Primality." *Amer. Math. Monthly* **103**, 40–50, 1996.

Kovalevskaya Exponent

see Leading Order Analysis

Kozyrev-Grinberg Theory

A theory of Hamiltonian Circuits.

see also Grinberg Formula, Hamiltonian Circuit

Kramers Rate

The characteristic escape rate from a stable state of a potential in the absence of signal.

see also Stochastic Resonance

References

Bulsara, A. R. and Gammaitoni, L. "Tuning in to Noise." *Phys. Today* **49**, 39–45, March 1996.

Krawtchouk Polynomial

Let $\alpha(x)$ be a Step Function with the Jump

$$j(x) = \binom{N}{x} p^x q^{N-x} \qquad (1)$$

at $x = 0, 1, \ldots, N$, where $p > 0, q > 0$, and $p + q = 1$. Then

$$k_n^{(p)}(x) = \left[\binom{N}{n}\right]^{-1/2} (pq)^{-n/2} \times \sum_{\nu=0}^{n} (-1)^{n-\nu} \binom{N-x}{n-\nu} \binom{x}{\nu} p^{n-\nu} q^{\nu}, \qquad (2)$$

for $n = 0, 1, \ldots, N$. It has WEIGHT FUNCTION

$$w = \frac{N! p^x q^{N-x}}{\Gamma(1+x)\Gamma(N+1-x)}, \tag{3}$$

where $\Gamma(x)$ is the GAMMA FUNCTION, RECURRENCE RELATION

$$\begin{aligned}(n+1)k_{n+1}^{(p)}(x) + pq(N-n+1)k_{n-1}^{(p)}(x) \\ = [x-n-(N-2)]k_n^{(p)}(x),\end{aligned} \tag{4}$$

and squared norm

$$\frac{N!}{n!(N-n)!}(pq)^n. \tag{5}$$

It has the limit

$$\lim_{n\to\infty}\left(\frac{2}{Npq}\right)^{n/2} n! k_n^{(p)}(Np + \sqrt{2Npqs}) = H_n(s), \tag{6}$$

where $H_n(x)$ is a HERMITE POLYNOMIAL, and is related to the HYPERGEOMETRIC FUNCTION by

$$\begin{aligned}k_n^{(p)}(x,N) &= k_n^{(p)}(x,N) \\ &= (-1)^n \binom{N}{n} p^n {}_2F_1(-n,-x;-N;1/p) \\ &\quad \frac{(-1)^n p^n}{n!} \frac{\Gamma(N-x+1)}{\Gamma(N-x-n+1)} \\ &\quad \times {}_2F_1(-n,-x;N-x-n+1;-q/p).\end{aligned} \tag{7}$$

see also ORTHOGONAL POLYNOMIALS

References

Nikiforov, A. F.; Uvarov, V. B.; and Suslov, S. S. *Classical Orthogonal Polynomials of a Discrete Variable.* New York: Springer-Verlag, 1992.

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., pp. 35–37, 1975.

Zelenkov, V. "Krawtchouk Polynomial Home Page." `http://www.isir.minsk.by/~zelenkov/physmath/kr_polyn/`.

Kreisel Conjecture

A CONJECTURE in DECIDABILITY theory which postulates that, if there is a uniform bound to the lengths of shortest proofs of instances of $S(n)$, then the universal generalization is necessarily provable in PEANO ARITHMETIC. The CONJECTURE was proven true by M. Baaz in 1988 (Baaz and Pudlák 1993).

see also DECIDABLE

References

Baaz, M. and Pudlák P. "Kreisel's Conjecture for $L\exists_1$. In *Arithmetic, Proof Theory, and Computational Complexity, Papers from the Conference Held in Prague, July 2–5, 1991* (Ed. P. Clote and J. Krajíček). New York: Oxford University Press, pp. 30–60, 1993.

Dawson, J. "The Gödel Incompleteness Theorem from a Length of Proof Perspective." *Amer. Math. Monthly* **86**, 740–747, 1979.

Kreisel, G. "On the Interpretation of Nonfinitistic Proofs, II." *J. Symbolic Logic* **17**, 43–58, 1952.

Kronecker Decomposition Theorem

Every FINITE ABELIAN GROUP can be written as a DIRECT PRODUCT of CYCLIC GROUPS of PRIME POWER ORDERS. In fact, the number of nonisomorphic ABELIAN FINITE GROUPS $a(n)$ of any given ORDER n is given by writing n as

$$n = \prod_i {p_i}^{\alpha_i},$$

where the p_i are distinct PRIME FACTORS, then

$$a(n) = \prod_i P(\alpha_i),$$

where P is the PARTITION FUNCTION. This gives 1, 1, 1, 2, 1, 1, 1, 3, 2, ... (Sloane's A000688).

see also ABELIAN GROUP, FINITE GROUP, ORDER (GROUP), PARTITION FUNCTION P

References

Sloane, N. J. A. Sequence A000688/M0064 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Kronecker Delta

The simplest interpretation of the Kronecker delta is as the discrete version of the DELTA FUNCTION defined by

$$\delta_{ij} \equiv \begin{cases} 0 & \text{for } i \neq j \\ 1 & \text{for } i = j. \end{cases} \tag{1}$$

It has the COMPLEX GENERATING FUNCTION

$$\delta_{mn} = \frac{1}{2\pi i}\int z^{m-n-1}\,dz, \tag{2}$$

where m and n are INTEGERS. In 3-space, the Kronecker delta satisfies the identities

$$\delta_{ii} = 3 \tag{3}$$

$$\delta_{ij}\epsilon_{ijk} = 0 \tag{4}$$

$$\epsilon_{ipq}\epsilon_{jpq} = 2\delta_{ij} \tag{5}$$

$$\epsilon_{ijk}\epsilon_{pqk} = \delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}, \tag{6}$$

where EINSTEIN SUMMATION is implicitly assumed, $i, j = 1, 2, 3$, and ϵ is the PERMUTATION SYMBOL.

Technically, the Kronecker delta is a TENSOR defined by the relationship

$$\delta_l^k \frac{\partial x'_i}{\partial x_k}\frac{\partial x_l}{\partial x'_j} = \frac{\partial x'_i}{\partial x_k}\frac{\partial x_k}{\partial x'_j} = \frac{\partial x'_i}{\partial x'_j}. \tag{7}$$

Since, by definition, the coordinates x_i and x_j are independent for $i \neq j$,

$$\frac{\partial x'_i}{\partial x'_j} = {\delta'}^i_j, \tag{8}$$

so

$$\delta'^{i}_{j} = \frac{\partial x'_i}{\partial x_k}\frac{\partial x_l}{\partial x'_j}\delta^k_l, \tag{9}$$

and δ^i_j is really a mixed second RANK TENSOR. It satisfies

$$\delta_{ab}{}^{jk} = \epsilon_{abi}\epsilon^{jki} = \delta^j_a\delta^k_b - \delta^k_a\delta^j_b \tag{10}$$

$$\delta_{abjk} = g_{aj}g_{bk} - g_{ak}g_{bj} \tag{11}$$

$$\epsilon_{aij}\epsilon^{bij} = \delta_{ai}{}^{bi} = 2\delta^b_a. \tag{12}$$

see also DELTA FUNCTION, PERMUTATION SYMBOL

Kronecker's Polynomial Theorem

An algebraically soluble equation of ODD PRIME degree which is irreducible in the natural FIELD possesses either

1. Only a single REAL ROOT, or
2. All REAL ROOTS.

see also ABEL'S IRREDUCIBILITY THEOREM, ABEL'S LEMMA, SCHOENEMANN'S THEOREM

References

Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover p. 127, 1965.

Kronecker Product

see DIRECT PRODUCT (MATRIX)

Kronecker Symbol

An extension of the JACOBI SYMBOL (n/m) to all INTEGERS. It can be computed using the normal rules for the JACOBI SYMBOL

$$\left(\frac{ab}{cd}\right) = \left(\frac{a}{cd}\right)\left(\frac{b}{cd}\right) = \left(\frac{ab}{c}\right)\left(\frac{ab}{d}\right)$$
$$= \left(\frac{a}{c}\right)\left(\frac{b}{c}\right)\left(\frac{a}{d}\right)\left(\frac{b}{d}\right)$$

plus additional rules for $m = -1$,

$$(n/-1) = \begin{cases} -1 & \text{for } n < 0 \\ 1 & \text{for } n > 0, \end{cases}$$

and $m = 2$. The definition for $(n/2)$ is variously written as

$$(n/2) \equiv \begin{cases} 0 & \text{for } n \text{ even} \\ 1 & \text{for } n \text{ odd, } n \equiv \pm 1 \pmod 8 \\ -1 & \text{for } n \text{ odd, } n \equiv \pm 3 \pmod 8 \end{cases}$$

or

$$(n/2) \equiv \begin{cases} 0 & \text{for } 4|n \\ 1 & \text{for } n \equiv 1 \pmod 8 \\ -1 & \text{for } n \equiv 5 \pmod 8 \\ \text{undefined} & \text{otherwise} \end{cases}$$

(Cohn 1980). Cohn's form "undefines" $(n/2)$ for SINGLY EVEN NUMBERS $n \equiv 4 \pmod 2$ and $n \equiv -1, 3 \pmod 8$, probably because no other values are needed in applications of the symbol involving the DISCRIMINANTS d of QUADRATIC FIELDS, where $m > 0$ and d always satisfies $d \equiv 0, 1 \pmod 4$.

The KRONECKER SYMBOL is a REAL CHARACTER modulo n, and is, in fact, essentially the only type of REAL primitive character (Ayoub 1963).

see also CHARACTER (NUMBER THEORY), CLASS NUMBER, DIRICHLET L-SERIES, JACOBI SYMBOL, LEGENDRE SYMBOL

References

Ayoub, R. G. *An Introduction to the Analytic Theory of Numbers.* Providence, RI: Amer. Math. Soc., 1963.

Cohn, H. *Advanced Number Theory.* New York: Dover, p. 35, 1980.

Krull Dimension

If R is a RING (commutative with 1), the height of a PRIME IDEAL p is defined as the SUPREMUM of all n so that there is a chain $p_0 \subset \cdots p_{n-1} \subset p_n = p$ where all p_i are distinct PRIME IDEALS. Then, the Krull dimension of R is defined as the SUPREMUM of all the heights of all its PRIME IDEALS.

see also PRIME IDEAL

References

Eisenbud, D. *Commutative Algebra with a View Toward Algebraic Geometry.* New York: Springer-Verlag, 1995.

Macdonald, I. G. and Atiyah, M. F. *Introduction to Commutative Algebra.* Reading, MA: Addison-Wesley, 1969.

Kruskal's Algorithm

An ALGORITHM for finding a GRAPH's spanning TREE of minimum length.

see also KRUSKAL'S TREE THEOREM

References

Gardner, M. *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, pp. 248–249, 1978.

Kruskal's Tree Theorem

A theorem which plays a fundamental role in computer science because it is one of the main tools for showing that certain orderings on TREES are well-founded. These orderings play a crucial role in proving the termination of rewriting rules and the correctness of the Knuth-Bendix equational completion procedures.

see also KRUSKAL'S ALGORITHM, NATURAL INDEPENDENCE PHENOMENON, TREE

References

Gallier, J. "What's so Special about Kruskal's Theorem and the Ordinal Gamma[0]? A Survey of Some Results in Proof Theory." *Ann. Pure and Appl. Logic* **53**, 199–260, 1991.

KS Entropy

see METRIC ENTROPY

Kuen Surface

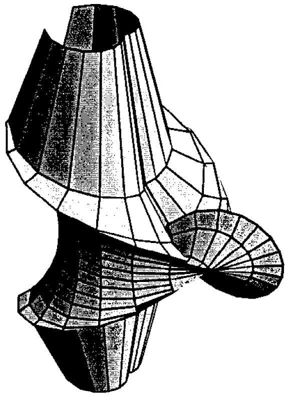

A special case of ENNEPER'S SURFACES which can be given parametrically by

$$x = \frac{2(\cos u + u \sin u)\sin v}{1 + u^2 \sin^2 v} \tag{1}$$

$$= \frac{2\sqrt{1+u^2}\,\cos(u - \tan^{-1} u)\sin v}{1 + u^2 \sin^2 v} \tag{2}$$

$$y = \frac{2(\sin u - u \cos u)\sin v}{1 + u^2 \sin^2 v} \tag{3}$$

$$= \frac{2\sqrt{1+u^2}\,\sin(u - \tan^{-1} u)\sin v}{1 + u^2 \sin^2 v} \tag{4}$$

$$z = \ln[\tan(\tfrac{1}{2}v)] + \frac{2\cos v}{1 + u^2 \sin^2 v} \tag{5}$$

for $v \in [0, \pi)$, $u \in [0, 2\pi)$ (Reckziegel *et al.* 1986). The Kuen surface has constant NEGATIVE GAUSSIAN CURVATURE of $K = -1$. The PRINCIPAL CURVATURES are given by

$$\kappa_1 = -\frac{u \cos(\frac{1}{2}v)[-2 - u^2 + u^2\cos(2v)]^4 \sin(\frac{1}{2}v)}{2[2 - u^2 + u^2\cos(2v)](1 + u^2 \sin^2 v)^4} \tag{6}$$

$$\kappa_2 = \frac{[-2 - u^2 + u^2\cos(2v)]^4 [2 - u^2 + u^2\cos(2v)]\csc(v)}{64u(1 + u^2 \sin^2 v)^4}. \tag{7}$$

see also ENNEPER'S SURFACES, REMBS' SURFACES, SIEVERT'S SURFACE

References

Fischer, G. (Ed.). Plate 86 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 82, 1986.

Gray, A. "Kuen's Surface." §19.4 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 384–386, 1993.

Kuen, T. "Ueber Flächen von constantem Krümmungsmaass." *Sitzungsber. d. königl. Bayer. Akad. Wiss. Math.-phys. Classe,* Heft II, 193–206, 1884.

Nordstrand, T. "Kuen's Surface." `http://www.uib.no/people/nfytn/kuentxt.htm`.

Reckziegel, H. "Kuen's Surface." §3.4.4.2 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, p. 38, 1986.

Kuhn-Tucker Theorem

A theorem in nonlinear programming which states that if a regularity condition holds and f and the functions h_j are convex, then a solution x^0 which satisfies the conditions h_j for a VECTOR of multipliers λ is a GLOBAL MINIMUM. The Kuhn-Tucker theorem is a generalization of LAGRANGE MULTIPLIERS. FARKAS'S LEMMA is key in proving this theorem.

see also FARKAS'S LEMMA, LAGRANGE MULTIPLIER

Kuiper Statistic

A statistic defined to improve the KOLMOGOROV-SMIRNOV TEST in the TAILS.

see also ANDERSON-DARLING STATISTIC

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 621, 1992.

Kulikowski's Theorem

For every POSITIVE INTEGER n, there exists a SPHERE which has exactly n LATTICE POINTS on its surface. The SPHERE is given by the equation

$$(x-a)^2 + (y-b)^2 + (z-\sqrt{2}\,)^2 = c^2 + 2,$$

where a and b are the coordinates of the center of the so-called SCHINZEL CIRCLE

$$\begin{cases} (x-\frac{1}{2})^2 + y^2 = \frac{1}{4}5^{k-1} & \text{for } n = 2k \text{ even} \\ (x-\frac{1}{3})^2 + y^2 = \frac{1}{9}5^{2k} & \text{for } n = 2k+1 \text{ odd} \end{cases}$$

and c is its RADIUS.

see also CIRCLE LATTICE POINTS, LATTICE POINT, SCHINZEL'S THEOREM

References

Honsberger, R. "Circles, Squares, and Lattice Points." Ch. 11 in *Mathematical Gems I.* Washington, DC: Math. Assoc. Amer., pp. 117–127, 1973.

Kulikowski, T. "Sur l'existence d'une sphère passant par un nombre donné aux coordonnées entières." *L'Enseignement Math. Ser. 2* **5**, 89–90, 1959.

Schinzel, A. "Sur l'existence d'un cercle passant par un nombre donné de points aux coordonnées entières." *L'Enseignement Math. Ser. 2* **4**, 71–72, 1958.

Sierpiński, W. "Sur quelques problèmes concernant les points aux coordonnées entières." *L'Enseignement Math. Ser. 2* **4**, 25–31, 1958.

Sierpiński, W. "Sur un problème de H. Steinhaus concernant les ensembles de points sur le plan." *Fund. Math.* **46**, 191–194, 1959.

Sierpiński, W. *A Selection of Problems in the Theory of Numbers.* New York: Pergamon Press, 1964.

Kummer's Conjecture

A conjecture concerning PRIMES.

Kummer's Differential Equation

see CONFLUENT HYPERGEOMETRIC DIFFERENTIAL EQUATION

Kummer's Formulas

Kummer's first formula is

$$ {}_2F_1(\tfrac{1}{2}+m-k,-n;2m+1;1) = \frac{\Gamma(2m+1)\Gamma(m+\frac{1}{2}+k+n)}{\Gamma(m+\frac{1}{2}+k)\Gamma(2m+1+n)}, \qquad (1) $$

where ${}_2F_1(a,b;c;z)$ is the HYPERGEOMETRIC FUNCTION with $m \neq -1/2, -1, -3/2, \ldots$, and $\Gamma(z)$ is the GAMMA FUNCTION. The identity can be written in the more symmetrical form as

$$ {}_2F_1(a,b;c;-1) = \frac{\Gamma(\frac{1}{2}b+1)\Gamma(b-a+1)}{\Gamma(b+1)\Gamma(\frac{1}{2}b-a+1)}, \qquad (2) $$

where $a-b+c=1$ and b is a positive integer. If b is a negative integer, the identity takes the form

$$ {}_2F_1(a,b;c;-1) = 2\cos(\tfrac{1}{2}\pi b)\frac{\Gamma(|b|)\Gamma(b-a+1)}{\Gamma(\frac{1}{2}b-a+1)} \qquad (3) $$

(Petkovšek *et al.* 1996).

Kummer's second formula is

$$ {}_1F_1(\tfrac{1}{2}+m;2m+1;z) = M_{0,m}(z) = z^{m+1/2}\left[1+\sum_{p=1}^{\infty}\frac{z^{2p}}{2^{4p}p!(m+1)(m+2)\cdots(m+p)}\right], \qquad (4) $$

where ${}_1F_1(a;b;z)$ is the CONFLUENT HYPERGEOMETRIC FUNCTION and $m \neq -1/2, -1, -3/2, \ldots$.

References

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B*. Wellesley, MA: A. K. Peters, pp. 42–43 and 126, 1996.

Kummer's Function

see CONFLUENT HYPERGEOMETRIC FUNCTION

Kummer Group

A GROUP of LINEAR FRACTIONAL TRANSFORMATIONS which transform the arguments of Kummer solutions to the HYPERGEOMETRIC DIFFERENTIAL EQUATION into each other. Define

$$ A(z) = 1-z $$
$$ B(z) = 1/z, $$

then the elements of the group are $\{I, A, B, AB, BA, ABA = BAB\}$.

Kummer's Quadratic Transformation

A transformation of a HYPERGEOMETRIC FUNCTION,

$$ {}_2F_1\left(\alpha,\beta;2\beta;\frac{4z}{(1+z)^2}\right) = (1+z)^{2\alpha}{}_2F_1(\alpha,\alpha+\tfrac{1}{2}-\beta;\beta+\tfrac{1}{2};z^2). $$

Kummer's Relation

An identity which relates HYPERGEOMETRIC FUNCTIONS,

$$ {}_2F_1(2a,2b;a+b+\tfrac{1}{2};x) = {}_2F_1(a,b;a+b+\tfrac{1}{2},4x(1-x)). $$

Kummer's Series

see HYPERGEOMETRIC FUNCTION

Kummer's Series Transformation

Let $\sum_{k=0}^{\infty} a_k = a$ and $\sum_{k=0}^{\infty} c_k = c$ be convergent series such that

$$ \lim_{k\to\infty}\frac{a_k}{c_k} = \lambda \neq 0. $$

Then

$$ a = \lambda c + \sum_{k=0}^{\infty}\left(1-\lambda\frac{c_k}{a_k}\right)a_k. $$

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 16, 1972.

Kummer Surface

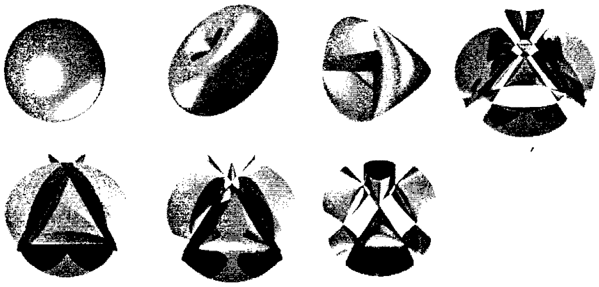

The Kummer surfaces are a family of QUARTIC SURFACES given by the algebraic equation

$$ (x^2+y^2+z^2-\mu^2w^2)^2 - \lambda pqrs = 0, \qquad (1) $$

where

$$ \lambda \equiv \frac{3\mu^2-1}{3-\mu^2}, \qquad (2) $$

p, q, r, and s are the TETRAHEDRAL COORDINATES

$$ p = w - z - \sqrt{2}\,x \qquad (3) $$
$$ q = w - z + \sqrt{2}\,x \qquad (4) $$
$$ r = w + z + \sqrt{2}\,y \qquad (5) $$
$$ s = w + z - \sqrt{2}\,y, \qquad (6) $$

and w is a parameter which, in the above plots, is set to $w = 1$. The above plots correspond to $\mu^2 = 1/3$

$$(3x^2 + 3y^2 + 3z^2 + 1)^2 = 0,$$

(double sphere), 2/3, 1

$$x^4 - 2x^2y^2 + y^4 + 4x^2z + 4y^2z + 4x^2z^2 + 4y^2z^2 = 0 \quad (7)$$

(ROMAN SURFACE), $\sqrt{2}$, $\sqrt{3}$

$$[(z-1)^2 - 2x^2][y^2 - (z+1)^2] = 0 \quad (8)$$

(four planes), 2, and 5. The case $0 \leq \mu^2 \leq 1/3$ corresponds to four real points.

The following table gives the number of ORDINARY DOUBLE POINTS for various ranges of μ^2, corresponding to the preceding illustrations.

Range	Real Nodes	Complex Nodes
$0 \leq \mu^2 \leq \frac{1}{3}$	4	12
$\mu^2 = \frac{1}{3}$		
$\frac{1}{3} \leq \mu^2 < 1$	4	12
$\mu^2 = 1$		
$1 < \mu^2 < 3$	16	0
$\mu^2 = 3$		
$\mu^2 > 3$	16	0

The Kummer surfaces can be represented parametrically by hyperelliptic THETA FUNCTIONS. Most of the Kummer surfaces admit 16 ORDINARY DOUBLE POINTS, the maximum possible for a QUARTIC SURFACE. A special case of a Kummer surface is the TETRAHEDROID.

Nordstrand gives the implicit equations as

$$x^4 + y^4 + z^4 - x^2 - y^2 - z^2 - x^2y^2 - x^2z^2 - y^2z^2 + 1 = 0 \quad (9)$$

or

$$x^4 + y^4 + z^4 + a(x^2 + y^2 + z^2) + b(x^2y^2 + x^2z^2 + y^2z^2) + cxyz - 1 = 0. \quad (10)$$

see also QUARTIC SURFACE, ROMAN SURFACE, TETRAHEDROID

References

Endraß, S. "Flächen mit vielen Doppelpunkten." *DMV-Mitteilungen* **4**, 17–20, Apr. 1995.

Endraß, S. "Kummer Surfaces." `http://www.mathematik.uni-mainz.de/Algebraische Geometrie/docs/Ekummer.shtml`.

Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, pp. 14–19, 1986.

Fischer, G. (Ed.). Plates 34–37 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 33–37, 1986.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 183, 1994.

Hudson, R. *Kummer's Quartic Surface.* Cambridge, England: Cambridge University Press, 1990.

Kummer, E. "Über die Flächen vierten Grades mit sechszehn singulären Punkten." *Ges. Werke* **2**, 418–432.

Kummer, E. "Über Strahlensysteme, deren Brennflächen Flächen vierten Grades mit sechszehn singulären Punkten sind." *Ges. Werke* **2**, 418–432.

Nordstrand, T. "Kummer's Surface." `http://www.uib.no/people/nfytn/kummtxt.htm`.

Kummer's Test

Given a SERIES of POSITIVE terms u_i and a sequence of finite POSITIVE constants a_i, let

$$\rho \equiv \lim_{n\to\infty} \left(a_n \frac{u_n}{u_{n+1}} - a_{n+1} \right).$$

1. If $\rho > 0$, the series converges.
2. If $\rho < 0$, the series diverges.
3. If $\rho = 0$, the series may converge or diverge.

The test is a general case of BERTRAND'S TEST, the ROOT TEST, GAUSS'S TEST, and RAABE'S TEST. With $a_n = n$ and $a_{n+1} = n + 1$, the test becomes RAABE'S TEST.

see also CONVERGENCE TESTS, RAABE'S TEST

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 285–286, 1985.

Jingcheng, T. "Kummer's Test Gives Characterizations for Convergence or Divergence of All Series." *Amer. Math. Monthly* **101**, 450–452, 1994.

Samelson, H. "More on Kummer's Test." *Amer. Math. Monthly* **102**, 817–818, 1995.

Kummer's Theorem

$${}_2F_1(x, -x; x + n + 1; -1) = \frac{\Gamma(x + n + 1)\Gamma(\frac{1}{2}n + 1)}{\Gamma(x + \frac{1}{2}n + 1)\Gamma(n + 1)}$$

$${}_2F_1(\alpha, \beta; 1 + \alpha - \beta; -1) = \frac{\Gamma(1 + \alpha - \beta)\Gamma(1 + \frac{1}{2}\alpha)}{\Gamma(1 + \alpha)\Gamma(1 + \frac{1}{2}\alpha - \beta)},$$

where ${}_2F_1$ is a HYPERGEOMETRIC FUNCTION and $\Gamma(z)$ is the GAMMA FUNCTION.

Kuratowski's Closure-Component Problem

Let X be an arbitrary TOPOLOGICAL SPACE. Denote the CLOSURE of a SUBSET A of X by A^- and the complement of A by A'. Then at most 14 different SETS can be derived from A by repeated application of closure and complementation (Berman and Jordan 1975, Fife 1991). The problem was first proved by Kuratowski (1922) and popularized by Kelley (1955).

see also KURATOWSKI REDUCTION THEOREM

References

Anusiak, J. and Shum, K. P. "Remarks on Finite Topological Spaces." *Colloq. Math.* **23**, 217–223, 1971.

Aull, C. E. "Classification of Topological Spaces." *Bull. de l'Acad. Pol. Sci. Math. Astron. Phys.* **15**, 773–778, 1967.
Baron, S. Advanced Problem 5569. *Amer. Math. Monthly* **75**, 199, 1968.
Berman, J. and Jordan, S. L. "The Kuratowski Closure-Complement Problem." *Amer. Math. Monthly* **82**, 841–842, 1975.
Buchman, E. "Problem E 3144." *Amer. Math. Monthly* **93**, 299, 1986.
Chagrov, A. V. "Kuratowski Numbers, Application of Functional Analysis in Approximation Theory." Kalinin: Kalinin Gos. Univ., pp. 186–190, 1982.
Chapman, T. A. "A Further Note on Closure and Interior Operators." *Amer. Math. Monthly* **69**, 524–529, 1962.
Fife, J. H. "The Kuratowski Closure-Complement Problem." *Math. Mag.* **64**, 180–182, 1991.
Fishburn, P. C. "Operations on Binary Relations." *Discrete Math.* **21**, 7–22, 1978.
Graham, R. L.; Knuth, D. E.; and Motzkin, T. S. "Complements and Transitive Closures." *Discrete Math.* **2**, 17–29, 1972.
Hammer, P. C. "Kuratowski's Closure Theorem." *Nieuw Arch. Wisk.* **8**, 74–80, 1960.
Herda, H. H. and Metzler, R. C. "Closure and Interior in Finite Topological Spaces." *Colloq. Math.* **15**, 211–216, 1966.
Kelley, J. L. *General Topology.* Princeton: Van Nostrand, p. 57, 1955.
Koenen, W. "The Kuratowski Closure Problem in the Topology of Convexity." *Amer. Math. Monthly* **73**, 704–708, 1966.
Kuratowski, C. "Sur l'operation A de l'analysis situs." *Fund. Math.* **3**, 182–199, 1922.
Langford, E. "Characterization of Kuratowski 14-Sets." *Amer. Math. Monthly* **78**, 362–367, 1971.
Levine, N. "On the Commutativity of the Closure and Interior Operators in Topological Spaces." *Amer. Math. Monthly* **68**, 474–477, 1961.
Moser, L. E. "Closure, Interior, and Union in Finite Topological Spaces." *Colloq. Math.* **38**, 41–51, 1977.
Munkres, J. R. *Topology: A First Course.* Englewood Cliffs, NJ: Prentice-Hall, 1975.
Peleg, D. "A Generalized Closure and Complement Phenomenon." *Discrete Math.* **50**, 285–293, 1984.
Shum, K. P. "On the Boundary of Kuratowski 14-Sets in Connected Spaces." *Glas. Mat. Ser. III* **19**, 293–296, 1984.
Shum, K. P. "The Amalgamation of Closure and Boundary Functions on Semigroups and Partially Ordered Sets." In *Proceedings of the Conference on Ordered Structures and Algebra of Computer Languages.* Singapore: World Scientific, pp. 232–243, 1993.
Smith, A. Advanced Problem 5996. *Amer. Math. Monthly* **81**, 1034, 1974.
Soltan, V. P. "On Kuratowski's Problem." *Bull. Acad. Polon. Sci. Ser. Sci. Math.* **28**, 369–375, 1981.
Soltan, V. P. "Problems of Kuratowski Type." *Mat. Issled.* **65**, 121–131 and 155, 1982.

Kuratowski Reduction Theorem

Every nonplanar graph is a SUPERGRAPH of an expansion of the UTILITY GRAPH $UG = K_{3,3}$ or the COMPLETE GRAPH K_5. This theorem was also proven earlier by Pontryagin (1927–1928), and later by Frink and Smith (1930). Kennedy *et al.* (1985) give a detailed history of the theorem, and there exists a generalization known as the ROBERTSON-SEYMOUR THEOREM.

see also COMPLETE GRAPH, PLANAR GRAPH, ROBERTSON-SEYMOUR THEOREM, UTILITY GRAPH

References

Kennedy, J. W.; Quintas, L. V.; and Syslo, M. M. "The Theorem on Planar Graphs." *Historia Math.* **12**, 356–368, 1985.
Kuratowski, C. "Sur l'operation A de l'analysis situs." *Fund. Math.* **3**, 182–199, 1922.
Thomassen, C. "Kuratowski's Theorem." *J. Graph Th.* **5**, 225–241, 1981.
Thomassen, C. "A Link Between the Jordan Curve Theorem and the Kuratowski Planarity Criterion." *Amer. Math. Monthly* **97**, 216–218, 1990.

Kuratowski's Theorem

see KURATOWSKI REDUCTION THEOREM

Kürschák's Tile

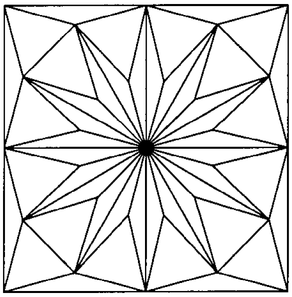

An attractive tiling of the SQUARE composed of two types of triangular tiles.

References

Alexanderson, G. L. and Seydel, K. "Kürschák's Tile." *Math. Gaz.* **62**, 192–196, 1978.
Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 30–32, 1985.
Schoenberg, I. *Mathematical Time Exposures.* Washington, DC: Math. Assoc. Amer., p. 7, 1982.
Weisstein, E. W. "Kürschák's Tile." `http://www.astro.virginia.edu/~eww6n/math/notebooks/KurschaksTile.m`.

Kurtosis

The degree of peakedness of a distribution, also called the EXCESS or EXCESS COEFFICIENT. Kurtosis is denoted γ_2 (or b_2) or β_2 and computed by taking the fourth MOMENT of a distribution. A distribution with a high peak ($\gamma_2 > 0$) is called LEPTOKURTIC, a flat-topped curve ($\gamma_2 < 0$) is called PLATYKURTIC, and the normal distribution ($\gamma_2 = 0$) is called MESOKURTIC. Let μ_i denote the ith MOMENT $\langle x^i \rangle$. The FISHER KURTOSIS is defined by

$$\gamma_2 \equiv b_2 \equiv \frac{\mu_4}{{\mu_2}^2} - 3 = \frac{\mu_4}{\sigma^4} - 3, \qquad (1)$$

and the PEARSON KURTOSIS is defined by

$$\beta_2 \equiv \alpha_4 \equiv \frac{\mu_4}{\sigma^4}. \qquad (2)$$

An ESTIMATOR for the γ_2 FISHER KURTOSIS is given by

$$g_2 = \frac{k_4}{{k_2}^2}, \qquad (3)$$

where the ks are k-STATISTICS. The STANDARD DEVIATION of the estimator is

$$\sigma_{g_2}{}^2 \approx \frac{24}{N}. \qquad (4)$$

see also FISHER KURTOSIS, MEAN, PEARSON KURTOSIS, SKEWNESS, STANDARD DEVIATION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 928, 1972.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Moments of a Distribution: Mean, Variance, Skewness, and So Forth." §14.1 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 604–609, 1992.

L

L_1-Norm

A VECTOR NORM defined for a VECTOR

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

with COMPLEX entries by

$$||\mathbf{x}||_1 = \sum_{r=1}^{n} |x_r|.$$

see also L_2-NORM, L_∞-NORM, VECTOR NORM

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 1114–1125, 1979.

L_2-Norm

A VECTOR NORM defined for a VECTOR

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

with COMPLEX entries by

$$||\mathbf{x}||_2 = \sqrt{\sum_{r=1}^{n} |x_r|^2}.$$

The L_2-norm is also called the EUCLIDEAN NORM. The L_2-norm is defined for a function $\phi(x)$ by

$$||\phi(x)|| \equiv \phi(x) \cdot \phi(x) \equiv \left\langle [\phi(x)]^2 \right\rangle \equiv \int_a^b [\phi(x)]^2 \, dx.$$

see also L_1-NORM, L_2-SPACE, L_∞-NORM, PARALLELOGRAM LAW, VECTOR NORM

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 1114–1125, 1979.

L_2-Space

A HILBERT SPACE in which a BRACKET PRODUCT is defined by

$$\langle \phi | \psi \rangle \equiv \int \psi^* \phi \, dx \tag{1}$$

and which satisfies the following conditions

$$\langle \phi | \psi \rangle^* = \langle \psi | \phi \rangle \, e \tag{2}$$

$$\langle \phi | \lambda_1 \psi_1 + \lambda_2 \psi_2 \rangle = \lambda_1 \langle \phi | \psi_1 \rangle + \lambda_2 \langle \phi | \psi_2 \rangle \tag{3}$$

$$\langle \lambda_1 \phi_1 + \lambda_2 \phi_2 | \psi \rangle = {\lambda_1}^* \langle \phi_1 | \psi \rangle + {\lambda_2}^* \langle \phi_2 | \psi \rangle \tag{4}$$

$$\langle \psi | \psi \rangle \in \mathbb{R} \geq 0 \tag{5}$$

$$| \langle \psi_1 | \psi_2 \rangle |^2 \leq \langle \psi_1 | \psi_1 \rangle \langle \psi_2 | \psi_2 \rangle . \tag{6}$$

The last of these is SCHWARZ'S INEQUALITY.

see also BRACKET PRODUCT, HILBERT SPACE, L_2-NORM, RIESZ-FISCHER THEOREM, SCHWARZ'S INEQUALITY

L_∞-Norm

A VECTOR NORM defined for a VECTOR

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

with COMPLEX entries by

$$||\mathbf{x}||_\infty = \max_i |x_i|.$$

see also L_1-NORM, L_2-NORM, VECTOR NORM

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 1114–1125, 1979.

$L_{p'}$-Balance Theorem

If every component L of $X/O_{p'}(X)$ satisfies the "Schreler property," then

$$L_{p'}(Y) \leq L_{p'}(X)$$

for every p-local SUBGROUP Y of X, where $L_{p'}$ is the p-LAYER.

see also p-LAYER, SUBGROUP

L-Estimate

A ROBUST ESTIMATION based on linear combinations of ORDER STATISTICS. Examples include the MEDIAN and TUKEY'S TRIMEAN.

see also M-ESTIMATE, R-ESTIMATE

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Robust Estimation." §15.7 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 694–700, 1992.

L-Function

see ARTIN L-FUNCTION, DIRICHLET L-SERIES, EULER L-FUNCTION, HECKE L-FUNCTION

L-Polyomino

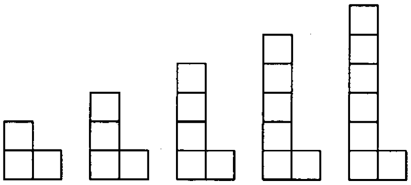

The order $n \geq 2$ L-polyomino consists of a vertical line of n SQUARES with a single additional SQUARE attached at the bottom.

see also L-POLYOMINO, SKEW POLYOMINO, SQUARE, SQUARE POLYOMINO, STRAIGHT POLYOMINO

L-Series

see DIRICHLET L-SERIES

L-System

see LINDENMAYER SYSTEM

L'Hospital's Cubic

see TSCHIRNHAUSEN CUBIC

L'Hospital's Rule

Let lim stand for the LIMIT $\lim_{x\to c}$, $\lim_{x\to c^-}$, $\lim_{x\to c^+}$, $\lim_{x\to\infty}$, or $\lim_{x\to-\infty}$, and suppose that $\lim f(x)$ and $\lim g(x)$ are both ZERO or are both $\pm\infty$. If

$$\lim \frac{f'(x)}{g'(x)}$$

has a finite value or if the LIMIT is $\pm\infty$, then

$$\lim \frac{f(x)}{g(x)} = \lim \frac{f'(x)}{g'(x)}.$$

L'Hospital's rule occasionally fails to yield useful results, as in the case of the function $\lim_{u\to\infty} u(u^2+1)^{-1/2}$. Repeatedly applying the rule in this case gives expressions which oscillate and never converge,

$$\begin{aligned}\lim_{u\to\infty} \frac{u}{(u^2+1)^{1/2}} &= \lim_{u\to\infty} \frac{1}{u(u^2+1)^{-1/2}} \\ &= \lim_{u\to\infty} \frac{(u^2+1)^{1/2}}{u} = \lim_{u\to\infty} \frac{u(u^2+1)^{-1/2}}{1} \\ &= \lim_{u\to\infty} \frac{u}{(u^2+1)^{1/2}}.\end{aligned}$$

(The actual LIMIT is 1.)

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 13, 1972.

L'Hospital, G. de *L'analyse des infiniment petits pour l'intelligence des lignes courbes.* 1696.

L'Huilier's Theorem

Let a SPHERICAL TRIANGLE have sides of length a, b, and c, and SEMIPERIMETER s. Then the SPHERICAL EXCESS Δ is given by

$$\begin{aligned}&\tan(\tfrac{1}{4}\Delta) \\ &\quad = \sqrt{\tan(\tfrac{1}{2}s)\tan[\tfrac{1}{2}(s-a)]\tan[\tfrac{1}{2}(s-b)]\tan[\tfrac{1}{2}(s-c)]}.\end{aligned}$$

see also GIRARD'S SPHERICAL EXCESS FORMULA, SPHERICAL EXCESS,SPHERICAL TRIANGLE

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 148, 1987.

Labelled Graph

A labelled graph $G = (V, E)$ is a finite series of VERTICES V with a set of EDGES E of 2-SUBSETS of V. Given a VERTEX set $V_n = \{1, 2, \ldots, n\}$, the number of labelled graphs is given by $2^{n(n-1)/2}$. Two graphs G and H with VERTICES $V_n = \{1, 2, \ldots, n\}$ are said to be ISOMORPHIC if there is a PERMUTATION p of V_n such that $\{u, v\}$ is in the set of EDGES $E(G)$ IFF $\{p(u), p(v)\}$ is in the set of EDGES $E(H)$.

see also CONNECTED GRAPH, GRACEFUL GRAPH, GRAPH (GRAPH THEORY), HARMONIOUS GRAPH, MAGIC GRAPH, TAYLOR'S CONDITION, WEIGHTED TREE

References

Cahit, I. "Homepage for the Graph Labelling Problems and New Results." `http://193.140.42.134/~cahit/CORDIAL.html`.

Gallian, J. A. "Graph Labelling." *Elec. J. Combin.* DS6, 1–43, Mar. 5, 1998. `http://www.combinatorics.org/Surveys/`.

Lacunarity

Quantifies deviation from translational invariance by describing the distribution of gaps within a set at multiple scales. The more lacunar a set, the more heterogeneous the spatial arrangement of gaps.

Ladder

see ASTROID, CROSSED LADDERS PROBLEM, LADDER GRAPH

Ladder Graph

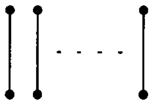

A GRAPH consisting of two rows of paired nodes each connected by an EDGE. Its complement is the COCKTAIL PARTY GRAPH.

see also COCKTAIL PARTY GRAPH

Lagrange Bracket

Let F and G be infinitely differentiable functions of x, u, and p. Then the Lagrange bracket is defined by

$$[F,G]=\sum_{\nu=1}^{n}\left[\frac{\partial F}{\partial p_\nu}\left(\frac{\partial G}{\partial x_p}+p_\nu\frac{\partial G}{\partial u}\right)\right.$$
$$\left.-\frac{\partial G}{\partial p_\nu}\left(\frac{\partial F}{\partial x_\nu}+p_\nu\frac{\partial F}{\partial u}\right)\right]. \quad (1)$$

The Lagrange bracket satisfies

$$[F,G]=-[G,F] \quad (2)$$

$$[[F,G],H]+[[G,H],F]+[[H,F],G]$$
$$=\frac{\partial F}{\partial u}[G,H]+\frac{\partial G}{\partial u}[H,F]+\frac{\partial H}{\partial u}[F,G]. \quad (3)$$

If F and G are functions of x and p only, then the Lagrange bracket $[F,G]$ collapses the Poisson Bracket (F,G).

see also Lie Bracket, Poisson Bracket

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1004, 1980.

Lagrange-Bürmann Theorem

see Lagrange Inversion Theorem

Lagrangian Coefficient

Coefficients which appear in Lagrange Interpolating Polynomials where the points are equally spaced along the Abscissa.

Lagrange's Continued Fraction Theorem

The Real Roots of quadratic expressions with integral Coefficients have periodic Continued Fractions, as first proved by Lagrange.

Lagrangian Derivative

see Convective Derivative

Lagrange's Equation

The Partial Differential Equation

$$(1+{f_y}^2)f_{xx}+2f_xf_yf_{xy}+(1+{f_x}^2)f_{yy}=0,$$

whose solutions are called Minimal Surfaces.

see also Minimal Surface

References

do Carmo, M. P. "Minimal Surfaces." §3.5 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 41–43, 1986.

Lagrange Expansion

Let $y=f(x)$ and $y_0=f(x_0)$ where $f'(x_0)\neq 0$, then

$$x=x_0+\sum_{k=1}^{\infty}\frac{(y-y_0)^k}{k!}\left\{\frac{d^{k-1}}{dx^{k-1}}\left[\frac{x-x_0}{f(x)-y_0}\right]^k\right\}_{x=x_0}$$

$$g(x)=g(x_0)$$
$$+\sum_{k=1}^{\infty}\frac{(y-y_0)^k}{k!}\left\{\frac{d^{k-1}}{dx^{k-1}}\left[g'(x)\left(\frac{x-x_0}{f(x)-y_0}\right)^k\right]\right\}_{x=x_0}$$

see also Maclaurin Series, Taylor Series

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 14, 1972.

Lagrange's Four-Square Theorem

A theorem also known as Bachet's Conjecture which was stated but not proven by Diophantus. It states that every Positive Integer can be written as the Sum of at most four Squares. Although the theorem was proved by Fermat using infinite descent, the proof was suppressed. Euler was unable to prove the theorem. The first published proof was given by Lagrange in 1770 and made use of the Euler Four-Square Identity.

see also Euler Four-Square Identity, Fermat's Polygonal Number Theorem, Fifteen Theorem, Vinogradov's Theorem, Waring's Problem

Lagrange's Group Theorem

Also known as Lagrange's Lemma. If A is an Element of a Finite Group of order n, then $A^n=1$. This implies that $e|n$ where e is the smallest exponent such that $A^e=1$. Stated another way, the Order of a Subgroup divides the Order of the Group. The converse of Lagrange's theorem is not, in general, true (Gallian 1993, 1994).

References

Birkhoff, G. and Mac Lane, S. *A Brief Survey of Modern Algebra, 2nd ed.* New York: Macmillan, p. 111, 1965.
Gallian, J. A. "On the Converse of Lagrange's Theorem." *Math. Mag.* **63**, 23, 1993.
Gallian, J. A. *Contemporary Abstract Algebra, 3rd ed.* Lexington, MA: D. C. Heath, 1994.
Herstein, I. N. *Abstract Algebra, 2nd ed.* New York: Macmillan, p. 66, 1990.
Hogan, G. T. "More on the Converse of Lagrange's Theorem." *Math. Mag.* **69**, 375–376, 1996.
Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 86, 1993.

Lagrange's Identity

The vector identity

$$(\mathbf{A}\times\mathbf{B})\cdot(\mathbf{C}\times\mathbf{D})=(\mathbf{A}\cdot\mathbf{C})(\mathbf{B}\cdot\mathbf{D})-(\mathbf{A}\cdot\mathbf{D})(\mathbf{B}\cdot\mathbf{C}). \quad (1)$$

This identity can be generalized to n-D,

$$(\mathbf{a}_1 \times \cdots \times \mathbf{a}_{n-1}) \cdot (\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}) = \begin{vmatrix} \mathbf{a}_1 \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_1 \cdot \mathbf{b}_{n-1} \\ \vdots & \ddots & \vdots \\ \mathbf{a}_{n-1} \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_{n-1} \cdot \mathbf{b}_{n-1} \end{vmatrix}, \quad (2)$$

where $|\mathsf{A}|$ is the DETERMINANT of A, or

$$\left(\sum_{k=1}^{n} a_k b_k\right)^2 = \left(\sum_{k=1}^{n} a_k{}^2\right)\left(\sum_{k=1}^{n} b_k{}^2\right) - \sum_{1 \leq k \leq j \leq n} (a_k b_j - a_j b_k)^2. \quad (3)$$

see also VECTOR TRIPLE PRODUCT, VECTOR QUADRUPLE PRODUCT

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1093, 1979.

Lagrange's Interpolating Fundamental Polynomial

Let $l(x)$ be an nth degree POLYNOMIAL with zeros at $x_1, \ldots, x_m$. Then the fundamental POLYNOMIALS are

$$l_\nu(x) = \frac{l(x)}{l'(x_\nu)(x - x_\nu)}. \quad (1)$$

They have the property

$$l_\nu(x) = \delta_{\nu\mu}, \quad (2)$$

where $\delta_{\nu\mu}$ is the KRONECKER DELTA. Now let $f_1, \ldots, f_n$ be values. Then the expansion

$$L_n(x) = \sum_{\nu=1}^{n} f_\nu l_\nu(x) \quad (3)$$

gives the unique LAGRANGE INTERPOLATING POLYNOMIAL assuming the values f_ν at x_ν. Let $d\alpha(x)$ be an arbitrary distribution on the interval $[a, b]$, $\{p_n(x)\}$ the associated ORTHOGONAL POLYNOMIALS, and $l_1(x), \ldots, l_n(x)$ the fundamental POLYNOMIALS corresponding to the set of zeros of $p_n(x)$. Then

$$\int_a^b l_\nu(x) l_\mu(x)\, d\alpha(x) = \lambda_\mu \delta_{\nu\mu} \quad (4)$$

for $\nu, \mu = 1, 2, \ldots, n$, where λ_ν are CHRISTOFFEL NUMBERS.

References

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., pp. 329 and 332, 1975.

Lagrange Interpolating Polynomial

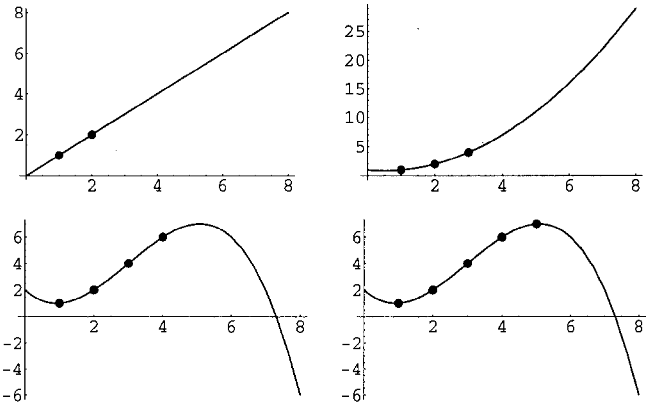

The Lagrange interpolating polynomial is the POLYNOMIAL of degree $n-1$ which passes through the n points $y_1 = f(x_1)$, $y_2 = f(x_2)$, ..., $y_n = f(x_n)$. It is given by

$$P(x) = \sum_{j=1}^{n} P_j(x), \quad (1)$$

where

$$P_j(x) = \prod_{\substack{k=1 \\ k \neq j}}^{n} \frac{x - x_k}{x_j - x_k} y_j. \quad (2)$$

Written explicitly,

$$\begin{aligned} P(x) = {} & \frac{(x-x_2)(x-x_3)\cdots(x-x_n)}{(x_1-x_2)(x_1-x_3)\cdots(x_1-x_n)} y_1 \\ & + \frac{(x-x_1)(x-x_3)\cdots(x-x_n)}{(x_2-x_1)(x_2-x_3)\cdots(x_2-x_n)} y_2 + \cdots \\ & + \frac{(x-x_1)(x-x_2)\cdots(x-x_{n-1})}{(x_n-x_1)(x_n-x_2)\cdots(x_n-x_{n-1})} y_n. \quad (3) \end{aligned}$$

For $n = 3$ points,

$$\begin{aligned} P(x) = {} & \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)} y_1 + \frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)} y_2 \\ & + \frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)} y_3 \quad (4) \end{aligned}$$

$$\begin{aligned} P'(x) = {} & \frac{2x-x_2-x_3}{(x_1-x_2)(x_1-x_3)} y_1 + \frac{2x-x_1-x_3}{(x_2-x_1)(x_2-x_3)} y_2 \\ & + \frac{2x-x_1-x_2}{(x_3-x_1)(x_3-x_2)} y_3. \quad (5) \end{aligned}$$

Note that the function $P(x)$ passes through the points (x_i, y_i), as can be seen for the case $n = 3$,

$$\begin{aligned} P(x_1) = {} & \frac{(x_1-x_2)(x_1-x_3)}{(x_1-x_2)(x_1-x_3)} y_1 + \frac{(x_1-x_1)(x_1-x_3)}{(x_2-x_1)(x_2-x_3)} y_2 \\ & + \frac{(x_1-x_1)(x_1-x_2)}{(x_3-x_1)(x_3-x_2)} y_3 = y_1 \quad (6) \end{aligned}$$

$$P(x_2) = \frac{(x_2-x_2)(x_2-x_3)}{(x_1-x_2)(x_1-x_3)} y_1 + \frac{(x_2-x_1)(x_2-x_3)}{(x_2-x_1)(x_2-x_3)} y_2$$

$$+\frac{(x_2-x_1)(x_2-x_2)}{(x_3-x_1)(x_3-x_2)}y_3 = y_2 \qquad (7)$$

$$P(x_3) = \frac{(x_3-x_2)(x_3-x_3)}{(x_1-x_2)(x_1-x_3)}y_1 + \frac{(x_3-x_1)(x_3-x_3)}{(x_2-x_1)(x_2-x_3)}y_2$$
$$+\frac{(x_3-x_1)(x_3-x_2)}{(x_3-x_1)(x_3-x_2)}y_3 = y_3. \qquad (8)$$

Generalizing to arbitrary n,

$$P(x_j) = \sum_{k=1}^{n} P_k(x_j) = \sum_{k=1}^{n} \delta_{jk} y_k = y_j. \qquad (9)$$

The Lagrange interpolating polynomials can also be written using

$$\pi(x) \equiv \prod_{k=1}^{n}(x-x_k), \qquad (10)$$

$$\pi(x_j) = \prod_{k=1}^{n}(x_j-x_k), \qquad (11)$$

$$\pi'(x_j) = \left[\frac{d\pi}{dx}\right]_{x=x_j} = \prod_{\substack{k=1\\k\neq j}}^{n}(x_j-x_k), \qquad (12)$$

so

$$P(x) = \sum_{k=1}^{n} \frac{\pi(x)}{(x-x_k)\pi'(x_k)} y_k. \qquad (13)$$

Lagrange interpolating polynomials give no error estimate. A more conceptually straightforward method for calculating them is NEVILLE'S ALGORITHM.

see also AITKEN INTERPOLATION, LEBESGUE CONSTANTS (LAGRANGE INTERPOLATION), NEVILLE'S ALGORITHM, NEWTON'S DIVIDED DIFFERENCE INTERPOLATION FORMULA

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 878–879 and 883, 1972.

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 439, 1987.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Polynomial Interpolation and Extrapolation" and "Coefficients of the Interpolating Polynomial." §3.1 and 3.5 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 102–104 and 113–116, 1992.

Lagrange Inversion Theorem

Let z be defined as a function of w in terms of a parameter α by

$$z = w + \alpha\phi(z).$$

Then any function of z can be expressed as a POWER SERIES in α which converges for sufficiently small α and has the form

$$F(z) = F(w) + \frac{\alpha}{1}\phi(w)F'(w) + \frac{\alpha^2}{1\cdot 2}\frac{\partial}{\partial w}\{[\phi(w)]^2 F'(w)\}$$
$$+\ldots+\frac{\alpha^{n+1}}{(n+1)!}\frac{\partial^n}{\partial w^n}\{[\phi(w)]^{n+1}F'(w)\}+\ldots.$$

References

Goursat, E. *Functions of a Complex Variable, Vol. 2, Pt. 1.* New York: Dover, 1959.

Moulton, F. R. *An Introduction to Celestial Mechanics, 2nd rev. ed.* New York: Dover, p. 161, 1970.

Williamson, B. "Remainder in Lagrange's Series." §119 in *An Elementary Treatise on the Differential Calculus, 9th ed.* London: Longmans, pp. 158–159, 1895.

Lagrange's Lemma

see LAGRANGE'S FOUR-SQUARE THEOREM

Lagrange Multiplier

Used to find the EXTREMUM of $f(x_1, x_2, \ldots, x_n)$ subject to the constraint $g(x_1, x_2, \ldots, x_n) = C$, where f and g are functions with continuous first PARTIAL DERIVATIVES on the OPEN SET containing the curve $g(x_1, x_2, \ldots, x_n) = 0$, and $\nabla g \neq \mathbf{0}$ at any point on the curve (where ∇ is the GRADIENT). For an EXTREMUM to exist,

$$df = \frac{\partial f}{\partial x_1}\,dx_1 + \frac{\partial f}{\partial x_2}\,dx_2 + \ldots + \frac{\partial f}{\partial x_n}\,dx_n = 0. \qquad (1)$$

But we also have

$$dg = \frac{\partial g}{\partial x_1}\,dx_1 + \frac{\partial g}{\partial x_2}\,dx_2 + \ldots + \frac{\partial g}{\partial x_n}\,dx_n = 0. \qquad (2)$$

Now multiply (2) by the as yet undetermined parameter λ and add to (1),

$$\left(\frac{\partial f}{\partial x_1} + \lambda\frac{\partial g}{\partial x_1}\right)dx_1 + \left(\frac{\partial f}{\partial x_2} + \lambda\frac{\partial g}{\partial x_2}\right)dx_2$$
$$+\ldots+\left(\frac{\partial f}{\partial x_n} + \lambda\frac{\partial g}{\partial x_n}\right)dx_n = 0. \qquad (3)$$

Note that the differentials are all independent, so we can set any combination equal to 0, and the remainder must still give zero. This requires that

$$\frac{\partial f}{\partial x_k} + \lambda\frac{\partial g}{\partial x_k} = 0 \qquad (4)$$

for all $k = 1, \ldots, n$. The constant λ is called the Lagrange multiplier. For multiple constraints, $g_1 = 0$, $g_2 = 0, \ldots,$

$$\nabla f = \lambda_1\nabla g_1 + \lambda_2\nabla g_2 + \ldots. \qquad (5)$$

see also KUHN-TUCKER THEOREM

References

Arfken, G. "Lagrange Multipliers." §17.6 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 945–950, 1985.

Lagrange Number (Diophantine Equation)

Given a Fermat Difference Equation (a quadratic Diophantine Equation)

$$x^2 - r^2y^2 = 4$$

with r a Quadratic Surd, assign to each solution $x|y$ the Lagrange number

$$z \equiv \tfrac{1}{2}(x + yr).$$

The product and quotient of two Lagrange numbers are also Lagrange numbers. Furthermore, every Lagrange number is a Power of the smallest Lagrange number with an integral exponent.

see also Pell Equation

References

Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 94–95, 1965.

Lagrange Number (Rational Approximation)

Hurwitz's Irrational Number Theorem gives the best rational approximation possible for an arbitrary irrational number α as

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{L_n q^2}.$$

The L_n are called Lagrange numbers and get steadily larger for each "bad" set of irrational numbers which is excluded.

n	Exclude	L_n
1	none	$\sqrt{5}$
2	ϕ	$\sqrt{8}$
3	$\sqrt{2}$	$\frac{\sqrt{221}}{5}$

Lagrange numbers are of the form

$$\sqrt{9 - \frac{4}{m^2}},$$

where m is a Markov Number. The Lagrange numbers form a Spectrum called the Lagrange Spectrum.

see also Hurwitz's Irrational Number Theorem, Liouville's Rational Approximation Theorem, Liouville-Roth Constant, Markov Number, Roth's Theorem, Spectrum Sequence, Thue-Siegel-Roth Theorem

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 187–189, 1996.

Lagrange Polynomial

see Lagrange Interpolating Polynomial

Lagrange Remainder

Given a Taylor Series, the error after n terms is bounded by

$$R_n = \frac{f^{(n)}(\xi)}{n!}(x-a)^n$$

for some $\xi \in (a, x)$.

see also Cauchy Remainder Form, Taylor Series

Lagrange Resolvent

A quantity involving primitive cube roots of unity which can be used to solve the Cubic Equation.

References

Faucette, W. M. "A Geometric Interpretation of the Solution of the General Quartic Polynomial." *Amer. Math. Monthly* **103**, 51–57, 1996.

Lagrange Spectrum

A Spectrum formed by the Lagrange Numbers. The only ones less than three are the Lagrange Numbers, but the last gaps end at Freiman's Constant. Real Numbers larger than Freiman's Constant are in the Markov Spectrum.

see also Freiman's Constant, Lagrange Number (Rational Approximation), Markov Spectrum, Spectrum Sequence

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 187–189, 1996.

Laguerre Differential Equation

$$xy'' + (1-x)y' + \lambda y = 0. \tag{1}$$

The Laguerre differential equation is a special case of the more general "associated Laguerre differential equation"

$$xy'' + (\nu + 1 - x)y' + \lambda y = 0 \tag{2}$$

with $\nu = 0$. Note that if $\lambda = 0$, then the solution to the associated Laguerre differential equation is of the form

$$y''(x) + P(x)y'(x) = 0, \tag{3}$$

and the solution can be found using an Integrating Factor

$$\begin{aligned}\mu &= \exp\left(\int P(x)\,dx\right) = \exp\left(\int \frac{\nu+1-x}{x}\,dx\right)\\ &= \exp[(\nu+1)\ln x - x] = x^{\nu+1}e^{-x},\end{aligned} \tag{4}$$

so

$$y = C_1 \int \frac{dx}{\mu} + C_2 = C_1 \int \frac{e^x}{x^{\nu+1}}\,dx + C_2. \tag{5}$$

The associated Laguerre differential equation has a Regular Singular Point at 0 and an Irregular

Singularity at ∞. It can be solved using a series expansion,

$$x\sum_{n=2}^{\infty} n(n-1)a_n x^{n-2} + (\nu+1)\sum_{n=1}^{\infty} n a_n x^{n-1} - x\sum_{n=1}^{\infty} n a_n x^{n-1} + \lambda\sum_{n=0}^{\infty} a_n x^n = 0 \tag{6}$$

$$\sum_{n=2}^{\infty} n(n-1)a_n x^{n-1} + (\nu+1)\sum_{n=1}^{\infty} n a_n x^{n-1} - \sum_{n=1}^{\infty} n a_n x^{n} + \lambda\sum_{n=0}^{\infty} a_n x^n = 0 \tag{7}$$

$$\sum_{n=1}^{\infty} (n+1)n a_{n+1} x^{n} + (\nu+1)\sum_{n=0}^{\infty} (n+1) a_{n+1} x^{n} - \sum_{n=1}^{\infty} n a_n x^{n} + \lambda\sum_{n=0}^{\infty} a_n x^n = 0 \tag{8}$$

$$[(n+1)a_1 + \lambda a_0] + \sum_{n=1}^{\infty}\{[(n+1)n + (\nu+1)(n+1)]a_{n+1} - n a_n + \lambda a_n\}x^n = 0 \tag{9}$$

$$[(n+1)a_1 + \lambda a_0] + \sum_{n=1}^{\infty}[(n+1)(n+\nu+1)a_{n+1} + (\lambda-n)a_n]x^n = 0. \tag{10}$$

This requires

$$a_1 = -\frac{\lambda}{\nu+1}a_0 \tag{11}$$

$$a_{n+1} = \frac{n-\lambda}{(n+1)(n+\nu+1)}a_n \tag{12}$$

for $n > 1$. Therefore,

$$a_{n+1} = \frac{n-\lambda}{(n+1)(n+\nu+1)}a_n \tag{13}$$

for $n = 1, 2, \ldots$, so

$$y = a_0\left[1 - \frac{\lambda}{\nu+1}x - \frac{\lambda(1-\lambda)}{2(\nu+1)(\nu+2)}x^2 - \frac{\lambda(1-\lambda)(2-\lambda)}{2\cdot 3(\nu+1)(\nu+2)(\nu+3)} + \cdots\right]. \tag{14}$$

If λ is a Positive Integer, then the series terminates and the solution is a Polynomial, known as an associated Laguerre Polynomial (or, if $\nu = 0$, simply a Laguerre Polynomial).

see also Laguerre Polynomial

Laguerre-Gauss Quadrature

Also called Gauss-Laguerre Quadrature or Laguerre Quadrature. A Gaussian Quadrature over the interval $[0, \infty)$ with Weighting Function $W(x) = e^{-x}$. The Abscissas for quadrature order n are given by the Roots of the Laguerre Polynomials $L_n(x)$. The weights are

$$w_i = -\frac{A_{n+1}\gamma_n}{A_n L_n'(x_i) L_{n+1}(x_i)} = \frac{A_n}{A_{n-1}}\frac{\gamma_{n-1}}{L_{n-1}(x_i)L_n'(x_i)}, \tag{1}$$

where A_n is the Coefficient of x^n in $L_n(x)$. For Laguerre Polynomials,

$$A_n = (-1)^n n!, \tag{2}$$

where $n!$ is a Factorial, so

$$\frac{A_{n+1}}{A_n} = -(n+1). \tag{3}$$

Additionally,

$$\gamma_n = 1, \tag{4}$$

so

$$w_i = \frac{n+1}{L_{n+1}(x_i)L_n'(x_i)} = -\frac{n}{L_{n-1}(x_i)L_n'(x_i)}. \tag{5}$$

(Note that the normalization used here is different than that in Hildebrand 1956.) Using the recurrence relation

$$\begin{aligned} xL_n'(x) &= nL_n(x) - nL_{n-1}(x) \\ &= (x-n-1)L_n(x) + (n+1)L_{n+1}(x) \end{aligned} \tag{6}$$

which implies

$$x_i L_n'(x_i) = -nL_{n-1}(x_i) = (n+1)L_{n+1}(x_i) \tag{7}$$

gives

$$w_i = \frac{1}{x_i[L_n'(x_i)]^2} = \frac{x_i}{(n+1)^2[L_{n+1}(x_i)]^2}. \tag{8}$$

The error term is

$$E = \frac{(n!)^2}{(2n)!}f^{(2n)}(\xi). \tag{9}$$

Beyer (1987) gives a table of Abscissas and weights up to $n = 6$.

n	x_i	w_i
2	0.585786	0.853553
	3.41421	0.146447
3	0.415775	0.711093
	2.29428	0.278518
	6.28995	0.0103893
4	0.322548	0.603154
	1.74576	0.357419
	4.53662	0.0388879
	9.39507	0.000539295
5	0.26356	0.521756
	1.4134	0.398667
	3.59643	0.0759424
	7.08581	0.00361176
	12.6408	0.00002337

The ABSCISSAS and weights can be computed analytically for small n.

n	x_i	w_i
2	$2-\sqrt{2}$	$\frac{1}{4}(2+\sqrt{2})$
	$2+\sqrt{2}$	$\frac{1}{4}(2-\sqrt{2})$

For the associated Laguerre polynomial $L_n^\beta(x)$ with WEIGHTING FUNCTION $w(x) = x^\beta e^{-x}$,

$$A_n = (-1)^n \tag{10}$$

and

$$\gamma_n = n! \int_0^\infty x^{\beta+n} e^{-x}\, dx = n!\Gamma(n+\beta+1). \tag{11}$$

The weights are

$$w_i = \frac{n!\Gamma(n+\beta+1)}{x_i[{L_m^\beta}'(x_i)]^2} = \frac{n!\Gamma(n+\beta+1)x_i}{[L_{n+1}^\beta(x_i)]^2}, \tag{12}$$

where $\Gamma(z)$ is the GAMMA FUNCTION, and the error term is

$$E_n = \frac{n!\Gamma(n+\beta+1)}{(2n)!} f^{(2n)}(\xi). \tag{13}$$

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 463, 1987.

Chandrasekhar, S. *Radiative Transfer.* New York: Dover, pp. 64–65, 1960.

Hildebrand, F. B. *Introduction to Numerical Analysis.* New York: McGraw-Hill, pp. 325–327, 1956.

Laguerre's Method

A ROOT-finding algorithm which converges to a COMPLEX ROOT from any starting position.

$$P_n(x) = (x-x_1)(x-x_2)\cdots(x-x_n) \tag{1}$$

$$\ln|P_n(x)| = \ln|x-x_1| + \ln|x-x_2| + \ldots + \ln|x-x_n| \tag{2}$$

$$\begin{aligned} P_n'(x) &= (x-x_2)\cdots(x-x_n) + (x-x_1)\cdots(x-x_n) + \ldots \\ &= P_n(x)\left(\frac{1}{x-x_1} + \ldots + \frac{1}{x-x_n}\right) \end{aligned} \tag{3}$$

$$\begin{aligned} \frac{d\ln|P_n(x)|}{dx} &= \frac{1}{x-x_1} + \frac{1}{x-x_2} + \ldots + \frac{1}{x-x_n} \\ &= \frac{P_n'(x)}{P_n(x)} \equiv G(x) \end{aligned} \tag{4}$$

$$\begin{aligned} &-\frac{d^2\ln|P_n(x)|}{dx^2} \\ &= \frac{1}{(x-x_1)^2} + \frac{1}{(x-x_2)^2} + \ldots + \frac{1}{(x-x_n)^2} \\ &= \left[\frac{P_n'(x)}{P_n(x)}\right]^2 - \frac{P_n''(x)}{P_n(x)} \equiv H(x). \end{aligned} \tag{5}$$

Now let $a \equiv x - x_1$ and $b \equiv x - x_1$. Then

$$G \equiv \frac{1}{a} + \frac{n-1}{b} \tag{6}$$

$$H \equiv \frac{1}{a^2} + \frac{n-1}{b^2}, \tag{7}$$

so

$$a = \frac{n}{\max\left[G \pm \sqrt{(n-1)(nH-G^2)}\right]}. \tag{8}$$

Setting $n = 2$ gives HALLEY'S IRRATIONAL FORMULA.

see also HALLEY'S IRRATIONAL FORMULA, HALLEY'S METHOD, NEWTON'S METHOD, ROOT

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 365–366, 1992.

Ralston, A. and Rabinowitz, P. §8.9–8.13 in *A First Course in Numerical Analysis, 2nd ed.* New York: McGraw-Hill, 1978.

Laguerre Polynomial

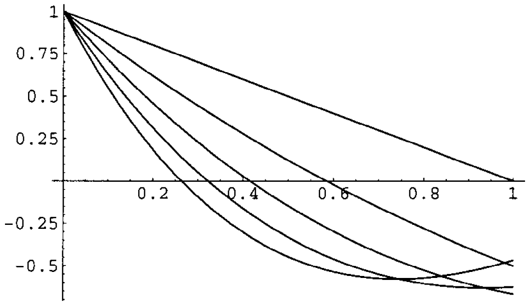

Solutions to the LAGUERRE DIFFERENTIAL EQUATION with $\nu = 0$ are called Laguerre polynomials. The Laguerre polynomials $L_n(x)$ are illustrated above for $x \in [0,1]$ and $n = 1, 2, \ldots, 5$.

The Rodrigues formula for the Laguerre polynomials is

$$L_n(x) = \frac{e^x}{n!}\frac{d^n}{dx^n}(x^n e^{-x}) \tag{1}$$

and the GENERATING FUNCTION for Laguerre polynomials is

$$\begin{aligned} g(x,z) &= \frac{\exp\left(-\frac{xz}{1-z}\right)}{1-z} = 1 + (-x+1)z \\ &+ (\tfrac{1}{2}x^2 - 2x + 1)z^2 + (-\tfrac{1}{6}x^3 + \tfrac{3}{2}x^2 - 3x + 1)z^3 + \ldots. \end{aligned} \tag{2}$$

A CONTOUR INTEGRAL is given by

$$L_n(x) = \frac{1}{2\pi i}\int \frac{e^{-xz/(1-z)}}{(1-z)z^{n+1}}\, dz. \tag{3}$$

The Laguerre polynomials satisfy the RECURRENCE RELATIONS

$$(n+1)L_{n+1}(x) = (2n+1-x)L_n(x) - nL_{n-1}(x) \quad (4)$$

(Petkovšek *et al.* 1996) and

$$xL_n'(x) = nL_n(x) - nL_{n-1}(x). \quad (5)$$

The first few Laguerre polynomials are

$$\begin{aligned} L_0(x) &= 1 \\ L_1(x) &= -x+1 \\ L_2(x) &= \tfrac{1}{2}(x^2-4x+2) \\ L_3(x) &= \tfrac{1}{6}(-x^3+9x^2-18x+6). \end{aligned}$$

Solutions to the associated LAGUERRE DIFFERENTIAL EQUATION with $\nu \neq 0$ are called associated Laguerre polynomials $L_n^k(x)$. In terms of the normal Laguerre polynomials,

$$L_n(x) = L_n^0(x). \quad (6)$$

The Rodrigues formula for the associated Laguerre polynomials is

$$\begin{aligned} L_n^k(x) &= \frac{e^x x^{-k}}{n!}\frac{d^n}{dx^n}(e^{-x}x^{n+k}) \\ &= (-1)^n \frac{d^n}{dx^n}[L_{n+k}(x)] \quad (7) \\ &= \sum_{m=0}^{\infty}(-1)^m \frac{(n+k)!}{(n-m)!(k+m)!m!}x^m \quad (8) \end{aligned}$$

and the GENERATING FUNCTION is

$$g(x,z) = \frac{\exp\left(-\frac{xz}{1-z}\right)}{(1-z)^{k+1}}$$

$$1+(k+1-x)z+\tfrac{1}{2}[x^2-2(k+2)x+(k+1)(k+2)]z^2+\ldots. \quad (9)$$

The associated Laguerre polynomials are orthogonal over $[0,\infty)$ with respect to the WEIGHTING FUNCTION $x^n e^{-x}$.

$$\int_0^\infty e^{-x}x^k L_n^k(x)L_m^k(x)\,dx = \frac{(n+k)!}{n!}\delta_{mn}, \quad (10)$$

where δ_{mn} is the KRONECKER DELTA. They also satisfy

$$\int_0^\infty e^{-x}x^{k+1}[L_n^k(x)]^2\,dx = \frac{(n+k)!}{n!}(2n+k+1). \quad (11)$$

RECURRENCE RELATIONS include

$$\sum_{\nu=0}^{n} L_\nu^{(\alpha)}(x) = L_n^{(\alpha+1)}(x) \quad (12)$$

and

$$L_n^{(\alpha)}(x) = L_n^{(\alpha+1)}(x) - L_{n-1}^{(\alpha+1)}(x). \quad (13)$$

The DERIVATIVE is given by

$$\begin{aligned} \frac{d}{dx}L_n^{(\alpha)}(x) &= -L_{n-1}^{(\alpha+1)}(x) \\ &= x^{-1}[nL_n^{(\alpha)}(x) - (n+\alpha)L_{n-1}^{(\alpha)}(x). \quad (14) \end{aligned}$$

In terms of the CONFLUENT HYPERGEOMETRIC FUNCTION,

$$L_n^k(x) = \frac{(k+1)_n}{n!}\,{}_1F_1(-b;k+1;x). \quad (15)$$

An interesting identity is

$$\sum_{n=0}^{\infty}\frac{L_n^{(\alpha)}(x)}{\Gamma(n+\alpha+1)}w^n = e^w(xw)^{-\alpha/2}J_\alpha(2\sqrt{xw}\,), \quad (16)$$

where $\Gamma(z)$ is the GAMMA FUNCTION and $J_\alpha(z)$ is the BESSEL FUNCTION OF THE FIRST KIND (Szegő 1975, p. 102). An integral representation is

$$e^{-x}x^{\alpha/2}L_n^{(\alpha)}(x) = \frac{1}{n!}\int_0^\infty e^{-t}t^{n+\alpha/2}J_\alpha(2\sqrt{tx}\,)\,dt \quad (17)$$

for $n = 0, 1, \ldots$ and $\alpha > -1$. The DISCRIMINANT is

$$D_n^{(\alpha)} = \prod_{\nu=1}^{n}\nu^{\nu-2n+2}(\nu+\alpha)^{\nu-1} \quad (18)$$

(Szegő 1975, p. 143). The KERNEL POLYNOMIAL is

$$K_n^{(\alpha)}(x,y) = \frac{n+1}{\Gamma(\alpha+1)}\binom{n+\alpha}{n}^{-1} \frac{L_n^{(\alpha)}(x)L_{n+1}^{(\alpha)}(y) - L_{n+1}^{(\alpha)}(x)L_n(\alpha)(y)}{x-y}, \quad (19)$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT (Szegő 1975, p. 101).

The first few associated Laguerre polynomials are

$$\begin{aligned} L_0^k(x) &= 1 \\ L_1^k(x) &= -x+k+1 \\ L_2^k(x) &= \tfrac{1}{2}[x^2-2(k+2)x+(k+1)(k+2)] \\ L_3^k(x) &= \tfrac{1}{6}[-x^3+3(k+3)x^2-3(k+2)(k+3)x \\ &\quad +(k+1)(k+2)(k+3)]. \end{aligned}$$

see also SONINE POLYNOMIAL

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Orthogonal Polynomials." Ch. 22 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 771–802, 1972.

Arfken, G. "Laguerre Functions." §13.2 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 721–731, 1985.

Chebyshev, P. L. "Sur le développement des fonctions à une seule variable." *Bull. Ph.-Math., Acad. Imp. Sc. St. Pétersbourg* **1**, 193–200, 1859.

Chebyshev, P. L. *Oeuvres, Vol. 1.* New York: Chelsea, pp. 499–508, 1987.

Iyanaga, S. and Kawada, Y. (Eds.). "Laguerre Functions." Appendix A, Table 20.VI in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1481, 1980.

Laguerre, E. de. "Sur l'intégrale $\int_x^{+\infty} x^{-1}e^{-x}\,dx$." *Bull. Soc. math. France* **7**, 72–81, 1879. Reprinted in *Oeuvres, Vol. 1.* New York: Chelsea, pp. 428–437, 1971.

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, pp. 61–62, 1996.

Sansone, G. "Expansions in Laguerre and Hermite Series." Ch. 4 in *Orthogonal Functions, rev. English ed.* New York: Dover, pp. 295–385, 1991.

Spanier, J. and Oldham, K. B. "The Laguerre Polynomials $L_n(x)$." Ch. 23 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 209–216, 1987.

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., 1975.

Laguerre Quadrature

A GAUSSIAN QUADRATURE-like FORMULA for numerical estimation of integrals. It fits exactly all POLYNOMIALS of degree $2m-1$.

References

Chandrasekhar, S. *Radiative Transfer.* New York: Dover, p. 61, 1960.

Laguerre's Repeated Fraction

The CONTINUED FRACTION

$$\frac{(x+1)^n-(x-1)^n}{(x+1)^n+(x-1)^n}=\frac{n}{x+}\,\frac{n^2-1}{3x+}\,\frac{n^2-2^2}{5x+\ldots}.$$

References

Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, p. 13, 1959.

Laisant's Recurrence Formula

The RECURRENCE RELATION

$$(n-1)A_{n+1}=(n^2-1)A_n+(n+1)A_{n-1}+4(-1)^n$$

with $A(1)=A(2)=1$ which solves the MARRIED COUPLES PROBLEM.

see also MARRIED COUPLES PROBLEM

Lakshmi Star

see STAR OF LAKSHMI

Lal's Constant

Let $P(N)$ denote the number of PRIMES of the form n^2+1 for $1\leq n\leq N$, then

$$P(N)\sim 0.68641\,\mathrm{li}(N), \tag{1}$$

where $\mathrm{li}(N)$ is the LOGARITHMIC INTEGRAL (Shanks 1960, pp. 321–332). Let $Q(N)$ denote the number of PRIMES of the form n^4+1 for $1\leq n\leq N$, then

$$Q(N)\sim \tfrac{1}{4}s_1\,\mathrm{li}(N)=0.66974\,\mathrm{li}(N) \tag{2}$$

(Shanks 1961, 1962). Let $R(N)$ denote the number of pairs of PRIMES $(n-1)^2+1$ and $(n+1)^2+1$ for $n\leq N-1$, then

$$R(N)\sim 0.48762\,\mathrm{li}_2(N), \tag{3}$$

where

$$\mathrm{li}_2(N)\equiv\int_2^N\frac{dn}{(\ln n)^2} \tag{4}$$

(Shanks 1960, pp. 201–203). Finally, let $S(N)$ denote the number of pairs of PRIMES $(n-1)^4+1$ and $(n+1)^4+1$ for $n\leq N-1$, then

$$S(N)\sim\lambda\,\mathrm{li}_2(N) \tag{5}$$

(Lal 1967), where λ is called Lal's constant. Shanks (1967) showed that $\lambda\approx 0.79220$.

References

Lal, M. "Primes of the Form n^4+1." *Math. Comput.* **21**, 245–247, 1967.

Shanks, D. "On the Conjecture of Hardy and Littlewood Concerning the Number of Primes of the Form n^2+a." *Math. Comput.* **14**, 321–332, 1960.

Shanks, D. "On Numbers of the Form n^4+1." *Math. Comput.* **15**, 186–189, 1961.

Shanks, D. Corrigendum to "On the Conjecture of Hardy and Littlewood Concerning the Number of Primes of the Form n^2+a." *Math. Comput.* **16**, 513, 1962.

Shanks, D. "Lal's Constant and Generalization." *Math. Comput.* **21**, 705–707, 1967.

Lam's Problem

Given an 111×111 MATRIX, fill 11 spaces in each row in such a way that all columns also have 11 spaces filled. Furthermore, each pair of rows must have exactly one filled space in the same column. This problem is equivalent to finding a PROJECTIVE PLANE of order 10. Using a computer program, Lam showed that no such arrangement exists.

see also PROJECTIVE PLANE

Laman's Theorem

Let a GRAPH G have exactly $2n-3$ EDGES, where n is the number of VERTICES in G. Then G is "generically" RIGID in $\mathbb{R}^2$ IFF $e'\leq 2n'-3$ for every SUBGRAPH of G having n' VERTICES and r' EDGES.

see also RIGID

References

Laman, G. "On Graphs and Rigidity of Plane Skeletal Structures." *J. Engineering Math.* **4**, 331–340, 1970.

Lambda Calculus

Developed by Alonzo Church and Stephen Kleene to address the COMPUTABLE NUMBER problem. In the lambda calculus, λ is defined as the ABSTRACTION OPERATOR. Three theorems of lambda calculus are λ-conversion, α-conversion, and η-conversion.

see also ABSTRACTION OPERATOR, COMPUTABLE NUMBER

References

Penrose, R. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics.* Oxford, England: Oxford University Press, pp. 66–70, 1989.

Lambda Function

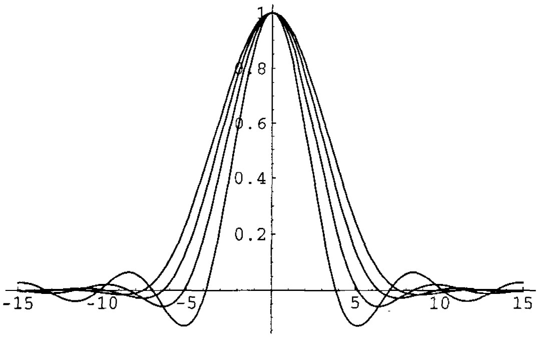

The lambda function defined by Jahnke and Emden (1945) is

$$\Lambda_\nu(z) \equiv \Gamma(\nu+1)\frac{J_\nu(z)}{(\frac{1}{2}z)^\nu} \tag{1}$$

$$\Lambda_1(z) \equiv \frac{J_1(z)}{\frac{z}{2}} = 2\,\mathrm{jinc}(z), \tag{2}$$

where $J_1(z)$ is a BESSEL FUNCTION OF THE FIRST KIND and $\mathrm{jinc}(z)$ is the JINC FUNCTION.

A two-variable lambda function defined by Gradshteyn and Ryzhik (1979) is

$$\lambda(x,y) \equiv \int_0^y \frac{\Gamma(u+1)\,du}{x^u}, \tag{3}$$

where $\Gamma(z)$ is the GAMMA FUNCTION.

see also AIRY FUNCTIONS, DIRICHLET LAMBDA FUNCTION, ELLIPTIC LAMBDA FUNCTION, JINC FUNCTION, LAMBDA HYPERGEOMETRIC FUNCTION, MANGOLDT FUNCTION, MU FUNCTION, NU FUNCTION

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1079, 1979.

Jahnke, E. and Emde, F. *Tables of Functions with Formulae and Curves, 4th ed.* New York: Dover, 1945.

Lambda Group

The set of linear fractional transformations w which satisfy

$$w(t) = \frac{at+b}{ct+d},$$

where a and d are ODD and b and c are EVEN. Also called the THETA SUBGROUP. It is a SUBGROUP of the GAMMA GROUP.

see also GAMMA GROUP

Lambda Hypergeometric Function

$$\lambda(t) = 16q\prod_{n=1}^{\infty}\left(\frac{1+q^{2n}}{1+q^{2n-1}}\right)^8, \tag{1}$$

where q is the NOME. The lambda hypergeometric functions satisfy the recurrence relationships

$$\lambda(t+2) = \lambda(t) \tag{2}$$

$$\lambda\left(\frac{t}{2t+1}\right) = \lambda(t). \tag{3}$$

Lambert Azimuthal Equal-Area Projection

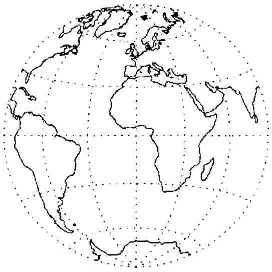

$$x = k'\cos\phi\sin(\lambda-\lambda_0) \tag{1}$$

$$y = k'[\cos\phi_1\sin\phi - \sin\phi_1\cos\phi\cos(\lambda-\lambda_0)], \tag{2}$$

where

$$k' = \sqrt{\frac{2}{1+\sin\phi_1\sin\phi+\cos\phi_1\cos\phi\cos(\lambda-\lambda_0)}}. \tag{3}$$

The inverse FORMULAS are

$$\phi = \sin^{-1}\left(\cos c\sin\phi_1 + \frac{y\sin c\cos\phi_1}{\rho}\right) \tag{4}$$

$$\lambda = \lambda_0 + \tan^{-1}\left(\frac{x\sin c}{\rho\cos\phi_1\cos c - y\sin\phi_1\sin c}\right), \tag{5}$$

where

$$\rho = \sqrt{x^2+y^2} \tag{6}$$

$$c = 2\sin^{-1}(\tfrac{1}{2}\rho). \tag{7}$$

References

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 182–190, 1987.

Lambert Conformal Conic Projection

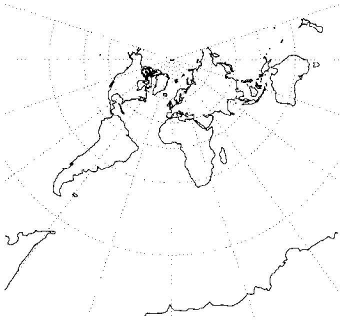

$$x = \rho \sin[n(\lambda - \lambda_0)] \tag{1}$$

$$y = \rho_0 - \rho \cos[n(\lambda - \lambda_0)], \tag{2}$$

where

$$\rho = F \cot^n(\tfrac{1}{4}\pi + \tfrac{1}{2}\phi) \tag{3}$$

$$\rho_0 = F \cot^n(\tfrac{1}{4}\pi + \tfrac{1}{2}\phi_0) \tag{4}$$

$$F = \frac{\cos\phi_1 \tan^n(\frac{1}{4}\pi + \frac{1}{2}\phi_1)}{n} \tag{5}$$

$$n = \frac{\ln(\cos\phi_1 \sec\phi_2)}{\ln[\tan(\frac{1}{4}\pi + \frac{1}{2}\phi_2)\cot(\frac{1}{4}\pi + \frac{1}{2}\phi_1)]}. \tag{6}$$

The inverse FORMULAS are

$$\phi = 2\tan^{-1}\left[\left(\frac{F}{\rho}\right)^{1/n}\right] - \tfrac{1}{2}\pi \tag{7}$$

$$\lambda = \lambda_0 + \frac{\theta}{n}, \tag{8}$$

where

$$\rho = \operatorname{sgn}(n)\sqrt{x^2 + (\rho_0 - y)^2} \tag{9}$$

$$\theta = \tan^{-1}\left(\frac{x}{\rho_0 - y}\right). \tag{10}$$

References

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 104–110, 1987.

Lambert's Method

A ROOT-finding method also called BAILEY'S METHOD and HUTTON'S METHOD. If $g(x) = x^d - r$, then

$$H_g(x) = \frac{(d-1)x^d + (d+1)r}{(d+1)x^d + (d-1)r}x.$$

References

Scavo, T. R. and Thoo, J. B. "On the Geometry of Halley's Method." *Amer. Math. Monthly* **102**, 417–426, 1995.

Lambert Series

A series of the form

$$F(x) \equiv \sum_{n=1}^{\infty} a_n \frac{x^n}{1 - x^n} \tag{1}$$

for $|x| < 1$. Then

$$F(x) = \sum_{n=1}^{\infty} a_n \sum_{m=1}^{\infty} x^{mn} = \sum_{N=1}^{\infty} b_N x^N, \tag{2}$$

where

$$b_N \equiv \sum_{n|N} a_n. \tag{3}$$

Some beautiful series of this type include

$$\sum_{n=1}^{\infty} \frac{\mu(n)x^n}{1 - x^n} = x \tag{4}$$

$$\sum_{n=1}^{\infty} \frac{\phi(n)x^n}{1 - x^n} = \frac{x}{(1-x)^2} \tag{5}$$

$$\sum_{n=1}^{\infty} \frac{x^n}{1 - x^n} = \sum_{n=1}^{\infty} d(n)x^n \tag{6}$$

$$\sum_{n=1}^{\infty} \frac{n^k x^n}{1 - x^n} = \sum_{n=1}^{\infty} \sigma_k(n)x^n \tag{7}$$

$$\sum_{n=1}^{\infty} \frac{4(-1)^{n+1}x^n}{1 - x^n} = \sum_{n=1}^{\infty} r(n)x^n, \tag{8}$$

where $\mu(n)$ is the MÖBIUS FUNCTION, $\phi(n)$ is the TOTIENT FUNCTION, $d(n) = \sigma_0(n)$ is the number of divisors of n, $\sigma_k(n)$ is the DIVISOR FUNCTION, and $r(n)$ is the number of representations of n in the form $n = A^2 + B^2$ where A and B are rational integers (Hardy and Wright 1979).

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Number Theoretic Functions." §24.3.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 826–827, 1972.

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, pp. 257–258, 1979.

Lambert's Transcendental Equation

An equation proposed by Lambert (1758) and studied by Euler in 1779 (Euler 1921).

$$x^\alpha - x^\beta = (\alpha - \beta)vx^{\alpha+\beta}.$$

When $\alpha \to \beta$, the equation becomes

$$\ln x = vx^\beta,$$

which has the solution

$$x = \exp\left[-\frac{W(-\beta v)}{\beta}\right],$$

where W is LAMBERT'S W-FUNCTION.

References

Corless, R. M.; Gonnet, G. H.; Hare, D. E. G.; and Jeffrey, D. J. "On Lambert's *W* Function." `ftp://watdragon.uwaterloo.ca/cs-archive/CS-93-03/W.ps.Z`.
de Bruijn, N. G. *Asymptotic Methods in Analysis.* Amsterdam, Netherlands: North-Holland, pp. 27–28, 1961.
Euler, L. "De Serie Lambertina Plurismique Eius Insignibus Proprietatibus." *Leonhardi Euleri Opera Omnia, Ser. 1. Opera Mathematica,* Bd. 6, 1921.
Lambert, J. H. "Observations variae in Mathesin Puram." *Acta Helvitica, physico-mathematico-anatomico-botanico-medica* **3**, 128–168, 1758.

Lambert's *W*-Function

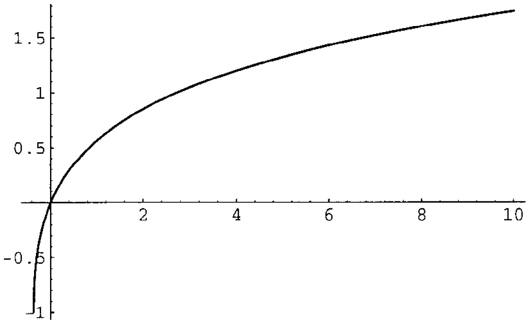

The inverse of the function

$$f(W) = We^W, \tag{1}$$

also called the OMEGA FUNCTION. The function is implemented as the *Mathematica*® (Wolfram Research, Champaign, IL) function `ProductLog[z]`. $W(1)$ is called the OMEGA CONSTANT and can be considered a sort of "GOLDEN RATIO" of exponentials since

$$\exp[-W(1)] = W(1), \tag{2}$$

giving

$$\ln\left[\frac{1}{W(1)}\right] = W(1). \tag{3}$$

Lambert's W-Function has the series expansion

$$W(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1} n^{n-2}}{(n-1)!} x^n = x - x^2 + \tfrac{3}{2}x^3 - \tfrac{8}{3}x^4 + \tfrac{125}{24}x^5 - \tfrac{54}{5}x^6 + \tfrac{16807}{720}x^7 + \ldots. \tag{4}$$

The LAGRANGE INVERSION THEOREM gives the equivalent series expansion

$$W_0(z) = \sum_{n=1}^{\infty} \frac{(-n)^{n-1}}{n!} z^n, \tag{5}$$

where $n!$ is a FACTORIAL. However, this series oscillates between ever larger POSITIVE and NEGATIVE values for REAL $z \gtrsim 0.4$, and so cannot be used for practical numerical computation. An asymptotic FORMULA which yields reasonably accurate results for $z \gtrsim 3$ is

$$\begin{aligned} W(z) &= \operatorname{Ln} z - \ln \operatorname{Ln} z \\ &\quad + \sum_{k=0}^{\infty}\sum_{m=0}^{\infty} c_{km} (\ln \operatorname{Ln} z)^{m+1} (\operatorname{Ln} z)^{-k-m-1} \\ &= L_1 - L_2 + \frac{L_2}{L_1} + \frac{L_2(-2+L_2)}{2{L_1}^2} \\ &\quad + \frac{L_2(6 - 9L_2 + 2{L_2}^2}{6{L_1}^2} \\ &\quad + \frac{L_2(-12 + 36L_2 - 22{L_2}^2 + 3{L_2}^3)}{12{L_1}^4} \\ &\quad + \frac{L_2(60 - 300L_2 + 350{L_2}^2 - 125{L_2}^3 + 12{L_2}^4)}{60{L_1}^5} \\ &\quad + \mathcal{O}\left[\left(\frac{L_2}{L_1}\right)^6\right], \end{aligned} \tag{6}$$

where

$$L_1 = \operatorname{Ln} z \tag{7}$$

$$L_2 = \ln \operatorname{Ln} z \tag{8}$$

(Corless *et al.*), correcting a typographical error in de Bruijn (1961). Another expansion due to Gosper is the DOUBLE SUM

$$W(x) = a + \sum_{n=0}^{\infty} \left\{ \sum_{k=0}^{n} \frac{S_1(n,k)}{\left[\ln\left(\frac{x}{a}\right) - a\right]^{k-1} (n-k+1)!} \right\} \times \left[1 - \frac{\ln\left(\frac{x}{a}\right)}{a}\right]^n, \tag{9}$$

where S_1 is a *nonnegative* STIRLING NUMBER OF THE FIRST KIND and a is a first approximation which can be used to select between branches. Lambert's W-function is two-valued for $-1/e \le x < 0$. For $W(x) \ge -1$, the function is denoted $W_0(x)$ or simply $W(x)$, and this is called the principal branch. For $W(x) \le -1$, the function is denoted $W_{-1}(x)$. The DERIVATIVE of W is

$$W'(x) = \frac{1}{[1 + W(x)]\exp[W(x)]} = \frac{W(x)}{x[1 + W(x)]} \tag{10}$$

for $x \neq 0$. For the principal branch when $z > 0$,

$$\ln W(z) = \ln z - W(z). \tag{11}$$

see also ITERATED EXPONENTIAL CONSTANTS, OMEGA CONSTANT

References

Corless, R. M.; Gonnet, G. H.; Hare, D. E. G.; and Jeffrey, D. J. "On Lambert's *W* Function." `ftp://watdragon.uwaterloo.ca/cs-archive/CS-93-03/W.ps.Z`.
de Bruijn, N. G. *Asymptotic Methods in Analysis.* Amsterdam, Netherlands: North-Holland, pp. 27–28, 1961.

Lamé Curve

A curve with Cartesian equation

$$\left(\frac{x}{a}\right)^n + \left(\frac{y}{b}\right)^n = c$$

first discussed in 1818 by Lamé. If n is a rational, then the curve is algebraic. However, for irrational n, the curve is transcendental. For EVEN INTEGERS n, the curve becomes closer to a rectangle as n increases. For ODD INTEGER values of n, the curve looks like the EVEN case in the POSITIVE quadrant but goes to infinity in both the second and fourth quadrants (MacTutor Archive). The EVOLUTE of an ELLIPSE,

$$(ax)^{2/3} + (by)^{2/3} = (a^2 - b^2)^{2/3}.$$

n	Curve
$\frac{2}{3}$	astroid
$\frac{5}{2}$	superellipse
3	witch of Agnesi

see also ASTROID, SUPERELLIPSE, WITCH OF AGNESI

References

MacTutor History of Mathematics Archive. "Lamé Curves." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Lame.html`.

Lamé's Differential Equation

$$(x^2 - b^2)(x^2 - c^2)\frac{d^2z}{dx^2} + x(x^2 - b^2 + x^2 - c^2)\frac{dz}{dx} - [m(m+1)x^2 - (b^2 + c^2)p]z = 0. \quad (1)$$

(Byerly 1959, p. 255). The solution is denoted $E_m^p(x)$ and is known as a LAMÉ FUNCTION or an ELLIPSOIDAL HARMONIC. Whittaker and Watson (1990, pp. 554–555) give the alternative forms

$$4\Delta_\lambda \frac{d}{d\lambda}\left[\Delta_\lambda \frac{d\Lambda}{d\lambda}\right] = [n(n+1)\lambda + C]\Lambda \quad (2)$$

$$\frac{d^2\Lambda}{d\lambda^2} + \left[\frac{\frac{1}{2}}{a^2+\lambda} + \frac{\frac{1}{2}}{b^2+\lambda} + \frac{\frac{1}{2}}{c^2}\right]\frac{d\Lambda}{d\lambda} = \frac{[n(n+1)\lambda + C]\Lambda}{4\Delta_\lambda} \quad (3)$$

$$\frac{d^2\Lambda}{du^2} = [n(n+1)\wp(u) + C - \tfrac{1}{3}n(n+1)(a^2+b^2+c^2)]\Lambda \quad (4)$$

$$\frac{d^2\Lambda}{d{z_1}^2} = [n(n+1)k^2 \operatorname{sn}^2 \alpha + A]\Lambda, \quad (5)$$

where $\wp$ is a WEIERSTRAß ELLIPTIC FUNCTION and

$$\Lambda(\theta) \equiv \prod_{q=1}^{m}(\theta - \theta_q) \quad (6)$$

$$\Delta_\lambda \equiv \sqrt{(a^2+\lambda)(b^2+\lambda)(c^2+\lambda)} \quad (7)$$

$$A \equiv \frac{C - \frac{1}{3}n(n+1)(a^2+b^2+c^2) + e_3 n(n+1)}{e_1 - e_3}. \quad (8)$$

References

Byerly, W. E. *An Elementary Treatise on Fourier's Series, and Spherical, Cylindrical, and Ellipsoidal Harmonics, with Applications to Problems in Mathematical Physics.* New York: Dover, 1959.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Lamé's Differential Equation (Types)

Whittaker and Watson (1990, pp. 539–540) write Lamé's differential equation for ELLIPSOIDAL HARMONICS of the four types as

$$4\Delta(\theta)\frac{d}{d\theta}\left[F(\theta)\frac{d\Lambda(\theta)}{d\theta}\right] = [2m(2m+1)\theta + C]\Lambda(\theta) \quad (1)$$

$$4\Delta(\theta)\frac{d}{d\theta}\left[F(\theta)\frac{d\Lambda(\theta)}{d\theta}\right] = [(2m+1)(2m+2)\theta + C]\Lambda(\theta) \quad (2)$$

$$4\Delta(\theta)\frac{d}{d\theta}\left[F(\theta)\frac{d\Lambda(\theta)}{d\theta}\right] = [(2m+2)(2m+3)\theta + C]\Lambda(\theta) \quad (3)$$

$$4\Delta(\theta)\frac{d}{d\theta}\left[F(\theta)\frac{d\Lambda(\theta)}{d\theta}\right] = [(2m+3)(2m+4)\theta + C]\Lambda(\theta), \quad (4)$$

where

$$\Delta(\theta) \equiv \sqrt{(a^2+\theta)(b^2+\theta)(c^2+\theta)} \quad (5)$$

$$\Lambda(\theta) \equiv \prod_{q=1}^{m}(\theta - \theta_q). \quad (6)$$

References

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Lamé Function

see ELLIPSOIDAL HARMONIC

Lamé's Theorem

If a is the smallest INTEGER for which there is a smaller INTEGER b such that a and b generate a EUCLIDEAN ALGORITHM remainder sequence with n steps, then a is the FIBONACCI NUMBER F_{n+2}. Furthermore, the number of steps in the EUCLIDEAN ALGORITHM never exceeds 5 times the number of digits in the smaller number.

see also EUCLIDEAN ALGORITHM

References

Honsberger, R. "A Theorem of Gabriel Lamé." Ch. 7 in *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., pp. 54–57, 1976.

Lamina

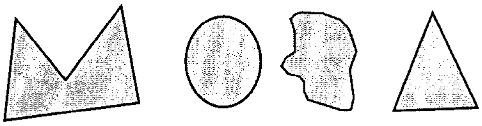

A 2-D planar closed surface L which has a mass M and a surface density $\sigma(x,y)$ (in units of mass per areas squared) such that

$$M = \int_L \sigma(x,y)\,dx\,dy.$$

The CENTER OF MASS of a lamina is called its CENTROID.

see also CENTROID (GEOMETRIC), CROSS-SECTION, SOLID

Laminated Lattice

A LATTICE which is built up of layers of n-D lattices in $(n+1)$-D space. The VECTORS specifying how layers are stacked are called GLUE VECTORS.

see also GLUE VECTOR, LATTICE

References

Conway, J. H. and Sloane, N. J. A. "Laminated Lattices." Ch. 6 in *Sphere Packings, Lattices, and Groups, 2nd ed.* New York: Springer-Verlag, pp. 157–180, 1993.

Lancret Equation

$$ds_N{}^2 = ds_T{}^2 + ds_B{}^2,$$

where N is the NORMAL VECTOR, T is the TANGENT, and B is the BINORMAL VECTOR.

Lancret's Theorem

A NECESSARY and SUFFICIENT condition for a curve to be a HELIX is that the ratio of CURVATURE to TORSION be constant.

Lanczos Approximation

see GAMMA FUNCTION

Lanczos σ Factor

Writing a FOURIER SERIES as

$$f(\theta) = \tfrac{1}{2}a_0 + \sum_{n=1}^{m} \operatorname{sinc}\left(\frac{n\pi}{2m}\right)[a_n\cos(n\theta) + b_n\sin(n\theta)],$$

where m is the last term and the $\operatorname{sinc} x$ terms are the Lanczos σ factor, removes the GIBBS PHENOMENON (Acton 1990).

see also FOURIER SERIES, GIBBS PHENOMENON, SINC FUNCTION

References

Acton, F. S. *Numerical Methods That Work, 2nd printing.* Washington, DC: Math. Assoc. Amer., p. 228, 1990.

Landau Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let F be the set of COMPLEX analytic functions f defined on an open region containing the closure of the unit disk $D = \{z : |z| < 1\}$ satisfying $f(0) = 0$ and $df/dz(0) = 1$. For each f in F, let $l(f)$ be the SUPREMUM of all numbers r such that $f(D)$ contains a disk of radius r. Then

$$L \equiv \inf\{l(f) : f \in F\}.$$

This constant is called the Landau constant, or the BLOCH-LANDAU CONSTANT. Robinson (1938, unpublished) and Rademacher (1943) derived the bounds

$$\tfrac{1}{2} < L \le \frac{\Gamma(\frac{1}{3})\Gamma(\frac{5}{6})}{\Gamma(\frac{1}{6})} = 0.5432588\ldots,$$

where $\Gamma(z)$ is the GAMMA FUNCTION, and conjectured that the second inequality is actually an equality,

$$L = \frac{\Gamma(\frac{1}{3})\Gamma(\frac{5}{6})}{\Gamma(\frac{1}{6})} = 0.5432588\ldots.$$

see also BLOCH CONSTANT

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/bloch/bloch.html`.

Rademacher, H. "On the Bloch-Landau Constant." *Amer. J. Math.* **65**, 387–390, 1943.

Landau-Kolmogorov Constants

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let $||f||$ be the SUPREMUM of $|f(x)|$, a real-valued function f defined on $(0,\infty)$. If f is twice differentiable and both f and f'' are bounded, Landau (1913) showed that

$$||f'|| \le 2||f||^{1/2}||f''||^{1/2}, \tag{1}$$

where the constant 2 is the best possible. Schoenberg (1973) extended the result to the nth derivative of f defined on $(0,\infty)$ if both f and $f^{(n)}$ are bounded,

$$||f^{(k)}|| \le C(n,k)||f||^{1-k/n}||f^{(n)}||^{k/n}. \tag{2}$$

An explicit FORMULA for $C(n,k)$ is not known, but particular cases are

$$C(3,1) = \left(\frac{243}{8}\right)^{1/3} \tag{3}$$

$$C(3,2) = 24^{1/3} \tag{4}$$

$$C(4,1) = 4.288\ldots \tag{5}$$

$$C(4,2) = 5.750\ldots \tag{6}$$

$$C(4,3) = 3.708\ldots. \tag{7}$$

Let $||f||$ be the SUPREMUM of $|f(x)|$, a real-valued function f defined on $(-\infty, \infty)$. If f is twice differentiable and both f and f'' are bounded, Hadamard (1914) showed that

$$||f'|| \leq \sqrt{2}||f||^{1/2}||f''||^{1/2}, \tag{8}$$

where the constant $\sqrt{2}$ is the best possible. Kolmogorov (1962) determined the best constants $C(n,k)$ for

$$||f^{(k)}|| \leq C(n,k)||f||^{1-k/n}||f^{(n)}||^{k/n} \tag{9}$$

in terms of the FAVARD CONSTANTS

$$a_n = \frac{4}{\pi}\sum_{j=0}^{\infty}\left[\frac{(-1)^j}{2j+1}\right]^{n+1} \tag{10}$$

by

$$C(n,k) = a_{n-k}{a_n}^{-1+k/n}. \tag{11}$$

Special cases derived by Shilov (1937) are

$$C(3,1) = \left(\frac{9}{8}\right)^{1/3} \tag{12}$$

$$C(3,2) = 3^{1/3} \tag{13}$$

$$C(4,1) = \left(\frac{512}{375}\right)^{1/4} \tag{14}$$

$$C(4,2) = \sqrt{\frac{6}{5}} \tag{15}$$

$$C(4,3) = \left(\frac{24}{5}\right)^{1/4} \tag{16}$$

$$C(5,1) = \left(\frac{1953125}{1572864}\right)^{1/5} \tag{17}$$

$$C(5,2) = \left(\frac{125}{72}\right)^{1/5}. \tag{18}$$

For a real-valued function f defined on $(-\infty, \infty)$, define

$$||f|| = \sqrt{\int_{-\infty}^{\infty}[f(x)]^2\,dx}. \tag{19}$$

If f is n differentiable and both f and $f^{(n)}$ are bounded, Hardy *et al.* (1934) showed that

$$||f^{(k)}|| \leq ||f||^{1-k/n}\,||f^{(n)}||^{k/n}, \tag{20}$$

where the constant 1 is the best possible for all n and $0 < k < n$.

For a real-valued function f defined on $(0, \infty)$, define

$$||f|| = \sqrt{\int_0^{\infty}[f(x)]^2\,dx}. \tag{21}$$

If f is twice differentiable and both f and f'' are bounded, Hardy *et al.* (1934) showed that

$$||f'|| \leq \sqrt{2}\,||f||^{1/2}||f^{(n)}||^{1/2}, \tag{22}$$

where the constant $\sqrt{2}$ is the best possible. This inequality was extended by Ljubic (1964) and Kupcov (1975) to

$$||f^{(k)}|| \leq C(n,k)\,||f||^{1-k/n}||f^{(n)}||^{k/n} \tag{23}$$

where $C(n,k)$ are given in terms of zeros of POLYNOMIALS. Special cases are

$$\begin{aligned} C(3,1) = C(3,2) &= 3^{1/2}[2(2^{1/2}-1)]^{-1/3} \\ &= 1.84420\ldots \end{aligned} \tag{24}$$

$$\begin{aligned} C(4,1) = C(4,3) &= \sqrt{\frac{3^{1/4}+3^{-3/4}}{a}} \\ &= 2.27432\ldots \end{aligned} \tag{25}$$

$$C(4,2) = \sqrt{\frac{2}{b}} = 2.97963\ldots \tag{26}$$

$$C(4,3) = \left(\frac{24}{5}\right)^{1/4} \tag{27}$$

$$C(5,1) = C(5,4) = 2.70247\ldots \tag{28}$$

$$C(5,2) = C(5,3) = 4.37800\ldots, \tag{29}$$

where a is the least POSITIVE ROOT of

$$x^8 - 6x^4 - 8x^2 + 1 = 0 \tag{30}$$

and b is the least POSITIVE ROOT of

$$x^4 - 2x^2 - 4x + 1 = 0 \tag{31}$$

(Franco *et al.* 1985, Neta 1980). The constants $C(n,1)$ are given by

$$C(n,1) = \sqrt{\frac{(n-1)^{1/n}+(n+1)^{-1+1/n}}{c}}, \tag{32}$$

where c is the least POSITIVE ROOT of

$$\int_0^c\int_0^{\infty}\frac{dx\,dy}{(x^{2n}-yx^2+1)\sqrt{y}} = \frac{\pi^2}{2n}. \tag{33}$$

An explicit FORMULA of this type is not known for $k > 1$.

The cases $p = 1,\ 2,\ \infty$ are the only ones for which the best constants have exact expressions (Kwong and Zettl 1992, Franco *et al.* 1983).

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/lk/lk.html`.

Franco, Z. M.; Kaper, H. G.; Kwong, M. N.; and Zettl, A. "Bounds for the Best Constants in Landau's Inequality on the Line." *Proc. Roy. Soc. Edinburgh* **95A**, 257–262, 1983.

Franco, Z. M.; Kaper, H. G.; Kwong, M. N.; and Zettl, A. "Best Constants in Norm Inequalities for Derivatives on a Half Line." *Proc. Roy. Soc. Edinburgh* **100A**, 67–84, 1985.

Hardy, G. H.; Littlewood, J. E.; and Pólya, G. *Inequalities.* Cambridge, England: Cambridge University Press, 1934.

Kolmogorov, A. "On Inequalities Between the Upper Bounds of the Successive Derivatives of an Arbitrary Function on an Infinite Integral." *Amer. Math. Soc. Translations, Ser. 1* **2**, 233–243, 1962.

Kupcov, N. P. "Kolmogorov Estimates for Derivatives in $L_2(0,\infty)$." *Proc. Steklov Inst. Math.* **138**, 101–125, 1975.

Kwong, M. K. and Zettl, A. *Norm Inequalities for Derivatives and Differences.* New York: Springer-Verlag, 1992.

Landau, E. "Einige Ungleichungen für zweimal differentzierbare Funktionen." *Proc. London Math. Soc. Ser. 2* **13**, 43–49, 1913.

Landau, E. "Die Ungleichungen für zweimal differentzierbare Funktionen." *Danske Vid. Selsk. Math. Fys. Medd.* **6**, 1–49, 1925.

Ljubic, J. I. "On Inequalities Between the Powers of a Linear Operator." *Amer. Math. Soc. Trans. Ser. 2* **40**, 39–84, 1964.

Neta, B. "On Determinations of Best Possible Constants in Integral Inequalities Involving Derivatives." *Math. Comput.* **35**, 1191–1193, 1980.

Schoenberg, I. J. "The Elementary Case of Landau's Problem of Inequalities Between Derivatives." *Amer. Math. Monthly* **80**, 121–158, 1973.

Landau-Ramanujan Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let $S(x)$ denote the number of POSITIVE INTEGERS not exceeding x which can be expressed as a sum of two squares, then

$$\lim_{x\to\infty}\frac{\sqrt{\ln x}}{x}S(x)=K, \tag{1}$$

as proved by Landau (1908) and stated by Ramanujan. The value of K (also sometimes called λ) is

$$K=\sqrt{\tfrac{1}{2}\prod_{\substack{p\text{ a prime}\\ \equiv 3\ (\mathrm{mod}\ 4)}}\frac{1}{1-p^{-2}}}=0.764223653\ldots \tag{2}$$

(Hardy 1940, Berndt 1994). Ramanujan found the approximate value $K=0.764$. Flajolet and Vardi (1996) give a beautiful FORMULA with fast convergence

$$K=\frac{1}{\sqrt{2}}\prod_{n=1}^{\infty}\left[\left(1-\frac{1}{2^{2^n}}\right)\frac{\zeta(2^n)}{\beta(2^n)}\right]^{1/(2^n+1)}, \tag{3}$$

where

$$\beta(s)\equiv\frac{1}{4^s}[\zeta(s,\tfrac{1}{4})-\zeta(s,\tfrac{3}{4})] \tag{4}$$

is the DIRICHLET BETA FUNCTION, and $\zeta(z,a)$ is the HURWITZ ZETA FUNCTION. Landau proved the even stronger fact

$$\lim_{x\to\infty}\frac{(\ln x)^{3/2}}{Kx}\left[S(x)-\frac{Kx}{\sqrt{\ln x}}\right]=C, \tag{5}$$

where

$$C\equiv\frac{1}{2}\left[1-\ln\left(\frac{\pi e^{\gamma}}{L}\right)\right]-\frac{1}{4}\frac{d}{ds}\left[\ln\left(\prod_{\substack{p\text{ prime}\\ p=4k+3}}\frac{1}{p^{-2s}}\right)\right]_{s=1}$$
$$=0.581948659\ldots. \tag{6}$$

Here,

$$L=5.2441151086\ldots \tag{7}$$

is the ARC LENGTH of a LEMNISCATE with $a=1$ (the LEMNISCATE CONSTANT to within a factor of 2 or 4), and γ is the EULER-MASCHERONI CONSTANT.

References

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 60–66, 1994.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/lr/lr.html`.

Flajolet, P. and Vardi, I. "Zeta Function Expansions of Classical Constants." Unpublished manuscript. 1996. `http://pauillac.inria.fr/algo/flajolet/Publications/landau.ps`.

Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, pp. 61–63, 1940.

Landau, E. "Über die Einteilung der positiven ganzen Zahlen in vier Klassen nach der Mindeszahl der zu ihrer additiven Zusammensetzung erforderlichen Quadrate." *Arch. Math. Phys.* **13**, 305–312, 1908.

Shanks, D. "The Second-Order Term in the Asymptotic Expansion of $B(x)$." *Math. Comput.* **18**, 75–86, 1964.

Shanks, D. "Non-Hypotenuse Numbers." *Fibonacci Quart.* **13**, 319–321, 1975.

Shanks, D. and Schmid, L. P. "Variations on a Theorem of Landau. I." *Math. Comput.* **20**, 551–569, 1966.

Shiu, P. "Counting Sums of Two Squares: The Meissel-Lehmer Method." *Math. Comput.* **47**, 351–360, 1986.

Landau Symbol

Let $f(z)$ be a function $\neq 0$ in an interval containing $z=0$. Let $g(z)$ be another function also defined in this interval such that $g(z)/f(z)\to 0$ as $z\to 0$. Then $g(z)$ is said to be $\mathcal{O}(f(z))$.

Landen's Formula

$$\frac{\vartheta_3(z,t)\vartheta_4(z,t)}{\vartheta_4(2z,2t)}=\frac{\vartheta_3(0,t)\vartheta_4(0,t)}{\vartheta_4(0,2t)}=\frac{\vartheta_2(z,t)\vartheta_1(z,t)}{\vartheta_1(2z,2t)},$$

where ϑ_i are THETA FUNCTIONS. This transformation was used by Gauss to show that ELLIPTIC INTEGRALS could be computed using the ARITHMETIC-GEOMETRIC MEAN.

Landen's Transformation

If $x\sin\alpha = \sin(2\beta-\alpha)$, then

$$(1+x)\int_0^\alpha \frac{d\phi}{\sqrt{1-x^2\sin^2\phi}} = 2\int_0^\beta \frac{d\phi}{\sqrt{1-\frac{4x}{(1+x)^2}\sin^2\phi}}.$$

see also ELLIPTIC INTEGRAL OF THE FIRST KIND, GAUSS'S TRANSFORMATION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Ascending Landen Transformation" and "Landen's Transformation." §16.14 and 17.5 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 573–574 and 597–598, 1972.

Lane-Emden Differential Equation

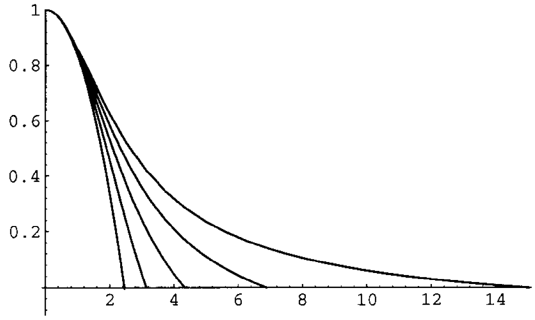

A second-order ORDINARY DIFFERENTIAL EQUATION arising in the study of stellar interiors. It is given by

$$\frac{1}{\xi^2}\frac{d}{d\xi}\left(\xi^2\frac{d\theta}{d\xi}\right)+\theta^n=0 \tag{1}$$

$$\frac{1}{\xi^2}\left(2\xi\frac{d\theta}{d\xi}+\xi^2\frac{d^2\theta}{d\xi^2}\right)+\theta^n=\frac{d^2\theta}{d\xi^2}+\frac{2}{\xi}\frac{d\theta}{d\xi}+\theta^n=0. \tag{2}$$

It has the BOUNDARY CONDITIONS

$$\theta(0)=1 \tag{3}$$

$$\left[\frac{d\theta}{d\xi}\right]_{\xi=0}=0. \tag{4}$$

Solutions $\theta(\xi)$ for $n = 0$, 1, 2, 3, and 4 are shown above. The cases $n = 0$, 1, and 5 can be solved analytically (Chandrasekhar 1967, p. 91); the others must be obtained numerically.

For $n = 0$ ($\gamma = \infty$), the LANE-EMDEN DIFFERENTIAL EQUATION is

$$\frac{1}{\xi^2}\frac{d}{d\xi}\left(\xi^2\frac{d\theta}{d\xi}\right)+1=0 \tag{5}$$

(Chandrasekhar 1967, pp. 91–92). Directly solving gives

$$\frac{d}{d\xi}\left(\xi^2\frac{d\theta}{d\xi}\right)=-\xi^2 \tag{6}$$

$$\int d\left(\xi^2\frac{d\theta}{d\xi^2}\right)=-\int\xi^2\,d\xi \tag{7}$$

$$\xi^2\frac{d\theta}{d\xi}=c_1-\tfrac{1}{3}\xi^3 \tag{8}$$

$$\frac{d\theta}{d\xi}=\frac{c_1-\frac{1}{3}\xi^3}{\xi^2} \tag{9}$$

$$\theta(\xi)=\int d\theta=\int\frac{c_1-\frac{1}{3}\xi^3}{\xi^2}\,d\xi \tag{10}$$

$$\theta(\xi)=\theta_0-c_1\xi^{-1}-\tfrac{1}{6}\xi^2. \tag{11}$$

The BOUNDARY CONDITION $\theta(0) = 1$ then gives $\theta_0 = 1$ and $c_1 = 0$, so

$$\theta_1(\xi)=1-\tfrac{1}{6}\xi^2, \tag{12}$$

and $\theta_1(\xi)$ is PARABOLIC.

For $n = 1$ ($\gamma = 2$), the differential equation becomes

$$\frac{1}{\xi^2}\frac{d}{d\xi}\left(\xi^2\frac{d\theta}{d\xi}\right)+\theta=0 \tag{13}$$

$$\frac{d}{d\xi}\left(\xi^2\frac{d\theta}{d\xi}\right)+\theta\xi^2=0, \tag{14}$$

which is the SPHERICAL BESSEL DIFFERENTIAL EQUATION

$$\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)+[k^2r^2-n(n+1)]R=0 \tag{15}$$

with $k = 1$ and $n = 0$, so the solution is

$$\theta(\xi)=Aj_0(\xi)+Bn_0(\xi). \tag{16}$$

Applying the BOUNDARY CONDITION $\theta(0) = 1$ gives

$$\theta_2(\xi)=j_0(\xi)=\frac{\sin\xi}{\xi}, \tag{17}$$

where $j_0(x)$ is a SPHERICAL BESSEL FUNCTION OF THE FIRST KIND (Chandrasekhar 1967, pp. 92).

For $n = 5$, make Emden's transformation

$$\theta=Ax^\omega z \tag{18}$$

$$\omega=\frac{2}{n-1}, \tag{19}$$

which reduces the Lane-Emden equation to

$$\frac{d^2z}{dt^2}+(2\omega-1)\frac{dz}{dt}+\omega(\omega-1)z+A^{n-1}z^n=0 \tag{20}$$

(Chandrasekhar 1967, p. 90). After further manipulation (not reproduced here), the equation becomes

$$\frac{d^2z}{dt^2}=\tfrac{1}{4}z(1-z^4) \tag{21}$$

and then, finally,

$$\theta_3(\xi)(1+\tfrac{1}{3}\xi^2)^{-1/2}. \tag{22}$$

References

Chandrasekhar, S. *An Introduction to the Study of Stellar Structure.* New York: Dover, pp. 84–182, 1967.

Langford's Problem

Arrange copies of the n digits 1, ..., n such that there is one digit between the 1s, two digits between the 2s, etc. For example, the $n = 3$ solution is 312132 and the $n = 4$ solution is 41312432. Solutions exist only if $n \equiv 0, 3 \pmod 4$. The number of solutions for $n = 3$, 4, 5, ... are 1, 1, 0, 0, 26, 150, 0, 0, 17792, 108144, ... (Sloane's A014552).

References

Gardner, M. *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, pp. 70 and 77–78, 1978.

Sloane, N. J. A. Sequence A014551 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Langlands Program

A grand unified theory of mathematics which includes the search for a generalization of ARTIN RECIPROCITY (known as LANGLANDS RECIPROCITY) to non-Abelian Galois extensions of NUMBER FIELDS. Langlands proposed in 1970 that the mathematics of algebra and analysis are intimately related. He was a co-recipient of the 1996 Wolf Prize for this formulation.

see also ARTIN RECIPROCITY, LANGLANDS RECIPROCITY

References

American Mathematical Society. "Langlands and Wiles Share Wolf Prize." *Not. Amer. Math. Soc.* **43**, 221–222, 1996.

Knapp, A. W. "Group Representations and Harmonic Analysis from Euler to Langlands." *Not. Amer. Math. Soc.* **43**, 410–415, 1996.

Langlands Reciprocity

The conjecture that the ARTIN L-FUNCTION of any n-D GALOIS GROUP representation is an L-FUNCTION obtained from the GENERAL LINEAR GROUP $GL_1(\mathbb{A})$.

References

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Langton's Ant

A CELLULAR AUTOMATON. The COHEN-KUNG THEOREM guarantees that the ant's trajectory is unbounded.

see also CELLULAR AUTOMATON, COHEN-KUNG THEOREM

References

Stewart, I. "The Ultimate in Anty-Particles." *Sci. Amer.* **271**, 104–107, 1994.

Laplace-Beltrami Operator

A self-adjoint elliptic differential operator defined somewhat technically as

$$\Delta = d\delta + \delta d,$$

where d is the EXTERIOR DERIVATIVE and d and δ are adjoint to each other with respect to the INNER PRODUCT.

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 628, 1980.

Laplace Distribution

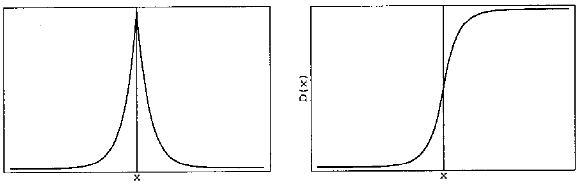

Also called the DOUBLE EXPONENTIAL DISTRIBUTION. It is the distribution of differences between two independent variates with identical EXPONENTIAL DISTRIBUTIONS (Abramowitz and Stegun 1972, p. 930).

$$P(x) = \frac{1}{2b} e^{-|x-\mu|/b} \tag{1}$$

$$D(x) = \tfrac{1}{2}[1 + \operatorname{sgn}(x-\mu)(1 - e^{-|x-\mu|/b})]. \tag{2}$$

The MOMENTS about the MEAN μ_n are related to the MOMENTS about 0 by

$$\mu_n = \sum_{j=0}^{n} \binom{n}{j} (-1)^{n-j} \mu'_j \mu^{n-j}, \tag{3}$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT, so

$$\begin{aligned}\mu_n &= \sum_{j=0}^{n} \sum_{k=0}^{\lfloor j/2 \rfloor} (-1)^{n-j} \binom{n}{j} \binom{j}{2k} b^{2k} \mu^{n-2k} \Gamma(2k+1) \\ &= \begin{cases} n!b^n & \text{for } n \text{ even} \\ 0 & \text{for } n \text{ odd,} \end{cases}\end{aligned} \tag{4}$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION and $\Gamma(2k+1)$ is the GAMMA FUNCTION.

The MOMENTS can also be computed using the CHARACTERISTIC FUNCTION,

$$\phi(t) \equiv \int_{-\infty}^{\infty} e^{itx} P(x)\, dx = \frac{1}{2b} \int_{-\infty}^{\infty} e^{itx} e^{-|x-\mu|/b}\, dx. \tag{5}$$

Using the FOURIER TRANSFORM OF THE EXPONENTIAL FUNCTION

$$\mathcal{F}[e^{-2\pi k_0 |x|}] = \frac{1}{\pi} \frac{k_0}{k^2 + {k_0}^2} \tag{6}$$

gives

$$\phi(t) = \frac{e^{i\mu t}}{2b} \frac{\frac{2}{b}}{t^2 + \left(\frac{1}{b}\right)^2} = \frac{e^{i\mu t}}{1 + b^2 t^2}. \tag{7}$$

The MOMENTS are therefore

$$\mu_n = (-i)^n \phi(0) = (-i)^n \left[\frac{d^n\phi}{dt^n}\right]_{t=0}. \qquad (8)$$

The MEAN, VARIANCE, SKEWNESS, and KURTOSIS are

$$\mu = \mu \qquad (9)$$

$$\sigma^2 = 2b^2 \qquad (10)$$

$$\gamma_1 = 0 \qquad (11)$$

$$\gamma_2 = 3. \qquad (12)$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, 1972.

Laplace's Equation

The scalar form of Laplace's equation is the PARTIAL DIFFERENTIAL EQUATION

$$\nabla^2\psi = 0. \qquad (1)$$

It is a special case of the HELMHOLTZ DIFFERENTIAL EQUATION

$$\nabla^2\psi + k^2\psi = 0 \qquad (2)$$

with $k = 0$, or POISSON'S EQUATION

$$\nabla^2\psi = -4\pi\rho \qquad (3)$$

with $\rho = 0$. The vector Laplace's equation is given by

$$\nabla^2\mathbf{F} = \mathbf{0}. \qquad (4)$$

A FUNCTION ψ which satisfies Laplace's equation is said to be HARMONIC. A solution to Laplace's equation has the property that the average value over a spherical surface is equal to the value at the center of the SPHERE (GAUSS'S HARMONIC FUNCTION THEOREM). Solutions have no local maxima or minima. Because Laplace's equation is linear, the superposition of any two solutions is also a solution.

A solution to Laplace's equation is uniquely determined if (1) the value of the function is specified on all boundaries (DIRICHLET BOUNDARY CONDITIONS) or (2) the normal derivative of the function is specified on all boundaries (NEUMANN BOUNDARY CONDITIONS).

Laplace's equation can be solved by SEPARATION OF VARIABLES in all 11 coordinate systems that the HELMHOLTZ DIFFERENTIAL EQUATION can. In addition, separation can be achieved by introducing a multiplicative factor in two additional coordinate systems. The separated form is

$$\psi = \frac{X_1(u_1)X_2(u_2)X_3(u_3)}{R(u_1,u_2,u_3)}, \qquad (5)$$

and setting

$$\frac{h_1h_2h_3}{{h_i}^2} = g_i(u_{i+1},u_{i+2})f_i(u_i)R^2, \qquad (6)$$

where h_i are SCALE FACTORS, gives the Laplace's equation

$$\sum_{i=1}^{3} \frac{1}{{h_i}^2X_i}\left[\frac{1}{f_i}\frac{d}{du_i}\left(f_i\frac{dX_i}{du_i}\right)\right] = \sum_{i=1}^{3} \frac{1}{{h_i}^2R}\left[\frac{1}{f_i}\frac{\partial}{\partial u_i}\left(f_i\frac{\partial R}{\partial u_i}\right)\right]. \qquad (7)$$

If the right side is equal to $-{k_1}^2/F(u_1,u_2,u_3)$, where k_1 is a constant and F is any function, and if

$$h_1h_2h_3 = Sf_1f_2f_3R^2F, \qquad (8)$$

where S is the STÄCKEL DETERMINANT, then the equation can be solved using the methods of the HELMHOLTZ DIFFERENTIAL EQUATION. The two systems where this is the case are BISPHERICAL and TOROIDAL, bringing the total number of separable systems for Laplace's equation to 13 (Morse and Feshbach 1953, pp. 665–666).

In 2-D BIPOLAR COORDINATES, Laplace's equation is separable, although the HELMHOLTZ DIFFERENTIAL EQUATION is not.

see also BOUNDARY CONDITIONS, HARMONIC EQUATION, HELMHOLTZ DIFFERENTIAL EQUATION, PARTIAL DIFFERENTIAL EQUATION, POISSON'S EQUATION, SEPARATION OF VARIABLES, STÄCKEL DETERMINANT

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 17, 1972.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 125–126, 1953.

Laplace's Equation—Bipolar Coordinates

In 2-D BIPOLAR COORDINATES, LAPLACE'S EQUATION is

$$\frac{(\cosh v - \cos u)^2}{a^2}\left(\frac{\partial F^2}{\partial u^2} + \frac{\partial F^2}{\partial v^2}\right) = 0, \qquad (1)$$

which simplifies to

$$\frac{\partial F^2}{\partial u^2} + \frac{\partial F^2}{\partial v^2} = 0, \qquad (2)$$

so LAPLACE'S EQUATION is separable, although the HELMHOLTZ DIFFERENTIAL EQUATION is not.

Laplace's Equation—Bispherical Coordinates

$$\left[\frac{-\cos u\cot^2 u+3\cosh v\cot^2 u}{\cosh v-\cos u}\right.$$
$$\left.-\frac{3\cosh^2 v\cot u\csc u+\cosh^3 v\csc^2 u}{\cosh v-\cos u}\right]\frac{\partial}{\partial\phi^2}$$
$$+(\cos u-\cosh v)\sinh v\frac{\partial}{\partial v}+(\cosh^2 v-\cos u)^2\frac{\partial^2}{\partial v^2}$$
$$+(\cosh v-\cos u)(\cosh v\cot u-\sin u-\cos u\cot u)\frac{\partial}{\partial u}$$
$$+(\cosh^2 v-\cos u)^2\frac{\partial^2}{\partial u^2}=0. \tag{1}$$

Let

$$F(u,v,\phi)=\sqrt{\cosh u-\cos v}\,U(u)V(v)\Phi(\phi), \tag{2}$$

then LAPLACE'S EQUATION is partially separable, although the HELMHOLTZ DIFFERENTIAL EQUATION is not.

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 665–666, 1953.

Laplace's Equation—Toroidal Coordinates

$$\nabla^2 f=\frac{(\cosh v-\cos u)^3}{a^2}\frac{\partial}{\partial u}\left(\frac{1}{\cosh v-\cos u}\frac{\partial f}{\partial u}\right)$$
$$+\frac{(\cosh v-\cos u)^3}{a^2\sinh v}\frac{\partial}{\partial v}\left(\frac{\sinh v}{\cosh v-\cos u}\frac{\partial f}{\partial v}\right)$$
$$+\frac{(\cosh v-\cos u)^2}{a^2\sinh v}\frac{\partial^2 f}{\partial\phi^2} \tag{1}$$
$$=\left[\frac{-3\cos\coth^2 v+\cosh v\coth^2 v}{\cosh v-\cos u}\right.$$
$$\left.+\frac{3\cos^2 u\coth v\operatorname{csch} v-\cos^3 u\operatorname{csch}^2 v}{\cosh v-\cos u}\right]\frac{\partial^2}{\partial\phi^2}$$
$$+(\cos u-\cosh v)\sin u\frac{\partial}{\partial u}+(\cosh v-\cos u)^2\frac{\partial^2}{\partial u^2}$$
$$+(\cosh v-\cos u)(\cosh v\coth v-\sinh v-\cos u\coth v)\frac{\partial}{\partial v}$$
$$+(\cosh^2 v-\cos u)^2\frac{\partial^2}{\partial v^2}. \tag{2}$$

Let

$$f(\xi,\eta,\phi)=\sqrt{\cosh\eta-\cos\xi}\,X(\xi)H(\eta)\Psi(\psi), \tag{3}$$

then

$$X(\xi)=\begin{matrix}\sin\\ \cos\end{matrix}(n\xi) \tag{4}$$

$$\Psi(\psi)=\begin{matrix}\sin\\ \cos\end{matrix}(m\phi), \tag{5}$$

and the equation in η becomes

$$\frac{1}{\sinh\eta}\frac{d}{d\eta}\left(\sinh\eta\frac{dH}{d\eta}\right)-\frac{m^2}{\sinh^2\eta}H-(n^2-\tfrac{1}{4})H=0. \tag{6}$$

LAPLACE'S EQUATION is partially separable, although the HELMHOLTZ DIFFERENTIAL EQUATION is not.

References

Arfken, G. "Toroidal Coordinates (ξ,η,ϕ)." §2.13 in *Mathematical Methods for Physicists, 2nd ed.* Orlando, FL: Academic Press, pp. 112–114, 1970.

Byerly, W. E. *An Elementary Treatise on Fourier's Series, and Spherical, Cylindrical, and Ellipsoidal Harmonics, with Applications to Problems in Mathematical Physics.* New York: Dover, p. 264, 1959.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 666, 1953.

Laplace's Integral

$$P_n(x)=\frac{1}{\pi}\int_0^p i\frac{du}{\left(x+\sqrt{x^2-1}\cos u\right)^{n+1}}$$
$$=\frac{1}{\pi}\int_0^\pi(x+\sqrt{x^2-1}\cos u)^n\,du.$$

Laplace Limit

The value $e=0.6627434193\ldots$ (Sloane's A033259) for which Laplace's formula for solving KEPLER'S EQUATION begins diverging. The constant is defined as the value e at which the function

$$f(x)=\frac{x\exp(\sqrt{1+x^2})}{1+\sqrt{1+x^2}}$$

equals $f(\lambda)=1$. The CONTINUED FRACTION of e is given by [0, 1, 1, 1, 27, 1, 1, 1, 8, 2, 154, ...] (Sloane's A033260). The positions of the first occurrences of n in the CONTINUED FRACTION of e are 2, 10, 35, 13, 15, 32, 101, 9, ... (Sloane's A033261). The incrementally largest terms in the CONTINUED FRACTION are 1, 27, 154, 1601, 2135, ... (Sloane's A033262), which occur at positions 2, 5, 11, 19, 1801, ... (Sloane's A033263).

see also ECCENTRIC ANOMALY, KEPLER'S EQUATION

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/lpc/lpc.html`.

Plouffe, S. "Laplace Limit Constant." `http://lacim.uqam.ca/piDATA/laplace.txt`.

Sloane, N. J. A. Sequences A033259, A033260, A033261, A033262, and A033263 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Laplace-Mehler Integral

$$\begin{aligned} P_n(\cos\theta) &= \frac{1}{\pi}\int_0^{2\pi} (\cos\theta + i\sin\theta\cos\phi)^n \, d\phi \\ &= \frac{\sqrt{2}}{\pi}\int_0^{\theta} \frac{\cos[(n+\frac{1}{2})\phi]}{\sqrt{\cos\phi - \cos\theta}} \, d\phi \\ &= \frac{\sqrt{2}}{\pi}\int_\theta^{\pi} \frac{\sin[(n+\frac{1}{2})\phi]}{\sqrt{\cos\theta - \cos\phi}} \, d\phi. \end{aligned}$$

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1463, 1980.

Laplace Series

A function $f(\theta, \phi)$ expressed as a double sum of SPHERICAL HARMONICS is called a Laplace series. Taking f as a COMPLEX FUNCTION,

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-1}^{l} a_{lm} Y_l^m(\theta, \phi). \tag{1}$$

Now multiply both sides by $Y_{l'}^{m'*} \sin\theta$ and integrate over $d\theta$ and $d\phi$.

$$\begin{aligned} &\int_0^{2\pi}\int_0^{\pi} f(\theta,\phi) Y_{l'}^{m'*} \sin\theta \, d\theta \, d\phi \\ &= \sum_{l=0}^{\infty}\sum_{m=-l}^{l} a_{lm} \int_0^{2\pi}\int_0^{\pi} Y_{l'}^{m'*}(\theta,\phi) Y_l^m(\theta,\phi) \sin\theta \, d\theta \, d\phi. \end{aligned} \tag{2}$$

Now use the ORTHOGONALITY of the SPHERICAL HARMONICS

$$\int_0^{2\pi}\int_0^{\pi} Y_l^m(\theta,\phi) Y_{l'}^{m'*} \sin\theta \, d\theta \, d\phi = \delta_{mm'}\delta_{ll'}, \tag{3}$$

so (2) becomes

$$\begin{aligned} &\int_0^{2\pi}\int_0^{\pi} f(\theta,\phi) Y_{l'}^{m'*} \sin\theta \, d\theta \, d\phi \\ &\qquad = \sum_{l=0}^{\infty}\sum_{m=-1}^{l} a_{lm}\delta_{mm'}\delta_{ll'} = a_{lm}, \end{aligned} \tag{4}$$

where δ_{mn} is the KRONECKER DELTA.

For a REAL series, consider

$$\begin{aligned} &f(\theta,\phi) \\ &= \sum_{l=0}^{\infty}\sum_{m=-1}^{l} [C_l^m \cos(m\phi) + S_l^m \sin(m\phi)] P_l^m(\cos\theta). \end{aligned} \tag{5}$$

Proceed as before, using the orthogonality relationships

$$\begin{aligned} &\int_0^{2\pi}\int_0^{\pi} P_l^m(\cos\theta)\cos(m\phi) P_{l'}^{m'}(\cos\theta) \\ &\quad \times \cos(m'\phi)\sin(\theta) \, d\theta \, d\phi = -\frac{2\pi(l+m)!}{(2l+1)(l-m)!}\delta_{mm'}\delta_{ll'} \end{aligned} \tag{6}$$

$$\begin{aligned} &\int_0^{2\pi}\int_0^{\pi} P_l^m(\cos\theta)\sin(m\phi) P_{l'}^{m'}(\cos\theta) \\ &\quad \times \sin(m'\phi)\sin\theta \, d\theta \, d\phi = -\frac{2\pi(l+m)!}{(2l+1)(l-m)!}\delta_{mm'}\delta_{ll'}. \end{aligned} \tag{7}$$

So C_l^m and S_l^m are given by

$$\begin{aligned} C_l^m &= -\frac{(2l+1)(l-m)!}{2\pi(l+m)!} \\ &\quad \times \int_0^{2\pi}\int_0^{\pi} f(\theta,\phi) P_l^m \cos\theta \cos(m\phi) \sin\theta \, d\theta \, d\phi \end{aligned} \tag{8}$$

$$\begin{aligned} S_l^m &= -\frac{(2l+1)(l-m)!}{2\pi(l+m)!} \\ &\quad \times \int_0^{2\pi}\int_0^{\pi} f(\theta,\phi) P_l^m \cos\theta \sin(m\phi) \sin\theta \, d\theta \, d\phi. \end{aligned} \tag{9}$$

Laplace-Stieltjes Transform

An integral transform which is often written as an ordinary LAPLACE TRANSFORM involving the DELTA FUNCTION.

see also LAPLACE TRANSFORM

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 1029, 1972.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, 1953.

Widder, D. V. *The Laplace Transform.* Princeton, NJ: Princeton University Press, 1941.

Laplace Transform

The Laplace transform is an INTEGRAL TRANSFORM perhaps second only to the FOURIER TRANSFORM in its utility in solving physical problems. Due to its useful properties, the Laplace transform is particularly useful in solving linear ORDINARY DIFFERENTIAL EQUATIONS such as those arising in the analysis of electronic circuits.

The (one-sided) Laplace transform $\mathcal{L}$ (not to be confused with the LIE DERIVATIVE) is defined by

$$\mathcal{L}(s) = \mathcal{L}(f(t)) \equiv \int_0^{\infty} f(t)e^{-st} \, dt, \tag{1}$$

where $f(t)$ is defined for $t \geq 0$. A two-sided Laplace transform is sometimes also defined by

$$\mathcal{L}(s) = \mathcal{L}(f(t)) = \int_{-\infty}^{\infty} f(t) e^{-st}\, dt. \qquad (2)$$

The Laplace transform existence theorem states that, if $f(t)$ is piecewise Continuous on every finite interval in $[0, \infty)$ satisfying

$$|f(t)| \leq M e^{at} \qquad (3)$$

for all $t \in [0, \infty)$, then $\mathcal{L}(f(t))$ exists for all $s > a$. The Laplace transform is also Unique, in the sense that, given two functions $F_1(t)$ and $F_2(t)$ with the same transform so that

$$\mathcal{L}[F_1(t)] = \mathcal{L}[F_2(t)] \equiv f(s), \qquad (4)$$

then Lerch's Theorem guarantees that the integral

$$\int_0^a N(t)\, dt = 0 \qquad (5)$$

vanishes for all $a > 0$ for a Null Function defined by

$$N(t) \equiv F_1(t) - F_2(t). \qquad (6)$$

The Laplace transform is Linear since

$$\begin{aligned}\mathcal{L}[af(t) + bg(t)] &= \int_0^{\infty} [af(t) + bg(t)] e^{-st}\, dt \\ &= a \int_0^{\infty} f(t) e^{-st}\, dt + b \int_0^{\infty} g(t) e^{-st}\, dt \\ &= a\mathcal{L}[f(t)] + b\mathcal{L}[g(t)]. \qquad (7)\end{aligned}$$

The inverse Laplace transform is given by the Bromwich Integral (see also Duhamel's Convolution Principle). A table of several important Laplace transforms follows.

$f(t)$	$\mathcal{L}[f(t)]$	Range
1	$\frac{1}{s}$	$s > 0$
t	$\frac{1}{s^2}$	$s > 0$
t^n	$\frac{n!}{s^{n+1}}$	$n \in \mathbb{Z} > 0$
t^a	$\frac{\Gamma(a+1)}{s^{a+1}}$	$a > 0$
e^{at}	$\frac{1}{s-a}$	$s > a$
$\cos(\omega t)$	$\frac{s}{s^2+\omega^2}$	$s > 0$
$\sin(\omega t)$	$\frac{\omega}{s^2+\omega^2}$	$s > 0$
$\cosh(\omega t)$	$\frac{s}{s^2-\omega^2}$	$s > \|a\|$
$\sinh(\omega t)$	$\frac{\omega}{s^2-\omega^2}$	$s > \|a\|$
$e^{at}\sin(bt)$	$\frac{b}{(s-a)^2+b^2}$	$s > a$
$e^{at}\cos(bt)$	$\frac{s-a}{(s-a)^2+b^2}$	$s > a$
$\delta(t-c)$	e^{-cs}	
$H_c(t)$	$\frac{e^{-cs}}{s}$	$s > 0$
$J_0(t)$	$\frac{1}{\sqrt{s^2+1}}$	

In the above table, $J_0(t)$ is the zeroth order Bessel Function of the First Kind, $\delta(t)$ is the Delta Function, and $H_c(t)$ is the Heaviside Step Function. The Laplace transform has many important properties.

The Laplace transform of a Convolution is given by

$$\mathcal{L}(f(t) * g(t)) = \mathcal{L}(f(t))\mathcal{L}(g(t)) \qquad (8)$$

$$\mathcal{L}^{-1}(F(s)G(s)) = \mathcal{L}^{-1}(F(s)) * \mathcal{L}^{-1}(G(s)). \qquad (9)$$

Now consider Differentiation. Let $f(t)$ be continuously differentiable $n-1$ times in $[0, \infty)$. If $|f(t)| \leq Me^{at}$, then

$$\begin{aligned}\mathcal{L}[f^{(n)}(t)] &= s^n \mathcal{L}(f(t)) - s^{n-1} f(0) \\ &\quad - s^{n-2} f'(0) - \ldots - f^{(n-1)}(0). \qquad (10)\end{aligned}$$

This can be proved by Integration by Parts,

$$\begin{aligned}\mathcal{L}[f'(t)] &= \lim_{a \to 0} \int_0^a e^{-st} f'(t)\, dt \\ &= \lim_{a \to 0} [e^{-st} f(t)]_0^a + s \int_0^a e^{-st} f(t)\, dt] \\ &= \lim_{a \to 0} [e^{-sa} f(a) - f(0) + s \int_0^a e^{-st} f(t)\, dt] \\ &= s\mathcal{L}[f(t)] - f(0). \qquad (11)\end{aligned}$$

Continuing for higher order derivatives then gives

$$\mathcal{L}[f''(t)] = s^2 \mathcal{L}[f(t)] - sf(0) - f'(0). \qquad (12)$$

This property can be used to transform differential equations into algebraic equations, a procedure known as the Heaviside Calculus, which can then be inverse transformed to obtain the solution. For example, applying the Laplace transform to the equation

$$f''(t) + a_1 f'(t) + a_0 f(t) = 0 \qquad (13)$$

gives

$$\begin{aligned}\{s^2 \mathcal{L}[f(t)] - sf(0) - f'(0)\} + a_1 \{s\mathcal{L}[f(t)] - f(0)\} \\ + a_0 \mathcal{L}[f(t)] = 0 \qquad (14)\end{aligned}$$

$$\mathcal{L}[f(t)](s^2 + a_1 s + a_0) - sf(0) - f'(0) - a_1 f(0) = 0, \qquad (15)$$

which can be rearranged to

$$\mathcal{L}[f(t)] = \frac{sf(0) + [f'(0) + a_1 f(0)]}{s^2 + a_1 s + a_0}. \qquad (16)$$

If this equation can be inverse Laplace transformed, then the original differential equation is solved.

Consider EXPONENTIATION. If $\mathcal{L}(f(t)) = F(s)$ for $s > \alpha$, then $\mathcal{L}(e^{at}f(t)) = F(s-a)$ for $s > a + \alpha$.

$$F(s-a) = \int_0^\infty f(t)e^{-(s-a)t}\,dt = \int_0^\infty [f(t)e^{at}]e^{-st}\,dt = \mathcal{L}(e^{at}f(t)). \tag{17}$$

Consider INTEGRATION. If $f(t)$ is piecewise continuous and $|f(t)| \leq Me^{at}$, then

$$\mathcal{L}\left[\int_0^t f(t)\,dt\right] = \frac{1}{s}\mathcal{L}[f(t)]. \tag{18}$$

The inverse transform is known as the BROMWICH INTEGRAL, or sometimes the FOURIER-MELLIN INTEGRAL.

see also BROMWICH INTEGRAL, FOURIER-MELLIN INTEGRAL, FOURIER TRANSFORM, INTEGRAL TRANSFORM, LAPLACE-STIELTJES TRANSFORM, OPERATIONAL MATHEMATICS

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Laplace Transforms." Ch. 29 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 1019–1030, 1972.

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 824–863, 1985.

Churchill, R. V. *Operational Mathematics.* New York: McGraw-Hill, 1958.

Doetsch, G. *Introduction to the Theory and Application of the Laplace Transformation.* Berlin: Springer-Verlag, 1974.

Franklin, P. *An Introduction to Fourier Methods and the Laplace Transformation.* New York: Dover, 1958.

Jaeger, J. C. and Newstead, G. H. *An Introduction to the Laplace Transformation with Engineering Applications.* London: Methuen, 1949.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 467–469, 1953.

Spiegel, M. R. *Theory and Problems of Laplace Transforms.* New York: McGraw-Hill, 1965.

Widder, D. V. *The Laplace Transform.* Princeton, NJ: Princeton University Press, 1941.

Laplacian

The Laplacian operator for a SCALAR function ϕ is defined by

$$\nabla^2\phi = \frac{1}{h_1h_2h_3}\left[\frac{\partial}{\partial u_1}\left(\frac{h_2h_3}{h_1}\frac{\partial}{\partial u_1}\right) + \frac{\partial}{\partial u_2}\left(\frac{h_1h_3}{h_2}\frac{\partial}{\partial u_2}\right) + \frac{\partial}{\partial u_3}\left(\frac{h_1h_2}{h_3}\frac{\partial}{\partial u_3}\right)\right]\phi \tag{1}$$

in VECTOR notation, where the h_i are the SCALE FACTORS of the coordinate system. In TENSOR notation, the Laplacian is written

$$\begin{aligned}\nabla^2\phi &= (g^{\lambda\kappa}\phi_{;\lambda})_{;\kappa} = g^{\lambda\kappa}\frac{\partial^2\phi}{\partial x^\lambda \partial x^\kappa} - \Gamma^\lambda \frac{\partial\phi}{\partial x^\lambda}\\ &= \frac{1}{\sqrt{g}}\frac{\partial}{\partial x^j}\left(\sqrt{g}g^{ij}\frac{\partial\phi}{\partial x^i}\right),\end{aligned} \tag{2}$$

where $g_{;\kappa}$ is a COVARIANT DERIVATIVE and

$$\Gamma^\lambda \equiv \tfrac{1}{2}g^{\mu\nu}g^{\lambda\kappa}\left(\frac{\partial g_{\kappa\mu}}{\partial x^\nu} + \frac{\partial g_{\kappa\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial x^\kappa}\right). \tag{3}$$

The finite difference form is

$$\begin{aligned}\nabla^2\psi(x,y,z) = \frac{1}{h^2}[&\psi(x+h,y,z) + \psi(x-h,y,z)\\ &+\psi(x,y+h,z) + \psi(x,y-h,z) + \psi(x,y,z+h)\\ &+\psi(x,y,z-h) - 6\psi(x,y,z)].\end{aligned} \tag{4}$$

For a pure radial function $g(r)$,

$$\begin{aligned}\nabla^2 g(r) &\equiv \nabla\cdot[\nabla g(r)]\\ &= \nabla\cdot\left[\frac{\partial g(r)}{\partial r}\hat{\mathbf{r}} + \frac{1}{r}\frac{\partial g(r)}{\partial\theta}\hat{\boldsymbol{\theta}} + \frac{1}{r\sin\theta}\frac{\partial g(r)}{\partial\phi}\hat{\boldsymbol{\phi}}\right]\\ &= \nabla\cdot\left(\hat{\mathbf{r}}\frac{dg}{dr}\right).\end{aligned} \tag{5}$$

Using the VECTOR DERIVATIVE identity

$$\nabla\cdot(f\mathbf{A}) = f(\nabla\cdot\mathbf{A}) + (\nabla f)\cdot(\mathbf{A}), \tag{6}$$

so

$$\begin{aligned}\nabla^2 g(r) \equiv \nabla\cdot[\nabla g(r)] &= \frac{dg}{dr}\nabla\cdot\hat{\mathbf{r}} + \nabla\left(\frac{dg}{dr}\right)\cdot\hat{\mathbf{r}}\\ &= \frac{2}{r}\frac{dg}{dr} + \frac{d^2g}{dr^2}.\end{aligned} \tag{7}$$

Therefore, for a radial POWER law,

$$\begin{aligned}\nabla^2 r^n &= \frac{2}{r}nr^{n-1} + n(n-1)r^{n-2} = [2n + n(n-1)]r^{n-2}\\ &= n(n+1)r^{n-2}.\end{aligned} \tag{8}$$

A VECTOR Laplacian can also be defined for a VECTOR $\mathbf{A}$ by

$$\nabla^2\mathbf{A} = \nabla(\nabla\cdot\mathbf{A}) - \nabla\times(\nabla\times\mathbf{A}) \tag{9}$$

in vector notation. In tensor notation, $\mathbf{A}$ is written A_μ, and the identity becomes

$$\begin{aligned}\nabla^2 A_\mu &= A_{\mu;\lambda}{}^{;\lambda} = (g^{\lambda\kappa}A_{\mu;\lambda})_{;\kappa}\\ &= g^\lambda{}_{\kappa;\kappa}A_{\mu;\lambda} + g^{\lambda\kappa}A_{\mu;\lambda\kappa}.\end{aligned} \tag{10}$$

Similarly, a TENSOR Laplacian can be given by

$$\nabla^2 A_{\alpha\beta} = A_{\alpha\beta;\lambda}{}^{;\lambda}. \tag{11}$$

An identity satisfied by the Laplacian is

$$\nabla^2|\mathbf{x}\mathsf{A}| = \frac{{|\mathsf{A}|_2}^2 - |(\mathbf{x}\mathsf{A})\mathsf{A}^{\mathrm{T}}|^2}{|\mathbf{x}\mathsf{A}|^3}, \tag{12}$$

where $|\mathsf{A}|_2$ is the HILBERT-SCHMIDT NORM, $\mathbf{x}$ is a row VECTOR, and A^{T} is the MATRIX TRANSPOSE of A.

To compute the LAPLACIAN of the inverse distance function $1/r$, where $r \equiv |\mathbf{r}-\mathbf{r}'|$, and integrate the LAPLACIAN over a volume,

$$\int_V \nabla^2 \left(\frac{1}{|\mathbf{r}-\mathbf{r}'|}\right) d^3\mathbf{r}. \tag{13}$$

This is equal to

$$\begin{aligned}\int_V \nabla^2 \frac{1}{r} d^3\mathbf{r} &= \int_V \nabla\cdot\left(\nabla\frac{1}{r}\right) d^3\mathbf{r} = \int_S \left(\nabla\frac{1}{r}\right)\cdot d\mathbf{a}\\ &= \int_S \frac{\partial}{\partial r}\left(\frac{1}{r}\right)\hat{\mathbf{r}}\cdot d\mathbf{a} = \int_S -\frac{1}{r^2}\hat{\mathbf{r}}\cdot d\mathbf{a}\\ &= -4\pi\frac{R^2}{r^2}, \end{aligned} \tag{14}$$

where the integration is over a small SPHERE of RADIUS R. Now, for $r > 0$ and $R \to 0$, the integral becomes 0. Similarly, for $r = R$ and $R \to 0$, the integral becomes -4π. Therefore,

$$\nabla^2 \left(\frac{1}{|\mathbf{r}-\mathbf{r}'|}\right) = -4\pi\delta^3(\mathbf{r}-\mathbf{r}'), \tag{15}$$

where δ is the DELTA FUNCTION.

see also ANTILAPLACIAN

Laplacian Determinant Expansion by Minors

see DETERMINANT EXPANSION BY MINORS

Large Number

There are a wide variety of large numbers which crop up in mathematics. Some are contrived, but some actually arise in proofs. Often, it is possible to prove existence theorems by deriving some potentially huge upper limit which is frequently greatly reduced in subsequent versions (e.g., GRAHAM'S NUMBER, KOLMOGOROV-ARNOLD-MOSER THEOREM, MERTENS CONJECTURE, SKEWES NUMBER, WANG'S CONJECTURE).

Large decimal numbers beginning with 10^9 are named according to two mutually conflicting nomenclatures: the American system (in which the prefix stands for n in 10^{3+3n}) and the British system (in which the prefix stands for n in 10^{6n}). The following table gives the names assigned to various POWERS of 10 (Woolf 1982).

American	British	Power of 10
million	million	10^6
billion	milliard	10^9
trillion	billion	10^{12}
quadrillion		10^{15}
quintillion	trillion	10^{18}
sextillion		10^{21}
septillion	quadrillion	10^{24}
octillion		10^{27}
nonillion	quintillion	10^{30}
decillion		10^{33}
undecillion	sexillion	10^{36}
duodecillion		10^{39}
tredecillion	septillion	10^{42}
quattuordecillion		10^{45}
quindecillion	octillion	10^{48}
sexdecillion		10^{51}
septendecillion	nonillion	10^{54}
octodecillion		10^{57}
novemdecillion	decillion	10^{60}
vigintillion		10^{63}
	undecillion	10^{66}
	duodecillion	10^{72}
	tredecillion	10^{78}
	quattuordecillion	10^{84}
	quindecillion	10^{90}
	sexdecillion	10^{96}
	septendecillion	10^{102}
	octodecillion	10^{108}
	novemdecillion	10^{114}
	vigintillion	10^{120}
centillion		10^{303}
	centillion	10^{600}

see also 10, ACKERMANN NUMBER, ARROW NOTATION, BILLION, CIRCLE NOTATION, EDDINGTON NUMBER, G-FUNCTION, GÖBEL'S SEQUENCE, GOOGOL, GOOGOLPLEX, GRAHAM'S NUMBER, HUNDRED, HYPERFACTORIAL, JUMPING CHAMPION, LAW OF TRULY LARGE NUMBERS, MEGA, MEGISTRON, MILLION, MONSTER GROUP, MOSER, n-PLEX, POWER TOWER, SKEWES NUMBER, SMALL NUMBER, STEINHAUS-MOSER NOTATION, STRONG LAW OF LARGE NUMBERS, SUPERFACTORIAL, THOUSAND, WEAK LAW OF LARGE NUMBERS, ZILLION

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 59–62, 1996.

Crandall, R. E. "The Challenge of Large Numbers." *Sci. Amer.* **276**, 74–79, Feb. 1997.

Davis, P. J. *The Lore of Large Numbers.* New York: Random House, 1961.

Knuth, D. E. "Mathematics and Computer Science: Coping with Finiteness. Advances in Our Ability to Compute Are Bringing Us Substantially Closer to Ultimate Limitations." *Science* **194**, 1235–1242, 1976.

Munafo, R. "Large Numbers." `http://home.earthlink.net/~mrob/largenum`.

Spencer, J. "Large Numbers and Unprovable Theorems." *Amer. Math. Monthly* **90**, 669–675, 1983.

Woolf, H. B. (Ed. in Chief). *Webster's New Collegiate Dictionary.* Springfield, MA: Merriam, p. 782, 1980.

Large Prime

see GIGANTIC PRIME, LARGE NUMBER, TITANIC PRIME

Laspeyres' Index

The statistical INDEX

$$P_L \equiv \frac{\sum p_n q_0}{\sum p_0 q_0},$$

where p_n is the price per unit in period n and q_n is the quantity produced in period n.

see also INDEX

References

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, pp. 65–67, 1962.

Latin Cross

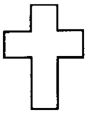

An irregular DODECAHEDRON CROSS in the shape of a dagger †. The six faces of a CUBE can be cut along seven EDGES and unfolded into a Latin cross (i.e., the Latin cross is the NET of the CUBE). Similarly, eight hypersurfaces of a HYPERCUBE can be cut along 17 SQUARES and unfolded to form a 3-D Latin cross.

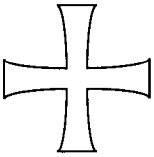

Another cross also called the Latin cross is illustrated above. It is a GREEK CROSS with flared ends, and is also known as the crux immissa or cross patée.

see also CROSS, DISSECTION, DODECAHEDRON, GREEK CROSS

Latin Rectangle

A $k \times n$ Latin rectangle is a $k \times n$ MATRIX with elements $a_{ij} \in \{1, 2, \ldots, n\}$ such that entries in each row and column are distinct. If $k = n$, the special case of a LATIN SQUARE results. A normalized Latin rectangle has first row $\{1, 2, \ldots, n\}$ and first column $\{1, 2, \ldots, k\}$. Let $L(k, n)$ be the number of normalized $k \times n$ Latin rectangles, then the total number of $k \times n$ Latin rectangles is

$$N(k, n) = \frac{n!(n-1)!L(k, n)}{(n-k)!}$$

(McKay and Rogoyski 1995), where $n!$ is a FACTORIAL. Kerewala (1941) found a RECURRENCE RELATION for $L(3, n)$, and Athreya, Pranesachar, and Singhi (1980) found a summation FORMULA for $L(4, n)$.

The asymptotic value of $L(o(n^{6/7}), n)$ was found by Godsil and McKay (1990). The numbers of $k \times n$ Latin rectangles are given in the following table from McKay and Rogoyski (1995). The entries $L(1, n)$ and $L(n, n)$ are omitted, since

$$L(1, n) = 1$$
$$L(n, n) = L(n-1, n),$$

but $L(1, 1)$ and $L(2, 1)$ are included for clarity. The values of $L(k, n)$ are given as a "wrap-around" series by Sloane's A001009.

n	k	$L(k, n)$
1	1	1
2	1	1
3	2	1
4	2	3
4	3	4
5	2	11
5	3	46
5	4	56
6	2	53
6	3	1064
6	4	6552
6	5	9408
7	2	309
7	3	36792
7	4	1293216
7	5	11270400
7	6	16942080
8	2	2119
8	3	1673792
8	4	420906504
8	5	27206658048
8	6	335390189568
8	7	535281401856
9	2	16687
9	3	103443808
9	4	207624560256
9	5	112681643983776
9	6	12962605404381184
9	7	224382967916691456
9	8	377597570964258816
10	2	148329
10	3	8154999232
10	4	147174521059584
10	5	746988383076286464
10	6	870735405591003709440
10	7	177144296983054185922560
10	8	4292039421591854273003520
10	9	7580721482160132811489280

References

Athreya, K. B.; Pranesachar, C. R.; and Singhi, N. M. "On the Number of Latin Rectangles and Chromatic Polynomial of $L(K_{r,s})$." *Europ. J. Combin.* **1**, 9–17, 1980.

Colbourn, C. J. and Dinitz, J. H. (Eds.) *CRC Handbook of Combinatorial Designs.* Boca Raton, FL: CRC Press, 1996.
Godsil, C. D. and McKay, B. D. "Asymptotic Enumeration of Latin Rectangles." *J. Combin. Th. Ser. B* **48**, 19–44, 1990.
Kerawla, S. M. "The Enumeration of Latin Rectangle of Depth Three by Means of Difference Equation" [sic]. *Bull. Calcutta Math. Soc.* **33**, 119–127, 1941.
McKay, B. D. and Rogoyski, E. "Latin Squares of Order 10." *Electronic J. Combinatorics* **2**, N3, 1–4, 1995. `http://www.combinatorics.org/Volume_2/volume2.html#N3`.
Ryser, H. J. "Latin Rectangles." §3.3 in *Combinatorial Mathematics.* Buffalo, NY: Math. Assoc. of Amer., pp. 35–37, 1963.
Sloane, N. J. A. Sequence A001009 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Latin Square

An $n \times n$ Latin square is a LATIN RECTANGLE with $k = n$. Specifically, a Latin square consists of n sets of the numbers 1 to n arranged in such a way that no orthogonal (row or column) contains the same two numbers. The numbers of Latin squares of order $n = 1$, 2, ... are 1, 2, 12, 576, ... (Sloane's A002860). A pair of Latin squares is said to be orthogonal if the n^2 pairs formed by juxtaposing the two arrays are all distinct.

Two of the Latin squares of order 3 are

$$\begin{bmatrix} 3 & 2 & 1 \\ 2 & 1 & 3 \\ 1 & 3 & 2 \end{bmatrix} \quad \begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix},$$

which are orthogonal. Two of the 576 Latin squares of order 4 are

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \\ 3 & 4 & 1 & 2 \\ 4 & 3 & 2 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \\ 4 & 3 & 2 & 1 \\ 2 & 1 & 4 & 3 \end{bmatrix}.$$

A normalized, or reduced, Latin square is a Latin square with the first row and column given by $\{1, 2, \ldots, n\}$. General FORMULAS for the number of *normalized* Latin squares $L(n,n)$ are given by Nechvatal (1981), Gessel (1987), and Shao and Wei (1992). The total number of Latin squares of order n can then be computed from

$$N(n,n) = n!(n-1)!L(n,n) = n!(n-1)!L(n-1,n).$$

The numbers of normalized Latin square of order $n = 1$, 2, ..., are 1, 1, 1, 4, 56, 9408, ... (Sloane's A000315). McKay and Rogoyski (1995) give the number of normalized LATIN RECTANGLES $L(k,n)$ for $n = 1$, ..., 10, as well as estimates for $L(n,n)$ with $n = 11$, 12, ..., 15.

n	$L(n,n)$
11	5.36×10^{33}
12	1.62×10^{44}
13	2.51×10^{56}
14	2.33×10^{70}
15	1.5×10^{86}

see also EULER SQUARE, KIRKMAN TRIPLE SYSTEM, PARTIAL LATIN SQUARE, QUASIGROUP

References

Colbourn, C. J. and Dinitz, J. H. *CRC Handbook of Combinatorial Designs.* Boca Raton, FL: CRC Press, 1996.
Gessel, I. "Counting Latin Rectangles." *Bull. Amer. Math. Soc.* **16**, 79–83, 1987.
Hunter, J. A. H. and Madachy, J. S. *Mathematical Diversions.* New York: Dover, pp. 33–34, 1975.
Kraitchik, M. "Latin Squares." §7.11 in *Mathematical Recreations.* New York: W. W. Norton, p. 178, 1942.
Lindner, C. C. and Rodger, C. A. *Design Theory.* Boca Raton, FL: CRC Press, 1997.
McKay, B. D. and Rogoyski, E. "Latin Squares of Order 10." *Electronic J. Combinatorics* **2**, N3, 1–4, 1995. `http://www.combinatorics.org/Volume_2/volume2.html#N3`.
Nechvatal, J. R. "Asymptotic Enumeration of Generalised Latin Rectangles." *Util. Math.* **20**, 273–292, 1981.
Ryser, H. J. "Latin Rectangles." §3.3 in *Combinatorial Mathematics.* Buffalo, NY: Math. Assoc. Amer., pp. 35–37, 1963.
Shao, J.-Y. and Wei, W.-D. "A Formula for the Number of Latin Squares." *Disc. Math.* **110**, 293–296, 1992.
Sloane, N. J. A. Sequences A002860/M2051 and A000315/M3690 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Latin-Graeco Square

see EULER SQUARE

Latitude

The latitude of a point on a SPHERE is the elevation of the point from the PLANE of the equator. The latitude δ is related to the COLATITUDE (the polar angle in SPHERICAL COORDINATES) by $\delta = \phi - 90°$. More generally, the latitude of a point on an ELLIPSOID is the ANGLE between a LINE PERPENDICULAR to the surface of the ELLIPSOID at the given point and the PLANE of the equator (Snyder 1987).

The equator therefore has latitude 0°, and the north and south poles have latitude ±90°, respectively. Latitude is also called GEOGRAPHIC LATITUDE or GEODETIC LATITUDE in order to distinguish it from several subtly different varieties of AUXILIARY LATITUDES.

The shortest distance between any two points on a SPHERE is the so-called GREAT CIRCLE distance, which can be directly computed from the latitudes and LONGITUDES of the two points.

see also AUXILIARY LATITUDE, COLATITUDE, CONFORMAL LATITUDE, GREAT CIRCLE, ISOMETRIC LATITUDE, LATITUDE, LONGITUDE, SPHERICAL COORDINATES

References

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, p. 13, 1987.

Lattice

A lattice is a system K such that $\forall A \in K$, $A \subset A$, and if $A \subset B$ and $B \subset A$, then $A = B$, where $=$ here means "is included in." Lattices offer a natural way to formalize and study the ordering of objects using a general concept known as the POSET (partially ordered set). The study of lattices is called LATTICE THEORY. Note that this type of lattice is an abstraction of the regular array of points known as LATTICE POINTS.

The following inequalities hold for any lattice:

$$(x \wedge y) \vee (x \wedge z) \leq x \wedge (y \vee z)$$

$$x \vee (y \wedge z) \leq (x \vee y) \wedge (x \vee z)$$

$$(x \wedge y) \vee (y \wedge z) \vee (z \wedge x) \leq (x \vee y) \wedge (y \vee z) \wedge (z \vee x)$$

$$(x \wedge y) \vee (x \wedge z) \leq x \wedge (y \vee (x \wedge z))$$

(Grätzer 1971, p. 35). The first three are the distributive inequalities, and the last is the modular identity.

see also DISTRIBUTIVE LATTICE, INTEGRATION LATTICE, LATTICE THEORY, MODULAR LATTICE, TORIC VARIETY

References
Grätzer, G. *Lattice Theory: First Concepts and Distributive Lattices.* San Francisco, CA: W. H. Freeman, 1971.

Lattice Algebraic System

A generalization of the concept of set unions and intersections.

Lattice Animal

A distinct (including reflections and rotations) arrangement of adjacent squares on a grid, also called fixed POLYOMINOES.

see also PERCOLATION THEORY, POLYOMINO

Lattice Distribution

A DISCRETE DISTRIBUTION of a random variable such that every possible value can be represented in the form $a + bn$, where $a, b \neq 0$ and n is an INTEGER.

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 927, 1972.

Lattice Graph

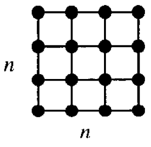

The lattice graph with n nodes on a side is denoted $L(n)$.

see also TRIANGULAR GRAPH

Lattice Groups

In the plane, there are 17 lattice groups, eight of which are pure translation. In $\mathbb{R}^3$, there are 32 POINT GROUPS and 230 SPACE GROUPS. In $\mathbb{R}^4$, there are 4783 space lattice groups.

see also POINT GROUPS, SPACE GROUPS, WALLPAPER GROUPS

Lattice Path

A path composed of connected horizontal and vertical line segments, each passing between adjacent LATTICE POINTS. A lattice path is therefore a SEQUENCE of points P_0, P_1, ..., P_n with $n \geq 0$ such that each P_i is a LATTICE POINT and P_{i+1} is obtained by offsetting one unit east (or west) or one unit north (or south).

The number of paths of length $a + b$ from the ORIGIN (0,0) to a point (a, b) which are restricted to east and north steps is given by the BINOMIAL COEFFICIENT $\binom{a+b}{a}$.

see also BALLOT PROBLEM, GOLYGON, KINGS PROBLEM, LATTICE POINT, p-GOOD PATH, RANDOM WALK

References
Dickau, R. M. "Shortest-Path Diagrams." `http://forum.swarthmore.edu/advanced/robertd/manhattan.html`.
Hilton, P. and Pederson, J. "Catalan Numbers, Their Generalization, and Their Uses." *Math. Intel.* **13**, 64–75, 1991.

Lattice Point

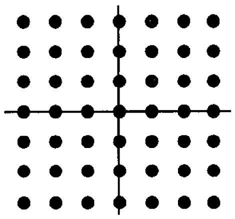

A POINT at the intersection of two or more grid lines in a ruled array. (The array of grid lines could be oriented to form unit cells in the shape of a square, rectangle, hexagon, etc.) However, unless otherwise specified, lattice points are generally taken to refer to points in a square array, i.e., points with coordinates $(m, n, \ldots)$, where m, n, ... are INTEGERS.

An n-D $\mathbb{Z}[\omega]$-lattice L_n lattice can be formally defined as a free $\mathbb{Z}[\omega]$-MODULE in complex n-D space $\mathbb{C}^n$.

The FRACTION of lattice points VISIBLE from the ORIGIN, as derived in Castellanos (1988, pp. 155–156), is

$$\frac{N'(r)}{N(r)} = \frac{\frac{24}{\pi^2} r^2 + \mathcal{O}(r \ln r)}{4r^2 + \mathcal{O}(r)} = \frac{\frac{6}{\pi^2} + \mathcal{O}\left(\frac{\ln r}{r}\right)}{1 + \mathcal{O}\left(\frac{1}{r}\right)} = \frac{6}{\pi^2}.$$

Therefore, this is also the probability that two randomly picked integers will be RELATIVELY PRIME to one another.

For $2 \leq n \leq 32$, it is possible to select $2n$ lattice points with $x, y \in [1, n]$ such that no three are in a straight

LINE. The number of distinct solutions (not counting reflections and rotations) for $n = 1, 2, \ldots,$ are 1, 1, 4, 5, 11, 22, 57, 51, 156 ... (Sloane's A000769). For large n, it is conjectured that it is only possible to select at most $(c+\epsilon)n$ lattice points with no three COLLINEAR, where

$$c = (2\pi^2/3)^{1/3} \approx 1.85$$

(Guy and Kelly 1968; Guy 1994, p. 242). The number of the n^2 lattice points $x, y \in [1, n]$ which can be picked with no four CONCYCLIC is $\mathcal{O}(n^{2/3} - \epsilon)$ (Guy 1994, p. 241).

A special set of POLYGONS defined on the regular lattice are the GOLYGONS. A NECESSARY and SUFFICIENT condition that a linear transformation transforms a lattice to itself is that it be UNIMODULAR. M. Ajtai has shown that there is no efficient ALGORITHM for finding any fraction of a set of spanning vectors in a lattice having the shortest lengths unless there is an efficient algorithm for all of them (of which none is known). This result has potential applications to cryptography and authentication (Cipra 1996).

see also BARNES-WALL LATTICE, BLICHFELDT'S THEOREM, BROWKIN'S THEOREM, CIRCLE LATTICE POINTS, COXETER-TODD LATTICE, EHRHART POLYNOMIAL, GAUSS'S CIRCLE PROBLEM, GOLYGON, INTEGRATION LATTICE, JARNICK'S INEQUALITY, LATTICE PATH, LATTICE SUM, LEECH LATTICE, MINKOWSKI CONVEX BODY THEOREM, MODULAR LATTICE, N-CLUSTER, NOSARZEWSKA'S INEQUALITY, PICK'S THEOREM, POSET, RANDOM WALK, SCHINZEL'S THEOREM, SCHRÖDER NUMBER, VISIBLE POINT, VORONOI POLYGON

References

Apostol, T. *Introduction to Analytic Number Theory.* New York: Springer-Verlag, 1995.

Castellanos, D. "The Ubiquitous Pi." *Math. Mag.* **61**, 67–98, 1988.

Cipra, B. "Lattices May Put Security Codes on a Firmer Footing." *Science* **273**, 1047–1048, 1996.

Eppstein, D. "Lattice Theory and Geometry of Numbers." `http://www.ics.uci.edu/~eppstein/junkyard/lattice.html`.

Guy, R. K. "Gauß's Lattice Point Problem," "Lattice Points with Distinct Distances," "Lattice Points, No Four on a Circle," and "The No-Three-in-a-Line Problem." §F1, F2, F3, and F4 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 240–244, 1994.

Guy, R. K. and Kelly, P. A. "The No-Three-in-Line-Problem." *Canad. Math. Bull.* **11**, 527–531, 1968.

Hammer, J. *Unsolved Problems Concerning Lattice Points.* London: Pitman, 1977.

Sloane, N. J. A. Sequence A000769/M3252 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Lattice Reduction

The process finding a reduced set of basis vectors for a given LATTICE having certain special properties. Lattice reduction is implemented in Mathematica® (Wolfram Research, Champaign, IL) using the function `LatticeReduce`. Lattice reduction algorithms are used in a number of modern number theoretical applications, including in the discovery of a SPIGOT ALGORITHM for PI.

see also INTEGER RELATION, PSLQ ALGORITHM

Lattice Sum

Cubic lattice sums include the following:

$$b_2(2s) \equiv \sum_{i,j=-\infty}^{\infty}{}' \frac{(-1)^{i+j}}{(i^2+j^2)^s} \tag{1}$$

$$b_3(2s) \equiv \sum_{i,j,k=-\infty}^{\infty}{}' \frac{(-1)^{i+j+k}}{(i^2+j^2+k^2)^s} \tag{2}$$

$$b_n(2s) \equiv \sum_{k_1,\ldots,k_n=-\infty}^{\infty}{}' \frac{(-1)^{k_1+\ldots+k_n}}{({k_1}^2+\ldots+{k_n}^2)^s}, \tag{3}$$

where the prime indicates that summation over $(0,0,0)$ is excluded. As shown in Borwein and Borwein (1987, pp. 288–301), these have closed forms for even n

$$b_2(2s) = -4\beta(s)\eta(s) \tag{4}$$

$$b_4(2s) = -8\eta(s)\eta(s-1) \tag{5}$$

$$b_8(2s) = -16\zeta(s)\eta(s-3), \quad \text{for } \Re[s] > 1 \tag{6}$$

where $\beta(z)$ is the DIRICHLET BETA FUNCTION, $\eta(z)$ is the DIRICHLET ETA FUNCTION, and $\zeta(z)$ is the RIEMANN ZETA FUNCTION. The lattice sums evaluated at $s = 1$ are called the MADELUNG CONSTANTS. Borwein and Borwein (1986) prove that $b_8(2)$ converges (the closed form for $b_8(2s)$ above does not apply for $s = 1$), but its value has not been computed.

For hexagonal sums, Borwein and Borwein (1987, p. 292) give

$$h_2(2s) \equiv \frac{4}{3} \sum_{m,n=-\infty}^{\infty} \frac{\sin[(n+1)\theta]\sin[(m+1)\theta] - \sin(n\theta)\sin[(m-1)\theta]}{\left[(n+\frac{1}{2}m)^2 + 3(\frac{1}{2}m)^2\right]^s}, \tag{7}$$

where $\theta \equiv 2\pi/3$. This MADELUNG CONSTANT is expressible in closed form for $s = 1$ as

$$h_2(2) = \pi \ln 3\sqrt{3}. \tag{8}$$

see also BENSON'S FORMULA, MADELUNG CONSTANTS

References

Borwein, D. and Borwein, J. M. "On Some Trigonometric and Exponential Lattice Sums." *J. Math. Anal.* **188**, 209–218, 1994.

Borwein, D.; Borwein, J. M.; and Shail, R. "Analysis of Certain Lattice Sums." *J. Math. Anal.* **143**, 126–137, 1989.

Borwein, D. and Borwein, J. M. "A Note on Alternating Series in Several Dimensions." *Amer. Math. Monthly* **93**, 531–539, 1986.

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, 1987.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/mdlung/mdlung.html`.

Glasser, M. L. and Zucker, I. J. "Lattice Sums." In *Perspectives in Theoretical Chemistry: Advances and Perspectives* **5**, 67–139, 1980.

Lattice Theory

Lattice theory is the study of sets of objects known as LATTICES. It is an outgrowth of the study of BOOLEAN ALGEBRAS, and provides a framework for unifying the study of classes or ordered sets in mathematics. Its study was given a great boost by a series of papers and subsequent textbook written by Birkhoff (1967).

see also LATTICE

References

Birkhoff, G. *Lattice Theory, 3rd ed.* Providence, RI: Amer. Math. Soc., 1967.

Grätzer, G. *Lattice Theory: First Concepts and Distributive Lattices.* San Francisco, CA: W. H. Freeman, 1971.

Latus Rectum

Twice the SEMILATUS RECTUM.

see also PARABOLA

Laurent Polynomial

A Laurent polynomial with COEFFICIENTS in the FIELD $\mathbb{F}$ is an algebraic object that is typically expressed in the form

$$\ldots + a_{-n}t^{-n} + a_{-(n-1)}t^{-(n-1)} + \ldots + a_{-1}t^{-1} + a_0 + a_1t + \ldots + a_nt^n + \ldots,$$

where the a_i are elements of $\mathbb{F}$, and only finitely many of the a_i are NONZERO. A Laurent polynomial is an algebraic object in the sense that it is treated as a POLYNOMIAL except that the indeterminant "t" can also have NEGATIVE POWERS.

Expressed more precisely, the collection of Laurent polynomials with COEFFICIENTS in a FIELD $\mathbb{F}$ form a RING, denoted $\mathbb{F}[t, t^{-1}]$, with RING operations given by componentwise addition and multiplication according to the relation

$$at^n \cdot bt^m = abt^{n+m}$$

for all n and m in the INTEGERS. Formally, this is equivalent to saying that $\mathbb{F}[t, t^{-1}]$ is the GROUP RING of the INTEGERS and the FIELD $\mathbb{F}$. This corresponds to $\mathbb{F}[t]$ (the POLYNOMIAL ring in one variable for $\mathbb{F}$) being the GROUP RING or MONOID ring for the MONOID of natural numbers and the FIELD $\mathbb{F}$.

see also POLYNOMIAL

References

Lang, S. *Undergraduate Algebra, 2nd ed.* New York: Springer-Verlag, 1990.

Laurent Series

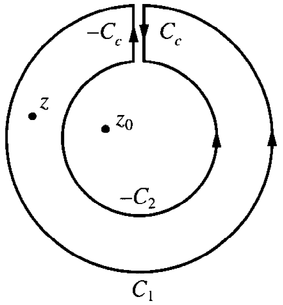

Let there be two circular contours C_2 and C_1, with the radius of C_1 larger than that of C_2. Let z_0 be interior to C_1 and C_2, and z be between C_1 and C_2. Now create a cut line C_c between C_1 and C_2, and integrate around the path $C \equiv C_1 + C_c - C_2 - C_c$, so that the plus and minus contributions of C_c cancel one another, as illustrated above. From the CAUCHY INTEGRAL FORMULA,

$$\begin{aligned} f(z) &= \frac{1}{2\pi i}\int_C \frac{f(z')}{z'-z}\,dz' \\ &= \frac{1}{2\pi i}\int_{C_1} \frac{f(z')}{z'-z}\,dz' + \frac{1}{2\pi i}\int_{C_c} \frac{f(z')}{z'-z}\,dz' \\ &\quad - \frac{1}{2\pi i}\int_{C_2} \frac{f(z')}{z'-z}\,dz' - \frac{1}{2\pi i}\int_{C_c} \frac{f(z')}{z'-z}\,dz' \\ &= \frac{1}{2\pi i}\int_{C_1} \frac{f(z')}{z'-z}\,dz' - \frac{1}{2\pi i}\int_{C_2} \frac{f(z')}{z'-z}\,dz'. \end{aligned} \tag{1}$$

Now, since contributions from the cut line in opposite directions cancel out,

$$\begin{aligned} f(z) &= \frac{1}{2\pi i}\int_{C_1} \frac{f(z')}{(z'-z_0)-(z-z_0)}\,dz' \\ &\quad - \frac{1}{2\pi i}\int_{C_2} \frac{f(z')}{(z'-z_0)-(z-z_0)}\,dz' \\ &= \frac{1}{2\pi i}\int_{C_1} \frac{f(z')}{(z'-z_0)\left(1-\frac{z-z_0}{z'-z_0}\right)}\,dz' \\ &\quad - \frac{1}{2\pi i}\int_{C_2} \frac{f(z')}{(z-z_0)\left(\frac{z'-z_0}{z-z_0}-1\right)}\,dz' \\ &= \frac{1}{2\pi i}\int_{C_1} \frac{f(z')}{(z'-z_0)\left(1-\frac{z-z_0}{z'-z_0}\right)}\,dz' \\ &\quad + \frac{1}{2\pi i}\int_{C_2} \frac{f(z')}{(z-z_0)\left(1-\frac{z'-z_0}{z-z_0}\right)}\,dz'. \end{aligned} \tag{2}$$

For the first integral, $|z'-z_0| > |z-z_0|$. For the second, $|z'-z_0| < |z-z_0|$. Now use the TAYLOR EXPANSION (valid for $|t| < 1$)

$$\frac{1}{1-t} = \sum_{n=0}^{\infty} t^n \tag{3}$$

to obtain

$$\begin{aligned} f(z) &= \frac{1}{2\pi i}\left[\int_{C_1} \frac{f(z')}{z'-z_0}\sum_{n=0}^{\infty}\left(\frac{z-z_0}{z'-z_0}\right)^n dz' \right. \\ &\quad \left. + \int_{C_2} \frac{f(z')}{z-z_0}\sum_{n=0}^{\infty}\left(\frac{z'-z_0}{z-z_0}\right)^n dz'\right] \\ &= \frac{1}{2\pi i}\sum_{n=0}^{\infty}(z-z_0)^n \int_{C_1}\frac{f(z')}{(z'-z_0)^{n+1}}\,dz' \\ &\quad + \frac{1}{2\pi i}\sum_{n=0}^{\infty}(z-z_0)^{-n-1}\int_{C_2}(z'-z_0)^n f(z')\,dz' \\ &= \frac{1}{2\pi i}\sum_{n=0}^{\infty}(z-z_0)^n \int_{C_1}\frac{f(z')}{(z'-z_0)^{n+1}}\,dz' \\ &\quad + \frac{1}{2\pi i}\sum_{n=1}^{\infty}(z-z_0)^{-n}\int_{C_2}(z'-z_0)^{n+1} f(z')\,dz', \end{aligned} \tag{4}$$

where the second term has been re-indexed. Re-indexing again,

$$\begin{aligned} f(z) &= \frac{1}{2\pi i}\sum_{n=0}^{\infty}(z-z_0)^n \int_{C_1}\frac{f(z')}{(z'-z_0)^{n+1}}\,dz' \\ &\quad + \frac{1}{2\pi i}\sum_{n=-\infty}^{-1}(z-z_0)^n \int_{C_2}\frac{f(z')}{(z'-z_0)^{n+1}}\,dz'. \end{aligned} \tag{5}$$

Now, use the CAUCHY INTEGRAL THEOREM, which requires that any CONTOUR INTEGRAL of a function which encloses no POLES has value 0. But $1/(z'-z_0)^{n+1}$ is never singular inside C_2 for $n \geq 0$, and $1/(z'-z_0)^{n+1}$ is never singular inside C_1 for $n \leq -1$. Similarly, there are no POLES in the closed cut $C_c - C_c$. We can therefore replace C_1 and C_2 in the above integrals by C without altering their values, so

$$\begin{aligned} f(z) &= \frac{1}{2\pi i}\sum_{n=0}^{\infty}(z-z_0)^n \int_{C}\frac{f(z')}{(z'-z_0)^{n+1}}\,dz' \\ &\quad + \frac{1}{2\pi i}\sum_{n=-\infty}^{-1}(z-z_0)^n \int_{C}\frac{f(z')}{(z'-z_0)^{n+1}}\,dz' \\ &= \frac{1}{2\pi i}\sum_{n=-\infty}^{\infty}(z-z_0)^n \int_{C}\frac{f(z')}{(z'-z_0)^{n+1}}\,dz' \\ &\equiv \sum_{n=-\infty}^{\infty} a_n(z-z_0)^n. \end{aligned} \tag{6}$$

The only requirement on C is that it encloses z, so we are free to choose any contour γ that does so. The RESIDUES a_n are therefore defined by

$$a_n \equiv \frac{1}{2\pi i}\int_{\gamma}\frac{f(z')}{(z'-z_0)^{n+1}}\,dz'. \tag{7}$$

see also MACLAURIN SERIES, RESIDUE (COMPLEX ANALYSIS), TAYLOR SERIES

References

Arfken, G. "Laurent Expansion." §6.5 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 376–384, 1985.

Morse, P. M. and Feshbach, H. "Derivatives of Analytic Functions, Taylor and Laurent Series." §4.3 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 374–398, 1953.

Law

A law is a mathematical statement which always holds true. Whereas "laws" in physics are generally experimental observations backed up by theoretical underpinning, laws in mathematics are generally THEOREMS which can formally be proven true under the stated conditions. However, the term is also sometimes used in the sense of an empirical observation, e.g., BENFORD'S LAW.

see also ABSORPTION LAW, BENFORD'S LAW, CONTRADICTION LAW, DE MORGAN'S DUALITY LAW, DE MORGAN'S LAWS, ELLIPTIC CURVE GROUP LAW, EXCLUDED MIDDLE LAW, EXPONENT LAWS, GIRKO'S CIRCULAR LAW, LAW OF COSINES, LAW OF SINES, LAW OF TANGENTS, LAW OF TRULY LARGE NUMBERS, MORRIE'S LAW, PARALLELOGRAM LAW, PLATEAU'S LAWS, QUADRATIC RECIPROCITY LAW, STRONG LAW OF LARGE NUMBERS, STRONG LAW OF SMALL NUMBERS, SYLVESTER'S INERTIA LAW, TRICHOTOMY LAW, VECTOR TRANSFORMATION LAW, WEAK LAW OF LARGE NUMBERS, ZIPF'S LAW

Law of Anomalous Numbers

see BENFORD'S LAW

Law of Cancellation

see CANCELLATION LAW

Law of Cosines

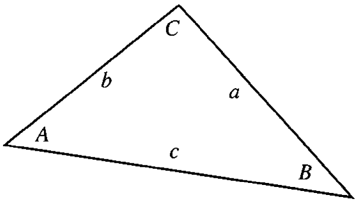

Let a, b, and c be the lengths of the legs of a TRIANGLE opposite ANGLES A, B, and C. Then the law of cosines states

$$c^2 = a^2 + b^2 - 2ab\cos C. \tag{1}$$

This law can be derived in a number of ways. The definition of the DOT PRODUCT incorporates the law of cosines, so that the length of the VECTOR from $\mathbf{X}$ to $\mathbf{Y}$ is given by

$$\begin{aligned} |\mathbf{X}-\mathbf{Y}|^2 &= (\mathbf{X}-\mathbf{Y})\cdot(\mathbf{X}-\mathbf{Y}) && (2) \\ &= \mathbf{X}\cdot\mathbf{X} - 2\mathbf{X}\cdot\mathbf{Y} + \mathbf{Y}\cdot\mathbf{Y} && (3) \\ &= |\mathbf{X}|^2 + |\mathbf{Y}|^2 - 2|\mathbf{X}|\,|\mathbf{Y}|\cos\theta, && (4) \end{aligned}$$

where θ is the ANGLE between $\mathbf{X}$ and $\mathbf{Y}$.

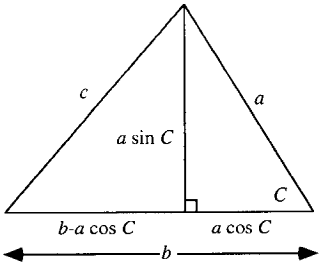

The formula can also be derived using a little geometry and simple algebra. From the above diagram,

$$\begin{aligned} c^2 &= (a\sin C)^2 + (b - a\cos C)^2 \\ &= a^2\sin^2 c + b^2 - 2ab\cos C + a^2\cos^2 C \\ &= a^2 + b^2 - 2ab\cos C. \end{aligned} \tag{5}$$

The law of cosines for the sides of a SPHERICAL TRIANGLE states that

$$\cos a = \cos b\cos c + \sin b\sin c\cos A \tag{6}$$

$$\cos b = \cos c\cos a + \sin c\sin a\cos B \tag{7}$$

$$\cos c = \cos a\cos b + \sin a\sin b\cos C \tag{8}$$

(Beyer 1987). The law of cosines for the angles of a SPHERICAL TRIANGLE states that

$$\cos A = -\cos B\cos C + \sin B\sin C\cos a \tag{9}$$

$$\cos B = -\cos C\cos A + \sin C\sin A\cos b \tag{10}$$

$$\cos C = -\cos A\cos B + \sin A\sin B\cos c \tag{11}$$

(Beyer 1987).

see also LAW OF SINES, LAW OF TANGENTS

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 79, 1972.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 148–149, 1987.

Law of Large Numbers

see LAW OF TRULY LARGE NUMBERS, STRONG LAW OF LARGE NUMBERS, WEAK LAW OF LARGE NUMBERS

Law of Sines

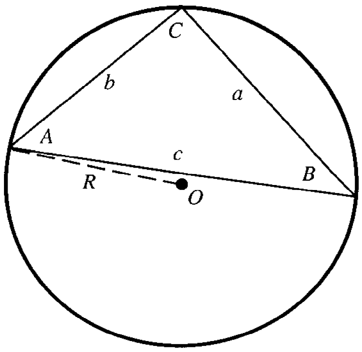

Let a, b, and c be the lengths of the LEGS of a TRIANGLE opposite ANGLES A, B, and C. Then the law of sines states that

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C} = 2R, \tag{1}$$

where R is the radius of the CIRCUMCIRCLE. Other related results include the identities

$$a(\sin B - \sin C) + b(\sin C - \sin A) + c(\sin A - \sin B) = 0 \tag{2}$$

$$a = b\cos C + c\cos B, \tag{3}$$

the LAW OF COSINES

$$\cos A = \frac{c^2 + b^2 - a^2}{2bc}, \tag{4}$$

and the LAW OF TANGENTS

$$\frac{a+b}{a-b} = \frac{\tan[\frac{1}{2}(A+B)]}{\tan[\frac{1}{2}(A-B)]}. \tag{5}$$

The law of sines for oblique SPHERICAL TRIANGLES states that

$$\frac{\sin a}{\sin A} = \frac{\sin b}{\sin B} = \frac{\sin c}{\sin C}. \tag{6}$$

see also LAW OF COSINES, LAW OF TANGENTS

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 79, 1972.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 148, 1987.

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 1–3, 1967.

Law of Small Numbers

see STRONG LAW OF SMALL NUMBERS

Law of Tangents

Let a TRIANGLE have sides of lengths a, b, and c and let the ANGLES opposite these sides by A, B, and C. The law of tangents states

$$\frac{a-b}{a+b} = \frac{\tan[\frac{1}{2}(A-B)]}{\tan[\frac{1}{2}(A+B)]}.$$

An analogous result for oblique SPHERICAL TRIANGLES states that

$$\frac{\tan[\frac{1}{2}(a-b)]}{\tan[\frac{1}{2}(a+b)]} = \frac{\tan[\frac{1}{2}(A-B)]}{\tan[\frac{1}{2}(A+B)]}.$$

see also LAW OF COSINES, LAW OF SINES

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 79, 1972.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 145 and 149, 1987.

Law of Truly Large Numbers

With a large enough sample, any outrageous thing is likely to happen (Diaconis and Mosteller 1989). Littlewood (1953) considered an event which occurs one in a million times to be "surprising." Taking this definition, close to 100,000 surprising events are "expected" each year in the United States alone and, in the world at large, "we can be absolutely sure that we will see incredibly remarkable events" (Diaconis and Mosteller 1989).

see also COINCIDENCE, STRONG LAW OF LARGE NUMBERS, STRONG LAW OF SMALL NUMBERS, WEAK LAW OF LARGE NUMBERS

References
Diaconis, P. and Mosteller, F. "Methods of Studying Coincidences." *J. Amer. Statist. Assoc.* **84**, 853–861, 1989.
Littlewood, J. E. *Littlewood's Miscellany.* Cambridge, England: Cambridge University Press, 1986.

Lax-Milgram Theorem

Let ϕ be a bounded COERCIVE bilinear FUNCTIONAL on a HILBERT SPACE H. Then for every bounded linear FUNCTIONAL f on H, there exists a unique $x_f \in H$ such that

$$f(x) = \phi(x, x_f)$$

for all $x \in H$.

References
Debnath, L. and Mikusiński, P. *Introduction to Hilbert Spaces with Applications.* San Diego, CA: Academic Press, 1990.
Zeidler, E. *Applied Functional Analysis: Applications to Mathematical Physics.* New York: Springer-Verlag, 1995.

Lax Pair

A pair of linear OPERATORS L and A associated with a given PARTIAL DIFFERENTIAL EQUATION which can be used to solve the equation. However, it turns out to be very difficult to find the L and A corresponding to a given equation, so it is actually simpler to postulate a given L and A and determine to which PARTIAL DIFFERENTIAL EQUATION they correspond (Infeld and Rowlands 1990).

see also PARTIAL DIFFERENTIAL EQUATION

References
Infeld, E. and Rowlands, G. "Integrable Equations in Two Space Dimensions as Treated by the Zakharov Shabat Methods." §7.10 in *Nonlinear Waves, Solitons, and Chaos.* Cambridge, England: Cambridge University Press, pp. 216–223, 1990.

Layer

see p-LAYER

Le Cam's Identity

Let S_n be the sum of n random variates X_i with a BERNOULLI DISTRIBUTION with $P(X_i = 1) = p_i$. Then

$$\sum_{k=0}^{\infty} \left| P(S_n = k) - \frac{e^{-\lambda}\lambda^k}{k!} \right| < 2 \sum_{i=1}^{n} {p_i}^2,$$

where

$$\lambda \equiv \sum_{i=1}^{n} p_i.$$

see also BERNOULLI DISTRIBUTION

References
Cox, D. A. "Introduction to Fermat's Last Theorem." *Amer. Math. Monthly* **101**, 3–14, 1994.

Leading Digit Phenomenon

see BENFORD'S LAW

Leading Order Analysis

A procedure for determining the behavior of an nth order ORDINARY DIFFERENTIAL EQUATION at a REMOVABLE SINGULARITY without actually solving the equation. Consider

$$\frac{d^n y}{dz^n} = F\left(\frac{d^{n-1}y}{dz^{n-1}}, \ldots, \frac{dy}{dx}, y, z\right), \tag{1}$$

where F is ANALYTIC in z and rational in its other arguments. Proceed by making the substitution

$$y(z) \equiv a(z - z_0)^\alpha \tag{2}$$

with $\alpha < 1$. For example, in the equation

$$\frac{d^2 y}{dz^2} = 6y^2 + Ay, \tag{3}$$

making the substitution gives

$$a\alpha(\alpha-1)(z-z_0)^{\alpha-2} = 6a^2(z-z_0)^{2\alpha} + Aa(az-z_0)^\alpha. \tag{4}$$

The most singular terms (those with the most NEGATIVE exponents) are called the "dominant balance terms," and must balance exponents and COEFFICIENTS at the SINGULARITY. Here, the first two terms are dominant, so

$$\alpha - 2 = 2\alpha \Rightarrow \alpha = -2 \tag{5}$$

$$6a = 6a^2 \Rightarrow a = 1, \tag{6}$$

and the solution behaves as $y(z) = (z - z_0)^{-2}$. The behavior in the NEIGHBORHOOD of the SINGULARITY is given by expansion in a LAURENT SERIES, in this case,

$$y(z) = \sum_{j=0}^{\infty} a_j (z - z_0)^{j-2}. \tag{7}$$

Plugging this series in yields

$$\sum_{j=0}^{\infty} a_j(j-2)(j-3)(z-z_0)^{j-4}$$
$$= 6\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} a_j a_k (z-z_0)^{j+k-4} + A\sum_{j=0}^{\infty} a_j (z-z_0)^{j-2}. \tag{8}$$

This gives Recurrence Relations, in this case with a_6 arbitrary, so the $(z-z_0)^6$ term is called the resonance or Kovalevskaya Exponent. At the resonances, the Coefficient will always be arbitrary. If no resonance term is present, the Pole present is not ordinary, and the solution must be investigated using a Psi Function.

see also Psi Function

References
Tabor, M. *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, p. 330, 1989.

Leaf (Foliation)

Let M^n be an n-Manifold and let $\mathsf{F} = \{F_\alpha\}$ denote a Partition of M into Disjoint path-connected Subsets. Then if F is a Foliation of M, each F_α is called a leaf and is not necessarily closed or compact.

see also Foliation

References
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 284, 1976.

Leaf (Tree)

An unconnected end of a Tree.

see also Branch, Child, Fork, Root (Tree), Tree

Leakage

see Aliasing

Least Bound

see Supremum

Least Common Multiple

The least common multiple of two numbers n_1 and n_2 is denoted $\mathrm{LCM}(n_1, n_2)$ or $[n_1, n_2]$ and can be obtained by finding the Prime factorization of each

$$n_1 = {p_1}^{a_1} \cdots {p_n}^{a_n} \tag{1}$$

$$n_2 = {p_1}^{b_1} \cdots {p_n}^{b_n}, \tag{2}$$

where the p_is are all Prime Factors of n_1 and n_2, and if p_i does not occur in one factorization, then the corresponding exponent is 0. The least common multiple is then

$$\mathrm{LCM}(n_1, n_2) = [n_1, n_2] = \prod_{i=1}^{n} {p_i}^{\max(a_i, b_i)}. \tag{3}$$

Let m be a common multiple of a and b so that

$$m = ha = kb. \tag{4}$$

Write $a = a_1(a,b)$ and $b = b_1(a,b)$, where a_1 and b_1 are Relatively Prime by definition of the Greatest Common Divisor $(a_1, b_1) = 1$. Then $ha_1 = kb_1$, and from the Division Lemma (given that ha_1 is Divisible by b and $(b_1, a_1) = 0$), we have h is Divisible by b_1, so

$$h = nb_1 \tag{5}$$

$$m = ha = nb_1 a = n\frac{ab}{(a,b)}. \tag{6}$$

The smallest m is given by $n = 1$,

$$\mathrm{LCM}(a,b) = \frac{ab}{\mathrm{GCD}(a,b)}, \tag{7}$$

so

$$\mathrm{GCD}(a,b)\ \mathrm{LCM}(a,b) = ab \tag{8}$$

$$(a,b)[a,b] = ab. \tag{9}$$

The LCM is Idempotent

$$[a,a] = a, \tag{10}$$

Commutative

$$[a,b] = [b,a], \tag{11}$$

Associative

$$[a,b,c] = [[a,b],c] = [a,[b,c]], \tag{12}$$

Distributive

$$[ma, mb, mc] = m[a,b,c], \tag{13}$$

and satisfies the Absorption Law

$$(a,[a,b]) = a. \tag{14}$$

It is also true that

$$[ma, mb] = \frac{(ma)(mb)}{(ma, mb)} = m\frac{ab}{(a,b)} = m[a,b]. \tag{15}$$

see also Greatest Common Divisor, Mangoldt Function, Relatively Prime

References
Guy, R. K. "Density of a Sequence with L.C.M. of Each Pair Less than x." §E2 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 200–201, 1994.

Least Deficient Number

A number for which

$$\sigma(n) = 2n - 1.$$

All POWERS of 2 are least deficient numbers.

see also DEFICIENT NUMBER, QUASIPERFECT NUMBER

Least Period

The smallest n for which a point x_0 is a PERIODIC POINT of a function f so that $f^n(x_0) = x_0$. For example, for the FUNCTION $f(x) = -x$, all points x have period 2 (including $x = 0$). However, $x = 0$ has a *least* period of 1. The analogous concept exists for a PERIODIC SEQUENCE, but not for a PERIODIC FUNCTION. The least period is also called the EXACT PERIOD.

Least Prime Factor

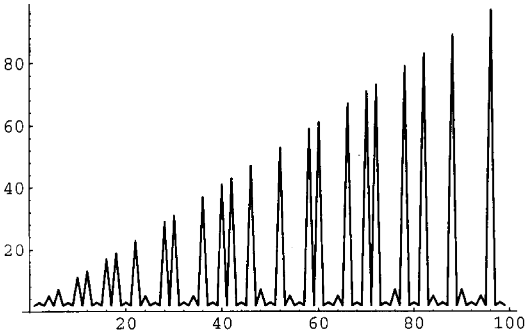

For an INTEGER $n \geq 2$, let $\operatorname{lpf}(x)$ denote the LEAST PRIME FACTOR of n, i.e., the number p_1 in the factorization

$$n = {p_1}^{a_1} \cdots {p_k}^{a_k},$$

with $p_i < p_j$ for $i < j$. For $n = 2, 3, \ldots$, the first few are 2, 3, 2, 5, 2, 7, 2, 3, 2, 11, 2, 13, 2, 3, ... (Sloane's A020639). The above plot of the least prime factor function can be seen to resemble a jagged terrain of mountains, which leads to the appellation of "TWIN PEAKS" to a PAIR of INTEGERS (x, y) such that

1. $x < y$,
2. $\operatorname{lpf}(x) = \operatorname{lpf}(y)$,
3. For all z, $x < z < y$ IMPLIES $\operatorname{lpf}(z) < \operatorname{lpf}(x)$.

The least *multiple* prime factors for SQUAREFUL integers are 2, 2, 3, 2, 2, 3, 2, 2, 5, 3, 2, 2, 2, ... (Sloane's A046027).

see also ALLADI-GRINSTEAD CONSTANT, DISTINCT PRIME FACTORS, ERDŐS-SELFRIDGE FUNCTION, FACTOR, GREATEST PRIME FACTOR, LEAST COMMON MULTIPLE, MANGOLDT FUNCTION, PRIME FACTORS, TWIN PEAKS

References

Sloane, N. J. A. Sequence A020639 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Least Squares Fitting

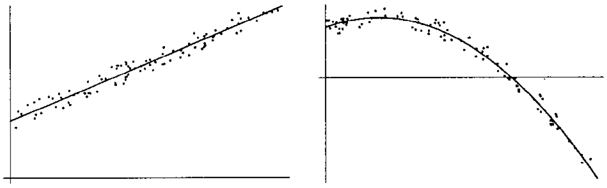

A mathematical procedure for finding the best fitting curve to a given set of points by minimizing the sum of the squares of the offsets ("the residuals") of the points from the curve. The sum of the *squares* of the offsets is used instead of the offset absolute values because this allows the residuals to be treated as a continuous differentiable quantity. However, because squares of the offsets are used, outlying points can have a disproportionate effect on the fit, a property which may or may not be desirable depending on the problem at hand.

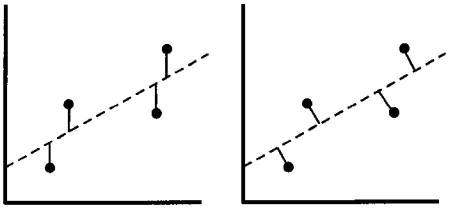

vertical offsets *perpendicular offsets*

In practice, the *vertical* offsets from a line are almost always minimized instead of the *perpendicular* offsets. This allows uncertainties of the data points along the x- and y-axes to be incorporated simply, and also provides a much simpler analytic form for the fitting parameters than would be obtained using a fit based on perpendicular distances. In addition, the fitting technique can be easily generalized from a best-fit *line* to a best-fit *polynomial* when sums of vertical distances are used (which is not the case using perpendicular distances). For a reasonable number of noisy data points, the difference between vertical and perpendicular fits is quite small.

The *linear* least squares fitting technique is the simplest and most commonly applied form of LINEAR REGRESSION and provides a solution to the problem of finding the best fitting *straight* line through a set of points. In fact, if the functional relationship between the two quantities being graphed is known to within additive or multiplicative constants, it is common practice to transform the data in such a way that the resulting line *is* a straight line, say by plotting T vs. $\sqrt{\ell}$ instead of t vs. ℓ. For this reason, standard forms for EXPONENTIAL, LOGARITHMIC, and POWER laws are often explicitly computed. The formulas for linear least squares fitting were independently derived by Gauss and Legendre.

For NONLINEAR LEAST SQUARES FITTING to a number of unknown parameters, linear least squares fitting may be applied iteratively to a linearized form of the function until convergence is achieved. Depending on the type of fit and initial parameters chosen, the nonlinear fit may have good or poor convergence properties. If uncertainties (in the most general case, error ellipses) are given for the points, points can be weighted differently in order to give the high-quality points more weight.

The residuals of the best-fit line for a set of n points using *unsquared perpendicular* distances d_i of points (x_i, y_i) are given by

$$R_\perp \equiv \sum_{i=1}^{n} d_i. \tag{1}$$

Since the perpendicular distance from a line $y = a + bx$ to point i is given by

$$d_i = \frac{|y_i - (a + bx_i)|}{\sqrt{1+b^2}}, \tag{2}$$

the function to be minimized is

$$R_\perp \equiv \sum_{i=1}^{n} \frac{|y_i - (a + bx_i)|}{\sqrt{1+b^2}}. \tag{3}$$

Unfortunately, because the absolute value function does not have continuous derivatives, minimizing $R_\perp$ is not amenable to analytic solution. However, if the *square* of the perpendicular distances

$$R_\perp^2 \equiv \sum_{i=1}^{n} \frac{[y_i - (a + bx_i)]^2}{1+b^2} \tag{4}$$

is minimized instead, the problem can be solved in closed form. $R_\perp^2$ is a minimum when (suppressing the indices)

$$\frac{\partial R_\perp^2}{\partial a} = \frac{2}{1+b^2} \sum [y - (a + bx)](-1) = 0 \tag{5}$$

and

$$\begin{aligned} \frac{\partial R_\perp^2}{\partial b} &= \frac{2}{1+b^2} \sum [y - (a + bx)](-x) \\ &\quad + \sum \frac{[y - (a + bx)]^2 (-1)(2b)}{(1+b^2)^2} = 0. \end{aligned} \tag{6}$$

The former gives

$$a = \frac{\sum y - b \sum x}{n} = \bar{y} - b\bar{x}, \tag{7}$$

and the latter

$$(1+b^2) \sum [y - (a+bx)]x + b \sum [y - (a+bx)]^2 = 0. \tag{8}$$

But

$$\begin{aligned} [y - (a + bx)]^2 &= y^2 - 2(a + bx)y + (a + bx)^2 \\ &= y^2 - 2ay - 2bxy + a^2 + 2abx + b^2x^2, \end{aligned} \tag{9}$$

so (8) becomes

$$\begin{aligned} &(1+b^2)\left(\sum xy - a \sum x - b \sum x^2\right) \\ &+b\left(\sum y^2 - 2a \sum y - 2b \sum xy + a^2 \sum 1\right. \\ &\left.+2ab \sum x + b^2 \sum x^2\right) = 0 \end{aligned} \tag{10}$$

$$\begin{aligned} &[(1+b^2)(-b) + b(b^2)] \sum x^2 + [(1+b^2) - 2b^2] \sum xy \\ &+b \sum y^2 + [-a(1+b^2) + 2ab^2] \sum x - 2ab \sum y \\ &+ba^2 \sum 1 = 0 \end{aligned} \tag{11}$$

$$\begin{aligned} -b \sum x^2 + (1-b^2) \sum xy + b \sum y^2 + a(b^2 - 1) \sum x \\ -2ab \sum y + ba^2 n = 0. \end{aligned} \tag{12}$$

Plugging (7) into (12) then gives

$$\begin{aligned} &-b \sum x^2 + (1-b^2) \sum xy + b \sum y^2 \\ &+\frac{1}{n}(b^2 - 1)\left(\sum y - b \sum x\right) \sum x \\ &-\frac{2}{n}\left(\sum y - b \sum x\right) b \sum y + \frac{1}{n} b \left(\sum y - b \sum x\right)^2 \\ &= 0 \end{aligned} \tag{13}$$

After a fair bit of algebra, the result is

$$b^2 + \frac{\sum y^2 - \sum x^2 + \frac{1}{n}\left[\left(\sum x\right)^2 - \left(\sum y\right)^2\right]}{\frac{1}{n} \sum x \sum y - \sum xy} b - 1 = 0. \tag{14}$$

So define

$$\begin{aligned} B &\equiv \frac{1}{2} \frac{\left[\sum y^2 - \frac{1}{n}\left(\sum y\right)^2\right] - \left[\sum x^2 - \frac{1}{n}\left(\sum x\right)^2\right]}{\frac{1}{n} \sum x \sum y - \sum xy} \\ &= \frac{1}{2} \frac{\left(\sum y^2 - n\bar{y}^2\right) - \left(\sum x^2 - n\bar{x}^2\right)}{n \sum x \sum y - \sum xy}, \end{aligned} \tag{15}$$

and the QUADRATIC FORMULA gives

$$b = -B \pm \sqrt{B^2 + 1}, \tag{16}$$

with a found using (7). Note the rather unwieldy form of the best-fit parameters in the formulation. In addition, minimizing $R_\perp^2$ for a second- or higher-order POLYNOMIAL leads to polynomial equations having *higher* order, so this formulation cannot be extended.

Vertical least squares fitting proceeds by finding the sum of the *squares* of the *vertical* deviations R^2 of a set of n data points

$$R^2 \equiv \sum [y_i - f(x_i, a_1, a_2, \ldots, a_n)]^2 \tag{17}$$

from a function f. Note that this procedure does *not* minimize the actual deviations from the line (which would be measured perpendicular to the given function). In addition, although the *unsquared* sum of distances might seem a more appropriate quantity to minimize, use of the absolute value results in discontinuous derivatives which cannot be treated analytically. The square deviations from each point are therefore summed, and the resulting residual is then minimized to find the best fit line. This procedure results in outlying points being given disproportionately large weighting.

The condition for R^2 to be a minimum is that

$$\frac{\partial(R^2)}{\partial a_i} = 0 \tag{18}$$

for $i = 1, \ldots, n$. For a linear fit,

$$f(a, b) = a + bx, \tag{19}$$

so

$$R^2(a, b) \equiv \sum_{i=1}^{n} [y_i - (a + bx_i)]^2 \tag{20}$$

$$\frac{\partial(R^2)}{\partial a} = -2 \sum_{i=1}^{n} [y_i - (a + bx_i)] = 0 \tag{21}$$

$$\frac{\partial(R^2)}{\partial b} = -2 \sum_{i=1}^{n} [y_i - (a + bx_i)] x_i = 0. \tag{22}$$

These lead to the equations

$$na + b \sum x = \sum y \tag{23}$$

$$a \sum x + b \sum x^2 = \sum xy, \tag{24}$$

where the subscripts have been dropped for conciseness. In MATRIX form,

$$\begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}, \tag{25}$$

so

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}. \tag{26}$$

The 2×2 MATRIX INVERSE is

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{n \sum x^2 - \left(\sum x\right)^2} \begin{bmatrix} \sum y \sum x^2 - \sum x \sum xy \\ n \sum xy - \sum x \sum y \end{bmatrix}, \tag{27}$$

so

$$a = \frac{\sum y \sum x^2 - \sum x \sum xy}{n \sum x^2 - \left(\sum x\right)^2} \tag{28}$$

$$= \frac{\bar{y} \sum x^2 - \bar{x} \sum xy}{\sum x^2 - n\bar{x}^2} \tag{29}$$

$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - \left(\sum x\right)^2} \tag{30}$$

$$= \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2} \tag{31}$$

(Kenney and Keeping 1962). These can be rewritten in a simpler form by defining the sums of squares

$$\mathrm{ss}_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \left(\sum x^2 - n\bar{x}^2\right) \tag{32}$$

$$\mathrm{ss}_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \left(\sum y^2 - n\bar{y}^2\right) \tag{33}$$

$$\mathrm{ss}_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \left(\sum xy - n\bar{x}\bar{y}\right), \tag{34}$$

which are also written as

$$\sigma_x^2 = \mathrm{ss}_{xx} \tag{35}$$

$$\sigma_y^2 = \mathrm{ss}_{yy} \tag{36}$$

$$\mathrm{cov}(x, y) = \mathrm{ss}_{xy}. \tag{37}$$

Here, $\mathrm{cov}(x, y)$ is the COVARIANCE and σ_x^2 and σ_y^2 are variances. Note that the quantities $\sum xy$ and $\sum x^2$ can also be interpreted as the DOT PRODUCTS

$$\sum x^2 = \mathbf{x} \cdot \mathbf{x} \tag{38}$$

$$\sum xy = \mathbf{x} \cdot \mathbf{y}. \tag{39}$$

In terms of the sums of squares, the REGRESSION COEFFICIENT b is given by

$$b = \frac{\mathrm{cov}(x, y)}{{\sigma_x}^2} = \frac{\mathrm{ss}_{xy}}{\mathrm{ss}_{xx}}, \tag{40}$$

and a is given in terms of b using (24) as

$$a = \bar{y} - b\bar{x}. \tag{41}$$

The overall quality of the fit is then parameterized in terms of a quantity known as the CORRELATION COEFFICIENT, defined by

$$r^2 = \frac{{\mathrm{ss}_{xy}}^2}{\mathrm{ss}_{xx}\mathrm{ss}_{yy}}, \tag{42}$$

which gives the proportion of ss_{yy} which is accounted for by the regression.

The STANDARD ERRORS for a and b are

$$\mathrm{SE}(a) = s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\mathrm{ss}_{xx}}} \tag{43}$$

$$\mathrm{SE}(b) = \frac{s}{\sqrt{\mathrm{ss}_{xx}}}. \tag{44}$$

Let $\hat{y}_i$ be the vertical coordinate of the best-fit line with x-coordinate x_i, so

$$\hat{y}_i \equiv a + bx_i, \tag{45}$$

then the error between the actual vertical point y_i and the fitted point is given by

$$e_i \equiv y_i - \hat{y}_i. \tag{46}$$

Now define s^2 as an estimator for the variance in e_i,

$$s^2 = \sum_{i=1}^{n} \frac{{e_i}^2}{n-2}. \tag{47}$$

Then s can be given by

$$s = \sqrt{\frac{\mathrm{ss}_{yy} - b\,\mathrm{ss}_{xy}}{n-2}} = \sqrt{\frac{\mathrm{ss}_{yy} - \frac{s_{xy}^2}{s_{xx}}}{n-2}} \tag{48}$$

(Acton 1966, pp. 32–35; Gonick and Smith 1993, pp. 202–204).

Generalizing from a straight line (i.e., first degree polynomial) to a kth degree POLYNOMIAL

$$y = a_0 + a_1 x + \ldots + a_k x^k, \tag{49}$$

the residual is given by

$$R^2 \equiv \sum_{i=1}^{n} [y_i - (a_0 + a_1 x_i + \ldots + a_k {x_i}^k)]^2. \tag{50}$$

The PARTIAL DERIVATIVES (again dropping superscripts) are

$$\frac{\partial(R^2)}{\partial a_0} = -2\sum [y - (a_0 + a_1 x + \ldots + a_k x^k)] = 0 \tag{51}$$

$$\frac{\partial(R^2)}{\partial a_1} = -2\sum [y - (a_0 + a_1 x + \ldots + a_k x^k)]x = 0 \tag{52}$$

$$\frac{\partial(R^2)}{\partial a_k} = -2\sum [y - (a_0 + a_1 x + \ldots + a_k x^k)]x^k = 0. \tag{53}$$

These lead to the equations

$$a_0 n + a_1 \sum x + \ldots + a_k \sum x^k = \sum y \tag{54}$$

$$a_0 \sum x + a_1 \sum x^2 + \ldots + a_k \sum x^{k+1} = \sum xy \tag{55}$$

$$a_0 \sum x^k + a_1 \sum x^{k+1} + \ldots + a_k \sum x^{2k} = \sum x^k y \tag{56}$$

or, in MATRIX form

$$\begin{bmatrix} n & \sum x & \cdots & \sum x^k \\ \sum x & \sum x^2 & \cdots & \sum x^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum x^k & \sum x^{k+1} & \cdots & \sum x^{2k} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum xy \\ \vdots \\ \sum x^k y \end{bmatrix}. \tag{57}$$

This is a VANDERMONDE MATRIX. We can also obtain the MATRIX for a least squares fit by writing

$$\begin{bmatrix} 1 & x_1 & \cdots & {x_1}^k \\ 1 & x_2 & \cdots & {x_2}^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & {x_n}^k \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}. \tag{58}$$

Premultiplying both sides by the TRANSPOSE of the first MATRIX then gives

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ {x_1}^k & {x_2}^k & \cdots & {x_n}^k \end{bmatrix} \begin{bmatrix} 1 & x_1 & \cdots & {x_1}^k \\ 1 & x_2 & \cdots & {x_2}^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & {x_n}^k \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ {x_1}^k & {x_2}^k & \cdots & {x_n}^k \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \tag{59}$$

so

$$\begin{bmatrix} n & \sum x & \cdots & \sum x^n \\ \sum x & \sum x^2 & \cdots & \sum x^{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum x^n & \sum x^{n+1} & \cdots & \sum x^{2n} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum xy \\ \vdots \\ \sum x^k y \end{bmatrix}. \tag{60}$$

As before, given m points (x_i, y_i) and fitting with POLYNOMIAL COEFFICIENTS $a_0, \ldots, a_n$ gives

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} 1 & x_1 & {x_1}^2 & \cdots & {x_1}^n \\ 1 & x_2 & {x_2}^2 & \cdots & {x_2}^n \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & {x_m}^2 & \cdots & {x_m}^n \end{bmatrix} \begin{bmatrix} a_0 \\ a_0 \\ \vdots \\ a_n \end{bmatrix}, \tag{61}$$

In MATRIX notation, the equation for a polynomial fit is given by

$$\mathsf{y} = \mathsf{Xa}. \tag{62}$$

This can be solved by premultiplying by the MATRIX TRANSPOSE X^T,

$$\mathsf{X}^\mathrm{T}\mathsf{y} = \mathsf{X}^\mathrm{T}\mathsf{Xa}. \tag{63}$$

This MATRIX EQUATION can be solved numerically, or can be inverted directly if it is well formed, to yield the solution vector

$$\mathbf{a} = (\mathsf{X}^{\mathrm{T}}\mathsf{X})^{-1}\mathsf{X}^{\mathrm{T}}\mathbf{y}. \qquad (64)$$

Setting $m = 1$ in the above equations reproduces the linear solution.

see also CORRELATION COEFFICIENT, INTERPOLATION, LEAST SQUARES FITTING—EXPONENTIAL, LEAST SQUARES FITTING—LOGARITHMIC, LEAST SQUARES FITTING—POWER LAW, MOORE-PENROSE GENERALIZED MATRIX INVERSE, NONLINEAR LEAST SQUARES FITTING, REGRESSION COEFFICIENT, SPLINE

References

Acton, F. S. *Analysis of Straight-Line Data.* New York: Dover, 1966.

Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences.* New York: McGraw-Hill, 1969.

Gonick, L. and Smith, W. *The Cartoon Guide to Statistics.* New York: Harper Perennial, 1993.

Kenney, J. F. and Keeping, E. S. "Linear Regression, Simple Correlation, and Contingency." Ch. 8 in *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, pp. 199–237, 1951.

Kenney, J. F. and Keeping, E. S. "Linear Regression and Correlation." Ch. 15 in *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, pp. 252–285, 1962.

Lancaster, P. and Šalkauskas, K. *Curve and Surface Fitting: An Introduction.* London: Academic Press, 1986.

Lawson, C. and Hanson, R. *Solving Least Squares Problems.* Englewood Cliffs, NJ: Prentice-Hall, 1974.

Nash, J. C. *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation, 2nd ed.* Bristol, England: Adam Hilger, pp. 21–24, 1990.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Fitting Data to a Straight Line" "Straight-Line Data with Errors in Both Coordinates," and "General Linear Least Squares." §15.2, 15.3, and 15.4 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 655–675, 1992.

York, D. "Least-Square Fitting of a Straight Line." *Canad. J. Phys.* **44**, 1079–1086, 1966.

Least Squares Fitting—Exponential

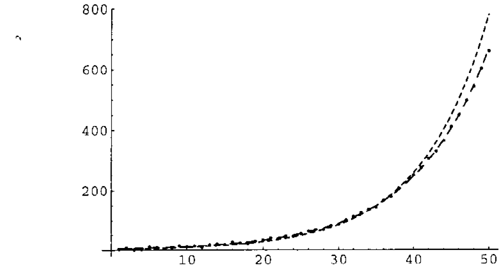

To fit a functional form

$$y = Ae^{Bx}, \qquad (1)$$

take the LOGARITHM of both sides

$$\ln y = \ln A + B\ln x. \qquad (2)$$

The best-fit values are then

$$a = \frac{\sum \ln y \sum x^2 - \sum x \sum x\ln y}{n\sum x^2 - \left(\sum x\right)^2} \qquad (3)$$

$$b = \frac{n\sum x\ln y - \sum x \sum \ln y}{n\sum x^2 - \left(\sum x\right)^2}, \qquad (4)$$

where $B \equiv b$ and $A \equiv \exp(a)$.

This fit gives greater weights to small y values so, in order to weight the points equally, it is often better to minimize the function

$$\sum y(\ln y - a - bx)^2. \qquad (5)$$

Applying LEAST SQUARES FITTING gives

$$a\sum y + b\sum xy = \sum y\ln y \qquad (6)$$

$$a\sum xy + b\sum x^2 y = \sum xy\ln y \qquad (7)$$

$$\begin{bmatrix} \sum y & \sum xy \\ \sum xy & \sum x^2 y \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum y\ln y \\ \sum xy\ln y \end{bmatrix}. \qquad (8)$$

Solving for a and b,

$$a = \frac{\sum(x^2 y)\sum(y\ln y) - \sum(xy)\sum(xy\ln y)}{\sum y\sum(x^2 y) - \left(\sum xy\right)^2} \qquad (9)$$

$$b = \frac{\sum y\sum(xy\ln y) - \sum(xy)\sum(y\ln y)}{\sum y\sum(x^2 y) - \left(\sum xy\right)^2}. \qquad (10)$$

In the plot above, the short-dashed curve is the fit computed from (3) and (4) and the long-dashed curve is the fit computed from (9) and (10).

see also LEAST SQUARES FITTING, LEAST SQUARES FITTING—LOGARITHMIC, LEAST SQUARES FITTING—POWER LAW

Least Squares Fitting—Logarithmic

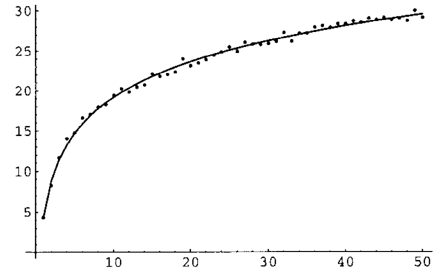

Given a function of the form

$$y = a + b\ln x, \qquad (1)$$

the COEFFICIENTS can be found from LEAST SQUARES FITTING as

$$b = \frac{n\sum(y\ln x) - \sum y\sum(\ln x)}{n\sum[(\ln x)^2] - \left[\sum(\ln x)\right)^2} \qquad (2)$$

$$a = \frac{\sum y - b\sum(\ln x)}{n}. \qquad (3)$$

see also LEAST SQUARES FITTING, LEAST SQUARES FITTING—EXPONENTIAL, LEAST SQUARES FITTING—POWER LAW

Least Squares Fitting—Power Law

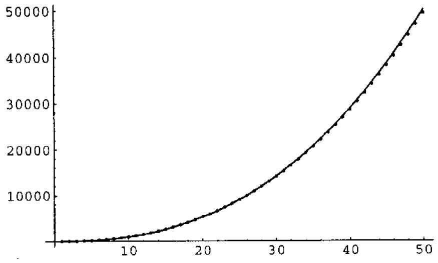

Given a function of the form

$$y = Ax^B, \tag{1}$$

LEAST SQUARES FITTING gives the COEFFICIENTS as

$$b = \frac{n\sum(\ln x \ln y) - \sum(\ln x)\sum(\ln y)}{n\sum[(\ln x)^2] - \left(\sum \ln x\right)^2} \tag{2}$$

$$a = \frac{\sum(\ln y) - b\sum(\ln x)}{n}, \tag{3}$$

where $B \equiv b$ and $A \equiv \exp(a)$.

see also LEAST SQUARES FITTING, LEAST SQUARES FITTING—EXPONENTIAL, LEAST SQUARES FITTING—LOGARITHMIC

Least Upper Bound

see SUPREMUM

Lebesgue Constants (Fourier Series)

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Assume a function f is integrable over the interval $[-\pi, \pi]$ and $S_n(f, x)$ is the nth partial sum of the FOURIER SERIES of f, so that

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos(kt)\,dt \tag{1}$$

$$b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin(kt)\,dt \tag{2}$$

and

$$S_n(f,x) = \tfrac{1}{2}a_0 + \left\{\sum_{k=1}^{n}[a_k\cos(kx) + b_k\sin(kx)]\right\}. \tag{3}$$

If

$$|f(x)| \leq 1 \tag{4}$$

for all x, then

$$S_n(f,x) \leq \frac{1}{\pi}\int_0^{\pi} \frac{|\sin[\frac{1}{2}(2n+1)\theta]|}{\sin(\frac{1}{2}\theta)}\,d\theta = L_n, \tag{5}$$

and L_n is the smallest possible constant for which this holds for all continuous f. The first few values of L_n are

$$L_0 = 1 \tag{6}$$

$$L_1 = \frac{1}{3} + \frac{2\sqrt{3}}{\pi} = 1.435991124\ldots \tag{7}$$

$$L_2 = 1.642188435\ldots \tag{8}$$

$$L_3 = 1.778322862. \tag{9}$$

Some FORMULAS for L_n include

$$\begin{aligned} L_n &= \frac{1}{2n+1} + \frac{2}{\pi}\sum_{k=1}^{n}\frac{1}{k}\tan\left(\frac{\pi k}{2n+1}\right) \\ &= \frac{16}{\pi^2}\sum_{k=1}^{\infty}\sum_{j=1}^{(2n+1)k}\frac{1}{4k^2-1}\frac{1}{2j-1} \end{aligned} \tag{10}$$

(Zygmund 1959) and integral FORMULAS include

$$\begin{aligned} L_n &= 4\int_0^{\infty}\frac{\tanh[(2n+1)x]}{\tanh x}\frac{dx}{\pi^2+4x^2} \\ &= \frac{4}{\pi^2}\int_0^{\infty}\frac{\sinh[(2n+1)x]}{\sinh x}\ln\{\coth[\tfrac{1}{2}(2n+1)x]\}\,dx \end{aligned} \tag{11}$$

(Hardy 1942). For large n,

$$\frac{4}{\pi^2}\ln n < L_n < 3 + \frac{4}{\pi^2}\ln n. \tag{12}$$

This result can be generalized for an r-differentiable function satisfying

$$\left|\frac{d^r f}{dx^r}\right| \leq 1 \tag{13}$$

for all x. In this case,

$$|f(x) - S_n(f,x)| \leq L_{n,r} = \frac{4}{\pi^2}\frac{\ln n}{n^r} + \mathcal{O}\left(\frac{1}{n^r}\right), \tag{14}$$

where

$$L_{n,r} = \begin{cases} \frac{1}{\pi}\int_{-\pi}^{\pi}\left|\sum_{k=n+1}^{\infty}\frac{\sin(kx)}{k^r}\right|\,dx & \text{for } r \geq 1 \text{ odd} \\ \frac{1}{\pi}\int_{-\pi}^{\pi}\left|\sum_{k=n+1}^{\infty}\frac{\cos(kx)}{k^r}\right|\,dx & \text{for } r \geq 1 \text{ even} \end{cases} \tag{15}$$

(Kolmogorov 1935, Zygmund 1959).

Watson (1930) showed that

$$\lim_{n\to\infty}\left[L_n - \frac{4}{\pi^2}\ln(2n+1)\right] = c, \tag{16}$$

where

$$c = \frac{8}{\pi^2}\left(\sum_{k=1}^{\infty}\frac{\ln k}{4k^2-1}\right) - \frac{4}{\pi^2}\frac{\Gamma'(\frac{1}{2})}{\Gamma(\frac{1}{2})} \tag{17}$$

$$= \frac{8}{\pi^2}\left[\sum_{j=0}^{\infty}\frac{\lambda(2j+2)-1}{2j+1}\right] + \frac{4}{\pi^2}(2\ln 2+\gamma) \tag{18}$$

$$= 0.9894312738\ldots, \tag{19}$$

where $\Gamma(z)$ is the Gamma Function, $\lambda(z)$ is the Dirichlet Lambda Function, and γ is the Euler-Mascheroni Constant.

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/lbsg/lbsg.html`.

Hardy, G. H. "Note on Lebesgue's Constants in the Theory of Fourier Series." *J. London Math. Soc.* **17**, 4–13, 1942.

Kolmogorov, A. N. "Zur Grössenordnung des Restgliedes Fourierscher reihen differenzierbarer Funktionen." *Ann. Math.* **36**, 521–526, 1935.

Watson, G. N. "The Constants of Landau and Lebesgue." *Quart. J. Math. Oxford* **1**, 310–318, 1930.

Zygmund, A. G. *Trigonometric Series, 2nd ed., Vols. 1-2.* Cambridge, England: Cambridge University Press, 1959.

Lebesgue Constants (Lagrange Interpolation)

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Define the nth Lebesgue constant for the Lagrange Interpolating Polynomial by

$$\Lambda_n(X) \equiv \max_{-1\leq x\leq 1}\sum_{k=1}^{n}\left|\prod_{j\neq k}\frac{x-x_j}{x_k-x_j}\right|. \tag{1}$$

It is true that

$$\Lambda_n > \frac{4}{\pi^2}\ln n - 1. \tag{2}$$

The efficiency of a Lagrange interpolation is related to the rate at which Λ_n increases. Erdős (1961) proved that there exists a Positive constant such that

$$\Lambda_n > \frac{2}{\pi}\ln n - C \tag{3}$$

for all n. Erdős (1961) further showed that

$$\Lambda_n < \frac{2}{\pi}\ln n + 4, \tag{4}$$

so (3) cannot be improved upon.

References

Erdős, P. "Problems and Results on the Theory of Interpolation, II." *Acta Math. Acad. Sci. Hungary* **12**, 235–244, 1961.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/lbsg/lbsg.html`.

Lebesgue Covering Dimension

An important Dimension and one of the first dimensions investigated. It is defined in terms of covering sets, and is therefore also called the Covering Dimension. Another name for the Lebesgue covering dimension is the Topological Dimension.

A Space has Lebesgue covering dimension m if for every open Cover of that space, there is an open Cover that refines it such that the refinement has order at most $m+1$. Consider how many elements of the cover contain a given point in a base space. If this has a maximum over all the points in the base space, then this maximum is called the order of the cover. If a Space does not have Lebesgue covering dimension m for any m, it is said to be infinite dimensional.

Results of this definition are:

1. Two homeomorphic spaces have the same dimension,
2. $\mathbb{R}^n$ has dimension n,
3. A Topological Space can be embedded as a closed subspace of a Euclidean Space Iff it is locally compact, Hausdorff, second countable, and is finite dimensional (in the sense of the Lebesgue Dimension), and
4. Every compact metrizable m-dimensional Topological Space can be embedded in $\mathbb{R}^{2m+1}$.

see also Lebesgue Minimal Problem

References

Dieudonne, J. A. *A History of Algebraic and Differential Topology.* Boston, MA: Birkhäuser, 1994.

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 414, 1980.

Munkres, J. R. *Topology: A First Course.* Englewood Cliffs, NJ: Prentice-Hall, 1975.

Lebesgue Dimension

see Lebesgue Covering Dimension

Lebesgue Integrable

A real-valued function f defined on the reals $\mathbb{R}$ is called Lebesgue integrable if there exists a Sequence of Step Functions $\{f_n\}$ such that the following two conditions are satisfied:

1. $\sum_{n=1}^{\infty}\int|f_n| < \infty$,
2. $f(x) = \sum_{n=1}^{\infty}$ for every $x \in \mathbb{R}$ such that $\sum_{n=1}^{\infty}\int|f_n| < \infty$.

Here, the above integral denotes the ordinary Riemann Integral. Note that this definition avoids explicit use of the Lebesgue Measure.

see also Integral, Lebesgue Integral, Riemann Integral, Step Function

Lebesgue Integral

The LEBESGUE INTEGRAL is defined in terms of upper and lower bounds using the LEBESGUE MEASURE of a SET. It uses a LEBESGUE SUM $S_n = \eta_i \mu(E_i)$ where η_i is the value of the function in subinterval i, and $\mu(E_i)$ is the LEBESGUE MEASURE of the SET E_i of points for which values are approximately η_i. This type of integral covers a wider class of functions than does the RIEMANN INTEGRAL.

see also A-INTEGRABLE, COMPLETE FUNCTIONS, INTEGRAL

References
Kestelman, H. "Lebesgue Integral of a Non-Negative Function" and "Lebesgue Integrals of Functions Which Are Sometimes Negative." Chs. 5–6 in *Modern Theories of Integration, 2nd rev. ed.* New York: Dover, pp. 113–160, 1960.

Lebesgue Measurability Problem

A problem related to the CONTINUUM HYPOTHESIS which was solved by Solovay (1970) using the INACCESSIBLE CARDINALS AXIOM. It has been proven by Shelah and Woodin (1990) that use of this AXIOM is essential to the proof.

see also CONTINUUM HYPOTHESIS, INACCESSIBLE CARDINALS AXIOM, LEBESGUE MEASURE

References
Shelah, S. and Woodin, H. "Large Cardinals Imply that Every Reasonable Definable Set of Reals is Lebesgue Measurable." *Israel J. Math.* **70**, 381–394, 1990.
Solovay, R. M. "A Model of Set-Theory in which Every Set of Reals is Lebesgue Measurable." *Ann. Math.* **92**, 1–56, 1970.

Lebesgue Measure

An extension of the classical notions of length and AREA to more complicated sets. Given an open set $S \equiv \sum_k (a_k, b_k)$ containing DISJOINT intervals,

$$\mu_L(S) \equiv \sum_k (b_k - a_k).$$

Given a CLOSED SET $S' \equiv [a, b] - \sum_k (a_k, b_k)$,

$$\mu_L(S') \equiv (b - a) - \sum_k (b_k - a_k).$$

A LINE SEGMENT has Lebesgue measure 1; the CANTOR SET has Lebesgue measure 0. The MINKOWSKI MEASURE of a bounded, CLOSED SET is the same as its Lebesgue measure (Ko 1995).

see also CANTOR SET, MEASURE, RIESZ-FISCHER THEOREM

References
Kestelman, H. "Lebesgue Measure." Ch. 3 in *Modern Theories of Integration, 2nd rev. ed.* New York: Dover, pp. 67–91, 1960.
Ko, K.-I. "A Polynomial-Time Computable Curve whose Interior has a Nonrecursive Measure." *Theoret. Comput. Sci.* **145**, 241–270, 1995.

Lebesgue Minimal Problem

Find the plane LAMINA of least AREA A which is capable of covering any plane figure of unit GENERAL DIAMETER. A UNIT CIRCLE is too small, but a HEXAGON circumscribed on the UNIT CIRCLE is too large. More specifically, the AREA is bounded by

$$0.8257\ldots = \tfrac{1}{8}\pi + \tfrac{1}{4}\sqrt{3} < A < \tfrac{2}{3}(3 - \sqrt{3}) = 0.8454\ldots$$

(Pal 1920).

see also AREA, BORSUK'S CONJECTURE, DIAMETER (GENERAL), KAKEYA NEEDLE PROBLEM

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 99, 1987.
Coxeter, H. S. M. "Lebesgue's Minimal Problem." *Eureka* **21**, 13, 1958.
Grünbaum, B. "Borsuk's Problem and Related Questions." *Proc. Sympos. Pure Math, Vol. 7.* Providence, RI: Amer. Math. Soc., pp. 271–284, 1963.
Kakeya, S. "Some Problems on Maxima and Minima Regarding Ovals." *Sci. Reports Tôhoku Imperial Univ., Ser. 1 (Math., Phys., Chem.)* **6**, 71–88, 1917.
Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 142–144, 1990.
Pál, J. *Danske videnkabernes selskab, Copenhagen Math.-fys. maddelelser* **3**, 1–35, 1920.
Yaglom, I. M. and Boltyanskii, V. G. *Convex Figures.* New York: Holt, Rinehart, & Winston, pp. 18 and 100, 1961.

Lebesgue-Radon Integral

see LEBESGUE-STIELTJES INTEGRAL

Lebesgue Singular Integrals

$$\mathcal{U}_n(f) = \int_a^b f(x) K_n(x)\, dx,$$

where $\{K_n(x)\}$ is a SEQUENCE of CONTINUOUS FUNCTIONS.

Lebesgue-Stieltjes Integral

Let $\alpha(x)$ be a monotone increasing function and define an INTERVAL $I = (x_1, x_2)$. Then define the NONNEGATIVE function

$$U(I) = \alpha(x_2 + 0) - \alpha(x_1 + 0).$$

The LEBESGUE INTEGRAL with respect to a MEASURE constructed using $U(I)$ is called the Lebesgue-Stieltjes integral, or sometimes the LEBESGUE-RADON INTEGRAL.

References
Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 326, 1980.

Lebesgue Sum

$$S_n = \eta_i \mu(E_i),$$

where $\mu(E_i)$ is the MEASURE of the SET E_i of points on the x-axis for which $f(x) \approx \eta_i$.

Leech Lattice

A 24-D Euclidean lattice. An AUTOMORPHISM of the Leech lattice modulo a center of two leads to the CONWAY GROUP Co_1. Stabilization of the 1- and 2-D sublattices leads to the CONWAY GROUPS Co_2 and Co_3, the HIGMAN-SIMS GROUP *HS* and the MCLAUGHLIN GROUP *McL*.

The Leech lattice appears to be the densest HYPERSPHERE PACKING in 24-D, and results in each HYPERSPHERE touching 195,560 others.

see also BARNES-WALL LATTICE, CONWAY GROUPS, COXETER-TODD LATTICE, HIGMAN-SIMS GROUP, HYPERSPHERE, HYPERSPHERE PACKING, KISSING NUMBER, MCLAUGHLIN GROUP

References

Conway, J. H. and Sloane, N. J. A. "The 24-Dimensional Leech Lattice Λ_{24}," "A Characterization of the Leech Lattice," "The Covering Radius of the Leech Lattice," "Twenty-Three Constructions for the Leech Lattice," "The Cellular of the Leech Lattice," "Lorentzian Forms for the Leech Lattice." §4.11, Ch. 12, and Chs. 23–26 in *Sphere Packings, Lattices, and Groups, 2nd ed.* New York: Springer-Verlag, pp. 131–135, 331–336, and 478–526, 1993.

Leech, J. "Notes on Sphere Packings." *Canad. J. Math.* **19**, 251–267, 1967.

Wilson, R. A. "Vector Stabilizers and Subgroups of Leech Lattice Groups." *J. Algebra* **127**, 387–408, 1989.

Lefshetz Fixed Point Formula

see LEFSHETZ TRACE FORMULA

Lefshetz's Theorem

Each DOUBLE POINT assigned to an irreducible curve whose GENUS is NONNEGATIVE imposes exactly one condition.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 104, 1959.

Lefshetz Trace Formula

A formula which counts the number of FIXED POINTS for a topological transformation.

Leg

The leg of a TRIANGLE is one of its sides.

see also HYPOTENUSE, TRIANGLE

Legendre Addition Theorem

see SPHERICAL HARMONIC ADDITION THEOREM

Legendre's Chi-Function

The function defined by

$$\chi_\nu(z) = \sum_{k=0}^\infty \frac{z^{2k+1}}{(2k+1)^\nu}$$

for integral $\nu = 2, 3, \ldots$. It is related to the POLYLOGARITHM by

$$\begin{aligned}\chi_\nu(z) &= \tfrac{1}{2}[\operatorname{Li}_\nu(z) - \operatorname{Li}_\nu(-z)]\\ &= \operatorname{Li}_\nu(z) - 2^{-\nu}\operatorname{Li}_\nu(z^2).\end{aligned}$$

see also POLYLOGARITHM

References

Cvijović, D. and Klinowski, J. "Closed-Form Summation of Some Trigonometric Series." *Math. Comput.* **64**, 205–210, 1995.

Lewin, L. *Polylogarithms and Associated Functions.* Amsterdam, Netherlands: North-Holland, pp. 282–283, 1981.

Legendre's Constant

The number 1.08366 in Legendre's guess at the PRIME NUMBER THEOREM

$$\pi(n) \sim \frac{n}{\ln n - 1.08366}.$$

This expression is correct to leading term only.

References

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 147, 1983.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 28–29, 1991.

Legendre Differential Equation

The second-order ORDINARY DIFFERENTIAL EQUATION

$$(1-x^2)\frac{d^2y}{dx^2} - 2x\frac{dy}{dx} + l(l+1)y = 0, \tag{1}$$

which can be rewritten

$$\frac{d}{dx}\left[(1-x^2)\frac{dy}{dx}\right] + l(l+1)y = 0. \tag{2}$$

The above form is a special case of the associated Legendre differential equation with $m = 0$. The Legendre differential equation has REGULAR SINGULAR POINTS at -1, 1, and ∞. It can be solved using a series expansion,

$$y = \sum_{n=0}^\infty a_n x^n \tag{3}$$

$$y' = \sum_{n=0}^\infty n a_n x^{n-1} \tag{4}$$

$$y'' = \sum_{n=0}^\infty n(n-1) a_n x^{n-2}. \tag{5}$$

Plugging in,

$$(1-x^2)\sum_{n=0}^\infty n(n-1)a_n x^{n-2} - 2x\sum_{n=0}^\infty n a_n x^{n-1} + l(l+1)\sum_{n=0}^\infty a_n x^n = 0 \tag{6}$$

$$\sum_{n=0}^{\infty} n(n-1)a_n x^{n-2} - \sum_{n=0}^{\infty} n(n-1)a_n x^n$$
$$-2x\sum_{n=0}^{\infty} n a_n x^{n-1} + l(l+1)\sum_{n=0}^{\infty} a_n x^n = 0 \quad (7)$$

$$\sum_{n=2}^{\infty} n(n-1)a_n x^{n-2} - \sum_{n=0}^{\infty} n(n-1)a_n x^n$$
$$-2\sum_{n=0}^{\infty} n a_n x^n + l(l+1)\sum_{n=0}^{\infty} a_n x^n = 0 \quad (8)$$

$$\sum_{n=0}^{\infty} (n+2)(n+1)a_{n+2} x^n - \sum_{n=0}^{\infty} n(n-1)a_n x^n$$
$$-2\sum_{n=0}^{\infty} n a_n x^n + l(l+1)\sum_{n=0}^{\infty} a_n x^n = 0 \quad (9)$$

$$\sum_{n=0}^{\infty} \{(n+1)(n+2)a_{n+2}$$
$$+[-n(n-1) - 2n + l(l+1)]a_n\} = 0, \quad (10)$$

so each term must vanish and

$$(n+1)(n+2)a_{n+2} - n(n+1) + l(l+1)]a_n = 0 \quad (11)$$

$$\begin{aligned} a_{n+2} &= \frac{n(n+1) - l(l+1)}{(n+1)(n+2)} a_n \\ &= -\frac{[l+(n+1)](l-n)}{(n+1)(n+2)} a_n. \end{aligned} \quad (12)$$

Therefore,

$$a_2 = -\frac{l(l+1)}{1\cdot 2} a_0 \quad (13)$$

$$\begin{aligned} a_4 &= -\frac{(l-2)(l+3)}{3\cdot 4} a_2 \\ &= (-1)^2 \frac{[(l-2)l][(l+1)(l+3)]}{1\cdot 2\cdot 3\cdot 4} a_0 \end{aligned} \quad (14)$$

$$\begin{aligned} a_6 &= -\frac{(l-4)(l+5)}{5\cdot 6} a_4 \\ &= (-1)^3 \frac{[(l-4)(l-2)l][(l+1)(l+3)(l+5)]}{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6} a_0, \end{aligned} \quad (15)$$

so the Even solution is

$$y_1(x) = 1 + \sum_{n=1}^{\infty} (-1)^n \frac{[(l-2n+2)\cdots(l-2)l][(l+1)(l+3)\cdots(l+2n-1)]}{(2n)!} x^{2n}. \quad (16)$$

Similarly, the Odd solution is

$$y_2(x) = x + \sum_{n=1}^{\infty} (-1)^n \times \frac{[(l-2n+1)\cdots(l-3)(l-1)][(l+2)(l+4)\cdots(l+2n)]}{(2n+1)!} x^{2m+1}. \quad (17)$$

If l is an Even Integer, the series y_1 reduces to a Polynomial of degree l with only Even Powers of x and the series y_2 diverges. If l is an Odd Integer, the series y_2 reduces to a Polynomial of degree l with only Odd Powers of x and the series y_1 diverges. The general solution for an Integer l is given by the Legendre Polynomials

$$P_n(x) = c_n \begin{cases} y_1(x) & \text{for } l \text{ even} \\ y_2(x) & \text{for } l \text{ odd}, \end{cases} \quad (18)$$

where c_n is chosen so that $P_n(1) = 1$. If the variable x is replaced by $\cos\theta$, then the Legendre differential equation becomes

$$\frac{d^2y}{d\theta^2} + \frac{\cos\theta}{\sin\theta}\frac{dy}{dx} + l(l+1)y = 0, \quad (19)$$

as is derived for the associated Legendre differential equation with $m = 0$.

The *associated* Legendre differential equation is

$$\frac{d}{dx}\left[(1-x^2)\frac{dy}{dx}\right] + \left[l(l+1) - \frac{m^2}{1-x^2}\right] y = 0 \quad (20)$$

$$(1-x^2)\frac{d^2y}{dx^2} - 2x\frac{dy}{dx} + \left[l(l+1) - \frac{m^2}{1-x^2}\right] y = 0. \quad (21)$$

The solutions to this equation are called the associated Legendre polynomials. Writing $x \equiv \cos\theta$, first establish the identities

$$\frac{dy}{dx} = \frac{dy}{d(\cos\theta)} = -\frac{1}{\sin\theta}\frac{dy}{d\theta} \quad (22)$$

$$x\frac{dy}{dx} = -\frac{\cos\theta}{\sin\theta}\frac{dy}{d\theta}, \quad (23)$$

$$\begin{aligned} \frac{d^2y}{dx^2} &= \frac{1}{\sin\theta}\frac{d}{d\theta}\left(\frac{1}{\sin\theta}\frac{dy}{d\theta}\right) \\ &= \frac{1}{\sin\theta}\left(\frac{-\cos\theta}{\sin^2\theta}\right)\frac{dy}{d\theta} + \frac{1}{\sin^2\theta}\frac{d^2y}{d\theta^2}, \end{aligned} \quad (24)$$

and

$$1 - x^2 = 1 - \cos^2\theta = \sin^2\theta. \quad (25)$$

Therefore,

$$(1-x^2)\frac{d^2y}{dx^2} = \sin^2\theta\frac{1}{\sin\theta}\left(\frac{-\cos\theta}{\sin^2\theta}\right)\frac{dy}{d\theta} + \frac{1}{\sin^2\theta}\frac{d^2y}{d\theta^2}$$
$$= \frac{d^2y}{d\theta^2} - \frac{\cos\theta}{\sin\theta}\frac{dy}{d\theta}. \qquad (26)$$

Plugging (22) into (26) and the result back into (21) gives

$$\left(\frac{d^2y}{d\theta^2} - \frac{\cos\theta}{\sin\theta}\frac{dy}{d\theta}\right)$$
$$+2\frac{\cos\theta}{\sin\theta}\frac{dy}{d\theta} + \left[l(l+1) - \frac{m^2}{\sin^2\theta}\right]y = 0 \qquad (27)$$

$$\frac{d^2y}{d\theta^2} + \frac{\cos\theta}{\sin\theta}\frac{dy}{dx} + \left[l(l+1) - \frac{m^2}{\sin^2\theta}\right]y = 0. \qquad (28)$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 332, 1972.

Legendre Duplication Formula

Gamma Functions of argument $2z$ can be expressed in terms of Gamma Functions of smaller arguments. From the definition of the Beta Function,

$$B(m,n) = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)} = \int_0^1 u^{m-1}(1-u)^{n-1}\,du. \qquad (1)$$

Now, let $m = n \equiv z$, then

$$\frac{\Gamma(z)\Gamma(z)}{\Gamma(2z)} = \int_0^1 u^{z-1}(1-u)^{z-1}\,du \qquad (2)$$

and $u \equiv (1+x)/2$, so $du = dx/2$ and

$$\frac{\Gamma(z)\Gamma(z)}{\Gamma(2z)} = \int_0^1 \left(\frac{1+x}{2}\right)^{z-1}\left(1-\frac{1+x}{2}\right)^{z-1}\left(\tfrac{1}{2}\,dx\right)$$
$$= \frac{1}{2}\int_0^1 \left(\frac{1+x}{2}\right)^{z-1}\left(\frac{1-x}{2}\right)^{z-1} dx$$
$$= \frac{1}{2^{1+2(z-1)}}\int_0^1 (1-x^2)^{z-1}\,dx$$
$$= 2^{1-2z}\int_0^1 (1-x^2)^{z-1}\,dx. \qquad (3)$$

Now, use the Beta Function identity

$$B(m,n) = 2\int_0^1 x^{2z-1}(1-x^2)^{z-1}\,dx \qquad (4)$$

to write the above as

$$\frac{\Gamma(z)\Gamma(z)}{\Gamma(2z)} = 2^{1-2z}B(\tfrac{1}{2},z) = 2^{1-2z}\frac{\Gamma(\frac{1}{2})\Gamma(z)}{\Gamma(z+\frac{1}{2})}. \qquad (5)$$

Solving for $\Gamma(2z)$,

$$\Gamma(2z) = \frac{\Gamma(z)\Gamma(z+\frac{1}{2})2^{2z-1}}{\Gamma(\frac{1}{2})} = \frac{\Gamma(z)\Gamma(z+\frac{1}{2})2^{2z-1}}{\sqrt{\pi}}$$
$$= (2\pi)^{-1/2}2^{2z-1/2}\Gamma(z)\Gamma(z+\tfrac{1}{2}), \qquad (6)$$

since $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

see also Gamma Function, Gauss Multiplication Formula

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 256, 1972.

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 561–562, 1985.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 424–425, 1953.

Legendre's Factorization Method

A Prime Factorization Algorithm in which a sequence of Trial Divisors is chosen using a Quadratic Sieve. By using Quadratic Residues of N, the Quadratic Residues of the factors can also be found.

see also Prime Factorization Algorithms, Quadratic Residue, Quadratic Sieve Factorization Method, Trial Divisor

Legendre's Formula

Counts the number of Positive Integers less than or equal to a number x which are not divisible by any of the first a Primes,

$$\phi(x,a) = \lfloor x\rfloor - \sum\left\lfloor\frac{x}{p_i}\right\rfloor + \sum\left\lfloor\frac{x}{p_ip_j}\right\rfloor$$
$$-\sum\left\lfloor\frac{x}{p_ip_jp_k}\right\rfloor + \ldots, \qquad (1)$$

where $\lfloor x\rfloor$ is the Floor Function. Taking $a = x$ gives

$$\phi(x,x) = \pi(x) - \pi(\sqrt{x}) + 1 = \lfloor x\rfloor - \sum_{p_i\leq\sqrt{x}}\left\lfloor\frac{x}{p_i}\right\rfloor$$
$$+ \sum_{p_i<p_j\leq\sqrt{x}}\left\lfloor\frac{x}{p_ip_j}\right\rfloor - \sum_{p_i<p_j<p_k\leq\sqrt{x}}\left\lfloor\frac{x}{p_ip_jp_k}\right\rfloor + \ldots,$$
$$(2)$$

where $\pi(n)$ is the Prime Counting Function. Legendre's formula holds since one more than the number of Primes in a range equals the number of Integers minus the number of composites in the interval.

Legendre's formula satisfies the Recurrence Relation

$$\phi(x,a) = \phi(x,a-1) - \phi\left(\frac{x}{p_a},a-1\right). \qquad (3)$$

Let $m_k \equiv p_1 p_2 \cdots p_k$, then

$$\begin{aligned}\phi(m_k, k) &= \lfloor m_k \rfloor - \sum \left\lfloor \frac{m_k}{p_i} \right\rfloor + \sum \left\lfloor \frac{m_k}{p_i p_j} \right\rfloor - \ldots \\ &= m_k - \sum \frac{m_k}{p_i} + \sum \frac{m_k}{p_i p_j} - \ldots \\ &= m_k \left(1 - \frac{1}{p-1}\right)\left(1 - \frac{1}{p_2}\right) \cdots \left(1 - \frac{1}{p_k}\right) \\ &= \prod_{i=1}^{k} (p_i - 1) = \phi(m_k), \qquad (4)\end{aligned}$$

where $\phi(n)$ is the TOTIENT FUNCTION, and

$$\phi(s m_k + t, k) = s\phi(m_k) + \phi(t, k), \qquad (5)$$

where $0 \leq t \leq m_k$. If $t > m_k/2$, then

$$\phi(t, k) = \phi(m_k) - \phi(m_k - t - 1, k). \qquad (6)$$

Note that $\phi(n, n)$ is not practical for computing $\pi(n)$ for large arguments. A more efficient modification is MEISSEL'S FORMULA.

see also LEHMER'S FORMULA, MAPES' METHOD, MEISSEL'S FORMULA, PRIME COUNTING FUNCTION

Legendre Function of the First Kind

see LEGENDRE POLYNOMIAL

Legendre Function of the Second Kind

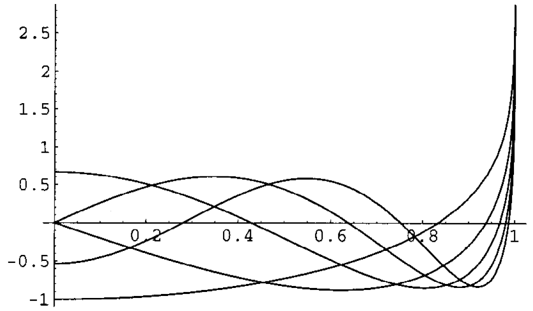

A solution to the LEGENDRE DIFFERENTIAL EQUATION which is singular at the origin. The Legendre functions of the second kind satisfy the same RECURRENCE RELATION as the LEGENDRE FUNCTIONS OF THE FIRST KIND. The first few are

$$\begin{aligned}Q_0 &= \tfrac{1}{2} \ln\left(\frac{1+x}{1-x}\right) \\ Q_1 &= \frac{x}{2} \ln\left(\frac{1+x}{1-x}\right) - 1 \\ Q_2 &= \frac{3x^2 - 1}{4} \ln\left(\frac{1+x}{1-x}\right) - \frac{3x}{2} \\ Q_3 &= \frac{5x^3 - 3x}{4} \ln\left(\frac{1+x}{1-x}\right) - \frac{5x^2}{2} + \frac{2}{3}.\end{aligned}$$

The associated Legendre functions of the second kind have DERIVATIVE about 0 of

$$\left[\frac{dQ_\nu^\mu(x)}{dx}\right]_{x=0} = \frac{2^\mu \sqrt{\pi}\,\cos[\tfrac{1}{2}\pi(\nu+\mu)]\Gamma(\tfrac{1}{2}\nu + \tfrac{1}{2}\mu + 1)}{\Gamma(\tfrac{1}{2}\nu - \tfrac{1}{2}\mu + \tfrac{1}{2})}$$

(Abramowitz and Stegun 1972, p. 334). The logarithmic derivative is

$$\begin{aligned}&\left[\frac{d \ln Q_\lambda^\mu(z)}{dz}\right]_{z=0} \\ &\quad = 2\exp\{\tfrac{1}{2}\pi i \operatorname{sgn}(\Im[z])\}\frac{[\tfrac{1}{2}(\lambda+\mu)]![\tfrac{1}{2}(\lambda-\mu)]!}{[\tfrac{1}{2}(\lambda+\mu-1)]![\tfrac{1}{2}(\lambda-\mu-1)]!}\end{aligned}$$

(Binney and Tremaine 1987, p. 654).

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Legendre Functions." Ch. 8 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 331–339, 1972.

Arfken, G. "Legendre Functions of the Second Kind, $Q_n(x)$." *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 701–707, 1985.

Binney, J. and Tremaine, S. "Associated Legendre Functions." Appendix 5 in *Galactic Dynamics.* Princeton, NJ: Princeton University Press, pp. 654–655, 1987.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 597–600, 1953.

Snow, C. *Hypergeometric and Legendre Functions with Applications to Integral Equations of Potential Theory.* Washington, DC: U. S. Government Printing Office, 1952.

Spanier, J. and Oldham, K. B. "The Legendre Functions $P_\nu(x)$ and $Q_\nu(x)$." Ch. 59 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 581–597, 1987.

Legendre-Gauss Quadrature

Also called "the" GAUSSIAN QUADRATURE or LEGENDRE QUADRATURE. A GAUSSIAN QUADRATURE over the interval $[-1, 1]$ with WEIGHTING FUNCTION $W(x) = 1$. The ABSCISSAS for quadrature order n are given by the roots of the LEGENDRE POLYNOMIALS $P_n(x)$, which occur symmetrically about 0. The weights are

$$w_i = -\frac{A_{n+1}\gamma_n}{A_n P_n'(x_i) P_{n+1}(x_i)} = \frac{A_n}{A_{n-1}} \frac{\gamma_{n-1}}{P_{n-1}(x_i) P_n'(x_i)}, \qquad (1)$$

where A_n is the COEFFICIENT of x^n in $P_n(x)$. For LEGENDRE POLYNOMIALS,

$$A_n = \frac{(2n)!}{2^n (n!)^2}, \qquad (2)$$

so

$$\begin{aligned}\frac{A_{n+1}}{A_n} &= \frac{[2(n+1)]!}{2^{n+1}[(n+1)!]^2} \frac{2^n (n!)^2}{(2n)!} \\ &= \frac{(2n+1)(2n+2)}{2(n+1)^2} = \frac{2n+1}{n+1}. \qquad (3)\end{aligned}$$

Additionally,

$$\gamma_n = \frac{2}{2n+1}, \qquad (4)$$

so

$$w_i = -\frac{2}{(n+1)P_{n+1}(x_i)P_n'(x_i)} = \frac{2}{nP_{n-1}(x_i)P_n'(x_i)}. \tag{5}$$

Using the RECURRENCE RELATION

$$\begin{aligned}(1-x^2)P_n'(x) &= nxP_n(x) + nP_{n-1}(x) & (6)\\ &= (n+1)xP_n(x) - (n+1)P_{n+1}(x) & (7)\end{aligned}$$

gives

$$w_i = \frac{2}{(1-x_i{}^2)[P_n'(x_i)]^2} = \frac{2(1-x_i{}^2)}{(n+1)^2[P_{n+1}(x_i)]^2}. \tag{8}$$

The error term is

$$E = \frac{2^{2n+1}(n!)^4}{(2n+1)[(2n)!]^3} f^{(2n)}(\xi). \tag{9}$$

Beyer (1987) gives a table of ABSCISSAS and weights up to $n = 16$, and Chandrasekhar (1960) up to $n = 8$ for n EVEN.

n	x_i	w_i
2	±0.57735	1.000000
3	0	0.888889
	±0.774597	0.555556
4	±0.339981	0.652145
	±0.861136	0.347855
5	0	0.568889
	±0.538469	0.478629
	±0.90618	0.236927

The ABSCISSAS and weights can be computed analytically for small n.

n	x_i	w_i
2	$\pm\frac{1}{3}\sqrt{3}$	1
3	0	$\frac{8}{9}$
	$\pm\frac{1}{5}\sqrt{15}$	$\frac{5}{9}$
4	$\pm\sqrt{\frac{3-2\sqrt{\frac{6}{5}}}{7}}$	
	$\pm\sqrt{\frac{3+2\sqrt{\frac{6}{5}}}{7}}$	

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 462–463, 1987.

Chandrasekhar, S. *Radiative Transfer.* New York: Dover, pp. 56–62, 1960.

Hildebrand, F. B. *Introduction to Numerical Analysis.* New York: McGraw-Hill, pp. 323–325, 1956.

Legendre-Jacobi Elliptic Integral

Any of the three standard forms in which an ELLIPTIC INTEGRAL can be expressed.

see also ELLIPTIC INTEGRAL OF THE FIRST KIND, ELLIPTIC INTEGRAL OF THE SECOND KIND, ELLIPTIC INTEGRAL OF THE THIRD KIND

Legendre Polynomial

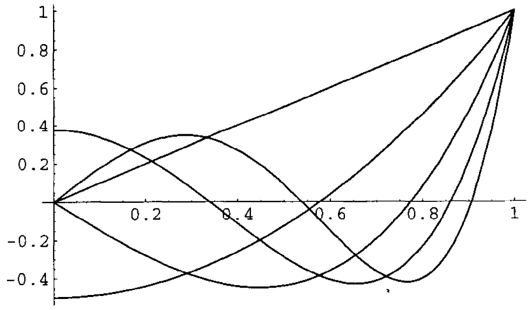

The LEGENDRE FUNCTIONS OF THE FIRST KIND are solutions to the LEGENDRE DIFFERENTIAL EQUATION. If l is an INTEGER, they are POLYNOMIALS. They are a special case of the ULTRASPHERICAL FUNCTIONS with $\alpha = 1/2$. The Legendre polynomials $P_n(x)$ are illustrated above for $x \in [0,1]$ and $n = 1, 2, \ldots, 5$.

The Rodrigues FORMULA provides the GENERATING FUNCTION

$$P_l(x) = \frac{1}{2^l l!}\frac{dl}{dx^l}(x^2-1)^l, \tag{1}$$

which yields upon expansion

$$P_l(x) = \frac{1}{2^l}\sum_{k=0}^{\lfloor n/2 \rfloor} \frac{(-1)^k(2l-2k)!}{k!(l-k)!(l-2k)!}x^{l-2k}, \tag{2}$$

where $\lfloor r \rfloor$ is the FLOOR FUNCTION. The GENERATING FUNCTION is

$$g(t,x) = (1-2xt+t^2)^{-1/2} = \sum_{n=0}^{\infty} P_n(x)t^n. \tag{3}$$

Take $\partial g/\partial t$,

$$-\tfrac{1}{2}(1-2xt+t^2)^{-3/2}(-2x+2t) = \sum_{n=0}^{\infty} nP_n(x)t^{n-1}. \tag{4}$$

Multiply (4) by $2t$,

$$-t(1-2xt+t^2)^{-3/2}(-2x+2t) = \sum_{n=0}^{\infty} 2nP_n(x)t^n \tag{5}$$

and add (3) and (5),

$$(1-2xt+t^2)^{-3/2}[(2xt-2t^2)+(1-2xt+t^2)] = \sum_{n=0}^{\infty}(2n+1)P_n(x)t^n \tag{6}$$

$$(1-2xt+t^2)^{-3/2}(1-t^2) = \sum_{n=0}^{\infty}(2n+1)P_n(x)t^n. \tag{7}$$

This expansion is useful in some physical problems, including expanding the Heyney-Greenstein phase function and computing the charge distribution on a Sphere. They satisfy the Recurrence Relation

$$(l+1)P_{l+1}(x)-(2l+1)xP_l(x)+lP_{l-1}(x)=0. \qquad (8)$$

The Legendre polynomials are orthogonal over $(-1,1)$ with Weighting Function 1 and satisfy

$$\int_{-1}^{1} P_n(x)P_m(x)\,dx=\frac{2}{2n+1}\delta_{mn}, \qquad (9)$$

where δ_{mn} is the Kronecker Delta.

A Complex Generating Function is

$$P_l(x)=\frac{1}{2\pi i}\int(1-2zx+z^2)^{-1/2}z^{-l-1}\,dz, \qquad (10)$$

and the Schläfli integral is

$$P_l(x)=\frac{(-1)^l}{2^l}\frac{1}{2\pi i}\int\frac{(1-z^2)^l}{(z-x)^{l+1}}\,dz. \qquad (11)$$

Additional integrals (Byerly 1959, p. 172) include

$$\int_0^1 P_m(x)\,dx = \begin{cases} 0 & m \text{ even} \neq 0 \\ (-1)^{(m-1)/2}\frac{m!!}{m(m+1)(m-1)!!} & m \text{ odd} \end{cases} \qquad (12)$$

$$\int_0^1 P_m(x)P_n(x)\,dx = \begin{cases} 0 \\ \quad m,n \text{ both even or odd } m\neq n \\ (-1)^{(m+n+1)/2}\frac{m!n!}{2^{m+n+1}(m-n)(m+n+1)(\frac{1}{2}m)!\{[\frac{1}{2}(n-1)]!\}^2} \\ \quad m \text{ even, } n \text{ odd} \\ \frac{1}{2n+1}, \\ \quad m=n. \end{cases} \qquad (13)$$

An additional identity is

$$1-[P_n(x)]^2=\sum_{\nu=1}^{n}\frac{1-x^2}{1-{x_\nu}^2}\left[\frac{P_n(x)}{P_n'(x_\nu)(x-x_\nu)}\right]^2 \qquad (14)$$

(Szegő 1975, p. 348).

The first few Legendre polynomials are

$$\begin{aligned} P_0(x)&=1\\ P_1(x)&=x\\ P_2(x)&=\tfrac{1}{2}(3x^2-1)\\ P_3(x)&=\tfrac{1}{2}(5x^3-3x)\\ P_4(x)&=\tfrac{1}{8}(35x^4-30x^2+3)\\ P_5(x)&=\tfrac{1}{8}(63x^5-70x^3+15x)\\ P_6(x)&=\tfrac{1}{16}(231x^6-315x^4+105x^2-5). \end{aligned}$$

The first few Powers in terms of Legendre polynomials are

$$\begin{aligned} x&=P_1\\ x^2&=\tfrac{1}{3}(P_0+2P_2)\\ x^3&=\tfrac{1}{5}(3P_1+2P_3)\\ x^4&=\tfrac{1}{35}(7P_0+20P_2+8P_4)\\ x^5&=\tfrac{1}{63}(27P_1+28P_3+8P_5)\\ x^6&=\tfrac{1}{231}(33P_0+110P_2+72P_4+16P_5). \end{aligned}$$

For Legendre polynomials and Powers up to exponent 12, see Abramowitz and Stegun (1972, p. 798).

The Legendre Polynomials can also be generated using Gram-Schmidt Orthonormalization in the Open Interval $(-1,1)$ with the Weighting Function 1.

$$P_0(x)=1 \qquad (15)$$

$$\begin{aligned} P_1(x)&=\left[x-\frac{\int_{-1}^{1}x\,dx}{\int_{-1}^{1}dx}\right]\cdot 1\\ &=x-\frac{\frac{1}{2}[x^2]_{-1}^{1}}{[x]_{-1}^{1}}=x-\frac{\frac{1}{2}(1-1)}{1-(-1)}=x \qquad (16) \end{aligned}$$

$$\begin{aligned} P_2(x)&=\left[x-\frac{\int_{-1}^{1}x^3\,dx}{\int_{-1}^{1}x^2\,dx}\right]-\left[\frac{\int_{-1}^{1}x^2\,dx}{\int_{-1}^{1}dx}\right]\cdot 1\\ &=\left[x-\frac{\frac{1}{4}[x^4]_{-1}^{1}}{\frac{1}{3}[x^3]_{-1}^{1}}\right]x-\frac{\frac{1}{3}[x^3]_{-1}^{1}}{[x]_{-1}^{1}}=x^2-\tfrac{1}{3} \qquad (17) \end{aligned}$$

$$\begin{aligned} P_3(x)&=\left[x-\frac{\int_{-1}^{1}x(x^2-\frac{1}{3})^2\,dx}{\int_{-1}^{1}(x^2-\frac{1}{3})^2\,dx}\right](x^2-\tfrac{1}{3})\\ &\quad-\left[\frac{\int_{-1}^{1}(x^2-\frac{1}{3})^2\,dx}{\int_{-1}^{1}x^2\,dx}\right]x\\ &=x\left[x^2-\tfrac{1}{3}-\frac{(\frac{1}{5}-\frac{2}{9}+\frac{1}{9})x}{\frac{1}{3}}\right]\\ &=x^3-\tfrac{1}{3}x-3(\tfrac{1}{5}-\tfrac{1}{9})\\ &=x^3-x\left(\tfrac{1}{3}+\tfrac{3}{5}-\tfrac{1}{3}\right)=x^3-\tfrac{3}{5}x. \qquad (18) \end{aligned}$$

Normalizing so that $P_n(1)=1$ gives the expected Legendre polynomials.

The "shifted" Legendre polynomials are a set of functions analogous to the Legendre polynomials, but defined on the interval (0, 1). They obey the Orthogonality relationship

$$\int_0^1 \bar{P}_m(x)\bar{P}_n(x)\,dx=\frac{1}{2n+1}\delta_{mn}. \qquad (19)$$

The first few are

$$\begin{aligned} \bar{P}_0(x)&=1\\ \bar{P}_1(x)&=2x-1\\ \bar{P}_2(x)&=6x^2-6x+1\\ \bar{P}_3(x)&=20x^3-30x^2+12x-1. \end{aligned}$$

The associated Legendre polynomials $P_l^m(x)$ are solutions to the associated LEGENDRE DIFFERENTIAL EQUATION, where l is a POSITIVE INTEGER and $m = 0, \ldots, l$. They can be given in terms of the unassociated polynomials by

$$\begin{aligned} P_l^m(x) &= (-1)^m(1-x^2)^{m/2}\frac{d^m}{dx^m}P_l(x) \\ &= \frac{(-1)^m}{2^l l!}(1-x^2)^{m/2}\frac{d^{l+m}}{dx^{l+m}}(x^2-1)^l, \end{aligned} \tag{20}$$

where $P_l(x)$ are the unassociated LEGENDRE POLYNOMIALS. Note that some authors (e.g., Arfken 1985, p. 668) omit the CONDON-SHORTLEY PHASE $(-1)^m$, while others include it (e.g., Abramowitz and Stegun 1972, Press *et al.* 1992, and the `LegendreP[l,m,z]` command of *Mathematica®*). Abramowitz and Stegun (1972, p. 332) use the notation

$$P_{lm}(x) \equiv (-1)^m P_m^l(x) \tag{21}$$

to distinguish these two cases.

Associated polynomials are sometimes called FERRERS' FUNCTIONS (Sansone 1991, p. 246). If $m = 0$, they reduce to the unassociated POLYNOMIALS. The associated Legendre functions are part of the SPHERICAL HARMONICS, which are the solution of LAPLACE'S EQUATION in SPHERICAL COORDINATES. They are ORTHOGONAL over $[-1, 1]$ with the WEIGHTING FUNCTION 1

$$\int_{-1}^{1} P_l^m(x)P_{l'}^m(x)\,dx = \frac{2}{2l+1}\frac{(l+m)!}{(l-m)!}\delta_{ll'}, \tag{22}$$

ORTHOGONAL over $[-1, 1]$ with respect to m with the WEIGHTING FUNCTION $(1-x^2)^{-2}$

$$\int_{-1}^{1} P_l^m(x)P_l^{m'}(x)\frac{dx}{1-x^2} = \frac{(l+m)!}{m(l-m)!}\delta_{mm'}. \tag{23}$$

They obey the RECURRENCE RELATIONS

$$(l-m)P_l^m(x) = x(2l-1)P_{l-1}^m(x) - (l+m-1)P_{l-2}^m(x) \tag{24}$$

$$\begin{aligned} \frac{dP_l^m}{d\theta} &= -\sqrt{1-\mu^2}\,\frac{dP_l^m}{d\mu} \\ &= \tfrac{1}{2}(l-m+1)(l+m+P_l^{m-1}-P_l^{m+1}) \end{aligned} \tag{25}$$

$$(2l+1)\mu P_l^m = (l+m)P_{l-1}^m + (l-m+1)P_{l+1}^m \tag{26}$$

$$(2l+1)\sqrt{1-\mu^2}\,P_l^m = P_{l+1}^{m+1} - P_{l-1}^{m+1}. \tag{27}$$

An identity relating associated POLYNOMIALS with NEGATIVE m to the corresponding functions with POSITIVE m is

$$P_l^{-m} = (-1)^m\frac{(l-m)!}{(l+m)!}P_l^m. \tag{28}$$

Additional identities are

$$P_l^l(x) = (-1)^l(2l-1)!!(1-x^2)^{l/2} \tag{29}$$

$$P_{l+1}^l(x) = x(2l+1)P_l^l(x). \tag{30}$$

Written in terms of x and using the convention without a leading factor of $(-1)^m$ (Arfken 1985, p. 669), the first few associated Legendre polynomials are

$$\begin{aligned} P_0^0(x) &= 1 \\ P_1^0(x) &= x \\ P_1^1(x) &= -(1-x^2)^{1/2} \\ P_2^0(x) &= \tfrac{1}{2}(3x^2-1) \\ P_2^1(x) &= -3x(1-x^2)^{1/2} \\ P_2^2(x) &= 3(1-x^2) \\ P_3^0(x) &= \tfrac{1}{2}x(5x^2-3) \\ P_3^1(x) &= \tfrac{3}{2}(1-5x^2)(1-x^2)^{1/2} \\ P_3^2(x) &= 15x(1-x^2) \\ P_3^3(x) &= -15(1-x^2)^{3/2} \\ P_4^0(x) &= \tfrac{1}{8}(35x^4-30x^2+3) \\ P_4^1(x) &= \tfrac{5}{2}x(3-7x^2)(1-x^2)^{1/2} \\ P_4^2(x) &= \tfrac{15}{2}(7x^2-1)(1-x^2) \\ P_4^3(x) &= -105x(1-x^2)^{3/2} \\ P_4^4(x) &= 105(1-x^2)^2 \\ P_5^0(x) &= \tfrac{1}{8}x(63x^4-70x^2+15). \end{aligned}$$

Written in terms $x \equiv \cos\theta$, the first few become

$$\begin{aligned} P_0^0(\cos\theta) &= 1 \\ P_1^{-1}(\cos\theta) &= \tfrac{1}{2}\sin\theta \\ P_1^0(\cos\theta) &= \cos\theta = \mu \\ P_1^1(\cos\theta) &= \sin\theta \\ P_2^{-2}(\cos\theta) &= \tfrac{1}{8}\sin^2\theta \\ P_2^{-1}(\cos\theta) &= \tfrac{1}{2}\sin\theta\cos\theta \\ P_2^0(\cos\theta) &= \tfrac{1}{2}(3\cos^2\theta-1) \\ P_2^1(\cos\theta) &= 3\sin\theta\cos\theta \\ &= \tfrac{3}{2}\sin^2\theta \\ P_2^2(\cos\theta) &= 3\sin^2\theta \\ &= \tfrac{3}{2}(1-\cos^2\theta) \\ P_3^0(\cos\theta) &= \tfrac{1}{2}\cos\theta(5\cos^2\theta-3) \\ &= \tfrac{1}{2}\cos\theta(2-5\sin^2\theta) \\ P_3^1(\cos\theta) &= \tfrac{3}{2}(5\cos^2\theta-1)\sin\theta \\ &= \tfrac{3}{8}(\sin\theta+5\sin^3\theta). \end{aligned}$$

The derivative about the origin is

$$\left[\frac{dP_\nu^\mu(x)}{dx}\right]_{x=0} = \frac{2^{\mu+1}\sin[\frac{1}{2}\pi(\nu+\mu)]\Gamma(\frac{1}{2}\nu+\frac{1}{2}\mu+1)}{\pi^{1/2}\Gamma(\frac{1}{2}\nu-\frac{1}{2}\mu+\frac{1}{2})} \tag{31}$$

(Abramowitz and Stegun 1972, p. 334), and the logarithmic derivative is

$$\left[\frac{d\ln P_\lambda^\mu(z)}{dz}\right]_{z=0} = 2\tan[\tfrac{1}{2}\pi(\lambda+\mu)]\frac{[\frac{1}{2}(\lambda+\mu)]![\frac{1}{2}(\lambda-\mu)]!}{[\frac{1}{2}(\lambda+\mu-1)]![\frac{1}{2}(\lambda-\mu-1)]!}. \tag{32}$$

(Binney and Tremaine 1987, p. 654).

see also CONDON-SHORTLEY PHASE, CONICAL FUNCTION, GEGENBAUER POLYNOMIAL, KINGS PROBLEM, LAPLACE'S INTEGRAL, LAPLACE-MEHLER INTEGRAL, SUPER CATALAN NUMBER, TOROIDAL FUNCTION, TURÁN'S INEQUALITIES

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Legendre Functions" and "Orthogonal Polynomials." Ch. 22 in Chs. 8 and 22 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 331–339 and 771–802, 1972.

Arfken, G. "Legendre Functions." Ch. 12 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 637–711, 1985.

Binney, J. and Tremaine, S. "Associated Legendre Functions." Appendix 5 in *Galactic Dynamics.* Princeton, NJ: Princeton University Press, pp. 654–655, 1987.

Byerly, W. E. *An Elementary Treatise on Fourier's Series, and Spherical, Cylindrical, and Ellipsoidal Harmonics, with Applications to Problems in Mathematical Physics.* New York: Dover, 1959.

Iyanaga, S. and Kawada, Y. (Eds.). "Legendre Function" and "Associated Legendre Function." Appendix A, Tables 18.II and 18.III in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 1462–1468, 1980.

Legendre, A. M. "Sur l'attraction des Sphéroides." *Mém. Math. et Phys. présentés à l'Ac. r. des. sc. par divers savants* **10**, 1785.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 593–597, 1953.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 252, 1992.

Sansone, G. "Expansions in Series of Legendre Polynomials and Spherical Harmonics." Ch. 3 in *Orthogonal Functions, rev. English ed.* New York: Dover, pp. 169–294, 1991.

Snow, C. *Hypergeometric and Legendre Functions with Applications to Integral Equations of Potential Theory.* Washington, DC: U. S. Government Printing Office, 1952.

Spanier, J. and Oldham, K. B. "The Legendre Polynomials $P_n(x)$" and "The Legendre Functions $P_\nu(x)$ and $Q_\nu(x)$." Chs. 21 and 59 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 183–192 and 581–597, 1987.

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., 1975.

Legendre Polynomial of the Second Kind

see LEGENDRE FUNCTION OF THE SECOND KIND

Legendre's Quadratic Reciprocity Law

see QUADRATIC RECIPROCITY LAW

Legendre Quadrature

see LEGENDRE-GAUSS QUADRATURE

Legendre Relation

Let $E(k)$ and $K(k)$ be complete ELLIPTIC INTEGRALS OF THE FIRST and SECOND KINDS, with $E'(k)$ and $K'(k)$ the complementary integrals. Then

$$E(k)K'(k) + E'(k)K(k) - K(k)K'(k) = \tfrac{1}{2}\pi.$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 591, 1972.

Legendre Series

Because the LEGENDRE FUNCTIONS OF THE FIRST KIND form a COMPLETE ORTHOGONAL BASIS, any FUNCTION may be expanded in terms of them

$$f(x) = \sum_{n=0}^\infty a_n P_n(x). \tag{1}$$

Now, multiply both sides by $P_m(x)$ and integrate

$$\int_{-1}^1 P_m(x)f(x)\,dx = \sum_{n=0}^\infty a_n \int_{-1}^1 P_n(x)P_m(x)\,dx. \tag{2}$$

But

$$\int_{-1}^1 P_n(x)P_m(x)\,dx = \frac{2}{2m+1}\delta_{mn}, \tag{3}$$

where δ_{mn} is the KRONECKER DELTA, so

$$\int_{-1}^1 P_m(x)f(x)\,dx = \sum_{n=0}^\infty a_n \frac{2}{2m+1}\delta_{mn} = \frac{2}{2m+1}a_m \tag{4}$$

and

$$a_m = \frac{2m+1}{2}\int_{-1}^1 P_m(x)f(x)\,dx. \tag{5}$$

see also FOURIER SERIES, JACKSON'S THEOREM, LEGENDRE POLYNOMIAL, MACLAURIN SERIES, PICONE'S THEOREM, TAYLOR SERIES

Legendre Sum

see LEGENDRE'S FORMULA

Legendre Symbol

$$\left(\frac{m}{n}\right) = (m|n)$$
$$\equiv \begin{cases} 0 & \text{if } m|n \\ 1 & \text{if } n \text{ is a quadratic residue modulo } m \\ -1 & \text{if } n \text{ is a quadratic nonresidue modulo } m. \end{cases}$$

If m is an ODD PRIME, then the JACOBI SYMBOL reduces to the Legendre symbol. The Legendre symbol obeys $(ab|p) = (a|p)(b|p)$.

$$\left(\frac{3}{p}\right) = \begin{cases} 1 & \text{if } p \equiv \pm 1 \pmod{12} \\ -1 & \text{if } p \equiv \pm 5 \pmod{12}. \end{cases}$$

see also JACOBI SYMBOL, KRONECKER SYMBOL, QUADRATIC RECIPROCITY THEOREM

References

Guy, R. K. "Quadratic Residues. Schur's Conjecture." §F5 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 244–245, 1994.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 33–34 and 40–42, 1993.

Legendre Transformation

Given a function of two variables

$$df = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy \equiv u\,dx + v\,dy, \qquad (1)$$

change the differentials from dx and dy to du and dy with the transformation

$$g \equiv f - ux \qquad (2)$$

$$\begin{aligned} dg &= df - u\,dx - x\,du = u\,dx + v\,dy - u\,dx - x\,du \\ &= v\,dy - x\,du. \end{aligned} \qquad (3)$$

Then

$$x \equiv -\frac{\partial g}{\partial u} \qquad (4)$$

$$v \equiv \frac{\partial g}{\partial y}. \qquad (5)$$

Lehmer's Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Lehmer (1938) showed that every POSITIVE IRRATIONAL NUMBER x has a unique infinite continued cotangent representation of the form

$$x = \cot\left[\sum_{k=0}^{\infty}(-1)^k \cot^{-1} b_k\right],$$

where the b_ks are NONNEGATIVE and

$$b_k \geq (b_{k-1})^2 + b_{k-1} + 1.$$

The case for which the convergence is slowest occurs when the inequality is replaced by equality, giving $c_0 = 0$ and

$$c_k = (c_{k-1})^2 + c_{k-1} + 1$$

for $k \geq 1$. The first few values are c_k are 0, 1, 3, 13, 183, 33673, ... (Sloane's A024556), resulting in the constant

$$\begin{aligned} \xi &= \cot(\cot^{-1} 0 - \cot^{-1} 1 + \cot^{-1} 3 - \cot^{-1} 13 \\ &\quad + \cot^{-1} 183 - \cot^{-1} 33673 + \cot^{-1} 1133904603 \\ &\quad - \cot^{-1} 1285739649838492213 + \ldots + (-1)^k c_k + \ldots) \\ &= \cot(\tfrac{1}{4}\pi + \cot^{-1} 3 - \cot^{-1} 13 \\ &\quad + \cot^{-1} 183 - \cot^{-1} 33673 + \cot^{-1} 1133904603 \\ &\quad - \cot^{-1} 1285739649838492213 + \ldots + (-1)^k c_k + \ldots) \\ &= 0.59263271\ldots \end{aligned}$$

(Sloane's A030125). ξ is not an ALGEBRAIC NUMBER of degree less than 4, but Lehmer's approach cannot show whether or not ξ is TRANSCENDENTAL.

see also ALGEBRAIC NUMBER, TRANSCENDENTAL NUMBER

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/lehmer/lehmer.html`.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 29, 1983.

Lehmer, D. H. "A Cotangent Analogue of Continued Fractions." *Duke Math. J.* **4**, 323–340, 1938.

Plouffe, S. "The Lehmer Constant." `http://lacim.uqam.ca/piDATA/lehmer.txt`.

Sloane, N. J. A. Sequences A024556 and A030125 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Lehmer's Formula

A FORMULA related to MEISSEL'S FORMULA.

$$\begin{aligned} \pi(x) &= \lfloor x \rfloor - \sum_{i=1}^{a}\left\lfloor \frac{x}{p_i} \right\rfloor + \sum_{1\leq i\leq j\leq a}\left\lfloor \frac{x}{p_i p_j} \right\rfloor - \ldots \\ &\quad + \tfrac{1}{2}(b+a-2)(b-a+1) - \sum_{a<i\leq b}\pi\left(\frac{x}{p_i}\right) \\ &\quad - \sum_{i=a+1}^{c}\sum_{j=i}^{b_i}\left[\pi\left(\frac{x}{p_i p_j}\right) - (j-1)\right], \end{aligned}$$

where

$$\begin{aligned} a &\equiv \pi(x^{1/4}) \\ b &\equiv \pi(x^{1/2}) \\ b_i &\equiv \pi(\sqrt{x/p_i}) \\ c &\equiv \pi(x^{1/3}), \end{aligned}$$

and $\pi(n)$ is the PRIME COUNTING FUNCTION.

References

Riesel, H. "Lehmer's Formula." *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, pp. 13–14, 1994.

Lehmer Method

see LEHMER-SCHUR METHOD

Lehmer Number

A number generated by a generalization of a LUCAS SEQUENCE. Let α and β be COMPLEX NUMBERS with

$$\alpha+\beta=\sqrt{R} \tag{1}$$

$$\alpha\beta=Q, \tag{2}$$

where Q and R are RELATIVELY PRIME NONZERO INTEGERS and α/β is a ROOT OF UNITY. Then the Lehmer numbers are

$$U_n(\sqrt{R},Q)=\frac{\alpha^n-\beta^n}{\alpha-\beta}, \tag{3}$$

and the companion numbers

$$V_n(\sqrt{R},Q)=\begin{cases}\frac{\alpha^n+\beta^n}{\alpha+\beta} & \text{for } n \text{ odd}\\ \alpha^n+\beta^n & \text{for } n \text{ even.}\end{cases} \tag{4}$$

References

Lehmer, D. H. "An Extended Theory of Lucas' Functions." *Ann. Math.* **31**, 419–448, 1930.

Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, pp. 61 and 70, 1989.

Williams, H. C. "The Primality of $N = 2A3^n - 1$." *Canad. Math. Bull.* **15**, 585–589, 1972.

Lehmer's Phenomenon

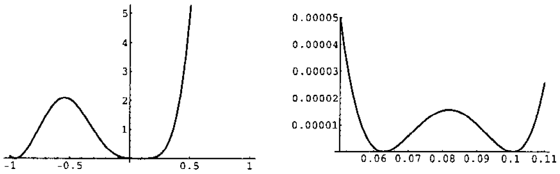

The appearance of nontrivial zeros (i.e., those along the CRITICAL STRIP with $\Re[z]=1/2$) of the RIEMANN ZETA FUNCTION $\zeta(z)$ very close together. An example is the pair of zeros $\zeta(\frac{1}{2}+(7005+t)i)$ given by $t_1 \approx 0.0606918$ and $t_2 \approx 0.100055$, illustrated above in the plot of $|\zeta(\frac{1}{2}+(7005+t)i)|^2$.

see also CRITICAL STRIP, RIEMANN ZETA FUNCTION

References

Csordas, G.; Odlyzko, A. M.; Smith, W.; and Varga, R. S. "A New Lehmer Pair of Zeros and a New Lower Bound for the de Bruijn-Newman Constant." *Elec. Trans. Numer. Analysis* **1**, 104–111, 1993.

Csordas, G.; Smith, W.; and Varga, R. S. "Lehmer Pairs of Zeros, the de Bruijn-Newman Constant and the Riemann Hypothesis." *Constr. Approx.* **10**, 107–129, 1994.

Csordas, G.; Smith, W.; and Varga, R. S. "Lehmer Pairs of Zeros and the Riemann ξ-Function." In *Mathematics of Computation 1943–1993: A Half-Century of Computational Mathematics* (Vancouver, BC, 1993). *Proc. Sympos. Appl. Math.* **48**, 553–556, 1994.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 357–358, 1991.

Lehmer's Problem

Do there exist any COMPOSITE NUMBERS n such that $\phi(n)|(n-1)$? No such numbers are known. In 1932, Lehmer showed that such an n must be ODD and SQUAREFREE, and that the number of distinct PRIME factors $d(7) \geq 7$. This was subsequently extended to $d(n) \geq 11$. The best current results are $n > 10^{20}$ and $d(n) \geq 14$ (Cohen and Hagis 1980), if $30 \nmid n$, then $d(n) \geq 26$ (Wall 1980), and if $3|n$ then $d(n) \geq 213$ and 5.5×10^{570} (Lieuwens 1970).

References

Cohen, G. L. and Hagis, P. Jr. "On the Number of Prime Factors of n is $\phi(n)|(n-1)$." *Nieuw Arch. Wisk.* **28**, 177–185, 1980.

Lieuwens, E. "Do There Exist Composite Numbers for which $k\phi(M) = M-1$ Holds?" *Nieuw. Arch. Wisk.* **18**, 165–169, 1970.

Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, pp. 27–28, 1989.

Wall, D. W. "Conditions for $\phi(N)$ to Properly Divide $N-1$." In *A Collection of of Manuscripts Related to the Fibonacci Sequence* (Ed. V. E. Hoggatt and M. V. E. Bicknell-Johnson). San Jose, CA: Fibonacci Assoc., pp. 205–208, 1980.

Lehmer-Schur Method

An ALGORITHM which isolates ROOTS in the COMPLEX PLANE by generalizing 1-D bracketing.

References

Acton, F. S. *Numerical Methods That Work, 2nd printing.* Washington, DC: Math. Assoc. Amer., pp. 196–198, 1990.

Lehmer's Theorem

see FERMAT'S LITTLE THEOREM CONVERSE

Lehmus' Theorem

see STEINER-LEHMUS THEOREM

Leibniz Criterion

Also known as the ALTERNATING SERIES TEST. Given a SERIES

$$\sum_{n=1}^{\infty}(-1)^{n+1}a_n$$

with $a_n > 0$, if a_n is monotonic decreasing as $n \to \infty$ and

$$\lim_{n\to\infty} a_n = 0,$$

then the series CONVERGES.

Leibniz Harmonic Triangle

$$\begin{array}{ccccccccc} & & & & \frac{1}{1} & & & & \\ & & & \frac{1}{2} & & \frac{1}{2} & & & \\ & & \frac{1}{3} & & \frac{1}{6} & & \frac{1}{3} & & \\ & \frac{1}{4} & & \frac{1}{12} & & \frac{1}{12} & & \frac{1}{4} & \\ \frac{1}{5} & & \frac{1}{20} & & \frac{1}{30} & & \frac{1}{20} & & \frac{1}{5} \end{array}$$

In the Leibniz harmonic triangle, each FRACTION is the sum of numbers below it, with the initial and final entry on each row one over the corresponding entry in

PASCAL'S TRIANGLE. The DENOMINATORS in the second diagonals are 6, 12, 20, 30, 42, 56, ... (Sloane's A007622).

see also CATALAN'S TRIANGLE, CLARK'S TRIANGLE, EULER'S TRIANGLE, NUMBER TRIANGLE, PASCAL'S TRIANGLE, SEIDEL-ENTRINGER-ARNOLD TRIANGLE

References

Sloane, N. J. A. Sequence A007622/M4096 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Leibniz Identity

$$\frac{d^n}{dx^n}(uv) = \frac{d^n u}{dx^n}v + \binom{n}{1}\frac{d^{n-1}u}{dx^{n-1}}\frac{dv}{dx} + \ldots + \binom{n}{r}\frac{d^{n-r}u}{dx^{n-r}}\frac{d^r v}{dx^r} + ud^n v + dx^n. \qquad (1)$$

Therefore,

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} \qquad (2)$$

$$\frac{d^2x}{dy^2} = -\frac{d^2y}{dx^2}\left(\frac{dy}{dx}\right)^{-3} \qquad (3)$$

$$\frac{d^3x}{dy^3} = \left[3\left(\frac{d^2y}{dx^2}\right)^2 - \frac{d^3y}{dx^3}\frac{dy}{dx}\right]\left(\frac{dy}{dx}\right)^{-5}. \qquad (4)$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 12, 1972.

Leibniz Integral Rule

$$\frac{\partial}{\partial z}\int_{a(z)}^{b(z)} f(x,z)\,dx = \int_{a(z)}^{b(z)} \frac{\partial f}{\partial z}\,dx + f(b(z),z)\frac{\partial b}{\partial z} - f(a(z),z)\frac{\partial a}{\partial z}.$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 11, 1972.

Leibniz Series

The SERIES for the INVERSE TANGENT,

$$\tan^{-1} x = x - \tfrac{1}{3}x^3 + \tfrac{1}{5}x^5 + \ldots.$$

Lemarié's Wavelet

A wavelet used in multiresolution representation to analyze the information content of images. The WAVELET is defined by

$$H(\omega) = \left[2(1-u)^4\frac{315 - 420u + 126u^2 - 4u^3}{315 - 420v + 126v^2 - 4v^3}\right]^{1/2},$$

where

$$u \equiv \sin^2(\tfrac{1}{2}\omega)$$

$$v \equiv \sin^2 \omega$$

(Mallat 1989).

see also WAVELET

References

Mallat, S. G. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation." *IEEE Trans. Pattern Analysis Machine Intel.* **11**, 674–693, 1989.

Mallat, S. G. "Multiresolution Approximation and Wavelet Orthonormal Bases of $L^2(\mathbb{R})$." *Trans. Amer. Math. Soc.* **315**, 69–87, 1989.

Lemma

A short THEOREM used in proving a larger THEOREM. Related concepts are the AXIOM, PORISM, POSTULATE, PRINCIPLE, and THEOREM.

see also ABEL'S LEMMA, ARCHIMEDES' LEMMA, BARNES' LEMMA, BLICHFELDT'S LEMMA, BOREL-CANTELLI LEMMA, BURNSIDE'S LEMMA, DANIELSON-LANCZOS LEMMA, DEHN'S LEMMA, DILWORTH'S LEMMA, DIRICHLET'S LEMMA, DIVISION LEMMA, FARKAS'S LEMMA, FATOU'S LEMMA, FUNDAMENTAL LEMMA OF CALCULUS OF VARIATIONS, GAUSS'S LEMMA, HENSEL'S LEMMA, ITÔ'S LEMMA, JORDAN'S LEMMA, LAGRANGE'S LEMMA, NEYMAN-PEARSON LEMMA, POINCARÉ'S HOLOMORPHIC LEMMA, POINCARÉ'S LEMMA, PÓLYA-BURNSIDE LEMMA, RIEMANN-LEBESGUE LEMMA, SCHUR'S LEMMA, SCHUR'S REPRESENTATION LEMMA, SCHWARZ-PICK LEMMA, SPIJKER'S LEMMA, ZORN'S LEMMA

Lemniscate

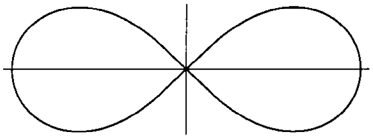

A polar curve also called LEMNISCATE OF BERNOULLI which is the LOCUS of points the product of whose distances from two points (called the FOCI) is a constant. Letting the FOCI be located at $(\pm a, 0)$, the Cartesian equation is

$$[(x-a)^2 + y^2][(x+a)^2 + y^2] = a^4, \qquad (1)$$

which can be rewritten

$$x^4 + y^4 + 2x^2y^2 = 2a^2(x^2 - y^2). \qquad (2)$$

Letting $a' \equiv \sqrt{2}a$, the POLAR COORDINATES are given by

$$r^2 = a^2 \cos(2\theta). \tag{3}$$

An alternate form is

$$r^2 = a^2 \sin(2\theta). \tag{4}$$

The parametric equations for the lemniscate are

$$x = \frac{a \cos t}{1 + \sin^2 t} \tag{5}$$

$$y = \frac{a \sin t \cos t}{1 + \sin^2 t}. \tag{6}$$

The bipolar equation of the lemniscate is

$$rr' = \tfrac{1}{2}a^2, \tag{7}$$

and in PEDAL COORDINATES with the PEDAL POINT at the center, the equation is

$$pa^2 = r^3. \tag{8}$$

The two-center BIPOLAR COORDINATES equation with origin at a FOCUS is

$$r_1 r_2 = c^2. \tag{9}$$

Jakob Bernoulli published an article in *Acta Eruditorum* in 1694 in which he called this curve the lemniscus ("a pendant ribbon"). Jakob Bernoulli was not aware that the curve he was describing was a special case of CASSINI OVALS which had been described by Cassini in 1680. The general properties of the lemniscate were discovered by G. Fagnano in 1750 (MacTutor Archive). Gauss's and Euler's investigations of the ARC LENGTH of the curve led to later work on ELLIPTIC FUNCTIONS.

The CURVATURE of the lemniscate is

$$\kappa = \frac{3\sqrt{2}\cos t}{\sqrt{3 - \cos(2t)}}. \tag{10}$$

The ARC LENGTH is more problematic. Using the polar form,

$$ds^2 = dr^2 + r^2\,d\theta^2 \tag{11}$$

so

$$ds = \sqrt{1 + \left(r\frac{d\theta}{dr}\right)^2}\,dr. \tag{12}$$

But we have

$$2r\,dr = 2a^2 \sin(2\theta)\,d\theta \tag{13}$$

$$r\frac{dr}{d\theta} = \frac{r^2}{a^2 \sin(2\theta)} \tag{14}$$

$$\left(r\frac{d\theta}{dr}\right)^2 = \frac{r^4}{a^4 \sin^2(2\theta)} = \frac{r^4}{a^4[1 - \cos^2(2\theta)]} = \frac{r^4}{a^4 - r^4}, \tag{15}$$

so

$$ds = \sqrt{1 + \frac{r^4}{a^4 - r^4}}\,dr = \sqrt{\frac{a^4}{a^4 - r^4}}\,dr = \frac{a^2}{\sqrt{a^4 - r^4}}\,dr = \frac{dr}{\sqrt{1 - \left(\frac{r}{a}\right)^4}}, \tag{16}$$

and

$$L = \int_0^a ds = 2\int_0^a \frac{ds}{dr}\,dr = 2\int_0^a \frac{dr}{\sqrt{1 - \left(\frac{r}{a}\right)^4}}. \tag{17}$$

Let $t \equiv r/a$, so $dt = dr/a$, and

$$L = 2a\int_0^1 (1 - t^4)^{-1/2}\,dt, \tag{18}$$

which, as shown in LEMNISCATE FUNCTION, is given analytically by

$$L = \sqrt{2}\,aK\left(\frac{1}{\sqrt{2}}\right) = \frac{\Gamma^2(\frac{1}{4})}{2^{3/2}\sqrt{\pi}}\,a. \tag{19}$$

If $a = 1$, then

$$L = 5.2441151086\ldots, \tag{20}$$

which is related to GAUSS'S CONSTANT M by

$$L = \frac{2\pi}{M}. \tag{21}$$

The quantity $L/2$ or $L/4$ is called the LEMNISCATE CONSTANT and plays a role for the lemniscate analogous to that of π for the CIRCLE.

The AREA of one loop of the lemniscate is

$$A = \tfrac{1}{2}\int r^2\,d\theta = \tfrac{1}{2}a^2\int_{-\pi/4}^{\pi/4} \cos(2\theta)\,d\theta = \tfrac{1}{4}a^2[\sin(2\theta)]_{-\pi/4}^{\pi/4} = \tfrac{1}{2}a^2[\sin(2\theta)]_0^{\pi/4} = \tfrac{1}{2}a^2[\sin(\tfrac{\pi}{2}) - \sin 0] = \tfrac{1}{2}a^2. \tag{22}$$

see also LEMNISCATE FUNCTION

References

Ayoub, R. "The Lemniscate and Fagnano's Contributions to Elliptic Integrals." *Arch. Hist. Exact Sci.* **29**, 131–149, 1984.

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, 1987.

Gray, A. "Lemniscates of Bernoulli." §3.2 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 39–41, 1993.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 120–124, 1972.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 37, 1983.

Lee, X. "Lemniscate of Bernoulli." `http://www.best.com/~xah/SpecialPlaneCurves_dir/LemniscateOfBernoulli_dir/lemniscateOfBernoulli.html`.

Lockwood, E. H. *A Book of Curves.* Cambridge, England: Cambridge University Press, 1967.

MacTutor History of Mathematics Archive. "Lemniscate of Bernoulli." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Lemniscate.html`.

Yates, R. C. "Lemniscate." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 143–147, 1952.

Lemniscate of Bernoulli

see LEMNISCATE

Lemniscate Case

The case of the WEIERSTRAß ELLIPTIC FUNCTION with invariants $g_2 = 1$ and $g_3 = 0$.

see also EQUIANHARMONIC CASE, WEIERSTRAß ELLIPTIC FUNCTION, PSEUDOLEMNISCATE CASE

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Lemniscate Case ($g_2 = 1$, $g_3 = 0$)." §18.14 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 658–662, 1972.

Lemniscate Constant

Let

$$L = \frac{1}{\sqrt{2\pi}}[\Gamma(\tfrac{1}{4})]^2 = 5.2441151086\ldots$$

be the ARC LENGTH of a LEMNISCATE with $a = 1$. Then the lemniscate constant is the quantity $L/2$ (Abramowitz and Stegun 1972), or $L/4 = 1.311028777\ldots$ (Todd 1975, Le Lionnais 1983). Todd (1975) cites T. Schneider (1937) as proving L to be a TRANSCENDENTAL NUMBER.

see also LEMNISCATE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, 1972.

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, 1987.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/gauss/gauss.html`.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 37, 1983.

Todd, J. "The Lemniscate Constant." *Comm. ACM* **18**, 14–19 and 462, 1975.

Lemniscate Function

The lemniscate functions arise in rectifying the ARC LENGTH of the LEMNISCATE. The lemniscate functions were first studied by Jakob Bernoulli and G. Fagnano. A historical account is given by Ayoub (1984), and an extensive discussion by Siegel (1969). The lemniscate functions were the first functions defined by inversion of an integral, which was first done by Gauss.

$$L = 2a \int_0^1 (1-t^4)^{-1/2}\,dt. \tag{1}$$

Define the functions

$$\phi(x) \equiv \operatorname{arcsinlemn} x = \int_0^x (1-t^4)^{-1/2}\,dt \tag{2}$$

$$\phi'(x) \equiv \operatorname{arccoslemn} x = \int_x^1 (1-t^4)^{-1/2}\,dt, \tag{3}$$

where

$$\varpi \equiv \frac{L}{a}, \tag{4}$$

and write

$$x = \operatorname{sinlemn} \phi \tag{5}$$

$$x = \operatorname{coslemn} \phi'. \tag{6}$$

There is an identity connecting ϕ and ϕ' since

$$\phi(x) + \phi'(x) = \frac{L}{2a} = \tfrac{1}{2}\varpi, \tag{7}$$

so

$$\operatorname{sinlemn} \phi = \operatorname{coslemn}(\tfrac{1}{2}\varpi - \phi). \tag{8}$$

These functions can be written in terms of JACOBI ELLIPTIC FUNCTIONS,

$$u = \int_0^{\operatorname{sd}(u,k)} [(1-k'^2y^2)(1+k^2y^2)]^{-1/2}\,dy. \tag{9}$$

Now, if $k = k' = 1/\sqrt{2}$, then

$$\begin{aligned} u &= \int_0^{\operatorname{sd}(u,1/\sqrt{2})} [(1-\tfrac{1}{2}y^2)(1+\tfrac{1}{2}y^2)]^{-1/2}\,dy \\ &= \int_0^{\operatorname{sd}(u,1/\sqrt{2})} (1-\tfrac{1}{4}y^4)^{-1/2}\,dy. \end{aligned} \tag{10}$$

Let $t \equiv y/\sqrt{2}$ so $dy = \sqrt{2}\,dt$,

$$u = \sqrt{2} \int_0^{\operatorname{sd}(u,1/\sqrt{2})/\sqrt{2}} (1-t^4)^{-1/2}\,dt \tag{11}$$

$$\frac{u}{\sqrt{2}} = \int_0^{\operatorname{sd}(u,1/\sqrt{2})/\sqrt{2}} (1-t^4)^{-1/2}\,dt \tag{12}$$

$$u = \int_0^{\operatorname{sd}(u\sqrt{2},1/\sqrt{2})/\sqrt{2}} (1-t^4)^{-1/2}\,dt, \tag{13}$$

and

$$\operatorname{sinlemn} \phi = \frac{1}{\sqrt{2}} \operatorname{sd}\left(\phi\sqrt{2}, \frac{1}{\sqrt{2}}\right). \tag{14}$$

Similarly,

$$\begin{aligned} u &= \int_{\operatorname{cn}(u,k)}^1 (1-t^2)^{-1/2}(k'^2+k^2t^2)^{-1/2}\,dt \\ &= \int_{\operatorname{cn}(u,1/\sqrt{2})}^1 (1-t^2)^{-1/2}\left(\tfrac{1}{2}+\tfrac{1}{2}t^2\right)^{-1/2}\,dt \\ &= \sqrt{2} \int_{\operatorname{cn}(u,1/\sqrt{2})}^1 (1-t^4)^{-1/2}\,dt \end{aligned} \tag{15}$$

$$\frac{u}{\sqrt{2}} = \int_{\operatorname{cn}(u,1/\sqrt{2})}^1 (1-t^4)^{-1/2}\,dt \tag{16}$$

$$u = \int_{\operatorname{cn}(u\sqrt{2},1/\sqrt{2})}^{1} (1-t^4)^{-1/2}\,dt, \tag{17}$$

and

$$\operatorname{coslemn}\phi = \operatorname{cn}\left(\phi\sqrt{2}, \frac{1}{\sqrt{2}}\right). \tag{18}$$

We know

$$\operatorname{coslemn}(\tfrac{1}{2}\varpi) = \operatorname{cn}\left(\tfrac{1}{2}\varpi\sqrt{2}, \frac{1}{\sqrt{2}}\right) = 0. \tag{19}$$

But it is true that

$$\operatorname{cn}(K,k) = 0, \tag{20}$$

so

$$K\left(\frac{1}{\sqrt{2}}\right) = \tfrac{1}{2}\sqrt{2}\,\varpi = \frac{1}{\sqrt{2}}\varpi \tag{21}$$

$$\frac{\Gamma^2(\frac{1}{4})}{4\sqrt{\pi}} = \frac{1}{\sqrt{2}}\varpi \tag{22}$$

$$L = a\varpi = a\sqrt{2}\,\frac{\Gamma^2(\frac{1}{4})}{4\sqrt{\pi}} = \frac{\Gamma^2(\frac{1}{4})}{2^{3/2}\sqrt{\pi}}\,a. \tag{23}$$

By expanding $(1-t^4)^{-1/2}$ in a Binomial Series and integrating term by term, the arcsinlemn function can be written

$$\phi(x) = \int_0^v \frac{dt}{\sqrt{1-t^4}} = \sum_{n=0}^{\infty} \frac{(\frac{1}{2})_n x^{4n+1}}{n!(4n+1)}, \tag{24}$$

where $(a)_n$ is the Rising Factorial (Berndt 1994). Ramanujan gave the following inversion Formula for $\phi(x)$. If

$$\frac{\theta\mu}{\sqrt{2}} = \sum_{n=0}^{\infty} \frac{(\frac{1}{2})_n x^{4n+1}}{n!(4n+1)}, \tag{25}$$

where

$$\mu = \frac{\Gamma^2(\frac{1}{4})}{2\pi^{3/2}} \tag{26}$$

is the constant obtained by letting $x = 1$ and $\theta = \pi/2$, and

$$v = 2^{-1/2}\operatorname{sd}(\mu\theta), \tag{27}$$

then

$$\frac{\mu^2}{2x^2} = \csc^2\theta - \frac{1}{\pi} - 8\sum_{n=1}^{\infty} \frac{n\cos(2n\theta)}{e^{2\pi n}-1} \tag{28}$$

(Berndt 1994). Ramanujan also showed that if $0 < \theta < \pi/2$, then

$$-\frac{\mu}{\sqrt{2}}\sum_{n=0}^{\infty} \frac{(\frac{1}{2})_n v^{4n-1}}{n!(4n-1)} = \cot\theta + \frac{\theta}{\pi} + 4\sum_{n=1}^{\infty} \frac{\sin(2n\theta)}{2^{2\pi n}-1}, \tag{29}$$

$$\ln v + \tfrac{1}{6}\pi - \tfrac{1}{2}\ln 2 + \sum_{n=1}^{\infty} \frac{(\frac{1}{4})_n v^{4n}}{(\frac{3}{4})_n 4n} = \ln(\sin\theta) + \frac{\theta^2}{2\pi} - 2\sum_{n=1}^{\infty} \frac{\cos(2n\theta)}{n(e^{2\pi n}-1)}, \tag{30}$$

$$\tfrac{1}{2}\tan^{-1} v = \sum_{n=0}^{\infty} \frac{\sin[(2n+1)\theta]}{(2n+1)\cosh[\frac{1}{2}(2n+1)\pi]}, \tag{31}$$

$$\tfrac{1}{4}\cos^{-1}(v^2) = \sum_{n=0}^{\infty} \frac{(-1)^n\cos[(2n+1)\theta]}{(2n+1)\cosh[\frac{1}{2}(2n+1)\pi]}, \tag{32}$$

and

$$\frac{\sqrt{2}}{4\mu}\sum_{n=0}^{\infty} \frac{2^{2n}(n!)^2}{(2n+1)!(4n+3)} v^{4n+3} = \frac{\pi\theta}{8} - \sum_{n=0}^{\infty} \frac{(-1)^n\sin[(2n+1)\theta]}{(2n+1)^2\cosh[\frac{1}{2}(2n+1)\pi]} \tag{33}$$

(Berndt 1994).

A generalized version of the lemniscate function can be defined by letting $0 \leq \theta \leq \pi/2$ and $0 \leq v \leq 1$. Write

$$\tfrac{2}{3}\theta\mu = \int_0^v \frac{dt}{\sqrt{1-t^6}}, \tag{34}$$

where μ is the constant obtained by setting $\theta = \pi/2$ and $v = 1$. Then

$$\mu = \frac{\sqrt{\pi}}{\Gamma(\frac{2}{3})\Gamma(\frac{5}{6})}, \tag{35}$$

and Ramanujan showed

$$\frac{4\mu^2}{9v^2} = \csc^2\theta - \frac{2}{\pi\sqrt{3}} + 8\sum_{n=1}^{\infty} \frac{(-1)^{n-1}n\cos(2n\theta)}{e^{\pi n\sqrt{3}} - (-1)^n} \tag{36}$$

(Berndt 1994).

see also Hyperbolic Lemniscate Function

References

Ayoub, R. "The Lemniscate and Fagnano's Contributions to Elliptic Integrals." *Arch. Hist. Exact Sci.* **29**, 131–149, 1984.

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 245, and 247–255, 258–260, 1994.

Siegel, C. L. *Topics in Complex Function Theory, Vol. 1.* New York: Wiley, 1969.

Lemniscate of Gerono

see Eight Curve

Lemniscate Inverse Curve

The Inverse Curve of a Lemniscate in a Circle centered at the origin and touching the Lemniscate where it crosses the x-Axis produces a Rectangular Hyperbola.

Lemniscate (Mandelbrot Set)

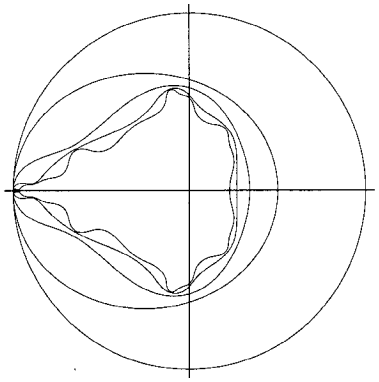

A curve on which points of a MAP z_n (such as the MANDELBROT SET) diverge to a given value $r_{\max}$ at the same rate. A common method of obtaining lemniscates is to define an INTEGER called the COUNT which is the largest n such that $|z_n| < r$ where r is usually taken as $r = 2$. Successive COUNTS then define a series of lemniscates, which are called EQUIPOTENTIAL CURVES by Peitgen and Saupe (1988).

see also COUNT, MANDELBROT SET

References

Peitgen, H.-O. and Saupe, D. (Eds.). *The Science of Fractal Images.* New York: Springer-Verlag, pp. 178–179, 1988.

Lemoine Axis

see LEMOINE LINE

Lemoine Circle

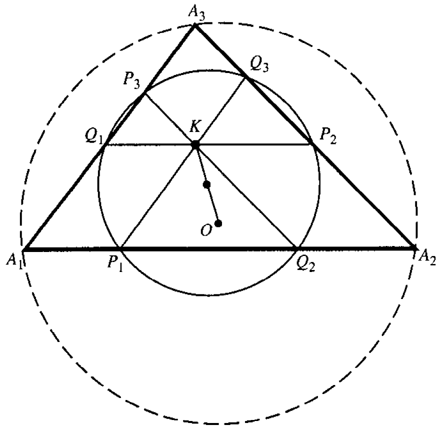

Also called the TRIPLICATE-RATIO CIRCLE. Draw lines through the LEMOINE POINT K and parallel to the sides of the triangle. The points where the parallel lines intersect the sides then lie on a CIRCLE known as the Lemoine circle. This circle has center at the MIDPOINT of OK, where O is the CIRCUMCENTER. The circle has radius

$$\tfrac{1}{2}\sqrt{R^2 + r^2} = \tfrac{1}{2}R\sec\omega,$$

where R is the CIRCUMRADIUS, r is the INRADIUS, and ω is the BROCARD ANGLE. The Lemoine circle divides any side into segments proportional to the squares of the sides

$$\overline{A_2P_2} : \overline{P_2Q_3} : \overline{Q_3A_3} = {a_3}^2 : {a_1}^2 : {a_2}^2.$$

Furthermore, the chords cut from the sides by the Lemoine circle are proportional to the squares of the sides.

The COSINE CIRCLE is sometimes called the second Lemoine circle.

see also COSINE CIRCLE, LEMOINE LINE, LEMOINE POINT, TUCKER CIRCLES

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 273–275, 1929.

Lemoine Line

The Lemoine line, also called the LEMOINE AXIS, is the perspectivity axis of a TRIANGLE and its TANGENTIAL TRIANGLE, and also the TRILINEAR POLAR of the CENTROID of the triangle vertices. It is also the POLAR of K with regard to its CIRCUMCIRCLE, and is PERPENDICULAR to the BROCARD AXIS.

The centers of the APOLLONIUS CIRCLES L_1, L_2, and L_3 are COLLINEAR on the LEMOINE LINE. This line is PERPENDICULAR to the BROCARD AXIS OK and is the RADICAL AXIS of the CIRCUMCIRCLE and the BROCARD CIRCLE. It has equation

$$\frac{\alpha}{a} + \frac{\beta}{b} + \frac{\gamma}{c}$$

in terms of TRILINEAR COORDINATES (Oldknow 1996).

see also APOLLONIUS CIRCLES, BROCARD AXIS, CENTROID (TRIANGLE), CIRCUMCIRCLE, COLLINEAR, LEMOINE CIRCLE, LEMOINE POINT, POLAR, RADICAL AXIS, TANGENTIAL TRIANGLE, TRILINEAR POLAR

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, p. 295, 1929.

Oldknow, A. "The Euler-Gergonne-Soddy Triangle of a Triangle." *Amer. Math. Monthly* **103**, 319–329, 1996.

Lemoine Point

The point of concurrence K of the SYMMEDIAN LINES, sometimes also called the SYMMEDIAN POINT and GREBE POINT.

Let G be the CENTROID of a TRIANGLE ΔABC, L_A, L_B, and L_C the ANGLE BISECTORS of ANGLES A, B, C, and G_A, G_B, and G_C the reflections of AG, BG, and CG about L_A, L_B, and L_C. Then K is the point of concurrence of the lines G_A, G_B, and G_C. It is the perspectivity center of a TRIANGLE and its TANGENTIAL TRIANGLE.

In AREAL COORDINATES (actual TRILINEAR COORDINATES), the Lemoine point is the point for which $\alpha^2+\beta^2+\gamma^2$ is a minimum. A center X is the CENTROID of its own PEDAL TRIANGLE IFF it is the Lemoine point.

The Lemoine point lies on the BROCARD AXIS, and its distances from the Lemoine point K to the sides of the TRIANGLE are

$$\overline{KK_i} = \tfrac{1}{2}a_i \tan\omega,$$

where ω is the BROCARD ANGLE. A BROCARD LINE, MEDIAN, and Lemoine point are concurrent, with $A_1\Omega_1$, A_2K, and A_3M meeting at a point. Similarly, $A_1\Omega'$, A_2M, and A_3K meet at a point which is the ISOGONAL CONJUGATE of the first (Johnson 1929, pp. 268–269). The line joining the MIDPOINT of any side to the midpoint of the ALTITUDE on that side passes through the Lemoine point K. The Lemoine point K is the STEINER POINT of the first BROCARD TRIANGLE.

see also ANGLE BISECTOR, BROCARD ANGLE, BROCARD AXIS, BROCARD DIAMETER, CENTROID (TRIANGLE), COSYMMEDIAN TRIANGLES, GREBE POINT, ISOGONAL CONJUGATE, LEMOINE CIRCLE, LEMOINE LINE, LINE AT INFINITY, MITTENPUNKT, PEDAL TRIANGLE, STEINER POINTS, SYMMEDIAN LINE, TANGENTIAL TRIANGLE

References
Gallatly, W. *The Modern Geometry of the Triangle, 2nd ed.* London: Hodgson, p. 86, 1913.
Honsberger, R. *Episodes in Nineteenth and Twentieth Century Euclidean Geometry.* Washington, DC: Math. Assoc. Amer., 1995.
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 217, 268–269, and 271–272, 1929.
Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.
Kimberling, C. "Symmedian Point." `http://www.evansville.edu/~ck6/tcenters/class/sympt.html`.
Mackay, J. S. "Early History of the Symmedian Point." *Proc. Edinburgh Math. Soc.* **11**, 92–103, 1892–1893.

Lemoine's Problem

Given the vertices of the three EQUILATERAL TRIANGLES placed on the sides of a TRIANGLE T, construct T. The solution can be given using KIEPERT'S HYPERBOLA.

see also KIEPERT'S HYPERBOLA

Lemon

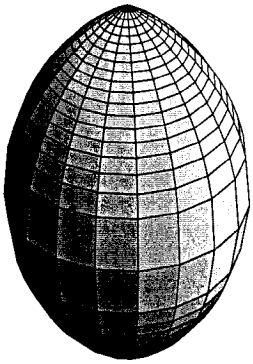

A SURFACE OF REVOLUTION defined by Kepler. It consists of less than half of a circular ARC rotated about an axis passing through the endpoints of the ARC. The equations of the upper and lower boundaries in the xz plane are

$$z_\pm = \pm\sqrt{R^2-(x+r)^2}$$

for $R > r$ and $x \in [-(R-r), R-r]$. The CROSS-SECTION of a lemon is a LENS. The lemon is the inside surface of a SPINDLE TORUS.

see also APPLE, LENS, SPINDLE TORUS

Length (Curve)

Let $\gamma(t)$ be a smooth curve in a MANIFOLD M from x to y with $\gamma(0) = x$ and $\gamma(1) = y$. Then $\gamma'(t) \in T_{\gamma(t)}$, where T_x is the TANGENT SPACE of M at x. The length of γ with respect to the Riemannian structure is given by

$$\int_0^1 ||\gamma'(t)||_{\gamma(t)}\,dt.$$

see also ARC LENGTH, DISTANCE

Length Distribution Function

A function giving the distribution of the interpoint distances of a curve. It is defined by

$$p(r) = \frac{1}{N}\sum_{ij}\delta_{r_{ij}=r}.$$

see also RADIUS OF GYRATION

References
Pickover, C. A. *Keys to Infinity.* New York: W. H. Freeman, pp. 204–206, 1995.

Length (Number)

The length of a number n in base b is the number of DIGITS in the base-b numeral for n, given by the formula

$$L(n,b) = \lfloor \log_b(n) \rfloor + 1,$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION.

The MULTIPLICATIVE PERSISTENCE of an n-DIGIT is sometimes also called its length.

see also CONCATENATION, DIGIT, FIGURES, MULTIPLICATIVE PERSISTENCE

Length (Partial Order)

For a PARTIAL ORDER, the size of the longest CHAIN is called the length.

see also WIDTH (PARTIAL ORDER)

Length (Size)

The longest dimension of a 3-D object.

see also HEIGHT, WIDTH (SIZE)

Lengyel's Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let L denote the partition lattice of the SET $\{1, 2, \ldots, n\}$. The MAXIMUM element of L is

$$M = \{\{1, 2, \ldots, n\}\} \tag{1}$$

and the MINIMUM element is

$$m = \{\{1\}, \{2\}, \ldots, \{n\}\}. \tag{2}$$

Let Z_n denote that number of chains of any length in L containing both M and m. Then Z_n satisfies the RECURRENCE RELATION

$$Z_n = \sum_{k=1}^{n-1} s(n,k) Z_k, \tag{3}$$

where $s(n,k)$ is a STIRLING NUMBER OF THE SECOND KIND. Lengyel (1984) proved that the QUOTIENT

$$r(n) = \frac{Z_n}{(n!)^2 (2\ln 2)^{-n} n^{1-(\ln 2)/3}} \tag{4}$$

is bounded between two constants as $n \to \infty$, and Flajolet and Salvy (1990) improved the result of Babai and Lengyel (1992) to show that

$$\Lambda \equiv \lim_{n\to\infty} r(n) = 1.0986858055\ldots. \tag{5}$$

References

Babai, L. and Lengyel, T. "A Convergence Criterion for Recurrent Sequences with Application to the Partition Lattice." *Analysis* **12**, 109–119, 1992.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/lngy/lngy.html`.

Flajolet, P. and Salvy, B. "Hierarchal Set Partitions and Analytic Iterates of the Exponential Function." Unpublished manuscript, 1990.

Lengyel, T. "On a Recurrence Involving Stirling Numbers." *Europ. J. Comb.* **5**, 313–321, 1984.

Plouffe, S. "The Lengyel Constant." `http://lacim.uqam.ca/piDATA/lengyel.txt`.

Lens

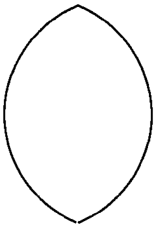

A figure composed of two equal and symmetrically placed circular ARCS. It is also known as the FISH BLADDER (Pedoe 1995, p. xii) or VESICA PISCIS. The latter term is often used for the particular lens formed by the intersection of two unit CIRCLES whose centers are offset by a unit distance (Rawles 1997). In this case, the height of the lens is given by letting $d = r = R = 1$ in the equation for a CIRCLE-CIRCLE INTERSECTION

$$a = \frac{1}{d}\sqrt{4d^2R^2 - (d^2 - r^2 + R^2)^2}, \tag{1}$$

giving $a = \sqrt{3}$. The AREA of the VESICA PISCIS is given by plugging $d = R$ into the CIRCLE-CIRCLE INTERSECTION area equation with $r = R$,

$$A = 2R^2 \cos^{-1}\left(\frac{d}{2R}\right) - \tfrac{1}{2}d\sqrt{4R^2 - d^2}, \tag{2}$$

giving

$$A = \tfrac{1}{6}(4\pi - 3\sqrt{3}) \approx 1.22837. \tag{3}$$

Renaissance artists frequently surrounded images of Jesus with the vesica piscis (Rawles 1997). An asymmetrical lens is produced by a CIRCLE-CIRCLE INTERSECTION for unequal CIRCLES.

see also CIRCLE, CIRCLE-CIRCLE INTERSECTION, FLOWER OF LIFE, LEMON, LUNE (PLANE), REULEAUX TRIANGLE, SECTOR, SEED OF LIFE, SEGMENT, VENN DIAGRAM

References

Pedoe, D. *Circles: A Mathematical View, rev. ed.* Washington, DC: Math. Assoc. Amer., 1995.

Rawles, B. *Sacred Geometry Design Sourcebook: Universal Dimensional Patterns.* Nevada City, CA: Elysian Pub., p. 11, 1997.

Lens Space

A lens space $L(p,q)$ is the 3-MANIFOLD obtained by gluing the boundaries of two solid TORI together such that the meridian of the first goes to a (p,q)-curve on the second, where a (p,q)-curve has p meridians and q longitudes.

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, 1976.

Lenstra Elliptic Curve Method

A method of factoring INTEGERS using ELLIPTIC CURVES.

References

Montgomery, P. L. "Speeding up the Pollard and Elliptic Curve Methods of Factorization." *Math. Comput.* **48**, 243–264, 1987.

Léon Anne's Theorem

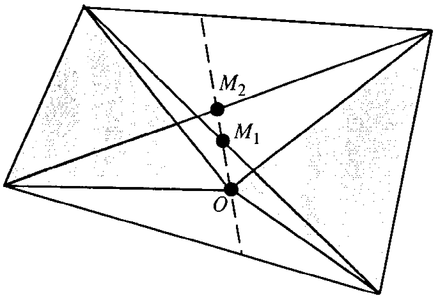

Pick a point O in the interior of a QUADRILATERAL which is not a PARALLELOGRAM. Join this point to each of the four VERTICES, then the LOCUS of points O for which the sum of opposite TRIANGLE areas is half the QUADRILATERAL AREA is the line joining the MIDPOINTS M_1 and M_2 of the DIAGONALS.

see also DIAGONAL (POLYGON), MIDPOINT, QUADRILATERAL

References

Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., pp. 174–175, 1991.

Leonardo's Paradox

In the depiction of a row of identical columns parallel to the plane of a PERSPECTIVE drawing, the outer columns should appear wider even though they are farther away.

see also PERSPECTIVE, VANISHING POINT, ZEEMAN'S PARADOX

References

Dixon, R. *Mathographics.* New York: Dover, p. 82, 1991.

Leptokurtic

A distribution with a high peak so that the KURTOSIS satisfies $\gamma_2 > 0$.

see also KURTOSIS

Lerch's Theorem

If there are two functions $F_1(t)$ and $F_2(t)$ with the same integral transform

$$\mathcal{T}[F_1(t)] = \mathcal{T}[F_2(t)] \equiv f(s), \tag{1}$$

then a NULL FUNCTION can be defined by

$$\delta_0(t) \equiv F_1(t) - F_2(t) \tag{2}$$

so that the integral

$$\int_0^a \delta_0(t)\,dt = 0 \tag{3}$$

vanishes for all $a > 0$.

see also NULL FUNCTION

Lerch Transcendent

A generalization of the HURWITZ ZETA FUNCTION and POLYLOGARITHM function. Many sums of reciprocal POWERS can be expressed in terms of it. It is defined by

$$\Phi(z,s,a) \equiv \sum_{k=0}^{\infty} \frac{z^k}{(a+k)^s}, \tag{1}$$

where any term with $a + k = 0$ is excluded.

The Lerch transcendent can be used to express the DIRICHLET BETA FUNCTION

$$\beta(s) \equiv \sum_{k=0}^{\infty} (-1)^k (2k+1)^{-s} 2^{-s} \Phi(-1, s, \tfrac{1}{2}), \tag{2}$$

the integral of the FERMI-DIRAC DISTRIBUTION

$$\int_0^{\infty} \frac{k^s}{e^{k-\mu}+1}\,dk = e^{\mu}\Gamma(s+1)\Phi(-e^{\mu}, s+1, 1), \tag{3}$$

where $\Gamma(z)$ is the GAMMA FUNCTION, and to evaluate the DIRICHLET L-SERIES.

see also DIRICHLET BETA FUNCTION, DIRICHLET L-SERIES, FERMI-DIRAC DISTRIBUTION, HURWITZ ZETA FUNCTION, POLYLOGARITHM

Less

A quantity a is said to be less than b if a is smaller than b, written $a < b$. If a is less than or EQUAL to b, the relationship is written $a \leq b$. If a is MUCH LESS than b, this is written $a \ll b$. Statements involving GREATER than and less than symbols are called INEQUALITIES.

see also EQUAL, GREATER, INEQUALITY, MUCH GREATER, MUCH LESS

Letter-Value Display

A method of displaying simple statistical parameters including HINGES, MEDIAN, and upper and lower values.

References

Tukey, J. W. *Explanatory Data Analysis.* Reading, MA: Addison-Wesley, p. 33, 1977.

Leudesdorf Theorem

Let $t(m)$ denote the set of the $\phi(m)$ numbers less than and RELATIVELY PRIME to m, where $\phi(n)$ is the TOTIENT FUNCTION. Then if

$$S_m \equiv \sum_{t(m)} \frac{1}{t},$$

then

$$\begin{cases} S_m \equiv 0 \pmod{m^2} & \text{if } 2\nmid m,\ 3\nmid m \\ S_m \equiv 0 \pmod{\frac{1}{3}m^2} & \text{if } 2\nmid m,\ 3|m \\ S_m \equiv 0 \pmod{\frac{1}{2}m^2} & 2|m,\ \nmid m,\ m \text{ not a power of 2} \\ S_m \equiv 0 \pmod{\frac{1}{6}m^2} & \text{if } 2|m,\ 3|m \\ S_m \equiv 0 \pmod{\frac{1}{4}m^2} & \text{if } m = 2^a. \end{cases}$$

see also BAUER'S IDENTICAL CONGRUENCE, TOTIENT FUNCTION

References

Hardy, G. H. and Wright, E. M. "A Theorem of Leudesdorf." §8.7 in *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, pp. 100–102, 1979.

Level Curve

A LEVEL SET in 2-D.

Level Set

The level set of c is the SET of points

$$\{(x_1,\ldots,x_n) \in U : f(x_1,\ldots,x_n) = c\} \in \mathbb{R}^n,$$

and is in the DOMAIN of the function. If $n = 2$, the level set is a plane curve (a level curve). If $n = 3$, the level set is a surface (a level surface).

References

Gray, A. "Level Surfaces in $\mathbb{R}^3$." §10.7 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 204–207, 1993.

Level Surface

A LEVEL SET in 3-D.

Levi-Civita Density

see PERMUTATION SYMBOL

Levi-Civita Symbol

see PERMUTATION SYMBOL

Levi-Civita Tensor

see PERMUTATION TENSOR

Leviathan Number

The number $(10^{666})!$, where 666 is the BEAST NUMBER and $n!$ denotes a FACTORIAL. The number of trailing zeros in the Leviathan number is $25 \times 10^{664} - 143$ (Pickover 1995).

see also 666, APOCALYPSE NUMBER, APOCALYPTIC NUMBER, BEAST NUMBER

References

Pickover, C. A. *Keys to Infinity.* New York: Wiley, pp. 97–102, 1995.

Levine-O'Sullivan Greedy Algorithm

For a sequence $\{\chi_i\}$, the Levine-O'Sullivan greedy algorithm is given by

$$\begin{aligned} \chi_1 &= 1 \\ \chi_i &= \max_{1\leq j\leq i-1} (j+1)(i-\chi_j) \end{aligned}$$

for $i > 1$.

see also GREEDY ALGORITHM, LEVINE-O'SULLIVAN SEQUENCE

References

Levine, E. and O'Sullivan, J. "An Upper Estimate for the Reciprocal Sum of a Sum-Free Sequence." *Acta Arith.* **34**, 9–24, 1977.

Levine-O'Sullivan Sequence

The sequence generated by the LEVINE-O'SULLIVAN GREEDY ALGORITHM: 1, 2, 4, 6, 9, 12, 15, 18, 21, 24, 28, 32, 36, 40, 45, 50, 55, 60, 65, ... (Sloane's A014011). The reciprocal sum of this sequence is conjectured to bound the reciprocal sum of all A-SEQUENCES.

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/erdos/erdos.html`.

Levine, E. and O'Sullivan, J. "An Upper Estimate for the Reciprocal Sum of a Sum-Free Sequence." *Acta Arith.* **34**, 9–24, 1977.

Sloane, N. J. A. Sequence A014011 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Lévy Constant

Let p_n/q_n be the nth CONVERGENT of a REAL NUMBER x. Then almost all REAL NUMBERS satisfy

$$L \equiv \lim_{n\to\infty} (q_n)^{1/n} = e^{\pi^2/(12\ln 2)} = 3.27582291872\ldots.$$

see also KHINTCHINE'S CONSTANT, KHINTCHINE-LÉVY CONSTANT

References

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 51, 1983.

Lévy Distribution

$$\mathcal{F}[P_N(k)] = \exp(-N|k|^\beta),$$

where $\mathcal{F}$ is the FOURIER TRANSFORM of the probability $P_N(k)$ for N-step addition of random variables. Lévy showed that $\beta \in (0, 2)$ for $P(x)$ to be NONNEGATIVE. The Lévy distribution has infinite variance and sometimes infinite mean. The case $\beta = 1$ gives a CAUCHY DISTRIBUTION, while $\beta = 2$ gives a GAUSSIAN DISTRIBUTION.

see also CAUCHY DISTRIBUTION, GAUSSIAN DISTRIBUTION

Lévy Flight

RANDOM WALK trajectories which are composed of self-similar jumps. They are described by the LÉVY DISTRIBUTION.

see also LÉVY DISTRIBUTION

References

Shlesinger, M.; Zaslavsky, G. M.; and Frisch, U. (Eds.). *Lévy Flights and Related Topics in Physics.* New York: Springer-Verlag, 1995.

Lévy Fractal

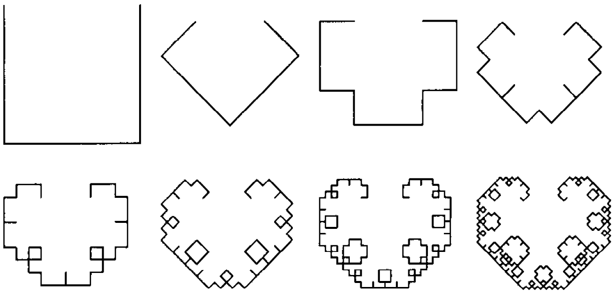

A FRACTAL curve, also called the C-CURVE (Beeler *et al.* 1972, Item 135). The base curve and motif are illustrated below.

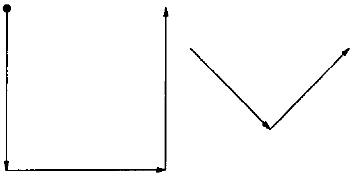

see also LÉVY TAPESTRY

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Dixon, R. *Mathographics.* New York: Dover, pp. 182–183, 1991.

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 45–48, 1991.

Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m`.

Lévy Function

see BROWN FUNCTION

Lévy Tapestry

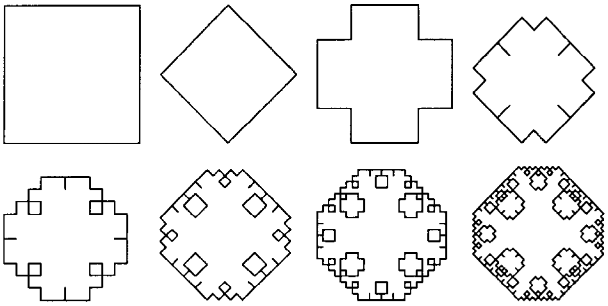

The FRACTAL curve illustrated above, with base curve and motif illustrated below.

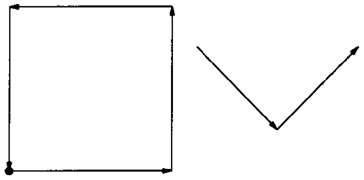

see also LÉVY FRACTAL

References

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 45–48, 1991.

Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m`.

Lew k-gram

Diagrams invented by Lewis Carroll which can be used to determine the number of minimal MINIMAL COVERS of n numbers with k members.

References

Macula, A. J. "Lewis Carroll and the Enumeration of Minimal Covers." *Math. Mag.* **68**, 269–274, 1995.

Lexicographic Order

An ordering of PERMUTATIONS in which they are listed in increasing numerical order. For example, the PERMUTATIONS of $\{1, 2, 3\}$ in lexicographic order are 123, 132, 213, 231, 312, and 321.

see also TRANSPOSITION ORDER

References

Ruskey, F. "Information on Combinations of a Set." `http://sue.csc.uvic.ca/~cos/inf/comb/CombinationsInfo.html`.

Lexis Ratio

$$L \equiv \frac{\sigma}{\sigma_B},$$

where σ is the VARIANCE in a set of s LEXIS TRIALS and σ_B is the VARIANCE assuming BERNOULLI TRIALS.

If $L < 1$, the trials are said to be SUBNORMAL, and if $L > 1$, the trials are said to be SUPERNORMAL.

see also BERNOULLI TRIAL, LEXIS TRIALS, SUBNORMAL, SUPERNORMAL

Lexis Trials

n sets of s trials each, with the probability of success p constant in each set.

$$\operatorname{var}\left(\frac{x}{n}\right) = spq + s(s-1){\sigma_p}^2,$$

where ${\sigma_p}^2$ is the VARIANCE of p_i.

see also BERNOULLI TRIAL, LEXIS RATIO

Lg

The LOGARITHM to BASE 2 is denoted lg, i.e.,

$$\lg x \equiv \log_2 x.$$

see also BASE (LOGARITHM), E, LN, LOGARITHM, NAPIERIAN LOGARITHM, NATURAL LOGARITHM

Liar's Paradox

see EPIMENIDES PARADOX

Lichnerowicz Conditions

Second and higher derivatives of the METRIC TENSOR g_{ab} need not be continuous across a surface of discontinuity, but g_{ab} and $g_{ab,c}$ must be continuous across it.

Lichnerowicz Formula

$$D^*D\psi = \nabla^*\nabla\psi + \tfrac{1}{4}R\psi - \tfrac{1}{2}F_L^+(\psi),$$

where D is the Dirac operator $D : \Gamma(W^+) \to \Gamma(W^-)$, ∇ is the COVARIANT DERIVATIVE on SPINORS, R is the CURVATURE SCALAR, and F_L^+ is the self-dual part of the curvature of L.

see also LICHNEROWICZ-WEITZENBOCK FORMULA

References

Donaldson, S. K. "The Seiberg-Witten Equations and 4-Manifold Topology." *Bull. Amer. Math. Soc.* **33**, 45–70, 1996.

Lichnerowicz-Weitzenbock Formula

$$D^*D\psi = \nabla^*\nabla\psi + \tfrac{1}{4}R\psi,$$

where D is the Dirac operator $D : \Gamma(S^+) \to \Gamma(S^-)$, ∇ is the COVARIANT DERIVATIVE on SPINORS, and R is the CURVATURE SCALAR.

see also LICHNEROWICZ FORMULA

References

Donaldson, S. K. "The Seiberg-Witten Equations and 4-Manifold Topology." *Bull. Amer. Math. Soc.* **33**, 45–70, 1996.

Lichtenfels Surface

A MINIMAL SURFACE given by the parametric equation

$$x = \Re\left[\sqrt{2}\cos(\tfrac{1}{3}\zeta)\sqrt{\cos(\tfrac{2}{3}\zeta)}\right]$$
$$y = \Re\left[-\sqrt{2}\cos(\tfrac{1}{3}\zeta)\sqrt{\cos(\tfrac{2}{3}\zeta)}\right]$$
$$z = \Re\left[-\tfrac{1}{3}\sqrt{2}i\int_0^t \frac{d\zeta}{\sqrt{\cos(\tfrac{2}{3}\zeta)}}\right].$$

References

do Carmo, M. P. "The Helicoid." §3.5F in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, p. 47, 1986.

Lichtenfels, O. von. "Notiz über eine transcendente Minimalfläche." *Sitzungsber. Kaiserl. Akad. Wiss. Wien* **94**, 41–54, 1889.

Lie Algebra

A NONASSOCIATIVE ALGEBRA obeyed by objects such as the LIE BRACKET and POISSON BRACKET. Elements f, g, and h of a Lie algebra satisfy

$$[f,g] = -[g,f], \tag{1}$$

$$[f+g,h] = [f,h] + [g,h], \tag{2}$$

and

$$[f,[g,h]] + [g,[h,f]] + [h,[f,g]] = 0 \tag{3}$$

(the JACOBI IDENTITY), and are *not* ASSOCIATIVE. The binary operation of a Lie algebra is the bracket

$$[fg,h] = f[g,h] + g[f,h]. \tag{4}$$

see also JACOBI IDENTITIES, LIE ALGEBROID, LIE BRACKET, IWASAWA'S THEOREM, POISSON BRACKET

References

Jacobson, N. *Lie Algebras.* New York: Dover, 1979.

Lie Algebroid

The infinitesimal algebraic object associated with a LIE GROUPOID. A Lie algebroid over a MANIFOLD B is a VECTOR BUNDLE A over B with a LIE ALGEBRA structure $[\,,\,]$ (LIE BRACKET) on its SPACE of smooth sections together with its ANCHOR ρ.

see also LIE ALGEBRA

References

Weinstein, A. "Groupoids: Unifying Internal and External Symmetry." *Not. Amer. Math. Soc.* **43**, 744–752, 1996.

Lie Bracket

The commutation operation

$$[a,b] = ab - ba$$

corresponding to the LIE PRODUCT.

see also LAGRANGE BRACKET, POISSON BRACKET

Lie Commutator

see LIE PRODUCT

Lie Derivative

$$\mathcal{L}_x T^{ab} \equiv \lim_{\delta u \to 0} \frac{T^{ab}(x') - T'^{ab}(x)}{\delta u}.$$

Lie Group

A continuous GROUP with an infinite number of elements such that the parameters of a product element are ANALYTIC FUNCTIONS. Lie groups are also C^∞ MANIFOLDS with the restriction that the group operation maps a C^∞ map of the MANIFOLD into itself. Examples include O_3, $SU(n)$, and the LORENTZ GROUP.

see also COMPACT GROUP, LIE ALGEBRA, LIE GROUPOID, LIE-TYPE GROUP, NIL GEOMETRY, SOL GEOMETRY

References

Arfken, G. "Infinite Groups, Lie Groups." *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 251–252, 1985.

Chevalley, C. *Theory of Lie Groups.* Princeton, NJ: Princeton University Press, 1946.

Knapp, A. W. *Lie Groups Beyond an Introduction.* Boston, MA: Birkhäuser, 1996.

Lipkin, H. J. *Lie Groups for Pedestrians, 2nd ed.* Amsterdam, Netherlands: North-Holland, 1966.

Lie Groupoid

A GROUPOID G over B for which G and B are differentiable manifolds and α, β, and multiplication are differentiable maps. Furthermore, the derivatives of α and β are required to have maximal RANK everywhere. Here, α and β are maps from G onto $\mathbb{R}^2$ with $\alpha : (x, \gamma, y) \mapsto x$ and $\beta : (x, \gamma, y) \mapsto y$.

see also LIE ALGEBROID, NILPOTENT LIE GROUP, SEMISIMPLE LIE GROUP, SOLVABLE LIE GROUP

References

Weinstein, A. "Groupoids: Unifying Internal and External Symmetry." *Not. Amer. Math. Soc.* **43**, 744–752, 1996.

Lie Product

The multiplication operation corresponding to the LIE BRACKET.

Lie-Type Group

A finite analog of LIE GROUPS. The Lie-type groups include the CHEVALLEY GROUPS [$PSL(n,q)$, $PSU(n,q)$, $PSp(2n,q)$, $P\Omega^\epsilon(n,q)$], TWISTED CHEVALLEY GROUPS, and the TITS GROUP.

see also CHEVALLEY GROUPS, FINITE GROUP, LIE GROUP, LINEAR GROUP, ORTHOGONAL GROUP, SIMPLE GROUP, SYMPLECTIC GROUP, TITS GROUP, TWISTED CHEVALLEY GROUPS, UNITARY GROUP

References

Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas#lie`.

Liebmann's Theorem

A SPHERE is RIGID.

see also RIGID

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 377, 1993.

O'Neill, B. *Elementary Differential Geometry, 2nd ed.* New York: Academic Press, p. 262, 1997.

Life

The most well-known CELLULAR AUTOMATON, invented by John Conway and popularized in Martin Gardner's *Scientific American* column starting in October 1970. The game was originally played (i.e., successive generations were produced) by hand with counters, but implementation on a computer greatly increased the ease of exploring patterns.

The Life AUTOMATON is run by placing a number of filled cells on a 2-D grid. Each generation then switches cells on or off depending on the state of the cells that surround it. The rules are defined as follows. All eight of the cells surrounding the current one are checked to see if they are on or not. Any cells that are on are counted, and this count is then used to determine what will happen to the current cell.

1. Death: if the count is less than 2 or greater than 3, the current cell is switched off.
2. Survival: if (a) the count is exactly 2, or (b) the count is exactly 3 and the current cell is on, the current cell is left unchanged.
3. Birth: if the current cell is off and the count is exactly 3, the current cell is switched on.

Hensel gives a Java applet (`http://www.mindspring.com/~alanh/life/`) implementing the Game of Life on his web page.

A pattern which does not change from one generation to the next is known as a Still Life, and is said to have period 1. Conway originally believed that no pattern could produce an infinite number of cells, and offered a $50 prize to anyone who could find a counterexample before the end of 1970 (Gardner 1983, p. 216). Many counterexamples were subsequently found, including Guns and Puffer Trains.

A Life pattern which has no Father Pattern is known as a Garden of Eden (for obvious biblical reasons). The first such pattern was not found until 1971, and at least 3 are now known. It is not, however, known if a pattern exists which has a Father Pattern, but no Grandfather Pattern (Gardner 1983, p. 249).

Rather surprisingly, Gosper and J. H. Conway independently showed that Life can be used to generate a UNIVERSAL TURING MACHINE (Berlekamp *et al.* 1982, Gardner 1983, pp. 250–253).

Similar CELLULAR AUTOMATON games with different rules are HASHLIFE, HEXLIFE, and HIGHLIFE.

see also CELLULAR AUTOMATON, HASHLIFE, HEXLIFE, HIGHLIFE

References

"Alife online." `http://alife.santafe.edu/alife/topics/cas/ca-faq/lifefaq/lifefaq.html`.

Berlekamp, E. R.; Conway, J. H.; and Guy, R. K. "What Is Life." Ch. 25 in *Winning Ways, For Your Mathematical Plays, Vol. 2: Games in Particular*. London: Academic Press, 1982.

Callahan, P. "Patterns, Programs, and Links for Conway's Game of Life." `http://www.cs.jhu.edu/~callahan/lifepage.html`.

Flammenkamp, A. "Game of Life." `http://www.minet.uni-jena.de/~achim/gol.html`.

"The Game of Life." *Math Horizons*. p. 9, Spring 1994.

Gardner, M. "The Game of Life, Parts I–III." Chs. 20–22 in *Wheels, Life, and other Mathematical Amusements*. New York: W. H. Freeman, 1983.

Hensel, A. "A Brief Illustrated Glossary of Terms in Conway's Game of Life." `http://www.cs.jhu.edu/~callahan/glossary.html`.

Hensel, A. "PC Life Distribution." `http://www.mindspring.com/~alanh/lifep.zip`.

Hensel, A. "Conway's Game of Life." Includes a Java applet for the Game of Life. `http://www.mindspring.com/~alanh/life/`.

Koenig, H. "Game of Life Information." `http://www.halcyon.com/hkoenig/LifeInfo/LifeInfo.html`.

McIntosh, H. V. "A Zoo of Life Forms." `http://www.cs.cinvestav.mx/mcintosh/life.html`.

Poundstone, W. *The Recursive Universe: Cosmic Complexity and the Limits of Scientific Knowledge*. New York: Morrow, 1985.

Toffoli, T. and Margolus, N. *Cellular Automata Machines: A New Environment for Modeling*. Cambridge, MA: MIT Press, 1987.

Wainwright, R. T. "LifeLine." `http://members.aol.com/lifeline/life/lifepage.htm`.

Wainwright, R. T. *LifeLine: A Quarterly Newsletter for Enthusiasts of John Conway's Game of Life*. Nos. 1–11, 1971–1973.

Life Expectancy

An l_x table is a tabulation of numbers which is used to calculate life expectancies.

x	n_x	d_x	l_x	q_x	L_x	T_x	e_x
0	1000	200	1.00	0.20	0.90	2.70	2.70
1	800	100	0.80	0.12	0.75	1.80	2.25
2	700	200	0.70	0.29	0.60	1.05	1.50
3	500	300	0.50	0.60	0.35	0.45	0.90
4	200	200	0.20	1.00	0.10	0.10	0.50
5	0	0	0.00	—	0.00	0.00	—
$\sum$		1000	2.70				

x : Age category ($x = 0, 1, \ldots, k$). These values can be in any convenient units, but must be chosen so that no observed lifespan extends past category $k-1$.

n_x : Census size, defined as the number of individuals in the study population who survive to the beginning of age category x. Therefore, $n_0 = N$ (the total population size) and $n_k = 0$.

d_x : $= n_x - n_{x+1}$; $\sum_{i=0}^k d_i = n_0$. Crude death rate, which measures the number of individuals who die within age category x.

l_x : $= n_x/n_0$. Survivorship, which measures the *proportion* of individuals who survive to the beginning of age category x.

q_x : $= d_x/n_x$; $q_{k-1} = 1$. Proportional death rate, or "risk," which measures the proportion of individuals surviving to the beginning of age category x who die within that category.

L_x : $= (l_x + l_{x+1})/2$. Midpoint survivorship, which measures the proportion of individuals surviving to the *midpoint* of age category x. Note that the simple averaging formula must be replaced by a more complicated expression if survivorship is nonlinear within age categories. The sum $\sum_{i=0}^k L_x$ gives the total number of age categories lived by the entire study population.

T_x : $= T_{x-1} - L_{x-1}$; $T_0 = \sum_{i=0}^k L_x$. Measures the total number of age categories left to be lived by all individuals who survive to the beginning of age category x.

e_x : $= T_x/l_x$; $e_{k-1} = 1/2$. Life expectancy, which is the mean number of age categories remaining until death for individuals surviving to the beginning of age category x.

For all x, $e_{x+1} + 1 > e_x$. This means that the total expected lifespan increases monotonically. For instance, in the table above, the one-year-olds have an average age at death of $2.25 + 1 = 3.25$, compared to 2.70 for newborns. In effect, the age of death of older individuals is a distribution conditioned on the fact that they have survived to their present age.

It is common to study survivorship as a semilog plot of l_x vs. x, known as a SURVIVORSHIP CURVE. A so-called $l_x m_x$ table can be used to calculate the mean generation time of a population. Two $l_x m_x$ tables are illustrated below.

Population 1

x	l_x	m_x	$l_x m_x$	$x l_x m_x$
0	1.00	0.00	0.00	0.00
1	0.70	0.50	0.35	0.35
2	0.50	1.50	0.75	1.50
3	0.20	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00
			$R_0 = 1.10$	$\sum = 1.85$

$$T = \frac{\sum x l_x m_x}{\sum l_x m_x} = \frac{1.85}{1.10} = 1.68$$

$$r = \frac{\ln R_0}{T} = \frac{\ln 1.10}{1.68} = 0.057.$$

Population 2

x	l_x	m_x	$l_x m_x$	$x l_x m_x$
0	1.00	0.00	0.00	0.00
1	0.70	0.00	0.00	0.00
2	0.50	2.00	1.00	2.00
3	0.20	0.50	0.10	0.30
4	0.00	0.00	0.00	0.00
			$R_0 = 1.10$	$\sum = 2.30$

$$T = \frac{\sum x l_x m_x}{\sum l_x m_x} = \frac{2.30}{1.10} = 2.09$$

$$r = \frac{\ln R_0}{T} = \frac{\ln 1.10}{2.09} = 0.046.$$

x : Age category ($x = 0, 1, \ldots, k$). These values can be in any convenient units, but must be chosen so that no observed lifespan extends past category $k - 1$ (as in an l_x table).

l_x : $= n_x/n_0$. Survivorship, which measures the *proportion* of individuals who survive to the beginning of age category x (as in an l_x table).

m_x : The average number of offspring produced by an individual in age category x *while in that age category.* $\sum_{i=0}^{k} m_x$ therefore represents the average lifetime number of offspring produced by an individual of maximum lifespan.

$l_x m_x$: The average number of offspring produced by an individual within age category x weighted by the probability of surviving to the beginning of that age category. $\sum_{i=0}^{k} l_x m_x$ therefore represents the average lifetime number of offspring produced by a member of the study population. It is called the net reproductive rate per generation and is often denoted R_0.

$x l_x m_x$: A column weighting the offspring counted in the previous column by their parents' age when they were born. Therefore, the ratio $T = \sum(x l_x m_x)/\sum(l_x m_x)$ is the mean generation time of the population.

The MALTHUSIAN PARAMETER r measures the reproductive rate per unit time and can be calculated as $r = (\ln R_0)/T$. For an exponentially increasing population, the population size $N(t)$ at time t is then given by

$$N(t) = N_0 e^{rt}.$$

In the above two tables, the populations have identical reproductive rates of $R_0 = 1.10$. However, the shift toward later reproduction in population 2 increases the generation time, thus slowing the rate of POPULATION GROWTH. Often, a slight delay of reproduction decreases POPULATION GROWTH more strongly than does even a fairly large reduction in reproductive rate.

see also GOMPERTZ CURVE, LOGISTIC GROWTH CURVE, MAKEHAM CURVE, MALTHUSIAN PARAMETER, POPULATION GROWTH, SURVIVORSHIP CURVE

Lift

Given a MAP f from a SPACE X to a SPACE Y and another MAP g from a SPACE Z to a SPACE Y, a lift is a MAP h from X to Z such that $gh = f$. In other words, a lift of f is a MAP h such that the diagram (shown below) commutes.

$$\begin{tikzcd} & Z \arrow[d, "g"] \\ X \arrow[ur, dashed, "h"] \arrow[r, "f"] & Y \end{tikzcd}$$

If f is the identity from Y to Y, a MANIFOLD, and if g is the bundle projection from the TANGENT BUNDLE to Y, the lifts are precisely VECTOR FIELDS. If g is a bundle projection from any FIBER BUNDLE to Y, then lifts are precisely sections. If f is the identity from Y to Y, a MANIFOLD, and g a projection from the orientation double cover of Y, then lifts exist IFF Y is an orientable MANIFOLD.

If f is a MAP from a CIRCLE to Y, an n-MANIFOLD, and g the bundle projection from the FIBER BUNDLE of alternating n-FORMS on Y, then lifts always exist IFF Y is orientable. If f is a MAP from a region in the COMPLEX PLANE to the COMPLEX PLANE (complex analytic), and if g is the exponential MAP, lifts of f are precisely LOGARITHMS of f.

see also LIFTING PROBLEM

Lifting Problem

Given a MAP f from a SPACE X to a SPACE Y and another MAP g from a SPACE Z to a SPACE Y, does there exist a MAP h from X to Z such that $gh = f$? If such a map h exists, then h is called a LIFT of f.

see also EXTENSION PROBLEM, LIFT

Ligancy

see KISSING NUMBER

Likelihood

The hypothetical PROBABILITY that an event which has already occurred would yield a specific outcome. The concept differs from that of a probability in that a probability refers to the occurrence of future events, while a likelihood refers to past events with known outcomes.

see also LIKELIHOOD RATIO, MAXIMUM LIKELIHOOD, NEGATIVE LIKELIHOOD RATIO, PROBABILITY

Likelihood Ratio

A quantity used to test NESTED HYPOTHESES. Let H' be a NESTED HYPOTHESIS with n' DEGREES OF FREEDOM within H (which has n DEGREES OF FREEDOM), then calculate the MAXIMUM LIKELIHOOD of a given outcome, first given H', then given H. Then

$$\mathrm{LR} = \frac{[\text{likelihood H'}]}{[\text{likelihood H}]}.$$

Comparison of this ratio to the critical value of the CHI-SQUARED DISTRIBUTION with $n - n'$ DEGREES OF FREEDOM then gives the SIGNIFICANCE of the increase in LIKELIHOOD.

The term likelihood ratio is also used (especially in medicine) to test nonnested complementary hypotheses as follows,

$$\mathrm{LR} = \frac{[\text{true positive rate}]}{[\text{false positive rate}]} = \frac{[\text{sensitivity}]}{1 - [\text{specificity}]}.$$

see also NEGATIVE LIKELIHOOD RATIO, SENSITIVITY, SPECIFICITY

Limaçon

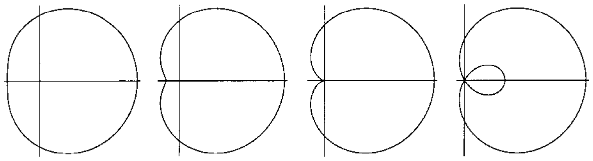

The limaçon is a polar curve of the form

$$r = b + a\cos\theta$$

also called the LIMAÇON OF PASCAL. It was first investigated by Dürer, who gave a method for drawing it in *Underweysung der Messung* (1525). It was rediscovered by Étienne Pascal, father of Blaise Pascal, and named by Gilles-Personne Roberval in 1650 (MacTutor Archive). The word "limaçon" comes from the Latin *limax*, meaning "snail."

If $b \geq 2a$, we have a convex limaçon. If $2a > b > a$, we have a dimpled limaçon. If $b = a$, the limaçon degenerates to a CARDIOID. If $b < a$, we have limaçon with an inner loop. If $b = a/2$, it is a TRISECTRIX (but *not* the MACLAURIN TRISECTRIX) with inner loop of AREA

$$A_{\text{inner loop}} = \tfrac{1}{4}a^2\left(\pi - 3\sqrt{\frac{3}{2}}\right),$$

and AREA between the loops of

$$A_{\text{between loops}} = \tfrac{1}{4}a^2(\pi + 3\sqrt{3})$$

(MacTutor Archive). The limaçon is an ANALLAGMATIC CURVE, and is also the CATACAUSTIC of a CIRCLE when the RADIANT POINT is a finite (NONZERO) distance from the CIRCUMFERENCE, as shown by Thomas de St. Laurent in 1826 (MacTutor Archive).

see also CARDIOID

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 113–117, 1972.

Lee, X. "Limacon of Pascal." `http://www.best.com/~xah/SpecialPlaneCurves_dir/LimaconOfPascal_dir/limaconOfPascal.html`.

Lee, X. "Limacon Graphics Gallery." `http://www.best.com/~xah/SpecialPlaneCurves_dir/LimaconGGallery_dir/limaconGGallery.html`.

Lockwood, E. H. "The Limaçon." Ch. 5 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 44–51, 1967.

MacTutor History of Mathematics Archive. "Limacon of Pascal." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Limacon.html`.

Yates, R. C. "Limacon of Pascal." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 148–151, 1952.

Limaçon Evolute

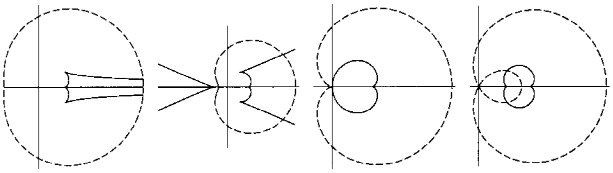

The CATACAUSTIC of a CIRCLE for a RADIANT POINT is the limaçon evolute. It has parametric equations

$$x = \frac{a[4a^2 + 4b^2 + 9ab\cos t - ab\cos(3t)]}{4(2a^2 + b^2 + 3ab\cos t)}$$

$$y = \frac{a^2 b\sin^3 t}{2a^2 + b^2 + 3ab\cos t}.$$

Limaçon of Pascal

see LIMAÇON

Limit

A function $f(z)$ is said to have a limit $\lim_{z\to a} f(z) = c$ if, for all $\epsilon > 0$, there exists a $\delta > 0$ such that $|f(z) - c| < \epsilon$ whenever $0 < |z - a| < \delta$.

A LOWER LIMIT

$$\text{lower} \lim_{n\to\infty} S_n = \varliminf_{n\to\infty} S_n = h$$

is said to exist if, for every $\epsilon > 0$, $|S_n - h| < \epsilon$ for infinitely many values of n and if no number less than h has this property.

An UPPER LIMIT

$$\text{upper} \lim_{n\to\infty} S_n = \varlimsup_{n\to\infty} S_n = k$$

is said to exist if, for every $\epsilon > 0$, $|S_n - k| < \epsilon$ for infinitely many values of n and if no number larger than k has this property.

Indeterminate limit forms of types ∞/∞ and $0/0$ can be computed with L'HOSPITAL'S RULE. Types $0 \cdot \infty$ can be converted to the form $0i/0$ by writing

$$f(x)g(x) = \frac{f(x)}{1/g(x)}.$$

Types 0^0, ∞^0, and 1^∞ are treated by introducing a dependent variable $y = f(x)g(x)$, then calculating $\lim \ln y$. The original limit then equals $e^{\lim \ln y}$.

see also CENTRAL LIMIT THEOREM, CONTINUOUS, DISCONTINUITY, L'HOSPITAL'S RULE, LOWER LIMIT, UPPER LIMIT

References

Courant, R. and Robbins, H. "Limits. Infinite Geometrical Series." §2.2.3 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 63–66, 1996.

Limit Comparison Test

Let $\sum a_k$ and $\sum b_k$ be two SERIES with POSITIVE terms and suppose

$$\lim_{k\to\infty} \frac{a_k}{b_k} = \rho.$$

If ρ is finite and $\rho > 0$, then the two SERIES both CONVERGE or DIVERGE.

see also CONVERGENCE TESTS

Limit Cycle

An attracting set to which orbits or trajectories converge and upon which trajectories are periodic.

see also HOPF BIFURCATION

Limit Point

A number x such that for all $\epsilon > 0$, there exists a member of the SET y different from x such that $|y - x| < \epsilon$. The topological definition of limit point P of A is that P is a point such that every OPEN SET around it intersects A.

see also CLOSED SET, OPEN SET

References

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 25–26, 1991.

Lin's Method

An ALGORITHM for finding ROOTS for QUARTIC EQUATIONS with COMPLEX ROOTS.

References

Acton, F. S. *Numerical Methods That Work, 2nd printing.* Washington, DC: Math. Assoc. Amer., pp. 198–199, 1990.

Lindeberg Condition

A SUFFICIENT condition on the LINDEBERG-FELLER CENTRAL LIMIT THEOREM. Given random variates X_1, X_2, ..., let $\langle X_i \rangle = 0$, the VARIANCE ${\sigma_i}^2$ of X_i be finite, and VARIANCE of the distribution consisting of a sum of X_is

$$S_n \equiv X_1 + X_2 + \ldots + X_n \tag{1}$$

be

$${s_n}^2 \equiv \sum_{i=1}^n {\sigma_i}^2. \tag{2}$$

Let

$$\Lambda_n(\epsilon) \equiv \sum_{k=1}^n \left\langle \left(\frac{X_k}{s_n}\right)^2 \middle| \frac{|X_k|}{s_n} \geq \epsilon \right\rangle, \tag{3}$$

then the Lindeberg condition is

$$\lim_{n\to\infty} \Lambda_n(\epsilon) = 0 \tag{4}$$

for all $\epsilon > 0$.

see also FELLER-LÉVY CONDITION

References

Zabell, S. L. "Alan Turing and the Central Limit Theorem." *Amer. Math. Monthly* **102**, 483–494, 1995.

Lindeberg-Feller Central Limit Theorem

If the random variates X_1, X_2, ... satisfy the LINDEBERG CONDITION, then for all $a < b$,

$$\lim_{n\to\infty} P\left(a < \frac{S_n}{s_n} < b\right) = \Phi(b) - \Phi(a),$$

where Φ is the NORMAL DISTRIBUTION FUNCTION.

see also CENTRAL LIMIT THEOREM, FELLER-LÉVY CONDITION, NORMAL DISTRIBUTION FUNCTION

References

Zabell, S. L. "Alan Turing and the Central Limit Theorem." *Amer. Math. Monthly* **102**, 483–494, 1995.

Lindelof's Theorem

The SURFACE OF REVOLUTION generated by the external CATENARY between a fixed point a and its conjugate on the ENVELOPE of the CATENARY through the fixed point is equal in AREA to the surface of revolution generated by its two Lindelof TANGENTS, which cross the axis of rotation at the point a and are calculable from the position of the points and CATENARY.

see also CATENARY, ENVELOPE, SURFACE OF REVOLUTION

Lindemann-Weierstraß Theorem

If α_1, ..., α_n are linearly independent over $\mathbb{Q}$, then e^{α_1}, ..., e^{α_n} are algebraically independent over $\mathbb{Q}$.

see also HERMITE-LINDEMANN THEOREM

Lindenmayer System

A STRING REWRITING system which can be used to generate FRACTALS with DIMENSION between 1 and 2. The term L-SYSTEM is often used as an abbreviation.

see also ARROWHEAD CURVE, DRAGON CURVE EXTERIOR SNOWFLAKE, FRACTAL, HILBERT CURVE, KOCH SNOWFLAKE, PEANO CURVE, PEANO-GOSPER CURVE, SIERPIŃSKI CURVE, STRING REWRITING

References

Dickau, R. M. "Two-dimensional L-systems." `http://forum.swarthmore.edu/advanced/robertd/lsys2d.html`.

Prusinkiewicz, P. and Hanan, J. *Lindenmayer Systems, Fractal, and Plants.* New York: Springer-Verlag, 1989.

Prusinkiewicz, P. and Lindenmayer, A. *The Algorithmic Beauty of Plants.* New York: Springer-Verlag, 1990.

Stevens, R. T. *Fractal Programming in C.* New York: Holt, 1989.

Wagon, S. "Recursion via String Rewriting." §6.2 in *Mathematica in Action.* New York: W. H. Freeman, pp. 190–196, 1991.

Line

Euclid defined a line as a "breadthless length," and a straight line as a line which "lies evenly with the points on itself" (Kline 1956, Dunham 1990). Lines are intrinsically 1-dimensional objects, but may be embedded in higher dimensional SPACES. An infinite line passing through points A and B is denoted $\overleftrightarrow{AB}$. A LINE SEGMENT terminating at these points is denoted $\overline{AB}$. A line is sometimes called a STRAIGHT LINE or, more archaically, a RIGHT LINE (Casey 1893), to emphasize that it has no curves anywhere along its length.

Consider first lines in a 2-D PLANE. The line with x-INTERCEPT a and y-INTERCEPT b is given by the *intercept form*

$$\frac{x}{a}+\frac{y}{b}=1. \tag{1}$$

The line through (x_1, y_1) with SLOPE m is given by the *point-slope form*

$$y-y_1=m(x-x_1). \tag{2}$$

The line with y-intercept b and slope m is given by the *slope-intercept form*

$$y=mx+b. \tag{3}$$

The line through (x_1, y_1) and (x_2, y_2) is given by the *two point form*

$$y-y_1=\frac{y_2-y_1}{x_2-x_1}(x-x_1). \tag{4}$$

Other forms are

$$a(x-x_1)+b(y-y_1)=0 \tag{5}$$

$$ax+by+c=0 \tag{6}$$

$$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix}=0. \tag{7}$$

A line in 2-D can also be represented as a VECTOR. The VECTOR along the line

$$ax+by=0 \tag{8}$$

is given by

$$t\begin{bmatrix} -b \\ a \end{bmatrix}, \tag{9}$$

where $t \in \mathbb{R}$. Similarly, VECTORS of the form

$$t\begin{bmatrix} a \\ b \end{bmatrix} \tag{10}$$

are PERPENDICULAR to the line. Three points lie on a line if

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}=0. \tag{11}$$

The ANGLE between lines

$$A_1x+B_1y+C_1=0 \tag{12}$$

$$A_2x+B_2y+C_2=0 \tag{13}$$

is

$$\tan\theta=\frac{A_1B_2-A_2B_1}{A_1A_2+B_1B_2}. \tag{14}$$

The line joining points with TRILINEAR COORDINATES $\alpha_1 : \beta_1 : \gamma_1$ and $\alpha_2 : \beta_2 : \gamma_2$ is the set of point $\alpha : \beta : \gamma$ satisfying

$$\begin{vmatrix} \alpha & \beta & \gamma \\ \alpha_1 & \beta_1 & \gamma_1 \\ \alpha_2 & \beta_2 & \gamma_2 \end{vmatrix}=0 \tag{15}$$

$$(\beta_1\gamma_2-\gamma_1\beta_2)\alpha+(\gamma_1\alpha_2-\alpha_1\gamma_2)\beta+(\alpha_1\beta_2-\beta_1\alpha_2)\gamma=0. \tag{16}$$

Three lines CONCUR if their TRILINEAR COORDINATES satisfy

$$l_1\alpha+m_1\beta+n_1\gamma=0 \tag{17}$$

$$l_2\alpha+m_2\beta+n_2\gamma=0 \tag{18}$$

$$l_3\alpha+m_3\beta+n_3\gamma=0, \tag{19}$$

in which case the point is

$$m_2n_3-n_2m_3 : n_2l_3-l_2n_3 : l_2m_3-m_2l_3, \tag{20}$$

or if the COEFFICIENTS of the lines

$$A_1x+B_1y+C_1=0 \tag{21}$$

$$A_2x+B_2y+C_2=0 \tag{22}$$

$$A_3x+B_3y+C_3=0 \tag{23}$$

satisfy

$$\begin{vmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A_3 & B_3 & C_3 \end{vmatrix}=0. \tag{24}$$

Two lines CONCUR if their TRILINEAR COORDINATES satisfy

$$\begin{vmatrix} l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \\ l_3 & m_3 & n_3 \end{vmatrix}=0. \tag{25}$$

The line through P_1 is the direction (a_1, b_1, c_1) and the line through P_2 in direction (a_2, b_2, c_2) intersect IFF

$$\begin{vmatrix} x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix} = 0. \tag{26}$$

The line through a point $\alpha' : \beta' : \gamma'$ PARALLEL to

$$l\alpha + m\beta + n\gamma = 0 \tag{27}$$

is

$$\begin{vmatrix} \alpha & \beta & \gamma \\ \alpha' & \beta' & \gamma' \\ bn - cm & cl - an & am - bl \end{vmatrix} = 0. \tag{28}$$

The lines

$$l\alpha + m\beta + n\gamma = 0 \tag{29}$$

$$l'\alpha + m'\beta + n'\gamma = 0 \tag{30}$$

are PARALLEL if

$$a(mn' - nm') + b(nl' - ln') + c(lm' - ml') = 0 \tag{31}$$

for all (a, b, c), and PERPENDICULAR if

$$\begin{aligned} &2abc(ll' + mm' + nn') - (mn' + m'm)\cos A \\ &\quad -(nl' + n'l)\cos B - (lm' + l'm)\cos C = 0 \end{aligned} \tag{32}$$

for all (a, b, c) (Sommerville 1924). The line through a point $\alpha' : \beta' : \gamma'$ PERPENDICULAR to (32) is given by

$$\begin{vmatrix} \alpha & \beta & \gamma \\ \alpha' & \beta' & \gamma' \\ l - m\cos C & m - n\cos A & n - l\cos B \\ -n\cos B & -l\cos C & -m\cos A \end{vmatrix} = 0. \tag{33}$$

In 3-D SPACE, the line passing through the point (x_0, y_0, z_0) and PARALLEL to the NONZERO VECTOR

$$\mathbf{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \tag{34}$$

has parametric equations

$$\begin{cases} x = x_0 + at \\ y = y_0 + bt \\ z = z_0 + ct. \end{cases} \tag{35}$$

see also ASYMPTOTE, BROCARD LINE, COLLINEAR, CONCUR, CRITICAL LINE, DESARGUES' THEOREM, ERDŐS-ANNING THEOREM, LINE SEGMENT, ORDINARY LINE, PENCIL, POINT, POINT-LINE DISTANCE—2-D, POINT-LINE DISTANCE—3-D, PLANE, RANGE (LINE SEGMENT), RAY, SOLOMON'S SEAL LINES, STEINER SET, STEINER'S THEOREM, SYLVESTER'S LINE PROBLEM

References

Casey, J. "The Right Line." Ch. 2 in *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions, with Numerous Examples, 2nd ed., rev. enl.* Dublin: Hodges, Figgis, & Co., pp. 30–95, 1893.

Dunham, W. *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, p. 32, 1990.

Kline, M. "The Straight Line." *Sci. Amer.* **156**, 105–114, Mar. 1956.

MacTutor History of Mathematics Archive. "Straight Line." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Straight.html`.

Sommerville, D. M. Y. *Analytical Conics.* London: G. Bell, p. 186, 1924.

Spanier, J. and Oldham, K. B. "The Linear Function $bx + c$ and Its Reciprocal." Ch. 7 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 53–62, 1987.

Line Bisector

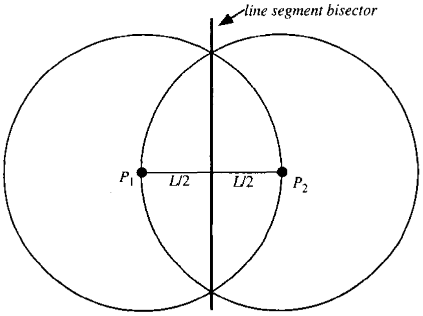

The line bisecting a given LINE SEGMENT P_1P_2 can be constructed geometrically, as illustrated above.

References

Courant, R. and Robbins, H. "How to Bisect a Segment and Find the Center of a Circle with the Compass Alone." §3.4.4 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 145–146, 1996.

Dixon, R. *Mathographics.* New York: Dover, p. 22, 1991.

Line of Curvature

A curve on a surface whose tangents are always in the direction of PRINCIPAL CURVATURE. The equation of the lines of curvature can be written

$$\begin{vmatrix} g_{11} & g_{12} & g_{22} \\ b_{11} & b_{12} & b_{22} \\ du^2 & -du\,dv & dv^2 \end{vmatrix} = 0,$$

where g and b are the COEFFICIENTS of the first and second FUNDAMENTAL FORMS.

see also DUPIN'S THEOREM, FUNDAMENTAL FORMS, PRINCIPAL CURVATURES

Line Element

Also known as the first FUNDAMENTAL FORM

$$ds^2 = g_{ab}\,dx^a\,dx^b.$$

In the principal axis frame for 3-D,

$$ds^2 = g_{aa}(dx^a)^2 + g_{bb}(dx^b)^2 + g_{cc}(dx^c)^2.$$

At ORDINARY POINTS on a surface, the line element is positive definite.

see also AREA ELEMENT, FUNDAMENTAL FORMS, VOLUME ELEMENT

Line Graph

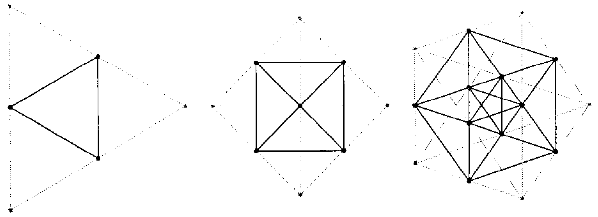

A LINE GRAPH $L(G)$ (also called an INTERCHANGE GRAPH) of a graph G is obtained by associating a vertex with each edge of the graph and connecting two vertices with an edge IFF the corresponding edges of G meet at one or both endpoints. In the three examples above, the original graphs are the COMPLETE GRAPHS K_3, K_4, and K_5 shown in gray, and their line graphs are shown in black.

References

Saaty, T. L. and Kainen, P. C. "Line Graphs." §4-3 in *The Four-Color Problem: Assaults and Conquest.* New York: Dover, pp. 108–112, 1986.

Line at Infinity

The straight line on which all POINTS AT INFINITY lie. The line at infinity is given in terms of TRILINEAR COORDINATES by

$$a\alpha + b\beta + c\gamma = 0,$$

which follows from the fact that a REAL TRIANGLE will have POSITIVE AREA, and therefore that

$$2\Delta = a\alpha + b\beta + c\gamma > 0.$$

Instead of the three reflected segments concurring for the ISOGONAL CONJUGATE of a point X on the CIRCUMCIRCLE of a TRIANGLE, they become parallel (and can be considered to meet at infinity). As X varies around the CIRCUMCIRCLE, X^{-1} varies through a line called the line at infinity. Every line is PERPENDICULAR to the line at infinity.

see also POINT AT INFINITY

Line Integral

The line integral on a curve $\boldsymbol{\sigma}$ is defined by

$$\int_{\boldsymbol{\sigma}} \mathbf{F} \cdot d\mathbf{s} = \int_a^b \mathbf{F}(\boldsymbol{\sigma}(t)) \cdot \boldsymbol{\sigma}'(t)\,dt \tag{1}$$

$$= \int_C F_1\,dx + F_2\,dy + F_3\,dz, \tag{2}$$

where

$$\mathbf{F} \equiv \begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix}. \tag{3}$$

If $\nabla \cdot \mathbf{F} = 0$ (i.e., it is a DIVERGENCELESS FIELD), then the line integral is path independent and

$$\int_{(a,b,c)}^{(x,y,z)} F_1\,dx + F_2\,dy + F_3\,dz$$

$$= \int_{(a,b,c)}^{(x,b,c)} F_1\,dx + \int_{(x,b,c)}^{(x,y,c)} F_2\,dy + \int_{(x,y,c)}^{(x,y,z)} F_3\,dz. \tag{4}$$

For z COMPLEX, $\gamma : z = z(t)$, and $t \in [a,b]$,

$$\int_\gamma f\,dz = \int_a^b f(z(t))z'(t)\,dt. \tag{5}$$

see also CONTOUR INTEGRAL, PATH INTEGRAL

Line Segment

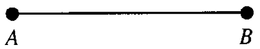

A closed interval corresponding to a FINITE portion of an infinite LINE. Line segments are generally labelled with two letters corresponding to their endpoints, say A and B, and then written AB. The length of the line segment is indicated with an overbar, so the length of the line segment AB would be written $\overline{AB}$.

Curiously, the number of points in a line segment (ALEPH-1; $\aleph_1$) is equal to that in an entire 1-D SPACE (a LINE), and also to the number of points in an n-D SPACE, as first recognized by Georg Cantor.

see also ALEPH-1 ($\aleph_1$), COLLINEAR, CONTINUUM, LINE, RAY

Line Space

see LIOUVILLE SPACE

Linear Algebra

The study of linear sets of equations and their transformation properties. Linear algebra allows the analysis of ROTATIONS in space, LEAST SQUARES FITTING, solution of coupled differential equations, determination of a circle passing through three given points, as well as many other other problems in mathematics, physics, and engineering.

The MATRIX and DETERMINANT are extremely useful tools of linear algebra. One central problem of linear algebra is the solution of the matrix equation

$$\mathsf{A}\mathbf{x} = \mathbf{b}$$

for $\mathbf{x}$. While this can, in theory, be solved using a MATRIX INVERSE

$$\mathbf{x} = \mathsf{A}^{-1}\mathbf{b},$$

other techniques such as GAUSSIAN ELIMINATION are numerically more robust.

see also CONTROL THEORY, CRAMER'S RULE, DETERMINANT, GAUSSIAN ELIMINATION, LINEAR TRANSFORMATION, MATRIX, VECTOR

References

Ayres, F. Jr. *Theory and Problems of Matrices.* New York: Schaum, 1962.

Banchoff, T. and Wermer, J. *Linear Algebra Through Geometry, 2nd ed.* New York: Springer-Verlag, 1992.

Bellman, R. E. *Introduction to Matrix Analysis, 2nd ed.* New York: McGraw-Hill, 1970.

Faddeeva, V. N. *Computational Methods of Linear Algebra.* New York: Dover, 1958.

Golub, G. and van Loan, C. *Matrix Computations, 3rd ed.* Baltimore, MD: Johns Hopkins University Press, 1996.

Halmos, P. R. *Linear Algebra Problem Book.* Providence, RI: Math. Assoc. Amer., 1995.

Lang, S. *Introduction to Linear Algebra, 2nd ed.* New York: Springer-Verlag, 1997.

Marcus, M. and Minc, H. *Introduction to Linear Algebra.* New York: Dover, 1988.

Marcus, M. and Minc, H. *A Survey of Matrix Theory and Matrix Inequalities.* New York: Dover, 1992.

Marcus, M. *Matrices and Matlab: A Tutorial.* Englewood Cliffs, NJ: Prentice-Hall, 1993.

Mirsky, L. *An Introduction to Linear Algebra.* New York: Dover, 1990.

Muir, T. *A Treatise on the Theory of Determinants.* New York: Dover, 1960.

Nash, J. C. *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation, 2nd ed.* Bristol, England: Adam Hilger, 1990.

Strang, G. *Linear Algebra and its Applications, 3rd ed.* Philadelphia, PA: Saunders, 1988.

Strang, G. *Introduction to Linear Algebra.* Wellesley, MA: Wellesley-Cambridge Press, 1993.

Strang, G. and Borre, K. *Linear Algebra, Geodesy, & GPS.* Wellesley, MA: Wellesley-Cambridge Press, 1997.

Linear Approximation

A linear approximation to a function $f(x)$ at a point x_0 can be computed by taking the first term in the TAYLOR SERIES

$$f(x_0+\Delta x)=f(x_0)+f'(x_0)\Delta x+\ldots.$$

see also MACLAURIN SERIES, TAYLOR SERIES

Linear Code

A linear code over a FINITE FIELD with q elements F_q is a linear SUBSPACE $C\subset {F_q}^n$. The vectors forming the SUBSPACE are called code words. When code words are chosen such that the distance between them is maximized, the code is called error-correcting since slightly garbled vectors can be recovered by choosing the nearest code word.

see also CODE, CODING THEORY, ERROR-CORRECTING CODE, GRAY CODE, HUFFMAN CODING, ISBN

Linear Congruence

A linear congruence

$$ax\equiv b \pmod m$$

is solvable IFF the CONGRUENCE

$$b\equiv 0 \pmod{(a,m)}$$

is solvable, where $d\equiv(a,m)$ is the GREATEST COMMON DIVISOR, in which case the solutions are x_0, x_0+m/d, x_0+2m/d, $\ldots$, $x_0+(d-1)m/d$, where $x_0<m/d$. If $d=1$, then there is only one solution.

see also CONGRUENCE, QUADRATIC CONGRUENCE

Linear Congruence Method

A METHOD for generating RANDOM (PSEUDORANDOM) numbers using the linear RECURRENCE RELATION

$$X_{n+1}=aX_n+c \pmod m,$$

where a and c must assume certain fixed values and X_0 is an initial number known as the SEED.

see also PSEUDORANDOM NUMBER, RANDOM NUMBER, SEED

References

Pickover, C. A. "Computers, Randomness, Mind, and Infinity." Ch. 31 in *Keys to Infinity.* New York: W. H. Freeman, pp. 233–247, 1995.

Linear Equation

An algebraic equation of the form

$$y=ax+b$$

involving only a constant and a first-order (linear) term.

see also LINE, POLYNOMIAL, QUADRATIC EQUATION

Linear Equation System

When solving a system of n linear equations with $k>n$ unknowns, use MATRIX operations to solve the system as far as possible. Then solve for the first $(k-n)$ components in terms of the last n components to find the solution space.

Linear Extension

A linear extension of a PARTIALLY ORDERED SET P is a PERMUTATION of the elements p_1, p_2, $\ldots$ of P such that $i<j$ IMPLIES $p_i<p_j$. For example, the linear extensions of the PARTIALLY ORDERED SET $((1,2),(3,4))$ are 1234, 1324, 1342, 3124, 3142, and 3412, all of which have 1 before 2 and 3 before 4.

References

Brightwell, G. and Winkler, P. "Counting Linear Extensions." *Order* **8**, 225–242, 1991.

Preusse, G. and Ruskey, F. "Generating Linear Extensions Fast." *SIAM J. Comput.* **23**, 373–386, 1994.

Ruskey, F. "Information on Linear Extension." `http://sue.csc.uvic.ca/~cos/inf/pose/LinearExt.html`.

Varol, Y. and Rotem, D. "An Algorithm to Generate All Topological Sorting Arrangements." *Comput. J.* **24**, 83–84, 1981.

Linear Fractional Transformation

see MÖBIUS TRANSFORMATION

Linear Group

see GENERAL LINEAR GROUP, LIE-TYPE GROUP, PROJECTIVE GENERAL LINEAR GROUP, PROJECTIVE SPECIAL LINEAR GROUP, SPECIAL LINEAR GROUP

References

Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas#lin`.

Linear Group Theorem

Any linear system of point-groups on a curve with only ordinary singularities may be cut by ADJOINT CURVES.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, pp. 122 and 251, 1959.

Linear Operator

An operator $\tilde{L}$ is said to be linear if, for every pair of functions f and g and SCALAR t,

$$\tilde{L}(f+g)=\tilde{L}f+\tilde{L}g$$

and

$$\tilde{L}(tf)=t\tilde{L}f.$$

see also LINEAR TRANSFORMATION, OPERATOR

Linear Ordinary Differential Equation

see ORDINARY DIFFERENTIAL EQUATION—FIRST-ORDER, ORDINARY DIFFERENTIAL EQUATION—SECOND-ORDER

Linear Programming

The problem of maximizing a linear function over a convex polyhedron, also known as OPERATIONS RESEARCH, OPTIMIZATION THEORY, or CONVEX OPTIMIZATION THEORY. It can be solved using the SIMPLEX METHOD (Wood and Dantzig 1949, Dantzig 1949) which runs along EDGES of the visualization solid to find the best answer.

In 1979, L. G. Khachian found a $\mathcal{O}(x^5)$ POLYNOMIAL-time ALGORITHM. A much more efficient POLYNOMIAL-time ALGORITHM was found by Karmarkar (1984). This method goes through the middle of the solid and then transforms and warps, and offers many advantages over the simplex method.

see also CRISS-CROSS METHOD, ELLIPSOIDAL CALCULUS, KUHN-TUCKER THEOREM, LAGRANGE MULTIPLIER, VERTEX ENUMERATION

References

Bellman, R. and Kalaba, R. *Dynamic Programming and Modern Control Theory.* New York: Academic Press, 1965.
Dantzig, G. B. "Programming of Interdependent Activities. II. Mathematical Model." *Econometrica* **17**, 200–211, 1949.
Dantzig, G. B. *Linear Programming and Extensions.* Princeton, NJ: Princeton University Press, 1963.
Greenberg, H. J. "Mathematical Programming Glossary." `http://www-math.cudenver.edu/~hgreenbe/glossary/glossary.html`.
Karloff, H. *Linear Programming.* Boston, MA: Birkhäuser, 1991.
Karmarkar, N. "A New Polynomial-Time Algorithm for Linear Programming." *Combinatorica* **4**, 373–395, 1984.
Pappas, T. "Projective Geometry & Linear Programming." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 216–217, 1989.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Linear Programming and the Simplex Method." §10.8 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 423–436, 1992.
Sultan, A. *Linear Programming: An Introduction with Applications.* San Diego, CA: Academic Press, 1993.
Tokhomirov, V. M. "The Evolution of Methods of Convex Optimization." *Amer. Math. Monthly* **103**, 65–71, 1996.
Wood, M. K. and Dantzig, G. B. "Programming of Interdependent Activities. I. General Discussion." *Econometrica* **17**, 193–199, 1949.
Yudin, D. B. and Nemirovsky, A. S. *Problem Complexity and Method Efficiency in Optimization.* New York: Wiley, 1983.

Linear Recurrence Sequence

see RECURRENCE SEQUENCE

Linear Regression

The fitting of a straight LINE through a given set of points according to some specified goodness-of-fit criterion. The most common form of linear regression is LEAST SQUARES FITTING.

see also LEAST SQUARES FITTING, MULTIPLE REGRESSION, NONLINEAR LEAST SQUARES FITTING

References

Edwards, A. L. *An Introduction to Linear Regression and Correlation.* San Francisco, CA: W. H. Freeman, 1976.
Edwards, A. L. *Multiple Regression and the Analysis of Variance and Covariance.* San Francisco, CA: W. H. Freeman, 1979.

Linear Space

see VECTOR SPACE

Linear Stability

Consider the general system of two first-order ORDINARY DIFFERENTIAL EQUATIONS

$$\dot{x}=f(x,y) \tag{1}$$
$$\dot{y}=g(x,y). \tag{2}$$

Let x_0 and y_0 denote FIXED POINTS with $\dot{x}=\dot{y}=0$, so

$$f(x_0,y_0)=0 \tag{3}$$
$$g(x_0,y_0)=0. \tag{4}$$

Then expand about (x_0, y_0) so

$$\delta\dot{x} = f_x(x_0,y_0)\delta x + f_y(x_0,y_0)\delta y + f_{xy}(x_0,y_0)\delta x\delta y + \ldots \tag{5}$$

$$\delta\dot{y} = g_x(x_0,y_0)\delta x + g_y(x_0,y_0)\delta y + g_{xy}(x_0,y_0)\delta x\delta y + \ldots. \tag{6}$$

To first-order, this gives

$$\frac{d}{dt}\begin{bmatrix}\delta x\\ \delta y\end{bmatrix} = \begin{bmatrix} f_x(x_0,y_0) & f_y(x_0,y_0)\\ g_x(x_0,y_0) & g_y(x_0,y_0)\end{bmatrix}\begin{bmatrix}\delta x\\ \delta y\end{bmatrix}, \tag{7}$$

where the 2×2 MATRIX is called the STABILITY MATRIX.

In general, given an n-D MAP $\mathbf{x}' = T(\mathbf{x})$, let $\mathbf{x}_0$ be a FIXED POINT, so that

$$T(\mathbf{x}_0) = \mathbf{x}_0. \tag{8}$$

Expand about the fixed point,

$$\begin{aligned} T(\mathbf{x}_0 + \delta\mathbf{x}) &= T(\mathbf{x}_0) + \frac{\partial T}{\partial \mathbf{x}}\delta\mathbf{x} + \mathcal{O}(\delta\mathbf{x})^2\\ &\equiv T(\mathbf{x}_0) + \delta T, \end{aligned} \tag{9}$$

so

$$\delta T = \frac{\partial T}{\partial \mathbf{x}}\delta\mathbf{x} \equiv \mathsf{A}\,\delta\mathbf{x}. \tag{10}$$

The map can be transformed into the principal axis frame by finding the EIGENVECTORS and EIGENVALUES of the MATRIX A

$$(\mathsf{A} - \lambda\mathsf{I})\,\delta\mathbf{x} = \mathbf{0}, \tag{11}$$

so the DETERMINANT

$$|\mathsf{A} - \lambda\mathsf{I}| = 0. \tag{12}$$

The mapping is

$$\delta\mathbf{x}_{\text{princ}}{}' = \begin{bmatrix}\lambda_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \lambda_n\end{bmatrix}. \tag{13}$$

When iterated a large number of times,

$$\delta T'_{\text{princ}} \to 0 \tag{14}$$

only if $|\Re(\lambda_i)| < 1$ for $i = 1, \ldots, n$ but $\to \infty$ if any $|\lambda_i| > 1$. Analysis of the EIGENVALUES (and EIGENVECTORS) of A therefore characterizes the type of FIXED POINT. The condition for stability is $|\Re(\lambda_i)| < 1$ for $i = 1, \ldots, n$.

see also FIXED POINT, STABILITY MATRIX

References

Tabor, M. "Linear Stability Analysis." §1.4 in *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, pp. 20–31, 1989.

Linear Transformation

An $n\times n$ MATRIX A is a linear transformation (linear MAP) IFF, for every pair of n-VECTORS $\mathbf{X}$ and $\mathbf{Y}$ and every SCALAR t,

$$\mathsf{A}(\mathbf{X} + \mathbf{Y}) = \mathsf{A}(\mathbf{X}) + \mathsf{A}(\mathbf{Y}) \tag{1}$$

and

$$\mathsf{A}(t\mathbf{X}) = t\mathsf{A}(\mathbf{X}). \tag{2}$$

Consider the 2-D transformation

$$\rho x_1' = a_{11}x_1 + a_{12}x_2 \tag{3}$$

$$\rho x_2' = a_{21}x_2 + a_{22}x_2. \tag{4}$$

Rescale by defining $\lambda \equiv x_1/x_2$ and $\lambda' \equiv x_1'/x_2'$, then the above equations become

$$\lambda' = \frac{\alpha\lambda + \beta}{\gamma\lambda + \delta}, \tag{5}$$

where $\alpha\delta - \beta\gamma \neq 0$ and α, β, γ and δ are defined in terms of the old constants. Solving for λ gives

$$\lambda = \frac{\delta\lambda' - \beta}{-\gamma\lambda' + \alpha}, \tag{6}$$

so the transformation is ONE-TO-ONE. To find the FIXED POINTS of the transformation, set $\lambda = \lambda'$ to obtain

$$\gamma\lambda^2 + (\delta - \alpha)\lambda - \beta = 0. \tag{7}$$

This gives two fixed points which may be distinct or coincident. The fixed points are classified as follows.

variables	type
$(\delta - \alpha)^2 + 4\beta\gamma > 0$	hyperbolic fixed point
$(\delta - \alpha)^2 + 4\beta\gamma < 0$	elliptic fixed point
$(\delta - \alpha)^2 + 4\beta\gamma = 0$	parabolic fixed point

see also ELLIPTIC FIXED POINT (MAP), HYPERBOLIC FIXED POINT (MAP), INVOLUNTARY, LINEAR OPERATOR, PARABOLIC FIXED POINT

References

Woods, F. S. *Higher Geometry: An Introduction to Advanced Methods in Analytic Geometry.* New York: Dover, pp. 13–15, 1961.

Linearly Dependent Curves

Two curves ϕ and ψ satisfying

$$\phi + \psi = 0$$

are said to be linearly dependent. Similarly, n curves ϕ_i, $i = 1, \ldots, n$ are said to be linearly dependent if

$$\sum_{i=1}^{n} \phi_i = 0.$$

see also BERTINI'S THEOREM, STUDY'S THEOREM

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, pp. 32–34, 1959.

Linearly Dependent Functions

The n functions $f_1(x)$, $f_2(x)$, ..., $f_n(x)$ are linearly dependent if, for some $c_1, c_2, \ldots, c_n \in \mathbb{R}$ not all zero,

$$c_i f_i(x) = 0 \qquad (1)$$

(where EINSTEIN SUMMATION is used) for all x in some interval I. If the functions are not linearly dependent, they are said to be linearly independent. Now, if the functions $\in \mathbb{R}^{n-1}$, we can differentiate (1) up to $n-1$ times. Therefore, linear dependence also requires

$$c_i f_i' = 0 \qquad (2)$$

$$c_i f_i'' = 0 \qquad (3)$$

$$c_i f_i^{(n-1)} = 0, \qquad (4)$$

where the sums are over $i = 1, \ldots, n$. These equations have a nontrivial solution IFF the DETERMINANT

$$\begin{vmatrix} f_1 & f_2 & \cdots & f_n \\ f_1' & f_2' & \cdots & f_n' \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)} & f_2^{(n-1)} & \cdots & f_n^{(n-1)} \end{vmatrix} = 0, \qquad (5)$$

where the DETERMINANT is conventionally called the WRONSKIAN and is denoted $W(f_1, f_2, \ldots, f_n)$. If the WRONSKIAN $\neq 0$ for any value c in the interval I, then the only solution possible for (2) is $c_i = 0$ ($i = 1, \ldots, n$), and the functions are linearly independent. If, on the other hand, $W = 0$ for a range, the functions are linearly dependent in the range. This is equivalent to stating that if the vectors $\mathbf{V}[f_1(c)], \ldots, \mathbf{V}[f_n(c)]$ defined by

$$\mathbf{V}[f_i(x)] = \begin{bmatrix} f_i(x) \\ f_i'(x) \\ f_i''(n) \\ \vdots \\ f_i^{(n-1)}(x) \end{bmatrix} \qquad (6)$$

are linearly independent for at least one $c \in I$, then the functions f_i are linearly independent in I.

References

Sansone, G. "Linearly Independent Functions." §1.2 in *Orthogonal Functions, rev. English ed.* New York: Dover, pp. 2–3, 1991.

Linearly Dependent Vectors

n VECTORS $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are linearly dependent IFF there exist SCALARS $c_1, c_2, \ldots, c_n$ such that

$$c_i \mathbf{X}_i = 0, \qquad (1)$$

where EINSTEIN SUMMATION is used and $i = 1, \ldots, n$. If no such SCALARS exist, then the vectors are said to be linearly independent. In order to satisfy the CRITERION for linear dependence,

$$c_1 \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix} + c_2 \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{bmatrix} + \cdots + c_n \begin{bmatrix} x_{1n} \\ x_{2n} \\ \vdots \\ x_{nn} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (2)$$

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \qquad (3)$$

In order for this MATRIX equation to have a nontrivial solution, the DETERMINANT must be 0, so the VECTORS are linearly dependent if

$$\begin{vmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{vmatrix} = 0, \qquad (4)$$

and linearly independent otherwise.

Let $\mathbf{p}$ and $\mathbf{q}$ be n-D VECTORS. Then the following three conditions are equivalent (Gray 1993).

1. $\mathbf{p}$ and $\mathbf{q}$ are linearly dependent.
2. $\begin{vmatrix} \mathbf{p}\cdot\mathbf{p} & \mathbf{p}\cdot\mathbf{q} \\ \mathbf{q}\cdot\mathbf{p} & \mathbf{q}\cdot\mathbf{q} \end{vmatrix} = 0.$
3. The $2 \times n$ MATRIX $\begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix}$ has rank less than two.

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 186–187, 1993.

Linearly Independent

Two or more functions, equations, or vectors which are not linearly dependent are said to be linearly independent.

see also LINEARLY DEPENDENT CURVES, LINEARLY DEPENDENT FUNCTIONS, LINEARLY DEPENDENT VECTORS, MAXIMALLY LINEAR INDEPENDENT

Link

Formally, a link is one or more disjointly embedded CIRCLES in 3-space. More informally, a link is an assembly of KNOTS with mutual entanglements. Kuperberg (1994) has shown that a nontrivial KNOT or link in $\mathbb{R}^3$ has four COLLINEAR points (Eppstein). Doll and Hoste (1991) list POLYNOMIALS for oriented links of nine or fewer crossings. A listing of the first few simple links follows, arranged by CROSSING NUMBER.

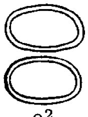

0_1^2

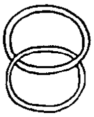
2^2_1

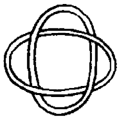
4^2_1

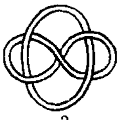
5^2_1

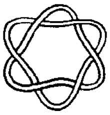
6^2_1

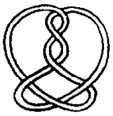
6^2_2

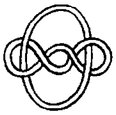
6^2_3

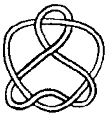
7^2_1

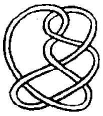
7^2_2

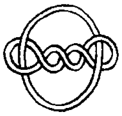
7^2_3

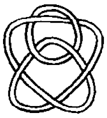
7^2_4

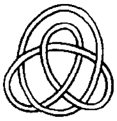
7^2_5

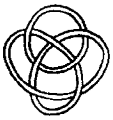
7^2_6

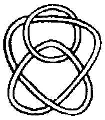
7^2_7

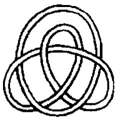
7^2_8

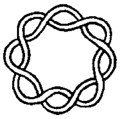
8^2_1

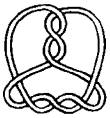
8^2_2

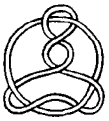
8^2_3

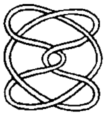
8^2_4

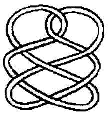
8^2_5

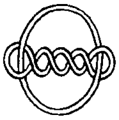
8^2_6

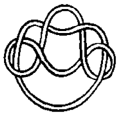
8^2_7

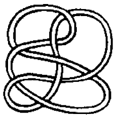
8^2_8

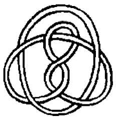
8^2_9

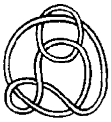
8^2_{10}

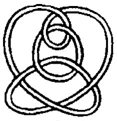
8^2_{11}

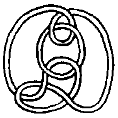
8^2_{12}

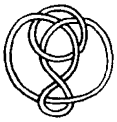
8^2_{13}

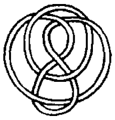
8^2_{14}

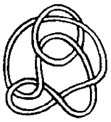
8^2_{15}

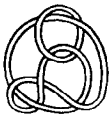
8^2_{16}

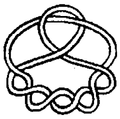
9^2_1

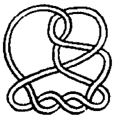
9^2_2

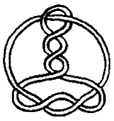
9^2_3

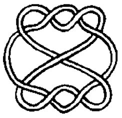
9^2_4

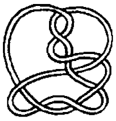
9^2_5

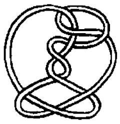
9^2_6

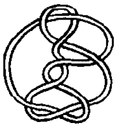
9^2_7

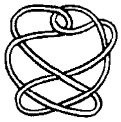
9^2_8

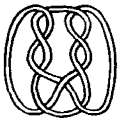
9^2_9

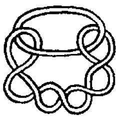
9^2_{10}

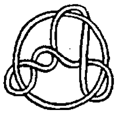
9^2_{11}

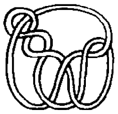
9^2_{12}

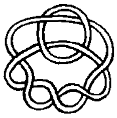
9^2_{13}

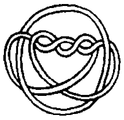
9^2_{14}

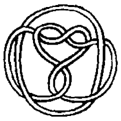
9^2_{15}

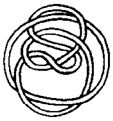
9^2_{16}

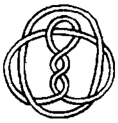
9^2_{17}

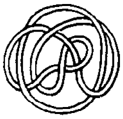
9^2_{18}

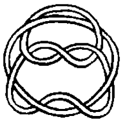
9^2_{19}

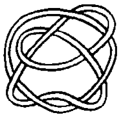
9^2_{20}

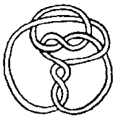
9^2_{21}

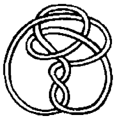
9^2_{22}

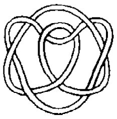
9^2_{23}

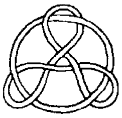
9^2_{24}

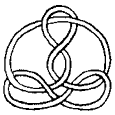
9^2_{25}

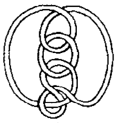
9^2_{26}

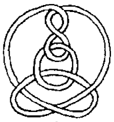
9^2_{27}

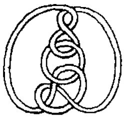
9^2_{28}

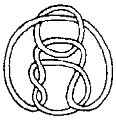
9^2_{29}

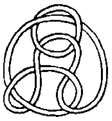
9^2_{30}

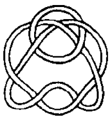
9^2_{31}

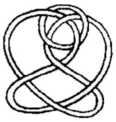
9^2_{32}

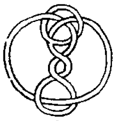
9^2_{33}

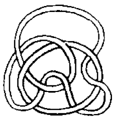
9^2_{34}

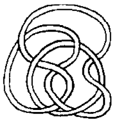
9^2_{35}

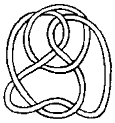
9^2_{36}

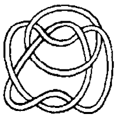
9^2_{37}

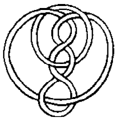
9^2_{38}

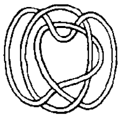
9^2_{39}

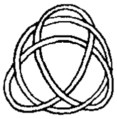
9^2_{40}

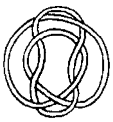
9^2_{41}

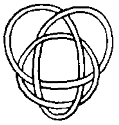
9^2_{42}

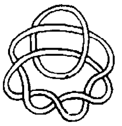
9^2_{43}

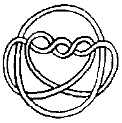
9^2_{44}

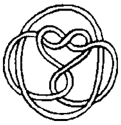
9^2_{45}

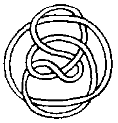
9^2_{46}

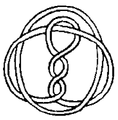
9^2_{47}

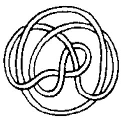
9^2_{48}

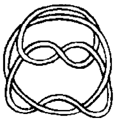
9^2_{49}

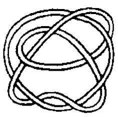
9^2_{50}

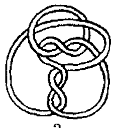
9^2_{51}

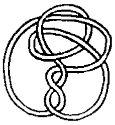
9^2_{52}

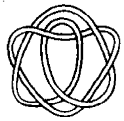
9^2_{53}

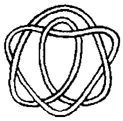
9^2_{54}

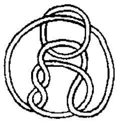
9^2_{55}

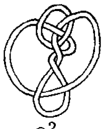
9^2_{56}

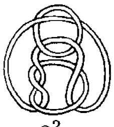
9^2_{57}

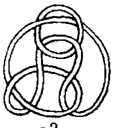
9^2_{58}

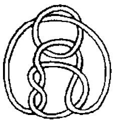
9^2_{59}

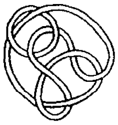
9^2_{60}

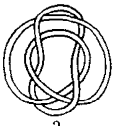
9^2_{61}

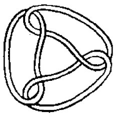
6^3_1

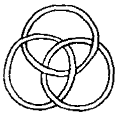
6^3_2

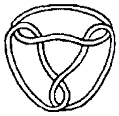
6^3_3

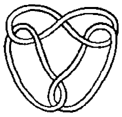
7^3_1

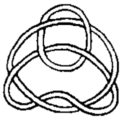
8^3_1

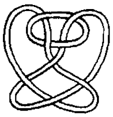
8^3_2

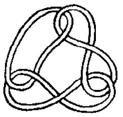
8^3_3

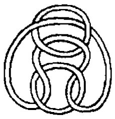
8^3_4

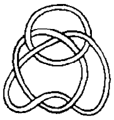
8^3_5

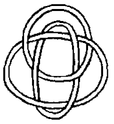
8^3_6

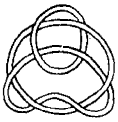
8^3_7

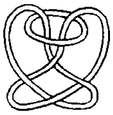
8^3_8

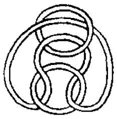
8^3_9

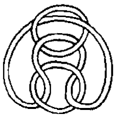
8^3_{10}

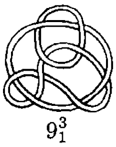
9^3_1

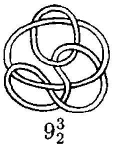
9^3_2

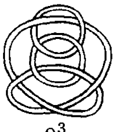
9^3_3

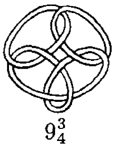
9^3_4

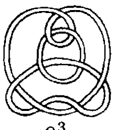
9^3_5

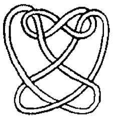
9^3_6

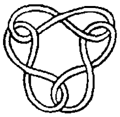
9^3_7

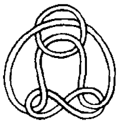
9^3_8

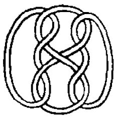
9^3_9

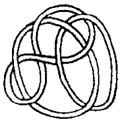
9^3_{10}

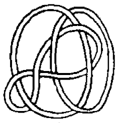
9^3_{11}

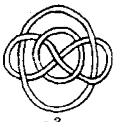
9^3_{12}

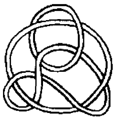
9^3_{13}

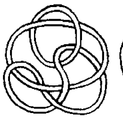
9^3_{14}

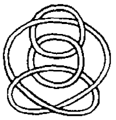
9^3_{15}

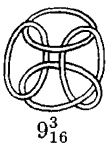
9^3_{16}

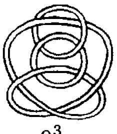
9^3_{17}

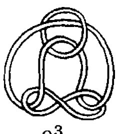
9^3_{18}

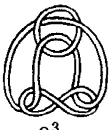
9^3_{19}

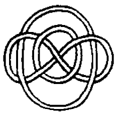
9^3_{20}

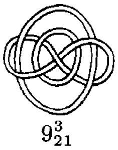
9^3_{21}

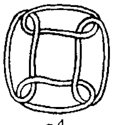
8^4_1

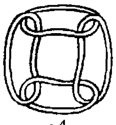
8^4_2

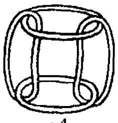
8^4_3

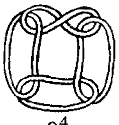
9^4_1

see also ANDREWS-CURTIS LINK, BORROMEAN RINGS, BRUNNIAN LINK, HOPF LINK, KNOT, WHITEHEAD LINK

References

Doll, H. and Hoste, J. "A Tabulation of Oriented Links." *Math. Comput.* **57**, 747–761, 1991.

Eppstein, D. "Colinear Points on Knots." `http://www.ics.uci.edu/~eppstein/junkyard/knot-colinear.html`.

Kuperberg, G. "Quadrisecants of Knots and Links." *J. Knot Theory Ramifications* **3**, 41–50, 1994.

Weisstein, E. W. "Knots." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Knots.m`.

Link Diagram

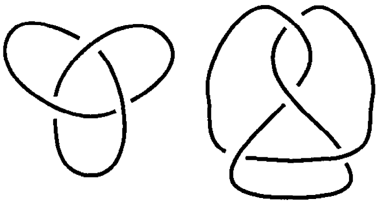

A planar diagram depicting a LINK (or KNOT) as a sequence of segments with gaps representing undercrossings and solid lines overcrossings. In such a diagram, only two segments should ever cross at a single point. Link diagrams for the TREFOIL KNOT and FIGURE-OF-EIGHT KNOT are illustrated above.

Linkage

Sylvester, Kempe and Cayley developed the geometry associated with the theory of linkages in the 1870s. Kempe proved that every finite segment of an algebraic curve can be generated by a linkage in the manner of WATT'S CURVE.

see also HART'S INVERSOR, KEMPE LINKAGE, PANTOGRAPH, PEAUCELLIER INVERSOR, SARRUS LINKAGE, WATT'S PARALLELOGRAM

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., 1989.

Linking Number

A LINK invariant. Given a two-component oriented LINK, take the sum of +1 crossings and −1 crossing over all crossings between the two links and divide by 2. For components α and β,

$$L(\alpha,\beta) \equiv \tfrac{1}{2} \sum_{p \in \alpha \sqcap \beta} \epsilon(p),$$

where $\alpha \sqcap \beta$ is the set of crossings of α with β and $\epsilon(p)$ is the sign of the crossing. The linking number of a splittable two-component link is always 0.

see also JONES POLYNOMIAL, LINK

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 132–133, 1976.

Links Curve

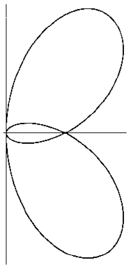

The curve given by the Cartesian equation

$$(x^2 + y^2 - 3x)^2 = 4x^2(2 - x).$$

The origin of the curve is a TACNODE.

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 72, 1989.

Linnik's Constant

The constant L in LINNIK'S THEOREM. Heath-Brown (1992) has shown that $L \leq 5.5$, and Schinzel, Sierpiński, and Kanold (Ribenboim 1989) have conjectured that $L = 2$.

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/linnik/linnik.html`.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 13, 1994.

Heath-Brown, D. R. "Zero-Free Regions for Dirichlet *L*-Functions and the Least Prime in an Arithmetic Progression." *Proc. London Math. Soc.* **64**, 265–338, 1992.

Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, 1989.

Linnik's Theorem

Let $p(d,a)$ be the smallest PRIME in the arithmetic progression $\{a + kd\}$ for k an INTEGER > 0. Let

$$p(d) \equiv \max p(d,a)$$

such that $1 \leq a < d$ and $(a,d) = 1$. Then there exists a $d_0 \geq 2$ and an $L > 1$ such that $p(d) < d^L$ for all $d > d_0$. L is known as LINNIK'S CONSTANT.

References

Linnik, U. V. "On the Least Prime in an Arithmetic Progression. I. The Basic Theorem." *Mat. Sbornik N. S.* **15 (57)**, 139–178, 1944.

Linnik, U. V. "On the Least Prime in an Arithmetic Progression. II. The Deuring-Heilbronn Phenomenon" *Mat. Sbornik N. S.* **15 (57)**, 347–368, 1944.

Liouville's Boundedness Theorem

A bounded ENTIRE FUNCTION in the COMPLEX PLANE $\mathbb{C}$ is constant. The FUNDAMENTAL THEOREM OF ALGEBRA follows as a simple corollary.

see also COMPLEX PLANE, ENTIRE FUNCTION, FUNDAMENTAL THEOREM OF ALGEBRA

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 381–382, 1953.

Liouville's Conformality Theorem

In SPACE, the only CONFORMAL TRANSFORMATIONS are inversions, SIMILARITY TRANSFORMATIONS, and CONGRUENCE TRANSFORMATIONS. Or, restated, every ANGLE-preserving transformation is a SPHERE-preserving transformation.

see also CONFORMAL MAP

Liouville's Conic Theorem

The lengths of the TANGENTS from a point P to a CONIC C are proportional to the CUBE ROOTS of the RADII OF CURVATURE of C at the corresponding points of contact.

see also CONIC SECTION

Liouville's Constant

$$L \equiv \sum_{n=1}^{\infty} 10^{-n!} = 0.110001000000000000000001\ldots$$

(Sloane's A012245). Liouville's constant is a decimal fraction with a 1 in each decimal place corresponding to a FACTORIAL $n!$, and ZEROS everywhere else. This number was among the first to be proven to be TRANSCENDENTAL. It nearly satisfies

$$10x^6 - 75x^3 - 190x + 21 = 0,$$

but with $x = L$, this equation gives $-0.0000000059\ldots$.

see also LIOUVILLE NUMBER

References
Conway, J. H. and Guy, R. K. "Liouville's Number." In *The Book of Numbers.* New York: Springer-Verlag, pp. 239–241, 1996.
Courant, R. and Robbins, H. "Liouville's Theorem and the Construction of Transcendental Numbers." §2.6.2 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 104–107, 1996.
Liouville, J. "Sur des classes très étendues de quantités dont la valeur n'est ni algébrique, ni même reductible à des irrationelles algébriques." *C. R. Acad. Sci. Paris* **18**, 883–885 and 993–995, 1844.
Liouville, J. "Sur des classes très-étendues de quantités dont la valeur n'est ni algébrique, ni même réductible à des irrationelles algébriques." *J. Math. pures appl.* **15**, 133–142, 1850.
Sloane, N. J. A. Sequence A012245 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Liouville's Elliptic Function Theorem

An ELLIPTIC FUNCTION with no POLES in a fundamental cell is a constant.

Liouville Function

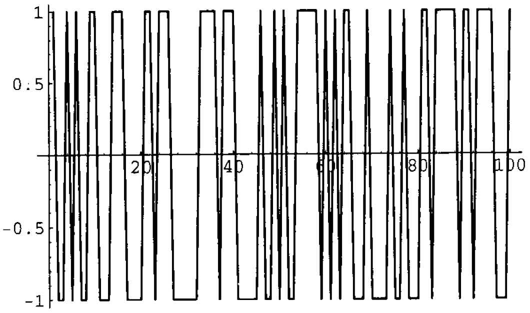

The function

$$\lambda(n) = (-1)^{r(n)}, \tag{1}$$

where $r(n)$ is the number of not necessarily distinct PRIME FACTORS of n, with $r(1) = 0$. The first few values of $\lambda(n)$ are 1, -1, -1, 1, -1, 1, -1, -1, 1, 1, -1, -1, ... The Liouville function is connected with the RIEMANN ZETA FUNCTION by the equation

$$\frac{\zeta(2s)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\lambda(n)}{n^s} \tag{2}$$

(Lehman 1960).

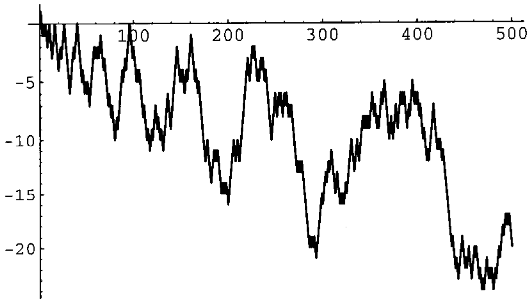

The CONJECTURE that the SUMMATORY FUNCTION

$$L(n) \equiv \sum_{k=1}^{n} \lambda(n) \tag{3}$$

satisfies $L(n) \leq 0$ for $n \geq 2$ is called the PÓLYA CONJECTURE and has been proved to be false. The first n for which $L(n) = 0$ are for n = 2, 4, 6, 10, 16, 26, 40, 96, 586, 906150256, ... (Sloane's A028488), and $n = 906150257$ is, in fact, the first counterexample to the PÓLYA CONJECTURE (Tanaka 1980). However, it is unknown if $L(x)$ changes sign infinitely often (Tanaka 1980). The first few values of $L(n)$ are 1, 0, -1, 0, -1, 0, -1, -2, -1, 0, -1, -2, -3, -2, -1, 0, -1, -2, -3, -4, ... (Sloane's A002819). $L(n)$ also satisfies

$$\sum_{n=1}^{x} L\left(\frac{x}{n}\right) = \left\lfloor \sqrt{x} \right\rfloor, \tag{4}$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION (Lehman 1960). Lehman (1960) also gives the formulas

$$\begin{aligned} L(x) = \sum_{m=1}^{x/w} \mu(m) \Bigg\{ \left\lfloor \sqrt{\frac{x}{m}} \right\rfloor & \\ - \sum_{k=1}^{v-1} \lambda(k) \left(\left\lfloor \frac{x}{km} \right\rfloor - \left\lfloor \frac{x}{mv} \right\rfloor \right) \Bigg\} & \\ - \sum_{l=x/w-1}^{x/v} L\left(\frac{x}{l}\right) \sum_{\substack{m\mid l \\ m=1}}^{x/w} \mu(m) & \end{aligned} \tag{5}$$

and

$$\begin{aligned} L(x) = \sum_{k=1}^{g} M\left(\frac{x}{k^2}\right) + \sum_{l=1}^{x/g^2} \mu(l) \left\lfloor \sqrt{\frac{x}{l}} \right\rfloor & \\ - M\left(\frac{x}{g^2}\right) \left\lfloor \sqrt{\frac{x}{g^2}} \right\rfloor, & \end{aligned} \tag{6}$$

where k, l, and m are variables ranging over the POSITIVE integers, $\mu(n)$ is the MÖBIUS FUNCTION, $M(x)$ is MERTENS FUNCTION, and v, w, and x are POSITIVE real numbers with $v < w < x$.

see also PÓLYA CONJECTURE, PRIME FACTORS, RIEMANN ZETA FUNCTION

References
Fawaz, A. Y. "The Explicit Formula for $L_0(x)$." *Proc. London Math. Soc.* **1**, 86–103, 1951.
Lehman, R. S. "On Liouville's Function." *Math. Comput.* **14**, 311–320, 1960.
Sloane, N. J. A. Sequences A028488 and A002819/M0042 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Tanaka, M. "A Numerical Investigation on Cumulative Sum of the Liouville Function." *Tokyo J. Math.* **3**, 187–189, 1980.

Liouville Measure

$$\prod_i dp_i\, dq_i,$$

where p_i and q_i are momenta and positions of particles.

see also LIOUVILLE'S PHASE SPACE THEOREM, PHASE SPACE

Liouville Number

A Liouville number is a TRANSCENDENTAL NUMBER which is very close to a RATIONAL NUMBER. An IRRATIONAL NUMBER β is a Liouville number if, for any n, there exist an infinite number of pairs of INTEGERS p and q such that

$$0 < \left|\beta - \frac{p}{q}\right| < \frac{1}{q^n}.$$

Mahler (1953) proved that π is not a Liouville number.

see also LIOUVILLE'S CONSTANT, LIOUVILLE'S RATIONAL APPROXIMATION THEOREM, ROTH'S THEOREM, TRANSCENDENTAL NUMBER

References

Mahler, K. "On the Approximation of π." *Nederl. Akad. Wetensch. Proc. Ser. A.* **56**/*Indagationes Math.* **15**, 30–42, 1953.

Liouville's Phase Space Theorem

States that for a nondissipative HAMILTONIAN SYSTEM, phase space density (the AREA between phase space contours) is constant. This requires that, given a small time increment dt,

$$q_1 \equiv q(t_0 + dt) = q_0 + \frac{\partial H(q_0, p_0, t)}{\partial p_0} dt + \mathcal{O}(dt^2) \qquad (1)$$

$$p_1 \equiv p(t_0 + dt) = p_0 - \frac{\partial H(q_0, p_0, t)}{\partial q_0} dt + \mathcal{O}(dt^2), \qquad (2)$$

the JACOBIAN be equal to one:

$$\begin{aligned} \frac{\partial(q_1, p_1)}{\partial(q_0, p_0)} &= \begin{vmatrix} \frac{\partial q_1}{\partial q_0} & \frac{\partial p_1}{\partial q_0} \\ \frac{\partial q_1}{\partial p_0} & \frac{\partial p_1}{\partial p_0} \end{vmatrix} \\ &= \begin{vmatrix} 1 + \frac{\partial^2 H}{\partial q_0 \partial p_0} dt & -\frac{\partial^2 H}{\partial {q_0}^2} dt \\ \frac{\partial^2 H}{\partial {p_0}^2} dt & 1 - \frac{\partial^2 H}{\partial q_0 \partial p_0} dt \end{vmatrix} + \mathcal{O}(dt^2) \\ &= 1 + \mathcal{O}(dt^2). \qquad (3) \end{aligned}$$

Expressed in another form, the integral of the LIOUVILLE MEASURE,

$$\prod_{i=1}^N \int dp_i\, dq_i, \qquad (4)$$

is a constant of motion. SYMPLECTIC MAPS of HAMILTONIAN SYSTEMS must therefore be AREA preserving (and have DETERMINANTS equal to 1).

see also LIOUVILLE MEASURE, PHASE SPACE

References

Chavel, I. *Riemannian Geometry: A Modern Introduction.* New York: Cambridge University Press, 1994.

Liouville Polynomial Identity

$$\begin{aligned} 6({x_1}^2 + {x_2}^2 + {x_3}^2 + {x_4}^2) &= (x_1 + x_2)^4 + (x_1 + x_3)^4 \\ &+ (x_2+x_3)^4 + (x_1+x_4)^4 + (x_2+x_4)^4 + (x_3+x_4)^4 + (x_1-x_2)^4 \\ &+ (x_1 - x_3)^4 + (x_2 - x_3)^4 + (x_1 - x_4)^4 + (x_2 - x_4)^4 \\ &+ (x_3 - x_4)^4. \end{aligned}$$

This is proven in Rademacher and Toeplitz (1957).

see also WARING'S PROBLEM

References

Rademacher, H. and Toeplitz, O. *The Enjoyment of Mathematics: Selections from Mathematics for the Amateur.* Princeton, NJ: Princeton University Press, pp. 55–56, 1957.

Liouville's Rational Approximation Theorem

For any ALGEBRAIC NUMBER x of degree $n > 1$, a RATIONAL approximation $x = p/q$ must satisfy

$$\left|x - \frac{p}{q}\right| > \frac{1}{q^{n+1}}$$

for sufficiently large q. Writing $r \equiv n + 1$ leads to the definition of the LIOUVILLE-ROTH CONSTANT of a given number.

see also LAGRANGE NUMBER (RATIONAL APPROXIMATION), LIOUVILLE'S CONSTANT, LIOUVILLE NUMBER, LIOUVILLE-ROTH CONSTANT, MARKOV NUMBER, ROTH'S THEOREM, THUE-SIEGEL-ROTH THEOREM

References

Courant, R. and Robbins, H. "Liouville's Theorem and the Construction of Transcendental Numbers." §2.6.2 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 104–107, 1996.

Liouville-Roth Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let x be a REAL NUMBER, and let R be the SET of POSITIVE REAL NUMBERS for which

$$\left|x - \frac{p}{q}\right| < \frac{1}{q^r} \qquad (1)$$

has (at most) finitely many solutions p/q for p and q INTEGERS. Then the Liouville-Roth constant (or IRRATIONALITY MEASURE) is defined as the threshold at which LIOUVILLE'S RATIONAL APPROXIMATION THEOREM kicks in and x is no longer approximable by RATIONAL NUMBERS,

$$r(x) \equiv \inf_{r \in R} r. \qquad (2)$$

There are three regimes:

$$\begin{cases} r(x) = 1 & x \text{ is rational} \\ r(x) = 2 & x \text{ is algebraic irrational} \\ r(x) \geq 2 & x \text{ is transcendental.} \end{cases} \tag{3}$$

The best known upper bounds for common constants are

$$r(L) = \infty \tag{4}$$

$$r(e) = 2 \tag{5}$$

$$r(\pi) < 8.0161 \tag{6}$$

$$r(\ln 2) < 4.13 \tag{7}$$

$$r(\pi^2) < 6.3489 \tag{8}$$

$$r(\zeta(3)) < 13.42, \tag{9}$$

where L is LIOUVILLE'S CONSTANT, $\zeta(3)$ is APÉRY'S CONSTANT, and the lower bounds are 2 for the inequalities.

see also LIOUVILLE'S RATIONAL APPROXIMATION THEOREM, ROTH'S THEOREM, THUE-SIEGEL-ROTH THEOREM

References

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, 1987.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/lvlrth/lvlrth.html`.

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford: Clarendon Press, 1979.

Hata, M. "Improvement in the Irrationality Measures of π and π^2." *Proc. Japan. Acad. Ser. A Math. Sci.* **68**, 283–286, 1992.

Hata, M. "Rational Approximations to π and Some Other Numbers." *Acta Arith.* **63** 335–349, 1993.

Hata, M. "A Note on Beuker's Integral." *J. Austral. Math. Soc.* **58**, 143–153, 1995.

Stark, H. M. *An Introduction to Number Theory.* Cambridge, MA: MIT Press, 1978.

Liouville Space

Also known as LINE SPACE or "extended" HILBERT SPACE, it is the DIRECT PRODUCT of two HILBERT SPACES.

see also DIRECT PRODUCT (SET), HILBERT SPACE

Liouville's Sphere-Preserving Theorem

see LIOUVILLE'S CONFORMALITY THEOREM

Lipschitz Condition

A function $f(x)$ satisfies the Lipschitz condition of order α at $x = 0$ if

$$|f(h) - f(0)| \leq B|h|^\beta$$

for all $|h| < \epsilon$, where B and β are independent of h, $\beta > 0$, and α is an UPPER BOUND for all β for which a finite B exists.

see also HILLAM'S THEOREM, LIPSCHITZ FUNCTION

Lipschitz Function

A function f such that

$$|f(x) - f(y)| \leq C|x - y|$$

is called a Lipschitz function.

see also LIPSCHITZ CONDITION

References

Morgan, F. "What Is a Surface?" *Amer. Math. Monthly* **103**, 369–376, 1996.

Lipschitz's Integral

$$\int_0^\infty e^{-ax} J_0(bx)\,dx = \frac{1}{\sqrt{a^2+b^2}},$$

where $J_0(z)$ is the zeroth order BESSEL FUNCTION OF THE FIRST KIND.

References

Bowman, F. *Introduction to Bessel Functions.* New York: Dover, p. 58, 1958.

Lissajous Curve

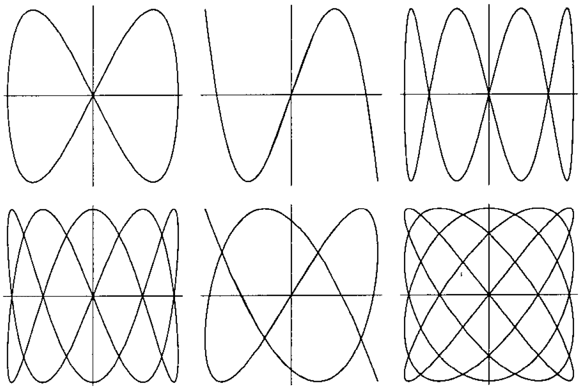

Lissajous curves are the family of curves described by the parametric equations

$$x(t) = A\cos(\omega_x t - \delta_x) \tag{1}$$

$$y(t) = B\cos(\omega_y t - \delta_y), \tag{2}$$

sometimes also written in the form

$$x(t) = a\sin(nt + c) \tag{3}$$

$$y(t) = b\sin t. \tag{4}$$

They are sometimes known as BOWDITCH CURVES after Nathaniel Bowditch, who studied them in 1815. They were studied in more detail (independently) by Jules-Antoine Lissajous in 1857 (MacTutor Archive). Lissajous curves have applications in physics, astronomy, and other sciences. The curves close IFF ω_x/ω_y is RATIONAL.

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 53–54, 1993.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 178–179 and 181–183, 1972.

MacTutor History of Mathematics Archive. "Lissajous Curves." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Lissajous.html`.

Lissajous Figure

see LISSAJOUS CURVE

List

A DATA STRUCTURE consisting of an order SET of elements, each of which may be a number, another list, etc. A list is usually denoted $(a_1, a_2, \ldots, a_n)$ or $\{a_1, a_2, \ldots, a_n\}$.

see also QUEUE, STACK

Lituus

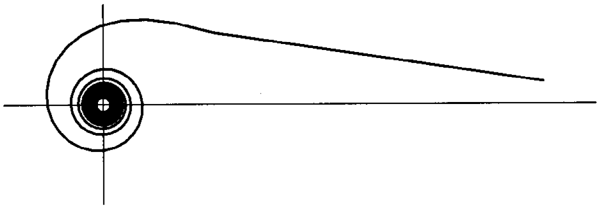

An ARCHIMEDEAN SPIRAL with $m = -2$, having polar equation

$$r^2\theta = a^2.$$

Lituus means a "crook," in the sense of a bishop's crosier. The lituus curve originated with Cotes in 1722. Maclaurin used the term lituus in his book *Harmonia Mensurarum* in 1722 (MacTutor Archive). The lituus is the locus of the point P moving such that the AREA of a circular SECTOR remains constant.

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 69–70, 1993.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 186 and 188, 1972.

Lockwood, E. H. *A Book of Curves.* Cambridge, England: Cambridge University Press, p. 175, 1967.

MacTutor History of Mathematics Archive. "Lituus." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Lituus.html`.

Lituus Inverse Curve

The INVERSE CURVE of the LITUUS is an ARCHIMEDEAN SPIRAL with $m = 2$, which is FERMAT'S SPIRAL.

see also ARCHIMEDEAN SPIRAL, FERMAT'S SPIRAL, LITUUS

LLL Algorithm

An INTEGER RELATION algorithm.

see also FERGUSON-FORCADE ALGORITHM, HJLS ALGORITHM, INTEGER RELATION, PSLQ ALGORITHM, PSOS ALGORITHM

References

Lenstra, A. K.; Lenstra, H. W.; and Lovasz, L. "Factoring Polynomials with Rational Coefficients." *Math. Ann.* **261**, 515–534, 1982.

Ln

The LOGARITHM to BASE e, also called the NATURAL LOGARITHM, is denoted ln, i.e.,

$$\ln x \equiv \log_e x.$$

see also BASE (LOGARITHM), E, LG, LOGARITHM, NAPIERIAN LOGARITHM, NATURAL LOGARITHM

Lo Shu

8	1	6
3	5	7
4	9	2

The unique MAGIC SQUARE of order three. The Lo Shu is an ASSOCIATIVE MAGIC SQUARE, but not a PANMAGIC SQUARE.

see also ASSOCIATIVE MAGIC SQUARE, MAGIC SQUARE, PANMAGIC SQUARE

References

Hunter, J. A. H. and Madachy, J. S. *Mathematical Diversions.* New York: Dover, pp. 23–24, 1975.

Lobachevsky-Bolyai-Gauss Geometry

see HYPERBOLIC GEOMETRY

Lobachevsky's Formula

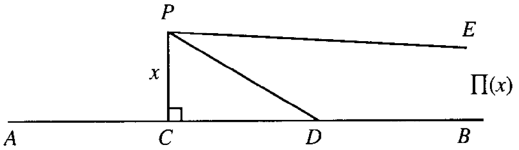

Given a point P and a LINE AB, draw the PERPENDICULAR through P and call it PC. Let PD be any other line from P which meets CB in D. In a HYPERBOLIC GEOMETRY, as D moves off to infinity along CB, then the line PD approaches the limiting line PE, which is said to be parallel to CB at P. The angle $\angle CPE$ which PE makes with PC is then called the ANGLE OF PARALLELISM for perpendicular distance x, and is given by

$$\Pi(x) = 2\tan^{-1}(e^{-x}),$$

which is called Lobachevsky's formula.

see also ANGLE OF PARALLELISM, HYPERBOLIC GEOMETRY

References

Manning, H. P. *Introductory Non-Euclidean Geometry.* New York: Dover, p. 58, 1963.

Lobatto Quadrature

Also called RADAU QUADRATURE (Chandrasekhar 1960). A GAUSSIAN QUADRATURE with WEIGHTING FUNCTION $W(x) = 1$ in which the endpoints of the interval $[-1, 1]$ are included in a total of n ABSCISSAS, giving $r = n - 2$ free abscissas. ABSCISSAS are symmetrical about the origin, and the general FORMULA is

$$\int_{-1}^{1} f(x)\,dx = w_1 f(-1) + w_n f(1) + \sum_{i=2}^{n-1} w_i f(x_i). \quad (1)$$

The free ABSCISSAS x_i for $i = 2, \ldots, n-1$ are the roots of the POLYNOMIAL $P'_{n-1}(x)$, where $P(x)$ is a LEGENDRE POLYNOMIAL. The weights of the free abscissas are

$$w_i = -\frac{2n}{(1-x_i{}^2)P''_{n-1}(x_i)P'_m(x_i)} \quad (2)$$

$$= \frac{2}{n(n-1)[P_{n-1}(x_i)]^2}, \quad (3)$$

and of the endpoints are

$$w_{1,n} = \frac{2}{n(n-1)}. \quad (4)$$

The error term is given by

$$E = -\frac{n(n-1)^3 2^{2n-1}[(n-2)!]^4}{(2n-1)[(2n-2)!]^3} f^{(2n-2)}(\xi), \quad (5)$$

for $\xi \in (-1, 1)$. Beyer (1987) gives a table of parameters up to n=11 and Chandrasekhar (1960) up to n=9 (although Chandrasekhar's $\mu_{3,4}$ for $m = 5$ is incorrect).

n	x_i	w_i
3	0	1.33333
	±1	0.333333
4	±0.447214	0.833333
	±1	0.166667
5	0	0.711111
	±0.654654	0.544444
	±1	0.100000
6	±0.285232	0.554858
	±0.765055	0.378475
	±1	0.0666667

The ABSCISSAS and weights can be computed analytically for small n.

n	x_i	w_i
3	0	$\frac{4}{3}$
	± 1	$\frac{1}{3}$
4	$\pm\frac{1}{5}\sqrt{5}$	$\frac{1}{6}$
	± 1	$\frac{5}{6}$
5	0	$\frac{32}{45}$
	$\pm\frac{1}{7}\sqrt{21}$	$\frac{49}{90}$
	± 1	$\frac{1}{10}$

see also CHEBYSHEV QUADRATURE, RADAU QUADRATURE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 888–890, 1972.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 465, 1987.

Chandrasekhar, S. *Radiative Transfer.* New York: Dover, pp. 63–64, 1960.

Hildebrand, F. B. *Introduction to Numerical Analysis.* New York: McGraw-Hill, pp. 343–345, 1956.

Lobster

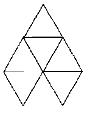

A 6-POLYIAMOND.

References

Golomb, S. W. *Polyominoes: Puzzles, Patterns, Problems, and Packings, 2nd ed.* Princeton, NJ: Princeton University Press, p. 92, 1994.

Local Cell

The POLYHEDRON resulting from letting each SPHERE in a SPHERE PACKING expand uniformly until it touches its neighbors on flat faces.

see also LOCAL DENSITY

Local Degree

The degree of a VERTEX of a GRAPH is the number of EDGES which touch the VERTEX, also called the LOCAL DEGREE. The VERTEX degree of a point A in a GRAPH, denoted $\rho(A)$, satisfies

$$\sum_{i=1}^{n} \rho(A_i) = 2E,$$

where E is the total number of EDGES. Directed graphs have two types of degrees, known as the INDEGREE and OUTDEGREE.

see also INDEGREE, OUTDEGREE

Local Density

Let each SPHERE in a SPHERE PACKING expand uniformly until it touches its neighbors on flat faces. Call the resulting POLYHEDRON the LOCAL CELL. Then the local density is given by

$$\rho \equiv \frac{V_{\text{sphere}}}{V_{\text{local cell}}}.$$

When the LOCAL CELL is a regular DODECAHEDRON, then

$$\rho_{\text{dodecahedron}} = \frac{\pi\sqrt{5+\sqrt{5}}}{5\sqrt{10}\,(\sqrt{5}-2)} = 0.7547\ldots.$$

see also LOCAL DENSITY CONJECTURE

Local Density Conjecture

The CONJECTURE that the maximum LOCAL DENSITY is given by $\rho_{\text{dodecahedron}}$.

see also LOCAL DENSITY

Local Extremum

A LOCAL MINIMUM or LOCAL MAXIMUM.

see also EXTREMUM, GLOBAL EXTREMUM

Local Field

A FIELD which is complete with respect to a discrete VALUATION is called a local field if its FIELD of RESIDUE CLASSES is FINITE. The HASSE PRINCIPLE is one of the chief applications of local field theory.

see also HASSE PRINCIPLE, VALUATION

References

Iyanaga, S. and Kawada, Y. (Eds.). "Local Fields." §257 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 811–815, 1980.

Local-Global Principle

see HASSE PRINCIPLE

Local Group Theory

The study of a FINITE GROUP G using the LOCAL SUBGROUPS of G. Local group theory plays a critical role in the CLASSIFICATION THEOREM.

see also SYLOW THEOREMS

Local Maximum

The largest value of a set, function, etc., within some local neighborhood.

see also GLOBAL MAXIMUM, LOCAL MINIMUM, MAXIMUM, PEANO SURFACE

Local Minimum

The smallest value of a set, function, etc., within some local neighborhood.

see also GLOBAL MINIMUM, LOCAL MAXIMUM, MINIMUM

Local Ring

A NOETHERIAN RING R with a Jacobson radical which has only a single maximal ideal.

References

Iyanaga, S. and Kawada, Y. (Eds.). "Local Rings." §281D in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 890–891, 1980.

Local Subgroup

A normalizer of a nontrivial SYLOW p-SUBGROUP of a GROUP G.

see also LOCAL GROUP THEORY

Local Surface

see PATCH

Locally Convex Space

see LOCALLY PATHWISE-CONNECTED SPACE

Locally Finite Space

A locally finite SPACE is one for which every point of a given space has a NEIGHBORHOOD that meets only finitely many elements of the COVER.

Locally Pathwise-Connected Space

A SPACE X is locally pathwise-connected if for every NEIGHBORHOOD around every point in X, there is a smaller, PATHWISE-CONNECTED NEIGHBORHOOD.

Loculus of Archimedes

see STOMACHION

Locus

The set of all points (usually forming a curve or surface) satisfying some condition. For example, the locus of points in the plane equidistant from a given point is a CIRCLE, and the set of points in 3-space equidistant from a given point is a SPHERE.

Log

The symbol $\log x$ is used by physicists, engineers, and calculator keypads to denote the BASE 10 LOGARITHM. However, mathematicians generally use the same symbol to mean the NATURAL LOGARITHM LN, $\ln x$. In this work, $\log x = \log_{10} x$, and $\ln x = \log_e x$ is used for the NATURAL LOGARITHM.

see also LG, LN, LOGARITHM, NATURAL LOGARITHM

Log Likelihood Procedure

A method for testing NESTED HYPOTHESES. To apply the procedure, given a specific model, calculate the LIKELIHOOD of observing the actual data. Then compare this likelihood to a nested model (i.e., one in which fewer parameters are allowed to vary independently).

Log Normal Distribution

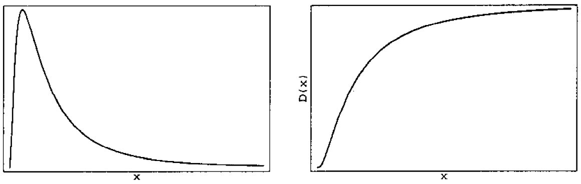

A CONTINUOUS DISTRIBUTION in which the LOGARITHM of a variable has a NORMAL DISTRIBUTION. It is a general case of GILBRAT'S DISTRIBUTION, to which the log normal distribution reduces with $S = 1$ and

$M = 0$. The probability density and cumulative distribution functions are log normal distribution

$$P(x) = \frac{1}{Sx\sqrt{2\pi}} e^{-(\ln x - M)^2/(2S^2)} \tag{1}$$

$$D(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\ln x - M}{S\sqrt{2}}\right)\right], \tag{2}$$

where $\operatorname{erf}(x)$ is the ERF function. This distribution is normalized, since letting $y \equiv \ln x$ gives $dy = dx/x$ and $x = e^y$, so

$$\int_0^\infty P(x)\,dx = \frac{1}{S\sqrt{2\pi}} \int_{-\infty}^\infty e^{-(y-M)^2/2S^2}\,dy = 1. \tag{3}$$

The MEAN, VARIANCE, SKEWNESS, and KURTOSIS are given by

$$\mu = e^{M+S^2/2} \tag{4}$$

$$\sigma^2 = e^{S^2+2M}(e^{S^2} - 1) \tag{5}$$

$$\gamma_1 = \sqrt{e^S - 1}\,(2 + e^{S^2}) \tag{6}$$

$$\gamma_2 = e^{2S^2}(3 + 2e^{S^2} + e^{2s^2}) - 3. \tag{7}$$

These can be found by direct integration

$$\begin{aligned}
\mu &= \frac{1}{S\sqrt{2\pi}} \int_0^\infty e^{-(\ln x - M)^2/2S^2}\,dx \\
&= \frac{1}{S\sqrt{2\pi}} \int_{-\infty}^\infty e^{(y-M)^2/2S^2} e^y\,dy \\
&= \frac{1}{S\sqrt{2\pi}} \int_{-\infty}^\infty e^{-[-y+(y-M)^2/2S^2]}\,dy \\
&= \frac{1}{S\sqrt{2\pi}} \int_{-\infty}^\infty e^{-(-2S^2y+y^2-2yM+M^2)/2S^2}\,dy \\
&= \frac{1}{S\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\{[y-(S^2+M)]^2+S^2(S^2+2M)\}/2S^2}\,dy \\
&= \frac{1}{S\sqrt{2\pi}} e^{M+S^2/2} \int_{-\infty}^\infty e^{-[y-(S^2+M)^2]/2S^2}\,dy \\
&= e^{M+S^2/2},
\end{aligned} \tag{8}$$

and similarly for σ^2. Examples of variates which have approximately log normal distributions include the size of silver particles in a photographic emulsion, the survival time of bacteria in disinfectants, the weight and blood pressure of humans, and the number of words written in sentences by George Bernard Shaw.

see also GILBRAT'S DISTRIBUTION, WEIBULL DISTRIBUTION

References

Aitchison, J. and Brown, J. A. C. *The Lognormal Distribution, with Special Reference to Its Use in Economics.* New York: Cambridge University Press, 1957.

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, p. 123, 1951.

Log-Series Distribution

The terms in the series expansion of $\ln(1-\theta)$ about $\theta = 0$ are proportional to this distribution.

$$P(n) = -\frac{\theta^n}{n\ln(1-\theta)} \tag{1}$$

$$D(n) \equiv \sum_{i=1}^n P(i) = \frac{\theta^{1+n}\Phi(\theta, 1, 1+n) + \ln(1-\theta)}{\ln(1-\theta)}, \tag{2}$$

where Φ is the LERCH TRANSCENDENT. The MEAN, VARIANCE, SKEWNESS, and KURTOSIS

$$\mu = \frac{\theta}{(\theta-1)\ln(1-\theta)} \tag{3}$$

$$\sigma^2 = -\frac{\theta[\theta + \ln(1-\theta)]}{(\theta-1)^2[\ln(1-\theta)]^2} \tag{4}$$

$$\gamma_1 = \frac{2\theta^2 + 3\theta\ln(1-\theta) + (1+\theta)\ln^2(1-\theta)}{\ln(1-\theta)[\theta + \ln(1-\theta)]\sqrt{-\theta[\theta + \ln(1-\theta)]}}\ln(1-\theta) \tag{5}$$

$$\gamma_2 = \frac{6\theta^3 + 12\theta^2\ln(1-\theta) + \theta(7+4\theta)\ln^2(1-\theta)}{\theta[\theta + \ln(1-\theta)]^2} + \frac{(1+4\theta+\theta^2)\ln^3(1-\theta)}{\theta[\theta + \ln(1-\theta)]^2}. \tag{6}$$

Log-Weibull Distribution

see FISHER-TIPPETT DISTRIBUTION

Logarithm

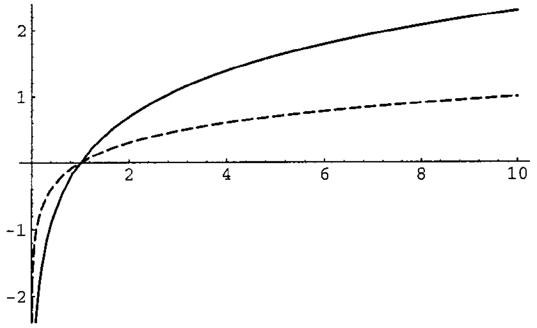

The logarithm is defined to be the INVERSE FUNCTION of taking a number to a given POWER. Therefore, for any x and b,

$$x = b^{\log_b x}, \tag{1}$$

or equivalently,

$$x = \log_b(b^x). \tag{2}$$

Here, the POWER b is known as the BASE of the logarithm. For any BASE, the logarithm function has a SINGULARITY at $x = 0$. In the above plot, the solid curve is the logarithm to BASE e (the NATURAL LOGARITHM), and the dotted curve is the logarithm to BASE 10 (LOG).

Logarithms are used in many areas of science and engineering in which quantities vary over a large range. For example, the decibel scale for the loudness of sound, the Richter scale of earthquake magnitudes, and the astronomical scale of stellar brightnesses are all logarithmic scales.

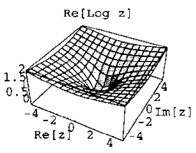

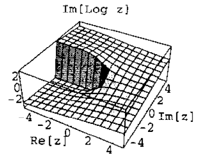

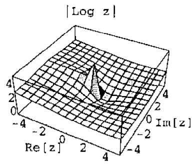

The logarithm can also be defined for COMPLEX arguments, as shown above. If the logarithm is taken as the forward function, the function taking the BASE to a given POWER is then called the ANTILOGARITHM.

For $x = \log N$, $\lfloor x \rfloor$ is called the CHARACTERISTIC and $x - \lfloor x \rfloor$ is called the MANTISSA. Division and multiplication identities follow from these

$$xy = b^{\log_b x} b^{\log_b y} = b^{\log_b x + \log_b y}, \tag{3}$$

from which it follows that

$$\log_b(xy) = \log_b x + \log_b y \tag{4}$$

$$\log_b \left(\frac{x}{y}\right) = \log_b x - \log_b y \tag{5}$$

$$\log_b x^n = n \log_b x. \tag{6}$$

There are a number of properties which can be used to change from one BASE to another

$$a = a^{\log_a b/\log_a b} = (a^{\log_a b})^{1/\log_a b} = b^{1/\log_a b} \tag{7}$$

$$\log_b a = \frac{1}{\log_a b} \tag{8}$$

$$\log_x z = \log_x(y^{\log_y z}) = \log_y z \log_x y \tag{9}$$

$$\log_y z = \frac{\log_x z}{\log_x y} \tag{10}$$

$$a^x = b^{x/\log_a b} = b^{x \log_b a}. \tag{11}$$

The logarithm BASE e is called the NATURAL LOGARITHM and is denoted $\ln x$ (LN). The logarithm BASE 10 is denoted $\log x$ (LOG), (although mathematics texts often use $\log x$ to mean $\ln x$). The logarithm BASE 2 is denoted $\lg x$ (LG).

An interesting property of logarithms follows from looking for a number y such that

$$\log_b(x + y) = -\log_b(x - y) \tag{12}$$

$$x + y = \frac{1}{x - y} \tag{13}$$

$$x^2 - y^2 = 1 \tag{14}$$

$$y = \sqrt{x^2 - 1}, \tag{15}$$

so

$$\log_b(x + \sqrt{x^2 - 1}) = -\log_b(x - \sqrt{x^2 - 1}). \tag{16}$$

Numbers of the form $\log_a b$ are IRRATIONAL if a and b are INTEGERS, one of which has a PRIME factor which the other lacks. A. Baker made a major step forward in TRANSCENDENTAL NUMBER theory by proving the transcendence of sums of numbers of the form $\alpha \ln \beta$ for α and β ALGEBRAIC NUMBERS.

see also ANTILOGARITHM, COLOGARITHM, e, EXPONENTIAL FUNCTION, HARMONIC LOGARITHM, LG, LN, LOG, LOGARITHMIC NUMBER, NAPIERIAN LOGARITHM, NATURAL LOGARITHM, POWER

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Logarithmic Function." §4.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 67–69, 1972.

Conway, J. H. and Guy, R. K. "Logarithms." *The Book of Numbers.* New York: Springer-Verlag, pp. 248–252, 1996.

Beyer, W. H. "Logarithms." *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 159–160, 1987.

Pappas, T. "Earthquakes and Logarithms." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 20–21, 1989.

Spanier, J. and Oldham, K. B. "The Logarithmic Function ln(x)." Ch. 25 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 225–232, 1987.

Logarithmic Binomial Formula

see LOGARITHMIC BINOMIAL THEOREM

Logarithmic Binomial Theorem

For all integers n and $|x| < a$,

$$\lambda_n^{(t)}(x + a) = \sum_{k=0}^{\infty} \left\lfloor {n \atop k} \right\rceil \lambda_{n-k}^{(t)}(a) x^k,$$

where $\lambda_n^{(t)}$ is the HARMONIC LOGARITHM and $\left\lfloor {n \atop k} \right\rceil$ is a ROMAN COEFFICIENT. For $t = 0$, the logarithmic binomial theorem reduces to the classical BINOMIAL THEOREM for POSITIVE n, since $\lambda_{n-k}^{(0)}(a) = a^{n-k}$ for $n \geq k$, $\lambda_{n-k}^{(0)}(a) = 0$ for $n < k$, and $\left\lfloor {n \atop k} \right\rceil = \binom{n}{k}$ when $n \geq k \geq 0$.

Similarly, taking $t = 1$ and $n < 0$ gives the NEGATIVE BINOMIAL SERIES. Roman (1992) gives expressions obtained for the case $t = 1$ and $n \geq 0$ which are not obtainable from the BINOMIAL THEOREM.

see also HARMONIC LOGARITHM, ROMAN COEFFICIENT

References

Roman, S. "The Logarithmic Binomial Formula." *Amer. Math. Monthly* **99**, 641–648, 1992.

Logarithmic Distribution

A CONTINUOUS DISTRIBUTION for a variate with probability function

$$P(x) = \frac{\log x}{b(\log b - 1) - a(\log a - 1)}$$

and distribution function

$$D(x) = \frac{a(1 - \log a) - x(1 - \log x)}{a(1 - \log a) - b(1 - \log b)}.$$

The MEAN is

$$\mu = \frac{a^2(1 - 2\log a) - b^2(1 - 2\log b)}{4[a(1 - \log a) - b(1 - \log b)]},$$

but higher order moments are rather messy.

Logarithmic Integral

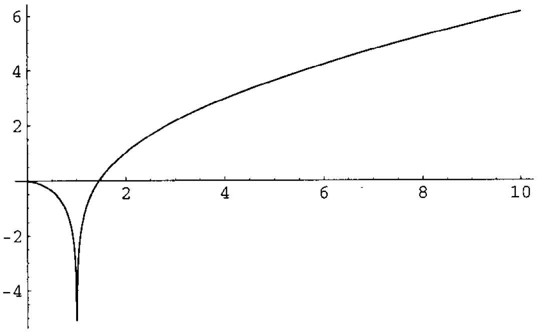

The logarithmic integral is defined by

$$\operatorname{li}(x) \equiv \int_0^x \frac{du}{\ln u}. \tag{1}$$

The offset form appearing in the PRIME NUMBER THEOREM is defined so that $\operatorname{Li}(2) = 0$:

$$\begin{aligned} \operatorname{Li}(x) &\equiv \int_2^x \frac{du}{\ln u} & (2)\\ &= \operatorname{li}(x) - \operatorname{li}(2) \approx \operatorname{li}(x) - 1.04516 & (3)\\ &= \operatorname{ei}(\ln x), & (4) \end{aligned}$$

where $\operatorname{ei}(x)$ is the EXPONENTIAL INTEGRAL. (Note that the NOTATION $\operatorname{Li}_n(z)$ is also used for the POLYLOGARITHM.) Nielsen (1965, pp. 3 and 11) showed and Ramanujan independently discovered (Berndt 1994) that

$$\int_\mu^x \frac{dt}{\ln t} = \gamma + \ln\ln x + \sum_{k=1}^\infty \frac{(\ln x)^k}{k!k}, \tag{5}$$

where γ is the EULER-MASCHERONI CONSTANT and μ is SOLDNER'S CONSTANT. Another FORMULA due to Ramanujan which converges more rapidly is

$$\begin{aligned} \int_\mu^x \frac{dt}{\ln t} &= \gamma + \ln\ln x\\ &+ \sqrt{x} \sum_{n=0}^\infty \frac{(-1)^{n-1}(\ln x)^n}{n!2^{n-1}} \sum_{k=0}^{\lfloor (n-1)/2 \rfloor} \frac{1}{2k+1} \end{aligned} \tag{6}$$

(Berndt 1994).

see also POLYLOGARITHM, PRIME CONSTELLATION, PRIME NUMBER THEOREM, SKEWES NUMBER

References

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 126–131, 1994.

Nielsen, N. *Theorie des Integrallogarithms.* New York: Chelsea, 1965.

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 151, 1991.

Logarithmic Number

A COEFFICIENT of the MACLAURIN SERIES of

$$\frac{1}{\ln(1+x)} = \frac{1}{x} + \tfrac{1}{2} + \tfrac{1}{12}x^2 - \tfrac{19}{720}x^3 + \tfrac{3}{160}x^4 + \ldots$$

(Sloane's A002206 and A002207), the multiplicative inverse of the MERCATOR SERIES function $\ln(1+x)$.

see also MERCATOR SERIES

References

Sloane, N. J. A. Sequences A002206/M5066 and A002207/M2017 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Logarithmic Spiral

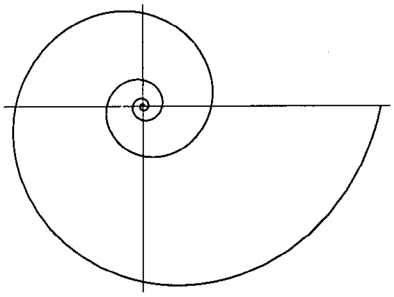

A curve whose equation in POLAR COORDINATES is given by

$$r = ae^{b\theta}, \tag{1}$$

where r is the distance from the ORIGIN, θ is the angle from the x-axis, and a and b are arbitrary constants. The logarithmic spiral is also known as the GROWTH SPIRAL, EQUIANGULAR SPIRAL, and SPIRA MIRABILIS. It can be expressed parametrically using

$$\cos\theta = \frac{1}{\sqrt{1 - \tan^2\theta}} = \frac{1}{\sqrt{1 + \frac{y^2}{x^2}}} = \frac{x}{\sqrt{x^2 + y^2}} = \frac{x}{r}, \tag{2}$$

which gives

$$\begin{aligned} x &= r\cos\theta = a\cos\theta e^{b\theta} & (3)\\ y &= x\tan\theta = r\sin\theta = a\sin\theta e^{b\theta}. & (4) \end{aligned}$$

The logarithmic spiral was first studied by Descartes in 1638 and Jakob Bernoulli. Bernoulli was so fascinated by the spiral that he had one engraved on his tombstone (although the engraver did not draw it true to form). Torricelli worked on it independently and found the length of the curve (MacTutor Archive).

The rate of change of RADIUS is

$$\frac{dr}{d\theta} = abe^{b\theta} = br, \tag{5}$$

and the ANGLE between the tangent and radial line at the point (r, θ) is

$$\psi = \tan^{-1}\left(\frac{r}{\frac{dr}{d\theta}}\right) = \tan^{-1}\left(\frac{1}{b}\right) = \cot^{-1} b. \tag{6}$$

So, as $b \to 0$, $\psi \to \pi/2$ and the spiral approaches a CIRCLE.

If P is any point on the spiral, then the length of the spiral from P to the origin is finite. In fact, from the point P which is at distance r from the origin measured along a RADIUS vector, the distance from P to the POLE along the spiral is just the ARC LENGTH. In addition, any RADIUS from the origin meets the spiral at distances which are in GEOMETRIC PROGRESSION (MacTutor Archive).

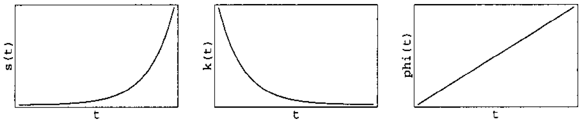

The ARC LENGTH, CURVATURE, and TANGENTIAL ANGLE of the logarithmic spiral are

$$\begin{aligned} s &= \int ds = \int \sqrt{x'^2 + y'^2}\, dt = \frac{a\sqrt{1+b^2}}{b} e^{b\theta} \\ &= \frac{r\sqrt{1+b^2}}{b} \end{aligned} \tag{7}$$

$$\kappa = \frac{x'y'' - y'x''}{(x'^2 + y'^2)^{3/2}} = (a\sqrt{1+b^2}e^{b\theta})^{-1} \tag{8}$$

$$\phi = \int \kappa(s)\, ds = \theta. \tag{9}$$

The CESÀRO EQUATION is

$$\kappa = \frac{1}{bs}. \tag{10}$$

On the surface of a SPHERE, the analog is a LOXODROME. This SPIRAL is related to FIBONACCI NUMBERS and the GOLDEN MEAN.

References
Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 184–186, 1972.
Lee, X. "EquiangularSpiral." `http://www.best.com/~xah/Special Plane Curves _ dir / Equiangular Spiral _ dir / equiangularSpiral.html`.
Lockwood, E. H. "The Equiangular Spiral." Ch. 11 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 98–109, 1967.
MacTutor History of Mathematics Archive. "Equiangular Spiral." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Equiangular.html`.

Logarithmic Spiral Caustic Curve

The CAUSTIC of a LOGARITHMIC SPIRAL, where the pole is taken as the RADIANT POINT, is an equal LOGARITHMIC SPIRAL.

Logarithmic Spiral Evolute

$$R = \frac{(r^2 + r_\theta{}^2)^{3/2}}{r^2 + 2r^2 r_\theta{}^2 - r r_{\theta\theta}}. \tag{1}$$

Using

$$r = ae^{b\theta} \quad r_\theta = abe^{b\theta} \quad r_{\theta\theta} = ab^2 e^{b\theta} \tag{2}$$

gives

$$\begin{aligned} R &= \frac{(a^2e^{2b\theta} + a^2b^2e^{2b\theta})^{3/2}}{(ae^{b\theta})^2 + 2(abe^{b\theta})^2 - (ab^{b\theta})(ab^2e^{b\theta})} \\ &= \frac{(1+b^2)^{3/2}a^3e^{3b\theta}}{2a^2b^2e^{2b\theta} + a^2e^{2b\theta} - a^2b^2e^{2b\theta}} \\ &= \frac{(1+b^2)^{3/2}a^3e^{3b\theta}}{a^2b^2e^{2b\theta} + a^2e^{2b\theta}} = \frac{(1+b^2)^{3/2}a^3e^{3b\theta}}{a^2(1+b^2)e^{2b\theta}} \\ &= a\sqrt{1+b^2}\, e^{b\theta} \end{aligned} \tag{3}$$

and

$$\begin{aligned} \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} ae^{b\theta}\cos\theta \\ ae^{b\theta}\sin\theta \end{bmatrix} \\ \begin{bmatrix} x' \\ y' \end{bmatrix} &= \begin{bmatrix} abe^{b\theta}\cos\theta - ae^{b\theta}\sin\theta \\ abe^{b\theta}\sin\theta + ae^{b\theta}\cos\theta \end{bmatrix} \\ &= ae^{b\theta}\begin{bmatrix} b\cos\theta - \sin\theta \\ b\sin\theta + \cos\theta \end{bmatrix}, \end{aligned} \tag{4}$$

so

$$\begin{aligned} |\mathbf{r}'| &= ae^{b\theta}\sqrt{(b\cos\theta - \sin\theta)^2 + (b\sin\theta + \cos\theta)^2} \\ &= ae^{b\theta}\sqrt{1+b^2}, \end{aligned} \tag{5}$$

and the TANGENT VECTOR is given by

$$\begin{aligned} \hat{\mathbf{T}} &= \frac{\mathbf{r}'}{|\mathbf{r}'|} = \frac{1}{ae^{b\theta}\sqrt{1+b^2}}\begin{bmatrix} ae^{b\theta}\cos\theta \\ ae^{b\theta}\sin\theta \end{bmatrix} \\ &= \frac{1}{\sqrt{1+b^2}}\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}. \end{aligned} \tag{6}$$

The coordinates of the EVOLUTE are therefore

$$\xi = -abe^{b\theta}\sin\theta \tag{7}$$

$$\eta = abe^{b\theta}\cos\theta. \tag{8}$$

So the EVOLUTE is another logarithmic spiral with $a' \equiv ab$, as first shown by Johann Bernoulli. However, in some cases, the EVOLUTE is identical to the original, as can be demonstrated by making the substitution to the new variable

$$\theta \equiv \phi - \tfrac{1}{2}\pi \pm 2n\pi. \tag{9}$$

Then the above equations become

$$\begin{aligned}\xi &= -abe^{b(\phi-\pi/2\pm 2n\pi)}\sin(\phi-\pi/2\pm 2n\pi)\\ &= abe^{b\phi}e^{b(-\pi/2\pm 2n\pi)}\cos\phi \qquad (10)\\ \eta &= abe^{b(\phi-\pi/2\pm 2n\pi)}\cos(\phi-\pi/2\pm 2n\pi)\\ &= abe^{b\phi}e^{b(-\pi/2\pm 2n\pi)}\sin\phi, \qquad (11)\end{aligned}$$

which are equivalent to the form of the original equation if

$$be^{b(-\frac{1}{2}\pi\pm 2n\pi)} = 1 \qquad (12)$$

$$\ln b + b(-\tfrac{1}{2}\pi \pm 2n\pi) = 0 \qquad (13)$$

$$\frac{\ln b}{b} = \tfrac{1}{2}\pi \mp 2n\pi = -(2n-\tfrac{1}{2})\pi, \qquad (14)$$

where only solutions with the minus sign in $\mp$ exist. Solving gives the values summarized in the following table.

n	b_n	$\psi = \cot^{-1} b_n$
1	0.2744106319...	74°39′18.53”
2	0.1642700512...	80°40′16.80”
3	0.1218322508...	83°03′13.53”
4	0.0984064967...	84°22′47.53”
5	0.0832810611...	85°14′21.60”
6	0.0725974881...	85°50′51.92”
7	0.0645958183...	86°18′14.64”
8	0.0583494073...	86°39′38.20”
9	0.0533203211...	86°56′52.30”
10	0.0491732529...	87°11′05.45”

References

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 60–64, 1991.

Logarithmic Spiral Inverse Curve

The INVERSE CURVE of the LOGARITHMIC SPIRAL

$$r = e^{a\theta}$$

with INVERSION CENTER at the origin and inversion radius k is the LOGARITHMIC SPIRAL

$$r = ke^{-a\theta}.$$

Logarithmic Spiral Pedal Curve

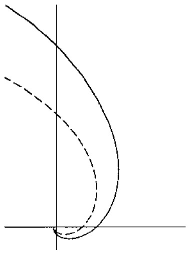

The PEDAL CURVE of a LOGARITHMIC SPIRAL with parametric equation

$$f = e^{at}\cos t \qquad (1)$$

$$g = e^{at}\sin t \qquad (2)$$

for a PEDAL POINT at the pole is an identical LOGARITHMIC SPIRAL

$$x = \frac{(a\sin t + \cos t)e^{at}}{1+a^2} \qquad (3)$$

$$y = \frac{(\sin t - a\cos t)e^{at}}{1+a^2}, \qquad (4)$$

so

$$r = \sqrt{x^2+y^2} = \frac{e^{at}}{\sqrt{1+a^2}}. \qquad (5)$$

Logarithmic Spiral Radial Curve

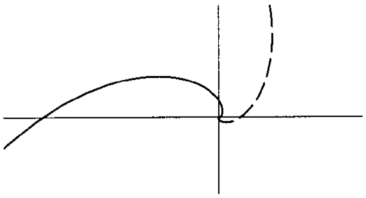

The RADIAL CURVE of the LOGARITHMIC SPIRAL is another LOGARITHMIC SPIRAL.

Logarithmically Convex Function

A function $f(x)$ is logarithmically convex on the interval $[a,b]$ if $f>0$ and $\ln f(x)$ is CONCAVE on $[a,b]$. If $f(x)$ and $g(x)$ are logarithmically convex on the interval $[a,b]$, then the functions $f(x)+g(x)$ and $f(x)g(x)$ are also logarithmically convex on $[a,b]$.

see also CONVEX FUNCTION

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1100, 1980.

Logic

The formal mathematical study of the methods, structure, and validity of mathematical deduction and proof. Formal logic seeks to devise a complete, consistent formulation of mathematics such that propositions can be formally stated and proved using a small number of symbols with well-defined meanings. While this sounds like an admirable pursuit in principle, in practice the study of mathematical logic can rapidly become bogged down in pages of dense and unilluminating mathematical symbols, of which Whitehead and Russell's *Principia Mathematica* (1925) is perhaps the best (or worst) example.

A very simple form of logic is the study of "TRUTH TABLES" and digital logic circuits in which one or more outputs depend on a combination of circuit elements (AND, NAND, OR, XOR, etc.; "gates") and the input values. In such a circuit, values at each point can take on values of only True (1) or False (0). de Morgan's Duality Law is a useful principle for the analysis and simplification of such circuits.

A generalization of this simple type of logic in which possible values are True, False, and "undecided" is called Three-Valued Logic. A further generalization called Fuzzy Logic treats "truth" as a continuous quantity ranging from 0 to 1.

see also Absorption Law, Alethic, Boolean Algebra, Boolean Connective, Bound, Caliban Puzzle, Contradiction Law, de Morgan's Duality Law, de Morgan's Laws, Deducible, Excluded Middle Law, Free, Fuzzy Logic, Gödel's Incompleteness Theorem, Khovanski's Theorem, Logical Paradox, Logos, Löwenheimer-Skolem Theorem, Metamathematics, Model Theory, Quantifier, Sentence, Tarski's Theorem, Tautology, Three-Valued Logic, Topos, Truth Table, Turing Machine, Universal Statement, Universal Turing Machine, Venn Diagram, Wilkie's Theorem

References
Adamowicz, Z. and Zbierski, P. *Logic of Mathematics: A Modern Course of Classical Logic.* New York: Wiley, 1997.
Bogomolny, A. "Falsity Implies Anything." `http://www.cut-the-knot.com/do_you_know/falsity.html`.
Carnap, R. *Introduction to Symbolic Logic and Its Applications.* New York: Dover, 1958.
Church, A. *Introduction to Mathematical Logic, Vol. 1.* Princeton, NJ: Princeton University Press, 1996.
Gödel, K. *On Formally Undecidable Propositions of Principia Mathematica and Related Systems.* New York: Dover, 1992.
Jeffrey, R. C. *Formal Logic: Its Scope and Limits.* New York: McGraw-Hill, 1967.
Kac, M. and Ulam, S. M. *Mathematics and Logic: Retrospect and Prospects.* New York: Dover, 1992.
Kleene, S. C. *Introduction to Metamathematics.* Princeton, NJ: Van Nostrand, 1971.
Whitehead, A. N. and Russell, B. *Principia Mathematica, 2nd ed.* Cambridge, England: Cambridge University Press, 1962.

Logical Paradox

see Paradox

Logistic Distribution

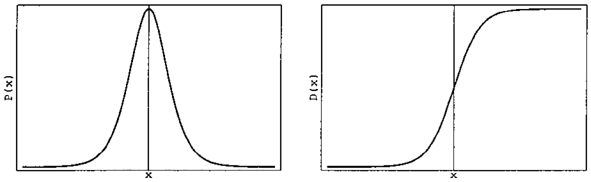

$$P(x) = \frac{e^{(x-m)/b}}{|b|[1+e^{(x-m)/b}]^2} \qquad (1)$$

$$D(x) = \frac{1}{1+e^{(m-x)/|b|}}, \qquad (2)$$

and the Mean, Variance, Skewness, and Kurtosis are

$$\mu = m \qquad (3)$$

$$\sigma^2 = \tfrac{1}{3}\pi^2\beta^2 \qquad (4)$$

$$\gamma_1 = 0 \qquad (5)$$

$$\gamma_2 = \tfrac{6}{5}. \qquad (6)$$

see also Logistic Equation, Logistic Growth Curve

References
von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 250, 1993.

Logistic Equation

The logistic equation (sometimes called the Verhulst Model since it was first published in 1845 by the Belgian P.-F. Verhulst) is defined by

$$x_{n+1} = rx_n(1-x_n), \qquad (1)$$

where r (sometimes also denoted μ) is a Positive constant (the "biotic potential"). We will start x_0 in the interval $[0,1]$. In order to keep points in the interval, we must find appropriate conditions on r. The maximum value x_{n+1} can take is found from

$$\frac{dx_{n+1}}{dx_n} = r(1-2x_n) = 0, \qquad (2)$$

so the largest value of x_{n+1} occurs for $x_n = 1/2$. Plugging this in, $\max(x_{n+1}) = r/4$. Therefore, to keep the Map in the desired region, we must have $r \in (0,4]$. The Jacobian is

$$J = \left|\frac{dx_{n+1}}{dx_n}\right| = |r(1-2x_n)|, \qquad (3)$$

and the Map is stable at a point x_0 if $J(x_0) < 1$. Now we wish to find the Fixed Points of the Map, which occur when $x_{n+1} = x_n$. Drop the n subscript on x_n

$$f(x) = rx(1-x) = x \qquad (4)$$

$$x[1-r(1-x)] = x(1-r+rx) = rx\left[x-\left(1-r^{-1}\right)\right] = 0, \qquad (5)$$

so the Fixed Points are $x_1^{(1)} = 0$ and $x_2^{(1)} = 1-r^{-1}$. An interesting thing happens if a value of r greater than 3 is chosen. The map becomes unstable and we get a Pitchfork Bifurcation with two stable orbits of period two corresponding to the two stable Fixed Points of $f^2(x)$. The fixed points of order two must satisfy $x_{n+2} = x_n$, so

$$\begin{aligned} x_{n+2} &= rx_{n+1}(1-x_{n+1}) \\ &= r[rx_n(1-x_n)][1-rx_n(1-x_n)] \\ &= r^2x_n(1-x_n)(1-rx_n+rx_n{}^2) = x_n. \end{aligned} \qquad (6)$$

Now, drop the n subscripts and rewrite

$$x\{r^2[1 - x(1+r) + 2rx^2 - rx^3] - 1\} = 0 \qquad (7)$$

$$x[-r^3x^3 + 2r^3x^2 - r^2(1+r)x + (r^2 - 1)] = 0 \qquad (8)$$

$$-r^3x[x - (1 - r^{-1})][x^2 - (1 + r^{-1})x + r^{-1}(1 + r^{-1})] = 0. \qquad (9)$$

Notice that we have found the first-order Fixed Points as well, since two iterations of a first-order Fixed Point produce a trivial second-order Fixed Point. The true 2-Cycles are given by solutions to the quadratic part

$$\begin{aligned} x_\pm^{(2)} &= \tfrac{1}{2}[(1 + r^{-1}) \pm \sqrt{(1 + r^{-1})^2 - 4r^{-1}(1 + r^{-1})}\,] \\ &= \tfrac{1}{2}[(1 + r^{-1}) \pm \sqrt{1 + 2r^{-1} + r^{-2} - 4r^{-1} - 4r^{-2}}\,] \\ &= \tfrac{1}{2}[(1 + r^{-1}) \pm \sqrt{1 - 2r^{-1} - 3r^{-2}}\,] \\ &= \tfrac{1}{2}[(1 + r^{-1}) \pm r^{-1}\sqrt{(r-3)(r+1)}\,]. \qquad (10) \end{aligned}$$

These solutions are only Real for $r \geq 3$, so this is where the 2-Cycle begins. Now look for the onset of the 4-Cycle. To eliminate the 2- and 1-Cycles, consider

$$\frac{f^4(x) - x}{f^2(x) - x} = 0. \qquad (11)$$

This gives

$$\begin{aligned} 1+r^2 &+ (-r^2 - r^3 - r^4 - r^5)x \\ &+ (2r^3 + r^4 + 4r^5 + r^6 + 2r^7)x^2 \\ &+ (-r^3 - 5r^5 - 4r^6 - 5r^7 - 4r^8 - r^9)x^3 \\ &+ (2r^5 + 6r^6 + 4r^7 + 14r^8 + 5r^9 + 3r^{10})x^4 \\ &+ (-4r^6 - r^7 - 18r^8 - 12r^9 - 12r^{10} - 3r^{11})x^5 \\ &+ (r^6 + 10r^8 + 17r^9 + 18r^{10} + 15r^{11} + r^{12})x^6 \\ &+ (-2r^8 - 14r^9 - 12r^{10} - 30r^{11} - 6r^{12})x^7 \\ &+ (6r^9 + 3r^{10} + 30r^{11} + 15r^{12})x^8 \\ &+ (-r^9 - 15r^{11} - 20r^{12})x^9 + (3r^{11} + 15r^{12})x^{10} \\ &- 6r^{12}x^{11} + r^{12}x^{12}. \qquad (12) \end{aligned}$$

The Roots of this equation are all Imaginary for $r < 1 + \sqrt{6}$, but two of them convert to Real roots at this value (although this is difficult to show except by plugging in). The 4-Cycle therefore starts at $1 + \sqrt{6} = 3.449490\ldots$. The Bifurcations come faster and faster (8, 16, 32, ...), then suddenly break off. Beyond a certain point known as the Accumulation Point, periodicity gives way to Chaos.

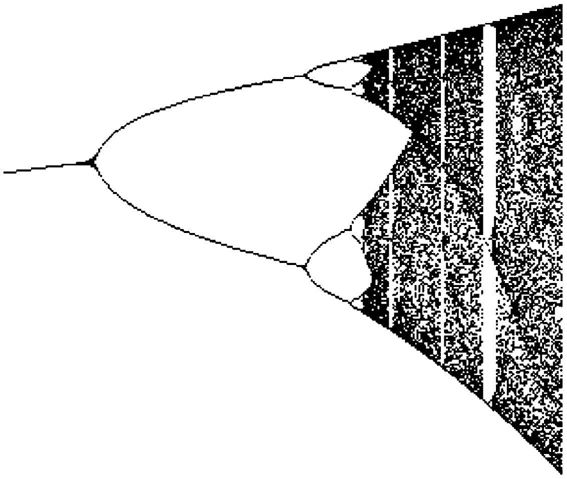

A table of the Cycle type and value of r_n at which the cycle 2^n appears is given below.

n	cycle (2^n)	r_n
1	2	3
2	4	3.449490
3	8	3.544090
4	16	3.564407
5	32	3.568750
6	64	3.56969
7	128	3.56989
8	256	3.569934
9	512	3.569943
10	1024	3.5699451
11	2048	3.569945557
∞	acc. pt.	3.569945672

For additional values, see Rasband (1990, p. 23). Note that the table in Tabor (1989, p. 222) is incorrect, as is the $n = 2$ entry in Lauweirer 1991. In the middle of the complexity, a window suddenly appears with a regular period like 3 or 7 as a result of Mode Locking. The period 3 Bifurcation occurs at $r = 1 + 2\sqrt{2} = 3.828427\ldots$, as is derived below. Following the 3-Cycle, the Period Doublings then begin again with Cycles of 6, 12, ... and 7, 14, 28, ..., and then once again break off to Chaos.

A set of $n+1$ equations which can be solved to give the onset of an arbitrary n-cycle (Saha and Strogatz 1995) is

$$\begin{cases} x_2 = rx_1(1 - x_1) \\ x_3 = rx_2(1 - x_2) \\ \vdots \\ x_n = rx_{n-1}(1 - x_{n-1}) \\ x_1 = rx_n(1 - x_n) \\ r^n \prod_{k=1}^n (1 - 2x_k) = 1. \end{cases} \qquad (13)$$

The first n of these give $f(x)$, $f^2(x)$, ..., $f^n(x)$, and the last uses the fact that the onset of period n occurs by a Tangent Bifurcation, so the nth Derivative is 1.

For $n = 2$, the solutions $(x_1, \ldots, x_n, r)$ are $(0, 0, \pm 1)$ and $(2/3, 2/3, 3)$, so the desired Bifurcation occurs at $r_2 = 3$. Taking $n = 3$ gives

$$\begin{aligned} \frac{d[f^3(x)]}{dx} &= \frac{d[f^3(x)]}{d[f^2(x)]}\frac{d[f^2(x)]}{d[f(x)]}\frac{d[f(x)]}{dx} \\ &= \frac{d[f(z)]}{dz}\frac{d[f(y)]}{dy}\frac{d[f(x)]}{dx} \\ &= r^3(1-2z)(1-2y)(1-2x). \qquad (14) \end{aligned}$$

Solving the resulting Cubic Equation using computer algebra gives

$$\begin{aligned} x_1 = &-\left(\frac{2^{5^6}}{63\cdot 7^{1/3}} + \frac{1}{63\cdot 28^{1/3}}\right)c^2 \\ &- \frac{1}{9\cdot 98^{1/3}}c + \frac{10+\sqrt{2}}{21} + \left(\frac{4\cdot 2^{5/6}}{9\cdot 7^{1/3}} - \frac{2^{1/3}}{7^{1/3}}\right)c^{-1} \\ &+ \frac{25\cdot 28^{1/3} - 44\cdot 2^{1/6}7^{1/3}}{9}c^{-2} \qquad (15) \\ x_2 = &\left(\frac{1}{63\cdot 28^{1/3}} + \frac{2^{5/6}}{63\cdot 7^{1/3}}\right)c^2 - \frac{2^{2/3}}{9\cdot 7^{2/3}}c + \frac{10+\sqrt{2}}{21} \\ &+ \left(\frac{8\cdot 2^{5/6}}{9\cdot 7^{1/3}} - \frac{2\cdot 2^{1/3}}{7^{1/3}}\right)c^{-1} \\ &+ \frac{44\cdot 2^{1/6}7^{1/3} - 25\cdot 28^{1/3}}{9}c^{-2} \qquad (16) \\ x_3 = &\frac{1}{3\cdot 98^{1/3}}c + \frac{10+\sqrt{2}}{21} + \frac{2^{1/3}(9-4\sqrt{2}\,)}{3\cdot 7^{1/3}}c^{-1} \qquad (17) \\ r = &1 + 2\sqrt{2}, \qquad (18) \end{aligned}$$

where

$$c \equiv (-25 + 22\sqrt{2} + 3\sqrt{3}\sqrt{1100\sqrt{2} - 1593}\,)^{1/3}. \qquad (19)$$

Numerically,

$$\begin{aligned} x_1 &= 0.514355\ldots \qquad (20) \\ x_2 &= 0.956318\ldots \qquad (21) \\ x_3 &= 0.159929\ldots \qquad (22) \\ r &= 3.828427\ldots. \qquad (23) \end{aligned}$$

Saha and Strogatz (1995) give a simplified algebraic treatment which involves solving

$$r^3(1 - 2\alpha + 4\beta - 8\gamma) = 1, \qquad (24)$$

together with three other simultaneous equations, where

$$\begin{aligned} \alpha &\equiv x_1 + x_2 + x_3 \qquad (25) \\ \beta &\equiv x_1x_2 + x_1x_3 + x_2x_3 \qquad (26) \\ \gamma &\equiv x_1x_2x_3. \qquad (27) \end{aligned}$$

Further simplifications still are provided in Bechhoeffer (1996) and Gordon (1996), but neither of these techniques generalizes easily to higher Cycles. Bechhoeffer (1996) expresses the three additional equations as

$$\begin{aligned} 2\alpha &= 3 + r^{-1} \qquad (28) \\ 4\beta &= \tfrac{3}{2} + 5r^{-1} + \tfrac{3}{2}r^{-2} \qquad (29) \\ 8\gamma &= -\tfrac{1}{2} + \tfrac{7}{2}r^{-1} + \tfrac{5}{2}r^{-2} + \tfrac{5}{2}r^{-3}, \qquad (30) \end{aligned}$$

giving

$$r^2 - 2r - 7 = 0. \qquad (31)$$

Gordon (1996) derives not only the value for the onset of the 3-Cycle, but also an upper bound for the r-values supporting stable period 3 orbits. This value is obtained by solving the Cubic Equation

$$s^3 - 11s^2 + 37s - 108 = 0 \qquad (32)$$

for s, then

$$\begin{aligned} r' &= 1 + \sqrt{s} \qquad (33) \\ &= 1 + \sqrt{\tfrac{11}{3} + (\tfrac{1915}{54} + \tfrac{5}{2}\sqrt{201}\,)^{1/3} + (\tfrac{1915}{54} - \tfrac{5}{2}\sqrt{201}\,)^{1/3}} \\ &= 3.841499007543\ldots. \qquad (34) \end{aligned}$$

The logistic equation has Correlation Exponent 0.500±0.005 (Grassberger and Procaccia 1983), Capacity Dimension 0.538 (Grassberger 1981), and Information Dimension 0.5170976 (Grassberger and Procaccia 1983).

see also Bifurcation, Feigenbaum Constant, Logistic Distribution, Logistic Equation—$r = 4$, Logistic Growth Curve, Period Three Theorem, Quadratic Map

References

Bechhoeffer, J. "The Birth of Period 3, Revisited." *Math. Mag.* **69**, 115–118, 1996.

Bogomolny, A. "Chaos Creation (There is Order in Chaos)." `http://www.cut-the-knot.com/blue/chaos.html`.

Dickau, R. M. "Bifurcation Diagram." `http://forum.swarthmore.edu/advanced/robertd/bifurcation.html`.

Gleick, J. *Chaos: Making a New Science.* New York: Penguin Books, pp. 69–80, 1988.

Gordon, W. B. "Period Three Trajectories of the Logistic Map." *Math. Mag.* **69**, 118–120, 1996.

Grassberger, P. "On the Hausdorff Dimension of Fractal Attractors." *J. Stat. Phys.* **26**, 173–179, 1981.

Grassberger, P. and Procaccia, I. "Measuring the Strangeness of Strange Attractors." *Physica D* **9**, 189–208, 1983.

Lauwerier, H. *Fractals: Endlessly Repeated Geometrical Figures.* Princeton, NJ: Princeton University Press, pp. 119–122, 1991.

May, R. M. "Simple Mathematical Models with Very Complicated Dynamics." *Nature* **261**, 459–467, 1976.

Peitgen, H.-O.; Jürgens, H.; and Saupe, D. *Chaos and Fractals: New Frontiers of Science.* New York: Springer-Verlag, pp. 585–653, 1992.

Rasband, S. N. *Chaotic Dynamics of Nonlinear Systems.* New York: Wiley, p. 23, 1990.

Russell, D. A.; Hanson, J. D.; and Ott, E. "Dimension of Strange Attractors." *Phys. Rev. Let.* **45**, 1175–1178, 1980.
Saha, P. and Strogatz, S. H. "The Birth of Period Three." *Math. Mag.* **68**, 42–47, 1995.
Strogatz, S. H. *Nonlinear Dynamics and Chaos.* Reading, MA: Addison-Wesley, 1994.
Tabor, M. *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, 1989.
Wagon, S. "The Dynamics of the Quadratic Map." §4.4 in *Mathematica in Action.* New York: W. H. Freeman, pp. 117–140, 1991.

Logistic Equation—$r = 4$

With $r = 4$, the LOGISTIC EQUATION becomes

$$x_{n+1} = 4x_n(1 - x_n). \qquad (1)$$

Now let

$$x \equiv \sin^2(\tfrac{1}{2}\pi y) = \tfrac{1}{2}[1 - \cos(\pi y)] \qquad (2)$$

$$\sqrt{x} = \sin(\tfrac{1}{2}\pi y) \qquad (3)$$

$$y = \frac{2}{\pi}\sin^{-1}(\sqrt{x}\,) \qquad (4)$$

$$\frac{dy}{dx} = \frac{2}{\pi}\frac{1}{\sqrt{1-x}}\tfrac{1}{2}x^{-1/2} = \frac{1}{\pi\sqrt{x(1-x)}}. \qquad (5)$$

Manipulating (2) gives

$$\begin{aligned}&\sin^2(\tfrac{1}{2}\pi y_{n+1})\\ &\quad = 4\tfrac{1}{2}[1 - \cos(\pi y_n)]\{1 - \tfrac{1}{2}[1 - \tfrac{1}{2}(1 - \cos(\pi y_n)]\}\\ &\quad = 2[1 - \cos(\pi y = 1 - \cos^2(\pi y_n)\sin^2(\pi y_n), \qquad (6)\end{aligned}$$

so

$$\tfrac{1}{2}\pi y_{n+1} = \pm y_n + s\pi \qquad (7)$$

$$y_{n+1} = \pm 2y_n + \tfrac{1}{2}s. \qquad (8)$$

But $y \in [0,1]$. Taking $y_n \in [0, 1/2]$, then $s = 0$ and

$$y_{n+1} = 2y_n. \qquad (9)$$

For $y \in [1/2, 1]$, $s = 1$ and

$$y_{n+1} = 2 - 2y_n. \qquad (10)$$

Combining

$$y_n = \begin{cases} 2y_n & \text{for } y_n \in [0, \tfrac{1}{2}] \\ 2 - 2y_n & \text{for } y_n \in [\tfrac{1}{2}, 1], \end{cases} \qquad (11)$$

which can be written

$$y_n = 1 - 2|x_n - h|, \qquad (12)$$

the TENT MAP with $\mu = 1$, so the NATURAL INVARIANT in y is

$$\rho(y) = 1. \qquad (13)$$

Transforming back to x gives

$$\begin{aligned}\rho(x) &= \left|\frac{dy}{dx}\right|\rho(y(x)) = \frac{2}{\pi}\frac{1}{\sqrt{1-x}}\tfrac{1}{2}x^{-1/2}\\ &= \frac{1}{\pi\sqrt{x(1-x)}}. \qquad (14)\end{aligned}$$

This can also be derived from

$$\rho(x) \equiv \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta(x_i - x) = \frac{1}{\pi\sqrt{x(1-x)}}, \qquad (15)$$

where $\delta(x)$ is the DELTA FUNCTION.

see also LOGISTIC EQUATION

Logistic Growth Curve

The POPULATION GROWTH law which arises frequently in biology and is given by the differential equation

$$\frac{dN}{dt} = \frac{r(K - N)}{K}, \qquad (1)$$

where r is the MALTHUSIAN PARAMETER and K is the so-called CARRYING CAPACITY (i.e., the maximum sustainable population). Rearranging and integrating both sides gives

$$\int_{N_0}^{N}\frac{dN}{K - N} = \frac{r}{K}\int_0^t dt \qquad (2)$$

$$\ln\left(\frac{N_0 - K}{N - K}\right) = \frac{r}{K}t \qquad (3)$$

$$N(t) = K + (N_0 - K)e^{-rt/K}. \qquad (4)$$

The curve

$$y = \frac{a}{1 + bq^x} \qquad (5)$$

is sometimes also known as the logical curve.

see also GOMPERTZ CURVE, LIFE EXPECTANCY, LOGISTIC EQUATION, MAKEHAM CURVE, MALTHUSIAN PARAMETER, POPULATION GROWTH

Logistic Map

see LOGISTIC EQUATION

Logit Transformation

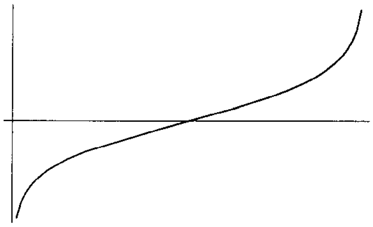

The function

$$z = f(x) = \ln\left(\frac{x}{1-x}\right).$$

This function has an inflection point at $x = 1/2$, where

$$f''(x) = \frac{2x-1}{x^2(x-1)^2} = 0.$$

Applying the logit transformation to values obtained by iterating the LOGISTIC EQUATION generates a sequence of RANDOM NUMBERS having distribution

$$P_z = \frac{1}{\pi(e^{x/2} + e^{-x/2})},$$

which is very close to a GAUSSIAN DISTRIBUTION.

References

Collins, J.; Mancilulli, M.; Hohlfeld, R.; Finch, D.; Sandri, G.; and Shtatland, E. "A Random Number Generator Based on the Logit Transform of the Logistic Variable." *Computers in Physics* **6**, 630–632, 1992.

Pickover, C. A. *Keys to Infinity.* New York: W. H. Freeman, pp. 244–245, 1995.

Logos

A generalization of a HEYTING ALGEBRA which replaces BOOLEAN ALGEBRA in "intuitionistic" LOGIC.

see also TOPOS

Lommel Differential Equation

A generalization of the BESSEL DIFFERENTIAL EQUATION (Watson 1966, p. 345),

$$z^2 \frac{d^2y}{dz^2} + z\frac{dy}{dz} - (z^2 + \nu^2)y = kz^{\mu+1}.$$

A further generalization gives

$$z^2 \frac{d^2y}{dz^2} + z\frac{dy}{dz} - (z^2 + \nu^2)y = \pm kz^{\mu+1}.$$

The solutions are LOMMEL FUNCTIONS.

see also LOMMEL FUNCTION

References

Watson, G. N. *A Treatise on the Theory of Bessel Functions, 2nd ed.* Cambridge, England: Cambridge University Press, 1966.

Lommel Function

There are several functions called "Lommel functions." One type of Lommel function is the solution to the LOMMEL DIFFERENTIAL EQUATION with a PLUS SIGN,

$$y = ks_{\mu,\nu}(z), \tag{1}$$

where

$$s_{\mu,\nu}^{(+)}(z) \equiv \tfrac{1}{2}\pi \left[Y_\nu(z) \int_0^z z^\mu J_\nu(z)\,dz - J_\nu(z) \int_0^z z^\mu Y_\nu(z)\,dz \right]. \tag{2}$$

Here, $J_\nu(z)$ and $Y_\nu(z)$ are BESSEL FUNCTIONS OF THE FIRST and SECOND KINDS (Watson 1966, p. 346). If a minus sign precedes k, then the solution is

$$s_{\mu,\nu}^{(-)} \equiv I_\nu(z) \int_z^{c_1} z^\mu K_\nu(z)\,dz - J_\nu(z) \int_{c_2}^z z^\mu I_\nu(z)\,dz, \tag{3}$$

where $K_\nu(z)$ and $I_\nu(z)$ are MODIFIED BESSEL FUNCTIONS OF THE FIRST and SECOND KINDS.

Lommel functions of two variables are related to the BESSEL FUNCTION OF THE FIRST KIND and arise in the theory of diffraction and, in particular, Mie scattering (Watson 1966, p. 537),

$$U_n(w,z) = \sum_{m=0}^{\infty} (-1)^m \left(\frac{w}{z}\right)^{n+2m} J_{n+2m}(z) \tag{4}$$

$$V_n(w,z) = \sum_{m=0}^{\infty} (-1)^m \left(\frac{w}{z}\right)^{-n-2m} J_{-n-2m}(z). \tag{5}$$

see also LOMMEL DIFFERENTIAL EQUATION, LOMMEL POLYNOMIAL

References

Chandrasekhar, S. *Radiative Transfer.* New York: Dover, p. 369, 1960.

Watson, G. N. *A Treatise on the Theory of Bessel Functions, 2nd ed.* Cambridge, England: Cambridge University Press, 1966.

Lommel's Integrals

$$(\beta^2 - \alpha^2) \int xJ_n(\alpha x)J_n(\beta x)\,dx = x[\alpha J_n'(\alpha x)J_n(\beta x) - \beta J_n'(\beta x)J_n(\alpha x)]$$

$$\int x{J_n}^2(\alpha x)\,dx = \tfrac{1}{2}x^2[{J_n}^2(\alpha x) + J_{n-1}(\alpha x)J_{n+1}(\alpha x)],$$

where $J_n(x)$ is a BESSEL FUNCTION OF THE FIRST KIND.

References

Bowman, F. *Introduction to Bessel Functions.* New York: Dover, p. 101, 1958.

Lommel Polynomial

$$\begin{aligned} &R_{m,\nu}(z) \\ &= \frac{\Gamma(\nu+m)}{\Gamma(\nu)(z/2)^m}\, {}_2F_3(\tfrac{1}{2}(1-m), -\tfrac{1}{2}m; \nu, -m, 1-\nu-m; z^2) \\ &\quad \times \frac{\pi z}{2\sin(\nu\pi)}[J_{\nu+m}(z)J_{-\nu+1}(z) \\ &\qquad +(-1)^m J_{-\nu-m}(z)J_{\nu-1}(z)], \end{aligned}$$

where $\Gamma(z)$ is a GAMMA FUNCTION, $J_n(x)$ is a BESSEL FUNCTION OF THE FIRST KIND, and ${}_2F_3(a, b; c, d, e; z)$ is a GENERALIZED HYPERGEOMETRIC FUNCTION.

see also LOMMEL FUNCTION

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1477, 1980.

Long Division

$$
\begin{array}{rr}
 & 7 \\ \cline{2-2}
17 & \big)\,123456. \\
 & -\underline{119} \\
 & 44
\end{array}
\qquad
\begin{array}{rr}
 & 72 \\ \cline{2-2}
17 & \big)\,123456. \\
 & -\underline{119} \\
 & 44 \\
 & -\underline{34} \\
 & 105
\end{array}
\qquad
\begin{array}{rr}
 & 726 \\ \cline{2-2}
17 & \big)\,123456. \\
 & -\underline{119} \\
 & 44 \\
 & -\underline{34} \\
 & 105 \\
 & -\underline{102} \\
 & 36
\end{array}
$$

$$
\begin{array}{rr}
 & 7262.1 \\ \cline{2-2}
17 & \big)\,123456.0 \\
 & -\underline{119} \\
 & 44 \\
 & -\underline{34} \\
 & 105 \\
 & -\underline{102} \\
 & 36 \\
 & -\underline{34} \\
 & 20
\end{array}
\qquad
\begin{array}{rr}
 & 7262.11\ldots \\ \cline{2-2}
17 & \big)\,123456.00 \\
 & -\underline{119} \\
 & 44 \\
 & -\underline{34} \\
 & 105 \\
 & -\underline{102} \\
 & 36 \\
 & -\underline{34} \\
 & 20 \\
 & -\underline{17} \\
 & 30
\end{array}
$$

Long division is an algorithm for dividing two numbers, obtaining the QUOTIENT one DIGIT at a time. The above example shows how the division of 123456/17 is performed to obtain the result 7262.11....

see also DIVISION

Long Exact Sequence of a Pair Axiom

One of the EILENBERG-STEENROD AXIOMS. It states that, for every pair (X, A), there is a natural long exact sequence

$$\ldots \to H_n(A) \to H_n(X) \to H_n(X, A) \to H_{n-1}(A) \to \ldots,$$

where the MAP $H_n(A) \to H_n(X)$ is induced by the INCLUSION MAP $A \to X$, $H_n(X) \to H_n(X, A)$ is induced by the INCLUSION MAP $(X, \phi)^{\check{}} \to (X, A)$. The MAP $H_n(X, A) \to H_{n-1}(A)$ is called the BOUNDARY MAP.

see also EILENBERG-STEENROD AXIOMS

Long Prime

see DECIMAL EXPANSION

Longitude

The azimuthal coordinate on the surface of a SPHERE (θ in SPHERICAL COORDINATES) or on a SPHEROID (in PROLATE or OBLATE SPHEROIDAL COORDINATES). Longitude is defined such that 0° = 360°. Lines of constant longitude are generally called MERIDIANS. The other angular coordinate on the surface of a SPHERE is called the LATITUDE.

The shortest distance between any two points on a SPHERE is the so-called GREAT CIRCLE distance, which can be directly computed from the LATITUDE and longitudes of two points.

see also GREAT CIRCLE, LATITUDE, MERIDIAN, OBLATE SPHEROIDAL COORDINATES, PROLATE SPHEROIDAL COORDINATES

Look and Say Sequence

The INTEGER SEQUENCE beginning with a single digit in which the next term is obtained by describing the previous term. Starting with 1, the sequence would be defined by "one 1, two 1s, one 2 two 1s," etc., and the result is 1, 11, 21, 1211, 111221, 312211, 13112221, 1113213211, ... (Sloane's A005150).

Starting the sequence instead with the digit d for $2 \leq d \leq 9$ gives d, $1d$, $111d$, $311d$, $13211d$, $111312211d$, $31131122211d$, $1321132132211d$, ... The sequences for $d = 2$ and 3 are Sloane's A006715 and A006751. The number of DIGITS in the nth term of both the sequences for $1 \leq n \leq 9$ is asymptotic to $C\lambda^n$, where C is a constant and

$$\lambda = 1.3035772690342 96\ldots$$

(Sloane's A014715) is CONWAY'S CONSTANT. λ is given by the largest ROOT of the POLYNOMIAL

$$
\begin{aligned}
0 = {} & x^{71} \\
& - x^{69} - 2x^{68} + 2x^{66} + 2x^{65} + x^{64} - x^{63} - x^{62} - x^{61} \\
& - x^{60} - x^{59} + 2x^{58} + 5x^{57} + 3x^{56} - 2x^{55} - 10x^{54} \\
& - 3x^{53} - 2x^{52} + 6x^{51} + 6x^{50} + x^{49} + 9x^{48} - 3x^{47} \\
& - 7x^{46} - 8x^{45} - 8x^{44} + 10x^{43} + 6x^{42} + 8x^{41} - 4x^{40} \\
& - 12x^{39} + 7x^{38} - 7x^{37} + 7x^{36} - 3x^{34} + x^{35} + 10x^{33} \\
& + x^{32} - 6x^{31} - 2x^{30} - 10x^{29} - 3x^{28} + 2x^{27} + 9x^{26} \\
& - 3x^{25} + 14x^{24} - 8x^{23} - 7x^{21} + 9x^{20} - 3x^{19} - 4x^{18} \\
& - 10x^{17} - 7x^{16} + 12x^{15} + 7x^{14} + 2x^{13} - 12x^{12} \\
& - 4x^{11} - 2x^{10} - 5x^{9} + x^{7} - 7x^{6} \\
& + 7x^{5} - 4x^{4} + 12x^{3} - 6x^{2} + 3x - 6.
\end{aligned}
$$

In fact, the constant is even more general than this, applying to *all* starting sequences (i.e., even those starting with arbitrary starting digits), with the exception of 22, a result which follows from the COSMOLOGICAL THEOREM. Conway discovered that strings sometimes factor as a concatenation of two strings whose descendants never interfere with one another. A string with no nontrivial splittings is called an "element," and other strings are called "compounds." Every string of 1s, 2s, and 3s eventually "decays" into a compound of 92 special elements, named after the chemical elements.

see also CONWAY'S CONSTANT, COSMOLOGICAL THEOREM

References

Conway, J. H. "The Weird and Wonderful Chemistry of Audioactive Decay." *Eureka,* 5–18, 1985.

Conway, J. H. "The Weird and Wonderful Chemistry of Audioactive Decay." §5.11 in *Open Problems in Communications and Computation.* (Ed. T. M. Cover and B. Gopinath). New York: Springer-Verlag, pp. 173–188, 1987.

Conway, J. H. and Guy, R. K. "The Look and Say Sequence." In *The Book of Numbers.* New York: Springer-Verlag, pp. 208–209, 1996.

Sloane, N. J. A. Sequences A005150/M4780, A006715/M2965, and A6751/M2052 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 13–14, 1991.

Loop (Algebra)

A nonassociative ALGEBRA (and QUASIGROUP) which has a single binary operation.

Loop Gain

The loop gain is usually assigned a value between 0.1 and 0.5. The CLEAN ALGORITHM performs better for extended structures if μ is set to the lower part of this range. However, the time required for the CLEAN ALGORITHM increases rapidly for small μ. From Thompson *et al.* (1986), the number of cycles needed for one point source is

$$[\text{cycles}] = -\frac{\ln(\text{SNR})}{\ln(1-\gamma)}.$$

see also CLEAN ALGORITHM

References

Thompson, A. R.; Moran, J. M.; and Swenson, G. W. Jr. *Interferometry and Synthesis in Radio Astronomy.* New York: Wiley, p. 348, 1986.

Loop (Graph)

A degenerate edge of a graph which joins a vertex to itself.

Loop (Knot)

A KNOT or HITCH which holds its form rigidly.

References

Owen, P. *Knots.* Philadelphia, PA: Courage, p. 35, 1993.

Loop Space

Let Y^X be the set of continuous mappings $f: X \to Y$. Then the TOPOLOGICAL SPACE for Y^X supplied with a compact-open topology is called a MAPPING SPACE, and if $Y = I$ is taken as the interval $(0,1)$, then $Y^I = \Omega(Y)$ is called a loop space (or SPACE OF CLOSED PATHS).

see also MACHINE, MAPPING SPACE, MAY-THOMASON UNIQUENESS THEOREM

References

Brylinski, J.-L. *Loop Spaces, Characteristic Classes and Geometric Quantization.* Boston, MA: Birkhäuser, 1993.

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 658, 1980.

Lorentz Group

The Lorentz group is the GROUP L of time-preserving linear ISOMETRIES of MINKOWSKI SPACE $\mathbb{R}^4$ with the pseudo-Riemannian metric

$$d\tau^2 = -dt^2 + dx^2 + dy^2 + dz^2.$$

It is also the GROUP of ISOMETRIES of 3-D HYPERBOLIC SPACE. It is time-preserving in the sense that the unit time VECTOR $(1,0,0,0)$ is sent to another VECTOR (t,x,y,z) such that $t > 0$.

A consequence of the definition of the Lorentz group is that the full GROUP of time-preserving isometries of MINKOWSKI $\mathbb{R}^4$ is the DIRECT PRODUCT of the group of translations of $\mathbb{R}^4$ (i.e., $\mathbb{R}^4$ itself, with addition as the group operation), with the Lorentz group, and that the full isometry group of the MINKOWSKI $\mathbb{R}^4$ is a group extension of $\mathbb{Z}_2$ by the product $L \otimes \mathbb{R}^4$.

The Lorentz group is invariant under space rotations and LORENTZ TRANSFORMATIONS.

see also LORENTZ TENSOR, LORENTZ TRANSFORMATION

References

Arfken, G. "Homogeneous Lorentz Group." §4.13 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 271–275, 1985.

Lorentz Tensor

The TENSOR in the LORENTZ TRANSFORMATION given by

$$\mathsf{L} \equiv \begin{bmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

where beta and gamma are defined by

$$\beta \equiv \frac{v}{c} \tag{2}$$

$$\gamma \equiv \frac{1}{\sqrt{1-\beta^2}}. \tag{3}$$

see also LORENTZ GROUP, LORENTZ TRANSFORMATION

Lorentz Transformation

A 4-D transformation satisfied by all FOUR-VECTORS a^ν,

$$a'^\mu = \Lambda^\mu_\nu a^\nu. \tag{1}$$

In the theory of special relativity, the Lorentz transformation replaces the GALILEAN TRANSFORMATION as the valid transformation law between reference frames moving with respect to one another at constant VELOCITY. Let x^ν be the POSITION FOUR-VECTOR with $x^0 = ct$, and let the relative motion be along the x^1 axis with VELOCITY v. Then (1) becomes

$$x'^\mu = \Lambda^\mu_\nu x^\nu, \tag{2}$$

where the LORENTZ TENSOR is given by

$$\mathsf{L} = \begin{bmatrix} \Lambda^0_0 & \Lambda^0_1 & \Lambda^0_2 & \Lambda^0_3 \\ \Lambda^1_0 & \Lambda^1_1 & \Lambda^1_2 & \Lambda^1_3 \\ \Lambda^2_0 & \Lambda^2_1 & \Lambda^2_2 & \Lambda^2_3 \\ \Lambda^3_0 & \Lambda^3_1 & \Lambda^3_2 & \Lambda^3_3 \end{bmatrix} \equiv \begin{bmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{3}$$

Here,

$$\beta \equiv \frac{v}{c} \tag{4}$$

$$\gamma \equiv \frac{1}{\sqrt{1-\beta^2}}. \tag{5}$$

Written explicitly, the transformation between x^ν and $x^{\nu\prime}$ coordinate is

$$x^{0\prime} = \gamma(x^0 - \beta x^1) \tag{6}$$

$$x^{1\prime} = \gamma(x^1 - \beta x^0) \tag{7}$$

$$x^{2\prime} = x^2 \tag{8}$$

$$x^{3\prime} = x^3. \tag{9}$$

The DETERMINANT of the upper left 2×2 MATRIX in (3) is

$$D = (\gamma)^2 - (-\gamma\beta)^2 = \gamma^2(1-\beta^2) = \frac{\gamma^2}{\gamma^2} = 1, \tag{10}$$

so

$$\mathsf{L}^{-1} = \begin{bmatrix} (\Lambda^{-1})^0_0 & (\Lambda^{-1})^0_1 & (\Lambda^{-1})^0_2 & (\Lambda^{-1})^0_3 \\ (\Lambda^{-1})^1_0 & (\Lambda^{-1})^1_1 & (\Lambda^{-1})^1_2 & (\Lambda^{-1})^1_3 \\ (\Lambda^{-1})^2_0 & (\Lambda^{-1})^2_1 & (\Lambda^{-1})^2_2 & (\Lambda^{-1})^2_3 \\ (\Lambda^{-1})^3_0 & (\Lambda^{-1})^3_1 & (\Lambda^{-1})^3_2 & (\Lambda^{-1})^3_3 \end{bmatrix} \equiv \begin{bmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{11}$$

A Lorentz transformation along the x_1-axis can also be written

$$\begin{bmatrix} x_1{}' \\ x_2{}' \\ x_3{}' \\ x_4{}' \end{bmatrix} = \begin{bmatrix} \cosh\theta & i\sinh\theta & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\sinh\theta & \cosh\theta & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}, \tag{12}$$

where θ is called the rapidity,

$$x_4 \equiv ict, \tag{13}$$

and

$$\tanh\theta \equiv \beta \equiv \frac{v}{c} \tag{14}$$

$$\cosh\theta \equiv \gamma \equiv \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \tag{15}$$

$$\sinh\theta = \beta\gamma. \tag{16}$$

see also HYPERBOLIC ROTATION, LORENTZ GROUP, LORENTZ TENSOR

References

Fraundorf, P. "Accel-1D: Frame-Dependent Relativity at UM-StL." `http://www.umsl.edu/~fraundor/altoc.html`.

Griffiths, D. J. *Introduction to Electrodynamics.* Englewood Cliffs, NJ: Prentice-Hall, pp. 412–414, 1981.

Morse, P. M. and Feshbach, H. "The Lorentz Transformation, Four-Vectors, Spinors." §1.7 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 93–107, 1953.

Lorentzian Distribution

see CAUCHY DISTRIBUTION

Lorentzian Function

The Lorentzian function is given by

$$L(x) = \frac{1}{\pi} \frac{\frac{1}{2}\Gamma}{(x-x_0)^2 + (\frac{1}{2}\Gamma)^2}.$$

Its FULL WIDTH AT HALF MAXIMUM is Γ. This function gives the shape of certain types of spectral lines and is the distribution function in the CAUCHY DISTRIBUTION. The Lorentzian function has FOURIER TRANSFORM

$$\mathcal{F}\left[\frac{1}{\pi} \frac{\frac{1}{2}\Gamma}{(x-x_0)^2 + (\frac{1}{2}\Gamma)^2}\right] = e^{-2\pi i k x_0 - \Gamma\pi|k|}.$$

see also DAMPED EXPONENTIAL COSINE INTEGRAL, FOURIER TRANSFORM—LORENTZIAN FUNCTION

Lorenz System

A simplified system of equations describing the 2-D flow of fluid of uniform depth H, with an imposed temperature difference ΔT, under gravity g, with buoyancy α, thermal diffusivity κ, and kinematic viscosity ν. The full equations are

$$\frac{\partial}{\partial t}(\nabla^2\phi) = \frac{\partial\psi}{\partial z}\frac{\partial}{\partial x}(\nabla^2\psi) - \frac{\partial\psi}{\partial x}\frac{\partial}{\partial z}(\nabla^2\psi) + \nu\nabla^2(\nabla^2\psi) + g\alpha\frac{dT}{dx} \tag{1}$$

$$\frac{\partial T}{\partial t} = \frac{\partial T}{\partial z}\frac{\partial\psi}{\partial x} - \frac{\partial\theta}{\partial x}\frac{\partial\psi}{\partial z} + \kappa\nabla^2 T + \frac{\Delta T}{H}\frac{\partial\psi}{\partial x}. \tag{2}$$

Here, ψ is the "stream function," as usual defined such that

$$u = \frac{\partial \psi}{\partial x}, \qquad v = \frac{\partial \psi}{\partial x}. \tag{3}$$

In the early 1960s, Lorenz accidentally discovered the chaotic behavior of this system when he found that, for a simplified system, periodic solutions of the form

$$\psi = \psi_0 \sin\left(\frac{\pi a x}{H}\right) \sin\left(\frac{\pi z}{H}\right) \tag{4}$$

$$\theta = \theta_0 \cos\left(\frac{\pi a x}{H}\right) \sin\left(\frac{\pi z}{H}\right) \tag{5}$$

grew for Rayleigh numbers larger than the critical value, $Ra > Ra_c$. Furthermore, vastly different results were obtained for very small changes in the initial values, representing one of the earliest discoveries of the so-called BUTTERFLY EFFECT.

Lorenz included the following terms in his system of equations,

$$X \equiv \psi_{11} \propto \text{ convective intensity} \tag{6}$$

$$Y \equiv T_{11} \propto \Delta T \text{ between descending and ascending currents} \tag{7}$$

$$Z \equiv T_{02} \propto \Delta \text{ vertical temperature profile from linearity,} \tag{8}$$

and obtained the simplified equations

$$\dot{X} = \sigma(Y - X) \tag{9}$$

$$\dot{Y} = -XZ + rX - Y \tag{10}$$

$$\dot{Z} = XY - bZ, \tag{11}$$

where

$$\sigma \equiv \frac{\nu}{\kappa} = \text{ Prandtl number} \tag{12}$$

$$r \equiv \frac{Ra}{Ra_c} = \text{ normalized Rayleigh number} \tag{13}$$

$$b \equiv \frac{4}{1 + a^2} = \text{ geometric factor.} \tag{14}$$

Lorenz took $b \equiv 8/3$ and $\sigma \equiv 10$.

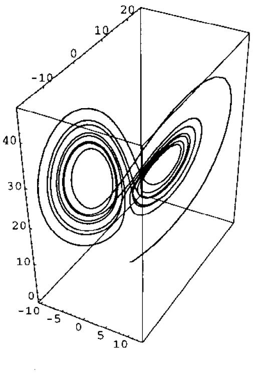

The CRITICAL POINTS at (0, 0, 0) correspond to no convection, and the CRITICAL POINTS at

$$(\sqrt{b(r-1)}, \sqrt{b(r-1)}, r-1) \tag{15}$$

and

$$(-\sqrt{b(r-1)}, -\sqrt{b(r-1)}, r-1) \tag{16}$$

correspond to steady convection. This pair is stable only if

$$r = \frac{\sigma(\sigma + b + 3)}{\sigma - b - 1}, \tag{17}$$

which can hold only for POSITIVE r if $\sigma > b + 1$. The Lorenz attractor has a CORRELATION EXPONENT of 2.05 ± 0.01 and CAPACITY DIMENSION 2.06 ± 0.01 (Grassberger and Procaccia 1983). For more details, see Lichtenberg and Lieberman (1983, p. 65) and Tabor (1989, p. 204).

see also BUTTERFLY EFFECT, RÖSSLER MODEL

References

Gleick, J. *Chaos: Making a New Science.* New York: Penguin Books, pp. 27–31, 1988.

Grassberger, P. and Procaccia, I. "Measuring the Strangeness of Strange Attractors." *Physica D* **9**, 189–208, 1983.

Lichtenberg, A. and Lieberman, M. *Regular and Stochastic Motion.* New York: Springer-Verlag, 1983.

Lorenz, E. N. "Deterministic Nonperiodic Flow." *J. Atmos. Sci.* **20**, 130–141, 1963.

Peitgen, H.-O.; Jürgens, H.; and Saupe, D. *Chaos and Fractals: New Frontiers of Science.* New York: Springer-Verlag, pp. 697–708, 1992.

Tabor, M. *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, 1989.

Lorraine Cross

see GAULLIST CROSS

Lotka-Volterra Equations

An ecological model which assumes that a population x increases at a rate $dx = Ax\,dt$, but is destroyed at a rate $dx = -Bxy\,dt$. Population y decreases at a rate $dy = -Cy\,dt$, but increases at $dy = Dxy\,dt$, giving the coupled differential equations

$$\frac{dx}{dt} = Ax - Bxy$$

$$\frac{dy}{dt} = -Cy + Dxy.$$

Critical points occur when $dx/dt = dy/dt = 0$, so

$$A - By = 0$$

$$-C + Dx = 0.$$

The sole STATIONARY POINT is therefore located at $(x, y) = (C/D, A/B)$.

Low-Dimensional Topology

Low-dimensional topology usually deals with objects that are 2-, 3-, or 4-dimensional in nature. Properly speaking, low-dimensional topology should be part of DIFFERENTIAL TOPOLOGY, but the general machinery of ALGEBRAIC and DIFFERENTIAL TOPOLOGY gives only limited information. This fact is particularly noticeable in dimensions three and four, and so alternative specialized methods have evolved.

see also ALGEBRAIC TOPOLOGY, DIFFERENTIAL TOPOLOGY, TOPOLOGY

Löwenheimer-Skolem Theorem

A fundamental result in MODEL THEORY which states that if a countable theory has a model, then it has a countable model. Furthermore, it has a model of every CARDINALITY greater than or equal to $\aleph_0$ (ALEPH-0). This theorem established the existence of "nonstandard" models of arithmetic.

see also ALEPH-0 ($\aleph_0$), CARDINALITY, MODEL THEORY

References

Chang, C. C. and Keisler, H. J. *Model Theory, 3rd enl. ed.* New York: Elsevier, 1990.

Lower Bound

see GREATEST LOWER BOUND

Lower Denjoy Sum

see LOWER SUM

Lower Integral

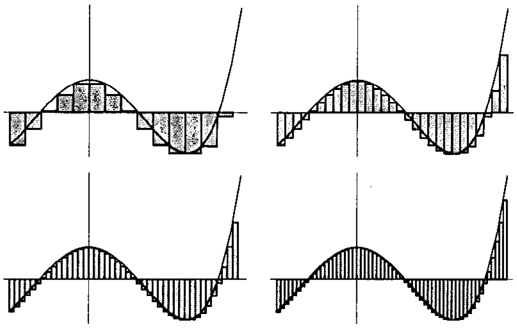

The limit of a LOWER SUM, when it exists, as the MESH SIZE approaches 0.

see also LOWER SUM, RIEMANN INTEGRAL, UPPER INTEGRAL

Lower Limit

Let the least term h of a SEQUENCE be a term which is smaller than all but a finite number of the terms which are equal to h. Then h is called the lower limit of the SEQUENCE.

A lower limit of a SERIES

$$\text{lower}\lim_{n\to\infty} S_n = \underline{\lim}_{n\to\infty} S_n = h$$

is said to exist if, for every $\epsilon > 0$, $|S_n - h| < \epsilon$ for infinitely many values of n and if no number less than h has this property.

see also LIMIT, UPPER LIMIT

References

Bromwich, T. J. I'a and MacRobert, T. M. "Upper and Lower Limits of a Sequence." §5.1 in *An Introduction to the Theory of Infinite Series, 3rd ed.* New York: Chelsea, p. 40 1991.

Lower Sum

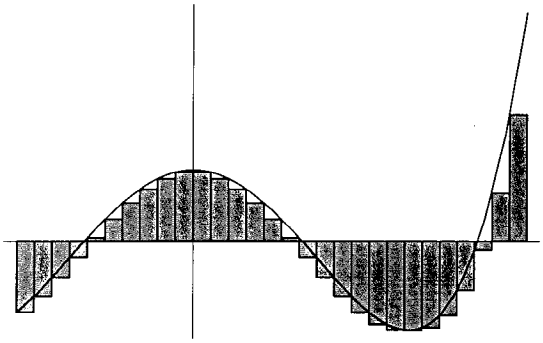

For a given function $f(x)$ over a partition of a given interval, the lower sum is the sum of box areas $f(x_k^*)\Delta x_k$ using the smallest value of the function $f(x_k^*)$ in each subinterval Δx_k.

see also LOWER INTEGRAL, RIEMANN INTEGRAL, UPPER SUM

Lower-Trimmed Subsequence

The lower-trimmed subsequence of $x = \{x_n\}$ is the sequence $V(x)$ obtained by subtracting 1 from each x_n and then removing all 0s. If x is a FRACTAL SEQUENCE, then $V(x)$ is a FRACTAL SEQUENCE. If x is a SIGNATURE SEQUENCE, then $V(x) = x$.

see also SIGNATURE SEQUENCE, UPPER-TRIMMED SUBSEQUENCE

References

Kimberling, C. "Fractal Sequences and Interspersions." *Ars Combin.* **45**, 157–168, 1997.

Lowest Terms Fraction

A FRACTION p/q for which $(p,q) = 1$, where (p,q) denotes the GREATEST COMMON DIVISOR.

Loxodrome

A path, also known as a RHUMB LINE, which cuts a MERIDIAN on a given surface (usually a SPHERE, in which case the loxodrome is also called a SPHERICAL HELIX) at any constant ANGLE but a RIGHT ANGLE. The loxodrome is the path taken when a compass is kept pointing in a constant direction. It is *not* the shortest distance between two points.

see also GREAT CIRCLE

Lozenge

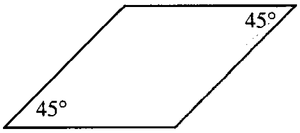

A PARALLELOGRAM whose ACUTE ANGLES are 45°.

see also DIAMOND, PARALLELOGRAM, QUADRILATERAL, RHOMBUS

Lozenge Method

A method for constructing MAGIC SQUARES of ODD order.

see also MAGIC SQUARE

Lozi Map

A 2-D map similar to the HÉNON MAP which is given by the equations

$$\begin{aligned} x_{n+1} &= 1 - \alpha|x_n| + y_n \\ y_{n+1} &= \beta x_n. \end{aligned}$$

see also HÉNON MAP

References

Dickau, R. M. "Lozi Attractor." `http://www.prairienet.org/~pops/lozi.html`.

Peitgen, H.-O.; Jürgens, H.; and Saupe, D. §12.1 in *Chaos and Fractals: New Frontiers of Science.* New York: Springer-Verlag, p. 672, 1992.

LU Decomposition

A procedure for decomposing an $N \times N$ matrix A into a product of a lower TRIANGULAR MATRIX L and an upper TRIANGULAR MATRIX U,

$$\mathsf{LU} = \mathsf{A}. \tag{1}$$

Written explicitly for a 3×3 MATRIX, the decomposition is

$$\begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{2}$$

$$\begin{bmatrix} l_{11}u_{11} & l_{11}u_{12} & l_{11}u_{13} \\ l_{21}u_{11} & l_{21}u_{22} + l_{22}u_{22} & l_{21}u_{13} + l_{22}u_{23} \\ l_{31}u_{11} & l_{31}u_{12} + l_{32}u_{22} & l_{31}u_{13} + l_{32}u_{23} + l_{33}u_{23} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}. \tag{3}$$

This gives three types of equations

$$i < j \qquad l_{i1}u_{ij} + l_{i2}u_{2j} + \ldots + l_{ii}u_{ij} = a_{ij} \tag{4}$$
$$i = j \qquad l_{i1}u_{ij} + l_{i2}u_{2j} + \ldots + l_{ii}u_{jj} = a_{ij} \tag{5}$$
$$i > j \qquad l_{i1}u_{1j} + l_{i2}u_{2j} + \ldots + l_{ij}u_{jj} = a_{ij}. \tag{6}$$

This gives N^2 equations for $N^2 + N$ unknowns (the decomposition is not unique), and can be solved using CROUT'S METHOD. To solve the MATRIX equation

$$\mathsf{A}\mathbf{x} = (\mathsf{LU})\mathbf{x} = \mathsf{L}(\mathsf{U}\mathbf{x}) = \mathbf{b}, \tag{7}$$

first solve $\mathsf{L}\mathbf{y} = \mathbf{b}$ for $\mathbf{y}$. This can be done by forward substitution

$$y_1 = \frac{b_1}{l_{11}} \tag{8}$$

$$y_i = \frac{1}{l_{ii}} \left(b_i - \sum_{j=1}^{i-1} l_{ij} y_j \right) \tag{9}$$

for $i = 2, \ldots, N$. Then solve $\mathsf{U}\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$. This can be done by back substitution

$$x_N = \frac{y_N}{u_{NN}} \tag{10}$$

$$x_i = \frac{1}{u_{ii}} \left(y_i - \sum_{j=i+1}^{N} u_{ij} x_j \right) \tag{11}$$

for $i = N - 1, \ldots, 1$.

see also CHOLESKY DECOMPOSITION, QR DECOMPOSITION, TRIANGULAR MATRIX

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "LU Decomposition and Its Applications." §2.3 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 34–42, 1992.

Lucas Correspondence

The correspondence which relates the HANOI GRAPH to the ISOMORPHIC GRAPH of the ODD BINOMIAL COEFFICIENTS in PASCAL'S TRIANGLE, where the adjacencies are determined by adjacency (either horizontal or diagonal) in PASCAL'S TRIANGLE. The proof that the correspondence is given by the LUCAS CORRESPONDENCE THEOREM.

see also BINOMIAL COEFFICIENT, HANOI GRAPH, PASCAL'S TRIANGLE

References

Poole, David G. "The Towers and Triangles of Professor Claus (or, Pascal Knows Hanoi)." *Math. Mag.* **67**, 323–344, 1994.

Lucas Correspondence Theorem

Let p be PRIME and

$$r = r_m p^m + \ldots + r_1 p + r_0 \qquad (0 \le r_i < p) \tag{1}$$
$$k = k_m p^m + \ldots + k_1 p + k_0 \qquad (0 \le k_i < p), \tag{2}$$

then

$$\binom{r}{k} = \prod_{i=0}^{m} \binom{r_i}{k_i} \pmod{p}. \tag{3}$$

This is proved in Fine (1947).

References

Fine, N. J. "Binomial Coefficients Modulo a Prime." *Amer. Math. Monthly* **54**, 589–592, 1947.

Lucas-Lehmer Residue

see LUCAS-LEHMER TEST

Lucas-Lehmer Test

A MERSENNE NUMBER M_p is prime IFF M_p divides s_{p-2}, where $s_0 \equiv 4$ and

$$s_i \equiv {s_{i-1}}^2 - 2 (\bmod\ 2^p - 1) \tag{1}$$

for $i \geq 1$. The first few terms of this series are 4, 14, 194, 37634, 1416317954, ... (Sloane's A003010). The remainder when s_{p-2} is divided by M_p is called the LUCAS-LEHMER RESIDUE for p. The LUCAS-LEHMER RESIDUE is 0 IFF M_p is PRIME. This test can also be extended to arbitrary INTEGERS.

A generalized version of the Lucas-Lehmer test lets

$$N + 1 = \prod_{j=1}^{n} {q_j}^{\beta_j}, \tag{2}$$

with q_j the distinct PRIME factors, and β_j their respective POWERS. If there exists a LUCAS SEQUENCE U_ν such that

$$\mathrm{GCD}(U_{(N+1)/q_j}, N) = 1 \tag{3}$$

for $j = 1, \ldots, n$ and

$$U_{N+1} \equiv 0 \pmod{N}, \tag{4}$$

then N is a PRIME. The test is particularly simple for MERSENNE NUMBERS, yielding the conventional Lucas-Lehmer test.

see also LUCAS SEQUENCE, MERSENNE NUMBER, RABIN-MILLER STRONG PSEUDOPRIME TEST

References

Sloane, N. J. A. Sequence A003010/M3494 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Lucas' Married Couples Problem

see MARRIED COUPLES PROBLEM

Lucas Number

The numbers produced by the V recurrence in the LUCAS SEQUENCE with $(P, Q) = (1, -1)$ are called Lucas numbers. They are the companions to the FIBONACCI NUMBERS F_n and satisfy the same recurrence

$$L_n = L_{n-1} + L_{n-2}, \tag{1}$$

where $L_1 = 1$, $L_2 = 3$. The first few are 1, 3, 4, 7, 11, 18, 29, 47, 76, 123, ... (Sloane's A000204).

In terms of the FIBONACCI NUMBERS,

$$L_n = F_{n-1} + F_{n+1}. \tag{2}$$

The analog of BINET'S FORMULA for Lucas numbers is

$$L_n = \left(\frac{1+\sqrt{5}}{2}\right)^n + \left(\frac{1-\sqrt{5}}{2}\right)^n. \tag{3}$$

Another formula is

$$L_n = [\phi^n], \tag{4}$$

where ϕ is the GOLDEN RATIO and $[x]$ denotes the NINT function. Given L_n,

$$L_{n+1} = \left\lfloor \frac{L_n(1+\sqrt{5})+1}{2} \right\rfloor, \tag{5}$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION,

$${L_n}^2 - L_{n-1}L_{n+1} = 5(-1)^n, \tag{6}$$

and

$$\sum_{k=0}^{n} {L_k}^2 = L_n L_{n+1} - 2. \tag{7}$$

Let p be a PRIME > 3 and k be a POSITIVE INTEGER. Then L_{2p^k} ends in a 3 (Honsberger 1985, p. 113). Analogs of the Cesàro identities for FIBONACCI NUMBERS are

$$\sum_{k=0}^{n} \binom{n}{k} L_k = L_{2n} \tag{8}$$

$$\sum_{k=0}^{n} \binom{n}{k} 2^k L_k = L_{3n}, \tag{9}$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT.

$L_n | F_m$ (L_n DIVIDES F_m) IFF n DIVIDES into m an EVEN number of times. $L_n | L_m$ IFF n divides into m an ODD number of times. $2^n L_n$ always ends in 2 (Honsberger 1985, p. 137).

Defining

$$D_n \equiv \begin{vmatrix} 3 & i & 0 & 0 & \cdots & 0 & 0 \\ i & 1 & i & 0 & \cdots & 0 & 0 \\ 0 & i & 1 & i & \cdots & 0 & 0 \\ 0 & 0 & i & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & i \\ 0 & 0 & 0 & 0 & \cdots & i & 1 \end{vmatrix} = L_{n+1} \tag{10}$$

gives

$$D_n = D_{n-1} + D_{n-2} \tag{11}$$

(Honsberger 1985, pp. 113–114).

The number of ways of picking a set (including the EMPTY SET) from the numbers 1, 2, ..., n without picking two consecutive numbers (where 1 and n are now consecutive) is L_n (Honsberger 1985, p. 122).

The only SQUARE NUMBERS in the Lucas sequence are 1 and 4, as proved by John H. E. Cohn (Alfred 1964). The only TRIANGULAR Lucas numbers are 1, 3, and 5778 (Ming 1991). The only Lucas CUBIC NUMBER is 1. The first few Lucas PRIMES L_n occur for $n = 2$, 4, 5, 7, 8, 11, 13, 16, 17, 19, 31, 37, 41, 47, 53, 61, 71, 79, 113, 313, 353, ... (Dubner and Keller 1998, Sloane's A001606).

see also FIBONACCI NUMBER

References

Alfred, Brother U. "On Square Lucas Numbers." *Fib. Quart.* **2**, 11–12, 1964.

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, pp. 94–101, 1987.

Brillhart, J.; Montgomery, P. L.; and Solverman, R. D. "Tables of Fibonacci and Lucas Factorizations." *Math. Comput.* **50**, 251–260 and S1–S15, 1988.

Brown, J. L. Jr. "Unique Representation of Integers as Sums of Distinct Lucas Numbers." *Fib. Quart.* **7**, 243–252, 1969.

Dubner, H. and Keller, W. "New Fibonacci and Lucas Primes." *Math. Comput.* 1998.

Guy, R. K. "Fibonacci Numbers of Various Shapes." §D26 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 194–195, 1994.

Hoggatt, V. E. Jr. *The Fibonacci and Lucas Numbers.* Boston, MA: Houghton Mifflin, 1969.

Honsberger, R. "A Second Look at the Fibonacci and Lucas Numbers." Ch. 8 in *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., 1985.

Leyland, P. `ftp://sable.ox.ac.uk/pub/math/factors/lucas.Z`.

Ming, L. "On Triangular Lucas Numbers." *Applications of Fibonacci Numbers, Vol. 4* (Ed. G. E. Bergum, A. N. Philippou, and A. F. Horadam). Dordrecht, Netherlands: Kluwer, pp. 231–240, 1991.

Sloane, N. J. A. Sequences A000692/M2341 and A001606/M0961 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Lucas Polynomial

The w POLYNOMIALS obtained by setting $p(x) = x$ and $q(x) = 1$ in the LUCAS POLYNOMIAL SEQUENCE. The first few are

$$\begin{aligned}F_1(x) &= x\\ F_2(x) &= x^2 + 2\\ F_3(x) &= 3x^3 + 3x\\ F_4(x) &= x^4 + 4x^2 + 2\\ F_5(x) &= x^5 + 5x^3 + 5x.\end{aligned}$$

The corresponding W POLYNOMIALS are called FIBONACCI POLYNOMIALS. The Lucas polynomials satisfy

$$L_n(1) = L_n,$$

where the L_ns are LUCAS NUMBERS.

see also FIBONACCI POLYNOMIAL, LUCAS NUMBER, LUCAS POLYNOMIAL SEQUENCE

Lucas Polynomial Sequence

A pair of generalized POLYNOMIALS which generalize the LUCAS SEQUENCE to POLYNOMIALS is given by

$$W_n^k(x) = \frac{\Delta^k(x)[a^n(x) - (-1)^k b^n(x)]}{\Delta(x)} \tag{1}$$

$$w_n^k(x) = \Delta^k(x)[a^n(x) + (-1)^k b^n(x)], \tag{2}$$

where

$$a(x) + b(x) = p(x) \tag{3}$$

$$a(x)b(x) = -q(x) \tag{4}$$

$$a(x) - b(x) = \sqrt{p^2(x) + 4q(x)} \equiv \Delta(x) \tag{5}$$

(Horadam 1996). Setting $n = 0$ gives

$$W_0^k(x) = \Delta^k(x)\frac{1 - (-1)^k}{\Delta(x)} \tag{6}$$

$$w_0^k(x) = \Delta^k(x)[1 + (-1)^k], \tag{7}$$

giving

$$W_0^0(x) = 0 \tag{8}$$

$$w_0^0(x) = 2. \tag{9}$$

When $k = 1$,

$$W_n^1(x) = w_n^0(x) = w_n(x) \tag{10}$$

$$W_n^1(x) = \Delta^2(x)W_n^0(x) = \Delta^2(x)W_n(x). \tag{11}$$

Special cases are given in the following table.

$p(x)$	$q(x)$	Polynomial 1	Polynomial 2
x	1	Fibonacci $F_n(x)$	Lucas $L_n(x)$
$2x$	1	Pell $P_n(x)$	Pell-Lucas $Q_n(x)$
1	$2x$	Jacobsthal $J_n(x)$	Jacobsthal-Lucas $j_n(x)$
$3x$	-2	Fermat $\mathcal{F}_n(x)$	Fermat-Lucas $f_n(x)$
$2x$	-1	Chebyshev $U_{n-1}(x)$	Chebyshev $2T_n(x)$

see also LUCAS SEQUENCE

References

Horadam, A. F. "Extension of a Synthesis for a Class of Polynomial Sequences." *Fib. Quart.* **34**, 68–74, 1996.

Lucas Pseudoprime

When P and Q are INTEGERS such that $D = P^2 - 4Q \neq 0$, define the LUCAS SEQUENCE $\{U_k\}$ by

$$U_k = \frac{a^k - b^k}{a - b}$$

for $k \geq 0$, with a and b the two ROOTS of $x^2 - Px + Q = 0$. Then define a Lucas pseudoprime as an ODD COMPOSITE number n such that $n \nmid Q$, the JACOBI SYMBOL $(D/n) = -1$, and $n | U_{n+1}$.

There are no EVEN Lucas pseudoprimes (Bruckman 1994). The first few Lucas pseudoprimes are 705, 2465, 2737, 3745, ... (Sloane's A005845).

see also EXTRA STRONG LUCAS PSEUDOPRIME, LUCAS SEQUENCE, PSEUDOPRIME, STRONG LUCAS PSEUDOPRIME

References

Bruckman, P. S. "Lucas Pseudoprimes are Odd." *Fib. Quart.* **32**, 155–157, 1994.

Ribenboim, P. "Lucas Pseudoprimes (lpsp(P, Q))." §2.X.B in *The New Book of Prime Number Records, 3rd ed.* New York: Springer-Verlag, p. 129, 1996.

Sloane, N. J. A. Sequence A005845/M5469 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Lucas Sequence

Let P, Q be POSITIVE INTEGERS. The ROOTS of

$$x^2 - Px + Q = 0 \tag{1}$$

are

$$a \equiv \tfrac{1}{2}(P + \sqrt{D}) \tag{2}$$

$$b \equiv \tfrac{1}{2}(P - \sqrt{D}), \tag{3}$$

where

$$D \equiv P^2 - 4Q, \tag{4}$$

so

$$a + b = P \tag{5}$$

$$ab = \tfrac{1}{4}(P^2 - D) = Q \tag{6}$$

$$a - b = \sqrt{D}. \tag{7}$$

Then define

$$U_n(P,Q) \equiv \frac{a^n - b^n}{a - b} \tag{8}$$

$$V_n(P,Q) \equiv a^n + b^n. \tag{9}$$

The first few values are therefore

$$U_0(P,Q) = 0 \tag{10}$$

$$U_1(P,Q) = 1 \tag{11}$$

$$V_0(P,Q) = 2 \tag{12}$$

$$V_1(P,Q) = P. \tag{13}$$

The sequences

$$U(P,Q) = \{U_n(P,Q) : n \geq 1\} \tag{14}$$

$$V(P,Q) = \{V_n(P,Q) : n \geq 1\} \tag{15}$$

are called Lucas sequences, where the definition is usually extended to include

$$U_{-1} = \frac{a^{-1} - b^{-1}}{a - b} = \frac{-1}{ab} = -\frac{1}{Q}. \tag{16}$$

For $(P, Q) = (1, -1)$, the U_n are the FIBONACCI NUMBERS and V_n are the LUCAS NUMBERS. For $(P, Q) = (2, -1)$, the PELL NUMBERS and Pell-Lucas numbers are obtained. $(P, Q) = (1, -2)$ produces the JACOBSTHAL NUMBERS and Pell-Jacobsthal Numbers.

The Lucas sequences satisfy the general RECURRENCE RELATIONS

$$\begin{aligned} U_{m+n} &= \frac{a^{m+n} - b^{m+n}}{a - b} \\ &= \frac{(a^m - b^m)(a^n + b^n)}{a - b} - \frac{a^n b^n (a^{m-n} - b^{m-n})}{a - b} \\ &= U_m V_n - a^n b^n U_{m-n} \end{aligned} \tag{17}$$

$$\begin{aligned} V_{m+n} &= a^{m+n} + b^{m+n} \\ &= (a^m + b^m)(a^n + b^n) - a^n b^n (a^{m-n} + b^{m-n}) \\ &= V_m V_n - a^n b^n V_{m-n}. \end{aligned} \tag{18}$$

Taking $n = 1$ then gives

$$U_m(P,Q) = PU_{m-1}(P,Q) - QU_{m-2}(P,Q) \tag{19}$$

$$V_m(P,Q) = PV_{m-1}(P,Q) - QV_{m-2}(P,Q). \tag{20}$$

Other identities include

$$U_{2n} = U_n V_n \tag{21}$$

$$U_{2n+1} = U_{n+1} V_n - Q^n \tag{22}$$

$$V_{2n} = {V_n}^2 - 2(ab)^n = {V_n}^2 - 2Q^n \tag{23}$$

$$V_{2n+1} = V_{n+1} V_n - PQ^n. \tag{24}$$

These formulas allow calculations for large n to be decomposed into a chain in which only four quantities must be kept track of at a time, and the number of steps needed is $\sim \lg n$. The chain is particularly simple if n has many 2s in its factorization.

The Us in a Lucas sequence satisfy the CONGRUENCE

$$U_{p^{n-1}[p-(D/p)]} \equiv 0 \pmod{p^n} \tag{25}$$

if

$$\mathrm{GCD}(2QcD, p) = 1, \tag{26}$$

where

$$P^2 - 4Q^2 = c^2 D. \tag{27}$$

This fact is used in the proof of the general LUCAS-LEHMER TEST.

see also FIBONACCI NUMBER, JACOBSTHAL NUMBER, LUCAS-LEHMER TEST, LUCAS NUMBER, LUCAS POLYNOMIAL SEQUENCE, PELL NUMBER, RECURRENCE SEQUENCE, SYLVESTER CYCLOTOMIC NUMBER

References

Dickson, L. E. "Recurring Series; Lucas' u_n, v_n." Ch. 17 in *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, pp. 393–411, 1952.

Ribenboim, P. *The Little Book of Big Primes.* New York: Springer-Verlag, pp. 35–53, 1991.

Lucas's Theorem

The primitive factors $Q_n(x,y)$ of x^n+y^n can be written in the form

$$Q_n(x,y)=U^2(x,y)\pm nxyV^2(x,y)$$

for SQUAREFREE n where U and V are HOMOGENEOUS POLYNOMIALS with the sign chosen according to

$$\begin{cases} + & \text{for } n=4l+1 \\ - & \text{for } n=4l+3 \\ \text{either} & \text{for } n=4l+2. \end{cases}$$

Lucky Number

Write out all the ODD numbers: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, ... The first ODD number > 1 is 3, so strike out every third number from the list: 1, 3, 7, 9, 13, 15, 19, ... The first ODD number greater than 3 in the list is 7, so strike out every seventh number: 1, 3, 7, 9, 13, 15, 21, 25, 31, ...

Numbers remaining after this procedure has been carried out completely are called lucky numbers. The first few are 1, 3, 7, 9, 13, 15, 21, 25, 31, 33, 37, ... (Sloane's A000959). Many asymptotic properties of the PRIME NUMBERS are shared by the lucky numbers. The asymptotic density is $1/\ln N$, just as the PRIME NUMBER THEOREM, and the frequency of TWIN PRIMES and twin lucky numbers are similar. A version of the GOLDBACH CONJECTURE also seems to hold.

It therefore appears that the SIEVING process accounts for many properties of the PRIMES.

see also GOLDBACH CONJECTURE, LUCKY NUMBER OF EULER, PRIME NUMBER, PRIME NUMBER THEOREM, SIEVE

References

Gardner, M. "Mathematical Games: Tests Show whether a Large Number can be Divided by a Number from 2 to 12." *Sci. Amer.* **207**, 232, Sep. 1962.

Gardner, M. "Lucky Numbers and 2187." *Math. Intell.* **19**, 26, 1997.

Guy, R. K. "Lucky Numbers." §C3 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 108–109, 1994.

Ogilvy, C. S. and Anderson, J. T. *Excursions in Number Theory.* New York: Dover, pp. 100–102, 1988.

Peterson, I. "MathTrek: Martin Gardner's Luck Number." `http://www.sciencenews.org/sn_arc97/9_6_97/mathland.htm`.

Sloane, N. J. A. Sequence A000959/M2616 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Ulam, S. M. *A Collection of Mathematical Problems.* New York: Interscience Publishers, p. 120, 1960.

Wells, D. G. *The Penguin Dictionary of Curious and Interesting Numbers.* London: Penguin, p. 32, 1986.

Lucky Number of Euler

A number p such that the PRIME-GENERATING POLYNOMIAL

$$n^2-n+p$$

is PRIME for $n=0, 1, \ldots, p-2$. Such numbers are related to the COMPLEX QUADRATIC FIELD in which the RING of INTEGERS is factorable. Specifically, the Lucky numbers of Euler (excluding the trivial case $p=3$) are those numbers p such that the QUADRATIC FIELD $\mathbb{Q}(\sqrt{1-4p})$ has CLASS NUMBER 1 (Rabinowitz 1913, Le Lionnais 1983, Conway and Guy 1996).

As established by Stark (1967), there are only nine numbers $-d$ such that $h(-d)=1$ (the HEEGNER NUMBERS -2, -3, -7, -11, -19, -43, -67, and -163), and of these, only 7, 11, 19, 43, 67, and 163 are of the required form. Therefore, the only Lucky numbers of Euler are 2, 3, 5, 11, 17, and 41 (Le Lionnais 1983, Sloane's A014556), and there does not exist a better PRIME-GENERATING POLYNOMIAL of Euler's form.

see also CLASS NUMBER, HEEGNER NUMBER, PRIME-GENERATING POLYNOMIAL

References

Conway, J. H. and Guy, R. K. "The Nine Magic Discriminants." In *The Book of Numbers.* New York: Springer-Verlag, pp. 224–226, 1996.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, pp. 88 and 144, 1983.

Rabinowitz, G. "Eindeutigkeit der Zerlegung in Primzahlfaktoren in quadratischen Zahlkörpern." *Proc. Fifth Internat. Congress Math. (Cambridge)* **1**, 418–421, 1913.

Sloane, N. J. A. Sequence A014556 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Stark, H. M. "A Complete Determination of the Complex Quadratic Fields of Class Number One." *Michigan Math. J.* **14**, 1–27, 1967.

LUCY

A nonlinear DECONVOLUTION technique used in deconvolving images from the Hubble Space Telescope before corrective optics were installed.

see also CLEAN ALGORITHM, DECONVOLUTION, MAXIMUM ENTROPY METHOD

Ludolph's Constant

see PI

Ludwig's Inversion Formula

Expresses a function in terms of its RADON TRANSFORM,

$$\begin{aligned} f(x,y)&=\mathcal{R}^{-1}(\mathcal{R}f)(x,y)\\ &=\frac{1}{\pi}\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\frac{\partial}{\partial p}(\mathcal{R}f)(p,\alpha)}{x\cos\alpha+y\sin\alpha-p}\,dp\,d\alpha. \end{aligned}$$

see also RADON TRANSFORM

Lukács Theorem

Let $\rho(x)$ be an mth degree POLYNOMIAL which is NONNEGATIVE in $[-1,1]$. Then $\rho(x)$ can be represented in the form

$$\begin{cases} [A(x)]^2+(1-x^2)[B(x)]^2 & \text{for } m \text{ even} \\ (1+x)[C(x)]^2+(1-x)[D(x)]^2 & \text{for } m \text{ odd,} \end{cases}$$

where $A(x)$, $B(x)$, $C(x)$, and $D(x)$ are REAL POLYNOMIALS whose degrees do not exceed m.

References

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., p. 4, 1975.

Lune (Plane)

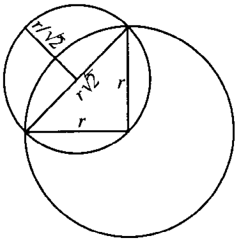

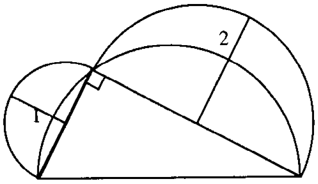

A figure bounded by two circular ARCS of unequal RADII. Hippocrates of Chios SQUARED the above left lune, as well as two others, in the fifth century BC. Two more SQUARABLE lunes were found by T. Clausen in the 19th century (Dunham 1990 attributes these discoveries to Euler in 1771). In the 20th century, N. G. Tschebatorew and A. W. Dorodnow proved that these are the only five squarable lunes (Shenitzer and Steprans 1994). The left lune above is squared as follows,

$$\begin{aligned} A_{\text{half small circle}} &= \tfrac{1}{2}\pi\left(\frac{r}{\sqrt{2}}\right)^2 = \tfrac{1}{4}\pi r^2 \\ A_{\text{lens}} &= A_{\text{quarter big circle}} - A_{\text{triangle}} \\ &= \tfrac{1}{4}\pi r^2 - \tfrac{1}{2}r^2 \\ A_{\text{lune}} &= A_{\text{half small circle}} - A_{\text{lens}} = \tfrac{1}{2}r^2 \\ &= A_{\text{triangle}}, \end{aligned}$$

so the lune and TRIANGLE have the same AREA. In the right figure, $A_1 + A_2 = A_\Delta$.

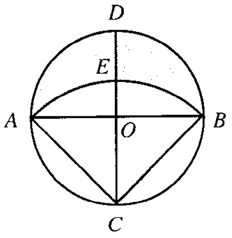

For the above lune,

$$A_{\text{lune}} = 2A_{\Delta OBC}.$$

see also ANNULUS, ARC, CIRCLE, LUNE (SURFACE)

References

Dunham, W. "Hippocrates' Quadrature of the Lune." Ch. 1 in *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 1–20, 1990.

Heath, T. L. *A History of Greek Mathematics.* New York: Dover, p. 185, 1981.

Pappas, T. "Lunes." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 72–73, 1989.

Shenitzer, A. and Steprans, J. "The Evolution of Integration." *Amer. Math. Monthly* **101**, 66–72, 1994.

Lune (Solid)

A geometric figure consisting of two TRIANGLES attached to opposite sides of a SQUARE.

see also SQUARE, TRIANGLE

Lune (Surface)

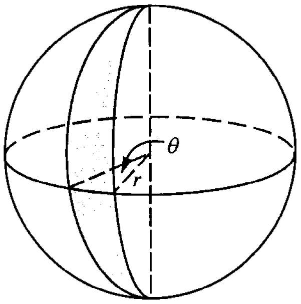

A sliver of the surface of a SPHERE of RADIUS r cut out by two planes through the azimuthal axis with DIHEDRAL ANGLE θ. The SURFACE AREA of the lune is

$$S = 2r^2\theta,$$

which is just the area of the SPHERE times $\theta/(2\pi)$.

see also LUNE (PLANE), SPHERE

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 130, 1987.

Lunule

see LUNE (PLANE)

Lüroth's Theorem

If x and y are nonconstant rational functions of a parameter, the curve so defined has GENUS 0. Furthermore, x and y may be expressed rationally in terms of a parameter which is rational in them.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 246, 1959.

Lusin's Theorem

Let $f(x)$ be a finite and MEASURABLE FUNCTION in $(-\infty, \infty)$, and let ϵ be freely chosen. Then there is a function $g(x)$ such that

1. $g(x)$ is continuous in $(-\infty, \infty)$,
2. The MEASURE of $\{x : f(x) \neq g(x)\}$ is $< \epsilon$,
3. $M(|g|; R_1) \leq M(|f|; R_1)$,

where $M(f; S)$ denotes the upper bound of the aggregate of the values of $f(P)$ as P runs through all values of S.

References

Kestelman, H. §4.4 in *Modern Theories of Integration, 2nd rev. ed.* New York: Dover, pp. 30 and 109–112, 1960.

LUX Method

A method for constructing MAGIC SQUARES of SINGLY EVEN order $n \geq 6$.

see also MAGIC SQUARE

Lyapunov Characteristic Exponent

The Lyapunov characteristic exponent [LCE] gives the rate of exponential divergence from perturbed initial conditions. To examine the behavior of an orbit around a point $\mathbf{X}^*(t)$, perturb the system and write

$$\mathbf{X}(t) = \mathbf{X}^*(t) + U(t), \tag{1}$$

where $U(t)$ is the average deviation from the unperturbed trajectory at time t. In a CHAOTIC region, the LCE σ is independent of $\mathbf{X}^*(0)$. It is given by the OSEDELEC THEOREM, which states that

$$\sigma_i = \lim_{t \to \infty} \ln |\mathbf{U}(t)|. \tag{2}$$

For an n-dimensional mapping, the Lyapunov characteristic exponents are given by

$$\sigma_i = \lim_{N \to \infty} \ln |\lambda_i(N)| \tag{3}$$

for $i = 1, \ldots, n$, where λ_i is the LYAPUNOV CHARACTERISTIC NUMBER.

One Lyapunov characteristic exponent is always 0, since there is never any divergence for a perturbed trajectory in the direction of the unperturbed trajectory. The larger the LCE, the greater the rate of exponential divergence and the wider the corresponding SEPARATRIX of the CHAOTIC region. For the STANDARD MAP, an analytic estimate of the width of the CHAOTIC zone by Chirikov (1979) finds

$$\delta I = Be^{-AK^{-1/2}}. \tag{4}$$

Since the Lyapunov characteristic exponent increases with increasing K, some relationship likely exists connecting the two. Let a trajectory (expressed as a MAP) have initial conditions (x_0, y_0) and a nearby trajectory have initial conditions $(x', y') = (x_0 + dx, y_0 + dy)$. The distance between trajectories at iteration k is then

$$dk = \|(x' - x_0, y' - y_0)\|, \tag{5}$$

and the mean exponential rate of divergence of the trajectories is defined by

$$\sigma_1 = \lim_{k \to \infty} \frac{1}{k} \ln \left(\frac{d_k}{d_0}\right). \tag{6}$$

For an n-dimensional phase space (MAP), there are n Lyapunov characteristic exponents $\sigma_1 \geq \sigma_2 \geq \ldots > \sigma_n$. However, because the largest exponent σ_1 will dominate, this limit is practically useful only for finding the largest exponent. Numerically, since d_k increases exponentially with k, after a few steps the perturbed trajectory is no longer nearby. It is therefore necessary to renormalize frequently every t steps. Defining

$$r_{k\tau} \equiv \frac{d_{k\tau}}{d_0}, \tag{7}$$

one can then compute

$$\sigma_1 = \lim_{n \to \infty} \frac{1}{n\tau} \sum_{k=1}^{n} \ln r_{k\tau}. \tag{8}$$

Numerical computation of the second (smaller) Lyapunov exponent may be carried by considering the evolution of a 2-D surface. It will behave as

$$e^{(\sigma_1 + \sigma_2)t}, \tag{9}$$

so σ_2 can be extracted if σ_1 is known. The process may be repeated to find smaller exponents.

For HAMILTONIAN SYSTEMS, the LCEs exist in additive inverse pairs, so if σ is an LCE, then so is $-\sigma$. One LCE is always 0. For a 1-D oscillator (with a 2-D phase space), the two LCEs therefore must be $\sigma_1 = \sigma_2 = 0$, so the motion is QUASIPERIODIC and cannot be CHAOTIC. For higher order HAMILTONIAN SYSTEMS, there are always at least two 0 LCEs, but other LCEs may enter in plus-and-minus pairs l and $-l$. If they, too, are both zero, the motion is integrable and not CHAOTIC. If they are NONZERO, the POSITIVE LCE l results in an exponential separation of trajectories, which corresponds to a CHAOTIC region. Notice that it is not possible to have all LCEs NEGATIVE, which explains why convergence of orbits is never observed in HAMILTONIAN SYSTEMS.

Now consider a dissipative system. For an arbitrary n-D phase space, there must always be one LCE equal to 0, since a perturbation along the path results in no divergence. The LCEs satisfy $\sum_i \sigma_i < 0$. Therefore, for a 2-D phase space of a dissipative system, $\sigma_1 = 0, \sigma_2 < 0$. For a 3-D phase space, there are three possibilities:

1. (Integrable): $\sigma_1 = 0, \sigma_2 = 0, \sigma_3 < 0$,

2. (Integrable): $\sigma_1 = 0, \sigma_2, \sigma_3 < 0$,
3. (Chaotic): $\sigma_1 = 0, \sigma_2 > 0, \sigma_3 < -\sigma_2 < 0$.

see also Chaos, Hamiltonian System, Lyapunov Characteristic Number, Osedelec Theorem

References

Chirikov, B. V. "A Universal Instability of Many-Dimensional Oscillator Systems." *Phys. Rep.* **52**, 264–379, 1979.

Lyapunov Characteristic Number

Given a Lyapunov Characteristic Exponent σ_i, the corresponding Lyapunov characteristic number λ_i is defined as

$$\lambda_i \equiv e^{\sigma_i}. \qquad (1)$$

For an n-dimensional linear Map,

$$\mathbf{X}_{n+1} = \mathsf{M}\mathbf{X}_n. \qquad (2)$$

The Lyapunov characteristic numbers $\lambda_1, \ldots, \lambda_n$ are the Eigenvalues of the Map Matrix. For an arbitrary Map

$$x_{n+1} = f_1(x_n, y_n) \qquad (3)$$

$$y_{n+1} = f_2(x_n, y_n), \qquad (4)$$

the Lyapunov numbers are the Eigenvalues of the limit

$$\lim_{n\to\infty} [J(x_n, y_n) J(x_{n-1}, y_{n-1}) \cdots J(x_1, y_1)]^{1/n}, \qquad (5)$$

where $J(x,y)$ is the Jacobian

$$J(x,y) \equiv \begin{vmatrix} \frac{\partial f_1(x,y)}{\partial x} & \frac{\partial f_1(x,y)}{\partial y} \\ \frac{\partial f_2(x,y)}{\partial x} & \frac{\partial f_2(x,y)}{\partial y} \end{vmatrix}. \qquad (6)$$

If $\lambda_i = 0$ for all i, the system is not Chaotic. If $\lambda \neq 0$ and the Map is Area-Preserving (Hamiltonian), the product of Eigenvalues is 1.

see also Adiabatic Invariant, Chaos, Lyapunov Characteristic Exponent

Lyapunov Condition

If the third Moment exists for a Distribution of x_i and the Lebesgue Integral is given by

$$r_n{}^3 = \sum_{i=1}^{n} \int_{-\infty}^{\infty} |x|^3 \, dF_i(x),$$

then if

$$\lim_{n\to\infty} \frac{r_n}{s_n} = 0,$$

the Central Limit Theorem holds.

see also Central Limit Theorem

Lyapunov Dimension

For a 2-D Map with $\sigma_2 > \sigma_1$,

$$d_{\text{Lya}} = 1 - \frac{\sigma_1}{\sigma_2},$$

where σ_n are the Lyapunov Characteristic Exponents.

see also Capacity Dimension, Kaplan-Yorke Conjecture

References

Frederickson, P.; Kaplan, J. L.; Yorke, E. D.; and Yorke, J. A. "The Liapunov Dimension of Strange Attractors." *J. Diff. Eq.* **49**, 185–207, 1983.

Nayfeh, A. H. and Balachandran, B. *Applied Nonlinear Dynamics: Analytical, Computational, and Experimental Methods.* New York: Wiley, p. 549, 1995.

Lyapunov's First Theorem

A Necessary and Sufficient condition for all the Eigenvalues of a Real $n \times n$ matrix A to have Negative Real Parts is that the equation

$$\mathsf{A}^{\mathrm{T}}\mathsf{V} + \mathsf{V}\mathsf{A} = -\mathsf{I}$$

has as a solution where V is an $n \times n$ matrix and $(\mathbf{x}, \mathsf{V}\mathbf{x})$ is a positive definite quadratic form.

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1122, 1979.

Lyapunov Function

A function which is continuous, nonnegative, and has continuous Partial Derivatives. The existence of a Lyapunov function guarantees the Nonlinear Stability of a Fixed Point.

References

Jordan, D. W. and Smith, P. *Nonlinear Ordinary Differential Equations.* Oxford, England: Clarendon Press, p. 283, 1977.

Lyapunov's Second Theorem

If all the Eigenvalues of a Real Matrix A have Real Parts, then to an arbitrary negative definite quadratic form $(\mathbf{x}, \mathsf{W}\mathbf{x})$ with $\mathbf{x} = \mathbf{x}(t)$ there corresponds a positive definite quadratic form $(\mathbf{x}, \mathsf{V}\mathbf{x})$ such that if one takes

$$\frac{d\mathbf{x}}{dt} = \mathsf{A}\mathbf{x},$$

then $(\mathbf{x}, \mathsf{W}\mathbf{x})$ and $(\mathbf{x}, \mathsf{W}\mathbf{x})$ satisfy

$$\frac{d}{dt}(\mathbf{x}, \mathsf{V}\mathbf{x}) = (\mathbf{x}, \mathsf{W}\mathbf{x}).$$

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1122, 1979.

Lyndon Word

A Lyndon word is an aperiodic notation for representing a NECKLACE.

see also DE BRUIJN SEQUENCE, NECKLACE

References

Ruskey, F. "Information on Necklaces, Lyndon Words, de Bruijn Sequences." `http://sue.csc.uvic.ca/~cos/inf/neck/NecklaceInfo.html`.

Sloane, N. J. A. Sequence A001037/M0116 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Lyons Group

The SPORADIC GROUP *Ly*.

see also SPORADIC GROUP

References

Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas/Ly.html`.

M

M-Estimate

A ROBUST ESTIMATION based on maximum likelihood argument.

see also L-ESTIMATE, R-ESTIMATE

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Robust Estimation." §15.7 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 694–700, 1992.

Mac Lane's Theorem

A theorem which treats constructions of FIELDS of CHARACTERISTIC p.

see also CHARACTERISTIC (FIELD), FIELD

Machin's Formula

$$\tfrac{1}{4}\pi = 4\tan^{-1}(\tfrac{1}{5}) - \tan^{-1}(\tfrac{1}{239}).$$

There are a whole class of MACHIN-LIKE FORMULAS with various numbers of terms (although only four such formulas with only two terms). The properties of these formulas are intimately connected with COTANGENT identities.

see also 196-ALGORITHM, GREGORY NUMBER, MACHIN-LIKE FORMULAS, PI

Machin-Like Formulas

Machin-like formulas have the form

$$m\cot^{-1}u + n\cot^{-1}v = \tfrac{1}{4}k\pi, \qquad (1)$$

where u, v, and k are POSITIVE INTEGERS and m and n are NONNEGATIVE INTEGERS. Some such FORMULAS can be found by converting the INVERSE TANGENT decompositions for which $c_n \neq 0$ in the table of Todd (1949) to INVERSE COTANGENTS. However, this gives only Machin-like formulas in which the smallest term is ± 1.

Maclaurin-like formulas can be derived by writing

$$\cot^{-1}z = \frac{1}{2i}\ln\left(\frac{z+i}{z-i}\right) \qquad (2)$$

and looking for a_k and u_k such that

$$\sum_k a_k \cot^{-1}u_k = \tfrac{1}{4}\pi, \qquad (3)$$

so

$$\prod_k \left(\frac{u_k+i}{u_k-i}\right)^{a_k} = e^{2\pi i/4} = i. \qquad (4)$$

Machin-like formulas exist IFF (4) has a solution in INTEGERS. This is equivalent to finding INTEGER values such that

$$(1-i)^k(u+i)^m(v+i)^n \qquad (5)$$

is REAL (Borwein and Borwein 1987, p. 345). An equivalent formulation is to find all integral solutions to one of

$$1 + x^2 = 2y^n \qquad (6)$$

$$1 + x^2 = y^n \qquad (7)$$

for $n = 3, 5, \ldots$.

There are only four such FORMULAS,

$$\tfrac{1}{4}\pi = 4\tan^{-1}(\tfrac{1}{5}) - \tan^{-1}(\tfrac{1}{239}) \qquad (8)$$

$$\tfrac{1}{4}\pi = \tan^{-1}(\tfrac{1}{2}) + \tan^{-1}(\tfrac{1}{3}) \qquad (9)$$

$$\tfrac{1}{4}\pi = 2\tan^{-1}(\tfrac{1}{2}) - \tan^{-1}(\tfrac{1}{7}) \qquad (10)$$

$$\tfrac{1}{4}\pi = 2\tan^{-1}(\tfrac{1}{3}) + \tan^{-1}(\tfrac{1}{7}), \qquad (11)$$

known as MACHIN'S FORMULA, EULER'S MACHIN-LIKE FORMULA, HERMANN'S FORMULA, and HUTTON'S FORMULA. These follow from the identities

$$\left(\frac{5+i}{5-i}\right)^4\left(\frac{239+i}{239-i}\right)^{-1} = i \qquad (12)$$

$$\left(\frac{2+i}{2-i}\right)\left(\frac{3+i}{3-i}\right) = i \qquad (13)$$

$$\left(\frac{2+i}{2-i}\right)^2\left(\frac{7+i}{7-i}\right)^{-1} = i \qquad (14)$$

$$\left(\frac{3+i}{3-i}\right)^2\left(\frac{7+i}{7-i}\right) = i. \qquad (15)$$

Machin-like formulas with two terms can also be generated which do not have integral arc cotangent arguments such as Euler's

$$\tfrac{1}{4}\pi = 5\tan^{-1}(\tfrac{1}{7}) + 2\tan^{-1}(\tfrac{3}{79}) \qquad (16)$$

(Wetherfield 1996), and which involve inverse SQUARE ROOTS, such as

$$\frac{\pi}{2} = 2\tan^{-1}\left(\frac{1}{\sqrt{2}}\right) + \tan^{-1}\left(\frac{1}{\sqrt{8}}\right). \qquad (17)$$

Three-term Machin-like formulas include GAUSS'S MACHIN-LIKE FORMULA

$$\tfrac{1}{4}\pi = 12\cot^{-1}18 + 8\cot^{-1}57 - 5\cot^{-1}239, \qquad (18)$$

STRASSNITZKY'S FORMULA

$$\tfrac{1}{4}\pi = \cot^{-1}2 + \cot^{-1}5 + \cot^{-1}8, \qquad (19)$$

and the following,

$$\tfrac{1}{4}\pi = 6\cot^{-1}8 + 2\cot^{-1}57 + \cot^{-1}239 \qquad (20)$$

$$\tfrac{1}{4}\pi = 4\cot^{-1}5 - 1\cot^{-1}70 + \cot^{-1}99 \qquad (21)$$

$$\tfrac{1}{4}\pi = 1\cot^{-1}2 + 1\cot^{-1}5 + \cot^{-1}8 \qquad (22)$$

$$\tfrac{1}{4}\pi = 8\cot^{-1}10 - 1\cot^{-1}239 - 4\cot^{-1}515 \qquad (23)$$

$$\tfrac{1}{4}\pi = 5\cot^{-1}7 + 4\cot^{-1}53 + 2\cot^{-1}4443. \qquad (24)$$

The first is due to Størmer, the second due to Rutherford, and the third due to Dase.

Using trigonometric identities such as

$$\cot^{-1}x = 2\cot^{-1}(2x) - \cot^{-1}(4x^3 + 3x), \qquad (25)$$

it is possible to generate an infinite sequence of Machin-like formulas. Systematic searches therefore most often concentrate on formulas with particularly "nice" properties (such as "efficiency").

The efficiency of a FORMULA is the time it takes to calculate π with the POWER series for arctangent

$$\pi = a_1 \cot(b_1) + a_2 \cot(b_2) + \ldots, \qquad (26)$$

and can be roughly characterized using Lehmer's "measure" formula

$$e \equiv \sum \frac{1}{\log_{10} b_i}. \qquad (27)$$

The number of terms required to achieve a given precision is roughly proportional to e, so lower e-values correspond to better sums. The best currently known efficiency is 1.51244, which is achieved by the 6-term series

$$\begin{aligned}\tfrac{1}{4}\pi = 183\cot^{-1}239 + 32\cot^{-1}1023 - 68\cot^{-1}5832 \\ +12\cot^{-1}110443 - 12\cot^{-1}4841182 \\ -100\cot^{-1}6826318 \qquad (28)\end{aligned}$$

discovered by C.-L. Hwang (1997). Hwang (1997) also discovered the remarkable identities

$$\begin{aligned}\tfrac{1}{4}\pi = {} & P\cot^{-1}2 - M\cot^{-1}3 + L\cot^{-1}5 + K\cot^{-1}7 \\ & +(N + K + L - 2M + 3P - 5)\cot^{-1}8 \\ & +(2N + M - P + 2 - L)\cot^{-1}18 \\ & -(2P - 3 - M + L + K - N)\cot^{-1}57 - N\cot^{-1}239, \\ & \qquad (29)\end{aligned}$$

where K, L, M, N, and P are POSITIVE INTEGERS, and

$$\tfrac{1}{4}\pi = (N+2)\cot^{-1}2 - N\cot^{-1}3 - (N+1)\cot^{-1}N. \quad (30)$$

The following table gives the number $N(n)$ of Machin-like formulas of n terms in the compilation by Wetherfield and Hwang. Except for previously known identities (which are included), the criteria for inclusion are the following:

1. first term < 8 digits: measure < 1.8.
2. first term = 8 digits: measure < 1.9.
3. first term = 9 digits: measure < 2.0.
4. first term =10 digits: measure < 2.0.

n	$N(n)$	min e
1	1	0
2	4	1.85113
3	106	1.78661
4	39	1.58604
5	90	1.63485
6	120	1.51244
7	113	1.54408
8	18	1.65089
9	4	1.72801
10	78	1.63086
11	34	1.6305
12	188	1.67458
13	37	1.71934
14	5	1.75161
15	24	1.77957
16	51	1.81522
17	5	1.90938
18	570	1.87698
19	1	1.94899
20	11	1.95716
21	1	1.98938
Total	1500	1.51244

see also EULER'S MACHIN-LIKE FORMULA, GAUSS'S MACHIN-LIKE FORMULA, GREGORY NUMBER, HERMANN'S FORMULA, HUTTON'S FORMULA, INVERSE COTANGENT, MACHIN'S FORMULA, PI, STØRMER NUMBER, STRASSNITZKY'S FORMULA

References

Arndt, J. "Arctan Formulas." `http://jjj.spektracom.de/jjf.dvi`.

Arndt, J. "Big ArcTan Formula Bucket." `http://jjj.spektracom.de/fox.dvi`.

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 347–359, 1987.

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, 1987.

Castellanos, D. "The Ubiquitous Pi. Part I." *Math. Mag.* **61**, 67–98, 1988.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 241–248, 1996.

Hwang, C.-L. "More Machin-Type Identities." *Math. Gaz.*, 120–121, March 1997.

Lehmer, D. H. "On Arccotangent Relations for π." *Amer. Math. Monthly* **45**, 657–664, 1938.

Lewin, L. *Polylogarithms and Associated Functions.* New York: North-Holland, 1981.

Lewin, L. *Structural Properties of Polylogarithms.* Providence, RI: Amer. Math. Soc., 1991.

Nielsen, N. *Der Euler'sche Dilogarithms.* Leipzig, Germany: Halle, 1909.

Størmer, C. "Sur l'Application de la Théorie des Nombres Entiers Complexes à la Solution en Nombres Rationels x_1,

$x_2, \ldots, c_1, c_2, \ldots, k$ de l'Equation...." *Archiv for Mathematik og Naturvidenskab* **B 19**, 75–85, 1896.
Todd, J. "A Problem on Arc Tangent Relations." *Amer. Math. Monthly* **56**, 517–528, 1949.
Weisstein, E. W. "Machin-Like Formulas." `http://www.astro.virginia.edu/~eww6n/math/notebooks/MachinFormulas.m`.
Wetherfield, M. "The Enhancement of Machin's Formula by Todd's Process." *Math. Gaz.* **80**, 333–344, 1996.
Wetherfield, M. "Machin Revisited." *Math. Gaz.*, 121–123, March 1997.
Williams, R. "Arctangent Formulas for Pi." `http://www.cacr.caltech.edu/~roy/pi.formulas.html`. [Contains errors].

Machine

A method for producing infinite LOOP SPACES and spectra.

see also GADGET, LOOP SPACE, MAY-THOMASON UNIQUENESS THEOREM, TURING MACHINE

Mackey's Theorem

Let E and F be paired spaces with S a family of absolutely convex bounded sets of F such that the sets of S generate F and, if $B_1, B_2 \in S$, there exists a $B_3 \in S$ such that $B_3 \supset B_1$ and $B_3 \supset B_2$. Then the dual space of E_S is equal to the union of the weak completions of λB, where $\lambda > 0$ and $B \in S$.

see also GROTHENDIECK'S THEOREM

References
Iyanaga, S. and Kawada, Y. (Eds.). "Mackey's Theorem." §407M in *Encyclopedic Dictionary of Mathematics*. Cambridge, MA: MIT Press, p. 1274, 1980.

Maclaurin-Bezout Theorem

The Maclaurin-Bèzout theorem says that two curves of degree n intersect in n^2 points, so two CUBICS intersect in nine points. This means that $n(n+3)/2$ points do not always uniquely determine a single curve of order n.

see also CRAMÉR-EULER PARADOX

Maclaurin-Cauchy Theorem

If $f(x)$ is POSITIVE and decreases to 0, then an EULER CONSTANT

$$\gamma_f \equiv \lim_{n\to\infty} \left[\sum_{k=1}^{n} f(k) - \int_a^n f(x)\,dx\right]$$

can be defined. If $f(x) = 1/x$, then

$$\gamma = \lim_{n\to\infty} \left(\sum_{k=1}^{n} \frac{1}{k} - \int_1^n \frac{dx}{x}\right) = \lim_{n\to\infty} \left(\sum_{k=1}^{n} \frac{1}{k} - \ln n\right),$$

where γ is the EULER-MASCHERONI CONSTANT.

Maclaurin Integral Test

see INTEGRAL TEST

Maclaurin Polynomial

see MACLAURIN SERIES

Maclaurin Series

A series expansion of a function about 0,

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \ldots + \frac{f^{(n)}(0)}{n!}x^n + \ldots, \tag{1}$$

named after the Scottish mathematician Maclaurin. Maclaurin series for common functions include

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + x^5 + \ldots \qquad -1 < x < 1 \tag{2}$$

$$\mathrm{cn}(x, k^2) = 1 - \tfrac{1}{2!}x^2 + \tfrac{1}{4!}(1 + 4k^2)x^4 + \ldots \tag{3}$$

$$\cos x = 1 - \tfrac{1}{2!}x^2 + \tfrac{1}{4!}x^4 - \tfrac{1}{6!}x^6 - \ldots \qquad -\infty < x < \infty \tag{4}$$

$$\cos^{-1} x = \tfrac{1}{2}\pi - x - \tfrac{1}{6}x^3 - \tfrac{3}{40}x^5 - \tfrac{5}{112}x^7 - \ldots \qquad -1 < x < 1 \tag{5}$$

$$\cosh x = 1 + \tfrac{1}{2}x^2 + \tfrac{1}{24}x^4 + \tfrac{1}{720}x^6 + \tfrac{1}{40{,}320}x^8 + \ldots \tag{6}$$

$$\cosh^{-1}(1+x) = \sqrt{2x}\,(1 - \tfrac{1}{2}x + \tfrac{3}{160}x^2 - \tfrac{5}{896}x^3 + \ldots) \tag{7}$$

$$\cot x = x^{-1} - \tfrac{1}{3}x - \tfrac{1}{45}x^3 - \tfrac{2}{945}x^5 - \tfrac{1}{4725}x^7 - \ldots \tag{8}$$

$$\cot^{-1} x = \tfrac{1}{2}\pi - x + \tfrac{1}{3}x^3 - \tfrac{1}{5}x^5 + \tfrac{1}{7}x^7 - \tfrac{1}{9}x^9 + \ldots \tag{9}$$

$$= x^{-1} - \tfrac{1}{3}x^{-3} + \tfrac{1}{5}x^{-5} - \tfrac{1}{7}x^{-7} + \tfrac{1}{9}x^{-9} + \ldots \tag{10}$$

$$\coth x = x^{-1} + \tfrac{1}{3}x - \tfrac{1}{45}x^4 + \tfrac{2}{945}x^5 - \tfrac{1}{4725}x^7 + \ldots \tag{11}$$

$$\coth^{-1}(1+x) = \tfrac{1}{2}\ln 2 - \tfrac{1}{2}\ln x + \tfrac{1}{4}x - \tfrac{1}{16}x^2 + \ldots \tag{12}$$

$$\csc x = x^{-1} + \tfrac{1}{6}x + \tfrac{7}{360}x^3 + \tfrac{31}{15120}x^5 + \ldots \tag{13}$$

$$\mathrm{csch}\, x = x^{-1} - \tfrac{1}{6}x + \tfrac{7}{360}x^3 + \tfrac{31}{15120}x^5 + \ldots \tag{14}$$

$$\mathrm{csch}^{-1}\, x = \ln 2 - \ln x + \tfrac{1}{4}x^2 - \tfrac{3}{32}x^4 + \tfrac{5}{96}x^6 - \ldots \tag{15}$$

$$\mathrm{dn}(x, k^2)x = 1 - \tfrac{1}{2!}k^2x^2 + \tfrac{1}{4!}k^2(4 + k^2)x^4 + \ldots \tag{16}$$

$$\mathrm{erf}\, x = \frac{1}{\sqrt{\pi}}\left(2x - \tfrac{2}{3}x^3 + \tfrac{1}{5}x^5 - \tfrac{1}{21}x^7 + \ldots\right) \tag{17}$$

$$e^x = 1 + x + \tfrac{1}{2!}x^2 + \tfrac{1}{3!}x^3 + \tfrac{1}{4!}x^4 + \ldots \qquad -\infty < x < \infty \tag{18}$$

$${}_2F_1(\alpha, \beta, \gamma; x) = 1 + \frac{\alpha\beta}{1!\gamma}x + \frac{\alpha(\alpha+1)\beta(\beta+1)}{2!\gamma(\gamma+1)}x^2 + \ldots \tag{19}$$

$$\ln(1+x) = x - \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 - \tfrac{1}{4}x^4 + \ldots \qquad -1 < x < 1 \tag{20}$$

$$\ln\left(\frac{1+x}{1-x}\right) = 2x + \tfrac{2}{3}x^3 + \tfrac{2}{5}x^5 + \tfrac{2}{7}x^7 + \ldots \qquad -1 < x < 1 \tag{21}$$

$$\sec x = 1 + \tfrac{1}{2}x^2 + \tfrac{5}{24}x^4 + \tfrac{61}{720}x^6 + \tfrac{277}{8064}x^8 + \ldots \tag{22}$$

$$\mathrm{sech}\, x = 1 - \tfrac{1}{2}x^2 + \tfrac{5}{24}x^4 - \tfrac{61}{720}x^6 + \tfrac{277}{8064}x^8 + \ldots \tag{23}$$

$$\mathrm{sech}^{-1}\, x = \ln 2 - \ln x - \tfrac{1}{4}x^2 - \tfrac{3}{32}x^4 - \ldots \tag{24}$$

$$\sin x = x - \tfrac{1}{3!}x^3 + \tfrac{1}{5!}x^5 - \tfrac{1}{7!}x^7 + \ldots$$

$$-\infty < x < \infty \qquad (25)$$

$$\sin^{-1} x = x + \tfrac{1}{6}x^3 + \tfrac{3}{40}x^5 + \tfrac{5}{112}x^7 + \tfrac{35}{1152}x^9 + \ldots \qquad (26)$$

$$\sinh x = x + \tfrac{1}{6}x^3 + \tfrac{1}{120}x^5 + \tfrac{1}{5040}x^7 + \tfrac{1}{362{,}880}x^9 + \ldots \qquad (27)$$

$$\sinh^{-1} x = x - \tfrac{1}{6}x^3 + \tfrac{3}{40}x^5 - \tfrac{5}{112}x^7 + \tfrac{35}{1152}x^9 - \ldots \qquad (28)$$

$$\operatorname{sn}(x, k^2) = \tfrac{1}{3!}(1+k^2)x^3 + \tfrac{1}{5!}(1+14k^2+k^4)x^5 + \ldots \qquad (29)$$

$$\tan x = x + \tfrac{1}{3}x^3 + \tfrac{2}{15}x^5 + \tfrac{17}{315}x^7 + \tfrac{62}{2835}x^9 + \ldots \qquad (30)$$

$$\tan^{-1} x = x - \tfrac{1}{3}x^3 + \tfrac{1}{5}x^5 - \tfrac{1}{7}x^7 + \ldots$$

$$-1 < x < 1 \qquad (31)$$

$$\tan^{-1}(1+x) = \tfrac{1}{4}\pi + \tfrac{1}{2}x - \tfrac{1}{4}x^2 + \tfrac{1}{12}x^3 + \tfrac{1}{40}x^5 + \ldots \qquad (32)$$

$$\tanh x = x - \tfrac{1}{3}x^3 + \tfrac{2}{15}x^5 - \tfrac{17}{315}x^7 + \tfrac{62}{2835}x^9 + \ldots \qquad (33)$$

$$\tanh^{-1} x = x + \tfrac{1}{3}x^3 + \tfrac{1}{5}x^5 + \tfrac{1}{7}x^7 + \tfrac{1}{9}x^9 + \ldots. \qquad (34)$$

The explicit forms for some of these are

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n \qquad (35)$$

$$\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} \qquad (36)$$

$$\csc x = \sum_{n=0}^{\infty} \frac{(-1)^{n+1} 2(2^{2n-1}-1)B_{2n}}{(2n)!} x^{2n-1} \qquad (37)$$

$$e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n \qquad (38)$$

$$\ln(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n \qquad (39)$$

$$\ln\left(\frac{1+x}{1-x}\right) = \sum_{n=1}^{\infty} \frac{2}{(2n-1)} x^{2n-1} \qquad (40)$$

$$\sec x = \sum_{n=0}^{\infty} \frac{(-1)^n E_{2n}}{(2n)!} x^{2n} \qquad (41)$$

$$\sin x = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{(2n-1)!} x^{2n-1} \qquad (42)$$

$$\tan x = \sum_{n=1}^{\infty} \frac{(-1)^{n+1} 2^{2n}(2^{2n}-1)B_{2n}}{(2n)!} x^{2n-1} \qquad (43)$$

$$\tan^{-1} x = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{2n-1} x^{2n-1} \qquad (44)$$

$$\tanh^{-1} x = \sum_{n=1}^{\infty} \frac{1}{2n-1} x^{2n-1}, \qquad (45)$$

where B_n are BERNOULLI NUMBERS and E_n are EULER NUMBERS.

see also ALCUIN'S SEQUENCE, LAGRANGE EXPANSION, LEGENDRE SERIES, TAYLOR SERIES

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 299-300, 1987.

Maclaurin Trisectrix

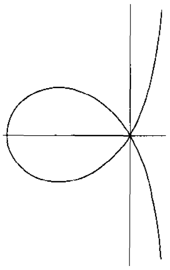

A curve first studied by Colin Maclaurin in 1742. It was studied to provide a solution to one of the GEOMETRIC PROBLEMS OF ANTIQUITY, in particular TRISECTION of an ANGLE, whence the name trisectrix. The Maclaurin trisectrix is an ANALLAGMATIC CURVE, and the origin is a CRUNODE.

The Maclaurin trisectrix has CARTESIAN equation

$$y^2 = \frac{x^2(x+3a)}{a-x}, \qquad (1)$$

or the parametric equations

$$x = a\frac{t^2-3}{t^2+1} \qquad (2)$$

$$y = a\frac{t(t^2-3)}{t^2+1}. \qquad (3)$$

The ASYMPTOTE has equation $x = a$, and the center of the loop is as $(2a, 0)$. Draw a line L with ANGLE 3α through the loop center. Then the angle made by the line through the center and point of intersection of L with the trisectrix is α. The trisectrix is sometimes defined instead as

$$x(x^2+y^2) = a(y^2-3x^2) \qquad (4)$$

$$y^2 = \frac{x^2(3a+x)}{a-x} \qquad (5)$$

$$r = \frac{2a\sin(3\theta)}{\sin(2\theta)}. \qquad (6)$$

Another form of the equation is the POLAR EQUATION

$$r = a\sec(\tfrac{1}{3}\theta), \qquad (7)$$

where the origin is inside the loop and the crossing point is on the NEGATIVE x-AXIS.

The tangents to the curve at the origin make angles of ±60° with the x-AXIS. The AREA of the loop is

$$A_{\text{loop}} = 3\sqrt{3}\,a^2, \qquad (8)$$

and the NEGATIVE x-intercept is $(-3a, 0)$ (MacTutor Archive).

The Maclaurin trisectrix is the PEDAL CURVE of the PARABOLA where the PEDAL POINT is taken as the reflection of the FOCUS in the DIRECTRIX.

see also CATALAN'S TRISECTRIX, STROPHOID

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 103–106, 1972.

Lee, X. "Trisectrix." `http://www.best.com/~xah/Special PlaneCurves_dir/Trisectrix_dir/trisectrix.html`.

Lee, X. "Trisectrix of Maclaurin." `http://www.best.com/~ xah / Special Plane Curves _ dir / TriOf Maclaurin _ dir / triOf Maclaurin.html`.

MacTutor History of Mathematics Archive. "Trisectrix of Maclaurin." `http: //www-groups .dcs . st-and . ac . uk/ ~history/Curves/Trisectrix.html`.

Maclaurin Trisectrix Inverse Curve

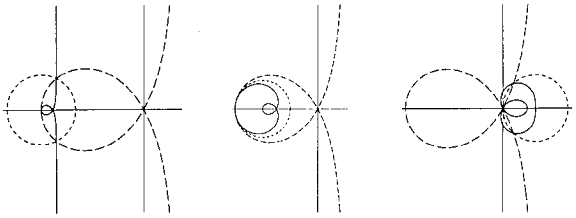

The INVERSE CURVE of the MACLAURIN TRISECTRIX with INVERSION CENTER at the NEGATIVE x-intercept is a TSCHIRNHAUSEN CUBIC.

MacMahon's Prime Number of Measurement

see PRIME NUMBER OF MEASUREMENT

MacRobert's E-Function

$$\begin{aligned}&E(p;\alpha_r:\rho_s:x)\\&\quad\equiv\frac{\Gamma(\alpha_{q+1})}{\Gamma(\rho_1-\alpha_1)\Gamma(\rho_2-\alpha_2)\cdots\Gamma(\rho_q-\alpha_q)}\\&\quad\times\prod_{\mu=1}^{q}\int_0^\infty {\lambda_\mu}^{\rho_\mu-\alpha_\mu-1}(1+\lambda_\mu)^{-\rho_\mu}\,d\lambda_\mu\\&\quad\times\prod_{\nu=2}^{p-q-1}\int_0^\infty e^{-\lambda_{q+\nu}}{\lambda_{q+\nu}}^{\alpha_{q+\nu}-1}\,d\lambda_{q+\nu}\times\\&\int_0^\infty e^{-\lambda_p}{\lambda_p}^{\alpha_p-1}\left[1+\frac{\lambda_{q+2}\lambda_{q+3}\cdots\lambda_p}{(1+\lambda_1)\cdots(1+\lambda_q)x}\right]^{-\alpha_q+1}d\lambda_p,\end{aligned}$$

where $\Gamma(z)$ is the GAMMA FUNCTION and other details are discussed by Gradshteyn and Ryzhik (1980).

see also FOX'S H-FUNCTION, MEIJER'S G-FUNCTION

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 896–903 and 1071–1072, 1979.

Madelung Constants

The quantities obtained from cubic, hexagonal, etc., LATTICE SUMS, evaluated at $s=1$, are called Madelung constants. For cubic LATTICE SUMS, they are expressible in closed form for EVEN indices,

$$b_2(2)=-4\beta(1)\eta(1)=-4\tfrac{\pi}{4}\ln 2=-\pi\ln 2 \qquad (1)$$

$$b_4(2)=-8\eta(1)\eta(0)=-8\ln 2\cdot\tfrac{1}{2}=-4\ln 2. \qquad (2)$$

$b_3(1)$ is given by BENSON'S FORMULA,

$$\begin{aligned}-b_3(1)&=\sum_{i,\,j,\,k=-\infty}^{\infty}{}' \frac{(-1)^{i+j+k+1}}{\sqrt{i^2+j^2+k^2}}\\&=12\pi\sum_{m,\,n=1,\,3,\,\ldots}^{\infty}\operatorname{sech}^2\left(\tfrac{1}{2}\pi\sqrt{m^2+n^2}\right),\end{aligned} \qquad (3)$$

where the prime indicates that summation over (0, 0, 0) is excluded. $b_3(1)$ is sometimes called "the" Madelung constant, corresponds to the Madelung constant for a 3-D NaCl crystal, and is numerically equal to $-1.74756\ldots$.

For hexagonal LATTICE SUM, $h_2(2)$ is expressible in closed form as

$$h_2(2)=\pi\ln 3\sqrt{3}. \qquad (4)$$

see also BENSON'S FORMULA, LATTICE SUM

References

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, 1987.

Buhler, J. and Wagon, S. "Secrets of the Madelung Constant." *Mathematica in Education and Research* **5**, 49–55, Spring 1996.

Crandall, R. E. and Buhler, J. P. "Elementary Function Expansions for Madelung Constants." *J. Phys. Ser. A: Math. and Gen.* **20**, 5497–5510, 1987.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/mdlung/mdlung.html`.

Maeder's Owl Minimal Surface

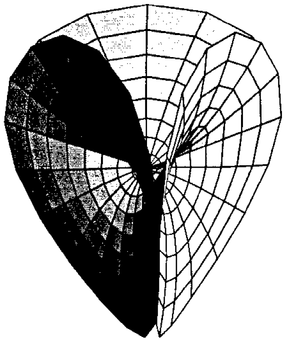

A MINIMAL SURFACE which resembles a CROSS-CAP. It is given by the polar equations

$$x=1 \qquad (1)$$

$$y=\sqrt{z} \qquad (2)$$

$$z=z, \qquad (3)$$

or the parametric equations

$$x = r\cos\theta - \tfrac{1}{2}r^2\cos(2\theta) \qquad (4)$$

$$y = -r\sin\theta - \tfrac{1}{2}r^2\sin(2\theta), \qquad (5)$$

$$z = \tfrac{4}{3}r^{3/2}\cos(\tfrac{3}{2}\theta). \qquad (6)$$

see also CROSS-CAP, MINIMAL SURFACE

References

Maeder, R. *Programming in Mathematica, 3rd ed.* Reading, MA: Addison-Wesley, pp. 29–30, 1997.

Maehly's Procedure

A method for finding ROOTS which defines

$$P_j(x) = \frac{P(x)}{(x-x_1)\cdots(x-x_j)}, \qquad (1)$$

so the derivative is

$$P_j'(x) = \frac{P'(x)}{(x-x_1)\cdots(x-x_j)} - \frac{P(x)}{(x-x_1)\cdots(x-x_j)}\sum_{i=1}^{j}(x-x_i)^{-1}. \qquad (2)$$

One step of NEWTON'S METHOD can then be written as

$$x_{k+1} = x_k - \frac{P(x_k)}{P'(x_k) - P(x_k)\sum_{i=1}^{j}(x_k-x_i)^{-1}}. \qquad (3)$$

Mainardi-Codazzi Equations

$$\frac{\partial e}{\partial v} - \frac{\partial f}{\partial u} = e\Gamma^1_{12} + f(\Gamma^2_{12} - \Gamma^1_{11}) - g\Gamma^2_{11} \qquad (1)$$

$$\frac{\partial f}{\partial v} - \frac{\partial g}{\partial u} = e\Gamma^1_{22} + f(\Gamma^2_{22} - \Gamma^1_{12}) - g\Gamma^2_{12}, \qquad (2)$$

where e, f, and g are coefficients of the second FUNDAMENTAL FORM and Γ^k_{ij} are CHRISTOFFEL SYMBOLS OF THE SECOND KIND. Therefore,

$$\frac{\partial e}{\partial v} = \tfrac{1}{2}E_v\left(\frac{e}{E} + \frac{g}{G}\right) \qquad (3)$$

$$\frac{\partial g}{\partial u} = \tfrac{1}{2}G_u\left(\frac{e}{E} + \frac{g}{G}\right) \qquad (4)$$

$$\frac{\partial(\ln f)}{\partial u} = \Gamma^1_{11} - \Gamma^2_{12} \qquad (5)$$

$$\frac{\partial(\ln f)}{\partial v} = \Gamma^2_{22} - \Gamma^1_{12} \qquad (6)$$

$$\frac{\partial}{\partial u}\left(\frac{\ln f}{\sqrt{EG-F^2}}\right) = -2\Gamma^2_{12} \qquad (7)$$

$$\frac{\partial}{\partial v}\left(\frac{\ln f}{\sqrt{EG-F^2}}\right) = -2\Gamma^1_{12}, \qquad (8)$$

where E, F, and G are coefficients of the first FUNDAMENTAL FORM.

References

Gray, A. "The Mainardi-Codazzi Equations." §20.4 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 401–402, 1993.

Green, A. E. and Zerna, W. *Theoretical Elasticity, 2nd ed.* New York: Dover, p. 37, 1992.

Magic Circles

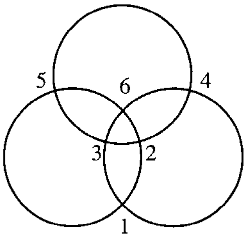

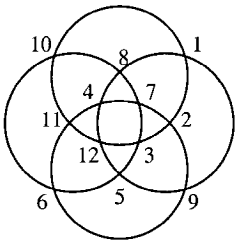

A set of n magic circles is a numbering of the intersection of the n CIRCLES such that the sum over all intersections is the same constant for all circles. The above sets of three and four magic circles have magic constants 14 and 39 (Madachy 1979).

see also MAGIC GRAPH, MAGIC SQUARE

References

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, p. 86, 1979.

Magic Constant

The number

$$M_2(n) = \frac{1}{n}\sum_{k=1}^{n^2} k = \tfrac{1}{2}n(n^2+1)$$

to which the n numbers in any horizontal, vertical, or *main* diagonal line must sum in a MAGIC SQUARE. The first few values are 1, 5 (no such magic square), 15, 34, 65, 111, 175, 260, ... (Sloane's A006003). The magic constant for an nth order magic square starting with an INTEGER A and with entries in an increasing ARITHMETIC SERIES with difference D between terms is

$$M_2(n; A, D) = \tfrac{1}{2}n[2a + D(n^2-1)]$$

(Hunter and Madachy 1975, Madachy 1979). In a PANMAGIC SQUARE, in addition to the main diagonals, the broken diagonals also sum to $M_2(n)$.

For a MAGIC CUBE, the magic constant is

$$M_3(n) = \frac{1}{n^2}\sum_{k=1}^{n^3} k = \tfrac{1}{2}n(n^3+1) = \tfrac{1}{2}n(1+n)(n^2-n+1).$$

The first few values are 1, 9, 42, 130, 315, 651, 1204, ... (Sloane's A027441).

There is a corresponding multiplicative magic constant for MULTIPLICATION MAGIC SQUARES.

see also MAGIC CUBE, MAGIC GEOMETRIC CONSTANTS, MAGIC HEXAGON, MAGIC SQUARE, MULTIPLICATION MAGIC SQUARE, PANMAGIC SQUARE

References

Hunter, J. A. H. and Madachy, J. S. "Mystic Arrays." Ch. 3 in *Mathematical Diversions.* New York: Dover, pp. 23–34, 1975.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, p. 86, 1979.

Sloane, N. J. A. Sequences A027441 and A006003/M3849 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Magic Cube

An $n \times n \times n$ 3-D version of the MAGIC SQUARE in which the n^2 rows, n^2 columns, n^2 pillars (or "files"), and four space diagonals each sum to a single number $M_3(n)$ known as the MAGIC CONSTANT. If the cross-section diagonals also sum to $M_3(n)$, the magic cube is called a PERFECT MAGIC CUBE; if they do not, the cube is called a SEMIPERFECT MAGIC CUBE, or sometimes an ANDREWS CUBE (Gardner 1988). A pandiagonal cube is a perfect or semiperfect magic cube which is magic not only along the main space diagonals, but also on the broken space diagonals.

A magic cube using the numbers 1, 2, ..., n^3, if it exists, has MAGIC CONSTANT

$$M_3(n) = \frac{1}{n^2}\sum_{k=1}^{n^3} k = \tfrac{1}{2}n(n^3+1) = \tfrac{1}{2}n(n+1)(n^2-n+1).$$

For $n = 1, 2, \ldots$, the magic constants are 1, 9, 42, 130, 315, 651, ... (Sloane's A027441).

4	12	26
11	25	6
27	5	10

20	7	15
9	14	19
13	21	8

18	23	1
22	3	17
2	16	24

60	37	12	21
13	20	61	36
56	41	8	25
1	32	49	48

7	26	55	42
50	47	2	31
11	22	59	38
62	35	14	19

57	40	9	24
16	17	64	33
53	44	5	28
4	29	52	45

6	27	54	43
51	46	3	30
10	23	58	39
63	34	15	18

The above semiperfect magic cubes of orders three (Hunter and Madachy 1975, p. 31; Ball and Coxeter 1987, p. 218) and four (Ball and Coxeter 1987, p. 220) have magic constants 42 and 130, respectively. There is a trivial semiperfect magic cube of order one, but no semiperfect cubes of orders two or three exist. Semiperfect cubes of ODD order with $n \geq 5$ and DOUBLY EVEN order can be constructed by extending the methods used for MAGIC SQUARES.

19	497	255	285	432	78	324	162
303	205	451	33	148	370	128	414
336	174	420	66	243	273	31	509
116	402	160	382	463	45	291	193
486	8	266	236	89	443	181	343
218	316	54	472	357	135	393	107
185	347	85	439	262	232	490	12
389	103	361	139	58	476	214	312

134	360	106	396	313	219	469	55
442	92	342	184	5	487	233	267
473	59	309	215	102	392	138	364
229	263	9	491	346	188	438	88
371	145	415	125	208	302	36	450
79	429	163	321	500	18	288	254
48	462	196	290	403	113	383	157
276	242	512	30	175	333	67	417

306	212	478	64	141	367	97	387
14	496	226	260	433	83	349	191
109	399	129	355	466	52	318	224
337	179	445	95	238	272	2	484
199	293	43	457	380	154	408	118
507	25	279	245	72	422	172	330
412	122	376	150	39	453	203	297
168	326	76	426	283	249	503	21

423	69	331	169	28	506	248	278
155	377	119	405	296	198	460	42
252	282	24	502	327	165	427	73
456	38	300	202	123	409	151	373
82	436	190	352	493	15	257	227
366	144	386	100	209	307	61	479
269	239	481	3	178	340	94	448
49	467	221	319	398	112	354	132

381	159	401	115	194	292	46	464
65	419	173	335	510	32	274	244
34	452	206	304	413	127	369	147
286	256	498	20	161	323	77	431
140	362	104	390	311	213	475	57
440	86	348	186	11	489	231	261
471	53	315	217	108	394	136	358
235	265	7	485	344	182	444	90

492	10	264	230	87	437	187	345
216	310	60	474	363	137	391	101
183	341	91	441	268	234	488	6
395	105	359	133	56	470	220	314
29	511	241	275	418	68	334	176
289	195	461	47	158	384	114	404
322	164	430	80	253	287	17	499
126	416	146	372	449	35	301	207

96	446	180	338	483	1	271	237
356	130	400	110	223	317	51	465
259	225	495	13	192	350	84	434
63	477	211	305	388	98	368	142
425	75	325	167	22	504	250	284
149	375	121	411	298	204	454	40
246	280	26	508	329	171	421	71
458	44	294	200	117	407	153	379

201	299	37	455	374	152	410	124
501	23	281	251	74	428	166	328
406	120	378	156	41	459	197	295
170	332	70	424	277	247	505	27
320	222	468	50	131	353	111	397
4	482	240	270	447	93	339	177
99	385	143	365	480	62	308	210
351	189	435	81	228	258	16	494

There are no perfect magic cubes of order four (Beeler *et al.* 1972, Item 50; Gardner 1988). No perfect magic cubes of order five are known, although it is known that such a cube must have a central value of 63 (Beeler *et al.* 1972, Item 51; Gardner 1988). No order-six perfect magic cubes are known, but Langman (1962) constructed a perfect magic cube of order seven. An order-eight perfect magic cube was published anonymously in 1875 (Barnard 1888, Benson and Jacoby 1981, Gardner 1988). The construction of such a cube is discussed in Ball and Coxeter (1987). Rosser and Walker rediscovered the order-eight cube in the late 1930s (but did not publish it), and Myers independently discovered the cube illustrated above in 1970 (Gardner 1988). Order 9 and 11 magic cubes have also been discovered, but none of order 10 (Gardner 1988).

Semiperfect pandiagonal cubes exist for all orders $8n$ and all ODD $n > 8$ (Ball and Coxeter 1987). A perfect pandiagonal magic cube has been constructed by Planck (1950), cited in Gardner (1988).

Berlekamp *et al.* (1982, p. 783) give a magic TESSERACT.

see also MAGIC CONSTANT, MAGIC GRAPH, MAGIC HEXAGON, MAGIC SQUARE

References

Adler, A. and Li, S.-Y. R. "Magic Cubes and Prouhet Sequences." *Amer. Math. Monthly* **84**, 618–627, 1977.

Andrews, W. S. *Magic Squares and Cubes, 2nd rev. ed.* New York: Dover, 1960.

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 216–224, 1987.

Barnard, F. A. P. "Theory of Magic Squares and Cubes." *Mem. Nat. Acad. Sci.* **4**, 209–270, 1888.

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Benson, W. H. and Jacoby, O. *Magic Cubes: New Recreations.* New York: Dover, 1981.

Berlekamp, E. R.; Conway, J. H; and Guy, R. K. *Winning Ways, For Your Mathematical Plays, Vol. 2: Games in Particular.* London: Academic Press, 1982.

Gardner, M. "Magic Squares and Cubes." Ch. 17 in *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, pp. 213–225, 1988.

Hendricks, J. R. "Ten Magic Tesseracts of Order Three." *J. Recr. Math.* **18**, 125–134, 1985–1986.

Hirayama, A. and Abe, G. *Researches in Magic Squares.* Osaka, Japan: Osaka Kyoikutosho, 1983.

Hunter, J. A. H. and Madachy, J. S. "Mystic Arrays." Ch. 3 in *Mathematical Diversions.* New York: Dover, p. 31, 1975.

Langman, H. *Play Mathematics.* New York: Hafner, pp. 75–76, 1962.

Lei, A. "Magic Cube and Hypercube." `http://www.cs.ust.hk/~philipl/magic/mcube2.html`.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 99–100, 1979.

Pappas, T. "A Magic Cube." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 77, 1989.

Planck, C. *Theory of Path Nasiks.* Rugby, England: Privately Published, 1905.

Rosser, J. B. and Walker, R. J. "The Algebraic Theory of Diabolical Squares." *Duke Math. J.* **5**, 705–728, 1939.

Sloane, N. J. A. Sequence A027441 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Wynne, B. E. "Perfect Magic Cubes of Order 7." *J. Recr. Math.* **8**, 285–293, 1975–1976.

Magic Geometric Constants

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let E be a compact connected subset of d-dimensional EUCLIDEAN SPACE. Gross (1964) and Stadje (1981) proved that there is a unique REAL NUMBER $a(E)$ such that for all $x_1, x_2, \ldots, x_n \in E$, there exists $y \in E$ with

$$\frac{1}{n}\sum_{j=1}^{n}\sqrt{\sum_{k=1}^{d}(x_{j,k}-y_k)^2} = a(E). \qquad (1)$$

The magic constant $m(E)$ of E is defined by

$$m(E) = \frac{a(E)}{\operatorname{diam}(E)}, \qquad (2)$$

where

$$\operatorname{diam}(E) \equiv \max_{u,v\in E}\sqrt{\sum_{k=1}^{d}(u_k - v_k)^2}. \qquad (3)$$

These numbers are also called DISPERSION NUMBERS and RENDEZVOUS VALUES. For any E, Gross (1964) and Stadje (1981) proved that

$$\tfrac{1}{2} \leq m(E) < 1. \qquad (4)$$

If I is a subinterval of the LINE and D is a circular DISK in the PLANE, then

$$m(I) = m(D) = \tfrac{1}{2}. \qquad (5)$$

If C is a CIRCLE, then

$$m(C) = \frac{2}{\pi} = 0.6366\ldots. \qquad (6)$$

An expression for the magic constant of an ELLIPSE in terms of its SEMIMAJOR and SEMIMINOR AXES lengths is not known. Nikolas and Yost (1988) showed that for a REULEAUX TRIANGLE T

$$0.6675276 \leq m(T) \leq 0.6675284. \qquad (7)$$

Denote the MAXIMUM value of $m(E)$ in n-D space by $M(n)$. Then

$$M(1) = \tfrac{1}{2} \qquad (8)$$

$$M(2) : m(T) \leq M(2) \leq \frac{2+\sqrt{3}}{3\sqrt{3}} < 0.7182336 \qquad (9)$$

$$M(d) : \frac{d}{d+1} \leq M(d) \leq \frac{[\Gamma(\frac{1}{2}d)]^2 2^{d-2}\sqrt{2d}}{\Gamma(d-\frac{1}{2})\sqrt{(d+1)\pi}} < \sqrt{\frac{d}{d+1}}, \qquad (10)$$

where $\Gamma(z)$ is the GAMMA FUNCTION (Nikolas and Yost 1988).

An unrelated quantity characteristic of a given MAGIC SQUARE is also known as a MAGIC CONSTANT.

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/magic/magic.html`.

Cleary, J.; Morris, S. A.; and Yost, D. "Numerical Geometry—Numbers for Shapes." *Amer. Math. Monthly* **95**, 260–275, 1986.

Croft, H. T.; Falconer, K. J.; and Guy, R. K. *Unsolved Problems in Geometry.* New York: Springer-Verlag, 1994.

Gross, O. *The Rendezvous Value of Metric Space.* Princeton, NJ: Princeton University Press, pp. 49–53, 1964.

Nikolas, P. and Yost, D. "The Average Distance Property for Subsets of Euclidean Space." *Arch. Math. (Basel)* **50**, 380–384, 1988.

Stadje, W. "A Property of Compact Connected Spaces." *Arch. Math. (Basel)* **36**, 275–280, 1981.

Magic Graph

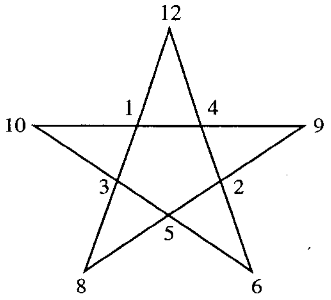

A LABELLED GRAPH with e EDGES labeled with distinct elements $\{1, 2, \ldots, e\}$ so that the sum of the EDGE labels at each VERTEX is the same. Another type of magic graph, such as the PENTAGRAM shown above, has labelled VERTICES which give the same sum along every straight line segment (Madachy 1979).

see also ANTIMAGIC GRAPH, LABELLED GRAPH, MAGIC CIRCLES, MAGIC CONSTANT, MAGIC CUBE, MAGIC HEXAGON, MAGIC SQUARE

References

Hartsfield, N. and Ringel, G. *Pearls in Graph Theory: A Comprehensive Introduction.* San Diego, CA: Academic Press, 1990.

Heinz, H. "Magic Stars." `http://www.geocities.com/CapeCanaveral/Launchpad/4057/magicstar.htm`.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 98–99, 1979.

Magic Hexagon

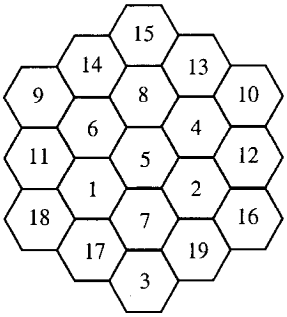

An arrangement of close-packed HEXAGONS containing the numbers 1, 2, ..., $H_n = 3n(n-1)+1$, where H_n is the nth HEX NUMBER, such that the numbers along each straight line add up to the same sum. In the above magic hexagon, each line (those of lengths both 3 and 4) adds up to 38. This is the only magic hexagon of the counting numbers for any size hexagon. It was discovered by C. W. Adam, who worked on the problem from 1910 to 1957.

see also HEX NUMBER, HEXAGON, MAGIC GRAPH, MAGIC SQUARE, TALISMAN HEXAGON

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Item 49, Feb. 1972.

Gardner, M. "Permutations and Paradoxes in Combinatorial Mathematics." *Sci. Amer.* **209**, 112–119, Aug. 1963.

Honsberger, R. *Mathematical Gems I.* Washington, DC: Math. Assoc. Amer., pp. 69–76, 1973.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 100–101, 1979.

Magic Labelling

It is conjectured that every TREE with e edges whose nodes are all trivalent or monovalent can be given a "magic" labelling such that the INTEGERS 1, 2, ..., e can be assigned to the edges so that the SUM of the three meeting at a node is constant.

see also MAGIC CONSTANT, MAGIC CUBE, MAGIC GRAPH, MAGIC HEXAGON, MAGIC SQUARE

References

Guy, R. K. "Unsolved Problems Come of Age." *Amer. Math. Monthly* **96**, 903–909, 1989.

Magic Number

see MAGIC CONSTANT

Magic Series

n numbers form a magic series of degree p if the sum of their kth POWERS is the MAGIC CONSTANT of degree k for all $k \in [1, p]$.

see also MAGIC CONSTANT, MAGIC SQUARE

References

Kraitchik, M. "Magic Series." §7.13.3 in *Mathematical Recreations.* New York: W. W. Norton, pp. 183–186, 1942.

Magic Square

8	1	6
3	5	7
4	9	2

16	2	3	13
5	11	10	8
9	7	6	12
4	14	15	1

17	24	1	8	15
23	5	7	14	16
4	6	13	20	22
10	12	19	21	3
11	18	25	2	9

32	29	4	1	24	21
30	31	2	3	22	23
12	9	17	20	28	25
10	11	18	19	26	27
13	16	36	33	5	8
14	15	34	35	6	7

30	39	48	1	10	19	28
38	47	7	9	18	27	29
46	6	8	17	26	35	37
5	14	16	25	34	36	45
13	15	24	33	42	44	4
21	23	32	41	43	3	12
22	31	40	49	2	11	20

64	2	3	61	60	6	7	57
9	55	54	12	13	51	50	16
17	47	46	20	21	43	42	24
40	26	27	37	36	30	31	33
32	34	35	29	28	38	39	25
41	23	22	44	45	19	18	48
49	15	14	52	53	11	10	56
8	58	59	5	4	62	63	1

A (normal) magic square consists of the distinct POSITIVE INTEGERS 1, 2, ..., n^2 such that the sum of the n numbers in any horizontal, vertical, or *main* diagonal line is always the same MAGIC CONSTANT

$$M_2(n) = \frac{1}{n}\sum_{k=1}^{n^2} k = \tfrac{1}{2}n(n^2+1).$$

The unique normal square of order three was known to the ancient Chinese, who called it the LO SHU. A version of the order 4 magic square with the numbers 15 and 14 in adjacent middle columns in the bottom row is called DÜRER'S MAGIC SQUARE. Magic squares of order 3 through 8 are shown above.

The MAGIC CONSTANT for an nth order magic square starting with an INTEGER A and with entries in an increasing ARITHMETIC SERIES with difference D between terms is

$$M_2(n; A, D) = \tfrac{1}{2}n[2a + D(n^2-1)]$$

(Hunter and Madachy 1975). If every number in a magic square is subtracted from n^2+1, another magic square is obtained called the complementary magic square. Squares which are magic under multiplication instead of addition can be constructed and are known as MULTIPLICATION MAGIC SQUARES. In addition, squares which are magic under both addition *and* multiplication can be constructed and are known as ADDITION-MULTIPLICATION MAGIC SQUARES (Hunter and Madachy 1975).

A square that fails to be magic only because one or both of the main diagonal sums do not equal the MAGIC CONSTANT is called a SEMIMAGIC SQUARE. If *all* diagonals (including those obtained by wrapping around) of a magic square sum to the MAGIC CONSTANT, the square is said to be a PANMAGIC SQUARE (also called a DIABOLICAL SQUARE or PANDIAGONAL SQUARE). If replacing each number n_i by its square ${n_i}^2$ produces another magic square, the square is said to be a BIMAGIC SQUARE (or DOUBLY MAGIC SQUARE). If a square is magic for n_i, ${n_i}^2$, and ${n_i}^3$, it is called a TREBLY MAGIC SQUARE. If all pairs of numbers symmetrically opposite the center sum to $n^2 + 1$, the square is said to be an ASSOCIATIVE MAGIC SQUARE.

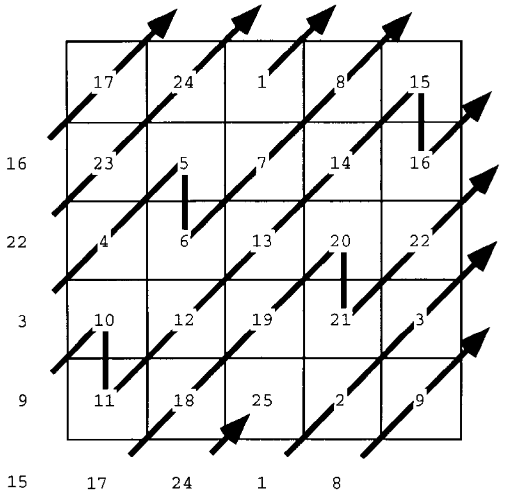

Kraitchik (1942) gives general techniques of constructing EVEN and ODD squares of order n. For n ODD, a very straightforward technique known as the Siamese method can be used, as illustrated above (Kraitchik 1942, pp. 148–149). It begins by placing a 1 in any location (in the center square of the top row in the above example), then incrementally placing subsequent numbers in the square one unit above and to the right. The counting is wrapped around, so that falling off the top returns on the bottom and falling off the right returns on the left. When a square is encountered which is already filled, the next number is instead placed *below* the previous one and the method continues as before. The method, also called de la Loubere's method, is purported to have been first reported in the West when de la Loubere returned to France after serving as ambassador to Siam.

A generalization of this method uses an "ordinary vector" (x, y) which gives the offset for each noncolliding move and a "break vector" (u, v) which gives the offset to introduce upon a collision. The standard Siamese method therefore has ordinary vector (1, −1) and break vector (0, 1). In order for this to produce a magic square, each break move must end up on an unfilled cell. Special classes of magic squares can be constructed by considering the absolute sums $|u+v|$, $|(u-x)+(v-y)|$, $|u-v|$, and $|(u-x)-(v-y)| = |u+y-x-v|$. Call the set of these numbers the sumdiffs (sums and differences). If all sumdiffs are RELATIVELY PRIME to n and the square is a magic square, then the square is also a PANMAGIC SQUARE. This theory originated with de la Hire. The following table gives the sumdiffs for particular choices of ordinary and break vectors.

Ordinary Vector	Break Vector	Sumdiffs	Magic Squares	Panmagic Squares
(1, −1)	(0, 1)	(1, 3)	$2k+1$	none
(1, −1)	(0, 2)	(0, 2)	$6k \pm 1$	none
(2, 1)	(1, −2)	(1, 2, 3, 4)	$6k \pm 1$	none
(2, 1)	(1, −1)	(0, 1, 2, 3)	$6k \pm 1$	$6k \pm 1$
(2, 1)	(1, 0)	(0, 1, 2)	$2k+1$	none
(2, 1)	(1, 2)	(0, 1, 2, 3)	$6k \pm 1$	none

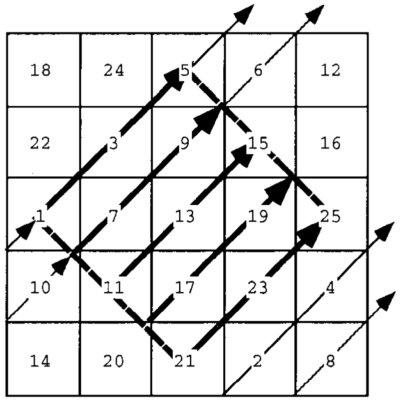

A second method for generating magic squares of ODD order has been discussed by J. H. Conway under the name of the "lozenge" method. As illustrated above, in this method, the ODD numbers are built up along diagonal lines in the shape of a DIAMOND in the central part of the square. The EVEN numbers which were missed are then added sequentially along the continuation of the diagonal obtained by wrapping around the square until the wrapped diagonal reaches its initial point. In the above square, the first diagonal therefore fills in 1, 3, 5, 2, 4, the second diagonal fills in 7, 9, 6, 8, 10, and so on.

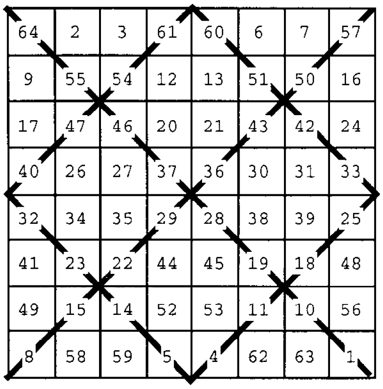

64	2	3	61	60	6	7	57
9	55	54	12	13	51	50	16
17	47	46	20	21	43	42	24
40	26	27	37	36	30	31	33
32	34	35	29	28	38	39	25
41	23	22	44	45	19	18	48
49	15	14	52	53	11	10	56
8	58	59	5	4	62	63	1

An elegant method for constructing magic squares of DOUBLY EVEN order $n = 4m$ is to draw xs through each 4×4 subsquare and fill all squares in sequence. Then replace each entry a_{ij} on a crossed-off diagonal by $(n^2 + 1) - a_{ij}$ or, equivalently, reverse the order of the crossed-out entries. Thus in the above example for $n = 8$, the crossed-out numbers are originally 1, 4, ..., 61, 64, so entry 1 is replaced with 64, 4 with 61, etc.

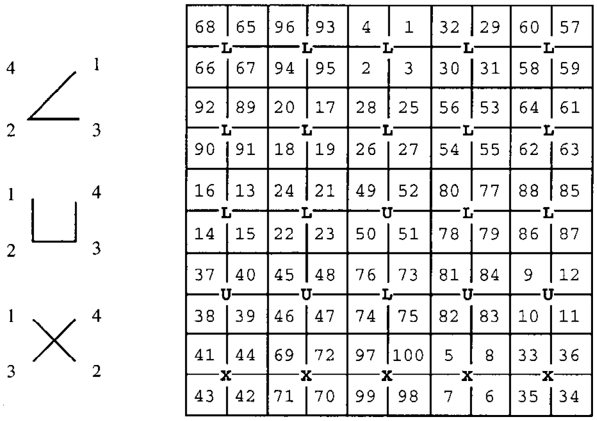

68	65	96	93	4	1	32	29	60	57
66	67	94	95	2	3	30	31	58	59
92	89	20	17	28	25	56	53	64	61
90	91	18	19	26	27	54	55	62	63
16	13	24	21	49	52	80	77	88	85
14	15	22	23	50	51	78	79	86	87
37	40	45	48	76	73	81	84	9	12
38	39	46	47	74	75	82	83	10	11
41	44	69	72	97	100	5	8	33	36
43	42	71	70	99	98	7	6	35	34

A very elegant method for constructing magic squares of SINGLY EVEN order $n = 4m + 2$ with $m \geq 1$ (there is no magic square of order 2) is due to J. H. Conway, who calls it the "LUX" method. Create an array consisting of $m + 1$ rows of Ls, 1 row of Us, and $m - 1$ rows of Xs, all of length $n/2 = 2m + 1$. Interchange the middle U with the L above it. Now generate the magic square of order $2m + 1$ using the Siamese method centered on the array of letters (starting in the center square of the top row), but fill each set of four squares surrounding a letter sequentially according to the order prescribed by the the letter. That order is illustrated on the left side of the above figure, and the completed square is illustrated to the right. The "shapes" of the letters L, U, and X naturally suggest the filling order, hence the name of the algorithm.

It is an unsolved problem to determine the number of magic squares of an arbitrary order, but the number of distinct magic squares (excluding those obtained by rotation and reflection) of order $n = 1, 2, \ldots$ are 1, 0, 1, 880, 275305224, ... (Sloane's A006052; Madachy 1979, p. 87). The 880 squares of order four were enumerated by Frenicle de Bessy in the seventeenth century, and are illustrated in Berlekamp *et al.* (1982, pp. 778–783). The number of 6×6 squares is not known.

67	1	43
13	37	61
31	73	7

3	61	19	37
43	31	5	41
7	11	73	29
67	17	23	13

The above magic squares consist only of PRIMES and were discovered by E. Dudeney (1970) and A. W. Johnson, Jr. (Dewdney 1988). Madachy (1979, pp. 93–96) and Rivera discuss other magic squares composed of PRIMES.

52	61	4	13	20	29	36	45
14	3	62	51	46	35	30	19
53	60	5	12	21	28	37	44
11	6	59	54	43	38	27	22
55	58	7	10	23	26	39	42
9	8	57	56	41	40	25	24
50	63	2	15	18	31	34	47
16	1	64	49	48	33	32	17

Benjamin Franklin constructed the above 8×8 PANMAGIC SQUARE having MAGIC CONSTANT 260. Any half-row or half-column in this square totals 130, and the four corners plus the middle total 260. In addition, bent diagonals (such as 52-3-5-54-10-57-63-16) also total 260 (Madachy 1979, p. 87).

1480028159	1480028153	1480028201
1480028213	1480028171	1480028129
1480028141	1480028189	1480028183

In addition to other special types of magic squares, a 3×3 square whose entries are consecutive PRIMES, illustrated above, has been discovered by H. Nelson (Rivera). Variations on magic squares can also be constructed using letters (either in defining the square or as entries in it), such as the ALPHAMAGIC SQUARE and TEMPLAR MAGIC SQUARE.

4	9	2
3	5	7
8	1	6

4	14	15	1
9	7	6	12
5	11	10	8
16	2	3	13

11	24	7	20	3
4	12	25	8	16
17	5	13	21	9
10	18	1	14	22
23	6	19	2	15

6	32	3	34	35	1
7	11	27	28	8	30
19	14	16	15	23	24
18	20	22	21	17	13
25	29	10	9	26	12
36	5	33	4	2	31

22	47	16	41	10	35	4
5	23	48	17	42	11	29
30	6	24	49	18	36	12
13	31	7	25	43	19	37
38	14	32	1	26	44	20
21	39	8	33	2	27	45
46	15	40	9	34	3	28

8	58	59	5	4	62	63	1
49	15	14	52	53	11	10	56
41	23	22	44	45	19	18	48
32	34	35	29	28	38	39	25
40	26	27	37	36	30	31	33
17	47	46	20	21	43	42	24
9	55	54	12	13	51	50	16
64	2	3	61	60	6	7	57

37	78	29	70	21	62	13	54	5
6	38	79	30	71	22	63	14	46
47	7	39	80	31	72	23	55	15
16	48	8	40	81	32	64	24	56
57	17	49	9	41	73	33	65	25
26	58	18	50	1	42	74	34	66
67	27	59	10	51	2	43	75	35
36	68	19	60	11	52	3	44	76
77	28	69	20	61	12	53	4	45

Various numerological properties have also been associated with magic squares. Pivari associates the squares illustrated above with Saturn, Jupiter, Mars, the Sun, Venus, Mercury, and the Moon, respectively. Attractive patterns are obtained by connecting consecutive numbers in each of the squares (with the exception of the Sun magic square).

see also ADDITION-MULTIPLICATION MAGIC SQUARE ALPHAMAGIC SQUARE, ANTIMAGIC SQUARE, ASSOCIATIVE MAGIC SQUARE, BIMAGIC SQUARE, BORDER SQUARE, DÜRER'S MAGIC SQUARE, EULER SQUARE, FRANKLIN MAGIC SQUARE, GNOMON MAGIC SQUARE, HETEROSQUARE, LATIN SQUARE, MAGIC CIRCLES, MAGIC CONSTANT, MAGIC CUBE, MAGIC HEXAGON, MAGIC LABELLING, MAGIC SERIES, MAGIC TOUR, MULTIMAGIC SQUARE, MULTIPLICATION MAGIC SQUARE, PANMAGIC SQUARE, SEMIMAGIC SQUARE, TALISMAN SQUARE, TEMPLAR MAGIC SQUARE, TRIMAGIC SQUARE

References

Abe, G. "Unsolved Problems on Magic Squares." *Disc. Math.* **127**, 3–13, 1994.

Alejandre, S. "Suzanne Alejandre's Magic Squares." `http://forum.swarthmore.edu/alejandre/magic.square.html`.

Andrews, W. S. *Magic Squares and Cubes, 2nd rev. ed.* New York: Dover, 1960.

Ball, W. W. R. and Coxeter, H. S. M. "Magic Squares." Ch. 7 in *Mathematical Recreations and Essays, 13th ed.* New York: Dover, 1987.

Barnard, F. A. P. "Theory of Magic Squares and Cubes." *Memoirs Natl. Acad. Sci.* **4**, 209–270, 1888.

Benson, W. H. and Jacoby, O. *New Recreations with Magic Squares.* New York: Dover, 1976.

Berlekamp, E. R.; Conway, J. H; and Guy, R. K. *Winning Ways, For Your Mathematical Plays, Vol. 2: Games in Particular.* London: Academic Press, 1982.

Dewdney, A. K. "Computer Recreations: How to Pan for Primes in Numerical Gravel." *Sci. Amer.* **259**, pp. 120–123, July 1988.

Dudeney, E. *Amusements in Mathematics.* New York: Dover, 1970.

Fults, J. L. *Magic Squares.* Chicago, IL: Open Court, 1974.

Gardner, M. "Magic Squares." Ch. 12 in *The Second Scientific American Book of Mathematical Puzzles & Diversions: A New Selection.* New York: Simon and Schuster, 1961.

Gardner, M. "Magic Squares and Cubes." Ch. 17 in *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, 1988.

Grogono, A. W. "Magic Squares by Grog." `http://www.grogono.com/magic/`.

Heinz, H. "Magic Squares." `http://www.geocities.com/CapeCanaveral/Launchpad/4057/magicsquare.htm`.

Hirayama, A. and Abe, G. *Researches in Magic Squares.* Osaka, Japan: Osaka Kyoikutosho, 1983.

Horner, J. "On the Algebra of Magic Squares, I., II., and III." *Quart. J. Pure Appl. Math.* **11**, 57–65, 123–131, and 213–224, 1871.

Hunter, J. A. H. and Madachy, J. S. "Mystic Arrays." Ch. 3 in *Mathematical Diversions.* New York: Dover, pp. 23–34, 1975.

Kanada, Y. "Magic Square Page." `http://www.st.rim.or.jp/~kanada/puzzles/magic-square.html`.

Kraitchik, M. "Magic Squares." Ch. 7 in *Mathematical Recreations.* New York: Norton, pp. 142–192, 1942.

Lei, A. "Magic Square, Cube, Hypercube." `http://www.cs.ust.hk/~philipl/magic/magic.html`.

Madachy, J. S. "Magic and Antimagic Squares." Ch. 4 in *Madachy's Mathematical Recreations.* New York: Dover, pp. 85–113, 1979.

Moran, J. *The Wonders of Magic Squares.* New York: Vintage, 1982.

Pappas, T. "Magic Squares," "The 'Special' Magic Square," "The Pyramid Method for Making Magic Squares," "Ancient Tibetan Magic Square," "Magic 'Line.'," and "A Chinese Magic Square." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 82–87, 112, 133, 169, and 179, 1989.

Pivari, F. "Nice Examples." `http://www.geocities.com/CapeCanaveral/Lab/3469/examples.html`.

Pivari, F. "Simple Magic Square Checker and GIF Maker." `http://www.geocities.com/CapeCanaveral/Lab/3469/squaremaker.html`.

Rivera, C. "Problems & Puzzles (Puzzles): Magic Squares with Consecutive Primes." `http://www.sci.net.mx/~crivera/ppp/puzz_003.htm`.

Rivera, C. "Problems & Puzzles (Puzzles): Prime-Magical Squares." `http://www.sci.net.mx/~crivera/ppp/puzz_004.htm`.

Sloane, N. J. A. Sequence A006052/M5482 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Suzuki, M. "Magic Squares." `http://www.pse.che.tohoku.ac.jp/~msuzuki/MagicSquare.html`.

Weisstein, E. W. "Magic Squares." `http://www.astro.virginia.edu/~eww6n/math/notebooks/MagicSquares.m`.

Magic Star

see MAGIC GRAPH

Magic Tour

Let a chess piece make a TOUR on an $n \times n$ CHESSBOARD whose squares are numbered from 1 to n^2 along the path of the chess piece. Then the TOUR is called a magic tour if the resulting arrangement of numbers is a MAGIC SQUARE. If the first and last squares traversed are connected by a move, the tour is said to be closed (or "re-entrant"); otherwise it is open. The MAGIC CONSTANT for the 8 × 8 CHESSBOARD is 260.

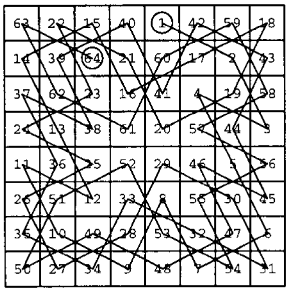

63	22	15	40	1	42	59	18
14	39	64	21	60	17	2	43
37	62	23	16	41	4	19	58
24	13	38	61	20	57	44	3
11	36	25	52	29	46	5	56
26	51	12	33	8	55	30	45
35	10	49	28	53	32	47	6
50	27	34	9	48	7	54	31

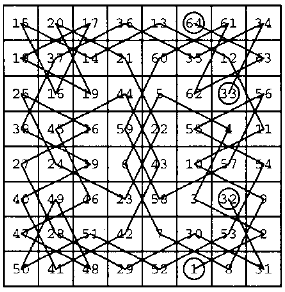

15	20	17	36	13	64	61	34
18	37	14	21	60	35	12	63
25	16	19	44	5	62	33	56
38	45	26	59	22	55	4	11
27	24	39	6	43	10	57	54
40	49	46	23	58	3	32	9
47	28	51	42	7	30	53	2
50	41	48	29	52	1	8	31

Magic KNIGHT'S TOURS are not possible on $n \times n$ boards for n ODD, and are believed to be impossible for $n = 8$. The "most magic" knight tour known on the 8 × 8 board is the SEMIMAGIC SQUARE illustrated in the above left figure (Ball and Coxeter 1987, p. 185) having main diagonal sums of 348 and 168. Combining two half-knights' tours one above the other as in the above right figure does, however, give a MAGIC SQUARE (Ball and Coxeter 1987, p. 185).

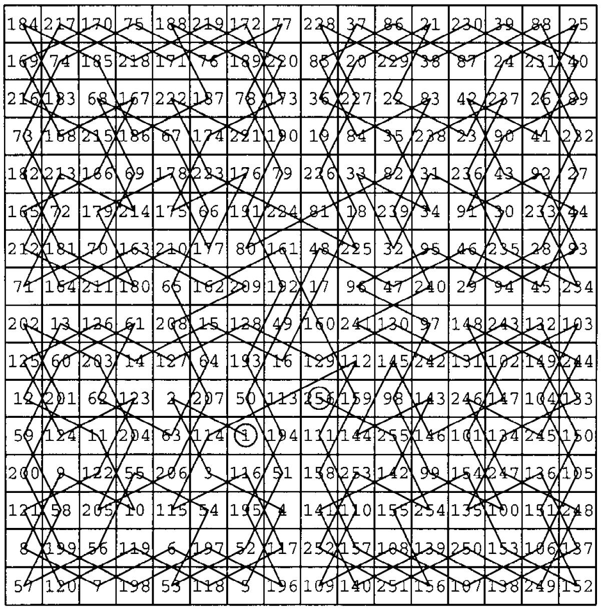

184	217	170	75	188	219	172	77	228	37	86	21	230	39	88	25
169	74	185	218	171	76	189	220	85	20	229	38	87	24	231	40
216	183	68	167	222	187	78	173	36	227	22	83	42	237	26	89
73	168	215	186	67	174	221	190	19	84	35	238	23	90	41	232
182	213	166	69	178	223	176	79	226	33	82	31	236	43	92	27
165	72	179	214	175	66	191	224	81	18	239	34	91	30	233	44
212	181	70	163	210	177	80	161	48	225	32	95	46	235	28	93
71	164	211	180	65	162	209	192	17	96	47	240	29	94	45	234
202	13	126	61	208	15	128	49	160	241	130	97	148	243	132	103
125	60	203	14	127	64	193	16	129	112	145	242	131	102	149	244
12	201	62	123	2	207	50	113	256	159	98	143	246	147	104	133
59	124	11	204	63	114	1	194	111	144	255	146	101	134	245	150
200	9	122	55	206	3	116	51	158	253	142	99	154	247	136	105
121	58	205	10	115	54	195	4	141	110	155	254	135	100	151	248
8	199	56	119	6	197	52	117	252	157	108	139	250	153	106	137
57	120	7	198	53	118	5	196	109	140	251	156	107	138	249	152

The above illustration shows a 16 × 16 closed magic KNIGHT'S TOUR (Madachy 1979).

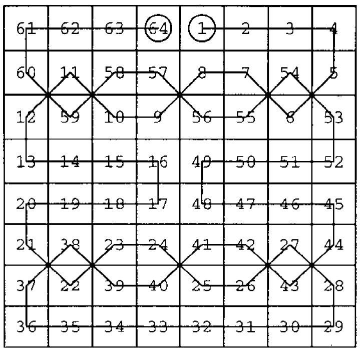

61	62	63	64	1	2	3	4
60	11	58	57	8	7	54	5
12	59	10	9	56	55	6	53
13	14	15	16	49	50	51	52
20	19	18	17	48	47	46	45
21	38	23	24	41	42	27	44
37	22	39	40	25	26	43	28
36	35	34	33	32	31	30	29

A magic tour for king moves is illustrated above (Coxeter 1987, p. 186).

see also CHESSBOARD, KNIGHT'S TOUR, MAGIC SQUARE, SEMIMAGIC SQUARE, TOUR

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 185–187, 1987.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 87–89, 1979.

Mahler-Lech Theorem

Let K be a FIELD of CHARACTERISTIC 0 (e.g., the rationals $\mathbb{Q}$) and let $\{u_n\}$ be a SEQUENCE of elements of K which satisfies a difference equation of the form

$$u_n = c_0 u_n + c_1 u_{n+1} + \ldots + c_k u_{n+k},$$

where the COEFFICIENTS c_i are fixed elements of K. Then, for any $c \in K$, we have *either* $u_n = c$ for only finitely many values of n, *or* $u_n = c$ for the values of n in some ARITHMETIC PROGRESSION.

The proof involves embedding certain fields inside the p-ADIC NUMBERS $\mathbb{Q}_p$ for some PRIME p, and using properties of zeros of POWER series over $\mathbb{Q}_p$ (STRASSMAN'S THEOREM).

see also ARITHMETIC PROGRESSION, p-ADIC NUMBER, STRASSMAN'S THEOREM

Mahler's Measure

For a POLYNOMIAL P,

$$M(P) = \exp \int_0^{2\pi} \ln |P(e^{i\theta})| \frac{d\theta}{2\pi}.$$

It is related to JENSEN'S INEQUALITY.

see also JENSEN'S INEQUALITY

Major Axis

see SEMIMAJOR AXIS

Major Triangle Center

A TRIANGLE CENTER $\alpha : \beta : \gamma$ is called a major center if the TRIANGLE CENTER FUNCTION $\alpha = f(a, b, c, A, B, C)$ is a function of ANGLE A alone, and therefore β and γ of B and C alone, respectively.

see also REGULAR TRIANGLE CENTER, TRIANGLE CENTER

References

Kimberling, C. "Major Centers of Triangles." *Amer. Math. Monthly* **104**, 431–438, 1997.

Majorant

A function used to study ORDINARY DIFFERENTIAL EQUATIONS.

Makeham Curve

The function defined by

$$y = ks^x b^{q^x}$$

which is used in actuarial science for specifying a simplified mortality law. Using $s(x)$ as the probability that a newborn will achieve age x, the Makeham law (1860) uses

$$s(x) = \exp(-Ax - m(c^x - 1))$$

for $B > 0$, $A \geq -B$, $c > 1$, $x \geq 0$.

see also GOMPERTZ CURVE, LIFE EXPECTANCY, LOGISTIC GROWTH CURVE, POPULATION GROWTH

References

Bowers, N. L. Jr.; Gerber, H. U.; Hickman, J. C.; Jones, D. A.; and Nesbitt, C. J. *Actuarial Mathematics.* Itasca, IL: Society of Actuaries, p. 71, 1997.

Makeham, W. M. "On the Law of Mortality, and the Construction of Annuity Tables." *JIA* **8**, 1860.

Malfatti Circles

Three circles packed inside a RIGHT TRIANGLE which are tangent to each other and to two sides of the TRIANGLE.

see also MALFATTI'S RIGHT TRIANGLE PROBLEM

Malfatti Points

see AJIMA-MALFATTI POINTS

Malfatti's Right Triangle Problem

Find the maximum total AREA of three CIRCLES (of possibly different sizes) which can be packed inside a RIGHT TRIANGLE of any shape without overlapping. In 1803, Malfatti gave the solution as three CIRCLES (the MALFATTI CIRCLES) tangent to each other and to two sides of the TRIANGLE. In 1929, it was shown that the MALFATTI CIRCLES were not always the best solution. Then Goldberg (1967) showed that, even worse, they are *never* the best solution.

see also MALFATTI'S TANGENT TRIANGLE PROBLEM

References

Eves, H. *A Survey of Geometry, Vol. 2.* Boston: Allyn & Bacon, p. 245, 1965.

Goldberg, M. "On the Original Malfatti Problem." *Math. Mag.* **40**, 241–247, 1967.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 145–147, 1990.

Malfatti's Tangent Triangle Problem

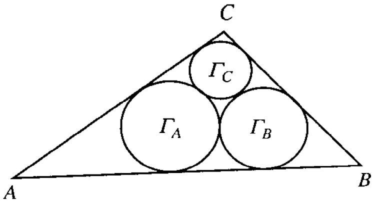

Draw within a given TRIANGLE three CIRCLES, each of which is TANGENT to the other two and to two sides of the TRIANGLE. Denote the three CIRCLES so constructed Γ_A, Γ_B, and Γ_C. Then Γ_A is tangent to AB and AC, Γ_B is tangent to BC and BA, and Γ_C is tangent to AC and BC.

see also AJIMA-MALFATTI POINTS, MALFATTI'S RIGHT TRIANGLE PROBLEM

References

Dörrie, H. "Malfatti's Problem." §30 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 147–151, 1965.

Forder, H. G. *Higher Course Geometry.* Cambridge, England: Cambridge University Press, pp. 244–245, 1931.

Fukagawa, H. and Pedoe, D. *Japanese Temple Geometry Problems (San Gaku).* Winnipeg: The Charles Babbage Research Centre, pp. 106–120, 1989.

Gardner, M. *Fractal Music, Hypercards, and More Mathematical Recreations from Scientific American Magazine.* New York: W. H. Freeman, pp. 163–165, 1992.

Goldberg, M. "On the Original Malfatti Problem." *Math. Mag.* **40**, 241–247, 1967.

Lob, H. and Richmond, H. W. "On the Solution of Malfatti's Problem for a Triangle." *Proc. London Math. Soc.* **2**, 287–304, 1930.

Woods, F. S. *Higher Geometry.* New York: Dover, pp. 206–209, 1961.

Malliavin Calculus

An infinite-dimensional DIFFERENTIAL CALCULUS on the WIENER SPACE. Also called STOCHASTIC CALCULUS OF VARIATIONS.

Mallow's Sequence

An INTEGER SEQUENCE given by the recurrence relation

$$a(n) = a(a(n-2)) + a(n - a(n-2))$$

with $a(1) = a(2) = 1$. The first few values are 1, 1, 2, 3, 3, 4, 5, 6, 6, 7, 7, 8, 9, 10, 10, 11, 12, 12, 13, 14, ... (Sloane's A005229).

see also HOFSTADTER-CONWAY $10,000 SEQUENCE, HOFSTADTER'S Q-SEQUENCE

References

Mallows, C. "Conway's Challenging Sequence." *Amer. Math. Monthly* **98**, 5–20, 1991.

Sloane, N. J. A. Sequence A005229/M0441 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Malmstén's Differential Equation

$$y'' + \frac{r}{z} y' = \left(Az^m + \frac{s}{z^2}\right) y.$$

References

Watson, G. N. *A Treatise on the Theory of Bessel Functions, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 99–100, 1966.

Maltese Cross

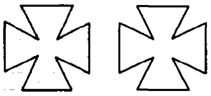

An irregular DODECAHEDRON CROSS shaped like a + sign but whose points flange out at the end: ✠. The conventional proportions as computed on a 5×5 grid as illustrated above.

see also CROSS, DISSECTION, DODECAHEDRON

References

Frederickson, G. "Maltese Crosses." Ch. 14 in *Dissections: Plane and Fancy.* New York: Cambridge University Press, pp. 157–162, 1997.

Maltese Cross Curve

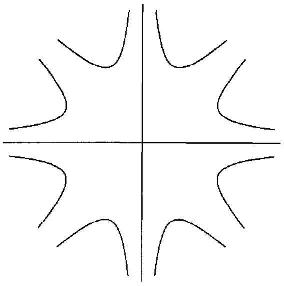

The plane curve with Cartesian equation

$$xy(x^2 - y^2) = x^2 + y^2$$

and polar equation

$$r^2 = \frac{1}{\cos\theta \sin\theta(\cos^2\theta - \sin^2\theta)}.$$

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 71, 1989.

Malthusian Parameter

The parameter α in the exponential POPULATION GROWTH equation

$$N_1(t) = N_0 e^{\alpha t}.$$

see also LIFE EXPECTANCY, POPULATION GROWTH

Mandelbar Set

A FRACTAL set analogous to the MANDELBROT SET or its generalization to a higher power with the variable z replaced by its COMPLEX CONJUGATE z^*.

see also MANDELBROT SET

Mandelbrot Set

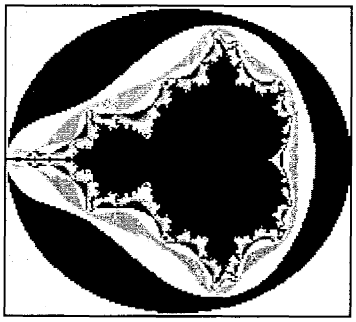

The set obtained by the QUADRATIC RECURRENCE

$$z_{n+1} = {z_n}^2 + C, \qquad (1)$$

where points C for which the orbit $z = 0$ does not tend to infinity are in the SET. It marks the set of points in the COMPLEX PLANE such that the corresponding JULIA SET is CONNECTED and not COMPUTABLE. The Mandelbrot set was originally called a μ MOLECULE by Mandelbrot.

J. Hubbard and A. Douady proved that the Mandelbrot set is CONNECTED. Shishikura (1994) proved that the boundary of the Mandelbrot set is a FRACTAL with HAUSDORFF DIMENSION 2. However, it is not yet known if the Mandelbrot set is pathwise-connected. If it is pathwise-connected, then Hubbard and Douady's proof implies that the Mandelbrot set is the image of a CIRCLE and can be constructed from a DISK by collapsing certain arcs in the interior (Douady 1986).

The AREA of the set is known to lie between 1.5031 and 1.5702; it is estimated as 1.50659....

Decomposing the COMPLEX coordinate $z = x + iy$ and $z_0 = a + ib$ gives

$$x' = x^2 - y^2 + a \qquad (2)$$

$$y' = 2xy + b. \qquad (3)$$

In practice, the limit is approximated by

$$\lim_{n\to\infty} |z_n| \approx \lim_{n\to n_{\max}} |z_n| < r_{\max}. \qquad (4)$$

Beautiful computer-generated plots can be created by coloring nonmember points depending on how quickly they diverge to $r_{\max}$. A common choice is to define an INTEGER called the COUNT to be the largest n such that $|z_n| < r$, where r is usually taken as $r = 2$, and to color points of different COUNT different colors. The boundary between successive COUNTS defines a series of "LEMNISCATES," called EQUIPOTENTIAL CURVES by Peitgen and Saupe (1988), $|L_n(C)| = r$ which have distinctive shapes. The first few LEMNISCATES are

$$L_1(C) = C \qquad (5)$$

$$L_2(C) = C(C+1) \qquad (6)$$

$$L_3(C) = C + (C + C^2)^2 \qquad (7)$$

$$L_4(C) = C + [C + (C + C^2)^2]^2. \qquad (8)$$

When written in CARTESIAN COORDINATES, the first three of these are

$$r^2 = x^2 + y^2 \tag{9}$$

$$r^2 = (x^2 + y^2)[(x+1)^2 + y^2] \tag{10}$$

$$\begin{aligned} r^2 = {} & (x^2 + y^2)(1 + 2x + 5x^2 + 6x^3 + 6x^4 + 4x^5 + x^6 \\ & - 3y^2 - 2xy^2 + 8x^2y^2 + 8x^3y^2 \\ & + 3x^4y^2 + 2y^4 + 4xy^4 + 3x^2y^4 + y^6), \end{aligned} \tag{11}$$

which are a CIRCLE, an OVAL, and a PEAR CURVE. In fact, the second LEMNISCATE L_2 can be written in terms of a new coordinate system with $x' \equiv x - 1/2$ as

$$[(x' - \tfrac{1}{2})^2 + y^2][(x' + \tfrac{1}{2})^2 + y^2] = r^2, \tag{12}$$

which is just a CASSINI OVAL with $a = 1/2$ and $b^2 = r$. The LEMNISCATES grow increasingly convoluted with higher COUNT and approach the Mandelbrot set as the COUNT tends to infinity.

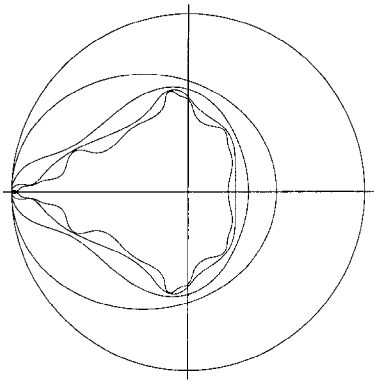

The kidney bean-shaped portion of the Mandelbrot set is bordered by a CARDIOID with equations

$$4x = 2\cos t - \cos(2t) \tag{13}$$

$$4y = 2\sin t - \sin(2t). \tag{14}$$

The adjoining portion is a CIRCLE with center at $(-1, 0)$ and RADIUS 1/4. One region of the Mandelbrot set containing spiral shapes is known as SEA HORSE VALLEY because the shape resembles the tail of a sea horse.

Generalizations of the Mandelbrot set can be constructed by replacing ${z_n}^2$ with ${z_n}^k$ or ${z_n^*}^k$, where k is a POSITIVE INTEGER and z^* denotes the COMPLEX CONJUGATE of z. The following figures show the FRACTALS obtained for $k = 2$, 3, and 4 (Dickau). The plots on the right have z replaced with z^* and are sometimes called "MANDELBAR SETS."

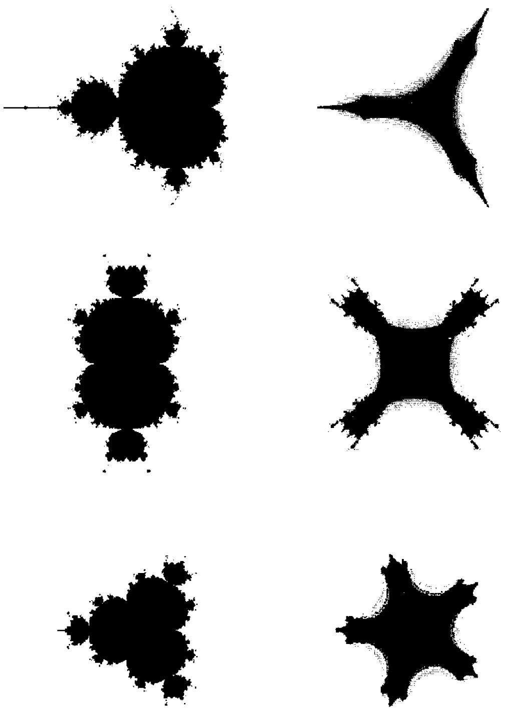

see also CACTUS FRACTAL, FRACTAL, JULIA SET, LEMNISCATE (MANDELBROT SET), MANDELBAR SET, QUADRATIC MAP, RANDELBROT SET, SEA HORSE VALLEY

References

Alfeld, P. "The Mandelbrot Set." `http://www.math.utah.edu/~alfeld/math/mandelbrot/mandelbrot1.html`.

Branner, B. "The Mandelbrot Set." In *Chaos and Fractals: The Mathematics Behind the Computer Graphics, Proc. Sympos. Appl. Math., Vol. 39* (Ed. R. L. Devaney and L. Keen). Providence, RI: Amer. Math. Soc., 75–105, 1989.

Dickau, R. M. "Mandelbrot (and Similar) Sets." `http://forum.swarthmore.edu/advanced/robertd/mandelbrot.html`.

Douady, A. "Julia Sets and the Mandelbrot Set." In *The Beauty of Fractals: Images of Complex Dynamical Systems* (Ed. H.-O. Peitgen and D. H. Richter). Berlin: Springer-Verlag, p. 161, 1986.

Eppstein, D. "Area of the Mandelbrot Set." `http://www.ics.uci.edu/~eppstein/junkyard/mand-area.html`.

Fisher, Y. and Hill, J. "Bounding the Area of the Mandelbrot Set." Submitted.

Hill, J. R. "Fractals and the Grand Internet Parallel Processing Project." Ch. 15 in *Fractal Horizons: The Future Use of Fractals.* New York: St. Martin's Press, pp. 299–323, 1996.

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 148–151 and 179–180, 1991.

Munafo, R. "Mu-Ency—The Encyclopedia of the Mandelbrot Set." `http://home.earthlink.net/~mrob/muency.html`.

Peitgen, H.-O. and Saupe, D. (Eds.). *The Science of Fractal Images.* New York: Springer-Verlag, pp. 178–179, 1988.

Shishikura, M. "The Boundary of the Mandelbrot Set has Hausdorff Dimension Two." *Astérisque*, No. 222, **7**, 389–405, 1994.

Mandelbrot Tree

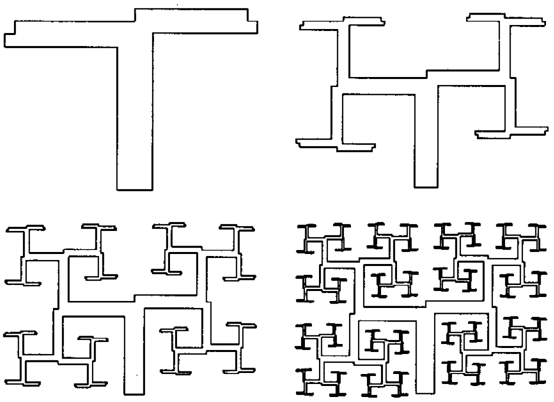

The FRACTAL illustrated above.

References

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 71–73, 1991.

Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m`.

Mangoldt Function

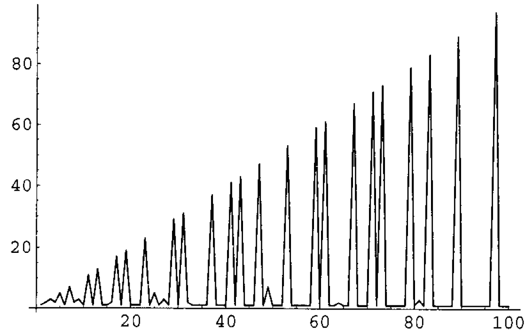

The function defined by

$$\Lambda(n) \equiv \begin{cases} \ln p & \text{if } n = p^k \text{ for } p \text{ a prime} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

$\exp(\Lambda(n))$ is also given by $[1, 2, \ldots, n]/[1, 2, \ldots, n-1]$, where $[a, b, c, \ldots]$ denotes the LEAST COMMON MULTIPLE. The first few values of $\exp(\Lambda(n))$ for $n = 1, 2, \ldots$, plotted above, are 1, 2, 3, 2, 5, 1, 7, 2, ... (Sloane's A014963). The Mangoldt function is related to the RIEMANN ZETA FUNCTION $\zeta(z)$ by

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s}, \tag{2}$$

where $\Re[s] > 1$.

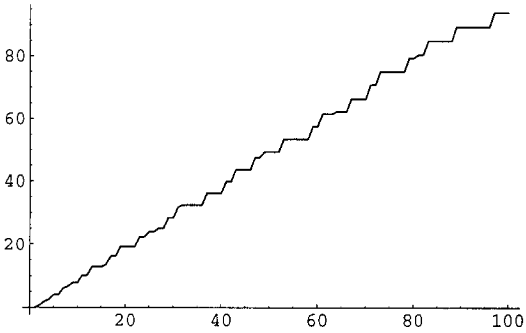

The SUMMATORY Mangoldt function, illustrated above, is defined by

$$\psi(x) \equiv \sum_{n \leq x} \Lambda(n), \tag{3}$$

where $\Lambda(n)$ is the MANGOLDT FUNCTION. This has the explicit formula

$$\psi(x) = x - \sum_{\rho} \frac{x^\rho}{\rho} - \ln(2\pi) - \tfrac{1}{2}\ln(1 - x^2), \tag{4}$$

where the second SUM is over all complex zeros ρ of the RIEMANN ZETA FUNCTION $\zeta(s)$ and interpreted as

$$\lim_{t \to \infty} \sum_{|\Im(\rho)| < t} \frac{x^\rho}{\rho}. \tag{5}$$

Vardi (1991, p. 155) also gives the interesting formula

$$\ln([x]!) = \psi(x) + \psi(\tfrac{1}{2}x) + \psi(\tfrac{1}{3}x) + \ldots, \tag{6}$$

where $[x]$ is the NINT function and $n!$ is a FACTORIAL.

Vallée Poussin's version of the PRIME NUMBER THEOREM states that

$$\psi(x) = x + \mathcal{O}(xe^{-a\sqrt{\ln x}}) \tag{7}$$

for some a (Davenport 1980, Vardi 1991). The RIEMANN HYPOTHESIS is equivalent to

$$\psi(x) = x + \mathcal{O}(\sqrt{x}\,(\ln x)^2) \tag{8}$$

(Davenport 1980, p. 114; Vardi 1991).

see also BOMBIERI'S THEOREM, GREATEST PRIME FACTOR, LAMBDA FUNCTION, LEAST COMMON MULTIPLE, LEAST PRIME FACTOR, RIEMANN FUNCTION

References

Davenport, H. *Multiplicative Number Theory, 2nd ed.* New York: Springer-Verlag, p. 110, 1980.

Sloane, N. J. A. Sequence A014963 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 146–147, 152–153, and 249, 1991.

Manifold

Rigorously, an n-D (topological) manifold is a Topological Space M such that any point in M has a Neighborhood $U \subset M$ which is Homeomorphic to n-D Euclidean Space. The Homeomorphism is called a chart, since it lays that part of the manifold out flat, like charts of regions of the Earth. So a preferable statement is that any object which can be "charted" is a manifold.

The most important manifolds are Differentiable Manifolds. These are manifolds where overlapping charts "relate smoothly" to each other, meaning that the inverse of one followed by the other is an infinitely differentiable map from Euclidean Space to itself.

Manifolds arise naturally in a variety of mathematical and physical applications as "global objects." For example, in order to precisely describe all the configurations of a robot arm or all the possible positions and momenta of a rocket, an object is needed to store all of these parameters. The objects that crop up are manifolds. From the geometric perspective, manifolds represent the profound idea having to do with global versus local properties.

Consider the ancient belief that the Earth was flat compared to the modern evidence that it is round. The discrepancy arises essentially from the fact that on the small scales that we see, the Earth does look flat. We cannot see it curve because we are too small (although the Greeks did notice that the last part of a ship to disappear over the horizon was the mast). We can detect curvature only indirectly from our vantage point on the Earth. The basic idea for this "problem" was codified by Poincaré. The problem is that on a small scale, the Earth is nearly flat. In general, any object which is nearly "flat" on small scales is a manifold, and so manifolds constitute a generalization of objects we could live on in which we would encounter the round/flat Earth problem.

see also Cobordant Manifold, Compact Manifold, Connected Sum Decomposition, Differentiable Manifold, Flag Manifold, Grassmann Manifold, Heegaard Splitting, Isospectral Manifolds, Jaco-Shalen-Johannson Torus Decomposition, Kähler Manifold, Poincaré Conjecture, Poisson Manifold, Prime Manifold, Riemannian Manifold, Set, Smooth Manifold, Space, Stiefel Manifold, Stratified Manifold, Submanifold, Surgery, Symplectic Manifold, Thurston's Geometrization Conjecture, Topological Manifold, Topological Space, Whitehead Manifold, Wiedersehen Manifold

References

Conlon, L. *Differentiable Manifolds: A First Course.* Boston, MA: Birkhäuser, 1993.

Mantissa

For a Real Number x, the mantissa is defined as the Positive fractional part $x - \lfloor x \rfloor = \texttt{frac(x)}$, where $\lfloor x \rfloor$ denotes the Floor Function.

see also Characteristic (Real Number), Floor Function, Scientific Notation

Map

A way of associating unique objects to every point in a given Set. So a map from $A \mapsto B$ is an object f such that for every $a \in A$, there is a unique object $f(a) \in B$. The terms Function and Mapping are synonymous with map.

The following table gives several common types of complex maps.

Mapping	Formula	Domain
inversion	$f(z) = \frac{1}{z}$	
magnification	$f(z) = az$	$a \in \mathbb{R} \neq 0$
magnification+rotation	$f(z) = az$	$a \in \mathbb{C} \neq 0$
Möbius	$f(z) = \frac{az+b}{cz+d}$	$a, b, c, d \in \mathbb{C}$
rotation	$f(z) = e^{i\theta z}$	
translation	$f(z) = z + a$	$a \in \mathbb{C}$

see also $2x$ mod 1 Map, Arnold's Cat Map, Baker's Map, Boundary Map, Conformal Map, Function, Gauss Map, Gingerbreadman Map, Harmonic Map, Hénon Map, Identity Map, Inclusion Map, Kaplan-Yorke Map, Logistic Map, Mandelbrot Set, Map Projection, Pullback Map, Quadratic Map, Tangent Map, Tent Map, Transformation, Zaslavskii Map

References

Arfken, G. "Mapping." §6.6 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 384–392, 1985.

Lee, X. "Transformation of the Plane." `http://www.best.com/~xah/MathGraphicsGallery_dir/Transform2DPlot_dir/transform2DPlot.html`.

Map Coloring

Given a map with Genus $g > 0$, Heawood showed in 1890 that the maximum number N_u of colors necessary to color a map (the Chromatic Number) on an unbounded surface is

$$N_u \equiv \left\lfloor \tfrac{1}{2}(7 + \sqrt{48g + 1}\,) \right\rfloor = \left\lfloor \tfrac{1}{2}(7 + \sqrt{49 - 24\chi}\,) \right\rfloor,$$

where $\lfloor x \rfloor$ is the Floor Function, g is the Genus, and χ is the Euler Characteristic. This is the Heawood Conjecture. In 1968, for any orientable surface other than the Sphere (or equivalently, the Plane) and any nonorientable surface other than the Klein Bottle, N_u was shown to be not merely a maximum, but the actual number needed (Ringel and Youngs 1968).

When the Four-Color Theorem was proven, the Heawood Formula was shown to hold also for all orientable and nonorientable surfaces with the exception of the

Klein Bottle. For this case (which has Euler Characteristic 1, and therefore can be considered to have $g = 1/2$), the actual number of colors N needed is six—*one less than* $N_u = 7$ (Franklin 1934; Saaty 1986, p. 45).

surface	g	N_u	N
Klein bottle	1	7	6
Möbius strip	$\frac{1}{2}$	6	6
plane	0	4	4
projective plane	$\frac{1}{2}$	6	6
sphere	0	4	4
torus	1	7	7

see also Chromatic Number, Four-Color Theorem, Heawood Conjecture, Six-Color Theorem, Torus Coloring

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 237–238, 1987.

Barnette, D. *Map Coloring, Polyhedra, and the Four-Color Problem.* Washington, DC: Math. Assoc. Amer., 1983.

Franklin, P. "A Six Colour Problem." *J. Math. Phys.* **13**, 363–369, 1934.

Franklin, P. *The Four-Color Problem.* New York: Scripta Mathematica, Yeshiva College, 1941.

Ore, Ø. *The Four-Color Problem.* New York: Academic Press, 1967.

Ringel, G. and Youngs, J. W. T. "Solution of the Heawood Map-Coloring Problem." *Proc. Nat. Acad. Sci. USA* **60**, 438–445, 1968.

Saaty, T. L. and Kainen, P. C. *The Four-Color Problem: Assaults and Conquest.* New York: Dover, 1986.

Map Folding

A general Formula giving the number of distinct ways of folding an $N = m \times n$ rectangular map is not known. A distinct folding is defined as a permutation of N numbered cells reading from the top down. Lunnon (1971) gives values up to $n = 28$.

n	$1 \times n$	$2 \times n$	$3 \times n$	$4 \times n$	$5 \times n$
1	1	1			
2	2	8			
3	6	60	1368		
4	16	1980		300608	
5	59	19512			18698669
6	144	15552			

The limiting ratio of the number of $1 \times (n+1)$ strips to the number of $1 \times n$ strips is given by

$$\lim_{n\to\infty} \frac{[1 \times (n+1)]}{[1 \times n]} \in [3.3868, 3.9821].$$

see also Stamp Folding

References

Gardner, M. "The Combinatorics of Paper Folding." Ch. 7 in *Wheels, Life, and Other Mathematical Amusements.* New York: W. H. Freeman, 1983.

Koehler, J. E. "Folding a Strip of Stamps." *J. Combin. Th.* **5**, 135–152, 1968.

Lunnon, W. F. "A Map-Folding Problem." *Math. Comput.* **22**, 193–199, 1968.

Lunnon, W. F. "Multi-Dimensional Strip Folding." *Computer J.* **14**, 75–79, 1971.

Map Projection

A projection which maps a Sphere (or Spheroid) onto a Plane. No projection can be simultaneously Conformal and Area-Preserving.

see also Airy Projection, Albers Equal-Area Conic Projection, Axonometry, Azimuthal Equidistant Projection, Azimuthal Projection, Behrmann Cylindrical Equal-Area Projection, Bonne Projection, Cassini Projection, Chromatic Number, Conic Equidistant Projection, Conic Projection, Cylindrical Equal-Area Projection, Cylindrical Equidistant Projection, Cylindrical Projection, Eckert IV Projection, Eckert VI Projection, Four-Color Theorem, Gnomic Projection, Guthrie's Problem, Hammer-Aitoff Equal-Area Projection, Lambert Azimuthal Equal-Area Projection, Lambert Conformal Conic Projection, Map Coloring, Mercator Projection, Miller Cylindrical Projection, Mollweide Projection, Orthographic Projection, Polyconic Projection, Pseudocylindrical Projection, Rectangular Projection, Sinusoidal Projection, Six-Color Theorem, Stereographic Projection, van der Grinten Projection, Vertical Perspective Projection

References

Dana, P. H. "Map Projections." `http://www.utexas.edu/depts/grg/gcraft/notes/mapproj/mapproj.html`.

Hunter College Geography. "The Map Projection Home Page." `http://everest.hunter.cuny.edu/mp/`.

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, 1987.

Mapes' Method

A method for computing the Prime Counting Function. Define the function

$$T_k(x,a) = (-1)^{\beta_0+\beta_1+\ldots+\beta_{a-1}} \left\lfloor \frac{x}{{p_1}^{\beta_0}{p_2}^{\beta_1}\cdots{p_a}^{\beta_{a-1}}} \right\rfloor, \tag{1}$$

where $\lfloor x \rfloor$ is the Floor Function and the β_i are the binary digits (0 or 1) in

$$k = 2^{a-1}\beta_{a-1} + 2^{a-2}\beta_{a-2} + \ldots + 2^1\beta_1 + 2^0\beta_0. \tag{2}$$

The Legendre Sum can then be written

$$\phi(x,a) = \sum_{k=0}^{2^a-1} T_k(x,a). \tag{3}$$

The first few values of $T_k(x,3)$ are

$$T_0(x,3) = \lfloor x \rfloor \tag{4}$$

$$T_1(x,3) = -\left\lfloor \frac{x}{p_1} \right\rfloor \tag{5}$$

$$T_2(x,3) = -\left\lfloor \frac{x}{p_2} \right\rfloor \tag{6}$$

$$T_3(x,3) = \left\lfloor \frac{x}{p_1 p_2} \right\rfloor \tag{7}$$

$$T_4(x,3) = -\left\lfloor \frac{x}{p_3} \right\rfloor \tag{8}$$

$$T_5(x,3) = \left\lfloor \frac{x}{p_1 p_3} \right\rfloor \tag{9}$$

$$T_6(x,3) = \left\lfloor \frac{x}{p_2 p_3} \right\rfloor \tag{10}$$

$$T_7(x,3) = -\left\lfloor \frac{x}{p_1 p_2 p_3} \right\rfloor. \tag{11}$$

Mapes' method takes time $\sim x^{0.7}$, which is slightly faster than the LEHMER-SCHUR METHOD.

see also LEHMER-SCHUR METHOD, PRIME COUNTING FUNCTION

References

Mapes, D. C. "Fast Method for Computing the Number of Primes Less than a Given Limit." *Math. Comput.* **17**, 179–185, 1963.

Riesel, H. "Mapes' Method." *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, p. 23, 1994.

Mapping (Function)

see MAP

Mapping Space

Let Y^X be the set of continuous mappings $f : X \to Y$. Then the TOPOLOGICAL SPACE for Y^X supplied with a compact-open topology is called a mapping space.

see also LOOP SPACE

References

Iyanaga, S. and Kawada, Y. (Eds.). "Mapping Spaces." §204B in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 658, 1980.

Marginal Analysis

Let $R(x)$ be the revenue for a production x, $C(x)$ the cost, and $P(x)$ the profit. Then

$$P(x) = R(x) - C(x),$$

and the marginal profit for the x_0th unit is defined by

$$P'(x_0) = R'(x_0) - C'(x_0),$$

where $P'(x)$, $R'(x)$, and $C'(x)$ are the DERIVATIVES of $P(x)$, $R(x)$, and $C(x)$, respectively.

see also DERIVATIVE

Marginal Probability

Let S be partitioned into $r \times s$ disjoint sets E_i and F_j where the general subset is denoted $E_i \cap F_j$. Then the marginal probability of E_i is

$$P(E_i) = \sum_{j=1}^{s} P(E_i \cap F_j).$$

Markoff's Formulas

Formulas obtained from differentiating NEWTON'S FORWARD DIFFERENCE FORMULA,

$$f'(a_0 + ph) = \frac{1}{h}[\Delta_0 + \tfrac{1}{2}(2p-1)\Delta_0^2 + \tfrac{1}{6}(3p^2 - 6p + 2)\Delta_0^3 + \ldots + \frac{d}{dp}\binom{p}{n}\Delta_0^n\biggr] + R_n',$$

where

$$R_n' = h^n f^{(n+1)}(\xi)\frac{d}{dp}\binom{p}{n+1} + h^{n+1}\binom{p}{n+1}\frac{d}{dx}f^{(n+1)}(\xi),$$

$\binom{n}{k}$ is a BINOMIAL COEFFICIENT, and $a_0 < \xi < a_n$. Abramowitz and Stegun (1972) and Beyer (1987) give derivatives $h^n f_0^{(n)}$ in terms of Δ^k and derivatives in terms of δ^k and ∇^k.

see also FINITE DIFFERENCE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 883, 1972.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 449–450, 1987.

Markoff Number

see MARKOV NUMBER

Markov Algorithm

An ALGORITHM which constructs allowed mathematical statements from simple ingredients.

Markov Chain

A collection of random variables $\{X_t\}$, where the index t runs through 0, 1, ...

References

Kemeny, J. G. and Snell, J. L. *Finite Markov Chains.* New York: Springer-Verlag, 1976.

Stewart, W. J. *Introduction to the Numerical Solution of Markov Chains.* Princeton, NJ: Princeton University Press, 1995.

Markov's Inequality

If x takes only NONNEGATIVE values, then

$$P(x \geq a) \leq \frac{\langle x \rangle}{a}.$$

To prove the theorem, write

$$\langle x \rangle = \int_0^\infty x f(x)\,dx = \int_0^a x f(x)\,dx + \int_a^\infty x f(x)\,dx.$$

Since $P(x)$ is a probability density, it must be ≥ 0. We have stipulated that $x \geq 0$, so

$$\langle x \rangle \geq \int^a x f(x)\,dx + \int^\infty x f(x)\,dx$$

$$= a \int_a^\infty f(x)\,dx = aP(x \geq a),$$

Q. E. D.

Markov Matrix

see STOCHASTIC MATRIX

Markov Moves

A type I move (CONJUGATION) takes $AB \to BA$ for A, $B \in B_n$ where B_n is a BRAID GROUP.

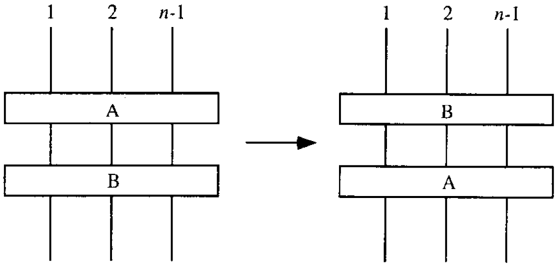

A type II move (STABILIZATION) takes $A \to Ab_n$ or $A \to Ab_n{}^{-1}$ for $A \in B_n$, and b_n, Ab_n, and $Ab_n{}^{-1} \in B_{n+1}$.

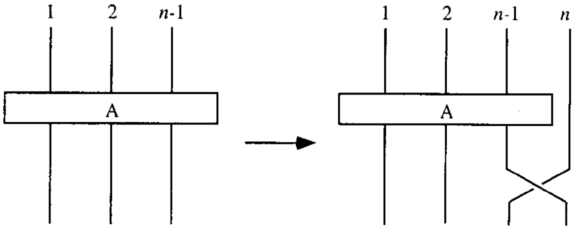

see also BRAID GROUP, CONJUGATION, REIDEMEISTER MOVES, STABILIZATION

Markov Number

The Markov numbers m occur in solutions to the DIOPHANTINE EQUATION

$$x^2 + y^2 + z^2 = 3xyz,$$

and are related to LAGRANGE NUMBERS L_n by

$$L_n = \sqrt{9 - \frac{4}{m^2}}.$$

The first few solutions are $(x, y, z) = (1, 1, 1)$, (1, 1, 2), (1, 2, 5), (1, 5, 13), (2, 5, 29), ... The solutions can be arranged in an infinite tree with two smaller branches on each trunk. It is not known if two different regions can have the same label. Strangely, the regions adjacent to 1 have alternate FIBONACCI NUMBERS 1, 2, 5, 13, 34,

N, then

$$M(n) = C(\ln N) + \mathcal{O}((\ln N)^{1+\epsilon}),$$

where $C \approx 0.180717105$ (Guy 1994, p. 166).

see also HURWITZ EQUATION, HURWITZ'S IRRATIONAL NUMBER THEOREM, LAGRANGE NUMBER (RATIONAL APPROXIMATION) LIOUVILLE'S RATIONAL APPROXIMATION THEOREM, LIOUVILLE-ROTH CONSTANT, ROTH'S THEOREM, SEGRE'S THEOREM, THUE-SIEGEL-ROTH THEOREM

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 187–189, 1996.

Guy, R. K. "Markoff Numbers." §D12 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 166–168, 1994.

Markov Process

A random process whose future probabilities are determined by its most recent values.

see also DOOB'S THEOREM

Markov Spectrum

A SPECTRUM containing the REAL NUMBERS larger than FREIMAN'S CONSTANT.

see also FREIMAN'S CONSTANT, SPECTRUM SEQUENCE

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 188–189, 1996.

Markov's Theorem

Published by A. A. Markov in 1935, Markov's theorem states that equivalent BRAIDS expressing the same LINK are mutually related by successive applications of two types of MARKOV MOVES. Markov's theorem is difficult to apply in practice, so it is difficult to establish the equivalence or nonequivalence of LINKS having different BRAID representations.

see also BRAID, LINK, MARKOV MOVES

Marriage Theorem

If a group of men and women may date only if they have previously been introduced, then a complete set of dates is possible IFF every subset of men has collectively been introduced to at least as many women, and vice versa.

References
Chartrand, G. *Introductory Graph Theory.* New York: Dover, p. 121, 1985.

Married Couples Problem

Also called the MÉNAGE PROBLEM. In how many ways can n married couples be seated around a circular table in such a manner than there is always one man between two women and none of the men is next to his own wife? The solution (Ball and Coxeter 1987, p. 50) uses DISCORDANT PERMUTATIONS and can be given in terms of LAISANT'S RECURRENCE FORMULA

$$(n-1)A_{n+1}=(n^2-1)A_n+(n+1)A_{n-1}+4(-1)^n,$$

with $A_1 = A_2 = 1$. A closed form expression due to Touchard (1934) is

$$A_n=\sum_{k=0}^{n}\frac{2n}{2n-k}\binom{2n-k}{k}(n-k)!(-1)^k,$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT (Vardi 1991).

The first few values of A_n are -1, 1, 0, 2, 13, 80, 579, ... (Sloane's A000179), which are sometimes called MÉNAGE NUMBERS. The desired solution is then $2n!A_n$ The numbers A_n can be considered a special case of a restricted ROOKS PROBLEM.

see also DISCORDANT PERMUTATION, LAISANT'S RECURRENCE FORMULA, ROOKS PROBLEM

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 50, 1987.
Dörrie, H. §8 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 27–33, 1965.
Halmos, P. R.; Vaughan, H. E. "The Marriage Problem." *Amer. J. Math.* **72**, 214–215, 1950.
Lucas, E. *Théorie des Nombres.* Paris, pp. 215 and 491–495, 1891.
MacMahon, P. A. *Combinatory Analysis, Vol. 1.* London: Cambridge University Press, pp. 253–256, 1915.
Newman, D. J. "A Problem in Graph Theory." *Amer. Math. Monthly* **65**, 611, 1958.
Sloane, N. J. A. Sequence A000179/M2062 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Touchard, J. "Sur un probléme de permutations." *C. R. Acad. Sci. Paris* **198**, 631–633, 1934.
Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 123, 1991.

Marshall-Edgeworth Index

The statistical INDEX

$$P_{ME}\equiv\frac{\sum p_n(q_0+q_n)}{\sum(v_0+v_n)},$$

where p_n is the price per unit in period n, q_n is the quantity produced in period n, and $v_n \equiv p_n q_n$ is the value of the n units.

see also INDEX

References
Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, pp. 66–67, 1962.

Martingale

A sequence of random variates such that the CONDITIONAL PROBABILITY of x_{n+1} given x_1, x_2, ..., x_n is x_n. The term was first used to describe a type of wagering in which the bet is doubled or halved after a loss or win, respectively.

see also GAMBLER'S RUIN, SAINT PETERSBURG PARADOX

Mascheroni Constant

see EULER-MASCHERONI CONSTANT

Mascheroni Construction

A geometric construction done with a movable COMPASS alone. All constructions possible with a COMPASS and STRAIGHTEDGE are possible with a *movable* COMPASS alone, as was proved by Mascheroni (1797). Mascheroni's results are now known to have been anticipated largely by Mohr (1672).

see also COMPASS, GEOMETRIC CONSTRUCTION, NEUSIS CONSTRUCTION, STRAIGHTEDGE

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 96–97, 1987.
Bogomolny, A. "Geometric Constructions with the Compass Alone." `http://www.cut-the-knot.com/do_you_know/compass.html`.
Courant, R. and Robbins, H. "Constructions with Other Tools. Mascheroni Constructions with Compass Alone." §3.5 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 146–158, 1996.
Dörrie, H. "Mascheroni's Compass Problem." §33 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 160–164, 1965.
Gardner, M. "Mascheroni Constructions." Ch. 17 in *Mathematical Circus: More Puzzles, Games, Paradoxes and Other Mathematical Entertainments from Scientific American.* New York: Knopf, pp. 216–231, 1979.
Mascheroni, L. *Geometry of Compass.* Pavia, Italy, 1797.
Mohr, G. *Euclides Danicus.* Amsterdam, Netherlands, 1672.

Maschke's Theorem

If a MATRIX GROUP is reducible, then it is completely reducible, i.e., if the MATRIX GROUP is equivalent to the MATRIX GROUP in which every MATRIX has the reduced form

$$\begin{bmatrix} D_i^{(1)} & X_i \\ 0 & D_i^{(2)} \end{bmatrix},$$

then it is equivalent to the MATRIX GROUP obtained by putting $X_i = 0$.

see also MATRIX GROUP

References

Lomont, J. S. *Applications of Finite Groups.* New York: Dover, p. 49, 1987.

Mason's abc Theorem

see MASON'S THEOREM

Mason's Theorem

Let there be three POLYNOMIALS $a(x)$, $b(x)$, and $c(x)$ with no common factors such that

$$a(x) + b(x) = c(x).$$

Then the number of distinct ROOTS of the three POLYNOMIALS is one or more greater than their largest degree. The theorem was first proved by Stothers (1981).

Mason's theorem may be viewed as a very special case of a Wronskian estimate (Chudnovsky and Chudnovsky 1984). The corresponding Wronskian identity in the proof by Lang (1993) is

$$c^3 * W(a, b, c) = W(W(a, c), W(b, c)),$$

so if a, b, and c are linearly dependent, then so are $W(a, c)$ and $W(b, c)$. More powerful Wronskian estimates with applications toward diophantine approximation of solutions of linear differential equations may be found in Chudnovsky and Chudnovsky (1984) and Osgood (1985).

The rational function case of FERMAT'S LAST THEOREM follows trivially from Mason's theorem (Lang 1993, p. 195).

see also ABC CONJECTURE

References

Chudnovsky, D. V. and Chudnovsky, G. V. "The Wronskian Formalism for Linear Differential Equations and Padé Approximations." *Adv. Math.* **53**, 28–54, 1984.

Lang, S. "Old and New Conjectured Diophantine Inequalities." *Bull. Amer. Math. Soc.* **23**, 37–75, 1990.

Lang, S. *Algebra, 3rd ed.* Reading, MA: Addison-Wesley, 1993.

Osgood, C. F. "Sometimes Effective Thue-Siegel-Roth-Schmidt-Nevanlinna Bounds, or Better." *J. Number Th.* **21**, 347–389, 1985.

Stothers, W. W. "Polynomial Identities and Hauptmodulen." *Quart. J. Math. Oxford Ser. II* **32**, 349–370, 1981.

Masser-Gramain Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

If $f(z)$ is an ENTIRE FUNCTION such that $f(n)$ is an INTEGER for each POSITIVE INTEGER n. Pólya (1915) showed that if

$$\limsup_{r\to\infty} \frac{\ln M_r}{r} < \ln 2 = 0.693\ldots, \qquad (1)$$

where

$$M_r = \sup_{|z|\leq r} |f(x)| \qquad (2)$$

is the SUPREMUM, then f is a POLYNOMIAL. Furthermore, $\ln 2$ is the best constant (i.e., counterexamples exist for every smaller value).

If $f(z)$ is an ENTIRE FUNCTION with $f(n)$ a GAUSSIAN INTEGER for each GAUSSIAN INTEGER n, then Gelfond (1929) proved that there exists a constant α such that

$$\limsup_{r\to\infty} \frac{\ln M_r}{r^2} < \alpha \qquad (3)$$

implies that f is a POLYNOMIAL. Gramain (1981, 1982) showed that the best such constant is

$$\alpha = \frac{\pi}{2e} = 0.578\ldots. \qquad (4)$$

Maser (1980) proved the weaker result that f must be a POLYNOMIAL if

$$\limsup_{r\to\infty} \frac{\ln M_r}{r^2} < \alpha_0 = \tfrac{1}{2}\exp\left(-\delta + \frac{4c}{\pi}\right), \qquad (5)$$

where

$$c = \gamma\beta(1) + \beta'(1) = 0.6462454398948114\ldots, \qquad (6)$$

γ is the EULER-MASCHERONI CONSTANT, $\beta(z)$ is the DIRICHLET BETA FUNCTION,

$$\delta \equiv \lim_{n\to\infty}\left(\sum_{k=2}^{\infty} \frac{1}{\pi {r_k}^2} - \ln n\right), \qquad (7)$$

and r_k is the minimum NONNEGATIVE r for which there exists a COMPLEX NUMBER z for which the CLOSED DISK with center z and radius r contains at least k distinct GAUSSIAN INTEGER. Gosper gave

$$c = \pi\{-\ln[\Gamma(\tfrac{1}{4})] + \tfrac{3}{4}\pi + \tfrac{1}{2}\ln 2 + \tfrac{1}{2}\gamma\}. \qquad (8)$$

Gramain and Weber (1985, 1987) have obtained

$$1.811447299 < \delta < 1.897327117, \qquad (9)$$

which implies

$$0.1707339 < \alpha_0 < 0.1860446. \qquad (10)$$

Gramain (1981, 1982) conjectured that

$$\alpha_0 = \frac{1}{2e}, \tag{11}$$

which would imply

$$\delta = 1 + \frac{4c}{\pi} = 1.822825249\ldots. \tag{12}$$

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/masser/masser.html`.

Gramain, F. "Sur le théorème de Fukagawa-Gel'fond." *Invent. Math.* **63**, 495–506, 1981.

Gramain, F. "Sur le théorème de Fukagawa-Gel'fond-Gruman-Masser." *Séminaire Delange-Pisot-Poitou (Théorie des Nombres), 1980–1981.* Boston, MA: Birkhäuser, 1982.

Gramain, F. and Weber, M. "Computing and Arithmetic Constant Related to the Ring of Gaussian Integers." *Math. Comput.* **44**, 241–245, 1985.

Gramain, F. and Weber, M. "Computing and Arithmetic Constant Related to the Ring of Gaussian Integers." *Math. Comput.* **48**, 854, 1987.

Masser, D. W. "Sur les fonctions entières à valeurs entières." *C. R. Acad. Sci. Paris Sér. A-B* **291**, A1–A4, 1980.

Match Problem

Given n matches, find the number of topologically distinct planar arrangements $T(n)$ which can be made. The first few values are 1, 1, 3, 5, 10, 19, 39, ... (Sloane's A003055).

see also CIGARETTES, MATCHSTICK GRAPH

References

Gardner, M. "The Problem of the Six Matches." In *The Unexpected Hanging and Other Mathematical Diversions.* Chicago, IL: Chicago University Press, pp. 79–81, 1991.

Sloane, N. J. A. Sequence A003055/M2464 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Matchstick Graph

A PLANAR GRAPH whose EDGES are all unit line segments. The minimal number of EDGES for matchstick graphs of various degrees are given in the table below. The minimal degree 1 matchstick graph is a single EDGE, and the minimal degree 2 graph is an EQUILATERAL TRIANGLE.

n	e	v
1	1	2
2	3	3
3	12	8
4		≤ 42

Mathematical Induction

see INDUCTION

Mathematics

Mathematics is a broad-ranging field of study in which the properties and interactions of idealized objects are examined. Whereas mathematics began merely as a calculational tool for computation and tabulation of quantities, it has blossomed into an extremely rich and diverse set of tools, terminologies, and approaches which range from the purely abstract to the utilitarian.

Bertrand Russell once whimsically defined mathematics as, "The subject in which we never know what we are talking about nor whether what we are saying is true" (Bergamini 1969).

The term "mathematics" is often shortened to "math" in informal American speech and, consistent with the British penchant for adding superfluous letters, "maths" in British English.

see also METAMATHEMATICS

References

Bergamini, D. *Mathematics.* New York: Time-Life Books, p. 9, 1969.

Mathematics Prizes

Several prizes are awarded periodically for outstanding mathematical achievement. There is no Nobel Prize in mathematics, and the most prestigious mathematical award is known as the FIELDS MEDAL. In rough order of importance, other awards are the $100,000 Wolf Prize of the Wolf Foundation of Israel, the Leroy P. Steele Prize of the American Mathematical Society, followed by the Bôcher Memorial Prize, Frank Nelson Cole Prizes in Algebra and Number Theory, and the Delbert Ray Fulkerson Prize, all presented by the American Mathematical Society.

see also FIELDS MEDAL

References

"AMS Funds and Prizes." `http://www.ams.org/ams/prizes.html`.

MacTutor History of Mathematics Archives. "The Fields Medal." `http://www-groups.dcs.st-and.ac.uk/~history/Societies/FieldsMedal.html`. "Winners of the Bôcher Prize of the AMS." `http://www-groups.dcs.st-and.ac.uk/~history/Societies/AMSBocherPrize.html`. "Winners of the Frank Nelson Cole Prize of the AMS." `http://www-groups.dcs.st-and.ac.uk/~history/Societies/AMSColePrize.html`.

MacTutor History of Mathematics Archives. "Mathematical Societies, Medals, Prizes, and Other Honours." `http://www-groups.dcs.st-and.ac.uk/~history/Societies/`.

Monastyrsky, M. *Modern Mathematics in the Light of the Fields Medals.* Wellesley, MA: A. K. Peters, 1997.

"Wolf Prize Recipients in Mathematics." `http://www.aquanet.co.il/wolf/wolf5.html`.

Mathieu Differential Equation

$$\frac{d^2V}{dv^2} + [b - 2q\cos(2v)]V = 0.$$

It arises in separation of variables of LAPLACE'S EQUATION in ELLIPTIC CYLINDRICAL COORDINATES. Whittaker and Watson (1990) use a slightly different form to define the MATHIEU FUNCTIONS.

The modified Mathieu differential equation

$$\frac{d^2U}{du^2} - [b - 2q\cosh(2u)]U = 0$$

arises in SEPARATION OF VARIABLES of the HELMHOLTZ DIFFERENTIAL EQUATION in ELLIPTIC CYLINDRICAL COORDINATES.

see also MATHIEU FUNCTION

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 722, 1972.
Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 556–557, 1953.
Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Mathieu Function

The form given by Whittaker and Watson (1990, p. 405) defines the Mathieu function based on the equation

$$\frac{d^2u}{dz^2} + [a + 16q\cos(2z)]u = 0. \tag{1}$$

This equation is closely related to HILL'S DIFFERENTIAL EQUATION. For an EVEN Mathieu function,

$$G(\eta) = \lambda \int_{-\pi}^{\pi} e^{k\cos\eta\cos\theta} G(\theta)\,d\theta, \tag{2}$$

where $k \equiv \sqrt{32q}$. For an ODD Mathieu function,

$$G(\eta) = \lambda \int_{-\pi}^{\pi} \sin(k\sin\eta\sin\theta) G(\theta)\,d\theta. \tag{3}$$

Both EVEN and ODD functions satisfy

$$G(\eta) = \lambda \int_{-\pi}^{\pi} e^{ik\sin\eta\sin\theta} G(\theta)\,d\theta. \tag{4}$$

Letting $\zeta \equiv \cos^2 z$ transforms the MATHIEU DIFFERENTIAL EQUATION to

$$4\zeta(1-\zeta)\frac{d^2u}{d\zeta^2} + 2(1-2\zeta)\frac{du}{d\zeta} + (a - 16q + 32q\zeta)u = 0. \tag{5}$$

see also MATHIEU DIFFERENTIAL EQUATION

References
Abramowitz, M. and Stegun, C. A. (Eds.). "Mathieu Functions." Ch. 20 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 721–746, 1972.
Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 562–568 and 633–642, 1953.
Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Mathieu Groups

The first SIMPLE SPORADIC GROUPS discovered. M_{11}, M_{12}, M_{22}, M_{23}, M_{24} were discovered in 1861 and 1873 by Mathieu. Frobenius showed that all the Mathieu groups are SUBGROUPS of M_{24}.

The Mathieu groups are most simply defined as AUTOMORPHISM groups of STEINER SYSTEMS. M_{11} corresponds to $S(4,5,11)$ and M_{23} corresponds to $S(4,7,23)$. M_{11} and M_{23} are TRANSITIVE PERMUTATION GROUPS of 11 and 23 elements.

The ORDERS of the Mathieu groups are

$$\begin{aligned}|M_{11}| &= 2^4 \cdot 3^2 \cdot 5 \cdot 11\\ |M_{12}| &= 2^6 \cdot 3^3 \cdot 5 \cdot 11\\ |M_{22}| &= 2^7 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11\\ |M_{23}| &= 2^7 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11 \cdot 23.\end{aligned}$$

see also SPORADIC GROUP

References
Conway, J. H. and Sloane, N. J. A. "The Golay Codes and the Mathieu Groups." Ch. 11 in *Sphere Packings, Lattices, and Groups, 2nd ed.* New York: Springer-Verlag, pp. 299–330, 1993.
Rotman, J. J. Ch. 9 in *An Introduction to the Theory of Groups, 4th ed.* New York: Springer-Verlag, 1995.
Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas/M11.html`, `M12.html`, `M22.html`, `M23.html`, and `M24.html`.

Matrix

The TRANSFORMATION given by the system of equations

$$\begin{aligned}x_1' &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n\\ x_2' &= a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n\\ &\vdots\\ x_m' &= a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n\end{aligned}$$

is denoted by the MATRIX EQUATION

$$\begin{bmatrix}x_1'\\ x_2'\\ \vdots\\ x_n'\end{bmatrix} = \begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix} \begin{bmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix}.$$

In concise notation, this could be written

$$\mathbf{x}' = \mathsf{A}\mathbf{x},$$

where $\mathbf{x}'$ and $\mathbf{x}$ are VECTORS and A is called an $n \times m$ matrix. A matrix is said to be SQUARE if $m = n$. Special types of SQUARE MATRICES include the IDENTITY MATRIX I, with $a_{ij} = \delta_{ij}$ (where δ_{ij} is the KRONECKER DELTA) and the DIAGONAL MATRIX $a_{ij} = c_i\delta_{ij}$ (where c_i are a set of constants).

For every linear transformation there exists one and only one corresponding matrix. Conversely, every matrix corresponds to a unique linear transformation. The matrix is an important concept in mathematics, and was first formulated by Sylvester and Cayley.

Two matrices may be added (MATRIX ADDITION) or multiplied (MATRIX MULTIPLICATION) together to yield a new matrix. Other common operations on a single matrix are diagonalization, inversion (MATRIX INVERSE), and transposition (MATRIX TRANSPOSE). The DETERMINANT $\det(\mathsf{A})$ or $|\mathsf{A}|$ of a matrix A is an very important quantity which appears in many diverse applications. Matrices provide a concise notation which is extremely useful in a wide range of problems involving linear equations (e.g., LEAST SQUARES FITTING).

see also ADJACENCY MATRIX, ADJUGATE MATRIX, ANTISYMMETRIC MATRIX, BLOCK MATRIX, CARTAN MATRIX, CIRCULANT MATRIX, CONDITION NUMBER, CRAMER'S RULE, DETERMINANT, DIAGONAL MATRIX, DIRAC MATRICES, EIGENVECTOR, ELEMENTARY MATRIX, EQUIVALENT MATRIX, FOURIER MATRIX, GRAM MATRIX, HILBERT MATRIX, HYPERMATRIX, IDENTITY MATRIX, INCIDENCE MATRIX, IRREDUCIBLE MATRIX, KAC MATRIX, LU DECOMPOSITION, MARKOV MATRIX, MATRIX ADDITION, MATRIX DECOMPOSITION THEOREM, MATRIX INVERSE, MATRIX MULTIPLICATION, MCCOY'S THEOREM, MINIMAL MATRIX, NORMAL MATRIX, PAULI MATRICES, PERMUTATION MATRIX, POSITIVE DEFINITE MATRIX, RANDOM MATRIX, RATIONAL CANONICAL FORM, REDUCIBLE MATRIX, ROTH'S REMOVAL RULE, SHEAR MATRIX, SKEW SYMMETRIC MATRIX, SMITH NORMAL FORM, SPARSE MATRIX, SPECIAL MATRIX, SQUARE MATRIX, STOCHASTIC MATRIX, SUBMATRIX, SYMMETRIC MATRIX, TOURNAMENT MATRIX

References

Arfken, G. "Matrices." §4.2 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 176–191, 1985.

Matrix Addition

Denote the sum of two MATRICES A and B (of the same dimensions) by $\mathsf{C} = \mathsf{A}+\mathsf{B}$. The sum is defined by adding entries with the same indices

$$c_{ij} \equiv a_{ij} + b_{ij}$$

over all i and j. For example,

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{bmatrix}.$$

Matrix addition is therefore both COMMUTATIVE and ASSOCIATIVE.

see also MATRIX, MATRIX MULTIPLICATION

Matrix Decomposition Theorem

Let P be a MATRIX of EIGENVECTORS of a given MATRIX A and D a MATRIX of the corresponding EIGENVALUES. Then A can be written

$$\mathsf{A} = \mathsf{PDP}^{-1}, \tag{1}$$

where D is a DIAGONAL MATRIX and the columns of P are ORTHOGONAL VECTORS. If P is not a SQUARE MATRIX, then it cannot have a MATRIX INVERSE. However, if P is $m \times n$ (with $m > n$), then A can be written

$$\mathsf{A} = \mathsf{UDV}^{\mathrm{T}}, \tag{2}$$

where U and V are $n \times n$ SQUARE MATRICES with ORTHOGONAL columns,

$$\mathsf{U}^{\mathrm{T}}\mathsf{U} = \mathsf{V}^{\mathrm{T}} = \mathsf{I}. \tag{3}$$

Matrix Diagonalization

Diagonalizing a MATRIX is equivalent to finding the EIGENVECTORS and EIGENVALUES. The EIGENVALUES make up the entries of the diagonalized MATRIX, and the EIGENVECTORS make up the new set of axes corresponding to the DIAGONAL MATRIX.

see also DIAGONAL MATRIX, EIGENVALUE, EIGENVECTOR

References

Arfken, G. "Diagonalization of Matrices." §4.6 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 217–229, 1985.

Matrix Direct Product

see DIRECT PRODUCT (MATRIX)

Matrix Equality

Two MATRICES A and B are said to be equal IFF

$$a_{ij} \equiv b_{ij}$$

for all i, j. Therefore,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix},$$

while

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \neq \begin{bmatrix} 0 & 2 \\ 3 & 4 \end{bmatrix}.$$

Matrix Equation

Nonhomogeneous matrix equations of the form

$$\mathsf{A}\mathbf{x} = \mathbf{b} \tag{1}$$

can be solved by taking the MATRIX INVERSE to obtain

$$\mathbf{x} = \mathsf{A}^{-1}\mathbf{b}. \tag{2}$$

This equation will have a nontrivial solution IFF the DETERMINANT $\det(\mathsf{A}) \neq 0$. In general, more numerically stable techniques of solving the equation include GAUSSIAN ELIMINATION, LU DECOMPOSITION, or the SQUARE ROOT METHOD.

For a homogeneous $n \times n$ MATRIX equation

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3}$$

to be solved for the x_is, consider the DETERMINANT

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}. \tag{4}$$

Now multiply by x_1, which is equivalent to multiplying the first row (or any row) by x_1,

$$x_1 \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11}x_1 & a_{12} & \cdots & a_{1n} \\ a_{21}x_1 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}x_1 & a_{n2} & \cdots & a_{nn} \end{vmatrix}. \tag{5}$$

The value of the DETERMINANT is unchanged if multiples of columns are added to other columns. So add x_2 times column 2, ..., and x_n times column n to the first column to obtain

$$x_1 \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n & a_{12} & \cdots & a_{1n} \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n & a_{n2} & \cdots & a_{nn} \end{vmatrix}. \tag{6}$$

But from the original MATRIX, each of the entries in the first columns is zero since

$$a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n = 0, \tag{7}$$

so

$$\begin{vmatrix} 0 & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} \end{vmatrix} = 0. \tag{8}$$

Therefore, if there is an $x_1 \neq 0$ which is a solution, the DETERMINANT is zero. This is also true for x_2, ..., x_n, so the original homogeneous system has a nontrivial solution for all x_is only if the DETERMINANT is 0. This approach is the basis for CRAMER'S RULE.

Given a numerical solution to a matrix equation, the solution can be iteratively improved using the following technique. Assume that the numerically obtained solution to

$$\mathsf{A}\mathbf{x} = \mathbf{b} \tag{9}$$

is $\mathbf{x}_1 = \mathbf{x} + \delta\mathbf{x}_1$, where $\delta\mathbf{x}_1$ is an error term. The first solution therefore gives

$$\mathsf{A}(\mathbf{x} + \delta\mathbf{x}_1) = \mathbf{b} + \delta\mathbf{b} \tag{10}$$

$$\mathsf{A}\delta\mathbf{x}_1 = \delta\mathbf{b}, \tag{11}$$

where $\delta\mathbf{b}$ is found by solving (10)

$$\delta\mathbf{b} = \mathsf{A}\mathbf{x}_1 - \mathbf{b}. \tag{12}$$

Combining (11) and (12) then gives

$$\delta\mathbf{x}_1 = \mathsf{A}^{-1}\delta\mathbf{b} = \mathsf{A}^{-1}(\mathsf{A}\mathbf{x}_1 - \mathbf{b}) = \mathbf{x}_1 - \mathsf{A}^{-1}\mathbf{b}, \tag{13}$$

so the next iteration to obtain $\mathbf{x}$ accurately should be

$$\mathbf{x}_2 = \mathbf{x}_1 - \delta\mathbf{x}_1. \tag{14}$$

see also CRAMER'S RULE, GAUSSIAN ELIMINATION, LU DECOMPOSITION, MATRIX, MATRIX ADDITION, MATRIX INVERSE, MATRIX MULTIPLICATION, NORMAL EQUATION, SQUARE ROOT METHOD

Matrix Exponential

Given a SQUARE MATRIX A, the matrix exponential is defined by

$$\exp(\mathsf{A}) \equiv e^{\mathsf{A}} = \sum_{n=0}^{\infty} \frac{\mathsf{A}^n}{n!} = \mathsf{I} + \mathsf{A} + \frac{\mathsf{AA}}{2!} + \frac{\mathsf{AAA}}{3!} + \ldots,$$

where I is the IDENTITY MATRIX.

see also EXPONENTIAL FUNCTION, MATRIX

Matrix Group

A GROUP in which the elements are SQUARE MATRICES, the group multiplication law is MATRIX MULTIPLICATION, and the group inverse is simply the MATRIX INVERSE. Every matrix group is equivalent to a unitary matrix group (Lomont 1987, pp. 47–48).

see also MASCHKE'S THEOREM

References

Lomont, J. S. "Matrix Groups." §3.1 in *Applications of Finite Groups*. New York: Dover, pp. 46–52, 1987.

Matrix Inverse

A Matrix A has an inverse Iff the Determinant $|\mathsf{A}| \neq 0$. For a 2×2 Matrix

$$\mathsf{A} \equiv \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \tag{1}$$

the inverse is

$$\mathsf{A}^{-1} = \frac{1}{|\mathsf{A}|} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{2}$$

For a 3×3 Matrix,

$$\mathsf{A}^{-1} = \frac{1}{|\mathsf{A}|} \begin{bmatrix} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{13} & a_{12} \\ a_{33} & a_{32} \end{vmatrix} & \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \\ \begin{vmatrix} a_{23} & a_{21} \\ a_{33} & a_{31} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{13} & a_{11} \\ a_{23} & a_{21} \end{vmatrix} \\ \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} & \begin{vmatrix} a_{12} & a_{11} \\ a_{32} & a_{31} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{bmatrix}. \tag{3}$$

A general $n \times n$ matrix can be inverted using methods such as the Gauss-Jordan Elimination, Gaussian Elimination, or LU Decomposition.

The inverse of a Product AB of Matrices A and B can be expressed in terms of A^{-1} and B^{-1}. Let

$$\mathsf{C} \equiv \mathsf{AB}. \tag{4}$$

Then

$$\mathsf{B} = \mathsf{A}^{-1}\mathsf{AB} = \mathsf{A}^{-1}\mathsf{C} \tag{5}$$

and

$$\mathsf{A} = \mathsf{ABB}^{-1} = \mathsf{CB}^{-1}. \tag{6}$$

Therefore,

$$\mathsf{C} = \mathsf{AB} = (\mathsf{CB}^{-1})(\mathsf{A}^{-1}\mathsf{C}) = \mathsf{CB}^{-1}\mathsf{A}^{-1}\mathsf{C}, \tag{7}$$

so

$$\mathsf{CB}^{-1}\mathsf{A}^{-1} = \mathsf{I}, \tag{8}$$

where I is the Identity Matrix, and

$$\mathsf{B}^{-1}\mathsf{A}^{-1} = \mathsf{C}^{-1} = (\mathsf{AB})^{-1}. \tag{9}$$

see also Matrix, Matrix Addition, Matrix Multiplication, Moore-Penrose Generalized Matrix Inverse, Strassen Formulas

References

Ben-Israel, A. and Greville, T. N. E. *Generalized Inverses: Theory and Applications.* New York: Wiley, 1977.

Nash, J. C. *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation, 2nd ed.* Bristol, England: Adam Hilger, pp. 24–26, 1990.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Is Matrix Inversion an N^3 Process?" §2.11 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 95–98, 1992.

Matrix Multiplication

The product C of two Matrices A and B is defined by

$$c_{ik} = a_{ij} b_{jk}, \tag{1}$$

where j is summed over for all possible values of i and k. Therefore, in order for multiplication to be defined, the dimensions of the Matrices must satisfy

$$(n \times m)(m \times p) = (n \times p), \tag{2}$$

where $(a \times b)$ denotes a Matrix with a rows and b columns. Writing out the product explicitly,

$$\begin{aligned} &\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{np} \end{bmatrix} \\ &\quad = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mp} \end{bmatrix}, \end{aligned} \tag{3}$$

where

$$\begin{aligned} c_{11} &= a_{11}b_{11} + a_{12}b_{21} + \ldots + a_{1m}b_{m1} \\ c_{12} &= a_{11}b_{12} + a_{12}b_{22} + \ldots + a_{1m}b_{m2} \\ c_{1p} &= a_{11}b_{1p} + a_{12}b_{2p} + \ldots + a_{1m}b_{mp} \\ c_{21} &= a_{21}b_{11} + a_{22}b_{21} + \ldots + a_{2m}b_{m1} \\ c_{22} &= a_{21}b_{12} + a_{22}b_{22} + \ldots + a_{2m}b_{m2} \\ c_{2p} &= a_{21}b_{1p} + a_{22}b_{2p} + \ldots + a_{2m}b_{mp} \\ c_{n1} &= a_{n1}b_{11} + a_{n2}b_{21} + \ldots + a_{nm}b_{m1} \\ c_{n2} &= a_{n1}b_{12} + a_{n2}b_{22} + \ldots + a_{nm}b_{m2} \\ c_{np} &= a_{n1}b_{1p} + a_{n2}b_{2p} + \ldots + a_{nm}b_{mp}. \end{aligned}$$

Matrix Multiplication is Associative, as can be seen by taking

$$[(ab)c]_{ij} = (ab)_{ik}c_{kj} = (a_{il}b_{lk})c_{kj}. \tag{4}$$

Now, since a_{il}, b_{lk}, and c_{kj} are Scalars, use the Associativity of Scalar Multiplication to write

$$(a_{il}b_{lk})c_{kj} = a_{il}(b_{lk}c_{kj}) = a_{il}(bc)_{lj} = [a(bc)]_{ij}. \tag{5}$$

Since this is true for all i and j, it must be true that

$$[(ab)c]_{ij} = [a(bc)]_{ij}. \tag{6}$$

That is, Matrix multiplication is Associative. However, Matrix Multiplication is *not* Commutative unless A and B are Diagonal (and have the same dimensions).

The product of two BLOCK MATRICES is given by multiplying each block

$$\begin{bmatrix} o & o & & & & \\ o & o & & & & \\ & & o & & & \\ & & & o & o & o \\ & & & o & o & o \\ & & & o & o & o \end{bmatrix}\begin{bmatrix} x & x & & & & \\ x & x & & & & \\ & & x & & & \\ & & & x & x & x \\ & & & x & x & x \\ & & & x & x & x \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} o & o \\ o & o \end{bmatrix}\begin{bmatrix} x & x \\ x & x \end{bmatrix} & & \\ & [o][x] & \\ & & \begin{bmatrix} o & o & o \\ o & o & o \\ o & o & o \end{bmatrix}\begin{bmatrix} x & x & x \\ x & x & x \\ x & x & x \end{bmatrix} \end{bmatrix}. \tag{7}$$

see also MATRIX, MATRIX ADDITION, MATRIX INVERSE, STRASSEN FORMULAS

References
Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 178–179, 1985.

Matrix Norm

Given a SQUARE MATRIX A with COMPLEX (or REAL) entries, a MATRIX NORM $||\mathsf{A}||$ is a NONNEGATIVE number associated with A having the properties

1. $||\mathsf{A}|| > 0$ when $\mathsf{A} \neq 0$ and $||\mathsf{A}|| = 0$ IFF $\mathsf{A} = 0$,
2. $||k\mathsf{A}|| = |k|\,||\mathsf{A}||$ for any SCALAR k,
3. $||\mathsf{A}+\mathsf{B}|| \leq ||\mathsf{A}|| + ||\mathsf{B}||$,
4. $||\mathsf{AB}|| \leq ||\mathsf{A}||\,||\mathsf{B}||$.

For an $n \times n$ MATRIX A and an $n \times n$ UNITARY MATRIX U,

$$||\mathsf{AU}|| = ||\mathsf{UA}|| = ||\mathsf{A}||.$$

Let $\lambda_1, \ldots, \lambda_n$ be the EIGENVALUES of A, then

$$\frac{1}{||\mathsf{A}^{-1}||} \leq |\lambda| \leq ||\mathsf{A}||.$$

The MAXIMUM ABSOLUTE COLUMN SUM NORM $||\mathsf{A}||_1$, SPECTRAL NORM $||\mathsf{A}||_2$, and MAXIMUM ABSOLUTE ROW SUM NORM $||\mathsf{A}||_\infty$ satisfy

$$(||\mathsf{A}||_2)^2 \leq ||\mathsf{A}||_1\,||\mathsf{A}||_\infty.$$

For a SQUARE MATRIX, the SPECTRAL NORM, which is the SQUARE ROOT of the maximum EIGENVALUE of $\mathsf{A}^\dagger\mathsf{A}$ (where $\mathsf{A}^\dagger$ is the ADJOINT MATRIX), is often referred to as "the" matrix norm.

see also COMPATIBLE, HILBERT-SCHMIDT NORM, MAXIMUM ABSOLUTE COLUMN SUM NORM, MAXIMUM ABSOLUTE ROW SUM NORM, NATURAL NORM, NORM, POLYNOMIAL NORM, SPECTRAL NORM, SPECTRAL RADIUS, VECTOR NORM

References
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 1114–1125, 1979.

Matrix Polynomial Identity

see CAYLEY-HAMILTON THEOREM

Matrix Product

The result of a MATRIX MULTIPLICATION.

see also PRODUCT

Matrix Transpose

see TRANSPOSE

Matroid

Roughly speaking, a finite set together with a generalization of a concept from linear algebra that satisfies a natural set of properties for that concept. For example, the finite set could be the rows of a MATRIX, and the generalizing concept could be linear dependence and independence of any subset of rows of the MATRIX. The number of matroids with n points are 1, 1, 2, 4, 9, 26, 101, 950, ... (Sloane's A002773).

References
Sloane, N. J. A. Sequences A002773/M1197 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Sloane, N. J. A. and Plouffe, S. Extended entry in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.
Whitely, W. "Matroids and Rigid Structures." In Matroid Applications, *Encyclopedia of Mathematics and Its Applications* (Ed. N. White), Vol. 40. New York: Cambridge University Press, pp. 1–53, 1992.

Maurer Rose

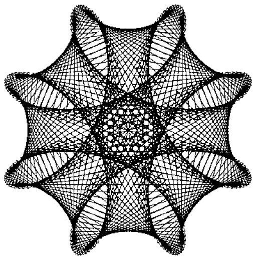

$n = 4, d = 120$

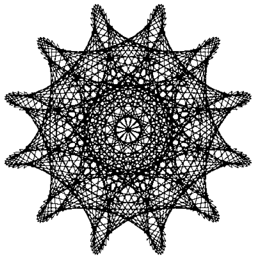

$n = 6, d = 72$

A Maurer rose is a plot of a "walk" along an n- (or $2n$-) leafed ROSE in steps of a fixed number d degrees, including all cosets.

see also STARR ROSE

References
Maurer, P. "A Rose is a Rose..." *Amer. Math. Monthly* **94**, 631–645, 1987.
Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 96–102, 1991.

Max Sequence

A sequence defined from a FINITE sequence $a_0, a_1, \ldots, a_n$ by defining $a_{n+1} = \max_i(a_i + a_{n-i})$.

see also MEX SEQUENCE

References
Guy, R. K. "Max and Mex Sequences." §E27 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 227–228, 1994.

Maximal Ideal

A maximal ideal of a Ring R is an Ideal I, not equal to R, such that there are no Ideals "in between" I and R. In other words, if J is an Ideal which contains I as a Subset, then either $J = I$ or $J = R$. For example, $n\mathbb{Z}$ is a maximal ideal of $\mathbb{Z}$ Iff n is Prime, where $\mathbb{Z}$ is the Ring of Integers.

see also Ideal, Prime Ideal, Regular Local Ring, Ring

Maximal Sum-Free Set

A maximal sum-free set is a set $\{a_1, a_2, \ldots, a_n\}$ of distinct Natural Numbers such that a maximum l of them satisfy $a_{i_j} + a_{i_k} \neq a_m$, for $1 \leq j < k \leq l$, $1 \leq m \leq n$.

see also Maximal Zero-Sum-Free Set

References

Guy, R. K. "Maximal Sum-Free Sets." §C14 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 128–129, 1994.

Maximal Zero-Sum-Free Set

A set having the largest number k of distinct residue classes modulo m so that no Subset has zero sum.

see also Maximal Sum-Free Set

References

Guy, R. K. "Maximal Zero-Sum-Free Sets." §C15 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 129–131, 1994.

Maximally Linear Independent

A set of Vectors is maximally linearly independent if including any other Vector in the Vector Space would make it Linearly Dependent (i.e., if any other Vector in the Space can be expressed as a linear combination of elements of a maximal set—the Basis).

Maximum

The largest value of a set, function, etc. The maximum value of a set of elements $A = \{a_i\}_{i=1}^N$ is denoted $\max A$ or $\max_i a_i$, and is equal to the last element of a sorted (i.e., ordered) version of A. For example, given the set {3, 5, 4, 1}, the sorted version is {1, 3, 4, 5}, so the maximum is 5. The maximum and Minimum are the simplest Order Statistics.

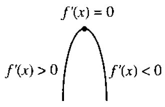

maximum

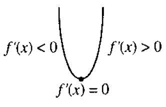

minimum

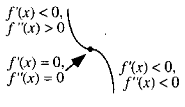
$f'(x) < 0,$ $f''(x) > 0$
$f'(x) = 0,$ $f''(x) = 0$
$f'(x) < 0,$ $f''(x) < 0$

stationary point

A continuous Function may assume a maximum at a single point or may have maxima at a number of points. A Global Maximum of a Function is the largest value in the entire Range of the Function, and a Local Maximum is the largest value in some local neighborhood.

For a function $f(x)$ which is Continuous at a point x_0, a Necessary but not Sufficient condition for $f(x)$ to have a Relative Maximum at $x = x_0$ is that x_0 be a Critical Point (i.e., $f(x)$ is either not Differentiable at x_0 or x_0 is a Stationary Point, in which case $f'(x_0) = 0$).

The First Derivative Test can be applied to Continuous Functions to distinguish maxima from Minima. For twice differentiable functions of one variable, $f(x)$, or of two variables, $f(x, y)$, the Second Derivative Test can sometimes also identify the nature of an Extremum. For a function $f(x)$, the Extremum Test succeeds under more general conditions than the Second Derivative Test.

see also Critical Point, Extremum, Extremum Test, First Derivative Test, Global Maximum, Inflection Point, Local Maximum, Midrange, Minimum, Order Statistic, Saddle Point (Function), Second Derivative Test, Stationary Point

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 14, 1972.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Minimization or Maximization of Functions." Ch. 10 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 387–448, 1992.

Tikhomirov, V. M. *Stories About Maxima and Minima.* Providence, RI: Amer. Math. Soc., 1991.

Maximum Absolute Column Sum Norm

The Natural Norm induced by the L_1-Norm is called the maximum absolute column sum norm and is defined by

$$||\mathsf{A}||_1 = \max_j \sum_{i=1}^n |a_{ij}|$$

for a Matrix A.

see also L_1-Norm, Maximum Absolute Row Sum Norm

Maximum Absolute Row Sum Norm

The Natural Norm induced by the L_∞-Norm is called the maximum absolute row sum norm and is defined by

$$||\mathsf{A}||_\infty = \max_i \sum_{j=1}^n |a_{ij}|$$

for a Matrix A.

see also L_∞-Norm, Maximum Absolute Column Sum Norm

Maximum Clique Problem

see Party Problem

Maximum Entropy Method

A DECONVOLUTION ALGORITHM (sometimes abbreviated MEM) which functions by minimizing a smoothness function ("ENTROPY") in an image. Maximum entropy is also called the ALL-POLES MODEL or AUTOREGRESSIVE MODEL. For images with more than a million pixels, maximum entropy is faster than the CLEAN ALGORITHM.

MEM is commonly employed in astronomical synthesis imaging. In this application, the resolution depends on the signal to NOISE ratio, which must be specified. Therefore, resolution is image dependent and varies across the map. MEM is also biased, since the ensemble average of the estimated noise is NONZERO. However, this bias is much smaller than the NOISE for pixels with a SNR $\gg 1$. It can yield super-resolution, which can usually be trusted to an order of magnitude in SOLID ANGLE.

Several definitions of "ENTROPY" normalized to the flux in the image are

$$H_1 \equiv \sum_k \ln\left(\frac{I_k}{M_k}\right) \tag{1}$$

$$H_2 \equiv -\sum_k I_k \ln\left(\frac{I_k}{M_k e}\right), \tag{2}$$

where M_k is a "default image" and I_k is the smoothed image. Some unnormalized entropy measures (Cornwell 1982, p. 3) are given by

$$H_1 \equiv -\sum f_i \ln(f_i) \tag{3}$$

$$H_2 \equiv \sum \ln(f_i) \tag{4}$$

$$H_3 \equiv -\sum \frac{1}{\ln(f_i)} \tag{5}$$

$$H_4 \equiv -\sum \frac{1}{[\ln(f_i)]^2} \tag{6}$$

$$H_5 \equiv \sum \sqrt{\ln(f_i)}\,. \tag{7}$$

see also CLEAN ALGORITHM, DECONVOLUTION, LUCY

References

Cornwell, T. J. "Can CLEAN be Improved?" VLA Scientific Memorandum No. 141, March 1982.

Cornwell, T. and Braun, R. "Deconvolution." Ch. 8 in *Synthesis Imaging in Radio Astronomy: Third NRAO Summer School, 1988* (Ed. R. A. Perley, F. R. Schwab, and A. H. Bridle). San Francisco, CA: Astronomical Society of the Pacific, pp. 167–183, 1989.

Christiansen, W. N. and Högbom, J. A. *Radiotelescopes, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 217–218, 1985.

Narayan, R. and Nityananda, R. "Maximum Entropy Restoration in Astronomy." *Ann. Rev. Astron. Astrophys.* **24**, 127–170, 1986.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Power Spectrum Estimation by the Maximum Entropy (All Poles) Method" and "Maximum Entropy Image Restoration." §13.7 and 18.7 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 565–569 and 809–817, 1992.

Thompson, A. R.; Moran, J. M.; and Swenson, G. W. Jr. §3.2 in *Interferometry and Synthesis in Radio Astronomy.* New York: Wiley, pp. 349–352, 1986.

Maximum Likelihood

The procedure of finding the value of one or more parameters for a given statistic which makes the *known* LIKELIHOOD distribution a MAXIMUM. The maximum likelihood estimate for a parameter μ is denoted $\hat{\mu}$.

For a BERNOULLI DISTRIBUTION,

$$\frac{d}{d\theta}\left[\binom{N}{Np}\theta^{Np}(1-\theta)^{Nq}\right] = Np(1-\theta)-\theta Nq = 0, \tag{1}$$

so maximum likelihood occurs for $\theta = p$. If p is not known ahead of time, the likelihood function is

$$\begin{aligned} f(x_1,\ldots,x_n|p) &= P(X_1 = x_1,\ldots,X_n = x_n|p) \\ &= p^{x_1}(1-p)^{1-x_1}\cdots p^{x_n}(1-p)^{1-x_1 n} \\ &= p^{\Sigma x_i}(1-p)^{\Sigma(1-x_i)} = p^{\Sigma x_i}(1-p)^{n-\Sigma x_i}, \end{aligned} \tag{2}$$

where $x = 0$ or 1, and $i = 1, \ldots, n$.

$$\ln f = \sum x_i \ln p + \left(n - \sum x_i\right)\ln(1-p) \tag{3}$$

$$\frac{d(\ln f)}{dp} = \frac{\sum x_i}{p} - \frac{n-\sum x_i}{1-p} = 0 \tag{4}$$

$$\sum x_i - p\sum x_i = np - p\sum x_i \tag{5}$$

$$\hat{p} = \frac{\sum x_i}{n}. \tag{6}$$

For a GAUSSIAN DISTRIBUTION,

$$\begin{aligned} f(x_1,\ldots,x_n|\mu,\sigma) &= \prod \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/2\sigma^2} \\ &= \frac{(2\pi)^{-n/2}}{\sigma^n}\exp\left[-\frac{\sum(x_i-\mu)^2}{2\sigma^2}\right] \end{aligned} \tag{7}$$

$$\ln f = -\tfrac{1}{2}n\ln(2\pi) - n\ln\sigma - \frac{\sum(x_i-\mu)^2}{2\sigma^2} \tag{8}$$

$$\frac{\partial(\ln f)}{\partial\mu} = \frac{\sum(x_i-\mu)}{\sigma^2} = 0 \tag{9}$$

gives

$$\hat{\mu} = \frac{\sum x_i}{n}. \tag{10}$$

$$\frac{\partial(\ln f)}{\partial\sigma} = -\frac{n}{\sigma} + \frac{\sum(x_i-\mu)^2}{\sigma^3} \tag{11}$$

gives

$$\hat{\sigma} = \sqrt{\frac{\sum (x_i - \hat{\mu})^2}{n}}. \tag{12}$$

Note that in this case, the maximum likelihood STANDARD DEVIATION is the sample STANDARD DEVIATION, which is a BIASED ESTIMATOR for the population STANDARD DEVIATION.

For a weighted GAUSSIAN DISTRIBUTION,

$$\begin{aligned} f(x_1, \ldots, x_n | \mu, \sigma) &= \prod \frac{1}{\sigma_i \sqrt{2\pi}} e^{-(x_i - \mu)^2 / 2{\sigma_i}^2} \\ &= \frac{(2\pi)^{-n/2}}{\sigma^n} \exp\left[-\frac{\sum (x_i - \mu)^2}{2\sigma^2} \right] \end{aligned} \tag{13}$$

$$\ln f = -\tfrac{1}{2} n \ln(2\pi) - n \sum \ln \sigma_i - \sum \frac{(x_i - \mu)^2}{2{\sigma_i}^2} \tag{14}$$

$$\frac{\partial (\ln f)}{\partial \mu} = \sum \frac{(x_i - \mu)}{{\sigma_i}^2} = \sum \frac{x_i}{{\sigma_i}^2} - \mu \sum \frac{1}{{\sigma_i}^2} = 0 \tag{15}$$

gives

$$\hat{\mu} = \frac{\sum \frac{x_i}{{\sigma_i}^2}}{\sum \frac{1}{{\sigma_i}^2}}. \tag{16}$$

The VARIANCE of the MEAN is then

$${\sigma_\mu}^2 = \sum {\sigma_i}^2 \left(\frac{\partial \mu}{\partial x_i} \right)^2. \tag{17}$$

But

$$\frac{\partial \mu}{\partial x_i} = \frac{\partial}{\partial x_i} \frac{\sum (x_i / {\sigma_i}^2)}{\sum (1 / {\sigma_i}^2)} = \frac{1 / {\sigma_i}^2}{\sum (1 / {\sigma_i}^2)}, \tag{18}$$

so

$$\begin{aligned} {\sigma_\mu}^2 &= \sum {\sigma_i}^2 \left(\frac{1/{\sigma_i}^2}{\sum (1/{\sigma_i}^2)} \right)^2 \\ &= \sum \frac{1/{\sigma_i}^2}{\left[\sum (1/{\sigma_i}^2) \right]^2} = \frac{1}{\sum (1/{\sigma_i}^2)}. \end{aligned} \tag{19}$$

For a POISSON DISTRIBUTION,

$$f(x_1, \ldots, x_n | \lambda) = \frac{e^{-\lambda} \lambda^{x_1}}{x_1!} \cdots \frac{e^{-\lambda} \lambda^{x_n}}{x_n!} = \frac{e^{-n\lambda} \lambda^{\sum x_i}}{x_1! \cdots x_n!} \tag{20}$$

$$\ln f = -n\lambda + (\ln \lambda) \sum x_i - \ln \left(\prod x_i! \right) \tag{21}$$

$$\frac{d(\ln f)}{\lambda} = -n + \frac{\sum x_i}{\lambda} = 0 \tag{22}$$

$$\hat{\lambda} = \frac{\sum x_i}{n}. \tag{23}$$

see also BAYESIAN ANALYSIS

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Least Squares as a Maximum Likelihood Estimator." §15.1 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 651–655, 1992.

Maxwell Distribution

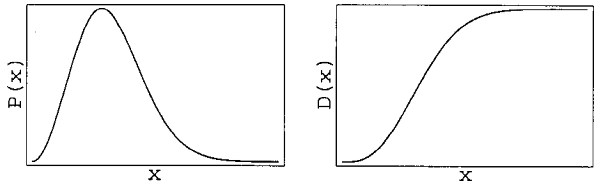

The distribution of speeds of molecules in thermal equilibrium as given by statistical mechanics. The probability and cumulative distributions are

$$P(x) = \sqrt{\frac{2}{\pi}} a^{3/2} x^2 e^{-ax^2/2} \tag{1}$$

$$D(x) = \frac{2\gamma(\frac{3}{2}, \frac{1}{2}ax^2)}{\sqrt{\pi}}, \tag{2}$$

where $\gamma(a, x)$ is an incomplete GAMMA FUNCTION and $x \in [0, \infty)$. The moments are

$$\mu = 2\sqrt{\frac{2}{\pi a}} \tag{3}$$

$$\mu_2 = \frac{3}{a} \tag{4}$$

$$\mu_3 = 8\sqrt{\frac{2}{a^3 \pi}} \tag{5}$$

$$\mu_4 = \tfrac{15}{2}, \tag{6}$$

and the MEAN, VARIANCE, SKEWNESS, and KURTOSIS are

$$\mu = 2\sqrt{\frac{2}{\pi a}} \tag{7}$$

$$\sigma^2 = \frac{3\pi - 8}{\pi a} \tag{8}$$

$$\gamma_1 = \frac{8}{3}\sqrt{\frac{2}{3\pi}} \tag{9}$$

$$\gamma_2 = -\tfrac{4}{3}. \tag{10}$$

see also EXPONENTIAL DISTRIBUTION, GAUSSIAN DISTRIBUTION, RAYLEIGH DISTRIBUTION

References

Spiegel, M. R. *Theory and Problems of Probability and Statistics.* New York: McGraw-Hill, p. 119, 1992.

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 252, 1993.

May's Theorem

Simple majority vote is the only procedure which is ANONYMOUS, DUAL, and MONOTONIC.

References

May, K. "A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decision." *Econometrica* **20**, 680–684, 1952.

May-Thomason Uniqueness Theorem

For every infinite Loop Space Machine E, there is a natural equivalence of spectra between EX and Segal's spectrum $\mathbf{B}X$.

References

May, J. P. and Thomason, R. W. "The Uniqueness of Infinite Loop Space Machines." *Topology* **17**, 205–224, 1978.

Weibel, C. A. "The Mathematical Enterprises of Robert Thomason." *Bull. Amer. Math. Soc.* **34**, 1–13, 1996.

Maze

A maze is a drawing of impenetrable line segments (or curves) with "paths" between them. The goal of the maze is to start at one given point and find a path which reaches a second given point.

References

Gardner, M. "Mazes." Ch. 10 in *The Second Scientific American Book of Mathematical Puzzles & Diversions: A New Selection.* New York: Simon and Schuster, pp. 112–118, 1961.

Jablan, S. "Roman Mazes." `http://members.tripod.com/~modularity/mazes.htm`.

Matthews, W. H. *Mazes and Labyrinths: Their History and Development.* New York: Dover, 1970.

Pappas, T. "Mazes." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 192–194, 1989.

Phillips, A. "The Topology of Roman Mazes." *Leonardo* **25**, 321–329, 1992.

Shepard, W. *Mazes and Labyrinths: A Book of Puzzles.* New York: Dover, 1961.

Mazur's Theorem

The generalization of the Schönflies Theorem to n-D. A smoothly embedded n-Hypersphere in an $(n+1)$-Hypersphere separates the $(n+1)$-Hypersphere into two components, each Homeomorphic to $(n+1)$-Balls. It can be proved using Morse Theory.

see also Ball, Hypersphere

McCay Circle

If the Vertex A_1 of a Triangle describes the Neuberg Circle n_1, its Median Point describes a circle whose radius is 1/3 that of the Neuberg Circle. Such a Circle is known as a McKay circle, and the three McCay circles are Concurrent at the Median Point M. Three homologous collinear points lie on the McCay circles.

see also Circle, Concurrent, Median Point, Neuberg Circles

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 290 and 306, 1929.

McCoy's Theorem

If two Square $n \times n$ Matrices A and B are simultaneously upper triangularizable by similarity transforms, then there is an ordering $a_1, \ldots, a_n$ of the Eigenvalues of A and $b_1, \ldots, b_n$ of the Eigenvalues of B so that, given any Polynomial $p(x,y)$ in noncommuting variables, the Eigenvalues of $p(A,B)$ are the numbers $p(a_i,b_i)$ with $i = 1, \ldots, n$. McCoy's theorem states the converse: If every Polynomial exhibits the correct Eigenvalues in a consistent ordering, then A and B are simultaneously triangularizable.

References

Luchins, E. H. and McLoughlin, M. A. "In Memoriam: Olga Taussky-Todd." *Not. Amer. Math. Soc.* **43**, 838–847, 1996.

McLaughlin Group

The Sporadic Group *McL*.

References

Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas/McL.html`.

McMohan's Theorem

Consider a Gaussian Bivariate Distribution. Let $f(x_1,x_2)$ be an arbitrary Function. Then

$$\frac{\partial^2\langle f\rangle}{\partial\rho^n} = \left\langle \frac{\partial^{2n} f}{\partial {x_1}^n\,\partial {x_2}^n} \right\rangle.$$

see also Gaussian Bivariate Distribution

McNugget Number

A number which can be obtained from an order of McDonald's® Chicken McNuggets™ (prior to consuming any), which originally came in boxes of 6, 9, and 20. All integers are McNugget numbers except 1, 2, 3, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 22, 23, 25, 28, 31, 34, 37, and 43. Since the Happy Meal™-sized nugget box (4 to a box) can now be purchased separately, the modern McNugget numbers are a linear combination of 4, 6, 9, and 20. These new-fangled numbers are much less interesting than before, with only 1, 2, 3, 5, 7, and 11 remaining as non-McNugget numbers.

The Greedy Algorithm can be used to find a McNugget expansion of a given Integer.

see also Complete Sequence, Greedy Algorithm

References

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 19–20 and 233–234, 1991.

Wilson, D. `rec.puzzles` newsgroup posting, March 20, 1990.

Mean

A mean is HOMOGENEOUS and has the property that a mean μ of a set of numbers x_i satisfies

$$\min(x_1, \ldots, x_n) \leq \mu \leq \max(x_1, \ldots, x_n).$$

There are several statistical quantities called means, e.g., ARITHMETIC-GEOMETRIC MEAN, GEOMETRIC MEAN, HARMONIC MEAN, QUADRATIC MEAN, ROOT-MEAN-SQUARE. However, the quantity referred to as "the" mean is the ARITHMETIC MEAN, also called the AVERAGE.

see also ARITHMETIC-GEOMETRIC MEAN, AVERAGE, GENERALIZED MEAN, GEOMETRIC MEAN, HARMONIC MEAN, QUADRATIC MEAN, ROOT-MEAN-SQUARE

Mean Cluster Count Per Site

see s-CLUSTER

Mean Cluster Density

see s-CLUSTER

Mean Curvature

Let κ_1 and κ_2 be the PRINCIPAL CURVATURES, then their MEAN

$$H = \tfrac{1}{2}(\kappa_1 + \kappa_2) \tag{1}$$

is called the mean curvature. Let R_1 and R_2 be the radii corresponding to the PRINCIPAL CURVATURES, then the multiplicative inverse of the mean curvature H is given by the multiplicative inverse of the HARMONIC MEAN,

$$H \equiv \frac{1}{2}\left(\frac{1}{R_1} + \frac{1}{R_2}\right) = \frac{R_1 + R_2}{2R_1R_2}. \tag{2}$$

In terms of the GAUSSIAN CURVATURE K,

$$H = \tfrac{1}{2}(R_1 + R_2)K. \tag{3}$$

The mean curvature of a REGULAR SURFACE in $\mathbb{R}^3$ at a point $\mathbf{p}$ is formally defined as

$$H(\mathbf{p}) = \tfrac{1}{2}\operatorname{tr}(S(\mathrm{p})), \tag{4}$$

where S is the SHAPE OPERATOR and $\operatorname{tr}(S)$ denotes the TRACE. For a MONGE PATCH with $z = h(x, y)$,

$$H = \frac{(1 + {h_v}^2)h_{uu} - 2h_u h_v h_{uv} + (1 + {h_u}^2)h_{vv}}{(1 + {h_u}^2 + {h_v}^2)^{3/2}} \tag{5}$$

(Gray 1993, p. 307).

If $\mathbf{x} : U \to \mathbb{R}^3$ is a REGULAR PATCH, then the mean curvature is given by

$$H = \frac{eG - 2fF + gE}{2(EG - F^2)}, \tag{6}$$

where E, F, and G are coefficients of the first FUNDAMENTAL FORM and e, f, and g are coefficients of the second FUNDAMENTAL FORM (Gray 1993, p. 282). It can also be written

$$H = \frac{\det(\mathbf{x}_{uu}\mathbf{x}_u\mathbf{x}_v)|\mathbf{x}_v|^2 - 2\det(\mathbf{x}_{uv}\mathbf{x}_u\mathbf{x}_v)(\mathbf{x}_u \cdot \mathbf{x}_v)}{2[|\mathbf{x}_u|^2|\mathbf{x}_v|^2 - (\mathbf{x}_u \cdot \mathbf{x}_v)^2]^{3/2}} + \frac{\det(\mathbf{x}_{vv}\mathbf{x}_u\mathbf{x}_v)|\mathbf{x}_u|^2}{2[|\mathbf{x}_u|^2|\mathbf{x}_v|^2 - (\mathbf{x}_u \cdot \mathbf{x}_v)^2]^{3/2}} \tag{7}$$

Gray (1993, p. 285).

The GAUSSIAN and mean curvature satisfy

$$H^2 \geq K, \tag{8}$$

with equality only at UMBILIC POINTS, since

$$H^2 - K^2 = \tfrac{1}{4}(\kappa_1 - \kappa_2)^2. \tag{9}$$

If $\mathbf{p}$ is a point on a REGULAR SURFACE $M \subset \mathbb{R}^3$ and $\mathbf{v_p}$ and $\mathbf{w_p}$ are tangent vectors to M at $\mathbf{p}$, then the mean curvature of M at $\mathbf{p}$ is related to the SHAPE OPERATOR S by

$$S(\mathbf{v_p}) \times \mathbf{w_p} + \mathbf{v_p} \times S(\mathbf{w_p}) = 2H(\mathbf{p})\mathbf{v_p} \times \mathbf{w_p}. \tag{10}$$

Let $\mathbf{Z}$ be a nonvanishing VECTOR FIELD on M which is everywhere PERPENDICULAR to M, and let V and W be VECTOR FIELDS tangent to M such that $V \times W = \mathbf{Z}$, then

$$H = -\frac{\mathbf{Z} \cdot (D_V\mathbf{Z} \times W + V \times D_W\mathbf{Z})}{2|\mathbf{Z}|^3} \tag{11}$$

(Gray 1993, pp. 291–292).

Wente (1985, 1986, 1987) found a nonspherical finite surface with constant mean curvature, consisting of a self-intersecting three-lobed toroidal surface. A family of such surfaces exists.

see also GAUSSIAN CURVATURE, PRINCIPAL CURVATURES, SHAPE OPERATOR

References

Gray, A. "The Gaussian and Mean Curvatures." §14.5 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 279–285, 1993.

Isenberg, C. *The Science of Soap Films and Soap Bubbles.* New York: Dover, p. 108, 1992.

Peterson, I. *The Mathematical Tourist: Snapshots of Modern Mathematics.* New York: W. H. Freeman, pp. 69–70, 1988.

Wente, H. C. "A Counterexample in 3-Space to a Conjecture of H. Hopf." In *Workshop Bonn 1984, Proceedings of the 25th Mathematical Workshop Held at the Max-Planck Institut für Mathematik, Bonn, June 15–22, 1984* (Ed. F. Hirzebruch, J. Schwermer, and S. Suter). New York: Springer-Verlag, pp. 421–429, 1985.

Wente, H. C. "Counterexample to a Conjecture of H. Hopf." *Pac. J. Math.* **121**, 193–243, 1986.

Wente, H. C. "Immersed Tori of Constant Mean Curvature in $\mathbb{R}^3$." In *Variational Methods for Free Surface Interfaces, Proceedings of a Conference Held in Menlo Park, CA, Sept. 7–12, 1985* (Ed. P. Concus and R. Finn). New York: Springer-Verlag, pp. 13–24, 1987.

Mean Deviation

The MEAN of the ABSOLUTE DEVIATIONS,

$$\mathrm{MD} \equiv \frac{1}{N}\sum_{i=1}^{N}|x_i - \bar{x}|,$$

where $\bar{x}$ is the MEAN of the distribution.

see also ABSOLUTE DEVIATION

Mean Distribution

For an infinite population with MEAN μ, STANDARD DEVIATION σ^2, SKEWNESS γ_1, and KURTOSIS γ_2, the corresponding quantities for the *distribution of means* are

$$\mu_{\bar{x}} = \mu \tag{1}$$

$$\sigma_{\bar{x}}{}^2 = \frac{\sigma^2}{N} \tag{2}$$

$$\gamma_{1,\bar{x}} = \frac{\gamma_1}{\sqrt{N}} \tag{3}$$

$$\gamma_{2,\bar{x}} = \frac{\gamma_2}{N}. \tag{4}$$

For a population of M (Kenney and Keeping 1962, p. 181),

$$\mu_{\bar{x}}^{(M)} = \mu \tag{5}$$

$$\sigma^{2(M)} = \frac{\sigma^2}{N}\frac{M-N}{M-1}. \tag{6}$$

References

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, 1962.

Mean Run Count Per Site

see s-RUN

Mean Run Density

see s-RUN

Mean Square Error

see ROOT-MEAN-SQUARE

Mean-Value Theorem

Let $f(x)$ be DIFFERENTIABLE on the OPEN INTERVAL (a,b) and CONTINUOUS on the CLOSED INTERVAL $[a,b]$. Then there is at least one point c in (a,b) such that

$$f'(c) = \frac{f(b)-f(a)}{b-a}.$$

see also EXTENDED MEAN-VALUE THEOREM, GAUSS'S MEAN-VALUE THEOREM

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 1097–1098, 1993.

Measurable Function

A function $f : X \to Y$ for which the pre-image of every measurable set in Y is measurable in X. For a BOREL MEASURE, all continuous functions are measurable.

Measurable Set

If F is a SIGMA ALGEBRA and A is a SUBSET of X, then A is called measurable if A is a member of F. X need not have, a priori, a topological structure. Even if it does, there may be no connection between the open sets in the topology and the given SIGMA ALGEBRA.

see also MEASURABLE SPACE, SIGMA ALGEBRA

Measurable Space

A SET considered together with the SIGMA ALGEBRA on the SET.

see also MEASURABLE SET, MEASURE SPACE, SIGMA ALGEBRA

Measure

The terms "measure," "measurable," etc., have very precise technical definitions (usually involving SIGMA ALGEBRAS) which makes them a little difficult to understand. However, the technical nature of the definitions is extremely important, since it gives a firm footing to concepts which are the basis for much of ANALYSIS (including some of the slippery underpinnings of CALCULUS).

For example, every definition of an INTEGRAL is based on a particular measure: the RIEMANN INTEGRAL is based on JORDAN MEASURE, and the LEBESGUE INTEGRAL is based on LEBESGUE MEASURE. The study of measures and their application to INTEGRATION is known as MEASURE THEORY.

A measure is formally defined as a MAP $m : F \to \mathbb{R}$ (the reals) such that $m(\varnothing) = 0$ and, if A_n is a COUNTABLE SEQUENCE in F and the A_n are pairwise DISJOINT, then

$$m\left(\bigcup_n A_n\right) = \sum_n m(A_n).$$

If, in addition, $m(X) = 1$, then m is said to be a PROBABILITY MEASURE.

A measure m may also be defined on SETS other than those in the SIGMA ALGEBRA F. By adding to F all sets to which m assigns measure zero, we again obtain a SIGMA ALGEBRA and call this the "completion" of F with respect to m. Thus, the completion of a SIGMA ALGEBRA is the smallest SIGMA ALGEBRA containing F and all sets of measure zero.

see also ALMOST EVERYWHERE, BOREL MEASURE, ERGODIC MEASURE, EULER MEASURE, GAUSS MEASURE, HAAR MEASURE, HAUSDORFF MEASURE, HELSON-SZEGŐ MEASURE, INTEGRAL, JORDAN MEASURE, LEBESGUE MEASURE, LIOUVILLE MEASURE, MAHLER'S

Measure, Measurable Space, Measure Algebra, Measure Space, Minkowski Measure, Natural Measure, Probability Measure, Wiener Measure

Measure Algebra

A Boolean Sigma Algebra which possesses a Measure.

Measure Polytope

see Hypercube

Measure-Preserving Transformation

see Endomorphism

Measure Space

A measure space is a Measurable Space possessing a Nonnegative Measure. Examples of measure spaces include n-D Euclidean Space with Lebesgue Measure and the unit interval with Lebesgue Measure (i.e., probability).

see also Lebesgue Measure, Measurable Space

Measure Theory

The mathematical theory of how to perform Integration in arbitrary Measure Spaces.

see also Cantor Set, Fractal, Integral, Measurable Function, Measurable Set, Measurable Space, Measure, Measure Space

References

Doob, J. L. *Measure Theory.* New York: Springer-Verlag, 1994.

Evans, L. C. and Gariepy, R. F. *Measure Theory and Finite Properties of Functions.* Boca Raton, FL: CRC Press, 1992.

Gordon, R. A. *The Integrals of Lebesgue, Denjoy, Perron, and Henstock.* Providence, RI: Amer. Math. Soc., 1994.

Halmos, P. R. *Measure Theory.* New York: Springer-Verlag, 1974.

Henstock, R. *The General Theory of Integration.* Oxford, England: Clarendon Press, 1991.

Kestelman, H. *Modern Theories of Integration, 2nd rev. ed.* New York: Dover, 1960.

Rao, M. M. *Measure Theory And Integration.* New York: Wiley, 1987.

Strook, D. W. *A Concise Introduction to the Theory of Integration, 2nd ed.* Boston, MA: Birkhäuser, 1994.

Mechanical Quadrature

see Gaussian Quadrature

Mecon

Buckminster Fuller's term for the Truncated Octahedron.

see also Dymaxion

Medial Axis

The boundaries of the cells of a Voronoi Diagram.

Medial Deltoidal Hexecontahedron

The Dual of the Rhombidodecadodecahedron.

Medial Disdyakis Triacontahedron

The Dual of the Truncated Dodecadodecahedron.

Medial Hexagonal Hexecontahedron

The Dual of the Snub Icosidodecadodecahedron.

Medial Icosacronic Hexecontahedron

The Dual of the Icosidodecadodecahedron.

Medial Inverted Pentagonal Hexecontahedron

The Dual of the Inverted Snub Dodecadodecahedron.

Medial Pentagonal Hexecontahedron

The Dual of the Snub Dodecadodecahedron.

Medial Rhombic Triacontahedron

A Zonohedron which is the Dual of the Dodecadodecahedron. It is also called the Small Stellated Triacontahedron.

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 125, 1989.

Medial Triambic Icosahedron

The Dual of the Ditrigonal Dodecadodecahedron.

Medial Triangle

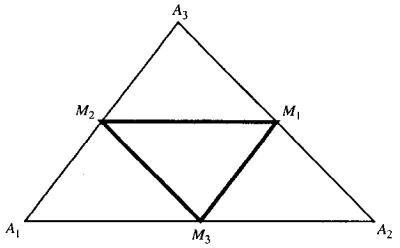

The Triangle $\Delta M_1M_2M_3$ formed by joining the Midpoints of the sides of a Triangle $\Delta A_1A_2A_3$. The medial triangle is sometimes also called the Auxiliary Triangle (Dixon 1991). The medial triangle has Trilinear Coordinates

$$\begin{aligned} A' &= 0 : b^{-1} : c^{-1} \\ B' &= a^{-1} : 0 : c^{-1} \\ C' &= a^{-1} : b^{-1} : 0. \end{aligned}$$

The medial triangle $\Delta M_1'M_2'M_3'$ of the medial triangle $\Delta M_1M_2M_3$ of a Triangle $\Delta A_1A_2A_3$ is similar to $\Delta A_1A_2A_3$.

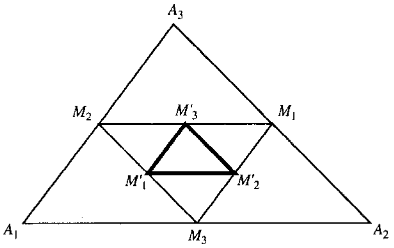

see also ANTICOMPLEMENTARY TRIANGLE

References
Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 18–20, 1967.
Dixon, R. *Mathographics.* New York: Dover, p. 56, 1991.

Medial Triangle Locus Theorem

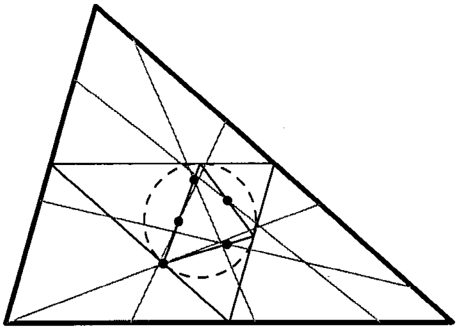

Given an original triangle (thick line), find the MEDIAL TRIANGLE (outer thin line) and its INCIRCLE. Take the PEDAL TRIANGLE (inner thin line) of the MEDIAL TRIANGLE with the INCENTER as the PEDAL POINT. Now pick any point on the original triangle, and connect it to the point located a half-PERIMETER away (gray lines). Then the locus of the MIDPOINTS of these lines (the •s in the above diagram) is the PEDAL TRIANGLE.

References
Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., pp. 261–267, 1991.
Tsintsifas, G. "Problem 674." *Crux Math.,* p. 256, 1982.

Median Point

see CENTROID (GEOMETRIC)

Median (Statistics)

The middle value of a distribution or average of the two middle items, denoted $\mu_{1/2}$ or $\tilde{x}$. For small samples, the MEAN is more efficient than the median and approximately $\pi/2$ less. It is less sensitive to outliers than the MEAN (Kenney and Keeping 1962, p. 211).

For large N samples with population median $\tilde{x}_0$,

$$\mu_{\tilde{x}} = \tilde{x}_0$$
$$\sigma_{\tilde{x}}{}^2 = \frac{1}{8Nf^2(\tilde{x}_0)}.$$

The median is an L-ESTIMATE (Press *et al.* 1992).

see also MEAN, MIDRANGE, MODE

References
Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, 1962.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 694, 1992.

Median (Triangle)

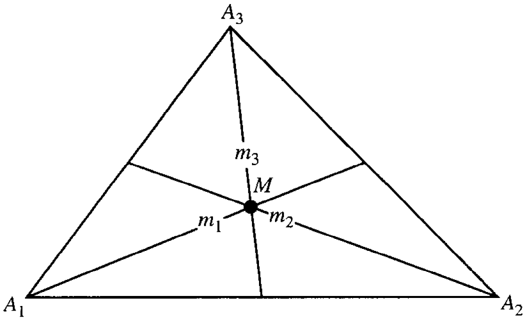

The CEVIAN from a TRIANGLE'S VERTEX to the MIDPOINT of the opposite side is called a median of the TRIANGLE. The three medians of any TRIANGLE are CONCURRENT, meeting in the TRIANGLE'S CENTROID (which has TRILINEAR COORDINATES $1/a : 1/b : 1/c$). In addition, the medians of a TRIANGLE divide one another in the ratio 2:1. A median also bisects the AREA of a TRIANGLE.

Let m_i denote the length of the median of the ith side a_i. Then

$$m_1{}^2 = \tfrac{1}{4}(2a_2{}^2 + 2a_3{}^2 - a_1{}^2) \tag{1}$$
$$m_1{}^2 + m_2{}^2 + m_3{}^2 = \tfrac{3}{4}(a_1{}^2 + a_2{}^2 + a_3{}^2) \tag{2}$$

(Johnson 1929, p. 68). The AREA of a TRIANGLE can be expressed in terms of the medians by

$$A = \tfrac{4}{3}\sqrt{s_m(s_m - m_1)(s_m - m_2)(s_m - m_3)}, \tag{3}$$

where

$$s_m \equiv \tfrac{1}{2}(m_1 + m_2 + m_3). \tag{4}$$

see also BIMEDIAN, EXMEDIAN, EXMEDIAN POINT, HERONIAN TRIANGLE, MEDIAL TRIANGLE

References
Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 7–8, 1967.
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 68 and 173–175, 1929.

Median Triangle

A TRIANGLE whose sides are equal and PARALLEL to the MEDIANS of a given TRIANGLE. The median triangle of the median triangle is similar to the given TRIANGLE in the ratio 3/4.

References
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 282–283, 1929.

Mediant

Given a FAREY SEQUENCE with consecutive terms h/k and h'/k', then the mediant is defined as the reduced form of the fraction $(h+h')/(k+k')$.

see also FAREY SEQUENCE

References

Conway, J. H. and Guy, R. K. "Farey Fractions and Ford Circles." *The Book of Numbers.* New York: Springer-Verlag, pp. 152–154, 1996.

Mega

Defined in terms of CIRCLE NOTATION by Steinhaus (1983, pp. 28–29) as

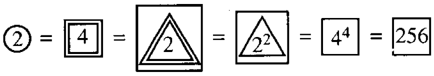

where STEINHAUS-MOSER NOTATION has also been used.

see also MEGISTRON, MOSER, STEINHAUS-MOSER NOTATION

References

Steinhaus, H. *Mathematical Snapshots, 3rd American ed.* New York: Oxford University Press, 1983.

Megistron

A very LARGE NUMBER defined in terms of CIRCLE NOTATION by Steinhaus (1983) as ⑩.

see also MEGA, MOSER

References

Steinhaus, H. *Mathematical Snapshots, 3rd American ed.* New York: Oxford University Press, pp. 28-29, 1983.

Mehler's Bessel Function Formula

$$J_0(x) = \frac{2}{\pi}\int_0^\infty \sin(x\cosh t)\,dt,$$

where $J_0(x)$ is a zeroth order BESSEL FUNCTION OF THE FIRST KIND.

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1472, 1980.

Mehler-Dirichlet Integral

$$P_n(\cos\alpha) = \frac{\sqrt{2}}{\pi}\int_0^\alpha \frac{\cos[(n+\frac{1}{2})\phi]}{\sqrt{\cos\phi - \cos\alpha}}\,d\phi,$$

where $P_n(x)$ is a LEGENDRE POLYNOMIAL.

Mehler's Hermite Polynomial Formula

$$\sum_{n=0}^\infty \frac{H_n(x)H_n(y)}{n!}(\tfrac{1}{2}w)^n = (1+4w^2)^{-1/2}\exp\left[\frac{2xyw-(x^2+y^2)w^2}{1-w^2}\right],$$

where $H_n(x)$ is a HERMITE POLYNOMIAL.

References

Almqvist, G. and Zeilberger, D. "The Method of Differentiating Under the Integral Sign." *J. Symb. Comput.* **10**, 571–591, 1990.

Foata, D. "A Combinatorial Proof of the Mehler Formula." *J. Comb. Th. Ser. A* **24**, 250–259, 1978.

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, pp. 194–195, 1996.

Rainville, E. D. *Special Functions.* New York: Chelsea, p. 198, 1971.

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., p. 380, 1975.

Mehler Quadrature

see JACOBI-GAUSS QUADRATURE

Meijer's *G*-Function

$$G_{p,q}^{m,n}\left(x\middle|{}^{a_1,\ldots,a_p}_{b_1,\ldots,b_p}\right) \equiv \frac{1}{2\pi i} \times \int_{\gamma_L} \frac{\prod_{j=1}^m \Gamma(b_j - z)\prod_{j=1}^n (1-a_j+s)}{\prod_{j=m+1}^q \Gamma(1-b_j+z)\prod_{j=n+1}^q \Gamma(q_j - z)} x^z\,dz,$$

where $\Gamma(z)$ is the GAMMA FUNCTION. The CONTOUR γ_L and other details are discussed by Gradshteyn and Ryzhik (1980, pp. 896–903 and 1068–1071). Prudnikov *et al.* (1990) contains an extensive nearly 200-page listing of formulas for the Meijer *G*-function.

see also FOX'S *H*-FUNCTION, *G*-FUNCTION, MACROBERT'S *E*-FUNCTION, RAMANUJAN *g*- AND *G*-FUNCTIONS

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, 1979.

Luke, Y. L. *The Special Functions and Their Approximations, 2 vols.* New York: Academic Press, 1969.

Mathai, A. M. *A Handbook of Generalized Special Functions for Statistical and Physical Sciences.* New York: Oxford University Press, 1993.

Prudnikov, A. P.; Marichev, O. I.; and Brychkov, Yu. A.; *Integrals and Series, Vol. 3: More Special Functions.* Newark, NJ: Gordon and Breach, 1990.

Meissel's Formula

A modification of LEGENDRE'S FORMULA for the PRIME COUNTING FUNCTION $\pi(x)$. It starts with

$$\begin{aligned}\lfloor x\rfloor = 1 + \sum_{1\leq i\leq a}\left\lfloor \frac{x}{p_i}\right\rfloor - \sum_{1\leq i<j\leq a}\left\lfloor\frac{x}{p_ip_j}\right\rfloor \\ + \sum_{1\leq i<j<k\leq a}\left\lfloor\frac{x}{p_ip_jp_k}\right\rfloor - \ldots \\ + \pi(x) - a + P_2(x,a) + P_3(x,a) + \ldots,\end{aligned} \qquad (1)$$

where $\lfloor x\rfloor$ is the FLOOR FUNCTION, $P_2(x,a)$ is the number of INTEGERS $p_ip_j \leq x$ with $a+1 \leq j \leq j$, and $P_3(x,a)$ is the number of INTEGERS $p_ip_jp_k \leq x$ with $a+1 \leq i \leq j \leq k$. Identities satisfied by the Ps include

$$P_2(x,a) = \sum\left[\pi\left(\frac{x}{p_i}\right) - (i-1)\right] \qquad (2)$$

for $p_a < p_i \leq \sqrt{x}$ and

$$\begin{aligned}P_3(x,a) &= \sum_{i>a} P_2\left(\frac{x}{p_i}, a\right) \\ &= \sum_{i=a+1}^{c}\sum_{j=i}^{\pi(\sqrt{x/p_i})}\left[\pi\left(\frac{x}{p_ip_j}\right) - (j-1)\right]. \qquad (3)\end{aligned}$$

Meissel's formula is

$$\begin{aligned}\pi(x) = \lfloor x\rfloor - \sum_{i=1}^{c}\left\lfloor\frac{x}{p_i}\right\rfloor + \sum_{1\leq i\leq j\leq c}\left\lfloor\frac{x}{p_ip_j}\right\rfloor - \ldots \\ + \tfrac{1}{2}(b+c-2)(b-c+1) - \sum_{c\leq i\leq b}\pi\left(\frac{x}{p_i}\right), \qquad (4)\end{aligned}$$

where

$$b \equiv \pi(x^{1/2}) \qquad (5)$$

$$c \equiv \pi(x^{1/3}). \qquad (6)$$

Taking the derivation one step further yields LEHMER'S FORMULA.

see also LEGENDRE'S FORMULA, LEHMER'S FORMULA, PRIME COUNTING FUNCTION

References

Riesel, H. "Meissel's Formula." *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, p. 12, 1994.

Mellin Transform

$$\phi(z) = \int_0^\infty t^{z-1} f(t)\,dt$$

$$f(t) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} t^{-z}\phi(z)\,dz.$$

see also STRASSEN FORMULAS

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 795, 1985.

Bracewell, R. *The Fourier Transform and Its Applications.* New York: McGraw-Hill, pp. 254–257, 1965.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 469–471, 1953.

Melnikov-Arnold Integral

$$A_m(\lambda) \equiv \int_{-\infty}^{\infty}\cos\left[\tfrac{1}{2}m\phi(t) - \lambda t\right]dt,$$

where the function

$$\phi(t) \equiv 4\tan^{-1}(e^t) - \pi$$

describes the motion along the pendulum SEPARATRIX. Chirikov (1979) has shown that this integral has the approximate value

$$A_m(\lambda) \approx \begin{cases} \frac{4\pi(2\lambda)^{m-1}}{\Gamma(m)}e^{-\pi\lambda/2} & \text{for } \lambda > 0 \\ -\frac{4e^{-\pi|\lambda|/2}}{(2|l|)^{m+1}}\Gamma(m+1)\sin(\pi m) & \text{for } \lambda < 0. \end{cases}$$

References

Chirikov, B. V. "A Universal Instability of Many-Dimensional Oscillator Systems." *Phys. Rep.* **52**, 264–379, 1979.

Melodic Series

If a_1, a_2, a_3, ... is an ARTISTIC SERIES, then $1/a_1$, $1/a_2$, $1/a_3$, ... is a MELODIC SERIES. The RECURRENCE RELATION obeyed by melodic series is

$$b_{i+3} = \frac{b_ib_{i+2}{}^2}{b_{i+1}{}^2} + \frac{b_{i+2}{}^2}{b_{i+1}} - b_{i+2}.$$

see also ARTISTIC SERIES

References

Duffin, R. J. "On Seeing Progressions of Constant Cross Ratio." *Amer. Math. Monthly* **100**, 38–47, 1993.

MEM

see MAXIMUM ENTROPY METHOD

Memoryless

A variable x is memoryless with respect to t if, for all s with $t \neq 0$,

$$P(x > s + t | x > t) = P(x > s). \tag{1}$$

Equivalently,

$$\frac{P(x > s + t, x > t)}{P(x > t)} = P(x > s) \tag{2}$$

$$P(x > s + t) = P(x > s) P(x > t). \tag{3}$$

The EXPONENTIAL DISTRIBUTION, which satisfies

$$P(x > t) = e^{-\lambda t} \tag{4}$$

$$P(x > s + t) = e^{-\lambda(s+t)}, \tag{5}$$

and therefore

$$\begin{aligned} P(x > s + t) = P(x > s) P(x > t) &= e^{-\lambda s} e^{-\lambda t} \\ &= e^{-\lambda(s+t)}, \end{aligned} \tag{6}$$

is the only memoryless random distribution.

see also EXPONENTIAL DISTRIBUTION

Ménage Number

see MARRIED COUPLES PROBLEM

Ménage Problem

see MARRIED COUPLES PROBLEM

Menasco's Theorem

For a BRAID with M strands, R components, P positive crossings, and N negative crossings,

$$\begin{cases} P - N \le U_+ + M - R & \text{if } P \ge N \\ P - N \le U_- + M - R & \text{if } P \le N, \end{cases}$$

where $U_\pm$ are the smallest number of positive and negative crossings which must be changed to crossings of the opposite sign. These inequalities imply BENNEQUIN'S CONJECTURE. Menasco's theorem can be extended to arbitrary knot diagrams.

see also BENNEQUIN'S CONJECTURE, BRAID, UNKNOTTING NUMBER

References

Cipra, B. "From Knot to Unknot." *What's Happening in the Mathematical Sciences, Vol. 2.* Providence, RI: Amer. Math. Soc., pp. 8–13, 1994.

Menasco, W. W. "The Bennequin-Milnor Unknotting Conjectures." *C. R. Acad. Sci. Paris Sér. I Math.* **318**, 831–836, 1994.

Menelaus' Theorem

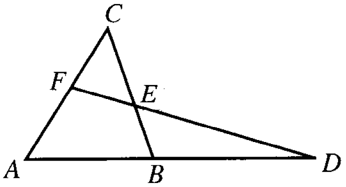

For TRIANGLES in the PLANE,

$$AD \cdot BE \cdot CF = BD \cdot CE \cdot AF. \tag{1}$$

For SPHERICAL TRIANGLES,

$$\sin AD \cdot \sin BE \cdot \sin CF = \sin BD \cdot \sin CE \cdot \sin AF. \tag{2}$$

This can be generalized to n-gons $P = [V_1, \ldots, V_n]$, where a transversal cuts the side V_iV_{i+1} in W_i for $i = 1, \ldots, n$, by

$$\prod_{i=1}^{n} \left[\frac{V_iW_i}{W_iV_{i+1}}\right] = (-1)^n. \tag{3}$$

Here, $AB || CD$ and

$$\left[\frac{AB}{CD}\right] \tag{4}$$

is the ratio of the lengths $[A, B]$ and $[C, D]$ with a PLUS or MINUS SIGN depending if these segments have the same or opposite directions (Grünbaum and Shepard 1995). The case $n = 3$ is PASCH'S AXIOM.

see also CEVA'S THEOREM, HOEHN'S THEOREM, PASCH'S AXIOM

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 122, 1987.

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 66–67, 1967.

Grünbaum, B. and Shepard, G. C. "Ceva, Menelaus, and the Area Principle." *Math. Mag.* **68**, 254–268, 1995.

Pedoe, D. *Circles: A Mathematical View, rev. ed.* Washington, DC: Math. Assoc. Amer., p. xxi, 1995.

Menger's n-Arc Theorem

Let G be a graph with A and B two disjoint n-tuples of VERTICES. Then either G contains n pairwise disjoint AB-paths, each connecting a point of A and a point of B, or there exists a set of fewer than n VERTICES that separate A and B.

References

Menger, K. *Kurventheorie.* Leipzig, Germany: Teubner, 1932.

Menger Sponge

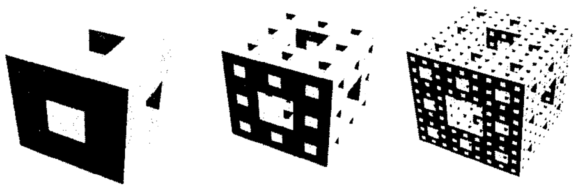

A FRACTAL which is the 3-D analog of the SIERPIŃSKI CARPET. Let N_n be the number of filled boxes, L_n the length of a side of a hole, and V_n the fractional VOLUME after the nth iteration.

$$N_n = 20^n \tag{1}$$

$$L_n = (\tfrac{1}{3})^n = 3^{-n} \tag{2}$$

$$V_n = {L_n}^3 N_n = (\tfrac{20}{27})^n. \tag{3}$$

The CAPACITY DIMENSION is therefore

$$\begin{aligned} d_{\text{cap}} &= -\lim_{n\to\infty} \frac{\ln N_n}{\ln L_n} = -\lim_{n\to\infty} \frac{\ln(20^n)}{\ln(3^{-n})} \\ &= \frac{\ln 20}{\ln 3} = \frac{\ln(2^2 \cdot 5)}{\ln 3} = \frac{2\ln 2 + \ln 5}{\ln 3} \\ &= 2.726833028\ldots. \end{aligned} \tag{4}$$

J. Mosely is leading an effort to construct a large Menger sponge out of old business cards.

see also SIERPIŃSKI CARPET, TETRIX

References

Dickau, R. M. "Menger (Sierpinski) Sponge." `http://forum.swarthmore.edu/advanced/robertd/sponge.html`.

Mosely, J. "Menger's Sponge (Depth 3)." `http://world.std.com/~j9/sponge/`.

Menn's Surface

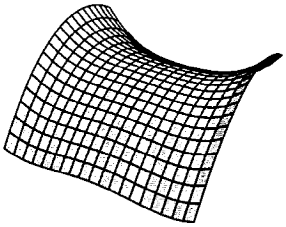

A surface given by the parametric equations

$$x(u,v) = u$$

$$y(u,v) = v$$

$$z(u,v) = au^4 + u^2v - v^2.$$

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 631, 1993.

Mensuration Formula

A mensuration formula is simply a formula for computing the length-related properties of an object (such as AREA, CIRCUMRADIUS, etc., of a POLYGON) based on other known lengths, areas, etc. Beyer (1987) gives a collection of such formulas for various plane and solid geometric figures.

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 121–133, 1987.

Mercator Projection

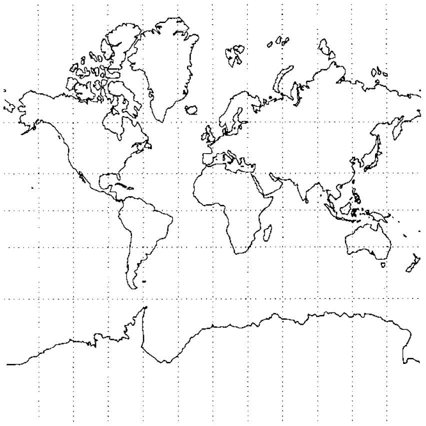

The following equations place the x-AXIS of the projection on the equator and the y-AXIS at LONGITUDE λ_0, where λ is the LONGITUDE and ϕ is the LATITUDE.

$$x = \lambda - \lambda_0 \tag{1}$$

$$y = \ln[\tan(\tfrac{1}{4}\pi + \tfrac{1}{2}\phi)] \tag{2}$$

$$= \tfrac{1}{2}\ln\left(\frac{1+\sin\phi}{1-\sin\phi}\right) \tag{3}$$

$$= \sinh^{-1}(\tan\phi) \tag{4}$$

$$= \tanh^{-1}(\sin\phi) \tag{5}$$

$$= \ln(\tan\phi + \sec\phi). \tag{6}$$

The inverse FORMULAS are

$$\phi = \tfrac{1}{2}\pi - 2\tan^{-1}(e^{-y}) = \tan^{-1}(\sinh y) \tag{7}$$

$$\lambda = x + \lambda_0. \tag{8}$$

LOXODROMES are straight lines and GREAT CIRCLES are curved.

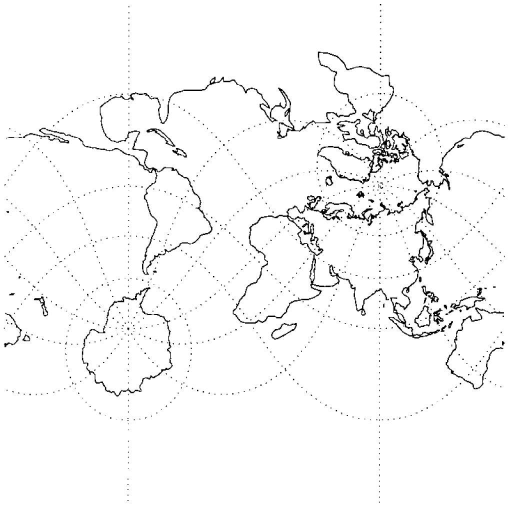

An oblique form of the Mercator projection is illustrated above. It has equations

$$x = \frac{\tan^{-1}[\tan\phi\cos\phi_p + \sin\phi_p\sin(\lambda-\lambda_0)]}{\cos(\lambda-\lambda_0)} \tag{9}$$

$$y = \tfrac{1}{2}\ln\left(\frac{1+A}{1-A}\right) = \tanh^{-1} A, \tag{10}$$

where

$$\lambda_p = \tan^{-1}\left(\frac{\cos\phi_1\sin\phi_2\cos\lambda_1 - \sin\phi_1\cos\phi_2\cos\lambda_2}{\sin\phi_1\cos\phi_2\sin\lambda_2 - \cos\phi_1\sin\phi_2\sin\lambda_1}\right) \tag{11}$$

$$\phi_p = \tan^{-1}\left(-\frac{\cos(\lambda_p-\lambda_1)}{\tan\phi_1}\right) \tag{12}$$

$$A = \sin\phi_p\sin\phi - \cos\phi_p\cos\phi\sin(\lambda-\lambda_0). \tag{13}$$

The inverse FORMULAS are

$$\phi = \sin^{-1}\left(\sin\phi_p\tanh y + \frac{\cos\phi_p\sin x}{\cosh y}\right) \tag{14}$$

$$\lambda = \lambda_0 + \tan^{-1}\left(\frac{\sin\phi_p\sin x - \cos\phi_p\sinh y}{\cos x}\right). \tag{15}$$

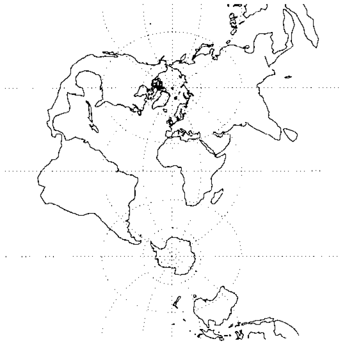

There is also a transverse form of the Mercator projection, illustrated above. It is given by the equations

$$x = \tfrac{1}{2}\ln\left(\frac{1+B}{1-B}\right) = \tanh^{-1} B \tag{16}$$

$$y = \tan^{-1}\left[\frac{\tan\phi}{\cos(\lambda-\lambda_0)}\right] - \phi_0 \tag{17}$$

$$\phi = \sin^{-1}\left(\frac{\sin D}{\cosh x}\right) \tag{18}$$

$$\lambda = \lambda_0 + \tan^{-1}\left(\frac{\sinh x}{\cos D}\right), \tag{19}$$

where

$$B \equiv \cos\phi\sin(\lambda-\lambda_0) \tag{20}$$

$$D \equiv y + \phi_0. \tag{21}$$

Finally, the "universal transverse Mercator projection" is a MAP PROJECTION which maps the SPHERE into 60 zones of 6° each, with each zone mapped by a transverse Mercator projection with central MERIDIAN in the center of the zone. The zones extend from 80° S to 84° N (Dana).

see also SPHERICAL SPIRAL

References

Dana, P. H. "Map Projections." `http://www.utexas.edu/depts/grg/gcraft/notes/mapporoj/mapproj.html`.

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 38–75, 1987.

Mercator Series

The TAYLOR SERIES for the NATURAL LOGARITHM

$$\ln(1+x) = x - \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 - \ldots$$

which was found by Newton, but independently discovered and first published by Mercator in 1668.

see also LOGARITHMIC NUMBER, NATURAL LOGARITHM

Mercer's Theorem

see RIEMANN-LEBESGUE LEMMA

Mergelyan-Wesler Theorem

Let $P = \{D_1, D_2, \ldots\}$ be an infinite set of disjoint open DISKS D_n of radius r_n such that the union is almost the unit DISK. Then

$$\sum_{n=1}^{\infty} r_n = \infty. \tag{1}$$

Define

$$M_x(P) \equiv \sum_{n=1}^{\infty} {r_n}^x. \tag{2}$$

Then there is a number $e(P)$ such that $M_x(P)$ diverges for $x < e(P)$ and converges for $x > e(P)$. The above theorem gives

$$1 < e(P) < 2. \tag{3}$$

There exists a constant which improves the inequality, and the best value known is

$$S = 1.306951\ldots. \tag{4}$$

References

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, pp. 36–37, 1983.

Mandelbrot, B. B. *Fractals.* San Francisco, CA: W. H. Freeman, p. 187, 1977.

Melzack, Z. A. "On the Solid Packing Constant for Circles." *Math. Comput.* **23**, 1969.

Meridian

A line of constant Longitude on a Spheroid (or Sphere). More generally, a meridian of a Surface of Revolution is the intersection of the surface with a Plane containing the axis of revolution.

see also Latitude, Longitude, Parallel (Surface of Revolution), Surface of Revolution

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 358, 1993.

Meromorphic

A meromorphic Function is complex analytic in all but a discrete subset of its domain, and at those singularities it must go to infinity like a Polynomial (i.e., have no Essential Singularities). An equivalent definition of a meromorphic function is a complex analytic Map to the Riemann Sphere.

see also Essential Singularity, Riemann Sphere

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 382–383, 1953.

Mersenne Number

A number of the form

$$M_n \equiv 2^n - 1 \tag{1}$$

for n an Integer is known as a Mersenne number. The Mersenne numbers are therefore 2-Repdigits, and also the numbers obtained by setting $x = 1$ in a Fermat Polynomial. The first few are 1, 3, 7, 15, 31, 63, 127, 255, ... (Sloane's A000225).

The number of digits D in the Mersenne number M_n is

$$D = \lfloor \log(2^n - 1) + 1 \rfloor, \tag{2}$$

where $\lfloor x \rfloor$ is the Floor Function, which, for large n, gives

$$D \approx \lfloor n \log 2 + 1 \rfloor \approx \lfloor 0.301029n + 1 \rfloor = \lfloor 0.301029n \rfloor + 1. \tag{3}$$

In order for the Mersenne number M_n to be Prime, n must be Prime. This is true since for Composite n with factors r and s, $n = rs$. Therefore, $2^n - 1$ can be written as $2^{rs} - 1$, which is a Binomial Number and can be factored. Since the most interest in Mersenne numbers arises from attempts to factor them, many authors prefer to define a Mersenne number as a number of the above form

$$M_p = 2^p - 1, \tag{4}$$

but with p restricted to Prime values.

The search for Mersenne Primes is one of the most computationally intensive and actively pursued areas of advanced and distributed computing.

see also Cunningham Number, Eberhart's Conjecture, Fermat Number, Lucas-Lehmer Test, Mersenne Prime, Perfect Number, Repunit, Riesel Number, Sierpiński Number of the Second Kind, Sophie Germain Prime, Superperfect Number, Wieferich Prime

References

Pappas, T. "Mersenne's Number." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 211, 1989.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 14, 18–19, 22, and 29–30, 1993.

Sloane, N. J. A. Sequence A000225/M2655 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Mersenne Prime

A Mersenne Number which is Prime is called a Mersenne prime. In order for the Mersenne number M_n defined by

$$M_n \equiv 2^n - 1$$

for n an Integer to be Prime, n must be Prime. This is true since for Composite n with factors r and s, $n = rs$. Therefore, $2^n - 1$ can be written as $2^{rs} - 1$, which is a Binomial Number and can be factored. Every Mersenne Prime gives rise to a Perfect Number.

If $n \equiv 3 \pmod 4$ is a Prime, then $2n + 1$ Divides M_n Iff $2n+1$ is Prime. It is also true that Prime divisors of $2^p - 1$ must have the form $2kp + 1$ where k is a Positive Integer and simultaneously of either the form $8n+1$ or $8n - 1$ (Uspensky and Heaslet). A Prime factor p of a Mersenne number $M_q = 2^q - 1$ is a Wieferich Prime Iff $p^2 | 2^q - 1$, Therefore, Mersenne Primes are *not* Wieferich Primes. All known Mersenne numbers M_p with p Prime are Squarefree. However, Guy (1994) believes that there are M_p which are not Squarefree.

Trial Division is often used to establish the Compositeness of a potential Mersenne prime. This test immediately shows M_p to be Composite for $p = 11$, 23, 83, 131, 179, 191, 239, and 251 (with small factors 23, 47, 167, 263, 359, 383, 479, and 503, respectively). A much more powerful primality test for M_p is the Lucas-Lehmer Test.

It has been conjectured that there exist an infinite number of Mersenne primes, although finding them is computationally very challenging. The table below gives the index p of known Mersenne primes (Sloane's A000043) M_p, together with the number of digits, discovery years, and discoverer. A similar table has been compiled by C. Caldwell. Note that the region after the 35th known Mersenne prime has not been completely searched, so identification of "the" 36th Mersenne prime is tentative. L. Welsh maintains an extensive bibliography and history of Mersenne numbers. G. Woltman has organized a distributed search program via the Internet in which hundreds of volunteers use their personal computers to perform pieces of the search.

#	p	Digits	Year	Published Reference
1	2	1	Anc.	
2	3	1	Anc.	
3	5	2	Anc.	
4	7	3	Anc.	
5	13	4	1461	Reguis 1536, Cataldi 1603
6	17	6	1588	Cataldi 1603
7	19	6	1588	Cataldi 1603
8	31	10	1750	Euler 1772
9	61	19	1883	Pervouchine 1883, Seelhoff 1886
10	89	27	1911	Powers 1911
11	107	33	1913	Powers 1914
12	127	39	1876	Lucas 1876
13	521	157	1952	Lehmer 1952-3
14	607	183	1952	Lehmer 1952-3
15	1279	386	1952	Lehmer 1952-3
16	2203	664	1952	Lehmer 1952-3
17	2281	687	1952	Lehmer 1952-3
18	3217	969	1957	Riesel 1957
19	4253	1281	1961	Hurwitz 1961
20	4423	1332	1961	Hurwitz 1961
21	9689	2917	1963	Gillies 1964
22	9941	2993	1963	Gillies 1964
23	11213	3376	1963	Gillies 1964
24	19937	6002	1971	Tuckerman 1971
25	21701	6533	1978	Noll and Nickel 1980
26	23209	6987	1979	Noll 1980
27	44497	13395	1979	Nelson and Slowinski 1979
28	86243	25962	1982	Slowinski 1982
29	110503	33265	1988	Colquitt and Welsh 1991
30	132049	39751	1983	Slowinski 1988
31	216091	65050	1985	Slowinski 1989
32	756839	227832	1992	Gage and Slowinski 1992
33	859433	258716	1994	Gage and Slowinski 1994
34	1257787	378632	1996	Slowinski and Gage
35	1398269	420921	1996	Armengaud, Woltman, *et al.*
36?	2976221	895832	1997	Spence
37?	3021377	909526	1998	Clarkson, Woltman, *et al.*

see also CUNNINGHAM NUMBER, FERMAT-LUCAS NUMBER, FERMAT NUMBER, FERMAT NUMBER (LUCAS), FERMAT POLYNOMIAL, LUCAS-LEHMER TEST, MERSENNE NUMBER, PERFECT NUMBER, REPUNIT, SUPERPERFECT NUMBER

References

Bateman, P. T.; Selfridge, J. L.; and Wagstaff, S. S. "The New Mersenne Conjecture." *Amer. Math. Monthly* **96**, 125–128, 1989.

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 66, 1987.

Beiler, A. H. Ch. 3 in *Recreations in the Theory of Numbers: The Queen of Mathematics Entertains.* New York: Dover, 1966.

Caldwell, C. "Mersenne Primes: History, Theorems and Lists." `http://www.utm.edu/research/primes/mersenne.shtml`.

Caldwell, C. "GIMPS Finds a Prime! $2^{1398269}-1$ is Prime." `http://www.utm.edu/research/primes/notes/1398269/`.

Colquitt, W. N. and Welsh, L. Jr. "A New Mersenne Prime." *Math. Comput.* **56**, 867–870, 1991.

Conway, J. H. and Guy, R. K. "Mersenne's Numbers." In *The Book of Numbers.* New York: Springer-Verlag, pp. 135–137, 1996.

Gillies, D. B. "Three New Mersenne Primes and a Statistical Theory." *Math Comput.* **18**, 93–97, 1964.

Guy, R. K. "Mersenne Primes. Repunits. Fermat Numbers. Primes of Shape $k\cdot 2^n+2$ [sic]." §A3 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 8–13, 1994.

Haghighi, M. "Computation of Mersenne Primes Using a Cray X-MP." *Intl. J. Comput. Math.* **41**, 251–259, 1992.

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, pp. 14–16, 1979.

Kraitchik, M. "Mersenne Numbers and Perfect Numbers." §3.5 in *Mathematical Recreations.* New York: W. W. Norton, pp. 70–73, 1942.

Kravitz, S. and Berg, M. "Lucas' Test for Mersenne Numbers $6000 < p < 7000$." *Math. Comput.* **18**, 148–149, 1964.

Lehmer, D. H. "On Lucas's Test for the Primality of Mersenne's Numbers." *J. London Math. Soc.* **10**, 162–165, 1935.

Leyland, P. `ftp://sable.ox.ac.uk/pub/math/factors/mersenne`.

Mersenne, M. *Cogitata Physico-Mathematica.* 1644.

Mersenne Organization. "GIMPS Discovers 36th Known Mersenne Prime, $2^{2976221}-1$ is Now the Largest Known Prime." `http://www.mersenne.org/2976221.htm`.

Mersenne Organization. "GIMPS Discovers 37th Known Mersenne Prime, $2^{3021377}-1$ is Now the Largest Known Prime." `http://www.mersenne.org/3021377.htm`.

Noll, C. and Nickel, L. "The 25th and 26th Mersenne Primes." *Math. Comput.* **35**, 1387–1390, 1980.

Powers, R. E. "The Tenth Perfect Number." *Amer. Math. Monthly* **18**, 195–196, 1911.

Powers, R. E. "Note on a Mersenne Number." *Bull. Amer. Math. Soc.* **40**, 883, 1934.

Sloane, N. J. A. Sequence A000043/M0672 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Slowinski, D. "Searching for the 27th Mersenne Prime." *J. Recreat. Math.* **11**, 258–261, 1978–1979.

Slowinski, D. *Sci. News* **139**, 191, 9/16/1989.

Tuckerman, B. "The 24th Mersenne Prime." *Proc. Nat. Acad. Sci. USA* **68**, 2319–2320, 1971.

Uhler, H. S. "A Brief History of the Investigations on Mersenne Numbers and the Latest Immense Primes." *Scripta Math.* **18**, 122–131, 1952.

Uspensky, J. V. and Heaslet, M. A. *Elementary Number Theory.* New York: McGraw-Hill, 1939.

Weisstein, E. W. "Mersenne Numbers." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Mersenne.m`.

Welsh, L. "Marin Mersenne." `http://www.scruznet.com/~luke/mersenne.htm`.

Welsh, L. "Mersenne Numbers & Mersenne Primes Bibliography." `http://www.scruznet.com/~luke/biblio.htm`.

Woltman, G. "The GREAT Internet Mersenne Prime Search." `http://www.mersenne.org/prime.htm`.

Mertens Conjecture

Given MERTENS FUNCTION defined by

$$M(n)\equiv\sum_{k=1}^{n}\mu(k), \tag{1}$$

where $\mu(n)$ is the MÖBIUS FUNCTION, Mertens (1897) conjecture states that

$$|M(x)| < x^{1/2} \tag{2}$$

for $x > 1$. The conjecture has important implications, since the truth of any equality of the form

$$|M(x)| \leq cx^{1/2} \tag{3}$$

for any fixed c (the form of Mertens conjecture with $c = 1$) would imply the RIEMANN HYPOTHESIS. In 1885, Stieltjes claimed that he had a proof that $M(x)x^{-1/2}$ always stayed between two fixed bounds. However, it seems likely that Stieltjes was mistaken.

Mertens conjecture was proved false by Odlyzko and te Riele (1985). Their proof is indirect and does not produce a specific counterexample, but it does show that

$$\limsup_{x\to\infty} M(x)x^{-1/2} > 1.06 \tag{4}$$

$$\liminf_{x\to\infty} M(x)x^{-1/2} < -1.009. \tag{5}$$

Odlyzko and te Riele (1985) believe that there are no counterexamples to Mertens conjecture for $x \leq 10^{20}$, or even 10^{30}. Pintz (1987) subsequently showed that at least one counterexample to the conjecture occurs for $x \leq 10^{65}$, using a weighted integral average of $M(x)/x$ and a discrete sum involving nontrivial zeros of the RIEMANN ZETA FUNCTION.

It is still not known if

$$\limsup_{x\to\infty} |M(x)|x^{-1/2} = \infty, \tag{6}$$

although it seems very probable (Odlyzko and te Riele 1985).

see also MERTENS FUNCTION, MÖBIUS FUNCTION, RIEMANN HYPOTHESIS

References

Anderson, R. J. "On the Mertens Conjecture for Cusp Forms." *Mathematika* **26**, 236–249, 1979.

Anderson, R. J. "Corrigendum: 'On the Mertens Conjecture for Cusp Forms.'" *Mathematika* **27**, 261, 1980.

Devlin, K. "The Mertens Conjecture." *Irish Math. Soc. Bull.* **17**, 29–43, 1986.

Grupp, F. "On the Mertens Conjecture for Cusp Forms." *Mathematika* **29**, 213–226, 1982.

Jurkat, W. and Peyerimhoff, A. "A Constructive Approach to Kronecker Approximation and Its Application to the Mertens Conjecture." *J. reine angew. Math.* **286/287**, 322–340, 1976.

Mertens, F. "Über eine zahlentheoretische Funktion." *Sitzungsber. Akad. Wiss. Wien IIa* **106**, 761–830, 1897.

Odlyzko, A. M. and te Riele, H. J. J. "Disproof of the Mertens Conjecture." *J. reine angew. Math.* **357**, 138–160, 1985.

Pintz, J. "An Effective Disproof of the Mertens Conjecture." *Astérique* **147–148**, 325–333 and 346, 1987.

te Riele, H. J. J. "Some Historical and Other Notes About the Mertens Conjecture and Its Recent Disproof." *Nieuw Arch. Wisk.* **3**, 237–243, 1985.

Mertens Constant

A constant related to the TWIN PRIMES CONSTANT which appears in the FORMULA for the sum of inverse PRIMES

$$\sum_{p \text{ prime}}^{x} \frac{1}{p} = \ln\ln x + B_1 + o(1) \tag{1}$$

which is given by

$$B_1 = \gamma + \sum_{p \text{ prime}} \left[\ln(1-p^{-1}) + \frac{1}{p}\right] \approx 0.261497. \tag{2}$$

Flajolet and Vardi (1996) show that

$$e^{B_1} = e^{\gamma} \prod_{m=2}^{\infty} \zeta(m)^{\mu(m)/m}, \tag{3}$$

where γ is the EULER-MASCHERONI CONSTANT, $\zeta(n)$ is the RIEMANN ZETA FUNCTION, and $\mu(n)$ is the MÖBIUS FUNCTION. The constant B_1 also occurs in the SUMMATORY FUNCTION of the number of DISTINCT PRIME FACTORS,

$$\sum_{k=2}^{n} \omega(k) = n\ln\ln n + B_1 n + o(n) \tag{4}$$

(Hardy and Wright 1979, p. 355).

The related constant

$$B_2 = \gamma + \sum_{p \text{ prime}} \left[\ln(1-p^{-1}) + \frac{1}{p-1}\right] \approx 1.034653 \tag{5}$$

appears in the SUMMATORY FUNCTION of the DIVISOR FUNCTION $\sigma_0(n) = \Omega(n)$,

$$\sum_{k=2}^{n} \Omega(k) = n\ln\ln n + B_2 + o(n) \tag{6}$$

(Hardy and Wright 1979, p. 355).

see also BRUN'S CONSTANT, PRIME NUMBER, TWIN PRIMES CONSTANT

References

Flajolet, P. and Vardi, I. "Zeta Function Expansions of Classical Constants." Unpublished manuscript. 1996. `http://pauillac.inria.fr/algo/flajolet/Publications/landau.ps`.

Hardy, G. H. and Weight, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Oxford University Press, pp. 351 and 355, 1979.

Mertens Function

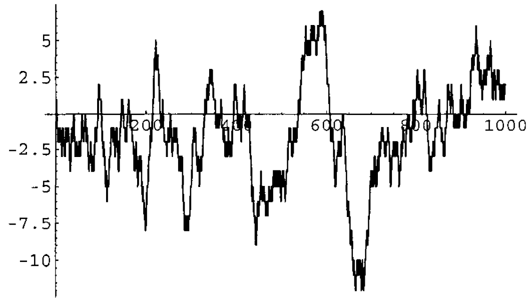

The summary function

$$M(n) \equiv \sum_{k=1}^{n} \mu(k) = \frac{6}{\pi^2} n + \mathcal{O}(\sqrt{n}),$$

where $\mu(n)$ is the MÖBIUS FUNCTION. The first few values are 1, 0, −1, −1, −2, −1, −2, −2, −2, −1, −2, −2, ... (Sloane's A002321). The first few values of n at which $M(n) = 0$ are 2, 39, 40, 58, 65, 93, 101, 145, 149, 150, ... (Sloane's A028442).

Mertens function obeys

$$\sum_{n=1}^{x} M\left(\frac{x}{n}\right) = 1$$

(Lehman 1960). The analytic form is unsolved, although MERTENS CONJECTURE that

$$|M(x)| < x^{1/2}$$

has been disproved.

Lehman (1960) gives an algorithm for computing $M(x)$ with $\mathcal{O}(x^{2/3+\epsilon})$ operations, while the Lagarias-Odlyzko (1987) algorithm for computing the PRIME COUNTING FUNCTION $\pi(x)$ can be modified to give $M(x)$ in $\mathcal{O}(x^{3/5+\epsilon})$ operations.

see also MERTENS CONJECTURE, MÖBIUS FUNCTION

References
Lagarias, J. and Odlyzko, A. "Computing $\pi(x)$: An Analytic Method." J. Algorithms **8**, 173–191, 1987.
Lehman, R. S. "On Liouville's Function." *Math. Comput.* **14**, 311–320, 1960.
Odlyzko, A. M. and te Riele, H. J. J. "Disproof of the Mertens Conjecture." *J. reine angew. Math.* **357**, 138–160, 1985.
Sloane, N. J. A. Sequence A028442/M002321 in "An On-Line Version of the Encyclopedia of Integer Sequences."0102

Mertens Theorem

$$\lim_{x\to\infty} \frac{\prod_{\substack{2\le p\le x\\ p\text{ prime}}} \left(1-\frac{1}{p}\right)}{\frac{e^{-\gamma}}{\ln x}} = 1,$$

where γ is the EULER-MASCHERONI CONSTANT and $e^{-\gamma} = 0.56145\ldots$.

References
Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Oxford University Press, p. 351, 1979.
Riesel, H. *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, pp. 66–67, 1994.

Mertz Apodization Function

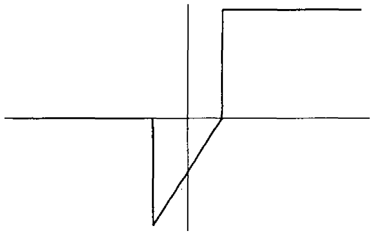

An asymmetrical APODIZATION FUNCTION defined by

$$M(x,b,d) = \begin{cases} 0 & \text{for } x < -b \\ (x-b)/(2b) & \text{for } -b < x < b \\ 1 & \text{for } b < x < b+2d \\ 0 & \text{for } x < b+2d, \end{cases}$$

where the two-sided portion is $2b$ long (total) and the one-sided portion is $b+2d$ long (Schnopper and Thompson 1974, p. 508). The APPARATUS FUNCTION is

$$M_A(k,b,d) = \frac{\sin[2\pi k(b+2d)}{2\pi k} + i\left\{\frac{\cos[2\pi k(b+2d)]}{2\pi k} - \frac{\sin(2b)}{4\pi^2 k^2 b}\right\}.$$

References
Schnopper, H. W. and Thompson, R. I. "Fourier Spectrometers." In *Methods of Experimental Physics* **12A**. New York: Academic Press, pp. 491–529, 1974.

Mesh Size

When a CLOSED INTERVAL $[a,b]$ is partitioned by points $a < x_1 < x_2 < \ldots < x_{n-1} < b$, the lengths of the resulting intervals between the points are denoted Δx_1, Δx_2, ..., Δx_n, and the value $\max \Delta x_k$ is called the mesh size of the partition.

see also INTEGRAL, LOWER SUM, RIEMANN INTEGRAL, UPPER SUM

Mesokurtic

A distribution with zero KURTOSIS ($\gamma_2 = 0$).

see also KURTOSIS, LEPTOKURTIC

Metabiaugmented Dodecahedron

see JOHNSON SOLID

Metabiaugmented Hexagonal Prism

see JOHNSON SOLID

Metabiaugmented Truncated Dodecahedron

see JOHNSON SOLID

Metabidiminished Icosahedron

see JOHNSON SOLID

Metabidiminished Rhombicosidodecahedron

see JOHNSON SOLID

Metabigyrate Rhombicosidodecahedron

see JOHNSON SOLID

Metadrome

A metadrome is a number whose HEXADECIMAL digits are in strict ascending order. The first few are 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, ... (Sloane's A023784).

see also HEXADECIMAL

References
Sloane, N. J. A. Sequence A023784 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Metagyrate Diminished Rhombicosidodecahedron

see JOHNSON SOLID

Metalogic

see METAMATHEMATICS

Metamathematics

The branch of LOGIC dealing with the study of the combination and application of mathematical symbols, sometimes called METALOGIC. Metamathematics is the study of MATHEMATICS itself, and one of its primary goals is to determine the nature of mathematical reasoning (Hofstadter 1989).

see also LOGIC, MATHEMATICS

References
Birkhoff, G. and Mac Lane, S. *A Survey of Modern Algebra, 3rd ed.* New York: Macmillan, p. 326, 1965.
Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, p. 23, 1989.

Method

A particular way of doing something, sometimes also called an ALGORITHM or PROCEDURE. (According to Petkovšek *et al.* (1996), "a method is a trick that has worked at least twice.")

see also ADAMS-BASHFORTH-MOULTON METHOD, ADAMS' METHOD, BACKUS-GILBERT METHOD, BADER-DEUFLHARD METHOD, BAILEY'S METHOD, BAIRSTOW'S METHOD, BRENT'S FACTORIZATION METHOD, BRENT'S METHOD, CIRCLE METHOD, CONJUGATE GRADIENT METHOD, CRISS-CROSS METHOD, CROUT'S METHOD, DE LA LOUBERE'S METHOD, DIXON'S FACTORIZATION METHOD, DIXON'S RANDOM SQUARES FACTORIZATION METHOD, ELLIPTIC CURVE FACTORIZATION METHOD, EULER'S FACTORIZATION METHOD, EXCLUDENT FACTORIZATION METHOD, EXHAUSTION METHOD, FALSE POSITION METHOD, FERMAT'S FACTORIZATION METHOD, FROBENIUS METHOD, GILL'S METHOD, GOSPER'S METHOD, GRAEFFE'S METHOD, GREENE'S METHOD, HALLEY'S METHOD, HORNER'S METHOD, HUTTON'S METHOD, JACOBI METHOD, KAPS-RENTROP METHODS, LAGUERRE'S METHOD, LAMBERT'S METHOD, LEGENDRE'S FACTORIZATION METHOD, LEHMER METHOD, LEHMER-SCHUR METHOD, LENSTRA ELLIPTIC CURVE METHOD, LIN'S METHOD, LOZENGE METHOD, LUX METHOD, MAPES' METHOD, MAXIMUM ENTROPY METHOD, MILNE'S METHOD, MULLER'S METHOD, NEWTON'S METHOD, NEWTON-RAPHSON METHOD, NUMBER FIELD SIEVE FACTORIZATION METHOD, OVERLAPPING RESONANCE METHOD, POLLARD MONTE CARLO FACTORIZATION METHOD, POLLARD ρ FACTORIZATION METHOD, POLLARD $p-1$ FACTORIZATION METHOD, PREDICTOR-CORRECTOR METHODS, QUADRATIC SIEVE FACTORIZATION METHOD, RESONANCE OVERLAP METHOD, ROSENBROCK METHODS, RUNGE-KUTTA METHOD, SCHRÖDER'S METHOD, SECANT METHOD, SIAMESE METHOD, SIMPLEX METHOD, SNAKE OIL METHOD, SQUARE ROOT METHOD, STEEPEST DESCENT METHOD, TANGENT HYPERBOLAS METHOD, UNDETERMINED COEFFICIENTS METHOD, WILLIAMS $p+1$ FACTORIZATION METHOD, WYNN'S EPSILON METHOD

References
Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, p. 117, 1996.

Metric

A NONNEGATIVE function $g(x,y)$ describing the "DISTANCE" between neighboring points for a given SET. A metric satisfies the TRIANGLE INEQUALITY

$$g(x,y)+g(y,z)\geq g(x,z),$$

with equality IFF $x=y$, and is symmetric, so

$$g(x,y)=g(y,x).$$

A SET possessing a metric is called a METRIC SPACE. When viewed as a TENSOR, the metric is called a METRIC TENSOR.

see also CAYLEY-KLEIN-HILBERT METRIC, DISTANCE, FUNDAMENTAL FORMS, HYPERBOLIC METRIC, METRIC ENTROPY, METRIC EQUIVALENCE PROBLEM, METRIC SPACE, METRIC TENSOR, PART METRIC, RIEMANNIAN METRIC, ULTRAMETRIC

References
Gray, A. "Metrics on Surfaces." Ch. 13 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 251–265, 1993.

Metric Entropy

Also known as Kolmogorov Entropy, Kolmogorov-Sinai Entropy, or KS Entropy. The metric entropy is 0 for nonchaotic motion and > 0 for Chaotic motion.

References

Ott, E. *Chaos in Dynamical Systems.* New York: Cambridge University Press, p. 138, 1993.

Metric Equivalence Problem

1. Find a complete system of invariants, or
2. decide when two Metrics differ only by a coordinate transformation.

The most common statement of the problem is, "Given Metrics g and g', does there exist a coordinate transformation from one to the other?" Christoffel and Lipschitz (1870) showed how to decide this question for two Riemannian Metrics.

The solution by E. Cartan requires computation of the 10th order Covariant Derivatives. The demonstration was simplified by A. Karlhede using the Tetrad formalism so that only seventh order Covariant Derivatives need be computed. However, in many common cases, the first or second-order Derivatives are Sufficient to answer the question.

References

Karlhede, A. and Lindström, U. "Finding Space-Time Geometries without Using a Metric." *Gen. Relativity Gravitation* **15**, 597–610, 1983.

Metric Space

A Set S with a global distance Function (the Metric g) which, for every two points x, y in S, gives the Distance between them as a Nonnegative Real Number $g(x,y)$. A metric space must also satisfy

1. $g(x,x) = 0$ Iff $x = y$,
2. $g(x,y) = g(y,x)$,
3. The Triangle Inequality $g(x,y) + g(y,z) \geq g(x,z)$.

References

Munkres, J. R. *Topology: A First Course.* Englewood Cliffs, NJ: Prentice-Hall, 1975.

Rudin, W. *Principles of Mathematical Analysis.* New York: McGraw-Hill, 1976.

Metric Tensor

A Tensor, also called a Riemannian Metric, which is symmetric and Positive Definite. Very roughly, the metric tensor g_{ij} is a function which tells how to compute the distance between any two points in a given Space. Its components can be viewed as multiplication factors which must be placed in front of the differential displacements dx_i in a generalized Pythagorean Theorem

$$ds^2 = g_{11} dx_1{}^2 + g_{12}\, dx_1\, dx_2 + g_{22}\, dx_2{}^2 + \ldots. \tag{1}$$

In Euclidean Space, $g_{ij} = \delta_{ij}$ where δ is the Kronecker Delta (which is 0 for $i \neq j$ and 1 for $i = j$), reproducing the usual form of the Pythagorean Theorem

$$ds^2 = dx_1{}^2 + dx_2{}^2 + \ldots. \tag{2}$$

The metric tensor is defined abstractly as an Inner Product of every Tangent Space of a Manifold such that the Inner Product is a symmetric, non-degenerate, bilinear form on a Vector Space. This means that it takes two Vectors $\mathbf{v}, \mathbf{w}$ as arguments and produces a Real Number $\langle \mathbf{v}, \mathbf{w} \rangle$ such that

$$\langle k\mathbf{v}, w \rangle = k \langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, k\mathbf{w} \rangle \tag{3}$$

$$\langle \mathbf{v} + \mathbf{w}, \mathbf{x} \rangle = \langle \mathbf{v}, \mathbf{x} \rangle + \langle \mathbf{w}, \mathbf{x} \rangle \tag{4}$$

$$\langle \mathbf{v}, \mathbf{w} + \mathbf{x} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{x} \rangle \tag{5}$$

$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle \tag{6}$$

$$\langle \mathbf{v}, \mathbf{v} \rangle \geq 0, \tag{7}$$

with equality Iff $\mathbf{v} = 0$.

In coordinate Notation (with respect to the basis),

$$g^{\alpha\beta} = \vec{e}^{\,\alpha} \cdot \vec{e}^{\,\beta} \tag{8}$$

$$g_{\alpha\beta} = \vec{e}_\alpha \cdot \vec{e}_\beta. \tag{9}$$

$$g_{\mu\nu} \equiv \frac{\partial \xi^\alpha}{\partial x^\mu} \frac{\partial \xi^\beta}{\partial x^\nu} \eta_{\alpha\beta}, \tag{10}$$

where $\eta_{\alpha\beta}$ is the Minkowski Metric. This can also be written

$$g = D^{\mathrm{T}} \eta D, \tag{11}$$

where

$$D_{\alpha\mu} \equiv \frac{\partial \xi^\alpha}{\partial x^\mu} \tag{12}$$

$$D_{\alpha\mu}{}^{\mathrm{T}} \equiv D_{\mu\alpha}. \tag{13}$$

$$\frac{\partial}{\partial x^m} g_{il} g^{lk} = \frac{\partial}{\partial x^m} \delta_i^k \tag{14}$$

gives

$$g_{il} \frac{\partial g^{lk}}{\partial x^m} = -g^{lk} \frac{\partial g_{il}}{\partial x^m}. \tag{15}$$

The metric is Positive Definite, so a metric's Discriminant is Positive. For a metric in 2-space,

$$g \equiv g_{11} g_{22} - g_{12}{}^2 > 0. \tag{16}$$

The Orthogonality of Contravariant and Covariant metrics stipulated by

$$g_{ik} g^{ij} = \delta_k^j \tag{17}$$

for $i = 1, \ldots, n$ gives n linear equations relating the $2n$ quantities g_{ij} and g^{ij}. Therefore, if n metrics are known, the others can be determined.

In 2-space,

$$g^{11} = \frac{g_{22}}{g} \tag{18}$$

$$g^{12} = g^{21} = -\frac{g_{12}}{g} \tag{19}$$

$$g^{22} = \frac{g_{11}}{g}. \tag{20}$$

If g is symmetric, then

$$g_{\alpha\beta} = g_{\beta\alpha} \tag{21}$$

$$g^{\alpha\beta} = g^{\beta\alpha}. \tag{22}$$

In Euclidean Space (and all other symmetric Spaces),

$$g_\alpha^\beta = g_\alpha^\beta = \delta_\alpha^\beta, \tag{23}$$

so

$$g_{\alpha\alpha} = \frac{1}{g^{\alpha\alpha}}. \tag{24}$$

The Angle ϕ between two parametric curves is given by

$$\cos\phi = \hat{\mathbf{r}}_1 \cdot \hat{\mathbf{r}}_2 = \frac{\mathbf{r}_1}{g_1} \cdot \frac{\mathbf{r}_2}{g_2} = \frac{g_{12}}{g_1 g_2}, \tag{25}$$

so

$$\sin\phi = \frac{\sqrt{g}}{g_1 g_2} \tag{26}$$

and

$$|\mathbf{r}_1 \times \mathbf{r}_2| = g_1 g_2 \sin\phi = \sqrt{g}. \tag{27}$$

The Line Element can be written

$$ds^2 = dx_i\,dx_i = g_{ij}\,dq_i\,dq_j \tag{28}$$

where Einstein Summation has been used. But

$$dx_i = \frac{\partial x_i}{\partial q_1}\,dq_1 + \frac{\partial x_i}{\partial q_2}\,dq_2 + \frac{\partial x_i}{\partial q_3}\,dq_3 = \frac{\partial x_i}{\partial q_j}\,dq_j, \tag{29}$$

so

$$g_{ij} = \sum_k \frac{\partial^2 x_k}{\partial q_i \partial q_j}. \tag{30}$$

For Orthogonal coordinate systems, $g_{ij} = 0$ for $i \neq j$, and the Line Element becomes (for 3-space)

$$\begin{aligned} ds^2 &= g_{11}\,dq_1{}^2 + g_{22}\,dq_2{}^2 + g_{33}\,dq_3{}^2 \\ &= (h_1\,dq_1)^2 + (h_2\,dq_2)^2 + (h_3\,dq_3)^2, \end{aligned} \tag{31}$$

where $h_i \equiv \sqrt{g_{ii}}$ are called the Scale Factors.

see also Curvilinear Coordinates, Discriminant (Metric), Lichnerowicz Conditions, Line Element, Metric, Metric Equivalence Problem, Minkowski Space, Scale Factor, Space

Mex

The Minimum excluded value. The mex of a Set S of Nonnegative Integers is the least Nonnegative Integer *not* in the set.

see also Mex Sequence

References

Guy, R. K. "Max and Mex Sequences." §E27 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 227–228, 1994.

Mex Sequence

A sequence defined from a Finite sequence a_0, a_1, ..., a_n by defining $a_{n+1} = \text{mex}_i(a_i + a_{n-i})$, where mex is the Mex (minimum excluded value).

see also Max Sequence, Mex

References

Guy, R. K. "Max and Mex Sequences." §E27 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 227–228, 1994.

Mian-Chowla Sequence

The sequence produced by starting with $a_1 = 1$ and applying the Greedy Algorithm in the following way: for each $k \geq 2$, let a_k be the least Integer exceeding a_{k-1} for which $a_j + a_k$ are all distinct, with $1 \leq j \leq k$. This procedure generates the sequence 1, 2, 4, 8, 13, 21, 31, 45, 66, 81, 97, 123, 148, 182, 204, 252, 290, ... (Sloane's A005282). The Reciprocal sum of the sequence,

$$S \equiv \sum_{i=1}^{\infty} \frac{1}{a_i},$$

satisfies

$$2.1568 \leq S \leq 2.1596.$$

see also A-Sequence, B_2-Sequence

References

Guy, R. K. "B_2-Sequences." §E28 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 228–229, 1994.

Sloane, N. J. A. Sequence A005282/M1094 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Mice Problem

n mice start at the corners of a regular n-gon of unit side length, each heading towards its closest neighboring mouse in a counterclockwise direction at constant speed. The mice each trace out a Spiral, meet in the center of the Polygon, and travel a distance

$$d_n = \frac{1}{1 - \cos\left(\frac{2\pi}{m}\right)}.$$

The first few values for $n = 2$, 3, ..., are

$$\tfrac{1}{2}, \tfrac{2}{3}, 1, \tfrac{1}{5}(5 + \sqrt{5}), 2, \frac{1}{1 - \cos\left(\frac{2\pi}{7}\right)},$$

$$2 + \sqrt{2}, \frac{1}{1 - \cos\left(\frac{2\pi}{9}\right)}, 3 + \sqrt{5}, \ldots,$$

giving the numerical values 0.5, 0.666667, 1, 1.44721, 2, 2.65597, 3.41421, 4.27432, 5.23607, ...

see also APOLLONIUS PURSUIT PROBLEM, PURSUIT CURVE, SPIRAL, TRACTRIX

References
Bernhart, A. "Polygons of Pursuit." *Scripta Math.* **24**, 23–50, 1959.
Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 201–204, 1979.

Mid-Arc Points

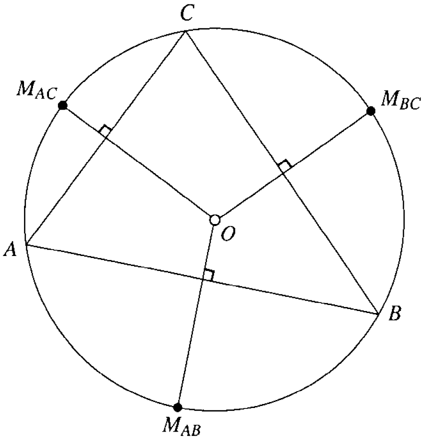

The mid-arc points M_{AB}, M_{AC}, and M_{BC} of a TRIANGLE ΔABC are the points on the CIRCUMCIRCLE of the triangle which lie half-way along each of the three ARCS determined by the vertices (Johnson 1929). These points arise in the definition of the FUHRMANN CIRCLE and FUHRMANN TRIANGLE, and lie on the extensions of the PERPENDICULAR BISECTORS of the triangle sides drawn from the CIRCUMCENTER O.

Kimberling (1988, 1994) and Kimberling and Veldkamp (1987) define the mid-arc points as the POINTS which have TRIANGLE CENTER FUNCTIONS

$$\alpha_1 = [\cos(\tfrac{1}{2}B) + \cos(\tfrac{1}{2}C)]\sec(\tfrac{1}{2}A)$$
$$\alpha_2 = [\cos(\tfrac{1}{2}B) + \cos(\tfrac{1}{2}C)]\csc(\tfrac{1}{2}A).$$

see also FUHRMANN CIRCLE, FUHRMANN TRIANGLE

References
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 228–229, 1929.
Kimberling, C. "Problem 804." *Nieuw Archief voor Wiskunde* **6**, 170, 1988.
Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.
Kimberling, C. and Veldkamp, G. R. "Problem 1160 and Solution." *Crux Math.* **13**, 298–299, 1987.

Midcircle

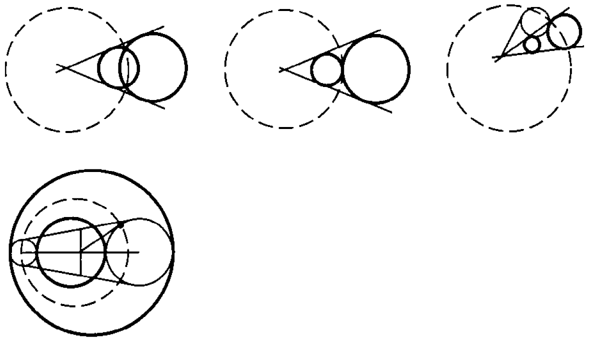

The midcircle of two given CIRCLES is the CIRCLE which would INVERT the circles into each other. Dixon (1991) gives constructions for the midcircle for four of the five possible configurations. In the case of the two given CIRCLES tangent to each other, there are two midcircles.

see also INVERSION, INVERSION CIRCLE

References
Dixon, R. *Mathographics.* New York: Dover, pp. 66–68, 1991.

Middlespoint

see MITTENPUNKT

Midpoint

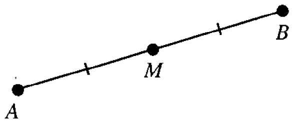

The point on a LINE SEGMENT dividing it into two segments of equal length. The midpoint of a line segment is easy to locate by first constructing a LENS using circular arcs, then connecting the cusps of the LENS. The point where the cusp-connecting line intersects the segment is then the midpoint (Pedoe 1995, p. xii). It is more challenging to locate the midpoint using only a COMPASS, but Pedoe (1995, pp. xviii–xix) gives one solution.

In a RIGHT TRIANGLE, the midpoint of the HYPOTENUSE is equidistant from the three VERTICES (Dunham 1990).

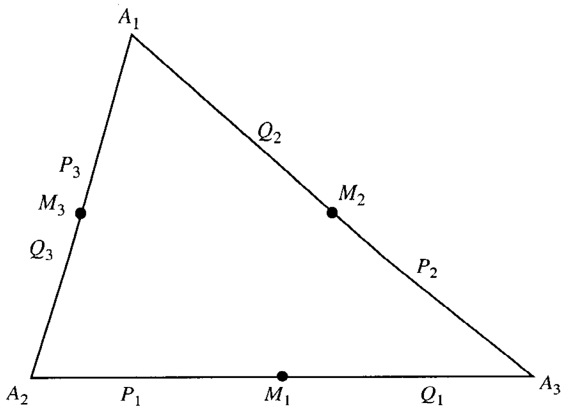

Given a Triangle $\Delta A_1A_2A_3$ with Area Δ, locate the midpoints M_i. Now inscribe two triangles $\Delta P_1P_2P_3$ and $\Delta Q_1Q_2Q_3$ with Vertices P_i and Q_i placed so that $\overline{P_iM_i} = \overline{Q_iM_i}$. Then $\Delta P_1P_2P_3$ and $\Delta Q_1Q_2Q_3$ have equal areas

$$\Delta_P = \Delta_Q = \Delta\left[1 - \left(\frac{m_1}{a_1} + \frac{m_2}{a_2} + \frac{m_3}{a_3}\right) + \frac{m_2m_2}{a_2a_3} + \frac{m_3m_1}{a_3a_1} + \frac{m_1m_2}{a_1a_2}\right],$$

where a_i are the sides of the original triangle and m_i are the lengths of the Medians (Johnson 1929).

see also Archimedes' Midpoint Theorem, Brocard Midpoint, Circle-Point Midpoint Theorem, Line Segment, Median (Triangle), Midpoint Ellipse

References

Dunham, W. *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 120–121, 1990.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, p. 80, 1929.

Pedoe, D. *Circles: A Mathematical View, rev. ed.* Washington, DC: Math. Assoc. Amer., 1995.

Midpoint Ellipse

The unique Ellipse tangent to the Midpoints of a Triangle's Legs. The midpoint ellipse has the maximum Area of any Inscribed Ellipse (Chakerian 1979). Under an Affine Transformation, the midpoint ellipse can be transformed into the Incircle of an Equilateral Triangle.

see also Affine Transformation, Ellipse, Incircle, Midpoint, Triangle

References

Central Similarities. University of Minnesota College Geometry Project. Distributed by International Film Bureau, Inc.

Chakerian, G. D. "A Distorted View of Geometry." Ch. 7 in *Mathematical Plums* (Ed. R. Honsberger). Washington, DC: Math. Assoc. Amer., pp. 135–136 and 145–146, 1979.

Pedoe, D. "Thinking Geometrically." *Amer. Math. Monthly* **77**, 711–721, 1970.

Midradius

The Radius of the Midsphere of a Polyhedron, also called the Interradius. For a Regular Polyhedron with Schläfli Symbol $\{q, p\}$, the Dual Polyhedron is $\{p, q\}$. Denote the Inradius r, midradius ρ, and Circumradius R, and let the side length be a. Then

$$r^2 = \left[a\csc\left(\frac{\pi}{p}\right)\right]^2 + R^2 = a^2 + \rho^2 \tag{1}$$

$$\rho^2 = \left[a\cot\left(\frac{\pi}{p}\right)\right]^2 + R^2. \tag{2}$$

For Regular Polyhedra and Uniform Polyhedra, the Dual Polyhedron has Circumradius ρ^2/r and Inradius ρ^2/R. Let θ be the Angle subtended by the Edge of an Archimedean Solid. Then

$$r = \tfrac{1}{2}a\cos(\tfrac{1}{2}\theta)\cot(\tfrac{1}{2}\theta) \tag{3}$$

$$\rho = \tfrac{1}{2}a\cot(\tfrac{1}{2}\theta) \tag{4}$$

$$R = \tfrac{1}{2}a\csc(\tfrac{1}{2}\theta), \tag{5}$$

so

$$r : \rho : R = \cos(\tfrac{1}{2}\theta) : 1 : \sec(\tfrac{1}{2}\theta) \tag{6}$$

(Cundy and Rollett 1989). Expressing the midradius in terms of the Inradius r and Circumradius R gives

$$\begin{aligned}\rho &= \tfrac{1}{2}\sqrt{2}\sqrt{r^2 + r\sqrt{r^2 + a^2}}\\ &= \sqrt{R^2 - \tfrac{1}{4}a^2}\end{aligned} \tag{7}$$

for an Archimedean Solid.

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 126–127, 1989.

Midrange

$$\text{midrange}[f(x)] \equiv \tfrac{1}{2}\{\max[f(x)] + \min[f(x)]\}.$$

see also Maximum, Mean, Median (Statistics), Minimum

Midsphere

The Sphere with respect to which the Vertices of a Polyhedron are the poles of the planes of the faces of the Dual Polyhedron (and vice versa). It touches all Edges of a Semiregular Polyhedron or Regular Polyhedron. It is also called the Intersphere or Reciprocating Sphere.

see also Circumsphere, Dual Polyhedron, Insphere

Midy's Theorem

If the period of a Repeating Decimal for a/p has an Even number of digits, the sum of the two halves is a string of 9s, where p is Prime and a/p is a Reduced Fraction.

see also Decimal Expansion, Repeating Decimal

References

Rademacher, H. and Toeplitz, O. *The Enjoyment of Mathematics: Selections from Mathematics for the Amateur.* Princeton, NJ: Princeton University Press, pp. 158–160, 1957.

Mikusiński's Problem

Is it possible to cover completely the surface of a SPHERE with congruent, nonoverlapping arcs of GREAT CIRCLES? Conway and Croft (1964) proved that it can be covered with half-open arcs, but not with open arcs. They also showed that the PLANE can be covered with congruent closed and half-open segments, but not with open ones.

References

Conway, J. H. and Croft, H. T. "Covering a Sphere with Great-Circle Arcs." *Proc. Cambridge Phil. Soc.* **60**, 787–900, 1964.

Gardner, M. "Point Sets on the Sphere." Ch. 12 in *Knotted Doughnuts and Other Mathematical Entertainments.* New York: W. H. Freeman, pp. 145–154, 1986.

Milin Conjecture

An INEQUALITY which IMPLIES the correctness of the ROBERTSON CONJECTURE (Milin 1971). de Branges (1985) proved this conjecture, which led to the proof of the full BIEBERBACH CONJECTURE.

see also BIEBERBACH CONJECTURE, ROBERTSON CONJECTURE

References

de Branges, L. "A Proof of the Bieberbach Conjecture." *Acta Math.* **154**, 137–152, 1985.

Milin, I. M. *Univalent Functions and Orthonormal Systems.* Providence, RI: Amer. Math. Soc., 1977.

Stewart, I. *From Here to Infinity: A Guide to Today's Mathematics.* Oxford, England: Oxford University Press, p. 165, 1996.

Mill

The n-roll mill curve is given by the equation

$$x^n - \binom{n}{2} x^{n-2} y^2 + \binom{n}{4} x^{n-4} y^4 - \ldots = a^n,$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT.

References

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 86, 1993.

Miller's Algorithm

For a catastrophically unstable recurrence in one direction, any seed values for consecutive x_j and x_{j+1} will converge to the desired sequence of functions in the opposite direction times an unknown normalization factor.

Miller-Aškinuze Solid

see ELONGATED SQUARE GYROBICUPOLA

Miller Cylindrical Projection

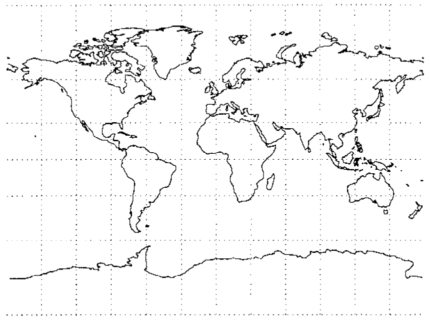

A MAP PROJECTION given by the following transformation,

$$x = \lambda - \lambda_0 \tag{1}$$

$$y = \tfrac{5}{4} \ln[\tan(\tfrac{1}{4}\pi + \tfrac{2}{5}\phi)] \tag{2}$$

$$= \tfrac{5}{4} \sinh^{-1}[\tan(\tfrac{4}{5}\phi)]. \tag{3}$$

Here x and y are the plane coordinates of a projected point, λ is the longitude of a point on the globe, λ_0 is central longitude used for the projection, and ϕ is the latitude of the point on the globe. The inverse FORMULAS are

$$\phi = \tfrac{5}{2} \tan^{-1}(e^{4y/5}) - \tfrac{5}{8}\pi = \tfrac{5}{4} \tan^{-1}[\sinh(\tfrac{4}{5}y)] \tag{4}$$

$$\lambda = \lambda_0 + x. \tag{5}$$

References

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 86–89, 1987.

Miller's Primality Test

If a number fails this test, it is not a PRIME. If the number passes, it *may* be a PRIME. A number passing Miller's test is called a STRONG PSEUDOPRIME to base a. If a number n does not pass the test, then it is called a WITNESS for the COMPOSITENESS of n. If n is an ODD, POSITIVE COMPOSITE NUMBER, then n passes Miller's test for at most $(n-1)/4$ bases with $1 \leq a \leq -1$ (Long 1995). There is no analog of CARMICHAEL NUMBERS for STRONG PSEUDOPRIMES.

The only COMPOSITE NUMBER less than 2.5×10^{13} which does not have 2, 3, 5, or 7 as a WITNESS is 3215031751. Miller showed that any composite n has a WITNESS less than $70(\ln n)^2$ if the RIEMANN HYPOTHESIS is true.

see also ADLEMAN-POMERANCE-RUMELY PRIMALITY TEST, STRONG PSEUDOPRIME

References

Long, C. T. Th. 4.21 in *Elementary Introduction to Number Theory, 3rd ed.* Prospect Heights, IL: Waveland Press, 1995.

Miller's Solid

see ELONGATED SQUARE GYROBICUPOLA

Milliard

In British, French, and German usage, one milliard equals 10^9. American usage does not have a number called the milliard, instead using the term BILLION to denote 10^9.

see also BILLION, LARGE NUMBER, MILLION, TRILLION

Millin Series

The series with sum

$$S' \equiv \sum_{n=0}^{\infty} \frac{1}{F_{2^n}} = \tfrac{1}{2}(7-\sqrt{5}),$$

where F_k is a FIBONACCI NUMBER (Honsberger 1985).

see also FIBONACCI NUMBER

References

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 135–137, 1985.

Million

The number $1,000,000 = 10^6$. While one million in America means the same thing as one million in Britain, the words BILLION, TRILLION, etc., refer to *different numbers* in the two naming systems. While Americans may say "Thanks a million" to express gratitude, Norwegians offer "Thanks a thousand" ("tusen takk").

see also BILLION, LARGE NUMBER, MILLIARD, THOUSAND, TRILLION

Mills' Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Mills (1947) proved the existence of a constant $\theta = 1.3064\ldots$ such that

$$\lfloor \theta^{3^n} \rfloor \tag{1}$$

is PRIME for all $n \geq 1$, where $\lfloor x \rfloor$ is the FLOOR FUNCTION. It is not, however, known if θ is IRRATIONAL. Mills' proof was based on the following theorem by Hoheisel (1930) and Ingham (1937). Let p_n be the nth PRIME, then there exists a constant K such that

$$p_{n+1} - p_n < K{p_n}^{5/8} \tag{2}$$

for all n. This has more recently been strengthened to

$$p_{n+1} - p_n < K{p_n}^{1051/1920} \tag{3}$$

(Mozzochi 1986). If the RIEMANN HYPOTHESIS is true, then Cramér (1937) showed that

$$p_{n+1} - p_n = \mathcal{O}(\ln p_n \sqrt{p_n}\,) \tag{4}$$

(Finch).

Hardy and Wright (1979) point out that, despite the beauty of such FORMULAS, they do not have any practical consequences. In fact, unless the exact value of θ is known, the PRIMES themselves must be known in advance to determine θ. A generalization of Mills' theorem to an arbitrary sequence of POSITIVE INTEGERS is given as an exercise by Ellison and Ellison (1985). Consequently, infinitely many values for θ other than the number $1.3064\ldots$ are possible.

References

Caldwell, C. "Mills' Theorem—A Generalization." `http://www.utm.edu/research/primes/notes/proofs/A3n.html`.

Ellison, W. and Ellison, F. *Prime Numbers.* New York: Wiley, pp. 31–32, 1985.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/mills/mills.html`.

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, 1979.

Mills, W. H. "A Prime-Representing Function." *Bull. Amer. Math. Soc.* **53**, 604, 1947.

Mozzochi, C. J. "On the Difference Between Consecutive Primes." *J. Number Th.* **24**, 181–187, 1986.

Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, pp. 135 and 191–193, 1989.

Ribenboim, P. *The Little Book of Big Primes.* New York: Springer-Verlag, pp. 109–110, 1991.

Milne's Method

A PREDICTOR-CORRECTOR METHOD for solution of ORDINARY DIFFERENTIAL EQUATIONS. The third-order equations for predictor and corrector are

$$\begin{aligned} y_{n+1} &= y_{n-3} + \tfrac{4}{3}h(2y'_n - y'_{n-1} + 2y'_{n-2}) + \mathcal{O}(h^5) \\ y_{n+1} &= y_{n-1} + \tfrac{1}{3}h(y'_{n-1} + 4y'_n + y'_{n+1}) + \mathcal{O}(h^5). \end{aligned}$$

Abramowitz and Stegun (1972) also give the fifth order equations and formulas involving higher derivatives.

see also ADAMS' METHOD, GILL'S METHOD, PREDICTOR-CORRECTOR METHODS, RUNGE-KUTTA METHOD

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 896–897, 1972.

Milnor's Conjecture

The UNKNOTTING NUMBER for a TORUS KNOT (p,q) is $(p-1)(q-1)/2$. This 40-year-old CONJECTURE was proved (Adams 1994) in Kronheimer and Mrowka (1993, 1995).

see also TORUS KNOT, UNKNOTTING NUMBER

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, p. 113, 1994.

Kronheimer, P. B. and Mrowka, T. S. "Gauge Theory for Embedded Surfaces. I." *Topology* **32**, 773–826, 1993.

Kronheimer, P. B. and Mrowka, T. S. "Gauge Theory for Embedded Surfaces. II." *Topology* **34**, 37–97, 1995.

Milnor's Theorem

If a COMPACT MANIFOLD M has NONNEGATIVE RICCI CURVATURE, then its FUNDAMENTAL GROUP has at most POLYNOMIAL growth. On the other hand, if M has NEGATIVE curvature, then its FUNDAMENTAL GROUP has exponential growth in the sense that $n(\lambda)$ grows exponentially, where $n(\lambda)$ is (essentially) the number of different "words" of length λ which can be made in the FUNDAMENTAL GROUP.

References
Chavel, I. *Riemannian Geometry: A Modern Introduction.* New York: Cambridge University Press, 1994.

Minimal Cover

A minimal cover is a COVER for which removal of one member destroys the covering property. Let $\mu(n,k)$ be the number of minimal covers of $\{1,\ldots,n\}$ with k members. Then

$$\mu(n,k)=\frac{1}{k!}\sum_{m=k}^{\alpha_k}\binom{2^k-k-1}{m-k}m!s(n,m),$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT, $s(n,m)$ is a STIRLING NUMBER OF THE SECOND KIND, and

$$\alpha_k=\min(n,2^k-1).$$

Special cases include $\mu(n,1)=1$ and $\mu(n,2)=s(n+1,3)$.

k	1	2	3	4	5	6	7
Sloane		000392	003468	016111			
n							
1	1						
2	1	1					
3	1	6	1				
4	1	25	22	1			
5	1	90	305	65	1		
6	1	301	3410	2540	171	1	
7	1	966	33621	77350	17066	420	1

see also COVER, LEW k-GRAM, STIRLING NUMBER OF THE SECOND KIND

References
Hearne, T. and Wagner, C. "Minimal Covers of Finite Sets." *Disc. Math.* **5**, 247–251, 1973.
Macula, A. J. "Lewis Carroll and the Enumeration of Minimal Covers." *Math. Mag.* **68**, 269–274, 1995.

Minimal Discriminant

see FREY CURVE

Minimal Matrix

A MATRIX with 0 DETERMINANT whose DETERMINANT becomes NONZERO when any element on or below the diagonal is changed from 0 to 1. An example is

$$\mathsf{M}=\begin{bmatrix}1 & -1 & 0 & 0\\ 0 & 0 & -1 & 0\\ 1 & 1 & 1 & -1\\ 0 & 0 & 1 & 0\end{bmatrix}.$$

There are 2^{n-1} minimal SPECIAL MATRICES of size $n\times n$.

see also SPECIAL MATRIX

References
Knuth, D. E. "Problem 10470." *Amer. Math. Monthly* **102**, 655, 1995.

Minimal Residue

The value b or $b-m$, whichever is smaller in ABSOLUTE VALUE, where $a\equiv b \pmod{m}$.

see also RESIDUE (CONGRUENCE)

Minimal Set

A SET for which the dynamics can be generated by the dynamics on any subset.

Minimal Surface

Minimal surfaces are defined as surfaces with zero MEAN CURVATURE, and therefore satisfy LAGRANGE'S EQUATION

$$(1+{f_y}^2)f_{xx}+2f_xf_yf_{xy}+(1+{f_x}^2)f_{yy}=0.$$

Minimal surfaces may also be characterized as surfaces of minimal AREA for given boundary conditions. A PLANE is a trivial MINIMAL SURFACE, and the first nontrivial examples (the CATENOID and HELICOID) were found by Meusnier in 1776 (Meusnier 1785).

Euler proved that a minimal surface is planar IFF its GAUSSIAN CURVATURE is zero at every point so that it is locally SADDLE-shaped. The EXISTENCE of a solution to the general case was independently proven by Douglas (1931) and Radó (1933), although their analysis could not exclude the possibility of singularities. Osserman (1970) and Gulliver (1973) showed that a minimizing solution *cannot* have singularities.

The only known complete (boundaryless), embedded (no self-intersections) minimal surfaces of finite topology known for 200 years were the CATENOID, HELICOID, and PLANE. Hoffman discovered a three-ended GENUS 1 minimal embedded surface, and demonstrated the existence of an infinite number of such surfaces. A four-ended embedded minimal surface has also been found. L. Bers proved that any finite isolated SINGULARITY of a single-valued parameterized minimal surface is removable.

A surface can be parameterized using a ISOTHERMAL PARAMETERIZATION. Such a parameterization is minimal if the coordinate functions x_k are HARMONIC, i.e., $\phi_k(\zeta)$ are ANALYTIC. A minimal surface can therefore be defined by a triple of ANALYTIC FUNCTIONS such that $\phi_k\phi_k=0$. The REAL parameterization is then obtained as

$$x_k=\Re\int\phi_k(\zeta)\,d\zeta. \tag{1}$$

But, for an ANALYTIC FUNCTION f and a MEROMORPHIC function g, the triple of functions

$$\phi_1(\zeta) = f(1-g^2) \tag{2}$$
$$\phi_2(\zeta) = if(1+g^2) \tag{3}$$
$$\phi_3(\zeta) = 2fg \tag{4}$$

are ANALYTIC as long as f has a zero of order $\geq m$ at every POLE of g of order m. This gives a minimal surface in terms of the ENNEPER-WEIERSTRAß PARAMETERIZATION

$$\Re \int \begin{bmatrix} f(1-g^2) \\ if(1+g^2) \\ 2fg \end{bmatrix} d\zeta. \tag{5}$$

see also BERNSTEIN MINIMAL SURFACE THEOREM, CALCULUS OF VARIATIONS, CATALAN'S SURFACE, CATENOID, COSTA MINIMAL SURFACE, ENNEPER-WEIERSTRAß PARAMETERIZATION, FLAT SURFACE, HENNEBERG'S MINIMAL SURFACE, HOFFMAN'S MINIMAL SURFACE, IMMERSED MINIMAL SURFACE, LICHTENFELS SURFACE, MAEDER'S OWL MINIMAL SURFACE, NIRENBERG'S CONJECTURE, PARAMETERIZATION, PLATEAU'S PROBLEM, SCHERK'S MINIMAL SURFACES, TRINOID, UNDULOID

References

Dickson, S. "Minimal Surfaces." *Mathematica J.* **1**, 38–40, 1990.

Dierkes, U.; Hildebrandt, S.; Küster, A.; and Wohlraub, O. *Minimal Surfaces, 2 vols. Vol. 1: Boundary Value Problems. Vol. 2: Boundary Regularity.* Springer-Verlag, 1992.

do Carmo, M. P. "Minimal Surfaces." §3.5 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 41–43, 1986.

Douglas, J. "Solution of the Problem of Plateau." *Trans. Amer. Math. Soc.* **33**, 263–321, 1931.

Fischer, G. (Ed.). Plates 93 and 96 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 89 and 96, 1986.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 280, 1993.

Gulliver, R. "Regularity of Minimizing Surfaces of Prescribed Mean Curvature." *Ann. Math.* **97**, 275–305, 1973.

Hoffman, D. "The Computer-Aided Discovery of New Embedded Minimal Surfaces." *Math. Intell.* **9**, 8–21, 1987.

Hoffman, D. and Meeks, W. H. III. *The Global Theory of Properly Embedded Minimal Surfaces.* Amherst, MA: University of Massachusetts, 1987.

Lagrange. "Essai d'une nouvelle méthode pour déterminer les maxima et les minima des formules intégrales indéfinies." 1776.

Meusnier, J. B. "Mémoire sur la courbure des surfaces." *Mém. des savans étrangers* **10** (lu 1776), 477–510, 1785.

Nitsche, J. C. C. *Introduction to Minimal Surfaces.* Cambridge, England: Cambridge University Press, 1989.

Osserman, R. *A Survey of Minimal Surfaces.* New York: Van Nostrand Reinhold, 1969.

Osserman, R. "A Proof of the Regularity Everywhere of the Classical Solution to Plateau's Problem." *Ann. Math.* **91**, 550–569, 1970.

Radó, T. "On the Problem of Plateau." *Ergeben. d. Math. u. ihrer Grenzgebiete.* Berlin: Springer-Verlag, 1933.

Minimax Approximation

A minimization of the MAXIMUM error for a fixed number of terms.

Minimax Polynomial

The approximating POLYNOMIAL which has the smallest maximum deviation from the true function. It is closely approximated by the CHEBYSHEV POLYNOMIALS OF THE FIRST KIND.

Minimax Theorem

The fundamental theorem of GAME THEORY which states that every FINITE, ZERO-SUM, two-person GAME has optimal MIXED STRATEGIES. It was proved by John von Neumann in 1928.

Formally, let **X** and **Y** be MIXED STRATEGIES for players A and B. Let A be the PAYOFF MATRIX. Then

$$\max_X \min_Y \mathbf{X}^{\mathrm{T}}\mathsf{A}\mathbf{Y} = \min_Y \max_X \mathbf{X}^{\mathrm{T}}\mathsf{A}\mathbf{Y} = v,$$

where v is called the VALUE of the GAME and **X** and **Y** are called the solutions. It also turns out that if there is more than one optimal MIXED STRATEGY, there are infinitely many.

see also MIXED STRATEGY

References

Willem, M. *Minimax Theorem.* Boston, MA: Birkhäuser, 1996.

Minimum

The smallest value of a set, function, etc. The minimum value of a set of elements $A = \{a_i\}_{i=1}^N$ is denoted $\min A$ or $\min_i a_i$, and is equal to the first element of a sorted (i.e., ordered) version of A. For example, given the set {3, 5, 4, 1}, the sorted version is {1, 3, 4, 5}, so the minimum is 1. The MAXIMUM and minimum are the simplest ORDER STATISTICS.

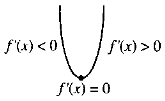

minimum

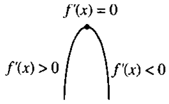

maximum

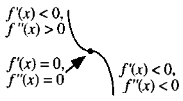

stationary point

A continuous FUNCTION may assume a minimum at a single point or may have minima at a number of points. A GLOBAL MINIMUM of a FUNCTION is the smallest value in the entire RANGE of the FUNCTION, while a LOCAL MINIMUM is the smallest value in some local neighborhood.

For a function $f(x)$ which is CONTINUOUS at a point x_0, a NECESSARY but not SUFFICIENT condition for $f(x)$ to have a RELATIVE MINIMUM at $x = x_0$ is that x_0 be a CRITICAL POINT (i.e., $f(x)$ is either not DIFFERENTIABLE at x_0 or x_0 is a STATIONARY POINT, in which case $f'(x_0) = 0$).

The FIRST DERIVATIVE TEST can be applied to CONTINUOUS FUNCTIONS to distinguish minima from MAXIMA. For twice differentiable functions of one variable, $f(x)$, or of two variables, $f(x, y)$, the SECOND DERIVATIVE TEST can sometimes also identify the nature of an EXTREMUM. For a function $f(x)$, the EXTREMUM TEST succeeds under more general conditions than the SECOND DERIVATIVE TEST.

see also CRITICAL POINT, EXTREMUM, FIRST DERIVATIVE TEST, GLOBAL MAXIMUM, INFLECTION POINT, LOCAL MAXIMUM, MAXIMUM, MIDRANGE, ORDER STATISTIC, SADDLE POINT (FUNCTION), SECOND DERIVATIVE TEST, STATIONARY POINT

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 14, 1972.
Brent, R. P. *Algorithms for Minimization Without Derivatives.* Englewood Cliffs, NJ: Prentice-Hall, 1973.
Nash, J. C. "Descent to a Minimum I–II: Variable Metric Algorithms." Chs. 15–16 in *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation, 2nd ed.* Bristol, England: Adam Hilger, pp. 186–206, 1990.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Minimization or Maximization of Functions." Ch. 10 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 387–448, 1992.
Tikhomirov, V. M. *Stories About Maxima and Minima.* Providence, RI: Amer. Math. Soc., 1991.

Minkowski-Bouligand Dimension

In many cases, the HAUSDORFF DIMENSION correctly describes the correction term for a resonator with FRACTAL PERIMETER in Lorentz's conjecture. However, in general, the proper dimension to use turns out to be the Minkowski-Bouligand dimension (Schroeder 1991).

Let $F(r)$ be the AREA traced out by a small CIRCLE with RADIUS r following a fractal curve. Then, providing the LIMIT exists,

$$D_M \equiv \lim_{r\to 0} \frac{\ln F(r)}{-\ln r} + 2$$

(Schroeder 1991). It is conjectured that for all strictly self-similar fractals, the Minkowski-Bouligand dimension is equal to the HAUSDORFF DIMENSION D; otherwise $D_M > D$.

see also HAUSDORFF DIMENSION

References
Berry, M. V. "Diffractals." *J. Phys.* **A12**, 781–797, 1979.
Hunt, F. V.; Beranek, L. L.; and Maa, D. Y. "Analysis of Sound Decay in Rectangular Rooms." *J. Acoust. Soc. Amer.* **11**, 80–94, 1939.
Lapidus, M. L. and Fleckinger-Pellé, J. "Tambour fractal: vers une résolution de la conjecture de Weyl-Berry put les valeurs propres du laplacien." *Compt. Rend. Acad. Sci. Paris Math. Sér 1* **306**, 171–175, 1988.
Schroeder, M. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise.* New York: W. H. Freeman, pp. 41–45, 1991.

Minkowski Convex Body Theorem

A bounded plane convex region symmetric about a LATTICE POINT and with AREA > 4 must contain at least three LATTICE POINTS in the interior. In n-D, the theorem can be generalized to a region with AREA $> 2^n$, which must contain at least three LATTICE POINTS. The theorem can be derived from BLICHFELDT'S THEOREM.

see also BLICHFELDT'S THEOREM

Minkowski Geometry

see MINKOWSKI SPACE

Minkowski-Hlawka Theorem

There exist lattices in n-D having HYPERSPHERE PACKING densities satisfying

$$\eta \geq \frac{\zeta(n)}{2^{n-1}},$$

where $\zeta(n)$ is the RIEMANN ZETA FUNCTION. However, the proof of this theorem is nonconstructive and it is still not known how to actually construct packings that are this dense.

see also HERMITE CONSTANTS, HYPERSPHERE PACKING

References
Conway, J. H. and Sloane, N. J. A. *Sphere Packings, Lattices, and Groups, 2nd ed.* New York: Springer-Verlag, pp. 14–16, 1993.

Minkowski Integral Inequality

If $p > 1$, then

$$\left[\int_a^b |f(x)+g(x)|^p\,dx\right]^{1/p} \leq \left[\int_a^b |f(x)|^p\,dx\right]^{1/p} + \left[\int_a^b |g(x)|^p\,dx\right]^{1/p}.$$

see also MINKOWSKI SUM INEQUALITY

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 11, 1972.
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1099, 1993.
Hardy, G. H.; Littlewood, J. E.; and Pólya, G. *Inequalities, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 146–150, 1988.
Minkowski, H. *Geometrie der Zahlen, Vol. 1.* Leipzig, Germany: pp. 115–117, 1896.
Sansone, G. *Orthogonal Functions, rev. English ed.* New York: Dover, p. 33, 1991.

Minkowski Measure

The Minkowski measure of a bounded, CLOSED SET is the same as its LEBESGUE MEASURE.

References

Ko, K.-I. "A Polynomial-Time Computable Curve whose Interior has a Nonrecursive Measure." *Theoret. Comput. Sci.* **145**, 241–270, 1995.

Minkowski Metric

In CARTESIAN COORDINATES,

$$ds^2 = dx^2 + dy^2 + dz^2 \tag{1}$$

$$d\tau^2 = -c^2\,dt^2 + dx^2 + dy^2 + dz^2, \tag{2}$$

and

$$g_{\alpha\beta} \equiv \eta_{\alpha\beta} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{3}$$

In SPHERICAL COORDINATES,

$$ds^2 = dr^2 + r^2\,d\theta^2 + r^2 \sin^2\theta\,d\phi^2 \tag{4}$$

$$d\tau^2 = -c^2\,dt^2 + dr^2 + r^2\,d\theta^2 + r^2 \sin^2\theta\,d\phi^2, \tag{5}$$

and

$$g = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2\sin^2\theta \end{bmatrix}. \tag{6}$$

see also LORENTZ TRANSFORMATION, MINKOWSKI SPACE

Minkowski Sausage

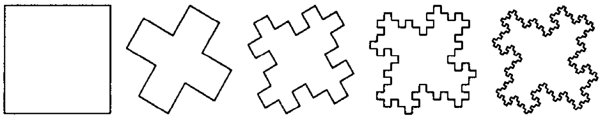

A FRACTAL created from the base curve and motif illustrated below.

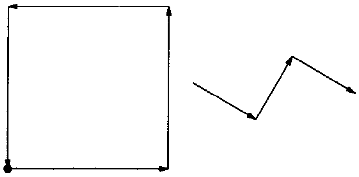

The number of segments after the nth iteration is

$$N_n = 8^n,$$

and

$$\epsilon_n = \left(\tfrac{1}{4}\right)^n,$$

so the CAPACITY DIMENSION is

$$D \equiv -\lim_{n\to\infty} \frac{\ln N_n}{\ln \epsilon_n} = -\lim_{n\to\infty} \frac{\ln 8^n}{\ln 4^n} = \frac{\ln 8}{\ln 4} = \frac{3\ln 2}{2\ln 2} = \frac{3}{2}.$$

References

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 37–38 and 42, 1991.

Peitgen, H.-O. and Saupe, D. (Eds.). *The Science of Fractal Images.* New York: Springer-Verlag, p. 283, 1988.

Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m`.

Minkowski Space

A 4-D space with the MINKOWSKI METRIC. Alternatively, it can be considered to have a EUCLIDEAN METRIC, but with its VECTORS defined by

$$\begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} ict \\ x \\ y \\ z \end{bmatrix}, \tag{1}$$

where c is the speed of light. The METRIC is DIAGONAL with

$$g_{\alpha\alpha} = \frac{1}{g_{\alpha\alpha}}, \tag{2}$$

so

$$\eta^{\beta\delta} = \eta_{\beta\delta}. \tag{3}$$

Let Λ be the TENSOR for a LORENTZ TRANSFORMATION. Then

$$\eta^{\beta\delta}\Lambda^{\gamma}{}_{\delta} = \Lambda^{\beta\gamma} \tag{4}$$

$$\eta_{\alpha\gamma}\Lambda^{\beta\gamma} = \Lambda_{\alpha}{}^{\beta} \tag{5}$$

$$\Lambda_{\alpha}{}^{\beta} = \eta_{\alpha\gamma}\Lambda^{\beta\gamma} = \eta_{\alpha\gamma}\eta^{\beta\delta}\Lambda^{\gamma}{}_{\delta}. \tag{6}$$

The NECESSARY and SUFFICIENT conditions for a metric $g_{\mu\nu}$ to be equivalent to the Minkowski metric $\eta_{\alpha\beta}$ are that the RIEMANN TENSOR vanishes everywhere ($R^{\lambda}{}_{\mu\nu\kappa} = 0$) and that at some point $g^{\mu\nu}$ has three POSITIVE and one NEGATIVE EIGENVALUES.

see also LORENTZ TRANSFORMATION, MINKOWSKI METRIC

References

Thompson, A. C. *Minkowski Geometry.* New York: Cambridge University Press, 1996.

Minkowski Sum

The sum of sets A and B in a VECTOR SPACE, equal to $\{a + b : a \in A, b \in B\}$.

Minkowski Sum Inequality

If $p > 1$ and a_k, $b_k > 0$, then

$$\left[\sum_{k=1}^{n} (a_k + b_k)^p\right]^{1/p} \leq \left(\sum_{k=1}^{n} {a_k}^p\right)^{1/p} + \left(\sum_{k=1}^{n} {b_k}^p\right)^{1/p}.$$

Equality holds IFF the sequences a_1, a_2, ... and b_1, b_2, ... are proportional.

see also MINKOWSKI INTEGRAL INEQUALITY

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 11, 1972.

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1092, 1979.

Hardy, G. H.; Littlewood, J. E.; and Pólya, G. *Inequalities, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 24–26, 1988.

Minor

The reduced DETERMINANT of a DETERMINANT EXPANSION, denoted M_{ij}, which is formed by omitting the ith row and jth column.

see also COFACTOR, DETERMINANT, DETERMINANT EXPANSION BY MINORS

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 169–170, 1985.

Minor Axis

see SEMIMINOR AXIS

Minor Graph

A "minor" is a sort of SUBGRAPH and is what Kuratowski means when he says "contain." It is roughly a small graph which can be mapped into the big one without merging VERTICES.

Minus

The operation of SUBTRACTION, i.e., a minus b. The operation is denoted $a - b$. The MINUS SIGN "$-$" is also used to denote a NEGATIVE number, i.e., $-x$.

see also MINUS SIGN, NEGATIVE, PLUS, PLUS OR MINUS, TIMES

Minus or Plus

see PLUS OR MINUS

Minus Sign

The symbol "$-$" which is used to denote a NEGATIVE number or SUBTRACTION.

see also MINUS, PLUS SIGN, SIGN, SUBTRACTION

Minute

see ARC MINUTE

Miquel Circles

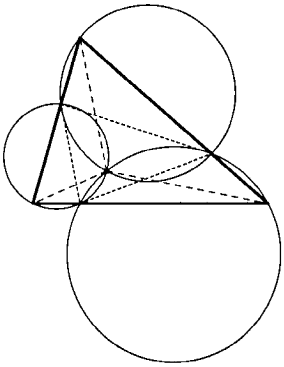

For a TRIANGLE ΔABC and three points A', B', and C', one on each of its sides, the three Miquel circles are the circles passing through each VERTEX and its neighboring side points (i.e., $AC'B'$, $BA'C'$, and $CB'A'$). According to MIQUEL'S THEOREM, the Miquel circles are CONCURRENT in a point M known as the MIQUEL POINT. Similarly, there are n Miquel circles for n lines taken $(n-1)$ at a time.

see also MIQUEL POINT, MIQUEL'S THEOREM, MIQUEL TRIANGLE

Miquel Equation

$$\measuredangle A_2 M A_3 = \measuredangle A_2 A_1 A_3 + \measuredangle P_2 P_1 P_3,$$

where $\measuredangle$ is a DIRECTED ANGLE.

see also DIRECTED ANGLE, MIQUEL'S THEOREM

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 131–144, 1929.

Miquel Point

The point of CONCURRENCE of the MIQUEL CIRCLES.

see also MIQUEL CIRCLES, MIQUEL'S THEOREM, MIQUEL TRIANGLE

Miquel's Theorem

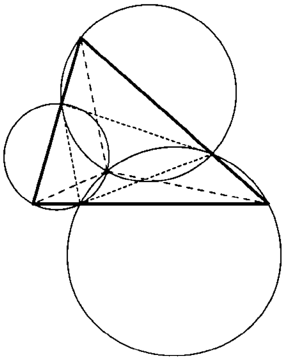

If a point is marked on each side of a TRIANGLE ΔABC, then the three MIQUEL CIRCLES (each through a VERTEX and the two marked points on the adjacent sides) are CONCURRENT at a point M called the MIQUEL POINT. This result is a slight generalization of the so-called PIVOT THEOREM.

If M lies in the interior of the triangle, then it satisfies

$$\angle P_2 M P_3 = 180^\circ - \alpha_1$$

$$\angle P_3 M P_1 = 180^\circ - \alpha_2$$

$$\angle P_1 M P_2 = 180^\circ - \alpha_3.$$

The lines from the MIQUEL POINT to the marked points make equal angles with the respective sides. (This is a by-product of the MIQUEL EQUATION.)

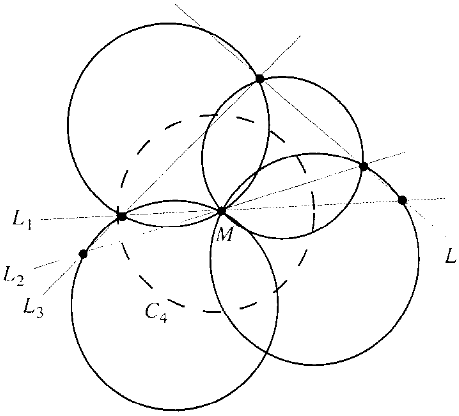

Given four lines L_1, ..., L_4 each intersecting the other three, the four MIQUEL CIRCLES passing through each subset of three intersection points of the lines meet in a point known as the 4-Miquel point M. Furthermore, the centers of these four MIQUEL CIRCLES lie on a CIRCLE C_4 (Johnson 1929, p. 139). The lines from M to given points on the sides make equal ANGLES with respect to the sides.

Similarly, given n lines taken by $(n-1)$s yield n MIQUEL CIRCLES like C_4 passing through a point P_n, and their centers lie on a CIRCLE C_{n+1}.

see also MIQUEL CIRCLES, MIQUEL EQUATION, MIQUEL TRIANGLE, NINE-POINT CIRCLE, PEDAL CIRCLE, PIVOT THEOREM

References
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 131–144, 1929.

Miquel Triangle

Given a point P and a triangle $\Delta A_1 A_2 A_3$, the Miquel triangle is the triangle connecting the side points P_1, P_2, and P_3 of $\Delta A_1 A_2 A_3$ with respect to which P is the MIQUEL POINT. All Miquel triangles of a given point M are directly similar, and M is the SIMILITUDE CENTER in every case.

Mira Fractal

A FRACTAL based on the map

$$F(x) = ax + \frac{2(1-a)x^2}{1+x^2}.$$

References
Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, p. 136, 1991.

Mirimanoff's Congruence

If the first case of FERMAT'S LAST THEOREM is false for the PRIME exponent p, then $3^{p-1} \equiv 1 \pmod{p^2}$.

see also FERMAT'S LAST THEOREM

Mirror Image

An image of an object obtained by reflecting it in a mirror so that the signs of one of its coordinates are reversed.

see AMPHICHIRAL, CHIRAL, ENANTIOMER, HANDEDNESS

Mirror Plane

The SYMMETRY OPERATION $(x, y, z) \to (x, y, -z)$, etc., which is equivalent to $\bar{2}$, where the bar denotes an IMPROPER ROTATION.

Misère Form

A version of NIM-like GAMES in which the player taking the last piece is the loser. For most IMPARTIAL GAMES, this form is much harder to analyze, but it requires only a trivial modification for the game of NIM.

Mitchell Index

The statistical INDEX

$$P_M \equiv \frac{\sum p_n q_a}{\sum p_0 q_a},$$

where p_n is the price per unit in period n and q_n is the quantity produced in period n.

see also INDEX

References
Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, pp. 66–67, 1962.

Miter Surface

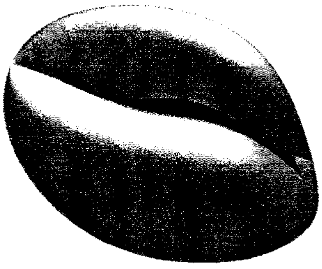

A Quartic Surface named after its resemblance to the liturgical headdress worn by bishops and given by the equation

$$4x^2(x^2+y^2+z^2)-y^2(1-y^2-z^2)=0.$$

see also Quartic Surface

References

Nordstrand, T. "Surfaces." `http://www.uib.no/people/nfytn/surfaces.htm`.

Mittag-Leffler Function

$$E_\gamma(x)\equiv\sum_{k=0}^{\infty}\frac{x^k}{\Gamma(\gamma k+1)}.$$

It is related to the Generalized Hyperbolic Functions by

$$F_{n,0}^1(x)=E_n(x^n).$$

References

Muldoon, M. E. and Ungar, A. A. "Beyond Sin and Cos." *Math. Mag.* **69**, 3–14, 1996.

Mittenpunkt

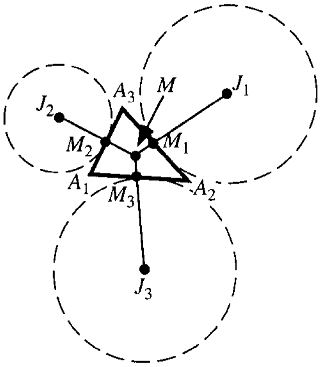

The Lemoine Point of the Excentral Triangle, i.e., the point of concurrence M of the lines from the Excenters J_i through the corresponding Triangle side Midpoint M_i. It is also called the Middlespoint and has Triangle Center Function

$$\alpha=b+c-a=\tfrac{1}{2}\cot A.$$

see also Excenter, Excentral Triangle, Nagel Point

References

Baptist, P. *Die Entwicklung der Neueren Dreiecksgeometrie.* Mannheim: Wissenschaftsverlag, p. 72, 1992.
Eddy, R. H. "A Generalization of Nagel's Middlespoint." *Elem. Math.* **45**, 14–18, 1990.
Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.
Kimberling, C. "Mittenpunkt." `http://www.evansville.edu/~ck6/tcenters/class/mitten.html`.

Mixed Partial Derivative

A Partial Derivative of second or greater order with respect to two or more different variables, for example

$$f_{xy}=\frac{\partial^2 f}{\partial x\partial y}.$$

If the mixed partial derivatives exist and are continuous at a point $\mathbf{x}_0$, then they are equal at $\mathbf{x}_0$ regardless of the order in which they are taken.

see also Partial Derivative

Mixed Strategy

A collection of moves together with a corresponding set of weights which are followed probabilistically in the playing of a Game. The Minimax Theorem of Game Theory states that every finite, zero-sum, two-person game has optimal mixed strategies.

see also Game Theory, Minimax Theorem, Strategy

Mixed Tensor

A Tensor having Contravariant and Covariant indices.

see also Contravariant Tensor, Covariant Tensor, Tensor

Mnemonic

A mental device used to aid memorization. Common mnemonics for mathematical constants such as e and Pi consist of sentences in which the number of letters in each word give successive digits.

see also e, Josephus Problem, Pi

References

Luria, A. R. *The Mind of a Mnemonist: A Little Book about a Vast Memory.* Cambridge, MA: Harvard University Press, 1987.

Möbius Band

see Möbius Strip

Möbius Function

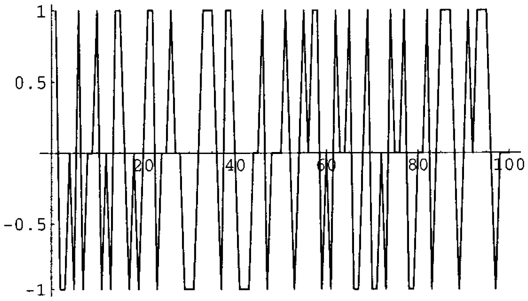

$$\mu(n) \equiv \begin{cases} 0 & \text{if } n \text{ has one or more repeated prime factors} \\ 1 & \text{if } n = 1 \\ (-1)^k & \text{if } n \text{ is a product of } k \text{ distinct primes,} \end{cases}$$

so $mu(n) \neq 0$ indicates that n is SQUAREFREE. The first few values are 1, −1, −1, 0, −1, 1, −1, 0, 0, 1, −1, 0, ... (Sloane's A008683).

The SUMMATORY FUNCTION of the Möbius function is called MERTENS FUNCTION.

see also BRAUN'S CONJECTURE, MERTENS FUNCTION, MÖBIUS INVERSION FORMULA, MÖBIUS PERIODIC FUNCTION, PRIME ZETA FUNCTION, RIEMANN FUNCTION, SQUAREFREE

References

Abramowitz, M. and Stegun, C. A. (Eds.). "The Möbius Function." §24.3.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 826, 1972.

Deléglise, M. and Rivat, J. "Computing the Summation of the Möbius Function." *Experiment. Math.* **5**, 291–295, 1996.

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford: Clarendon Press, p. 236, 1979.

Sloane, N. J. A. Sequence A008683 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Vardi, I. *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 7–8 and 223–225, 1991.

Möbius Group

The equation

$$x_1{}^2 + x_2{}^2 + \ldots + x_n{}^2 - 2x_0 x_\infty = 0$$

represents an n-D HYPERSPHERE $\mathbb{S}^n$ as a quadratic hypersurface in an $(n+1)$-D real projective space $\mathbb{P}^{n+1}$, where x_a are homogeneous coordinates in $\mathbb{P}^{n+1}$. Then the GROUP $M(n)$ of projective transformations which leave $\mathbb{S}^n$ invariant is called the Möbius group.

References

Iyanaga, S. and Kawada, Y. (Eds.). "Möbius Geometry." §78A in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 265–266, 1980.

Möbius Inversion Formula

If $g(n) \equiv \sum_{d|n} f(d)$, then

$$f(n) = \sum_{d|n} \mu(d) g\left(\frac{n}{d}\right),$$

where the sums are over all possible INTEGERS d that DIVIDE n and $\mu(d)$ is the MÖBIUS FUNCTION. The LOGARITHM of the CYCLOTOMIC POLYNOMIAL

$$\Phi_n(x) = \prod_{d|n} (1 - x^{n/d})^{\mu(d)}$$

is the Möbius inversion formula.

see also CYCLOTOMIC POLYNOMIAL, MÖBIUS FUNCTION

References

Hardy, G. H. and Wright, W. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Oxford University Press, pp. 91–93, 1979.

Schroeder, M. R. *Number Theory in Science and Communication, 3rd ed.* New York: Springer-Verlag, 1997.

Vardi, I. *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 7–8 and 223–225, 1991.

Möbius Periodic Function

A function periodic with period 2π such that

$$p(\theta + \pi) = -p(\theta)$$

for all θ is said to be Möbius periodic.

Möbius Problem

Let $A = \{a_1, a_2, \ldots\}$ be a free Abelian SEMIGROUP, where a_1 is the unit element. Then do the following properties,

1. $a < b$ IMPLIES $ac < bc$ for $a, b, c \in A$, where A has the linear order $a_1 < a_2 < \ldots$,
2. $\mu(a_n) = \mu(n)$ for all n,

imply that

$$a_{mn} = a_m a_n$$

for all $m, n \geq 1$? The problem is known to be true for $mn \leq 74$ for all $n \leq 240$.

see also BRAUN'S CONJECTURE, MÖBIUS FUNCTION

References

Flath, A. and Zulauf, A. "Does the Möbius Function Determine Multiplicative Arithmetic?" *Amer. Math. Monthly* **102**, 354–256, 1995.

Möbius Shorts

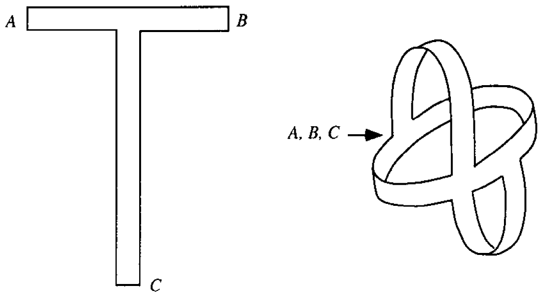

A one-sided surface reminiscent of the MÖBIUS STRIP.

see also MÖBIUS STRIP

References

Boas, R. P. Jr. "Möbius Shorts." *Math. Mag.* **68**, 127, 1995.

Möbius Strip

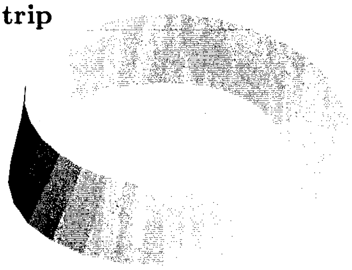

A one-sided surface obtained by cutting a band widthwise, giving it a half twist, and re-attaching the two ends. According to Madachy (1979), the B. F.Goodrich Company patented a conveyor belt in the form of a Möbius strip which lasts twice as long as conventional belts.

A Möbius strip can be represented parametrically by

$$\begin{aligned} x &= [R + s\cos(\tfrac{1}{2}\theta)]\cos\theta \\ y &= [R + s\cos(\tfrac{1}{2}\theta)]\sin\theta \\ z &= s\sin(\tfrac{1}{2}\theta), \end{aligned}$$

for $s \in [-1, 1]$ and $\theta \in [0, 2\pi)$. Cutting a Möbius strip, giving it extra twists, and reconnecting the ends produces unexpected figures called PARADROMIC RINGS (Listing and Tait 1847, Ball and Coxeter 1987) which are summarized in the table below.

half-twists	cuts	divs.	result
1	1	2	1 band, length 2
1	1	3	1 band, length 2 1 Möbius strip, length 1
1	2	4	2 bands, length 2
1	2	5	2 bands, length 2 1 Möbius strip, length 1
1	3	6	3 bands, length 2
1	3	7	3 bands, length 2 1 Möbius strip, length 1
2	1	2	2 bands, length 1
2	2	3	3 bands, length 1
2	3	4	4 bands, length 1

A TORUS can be cut into a Möbius strip with an EVEN number of half-twists, and a KLEIN BOTTLE can be cut in half along its length to make two Möbius strips. In addition, two strips on top of each other, each with a half-twist, give a single strip with four twists when disentangled.

There are three possible SURFACES which can be obtained by sewing a Möbius strip to the edge of a DISK: the BOY SURFACE, CROSS-CAP, and ROMAN SURFACE.

The Möbius strip has EULER CHARACTERISTIC 1, and the HEAWOOD CONJECTURE therefore shows that any set of regions on it can be colored using six-colors only.

see also BOY SURFACE, CROSS-CAP, MAP COLORING, PARADROMIC RINGS, PRISMATIC RING, ROMAN SURFACE

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 127–128, 1987.

Bogomolny, A. "Möbius Strip." `http://www.cut-the-knot.com/do_you_know/moebius.html`.

Gardner, M. "Möbius Bands." Ch. 9 in *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, pp. 123–136, 1978.

Geometry Center. "The Klein Bottle." `http://www.geom.umn.edu/zoo/features/mobius/`.

Gray, A. "The Möbius Strip." §12.3 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 236–238, 1993.

Hunter, J. A. H. and Madachy, J. S. *Mathematical Diversions.* New York: Dover, pp. 41–45, 1975.

Kraitchik, M. §8.4.3 in *Mathematical Recreations.* New York: W. W. Norton, pp. 212–213, 1942.

Listing and Tait. *Vorstudien zur Topologie, Göttinger Studien,* Pt. 10, 1847.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, p. 7, 1979.

Nordstrand, T. "Mobiusband." `http://www.uib.no/people/nfytn/moebtxt.htm`.

Pappas, T. "The Moebius Strip & the Klein Bottle," "A Twist to the Moebius Strip," "The 'Double' Moebius Strip." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 207, 1989.

Steinhaus, H. *Mathematical Snapshots, 3rd American ed.* New York: Oxford University Press, pp. 269–274, 1983.

Wagon, S. "Rotating Circles to Produce a Torus or Möbius Strip." §7.4 in *Mathematica in Action.* New York: W. H. Freeman, pp. 229–232, 1991.

Wang, P. "Renderings." `http://www.ugcs.caltech.edu/~peterw/portfolio/renderings/`.

Möbius Transformation

A transformation of the form

$$w = f(z) = \frac{az + b}{cz + d},$$

where a, b, c, $d \in \mathbb{C}$ and

$$ad - bc \neq 0,$$

is a CONFORMAL TRANSFORMATION and is called a Möbius transformation. It is linear in both w and z.

Every Möbius transformation except $f(z) = z$ has one or two FIXED POINTS. The Möbius transformation sends CIRCLES and lines to CIRCLES or lines. Möbius transformations preserve symmetry. The CROSS-RATIO is invariant under a Möbius transformation. A Möbius transformation is a composition of translations, rotations, magnifications, and inversions.

To determine a particular Möbius transformation, specify the map of three points which preserve orientation. A particular Möbius transformation is then uniquely determined. To determine a general Möbius transformation, pick two symmetric points α and α_S. Define $\beta \equiv f(\alpha)$, restricting β as required. Compute β_S. $f(\alpha_S)$ then equals β_S since the Möbius transformation preserves symmetry (the SYMMETRY PRINCIPLE). Plug in α and α_S into the general Möbius transformation and set equal to β and β_S. Without loss of generality, let $c = 1$ and solve for a and b in terms of β. Plug back into the general expression to obtain a Möbius transformation.

see also SYMMETRY PRINCIPLE

Möbius Triangles

SPHERICAL TRIANGLES into which a SPHERE is divided by the planes of symmetry of a UNIFORM POLYHEDRON.

see also SPHERICAL TRIANGLE, UNIFORM POLYHEDRON

Mock Theta Function

Ramanujan was the first to extensively study these THETA FUNCTION-like functions

$$f(q) = \sum_{n=0}^{\infty} \frac{q^{n^2}}{(1+q)^2(1+q^2)^2\cdots(1+q^n)^2}$$

$$\phi(q) = \sum_{n=0}^{\infty} \frac{q^{n^2}}{(1+q^2)(1+q^4)\cdots(1+q^{2n})}.$$

see also q-SERIES, THETA FUNCTION

References

Bellman, R. E. *A Brief Introduction to Theta Functions.* New York: Holt, Rinehart, and Winston, 1961.

Mod

see CONGRUENCE

Mode

The most common value obtained in a set of observations.

see also MEAN, MEDIAN (STATISTICS), ORDER STATISTIC

Mode Locking

A phenomenon in which a system being forced at an IRRATIONAL period undergoes rational, periodic motion which persists for a finite range of forcing values. It may occur for strong couplings between natural and forcing oscillation frequencies.

The phenomenon can be exemplified in the CIRCLE MAP when, after q iterations of the map, the new angle differs from the initial value by a RATIONAL NUMBER

$$\theta_{n+q} = \theta_n + \frac{p}{q}.$$

This is the form of the unperturbed CIRCLE MAP with the WINDING NUMBER

$$\Omega = \frac{p}{q}.$$

For Ω not a RATIONAL NUMBER, the trajectory is QUASIPERIODIC.

see also CHAOS, QUASIPERIODIC FUNCTION

Model Completion

Model completion is a term employed when EXISTENTIAL CLOSURE is successful. The formation of the COMPLEX NUMBERS, and the move from affine to projective geometry, are successes of this kind. The theory of existential closure gives a theoretical basis of Hilbert's "method of ideal elements."

References

Manders, K. L. "Interpretations and the Model Theory of the Classical Geometries." In *Models and Sets.* Berlin: Springer-Verlag, pp. 297–330, 1984.

Manders, K. L. "Domain Extension and the Philosophy of Mathematics." *J. Philos.* **86**, 553–562, 1989.

Model Theory

Model theory is a general theory of interpretations of an AXIOMATIC SET THEORY. It is the branch of LOGIC studying mathematical structures by considering first-order sentences which are true of those structures and the sets which are definable in those structures by first-order FORMULAS (Marker 1996).

Mathematical structures obeying axioms in a system are called "models" of the system. The usual axioms of ANALYSIS are second order and are known to have the REAL NUMBERS as their unique model. Weakening the axioms to include only the first-order ones leads to a new type of model in what is called NONSTANDARD ANALYSIS.

see also KHOVANSKI'S THEOREM, NONSTANDARD ANALYSIS, WILKIE'S THEOREM

References

Doets, K. *Basic Model Theory.* New York: Cambridge University Press, 1996.

Marker, D. "Model Theory and Exponentiation." *Not. Amer. Math. Soc.* **43**, 753–759, 1996.

Stewart, I. "Non-Standard Analysis." In *From Here to Infinity: A Guide to Today's Mathematics.* Oxford, England: Oxford University Press, pp. 80–81, 1996.

Modified Bessel Differential Equation

The second-order ordinary differential equation

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} - (x^2+n^2)y = 0.$$

The solutions are the MODIFIED BESSEL FUNCTIONS OF THE FIRST and SECOND KINDS. If $n = 0$, the modified Bessel differential equation becomes

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} - x^2y = 0,$$

which can also be written

$$\frac{d}{dx}\left(x\frac{dy}{dx}\right) = xy.$$

Modified Bessel Function of the First Kind

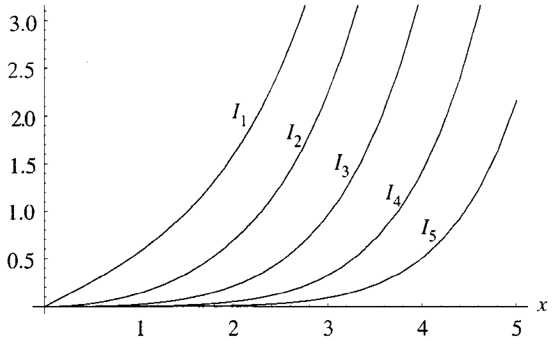

A function $I_n(x)$ which is one of the solutions to the MODIFIED BESSEL DIFFERENTIAL EQUATION and is closely related to the BESSEL FUNCTION OF THE FIRST KIND $J_n(x)$. The above plot shows $I_n(x)$ for $n = 1, 2, \ldots, 5$. In terms of $J_n(x)$,

$$I_n(x) \equiv i^{-n}J_n(ix) = e^{-n\pi i/2}J_n(xe^{i\pi/2}). \tag{1}$$

For a REAL NUMBER ν, the function can be computed using

$$I_\nu(z) = (\tfrac{1}{2}z)^\nu \sum_{k=0}^\infty \frac{(\frac{1}{4}z^2)^k}{k!\Gamma(\nu+k+1)}, \tag{2}$$

where $\Gamma(z)$ is the GAMMA FUNCTION. An integral formula is

$$I_\nu(z) = \frac{1}{\pi}\int_0^\pi e^{z\cos\theta}\cos(\nu\theta)\,d\theta - \frac{\sin(\nu\pi)}{\pi}\int_0^\infty e^{-z\cosh t-\nu t}\,dt, \tag{3}$$

which simplifies for ν an INTEGER n to

$$I_n(z) = \frac{1}{\pi}\int_0^\pi e^{z\cos\theta}\cos(n\theta)\,d\theta \tag{4}$$

(Abramowitz and Stegun 1972, p. 376).

A derivative identity for expressing higher order modified Bessel functions in terms of $I_0(x)$ is

$$I_n(x) = T_n\left(\frac{d}{dx}\right)I_0(x), \tag{5}$$

where $T_n(x)$ is a CHEBYSHEV POLYNOMIAL OF THE FIRST KIND.

see also BESSEL FUNCTION OF THE FIRST KIND, MODIFIED BESSEL FUNCTION OF THE FIRST KIND, WEBER'S FORMULA

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Modified Bessel Functions I and K." §9.6 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 374–377, 1972.

Arfken, G. "Modified Bessel Functions, $I_\nu(x)$ and $K_\nu(x)$." §11.5 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 610–616, 1985.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/cntfrc/cntfrc.html`.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Bessel Functions of Fractional Order, Airy Functions, Spherical Bessel Functions." §6.7 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 234–245, 1992.

Spanier, J. and Oldham, K. B. "The Hyperbolic Bessel Functions $I_0(x)$ and $I_1(x)$" and "The General Hyperbolic Bessel Function $I_\nu(x)$." Chs. 49–50 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 479–487 and 489–497, 1987.

Modified Bessel Function of the Second Kind

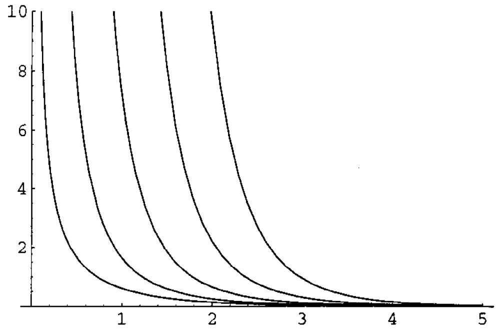

The function $K_n(x)$ which is one of the solutions to the MODIFIED BESSEL DIFFERENTIAL EQUATION. The above plot shows $K_n(x)$ for $n = 1, 2, \ldots, 5$. $K_n(x)$ is closely related to the MODIFIED BESSEL FUNCTION OF THE FIRST KIND $I_n(x)$ and HANKEL FUNCTION $H_n(x)$,

$$\begin{aligned} K_n(x) &\equiv \tfrac{1}{2}\pi i^{n+1}H_n^{(1)}(ix) && (1)\\ &= \tfrac{1}{2}\pi i^{n+1}[J_n(ix)+iN_n(ix)] && (2)\\ &= \frac{\pi}{2}\frac{I_{-n}(x)-I_n(x)}{\sin(n\pi)} && (3) \end{aligned}$$

(Watson 1966, p. 185). A sum formula for K_n is

$$\begin{aligned} K_n(z) = {} & \tfrac{1}{2}(\tfrac{1}{2}z)^{-n}\sum_{k=0}^{n-1}\frac{(n-k-1)!}{k!}(-\tfrac{1}{4}z^2)^k \\ & +(-1)^{n+1}\ln(\tfrac{1}{2}z)I_n(z) \\ & +(-1)^n\tfrac{1}{2}(\tfrac{1}{2}z)^n\sum_{k=0}^\infty[\psi(k+1)+\psi(n+k+1)]\frac{(\frac{1}{4}z^2)^k}{k!(n+k)!}, \end{aligned} \tag{4}$$

where ψ is the DIGAMMA FUNCTION (Abramowitz and Stegun 1972). An integral formula is

$$K_\nu(z) = \frac{\Gamma(\nu+\frac{1}{2})(2z)^\nu}{\sqrt{\pi}}\int_0^\infty \frac{\cos t\,dt}{(t^2+z^2)^{\nu+1/2}} \tag{5}$$

which, for $\nu = 0$, simplifies to

$$K_0(x) = \int_0^\infty \cos(x \sinh t)\,dt = \int_0^\infty \frac{\cos(xt)\,dt}{\sqrt{t^2+1}}. \qquad (6)$$

Other identities are

$$K_n(z) = \frac{\sqrt{\pi}}{(n-\frac{1}{2})!}\left(\tfrac{1}{2}z\right)^n \int_1^\infty e^{-zx}(x^2-1)^{n-1/2}\,dx \quad (7)$$

for $n > -1/2$ and

$$K_n(z) = \sqrt{\frac{\pi}{2z}}\frac{e^{-z}}{(n-\frac{1}{2})!}\int_0^\infty e^{-t}t^{n-1/2}\left(1-\frac{t}{2z}\right)^{n-1/2}\,dt \qquad (8)$$

$$= \sqrt{\frac{\pi}{2z}}\frac{e^{-z}}{(n-\frac{1}{2})!}\sum_{r=0}^\infty \frac{(n-\frac{1}{2})!}{r!(n-r-\frac{1}{2})!}(2z)^{-r} \times \int_0^\infty e^{-t}t^{n+r-1/2}\,dt. \qquad (9)$$

The modified Bessel function of the second kind is sometimes called the BASSET FUNCTION.

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Modified Bessel Functions I and K." §9.6 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 374–377, 1972.

Arfken, G. "Modified Bessel Functions, $I_\nu(x)$ and $K_\nu(x)$." §11.5 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 610–616, 1985.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Modified Bessel Functions of Integral Order" and "Bessel Functions of Fractional Order, Airy Functions, Spherical Bessel Functions." §6.6 and 6.7 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 229–245, 1992.

Spanier, J. and Oldham, K. B. "The Basset $K_\nu(x)$." Ch. 51 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 499–507, 1987.

Watson, G. N. *A Treatise on the Theory of Bessel Functions, 2nd ed.* Cambridge, England: Cambridge University Press, 1966.

Modified Spherical Bessel Differential Equation

The SPHERICAL BESSEL DIFFERENTIAL EQUATION with a NEGATIVE separation constant, given by

$$r^2\frac{d^2R}{dr^2} + 2r\frac{dR}{dr} - [k^2r^2 + n(n+1)]R = 0.$$

The solutions are called MODIFIED SPHERICAL BESSEL FUNCTIONS.

Modified Spherical Bessel Function

Solutions to the MODIFIED SPHERICAL BESSEL DIFFERENTIAL EQUATION, given by

$$i_n(x) \equiv \sqrt{\frac{\pi}{2x}}\,I_{n+1/2}(x) \qquad (1)$$

$$i_0(x) = \frac{\sinh(x)}{x} \qquad (2)$$

$$k_n(x) \equiv \sqrt{\frac{2\pi}{x}}\,K_{n+1/2}(x) \qquad (3)$$

$$k_0(x) = \frac{e^{-x}}{x}, \qquad (4)$$

where $I_n(x)$ is a MODIFIED BESSEL FUNCTION OF THE FIRST KIND and $K_n(x)$ is a MODIFIED BESSEL FUNCTION OF THE SECOND KIND.

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Modified Spherical Bessel Functions." §10.2 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 443–445, 1972.

Modified Struve Function

$$\mathcal{L}_\nu(z) = \left(\tfrac{1}{2}z\right)^{\nu+1}\sum_{k=0}^\infty \frac{\left(\frac{z}{2}\right)^{2k}}{\Gamma\left(k+\frac{3}{2}\right)\Gamma\left(k+\nu+\frac{3}{2}\right)}$$

$$= \frac{2\left(\frac{z}{2}\right)^\nu}{\sqrt{\pi}\Gamma(\nu+\frac{1}{2})}\int_0^{\pi/2}\sinh(z\cos\theta)\sin^{2\nu}\theta\,d\theta,$$

where $\Gamma(z)$ is the GAMMA FUNCTION.

see also ANGER FUNCTION, STRUVE FUNCTION, WEBER FUNCTIONS

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Modified Struve Function $\mathbf{L}_\nu(x)$." §12.2 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 498, 1972.

Modular Angle

Given a MODULUS k in an ELLIPTIC INTEGRAL, the modular angle is defined by $k \equiv \sin\alpha$. An ELLIPTIC INTEGRAL is written $I(\phi|m)$ when the PARAMETER is used, $I(\phi, k)$ when the MODULUS is used, and $I(\phi\backslash\alpha)$ when the modular angle is used.

see also AMPLITUDE, CHARACTERISTIC (ELLIPTIC INTEGRAL), ELLIPTIC INTEGRAL, MODULUS (ELLIPTIC INTEGRAL), NOME, PARAMETER

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 590, 1972.

Modular Equation

The modular equation of degree n gives an algebraic connection of the form

$$\frac{K'(l)}{K(l)} = n\frac{K'(k)}{K(k)} \tag{1}$$

between the TRANSCENDENTAL COMPLETE ELLIPTIC INTEGRALS OF THE FIRST KIND with moduli k and l. When k and l satisfy a modular equation, a relationship of the form

$$\frac{M(l,k)\,dy}{\sqrt{(1-y^2)(1-l^2y^2)}} = \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}} \tag{2}$$

exists, and M is called the MODULAR FUNCTION MULTIPLIER. In general, if p is an ODD PRIME, then the modular equation is given by

$$\Omega_p(u,v) = (v-u_0)(v-u_1)\cdots(v-u_p), \tag{3}$$

where

$$u_p \equiv (-1)^{(p^2-1)/8}[\lambda(q^p)]^{1/8} \equiv (-1)^{(p^2-1)/8}u(q^p), \tag{4}$$

λ is a ELLIPTIC LAMBDA FUNCTION, and

$$q \equiv e^{\imath\pi t} \tag{5}$$

(Borwein and Borwein 1987, p. 126). An ELLIPTIC INTEGRAL identity gives

$$\frac{K'(k)}{K(k)} = 2\frac{K'\left(\frac{2\sqrt{k}}{1+k}\right)}{K\left(\frac{2\sqrt{k}}{1+k}\right)}, \tag{6}$$

so the modular equation of degree 2 is

$$l = \frac{2\sqrt{k}}{1+k} \tag{7}$$

which can be written as

$$l^2(1+k)^2 = 4k. \tag{8}$$

A few low order modular equations written in terms of k and l are

$$\Omega_2 = l^2(1+k)^2 - 4k = 0 \tag{9}$$

$$\Omega_7 = (kl)^{1/4} + (k'l')^{1/4} - 1 = 0 \tag{10}$$

$$\Omega_{23} = (kl)^{1/4} + (k'l')^{1/4} + 2^{2/3}(klk'l')^{1/12} - 1 = 0. \tag{11}$$

In terms of u and v,

$$\Omega_3(u,v) = u^4 - v^4 + 2uv(1-u^2v^2) = 0 \tag{12}$$

$$\begin{aligned}\Omega_5(u,v) &= v^6 - u^6 + 5u^2v^2(v^2-u^2) + 4uv(u^4v^4-1)\\ &= \left(\frac{u}{v}\right)^3 + \left(\frac{v}{u}\right)^3 = 2\left(u^2v^2 - \frac{1}{u^2v^2}\right) = 0\end{aligned} \tag{13}$$

$$\Omega_7(u,v) = (1-u^8)(1-v^8) - (1-uv)^8 = 0, \tag{14}$$

where

$$u^2 \equiv \sqrt{k} = \frac{\vartheta_2(q)}{\vartheta_3(q)} \tag{15}$$

and

$$v^2 \equiv \sqrt{l} = \frac{\vartheta_2(q^p)}{\vartheta_3(q^p)}. \tag{16}$$

Here, ϑ_i are THETA FUNCTIONS.

A modular equation of degree 2^r for $r \geq 2$ can be obtained by iterating the equation for 2^{r-1}. Modular equations for PRIME p from 3 to 23 are given in Borwein and Borwein (1987).

Quadratic modular identities include

$$\frac{\vartheta_3(q)}{\vartheta_3(q^4)} - 1 = \left[\frac{\vartheta_3{}^2(q^2)}{\vartheta_3{}^2(q^4)} - 1\right]^{1/2}. \tag{17}$$

Cubic identities include

$$\left[3\frac{\vartheta_2(q^9)}{\vartheta_2(q)} - 1\right]^3 = 9\frac{\vartheta_2{}^4(q^3)}{\vartheta_2{}^4(q)} - 1 \tag{18}$$

$$\left[3\frac{\vartheta_3(q^9)}{\vartheta_3(q)} - 1\right]^3 = 9\frac{\vartheta_3{}^4(q^3)}{\vartheta_3{}^4(q)} - 1 \tag{19}$$

$$\left[3\frac{\vartheta_4(q^9)}{\vartheta_4(q)} - 1\right]^3 = 9\frac{\vartheta_4{}^4(q^3)}{\vartheta_4{}^4(q)} - 1. \tag{20}$$

A seventh-order identity is

$$\sqrt{\vartheta_3(q)\vartheta_3(q^7)} - \sqrt{\vartheta_4(q)\vartheta_4(q^7)} = \sqrt{\vartheta_2(q)\vartheta_2(q^7)}. \tag{21}$$

From Ramanujan (1913–1914),

$$(1+q)(1+q^3)(1+q^5)\cdots = 2^{1/6}q^{1/24}(kk')^{-1/12} \tag{22}$$

$$(1-q)(1-q^3)(1-q^5)\cdots = 2^{1/6}q^{1/24}k^{-1/12}k'^{1/6}. \tag{23}$$

see also SCHLÄFLI'S MODULAR FORM

References

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, pp. 127–132, 1987.

Hanna, M. "The Modular Equations." *Proc. London Math. Soc.* **28**, 46–52, 1928.

Ramanujan, S. "Modular Equations and Approximations to π." *Quart. J. Pure. Appl. Math.* **45**, 350–372, 1913–1914.

Modular Form

A modular form is a function in the COMPLEX PLANE with rather spectacular and special properties resulting from a surprising array of internal symmetries. If

$$F\left(\frac{az+b}{cz+d}\right) = (cz+d)^2F(z),$$

then $F(z)$ is said to be a modular form of weight 2 and level N. If it is correctly parameterized, a modular form is ANALYTIC and vanishes at the cusps, so it is called a CUSP FORM. It is also an eigenform under a certain HECKE ALGEBRA.

A remarkable connection between rational ELLIPTIC CURVES and modular forms is given by the TANIYAMA-SHIMURA CONJECTURE, which states that any rational ELLIPTIC CURVE is a modular form in disguise. This result was the one proved by Andrew Wiles in his celebrated proof of FERMAT'S LAST THEOREM.

see also CUSP FORM, ELLIPTIC CURVE, ELLIPTIC FUNCTION, FERMAT'S LAST THEOREM, HECKE ALGEBRA, MODULAR FUNCTION, MODULAR FUNCTION MULTIPLIER, SCHLÄFLI'S MODULAR FORM, TANIYAMA-SHIMURA CONJECTURE

References

Knopp, M. I. *Modular Functions, 2nd ed.* New York: Chelsea, 1993.

Koblitz, N. *Introduction to Elliptic Curves and Modular Forms.* New York: Springer-Verlag, 1993.

Rankin, R. A. *Modular Forms and Functions.* Cambridge, England: Cambridge University Press, 1977.

Sarnack, P. *Some Applications of Modular Forms.* Cambridge, England: Cambridge University Press, 1993.

Modular Function

f is a modular function of level N on the upper half H of the COMPLEX PLANE if it is MEROMORPHIC (even at the CUSPS), $ad - bc = 1$ for all a, b, c, d, and $N|c$.

see also ELLIPTIC FUNCTION, ELLIPTIC MODULAR FUNCTION, MODULAR FORM

References

Apostol, T. M. *Modular Functions and Dirichlet Series in Number Theory.* New York: Springer-Verlag, 1976.

Askey, R. In *Ramanujan International Symposium* (Ed. N. K Thakare). pp. 1–83.

Borwein, J. M. and Borwein, P. B. *Pi and the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, 1987.

Rankin, R. A. *Modular Forms and Functions.* Cambridge, England: Cambridge University Press, 1977.

Schoeneberg, B. *Elliptic Modular Functions: An Introduction.* Berlin: New York: Springer-Verlag, 1974.

Modular Function Multiplier

When k and l satisfy a MODULAR EQUATION, a relationship of the form

$$\frac{M(l,k)\,dy}{\sqrt{(1-y^2)(1-l^2y^2)}} = \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}} \tag{1}$$

exists, and M is called the multiplier. The multiplier of degree n can be given by

$$M_n(l,k) \equiv \frac{{\vartheta_3}^2(q)}{{\vartheta_3}^2(q^{1/p})} = \frac{K(k)}{K(l)}, \tag{2}$$

where ϑ_i is a THETA FUNCTION and $K(k)$ is a complete ELLIPTIC INTEGRAL OF THE FIRST KIND.

The first few multipliers in terms of l and k are

$$M_2(l,k) = \frac{1}{1+k} = \frac{1+l'}{2} \tag{3}$$

$$M_3(l,k) = \frac{1-\sqrt{\frac{l^3}{k}}}{1-\sqrt{\frac{k^3}{l}}}. \tag{4}$$

In terms of the u and v defined for MODULAR EQUATIONS,

$$M_3 = \frac{v}{v+2u^3} = \frac{2v^3-u}{3u} \tag{5}$$

$$M_5 = \frac{v(1-uv^3)}{v-u^5} = \frac{u+v^5}{5u(1+u^3v)} \tag{6}$$

$$\begin{aligned} M_7 &= \frac{v(1-uv)[1-uv+(uv)^2]}{v-u^7} \\ &= \frac{v^7-u}{7u(1-uv)[1-uv+(uv)^2]}. \end{aligned} \tag{7}$$

Modular Gamma Function

The GAMMA GROUP Γ is the set of all transformations w of the form

$$w(t) = \frac{at+b}{ct+d},$$

where a, b, c, and d are INTEGERS and $ad - bc = 1$. Γ-modular functions are then defined as in Borwein and Borwein (1987, p. 114).

see also KLEIN'S ABSOLUTE INVARIANT, LAMBDA GROUP, THETA FUNCTION

References

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, pp. 127–132, 1987.

Modular Group

The GROUP of all MÖBIUS TRANSFORMATIONS having INTEGER coefficients and DETERMINANT equal to 1.

Modular Lambda Function

see ELLIPTIC LAMBDA FUNCTION

Modular Lattice

A LATTICE which satisfies the identity

$$(x \wedge y) \vee (x \wedge z) = x \wedge (y \vee (x \wedge z))$$

is said to be modular.

see also DISTRIBUTIVE LATTICE

References

Grätzer, G. *Lattice Theory: First Concepts and Distributive Lattices.* San Francisco, CA: W. H. Freeman, pp. 35–36, 1971.

Modular System

A set M of all POLYNOMIALS in s variables, $x_1, \ldots, x_s$ such that if P, P_1, and P_2 are members, then so are $P_1 + P_2$ and QP, where Q is any POLYNOMIAL in $x_1, \ldots, x_s$.

see also HILBERT'S THEOREM, MODULAR SYSTEM BASIS

Modular System Basis

A basis of a MODULAR SYSTEM M is any set of POLYNOMIALS $B_1, B_2, \ldots$ of M such that every POLYNOMIAL of M is expressible in the form

$$R_1B_1 + R_2B_2 + \ldots,$$

where $R_1, R_2, \ldots$ are POLYNOMIALS.

Modular Transformation

see MODULAR EQUATION

Modulation Theorem

The important property of FOURIER TRANSFORMS that $\mathcal{F}[cos(2\pi k_0 x)f(x)]$ can be expressed in terms of $\mathcal{F}[f(x)] = F(k)$ as follows,

$$\mathcal{F}[cos(2\pi k_0 x)f(x)] = \tfrac{1}{2}[F(k-k_0) + F(k+k_0)].$$

see also FOURIER TRANSFORM

References

Bracewell, R. "Modulation Theorem." *The Fourier Transform and Its Applications.* New York: McGraw-Hill, p. 108, 1965.

Module

A mathematical object in which things can be added together COMMUTATIVELY by multiplying COEFFICIENTS and in which most of the rules of manipulating VECTORS hold. A module is abstractly very similar to a VECTOR SPACE, although modules have COEFFICIENTS in much more general algebraic objects and use RINGS as the COEFFICIENTS instead of FIELDS.

The additive submodule of the INTEGERS is a set of quantities closed under ADDITION and SUBTRACTION (although it is SUFFICIENT to require closure under SUBTRACTION). Numbers of the form $n\alpha \pm m\alpha$ for $n, m \in \mathbb{Z}$ form a module since,

$$n\alpha \pm m\alpha = (n \pm m)\alpha.$$

Given two INTEGERS a and b, the smallest module containing a and b is $\mathrm{GCD}(a, b)$.

References

Foote, D. and Dummit, D. *Abstract Algebra.* Englewood Cliffs, NJ: Prentice-Hall, 1990.

Modulo

see CONGRUENCE

Modulo Multiplication Group

A FINITE GROUP M_m of RESIDUE CLASSES prime to m under multiplication mod m. M_m is ABELIAN of ORDER $\phi(m)$, where $\phi(m)$ is the TOTIENT FUNCTION. The following table gives the modulo multiplication groups of small orders.

M_m	Group	$\phi(m)$	Elements
M_2	$\langle e\rangle$	1	1
M_3	Z_2	2	1, 2
M_4	Z_2	2	1, 3
M_5	Z_4	4	1, 2, 3, 4
M_6	Z_2	2	1, 5
M_7	Z_6	6	1, 2, 3, 4, 5, 6
M_8	$Z_2 \otimes Z_2$	4	1, 3, 5, 7
M_9	Z_6	6	1, 2, 4, 5, 7, 8
M_{10}	Z_4	4	1, 3, 7, 9
M_{11}	Z_{10}	10	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
M_{12}	$Z_2 \otimes Z_2$	4	1, 5, 7, 11
M_{13}	Z_{12}	12	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11
M_{14}	Z_6	6	1, 3, 5, 9, 11, 13
M_{15}	$Z_2 \otimes Z_4$	8	1, 2, 4, 7, 8, 11, 13, 14
M_{16}	$Z_2 \otimes Z_4$	8	1, 3, 5, 7, 9, 11, 13, 15
M_{17}	Z_{16}	16	1, 2, 3, ..., 16
M_{18}	Z_6	6	1, 5, 7, 11, 13, 17
M_{19}	Z_{18}	18	1, 2, 3, ..., 18
M_{20}	$Z_2 \otimes Z_4$	8	1, 3, 7, 9, 11, 13, 17, 19
M_{21}	$Z_2 \otimes Z_6$	12	1, 2, 4, 5, 7, 8, 10, 11, 13, 16, 17, 19
M_{22}	Z_{10}	10	1, 3, 5, 7, 9, 13, 15, 17, 19, 21
M_{23}	Z_{22}	22	1, 2, 3, ..., 22
M_{24}	$Z_2 \otimes Z_2 \otimes Z_2$	8	1, 5, 7, 11, 13, 17, 19, 23

M_m is a CYCLIC GROUP (which occurs exactly when m has a PRIMITIVE ROOT) IFF m is of one of the forms $m = 2$, 4, p^n, or $2p^n$, where p is an ODD PRIME and $n \geq 1$ (Shanks 1993, p. 92).

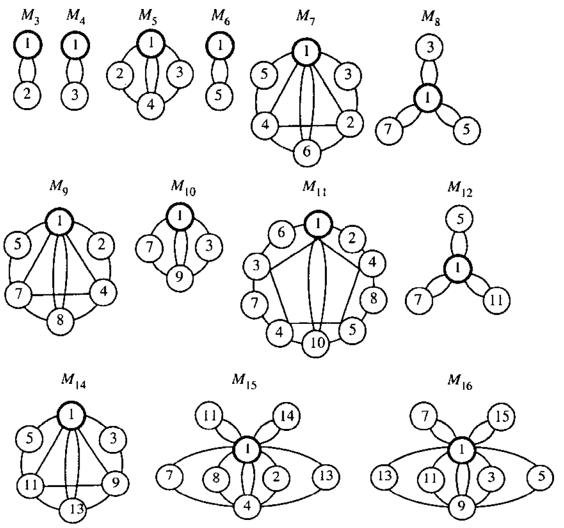

ISOMORPHIC modulo multiplication groups can be determined using a particular type of factorization of $\phi(m)$ as described by Shanks (1993, pp. 92–93). To perform this factorization (denoted ϕ_m), factor m in the standard form

$$m = {p_1}^{a_1} {p_2}^{a_2} \cdots {p_n}^{a_n}. \tag{1}$$

Now write the factorization of the TOTIENT FUNCTION involving each power of an ODD PRIME

$$\phi({p_i}^{a_i}) = (p_i - 1){p_i}^{a_i - 1} \tag{2}$$

as

$$\phi({p_i}^{a_i}) = \left\langle {q_1}^{b_1} \right\rangle \left\langle {q_2}^{b_2} \right\rangle \cdots \left\langle {q_s}^{b_s} \right\rangle \left\langle {p_i}^{a_i - 1} \right\rangle, \tag{3}$$

where

$$p_i - 1 = {q_1}^{b_1} {q_2}^{b_2} \cdots {q_s}^{b_s}, \tag{4}$$

$\left\langle q^b \right\rangle$ denotes the explicit expansion of q^b (i.e., $5^2 = 25$), and the last term is omitted if $a_i = 1$. If $p_1 = 2$, write

$$\phi(2^{a_1}) = \begin{cases} 2 & \text{for } a_1 = 2 \\ 2\left\langle 2^{a_1 - 2} \right\rangle & \text{for } a_1 > 2. \end{cases} \tag{5}$$

Now combine terms from the odd and even primes. For example, consider $m = 104 = 2^3 \cdot 13$. The only odd prime factor is 13, so factoring gives $13 - 1 = 12 = \left\langle 2^2 \right\rangle \langle 3 \rangle = 3 \cdot 4$. The rule for the powers of 2 gives $2^3 = 2\left\langle 2^{3-2} \right\rangle = 2\langle 2 \rangle = 2 \cdot 2$. Combining these two gives $\phi_{104} = 2 \cdot 2 \cdot 3 \cdot 4$. Other explicit values of ϕ_m are given below.

$$\begin{aligned} \phi_3 &= 2 \\ \phi_4 &= 2 \\ \phi_5 &= 4 \\ \phi_6 &= 2 \\ \phi_{15} &= 2 \cdot 4 \\ \phi_{16} &= 2 \cdot 4 \\ \phi_{17} &= 16 \\ \phi_{104} &= 2 \cdot 2 \cdot 3 \cdot 4 \\ \phi_{105} &= 2 \cdot 2 \cdot 3 \cdot 4. \end{aligned}$$

M_m and M_n are isomorphic IFF ϕ_m and ϕ_n are identical. More specifically, the abstract GROUP corresponding to a given M_m can be determined explicitly in terms of a DIRECT PRODUCT of CYCLIC GROUPS of the so-called CHARACTERISTIC FACTORS, whose product is denoted Φ_n. This representation is obtained from ϕ_m as the set of products of largest powers of each factor of ϕ_m. For example, for ϕ_{104}, the largest power of 2 is $4 = 2^2$ and the largest power of 3 is $3 = 3^1$, so the first characteristic factor is $4 \times 3 = 12$, leaving $2 \cdot 2$ (i.e., only powers of two). The largest power remaining is $2 = 2^1$, so the second CHARACTERISTIC FACTOR is 2, leaving 2, which is the third and last CHARACTERISTIC FACTOR. Therefore, $\Phi_{104} = 2 \cdot 2 \cdot 4$, and the group M_m is isomorphic to $Z_2 \otimes Z_2 \otimes Z_4$.

The following table summarizes the isomorphic modulo multiplication groups M_n for the first few n and identifies the corresponding abstract GROUP. No M_m is ISOMORPHIC to Z_8, Q_8, or D_4. However, every finite ABELIAN GROUP is isomorphic to a SUBGROUP of M_m for infinitely many different values of m (Shanks 1993, p. 96). CYCLE GRAPHS corresponding to M_n for small n are illustrated above, and more complicated CYCLE GRAPHS are illustrated by Shanks (1993, pp. 87–92).

Group	Isomorphic M_m
$\langle e \rangle$	M_2
Z_2	M_3, M_4, M_6
Z_4	M_5, M_{10}
$Z_2 \otimes Z_2$	M_8, M_{12}
Z_6	M_7, M_9, M_{14}, M_{18}
$Z_2 \otimes Z_4$	M_{15}, M_{16}, M_{20}, M_{30}
$Z_2 \otimes Z_2 \otimes Z_2$	M_{24}
Z_{10}	M_{11}, M_{22}
Z_{12}	M_{13}, M_{26}
$Z_2 \otimes Z_6$	M_{21}, M_{28}, M_{36}, M_{42}
Z_{16}	M_{17}, M_{34}
$Z_2 \otimes Z_8$	M_{32}
$Z_2 \otimes Z_2 \otimes Z_4$	M_{40}, M_{48}, M_{60}
Z_{18}	M_{19}, M_{27}, M_{38}, M_{54}
Z_{20}	M_{25}, M_{50}
$Z_2 \otimes Z_{10}$	M_{33}, M_{44}, M_{66}
Z_{22}	M_{23}, M_{46}
$Z_2 \otimes Z_{12}$	M_{35}, M_{39}, M_{45}, M_{52}, M_{70}, M_{78}, M_{90}
Z_{28}	M_{29}, M_{58}
Z_{30}	M_{31}, M_{62}
Z_{36}	M_{37}, M_{74}

The number of CHARACTERISTIC FACTORS r of M_m for $m = 1, 2, \ldots$ are 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 2, ... (Sloane's A046072). The number of QUADRATIC RESIDUES in M_m for $m > 2$ are given by $\phi(m)/2^r$ (Shanks 1993, p. 95). The first few for $m = 1, 2, \ldots$ are 0, 1, 1, 1, 2, 1, 3, 1, 3, 2, 5, 1, 6, ... (Sloane's A046073).

In the table below, $\phi(n)$ is the TOTIENT FUNCTION (Sloane's A000010) factored into CHARACTERISTIC FACTORS, $\lambda(n)$ is the CARMICHAEL FUNCTION (Sloane's A011773), and g_i are the smallest generators of the group M_n (of which there is a number equal to the number of CHARACTERISTIC FACTORS).

n	$\phi(n)$	$\lambda(n)$	g_i
3	2	2	2
4	2	2	3
5	4	2	2
6	2	2	5
7	6	6	3
8	$2 \cdot 2$	2	7, 3
9	6	6	2
10	4	4	3
11	10	10	2
12	$2 \cdot 2$	2	5, 7
13	12	12	2
14	6	6	3
15	$2 \cdot 4$	4	14, 2
16	$2 \cdot 4$	4	15, 3
17	16	16	3
18	6	6	5
19	18	18	2
20	$2 \cdot 4$	4	19, 3
21	$2 \cdot 6$	6	20, 2
22	10	10	7
23	22	22	5
24	$2 \cdot 2 \cdot 2$	2	5, 7, 13
25	20	20	2
26	12	12	7
27	18	18	2
28	$2 \cdot 6$	6	13, 3
29	28	28	2
30	$2 \cdot 4$	4	11, 7
31	30	30	3
32	$2 \cdot 8$	8	31, 3
33	$2 \cdot 10$	10	10, 2
34	16	16	3
35	$2 \cdot 12$	12	6, 2
36	$2 \cdot 6$	6	19,5
37	36	36	2
38	18	18	3
39	$2 \cdot 12$	12	38, 2
40	$2 \cdot 2 \cdot 4$	4	39, 11, 3
41	40	40	6
42	$2 \cdot 6$	6	13, 5
43	42	42	3
44	$2 \cdot 10$	10	43, 3
45	$2 \cdot 12$	12	44, 2
46	22	22	5
47	46	46	5
48	$2 \cdot 2 \cdot 4$	4	47, 7, 5
49	42	42	3
50	20	20	3

see also CHARACTERISTIC FACTOR, CYCLE GRAPH, FINITE GROUP, RESIDUE CLASS

References

Riesel, H. "The Structure of the Group M_n." *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, pp. 270–272, 1994.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 61–62 and 92, 1993.

Sloane, N. J. A. Sequences A011773, A046072, A046073, and A000010/M0299 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Weisstein, E. W. "Groups." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Groups.m`.

Modulus (Complex Number)

The modulus of a COMPLEX NUMBER z is denoted $|z|$.

$$|x+iy| \equiv \sqrt{x^2+y^2} \tag{1}$$

$$|re^{i\phi}| = |r|. \tag{2}$$

Let $c_1 \equiv Ae^{i\phi_1}$ and $c_2 \equiv Be^{i\phi_2}$ be two COMPLEX NUMBERS. Then

$$\left|\frac{c_1}{c_2}\right| = \left|\frac{Ae^{i\phi_1}}{Be^{i\phi_2}}\right| = \frac{A}{B}|e^{i(\phi_1-\phi_2)}| = \frac{A}{B} \tag{3}$$

$$\frac{|c_1|}{|c_2|} = \frac{|Ae^{i\phi_1}|}{|Be^{i\phi_2}|} = \frac{A}{B}\frac{|e^{i\phi_1}|}{|e^{i\phi_2}|} = \frac{A}{B}, \tag{4}$$

so

$$\left|\frac{c_1}{c_2}\right| = \frac{|c_1|}{|c_2|}. \tag{5}$$

Also,

$$|c_1 c_2| = |(Ae^{i\phi_1})(Be^{i\phi_2})| = AB|e^{i(\phi_1+\phi_2)}| = AB \tag{6}$$

$$|c_1|\,|c_2| = |Ae^{i\phi_1}|\,|Be^{i\phi_2}| = AB|e^{i\phi_1}|\,|e^{i\phi_2}| = AB, \tag{7}$$

so

$$|c_1 c_2| = |c_1|\,|c_2| \tag{8}$$

and, by extension,

$$|z^n| = |z|^n. \tag{9}$$

The only functions satisfying identities of the form

$$|f(x+iy)| = |f(x)+f(iy)| \tag{10}$$

are $f(z) = Az$, $f(z) = A\sin(bz)$, and $f(z) = A\sinh(bz)$ (Robinson 1957).

see also ABSOLUTE SQUARE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 16, 1972.

Robinson, R. M. "A Curious Mathematical Identity." *Amer. Math. Monthly* **64**, 83–85, 1957.

Modulus (Congruence)

see CONGRUENCE

Modulus (Elliptic Integral)

A parameter k used in ELLIPTIC INTEGRALS and ELLIPTIC FUNCTIONS defined to be $k \equiv \sqrt{m}$, where m is the PARAMETER. An ELLIPTIC INTEGRAL is written $I(\phi, k)$ when the modulus is used. It can be computed explicitly in terms of THETA FUNCTIONS of zero argument:

$$k = \frac{{\vartheta_2}^2(0|\tau)}{{\vartheta_3}^2(0|\tau)}. \tag{1}$$

The REAL period $K(k)$ and IMAGINARY period $K'(k) = K(k') = K(\sqrt{1-k^2}\,)$ are given by

$$4K(k) = 2\pi{\vartheta_3}^2(0|\tau) \tag{2}$$

$$2iK'(k) = \pi\tau{\vartheta_3}^2(0|\tau), \tag{3}$$

where $K(k)$ is a complete ELLIPTIC INTEGRAL OF THE FIRST KIND and the complementary modulus is defined by

$$k'^2 \equiv 1-k^2, \tag{4}$$

with k the modulus.

see also AMPLITUDE, CHARACTERISTIC (ELLIPTIC INTEGRAL), ELLIPTIC FUNCTION, ELLIPTIC INTEGRAL, ELLIPTIC INTEGRAL SINGULAR VALUE, MODULAR ANGLE, NOME, PARAMETER, THETA FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 590, 1972.

Modulus (Quadratic Invariants)

The quantity $ps - rq$ obtained by letting

$$x = pX + qY \tag{1}$$
$$y = rX + sY \tag{2}$$

in

$$ax^2 + 2bxy + cy^2 \tag{3}$$

so that

$$A = ap^2 + 2bpr + cr^2 \tag{4}$$
$$B = apq + b(ps + qr) + crs \tag{5}$$
$$C = aq^2 + 2bqs + cs^2 \tag{6}$$

and

$$B^2 - AC = (ps - rq)^2(b^2 - ac), \tag{7}$$

is called the modulus.

Modulus (Set)

The name for the SET of INTEGERS modulo m, denoted $\mathbb{Z}\backslash m\mathbb{Z}$. If m is a PRIME p, then the modulus is a FINITE FIELD $\mathbb{F}_p = \mathbb{Z}\backslash p\mathbb{Z}$.

Moessner's Theorem

Write down the POSITIVE INTEGERS in row one, cross out every k_1th number, and write the partial sums of the remaining numbers in the row below. Now cross off every k_2th number and write the partial sums of the remaining numbers in the row below. Continue. For every POSITIVE INTEGER $k > 1$, if every kth number is ignored in row 1, every $(k-1)$th number in row 2, and every $(k+1-i)$th number in row i, then the kth row of partial sums will be the kth POWERS 1^k, 2^k, 3^k, ...

References

Conway, J. H. and Guy, R. K. "Moessner's Magic." In *The Book of Numbers.* New York: Springer-Verlag, pp. 63–65, 1996.

Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., pp. 268–277, 1991.

Long, C. T. "On the Moessner Theorem on Integral Powers." *Amer. Math. Monthly* **73**, 846–851, 1966.

Long, C. T. "Strike it Out—Add it Up." *Math. Mag.* **66**, 273–277, 1982.

Moessner, A. "Eine Bemerkung über die Potenzen der natürlichen Zahlen." *S.-B. Math.-Nat. Kl. Bayer. Akad. Wiss.* **29**, 1952.

Paasche, I. "Ein neuer Beweis des moessnerischen Satzes." *S.-B. Math.-Nat. Kl. Bayer. Akad. Wiss.* **1952**, 1–5, 1953.

Paasche, I. "Ein zahlentheoretische-logarithmischer 'Rechenstab'." *Math. Naturwiss. Unterr.* **6**, 26–28, 1953–54.

Paasche, I. "Eine Verallgemeinerung des moessnerschen Satzes." *Compositio Math.* **12**, 263–270, 1956.

Mohammed Sign

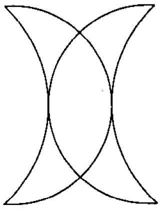

A curve consisting of two mirror-reversed intersecting crescents. This curve can be traced UNICURSALLY.

see also UNICURSAL CIRCUIT

Møiré Pattern

An interference pattern produced by overlaying similar but slightly offset templates. Møiré patterns can also be created be plotting series of curves on a computer screen. Here, the interference is provided by the discretization of the finite-sized pixels.

see also CIRCLES-AND-SQUARES FRACTAL

References

Cassin, C. *Visual Illusions in Motion with Møiré Screens: 60 Designs and 3 Plastic Screens.* New York: Dover, 1997.

Grafton, C. B. *Optical Designs in Motion with Møiré Overlays.* New York: Dover, 1976.

Mollweide's Formulas

Let a TRIANGLE have side lengths a, b, and c with opposite angles A, B, and C. Then

$$\frac{b-c}{a} = \frac{\sin[\frac{1}{2}(B-C)]}{\cos(\frac{1}{2}A)}$$
$$\frac{c-a}{b} = \frac{\sin[\frac{1}{2}(C-A)]}{\cos(\frac{1}{2}B)}$$
$$\frac{a-b}{c} = \frac{\sin[\frac{1}{2}(A-B)]}{\cos(\frac{1}{2}C)}.$$

see also NEWTON'S FORMULAS, TRIANGLE

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 146, 1987.

Mollweide Projection

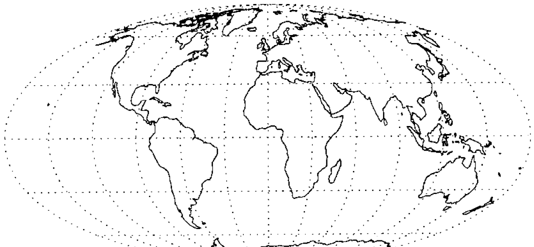

A MAP PROJECTION also called the ELLIPTICAL PROJECTION or HOMOLOGRAPHIC EQUAL AREA PROJECTION. The forward transformation is

$$x = \frac{2\sqrt{2}\,(\lambda - \lambda_0)\cos\theta}{\pi} \tag{1}$$

$$y = 2^{1/2}\sin\theta, \tag{2}$$

where θ is given by

$$2\theta + \sin(2\theta) = \pi\sin\phi. \tag{3}$$

NEWTON'S METHOD can then be used to compute θ' iteratively from

$$\Delta\theta' = -\frac{\theta' + \sin\theta' - \pi\sin\phi}{1 + \cos\theta'}, \tag{4}$$

where

$$\theta' = \tfrac{1}{2}\theta' \tag{5}$$

or, better yet,

$$\theta' = 2\sin^{-1}\left(\frac{2\phi}{\pi}\right) \tag{6}$$

can be used as a first guess.

The inverse FORMULAS are

$$\phi = \sin^{-1}\left[\frac{2\theta + \sin(2\theta)}{\pi}\right] \tag{7}$$

$$\lambda = \lambda_0 + \frac{\pi x}{2\sqrt{2}\,\cos\theta}, \tag{8}$$

where

$$\theta = \sin^{-1}\left(\frac{y}{\sqrt{2}}\right). \tag{9}$$

References

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 249–252, 1987.

Moment

The nth moment of a distribution about zero μ'_n is defined by

$$\mu'_n = \langle x^n \rangle, \tag{1}$$

where

$$\langle f(x) \rangle = \begin{cases} \sum f(x)P(x) & \text{discrete distribution} \\ \int f(x)P(x)\,dx & \text{continuous distribution.} \end{cases} \tag{2}$$

μ'_1, the MEAN, is usually simply denoted $\mu = \mu_1$. If the moment is instead taken about a point a,

$$\mu_n(a) = \langle (x-a)^n \rangle = \sum (x-a)^n P(x). \tag{3}$$

The moments are most commonly taken about the MEAN. These moments are denoted μ_n and are defined by

$$\mu_n \equiv \langle (x-\mu)^n \rangle, \tag{4}$$

with $\mu_1 = 0$. The moments about zero and about the MEAN are related by

$$\mu_2 = \mu'_2 - (\mu'_1)^2 \tag{5}$$

$$\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2(\mu'_1)^3 \tag{6}$$

$$\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2(\mu'_1)^2 - 3(\mu'_1)^4. \tag{7}$$

The second moment about the MEAN is equal to the VARIANCE

$$\mu_2 = \sigma^2, \tag{8}$$

where $\sigma = \sqrt{\mu_2}$ is called the STANDARD DEVIATION.

The related CHARACTERISTIC FUNCTION is defined by

$$\phi^{(n)}(0) \equiv \left[\frac{d^n\phi}{dt^n}\right]_{t=0} = i^n \mu_n(0). \tag{9}$$

The moments may be simply computed using the MOMENT-GENERATING FUNCTION,

$$\mu'_n = M^{(n)}(0). \tag{10}$$

A DISTRIBUTION is not uniquely specified by its moments, although it is by its CHARACTERISTIC FUNCTION.

see also CHARACTERISTIC FUNCTION, CHARLIER'S CHECK, CUMULANT-GENERATING FUNCTION, FACTORIAL MOMENT, KURTOSIS, MEAN, MOMENT-GENERATING FUNCTION, SKEWNESS, STANDARD DEVIATION, STANDARDIZED MOMENT, VARIANCE

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Moments of a Distribution: Mean, Variance, Skewness, and So Forth." §14.1 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 604–609, 1992.

Moment-Generating Function

Given a RANDOM VARIABLE $x \in R$, if there exists an $h > 0$ such that

$$M(t) \equiv \langle e^{tx} \rangle = \begin{cases} \sum_R e^{tx}P(x) & \text{for a discrete distribution} \\ \int_{-\infty}^{\infty} e^{tx}P(x)\,dx & \text{for a continuous distribution} \end{cases} \tag{1}$$

for $|t| < h$, then

$$M(t) \equiv \langle e^{tx} \rangle \tag{2}$$

is the moment-generating function.

$$\begin{aligned} M(t) &= \int_{-\infty}^{\infty} (1 + tx + \tfrac{1}{2!}t^2x^2 + \ldots)P(x)\,dx \\ &= 1 + tm_1 + \tfrac{1}{2!}t^2m_2 + \ldots, \end{aligned} \tag{3}$$

where m_r is the rth MOMENT about zero. The moment-generating function satisfies

$$\begin{aligned} M_{x+y}(t) &= \langle e^{t(x+y)} \rangle = \langle e^{tx} e^{ty} \rangle \\ &= \langle e^{tx} \rangle \langle e^{ty} \rangle = M_x(t) M_y(t). \end{aligned} \tag{4}$$

If $M(t)$ is differentiable at zero, then the nth MOMENTS about the ORIGIN are given by $M^n(0)$

$$M(t) = \langle e^{tx} \rangle \qquad M(0) = 1 \tag{5}$$

$$M'(t) = \langle x e^{tx} \rangle \qquad M'(0) = \langle x \rangle \tag{6}$$

$$M''(t) = \langle x^2 e^{tx} \rangle \qquad M''(0) = \langle x^2 \rangle \tag{7}$$

$$M^{(n)}(t) = \langle x^n e^{tx} \rangle \qquad M^{(n)}(0) = \langle x^n \rangle. \tag{8}$$

The MEAN and VARIANCE are therefore

$$\mu \equiv \langle x \rangle = M'(0) \tag{9}$$

$$\sigma^2 \equiv \langle x^2 \rangle - \langle x \rangle^2 = M''(0) - [M'(0)]^2. \tag{10}$$

It is also true that

$$\mu_n = \sum_{j=0}^{n} \binom{n}{j} (-1)^{n-j} \mu'_j (\mu'_1)^{n-j}, \tag{11}$$

where $\mu'_0 = 1$ and μ'_j is the jth moment about the origin.

It is sometimes simpler to work with the LOGARITHM of the moment-generating function, which is also called the CUMULANT-GENERATING FUNCTION, and is defined by

$$R(t) \equiv \ln[M(t)] \tag{12}$$

$$R'(t) = \frac{M'(t)}{M(t)} \tag{13}$$

$$R''(t) = \frac{M(t)M''(t) - [M'(t)]^2}{[M(t)]^2}. \tag{14}$$

But $M(0) = \langle 1 \rangle = 1$, so

$$\mu = M'(0) = R'(0) \tag{15}$$

$$\sigma^2 = M''(0) - [M'(0)]^2 = R''(0). \tag{16}$$

see also CHARACTERISTIC FUNCTION, CUMULANT, CUMULANT-GENERATING FUNCTION, MOMENT

References

Kenney, J. F. and Keeping, E. S. "Moment-Generating and Characteristic Functions," "Some Examples of Moment-Generating Functions," and "Uniqueness Theorem for Characteristic Functions." §4.6–4.8 in *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, pp. 72–77, 1951.

Momental Skewness

$$\alpha^{(m)} \equiv \tfrac{1}{2}\gamma_1 = \frac{\mu_3}{2\sigma^3},$$

where γ_1 is the FISHER SKEWNESS.

see also FISHER SKEWNESS, SKEWNESS

Monad

A mathematical object which consists of a set of a single element. The YIN-YANG is also known as the monad.

see also HEXAD, QUARTET, QUINTET, TETRAD, TRIAD, YIN-YANG

Money-Changing Problem

see COIN PROBLEM

Monge-Ampère Differential Equation

A second-order PARTIAL DIFFERENTIAL EQUATION of the form

$$Hr + 2Ks + Lt + M + N(rt - s^2) = 0,$$

where H, K, L, M, and N are functions of x, y, z, p, and q, and r, s, t, p, and q are defined by

$$\begin{aligned} r &= \frac{\partial^2 z}{\partial x^2} \\ s &= \frac{\partial^2 z}{\partial x \partial y} \\ t &= \frac{\partial^2 z}{\partial y^2} \\ p &= \frac{\partial z}{\partial x} \\ q &= \frac{\partial z}{\partial y}. \end{aligned}$$

The solutions are given by a system of differential equations given by Iyanaga and Kawada (1980).

References

Iyanaga, S. and Kawada, Y. (Eds.). "Monge-Ampère Equations." §276 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 879–880, 1980.

Monge's Chordal Theorem

see RADICAL CENTER

Monge's Form

A surface given by the form $z = F(x, y)$.

see also MONGE PATCH

Monge Patch

A Monge patch is a PATCH $\mathbf{x}: U \to \mathbb{R}^3$ of the form

$$\mathbf{x}(u, v) = (u, v, h(u, v)), \tag{1}$$

where U is an OPEN SET in $\mathbb{R}^2$ and $h: U \to \mathbb{R}$ is a differentiable function. The coefficients of the first FUNDAMENTAL FORM are given by

$$E = 1 + {h_u}^2 \tag{2}$$

$$F = h_u h_v \tag{3}$$

$$G = 1 + {h_v}^2 \tag{4}$$

and the second Fundamental Form by

$$e = \frac{h_{uu}}{\sqrt{1+{h_u}^2+{h_v}^2}} \tag{5}$$

$$f = \frac{h_{uv}}{\sqrt{1+{h_u}^2+{h_v}^2}} \tag{6}$$

$$g = \frac{g_{vv}}{\sqrt{1+{h_u}^2+{h_v}^2}}. \tag{7}$$

For a Monge patch, the Gaussian Curvature and Mean Curvature are

$$K = \frac{h_{uu}h_{vv}-{h_{uv}}^2}{(1+{h_u}^2+{h_v}^2)^2} \tag{8}$$

$$H = \frac{(1+{h_v}^2)h_{uu}-2h_uh_vh_{uv}+(1+{h_u}^2)h_{vv}}{(1+{h_u}^2+{h_v}^2)^{3/2}}. \tag{9}$$

see also Monge's Form, Patch

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 305–306, 1993.

Monge's Problem

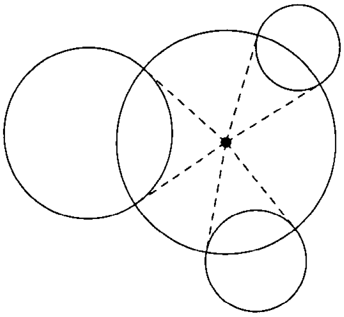

Draw a Circle that cuts three given Circles Perpendicularly. The solution is obtained by drawing the Radical Center R of the given three Circles. If it lies outside the three Circles, then the Circle with center R and Radius formed by the tangent from R to one of the given Circles intersects the given Circles perpendicularly. Otherwise, if R lies inside one of the circles, the problem is unsolvable.

see also Circle Tangents, Radical Center

References

Dörrie, H. "Monge's Problem." §31 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 151–154, 1965.

Monge's Shuffle

A Shuffle in which Cards from the top of the deck in the left hand are alternatively moved to the bottom and top of the deck in the right hand. If the deck is shuffled m times, the final position x_m and initial position x_0 of a card are related by

$$2^{m+1}x_m = (4p+1)[2^{m-1}+(-1)^{m-1}(2^{m-2}+\ldots+2+1)] +(-1)^{m-1}2x_0+2^m+(-1)^{m-1}$$

for a deck of $2p$ cards (Kraitchik 1942).

see also Cards, Shuffle

References

Conway, J. H. and Guy, R. K. "Fractions Cycle into Decimals." In *The Book of Numbers.* New York: Springer-Verlag, pp. 157–163, 1996.

Kraitchik, M. "Monge's Shuffle." §12.2.14 in *Mathematical Recreations.* New York: W. W. Norton, pp. 321–323, 1942.

Monge's Theorem

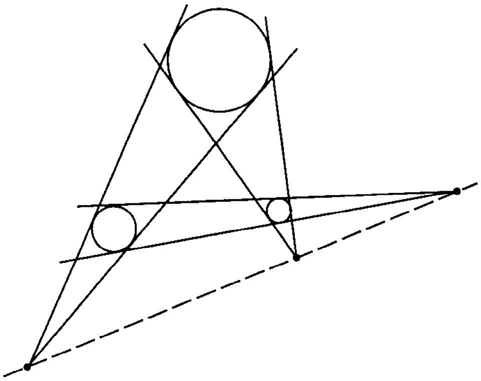

Draw three nonintersecting Circles in the plane, and the common tangent line for each pair of two. The points of intersection of the three pairs of tangent lines lie on a straight line.

References

Coxeter, H. S. M. "The Problem of Apollonius." *Amer. Math. Monthly* **75**, 5–15, 1968.

Graham, L. A. Problem 62 in *Ingenious Mathematical Problems and Methods.* New York: Dover, 1959. Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 115–117, 1990.

Walker, W. "Monge's Theorem in Many Dimensions." *Math. Gaz.* **60**, 185–188, 1976.

Monic Polynomial

A Polynomial in which the Coefficient of the highest Order term is 1.

see also Monomial

Monica Set

The nth Monica set M_n is defined as the set of Composite Numbers x for which $n|S(x)-S_p(x)$, where

$$x = a_0+a_1(10^1)+\ldots+a_d(10^d) = p_1p_2\cdots p_n, \tag{1}$$

and

$$S(x) = \sum_{j=0}^{d} a_j \tag{2}$$

$$S_p(x) = \sum_{i=1}^{m} S(p_i). \tag{3}$$

Every Monica set has an infinite number of elements. The Monica set M_n is a subset of the Suzanne Set S_n.

If x is a SMITH NUMBER, then it is a member of the Monica set M_n for all $n \in \mathbb{N}$. For any INTEGER $k > 1$, if x is a k-SMITH NUMBER, then $x \in M_{k-1}$.

see also SUZANNE SET

References

Smith, M. "Cousins of Smith Numbers: Monica and Suzanne Sets." *Fib. Quart.* **34**, 102–104, 1996.

Monkey and Coconut Problem

A DIOPHANTINE problem (i.e., one whose solution must be given in terms of INTEGERS) which seeks a solution to the following problem. Given n men and a pile of coconuts, each man in sequence takes $(1/n)$th of the coconuts and gives the m coconuts which do not divide equally to a monkey. When all n men have so divided, they divide the remaining coconuts five ways, and give the m coconuts which are left-over to the monkey. How many coconuts N were there originally? The solution is equivalent to solving the $n+1$ DIOPHANTINE EQUATIONS

$$\begin{aligned} N &= nA + m \\ (n-1)A &= nB + m \\ (n-1)B &= nC + m \\ &\vdots \\ (n-1)X &= nY + m \\ (n-1)Y &= nZ + m, \end{aligned}$$

and is given by

$$N = kn^{n+1} - m(n-1),$$

where k is an an arbitrary INTEGER (Gardner 1961).

For the particular case of $n = 5$ men and $m = 1$ left over coconuts, the 6 equations can be combined into the single DIOPHANTINE EQUATION

$$1,024N = 15,625F + 11,529,$$

where F is the number given to each man in the last division. The smallest POSITIVE solution in this case is $N = 15,621$ coconuts, corresponding to $k = 1$ and $F = 1,023$ (Gardner 1961). The following table shows how this rather large number of coconuts is divided under the scheme described above.

Removed	Given to Monkey	Left
		15,621
3,124	1	12,496
2,499	1	9,996
1,999	1	7,996
1,599	1	6,396
1,279	1	5,116
5 × 1023	1	0

If no coconuts are left for the monkey after the final n-way division (Williams 1926), then the original number of coconuts is

$$\begin{cases} (1+nk)n^n - (n-1) & n \text{ odd} \\ (n-1+nk)n^n - (n-1) & n \text{ even.} \end{cases}$$

The smallest POSITIVE solution for case $n = 5$ and $m = 1$ is $N = 3,121$ coconuts, corresponding to $k = 1$ and 1,020 coconuts in the final division (Gardner 1961). The following table shows how these coconuts are divided.

Removed	Given to Monkey	Left
		3,121
624	1	2,496
499	1	1,996
399	1	1,596
319	1	1,276
255	1	1,020
5 × 204	0	0

A different version of the problem having a solution of 79 coconuts is considered by Pappas (1989).

see also DIOPHANTINE EQUATION—LINEAR, PELL EQUATION

References

Anning, N. "Monkeys and Coconuts." *Math. Teacher* **54**, 560–562, 1951.

Bowden, J. "The Problem of the Dishonest Men, the Monkeys, and the Coconuts." In *Special Topics in Theoretical Arithmetic.* Lancaster, PA: Lancaster Press, pp. 203–212, 1936.

Gardner, M. "The Monkey and the Coconuts." Ch. 9 in *The Second Scientific American Book of Puzzles & Diversions: A New Selection.* New York: Simon and Schuster, 1961.

Kirchner, R. B. "The Generalized Coconut Problem." *Amer. Math. Monthly* **67**, 516–519, 1960.

Moritz, R. E. "Solution to Problem 3,242." *Amer. Math. Monthly* **35**, 47–48, 1928.

Ogilvy, C. S. and Anderson, J. T. *Excursions in Number Theory.* New York: Dover, pp. 52–54, 1988.

Olds, C. D. *Continued Fractions.* New York: Random House, pp. 48–50, 1963.

Pappas, T. "The Monkey and the Coconuts." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 226–227 and 234, 1989.

Williams, B. A. "Coconuts." *The Saturday Evening Post,* Oct. 9, 1926.

Monkey Saddle

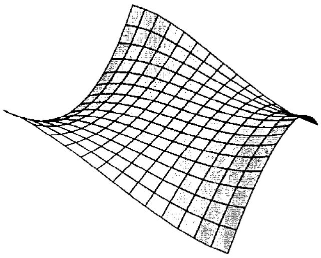

A SURFACE which a monkey can straddle with both his two legs and his tail. A simply Cartesian equation for such a surface is

$$z = x(x^2 - 3y^2), \tag{1}$$

which can also be given by the parametric equations

$$x(u,v) = u \tag{2}$$

$$y(u,v) = v \tag{3}$$

$$z(u,v) = u^3 - 3uv^2. \tag{4}$$

The coefficients of the first and second FUNDAMENTAL FORMS of the monkey saddle are given by

$$e = \frac{6u}{\sqrt{1+9u^4+18u^2v^2+9v^4}} \tag{5}$$

$$f = -\frac{6v}{\sqrt{1+9u^4+18u^2v^2+9v^4}} \tag{6}$$

$$g = -\frac{6u}{\sqrt{1+9u^4+18u^2v^2+9v^4}} \tag{7}$$

$$E = 1 + 9(u^2 - v^2)^2 \tag{8}$$

$$F = -18uv(u^2 - v^2) \tag{9}$$

$$G = 1 + 36u^2v^2, \tag{10}$$

giving RIEMANNIAN METRIC

$$ds^2 = [1 + (3u^2 - 3v^2)^2]\,du^2 - 2[18uv(u^2 - v^2)]\,du\,dv + (1 + 36u^2v^2)\,dv^2, \tag{11}$$

AREA ELEMENT

$$dA = \sqrt{1+9u^4+18u^2v^2+9v^4}\,du \wedge dv, \tag{12}$$

and GAUSSIAN and MEAN CURVATURES

$$K = -\frac{36(u^2+v^2)}{(1+9u^4+18u^2v^2+9v^4)^2} \tag{13}$$

$$H = \frac{27u(-u^4+2u^2v^2+3v^4)}{(1+9u^4+18u^2v^2+9v^4)^{3/2}} \tag{14}$$

(Gray 1993). Every point of the monkey saddle except the origin has NEGATIVE GAUSSIAN CURVATURE.

see also CROSSED TROUGH, PARTIAL DERIVATIVE

References

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, p. 365, 1969.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 213–215, 262–263, and 288–289, 1993.

Hilbert, D. and Cohn-Vossen, S. *Geometry and the Imagination.* New York: Chelsea, p. 202, 1952.

Monochromatic Forced Triangle

Given a COMPLETE GRAPH K_n which is two-colored, the number of forced monochromatic TRIANGLES is at least

$$\begin{cases} \frac{1}{3}u(u-1)(u-2) & \text{for } n = 2u \\ \frac{2}{3}u(u-1)(4u+1) & \text{for } n = 4u+1 \\ \frac{2}{3}u(u+1)(4u-1) & \text{for } n = 4u+3. \end{cases}$$

The first few numbers of monochromatic forced triangles are 0, 0, 0, 0, 0, 2, 4, 8, 12, 20, 28, 40, ... (Sloane's A014557).

see also COMPLETE GRAPH, EXTREMAL GRAPH

References

Goodman, A. W. "On Sets of Acquaintances and Strangers at Any Party." *Amer. Math. Monthly* **66**, 778–783, 1959.

Sloane, N. J. A. Sequence A014553 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Monodromy

A general concept in CATEGORY THEORY involving the globalization of local MORPHISMS.

see also HOLONOMY

Monodromy Group

A technically defined GROUP characterizing a system of linear differential equations

$$y_j' = \sum_{k=1}^{n} a_{jk}(x) y_k$$

for $j = 1, \ldots, n$, where a_{jk} are COMPLEX ANALYTIC FUNCTIONS of x in a given COMPLEX DOMAIN.

see also HILBERT'S 21ST PROBLEM, RIEMANN P-SERIES

References

Iyanaga, S. and Kawada, Y. (Eds.). "Monodromy Groups." §253B in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 793, 1980.

Monodromy Theorem

If a COMPLEX function f is ANALYTIC in a DISK contained in a simply connected DOMAIN D and f can be ANALYTICALLY CONTINUED along every polygonal arc in D, then f can be ANALYTICALLY CONTINUED to a single-valued ANALYTIC FUNCTION on all of D!

see also ANALYTIC CONTINUATION

Monogenic Function

If

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

is the same for all paths in the COMPLEX PLANE, then $f(z)$ is said to be monogenic at z_0. Monogenic therefore essentially means having a single DERIVATIVE at a point. Functions are either monogenic or have infinitely many DERIVATIVES (in which case they are called POLYGENIC); intermediate cases are not possible.

see also POLYGENIC FUNCTION

References
Newman, J. R. *The World of Mathematics, Vol. 3.* New York: Simon & Schuster, p. 2003, 1956.

Monohedral Tiling

A TILING is which all tiles are congruent.

see also ANISOHEDRAL TILING, ISOHEDRAL TILING

References
Berglund, J. "Is There a k-Anisohedral Tile for $k \geq 5$?" *Amer. Math. Monthly* **100**, 585-588, 1993.
Grünbaum, B. and Shephard, G. C. "The 81 Types of Isohedral Tilings of the Plane." *Math. Proc. Cambridge Philos. Soc.* **82**, 177-196, 1977.

Monoid

A GROUP-like object which fails to be a GROUP because elements need not have an inverse within the object. A monoid S must also be ASSOCIATIVE and an IDENTITY ELEMENT $I \in S$ such that for all $a \in S$, $1a = a1 = a$. A monoid is therefore a SEMIGROUP with an identity element. A monoid must contain at least one element.

The numbers of free idempotent monoids on n letters are 1, 2, 7, 160, 332381, ... (Sloane's A005345).

see also BINARY OPERATOR, GROUP, SEMIGROUP

References
Rosenfeld, A. *An Introduction to Algebraic Structures.* New York: Holden-Day, 1968.
Sloane, N. J. A. Sequence A005345/M1820 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Monomial

A POLYNOMIAL consisting of a single term.

see also BINOMIAL, MONIC POLYNOMIAL, POLYNOMIAL, TRINOMIAL

Monomino

The unique 1-POLYOMINO, consisting of a single SQUARE.

see also DOMINO, TRIOMINO

References
Gardner, M. "Polyominoes." Ch. 13 in *The Scientific American Book of Mathematical Puzzles & Diversions.* New York: Simon and Schuster, pp. 124–140, 1959.

Monomorph

An INTEGER which is expressible in only one way in the form $x^2 + Dy^2$ or $x^2 - Dy^2$ where x^2 is RELATIVELY PRIME to Dy^2. If the INTEGER is expressible in more than one way, it is called a POLYMORPH.

see also ANTIMORPH, IDONEAL NUMBER, POLYMORPH

Monomorphism

An INJECTIVE MORPHISM.

Monotone

Another word for monotonic.

see also MONOTONIC FUNCTION, MONOTONIC SEQUENCE, MONOTONIC VOTING

Monotone Decreasing

Always decreasing; never remaining constant or increasing.

Monotone Increasing

Always increasing; never remaining constant or decreasing.

Monotonic Function

A function which is either entirely nonincreasing or nondecreasing. A function is monotonic if its first DERIVATIVE (which need not be continuous) does not change sign.

Monotonic Sequence

A SEQUENCE $\{a_n\}$ such that either (1) $a_{i+1} \geq a_i$ for every $i \geq 1$, or (2) $a_{i+1} \leq a_i$ for every $i \geq 1$.

Monotonic Voting

A term in SOCIAL CHOICE THEORY meaning a change favorable for X does not hurt X.

see also ANONYMOUS, DUAL VOTING

Monster Group

The highest order SPORADIC GROUP M. It has ORDER

$$2^{46} \cdot 3^{20} \cdot 5^9 \cdot 7^6 \cdot 11^2 \cdot 13^3 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71,$$

and is also called the FRIENDLY GIANT GROUP. It was constructed in 1982 by Robert Griess as a GROUP of ROTATIONS in 196,883-D space.

see also BABY MONSTER GROUP, BIMONSTER, LEECH LATTICE

References
Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, p. viii, 1985.
Conway, J. H. and Norton, S. P. "Monstrous Moonshine." *Bull. London Math. Soc.* **11**, 308–339, 1979.
Conway, J. H. and Sloane, N. J. A. "The Monster Group and its 196884-Dimensional Space" and "A Monster Lie Algebra?" Chs. 29–30 in *Sphere Packings, Lattices, and Groups, 2nd ed.* New York: Springer-Verlag, pp. 554–571, 1993.
Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas/M.html`.

Monte Carlo Integration

In order to integrate a function over a complicated DOMAIN D, Monte Carlo integration picks random points over some simple DOMAIN D' which is a superset of D, checks whether each point is within D, and estimates the AREA of D (VOLUME, n-D CONTENT, etc.) as the AREA of D' multiplied by the fraction of points falling within D'.

An estimate of the uncertainty produced by this technique is given by

$$\int f\,dV \approx V\langle f\rangle \pm \sqrt{\frac{\langle f^2\rangle - \langle f\rangle^2}{N}}.$$

see also MONTE CARLO METHOD

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Simple Monte Carlo Integration" and "Adaptive and Recursive Monte Carlo Methods." §7.6 and 7.8 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 295–299 and 306–319, 1992.

Monte Carlo Method

Any method which solves a problem by generating suitable random numbers and observing that fraction of the numbers obeying some property or properties. The method is useful for obtaining numerical solutions to problems which are too complicated to solve analytically. The most common application of the Monte Carlo method is MONTE CARLO INTEGRATION.

see also MONTE CARLO INTEGRATION

References

Sobol, I. M. *A Primer for the Monte Carlo Method.* Boca Raton, FL: CRC Press, 1994.

Montel's Theorem

Let $f(z)$ be an analytic function of z, regular in the half-strip S defined by $a < x < b$ and $y > 0$. If $f(z)$ is bounded in S and tends to a limit l as $y \to \infty$ for a certain fixed value ξ of x between a and b, then $f(z)$ tends to this limit l on every line $x = x_0$ in S, and $f(z) \to l$ uniformly for $a + \delta \leq x_0 \leq b - \delta$.

see also VITALI'S CONVERGENCE THEOREM

References

Titchmarsh, E. C. *The Theory of Functions, 2nd ed.* Oxford, England: Oxford University Press, p. 170, 1960.

Monty Hall Dilemma

see MONTY HALL PROBLEM

Monty Hall Problem

The Monty Hall problem is named for its similarity to the *Let's Make a Deal* television game show hosted by Monty Hall. The problem is stated as follows. Assume that a room is equipped with three doors. Behind two are goats, and behind the third is a shiny new car. You are asked to pick a door, and will win whatever is behind it. Let's say you pick door 1. Before the door is opened, however, someone who knows what's behind the doors (Monty Hall) opens *one of the other* two doors, revealing a goat, and asks you if you wish to change your selection to the third door (i.e., the door which neither you picked nor he opened). The Monty Hall problem is deciding whether you do.

The correct answer is that you *do* want to switch. If you do not switch, you have the expected 1/3 chance of winning the car, since no matter whether you initially picked the correct door, Monty will show you a door with a goat. But after Monty has eliminated one of the doors for you, you obviously do not improve your chances of winning to better than 1/3 by sticking with your original choice. If you now switch doors, however, there is a 2/3 chance you will win the car (counterintuitive though it seems).

d_1	d_2	Winning Probability
pick	stick	1/3
pick	switch	2/3

The problem can be generalized to four doors as follows. Let one door conceal the car, with goats behind the other three. Pick a door d_1. Then the host will open one of the nonwinners and give you the option of switching. Call your new choice (which could be the same as d_1 if you don't switch) d_2. The host will then open a second nonwinner, and you must decide for choice d_3 if you want to stick to d_2 or switch to the remaining door. The probabilities of winning are shown below for the four possible strategies.

d_1	d_2	d_3	Winning Probability
pick	stick	stick	4/8
pick	switch	stick	3/8
pick	stick	switch	6/8
pick	switch	switch	5/8

The above results are characteristic of the best strategy for the n-stage Monty Hall problem: stick until the last choice, then switch.

see also ALIAS' PARADOX

References

Barbeau, E. "The Problem of the Car and Goats." *CMJ* **24**, 149, 1993.

Bogomolny, A. "Monty Hall Dilemma." `http://www.cut-the-knot.com/hall.html`.

Dewdney, A. K. *200% of Nothing.* New York: Wiley, 1993.

Donovan, D. "The WWW Tackles the Monty Hall Problem." `http://math.rice.edu/~ddonovan/montyurl.html`.

Ellis, K. M. "The Monty Hall Problem." `http://www.io.com/~kmellis/monty.html`.

Gardner, M. *Aha! Gotcha: Paradoxes to Puzzle and Delight.* New York: W. H. Freeman, 1982.
Gillman, L. "The Car and the Goats." *Amer. Math. Monthly* **99**, 3, 1992.
Selvin, S. "A Problem in Probability." *Amer. Stat.* **29**, 67, 1975.
vos Savant, M. *The Power of Logical Thinking.* New York: St. Martin's Press, 1996.

Moore Graph

A Graph with Diameter d and Girth $2d+1$. Moore graphs have Diameter of at most 2. Every Moore graph is both Regular and distance regular. Hoffman and Singleton (1960) show that k-regular Moore graphs with Diameter 2 have $k \in \{2, 3, 7, 57\}$.

References

Godsil, C. D. "Problems in Algebraic Combinatorics." *Electronic J. Combinatorics* **2**, F1, 1–20, 1995. `http://www.combinatorics.org/Volume_2/volume2.html#F1`.
Hoffman, A. J. and Singleton, R. R. "On Moore Graphs of Diameter Two and Three." *IBM J. Res. Develop.* **4**, 497–504, 1960.

Moore-Penrose Generalized Matrix Inverse

Given an $m \times n$ Matrix B, the Moore-Penrose generalized Matrix Inverse is a unique $n \times m$ Matrix B^+ which satisfies

$$\mathsf{B}\mathsf{B}^+\mathsf{B} = \mathsf{B} \tag{1}$$

$$\mathsf{B}^+\mathsf{B}\mathsf{B}^+ = \mathsf{B}^+ \tag{2}$$

$$(\mathsf{B}\mathsf{B}^+)^{\mathrm{T}} = \mathsf{B}\mathsf{B}^+ \tag{3}$$

$$(\mathsf{B}^+\mathsf{B})^{\mathrm{T}} = \mathsf{B}^+\mathsf{B}. \tag{4}$$

It is also true that

$$\mathbf{z} = \mathsf{B}^+\mathbf{c} \tag{5}$$

is the shortest length Least Squares solution to the problem

$$\mathsf{B}\mathbf{z} = \mathbf{c}. \tag{6}$$

If the inverse of $(\mathsf{B}^{\mathrm{T}}\mathsf{B})$ exists, then

$$\mathsf{B}^+ = (\mathsf{B}^{\mathrm{T}}\mathsf{B})^{-1}\mathsf{B}^{\mathrm{T}}, \tag{7}$$

where B^{T} is the Matrix Transpose, as can be seen by premultiplying both sides of (7) by B^{T} to create a Square Matrix which can then be inverted,

$$\mathsf{B}^{\mathrm{T}}\mathsf{B}\mathbf{z} = \mathsf{B}^{\mathrm{T}}\mathbf{c}, \tag{8}$$

giving

$$\begin{aligned} \mathbf{z} &= (\mathsf{B}^{\mathrm{T}}\mathsf{B})^{-1}\mathsf{B}^{\mathrm{T}}\mathbf{c} \\ &\equiv \mathsf{B}^+\mathbf{c}. \end{aligned} \tag{9}$$

see also Least Squares Fitting, Matrix Inverse

References

Ben-Israel, A. and Greville, T. N. E. *Generalized Inverses: Theory and Applications.* New York: Wiley, 1977.
Lawson, C. and Hanson, R. *Solving Least Squares Problems.* Englewood Cliffs, NJ: Prentice-Hall, 1974.
Penrose, R. "A Generalized Inverse for Matrices." *Proc. Cambridge Phil. Soc.* **51**, 406–413, 1955.

Mordell Conjecture

Diophantine Equations that give rise to surfaces with two or more holes have only finite many solutions in Gaussian Integers with no common factors. Fermat's equation has $(n-1)(n-2)/2$ Holes, so the Mordell conjecture implies that for each Integer $n \geq 3$, the Fermat Equation has at most a finite number of solutions. This conjecture was proved by Faltings (1984).

see also Fermat Equation, Fermat's Last Theorem, Safarevich Conjecture, Shimura-Taniyama Conjecture

References

Faltings, G. "Die Vermutungen von Tate und Mordell." *Jahresber. Deutsch. Math.-Verein* **86**, 1–13, 1984.
Ireland, K. and Rosen, M. "The Mordell Conjecture." §20.3 in *A Classical Introduction to Modern Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 340–342, 1990.

Mordell Integral

The integral

$$\phi(t,u) = \int \frac{e^{\pi i t x^2 + 2\pi i u x}}{e^{2\pi i x} - 1}\, dx$$

which is related to the Theta Functions, Mock Theta Functions, and Riemann Zeta Function.

Mordell-Weil Theorem

For Elliptic Curves over the Rationals, $\mathbb{Q}$, the number of generators of the set of Rational Points is always finite. This theorem was proved by Mordell in 1921 and extended by Weil in 1928 to Abelian Varieties over Number Fields.

References

Ireland, K. and Rosen, M. "The Mordell-Weil Theorem." Ch. 19 in *A Classical Introduction to Modern Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 319–338, 1990.

Morera's Theorem

If $f(z)$ is continuous in a simply connected region D and satisfies

$$\int_\gamma f\, dz = 0$$

for all closed Contours γ in D, then $f(z)$ is Analytic in D.

see also Cauchy Integral Theorem

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 373–374, 1985.

Morgado Identity

An identity satisfied by w GENERALIZED FIBONACCI NUMBERS:

$$\begin{aligned}4w_nw_{n+1}w_{n+2}w_{n+4}w_{n+5}w_{n+6}\\ +e^2q^{2n}(w_nU_4U_5-w_{n+1}U_2U_6-w_nU_1U_8)^2\\ =(w_{n+1}w_{n+2}w_{n+6}+w_nw_{n+4}w_{n+5})^2,\end{aligned} \quad (1)$$

where

$$e\equiv pab-qa^2-b^2 \quad (2)$$

$$U_n\equiv w_n(0,1;p,q). \quad (3)$$

see also GENERALIZED FIBONACCI NUMBER

References

Morgado, J. "Note on Some Results of A. F. Horadam and A. G. Shannon Concerning a Catalan's Identity on Fibonacci Numbers." *Portugaliae Math.* **44**, 243–252, 1987.

Morgan-Voyce Polynomial

Polynomials related to the BRAHMAGUPTA POLYNOMIALS. They are defined by the RECURRENCE RELATIONS

$$b_n(x)=xB_{n-1}(x)+b_{n-1}(x) \quad (1)$$

$$B_n(x)=(x+1)B_{n-1}(x)+b_{n-1}(x) \quad (2)$$

for $n\geq 1$, with

$$b_0(x)=B_0(x)=1. \quad (3)$$

Alternative recurrences are

$$B_{n+1}B_{n-1}-{B_n}^2=-1 \quad (4)$$

$$b_{n+1}b_{n-1}-{b_n}^2=x. \quad (5)$$

The polynomials can be given explicitly by the sums

$$B_n(x)=\sum_{k=0}^{n}\binom{n+k-1}{n-k} \quad (6)$$

$$b_n(x)=\sum_{k=0}^{n}\binom{n+k}{n-k}. \quad (7)$$

Defining the MATRIX

$$\mathsf{Q}=\begin{bmatrix}x+2 & -1\\ 1 & 0\end{bmatrix} \quad (8)$$

gives the identities

$$\mathsf{Q}^n=\begin{bmatrix}B_n & -B_{n-1}\\ B_{n-1} & -B_{n-2}\end{bmatrix} \quad (9)$$

$$\mathsf{Q}^n-\mathsf{Q}^{n-1}=\begin{bmatrix}b_n & -b_{n-1}\\ b_{n-1} & -b_{n-2}\end{bmatrix}. \quad (10)$$

Defining

$$\cos\theta=\tfrac{1}{2}(x+2) \quad (11)$$

$$\cosh\phi=\tfrac{1}{2}(x+2) \quad (12)$$

gives

$$B_n(x)=\frac{\sin[(n+1)\theta]}{\sin\theta} \quad (13)$$

$$B_n(x)=\frac{\sinh[(n+1)\phi]}{\sinh\phi} \quad (14)$$

and

$$b_n(x)=\frac{\cos[\frac{1}{2}(2n+1)\theta]}{\cos(\frac{1}{2}\theta)} \quad (15)$$

$$b_n(x)=\frac{\cosh[\frac{1}{2}(2n+1)\phi]}{\cosh(\frac{1}{2}\theta)}. \quad (16)$$

The Morgan-Voyce polynomials are related to the FIBONACCI POLYNOMIALS $F_n(x)$ by

$$b_n(x^2)=F_{2n+1}(x) \quad (17)$$

$$B_n(x^2)=\frac{1}{x}F_{2n+2}(x) \quad (18)$$

(Swamy 1968).

$B_n(x)$ satisfies the ORDINARY DIFFERENTIAL EQUATION

$$x(x+4)y''+3(x+2)y'-n(n+2)y=0, \quad (19)$$

and $b_n(x)$ the equation

$$x(x+4)y''+2(x+1)y'-n(n+1)y=0. \quad (20)$$

These and several other identities involving derivatives and integrals of the polynomials are given by Swamy (1968).

see also BRAHMAGUPTA POLYNOMIAL, FIBONACCI POLYNOMIAL

References

Lahr, J. "Fibonacci and Lucas Numbers and the Morgan-Voyce Polynomials in Ladder Networks and in Electric Line Theory." In *Fibonacci Numbers and Their Applications* (Ed. G. E. Bergum, A. N. Philippou, and A. F. Horadam). Dordrecht, Netherlands: Reidel, 1986.

Morgan-Voyce, A. M. "Ladder Network Analysis Using Fibonacci Numbers." *IRE Trans. Circuit Th.* **CT-6**, 321–322, Sep. 1959.

Swamy, M. N. S. "Properties of the Polynomials Defined by Morgan-Voyce." *Fib. Quart.* **4**, 73–81, 1966.

Swamy, M. N. S. "More Fibonacci Identities." *Fib. Quart.* **4**, 369–372, 1966.

Swamy, M. N. S. "Further Properties of Morgan-Voyce Polynomials." *Fib. Quart.* **6**, 167–175, 1968.

Morley Centers

The CENTROID of MORLEY'S TRIANGLE is called Morley's first center. It has TRIANGLE CENTER FUNCTION

$$\alpha = \cos(\tfrac{1}{3}A) + 2\cos(\tfrac{1}{3}B)\cos(\tfrac{1}{3}C).$$

The PERSPECTIVE CENTER of MORLEY'S TRIANGLE with reference TRIANGLE ABC is called Morley's second center. The TRIANGLE CENTER FUNCTION is

$$\alpha = \sec(\tfrac{1}{3}A).$$

see also CENTROID (GEOMETRIC), MORLEY'S THEOREM, PERSPECTIVE CENTER

References

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Kimberling, C. "1st and 2nd Morley Centers." `http://www.evansville.edu/~ck6/tcenters/recent/morley.html`.

Oakley, C. O. and Baker, J. C. "The Morley Trisector Theorem." *Amer. Math. Monthly* **85**, 737–745, 1978.

Morley's Formula

$$\sum_{k=0}^{\infty} \binom{m}{k}^3 = 1 + \binom{m}{1}^3 + \left[\frac{m(m+1)}{1\cdot 2}\right]^3 + \ldots$$
$$= \frac{\Gamma(1-\frac{3}{2}m)}{[\Gamma(1-\frac{1}{2}m)]^3}\cos(\tfrac{1}{2}m\pi),$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT and $\Gamma(z)$ is the GAMMA FUNCTION.

Morley's Theorem

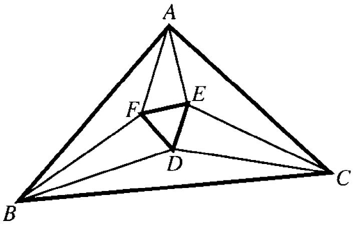

The points of intersection of the adjacent TRISECTORS of the ANGLES of any TRIANGLE ΔABC are the VERTICES of an EQUILATERAL TRIANGLE ΔDEF known as MORLEY'S TRIANGLE. Taylor and Marr (1914) give two geometric proofs and one trigonometric proof.

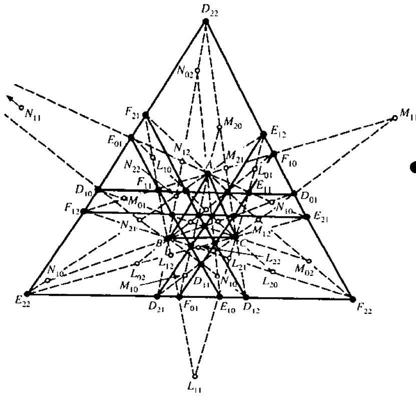

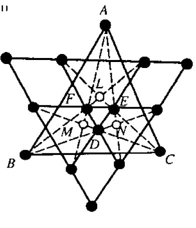

A generalization of MORLEY'S THEOREM was discovered by Morley in 1900 but first published by Taylor and Marr (1914). Each ANGLE of a TRIANGLE ΔABC has six trisectors, since each interior angle trisector has two associated lines making angles of 120° with it. The generalization of Morley's theorem states that these trisectors intersect in 27 points (denoted D_{ij}, E_{ij}, F_{ij}, for $i, j = 0$, 1, 2) which lie six by six on nine lines. Furthermore, these lines are in three triples of PARALLEL lines, ($D_{22}E_{22}$, $E_{12}D_{21}$, $F_{10}F_{01}$), ($D_{22}F_{22}$, $F_{21}D_{12}$, $E_{01}E_{10}$), and ($E_{22}F_{22}$, $F_{12}E_{21}$, $D_{10}D_{01}$), making ANGLES of 60° with one another (Taylor and Marr 1914, Johnson 1929, p. 254).

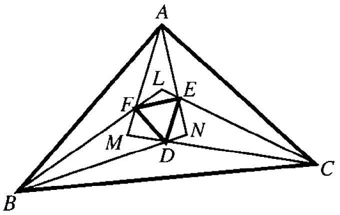

Let L, M, and N be the other trisector-trisector intersections, and let the 27 points L_{ij}, M_{ij}, N_{ij} for $i, j = 0$, 1, 2 be the ISOGONAL CONJUGATES of D, E, and F. Then these points lie 6 by 6 on 9 CONICS through ΔABC. In addition, these CONICS meet 3 by 3 on the CIRCUMCIRCLE, and the three meeting points form an EQUILATERAL TRIANGLE whose sides are PARALLEL to those of ΔDEF.

see also CONIC SECTION, MORLEY CENTERS, TRISECTION

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 47–50, 1967.

Gardner, M. *Martin Gardner's New Mathematical Diversions from Scientific American.* New York: Simon and Schuster, pp. 198 and 206, 1966.

Honsberger, R. "Morley's Theorem." Ch. 8 in *Mathematical Gems I.* Washington, DC: Math. Assoc. Amer., pp. 92–98, 1973.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 253–256, 1929.

Kimberling, C. "Hofstadter Points." *Nieuw Arch. Wiskunder* **12**, 109–114, 1994.

Marr, W. L. "Morley's Trisection Theorem: An Extension and Its Relation to the Circles of Apollonius." *Proc. Edinburgh Math. Soc.* **32**, 136–150, 1914.

Oakley, C. O. and Baker, J. C. "The Morley Trisector Theorem." *Amer. Math. Monthly* **85**, 737–745, 1978.

Pappas, T. "Trisecting & the Equilateral Triangle." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 174, 1989.

Taylor, F. G. "The Relation of Morley's Theorem to the Hessian Axis and Circumcentre." *Proc. Edinburgh Math. Soc.* **32**, 132–135, 1914.

Taylor, F. G. and Marr, W. L. "The Six Trisectors of Each of the Angles of a Triangle." *Proc. Edinburgh Math. Soc.* **32**, 119–131, 1914.

Morley's Triangle

An EQUILATERAL TRIANGLE considered by MORLEY'S THEOREM with side lengths

$$8R\sin(\tfrac{1}{3}A)\sin(\tfrac{1}{3}B)\sin(\tfrac{1}{3}C),$$

where R is the CIRCUMRADIUS of the original TRIANGLE.

Morphism

A map between two objects in an abstract CATEGORY.

1. A general morphism is called a HOMOMORPHISM,
2. An injective morphism is called a MONOMORPHISM,
3. A surjective morphism is an EPIMORPHISM,
4. A bijective morphism is called an ISOMORPHISM (if there is an isomorphism between two objects, then we say they are isomorphic),
5. A surjective morphism from an object to itself is called an ENDOMORPHISM, and
6. An ISOMORPHISM between an object and itself is called an AUTOMORPHISM.

see also AUTOMORPHISM, EPIMORPHISM, HOMEOMORPHISM, HOMOMORPHISM, ISOMORPHISM, MONOMORPHISM, OBJECT

Morrie's Law

$$\cos(20^\circ)\cos(40^\circ)\cos(80^\circ)=\tfrac{1}{8}.$$

This identity was referred to by Feynman (Gleick 1992). It is a special case of the general identity

$$2^k\prod_{j=0}^{k-1}\cos(2^j a)=\frac{\sin(2^k a)}{\sin a},$$

with $k = 3$ and $a = 20^\circ$ (Beyer *et al.* 1996).

References

Anderson, E. C. "Morrie's Law and Experimental Mathematics." To appear in *J. Recr. Math.*

Beyer, W. A.; Louck, J. D.; Zeilberger, D. "A Generalization of a Curiosity that Feynman Remembered All His Life." *Math. Mag.* **69**, 43–44, 1996.

Gleick, J. *Genius: The Life and Science of Richard Feynman.* New York: Pantheon Books, p. 47, 1992.

Morse Inequalities

Topological lower bounds in terms of BETTI NUMBERS for the number of critical points form a smooth function on a smooth MANIFOLD.

Morse Theory

"CALCULUS OF VARIATIONS in the large" which uses nonlinear techniques to address problems in the CALCULUS OF VARIATIONS. Morse theory applied to a FUNCTION g on a MANIFOLD W with $g(M) = 0$ and $g(M') = 1$ shows that every COBORDISM can be realized as a finite sequence of SURGERIES. Conversely, a sequence of SURGERIES gives a COBORDISM.

see also CALCULUS OF VARIATIONS, COBORDISM, SURGERY

Morse-Thue Sequence

see THUE-MORSE SEQUENCE

Mortal

A nonempty finite set of $n \times n$ MATRICES with INTEGER entries for which there exists some product of the MATRICES in the set which is equal to the zero MATRIX.

Mortality Problem

For a given n, is the problem of determining if a set is MORTAL solvable? $n = 1$ is solvable, $n = 2$ is unknown, and $n \geq 3$ is unsolvable.

see also LIFE EXPECTANCY

Morton-Franks-Williams Inequality

Let E be the largest and e the smallest POWER of ℓ in the HOMFLY POLYNOMIAL of an oriented LINK, and i be the BRAID INDEX. Then the MORTON-FRANKS-WILLIAMS INEQUALITY holds,

$$i \geq \tfrac{1}{2}(E-e)+1$$

(Franks and Williams 1985, Morton 1985). The inequality is sharp for all PRIME KNOTS up to 10 crossings with the exceptions of 09_{042}, 09_{049}, 10_{132}, 10_{150}, and 10_{156}.

see also BRAID INDEX

References

Franks, J. and Williams, R. F. "Braids and the Jones Polynomial." *Trans. Amer. Math. Soc.* **303**, 97–108, 1987.

Mosaic

see TESSELLATION

Moser

The very LARGE NUMBER consisting of the number 2 inside a MEGA-gon.

see also MEGA, MEGISTRON

Moser's Circle Problem

see CIRCLE CUTTING

Moss's Egg

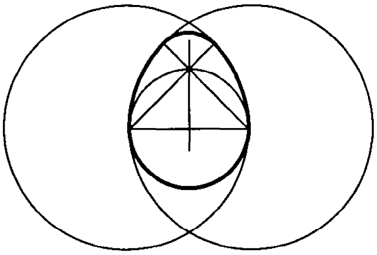

An OVAL whose construction is illustrated in the above diagram.

see also EGG, OVAL

References

Dixon, R. *Mathographics.* New York: Dover, p. 5, 1991.

Motzkin Number

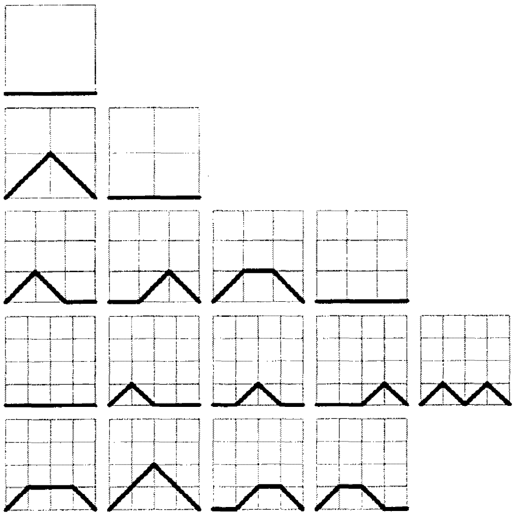

The Motzkin numbers enumerate various combinatorial objects. Donaghey and Shapiro (1977) give 14 different manifestations of these numbers. In particular, they give the number of paths from (0, 0) to (n, 0) which never dip below $y = 0$ and are made up only of the steps (1, 0), (1, 1), and (1, −1), i.e., $\rightarrow$, $\nearrow$, and $\searrow$. The first are 1, 2, 4, 9, 21, 51, ... (Sloane's A001006). The Motzkin number GENERATING FUNCTION $M(z)$ satisfies

$$M = 1 + xM + x^2M^2 \tag{1}$$

and is given by

$$\begin{aligned} M(x) &= \frac{1 - x - \sqrt{1 - 2x - 3x^2}}{2x^2} \\ &= 1 + x + 2x^2 + 4x^3 + 9x^4 + 21x^5 + \ldots, \end{aligned} \tag{2}$$

or by the RECURRENCE RELATION

$$M_n = a_{n-1} + \sum_{k=0}^{n-2} M_k M_{n-2-k} \tag{3}$$

with $M_0 = 1$. The Motzkin number M_n is also given by

$$M_n = -\frac{1}{2} \sum_{\substack{a+b=n+2 \\ a\geq 0, b\geq 0}} (-3)^a \binom{\frac{1}{2}}{a}\binom{\frac{1}{2}}{b} \tag{4}$$

$$= \frac{(-1)^{n+1}}{2^{2n+5}} \sum_{\substack{a+b=n+2 \\ a\geq 0, b\geq 0}} \frac{(-3)^a}{(2a-1)(2b-1)} \binom{2a}{a}\binom{2b}{b}, \tag{5}$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT.

see also CATALAN NUMBER, KING WALK, SCHRÖDER NUMBER

References

Barcucci, E.; Pinzani, R.; and Sprugnoli, R. "The Motzkin Family." *Pure Math. Appl. Ser. A* **2**, 249–279, 1991.

Donaghey, R. "Restricted Plane Tree Representations of Four Motzkin-Catalan Equations." *J. Combin. Th. Ser. B* **22**, 114–121, 1977.

Donaghey, R. and Shapiro, L. W. "Motzkin Numbers." *J. Combin. Th. Ser. A* **23**, 291–301, 1977.

Kuznetsov, A.; Pak, I.; and Postnikov, A. "Trees Associated with the Motzkin Numbers." *J. Combin. Th. Ser. A* **76**, 145–147, 1996.

Motzkin, T. "Relations Between Hypersurface Cross Ratios, and a Combinatorial Formula for Partitions of a Polygon, for Permanent Preponderance, and for Nonassociative Products." *Bull. Amer. Math. Soc.* **54**, 352–360, 1948.

Sloane, N. J. A. Sequence A001006/M1184 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Moufang Plane

A PROJECTIVE PLANE in which every line is a translation line is called a Moufang plane.

References

Colbourn, C. J. and Dinitz, J. H. (Eds.) *CRC Handbook of Combinatorial Designs.* Boca Raton, FL: CRC Press, p. 710, 1996.

Mousetrap

A PERMUTATION problem invented by Cayley.

References

Guy, R. K. "Mousetrap." §E37 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 237–238, 1994.

Mouth

A PRINCIPAL VERTEX x_i of a SIMPLE POLYGON P is called a mouth if the diagonal $[x_{i-1}, x_{i+1}]$ is an extremal diagonal (i.e., the interior of $[x_{i-1}, x_{i+1}]$ lies in the exterior of P).

see also ANTHROPOMORPHIC POLYGON, EAR, ONE-MOUTH THEOREM

References

Toussaint, G. "Anthropomorphic Polygons." *Amer. Math. Monthly* **122**, 31–35, 1991.

Moving Average

Given a SEQUENCE $\{a_i\}_{i=1}^{N}$, an n-moving average is a new sequence $\{s_i\}_{i=1}^{N-n+1}$ defined from the a_i by taking the AVERAGE of subsequences of n terms,

$$s_i = \frac{1}{n}\sum_{j=i}^{i+n-1} a_j.$$

see also AVERAGE, SPENCER'S 15-POINT MOVING AVERAGE

References

Kenney, J. F. and Keeping, E. S. "Moving Averages." §14.2 in *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, pp. 221–223, 1962.

Moving Ladder Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

What is the longest ladder which can be moved around a right-angled hallway of unit width? For a straight, rigid ladder, the answer is $2\sqrt{2}$. For a smoothly-shaped ladder, the largest diameter is $\geq 2(1+\sqrt{2}\,)$ (Finch).

see also MOVING SOFA CONSTANT, PIANO MOVER'S PROBLEM

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/sofa/sofa.html`.

Moving Sofa Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

What is the sofa of greatest AREA S which can be moved around a right-angled hallway of unit width? Hammersley (Croft *et al.* 1994) showed that

$$S \geq \frac{\pi}{2} + \frac{2}{\pi} = 2.2074\ldots. \qquad (1)$$

Gerver (1992) found a sofa with larger AREA and provided arguments indicating that it is either optimal or close to it. The boundary of Gerver's sofa is a complicated shape composed of 18 ARCS. Its AREA can be given by defining the constants A, B, ϕ, and θ by solving

$$A(\cos\theta - \cos\phi) - 2B\sin\phi + (\theta - \phi - 1)\cos\theta - \sin\theta + \cos\phi + \sin\phi = 0 \qquad (2)$$

$$A(3\sin\theta + \sin\phi) - 2B\cos\phi + 3(\theta - \phi - 1)\sin\theta + 3\cos\theta - \sin\phi + \cos\phi = 0 \qquad (3)$$

$$A\cos\phi - (\sin\phi + \tfrac{1}{2} - \tfrac{1}{2}\cos\phi + B\sin\phi) = 0 \qquad (4)$$

$$(A + \tfrac{1}{2}\pi - \phi - \theta) - [B - \tfrac{1}{2}(\theta - \phi)(1 + A) - \tfrac{1}{4}(\theta - \phi)^2] = 0. \qquad (5)$$

This gives

$$A = 0.094426560843653\ldots \qquad (6)$$

$$B = 1.399203727333547\ldots \qquad (7)$$

$$\phi = 0.039177364790084\ldots \qquad (8)$$

$$\theta = 0.681301509382725\ldots. \qquad (9)$$

Now define

$$r(\alpha) \equiv \begin{cases} \frac{1}{2} & \text{for } 0 \leq \alpha < \phi \\ \frac{1}{2}(1 + A + \alpha - \phi) & \text{for } \phi \leq \alpha < \theta \\ A + \alpha - \phi & \text{for } \theta \leq \alpha < \frac{1}{2}\pi - \theta \\ B - \frac{1}{2}(\frac{1}{2}\pi - \alpha - \phi)(1 + A) - \frac{1}{4}(\frac{1}{2}\pi - \alpha - \phi)^2, & \text{for } \frac{1}{2}\pi - \theta \leq \alpha < \frac{1}{2}\pi - \phi, \end{cases} \qquad (10)$$

where

$$s(\alpha) \equiv 1 - r(\alpha) \qquad (11)$$

$$u(\alpha) \equiv \begin{cases} B - \frac{1}{2}(\alpha - \phi)(1 + A) - \frac{1}{4}(\alpha - \phi)^2 & \text{for } \phi \leq \alpha < \theta \\ A + \frac{1}{2}\pi - \phi - \alpha & \text{for } \theta \leq \alpha < \frac{1}{4}\pi \end{cases} \qquad (12)$$

$$D_u(\alpha) = \frac{du}{d\alpha} = \begin{cases} -\frac{1}{2}(1 + A) - \frac{1}{2}(\alpha - \phi) & \text{for } \phi \leq \alpha \leq \theta \\ -1 & \text{if } \theta \leq \alpha < \frac{1}{4}\pi. \end{cases} \qquad (13)$$

Finally, define the functions

$$y_1(\alpha) \equiv 1 - \int_0^\alpha r(t)\sin t\,dt \qquad (14)$$

$$y_2(\alpha) \equiv 1 - \int_0^\alpha s(t)\sin t\,dt \qquad (15)$$

$$y_3(\alpha) \equiv 1 - \int_0^\alpha s(t)\sin t\,dt - u(\alpha)\sin\alpha. \qquad (16)$$

The AREA of the optimal sofa is given by

$$A = 2\int_0^{\pi/2-\phi} y_1(\alpha)r(\alpha)\cos\alpha\,d\alpha + 2\int_0^\theta y_2(\alpha)s(\alpha)\cos\alpha\,d\alpha + 2\int_\phi^{\pi/4} y_3(\alpha)[u(\alpha)\sin\alpha - D_u(\alpha)\cos\alpha - s(\alpha)\cos\alpha]\,d\alpha = 2.21953166887197\ldots \qquad (17)$$

(Finch).

see also PIANO MOVER'S PROBLEM

References
Croft, H. T.; Falconer, K. J.; and Guy, R. K. *Unsolved Problems in Geometry.* New York: Springer-Verlag, 1994.
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/sofa/sofa.html`.
Gerver, J. L. "On Moving a Sofa Around a Corner." *Geometriae Dedicata* **42**, 267–283, 1992.
Stewart, I. *Another Fine Math You've Got Me Into....* New York: W. H. Freeman, 1992.

Mrs. Perkins' Quilt

The DISSECTION of a SQUARE of side n into a number S_n of smaller squares. Unlike a PERFECT SQUARE DISSECTION, however, the smaller SQUARES need not be all different sizes. In addition, only prime dissections are considered so that patterns which can be dissected on lower order SQUARES are not permitted. The following table gives the smallest number of coprime dissections of an $n \times n$ quilt (Sloane's A005670).

n	S_n
1	1
2	4
3	6
4	7
5	8
6–7	9
8–9	10
10–13	11
14–17	12
18–23	13
24–29	14
30-39	15
40	16
41	15
42–100	[17, 19]

see also PERFECT SQUARE DISSECTION

References
Conway, J. H. "Mrs. Perkins's Quilt." *Proc. Cambridge Phil. Soc.* **60**, 363–368, 1964.
Dudeney, H. E. Problem 173 in *Amusements in Mathematics.* New York: Dover, 1917.
Dudeney, H. E. Problem 177 in *536 Puzzles & Curious Problems.* New York: Scribner, 1967.
Gardner, M. "Mrs. Perkins' Quilt and Other Square-Packing Problems." Ch. 11 in *Mathematical Carnival: A New Round-Up of Tantalizers and Puzzles from Scientific American.* New York: Vintage, 1977.
Sloane, N. J. A. Sequence A005670/M3267 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Trustrum, G. B. "Mrs. Perkins's Quilt." *Proc. Cambridge Phil. Soc.* **61**, 7–11, 1965.

Mu Function

$$\mu(x,\beta) \equiv \int_0^\infty \frac{x^t t^\beta \, dt}{\Gamma(\beta+1)\Gamma(t+1)}$$

$$\mu(x,\beta,\alpha) \equiv \int_0^\infty \frac{x^{\alpha+t} t^\beta \, dt}{\Gamma(\beta+1)\Gamma(\alpha+t+1)},$$

where $\Gamma(z)$ is the GAMMA FUNCTION (Gradshteyn and Ryzhik 1980, p. 1079).

see also LAMBDA FUNCTION, NU FUNCTION

References
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, 1979.

μ Molecule

see MANDELBROT SET

Much Greater

A strong INEQUALITY in which a is not only GREATER than b, but *much* greater (by some convention), is denoted $a \gg b$. For an astronomer, "much" may mean by a factor of 100 (or even 10), while for a mathematician, it might mean by a factor of 10^4 (or even much more).

see also GREATER, MUCH LESS

Much Less

A strong INEQUALITY in which a is not only LESS than b, but *much* less (by some convention) is denoted $a \ll b$.

see also LESS, MUCH GREATER

Muirhead's Theorem

A NECESSARY and SUFFICIENT condition that $[\alpha']$ should be comparable with $[\alpha]$ for all POSITIVE values of the a is that one of (α') and (α) should be majorized by the other. If $(\alpha') \prec (\alpha)$, then

$$[\alpha'] \leq [\alpha],$$

with equality only when (α') and (α) are identical or when all the a are equal. See Hardy *et al.* (1988) for a definition of notation.

References
Hardy, G. H.; Littlewood, J. E.; and Pólya, G. *Inequalities, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 44–48, 1988.

Müller-Lyer Illusion

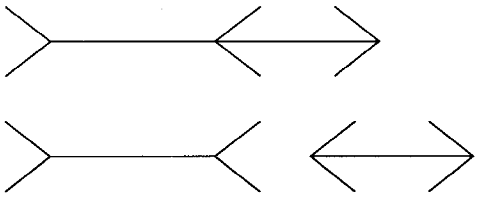

An optical ILLUSION in which the orientation of arrowheads makes one LINE SEGMENT look longer than another. In the above figure, the LINE SEGMENTS on the left and right are of equal length in both cases.

see also ILLUSION, POGGENDORFF ILLUSION, PONZO'S ILLUSION, VERTICAL-HORIZONTAL ILLUSION

References
Fineman, M. *The Nature of Visual Illusion.* New York: Dover, p. 153, 1996.
Luckiesh, M. *Visual Illusions: Their Causes, Characteristics & Applications.* New York: Dover, p. 93, 1965.

Muller's Method

Generalizes the Secant Method of root finding by using quadratic 3-point interpolation

$$q \equiv \frac{x_n - x_{n-1}}{x_{n-1} - x_{n-2}}. \tag{1}$$

Then define

$$A \equiv qP(x_n) - q(1+q)P(x_{n-1}) + q^2 P(x_{n-2}) \tag{2}$$

$$B \equiv (2q+1)P(x_n) - (1+q)^2 P(x_{n-1}) + q^2 P(x_{n-2}) \tag{3}$$

$$C \equiv (1+q)P(x_n), \tag{4}$$

and the next iteration is

$$x_{n+1} = x_n - (x_n - x_{n-1}) \frac{2C}{\max(B \pm \sqrt{B^2 - 4AC})}. \tag{5}$$

This method can also be used to find Complex zeros of Analytic Functions.

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 364, 1992.

Mulliken Symbols

Symbols used to identify irreducible representations of Groups:

$A =$ singly degenerate state which is symmetric with respect to Rotation about the principal C_n axis,

$B =$ singly Degenerate state which is antisymmetric with respect to Rotation about the principal C_n axis,

$E =$ doubly Degenerate,

$T =$ triply Degenerate,

$X_g =$ (gerade, symmetric) the sign of the wavefunction does not change on Inversion through the center of the atom,

$X_u =$ (ungerade, antisymmetric) the sign of the wavefunction changes on Inversion through the center of the atom,

$X_1 =$ (on a or b) the sign of the wavefunction does not change upon Rotation about the center of the atom,

$X_2 =$ (on a or b) the sign of the wavefunction changes upon Rotation about the center of the atom,

$' =$ symmetric with respect to a horizontal symmetry plane σ_h,

$'' =$ antisymmetric with respect to a horizontal symmetry plane σ_h.

see also Group Theory

Multiamicable Numbers

Two integers n and $m < n$ are (α, β)-multiamicable if

$$\sigma(m) - m = \alpha n$$

and

$$\sigma(n) - n = \beta m,$$

where $\sigma(n)$ is the Divisor Function and α, β are Positive integers. If $\alpha = \beta = 1$, (m, n) is an Amicable Pair.

m cannot have just one distinct prime factor, and if it has precisely two prime factors, then $\alpha = 1$ and m is Even. Small multiamicable numbers for small α, β are given by Cohen *et al.* (1995). Several of these numbers are reproduced in the below table.

α	β	m	n
1	6	76455288	183102192
1	7	52920	152280
1	7	16225560	40580280
1	7	90863136	227249568
1	7	16225560	40580280
1	7	70821324288	177124806144
1	7	199615613902848	499240550375424

see also Amicable Pair, Divisor Function

References

Cohen, G. L; Gretton, S.; and Hagis, P. Jr. "Multiamicable Numbers." *Math. Comput.* **64**, 1743–1753, 1995.

Multifactorial

A generalization of the Factorial and Double Factorial,

$$n! = n(n-1)(n-2)\cdots 2 \cdot 1$$

$$n!! = n(n-2)(n-4)\cdots$$

$$n!!! = n(n-3)(n-6)\cdots,$$

etc., where the product runs through positive integers. The Factorials $n!$ for $n = 1, 2, \ldots$, are 1, 2, 6, 24, 120, 720, ... (Sloane's A000142); the Double Factorials $n!!$ are 1, 2, 3, 8, 15, 48, 105, ... (Sloane's A006882); the triple factorials $n!!!$ are 1, 2, 3, 4, 10, 18, 28, 80, 162, 280, ... (Sloane's A007661); and the quadruple factorials $n!!!!$ are 1, 2, 3, 4, 5, 12, 21, 32, 45, 120, ... (Sloane's A007662).

see also Factorial, Gamma Function

References

Sloane, N. J. A. Sequences A000142/M1675, A006882/M0876, A007661/M0596, and A007662/M0534 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Multifractal Measure

A Measure for which the q-Dimension D_q varies with q.

References

Ott, E. *Chaos in Dynamical Systems.* New York: Cambridge University Press, 1993.

Multigrade Equation

A (k, l)-multigrade equation is a DIOPHANTINE EQUATION of the form

$$\sum_{i=1}^{l} n_i^j = \sum_{i=1}^{l} m_i^j$$

for $j = 1, \ldots, k$, where $\mathbf{m}$ and $\mathbf{n}$ are l-VECTORS. Multigrade identities remain valid if a constant is added to each element of $\mathbf{m}$ and $\mathbf{n}$ (Madachy 1979), so multigrades can always be put in a form where the minimum component of one of the vectors is 1.

Small-order examples are the (2, 3)-multigrade with $\mathbf{m} = \{1, 6, 8\}$ and $\mathbf{n} = \{2, 4, 9\}$:

$$\sum_{i=1}^{3} m_i^1 = \sum_{i=1}^{3} n_i^1 = 15$$

$$\sum_{i=1}^{3} m_i^2 = \sum_{i=1}^{3} n_i^2 = 101,$$

the (3, 4)-multigrade with $\mathbf{m} = \{1, 5, 8, 12\}$ and $\mathbf{n} = \{2, 3, 10, 11\}$:

$$\sum_{i=1}^{4} m_i^1 = \sum_{i=1}^{4} n_i^1 = 26$$

$$\sum_{i=1}^{4} m_i^2 = \sum_{i=1}^{4} n_i^2 = 234$$

$$\sum_{i=1}^{4} m_i^3 = \sum_{i=1}^{4} n_i^3 = 2366,$$

and the (4, 6)-multigrade with $\mathbf{m} = \{1, 5, 8, 12, 18, 19\}$ and $\mathbf{n} = \{2, 3, 9, 13, 16, 20\}$:

$$\sum_{i=1}^{6} m_i^1 = \sum_{i=1}^{6} n_i^1 = 63$$

$$\sum_{i=1}^{6} m_i^2 = \sum_{i=1}^{6} n_i^2 = 919$$

$$\sum_{i=1}^{6} m_i^3 = \sum_{i=1}^{6} n_i^3 = 15057$$

$$\sum_{i=1}^{6} m_i^4 = \sum_{i=1}^{6} n_i^4 = 260755$$

(Madachy 1979).

A spectacular example with $k = 9$ and $l = 10$ is given by $\mathbf{n} = \{\pm 12, \pm 11881, \pm 20231, \pm 20885, \pm 23738\}$ and $\mathbf{m} = \{\pm 436, \pm 11857, \pm 20449, \pm 20667, \pm 23750\}$ (Guy 1994), which has sums

$$\sum_{i=1}^{9} m_i^1 = \sum_{i=1}^{9} n_i^1 = 0$$

$$\sum_{i=1}^{9} m_i^2 = \sum_{i=1}^{9} n_i^2 = 3100255070$$

$$\sum_{i=1}^{9} m_i^3 = \sum_{i=1}^{9} n_i^3 = 0$$

$$\sum_{i=1}^{9} m_i^4 = \sum_{i=1}^{9} n_i^4 = 1390452894778220678$$

$$\sum_{i=1}^{9} m_i^5 = \sum_{i=1}^{9} n_i^5 = 0$$

$$\sum_{i=1}^{9} m_i^6 = \sum_{i=1}^{9} n_i^6 = 666573454337853049941719510$$

$$\sum_{i=1}^{9} m_i^7 = \sum_{i=1}^{9} n_i^7 = 0$$

$$\sum_{i=1}^{9} m_i^8 = \sum_{i=1}^{9} n_i^8$$

$$= 330958142560259813821203262692838598$$

$$\sum_{i=1}^{9} m_i^9 = \sum_{i=1}^{9} n_i^9 = 0.$$

see also DIOPHANTINE EQUATION

References

Chen, S. "Equal Sums of Like Powers: On the Integer Solution of the Diophantine System." `http://www.nease.net/~chin/eslp/`

Gloden, A. *Mehrgeradige Gleichungen.* Groningen, Netherlands: Noordhoff, 1944.

Gloden, A. "Sur la multigrade A_1, A_2, A_3, A_4, $A_5 =^k B_1$, B_2, B_3, B_4, B_5 (k = 1, 3, 5, 7)." *Revista Euclides* **8**, 383–384, 1948.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 143, 1994.

Kraitchik, M. "Multigrade." §3.10 in *Mathematical Recreations.* New York: W. W. Norton, p. 79, 1942.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 171–173, 1979.

Multilinear

A function, form, etc., in two or more variables is said to be multilinear if it is linear in each variable separately.

see also BILINEAR, LINEAR OPERATOR

Multimagic Series

n numbers form a p-multimagic series if the sum of their kth powers is the MAGIC CONSTANT of degree k for every $k = 1, \ldots, p$. The following table gives the number of p-multimagic series N_p of given orders n (Kraitchik 1942).

n	N_1	N_2	N_3
2	2		
3	8		
4	86	2	2
5	1,394	8	2
6	0	98	0
7	0	1,844	0
8	0	38,039	115
9	0	0	41
10	0	0	0
11	0	0	961

References

Kraitchik, M. "Multimagic Squares." §7.10 in *Mathematical Recreations.* New York: W. W. Norton, pp. 176–178, 1942.

Multimagic Square

A MAGIC SQUARE is p-multimagic if the square formed by replacing each element by its kth power for $k = 1, 2, \ldots, p$ is also magic. A 2-multimagic square is called a BIMAGIC SQUARE, and a 3-multimagic square is called a TRIMAGIC SQUARE.

see also BIMAGIC SQUARE, MAGIC SQUARE, TRIMAGIC SQUARE

References

Kraitchik, M. "Multimagic Squares." §7.10 in *Mathematical Recreations.* New York: W. W. Norton, pp. 176–178, 1942.

Multinomial Coefficient

The multinomial COEFFICIENTS

$$(x_1, x_2, \ldots) = \frac{x_1 + x_2 + \ldots}{x_1! x_2! \cdots}$$

are the terms in the MULTINOMIAL SERIES expansion. They satisfy

$$\begin{aligned}(x_1, x_2, x_3, \ldots) &= (x_1 + x_2, x_3, \ldots)(x_1, x_2)\\ &= (x_1 + x_2 + x_3, \ldots)(x_1, x_2, x_3) = \ldots\end{aligned}$$

(Beeler *et al.* 1972, Item 44).

see also BINOMIAL COEFFICIENT, MULTINOMIAL SERIES

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Multinomial Coefficients." §24.1.2 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 823–824, 1972.

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Spiegel, M. R. *Theory and Problems of Probability and Statistics.* New York: McGraw-Hill, p. 113, 1992.

Multinomial Distribution

Let a set of random variates $X_1, X_2, \ldots, X_n$ have a probability function

$$P(X_1 = x_1, \ldots, X_n = x_n) = \frac{N!}{\prod_{i=1}^n x_i!} \prod_{i=1}^n {\theta_i}^{x_i} \tag{1}$$

where x_i are POSITIVE INTEGERS, $\theta_i > 0$, and

$$\sum_{i=1}^n \theta_i = 1 \tag{2}$$

$$\sum_{i=1}^n x_i = N. \tag{3}$$

Then the joint distribution of $X_1, \ldots, X_n$ is a multinomial distribution and $P(X_1 = x_1, \ldots, X_n = x_n)$ is given by the corresponding coefficient of the MULTINOMIAL SERIES

$$(\theta_1 + \theta_2 + \ldots + \theta_n)^N. \tag{4}$$

The MEAN and VARIANCE of X_i are

$$\mu_i = N\theta_i \tag{5}$$

$${\sigma_i}^2 = N\theta_i(1 - \theta_i). \tag{6}$$

The COVARIANCE of X_i and X_j is

$${\sigma_{ij}}^2 = -N\theta_i\theta_j. \tag{7}$$

see also BINOMIAL DISTRIBUTION

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 532, 1987.

Multinomial Series

A generalization of the BINOMIAL SERIES discovered by Johann Bernoulli and Leibniz.

$$(a_1 + a_2 + \ldots + a_k)^n = \sum_{n_1, n_2, \ldots, n_k}^{n} \frac{n!}{n_1! n_2! \cdots n_k!} {a_1}^{n_1} {a_2}^{n_2} \cdots {a_k}^{n_k},$$

where $n \equiv n_1 + n_2 + \ldots + n_k$. The multinomial series arises in a generalization of the BINOMIAL DISTRIBUTION called the MULTINOMIAL DISTRIBUTION.

see also BINOMIAL SERIES, MULTINOMIAL DISTRIBUTION

Multinomial Theorem

see MULTINOMIAL SERIES

Multiperfect Number

A number n is k-multiperfect (also called a k-MULTIPLY PERFECT NUMBER or k-PLUPERFECT NUMBER) if

$$\sigma(n) = kn$$

for some INTEGER $k > 2$, where $\sigma(n)$ is the DIVISOR FUNCTION. The value of k is called the CLASS. The special case $k = 2$ corresponds to PERFECT NUMBERS P_2, which are intimately connected with MERSENNE PRIMES (Sloane's A000396). The number 120 was long known to be 3-multiply perfect (P_3) since

$$\sigma(120) = 3 \cdot 120.$$

The following table gives the first few P_n for $n = 2, 3, \ldots, 6$.

n	Sloane	P_n
2	000396	6, 28, 496, 8128, ...,
3	005820	120, 672, 523776, 459818240, ...
4	027687	30240, 32760, 2178540, 23569920, ...
5	046060	14182439040, 31998395520, ...
6	046061	154345556085770649600, ...

In 1900–1901, Lehmer proved that P_3 has at least three distinct PRIME factors, P_4 has at least four, P_5 at least six, P_6 at least nine, and P_7 at least 14.

As of of 1911, 251 pluperfect numbers were known (Carmichael and Mason 1911). As of 1929, 334 pluperfect numbers were known, many of them found by Poulet. Franqui and García (1953) found 63 additional ones (five P_5s, 29 P_6s, and 29 P_7s), several of which were known to Poulet but had not been published, bringing the total to 397. Brown (1954) discovered 110 pluperfects, including 31 discovered but not published by Poulet and 25 previously published by Franqui and García (1953), for a total of 482. Franqui and García (1954) subsequently discovered 57 additional pluperfects (3 P_6s, 52 P_7s, and 2 P_8s), increasing the total known to 539.

An outdated database is maintained by R. Schroeppel, who lists 2,094 multiperfects, and an up-to-date list by J. L. Moxham (1998). It is believed that all multiperfect numbers of index 3, 4, 5, 6, and 7 are known. The number of known n-multiperfect numbers are 1, 37, 6, 36, 65, 245, 516, 1101, 1129, 46, 0, 0, ...

If n is a P_5 number such that $3 \nmid n$, then $3n$ is a P_4 number. If $3n$ is a P_{4k} number such that $3 \nmid n$, then n is a P_{3k} number. If n is a P_3 number such that 3 (but not 5 and 9) DIVIDES n, then $45n$ is a P_4 number.

see also e-MULTIPERFECT NUMBER, FRIENDLY PAIR, HYPERPERFECT NUMBER, INFINARY MULTIPERFECT NUMBER, MERSENNE PRIME, PERFECT NUMBER, UNITARY MULTIPERFECT NUMBER

References

Brown, A. L. "Multiperfect Numbers." *Scripta Math.* **20**, 103–106, 1954.

Dickson, L. E. *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, pp. 33–38, 1952.

Flammenkamp, A. "Multiply Perfect Numbers." `http://www.uni-bielefeld.de/~achim/mpn.html`.

Franqui, B. and García, M. "Some New Multiply Perfect Numbers." *Amer. Math. Monthly* **60**, 459–462, 1953.

Franqui, B. and García, M. "57 New Multiply Perfect Numbers." *Scripta Math.* **20**, 169–171, 1954.

Guy, R. K. "Almost Perfect, Quasi-Perfect, Pseudoperfect, Harmonic, Weird, Multiperfect and Hyperperfect Numbers." §B2 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 45–53, 1994.

Helenius, F. W. "Multiperfect Numbers (MPFNs)." `http://www.netcom.com/~fredh/mpfn`.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 149–151, 1979.

Moxham, J. L. "13 New MPFN's." math-fun@cs.arizona.edu posting, Aug 13, 1998.

Poulet, P. *La Chasse aux nombres,* Vol. 1. Brussels, pp. 9–27, 1929.

Schroeppel, R. "Multiperfect Numbers–Multiply Perfect Numbers–Pluperfect Numbers–MPFNs." Rev. Dec. 13, 1995. `ftp://ftp.cs.arizona.edu/xkernel/rcs/mpfn.html`.

Schroeppel, R. (moderator). mpfn mailing list. e-mail `rcs@cs.arizona.edu` to subscribe.

Sloane, N. J. A. Sequences A000396/M4186 and A005820/M5376 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Multiple Integral

A repeated integral over $n > 1$ variables

$$\underbrace{\int \cdots \int}_{n} f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n$$

is called a multiple integral. An nth order integral corresponds, in general, to an n-D VOLUME (CONTENT), with $n = 2$ corresponding to an AREA. In an indefinite multiple integral, the order in which the integrals are carried out can be varied at will; for definite multiple integrals, care must be taken to correctly transform the limits if the order is changed.

see also INTEGRAL, MONTE CARLO INTEGRATION

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Multidimensional Integrals." §4.6 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 155–158, 1992.

Multiple Regression

A REGRESSION giving conditional expectation values of a given variable in terms of two or more other variables.

see also LEAST SQUARES FITTING, MULTIVARIATE ANALYSIS, NONLINEAR LEAST SQUARES FITTING

References

Edwards, A. L. *Multiple Regression and the Analysis of Variance and Covariance.* San Francisco, CA: W. H. Freeman, 1979.

Multiplication

In simple algebra, multiplication is the process of calculating the result when a number a is taken b times. The result of a multiplication is called the Product of a and b. It is denoted $a \times b$, $a \cdot b$, $(a)(b)$, or simply ab. The symbol $\times$ is known as the Multiplication Sign. Normal multiplication is Associative, Commutative, and Distributive.

More generally, multiplication can also be defined for other mathematical objects such as Groups, Matrices, Sets, and Tensors.

Karatsuba and Ofman (1962) discovered that multiplication of two n digit numbers can be done with a Bit Complexity of less than n^2 using an algorithm now known as Karatsuba Multiplication.

see also Addition, Bit Complexity, Complex Multiplication, Division, Karatsuba Multiplication, Matrix Multiplication, Product, Russian Multiplication, Subtraction, Times

References

Karatsuba, A. and Ofman, Yu. "Multiplication of Many-Digital Numbers by Automatic Computers." *Doklady Akad. Nauk SSSR* **145**, 293–294, 1962. Translation in *Physics–Doklady* **7**, 595–596, 1963.

Multiplication Magic Square

128	1	32
4	16	64
8	256	2

A square which is magic under multiplication instead of addition (the operation used to define a conventional Magic Square) is called a multiplication magic square. Unlike (normal) Magic Squares, the n^2 entries for an nth order multiplicative magic square are not required to be consecutive. The above multiplication magic square has a multiplicative magic constant of 4,096.

see also Addition-Multiplication Magic Square, Magic Square

References

Hunter, J. A. H. and Madachy, J. S. "Mystic Arrays." Ch. 3 in *Mathematical Diversions.* New York: Dover, pp. 30–31, 1975.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 89–91, 1979.

Multiplication Principle

If one event can occur in m ways and a second can occur independently of the first in n ways, then the two events can occur in mn ways.

Multiplication Sign

The symbol $\times$ used to denote Multiplication, i.e., $a \times b$ denotes a times b.

Multiplication Table

A multiplication table is an array showing the result of applying a Binary Operator to elements of a given set S.

×	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4	5	6	7	8	9	10
2	2	4	6	8	10	12	14	16	18	20
3	3	6	9	12	15	18	21	24	27	30
4	4	8	12	16	20	24	28	32	36	40
5	5	10	15	20	25	30	35	40	45	50
6	6	12	18	24	30	36	42	48	54	60
7	7	14	21	28	35	42	49	56	63	70
8	8	16	24	32	40	48	56	64	72	80
9	9	18	27	36	45	54	63	72	81	90
10	10	20	30	40	50	60	70	80	90	100

see also Binary Operator, Truth Table

Multiplicative Character

see Character (Multiplicative)

Multiplicative Digital Root

Consider the process of taking a number, multiplying its Digits, then multiplying the Digits of numbers derived from it, etc., until the remaining number has only one Digit. The number of multiplications required to obtain a single Digit from a number n is called the Multiplicative Persistence of n, and the Digit obtained is called the multiplicative digital root of n.

For example, the sequence obtained from the starting number 9876 is (9876, 3024, 0), so 9876 has a Multiplicative Persistence of two and a multiplicative digital root of 0. The multiplicative digital roots of the first few positive integers are 1, 2, 3, 4, 5, 6, 7, 8, 9, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 0, 2, 4, 6, 8, 0, 2, 4, 6, 8, 0, 3, 6, 9, 2, 5, 8, 2, ... (Sloane's A031347).

see also Additive Persistence, Digitadition, Digital Root, Multiplicative Persistence

References

Sloane, N. J. A. Sequence A031347 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Multiplicative Function

A function $f(m)$ is called multiplicative if $(m, m') = 1$ (i.e., the statement that m and m' are Relatively Prime) implies

$$f(mm') = f(m)f(m').$$

see also Quadratic Residue, Totient Function

Multiplicative Inverse

The multiplicative of a Real or Complex Number z is its Reciprocal $1/z$. For complex $z = x + iy$,

$$\frac{1}{z} = \frac{1}{x+iy} = \frac{x}{x^2-y^2} - i\frac{y}{x^2-y^2}.$$

Multiplicative Perfect Number

A number n for which the PRODUCT of DIVISORS is equal to n^2. The first few are 1, 6, 8, 10, 14, 15, 21, 22, ... (Sloane's A007422).

see also PERFECT NUMBER

References

Sloane, N. J. A. Sequence A007422/M4068 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Multiplicative Persistence

Multiply all the digits of a number n by each other, repeating with the product until a single DIGIT is obtained. The number of steps required is known as the multiplicative persistence, and the final DIGIT obtained is called the MULTIPLICATIVE DIGITAL ROOT of n.

For example, the sequence obtained from the starting number 9876 is (9876, 3024, 0), so 9876 has an multiplicative persistence of two and a MULTIPLICATIVE DIGITAL ROOT of 0. The multiplicative persistences of the first few positive integers are 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 1, 1, ... (Sloane's A031346). The smallest numbers having multiplicative persistences of 1, 2, ... are 10, 25, 39, 77 679 6788 68889 267889 26888999 3778888999 277777788888899 (Sloane's A003001). There is no number $< 10^{50}$ with multiplicative persistence > 11.

The multiplicative persistence of an n-DIGIT number is also called its LENGTH. The maximum lengths for $n =$ 2-, 3-, ..., digit numbers are 4, 5, 6, 7, 7, 8, 9, 9, 10, 10, 10, ... (Sloane's A014553; (Beeler *et al.* 1972, Item 56; Gottlieb 1969–1970).

The concept of multiplicative persistence can be generalized to multiplying the kth powers of the digits of a number and iterating until the result remains constant. All numbers other than REPUNITS, which converge to 1, converge to 0. The number of iterations required for the kth powers of a number's digits to converge to 0 is called its k-multiplicative persistence. The following table gives the n-multiplicative persistences for the first few positive integers.

n	Sloane	n-Persistences
2	031348	0, 7, 6, 6, 3, 5, 5, 4, 5, 1, ...
3	031349	0, 4, 5, 4, 3, 4, 4, 3, 3, 1, ...
4	031350	0, 4, 3, 3, 3, 3, 2, 2, 3, 1, ...
5	031351	0, 4, 4, 2, 3, 3, 2, 3, 2, 1, ...
6	031352	0, 3, 3, 2, 3, 3, 3, 3, 3, 1, ...
7	031353	0, 4, 3, 3, 3, 3, 3, 2, 3, 1, ...
8	031354	0, 3, 3, 3, 2, 4, 2, 3, 2, 1, ...
9	031355	0, 3, 3, 3, 3, 2, 2, 3, 2, 1, ...
10	031356	0, 2, 2, 2, 3, 2, 3, 2, 2, 1, ...

see also 196-ALGORITHM, ADDITIVE PERSISTENCE, DIGITADITION, DIGITAL ROOT, KAPREKAR NUMBER, LENGTH (NUMBER), MULTIPLICATIVE DIGITAL ROOT, NARCISSISTIC NUMBER, RECURRING DIGITAL INVARIANT

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Gottlieb, A. J. Problems 28–29 in "Bridge, Group Theory, and a Jigsaw Puzzle." *Techn. Rev.* **72**, unpaginated, Dec. 1969.

Gottlieb, A. J. Problem 29 in "Integral Solutions, Ladders, and Pentagons." *Techn. Rev.* **72**, unpaginated, Apr. 1970.

Sloane, N. J. A. "The Persistence of a Number." *J. Recr. Math.* **6**, 97–98, 1973.

Sloane, N. J. A. Sequences A014553 and A003001/M4687 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Multiplicative Primitive Residue Class Group

see MODULO MULTIPLICATION GROUP

Multiplicity

The word multiplicity is a general term meaning "the number of values for which a given condition holds." The most common use of the word is as the value of the TOTIENT VALENCE FUNCTION.

see also DEGENERATE, NOETHER'S FUNDAMENTAL THEOREM, TOTIENT VALENCE FUNCTION

Multiplier

see MODULAR FUNCTION MULTIPLIER

Multiply Connected

A set which is CONNECTED but not SIMPLY CONNECTED is called multiply connected. A SPACE is n-MULTIPLY CONNECTED if it is $(n-1)$-connected and if every MAP from the n-SPHERE into it extends continuously over the $(n+1)$-DISK

A theorem of Whitehead says that a SPACE is infinitely connected IFF it is contractible.

see also CONNECTIVITY, LOCALLY PATHWISE-CONNECTED SPACE, PATHWISE-CONNECTED, SIMPLY CONNECTED

Multiply Perfect Number

see MULTIPERFECT NUMBER

Multisection

see SERIES MULTISECTION

Multivalued Function

A FUNCTION which assumes two or more distinct values at one or more points in its DOMAIN.

see also BRANCH CUT, BRANCH POINT

References

Morse, P. M. and Feshbach, H. "Multivalued Functions." §4.4 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 398–408, 1953.

Multivariate Analysis

The study of random distributions involving more than one variable.

see also GAUSSIAN JOINT VARIABLE THEOREM, MULTIPLE REGRESSION, MULTIVARIATE FUNCTION

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 927–928, 1972.
Feinstein, A. R. *Multivariable Analysis.* New Haven, CT: Yale University Press, 1996.
Hair, J. F. Jr. *Multivariate Data Analysis with Readings, 4th ed.* Englewood Cliffs, NJ: Prentice-Hall, 1995.
Sharma, S. *Applied Multivariate Techniques.* New York: Wiley, 1996.

Multivariate Function

A FUNCTION of more than one variable.

see also MULTIVARIATE ANALYSIS, UNIVARIATE FUNCTION

Multivariate Theorem

see GAUSSIAN JOINT VARIABLE THEOREM

Müntz Space

A Müntz space is a technically defined SPACE

$$M(\Lambda) \equiv \operatorname{span}\{x^{\lambda_0}, x^{\lambda_1}, \ldots\}$$

which arises in the study of function approximations.

Müntz's Theorem

Müntz's theorem is a generalization of the WEIERSTRAß APPROXIMATION THEOREM, which states that any continuous function on a closed and bounded interval can be uniformly approximated by POLYNOMIALS involving constants and any INFINITE SEQUENCE of POWERS whose RECIPROCALS diverge.

In technical language, Müntz's theorem states that the MÜNTZ SPACE $M(\Lambda)$ is dense in $C[0,1]$ IFF

$$\sum_{i=1}^{\infty} \frac{1}{\lambda_i} = \infty.$$

see also WEIERSTRAß APPROXIMATION THEOREM

Mutant Knot

Given an original KNOT K, the three knots produced by MUTATION together with K itself are called mutant knots. Mutant knots are often difficult to distinguish. For instant, mutants have the same HOMFLY POLYNOMIALS and HYPERBOLIC KNOT volume. Many but not all mutants also have the same GENUS (KNOT).

Mutation

Consider a KNOT as being formed from two TANGLES. The following three operations are called mutations.

1. Cut the knot open along four points on each of the four strings coming out of T_2, flipping T_2 over, and gluing the strings back together.
2. Cut the knot open along four points on each of the four strings coming out of T_2, flipping T_2 to the right, and gluing the strings back together.
3. Cut the knot, rotate it by 180°, and reglue. This is equivalent to performing (1), then (2).

Mutations applied to an alternating KNOT projection always yield an ALTERNATING KNOT. The mutation of a KNOT is always another KNOT (a opposed to a LINK).

References
Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, p. 49, 1994.

Mutual Energy

Let Ω be a SPACE with MEASURE $\mu \geq 0$, and let $\Phi(P,Q)$ be a real function on the PRODUCT SPACE $\Omega \times \Omega$. When

$$\begin{aligned}(\mu, nu) &= \iint \Phi(P,Q)\,d\mu(Q)\,d\nu(P)\\ &= \int \Phi(P,\mu)\,d\nu(P)\end{aligned}$$

exists for measures $\mu, \nu \geq 0$, (μ,ν) is called the mutual energy. (μ,μ) is then called the ENERGY.

see also ENERGY

References
Iyanaga, S. and Kawada, Y. (Eds.). "General Potential." §335.B in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1038, 1980.

Mutually Exclusive

Two events E_1 and E_2 are mutually exclusive if $E_1 \cap E_2 \equiv \varnothing$. n events $E_1, E_2, \ldots, E_n$ are mutually exclusive if $E_i \cap E_j \equiv \varnothing$ for $i \neq j$.

Mutually Singular

Let M be a SIGMA ALGEBRA M, and let λ_1 and λ_2 be MEASURES on M. If there EXISTS a pair of disjoint SETS A and B such that λ_1 is CONCENTRATED on A and λ_2 is CONCENTRATED on B, then λ_1 and λ_2 are said to be mutually singular, written $\lambda_1 \perp \lambda_2$.

see also ABSOLUTELY CONTINUOUS, CONCENTRATED, SIGMA ALGEBRA

References
Rudin, W. *Functional Analysis.* New York: McGraw-Hill, p. 121, 1991.

Myriad

The Greek word for 10,000.

Myriagon
A 10,000-sided POLYGON.

Mystic Pentagram
see PENTAGRAM

N

$\mathbb{N}$

The SET of NATURAL NUMBERS (the POSITIVE INTEGERS $\mathbb{Z}^+$ 1, 2, 3, ...; Sloane's A000027), denoted $\mathbb{N}$, also called the WHOLE NUMBERS. Like whole numbers, there is no general agreement on whether 0 should be included in the list of natural numbers.

Due to lack of standard terminology, the following terms are recommended in preference to "COUNTING NUMBER," "natural number," and "WHOLE NUMBER."

Set	Name	Symbol
$\ldots, -2, -1, 0, 1, 2, \ldots$	integers	$\mathbb{Z}$
$1, 2, 3, 4, \ldots$	positive integers	$\mathbb{Z}^+$
$0, 1, 2, 3, 4 \ldots$	nonnegative integers	$\mathbb{Z}^*$
$-1, -2, -3, -4, \ldots$	negative integers	$\mathbb{Z}^-$

see also $\mathbb{C}$, CARDINAL NUMBER, COUNTING NUMBER, $\mathbb{I}$, INTEGER, $\mathbb{Q}$, $\mathbb{R}$, WHOLE NUMBER, $\mathbb{Z}$, $\mathbb{Z}^+$

References
Sloane, N. J. A. Sequence A000027/M0472 in "An On-Line Version of the Encyclopedia of Integer Sequences."

N-Cluster

A LATTICE POINT configuration with no three points COLLINEAR and no four CONCYCLIC. An example is the 6-cluster (0, 0), (132, −720), (546, −272), (960, −720), (1155, 540), (546, 1120). Call the RADIUS of the smallest CIRCLE centered at one of the points of an N-cluster which contains all the points in the N-cluster the EXTENT. Noll and Bell (1989) found 91 nonequivalent prime 6-clusters of EXTENT less than 20937, but found no 7-clusters.

References
Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 187, 1994.
Noll, L. C. and Bell, D. I. "n-clusters for $1 < n < 7$." *Math. Comput.* **53**, 439–444, 1989.

n-Cube

see HYPERCUBE, POLYCUBE

n-in-a-Row

see TIC-TAC-TOE

n-minex

n-minex is defined as 10^{-n}.

see also n-PLEX

References
Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 16, 1996.

n-Omino

see POLYOMINO

n-plex

n-plex is defined as 10^n.

see also GOOGOLPLEX, n-MINEX

References
Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 16, 1996.

n-Sphere

see HYPERSPHERE

Nabla

see DEL, LAPLACIAN

Nagel Point

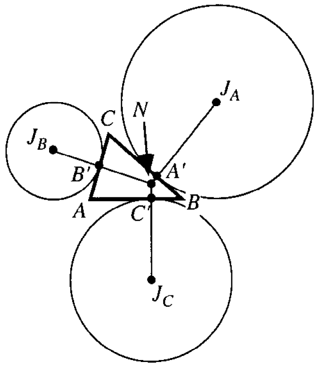

Let A' be the point at which the A-EXCIRCLE meets the side BC of a TRIANGLE ΔABC, and define B' and C' similarly. Then the lines AA', BB', and CC' CONCUR in the NAGEL POINT.

The Nagel point can also be constructed by letting A'' be the point half way around the PERIMETER of ΔABC starting at A, and B'' and C'' similarly defined. Then the lines AA'', BB'', and CC'' concur in the Nagel point. It is therefore sometimes known as the BISECTED PERIMETER POINT (Bennett *et al.* 1988, Chen *et al.* 1992, Kimberling 1994).

The Nagel point has TRIANGLE CENTER FUNCTION

$$\alpha = \frac{b+c-a}{a}.$$

It is the ISOTOMIC CONJUGATE POINT of the GERGONNE POINT.

see also EXCENTER, EXCENTRAL TRIANGLE, EXCIRCLE, MITTENPUNKT, TRISECTED PERIMETER POINT

References
Altshiller-Court, N. *College Geometry: A Second Course in Plane Geometry for Colleges and Normal Schools, 2nd ed.* New York: Barnes and Noble, pp. 160–164, 1952.
Bennett, G.; Glenn, J.; Kimberling, C.; and Cohen, J. M. "Problem E 3155 and Solution." *Amer. Math. Monthly* **95**, 874, 1988.
Chen, J.; Lo, C.-H.; and Lossers, O. P. "Problem E 3397 and Solution." *Amer. Math. Monthly* **99**, 70–71, 1992.

Eves, H. W. *A Survey of Geometry, rev. ed.* Boston, MA: Allyn and Bacon, p. 83, 1972.
Gallatly, W. *The Modern Geometry of the Triangle, 2nd ed.* London: Hodgson, p. 20, 1913.
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 184 and 225–226, 1929.
Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.
Kimberling, C. "Nagel Point." `http://www.evansville.edu/~ck6/tcenters/class/nagel.html`.

Naive Set Theory

A branch of mathematics which attempts to formalize the nature of the SET using a minimal collection of independent axioms. Unfortunately, as discovered by its earliest proponents, naive set theory quickly runs into a number of PARADOXES (such as RUSSELL'S PARADOX), so a less sweeping and more formal theory known as AXIOMATIC SET THEORY must be used.

see also AXIOMATIC SET THEORY, RUSSELL'S PARADOX, SET THEORY

Napier's Analogies

Let a SPHERICAL TRIANGLE have sides a, b, and c with A, B, and C the corresponding opposite angles. Then

$$\frac{\sin[\frac{1}{2}(A-B)]}{\sin[\frac{1}{2}(A+B)]}=\frac{\tan[\frac{1}{2}(a-b)]}{\tan(\frac{1}{2}c)} \qquad (1)$$

$$\frac{\cos[\frac{1}{2}(A-B)]}{\cos[\frac{1}{2}(A+B)]}=\frac{\tan[\frac{1}{2}(a+b)]}{\tan(\frac{1}{2}c)} \qquad (2)$$

$$\frac{\sin[\frac{1}{2}(a-b)]}{\sin[\frac{1}{2}(a+b)]}=\frac{\tan[\frac{1}{2}(A-B)]}{\cot(\frac{1}{2}C)} \qquad (3)$$

$$\frac{\cos[\frac{1}{2}(a-b)]}{\cos[\frac{1}{2}(a+b)]}=\frac{\tan[\frac{1}{2}(A+B)]}{\cot(\frac{1}{2}C)}. \qquad (4)$$

see also SPHERICAL TRIGONOMETRY

Napier's Bones

Numbered rods which can be used to perform MULTIPLICATION. This process is also called RABDOLOGY.

see also GENAILLE RODS

References

Gardner, M. "Napier's Bones." Ch. 7 in *Knotted Doughnuts and Other Mathematical Entertainments.* New York: W. H. Freeman, 1986.
Pappas, T. "Napier's Bones." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 64–65, 1989.

Napier's Constant

see e

Napier's Inequality

For $b > a > 0$,

$$\frac{1}{b}<\frac{\ln b-\ln a}{b-a}<\frac{1}{a}.$$

References

Nelsen, R. B. "Napier's Inequality (Two Proofs)." *College Math. J.* **24**, 165, 1993.

Napierian Logarithm

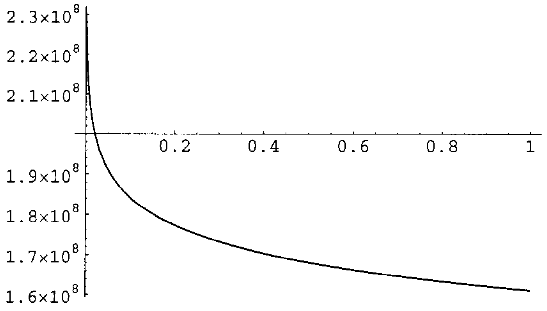

Write a number N as

$$N=10^7(1-10^{-7})^L,$$

then L is the Napierian logarithm of N. This was the original definition of a LOGARITHM, and can be given in terms of the modern LOGARITHM as

$$L(N)=-\frac{\log\left(\frac{n}{10^7}\right)}{\log\left(\frac{10^7}{10^7-1}\right)}.$$

The Napierian logarithm decreases with increasing numbers and does not satisfy many of the fundamental properties of the modern LOGARITHM, e.g.,

$$\text{Nlog}(xy)\neq\text{Nlog}\,x+\text{Nlog}\,y.$$

Napkin Ring

see SPHERICAL RING

Napoleon Points

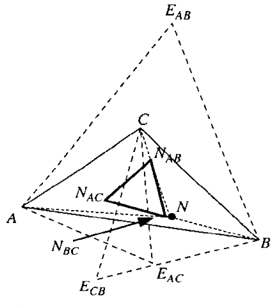

The inner Napoleon point N is the CONCURRENCE of lines drawn between VERTICES of a given TRIANGLE

ΔABC and the opposite VERTICES of the corresponding inner NAPOLEON TRIANGLE $\Delta N_{AB}N_{AC}N_{BC}$. The TRIANGLE CENTER FUNCTION of the inner Napoleon point is

$$\alpha = \csc(A - \tfrac{1}{6}\pi).$$

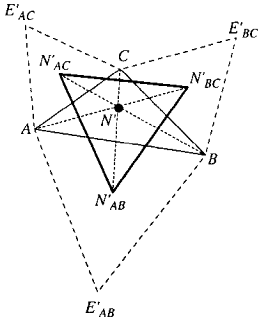

The outer Napoleon point N' is the CONCURRENCE of lines drawn between VERTICES of a given TRIANGLE ΔABC and the opposite VERTICES of the corresponding outer NAPOLEON TRIANGLE $\Delta N'_{AB}N'_{AC}N'_{BC}$. The TRIANGLE CENTER FUNCTION of the point is

$$\alpha = \csc(A + \tfrac{1}{6}\pi).$$

see also NAPOLEON'S THEOREM, NAPOLEON TRIANGLES

References
Casey, J. *Analytic Geometry, 2nd ed.* Dublin: Hodges, Figgis, & Co., pp. 442–444, 1893.
Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Napoleon's Problem

Given the center of a CIRCLE, divide the CIRCLE into four equal arcs using a COMPASS alone (a MASCHERONI CONSTRUCTION).

see also CIRCLE, COMPASS, MASCHERONI CONSTRUCTION

Napoleon's Theorem

If EQUILATERAL TRIANGLES are erected externally on the sides of any TRIANGLE, then the centers form an EQUILATERAL TRIANGLE (the outer NAPOLEON TRIANGLE). Furthermore, the inner NAPOLEON TRIANGLE is also EQUILATERAL and the difference between the areas of the outer and inner Napoleon triangles equals the AREA of the original TRIANGLE.

see also NAPOLEON POINTS, NAPOLEON TRIANGLES

References
Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 60–65, 1967.
Pappas, T. "Napoleon's Theorem." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 57, 1989.
Schmidt, F. "200 Jahre französische Revolution—Problem und Satz von Napoleon." *Didaktik der Mathematik* **19**, 15–29, 1990.
Wentzel, J. E. "Converses of Napoleon's Theorem." *Amer. Math. Monthly* **99**, 339–351, 1992.

Napoleon Triangles

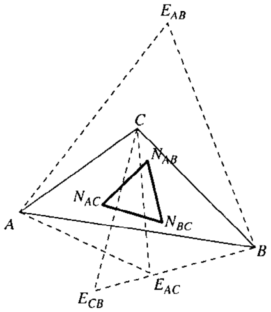

The inner Napoleon triangle is the TRIANGLE $\Delta N_{AB}N_{AC}N_{BC}$ formed by the centers of internally erected EQUILATERAL TRIANGLES ΔABE_{AB}, ΔACE_{AC}, and ΔBCE_{BC} on the sides of a given TRIANGLE ΔABC. It is an EQUILATERAL TRIANGLE.

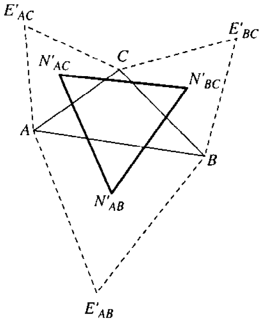

The outer Napoleon triangle is the TRIANGLE $\Delta N'_{AB}N'_{AC}N'_{BC}$ formed by the centers of externally erected EQUILATERAL TRIANGLES $\Delta ABE'_{AB}$, $\Delta ACE'_{AC}$, and $\Delta BCE'_{BC}$ on the sides of a given TRIANGLE ΔABC. It is also an EQUILATERAL TRIANGLE.

see also EQUILATERAL TRIANGLE, NAPOLEON POINTS, NAPOLEON'S THEOREM

References
Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 60–65, 1967.

Nappe

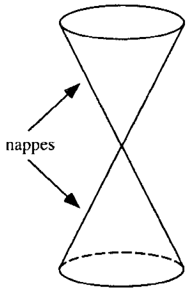

One of the two pieces of a double CONE (i.e., two CONES placed apex to apex).

see also CONE

Narcissistic Number

An n-Digit number which is the Sum of the nth Powers of its Digits is called an n-narcissistic number, or sometimes an Armstrong Number or Perfect Digital Invariant (Madachy 1979). The smallest example other than the trivial 1-Digit numbers is

$$153 = 1^3 + 5^3 + 3^3.$$

The series of smallest narcissistic numbers of n digits are 0, (none), 153, 1634, 54748, 548834, ... (Sloane's A014576). Hardy (1993) wrote, "There are just four numbers, after unity, which are the sums of the cubes of their digits: $153 = 1^3+5^3+3^3$, $370 = 3^3+7^3+0^3$, $371 = 3^3+7^3+1^3$, and $407 = 4^3+0^3+7^3$. These are odd facts, very suitable for puzzle columns and likely to amuse amateurs, but there is nothing in them which appeals to the mathematician." The following table gives the generalization of these "unappealing" numbers to other Powers (Madachy 1979, p. 164).

n	n-Narcissistic Numbers
1	0, 1, 2, 3, 4, 5, 6, 7, 8, 9
2	none
3	153, 370, 371, 407
4	1634, 8208, 9474
5	54748, 92727, 93084
6	548834
7	1741725, 4210818, 9800817, 9926315
8	24678050, 24678051, 88593477
9	146511208, 472335975, 534494836, 912985153
10	4679307774

A total of 88 narcissistic numbers exist in base-10, as proved by D. Winter in 1985 and verified by D. Hoey. These numbers exist for only 1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 16, 17, 19, 20, 21, 23, 24, 25, 27, 29, 31, 32, 33, 34, 35, 37, 38, and 39 digits. It can easily be shown that base-10 n-narcissistic numbers can exist only for $n \leq 60$, since

$$n \cdot 9^n < 10^{n-1}$$

for $n > 60$. The largest base-10 narcissistic number is the 39-narcissistic

$$115132219018736992565095597973971522401$$

A table of the largest known narcissistic numbers in various Bases is given by Pickover (1995). A tabulation of narcissistic numbers in various bases is given by (Corning).

A closely related set of numbers generalize the narcissistic number to n-Digit numbers which are the sums of *any* single Power of their Digits. For example, 4150 is a 4-Digit number which is the sum of fifth Powers of its Digits. Since the number of digits is not equal to the power to which they are taken for such numbers, it is *not* a narcissistic number. The smallest numbers which are sums of *any* single positive power of their digits are 1, 2, 3, 4, 5, 6, 7, 8, 9, 153, 370, 371, 407, 1634, 4150, 4151, 8208, 9474, ... (Sloane's A023052), with powers 1, 1, 1, 1, 1, 1, 1, 1, 1, 3, 3, 3, 3, 4, 5, 5, 4, 4, ... (Sloane's A046074).

The smallest numbers which are equal to the nth powers of their digits for $n = 3, 4, \ldots$, are 153, 1634, 4150, 548834, 1741725, ... (Sloane's A003321). The n-digit numbers equal to the sum of nth powers of their digits (a finite sequence) are called Armstrong Numbers or Plus Perfect Numbers and are given by 1, 2, 3, 4, 5, 6, 7, 8, 9, 153, 370, 371, 407, 1634, 8208, 9474, 54748, ... (Sloane's A005188).

If the sum-of-kth-powers-of-digits operation applied iteratively to a number n eventually returns to n, the smallest number in the sequence is called a k-Recurring Digital Invariant.

see also Additive Persistence, Digital Root, Digitadition, Kaprekar Number, Multiplicative Digital Root, Multiplicative Persistence, Recurring Digital Invariant, Vampire Number

References

Corning, T. "Exponential Digital Invariants." `http://members.aol.com/tec153/Edi4web/Edi.html`.

Hardy, G. H. *A Mathematician's Apology.* New York: Cambridge University Press, p. 105, 1993.

Madachy, J. S. "Narcissistic Numbers." *Madachy's Mathematical Recreations.* New York: Dover, pp. 163–173, 1979.

Pickover, C. A. *Keys to Infinity.* New York: W. H. Freeman, pp. 169–170, 1995.

Rumney, M. "Digital Invariants." *Recr. Math. Mag.* No. 12, 6–8, Dec. 1962.

Sloane, N. J. A. Sequences A014576, A023052, A005188/M0488, and A003321/M5403 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Weisstein, E. W. "Narcissistic Numbers." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Narcissistic.dat`.

Nash Equilibrium

A set of Mixed Strategies for finite, noncooperative Games of two or more players in which no player can improve his payoff by unilaterally changing strategy.

see also Fixed Point, Game, Mixed Strategy, Nash's Theorem

Nash's Theorem

A theorem in Game Theory which guarantees the existence of a Nash Equilibrium for Mixed Strategies in finite, noncooperative Games of two or more players.

see also Mixed Strategy, Nash Equilibrium

Nasik Square

see Panmagic Square

Nasty Knot

An UNKNOT which can only be unknotted by first increasing the number of crossings.

Natural Density

see NATURAL INVARIANT

Natural Equation

A natural equation is an equation which specifies a curve independent of any choice of coordinates or parameterization. The study of natural equations began with the following problem: given two functions of one parameter, find the SPACE CURVE for which the functions are the CURVATURE and TORSION.

Euler gave an integral solution for plane curves (which always have TORSION $\tau = 0$). Call the ANGLE between the TANGENT line to the curve and the x-AXIS ϕ the TANGENTIAL ANGLE, then

$$\phi = \int \kappa(s)\, ds, \tag{1}$$

where κ is the CURVATURE. Then the equations

$$\kappa = \kappa(s) \tag{2}$$

$$\tau = 0, \tag{3}$$

where τ is the TORSION, are solved by the curve with parametric equations

$$x = \int \cos\phi\, ds \tag{4}$$

$$y = \int \sin\phi\, ds. \tag{5}$$

The equations $\kappa = \kappa(s)$ and $\tau = \tau(s)$ are called the natural (or INTRINSIC) equations of the space curve. An equation expressing a plane curve in terms of s and RADIUS OF CURVATURE R (or κ) is called a CESÀRO EQUATION, and an equation expressing a plane curve in terms of s and ϕ is called a WHEWELL EQUATION.

Among the special planar cases which can be solved in terms of elementary functions are the CIRCLE, LOGARITHMIC SPIRAL, CIRCLE INVOLUTE, and EPICYCLOID. Enneper showed that each of these is the projection of a HELIX on a CONIC surface of revolution along the axis of symmetry. The above cases correspond to the CYLINDER, CONE, PARABOLOID, and SPHERE.

see also CESÀRO EQUATION, INTRINSIC EQUATION, WHEWELL EQUATION

References

Cesàro, E. *Lezioni di Geometria Intrinseca.* Napoli, Italy, 1896.
Euler, L. *Comment. Acad. Petropolit.* **8**, 66–85, 1736.
Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 111–112, 1993.
Melzak, Z. A. *Companion to Concrete Mathematics, Vol. 2.* New York: Wiley, 1976.
Struik, D. J. *Lectures on Classical Differential Geometry.* New York: Dover, pp. 26–28, 1988.

Natural Independence Phenomenon

A type of mathematical result which is considered by most logicians as more natural than the METAMATHEMATICAL incompleteness results first discovered by Gödel. Finite combinatorial examples include GOODSTEIN'S THEOREM, a finite form of RAMSEY'S THEOREM, and a finite form of KRUSKAL'S TREE THEOREM (Kirby and Paris 1982; Smorynski 1980, 1982, 1983; Gallier 1991).

see also GÖDEL'S INCOMPLETENESS THEOREM, GOODSTEIN'S THEOREM, KRUSKAL'S TREE THEOREM, RAMSEY'S THEOREM

References

Gallier, J. "What's so Special about Kruskal's Theorem and the Ordinal Gamma[0]? A Survey of Some Results in Proof Theory." *Ann. Pure and Appl. Logic* **53**, 199–260, 1991.
Kirby, L. and Paris, J. "Accessible Independence Results for Peano Arithmetic." *Bull. London Math. Soc.* **14**, 285–293, 1982.
Smorynski, C. "Some Rapidly Growing Functions." *Math. Intell.* **2**, 149–154, 1980.
Smorynski, C. "The Varieties of Arboreal Experience." *Math. Intell.* **4**, 182–188, 1982.
Smorynski, C. "'Big' News from Archimedes to Friedman." *Not. Amer. Math. Soc.* **30**, 251–256, 1983.

Natural Invariant

Let $\rho(x)\, dx$ be the fraction of time a typical dynamical ORBIT spends in the interval $[x, x + dx]$, and let $\rho(x)$ be normalized such that

$$\int \rho(x)\, dx = 1$$

over the entire interval of the map. Then the fraction the time an ORBIT spends in a finite interval $[a, b]$, is given by

$$\int_a^b \rho(x)\, dx.$$

The natural invariant is also called the INVARIANT DENSITY or NATURAL DENSITY.

Natural Logarithm

The LOGARITHM having base e, where

$$e = 2.718281828\ldots, \tag{1}$$

which can be defined

$$\ln x \equiv \int_1^x \frac{dt}{t} \tag{2}$$

for $x > 0$. The natural logarithm can also be defined for COMPLEX NUMBERS as

$$\ln z \equiv \ln|z| + i\arg(z), \tag{3}$$

where $|z|$ is the MODULUS and $\arg(z)$ is the ARGUMENT. The natural logarithm is especially useful in CALCULUS because its DERIVATIVE is given by the simple equation

$$\frac{d}{dx}\ln x = \frac{1}{x}, \tag{4}$$

whereas logarithms in other bases have the more complicated DERIVATIVE

$$\frac{d}{dx}\log_b x = \frac{1}{x\ln b}. \tag{5}$$

The MERCATOR SERIES

$$\ln(1+x) = x - \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 - \ldots \tag{6}$$

gives a TAYLOR SERIES for the natural logarithm.

CONTINUED FRACTION representations of logarithmic functions include

$$\ln(1+x) = \cfrac{x}{1+\cfrac{1^2x}{2+\cfrac{1^2x}{3+\cfrac{2^2x}{4+\cfrac{2^2x}{5+\cfrac{3^2x}{6+\cfrac{3^2x}{7+\ldots}}}}}}} \tag{7}$$

$$\ln\left(\frac{1+x}{1-x}\right) = \cfrac{2x}{1-\cfrac{x^2}{3-\cfrac{4x^2}{5-\cfrac{9x^2}{7-\cfrac{16x^2}{9-\ldots}}}}}. \tag{8}$$

For a COMPLEX NUMBER z, the natural logarithm satisfies

$$\ln z = \ln[re^{i(\theta+2n\pi)}] = \ln r + i(\theta + 2n\pi) \tag{9}$$

$$PV(\ln z) = \ln r + i\theta, \tag{10}$$

where PV is the PRINCIPAL VALUE.

Some special values of the natural logarithm are

$$\ln 1 = 0 \tag{11}$$

$$\ln 0 = -\infty \tag{12}$$

$$\ln(-1) = \pi i \tag{13}$$

$$\ln(\pm i) = \pm\tfrac{1}{2}\pi i. \tag{14}$$

An identity for the natural logarithm of 2 discovered using the PSLQ ALGORITHM is

$$\begin{aligned}(\ln 2)^2 = -\frac{1}{6}\sum_{k=0}^{\infty}\frac{1}{16^k}\Bigg[&-\frac{3}{(8k)^2} - \frac{16}{(8k+1)^2}\\ &-\frac{40}{(8k+2)^2} - \frac{8}{(8k+3)^2} - \frac{28}{(8k+4)^2} + \frac{4}{(8k+5)^2}\\ &-\frac{28}{(8k+4)^2} - \frac{4}{(8k+5)^2} + \frac{10}{(8k+5)^2} - \frac{2}{(8k+7)^2}\Bigg]\end{aligned} \tag{15}$$

(Bailey *et al.* 1995, Bailey and Plouffe).

see also e, JENSEN'S FORMULA, LG, LOGARITHM

References

Bailey, D.; Borwein, P.; and Plouffe, S. "On the Rapid Computation of Various Polylogarithmic Constants." `http://www.cecm.sfu.ca/~pborwein/PAPERS/P123.ps.`

Bailey, D. and Plouffe, S. "Recognizing Numerical Constants." `http://www.cecm.sfu.ca/organics/papers/bailey.`

Natural Measure

$\mu_i(\epsilon)$, sometimes denoted $P_i(\epsilon)$, is the probability that element i is populated, normalized such that

$$\sum_{i=1}^{N}\mu_i(\epsilon) = 1.$$

see also INFORMATION DIMENSION, q-DIMENSION

Natural Norm

Let $||\mathbf{z}||$ be a VECTOR NORM of $\mathbf{z}$ such that

$$||\mathsf{A}|| = \max_{||\mathbf{z}||=1}||\mathsf{A}\mathbf{z}||.$$

Then $||\mathsf{A}||$ is a MATRIX NORM which is said to be the natural norm INDUCED (or SUBORDINATE) to the VECTOR NORM $||\mathbf{z}||$. For any natural norm,

$$||\mathsf{I}|| = 1,$$

where I is the IDENTITY MATRIX. The natural matrix norms induced by the L_1-NORM, L_2-NORM, and L_∞-NORM are called the MAXIMUM ABSOLUTE COLUMN SUM NORM, SPECTRAL NORM, and MAXIMUM ABSOLUTE ROW SUM NORM, respectively.

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1115, 1979.

Natural Number

A POSITIVE INTEGER 1, 2, 3, ... (Sloane's A000027). The set of natural numbers is denoted $\mathbb{N}$ or $\mathbb{Z}^+$. Unfortunately, 0 is sometimes also included in the list of "natural" numbers (Bourbaki 1968, Halmos 1974), and there seems to be no general agreement about whether to include it.

Due to lack of standard terminology, the following terms are recommended in preference to "COUNTING NUMBER," "natural number," and "WHOLE NUMBER."

Set	Name	Symbol
$\ldots, -2, -1, 0, 1, 2, \ldots$	integers	$\mathbb{Z}$
$1, 2, 3, 4, \ldots$	positive integers	$\mathbb{Z}^+$
$0, 1, 2, 3, 4 \ldots$	nonnegative integers	$\mathbb{Z}^*$
$-1, -2, -3, -4, \ldots$	negative integers	$\mathbb{Z}^-$

see also COUNTING NUMBER, INTEGER, $\mathbb{N}$, POSITIVE, $\mathbb{Z}$, $\mathbb{Z}^-$, $\mathbb{Z}^+$, $\mathbb{Z}^*$

References

Bourbaki, N. *Elements of Mathematics: Theory of Sets.* Paris, France: Hermann, 1968.

Courant, R. and Robbins, H. "The Natural Numbers." Ch. 1 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 1–20, 1996.

Halmos, P. R. *Naive Set Theory.* New York: Springer-Verlag, 1974.

Sloane, N. J. A. Sequence A000027/M0472 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Naught

The British word for "ZERO." It is often used to indicate 0 subscripts, so a_0 would be spoken as "a naught."

see also ZERO

Navigation Problem

A problem in the CALCULUS OF VARIATIONS. Let a vessel traveling at constant speed c navigate on a body of water having surface velocity

$$\begin{aligned} u &= u(x,y) \\ v &= v(x,y). \end{aligned}$$

The navigation problem asks for the course which travels between two points in minimal time.

References

Sagan, H. *Introduction to the Calculus of Variations.* New York: Dover, pp. 226–228, 1992.

Near-Integer

see ALMOST INTEGER

Near Noble Number

A REAL NUMBER $0 < \nu < 1$ whose CONTINUED FRACTION is periodic, and the periodic sequence of terms is composed of a string of 1s followed by an INTEGER $n > 1$,

$$\nu = [\overline{\underbrace{1,1,\ldots,1}_{P},n}]. \tag{1}$$

This can be written in the form

$$\nu = [\underbrace{1,1,\ldots,1}_{P},n,\nu^{-1}], \tag{2}$$

which can be solved to give

$$\nu = \tfrac{1}{2}n\left(\sqrt{1+4\frac{nF_{P-1}+F_{P-2}}{n^2F_P}}-1\right), \tag{3}$$

where F_n is a FIBONACCI NUMBER. The special case $n = 2$ gives

$$\nu = \sqrt{\frac{F_{P+2}}{F_P}}-1. \tag{4}$$

see also NOBLE NUMBER

References

Schroeder, M. R. *Number Theory in Science and Communication: With Applications in Cryptography, Physics, Digital Information, Computing, and Self-Similarity, 2nd enl. ed., corr. printing.* Berlin: Springer-Verlag, 1990.

Schroeder, M. "Noble and Near Noble Numbers." In *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise.* New York: W. H. Freeman, pp. 392–394, 1991.

Near-Pencil

An arrangement of $n \geq 3$ points such that $n-1$ of them are COLLINEAR.

see also GENERAL POSITION, ORDINARY LINE, PENCIL

References

Guy, R. K. "Unsolved Problems Come of Age." *Amer. Math. Monthly* **96**, 903–909, 1989.

Nearest Integer Function

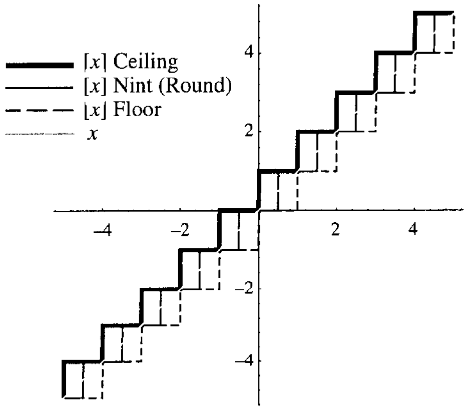

The nearest integer function nint(x) of x, also called NINT or the ROUND function, is defined such that $[x]$ is the INTEGER closest to x. It is shown as the thin solid curve in the above plot. Note that while $[x]$ is used to denote the nearest integer function in this work, $[x]$ is more commonly used to denote the FLOOR FUNCTION $\lfloor x \rfloor$.

see also CEILING FUNCTION, FLOOR FUNCTION

Nearest Neighbor Problem

The problem of identifying the point from a set of points which is nearest to a given point according to some measure of distance. The nearest neighborhood problem involves identifying the locus of points lying nearer to the query point than to any other point in the set.

References

Martin, E. C. "Computational Geometry." `http://www.mathsource.com/cgi-bin/MathSource/Enhancements/DiscreteMath/0200-181.`

Necessary

A CONDITION which must hold for a result to be true, but which does not guarantee it to be true. If a CONDITION is both NECESSARY and SUFFICIENT, then the result is said to be true IFF the CONDITION holds.

see also SUFFICIENT

Necker Cube

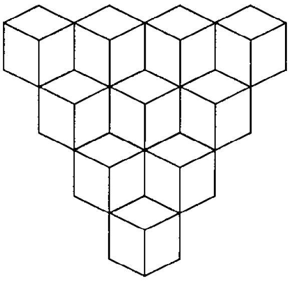

An ILLUSION in which a 2-D drawing of an array of CUBES appear to simultaneously protrude and intrude into the page.

References

Fineman, M. *The Nature of Visual Illusion.* New York: Dover, pp. 25 and 118, 1996.

Jablan, S. "Impossible Figures." `http://members.tripod.com/~modularity/impos.htm.`

Newbold, M. "Animated Necker Cube." `http://www.sover.net/~manx/necker.html.`

Necklace

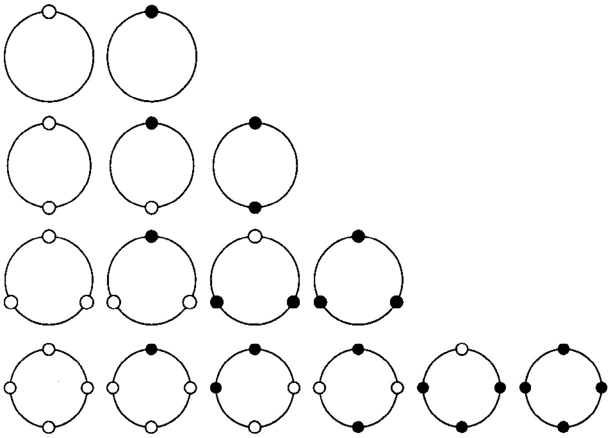

In the technical COMBINATORIAL sense, an a-ary necklace $N(n,a)$ of length n is a string of n characters, each of a possible types. Rotation is ignored, in the sense that $b_1b_2\ldots b_n$ is equivalent to $b_kb_{k+1}\cdots b_1b_2\cdots b_{k-1}$ for any k, but reversal of strings is respected. Necklaces therefore correspond to circular collections of beads in which the FIXED necklace may not be picked up out of the PLANE (so that opposite orientations are not considered equivalent).

The number of distinct FREE necklaces $N'(n,a)$ of n beads, each of a possible colors, in which opposite orientations (MIRROR IMAGES) are regarded as equivalent (so the necklace *can* be picked up out of the PLANE and flipped over) can be found as follows. Find the DIVISORS of n and label them $d_1 \equiv 1$, d_2, ..., $d_\nu(n) \equiv n$ where $\nu(n)$ is the number of DIVISORS of n. Then

$$N'(n,a) = \frac{1}{2n}\begin{cases} \sum_{i=1}^{\nu(n)} \phi(d_i)a^{n/d_i} + na^{(n+1)/2} \\ \quad \text{for } n \text{ odd} \\ \sum_{i=1}^{\nu(n)} \phi(d_i)a^{n/d_i} + \frac{1}{2}n(1+a)a^{n/2} \\ \quad \text{for } n \text{ even,} \end{cases}$$

where $\phi(x)$ is the TOTIENT FUNCTION. For $a = 2$ and $n = p$ an ODD PRIME, this simplifies to

$$N'(p,2) = \frac{2^{p-1}-1}{p} + 2^{(p-1)/2} + 1.$$

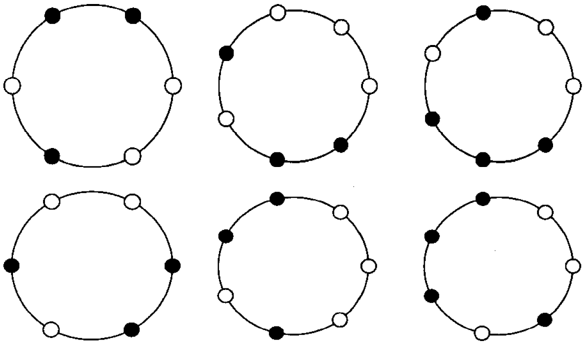

A table of the first few numbers of necklaces for $a = 2$ and $a = 3$ follows. Note that $N(n, 2)$ is larger than $N'(n, 2)$ for $n \geq 6$. For $n = 6$, the necklace 110100 is inequivalent to its MIRROR IMAGE 0110100, accounting for the difference of 1 between $N(6, 2)$ and $N'(6, 2)$. Similarly, the two necklaces 0010110 and 0101110 are inequivalent to their reversals, accounting for the difference of 2 between $N(7, 2)$ and $N'(7, 2)$.

n	$N(n,2)$	$N'(n,2)$	$N'(n,3)$
Sloane	000031	000029	027671
1	2	2	3
2	3	3	6
3	4	4	10
4	6	6	21
5	8	8	39
6	14	13	92
7	20	18	198
8	36	30	498
9	60	46	1219
10	108	78	3210
11	188	126	8418
12	352	224	22913
13	632	380	62415
14	1182	687	173088
15	2192	1224	481598

Ball and Coxeter (1987) consider the problem of finding the number of distinct arrangements of n people in a ring such that no person has the same two neighbors two or more times. For 8 people, there are 21 such arrangements.

see also ANTOINE'S NECKLACE, DE BRUIJN SEQUENCE, FIXED, FREE, IRREDUCIBLE POLYNOMIAL, JOSEPHUS PROBLEM, LYNDON WORD

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 49–50, 1987.

Dudeney, H. E. Problem 275 in *536 Puzzles & Curious Problems.* New York: Scribner, 1967.

Gardner, M. *Martin Gardner's New Mathematical Diversions from Scientific American.* New York: Simon and Schuster, pp. 240–246, 1966.

Gilbert, E. N. and Riordan, J. "Symmetry Types of Periodic Sequences." *Illinois J. Math.* **5**, 657–665, 1961.

Riordan, J. "The Combinatorial Significance of a Theorem of Pólya." *J. SIAM* **4**, 232–234, 1957.

Riordan, J. *An Introduction to Combinatorial Analysis.* New York: Wiley, p. 162, 1980.

Ruskey, F. "Information on Necklaces, Lyndon Words, de Bruijn Sequences." `http://sue.csc.uvic.ca/~cos/inf/neck/NecklaceInfo.html`.

Sloane, N. J. A. Sequences A000029/M0563, A000031/M0564, and A001869/M3860 in "An On-Line Version of the Encyclopedia of Integer Sequences." `http://www.research.att.com/~njas/sequences/eisonline.html`.

Sloane, N. J. A. and Plouffe, S. Extended entry for M3860 in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.

Needle

see BUFFON-LAPLACE NEEDLE PROBLEM, BUFFON'S NEEDLE PROBLEM, KAKEYA NEEDLE PROBLEM

Negation

see NOT

Negative

A quantity less than ZERO (< 0), denoted with a MINUS SIGN, i.e., $-x$.

see also NONNEGATIVE, NONPOSITIVE, NONZERO, POSITIVE, ZERO

Negative Binomial Distribution

Also known as the PASCAL DISTRIBUTION and PÓLYA DISTRIBUTION. The probability of $r - 1$ successes and x failures in $x + r - 1$ trials, and success on the $(x + r)$th trial is

$$\begin{aligned} p\left[\binom{x+r-1}{r-1} p^{r-1}(1-p)^{[(x+r-1)-(r-1)]}\right] &= \left[\binom{x+r-1}{r-1} p^{r-1}(1-p)^{x}\right] p \\ &= \binom{x+r-1}{r-1} p^{r}(1-p)^{x}, \quad (1) \end{aligned}$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT. Let

$$P = \frac{1-p}{p} \quad (2)$$

$$Q = \frac{1}{p}. \quad (3)$$

The CHARACTERISTIC FUNCTION is given by

$$\phi(t) = (Q - Pe^{it})^{-r}, \quad (4)$$

and the MOMENT-GENERATING FUNCTION by

$$M(t) = \langle e^{tx} \rangle = \sum_{x=0}^{\infty} e^{tx} \binom{x+r-1}{r-1} p^{r}(1-p)^{x}, \quad (5)$$

but, since $\binom{N}{n} = \binom{N}{N-m}$,

$$\begin{aligned} M(t) &= p^{r} \sum_{x=0}^{\infty} \binom{x+r-1}{x} [(1-p)e^{t}]^{x} \\ &= p^{r}[1-(1-p)e^{t}]^{-r} \quad (6) \\ M'(t) &= p^{r}(-r)[1-(1-p)e^{t}]^{-r-1}(p-1)e^{t} \\ &= p^{r}(1-p)r[1-(1-p)e^{t}]^{-r-1}e^{t} \quad (7) \\ M''(t) &= (1-p)rp^{r}(1-e^{t}+pe^{t})^{-r-2} \\ &\quad \times (-1-e^{t}r+e^{t}pr)e^{t} \quad (8) \\ M'''(t) &= (1-p)rp^{r}(1-e^{t}+e^{t}p)^{-r-3} \\ &\quad \times [1+e^{t}(1-p+3r-3pr) \\ &\quad + r^{2}e^{2t}(1-p)^{2}]e^{t}. \quad (9) \end{aligned}$$

The MOMENTS about zero $\mu'_n = M^n(0)$ are therefore

$$\mu'_1 = \mu = \frac{r(1-p)}{p} = \frac{rq}{p} \tag{10}$$

$$\mu'_2 = \frac{r(1-p)[1-r(p-1)]}{p^2} = \frac{rq(1-rq)}{p^2} \tag{11}$$

$$\mu'_3 = \frac{(1-p)r(2-p+3r-3pr+r^2-2pr^2+p^2r^2}{p^3} \tag{12}$$

$$\mu'_4 = \frac{(-1+p)r(-6+6p-p^2-11r+15pr-4p^2r-6r^2}{p^4} + \frac{12pr^2-6p^2r^2-r^3+3pr^3-3p^2r^3+p^3r^3)}{p^4}. \tag{13}$$

(Beyer 1987, p. 487, apparently gives the MEAN incorrectly.) The MOMENTS about the mean are

$$\mu_2 = \sigma^2 = \frac{r(1-p)}{p^2} \tag{14}$$

$$\mu_3 = \frac{r(2-3p+p^2)}{p^3} = \frac{r(p-1)(p-2)}{p^3} \tag{15}$$

$$\mu_4 = \frac{r(1-p)(6-6p+p^2+3r-3pr)}{p^4}. \tag{16}$$

The MEAN, VARIANCE, SKEWNESS and KURTOSIS are then

$$\mu = \frac{r(1-p)}{p} \tag{17}$$

$$\begin{aligned}\gamma_1 &= \frac{\mu_3}{\sigma^3} = \frac{r(p-1)(p-2)}{p^3}\left[\frac{p^2}{r(1-p)}\right]^{3/2} \\ &= \frac{r(2-p)(1-p)}{p^3}\,\frac{p^3}{r(1-p)\sqrt{1-p}} \\ &= \frac{2-p}{\sqrt{r(1-p)}}\end{aligned} \tag{18}$$

$$\begin{aligned}\gamma_2 &= \frac{\mu_4}{\sigma^4} - 3 \\ &= \frac{-6+6p-p^2-3r+3pr}{(p-1)r},\end{aligned} \tag{19}$$

which can also be written

$$\mu = nP \tag{20}$$

$$\mu_2 = nPQ \tag{21}$$

$$\gamma_1 = \frac{Q+P}{\sqrt{rPQ}} \tag{22}$$

$$\gamma_2 = \frac{1+6PQ}{rPQ} - 3. \tag{23}$$

The first CUMULANT is

$$\kappa_1 = nP, \tag{24}$$

and subsequent CUMULANTS are given by the recurrence relation

$$\kappa_{r+1} = PQ\frac{d\kappa_r}{dQ}. \tag{25}$$

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 533, 1987.

Spiegel, M. R. *Theory and Problems of Probability and Statistics.* New York: McGraw-Hill, p. 118, 1992.

Negative Binomial Series

The SERIES which arises in the BINOMIAL THEOREM for NEGATIVE integral n,

$$\begin{aligned}(x+a)^{-n} &= \sum_{k=0}^{\infty}\binom{-n}{k}x^k a^{-n-k} \\ &= \sum_{k=0}^{\infty}(-1)^k\binom{n+k-1}{k}x^k a^{-n-k}.\end{aligned}$$

For $a = 1$, the negative binomial series simplifies to

$$(x+1)^{-n} = 1 - nx + \tfrac{1}{2}n(n+1)x^2 - \tfrac{1}{6}n(n+1)(n+2) + \ldots.$$

see also BINOMIAL SERIES, BINOMIAL THEOREM

Negative Likelihood Ratio

The term NEGATIVE likelihood ratio is also used (especially in medicine) to test nonnested complementary hypotheses as follows,

$$\text{NLR} = \frac{[\text{true negative rate}]}{[\text{false negative rate}]} = \frac{[\text{specificity}]}{1-[\text{sensitivity}]}.$$

see also LIKELIHOOD RATIO, SENSITIVITY, SPECIFICITY

Negative Integer

see $\mathbb{Z}^-$

Negative Pedal Curve

Given a curve C and O a fixed point called the PEDAL POINT, then for a point P on C, draw a LINE PERPENDICULAR to OP. The ENVELOPE of these LINES as P describes the curve C is the negative pedal of C.

see also PEDAL CURVE

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 46–49, 1972.

Lockwood, E. H. "Negative Pedals." Ch. 19 in *A Bcok of Curves.* Cambridge, England: Cambridge University Press, pp. 156–159, 1967.

Neighborhood

The word neighborhood is a word with many different levels of meaning in mathematics. One of the most general concepts of a neighborhood of a point $x \in \mathbb{R}^n$ (also called an EPSILON-NEIGHBORHOOD or infinitesimal OPEN SET) is the set of points inside an n-BALL with center x and RADIUS $\epsilon > 0$.

Neile's Parabola

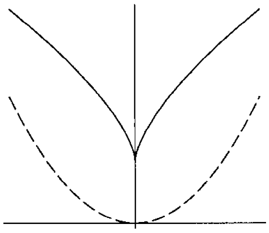

The solid curve in the above figure which is the EVOLUTE of the PARABOLA (dashed curve). In CARTESIAN COORDINATES,

$$y = \tfrac{3}{4}(2x)^{2/3} + \tfrac{1}{2}.$$

Neile's parabola is also called the SEMICUBICAL PARABOLA, and was discovered by William Neile in 1657. It was the first nontrivial ALGEBRAIC CURVE to have its ARC LENGTH computed. Wallis published the method in 1659, giving Neile the credit (MacTutor Archive).

see also PARABOLA EVOLUTE

References

Lee, X. "Semicubic Parabola." `http://www.best.com/~xah/Special Plane Curves _ dir / Semicubic Parabola _ dir / semicubicParabola.html`.

MacTutor History of Mathematics Archive. "Neile's Semi-Cubical Parabola." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Neiles.html`.

Nephroid

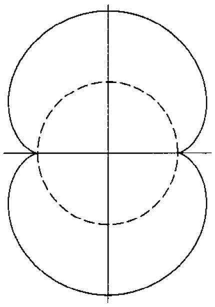

The 2-CUSPED EPICYCLOID is called a nephroid. Since $n = 2$, $a = b/2$, and the equation for r^2 in terms of the parameter ϕ is given by EPICYCLOID equation

$$r^2 = \frac{a^2}{n^2}\left[(n^2 + 2n + 2) - 2(n+1)\cos(n\phi)\right] \qquad (1)$$

with $n = 2$,

$$\begin{aligned} r^2 &= \frac{a^2}{2^2}[(2^2 + 2\cdot 2 + 2) - 2(2+1)\cos(2\phi)] \\ &= \tfrac{1}{4}a^2[10 - 6\cos(2\phi)] = \tfrac{1}{2}a^2[5 - 3\cos(2\phi)], \qquad (2) \end{aligned}$$

where

$$\tan\theta = \frac{3\sin\phi - \sin(3\phi)}{3\cos\phi - \cos(3\phi)}. \qquad (3)$$

This can be written

$$\left(\frac{r}{2a}\right)^{2/3} = [\sin(\tfrac{1}{2}\theta)]^{2/3} + [\cos(\tfrac{1}{2}\theta)]^{2/3}. \qquad (4)$$

The parametric equations are

$$x = a[3\cos t - \cos(3t)] \qquad (5)$$
$$y = a[3\sin t - \sin(3t)]. \qquad (6)$$

The Cartesian equation is

$$(x^2 + y^2 - 4a^2)^3 = 108a^4y^2. \qquad (7)$$

The name nephroid means "kidney shaped" and was first used for the two-cusped EPICYCLOID by Proctor in 1878 (MacTutor Archive). The nephroid has ARC LENGTH $24a$ and AREA $12\pi^2a^2$. The CATACAUSTIC for rays originating at the CUSP of a CARDIOID and reflected by it is a nephroid. Huygens showed in 1678 that the nephroid is the CATACAUSTIC of a CIRCLE when the light source is at infinity. He published this fact in *Traité de la luminère* in 1690 (MacTutor Archive).

see also ASTROID, DELTOID, FREETH'S NEPHROID

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 169–173, 1972.

Lee, X. "Nephroid." `http://www.best.com/~xah/Special PlaneCurves_dir/Nephroid_dir/nephroid.html`.

Lockwood, E. H. "The Nephroid." Ch. 7 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 62–71, 1967.

MacTutor History of Mathematics Archive. "Nephroid." `http://www-groups.dcs.st-and.ac.uk/~history/Curves /Nephroid.html`.

Yates, R. C. "Nephroid." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 152–154, 1952.

Nephroid Evolute

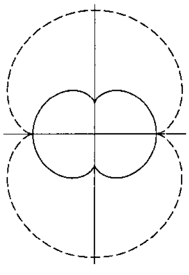

The EVOLUTE of the NEPHROID given by

$$x = \tfrac{1}{2}[3\cos t - \cos(3t)]$$
$$y = \tfrac{1}{2}[3\sin t - \sin(3t)]$$

is given by

$$x = \cos^3 t$$
$$y = \tfrac{1}{4}[3\sin t + \sin(3t)],$$

which is another NEPHROID.

Nephroid Involute

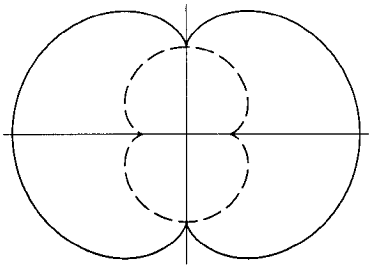

The INVOLUTE of the NEPHROID given by

$$\begin{aligned} x &= \tfrac{1}{2}[3\cos t - \cos(3t)] \\ y &= \tfrac{1}{2}[3\sin t - \sin(3t)] \end{aligned}$$

beginning at the point where the nephroid cuts the y-AXIS is given by

$$\begin{aligned} x &= 4\cos^3 t \\ y &= 3\sin t + \sin(3t), \end{aligned}$$

another NEPHROID. If the INVOLUTE is begun instead at the CUSP, the result is CAYLEY'S SEXTIC.

Néron-Severi Group

Let V be a complete normal VARIETY, and write $G(V)$ for the group of divisors, $G_n(V)$ for the group of divisors numerically equal to 0, and $G_a(V)$ the group of divisors algebraically equal to 0. Then the finitely generated QUOTIENT GROUP $NS(V) = G(V)/G_a(V)$ is called the Néron-Severi group.

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 75, 1980.

Nerve

The SIMPLICIAL COMPLEX formed from a family of objects by taking sets that have nonempty intersections.

see also DELAUNAY TRIANGULATION, SIMPLICIAL COMPLEX

Nested Hypothesis

Let S be the set of all possibilities that satisfy HYPOTHESIS H, and let S' be the set of all possibilities that satisfy HYPOTHESIS H'. Then H' is a nested hypothesis within H IFF $S' \subset S$, where $\subset$ denotes the PROPER SUBSET.

see also LOG LIKELIHOOD PROCEDURE

Nested Radical

A RADICAL of the form

$$\sqrt{n+\sqrt{n+\sqrt{n+\ldots}}}. \tag{1}$$

For this to equal a given INTEGER x, it must be true that

$$x = \sqrt{n+\sqrt{n+\sqrt{n+\ldots}}} = \sqrt{n+x}, \tag{2}$$

so

$$x^2 = n + x \tag{3}$$

and

$$n = x(x-1). \tag{4}$$

Nested radicals in the computation of PI,

$$\frac{2}{\pi} = \sqrt{\tfrac{1}{2}}\sqrt{\tfrac{1}{2}+\sqrt{\tfrac{1}{2}}}\sqrt{\tfrac{1}{2}+\sqrt{\tfrac{1}{2}+\sqrt{\tfrac{1}{2}}}}\cdots \tag{5}$$

and in TRIGONOMETRICAL values of COSINE and SINE for arguments of the form $\pi/2^n$, e.g.,

$$\sin\left(\frac{\pi}{8}\right) = \tfrac{1}{2}\sqrt{2-\sqrt{2}} \tag{6}$$

$$\cos\left(\frac{\pi}{8}\right) = \tfrac{1}{2}\sqrt{2+\sqrt{2}} \tag{7}$$

$$\sin\left(\frac{\pi}{16}\right) = \tfrac{1}{2}\sqrt{2-\sqrt{2+\sqrt{2}}} \tag{8}$$

$$\cos\left(\frac{\pi}{16}\right) = \tfrac{1}{2}\sqrt{2+\sqrt{2+\sqrt{2}}}. \tag{9}$$

see also SQUARE ROOT

References

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 14–20, 1994.

Net

A generalization of a SEQUENCE used in general topology and ANALYSIS when the spaces being dealt with are not FIRST-COUNTABLE. (Sequences provide an adequate way of dealing with CONTINUITY for FIRST-COUNTABLE SPACES.) Nets are used in the study of the RIEMANN INTEGRAL.

see also FIBER BUNDLE, FIBER SPACE, FIBRATION

Net (Polyhedron)

A plane diagram in which the EDGES of a POLYHEDRON are shown. All convex POLYHEDRA have nets, but not all concave polyhedra do (the constituent POLYGONS can overlap one another when a concave POLYHEDRON is flattened out). The GREAT DODECAHEDRON and STELLA OCTANGULA are examples of a concave polyhedron which have nets.

Netto's Conjecture

The probability that two elements P_1 and P_2 of a SYMMETRIC GROUP generate the entire GROUP tends to 3/4 as $n \to \infty$. This was proven by Dixon in 1967.

References

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 31, 1983.

Network

A DIRECTED GRAPH having a SOURCE, SINK, and a bound on each edge.

see also GRAPH (GRAPH THEORY), SINK (DIRECTED GRAPH), SMITH'S NETWORK THEOREM, SOURCE

Neuberg Circles

The LOCUS of the VERTEX A_1 of a TRIANGLE on a given base A_2A_3 and with a given BROCARD ANGLE ω is a CIRCLE on either side of A_2A_3. From the center N_1, the base A_2A_3 subtends the ANGLE 2ω. The RADIUS of the CIRCLE is

$$r = \tfrac{1}{2}a_1\sqrt{\cot^2\omega - 3}.$$

see also BROCARD ANGLE

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 287–290, 1929.

Neumann Algebra

see VON NEUMANN ALGEBRA

Neumann Boundary Conditions

PARTIAL DIFFERENTIAL EQUATION BOUNDARY CONDITIONS which give the normal derivative on a surface.

see also BOUNDARY CONDITIONS, CAUCHY BOUNDARY CONDITIONS

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 679, 1953.

Neumann Function

see BESSEL FUNCTION OF THE SECOND KIND

Neumann Polynomial

Polynomials which obey the RECURRENCE RELATION

$$O_{n+1}(x) = (n+1)\frac{2}{x}O_n(x) - \frac{n+1}{n-1}O_{n-1}(x) + \frac{2n}{x}\sin^2(\tfrac{1}{2}n\pi).$$

The first few are

$$\begin{aligned} O_0(x) &= \frac{1}{x} \\ O_1(x) &= \frac{1}{x^2} \\ O_2(x) &= \frac{1}{x} + \frac{4}{x^3}. \end{aligned}$$

see also SCHLÄFLI POLYNOMIAL

References

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 196, 1993.

Neumann Series (Bessel Function)

A series of the form

$$\sum_{n=0}^{\infty} a_n J_{\nu+n}(z), \tag{1}$$

where ν is a REAL and $J_{\nu+n}(z)$ is a BESSEL FUNCTION OF THE FIRST KIND. Special cases are

$$z^\nu = 2^\nu\Gamma(\tfrac{1}{2}\nu+1)\sum_{n=0}^{\infty}\frac{(\frac{1}{2}z)^{\nu/2+n}}{n!}J_{\nu/2+n}(z), \tag{2}$$

where $\Gamma(z)$ is the GAMMA FUNCTION, and

$$\sum_{n=0}^{\infty} b_n z^{\nu+n} = \sum_{n=0}^{\infty} a_n\left(\tfrac{1}{2}z\right)^{(\nu+n)/2} J_{(\nu+n)/2}(z), \tag{3}$$

where

$$a_n \equiv \sum_{m=0}^{\lfloor n/2\rfloor} \frac{2^{\nu+n-2m}\Gamma(\frac{1}{2}\nu+\frac{1}{2}n-m+1)}{m!} b_{n-2m}, \tag{4}$$

and $\lfloor x\rfloor$ is the FLOOR FUNCTION.

see also KAPTEYN SERIES

References

Watson, G. N. *A Treatise on the Theory of Bessel Functions, 2nd ed.* Cambridge, England: Cambridge University Press, 1966.

Neumann Series (Integral Equation)

A FREDHOLM INTEGRAL EQUATION OF THE SECOND KIND

$$\phi(x) = f(x) + \int_a^b K(x,t)\phi(t)\,dt \tag{1}$$

may be solved as follows. Take

$$\phi_0(x) \equiv f(x) \tag{2}$$

$$\phi_1(x) = f(x) + \lambda\int_a^b K(x,t)f(t)\,dt \tag{3}$$

$$\phi_2(x) = f(x) + \lambda\int_a^b K(x,t_1)f(t_1)\,dt_1 + \lambda^2\int_a^b\int_a^b K(x,t_1)K(t_1,t_2)f(t_2)\,dt_2\,dt_1 \tag{4}$$

$$\phi_n(x) = \sum_{i=0}^{n}\lambda^i u_i(x), \tag{5}$$

where

$$u_0(x) = f(x) \tag{6}$$

$$u_1(x) = \int_a^b K(x,t)f(t_1)\,dt_1 \tag{7}$$

$$u_2(x) = \int_a^b\int_a^b K(x,t_1)K(t_1,t_2)f(t_2)\,dt_2\,dt_1 \tag{8}$$

$$u_n(x) = \int_a^b\int_a^b\int_a^b K(x,t_1)K(t_1,t_2)\cdots \times K(t_{n-1},t_n)f(t_n)\,dt_n\cdots dt_1. \tag{9}$$

The Neumann series solution is then

$$\phi(x) = \lim_{n\to\infty} \phi_n(x) = \lim_{n\to\infty} \sum_{i=0}^{n} \lambda^i u_i(x). \qquad (10)$$

References

Arfken, G. "Neumann Series, Separable (Degenerate) Kernels." §16.3 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 879–890, 1985.

Neusis Construction

A geometric construction, also called a VERGING CONSTRUCTION, which allows the classical GEOMETRIC CONSTRUCTION rules to be bent in order to permit sliding of a marked RULER. Using a Neusis construction, CUBE DUPLICATION and angle TRISECTION are soluble. Conway and Guy (1996) give Neusis constructions for the 7-, 9-, and 13-gons which are based on angle TRISECTION.

see also CUBE DUPLICATION, GEOMETRIC CONSTRUCTION, MASCHERONI CONSTRUCTION, RULER, TRISECTION

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 194–200, 1996.

Neville's Algorithm

An interpolation ALGORITHM which proceeds by first fitting a POLYNOMIAL P_k of degree 0 through the points (x_k, y_k) for $k = 0, \ldots, n$, i.e., $P_k = y_k$. A second iteration is then performed in which P_{12} is fit through pairs of points, yielding P_{12}, P_{23}, ... The procedure is repeated, generating a "pyramid" of approximations until the final result is reached

$$\begin{array}{lllll} P_1 & & & & \\ & P_{12} & & & \\ P_2 & & P_{123} & & \\ & P_{23} & & P_{1234}. & \\ P_3 & & P_{234} & & \\ & P_{34} & & & \\ P_4 & & & & \end{array}$$

The final result is

$$P_{i(i+1)\cdots(i+m)} = \frac{(x - x_{i+m})P_{i(i+1)\cdots(i+m-1)}}{x_i - x_{i+m}} + \frac{(x_i - x)P_{(i+1)(i+2)\cdots(i+m)}}{x_i - x_{i+m}}.$$

see also BULIRSCH-STOER ALGORITHM

Neville Theta Function

The functions

$$\vartheta_s(u) = \frac{H(u)}{H'(0)} \qquad (1)$$

$$\vartheta_d(u) = \frac{\Theta(u+K)}{\Theta(k)} \qquad (2)$$

$$\vartheta_c(u) = \frac{H(u)}{H(K)} \qquad (3)$$

$$\vartheta_n(u) = \frac{\Theta(u)}{\Theta(0)}, \qquad (4)$$

where H and Θ are the JACOBI THETA FUNCTIONS and $K(u)$ is the complete ELLIPTIC INTEGRAL OF THE FIRST KIND.

see also JACOBI THETA FUNCTION, THETA FUNCTION

Newcomb's Paradox

A paradox in DECISION THEORY. Given two boxes, B1 which contains \$1000 and B2 which contains either nothing or a million dollars, you may pick either B2 or both. However, at some time before the choice is made, an omniscient Being has predicted what your decision will be and filled B2 with a million dollars if he expects you to take it, or with nothing if he expects you to take both.

see also ALIAS' PARADOX

References

Gardner, M. *The Unexpected Hanging and Other Mathematical Diversions.* Chicago, IL: Chicago University Press, 1991.
Gardner, M. "Newcomb's Paradox." Ch. 13 in *Knotted Doughnuts and Other Mathematical Entertainments.* New York: W. H. Freeman, 1986.
Nozick, R. "Reflections on Newcomb's Paradox." Ch. 14 in Gardner, M. *Knotted Doughnuts and Other Mathematical Entertainments.* New York: W. H. Freeman, 1986.

Newman-Conway Sequence

The sequence 1, 1, 2, 2, 3, 4, 4, 4, 5, 6, 7, 7, ... (Sloane's A004001) defined by the recurrence $P(1) = P(2) = 1$,

$$P(n) = P(P(n-1)) + P(n - P(n-1)).$$

It satisfies

$$P(2^k) = 2^{k-1}$$

and

$$P(2n) \leq 2P(n).$$

References

Bloom, D. M. "Newman-Conway Sequence." Solution to Problem 1459. *Math. Mag.* **68**, 400–401, 1995.
Sloane, N. J. A. Sequence A004001/M0276 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Newton's Backward Difference Formula

$$f_p = f_0 + p\nabla_0 + \tfrac{1}{2!}p(p+1)\nabla_0^2 + \tfrac{1}{3!}p(p+1)(p+2)\nabla_0^3 + \ldots,$$

for $p \in [0,1]$, where ∇ is the BACKWARD DIFFERENCE.

see also NEWTON'S FORWARD DIFFERENCE FORMULA

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 433, 1987.

Newton-Cotes Formulas

The Newton-Cotes formulas are an extremely useful and straightforward family of NUMERICAL INTEGRATION techniques.

To integrate a function $f(x)$ over some interval $[a, b]$, divide it into n equal parts such that $f_n = f(x_n)$ and $h \equiv (b-a)/n$. Then find POLYNOMIALS which approximate the tabulated function, and integrate them to approximate the AREA under the curve. To find the fitting POLYNOMIALS, use LAGRANGE INTERPOLATING POLYNOMIALS. The resulting formulas are called Newton-Cotes formulas, or QUADRATURE FORMULAS.

Newton-Cotes formulas may be "closed" if the interval $[x_1, x_n]$ is included in the fit, "open" if the points $[x_2, x_{n-1}]$ are used, or a variation of these two. If the formula uses n points (closed or open), the COEFFICIENTS of terms sum to $n-1$.

If the function $f(x)$ is given explicitly instead of simply being tabulated at the values x_i, the best numerical method of integration is called GAUSSIAN QUADRATURE. By picking the intervals at which to sample the function, this procedure produces more accurate approximations (but is significantly more complicated to implement).

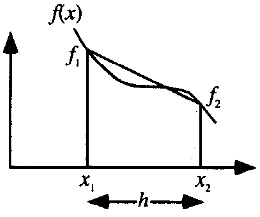

The 2-point closed Newton-Cotes formula is called the TRAPEZOIDAL RULE because it approximates the area under a curve by a TRAPEZOID with horizontal base and sloped top (connecting the endpoints x_1 and x_2). If the first point is x_1, then the other endpoint will be located at

$$x_2 = x_1 + h, \tag{1}$$

and the LAGRANGE INTERPOLATING POLYNOMIAL through the points (x_1, f_1) and (x_2, f_2) is

$$\begin{aligned} P_2(x) &= \frac{x-x_2}{x_1-x_2} f_1 + \frac{x-x_1}{x_2-x_1} f_2 \\ &= \frac{x-x_1-h}{-h} f_1 + \frac{x-x_1}{h} f_2 \\ &= \frac{x}{h}(f_2-f_1) + \left(f_1 + \frac{x_1}{h} f_1 - \frac{x_1}{h} f_2\right). \end{aligned} \tag{2}$$

Integrating over the interval (i.e., finding the area of the trapezoid) then gives

$$\begin{aligned} \int_{x_1}^{x_2} f(x)\,dx &= \int_{x_1}^{x_1+h} P_2(x)\,dx \\ &= \frac{1}{2h}(f_2-f_1)[x^2]_{x_1}^{x_2} \\ &\quad + \left(f_1 + \frac{x_1}{h} f_1 - \frac{x_1}{h} f_2\right)[x]_{x_1}^{x_2} \\ &= \frac{1}{2h}(f_2-f_1)(x_2+x_1)(x_2-x_1) \\ &\quad + (x_2-x_1)\left(f_1 + \frac{x_1}{h} f_1 - \frac{x_1}{h} f_2\right) \\ &= \tfrac{1}{2}(f_2-f_1)(2x_1+h) + f_1 h + x_1(f_1-f_2) \\ &= x_1(f_2-f_1) + \tfrac{1}{2}h(f_2-f_1) + hf_1 - x_1(f_2-f_1) \\ &= \tfrac{1}{2}h(f_1+f_2) - \tfrac{1}{2}h^3 f''(\xi). \end{aligned} \tag{3}$$

This is the trapezoidal rule, with the final term giving the amount of error (which, since $x_1 \le \xi \le x_2$, is no worse than the maximum value of $f''(\xi)$ in this range).

The 3-point rule is known as SIMPSON'S RULE. The ABSCISSAS are

$$x_2 = x_1 + h \tag{4}$$

$$x_3 = x_1 + 2h \tag{5}$$

and the LAGRANGE INTERPOLATING POLYNOMIAL is

$$\begin{aligned} P_3(x) &= \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)} f_1 \\ &\quad + \frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)} f_2 + \frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)} f_3 \\ &= \frac{x^2 - x(x_2+x_3) + x_2x_3}{h(2h)} f_1 \\ &\quad + \frac{x^2 - x(x_1+x_3) + x_1x_3}{h(-h)} f_2 + \frac{x^2 - x(x_1+x_2) + x_1x_2}{2h(h)} f_3 \\ &= \frac{1}{h^2}\{x^2(\tfrac{1}{2}f_1 - f_2 - \tfrac{1}{2}f_3) \\ &\quad + x[-\tfrac{1}{2}(2x_1+3h)f_1 + (2x_1+2h)f_2 - \tfrac{1}{2}(2x_1+h)] \\ &\quad + [\tfrac{1}{2}(x_1+h)(x_1+2h)f_1 - x_1(x_1+2h)f_2 + \tfrac{1}{2}x_1(x_1+h)f_3]\}. \end{aligned} \tag{6}$$

Integrating and simplifying gives

$$\begin{aligned} \int_{x_1}^{x_3} f(x)\,dx &= \int_{x_1}^{x_1+2h} P_3(x)\,dx \\ &= \tfrac{1}{3}h(f_1+4f_2+f_3) - \tfrac{1}{90}h^5 f^{(4)}(\xi). \end{aligned} \tag{7}$$

The 4-point closed rule is SIMPSON'S 3/8 RULE,

$$\int_{x_1}^{x_4} f(x)\,dx = \tfrac{3}{8}h(f_1+3f_2+3f_3+f_4) - \tfrac{3}{80}h^5 f^{(4)}(\xi). \tag{8}$$

The 5-point closed rule is Bode's Rule,

$$\int_{x_1}^{x_5} f(x)\,dx = \tfrac{2}{45}h(7f_1 + 32f_2 + 12f_3 + 32f_4 + 7f_5) - \tfrac{8}{945}h^7 f^{(6)}(\xi) \quad (9)$$

(Abramowitz and Stegun 1972, p. 886). Higher order rules include the 6-point

$$\int_{x_1}^{x_6} f(x)\,dx = \tfrac{5}{288}h(19f_1 + 75f_2 + 50f_3 + 50f_4 + 75f_5 + 19f_6) - \tfrac{275}{12096}h^7 f^{(6)}(\xi), \quad (10)$$

7-point

$$\int_{x_1}^{x_7} f(x)\,dx = \tfrac{1}{140}h(41f_1 + 216f_2 + 27f_3 + 272f_4 + 27f_5 + 216f_6 + 41f_7) - \tfrac{9}{1400}h^9 f^{(8)}(\xi), \quad (11)$$

8-point

$$\int_{x_1}^{x_8} f(x)\,dx = \tfrac{7}{17280}h(751f_1 + 3577f_2 + 1323f_2 + 2989f_3 + 2989f_5 + 1323f_6 + 3577f_7 + 751f_8) - \tfrac{8183}{518400}h^9 f^{(8)}(\xi), \quad (12)$$

9-point

$$\int_{x_1}^{x_9} f(x)\,dx = \tfrac{4}{14175}h(989f_1 + 5888f_2 - 928f_3 + 10496f_4 - 4540f_5 + 10496f_6 - 928f_7 + 5888f_8 + 989f_9) - \tfrac{2368}{467775}h^{11} f^{(10)}(\xi), \quad (13)$$

10-point

$$\int_{x_1}^{x_{10}} f(x)\,dx = \tfrac{9}{89600}h[2857(f_1 + f_{10}) + 15741(f_2 + f_9) + 1080(f_3 + f_8 + 19344(f_4 + f_7) + 5788(f_5 + f_6)] - \tfrac{173}{14620}h^{11} f^{(10)}(\xi), \quad (14)$$

and 11-point

$$\int_{x_1}^{x_{11}} f(x)\,dx = \tfrac{5}{299376}h[16067(f_1 + f_{11}) + 106300(f_2 + f_{10}) - 48525(f_3 + f_9) + 272400(f_4 + f_8) - 260550(f_5 + f_7) + 427368f_6] - \tfrac{1346350}{326918592}h^{13} f^{(12)}(\xi) \quad (15)$$

rules.

Closed "extended" rules use multiple copies of lower order closed rules to build up higher order rules. By appropriately tailoring this process, rules with particularly nice properties can be constructed. For n tabulated points, using the Trapezoidal Rule $(n-1)$ times and adding the results gives

$$\begin{aligned}\int_{x_1}^{x_n} f(x)\,dx &= \left(\int_{x_1}^{x_2} + \int_{x_2}^{x_3} + \ldots + \int_{x_{n-1}}^{x_n}\right) f(x)\,dx \\ &= \tfrac{1}{2}h[(f_1 + f_2) + (f_2 + f_3) + \ldots + (f_{n-2} + f_{n-1}) \\ &\quad +(f_{n-1}+f_n)] = h(\tfrac{1}{2}f_1+f_2+f_3+\ldots+f_{n-2}+f_{n-1}+\tfrac{1}{2}f_n) \\ &\quad - \tfrac{1}{12}nh^3 f''(\xi). \quad (16)\end{aligned}$$

Using a series of refinements on the extended Trapezoidal Rule gives the method known as Romberg Integration. A 3-point extended rule for Odd n is

$$\begin{aligned}\int_{x_1}^{x_n} f(x)\,dx &= h[(\tfrac{1}{3}f_1 + \tfrac{4}{3}f_2 + \tfrac{1}{3}f_3) + (\tfrac{1}{3}f_3 + \tfrac{4}{3}f_4 + \tfrac{1}{3}f_5) \\ &\quad + \ldots + (\tfrac{1}{3}f_{n-4} + \tfrac{4}{3}f_{n-3} + \tfrac{1}{3}f_{n-2}) \\ &\quad +(\tfrac{1}{3}f_{n-2} + \tfrac{4}{3}f_{n-1} + \tfrac{1}{3}f_n)] \\ &= \tfrac{1}{3}h(f_1 + 4f_2 + 2f_3 + 4f_4 + 2f_5 + \ldots + 4f_{n-1} + f_n) \\ &\quad - \frac{n-1}{2}\tfrac{1}{90}h^5 f^{(4)}(\xi). \quad (17)\end{aligned}$$

Applying Simpson's 3/8 Rule, then Simpson's Rule (3-point) twice, and adding gives

$$\begin{aligned}&\left[\int_{x_1}^{x_4} + \int_{x_4}^{x_6} + \int_{x_6}^{x_8}\right] f(x)\,dx \\ &\quad = h[(\tfrac{3}{8}f_1 + \tfrac{9}{8}f_2 + \tfrac{9}{8}f_3 + \tfrac{3}{8}f_4) \\ &\qquad +(\tfrac{1}{3}f_4 + \tfrac{4}{3}f_5 + \tfrac{1}{3}f_6) + (\tfrac{1}{3}f_6 + \tfrac{4}{3}f_7 + \tfrac{1}{3}f_8)] \\ &= h[\tfrac{3}{8}f_1 + \tfrac{9}{8}f_2 + \tfrac{9}{8}f_3 + (\tfrac{3}{8} + \tfrac{1}{3})f_4 + \tfrac{4}{3}f_5 \\ &\qquad +(\tfrac{1}{3} + \tfrac{1}{3})f_6 + \tfrac{4}{3}f_7 + \tfrac{1}{3}f_8] \\ &= h(\tfrac{3}{8}f_1 + \tfrac{9}{8}f_2 + \tfrac{9}{8}f_3 + \tfrac{17}{24}f_4 \\ &\qquad + \tfrac{4}{3}f_5 + \tfrac{2}{3}f_6 + \tfrac{4}{3}f_7 + \tfrac{1}{3}f_8). \quad (18)\end{aligned}$$

Taking the next Simpson's 3/8 step then gives

$$\int_{x_8}^{x_{11}} f(x)\,dx = h(\tfrac{3}{8}f_8 + \tfrac{9}{8}f_9 + \tfrac{9}{8}f_{10} + \tfrac{3}{8}f_{11}). \quad (19)$$

Combining with the previous result gives

$$\begin{aligned}\int_{x_1}^{x_{11}} f(x)\,dx &= h[\tfrac{3}{8}f_1 + \tfrac{9}{8}f_2 + \tfrac{9}{8}f_3 + \tfrac{17}{24}f_4 + \tfrac{4}{3}f_5 \\ &\quad + \tfrac{2}{3}f_6 + \tfrac{4}{3}f_7 + (\tfrac{1}{3} + \tfrac{3}{8})f_8 + \tfrac{9}{8}f_9 + \tfrac{9}{8}f_{10} + \tfrac{3}{8}f_{11}] \\ &= h(\tfrac{3}{8}f_1 + \tfrac{9}{8}f_2 + \tfrac{9}{8}f_3 + \tfrac{17}{24}f_4 + \tfrac{4}{3}f_5 + \tfrac{2}{3}f_6 + \tfrac{4}{3}f_7 \\ &\quad + \tfrac{17}{24}f_8 + \tfrac{9}{8}f_9 + \tfrac{9}{8}f_{10} + \tfrac{3}{8}f_{11}), \quad (20)\end{aligned}$$

where terms up to f_{10} have now been completely determined. Continuing gives

$$\begin{aligned}&h(\tfrac{3}{8}f_1 + \tfrac{9}{8}f_2 + \tfrac{9}{8}f_3 + \tfrac{17}{24}f_4 + \tfrac{4}{3}f_5 + \tfrac{2}{3}f_6 + \ldots \\ &+ \tfrac{2}{3}f_{n-5} + \tfrac{4}{3}f_{n-4} + \tfrac{17}{24}f_{n-3} + \tfrac{9}{8}f_{n-2} + \tfrac{9}{8}f_{n-1} + \tfrac{3}{8}f_n). \quad (21)\end{aligned}$$

Now average with the 3-point result

$$h(\tfrac{1}{3}f_1+\tfrac{4}{3}f_2+\tfrac{2}{3}f_3+\tfrac{4}{3}f_4+\tfrac{2}{3}f_5+\tfrac{4}{3}f_{n-1}+\tfrac{1}{3}f_n) \quad (22)$$

to obtain

$$h[\tfrac{17}{48}f_1+\tfrac{59}{48}f_2+\tfrac{43}{48}f_4+\tfrac{49}{48}f_4+(f_5+f_6+\ldots+f_{n-5}+f_{n-4}) \\ +\tfrac{49}{48}f_{n-3}+\tfrac{43}{38}f_{n-2}+\tfrac{59}{48}f_{n-1}+\tfrac{17}{48}f_n]+\mathcal{O}(n^{-4}). \quad (23)$$

Note that all the middle terms now have unity Coefficients. Similarly, combining a 4-point with the (2+4)-point rule gives

$$h(\tfrac{5}{12}f_1+\tfrac{13}{12}f_2+f_3+f_4+\ldots+f_{n-3}+f_{n-2}+\tfrac{13}{12}f_{n-1}+\tfrac{5}{12}) \\ +\mathcal{O}(n^{-3}). \quad (24)$$

Other Newton-Cotes rules occasionally encountered include Durand's Rule

$$\int_{x_1}^{x_n} f(x)\,dx \\ = h(\tfrac{2}{5}f_1+\tfrac{11}{10}f_2+f_3+\ldots+f_{n-2}+\tfrac{11}{10}f_{n-1}+\tfrac{2}{5}f_n) \quad (25)$$

(Beyer 1987), Hardy's Rule

$$\int_{x_0-3h}^{x_0+3h} f(x)\,dx = \tfrac{1}{100}h(28f_{-3}+162f_{-2}+22f_0+162f_2 \\ +28f_3)+\tfrac{9}{1400}h^7[2f^{(4)}(\xi_2)-h^2f^{(8)}(\xi_1)], \quad (26)$$

and Weddle's Rule

$$\int_{x_1}^{x_{6n}} f(x)\,dx = \tfrac{3}{10}h(f_1 \\ +5f_2+f_3+6f_4+5f_5+f_6+\ldots+5f_{6n-1}+f_{6n}) \quad (27)$$

(Beyer 1987).

The open Newton-Cotes rules use points outside the integration interval, yielding the 1-point

$$\int_{x_0}^{x_2} f(x)\,dx = 2hf_1, \quad (28)$$

2-point

$$\int_{x_0}^{x_3} f(x)\,dx = \int_{x_1-h}^{x_1+2h} P_2(x)\,dx \\ = \frac{1}{2h}(f_2-f_1)[x^2]_{x_1-h}^{x_1+2h}+\left(f_1+\frac{x_1}{h}f_1-\frac{x_1}{h}f_2\right)[x]_{x_1-h}^{x_1+2h} \\ = \tfrac{3}{2}h(f_1+f_2)+\tfrac{1}{4}h^3f''(\xi), \quad (29)$$

3-point

$$\int_{x_0}^{x_4} f(x)\,dx = \tfrac{4}{3}h(2f_1-f_2+2f_3)+\tfrac{28}{90}h^5f^{(4)}(\xi), \quad (30)$$

4-point

$$\int_{x_0}^{x_5} f(x)\,dx = \tfrac{5}{24}h(11f_1+f_2+f_3+11f_4)+\tfrac{95}{144}h^5f^{(4)}(\xi), \quad (31)$$

5-point

$$\int_{x_0}^{x_6} f(x)\,dx = \tfrac{6}{20}h(11f_1 \\ -14f_2+26f_3-14f_4+11f_5)-\tfrac{41}{140}h^7f^{(6)}(\xi), \quad (32)$$

6-point

$$\int_{x_0}^{x_7} f(x)\,dx = \tfrac{7}{1440}h(611f_1-453f_2+562f_3+562f_4 \\ -453f_5+611f_6)-\tfrac{5257}{8640}h^7f^{(6)}(\xi), \quad (33)$$

and 7-point

$$\int_{x_0}^{x_8} f(x)\,dx = \tfrac{8}{945}h(460f_1-954f_2+2196f_3-2459f_4 \\ +2196f_5-954f_6+460f_7)-\tfrac{3956}{14175}h^9f^{(8)}(\xi) \quad (34)$$

rules.

A 2-point open extended formula is

$$\int_{x_1}^{x_n} f(x)\,dx = h[(\tfrac{1}{2}f_1+f_2+\ldots+f_{n-1}+\tfrac{1}{2}f_n) \\ +\tfrac{1}{24}(-f_0+f_2+f_{n-1}+f_{n+1})]+\frac{11(n+1)}{720}h^5f^{(4)}(\xi). \quad (35)$$

Single interval extrapolative rules estimate the integral in an interval based on the points around it. An example of such a rule is

$$hf_1+\mathcal{O}(h^2f') \quad (36)$$

$$\tfrac{1}{2}h(3f_1-f_2)+\mathcal{O}(h^3f'') \quad (37)$$

$$\tfrac{1}{12}h(23f_1-16f_2+5f_3)+\mathcal{O}(h^4f^{(3)}) \quad (38)$$

$$\tfrac{1}{24}h(55f_1-59f_2+37f_3-9f_4)+\mathcal{O}(h^5f^{(4)}). \quad (39)$$

see also Bode's Rule, Difference Equation, Durand's Rule, Finite Difference, Gaussian Quadrature, Hardy's Rule, Lagrange Interpolating Polynomial, Numerical Integration, Simpson's Rule, Simpson's 3/8 Rule, Trapezoidal Rule, Weddle's Rule

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Integration." §25.4 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 885–887, 1972.

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 127, 1987.

Hildebrand, F. B. *Introduction to Numerical Analysis.* New York: McGraw-Hill, pp. 160–161, 1956.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Classical Formulas for Equally Spaced Abscissas." §4.1 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 124–130, 1992.

Newton's Diverging Parabolas

Curves with CARTESIAN equation

$$ay^2 = x(x^2 - 2bx + c)$$

with $a > 0$. The above equation represents the third class of Newton's classification of CUBIC CURVES, which Newton divided into five species depending on the ROOTS of the cubic in x on the right-hand side of the equation. Newton described these cases as having the following characteristics:

1. "All the ROOTS are REAL and unequal. Then the Figure is a diverging Parabola of the Form of a Bell, with an Oval at its Vertex.
2. Two of the ROOTS are equal. A PARABOLA will be formed, either Nodated by touching an Oval, or Punctate, by having the Oval infinitely small.
3. The three ROOTS are equal. This is the NEILIAN PARABOLA, commonly called SEMI-CUBICAL.
4. Only one REAL ROOT. If two of the ROOTS are impossible, there will be a Pure PARABOLA of a Bell-like Form"

(MacTutor Archive).

References

MacTutor History of Mathematics Archive. "Newton's Diverging Parabolas." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Newtons.html`.

Newton's Divided Difference Interpolation Formula

Let

$$\pi_n(x) \equiv \prod_{i=1}^{n}(x - x_n), \qquad (1)$$

then

$$f(x) = f_0 + \sum_{k=1}^{n} x_{k-1}(x)[x_0, x_1, \ldots, x_k] + R_n, \qquad (2)$$

where $[x_1, \ldots]$ is a DIVIDED DIFFERENCE, and the remainder is

$$R_n(x) = \pi_n(x)[x_0, \ldots, x_n, x] = \pi_n(x)\frac{f^{(n+1)}(\xi)}{(n+1)} \qquad (3)$$

for $x_0 < \xi < x_n$.

see also DIVIDED DIFFERENCE, FINITE DIFFERENCE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 880, 1972.

Hildebrand, F. B. *Introduction to Numerical Analysis.* New York: McGraw-Hill, pp.43–44 and 62–63, 1956.

Newton's Forward Difference Formula

A FINITE DIFFERENCE identity giving an interpolated value between tabulated points $\{f_p\}$ in terms of the first value f_0 and the POWERS of the FORWARD DIFFERENCE Δ. For $a \in [0, 1]$, the formula states

$$f_a = f_0 + a\Delta + \tfrac{1}{2!}a(a-1)\Delta^2 + \tfrac{1}{3!}a(a-1)(a-2)\Delta^3 + \ldots.$$

When written in the form

$$f(x+a) = \sum_{n=0}^{\infty} \frac{(a)_n \Delta^n f(x)}{n!}$$

with $(a)_n$ the POCHHAMMER SYMBOL, the formula looks suspiciously like a finite analog of a TAYLOR SERIES expansion. This correspondence was one of the motivating forces for the development of UMBRAL CALCULUS.

The DERIVATIVE of Newton's forward difference formula gives MARKOFF'S FORMULAS.

see also FINITE DIFFERENCE, MARKOFF'S FORMULAS, NEWTON'S BACKWARD DIFFERENCE FORMULA, NEWTON'S DIVIDED DIFFERENCE INTERPOLATION FORMULA

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 880, 1972.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 432, 1987.

Newton's Formulas

Let a TRIANGLE have side lengths a, b, and c with opposite angles A, B, and C. Then

$$\frac{b+c}{a} = \frac{\cos[\frac{1}{2}(B-C)]}{\sin(\frac{1}{2}A)}$$

$$\frac{c+a}{b} = \frac{\cos[\frac{1}{2}(C-A)]}{\sin(\frac{1}{2}B)}$$

$$\frac{a+b}{c} = \frac{\cos[\frac{1}{2}(A-B)]}{\sin(\frac{1}{2}C)}.$$

see also MOLLWEIDE'S FORMULAS, TRIANGLE

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 146, 1987.

Newton's Identities

see also NEWTON'S RELATIONS

Newton's Iteration

An algorithm for the SQUARE ROOT of a number r quadratically as $\lim_{n\to\infty} x_n$,

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{r}{x_n}\right),$$

where $x_0 = 1$. The first few approximants to $\sqrt{n}$ are given by

$$1, \tfrac{1}{2}(1+n), \frac{1+6n+n^2}{4(n+1)}, \frac{1+26n+70n^2+28n^3+n^4}{8(1+n)(1+6n+n^2)}, \ldots.$$

For $\sqrt{2}$, this gives the convergents as 1, 3/2, 17/12, 577/408, 665857/470832, ...

see also SQUARE ROOT

Newton's Method

A ROOT-finding ALGORITHM which uses the first few terms of the TAYLOR SERIES in the vicinity of a suspected ROOT to zero in on the root. The TAYLOR SERIES of a function $f(x)$ about the point $x + \epsilon$ is given by

$$f(x+\epsilon) = f(x) + f'(x)\epsilon + \tfrac{1}{2}f''(x)\epsilon^2 + \ldots. \qquad (1)$$

Keeping terms only to first order,

$$f(x+\epsilon) \approx f(x) + f'(x)\epsilon. \qquad (2)$$

This expression can be used to estimate the amount of offset ϵ needed to land closer to the root starting from an initial guess x_0. Setting $f(x_0 + \epsilon) = 0$ and solving (2) for ϵ gives

$$\epsilon_0 = -\frac{f(x_0)}{f'(x_0)}, \qquad (3)$$

which is the first-order adjustment to the ROOT's position. By letting $x_1 = x_0 + \epsilon_0$, calculating a new ϵ_1, and so on, the process can be repeated until it converges to a root.

Unfortunately, this procedure can be unstable near a horizontal ASYMPTOTE or a LOCAL MINIMUM. However, with a good initial choice of the ROOT's position, the algorithm can by applied iteratively to obtain

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \qquad (4)$$

for $n = 1, 2, 3, \ldots$.

The error ϵ_{n+1} after the $(n+1)$st iteration is given by

$$\begin{aligned}\epsilon_{n+1} &= \epsilon_n + (x_{n+1} - x_n)\\ &= \epsilon_n - \frac{f(x_n)}{f'(x_n)}. \qquad (5)\end{aligned}$$

But

$$\begin{aligned}f(x_n) &= f(x) + f'(x)\epsilon_n + \tfrac{1}{2}f''(x){\epsilon_n}^2 + \ldots\\ &= f'(x)\epsilon_n + \tfrac{1}{2}f''(x){\epsilon_n}^2 + \ldots \qquad (6)\\ f'(x_n) &= f'(x) + f''(x)\epsilon_n + \ldots, \qquad (7)\end{aligned}$$

so

$$\begin{aligned}\frac{f(x_n)}{f'(x_x)} &= \frac{f'(x)\epsilon_n + \frac{1}{2}f''(x){\epsilon_n}^2 + \ldots}{f'(x) + f''(x)\epsilon_n + \ldots}\\ &\approx \frac{f'(x)\epsilon + \frac{1}{2}f''(x){\epsilon_n}^2}{f'(x) + f''(x)\epsilon_n} = \epsilon_n + \frac{f''(x)}{2f'(x)}{\epsilon_n}^2, \qquad (8)\end{aligned}$$

and (5) becomes

$$\epsilon_{n+1} = \epsilon_n - \left[\epsilon_n + \frac{f''(x)}{2f'(x)}{\epsilon_n}^2\right] = -\frac{f''(x)}{2f'(x)}{\epsilon_n}^2. \qquad (9)$$

Therefore, when the method converges, it does so quadratically.

A FRACTAL is obtained by applying Newton's method to finding a ROOT of $z^n - 1 = 0$ (Mandelbrot 1983, Gleick 1988, Peitgen and Saupe 1988, Press *et al.* 1992, Dickau 1997). Iterating for a starting point z_0 gives

$$z_{i+1} = z_i - \frac{{z_i}^n - 1}{n{z_i}^{n-1}}. \qquad (10)$$

Since this is an nth order POLYNOMIAL, there are n ROOTS to which the algorithm can converge.

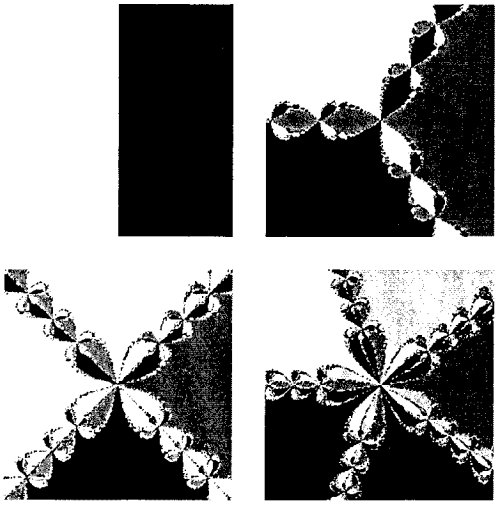

Coloring the BASIN OF ATTRACTION (the set of initial points z_0 which converge to the same ROOT) for each ROOT a different color then gives the above plots, corresponding to $n = 2$, 3, 4, and 5.

see also HALLEY'S IRRATIONAL FORMULA, HALLEY'S METHOD, HOUSEHOLDER'S METHOD, LAGUERRE'S METHOD

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 18, 1972.

Acton, F. S. Ch. 2 in *Numerical Methods That Work.* Washington, DC: Math. Assoc. Amer., 1990.

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 963–964, 1985.

Dickau, R. M. "Basins of Attraction for $z^5 = 1$ Using Newton's Method in the Complex Plane." `http://forum.swarthmore.edu/advanced/robertd/newtons.html`.

Dickau, R. M. "Variations on Newton's Method." `http://forum.swarthmore.edu/advanced/robertd/newnewton.html`.

Dickau, R. M. "Compilation of Iterative and List Operations." *Mathematica J.* **7**, 14–15, 1997.

Gleick, J. *Chaos: Making a New Science.* New York: Penguin Books, plate 6 (following pp. 114) and p. 220, 1988.

Householder, A. S. *Principles of Numerical Analysis.*ew York: McGraw-Hill, pp. 135–138, 1953.

Mandelbrot, B. B. *The Fractal Geometry of Nature.* San Francisco, CA: W. H. Freeman, 1983.

Ortega, J. M. and Rheinboldt, W. C. *Iterative Solution of Nonlinear Equations in Several Variables.* New York: Academic Press, 1970.

Peitgen, H.-O. and Saupe, D. *The Science of Fractal Images.* New York: Springer-Verlag, 1988.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Newton-Raphson Method Using Derivatives" and "Newton-Raphson Methods for Nonlinear Systems of Equations." §9.4 and 9.6 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 355–362 and 372–375, 1992.

Ralston, A. and Rabinowitz, P. §8.4 in *A First Course in Numerical Analysis, 2nd ed.* New York: McGraw-Hill, 1978.

Newton Number

see KISSING NUMBER

Newton's Parallelogram

Approximates the possible values of y in terms of x if

$$\sum_{i,j=0}^{n} a_{ij} x^i y^i = 0.$$

Newton-Raphson Fractal

see NEWTON'S METHOD

Newton-Raphson Method

see NEWTON'S METHOD

Newton's Relations

Let s_i be the sum of the products of distinct ROOTS r_j of the POLYNOMIAL equation of degree n

$$a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 = 0, \tag{1}$$

where the roots are taken i at a time (i.e., s_i is defined as the ELEMENTARY SYMMETRIC FUNCTION $\Pi_i(r_1, \ldots, r_n)$) s_i is defined for $i = 1, \ldots, n$. For example, the first few values of s_i are

$$s_1 = r_1 + r_2 + r_3 + r_4 + \ldots \tag{2}$$

$$s_2 = r_1 r_2 + r_1 r_3 + r_1 r_4 + r_2 r_3 + \ldots \tag{3}$$

$$s_3 = r_1 r_2 r_3 + r_1 r_2 r_4 + r_2 r_3 r_4 + \ldots, \tag{4}$$

and so on. Then

$$s_i = (-1)^i \frac{a_{n-i}}{a_n}. \tag{5}$$

This can be seen for a second DEGREE POLYNOMIAL by multiplying out,

$$\begin{aligned} a_2 x^2 + a_1 x + a_0 &= a_2 (x - r_1)(x - r_2) \\ &= a_2 [x^2 - (r_1 + r_2) x + r_1 r_2], \end{aligned} \tag{6}$$

so

$$s_1 = \sum_{i=1}^{2} r_i = r_1 + r_2 = -\frac{a_1}{a_2} \tag{7}$$

$$s_2 = \sum_{\substack{i,j=1 \\ i \neq j}}^{2} r_i r_j = r_1 r_2 = \frac{a_0}{a_2}, \tag{8}$$

and for a third DEGREE POLYNOMIAL,

$$\begin{aligned} &a_3 x^3 + a_2 x^2 + a_1 x + a_0 = a_3 (x - r_1)(x - r_2)(x - r_3) \\ &= a_3 [x^3 - (r_1 + r_2 + r_3) x^2 + (r_1 r_2 + r_1 r_3 + r_2 r_3) x - r_1 r_2 r_3], \end{aligned} \tag{9}$$

so

$$s_1 = \sum_{i=1}^{3} r_i = -\frac{a_2}{a_3} \tag{10}$$

$$s_2 = \sum_{\substack{i,j \\ i \neq j}}^{3} r_i r_j = r_1 r_2 + r_1 r_3 + r_2 r_3 = \frac{a_1}{a_3} \tag{11}$$

$$s_3 = \sum_{\substack{i,j,k \\ i \neq j \neq k}}^{3} r_i r_j r_k = r_1 r_2 r_3 = -\frac{a_0}{a_3}. \tag{12}$$

see also ELEMENTARY SYMMETRIC FUNCTION

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, pp. 1–2, 1959.

Newton's Theorem

If each of two nonparallel transversals with nonminimal directions meets a given curve in finite points only, then the ratio of products of the distances from the two sets of intersections to the intersection of the lines is independent of the position of the latter point.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 189, 1959.

Newtonian Form

see NEWTON'S DIVIDED DIFFERENCE INTERPOLATION FORMULA

Next Prime

The next prime function $NP(n)$ gives the smallest PRIME larger than n. The function can be given explicitly as

$$NP(n) = p_{1+\pi(n)},$$

where p_i is the ith PRIME and $\pi(n)$ is the PRIME COUNTING FUNCTION. For $n = 1, 2, \ldots$ the values are 2, 3, 5, 5, 7, 7, 11, 11, 11, 11, 13, 13, 17, 17, 17, 17, 19, ... (Sloane's A007918).

see also FORTUNATE PRIME, PRIME COUNTING FUNCTION, PRIME NUMBER

References
Sloane, N. J. A. Sequence A007918 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Nexus Number

A FIGURATE NUMBER built up of the nexus of cells less than n steps away from a given cell. In k-D, the $(n+1)$th nexus number is given by

$$N_{n+1}(k) = \sum_{i=0}^{k} \binom{k}{i} n^i,$$

where $\binom{n}{n}$ is a BINOMIAL COEFFICIENT. The first few k-dimensional nexus numbers are given in the table below.

k	N_{n+1}	name
0	1	unit
1	$1+2n$	odd number
2	$1+3n+3n^2$	hex number
3	$1+4n+6n^2+4n^3$	rhombic dodecahedral number

see also HEX NUMBER, ODD NUMBER, RHOMBIC DODECAHEDRAL NUMBER

References
Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 53–54, 1996.

Neyman-Pearson Lemma

If there exists a critical region C of size α and a NONNEGATIVE constant k such that

$$\frac{\prod_{i=1}^{n} f(x_i|\theta_1)}{\prod_{i=1}^{n} f(x_i|\theta_0)} \geq k$$

for points in C and

$$\frac{\prod_{i=1}^{n} f(x_i|\theta_1)}{\prod_{i=1}^{n} f(x_i|\theta_0)} \leq k$$

for points not in C, then C is a best critical region of size α.

References
Hoel, P. G.; Port, S. C.; and Stone, C. J. "Testing Hypotheses." Ch. 3 in *Introduction to Statistical Theory.* New York: Houghton Mifflin, pp. 56–67, 1971.

Nicholson's Formula

Let $J_\nu(z)$ be a BESSEL FUNCTION OF THE FIRST KIND, $Y_\nu(z)$ a BESSEL FUNCTION OF THE SECOND KIND, and $K_\nu(z)$ a MODIFIED BESSEL FUNCTION OF THE FIRST KIND. Also let $\Re[z] > 0$. Then

$$J_\nu^2(z) + Y_\nu^2(z) = \frac{8}{\pi^2} \int_0^\infty K_0(2z \sinh t) \cos(2\nu t)\, dt.$$

see also DIXON-FERRAR FORMULA, WATSON'S FORMULA

References
Gradshteyn, I. S. and Ryzhik, I. M. Eqn. 6.664.4 in *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 727, 1979.
Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1476, 1980.

Nicomachus's Theorem

The nth CUBIC NUMBER n^3 is a sum of n consecutive ODD NUMBERS, for example

$$\begin{aligned} 1^3 &= 1 \\ 2^3 &= 3+5 \\ 3^3 &= 7+9+11 \\ 4^3 &= 13+15+17+19, \end{aligned}$$

etc. This identity follows from

$$\sum_{i=1}^{n} [n(n-1) - 1 + 2i] = n^3.$$

It also follows from this fact that

$$\sum_{k=1}^{n} k^3 = \left(\sum_{k=1}^{n} k \right)^2.$$

see also ODD NUMBER THEOREM

Nicomedes' Conchoid

see CONCHOID OF NICOMEDES

Nielsen-Ramanujan Constants

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

N. Nielsen (1909) and Ramanujan (Berndt 1985) considered the integrals

$$a_k = \int_1^2 \frac{(\ln x)^k}{x-1}\, dx. \tag{1}$$

They found the values for $k = 1$ and 2. The general constants for $k > 3$ were found by V. Adamchik (Finch)

$$a_p = p!\zeta(p+1) - \frac{p(\ln 2)^{p+1}}{p+1} - p!\sum_{k=0}^{p-1} \frac{\mathrm{Li}_{p+1-k}(\frac{1}{2})(\ln 2)^k}{k!}, \tag{2}$$

where $\zeta(z)$ is the RIEMANN ZETA FUNCTION and $\mathrm{Li}_n(x)$ is the POLYLOGARITHM. The first few values are

$$a_1 = \tfrac{1}{2}\zeta(2) = \tfrac{1}{12}\pi^2 \tag{3}$$

$$a_2 = \tfrac{1}{4}\zeta(3) \tag{4}$$

$$a_3 = \tfrac{1}{15}\pi^4 + \tfrac{1}{4}\pi^2(\ln 2)^2 - \tfrac{1}{4}(\ln 2)^4 - 6\,\mathrm{Li}_4(\tfrac{1}{2}) - \tfrac{21}{4}\ln 2\zeta(3) \tag{5}$$

$$a_4 = \tfrac{2}{3}\pi^2(\ln 2)^3 - \tfrac{4}{5}(\ln 2)^5 - 24\ln 2\,\mathrm{Li}_4(\tfrac{1}{2}) - 24\,\mathrm{Li}_5(\tfrac{1}{2}) - \tfrac{21}{2}(\ln 2)^2\zeta(3) + 24\zeta(5). \tag{6}$$

see also POLYLOGARITHM, RIEMANN ZETA FUNCTION

References

Berndt, B. C. *Ramanujan's Notebooks, Part I.* New York: Springer-Verlag, 1985.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/nielram/nielram.html`.

Flajolet, P. and Salvy, B. "Euler Sums and Contour Integral Representation." Submitted to *Experim. Math* 1997. `http://pauillac.inria.fr/algo/flajolet/Publications/publist.html`.

Nielsen's Spiral

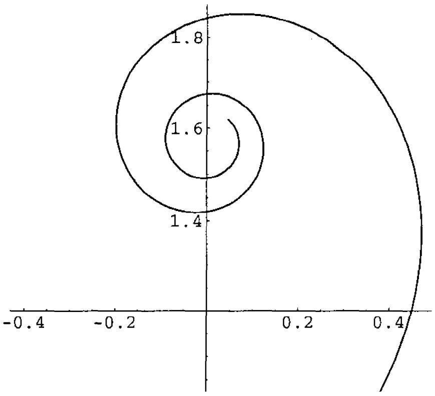

The SPIRAL with parametric equations

$$x(t) = a\,\mathrm{ci}(t) \tag{1}$$

$$y(t) = a\,\mathrm{si}(t), \tag{2}$$

where $\mathrm{ci}(t)$ is the COSINE INTEGRAL and $\mathrm{si}(t)$ is the SINE INTEGRAL. The CESÀRO EQUATION is

$$\kappa = \frac{e^{s/a}}{a}. \tag{3}$$

see also CORNU SPIRAL, COSINE INTEGRAL, SINE INTEGRAL

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 119, 1993.

Nil Geometry

The GEOMETRY of the LIE GROUP consisting of REAL MATRICES of the form

$$\begin{bmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{bmatrix},$$

i.e., the HEISENBERG GROUP.

see also HEISENBERG GROUP, LIE GROUP, THURSTON'S GEOMETRIZATION CONJECTURE

Nilmanifold

Let N be a NILPOTENT, connected, SIMPLY CONNECTED LIE GROUP, and let D be a discrete SUBGROUP of N with compact right QUOTIENT SPACE. Then N/D is called a nilmanifold.

Nilpotent Element

An element B of a RING is nilpotent if there exists a POSITIVE INTEGER k for which $B^k = 0$.

see also ENGEL'S THEOREM

Nilpotent Group

A GROUP G for which the chain of groups

$$I = Z_0 \subseteq Z_1 \subseteq \ldots \subseteq Z_n$$

with Z_{k+1}/Z_k (equal to the CENTER of G/Z_k) terminates finitely with $G = Z_n$ is called a nilpotent group.

see also CENTER (GROUP), NILPOTENT LIE GROUP

Nilpotent Lie Group

A LIE GROUP which has a simply connected covering group HOMEOMORPHIC to $\mathbb{R}^n$. The prototype is any connected closed subgroup of upper triangular COMPLEX matrices with 1s on the diagonal. The HEISENBERG GROUP is such a group.

References

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Nilpotent Matrix

A SQUARE MATRIX whose EIGENVALUES are all 0. A related definition is a SQUARE MATRIX M such that M^n is 0 for some POSITIVE integral POWER.

see also EIGENVALUE, SQUARE MATRIX

Nim

A game, also called TACTIX, which is played by the following rules. Given one or more piles (NIM-HEAPS), players alternate by taking all or some of the counters in a single heap. The player taking the last counter or stack of counters is the winner. Nim-like games are also called TAKE-AWAY GAMES and DISJUNCTIVE GAMES.

If optimal strategies are used, the winner can be determined from any intermediate position by its associated NIM-VALUE.

see also MISÈRE FORM, NIM-VALUE, WYTHOFF'S GAME

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 36–38, 1987.

Bogomolny, A. "The Game of Nim." `http://www.cut-the-knot.com/bottom_nim.html`.

Bouton, C. L. "Nim, A Game with a Complete Mathematical Theory." *Ann. Math. Princeton* **3**, 35–39, 1901–1902.

Gardner, M. "Nim and Hackenbush." Ch. 14 in *Wheels, Life, and other Mathematical Amusements.* New York: W. H. Freeman, 1983.

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Oxford University Press, pp. 117–120, 1990.

Kraitchik, M. "Nim." §3.12.2 in *Mathematical Recreations.* New York: W. W. Norton, pp. 86–88, 1942.

Nim-Heap

A pile of counters in a game of NIM.

Nim-Sum

see NIM-VALUE

Nim-Value

Every position of every IMPARTIAL GAME has a nim-value, making it equivalent to a NIM-HEAP. To find the nim-value (also called the SPRAGUE-GRUNDY NUMBER), take the MEX of the nim-values of the possible moves. The nim-value can also be found by writing the number of counters in each heap in binary, adding without carrying, and replacing the digits with their values mod 2. If the nim-value is 0, the position is SAFE; otherwise, it is UNSAFE. With two heaps, safe positions are (x, x) where $x \in [1, 7]$. With three heaps, (1, 2, 3), (1, 4, 5), (1, 6, 7), (2, 4, 6), (2, 5, 7), and (3, 4, 7).

see also GRUNDY'S GAME, IMPARTIAL GAME, MEX, NIM, SAFE, UNSAFE

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 36–38, 1987.

Grundy, P. M. "Mathematics and Games." *Eureka* **2**, 6–8, 1939.

Sprague, R. "Über mathematische Kampfspiele." *Tôhoku J. Math.* **41**, 438–444, 1936.

Nine-Point Center

The center F (or N) of the NINE-POINT CIRCLE. It has TRIANGLE CENTER FUNCTION

$$\begin{aligned}\alpha &= \cos(B-C) = \cos A + 2\cos B\cos C\\ &= bc[a^2b^2 + a^2c^2 + (b^2-c^2)^2],\end{aligned}$$

and is the MIDPOINT of the line between the CIRCUMCENTER C and ORTHOCENTER H. It lies on the EULER LINE.

see also EULER LINE, NINE-POINT CIRCLE, NINE-POINT CONIC

References

Carr, G. S. *Formulas and Theorems in Pure Mathematics, 2nd ed.* New York: Chelsea, p. 624, 1970.

Dixon, R. *Mathographics.* New York: Dover, pp. 57–58, 1991.

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Kimberling, C. "Nine-Point Center." `http://www.evansville.edu/~ck6/tcenters/class/npcenter.html`.

Nine-Point Circle

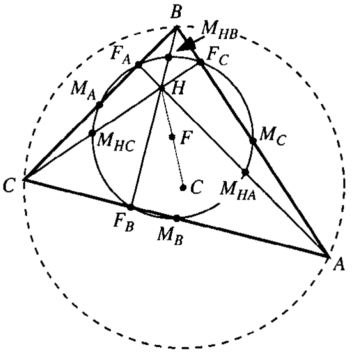

The CIRCLE, also called EULER'S CIRCLE and the FEUERBACH CIRCLE, which passes through the feet of the PERPENDICULAR F_A, F_B, and F_C dropped from the VERTICES of any TRIANGLE ΔABC on the sides opposite them. Euler showed in 1765 that it also passes through the MIDPOINTS M_A, M_B, M_C of the sides of ΔABC.

By FEUERBACH'S THEOREM, the nine-point circle also passes through the MIDPOINTS M_{HA}, M_{HB}, M_{HC} of the segments which join the VERTICES and the ORTHOCENTER H. These three triples of points make nine in all, giving the circle its name. The center F of the nine-point circle is called the NINE-POINT CENTER.

The RADIUS of the nine-point circle is $R/2$, where R is the CIRCUMRADIUS. The center of KIEPERT'S HYPERBOLA lies on the nine-point circle. The nine-point circle bisects any line from the ORTHOCENTER to a point on the CIRCUMCIRCLE. The nine-point circle of the INCENTER and EXCENTERS of a TRIANGLE is the CIRCUMCIRCLE.

The sum of the powers of the VERTICES with regard to the nine-point circle is

$$\tfrac{1}{4}({a_1}^2 + {a_2}^2 + {a_3}^2).$$

Also,

$$\overline{FA_1}^2 + \overline{FA_2}^2 + \overline{FA_3}^2 + \overline{FH}^2 = 3R^2,$$

where F is the NINE-POINT CENTER, A_i are the VERTICES, H is the ORTHOCENTER, and R is the CIRCUMRADIUS. All triangles inscribed in a given CIRCLE and having the same ORTHOCENTER have the same nine-point circle.

see also COMPLETE QUADRILATERAL, EIGHT-POINT CIRCLE THEOREM, FEUERBACH'S THEOREM, FONTENÉ THEOREMS, GRIFFITHS' THEOREM, NINE-POINT CENTER, NINE-POINT CONIC, ORTHOCENTRIC SYSTEM

References

Altshiller-Court, N. *College Geometry: A Second Course in Plane Geometry for Colleges and Normal Schools, 2nd ed., rev. enl.* New York: Barnes and Noble, pp. 93–97, 1952.

Brand, L. "The Eight-Point Circle and the Nine-Point Circle." *Amer. Math. Monthly* **51**, 84–85, 1944.

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* New York: Random House, pp. 20–22, 1967.

Dörrie, H. "The Feuerbach Circle." §28 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 142–144, 1965.

Gardner, M. *Mathematical Carnival: A New Round-Up of Tantalizers and Puzzles from Scientific American.* New York: Vintage Books, p. 59, 1977.

Guggenbuhl, L. "Karl Wilhelm Feuerbach, Mathematician." Appendix to *Circles: A Mathematical View, rev. ed.* Washington, DC: Math. Assoc. Amer., pp. 89–100, 1995.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 165 and 195–212, 1929.

Lange, J. *Geschichte des Feuerbach'schen Kreises.* Berlin, 1894.

Mackay, J. S. "History of the Nine-Point Circle." *Proc. Edinburgh Math. Soc.* **11**, 19–61, 1892.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 119–120, 1990.

Pedoe, D. *Circles: A Mathematical View, rev. ed.* Washington, DC: Math. Assoc. Amer., pp. 1–4, 1995.

Nine-Point Conic

A CONIC SECTION on which the MIDPOINTS of the sides of any COMPLETE QUADRANGLE lie. The three diagonal points also lie on this conic.

see also COMPLETE QUADRANGLE, CONIC SECTION, NINE-POINT CIRCLE

Nint

see NEAREST INTEGER FUNCTION

Nint Zeta Function

Let

$$S_N(s) = \sum_{n=1}^{\infty} [(n^{1/N})]^{-s}, \tag{1}$$

where $[x]$ denotes NINT, the INTEGER closest to x. For $s > 3$,

$$S_2(s) = 2\zeta(s-1) \tag{2}$$

$$S_3(s) = 3\zeta(s-2) + 4^{-s}\zeta(s) \tag{3}$$

$$S_4(s) = 4\zeta(s-3) + \zeta(s-1). \tag{4}$$

$S_N(n)$ is a POLYNOMIAL in π whose COEFFICIENTS are ALGEBRAIC NUMBERS whenever $n - N$ is ODD. The first few values are given explicitly by

$$S_3(4) = \frac{\pi^2}{2} + \frac{\pi^4}{23046} \tag{5}$$

$$S_5(6) = \frac{5\pi^2}{6} + \frac{\pi^4}{36} + \frac{\pi^6}{4^{12}} \times \left(\frac{1}{945} - \frac{170912 + 49928\sqrt{2}}{25} \sqrt{1 - \sqrt{\tfrac{1}{2}}} \right) \tag{6}$$

$$S_6(7) = \pi^2 + \frac{\pi^4}{18} + \frac{\pi^6}{2520} + \frac{246013 + 353664\sqrt{2}}{45} \frac{\pi^7}{2^{27}}. \tag{7}$$

References

Borwein, J. M.; Hsu, L. C.; Mabry, R.; Neu, K.; Roppert, J.; Tyler, D. B.; and de Weger, B. M. M. "Nearest Integer Zeta-Functions." *Amer. Math. Monthly* **101**, 579–580, 1994.

Nirenberg's Conjecture

If the GAUSS MAP of a complete minimal surface omits a NEIGHBORHOOD of the SPHERE, then the surface is a PLANE. This was proven by Osserman (1959). Xavier (1981) subsequently generalized the result as follows. If the GAUSS MAP of a complete MINIMAL SURFACE omits ≥ 7 points, then the surface is a PLANE.

see also GAUSS MAP, MINIMAL SURFACE, NEIGHBORHOOD

References

do Carmo, M. P. *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, p. 42, 1986.

Osserman, R. "Proof of a Conjecture of Nirenberg." *Comm. Pure Appl. Math.* **12**, 229–232, 1959.

Xavier, F. "The Gauss Map of a Complete Nonflat Minimal Surface Cannot Omit 7 Points on the Sphere." *Ann. Math.* **113**, 211–214, 1981.

Niven's Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Given a POSITIVE INTEGER $m > 1$, let its PRIME FACTORIZATION be written

$$m = p_1{}^{a_1} p_2{}^{a_2} p_3{}^{a_3} \cdots p_k{}^{a_k}. \tag{1}$$

Define the functions h and H by $h(1) = 1$, $H(1) = 1$, and

$$h(m) = \min(a_1, a_2, \ldots, a_k) \tag{2}$$

$$H(m) = \max(a_1, a_2, \ldots, a_k). \tag{3}$$

Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} h(m) = 1 \tag{4}$$

$$\lim_{n\to\infty}\frac{\sum_{m=1}^{n}h(m)-n}{\sqrt{n}}=\frac{\zeta(\frac{3}{2})}{\zeta(3)}, \tag{5}$$

where $\zeta(z)$ is the RIEMANN ZETA FUNCTION (Niven 1969). Niven (1969) also proved that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^{n}H(m)=C, \tag{6}$$

where

$$C=1+\left\{\sum_{j=2}^{\infty}\left[1-\frac{1}{\zeta(j)}\right]\right\}=1.705221\ldots \tag{7}$$

(Sloane's A033150).

The CONTINUED FRACTION of Niven's constant is 1, 1, 2, 2, 1, 1, 4, 1, 1, 3, 4, 4, 8, 4, 1, ... (Sloane's A033151). The positions at which the digits 1, 2, ... first occur in the CONTINUED FRACTION are 1, 3, 10, 7, 47, 41, 34, 13, 140, 252, 20, ... (Sloane's A033152). The sequence of largest terms in the CONTINUED FRACTION is 1, 2, 4, 8, 11, 14, 29, 372, 559, ... (Sloane's A0033153), which occur at positions 1, 3, 7, 13, 20, 35, 51, 68, 96, ... (Sloane's A033154).

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/niven/niven.html`.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 41, 1983.

Niven, I. "Averages of Exponents in Factoring Integers." *Proc. Amer. Math. Soc.* **22**, 356–360, 1969.

Plouffe, S. "The Niven Constant." `http://www.lacim.uqam.ca/piDATA/niven.txt`.

Niven Number

see HARSHAD NUMBER

Nobbs Points

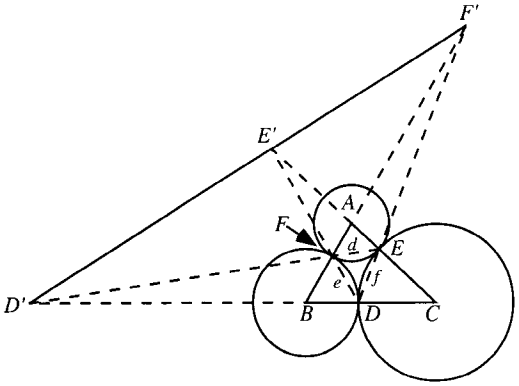

Given a TRIANGLE ΔABC, construct the CONTACT TRIANGLE ΔDEF. Then the Nobbs points are the three points D', E', and F' from which ΔABC and ΔDEF are PERSPECTIVE, as illustrated above. The Nobbs points are COLLINEAR and fall along the GERGONNE LINE.

see also COLLINEAR, CONTACT TRIANGLE, EVANS POINT, FLETCHER POINT, GERGONNE LINE, PERSPECTIVE TRIANGLES

References

Oldknow, A. "The Euler-Gergonne-Soddy Triangle of a Triangle." *Amer. Math. Monthly* **103**, 319–329, 1996.

Noble Number

A noble number is defined as an IRRATIONAL NUMBER which has a CONTINUED FRACTION which becomes an infinite sequence of 1s at some point,

$$\nu\equiv[a_1,a_2,\ldots,a_n,\bar{1}].$$

The prototype is the GOLDEN RATIO ϕ whose CONTINUED FRACTION is composed *entirely* of 1s, $[\bar{1}]$. Any noble number can written as

$$\nu=\frac{A_n+\phi A_{n-1}}{B_n+\phi B_{n+1}},$$

where A_k and B_k are the NUMERATOR and DENOMINATOR of the kth CONVERGENT of $[a_1,a_2,\ldots,a_n]$. The noble numbers are a SUBFIELD of $\mathbb{Q}(\sqrt{5})$.

see also NEAR NOBLE NUMBER

References

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, p. 236, 1979.

Schroeder, M. "Noble and Near Noble Numbers." In *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise.* New York: W. H. Freeman, pp. 392–394, 1991.

Node (Algebraic Curve)

see ORDINARY DOUBLE POINT

Node (Fixed Point)

A FIXED POINT for which the STABILITY MATRIX has both EIGENVALUES of the same sign (i.e., both are POSITIVE or both are NEGATIVE). If $\lambda_1<\lambda_2<0$, then the node is called STABLE; if $\lambda_1>\lambda_2>0$, then the node is called an UNSTABLE NODE.

see also STABLE NODE, UNSTABLE NODE

Node (Graph)

Synonym for the VERTICES of a GRAPH, i.e., the points connected by EDGES.

see also ACNODE, CRUNODE, TACNODE

Noether's Fundamental Theorem

If two curves ϕ and ψ of MULTIPLICITIES $r_i\neq 0$ and $s_i\neq 0$ have only ordinary points or ordinary singular points and CUSPS in common, then every curve which has at least MULTIPLICITY

$$r_i+s_i-1$$

at every point (distinct or infinitely near) can be written

$$f \equiv \phi\psi' + \psi\phi' = 0,$$

where the curves ϕ' and ψ' have MULTIPLICITIES at least $r_i - 1$ and $s_i - 1$.

References
Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, pp. 29–30, 1959.

Noether-Lasker Theorem

Let M be a finitely generated MODULE over a commutative NOETHERIAN RING R. Then there exists a finite set $\{N_i | 1 \leq i \leq l\}$ of submodules of M such that

1. $\cap_{i=1}^{l} N_i = 0$ and $\cap_{i \neq i_0} N_i$ is not contained in N_{i_0} for all $1 \leq i_0 \leq l$.
2. Each quotient M/N_i is primary for some prime P_i.
3. The P_i are all distinct for $1 \leq i \leq l$.
4. Uniqueness of the primary component N_i is equivalent to the statement that P_i does not contain P_j for any $j \neq i$.

Noether's Transformation Theorem

Any irreducible curve may be carried by a factorable CREMONA TRANSFORMATION into one with none but ordinary singular points.

References
Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 207, 1959.

Noetherian Module

A MODULE M is Noetherian if every submodule is finitely generated.

see also NOETHERIAN RING

Noetherian Ring

An abstract commutative RING satisfying the abstract chain condition.

see also LOCAL RING, NOETHER-LASKER THEOREM

Noise

An error which is superimposed on top of a true signal. Noise may be random or systematic. Noise can be greatly reduced by transmitting signals digitally instead of in analog form because each piece of information is allowed only discrete values which are spaced farther apart than the contribution due to noise.

CODING THEORY studies how to encode information efficiently, and ERROR-CORRECTING CODES devise methods for transmitting and reconstructing information in the presence of noise.

see also ERROR

References
Davenport, W. B. and Root, W. L. *An Introduction to the Theory of Random Signals and Noise.* New York: IEEE Press, 1987.
McDonough, R. N. and Whalen, A. D. *Detection of Signals in Noise, 2nd ed.* Orlando, FL: Academic Press, 1995.
Pierce, J. R. *Symbols, Signals and Noise: The Nature and Process of Communication.* New York: Harper & Row, 1961.
Vainshtein, L. A. and Zubakov, V. D. *Extraction of Signals from Noise.* New York: Dover, 1970.
van der Ziel, A. *Noise: Sources, Characterization, Measurement.* New York: Prentice-Hall, 1954.
van der Ziel, A. *Noise in Measurement.* New York: Wiley, 1976.
Wax, N. *Selected Papers on Noise and Stochastic Processes.* New York: Dover, 1954.

Noise Sphere

A mapping of RANDOM NUMBER TRIPLES to points in SPHERICAL COORDINATES,

$$\begin{aligned}\theta &= 2\pi X_n \\ \phi &= \pi X_{n+1} \\ r &= \sqrt{X_{n+2}}.\end{aligned}$$

The graphical result can yield unexpected structure which indicates correlations between triples and therefore that the numbers are not truly RANDOM.

References
Pickover, C. A. *Computers and the Imagination.* New York: St. Martin's Press, 1991.
Pickover, C. A. "Computers, Randomness, Mind, and Infinity." Ch. 31 in *Keys to Infinity.* New York: W. H. Freeman, pp. 233–247, 1995.
Richards, T. "Graphical Representation of Pseudorandom Sequences." *Computers and Graphics* **13**, 261–262, 1989.

Nolid

An assemblage of faces forming a POLYHEDRON of zero VOLUME (Holden 1991, p. 124).

see also ACOPTIC POLYHEDRON

References
Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, 1991.

Nome

Given a THETA FUNCTION, the nome is defined as

$$q(m) \equiv e^{\pi\tau i} = e^{-\pi K(1-m)/K(m)} \equiv e^{-\pi K'(m)/K(m)}, \quad (1)$$

where $K(k)$ is the complete ELLIPTIC INTEGRAL OF THE FIRST KIND, and m is the PARAMETER.

$$\vartheta_i(z,q) \equiv \vartheta(z|\tau) \quad (2)$$

$$\vartheta_i \equiv \vartheta(0,q). \quad (3)$$

Solving the nome for the PARAMETER m gives

$$m(q) = \frac{\vartheta_2{}^4(0,q)}{\vartheta_3{}^4(0,q)}, \tag{4}$$

where $\vartheta_i(z,q)$ is a THETA FUNCTION.

see also AMPLITUDE, CHARACTERISTIC (ELLIPTIC INTEGRAL), ELLIPTIC INTEGRAL, MODULAR ANGLE, MODULUS (ELLIPTIC INTEGRAL), PARAMETER

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 591, 1972.

Nomogram

A graphical plot which can be used for solving certain types of equations.

References

Iyanaga, S. and Kawada, Y. (Eds.). "Nomograms." §282 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 891–893, 1980.

Menzel, D. (Ed.). *Fundamental Formulas of Physics, Vol. 1.* New York: Dover, p. 141, 1960.

Nonagon

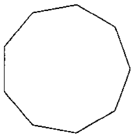

The unconstructible regular POLYGON with nine sides and SCHLÄFLI SYMBOL $\{9\}$. It is sometimes called an ENNEAGON.

Although the regular nonagon is not a CONSTRUCTIBLE POLYGON, Dixon (1991) gives several close approximations. While the ANGLE subtended by a side is $360°/9 = 40°$, Dixon gives constructions containing angles of $\tan^{-1}(5/6) \approx 39.8805571°$ and $2\tan^{-1}((\sqrt{3}-1)/2) \approx 40.207818°$.

Madachy (1979) illustrates how to construct a nonagon by folding and knotting a strip of paper.

see also NONAGRAM, TRIGONOMETRY VALUES—$\pi/9$

References

Dixon, R. *Mathographics.* New York: Dover, pp. 40–44, 1991.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 60–61, 1979.

Nonagonal Number

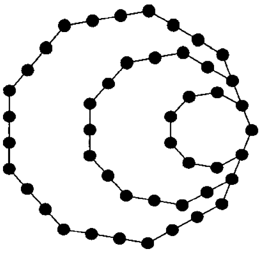

A FIGURATE NUMBER of the form $n(7n-5)/2$, also called an ENNEAGONAL NUMBER. The first few are 1, 9, 24, 46, 75, 111, 154, 204, ... (Sloane's A001106).

References

Sloane, N. J. A. Sequence A001106/M4604 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Nonagram

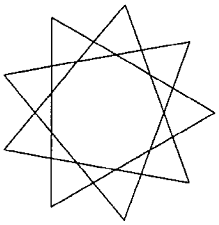

A STAR POLYGON composed of three EQUILATERAL TRIANGLES rotated at angles 0°, 40°, and 80°. It has been called the STAR OF GOLIATH by analogy with the STAR OF DAVID (HEXAGRAM).

see also HEXAGRAM, NONAGON, TRIGONOMETRY VALUES—$\pi/9$

Nonassociative Algebra

An ALGEBRA which does not satisfy

$$a(bc) = (ab)c$$

is called a nonassociative algebra. Bott and Milnor (1958) proved that the only nonassociative DIVISION ALGEBRAS are for $n = 1$, 2, 4, and 8. Each gives rise to an ALGEBRA with particularly useful physical applications (which, however, is not itself necessarily nonassociative), and these four cases correspond to REAL NUMBERS, COMPLEX NUMBERS, QUATERNIONS, and CAYLEY NUMBERS, respectively.

see also ALGEBRA, CAYLEY NUMBER, COMPLEX NUMBER, DIVISION ALGEBRA, QUATERNION, REAL NUMBER

References

Bott, R. and Milnor, J. "On the Parallelizability of the Spheres." *Bull. Amer. Math. Soc.* **64**, 87–89, 1958.

Nonassociative Product

The number of nonassociative n-products with k elements preceding the rightmost left parameter is

$$\begin{aligned} F(n,k) &= F(n-1,k) + F(n-1,k-1) \\ &= \binom{n+k-2}{k} - \binom{n+k-1}{k-1}, \end{aligned}$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT. The number of n-products in a nonassociative algebra is

$$F(n) = \sum_{j=0}^{n-2} F(n,j) = \frac{(2n-2)!}{n!(n-1)!}.$$

References

Niven, I. M. *Mathematics of Choice: Or, How to Count Without Counting.* Washington, DC: Math. Assoc. Amer., pp. 140–152, 1965.

Nonaveraging Sequence

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

An infinite sequence of POSITIVE INTEGERS

$$1 \leq a_1 < a_2 < a_3 < \ldots$$

is a nonaveraging sequence if it contains no three terms which are in an ARITHMETIC PROGRESSION, so that

$$a_i + a_j \neq 2a_k$$

for all distinct a_i, a_j, a_k. Wróblewski (1984) showed that

$$S(A) \equiv \sup_{\substack{\text{all nonaveraging} \\ \text{sequences}}} \sum_{k=1}^{\infty} \frac{1}{a_k} > 3.00849.$$

References

Behrend, F. "On Sets of Integers which Contain no Three Terms in an Arithmetic Progression." *Proc. Nat. Acad. Sci. USA* **32**, 331–332, 1946.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/erdos/erdos.html`.

Gerver, J. L. "The Sum of the Reciprocals of a Set of Integers with No Arithmetic Progression of k Terms." *Proc. Amer. Math. Soc.* **62**, 211–214, 1977.

Gerver, J. L. and Ramsey, L. "Sets of Integers with no Long Arithmetic Progressions Generated by the Greedy Algorithm." *Math. Comput.* **33**, 1353–1360, 1979.

Guy, R. K. "Nonaveraging Sets. Nondividing Sets." §C16 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 131–132, 1994.

Wróblewski, J. "A Nonaveraging Set of Integers with a Large Sum of Reciprocals." *Math. Comput.* **43**, 261–262, 1984.

Noncentral Distribution

see CHI-SQUARED DISTRIBUTION, F-DISTRIBUTION, STUDENT'S t-DISTRIBUTION

Noncommutative Group

A group whose elements do not commute. The simplest noncommutative GROUP is the DIHEDRAL GROUP D_3 of ORDER six.

see also COMMUTATIVE, FINITE GROUP—D_3

Nonconformal Mapping

Let γ be a path in $\mathbb{C}$, $w = f(z)$, and θ and ϕ be the tangents to the curves γ and $f(\gamma)$ at z_0 and w_0. If there is an N such that

$$f^{(N)}(z_0) \neq 0 \tag{1}$$

$$f^{(n)}(z_0) = 0 \tag{2}$$

for all $n < N$ (or, equivalently, if $f'(z)$ has a zero of order $N-1$), then

$$f(z) = f(z_0) + \frac{f^{(N)}(z_0)}{N!}(z-z_0)^N + \frac{f^{(N+1)}(z_0)}{(N+1)!}(z-z_0)^{N+1} + \ldots \tag{3}$$

$$f(z) - f(z_0) = (z-z_0)^N \left[\frac{f(N)(z_0)}{N!} + \frac{f^{(N+1)}(z_0)}{(N+1)!}(z-z_0) + \ldots\right], \tag{4}$$

so the ARGUMENT is

$$\arg[f(z) - f(z_0)] = N \arg(z-z_0) + \arg\left[\frac{f(N)(z_0)}{N!} + \frac{f^{(N+1)}(z_0)}{(N+1)!}(z-z_0) + \ldots\right]. \tag{5}$$

As $z \to z_0$, $\arg(z-z_0) \to \theta$ and $|\arg[f(z) - f(z_0)]| \to \phi$,

$$\phi = N\theta + \arg\left[\frac{f(N)(z_0)}{N!}\right] = N\theta + \arg[f(N)(z_0)]. \tag{6}$$

see also CONFORMAL TRANSFORMATION

Nonconstructive Proof

A PROOF which indirectly shows a mathematical object exists without providing a specific example or algorithm for producing an example.

see also PROOF

References

Courant, R. and Robbins, H. "The Indirect Method of Proof." §2.4.4 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 86–87, 1996.

Noncototient

A POSITIVE value of n for which $x - \phi(x) = n$ has no solution, where $\phi(x)$ is the TOTIENT FUNCTION. The first few are 10, 26, 34, 50, 52, ... (Sloane's A005278).

see also NONTOTIENT, TOTIENT FUNCTION

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 91, 1994.

Sloane, N. J. A. Sequence A005278/M4688 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Noncylindrical Ruled Surface

A RULED SURFACE parameterization $\mathbf{x}(u,v) = \mathbf{b}(u) + v\mathbf{g}(u)$ is called noncylindrical if $\mathbf{g} \times \mathbf{g}'$ is nowhere $\mathbf{0}$. A noncylindrical ruled surface always has a parameterization of the form

$$\mathbf{x}(u,v) = \boldsymbol{\sigma}(u) + v\boldsymbol{\delta}(u),$$

where $|\boldsymbol{\delta}| = 1$ and $\boldsymbol{\sigma}' \cdot \boldsymbol{\delta}' = 0$, where $\boldsymbol{\sigma}$ is called the STRICTION CURVE of $\mathbf{x}$ and $\boldsymbol{\delta}$ the DIRECTOR CURVE.

see also DISTRIBUTION PARAMETER, RULED SURFACE, STRICTION CURVE

References

Gray, A. "Noncylindrical Ruled Surfaces." §17.3 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 345–349, 1993.

Nondecreasing Function

A function $f(x)$ is said to be nondecreasing on an INTERVAL I if $f(b) \geq f(a)$ for all $b > a$, where $a, b \in I$. Conversely, a function $f(x)$ is said to be nonincreasing on an INTERVAL I if $f(b) \leq f(a)$ for all $b > a$ with $a, b \in I$.

see also DECREASING FUNCTION, NONINCREASING FUNCTION

Nondividing Set

A SET in which no element divides the SUM of any other.

References

Guy, R. K. "Nonaveraging Sets. Nondividing Sets." §C16 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 131–132, 1994.

Nonessential Singularity

see REGULAR SINGULAR POINT

Non-Euclidean Geometry

In 3 dimensions, there are three classes of constant curvature GEOMETRIES. All are based on the first four of EUCLID'S POSTULATES, but each uses its own version of the PARALLEL POSTULATE. The "flat" geometry of everyday intuition is called EUCLIDEAN GEOMETRY (or PARABOLIC GEOMETRY), and the non-Euclidean geometries are called HYPERBOLIC GEOMETRY (or LOBACHEVSKY-BOLYAI-GAUSS GEOMETRY) and ELLIPTIC GEOMETRY (or RIEMANNIAN GEOMETRY). It was not until 1868 that Beltrami proved that non-Euclidean geometries were as logically consistent as EUCLIDEAN GEOMETRY.

see also ABSOLUTE GEOMETRY, ELLIPTIC GEOMETRY, EUCLID'S POSTULATES, EUCLIDEAN GEOMETRY, HYPERBOLIC GEOMETRY, PARALLEL POSTULATE

References

Borsuk, K. *Foundations of Geometry: Euclidean and Bolyai-Lobachevskian Geometry. Projective Geometry.* Amsterdam, Netherlands: North-Holland, 1960.

Carslaw, H. S. *The Elements of Non-Euclidean Plane Geometry and Trigonometry.* London: Longmans, 1916.

Coxeter, H. S. M. *Non-Euclidean Geometry, 5th ed.* Toronto: University of Toronto Press, 1965.

Dunham, W. *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 53–60, 1990.

Iversen, B. *An Invitation to Hyperbolic Geometry.* Cambridge, England: Cambridge University Press, 1993.

Iyanaga, S. and Kawada, Y. (Eds.). "Non-Euclidean Geometry." §283 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 893–896, 1980.

Martin, G. E. *The Foundations of Geometry and the Non-Euclidean Plane.* New York: Springer-Verlag, 1975.

Pappas, T. "A Non-Euclidean World." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 90–92, 1989.

Ramsay, A. and Richtmeyer, R. D. *Introduction to Hyperbolic Geometry.* New York: Springer-Verlag, 1995.

Sommerville, D. Y. *The Elements of Non-Euclidean Geometry.* London: Bell, 1914.

Sommerville, D. Y. *Bibliography of Non-Euclidean Geometry, 2nd ed.* New York: Chelsea, 1960.

Sved, M. *Journey into Geometries.* Washington, DC: Math. Assoc. Amer., 1991.

Trudeau, R. J. *The Non-Euclidean Revolution.* Boston, MA: Birkhäuser, 1987.

Nonillion

In the American system, 10^{30}.

see also LARGE NUMBER

Nonincreasing Function

A function $f(x)$ is said to be nonincreasing on an INTERVAL I if $f(b) \leq f(a)$ for all $b > a$, where $a, b \in I$. Conversely, a function $f(x)$ is said to be nondecreasing on an INTERVAL I if $f(b) \geq f(a)$ for all $b > a$ with $a, b \in I$.

see also INCREASING FUNCTION, NONDECREASING FUNCTION

Nonlinear Least Squares Fitting

Given a function $f(x)$ of a variable x tabulated at m values $y_1 = f(x_1)$, ..., $y_m = f(x_m)$, assume the function is of known analytic form depending on n parameters $f(x; \lambda_1, \ldots, \lambda_n)$, and consider the overdetermined set of m equations

$$y_1 = f(x_1; \lambda_1, \lambda_2, \ldots, \lambda_n) \tag{1}$$

$$y_m = f(x_m; \lambda_1, \lambda_2, \ldots, \lambda_n). \tag{2}$$

We desire to solve these equations to obtain the values λ_1, ..., λ_n which best satisfy this system of equations. Pick an initial guess for the λ_i and then define

$$d\beta_i = y_i - f(x_i; \lambda_1, \ldots, \lambda_n). \tag{3}$$

Now obtain a linearized estimate for the changes $d\lambda_i$ needed to reduce $d\beta_i$ to 0,

$$d\beta_i = \sum_{j=1}^{n} \frac{\partial f}{\partial \lambda_j}\, d\lambda_j \bigg|_{x_j,\lambda} \tag{4}$$

for $i = 1, \ldots, n$. This can be written in component form as

$$d\beta_i = A_{ij}\, d\lambda_i, \tag{5}$$

where A is the $m \times n$ MATRIX

$$A_{ij} = \begin{bmatrix} \left.\frac{\partial f}{\partial \lambda_1}\right|_{x_1,\lambda} & \left.\frac{\partial f}{\partial \lambda_n}\right|_{x_1,\lambda} & \cdots \\ \left.\frac{\partial f}{\partial \lambda_2}\right|_{x_2,\lambda} & \left.\frac{\partial f}{\partial \lambda_2}\right|_{x_2,\lambda} & \cdots \\ \vdots & \vdots & \ddots \\ \left.\frac{\partial f}{\partial \lambda_1}\right|_{x_m,\lambda} & \left.\frac{\partial f}{\partial \lambda_n}\right|_{x_m,\lambda} & \cdots \end{bmatrix}. \tag{6}$$

In more concise MATRIX form,

$$d\boldsymbol{\beta} = \mathsf{A}\, d\lambda, \tag{7}$$

where $d\boldsymbol{\beta}$ and $d\lambda$ are m-VECTORS. Applying the MATRIX TRANSPOSE of A to both sides gives

$$\mathsf{A}^{\mathrm{T}}\, d\boldsymbol{\beta} = (\mathsf{A}^{\mathrm{T}}\mathsf{A})\, d\lambda. \tag{8}$$

Defining

$$\mathsf{a} \equiv \mathsf{A}^{\mathrm{T}}\mathsf{A} \tag{9}$$

$$\mathbf{b} \equiv \mathsf{A}^{\mathrm{T}}\, d\boldsymbol{\beta} \tag{10}$$

in terms of the known quantities A and $d\boldsymbol{\beta}$ then gives the MATRIX EQUATION

$$\mathsf{a} d\lambda = \mathbf{b}, \tag{11}$$

which can be solved for $d\lambda$ using standard matrix techniques such as GAUSSIAN ELIMINATION. This offset is then applied to λ and a new $d\boldsymbol{\beta}$ is calculated. By iteratively applying this procedure until the elements of $d\lambda$ become smaller than some prescribed limit, a solution is obtained. Note that the procedure may not converge very well for some functions and also that convergence is often greatly improved by picking initial values close to the best-fit value. The sum of square residuals is given by $R^2 = d\boldsymbol{\beta} \cdot d\boldsymbol{\beta}$ after the final iteration.

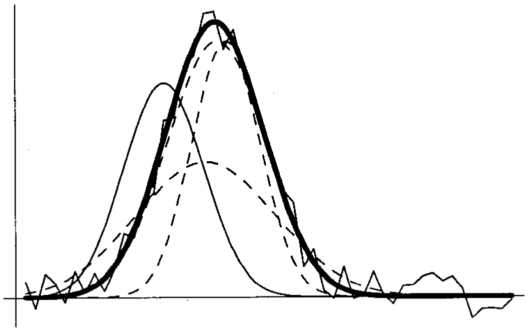

An example of a nonlinear least squares fit to a noisy GAUSSIAN FUNCTION

$$f(A, x_0, \sigma; x) = Ae^{-(x-x_0)^2/(2\sigma^2)} \tag{12}$$

is shown above, where the thin solid curve is the initial guess, the dotted curves are intermediate iterations, and the heavy solid curve is the fit to which the solution converges. The actual parameters are $(A, x_0, \sigma) = (1, 20, 5)$, the initial guess was (0.8, 15, 4), and the converged values are (1.03105, 20.1369, 4.86022), with $R^2 = 0.148461$. The PARTIAL DERIVATIVES used to construct the matrix A are

$$\frac{\partial f}{\partial A} = e^{-(x-x_0)^2/(2\sigma^2)} \tag{13}$$

$$\frac{\partial f}{\partial x_0} = \frac{A(x-x_0)}{\sigma^2} e^{-(x-x_0)^2/(2\sigma^2)} \tag{14}$$

$$\frac{\partial f}{\partial \sigma} = \frac{A(x-x_0)^2}{\sigma^3} e^{-(x-x_0)^2/(2\sigma^2)}. \tag{15}$$

The technique could obviously be generalized to multiple Gaussians, to include slopes, etc., although the convergence properties generally worsen as the number of free parameters is increased.

An analogous technique can be used to solve an overdetermined set of equations. This problem might, for example, arise when solving for the best-fit EULER ANGLES corresponding to a noisy ROTATION MATRIX, in which case there are three unknown angles, but nine correlated matrix elements. In such a case, write the n *different* functions as $f_i(\lambda_1, \ldots, \lambda_n)$ for $i = 1, \ldots, n$, call their actual values y_i, and define

$$\mathsf{A} = \begin{bmatrix} \left.\frac{\partial f_1}{\partial \lambda_1}\right|_{\lambda_i} & \left.\frac{\partial f_1}{\partial \lambda_2}\right|_{\lambda_i} & \cdots & \left.\frac{\partial f_1}{\partial \lambda_n}\right|_{\lambda_i} \\ \vdots & \vdots & \ddots & \vdots \\ \left.\frac{\partial f_m}{\partial \lambda_1}\right|_{\lambda_i} & \left.\frac{\partial f_m}{\partial \lambda_2}\right|_{\lambda_i} & \cdots & \left.\frac{\partial f_m}{\partial \lambda_n}\right|_{\lambda_i} \end{bmatrix}, \tag{16}$$

and

$$d\boldsymbol{\beta} = \mathbf{y} - f_i(\lambda_1, \ldots, \lambda_n), \tag{17}$$

where λ_i are the numerical values obtained after the ith iteration. Again, set up the equations as

$$\mathsf{A}\, d\lambda = d\boldsymbol{\beta}, \tag{18}$$

and proceed exactly as before.

see also LEAST SQUARES FITTING, LINEAR REGRESSION, MOORE-PENROSE GENERALIZED MATRIX INVERSE

Nonnegative

A quantity which is either 0 (ZERO) or POSITIVE, i.e., ≥ 0.

see also NEGATIVE, NONNEGATIVE INTEGER, NONPOSITIVE, NONZERO, POSITIVE, ZERO

Nonnegative Integer

see $\mathbb{Z}^*$

Nonnegative Partial Sum

The number of sequences with NONNEGATIVE partial sums which can be formed from n 1s and n -1s (Bailey 1996, Buraldi 1992) is given by the CATALAN NUMBERS. Bailey (1996) gives the number of NONNEGATIVE partial sums of n 1s and k -1s $a_1, a_2, \ldots, a_{n+k}$, so that

$$a_1 + a_2 + \ldots + a_i \geq 0 \quad (1)$$

for all $1 \leq i \leq n+k$. The closed form expression is

$$\left\{ \begin{matrix} n \\ 0 \end{matrix} \right\} = 1 \quad (2)$$

for $n \geq 0$,

$$\left\{ \begin{matrix} n \\ 1 \end{matrix} \right\} = n \quad (3)$$

for $n \geq 1$, and

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = \frac{(n+1-k)(n+2)(n+3)\cdots(n+k)}{k!}, \quad (4)$$

for $n \geq k \geq 2$. Setting $k = n$ then recovers the CATALAN NUMBERS

$$C_n = \left\{ \begin{matrix} n \\ n \end{matrix} \right\} = \frac{1}{n+1}\binom{2n}{n}. \quad (5)$$

see also CATALAN NUMBER

References

Bailey, D. F. "Counting Arrangements of 1's and −1's." *Math. Mag.* **69**, 128–131, 1996.

Buraldi, R. A. *Introductory Combinatorics, 2nd ed.* New York: Elsevier, 1992.

Nonorientable Surface

A surface such as the MÖBIUS STRIP on which there exists a closed path such that the directrix is reversed when moved around this path. The EULER CHARACTERISTIC of a nonorientable surface is ≤ 0. The real PROJECTIVE PLANE is also a nonorientable surface, as are the BOY SURFACE, CROSS-CAP, and ROMAN SURFACE, all of which are homeomorphic to the REAL PROJECTIVE PLANE (Pinkall 1986). There is a general method for constructing nonorientable surfaces which proceeds as follows (Banchoff 1984, Pinkall 1986). Choose three HOMOGENEOUS POLYNOMIALS of POSITIVE EVEN degree and consider the MAP

$$\mathbf{f} = (f_1(x,y,z), f_2(x,y,z), f_3(x,y,z)) : \mathbb{R}^3 \to \mathbb{R}^3. \quad (1)$$

Then restricting x, y, and z to the surface of a sphere by writing

$$x = \cos\theta \sin\phi \quad (2)$$

$$y = \sin\theta \sin\phi \quad (3)$$

$$z = \cos\phi \quad (4)$$

and restricting θ to $[0, 2\pi)$ and ϕ to $[0, \pi/2]$ defines a map of the REAL PROJECTIVE PLANE to $\mathbb{R}^3$.

In 3-D, there is no unbounded nonorientable surface which does not intersect itself (Kuiper 1961, Pinkall 1986).

see also BOY SURFACE, CROSS-CAP, MÖBIUS STRIP, ORIENTABLE SURFACE, PROJECTIVE PLANE, ROMAN SURFACE

References

Banchoff, T. "Differential Geometry and Computer Graphics." In *Perspectives of Mathematics: Anniversary of Oberwolfach* (Ed. W. Jager, R. Remmert, and J. Moser). Basel, Switzerland: Birkhäuser, 1984.

Gray, A. "Nonorientable Surfaces." Ch. 12 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 229–249, 1993.

Kuiper, N. H. "Convex Immersion of Closed Surfaces in E^3." *Comment. Math. Helv.* **35**, 85–92, 1961.

Pinkall, U. "Models of the Real Projective Plane." Ch. 6 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 63–67, 1986.

Nonpositive

A quantity which is either 0 (ZERO) or NEGATIVE, i.e., ≤ 0.

see also NEGATIVE, NONNEGATIVE, NONZERO, POSITIVE, ZERO

Nonsquarefree

see SQUAREFUL

Nonstandard Analysis

Nonstandard analysis is a branch of mathematical LOGIC which weakens the axioms of usual ANALYSIS to include only the first-order ones. It also introduces HYPERREAL NUMBERS to allow for the existence of "genuine INFINITESIMALS," numbers which are less than 1/2, 1/3, 1/4, 1/5, ..., but greater than 0. Abraham Robinson developed nonstandard analysis in the 1960s. The theory has since been investigated for its own sake and has been applied in areas such as BANACH SPACES, differential equations, probability theory, microeconomic theory, and mathematical physics (Apps).

see also AX-KOCHEN ISOMORPHISM THEOREM, LOGIC, MODEL THEORY

References

Albeverio, S.; Fenstad, J.; Hoegh-Krohn, R.; and Lindstrøom, T. *Nonstandard Methods in Stochastic Analysis and Mathematical Physics.* New York: Academic Press, 1986.

Anderson, R. "Nonstandard Analysis with Applications to Economics." In *Handbook of Mathematical Economics, Vol. 4.* New York: Elsevier, 1991.

Apps, P. "What is Nonstandard Analysis?" `http://www.math.wisc.edu/~apps/nonstandard.html`.

Dauben, J. W. *Abraham Robinson: The Creation of Nonstandard Analysis, A Personal and Mathematical Odyssey.* Princeton, NJ: Princeton University Press, 1998.

Davis, P. J. and Hersch, R. *The Mathematical Experience.* Boston: Birkhäuser, 1981.
Keisler, H. J. *Elementary Calculus: An Infinitesimal Approach.* Boston: PWS, 1986.
Lindstrøom, T. "An Invitation to Nonstandard Analysis." In *Nonstandard Analysis and Its Applications* (Ed. N. Cutland). New York: Cambridge University Press, 1988.
Robinson, A. *Non-Standard Analysis.* Princeton, NJ: Princeton University Press, 1996.
Stewart, I. "Non-Standard Analysis." In *From Here to Infinity: A Guide to Today's Mathematics.* Oxford, England: Oxford University Press, pp. 80–81, 1996.

Nontotient

A POSITIVE EVEN value of n for which $\phi(x) = n$, where $\phi(x)$ is the TOTIENT FUNCTION, has no solution. The first few are 14, 26, 34, 38, 50, ... (Sloane's A005277).

see also NONCOTOTIENT, TOTIENT FUNCTION

References
Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 91, 1994.
Sloane, N. J. A. Sequence A005277/M4927 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Nonwandering

A point x in a MANIFOLD M is said to be nonwandering if, for every open NEIGHBORHOOD U of x, it is true that $\phi^{-n}U \cup U \neq \varnothing$ for a MAP ϕ for some $n > 0$. In other words, every point close to x has some iterate under ϕ which is also close to x. The set of all nonwandering points is denoted $\Omega(\phi)$, which is known as the nonwandering set of ϕ.

see also ANOSOV DIFFEOMORPHISM, AXIOM A DIFFEOMORPHISM, SMALE HORSESHOE MAP

Nonzero

A quantity which does not equal ZERO is said to be nonzero. A REAL nonzero number must be either POSITIVE or NEGATIVE, and a COMPLEX nonzero number can have either REAL or IMAGINARY PART nonzero.

see also NEGATIVE, NONNEGATIVE, NONPOSITIVE, POSITIVE, ZERO

Nordstrand's Weird Surface

An attractive CUBIC SURFACE defined by Nordstrand. It is given by the implicit equation

$$\begin{aligned}25[x^3(y+z)+y^3(x+z)+z^3(x+y)]+50(x^2y^2+x^2z^2\\+y^2z^2)-125(x^2yz+y^2xz+z^2xy)+60xyz\\-4(xy+xz+yz)=0.\end{aligned}$$

References
Nordstrand, T. "Weird Cube." `http://www.uib.no/people/nfytn/weirdtxt.htm`.

Norm

Given a n-D VECTOR

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

a VECTOR NORM $||\mathbf{x}||$ is a NONNEGATIVE number satisfying

1. $||x|| > 0$ when $\mathbf{x} \neq \mathbf{0}$ and $||\mathbf{x}|| = 0$ IFF $\mathbf{x} = \mathbf{0}$,
2. $||k\mathbf{x}|| = |k|\,||\mathbf{x}||$ for any SCALAR k,
3. $||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$.

The most common norm is the vector L_2-NORM, defined by

$$||\mathbf{x}||_2 = |\mathbf{x}| = \sqrt{{x_1}^2 + {x_2}^2 + \ldots + {x_n}^2}.$$

Given a SQUARE MATRIX A, a MATRIX NORM $||\mathsf{A}||$ is a NONNEGATIVE number associated with A having the properties

1. $||\mathsf{A}|| > 0$ when $\mathsf{A} \neq 0$ and $||\mathsf{A}|| = 0$ IFF $\mathsf{A} = 0$,
2. $||k\mathsf{A}|| = |k|\,||\mathsf{A}||$ for any SCALAR k,
3. $||\mathsf{A}+\mathsf{B}|| \leq ||\mathsf{A}|| + ||\mathsf{B}||$,
4. $||\mathsf{A}\mathsf{B}|| \leq ||\mathsf{A}||\,||\mathsf{B}||$.

see also BOMBIERI NORM, COMPATIBLE, EUCLIDEAN NORM, HILBERT-SCHMIDT NORM, INDUCED NORM, L_1-NORM, L_2-NORM, L_∞-NORM, MATRIX NORM, MAXIMUM ABSOLUTE COLUMN SUM NORM, MAXIMUM ABSOLUTE ROW SUM NORM, NATURAL NORM, NORMALIZED VECTOR, NORMED SPACE, PARALLELOGRAM LAW, POLYNOMIAL NORM, SPECTRAL NORM, SUBORDINATE NORM, VECTOR NORM

References
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 1114–1125, 1979.

Norm Theorem

If a PRIME number divides a norm but not the bases of the norm, it is itself a norm.

Normal

see NORMAL CURVE, NORMAL DISTRIBUTION, NORMAL DISTRIBUTION FUNCTION, NORMAL EQUATION, NORMAL FORM, NORMAL GROUP, NORMAL MAGIC SQUARE, NORMAL MATRIX, NORMAL NUMBER, NORMAL PLANE, NORMAL SUBGROUP, NORMAL VECTOR

Normal (Algebraically)

see GALOISIAN

Normal Curvature

Let $\mathbf{u_p}$ be a unit TANGENT VECTOR of a REGULAR SURFACE $M \subset \mathbb{R}^3$. Then the normal curvature of M in the direction $\mathbf{u_p}$ is

$$\kappa(\mathbf{u_p}) = S(\mathbf{u_p}) \cdot \mathbf{u_p}, \qquad (1)$$

where S is the SHAPE OPERATOR. Let $M \subset \mathbb{R}^3$ be a REGULAR SURFACE, $\mathbf{p} \in M$, $\mathbf{x}$ be an injective REGULAR PATCH of M with $\mathbf{p} = \mathbf{x}(u_0, v_0)$, and

$$\mathbf{v_p} = a\mathbf{x}_u(u_0, v_0) + b\mathbf{x}_v(u_0, v_0), \qquad (2)$$

where $\mathbf{v_p} \in M_\mathbf{p}$. Then the normal curvature in the direction $\mathbf{v_p}$ is

$$\kappa(v\mathbf{p}) = \frac{ea^2 + 2fab + gb^2}{Ea^2 + 2Fab + Gb^2}, \qquad (3)$$

where E, F, and G are first FUNDAMENTAL FORMS and e, f, and g second FUNDAMENTAL FORMS.

The MAXIMUM and MINIMUM values of the normal curvature on a REGULAR SURFACE at a point on the surface are called the PRINCIPAL CURVATURES κ_1 and κ_2.

see also CURVATURE, FUNDAMENTAL FORMS, GAUSSIAN CURVATURE, MEAN CURVATURE, PRINCIPAL CURVATURES, SHAPE OPERATOR, TANGENT VECTOR

References

Euler, L. "Récherches sur la coubure des surfaces." *Mém. de l'Acad. des Sciences, Berlin* **16**, 119–143, 1760.

Gray, A. "Normal Curvature." §14.2 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 270-273 and 277, 1993.

Meusnier, J. B. "Mémoire sur la courbure des surfaces." *Mém. des savans étrangers* **10** (lu 1776), 477–510, 1785.

Normal Curve

see GAUSSIAN DISTRIBUTION

Normal Developable

A RULED SURFACE M is a normal developable of a curve $\mathbf{y}$ if M can be parameterized by $\mathbf{x}(u,v) = \mathbf{y}(u) + v\hat{\mathbf{N}}(u)$, where $\mathbf{N}$ is the NORMAL VECTOR.

see also BINORMAL DEVELOPABLE, TANGENT DEVELOPABLE

References

Gray, A. "Developables." §17.6 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 352–354, 1993.

Normal Distribution

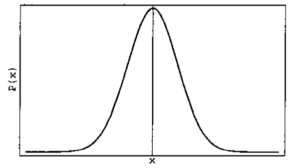

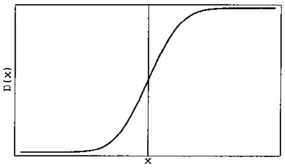

Another name for a GAUSSIAN DISTRIBUTION. Given a normal distribution in a VARIATE x with MEAN μ and VARIANCE σ^2,

$$P(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}\,dx,$$

the so-called "STANDARD NORMAL DISTRIBUTION" is given by taking $\mu = 0$ and $\sigma^2 = 1$. An arbitrary normal distribution can be converted to a STANDARD NORMAL DISTRIBUTION by changing variables to $z \equiv (x - \mu)/\sigma$, so $dz = dx/\sigma$, yielding

$$P(x)\,dx = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz.$$

The FISHER-BEHRENS PROBLEM is the determination of a test for the equality of MEANS for two normal distributions with different VARIANCES.

see also FISHER-BEHRENS PROBLEM, GAUSSIAN DISTRIBUTION, HALF-NORMAL DISTRIBUTION, KOLMOGOROV-SMIRNOV TEST, NORMAL DISTRIBUTION FUNCTION, STANDARD NORMAL DISTRIBUTION, TETRACHORIC FUNCTION

Normal Distribution Function

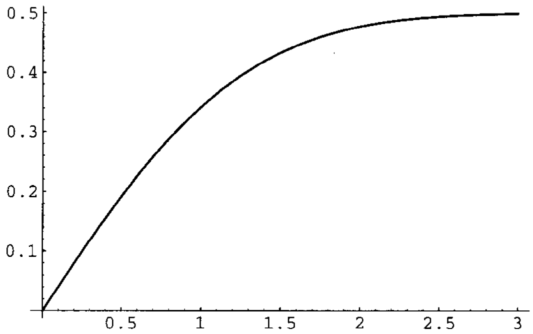

A normalized form of the cumulative GAUSSIAN DISTRIBUTION function giving the probability that a variate assumes a value in the range $[0, x]$,

$$\Phi(x) \equiv Q(x) \equiv \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2}\,dt. \qquad (1)$$

It is related to the PROBABILITY INTEGRAL

$$\alpha(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{-x}^x e^{-t^2/2}\,dt \qquad (2)$$

by

$$\Phi(x) = \tfrac{1}{2}\alpha(x). \qquad (3)$$

Let $u \equiv t/\sqrt{2}$ so $du = dt/\sqrt{2}$. Then

$$\Phi(x) = \frac{1}{\sqrt{\pi}} \int_0^{x/\sqrt{2}} e^{-u^2}\,du = \tfrac{1}{2}\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right). \qquad (4)$$

Here, ERF is a function sometimes called the error function. The probability that a normal variate assumes a value in the range $[x_1, x_2]$ is therefore given by

$$\Phi(x_1, x_2) = \frac{1}{2}\left[\operatorname{erf}\left(\frac{x_2}{\sqrt{2}}\right) - \operatorname{erf}\left(\frac{x_1}{\sqrt{2}}\right)\right]. \tag{5}$$

Neither $\Phi(z)$ nor ERF can be expressed in terms of finite additions, subtractions, multiplications, and root extractions, and so must be either computed numerically or otherwise approximated.

Note that a function different from $\Phi(x)$ is sometimes defined as "the" normal distribution function

$$\Phi'(x) \equiv \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right] = \tfrac{1}{2} + \Phi(x) \tag{6}$$

(Beyer 1987, p. 551), although this function is less widely encountered than the usual $\Phi(x)$.

The value of a for which $P(x)$ falls within the interval $[-a, a]$ with a given probability P is a related quantity called the CONFIDENCE INTERVAL.

For small values $x \ll 1$, a good approximation to $\Phi(x)$ is obtained from the MACLAURIN SERIES for ERF,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}(2x - \tfrac{1}{3}x^3 + \tfrac{1}{20}x^5 - \tfrac{1}{168}x^7 + \ldots). \tag{7}$$

For large values $x \gg 1$, a good approximation is obtained from the asymptotic series for ERF,

$$\Phi(x) = \frac{1}{2} + \frac{e^{-x^2/2}}{\sqrt{2\pi}}(x^{-1} - x^{-3} + 3x^{-5} - 15x^{-7} + 105x^{-9} + \ldots). \tag{8}$$

The value of $\Phi(x)$ for intermediate x can be computed using the CONTINUED FRACTION identity

$$\int_0^x e^{-u^2}\,du = \frac{\sqrt{\pi}}{2} - \cfrac{\tfrac{1}{2}e^{-x^2}}{x + \cfrac{1}{2x + \cfrac{2}{x + \cfrac{3}{2x + \cfrac{4}{x + \ldots}}}}}. \tag{9}$$

A simple approximation of $\Phi(x)$ which is good to two decimal places is given by

$$\Phi_1(x) \approx \begin{cases} 0.1x(4.4 - x) & \text{for } 0 \le x \le 2.2 \\ 0.49 & \text{for } 2.2 < x < 2.6 \\ 0.50 & \text{for } x \ge 2.6. \end{cases} \tag{10}$$

Abramowitz and Stegun (1972) and Johnson and Kotz (1970) give other functional approximations. An approximation due to Bagby (1995) is

$$\Phi_2(x) = \tfrac{1}{2}\{1 - \tfrac{1}{30}[7e^{-x^2/2} + 16e^{-x^2(2-\sqrt{2})} + (7 + \tfrac{1}{4}\pi x^2)e^{-x^2}]\}^{1/2}. \tag{11}$$

The plots below show the differences between Φ and the two approximations.

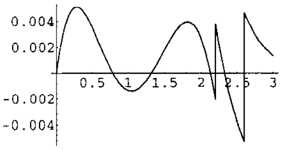

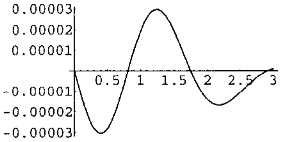

The first QUARTILE of a standard NORMAL DISTRIBUTION occurs when

$$\int_0^t \Phi(z)\,dz = \tfrac{1}{4}. \tag{12}$$

The solution is $t = 0.6745\ldots$. The value of t giving $\frac{1}{4}$ is known as the PROBABLE ERROR of a normally distributed variate.

see also CONFIDENCE INTERVAL, ERF, ERFC, FISHER-BEHRENS PROBLEM, GAUSSIAN DISTRIBUTION, GAUSSIAN INTEGRAL, HH FUNCTION, NORMAL DISTRIBUTION, PROBABILITY INTEGRAL, TETRACHORIC FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 931–933, 1972.

Bagby, R. J. "Calculating Normal Probabilities." *Amer. Math. Monthly* **102**, 46–49, 1995.

Beyer, W. H. (Ed.). *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, 1987.

Johnson, N.; Kotz, S.; and Balakrishnan, N. *Continuous Univariate Distributions, Vol. 1, 2nd ed.* Boston, MA: Houghton Mifflin, 1994.

Normal Equation

Given an overdetermined MATRIX EQUATION

$$\mathsf{Ax} = \mathsf{b},$$

the normal equation is that which minimizes the sum of the square differences between left and right sides

$$\mathsf{A}^{\mathrm{T}}\mathsf{Ax} = \mathsf{A}^{\mathrm{T}}\mathsf{b}.$$

see also LEAST SQUARES FITTING, MOORE-PENROSE GENERALIZED MATRIX INVERSE, NONLINEAR LEAST SQUARES FITTING

Normal Form

A way of representing objects so that, although each may have many different names, every possible name corresponds to exactly one object.

see also CANONICAL FORM

References

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, p. 7, 1996.

Normal Function

A SQUARE INTEGRABLE function ϕ is said to be normal if

$$\int \phi^2 \, dt = 1$$

However, the NORMAL DISTRIBUTION FUNCTION is also sometimes called "the normal function."

see also NORMAL DISTRIBUTION FUNCTION, SQUARE INTEGRABLE

References

Sansone, G. *Orthogonal Functions, rev. English ed.* New York: Dover, p. 6, 1991.

Normal Group

see NORMAL SUBGROUP

Normal Magic Square

see MAGIC SQUARE

Normal Matrix

A normal matrix A is a MATRIX for which

$$[\mathsf{A}, \mathsf{A}^\dagger] = 0,$$

where $[a, b]$ is the COMMUTATOR and † denotes the ADJOINT OPERATOR.

Normal Number

An IRRATIONAL NUMBER for which any FINITE pattern of numbers occurs with the expected limiting frequency in the expansion in any base. It is not known if π or e are normal. Tests of $\sqrt{n}$ for n = 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15 indicate that these SQUARE ROOTS may be normal. The only numbers known to be normal are artificially constructed ones such as the CHAMPERNOWNE CONSTANT and the COPELAND-ERDŐS CONSTANT.

see also CHAMPERNOWNE CONSTANT, COPELAND-ERDŐS CONSTANT, e, PI

Normal Order

$f(n)$ has the normal order $F(n)$ if $f(n)$ is approximately $F(n)$ for ALMOST ALL values of n. More precisely, if

$$(1 - \epsilon)F(n) < f(n) < (1 + \epsilon)F(n)$$

for every positive ϵ and ALMOST ALL values of n, then the normal order of $f(n)$ is $F(n)$.

see also ALMOST ALL

References

Hardy, G. H. and Weight, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Oxford University Press, p. 356, 1979.

Normal Plane

The PLANE spanned by **N** and **B** (the NORMAL VECTOR and BINORMAL VECTOR).

see also BINORMAL VECTOR, NORMAL VECTOR, PLANE

Normal to a Plane

see NORMAL VECTOR

Normal Section

Let $M \subset \mathbb{R}^3$ be a REGULAR SURFACE and $\mathbf{u_p}$ a unit TANGENT VECTOR to M, and let $\Pi(\mathbf{u_p}, \mathbf{N}(\mathbf{p}))$ be the PLANE determined by $\mathbf{u_p}$ and the normal to the surface $\mathbf{N}(\mathbf{p})$. Then the normal section of M is defined as the intersection of $\Pi(\mathbf{u_p}, \mathbf{N}(\mathbf{p}))$ and M.

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 271, 1993.

Normal Subgroup

Let H be a SUBGROUP of a GROUP G. Then H is a normal subgroup of G, written $H \lhd G$, if

$$xHx^{-1} = H$$

for every element x in H. Normal subgroups are also known as INVARIANT SUBGROUPS.

see also GROUP, SUBGROUP

Normal Vector

The normal to a PLANE specified by

$$f(x, y, z) = ax + by + cz + d = 0 \tag{1}$$

is given by

$$\mathbf{N} = \nabla f = \begin{bmatrix} a \\ b \\ c \end{bmatrix}. \tag{2}$$

The normal vector at a point (x_0, y_0) on a surface $z = f(x, y)$ is

$$\mathbf{N} = \begin{bmatrix} f_x(x_0, y_0) \\ f_y(x_0, y_0) \\ -1 \end{bmatrix}. \tag{3}$$

In the PLANE, the unit normal vector is defined by

$$\hat{\mathbf{N}} \equiv \frac{d\hat{\mathbf{T}}}{d\phi}, \tag{4}$$

where $\hat{\mathbf{T}}$ is the unit TANGENT VECTOR and ϕ is the polar angle. Given a unit TANGENT VECTOR

$$\hat{\mathbf{T}} \equiv u_1\hat{\mathbf{x}} + u_2\hat{\mathbf{y}} \tag{5}$$

with $\sqrt{u_1{}^2 + u_2{}^2} = 1$, the normal is

$$\hat{\mathbf{N}} = -u_2\hat{\mathbf{x}} + u_1\hat{\mathbf{y}}. \tag{6}$$

For a function given parametrically by $(f(t), g(t))$, the normal vector relative to the point $(f(t), g(t))$ is therefore given by

$$x(t) = -\frac{g'}{\sqrt{f'^2 + g'^2}} \tag{7}$$

$$y(t) = \frac{f'}{\sqrt{f'^2 + g'^2}}. \tag{8}$$

To actually place the vector normal to the curve, it must be displaced by $(f(t), g(t))$.

In 3-D SPACE, the unit normal is

$$\hat{\mathbf{N}} \equiv \frac{\frac{d\hat{\mathbf{T}}}{ds}}{\left|\frac{d\hat{\mathbf{T}}}{ds}\right|} = \frac{\frac{d\hat{\mathbf{T}}}{dt}}{\left|\frac{d\hat{\mathbf{T}}}{dt}\right|} = \frac{1}{\kappa}\frac{d\mathbf{T}}{ds}, \tag{9}$$

where κ is the CURVATURE. Given a 3-D surface $F(x, y, z) = 0$,

$$\hat{\mathbf{n}} = \frac{F_x + F_y + F_z}{\sqrt{{F_x}^2 + {F_y}^2 + {F_z}^2}}. \tag{10}$$

If the surface is defined parametrically in the form

$$x = x(\phi, \psi) \tag{11}$$

$$y = y(\phi, \psi) \tag{12}$$

$$z = z(\phi, \psi), \tag{13}$$

define the VECTORS

$$\mathbf{a} \equiv \begin{bmatrix} x_\phi \\ y_\phi \\ z_\phi \end{bmatrix} \tag{14}$$

$$\mathbf{b} \equiv \begin{bmatrix} x_\psi \\ y_\psi \\ z_\psi \end{bmatrix}. \tag{15}$$

Then the unit normal vector is

$$\hat{\mathbf{N}} = \frac{\mathbf{a} \times \mathbf{b}}{\sqrt{|\mathbf{a}|^2|\mathbf{b}|^2 - |\mathbf{a} \cdot \mathbf{b}|^2}}. \tag{16}$$

Let g be the discriminant of the METRIC TENSOR. Then

$$\mathbf{N} = \frac{\mathbf{r}_1 \times \mathbf{r}_2}{\sqrt{g}} = \epsilon_{ij}\mathbf{r}^j. \tag{17}$$

see also BINORMAL VECTOR, CURVATURE, FRENET FORMULAS, TANGENT VECTOR

References

Gray, A. "Tangent and Normal Lines to Plane Curves." §5.5 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 85–90, 1993.

Normalized Vector

The normalized vector of $\mathbf{X}$ is a VECTOR in the same direction but with NORM (length) 1. It is denoted $\hat{\mathbf{X}}$ and given by

$$\hat{\mathbf{X}} \equiv \frac{\mathbf{X}}{|\mathbf{X}|},$$

where $|\mathbf{X}|$ is the NORM of $\mathbf{X}$. It is also called a UNIT VECTOR.

see also UNIT VECTOR

Normalizer

A set of elements g of a GROUP such that

$$g^{-1}Hg = H,$$

is said to be the normalizer $N_G(H)$ with respect to a subset of group elements H.

see also CENTRALIZER, TIGHTLY EMBEDDED

Normed Space

A VECTOR SPACE possessing a NORM.

Nosarzewska's Inequality

Given a convex PLANE region with AREA A and PERIMETER p,

$$A - \tfrac{1}{2}p < N \leq A + \tfrac{1}{2}p + 1,$$

where N is the number of enclosed LATTICE POINTS (Nosarzewska 1948). This improves on JARNICK'S INEQUALITY

$$|N - A| < p.$$

see also JARNICK'S INEQUALITY, LATTICE POINT

References

Nosarzewska, M. "Évaluation de la différence entre l'aire d'une région plane convexe et le nombre des points aux coordonnées entières couverts par elle." *Colloq. Math.* **1**, 305–311, 1948.

Not

An operation in LOGIC which converts TRUE to FALSE and FALSE to TRUE. NOT A is denoted $!A$ or $\neg A$.

A	$\neg A$
F	T
T	F

see also AND, OR, TRUTH TABLE, XOR

Notation

A NOTATION is a set of well-defined rules for representing quantities and operations with symbols.

see also ARROW NOTATION, CHAINED ARROW NOTATION, CIRCLE NOTATION, CLEBSCH-ARONHOLD NOTATION, CONWAY'S KNOT NOTATION, DOWKER NOTATION, DOWN ARROW NOTATION, PETROV NOTATION, SCIENTIFIC NOTATION, STEINHAUS-MOSER NOTATION

References

Cajori, F. *A History of Mathematical Notations, Vols. 1–2.* New York: Dover, 1993.

Miller, J. "Earliest Uses of Various Mathematical Symbols." `http://members.aol.com/jeff570/mathsym.html`.

Miller, J. "Earliest Uses of Some of the Words of Mathematics." `http://members.aol.com/jeff570/mathword.html`.

Nöther

see NOETHER'S FUNDAMENTAL THEOREM, NOETHER-LASKER THEOREM, NOETHER'S TRANSFORMATION THEOREM, NOETHERIAN MODULE, NOETHERIAN RING

Novemdecillion

In the American system, 10^{60}.

see also LARGE NUMBER

NP-Complete Problem

A problem which is both NP (solvable in nondeterministic POLYNOMIAL time) and NP-HARD (can be translated into any other NP-PROBLEM). Examples of NP-hard problems include the HAMILTONIAN CYCLE and TRAVELING SALESMAN PROBLEMS.

In a landmark paper, Karp (1972) showed that 21 intractable combinatorial computational problems are all NP-complete.

see also HAMILTONIAN CYCLE, NP-HARD PROBLEM, NP-PROBLEM, P-PROBLEM, TRAVELING SALESMAN PROBLEM

References

Karp, R. M. "Reducibility Among Combinatorial Problems." In *Complexity of Computer Computations,* (Proc. Sympos. IBM Thomas J. Watson Res. Center, Yorktown Heights, N.Y., 1972). New York: Plenum, pp. 85–103, 1972.

NP-Hard Problem

A problem is NP-hard if an ALGORITHM for solving it can be translated into one for solving any other NP-PROBLEM (nondeterministic POLYNOMIAL time) problem. NP-hard therefore means "at least as hard as any NP-PROBLEM," although it might, in fact, be harder.

see also COMPLEXITY THEORY, HITTING SET, NP-COMPLETE PROBLEM, NP-PROBLEM, P-PROBLEM, SATISFIABILITY PROBLEM

NP-Problem

A problem is assigned to the NP (nondeterministic POLYNOMIAL time) class if it is solvable in polynomial time by a nondeterministic TURING MACHINE. (A nondeterministic TURING MACHINE is a "parallel" TURING MACHINE which can take many computational paths simultaneously, with the restriction that the parallel Turing machines cannot communicate.) A P-PROBLEM (whose solution time is bounded by a polynomial) is always also NP. If a solution to an NP problem is known, it can be reduced to a single P (POLYNOMIAL time) verification.

LINEAR PROGRAMMING, long known to be NP and thought *not* to be P, was shown to be P by L. Khachian in 1979. It is not known if all apparently NP problems are actually P.

A problem is said to be NP-HARD if an ALGORITHM for solving it can be translated into one for solving any other NP-problem problem. It is much easier to show that a problem is NP than to show that it is NP-HARD. A problem which is both NP and NP-HARD is called an NP-COMPLETE PROBLEM.

see also COMPLEXITY THEORY, NP-COMPLETE PROBLEM, NP-HARD PROBLEM, P-PROBLEM, TURING MACHINE

References

Borwein, J. M. and Borwein, P. B. *Pi and the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, 1987.

Greenlaw, R.; Hoover, H. J.; and Ruzzo, W. L. *Limits to Parallel Computation: P-Completeness Theory.* Oxford, England: Oxford University Press, 1995.

NSW Number

The numbers

$$S_{2m+1} = \frac{(1+\sqrt{2}\,)^{2m+1} + (1-\sqrt{2}\,)^{2m+1}}{2}$$

for positive integer m. The first few terms are 1, 7, 41, 239, 1393, ... (Sloane's A002315). The indices giving PRIME NSW numbers are 3, 5, 7, 19, 29, 47, 59, 163, 257, 421, 937, 947, 1493, 1901, ... (Sloane's A005850).

References

Ribenboim, P. "The NSW Primes." §5.9 in *The New Book of Prime Number Records.* New York: Springer-Verlag, pp. 367–369, 1996.

Sloane, N. J. A. Sequences A002315/M4423 and A005850/M2426 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Nu Function

$$\nu(x) \equiv \int_0^\infty \frac{x^t\,dt}{\Gamma(t+1)}$$

$$\nu(x,\alpha) \equiv \int_0^\infty \frac{x^{\alpha+t}\,dt}{\Gamma(\alpha+t+1)},$$

where $\Gamma(z)$ is the GAMMA FUNCTION. See Gradshteyn and Ryzhik (1980, p. 1079).

see also LAMBDA FUNCTION, MU FUNCTION

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, 1979.

Null Function

A null function $\delta^0(x)$ satisfies

$$\int_a^b \delta^0(x)\,dx = 0 \tag{1}$$

for all a, b, so

$$\int_{-\infty}^\infty |\delta^0(x)|\,dx = 0. \tag{2}$$

Like a DELTA FUNCTION, they satisfy

$$\delta_0(x) = \begin{cases} 0 & x \neq 0 \\ 1 & x = 0. \end{cases} \tag{3}$$

see also DELTA FUNCTION, LERCH'S THEOREM

Null Graph

A Graph containing only Vertices and no Edges.

Null Hypothesis

A hypothesis which is tested for possible rejection under the assumption that it is true (usually that observations are the result of chance). The concept was introduced by R. A. Fisher.

Null Tetrad

$$g_{ij} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{bmatrix}.$$

It can be expressed as

$$g_{ab} = l_a n_b + l_b n_a - m_a \bar{m}_b - m_b \bar{m}_a.$$

see also Tetrad

References

d'Inverno, R. *Introducing Einstein's Relativity.* Oxford, England: Oxford University Press, pp. 248–249, 1992.

Nullspace

Also called the Kernel. If T is a linear transformation of $\mathbb{R}^n$, then Null(T) is the set of all Vectors $\mathbf{X}$ such that T(X) = 0, i.e.,

$$\text{Null}(T) \equiv \{\mathbf{X} : T(\mathbf{X}) = 0\}.$$

Nullstellansatz

see Hilbert's Nullstellansatz

Number

The word "number" is a general term which refers to a member of a given (possibly ordered) Set. The meaning of "number" is often clear from context (i.e., does it refer to a Complex Number, Integer, Real Number, etc.?). Wherever possible in this work, the word "number" is used to refer to quantities which are Integers, and "Constant" is reserved for nonintegral numbers which have a fixed value. Because terms such as Real Number, Bernoulli Number, and Irrational Number are commonly used to refer to nonintegral quantities, however, it is not possible to be entirely consistent in nomenclature.

see also Abundant Number, Ackermann Number, Algebraic Number, Almost Perfect Number, Amenable Number, Amicable Numbers, Antimorphic Number, Apocalypse Number, Apocalyptic Number, Armstrong Number, Arrangement Number, Bell Number, Bernoulli Number, Bertelsen's Number, Betrothed Numbers, Betti Number, Bezout Numbers, Binomial Number, Brauer Number, Brown Numbers, Cardinal Number, Carmichael Number, Catalan Number, Cayley Number, Centered Cube Number, Centered Square Number, Chaitin's Number, Chern Number, Choice Number, Christoffel Number, Clique Number, Columbian Number, Complex Number, Computable Number, Condition Number, Congruent Numbers, Constructible Number, Cotes Number, Crossing Number (Graph), Crossing Number (Link), Cubic Number, Cullen Number, Cunningham Number, Cyclic Number, Cyclomatic Number, D-Number, de Moivre Number, Deficient Number, Delannoy Number, Demlo Number, Diagonal Ramsey Number, e-Perfect Number, Eban Number, Eddington Number, Edge Number, Enneagonal Number, Entringer Number, Erdős Number, Euclid Number, Euler's Idoneal Number, Euler Number, Eulerian Number, Euler Zigzag Number, Even Number, Factorial Number, Fermat Number, Fibonacci Number, Figurate Number, G-Number, Genocchi Number, Giuga Number, Gnomic Number, Gonal Number, Graham's Number, Gregory Number, Hailstone Number, Hansen Number, Happy Number, Harmonic Divisor Number, Harmonic Number, Harshad Number, Heegner Number, Heesch Number, Helly Number, Heptagonal Number, Heterogeneous Numbers, Hex Number, Hex Pyramidal Number, Hexagonal Number, Homogeneous Numbers, Hurwitz Number, Hypercomplex Number, Hyperperfect Number, i, Idoneal Number, Imaginary Number, Independence Number, Infinary Multiperfect Number, Infinary Perfect Number, Irrational Number, Irreducible Semiperfect Number, Irredundant Ramsey Number, j, Kaprekar Number, Keith Number, Kissing Number, Knödel Numbers, Lagrange Number (Diophantine Equation), Lagrange Number (Rational Approximation), Large Number, Least Deficient Number, Lehmer Number, Leviathan Number, Liouville Number, Logarithmic Number, Lucas Number, Lucky Number, MacMahon's Prime Number of Measurement, Markov Number, McNugget Number, Ménage Number, Mersenne Number, Motzkin Number, Multiplicative Perfect Number, Multiply Perfect Number, Narcissistic Number, Natural Number, Near Noble Number, Nexus Number, Niven Number, Noble Number, Nonagonal Number, Normal Number, NSW Number, Number Guessing, Oblong Number, Octagonal Number, Octahedral Number, Odd Number, Ore Number, Ordinal Number, Pentagonal Number, Pentatope Number, Perfect Digital Invariant, Perfect Number, Persistent Number, Pluperfect Number, Plus Perfect Number, Plutarch Numbers, Polygonal Number, PONTRYAGIN NUMBER, POULET NUMBER, POWERFUL NUMBER, PRACTICAL NUMBER, PRIMARY, PRIMITIVE ABUNDANT NUMBER, PRIMITIVE PSEUDOPERFECT NUMBER, PRIMITIVE SEMIPERFECT NUMBER, PSEUDOPERFECT NUMBER, PSEUDORANDOM NUMBER, PSEUDOSQUARE, PYRAMIDAL NUMBER, Q-NUMBER, QUASIPERFECT NUMBER, RAMSEY NUMBER, RATIONAL NUMBER, REAL NUMBER, RENCONTRES NUMBER, RECURRING DIGITAL INVARIANT, REPFIGIT NUMBER, RHOMBIC DODECAHEDRAL NUMBER, RIESEL NUMBER, ROTATION NUMBER, RSA NUMBER, SARRUS NUMBER, SCHRÖDER NUMBER, SCHUR NUMBER, SECANT NUMBER, SEGMENTED NUMBER, SELF-DESCRIPTIVE NUMBER, SELF NUMBER, SEMIPERFECT NUMBER, SIERPIŃSKI NUMBER OF THE FIRST KIND, SIERPIŃSKI NUMBER OF THE SECOND KIND, SINGLY EVEN NUMBER, SKEWES NUMBER, SMALL NUMBER, SMITH NUMBER, SMOOTH NUMBER, SOCIABLE NUMBERS, SPRAGUE-GRUNDY NUMBER, SQUARE NUMBER, SQUARE PYRAMIDAL NUMBER, STAR NUMBER, STELLA OCTANGULA NUMBER, STIEFEL-WHITNEY NUMBER, STIRLING CYCLE NUMBER, STIRLING SET NUMBER, STØRMER NUMBER, SUBLIME NUMBER, SUITABLE NUMBER, SUM-PRODUCT NUMBER, SUPER-3 NUMBER, SUPER CATALAN NUMBER, SUPERABUNDANT NUMBER, SUPERPERFECT NUMBER, SUPER-POULET NUMBER, TANGENT NUMBER, TAXICAB NUMBER, TETRAHEDRAL NUMBER, TRANSCENDENTAL NUMBER, TRANSFINITE NUMBER, TRIANGULAR NUMBER, TRIBONACCI NUMBER, TRIMORPHIC NUMBER, TRUNCATED OCTAHEDRAL NUMBER, TRUNCATED TETRAHEDRAL NUMBER, TWIST NUMBER, U-NUMBER, ULAM NUMBER, UNDULATING NUMBER, UNHAPPY NUMBER, UNITARY MULTIPERFECT NUMBER, UNITARY PERFECT NUMBER, UNTOUCHABLE NUMBER, VAMPIRE NUMBER, VAN DER WAERDEN NUMBER, VR NUMBER, WEIRD NUMBER, WHOLE NUMBER, WOODALL NUMBER, Z-NUMBER, ZAG NUMBER, ZEISEL NUMBER, ZIG NUMBER

References

Barbeau, E. J. *Power Play: A Country Walk through the Magical World of Numbers.* Providence, RI: Amer. Math. Soc., 1997.

Bogomolny, A. "What is a Number." `http://www.cut-the-knot.com/do_you_know/numbers.html`.

Borwein, J. and Borwein, P. *A Dictionary of Real Numbers.* London: Chapman & Hall, 1990.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, 1996.

Dantzig, T. *Number: The Language of Science, 4th rev. ed.* New York: Free Press, 1985.

Davis, P. J. *The Lore of Large Numbers.* New York: Random House, 1961.

Frege, G. *Grundlagen der Arithmetik: Eine logisch mathematische Untersuchung über den Begriff der Zahl.* New York: Georg Olms, 1997.

Frege, G. *Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number.* Evanston, IL: Northwestern University Press, 1968.

Ifrah, G. *From One to Zero: A Universal History of Numbers.* New York: Viking, 1987.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, 1983.

Phillips, R. *Numbers: Facts, Figures & Fiction.* Cambridge, England: Cambridge University Press, 1994.

Russell, B. "Definition of Number." *Introduction to Mathematical Philosophy.* New York: Simon and Schuster, 1971.

Smeltzer, D. *Man and Number.* Buchanan, NY: Emerson Books, 1974.

Wells, D. W. *The Penguin Dictionary of Curious and Interesting Numbers.* Harmondsworth, England: Penguin Books, 1986.

Number Axis

see REAL LINE

Number Field

If r is an ALGEBRAIC NUMBER of degree n, then the totality of all expressions that can be constructed from r by repeated additions, subtractions, multiplications, and divisions is called a number field (or an ALGEBRAIC NUMBER FIELD) generated by r, and is denoted $F[r]$. Formally, a number field is a finite extension $\mathbb{Q}(\alpha)$ of the FIELD $\mathbb{Q}$ of RATIONAL NUMBERS.

The numbers of a number field which are ROOTS of a POLYNOMIAL

$$z^n + a_{n-1}z^{n-1} + \ldots + a_0 = 0$$

with integral coefficients and leading coefficient 1 are called the ALGEBRAIC INTEGERS of that field.

see also ALGEBRAIC FUNCTION FIELD, ALGEBRAIC INTEGER, ALGEBRAIC NUMBER, FIELD, FINITE FIELD, $\mathbb{Q}$, QUADRATIC FIELD

References

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 127, 1996.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 151–152, 1993.

Number Field Sieve Factorization Method

An extremely fast factorization method developed by Pollard which was used to factor the RSA-130 NUMBER. This method is the most powerful known for factoring general numbers, and has complexity

$$\mathcal{O}\{\exp[c(\log n)^{1/3}(\log\log n)^{2/3}]\},$$

reducing the *exponent* over the CONTINUED FRACTION FACTORIZATION ALGORITHM and QUADRATIC SIEVE FACTORIZATION METHOD. There are three values of c relevant to different flavors of the method (Pomerance 1996). For the "special" case of the algorithm applied to numbers near a large POWER,

$$c = (\tfrac{32}{9})^{1/3} = 1.523\ldots,$$

for the "general" case applicable to any ODD POSITIVE number which is not a POWER,

$$c = (\tfrac{64}{9})^{1/3} = 1.923\ldots,$$

and for a version using many POLYNOMIALS (Coppersmith 1993),

$$c = \tfrac{1}{3}(92 + 26\sqrt{13})^{1/3} = 1.902\ldots.$$

References
Coppersmith, D. "Modifications to the Number Field Sieve." *J. Cryptology* **6**, 169–180, 1993.
Coppersmith, D.; Odlyzko, A. M.; and Schroeppel, R. "Discrete Logarithms in GF(p)." *Algorithmics* **1**, 1–15, 1986.
Cowie, J.; Dodson, B.; Elkenbracht-Huizing, R. M.; Lenstra, A. K.; Montgomery, P. L.; Zayer, J. A. "World Wide Number Field Sieve Factoring Record: On to 512 Bits." In *Advances in Cryptology—ASIACRYPT '96 (Kyongju)* (Ed. K. Kim and T. Matsumoto.) New York: Springer-Verlag, pp. 382–394, 1996.
Elkenbracht-Huizing, M. "A Multiple Polynomial General Number Field Sieve." *Algorithmic Number Theory (Talence, 1996).* New York: Springer-Verlag, pp. 99–114, 1996.
Elkenbracht-Huizing, M. "An Implementation of the Number Field Sieve." *Experiment. Math.* **5**, 231–253, 1996.
Elkenbracht-Huizing, R.-M. "Historical Background of the Number Field Sieve Factoring Method." *Nieuw Arch. Wisk.* **14**, 375–389, 1996.
Lenstra, A. K. and Lenstra, H. W. Jr. "Algorithms in Number Theory." In *Handbook of Theoretical Computer Science, Volume A: Algorithms and Complexity* (Ed. J. van Leeuwen). New York: Elsevier, pp. 673–715, 1990.
Pomerance, C. "A Tale of Two Sieves." *Not. Amer. Math. Soc.* **43**, 1473–1485, 1996.

Number Group

see FIELD

Number Guessing

By asking a small number of innocent-sounding questions about an unknown number, it is possible to reconstruct the number with absolute certainty (assuming that the questions are answered correctly). Ball and Coxeter (1987) give a number of sets of questions which can be used.

One of the simplest algorithms uses only three questions to determine an unknown number n:

1. Triple n and announce if the result $n' = 3n$ is EVEN or ODD.
2. If you were told that n' is EVEN, ask the person to reveal the number n'' which is half of n'. If you were told that n' is ODD, ask the person to reveal the number n'' which is half of $n' + 1$.
3. Ask the person to reveal the number of times k which 9 divides evenly into $n''' = 3n''$.

The original number n is then given by $2k$ if n' was EVEN, or $2k + 1$ if n' was ODD. For $n = 2m$ even, $n' = 6m$, $n'' = 3m$, $n''' = 9m$, $k = m$, so $2k = 2m = n$. For $n = 2m + 1$ odd, $n' = 6m + 3$, $n'' = 3m + 2$, $n''' = 9m + 6$, $k = m$, so $2k + 1 = 2m + 1 = n$.

Another method asks:

1. Multiply the number n by 5.
2. Add 6 to the product.
3. Multiply the sum by 4.
4. Add 9 to the product.
5. Multiply the sum by 5 and reveal the result n'.

The original number is then given by $n = (n' - 165)/100$, since the above steps give $n' = 5(4(5n+6)+9) = 100n + 165$.

References
Bachet, C. G. *Problèmes plaisans et délectables, 2nd ed.* 1624.
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 5–20, 1987.
Kraitchik, M. "To Guess a Selected Number." §3.3 in *Mathematical Recreations.* New York: W. W. Norton, pp. 58–66, 1942.

Number Pyramid

A set of numbers obeying a pattern like the following.

$$\begin{aligned} 91 \cdot 37 &= 3367 \\ 9901 \cdot 3367 &= 33336667 \\ 999001 \cdot 333667 &= 333333666667 \\ 99990001 \cdot 33336667 &= 3333333366666667 \end{aligned}$$

$$\begin{aligned} 4^2 &= 16 \\ 34^2 &= 1156 \\ 334^2 &= 111556 \end{aligned}$$

$$\begin{aligned} 7^2 &= 49 \\ 67^2 &= 4489 \\ 667^2 &= 444889. \end{aligned}$$

see also AUTOMORPHIC NUMBER

References
Heinz, H. "Miscellaneous Number Patters." `http://www.geocities.com/CapeCanaveral/Launchpad/4057/miscnum.htm`.

Number System

see BASE (NUMBER)

Number Theoretic Transform

Simplemindedly, a number theoretic transform is a generalization of a FAST FOURIER TRANSFORM obtained by replacing $e^{-2\pi ik/N}$ with an nth PRIMITIVE ROOT OF UNITY. This effectively means doing a transform over the QUOTIENT RING $\mathbb{Z}/p\mathbb{Z}$ instead of the COMPLEX NUMBERS $\mathbb{C}$. The theory is rather elegant and uses the language of FINITE FIELDS and NUMBER THEORY.

see also FAST FOURIER TRANSFORM, FINITE FIELD

References
Arndt, J. "Numbertheoretic Transforms (NTTs)." Ch. 4 in "Remarks on FFT Algorithms." `http://www.jjj.de/fxt/`.
Cohen, H. *A Course in Computational Algebraic Number Theory.* New York: Springer-Verlag, 1993.

Number Theory

A vast and fascinating field of mathematics consisting of the study of the properties of whole numbers. PRIMES and PRIME FACTORIZATION are especially important in number theory, as are a number of functions such as the DIVISOR FUNCTION, RIEMANN ZETA FUNCTION, and TOTIENT FUNCTION. Excellent introductions to number theory may be found in Ore (1988) and Beiler (1966). The classic history on the subject (now slightly dated) is that of Dickson (1952).

see also ARITHMETIC, CONGRUENCE, DIOPHANTINE EQUATION, DIVISOR FUNCTION, GÖDEL'S INCOMPLETENESS THEOREM, PEANO'S AXIOMS, PRIME COUNTING FUNCTION, PRIME FACTORIZATION, PRIME NUMBER, QUADRATIC RECIPROCITY THEOREM, RIEMANN ZETA FUNCTION, TOTIENT FUNCTION

References

Andrews, G. E. *Number Theory.* New York: Dover, 1994.

Andrews, G. E.; Berndt, B. C.; and Rankin, R. A. (Ed.). *Ramanujan Revisited: Proceedings of the Centenary Conference, University of Illinois at Urbana-Champaign, June 1-5, 1987.* Boston, MA: Academic Press, 1988.

Apostol, T. M. *Introduction to Analytic Number Theory.* New York: Springer-Verlag, 1976.

Ayoub, R. G. *An Introduction to the Analytic Theory of Numbers.* Providence, RI: Amer. Math. Soc., 1963.

Beiler, A. H. *Recreations in the Theory of Numbers: The Queen of Mathematics Entertains, 2nd ed.* New York: Dover, 1966.

Bellman, R. E. *Analytic Number Theory: An Introduction.* Reading, MA: Benjamin/Cummings, 1980.

Berndt, B. C. *Ramanujan's Notebooks, Part I.* New York: Springer-Verlag, 1985.

Berndt, B. C. *Ramanujan's Notebooks, Part II.* New York: Springer-Verlag, 1988.

Berndt, B. C. *Ramanujan's Notebooks, Part III.* New York: Springer-Verlag, 1997.

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, 1993.

Berndt, B. C. *Ramanujan's Notebooks, Part V.* New York: Springer-Verlag, 1997.

Berndt, B. C. and Rankin, R. A. *Ramanujan: Letters and Commentary.* Providence, RI: Amer. Math. Soc, 1995.

Borwein, J. M. and Borwein, P. B. *Pi and the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, 1987.

Brown, K. S. "Number Theory." `http://www.seanet.com/~ksbrown/inumber.htm`.

Burr, S. A. *The Unreasonable Effectiveness of Number Theory.* Providence, RI: Amer. Math. Soc., 1992.

Burton, D. M. *Elementary Number Theory, 4th ed.* Boston, MA: Allyn and Bacon, 1989.

Carmichael, R. D. *The Theory of Numbers, and Diophantine Analysis.* New York: Dover, 1959.

Cohn, H. *Advanced Number Theory.* New York: Dover, 1980.

Courant, R. and Robbins, H. "The Theory of Numbers." Supplement to Ch. 1 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 21-51, 1996.

Davenport, H. *The Higher Arithmetic: An Introduction to the Theory of Numbers, 6th ed.* Cambridge, England: Cambridge University Press, 1992.

Davenport, H. and Montgomery, H. L. *Multiplicative Number Theory, 2nd ed.* New York: Springer-Verlag, 1980.

Dickson, L. E. *History of the Theory of Numbers, 3 vols.* New York: Chelsea, 1952.

Dudley, U. *Elementary Number Theory.* San Francisco, CA: W. H. Freeman, 1978.

Friedberg, R. *An Adventurer's Guide to Number Theory.* New York: Dover, 1994.

Gauss, C. F. *Disquisitiones Arithmeticae.* New Haven, CT: Yale University Press, 1966.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, 1994.

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, 1979.

Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, 1959.

Hasse, H. *Number Theory.* Berlin: Springer-Verlag, 1980.

Ireland, K. F. and Rosen, M. I. *A Classical Introduction to Modern Number Theory, 2nd ed.* New York: Springer-Verlag, 1995.

Klee, V. and Wagon, S. *Old and New Unsolved Problems in Plane Geometry and Number Theory.* Washington, DC: Math. Assoc. Amer., 1991.

Koblitz, N. *A Course in Number Theory and Cryptography.* New York: Springer-Verlag, 1987.

Landau, E. *Elementary Number Theory, 2nd ed.* New York: Chelsea, 1988.

Lang, S. *Algebraic Number Theory, 2nd ed.* New York: Springer-Verlag, 1994.

Lenstra, H. W. and Tijdeman, R. (Eds.). *Computational Methods in Number Theory, 2 vols.* Amsterdam: Mathematisch Centrum, 1982.

LeVeque, W. J. *Fundamentals of Number Theory.* New York: Dover, 1996.

Mitrinovic, D. S. and Sandor, J. *Handbook of Number Theory.* Dordrecht, Netherlands: Kluwer, 1995.

Niven, I. M.; Zuckerman, H. S.; and Montgomery, H. L. *An Introduction to the Theory of Numbers, 5th ed.* New York: Wiley, 1991.

Ogilvy, C. S. and Anderson, J. T. *Excursions in Number Theory.* New York: Dover, 1988.

Ore, Ø. *Invitation to Number Theory.* Washington, DC: Math. Assoc. Amer., 1967.

Ore, Ø. *Number Theory and Its History.* New York: Dover, 1988.

Rose, H. E. *A Course in Number Theory, 2nd ed.* Oxford, England: Clarendon Press, 1995.

Rosen, K. H. *Elementary Number Theory and Its Applications, 3rd ed.* Reading, MA: Addison-Wesley, 1993.

Schroeder, M. R. *Number Theory in Science and Communication: With Applications in Cryptography, Physics, Digital Information, Computing, and Self-Similarity, 3rd ed.* New York: Springer-Verlag, 1997.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, 1993.

Sierpinski, W. *250 Problems in Elementary Number Theory.* New York: American Elsevier, 1970.

Uspensky, J. V. and Heaslet, M. A. *Elementary Number Theory.* New York: McGraw-Hill, 1939.

Vinogradov, I. M. *Elements of Number Theory, 5th rev. ed.* New York: Dover, 1954.

Weil, A. *Basic Number Theory, 3rd ed.* Berlin: Springer-Verlag, 1995.

Weil, A. *Number Theory: An Approach Through History From Hammurapi to Legendre.* Boston, MA: Birkhäuser, 1984.

Weyl, H. *Algebraic Theory of Numbers.* Princeton, NJ: Princeton University Press, 1998.

Number Triangle

see Bell Triangle, Clark's Triangle, Euler's Triangle, Leibniz Harmonic Triangle, Pascal's Triangle, Seidel-Entringer-Arnold Triangle, Trinomial Triangle

Number Wall

see Quotient-Difference Table

Numerator

The number p in a Fraction p/q.

see also Denominator, Fraction, Rational Number

Numeric Function

A Function $f : A \to B$ such that B is a Set of numbers.

Numerical Derivative

While it is usually much easier to compute a Derivative instead of an Integral (which is a little strange, considering that "more" functions have integrals than derivatives), there are still many applications where derivatives need to be computed numerically. The simplest approach simply uses the definition of the Derivative

$$f'(x) \equiv \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$$

for some small numerical value of $h \ll 1$.

see also Numerical Integration

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Numerical Derivatives." §5.7 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 180–184, 1992.

Numerical Integration

The approximate computation of an Integral. The numerical computation of an Integral is sometimes called Quadrature. There are a wide range of methods available for numerical integration. A good source for such techniques is Press *et al.* (1992).

The most straightforward numerical integration technique uses the Newton-Cotes Formulas (also called Quadrature Formulas), which approximate a function tabulated at a sequent of regularly spaced Intervals by various degree Polynomials. If the endpoints are tabulated, then the 2- and 3-point formulas are called the Trapezoidal Rule and Simpson's Rule, respectively. The 5-point formula is called Bode's Rule. A generalization of the Trapezoidal Rule is Romberg Integration, which can yield accurate results for many fewer function evaluations.

If the functions are known analytically instead of being tabulated at equally spaced intervals, the best numerical method of integration is called Gaussian Quadrature. By picking the abscissas at which to evaluate the function, Gaussian quadrature produces the most accurate approximations possible. However, given the speed of modern computers, the additional complication of the Gaussian Quadrature formalism often makes it less desirable than simply brute-force calculating twice as many points on a regular grid (which also permits the already computed values of the function to be re-used). An excellent reference for Gaussian Quadrature is Hildebrand (1956).

see also Double Exponential Integration, Filon's Integration Formula, Integral, Integration, Numerical Derivative, Quadrature

References

Davis, P. J. and Rabinowitz, P. *Methods of Numerical Integration, 2nd ed.* New York: Academic Press, 1984.

Hildebrand, F. B. *Introduction to Numerical Analysis.* New York: McGraw-Hill, pp. 319–323, 1956.

Milne, W. E. *Numerical Calculus: Approximations, Interpolation, Finite Differences, Numerical Integration and Curve Fitting.* Princeton, NJ: Princeton University Press, 1949.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, 1992.

Numerology

The study of numbers for the supposed purpose of predicting future events or seeking connections with the occult.

see also Beast Number, Number Theory

NURBS Curve

A nonuniform rational B-Spline curve defined by

$$\mathbf{C}(t) = \frac{\sum_{i=0}^{n} N_{i,p}(t) w_i \mathbf{P}_i}{\sum_{i=0}^{n} N_{i,p}(t) w_i},$$

where p is the order, $N_{i,p}$ are the B-Spline basis functions, $\mathbf{P}_i$ are control points, and the weight w_i of $\mathbf{P}_i$ is the last ordinate of the homogeneous point $\mathbf{P}_i^w$. These curves are closed under perspective transformations and can represent Conic Sections exactly.

see also B-Spline, Bézier Curve, NURBS Surface

References

Piegl, L. and Tiller, W. *The NURBS Book, 2nd ed* New York: Springer-Verlag, 1997.

NURBS Surface

A nonuniform rational B-Spline surface of degree (p, q) is defined by

$$\mathbf{S}(u,v) = \frac{\sum_{i=0}^{m} \sum_{j=0}^{n} N_{i,p}(u) N_{j,q}(v) w_{i,j} \mathbf{P}_{i,j}}{\sum_{i=0}^{m} \sum_{j=0}^{n} N_{i,p}(u) N_{j,q}(v) w_{i,j}},$$

where $N_{i,p}$ and $N_{j,q}$ are the B-SPLINE basis functions, $\mathbf{P}_{i,j}$ are control points, and the weight $w_{i,j}$ of $\mathbf{P}_{i,j}$ is the last ordinate of the homogeneous point $\mathbf{P}^w_{i,j}$.

see also B-SPLINE, BÉZIER CURVE, NURBS CURVE

Nyquist Frequency

In order to recover all FOURIER components of a periodic waveform, it is necessary to sample twice as fast as the *highest* waveform frequency ν,

$$f_{\text{Nyquist}} = 2\nu.$$

The minimum sampling frequency is called the Nyquist frequency.

see also FOURIER SERIES, FOURIER TRANSFORM, NYQUIST SAMPLING, OVERSAMPLING, SAMPLING THEOREM

Nyquist Sampling

Sampling at the NYQUIST FREQUENCY.

O

Obelus

The symbol ÷ used to indicate Division. In typography, an obelus has a more general definition as any symbol, such as the dagger (†), used to indicate a footnote.

see also Division, Solidus

Object

A mathematical structure (e.g., a Group, Vector Space, or Differentiable Manifold) in a Category.

see also Morphism

Oblate Spheroid

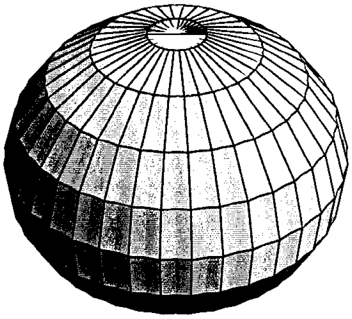

A "squashed" Spheroid for which the equatorial radius a is greater than the polar radius c, so $a > c$. To first approximation, the shape assumed by a rotating fluid (including the Earth, which is "fluid" over astronomical time scales) is an oblate spheroid. The oblate spheroid can be specified parametrically by the usual Spheroid equations (for a Spheroid with z-Axis as the symmetry axis),

$$x = a \sin v \cos u \tag{1}$$

$$y = a \sin v \sin u \tag{2}$$

$$z = c \cos v, \tag{3}$$

with $a > c$, $u \in [0, 2\pi)$, and $v \in [0, \pi]$. Its Cartesian equation is

$$\frac{x^2 + y^2}{a^2} + \frac{z^2}{c^2} = 1. \tag{4}$$

The Ellipticity of an oblate spheroid is defined by

$$e \equiv \sqrt{\frac{a^2 - c^2}{a^2}}, \tag{5}$$

so that

$$1 - e^2 = \frac{c^2}{a^2}. \tag{6}$$

Then the radial distance from the rotation axis is given by

$$r(\delta) = a \left(1 + \frac{e^2}{1 - e^2} \sin^2 \delta\right)^{-1/2} \tag{7}$$

as a function of the Latitude δ.

The Surface Area and Volume of an oblate spheroid are

$$S = 2\pi a^2 + \pi \frac{c^2}{e} \ln\left(\frac{1+e}{1-e}\right) \tag{8}$$

$$V = \tfrac{4}{3}\pi a^2 c. \tag{9}$$

An oblate spheroid with its origin at a Focus has equation

$$r = \frac{a(1 - e^2)}{1 + e \cos \phi}. \tag{10}$$

Define k and expand up to Powers of e^6,

$$\begin{aligned} k &\equiv e^2(1 - e^2)^{-1} = e^2(1 + e^2 - 2e^4 + 6e^6 + \ldots) \\ &= e^2 + e^4 - 2e^6 + \ldots \end{aligned} \tag{11}$$

$$k^2 = e^4 + e^6 + \ldots \tag{12}$$

$$k^3 = e^6 + \ldots. \tag{13}$$

Expanding r in Powers of Ellipticity to e^6 therefore yields

$$\begin{aligned} \frac{r}{a} &= 1 - \tfrac{1}{2}(e^2 + e^4 - 2e^4 + 6e^6) \sin^2 \delta \\ &\quad + \tfrac{3}{4}(e^4 + e^6) \sin^4 \delta - \tfrac{15}{8} e^6 \sin^6 \delta + \ldots. \end{aligned} \tag{14}$$

In terms of Legendre Polynomials,

$$\begin{aligned} \frac{r}{a} &= (1 - \tfrac{1}{6}e^2 - \tfrac{11}{20}e^4 - \tfrac{103}{1680}e^6) \\ &\quad + (-\tfrac{1}{3}e^2 - \tfrac{5}{42}e^4 - \tfrac{3}{56}e^6)P_2 \\ &\quad + (\tfrac{3}{35}e^4 + \tfrac{57}{770}e^6)P_4 - \tfrac{5}{231}e^6 P_6 + \ldots. \end{aligned} \tag{15}$$

The Ellipticity may also be expressed in terms of the Oblateness (also called Flattening), denoted ϵ or f.

$$\epsilon \equiv \frac{a - c}{a} \tag{16}$$

$$c = a(1 - \epsilon) \tag{17}$$

$$c^2 = a^2(1 - \epsilon)^2 \tag{18}$$

$$(1 - \epsilon)^2 = 1 - e^2, \tag{19}$$

so

$$\epsilon = 1 - \sqrt{1 - e^2} \tag{20}$$

and

$$e^2 = 1 - (1 - \epsilon)^2 = 1 - (1 - 2\epsilon + \epsilon^2) = 2\epsilon - \epsilon^2 \tag{21}$$

$$r = a\left[1 + \frac{2\epsilon - \epsilon^2}{(1 - \epsilon)^2} \sin^2 \delta\right]^{-1/2}. \tag{22}$$

Define k and expand up to POWERS of ϵ^6

$$\begin{aligned} k &\equiv (2\epsilon-\epsilon)(1-\epsilon)^{-2} = (2\epsilon-\epsilon^2)(1+2\epsilon-6\epsilon^2+\ldots) \\ &= 2\epsilon+4\epsilon^4-12\epsilon^3-\epsilon^2-2\epsilon^3+\ldots \\ &= 2\epsilon+3\epsilon^2-14\epsilon^3+\ldots \qquad (23) \\ k^2 &= 4\epsilon^2+6\epsilon^3+\ldots \qquad (24) \\ k^3 &= 8\epsilon^3+\ldots. \qquad (25) \end{aligned}$$

Expanding r in POWERS of the OBLATENESS to ϵ^3 yields

$$\begin{aligned} \frac{r}{a} &= 1-\tfrac{1}{2}(2\epsilon+3\epsilon^2-14\epsilon^3)\sin^2\delta \\ &\quad +\tfrac{3}{4}(4\epsilon^2+6\epsilon^3)\sin^4\delta+8\epsilon^3\sin^6\delta+\ldots. \qquad (26) \end{aligned}$$

In terms of LEGENDRE POLYNOMIALS,

$$\begin{aligned} \frac{r}{a} &= (1-\tfrac{1}{3}\epsilon-\tfrac{2}{5}\epsilon^2-\tfrac{13}{105}\epsilon^3)+(-\tfrac{2}{3}\epsilon-\tfrac{1}{7}\epsilon^2-\tfrac{1}{21}\epsilon^3)P_2 \\ &\quad +(\tfrac{12}{35}\epsilon^2-\tfrac{96}{385}\epsilon^3)P_4-\tfrac{40}{231}\epsilon^3P_6+\ldots. \qquad (27) \end{aligned}$$

To find the projection of an oblate spheroid onto a PLANE, set up a coordinate system such that the z-AXIS is towards the observer, and the x-axis is in the PLANE of the page. The equation for an oblate spheroid is

$$r(\theta) = a\left[1+\frac{2\epsilon-\epsilon^2}{(1-\epsilon)^2}\cos^2\theta\right]^{-1/2}. \qquad (28)$$

Define

$$k \equiv \frac{2\epsilon-\epsilon^2}{(1-\epsilon)^2}, \qquad (29)$$

and $x \equiv \sin\theta$. Then

$$r(\theta) = a[1+k(1-x^2)]^{-1/2} = a(1+k-kx^2)^{-1/2}. \qquad (30)$$

Now rotate that spheroid about the x-axis by an ANGLE B so that the new symmetry axes for the spheroid are $x' \equiv x$, y', and z'. The projected height of a point in the $x=0$ PLANE on the y-axis is

$$\begin{aligned} y &= r(\theta)\cos(\theta-B) = r(\theta)(\cos\theta\cos B-\sin\theta\sin B) \\ &= r(\theta)(\sqrt{1-x^2}\cos B+x\sin B). \qquad (31) \end{aligned}$$

To find the highest projected point,

$$\frac{dy}{d\theta} = \frac{a\sin(B-\theta)}{(1+k\cos^2\theta)^{1/2}}+ak\frac{\cos(B-\theta)\cos\theta\sin\theta}{(1+k\cos^2\theta)^{3/2}} = 0. \qquad (32)$$

Simplifying,

$$\tan(B-\theta)(1+k\cos^2\theta)+k\cos\theta\sin\theta = 0. \qquad (33)$$

But

$$\begin{aligned} \tan(B-\theta) &= \frac{\tan B-\tan\theta}{1+\tan B\tan\theta} = \frac{\tan B-\frac{\sin\theta}{\sqrt{1-\sin^2\theta}}}{1+\tan B\frac{\sin\theta}{\sqrt{1-\sin^2\theta}}} \\ &= \frac{\sqrt{1-\sin^2\theta}\tan B-\sin\theta}{\sqrt{1-\sin^2\theta}+\tan B\sin\theta}. \qquad (34) \end{aligned}$$

Plugging (34) into (33),

$$\frac{\sqrt{1-x^2}\tan B-x}{\sqrt{1-x^2}+x\tan B}[1+k(1-x^2)]+kx\sqrt{1-x^2} = 0 \qquad (35)$$

and performing a number of algebraic simplifications

$$\begin{aligned} &(\sqrt{1-x^2}\tan B-x)(1+k-kx^2) \\ &\qquad +kx\sqrt{1-x^2}(\sqrt{1-x^2}+x\tan B) = 0 \qquad (36) \end{aligned}$$

$$\begin{aligned} &[(1+k)\sqrt{1-x^2}\tan B-kx^2\sqrt{1-x^2}\tan B \\ &\quad -x-kx+kx^3]+[kx(1-x^2)+kx^2\sqrt{1-x^2}\tan B] \qquad (37) \end{aligned}$$

$$(1+k)\tan B\sqrt{1-x^2}-kx(1-x^2)-x+kx(1-x^2) = 0 \qquad (38)$$

$$(1+k)\tan B\sqrt{1-x^2} = x \qquad (39)$$

$$(1+k)^2\tan^2 B(1-x^2) = x^2 \qquad (40)$$

$$x^2[1+(1+k)^2\tan^2 B] = (1+k)^2\tan^2 B \qquad (41)$$

finally gives the expression for x in terms of B and k,

$$x^2 = \frac{\tan^2 B(1+k)^2}{1+(1+k)^2\tan^2 B}. \qquad (42)$$

Combine (30) and (31) and plug in for x,

$$\begin{aligned} y &= a\frac{\sqrt{1-x^2}\cos B+x\sin B}{\sqrt{1+k-kx^2}} \\ &= a\frac{\cos B+(1+k)\frac{\sin^2 B}{\cos B}}{\sqrt{(1+k)[1+(1+k)\tan^2 B]}} \\ &= a\frac{\cos^2 B+(1+k)\sin^2 B}{\cos B\sqrt{(1+k)[1+(1+k)\tan^2 B]}}. \qquad (43) \end{aligned}$$

Now re-express k in terms of a and c, using $\epsilon \equiv 1-c/a$,

$$\begin{aligned} k &\equiv \frac{(2-\epsilon)\epsilon}{(1-\epsilon)^2} = \frac{\left(1+\frac{c}{a}\right)\left(1-\frac{c}{a}\right)}{\left(\frac{c}{a}\right)^2} \\ &= \frac{1-\left(\frac{c}{a}\right)^2}{\left(\frac{c}{a}\right)^2} = \left(\frac{a}{c}\right)^2-1, \qquad (44) \end{aligned}$$

so

$$1+k = \left(\frac{a}{c}\right)^2. \qquad (45)$$

Plug (44) and (45) into (43) to obtain the SEMIMINOR AXIS of the projected oblate spheroid,

$$\begin{aligned} c' &= a\frac{\cos^2 B+\left(\frac{a}{c}\right)^2\sin^2 B}{\cos B\sqrt{\left(\frac{a}{c}\right)^2\left[1+\left(\frac{a}{c}\right)^2\tan^2 B\right]}} \\ &= a\frac{\cos^2 B+\left(\frac{a}{c}\right)^2\sin^2 B}{\frac{a}{c}\sqrt{\cos^2 B+\left(\frac{a}{c}\right)^2\sin^2 B}} \\ &= c\sqrt{\cos^2 B+\left(\frac{a}{c}\right)^2\sin^2 B} = \sqrt{c^2\cos^2 B+a^2\sin^2 B} \\ &= a\sqrt{(1-\epsilon)^2\cos^2 B+\sin^2 B}. \qquad (46) \end{aligned}$$

We wish to find the equation for a spheroid which has been rotated about the $x \equiv x'$-axis by ANGLE B, then the z-axis by ANGLE P

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos B & \sin B \\ 0 & -\sin B & \cos B \end{bmatrix} \begin{bmatrix} \cos P & 0 & \sin P \\ 0 & 1 & 0 \\ -\sin P & 0 & \cos P \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$= \begin{bmatrix} \cos P & 0 & \sin P \\ -\sin B \sin P & \cos B & \sin B \cos P \\ -\cos B \sin P & -\sin B & \cos B \cos P \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \qquad (47)$$

Now, in the original coordinates (x', y', z'), the spheroid is given by the equation

$$\frac{x'^2}{a^2} + \frac{y'^2}{c^2} + \frac{z'^2}{a^2} = 1, \qquad (48)$$

which becomes in the new coordinates,

$$\frac{(x \cos P + y \sin P)^2}{a^2} + \frac{(-x \sin B \sin P + z \cos B + y \sin B \cos P)^2}{a^2} + \frac{(-x \cos B \sin P - z \sin B + y \cos B \cos P)^2}{c^2} = 1. \qquad (49)$$

Collecting COEFFICIENTS,

$$Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz = 1, \qquad (50)$$

where

$$A \equiv \frac{\cos^2 P + \sin^2 B \sin^2 P}{a^2} + \frac{\cos^2 B \sin^2 P}{c^2} \qquad (51)$$

$$B \equiv \frac{\sin^2 P + \sin^2 B \cos^2 P}{a^2} + \frac{\cos^2 B \cos^2 P}{c^2} \qquad (52)$$

$$C \equiv \frac{\cos^2 B}{a^2} + \frac{\sin^2 B}{c^2} \qquad (53)$$

$$D \equiv 2 \cos P \sin P \left(\frac{1 - \sin^2 B}{a^2} - \frac{\cos^2 B}{c^2} \right)$$

$$= 2 \cos P \sin P \cos^2 B \left(\frac{1}{a^2} - \frac{1}{c^2} \right) \qquad (54)$$

$$E \equiv 2 \sin B \cos B \sin P \left(\frac{1}{b^2} - \frac{1}{a^2} \right) \qquad (55)$$

$$F \equiv 2 \sin B \cos B \cos P \left(\frac{1}{a^2} - \frac{1}{b^2} \right). \qquad (56)$$

If we are interested in computing z, the radial distance from the symmetry axis of the spheroid (y) corresponding to a point

$$Cz^2 + (Ex + Fy)z + (Ax^2 + By^2 + Dxy - 1) = Cz^2 + G(x, y)z + H(x, y) = 0, \qquad (57)$$

where

$$G(x, y) \equiv Ex + Fy \qquad (58)$$

$$H(x, y) \equiv Ax^2 + By^2 + Dxy - 1. \qquad (59)$$

z can now be computed using the quadratic equation when (x, y) is given,

$$z = \frac{-G(x, y) \pm \sqrt{G^2(x, y) - 4CG(x, y)}}{2C}. \qquad (60)$$

If $P = 0$, then we have $\sin P = 0$ and $\cos P = 1$, so (51) to (56) and (58) to (59) become

$$A \equiv \frac{1}{a^2} \qquad (61)$$

$$B \equiv \frac{\sin^2 B}{a^2} + \frac{\cos^2 B}{b^2} \qquad (62)$$

$$C \equiv \frac{\cos^2 B}{a^2} + \frac{\sin^2 B}{b^2} \qquad (63)$$

$$D \equiv 0 \qquad (64)$$

$$E \equiv 0 \qquad (65)$$

$$F \equiv 2 \sin B \cos B \left(\frac{1}{a^2} - \frac{1}{b^2} \right) \qquad (66)$$

$$G(x, y) \equiv Fy = 2y \sin B \cos B \left(\frac{1}{a^2} - \frac{1}{b^2} \right) \qquad (67)$$

$$H(x, y) \equiv Ax^2 + By^2 - 1 = \frac{x^2}{a^2} + y^2 \left(\frac{\sin^2 B}{a^2} + \frac{\cos^2 B}{b^2} \right) - 1. \qquad (68)$$

see also DARWIN-DE SITTER SPHEROID, ELLIPSOID, OBLATE SPHEROIDAL COORDINATES, PROLATE SPHEROID, SPHERE, SPHEROID

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 131, 1987.

Oblate Spheroid Geodesic

The GEODESIC on an OBLATE SPHEROID can be computed analytically for a spheroid specified parametrically by

$$x = a \sin v \cos u \qquad (1)$$

$$y = a \sin v \sin u \qquad (2)$$

$$z = c \cos v, \qquad (3)$$

with $a > c$, although it is much more unwieldy than for a simple SPHERE. Using the first PARTIAL DERIVATIVES

$$\frac{\partial x}{\partial u} = -a \sin v \sin u \qquad \frac{\partial x}{\partial v} = a \cos v \cos u \qquad (4)$$

$$\frac{\partial y}{\partial u} = a \sin v \cos u \qquad \frac{\partial y}{\partial v} = a \cos v \sin u \qquad (5)$$

$$\frac{\partial z}{\partial u} = 0 \qquad \frac{\partial z}{\partial v} = -c \sin v, \qquad (6)$$

and second PARTIAL DERIVATIVES

$$\frac{\partial^2 x}{\partial u^2} = -a \sin v \cos u \qquad \frac{\partial^2 x}{\partial v^2} = -a \sin v \cos u \qquad (7)$$

$$\frac{\partial^2 y}{\partial u^2} = -a \sin v \sin u \qquad \frac{\partial^2 y}{\partial v^2} = -a \sin v \sin u \tag{8}$$

$$\frac{\partial^2 z}{\partial u^2} = 0 \qquad \frac{\partial^2 z}{\partial v^2} = -z \cos v, \tag{9}$$

gives the GEODESICS functions as

$$\begin{aligned} P &\equiv \left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2 + \left(\frac{\partial z}{\partial u}\right)^2 \\ &= a^2(\sin^2 v \cos^2 u + \sin^2 v \sin^2 u) \\ &= a^2 \sin^2 v && (10) \\ Q &\equiv \frac{\partial x}{\partial u}\frac{\partial x}{\partial v} + \frac{\partial y}{\partial u}\frac{\partial y}{\partial v} + \frac{\partial z}{\partial u}\frac{\partial z}{\partial v} = 0 && (11) \\ R &\equiv \left(\frac{\partial x}{\partial v}\right)^2 + \left(\frac{\partial y}{\partial v}\right)^2 + \left(\frac{\partial z}{\partial v}\right)^2 \\ &= a^2 + (c^2 - a^2)\sin^2 v = a^2(1 - e^2 \sin^2 v). && (12) \end{aligned}$$

Since $Q = 0$ and P and R are explicit functions of v only, we can use the special form of the GEODESIC equation.

$$\begin{aligned} u &= \int \sqrt{\frac{R}{P^2 - c_1{}^2 P}}\, dv \\ &= \int \sqrt{\frac{a^2(1 - e^2 \sin^2 v)}{a^4 \sin^4 v - c_1{}^2 a^2 \sin^2 v}}\, dv \\ &= c_1 \int \sqrt{\frac{1 - e^2 \sin^2 v}{\left(\frac{a}{c_1}\right)^2 \sin^2 v - 1}}\, \frac{dv}{\sin v}. && (13) \end{aligned}$$

Integrating gives

$$u = -c_1 \frac{e^2 F\left(\phi \middle| \frac{(d^2-1)e^2}{d^2-e^2}\right) - b^2 \Pi\left(d^2 - 1, \phi \middle| \frac{(d^2-1)e^2}{d^2-e^2}\right)}{\sqrt{d^2 - e^2}}, \tag{14}$$

where

$$d \equiv \frac{a}{c_1} \tag{15}$$

$$\cos\phi \equiv \frac{d \cos v}{\sqrt{d^2 - 1}}, \tag{16}$$

$F(\phi|m)$ is an ELLIPTIC INTEGRAL OF THE FIRST KIND with PARAMETER m, and $\Pi(\phi|m, k)$ is an ELLIPTIC INTEGRAL OF THE THIRD KIND.

GEODESICS other than MERIDIANS of an OBLATE SPHEROID undulate between two parallels with latitudes equidistant from the equator. Using the WEIERSTRAß SIGMA FUNCTION and WEIERSTRAß ZETA FUNCTION, the GEODESIC on the OBLATE SPHEROID can be written as

$$x + iy = \kappa \frac{\sigma(a + u)}{\sigma(u)\sigma(a)} e^{u[\eta - \zeta(\omega + a)]} \tag{17}$$

$$x - iy = \kappa \frac{\sigma(a - u)}{\sigma(u)\sigma(a)} e^{-u[\eta - \zeta(\omega + a)]} \tag{18}$$

$$z^2 = \lambda^2 \frac{\sigma(\omega'' + u)\sigma(\omega'' - u)}{\sigma^2(u)\sigma^2(a)} \tag{19}$$

(Forsyth 1960, pp. 108–109; Halphen 1886–1891).

The equation of the GEODESIC can be put in the form

$$d\phi = \frac{\sqrt{1 - e^2 \sin^2 v}\, \sin a}{\sqrt{\sin^2 v - \sin^2 a}\, \sin v}\, dv, \tag{20}$$

where a is the smallest value of v on the curve. Furthermore, the difference in longitude between points of highest and next lowest latitude on the curve is

$$\pi - 2\frac{\sqrt{1 - e^2 \sin^2 a}}{\sin a} \int_0^{\kappa} \frac{\operatorname{dn} u - \operatorname{dn}^2 u}{1 + \cot^2 a \operatorname{sn}^2 u}\, du, \tag{21}$$

where the MODULUS of the ELLIPTIC FUNCTION is

$$k = \frac{e \cos a}{\sqrt{1 - e^2 \sin^2 a}} \tag{22}$$

(Forsyth 1960, p. 446).

see also ELLIPSOID GEODESIC, OBLATE SPHEROID, SPHERE GEODESIC

References

Forsyth, A. R. *Calculus of Variations.* New York: Dover, 1960.

Halphen, G. H. *Traité des fonctions elliptiques et de leurs applications fonctions elliptiques, Vol. 2.* Paris: Gauthier-Villars, pp. 238–243, 1886–1891.

Oblate Spheroidal Coordinates

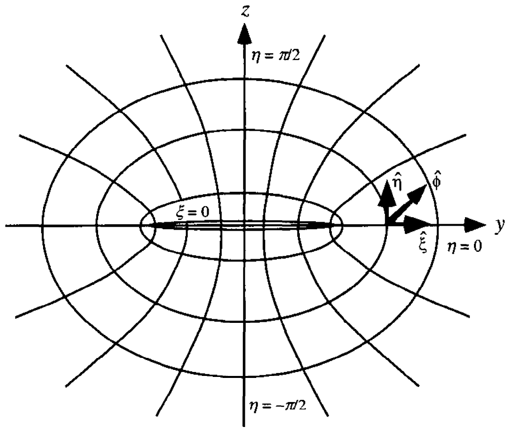

A system of CURVILINEAR COORDINATES in which two sets of coordinate surfaces are obtained by revolving the curves of the ELLIPTIC CYLINDRICAL COORDINATES about the y-AXIS which is relabeled the z-AXIS. The third set of coordinates consists of planes passing through this axis.

$$x = a \cosh\xi \cos\eta \cos\phi \tag{1}$$

$$y = a \cosh\xi \cos\eta \sin\phi \tag{2}$$

$$z = a \sinh\xi \sin\eta, \tag{3}$$

where $\xi \in [0, \infty)$, $\eta \in [-\pi/2, \pi/2]$, and $\phi \in [0, 2\pi)$. Arfken (1970) uses (u, v, φ) instead of (ξ, η, ϕ). The SCALE FACTORS are

$$h_\xi = a\sqrt{\sinh^2\xi + \sin^2\eta} \tag{4}$$

$$h_\eta = a\sqrt{\sinh^2\xi + \sin^2\eta} \tag{5}$$

$$h_\phi = a\cosh\xi\cos\eta. \tag{6}$$

The LAPLACIAN is

$$\begin{aligned}
\nabla^2 f &= \frac{1}{a^3(\sinh^2\xi + \sin^2\eta)\cosh\xi\cos\eta} \\
&\quad \times \left[\frac{\partial f}{\partial\xi}\left(a\cosh\xi\cos\eta\frac{\partial f}{\partial\xi}\right)\right. \\
&\quad \left. + \frac{\partial f}{\partial\eta}\left(a\cosh\xi\cos\eta\frac{\partial f}{\partial\eta}\right) + \frac{a^2(\sinh^2\xi + \sin^2\eta)}{a\cosh\xi\cos\eta}\frac{\partial^2 f}{\partial\phi^2}\right] \\
&= \frac{1}{a^3(\sinh^2\xi + \sin^2\eta)\cosh\xi\cos\eta}\left[a\sinh\xi\cos\eta\frac{\partial f}{\partial\xi}\right. \\
&\quad a\cosh\xi\cos\eta\frac{\partial^2 f}{\partial\xi^2} + a\sinh\xi\cos\eta\frac{\partial f}{\partial\eta} \\
&\quad \left. + a\cosh\xi\cos\eta\frac{\partial^2 f}{\partial\eta^2}\right] + \frac{1}{a^2(\sinh^2\xi + \sin^2\eta)}\frac{\partial^2 f}{\partial\phi^2} \\
&= \frac{1}{a^2(\sinh^2\xi + \sin^2\eta)}\left[\frac{1}{\cosh\xi}\frac{\partial}{\partial\xi}\left(\cosh\xi\frac{\partial f}{\partial\xi}\right)\right. \\
&\quad \left. + \frac{1}{\cos\eta}\frac{\partial}{\partial\eta}\left(\cos\eta\frac{\partial f}{\partial\eta}\right)\right] + \frac{1}{a^2(\cosh^2\xi + \cos^2\eta)}\frac{\partial^2 f}{\partial\phi^2}
\end{aligned} \tag{7}$$

$$\begin{aligned}
&= \frac{1}{\sin^2\eta + \sinh^2\xi}\left[(\operatorname{sech}^2\xi\tan^2\eta + \sec^2\tanh^2\xi)\frac{\partial^2}{\partial\phi^2}\right. \\
&\quad \left. + \tanh\xi\frac{\partial}{\partial\xi} + \frac{\partial^2}{\partial\xi^2} - \tan\eta\frac{\partial}{\eta} + \frac{\partial^2}{\eta^2}\right].
\end{aligned} \tag{8}$$

An alternate form useful for "two-center" problems is defined by

$$\xi_1 = \sinh\xi \tag{9}$$

$$\xi_1' = \cosh\xi \tag{10}$$

$$\xi_2 = \cos\eta \tag{11}$$

$$\xi_3 = \phi, \tag{12}$$

where $\xi_1 \in [1, \infty]$, $\xi_2 \in [-1, 1]$, and $\xi_3 \in [0, 2\pi)$. In these coordinates,

$$y = a\xi_1'\xi_2\sin\xi_3 \tag{13}$$

$$z = a\sqrt{({\xi_1'}^2 - 1)(1 - {\xi_2}^2)} \tag{14}$$

$$x = a\xi_1'\xi_2\cos\xi_3 \tag{15}$$

(Abramowitz and Stegun 1972). The SCALE FACTORS are

$$h_{\xi_1} = a\sqrt{\frac{{\xi_1}^2 - {\xi_2}^2}{{\xi_1}^2 - 1}} \tag{16}$$

$$h_{\xi_2} = a\sqrt{\frac{{\xi_1}^2 - {\xi_2}^2}{1 - {\xi_2}^2}} \tag{17}$$

$$h_{\xi_3} = a\xi\eta, \tag{18}$$

and the LAPLACIAN is

$$\begin{aligned}
\nabla^2 f = \frac{1}{a^2}&\left\{\frac{1}{{\xi_1}^2 + {\xi_2}^2}\frac{\partial}{\partial\xi_1}\left[({\xi_1}^2 + 1)\frac{\partial f}{\partial\xi_1}\right]\right. \\
&+ \frac{1}{{\xi_1}^2 + {\xi_2}^2}\frac{\partial}{\partial\xi_2}\left[(1 - {\xi_2}^2)\frac{\partial f}{\partial\xi_2}\right] \\
&\left. + \frac{1}{({\xi_1}^2 + 1)(1 - {\xi_2}^2)}\frac{\partial^2 f}{\partial{\xi_3}^2}\right\}.
\end{aligned} \tag{19}$$

The HELMHOLTZ DIFFERENTIAL EQUATION is separable.

see also HELMHOLTZ DIFFERENTIAL EQUATION—OBLATE SPHEROIDAL COORDINATES, LATITUDE, LONGITUDE, PROLATE SPHEROIDAL COORDINATES, SPHERICAL COORDINATES

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Definition of Oblate Spheroidal Coordinates." §21.2 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 752, 1972.

Arfken, G. "Prolate Spheroidal Coordinates (u, v, ϕ)." §2.11 in *Mathematical Methods for Physicists, 2nd ed.* Orlando, FL: Academic Press, pp. 107–109, 1970.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 663, 1953.

Oblate Spheroidal Wave Function

The wave equation in OBLATE SPHEROIDAL COORDINATES is

$$\begin{aligned}
\nabla^2\Phi + k^2\Phi &= \frac{\partial}{\partial\xi_1}\left[({\xi_1}^2 + 1)\frac{\partial\Phi}{\partial\xi_1}\right] \\
&+ \frac{\partial}{\partial\xi_2}\left[(1 - {\xi_2}^2)\frac{\partial\Phi}{\partial\xi_2}\right] + \frac{{\xi_1}^2 + {\xi_2}^2}{({\xi_1}^2 + 1)(1 - {x_2}^2)}\frac{\partial^2\Phi}{\partial\phi^2} \\
&+ c^2({\xi_1}^2 + {\xi_2}^2)\Phi = 0,
\end{aligned} \tag{1}$$

where

$$c \equiv \tfrac{1}{2}ak. \tag{2}$$

Substitute in a trial solution

$$\Phi = R_{mn}(c, \xi_1)S_{mn}(c, \xi_2)\genfrac{}{}{0pt}{}{\cos}{\sin}(m\phi). \tag{3}$$

The radial differential equation is

$$\begin{aligned}
&\frac{d}{d\xi_2}\left[(1 + {\xi_2}^2)\frac{d}{d\xi_2}S_{mn}(c, \xi_2)\right] \\
&\quad - \left(\lambda_{mn} - c^2{\xi_2}^2 + \frac{m^2}{1 + {\xi_2}^2}\right)R_{mn}(c, \xi_2) = 0,
\end{aligned} \tag{4}$$

and the angular differential equation is

$$\frac{d}{d\xi_2}\left[(1-{\xi_2}^2)\frac{d}{d\xi_2}S_{mn}(c,\xi_2)\right] - \left(\lambda_{mn} - c^2{\xi_2}^2 + \frac{m^2}{1-{\xi_2}^2}\right)R_{mn}(c,\xi_2) = 0 \quad (5)$$

(Abramowitz and Stegun 1972, pp. 753–755).

see also PROLATE SPHEROIDAL WAVE FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Spheroidal Wave Functions." Ch. 21 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 751–759, 1972.

Oblateness

see FLATTENING

Oblique Angle

An ANGLE which is not a RIGHT ANGLE.

Oblong Number

see PRONIC NUMBER

Obstruction

Obstruction theory studies the extentability of MAPS using algebraic GADGETS. While the terminology rapidly becomes technical and convoluted (as Iyanaga and Kawada note, "It is extremely difficult to discuss higher obstructions in general since they involve many complexities"), the ideas associated with obstructions are very important in modern ALGEBRAIC TOPOLOGY.

see also ALGEBRAIC TOPOLOGY, CHERN CLASS, EILENBERG-MAC LANE SPACE, STIEFEL-WHITNEY CLASS

References

Iyanaga, S. and Kawada, Y. (Eds.). "Obstructions." §300 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 948–950, 1980.

Obtuse Angle

An ANGLE greater than $\pi/2$ RADIANS (90°).

see also ACUTE ANGLE, OBTUSE TRIANGLE, RIGHT ANGLE, STRAIGHT ANGLE

Obtuse Triangle

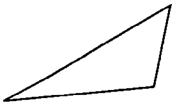

An obtuse triangle is a TRIANGLE in which one of the ANGLES is an OBTUSE ANGLE. (Obviously, only a single ANGLE in a TRIANGLE can be OBTUSE or it wouldn't be a TRIANGLE.) A triangle must be either obtuse, ACUTE, or RIGHT.

A famous problem is to find the chance that three points picked randomly in a PLANE are the VERTICES of an obtuse triangle (Eisenberg and Sullivan 1996). Unfortunately, the solution of the problem depends on the procedure used to pick the "random" points (Portnoy 1994). In fact, it is impossible to pick random variables which are uniformly distributed in the plane (Eisenberg and Sullivan 1996). Guy (1993) gives a variety of solutions to the problem. Woolhouse (1886) solved the problem by picking uniformly distributed points in the unit DISK, and obtained

$$P_2 = 1 - \left(\frac{4}{\pi^2} - \frac{1}{8}\right) = \frac{9}{8} - \frac{4}{\pi^2} = 0.719715\ldots. \quad (1)$$

The problem was generalized by Hall (1982) to n-D BALL TRIANGLE PICKING, and Buchta (1986) gave closed form evaluations for Hall's integrals.

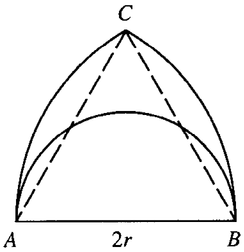

Lewis Carroll (1893) posed and gave another solution to the problem as follows. Call the longest side of a TRIANGLE AB, and call the DIAMETER $2r$. Draw arcs from A and B of RADIUS $2r$. Because the longest side of the TRIANGLE is defined to be AB, the third VERTEX of the TRIANGLE must lie within the region $ABCA$. If the third VERTEX lies within the SEMICIRCLE, the TRIANGLE is an obtuse triangle. If the VERTEX lies *on* the SEMICIRCLE (which will happen with probability 0), the TRIANGLE is a RIGHT TRIANGLE. Otherwise, it is an ACUTE TRIANGLE. The chance of obtaining an obtuse triangle is then the ratio of the AREA of the SEMICIRCLE to that of $ABCA$. The AREA of $ABCA$ is then twice the AREA of a SECTOR minus the AREA of the TRIANGLE.

$$A_{\text{whole figure}} = 2\left(\frac{4\pi r^2}{6}\right) - \sqrt{3}\,r^2 = r^2\left(\tfrac{4}{3}\pi - \sqrt{3}\right). \quad (2)$$

Therefore,

$$P = \frac{\frac{1}{2}\pi r^2}{r^2\left(\frac{4}{3}\pi - \sqrt{3}\right)} = \frac{3\pi}{8\pi - 6\sqrt{3}} = 0.63938\ldots. \quad (3)$$

Let the VERTICES of a triangle in n-D be NORMAL (GAUSSIAN) variates. The probability that a Gaussian triangle in n-D is obtuse is

$$\begin{aligned} P_n &= \frac{3\Gamma(n)}{\Gamma^2(\frac{1}{2}n)}\int_0^{1/3}\frac{x^{(n-2)/2}}{(1+x)^n}\,dx \\ &= \frac{3\Gamma(n)}{\Gamma^2(\frac{1}{2}n)2^{n-1}}\int_0^{\pi/3}\sin^{n-1}\theta\,d\theta \\ &= \frac{6\Gamma(n)\,{}_2F_1(\frac{1}{2}n, n, 1+\frac{1}{2}n; -\frac{1}{3})}{3^{n/2}n\Gamma^2(\frac{1}{2}n)}, \end{aligned} \quad (4)$$

where $\Gamma(n)$ is the GAMMA FUNCTION and ${}_2F_1(a,b;c;x)$ is the HYPERGEOMETRIC FUNCTION. For EVEN $n \equiv 2k$,

$$P_{2k} = 3\sum_{j=k}^{2k-1} \binom{2k-1}{j} \left(\frac{1}{4}\right)^j \left(\frac{3}{4}\right)^{2k-1-j} \tag{5}$$

(Eisenberg and Sullivan 1996). The first few cases are explicitly

$$P_2 = \tfrac{3}{4} = 0.75 \tag{6}$$

$$P_3 = 1 - \frac{3\sqrt{3}}{4\pi} = 0.586503\ldots \tag{7}$$

$$P_4 = \tfrac{15}{32} = 0.46875\ldots \tag{8}$$

$$P_5 = 1 - \frac{9\sqrt{3}}{8\pi} = 0.379755\ldots. \tag{9}$$

see also ACUTE ANGLE, ACUTE TRIANGLE, BALL TRIANGLE PICKING, OBTUSE ANGLE, RIGHT TRIANGLE, TRIANGLE

References
Buchta, C. "A Note on the Volume of a Random Polytope in a Tetrahedron." *Ill. J. Math.* **30**, 653–659, 1986.
Carroll, L. *Pillow Problems & A Tangled Tale.* New York: Dover, 1976.
Eisenberg, B. and Sullivan, R. "Random Triangles n Dimensions." *Amer. Math. Monthly* **103**, 308–318, 1996.
Guy, R. K. "There are Three Times as Many Obtuse-Angled Triangles as There are Acute-Angled Ones." *Math. Mag.* **66**, 175–178, 1993.
Hall, G. R. "Acute Triangles in the n-Ball." *J. Appl. Prob.* **19**, 712–715, 1982.
Portnoy, S. "A Lewis Carroll Pillow Problem: Probability on at Obtuse Triangle." *Statist. Sci.* **9**, 279–284, 1994.
Wells, D. G. *The Penguin Book of Interesting Puzzles.* London: Penguin Books, pp. 67 and 248–249, 1992.
Woolhouse, W. S. B. Solution to Problem 1350. *Mathematical Questions, with Their Solutions, from the Educational Times, 1.* London: F. Hodgson and Son, 49–51, 1886.

Ochoa Curve

The ELLIPTIC CURVE

$$3Y^2 = 2X^3 + 386X^2 + 256X - 58195,$$

given in Weierstraß form as

$$y^2 = x^3 - 440067x + 106074110.$$

The complete set of solutions to this equation consists of $(x, y) = (-761, 504)$, $(-745, 4520)$, $(-557, 13356)$, $(-446, 14616)$, $(-17, 10656)$, $(91, 8172)$, $(227, 4228)$, $(247, 3528)$, $(271, 2592)$, $(455, 200)$, $(499, 3276)$, $(523, 4356)$, $(530, 4660)$, $(599, 7576)$, $(751, 14112)$, $(1003, 25956)$, $(1862, 75778)$, $(3511, 204552)$, $(5287, 381528)$, $(23527, 3607272)$, $(64507, 16382772)$, $(100102, 31670478)$, and $(1657891, 2134685628)$ (Stroeker and de Weger 1994).

References
Guy, R. K. "The Ochoa Curve." *Crux Math.* **16**, 65–69, 1990.
Ochoa Melida, J. "La ecuacion diofántica $b_0y^3 - b_1y^2 + b_2y - b_3 = z^2$." *Gaceta Math.* 139–141, 1978.
Stroeker, R. J. and de Weger, B. M. M. "On Elliptic Diophantine Equations that Defy Thue's Method: The Case of the Ochoa Curve." *Experiment. Math.* **3**, 209–220, 1994.

Octacontagon

An 80-sided POLYGON.

Octadecagon

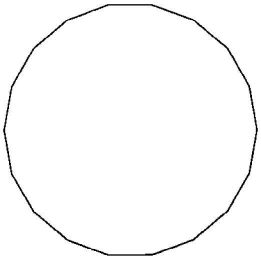

An 18-sided POLYGON, sometimes also called an OCTAKAIDECAGON.

see also POLYGON, REGULAR POLYGON, TRIGONOMETRY VALUES—$\pi/18$

Octagon

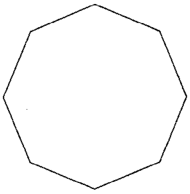

The regular 8-sided POLYGON. The INRADIUS r, CIRCUMRADIUS R, and AREA A can be computed directly from the formulas for a general regular POLYGON with side length s and $n = 8$ sides,

$$r = \tfrac{1}{2}s\cot\left(\frac{\pi}{8}\right) = \tfrac{1}{2}(1+\sqrt{2})s \tag{1}$$

$$R = \tfrac{1}{2}s\csc\left(\frac{\pi}{8}\right) = \tfrac{1}{2}\sqrt{4+2\sqrt{2}}\,s \tag{2}$$

$$A = \tfrac{1}{4}ns^2\cot\left(\frac{\pi}{8}\right) = 2(1+\sqrt{2})s^2. \tag{3}$$

see also OCTAHEDRON, POLYGON, REGULAR POLYGON, TRIGONOMETRY VALUES—$\pi/8$

Octagonal Number

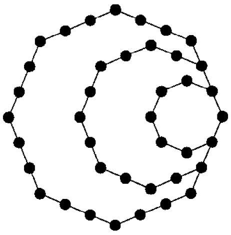

A POLYGONAL NUMBER of the form $n(3n-2)$. The first few are 1, 8, 21, 40, 65, 96, 133, 176, ... (Sloane's A000567). The GENERATING FUNCTION for the octagonal numbers is

$$\frac{x(5x+1)}{(1-x)^3} = x + 8x^2 + 21x^3 + 40x^4 + \ldots.$$

References
Sloane, N. J. A. Sequence A000567/M4493 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Octagram

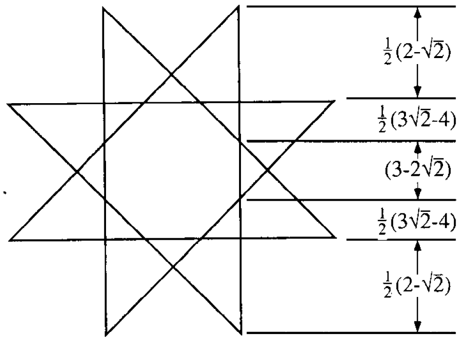
$\frac{1}{2}(2-\sqrt{2})$
$\frac{1}{2}(3\sqrt{2}-4)$
$(3-2\sqrt{2})$
$\frac{1}{2}(3\sqrt{2}-4)$
$\frac{1}{2}(2-\sqrt{2})$

The STAR POLYGON $\{8,3\}$.

Octahedral Graph

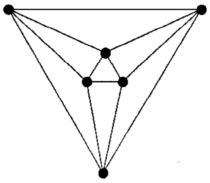

The POLYHEDRAL GRAPH having the topology of the OCTAHEDRON.

see also CUBICAL GRAPH, DODECAHEDRAL GRAPH, ICOSAHEDRAL GRAPH, OCTAHEDRON, TETRAHEDRAL GRAPH

Octahedral Group

The POINT GROUP of symmetries of the OCTAHEDRON, denoted O_h. It is also the symmetry group of the CUBE, CUBOCTAHEDRON, and TRUNCATED OCTAHEDRON. It has symmetry operations E, $8C_3$, $6C_4$, $6C_2$, $3C_2 = C_4^2$, i, $6S_4$, $8S_6$, $3\sigma_h$, and $6\sigma_4$ (Cotton 1990).

see also CUBE, CUBOCTAHEDRON, ICOSAHEDRAL GROUP, OCTAHEDRON, POINT GROUPS, TETRAHEDRAL GROUP, TRUNCATED OCTAHEDRON

References

Cotton, F. A. *Chemical Applications of Group Theory, 3rd ed.* New York: Wiley, p. 47–49, 1990.

Lomont, J. S. "Octahedral Group." §3.10.D in *Applications of Finite Groups.* New York: Dover, p. 81, 1987.

Octahedral Number

A FIGURATE NUMBER which is the sum of two consecutive PYRAMIDAL NUMBERS,

$$O_n = P_{n-1} + P_n = \tfrac{1}{3}n(2n^2+1).$$

The first few are 1, 6, 19, 44, 85, 146, 231, 344, 489, 670, 891, 1156, ... (Sloane's A005900). The GENERATING FUNCTION for the octahedral numbers is

$$\frac{x(x+1)^2}{(x-1)^4} = x + 6x^2 + 19x^3 + 44x^4 + \ldots.$$

see also TRUNCATED OCTAHEDRAL NUMBER

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 50, 1996.

Sloane, N. J. A. Sequence A005900/M4128 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Octahedron

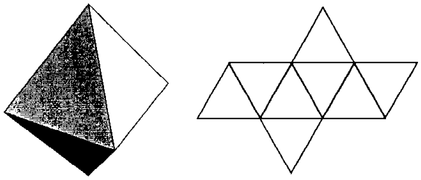

A PLATONIC SOLID (P_3) with six VERTICES, 12 EDGES, and eight equivalent EQUILATERAL TRIANGULAR faces ($8\{3\}$), given by the SCHLÄFLI SYMBOL $\{3,4\}$. It is also UNIFORM POLYHEDRON U_5 with the WYTHOFF SYMBOL $4\,|\,2\,3$. Its DUAL POLYHEDRON is the CUBE. Like the CUBE, it has the O_h OCTAHEDRAL GROUP of symmetries. The octahedron can be STELLATED to give the STELLA OCTANGULA.

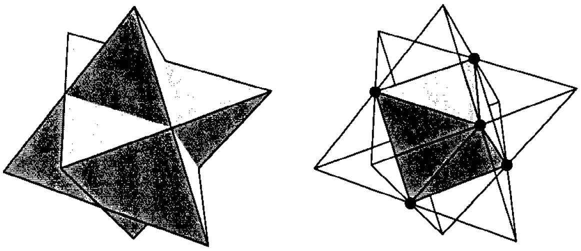

The solid bounded by the two TETRAHEDRA of the STELLA OCTANGULA (left figure) is an octahedron (right figure; Ball and Coxeter 1987).

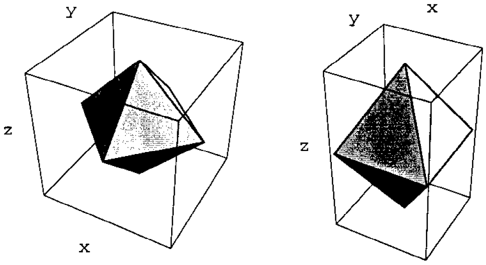

In one orientation (left figure), the VERTICES are given by $(\pm1,0,0)$, $(0,\pm1,0)$, $(0,0,\pm1)$. In another orientation (right figure), the vertices are $(\pm1,\pm1,0)$ and $(0,0,\pm\sqrt{3})$. In the latter, the constituent TRIANGLES are specified by

$$\begin{aligned}
T_1 &= \{(-1,-1,0),(1,-1,0),(0,0,\sqrt{3})\}\\
T_2 &= \{(-1,-1,0),(1,-1,0),(0,0,-\sqrt{3})\}\\
T_3 &= \{(-1,1,0),(1,1,0),(0,0,\sqrt{3})\}\\
T_4 &= \{(-1,1,0),(1,1,0),(0,0,-\sqrt{3})\}\\
T_5 &= \{(1,-1,0),(1,1,0),(0,0,\sqrt{3})\}\\
T_6 &= \{(-1,-1,0),(-1,1,0),(0,0,\sqrt{3})\}\\
T_7 &= \{(1,-1,0),(1,1,0),(0,0,-\sqrt{3})\}\\
T_8 &= \{(-1,-1,0),(-1,1,0),(0,0,-\sqrt{3})\}.
\end{aligned}$$

The face planes are $\pm x \pm y \pm z = 1$, so a solid octahedron is given by the equation

$$|x| + |y| + |z| \leq 1. \tag{1}$$

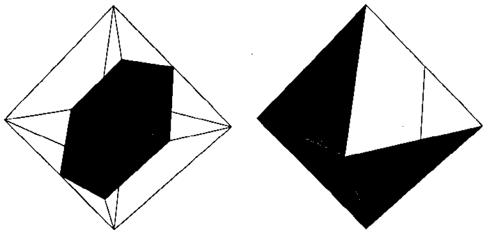

A plane PERPENDICULAR to a C_3 axis of an octahedron cuts the solid in a regular HEXAGONAL CROSS-SECTION (Holden 1991, pp. 22–23). Since there are four such axes, there are four possibly HEXAGONAL CROSS-SECTIONS. Faceted forms are the CUBOCTATRUNCATED CUBOCTAHEDRON and TETRAHEMIHEXAHEDRON.

Let an octahedron be length a on a side. The height of the top VERTEX from the square plane is also the CIRCUMRADIUS

$$R = \sqrt{a^2 - d^2}, \tag{2}$$

where

$$d = \tfrac{1}{2}\sqrt{2}\,a \tag{3}$$

is the diagonal length, so

$$R = \sqrt{a^2 - \tfrac{1}{2}a^2} = \tfrac{1}{2}\sqrt{2}\,a \approx 0.70710a. \tag{4}$$

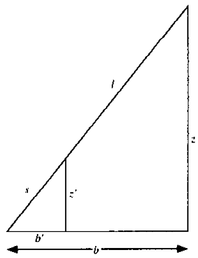

Now compute the INRADIUS.

$$\ell = \tfrac{1}{2}\sqrt{3}\,a \tag{5}$$

$$b = \tfrac{1}{2}a \tag{6}$$

$$s = \tfrac{1}{2}a\tan 30^\circ = \frac{a}{2\sqrt{3}}, \tag{7}$$

so

$$\frac{s}{\ell} = \frac{1}{2\sqrt{3}}\frac{2}{\sqrt{3}} = \tfrac{1}{3}. \tag{8}$$

Now use similar TRIANGLES to obtain

$$b' = \frac{s}{\ell}b = \tfrac{1}{6}a \tag{9}$$

$$z' = \frac{s}{\ell}z = \frac{a}{3\sqrt{2}} \tag{10}$$

$$x = b - b' = \tfrac{1}{2}a - \tfrac{1}{6}a = \tfrac{1}{3}a, \tag{11}$$

so the INRADIUS is

$$r = \sqrt{x^2 + z'^2} = a\sqrt{\tfrac{1}{9} + \tfrac{1}{18}} = \tfrac{1}{6}\sqrt{6}\,a \approx 0.40824a. \tag{12}$$

The INTERRADIUS is

$$\rho = \tfrac{1}{2}a = 0.5a. \tag{13}$$

The AREA of one face is the AREA of an EQUILATERAL TRIANGLE

$$A = \tfrac{1}{4}\sqrt{3}\,a^2. \tag{14}$$

The volume is two times the volume of a square-base pyramid,

$$V = 2(\tfrac{1}{3}a^2R) = 2(\tfrac{1}{3})(a^2)(\tfrac{1}{2}\sqrt{2}\,a) = \tfrac{1}{3}\sqrt{2}\,a^3. \tag{15}$$

The DIHEDRAL ANGLE is

$$\alpha = \cos^{-1}(-\tfrac{1}{3}) \approx 70.528779^\circ. \tag{16}$$

see also OCTAHEDRAL GRAPH, OCTAHEDRAL GROUP, OCTAHEDRON 5-COMPOUND, STELLA OCTANGULA, TRUNCATED OCTAHEDRON

References
Davie, T. "The Octahedron." `http://www.dcs.st-and.ac.uk/~ad/mathrecs/polyhedra/octahedron.html`.
Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, 1991.

Octahedron 5-Compound

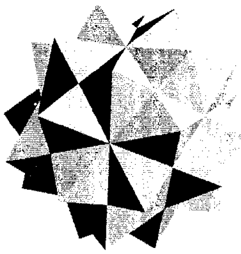

A POLYHEDRON COMPOUND composed of five OCTAHEDRA occupying the VERTICES of an ICOSAHEDRON. The 30 VERTICES of the compound form an ICOSIDODECAHEDRON (Ball and Coxeter 1987).

see also ICOSIDODECAHEDRON, OCTAHEDRON, POLYHEDRON COMPOUND

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 135 and 137, 1987.
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 137–138, 1989.
Wenninger, M. J. *Polyhedron Models.* New York: Cambridge University Press, p. 43, 1989.

Octahemioctacron

The DUAL POLYHEDRON of the OCTAHEMIOCTAHEDRON.

Octahemioctahedron

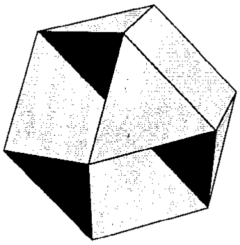

The UNIFORM POLYHEDRON U_3, also called the OCTATETRAHEDRON, whose DUAL POLYHEDRON is the OCTAHEMIOCTACRON. It has WYTHOFF SYMBOL $\frac{3}{2}\,3\,|\,3$. Its faces are $8\{3\}+4\{6\}$. It is a FACETED CUBOCTAHEDRON. For unit edge length, its CIRCUMRADIUS is

$$R = 1.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 103, 1989.

Octakaidecagon

see OCTADECAGON

Octal

The base 8 notational system for representing REAL NUMBERS. The digits used are 0, 1, 2, 3, 4, 5, 6, and 7, so that 8_{10} (8 in base 10) is represented as 10_8 ($10 = 1\cdot 8^1 + 0\cdot 8^0$) in base 8.

see also BASE (NUMBER), BINARY, DECIMAL, HEXADECIMAL, QUATERNARY, TERNARY

References

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 9–10, 1991.

Weisstein, E. W. "Bases." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Bases.m`.

Octant

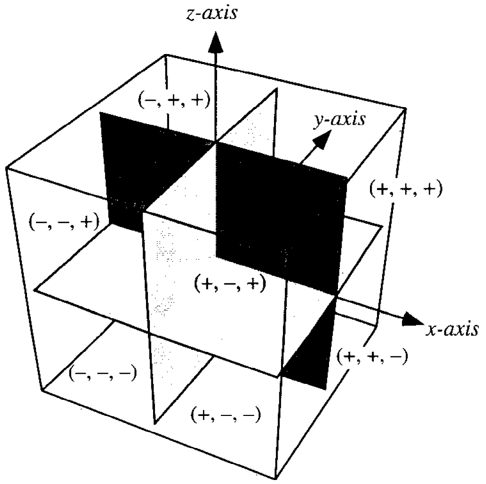

One of the eight regions of SPACE defined by the eight possible combinations of SIGNS $(\pm,\pm,\pm)$ for x, y, and z.

see also QUADRANT

Octatetrahedron

see OCTAHEMIOCTAHEDRON

Octic Surface

An ALGEBRAIC SURFACE of degree eight. The maximum number of ORDINARY DOUBLE POINTS known to exist on an octic surface is 168 (the ENDRASS OCTICS), although the rigorous upper bound is 174.

see also ALGEBRAIC SURFACE, ENDRASS OCTIC

Octillion

In the American system, 10^{27}.

see also LARGE NUMBER

Octodecillion

In the American system, 10^{57}.

see also LARGE NUMBER

Octonion

see CAYLEY NUMBER

Odd Function

An odd function is a function for which $f(x) = -f(-x)$. An EVEN FUNCTION times an odd function is odd.

Odd Number

An INTEGER of the form $N = 2n+1$, where n is an INTEGER. The odd numbers are therefore $\ldots, -3, -1, 1, 3, 5, 7, \ldots$ (Sloane's A005408), which are also the GNOMIC NUMBERS. The GENERATING FUNCTION for the odd numbers is

$$\frac{x(1+x)}{(x-1)^2} = x + 3x^2 + 5x^3 + 7x^4 + \ldots.$$

Since the odd numbers leave a remainder of 1 when divided by two, $N \equiv 1 \pmod 2$ for odd N. Integers which are not odd are called EVEN.

see also EVEN NUMBER, GNOMIC NUMBER, NICOMACHUS'S THEOREM, ODD NUMBER THEOREM, ODD PRIME

References

Sloane, N. J. A. Sequence A005408/M2400 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Odd Number Theorem

The sum of the first n ODD NUMBERS is a SQUARE NUMBER:

$$\sum_{k=1}^{n}(2k-1) = 2\sum_{k=1}^{n} k - \sum_{k=1}^{n} 1 = 2\left[\frac{n(n+1)}{2}\right] - n$$
$$= n(n+1) - n = n^2.$$

see also NICOMACHUS'S THEOREM

Odd Order Theorem

see FEIT-THOMPSON THEOREM

Odd Prime

Any PRIME NUMBER other than 2 (which is the only EVEN PRIME).

see also PRIME NUMBER

Odd Sequence

A SEQUENCE of n 0s and 1s is called an odd sequence if each of the n SUMS $\sum_{i=1}^{n-k} a_i a_{i+k}$ for $k = 0, 1, \ldots, n-1$.

References

Guy, R. K. "Odd Sequences." §E38 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 238–239, 1994.

Odds

Betting odds are written in the form $r : s$ ("r to s") and correspond to the probability of winning $P = s/(r+s)$. Therefore, given a probability P, the odds of winning are $(1/P) - 1 : 1$.

see also FRACTION, RATIO, RATIONAL NUMBER

References

Kraitchik, M. "The Horses." §6.17 in *Mathematical Recreations.* New York: W. W. Norton, pp. 134–135, 1942.

ODE

see ORDINARY DIFFERENTIAL EQUATION

Offset Rings

see SURFACE OF REVOLUTION

Ogive

Any cumulative frequency curve.

see also HISTOGRAM

References

Kenney, J. F. and Keeping, E. S. "Ogive Curves." §2.7 in *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, pp. 29–31, 1962.

Oldknow Points

The PERSPECTIVE CENTERS of a triangle and the TANGENTIAL TRIANGLES of its inner and outer SODDY CIRCLES, given by

$$Ol = I + 2Ge$$
$$Ol' = I - 2Ge,$$

where I is the INCENTER and Ge is the GERGONNE POINT.

see also GERGONNE POINT, INCENTER, PERSPECTIVE CENTER, SODDY CIRCLES, TANGENTIAL TRIANGLE

References

Oldknow, A. "The Euler-Gergonne-Soddy Triangle of a Triangle." *Amer. Math. Monthly* **103**, 319–329, 1996.

Omega Constant

$$W(1) \equiv 0.5671432904\ldots, \tag{1}$$

where $W(x)$ is LAMBERT'S W-FUNCTION. It is available in *Mathematica*® (Wolfram Research, Champaign, IL) using the function `ProductLog[1]`. $W(1)$ can be considered a sort of "GOLDEN RATIO" for exponentials since

$$\exp[-W(1)] = W(1), \tag{2}$$

giving

$$\ln\left[\frac{1}{W(1)}\right] = W(1). \tag{3}$$

see also GOLDEN RATIO, LAMBERT'S W-FUNCTION

References

Corless, R. M.; Gonnet, G. H.; Hare, D. E. G.; and Jeffrey, D. J. "On Lambert's W Function." `ftp://watdragon.uwaterloo.ca/cs-archive/CS-93-03/W.ps.Z`.

Plouffe, S. "The Omega Constant or $W(1)$." `http://lacim.uqam.ca/piDATA/omega.txt`.

Omega Function

see LAMBERT'S W-FUNCTION

Omino

see POLYOMINO

Omnific Integer

The appropriate notion of INTEGER for SURREAL NUMBERS.

O'Nan Group

The SPORADIC GROUP *O'N*.

References

Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas/ON.html`.

Onduloid

see UNDULOID

One

see 1

One-Form

A linear, real-valued FUNCTION of VECTORS such that $\omega^1(\mathbf{v}) \mapsto \mathbb{R}$. VECTORS and one-forms are DUAL to each other because VECTORS are CONTRAVARIANT ("KETS": $|\psi\rangle$) and one-forms are COVARIANT VECTORS ("BRAS": $\langle\phi|$), so

$$\omega^1(\mathbf{v}) \equiv \mathbf{v}(\omega^1) \equiv \left\langle \omega^1, \mathbf{v} \right\rangle = \langle\phi|\psi\rangle .$$

The operation of applying the one-form to a VECTOR $\omega^1(\mathbf{v})$ is called CONTRACTION.

see also ANGLE BRACKET, BRA, DIFFERENTIAL k-FORM, KET

One-Mouth Theorem

Except for convex polygons, every SIMPLE POLYGON has at least one MOUTH.

see also MOUTH, PRINCIPAL VERTEX, TWO-EARS THEOREM

References

Toussaint, G. "Anthropomorphic Polygons." *Amer. Math. Monthly* **122**, 31–35, 1991.

One-Ninth Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let $\lambda_{m,n}$ be CHEBYSHEV CONSTANTS. Schönhage (1973) proved that

$$\lim_{n\to\infty} (\lambda_{0,n})^{1/n} = \tfrac{1}{3}. \tag{1}$$

It was conjectured that

$$\Lambda \equiv \lim_{n\to\infty} (\lambda_{n,n})^{1/n} = \tfrac{1}{9}. \tag{2}$$

Carpenter *et al.* (1984) obtained

$$\Lambda = 0.1076539192\ldots \tag{3}$$

numerically. Gonchar and Rakhmanov (1980) showed that the limit exists and disproved the 1/9 conjecture, showing that Λ is given by

$$\Lambda = \exp\left[-\frac{\pi K(\sqrt{1-c^2}\,)}{K(c)}\right], \tag{4}$$

where K is the complete ELLIPTIC INTEGRAL OF THE FIRST KIND, and $c = 0.9089085575485414\ldots$ is the PARAMETER which solves

$$K(k) = 2E(k), \tag{5}$$

and E is the complete ELLIPTIC INTEGRAL OF THE SECOND KIND. This gives the value for Λ computed by Carpenter *et al.* (1984) Λ is also given by the unique POSITIVE ROOT of

$$f(z) = \tfrac{1}{8}, \tag{6}$$

where

$$f(z) \equiv \sum_{j=1}^{\infty} a_j z^j \tag{7}$$

and

$$a_j = \left| \sum_{d|j} (-1)^d d \right| \tag{8}$$

(Gonchar and Rakhmanov 1980). a_j may also be computed by writing j as

$$j = 2^m {p_1}^{m_1} {p_2}^{m_2} \cdots {p_k}^{m_k}, \tag{9}$$

where $m \geq 0$ and $m_i \geq 1$, then

$$a_j = |2^{m+1} - 3| \frac{{p_1}^{m_1+1} - 1}{p_1 - 1} \frac{{p_2}^{m_2+1} - 1}{p_2 - 1} \cdots \frac{{p_k}^{m_k+1} - 1}{p_k - 1} \tag{10}$$

(Gonchar 1990). Yet another equation for Λ is due to Magnus (1986). Λ is the unique solution with $x \in (0,1)$ of

$$\sum_{k=0}^{\infty} (2k+1)^2 (-x)^{k(k+1)/2} = 0, \tag{11}$$

an equation which had been studied and whose root had been computed by Halphen (1886). It has therefore been suggested (Varga 1990) that the constant be called the HALPHEN CONSTANT. $1/\Lambda$ is sometimes called VARGA'S CONSTANT.

see also CHEBYSHEV CONSTANTS, HALPHEN CONSTANT, VARGA'S CONSTANT

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/onenin/onenin.html`.

Carpenter, A. J.; Ruttan, A.; and Varga, R. S. "Extended Numerical Computations on the '1/9' Conjecture in Rational Approximation Theory." In *Rational approximation and interpolation (Tampa, Fla., 1983)* (Ed. P. R. Graves-Morris, E. B. Saff, and R. S. Varga). New York: Springer-Verlag, pp. 383–411, 1984.

Cody, W. J.; Meinardus, G.; and Varga, R. S. "Chebyshev Rational Approximations to e^{-x} in $[0, +\infty)$ and Applications to Heat-Conduction Problems." *J. Approx. Th.* **2**, 50–65, 1969.

Dunham, C. B. and Taylor, G. D. "Continuity of Best Reciprocal Polynomial Approximation on $[0, \infty)$." *J. Approx. Th.* **30**, 71–79, 1980.

Gonchar, A. A. "Rational Approximations of Analytic Functions." *Amer. Math. Soc. Transl. Ser. 2* **147**, 25–34, 1990.

Gonchar, A. A. and Rakhmanov, E. A. "Equilibrium Distributions and Degree of Rational Approximation of Analytic Functions." *Math. USSR Sbornik* **62**, 305–348, 1980.

Magnus, A. P. "On Freud's Equations for Exponential Weights, Papers Dedicated to the Memory of Géza Freud." *J. Approx. Th.* **46**, 65–99, 1986.

Rahman, Q. I. and Schmeisser, G. "Rational Approximation to the Exponential Function." In *Padé and Rational Approximation, (Proc. Internat. Sympos., Univ. South Florida, Tampa, Fla., 1976)* (Ed. E. B. Saff and R. S. Varga). New York: Academic Press, pp. 189–194, 1977.

Schönhage, A. "Zur rationalen Approximierbarkeit von e^{-x} über $[0, \infty)$." *J. Approx. Th.* **7**, 395–398, 1973.

Varga, R. S. *Scientific Computations on Mathematical Problems and Conjectures.* Philadelphia, PA: SIAM, 1990.

One-to-One

Let f be a FUNCTION defined on a SET S and taking values in a set T. Then f is said to be one-to-one (a.k.a. an INJECTION or EMBEDDING) if, whenever $f(x) = f(y)$, it must be the case that $x = y$. In other words, f is one-to-one if it MAPS distinct objects to distinct objects.

If the function is a linear OPERATOR which assigns a unique MAP to each value in a VECTOR SPACE, it is called one-to-one. Specifically, given a VECTOR SPACE $\mathbb{V}$ with $\mathbf{X}, \mathbf{Y} \in \mathbb{V}$, then a TRANSFORMATION T defined on $\mathbb{V}$ is one-to-one if $T(\mathbf{X}) \neq T(\mathbf{Y})$ for all $\mathbf{X} \neq \mathbf{Y}$.

see also BIJECTION, ONTO

One-Way Function

Consider straight-line algorithms over a FINITE FIELD with q elements. Then the ϵ-straight line complexity $C_\epsilon(\phi)$ of a function ϕ is defined as the length of the shortest straight-line algorithm which computes a function f such that $f(x) = x$ is satisfied for at least $(1-\epsilon)q$ elements of F. A function ϕ is straight-line "one way" of range $0 \leq \delta \leq 1$ if ϕ satisfies the properties:

1. There exists an infinite set S of finite fields such that ϕ is defined in every $F \in S$ and ϵ is ONE-TO-ONE in every $F \in S$.
2. For every ϵ such that $0 \leq \epsilon \leq \delta$, $C_\epsilon(\phi^{-1})$ tends to infinity as the cardinality q of F approaches infinity.
3. For every ϵ such that $0 \leq \epsilon \leq \delta$, the "work function" η satisfies

$$\eta \equiv \liminf_{q\to\infty} \eta \equiv \liminf_{q\to\infty} \frac{\ln C\epsilon(\phi^{-1}) - \ln C\epsilon(\phi)}{\ln C\epsilon(\phi)} > 1.$$

It is not known if there is a one-way function with work factor $\eta > (\ln q)^3$.

References

Ziv, J. "In Search of a One-Way Function" §4.1 in *Open Problems in Communication and Computation* (Ed. T. M. Cover and B. Gopinath). New York: Springer-Verlag, pp. 104–105, 1987.

Only Critical Point in Town Test

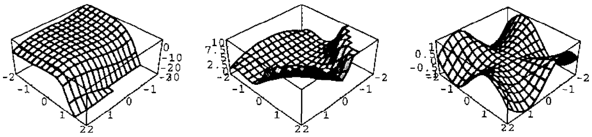

If there is only one CRITICAL POINT at an EXTREMUM, the CRITICAL POINT must be the EXTREMUM for functions of one variable. There are exceptions for two variables, but none of degree ≤ 4. Such exceptions include

$$z = 3xe^y - x^3 - e^{3y}$$

$$z = x^2(1+y)^3 + y^2$$

$$z = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} & \text{for } (x,y) \neq (0,0) \\ 0 & \text{for } (x,y) = (0,0) \end{cases}$$

(Wagon 1991). This latter function has discontinuous z_{xy} and z_{yx}, and $z_{yx}(0,0) = 1$ and $z_{xy}(0,0) = 1$.

References

Ash, A. M. and Sexton, H. "A Surface with One Local Minimum." *Math. Mag.* **58**, 147–149, 1985.

Calvert, B. and Vamanamurthy, M. K. "Local and Global Extrema for Functions of Several Variables." *J. Austral. Math. Soc.* **29**, 362–368, 1980.

Davies, R. Solution to Problem 1235. *Math. Mag.* **61**, 59, 1988.

Wagon, S. "Failure of the Only-Critical-Point-in-Town Test." §3.4 in *Mathematica in Action.* New York: W. H. Freeman, pp. 87–91 and 228, 1991.

Onto

Let f be a FUNCTION defined on a SET S and taking values in a set T. Then f is said to be onto (a.k.a. a SURJECTION) if, for any $t \in T$, there exists an $s \in S$ for which $t = f(s)$.

Let the function be an OPERATOR which MAPS points in the DOMAIN to every point in the RANGE and let $\mathbb{V}$ be a VECTOR SPACE with $\mathbf{X}, \mathbf{Y} \in \mathbb{V}$. Then a TRANSFORMATION T defined on $\mathbb{V}$ is onto if there is an $\mathbf{X} \in \mathbb{V}$ such that $T(\mathbf{X}) = \mathbf{Y}$ for all $\mathbf{Y}$.

see also BIJECTION, ONE-TO-ONE

Open Disk

An n-D open disk of RADIUS r is the collection of points of distance less than r from a fixed point in EUCLIDEAN n-space.

see also CLOSED DISK, DISK

Open Interval

An INTERVAL which does not include its LIMIT POINTS, denoted (a, b).

see also CLOSED INTERVAL, HALF-CLOSED INTERVAL

Open Map

A MAP which sends OPEN SETS to OPEN SETS.

see also OPEN MAPPING THEOREM

Open Mapping Theorem

There are several flavors of this theorem.

1. A continuous surjective linear mapping between Banach Spaces is an Open Map.
2. A nonconstant Analytic Function on a Domain D is an Open Map.

References

Zeidler, E. *Applied Functional Analysis: Applications to Mathematical Physics.* New York: Springer-Verlag, 1995.

Open Set

A Set is open if every point in the set has a Neighborhood lying in the set. An open set of Radius r and center $\mathbf{x}_0$ is the set of all points $\mathbf{x}$ such that $|\mathbf{x}-\mathbf{x}_0|<r$, and is denoted $D_r(\mathbf{x}_0)$. In 1-space, the open set is an Open Interval. In 2-space, the open set is a Disk. In 3-space, the open set is a Ball.

More generally, given a Topology (consisting of a Set X and a collection of Subsets T), a Set is said to be open if it is in T. Therefore, while it is not possible for a set to be both finite and open in the Topology of the Real Line (a single point is a Closed Set), it is possible for a more general topological Set to be both finite and open.

The complement of an open set is a Closed Set. It is possible for a set to be neither open nor Closed, e.g., the interval $(0,1]$.

see also Ball, Closed Set, Empty Set, Open Interval

Operad

A system of parameter chain complexes used for Multiplication on differential Graded Algebras up to Homotopy.

Operand

A mathematical object upon which an Operator acts. For example, in the expression 1×2, the Multiplication Operator acts upon the operands 1 and 2.

see also Operad, Operator

Operational Mathematics

The theory and applications of Laplace Transforms and other Integral Transforms.

References

Churchill, R. V. *Operational Mathematics, 3rd ed.* New York: McGraw-Hill, 1958.

Operations Research

A branch of mathematics which encompasses many diverse areas of minimization and optimization. Bronson (1982) describes operations research as being "concerned with the efficient allocation of scarce resources." It includes the Calculus of Variations, Control Theory, Convex Optimization Theory, Decision Theory, Game Theory, Linear Programming, Markov Chains, network analysis, Optimization Theory, queuing systems, etc. The more modern term for operations research is Optimization Theory.

see also Calculus of Variations, Control Theory, Convex Optimization Theory, Decision Theory, Game Theory, Linear Programming, Markov Chain, Optimization Theory, Queue

References

Bronson, R. *Schaum's Outline of Theory and Problems of Operations Research.* New York: McGraw-Hill, 1982.

Hiller, F. S. and Lieberman, G. J. *Introduction to Operations Research, 5th ed.* New York: McGraw-Hill, 1990.

Trick, M. "Michael Trick's Operations Research Page." `http://mat.gsia.cmu.edu`

Operator

An operator $A: f^{(n)}(I)\mapsto f(I)$ assigns to every function $f\in f^{(n)}(I)$ a function $A(f)\in f(I)$. It is therefore a mapping between two Function Spaces. If the range is on the Real Line or in the Complex Plane, the mapping is usually called a Functional instead.

see also Abstraction Operator, Adjoint Operator, Antilinear Operator, Biharmonic Operator, Binary Operator, Casimir Operator, Convective Operator, d'Alembertian Operator, Difference Operator, Functional Analysis, Hecke Operator, Hermitian Operator, Identity Operator, Laplace-Beltrami Operator, Linear Operator, Operand, Perron-Frobenius Operator, Projection Operator, Rotation Operator, Scattering Operator, Self-Adjoint Operator, Spectrum (Operator), Theta Operator, Wave Operator

References

Gohberg, I.; Lancaster, P.; and Shivakuar, P. N. (Eds.). *Recent Developments in Operator Theory and Its Applications.* Boston, MA: Birkhäuser, 1996.

Hutson, V. and Pym, J. S. *Applications of Functional Analysis and Operator Theory.* New York: Academic Press, 1980.

Optimization Theory

see Operations Research

Or

A term in Logic which yields True if any one of a sequence conditions is True, and False if *all* conditions are False. A OR B is denoted $A|B$, $A+B$, or $A\vee B$. The symbol $\vee$ derives from the first letter of the Latin word "vel" meaning "or." The Binary OR operator has the following Truth Table.

A	B	$A\vee B$
F	F	F
F	T	T
T	F	T
T	T	T

A product of ORs is called a DISJUNCTION and is denoted

$$\bigvee_{k=1}^{n} A_k.$$

Two BINARY numbers can have the operation OR performed bitwise. This operation is sometimes denoted $A||B$.

see also AND, BINARY OPERATOR, LOGIC, NOT, PREDICATE, TRUTH TABLE, UNION, XOR

Orbifold

The object obtained by identifying any two points of a MAP which are equivalent under some symmetry of the MAP'S GROUP.

Orbison's Illusion

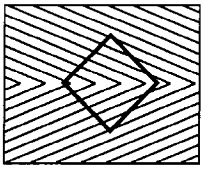

The illusion illustrated above in which the bounding RECTANGLE and inner SQUARE both appear distorted.

see also ILLUSION, MÜLLER-LYER ILLUSION, PONZO'S ILLUSION, VERTICAL-HORIZONTAL ILLUSION

References

Fineman, M. *The Nature of Visual Illusion.* New York: Dover, p. 153, 1996.

Orbit (Group)

Given a PERMUTATION GROUP G on a set S, the orbit of an element $s \in S$ is the subset of S consisting of elements to which some element G can send s.

Orbit (Map)

The SEQUENCE generated by repeated application of a MAP. The MAP is said to have a closed orbit if it has a finite number of elements.

see also DYNAMICAL SYSTEM, SINK (MAP)

Orchard-Planting Problem

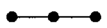

$n = 3, r = 1$

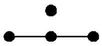

$n = 4, r = 1$

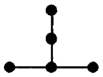

$n = 5, r = 2$

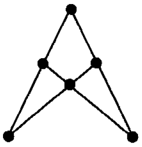

$n = 6, r = 4$

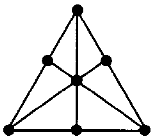

$n = 7, r = 6$

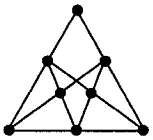

$n = 8, r = 7$

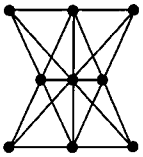

$n = 9, r = 10$

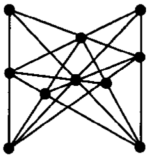

$n = 10, r = 12$

Also known as the TREE-PLANTING PROBLEM. Plant n trees so that there will be r straight rows with k trees in each row. The following table gives max(r) for various k. $k = 3$ is Sloane's A003035 and $k = 4$ is Sloane's A006065.

n	$k=3$	$k=4$	$k=5$
3	1	—	—
4	1	1	—
5	2	1	1
6	4	1	1
7	6	2	1
8	7	2	1
9	10	3	2
10	12	5	2
11	16	6	2
12	19	7	3
13	[22, 24]	≥ 9	3
14	[26, 27]	≥ 10	4
15	[31, 32]	≥ 12	≥ 6
16	37	≥ 15	≥ 6
17	[40, 42]	≥ 15	≥ 7
18	[46, 48]	≥ 18	≥ 9
19	[52, 54]	≥ 19	≥ 10
20	[57, 60]	≥ 21	≥ 11
21	[64, 67]		
22	[70, 73]		
23	[77, 81]		
24	[85, 88]		
25	[92, 96]		

Sylvester showed that

$$r(k=3) \geq \left\lfloor \tfrac{1}{6}(n-1)(n-2) \right\rfloor,$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION (Ball and Coxeter 1987). Burr, Grünbaum and Sloane (1974) have shown using cubic curves that

$$r(k=3) \leq 1 + \left\lfloor \tfrac{1}{6}n(n-3) \right\rfloor,$$

except for $n = 7$, 11, 16, and 19, and conjecture that the inequality is an equality with the exception of the preceding cases. For $n \geq 4$,

$$r(k=3) \geq \left\lfloor \tfrac{1}{3}\left[\tfrac{1}{2}n(n-1) - \left\lceil \tfrac{3}{7}n \right\rceil\right]\right\rfloor,$$

where $\lceil x \rceil$ is the CEILING FUNCTION.

see also ORCHARD VISIBILITY PROBLEM

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 104–105 and 129, 1987.
Burr, S. A. "Planting Trees." In *The Mathematical Gardner* (Ed. David Klarner). Boston, MA: Prindle, Weber, and Schmidt, pp. 90–99, 1981.
Dudeney, H. E. Problem 435 in *536 Puzzles & Curious Problems.* New York: Scribner, 1967.
Dudeney, H. E. *The Canterbury Puzzles and Other Curious Problems, 7th ed.* London: Thomas Nelson and Sons, p. 175, 1949.
Dudeney, H. E. §213 in *Amusements in Mathematics.* New York: Dover, 1970.
Gardner, M. Ch. 2 in *Mathematical Carnival: A New Round-Up of Tantalizers and Puzzles from Scientific American.* New York: Vintage Books, 1977.
Gardner, M. "Tree-Plant Problems." Ch. 22 in *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, pp. 277–290, 1988.
Grünbaum, B. "New Views on Some Old Questions of Combinatorial Geometry." *Teorie Combin.* **1**, 451–468, 1976.
Grünbaum, B. and Sloane, N. J. A. "The Orchard Problem." *Geom. Dedic.* **2**, 397–424, 1974.
Jackson, J. *Rational Amusements for Winter Evenings.* London, 1821.
Macmillan, R. H. "An Old Problem." *Math. Gaz.* **30**, 109, 1946.
Sloane, N. J. A. Sequences A006065/M0290 and A003035/M0982 in "An On-Line Version of the Encyclopedia of Integer Sequences." `http://www.research.att.com/~njas/sequences/eisonline.html`.
Sloane, N. J. A. and Plouffe, S. Extended entry for M0982 in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.

Orchard Visibility Problem

A tree is planted at each LATTICE POINT in a circular orchard which has CENTER at the ORIGIN and RADIUS r. If the radius of trees exceeds $1/r$ units, one is unable to see out of the orchard in any direction. However, if the RADII of the trees are $< 1/\sqrt{r^2+1}$, one can see out at certain ANGLES.

see also LATTICE POINT, ORCHARD-PLANTING PROBLEM, VISIBILITY

References
Honsberger, R. "The Orchard Problem." Ch. 4 in *Mathematical Gems I.* Washington, DC: Math. Assoc. Amer., pp. 43–52, 1973.

Order (Algebraic Curve)

The order of the POLYNOMIAL defining the curve.

Order (Algebraic Surface)

The order n of an ALGEBRAIC SURFACE is the order of the POLYNOMIAL defining a surface, which can be geometrically interpreted as the maximum number of points in which a line meets the surface.

Order	Surface
3	cubic surface
4	quartic surface
5	quintic surface
6	sextic surface
7	heptic surface
8	octic surface
9	nonic surface
10	decic surface

see also ALGEBRAIC SURFACE

References
Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, p. 8, 1986.

Order (Conjugacy Class)

The number of elements of a GROUP in a given CONJUGACY CLASS.

Order (Difference Set)

Let G be GROUP of ORDER h and D be a set of k elements of G. If the set of differences $d_i - d_j$ contains every NONZERO element of G exactly λ times, then D is a (h, k, λ)-difference set in G of order $n = k - \lambda$.

Order (Field)

The number of elements in a FINITE FIELD.

Order (Group)

The number of elements in a GROUP G, denoted $|G|$. The order of an element g of a finite group G is the smallest POWER of n such that $g^n = I$, where I is the IDENTITY ELEMENT. In general, finding the order of the element of a group is at least as hard as factoring (Meijer 1996). However, the problem becomes significantly easier if $|G|$ and the factorization of $|G|$ are known. Under these circumstances, efficient ALGORITHMS are known (Cohen 1993).

see also ABELIAN GROUP, FINITE GROUP

References
Cohen, H. *A Course in Computational Algebraic Number Theory.* New York: Springer-Verlag, 1993.
Meijer, A. R. "Groups, Factoring, and Cryptography." *Math. Mag.* **69**, 103–109, 1996.

Order (Modulo)

For any INTEGER a which is not a multiple of a PRIME p, there exists a smallest exponent $h \geq 1$ such that $a^h \equiv 1 \pmod{p}$ IFF $h|k$. In that case, h is called the order of a modulo p.

see also CARMICHAEL FUNCTION

Order (Ordinary Differential Equation)

An ORDINARY DIFFERENTIAL EQUATION of order n is an equation of the form

$$F(x, y, y', \ldots, y^{(n)}) = 0.$$

Order (Permutation)

see PERMUTATION

Order (Polynomial)

The highest order POWER in a one-variable POLYNOMIAL is known as its order (or sometimes its DEGREE). For example, the POLYNOMIAL

$$a_n x^n + \ldots + a_2 x^2 + a_1 x + a_0$$

is of order n.

Order Statistic

Given a sample of n variates $X_1, \ldots, X_n$, reorder them so that $X_1' < X_2' < \ldots < X_n'$. Then the ith order statistic $X^{(i)}$ is defined as X_i', with the special cases

$$m_n = X^{(1)} = \min_j(X_j)$$
$$M_n = X^{(n)} = \max_j(X_j).$$

A ROBUST ESTIMATION technique based on linear combinations of order statistics is called an L-ESTIMATE.

see also EXTREME VALUE DISTRIBUTION, HINGE, MAXIMUM, MINIMUM, MODE, ORDINAL NUMBER

References
Balakrishnan, N. and Cohen, A. C. *Order Statistics and Inference.* New York: Academic Press, 1991.
David, H. A. *Order Statistics, 2nd ed.* New York: Wiley, 1981.
Gibbons, J. D. and Chakraborti, S. (Eds.). *Nonparametric Statistic Inference, 3rd ed. exp. rev.* New York: Marcel Dekker, 1992.

Order (Vertex)

The number of EDGES meeting at a given node in a GRAPH is called the order of that VERTEX.

Ordered Geometry

A GEOMETRY constructed without reference to measurement. The only primitive concepts are those of points and intermediacy. There are 10 AXIOMS underlying ordered GEOMETRY.

see also ABSOLUTE GEOMETRY, AFFINE GEOMETRY, GEOMETRY

Ordered Pair

A PAIR of quantities (a, b) where ordering is significant, so (a, b) is considered distinct from (b, a) for $a \neq b$.

see also PAIR

Ordered Tree

A ROOTED TREE in which the order of the subtrees is significant. There is a ONE-TO-ONE correspondence between ordered FORESTS with n nodes and BINARY TREES with n nodes.

see also BINARY TREE, FOREST, ROOTED TREE

Ordering

The number of "ARRANGEMENTS" in an ordering of n items is given by either a COMBINATION (order is ignored) or a PERMUTATION (order is significant).

see also ARRANGEMENT, COMBINATION, CUTTING, DERANGEMENT, PARTIAL ORDER, PERMUTATION, SORTING, TOTAL ORDER

Ordering Axioms

The four of HILBERT'S AXIOMS which concern the arrangement of points.

see also CONGRUENCE AXIOMS, CONTINUITY AXIOMS, HILBERT'S AXIOMS, INCIDENCE AXIOMS, PARALLEL POSTULATE

References
Hilbert, D. *The Foundations of Geometry, 2nd ed.* Chicago, IL: Open Court, 1980.
Iyanaga, S. and Kawada, Y. (Eds.). "Hilbert's System of Axioms." §163B in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 544–545, 1980.

Ordinal Number

In informal usage, an ordinal number is an adjective which describes the numerical position of an object, e.g., first, second, third, etc.

In technical mathematics, an ordinal number is one of the numbers in Georg Cantor's extension of the WHOLE NUMBERS. The ordinal numbers are $0, 1, 2, \ldots, \omega, \omega+1, \omega+2, \ldots, \omega+\omega, \omega+\omega+1, \ldots$. Cantor's "smallest" TRANSFINITE NUMBER ω is defined to be the earliest number greater than all WHOLE NUMBERS, and is denoted by Conway and Guy (1996) as $\omega = \{0, 1, \ldots |\}$. The notation of ordinal numbers can be a bit counterintuitive, e.g., even though $1 + \omega = \omega$, $\omega + 1 > \omega$.

Ordinal numbers have some other rather peculiar properties. The sum of two ordinal numbers can take on two different values, the sum of three can take on five values. The first few terms of this sequence are 2, 5, 13, 33, 81, 193, 449, 33^2, $33 \cdot 81$, 81^2, $81 \cdot 193$, 192^2, ... (Conway and Guy 1996, Sloane's A005348). The sum of n ordinals has either $193^a 81^b$ or $33 \cdot 81^a$ possible answers for $n \geq 15$ (Conway and Guy 1996).

$r \times \omega$ is the same as ω, but $\omega \times r$ is equal to $\underbrace{\omega + \ldots + \omega}_{r}$. ω^2 is larger than any number of the form $\omega \times r$, ω^3 is larger than ω^2, and so on.

There exist ordinal numbers which cannot be constructed from smaller ones by finite additions, multiplications, and exponentiations. These ordinals obey Cantor's Equation. The first such ordinal is

$$\epsilon_0 = \underbrace{\omega^{\omega^{\cdot^{\cdot^{\cdot^{\omega}}}}}}_{\omega} = 1 + \omega + \omega^{\omega} + \omega^{\omega^{\omega}} + \ldots.$$

The next is

$$\epsilon_1 = (\epsilon_0 + 1) + \omega^{\epsilon_0 + 1} + \omega^{\omega\,\epsilon_0 + 1} + \ldots,$$

then follow ϵ_2, ϵ_3, ..., ϵ_ω, $\epsilon_{\omega+1}$, ..., $\epsilon_{\omega\times 2}$, ..., ϵ_{ω^2}, ϵ_{ω^ω}, ..., ϵ_{ϵ_0}, ϵ_{ϵ_0+1}, ..., $\epsilon_{\epsilon_0+\omega}$, ..., $\epsilon_{\epsilon_0+\omega^\omega}$, ..., $\epsilon_{\epsilon_0\times 2}$, ..., ϵ_{ϵ_1}, ..., ϵ_{ϵ_2}, ..., $\epsilon_{\epsilon_\omega}$, ..., $\epsilon_{\epsilon_{\epsilon_0}}$, ..., $\epsilon_{\epsilon_{\epsilon_1}}$, ..., $\epsilon_{\epsilon_{\epsilon_\omega}}$, ..., $\epsilon_{\epsilon_{\epsilon_{\epsilon_0}}}$, ... (Conway and Guy 1996).

see also Axiom of Choice, Cantor's Equation, Cardinal Number, Order Statistic, Power Set, Surreal Number

References

Cantor, G. *Über unendliche, lineare Punktmannigfältigkeiten, Arbeiten zur Mengenlehre aus dem Jahren 1872–1884.* Leipzig, Germany: Teubner-Archiv zur Mathematik, 1884.

Conway, J. H. and Guy, R. K. "Cantor's Ordinal Numbers." In *The Book of Numbers.* New York: Springer-Verlag, pp. 266–267 and 274, 1996.

Sloane, N. J. A. Sequence A005348/M1435 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Ordinary Differential Equation

An ordinary differential equation (frequently abbreviated ODE) is an equality involving a function and its Derivatives. An ODE of order n is an equation of the form

$$F(x, y, y', \ldots, y^{(n)}) = 0, \tag{1}$$

where $y' = dy/dx$ is a first Derivative with respect to x and $y^{(n)} = d^n y/dx^n$ is an nth Derivative with respect to x. An ODE of order n is said to be linear if it is of the form

$$a_n(x)y^{(n)} + a_{n-1}(x)y^{(n-1)} + \ldots + a_1(x)y' + a_0(x)y = Q(x). \tag{2}$$

A linear ODE where $Q(x) = 0$ is said to be homogeneous. Confusingly, an ODE of the form

$$\frac{dy}{dx} = f\left(\frac{y}{x}\right) \tag{3}$$

is also sometimes called "homogeneous."

Simple theories exist for first-order (Integrating Factor) and second-order (Sturm-Liouville Theory) ordinary differential equations, and arbitrary ODEs with linear constant Coefficients can be solved when they are of certain factorable forms. Integral transforms such as the Laplace Transform can also be used to solve classes of linear ODEs. Morse and Feshbach (1953, pp. 667–674) give canonical forms and solutions for second-order ODEs.

While there are many general techniques for analytically solving classes of ODEs, the only practical solution technique for complicated equations is to use numerical methods (Milne 1970). The most popular of these is the Runge-Kutta Method, but many others have been developed. A vast amount of research and huge numbers of publications have been devoted to the numerical solution of differential equations, both ordinary and Partial (PDEs) as a result of their importance in fields as diverse as physics, engineering, economics, and electronics.

The solutions to an ODE satisfy Existence and Uniqueness properties. These can be formally established by Picard's Existence Theorem for certain classes of ODEs. Let a system of first-order ODE be given by

$$\frac{dx_i}{dt} = f_i(x_1, \ldots, x_n, t), \tag{4}$$

for $i = 1$, ..., n and let the functions $f_i(x_1, \ldots, x_n, t)$, where $i = 1$, ..., n, all be defined in a Domain D of the $(n+1)$-D space of the variables x_1, ..., x_n, t. Let these functions be continuous in D and have continuous first Partial Derivatives $\partial f_i/\partial x_j$ for $i = 1$, ..., n and $j = 1$, ..., n in D. Let $(x_1^0, \ldots, x_n^0)$ be in D. Then there exists a solution of (4) given by

$$x_1 = x_1(t), \ldots, x_n = x_n(t) \tag{5}$$

for $t_0 - \delta < t < t_0 + \delta$ (where $\delta > 0$) satisfying the initial conditions

$$x_1(t_0) = x_1^0, \ldots, x_n(t_0) = x_n^0. \tag{6}$$

Furthermore, the solution is unique, so that if

$$x_1 = x_1^*(t), \ldots, x_n = x_n^*(t) \tag{7}$$

is a second solution of (4) for $t_0 - \delta < t < t_0 + \delta$ satisfying (6), then $x_i(t) \equiv x_i^*(t)$ for $t_0 - \delta < t < t_0 + \delta$. Because every nth-order ODE can be expressed as a system of n first-order differential equations, this theorem also applies to the single nth-order ODE.

In general, an nth-order ODE has n linearly independent solutions. Furthermore, any linear combination of Linearly Independent Functions solutions is also a solution.

An exact First-Order ODEs is one of the form

$$p(x, y)\,dx + q(x, y)\,dy = 0, \tag{8}$$

where

$$\frac{\partial p}{\partial y} = \frac{\partial q}{\partial x}. \tag{9}$$

An equation of the form (8) with

$$\frac{\partial p}{\partial y} \neq \frac{\partial q}{\partial x} \tag{10}$$

is said to be nonexact. If

$$\frac{\frac{\partial p}{\partial y} - \frac{\partial q}{\partial x}}{q} = f(x) \tag{11}$$

in (8), it has an x-dependent integrating factor. If

$$\frac{\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y}}{xp - yq} = f(xy) \tag{12}$$

in (8), it has an xy-dependent integrating factor. If

$$\frac{\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y}}{p} = f(y) \tag{13}$$

in (8), it has a y-dependent integrating factor.

Other special first-order types include cross multiple equations

$$yf(xy)\,dx + xg(xy)\,dy = 0, \tag{14}$$

homogeneous equations

$$\frac{dy}{dx} = f\left(\frac{y}{x}\right), \tag{15}$$

linear equations

$$\frac{dy}{dx} + p(x)y = q(x), \tag{16}$$

and separable equations

$$\frac{dy}{dx} = X(x)Y(y). \tag{17}$$

Special classes of SECOND-ORDER ODEs include

$$\frac{d^2y}{dx^2} = f(y, y') \tag{18}$$

(x missing) and

$$\frac{d^2y}{dx^2} = f(x, y') \tag{19}$$

(y missing). A second-order linear homogeneous ODE

$$\frac{d^2y}{dx^2} + P(x)\frac{dy}{dx} + Q(x)y = 0 \tag{20}$$

for which

$$\frac{Q'(x) + 2P(x)Q(x)}{2[Q(x)]^{3/2}} = [\text{constant}] \tag{21}$$

can be transformed to one with constant coefficients.

The undamped equation of SIMPLE HARMONIC MOTION is

$$\frac{d^2y}{dx^2} + {\omega_0}^2 y = 0, \tag{22}$$

which becomes

$$\frac{d^2y}{dx^2} + \beta\frac{dy}{dx} + {\omega_0}^2 y = 0 \tag{23}$$

when damped, and

$$\frac{d^2y}{dx^2} + \beta\frac{dy}{dx} + {\omega_0}^2 y = A\cos(\omega t) \tag{24}$$

when both forced and damped.

SYSTEMS WITH CONSTANT COEFFICIENTS are of the form

$$\frac{d\mathbf{x}}{dt} = \mathsf{A}\mathbf{x}(t) + \mathbf{p}(t). \tag{25}$$

The following are examples of important ordinary differential equations which commonly arise in problems of mathematical physics.

AIRY DIFFERENTIAL EQUATION

$$\frac{d^2y}{dx^2} - xy = 0. \tag{26}$$

BERNOULLI DIFFERENTIAL EQUATION

$$\frac{dy}{dx} + p(x)y = q(x)y^n. \tag{27}$$

BESSEL DIFFERENTIAL EQUATION

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + (\lambda^2x^2 - n^2)y = 0. \tag{28}$$

CHEBYSHEV DIFFERENTIAL EQUATION

$$(1 - x^2)\frac{d^2y}{dx^2} - x\frac{dy}{dx} + \alpha^2 y = 0. \tag{29}$$

CONFLUENT HYPERGEOMETRIC DIFFERENTIAL EQUATION

$$x\frac{d^2y}{dx^2} + (\gamma - x)\frac{dy}{dx} + \alpha y = 0. \tag{30}$$

EULER DIFFERENTIAL EQUATION

$$x^2\frac{d^2y}{dx^2} + ax\frac{dy}{dx} + by = S(x). \tag{31}$$

HERMITE DIFFERENTIAL EQUATION

$$\frac{d^2y}{dx^2} - 2x\frac{dy}{dx} + \lambda y = 0. \tag{32}$$

HILL'S DIFFERENTIAL EQUATION

$$\frac{d^2y}{dx^2} + \left[\theta_0 + 2\sum_{n=1}^{\infty}\theta_n\cos(2nz)\right] = 0. \tag{33}$$

HYPERGEOMETRIC DIFFERENTIAL EQUATION

$$x(x-1)\frac{d^2y}{dx^2}+[(1+\alpha+\beta)x-\gamma]\frac{dy}{dx}+\alpha\beta y=0. \quad (34)$$

JACOBI DIFFERENTIAL EQUATION

$$(1-x^2)y''+[\beta-\alpha-(\alpha+\beta+2)x]y'+n(n+\alpha+\beta+1)y=0. \quad (35)$$

LAGUERRE DIFFERENTIAL EQUATION

$$x\frac{d^2y}{dx^2}+(1-x)\frac{dy}{dx}+\lambda y=0. \quad (36)$$

LANE-EMDEN DIFFERENTIAL EQUATION

$$\frac{1}{\xi^2}\frac{d}{d\xi}\left(\xi^2\frac{d\theta}{d\xi}\right)+\theta^n=0. \quad (37)$$

LEGENDRE DIFFERENTIAL EQUATION

$$(1-x^2)\frac{d^2y}{dx^2}-2x\frac{dy}{dx}+\alpha(\alpha+1)y=0. \quad (38)$$

LINEAR CONSTANT COEFFICIENTS

$$a_0\frac{d^ny}{dx^n}+\ldots+a_{n-1}\frac{dy}{dx}+a_ny=p(x). \quad (39)$$

MALMSTÉN'S DIFFERENTIAL EQUATION

$$y''+\frac{r}{z}y'=\left(Az^m+\frac{s}{z^2}\right)y. \quad (40)$$

RICCATI DIFFERENTIAL EQUATION

$$\frac{dw}{dx}=q_0(x)+q_1(x)w+q_2(x)w^2. \quad (41)$$

RIEMANN P-DIFFERENTIAL EQUATION

$$\begin{aligned}\frac{d^2u}{dz^2}&+\left[\frac{1-\alpha-\alpha'}{z-a}+\frac{1-\beta-\beta'}{z-b}+\frac{1-\gamma-\gamma'}{z-c}\right]\frac{du}{dz}\\&+\left[\frac{\alpha\alpha'(a-b)(a-c)}{z-a}+\frac{\beta\beta'(b-c)(b-a)}{z-b}\right.\\&\left.+\frac{\gamma\gamma'(c-a)(c-b)}{z-c}\right]\frac{u}{(z-a)(z-b)(z-c)}=0. \quad (42)\end{aligned}$$

see also ADAMS' METHOD, GREEN'S FUNCTION, ISOCLINE, LAPLACE TRANSFORM, LEADING ORDER ANALYSIS, MAJORANT, ORDINARY DIFFERENTIAL EQUATION—FIRST-ORDER, ORDINARY DIFFERENTIAL EQUATION—SECOND-ORDER, PARTIAL DIFFERENTIAL EQUATION, RELAXATION METHODS, RUNGE-KUTTA METHOD, SIMPLE HARMONIC MOTION

References

Boyce, W. E. and DiPrima, R. C. *Elementary Differential Equations and Boundary Value Problems, 5th ed.* New York: Wiley, 1992.

Braun, M. *Differential Equations and Their Applications, 3rd ed.* New York: Springer-Verlag, 1991.

Forsyth, A. R. *Theory of Differential Equations, 6 vols.* New York: Dover, 1959.

Forsyth, A. R. *A Treatise on Differential Equations.* New York: Dover, 1997.

Guterman, M. M. and Nitecki, Z. H. *Differential Equations: A First Course, 3rd ed.* Philadelphia, PA: Saunders, 1992.

Ince, E. L. *Ordinary Differential Equations.* New York: Dover, 1956.

Milne, W. E. *Numerical Solution of Differential Equations.* New York: Dover, 1970.

Morse, P. M. and Feshbach, H. "Ordinary Differential Equations." Ch. 5 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 492–675, 1953.

Moulton, F. R. *Differential Equations.* New York: Dover, 1958.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Integration of Ordinary Differential Equations." Ch. 16 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 701–744, 1992.

Simmons, G. F. *Differential Equations, with Applications and Historical Notes, 2nd ed.* New York: McGraw-Hill, 1991.

Zwillinger, D. *Handbook of Differential Equations, 3rd ed.* Boston, MA: Academic Press, 1997.

Ordinary Differential Equation—First-Order

Given a first-order ORDINARY DIFFERENTIAL EQUATION

$$\frac{dy}{dx}=F(x,y), \quad (1)$$

if $F(x,y)$ can be expressed using SEPARATION OF VARIABLES as

$$F(x,y)=X(x)Y(y), \quad (2)$$

then the equation can be expressed as

$$\frac{dy}{Y(y)}=X(x)\,dx \quad (3)$$

and the equation can be solved by integrating both sides to obtain

$$\int\frac{dy}{Y(y)}=\int X(x)\,dx. \quad (4)$$

Any first-order ODE of the form

$$\frac{dy}{dx}+p(x)y=q(x) \quad (5)$$

can be solved by finding an INTEGRATING FACTOR $\mu=\mu(x)$ such that

$$\frac{d}{dx}(\mu y)=\mu\frac{dy}{dx}+y\frac{d\mu}{dx}=\mu q(x). \quad (6)$$

Dividing through by μy yields

$$\frac{1}{y}\frac{dy}{dx}+\frac{1}{\mu}\frac{d\mu}{dx}=\frac{q(x)}{y}. \quad (7)$$

However, this condition enables us to explicitly determine the appropriate μ for arbitrary p and q. To accomplish this, take

$$p(x) = \frac{1}{\mu}\frac{d\mu}{dx} \tag{8}$$

in the above equation, from which we recover the original equation (5), as required, in the form

$$\frac{1}{y}\frac{dy}{dx} + p(x) = \frac{q(x)}{y}. \tag{9}$$

But we can integrate both sides of (8) to obtain

$$\int p(x)\,dx = \int \frac{d\mu}{\mu} = \ln\mu + c \tag{10}$$

$$\mu = e^{\int p(x)\,dx}. \tag{11}$$

Now integrating both sides of (6) gives

$$\mu y = \int \mu q(x)\,dx + c \tag{12}$$

(with μ now a known function), which can be solved for y to obtain

$$y = \frac{\int \mu q(x)\,dx + c}{\mu} = \frac{\int e^{\int^x p(x')\,dx'} q(x)\,dx + c}{e^{\int^x p(x')\,dx'}}, \tag{13}$$

where c is an arbitrary constant of integration.

Given an nth-order linear ODE with constant COEFFICIENTS

$$\frac{d^n y}{dx^n} + a_{n-1}\frac{d^{n-1}y}{dx^{n-1}} + \ldots + a_1\frac{dy}{dx} + a_0 y = Q(x), \tag{14}$$

first solve the characteristic equation obtained by writing

$$y \equiv e^{rx} \tag{15}$$

and setting $Q(x) = 0$ to obtain the n COMPLEX ROOTS.

$$r^n e^{rx} + a_{n-1}r^{n-1}e^{rx} + \ldots + a_1 r e^{rx} + a_0 e^{rx} = 0 \tag{16}$$

$$r^n + a_{n-1}r^{n-1} + \ldots + a_1 r + a_0 = 0. \tag{17}$$

Factoring gives the ROOTS r_i,

$$(r - r_1)(r - r_2)\cdots(r - r_n) = 0. \tag{18}$$

For a nonrepeated REAL ROOT r, the corresponding solution is

$$y = e^{rx}. \tag{19}$$

If a REAL ROOT r is repeated k times, the solutions are degenerate and the linearly independent solutions are

$$y = e^{rx}, y = xe^{rx}, \ldots, y = x^{k-1}e^{rx}. \tag{20}$$

Complex ROOTS always come in COMPLEX CONJUGATE pairs, $r_\pm = a \pm ib$. For nonrepeated COMPLEX ROOTS, the solutions are

$$y = e^{ax}\cos(bx), y = e^{ax}\sin(bx). \tag{21}$$

If the COMPLEX ROOTS are repeated k times, the linearly independent solutions are

$$\begin{aligned} &y = e^{ax}\cos(bx), y = e^{ax}\sin(bx), \ldots, \\ &\qquad y = x^{k-1}e^{ax}\cos(bx), y = x^{k-1}e^{ax}\sin(bx). \end{aligned} \tag{22}$$

Linearly combining solutions of the appropriate types with arbitrary multiplicative constants then gives the complete solution. If initial conditions are specified, the constants can be explicitly determined. For example, consider the sixth-order linear ODE

$$(\tilde{D} - 1)(\tilde{D} - 2)^3(\tilde{D}^2 + \tilde{D} + 1)y = 0, \tag{23}$$

which has the characteristic equation

$$(r - 1)(r - 2)^3(r^2 + r + 1) = 0. \tag{24}$$

The roots are 1, 2 (three times), and $(-1 \pm \sqrt{3}\,i)/2$, so the solution is

$$\begin{aligned} y = Ae^x + Be^{2x} + Cxe^{2x} + Dx^2e^{3x} + Ee^{-x/2}\cos(\tfrac{1}{2}\sqrt{3}\,x) \\ +Fe^{-x}\sin(\tfrac{1}{2}\sqrt{3}\,x). \end{aligned} \tag{25}$$

If the original equation is nonhomogeneous ($Q(x) \neq 0$), now find the particular solution y^* by the method of VARIATION OF PARAMETERS. The general solution is then

$$y(x) = \sum_{i=1}^{n} c_i y_i(x) + y^*(x), \tag{26}$$

where the solutions to the linear equations are $y_1(x)$, $y_2(x)$, ..., $y_n(x)$, and $y^*(x)$ is the particular solution.

see also INTEGRATING FACTOR, ORDINARY DIFFERENTIAL EQUATION—FIRST-ORDER EXACT, SEPARATION OF VARIABLES, VARIATION OF PARAMETERS

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 440–445, 1985.

Ordinary Differential Equation—First-Order Exact

Consider a first-order ODE in the slightly different form

$$p(x, y)\,dx + q(x, y)\,dy = 0. \tag{1}$$

Such an equation is said to be exact if

$$\frac{\partial p}{\partial y} = \frac{\partial q}{\partial x}. \tag{2}$$

This statement is equivalent to the requirement that a CONSERVATIVE FIELD exists, so that a scalar potential can be defined. For an exact equation, the solution is

$$\int_{(x_0,y_0)}^{(x,y)} p(x,y)\,dx + q(x,y)\,dy = c, \tag{3}$$

where c is a constant.

A first-order ODE (1) is said to be inexact if

$$\frac{\partial p}{\partial y} \neq \frac{\partial q}{\partial x}. \tag{4}$$

For a nonexact equation, the solution may be obtained by defining an INTEGRATING FACTOR μ of (6) so that the new equation

$$\mu p(x,y)\,dx + \mu q(x,y)\,dy = 0 \tag{5}$$

satisfies

$$\frac{\partial}{\partial y}(\mu p) = \frac{\partial}{\partial x}(\mu q), \tag{6}$$

or, written out explicitly,

$$p\frac{\partial \mu}{\partial y} + \mu\frac{\partial p}{\partial y} = q\frac{\partial \mu}{\partial x} + \mu\frac{\partial p}{\partial x}. \tag{7}$$

This transforms the nonexact equation into an exact one. Solving (7) for μ gives

$$\mu = \frac{q\frac{\partial \mu}{\partial x} - p\frac{\partial \mu}{\partial y}}{\frac{\partial p}{\partial y} - \frac{\partial q}{\partial x}}. \tag{8}$$

Therefore, if a function μ satisfying (8) can be found, then writing

$$P(x,y) = \mu p \tag{9}$$

$$Q(x,y) = \mu q \tag{10}$$

in equation (5) then gives

$$P(x,y)\,dx + Q(x,y)\,dy = 0, \tag{11}$$

which is then an exact ODE. Special cases in which μ can be found include x-dependent, xy-dependent, and y-dependent integrating factors.

Given an inexact first-order ODE, we can also look for an INTEGRATING FACTOR $\mu(x)$ so that

$$\frac{\partial \mu}{\partial y} = 0. \tag{12}$$

For the equation to be exact in μp and μq, the equation for a first-order nonexact ODE

$$p\frac{\partial \mu}{\partial y} + \mu\frac{\partial p}{\partial y} = q\frac{\partial \mu}{\partial x} + \mu\frac{\partial p}{\partial x} \tag{13}$$

becomes

$$\mu\frac{\partial p}{\partial y} = q\frac{\partial \mu}{\partial x} + \mu\frac{\partial p}{\partial x}. \tag{14}$$

Solving for $\partial\mu/\partial x$ gives

$$\frac{\partial \mu}{\partial x} = \mu(x)\frac{\frac{\partial p}{\partial y} - \frac{\partial q}{\partial x}}{q} \equiv f(x,y)\mu(x), \tag{15}$$

which will be integrable if

$$f(x,y) \equiv \frac{\frac{\partial p}{\partial y} - \frac{\partial q}{\partial x}}{q} = f(x), \tag{16}$$

in which case

$$\frac{d\mu}{\mu} = f(x)\,dx, \tag{17}$$

so that the equation is integrable

$$\mu(x) = e^{\int f(x)\,dx}, \tag{18}$$

and the equation

$$[\mu p(x,y)]dx + [\mu q(x,y)]dy = 0 \tag{19}$$

with known $\mu(x)$ is now exact and can be solved as an exact ODE.

Given in an exact first-order ODE, look for an INTEGRATING FACTOR $\mu(x,y) = g(xy)$. Then

$$\frac{\partial \mu}{\partial x} = \frac{\partial g}{\partial x}y \tag{20}$$

$$\frac{\partial \mu}{\partial y} = \frac{\partial g}{\partial y}x. \tag{21}$$

Combining these two,

$$\frac{\partial \mu}{\partial x} = \frac{y}{x}\frac{\partial \mu}{\partial y}. \tag{22}$$

For the equation to be exact in μp and μq, the equation for a first-order nonexact ODE

$$p\frac{\partial \mu}{\partial y} + \mu\frac{\partial p}{\partial y} = q\frac{\partial \mu}{\partial x} + \mu\frac{\partial p}{\partial x} \tag{23}$$

becomes

$$\frac{\partial \mu}{\partial y}\left(p - \frac{y}{x}q\right) = \left(\frac{\partial p}{\partial x} - \frac{\partial p}{\partial y}\right)\mu. \tag{24}$$

Therefore,

$$\frac{1}{x}\frac{\partial \mu}{\partial y} = \frac{\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y}}{xp - yq}\mu. \tag{25}$$

Define a new variable

$$t(x,y) \equiv xy, \tag{26}$$

then $\partial t/\partial y = x$, so

$$\frac{\partial \mu}{\partial t} = \frac{\partial \mu}{\partial y}\frac{\partial y}{\partial t} = \frac{\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y}}{xp - yq}\mu(t) \equiv f(x,y)\mu(t). \quad (27)$$

Now, if

$$f(x,y) \equiv \frac{\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y}}{xp - yq} = f(xy) = f(t), \quad (28)$$

then

$$\frac{\partial \mu}{\partial t} = f(t)\mu(t), \quad (29)$$

so that

$$\mu = e^{\int f(t)\,dt} \quad (30)$$

and the equation

$$[\mu p(x,y)]\,dx + [\mu q(x,y)]\,dy = 0 \quad (31)$$

is now exact and can be solved as an exact ODE.

Given an inexact first-order ODE, assume there exists an integrating factor

$$\mu = f(y), \quad (32)$$

so $\partial \mu/\partial x = 0$. For the equation to be exact in μp and μq, equation (7) becomes

$$\frac{\partial \mu}{\partial y} = \frac{\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y}}{p}\mu = f(x,y)\mu(y). \quad (33)$$

Now, if

$$f(x,y) \equiv \frac{\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y}}{p} = f(y), \quad (34)$$

then

$$\frac{d\mu}{\mu} = f(y)\,dy, \quad (35)$$

so that

$$\mu(y) = e^{\int f(y)\,dy}, \quad (36)$$

and the equation

$$\mu p(x,y)\,dx + \mu q(x,y)\,dy = 0 \quad (37)$$

is now exact and can be solved as an exact ODE.

Given a first-order ODE of the form

$$yf(xy)\,dx + xg(xy)\,dy = 0, \quad (38)$$

define

$$v \equiv xy. \quad (39)$$

Then the solution is

$$\begin{cases} \ln x = \int \frac{g(v)\,dv}{c[g(v)-f(v)]} + c & \text{for } g(v) \neq f(v) \\ xy = c & \text{for } g(v) = f(v). \end{cases} \quad (40)$$

If

$$\frac{dy}{dx} = F(x,y) = G(v), \quad (41)$$

where

$$v \equiv \frac{y}{x}, \quad (42)$$

then letting

$$y \equiv xv \quad (43)$$

gives

$$\frac{dy}{dx} = x\,dv/dx + v \quad (44)$$

$$x\frac{dv}{dx} + v = G(v). \quad (45)$$

This can be integrated by quadratures, so

$$\ln x = \int \frac{dv}{f(v) - v} + c \qquad \text{for } f(v) \neq v \quad (46)$$

$$y = cx \qquad \text{for} f(v) = v. \quad (47)$$

References

Boyce, W. E. and DiPrima, R. C. *Elementary Differential Equations and Boundary Value Problems, 4th ed.* New York: Wiley, 1986.

Ordinary Differential Equation—Second-Order

An ODE

$$y'' + P(x)y' + Q(x)y = 0 \quad (1)$$

has singularities for finite $x = x_0$ under the following conditions: (a) If either $P(x)$ or $Q(x)$ diverges as $x \to x_0$, but $(x - x_0)P(x)$ and $(x - x_0)^2Q(x)$ remain finite as $x \to x_0$, then x_0 is called a regular or nonessential singular point. (b) If $P(x)$ diverges faster than $(x - x_0)^{-1}$ so that $(x - x_0)P(x) \to \infty$ as $x \to x_0$, or $Q(x)$ diverges faster than $(x - x_0)^{-2}$ so that $(x - x_0)^2Q(x) \to \infty$ as $x \to x_0$, then x_0 is called an irregular or essential singularity.

Singularities of equation (1) at infinity are investigated by making the substitution $x \equiv z^{-1}$, so $dx = -z^{-2}\,dz$, giving

$$\frac{dy}{dx} = -z^2\frac{dy}{dz} \quad (2)$$

$$\begin{aligned} \frac{d^2y}{dx^2} &= -z^2\frac{d}{dz}\left(-z^2\frac{dy}{dz}\right) = -z^2\left(-2z\frac{dy}{dz} - z^2\frac{d^2y}{dz^2}\right) \\ &= 2z^3\frac{dy}{dz} + z^4\frac{d^2y}{dz^2}. \end{aligned} \quad (3)$$

Then (1) becomes

$$z^4\frac{d^2y}{dz^2} + [2z^3 - z^2P(z)]\frac{dy}{dz} + Q(z)y = 0. \quad (4)$$

Case (a): If

$$\alpha(z) \equiv \frac{2z - P(z)}{z^2} \tag{5}$$

$$\beta(z) \equiv \frac{Q(z)}{z^4} \tag{6}$$

remain finite at $x = \pm\infty$ ($y = 0$), then the point is ordinary. Case (b): If either $\alpha(z)$ diverges no more rapidly than $1/z$ or $\beta(z)$ diverges no more rapidly than $1/z^2$, then the point is a regular singular point. Case (c): Otherwise, the point is an irregular singular point.

Morse and Feshbach (1953, pp. 667–674) give the canonical forms and solutions for second-order ODEs classified by types of singular points.

For special classes of second-order linear ordinary differential equations, variable COEFFICIENTS can be transformed into constant COEFFICIENTS. Given a second-order linear ODE with variable COEFFICIENTS

$$\frac{d^2y}{dx^2} + p(x)\frac{dy}{dx} + q(x)y = 0. \tag{7}$$

Define a function $z \equiv y(x)$,

$$\frac{dy}{dx} = \frac{dz}{dx}\frac{dy}{dz} \tag{8}$$

$$\frac{d^2y}{dx^2} = \left(\frac{dz}{dx}\right)^2 \frac{d^2y}{dz^2} + \frac{d^2z}{dx^2}\frac{dy}{dz} \tag{9}$$

$$\left(\frac{dz}{dx}\right)^2 \frac{d^2y}{dz^2} + \left[\frac{d^2z}{dx^2} + p(x)\frac{dz}{dx}\right]\frac{dy}{dz} + q(x)y = 0 \tag{10}$$

$$\frac{d^2y}{dz^2} + \left[\frac{\frac{d^2z}{dx^2} + p(x)\frac{dz}{dx}}{\left(\frac{dz}{dx}\right)^2}\right]\frac{dy}{dz} + \left[\frac{q(x)}{\left(\frac{dz}{dx}\right)^2}\right] y \equiv \frac{d^2y}{dz^2} + A\frac{dy}{dz} + By = 0. \tag{11}$$

This will have constant COEFFICIENTS if A and B are not functions of x. But we are free to set B to an arbitrary POSITIVE constant for $q(x) \geq 0$ by defining z as

$$z \equiv B^{-1/2}\int [q(x)]^{1/2}\,dx. \tag{12}$$

Then

$$\frac{dz}{dx} = B^{-1/2}[q(x)]^{1/2} \tag{13}$$

$$\frac{d^2z}{dx^2} = \tfrac{1}{2}B^{-1/2}[q(x)]^{-1/2}q'(x), \tag{14}$$

and

$$A = \frac{\frac{1}{2}B^{-1/2}[q(x)]^{-1/2}q'(x) + B^{-1/2}p(x)[q(x)]^{1/2}}{B^{-1}q(x)} = \frac{q'(x) + 2p(x)q(x)}{2[q(x)]^{3/2}}B^{1/2}. \tag{15}$$

Equation (11) therefore becomes

$$\frac{d^2y}{dz^2} + \frac{q'(x) + 2p(x)q(x)}{2[q(x)]^{3/2}}B^{1/2}\frac{dy}{dz} + By = 0, \tag{16}$$

which has constant COEFFICIENTS provided that

$$A \equiv \frac{q'(x) + 2p(x)q(x)}{2[q(x)]^{3/2}}B^{1/2} = [\text{constant}]. \tag{17}$$

Eliminating constants, this gives

$$A' \equiv \frac{q'(x) + 2p(x)q(x)}{[q(x)]^{3/2}} = [\text{constant}]. \tag{18}$$

So for an ordinary differential equation in which A' is a constant, the solution is given by solving the second-order linear ODE with constant COEFFICIENTS

$$\frac{d^2y}{dz^2} + A\frac{dy}{dz} + By = 0 \tag{19}$$

for z, where z is defined as above.

A linear second-order homogeneous differential equation of the general form

$$y''(x) + P(x)y' + Q(x)y = 0 \tag{20}$$

can be transformed into standard form

$$z''(x) + q(x)z = 0 \tag{21}$$

with the first-order term eliminated using the substitution

$$\ln y \equiv \ln z - \tfrac{1}{2}\int P(x)\,dx. \tag{22}$$

Then

$$\frac{y'}{y} = \frac{z'}{z} - \tfrac{1}{2}P(x) \tag{23}$$

$$\frac{yy'' - y'^2}{y^2} = \frac{zz'' - z'^2}{z^2} - \tfrac{1}{2}P'(x) \tag{24}$$

$$\frac{y''}{y} - \left(\frac{y'}{y}\right)^2 = \frac{z''}{z} - \frac{z'^2}{z} - \frac{z'^2}{z^2} - \tfrac{1}{2}P'(x) \tag{25}$$

$$\frac{y''}{y} = \left[\frac{z'}{z} - \tfrac{1}{2}P(x)\right]^2 + \frac{z''}{z} - \frac{z'^2}{z^2} - \tfrac{1}{2}P'(x)$$
$$= \frac{z'^2}{z^2} - \frac{z'}{z}P(x) + \tfrac{1}{4}P^2(x) + \frac{z''}{z} - \frac{z'^2}{z^2} - \tfrac{1}{2}P'(x), \tag{26}$$

so

$$\frac{y''}{y} + P(x)\frac{y'}{y} + Q(x) = -\frac{z'}{z}P(x) + \tfrac{1}{4}P^2(x) + \frac{z''}{z} - \tfrac{1}{2}P'(x) + P(x)\left[\frac{z'}{z} - \tfrac{1}{2}P(x)\right] + Q(x) = \frac{z''}{z} - \tfrac{1}{2}P'(x) - \tfrac{1}{4}P^2(x) + Q(x) = 0. \tag{27}$$

Therefore,

$$z'' + [Q(x) - \tfrac{1}{2}P'(x) - \tfrac{1}{4}P^2(x)]z \equiv z''(x) + q(x)z = 0, \quad (28)$$

where

$$q(x) \equiv Q(x) - \tfrac{1}{2}P'(x) - \tfrac{1}{4}P^2(x). \quad (29)$$

If $Q(x) = 0$, then the differential equation becomes

$$y'' + P(x)y' = 0, \quad (30)$$

which can be solved by multiplying by

$$\exp\left[\int^x P(x')\,dx'\right] \quad (31)$$

to obtain

$$0 = \frac{d}{dx}\left\{\exp\left[\int^x P(x')\,dx'\right]\frac{dy}{dx}\right\} \quad (32)$$

$$c_1 = \exp\left[\int^x P(x')\,dx'\right]\frac{dy}{dx} \quad (33)$$

$$y = c_1 \int^x \frac{dx}{\exp\left[\int^x P(x')\,dx'\right]} + c_2. \quad (34)$$

If one solution (y_1) to a second-order ODE is known, the other (y_2) may be found using the REDUCTION OF ORDER method. From the ABEL'S IDENTITY

$$\frac{dW}{W} = -P(x)\,dx, \quad (35)$$

where

$$W \equiv y_1 y_2' - y_1' y_2 \quad (36)$$

$$\int_a^x \frac{dW}{W} = \int_a^x P(x')\,dx' \quad (37)$$

$$\ln\left[\frac{W(x)}{W(a)}\right] = \int_a^x P(x')\,dx' \quad (38)$$

$$W(x) = W(a)\exp\left[-\int_a^x P(x')\,dx'\right]. \quad (39)$$

But

$$W \equiv y_1 y_2' - y_1' y_2 = y_1^2 \frac{d}{dx}\left(\frac{y_2}{y_1}\right). \quad (40)$$

Combining (39) and (40) yields

$$\frac{d}{dx}\left(\frac{y_2}{y1}\right) = W(a)\frac{\exp[-\int_a^x P(x')\,dx']}{y_1^2} \quad (41)$$

$$y_2(x) = y_1(x)W(a)\int_b^x \frac{\exp[-\int_a^{x'} P(x'')\,dx'']}{[y_1(x')]^2}\,dx'. \quad (42)$$

Disregarding $W(a)$, since it is simply a multiplicative constant, and the constants a and b, which will contribute a solution which is not linearly independent of y_1,

$$y_2(x) = y_1(x)\int^x \frac{\exp\left[-\int^{x'} P(x'')\,dx''\right]}{[y_1(x')]^2}\,dx'. \quad (43)$$

If $P(x) = 0$, this simplifies to

$$y_2(x) = y_1(x)\int^x \frac{dx'}{[y_1(x')]^2}. \quad (44)$$

For a nonhomogeneous second-order ODE in which the x term does not appear in the function $f(x, y, y')$,

$$\frac{d^2y}{dx^2} = f(y, y'), \quad (45)$$

let $v \equiv y'$, then

$$\frac{dv}{dx} = f(v, y) = \frac{dv}{dy}\frac{dy}{dx} = v\frac{dv}{dy}. \quad (46)$$

So the first-order ODE

$$v\frac{dv}{dy} = f(y, v), \quad (47)$$

if linear, can be solved for v as a linear first-order ODE. Once the solution is known,

$$\frac{dy}{dx} = v(y) \quad (48)$$

$$\int \frac{dy}{v(y)} = \int dx. \quad (49)$$

On the other hand, if y is missing from $f(x, y, y')$,

$$\frac{d^2y}{dx^2} = f(x, y'), \quad (50)$$

let $v \equiv y'$, then $v' = y''$, and the equation reduces to

$$v' = f(x, v), \quad (51)$$

which, if linear, can be solved for v as a linear first-order ODE. Once the solution is known,

$$y = \int v(x)\,dx. \quad (52)$$

see also ABEL'S IDENTITY, ADJOINT OPERATOR

References

Arfken, G. "A Second Solution." §8.6 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 467–480, 1985.

Boyce, W. E. and DiPrima, R. C. *Elementary Differential Equations and Boundary Value Problems, 4th ed.* New York: Wiley, 1986.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 667–674, 1953.

Ordinary Differential Equation—System with Constant Coefficients

To solve the system of differential equations

$$\frac{d\mathbf{x}}{dt} = \mathsf{A}\mathbf{x}(t) + \mathbf{p}(t), \tag{1}$$

where A is a MATRIX and $\mathbf{x}$ and $\mathbf{p}$ are VECTORS, first consider the homogeneous case with $\mathbf{p} = \mathbf{0}$. Then the solutions to

$$\frac{d\mathbf{x}}{dt} = \mathsf{A}\mathbf{x}(t) \tag{2}$$

are given by

$$\mathbf{x}(t) = e^{\mathsf{A}t}\mathbf{x}(\mathbf{t}). \tag{3}$$

But, by the MATRIX DECOMPOSITION THEOREM, the MATRIX EXPONENTIAL can be written as

$$e^{\mathsf{A}t} = \mathsf{uDu}^{-1}, \tag{4}$$

where the EIGENVECTOR MATRIX is

$$\mathsf{u} = [\,\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_n\,] \tag{5}$$

and the EIGENVALUE MATRIX is

$$\mathsf{D} = \begin{bmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & e^{\lambda_n t} \end{bmatrix}. \tag{6}$$

Now consider

$$\begin{aligned} e^{\mathsf{A}t}\mathsf{u} &= \mathsf{uDu}^{-1}\mathsf{u} = \mathsf{uD} \\ &= \begin{bmatrix} u_{11} & u_{21} & \cdots & u_{n1} \\ u_{12} & u_{22} & \cdots & u_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ u_{1n} & u_{2n} & \cdots & u_{nn} \end{bmatrix} \begin{bmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & e^{\lambda_n t} \end{bmatrix} \\ &= \begin{bmatrix} u_{11}e^{\lambda_1 t} & \cdots & u_{n1}e^{\lambda_n t} \\ u_{11}e^{\lambda_1 t} & \cdots & u_{n2}e^{\lambda_n t} \\ \vdots & \ddots & \vdots \\ u_{n1}e^{\lambda_1 t} & \cdots & u_{n2}e^{\lambda_n t} \end{bmatrix}. \end{aligned} \tag{7}$$

The individual solutions are then

$$\mathbf{x}_i = (e^{\mathsf{A}t}\mathsf{u}) \cdot \hat{\mathbf{e}}_i = \mathbf{u}_i e^{\lambda_i t}, \tag{8}$$

so the homogeneous solution is

$$\mathbf{x} = \sum_{i=1}^{n} c_i \mathbf{u}_i e^{\lambda_i t}, \tag{9}$$

where the c_is are arbitrary constants.

The general procedure is therefore

1. Find the EIGENVALUES of the MATRIX A ($\lambda_1, \ldots, \lambda_n$) by solving the CHARACTERISTIC EQUATION.
2. Determine the corresponding EIGENVECTORS $\mathbf{u}_1, \ldots, \mathbf{u}_n$.
3. Compute
   $$\mathbf{x}_i \equiv e^{\lambda_i t}\mathbf{u}_i \tag{10}$$
   for $i = 1, \ldots, n$. Then the VECTORS $\mathbf{x}_i$ which are REAL are solutions to the homogeneous equation. If A is a 2×2 matrix, the COMPLEX vectors $\mathbf{x}_j$ correspond to REAL solutions to the homogeneous equation given by $\Re(\mathbf{x}_j)$ and $\Im(\mathbf{x}_j)$.
4. If the equation is nonhomogeneous, find the particular solution given by
   $$\mathbf{x}^*(t) = \mathsf{X}(t) \int \mathsf{X}^{-1}(t)\mathbf{p}(t)\,dt, \tag{11}$$
   where the MATRIX X is defined by
   $$\mathsf{X}(t) \equiv [\,\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n\,]. \tag{12}$$
   If the equation is homogeneous so that $\mathbf{p}(t) = \mathbf{0}$, then look for a solution of the form
   $$\mathbf{x} = \boldsymbol{\xi} e^{\lambda t}. \tag{13}$$
   This leads to an equation
   $$(\mathsf{A} - \lambda\mathsf{I})\boldsymbol{\xi} = \mathbf{0}, \tag{14}$$
   so ξ is an EIGENVECTOR and λ an EIGENVALUE.
5. The general solution is
   $$\mathbf{x}(t) = \mathbf{x}^*(t) + \sum_{i=1}^{n} c_i \mathbf{x}_i. \tag{15}$$

Ordinary Double Point

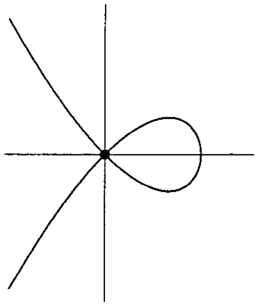

A RATIONAL DOUBLE POINT of CONIC DOUBLE POINT type, known as "A_1." An ordinary DOUBLE POINT is called a NODE. The above plot shows the curve $x^3 - x^2 + y^2 = 0$, which has an ordinary double point at the ORIGIN.

A surface in complex 3-space admits at most finitely many ordinary double points. The maximum possible number of ordinary double points $\mu(d)$ for a surface of degree $d = 1, 2, \ldots$, are 0, 1, 4, 16, 31, 65, $93 \le \mu(7) \le 104$, $168 \le \mu(8) \le 174$, $216 \le \mu(8) \le 246$, $345 \le \mu(10) \le 360$, $425 \le \mu(11) \le 480$, $576 \le \mu(12) \le 645 \ldots$ (Sloane's A046001; Chmutov 1992, Endraß 1995). The fact that $\mu(5) = 31$ was proved by Beauville (1980), and $\mu(6) = 65$ was proved by Jaffe and Ruberman (1994). For $d \ge 3$, the following inequality holds:

$$\mu(d) \le \tfrac{1}{2}[d(d-1)-3]$$

(Endraß 1995). Examples of ALGEBRAIC SURFACES having the maximum (known) number of ordinary double points are given in the following table.

d	$\mu(d)$	Surface
3	4	Cayley cubic
4	16	Kummer surface
5	31	dervish
6	65	Barth sextic
8	168	Endraß octic
10	345	Barth decic

see also ALGEBRAIC SURFACE, BARTH DECIC, BARTH SEXTIC, CAYLEY CUBIC, CUSP, DERVISH, ENDRASS OCTIC, KUMMER SURFACE, RATIONAL DOUBLE POINT

References

Basset, A. B. "The Maximum Number of Double Points on a Surface." *Nature* **73**, 246, 1906.
Beauville, A. "Sur le nombre maximum de points doubles d'une surface dans $\mathbb{P}^3$ ($\mu(5) = 31$)." *Journées de géométrie algébrique d'Angers (1979).* Sijthoff & Noordhoff, pp. 207–215, 1980.
Chmutov, S. V. "Examples of Projective Surfaces with Many Singularities." *J. Algebraic Geom.* **1**, 191–196, 1992.
Endraß, S. "Surfaces with Many Ordinary Nodes." `http://www.mathematik.uni-mainz.de/AlgebraischeGeometrie/docs/Eflaechen.shtml`.
Endraß, S. "Flächen mit vielen Doppelpunkten." *DMV-Mitteilungen* **4**, 17–20, Apr. 1995.
Endraß, S. *Symmetrische Fläche mit vielen gewöhnlichen Doppelpunkten.* Ph.D. thesis. Erlangen, Germany, 1996.
Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, pp. 12–13, 1986.
Jaffe, D. B. and Ruberman, D. "A Sextic Surface Cannot have 66 Nodes." *J. Algebraic Geom.* **6**, 151–168, 1997.
Miyaoka, Y. "The Maximal Number of Quotient Singularities on Surfaces with Given Numerical Invariants." *Math. Ann.* **268**, 159–171, 1984.
Sloane, N. J. A. Sequence A046001 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Togliatti, E. G. "Sulle superficie algebriche col massimo numero di punti doppi." *Rend. Sem. Mat. Torino* **9**, 47–59, 1950.
Varchenko, A. N. "On the Semicontinuity of Spectrum and an Upper Bound for the Number of Singular Points on a Projective Hypersurface." *Dokl. Acad. Nauk SSSR* **270**, 1309–1312, 1983.
Walker, R. J. *Algebraic Curves.* New York: Springer-Verlag, pp. 56–57, 1978.

Ordinary Line

Given an arrangement of $n \ge 3$ points, a LINE containing just two of them is called an ordinary line. Moser (1958) proved that at least $3n/7$ lines must be ordinary (Guy 1989, p. 903).

see also GENERAL POSITION, NEAR-PENCIL, ORDINARY POINT, SPECIAL POINT, SYLVESTER GRAPH

References

Guy, R. K. "Unsolved Problems Come of Age." *Amer. Math. Monthly* **96**, 903–909, 1989.

Ordinary Point

A POINT which lies on at least one ORDINARY LINE.

see also ORDINARY LINE, SPECIAL POINT, SYLVESTER GRAPH

References

Guy, R. K. "Unsolved Problems Come of Age." *Amer. Math. Monthly* **96**, 903–909, 1989.

Ordinate

The y- (vertical) axis of a GRAPH.

see also ABSCISSA, x-AXIS, y-AXIS, z-AXIS

Ore's Conjecture

Define the HARMONIC MEAN of the DIVISORS of n

$$H(n) \equiv \frac{\tau(n)}{\sum_{d|n} \frac{1}{d}},$$

where $\tau(n)$ is the TAU FUNCTION (the number of DIVISORS of n). If n is a PERFECT NUMBER, $H(n)$ is an INTEGER. Ore conjectured that if n is ODD, then $H(n)$ is not an INTEGER. This implies that no ODD PERFECT NUMBERS exist.

see also HARMONIC DIVISOR NUMBER, HARMONIC MEAN, PERFECT NUMBER, TAU FUNCTION

Ore Number

see HARMONIC DIVISOR NUMBER

Ore's Theorem

If a GRAPH G has $n \ge 3$ VERTICES such that every pair of the n VERTICES which are not joined by an EDGE has a sum of VALENCES which is $\ge n$, then G is HAMILTONIAN.

see also HAMILTONIAN GRAPH

Orientable Surface

A REGULAR SURFACE $M \subset \mathbb{R}^n$ is called orientable if each TANGENT SPACE M_p has a COMPLEX STRUCTURE $J_p : M_p \to M_p$ such that $p \to J_p$ is a continuous function.

see also NONORIENTABLE SURFACE, REGULAR SURFACE

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 230, 1993.

Orientation (Plane Curve)

A curve has positive orientation if a region R is on the left when traveling around the outside of R, or on the right when traveling around the inside of R.

Orientation-Preserving

A nonsingular linear MAP $A : \mathbb{R}^n \to \mathbb{R}^n$ is orientation-preserving if $\det(A) > 0$.

see also ORIENTATION-REVERSING, ROTATION

Orientation-Reversing

A nonsingular linear MAP $A : \mathbb{R}^n \to \mathbb{R}^n$ is orientation-reversing if $\det(A) < 0$.

see also ORIENTATION-PRESERVING

Orientation (Vectors)

Let θ be the ANGLE between two VECTORS. If $0 < \theta < \pi$, the VECTORS are positively oriented. If $\pi < \theta < 2\pi$, the vectors are negatively oriented.

Two vectors in the plane

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ and } \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

are positively oriented IFF the DETERMINANT

$$D \equiv \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} > 0,$$

and are negatively oriented IFF the DETERMINANT $D < 0$.

Origami

The Japanese art of paper folding to make 3-dimensional objects. CUBE DUPLICATION and TRISECTION of an ANGLE can be solved using origami, although they cannot be solved using the traditional rules for GEOMETRIC CONSTRUCTIONS.

see also FOLDING, GEOMETRIC CONSTRUCTION, STOMACHION, TANGRAM

References

Andersen, E. "Origami on the Web." `http://www.netspace.org/users/ema/oriweb.html`.

Eppstein, D. "Origami." `http://www.ics.uci.edu/~eppstein/junkyard/origami.html`.

Geretschläger, R. "Euclidean Constructions and the Geometry of Origami." *Math. Mag.* **68**, 357–371, 1995.

Gurkewitz, R. and Arnstein, B. *3-D Geometric Origami.* New York: Dover, 1996.

Kasahara, K. *Origami Omnibus.* Tokyo: Japan Publications, 1988.

Kasahara, K. and Takahara, T. *Origami for the Connoisseur.* Tokyo: Japan Publications, 1987.

Palacios, V. *Fascinating Origami: 101 Models by Alfredo Cerceda.* New York: Dover, 1997.

Pappas, T. "Mathematics & Paperfolding." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 48–50, 1989.

Row, T. S. *Geometric Exercises in Paper Folding.* New York: Dover, 1966.

Tomoko, F. *Unit Origami.* Tokyo: Japan Publications, 1990.

Wu, J. "Joseph Wu's Origami Page." `http://www.datt.co.jp/Origami`.

Origin

The central point ($r = 0$) in POLAR COORDINATES, or the point with all zero coordinates $(0, \ldots, 0)$ in CARTESIAN COORDINATES. In 3-D, the x-AXIS, y-AXIS, and z-AXIS meet at the origin.

see also OCTANT, QUADRANT, x-AXIS, y-AXIS, z-AXIS

Ornstein's Theorem

An important result in ERGODIC THEORY. It states that any two "Bernoulli schemes" with the same MEASURE-THEORETIC ENTROPY are MEASURE-THEORETICALLY ISOMORPHIC.

see also ERGODIC THEORY, ISOMORPHISM, MEASURE THEORY

Orr's Theorem

If

$$(1-z)^{\alpha+\beta+\gamma-1/2}\,{}_2F_1(2\alpha, 2\beta; 2\gamma; z) = \sum a_n z^n, \quad (1)$$

where ${}_2F_1(a, b; c; z)$ is a HYPERGEOMETRIC FUNCTION, then

$$\begin{aligned} &{}_2F_1(\alpha, \beta; \gamma; z)\,{}_2F_1(\gamma-\alpha+\tfrac{1}{2}, \gamma-\beta+\tfrac{1}{2}; \gamma+1; z) \\ &\qquad = \sum_{(\gamma+\frac{1}{2})_n/(\gamma+1)_n} a_n z^n. \quad (2) \end{aligned}$$

Furthermore, if

$$(1-z)^{\alpha+\beta-\gamma-1/2}\,{}_2F_1(2\alpha-1, 2\beta; 2\gamma-1; z) = \sum a_n z^n, \quad (3)$$

then

$$\begin{aligned} &{}_2F_1(\alpha, \beta; \gamma; z)\Gamma(\gamma-\alpha+\tfrac{1}{2}, \gamma-\beta-\tfrac{1}{2}; \gamma; z) \\ &\qquad = \sum_{(\gamma-\frac{1}{2})_n/(\gamma)_n} a_n z^n, \quad (4) \end{aligned}$$

where $\Gamma(z)$ is the GAMMA FUNCTION.

Orthic Triangle

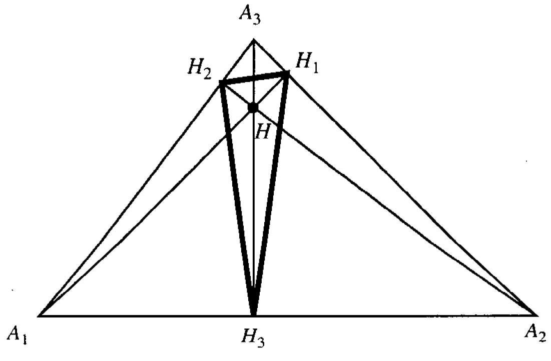

Given a TRIANGLE $\Delta A_1A_2A_3$, the TRIANGLE $\Delta H_1H_2H_3$ with VERTICES at the feet of the ALTITUDES (perpendiculars from a point to the sides) is called the orthic triangle. The three lines A_iH_i are CONCURRENT at the ORTHOCENTER H of $\Delta A_1A_2A_3$.

The centroid of the orthic triangle has TRIANGLE CENTER FUNCTION

$$\alpha = a^2 \cos(B - C)$$

(Casey 1893, Kimberling 1994). The ORTHOCENTER of the orthic triangle has TRIANGLE CENTER FUNCTION

$$\alpha = \cos(2A)\cos(B - C)$$

(Casey 1893, Kimberling 1994). The SYMMEDIAN POINT of the orthic triangle has TRIANGLE CENTER FUNCTION

$$\alpha = \tan A \cos(B - C)$$

(Casey 1893, Kimberling 1994).

see also ALTITUDE, FAGNANO'S PROBLEM, ORTHOCENTER, PEDAL TRIANGLE, SCHWARZ'S TRIANGLE PROBLEM, SYMMEDIAN POINT

References

Casey, J. *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions, with Numerous Examples, 2nd ed., rev. enl.* Dublin: Hodges, Figgis, & Co., p. 9, 1893.

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 9 and 16–18, 1967.

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Orthobicupola

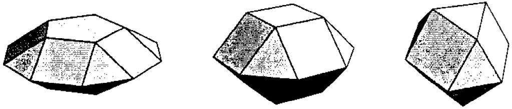

A BICUPOLA in which the bases are in the same orientation.

see also PENTAGONAL ORTHOBICUPOLA, SQUARE ORTHOBICUPOLA, TRIANGULAR ORTHOBICUPOLA

Orthobirotunda

A BIROTUNDA in which the bases are in the same orientation.

Orthocenter

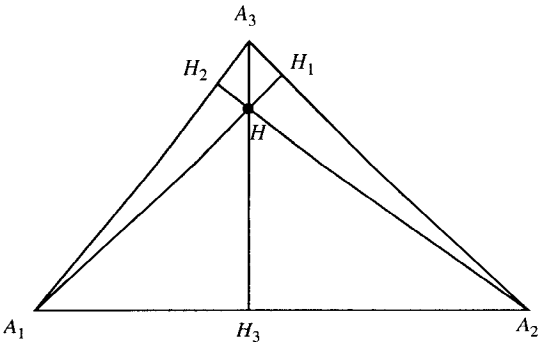

The intersection H of the three ALTITUDES of a TRIANGLE is called the orthocenter. Its TRILINEAR COORDINATES are

$$\cos B \cos C : \cos C \cos A : \cos A \cos B. \qquad (1)$$

If the TRIANGLE is not a RIGHT TRIANGLE, then (1) can be divided through by $\cos A \cos B \cos C$ to give

$$\sec A : \sec B : \sec C. \qquad (2)$$

If the triangle is ACUTE, the orthocenter is in the interior of the triangle. In a RIGHT TRIANGLE, the orthocenter is the VERTEX of the RIGHT ANGLE.

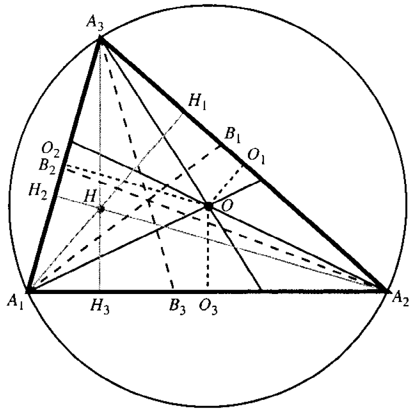

The CIRCUMCENTER O and orthocenter H are ISOGONAL CONJUGATE points. The orthocenter lies on the EULER LINE.

$$a_1{}^2 + a_2{}^2 + a_3{}^2 + \overline{A_1H}^2 + \overline{A_2H}^2 + \overline{A_3H}^2 = 12R^2 \qquad (3)$$

$$\overline{A_1H} + \overline{A_2H} + \overline{A_3H} = 2(r + R), \qquad (4)$$

$$\overline{A_1H}^2 + \overline{A_2H}^2 + \overline{A_3H}^2 = 4R^2 - 4Rr, \qquad (5)$$

where r is the INRADIUS and R is the CIRCUMRADIUS (Johnson 1929, p. 191).

Any HYPERBOLA circumscribed on a TRIANGLE and passing through the orthocenter is RECTANGULAR, and has its center on the NINE-POINT CIRCLE (Falisse 1920, Vandeghen 1965).

see also CENTROID (TRIANGLE), CIRCUMCENTER, EULER LINE, INCENTER, ORTHIC TRIANGLE, ORTHOCENTRIC COORDINATES, ORTHOCENTRIC QUADRILATERAL, ORTHOCENTRIC SYSTEM, POLAR CIRCLE

References

Altshiller-Court, N. *College Geometry: A Second Course in Plane Geometry for Colleges and Normal Schools, 2nd ed.* New York: Barnes and Noble, pp. 165–172, 1952.

Carr, G. S. *Formulas and Theorems in Pure Mathematics, 2nd ed.* New York: Chelsea, p. 622, 1970.

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 36–40, 1967.

Dixon, R. *Mathographics.* New York: Dover, p. 57, 1991.

Falisse, V. *Cours de géométrie analytique plane.* Brussels, Belgium: Office de Publicité, 1920.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 165–172 and 191, 1929.
Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.
Kimberling, C. "Orthocenter." `http://www.evansville.edu/~ck6/tcenters/class/orthocn.html`.
Vandeghen, A. "Some Remarks on the Isogonal and Cevian Transforms. Alignments of Remarkable Points of a Triangle." *Amer. Math. Monthly* **72**, 1091–1094, 1965.

Orthocentric Coordinates

Coordinates defined by an ORTHOCENTRIC SYSTEM.

see also TRILINEAR COORDINATES

Orthocentric Quadrilateral

If two pairs of opposite sides of a COMPLETE QUADRILATERAL are pairs of PERPENDICULAR lines, the QUADRILATERAL is said to be orthocentric. In such a case, the remaining sides are also PERPENDICULAR.

Orthocentric System

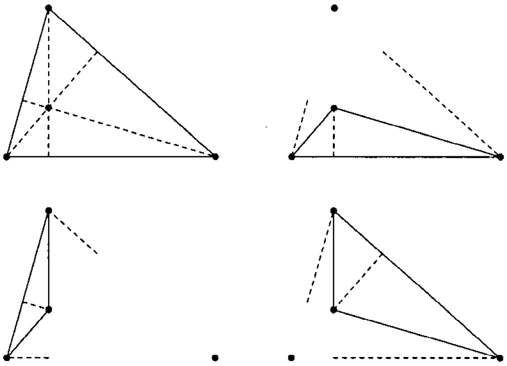

A set of four points, one of which is the ORTHOCENTER of the other three. In an orthocentric system, each point is the ORTHOCENTER of the TRIANGLE of the other three, as illustrated above. The INCENTER and EXCENTERS of a TRIANGLE are an orthocentric system. The centers of the CIRCUMCIRCLES of an orthocentric system form another orthocentric system congruent to the first. The sum of the squares of any nonadjacent pair of connectors of an orthocentric system equals the square of the DIAMETER of the CIRCUMCIRCLE. Orthocentric systems are used to define ORTHOCENTRIC COORDINATES.

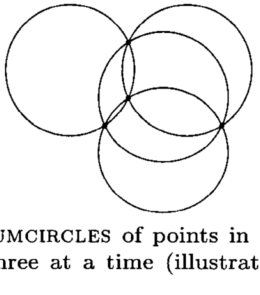

The four CIRCUMCIRCLES of points in an orthocentric system taken three at a time (illustrated above) have equal RADIUS.

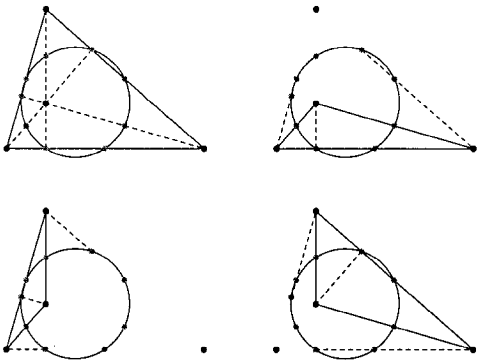

The four triangles of an orthocentric system have a common NINE-POINT CIRCLE, illustrated above.

see also ANGLE BISECTOR, CIRCUMCIRCLE, CYCLIC QUADRANGLE, NINE-POINT CIRCLE, ORTHIC TRIANGLE, ORTHOCENTER, ORTHOCENTRIC SYSTEM, POLAR CIRCLE

References
Altshiller-Court, N. *College Geometry: A Second Course in Plane Geometry for Colleges and Normal Schools, 2nd ed.* New York: Barnes and Noble, pp. 109–114, 1952.
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 165–176, 1929.

Orthocupolarotunda

A CUPOLAROTUNDA in which the bases are in the same orientation.

see also GYROCUPOLAROTUNDA, PENTAGONAL ORTHOCUPOLARONTUNDA

Orthodrome

see GREAT CIRCLE

Orthogonal Array

An orthogonal array OA(k, s) is a $k \times s^2$ ARRAY with entries taken from an s-set S having the property that in any two rows, each ordered pair of symbols from S occurs exactly once.

References
Colbourn, C. J. and Dinitz, J. H. (Eds.) *CRC Handbook of Combinatorial Designs.* Boca Raton, FL: CRC Press, p. 111, 1996.

Orthogonal Basis

A BASIS of vectors $\mathbf{x}$ which satisfy

$$x_j x_k = C_{jk}\delta_{jk}$$

$$x^\mu x_\nu = C^\mu_\nu \delta^\mu_\nu,$$

where C_{jk}, C^μ_ν are constants (not necessarily equal to 1) and δ_{jk} is the KRONECKER DELTA.

see also BASIS, ORTHONORMAL BASIS

Orthogonal Circles

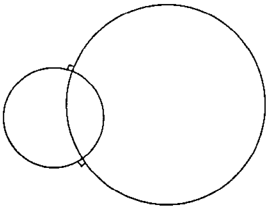

Orthogonal circles are ORTHOGONAL CURVES, i.e., they cut one another at RIGHT ANGLES. Two CIRCLES with equations

$$x^2+y^2+2gx+2fy+c=0 \tag{1}$$

$$x^2+y^2+2g'x+2f'y+c'=0 \tag{2}$$

are orthogonal if

$$2gg'+2ff'=c+c'. \tag{3}$$

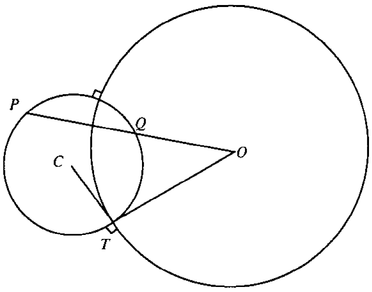

A theorem of Euclid states that, for the orthogonal circles in the above diagram,

$$OP\times OQ=OT^2 \tag{4}$$

(Dixon 1991, p. 65).

References
Dixon, R. *Mathographics.* New York: Dover, pp. 65–66, 1991.
Euclid. *The Thirteen Books of the Elements, 2nd ed. unabridged, Vol. 3: Books X–XIII* New York: Dover, p. 36, 1956.
Pedoe, D. *Circles: A Mathematical View, rev. ed.* Washington, DC: Math. Assoc. Amer., p. xxiv, 1995.

Orthogonal Curves

Two intersecting curves which are PERPENDICULAR at their INTERSECTION are said to be orthogonal.

Orthogonal Functions

Two functions $f(x)$ and $g(x)$ are orthogonal on the interval $a\leq x\leq b$ if

$$\langle f(x)|g(x)\rangle\equiv\int_a^b f(x)g(x)\,dx=0.$$

see also ORTHOGONAL POLYNOMIALS, ORTHONORMAL FUNCTIONS

Orthogonal Group

see GENERAL ORTHOGONAL GROUP, LIE-TYPE GROUP, ORTHOGONAL ROTATION GROUP, PROJECTIVE GENERAL ORTHOGONAL GROUP, PROJECTIVE SPECIAL ORTHOGONAL GROUP, SPECIAL ORTHOGONAL GROUP

References
Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas#orth`.

Orthogonal Group Representations

Two representations of a GROUP χ_i and χ_j are said to be orthogonal if

$$\sum_R \chi_i(R)\chi_j(R)=0$$

for $i\neq j$, where the sum is over all elements R of the representation.

see also GROUP

Orthogonal Lines

Two or more LINES or LINE SEGMENTS which are PERPENDICULAR are said to be orthogonal.

Orthogonal Matrix

Any ROTATION can be given as a composition of rotations about three axes (EULER'S ROTATION THEOREM), and thus can be represented by a 3×3 MATRIX operating on a VECTOR,

$$\begin{bmatrix}x_1'\\x_2'\\x_3'\end{bmatrix}=\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}. \tag{1}$$

We wish to place conditions on this matrix so that it is consistent with an ORTHOGONAL TRANSFORMATION (basically, a ROTATION or ROTOINVERSION).

In a ROTATION, a VECTOR must keep its original length, so it must be true that

$$x_i'x_i'=x_ix_i \tag{2}$$

for $i=1$, 2, 3, where EINSTEIN SUMMATION is being used. Therefore, from the transformation equation,

$$(a_{ij}x_j)(a_{ik}x_k)=x_ix_i. \tag{3}$$

This can be rearranged to

$$\begin{aligned}a_{ij}(x_ja_{ik})x_k&=a_{ij}(a_{ik}x_j)x_k\\&=a_{ij}a_{ik}x_jx_k=x_ix_i.\end{aligned} \tag{4}$$

In order for this to hold, it must be true that

$$a_{ij}a_{ik}=\delta_{jk} \tag{5}$$

for $j, k = 1, 2, 3$, where δ_{ij} is the KRONECKER DELTA. This is known as the ORTHOGONALITY CONDITION, and it guarantees that

$$\mathsf{A}^{-1} = \mathsf{A}^{\mathrm{T}}, \tag{6}$$

and

$$\mathsf{A}^{\mathrm{T}}\mathsf{A} = \mathsf{I}, \tag{7}$$

where A^{T} is the MATRIX TRANSPOSE and I is the IDENTITY MATRIX. Equation (7) is the identity which gives the orthogonal matrix its name. Orthogonal matrices have special properties which allow them to be manipulated and identified with particular ease.

Let A and B be two orthogonal matrices. By the ORTHOGONALITY CONDITION, they satisfy

$$a_{ij}a_{ik} = \delta_{jk}, \tag{8}$$

and

$$b_{ij}b_{ik} = \delta_{jk}, \tag{9}$$

where δ_{ij} is the KRONECKER DELTA. Now

$$\begin{aligned} c_{ij}c_{ik} &= (ab)_{ij}(ab)_{jk} = a_{is}b_{sj}a_{it}b_{tk} = a_{is}a_{it}b_{sj}b_{tk} \\ &= \delta_{st}b_{sj}b_{tk} = b_{tj}b_{tk} = \delta_{jk}, \end{aligned} \tag{10}$$

so the product $\mathsf{C} \equiv \mathsf{AB}$ of two orthogonal matrices is also orthogonal.

The EIGENVALUES of an orthogonal matrix must satisfy one of the following:

1. All EIGENVALUES are 1.
2. One EIGENVALUE is 1 and the other two are -1.
3. One EIGENVALUE is 1 and the other two are COMPLEX CONJUGATES of the form $e^{i\theta}$ and $e^{-i\theta}$.

An orthogonal MATRIX A is classified as proper (corresponding to pure ROTATION) if

$$\det(\mathsf{A}) = 1, \tag{11}$$

where $\det(\mathsf{A})$ is the DETERMINANT of A, or improper (corresponding to inversion with possible rotation; ROTOINVERSION) if

$$\det(\mathsf{A}) = -1. \tag{12}$$

see also EULER'S ROTATION THEOREM, ORTHOGONAL TRANSFORMATION, ORTHOGONALITY CONDITION, ROTATION, ROTATION MATRIX, ROTOINVERSION

References

Arfken, G. "Orthogonal Matrices." *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 191–205, 1985.

Goldstein, H. "Orthogonal Transformations." §4–2 in *Classical Mechanics, 2nd ed.* Reading, MA: Addison-Wesley, 132–137, 1980.

Orthogonal Polynomials

Orthogonal polynomials are classes of POLYNOMIALS $\{p_n(x)\}$ over a range $[a, b]$ which obey an ORTHOGONALITY relation

$$\int_a^b w(x)p_m(x)p_n(x)\,dx = \delta_{mn}c_n, \tag{1}$$

where $w(x)$ is a WEIGHTING FUNCTION and δ is the KRONECKER DELTA. If $c_m = 1$, then the POLYNOMIALS are not only orthogonal, but orthonormal.

Orthogonal polynomials have very useful properties in the solution of mathematical and physical problems. Just as FOURIER SERIES provide a convenient method of expanding a periodic function in a series of linearly independent terms, orthogonal polynomials provide a natural way to solve, expand, and interpret solutions to many types of important DIFFERENTIAL EQUATIONS. Orthogonal polynomials are especially easy to generate using GRAM-SCHMIDT ORTHONORMALIZATION. Abramowitz and Stegun (1972, pp. 774–775) give a table of common orthogonal polynomials.

Type	Interval	$w(x)$	c_n
Chebyshev First Kind	$[-1, 1]$	$(1-x^2)^{-1/2}$	$\begin{cases} \frac{1}{2}\pi \\ \pi \end{cases}$ $\begin{cases} \text{for } n = 0 \\ \text{otherwise} \end{cases}$
Chebyshev Second Kind	$[-1, 1]$	$\sqrt{1-x^2}$	$\frac{1}{2}\pi$
Hermite	$(-\infty, \infty)$	e^{-x^2}	$\sqrt{\pi}\,2^n n!$
Jacobi	$(-1, 1)$	$(1-x)^\alpha(1+x)^\beta$	h_n
Laguerre	$[0, \infty)$	e^{-x}	1
Laguerre (Associated)	$[0, \infty)$	$x^k e^{-x}$	$\frac{(n+k)!}{n!}$
Legendre	$[-1, 1]$	1	$\frac{2}{2n+1}$
Ultraspherical	$[-1, 1]$	$(1-x^2)^{\alpha-1/2}$	$\begin{cases} \frac{\pi 2^{1-2\alpha}\Gamma(n+2\alpha)}{n!(n+\alpha)[\Gamma(\alpha)]^2} \\ \frac{2\pi}{n^2} \end{cases}$ $\begin{cases} \text{for } \alpha \neq 0 \\ \text{for } \alpha = 0 \end{cases}$

In the above table, the normalization constant is the value of

$$c_n \equiv \int w(x)[p_n(x)]^2\,dx \tag{2}$$

and

$$h_n \equiv \frac{2^{\alpha+\beta+1}}{2n+\alpha+\beta+1}\frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{n!\Gamma(n+\alpha+\beta+1)}, \tag{3}$$

where $\Gamma(z)$ is a GAMMA FUNCTION.

The ROOTS of orthogonal polynomials possess many rather surprising and useful properties. For instance, let $x_1 < x_2 < \ldots < x_n$ be the ROOTS of the $p_n(x)$ with $x_0 = a$ and $x_{n+1} = b$. Then each interval $[x_\nu, x_{\nu+1}]$ for $\nu = 0, 1, \ldots, n$ contains exactly one ROOT of $p_{n+1}(x)$. Between two ROOTS of $p_n(x)$ there is at least one ROOT of $p_m(x)$ for $m > n$.

Let c be an arbitrary REAL constant, then the POLYNOMIAL

$$p_{n+1}(x) - cp_n(x) \tag{4}$$

has $n+1$ distinct REAL ROOTS. If $c > 0$ ($c < 0$), these ROOTS lie in the interior of $[a, b]$, with the exception of the greatest (least) ROOT which lies in $[a, b]$ only for

$$c \leq \frac{p_{n+1}(b)}{p_n(b)} \qquad \left(c \geq \frac{p_{n+1}(a)}{p_n(a)}\right). \tag{5}$$

The following decomposition into partial fractions holds

$$\frac{p_n(x)}{p_{n+1}(x)} = \sum_{\nu=0}^{n} \frac{l_\nu}{x - \xi}, \tag{6}$$

where $\{\xi_\nu\}$ are the ROOTS of $p_{n+1}(x)$ and

$$\begin{aligned} l_\nu &= \frac{p_n(\xi_\nu)}{p'_{n+1}(\xi_\nu)} \\ &= \frac{p'_{n+1}(\xi_\nu)p_n(\xi_\nu) - p'_n(\xi_\nu)'p_{n+1}(\xi_\nu)}{[p'_{n+1}(\xi_\nu)]^2} > 0. \end{aligned} \tag{7}$$

Another interesting property is obtained by letting $\{p_n(x)\}$ be the orthonormal set of POLYNOMIALS associated with the distribution $d\alpha(x)$ on $[a, b]$. Then the CONVERGENTS R_n/S_n of the CONTINUED FRACTION

$$\frac{1}{A_1x + B_1} - \frac{C_2}{A_2x + B_2} - \frac{C_3}{A_3x + B_3} - \ldots - \frac{C_n}{A_nx + B_n} + \ldots \tag{8}$$

are given by

$$\begin{aligned} R_n &= R_n(x) \\ &= {c_0}^{-3/2}\sqrt{c_0c_2 - {c_1}^2} \int_a^b \frac{p_n(x) - p_n(t)}{x - t}\,d\alpha(t) \quad (9) \\ S_n &= S_n(x) = \sqrt{c_0}\,p_n(x), \quad (10) \end{aligned}$$

where $n = 0, 1, \ldots$ and

$$c_n = \int_a^b x^n \, d\alpha(x). \tag{11}$$

Furthermore, the ROOTS of the orthogonal polynomials $p_n(x)$ associated with the distribution $d\alpha(x)$ on the interval $[a, b]$ are REAL and distinct and are located in the interior of the interval $[a, b]$.

see also CHEBYSHEV POLYNOMIAL OF THE FIRST KIND, CHEBYSHEV POLYNOMIAL OF THE SECOND KIND, GRAM-SCHMIDT ORTHONORMALIZATION, HERMITE POLYNOMIAL, JACOBI POLYNOMIAL, KRAWTCHOUK POLYNOMIAL, LAGUERRE POLYNOMIAL, LEGENDRE POLYNOMIAL, ORTHOGONAL FUNCTIONS, SPHERICAL HARMONIC, ULTRASPHERICAL POLYNOMIAL, ZERNIKE POLYNOMIAL

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Orthogonal Polynomials." Ch. 22 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 771–802, 1972.

Arfken, G. "Orthogonal Polynomials." *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 520–521, 1985.

Iyanaga, S. and Kawada, Y. (Eds.). "Systems of Orthogonal Functions." Appendix A, Table 20 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1477, 1980.

Nikiforov, A. F.; Uvarov, V. B.; and Suslov, S. S. *Classical Orthogonal Polynomials of a Discrete Variable.* New York: Springer-Verlag, 1992.

Sansone, G. *Orthogonal Functions.* New York: Dover, 1991.

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., pp. 44–47 and 54–55, 1975.

Orthogonal Projection

A PROJECTION of a figure by parallel rays. In such a projection, tangencies are preserved. Parallel lines project to parallel lines. The ratio of lengths of parallel segments is preserved, as is the ratio of areas.

Any TRIANGLE can be positioned such that its shadow under an orthogonal projection is EQUILATERAL. Also, the MEDIANS of a TRIANGLE project to the MEDIANS of the image TRIANGLE. ELLIPSES project to ELLIPSES, and any ELLIPSE can be projected to form a CIRCLE. The center of an ELLIPSE projects to the center of the image ELLIPSE. The CENTROID of a TRIANGLE projects to the CENTROID of its image. Under an ORTHOGONAL TRANSFORMATION, the MIDPOINT ELLIPSE can be transformed into a CIRCLE INSCRIBED in an EQUILATERAL TRIANGLE.

SPHEROIDS project to ELLIPSES (or CIRCLE in the DEGENERATE case).

see also PROJECTION

Orthogonal Rotation Group

Orthogonal rotation groups are LIE GROUPS. The orthogonal rotation group $O_3(n)$ is the set of $n \times n$ REAL ORTHOGONAL MATRICES.

The orthogonal rotation group $O_3^-(n)$ is the set of $n \times n$ REAL ORTHOGONAL MATRICES (having $n(n-1)/2$ independent parameters) with DETERMINANT -1.

The orthogonal rotation group $O_3^+(n)$ is the set of $n \times n$ REAL ORTHOGONAL MATRICES, having $n(n-1)/2$ independent parameters, with DETERMINANT $+1$. $O_3^+(n)$ is

HOMEOMORPHIC with $SU(2)$. Its elements can be written using EULER ANGLES and ROTATION MATRICES as

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

$$R_x(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix} \tag{2}$$

$$R_y(\theta) = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \tag{3}$$

$$R_z(\psi) = \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4}$$

References

Arfken, G. "Orthogonal Group, O_3^+." *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 252–253, 1985.

Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas#orth`.

Orthogonal Tensors

Orthogonal CONTRAVARIANT and COVARIANT satisfy

$$g_{ik}g^{ij} = \delta_k^j,$$

where δ_j^k is the KRONECKER DELTA.

see also CONTRAVARIANT TENSOR, COVARIANT TENSOR

Orthogonal Transformation

Any linear transformation

$$\begin{aligned} x_1' &= a_{11}x_1 + a_{12}x_2 + x_{13}x_3 \\ x_2' &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ x_3' &= a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \end{aligned}$$

satisfying the ORTHOGONALITY CONDITION

$$a_{ij}a_{ik} = \delta_{jk},$$

where EINSTEIN SUMMATION has been used and δ_{ij} is the KRONECKER DELTA, is called an orthogonal transformation.

Orthogonal transformations correspond to rigid ROTATIONS (or ROTOINVERSIONS), and may be represented using ORTHOGONAL MATRICES. If $A : \mathbb{R}^n \to \mathbb{R}^n$ is an orthogonal transformation, then $\det(A) = \pm 1$.

see also AFFINE TRANSFORMATION, ORTHOGONAL MATRIX, ORTHOGONALITY CONDITION, ROTATION, ROTOINVERSION

References

Goldstein, H. "Orthogonal Transformations." §4-2 in *Classical Mechanics, 2nd ed.* Reading, MA: Addison-Wesley, 132–137, 1980.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 104, 1993.

Orthogonal Vectors

Two vectors **u** and **v** whose DOT PRODUCT is $\mathbf{u} \cdot \mathbf{v} = 0$ (i.e., the vectors are PERPENDICULAR) are said to be orthogonal. The definition can be extended to three or more vectors which are mutually PERPENDICULAR.

see also DOT PRODUCT, PERPENDICULAR

Orthogonality Condition

A linear transformation

$$\begin{aligned} x_1' &= a_{11}x_1 + a_{12}x_2 + x_{13}x_3 \\ x_2' &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ x_3' &= a_{31}x_1 + a_{32}x_2 + a_{33}x_3, \end{aligned}$$

is said to be an ORTHOGONAL TRANSFORMATION if it satisfies the orthogonality condition

$$a_{ij}a_{ik} = \delta_{jk},$$

where EINSTEIN SUMMATION has been used and δ_{ij} is the KRONECKER DELTA.

see also ORTHOGONAL TRANSFORMATION

References

Goldstein, H. "Orthogonal Transformations." §4-2 in *Classical Mechanics, 2nd ed.* Reading, MA: Addison-Wesley, 132–137, 1980.

Orthogonality Theorem

see GROUP ORTHOGONALITY THEOREM

Orthographic Projection

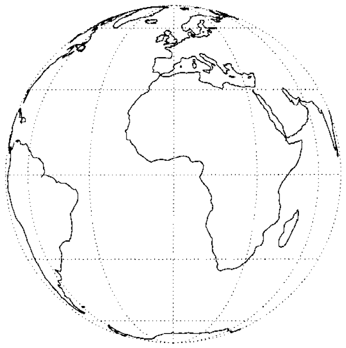

A projection from infinity which preserves neither AREA nor angle.

$$\begin{aligned} x &= \cos\phi \sin(\lambda - \lambda_0) && (1) \\ y &= \cos\phi_1 \sin\phi - \sin\phi_1 \cos\phi \cos(\lambda - \lambda_0). && (2) \end{aligned}$$

The inverse FORMULAS are

$$\phi = \sin^{-1}\left(\cos c \sin\phi_1 + \frac{y \sin c \cos\phi_1}{\rho}\right) \tag{3}$$

$$\lambda = \lambda_0 + \tan^{-1}\left(\frac{x \sin c}{\rho \cos\phi_1 \cos c - y \sin\phi_1 \sin c}\right), \tag{4}$$

where

$$\rho = \sqrt{x^2 + y^2} \tag{5}$$

$$c = \sin^{-1} \rho. \tag{6}$$

References

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 145–153, 1987.

Orthologic

Two TRIANGLES $A_1B_1C_1$ and $A_2B_2C_2$ are orthologic if the perpendiculars from the VERTICES A_1, B_1, C_1 on the sides B_2C_2, A_2C_2, and A_2B_2 pass through one point. This point is known as the orthology center of TRIANGLE 1 with respect to TRIANGLE 2.

Orthonormal Basis

A BASIS of VECTORS **x** which satisfy

$$x_j x_k = \delta_{jk}$$

and

$$x^\mu x_\nu = \delta^\mu_\nu,$$

where δ_{jk} is the KRONECKER DELTA. An orthonormal basis is a normalized ORTHOGONAL BASIS.

see also BASIS, ORTHOGONAL BASIS

Orthonormal Functions

A pair of functions ϕ_i and ϕ_j are orthonormal if they are ORTHOGONAL and each normalized. These two conditions can be succinctly written as

$$\int_a^b \phi_i(x)\phi_j(x)w(x)\,dx = \delta_{ij},$$

where $w(x)$ is a WEIGHTING FUNCTION and δ_{ij} is the KRONECKER DELTA.

see also ORTHOGONAL POLYNOMIALS

Orthonormal Vectors

UNIT VECTORS which are ORTHOGONAL are said to be orthonormal.

see also ORTHOGONAL VECTORS

Orthopole

If perpendiculars are dropped on any line from the vertices of a TRIANGLE, then the perpendiculars to the opposite sides from their FEET are CONCURRENT at a point called the orthopole.

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, p. 247, 1929.

Orthoptic Curve

An ISOPTIC CURVE formed from the locus of TANGENTS meeting at RIGHT ANGLES. The orthoptic of a PARABOLA is its DIRECTRIX. The orthoptic of a central CONIC was investigated by Monge and is a CIRCLE concentric with the CONIC SECTION. The orthoptic of an ASTROID is a CIRCLE.

Curve	Orthoptic
astroid	quadrifolium
cardioid	circle or limaçon
deltoid	circle
logarithmic spiral	equal logarithmic spiral
parabola	directrix

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 58 and 207, 1972.

Orthotomic

Given a source S and a curve γ, pick a point on γ and find its tangent T. Then the LOCUS of reflections of S about tangents T is the orthotomic curve (also known as the secondary CAUSTIC). The INVOLUTE of the orthotomic is the CAUSTIC. For a parametric curve $(f(t), g(t))$ with respect to the point (x_0, y_0), the orthotomic is

$$x = x_0 - \frac{2g'[f'(g-y_0) - g'(f-x_0)]}{f'^2 + g'^2}$$

$$y = y_0 + \frac{2f'[f'(g-y_0) - g'(f-x_0)]}{f'^2 + g'^2}.$$

see also CAUSTIC, INVOLUTE

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, p. 60, 1972.

Orthotope

A PARALLELOTOPE whose edges are all mutually PERPENDICULAR. The orthotope is a generalization of the RECTANGLE and RECTANGULAR PARALLELEPIPED.

see also RECTANGLE, RECTANGULAR PARALLELEPIPED

References

Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, pp. 122–123, 1973.

Osborne's Rule

The prescription that a TRIGONOMETRY identity can be converted to an analogous identity for HYPERBOLIC FUNCTIONS by expanding, exchanging trigonometric functions with their hyperbolic counterparts, and then flipping the sign of each term involving the product of two HYPERBOLIC SINES. For example, given the identity

$$\cos(x-y) = \cos x \cos y + \sin x \sin y,$$

Osborne's rule gives the corresponding identity

$$\cosh(x-y) = \cosh x \cosh y - \sinh x \sinh y.$$

see also HYPERBOLIC FUNCTIONS, TRIGONOMETRY

Oscillation

The variation of a FUNCTION which exhibits SLOPE changes, also called the SALTUS of a function.

Oscillation Land

see CAROTID-KUNDALINI FUNCTION

Osculating Circle

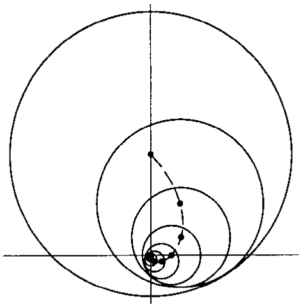

The CIRCLE which shares the same TANGENT as a curve at a given point. The RADIUS OF CURVATURE of the osculating circle is

$$\rho(t) = \frac{1}{|\kappa(t)|}, \tag{1}$$

where κ is the CURVATURE, and the center is

$$x = f - \frac{(f'^2 + g'^2)g'}{f'g'' - f''g'} \tag{2}$$

$$y = g + \frac{(f'^2 + g'^2)g'}{f'g'' - f''g'}, \tag{3}$$

i.e., the centers of the osculating circles to a curve form the EVOLUTE to that curve.

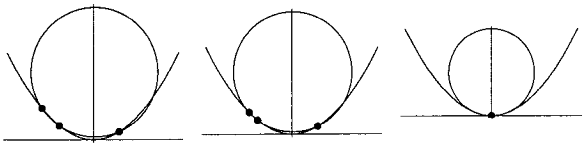

In addition, let $C(t_1, t_2, t_3)$ denote the CIRCLE passing through three points on a curve $(f(t), g(t))$ with $t_1 < t_2 < t_3$. Then the osculating circle C is given by

$$C = \lim_{t_1,t_2,t_3 \to t} C(t_1, t_2, t_3) \tag{4}$$

(Gray 1993).

see also CURVATURE, EVOLUTE, RADIUS OF CURVATURE, TANGENT

References

Gardner, M. "The Game of Life, Parts I–III." Chs. 20–22 in *Wheels, Life, and other Mathematical Amusements.* New York: W. H. Freeman, pp. 221, 237, and 243, 1983.

Gray, A. "Osculating Circles to Plane Curves." §5.6 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 90–95, 1993.

Osculating Curves

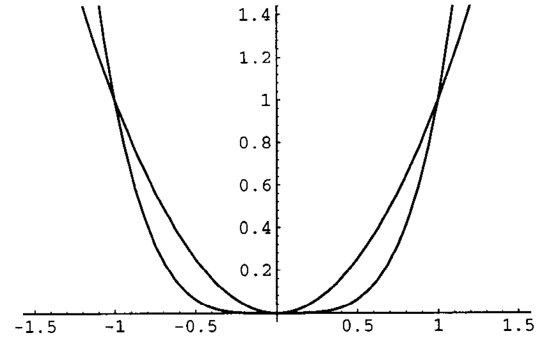

An osculating curve to $f(x)$ at x_0 is tangent at that point and has the same CURVATURE. It therefore satisfies

$$y^{(k)}(x_0) = f^{(k)}(x_0)$$

for $k = 0, 1, 2$. The point of tangency is called a TACNODE. The simplest example of osculating curves are x^2 and x^4, which osculate at the point $x_0 = 0$.

see also TACNODE

Osculating Interpolation

see HERMITE'S INTERPOLATING FUNDAMENTAL POLYNOMIAL

Osculating Plane

The PLANE spanned by the three points $\mathbf{x}(t)$, $\mathbf{x}(t+h_1)$, and $\mathbf{x}(t+h_2)$ on a curve as $h_1, h_2 \to 0$. Let $\mathbf{z}$ be a point on the osculating plane, then

$$[(\mathbf{z} - \mathbf{x}), \mathbf{x}', \mathbf{x}''] = 0,$$

where $[\mathbf{A}, \mathbf{B}, \mathbf{C}]$ denotes the SCALAR TRIPLE PRODUCT. The osculating plane passes through the tangent. The intersection of the osculating plane with the NORMAL PLANE is known as the PRINCIPAL NORMAL VECTOR. The VECTORS $\mathbf{T}$ and $\mathbf{N}$ (TANGENT VECTOR and NORMAL VECTOR) span the osculating plane.

see also NORMAL VECTOR, OSCULATING SPHERE, SCALAR TRIPLE PRODUCT, TANGENT VECTOR

Osculating Sphere

The center of any SPHERE which has a contact of (at least) first-order with a curve C at a point P lies in the normal plane to C at P. The center of any SPHERE which has a contact of (at least) second-order with C at point P, where the CURVATURE $\kappa > 0$, lies on the polar axis of C corresponding to P. All these SPHERES intersect the OSCULATING PLANE of C at P along a circle of curvature at P. The osculating sphere has center

$$\mathbf{a} = \mathbf{x} + \rho\hat{\mathbf{N}} + \frac{\dot{\rho}}{\tau}\hat{\mathbf{B}}$$

where $\hat{\mathbf{N}}$ is the unit NORMAL VECTOR, $\hat{\mathbf{B}}$ is the unit BINORMAL VECTOR, ρ is the RADIUS OF CURVATURE, and τ is the TORSION, and RADIUS

$$R = \sqrt{\rho^2 + \left(\frac{\dot{\rho}}{\tau}\right)^2},$$

and has contact of (at least) third order with C.

see also CURVATURE, OSCULATING PLANE, RADIUS OF CURVATURE, SPHERE, TORSION (DIFFERENTIAL GEOMETRY)

References
Kreyszig, E. *Differential Geometry.* New York: Dover, pp. 54–55, 1991.

Osedelec Theorem

For an n-D MAP, the LYAPUNOV CHARACTERISTIC EXPONENTS are given by

$$\sigma_i = \lim_{N\to\infty} \ln|\lambda_i(N)|$$

for $i = 1, \ldots, n$, where λ_i is the LYAPUNOV CHARACTERISTIC NUMBER.

see also LYAPUNOV CHARACTERISTIC EXPONENT, LYAPUNOV CHARACTERISTIC NUMBER

Ostrowski's Inequality

Let $f(x)$ be a monotonic function integrable on $[a, b]$ and let $f(a), f(b) \leq 0$ and $|f(a)| \geq |f(b)|$, then if g is a REAL function integrable on $[a, b]$,

$$\left|\int_a^b f(x)g(x)\,dx\right| \leq |f(a)| \max_{a\leq\xi\leq b}\left|\int_a^\xi g(x)\,dx\right|.$$

References
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1100, 1979.

Ostrowski's Theorem

Let $\mathsf{A} = a_{ij}$ be a MATRIX with POSITIVE COEFFICIENTS and λ_0 be the POSITIVE EIGENVALUE in the FROBENIUS THEOREM, then the $n-1$ EIGENVALUES $\lambda_j \neq \lambda_0$ satisfy the INEQUALITY

$$|\lambda_j| \leq \lambda_0 \frac{M^2 - m^2}{M^2 + m^2},$$

where

$$M = \max_{i,j} a_{ij}$$
$$m = \min_{i,j} a_{ij}$$

and $i, j = 1, 2, \ldots, n$.

see also FROBENIUS THEOREM

References
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1121, 1980.

Otter's Tree Enumeration Constants

see TREE

Outdegree

The number of outward directed EDGES from a given VERTEX in a DIRECTED GRAPH.

see also DIRECTED GRAPH, INDEGREE, LOCAL DEGREE

Outer Automorphism Group

A particular type of AUTOMORPHISM GROUP which exists only for GROUPS. For a GROUP G, the outer automorphism group is the QUOTIENT GROUP $\mathrm{Aut}(G)/\mathrm{Inn}(G)$, which is the AUTOMORPHISM GROUP of G modulo its INNER AUTOMORPHISM GROUP.

see also AUTOMORPHISM GROUP, INNER AUTOMORPHISM GROUP, QUOTIENT GROUP

Outer Product

see DIRECT PRODUCT (TENSOR)

Oval

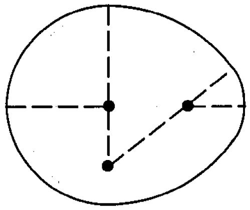

An oval is a curve resembling a squashed CIRCLE but, unlike the ELLIPSE, without a precise mathematical definition. The word oval derived from the Latin word "ovus" for egg. Unlike ellipses, ovals sometimes have only a single axis of reflection symmetry (instead of two).

Ovals can be constructed with a COMPASS by joining together arcs of different radii such that the centers of the arcs lie on a line passing through the join point (Dixon 1991). Albrecht Dürer used this method to design a Roman letter font.

see also CARTESIAN OVALS, CASSINI OVALS, EGG, ELLIPSE, OVOID, SUPERELLIPSE

References
Critchlow, K. *Time Stands Still.* London: Gordon Fraser, 1979.
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., 1989.
Dixon, R. *Mathographics.* New York: Dover, pp. 3–11, 1991.
Dixon, R. "The Drawing Out of an Egg." *New Sci.*, July 29, 1982.
Pedoe, D. *Geometry and the Liberal Arts.* London: Peregrine, 1976.

Oval of Descartes

see CARTESIAN OVALS

Ovals of Cassini

see CASSINI OVALS

Overlapping Resonance Method

see RESONANCE OVERLAP METHOD

Oversampling

A signal sampled at a frequency higher than the NYQUIST FREQUENCY is said to be oversampled β times, where the oversampling ratio is defined as

$$\beta \equiv \frac{\nu_{\text{sampling}}}{\nu_{\text{Nyquist}}}.$$

see also NYQUIST FREQUENCY, NYQUIST SAMPLING

Ovoid

An egg-shaped curve. Lockwood (1967) calls the NEGATIVE PEDAL CURVE of an ELLIPSE with ECCENTRICITY $e \leq 1/2$ an ovoid.

see also OVAL

References

Lockwood, E. H. *A Book of Curves.* Cambridge, England: Cambridge University Press, p. 157, 1967.

P

p-adic Number

A *p*-adic number is an extension of the FIELD of RATIONAL NUMBERS such that CONGRUENCES MODULO POWERS of a fixed PRIME *p* are related to proximity in the so called "*p*-adic metric."

Any NONZERO RATIONAL NUMBER x can be represented by

$$x = \frac{p^a r}{s}, \tag{1}$$

where p is a PRIME NUMBER, r and s are INTEGERS not DIVISIBLE by p, and a is a unique INTEGER. Then define the *p*-adic absolute value of x by

$$|x|_p = p^{-a}. \tag{2}$$

Also define the *p*-adic value

$$|0|_p = 0. \tag{3}$$

As an example, consider the FRACTION

$$\tfrac{140}{297} = 2^2 \cdot 3^{-3} \cdot 5 \cdot 7 \cdot 11^{-1}. \tag{4}$$

It has *p*-adic absolute values given by

$$|\tfrac{140}{297}|_2 = \tfrac{1}{4} \tag{5}$$

$$|\tfrac{140}{297}|_3 = 27 \tag{6}$$

$$|\tfrac{140}{297}|_5 = \tfrac{1}{5} \tag{7}$$

$$|\tfrac{140}{297}|_7 = \tfrac{1}{7} \tag{8}$$

$$|\tfrac{140}{297}|_{11} = 11. \tag{9}$$

The *p*-adic absolute value satisfies the relations

1. $|x|_p \geq 0$ for all x,
2. $|x|_p = 0$ IFF $x = 0$,
3. $|xy|_p = |x|_p\, |y|_p$ for all x and y,
4. $|x + y|_p \leq |x|_p + |y|_p$ for all x and y (the TRIANGLE INEQUALITY), and
5. $|x+y|_p \leq \max(|x|_p, |y|_p)$ for all x and y (the STRONG TRIANGLE INEQUALITY).

In the above, relation 4 follows trivially from relation 5, but relations 4 and 5 are relevant in the more general VALUATION THEORY.

The *p*-adics were probably first introduced by Hensel in 1902 in a paper which was concerned with the development of algebraic numbers in POWER SERIES. *p*-adic numbers were then generalized to VALUATIONS by Kürschák in 1913. In the early 1920s, Hasse formulated the LOCAL-GLOBAL PRINCIPLE (now usually called the HASSE PRINCIPLE), which is one of the chief applications of LOCAL FIELD theory. Skolem's *p*-adic method, which is used in attacking certain DIOPHANTINE EQUATIONS, is another powerful application of *p*-adic numbers. Another application is the theorem that the HARMONIC NUMBERS H_n are never INTEGERS (except for H_1). A similar application is the proof of the VON STAUDT-CLAUSEN THEOREM using the *p*-adic valuation, although the technical details are somewhat difficult. Yet another application is provided by the MAHLER-LECH THEOREM.

Every RATIONAL x has an "essentially" unique *p*-adic expansion ("essentially" since zero terms can always be added at the beginning)

$$x = \sum_{j=m}^{\infty} a_j p^j, \tag{10}$$

with m an INTEGER, a_j the INTEGERS between 0 and $p - 1$ inclusive, and where the sum is convergent with respect to *p*-adic valuation. If $x \neq 0$ and $a_m \neq 0$, then the expansion is unique. Burger and Struppeck (1996) show that for p a PRIME and n a POSITIVE INTEGER,

$$|n!|_p = p^{-(n-A_p(n))/(p-1)}, \tag{11}$$

where the *p*-adic expansion of n is

$$n = a_0 + a_1 p + a_2 p^2 + \ldots + a_L P^L, \tag{12}$$

and

$$A_p(n) = a_0 + a_1 + \ldots + a_L. \tag{13}$$

For sufficiently large n,

$$|n!|_p \leq p^{-n/(2p-2)}. \tag{14}$$

The *p*-adic valuation on $\mathbb{Q}$ gives rise to the *p*-adic metric

$$d(x, y) = |x - y|_p, \tag{15}$$

which in turn gives rise to the *p*-adic topology. It can be shown that the rationals, together with the *p*-adic metric, do not form a COMPLETE METRIC SPACE. The completion of this space can therefore be constructed, and the set of *p*-adic numbers $\mathbb{Q}_p$ is defined to be this completed space.

Just as the REAL NUMBERS are the completion of the RATIONALS $\mathbb{Q}$ with respect to the usual absolute valuation $|x - y|$, the *p*-adic numbers are the completion of $\mathbb{Q}$ with respect to the *p*-adic valuation $|x - y|_p$. The *p*-adic numbers are useful in solving DIOPHANTINE EQUATIONS. For example, the equation $X^2 = 2$ can easily be shown to have no solutions in the field of 2-adic numbers (we simply take the valuation of both sides). Because the 2-adic numbers contain the rationals as a subset, we can immediately see that the equation has no solutions in the RATIONALS. So we have an immediate proof of the irrationality of $\sqrt{2}$.

This is a common argument that is used in solving these types of equations: in order to show that an equation has no solutions in $\mathbb{Q}$, we show that it has no solutions in a FIELD EXTENSION. For another example, consider $X^2+1=0$. This equation has no solutions in $\mathbb{Q}$ because it has no solutions in the reals $\mathbb{R}$, and $\mathbb{Q}$ is a subset of $\mathbb{R}$.

Now consider the converse. Suppose we have an equation that does have solutions in $\mathbb{R}$ and in all the $\mathbb{Q}_p$. Can we conclude that the equation has a solution in $\mathbb{Q}$? Unfortunately, in general, the answer is no, but there are classes of equations for which the answer is yes. Such equations are said to satisfy the HASSE PRINCIPLE.

see also AX-KOCHEN ISOMORPHISM THEOREM, DIOPHANTINE EQUATION, HARMONIC NUMBER, HASSE PRINCIPLE, LOCAL FIELD, LOCAL-GLOBAL PRINCIPLE, MAHLER-LECH THEOREM, PRODUCT FORMULA, VALUATION, VALUATION THEORY, VON STAUDT-CLAUSEN THEOREM

References
Burger, E. B. and Struppeck, T. "Does $\sum_{n=0}^{\infty}\frac{1}{n!}$ Really Converge? Infinite Series and p-adic Analysis." *Amer. Math. Monthly* **103**, 565–577, 1996.
Cassels, J. W. S. and Scott, J. W. *Local Fields.* Cambridge, England: Cambridge University Press, 1986.
Gouvêa, F. Q. *P-adic Numbers: An Introduction, 2nd ed.* New York: Springer-Verlag, 1997.
Koblitz, N. *P-adic Numbers, P-adic Analysis, and Zeta-Functions, 2nd ed.* New York: Springer-Verlag, 1984.
Mahler, K. *P-adic Numbers and Their Functions, 2nd ed.* Cambridge, England: Cambridge University Press, 1981.

P-Circle

see SPIEKER CIRCLE

p-Element

see SEMISIMPLE

p-Good Path

A LATTICE PATH from one point to another is p-good if it lies completely below the line

$$y=(p-1)x.$$

Hilton and Pederson (1991) show that the number of p-good paths from $(1,\ q-1)$ to $(k,\ n-k)$ under the condition $2\leq k\leq n-p+1\leq p(k-1)$ is

$$\binom{n-q}{k-1}-\sum_{j=1}^{\ell} {}_p d_{qj}\binom{n-pj}{k-j},$$

where $\binom{a}{b}$ is a BINOMIAL COEFFICIENT, and

$$\ell\equiv\left\lfloor\frac{n-k}{p-1}\right\rfloor,$$

where $\lfloor x\rfloor$ is the FLOOR FUNCTION.

see also CATALAN NUMBER, LATTICE PATH, SCHRÖDER NUMBER

References
Hilton, P. and Pederson, J. "Catalan Numbers, Their Generalization, and Their Uses." *Math. Intel.* **13**, 64–75, 1991.

p-Group

A FINITE GROUP of ORDER p^a for p a PRIME is called a p-group. Sylow proved that every GROUP of this form has a POWER-commutator representation on n generators defined by

$$a_i^p=\prod_{k=i+1}^{n} a_k^{\beta(i,k)} \qquad (1)$$

for $0\leq\beta(i,k)<p$, $1\leq i\leq n$ and

$$[a_j,a_i]=\prod_{k=j+1}^{n} a_k^{\beta(i,j,k)} \qquad (2)$$

for $0\leq\beta(i,j,k)<p$, $1\leq i<j\leq n$. If p is PRIME and $f(p)$ the number of GROUPS of order p^m, then

$$f(p)=p^{Am^2}, \qquad (3)$$

where

$$\lim_{m\to\infty} A=\tfrac{2}{27} \qquad (4)$$

(Higman 1960a,b).

see also FINITE GROUP

References
Higman, G. "Enumerating p-Groups. I. Inequalities." *Proc. London Math. Soc.* **10**, 24–30, 1960a.
Higman, G. "Enumerating p-Groups. II. Problems Whose Solution is PORC." *Proc. London Math. Soc.* **10**, 566–582, 1960b.

p′-Group

X is a p'-group if p does not divide the ORDER of X.

p-Layer

The p-layer of H, $L_{p'}(H)$ is the unique minimal NORMAL SUBGROUP of H which maps onto $E(H/O_{p'}(H))$.

see also B_p-THEOREM, $L_{p'}$-BALANCE THEOREM, SIGNALIZER FUNCTOR THEOREM

P-Polynomial

see HOMFLY POLYNOMIAL

P-Problem

A problem is assigned to the P (POLYNOMIAL time) class if the number of steps is bounded by a POLYNOMIAL.

see also COMPLEXITY THEORY, NP-COMPLETE PROBLEM, NP-HARD PROBLEM, NP-PROBLEM

References

Borwein, J. M. and Borwein, P. B. *Pi and the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, 1987.

Greenlaw, R.; Hoover, H. J.; and Ruzzo, W. L. *Limits to Parallel Computation: P-Completeness Theory.* Oxford, England: Oxford University Press, 1995.

p-Series

A shorthand name for a POWER SERIES with a NEGATIVE exponent, $\sum_{k=1}^{\infty} k^{-p}$, where $p > 0$.

see also POWER SERIES, RIEMANN ZETA FUNCTION

p-Signature

Diagonalize a form over the rationals to

$$\operatorname{diag}[p^a \cdot A, p^b \cdot B, \ldots],$$

where all the entries are INTEGERS and A, B, ... are RELATIVELY PRIME to p. Then the p-signature of the form (for $p \neq -1, 2$) is

$$p^a + p^b + \ldots + 4k \pmod 8,$$

where k is the number of ANTISQUARES. For $p = -1$, the p-signature is SYLVESTER'S SIGNATURE.

see also SIGNATURE (QUADRATIC FORM)

P-Symbol

A symbol employed in a formal PROPOSITIONAL CALCULUS.

References

Nidditch, P. H. *Propositional Calculus.* New York: Free Press of Glencoe, p. 1, 1962.

P-Value

The PROBABILITY that a variate would assume a value greater than or equal to the observed value strictly by chance: $P(z \geq z_{\text{observed}})$.

see also ALPHA VALUE, SIGNIFICANCE

Paasche's Index

The statistical INDEX

$$P_P \equiv \frac{\sum p_n q_n}{\sum p_0 q_n},$$

where p_n is the price per unit in period n and q_n is the quantity produced in period n.

see also INDEX

References

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, p. 65, 1962.

Packing

The placement of objects so that they touch in some specified manner, often inside a container with specified properties.

see also BOX-PACKING THEOREM, CIRCLE PACKING, GROEMER PACKING, HYPERSPHERE PACKING, KEPLER PROBLEM, KISSING NUMBER PACKING DENSITY, POLYHEDRON PACKING, SPACE-FILLING POLYHEDRON, SPHERE PACKING

References

Eppstein, D. "Covering and Packing." `http://www.ics.uci.edu/~eppstein/junkyard/cover.html`.

Packing Density

The fraction of a volume filled by a given collection of solids.

see also HYPERSPHERE PACKING, PACKING, SPHERE PACKING

Padé Approximant

Approximants derived by expanding a function as a ratio of two POWER SERIES and determining both the NUMERATOR and DENOMINATOR COEFFICIENTS. Padé approximations are usually superior to TAYLOR EXPANSIONS when functions contain POLES, because the use of RATIONAL FUNCTIONS allows them to be well-represented.

The Padé approximant $R_{L/0}$ corresponds to the MACLAURIN SERIES. When it exists, the $R_{L/M} \equiv [L/M]$ Padé approximant to any POWER SERIES

$$A(x) = \sum_{j=0}^{\infty} a_j x^j \tag{1}$$

is unique. If $A(x)$ is a TRANSCENDENTAL FUNCTION, then the terms are given by the TAYLOR SERIES about x_0

$$a_n = \frac{1}{n!} A^{(n)}(x_0). \tag{2}$$

The COEFFICIENTS are found by setting

$$A(x) - \frac{P_L(x)}{Q_M(x)} = 0 \tag{3}$$

and equating COEFFICIENTS. $Q_M(x)$ can be multiplied by an arbitrary constant which will rescale the other COEFFICIENTS, so an addition constraint can be applied. The conventional normalization is

$$Q_M(0) = 1. \tag{4}$$

Expanding (3) gives

$$P_L(x) = p_0 + p_1 x + \ldots + p_L x^L \tag{5}$$

$$Q_M(x) = 1 + q_1 x + \ldots + q_M x^M. \tag{6}$$

These give the set of equations

$$a_0 = p_0 \tag{7}$$

$$a_1 + a_0 q_1 = p_1 \tag{8}$$

$$a_2 + a_1 q_1 + a_0 q_2 = p_2 \tag{9}$$

$$\vdots$$

$$a_L + a_{L-1} q_1 + \ldots + a_0 q_L = p_L \tag{10}$$

$$a_{L+1} + a_L q_1 + \ldots + a_{L-M+1} q_M = 0 \tag{11}$$

$$\vdots$$

$$q_{L+M} + a_{L+M-1} q_1 + \ldots + a_L q_M = 0, \tag{12}$$

where $a_n = 0$ for $n < 0$ and $q_j = 0$ for $j > M$. Solving these directly gives

$$[L/M] = \frac{\begin{vmatrix} a_{L-m+1} & a_{L-m+2} & \cdots & a_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ a_L & a_{L+1} & \cdots & a_{L+M} \\ \sum_{j=M}^{L} a_{j-M} x^j & \sum_{j=M-1}^{L} a_{j-M+1} x^j & \cdots & \sum_{j=0}^{L} a_j x^j \end{vmatrix}}{\begin{vmatrix} a_{L-M+1} & a_{L-M+2} & \cdots & a_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ a_L & a_{L+1} & \cdots & a_{L+M} \\ x^M & x^{M-1} & \cdots & 1 \end{vmatrix}}, \tag{13}$$

where sums are replaced by a zero if the lower index exceeds the upper. Alternate forms are

$$\begin{aligned}[L/M] &= \sum_{j=0}^{L-M} a_j x^j + x^{L-M+1} \mathbf{w}_{L/M}^{\mathrm{T}} \mathsf{W}_{L/M}^{-1} \mathbf{w}_{L/M} \\ &= \sum_{j=0}^{L+n} a_j x^j + x^{L+n+1} \mathbf{w}_{(L+M)/M}^{\mathrm{T}} \mathsf{W}_{L/M}^{-1} \mathbf{w}_{(L+n)/M}\end{aligned}$$

for

$$\mathsf{W}_{L/M} = \begin{bmatrix} a_{L-M+1} - x a_{L-M+2} & \cdots & a_L - x a_{L+1} \\ \vdots & \ddots & \vdots \\ a_L - x a_{L+1} & \cdots & a_{L+M-1} - x a_{L+M} \end{bmatrix} \tag{14}$$

$$\mathbf{w}_{L/M} = \begin{bmatrix} a_{L-M+1} \\ a_{L-M+2} \\ \vdots \\ a_L \end{bmatrix}, \tag{15}$$

and $0 \leq n \leq M$.

The first few Padé approximants for e^x are

$$\begin{aligned}
\exp_{0/0}(x) &= 1 \\
\exp_{0/1}(x) &= \frac{1}{1-x} \\
\exp_{0/2}(x) &= \frac{2}{2-2x+x^2} \\
\exp_{0/3}(x) &= \frac{6}{6-6x+3x^2-x^3} \\
\exp_{1/0}(x) &= 1+x \\
\exp_{1/1}(x) &= \frac{2+x}{2-x} \\
\exp_{1/2}(x) &= \frac{6+2x}{6-4x+x^2} \\
\exp_{1/3}(x) &= \frac{24+6x}{24-18x+6x^2-x^3} \\
\exp_{2/0}(x) &= \frac{2+2x+x^2}{2} \\
\exp_{2/1}(x) &= \frac{6+4x+x^2}{6-2x} \\
\exp_{2/2}(x) &= \frac{12+6x+x^2}{12-6x+x^2} \\
\exp_{2/3}(x) &= \frac{60+24x+3x^2}{60-36x+9x^2-x^3} \\
\exp_{3/0}(x) &= \frac{6+6x+3x^2+x^3}{6} \\
\exp_{3/1}(x) &= \frac{24+18x+16x^2+x^3}{24-6x} \\
\exp_{3/2}(x) &= \frac{60+36x+9x^2+x^3}{60-24x+3x^2} \\
\exp_{3/3}(x) &= \frac{120+60x+12x^2+x^3}{120-60x+12x^2-x^3}.
\end{aligned}$$

Two-term identities include

$$\frac{P_{L+1}(x)}{Q_{M+1}(x)} - \frac{P'_L(x)}{Q'_M(x)} = \frac{{C_{(L+1)/(M+1)}}^2 x^{L+M+1}}{Q_{M+1}(x) Q'_M(x)} \tag{16}$$

$$\frac{P_{L+1}(x)}{Q_M(x)} - \frac{P'_L(x)}{Q'_M(x)} = \frac{C_{(L+1)/M} C_{(L+1)/(M+1)} x^{L+M+1}}{Q_M(x) Q'_M(x)} \tag{17}$$

$$\frac{P_L(x)}{Q_{M+1}(x)} - \frac{P'_L(x)}{Q'_M(x)} = \frac{C_{L/(M+1)} C_{(L+1)/(M+1)} x^{L+M+1}}{Q_M(x) Q'_M(x)} \tag{18}$$

$$\frac{P_L(x)}{Q_{M+1}(x)} - \frac{P'_{L+1}(x)}{Q'_M} = \frac{{C_{(L+1)/(M+1)}}^2 x^{L+M+2}}{Q_{M+1} Q'_M} \tag{19}$$

$$\frac{P_{L+1}}{Q_M(x)} - \frac{P'_{L-1}(x)}{Q'_M(x)} = \frac{C_{L/(M+1)} C_{(L+1)/M} x^{L+M} + C_{L/M} C_{(L+1)/(M+1)} x^{L+M+1}}{Q_M(x) Q'_M(x)} \tag{20}$$

$$\frac{P_L(x)}{Q_{M+1}(x)} - \frac{P'_L(x)}{Q'_{M-1}(x)} = \frac{C_{L/(M+1)}C_{(L+1)/M}x^{L+M} - C_{L/M}C_{(L+1)/(M+1)}x^{L+M+1}}{Q_{M+1}(x)Q'_{M-1}(x)}, \tag{21}$$

where C is the C-DETERMINANT. Three-term identities can be derived using the FROBENIUS TRIANGLE IDENTITIES (Baker 1975, p. 32).

A five-term identity is

$$S_{(L+1)/M}S_{(L-1)/M} - S_{L/(M+1)}S_{L/(M-1)} = {S_{L/M}}^2. \tag{22}$$

Cross ratio identities include

$$\frac{(R_{L/M} - R_{L/(M+1)})(R_{(L+1)/M} - R_{(L+1)/(M+1)})}{(R_{L/M} - R_{(L+1)/M})(R_{L/(M+1)} - R_{(L+1)/(M+1)})} = \frac{C_{L/(M+1)}C_{(L+2)/(M+1)}}{C_{(L+1)/M}C_{(L+1)/(M+2)}} \tag{23}$$

$$\frac{(R_{L/M} - R_{(L+1)/(M+1)})(R_{(L+1)/M} - R_{L/(M+1)})}{(R_{L/M} - R_{L/(M+1)})(R_{(L+1)/M} - R_{(L+1)/(M+1)})} = \frac{{C_{(L+1)/(M+1)}}^2 x}{C_{L/(M+1)}C_{(L+2)/(M+1)}} \tag{24}$$

$$\frac{(R_{L/M} - R_{(L+1)/(M+1)})(R_{(L+1)/M} - R_{L/(M+1)})}{(R_{L/M} - R_{(L+1)/M})(R_{L/(M+1)} - R_{(L+1)/(M+1)})} = \frac{{C_{(L+1)/(M+1)}}^2 x}{C_{(L+1)/M}C_{(L+1)/(M+2)}} \tag{25}$$

$$\frac{(R_{L/M} - R_{(L+1)/(M-1)})(R_{L/(M+1)} - R_{(L+1)/M})}{(R_{L/M} - R_{L/(M+1)})(R_{(L+1)/(M+1)} - R_{(L+1)/M})} = \frac{C_{(L+1)/M}C_{(L+1)/(M+1)}x}{C_{L/(M+1)}C_{(L+2)/M}} \tag{26}$$

$$\frac{(R_{L/M} - R_{(L-1)/(M+1)})(R_{(L+1)/M} - R_{L/(M+1)})}{(R_{L/M} - R_{(L+1)/M})(R_{(L-1)/(M+1)} - R_{L/(M+1)})} = \frac{C_{L/(M+1)}C_{(L+1)/(M+1)}x}{C_{(L+1)/M}C_{L/(M+2)}}. \tag{27}$$

see also C-DETERMINANT, ECONOMIZED RATIONAL APPROXIMATION, FROBENIUS TRIANGLE IDENTITIES

References

Baker, G. A. Jr. "The Theory and Application of The Pade Approximant Method." In *Advances in Theoretical Physics, Vol. 1* (Ed. K. A. Brueckner). New York: Academic Press, pp. 1–58, 1965.

Baker, G. A. Jr. *Essentials of Padé Approximants in Theoretical Physics.* New York: Academic Press, pp. 27–38, 1975.

Baker, G. A. Jr. and Graves-Morris, P. *Padé Approximants.* New York: Cambridge University Press, 1996.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Padé Approximants." §5.12 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 194–197, 1992.

Padé Conjecture

If $P(z)$ is a POWER series which is regular for $|z| \leq 1$ except for m POLES within this CIRCLE and except for $z = +1$, at which points the function is assumed continuous when only points $|z| \leq 1$ are considered, then at least a subsequence of the $[N, N]$ PADÉ APPROXIMANTS are uniformly bounded in the domain formed by removing the interiors of small circles with centers at these POLES and uniformly continuous at $z = +1$ for $|z| \leq 1$.

see also PADÉ APPROXIMANT

References

Baker, G. A. Jr. "The Padé Conjecture and Some Consequences." §II.D in *Advances in Theoretical Physics, Vol. 1* (Ed. K. A. Brueckner). New York: Academic Press, pp. 23–27, 1965.

Padovan Sequence

The INTEGER SEQUENCE defined by the RECURRENCE RELATION

$$P(n) = P(n-2) + P(n-3)$$

with the initial conditions $P(0) = P(1) = P(2) = 1$. The first few terms are 1, 1, 2, 2, 3, 4, 5, 7, 9, 12, ... (Sloane's A000931). The ratio $\lim_{n\to\infty} P(n)/P(n-1)$ is called the PLASTIC CONSTANT.

see also PERRIN SEQUENCE, PLASTIC CONSTANT

References

Sloane, N. J. A. Sequence A000931/M0284 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Stewart, I. "Tales of a Neglected Number." *Sci. Amer.* **274**, 102–103, June 1996.

Painlevé Property

Following the work of Fuchs in classifying first-order ORDINARY DIFFERENTIAL EQUATIONS, Painlevé studied second-order ODEs of the form

$$\frac{d^2y}{dx^2} = F(y', y, x),$$

where F is ANALYTIC in x and rational in y and y'. Painlevé found 50 types whose only movable SINGULARITIES are ordinary POLES. This characteristic is known as the Painlevé property. Six of the transcendents define new transcendents known as PAINLEVÉ TRANSCENDENTS, and the remaining 44 can be integrated in terms of classical transcendents, quadratures, or the PAINLEVÉ TRANSCENDENTS.

see also PAINLEVÉ TRANSCENDENTS

Painlevé Transcendents

$$y'' = 6y^2 + x \tag{1}$$

$$y'' = 2y^3 + xy + \alpha \tag{2}$$

$$y'' = \frac{y'^2}{y} - \frac{1}{xy'} + \alpha y^3 + \frac{\beta}{xy^2} + \frac{\gamma}{x} + \frac{\delta}{y}. \tag{3}$$

Transcendents 4–6 do not have known first integrals, but all transcendents have first integrals for special values of their parameters except (1). Painlevé found the above transcendents (1) to (3), and the rest were investigated by his students. The sixth transcendent was found by Gambier and contains the other five as limiting cases.

see also PAINLEVÉ PROPERTY

Pair

A SET of two numbers or objects linked in some way are said to be a pair. The pair a and b are usually denoted $(a,\ b)$. In certain circumstances, pairs are also called BROTHERS or TWINS.

see also AMICABLE PAIR, AUGMENTED AMICABLE PAIR, BROWN NUMBERS, FRIENDLY PAIR, HEXAD, HOMOGENEOUS NUMBERS, IMPULSE PAIR, IRREGULAR PAIR, LAX PAIR, LONG EXACT SEQUENCE OF A PAIR AXIOM, MONAD, ORDERED PAIR, PERKO PAIR, QUADRUPLET, QUASIAMICABLE PAIR, QUINTUPLET, REDUCED AMICABLE PAIR, SMITH BROTHERS, TRIAD, TRIPLET, TWIN PEAKS, TWIN PRIMES, TWINS, UNITARY AMICABLE PAIR, WILF-ZEILBERGER PAIR

Pair Sum

Given an AMICABLE PAIR (m, n), the quantity

$$\sigma(m) = \sigma(n) = s(m) + s(n) = m + n$$

is called the pair sum, where $\sigma(n)$ is the DIVISOR FUNCTION and $s(n)$ is the RESTRICTED DIVISOR FUNCTION.

see also AMICABLE PAIR

Paired t-Test

Given two paired sets X_i and Y_i of n measured values, the paired t-test determines if they differ from each other in a significant way. Let

$$\begin{aligned}\hat{X}_i &= (X_i - \bar{X}_i)\\ \hat{Y}_i &= (Y_i - \bar{Y}_i),\end{aligned}$$

then define t by

$$t = (\bar{X} - \bar{Y})\sqrt{\frac{n(n-1)}{\sum_{i=1}^{n}(\hat{X}_i - \hat{Y}_i)^2}}.$$

This statistic has $n-1$ DEGREES OF FREEDOM.

A table of STUDENT'S t-DISTRIBUTION confidence interval can be used to determine the significance level at which two distributions differ.

see also FISHER SIGN TEST, HYPOTHESIS TESTING, STUDENT'S t-DISTRIBUTION, WILCOXON SIGNED RANK TEST

References

Goulden, C. H. *Methods of Statistical Analysis, 2nd ed.* New York: Wiley, pp. 50–55, 1956.

Paley Class

The Paley class of a POSITIVE INTEGER $m \equiv 0 \pmod 4$ is defined as the set of all possible QUADRUPLES (k, e, q, n) where

$$m = 2^e(q^n + 1),$$

q is an ODD PRIME, and

$$k = \begin{cases} 0 & \text{if } q = 0\\ 1 & \text{if } q^n - 3 \equiv 0 \pmod 4\\ 2 & \text{if } q^n - 1 \equiv 0 \pmod 4\\ \text{undefined} & \text{otherwise.}\end{cases}$$

see also HADAMARD MATRIX, PALEY CONSTRUCTION

Paley Construction

HADAMARD MATRICES H_n can be constructed using GALOIS FIELD GF(p^m) when $p = 4l - 1$ and m is ODD. Pick a representation r RELATIVELY PRIME to p. Then by coloring white $\lfloor (p-1)/2 \rfloor$ (where $\lfloor x \rfloor$ is the FLOOR FUNCTION) distinct equally spaced RESIDUES mod p (r^0, r, r^2, …; r^0, r^2, r^4, …; etc.) *in addition to* 0, a HADAMARD MATRIX is obtained if the POWERS of r (mod p) run through $< \lfloor (p-1)/2 \rfloor$. For example,

$$n = 12 = 11^1 + 1 = 2(5+1) = 2^2(2+1)$$

is of this form with $p = 11 = 4 \times 3 - 1$ and $m = 1$. Since $m = 1$, we are dealing with GF(11), so pick $p = 2$ and compute its RESIDUES (mod 11), which are

$$\begin{aligned} p^0 &\equiv 1\\ p^1 &\equiv 2\\ p^2 &\equiv 4\\ p^3 &\equiv 8\\ p^4 &\equiv 16 \equiv 5\\ p^5 &\equiv 10\\ p^6 &\equiv 20 \equiv 9\\ p^7 &\equiv 18 \equiv 7\\ p^8 &\equiv 14 \equiv 3\\ p^9 &\equiv 6\\ p^{10} &\equiv 12 \equiv 1.\end{aligned}$$

Picking the first $\lfloor 11/2 \rfloor = 5$ RESIDUES and adding 0 gives: 0, 1, 2, 4, 5, 8, which should then be colored in the MATRIX obtained by writing out the RESIDUES increasing to the left and up along the border (0 through $p-1$, followed by ∞), then adding horizontal and vertical coordinates to get the residue to place in each square.

$$\begin{bmatrix} \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty \\ 10 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & \infty \\ 9 & 10 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & \infty \\ 8 & 9 & 10 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & \infty \\ 7 & 8 & 9 & 10 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & \infty \\ 6 & 7 & 8 & 9 & 10 & 0 & 1 & 2 & 3 & 4 & 5 & \infty \\ 5 & 6 & 7 & 8 & 9 & 10 & 0 & 1 & 2 & 3 & 4 & \infty \\ 4 & 5 & 6 & 7 & 8 & 9 & 10 & 0 & 1 & 2 & 3 & \infty \\ 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 0 & 1 & 2 & \infty \\ 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 0 & 1 & \infty \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 0 & \infty \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & \infty \end{bmatrix}$$

H_{16} can be trivially constructed from $\mathsf{H}_4 \otimes \mathsf{H}_4$. H_{20} cannot be built up from smaller MATRICES, so use $n = 20 = 19 + 1 = 2(3^2 + 1) = 2^2(2^2 + 1)$. Only the first form can be used, with $p = 19 = 4 \times 5 - 1$ and $m = 1$. We therefore use GF(19), and color 9 RESIDUES plus 0 white. H_{24} can be constructed from $\mathsf{H}_2 \otimes \mathsf{H}_{12}$.

Now consider a more complicated case. For $n = 28 = 3^3 + 1 = 2(13+1)$, the only form having $p = 4l - 1$ is the first, so use the GF(3^3) field. Take as the modulus the IRREDUCIBLE POLYNOMIAL $x^3 + 2x + 1$, written 1021. A four-digit number can always be written using only three digits, since $1000 - 1021 \equiv 0012$ and $2000 - 2012 \equiv 0021$. Now look at the moduli starting with 10, where each digit is considered separately. Then

$$\begin{array}{lll} x^0 \equiv 1 & x^1 \equiv 10 & x^2 \equiv 100 \\ x^3 \equiv 1000 \equiv 12 & x^4 \equiv 120 & x^5 \equiv 1200 \equiv 212 \\ x^6 \equiv 2120 \equiv 111 & x^7 \equiv 1100 \equiv 122 & x^8 \equiv 1220 \equiv 202 \\ x^9 \equiv 2020 \equiv 11 & x^{10} \equiv 110 & x^{11} \equiv 1100 \equiv 112 \\ x^{12} \equiv 1120 \equiv 102 & x^{13} \equiv 1020 \equiv 2 & x^{14} \equiv 20 \\ x^{15} \equiv 200 & x^{16} \equiv 2000 \equiv 21 & x^{17} \equiv 210 \\ x^{18} \equiv 2100 \equiv 121 & x^{19} \equiv 1210 \equiv 222 & x^{20} \equiv 2220 \equiv 211 \\ x^{21} \equiv 2110 \equiv 101 & x^{22} \equiv 101 \equiv 22 & x^{23} \equiv 220 \\ x^{24} \equiv 2200 \equiv 221 & x^{25} \equiv 2210 \equiv 201 & x^{26} \equiv 2010 \equiv 1 \end{array}$$

Taking the alternate terms gives white squares as 000, 001, 020, 021, 022, 100, 102, 110, 111, 120, 121, 202, 211, and 221.

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 107–109 and 274, 1987.

Beth, T.; Jungnickel, D.; and Lenz, H. *Design Theory, 2nd ed. rev.* Cambridge, England: Cambridge University Press, 1998.

Geramita, A. V. *Orthogonal Designs: Quadratic Forms and Hadamard Matrices.* New York: Marcel Dekker, 1979.

Kitis, L. "Paley's Construction of Hadamard Matrices." `http://www.mathsource.com/cgi-bin/MathSource/Applications/Mathematics/0205-760`.

Paley's Theorem

Proved in 1933. If q is an ODD PRIME or $q = 0$ and n is any POSITIVE INTEGER, then there is a HADAMARD MATRIX of order

$$m = 2^e(q^n + 1),$$

where e is any POSITIVE INTEGER such that $m \equiv 0 \pmod 4$. If m is of this form, the matrix can be constructed with a PALEY CONSTRUCTION. If m is divisible by 4 but not of the form (1), the PALEY CLASS is undefined. However, HADAMARD MATRICES have been shown to exist for all $m \equiv 0 \pmod 4$ for $m < 428$.

see also HADAMARD MATRIX, PALEY CLASS, PALEY CONSTRUCTION

Palindrome Number

see PALINDROMIC NUMBER

Palindromic Number

A symmetrical number which is written in some base b as $a_1\,a_2\,\ldots\,a_2\,a_1$. The first few are 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 22, 33, 44, 55, 66, 77, 88, 99, 101, 111, 121, ... (Sloane's A002113).

The first few n for which the PRONIC NUMBER P_n is palindromic are 1, 2, 16, 77, 538, 1621, ... (Sloane's A028336), and the first few palindromic numbers which are PRONIC are 2, 6, 272, 6006, 289982, ... (Sloane's A028337). The first few numbers whose squares are palindromic are 1, 2, 3, 11, 22, 26, ... (Sloane's A002778), and the first few palindromic squares are 1, 4, 9, 121, 484, 676, ... (Sloane's A002779).

see also DEMLO NUMBER, PALINDROMIC NUMBER CONJECTURE, REVERSAL

References

de Geest, P. "Palindromic Products of Two Consecutive Integers." `http://www.ping.be/~ping6758/consec.htm`.

de Geest, P. "Palindromic Squares." `http://www.ping.be/~ping6758/square.htm`.

Pappas, T. "Numerical Palindromes." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 146, 1989.

Sloane, N. J. A. Sequences A028336, A028337, A002113/M0484, A0027778/M0807, and A002779/M3371 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Palindromic Number Conjecture

Apply the 196-ALGORITHM, which consists of taking any POSITIVE INTEGER of two digits or more, reversing the digits, and adding to the original number. Now sum the two and repeat the procedure with the sum. Of the first 10,000 numbers, only 251 do not produce a PALINDROMIC NUMBER in ≤ 23 steps (Gardner 1979).

It was therefore conjectured that *all* numbers will eventually yield a PALINDROMIC NUMBER. However, the conjecture has been proven false for bases which are a POWER of 2, and seems to be false for base 10 as well. Among the first 100,000 numbers, 5,996 numbers apparently never generate a PALINDROMIC NUMBER (Gruenberger 1984). The first few are 196, 887, 1675, 7436, 13783, 52514, 94039, 187088, 1067869, 10755470, ... (Sloane's A006960).

It is conjectured, but not proven, that there are an infinite number of palindromic PRIMES. With the exception of 11, palindromic PRIMES must have an ODD number of digits.

see also 196-ALGORITHM

References
Gardner, M. *Mathematical Circus: More Puzzles, Games, Paradoxes and Other Mathematical Entertainments from Scientific American.* New York: Knopf, pp. 242–245, 1979.
Gruenberger, F. "How to Handle Numbers with Thousands of Digits, and Why One Might Want to." *Sci. Amer.* **250**, 19–26, Apr. 1984.
Sloane, N. J. A. Sequence A006960/M5410 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Pancake Cutting

see CIRCLE CUTTING

Pancake Theorem

The 2-D version of the HAM SANDWICH THEOREM.

Pandiagonal Square

see PANMAGIC SQUARE

Pandigital

A decimal INTEGER which contains each of the digits from 0 to 9.

Panmagic Square

8	17	1	15	24
11	25	9	18	2
19	3	12	21	10
22	6	20	4	13
5	14	23	7	16

If all the diagonals (including those obtained by "wrapping around" the edges) of a MAGIC SQUARE, as well as the usual rows, columns, and *main diagonals* sum to the MAGIC CONSTANT, the square is said to be a PANMAGIC SQUARE (also called DIABOLICAL SQUARE, NASIK SQUARE, or PANDIAGONAL SQUARE). No panmagic squares exist of order 3 or any order $4k+2$ for k an INTEGER. The Siamese method for generating MAGIC SQUARES produces panmagic squares for orders $6k \pm 1$ with ordinary vector (2, 1) and break vector (1, -1).

1	15	24	8	17
23	7	16	5	14
20	4	13	22	6
12	21	10	19	3
9	18	2	11	25

The LO SHU is not panmagic, but it is an ASSOCIATIVE MAGIC SQUARE. Order four squares can be panmagic or ASSOCIATIVE, but not both. Order five squares are the smallest which can be both ASSOCIATIVE and panmagic, and 16 distinct ASSOCIATIVE panmagic squares exist, one of which is illustrated above (Gardner 1988).

The number of distinct panmagic squares of order 1, 2, ... are 1, 0, 0, 384, 3600, 0, ... (Sloane's A027567, Hunter and Madachy 1975). Panmagic squares are related to HYPERCUBES.

see also ASSOCIATIVE MAGIC SQUARE, HYPERCUBE, FRANKLIN MAGIC SQUARE, MAGIC SQUARE

References
Gardner, M. *The Second Scientific American Book of Mathematical Puzzles & Diversions: A New Selection.* New York: Simon and Schuster, pp. 135–137, 1961.
Gardner, M. "Magic Squares and Cubes." Ch. 17 in *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, pp. 213–225, 1988.
Hunter, J. A. H. and Madachy, J. S. "Mystic Arrays." Ch. 3 in *Mathematical Diversions.* New York: Dover, pp. 24–25, 1975.
Kraitchik, M. "Panmagic Squares." §7.9 in *Mathematical Recreations.* New York: W. W. Norton, pp. 174–176, 1942.
Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, p. 87, 1979.
Rosser, J. B. and Walker, R. J. "The Algebraic Theory of Diabolical Squares." *Duke Math. J.* **5**, 705–728, 1939.
Sloane, N. J. A. Sequence A027567 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Pantograph

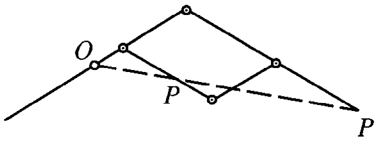

A LINKAGE invented in 1630 by Christoph Scheiner for making a scaled copy of a given figure. The linkage is pivoted at O; hinges are denoted ⊙. By placing a PENCIL at P (or P'), a DILATED image is obtained at P' (or P).

see also LINKAGE

Papal Cross

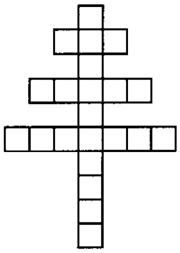

see also CROSS

Pappus's Centroid Theorem

The SURFACE AREA of a SURFACE OF REVOLUTION is given by

$$S_{\text{solid of rotation}} = [\text{perimenter}] \times [\text{distance traveled by centroid}],$$

and the VOLUME of a SOLID OF REVOLUTION is given by

$$V_{\text{solid of rotation}} = [\text{cross-section area}] \times [\text{distance traveled by centroid}].$$

see also CENTROID (GEOMETRIC), CROSS-SECTION, PERIMETER, SOLID OF REVOLUTION, SURFACE AREA, SURFACE OF REVOLUTION, TOROID, TORUS

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 132, 1987.

Pappus Chain

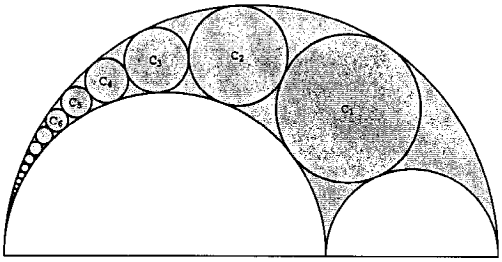

In the ARBELOS, construct a chain of TANGENT CIRCLES starting with the CIRCLE TANGENT to the two small interior semicircles and the large exterior one. Then the distance from the center of the first INSCRIBED CIRCLE to the bottom line is twice the CIRCLE'S RADIUS, from the second CIRCLE is four times the RADIUS, and for the nth CIRCLE is $2n$ times the RADIUS. The centers of the CIRCLES lie on an ELLIPSE, and the DIAMETER of the nth CIRCLE C_n is $(1/n)$th PERPENDICULAR distance to the base of the SEMICIRCLE. This result was known to Pappus, who referred to it as an ancient theorem (Hood 1961, Cadwell 1966, Gardner 1979, Bankoff 1981). The simplest proof is via INVERSIVE GEOMETRY.

If $r \equiv AB/AC$, then the radius of the nth circle in the pappus chain is

$$r_n = \frac{(1-r)r}{2[n^2(1-r)^2+r]}.$$

This equation can be derived by iteratively solving the QUADRATIC FORMULA generated by DESCARTES CIRCLE THEOREM for the radius of the SODDY CIRCLE. This general result simplifies to $r_n = 1/(6+n^2)$ for $r = 2/3$ (Gardner 1979). Further special cases when $AC = 1 + AB$ are considered by Gaba (1940).

If B divides AC in the GOLDEN RATIO ϕ, then the circles in the chain satisfy a number of other special properties (Bankoff 1955).

see also ARBELOS, COXETER'S LOXODROMIC SEQUENCE OF TANGENT CIRCLES, SODDY CIRCLES, STEINER CHAIN

References

Bankoff, L. "The Golden Arbelos." *Scripta Math.* **21**, 70–76, 1955.

Bankoff, L. "Are the Twin Circles of Archimedes Really Twins?" *Math. Mag.* **47**, 214–218, 1974.

Bankoff, L. "How Did Pappus Do It?" In *The Mathematical Gardner* (Ed. D. Klarner). Boston, MA: Prindle, Weber, and Schmidt, pp. 112–118, 1981.

Gaba, M. G. "On a Generalization of the Arbelos." *Amer. Math. Monthly* **47**, 19–24, 1940.

Gardner, M. "Mathematical Games: The Diverse Pleasures of Circles that Are Tangent to One Another." *Sci. Amer.* **240**, 18–28, Jan. 1979.

Hood, R. T. "A Chain of Circles." *Math. Teacher* **54**, 134–137, 1961.

Johnson, R. A. *Advanced Euclidean Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, p. 117, 1929.

Pappus-Guldinus Theorem

see PAPPUS'S CENTROID THEOREM

Pappus's Harmonic Theorem

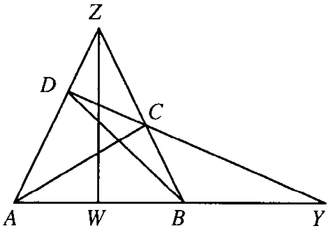

AW, AB, and AY in the above figure are in a HARMONIC RANGE.

see also CEVA'S THEOREM, MENELAUS' THEOREM

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 67–68, 1967.

Pappus's Hexagon Theorem

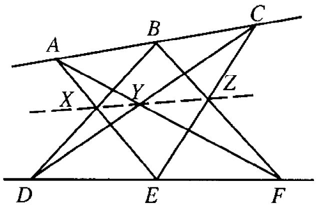

If A, B, and C are three points on one LINE, D, E, and F are three points on another LINE, and AE meets BD at X, AF meets CD at Y, and BF meets CE at Z, then the three points X, Y, and Z are COLLINEAR. Pappus's hexagon theorem is essentially its own dual according to the DUALITY PRINCIPLE of PROJECTIVE GEOMETRY.

see also CAYLEY-BACHARACH THEOREM, DESARGUES' THEOREM, DUALITY PRINCIPLE, PASCAL'S THEOREM, PROJECTIVE GEOMETRY

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 73–74, 1967.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 92–94, 1990.

Pappas, T. "Pappus' Theorem & the Nine Coin Puzzle." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./ Tetra, p. 163, 1989.

Pappus's Theorem

There are several THEOREMS that generally are known by the generic name "Pappus's Theorem."

see also PAPPUS'S CENTROID THEOREM, PAPPUS CHAIN, PAPPUS'S HARMONIC THEOREM, PAPPUS'S HEXAGON THEOREM

Parabiaugmented Dodecahedron

see JOHNSON SOLID

Parabiaugmented Hexagonal Prism

see JOHNSON SOLID

Parabiaugmented Truncated Dodecahedron

see JOHNSON SOLID

Parabidiminished Rhombicosidodecahedron

see JOHNSON SOLID

Parabigyrate Rhombicosidodecahedron

see JOHNSON SOLID

Parabola

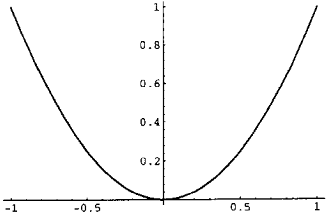

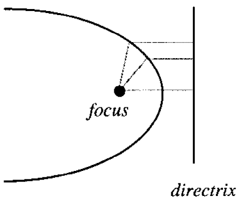

The set of all points in the PLANE equidistant from a given LINE (the DIRECTRIX) and a given point not on the line (the FOCUS).

The parabola was studied by Menaechmus in an attempt to achieve CUBE DUPLICATION. Menaechmus solved the problem by finding the intersection of the two parabolas $x^2 = y$ and $y^2 = 2x$. Euclid wrote about the parabola, and it was given its present name by Apollonius. Pascal considered the parabola as a projection of a CIRCLE, and Galileo showed that projectiles falling under uniform gravity follow parabolic paths. Gregory and Newton considered the CATACAUSTIC properties of a parabola which bring parallel rays of light to a focus (MacTutor Archive).

For a parabola opening to the right, the equation in CARTESIAN COORDINATES is

$$\sqrt{(x-p)^2+y^2} = x+p \tag{1}$$

$$(x-p)^2+y^2 = (x+p)^2 \tag{2}$$

$$x^2-2px+p^2+y^2 = x^2+2px+p^2 \tag{3}$$

$$y^2 = 4px. \tag{4}$$

If the VERTEX is at (x_0, y_0) instead of (0, 0), the equation is

$$(y-y_0)^2 = 4p(x-x_0). \tag{5}$$

If the parabola opens upwards,

$$x^2 = 4py \tag{6}$$

(which is the form shown in the above figure at left). The quantity $4p$ is known as the LATUS RECTUM. In POLAR COORDINATES,

$$r = \frac{2a}{1-\cos\theta}. \tag{7}$$

In PEDAL COORDINATES with the PEDAL POINT at the FOCUS, the equation is

$$p^2 = ar. \tag{8}$$

The parametric equations for the parabola are

$$x = 2at \tag{9}$$

$$y = at^2. \tag{10}$$

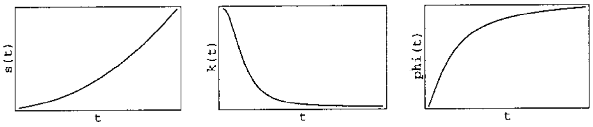

The CURVATURE, ARC LENGTH, and TANGENTIAL ANGLE are

$$\kappa(t) = \frac{1}{2(1+t^2)^{3/2}} \tag{11}$$

$$s(t) = t\sqrt{1+t^2} + \sinh^{-1} t \tag{12}$$

$$\phi(t) = \tan^{-1} t. \tag{13}$$

The TANGENT VECTOR of the parabola is

$$x_T(t) = \frac{1}{\sqrt{1+t^2}} \tag{14}$$

$$y_T(t) = \frac{t}{\sqrt{1+t^2}}. \tag{15}$$

The plots below show the normal and tangent vectors to a parabola.

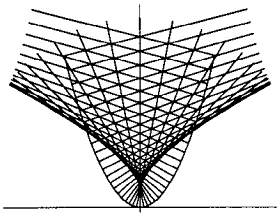

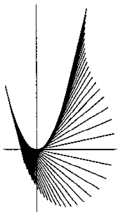

see also CONIC SECTION, ELLIPSE, HYPERBOLA, QUADRATIC CURVE, REFLECTION PROPERTY, TSCHIRNHAUSEN CUBIC PEDAL CURVE

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 198, 1987.

Casey, J. "The Parabola." Ch. 5 in *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions, with Numerous Examples, 2nd ed., rev. enl.* Dublin: Hodges, Figgis, & Co., pp. 173–200, 1893.

Coxeter, H. S. M. "Conics." §8.4 in *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 115–119, 1969.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 67–72, 1972.

Lee, X. "Parabola." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Parabola_dir/parabola.html`.

Lockwood, E. H. "The Parabola." Ch. 1 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 2–12, 1967.

MacTutor History of Mathematics Archive. "Parabola." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Parabola.html`.

Pappas, T. "The Parabolic Ceiling of the Capitol." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 22–23, 1989.

Parabola Caustic

The CAUSTIC of a PARABOLA with rays PERPENDICULAR to the axis of the PARABOLA is TSCHIRNHAUSEN CUBIC.

Parabola Evolute

Given a PARABOLA

$$y = x^2, \tag{1}$$

the parametric equation and its derivatives are

$$\begin{array}{lll} x = t & x' = t & y' = 2t \\ y = t^2 & x'' = 0 & y'' = 2. \end{array} \tag{2}$$

The RADIUS OF CURVATURE is

$$R = \frac{(x'^2 + y'^2)^{3/2}}{x'y'' - x''y'} = \frac{(1+4t^2)^{3/2}}{2}. \tag{3}$$

The TANGENT VECTOR is

$$\hat{\mathbf{T}} = \frac{1}{\sqrt{1+4t^2}} \begin{bmatrix} 1 \\ 2t \end{bmatrix}, \tag{4}$$

so the parametric equations of the evolute are

$$\xi = -4t^3 \tag{5}$$

$$\eta = \tfrac{1}{2} + 3t^2, \tag{6}$$

and

$$-\tfrac{1}{4}\xi = t^3 \tag{7}$$

$$\tfrac{1}{3}(\eta - \tfrac{1}{2}) = t^2 \tag{8}$$

$$\tfrac{1}{3}(\eta - \tfrac{1}{2}) = (-\tfrac{1}{4}\xi)^{2/3} \tag{9}$$

$$\tfrac{1}{3}(\eta - h) = \left(-\frac{2\xi}{8}\right)^{2/3} = \tfrac{1}{4}(2\xi)^{2/3}. \tag{10}$$

The EVOLUTE is therefore

$$\eta = \tfrac{3}{4}(2\xi)^{2/3} + \tfrac{1}{2}. \tag{11}$$

This is known as NEILE'S PARABOLA and is a SEMICUBICAL PARABOLA. From a point above the evolute three normals can be drawn to the PARABOLA, while only one normal can be drawn to the PARABOLA from a point below the EVOLUTE.

see also NEILE'S PARABOLA, PARABOLA, SEMICUBICAL PARABOLA

Parabola Inverse Curve

The INVERSE CURVE for a PARABOLA given by

$$x = at^2 \tag{1}$$

$$y = 2at \tag{2}$$

with INVERSION CENTER (x_0, y_0) and INVERSION RADIUS k is

$$x = x_0 + \frac{k(at^2 - x_0)}{(at^2 + x_0)^2 + (2at - y_0)^2} \tag{3}$$

$$y = y_0 + \frac{k(2at - y_0)}{(at^2 + x_0)^2 + (2at - y_0)^2}. \tag{4}$$

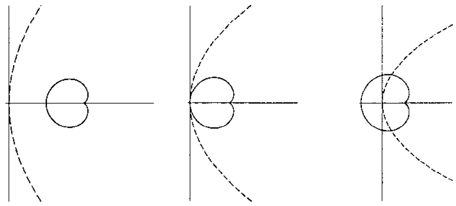

For $(x_0, y_0) = (a, 0)$ at the FOCUS, the INVERSE CURVE is the CARDIOID

$$x = a + \frac{k(t^2 - 1)}{a(1+t^2)^2} \tag{5}$$

$$y = \frac{2kt}{a(1+t^2)^2}. \tag{6}$$

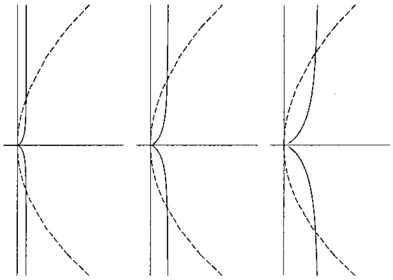

For $(x_0, y_0) = (0, 0)$ at the VERTEX, the INVERSE CURVE is the CISSOID OF DIOCLES

$$x = \frac{k}{a(4+t^2)} \tag{7}$$

$$y = \frac{2k}{at(4+t^2)}. \tag{8}$$

Parabola Involute

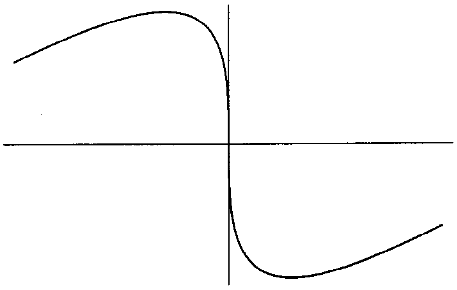

$$\frac{d\mathbf{r}}{dt} = \begin{bmatrix} 1 \\ 2t \end{bmatrix} \tag{1}$$

$$\hat{\mathbf{T}} = \frac{1}{\sqrt{1+4t^2}} \begin{bmatrix} 1 \\ 2t \end{bmatrix} \tag{2}$$

$$ds^2 = |d\mathbf{r}|^2 = (1+4t^2)\,dt^2 \tag{3}$$

$$ds = \sqrt{1+4t^2}\,dt \tag{4}$$

$$s = \int \sqrt{1+4t^2}\,dt = \tfrac{1}{2}t\sqrt{1+4t^2} + \tfrac{1}{4}\sinh^{-1}(2t). \tag{5}$$

So the equation of the INVOLUTE is

$$\mathbf{r}_i = \mathbf{r} - s\hat{\mathbf{T}} = \begin{bmatrix} t \\ t^2 \end{bmatrix} - \frac{\frac{1}{2}t\sqrt{1+4t^2} + \frac{1}{4}\sinh^{-1}(2t)}{\sqrt{1+4t^2}} \begin{bmatrix} 1 \\ 2t \end{bmatrix}$$

$$= \frac{1}{2\sqrt{1+4t^2}} \begin{bmatrix} t - \frac{1}{2}\sinh^{-1}(2t) \\ -\sinh^{-1}(2t) \end{bmatrix}. \tag{6}$$

Parabola Pedal Curve

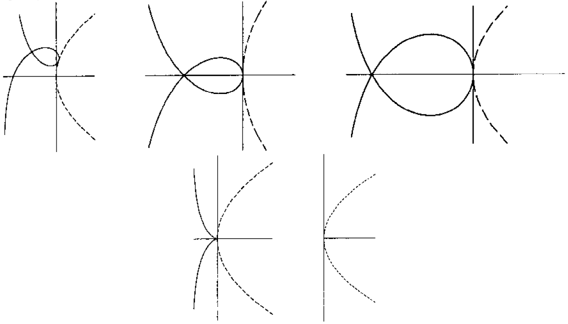

On the DIRECTRIX, the PEDAL CURVE of a PARABOLA is a STROPHOID (top left). On the foot of the DIRECTRIX, it is a RIGHT STROPHOID (top middle). On reflection of the FOCUS in the DIRECTRIX, it is a MACLAURIN TRISECTRIX (top right). On the VERTEX, it is a CISSOID OF DIOCLES (bottom left). On the FOCUS, it is a straight line (bottom right).

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 94–97, 1972.

Parabolic Coordinates

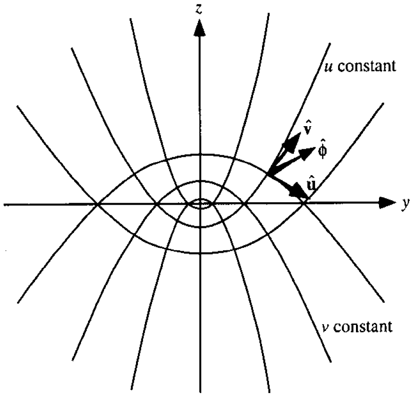

A system of CURVILINEAR COORDINATES in which two sets of coordinate surfaces are obtained by revolving the parabolas of PARABOLIC CYLINDRICAL COORDINATES about the x-AXIS, which is then relabeled the z-AXIS. There are several notational conventions. Whereas (u, v, θ) is used in this work, Arfken (1970) uses (ξ, η, φ).

The equations for the parabolic coordinates are

$$x = uv\cos\theta \tag{1}$$

$$y = uv\sin\theta \tag{2}$$

$$z = \tfrac{1}{2}(u^2 - v^2), \tag{3}$$

where $u \in [0, \infty)$, $v \in [0, \infty)$, and $\theta \in [0, 2\pi)$. To solve for u, v, and θ, examine

$$\begin{aligned} x^2 + y^2 + z^2 &= u^2v^2 + \tfrac{1}{4}(u^4 - 2u^2v^2 + v^4) \\ &= \tfrac{1}{4}(u^4 + 2u^2v^2 + v^4) \\ &= \tfrac{1}{4}(u^2 + v^2)^2, \end{aligned} \tag{4}$$

so

$$\sqrt{x^2 + y^2 + z^2} = \tfrac{1}{2}(u^2 + v^2) \tag{5}$$

and

$$\sqrt{x^2 + y^2 + z^2} + z = u^2 \tag{6}$$

$$\sqrt{x^2 + y^2 + z^2} - z = v^2. \tag{7}$$

We therefore have

$$u = \sqrt{\sqrt{x^2 + y^2 + z^2} + z} \tag{8}$$

$$v = \sqrt{\sqrt{x^2 + y^2 + z^2} - z} \tag{9}$$

$$\theta = \tan^{-1}\left(\frac{y}{x}\right). \tag{10}$$

The SCALE FACTORS are

$$h_u = \sqrt{u^2 + v^2} \tag{11}$$

$$h_v = \sqrt{u^2 + v^2} \tag{12}$$

$$h_\theta = uv. \tag{13}$$

The LINE ELEMENT is

$$ds^2 = (u^2 + v^2)(du^2 + dv^2) + u^2v^2\,d\theta^2, \tag{14}$$

and the VOLUME ELEMENT is

$$dV = uv(u^2 + v^2)\,du\,dv\,d\theta. \tag{15}$$

The LAPLACIAN is

$$\begin{aligned}\nabla^2 f &= \frac{1}{uv(u^2+v^2)}\left[\frac{\partial}{\partial u}\left(uv\frac{\partial f}{\partial u}\right) + \frac{\partial}{\partial v}\left(uv\frac{\partial f}{\partial v}\right)\right] \\ &\qquad + \frac{1}{u^2v^2}\frac{\partial^2 f}{\partial \theta^2} \\ &= \frac{1}{u^2+v^2}\left[\frac{1}{u}\frac{\partial}{\partial u}\left(u\frac{\partial f}{\partial u}\right) + \frac{1}{v}\frac{\partial}{\partial v}\left(v\frac{\partial f}{\partial v}\right)\right] + \frac{1}{u^2v^2}\frac{\partial^2 f}{\partial \theta^2} \\ &= \frac{1}{u^2+v^2}\left(\frac{1}{u}\frac{\partial f}{\partial u} + \frac{\partial^2 f}{\partial u^2} + \frac{1}{v}\frac{\partial f}{\partial v} + \frac{\partial^2 f}{\partial v^2}\right) + \frac{1}{u^2v^2}\frac{\partial^2 f}{\partial \theta^2}.\end{aligned} \tag{16}$$

The HELMHOLTZ DIFFERENTIAL EQUATION is SEPARABLE in parabolic coordinates.

see also CONFOCAL PARABOLOIDAL COORDINATES, HELMHOLTZ DIFFERENTIAL EQUATION—PARABOLIC COORDINATES, PARABOLIC CYLINDRICAL COORDINATES

References

Arfken, G. "Parabolic Coordinates (ξ, η, ϕ)." §2.12 in *Mathematical Methods for Physicists, 2nd ed.* Orlando, FL: Academic Press, pp. 109–112, 1970.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 660, 1953.

Parabolic Cyclide

A CYCLIDE formed by inversion of a STANDARD TORUS when the sphere of inversion is tangent to the torus.

see also PARABOLIC HORN CYCLIDE, PARABOLIC RING CYCLIDE, PARABOLIC SPINDLE CYCLIDE

Parabolic Cylinder

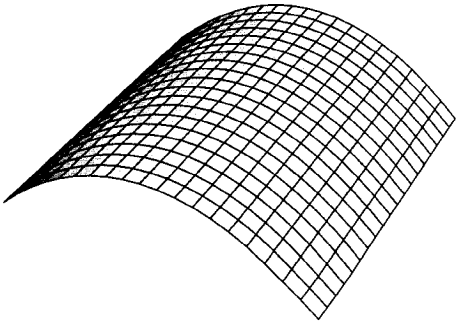

A QUADRATIC SURFACE given by the equation

$$x^2 + 2rz = 0.$$

Parabolic Cylinder Function

These functions are sometimes called WEBER FUNCTIONS. Whittaker and Watson (1990, p. 347) define the parabolic cylinder functions as solutions to the WEBER DIFFERENTIAL EQUATION

$$\frac{d^2D_n(z)}{dz^2} + (n + \tfrac{1}{2} - \tfrac{1}{4}z^2)D_n(z) = 0. \tag{1}$$

The two independent solutions are given by $D_n(z)$ and $D_{-n-1}(ze^{i\pi/2})$, where

$$\begin{aligned}D_n(z) &= 2^{n/2+1/4}z^{-1/2}W_{n/2+1/4,-1/4}(\tfrac{1}{2}z^2) \qquad (2)\\ &= \frac{\Gamma(\frac{1}{2})2^{n/2+1/4}z^{-1/2}}{\Gamma(\frac{1}{2}-\frac{1}{2}n)}M_{n/2+1/4,-1/4}(\tfrac{1}{2}z^2) \\ &\quad + \frac{\Gamma(-\frac{1}{2})2^{n/2+1/4}z^{-1/2}}{\Gamma(-\frac{1}{2}n)}M_{n/2+1/4,1/4}(\tfrac{1}{2}z^2). \qquad (3)\end{aligned}$$

Here, $W_{a,b}(z)$ is a WHITTAKER FUNCTION and $M_{a,b}(z) = {}_1F_1(a;b;z)$ are CONFLUENT HYPERGEOMETRIC FUNCTIONS.

Abramowitz and Stegun (1972, p. 686) define the parabolic cylinder functions as solutions to

$$y'' + (ax^2 + bx + c) = 0. \tag{4}$$

This can be rewritten by COMPLETING THE SQUARE,

$$y'' + \left[a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a} + c\right]y = 0. \tag{5}$$

Now letting

$$u = x + \frac{b}{2a} \tag{6}$$

$$du = dx \tag{7}$$

gives

$$\frac{d^2y}{du^2} + (au^2 + d)y = 0 \tag{8}$$

where

$$d \equiv \frac{b^2}{4a} + c. \tag{9}$$

Equation (4) has the two standard forms

$$y'' - (\tfrac{1}{4}x^2 + a)y = 0 \tag{10}$$

$$y'' + (\tfrac{1}{4}x^2 - a)y = 0. \tag{11}$$

For a general a, the EVEN and ODD solutions to (10) are

$$y_1(x) = e^{-x^2/4}\,{}_1F_1(\tfrac{1}{2}a + \tfrac{1}{4}; \tfrac{1}{2}; \tfrac{1}{2}x^2) \tag{12}$$

$$y_2(x) = xe^{-x^2/4}\,{}_1F_1(\tfrac{1}{2}a + \tfrac{3}{4}; \tfrac{3}{2}; \tfrac{1}{2}x^2), \tag{13}$$

where ${}_1F_1(a;b;z)$ is a CONFLUENT HYPERGEOMETRIC FUNCTION. If $y(a,x)$ is a solution to (10), then (11) has solutions

$$y(\pm ia, xe^{\mp i\pi/4}), y(\pm ia, -xe^{\mp i\pi/4}). \tag{14}$$

Abramowitz and Stegun (1972, p. 687) define standard solutions to (10) as

$$U(a,x) = \cos[\pi(\tfrac{1}{4}+\tfrac{1}{2}a)]Y_1 - \sin[\pi(\tfrac{1}{4}+\tfrac{1}{2}a)]Y_2 \tag{15}$$

$$V(a,x) = \frac{\sin[\pi(\tfrac{1}{4}+\tfrac{1}{2}a)]Y_1 + \cos[\pi(\tfrac{1}{4}+\tfrac{1}{2}a)]Y_2}{\Gamma(\tfrac{1}{2}-a)}, \tag{16}$$

where

$$\begin{aligned} Y_1 &\equiv \frac{1}{\sqrt{\pi}}\frac{\Gamma(\tfrac{1}{4}-\tfrac{1}{2}a)}{2^{a/2+1/4}}y_1 \\ &= \frac{1}{\sqrt{\pi}}\frac{\Gamma(\tfrac{1}{4}-\tfrac{1}{2}a)}{2^{a/2+1/4}}e^{-x^2/4}{}_1F_1(\tfrac{1}{2}a+\tfrac{1}{4};\tfrac{1}{2};\tfrac{1}{2}x^2) \end{aligned} \tag{17}$$

$$\begin{aligned} Y_2 &\equiv \frac{1}{\sqrt{\pi}}\frac{\Gamma(\tfrac{3}{4}-\tfrac{1}{2}a)}{2^{a/2+1/4}}y_2 \\ &= \frac{1}{\sqrt{\pi}}\frac{\Gamma(\tfrac{3}{4}-\tfrac{1}{2}a)}{2^{a/2+1/4}}xe^{-x^2/4}{}_1F_1(\tfrac{1}{2}a+\tfrac{3}{4};\tfrac{3}{2};\tfrac{1}{2}x^2). \end{aligned} \tag{18}$$

In terms of Whittaker and Watscn's functions,

$$U(a,x) = D_{-a-1/2}(x) \tag{19}$$

$$V(a,x) = \frac{\Gamma(\tfrac{1}{2}+a)[\sin(\pi a)D_{-a-1/2}(x) + D_{-a-1/2}(-x)]}{\pi}. \tag{20}$$

For NONNEGATIVE INTEGER n, the solution D_n reduces to

$$D_n(x) = 2^{-n/2}e^{-x^2/4}H_n\left(\frac{x}{\sqrt{2}}\right) = e^{-x^2/4}\mathrm{He}_n(x), \tag{21}$$

where $H_n(x)$ is a HERMITE POLYNOMIAL and He_n is a modified HERMITE POLYNOMIAL.

The parabolic cylinder functions D_ν satisfy the RECURRENCE RELATIONS

$$D_{\nu+1}(z) - zD_\nu(z) + \nu D_{\nu-1}(z) = 0 \tag{22}$$

$$D'_\nu(z) + \tfrac{1}{2}zD_\nu(z) - \nu D_{\nu-1}(z) = 0. \tag{23}$$

The parabolic cylinder function for integral n can be defined in terms of an integral by

$$D_n(z) = \frac{1}{\pi}\int_0^\pi \sin(n\theta - z\sin\theta)\,d\theta \tag{24}$$

(Watson 1966, p. 308), which is similar to the ANGER FUNCTION. The result

$$\int_{-\infty}^{\infty} D_m(x)D_n(x)\,dx = \delta_{mn}n!\sqrt{2\pi}, \tag{25}$$

where δ_{ij} is the KRONECKER DELTA, can also be used to determine the COEFFICIENTS in the expansion

$$f(z) = \sum_{n=0}^{\infty} a_nD_n \tag{26}$$

as

$$a_n = \frac{1}{n!\sqrt{2\pi}}\int_{-\infty}^{\infty} D_n(t)f(t)\,dt. \tag{27}$$

For ν real,

$$\int_0^\infty [D_\nu(t)]^2\,dt = \pi^{1/2}2^{-3/2}\frac{\phi_0(\tfrac{1}{2}-\tfrac{1}{2}\nu) - \phi_0(-\tfrac{1}{2}\nu)}{\Gamma(-\nu)} \tag{28}$$

(Gradshteyn and Ryzhik 1980, p. 885, 7.711.3), where $\Gamma(z)$ is the GAMMA FUNCTION and $\phi_0(z)$ is the POLYGAMMA FUNCTION of order 0.

see also ANGER FUNCTION, BESSEL FUNCTION, DARWIN'S EXPANSIONS, HH FUNCTION, STRUVE FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Parabolic Cylinder Function." Ch. 19 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 685–700, 1972.

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, 1979.

Iyanaga, S. and Kawada, Y. (Eds.). "Parabolic Cylinder Functions (Weber Functions)." Appendix A, Table 20.III in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1479, 1980.

Spanier, J. and Oldham, K. B. "The Parabolic Cylinder Function $D_\nu(x)$." Ch. 46 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 445–457, 1987.

Watson, G. N. *A Treatise on the Theory of Bessel Functions, 2nd ed.* Cambridge, England: Cambridge University Press, 1966.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Parabolic Cylindrical Coordinates

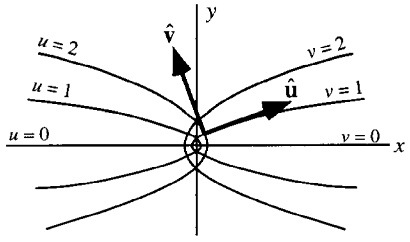

A system of CURVILINEAR COORDINATES. There are several different conventions for the orientation and designation of these coordinates. Arfken (1970) defines coordinates (ξ, η, z) such that

$$x = \xi\eta \quad (1)$$

$$y = \tfrac{1}{2}(\eta^2 - \xi^2) \quad (2)$$

$$z = z. \quad (3)$$

In this work, following Morse and Feshbach (1953), the coordinates (u, v, z) are used instead. In this convention, the traces of the coordinate surfaces of the xy-PLANE are confocal PARABOLAS with a common axis. The u curves open into the NEGATIVE x-AXIS; the v curves open into the POSITIVE x-AXIS. The u and v curves intersect along the y-AXIS.

$$x = \tfrac{1}{2}(u^2 - v^2) \quad (4)$$

$$y = uv \quad (5)$$

$$z = z, \quad (6)$$

where $u \in [0, \infty)$, $v \in [0, \infty)$, and $z \in (-\infty, \infty)$. The SCALE FACTORS are

$$h_1 = \sqrt{u^2 + v^2} \quad (7)$$

$$h_2 = \sqrt{u^2 + v^2} \quad (8)$$

$$h_3 = 1. \quad (9)$$

LAPLACE'S EQUATION is

$$\nabla^2 f = \frac{1}{u^2 + v^2}\left(\frac{\partial^2 f}{\partial u^2} + \frac{\partial^2 f}{\partial v^2}\right) + \frac{\partial^2 f}{\partial z^2}. \quad (10)$$

The HELMHOLTZ DIFFERENTIAL EQUATION is SEPARABLE in parabolic cylindrical coordinates.

see also CONFOCAL PARABOLOIDAL COORDINATES, HELMHOLTZ DIFFERENTIAL EQUATION—PARABOLIC CYLINDRICAL COORDINATES, PARABOLIC COORDINATES

References

Arfken, G. "Parabolic Cylinder Coordinates (ξ, η, z)." §2.8 in *Mathematical Methods for Physicists, 2nd ed.* Orlando, FL: Academic Press, p. 97, 1970.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 658, 1953.

Parabolic Fixed Point

A FIXED POINT of a LINEAR TRANSFORMATION for which the rescaled variables satisfy

$$(\delta - \alpha)^2 + 4\beta\gamma = 0.$$

see also ELLIPTIC FIXED POINT (MAP), HYPERBOLIC FIXED POINT (MAP), LINEAR TRANSFORMATION

Parabolic Geometry

see EUCLIDEAN GEOMETRY

Parabolic Horn Cyclide

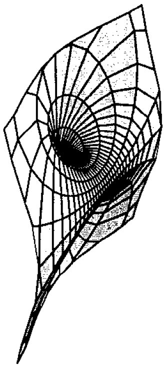

A PARABOLIC CYCLIDE formed by inversion of a HORN TORUS when the inversion sphere is tangent to the TORUS.

see also CYCLIDE, PARABOLIC RING CYCLIDE, PARABOLIC SPINDLE CYCLIDE

Parabolic Partial Differential Equation

A PARTIAL DIFFERENTIAL EQUATION of second-order, i.e., one of the form

$$Au_{xx} + 2Bu_{xy} + Cu_{yy} + Du_x + Eu_y + F = 0, \quad (1)$$

is called parabolic if the MATRIX

$$\mathsf{Z} \equiv \begin{bmatrix} A & B \\ B & C \end{bmatrix} \quad (2)$$

satisfies $\det(\mathsf{Z}) = 0$. The HEAT CONDUCTION EQUATION and other diffusion equations are examples. Initial-boundary conditions are used to give

$$u(x,t) = g(x,t) \quad \text{for } x \in \partial\Omega, t > 0 \quad (3)$$

$$u(x,0) = v(x) \quad \text{for } x \in \Omega, \quad (4)$$

where

$$u_{xx} = f(u_x, u_y, u, x, y) \quad (5)$$

holds in Ω.

see also ELLIPTIC PARTIAL DIFFERENTIAL EQUATION, HYPERBOLIC PARTIAL DIFFERENTIAL EQUATION, PARTIAL DIFFERENTIAL EQUATION

Parabolic Point

A point $\mathbf{p}$ on a REGULAR SURFACE $M \in \mathbb{R}^3$ is said to be parabolic if the GAUSSIAN CURVATURE $K(\mathbf{p}) = 0$ but $S(\mathbf{p}) \neq 0$ (where S is the SHAPE OPERATOR), or equivalently, exactly one of the PRINCIPAL CURVATURES κ_1 and κ_2 is 0.

see also ANTICLASTIC, ELLIPTIC POINT, GAUSSIAN CURVATURE, HYPERBOLIC POINT, PLANAR POINT, SYNCLASTIC

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 280, 1993.

Parabolic Ring Cyclide

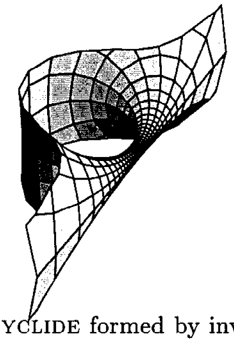

A PARABOLIC CYCLIDE formed by inversion of a RING TORUS when the inversion sphere is tangent to the TORUS.

see also CYCLIDE, PARABOLIC HORN CYCLIDE, PARABOLIC SPINDLE CYCLIDE

Parabolic Rotation

The MAP

$$x' = x + 1 \tag{1}$$
$$y' = 2x + y + 1, \tag{2}$$

which leaves the PARABOLA

$$x'^2 - y' = (x+1)^2 - (2x+y+1) = x^2 - y \tag{3}$$

invariant.

see also PARABOLA, ROTATION

Parabolic Segment

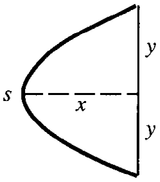

The ARC LENGTH of the parabolic segment shown above is given by

$$s = \sqrt{4x^2+y^2} + \frac{y^2}{2x}\ln\left(\frac{2x+\sqrt{4x^2+y^2}}{y}\right). \tag{1}$$

The AREA contained between the curves

$$y = x^2 \tag{2}$$
$$y = ax + b \tag{3}$$

can be found by eliminating y,

$$x^2 - ax - b = 0, \tag{4}$$

so the points of intersection are

$$x_\pm = \tfrac{1}{2}(a \pm \sqrt{a^2+4b}). \tag{5}$$

Therefore, for the AREA to be NONNEGATIVE, $a^2 + 4b > 0$, and

$$\begin{aligned} x_\pm &= \tfrac{1}{4}(a^2 \pm 2a\sqrt{a^2+b^2} + a^2 + 4b) \\ &= \tfrac{1}{4}(2a^2 + 4b \pm 2a\sqrt{a^2+4b}) \\ &= \tfrac{1}{2}(a^2 + 2b \pm a\sqrt{a^2+4b}), \end{aligned} \tag{6}$$

so the AREA is

$$\begin{aligned} A &= \int_{x_-}^{x_+} [(ax+b) - x^2]\,dx \\ &= \left[\tfrac{1}{2}ax^2 + bx - \tfrac{1}{3}x^3\right]_{(a-\sqrt{a^2+4b})/2}^{(a+\sqrt{a^2+4b})/2}. \end{aligned} \tag{7}$$

Now,

$$\begin{aligned} {x_+}^2 - {x_-}^2 &= \tfrac{1}{4}\Big[(a^2 + 2a\sqrt{a^2+4b} + a^2 + 4b) \\ &\quad - (a^2 - 2a\sqrt{a^2+4b} + a^2 + 4b)\Big] \\ &= \tfrac{1}{4}\left[4a\sqrt{a^2+4b}\right] = a\sqrt{a^2+4b} \end{aligned} \tag{8}$$

$$\begin{aligned} {x_+}^3 - {x_-}^3 &= (x_+ - x_-)({x_+}^2 + x_-x_+ + {x_-}^2) \\ &= \sqrt{a^2+4b}\Big\{\tfrac{1}{4}(a^2 + 2a\sqrt{a^2+4b} + a^2 + 4b) \\ &\quad + \tfrac{1}{4}[a^2 - (a^2+4b)] + \tfrac{1}{4}(a^2 - 2a\sqrt{a^2+4b} + a^2 + 4b)\Big\} \\ &= \tfrac{1}{4}\sqrt{a^2+4b}\,(4a^2+4b) = \sqrt{a^2+4b}\,(a^2+b). \end{aligned} \tag{9}$$

So

$$\begin{aligned} A &= \tfrac{1}{2}a^2\sqrt{a^2+4b} + b\sqrt{a^2+4b} = \tfrac{1}{3}(a^2+b)\sqrt{a^2+4b} \\ &= \sqrt{a^2+4b}\left[(\tfrac{1}{2} - \tfrac{1}{3})a^2 + b(1 - \tfrac{1}{3})\right] \\ &= (\tfrac{1}{6}a^2 + \tfrac{2}{3}b)\sqrt{a^2+4b} \\ &= \tfrac{1}{6}(a^2+4b)\sqrt{a^2+4b} = \tfrac{1}{6}(a^2+4b)^{3/2}. \end{aligned} \tag{10}$$

We now wish to find the maximum AREA of an inscribed TRIANGLE. This TRIANGLE will have two of its VERTICES at the intersections, and AREA

$$A_\triangle = \tfrac{1}{2}(x_-y_* - x_*y_- - x_+y_* + x_*y_+ + x_+y_- - x_-y_+). \tag{11}$$

But $y_* = {x_*}^2$, so

$$\begin{aligned} A_\triangle &= \tfrac{1}{2}(x_-{x_*}^2 - x_*y_- - x_+{x_*}^2 \\ &\quad + x_*y_* + x_+y_- - x_-y_+) \\ &= \tfrac{1}{2}[-{x_*}^2(x_+ - x_-) + x_*(y_+ - y_-) \\ &\quad + (x_+y_- - x_-y_+)]. \end{aligned} \tag{12}$$

The maximum AREA will occur when

$$\frac{\partial A_\triangle}{\partial x_*} = \tfrac{1}{2}[-2(x_+ - x_-)x_* + (y_+ - y_-)] = 0. \tag{13}$$

But

$$x_+ - x_- = \sqrt{a^2+4b} \tag{14}$$

$$y_+ - y_- = a\sqrt{a^2+4b}, \tag{15}$$

so

$$x_* = \frac{1}{2}\frac{y_+ - y_-}{x_+ - x_-} = \tfrac{1}{2}a \tag{16}$$

and

$$A_\Delta = \tfrac{1}{2}[-(\tfrac{1}{2}a)^2(x_+ - x_-) + (\tfrac{1}{2}a)(y_+ - y_-) + (x_+y_- - x_-y_+)]. \tag{17}$$

Working on the third term

$$\begin{aligned} x_+y_- &= \tfrac{1}{4}(a+\sqrt{a^2+4b})(a^2+2b-a\sqrt{a^2+4b}) \\ &= \tfrac{1}{4}\Big[a^3+2ab-a^2\sqrt{a^2+4b}+a^2\sqrt{a^2+4b} \\ &\quad +2b\sqrt{a^2+4b}-a(a^2+4b)\Big] \\ &= \tfrac{1}{4}[-2ab+2b\sqrt{a^2+4b}] \end{aligned} \tag{18}$$

$$\begin{aligned} x_-y_+ &= \tfrac{1}{4}(a-\sqrt{a^2+4b})(a^2+2b+a\sqrt{a^2+4b}) \\ &= \tfrac{1}{4}\Big[a^3+2ab+a^2\sqrt{a^2+4b}-a^2\sqrt{a^2+4b} \\ &\quad -2b\sqrt{a^2+4b}-a(a^2+4b)\Big] \\ &= \tfrac{1}{4}[-2ab-2b\sqrt{a^2+4b}], \end{aligned} \tag{19}$$

so

$$x_+y_- - x_-y_+ = \tfrac{1}{4}(4b\sqrt{a^2+4b}) = b\sqrt{a^2+4b} \tag{20}$$

and

$$\begin{aligned} A_\Delta &= \tfrac{1}{2}(-\tfrac{1}{4}a^2\sqrt{a^2+4b}+\tfrac{1}{2}a^2\sqrt{a^2+4b}+b\sqrt{a^2+b^2}) \\ &= \tfrac{1}{2}\sqrt{a^2+4b}\,[(\tfrac{1}{2}-\tfrac{1}{4})a^2+b] = \tfrac{1}{2}\sqrt{a^2+4b}\,(\tfrac{1}{4}a^2+b) \\ &= \tfrac{1}{8}\sqrt{a^2+4b}\,(a^2+4b) = \tfrac{1}{8}(a^2+4b)^{3/2}, \end{aligned} \tag{21}$$

which gives the result known to Archimedes in the third century BC that

$$\frac{A}{A_\Delta} = \frac{\frac{1}{6}}{\frac{1}{8}} = \frac{4}{3}. \tag{22}$$

The AREA of the parabolic segment of height h opening upward along the y-AXIS is

$$A = 2\int_0^h \sqrt{y}\,dy = \tfrac{1}{3}h^{3/2}. \tag{23}$$

The weighted mean of y is

$$\langle y\rangle = 2\int_0^h y\sqrt{y}\,dy = 2\int_0^h y^{3/2}\,dy = \tfrac{4}{5}h^{5/2}. \tag{24}$$

The CENTROID is then given by

$$\bar{y} = \frac{\langle y\rangle}{A} = \tfrac{3}{5}h. \tag{25}$$

see also CENTROID (GEOMETRIC), PARABOLA, SEGMENT

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 125, 1987.

Parabolic Spindle Cyclide

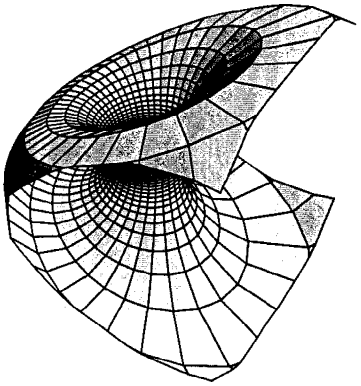

A PARABOLIC CYCLIDE formed by inversion of a SPINDLE TORUS when the inversion sphere is tangent to the TORUS.

see also CYCLIDE, PARABOLIC HORN CYCLIDE, PARABOLIC RING CYCLIDE

Parabolic Spiral

see FERMAT'S SPIRAL

Parabolic Umbilic Catastrophe

A CATASTROPHE which can occur for four control factors and two behavior axes.

Paraboloid

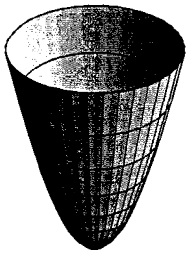

The SURFACE OF REVOLUTION of the PARABOLA. It is a QUADRATIC SURFACE which can be specified by the Cartesian equation

$$z = a(x^2+y^2), \tag{1}$$

or parametrically by

$$x(u,v) = \sqrt{u}\cos v \tag{2}$$

$$y(u,v) = \sqrt{u}\sin v \tag{3}$$

$$z(u,v) = u, \tag{4}$$

where $u \in [0,h]$, $v \in [0,2\pi)$, and h is the height.

The VOLUME of the paraboloid is

$$V = \pi\int_0^h z\,dz = \tfrac{1}{2}\pi h^2. \tag{5}$$

The weighted mean of z over the paraboloid is

$$\langle z\rangle = \pi\int_0^h z^2\,dz = \tfrac{1}{3}\pi h^3. \tag{6}$$

The Centroid is then given by

$$\bar z = \frac{\langle z\rangle}{V} = \tfrac{2}{3}h \tag{7}$$

(Beyer 1987).

see also Elliptic Paraboloid, Hyperbolic Paraboloid, Parabola

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 133, 1987.

Gray, A. "The Paraboloid." §11.5 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 221–222, 1993.

Paraboloid Geodesic

A Geodesic on a Paraboloid has differential parameters defined by

$$\begin{aligned} P &\equiv \left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2 + \left(\frac{\partial z}{\partial u}\right)^2 \\ &= 1 + \frac{\cos^2 v}{4u} + \frac{\sin^2 v}{4u} = 1 + \frac{1}{4u} && (1) \\ Q &\equiv \frac{\partial^2 x}{\partial u\,\partial v} + \frac{\partial^2 y}{\partial u\,\partial v} + \frac{\partial^2 z}{\partial u\,\partial v} \\ &= 0 + u\cos^2 v + u\sin^2 v = u && (2) \\ R &\equiv 0 - \frac{\sin v}{2\sqrt{u}} + \frac{\cos v}{2\sqrt{u}} = \frac{1}{2\sqrt{u}}(\cos v - \sin v). && (3) \end{aligned}$$

The Geodesic is then given by solving the Euler-Lagrange Differential Equation

$$\frac{\frac{\partial P}{\partial v} + 2v'\frac{\partial Q}{\partial v} + v'^2\frac{\partial R}{\partial v}}{2\sqrt{P + 2Qv' + Rv'^2}} - \frac{d}{du}\left(\frac{Q + Rv'}{\sqrt{P + 2Qv' + Rv'^2}}\right) = 0. \tag{4}$$

As given by Weinstock (1974), the solution simplifies to

$$\begin{aligned} &u - c^2 \\ &= u(1 + 4c^2)\sin^2\{v - 2c\ln[k(2\sqrt{u - c^2} + \sqrt{4u + 1})]\}. \end{aligned} \tag{5}$$

see also Geodesic

References

Weinstock, R. *Calculus of Variations, with Applications to Physics and Engineering.* New York: Dover, p. 45, 1974.

Paraboloidal Coordinates

see Confocal Paraboloidal Coordinates

Paracompact Space

A paracompact space is a Hausdorff Space such that every open Cover has a Locally Finite open Refinement. Paracompactness is a very common property that Topological Spaces satisfy. Paracompactness is similar to the compactness property, but generalized for slightly "bigger" Spaces. All Manifolds (e.g, second countable and Hausdorff) are paracompact.

see also Hausdorff Space, Locally Finite Space, Manifold, Topological Space

Paracycle

see Astroid

Paradox

A statement which appears self-contradictory or contrary to expectations, also known as an Antinomy. Bertrand Russell classified known logical paradoxes into seven categories.

Ball and Coxeter (1987) give several examples of geometrical paradoxes.

see also Alias' Paradox, Aristotle's Wheel Paradox, Arrow's Paradox, Banach-Tarski Paradox, Barber Paradox, Bernoulli's Paradox, Berry Paradox, Bertrand's Paradox, Cantor's Paradox, Coastline Paradox, Coin Paradox, Elevator Paradox, Epimenides Paradox, Eubulides Paradox, Grelling's Paradox, Hausdorff Paradox, Hempel's Paradox, Heterological Paradox, Leonardo's Paradox, Liar's Paradox, Logical Paradox, Potato Paradox, Richard's Paradox, Russell's Paradox, Saint Petersburg Paradox, Siegel's Paradox, Simpson's Paradox, Skolem Paradox, Smarandache Paradox, Socrates' Paradox, Sorites Paradox, Thomson Lamp Paradox, Unexpected Hanging Paradox, Zeeman's Paradox, Zeno's Paradoxes

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 84–86, 1987.

Bunch, B. *Mathematical Fallacies and Paradoxes.* New York: Dover, 1982.

Carnap, R. *Introduction to Symbolic Logic and Its Applications.* New York: Dover, 1958.

Curry, H. B. *Foundations of Mathematical Logic.* New York: Dover, 1977.

Kasner, E. and Newman, J. R. "Paradox Lost and Paradox Regained." In *Mathematics and the Imagination.* Redmond, WA: Tempus Books, pp. 193–222, 1989.

Northrop, E. P. *Riddles in Mathematics: A Book of Paradoxes.* Princeton, NJ: Van Nostrand, 1944.

O'Beirne, T. H. *Puzzles and Paradoxes.* New York: Oxford University Press, 1965.

Quine, W. V. "Paradox." *Sci. Amer.* **206**, 84–96, Apr. 1962.

Paradromic Rings

Rings produced by cutting a strip that has been given m half twists and been re-attached into n equal strips (Ball and Coxeter 1987, pp. 127–128).

see also MÖBIUS STRIP

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 127–128, 1987.

Paragyrate Diminished Rhombicosidodecahedron

see JOHNSON SOLID

Parallel

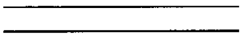

Two lines in 2-dimensional EUCLIDEAN SPACE are said to be parallel if they do not intersect. In 3-dimensional EUCLIDEAN SPACE, parallel lines not only fail to intersect, but also maintain a constant separation between points closest to each other on the two lines. (Lines in 3-space which are not parallel but do not intersect are called SKEW LINES.)

In a NON-EUCLIDEAN GEOMETRY, the concept of parallelism must be modified from its intuitive meaning. This is accomplished by changing the so-called PARALLEL POSTULATE. While this has counterintuitive results, the geometries so defined are still completely self-consistent.

see also ANTIPARALLEL, HYPERPARALLEL, LINE, NON-EUCLIDEAN GEOMETRY, PARALLEL CURVE, PARALLEL POSTULATE PERPENDICULAR, SKEW LINES

Parallel Axiom

see PARALLEL POSTULATE

Parallel Class

A set of blocks, also called a RESOLUTION CLASS, that partition the set V, where (V, B) is a balanced incomplete BLOCK DESIGN.

see also BLOCK DESIGN, RESOLVABLE

References

Abel, R. J. R. and Furino, S. C. "Resolvable and Near Resolvable Designs." §I.6 in *The CRC Handbook of Combinatorial Designs* (Ed. C. J. Colbourn and J. H. Dinitz). Boca Raton, FL: CRC Press, pp. 87–94, 1996.

Parallel Curve

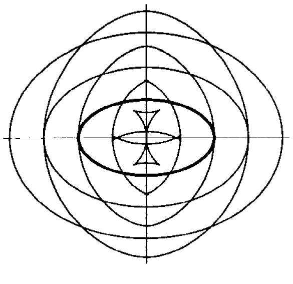

The two branches of the parallel curve a distance k away from a parametrically represented curve $(f(t), g(t))$ are

$$x = f \pm \frac{kg'}{\sqrt{f'^2 + g'^2}}$$

$$y = g \mp \frac{kf'}{\sqrt{f'^2 + g'^2}}.$$

The above figure shows the curves parallel to the ELLIPSE.

References

Gray, A. "Parallel Curves." §5.7 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 95–97, 1993.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 42–43, 1972.

Lee, X. "Parallel." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Parallel_dir/parallel.html`.

Yates, R. C. "Parallel Curves." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 155–159, 1952.

Parallel Postulate

Given any straight line and a point not on it, there "exists one and only one straight line which passes" through that point and never intersects the first line, no matter how far they are extended. This statement is equivalent to the fifth of EUCLID'S POSTULATES, which Euclid himself avoided using until proposition 29 in the *Elements*. For centuries, many mathematicians believed that this statement was not a true postulate, but rather a theorem which could be derived from the first four of EUCLID'S POSTULATES. (That part of geometry which could be derived using only postulates 1–4 came to be known as ABSOLUTE GEOMETRY.)

Over the years, many purported proofs of the parallel postulate were published. However, none were correct, including the 28 "proofs" G. S. Klügel analyzed in his dissertation of 1763 (Hofstadter 1989). In 1823, Janos Bolyai and Lobachevsky independently realized that entirely self-consistent "NON-EUCLIDEAN GEOMETRIES" could be created in which the parallel postulate *did not hold.* (Gauss had also discovered but suppressed the existence of non-Euclidean geometries.)

As stated above, the parallel postulate describes the type of geometry now known as PARABOLIC GEOMETRY. If, however, the phrase "exists one and only one straight line which passes" is replace by "exist no line which passes," or "exist at least two lines which pass," the postulate describes equally valid (though less intuitive) types of geometries known as ELLIPTIC and HYPERBOLIC GEOMETRIES, respectively.

The parallel postulate is equivalent to the EQUIDISTANCE POSTULATE, PLAYFAIR'S AXIOM, PROCLUS' AXIOM, TRIANGLE POSTULATE. There is also a single parallel axiom in HILBERT'S AXIOMS which is equivalent to Euclid's parallel postulate.

see also ABSOLUTE GEOMETRY, EUCLID'S AXIOMS, EUCLIDEAN GEOMETRY, HILBERT'S AXIOMS, NON-EUCLIDEAN GEOMETRY, PLAYFAIR'S AXIOM, TRIANGLE POSTULATE

References
Dixon, R. *Mathographics.* New York: Dover, p. 27, 1991.
Hilbert, D. *The Foundations of Geometry, 2nd ed.* Chicago, IL: Open Court, 1980.
Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, pp. 88–92, 1989.
Iyanaga, S. and Kawada, Y. (Eds.). "Hilbert's System of Axioms." §163B in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 544–545, 1980.

Parallel (Surface of Revolution)

A parallel of a SURFACE OF REVOLUTION is the intersection of the surface with a PLANE orthogonal to the axis of revolution.

see also MERIDIAN, SURFACE OF REVOLUTION

References
Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 358, 1993.

Parallelepiped

In 3-D, a parallelepiped is a PRISM whose faces are all PARALLELOGRAMS. The volume of a 3-D parallelepiped is given by the SCALAR TRIPLE PRODUCT

$$\begin{aligned}V_{\text{parallelepiped}} &= |\mathbf{B}\cdot(\mathbf{B}\times\mathbf{C})|\\ &= |\mathbf{C}\cdot(\mathbf{A}\times\mathbf{B})| = |\mathbf{B}\cdot(\mathbf{C}\times\mathbf{A})|.\end{aligned}$$

In n-D, a parallelepiped is the POLYTOPE spanned by n VECTORS $\mathbf{v}_1$, ..., $\mathbf{v}_n$ in a VECTOR SPACE over the reals,

$$\operatorname{span}(\mathbf{v}_1,\ldots,\mathbf{v}_n) = t_1\mathbf{v}_1 + \ldots + t_n\mathbf{v}_n,$$

where $t_i \in [0,1]$ for $i = 1$, ..., n. In the usual interpretation, the VECTOR SPACE is taken as EUCLIDEAN SPACE, and the CONTENT of this parallelepiped is given by

$$\operatorname{abs}(\det(\mathbf{v}_1,\ldots,\mathbf{v}_n)),$$

where the sign of the determinant is taken to be the "orientation" of the "oriented volume" of the parallelepiped.

see also PRISMATOID, RECTANGULAR PARALLELEPIPED, ZONOHEDRON

References
Phillips, A. W. and Fisher, I. *Elements of Geometry.* New York: Amer. Book Co., 1896.

Parallelism

see ANGLE OF PARALLELISM

Parallelizable

A sphere $\mathbb{S}^n$ is parallelizable if there exist n cuts containing linearly independent tangent vectors. There exist only three parallelizable spheres: $\mathbb{S}^1$, $\mathbb{S}^3$, and $\mathbb{S}^7$ (Adams 1962, Le Lionnais 1983).

see also SPHERE

References
Adams, J. F. "On the Non-Existence of Elements of Hopf Invariant One." *Bull. Amer. Math. Soc.* **64**, 279–282, 1958.
Adams, J. F. "On the Non-Existence of Elements of Hopf Invariant One." *Ann. Math.* **72**, 20–104, 1960.
Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 49, 1983.

Parallelogram

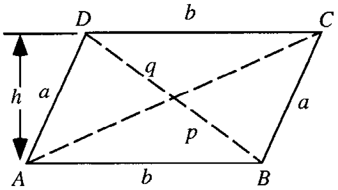

A QUADRILATERAL with opposite sides parallel (and therefore opposite angles equal). A quadrilateral with equal sides is called a RHOMBUS, and a parallelogram whose ANGLES are all RIGHT ANGLES is called a RECTANGLE.

A parallelogram of base b and height h has AREA

$$A = bh = ab\sin A = ab\sin B. \tag{1}$$

The height of a parallelogram is

$$h = a\sin A = a\sin B, \tag{2}$$

and the DIAGONALS are

$$\begin{aligned}p &= \sqrt{a^2+b^2-2ab\cos A} &(3)\\ q &= \sqrt{a^2+b^2-2ab\cos B} &(4)\\ &= \sqrt{a^2+b^2+2ab\cos A} &(5)\end{aligned}$$

(Beyer 1987).

The AREA of the parallelogram with sides formed by the VECTORS (a,c) and (b,d) is

$$A = \det\left(\begin{bmatrix}a & b\\ c & d\end{bmatrix}\right) = |ad-bc|. \tag{6}$$

Given a parallelogram P with area $A(P)$ and linear transformation T, the AREA of $T(P)$ is

$$A(T(P)) = \begin{vmatrix}a & b\\ c & d\end{vmatrix} A(P). \tag{7}$$

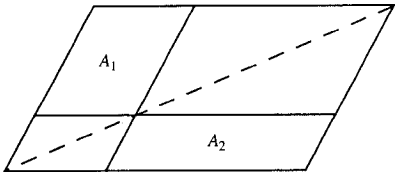

As shown by Euclid, if lines parallel to the sides are drawn through any point on a diagonal of a parallelogram, then the parallelograms not containing segments of that diagonal are equal in AREA (and conversely), so in the above figure, $A_1 = A_2$ (Johnson 1929).

see also DIAMOND, LOZENGE, PARALLELOGRAM ILLUSION, RECTANGLE, RHOMBUS, VARIGNON PARALLELOGRAM, WITTENBAUER'S PARALLELOGRAM

References
Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 123, 1987.
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, p. 61, 1929.

Parallelogram Illusion

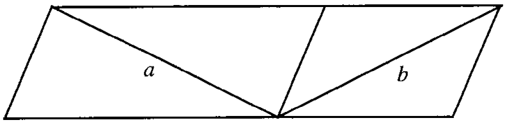

The sides a and b have the same length, appearances to the contrary.

Parallelogram Law

Let $|\cdot|$ denote the NORM of a quantity. Then the quantities x and y satisfy the parallelogram law if

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

If the NORM is defined as $|f| = \sqrt{\langle f|f\rangle}$ (the so-called L_2-NORM), then the law will always hold.

see also L_2-NORM, NORM

Parallelohedron

A special class of ZONOHEDRON. There are five parallelohedra with an infinity of equal and similarly situated replicas which are SPACE-FILLING POLYHEDRA: the CUBE, ELONGATED DODECAHEDRON, hexagonal PRISM, RHOMBIC DODECAHEDRON, and TRUNCATED OCTAHEDRON.

see also PARALLELOTOPE, SPACE-FILLING POLYHEDRON

References
Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, p. 29, 1973.

Parallelotope

Move a point Π_0 along a LINE for an initial point to a final point. It traces out a LINE SEGMENT Π_1. When Π_1 is translated from an initial position to a final position, it traces out a PARALLELOGRAM Π_2. When Π_2 is translated, it traces out a PARALLELEPIPED Π_3. The generalization of Π_n to n-D is then called a parallelotope. Π_n has 2^n vertices and

$$N_k = 2^{n-k}\binom{n}{k}$$

Π_ks, where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT and $k = 0$, 1, ..., n (Coxeter 1973). These are also the coefficients of $(2k+1)^n$.

see also HONEYCOMB, HYPERCUBE, ORTHOTOPE, PARALLELOHEDRON

References
Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, pp. 122–123, 1973.
Klee, V. and Wagon, S. *Old and New Unsolved Problems in Plane Geometry and Number Theory.* Washington, DC: Math. Assoc. Amer., 1991.
Zaks, J. "Neighborly Families of Congruent Convex Polytopes." *Amer. Math. Monthly* **94**, 151–155, 1987.

Paralogic Triangles

At the points where a line cuts the sides of a TRIANGLE $\Delta A_1A_2A_3$, perpendiculars to the sides are drawn, forming a TRIANGLE $\Delta B_1B_2B_3$ similar to the given TRIANGLE. The two triangles are also in perspective. One point of intersection of their CIRCUMCIRCLES is the SIMILITUDE CENTER, and the other is the PERSPECTIVE CENTER. The CIRCUMCIRCLES meet ORTHOGONALLY.

see also CIRCUMCIRCLE, ORTHOGONAL CIRCLES, PERSPECTIVE CENTER, SIMILITUDE CENTER

References
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 258–262, 1929.

Parameter

A parameter m used in ELLIPTIC INTEGRALS defined to be $m \equiv k^2$, where k is the MODULUS. An ELLIPTIC INTEGRAL is written $I(\phi|m)$ when the parameter is used. The complementary parameter is defined by

$$m' \equiv 1 - m, \qquad (1)$$

where m is the parameter. Let q be the NOME, k the MODULUS, and $m \equiv k^2$ the PARAMETER. Then

$$q(m) = e^{-\pi K'(m)/K(m)} \qquad (2)$$

where $K(m)$ is the complete ELLIPTIC INTEGRAL OF THE FIRST KIND. Then the inverse of $q(m)$ is given by

$$m(q) = \frac{\vartheta_2{}^4(q)}{\vartheta_3{}^4(q)}, \qquad (3)$$

where ϑ_i is a THETA FUNCTION.

see also AMPLITUDE, CHARACTERISTIC (ELLIPTIC INTEGRAL), ELLIPTIC INTEGRAL, ELLIPTIC INTEGRAL OF THE FIRST KIND, MODULAR ANGLE, MODULUS (ELLIPTIC INTEGRAL), NOME, PARAMETER, THETA FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 590, 1972.

Parameter (Quadric)

The number θ in the QUADRIC

$$\frac{x^2}{a^2+\theta}+\frac{y^2}{b^2+\theta}+\frac{z^2}{c^2+\theta}=1$$

is called the parameter.

see also QUADRIC

Parameterization

The specification of a curve, surface, etc., by means of one or more variables which are allowed to take on values in a given specified range.

see also ISOTHERMAL PARAMETERIZATION, REGULAR PARAMETERIZATION, SURFACE PARAMETERIZATION

Parametric Latitude

An AUXILIARY LATITUDE also called the REDUCED LATITUDE and denoted η or θ. It gives the LATITUDE on a SPHERE of RADIUS a for which the parallel has the same radius as the parallel of geodetic latitude ϕ and the ELLIPSOID through a given point. It is given by

$$\eta=\tan^{-1}(\sqrt{1-e^2}\,\tan\phi).$$

In series form,

$$\eta=\phi-e_1\sin(2\phi)+\tfrac{1}{2}{e_1}^2\sin(4\phi)-\tfrac{1}{3}{e_1}^3\sin(6\phi)+\ldots,$$

where

$$e_1\equiv\frac{1-\sqrt{1-e^2}}{1+\sqrt{1-e^2}}.$$

see also AUXILIARY LATITUDE, ELLIPSOID, LATITUDE, SPHERE

References

Adams, O. S. "Latitude Developments Connected with Geodesy and Cartography with Tables, Including a Table for Lambert Equal-Area Meridional Projections." Spec. Pub. No. 67. U. S. Coast and Geodetic Survey, 1921.

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, p. 18, 1987.

Parametric Test

A STATISTICAL TEST in which assumptions are made about the underlying distribution of observed data.

Pareto Distribution

The distribution

$$P(x)=\left(\frac{x}{b}\right)^{a+2}.$$

References

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 252, 1993.

Parity

The parity of a number n is the sum of the bits in BINARY representation (mod 2). The parities of the first few integers (starting with 0) are 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, ... (Sloane's A010060) summarized in the following table.

N	Binary	Parity	N	Binary	Parity
1	1	1	11	1011	1
2	10	1	12	1100	0
3	11	0	13	1101	1
4	100	1	14	1110	1
5	101	0	15	1111	0
6	110	0	16	10000	1
7	111	1	17	10001	0
8	1000	1	18	10010	0
9	1001	0	19	10011	1
10	1010	0	20	10100	0

The constant generated by the sequence of parity digits is called the THUE-MORSE CONSTANT.

see also BINARY, THUE-MORSE CONSTANT

References

Sloane, N. J. A. Sequence A010060 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Parity Constant

see THUE-MORSE CONSTANT

Parking Constant

see RÉNYI'S PARKING CONSTANTS

Parodi's Theorem

The EIGENVALUES λ satisfying $P(\lambda)=0$, where $P(\lambda)$ is the CHARACTERISTIC POLYNOMIAL, lie in the unions of the DISKS

$$|z|\leq 1$$

$$|z+b_1|\leq\sum_{j=1}^{n}|b_j|.$$

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1119, 1979.

Parry Circle

The CIRCLE passing through the ISODYNAMIC POINTS and the CENTROID of a TRIANGLE (Kimberling 1998, pp. 227–228).

see also CENTROID (TRIANGLE), ISODYNAMIC POINTS, PARRY POINT

References
Kimberling, C. "Triangle Centers and Central Triangles." *Congr. Numer.* **129**, 1–295, 1998.

Parry Point

The intersection of the PARRY CIRCLE and the CIRCUMCIRCLE of a TRIANGLE. The TRILINEAR COORDINATES of the Parry point are

$$\frac{a}{2a^2-b^2-c^2}:\frac{b}{2b^2-c^2-a^2}:\frac{c}{2c^2-a^2-b^2}$$

(Kimberling 1998, pp. 227–228).

see also PARRY CIRCLE

References
Kimberling, C. "Parry Point." `http://www.evansville.edu/~ck6/tcenters/recent/parry.html`.
Kimberling, C. "Triangle Centers and Central Triangles." *Congr. Numer.* **129**, 1–295, 1998.

Parseval's Integral

The POISSON INTEGRAL with $n=0$.

$$J_0(z)=\frac{1}{[\Gamma(n+\frac{1}{2})]^2}\int_0^\pi \cos(z\cos\theta)\,d\theta,$$

where $J_0(z)$ is a BESSEL FUNCTION OF THE FIRST KIND and $\Gamma(x)$ is a GAMMA FUNCTION.

Parseval's Relation

Let $F(\nu)$ and $G(\nu)$ be the FOURIER TRANSFORMS of $f(t)$ and $g(t)$, respectively. Then

$$\begin{aligned}&\int_{-\infty}^{\infty} f(t)g^*(t)\,dt\\&=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} F(\nu)e^{-2\pi i\nu t}\,d\nu\int_{-\infty}^{\infty} G^*(\nu')e^{2\pi i\nu' t}\,d\nu'\right]d\nu'\\&=\int_{-\infty}^{\infty} F(\nu)\int_{-\infty}^{\infty} G^*(\nu')\delta(\nu'-\nu)\,d\nu'\,d\nu\\&=\int_{-\infty}^{\infty} F(\nu)G^*(\nu)\,d\nu.\end{aligned}$$

see also FOURIER TRANSFORM, PARSEVAL'S THEOREM

References
Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 425, 1985.

Parseval's Theorem

Let $E(t)$ be a continuous function and $E(t)$ and E_ν be FOURIER TRANSFORM pairs so that

$$E(t)\equiv\int_{-\infty}^{\infty} E_\nu e^{-2\pi i\nu t}\,d\nu \tag{1}$$

$$E^*(t)\equiv\int_{-\infty}^{\infty} E_{\nu'}{}^* e^{2\pi i\nu' t}\,d\nu'. \tag{2}$$

Then

$$\begin{aligned}\int_{-\infty}^{\infty}|E(t)|^2\,dt&=\int_{-\infty}^{\infty}E(t)E^*(t)\,dt\\&=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}E_\nu e^{-2\pi i\nu t}\,d\nu\int_{-\infty}^{\infty}E_{\nu'}{}^*e^{2\pi i\nu' t}\,d\nu'\right]dt\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E_\nu E_{\nu'}{}^*e^{2\pi it(\nu'-\nu)}\,d\nu\,d\nu'\,dt\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E_\nu E_{\nu'}{}^*e^{2\pi it(\nu'-\nu)}\,dt\,d\nu\,d\nu'\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(\nu'-\nu)E_\nu E_{\nu'}{}^*\,d\nu\,d\nu'\\&=\int_{-\infty}^{\infty}E_\nu E_\nu{}^*\,d\nu=\int_{-\infty}^{\infty}|E_\nu|^2\,d\nu.\end{aligned} \tag{3}$$

where $\delta(x-x_0)$ is the DELTA FUNCTION.

For finite FOURIER TRANSFORM pairs h_k and H_n,

$$\sum_{k=0}^{N-1}|h_k|^2=\frac{1}{N}\sum_{n=0}^{N-1}|H_n|^2. \tag{4}$$

If a function has a FOURIER SERIES given by

$$f(x)=\tfrac{1}{2}a_0+\sum_{n=1}^{\infty}a_n\cos(nx)+\sum_{n=1}^{\infty}b_n\sin(nx), \tag{5}$$

then BESSEL'S INEQUALITY becomes an equality known as Parseval's theorem. From (5),

$$\begin{aligned}[f(x)]^2&=\tfrac{1}{4}a_0{}^2+a_0\sum_{n=1}^{\infty}[a_n\cos(nx)+b_n\sin(nx)]\\&+\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}[a_na_m\cos(nx)\cos(mx)\\&+a_nb_m\cos(nx)\sin(mx)+a_mb_n\sin(nx)\cos(mx)\\&+b_nb_m\sin(nx)\sin(mx)].\end{aligned} \tag{6}$$

Integrating

$$\begin{aligned}\int_{-\pi}^{\pi} [f(x)]^2\,dx &= \tfrac{1}{4}{a_0}^2 \int_{-\pi}^{\pi} dx \\ &\quad + a_0 \int_{-\pi}^{\pi} \sum_{n=1}^{\infty} [a_n \cos(nx) + b_n \sin(nx)]\,dx \\ &+ \int_{-\pi}^{\pi} \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} [a_n a_m \cos(nx)\cos(mx) \\ &+ a_n b_m \cos(nx)\sin(mx) + a_m b_n \sin(nx)\cos(mx) \\ &+ b_n b_m \sin(nx)\sin(mx)]\,dx = \tfrac{1}{4}{a_0}^2 (2\pi) + 0 \\ &+ \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} [a_n a_m \pi \delta_{nm} + 0 + 0 + b_n b_m \pi \delta_{nm}], \qquad (7)\end{aligned}$$

so

$$\frac{1}{\pi} \int_{-\pi}^{\pi} [f(x)]^2\,dx = \tfrac{1}{2}{a_0}^2 + \sum_{n=1}^{\infty} ({a_n}^2 + {b_n}^2). \qquad (8)$$

For a generalized FOURIER SERIES with a COMPLETE BASIS $\{\phi_i\}_{i=1}^{\infty}$, an analogous relationship holds. For a COMPLEX FOURIER SERIES,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2\,dx = \sum_{n=-\infty}^{\infty} |a_n|^2. \qquad (9)$$

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1101, 1979.

Part Metric

A METRIC defined by

$$d(z,w) = \sup \left[\left| \frac{\ln u(z)}{u(w)} \right| : u \in H^+ \right],$$

where H^+ denotes the POSITIVE HARMONIC FUNCTIONS on a DOMAIN. The part metric is invariant under CONFORMAL MAPS for any DOMAIN.

References

Bear, H. S. "Part Metric and Hyperbolic Metric." *Amer. Math. Monthly* **98**, 109–123, 1991.

Partial Derivative

Partial derivatives are defined as derivatives of a function of multiple variables when all but the variable of interest are held fixed during the differentiation.

$$\frac{\partial f}{\partial x_m} \equiv \lim_{h\to 0} \frac{f(x_1, \ldots, x_m + h, \ldots, x_n) - f(x_1, \ldots, x_m, \ldots, x_n)}{h}. \qquad (1)$$

The above partial derivative is sometimes denoted f_{x_m} for brevity. For a "nice" 2-D function $f(x,y)$ (i.e., one for which f, f_x, f_y, f_{xy}, f_{yx} exist and are continuous in a NEIGHBORHOOD (a,b)), then $f_{xy}(a,b) = f_{yx}(a,b)$. Partial derivatives involving more than one variable are called MIXED PARTIAL DERIVATIVES.

For nice functions, mixed partial derivatives must be equal regardless of the order in which the differentiation is performed so, for example,

$$f_{xy} = f_{yx} \qquad (2)$$

$$f_{xxy} = f_{xyx} = f_{yxx}. \qquad (3)$$

For an EXACT DIFFERENTIAL,

$$df = \left(\frac{\partial f}{\partial x}\right)_y dx + \left(\frac{\partial f}{\partial y}\right)_x dy, \qquad (4)$$

so

$$\left(\frac{\partial y}{\partial x}\right)_f = -\frac{\left(\frac{\partial f}{\partial x}\right)_y}{\left(\frac{\partial f}{\partial y}\right)_x}. \qquad (5)$$

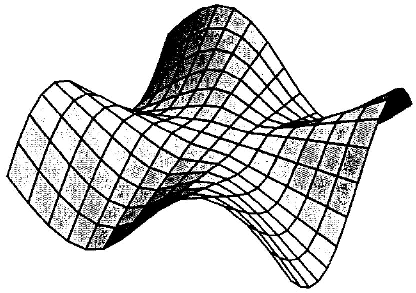

If the continuity requirement for MIXED PARTIALS is dropped, it is possible to construct functions for which MIXED PARTIALS are *not* equal. An example is the function

$$f(x,y) = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} & \text{for } (x,y) = 0 \\ 0 & \text{for } (x,y) = 0, \end{cases} \qquad (6)$$

which has $f_{xy}(0,0) = -1$ and $f_{yx}(0,0) = 1$ (Wagon 1991). This function is depicted above and by Fischer (1986).

Abramowitz and Stegun (1972) give FINITE DIFFERENCE versions for partial derivatives.

see also ABLOWITZ-RAMANI-SEGUR CONJECTURE, DERIVATIVE, MIXED PARTIAL DERIVATIVE, MONKEY SADDLE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 883–885, 1972.

Fischer, G. (Ed.). Plate 121 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 118, 1986.

Thomas, G. B. and Finney, R. L. §16.8 in *Calculus and Analytic Geometry, 9th ed.*0201531747 Reading, MA: Addison-Wesley, 1996.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 83–85, 1991.

Partial Differential Equation

A partial differential equation (PDE) is an equation involving functions and their PARTIAL DERIVATIVES; for example, the WAVE EQUATION

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}=\frac{1}{v^2}\frac{\partial^2\psi}{\partial t^2}. \tag{1}$$

In general, partial differential equations are much more difficult to solve analytically than are ORDINARY DIFFERENTIAL EQUATIONS. They may sometimes be solved using a BÄCKLUND TRANSFORMATION, CHARACTERISTIC, GREEN'S FUNCTION, INTEGRAL TRANSFORM, LAX PAIR, SEPARATION OF VARIABLES, or—when all else fails (which it frequently does)—numerical methods.

Fortunately, partial differential equations of second-order are often amenable to analytical solution. Such PDEs are of the form

$$Au_{xx}+2Bu_{xy}+Cu_{yy}+Du_x+Eu_y+F=0. \tag{2}$$

Second-order PDEs are then classified according to the properties of the MATRIX

$$\mathsf{Z}\equiv\begin{bmatrix}A & B\\ B & C\end{bmatrix} \tag{3}$$

as ELLIPTIC, HYPERBOLIC, or PARABOLIC.

If Z is a POSITIVE DEFINITE MATRIX, i.e., $\det(\mathsf{Z})>0$, the PDE is said to be ELLIPTIC. LAPLACE'S EQUATION and POISSON'S EQUATION are examples. Boundary conditions are used to give the constraint $u(x,y)=g(x,y)$ on $\partial\Omega$, where

$$u_{xx}+u_{yy}=f(u_x,u_y,u,x,y) \tag{4}$$

holds in Ω.

If $\det(\mathsf{Z})<0$, the PDE is said to be HYPERBOLIC. The WAVE EQUATION is an example of a hyperbolic partial differential equation. Initial-boundary conditions are used to give

$$u(x,y,t)=g(x,y,t)\quad\text{for } x\in\partial\Omega, t>0 \tag{5}$$

$$u(x,y,0)=v_0(x,y)\quad\text{in }\Omega \tag{6}$$

$$u_t(x,y,0)=v_1(x,y)\quad\text{in }\Omega, \tag{7}$$

where

$$u_{xy}=f(u_x,u_t,x,y) \tag{8}$$

holds in Ω.

If $\det(\mathsf{Z})=0$, the PDE is said to be parabolic. The HEAT CONDUCTION EQUATION equation and other diffusion equations are examples. Initial-boundary conditions are used to give

$$u(x,t)=g(x,t)\quad\text{for } x\in\partial\Omega, t>0 \tag{9}$$

$$u(x,0)=v(x)\quad\text{for } x\in\Omega, \tag{10}$$

where

$$u_{xx}=f(u_x,u_y,u,x,y) \tag{11}$$

holds in Ω.

see also BÄCKLUND TRANSFORMATION, BOUNDARY CONDITIONS, CHARACTERISTIC (PARTIAL DIFFERENTIAL EQUATION), ELLIPTIC PARTIAL DIFFERENTIAL EQUATION, GREEN'S FUNCTION, HYPERBOLIC PARTIAL DIFFERENTIAL EQUATION, INTEGRAL TRANSFORM, JOHNSON'S EQUATION, LAX PAIR, MONGE-AMPÈRE DIFFERENTIAL EQUATION, PARABOLIC PARTIAL DIFFERENTIAL EQUATION, SEPARATION OF VARIABLES

References

Arfken, G. "Partial Differential Equations of Theoretical Physics." §8.1 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 437–440, 1985.

Bateman, H. *Partial Differential Equations of Mathematical Physics.* New York: Dover, 1944.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Partial Differential Equations." Ch. 19 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 818–880, 1992.

Sobolev, S. L. *Partial Differential Equations of Mathematical Physics.* New York: Dover, 1989.

Sommerfeld, A. *Partial Differential Equations in Physics.* New York: Academic Press, 1964.

Webster, A. G. *Partial Differential Equations of Mathematical Physics, 2nd corr. ed.* New York: Dover, 1955.

Partial Fraction Decomposition

A RATIONAL FUNCTION $P(x)/Q(x)$ can be rewritten using what is known as partial fraction decomposition. This procedure often allows integration to be performed on each term separately by inspection. For each factor of $Q(x)$ the form $(ax+b)^m$, introduce terms

$$\frac{A_1}{ax+b}+\frac{A_2}{(ax+b)^2}+\ldots+\frac{A_m}{(ax+b)^m}. \tag{1}$$

For each factor of the form $(ax^2+bx+c)^m$, introduce terms

$$\frac{A_1x+B_1}{ax^2+bx+c}+\frac{A_2x+B_2}{(ax^2+bx+c)^2}+\ldots+\frac{A_mx+B_m}{(ax^2+bx+c)^m}. \tag{2}$$

Then write

$$\frac{P(x)}{Q(x)}=\frac{A_1}{ax+b}+\ldots+\frac{A_2x+B_2}{ax^2+bx+c}+\ldots \tag{3}$$

and solve for the A_is and B_is.

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 13–15, 1987.

Partial Latin Square

In a normal $n \times n$ LATIN SQUARE, the entries in each row and column are chosen from a "global" set of n objects. Like a Latin square, a partial Latin square has no two rows or columns which contain the same two symbols. However, in a partial Latin square, each cell is assigned one of its own set of n possible "local" (and distinct) symbols, chosen from an overall set of more than three distinct symbols, and these symbols may vary from location to location. For example, given the possible symbols $\{1, 2, \ldots, 6\}$ which must be arranged as

$$\begin{array}{ccc} \{1,2,3\} & \{1,3,4\} & \{2,5,6\} \\ \{2,3,5\} & \{1,2,3\} & \{4,5,6\} \\ \{4,3,6\} & \{3,5,6\} & \{2,3,5\}, \end{array}$$

the 3×3 partial Latin square

$$\begin{array}{ccc} 1 & 3 & 2 \\ 2 & 4 & 5 \\ 6 & 5 & 3 \end{array}$$

can be constructed.

see also DINITZ PROBLEM, LATIN SQUARE

References

Cipra, B. "Quite Easily Done." In *What's Happening in the Mathematical Sciences* **2**, pp. 41–46, 1994.

Partial Order

A RELATION "$\leq$" is a partial order on a SET S if it has:

1. Reflexivity: $a \leq a$ for all $a \in S$.
2. Antisymmetry: $a \leq b$ and $b \leq a$ implies $a = b$.
3. Transitivity: $a \leq b$ and $b \leq c$ implies $a \leq c$.

For a partial order, the size of the longest CHAIN (ANTICHAIN) is called the LENGTH (WIDTH). A partially ordered set is also called a POSET.

see also ANTICHAIN, CHAIN, FENCE POSET, IDEAL (PARTIAL ORDER), LENGTH (PARTIAL ORDER), LINEAR EXTENSION, PARTIALLY ORDERED SET, TOTAL ORDER, WIDTH (PARTIAL ORDER)

References

Ruskey, F. "Information on Linear Extension." `http://sue.csc.uvic.ca/~cos/inf/pose/LinearExt.html`.

Partial Quotient

If the SIMPLE CONTINUED FRACTION of a REAL NUMBER x is given by

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ldots}}},$$

then the quantities a_i are called partial quotients.

see also CONTINUED FRACTION, CONVERGENT, SIMPLE CONTINUED FRACTION

Partially Ordered Set

A partially ordered set (or POSET) is a SET taken together with a PARTIAL ORDER on it. Formally, a partially ordered set is defined as an ordered pair $P = (X, \leq)$, where X is called the GROUND SET of P and $\leq$ is the PARTIAL ORDER of P.

see also CIRCLE ORDER, COVER RELATION, DOMINANCE, GROUND SET, HASSE DIAGRAM, INTERVAL ORDER, ISOMORPHIC POSETS, PARTIAL ORDER, POSET DIMENSION, REALIZER, RELATION

References

Dushnik, B. and Miller, E. W. "Partially Ordered Sets." *Amer. J. Math.* **63**, 600–610, 1941.

Fishburn, P. C. *Interval Orders and Interval Sets: A Study of Partially Ordered Sets.* New York: Wiley, 1985.

Trotter, W. T. *Combinatorics and Partially Ordered Sets: Dimension Theory.* Baltimore, MD: Johns Hopkins University Press, 1992.

Particularly Well-Behaved Functions

Functions which have DERIVATIVES of all orders at all points and which, together with their DERIVATIVES, fall off at least as rapidly as $|x|^{-n}$ as $|x| \to \infty$, no matter how large n is.

see also REGULAR SEQUENCE

Partisan Game

A GAME for which each player has a different set of moves in any position. Every position in an IMPARTIAL GAME has a NIM-VALUE.

Partition

A partition is a way of writing an INTEGER n as a sum of POSITIVE INTEGERS without regard to order, possibly subject to one or more additional constraints. Particular types of partition functions include the PARTITION FUNCTION P, giving the number of partitions of a number without regard to order, and PARTITION FUNCTION Q, giving the number of ways of writing the INTEGER n as a sum of POSITIVE INTEGERS without regard to order with the constraint that all INTEGERS in each sum are distinct.

see also AMENABLE NUMBER, DURFEE SQUARE, ELDER'S THEOREM, FERRERS DIAGRAM, GRAPHICAL PARTITION, PARTITION FUNCTION P, Partition Function Q, PERFECT PARTITION, PLANE PARTITION, SET PARTITION, SOLID PARTITION, STANLEY'S THEOREM

References

Andrews, G. E. *The Theory of Partitions.* Cambridge, England: Cambridge University Press, 1998.

Dickson, L. E. "Partitions." Ch. 3 in *History of the Theory of Numbers, Vol. 2: Diophantine Analysis.* New York: Chelsea, pp. 101–164, 1952.

Partition Function P

$P(n)$ gives the number of ways of writing the INTEGER n as a sum of POSITIVE INTEGERS without regard to order. For example, since 4 can be written

$$\begin{aligned} 4 &= 4 \\ &= 3+1 \\ &= 2+2 \\ &= 2+1+1 \\ &= 1+1+1+1, \end{aligned} \tag{1}$$

so $P(4)=5$. $P(n)$ satisfies

$$P(n) \leq \tfrac{1}{2}[P(n+1)+P(n-1)] \tag{2}$$

(Honsberger 1991). The values of $P(n)$ for $n=1, 2, \ldots$, are 1, 2, 3, 5, 7, 11, 15, 22, 30, 42, ... (Sloane's A000041). The following table gives the value of $P(n)$ for selected small n.

n	$P(n)$
50	204226
100	190569292
200	3972999029388
300	9253082936723602
400	6727090051741041926
500	2300165032574323995027
600	458004788008144308553622
700	60378285202834474611028659
800	5733052172321422504456911979
900	415873681190459054784114365430
1000	24061467864032622473692149727991

n for which $P(n)$ is PRIME are 2, 3, 4, 5, 6, 13, 36, 77, 132, 157, 168, 186, ... (Sloane's A046063). Numbers which cannot be written as a PRODUCT of $P(n)$ are 13, 17, 19, 23, 26, 29, 31, 34, 37, 38, 39, ... (Sloane's A046064), corresponding to numbers of nonisomorphic ABELIAN GROUPS which are not possible for any group order.

When explicitly listing the partitions of a number n, the simplest form is the so-called *natural representation* which simply gives the sequence of numbers in the representation (e.g., (2, 1, 1) for the number $4=2+1+1$). The *multiplicity representation* instead gives the number of times each number occurs together with that number (e.g., (2, 1), (1, 2) for $4 = 2\cdot 1 + 1\cdot 2$). The FERRERS DIAGRAM is a pictorial representation of a partition.

Euler invented a GENERATING FUNCTION which gives rise to a POWER SERIES in $P(n)$,

$$P(n) = \sum_{m=1}^{\infty} (-1)^{m+1}[P(n-\tfrac{1}{2}m(3m-1)) + P(n-\tfrac{1}{2}m(3m+1))]. \tag{3}$$

A RECURRENCE RELATION is

$$P(n) = \frac{1}{n}\sum_{m=0}^{n-1} \sigma(n-m)P(m), \tag{4}$$

where $\sigma(n)$ is the DIVISOR FUNCTION (Berndt 1994, p. 108). Euler also showed that, for

$$f(x) \equiv \prod_{m=1}^{\infty}(1-x^m) = \sum_{n=-\infty}^{\infty}(-1)^n x^{n(3n+1)/2} \tag{5}$$

$$= 1 - x - x^2 + x^5 + x^7 - x^{12} - x^{15} + x^{22} + x^{26} + \ldots, \tag{6}$$

where the exponents are generalized PENTAGONAL NUMBERS 0, 1, 2, 5, 7, 12, 15, 22, 26, 35, ... (Sloane's A001318) and the sign of the kth term (counting 0 as the 0th term) is $(-1)^{\lfloor (k+1)/2 \rfloor}$ (with $\lfloor x \rfloor$ the FLOOR FUNCTION), the partition numbers $P(n)$ are given by the GENERATING FUNCTION

$$\frac{1}{f(x)} = \sum_{n=0}^{\infty} P(n)x^n. \tag{7}$$

MacMahon obtained the beautiful RECURRENCE RELATION

$$P(n) - P(n-1) - P(n-2) + P(n-5) + P(n-7) - P(n-12) - P(n-15) + \ldots = 0, \tag{8}$$

where the sum is over generalized PENTAGONAL NUMBERS $\leq n$ and the sign of the kth term is $(-1)^{\lfloor (k+1)/2 \rfloor}$, as above.

In 1916–1917, Hardy and Ramanujan used the CIRCLE METHOD and elliptic MODULAR FUNCTIONS to obtain the approximate solution

$$P(n) \sim \frac{1}{4n\sqrt{3}} e^{\pi\sqrt{2n/3}}. \tag{9}$$

Rademacher (1937) subsequently obtained an exact series solution which yields the Hardy-Ramanujan FORMULA (9) as the first term:

$$P(n) = \sum_{q=1}^{\infty} L_q(n)\psi_q(n), \tag{10}$$

where

$$K = \pi\sqrt{\frac{2}{3}} \tag{11}$$

$$L_q(n) = \sum_p \omega_{p,q} e^{-2np\pi i/q} \tag{12}$$

$$\omega_{p,q} = e^{\pi i s_{p,q}} \tag{13}$$

$$s_{p,q} = \frac{1}{q}\sum_{\mu=1}^{q-1}\mu\left(\frac{\mu p}{q} - \left\lfloor \frac{\mu p}{q} \right\rfloor - \frac{1}{2}\right) \tag{14}$$

$$\lambda_n = \sqrt{n - \tfrac{1}{24}} \tag{15}$$

$$\psi_q(n) = \frac{\sqrt{q}}{\pi\sqrt{2}} \left\{ \frac{d}{dm}\left[\frac{\sinh\left(\frac{K\lambda_m}{q}\right)}{\lambda_m}\right]\right\}_{m=n}, \tag{16}$$

$\lfloor x \rfloor$ is the Floor Function, and p runs through the Integers less than and Relatively Prime to q (when $q = 1$, $p = 0$). The remainder after Q terms is

$$R(Q) < CQ^{-1/2} + D\sqrt{\frac{Q}{n}}\sinh\left(\frac{K\sqrt{n}}{Q}\right), \qquad (17)$$

where C and D are fixed constants.

With $f(x)$ as defined above, Ramanujan also showed that

$$5\frac{f^5(x^5)}{f^6(x)} = \sum_{m=0}^{\infty} P(5m+4)x^m. \qquad (18)$$

Ramanujan also found numerous Congruences such as

$$P(5m+4) \equiv 0 \pmod{5} \qquad (19)$$

$$P(7m+5) \equiv 0 \pmod{7} \qquad (20)$$

$$P(11m+6) \equiv 0 \pmod{11}. \qquad (21)$$

Ramanujan's Identity gives the first of these.

Let $P_O(n)$ be the number of partitions of n containing Odd numbers only and $P_D(n)$ be the number of partitions of n without duplication, then

$$P_O(n) = P_D(n) = \prod_{k=1,3,\ldots}^{\infty}(1 + x^k + x^{2k} + x^{3k} + \ldots)$$

$$= \prod_{k=1}^{\infty}(1+x^k) = 1 + x + x^2 + 2x^3 + 2x^4 + 3x^5 + \ldots, \qquad (22)$$

as discovered by Euler (Honsberger 1985). The first few values of $P_O = P_D$ are 1, 1, 1, 2, 2, 3, 4, 5, 6, 8, 10, ... (Sloane's A000009).

Let $P_E(n)$ be the number of partitions of Even numbers only, and let $P_{EO}(n)$ $(P_{DO}(n))$ be the number of partitions in which the parts are all Even (Odd) and all different. The first few values of $P_{DO}(n)$ are 1, 1, 0, 1, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 4, ... (Sloane's A000700). Some additional Generating Functions are given by Honsberger (1985, pp. 241–242)

$$\sum_{n=1}^{\infty} P_{\text{no even part repeated}}(n)x^n$$

$$= \prod_{k=1}(1 - x^{2k-1})^{-1}(1 + x^{2k}) \qquad (23)$$

$$\sum_{n=1}^{\infty} P_{\text{no part occurs more than 3 times}}(n)x^n$$

$$= \prod_{k=1}(1 + x^k + x^{2k} + x^{3k}) \qquad (24)$$

$$\sum_{n=1}^{\infty} P_{\text{no part divisible by 4}}(n)x^n = \prod_{k=1}\frac{1-x^{4k}}{1-x^k} \qquad (25)$$

$$\sum_{n=1}^{\infty} P_{\text{no part occurs more than } d \text{ times}}(n)x^n$$

$$= \prod_{k=1}\sum_{i=0}^{d} x^{ik} = \prod_{k=1}\frac{1-x^{(d+1)k}}{1-x^k} \qquad (26)$$

$$\sum_{n=1}^{\infty} P_{\text{every part occurs 2, 3, or 5 times}}(n)x^n$$

$$= \prod_{k=1}(1 + x^{2k} + x^{3k} + x^{5k})$$

$$= \prod_{k=1}(1 + x^{2k})(1 + x^{3k}) = \prod_{k=1}\frac{1-x^{4k}}{1-x^{2k}}\frac{1-x^{6k}}{1-x^{3k}} \qquad (27)$$

$$\sum_{n=1}^{\infty} P_{\text{no part occurs exactly once}}(n)x^n$$

$$= (1 + x^{2k} + x^{3k} + \ldots) = \prod_{k}\frac{1+x^{6k}}{(1-x^{2k})(1-x^{3k})}. \qquad (28)$$

Some additional interesting theorems following from these (Honsberger 1985, pp. 64–68 and 143–146) are:

1. The number of partitions of n in which no Even part is repeated is the same as the number of partitions of n in which no part occurs more than three times and also the same as the number of partitions in which no part is divisible by four.
2. The number of partitions of n in which no part occurs more often than d times is the same as the number of partitions in which no term is a multiple of $d + 1$.
3. The number of partitions of n in which each part appears either 2, 3, or 5 times is the same as the number of partitions in which each part is Congruent mod 12 to either 2, 3, 6, 9, or 10.
4. The number of partitions of n in which no part appears exactly once is the same as the number of partitions of n in which no part is Congruent to 1 or 5 mod 6.
5. The number of partitions in which the parts are all Even and different is equal to the absolute difference of the number of partitions with Odd and Even parts.

$P(n,k)$, also written $P_k(n)$, is the number of ways of writing n as a sum of k terms, and can be computed from the Recurrence Relation

$$P(n,k) = P(n-1,k-1) + P(n-k,k) \qquad (29)$$

(Ruskey). The number of partitions of n with largest part k is the same as $P(n,k)$.

The function $P(n,k)$ can be given explicitly for the first few values of k,

$$P(n,2) = \left\lfloor \tfrac{1}{2}n \right\rfloor \qquad (30)$$

$$P(n,3) = \left[\tfrac{1}{12}n^2\right], \qquad (31)$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION and $[x]$ is the NINT function (Honsberger 1985, pp. 40–45).

see also ALCUIN'S SEQUENCE, ELDER'S THEOREM, EULER'S PENTAGONAL NUMBER THEOREM, FERRERS DIAGRAM, PARTITION FUNCTION Q, PENTAGONAL NUMBER, $r_k(n)$, ROGERS-RAMANUJAN IDENTITIES, STANLEY'S THEOREM

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Unrestricted Partitions." §24.2.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 825, 1972.

Adler, H. "Partition Identities—From Euler to the Present." *Amer. Math. Monthly* **76**, 733–746, 1969.

Adler, H. "The Use of Generating Functions to Discover and Prove Partition Identities." *Two-Year College Math. J.* **10**, 318–329, 1979.

Andrews, G. *Encyclopedia of Mathematics and Its Applications, Vol. 2: The Theory of Partitions.* Cambridge, England: Cambridge University Press, 1984.

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, 1994.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 94–96, 1996.

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 40–45 and 64–68, 1985.

Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., pp. 237–239, 1991.

Jackson, D. and Goulden, I. *Combinatorial Enumeration.* New York: Academic Press, 1983.

MacMahon, P. A. *Combinatory Analysis.* New York: Chelsea, 1960.

Rademacher, H. "On the Partition Function $p(n)$." *Proc. London Math. Soc.* **43**, 241–254, 1937.

Ruskey, F. "Information of Numerical Partitions." `http://sue.csc.uvic.ca/~cos/inf/nump/NumPartition.html`.

Sloane, N. J. A. Sequences A000009/M0281, A000041/M0663, and A000700/M0217 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Partition Function Q

$Q(n)$ gives the number of ways of writing the INTEGER n as a sum of POSITIVE INTEGERS without regard to order with the constraint that all INTEGERS in each sum are distinct. The values for $n = 1, 2, \ldots$ are 1, 1, 2, 2, 3, 4, 5, 6, 8, 10, ... (Sloane's A000009). The GENERATING FUNCTION for $Q(n)$ is

$$\begin{aligned}\prod_{m=1}^{\infty}(1+x^m) &= \frac{1}{\prod_{m=0}^{\infty}(1-x^{2m+1})}\\ &= 1+x+x^2+2x^3+2x^4+3x^5+\ldots.\end{aligned}$$

The values of n for which $Q(n)$ is PRIME are 3, 4, 5, 7, 22, 70, 100, 495, 1247, 2072, ... (Sloane's A046065), with no others for $n \leq 15{,}000$.

The number of PARTITIONS of n with $\leq k$ summands is denoted $q(n,k)$ or $q_k(n)$. Therefore, $q_n(n) = P(n)$ and

$$q_k(n) = q_{k-1}(n) + q_k(n-k).$$

see also PARTITION FUNCTION P

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Partitions into Distinct Parts." §24.2.2 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 823–824, 1972.

Sloane, N. J. A. Sequences A046065 and A000009/M0281 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Party Problem

Also known as the MAXIMUM CLIQUE PROBLEM. Find the minimum number of guests that must be invited so that at least m will know each other or at least n will not know each other. The solutions are known as RAMSEY NUMBERS.

see also CLIQUE, RAMSEY NUMBER

Parzen Apodization Function

An APODIZATION FUNCTION similar to the BARTLETT FUNCTION.

see also APODIZATION FUNCTION, BARTLETT FUNCTION

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 547, 1992.

Pascal Distribution

see NEGATIVE BINOMIAL DISTRIBUTION

Pascal's Formula

Each subsequent row of PASCAL'S TRIANGLE is obtained by adding the two entries diagonally above. This follows immediately from the BINOMIAL COEFFICIENT identity

$$\begin{aligned}\binom{n}{r} &\equiv \frac{n!}{(n-r)!r!} = \frac{(n-1)!n}{(n-r)!r!}\\ &= \frac{(n-1)!(n-r)}{(n-r)!r!} + \frac{(n-1)!r}{(n-r)!r!}\\ &= \frac{(n-1)!}{(n-r-1)!r!} + \frac{(n-1)!}{(n-r)!(r-1)!}\\ &= \binom{n-1}{r} + \binom{n-1}{r-1}.\end{aligned}$$

see also BINOMIAL COEFFICIENT, PASCAL'S TRIANGLE

Pascal's Hexagrammum Mysticum

see PASCAL'S THEOREM

Pascal's Limaçon

see LIMAÇON

Pascal Line

The line containing the three points of the intersection of the three pairs of opposite sides of two TRIANGLES.

see also PASCAL'S THEOREM

Pascal's Rule

see PASCAL'S FORMULA

Pascal's Theorem

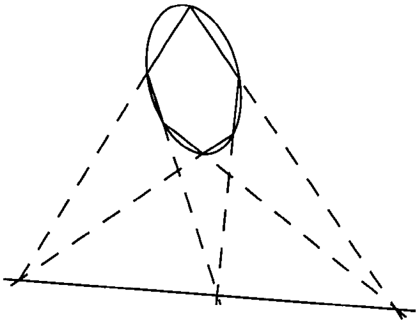

The dual of BRIANCHON'S THEOREM. It states that, given a (not necessarily REGULAR, or even CONVEX) HEXAGON inscribed in a CONIC SECTION, the three pairs of the continuations of opposite sides meet on a straight LINE, called the PASCAL LINE. There are 6! (6! means 6 FACTORIAL, where $6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$) possible ways of taking all VERTICES in any order, but among these are six equivalent CYCLIC PERMUTATIONS and two possible orderings, so the total number of different hexagons (not all simple) is

$$\frac{6!}{2 \cdot 6} = \frac{720}{12} = 60.$$

There are therefore a total of 60 PASCAL LINES created by connecting VERTICES in any order. These intersect three by three in 20 STEINER POINTS.

see also BRAIKENRIDGE-MACLAURIN CONSTRUCTION, BRIANCHON'S THEOREM, CAYLEY-BACHARACH THEOREM, CONIC SECTION, DUALITY PRINCIPLE, HEXAGON, PAPPUS'S HEXAGON THEOREM, PASCAL LINE, STEINER POINTS

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 73–76, 1967.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 105–106, 1990.

Pappas, T. "The Mystic Hexagram." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 118, 1989.

Pascal's Triangle

A TRIANGLE of numbers arranged in staggered rows such that

$$a_{nr} \equiv \frac{n!}{r!(n-r)!} \equiv \binom{n}{r}, \qquad (1)$$

where $\binom{n}{r}$ is a BINOMIAL COEFFICIENT. The triangle was studied by B. Pascal, although it had been described centuries earlier by Chinese mathematician Yanghui (about 500 years earlier, in fact) and the Arabian poet-mathematician Omar Khayyám. It is therefore known as the YANGHUI TRIANGLE in China. Starting with $n = 0$, the TRIANGLE is

$$\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1
\end{array}$$

(Sloane's A007318). PASCAL'S FORMULA shows that each subsequent row is obtained by adding the two entries diagonally above,

$$\binom{n}{r} = \frac{n!}{(n-r)!r!} = \binom{n-1}{r} + \binom{n-1}{r-1}. \qquad (2)$$

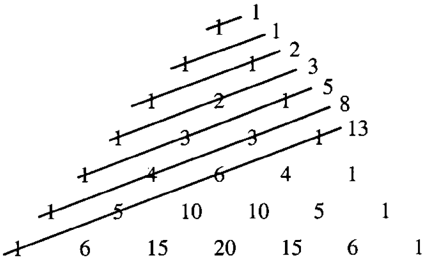

In addition, the "SHALLOW DIAGONALS" of Pascal's triangle sum to FIBONACCI NUMBERS,

$$\sum_{k=1}^{n} \binom{k}{n-k} = \frac{(-1)^n {}_3F_2(1,2,1-n;\tfrac{1}{2}(3-n),2-\tfrac{1}{2}n;-\tfrac{1}{4})}{\pi(2-3n+n^2)} = F_{n+1}, \qquad (3)$$

where ${}_3F_2(a,b,c;d,e;z)$ is a GENERALIZED HYPERGEOMETRIC FUNCTION.

Pascal's triangle contains the FIGURATE NUMBERS along its diagonals. It can be shown that

$$\sum_{i=1}^{n} a_{ij} = \frac{n+1}{j+1} a_{nj} = a_{(n+1),(j+1)} \qquad (4)$$

and

$$\binom{m+1}{1}\sum k^m + \binom{m+1}{2}\sum k^{m-1} + \ldots + \binom{m+1}{m}\sum k = (n+1)[(n+1)^m - 1]. \qquad (5)$$

The "shallow diagonals" sum to the FIBONACCI SEQUENCE, i.e.,

$$\begin{aligned} 1 &= 1 \\ 1 &= 1 \\ 2 &= 1+1 \\ 3 &= 2+1 \\ 5 &= 1+3+1 \\ 8 &= 3+4+1. \end{aligned} \tag{6}$$

In addition,

$$\sum_{j=1}^{i} a_{ij} = 2^i - 1. \tag{7}$$

It is also true that the first number after the 1 in each row divides all other numbers in that row IFF it is a PRIME. If P_n is the number of ODD terms in the first n rows of the Pascal triangle, then

$$0.812\ldots < P_n n^{-\ln 2/\ln 3} < 1 \tag{8}$$

(Harborth 1976, Le Lionnais 1983).

The BINOMIAL COEFFICIENT $\binom{m}{n}$ mod 2 can be computed using the XOR operation n XOR m, making Pascal's triangle mod 2 very easy to construct. Pascal's triangle is unexpectedly connected with the construction of regular POLYGONS and with the SIERPIŃSKI SIEVE.

see also BELL TRIANGLE, BINOMIAL COEFFICIENT, BINOMIAL THEOREM, BRIANCHON'S THEOREM, CATALAN'S TRIANGLE, CLARK'S TRIANGLE, EULER'S TRIANGLE, FIBONACCI NUMBER, FIGURATE NUMBER TRIANGLE, LEIBNIZ HARMONIC TRIANGLE, NUMBER TRIANGLE, PASCAL'S FORMULA, POLYGON, SEIDEL-ENTRINGER-ARNOLD TRIANGLE, SIERPIŃSKI SIEVE, TRINOMIAL TRIANGLE

References

Conway, J. H. and Guy, R. K. "Pascal's Triangle." In *The Book of Numbers.* New York: Springer-Verlag, pp. 68–70, 1996.

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 17, 1996.

Harborth, H. "Number of Odd Binomial Coefficients." *Not. Amer. Math. Soc.* **23**, 4, 1976.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 31, 1983.

Pappas, T. "Pascal's Triangle, the Fibonacci Sequence & Binomial Formula," "Chinese Triangle," and "Probability and Pascal's Triangle." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 40–41 88, and 184–186, 1989.

Sloane, N. J. A. Sequence A007318/M0082 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Smith, D. E. *A Source Book in Mathematics.* New York: Dover, p. 86, 1984.

Pascal's Wager

"God is or He is not... Let us weigh the gain and the loss in choosing... 'God is.' If you gain, you gain all, if you lose, you lose nothing. Wager, then, unhesitatingly, that He is."

Pasch's Axiom

In the plane, if a line intersects one side of a TRIANGLE and misses the three VERTICES, then it must intersect one of the other two sides. This is a special case of the generalized MENELAUS' THEOREM with $n = 3$.

see also HELLY'S THEOREM, MENELAUS' THEOREM, PASCH'S THEOREM

Pasch's Theorem

A theorem stated in 1882 which cannot be derived from EUCLID'S POSTULATES. Given points a, b, c, and d on a LINE, if it is known that the points are ordered as (a, b, c) and (b, c, d), it is also true that (a, b, d).

see also EUCLID'S POSTULATES, LINE, PASCH'S AXIOM

Pass Equivalent

Two KNOTS are pass equivalent if there exists a sequence of pass moves taking one to the other. Every KNOT is either pass equivalent to the UNKNOT or TREFOIL KNOT. These two knots are not pass equivalent to each other, but the ENANTIOMERS of the TREFOIL KNOT are pass equivalent. A KNOT has ARF INVARIANT 0 if the KNOT is pass equivalent to the UNKNOT and 1 if it is pass equivalent to the TREFOIL KNOT.

see also ARF INVARIANT, KNOT, PASS MOVE, TREFOIL KNOT, UNKNOT

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 223–228, 1994.

Pass Move

A change in a knot projection such that a pair of oppositely oriented strands are passed through another pair of oppositely oriented strands.

see also PASS EQUIVALENT

Patch

A patch (also called a LOCAL SURFACE) is a differentiable mapping $\mathbf{x} : U \to \mathbb{R}^n$, where U is an open subset of $\mathbb{R}^2$. More generally, if A is any SUBSET of $\mathbb{R}^2$, then a map $\mathbf{x} : A \to \mathbb{R}^n$ is a patch provided that $\mathbf{x}$ can be extended to a differentiable map from U into $\mathbb{R}^n$, where U is an open set containing A. Here, $\mathbf{x}(U)$ (or more generally, $\mathbf{x}(A)$) is called the TRACE of $\mathbf{x}$.

see also GAUSS MAP, INJECTIVE PATCH, MONGE PATCH, REGULAR PATCH, TRACE (MAP)

References

Gray, A. "Patches in $\mathbb{R}^3$." §10.2 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 183–184 and 192–193, 1993.

Path

A path γ is a continuous mapping $\gamma : [a, b] \mapsto \mathbb{C}$, where $\gamma(a)$ is the initial point and $\gamma(b)$ is the final point. It is often written parametrically as $\sigma(t)$.

Path Graph

The path P_n is a TREE with two nodes of valency 1, and the other $n - 2$ nodes of valency 2. Path graphs P_n are always GRACEFUL for $n > 4$.

see also CHAIN (GRAPH), GRACEFUL GRAPH, HAMILTONIAN PATH, TREE

Path Integral

Let γ be a PATH given parametrically by $\sigma(t)$. Let s denote ARC LENGTH from the initial point. Then

$$\begin{aligned}\int_\gamma f(s)\,ds &= \int_\gamma f(\sigma(t))\,|\sigma'(t)|\,dt\\ &= \int_\gamma f(x(t), y(t), z(t))\,|\sigma'(t)|\,dt.\end{aligned}$$

see also LINE INTEGRAL

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Evaluation of Functions by Path Integration." §5.14 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 201–204, 1992.

Pathwise-Connected

A TOPOLOGICAL SPACE X is pathwise-connected IFF for every two points $x, y \in X$, there is a CONTINUOUS FUNCTION f from [0,1] to X such that $f(0) = x$ and $f(1) = y$. Roughly speaking, a SPACE X is pathwise-connected if, for every two points in X, there is a path connecting them. For LOCALLY PATHWISE-CONNECTED SPACES (which include most "interesting spaces" such as MANIFOLDS and CW-COMPLEXES), being CONNECTED and being pathwise-connected are equivalent, although there are connected spaces which are not pathwise connected. Pathwise-connected spaces are also called 0-connected.

see also CONNECTED SPACE, CW-COMPLEX, LOCALLY PATHWISE-CONNECTED SPACE, TOPOLOGICAL SPACE

Patriarchal Cross

see GAULLIST CROSS

Pauli Matrices

Matrices which arise in Pauli's treatment of spin in quantum mechanics. They are defined by

$$\sigma_1 = \sigma_x \equiv \mathsf{P}_1 \equiv \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{1}$$

$$\sigma_2 = \sigma_y \equiv \mathsf{P}_2 \equiv \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix} \tag{2}$$

$$\sigma_3 = \sigma_z \equiv \mathsf{P}_3 \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{3}$$

The Pauli matrices plus the 2×2 IDENTITY MATRIX I form a complete set, so any 2×2 matrix A can be expressed as

$$\mathsf{A} = c_0\mathsf{I} + c_1\sigma_1 + c_2\sigma_2 + c_3\sigma_3. \tag{4}$$

The associated matrices

$$\sigma_+ \equiv 2\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \tag{5}$$

$$\sigma_- \equiv 2\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \tag{6}$$

$$\sigma^2 \equiv 3\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{7}$$

can also be defined. The Pauli spin matrices satisfy the identities

$$\sigma_i\sigma_j = \mathsf{I}\delta_{ij} + \epsilon_{ijk}i\sigma_k \tag{8}$$

$$\sigma_i\sigma_j + \sigma_j\sigma_i = 2\sigma_{ij} \tag{9}$$

$$\sigma_x p_x + \sigma_y p_y + \sigma_z p_z = \sqrt{{p_x}^2 + {p_y}^2 + {p_z}^2}. \tag{10}$$

see also DIRAC MATRICES, QUATERNION

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 211–212, 1985.

Goldstein, H. "The Cayley-Klein Parameters and Related Quantities." *Classical Mechanics, 2nd ed.* Reading, MA: Addison-Wesley, p. 156, 1980.

Pauli Spin Matrices

see PAULI MATRICES

Payoff Matrix

A $m \times n$ MATRIX which gives the possible outcome of a two-person ZERO-SUM GAME when player A has m possible strategies and player B n strategies. The analysis of the MATRIX in order to determine optimal strategies is the aim of GAME THEORY. The so-called "augmented" payoff matrix is defined as follows:

$$\mathsf{G} = \begin{array}{c} \begin{array}{cccccccc} P_0 & P_1 & P_2 & \cdots & P_n & P_{n+1} & P_{n+2} & \cdots & P_{n+m} \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ -1 & a_{11} & a_{12} & \cdots & a_{1n} & 1 & 0 & \cdots & 0 \\ -1 & a_{21} & a_{22} & \cdots & a_{2n} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & a_{m1} & a_{m2} & \cdots & a_{mn} & 0 & 0 & \cdots & 1 \end{bmatrix} \end{array}.$$

see also GAME THEORY, ZERO-SUM GAME

Peacock's Tail

One name for the figure used by Euclid to prove the PYTHAGOREAN THEOREM.

see also BRIDE'S CHAIR, WINDMILL

Peano Arithmetic

The theory of NATURAL NUMBERS defined by the five PEANO'S AXIOMS. Any universal statement which is undecidable in Peano arithmetic is necessarily TRUE. Undecidable statements may be either TRUE or FALSE. Paris and Harrington (1977) gave the first "natural" example of a statement which is true for the integers but unprovable in Peano arithmetic (Spencer 1983).

see also KREISEL CONJECTURE, NATURAL INDEPENDENCE PHENOMENON, NUMBER THEORY, PEANO'S AXIOMS

References

Kirby, L. and Paris, J. "Accessible Independence Results for Peano Arithmetic." *Bull. London Math. Soc.* **14**, 285–293, 1982.

Paris, J. and Harrington, L. "A Mathematical Incompleteness in Peano Arithmetic." In *Handbook of Mathematical Logic* (Ed. J. Barwise). Amsterdam, Netherlands: North-Holland, pp. 1133–1142, 1977.

Spencer, J. "Large Numbers and Unprovable Theorems." *Amer. Math. Monthly* **90**, 669–675, 1983.

Peano's Axioms

1. Zero is a number.
2. If a is a number, the successor of a is a number.
3. ZERO is not the successor of a number.
4. Two numbers of which the successors are equal are themselves equal.
5. (INDUCTION AXIOM.) If a set S of numbers contains ZERO and also the successor of every number in S, then every number is in S.

Peano's axioms are the basis for the version of NUMBER THEORY known as PEANO ARITHMETIC.

see also INDUCTION AXIOM, PEANO ARITHMETIC

Peano Curve

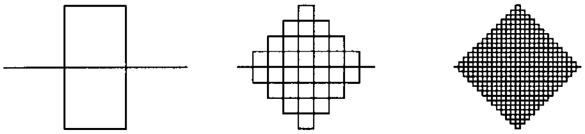

A FRACTAL curve which can be written as a LINDENMAYER SYSTEM.

see also DRAGON CURVE, HILBERT CURVE, LINDENMAYER SYSTEM, SIERPIŃSKI CURVE

References

Dickau, R. M. "Two-Dimensional L-Systems." `http://forum.swarthmore.edu/advanced/robertd/lsys2d.html`.

Hilbert, D. "Über die stetige Abbildung einer Linie auf ein Flachenstück." *Math. Ann.* **38**, 459–460, 1891.

Peano, G. "Sur une courbe, qui remplit une aire plane." *Math. Ann.* **36**, 157–160, 1890.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, p. 207, 1991.

Peano-Gosper Curve

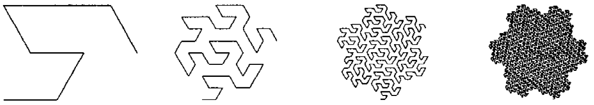

A PLANE-FILLING CURVE originally called a FLOWSNAKE by R. W. Gosper and M. Gardner. Mandelbrot (1977) subsequently coined the name Peano-Gosper curve. The GOSPER ISLAND bounds the space that the Peano-Gosper curve fills.

see also DRAGON CURVE, EXTERIOR SNOWFLAKE, GOSPER ISLAND, HILBERT CURVE, KOCH SNOWFLAKE, PEANO CURVE, SIERPIŃSKI ARROWHEAD CURVE, SIERPIŃSKI CURVE

References

Dickau, R. M. "Two-Dimensional L-Systems." `http://forum.swarthmore.edu/advanced/robertd/lsys2d.html`.

Mandelbrot, B. B. *Fractals: Form, Chance, & Dimension.* San Francisco, CA: W. H. Freeman, 1977.

Peano Surface

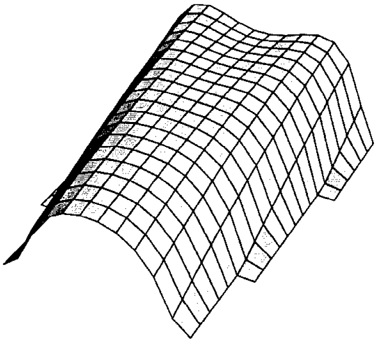

The function

$$f(x,y) = (2x^2 - y)(y - x^2)$$

which does *not* have a LOCAL MAXIMUM at (0, 0), despite criteria commonly touted in the second half of the 1800s which indicated the contrary.

see also LOCAL MAXIMUM

References

Fischer, G. (Ed.). Plate 122 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 119, 1986.

Leitere, J. "Functions." §7.1.2 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 70–71, 1986.

Pear Curve

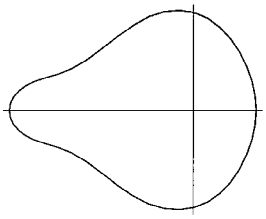

The Lemniscate L_3 in the iteration towards the Mandelbrot Set. In Cartesian Coordinates with a constant r, the equation is given by

$$\begin{aligned} r^2 = (x^2+y^2)(1+2x+5x^2+6x^3+6x^4+4x^5+x^6-3y^2 \\ -2xy^2+8x^2y^2+8x^3y^2+3x^4y^2+2y^4+4xy^4 \\ +3x^2y^4+y^6). \end{aligned}$$

see also Pear-Shaped Curve

Pear-Shaped Curve

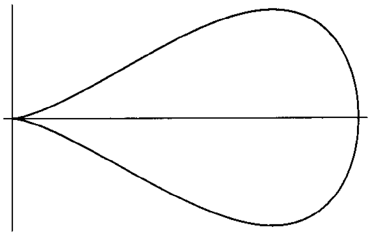

A curve given by the Cartesian equation

$$b^2y^2 = x^3(a-x).$$

see also Pear Curve, Teardrop Curve

References

MacTutor History of Mathematics Archive. "Pear-Shaped Cubic." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Pearshaped.html`.

Pearson's Correlation

see Correlation Coefficient

Pearson-Cunningham Function

see Cunningham Function

Pearson's Function

$$I\left(\frac{{\chi_s}^2}{\sqrt{2(k-1)}}, \frac{k-3}{2}\right) \equiv \frac{\Gamma\left(\frac{1}{2}{\chi_s}^2, \frac{k-1}{2}\right)}{\Gamma\left(\frac{k-1}{2}\right)},$$

where $\Gamma(x)$ is the Gamma Function.

see also Chi-Squared Test, Gamma Function

Pearson Kurtosis

Let μ_4 be the fourth Moment of a Distribution and σ its Variance. Then the Pearson kurtosis is defined by

$$\beta_2 \equiv \frac{\mu_4}{\sigma^4}.$$

see also Fisher Kurtosis, Kurtosis

Pearson Mode Skewness

Given a Distribution with measured Mean, Mode, and Standard Deviation s, the Pearson mode skewness is

$$\frac{\text{mean} - \text{mode}}{s}.$$

see also Mean, Mode, Pearson Skewness, Pearson's Skewness Coefficients, Skewness

Pearson Skewness

Let a Distribution have third Moment μ_3 and Standard Deviation σ, then the Pearson skewness is defined by

$$\beta_1 = \left(\frac{\mu_3}{\sigma^3}\right)^2.$$

see also Fisher Skewness, Pearson's Skewness Coefficients, Skewness

Pearson's Skewness Coefficients

Given a Distribution with measured Mean, Median, Mode, and Standard Deviation s, Pearson's first skewness coefficient is

$$\frac{3[\text{mean}] - [\text{mode}]}{s},$$

and the second coefficient is

$$\frac{3[\text{mean}] - [\text{median}]}{s}.$$

see also Fisher Skewness, Pearson Skewness, Skewness

Pearson System

Generalizes the differential equation for the Gaussian Distribution

$$\frac{dy}{dx} = \frac{y(m-x)}{a} \tag{1}$$

to

$$\frac{dy}{dx} = \frac{y(m-x)}{a+bx+cx^2}. \tag{2}$$

Let c_1, c_2 be the roots of $a+bx+cx^2$. Then the possible types of curves are

0. $b=c=0$, $a>0$. E.g., Normal Distribution.
I. $b^2/4ac<0$, $c_1 \le x \le c_2$. E.g., Beta Distribution.
II. $b^2/4ac=0$, $c<0$, $-c_1 \le x \le c_1$ where $c_1 \equiv \sqrt{-c/a}$.
III. $b^2/4ac=\infty$, $c=0$, $c_1 \le x < \infty$ where $c_1 \equiv -a/b$. E.g., Gamma Distribution. This case is intermediate to cases I and VI.
IV. $0 < b^2/4ac < 1$, $-\infty < x < \infty$.
V. $b^2/4ac=1$, $c_1 \le x < \infty$ where $c_1 \equiv -b/2a$. Intermediate to cases IV and VI.

VI. $b^2/4ac > 1$, $c_1 \leq x < \infty$ where c_1 is the larger root. E.g., BETA PRIME DISTRIBUTION.

VII. $b^2/4ac = 0$, $c > 0$, $-\infty < x < \infty$. E.g., STUDENT'S t-DISTRIBUTION.

Classes IX–XII are discussed in Pearson (1916). See also Craig (in Kenney and Keeping 1951). If a Pearson curve possesses a MODE, it will be at $x = m$. Let $y(x) = 0$ at c_1 and c_2, where these may be $-\infty$ or ∞. If yx^{r+2} also vanishes at c_1, c_2, then the rth MOMENT and $(r+1)$th MOMENTS exist.

$$\int_{c_1}^{c_2} \frac{dy}{dx}(ax^r+bx^{r+1}+cx^{r+2})\,dx = \int_{c_1}^{c_2} y(mx^r-x^{r+1})\,dx \quad (3)$$

giving

$$[y(ax^r+bx^{r+1}+cx^{r+2})]_{c_1}^{c_2} - \int_{c_1}^{c_2} y[arx^{r-1}+b(r+1)x^r+c(r+2)x^{r+1}]\,dx = \int_{c_1}^{c_2} y(mx^r-x^{r+1})\,dx \quad (4)$$

$$0 - \int_{c_1}^{c_2} y[arx^{r-1}+b(r+1)x^r+c(r+2)x^{r+1}]\,dx = \int_{c_1}^{c_2} y(mx^r-x^{r+1})\,dx \quad (5)$$

also,

$$\nu_r = \int_{c_1}^{c_2} yx^r\,dx, \quad (6)$$

so

$$ar\nu_{r-1} + b(r+1)\nu_r + c(r+2)\nu_{r+1} = -m\nu_r + \nu_{r+1}. \quad (7)$$

For $r = 0$,

$$b + 2c\nu_1 = -m + \nu_1, \quad (8)$$

so

$$\nu_1 = \frac{m+b}{1-2c}. \quad (9)$$

For $r = 1$,

$$a + 2b\nu_1 + 3c\nu_2 = -m\nu_1 + \nu_2, \quad (10)$$

so

$$\nu_2 = \frac{a+(m+2b)\nu_1}{1-3c}. \quad (11)$$

Now let $t \equiv (x-\nu_1)/\sigma$. Then

$$\nu_1 = 0 \quad (12)$$

$$\nu_2 = \mu_2 = 1 \quad (13)$$

$$\alpha_r = \mu_r = \nu_r. \quad (14)$$

Hence $b = -m$, and $a = 1 - c$ so

$$(1-3c)r\alpha_{r-1} - mr\alpha_r + [c(r+2)-1]\alpha_{r+1} = 0. \quad (15)$$

For $r = 2$,

$$2m + (1-4c)\alpha_3 = 0. \quad (16)$$

For $r = 3$,

$$3(1-3c) - 3m\alpha_3 - (1-5c)\alpha_4 = 0. \quad (17)$$

So the SKEWNESS and KURTOSIS are

$$\gamma_1 = \alpha_3 = \frac{2m}{4c-1} \quad (18)$$

$$\gamma_2 = \alpha_4 - 3 = \frac{6(m^2-4c^2+c)}{(4c-1)(5c-1)}. \quad (19)$$

So the parameters a, b, and c can be written

$$a = 1 - 3c \quad (20)$$

$$b = -m = \frac{\gamma_1}{2(1+2\delta)} \quad (21)$$

$$c = \frac{\delta}{2(1+2\delta)}, \quad (22)$$

where

$$\delta \equiv \frac{2\gamma_2 - 3\gamma_1{}^2}{\gamma_2 + 6}. \quad (23)$$

References

Craig, C. C. "A New Exposition and Chart for the Pearson System of Frequency Curves." *Ann. Math. Stat.* **7**, 16–28, 1936.

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, p. 107, 1951.

Pearson, K. "Second Supplement to a Memoir on Skew Variation." *Phil. Trans. A* **216**, 429–457, 1916.

Pearson Type III Distribution

A skewed distribution which is similar to the BINOMIAL DISTRIBUTION when $p \neq q$ (Abramowitz and Stegun 1972, p. 930).

$$y = k(t+A)^{A^2-1}e^{-At}, \quad (1)$$

for $t \in [0, \infty)$ where

$$A \equiv 2/\gamma \quad (2)$$

$$K \equiv \frac{A^{A^2}e^{-A^2}}{\Gamma(A^2)}, \quad (3)$$

$\Gamma(z)$ is the GAMMA FUNCTION, and t is a standardized variate. Another form is

$$P(x) = \frac{1}{\beta\Gamma(p)}\left(\frac{x-\alpha}{\beta}\right)^{p-1} \exp\left(\frac{x-\alpha}{\beta}\right). \quad (4)$$

For this distribution, the CHARACTERISTIC FUNCTION is

$$\phi(t) = e^{i\alpha t}(1 - i\beta t)^{-p}, \tag{5}$$

and the MEAN, VARIANCE, SKEWNESS, and KURTOSIS are

$$\mu = \alpha + p\beta \tag{6}$$
$$\sigma^2 = p\beta^2 \tag{7}$$
$$\gamma_1 = \frac{2}{\sqrt{p}} \tag{8}$$
$$\gamma_2 = \frac{6}{p}. \tag{9}$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, 1972.

Pearls of Sluze

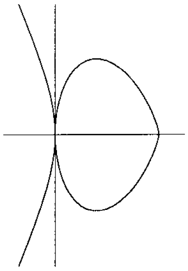

$$y^m = kx^n(a - x)^b.$$

The curves with integral n, p, and m were studied by de Sluze between 1657 and 1698. The name "Pearls of Sluze" was given to these curves by Blaise Pascal (MacTutor Archive).

References

MacTutor History of Mathematics Archive. "Pearls of Sluze." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Pearls.html`.

Peaucellier Cell

see PEAUCELLIER INVERSOR

Peaucellier Inversor

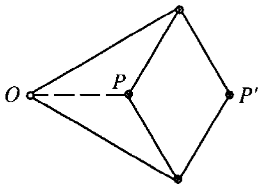

A LINKAGE with six rods which draws the inverse of a given curve. When a pencil is placed at P, the inverse is drawn at P' (or vice versa). If a seventh rod (dashed) is added (with an additional pivot), P is kept on a circle and the locus traced out by P' is a straight line. It therefore converts circular motion to linear motion without sliding, and was discovered in 1864. Another LINKAGE which performs this feat using hinged squares had been published by Sarrus in 1853 but ignored. Coxeter (1969, p. 428) shows that

$$OP \times OP' = OA^2 - PA^2.$$

see also HART'S INVERSOR, LINKAGE

References

Bogomolny, A. "Peaucellier Linkage." `http://www.cut-the-knot.com/pythagoras/invert.html`.

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods.* Oxford, England: Oxford University Press, p. 156, 1978.

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 82–83, 1969.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 46–48, 1990.

Rademacher, H. and Toeplitz, O. *The Enjoyment of Mathematics: Selections from Mathematics for the Amateur.* Princeton, NJ: Princeton University Press, pp. 121–126, 1957.

Smith, D. E. *A Source Book in Mathematics.* New York: Dover, p. 324, 1994.

Peaucellier's Linkage

see PEAUCELLIER INVERSOR

Pedal

The pedal of a curve with respect to a point P is the locus of the foot of the PERPENDICULAR from P to the TANGENT to the curve. When a CLOSED CURVE rolls on a straight line, the AREA between the line and ROULETTE after a complete revolution by any point on the curve is twice the AREA of the pedal (taken with respect to the generating point) of the rolling curve.

Pedal Circle

The pedal CIRCLE of a point P in a TRIANGLE is the CIRCLE through the feet of the perpendiculars from P to the sides of the TRIANGLE (the CIRCUMCIRCLE about the PEDAL TRIANGLE). When P is on a side of the TRIANGLE, the line between the two perpendiculars is called the PEDAL LINE. Given four points, no three of which are COLLINEAR, then the four PEDAL CIRCLES of each point for the TRIANGLE formed by the other three have a common point through which the NINE-POINT CIRCLES of the four TRIANGLES pass. The radius of the pedal circle of a point P is

$$r = \frac{\overline{A_1P}\cdot\overline{A_2P}\cdot\overline{A_3P}}{2(R^2 - \overline{OP}^2)}$$

(Johnson 1929, p. 141).

see also MIQUEL POINT, NINE-POINT CIRCLE, PEDAL TRIANGLE

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, 1929.

Pedal Coordinates

The pedal coordinates of a point P with respect to the curve C and the PEDAL POINT O are the radial distant r from O to P and the PERPENDICULAR distance p from O to the line L tangent to C at P.

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 2–3, 1972.

Yates, R. C. "Pedal Equations." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 166–169, 1952.

Pedal Curve

Given a curve C, the pedal curve of C with respect to a fixed point O (the PEDAL POINT) is the locus of the point P of intersection of the PERPENDICULAR from O to a TANGENT to C. The parametric equations for a curve $(f(t), g(t))$ relative to the PEDAL POINT (x_0, y_0) are

$$x = \frac{x_0 f'^2 + f g'^2 + (y_0 - g) f' g'}{f'^2 + g'^2}$$

$$y = \frac{g f'^2 + y_0 g'^2 + (x_0 - f) f' g'}{f'^2 + g'2^2}.$$

Curve	Pole	Pedal
astroid	center	quadrifolium
cardioid	cusp	Cayley's sextic
central conic	focus	circle
circle	any point	limaçon
circle	on circumference	cardioid
circle involute	center of circle	Archimedean spiral
cissoid of Diocles	focus	cardioid
deltoid	center	trifolium
deltoid	cusp	simple folium
deltoid	on the curve	unsymmetrical double folium
deltoid	vertex	double folium
epicycloid	center	rose
hypocycloid	center	rose
line	any point	point
logarithmic spiral	pole	logarithmic spiral
parabola	focus	line
parabola	foot of directrix	right strophoid
parabola	on directrix	strophoid
parabola	refl. of focus by dir.	Maclaurin trisectrix
parabola	vertex	cissoid of Diocles
sinusoidal spiral	pole	sinusoidal spiral
Tschirnhausen cubic	focus of pedal	parabola

see also NEGATIVE PEDAL CURVE

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 46–49 and 204, 1972.

Lee, X. "Pedal." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Pedal_dir/pedal.html`.

Lockwood, E. H. "Pedal Curves." Ch. 18 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 152–155, 1967.

Yates, R. C. "Pedal Curves." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 160–165, 1952.

Pedal Line

Mark a point P on a side of a TRIANGLE and draw the perpendiculars from the point to the two other sides. The line between the feet of these two perpendiculars is called the pedal line.

see also PEDAL TRIANGLE, SIMSON LINE

Pedal Point

The fixed point with respect to which a PEDAL CURVE is drawn.

Pedal Triangle

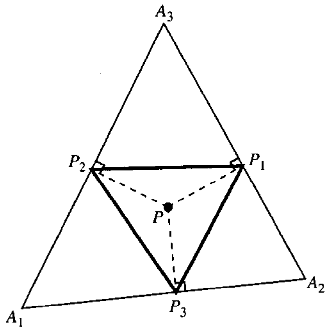

Given a point P, the pedal triangle of P is the TRIANGLE whose VERTICES are the feet of the perpendiculars from P to the side lines. The pedal triangle of a TRIANGLE with TRILINEAR COORDINATES $\alpha : \beta : \gamma$ and angles A, B, and C has VERTICES with TRILINEAR COORDINATES

$$0 : \beta + \alpha \cos C : \gamma + \alpha \cos B \tag{1}$$

$$\alpha + \beta \cos C : 0 : \gamma + \beta \cos A \tag{2}$$

$$\alpha + \gamma \cos B : \beta + \gamma \cos A : 0. \tag{3}$$

The third pedal triangle is similar to the original one. This theorem can be generalized to: the nth pedal n-gon of any n-gon is similar to the original one. It is also true that

$$P_2 P_3 = A_1 P \sin \alpha_1 \tag{4}$$

(Johnson 1929, pp. 135–136). The AREA A of the pedal triangle of a point P is proportional to the POWER of P with respect to the CIRCUMCIRCLE,

$$A = \tfrac{1}{2}(R^2 - \overline{OP}^2) \sin \alpha_1 \sin \alpha_2 \sin \alpha_3 = \frac{R^2 - \overline{OP}^2}{4R^2} \Delta \tag{5}$$

(Johnson 1929, pp. 139–141).

see also ANTIPEDAL TRIANGLE, FAGNANO'S PROBLEM, PEDAL CIRCLE, PEDAL LINE, SCHWARZ'S TRIANGLE PROBLEM

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 22–26, 1967.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, 1929.

Peg Knot

see Clove Hitch

Peg Solitaire

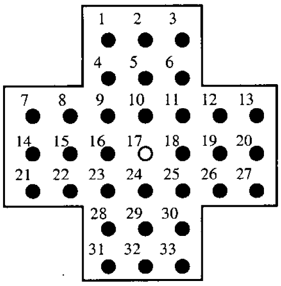

A game played on a cross-shaped board with 33 holes. All holes but the middle one are initially filled with pegs. The goal is to remove all pegs but one by jumping pegs from one side of an occupied peg hole to an empty space, removing the peg which was jumped over. Strategies and symmetries are discussed in Beeler *et al.* (1972, Item 75). A triangular version called Hi-Q also exists (Beeler *et al.* 1972, Item 76). Kraitchik (1942) considers a board with one additional hole placed at the vertices of the central right angles.

see also Hi-Q

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Gardner, M. "Peg Solitaire." Ch. 11 in *The Unexpected Hanging and Other Mathematical Diversions.* New York: Simon and Schuster, pp. 122–135 and 250–251, 1969.

Kraitchik, M. "Peg Solitaire." §12.19 in *Mathematical Recreations.* New York: W. W. Norton, pp. 297–298, 1942.

Peg Top

see Piriform

Peirce's Theorem

The only linear associative algebra in which the coordinates are Real Numbers and products vanish only if one factor is zero are the Field of Real Numbers, the Field of Complex Numbers, and the algebra of Quaternions with Real Coefficients.

see also Weierstraß's Theorem

Pell Equation

A special case of the quadratic Diophantine Equation having the form

$$x^2 - Dy^2 = 1, \tag{1}$$

where D is a nonsquare Natural Number. Dörrie (1965) defines the equation as

$$x^2 - Dy^2 = 4 \tag{2}$$

and calls it the Fermat Difference Equation. The general Pell equation was solved by the Indian mathematician Bhaskara.

Pell equations, as well as the analogous equation with a minus sign on the right, can be solved by finding the Continued Fraction $[a_1, a_2, \ldots]$ for $\sqrt{D}$. (The trivial solution $x = 1$, $y = 0$ is ignored in all subsequent discussion.) Let p_n/q_n denote the nth Convergent $[a_1, a_2, \ldots, a_n]$, then we are looking for a convergent which obeys the identity

$$p_n{}^2 - Dq_n{}^2 = (-1)^n, \tag{3}$$

which turns out to always be possible since the Continued Fraction of a Quadratic Surd always becomes periodic at some term a_{r+1}, where $a_{r+1} = 2a_1$, i.e.,

$$\sqrt{D} = [a_1, \overline{a_2, \ldots, a_r, 2a_1}]. \tag{4}$$

Writing $n = rk$ gives

$$p_{rk}{}^2 - Dq_{rk}{}^2 = (-1)^{rk}, \tag{5}$$

for k anPositive Integer. If r is Odd, solutions to

$$x^2 - Dy^2 = \pm 1 \tag{6}$$

can be obtained if k is chosen to be Even or Odd, but if r is Even, there are no values of k which can make the exponent Odd.

If r is Even, then $(-1)^r$ is Positive and the solution in terms of smallest Integers is $x = p_r$ and $y = q_r$, where p_r/q_r is the rth Convergent. If r is Odd, then $(-1)^r$ is Negative, but we can take $k = 2$ in this case, to obtain

$$p_{2r}{}^2 - Dq_{2r}{}^2 = 1, \tag{7}$$

so the solution in smallest Integers is $x = p_{2r}$, $y = q_{2r}$. Summarizing,

$$(x, y) = \begin{cases} (p_r, q_r) & \text{for } r \text{ even} \\ (p_{2r}, p_{2r}) & \text{for } r \text{ odd.} \end{cases} \tag{8}$$

Given one solution $(x, y) = (p, q)$ (which can be found as above), a whole family of solutions can be found by taking each side to the nth Power,

$$x^2 - Dy^2 = (p^2 - Dq^2)^n = 1. \tag{9}$$

Factoring gives

$$(x + \sqrt{D}\,y)(x - \sqrt{D}\,y) = (p + \sqrt{D}\,q)^n (p - \sqrt{D}\,q)^n \tag{10}$$

and

$$x + \sqrt{D}\,y = (p + \sqrt{D}\,q)^n \tag{11}$$

$$x - \sqrt{D}\,y = (p - \sqrt{D}\,q)^n, \tag{12}$$

which gives the family of solutions

$$x = \frac{(p+q\sqrt{D}\,)^n + (p-q\sqrt{D}\,)^n}{2} \qquad (13)$$

$$y = \frac{(p+q\sqrt{D}\,)^n - (p-q\sqrt{D}\,)^n}{2\sqrt{D}}. \qquad (14)$$

These solutions also hold for

$$x^2 - Dy^2 = -1, \qquad (15)$$

except that n can take on only ODD values.

The following table gives the smallest integer solutions (x, y) to the Pell equation with constant $D \leq 102$ (Beiler 1966, p. 254). SQUARE $D = d^2$ are not included, since they would result in an equation of the form

$$x^2 - d^2y^2 = x^2 - (dy)^2 = x^2 - y'^2 = 1, \qquad (16)$$

which has no solutions (since the difference of two SQUARES cannot be 1).

D	x	y	D	x	y
2	3	2	54	485	66
3	2	1	55	89	12
5	9	4	56	15	2
6	5	2	57	151	20
7	8	3	58	19603	2574
8	3	1	59	530	69
10	19	6	60	31	4
11	10	3	61	1766319049	226153980
12	7	2	62	63	8
13	649	180	63	8	1
14	15	4	65	129	16
15	4	1	66	65	8
17	33	8	67	48842	5967
18	17	4	68	33	4
19	170	39	69	7775	936
20	9	2	70	251	30
21	55	12	71	3480	413
22	197	42	72	17	2
23	24	5	73	2281249	267000
24	5	1	74	3699	430
26	51	10	75	26	3
27	26	5	76	57799	6630
28	127	24	77	351	40
29	9801	1820	78	53	6
30	11	2	79	80	9
31	1520	273	80	9	1
32	17	3	82	163	18
33	23	4	83	82	9
34	35	6	84	55	6
35	6	1	85	285769	30996
37	73	12	86	10405	1122
38	37	6	87	28	3
39	25	4	88	197	21
40	19	3	89	500001	53000
41	2049	320	90	19	2
42	13	2	91	1574	165
43	3482	531	92	1151	120
44	199	30	93	12151	1260
45	161	24	94	2143295	221064
46	24335	3588	95	39	4
47	48	7	96	49	5
48	7	1	97	62809633	6377352
50	99	14	98	99	10
51	50	7	99	10	1
52	649	90	101	201	20
53	66249	9100	102	101	10

The first few minimal values of x and y for nonsquare D are 3, 2, 9, 5, 8, 3, 19, 10, 7, 649, ... (Sloane's A033313) and 2, 1, 4, 2, 3, 1, 6, 3, 2, 180, ... (Sloane's A033317), respectively. The values of D having $x = 2$, 3, ... are 3, 2, 15, 6, 35, 12, 7, 5, 11, 30, ... (Sloane's A033314) and the values of D having $y = 1$, 2, ... are 3, 2, 7, 5, 23, 10, 47, 17, 79, 26, ... (Sloane's A033318). Values of the incrementally largest minimal x are 3, 9, 19, 649, 9801, 24335, 66249, ... (Sloane's A033315) which occur at $D = 2$, 5, 10, 13, 29, 46, 53, 61, 109, 181, ... (Sloane's A033316). Values of the incrementally largest minimal

y are 2, 4, 6, 180, 1820, 3588, 9100, 226153980, ... (Sloane's A033319), which occur at $D = 2$, 5, 10, 13, 29, 46, 53, 61, ... (Sloane's A033320).

see also DIOPHANTINE EQUATION, DIOPHANTINE EQUATION—QUADRATIC, LAGRANGE NUMBER (DIOPHANTINE EQUATION)

References
Beiler, A. H. "The Pellian." Ch. 22 in *Recreations in the Theory of Numbers: The Queen of Mathematics Entertains.* New York: Dover, pp. 248–268, 1966.
Degan, C. F. *Canon Pellianus.* Copenhagen, Denmark, 1817.
Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, 1965.
Lagarias, J. C. "On the Computational Complexity of Determining the Solvability or Unsolvability of the Equation $X^2-DY^2=-1$." *Trans. Amer. Math. Soc.* **260**, 485–508, 1980.
Smarandache, F. "Un metodo de resolucion de la ecuacion diofantica." *Gaz. Math.* **1**, 151–157, 1988.
Smarandache, F. " Method to Solve the Diophantine Equation $ax^2 - by^2 + c = 0$." In *Collected Papers, Vol. 1.* Lupton, AZ: Erhus University Press, 1996.
Stillwell, J. C. *Mathematics and Its History.* New York: Springer-Verlag, 1989.
Whitford, E. E. *Pell Equation.* New York: Columbia University Press, 1912.

Pell-Lucas Number

see PELL NUMBER

Pell-Lucas Polynomial

see PELL POLYNOMIAL

Pell Number

The numbers obtained by the U_ns in the LUCAS SEQUENCE with $P = 2$ and $Q = -1$. They and the Pell-Lucas numbers (the V_ns in the LUCAS SEQUENCE) satisfy the recurrence relation

$$P_n = 2P_{n-1} + P_{n-2}. \tag{1}$$

Using P_i to denote a Pell number and Q_i to denote a Pell-Lucas number,

$$P_{m+n} = P_m P_{n+1} + P_{m-1} P_n \tag{2}$$

$$P_{m+n} = 2P_m Q_n - (-1)^n P_{m-n}, \tag{3}$$

$$P_{2^t m} = P_m(2Q_m)(2Q_{2m})(2Q_{4m})\cdots(2Q_{2^{t-1}m}) \tag{4}$$

$$Q_m{}^2 = 2P_m{}^2 + (-1)^m \tag{5}$$

$$Q_{2m} = 2Q_m{}^2 - (-1)^m. \tag{6}$$

The Pell numbers have $P_0 = 0$ and $P_1 = 1$ and are 0, 1, 2, 5, 12, 29, 70, 169, 408, 985, 2378, ... (Sloane's A000129). The Pell-Lucas numbers have $Q_0 = 2$ and $Q_1 = 2$ and are 2, 2, 6, 14, 34, 82, 198, 478, 1154, 2786, 6726, ... (Sloane's A002203).

The only TRIANGULAR Pell number is 1 (McDaniel 1996).

see also BRAHMAGUPTA POLYNOMIAL, PELL POLYNOMIAL

References
McDaniel, W. L. "Triangular Numbers in the Pell Sequence." *Fib. Quart.* **34**, 105–107, 1996.
Sloane, N. J. A. Sequences A000129/M1413 and A002203/M0360 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Pell Polynomial

The Pell polynomials $P(x)$ and Lucas-Pell polynomials $Q(x)$ are generated by a LUCAS POLYNOMIAL SEQUENCE using generator $(2x, 1)$. This gives recursive equations for $P(x)$ from $P_0(x) = P_1(x) = 1$ and

$$P_{n+2}(x) = 2xP_{n+1}(x) + P_n(x). \tag{1}$$

The first few are

$$\begin{aligned} P_1 &= 1 \\ P_2 &= 2x \\ P_3 &= 4x^2 - 1 \\ P_4 &= 8x^3 - 4x \\ P_5 &= 16x^4 - 12x^2 + 1. \end{aligned}$$

The Pell-Lucas numbers are defined recursively by $q_0(x) = 1$, $q_1(x) = x$ and

$$q_{n+2}(x) = 2xq_{n+1}(x) + q_n(x), \tag{2}$$

together with

$$Q_n(x) \equiv 2q_n(x). \tag{3}$$

The first few are

$$\begin{aligned} Q_1 &= 2x \\ Q_2 &= 4x^2 - 2 \\ Q_3 &= 8x^3 - 6x \\ Q_4 &= 16x^4 - 16x^2 + 2 \\ Q_5 &= 32x^5 - 40x^3 + 10x. \end{aligned}$$

see also LUCAS POLYNOMIAL SEQUENCE

References
Horadam, A. F. and Mahon, J. M. "Pell and Pell-Lucas Polynomials." *Fib. Quart.* **23**, 7–20, 1985.
Mahon, J. M. M. A. (Honors) thesis, The University of New England. Armidale, Australia, 1984.
Sloane, N. J. A. Sequence A000129/M1413 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Pell Sequence

see PELL NUMBER

Pencil

The set of all LINES through a point. Woods (1961), however, uses this term as a synonym for RANGE.

see also NEAR-PENCIL, PERSPECTIVITY, RANGE (LINE SEGMENT), SECTION (PENCIL), SHEAF (GEOMETRY)

References

Woods, F. S. *Higher Geometry: An Introduction to Advanced Methods in Analytic Geometry.* New York: Dover, pp. 8 and 11–12, 1961.

Penrose Stairway

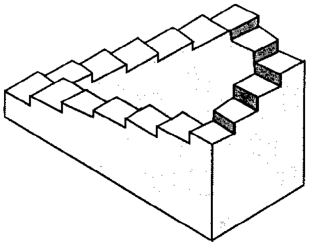

An IMPOSSIBLE FIGURE (also called the SCHROEDER STAIRS) in which a stairway in the shape of a square appears to circulate indefinitely while still possessing normal steps. The Dutch artist M. C. Escher included Penrose stairways in many of his mind-bending illustrations.

see also IMPOSSIBLE FIGURE

References

Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, p. 15, 1989.

Jablan, S. "Impossible Figures." `http://members.tripod.com/~modularity/impos.htm`.

Pappas, T. "Optical Illusions and Computer Graphics." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 5, 1989.

Robinson, J. O. and Wilson, J. A. "The Impossible Colonnade and Other Variations of a Well-Known Figure." *Brit. J. Psych.* **64**, 363–365, 1973.

Penrose Tiles

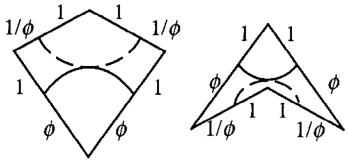

A pair of shapes which tile the plane only aperiodically (when the markings are constrained to match at borders). The two tiles, illustrated above, are called the "KITE" and "DART."

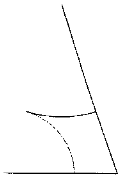

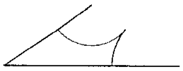

To see how the plane may be tiled aperiodically using the kite and dart, divide the kite into acute and obtuse tiles, shown above. Now define "deflation" and "inflation" operations. The deflation operator takes an acute TRIANGLE to the union of two ACUTE TRIANGLES and one OBTUSE, and the OBTUSE TRIANGLE goes to an ACUTE and an OBTUSE TRIANGLE. These operations are illustrated below.

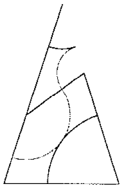

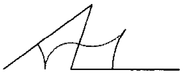

When applied to a collection of tiles, the deflation operator leads to a more refined collection. The operators do not respect tile boundaries, but do respect the half tiles defined above. There are two ways to obtain aperiodic TILINGS with 5-fold symmetry about a single point. They are known an the "star" and "sun" configurations, and are show below.

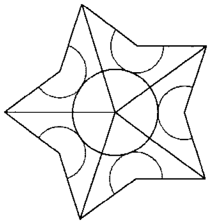

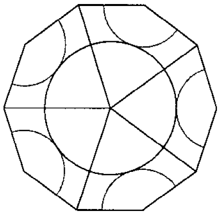

Higher order versions can then be obtained by deflation. For example, the following are third-order deflations:

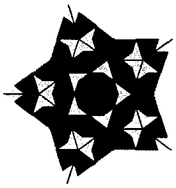

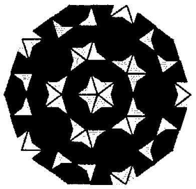

References

Gardner, M. Chs. 1–2 in *Penrose Tiles and Trapdoor Ciphers... and the Return of Dr. Matrix, reissue ed.* New York: W. H. Freeman, pp. 299–300, 1989.

Hurd, L. P. "PenroseTiles." `http://www.mathsource.com/cgi-bin/MathSource/Applications/Graphics/2D/0206-772`.

Peterson, I. *The Mathematical Tourist: Snapshots of Modern Mathematics.* New York: W. H. Freeman, pp. 86–95, 1988.

Wagon, S. "Penrose Tiles." §4.3 in *Mathematica in Action.* New York: W. H. Freeman, pp. 108–117, 1991.

Penrose Triangle

see TRIBAR

Penrose Tribar

see TRIBAR

Pentabolo

A 5-POLYABOLO.

Pentacle

see PENTAGRAM

Pentacontagon

A 50-sided POLYGON.

Pentad

A group of five elements.

see also MONAD, PAIR, QUADRUPLET, QUINTUPLET, TETRAD, TRIAD, TRIPLET, TWINS

Pentadecagon

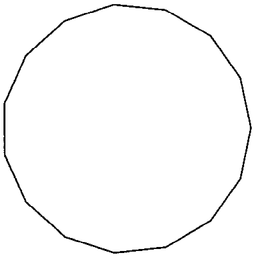

A 15-sided POLYGON, sometimes also called the PENTAKAIDECAGON.

see also POLYGON, REGULAR POLYGON, TRIGONOMETRY VALUES—$\pi/15$

Pentaflake

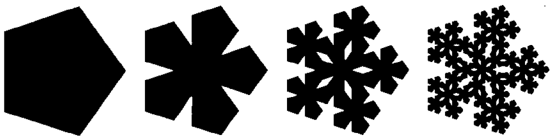

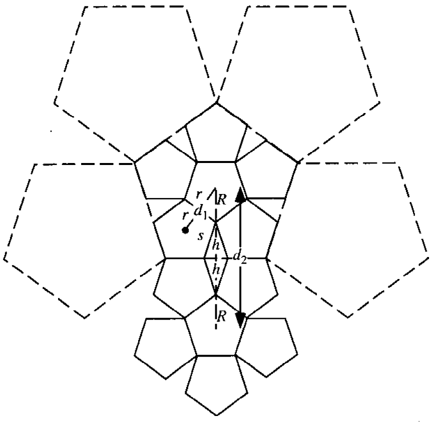

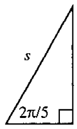

A FRACTAL with 5-fold symmetry. As illustrated above, five PENTAGONS can be arranged around an identical PENTAGON to form the first iteration of the pentaflake. This cluster of six pentagons has the shape of a pentagon with five triangular wedges removed. This construction was first noticed by Albrecht Dürer (Dixon 1991).

For a pentagon of side length 1, the first ring of pentagons has centers at RADIUS

$$d_1 = 2r = \tfrac{1}{2}(1+\sqrt{5})R = \phi R, \tag{1}$$

where ϕ is the GOLDEN RATIO. The INRADIUS r and CIRCUMRADIUS R are related by

$$r = R\cos(\tfrac{1}{5}\pi) = \tfrac{1}{4}(\sqrt{5}+1)R, \tag{2}$$

and these are related to the side length s by

$$s = 2\sqrt{R^2-r^2} = \tfrac{1}{2}R\sqrt{10-2\sqrt{5}}. \tag{3}$$

The height h is

$$h = s\sin(\tfrac{2}{5}\pi) = \tfrac{1}{4}s\sqrt{10+2\sqrt{5}} = \tfrac{1}{2}\sqrt{5}\,R, \tag{4}$$

giving a RADIUS of the second ring as

$$d_2 = 2(R+h) = (2+\sqrt{5})R = \phi^3 R. \tag{5}$$

Continuing, the nth pentagon ring is located at

$$d_n = \phi^{2n-1}. \tag{6}$$

Now, the length of the side of the first pentagon compound is given by

$$s_2 = 2\sqrt{(2r+R)^2-(h+R)^2} = R\sqrt{5+2\sqrt{5}}, \tag{7}$$

so the ratio of side lengths of the original pentagon to that of the compound is

$$\frac{s_2}{s} = \frac{R\sqrt{5+2\sqrt{5}}}{\tfrac{1}{2}R\sqrt{10-2\sqrt{5}}} = 1+\phi. \tag{8}$$

We can now calculate the dimension of the pentaflake fractal. Let N_n be the number of black pentagons and L_n the length of side of a pentagon after the n iteration,

$$N_n = 6^n \tag{9}$$

$$L_n = (1+\phi)^{-n}. \tag{10}$$

The CAPACITY DIMENSION is therefore

$$\begin{aligned} d_{\text{cap}} &= -\lim_{n\to\infty}\frac{\ln N_n}{\ln L_n} = \frac{\ln 6}{\ln(1+\phi)} = \frac{\ln 2+\ln 3}{\ln(1+\phi)} \\ &= 1.861715\ldots. \end{aligned} \tag{11}$$

see also PENTAGON

References

Dixon, R. *Mathographics.* New York: Dover, pp. 186–188, 1991.

Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m`.

Pentagon

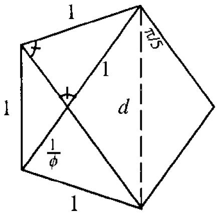

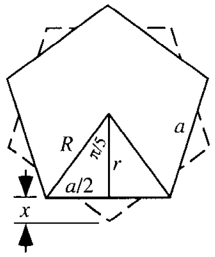

The regular convex 5-gon is called the pentagon. By SIMILAR TRIANGLES in the figure on the left,

$$\frac{d}{1} = \frac{1}{\frac{1}{\phi}} = \phi, \tag{1}$$

where d is the diagonal distance. But the dashed vertical line connecting two nonadjacent VERTICES is the same length as the diagonal one, so

$$\phi = 1 + \frac{1}{\phi} \tag{2}$$

$$\phi^2 - \phi - 1 \tag{3}$$

$$\phi = \frac{1 \pm \sqrt{1+4}}{2} = \frac{1+\sqrt{5}}{2}. \tag{4}$$

This number is the GOLDEN RATIO. The coordinates of the VERTICES relative to the center of the pentagon with unit sides, starting at the right VERTEX and moving clockwise, are $(\cos(\frac{1}{5}n\pi), \sin(\frac{1}{5}n\pi))$ for $n = 0, 1, \ldots, 4$, or

$$(1,0), (c_1, s_1), (c_2, s_2), (c_2, -s_2), (c_1, -s_1), \tag{5}$$

where

$$c_1 = \cos\left(\frac{\pi}{5}\right) = \tfrac{1}{4}(\sqrt{5}+1) \tag{6}$$

$$c_2 = \cos\left(\frac{2\pi}{5}\right) = \tfrac{1}{4}(\sqrt{5}-1) \tag{7}$$

$$s_1 = \sin\left(\frac{\pi}{5}\right) = \tfrac{1}{4}\sqrt{10-2\sqrt{5}} \tag{8}$$

$$s_2 = \sin\left(\frac{2\pi}{5}\right) = \tfrac{1}{4}\sqrt{10+2\sqrt{5}}. \tag{9}$$

For a regular POLYGON, the CIRCUMRADIUS, INRADIUS, SAGITTA, and AREA are given by

$$R_n = \tfrac{1}{2}a\csc\left(\frac{\pi}{n}\right) \tag{10}$$

$$r_n = \tfrac{1}{2}a\cot\left(\frac{\pi}{n}\right) \tag{11}$$

$$x_n = R_n - r_n = \tfrac{1}{2}a\tan\left(\frac{\pi}{2n}\right) \tag{12}$$

$$A_n = \tfrac{1}{4}na^2\cot\left(\frac{\pi}{n}\right). \tag{13}$$

Plugging in $n = 5$ gives

$$R = \tfrac{1}{2}a\csc(\tfrac{1}{5}\pi) = \tfrac{1}{10}a\sqrt{50+10\sqrt{5}} \tag{14}$$

$$r = \tfrac{1}{2}a\cot(\tfrac{1}{5}\pi) = \tfrac{1}{10}a\sqrt{25+10\sqrt{5}} \tag{15}$$

$$x = \tfrac{1}{10}a\sqrt{25-10\sqrt{5}} \tag{16}$$

$$A = \tfrac{5}{4}a^2\sqrt{25+10\sqrt{5}}. \tag{17}$$

Five pentagons can be arranged around an identical pentagon to form the first iteration of the "PENTAFLAKE," which itself has the shape of a pentagon with five triangular wedges removed. For a pentagon of side length 1, the first ring of pentagons has centers at radius ϕ, the second ring at ϕ^3, and the nth at ϕ^{2n-1}.

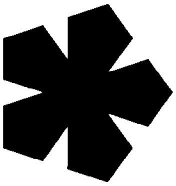

In proposition IV.11, Euclid showed how to inscribe a regular pentagon in a CIRCLE. Ptolemy also gave a RULER and COMPASS construction for the pentagon in his epoch-making work *The Almagest.* While Ptolemy's construction has a SIMPLICITY of 16, a GEOMETRIC CONSTRUCTION using CARLYLE CIRCLES can be made with GEOMETROGRAPHY symbol $2S_1 + S_2 + 8C_1 + 0C_2 + 4C_3$, which has SIMPLICITY 15 (De Temple 1991).

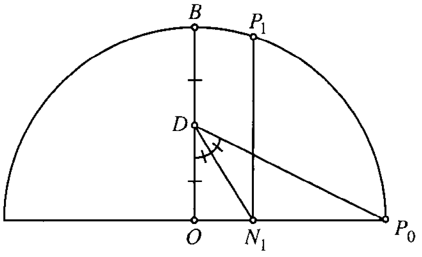

Pentagon

The following elegant construction for the pentagon is due to Richmond (1893). Given a point, a CIRCLE may be constructed of any desired RADIUS, and a DIAMETER drawn through the center. Call the center O, and the right end of the DIAMETER P_0. The DIAMETER PERPENDICULAR to the original DIAMETER may be constructed by finding the PERPENDICULAR BISECTOR. Call the upper endpoint of this PERPENDICULAR DIAMETER B. For the pentagon, find the MIDPOINT of OB and call it D. Draw DP_0, and BISECT $\angle ODP_0$, calling the intersection point with OP_0 N_1. Draw N_1P_1 PARALLEL to OB, and the first two points of the pentagon are P_0 and P_1 (Coxeter 1969).

Madachy (1979) illustrates how to construct a pentagon by folding and knotting a strip of paper.

see also CYCLIC PENTAGON, DECAGON, DISSECTION, FIVE DISKS PROBLEM, HOME PLATE, PENTAFLAKE, PENTAGRAM, POLYGON, TRIGONOMETRY VALUES—$\pi/5$

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 95–96, 1987.

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 26–28, 1969.

De Temple, D. W. "Carlyle Circles and the Lemoine Simplicity of Polygonal Constructions." *Amer. Math. Monthly* **98**, 97–108, 1991.

Dixon, R. *Mathographics.* New York: Dover, p. 17, 1991.

Dudeney, H. E. *Amusements in Mathematics.* New York: Dover, p. 38, 1970.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, p. 59, 1979.

Pappas, T. "The Pentagon, the Pentagram & the Golden Triangle." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 188–189, 1989.

Richmond, H. W. "A Construction for a Regular Polygon of Seventeen Sides." *Quart. J. Pure Appl. Math.* **26**, 206–207, 1893.

Wantzel, P. L. "Recherches sur les moyens de reconnaître si un Problème de Géométrie peut se résoudre avec la règle et le compas." *J. Math. pures appliq.* **1**, 366–372, 1836.

Pentagonal Antiprism

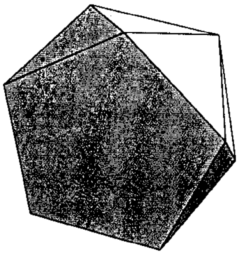

An ANTIPRISM and UNIFORM POLYHEDRON U_{77} whose DUAL POLYHEDRON is the PENTAGONAL DELTAHEDRON.

Pentagonal Cupola

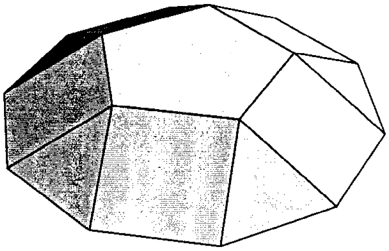

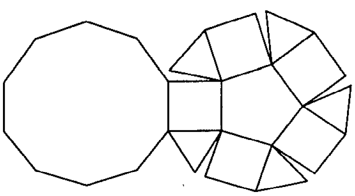

JOHNSON SOLID J_5. The bottom 10 VERTICES are

$$\left(\pm\frac{(1+\sqrt{5})\sqrt{5+\sqrt{5}}}{4\sqrt{2}},\pm\frac{1}{2},0\right),$$
$$\left(\pm\frac{(1+\sqrt{5})\sqrt{5-\sqrt{5}}}{4\sqrt{2}},\pm\frac{3+\sqrt{5}}{2},0\right),$$
$$(0,\pm\tfrac{1}{2}(1+\sqrt{5}),0)$$

and the top five VERTICES are

$$\left(\frac{\sqrt{5+\sqrt{5}}}{\sqrt{10}},0,\frac{\sqrt{5-\sqrt{5}}}{\sqrt{10}}\right),$$
$$\left(\frac{(\sqrt{5}-1)\sqrt{5+\sqrt{5}}}{4\sqrt{10}},\pm\tfrac{1}{4}(1+\sqrt{5}),\frac{\sqrt{5-\sqrt{5}}}{\sqrt{10}}\right),$$
$$\left(\frac{-(\sqrt{5}+1)\sqrt{5+\sqrt{5}}}{4\sqrt{10}},\pm\frac{1}{2},\frac{\sqrt{5-\sqrt{5}}}{\sqrt{10}}\right).$$

Pentagonal Deltahedron

A DELTAHEDRON which is the DUAL POLYHEDRON of the PENTAGONAL ANTIPRISM.

Pentagonal Dipyramid

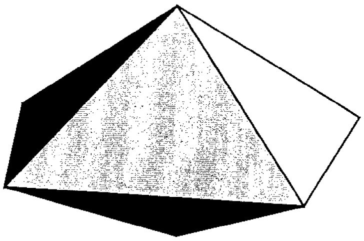

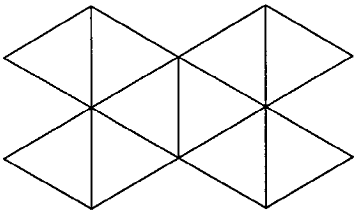

The pentagonal dipyramid is one of the convex DELTAHEDRA, and JOHNSON SOLID J_{13}. It is also the DUAL POLYHEDRON of the PENTAGONAL PRISM. The distance between two adjacent VERTICES on the base of the PENTAGON is

$$\begin{aligned}{d_{12}}^2 &= [1-\cos(\tfrac{2}{5}\pi)]^2+\sin^2(\tfrac{2}{5}\pi)\\ &= [1-\tfrac{1}{4}(\sqrt{5}-1)]^2+\left[\frac{(1+\sqrt{5})\sqrt{5-\sqrt{5}}}{4\sqrt{2}}\right]^2\\ &= \tfrac{1}{2}(5-\sqrt{5}),\end{aligned}\tag{1}$$

and the distance between the apex and one of the base points is

$$d_{1h}{}^2 = (0-1)^2 + (0-0)^2 + (h-0)^2 = 1 + h^2. \quad (2)$$

But

$$d_{12}{}^2 = d_{12}{}^2 \quad (3)$$

$$\tfrac{1}{2}(5-\sqrt{5}) = 1 + h^2 \quad (4)$$

$$h^2 = \tfrac{1}{2}(3-\sqrt{5}), \quad (5)$$

and

$$h = \sqrt{\frac{3-\sqrt{5}}{2}}. \quad (6)$$

This root is of the form $\sqrt{a+b\sqrt{c}}$, so applying SQUARE ROOT simplification gives

$$h = \tfrac{1}{2}(\sqrt{5}-1) \equiv \phi - 1, \quad (7)$$

where ϕ is the GOLDEN MEAN.

see also DELTAHEDRON, DIPYRAMID, GOLDEN MEAN, ICOSAHEDRON, JOHNSON SOLID, TRIANGULAR DIPYRAMID

Pentagonal Gyrobicupola

see JOHNSON SOLID

Pentagonal Gyrocupolarotunda

see JOHNSON SOLID

Pentagonal Hexecontahedron

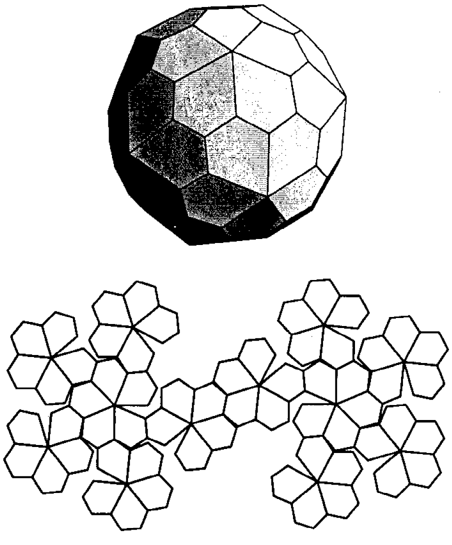

The DUAL POLYHEDRON of the SNUB DODECAHEDRON.

Pentagonal Icositetrahedron

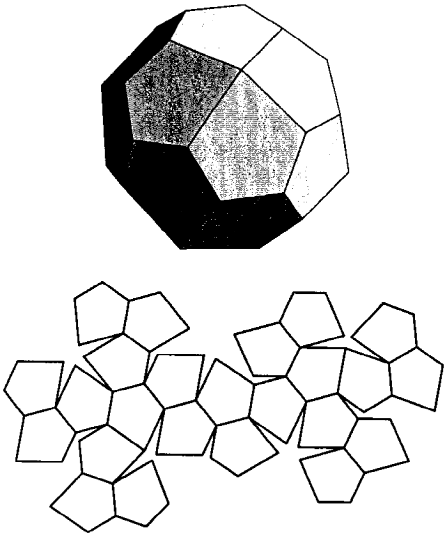

The DUAL POLYHEDRON of the SNUB CUBE.

Pentagonal Number

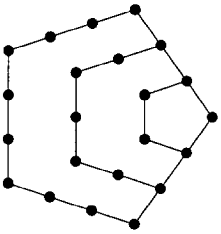

A POLYGONAL NUMBER of the form $n(3n-1)/2$. The first few are 1, 5, 12, 22, 35, 51, 70, ... (Sloane's A000326). The GENERATING FUNCTION for the pentagonal numbers is

$$\frac{x(2x+1)}{(1-x)^3} = x + 5x^2 + 12x^3 + 22x^4 + \ldots.$$

Every pentagonal number is 1/3 of a TRIANGULAR NUMBER.

The so-called generalized pentagonal numbers are given by $n(3n-1)/2$ with $n = 0, \pm1, \pm2, \ldots$, the first few of which are 0, 1, 2, 5, 7, 12, 15, 22, 26, 35, ... (Sloane's A001318).

see also EULER'S PENTAGONAL NUMBER THEOREM, PARTITION FUNCTION P, POLYGONAL NUMBER, TRIANGULAR NUMBER

References

Guy, R. K. "Sums of Squares." §C20 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 136–138, 1994.

Pappas, T. "Triangular, Square & Pentagonal Numbers." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 214, 1989.

Sloane, N. J. A. Sequences A000326/M3818 and A001318/M1336 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Pentagonal Orthobicupola

see Johnson Solid

Pentagonal Orthobirotunda

see Johnson Solid

Pentagonal Orthocupolarontunda

see Johnson Solid

Pentagonal Prism

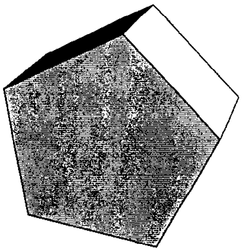

A Prism and Uniform Polyhedron U_{76} whose Dual Polyhedron is the Pentagonal Dipyramid.

see also Pentagrammic Prism

Pentagonal Pyramid

see Johnson Solid

Pentagonal Pyramidal Number

A Pyramidal Number of the form $n^2(n+1)/2$. The first few are 1, 6, 18, 40, 75, ... (Sloane's A002411). The Generating Function for the pentagonal pyramidal numbers is

$$\frac{x(2x+1)}{(x-1)^4} = x + 6x^2 + 18x^3 + 40x^4 + \ldots.$$

References

Sloane, N. J. A. Sequence A002411/M4116 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Pentagonal Rotunda

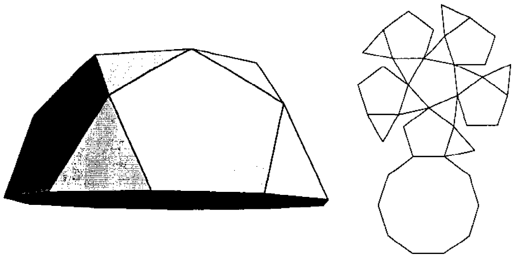

Half of an Icosidodecahedron, denoted R_5. It has 10 triangular and five pentagonal faces separating a Pentagonal ceiling and a Dodecahedral floor. It is Johnson Solid J_6, and the only true Rotunda.

see also Icosidodecahedron, Johnson Solid, Rotunda

Pentagonal Tiling

see Tiling

Pentagram

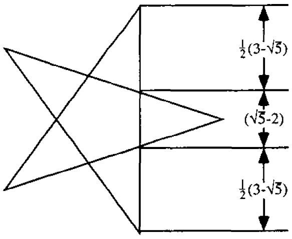

The Star Polygon $\{\frac{5}{2}\}$, also called the Pentacle, Pentalpha, or Pentangle.

see also Dissection, Hexagram, Hoehn's Theorem, Pentagon, Star Figure, Star of Lakshmi

References

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 122–125, 1990.

Pappas, T. "The Pentagon, the Pentagram & the Golden Triangle." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 188–189, 1989.

Schwartzman, S. *The Words of Mathematics: An Etymological Dictionary of Mathematical Terms Used in English.* Washington, DC: Math. Assoc. Amer., 1994.

Pentagrammic Antiprism

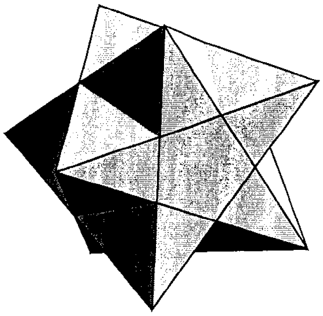

An Antiprism and Uniform Polyhedron U_{79} whose Dual Polyhedron is the Pentagrammic Deltahedron.

Pentagrammic Concave Deltahedron

The Dual Polyhedron of the Pentagrammic Crossed Antiprism.

Pentagrammic Crossed Antiprism

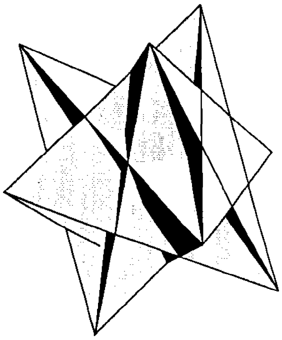

An ANTIPRISM and UNIFORM POLYHEDRON U_{80} whose DUAL POLYHEDRON is the PENTAGRAMMIC CONCAVE DELTAHEDRON.

Pentagrammic Deltahedron

The DUAL POLYHEDRON of the PENTAGRAMMIC ANTIPRISM.

Pentagrammic Dipyramid

The DUAL POLYHEDRON of the PENTAGRAMMIC PRISM.

Pentagrammic Prism

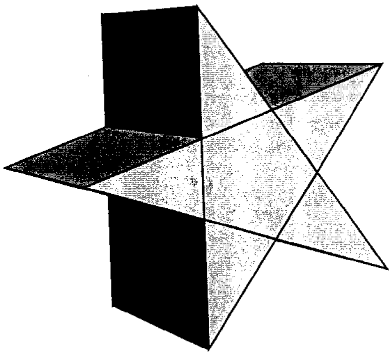

A PRISM and UNIFORM POLYHEDRON U_{78} whose DUAL POLYHEDRON is the PENTAGRAMMIC DIPYRAMID.

see also PENTAGONAL PRISM

Pentakaidecagon

see PENTADECAGON

Pentakis Dodecahedron

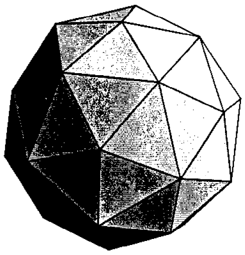

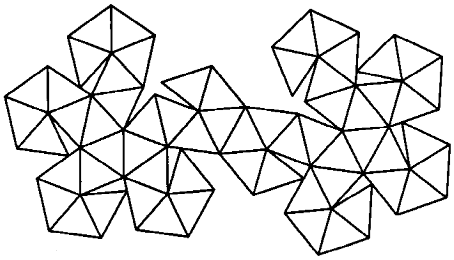

The DUAL POLYHEDRON of the TRUNCATED ICOSAHEDRON.

see also ARCHIMEDEAN SOLID, DUAL POLYHEDRON, TRUNCATED ICOSAHEDRON

Pentalpha

see PENTAGRAM

Pentangle

see PENTAGRAM

Pentatope

The simplest regular figure in 4-D.

Pentatope Number

A FIGURATE NUMBER which is given by

$$Ptop_n = \tfrac{1}{4} Te_n(n+3) = \tfrac{1}{24} n(n+1)(n+2)(n+3),$$

where Te_n is the nth TETRAHEDRAL NUMBER. The first few pentatope numbers are 1, 5, 15, 35, 70, 126, ... (Sloane's A000332). The GENERATING FUNCTION for the pentatope numbers is

$$\frac{x}{(1-x)^5} = x + 5x^2 + 15x^3 + 35x^4 + \ldots.$$

see also FIGURATE NUMBER, TETRAHEDRAL NUMBER

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 55–57, 1996.

Pentomino

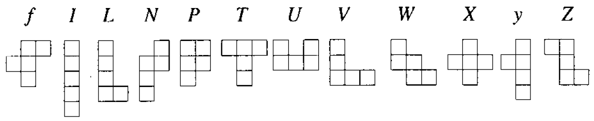

The twelve 5-POLYOMINOES illustrated above and known by the letters of the alphabet they most closely resemble: $f, I, L, N, P, T, U, V, W, X, y, Z$ (Gardner 1960).

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 110–111, 1987.

Dudeney, H. E. "The Broken Chessboard." Problem 74 in *The Canterbury Puzzles and Other Curious Problems, 7th ed.* London: Thomas Nelson and Sons, pp. 119–120, 1949.

Gardner, M. "Mathematical Games: More About the Shapes that Can Be Made with Complex Dominoes." *Sci. Amer.* **203**, 186–198, Nov. 1960.

Hunter, J. A. H. and Madachy, J. S. *Mathematical Diversions.* New York: Dover, pp. 80–86, 1975.

Lei, A. "Pentominoes." `http://www.cs.ust.hk/~philipl/omino/pento.html`.

Ruskey, F. "Information on Pentomino Puzzles." `http://sue.csc.uvic.ca/~cos/inf/misc/PentInfo.html`.

Pépin's Test

A test for the PRIMALITY of FERMAT NUMBERS $F_n = 2^{2^n} + 1$, with $n \geq 2$ and $k \geq 2$. Then the two following conditions are equivalent:

1. F_n is PRIME and $k/F_n = -1$.
2. $k^{(F_n-1)/2} \equiv -1 \pmod{F_n}$.

k is usually taken as 3 as a first test.

see also FERMAT NUMBER, PÉPIN'S THEOREM

References

Ribenboim, P. *The Little Book of Big Primes.* New York: Springer-Verlag, p. 62, 1991.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 119–120, 1993.

Pépin's Theorem

The FERMAT NUMBER F_n is PRIME IFF

$$3^{2^{2^n-1}} \equiv -1 \pmod{F_n}.$$

see also FERMAT NUMBER, PÉPIN'S TEST, SELFRIDGE-HURWITZ RESIDUE

Percent

The use of percentages is a way of expressing RATIOS in terms of whole numbers. Given a RATIO or FRACTION, it is converted to a percentage by multiplying by 100 and appending a "percentage sign" %. For example, if an investment grows from a number $P = 13.00$ to a number $A = 22.50$, then A is $22.50/13.00 = 1.7308$ times as much as P, or 173.08%, and the investment has grown by 73.08%.

see also PERCENTAGE ERROR, PERMIL

Percentage Error

The percentage error is 100% times the RELATIVE ERROR.

see also ABSOLUTE ERROR, ERROR PROPAGATION, PERCENT, RELATIVE ERROR

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 14, 1972.

Percolation Theory

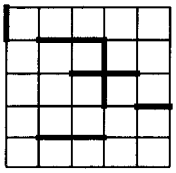

bond percolation

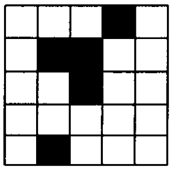

site percolation

Percolation theory deals with fluid flow (or any other similar process) in random media. If the medium is a set of regular LATTICE POINTS, then there are two types of percolation. A SITE PERCOLATION considers the lattice vertices as the relevant entities; a BOND PERCOLATION considers the lattice edges as the relevant entities.

see also BOND PERCOLATION, CAYLEY TREE, CLUSTER, CLUSTER PERIMETER, LATTICE ANIMAL, PERCOLATION THRESHOLD, POLYOMINO, s-CLUSTER, s-RUN, SITE PERCOLATION

References

Deutscher, G.; Zallen, R.; and Adler, J. (Eds.). *Percolation Structures and Processes.* Bristol: Adam Hilger, 1983.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/rndprc/rndprc.html`.

Grimmett, G. *Percolation.* New York: Springer-Verlag, 1989.

Kesten, H. *Percolation Theory for Mathematicians.* Boston, MA: Birkhäuser, 1982.

Stauffer, D. and Aharony, A. *Introduction to Percolation Theory, 2nd ed.* London: Taylor & Francis, 1992.

Percolation Threshold

The critical fraction of lattice points which must be filled to create a continuous path of nearest neighbors from one side to another. The following table is from Stauffer and Aharony (1992, p. 17).

Lattice	Site	Bond
Cubic (Body-Centered)	0.246	0.1803
Cubic (Face-Centered)	0.198	0.119
Cubic (Simple)	0.3116	0.2488
Diamond	0.43	0.388
Honeycomb	0.6962	0.65271
4-Hypercubic	0.197	0.1601
5-Hypercubic	0.141	0.1182
6-Hypercubic	0.107	0.0942
7-Hypercubic	0.089	0.0787
Square	0.592746	0.50000
Triangular	0.50000	0.34729

The square bond value is 1/2 exactly, as is the triangular site. $p_c = 2\sin(\pi/18)$ for the triangular bond and $p_c = 1 - 2\sin(\pi/18)$ for the honeycomb bond. An exact answer for the square site percolation threshold is not known.

see also PERCOLATION THEORY

References

Essam, J. W.; Gaunt, D. S.; and Guttmann, A. J. "Percolation Theory at the Critical Dimension." *J. Phys. A* **11**, 1983–1990, 1978.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/rndprc/rndprc.html`.

Kesten, H. *Percolation Theory for Mathematicians.* Boston, MA: Birkhäuser, 1982.

Stauffer, D. and Aharony, A. *Introduction to Percolation Theory, 2nd ed.* London: Taylor & Francis, 1992.

Perfect Box

see EULER BRICK

Perfect Cubic

A perfect cubic POLYNOMIAL can be factored into a linear and a quadratic term,

$$(a^3 - b^3) = (a - b)(a^2 + ab + b^2)$$

$$(a^3 + b^3) = (a + b)(a^2 - ab + b^2).$$

see also CUBIC EQUATION, PERFECT SQUARE, POLYNOMIAL

Perfect Cuboid

see EULER BRICK

Perfect Difference Set

A SET of RESIDUES $\{a_1, a_2, \ldots, a_{k+1}\}$ (mod n) such that every NONZERO RESIDUE can be uniquely expressed in the form $a_i - a_j$. Examples include {1, 2, 4} (mod 7) and {1, 2, 5, 7} (mod 13). A NECESSARY condition for a difference set to exist is that n be of the form $k^2 + k + 1$. A SUFFICIENT condition is that k be a PRIME POWER. Perfect sets can be used in the construction of PERFECT RULERS.

see also PERFECT RULER

References

Guy, R. K. "Modular Difference Sets and Error Correcting Codes." §C10 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 118–121, 1994.

Perfect Digital Invariant

see NARCISSISTIC NUMBER

Perfect Information

A class of GAME in which players move alternately and each player is completely informed of previous moves. FINITE, ZERO-SUM, two-player GAMES with perfect information (including checkers and chess) have a SADDLE POINT, and therefore one or more optimal strategies. However, the optimal strategy may be so difficult to compute as to be effectively impossible to determine (as in the game of CHESS).

see also FINITE GAME, GAME, ZERO-SUM GAME

Perfect Magic Cube

A perfect magic cube is a MAGIC CUBE for which the cross-section diagonals, as well as the space diagonals, sum to the MAGIC CONSTANT.

see also MAGIC CUBE, SEMIPERFECT MAGIC CUBE

References

Gardner, M. "Magic Squares and Cubes." Ch. 17 in *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, pp. 213–225, 1988.

Perfect Number

Perfect numbers are INTEGERS n such that

$$n = s(n), \tag{1}$$

where $s(n)$ is the RESTRICTED DIVISOR FUNCTION (i.e., the SUM of PROPER DIVISORS of n), or equivalently

$$\sigma(n) = 2n, \tag{2}$$

where $\sigma(n)$ is the DIVISOR FUNCTION (i.e., the SUM of DIVISORS of n including n itself). The first few perfect numbers are 6, 28, 496, 8128, ... (Sloane's A000396). This follows from the fact that

$$\begin{aligned} 6 &= \sum 1, 2, 3 \\ 28 &= \sum 1, 2, 4, 7, 14 \\ 496 &= \sum 1, 2, 4, 8, 16, 31, 62, 124, 248, \end{aligned}$$

etc.

Perfect numbers are intimately connected with a class of numbers known as MERSENNE PRIMES. This can be demonstrated by considering a perfect number P of the form $P = q2^{p-1}$ where q is PRIME. Then

$$\sigma(P) = 2P, \tag{3}$$

and using

$$\sigma(q) = q + 1 \tag{4}$$

for q prime, and

$$\sigma(2^\alpha) = 2^{\alpha+1} - 1 \tag{5}$$

gives

$$\begin{aligned} \sigma(q2^{p-1}) = \sigma(q)\sigma(2^{p-1}) &= (q + 1)(2^p - 1) \\ &= 2q2^{p-1} = q2^p \end{aligned} \tag{6}$$

$$q(2^p - 1) + 2^p - 1 = q2^p \tag{7}$$

$$q = 2^p - 1. \tag{8}$$

Therefore, if $M_p \equiv q = 2^p - 1$ is PRIME, then

$$P = \tfrac{1}{2}(M_p + 1)M_p = 2^{p-1}(2^p - 1) \tag{9}$$

is a perfect number, as was stated in Proposition IX.36 of Euclid's *Elements* (Dunham 1990). The first few perfect numbers are summarized in the following table.

#	p	P
1	2	6
2	3	28
3	5	496
4	7	8128
5	13	33550336
6	17	8589869056
7	19	137438691328
8	31	2305843008139952128

All EVEN perfect numbers are of this form, as was proved by Euler in a posthumous paper. The only even perfect number of the form x^3+1 is 28 (Mąkowski 1962).

It is not known if any ODD perfect numbers exist, although numbers up to 10^{300} have been checked (Brent *et al.* 1991, Guy 1994) without success, improving the result of Tuckerman (1973), who checked odd numbers up to 10^{36}. Euler showed that an ODD perfect number, if it exists, must be of the form

$$m = p^{4a+1}Q^2, \tag{10}$$

where p is an ODD PRIME RELATIVELY PRIME to Q. In 1887, Sylvester conjectured and in 1925, Gradshtein proved that any ODD perfect number must have at least six different prime aliquot factors (or eight if it is not divisible by 3; Ball and Coxeter 1987). Catalan (1888) proved that if an ODD perfect number is not divisible by 3, 5, or 7, it has at least 26 distinct prime aliquot factors. Stuyvaert (1896) proved that an ODD perfect number must be a sum of squares. All EVEN perfect numbers end in 16, 28, 36, 56, or 76 (Lucas 1891) and, with the exception of 6, have DIGITAL ROOT 1.

Every perfect number of the form $2^p(2^{p+1}-1)$ can be written

$$2^p(2^{p+1}-1) = \sum_{k=1}^{p/2}(2k-1)^3. \tag{11}$$

All perfect numbers are HEXAGONAL NUMBERS and therefore TRIANGULAR NUMBERS. It therefore follows that perfect numbers are always the sum of consecutive POSITIVE integers starting at 1, for example,

$$6 = \sum_{n=1}^{3} n \tag{12}$$

$$28 = \sum_{n=1}^{7} n \tag{13}$$

$$496 = \sum_{n=1}^{31} n \tag{14}$$

(Singh 1997). All EVEN perfect numbers $P > 6$ are of the form

$$P + 1 + 9T_n, \tag{15}$$

where T_n is a TRIANGULAR NUMBER

$$T_n = \tfrac{1}{2}n(n+1) \tag{16}$$

such that $n = 8j+2$ (Eaton 1995, 1996). The sum of reciprocals of all the divisors of a perfect number is 2, since

$$\underbrace{n + \ldots + c + b + a}_{n} = 2n \tag{17}$$

$$\frac{n}{a} + \frac{n}{b} + \ldots = 2n \tag{18}$$

$$\frac{1}{a} + \frac{1}{b} + \ldots = 2. \tag{19}$$

If $s(n) > n$, n is said to be an ABUNDANT NUMBER. If $s(n) < n$, n is said to be a DEFICIENT NUMBER. And if $s(n) = kn$ for a POSITIVE INTEGER $k > 1$, n is said to be a MULTIPERFECT NUMBER of order k.

see also ABUNDANT NUMBER, ALIQUOT SEQUENCE, AMICABLE NUMBERS, DEFICIENT NUMBER, DIVISOR FUNCTION, e-PERFECT NUMBER, HARMONIC NUMBER, HYPERPERFECT NUMBER, INFINARY PERFECT NUMBER, MERSENNE NUMBER, MERSENNE PRIME, MULTIPERFECT NUMBER, MULTIPLICATIVE PERFECT NUMBER, PLUPERFECT NUMBER, PSEUDOPERFECT NUMBER, QUASIPERFECT NUMBER, SEMIPERFECT NUMBER, SMITH NUMBER, SOCIABLE NUMBERS, SUBLIME NUMBER, SUPERPERFECT NUMBER, UNITARY PERFECT NUMBER, WEIRD NUMBER

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 66–67, 1987.

Brent, R. P.; Cohen, G. L. L.; and te Riele, H. J. J. "Improved Techniques for Lower Bounds for Odd Perfect Numbers." *Math. Comput.* **57**, 857–868, 1991.

Conway, J. H. and Guy, R. K. "Perfect Numbers." In *The Book of Numbers.* New York: Springer-Verlag, pp. 136–137, 1996.

Dickson, L. E. "Notes on the Theory of Numbers." *Amer. Math. Monthly* **18**, 109–111, 1911.

Dickson, L. E. *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, pp. 3–33, 1952.

Dunham, W. *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, p. 75, 1990.

Eaton, C. F. "Problem 1482." *Math. Mag.* **68**, 307, 1995.

Eaton, C. F. "Perfect Number in Terms of Triangular Numbers." Solution to Problem 1482. *Math. Mag.* **69**, 308–309, 1996.

Gardner, M. "Perfect, Amicable, Sociable." Ch. 12 in *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, pp. 160–171, 1978.

Guy, R. K. "Perfect Numbers." §B1 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 145, 1994.

Kraitchik, M. "Mersenne Numbers and Perfect Numbers." §3.5 in *Mathematical Recreations.* New York: W. W. Norton, pp. 70–73, 1942.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 145 and 147–151, 1979.

Mąkowski, A. "Remark on Perfect Numbers." *Elemente Math.* **17**, 109, 1962.

Powers, R. E. "The Tenth Perfect Number." *Amer. Math. Monthly* **18**, 195–196, 1911.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 1–13 and 25–29, 1993.

Singh, S. *Fermat's Enigma: The Epic Quest to Solve the World's Greatest Mathematical Problem.* New York: Walker, pp. 11–13, 1997.

Sloane, N. J. A. Sequence A000396/M4186 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Tuckerman, B. "Odd Perfect Numbers: A Search Procedure, and a New Lower Bound of 10^{36}." *Not. Amer. Math. Soc.* **15**, 226, 1968.

Tuckerman, B. "A Search Procedure and Lower Bound for Odd Perfect Numbers." *Math. Comp.* **27**, 943–949, 1973.
Zachariou, A. and Zachariou, E. "Perfect, Semi-Perfect and Ore Numbers." *Bull. Soc. Math. Grèce (New Ser.)* **13**, 12–22, 1972.

Perfect Partition

A PARTITION of n which can generate any number 1, 2, ..., n.

see also PARTITION

References
Cohen, D. I. A. *Basic Techniques of Combinatorial Theory.* New York: Wiley and Sons, p. 97, 1978.
Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 140–143, 1985.

Perfect Proportion

Since

$$\frac{2a}{a+b} = \frac{2ab}{(a+b)b}, \tag{1}$$

it follows that

$$\frac{a}{\frac{a+b}{2}} = \frac{\frac{2ab}{a+b}}{b}, \tag{2}$$

so

$$\frac{a}{A} = \frac{H}{b}, \tag{3}$$

where A and H are the ARITHMETIC MEAN and HARMONIC MEAN of a and b. This relationship was purportedly discovered by Pythagoras.

see also ARITHMETIC MEAN, HARMONIC MEAN

Perfect Rectangle

A RECTANGLE which cannot be built up of SQUARES all of different sizes is called an imperfect rectangle. A RECTANGLE which can be built up of SQUARES all of different sizes is called perfect.

order	perfect	imperfect
< 9	0	0
9	2	1
10	6	0
11	22	0
12	67	9
13	213	34
14	744	104
15	2609	282

Perfect Ruler

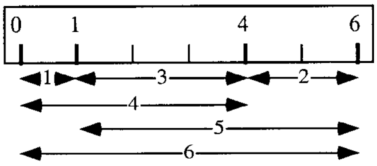

A type of RULER considered by Guy (1994) which has k distinct marks spaced such that the distances between marks can be used to measure all the distances 1, 2, 3, 4, ... up to some maximum distance $n > k$. Such a ruler can be constructed from a PERFECT DIFFERENCE SET by subtracting one from each element. For example, the PERFECT DIFFERENCE SET {1, 2, 5, 7} gives 0, 1, 4, 6, which can be used to measure $1 - 0 = 1$, $6 - 4 = 2$, $4 - 1 = 3$, $4 - 0 = 4$, $6 - 1 = 5$, $6 - 0 = 6$ (so we get 6 distances with only four marks).

see also PERFECT DIFFERENCE SET

References
Guy, R. K. "Modular Difference Sets and Error Correcting Codes." §C10 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 118–121, 1994.

Perfect Set

A SET P is called perfect if $P = P'$, where P' is the DERIVED SET of P.

see also DERIVED SET, SET

Perfect Square

The term perfect square is used to refer to a SQUARE NUMBER, a PERFECT SQUARE DISSECTION, or a factorable quadratic polynomial of the form $a^2 - b^2 = (a - b)(a + b)$.

see also PERFECT SQUARE DISSECTION, QUADRATIC EQUATION, SQUARE NUMBER, SQUAREFREE

Perfect Square Dissection

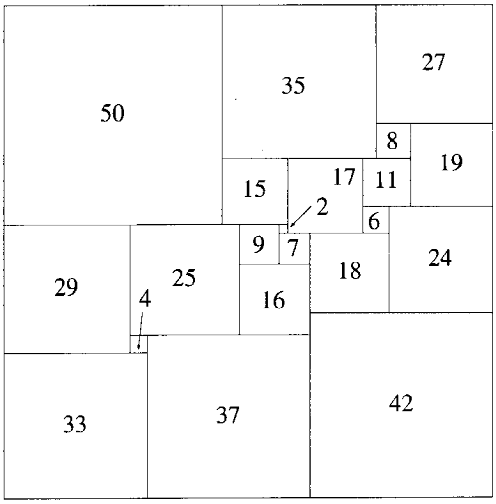

A SQUARE which can be DISSECTED into a number of smaller SQUARES with no two equal is called a PERFECT SQUARE DISSECTION (or a SQUARED SQUARE). Square dissections in which the squares need not be different sizes are called MRS. PERKINS' QUILTS. If no subset of the SQUARES forms a RECTANGLE, then the perfect square is called "simple." Lusin claimed that perfect squares were impossible to construct, but this assertion was proved erroneous when a 55-SQUARE perfect square was published by R. Sprague in 1939 (Wells 1991).

There is a unique simple perfect square of order 21 (the lowest possible order), discovered in 1978 by

A. J. W. Duijvestijn (Bouwkamp and Duijvestijn 1992). It is composed of 21 squares with total side length 112, and is illustrated above. There is a simple notation (sometimes called Bouwkamp code) used to describe perfect squares. In this notation, brackets are used to group adjacent squares with flush tops, and then the groups are sequentially placed in the highest (and leftmost) possible slots. For example, the 21-square illustrated above is denoted [50, 35, 27], [8, 19], [15, 17, 11], [6, 24], [29, 25, 9, 2], [7, 18], [16], [42], [4, 37], [33].

The number of simple perfect squares of order n for $n \geq 21$ are 1, 8, 12, 26, 160, 441, ... (Sloane's A006983). Duijvestijn's Table I gives a list of the 441 simple perfect squares of order 26, the smallest with side length 212 and the largest with side length 825. Skinner (1993) gives the smallest possible side length (and smallest order for each) as 110 (22), 112 (21), 120 (24), 139 (22), 140 (23), ... for simple perfect squared squares, and 175 (24), 235 (25), 288 (26), 324 (27), 325 (27), ... for compound perfect squared squares.

There are actually three simple perfect squares having side length 110. They are [60, 50], [23, 27], [24, 22, 14], [7, 16], [8, 6], [12, 15], [13], [2, 28], [26], [4, 21, 3], [18], [17] (order 22; discovered by A. J. W. Duijvestijn); [60, 50], [27, 23], [24, 22, 14], [4, 19], [8, 6], [3, 12, 16], [9], [2, 28], [26], [21], [1, 18], [17] (order 22; discovered by T. H. Willcocks); and [44, 29, 37], [21, 8], [13, 32], [28, 16], [15, 19], [12,4], [3, 1], [2, 14], [5], [10, 41], [38, 7], [31] (order 23; discovered by A. J. W. Duijvestijn).

D. Sleator has developed an efficient ALGORITHM for finding *non*-simple perfect squares using what he calls rectangle and "ell" grow sequences. This algorithm finds a slew of compound perfect squares of orders 24–32. Weisstein gives a partial list of known simple and compound perfect squares (where the number of simple perfect squares is exact for orders less than 27) as well as *Mathematica*® (Wolfram Research, Champaign, IL) algorithms for drawing them.

Order	# Simple	# Compound
21	1	0
22	8	0
23	12	0
24	26	1
25	160	1
26	441	2
27	?	2
28	?	4
29	?	2
30	?	3
31	?	2
32	?	2
38	1	0
69	1	0

see also MRS. PERKINS' QUILT

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 115–116, 1987.

Beiler, A. H. *Recreations in the Theory of Numbers: The Queen of Mathematics Entertains.* New York: Dover, pp. 157–161, 1966.

Bouwkamp, C. J. and Duijvestijn, A. J. W. "Catalogue of Simple Perfect Squared Squares of Orders 21 Through 25." Eindhoven Univ. Technology, Dept. Math, Report 92-WSK-03, Nov. 1992.

Brooks, R. L.; Smith, C. A. B.; Stone, A. H.; and Tutte, W. T. "The Dissection of Rectangles into Squares." *Duke Math. J.* **7**, 312–340, 1940.

Duijvestijn, A. J. W. "A Simple Perfect Square of Lowest Order." *J. Combin. Th. Ser. B* **25**, 240–243, 1978.

Duijvestijn, A. J. W. "A Lowest Order Simple Perfect 2 × 1 Squared Rectangle." *J. Combin. Th. Ser. B* **26**, 372–374, 1979.

Duijvestijn, A. J. W. `ftp://ftp.cs.utwente.nl/pub/doc/dvs/TableI`.

Gardner, M. "Squaring the Square." Ch. 17 in *The Second Scientific American Book of Mathematical Puzzles & Diversions: A New Selection.* New York: Simon and Schuster, 1961.

Gardner, M. *Fractal Music, Hypercards, and More: Mathematical Recreations from Scientific American Magazine.* New York: W. H. Freeman, pp. 172–174, 1992.

Kraitchik, M. *Mathematical Recreations.* New York: W. W. Norton, p. 198, 1942.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 15 and 32–33, 1979.

Mauldin, R. D. (Ed.) *The Scottish Book: Math at the Scottish Cafe* Boston, MA: Birkhäuser, 1982.

Moroń, Z. "O rozkładach prostokątów na kwadraty." *Przeglad matematyczno-fizyczny* **3**, 152–153, 1925.

Skinner, J. D. II. *Squares Squares: Who's Who & What's What.* Published by the author, 1993.

Sloane, N. J. A. Sequences A006983/M4482 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sloane, N. J. A. and Plouffe, S. Extended entry in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.

Sprague, R. "Beispiel einer Zerlegung des Quadrats in lauter verschiedene Quadrate." *Math. Z.* **45**, 607–608, 1939.

Weisstein, E. W. "Perfect Squares." `http://www.astro.virginia.edu/~eww6n/math/notebooks/PerfectSquare.m`.

Wells, D. *The Penguin Dictionary of Curious and Interesting Geometry.* London: Penguin, p. 242, 1991.

Periapsis

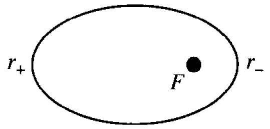

The smallest radial distance of an ELLIPSE as measured from a FOCUS. Taking $v = 0$ in the equation of an ELLIPSE

$$r = \frac{a(1-e^2)}{1+e\cos v}$$

gives the periapsis distance

$$r_- = a(1-e).$$

Periapsis for an orbit around the Earth is called perigee, and periapsis for an orbit around the Sun is called perihelion.

see also Apoapsis, Eccentricity, Ellipse, Focus

Perigon

An Angle of 2π radians $= 360°$ corresponding to the Central Angle of an entire Circle.

Perimeter

The Arc Length along the boundary of a closed 2-D region. The perimeter of a Circle is called the Circumference.

see also Circumference, Cluster Perimeter, Semiperimeter

Perimeter Polynomial

A sum over all Cluster Perimeters.

Period Doubling

A characteristic of some systems making a transition to Chaos. Doubling is followed by quadrupling, etc. An example of a map displaying period doubling is the Logistic Map.

see also Chaos, Logistic Map

Period Three Theorem

Li and Yorke (1975) proved that *any* 1-D system which exhibits a regular Cycle of period three will also display regular Cycles of every other length as well as completely Chaotic Cycles.

see also Chaos, Cycle (Map)

References

Li, T. Y. and Yorke, J. A. "Period Three Implies Chaos." *Amer. Math. Monthly* **82**, 985–992, 1975.

Periodic Function

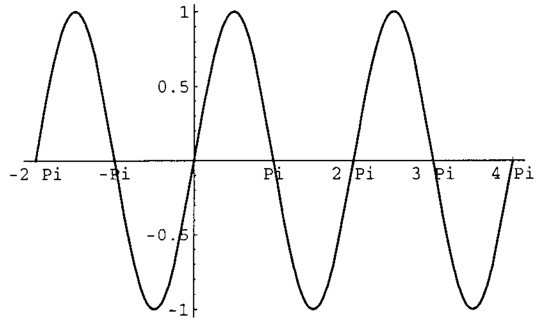

A Function $f(x)$ is said to be periodic with period p if $f(x) = f(x + np)$ for $n = 1, 2, \ldots$. For example, the Sine function $\sin x$ is periodic with period 2π (as well as with period -2π, 4π, 6π, etc.).

The Constant Function $f(x) = 0$ is periodic with any period R for all Nonzero Real Numbers R, so there is no concept analogous to the Least Period of a Periodic Point for functions.

see also Periodic Point, Periodic Sequence

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 425–427, 1953.

Spanier, J. and Oldham, K. B. "Periodic Functions." Ch. 36 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 343–349, 1987.

Periodic Point

A point x_0 is said to be a periodic point of a Function f of period n if $f^n(x_0) = x_0$, where $f_0(x) = x$ and $f^n(x)$ is defined recursively by $f^n(x) = f(f^{n-1}(x))$.

see also Least Period, Periodic Function, Periodic Sequence

Periodic Sequence

A Sequence $\{a_i\}$ is said to be periodic with period p with if it satisfies $a_i = a_{i+np}$ for $n = 1, 2, \ldots$. For example, $\{1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, \ldots\}$ is a periodic sequence with Least Period 2.

see also Eventually Periodic, Periodic Function, Periodic Point

Perkins' Quilt

see Mrs. Perkins' Quilt

Perko Pair

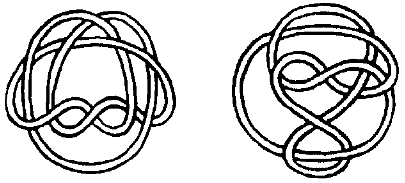

The Knots 10_{161} and 10_{162} illustrated above. They are listed as separate knots in the pictorial enumeration of Rolfsen (1976, Appendix C), but were identified as identical by Perko (1974).

References

Perko, K. A. Jr. "On the Classification of Knots." *Proc. Amer. Math. Soc.* **45**, 262–266, 1974.

Rolfsen, D. "Table of Knots and Links." Appendix C in *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 280–287, 1976.

Permanence of Algebraic Form

All Elementary Functions can be extended to the Complex Plane. Such definitions agree with the Real definitions on the x-Axis and constitute an Analytic Continuation.

see also Analytic Continuation, Elementary Function, Permanence of Mathematical Relations Principle

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 380, 1985.

Permanence of Mathematical Relations Principle

The metric properties discovered for a primitive figure remain applicable, without modifications other than changes of signs, to all correlative figures which can be considered to arise from the first.

This principle was formulated by Poncelet, and amounts to the statement that if an analytic identity in any finite number of variables holds for all real values of the variables, then it also holds by ANALYTIC CONTINUATION for all complex values (Bell 1945). This principle is also called PONCELET'S CONTINUITY PRINCIPLE.

see also ANALYTIC CONTINUATION, CONSERVATION OF NUMBER PRINCIPLE, DUALITY PRINCIPLE, PERMANENCE OF ALGEBRAIC FORM

References
Bell, E. T. *The Development of Mathematics, 2nd ed.* New York: McGraw-Hill, p. 340, 1945.

Permanent

An analog of a DETERMINANT where all the signs in the expansion by MINORS are taken as POSITIVE. The permanent of a MATRIX A is the coefficient of $x_1 \cdots x_n$ in

$$\prod_{i=1}^{n}(a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n)$$

(Vardi 1991). Another equation is the RYSER FORMULA

$$\operatorname{perm}(a_{ij}) = (-1)^n \sum_{s\subseteq\{1,\ldots,n\}} (-1)^{|s|} \prod_{i=1}^{n} \sum_{j\in s} a_{ij},$$

where the SUM is over all SUBSETS of $\{1, \ldots, n\}$, and $|s$ is the number of elements in s (Vardi 1991).

If M is a UNITARY MATRIX, then

$$|\operatorname{perm}(\mathsf{M})| \leq 1$$

(Minc 1978, p. 25; Vardi 1991).

see also DETERMINANT, FROBENIUS-KÖNIG THEOREM, IMMANANT, RYSER FORMULA, SCHUR MATRIX

References
Borovskikh, Y. V.; Korolyuk, V. S. *Random Permanents.* Philadelphia, PA: Coronet Books, 1994.
Minc, H. *Permanents.* Reading, MA: Addison-Wesley, 1978.
Vardi, I. "Permanents." §6.1 in *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 108 and 110–112, 1991.

Permil

The use of percentages is a way of expressing RATIOS in terms of whole numbers. Given a RATIO or FRACTION, it is converted to a permil-age by multiplying by 1000 and appending a "mil sign" ‰. For example, if an investment grows from a number $P = 13.00$ to a number $A = 22.50$, then A is $22.50/13.00 = 1.7308$ times as much as P, or 1730.8 ‰.

see also PERCENT

Permutation

The rearrangement of elements in a set into a ONE-TO-ONE correspondence with itself, also called an ARRANGEMENT or ORDER. The number of ways of obtaining r *ordered* outcomes from a permutation of n elements is

$$_nP_r \equiv \frac{n!}{(n-r)!} = r!\binom{n}{r}, \tag{1}$$

where $n!$ is n FACTORIAL and $\binom{a}{b}$ is a BINOMIAL COEFFICIENT. The total number of permutations for n elements is given by $n!$.

A representation of a permutation as a product of CYCLES is unique (up to the ordering of the cycles). An example of a cyclic decomposition is ({1, 3, 4}, {2}), corresponding to the permutations (1 → 3, 3 → 4, 4 → 1) and (2 → 2), which combine to give {4, 2, 1, 3}.

Any permutation is also a product of TRANSPOSITIONS. Permutations are commonly denoted in LEXICOGRAPHIC or TRANSPOSITION ORDER. There is a correspondence between a PERMUTATION and a pair of YOUNG TABLEAUX known as the SCHENSTED CORRESPONDENCE.

The number of wrong permutations of n objects is $[n!/e]$ where $[x]$ is the NINT function. A permutation of n ordered objects in which no object is in its natural place is called a DERANGEMENT (or sometimes, a COMPLETE PERMUTATION) and the number of such permutations is given by the SUBFACTORIAL $!n$.

Using

$$(x+y)^n = \sum_{r=0}^{n}\binom{n}{r}x^{n-r}y^r \tag{2}$$

with $x = y = 1$ gives

$$2^n = \sum_{r=0}^{n}\binom{n}{r}, \tag{3}$$

so the number of ways of choosing 0, 1, ..., or n at a time is 2^n.

The set of all permutations of a set of elements 1, ..., n can be obtained using the following recursive procedure

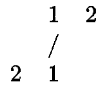

(4)

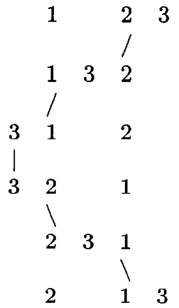

(5)

Let the set of INTEGERS 1, 2, ..., n be permuted and the resulting sequence be divided into increasing RUNS. As n approaches INFINITY, the average length of the nth RUN is denoted L_n. The first few values are

$$L_1 = e - 1 = 1.7182818\ldots \tag{6}$$

$$L_2 = e^2 - 2e = 1.9524\ldots \tag{7}$$

$$L_3 = e^3 - 3e^2 + \tfrac{3}{2}e = 1.9957\ldots, \tag{8}$$

where e is the base of the NATURAL LOGARITHM (Knuth 1973, Le Lionnais 1983).

see also ALTERNATING PERMUTATION, BINOMIAL COEFFICIENT, CIRCULAR PERMUTATION, COMBINATION, COMPLETE PERMUTATION, DERANGEMENT, DISCORDANT PERMUTATION, EULERIAN NUMBER, LINEAR EXTENSION, PERMUTATION MATRIX, SUBFACTORIAL, TRANSPOSITION

References

Bogomolny, A. "Graphs." `http://www.cut-the-knot.com/do_you_know/permutation.html`.

Conway, J. H. and Guy, R. K. "Arrangement Numbers." In *The Book of Numbers.* New York: Springer-Verlag, p. 66, 1996.

Dickau, R. M. "Permutation Diagrams." `http://forum.swarthmore.edu/advanced/robertd/permutations.html`.

Knuth, D. E. *The Art of Computer Programming, Vol. 1: Fundamental Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, 1973.

Kraitchik, M. "The Linear Permutations of n Different Things." §10.1 in *Mathematical Recreations.* New York: W. W. Norton, pp. 239–240, 1942.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, pp. 41–42, 1983.

Ruskey, F. "Information on Permutations." `http://sue.csc.uvic.ca/~cos/inf/perm/PermInfo.html`.

Sloane, N. J. A. Sequence A000142/M1675 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Permutation Group

A finite GROUP of substitutions of elements for each other. For instance, the order 4 permutation group {4, 2, 1, 3} would rearrange the elements {A, B, C, D} in the order {D, B, A, C}. A SUBSTITUTION GROUP of two elements is called a TRANSPOSITION. Every SUBSTITUTION GROUP with > 2 elements can be written as a product of transpositions. For example,

$$(abc) = (ab)(ac)$$

$$(abcde) = (ab)(ac)(ad)(ae).$$

CONJUGACY CLASSES of elements which are interchanged are called CYCLES (in the above example, the CYCLES are {{1, 3, 4}, {2}}).

see also CAYLEY'S GROUP THEOREM, CYCLE (PERMUTATION), GROUP, SUBSTITUTION GROUP, TRANSPOSITION

Permutation Matrix

A MATRIX p_{ij} obtained by permuting the ith and jth rows of the IDENTITY MATRIX with $i < j$. Every row and column therefore contain precisely a single 1, and every permutation corresponds to a unique permutation matrix. The matrix is nonsingular, so the DETERMINANT is always NONZERO. It satisfies

$$\mathsf{p}_{ij}{}^2 = \mathsf{I},$$

where I is the IDENTITY MATRIX. Applying to another MATRIX, $\mathsf{p}_{ij}\mathsf{A}$ gives A with the ith and jth rows interchanged, and $\mathsf{A}\mathsf{p}_{ij}$ gives A with the ith and jth columns interchanged.

Interpreting the 1s in an $n \times n$ permutation matrix as ROOKS gives an allowable configuration of nonattacking ROOKS on an $n \times n$ CHESSBOARD.

see also ELEMENTARY MATRIX, IDENTITY, PERMUTATION, ROOK NUMBER

Permutation Pseudotensor

see PERMUTATION TENSOR

Permutation Symbol

A three-index object sometimes called the LEVI-CIVITA SYMBOL defined by

$$\epsilon_{ijk} = \begin{cases} 0 & \text{for } i = j, j = k, \text{ or } k = i \\ +1 & \text{for } (i,j,k) \in \{(1,2,3),(2,3,1),(3,1,2)\} \\ -1 & \text{for } (i,j,k) \in \{(1,3,2),(3,2,1),(2,1,3)\}. \end{cases} \tag{1}$$

The permutation symbol satisfies

$$\delta_{ij}\epsilon_{ijk} = 0 \tag{2}$$

$$\epsilon_{ipq}\epsilon_{jpq} = 2\delta_{ij} \tag{3}$$

$$\epsilon_{ijk}\epsilon_{ijk} = 6 \tag{4}$$

$$\epsilon_{ijk}\epsilon_{pqk} = \delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}, \tag{5}$$

where δ_{ij} is the KRONECKER DELTA. The symbol can be defined as the SCALAR TRIPLE PRODUCT of unit vectors in a right-handed coordinate system,

$$\epsilon_{ijk} \equiv \hat{\mathbf{x}}_i \cdot (\hat{\mathbf{x}}_j \times \hat{\mathbf{x}}_k). \tag{6}$$

The symbol can also be interpreted as a TENSOR, in which case it is called the PERMUTATION TENSOR.

see also PERMUTATION TENSOR

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 132–133, 1985.

Permutation Tensor

A PSEUDOTENSOR which is ANTISYMMETRIC under the interchange of any two slots. Recalling the definition of the PERMUTATION SYMBOL in terms of a SCALAR TRIPLE PRODUCT of the Cartesian unit vectors,

$$\epsilon_{ijk} \equiv \hat{\mathbf{x}}_i \cdot (\hat{\mathbf{x}}_j \times \hat{\mathbf{x}}_k) = [\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j, \hat{\mathbf{x}}_k], \qquad (1)$$

the pseudotensor is a generalization to an arbitrary BASIS defined by

$$\epsilon_{\alpha\beta\cdots\mu} = \sqrt{|g|}\,[\alpha,\beta,\ldots,\mu] \qquad (2)$$

$$\epsilon^{\alpha\beta\cdots\mu} = \frac{[\alpha,\beta,\ldots,\mu]}{\sqrt{|g|}}, \qquad (3)$$

where

$$[\alpha,\beta,\ldots,\mu] = \begin{cases} 1 & \text{the arguments are an even permutation} \\ -1 & \text{the arguments are an odd permutation} \\ 0 & \text{two or more arguments are equal,} \end{cases} \qquad (4)$$

and $g \equiv \det(g_{\alpha\beta})$, where $g_{\alpha\beta}$ is the METRIC TENSOR. $\epsilon(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is NONZERO IFF the VECTORS are LINEARLY INDEPENDENT.

see also PERMUTATION SYMBOL, SCALAR TRIPLE PRODUCT

Peron Integral

see DENJOY INTEGRAL

Perpendicular

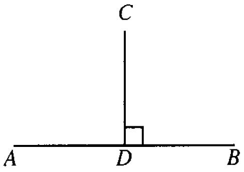

Two lines, vectors, planes, etc., are said to be perpendicular if they meet at a RIGHT ANGLE. In $\mathbb{R}^n$, two VECTORS **A** and **B** are PERPENDICULAR if their DOT PRODUCT

$$\mathbf{A} \cdot \mathbf{B} = 0.$$

In $\mathbb{R}^2$, a LINE with SLOPE $m_2 = -1/m_1$ is PERPENDICULAR to a LINE with SLOPE m_1. Perpendicular objects are sometimes said to be "orthogonal."

In the above figure, the LINE SEGMENT AB is perpendicular to the LINE SEGMENT CD. This relationship is commonly denoted with a small SQUARE at the vertex where perpendicular objects meet, as shown above.

see also ORTHOGONAL VECTORS, PARALLEL, PERPENDICULAR BISECTOR, PERPENDICULAR FOOT, RIGHT ANGLE

Perpendicular Bisector

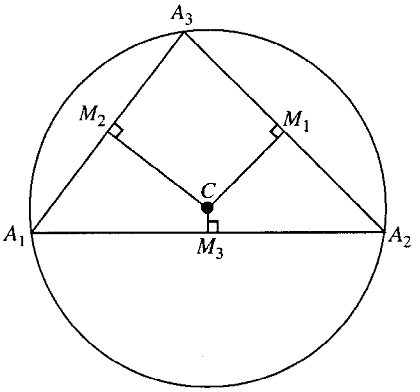

The perpendicular bisectors of a TRIANGLE $\Delta A_1A_2A_3$ are lines passing through the MIDPOINT M_i of each side which are PERPENDICULAR to the given side. A TRIANGLE'S three perpendicular bisectors meet at a point C known as the CIRCUMCENTER (which is also the center of the TRIANGLE'S CIRCUMCIRCLE).

see also CIRCUMCENTER, MIDPOINT, PERPENDICULAR, PERPENDICULAR FOOT

Perpendicular Foot

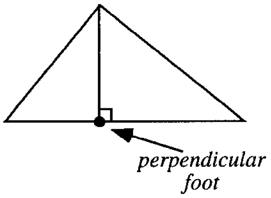

The FOOT of the PERPENDICULAR is the point on the leg opposite a given vertex of a TRIANGLE at which the PERPENDICULAR passing through that vertex intersects the side. The length of the LINE SEGMENT front vertex to perpendicular foot is called the ALTITUDE of the TRIANGLE.

see also ALTITUDE, FOOT, PERPENDICULAR, PERPENDICULAR BISECTOR

Perrin Pseudoprime

If p is PRIME, then $p|P(p)$, where $P(p)$ is a member of the PERRIN SEQUENCE 0, 2, 3, 2, 5, 5, 7, 10, 12, 17, ... (Sloane's A001608). A Perrin pseudoprime is a COMPOSITE NUMBER n such that $n|P(n)$. Several "unrestricted" Perrin pseudoprimes are known, the smallest of which are 271441, 904631, 16532714, 24658561, ... (Sloane's A013998).

Adams and Shanks (1982) discovered the smallest unrestricted Perrin pseudoprime after unsuccessful searches by Perrin (1899), Malo (1900), Escot (1901), and Jarden (1966). (Stewart's 1996 article stating no Perrin pseudoprimes were known was in error.)

Grantham (1996) generalized the definition of Perrin pseudoprime with parameters (r, s) to be an ODD COMPOSITE NUMBER n for which either

1. $(\Delta/n) = 1$ and n has an S-SIGNATURE, or
2. $(\Delta/n) = -1$ and n has a Q-SIGNATURE,

where (a/b) is the JACOBI SYMBOL. All the 55 Perrin pseudoprimes less than 50×10^9 have been computed by Kurtz *et al.* (1986). All have S-SIGNATURE, and form the sequence Sloane calls "restricted" Perrin pseudoprimes: 27664033, 46672291, 102690901, ... (Sloane's A018187).

see also PERRIN SEQUENCE, PSEUDOPRIME

References

Adams, W. W. "Characterizing Pseudoprimes for Third-Order Linear Recurrence Sequences." *Math Comput.* **48**, 1–15, 1987.

Adams, W. and Shanks, D. "Strong Primality Tests that Are Not Sufficient." *Math. Comput.* **39**, 255–300, 1982.

Bach, E. and Shallit, J. *Algorithmic Number Theory, Vol. 1: Efficient Algorithms.* Cambridge, MA: MIT Press, p. 305, 1996.

Escot, E.-B. "Solution to Item 1484." *L'Intermédiare des Math.* **8**, 63–64, 1901.

Grantham, J. "Frobenius Pseudoprimes." `http://www.clark.net/pub/grantham/pseudo/pseudo.ps`

Holzbaur, C. "Perrin Pseudoprimes." `http://ftp.ai.univie.ac.at/perrin.html`.

Jarden, D. *Recurring Sequences.* Jerusalem: Riveon Lematematika, 1966.

Kurtz, G. C.; Shanks, D.; and Williams, H. C. "Fast Primality Tests for Numbers Less than $50 \cdot 10^9$." *Math. Comput.* **46**, 691–701, 1986.

Malo, E. *L'Intermédiare des Math.* **7**, 281 and 312, 1900.

Perrin, R. "Item 1484." *L'Intermédiare des Math.* **6**, 76–77, 1899.

Ribenboim, P. *The New Book of Prime Number Records, 3rd ed.* New York: Springer-Verlag, p. 135, 1996.

Sloane, N. J. A. Sequences A013998, A018187, and A001608/M0429 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Stewart, I. "Tales of a Neglected Number." *Sci. Amer.* **274**, 102–103, June 1996.

Perrin Sequence

The INTEGER SEQUENCE defined by the recurrence

$$P(n) = P(n-2) + P(n-3) \tag{1}$$

with the initial conditions $P(0) = 3$, $P(1) = 0$, $P(2) = 2$. The first few terms are 0, 2, 3, 2, 5, 5, 7, 10, 12, 17, ... (Sloane's A001608). $P(n)$ is the solution of a third-order linear homogeneous DIFFERENCE EQUATION having characteristic equation

$$x^3 - x - 1 = 0, \tag{2}$$

discriminant -23, and ROOTS

$$\alpha \approx 1.324717957 \tag{3}$$

$$\beta \approx -0.6623589786 + 0.5622795121i \tag{4}$$

$$\gamma \approx -0.6623589786 - 0.5622795121i. \tag{5}$$

The solution is then

$$A(n) = \alpha^n + \beta^n + \gamma^n, \tag{6}$$

where

$$A(n) \sim \alpha^n. \tag{7}$$

Perrin (1899) investigated the sequence and noticed that if n is PRIME, then $n|P(n)$. The first statement of this fact is attributed to É. Lucas in 1876 by Stewart (1996). Perrin also searched for but did not find any COMPOSITE NUMBER n in the sequence such that $n|P(n)$. Such numbers are now known as PERRIN PSEUDOPRIMES. Malo (1900), Escot (1901), and Jarden (1966) subsequently investigated the series and also found no PERRIN PSEUDOPRIMES. Adams and Shanks (1982) subsequently found that 271,441 is such a number.

see also PADOVAN SEQUENCE, PERRIN PSEUDOPRIME, SIGNATURE (RECURRENCE RELATION)

References

Adams, W. and Shanks, D. "Strong Primality Tests that Are Not Sufficient." *Math. Comput.* **39**, 255–300, 1982.

Escot, E.-B. "Solution to Item 1484." *L'Intermédiare des Math.* **8**, 63–64, 1901.

Jarden, D. *Recurring Sequences.* Jerusalem: Riveon Lematematika, 1966.

Malo, E. *L'Intermédiare des Math.* **7**, 281 and 312, 1900.

Perrin, R. "Item 1484." *L'Intermédiare des Math.* **6**, 76–77, 1899.

Stewart, I. "Tales of a Neglected Number." *Sci. Amer.* **274**, 102–103, June 1996.

Sloane, N. J. A. Sequence A001608/M0429 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Perron-Frobenius Operator

An OPERATOR which describes the time evolution of densities in PHASE SPACE. The OPERATOR can be defined by

$$\rho_{n+1} = \tilde{L}\rho_n,$$

where ρ_n are the NATURAL DENSITIES after the nth iteration of a map f. This can be explicitly written as

$$\tilde{L}\rho(y) = \sum_{x \in f^{-1}(y)} \frac{\rho(x)}{|f'(x)|}.$$

References

Beck, C. and Schlögl, F. "Transfer Operator Methods." Ch. 17 in *Thermodynamics of Chaotic Systems.* Cambridge, England: Cambridge University Press, pp. 190–203, 1995.

Perron-Frobenius Theorem

If all elements a_{ij} of an IRREDUCIBLE MATRIX A are NONNEGATIVE, then $R = \min M_\lambda$ is an EIGENVALUE of A and all the EIGENVALUES of A lie on the DISK

$$|z| \leq R,$$

where, if $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a set of NONNEGATIVE numbers (which are not all zero),

$$M_\lambda = \inf \left\{ \mu : \mu\lambda_i > \sum_{j=1}^{n} |a_{ij}|\lambda_j, 1 \leq i \leq n \right\}$$

and $R = \min M_\lambda$. Furthermore, if A has exactly p EIGENVALUES ($p \le n$) on the CIRCLE $|z| = R$, then the set of all its EIGENVALUES is invariant under rotations by $2\pi/p$ about the ORIGIN.

see also WIELANDT'S THEOREM

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1121, 1979.

Perron's Theorem

If $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)$ is an arbitrary set of POSITIVE numbers, then all EIGENVALUES λ of the $n \times n$ MATRIX $\mathsf{A} = a_{ij}$ lie on the DISK $|z| \le M_\mu$, where

$$M_\mu = \max_{1\le i\le n} \sum_{j=1}^{n} \frac{\mu_j}{\mu_i}|a_{ij}|.$$

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1121, 1979.

Persistence

see ADDITIVE PERSISTENCE, MULTIPLICATIVE PERSISTENCE, PERSISTENT NUMBER, PERSISTENT PROCESS

Persistent Number

An n-persistent number is a POSITIVE INTEGER k which contains the digits 0, 1, ..., 9, and for which $2k$, ..., nk also share this property. No ∞-persistent numbers exist. However, the number $k = 1234567890$ is 2-persistent, since $2k = 2469135780$ but $3k = 3703703670$, and the number $k = 526315789473684210$ is 18-persistent. There exists at least one k-persistent number for each POSITIVE INTEGER k.

see also ADDITIVE PERSISTENCE, MULTIPLICATIVE PERSISTENCE

References

Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., pp. 15–18, 1991.

Persistent Process

A FRACTAL PROCESS for which $H > 1/2$, so $r > 0$.

see also ANTIPERSISTENT PROCESS, FRACTAL PROCESS

Perspective

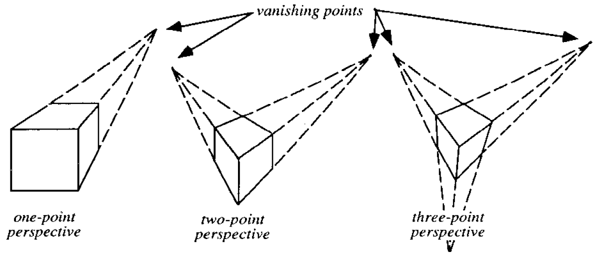

Perspective is the art and mathematics of realistically depicting 3-D objects in a 2-D plane. The study of the projection of objects in a plane is called PROJECTIVE GEOMETRY. The principles of perspective drawing were elucidated by the Florentine architect F. Brunelleschi (1377–1446). These rules are summarized by Dixon (1991):

1. The horizon appears as a line.
2. Straight lines in space appear as straight lines in the image.
3. Sets of PARALLEL lines meet at a VANISHING POINT.
4. Lines PARALLEL to the picture plane appear PARALLEL and therefore have no VANISHING POINT.

There is a graphical method for selecting vanishing points so that a CUBE or box appears to have the correct dimensions (Dixon 1991).

see also LEONARDO'S PARADOX, PERSPECTIVE AXIS, PERSPECTIVE CENTER, PERSPECTIVE COLLINEATION, PERSPECTIVE TRIANGLES, PERSPECTIVITY, PROJECTIVE GEOMETRY, VANISHING POINT, ZEEMAN'S PARADOX

References

de Vries, V. *Perspective.* New York: Dover, 1968.

Dixon, R. "Perspective Drawings." Ch. 3 in *Mathographics.* New York: Dover, pp. 79–88, 1991.

Parramon, J. M. *Perspective—How to Draw.* Barcelona, Spain: Parramon Editions, 1984.

Perspective Axis

The line joining the three collinear points of intersection of the extensions of corresponding sides in PERSPECTIVE TRIANGLES.

see also PERSPECTIVE CENTER, PERSPECTIVE TRIANGLES, SONDAT'S THEOREM

Perspective Center

The point at which the three LINES connecting the VERTICES of PERSPECTIVE TRIANGLES (from a point) CONCUR.

Perspective Collineation

A perspective collineation with center O and axis o is a COLLINEATION which leaves all lines through O and points of o invariant. Every perspective collineation is a PROJECTIVE COLLINEATION.

see also COLLINEATION, ELATION, HOMOLOGY (GEOMETRY), PROJECTIVE COLLINEATION

References

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 247–248, 1969.

Perspective Triangles

Two TRIANGLES are perspective from a line if the extensions of their three pairs of corresponding sides meet in COLLINEAR points. The line joining these points is called the PERSPECTIVE AXIS. Two TRIANGLES are perspective from a point if their three pairs of corresponding VERTICES are joined by lines which meet in a point of CONCURRENCE. This point is called the PERSPECTIVE CENTER. DESARGUES' THEOREM guarantees that if two TRIANGLES are perspective from a point, they are perspective from a line.

see also DESARGUES' THEOREM, HOMOTHETIC TRIANGLES, PARALOGIC TRIANGLES, PERSPECTIVE AXIS, PERSPECTIVE CENTER

Perspectivity

A correspondence between two RANGES that are sections of one PENCIL by two distinct lines.

see also PENCIL, PROJECTIVITY, RANGE (LINE SEGMENT)

Pesin Theory

A theory of linear HYPERBOLIC MAPS in which the leading constants do depend on the variable x.

Peter-Weyl Theorem

Establishes completeness for a REPRESENTATION.

References

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Peters Projection

A CYLINDRICAL equal-area projection that shifts the standard parallels to 45° or 47°.

see also CYLINDRICAL PROJECTION

References

Dana, P. H. "Map Projections." `http://www.utexas.edu/depts/grg/gcraft/notes/mapproj/mapproj.html`.

Petersen Graphs

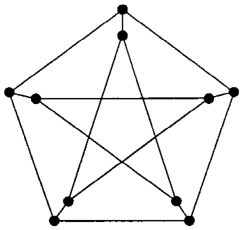

"The" Petersen graph is the GRAPH illustrated above possessing ten VERTICES all of whose nodes have DEGREE 3 (Saaty and Kainen 1986). The Petersen graph is the only smallest-girth graph which has no Tait coloring.

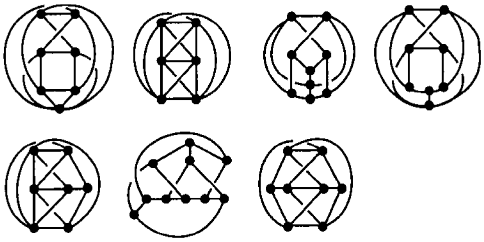

The seven graphs obtainable from the COMPLETE GRAPH K_6 by repeated triangle-Y exchanges are also called Petersen graphs, where the three EDGES forming the TRIANGLE are replaced by three EDGES and a new VERTEX that form a Y, and the reverse operation is also permitted. A GRAPH is intrinsically linked IFF it contains one of the seven Petersen graphs (Robertson *et al.* 1993).

see also HOFFMAN-SINGLETON GRAPH

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 221–222, 1994.

Robertson, N.; Seymour, P. D.; and Thomas, R. "Linkless Embeddings of Graphs in 3-Space." *Bull. Amer. Math. Soc.* **28**, 84–89, 1993.

Saaty, T. L. and Kainen, P. C. *The Four-Color Problem: Assaults and Conquest.* New York: Dover, p. 102, 1986.

Petersen-Shoute Theorem

1. If ΔABC and $\Delta A'B'C'$ are two directly similar triangles, while $\Delta AA'A''$, $\Delta BB'B''$, and $\Delta CC'C''$ are three directly similar triangles, then $\Delta A''B''C''$ is directly similar to ΔABC.
2. When all the points P on AB are related by a SIMILARITY TRANSFORMATION to all the points P' on $A'B'$, the points dividing the segment PP' in a given ratio are distant and collinear, or else they coincide.

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 95–100, 1967.

Petrie Polygon

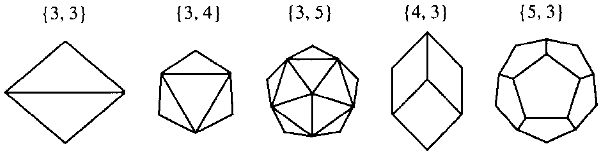

A skew POLYGON such that every two consecutive sides (but no three) belong to a face of a regular POLYHEDRON. Every finite POLYHEDRON can be orthogonally projected onto a plane in such a way that one Petrie polygon becomes a REGULAR POLYGON with the remainder of the projection interior to it. The Petrie polygon of the POLYHEDRON $\{p, q\}$ has h sides, where

$$\cos^2\left(\frac{\pi}{h}\right) = \cos^2\left(\frac{\pi}{p}\right) + \cos^2\left(\frac{\pi}{q}\right).$$

The Petrie polygons shown above correspond to the PLATONIC SOLIDS.

see also PLATONIC SOLID, REGULAR POLYGON

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 135, 1987.

Coxeter, H. S. M. "Petrie Polygons." §2.6 in *Regular Polytopes, 3rd ed.* New York: Dover, pp.24–25, 1973.

Petrov Notation

A TENSOR notation which considers the RIEMANN TENSOR $R_{\lambda\mu\nu\kappa}$ as a matrix $R_{(\lambda\mu)(\nu\kappa)}$ with indices $\lambda\mu$ and $\nu\kappa$.

References

Weinberg, S. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity.* New York: Wiley, p. 142, 1972.

Pfaffian Form

A 1-FORM

$$\omega = \sum_{i=1}^{n} a_i(x)\,dx_i$$

such that

$$\omega = 0.$$

References

Knuth, D. E. "Overlapping Pfaffians." *Electronic J. Combinatorics* **3**, No. 2, R5, 1–13, 1996. `http://www.combinatorics.org/Volume_3/volume3_2.html#R5`.

Phase

The angular position of a quantity. For example, the phase of a function $\cos(\omega t + \phi_0)$ as a function of time is

$$\phi(t) = \omega t + \phi_0.$$

The ARGUMENT of a COMPLEX NUMBER is sometimes also called the phase.

see also ARGUMENT (COMPLEX NUMBER), COMPLEX NUMBER, PHASOR, RETARDANCE

Phase Space

For a function or object with n DEGREES OF FREEDOM, the n-D SPACE which is accessible to the function or object is called its phase space.

see also WORLD LINE

Phase Transition

see RANDOM GRAPH

Phasor

The representation, beloved of engineers and physicists, of a COMPLEX NUMBER in terms of a COMPLEX exponential

$$x + iy = |z|e^{i\phi}, \tag{1}$$

where i (called j by engineers) is the IMAGINARY NUMBER and the MODULUS and ARGUMENT (also called PHASE) are

$$|z| = \sqrt{x^2 + y^2} \tag{2}$$

$$\phi = \tan^{-1}\left(\frac{y}{x}\right). \tag{3}$$

Here, ϕ is the counterclockwise ANGLE from the POSITIVE REAL axis. In the degenerate case when $x = 0$,

$$\phi = \begin{cases} -\frac{1}{2}\pi & \text{if } y < 0 \\ \text{undefined} & \text{if } y = 0 \\ \frac{1}{2}\pi & \text{if } y > 0. \end{cases} \tag{4}$$

It is trivially true that

$$\sum_i \Re[\psi_i] = \Re\left[\sum_i \psi_i\right]. \tag{5}$$

Now consider a SCALAR FUNCTION $\psi \equiv \psi_0 e^{i\phi}$. Then

$$\begin{aligned} I \equiv [\Re(\psi)]^2 &= [\tfrac{1}{2}(\psi + \psi^*)]^2 = \tfrac{1}{4}(\psi + \psi^*)^2 \\ &= \tfrac{1}{4}(\psi^2 + 2\psi\psi^* + \psi^{*2}). \end{aligned} \tag{6}$$

Look at the time averages of each term,

$$\left\langle \psi^2 \right\rangle = \left\langle {\psi_0}^2 e^{2i\phi} \right\rangle = {\psi_0}^2 \left\langle e^{2i\phi} \right\rangle = 0 \tag{7}$$

$$\left\langle \psi\psi^* \right\rangle = \left\langle \psi_0 e^{i\phi} \psi_0 e^{-i\phi} \right\rangle = {\psi_0}^2 = |\psi|^2 \tag{8}$$

$$\left\langle \psi^{*2} \right\rangle = \left\langle {\psi_0}^2 e^{-2i\phi} \right\rangle = {\psi_0}^2 \left\langle e^{-2i\phi} \right\rangle = 0. \tag{9}$$

Therefore,

$$\langle I \rangle = \tfrac{1}{2}|\psi|^2. \tag{10}$$

Consider now two scalar functions

$$\psi_1 \equiv \psi_{1,0} e^{i(kr_1 + \phi_1)} \tag{11}$$

$$\psi_2 \equiv \psi_{2,0} e^{i(kr_2 + \phi_2)}. \tag{12}$$

Then

$$\begin{aligned} I \equiv [\Re(\psi_1) + \Re(\psi_2)]^2 &= \tfrac{1}{4}[(\psi_1 + {\psi_1}^*) + (\psi_2 + {\psi_2}^*)]^2 \\ &= \tfrac{1}{4}[(\psi_1 + {\psi_1}^*)^2 + (\psi_2 + {\psi_2}^*)^2 \\ &\quad + 2(\psi_1\psi_2 + \psi_1{\psi_2}^* + {\psi_1}^*\psi_2 + {\psi_1}^*{\psi_2}^*)] \end{aligned} \tag{13}$$

$$\begin{aligned} \langle I \rangle &= \tfrac{1}{4}[2\psi_1{\psi_1}^* + 2\psi_2{\psi_2}^* + 2\psi_1{\psi_2}^* + 2{\psi_1}^*\psi_2] \\ &= \tfrac{1}{2}[\psi_1({\psi_1}^* + {\psi_2}^*) + \psi_2({\psi_1}^* + {\psi_2}^*)] \\ &= \tfrac{1}{2}(\psi_1 + \psi_2)({\psi_1}^* + {\psi_2}^*) = \tfrac{1}{2}|\psi_1 + \psi_2|^2. \end{aligned} \tag{14}$$

In general,

$$\langle I \rangle = \frac{1}{2}\left|\sum_{i=1}^{n} \psi_i\right|^2. \tag{15}$$

see also AFFIX, ARGUMENT (COMPLEX NUMBER), COMPLEX MULTIPLICATION, COMPLEX NUMBER, MODULUS (COMPLEX NUMBER), PHASE

Phi Curve

An ADJOINT CURVE which bears a special relation to the base curve.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 310, 1959.

Phi Number System

For every POSITIVE INTEGER n, there is a corresponding finite sequence of distinct INTEGERS $k_1, \ldots, k_m$ such that

$$n = \phi^{k_1} + \ldots + \phi^{k_m},$$

where ϕ is the GOLDEN MEAN.

References

Bergman, G. "A Number System with an Irrational Base." *Math. Mag.* **31**, 98–110, 1957.
Knuth, D. *The Art of Computer Programming, Vol. 1: Fundamental Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, 1973.
Rousseau, C. "The Phi Number System Revisited." *Math. Mag.* **68**, 283–284, 1995.

Phragmén-Lindêlöf Theorem

Let $f(z)$ be an ANALYTIC FUNCTION in an angular domain $W : |\arg z| < \alpha\pi/2$. Suppose there is a constant M such that for each $\epsilon > 0$, each finite boundary point has a NEIGHBORHOOD such that $|f(z)| < M + \epsilon$ on the intersection of D with this NEIGHBORHOOD, and that for some POSITIVE number $\beta > \alpha$ for sufficiently large $|z|$, the INEQUALITY $|f(z)| < \exp(|z|^{1/\beta})$ holds. Then $|f(z)| \leq M$ in D.

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 160, 1980.

Phyllotaxis

The beautiful arrangement of leaves in some plants, called phyllotaxis, obeys a number of subtle mathematical relationships. For instance, the florets in the head of a sunflower form two oppositely directed spirals: 55 of them clockwise and 34 counterclockwise. Surprisingly, these numbers are consecutive FIBONACCI NUMBERS. The ratios of alternate FIBONACCI NUMBERS are given by the convergents to ϕ^{-2}, where ϕ is the GOLDEN RATIO, and are said to measure the fraction of a turn between successive leaves on the stalk of a plant: 1/2 for elm and linden, 1/3 for beech and hazel, 2/5 for oak and apple, 3/8 for poplar and rose, 5/13 for willow and almond, etc. (Coxeter 1969, Ball and Coxeter 1987). A similar phenomenon occurs for DAISIES, pineapples, pinecones, cauliflowers, and so on.

Lilies, irises, and the trillium have three petals; columbines, buttercups, larkspur, and wild rose have five petals; delphiniums, bloodroot, and cosmos have eight petals; corn marigolds have 13 petals; asters have 21 petals; and daisies have 34, 55, or 84 petals—all FIBONACCI NUMBERS.

see also DAISY, FIBONACCI NUMBER, SPIRAL

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 56–57, 1987.
Church, A. H. *The Relation of Phyllotaxis to Mechanical Laws.* London: Williams and Norgate, 1904.
Church, A. H. *On the Interpretation of Phenomena of Phyllotaxis.* Riverside, NJ: Hafner, 1968.
Conway, J. H. and Guy, R. K. "Phyllotaxis." In *The Book of Numbers.* New York: Springer-Verlag, pp. 113–125, 1995.
Coxeter, H. S. M. "The Golden Section and Phyllotaxis." Ch. 11 in *Introduction to Geometry, 2nd ed.* New York: Wiley, 1969.
Coxeter, H. S. M. "The Golden Section, Phyllotaxis, and Wythoff's Game." *Scripta Mathematica* **19**, 135–143, 1953.
Dixon, R. *Mathographics.* New York: Dover, 1991.
Douady, S. and Couder, Y. "Phyllotaxis as a Self-Organized Growth Process." In *Growth Patterns in Physical Sciences and Biology* (Ed. Juan M. Garcia-Ruiz *et al.*). Plenum Press, 1993.
Hunter, J. A. H. and Madachy, J. S. *Mathematical Diversions.* New York: Dover, pp. 20–22, 1975.
Jean, R. V. *Phyllotaxis: A Systematic Study in Plant Morphogenesis.* New York: Cambridge University Press, 1994.
Pappas, T. "The Fibonacci Sequence & Nature." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 222–225, 1989.
Prusinkiewicz, P. and Lindenmayer, A. *The Algorithmic Beauty of Plants.* New York: Springer-Verlag, 1990.
Stewart, I. "Daisy, Daisy, Give Me Your Answer, Do." *Sci. Amer.* **200**, 96–99, Jan. 1995.
Thompson, D. W. *On Growth and Form.* Cambridge, England: Cambridge University Press, 1952.

Pi

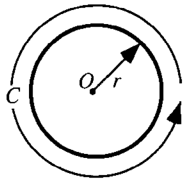

A REAL NUMBER denoted π which is defined as the ratio of a CIRCLE'S CIRCUMFERENCE C to its DIAMETER $d = 2r$,

$$\pi \equiv \frac{C}{d} = \frac{C}{2r} \tag{1}$$

It is equal to

$$\pi = 3.141592653589793238462643383279502884197\ldots \tag{2}$$

(Sloane's A000796). π has recently (August 1997) been computed to a world record $51,539,600,000 \approx 3 \cdot 2^{34}$ DECIMAL DIGITS by Y. Kanada. This calculation was done using Borwein's fourth-order convergent algorithm and required 29 hours on a massively parallel 1024-processor Hitachi SR2201 supercomputer. It was checked in 37 hours using the BRENT-SALAMIN FORMULA on the same machine.

The SIMPLE CONTINUED FRACTION for π, which gives the "best" approximation of a given order, is [3, 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, 1, 14, 2, 1, 1, 2, 2, 2, 2, ...] (Sloane's A001203). The very large term 292 means that the CONVERGENT

$$[3,7,15,1] = [3,7,16] = \tfrac{355}{113} = 3.14159292\ldots \tag{3}$$

is an extremely good approximation. The first few CONVERGENTS are 22/7, 333/106, 355/113, 103993/33102, 104348/33215, ... (Sloane's A002485 and A002486). The first occurrences of n in the CONTINUED FRACTION are 4, 9, 1, 30, 40, 32, 2, 44, 130, 100, ... (Sloane's A032523).

Gosper has computed 17,001,303 terms of π's CONTINUED FRACTION (Gosper 1977, Ball and Coxeter 1987), although the computer on which the numbers are stored may no longer be functional (Gosper, pers. comm., 1998). According to Gosper, a typical CONTINUED FRACTION term carries only slightly more significance than a decimal DIGIT. The sequence of increasing terms in the CONTINUED FRACTION is 3, 7, 15, 292, 436, 20776, ... (Sloane's A033089), occurring at positions 1, 2, 3, 5, 308, 432, ... (Sloane's A033090). In the first 26,491 terms of the CONTINUED FRACTION (counting 3 as the 0th), the only five-DIGIT terms are 20,776 (the 431st), 19,055 (15,543rd), and 19,308 (23,398th) (Beeler *et al.* 1972, Item 140). The first 6-DIGIT term is 528,210 (the 267,314th), and the first 8-DIGIT term is 12,996,958 (453,294th). The term having the largest known value is the whopping 9-DIGIT 87,878,3625 (the 11,504,931st term).

The SIMPLE CONTINUED FRACTION for π does not show any obvious patterns, but clear patterns *do* emerge in the beautiful non-simple CONTINUED FRACTIONS

$$\frac{4}{\pi} = 1 + \cfrac{1^2}{2 + \cfrac{3^2}{2 + \cfrac{5^2}{2 + \cfrac{7^2}{2 + \ldots}}}} \tag{4}$$

(Brouckner), giving convergents 1, 3/2, 15/13, 105/76, 315/263, ... (Sloane's A025547 and A007509) and

$$\frac{\pi}{2} = 1 - \cfrac{1}{3 - \cfrac{2\cdot 3}{1 - \cfrac{1\cdot 2}{3 - \cfrac{4\cdot 5}{1 - \cfrac{3\cdot 4}{3 - \cfrac{6\cdot 7}{1 - \cfrac{5\cdot 6}{3 - \ldots}}}}}}} \tag{5}$$

(Stern 1833), giving convergents 1, 2/3, 4/3, 16/15, 64/45, 128/105, ... (Sloane's A001901 and A046126).

π crops up in all sorts of unexpected places in mathematics besides CIRCLES and SPHERES. For example, it occurs in the normalization of the GAUSSIAN DISTRIBUTION, in the distribution of PRIMES, in the construction of numbers which are very close to INTEGERS (the RAMANUJAN CONSTANT), and in the probability that a pin dropped on a set of PARALLEL lines intersects a line (BUFFON'S NEEDLE PROBLEM). Pi also appears as the average ratio of the actual length and the direct distance between source and mouth in a meandering river (Støllum 1996, Singh 1997).

A brief history of NOTATION for pi is given by Castellanos (1988). π is sometimes known as LUDOLPH'S CONSTANT after Ludolph van Ceulen (1539–1610), a Dutch π calculator. The symbol π was first used by William Jones in 1706, and subsequently adopted by Euler. In *Measurement of a Circle,* Archimedes (ca. 225 BC) obtained the first rigorous approximation by INSCRIBING and CIRCUMSCRIBING $6\cdot 2^n$-gons on a CIRCLE using the ARCHIMEDES ALGORITHM. Using $n = 4$ (a 96-gon), Archimedes obtained

$$3 + \tfrac{10}{71} < \pi < 3 + \tfrac{1}{7} \tag{6}$$

(Shanks 1993, p. 140).

The Bible contains two references (I Kings 7:23 and Chronicles 4:2) which give a value of 3 for π. It should be mentioned, however, that both instances refer to a value obtained from physical measurements and, as such, are probably well within the bounds of experimental uncertainty. I Kings 7:23 states, "Also he made a molten sea of ten Cubits from brim to brim, round in compass, and five cubits in height thereof; and a line thirty cubits did compass it round about." This implies $\pi = C/d = 30/10 = 3$. The Babylonians gave an estimate of π as $3 + 1/8 = 3.125$. The Egyptians did better still, obtaining $2^8/3^4 = 3.1605\ldots$ in the Rhind papyrus, and 22/7 elsewhere. The Chinese geometers, however, did best of all, rigorously deriving π to 6 decimal places.

A method similar to Archimedes' can be used to estimate π by starting with an n-gon and then relating the AREA of subsequent $2n$-gons. Let β be the ANGLE from the center of one of the POLYGON's segments,

$$\beta = \tfrac{1}{4}(n-3)\pi. \tag{7}$$

Then

$$\pi = \frac{\frac{1}{2}n\sin(2\beta)}{\cos\beta\cos\left(\frac{\beta}{2}\right)\cos\left(\frac{\beta}{2^2}\right)\cos\left(\frac{\beta}{2^3}\right)\cdots} \tag{8}$$

(Beckmann 1989, pp. 92–94). Viète (1593) was the first to give an exact expression for π by taking $n = 4$ in the above expression, giving

$$\cos\beta = \sin\beta = \frac{1}{\sqrt{2}} = \tfrac{1}{2}\sqrt{2}, \tag{9}$$

which leads to an INFINITE PRODUCT of CONTINUED SQUARE ROOTS,

$$\frac{2}{\pi} = \sqrt{\tfrac{1}{2}}\sqrt{\tfrac{1}{2}+\sqrt{\tfrac{1}{2}}}\sqrt{\tfrac{1}{2}+\sqrt{\tfrac{1}{2}+\sqrt{\tfrac{1}{2}}}}\cdots \tag{10}$$

(Beckmann 1989, p. 95). However, this expression was not rigorously proved to converge until Rudio (1892). Another exact FORMULA is MACHIN'S FORMULA, which is

$$\frac{\pi}{4} = 4\tan^{-1}(\tfrac{1}{5}) - \tan^{-1}(\tfrac{1}{239}). \tag{11}$$

There are three other MACHIN-LIKE FORMULAS, as well as other FORMULAS with more terms. An interesting INFINITE PRODUCT formula due to Euler which relates π and the nth PRIME p_n is

$$\pi = \frac{2}{\prod_{i=n}^{\infty}\left[1+\frac{\sin(\frac{1}{2}\pi p_n)}{p_n}\right]} \tag{12}$$

$$= \frac{2}{\prod_{i=n}^{\infty}\left[1+\frac{(-1)^{(p_n-1)/2}}{p_n}\right]} \tag{13}$$

(Blatner 1997, p. 119), plotted below as a function of the number of terms in the product.

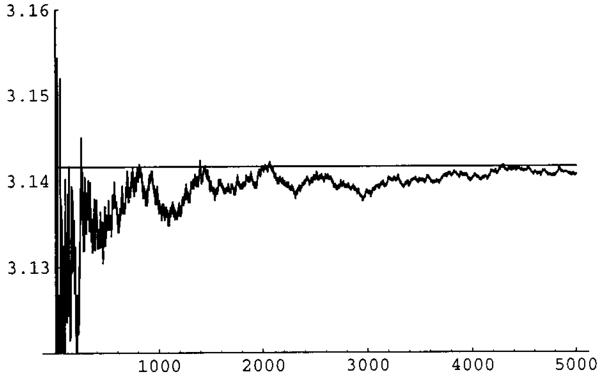

The AREA and CIRCUMFERENCE of the UNIT CIRCLE are given by

$$A = \pi = 4\int_0^1 \sqrt{1-x^2}\,dx \tag{14}$$

$$= \lim_{n\to\infty} \frac{4}{n^2}\sum_{k=0}^{n}\sqrt{n^2-k^2} \tag{15}$$

and

$$C = 2\pi = 4\int_0^1 \frac{dx}{\sqrt{1-x^2}} \tag{16}$$

$$= 4\int_0^1 \sqrt{1+\left(\frac{d}{dx}\sqrt{1-x^2}\right)^2}\,dx. \tag{17}$$

The SURFACE AREA and VOLUME of the unit SPHERE are

$$S = 4\pi \tag{18}$$

$$V = \tfrac{4}{3}\pi. \tag{19}$$

π is known to be IRRATIONAL (Lambert 1761, Legendre 1794) and even TRANSCENDENTAL (Lindemann 1882). Incidentally, Lindemann's proof of the transcendence of π also proved that the GEOMETRIC PROBLEM OF ANTIQUITY known as CIRCLE SQUARING is impossible. A simplified, but still difficult, version of Lindemann's proof is given by Klein (1955).

It is also known that π is not a LIOUVILLE NUMBER (Mahler 1953). In 1974, M. Mignotte showed that

$$\left|\pi - \frac{p}{q}\right| \leq q^{-20} \tag{20}$$

has only a finite number of solutions in INTEGERS (Le Lionnais 1983, p. 50). This result was subsequently improved by Chudnovsky and Chudnovsky (1984) who showed that

$$\left|\pi - \frac{p}{q}\right| > q^{-14.65}, \tag{21}$$

although it is likely that the exponent can be reduced to $2+\epsilon$, where ϵ is an infinitesimally small number (Borwein *et al.* 1989). It is *not* known if π is NORMAL (Wagon 1985), although the first 30 million DIGITS are very UNIFORMLY DISTRIBUTED (Bailey 1988). The following distribution is found for the first n DIGITS of $\pi-3$. It shows no statistically SIGNIFICANT departure from a UNIFORM DISTRIBUTION (technically, in the CHI-SQUARED TEST, it has a value of ${\chi_s}^2 = 5.60$ for the first 5×10^{10} terms).

digit	1×10^5	1×10^6	6×10^9	5×10^{10}
0	9,999	99,959	599,963,005	5,000,012,647
1	10,137	99,758	600,033,260	4,999,986,263
2	9,908	100,026	599,999,169	5,000,020,237
3	10,025	100,229	600,000,243	4,999,914,405
4	9,971	100,230	599,957,439	5,000,023,598
5	10,026	100,359	600,017,176	4,999,991,499
6	10,029	99,548	600,016,588	4,999,928,368
7	10,025	99,800	600,009,044	5,000,014,860
8	9,978	99,985	599,987,038	5,000,117,637
9	9,902	100,106	600,017,038	4,999,990,486

The digits of $1/\pi$ are also very uniformly distributed (${\chi_s}^2 = 7.04$), as shown in the following table.

digit	5×10^{10}
0	4,999,969,955
1	5,000,113,699
2	4,999,987,893
3	5,000,040,906
4	4,999,985,863
5	4,999,977,583
6	4,999,990,916
7	4,999,985,552
8	4,999,881,183
9	5,000,066,450

It is not known if $\pi + e$, π/e, or $\ln\pi$ are IRRATIONAL. However, it is known that they cannot satisfy any POLYNOMIAL equation of degree ≤ 8 with INTEGER COEFFICIENTS of average size 10^9 (Bailey 1988, Borwein *et al.* 1989).

π satisfies the INEQUALITY

$$\left(1+\frac{1}{\pi}\right)^{\pi+1} \approx 3.14097 < \pi. \tag{22}$$

Beginning with any POSITIVE INTEGER n, round up to the nearest multiple of $n-1$, then up to the nearest multiple of $n-2$, and so on, up to the nearest multiple of 1. Let $f(n)$ denote the result. Then the ratio

$$\lim_{n\to\infty} \frac{n^2}{f(n)} = \pi \tag{23}$$

(Brown). David (1957) credits this result to Jabotinski and Erdős and gives the more precise asymptotic result

$$f(n) = \frac{n^2}{\pi} + \mathcal{O}(n^{4/3}). \tag{24}$$

The first few numbers in the sequence $\{f(n)\}$ are 1, 2, 4, 6, 10, 12, 18, 22, 30, 34, ... (Sloane's A002491).

A particular case of the WALLIS FORMULA gives

$$\frac{\pi}{2} = \prod_{n=1}^{\infty} \left[\frac{(2n)^2}{(2n-1)(2n+1)}\right] = \frac{2\cdot 2}{1\cdot 3}\frac{4\cdot 4}{3\cdot 5}\frac{6\cdot 6}{5\cdot 7}\cdots. \tag{25}$$

This formula can also be written

$$\lim_{n\to\infty} \frac{2^{4n}}{n\binom{2n}{n}^2} = \pi \lim_{n\to\infty} \frac{n[\Gamma(n)]^2}{[\Gamma(\frac{1}{2}+n)]^2} = \pi, \tag{26}$$

where $\binom{n}{k}$ denotes a BINOMIAL COEFFICIENT and $\Gamma(x)$ is the GAMMA FUNCTION (Knopp 1990). Euler obtained

$$\pi = \sqrt{6\left(1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\ldots\right)}, \tag{27}$$

which follows from the special value of the RIEMANN ZETA FUNCTION $\zeta(2) = \pi^2/6$. Similar FORMULAS follow from $\zeta(2n)$ for all POSITIVE INTEGERS n. Gregory and Leibniz found

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} + \ldots, \tag{28}$$

which is sometimes known as GREGORY'S FORMULA. The error after the nth term of this series in GREGORY'S FORMULA is larger than $(2n)^{-1}$ so this sum converges so slowly that 300 terms are not sufficient to calculate π correctly to two decimal places! However, it can be transformed to

$$\pi = \sum_{k=1}^{\infty} \frac{3^k-1}{4^k}\zeta(k+1), \tag{29}$$

where $\zeta(z)$ is the RIEMANN ZETA FUNCTION (Vardi 1991, pp. 157–158; Flajolet and Vardi 1996), so that the error after k terms is $\approx (3/4)^k$. Newton used

$$\pi = \tfrac{3}{4}\sqrt{3} + 24\int_0^{1/4} \sqrt{x-x^2}\,dx \tag{30}$$

$$= \frac{3\sqrt{3}}{4} + 24\left(\frac{1}{12} - \frac{1}{5\cdot 2^5} - \frac{1}{28\cdot 2^7} - \frac{1}{72\cdot 2^9} - \ldots\right) \tag{31}$$

(Borwein *et al.* 1989). Using Euler's CONVERGENCE IMPROVEMENT transformation gives

$$\frac{\pi}{2} = \frac{1}{2}\sum_{n=0}^{\infty} \frac{(n!)^2 2^{n+1}}{(2n+1)!} = \sum_{n=0}^{\infty} \frac{n!}{(2n+1)!!}$$

$$= 1 + \frac{1}{3} + \frac{1\cdot 2}{3\cdot 5} + \frac{1\cdot 2\cdot 3}{3\cdot 5\cdot 7} + \ldots \tag{32}$$

$$= 1 + \frac{1}{3}\left(1+\frac{2}{5}\left(1+\frac{3}{7}\left(1+\frac{4}{9}(1+\ldots)\right)\right)\right) \tag{33}$$

(Beeler *et al.* 1972, Item 120). This corresponds to plugging $x = 1/\sqrt{2}$ into the POWER SERIES for the HYPERGEOMETRIC FUNCTION ${}_2F_1(a,b;c;x)$,

$$\frac{\sin^{-1} x}{\sqrt{1-x^2}} = \sum_{i=0}^{\infty} \frac{(2x)^{2i+1}(i!)^2}{2(2i+1)!} = {}_2F_1(1,1;\tfrac{3}{2};x^2)x. \tag{34}$$

Despite the convergence improvement, series (33) converges at only one bit/term. At the cost of a SQUARE ROOT, Gosper has noted that $x = 1/2$ gives 2 bits/term,

$$\tfrac{1}{9}\sqrt{3}\pi = \frac{1}{2}\sum_{i=1}^{\infty} \frac{(i!)^2}{(2i+1)!}, \tag{35}$$

and $x = \sin(\pi/10)$ gives almost 3.39 bits/term,

$$\frac{\pi}{5\sqrt{\phi+2}} = \frac{1}{2}\sum_{i=0}^{\infty} \frac{(i!)^2}{\phi^{2i+1}(2i+1)!}, \tag{36}$$

where ϕ is the GOLDEN RATIO. Gosper also obtained

$$\pi = 3 + \frac{1}{60}\left(8 + \frac{2\cdot 3}{7\cdot 8\cdot 3}\left(13 + \frac{3\cdot 5}{10\cdot 11\cdot 3}\right.\right.$$
$$\left.\left.\times\left(18 + \frac{4\cdot 7}{13\cdot 14\cdot 3}(23+\ldots)\right)\right)\right). \quad (37)$$

An infinite sum due to Ramanujan is

$$\frac{1}{\pi} = \sum_{n=0}^{\infty} \binom{2n}{n}^3 \frac{42n+5}{2^{12n+4}} \quad (38)$$

(Borwein *et al.* 1989). Further sums are given in Ramanujan (1913–14),

$$\frac{4}{\pi} = \sum_{n=0}^{\infty} \frac{(-1)^n(1123+21460n)(2n-1)!!(4n-1)!!}{882^{2n+1}32^n(n!)^3} \quad (39)$$

and

$$\frac{1}{\pi} = \sqrt{8}\sum_{n=0}^{\infty} \frac{(1103+26390n)(2n-1)!!(4n-1)!!}{99^{4n+2}32^n(n!)^3}$$
$$= \frac{\sqrt{8}}{9801}\sum_{n=0}^{\infty}\frac{(4n)!(1103+26390n)}{(n!)^4 396^{4n}} \quad (40)$$

(Beeler *et al.* 1972, Item 139; Borwein *et al.* 1989). Equation (40) is derived from a modular identity of order 58, although a first derivation was not presented prior to Borwein and Borwein (1987). The above series both give

$$\pi \approx \frac{2206\sqrt{2}}{9801} = 3.14159273001\ldots \quad (41)$$

as the first approximation and provide, respectively, about 6 and 8 decimal places per term. Such series exist because of the rationality of various modular invariants. The general form of the series is

$$\sum_{n=0}^{\infty}[a(t)+nb(t)]\frac{(6n)!}{(3n)!(n!)^3}\frac{1}{[j(t)]^n} = \frac{\sqrt{-j(t)}}{\pi}, \quad (42)$$

where t is a QUADRATIC FORM DISCRIMINANT, $j(t)$ is the j-FUNCTION,

$$b(t) = \sqrt{t[1728 - j(t)]} \quad (43)$$
$$a(t) = \frac{b(t)}{6}\left\{1 - \frac{E_4(t)}{E_6(t)}\left[E_2(t) - \frac{6}{\pi\sqrt{t}}\right]\right\}, \quad (44)$$

and the E_i are RAMANUJAN-EISENSTEIN SERIES. A CLASS NUMBER p field involves pth degree ALGEBRAIC INTEGERS of the constants $A = a(t)$, $B = b(t)$, and $C = c(t)$. The fastest converging series that uses only INTEGER terms corresponds to the largest CLASS NUMBER 1 discriminant of $d = -163$ and was formulated by the Chudnovsky brothers (1987). The 163 appearing here is the same one appearing in the fact that $e^{\pi\sqrt{163}}$ (the RAMANUJAN CONSTANT) is very nearly an INTEGER. The series is given by

$$\frac{1}{\pi} = 12\sum_{n=0}^{\infty}\frac{(-1)^n(6n)!(13591409+545140134n)}{(n!)^3(3n)!(640320^3)^{n+1/2}}$$
$$= \frac{163\cdot 8\cdot 27\cdot 7\cdot 11\cdot 19\cdot 127}{640320^{3/2}}$$
$$\times\sum_{n=0}^{\infty}\left(\frac{13591409}{163\cdot 2\cdot 9\cdot 7\cdot 11\cdot 19\cdot 127}+n\right)$$
$$\times\frac{(6n)!}{(3n)!(n!)^3}\frac{(-1)^n}{640320^{3n}} \quad (45)$$

(Borwein and Borwein 1993). This series gives 14 digits accurately per term. The same equation in another form was given by the Chudnovsky brothers (1987) and is used by *Mathematica*® (Wolfram Research, Champaign, IL) to calculate π (Vardi 1991),

$$\pi = \frac{426880\sqrt{10005}}{A[{}_3F_2(\frac{1}{6},\frac{1}{2},\frac{5}{6};1,1;B) - C\,{}_3F_2(\frac{7}{6},\frac{3}{2},\frac{11}{6};2,2;B)]}, \quad (46)$$

where

$$A \equiv 13591409 \quad (47)$$
$$B \equiv -\frac{1}{151931373056000} \quad (48)$$
$$C \equiv \frac{30285563}{1651969144908540723200}. \quad (49)$$

The best formula for CLASS NUMBER 2 (largest discriminant -427) is

$$\frac{1}{\pi} = 12\sum_{n=0}^{\infty}\frac{(-1)^n(6n)!(A+Bn)}{(n!)^3(3n)!C^{n+1/2}}, \quad (50)$$

where

$$A \equiv 212175710912\sqrt{61} + 1657145277365 \quad (51)$$
$$B \equiv 13773980892672\sqrt{61} + 107578229802750 \quad (52)$$
$$C \equiv [5280(236674 + 30303\sqrt{61}]^3 \quad (53)$$

(Borwein and Borwein 1993). This series adds about 25 digits for each additional term. The fastest converging series for CLASS NUMBER 3 corresponds to $d = -907$ and gives 37–38 digits per term. The fastest converging CLASS NUMBER 4 series corresponds to $d = -1555$ and is

$$\frac{\sqrt{-C^3}}{\pi} = \sum_{n=0}^{\infty}\frac{(6n)!}{(3n)!(n!)^3}\frac{A+nB}{C^{3n}}, \quad (54)$$

where

$$\begin{aligned}A &= 63365028312971999585426220\\ &\quad + 28337702140800842046825600\sqrt{5}\\ &\quad + 384\sqrt{5}(10891728551171178200467 4\cdots\\ &\quad \cdots 36212395209160385656017 + 487902908657881022\cdots\\ &\quad \cdots 5077338534541688721351255040\sqrt{5}\,)^{1/2} \qquad (55)\\ B &= 7849910453496627210289749000\\ &\quad + 3510586678260932028965606400\sqrt{5}\\ &\quad + 2515968\sqrt{3110}(6260208323789001 6\cdots\\ &\quad \cdots 3699332265444402088216 1 + 2799650273060444296\cdots\\ &\quad \cdots 5772068907188251902 35\sqrt{5}\,)^{1/2} \qquad (56)\\ C &= -214772995063512240 - 96049403338648032\sqrt{5}\\ &\quad - 1296\sqrt{5}(10985234579463550323713318473\\ &\quad + 4912746253692362754607395912\sqrt{5}\,)^{1/2}. \qquad (57)\end{aligned}$$

This gives 50 digits per term. Borwein and Borwein (1993) have developed a general ALGORITHM for generating such series for arbitrary CLASS NUMBER. Bellard gives the exotic formula

$$\pi = \frac{1}{740025}\left[\sum_{n=1}^{\infty}\frac{3P(n)}{\binom{7n}{2n}2^{n-1}} - 20379280\right], \qquad (58)$$

where

$$\begin{aligned}P(n) \equiv{}& -885673181n^5 + 3125347237n^4 - 2942969225n^3\\ &+1031962795n^2 - 196882274n + 10996648. \qquad (59)\end{aligned}$$

A complete listing of Ramanujan's series for $1/\pi$ found in his second and third notebooks is given by Berndt (1994, pp. 352–354),

$$\frac{4}{\pi} = \sum_{n=0}^{\infty}\frac{(6n+1){(\frac{1}{2})_n}^3}{4^n(n!)^3} \qquad (60)$$

$$\frac{16}{\pi} = \sum_{n=0}^{\infty}\frac{(42n+5){(\frac{1}{2})_n}^3}{(64)^n(n!)^3} \qquad (61)$$

$$\frac{32}{\pi} = \sum_{n=0}^{\infty}\frac{(42\sqrt{5}\,n+5\sqrt{5}+30n-1){(\frac{1}{2})_n}^3}{(64)^n(n!)^3} \times\left(\frac{\sqrt{5}-1}{2}\right)^{8n} \qquad (62)$$

$$\frac{27}{4\pi} = \sum_{n=0}^{\infty}\frac{(15n+2)(\frac{1}{2})_n(\frac{1}{3})_n(\frac{2}{3})_n}{(n!)^3}\left(\frac{2}{27}\right)^n \qquad (63)$$

$$\frac{15\sqrt{3}}{2\pi} = \sum_{n=0}^{\infty}\frac{(33n+4)(\frac{1}{2})_n(\frac{1}{3})_n(\frac{2}{3})_n}{(n!)^3}\left(\frac{4}{125}\right)^n \qquad (64)$$

$$\frac{5\sqrt{5}}{2\pi\sqrt{3}} = \sum_{n=0}^{\infty}\frac{(11n+1)(\frac{1}{2})_n(\frac{1}{6})_n(\frac{5}{6})_n}{(n!)^3}\left(\frac{4}{125}\right)^n \qquad (65)$$

$$\frac{85\sqrt{85}}{18\pi\sqrt{3}} = \sum_{n=0}^{\infty}\frac{(133n+8)(\frac{1}{2})_n(\frac{1}{6})_n(\frac{5}{6})_n}{(n!)^3}\left(\frac{4}{85}\right)^n \qquad (66)$$

$$\frac{4}{\pi} = \sum_{n=0}^{\infty}\frac{(-1)^n(20n+3)(\frac{1}{2})_n(\frac{1}{4})_n(\frac{3}{4})_n}{(n!)^3 2^{2n+1}} \qquad (67)$$

$$\frac{4}{\pi\sqrt{3}} = \sum_{n=0}^{\infty}\frac{(-1)^n(28n+3)(\frac{1}{2})_n(\frac{1}{4})_n(\frac{3}{4})_n}{(n!)^3 3^n 4^{n+1}} \qquad (68)$$

$$\frac{4}{\pi} = \sum_{n=0}^{\infty}\frac{(-1)^n(260n+23)(\frac{1}{2})_n(\frac{1}{4})_n(\frac{3}{4})_n}{(n!)^3(18)^{2n+1}} \qquad (69)$$

$$\frac{4}{\pi\sqrt{5}} = \sum_{n=0}^{\infty}\frac{(-1)^n(644n+41)(\frac{1}{2})_n(\frac{1}{4})_n(\frac{3}{4})_n}{(n!)^3 5^n(72)^{2n+1}} \qquad (70)$$

$$\frac{4}{\pi} = \sum_{n=0}^{\infty}\frac{(-1)^n(21460n+1123)(\frac{1}{2})_n(\frac{1}{4})_n(\frac{3}{4})_n}{(n!)^3(882)^{2n+1}} \qquad (71)$$

$$\frac{2\sqrt{3}}{\pi} = \sum_{n=0}^{\infty}\frac{(8n+1)(\frac{1}{2})_n(\frac{1}{4})_n(\frac{3}{4})_n}{(n!)^3 9^n} \qquad (72)$$

$$\frac{1}{2\pi\sqrt{2}} = \sum_{n=0}^{\infty}\frac{(10n+1)(\frac{1}{2})_n(\frac{1}{4})_n(\frac{3}{4})_n}{(n!)^3 9^{2n+1}} \qquad (73)$$

$$\frac{1}{3\pi\sqrt{3}} = \sum_{n=0}^{\infty}\frac{(40n+3)(\frac{1}{2})_n(\frac{1}{4})_n(\frac{3}{4})_n}{(n!)^3(49)^{2n+1}} \qquad (74)$$

$$\frac{2}{\pi\sqrt{11}} = \sum_{n=0}^{\infty}\frac{(280n+19)(\frac{1}{2})_n(\frac{1}{4})_n(\frac{3}{4})_n}{(n!)^3(99)^{2n+1}} \qquad (75)$$

$$\frac{1}{2\pi\sqrt{2}} = \sum_{n=0}^{\infty}\frac{(26390n+1103)(\frac{1}{2})_n(\frac{1}{4})_n(\frac{3}{4})_n}{(n!)^3(99)^{4n+2}}. \qquad (76)$$

These equations were first proved by Borwein and Borwein (1987, pp. 177–187). Borwein and Borwein (1987b, 1988, 1993) proved other equations of this type, and Chudnovsky and Chudnovsky (1987) found similar equations for other transcendental constants.

A SPIGOT ALGORITHM for π is given by Rabinowitz and Wagon (1995). Amazingly, a closed form expression giving a *digit extraction* algorithm which produces digits of π (or π^2) in base-16 was recently discovered by Bailey *et al.* (Bailey *et al.* 1995, Adamchik and Wagon 1997),

$$\pi = \sum_{n=0}^{\infty}\left(\frac{4}{8n+1}-\frac{2}{8n+4}-\frac{1}{8n+5}-\frac{1}{8n+6}\right)\left(\frac{1}{16}\right)^n, \qquad (77)$$

which can also be written using the shorthand notation

$$\pi = \sum_{i=1}^{\infty}\frac{p_i}{16^{\lfloor i/8\rfloor}i} \qquad \{p_i\} = \{\,\overline{4,0,0,-2,-1,-1,0,0}\,\}, \qquad (78)$$

where $\{p_i\}$ is given by the periodic sequence obtained by appending copies of $\{4,0,0,-2,-1,-1,0,0\}$ (in other words, $p_i \equiv p_{[(i-1)\pmod 8]+1}$ for $i > 8$) and $\lfloor x\rfloor$ is the FLOOR FUNCTION. This expression was discovered using the PSLQ ALGORITHM and is equivalent to

$$\pi = \int_0^1 \frac{16y-16}{y^4-2y^3+4y-4}\,dy. \qquad (79)$$

A similar formula was subsequently discovered by Ferguson, leading to a 2-D lattice of such formulas which can be generated by these two formulas. A related integral is

$$\pi = \tfrac{22}{7} - \int_0^1 \frac{x^4(1-x)^4}{1+x^2}\,dx \tag{80}$$

(Le Lionnais 1983, p. 22). F. Bellard found the more rapidly converging digit-extraction algorithm (in Hexadecimal)

$$\begin{aligned}\pi = \frac{1}{2^6}\sum_{n=0}^{\infty}\frac{(-1)^n}{2^{10n}}\biggl(&-\frac{2^5}{4n+1}-\frac{1}{4n+3}+\frac{2^8}{10n+1}\\ &-\frac{2^6}{10n+3}-\frac{2^2}{10n+5}-\frac{2^2}{10n+7}+\frac{1}{10n+9}\biggr).\end{aligned} \tag{81}$$

More amazingly still, S. Plouffe has devised an algorithm to compute the nth Digit of π in any base in $\mathcal{O}(n^3(\log n)^3)$ steps.

Another identity is

$$\pi^2 = 36\,\mathrm{Li}_2(\tfrac{1}{2}) - 36\,\mathrm{Li}_2(\tfrac{1}{4}) - 12\,\mathrm{Li}_2(\tfrac{1}{8}) + 6\,\mathrm{Li}_2(\tfrac{1}{64}), \tag{82}$$

where L_n is the Polylogarithm. (82) is equivalent to

$$\frac{\pi^2}{36} = \sum_{i=1}^{\infty}\frac{a_i}{2^i i^2} \qquad \{a_i\} = [\overline{1,-3,-2,-3,1,0}] \tag{83}$$

and

$$\pi^2 = 12L_2(\tfrac{1}{2}) + 6(\ln 2)^2 \tag{84}$$

(Bailey *et al.* 1995). Furthermore

$$\begin{aligned}\pi^2 = \frac{1}{8}\sum_{k=0}^{\infty}\frac{1}{64^k}\biggl[&\frac{144}{(6k+1)^2}-\frac{216}{(6k+2)^2}-\frac{72}{(6k+3)^2}\\ &-\frac{54}{(6k+4)^2}+\frac{9}{(6k+5)^2}\biggr]\end{aligned} \tag{85}$$

and

$$\begin{aligned}\pi^2 = \sum_{k=0}^{\infty}\frac{1}{16^k}\biggl[&\frac{16}{(8k+1)^2}-\frac{16}{(8k+2)^2}-\frac{8}{(8k+3)^2}\\ &-\frac{16}{(8k+4)^2}-\frac{4}{(8k+5)^2}-\frac{4}{(8k+6)^2}+\frac{2}{(8k+7)^2}\biggr]\end{aligned} \tag{86}$$

(Bailey *et al.* 1995, Bailey and Plouffe).

A slew of additional identities due to Ramanujan, Catalan, and Newton are given by Castellanos (1988, pp. 86–88), including several involving sums of Fibonacci Numbers.

Gasper quotes the result

$$\pi = \frac{16}{3}\left[\lim_{x\to\infty} x\,{}_1F_2(\tfrac{1}{2};2,3;-x^2)\right]^{-1}, \tag{87}$$

where ${}_1F_2$ is a Generalized Hypergeometric Function, and transforms it to

$$\pi = \lim_{x\to\infty} 4x\,{}_1F_2(\tfrac{1}{2};\tfrac{3}{2},\tfrac{3}{2};-x^2). \tag{88}$$

Fascinating results due to Gosper include

$$\lim_{n\to\infty}\prod_{i=n}^{2n}\frac{\pi}{2\tan^{-1} i} = 4^{1/\pi} = 1.554682275\ldots \tag{89}$$

and

$$\sum_{n=1}^{\infty}\frac{1}{n^2}\cos\left(\frac{9}{n\pi+\sqrt{n^2\pi^2-9}}\right) = -\frac{\pi^2}{12e^3} = -0.040948222\ldots. \tag{90}$$

Gosper also gives the curious identity

$$\begin{aligned}&\frac{1}{e}\prod_{n=1}^{\infty}\left(\frac{1}{3n}+1\right)^{3n+1/2}\\ &= \frac{3\cdot 3^{1/24}\sqrt{(\frac{1}{3})!}}{2^{5/6}\exp\left[\frac{\gamma}{3}-\frac{\pi\sqrt{3}}{18}+\frac{\sqrt{3}\psi_1(\frac{1}{3})}{12\pi}-\frac{2\zeta'(2)}{\pi^2}-1\right]\pi^{5/6}}\\ &= 1.012378552722912\ldots.\end{aligned} \tag{91}$$

Another curious fact is the Almost Integer

$$e^\pi - \pi = 19.999099979\ldots, \tag{92}$$

which can also be written as

$$(\pi+20)^i = -0.9999999992 - 0.0000388927i \approx -1 \tag{93}$$

$$\cos(\ln(\pi+20)) \approx -0.9999999992. \tag{94}$$

Applying Cosine a few more times gives

$$\cos(\pi\cos(\pi\cos(\ln(\pi+20)))) \approx -1 + 3.9321609261\times 10^{-35}. \tag{95}$$

π may also be computed using iterative Algorithms. A quadratically converging Algorithm due to Borwein is

$$x_0 = \sqrt{2} \tag{96}$$

$$\pi_0 = 2+\sqrt{2} \tag{97}$$

$$y_1 = 2^{1/4} \tag{98}$$

and

$$x_{n+1} = \frac{1}{2}\left(\sqrt{x_n}+\frac{1}{\sqrt{x_n}}\right) \tag{99}$$

$$y_{n+1} = \frac{y_n\sqrt{x_n}+\frac{1}{\sqrt{x_n}}}{y_n+1} \tag{100}$$

$$\pi_n = \pi_{n-1}\frac{x_n+1}{y_n+1}. \tag{101}$$

π_n decreases monotonically to π with

$$\pi_n - \pi < 10^{-2^{n+1}} \tag{102}$$

for $n \geq 2$. The BRENT-SALAMIN FORMULA is another quadratically converging algorithm which can be used to calculate π. A quadratically convergent algorithm for $\pi/\ln 2$ based on an observation by Salamin is given by defining

$$f(k) = k2^{-k/4}\left[\sum_{n=1}^{\infty} 2^{-k\binom{n}{2}}\right]^2, \tag{103}$$

then writing

$$g_0 \equiv \frac{f(n)}{f(2n)}. \tag{104}$$

Now iterate

$$g_k = \sqrt{\frac{1}{2}\left(g_{k-1} + \frac{1}{g_{k+1}}\right)} \tag{105}$$

to obtain

$$\pi = 2(\ln 2)f(n)\prod_{k=1}^{\infty} g_k. \tag{106}$$

A cubically converging ALGORITHM which converges to the nearest multiple of π to f_0 is the simple iteration

$$f_n = f_{n-1} + \sin(f_{n-1}) \tag{107}$$

(Beeler *et al.* 1972). For example, applying to 23 gives the sequence

$$\{23, 22.1537796, 21.99186453, 21.99114858, \ldots\}, \tag{108}$$

which converges to $7\pi \approx 21.99114858$.

A quartically converging ALGORITHM is obtained by letting

$$y_0 = \sqrt{2} - 1 \tag{109}$$

$$\alpha = 6 - 4\sqrt{2}, \tag{110}$$

then defining

$$y_{n+1} = \frac{1 - (1 - {y_n}^4)^{1/4}}{1 + (1 - {y_n}^4)^{1/4}} \tag{111}$$

$$\alpha_{n+1} = (1 + y_{n+1})^4 \alpha_n - 2^{2n+3} y_{n+1}(1 + y_{n+1} + {y_{n+1}}^2). \tag{112}$$

Then

$$\pi = \lim_{n\to\infty} \frac{1}{\alpha_n} \tag{113}$$

and α_n converges to $1/\pi$ quartically with

$$\alpha_n - \frac{1}{\pi} < 16 \cdot 4^n e^{-2\pi \cdot 4^n} \tag{114}$$

(Borwein and Borwein 1987, Bailey 1988, Borwein *et al.* 1989). This ALGORITHM rests on a MODULAR EQUATION identity of order 4.

A quintically converging ALGORITHM is obtained by letting

$$s_0 = 5(\sqrt{5} - 2) \tag{115}$$

$$\alpha_0 = \tfrac{1}{2}. \tag{116}$$

Then let

$$s_{n+1} = \frac{25}{(z + \frac{x}{z} + 1)^2 s_n}, \tag{117}$$

where

$$x = \frac{5}{s_n} - 1 \tag{118}$$

$$y = (x-1)^2 + 7 \tag{119}$$

$$z = [\tfrac{1}{2}x(y + \sqrt{y^2 - 4x^3})]^{1/5}. \tag{120}$$

Finally, let

$$\alpha_{n+1} = {s_n}^2 \alpha_n - 5^n[\tfrac{1}{2}({s_n}^2 - 5) + \sqrt{s_n({s_n}^2 - 2s_n + 5)}], \tag{121}$$

then

$$0 < \alpha_n - \frac{1}{\pi} < 16 \cdot 5^n e^{-\pi 5^n} \tag{122}$$

(Borwein *et al.* 1989). This ALGORITHM rests on a MODULAR EQUATION identity of order 5.

Another ALGORITHM is due to Woon (1995). Define $a(0) \equiv 1$ and

$$a(n) = \sqrt{1 + \left[\sum_{k=0}^{n-1} a(k)\right]^2}. \tag{123}$$

It can be proved by induction that

$$a(n) = \csc\left(\frac{\pi}{2^{n+1}}\right). \tag{124}$$

For $n = 0$, the identity holds. If it holds for $n \leq t$, then

$$a(t+1) = \sqrt{1 + \left[\sum_{k=0}^{t} \csc\left(\frac{\pi}{2^{k+1}}\right)\right]^2}, \tag{125}$$

but

$$\csc\left(\frac{\pi}{2^{k+1}}\right) = \cot\left(\frac{\pi}{2^{k+2}}\right) - \cot\left(\frac{\pi}{2^{k+1}}\right), \tag{126}$$

so

$$\sum_{k=0}^{t} \csc\left(\frac{\pi}{2^{k+1}}\right) = \cot\left(\frac{\pi}{2^{t+2}}\right). \tag{127}$$

Therefore,

$$a(t+1) = \csc\left(\frac{\pi}{2^{t+2}}\right), \tag{128}$$

so the identity holds for $n = t+1$ and, by induction, for all Nonnegative n, and

$$\begin{aligned}\lim_{n\to\infty}\frac{2^{n+1}}{a(n)} &= \lim_{n\to\infty} 2^{n+1}\sin\left(\frac{\pi}{2^{n+1}}\right)\\ &= \lim_{n\to\infty} 2^{n+1}\frac{\pi}{2^{n+1}}\frac{\sin\left(\frac{\pi}{2^{n+1}}\right)}{\frac{\pi}{2^{n+1}}}\\ &= \pi\lim_{\theta\to 0}\frac{\sin\theta}{\theta} = \pi.\end{aligned} \tag{129}$$

Other iterative Algorithms are the Archimedes Algorithm, which was derived by Pfaff in 1800, and the Brent-Salamin Formula. Borwein *et al.* (1989) discuss pth order iterative algorithms.

Kochansky's Approximation is the Root of

$$9x^4 - 240x^2 + 1492. \tag{130}$$

given by

$$\pi \approx \sqrt{\tfrac{40}{3} - \sqrt{12}} \approx 3.141533. \tag{131}$$

An approximation involving the Golden Mean is

$$\pi \approx \tfrac{6}{5}\phi^2 = \frac{6}{5}\left(\frac{\sqrt{5}+1}{2}\right)^2 = \tfrac{3}{5}(3+\sqrt{5}) = 3.14164\ldots. \tag{132}$$

Some approximations due to Ramanujan

$$\pi \approx \frac{19\sqrt{7}}{16} \tag{133}$$

$$\approx \tfrac{7}{3}(1 + \tfrac{1}{5}\sqrt{3}) \tag{134}$$

$$\approx \tfrac{9}{5} + \sqrt{\tfrac{9}{5}} \tag{135}$$

$$\approx \left(9^2 + \frac{19^2}{22}\right)^{1/4} = \left(102 - \frac{2222}{22^2}\right)^{1/4} \tag{136}$$

$$\approx (97 + \tfrac{1}{2} - \tfrac{1}{11})^{1/4} = (97 + \tfrac{9}{22})^{1/4} \tag{137}$$

$$\approx \frac{63}{25}\left(\frac{17+15\sqrt{5}}{7+15\sqrt{5}}\right) \tag{138}$$

$$\approx \frac{355}{113}\left(1 - \frac{0.0003}{3533}\right) \tag{139}$$

$$\approx \frac{12}{\sqrt{130}}\ln\left[\frac{(3+\sqrt{13})(\sqrt{8}+\sqrt{10})}{2}\right] \tag{140}$$

$$\approx \frac{24}{\sqrt{142}}\ln\left[\frac{\sqrt{10+11\sqrt{2}}+\sqrt{10+7\sqrt{2}}}{2}\right] \tag{141}$$

$$\approx \frac{12}{\sqrt{190}}\ln[(3+\sqrt{10})(\sqrt{8}+\sqrt{10})] \tag{142}$$

$$\begin{aligned}&\approx \frac{12}{\sqrt{310}}\ln\left[\tfrac{1}{4}(3+\sqrt{5})(2+\sqrt{2})\left(5+2\sqrt{10}\right.\right.\\ &\qquad \left.\left.+\sqrt{61+20\sqrt{10}}\right)\right]\end{aligned} \tag{143}$$

$$\begin{aligned}&\approx \frac{4}{\sqrt{522}}\ln\left[\left(\frac{5+\sqrt{29}}{\sqrt{2}}\right)^3(5\sqrt{29}+11\sqrt{6})\right.\\ &\qquad \left.\times\left(\sqrt{\frac{9+3\sqrt{6}}{4}}+\sqrt{\frac{5+3\sqrt{6}}{4}}\right)^6\right],\end{aligned} \tag{144}$$

which are accurate to 3, 4, 4, 8, 8, 9, 14, 15, 15, 18, 23, 31 digits, respectively (Ramanujan 1913–1914; Hardy 1952, p. 70; Berndt 1994, pp. 48–49 and 88–89).

Castellanos (1988) gives a slew of curious formulas:

$$\pi \approx (2e^3 + e^8)^{1/7} \tag{145}$$

$$\approx \left(\frac{553}{311+1}\right)^2 \tag{146}$$

$$\approx (\tfrac{3}{14})^4(\tfrac{193}{5})^2 \tag{147}$$

$$\approx (\tfrac{296}{167})^2 \tag{148}$$

$$\approx \left(\frac{66^3+86^2}{55^3}\right)^2 \tag{149}$$

$$\approx 1.09999901 \cdot 1.19999911 \cdot 1.39999931 \cdot 1.69999961 \tag{150}$$

$$\approx \frac{47^3+20^3}{30^3} - 1 \tag{151}$$

$$\approx 2 + \sqrt{1 + (\tfrac{413}{750})^2} \tag{152}$$

$$\approx (\tfrac{77729}{254})^{1/5} \tag{153}$$

$$\approx \left(31 + \frac{62^2+14}{28^4}\right)^{1/3} \tag{154}$$

$$\approx \frac{1700^3 + 82^3 - 10^3 - 9^3 - 6^3 - 3^3}{69^5} \tag{155}$$

$$\approx \left(95 + \frac{93^4+34^4+17^4+88}{75^4}\right)^{1/4} \tag{156}$$

$$\approx \left(100 - \frac{2125^3+214^3+30^3+37^2}{82^5}\right)^{1/4}, \tag{157}$$

which are accurate to 3, 4, 4, 5, 6, 7, 7, 8, 9, 10, 11, 12, and 13 digits, respectively. An extremely accurate approximation due to Shanks (1982) is

$$\pi \approx \frac{6}{\sqrt{3502}}\ln(2u) + 7.37\times 10^{-82}, \tag{158}$$

where u is the product of four simple quartic units. A sequence of approximations due to Plouffe includes

$$\pi \approx 43^{7/23} \tag{159}$$

$$\approx \frac{\ln 2198}{\sqrt{6}} \tag{160}$$

$$\approx (\tfrac{13}{4})^{1181/1216} \tag{161}$$

$$\approx \frac{689}{396\ln\left(\frac{689}{396}\right)} \tag{162}$$

$$\approx (\tfrac{2143}{22})^{1/4} \tag{163}$$

$$\approx \sqrt{\frac{9}{67}\ln 5280} \tag{164}$$

$$\approx (\tfrac{63023}{30510})^{1/3} + \tfrac{1}{4} + \tfrac{1}{2}(\sqrt{5}+1) \tag{165}$$

$$\approx \tfrac{48}{23}\ln\left(\frac{60318}{13387}\right) \tag{166}$$

$$\approx (228 + \tfrac{16}{1329})^{1/41} + 2 \tag{167}$$

$$\approx \tfrac{125}{124} \ln\left(\frac{28102}{1277}\right) \tag{168}$$

$$\approx \tfrac{276694819753963}{226588}^{1/158} + 2 \tag{169}$$

$$\approx \frac{\ln 262537412640768744}{\sqrt{163}}, \tag{170}$$

which are accurate to 4, 5, 7, 7, 8, 9, 10, 11, 11, 11, 23, and 30 digits, respectively.

Ramanujan (1913–14) and Olds (1963) give geometric constructions for 355/113. Gardner (1966, pp. 92–93) gives a geometric construction for $3 + 16/113 = 3.1415929\ldots$. Dixon (1991) gives constructions for $6/5(1+\phi) = 3.141640\ldots$ and $\sqrt{4 + (3 - \tan(30°))} = 3.141533\ldots$. Constructions for approximations of π are approximations to CIRCLE SQUARING (which is itself impossible).

A short mnemonic for remembering the first eight DECIMAL DIGITS of π is "May I have a large container of coffee?" giving 3.1415926 (Gardner 1959; Gardner 1966, p. 92; Eves 1990, p. 122, Davis 1993, p. 9). A more substantial mnemonic giving 15 digits (3.14159265358979) is "How I want a drink, alcoholic of course, after the heavy lectures involving quantum mechanics," originally due to Sir James Jeans (Gardner 1966, p. 92; Castellanos 1988, p. 152; Eves 1990, p. 122; Davis 1993, p. 9; Blatner 1997, p. 112). A slight extension of this adds the phrase "All of thy geometry, Herr Planck, is fairly hard," giving 24 digits in all (3.141592653589793238462643).

An even more extensive rhyming mnemonic giving 31 digits is "Now I will a rhyme construct, By chosen words the young instruct. Cunningly devised endeavour, Con it and remember ever. Widths in circle here you see, Sketched out in strange obscurity." (Note that the British spelling of "endeavour" is required here.)

The following stanzas are the first part of a poem written by M. Keith based on Edgar Allen Poe's "The Raven." The entire poem gives 740 digits; the fragment below gives only the first 80 (Blatner 1997, p. 113). Words with ten letters represent the digit 0, and those with 11 or more digits are taken to represent two digits.

Poe, E.: Near a Raven.

Midnights so dreary, tired and weary.
Silently pondering volumes extolling all by-now obsolete lore.
During my rather long nap-the weirdest tap!
An ominous vibrating sound disturbing my chamber's antedoor.
'This,' I whispered quietly, 'I ignore.'
Perfectly, the intellect remembers: the ghostly fires, a glittering ember.
Inflamed by lightning's outbursts, windows cast penumbras upon this floor.
Sorrowful, as one mistreated, unhappy thoughts I heeded:
That inimitable lesson in elegance—Lenore—
is delighting, exciting... nevermore.

An extensive collection of π mnemonics in many languages is maintained by A. P. Hatzipolakis. Other mnemonics in various languages are given by Castellanos (1988) and Blatner (1997, pp. 112–118).

In the following, the word "digit" refers to decimal digit after the decimal point. J. H. Conway has shown that there is a sequence of fewer than 40 FRACTIONS F_1, F_2, ... with the property that if you start with 2^n and repeatedly multiply by the first of the F_i that gives an integral answer, then the next POWER of 2 to occur will be the 2^nth decimal digit of π.

The first occurrence of n 0s appear at digits 32, 307, 601, 13390, 17534, ... The sequence 9999998 occurs at decimal 762 (which is sometimes called the FEYNMAN POINT). This is the largest value of any seven digits in the first million decimals. The first time the BEAST NUMBER 666 appears is decimal 2440. The digits 314159 appear at least six times in the first 10 million decimal places of π (Pickover 1995). In the following, "digit" means digit of $\pi - 3$. The sequence 0123456789 occurs beginning at digits 17,387,594,880, 26,852,899,245, 30,243,957,439, 34,549,153,953, 41,952,536,161, and 43,289,964,000. The sequence 9876543210 occurs beginning at digits 21,981,157,633, 29,832,636,867, 39,232,573,648, 42,140,457,481, and 43,065,796,214. The sequence 27182818284 (the digits of e) occur beginning at digit 45,111,908,393. There are also interesting patterns for $1/\pi$. 0123456789 occurs at 6,214,876,462, 9876543210 occurs at 15,603,388,145 and 51,507,034,812, and 999999999999 occurs at 12,479,021,132 of $1/\pi$.

Scanning the decimal expansion of π until all n-digit numbers have occurred, the last 1-, 2-, ... digit numbers appearing are 0, 68, 483, 6716, 33394, 569540, ... (Sloane's A032510). These end at digits 32, 606, 8555, 99849, 1369564, 14118312, ...

see also ALMOST INTEGER, ARCHIMEDES ALGORITHM, BRENT-SALAMIN FORMULA, BUFFON-LAPLACE NEEDLE PROBLEM, BUFFON'S NEEDLE PROBLEM, CIRCLE, DIRICHLET BETA FUNCTION, DIRICHLET ETA FUNCTION, DIRICHLET LAMBDA FUNCTION, e, EULER-MASCHERONI CONSTANT, GAUSSIAN DISTRIBUTION, MACLAURIN SERIES, MACHIN'S FORMULA, MACHIN-LIKE FORMULAS, RELATIVELY PRIME, RIEMANN ZETA FUNCTION, SPHERE, TRIGONOMETRY

References

Adamchik, V. and Wagon, S. "A Simple Formula for π." *Amer. Math. Monthly* **104**, 852–855, 1997.

Adamchik, V. and Wagon, S. "Pi: A 2000-Year Search Changes Direction." `http://www.wolfram.com/~victor/articles/pi/pi.html`.

Almkvist, G. "Many Correct Digits of π, Revisited." *Amer. Math. Monthly* **104**, 351–353, 1997.

Arndt, J. "Cryptic Pi Related Formulas." http://www.jjj.de/hfloat/pise.dvi.

Arndt, J. and Haenel, C. *Pi: Algorithmen, Computer, Arithmetik.* Berlin: Springer-Verlag, 1998.

Assmus, E. F. "Pi." *Amer. Math. Monthly* **92**, 213–214, 1985.

Bailey, D. H. "Numerical Results on the Transcendence of Constants Involving π, e, and Euler's Constant." *Math. Comput.* **50**, 275–281, 1988a.

Bailey, D. H. "The Computation of π to 29,360,00 Decimal Digit using Borwein's' Quartically Convergent Algorithm." *Math. Comput.* **50**, 283–296, 1988b.

Bailey, D.; Borwein, P.; and Plouffe, S. "On the Rapid Computation of Various Polylogarithmic Constants." http://www.cecm.sfu.ca/~pborwein/PAPERS/P123.ps.

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 55 and 274, 1987.

Beckmann, P. *A History of Pi, 3rd ed.* New York: Dorset Press, 1989.

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Berggren, L.; Borwein, J.; and Borwein, P. *Pi: A Source Book.* New York: Springer-Verlag, 1997.

Bellard, F. "Fabrice Bellard's Pi Page." http://www-stud.enst.fr/~bellard/pi/.

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, 1994.

Blatner, D. *The Joy of Pi.* New York: Walker, 1997.

Blatner, D. "The Joy of Pi." http://www.joyofpi.com.

Borwein, P. B. "Pi and Other Constants." http://www.cecm.sfu.ca/~pborwein/PISTUFF/Apistuff.html.

Borwein, J. M. "Ramanujan Type Series." http://www.cecm.sfu.ca/organics/papers/borwein/paper/html/local/omlink9/html/node1.html.

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, 1987a.

Borwein, J. M. and Borwein, P. B. "Ramanujan's Rational and Algebraic Series for $1/\pi$." *Indian J. Math.* **51**, 147–160, 1987b.

Borwein, J. M. and Borwein, P. B. "More Ramanujan-Type Series for $1/\pi$." In *Ramanujan Revisited.* Boston, MA: Academic Press, pp. 359–374, 1988.

Borwein, J. M. and Borwein, P. B. "Class Number Three Ramanujan Type Series for $1/\pi$." *J. Comput. Appl. Math.* **46**, 281–290, 1993.

Borwein, J. M.; Borwein, P. B.; and Bailey, D. H. "Ramanujan, Modular Equations, and Approximations to Pi, or How to Compute One Billion Digits of Pi." *Amer. Math. Monthly* **96**, 201–219, 1989.

Brown, K. S. "Rounding Up to Pi." http://www.seanet.com/~ksbrown/kmath001.htm.

Castellanos, D. "The Ubiquitous Pi. Part I." *Math. Mag.* **61**, 67–98, 1988.

Castellanos, D. "The Ubiquitous Pi. Part II." *Math. Mag.* **61**, 148–163, 1988.

Chan, J. "As Easy as Pi." *Math Horizons,* Winter 1993, pp. 18–19, 1993.

Chudnovsky, D. V. and Chudnovsky, G. V. *Padé and Rational Approximations to Systems of Functions and Their Arithmetic Applications.* Berlin: Springer-Verlag, 1984.

Chudnovsky, D. V. and Chudnovsky, G. V. "Approximations and Complex Multiplication According to Ramanujan." In *Ramanujan Revisited: Proceedings of the Centenary Conference* (Ed. G. E. Andrews, B. C. Berndt, and R. A. Rankin). Boston, MA: Academic Press, pp. 375–472, 1987.

Conway, J. H. and Guy, R. K. "The Number π." In *The Book of Numbers.* New York: Springer-Verlag, pp. 237–239, 1996.

David, Y. "On a Sequence Generated by a Sieving Process." *Riveon Lematematika* **11**, 26–31, 1957.

Davis, D. M. *The Nature and Power of Mathematics.* Princeton, NJ: Princeton University Press, 1993.

Dixon, R. "The Story of Pi (π)." §4.3 in *Mathographics.* New York: Dover, pp. 44–49 and 98–101, 1991.

Dunham, W. "A Gem from Isaac Newton." Ch. 7 in *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 106–112 and 155–183, 1990.

Eves, H. *An Introduction to the History of Mathematics, 6th ed.* Philadelphia, PA: Saunders, 1990.

Exploratorium. "π Page." http://www.exploratorium.edu/learning_studio/pi.

Finch, S. "Favorite Mathematical Constants." http://www.mathsoft.com/asolve/constant/pi/pi.html.

Flajolet, P. and Vardi, I. "Zeta Function Expansions of Classical Constants." Unpublished manuscript. 1996. http://pauillac.inria.fr/algo/flajolet/Publications/landau.ps.

Gardner, M. "Memorizing Numbers." Ch. 11 in *The Scientific American Book of Mathematical Puzzles and Diversions.* New York: Simon and Schuster, p. 103, 1959.

Gardner, M. "The Transcendental Number Pi." Ch. 8 in *Martin Gardner's New Mathematical Diversions from Scientific American.* New York: Simon and Schuster, 1966.

Gosper, R. W. *Table of Simple Continued Fraction for π and the Derived Decimal Approximation.* Stanford, CA: Artificial Intelligence Laboratory, Stanford University, Oct. 1975. Reviewed in *Math. Comput.* **31**, 1044, 1977.

Hardy, G. H. *A Course of Pure Mathematics, 10th ed.* Cambridge, England: Cambridge University Press, 1952.

Hatzipolakis, A. P. "PiPhilology." http://users.hol.gr/~xpolakis/piphil.html.

Hobsen, E. W. *Squaring the Circle.* New York: Chelsea, 1988.

Johnson-Hill, N. "Extraordinary Pi." http://www.users.globalnet.co.uk/~nickjh/Pi.htm.

Johnson-Hill, N. "The Biggest Selection of Pi Links on the Internet." http://www.users.globalnet.co.uk/~nickjh/pi_links.htm.

Kanada, Y. "New World Record of Pi: 51.5 Billion Decimal Digits." http://www.cecm.sfu.ca/personal/jborwein/Kanada_50b.html.

Klein, F. *Famous Problems.* New York: Chelsea, 1955.

Knopp, K. §32, 136, and 138 in *Theory and Application of Infinite Series.* New York: Dover, p. 238, 1990.

Laczkovich, M. "On Lambert's Proof of the Irrationality of π." *Amer. Math. Monthly* **104**, 439–443, 1997.

Lambert, J. H. "Mémoire sur quelques propriétés remarquables des quantités transcendantes circulaires et logarithmiques." *Mémoires de l'Academie des sciences de Berlin* **17**, 265–322, 1761.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, pp. 22 and 50, 1983.

Lindemann, F. "Über die Zahl π." *Math. Ann.* **20**, 213–225, 1882.

Lopez, A. "Indiana Bill Sets the Value of π to 3." http://daisy.uwaterloo.ca/~alopez-o/math-faq/mathtext/node19.html.

MacTutor Archive. "Pi Through the Ages." http://www-groups.dcs.st-and.ac.uk/~history/HistTopics/Pi_through_the_ages.html.

Mahler, K. "On the Approximation of π." *Nederl. Akad. Wetensch. Proc. Ser. A.* **56**/*Indagationes Math.* **15**, 30–42, 1953.

Ogilvy, C. S. "Pi and Pi-Makers." Ch. 10 in *Excursions in Mathematics.* New York: Dover, pp. 108–120, 1994.

Olds, C. D. *Continued Fractions.* New York: Random House, pp. 59–60, 1963.

Pappas, T. "Probability and π." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 18–19, 1989.

Peterson, I. *Islands of Truth: A Mathematical Mystery Cruise.* New York: W. H. Freeman, pp. 178–186, 1990.

Pickover, C. A. *Keys to Infinity.* New York: Wiley, p. 62, 1995.

Plouffe, S. "Plouffe's Inverter: Table of Current Records for the Computation of Constants." `http://lacim.uqam.ca/pi/records.html`.

Plouffe, S. "People Who Computed Pi." `http://www.cecm.sfu.ca/projects/ISC/records.html`.

Plouffe, S. "Plouffe's Inverter: A Few Approximations of Pi." `http://www.lacim.uqam.ca/pi/approxpi.html`.

Plouffe, S. "The π Page." `http://www.cecm.sfu.ca/pi/`.

Plouffe, S. "Table of Computation of Pi from 2000 BC to Now." `http://www.cecm.sfu.ca/projects/ISC/Pihistory.html`.

Preston, R. "Mountains of Pi." *New Yorker* **68**, 36–67, Mar. 2, 1992. `http://www.lacim.uqam.ca/plouffe/Chudnovsky.html`.

Project Mathematics! *The Story of Pi.* Videotape (24 minutes). California Institute of Technology. Available from the Math. Assoc. Amer.

Rabinowitz, S. and Wagon, S. "A Spigot Algorithm for the Digits of π." *Amer. Math. Monthly* **102**, 195–203, 1995.

Ramanujan, S. "Modular Equations and Approximations to π." *Quart. J. Pure. Appl. Math.* **45**, 350–372, 1913–1914.

Rudio, F. "Archimedes, Huygens, Lambert, Legendre." In *Vier Abhandlungen über die Kreismessung.* Leipzig, Germany, 1892.

Shanks, D. "Dihedral Quartic Approximations and Series for π." *J. Number. Th.* **14**, 397–423, 1982.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, 1993.

Singh, S. *Fermat's Enigma: The Epic Quest to Solve the World's Greatest Mathematical Problem.* New York: Walker, pp. 17–18, 1997.

Sloane, N. J. A. Sequences A000796/M2218, A001203/M2646, A002486/M4456, and A002491/M1009 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sloane, N. J. A. Sequences A025547, A001901, A046126, and A007509/M2061 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sloane, N. J. A. Sequences A032523, A033089, A033090, and A002485/M3097 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Støllum, H.-H. "River Meandering as a Self-Organization Process." *Science* **271**, 1710–1713, 1996.

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 159, 1991.

Viète, F. *Uriorum de rebus mathematicis responsorum,* liber VIII, 1593.

Wagon, S. "Is π Normal?" *Math. Intel.* **7**, 65–67, 1985.

Whitcomb, C. "Notes on Pi (π)." `http://witcombe.bcpw.sbc.edu/EMPi.html`.

Woon, S. C. "Problem 1441." *Math. Mag.* **68**, 72–73, 1995.

Pi Heptomino

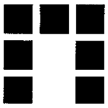

A HEPTOMINO in the shape of the Greek character π.

Piano Mover's Problem

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Given an open subset U in n-D space and two compact subsets C_0 and C_1 of U, where C_1 is derived from C_0 by a continuous motion, is it possible to move C_0 to C_1 while remaining entirely inside U?

see also MOVING LADDER CONSTANT, MOVING SOFA CONSTANT

References

Buchberger, B.; Collins, G. E.; and Kutzler, B. "Algebraic Methods in Geometry." *Annual Rev. Comput. Sci.* **3**, 85–119, 1988.

Feinberg, E. B. and Papadimitriou, C. H. "Finding Feasible Points for a Two-point Body." *J. Algorithms* **10**, 109–119, 1989.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/sofa/sofa.html`.

Leven, D. and Sharir, M. "An Efficient and Simple Motion Planning Algorithm for a Ladder Moving in Two-Dimensional Space Amidst Polygonal Barriers." *J. Algorithms* **8**, 192–215, 1987.

Picard's Existence Theorem

If f is a continuous function that satisfies the LIPSCHITZ CONDITION

$$|f(x,t) - f(y,t)| \leq L|x-y|$$

in a surrounding of $(x_0, t_0) \in \Omega \subset \mathbb{R} \times \mathbb{R}^n = \{(x,t) : |x - x_0| < b, |t - t_0| < a\}$, then the differential equation

$$\frac{df}{dx} = f(x,t)$$
$$x(t_0) = x_0$$

has a unique solution $x(t)$ in the interval $|t - t_0| < d$, where $d = \min(a, b/B)$, min denotes the MINIMUM, $B = \sup |f(t,x)|$, and sup denotes the SUPREMUM.

see also ORDINARY DIFFERENTIAL EQUATION

Picard's Little Theorem

Any ENTIRE ANALYTIC FUNCTION whose range omits two points must be a constant.

Picard's Theorem

An ANALYTIC FUNCTION assumes every COMPLEX NUMBER, with possibly one exception, infinitely often in any NEIGHBORHOOD of an ESSENTIAL SINGULARITY.

see also ANALYTIC FUNCTION, ESSENTIAL SINGULARITY, NEIGHBORHOOD

Picard Variety

Let V be a VARIETY, and write $G(V)$ for the set of divisors, $G_l(V)$ for the set of divisors linearly equivalent to 0, and $G_a(V)$ for the group of divisors algebraically equal to 0. Then $G_a(V)/G_l(V)$ is called the Picard variety. The ALBANESE VARIETY is dual to the Picard variety.

see also ALBANESE VARIETY

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 75, 1980.

Pick's Formula

see PICK'S THEOREM

Pick's Theorem

Let A be the AREA of a simply closed POLYGON whose VERTICES are lattice points. Let B denote the number of LATTICE POINTS on the EDGES and I the number of points in the interior of the POLYGON. Then

$$A = I + \tfrac{1}{2}B - 1.$$

The FORMULA has been generalized to 3-D and higher dimensions using EHRHART POLYNOMIALS.

see also BLICHFELDT'S THEOREM, EHRHART POLYNOMIAL, LATTICE POINT, MINKOWSKI CONVEX BODY THEOREM

References

Diaz, R. and Robins, S. "Pick's Formula via the Weierstraß ℘-Function." *Amer. Math. Monthly* **102**, 431–437, 1995.

Ewald, G. *Combinatorial Convexity and Algebraic Geometry.* New York: Springer-Verlag, 1996.

Hammer, J. *Unsolved Problems Concerning Lattice Points.* London: Pitman, 1977.

Morelli, R. "Pick's Theorem and the Todd Class of a Toric Variety." *Adv. Math.* **100**, 183–231, 1993.

Pick, G. "Geometrisches zur Zahlentheorie." *Sitzenber. Lotos (Prague)* **19**, 311–319, 1899.

Steinhaus, H. *Mathematical Snapshots, 3rd American ed.* New York: Oxford University Press, pp. 97–98, 1983.

Picone's Theorem

Let $f(x)$ be integrable in $[-1,1]$, let $(1-x^2)f(x)$ be of bounded variation in $[-1,1]$, let M' denote the least upper bound of $|f(x)(1-x^2)|$ in $[-1,1]$, and let V' denote the total variation of $f(x)(1-x^2)$ in $[-1,1]$. Given the function

$$F(x) = F(-1) + \int_1^x f(x)\,dx,$$

then the terms of its LEGENDRE SERIES

$$F(x) \sim \sum_{n=0}^{\infty} a_n P_n(x)$$

$$a_n = \tfrac{1}{2}(2n+1)\int_{-1}^{1} F(x)P_n(x)\,dx,$$

where $P_n(x)$ is a LEGENDRE POLYNOMIAL, satisfy the inequalities

$$|a_nP_n(x)| < \begin{cases} 8\sqrt{\frac{2}{\pi}}\frac{M'+V'}{(1-\delta^2)^{1/4}}n^{-3/2} & \text{for } |x| \leq \delta < 1 \\ 2(M'+V')n^{-1} & \text{for } |x| \leq 1 \end{cases}$$

for $n \geq 1$ (Sansone 1991).

see also JACKSON'S THEOREM, LEGENDRE SERIES

References

Picone, M. *Appunti di Analise Superiore.* Naples, Italy,, p. 260, 1940.

Sansone, G. *Orthogonal Functions, rev. English ed.* New York: Dover, pp. 203–205, 1991.

Pie Cutting

see CIRCLE CUTTING, CYLINDER CUTTING, PANCAKE THEOREM, PIZZA THEOREM

Piecewise Circular Curve

A curve composed exclusively of circular ARCS, e.g., the FLOWER OF LIFE, LENS, REULEAUX TRIANGLE, SEED OF LIFE, and YIN-YANG.

see also ARC, REULEAUX TRIANGLE, YIN-YANG FLOWER OF LIFE, LENS, REULEAUX POLYGON, REULEAUX TRIANGLE, SALINON, SEED OF LIFE, TRIANGLE ARCS, YIN-YANG

References

Banchoff, T. and Giblin, P. "On The Geometry Of Piecewise Circular Curves." *Amer. Math. Monthly* **101**, 403–416, 1994.

Pigeonhole Principle

see DIRICHLET'S BOX PRINCIPLE

Pillai's Conjecture

For every $k > 1$, there exist only finite many pairs of POWERS (p, p') with p and p' PRIME and $k = p' - p$.

References

Ribenboim, P. "Catalan's Conjecture." *Amer. Math. Monthly* **103**, 529–538, 1996.

Pilot Vector

see VECTOR SPHERICAL HARMONIC

Pinch Point

A singular point such that every NEIGHBORHOOD of the point intersects itself. Pinch points are also called Whitney singularities or branch points.

Pinching Theorem

Let $g(x) \leq f(x) \leq h(x)$ for all x in some open interval containing a. If

$$\lim_{\Delta x \to a} g(x) = \lim_{\Delta x \to a} h(x) = L,$$

then $\lim_{\Delta x \to a} f(x) = L$.

Pine Cone Number

see FIBONACCI NUMBER

Piriform

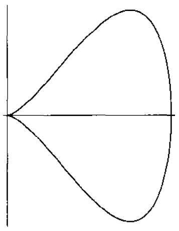

A plane curve also called the PEG TOP and given by the CARTESIAN equation

$$a^4y^2 = b^2x^3(2a - x) \tag{1}$$

and the parametric curves

$$x = a(1 + \sin t) \tag{2}$$

$$y = b\cos t(1 + \sin t) \tag{3}$$

for $t \in [-\pi/2, \pi/2]$. It was studied by G. de Longchamps in 1886. The generalization to a QUARTIC 3-D surface

$$(x^4 - x^3) + y^2 + z^2 = 0, \tag{4}$$

is shown below (Nordstrand).

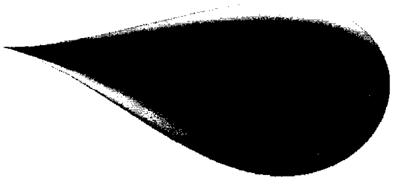

see also BUTTERFLY CURVE, DUMBBELL CURVE, EIGHT CURVE, HEART SURFACE, PEAR CURVE

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub. p. 71, 1989.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 148–150, 1972.

Nordstrand, T. "Surfaces." `http://www.uib.no/people/nfytn/surfaces.htm`.

Pisot-Vijayaraghavan Constants

Let θ be a number greater than 1, λ a POSITIVE number, and

$$(x) \equiv x - \lfloor x \rfloor \tag{1}$$

denote the fractional part of x. Then for a given λ, the sequence of numbers $(\lambda\theta^n)$ for $n = 1, 2, \ldots$ is uniformly distributed in the interval (0, 1) when θ does not belong to a λ-dependent exceptional set S of MEASURE zero (Koksma 1935). Pisot (1938) and Vijayaraghavan (1941) independently studied the exceptional values of θ, and Salem (1943) proposed calling such values Pisot-Vijayaraghavan numbers.

Pisot (1938) proved that if θ is such that there exists a $\lambda \neq 0$ such that the series $\sum_{n=0}^{\infty} \sin^2(\pi\lambda\theta)^n$ converges, then θ is an ALGEBRAIC INTEGER whose conjugates all (except for itself) have modulus < 1, and λ is an algebraic INTEGER of the FIELD $K(\theta)$. Vijayaraghavan (1940) proved that the set of Pisot-Vijayaraghavan numbers has infinitely many limit points. Salem (1944) proved that the set of Pisot-Vijayaraghavan constants is closed. The proof of this theorem is based on the LEMMA that for a Pisot-Vijayaraghavan constant θ, there always exists a number λ such that $1 \leq \lambda < \theta$ and the following inequality is satisfied,

$$\sum_{n=0}^{\infty} \sin^2(\pi\lambda\theta^n) \leq \frac{\pi^2(2\theta + 1)^2}{(\theta - 1)^2}. \tag{2}$$

The smallest Pisot-Vijayaraghavan constant is given by the POSITIVE ROOT θ_0 of

$$x^3 - x - 1 = 0. \tag{3}$$

This number was identified as the smallest known by Salem (1944), and proved to be the smallest possible by Siegel (1944). Siegel also identified the next smallest Pisot-Vijayaraghavan constant θ_1 as the root of

$$x^4 - x^3 - 1 = 0, \tag{4}$$

showed that θ_1 and θ_2 are isolated in S, and showed that the roots of each POLYNOMIAL

$$x^n(x^2 - x - 1) + x^2 - 1 \qquad n = 1, 2, 3, \ldots \tag{5}$$

$$x^n - \frac{x^{n+1} - 1}{x^2 - 1} \qquad n = 3, 5, 7, \ldots \tag{6}$$

$$x^n - \frac{x^{n-1} - 1}{x - 1} \qquad n = 3, 5, 7, \ldots \tag{7}$$

belong to S, where $\theta_0 = \phi$ (the GOLDEN MEAN) is the accumulation point of the set (in fact, the smallest; Le Lionnais 1983, p. 40). Some small Pisot-Vijayaraghavan constants and their POLYNOMIALS are given in the following table. The latter two entries are from Boyd (1977).

k	Number	Order	Polynomial
0	1.3247179572	3	1 0 −1 −1
1	1.3802775691	4	1 −1 0 0 −1
	1.6216584885	16	1 −2 2 −3 2 −2 1 0 0 1 −1 2 −2 2 −2 1 −1
	1.8374664495	20	1 −2 0 1 −1 0 1 −1 0 1 0 −1 0 1 −1 0 1 −1 0 1 −1

All the points in S less than ϕ are known (Dufresnoy and Pisot 1955). Each point of S is a limit point from both sides of the set T of SALEM CONSTANTS (Salem 1945).

see also SALEM CONSTANTS

References
Boyd, D. W. "Small Salem Numbers." *Duke Math. J.* **44**, 315–328, 1977.
Dufresnoy, J. and Pisot, C. "Étude de certaines fonctions méromorphes bornées sur le cercle unité, application à un ensemble fermé d'entiers algébriques." *Ann. Sci. École Norm. Sup.* **72**, 69–92, 1955.
Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, pp. 38 and 148, 1983.
Koksma, J. F. "Ein mengentheoretischer Satz über die Gleichverteilung modulo Eins." *Comp. Math.* **2**, 250–258, 1935.
Pisot, C. "La répartition modulo 1 et les nombres algébriques." *Annali di Pisa* **7**, 205–248, 1938.
Salem, R. "Sets of Uniqueness and Sets of Multiplicity." *Trans. Amer. Math. Soc.* **54**, 218–228, 1943.
Salem, R. "A Remarkable Class of Algebraic Numbers. Proof of a Conjecture of Vijayaraghavan." *Duke Math. J.* **11**, 103–108, 1944.
Salem, R. "Power Series with Integral Coefficients." *Duke Math. J.* **12**, 153–172, 1945.
Siegel, C. L. "Algebraic Numbers whose Conjugates Lie in the Unit Circle." *Duke Math. J.* **11**, 597–602, 1944.
Vijayaraghavan, T. "On the Fractional Parts of the Powers of a Number, II." *Proc. Cambridge Phil. Soc.* **37**, 349–357, 1941.

Pistol

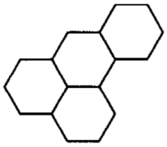

A 4-POLYHEX.

References
Gardner, M. *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, p. 147, 1978.

Pitchfork Bifurcation

Let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a one-parameter family of C^3 map satisfying

$$f(-x,\mu) = -f(x,\mu) \tag{1}$$

$$\left[\frac{\partial f}{\partial x}\right]_{\mu=0,x=0} = 1 \tag{2}$$

$$\left[\frac{\partial f}{\partial x}\right]_{\mu,x} = \left[\frac{\partial f}{\partial x}\right]_{\mu=0,x=\mu} \tag{3}$$

$$\left[\frac{\partial^2 f}{\partial x \partial \mu}\right]_{0,0} > 0 \tag{4}$$

$$\left[\frac{\partial^3 f}{\partial \mu^3}\right]_{\mu=0,x=0} < 0. \tag{5}$$

Then there are intervals having a single stable fixed point and three fixed points (two of which are stable and one of which is unstable). This BIFURCATION is called a pitchfork bifurcation. An example of an equation displaying a pitchfork bifurcation is

$$\dot{x} = \mu x - x^3 \tag{6}$$

(Guckenheimer and Holmes 1997, p. 145).

see also BIFURCATION

References
Guckenheimer, J. and Holmes, P. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields, 3rd ed.* New York: Springer-Verlag, pp. 145 and 149–150, 1997.
Rasband, S. N. *Chaotic Dynamics of Nonlinear Systems.* New York: Wiley, p. 31, 1990.

Pivot Theorem

If the VERTICES A, B, and C of TRIANGLE ΔABC lie on sides QR, RP, and PQ of the TRIANGLE ΔPQR, then the three CIRCLES CBP, ACQ, and BAR have a common point. In extended form, it is MIQUEL'S THEOREM.

see also MIQUEL'S THEOREM

References
Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* New York: Random House, pp. 61–62, 1967.
Forder, H. G. *Geometry.* London: Hutchinson, p. 17, 1960.

Pivoting

The element in the diagonal of a matrix by which other elements are divided in an algorithm such as GAUSS-JORDAN ELIMINATION is called the pivot element. Partial pivoting is the interchanging of rows and full pivoting is the interchanging of both rows and columns in order to place a particularly "good" element in the diagonal position prior to a particular operation.

see also GAUSS-JORDAN ELIMINATION

References
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 29–30, 1992.

Pizza Theorem

If a circular pizza is divided into 8, 12, 16, ... slices by making cuts at equal angles from an arbitrary point, then the sums of the areas of alternate slices are equal.

Place (Digit)

see DIGIT

Place (Field)

A place ν of a number FIELD k is an ISOMORPHISM class of field maps k onto a dense subfield of a nondiscrete locally compact FIELD k_ν.

In the function field case, let F be a function field of algebraic functions of one variable over a FIELD K. Then by a place in F, we mean a subset p of F which is the IDEAL of nonunits of some VALUATION RING O over K.

References

Chevalley, C. *Introduction to the Theory of Algebraic Functions of One Variable.* Providence, RI: Amer. Math. Soc., p. 2, 1951.

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Place (Game)

For n players, $n-1$ games are needed to fairly determine first place, and $n-1+\lg(n-1)$ are needed to fairly determine first and second place.

Planar Bubble Problem

see BUBBLE

Planar Distance

For n points in the PLANE, there are at least

$$N_1 = \sqrt{n - \tfrac{3}{4}} - \tfrac{1}{2}$$

different DISTANCES. The minimum DISTANCE can occur only $\leq 3n-6$ times, and the MAXIMUM DISTANCE can occur $\leq n$ times. Furthermore, no DISTANCE can occur as often as

$$N_2 = \tfrac{1}{4}n(1+\sqrt{8n-7}) < \frac{n^{3/2}}{\sqrt{2}} - \frac{n}{4}$$

times. No set of $n > 6$ points in the PLANE can determine only ISOSCELES TRIANGLES.

see also DISTANCE

References

Honsberger, R. "The Set of Distances Determined by n Points in the Plane." Ch. 12 in *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., pp. 111–135, 1976.

Planar Graph

A GRAPH is planar if it can be drawn in a PLANE without EDGES crossing (i.e., it has CROSSING NUMBER 0). Only planar graphs have DUALS. If G is planar, then G has VERTEX DEGREE ≤ 5. COMPLETE GRAPHS are planar only for $n \leq 4$. The complete BIPARTITE GRAPH $K(3,3)$ in nonplanar. More generally, Kuratowski proved in 1930 that a graph is planar IFF it does not contain within it any graph which can be CONTRACTED to the pentagonal graph $K(5)$ or the hexagonal graph $K(3,3)$. K_5 can be decomposed into a union of two planar graphs, giving it a "DEPTH" of $E(K_5) = 2$. Simple CRITERIA for determining the depth of graphs are not known. Beineke and Harary (1964, 1965) have shown that if $n \not\equiv 4 \pmod 6$, then

$$E(K_n) = \left\lfloor \tfrac{1}{6}(n+7) \right\rfloor .$$

The DEPTHS of the graphs K_n for $n = 4$, 10, 22, 28, 34, and 40 are 1, 3, 4, 5, 6, and 7 (Meyer 1970).

see also COMPLETE GRAPH, FABRY IMBEDDING, INTEGRAL DRAWING, PLANAR STRAIGHT LINE GRAPH

References

Beineke, L. W. and Harary, F. "On the Thickness of the Complete Graph." *Bull. Amer. Math. Soc.* **70**, 618–620, 1964.

Beineke, L. W. and Harary, F. "The Thickness of the Complete Graph." *Canad. J. Math.* **17**, 850–859, 1965.

Booth, K. S. and Lueker, G. S. "Testing for the Consecutive Ones Property, Interval Graphs, and Graph Planarity using PQ-Tree Algorithms." *J. Comput. System Sci.* **13**, 335–379, 1976.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 56, 1983.

Meyer, J. "L'épaisseur des graphes completes K_{34} et K_{40}." *J. Comp. Th.* **9**, 1970.

Planar Point

A point $\mathbf{p}$ on a REGULAR SURFACE $M \in \mathbb{R}^3$ is said to be planar if the GAUSSIAN CURVATURE $K(\mathbf{p}) = 0$ and $S(\mathbf{p}) = 0$ (where S is the SHAPE OPERATOR), or equivalently, both of the PRINCIPAL CURVATURES κ_1 and κ_2 are 0.

see also ANTICLASTIC, ELLIPTIC POINT, GAUSSIAN CURVATURE, HYPERBOLIC POINT, PARABOLIC POINT, SYNCLASTIC

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 280, 1993.

Planar Space

Let (ξ_1, ξ_2) be a locally EUCLIDEAN coordinate system. Then

$$ds^2 = d\xi_1{}^2 + d\xi_2{}^2. \tag{1}$$

Now plug in

$$d\xi_1 = \frac{\partial \xi_1}{\partial x_1} dx_1 + \frac{\partial \xi_1}{\partial x_2} dx_2 \tag{2}$$

$$d\xi_2 = \frac{\partial \xi_2}{\partial x_1} dx_1 + \frac{\partial \xi_2}{\partial x_2} dx_2 \tag{3}$$

to obtain

$$\begin{aligned} ds^2 = {} & \left[\left(\frac{\partial \xi_1}{\partial x_1}\right)^2 + \left(\frac{\partial \xi_2}{\partial x_1}\right)^2\right] dx_1{}^2 \\ & + 2\left[\frac{\partial \xi_1}{\partial x_1}\frac{\partial \xi_1}{\partial x_2} + \frac{\partial \xi_2}{\partial x_1}\frac{\partial \xi_2}{\partial x_2}\right] dx_1\, dx_2 \\ & + \left[\left(\frac{\partial \xi_1}{\partial x_2}\right)^2 + \left(\frac{\partial \xi_2}{\partial x_2}\right)^2\right] dx_2{}^2. \end{aligned} \tag{4}$$

Reading off the COEFFICIENTS from

$$ds^2 = g_{11}\, dx_1{}^2 + 2g_{12}\, dx_1\, dx_2 + g_{22}\,(dx_2)^2 \tag{5}$$

gives

$$g_{11} = \left(\frac{\partial \xi_1}{\partial x_1}\right)^2 + \left(\frac{\partial \xi_2}{\partial x_1}\right)^2 \tag{6}$$

$$g_{12} = \frac{\partial \xi_1}{\partial x_1}\frac{\partial \xi_1}{\partial x_2} + \frac{\partial \xi_2}{\partial x_1}\frac{\partial \xi_2}{\partial x_2} \tag{7}$$

$$g_{22} = \left(\frac{\partial \xi_1}{\partial x_2}\right)^2 + \left(\frac{\partial \xi_2}{\partial x_2}\right)^2 . \tag{8}$$

Making a change of coordinates $(x_1, x_2) \to (x_1', x_2')$ gives

$$\begin{aligned} g_{11}' &= \left(\frac{\partial \xi_1}{\partial x_1'}\right)^2 + \left(\frac{\partial \xi_2}{\partial x_1'}\right)^2 \\ &= \left(\frac{\partial \xi_1}{\partial x_1}\frac{\partial x_1}{\partial x_1'} + \frac{\partial \xi_1}{\partial x_2}\frac{\partial x_2}{\partial x_1'}\right)^2 \\ &\quad + \left(\frac{\partial \xi_2}{\partial x_1}\frac{\partial x_1}{\partial x_1'} + \frac{\partial \xi_2}{\partial x_2}\frac{\partial x_2}{\partial x_1'}\right)^2 \\ &= g_{11}\left(\frac{\partial x_1}{\partial x_1'}\right)^2 + 2g_{12}\frac{\partial x_1}{\partial x_1'}\frac{\partial x_2}{\partial x_1'} + g_{22}\left(\frac{\partial x_2}{\partial x_1'}\right)^2 \qquad (9) \\ g_{12}' &= \frac{\partial \xi_1}{\partial x_1}\frac{\partial x_1}{\partial x_1'}\frac{\partial \xi_1}{\partial x_2}\frac{\partial x_2}{\partial x_2'} + \frac{\partial \xi_2}{\partial x_1}\frac{\partial x_1}{\partial x_1'}\frac{\partial \xi_2}{\partial x_2}\frac{\partial x_2}{\partial x_2'} \\ &= g_{12}\frac{\partial x_1}{\partial x_1'}\frac{\partial x_2}{\partial x_2'} \qquad (10) \\ g_{22}' &= g_{11}\left(\frac{\partial x_1}{\partial x_1'}\right)^2 + 2g_{12}\frac{\partial x_1}{\partial x_2'}\frac{\partial x_2}{\partial x_2'} + g_{22}\left(\frac{\partial x_2}{\partial x_2'}\right)^2. \qquad (11) \end{aligned}$$

Planar Straight Line Graph

A PLANAR GRAPH in which only straight line segments are used to connect the VERTICES, where the EDGES may intersect.

see also PLANAR GRAPH

Plancherel's Theorem

$$\int_{-\infty}^{\infty} f(x)g^*(x)\,dx = \int_{-\infty}^{\infty} F(s)G^*(s)\,ds,$$

where $F(s) \equiv \mathcal{F}[f(x)]$ and $\mathcal{F}$ denotes a FOURIER TRANSFORM. If f and g are real

$$\int_{-\infty}^{\infty} f(x)g(-x)\,dx = \int_{-\infty}^{\infty} F(s)G(s)\,ds.$$

see also FOURIER TRANSFORM, PARSEVAL'S THEOREM

Planck's Radiation Function

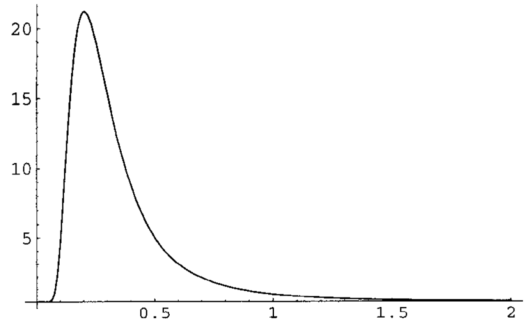

The function

$$f(x) = \frac{1}{x^5(e^{1/x} - 1)}.$$

It has a MAXIMUM at $x \approx 0.201405$, where

$$f'(x) = \frac{5x - e^{1/x}(5x-1)}{x^7(e^{1/x}-1)^2} = 0,$$

and inflection points at $x \approx 0.11842$ and $x \approx 0.283757$, where

$$f''(x) = \frac{e^{1/x}(1+e^{1/x}) + 6x(e^{1/x}-1)[e^{1/x}(5x-2) - 5x]}{(e^{1/x}-1)^3 x^9} = 0.$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Planck's Radiation Function." §27.2 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 999, 1972.

Plane

A plane is a 2-D SURFACE spanned by two linearly independent vectors. The generalization of the plane to higher DIMENSIONS is called a HYPERPLANE.

In intercept form, a plane passing through the points $(a, 0, 0)$, $(0, b, 0)$ and $(0, 0, c)$ is given by

$$\frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 1. \qquad (1)$$

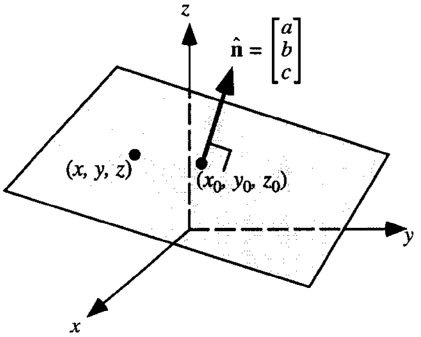

The equation of a plane PERPENDICULAR to the NONZERO VECTOR $\hat{\mathbf{n}} = (a, b, c)$ through the point (x_0, y_0, z_0) is

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} \cdot \begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix} = a(x - x_0) + b(y - y_0) + c(z - z_0) = 0, \qquad (2)$$

so

$$ax + by + cz + d = 0, \qquad (3)$$

where

$$d \equiv -ax_0 - by_0 - cz_0. \qquad (4)$$

A plane specified in this form therefore has x-, y-, and z-intercepts at

$$x = -\frac{d}{a} \tag{5}$$

$$y = -\frac{d}{b} \tag{6}$$

$$z = -\frac{d}{c}, \tag{7}$$

and lies at a DISTANCE

$$h = \frac{|d|}{\sqrt{a^2 + b^2 + c^2}} \tag{8}$$

from the ORIGIN.

The plane through P_1 and parallel to (a_1, b_1, c_1) and (a_2, b_2, c_2) is

$$\begin{vmatrix} x - x_1 & y - y_1 & z - z_1 \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix} = 0. \tag{9}$$

The plane through points P_1 and P_2 parallel to direction (a, b, c) is

$$\begin{vmatrix} x - x_1 & y - y_1 & z - z_1 \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ a & b & c \end{vmatrix} = 0. \tag{10}$$

The three-point form is

$$\begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix} = \begin{vmatrix} x - x_1 & y - y_1 & z - z_1 \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ x_3 - x_1 & y_3 - y_1 & z_3 - z_1 \end{vmatrix} = 0. \tag{11}$$

The DISTANCE from a point (x_1, y_1, z_1) to a plane

$$Ax + By + Cz + D = 0 \tag{12}$$

is

$$d = \frac{Ax_1 + By_1 + Cz_1 + D}{\pm\sqrt{A^2 + B^2 + C^2}}. \tag{13}$$

The DIHEDRAL ANGLE between the planes

$$A_1x + B_1y + C_1z + D_1 = 0 \tag{14}$$

$$A_2x + B_2y + C_2z + D_2 = 0 \tag{15}$$

is

$$\cos\theta = \frac{A_1A_2 + B_1B_2 + C_1C_2}{\sqrt{{A_1}^2 + {B_1}^2 + {C_1}^2}\sqrt{{A_2}^2 + {B_2}^2 + {C_2}^2}}. \tag{16}$$

In order to specify the relative distances of $n > 1$ points in the plane, $1 + 2(n - 2) = 2n - 3$ coordinates are needed, since the first can always be placed at (0, 0) and the second at $(x, 0)$, where it defines the x-AXIS. The remaining $n - 2$ points need two coordinates each. However, the total number of distances is

$${}_nC_2 = \binom{n}{2} = \frac{n!}{2!(n-2)!} = \tfrac{1}{2}n(n-1), \tag{17}$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT, so the distances between points are subject to m relationships, where

$$m \equiv \tfrac{1}{2}n(n-1) - (2n - 3) = \tfrac{1}{2}(n-2)(n-3). \tag{18}$$

For $n = 2$ and $n = 3$, there are no relationships. However, for a QUADRILATERAL (with $n = 4$), there is one (Weinberg 1972).

It is impossible to pick random variables which are uniformly distributed in the plane (Eisenberg and Sullivan 1996). In 4-D, it is possible for four planes to intersect in exactly one point. For every set of n points in the plane, there exists a point O in the plane having the property such that *every* straight line through O has at least 1/3 of the points on each side of it (Honsberger 1985).

Every RIGID motion of the plane is one of the following types (Singer 1995):

1. ROTATION about a fixed point P.
2. TRANSLATION in the direction of a line l.
3. REFLECTION across a line l.
4. Glide-reflections along a line l.

Every RIGID motion of the hyperbolic plane is one of the previous types or a

5. Horocycle rotation.

see also ARGAND PLANE, COMPLEX PLANE, DIHEDRAL ANGLE, ELLIPTIC PLANE, FANO PLANE, HYPERPLANE, MOUFANG PLANE, NIRENBERG'S CONJECTURE, NORMAL SECTION, POINT-PLANE DISTANCE, PROJECTIVE PLANE

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 208–209, 1987.

Eisenberg, B. and Sullivan, R. "Random Triangles n Dimensions." *Amer. Math. Monthly* **103**, 308–318, 1996.

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 189–191, 1985.

Singer, D. A. "Isometries of the Plane." *Amer. Math. Monthly* **102**, 628–631, 1995.

Weinberg, S. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity.* New York: Wiley, p. 7, 1972.

Plane Curve

see CURVE

Plane Cutting

see CIRCLE CUTTING

Plane Division

Consider n intersecting CIRCLES and ELLIPSES. The maximal number of regions in which these divide the PLANE are

$$\begin{aligned} N_{\text{circle}} &= n^2 - n + 2 \\ N_{\text{ellipse}} &= 2n^2 - 2n + 2. \end{aligned}$$

see also ARRANGEMENT, CIRCLE, CUTTING, ELLIPSE, SPACE DIVISION

Plane-Filling Curve

see PLANE-FILLING FUNCTION

Plane-Filling Function

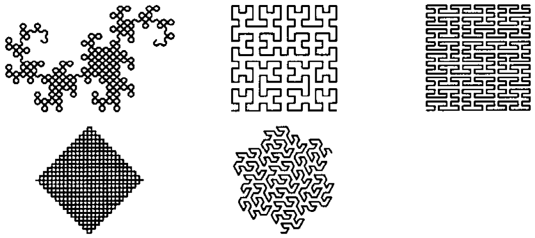

A SPACE-FILLING FUNCTION which maps a 1-D INTERVAL into a 2-D area. Plane-filling functions were thought to be impossible until Hilbert discovered the HILBERT CURVE in 1891.

Plane-filling functions are often (imprecisely) defined to be the "limit" of an infinite sequence of specified curves which "fill" the PLANE without "HOLES," hence the more popular term PLANE-FILLING CURVE. The term "plane-filling function" is preferable to "PLANE-FILLING CURVE" because "curve" informally connotes "GRAPH" (i.e., range) of some continuous function, but the GRAPH of a plane-filling function is a solid patch of 2-space with no evidence of the order in which it was traced (and, for a dense set, retraced). Actually, all that is needed to rigorously define a plane-filling function is an arbitrarily refinable correspondence between contiguous subintervals of the domain and contiguous subareas of the range.

True plane-filling functions are not ONE-TO-ONE. In fact, because they map closed intervals onto closed areas, they cannot help but overfill, revisiting at least twice a dense subset of the filled area. Thus, every point in the filled area has *at least* one inverse image.

see also HILBERT CURVE, PEANO CURVE, PEANO-GOSPER CURVE, SIERPIŃSKI CURVE, SPACE-FILLING FUNCTION, SPACE-FILLING POLYHEDRON

References

Bogomolny, A. "Plane Filling Curves." `http://www.cut-the-knot.com/do_you_know/hilbert.html`.

Wagon, S. "A Space-Filling Curve." §6.3 in *Mathematica in Action.* New York: W. H. Freeman, pp. 196–209, 1991.

Plane Geometry

That portion of GEOMETRY dealing with figures in a PLANE, as opposed to SOLID GEOMETRY. Plane geometry deals with the CIRCLE, LINE, POLYGON, etc.

see also CONSTRUCTIBLE POLYGON, GEOMETRIC CONSTRUCTION, GEOMETRY, SOLID GEOMETRY, SPHERICAL GEOMETRY

References

Altshiller-Court, N. *College Geometry: A Second Course in Plane Geometry for Colleges and Normal Schools, 2nd ed., rev. enl.* New York: Barnes and Noble, 1952.

Casey, J. *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions with Numerous Examples, 2nd rev. enl. ed.* Dublin: Hodges, Figgis, & Co., 1893.

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., 1967.

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, 1969.

Dixon, R. *Mathographics.* New York: Dover, 1991.

Gallatly, W. *The Modern Geometry of the Triangle, 2nd ed.* London: Hodgson, 1913.

Heath, T. L. *The Thirteen Books of the Elements, 2nd ed., Vol. 1: Books I and II.* New York: Dover, 1956.

Heath, T. L. *The Thirteen Books of the Elements, 2nd ed., Vol. 2: Books III–IX.* New York: Dover, 1956.

Heath, T. L. *The Thirteen Books of the Elements, 2nd ed., Vol. 3: Books X–XIII.* New York: Dover, 1956.

Hilbert, D. *The Foundations of Geometry.* Chicago, IL: Open Court, 1980.

Hilbert, D. and Cohn-Vossen, S. *Geometry and the Imagination.* New York: Chelsea, 1952.

Honsberger, R. *Episodes in Nineteenth and Twentieth Century Euclidean Geometry.* Washington, DC: Math. Assoc. Amer., 1995.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, 1929.

Kimberling, C. "Triangle Centers and Central Triangles." *Congr. Numer.* **129**, 1–295, 1998.

Klee, V. "Some Unsolved Problems in Plane Geometry." *Math. Mag.* **52**, 131–145, 1979.

Klee, V. and Wagon, S. *Old and New Unsolved Problems in Plane Geometry and Number Theory, rev. ed.* Washington, DC: Math. Assoc. Amer., 1991.

Pedoe, D. *Circles: A Mathematical View, rev. ed.* Washington, DC: Math. Assoc. Amer., 1995.

Plane Partition

A two-dimensional array of INTEGERS nonincreasing both left to right and top to bottom which add up to a given number, i.e., $n_{ij} \geq n_{i(j+1)}$ and $n_{ij} \geq n_{(i+1)j}$. For example, a planar partition of 2 is given by

$$\begin{array}{ccccc} 5 & 4 & 2 & 1 & 1 \\ 3 & 2 \\ 2 & 2. \end{array}$$

The GENERATING FUNCTION for the number $\mathrm{PL}(n)$ of planar partitions of n is

$$\sum_{n=0}^{\infty} \mathrm{PL}(n) x^n = \frac{1}{\prod_{k=1}^{\infty}(1-x^k)^k} = 1 + x + 3x^2 + 6x^3 + 13x^4 + 24x^5 + \ldots$$

(Sloane's A000219, MacMahon 1912b, Beeler *et al.* 1972, Bender and Knuth 1972). The concept of planar partitions can also be generalized to cubic partitions.

see also PARTITION, SOLID PARTITION

References
Beeler, M.; Gosper, R. W.; and Schroeppel, R. Item 18 in *HAKMEM*. Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.
Bender, E. A. and Knuth, D. E. "Enumeration of Plane Partitions." *J. Combin. Theory Ser. A.* **13**, 40–54, 1972.
Knuth, D. E. "A Note on Solid Partitions." *Math. Comput.* **24**, 955–961, 1970.
MacMahon, P. A. "Memoir on the Theory of the Partitions of Numbers. V: Partitions in Two-Dimensional Space." *Phil. Trans. Roy. Soc. London Ser. A* **211**, 75–110, 1912a.
MacMahon, P. A. "Memoir on the Theory of the Partitions of Numbers. VI: Partitions in Two-Dimensional Space, to which is Added an Adumbration of the Theory of Partitions in Three-Dimensional Space." *Phil. Trans. Roy. Soc. London Ser. A* **211**, 345–373, 1912b.
MacMahon, P. A. *Combinatory Analysis, Vol. 2.* New York: Chelsea, 1960.
Sloane, N. J. A. Sequence A000219/M2566 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Plane Symmetry Groups

see WALLPAPER GROUPS

Planted Planar Tree

A planted plane tree (V, E, v, α) is defined as a vertex set V, edges set E, ROOT v, and order relation α on V which satisfies

1. For $x, y \in V$ if $\rho(x) < \rho(y)$, then $x \,\alpha\, y$, where $\rho(x)$ is the length of the path from v to x,
2. If $\{r, s\}$, $\{x, y\} \in E$, $\rho(r) = \rho(x) = \rho(s) - 1 = \rho(y) - 1$ and $r \,\alpha\, x$, then $s \,\alpha\, y$

(Klarner 1969, Chorneyko and Mohanty 1975). The CATALAN NUMBERS give the number of planar trivalent planted trees.

see also CATALAN NUMBER, TREE

References
Chorneyko, I. Z. and Mohanty, S. G. "On the Enumeration of Certain Sets of Planted Plane Trees." *J. Combin. Th. Ser. B* **18**, 209–221, 1975.
Harary, F.; Prins, G.; and Tutte, W. T. "The Number of Plane Trees." *Indag. Math.* **26**, 319–327, 1964.
Klarner, D. A. "A Correspondence Between Sets of Trees." *Indag. Math.* **31**, 292–296, 1969.

Plastic Constant

The limiting ratio of the successive terms of the PADOVAN SEQUENCE, $P = 1.32471795\ldots$.

see also PADOVAN SEQUENCE

References
Stewart, I. "Tales of a Neglected Number." *Sci. Amer.* **274**, 102–103, Jun. 1996.

Plat

A BRAID in which strands are intertwined in the center and are free in "handles" on either side of the diagram.

Plateau Curves

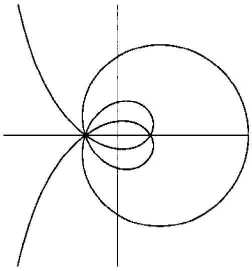

A curve studied by the Belgian physicist and mathematician Joseph Plateau. It has Cartesian equation

$$x = \frac{a \sin[(m+n)t]}{\sin[(m-n)t]}$$
$$y = \frac{2a \sin(mt)\sin(nt)}{\sin[(m-n)t]}.$$

If $m = 2n$, the Plateau curve degenerates to a CIRCLE with center $(1, 0)$ and radius 2.

References
MacTutor History of Mathematics Archive. "Plateau Curves." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Plateau.html`.

Plateau's Laws

BUBBLES can meet only at ANGLES of 120° (for two BUBBLES) and 109.5° (for three BUBBLES), where the exact value of 109.5° is the TETRAHEDRAL ANGLE. This was proved by Jean Taylor using MEASURE THEORY to study AREA minimization. The DOUBLE BUBBLE is AREA minimizing, but it is not known the triple BUBBLE is also AREA minimizing. It is also unknown if empty chambers trapped inside can minimize AREA for $n \geq 3$ BUBBLES.

see also BUBBLE, CALCULUS OF VARIATIONS, DOUBLE BUBBLE, PLATEAU'S PROBLEM

References
Morgan, F. "Mathematicians, including Undergraduates, Look at Soap Bubbles." *Amer. Math. Monthly* **101**, 343–351, 1994.
Taylor, J. E. "The Structure of Singularities in Soap-Bubble-Like and Soap-Film-Like Minimal Surfaces." *Ann. Math.* **103**, 489–539, 1976.

Plateau's Problem

The problem in CALCULUS OF VARIATIONS to find the MINIMAL SURFACE of a boundary with specified constraints. In general, there may be one, multiple, or no MINIMAL SURFACES spanning a given closed curve in space.

see also CALCULUS OF VARIATIONS, MINIMAL SURFACE

References
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 48–49, 1989.
Stuwe, M. *Plateau's Problem and the Calculus of Variations.* Princeton, NJ: Princeton University Press, 1989.

Plato's Number

A number appearing in *The Republic* which involves 216 and 12,960,000.

References

Plato. *The Republic.* New York: Oxford University Press, 1994.

Wells, D. G. *The Penguin Dictionary of Curious and Interesting Numbers.* London: Penguin, p. 144, 1986.

Platonic Solid

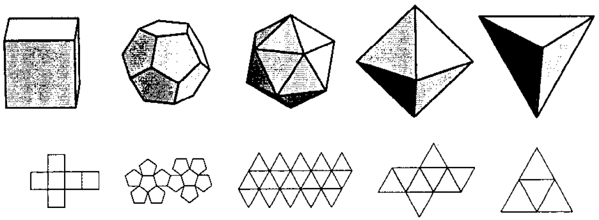

A solid with equivalent faces composed of congruent regular convex Polygons. There are exactly five such solids: the Cube, Dodecahedron, Icosahedron, Octahedron, and Tetrahedron, as was proved by Euclid in the last proposition of the *Elements.*

The Platonic solids were known to the ancient Greeks, and were described by Plato in his *Timaeus* ca. 350 BC. In this work, Plato equated the Tetrahedron with the "element" fire, the Cube with earth, the Icosahedron with water, the Octahedron with air, and the Dodecahedron with the stuff of which the constellations and heavens were made (Cromwell 1997).

The Platonic solids are sometimes also known as the Regular Polyhedra of Cosmic Figures (Cromwell 1997), although the former term is sometimes used to refer collectively to both the Platonic solids *and* Kepler-Poinsot Solids (Coxeter 1973).

If P is a Polyhedron with congruent (convex) regular polygonal faces, then Cromwell (1997, pp. 77–78) shows that the following statements are equivalent.

1. The vertices of P all lie on a Sphere.
2. All the Dihedral Angles are equal.
3. All the Vertex Figures are Regular Polygons.
4. All the Solid Angles are equivalent.
5. All the vertices are surrounded by the same number of Faces.

Let v (sometimes denoted N_0) be the number of Vertices, e (or N_1) the number of Edges, and f (or N_2) the number of Faces. The following table gives the Schläfli Symbol, Wythoff Symbol, and C&R symbol, the number of vertices v, edges e, and faces f, and the Point Groups for the Platonic solids (Wenninger 1989).

Solid	Schläfli	Wyth.	C&R	v	e	f	Grp
cube	$\{4,3\}$	$3 \mid 2\ 4$	4^3	8	12	6	O_h
dodecahedron	$\{5,3\}$	$3 \mid 2\ 5$	5^3	20	30	12	I_h
icosahedron	$\{3,5\}$	$5 \mid 2\ 3$	3^5	12	30	20	I_h
octahedron	$\{3,4\}$	$4 \mid 2\ 3$	3^4	6	12	8	O_h
tetrahedron	$\{3,3\}$	$3 \mid 2\ 3$	3^3	4	6	4	T_d

Let r be the Inradius, ρ the Midradius, and R the Circumradius. The following two tables give the analytic and numerical values of these distances for Platonic solids with unit side length.

Solid	r	ρ	R
cube	$\frac{1}{2}$	$\frac{1}{2}\sqrt{2}$	$\frac{1}{2}\sqrt{3}$
dodecahedron	$\frac{1}{20}\sqrt{250+110\sqrt{5}}$	$\frac{1}{4}(3+\sqrt{5})$	$\frac{1}{4}(\sqrt{15}+\sqrt{3})$
icosahedron	$\frac{1}{12}(3\sqrt{3}+\sqrt{15})$	$\frac{1}{4}(1+\sqrt{5})$	$\frac{1}{4}\sqrt{10+2\sqrt{5}}$
octahedron	$\frac{1}{6}\sqrt{6}$	$\frac{1}{2}$	$\frac{1}{2}\sqrt{2}$
tetrahedron	$\frac{1}{12}\sqrt{6}$	$\frac{1}{4}\sqrt{2}$	$\frac{1}{4}\sqrt{6}$

Solid	r	ρ	R
cube	0.5	0.70711	0.86603
dodecahedron	1.11352	1.30902	1.40126
icosahedron	0.75576	0.80902	0.95106
octahedron	0.40825	0.5	0.70711
tetrahedron	0.20412	0.35355	0.61237

Finally, let A be the Area of a single Face, V be the Volume of the solid, the Edges be of unit length on a side, and α be the Dihedral Angle. The following table summarizes these quantities for the Platonic solids.

Solid	A	V	α
cube	1	1	$\frac{1}{2}\pi$
dodecahedron	$\frac{1}{4}\sqrt{25+10\sqrt{5}}$	$\frac{1}{4}(15+7\sqrt{5})$	$\cos^{-1}(-\frac{1}{5}\sqrt{5})$
icosahedron	$\frac{1}{4}\sqrt{3}$	$\frac{5}{12}(3+\sqrt{5})$	$\cos^{-1}(-\frac{1}{3}\sqrt{5})$
octahedron	$\frac{1}{4}\sqrt{3}$	$\frac{1}{3}\sqrt{2}$	$\cos^{-1}(-\frac{1}{3})$
tetrahedron	$\frac{1}{4}\sqrt{3}$	$\frac{1}{12}\sqrt{2}$	$\cos^{-1}(\frac{1}{3})$

The number of Edges meeting at a Vertex is $2e/v$. The Schläfli Symbol can be used to specify a Platonic solid. For the solid whose faces are p-gons (denoted $\{p\}$), with q touching at each Vertex, the symbol is $\{p,q\}$. Given p and q, the number of Vertices, Edges, and faces are given by

$$N_0 = \frac{4p}{4-(p-2)(q-2)}$$
$$N_1 = \frac{2pq}{4-(p-2)(q-2)}$$
$$N_2 = \frac{4q}{4-(p-2)(q-2)}.$$

Minimal Surfaces for Platonic solid frames are illustrated in Isenberg (1992, pp. 82–83).

see also Archimedean Solid, Catalan Solid, Johnson Solid, Kepler-Poinsot Solid, Quasiregular Polyhedron, Uniform Polyhedron

References

Artmann, B. "Symmetry Through the Ages: Highlights from the History of Regular Polyhedra." In *In Eves' Circles* (Ed. J. M. Anthony). Washington, DC: Math. Assoc. Amer., pp. 139–148, 1994.

Ball, W. W. R. and Coxeter, H. S. M. "Polyhedra." Ch. 5 in *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 131–136, 1987.

Behnke, H.; Bachman, F.; Fladt, K.; and Kunle, H. (Eds.). *Fundamentals of Mathematics, Vol. 2.* Cambridge, MA: MIT Press, p. 272, 1974.

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 128–129, 1987.

Bogomolny, A. "Regular Polyhedra." `http://www.cut-the-knot.com/do_you_know/polyhedra.html`.

Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, pp. 1–17, 93, and 107–112, 1973.

Critchlow, K. *Order in Space: A Design Source Book.* New York: Viking Press, 1970.

Cromwell, P. R. *Polyhedra.* New York: Cambridge University Press, pp. 51–57, 66–70, and 77–78, 1997.

Dunham, W. *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 78–81, 1990.

Gardner, M. "The Five Platonic Solids." Ch. 1 in *The Second Scientific American Book of Mathematical Puzzles & Diversions: A New Selection.* New York: Simon and Schuster, pp. 13–23, 1961.

Heath, T. *A History of Greek Mathematics, Vol. 1.* Oxford, England: Oxford University Press, p. 162, 1921.

Isenberg, C. *The Science of Soap Films and Soap Bubbles.* New York: Dover, 1992.

Kepler, J. *Opera Omnia, Vol. 5.* Frankfort, p. 121, 1864.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 129–131, 1990.

Pappas, T. "The Five Platonic Solids." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 39 and 110–111, 1989.

Rawles, B. A. "Platonic and Archimedean Solids—Faces, Edges, Areas, Vertices, Angles, Volumes, Sphere Ratios." `http://www.intent.com/sg/polyhedra.html`.

Steinhaus, H. "Platonic Solids, Crystals, Bees' Heads, and Soap." Ch. 8 in *Mathematical Snapshots, 3rd American ed.* New York: Oxford University Press, 1960.

Waterhouse, W. "The Discovery of the Regular Solids." *Arch. Hist. Exact Sci.* **9**, 212–221, 1972–1973.

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, 1971.

Platykurtic

A distribution with FISHER KURTOSIS $\gamma_2 < 0$ (and therefore having a flattened shape).

see also FISHER KURTOSIS

Playfair's Axiom

Through any point in space, there is exactly one straight line PARALLEL to a given straight line. This AXIOM is equivalent to the PARALLEL AXIOM.

see also PARALLEL AXIOM

References

Dunham, W. "Hippocrates' Quadrature of the Lune." Ch. 1 in *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, p. 54, 1990.

Plethysm

A group theoretic operation which is useful in the study of complex atomic spectra. A plethysm takes a set of functions of a given symmetry type $\{\mu\}$ and forms from them symmetrized products of a given degree r and other symmetry type $\{\nu\}$. A plethysm

$$\{\mu\}\otimes\{\nu\} = \sum\{\lambda\}$$

satisfies the rules

$$A\otimes(BC) = (A\otimes B)(A\otimes C) = A\otimes BA\otimes C,$$

$$A\otimes(B\pm C) = A\otimes B\pm A\otimes C$$

$$(A\otimes B)\otimes C = A\otimes(B\otimes C)$$

$$(A+B)\otimes\{\lambda\} = \sum\Gamma_{\mu\nu\lambda}(A\otimes\{\mu\})(B\otimes\{\nu\}),$$

where $\Gamma_{\mu\nu\lambda}$ is the coefficient of $\{\lambda\}$ in $\{\mu\}\{\nu\}$,

$$(A-B)\otimes\{\lambda\} = \sum(-1)^r\Gamma_{\mu\nu\lambda}(A\otimes\{\mu\})(B\otimes\{\tilde{\nu}\}),$$

where $\{\tilde{\nu}\}$ is the partition of r conjugate to $\{\nu\}$, and

$$(AB)\otimes\{\lambda\} = \sum g_{\mu\nu\lambda}(A\otimes\{\mu\})(B\otimes\{\nu\}),$$

where $g_{\mu\nu\lambda}$ is the coefficient of $\{\lambda\}$ in the inner product $\{\mu\}\circ\{\nu\}$ (Wybourne 1970).

References

Littlewood, D. E. "Polynomial Concomitants and Invariant Matrices." *J. London Math. Soc.* **11**, 49–55, 1936.

Wybourne, B. G. "The Plethysm of S-Functions" and "Plethysm and Restricted Groups." Chs. 6–7 in *Symmetry Principles and Atomic Spectroscopy.* New York: Wiley, pp. 49–68, 1970.

Plot

see GRAPH (FUNCTION)

Plouffe's Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Define the function

$$\rho(x) \equiv \begin{cases} 1 & \text{for } x < 0 \\ 0 & \text{for } x \geq 0. \end{cases} \tag{1}$$

Let

$$a_n = \sin(2^n) = \begin{cases} \sin 1 & \text{for } n = 0 \\ 2a_0\sqrt{1-{a_0}^2} & \text{for } n = 1 \\ 2a_{n-1}(1-2{a_{n-2}}^2) & \text{for } n \geq 2, \end{cases} \tag{2}$$

then

$$\sum_{n=0}^{\infty}\frac{\rho(a_n)}{2^{n+1}} = \frac{1}{2\pi}. \tag{3}$$

For

$$b_n = \cos(2^n) = \begin{cases} \cos 1 & \text{for } n = 0 \\ 2{b_{n-1}}^2 - 1 & \text{for } n \geq 1, \end{cases} \tag{4}$$

and

$$\sum_{n=0}^{\infty} \frac{\rho(b_n)}{2^{n+1}} = 0.4756260767\ldots. \tag{5}$$

Letting

$$c_n = \tan(2^n) = \begin{cases} \tan 1 & \text{for } n = 0 \\ \frac{2c_{n-1}}{1-{c_{n-1}}^2} & \text{for } n \geq 1, \end{cases} \tag{6}$$

then

$$\sum_{n=0}^{\infty} \frac{\rho(c_n)}{2^{n+1}} = \frac{1}{\pi}. \tag{7}$$

Plouffe asked if the above processes could be "inverted." He considered

$$\begin{aligned} \alpha_n &= \sin(2^n \sin^{-1} \tfrac{1}{2}) \\ &= \begin{cases} \frac{1}{2} & \text{for } n = 0 \\ \frac{1}{2}\sqrt{3} & \text{for } n = 1 \\ 2\alpha_{n-1}(1 - 2{\alpha_{n-2}}^2) & \text{for } n \geq 2, \end{cases} \end{aligned} \tag{8}$$

giving

$$\sum_{n=0}^{\infty} \frac{\rho(\alpha_n)}{2^{n+1}} = \tfrac{1}{12}, \tag{9}$$

and

$$\beta_n = \cos(2^n \cos^{-1} \tfrac{1}{2}) = \begin{cases} \frac{1}{2} & \text{for } n = 0 \\ 2{\beta_{n-1}}^2 - 1 & \text{for } n \geq 1, \end{cases} \tag{10}$$

giving

$$\sum_{n=0}^{\infty} \frac{\rho(\beta_n)}{2^{n+1}} = \tfrac{1}{2}, \tag{11}$$

and

$$\gamma_n = \tan(2^n \tan^{-1} \tfrac{1}{2}) = \begin{cases} \frac{1}{2} & \text{for } n = 0 \\ \frac{2\gamma_{n-1}}{1-{\gamma_{n-1}}^2} & \text{for } n \geq 1, \end{cases} \tag{12}$$

giving

$$\sum_{n=0}^{\infty} \frac{\rho(\alpha_n)}{2^{n+1}} = \frac{1}{\pi} \tan^{-1}(\tfrac{1}{2}). \tag{13}$$

The latter is known as Plouffe's constant (Plouffe 1997). The positions of the 1s in the BINARY expansion of this constant are 3, 6, 8, 9, 10, 13, 21, 23, ... (Sloane's A004715).

Borwein and Girgensohn (1995) extended Plouffe's γ_n to *arbitrary* REAL x, showing that if

$$\xi_n = \tan(2^n \tan^{-1} x) = \begin{cases} x & \text{for } n = 0 \\ \frac{2\xi_{n-1}}{1-{\xi_{n-1}}^2} & \text{for } n \geq 1 \text{ and } |\xi_{n-1}| \neq 1 \\ -\infty & \text{for } n \geq 1 \text{ and } |\xi_{n-1}| = 1, \end{cases} \tag{14}$$

then

$$\sum_{n=0}^{\infty} \frac{\rho(\xi_n)}{2^{n+1}} = \begin{cases} \frac{\tan^{-1} x}{\pi} & \text{for } x \geq 0 \\ 1 + \frac{\tan^{-1} x}{\pi} & \text{for } x < 0. \end{cases} \tag{15}$$

Borwein and Girgensohn (1995) also give much more general recurrences and formulas.

References

Borwein, J. M. and Girgensohn, R. "Addition Theorems and Binary Expansions." *Canad. J. Math.* **47**, 262–273, 1995.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/plff/plff.html`.

Plouffe, S.. "The Computation of Certain Numbers Using a Ruler and Compass." Dec. 12, 1997. `http://www.research.att.com/~njas/sequences/JIS/compass.html`.

Plücker Characteristics

The CLASS m, ORDER n, number of NODES δ, number of CUSPS κ, number of STATIONARY TANGENTS (INFLECTION POINTS) ι, number of BITANGENTS τ, and GENUS p.

see also ALGEBRAIC CURVE, BITANGENT, CUSP, GENUS (SURFACE), INFLECTION POINT, NODE (ALGEBRAIC CURVE), STATIONARY TANGENT

Plücker's Conoid

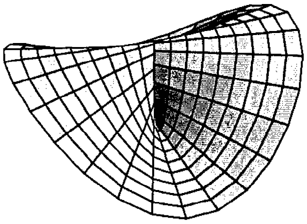

A RULED SURFACE sometimes also called the CYLINDROID. von Seggern (1993) gives the general functional form as

$$ax^2 + by^2 - zx^2 - zy^2 = 0, \tag{1}$$

whereas Fischer (1986) and Gray (1993) give

$$z = \frac{2xy}{(x^2 + y^2)}. \tag{2}$$

A polar parameterization therefore gives

$$x(r,\theta) = r\cos\theta \tag{3}$$

$$y(r,\theta) = r\sin\theta \tag{4}$$

$$z(r,\theta) = 2\cos\theta\sin\theta. \tag{5}$$

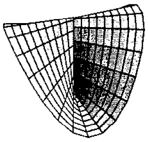

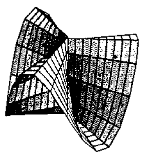

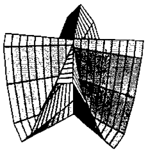

A generalization of Plücker's conoid to n folds is given by

$$x(r,\theta) = r\cos\theta \qquad (6)$$

$$y(r,\theta) = r\sin\theta \qquad (7)$$

$$z(r,\theta) = \sin(n\theta) \qquad (8)$$

(Gray 1993). The cylindroid is the inversion of the CROSS-CAP (Pinkall 1986).

see also CROSS-CAP, RIGHT CONOID, RULED SURFACE

References

Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, pp. 4–5, 1986.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 337–339, 1993.

Pinkall, U. *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, p. 64, 1986.

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 288, 1993.

Plücker's Equations

Relationships between the number of SINGULARITIES of plane algebraic curves. Given a PLANE CURVE,

$$m = n(n-1) - 2\delta - 3\kappa \qquad (1)$$

$$n = m(m-1) - 2\tau - 3\iota \qquad (2)$$

$$\iota = 3n(n-2) - 6\delta - 8\kappa \qquad (3)$$

$$\kappa = 3m(m-2) - 6\tau - 8\iota, \qquad (4)$$

where m is the CLASS, n the ORDER, δ the number of NODES, κ the number of CUSPS, ι the number of STATIONARY TANGENTS (INFLECTION POINTS), and τ the number of BITANGENTS. Only three of these equations are LINEARLY INDEPENDENT.

see also ALGEBRAIC CURVE, BIOCHE'S THEOREM, BITANGENT, CUSP, GENUS (SURFACE), INFLECTION POINT, KLEIN'S EQUATION, NODE (ALGEBRAIC CURVE), STATIONARY TANGENT

References

Boyer, C. B. *A History of Mathematics.* New York: Wiley, pp. 581–582, 1968.

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, pp. 99–118, 1959.

Plücker Relations

see PLÜCKER'S EQUATIONS

Plumbing

The plumbing of a p-sphere and a q-sphere is defined as the disjoint union of $\mathbb{S}^p \times \mathbb{S}^q$ and $\mathbb{D}^p \times \mathbb{S}^q$ with their common $\mathbb{D}^p \times \mathbb{D}^q$, identified via the identity homeomorphism.

see also HYPERSPHERE

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 180, 1976.

Pluperfect Number

see MULTIPLY PERFECT NUMBER

Plurisubharmonic Function

An upper semicontinuous function whose restrictions to all COMPLEX lines are subharmonic (where defined). These functions were introduced by P. Lelong and Oka in the early 1940s. Examples of such a function are the logarithms of moduli of holomorphic functions.

References

Range, R. M. and Anderson, R. W. "Hans-Joachim Bremermann, 1926–1996." *Not. Amer. Math. Soc.* **43**, 972–976, 1996.

Plus

The ADDITION of two quantities, i.e., a plus b. The operation is denoted $a+b$, and the symbol $+$ is called the PLUS SIGN. Floating point ADDITION is sometimes denoted $\oplus$.

see also ADDITION, MINUS, PLUS OR MINUS, TIMES

Plus or Minus

The symbol $\pm$ is used to denote a quantity which should be both added and subtracted, as in $a \pm b$. The symbol can be used to denote a range of uncertainty, or to denote a pair of quantities, such as the roots given by the QUADRATIC FORMULA

$$x_\pm = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

When order is relevant, the symbol $a \mp b$ is also used, so an expression of the form $x \pm y \mp z$ is interpreted as $x+y-z$ or $x-y+z$. In contrast, the expression $x \pm y \pm z$ is interpreted to mean the set of four quantities $x+y+z$, $x-y+z$, $x+y-z$, and $x-y-z$.

see also MINUS, MINUS SIGN, PLUS, PLUS SIGN, SIGN

Plus Perfect Number

see ARMSTRONG NUMBER

Plus Sign

The symbol "+" which is used to denote a POSITIVE number or to indicate ADDITION.

see also ADDITION, MINUS SIGN, SIGN

Plutarch Numbers

In *Moralia,* the Greek biographer and philosopher Plutarch states "Chrysippus says that the number of compound propositions that can be made from only ten simple propositions exceeds a million. (Hipparchus, to be sure, refuted this by showing that on the affirmative side there are 103,049 compound statements, and on the negative side 310,952.)" These numbers are known as the Plutarch numbers. 103,049 can be interpreted as the number s_{10} of BRACKETINGS on ten letters (Stanley 1997), Habsieger *et al.* 1998). Similarly, Plutarch's second number is given by $(s_{10}+s_{11})/2 = 310,954$ (Habsieger *et al.* 1998).

References

Biermann, K.-R. and Mau, J. "Überprüfung einer frühen Anwendung der Kombinatorik in der Logik." *J. Symbolic Logic* **23**, 129–132, 1958.

Biggs, N. L. "The Roots of Combinatorics." *Historia Mathematica* **6**, 109–136, 1979.

Habsieger, L.; Kazarian, M.; and Lando, S. "On the Second Number of Plutarch." *Amer. Math. Monthly* **105**, 446, 1998.

Heath, T. L. *A History of Greek Mathematics, Vol. 2: From Aristarchus to Diophantus.* New York: Dover, p. 256, 1981.

Kneale, W. and Kneale, M. *The Development of Logic.* Oxford, England: Oxford University Press, p. 162, 1971.

Neugebauer, O. *A History of Ancient Mathematical Astronomy, Vol. 1.* New York: Springer-Verlag, p. 338, 1975.

Plutarch. §VIII.9 in *Moralia, Vol. 9.* Cambridge, MA: Harvard University Press, p. 732, 1961.

Stanley, R. P. *Enumerative Combinatorics, Vol. 1.* Cambridge, England: Cambridge University Press, p. 63, 1996.

Stanley, R. P. "Hipparchus, Plutarch, Schröder, and Hough." *Amer. Math. Monthly* **104**, 344–350, 1997.

Pochhammer Symbol

A.k.a. RISING FACTORIAL. For an INTEGER $n > 0$,

$$(a)_n \equiv \frac{\Gamma(a+k)}{\Gamma(a)} = a(a+1)\cdots(a+n-1), \qquad (1)$$

where $\Gamma(z)$ is the GAMMA FUNCTION and

$$(a)_0 \equiv 1. \qquad (2)$$

The NOTATION conflicts with both that for q-SERIES and that for GAUSSIAN COEFFICIENTS, so context usually serves to distinguish the three. Additional identities are

$$\frac{d}{da}(a)_n = (a)_n[F(a+n-1) - F(a-1)] \qquad (3)$$

$$(a)_{n+k} = (a+n)_k(a)_n, \qquad (4)$$

where F is the DIGAMMA FUNCTION. The Pochhammer symbol arises in series expansions of HYPERGEOMETRIC FUNCTIONS and GENERALIZED HYPERGEOMETRIC FUNCTIONS.

see also FACTORIAL, GENERALIZED HYPERGEOMETRIC FUNCTION, HARMONIC LOGARITHM, HYPERGEOMETRIC FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 256, 1972.

Spanier, J. and Oldham, K. B. "The Pochhammer Polynomials $(x)_n$." Ch. 18 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 149–165, 1987.

Pocklington's Criterion

Let p be an ODD PRIME, k be an INTEGER such that $p\nmid k$ and $1 \leq k \leq 2(p+1)$, and

$$N \equiv 2kp + 1.$$

Then the following are equivalent

1. N is PRIME.
2. $\mathrm{GCD}(a^k + 1, N) = 1$.

This is a modified version of the original theorem due to Lehmer.

References

Pocklington, H. C. "The Determination of the Prime or Composite Nature of Large Numbers by Fermat's Theorem." *Proc. Cambridge Phil. Soc.* **18**, 29–30, 1914/16.

Pocklington-Lehmer Test

see POCKLINGTON'S THEOREM

Pocklington's Theorem

Let $n-1 = FR$ where F is the factored part of a number

$$F = {p_1}^{a_1}\cdots{p_r}^{a_r}, \qquad (1)$$

where $(R,F) = 1$, and $R < \sqrt{n}$. If there exists a b_i for $i = 1, \ldots, r$ such that

$${b_i}^{n-1} \equiv 1 \pmod{n} \qquad (2)$$

$$\mathrm{GCD}({b_i}^{(n-1)/p_i} - 1, n) = 1, \qquad (3)$$

then n is a PRIME.

Poggendorff Illusion

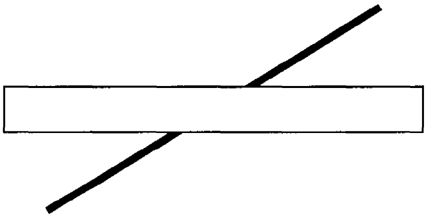

The illusion that the two ends of a straight LINE SEGMENT passing behind an obscuring RECTANGLE are offset when, in fact, they are aligned.

see also ILLUSION, MÜLLER-LYER ILLUSION, PONZO'S ILLUSION, VERTICAL-HORIZONTAL ILLUSION

References
Burmester, E. "Beiträge zu experimentellen Bestimmung geometrisch-optischer Täuschungen." *Z. Psychologie* **12**, 355–394, 1896.
Day, R. H. and Dickenson, R. G. "The Components of the Poggendorff Illusion." *Brit. J. Psychology* **67**, 537–552, 1976.
Fineman, M. "Poggendorff's Illusion." Ch. 19 in *The Nature of Visual Illusion.* New York: Dover, pp. 151–159, 1996.

Pohlke's Theorem

The principal theorem of AXONOMETRY. It states that three segments of arbitrary length $a'x'$, $a'y'$, and $a'z'$ which are drawn in a PLANE from a point a' under arbitrary ANGLES form a parallel projection of three equal segments ax, ay, and az from the ORIGIN of three PERPENDICULAR coordinate axes. However, only one of the segments or one of the ANGLES may vanish.

see also AXONOMETRY

Poincaré-Birkhoff Fixed Point Theorem

For the rational curve of an unperturbed system with ROTATION NUMBER r/s under a map T (for which every point is a FIXED POINT of T^s), only an even number of FIXED POINTS $2ks$ ($k = 1, 2, \ldots$) will remain under perturbation. These FIXED POINTS are alternately stable (ELLIPTIC) and unstable (HYPERBOLIC). Around each elliptic fixed point there is a simultaneous application of the Poincaré-Birkhoff fixed point theorem and the KAM THEOREM, which leads to a self-similar structure on all scales.

The original formulation was: Given a CONFORMAL ONE-TO-ONE transformation from an ANNULUS to itself that advances points on the outer edge positively and on the inner edge negatively, then there are at least two fixed points.

It was conjectured by Poincaré from a consideration of the three-body problem in celestial mechanics and proved by Birkhoff.

Poincaré Conjecture

A SIMPLY CONNECTED 3-MANIFOLD is HOMEOMORPHIC to the 3-SPHERE. The generalized Poincaré conjecture is that a COMPACT n-MANIFOLD is HOMOTOPY equivalent to the n-sphere IFF it is HOMEOMORPHIC to the n-SPHERE. This reduces to the original conjecture for $n = 3$.

The $n = 1$ case of the generalized conjecture is trivial, the $n = 2$ case is classical, $n = 3$ remains open, $n = 4$ was proved by Freedman (1982) (for which he was awarded the 1986 FIELDS MEDAL), $n = 5$ by Zeeman (1961), $n = 6$ by Stallings (1962), and $n \geq 7$ by Smale in 1961 (Smale subsequently extended this proof to include $n \geq 5$.)

see also COMPACT MANIFOLD, HOMEOMORPHIC, HOMOTOPY, MANIFOLD, SIMPLY CONNECTED, SPHERE, THURSTON'S GEOMETRIZATION CONJECTURE

References
Freedman, M. H. "The Topology of Four-Differentiable Manifolds." *J. Diff. Geom.* **17**, 357–453, 1982.
Stallings, J. "The Piecewise-Linear Structure of Euclidean Space." *Proc. Cambridge Philos. Soc.* **58**, 481–488, 1962.
Smale, S. "Generalized Poincaré's Conjecture in Dimensions Greater than Four." *Ann. Math.* **74**, 391–406, 1961.
Zeeman, E. C. "The Generalised Poincaré Conjecture." *Bull. Amer. Math. Soc.* **67**, 270, 1961.
Zeeman, E. C. "The Poincaré Conjecture for $n \geq 5$." In *Topology of 3-Manifolds and Related Topics, Proceedings of the University of Georgia Institute, 1961.* Englewood Cliffs, NJ: Prentice-Hall, pp. 198–204, 1961.

Poincaré Duality

The BETTI NUMBERS of a compact orientable n-MANIFOLD satisfy the relation

$$b_i = b_{n-i}.$$

see also BETTI NUMBER

Poincaré Formula

The POLYHEDRAL FORMULA generalized to a surface of GENUS p.

$$V - E + F = 2 - 2p$$

where V is the number of VERTICES, E is the number of EDGES, F is the number of faces, and

$$\chi \equiv 2 - 2p$$

is called the EULER CHARACTERISTIC.

see also EULER CHARACTERISTIC, GENUS (SURFACE), POLYHEDRAL FORMULA

References
Eppstein, D. "Fourteen Proofs of Euler's Formula: $V - E + F = 2$." `http://www.ics.uci.edu/~eppstein/junkyard/euler`.

Poincaré-Fuchs-Klein Automorphic Function

$$f(z) = \frac{k}{(cz+d)^r} f\left(\frac{az+b}{cz+d}\right)$$

where $\Im(z) > 0$.

see also AUTOMORPHIC FUNCTION

Poincaré Group

see LORENTZ GROUP

Poincaré's Holomorphic Lemma

Solutions to HOLOMORPHIC differential equations are themselves HOLOMORPHIC FUNCTIONS of time, initial conditions, and parameters.

Poincaré-Hopf Index Theorem

The index of a VECTOR FIELD with finitely many zeros on a compact, oriented MANIFOLD is the same as the EULER CHARACTERISTIC of the MANIFOLD.

see also GAUSS-BONNET FORMULA

Poincaré Hyperbolic Disk

A 2-D space having HYPERBOLIC GEOMETRY defined as the 2-BALL $\{x \in \mathbb{R}^2 : |x| < 1\}$, with HYPERBOLIC METRIC

$$\frac{dx^2 + dy^2}{(1 - r^2)^2}.$$

The Poincaré disk is a model for HYPERBOLIC GEOMETRY, and there is an isomorphism between the Poincaré disk model and the KLEIN-BELTRAMI MODEL.

see also ELLIPTIC PLANE, HYPERBOLIC GEOMETRY, HYPERBOLIC METRIC, KLEIN-BELTRAMI MODEL

Poincaré's Lemma

Let $\wedge$ denote the WEDGE PRODUCT and D the EXTERIOR DERIVATIVE. Then

$$D^2 t = \frac{\partial}{\partial x} \wedge \left(\frac{\partial}{\partial x} \wedge t\right) = \left(\frac{\partial}{\partial x} \wedge \frac{\partial}{\partial x}\right) \wedge t = 0.$$

see also DIFFERENTIAL FORM, EXTERIOR DERIVATIVE, POINCARÉ'S HOLOMORPHIC LEMMA, WEDGE PRODUCT

Poincaré Manifold

A nonsimply connected 3-manifold also called a DODECAHEDRAL SPACE.

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 245, 290, and 308, 1976.

Poincaré Metric

The METRIC

$$ds^2 = \frac{dx^2 + dy^2}{(1 - |z|^2)^2}$$

of the POINCARÉ HYPERBOLIC DISK.

see also POINCARÉ HYPERBOLIC DISK

Poincaré Separation Theorem

Let $\{\mathbf{y}^k\}$ be a set of orthonormal vectors with $k = 1, 2, \ldots, K$, such that the INNER PRODUCT $(\mathbf{y}^k, \mathbf{y}^k) = 1$. Then set

$$\mathbf{x} = \sum_{k=1}^{K} u_k \mathbf{y}^k \tag{1}$$

so that for any SQUARE MATRIX A for which the product $\mathsf{A}\mathbf{x}$ is defined, the corresponding QUADRATIC FORM is

$$(\mathbf{x}, \mathsf{A}\mathbf{x}) = \sum_{k,l=1}^{K} u_k u_l (\mathbf{y}^k, \mathsf{A}\mathbf{y}^l). \tag{2}$$

Then if

$$\mathbf{B}_k = (\mathbf{y}^k, \mathsf{A}\mathbf{y}^l) \tag{3}$$

for $k, l = 1, 2, \ldots, K$, it follows that

$$\lambda_i(\mathbf{B}_K) \leq \lambda_1(\mathsf{A}) \tag{4}$$

$$\lambda_{K-j}(\mathbf{B}_K) \geq \lambda_{N-j}(\mathsf{A}) \tag{5}$$

for $i = 1, 2, \ldots, K$ and $j = 0, 1, \ldots, K - 1$.

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1120, 1979.

Poinsot Solid

see KEPLER-POINSOT SOLID

Poinsot's Spirals

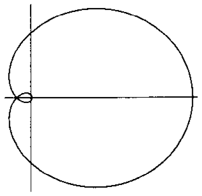

$r \sinh(n\theta) = a.$

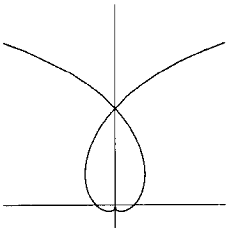

$r \operatorname{csch}(n\theta) = a.$

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 192 and 194, 1972.

Point

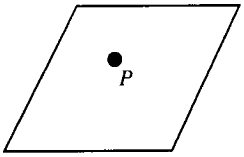

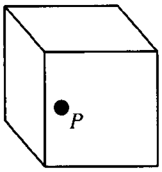

A 0-DIMENSIONAL mathematical object which can be specified in n-D space using n coordinates. Although the notion of a point is intuitively rather clear, the mathematical machinery used to deal with points and point-like objects can be surprisingly slippery. This difficulty was encountered by none other than Euclid himself who, in his *Elements*, gave the vague definition of a point as "that which has no part."

The basic geometric structures of higher DIMENSIONAL geometry—the LINE, PLANE, SPACE, and HYPERSPACE—are all built up of infinite numbers of points arranged in particular ways.

see also ACCUMULATION POINT, ANTIGONAL POINTS, ANTIHOMOLOGOUS POINTS, APOLLONIUS POINT, BOUNDARY POINT, BRANCH POINT, BRIANCHON POINT, BROCARD MIDPOINT, BROCARD POINTS, Cantor-Dedekind Axiom, Center, Circle Lattice Points, Concur, Concurrent, Congruent Incircles Point, Congruent Isoscelizers Point, Conjugate Points, Critical Point, Crucial Point, Cube Point Picking, Cusp Point, de Longchamps Point, Double Point, Eckardt Point, Elkies Point, Elliptic Fixed Point (Differential Equations), Elliptic Fixed Point (Map), Elliptic Point, Equal Detour Point, Equal Parallelians Point, Equichordal Point, Equilibrium Point, Equiproduct Point, Equireciprocal Point, Evans Point, Exeter Point, Exmedian Point, Fagnano's Point, Far-Out Point, Fejes Tóth's Problem, Fermat Point, Feuerbach Point, Feynman Point, Fixed Point, Fletcher Point, Gergonne Point, Grebe Point, Griffiths Points, Harmonic Conjugate Points, Hermit Point, Hofstadter Point, Homologous Points, Hyperbolic Fixed Point (Differential Equations), Hyperbolic Fixed Point (Map), Hyperbolic Point, Ideal Point, Imaginary Point, Invariant Point, Inverse Points, Isodynamic Points, Isolated Point, Isoperimetric Point, Isotomic Conjugate Point, Lattice Point, Lemoine Point, Limit Point, Malfatti Points, Median Point, Mid-Arc Points, Midpoint, Miquel Point, Nagel Point, Napoleon Points, Nobbs Points, Oldknow Points, Only Critical Point in Town Test, Ordinary Point, Parabolic Point, Parry Point, Pedal Point, Periodic Point, Planar Point, Point at Infinity, Point-Line Distance—2-D, Point-Line Distance—3-D, Point-Quadratic Distance, Point-Plane Distance, Point-Set Topology, Pointwise Dimension, Policeman on Point Duty Curve, Power Point, Radial Point, Radiant Point, Rational Point, Rigby Points, Saddle Point (Game), Saddle Point (Function), Salient Point, Schiffler Point, Self-Homologous Point, Similarity Point, Singular Point (Algebraic Curve), Singular Point (Function), Soddy Points, Special Point, Stationary Point, Steiner Points, Sylvester's Four-Point Problem, Symmedian Point, Symmetric Points, Tarry Point, Torricelli Point, Trisected Perimeter Point, Umbilic Point, Unit Point, Vanishing Point, Visible Point, Weierstraß Point, Wild Point, Yff Points

References

Casey, J. "The Point." Ch. 1 in *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions, with Numerous Examples, 2nd ed., rev. enl.* Dublin: Hodges, Figgis, & Co., pp. 1–29, 1893.

Point Estimator

An Estimator of the actual values of population.

Point Groups

The symmetry groups possible in a crystal lattice without the translation symmetry element. Although an isolated object may have an arbitrary Schönflies Symbol, the requirement that symmetry be present in a lattice requires that only 1, 2, 3, and 6-fold symmetry axes are possible (the Crystallography Restriction), which restricts the number of possible point groups to 32: C_i, C_s, C_1, C_2, C_3, C_4, C_6, C_{2h}, C_{3h}, C_{4h}, C_{6h}, C_{2v}, C_{3v}, C_{4v}, C_{6v}, D_2, D_3, D_4, D_6 (the Dihedral Groups), D_{2h}, D_{3h}, D_{4h}, D_{6h}, D_{2d}, D_{3d}, O, O_h (the Octahedral Group), S_4, S_6, T, T_h, and T_d (the Tetrahedral Group).

see also Crystallography Restriction, Dihedral Group, Group, Group Theory, Hermann-Mauguin Symbol, Lattice Groups, Octahedral Group, Schönflies Symbol, Space Groups, Tetrahedral Group

References

Arfken, G. "Crystallographic Point and Space Groups." *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 248–249, 1985.

Cotton, F. A. *Chemical Applications of Group Theory, 3rd ed.* New York: Wiley, p. 379, 1990.

Lomont, J. S. "Crystallographic Point Groups." §4.4 in *Applications of Finite Groups.* New York: Dover, pp. 132–146, 1993.

Point at Infinity

P is the point on the line AB such that $\overline{PA}/\overline{PB} = 1$. It can also be thought of as the point of intersection of two Parallel lines.

see also Line at Infinity

References

Behnke, H.; Bachmann, F.; Fladt, K.; and Suss, W. (Eds.). Ch. 7 in *Fundamentals of Mathematics, Vol. 3: Points at Infinity.* Cambridge, MA: MIT Press, 1974.

Point-Line Distance—2-D

Given a line $ax + by + c = 0$ and a point (x_0, y_0), in slope-intercept form, the equation of the line is

$$y = -\frac{a}{b}x - \frac{c}{b}, \tag{1}$$

so the line has Slope $-a/b$. Points on the line have the vector coordinates

$$\begin{bmatrix} x \\ -\frac{a}{b}x - \frac{c}{d} \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{c}{d} \end{bmatrix} - \frac{1}{b}\begin{bmatrix} -b \\ a \end{bmatrix}. \tag{2}$$

Therefore, the Vector

$$\begin{bmatrix} -b \\ a \end{bmatrix} \tag{3}$$

is Parallel to the line, and the Vector

$$\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix} \tag{4}$$

is PERPENDICULAR to it. Now, a VECTOR from the point to the line is given by

$$\mathbf{r} = \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}. \tag{5}$$

Projecting $\mathbf{r}$ onto $\mathbf{v}$,

$$\begin{aligned} |\operatorname{proj}_{\mathbf{v}}\mathbf{r}| &= \frac{|\mathbf{v}\cdot\mathbf{r}|}{|\mathbf{v}|} = |\hat{\mathbf{v}}\cdot\mathbf{r}| = \frac{|a(x-x_0)+b(y-y_0)|}{\sqrt{a^2+b^2}} \\ &= \frac{|ax+by-ax_0-by_0|}{\sqrt{a^2+b^2}} \\ &= \frac{|ax_0+by_0+c|}{\sqrt{a^2+b^2}}. \end{aligned} \tag{6}$$

If the line is represented by the endpoints of a VECTOR (x_1, y_1) and (x_2, y_2), then the PERPENDICULAR VECTOR is

$$\mathbf{v} = \begin{bmatrix} y_2 - y_1 \\ -(x_2 - x_1) \end{bmatrix} \tag{7}$$

$$\hat{\mathbf{v}} = \frac{1}{s}\begin{bmatrix} y_2 - y_1 \\ -(x_2 - x_1) \end{bmatrix}, \tag{8}$$

where

$$s = |\mathbf{v}| = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2}, \tag{9}$$

so the distance is

$$d = |\hat{\mathbf{v}}\cdot\mathbf{r}| = \frac{|(y_2-y_1)(x_0-x_1)-(x_2-x_1)(y_0-y_1)|}{s}. \tag{10}$$

The distance from a point (x_1,y_1) to the line $y = a + bx$ can be computed using VECTOR algebra. Let $\mathbf{L}$ be a VECTOR in the same direction as the line

$$\mathbf{L} = \begin{bmatrix} x \\ a + bx \end{bmatrix} - \begin{bmatrix} 0 \\ a \end{bmatrix} = \begin{bmatrix} x \\ bx \end{bmatrix} \tag{11}$$

$$\hat{\mathbf{L}} = \frac{1}{\sqrt{b^2+1}}\begin{bmatrix} 1 \\ b \end{bmatrix}. \tag{12}$$

A given point on the line is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} - \begin{bmatrix} 0 \\ -a \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 - a \end{bmatrix}, \tag{13}$$

so the point-line distance is

$$\begin{aligned} \mathbf{r} &= (\mathbf{x}\cdot\hat{\mathbf{L}})\hat{\mathbf{L}} - \mathbf{x} \\ &= \frac{1}{1+b^2}\left(\begin{bmatrix} x_1 \\ y_1 - a \end{bmatrix}\cdot\begin{bmatrix} 1 \\ v \end{bmatrix}\right)\begin{bmatrix} 1 \\ b \end{bmatrix} - \begin{bmatrix} x_1 \\ y_1 - a \end{bmatrix} \\ &= \frac{x_1 + b(y_1 - a)}{1+b^2}\begin{bmatrix} 1 \\ b \end{bmatrix} - \begin{bmatrix} x_1 \\ y_1 - a \end{bmatrix} \\ &= \frac{1}{1+b^2}\begin{bmatrix} b(y_1-a) - b^2x_1 \\ bx_1 + b^2y_1 - ab^2 - y_1 + a - b^2y_1 + ab^2 \end{bmatrix} \\ &= \frac{1}{1+b^2}\begin{bmatrix} b[(y_1-a) - bx_1] \\ -[(y_1-a) - bx_1] \end{bmatrix} \\ &= \frac{y_1 - (a + bx_1)}{1+b^2}\begin{bmatrix} b \\ -1 \end{bmatrix}. \end{aligned} \tag{14}$$

Therefore,

$$d = |\mathbf{r}| = \frac{|y_1 - (a+bx_1)|}{1+b^2}\sqrt{1+b^2} = \frac{|y_1 - (a+bx_1)|}{\sqrt{1+b^2}}. \tag{15}$$

This result can also be obtained much more simply by noting that the PERPENDICULAR distance is just $\cos\theta$ times the vertical distance $|y_1 - (a + bx_1)|$. But the SLOPE b is just $\tan\theta$, so

$$\sin^2\theta + \cos^2\theta = 1 \Rightarrow \tan^2\theta + 1 = \frac{1}{\cos^2\theta}, \tag{16}$$

and

$$\cos\theta = \frac{1}{\sqrt{1+\tan^2\theta}} = \frac{1}{\sqrt{1+b^2}}. \tag{17}$$

The PERPENDICULAR distance is then

$$d = \frac{|y_1 - (a+bx_1)|}{\sqrt{1+b^2}}, \tag{18}$$

the same result as before.

see also LINE, POINT, POINT-LINE DISTANCE—3-D

Point-Line Distance—3-D

A line in 3-D is given by the parametric VECTOR

$$\mathbf{v} = \begin{bmatrix} x_0 + at \\ y_0 + bt \\ z_0 + ct \end{bmatrix}. \tag{1}$$

The distance between a point on the line with parameter t and the point (x_1, y_1, z_1) is therefore

$$r^2 = (x_1 - x_0 - at)^2 + (y_1 - y_0 - bt)^2 + (z_1 - z_0 - ct)^2. \tag{2}$$

To minimize the distance, take

$$\begin{aligned} \frac{\partial(r^2)}{\partial t} &= -2a(x_1 - x_0 - at) - 2b(y_1 - y_0 - bt) \\ &\quad -2c(z_1 - z_0 - ct) = 0 \end{aligned} \tag{3}$$

$$a(x_1-x_0) + b(y_1-y_0) + c(z_1-z_0) - t(a^2+b^2+c^2) = 0 \tag{4}$$

$$t = \frac{a(x_1-x_0) + b(y_1-y_0) + c(z_1-z_0)}{a^2+b^2+c^2}, \tag{5}$$

so the minimum distance is found by plugging (5) into (2) and taking the SQUARE ROOT.

see also LINE, POINT, POINT-LINE DISTANCE—2-D

Point Picking

see 18-POINT PROBLEM, BALL TRIANGLE PICKING, CUBE POINT PICKING, CUBE TRIANGLE PICKING, DISCREPANCY THEOREM, ISOSCELES TRIANGLE, OBTUSE TRIANGLE, PLANAR DISTANCE, SYLVESTER'S FOUR-POINT PROBLEM

Point-Plane Distance

Given a PLANE

$$ax + by + cz + d = 0 \tag{1}$$

and a point (x_0, y_0, z_0), the NORMAL to the PLANE is given by

$$\mathbf{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \tag{2}$$

and a VECTOR from the plane to the point is given by

$$\mathbf{w} = \begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix}. \tag{3}$$

Projecting $\mathbf{w}$ onto $\mathbf{v}$,

$$\begin{aligned} |\mathrm{proj}_{\mathbf{v}}\mathbf{w}| &= \frac{|\mathbf{v}\cdot\mathbf{w}|}{|\mathbf{v}|} \\ &= \frac{|a(x-x_0) + b(y-y_0) + c(z-z_0) + d|}{\sqrt{a^2+b^2+c^2}} \\ &= \frac{|ax + by + cz - ax_0 - by_0 - cz_0|}{\sqrt{a^2+b^2+c^2}} \\ &= \frac{|ax_0 + by_0 + cz_0 + d|}{\sqrt{a^2+b^2+c^2}}. \end{aligned} \tag{4}$$

Point-Point Distance—1-D

Given a unit LINE SEGMENT $[0,1]$, pick two points at random on it. Call the first point x_1 and the second point x_2. Find the distribution of distances d between points. The probability of the points being a (POSITIVE) distance d apart (i.e., without regard to ordering) is given by

$$\begin{aligned} P(d) &= \frac{\int_0^1\int_0^1 \delta(d - |x_2 - x_1|)\,dx_1\,dx_2}{\int_0^1\int_0^1 dx_1\,dx_2} \\ &= (1-d)[H(1-d) - H(d-1) + H(d) - H(-d)] \\ &= \begin{cases} 2(1-d) & \text{for } 0 \leq d \leq 1 \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{1}$$

where δ is the DIRAC DELTA FUNCTION and H is the HEAVISIDE STEP FUNCTION. The MOMENTS are then

$$\begin{aligned} \mu'_m &= \int_0^1 d^m P(d)\,dd = 2\int_0^1 d^m(1-d)\,dd \\ &= 2\left[\frac{d^{m+1}}{m+1} - \frac{d^{m+2}}{m+2}\right]_0^1 \\ &= 2\left(\frac{1}{m+1} - \frac{1}{m+2}\right) = 2\left[\frac{(m+2)-(m+1)}{(m+1)(m+2)}\right] \\ &= \frac{2}{(m+1)(m+2)} \\ &= \begin{cases} \frac{1}{(n+1)(2n+1)} & \text{for } m = 2n \\ \frac{1}{(n+1)(2n+3)} & \text{for } m = 2n+1, \end{cases} \end{aligned} \tag{2}$$

giving MOMENTS about 0

$$\mu'_1 = \tfrac{1}{3} \tag{3}$$

$$\mu'_2 = \tfrac{1}{6} \tag{4}$$

$$\mu'_3 = \tfrac{1}{10} \tag{5}$$

$$\mu'_4 = \tfrac{1}{15}. \tag{6}$$

The MOMENTS can also be computed directly without explicit knowledge of the distribution

$$\begin{aligned} \mu'_1 &= \frac{\int_0^1\int_0^1 |x_2 - x_1|\,dx_1\,dx_2}{\int_0^1\int_0^1 dx_1\,dx_2} \\ &= \int_0^1\int_0^1 |x_2 - x_1|\,dx_1\,dx_2 \\ &= \underset{x_2 - x_1 > 0}{\int_0^1\int_0^1} (x_2 - x_1)\,dx_1\,dx_2 \\ &\quad + \underset{x_2 - x_1 < 0}{\int_0^1\int_0^1} (x_1 - x_2)\,dx_1\,dx_2 \\ &= \int_0^1\int_{x_1}^1 (x_2 - x_1)\,dx_1\,dx_2 \\ &\quad + \int_0^1\int_0^{x_1} (x_2 - x_1)\,dx_1\,dx_2 \\ &= \int_0^1 \left[\tfrac{1}{2}{x_2}^2 - x_1x_2\right]_{x_1}^1 dx_1 \\ &\quad + \int_0^1 \left[x_1x_2 - \tfrac{1}{2}{x_2}^2\right]_0^{x_1} dx_1 \\ &= \int_0^1 \left[(\tfrac{1}{2} - x_1) - (\tfrac{1}{2}{x_1}^2 - {x_1}^2)\right] dx_1 \\ &\quad + \int_0^1 \left[({x_1}^2 - \tfrac{1}{2}{x_1}^2) - (0 - 0)\right] dx_1 \\ &= \int_0^1 (\tfrac{1}{2} - x_1 + {x_1}^2)\,dx_1 = [\tfrac{1}{2}x_1 - \tfrac{1}{2}{x_1}^2 + \tfrac{1}{3}{x_1}^3]_0^1 \\ &= (\tfrac{1}{2} - \tfrac{1}{2} + \tfrac{1}{3}) - (0 - 0 + 0) = \tfrac{1}{3} \end{aligned} \tag{7}$$

$$\begin{aligned} \mu'_2 &= \int_0^1\int_0^1 (|x_2 - x_1|)^2\,dx_2\,dx_1 \\ &= \int_0^1\int_0^1 (x_2 - x_1)^2\,dx_1\,dx_2 \\ &= \int_0^1\int_0^1 ({x_2}^2 - 2x_1x_2 + {x_1}^2)\,dx_1\,dx_2 \\ &= \int_0^1 [\tfrac{1}{3}{x_2}^3 - x_1{x_2}^2 + {x_1}^2x_2]_0^1\,dx_1 \\ &= \int_0^1 (\tfrac{1}{3} - x_1 + {x_1}^2)\,dx_1 = [\tfrac{1}{3}{x_1}^3 - \tfrac{1}{2}{x_1}^2 + \tfrac{1}{3}x_1]_0^1 \\ &= \tfrac{1}{3} - \tfrac{1}{2} + \tfrac{1}{3} = \tfrac{1}{6}. \end{aligned} \tag{8}$$

The MOMENTS about the MEAN are therefore

$$\mu_2 = \mu_2' - {\mu_1'}^2 = \tfrac{1}{6} - (\tfrac{1}{3})^2 = \tfrac{1}{18} \tag{9}$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3 = \tfrac{1}{135} \tag{10}$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4 = \tfrac{1}{135}, \tag{11}$$

so the MEAN, VARIANCE, SKEWNESS, and KURTOSIS are

$$\mu = \mu_1' = \tfrac{1}{3} \tag{12}$$

$$\sigma^2 = \mu_2 = \tfrac{1}{18} \tag{13}$$

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \tfrac{2}{5}\sqrt{2} \tag{14}$$

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = -\tfrac{3}{5}. \tag{15}$$

The probability distribution of the distance between two points randomly picked on a LINE SEGMENT is germane to the problem of determining the access time of computer hard drives. In fact, the average access time for a hard drive is precisely the time required to seek across 1/3 of the tracks (Benedict 1995).

see also POINT-POINT DISTANCE—2-D, POINT-POINT DISTANCE—3-D, POINT-QUADRATIC DISTANCE, TETRAHEDRON INSCRIBING, TRIANGLE INSCRIBING IN A CIRCLE

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 930–931, 1985.

Benedict, B. *Using Norton Utilities for the Macintosh.* Indianapolis, IN: Que, pp. B-8–B-9, 1995.

Point-Point Distance—2-D

Given two points in the PLANE, find the curve which minimizes the distance between them. The LINE ELEMENT is given by

$$ds = \sqrt{dx^2 + dy^2}, \tag{1}$$

so the ARC LENGTH between the points x_1 and x_2 is

$$L = \int ds = \int_{x_1}^{x_2} \sqrt{1 + y'^2}\,dx, \tag{2}$$

where $y' \equiv dy/dx$ and the quantity we are minimizing is

$$f = \sqrt{1 + y'^2}. \tag{3}$$

Finding the derivatives gives

$$\frac{\partial f}{\partial y} = 0 \tag{4}$$

$$\frac{d}{dx}\frac{\partial f}{\partial y'} = \frac{d}{dx}[(1 + y'^2)^{-1/2}y'], \tag{5}$$

so the EULER-LAGRANGE DIFFERENTIAL EQUATION becomes

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'} = \frac{d}{dx}\left(\frac{y'}{\sqrt{1 + y'^2}}\right) = 0. \tag{6}$$

Integrating and rearranging,

$$\frac{y'}{\sqrt{1 + y'^2}} = c \tag{7}$$

$$y'^2 = c^2(1 + y'^2) \tag{8}$$

$$y'^2(1 - c^2) = c^2 \tag{9}$$

$$y' = \frac{c}{\sqrt{1 - c^2}} \equiv a. \tag{10}$$

The solution is therefore

$$y = ax + b, \tag{11}$$

which is a straight LINE. Now verify that the ARC LENGTH is indeed the straight-line distance between the points. a and b are determined from

$$y_1 = ax_1 + b \tag{12}$$

$$y_2 = ax_2 + b. \tag{13}$$

Writing (12) and (13) as a MATRIX EQUATION gives

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \tag{14}$$

$$\begin{aligned} \begin{bmatrix} a \\ b \end{bmatrix} &= \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \\ &= \frac{1}{x_1 - x_2}\begin{bmatrix} 1 & -1 \\ -x_2 & x_1 \end{bmatrix}^{-1} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \end{aligned} \tag{15}$$

so

$$a = \frac{y_1 - y_2}{x_1 - x_2} = \frac{y_2 - y_1}{x_2 - x_1} \tag{16}$$

$$b = \frac{x_1y_2 - x_2y_1}{x_1 - x_2} \tag{17}$$

$$\begin{aligned} L &= \int_{x_1}^{x_2} \sqrt{1 + y'^2}\,dy = (x_2 - x_1)\sqrt{1 + a^2} \\ &= (x_2 - x_1)\sqrt{1 + \left(\frac{y_2 - y_1}{x_2 - x_1}\right)^2} \\ &= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \end{aligned} \tag{18}$$

as expected.

The shortest distance between two points on a SPHERE is the so-called GREAT CIRCLE distance.

see also CALCULUS OF VARIATIONS, GREAT CIRCLE, POINT-POINT DISTANCE—1-D, POINT-POINT DISTANCE—3-D, POINT-QUADRATIC DISTANCE, TETRAHEDRON INSCRIBING, TRIANGLE INSCRIBING IN A CIRCLE

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 930–931, 1985.

Point-Point Distance—3-D

The LINE ELEMENT is

$$ds = \sqrt{dx^2 + dy^2 + dz^2}, \tag{1}$$

so the ARC LENGTH between the points x_1 and x_2 is

$$L = \int ds = \int_{x_1}^{x_2} \sqrt{1 + y'^2 + z'^2}\, dx \tag{2}$$

and the quantity we are minimizing is

$$f = \sqrt{1 + y'^2 + z'^2}. \tag{3}$$

Finding the derivatives gives

$$\frac{\partial f}{\partial y} = 0 \tag{4}$$

$$\frac{\partial f}{\partial z} = 0 \tag{5}$$

and

$$\frac{\partial f}{\partial y'} = \frac{y'}{\sqrt{1 + y'^2 + z'^2}} \tag{6}$$

$$\frac{\partial f}{\partial z'} = \frac{z'}{\sqrt{1 + y'^2 + z'^2}}, \tag{7}$$

so the EULER-LAGRANGE DIFFERENTIAL EQUATIONS become

$$\frac{d}{dx}\left(\frac{y'}{\sqrt{1 + y'^2 + z'^2}}\right) = 0 \tag{8}$$

$$\frac{d}{dx}\left(\frac{z'}{\sqrt{1 + y'^2 + z'^2}}\right) = 0. \tag{9}$$

These give

$$\frac{y'}{\sqrt{1 + y'^2 + z'^2}} = c_1 \tag{10}$$

$$\frac{z'}{\sqrt{1 + y'^2 + z'^2}} = c_2. \tag{11}$$

Taking the ratio,

$$z' = \frac{c_2}{c_1} y' \tag{12}$$

$$\frac{y'}{\sqrt{1 + y'^2 + \left(\frac{c_2}{c_1}\right)^2 y'^2}} = c_1 \tag{13}$$

$$y'^2 = {c_1}^2 \left[1 + y'^2 + \left(\frac{c_2}{c_1}\right)^2 y'^2\right] = {c_1}^2 + y'^2({c_1}^2 + {c_2}^2), \tag{14}$$

which gives

$$y'^2 = \frac{{c_1}^2}{1 - {c_1}^2 - {c_2}^2} \equiv {a_1}^2 \tag{15}$$

$$z'^2 = \left(\frac{c_2}{c_1}\right)^2 y'^2 = \frac{{c_2}^2}{1 - {c_1}^2 - {c_2}^2} \equiv {b_1}^2. \tag{16}$$

Therefore, $y' = a_1$ and $z' = b_1$, so the solution is

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ a_1 x + a_0 \\ b_1 x + b_0 \end{bmatrix}, \tag{17}$$

which is the parametric representation of a straight line with parameter $x \in [x_1, x_2]$. Verifying the ARC LENGTH gives

$$L = \sqrt{1 + {a_1}^2 + {b_1}^2}\,(x_2 - x_1) \tag{18}$$

where

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_0 \end{bmatrix} \tag{19}$$

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_0 \end{bmatrix}. \tag{20}$$

see also POINT-POINT DISTANCE—1-D, POINT-POINT DISTANCE—2-D, POINT-QUADRATIC DISTANCE

Point Probability

The portion of the probability distribution which has a P-VALUE *equal to* the observed P-VALUE.

see also TAIL PROBABILITY

Point-Quadratic Distance

Find the minimum distance between a point in the plane (x_0, y_0) and a quadratic PLANE CURVE

$$y = a_0 + a_1 x + a_2 x^2. \tag{1}$$

The square of the distance is

$$\begin{aligned} r^2 &= (x - x_0)^2 + (y - y_0)^2 \\ &= (x - x_0)^2 + (a_0 + a_1 x + a_2 x^2 - y_0)^2. \end{aligned} \tag{2}$$

Minimizing the distance squared is the equivalent to minimizing the distance (since r^2 and $|r|$ have minima at the same point), so take

$$\frac{\partial(r^2)}{\partial x} = 2(x - x_0) + 2(a_0 + a_1 x + a_2 x^2 - y_0)(a_1 + 2a_2 x) = 0 \tag{3}$$

$$\begin{aligned} x - x_0 + a_0 a_1 + {a_1}^2 + a_1 a_2 x^2 - a_1 y_0 + 2a_0 a_2 x \\ + 2a_1 a_2 x^2 + 2{a_2}^2 x^3 - 2a_2 y_0 x = 0 \end{aligned} \tag{4}$$

$$\begin{aligned} 2{a_2}^2 x^3 + 3a_1 a_2 x^2 + ({a_1}^2 + 2a_0 a_2 - 2a_2 y_0 + 1)x \\ + (a_0 a_1 - a_1 y_0 - x_0) = 0. \end{aligned} \tag{5}$$

Minimizing the distance therefore requires solution of a CUBIC EQUATION.

see also POINT-POINT DISTANCE—1-D, POINT-POINT DISTANCE—2-D, POINT-POINT DISTANCE—3-D

Point-Set Topology

The low-level language of TOPOLOGY, which is not really considered a separate "branch" of TOPOLOGY. Point-set topology, also called set-theoretic topology or general topology, is the study of the general abstract nature of continuity or "closeness" on SPACES. Basic point-set topological notions are ones like CONTINUITY, DIMENSION, COMPACTNESS, and CONNECTEDNESS. The INTERMEDIATE VALUE THEOREM (which states that if a path in the real line connects two numbers, then it passes over every point between the two) is a basic topological result. Others are that EUCLIDEAN n-space is HOMEOMORPHIC to EUCLIDEAN m-space IFF $m = n$, and that REAL valued functions achieve maxima and minima on COMPACT SETS.

Foundational point-set topological questions are ones like "when can a topology on a space be derived from a metric?" Point-set topology deals with differing notions of continuity and compares them, as well as dealing with their properties. Point-set topology is also the ground-level of inquiry into the geometrical properties of spaces and continuous functions between them, and in that sense, it is the foundation on which the remainder of topology (ALGEBRAIC, DIFFERENTIAL, and LOW-DIMENSIONAL) stands.

see also ALGEBRAIC TOPOLOGY, DIFFERENTIAL TOPOLOGY, LOW-DIMENSIONAL TOPOLOGY, TOPOLOGY

References

Sutherland, W. A. *An Introduction to Metric & Topological Spaces.* New York: Oxford University Press, 1975.

Points Problem

see SHARING PROBLEM

Pointwise Dimension

$$D_P(\mathbf{x}) \equiv \lim_{\epsilon \to 0} \frac{\ln \mu(B_\epsilon(\mathbf{x}))}{\ln \epsilon},$$

where $B_\epsilon(\mathbf{x})$ is an n-D BALL of RADIUS ϵ centered at $\mathbf{x}$ and μ is the PROBABILITY MEASURE.

see also BALL, PROBABILITY MEASURE

References

Nayfeh, A. H. and Balachandran, B. *Applied Nonlinear Dynamics: Analytical, Computational, and Experimental Methods.* New York: Wiley, pp. 541–545, 1995.

Poisson's Bessel Function Formula

For $\Re[\nu] > -1/2$,

$$J_\nu(z) = \left(\frac{z}{2}\right)^\nu \frac{2}{\sqrt{\pi}\,\Gamma(\nu + \frac{1}{2})} \int_0^{\pi/2} \cos(z \cos t) \sin^{2\nu} t \, dt,$$

where $J_\nu(z)$ is a BESSEL FUNCTION OF THE FIRST KIND, and $\Gamma(z)$ is the GAMMA FUNCTION.

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1472, 1980.

Poisson Bracket

Let F and G be infinitely differentiable functions of x and p. Then the Poisson bracket is defined by

$$(F, G) = \sum_{\nu=1}^{n} \left(\frac{\partial F}{\partial p_\nu} \frac{\partial G}{\partial x_p} - \frac{\partial G}{\partial p_\nu} \frac{\partial F}{\partial x_\nu} \right).$$

If F and G are functions of x and p only, then the LAGRANGE BRACKET $[F, G]$ collapses the Poisson bracket (F, G).

see also LAGRANGE BRACKET, LIE BRACKET

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1004, 1980.

Poisson-Charlier Function

$$\rho_n(\nu, x) \equiv \frac{(1 + \nu - n)_n}{\sqrt{n! x^n}} \, {}_1F_1(-n; 1 + \nu - n; x),$$

where $(a)_n$ is a POCHHAMMER SYMBOL and ${}_1F_1(a; b; z)$ is a CONFLUENT HYPERGEOMETRIC FUNCTION.

see also POISSON-CHARLIER POLYNOMIAL

Poisson-Charlier Polynomial

Polynomials $p_n(x)$ which belong to the distribution $d\alpha(x)$ where $\alpha(x)$ is a STEP FUNCTION with JUMP

$$j(x) = e^{-a} a^x (x!)^{-1} \tag{1}$$

at $x = 0, 1, \ldots$ for $a > 0$.

$$\begin{aligned} p_n(x) &= a^{n/2} (n!)^{-1/2} \sum_{\nu=0}^{n} (-1)^{n-\nu} \binom{n}{\nu} \nu! a^{-\nu} \binom{x}{\nu} && (2) \\ &= a^{n/2} (n!)^{-1/2} (-1)^n [j(x)]^{-1} \Delta^n j(x - n) && (3) \\ &= a^{-n/2} \sqrt{n!}\, L_n^{x-n}(a), && (4) \end{aligned}$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT, $L_n^k(x)$ is an associated LAGUERRE POLYNOMIAL, and

$$\begin{aligned} &\Delta f(x) = f(x + 1) - f(x) && (5) \\ &\Delta^n f(x) = \Delta[\Delta^{n-1} f(x)] \\ &= f(x + n) - \binom{n}{1} f(x + n - 1) + \ldots + (-1)^n f(x). && (6) \end{aligned}$$

see also POISSON-CHARLIER FUNCTION

References

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., pp. 34–35, 1975.

Poisson Distribution

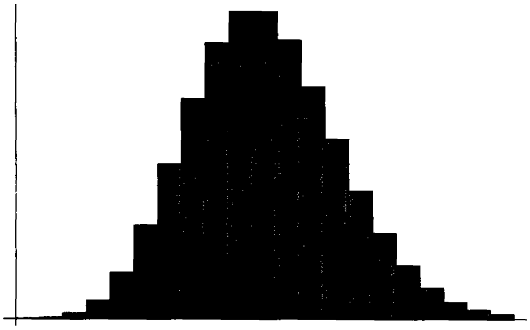

A Poisson distribution is a distribution with the following properties:

1. The number of changes in nonoverlapping intervals are independent for all intervals.
2. The probability of exactly one change in a sufficiently small interval $h \equiv 1/n$ is $P = \nu h \equiv \nu/n$, where ν is the probability of one change and n is the number of TRIALS.
3. The probability of two or more changes in a sufficiently small interval h is essentially 0.

The probability of k changes in a given interval is then given by the limit of the BINOMIAL DISTRIBUTION

$$P(k) = \frac{n!}{k!(n-k)!}\left(\frac{\nu}{n}\right)^k \left(1-\frac{\nu}{n}\right)^{n-k} \tag{1}$$

as the number of trials becomes very large,

$$\begin{aligned}\lim_{n\to\infty} P(k) &= \\ \lim_{n\to\infty} \frac{n(n-1)\cdots(n-k-1)}{n^k}\frac{\nu^k}{k!}&\left(1-\frac{\nu}{n}\right)^n\left(1-\frac{\nu}{n}\right)^{-k} \\ &= (1)\left(\frac{\nu^k}{k!}\right)(e^{-\nu})(1) = \frac{\nu^k e^{-\nu}}{k!}. \end{aligned}\tag{2}$$

This should be normalized so that the sum of probabilities equals 1. Indeed,

$$\sum_{k=0}^{\infty} P(k) = e^{\nu}\sum_{k=0}^{\infty}\frac{\nu^k}{k!} = e^{\nu}e^{-\nu} = 1, \tag{3}$$

as required. The ratio of probabilities is given by

$$\frac{P(k=i+1)}{P(k=i)} = \frac{\frac{\nu^{i+1}e^{-\nu}}{(i+1)!}}{\frac{i!}{e^{-\nu}\nu^i}} = \frac{\nu}{i+1}. \tag{4}$$

The MOMENT-GENERATING FUNCTION of this distribution is

$$\begin{aligned} M(t) &= \sum_{k=0}^{\infty} e^{tk}\frac{\nu^k e^{-\nu}}{k!} = e^{-\nu}\sum_{k=0}^{\infty}\frac{(\nu e^t)^k}{k!} \\ &= e^{-\nu}e^{\nu e^t} = e^{\nu(e^t-1)} && (5) \\ M'(t) &= \nu e^t e^{\nu(e^t-1)} && (6) \\ M''(t) &= (\nu e^t)^2 e^{\nu(e^t-1)} + \nu e^t e^{\nu(e^t-1)} && (7) \\ R(t) &\equiv \ln M(t) = \nu(e^t-1) && (8) \\ R'(t) &= \nu e^t && (9) \\ R''(t) &= \nu e^t, && (10) \end{aligned}$$

so

$$\mu = R'(0) = \nu \tag{11}$$

$$\sigma^2 = R''(0) = \nu. \tag{12}$$

The MOMENTS about zero can also be computed directly

$$\mu_2' = \nu(1+\nu) \tag{13}$$

$$\mu_3' = \nu(1+3\nu+\nu^2) \tag{14}$$

$$\mu_4' = \nu(1+7\nu+6\nu^2+\nu^3), \tag{15}$$

as can the MOMENTS about the MEAN.

$$\mu_1 = \nu \tag{16}$$

$$\mu_2 = \nu \tag{17}$$

$$\mu_3 = \nu \tag{18}$$

$$\mu_4 = \nu(1+3\nu), \tag{19}$$

so the MEAN, VARIANCE, SKEWNESS, and KURTOSIS are

$$\mu = \nu \tag{20}$$

$$\sigma^2 = \nu \tag{21}$$

$$\gamma_1 \equiv \frac{\mu_3}{\sigma^3} = \frac{\nu}{\nu^{3/2}} = \nu^{-1/2} \tag{22}$$

$$\begin{aligned}\gamma_2 &\equiv \frac{\mu_4}{\sigma^4} - 3 = \frac{\nu(1+3\nu)}{\nu} - 3 \\ &= \frac{\nu + 3\nu^2 - 3\nu^2}{\nu^2} = \nu^{-1}. \end{aligned}\tag{23}$$

The CHARACTERISTIC FUNCTION is

$$\phi(t) = e^{m(e^{it}-1)} \tag{24}$$

and the CUMULANT-GENERATING FUNCTION is

$$K(h) = \nu(e^h - 1) = \nu(h + \tfrac{1}{2!}h^2 + \tfrac{1}{3!}h^3 + \ldots), \tag{25}$$

so

$$\kappa_r = \nu. \tag{26}$$

The Poisson distribution can also be expressed in terms of

$$\lambda \equiv \frac{\nu}{x}, \tag{27}$$

the rate of changes, so that

$$P(k) = \frac{(\lambda x)^k e^{-\lambda x}}{k!}. \tag{28}$$

The MOMENT-GENERATING FUNCTION of a Poisson distribution in two variables is given by

$$M(t) = e^{(\nu_1+\nu_2)(e^t-1)}. \tag{29}$$

If the independent variables $x_1, x_2, \ldots, x_N$ have Poisson distributions with parameters $\mu_1, \mu_2, \ldots, \mu_N$, then

$$X = \sum_{j=1}^{N} x_j \tag{30}$$

has a Poisson distribution with parameter

$$\mu = \sum_{j=1}^{N} \mu_j. \tag{31}$$

This can be seen since the Cumulant-Generating Function is

$$K_j(h) = \mu_j(e^h - 1), \tag{32}$$

$$K \equiv \sum_j K_j(h) = (e^h - 1)\sum_j \mu_j = \mu(e^h - 1). \tag{33}$$

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 532, 1987.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Incomplete Gamma Function, Error Function, Chi-Square Probability Function, Cumulative Poisson Function." §6.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 209–214, 1992.

Spiegel, M. R. *Theory and Problems of Probability and Statistics.* New York: McGraw-Hill, p. 111–112, 1992.

Poisson's Equation

A second-order Partial Differential Equation arising in physics:

$$\nabla^2 \psi = -4\pi\rho.$$

If $\rho = 0$, it reduces Laplace's Equation. It is also related to the Helmholtz Differential Equation

$$\nabla^2\psi + k^2\psi = 0.$$

see also Helmholtz Differential Equation, Laplace's Equation

References

Arfken, G. "Gauss's Law, Poisson's Equation." §1.14 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 74–78, 1985.

Poisson's Harmonic Function Formula

Let $\phi(z)$ be a Harmonic Function. Then

$$\phi(z_0) = \frac{1}{2\pi}\int_0^{2\pi} K(r,\theta)\phi(z_0 + re^{i\theta})\,d\theta, \tag{1}$$

where $R = |z_0|$ and $K(r,\theta)$ is the Poisson Kernel. For a Circle,

$$u(x,y) = \frac{1}{2\pi}\int_0^{2\pi} u(a\cos\phi, a\sin\phi) \frac{a^2 - R^2}{a^2 + R^2 - 2ar\cos(\theta - \phi)}\,d\phi. \tag{2}$$

For a Sphere,

$$u(x,y,z) = \frac{1}{4\pi a}\iint_S u\frac{a^2 - R^2}{(a^2 + R^2 - 2aR\cos\theta)^{3/2}}\,dS, \tag{3}$$

where

$$\cos\theta \equiv \mathbf{x}\cdot\boldsymbol{\xi}. \tag{4}$$

see also Circle, Harmonic Function, Poisson Kernel, Sphere

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 373–374, 1953.

Poisson Integral

A.k.a. Bessel's Second Integral.

$$J_n(z) = \frac{\left(\frac{1}{2}\right)^n}{\Gamma(n + \frac{1}{2})\Gamma(\frac{1}{2})}\int_0^{\pi} \cos(z\cos\theta)\sin^{2n}\theta\,d\theta,$$

where $J_n(z)$ is a Bessel Function of the First Kind and $\Gamma(x)$ is a Gamma Function. It can be derived from Sonine's Integral. With $n = 0$, the integral becomes Parseval's Integral.

see also Bessel Function of the First Kind, Parseval's Integral, Sonine's Integral

Poisson Integral Representation

$$j_n(z) = \frac{z^n}{2^{n+1}n!}\int_0^{\pi}\cos(z\cos\theta)\sin^{2n+1}\theta\,d\theta,$$

where $j_n(z)$ is a Spherical Bessel Function of the First Kind.

Poisson Kernel

In 2-D,

$$\begin{aligned} K(r,\theta) &\equiv \Re\left(\frac{R + re^{i\theta}}{R - re^{i\theta}}\right) \\ &= \Re\left[\frac{(R + re^{i\theta})(R - re^{-i\theta})}{(R - re^{i\theta})(R - re^{-i\theta})}\right] \\ &= \Re\left[\frac{R^2 - rR(e^{i\theta} - e^{-i\theta}) - r^2}{R^2 - rR(e^{i\theta} + e^{-i\theta}) + r^2}\right] \\ &= \Re\left[\frac{R^2 + 2irR\sin\theta - r^2}{R^2 - 2Rr\cos\theta + r^2}\right] \\ &= \frac{R^2 - r^2}{R^2 - 2Rr\cos\theta + r^2}. \end{aligned} \tag{1}$$

In 3-D,

$$u(\mathbf{y}) = \frac{R(R^2 - a^2)}{4\pi} \times \int_0^{2\pi}\int_0^{\pi} \frac{f(\theta,\phi)\sin\theta\,d\theta\,d\phi}{(R^2 + a^2 - 2aR\cos\gamma)^{3/2}}, \tag{2}$$

where $a = |\mathbf{y}|$ and

$$\cos\gamma = \mathbf{y} \cdot \begin{bmatrix} R\cos\theta\sin\phi \\ R\sin\theta\sin\phi \\ R\cos\phi \end{bmatrix}. \quad (3)$$

The Poisson kernel for the n-BALL is

$$P(\mathbf{x},\mathbf{z}) = \frac{1}{2-n}(D_\mathbf{n}\mathbf{v})(\mathbf{z}), \quad (4)$$

where $D_\mathbf{n}$ is the outward normal derivative at point $\mathbf{z}$ on a unit n-SPHERE and

$$\mathbf{v}(\mathbf{z}) = |\mathbf{z}-\mathbf{x}|^{2-n} - |\mathbf{x}|^{2-n} \left|\frac{\mathbf{x}}{|\mathbf{x}|^2}\right|^{2-n}. \quad (5)$$

see also POISSON'S HARMONIC FUNCTION FORMULA

References
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1090, 1979.

Poisson Manifold

A smooth MANIFOLD with a POISSON BRACKET defined on its FUNCTION SPACE.

Poisson Sum Formula

A special case of the general result

$$\sum_{n=-\infty}^{\infty} f(x+n) = \sum_{k=-\infty}^{\infty} e^{2\pi ikx} \int_{-\infty}^{\infty} f(x_1)e^{-2\pi ikx}\,dx_1 \quad (1)$$

with $x = 0$, yielding

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1)e^{-2\pi ikx}\,dx_1. \quad (2)$$

An alternate form is

$$\frac{1}{2} + \sum_{n=1}^{\infty} e^{-(nx)^2} = \frac{\sqrt{\pi}}{x}\left[\frac{1}{2} + \sum_{n=1}^{\infty} e^{-(n\pi/x)^2}\right]. \quad (3)$$

Another formula called the Poisson summation formula is

$$\sqrt{\alpha}\,[\tfrac{1}{2}\phi(0) + \phi(\alpha) + \phi(2\alpha) + \ldots] = \sqrt{\beta}[\tfrac{1}{2}\psi(0) + \psi(\beta) + \psi(2\beta) + \ldots], \quad (4)$$

where

$$\psi(x) = \sqrt{\frac{2}{\pi}} \int_0^\infty \psi(t)\cos(xt)\,dt \quad (5)$$

$$\alpha\beta = 2\pi. \quad (6)$$

References
Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 466–467, 1953.

Poisson Trials

A number s TRIALS in which the probability of success p_i varies from trial to trial. Let x be the number of successes, then

$$\mathrm{var}(x) = spq - s{\sigma_p}^2, \quad (1)$$

where ${\sigma_p}^2$ is the VARIANCE of p_i and $q \equiv (1-p)$. Uspensky has shown that

$$P(s,x) = \beta\frac{m^x e^{-m}}{x!}, \quad (2)$$

where

$$\beta = [1 - \theta g(x)]e^{h(x)} \quad (3)$$

$$g(x) = \frac{(s-x)m^3}{3(s-m)^3} + \frac{x^3}{2s(s-x)} \quad (4)$$

$$h(x) = \frac{mx}{s} - \frac{m^2}{2s^2}(s-x) - \frac{x(x-1)}{2s} = p\left[\frac{x}{2}\left(1+\frac{1}{m}\right) - \frac{(x-m)^2}{2m}\right] \quad (5)$$

and $\theta \in (0,1)$. The probability that the number of successes is at least x is given by

$$Q_m(x) = \sum_{r=x}^{\infty} \frac{m^r e^{-m}}{r!}. \quad (6)$$

Uspensky gives the true probability that there are at least x successes in s trials as

$$P_{ms}(x) = Q_m(x) + \Delta, \quad (7)$$

where

$$|\Delta| < \begin{cases} (e^\chi - 1)Q_m(x+1) & \text{for } Q_m(x+1) \geq \frac{1}{2} \\ (e^\chi - 1)[1 - Q_m(x+1)] & \text{for } Q_m(x+1) \leq \frac{1}{2} \end{cases} \quad (8)$$

$$\chi = \frac{m + \frac{1}{4} + \frac{m^3}{s}}{2(s-m)}. \quad (9)$$

Poke Move

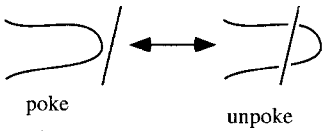

The REIDEMEISTER MOVE of type II.

see also REIDEMEISTER MOVES

Poker

Poker is a CARD game played with a normal deck of 52 CARDS. Sometimes, additional cards called "jokers" are also used. In straight or draw poker, each player is normally dealt a hand of five cards. Depending on the variant, players then discard and redraw CARDS, trying to improve their hands. Bets are placed at each discard step. The number of possible distinct five-card hands is

$$N = \binom{52}{5} = 2{,}598{,}960,$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT.

There are special names for specific types of hands. A royal flush is an ace, king, queen, jack, and 10, all of one suit. A straight flush is five consecutive cards all of the same suit (but not a royal flush), where an ace may count as either high or low. A full house is three-of-a-kind and a pair. A flush is five cards of the same suit (but not a royal flush or straight flush). A straight is five consecutive cards (but not a royal flush or straight flush), where an ace may again count as either high or low.

The probabilities of being dealt five-card poker hands of a given type (before discarding and with no jokers) on the initial deal are given below (Packel 1981). As usual, for a hand with probability P, the ODDS against being dealt it are $(1/r) - 1 : 1$.

Hand	Exact Probability
royal flush	$\frac{4}{N} = \frac{1}{649{,}740}$
straight flush	$\frac{4(10)-4}{N} = \frac{3}{216{,}580}$
four of a kind	$\frac{13(48)}{N} = \frac{1}{4{,}165}$
full house	$\frac{13\binom{4}{3}12\binom{4}{2}}{N} = \frac{6}{4{,}165}$
flush	$\frac{4\binom{13}{5}-36-4}{N} = \frac{1{,}277}{649{,}740}$
straight	$\frac{10(4^5)-36-4}{N} = \frac{5}{1{,}274}$
three of a kind	$\frac{13\binom{4}{3}\frac{(48)(44)}{2!}}{N} = \frac{88}{4{,}165}$
two pair	$\frac{\frac{13\binom{4}{2}12\binom{4}{2}}{2!}44}{N} = \frac{198}{4{,}165}$
one pair	$\frac{13\binom{4}{2}\frac{(48)(44)(40)}{3!}}{N} = \frac{352}{833}$

Hand	Probability	Odds
royal flush	1.54×10^{-6}	649,739.0:1
straight flush	1.39×10^{-5}	72,192.3:1
four of a kind	2.40×10^{-4}	4,164.0:1
full house	1.44×10^{-3}	693.2:1
flush	1.97×10^{-3}	507.8:1
straight	3.92×10^{-3}	253.8:1
three of a kind	0.0211	46.3:1
two pair	0.0475	20.0:1
one pair	0.423	1.366:1

Gadbois (1996) gives probabilities for hands if two jokers are included, and points out that it is *impossible* to rank hands in any single way which is consistent with the relative frequency of the hands.

see also BRIDGE CARD GAME, CARDS

References

Cheung, Y. L. "Why Poker is Played with Five Cards." *Math. Gaz.* **73**, 313–315, 1989.

Conway, J. H. and Guy, R. K. "Choice Numbers with Repetitions." In *The Book of Numbers.* New York: Springer-Verlag, pp. 70–71, 1996.

Gadbois, S. "Poker with Wild Cards—A Paradox?" *Math. Mag.* **69**, 283–285, 1996.

Jacoby, O. *Oswald Jacoby on Poker.* New York: Doubleday, 1981.

Packel, E. W. *The Mathematics of Games and Gambling.* Washington, DC: Math. Assoc. Amer., 1981.

Polar

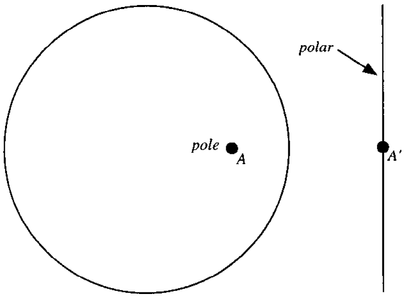

If two points A and A' are INVERSE with respect to a CIRCLE (the INVERSION CIRCLE), then the straight line through A' which is PERPENDICULAR to the line of the points AA' is called the polar of A with respect to the CIRCLE, and A is called the POLE of the polar.

see also APOLLONIUS' PROBLEM, INVERSE POINTS, INVERSION CIRCLE, POLARITY, POLE, TRILINEAR POLAR

References

Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, p. 157, 1965.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 100–106, 1929.

Polar Angle

The ANGLE a point makes from the ORIGIN as measured from the x-AXIS.

see also POLAR COORDINATES

Polar Circle

Given a TRIANGLE, the polar circle has center at the ORTHOCENTER H. Call H_i the FEET of the ALTITUDE. Then the RADIUS is

$$\begin{aligned} r^2 &= \overline{HA_1} \cdot \overline{HH_1} = \overline{HA_2} \cdot \overline{HH_2} = \overline{HA_2} \cdot \overline{HH_2} && (1) \\ &= -4R^2 \cos\alpha_1 \cos\alpha_2 \cos\alpha_3 && (2) \\ &= \tfrac{1}{2}({a_1}^2 + {a_2}^2 + {a_3}^2) - 4R^2, && (3) \end{aligned}$$

where R is the CIRCUMRADIUS, α_i the VERTEX angles, and a_i the corresponding side lengths.

A TRIANGLE is self-conjugate with respect to its polar circle. Also, the RADICAL AXIS of any two polar circles is the ALTITUDE from the third VERTEX. Any two polar circles of an ORTHOCENTRIC SYSTEM are orthogonal. The polar circles of the triangles of a COMPLETE QUADRILATERAL constitute a COAXAL SYSTEM conjugate to that of the circles on the diagonals.

see also COAXAL SYSTEM, ORTHOCENTRIC SYSTEM, POLAR, POLE, RADICAL AXIS

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 136–138, 1967.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 176–181, 1929.

Polar Coordinates

The polar coordinates r and θ are defined by

$$x = r\cos\theta \tag{1}$$

$$y = r\sin\theta. \tag{2}$$

In terms of x and y,

$$r = \sqrt{x^2+y^2} \tag{3}$$

$$\theta = \tan^{-1}\left(\frac{y}{x}\right). \tag{4}$$

The ARC LENGTH of a polar curve given by $r = r(\theta)$ is

$$s = \int_{\theta_1}^{\theta_2} \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}\, d\theta. \tag{5}$$

The LINE ELEMENT is given by

$$ds^2 = r^2\, d\theta^2, \tag{6}$$

and the AREA element by

$$dA = r\, dr\, d\theta. \tag{7}$$

The AREA enclosed by a polar curve $r = r(\theta)$ is

$$A = \tfrac{1}{2}\int_{\theta_1}^{\theta_2} r^2\, d\theta. \tag{8}$$

The SLOPE of a polar function $r = r(\theta)$ at the point (r, θ) is given by

$$m = \frac{r + \tan\theta\frac{dr}{d\theta}}{-r\tan\theta + \frac{dr}{d\theta}}. \tag{9}$$

The ANGLE between the tangent and radial line at the point (r, θ) is

$$\psi = \tan^{-1}\left(\frac{r}{\frac{dr}{d\theta}}\right). \tag{10}$$

A polar curve is symmetric about the x-axis if replacing θ by $-\theta$ in its equation produces an equivalent equation, symmetric about the y-axis if replacing θ by $\pi - \theta$ in its equation produces an equivalent equation, and symmetric about the origin if replacing r by $-r$ in its equation produces an equivalent equation.

In Cartesian coordinates, the POSITION VECTOR and its derivatives are

$$\mathbf{r} = \sqrt{x^2+y^2}\,\hat{\mathbf{r}} \tag{11}$$

$$\dot{\mathbf{r}} = \dot{\hat{\mathbf{r}}}\sqrt{x^2+y^2} + \hat{\mathbf{r}}\,(x^2+y^2)^{-1/2}(x\dot{x}+y\dot{y}) \tag{12}$$

$$\hat{\mathbf{r}} = \frac{x\hat{\mathbf{x}}+y\hat{\mathbf{y}}}{\sqrt{x^2+y^2}} \tag{13}$$

$$\begin{aligned}\dot{\hat{\mathbf{r}}} &= \frac{\dot{x}\hat{\mathbf{x}}+\dot{y}\hat{\mathbf{y}}}{\sqrt{x^2+y^2}}\\ &\quad - \tfrac{1}{2}(x^2+y^2)^{-3/2}(2)(x\dot{x}+y\dot{y})(x\hat{\mathbf{x}}+y\hat{\mathbf{y}})\\ &= \frac{(x\dot{y}-y\dot{x})(x\hat{\mathbf{y}}-y\hat{\mathbf{x}})}{(x^2+y^2)^{3/2}}.\end{aligned} \tag{14}$$

In polar coordinates, the UNIT VECTORS and their derivatives are

$$\mathbf{r} \equiv \begin{bmatrix} r\cos\theta \\ r\sin\theta \end{bmatrix} \tag{15}$$

$$\hat{\mathbf{r}} \equiv \frac{\frac{d\mathbf{r}}{dr}}{\left|\frac{d\mathbf{r}}{dr}\right|} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \tag{16}$$

$$\hat{\boldsymbol{\theta}} \equiv \frac{\frac{d\boldsymbol{\theta}}{d\theta}}{\left|\frac{d\boldsymbol{\theta}}{d\theta}\right|} = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix} \tag{17}$$

$$\dot{\hat{\mathbf{r}}} = \begin{bmatrix} -\sin\theta\dot{\theta} \\ \cos\theta\dot{\theta} \end{bmatrix} = \dot{\theta}\,\hat{\boldsymbol{\theta}} \tag{18}$$

$$\dot{\hat{\boldsymbol{\theta}}} = \begin{bmatrix} -\cos\theta\dot{\theta} \\ -\sin\theta\dot{\theta} \end{bmatrix} = -\dot{\theta}\hat{\mathbf{r}} \tag{19}$$

$$\dot{\mathbf{r}} = \begin{bmatrix} -r\sin\theta\dot{\theta} + \cos\theta\dot{r} \\ r\cos\theta\dot{\theta} + \sin\theta\dot{r} \end{bmatrix} = r\dot{\theta}\,\hat{\boldsymbol{\theta}} + \dot{r}\,\hat{\mathbf{r}} \tag{20}$$

$$\begin{aligned}\ddot{\mathbf{r}} &= \dot{r}\dot{\theta}\,\hat{\boldsymbol{\theta}} + r\ddot{\theta}\hat{\boldsymbol{\theta}} + r\dot{\theta}\dot{\hat{\boldsymbol{\theta}}} + \ddot{r}\hat{\mathbf{r}} + \dot{r}\dot{\hat{\mathbf{r}}}\\ &= \dot{r}\dot{\theta}\hat{\boldsymbol{\theta}} + \dot{r}\ddot{\theta}\hat{\boldsymbol{\theta}} + r\dot{\theta}(-\dot{\theta}\hat{\mathbf{r}}) + \ddot{r}\hat{\mathbf{r}} + \dot{r}\dot{\theta}\hat{\boldsymbol{\theta}}\\ &= (\ddot{r} - r\dot{\theta}^2)\hat{\mathbf{r}} + (2\dot{r}\dot{\theta} + r\ddot{\theta})\hat{\boldsymbol{\theta}}\\ &= (\ddot{r} - r\dot{\theta}^2)\hat{\mathbf{r}} + \frac{1}{r}\frac{d}{dt}(r^2\dot{\theta})\hat{\boldsymbol{\theta}}.\end{aligned} \tag{21}$$

see also CARDIOID, CIRCLE, CISSOID, CONCHOID, CURVILINEAR COORDINATES, CYLINDRICAL COORDINATES, EQUIANGULAR SPIRAL, LEMNISCATE, LIMAÇON, ROSE

Polar Line

see POLAR

Polarity

A projective CORRELATION of period two. In a polarity, a is called the POLAR of A, and A the POLE a.

see also CHASLES'S THEOREM, CORRELATION, POLAR, POLE (GEOMETRY)

Pole

A COMPLEX function f has a pole of order m at z_0 if, in the LAURENT SERIES, $a_n = 0$ for $n < -m$ and $a_m \neq 0$. Equivalently, f has a pole of order n at z_0 if n is the smallest POSITIVE INTEGER for which $(z-z_0)^n f(z)$ is differentiable at z_0. If $f(\pm\infty) \neq \pm\infty$, there is no pole at $\pm\infty$. Otherwise, the order of the pole is the greatest POSITIVE COEFFICIENT in the LAURENT SERIES.

This is equivalent to finding the smallest n such that

$$\frac{(z-z_0)^n}{f(z)}$$

is differentiable at 0.

see also LAURENT SERIES, RESIDUE (COMPLEX ANALYSIS)

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 396–397, 1985.

Pole (Geometry)

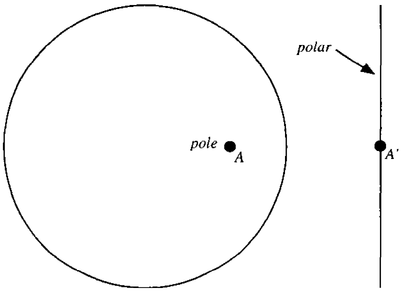

If two points A and A' are INVERSE with respect to a CIRCLE (the INVERSION CIRCLE), then the straight line through A' which is PERPENDICULAR to the line of the points AA' is called the POLAR of the A with respect to the CIRCLE, and A is called the pole of the POLAR.

see also INVERSE POINTS, INVERSION CIRCLE, POLAR, POLARITY, TRILINEAR POLAR

References

Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, p. 157, 1965.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 100–106, 1929.

Pole (Origin)

see ORIGIN

Policeman on Point Duty Curve

see CRUCIFORM

Polignac's Conjecture

see DE POLIGNAC'S CONJECTURE

Polish Space

The HOMEOMORPHIC image of a so-called "complete separable" METRIC SPACE. The continuous image of a Polish space is called a SOUSLIN SET.

see also DESCRIPTIVE SET THEORY, STANDARD SPACE

Pollaczek Polynomial

Let $a > |b|$, and write

$$h(\theta) = \frac{a\cos\theta + b}{2\sin\theta}. \tag{1}$$

Then define $P_n(x;a,b)$ by the GENERATING FUNCTION

$$\begin{aligned} f(x,w) = f(\cos\theta,w) &= \sum_{n=0}^{\infty} P_n(x;a,b)w^n \\ &= (1-we^{i\theta})^{-1/2+ih(\theta)}(1-we^{i\theta})^{-1/2-ih(\theta)}. \end{aligned} \tag{2}$$

The GENERATING FUNCTION may also be written

$$\begin{aligned} f(x,w) &= (1-2xw+w^2)^{-1/2} \\ &\quad \exp\left[(ax+b)\sum_{m=1}^{\infty}\frac{w^m}{m}U_{m-1}(x)\right], \end{aligned} \tag{3}$$

where $U_m(x)$ is a CHEBYSHEV POLYNOMIAL OF THE SECOND KIND. They satisfy the RECURRENCE RELATION

$$\begin{aligned} nP_n(x;a,b) &= [(2n-1+2a)x+2b]P_{n-1}(x;a,b) \\ &\quad -(n-1)P_{n-2}(x;a,b) \end{aligned} \tag{4}$$

for $n = 2, 3, \ldots$ with

$$P_0 = 1 \tag{5}$$

$$P_1 = (2a+1)x + 2b. \tag{6}$$

In terms of the HYPERGEOMETRIC FUNCTION ${}_2F_1(a,b;c;x)$,

$$P_n(\cos\theta;a;b) = e^{in\theta}{}_2F_1(-n,\tfrac{1}{2}+ih(\theta);1;1-e^{-2i\theta}). \tag{7}$$

They obey the orthogonality relation

$$\begin{aligned} \int_{-1}^{1} & P_n(x;a,b)P_m(x;a,b)w(x;a,b)\,dx \\ &= [n+\tfrac{1}{2}(a+1)]^{-1}\delta_{nm}, \end{aligned} \tag{8}$$

where δ_{nm} is the KRONECKER DELTA, for $n, m = 0, 1, \ldots$, with the WEIGHT FUNCTION

$$w(\cos\theta;a,b) = e^{(2\theta-\pi)h(\theta)}\{\cosh[\pi h(\theta)]\}^{-1}. \tag{9}$$

References

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., pp. 393–400, 1975.

Pollard Monte Carlo Factorization Method

see POLLARD ρ FACTORIZATION METHOD

Pollard $p-1$ Factorization Method

A PRIME FACTORIZATION ALGORITHM which can be implemented in a single-step or double-step form. In the single-step version, PRIMES p are found if $p-1$ is a product of small PRIMES by finding an m such that

$$m \equiv c^q \pmod{n},$$

where $p-1|q$, with q a large number and $(c,n)=1$. Then since $p-1|q$, $m \equiv 1 \pmod{p}$, so $p|m-1$. There is therefore a good chance that $n \nmid m-1$, in which case $\mathrm{GCD}(m-1,n)$ (where GCD is the GREATEST COMMON DIVISOR) will be a nontrivial divisor of n.

In the double-step version, a PRIMES p can be factored if $p-1$ is a product of small PRIMES and a single larger PRIME.

see also PRIME FACTORIZATION ALGORITHMS, WILLIAMS $p+1$ FACTORIZATION METHOD

References

Bressoud, D. M. *Factorization and Prime Testing.* New York: Springer-Verlag, pp. 67–69, 1989.

Pollard, J. M. "Theorems on Factorization and Primality Testing." *Proc. Cambridge Phil. Soc.* **76**, 521–528, 1974.

Pollard ρ Factorization Method

A PRIME FACTORIZATION ALGORITHM also known as POLLARD MONTE CARLO FACTORIZATION METHOD. Let $x_0=2$, then compute

$$x_{i+1} = {x_i}^2 - x_i + 1 \pmod{n}.$$

If $\mathrm{GCD}(x_{2i}-x_i,n)>1$, then n is COMPOSITE and its factors are found. In modified form, it becomes BRENT'S FACTORIZATION METHOD. In practice, almost any unfactorable POLYNOMIAL can be used for the iteration (x^2-2, however, cannot). Under worst conditions, the ALGORITHM can be very slow.

see also BRENT'S FACTORIZATION METHOD, PRIME FACTORIZATION ALGORITHMS

References

Brent, R. P. "Some Integer Factorization Algorithms Using Elliptic Curves." *Austral. Comp. Sci. Comm.* **8**, 149–163, 1986.

Bressoud, D. M. *Factorization and Prime Testing.* New York: Springer-Verlag, pp. 61–67, 1989.

Eldershaw, C. and Brent, R. P. "Factorization of Large Integers on Some Vector and Parallel Computers." `ftp://nimbus.anu.edu.au/pub/Brent/156tr.dvi.Z`.

Montgomery, P. L. "Speeding the Pollard and Elliptic Curve Methods of Factorization." *Math. Comput.* **48**, 243–264, 1987.

Pollard, J. M. "A Monte Carlo Method for Factorization." *Nordisk Tidskrift for Informationsbehandlung (BIT)* **15**, 331–334, 1975.

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 83 and 102–103, 1991.

Poloidal Field

A VECTOR FIELD resembling a magnetic multipole which has a component along the z-AXIS of a SPHERE and continues along lines of LONGITUDE.

see also DIVERGENCELESS FIELD, TOROIDAL FIELD

References

Stacey, F. D. *Physics of the Earth, 2nd ed.* New York: Wiley, p. 239, 1977.

Pólya-Burnside Lemma

see PÓLYA ENUMERATION THEOREM

Pólya Conjecture

Let n be a POSITIVE INTEGER and $r(n)$ the number of (not necessarily distinct) PRIME FACTORS of n (with $r(1)=0$). Let $O(m)$ be the number of POSITIVE INTEGERS $\leq m$ with an ODD number of PRIME factors, and $E(m)$ the number of POSITIVE INTEGERS $\leq m$ with an EVEN number of PRIME factors. Pólya conjectured that

$$L(m) \equiv E(m) - O(m) = \sum_{n=1}^{m} \lambda(n)$$

is ≤ 0, where $\lambda(n)$ is the LIOUVILLE FUNCTION.

The conjecture was made in 1919, and disproven by Haselgrove (1958) using a method due to Ingham (1942). Lehman (1960) found the first explicit counterexample, $L(906{,}180{,}359)=1$, and the smallest counterexample $m=906{,}150{,}257$ was found by Tanaka (1980). The first n for which $L(n)=0$ are $n=2$, 4, 6, 10, 16, 26, 40, 96, 586, 906150256, ... (Tanaka 1980, Sloane's A028488). It is unknown if $L(x)$ changes sign infinitely often (Tanaka 1980).

see also ANDRICA'S CONJECTURE, LIOUVILLE FUNCTION, PRIME FACTORS

References

Haselgrove, C. B. "A Disproof of a Conjecture of Pólya." *Mathematika* **5**, 141–145, 1958.

Ingham, A. E. "On Two Conjectures in the Theory of Numbers." *Amer. J. Math.* **64**, 313–319, 1942.

Lehman, R. S. "On Liouville's Function." *Math. Comput.* **14**, 311–320, 1960.

Sloane, N. J. A. Sequence A028488 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Tanaka, M. "A Numerical Investigation on Cumulative Sum of the Liouville Function" [sic]. *Tokyo J. Math.* **3**, 187–189, 1980.

Pólya Distribution

see NEGATIVE BINOMIAL DISTRIBUTION

Pólya Enumeration Theorem

A very general theorem which allows the number of discrete combinatorial objects of a given type to be enumerated (counted) as a function of their "order." The most common application is in the counting of the number of GRAPHS of n nodes, TREES and ROOTED TREES with n branches, GROUPS of order n, etc. The theorem is an extension of BURNSIDE'S LEMMA and is sometimes also called the PÓLYA-BURNSIDE LEMMA.

see also BURNSIDE'S LEMMA, GRAPH (GRAPH THEORY), GROUP, ROOTED TREE, TREE

References

Harary, F. "The Number of Linear, Directed, Rooted, and Connected Graphs." *Trans. Amer. Math. Soc.* **78**, 445–463, 1955.

Pólya, G. "Kombinatorische Anzahlbestimmungen für Gruppen, Graphen, und chemische Verbindungen." *Acta Math.* **68**, 145–254, 1937.

Pólya Polynomial

The POLYNOMIAL giving the number of colorings, with m colors, of a structure defined by a PERMUTATION GROUP.

see also PERMUTATION GROUP, PÓLYA ENUMERATION THEOREM

Pólya's Random Walk Constants

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let $p(d)$ be the probability that a RANDOM WALK on a d-D lattice returns to the origin. Pólya (1921) proved that

$$p(1) = p(2) = 1, \tag{1}$$

but

$$p(d) < 1 \tag{2}$$

for $d > 2$. Watson (1939), McCrea and Whipple (1940), Domb (1954), and Glasser and Zucker (1977) showed that

$$p(3) = 1 - \frac{1}{u(3)} = 0.3405373296\ldots, \tag{3}$$

where

$$u(3) = \frac{3}{(2\pi)^3}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{dx\,dy\,dz}{3-\cos x-\cos y-\cos z} \tag{4}$$

$$= \frac{12}{\pi^2}(18+12\sqrt{2}-10\sqrt{3}-7\sqrt{6}) \times \{K[(2-\sqrt{3})(\sqrt{3}-\sqrt{2})]\}^2 \tag{5}$$

$$= 3(18+12\sqrt{2}-10\sqrt{3}-7\sqrt{6}) \times \left[1+2\sum_{k=1}^{\infty}\exp(-k^2\pi\sqrt{6})\right]^4 \tag{6}$$

$$= \frac{\sqrt{6}}{32\pi^3}\Gamma(\tfrac{1}{24})\Gamma(\tfrac{5}{24})\Gamma(\tfrac{7}{24})\Gamma(\tfrac{11}{24}) \tag{7}$$

$$= 1.5163860592\ldots, \tag{8}$$

where $K(k)$ is a complete ELLIPTIC INTEGRAL OF THE FIRST KIND and $\Gamma(z)$ is the GAMMA FUNCTION. Closed forms for $d > 3$ are not known, but Montroll (1956) showed that

$$p(d) = 1 - [u(d)]^{-1}, \tag{9}$$

where

$$u(d) = \frac{d}{(2\pi)^d}\underbrace{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}}_{d}\left(d-\sum_{k=1}^{d}\cos x_k\right)^{-1} \times dx_1\,dx_2\cdots dx_d$$

$$= \int_0^{\infty}\left[I_0\left(\frac{t}{d}\right)\right]^d e^{-t}\,dt, \tag{10}$$

and $I_0(z)$ is a MODIFIED BESSEL FUNCTION OF THE FIRST KIND. Numerical values from Montroll (1956) and Flajolet (Finch) are

d	$p(d)$
4	0.20
5	0.136
6	0.105
7	0.0858
8	0.0729

see also RANDOM WALK

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/polya/polya.html`.

Domb, C. "On Multiple Returns in the Random-Walk Problem." *Proc. Cambridge Philos. Soc.* **50**, 586–591, 1954.

Glasser, M. L. and Zucker, I. J. "Extended Watson Integrals for the Cubic Lattices." *Proc. Nat. Acad. Sci. U.S.A.* **74**, 1800–1801, 1977.

McCrea, W. H. and Whipple, F. J. W. "Random Paths in Two and Three Dimensions." *Proc. Roy. Soc. Edinburgh* **60**, 281–298, 1940.

Montroll, E. W. "Random Walks in Multidimensional Spaces, Especially on Periodic Lattices." *J. SIAM* **4**, 241–260, 1956.

Watson, G. N. "Three Triple Integrals." *Quart. J. Math., Oxford Ser. 2* **10**, 266–276, 1939.

Pólya-Vinogradov Inequality

Let χ be a nonprincipal character (mod q). Then

$$\sum_{n=M+1}^{M+N}\chi(n) \ll \sqrt{q}\ln q,$$

where $\ll$ indicates MUCH LESS than.

References

Davenport, H. "The Pólya-Vinogradov Inequality." Ch. 23 in *Multiplicative Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 135–138, 1980.

Pólya, G. "Über die Verteilung der quadratischen Reste und Nichtreste." *Nachr. Königl. Gesell. Wissensch. Göttingen, Math.-Phys. Klasse,* 21–29, 1918.

Polyabolo

An analog of the POLYOMINO composed of n ISOSCELES RIGHT TRIANGLES joined along edges of the same length. The number of polyaboloes composed of n triangles are 1, 3, 4, 14, 30, 107, 318, 1106, 3671, ... (Sloane's A006074).

see also DIABOLO, HEXABOLO, PENTABOLO, TETRABOLO, TRIABOLO

References

Sloane, N. J. A. Sequence A006074/M2379 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Polyconic Projection

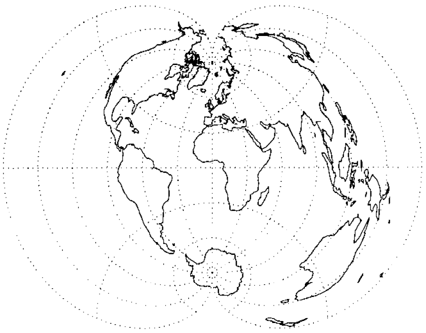

$$x = \cot\phi \sin E \tag{1}$$

$$y = (\phi - \phi_0) + \cot\phi(1 - \cos E), \tag{2}$$

where

$$E = (\lambda - \lambda_0)\sin\phi. \tag{3}$$

The inverse FORMULAS are

$$\lambda = \frac{\sin^{-1}(x\tan\phi)}{\sin\phi} + \lambda_0, \tag{4}$$

and ϕ is determined from

$$\Delta\phi = -\frac{A(\phi\tan\phi + 1) - \phi - \frac{1}{2}(\phi^2 + B)\tan\phi}{\frac{\phi - A}{\tan\phi} - 1}, \tag{5}$$

where $\phi_0 = A$ and

$$A = \phi_0 + y \tag{6}$$

$$B = x^2 + A^2. \tag{7}$$

References

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 124–137, 1987.

Polycube

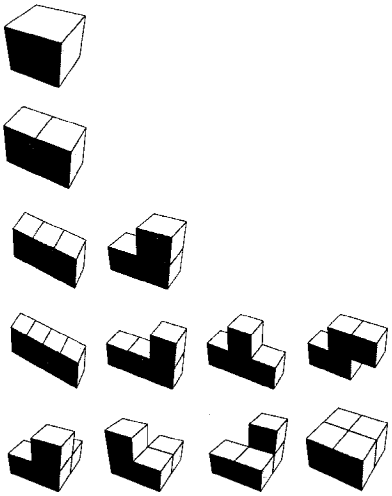

3-D generalization of the POLYOMINOES to n-D. The number of polycubes $N(n)$ composed of n CUBES are 1, 1, 2, 8, 29, 166, 1023, ... (Sloane's A000162, Ball and Coxeter 1987).

see also CONWAY PUZZLE, CUBE DISSECTION, DIABOLICAL CUBE, SLOTHOUBER-GRAATSMA PUZZLE, SOMA CUBE

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 112–113, 1987.

Gardner, M. *The Second Scientific American Book of Mathematical Puzzles & Diversions: A New Selection.* New York: Simon and Schuster, pp. 76–77, 1961.

Gardner, M. "Polycubes." Ch. 3 in *Knotted Doughnuts and Other Mathematical Entertainments.* New York: W. H. Freeman, 1986.

Sloane, N. J. A. Sequence A000162/M1845 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Polydisk

Let $\mathbf{c} = (c_1, \ldots, c_n)$ be a point in $\mathbb{C}^n$, then the open polydisk is defined by

$$S = \{z : |z_j - c_j| < |z_j^0 - c_j|\}$$

for $j = 1, \ldots, n$.

see also DISK, OPEN DISK

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 100, 1980.

Polygamma Function

The polygamma function is sometimes denoted $F_m(z)$, and sometimes $\psi_m(z)$. In $F_m(z)$ notation,

$$F_m(z) \equiv \frac{d^{m+1}}{dz^{m+1}} \ln z! \tag{1}$$

$$= (-1)^{m+1} m! \sum_{n=0}^{\infty} \frac{1}{(z+n)^{m+1}} \tag{2}$$

$$= (-1)^{m+1} m! \zeta(m+1, z), \tag{3}$$

where $\zeta(a,z)$ is the Hurwitz Zeta Function. In the ψ_m Notation (the form returned by the `PolyGamma[m,z]` function in *Mathematica*®; Wolfram Research, Champaign, IL),

$$\begin{aligned}\psi_m(z) &= \frac{d^{m+1}}{dz^{m+1}}\ln[\Gamma(z)]\\ &= \frac{d^m}{dz^m}\frac{\Gamma'(z)}{\Gamma(z)} = \frac{d^m}{dz^m}\Psi(z),\end{aligned} \tag{4}$$

where $\Gamma(z)$ is the Gamma Function and $\Psi(z)$ is the Digamma Function. $\psi_m(z)$ is therefore related to $F_m(z)$ by

$$\psi_m(z) = F_m(z-1). \tag{5}$$

The function $\psi_0(z)$ is equivalent to the Digamma Function $\Psi(z)$. Note that Morse and Feshbach (1953) adopt a notation no longer in standard use in which Morse and Feshbach's $\psi_m(z)$ is equal to the above $\psi_{m-1}(z)$.

The polygamma function obeys the Recurrence Relation

$$\psi_n(z+1) = \psi_n(z) + (-1)^n n! z^{-n-1}, \tag{6}$$

the reflection Formula

$$\psi_n(1-z) + (-1)^{n+1}\psi_n(z) = (-1)^n \pi \frac{d^n}{dz^n}\cot(\pi z), \tag{7}$$

and the multiplication Formula,

$$\psi_n(mz) = \delta_{n0}\ln m + \frac{1}{m^{n+1}}\sum_{k=1}^{m-1}\psi_n\left(z+\frac{k}{m}\right), \tag{8}$$

where δ_{mn} is the Kronecker Delta.

In general, special values for integral indices are given by

$$\psi_n(1) = (-1)^{n+1} n! \zeta(n+1) \tag{9}$$

$$\psi_n(\tfrac{1}{2}) = (-1)^{n+1} n! (2^{n+1}-1)\zeta(n+1), \tag{10}$$

giving

$$\psi_1(\tfrac{1}{2}) = \tfrac{1}{2}\pi^2 \tag{11}$$

$$\psi_1(1) = \zeta(2) = \tfrac{1}{6}\pi^2 \tag{12}$$

$$\psi_2(1) = -2\zeta(3), \tag{13}$$

$$\psi_3(\tfrac{1}{2}) = \pi^4 \tag{14}$$

and so on.

R. Manzoni has shown that the polygamma function can be expressed in terms of Clausen Functions for Rational arguments and integer index. Special cases are given by

$$\psi_1(\tfrac{1}{3}) = \tfrac{2}{3}\pi^2 + \tfrac{3}{2}\sqrt{3}[\mathrm{Cl}_2(\tfrac{2}{3}\pi) - \mathrm{Cl}_2(\tfrac{4}{3}\pi) \tag{15}$$

$$\psi_1(\tfrac{2}{3}) = \tfrac{2}{3}\pi^2 - \tfrac{3}{2}\sqrt{3}[\mathrm{Cl}_2(\tfrac{2}{3}\pi) - \mathrm{Cl}_2(\tfrac{4}{3}\pi) \tag{16}$$

$$\psi_1(\tfrac{1}{4}) = \pi^2 + 4[\mathrm{Cl}_2(\tfrac{1}{2}\pi) - \mathrm{Cl}_2(\tfrac{3}{2}\pi)] \tag{17}$$

$$\psi_1(\tfrac{3}{4}) = \pi^2 - 4[\mathrm{Cl}_2(\tfrac{1}{2}\pi) - \mathrm{Cl}_2(\tfrac{3}{2}\pi)]. \tag{18}$$

$$\psi_2(\tfrac{1}{2}) = -8[\mathrm{Cl}_3(0) - \mathrm{Cl}_3(\pi)]. \tag{19}$$

$$\psi_2(\tfrac{1}{3}) = -\frac{4\pi^3}{3\sqrt{3}} - 18\,\mathrm{Cl}_3(0) + 9[\mathrm{Cl}_3(\tfrac{2}{3}\pi) + \mathrm{Cl}_3(\tfrac{4}{3}\pi)] \tag{20}$$

$$\psi_2(\tfrac{2}{3}) = \frac{4\pi^3}{3\sqrt{3}} - 18\,\mathrm{Cl}_3(0) + 9[\mathrm{Cl}_3(\tfrac{2}{3}\pi) + \mathrm{Cl}_3(\tfrac{4}{3}\pi)] \tag{21}$$

$$\psi_2(\tfrac{1}{4}) = -2\pi^3 - 32[\mathrm{Cl}_3(0) - \mathrm{Cl}_3(\pi)] \tag{22}$$

$$\psi_3(\tfrac{3}{4}) = 2\pi^3 - 32[\mathrm{Cl}_3(0) - \mathrm{Cl}_3(\pi)] \tag{23}$$

$$\psi_3(\tfrac{1}{3}) = \tfrac{8}{3}\pi^4 + 81\sqrt{3}[\mathrm{Cl}_4(\tfrac{2}{3}\pi) - \mathrm{Cl}_4(\tfrac{4}{3}\pi)] \tag{24}$$

$$\psi_3(\tfrac{2}{3}) = \tfrac{8}{3}\pi^4 - 81\sqrt{3}[\mathrm{Cl}_4(\tfrac{2}{3}\pi) - \mathrm{Cl}_4(\tfrac{4}{3}\pi)] \tag{25}$$

$$\psi_3(\tfrac{1}{4}) = 8\pi^4 + 384[\mathrm{Cl}_4(\tfrac{1}{2}\pi) - \mathrm{Cl}_4(\tfrac{3}{2}\pi)] \tag{26}$$

$$\psi_3(\tfrac{3}{4}) = 8\pi^4 - 384[\mathrm{Cl}_4(\tfrac{1}{2}\pi) - \mathrm{Cl}_4(\tfrac{3}{2}\pi)]. \tag{27}$$

see also Clausen Function, Digamma Function, Gamma Function, Stirling's Series

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Polygamma Functions." §6.4 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing*. New York: Dover, p. 260, 1972.

Adamchik, V. S. "Polygamma Functions of Negative Order." Submitted to *J. Symb. Comput.* `http://www.wolfram.com/~victor/articles/polyg.html`.

Arfken, G. "Digamma and Polygamma Functions." §10.2 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 549–555, 1985.

Davis, H. T. *Tables of the Higher Mathematical Functions.* Bloomington, IN: Principia Press, 1933.

Kolbig, V. "The Polygamma Function $\psi_k(x)$ for $x = 1/4$ and $x = 3/4$." *J. Comp. Appl. Math.* **75**, 43–46, 1996.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 422–424, 1953.

Polygenic Function

A function which has infinitely many Derivatives at a point. If a function is not polygenic, it is Monogenic.

see also Monogenic Function

References

Newman, J. R. *The World of Mathematics, Vol. 3.* New York: Simon & Schuster, p. 2003, 1956.

Polygon

A closed plane figure with n sides. If all sides and angles are equivalent, the polygon is called regular. Regular polygons can be Convex or Star. The word derives from the Greek *poly* (many) and *gonu* (knee).

The AREA of a polygon with VERTICES (x_1, y_1), ..., (x_n, y_n) is

$$A = \frac{1}{2}\left(\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} + \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} + \ldots + \begin{vmatrix} x_n & y_n \\ x_1 & y_1 \end{vmatrix} \right), \quad (1)$$

which can be written

$$A = \tfrac{1}{2}(x_1y_2 + x_2y_1 + \ldots + x_{n-1}y_n + x_ny_1 - y_1x_2 - y_2x_3 - \ldots - y_{n+1}x_n - y_nx_1), \quad (2)$$

where the signs can be found from the following diagram.

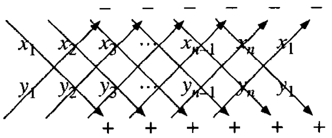
− − − − − −
x_1 x_2 x_3 ... x_{n-1} x_n x_1
y_1 y_2 y_3 ... y_{n-1} y_n y_1
+ + + + + +

The AREA of a polygon is defined to be POSITIVE if the points are arranged in a counterclockwise order, and NEGATIVE if they are in clockwise order (Beyer 1987).

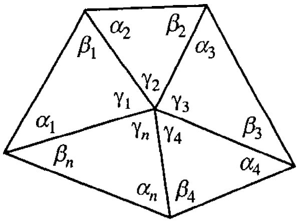

The sum I of internal angles in the above diagram of a dissected PENTAGON is

$$I = \sum_{i=1}^{n}(\alpha_i + \beta_i) = \sum_{i=1}^{n}(\alpha_i + \beta_i + \gamma_i) - \sum_{i=1}^{n}\gamma_i. \quad (3)$$

But

$$\sum_{i=1}^{n}\gamma_i = 360° \quad (4)$$

and the sum of ANGLES of the n TRIANGLES is

$$\sum_{i=1}^{n}(\alpha_i + \beta_i + \gamma_i) = \sum_{i=1}^{n}(180°) = n(180°). \quad (5)$$

Therefore,

$$I = n(180°) - 360° = (n-2)180°. \quad (6)$$

Let n be the number of sides. The *regular* n-gon is then denoted $\{n\}$.

n	$\{n\}$
2	digon
3	equilateral triangle (trigon)
4	square (quadrilateral, tetragon)
5	pentagon
6	hexagon
7	heptagon
8	octagon
9	nonagon (enneagon)
10	decagon
11	undecagon (hendecagon)
12	dodecagon
13	tridecagon (triskaidecagon)
14	tetradecagon (tetrakaidecagon)
15	pentadecagon (pentakaidecagon)
16	hexadecagon (hexakaidecagon)
17	heptadecagon (heptakaidecagon)
18	octadecagon (octakaidecagon)
19	enneadecagon (enneakaidecagon)
20	icosagon
30	triacontagon
40	tetracontagon
50	pentacontagon
60	hexacontagon
70	heptacontagon
80	octacontagon
90	enneacontagon
100	hectogon
10000	myriagon

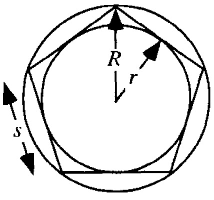

$n = 5$

Let s be the side length, r be the INRADIUS, and R the CIRCUMRADIUS. Then

$$s = 2r\tan\left(\frac{\pi}{n}\right) = 2R\sin\left(\frac{\pi}{n}\right) \quad (7)$$

$$r = \tfrac{1}{2}s\cot\left(\frac{\pi}{n}\right) \quad (8)$$

$$R = \tfrac{1}{2}s\csc\left(\frac{\pi}{n}\right) \quad (9)$$

$$A = \tfrac{1}{4}ns^2\cot\left(\frac{\pi}{n}\right) \quad (10)$$

$$= nr^2\tan\left(\frac{\pi}{n}\right) \quad (11)$$

$$= \tfrac{1}{2}nR^2\sin\left(\frac{2\pi}{n}\right). \quad (12)$$

If the number of sides is doubled, then

$$s_{2n} = \sqrt{2R^2 - R\sqrt{4R^2 - s_n{}^2}} \quad (13)$$

$$A_{2n} = \frac{4rA_n}{2r + \sqrt{4r^2 + s_n{}^2}}. \quad (14)$$

Furthermore, if p_k and P_k are the PERIMETERS of the regular polygons inscribed in and circumscribed around a given CIRCLE and a_k and A_k their areas, then

$$P_{2n} = \frac{2p_nP_n}{p_n + P_n} \tag{15}$$

$$p_{2n} = \sqrt{p_nP_{2n}}, \tag{16}$$

and

$$a_{2n} = \sqrt{a_nA_n} \tag{17}$$

$$A_{2n} = \frac{2a_{2n}A_n}{a_{2n} + A_n} \tag{18}$$

(Beyer 1987, p. 125).

COMPASS and STRAIGHTEDGE constructions dating back to Euclid were capable of inscribing regular polygons of 3, 4, 5, 6, 8, 10, 12, 16, 20, 24, 32, 40, 48, 64, ..., sides. However, this listing is not a complete enumeration of "constructible" polygons. In fact, a regular n-gon is constructible only if $\phi(n)$ is a POWER of 2, where ϕ is the TOTIENT FUNCTION (this is a NECESSARY but not SUFFICIENT condition). More specifically, a regular n-gon ($n \geq 3$) can be constructed by STRAIGHTEDGE and COMPASS (i.e., can have trigonometric functions of its ANGLES expressed in terms of finite SQUARE ROOT extractions) IFF

$$n = 2^k p_1 p_2 \cdots p_s, \tag{19}$$

where k is in INTEGER ≥ 0 and the p_i are distinct FERMAT PRIMES. FERMAT NUMBERS are of the form

$$F_m = 2^{2^m} + 1, \tag{20}$$

where m is an INTEGER ≥ 0. The only known PRIMES of this form are 3, 5, 17, 257, and 65537.

The fact that this condition was SUFFICIENT was first proved by Gauss in 1796 when he was 19 years old, and it relies on the property of IRREDUCIBLE POLYNOMIALS that ROOTS composed of a finite number of SQUARE ROOT extractions exist only if the order of the equation is of the form 2^h. That this condition was also NECESSARY was not explicitly proven by Gauss, and the first proof of this fact is credited to Wantzel (1836).

Constructible values of n for $n < 300$ were given by Gauss (Smith 1994), and the first few are 2, 3, 4, 5, 6, 8, 10, 12, 15, 16, 17, 20, 24, 30, 32, 34, 40, 48, 51, 60, 64, 68, 80, 85, 96, 102, 120, 128, 136, 160, 170, 192, ... (Sloane's A003401). Gardner (1977) and independently Watkins (Conway and Guy 1996) noticed that the number of sides for constructible polygons with an ODD number of sides is given by the first 32 rows of PASCAL'S TRIANGLE (mod 2) interpreted as BINARY numbers, giving 1, 3, 5, 15, 17, 51, 85, 255, ... (Sloane's A004729, Conway and Guy 1996, p. 140).

1	1	1
1 1	1 1	3
1 2 1	1 0 1	5
1 3 3 1	1 1 1 1	15
1 4 6 4 1	1 0 0 0 1	17
1 5 10 10 5 1	1 1 0 0 1 1	51
1 6 15 20 15 6 1	1 0 1 0 1 0 1	85
1 7 21 35 35 21 7 1	1 1 1 1 1 1 1 1	255
1 8 28 56 70 56 28 8 1	1 0 0 0 0 0 0 0 1	257

Although constructions for the regular TRIANGLE, SQUARE, PENTAGON, and their derivatives had been given by Euclid, constructions based on the FERMAT PRIMES $\geq$ 17 were unknown to the ancients. The first explicit construction of a HEPTADECAGON (17-gon) was given by Erchinger in about 1800. Richelot and Schwendenwein found constructions for the 257-GON in 1832, and Hermes spent 10 years on the construction of the 65537-GON at Göttingen around 1900 (Coxeter 1969). Constructions for the EQUILATERAL TRIANGLE and SQUARE are trivial (top figures below). Elegant constructions for the PENTAGON and HEPTADECAGON are due to Richmond (1893) (bottom figures below).

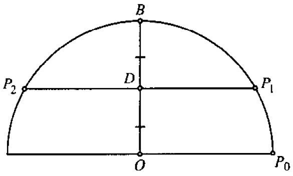

Equilateral Triangle

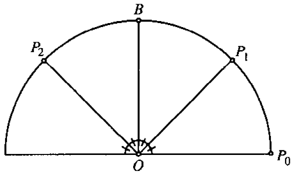

Square

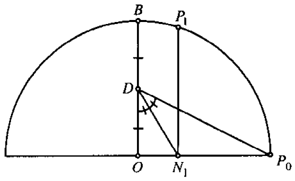

Pentagon

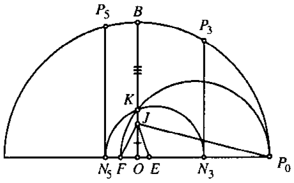

17-gon

Given a point, a CIRCLE may be constructed of any desired RADIUS, and a DIAMETER drawn through the center. Call the center O, and the right end of the DIAMETER P_0. The DIAMETER PERPENDICULAR to the original DIAMETER may be constructed by finding the PERPENDICULAR BISECTOR. Call the upper endpoint of this PERPENDICULAR DIAMETER B. For the PENTAGON, find the MIDPOINT of OB and call it D. Draw DP_0, and BISECT $\angle ODP_0$, calling the intersection point with OP_0 N_1. Draw N_1P_1 PARALLEL to OB, and the first two points of the PENTAGON are P_0 and P_1. The construction for the HEPTADECAGON is more complicated, but can be accomplished in 17 relatively simple steps. The construction problem has now been automated (Bishop 1978).

see also 257-GON, 65537-GON, ANTHROPOMORPHIC POLYGON, BICENTRIC POLYGON, CARNOT'S POLYGON THEOREM, CHAOS GAME, CONVEX POLYGON, CYCLIC POLYGON, DE MOIVRE NUMBER, DIAGONAL (POLYGON), EQUILATERAL TRIANGLE, EULER'S POLYGON DIVISION PROBLEM, HEPTADECAGON, HEXAGON, Hexagram, Illumination Problem, Jordan Polygon, Lozenge, Octagon, Parallelogram, Pascal's Theorem, Pentagon, Pentagram, Petrie Polygon, Polygon Circumscribing Constant, Polygon Inscribing Constant, Polygonal Knot, Polygonal Number, Polygonal Spiral, Polygon Triangulation, Polygram, Polyhedral Formula, Polyhedron, Polytope, Quadrangle, Quadrilateral, Regular Polygon, Reuleaux Polygon, Rhombus, Rotor, Simple Polygon, Simplicity, Square, Star Polygon, Trapezium, Trapezoid, Triangle, Visibility, Voronoi Polygon, Wallace-Bolyai-Gerwein Theorem

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 124–125 and 196, 1987.

Bishop, W. "How to Construct a Regular Polygon." *Amer. Math. Monthly* **85**, 186–188, 1978.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 140 and 197–202, 1996.

Courant, R. and Robbins, H. "Regular Polygons." §3.2 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 122–125, 1996.

Coxeter, H. S.M. *Introduction to Geometry, 2nd ed.* New York: Wiley, 1969.

De Temple, D. W. "Carlyle Circles and the Lemoine Simplicity of Polygonal Constructions." *Amer. Math. Monthly* **98**, 97–108, 1991.

Gardner, M. *Mathematical Carnival: A New Round-Up of Tantalizers and Puzzles from Scientific American.* New York: Vintage Books, p. 207, 1977.

Gauss, C. F. §365 and 366 in *Disquisitiones Arithmeticae.* Leipzig, Germany, 1801. Translated by A. A Clarke. New Haven, CT: Yale University Press, 1965.

The Math Forum. "Naming Polygons and Polyhedra." `http://forum.swarthmore.edu/dr.math/faq/faq.polygon.names.html`.

Rawles, B. *Sacred Geometry Design Sourcebook: Universal Dimensional Patterns.* Nevada City, CA: Elysian Pub., p. 238, 1997.

Richmond, H. W. "A Construction for a Regular Polygon of Seventeen Sides." *Quart. J. Pure Appl. Math.* **26**, 206–207, 1893.

Sloane, N. J. A. Sequences A004729 and A003401/M0505 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Smith, D. E. *A Source Book in Mathematics.* New York: Dover, p. 350, 1994.

Tietze, H. Ch. 9 in *Famous Problems of Mathematics.* New York: Graylock Press, 1965.

Wantzel, P. L. "Recherches sur les moyens de reconnaître si un Problème de Géométrie peut se résoudre avec la règle et le compas." *J. Math. pures appliq.* **1**, 366–372, 1836.

Polygon Circumscribing Constant

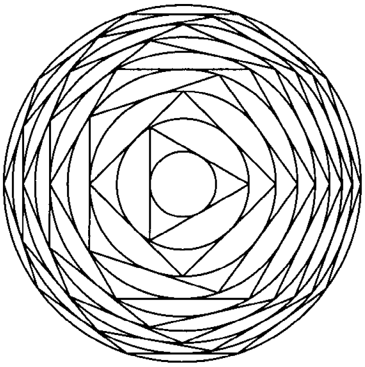

If a Triangle is Circumscribed about a Circle, another Circle around the Triangle, a Square outside the Circle, another Circle outside the Square, and so on. From Polygons, the Circumradius and Inradius for an n-gon are

$$R = \tfrac{1}{2} s \csc\left(\frac{\pi}{n}\right) \tag{1}$$

$$r = \tfrac{1}{2} s \cot\left(\frac{\pi}{n}\right), \tag{2}$$

where s is the side length. Therefore,

$$\frac{R}{r} = \frac{1}{\cos\left(\frac{\pi}{n}\right)} = \sec\left(\frac{\pi}{n}\right), \tag{3}$$

and an infinitely nested set of circumscribed polygons and circles has

$$K \equiv \frac{r_{\text{final circle}}}{r_{\text{initial circle}}} = \sec\left(\frac{\pi}{3}\right)\sec\left(\frac{\pi}{4}\right)\sec\left(\frac{\pi}{5}\right)\cdots. \tag{4}$$

Kasner and Newman (1989) and Haber (1964) state that $K = 12$, but this is incorrect. Write

$$K = \prod_{n=3}^{\infty} \frac{1}{\cos\left(\frac{\pi}{n}\right)} \tag{5}$$

$$\ln K = -\sum_{n=3}^{\infty} \ln(\cos x). \tag{6}$$

Define

$$y_0(x) \equiv -\ln(\cos x) = \tfrac{1}{2}x^2 + \tfrac{1}{12}x^4 + \tfrac{1}{45}x^6 + \tfrac{17}{2520}x^8 + \ldots. \tag{7}$$

Now define

$$y_1(x) = \tfrac{1}{2}ax^2, \tag{8}$$

with

$$y_1(\tfrac{\pi}{3}) = y_0(\tfrac{\pi}{3}) \tag{9}$$

$$\tfrac{1}{2}a\left(\frac{\pi}{3}\right)^2 = \ln 2, \tag{10}$$

so

$$a = 2\left(\frac{3}{\pi}\right)^2 \ln 2, \tag{11}$$

and

$$y_2(x) = \frac{9\ln 2}{\pi^2}x^2. \tag{12}$$

But $y_2(x) > y_1(x)$ for $x \in (0, \pi/3)$, so

$$\sum_{n=3}^{\infty} y_2\left(\frac{\pi}{n}\right) > -\sum_{n=3}^{\infty} \ln\left[\cos\left(\frac{\pi}{n}\right)\right] \tag{13}$$

$$\begin{aligned}
\ln K &< \sum_{n=3}^{\infty} y_2\left(\frac{\pi}{n}\right) \frac{9\ln 2}{\pi^2}\sum_{n=3}^{\infty}\left(\frac{\pi}{n}\right)^2 = 9\ln 2\sum_{n=3}^{\infty}\frac{1}{n^2} \\
&= 9\ln 2\left(\sum_{n=1}^{\infty}\frac{1}{n^2} - \sum_{n=1}^{2}\frac{1}{n^2}\right) = 9\ln 2[\zeta(2) - \tfrac{5}{4}] \\
&= 9\ln 2\left(\frac{\pi^2}{6} - \frac{5}{4}\right) = 2.4637
\end{aligned} \tag{14}$$

$$K < e^{2.4637} = 11.75. \tag{15}$$

If the next term is included,

$$y_2(x) = a(\tfrac{1}{2}x^2 + \tfrac{1}{12}x^4). \tag{16}$$

As before,

$$y_2(\tfrac{\pi}{3}) = y_0(\tfrac{\pi}{3}) \tag{17}$$

$$a = \frac{972\ln 2}{\pi^2(54+\pi^2)}, \tag{18}$$

so

$$y_2(x) = \frac{972\ln 2}{\pi^2(54+\pi^2)}(\tfrac{1}{2}x^2 + \tfrac{1}{12}x^4) \tag{19}$$

$$\begin{aligned}
\ln K &< \frac{972\ln 2}{\pi^2(54+\pi^2)}\sum_{n=3}^{\infty}\left[\frac{1}{2}\left(\frac{\pi}{n}\right)^2 + \frac{1}{12}\left(\frac{\pi}{n}\right)^4\right] \\
&= \frac{972\ln 2}{\pi^2(54+\pi^2)}\left\{\frac{1}{2}\left[\zeta(2) - \frac{5}{4}\right] + \frac{\pi^2}{12}\left[\zeta(4) - 1 - \frac{1}{2^4}\right]\right\} \\
&= \frac{972\ln 2}{\pi^2(54+\pi^2)}\left[\frac{1}{2}\left(\frac{\pi^2}{6} - \frac{5}{4}\right) + \frac{\pi^2}{12}\left(\frac{\pi^4}{90} - 1 - \frac{1}{2^4}\right)\right] \\
&= \frac{9(8\pi^6 - 45\pi^2 - 5400)\ln 2}{80(\pi^2+54)} = 2.255,
\end{aligned} \tag{20}$$

and

$$K < e^{2.255} = 9.535. \tag{21}$$

The process can be automated using computer algebra, and the first few bounds are 11.7485, 9.53528, 8.98034, 8.8016, 8.73832, 8.71483, 8.70585, 8.70235, 8.70097, and 8.70042. In order to obtain this accuracy by direct multiplication of the terms, more than 10,000 terms are needed. The limit is

$$K = 8.700036625\ldots. \tag{22}$$

Bouwkamp (1965) produced the following INFINITE PRODUCT formulas

$$K = \frac{2}{\pi}\prod_{m=1}^{\infty}\prod_{n=1}^{\infty}\left[1 - \frac{1}{m^2(n+\frac{1}{2})^2}\right] \tag{23}$$

$$= 6\exp\left\{\sum_{k=1}^{\infty}\frac{[\lambda(2k)-1]2^{2k}[\zeta(2k)-1-2^{-2k}]}{k}\right\}, \tag{24}$$

where $\zeta(x)$ is the RIEMANN ZETA FUNCTION and $\lambda(x)$ is the DIRICHLET LAMBDA FUNCTION. Bouwkamp (1965) also produced the formula with accelerated convergence

$$K = \tfrac{1}{12}\sqrt{6}\pi^4(1 - \tfrac{1}{2}\pi^2 + \tfrac{1}{24}\pi^4)(1 - \tfrac{1}{8}\pi^2 + \tfrac{1}{384}\pi^4) \times \csc\left(\frac{\pi^2}{\sqrt{6+2\sqrt{3}}}\right)\csc\left(\frac{\pi^2}{\sqrt{6-2\sqrt{3}}}\right)B, \tag{25}$$

where

$$B \equiv \prod_{n=3}^{\infty}\left(1 - \frac{\pi^2}{2n^2} + \frac{\pi^4}{24n^4}\right)\sec\left(\frac{\pi}{n}\right) \tag{26}$$

(cited in Pickover 1995).

see also POLYGON INSCRIBING CONSTANT

References

Bouwkamp, C. "An Infinite Product." *Indag. Math.* **27**, 40–46, 1965.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/infprd/infprd.html`.

Haber, H. "Das Mathematische Kabinett." *Bild der Wissenschaft* **2**, 73, Apr. 1964.

Kasner, E. and Newman, J. R. *Mathematics and the Imagination.* Redmond, WA: Microsoft Press, pp. 311–312, 1989.

Pappas, T. "Infinity & Limits." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 180, 1989.

Pickover, C. A. "Infinitely Exploding Circles." Ch. 18 in *Keys to Infinity.* New York: W. H. Freeman, pp. 147–151, 1995.

Pinkham, R. S. "Mathematics and Modern Technology." *Amer. Math. Monthly* **103**, 539–545, 1996.

Plouffe, S. "Product(cos(Pi/n),n=3..infinity)." `http://lacim.uqam.ca/piDATA/productcos.txt`.

Polygon Construction

see GEOMETRIC CONSTRUCTION, GEOMETROGRAPHY, POLYGON, SIMPLICITY

Polygon Division Problem

see EULER'S POLYGON DIVISION PROBLEM

Polygon Fractal

see CHAOS GAME

Polygon Inscribing Constant

If a TRIANGLE is inscribed in a CIRCLE, another CIRCLE inside the TRIANGLE, a SQUARE inside the CIRCLE, another CIRCLE inside the SQUARE, and so on,

$$K' \equiv \frac{r_{\text{final circle}}}{r_{\text{initial circle}}} = \cos\left(\frac{\pi}{3}\right)\cos\left(\frac{\pi}{4}\right)\cos\left(\frac{\pi}{5}\right)\cdots.$$

Numerically,

$$K' = \frac{1}{K} = \frac{1}{8.7000366252\ldots} = 0.1149420448\ldots,$$

where K is the POLYGON CIRCUMSCRIBING CONSTANT. Kasner and Newman's (1989) assertion that $K = 1/12$ is incorrect.

Let a convex POLYGON be inscribed in a CIRCLE and divided into TRIANGLES from diagonals from one VERTEX. The sum of the RADII of the CIRCLES inscribed in these TRIANGLES is the same independent of the VERTEX chosen (Johnson 1929, p. 193).

see also POLYGON CIRCUMSCRIBING CONSTANT

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/infprd/infprd.html`.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, 1929.

Kasner, E. and Newman, J. R. *Mathematics and the Imagination.* Redmond, WA: Microsoft Press, pp. 311–312, 1989.

Pappas, T. "Infinity & Limits." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 180, 1989.

Plouffe, S. "Product(cos(Pi/n),n=3..infinity)." `http://lacim.uqam.ca/piDATA/productcos.txt`.

Polygon Triangulation

see EULER'S POLYGON DIVISION PROBLEM

Polygonal Knot

A KNOT equivalent to a POLYGON in $\mathbb{R}^3$, also called a TAME KNOT. For a polygonal knot K, there exists a PLANE such that the orthogonal projection π on it satisfies the following conditions:

1. The image $\pi(K)$ has no multiple points other than a FINITE number of double points.
2. The projections of the vertices of K are not double points of $\pi(K)$.

Such a projection $\pi(K)$ is called a regular knot projection.

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 735, 1980.

Polygonal Number

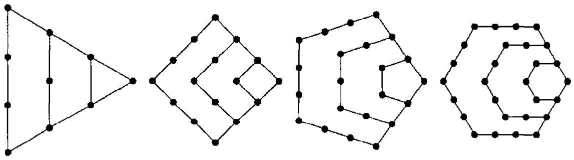

A type of FIGURATE NUMBER which is a generalization of TRIANGULAR, SQUARE, etc., numbers to an arbitrary n-gonal number. The above diagrams graphically illustrate the process by which the polygonal numbers are built up. Starting with the nth TRIANGULAR NUMBER T_n, then

$$n + T_{n-1} = T_n. \tag{1}$$

Now note that

$$n + 2T_{n-1} = n^2 = S_n \tag{2}$$

gives the nth SQUARE NUMBER,

$$n + 3T_{n-1} = \tfrac{1}{2}n(3n-1) = P_n, \tag{3}$$

gives the nth PENTAGONAL NUMBER, and so on. The general polygonal number can be written in the form

$$p_r^n = \tfrac{1}{2}r[(r-1)n-2(r-2)] = \tfrac{1}{2}r[(n-2)r-(n-4)], \tag{4}$$

where p_r^n is the rth n-gonal number. For example, taking $n = 3$ in (4) gives a TRIANGULAR NUMBER, $n = 4$ gives a SQUARE NUMBER, etc.

Fermat proposed that every number is expressible as *at most k k*-gonal numbers (FERMAT'S POLYGONAL NUMBER THEOREM). Fermat claimed to have a proof of this result, although this proof has never been found. Jacobi, Lagrange (1772), and Euler all proved the square case, and Gauss proved the triangular case in 1796. In 1813, Cauchy proved the proposition in its entirety.

An arbitrary number N can be checked to see if it is a n-gonal number as follows. Note the identity

$$\begin{aligned} 8(n-2)p_n^r + (n-4)^2 &= 4r(n-2)[(r-1)n-2(r-2)] \\ &\quad +(n-4)^2 = 4r(r-1)n^2 + r[-8(r-1)-8(r-2)]n \\ &\quad +16r(r-2) + (n^2-8n+16) \\ &= (4r^2-4r+1)n^2 + (-16r^2+24r-8)n \\ &\quad +(16r^2-32r+16) \\ &= (2r-1)^2n^2 - 8(2r^2-3r+1)n + 16(r^2-2r+1) \\ &= (2rn-4r-n+4)^2, \end{aligned} \tag{5}$$

so $8(n-2)N + (n-4)^2 = S^2$ must be a PERFECT SQUARE. Therefore, if it is not, the number cannot be n-gonal. If it is a PERFECT SQUARE, then solving

$$S = 2rn - 4r - n + 4 \tag{6}$$

for the rank r gives

$$r = \frac{S+n-4}{2(n-2)}. \tag{7}$$

An n-gonal number is equal to the sum of the $(n-1)$-gonal number of the same RANK and the TRIANGULAR NUMBER of the previous RANK.

see also CENTERED POLYGONAL NUMBER, DECAGONAL NUMBER, FERMAT'S POLYGONAL NUMBER THEOREM, FIGURATE NUMBER, HEPTAGONAL NUMBER, HEXAGONAL NUMBER, NONAGONAL NUMBER, OCTAGONAL NUMBER, PENTAGONAL NUMBER, PYRAMIDAL NUMBER, SQUARE NUMBER, TRIANGULAR NUMBER

References

Beiler, A. H. "Ball Games." Ch. 18 in *Recreations in the Theory of Numbers: The Queen of Mathematics Entertains.* New York: Dover, pp. 184–199, 1966.

Dickson, L. E. *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, pp. 3–33, 1952.

Guy, K. "Every Number is Expressible as a Sum of How Many Polygonal Numbers?" *Amer. Math. Monthly* **101**, 169–172, 1994.

Pappas, T. "Triangular, Square & Pentagonal Numbers." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 214, 1989.

Sloane, N. J. A. Sequences A000217/M2535 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sloane, N. J. A. and Plouffe, S. Extended entry in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.

Polygonal Spiral

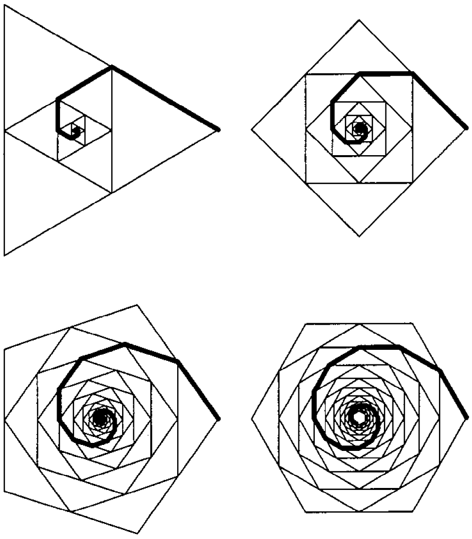

The length of the polygonal spiral is found by noting that the ratio of INRADIUS to CIRCUMRADIUS of a regular POLYGON of n sides is

$$\frac{r}{R} = \frac{\cot\left(\frac{\pi}{n}\right)}{\csc\left(\frac{\pi}{n}\right)} = \cos\left(\frac{\pi}{n}\right). \tag{1}$$

The total length of the spiral for an n-gon with side length s is therefore

$$L = \tfrac{1}{2}s\sum_{k=0}^{\infty}\cos^k\left(\frac{\pi}{n}\right) = \frac{s}{2\left[1-\cos\left(\frac{\pi}{n}\right)\right]}. \tag{2}$$

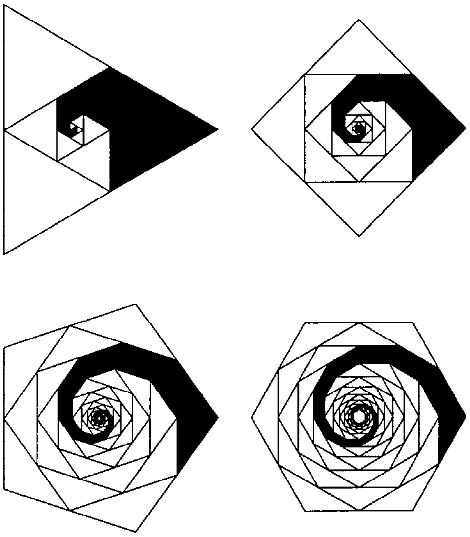

Consider the solid region obtained by filling in subsequent triangles which the spiral encloses. The AREA of this region, illustrated above for n-gons of side length s, is

$$A = \tfrac{1}{4}s^2\cot\left(\frac{\pi}{n}\right). \tag{3}$$

References

Sandefur, J. T. "Using Self-Similarity to Find Length, Area, and Dimension." *Amer. Math. Monthly* **103**, 107–120, 1996.

Polygram

A self-intersecting STAR FIGURE such as the PENTAGRAM or HEXAGRAM.

n	symbol	polygram
5	{5/2}	pentagram
6	{6/2}	hexagram
7	{7/2}	heptagram
8	{8/3}	octagram
	{8/4}	star of Lakshmi
10	{10/3}	decagram

Polyhedral Formula

A formula relating the number of VERTICES, FACES, and EDGES of a POLYHEDRON (or POLYGON). It was discovered independently by Euler and Descartes, so it is also known as the DESCARTES-EULER POLYHEDRAL FORMULA. The polyhedron need not be CONVEX, but the FORMULA does not hold for STELLATED POLYHEDRA.

$$V+F-E=2, \tag{1}$$

where $V = N_0$ is the number of VERTICES, $E = N_1$ is the number of EDGES, and $F = N_2$ is the number of FACES. For a proof, see Courant and Robbins (1978, pp. 239–240). The FORMULA can be generalized to n-D POLYTOPES.

$$\Pi_1 : N_0 = 2 \tag{2}$$

$$\Pi_2 : N_0 - N_1 = 0 \tag{3}$$

$$\Pi_3 : N_0 - N_1 + N_2 = 2 \tag{4}$$

$$\Pi_4 : N_0 - N_1 + N_2 - N_3 = 0 \tag{5}$$

$$\Pi_n : N_0 - N_1 + N_2 - \ldots + (-1)^{n-1} N_{n-1} = 1 - (-1)^n. \tag{6}$$

For a proof of this, see Coxeter (1973, pp. 166–171).

see also DEHN INVARIANT, DESCARTES TOTAL ANGULAR DEFECT

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 128, 1987.

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods.* Oxford, England: Oxford University Press, 1978.

Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, 1973.

Polyhedral Graph

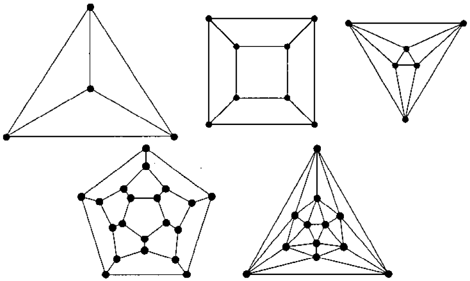

The graphs corresponding to the skeletons of PLATONIC SOLIDS. They are special cases of SCHLEGEL GRAPHS.

see also CUBICAL GRAPH, DODECAHEDRAL GRAPH, ICOSAHEDRAL GRAPH, OCTAHEDRAL GRAPH, SCHLEGEL GRAPH, TETRAHEDRAL GRAPH

Polyhedron

A 3-D solid which consists of a collection of POLYGONS, usually joined at their EDGES. The word derives from the Greek *poly* (many) plus the Indo-European *hedron* (seat). A polyhedron is the 3-D version of the more general POLYTOPE, which can be defined on arbitrary dimensions.

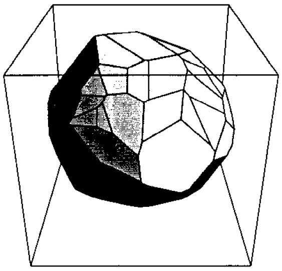

A CONVEX POLYHEDRON can be defined as the set of solutions to a system of linear inequalities

$$\mathbf{mx} \leq \mathbf{b},$$

where $\mathbf{m}$ is a real $s \times 3$ MATRIX and $\mathbf{b}$ is a real s-VECTOR. An example is illustrated above. The more simple DODECAHEDRON is given by a system with $s = 12$. In general, given the MATRICES, the VERTICES (and FACES) can be found using VERTEX ENUMERATION.

A polyhedron is said to be regular if its FACES and VERTEX FIGURES are REGULAR (not necessarily CONVEX) polygons (Coxeter 1973, p. 16). Using this definition, there are a total of nine REGULAR POLYHEDRA, five being the CONVEX PLATONIC SOLIDS and four being the CONCAVE (stellated) KEPLER-POINSOT SOLIDS. However, the term "regular polyhedra" is sometimes also used to refer exclusively to the PLATONIC SOLIDS (Cromwell 1997, p. 53). The DUAL POLYHEDRA of the PLATONIC SOLIDS are not new polyhedra, but are themselves PLATONIC SOLIDS.

A CONVEX polyhedron is called SEMIREGULAR if its FACES have a similar arrangement of nonintersecting regular plane CONVEX polygons of two or more different types about each VERTEX (Holden 1991, p. 41). These solids are more commonly called the ARCHIMEDEAN SOLIDS, and there are 13 of them. The DUAL POLYHEDRA of the ARCHIMEDEAN SOLIDS are 13 new (and beautiful) solids, sometimes called the CATALAN SOLIDS.

A QUASIREGULAR POLYHEDRON is the solid region interior to two DUAL REGULAR POLYHEDRA (Coxeter 1973, pp. 17–20). There are only two CONVEX QUASIREGULAR POLYHEDRA: the CUBOCTAHEDRON and ICOSIDODECAHEDRON. There are also infinite families of PRISMS and ANTIPRISMS.

There exist exactly 92 CONVEX POLYHEDRA with REGULAR POLYGONAL faces (and not necessary equivalent vertices). They are known as the JOHNSON SOLIDS. Polyhedra with identical VERTICES related by a symmetry operation are known as UNIFORM POLYHEDRA. There are 75 such polyhedra in which only two faces may meet at an EDGE, and 76 in which any EVEN number of faces may meet. Of these, 37 were discovered by Badoureau in 1881 and 12 by Coxeter and Miller ca. 1930.

Polyhedra can be superposed on each other (with the sides allowed to pass through each other) to yield additional POLYHEDRON COMPOUNDS. Those made from REGULAR POLYHEDRA have symmetries which are especially aesthetically pleasing. The graphs corresponding to polyhedra skeletons are called SCHLEGEL GRAPHS.

Behnke *et al.* (1974) have determined the symmetry groups of all polyhedra symmetric with respect to their VERTICES.

see also ACOPTIC POLYHEDRON, APEIROGON, ARCHIMEDEAN SOLID, CANONICAL POLYHEDRON, CATALAN SOLID, CUBE, DICE, DIGON, DODECAHEDRON, DUAL POLYHEDRON, ECHIDNAHEDRON, FLEXIBLE POLYHEDRON, HEXAHEDRON, HYPERBOLIC POLYHEDRON, ICOSAHEDRON, ISOHEDRON, JOHNSON SOLID, KEPLER-POINSOT SOLID, NOLID, OCTAHEDRON, PETRIE POLYGON, PLATONIC SOLID, POLYHEDRON COLORING, POLYHEDRON COMPOUND, PRISMATOID, QUADRICORN, QUASIREGULAR POLYHEDRON, RIGIDITY THEOREM, SEMIREGULAR POLYHEDRON, SKELETON, TETRAHEDRON, UNIFORM POLYHEDRON, ZONOHEDRON

References

Ball, W. W. R. and Coxeter, H. S. M. "Polyhedra." Ch. 5 in *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 130–161, 1987.

Behnke, H.; Bachman, F.; Fladt, K.; and Kunle, H. (Eds.). *Fundamentals of Mathematics, Vol. 2.* Cambridge, MA: MIT Press, 1974.

Bulatov, V. "Polyhedra Collection." `http://www.physics.orst.edu/~bulatov/polyhedra/`.

Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, 1973.

Critchlow, K. *Order in Space: A Design Source Book.* New York: Viking Press, 1970.

Cromwell, P. R. *Polyhedra.* New York: Cambridge University Press, 1997.

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., 1989.

Davie, T. "Books and Articles about Polyhedra and Polytopes." `http://www.dcs.st-andrews.ac.uk/~ad/mathrecs/polyhedra/polyhedrabooks.html`.

Davie, T. "The Regular (Platonic) and Semi-Regular (Archimedean) Solids." `http://www.dcs.st-andrews.ac.uk/~ad/mathrecs/polyhedra/polyhedratopic.html`.

Eppstein, D. "Geometric Models." `http://www.ics.uci.edu/~eppstein/junkyard/model.html`.

Eppstein, D. "Polyhedra and Polytopes." `http://www.ics.uci.edu/~eppstein/junkyard/polytope.html`.

Hart, G. W. "Virtual Polyhedra." `http://www.li.net/~george/virtual-polyhedra/vp.html`.

Hilton, P. and Pedersen, J. *Build Your Own Polyhedra.* Reading, MA: Addison-Wesley, 1994.

Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, 1991.

Lyusternik, L. A. *Convex Figures and Polyhedra.* New York: Dover, 1963.

Malkevitch, J. "Milestones in the History of Polyhedra." In *Shaping Space: A Polyhedral Approach* (Ed. M. Senechal and G. Fleck). Boston, MA: Birkhäuser, pp. 80–92, 1988.

Miyazaki, K. *An Adventure in Multidimensional Space: The Art and Geometry of Polygons, Polyhedra, and Polytopes.* New York: Wiley, 1983.

Paeth, A. W. "Exact Dihedral Metrics for Common Polyhedra." In *Graphic Gems II* (Ed. J. Arvo). New York: Academic Press, 1991.

Pappas, T. "Crystals–Nature's Polyhedra." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 38–39, 1989.

Pugh, A. *Polyhedra: A Visual Approach.* Berkeley: University of California Press, 1976.

Schaaf, W. L. "Regular Polygons and Polyhedra." Ch. 3, §4 in *A Bibliography of Recreational Mathematics.* Washington, DC: National Council of Teachers of Math., pp. 57–60, 1978.

Virtual Image. "Polytopia I" and "Polytopia II" CD-ROMs. `http://ourworld.compuserve.com/homepages/vir_image/html/polytopiai.html` and `polytopiaii.html`.

Polyhedron Coloring

Define a valid "coloring" to occur when no two faces with a common EDGE share the same color. Given two colors, there is a single way to color an OCTAHEDRON. Given three colors, there is one way to color a CUBE and 144 ways to color an ICOSAHEDRON. Given four-colors, there are two distinct ways to color a TETRAHEDRON and 4 ways to color a DODECAHEDRON. Given five colors, there are four ways to color an ICOSAHEDRON.

see also COLORING, POLYHEDRON

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, 238–242, 1987.

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 82–83, 1989.

Polyhedron Compound

Solid	Vertices	Symbol
cube-octahedron	both	
dodec.+icos.	both	
two cubes		
three cubes		
four cubes		
five cubes	dodecahedron	$2\{5,3\}[5\{4,3\}]$
five octahedra	icosidodeca.	$[5\{3,4\}]2\{3,5\}$
five tetrahedra	dodecahedron	$\{5,3\}[5\{3,3\}]2\{3,5\}$
two dodecahedra	both	
great dodecahedron-small stellated dodec.		
great icosahedron-great stellated dodec.	both	
stella octangula	cube	$\{4,3\}[2\{3,3\}]\{3,4\}$
ten tetrahedra	dodecahedron	$2\{5,3\}[10\{3,3\}]2\{3,5\}$

The above table gives some common polyhedron compounds. In Coxeter's NOTATION, d distinct VERTICES of $\{m,n\}$ taken c times are denoted

$$c\{m,n\}[d\{p,q\}],$$

or faces of $\{s,t\}$ e times

$$[d\{p,q\}]e\{s,t\},$$

or both

$$c\{m,n\}[d\{p,q\}]e\{s,t\}.$$

The five TETRAHEDRA can be arranged in a laevo or dextro configuration.

see also CUBE-OCTAHEDRON COMPOUND, DODECAHEDRON-ICOSAHEDRON COMPOUND, OCTAHEDRON 5-COMPOUND, STELLA OCTANGULA, TETRAHEDRON 5-COMPOUND

Polyhedron Dissection

A DISSECTION of one or more polyhedra into other shapes.

see also CUBE DISSECTION, DIABOLICAL CUBE, POLYCUBE, SOMA CUBE, WALLACE-BOLYAI-GERWEIN THEOREM

References

Bulatov, V.v "Compounds of Uniform Polyhedra." http://www.physics.orst.edu/~bulatov/polyhedra/uniform_compounds/.

Coffin, S. T. *The Puzzling World of Polyhedral Dissections.* New York: Oxford University Press, 1990.

Polyhedron Dual

see DUAL POLYHEDRON

Polyhedron Hinging

see RIGIDITY THEOREM

Polyhedron Packing

see KELVIN'S CONJECTURE, SPACE-FILLING POLYHEDRON

Polyhex

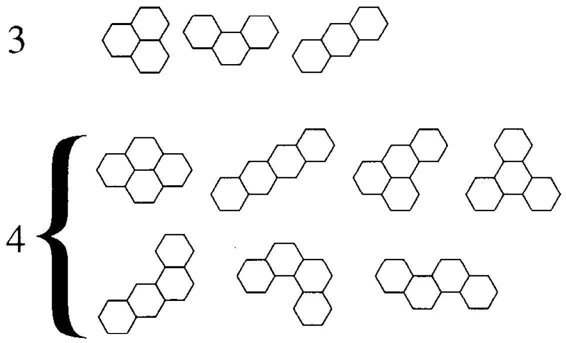

An analog of the POLYOMINOES and POLYIAMONDS in which collections of regular hexagons are arranged with adjacent sides. They are also called HEXES and HEXAS. The number of polyhexes of n hexagons are 1, 1, 2, 7, 22, 82, 333, 1448, 6572, 30490, 143552, 683101, ... (Sloane's A014558). For the 4-hexes (tetrahexes), the possible arrangements are known as the BEE, BAR, PISTOL, PROPELLER, WORM, ARCH, and WAVE.

References

Gardner, M. "Polyhexes and Polyaboloes." Ch. 11 in *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, pp. 146–159, 1978.

Gardner, M. "Tiling with Polyominoes, Polyiamonds, and Polyhexes." Ch. 14 in *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, pp. 175–187, 1988.

Golomb, S. W. *Polyominoes: Puzzles, Patterns, Problems, and Packings, 2nd ed.* Princeton, NJ: Princeton University Press, pp. 92–93, 1994.

Sloane, N. J. A. Sequence A014558 in "An On-Line Version of the Encyclopedia of Integer Sequences."

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 342–343, 1993.

Polyiamond

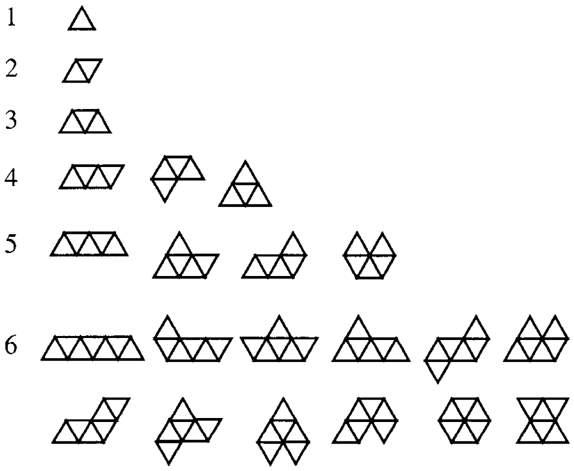

A generalization of the POLYOMINOES using a collection of equal-sized EQUILATERAL TRIANGLES (instead of SQUARES) arranged with coincident sides. Polyiamonds are sometimes simply known as IAMONDS.

The number of two-sided (i.e., can be picked up and flipped, so MIRROR IMAGE pieces are considered identical) polyiamonds made up of n triangles are 1, 1, 1, 3, 4, 12, 24, 66, 160, 448, ... (Sloane's A000577). The number of one-sided polyiamonds composed of n triangles are 1, 1, 1, 4, 6, 19, 43, 121, ... (Sloane's A006534). No HOLES are possible with fewer than seven triangles.

The top row of 6-polyiamonds in the above figure are known as the BAR, CROOK, CROWN, SPHINX, SNAKE, and YACHT. The bottom row of 6-polyiamonds are known as the CHEVRON, SIGNPOST, LOBSTER, HOOK, HEXAGON, and BUTTERFLY.

see also POLYABOLO, POLYHEX, POLYOMINO

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Gardner, M. "Mathematical Games." *Sci. Amer.*, Dec. 1964.

Golomb, S. W. *Polyominoes: Puzzles, Patterns, Problems, and Packings, 2nd ed.* Princeton, NJ: Princeton University Press, pp. 90–92, 1994.

Sloane, N. J. A. Sequences A000577/M2374 and A006534/M3287 in "An On-Line Version of the Encyclopedia of Integer Sequences."

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 342–343, 1993.

Polyking

see POLYPLET

Polylogarithm

The function

$$\mathrm{Li}_n(z) \equiv \sum_{k=1}^{\infty} \frac{z^k}{k^n}, \tag{1}$$

Also known as JONQUIÈRE'S FUNCTION. (Note that the NOTATION $\mathrm{Li}(z)$ is also used for the LOGARITHMIC INTEGRAL.) The polylogarithm arises in Feynman Diagram integrals, and the special case $n = 2$ is called the DILOGARITHM. The polylogarithm of NEGATIVE INTEGER order arises in sums of the form

$$\sum_{k=1}^{\infty} k^n r^k = \mathrm{Li}_{-n}(r) = \frac{r}{(1-r)^{n+1}} \sum_{i=1}^{n} \left\langle {n \atop i} \right\rangle r^{n-i}, \tag{2}$$

where $\left\langle {n \atop i} \right\rangle$ is an EULERIAN NUMBER.

The polylogarithm satisfies the fundamental identities

$$-\ln(1-2^{-n}) = \mathrm{Li}_1(2^{-n}) \tag{3}$$

$$\mathrm{Li}_s(-1) = -(1-2^{1-s})\zeta(s), \tag{4}$$

where $\zeta(s)$ is the RIEMANN ZETA FUNCTION. The derivative is therefore given by

$$\frac{d}{ds}\mathrm{Li}_s(-1) = -2^{1-s}\zeta(s)\ln 2 - (1-2^{1-s})\zeta'(s), \tag{5}$$

or in the special case $s = 0$, by

$$\begin{aligned}\left[\frac{d}{ds}\mathrm{Li}_s(-1)\right]_{s=0} &= \ln 2 + \zeta'(0) = \ln 2 - \tfrac{1}{2}\ln(2\pi)\\ &= \ln\left(\sqrt{\frac{2}{\pi}}\right).\end{aligned} \tag{6}$$

This latter fact provides a remarkable proof of the WALLIS FORMULA.

The polylogarithm identities lead to remarkable expressions. Ramanujan gave the polylogarithm identities

$$\mathrm{Li}_2(\tfrac{1}{3}) - \tfrac{1}{6}\mathrm{Li}_2(\tfrac{1}{9}) = \tfrac{1}{18}\pi^2 - \tfrac{1}{6}(\ln 3)^2 \tag{7}$$

$$\begin{aligned}&\mathrm{Li}_2(-\tfrac{1}{2}) + \tfrac{1}{6}\mathrm{Li}_2(\tfrac{1}{9})\\ &\quad= -\tfrac{1}{18}\pi^2 + \ln 2\ln 3 - \tfrac{1}{2}(\ln 2)^2 - \tfrac{1}{3}(\ln 3)^2\end{aligned} \tag{8}$$

$$\mathrm{Li}_2(\tfrac{1}{4}) + \tfrac{1}{3}\mathrm{Li}_2(\tfrac{1}{9}) = \tfrac{1}{18}\pi^2 + 2\ln 2\ln 3 - 2(\ln 2)^2 - \tfrac{2}{3}(\ln 3)^2 \tag{9}$$

$$\mathrm{Li}_2(-\tfrac{1}{3}) - \tfrac{1}{3}\mathrm{Li}_2(\tfrac{1}{9}) = -\tfrac{1}{18}\pi^2 + \tfrac{1}{6}(\ln 3)^2 \tag{10}$$

$$\mathrm{Li}_2(-\tfrac{1}{8}) + \mathrm{Li}_2(\tfrac{1}{9}) = -\tfrac{1}{2}(\ln \tfrac{9}{8})^2 \tag{11}$$

(Berndt 1994), and Bailey *et al.* show that

$$\pi^2 = 36\,\mathrm{Li}_2(\tfrac{1}{2}) - 36\,\mathrm{Li}_2(\tfrac{1}{4}) - 12\,\mathrm{Li}_2(\tfrac{1}{8}) + 6\,\mathrm{Li}_2(\tfrac{1}{64}) \tag{12}$$

$$12\,\mathrm{Li}_2(\tfrac{1}{2}) = \pi^2 - 6(\ln 2)^2 \tag{13}$$

$$\begin{aligned}&\tfrac{35}{2}\zeta(3) - \pi^2\ln 2\\ &\quad= 36\,\mathrm{Li}_3(\tfrac{1}{2}) - 18\,\mathrm{Li}_3(\tfrac{1}{4}) - 4\,\mathrm{Li}_3(\tfrac{1}{8}) + \mathrm{Li}_3(\tfrac{1}{64})\end{aligned} \tag{14}$$

$$\begin{aligned}&2(\ln 2)^3 - 7\zeta(3)\\ &\quad= -24\,\mathrm{Li}_3(\tfrac{1}{2}) + 18\,\mathrm{Li}_3(\tfrac{1}{4}) + 4\,\mathrm{Li}_3(\tfrac{1}{8}) - \mathrm{Li}_3(\tfrac{1}{64})\end{aligned} \tag{15}$$

$$\begin{aligned}10(\ln 2)^3 - 2\pi^2\ln 2 &= -48\,\mathrm{Li}_3(\tfrac{1}{2}) + 54\,\mathrm{Li}_3(\tfrac{1}{4})\\ &\quad+12\,\mathrm{Li}_3(\tfrac{1}{8}) - 3\,\mathrm{Li}_3(\tfrac{1}{64}),\end{aligned} \tag{16}$$

and

$$\begin{aligned}&\frac{\mathrm{Li}_m(\frac{1}{64})}{6^{m-1}} - \frac{\mathrm{Li}_m(\frac{1}{8})}{3^{m-1}} - \frac{2\,\mathrm{Li}_m(\frac{1}{4})}{2^{m-1}} + \frac{4\,\mathrm{Li}_m(\frac{1}{2})}{9} - \frac{5(-\ln 2)^m}{9m!}\\ &+\frac{\pi^2(-\ln 2)^{m-2}}{54(m-2)!} - \frac{\pi^4(-\ln 2)^{m-4}}{486(m-4)!} - \frac{403\zeta(5)(-\ln 2)^{m-5}}{1296(m-5)!}\\ &= 0.\end{aligned} \tag{17}$$

No general ALGORITHM is know for the integration of polylogarithms of functions.

see also DILOGARITHM, EULERIAN NUMBER, LEGENDRE'S CHI-FUNCTION, LOGARITHMIC INTEGRAL, NIELSEN-RAMANUJAN CONSTANTS

References

Bailey, D.; Borwein, P.; and Plouffe, S. "On the Rapid Computation of Various Polylogarithmic Constants." `http://www.cecm.sfu.ca/~pborwein/PAPERS/P123.ps`.

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 323–326, 1994.

Lewin, L. *Polylogarithms and Associated Functions.* New York: North-Holland, 1981.

Lewin, L. *Structural Properties of Polylogarithms.* Providence, RI: Amer. Math. Soc., 1991.

Nielsen, N. *Der Euler'sche Dilogarithms.* Leipzig, Germany: Halle, 1909.

Polymorph

An INTEGER which is expressible in more than one way in the form x^2+Dy^2 or x^2-Dy^2 where x^2 is RELATIVELY PRIME to Dy^2. If the INTEGER is expressible in only one way, it is called a MONOMORPH.

see also ANTIMORPH, IDONEAL NUMBER, MONOMORPH

Polynomial

A POLYNOMIAL is a mathematical expression involving a series of POWERS in one or more variables multiplied by COEFFICIENTS. A POLYNOMIAL in one variable with constant COEFFICIENTS is given by

$$a_n x^n + \ldots + a_2 x^2 + a_1 x + a_0. \tag{1}$$

The highest POWER in a one-variable POLYNOMIAL is called its ORDER. A POLYNOMIAL in two variables with constant COEFFICIENTS is given by

$$\begin{aligned}&a_{nm}x^n y^m + a_{22}x^2y^2 + a_{21}x^2y + a_{12}xy^2\\ &\quad+a_{11}xy + a_{10}x + a_{01}y + a_{00}.\end{aligned} \tag{2}$$

Exchanging the COEFFICIENTS of a one-variable POLYNOMIAL end-to-end produces a POLYNOMIAL

$$a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n = 0 \tag{3}$$

whose ROOTS are RECIPROCALS $1/x_i$ of the original ROOTS x_i.

The following table gives special names given to polynomials of low orders.

Order	Polynomial Name
1	linear equation
2	quadratic equation
3	cubic equation
4	quartic equation
5	quintic equation
6	sextic equation

Polynomials of fourth degree may be computed using three multiplications and five additions if a few quantities are calculated first (Press *et al.* 1989):

$$\begin{aligned} &a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 \\ &\quad = [(Ax+B)^2 + Ax + C][(Ax+B)^2 + D] + E, \end{aligned} \tag{4}$$

where

$$A \equiv (a_4)^{1/4} \tag{5}$$

$$B \equiv \frac{a_3 - A^3}{4A^3} \tag{6}$$

$$D \equiv 3B^2 + 8B^3 + \frac{a_1A - 2a_2B}{A^2} \tag{7}$$

$$C \equiv \frac{a_2}{A^2} - 2B - 6B^2 - D \tag{8}$$

$$E \equiv a_0 - B^4 - B^2(C+D) - CD. \tag{9}$$

Similarly, a POLYNOMIAL of fifth degree may be computed with four multiplications and five additions, and a POLYNOMIAL of sixth degree may be computed with four multiplications and seven additions.

Polynomials of orders 1 to 4 are solvable using only algebraic functions and finite square root extraction. A first-order equation is trivially solvable. A second-order equation is soluble using the QUADRATIC EQUATION. A third-order equation is solvable using the CUBIC EQUATION. A fourth-order equation is solvable using the QUARTIC EQUATION. It was proved by Abel using GROUP THEORY that higher order equations cannot be solved by finite root extraction.

However, the general QUINTIC EQUATION may be given in terms of the THETA FUNCTIONS, or HYPERGEOMETRIC FUNCTIONS in one variable. Hermite and Kronecker proved that higher order POLYNOMIALS are not soluble in the same manner. Klein showed that the work of Hermite was implicit in the GROUP properties of the ICOSAHEDRON. Klein's method of solving the quintic in terms of HYPERGEOMETRIC FUNCTIONS in one variable can be extended to the sextic, but for higher order POLYNOMIALS, either HYPERGEOMETRIC FUNCTIONS in several variables or "Siegel functions" must be used. In the 1880s, Poincaré created functions which give the solution to the nth order POLYNOMIAL equation in finite form. These functions turned out to be "natural" generalizations of the ELLIPTIC FUNCTIONS.

Given an nth degree POLYNOMIAL, the ROOTS can be found by finding the EIGENVALUES of the MATRIX

$$\begin{bmatrix} -a_0/a_n & -a_1/a_n & -a_2/a_n & \cdots & -1 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & 1 & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}. \tag{10}$$

This method can be computationally expensive, but is fairly robust at finding close and multiple roots.

Polynomial identities involving sums and differences of like POWERS include

$$x^2 - y^2 = (x-y)(x+y) \tag{11}$$

$$x^3 - y^3 = (x-y)(x^2+xy+y^2) \tag{12}$$

$$x^3 + y^3 = (x+y)(x^2-xy+y^2) \tag{13}$$

$$x^4 - y^4 = (x-y)(x+y)(x^2+y^2) \tag{14}$$

$$x^4 + 4y^4 = (x^2+2xy+2y^2)(x^2-2xy+2y^2) \tag{15}$$

$$x^5 - y^5 = (x-y)(x^4+x^3y+x^2y^2+xy^3+y^4) \tag{16}$$

$$x^5 + y^5 = (x+y)(x^4-x^3y+x^2y^2-xy^3+y^4) \tag{17}$$

$$x^6 - y^6 = (x-y)(x+y)(x^2+xy+y^2)(x^2-xy+y^2) \tag{18}$$

$$x^6 + y^6 = (x^2+y^2)(x^4-x^2y^2+y^4). \tag{19}$$

Further identities include

$$x^4 + x^2y^2 + y^4 = (x^2+xy+y^2)(x^2-xy+y^2) \tag{20}$$

$$\begin{aligned} &({x_1}^2 - D{y_1}^2)({x_2}^2 - D{y_2}^2) \\ &\quad = (x_1x_2 + Dy_1y_2)^2 - D(x_1y_2 + x_2y_1)^2 \end{aligned} \tag{21}$$

$$\begin{aligned} &({x_1}^2 + D{y_1}^2)({x_2}^2 + D{y_2}^2) \\ &\quad = (x_1x_2 \pm Dy_1y_2)^2 + D(x_1y_2 \mp x_2y_1)^2. \end{aligned} \tag{22}$$

The identity

$$\begin{aligned} &(X+Y+Z)^7 - (X^7+Y^7+Z^7) = 7(X+Y)(X+Z)(Y+Z) \\ &\times[(X^2+Y^2+Z^2+XY+XZ+YZ)^2 + XYZ(X+Y+Z)] \end{aligned} \tag{23}$$

was used by Lamé in his proof that FERMAT'S LAST THEOREM was true for $n = 7$.

see also APPELL POLYNOMIAL, BERNSTEIN POLYNOMIAL, BESSEL POLYNOMIAL, BEZOUT'S THEOREM, BINOMIAL, BOMBIERI INNER PRODUCT, BOMBIERI NORM, CHEBYSHEV POLYNOMIAL OF THE FIRST KIND, CHEBYSHEV POLYNOMIAL OF THE SECOND KIND, CHRISTOFFEL-DARBOUX FORMULA, CHRISTOFFEL NUMBER, COMPLEX NUMBER, CYCLOTOMIC POLYNOMIAL, DESCARTES' SIGN RULE, DISCRIMINANT (POLYNOMIAL), DURFEE POLYNOMIAL, EHRHART POLYNOMIAL, EULER FOUR-SQUARE IDENTITY, FIBONACCI IDENTITY, FUNDAMENTAL THEOREM OF ALGEBRA, FUNDAMENTAL THEOREM OF SYMMETRIC FUNCTIONS, GAUSS-JACOBI MECHANICAL QUADRATURE, GEGENBAUER POLYNOMIAL, GRAM-SCHMIDT ORTHONORMALIZATION, GREATEST LOWER BOUND, HERMITE POLYNOMIAL, HILBERT POLYNOMIAL, IRREDUCIBLE POLYNOMIAL, ISOBARIC POLYNOMIAL, ISOGRAPH, JENSEN POLYNOMIAL, KERNEL POLYNOMIAL, KRAWTCHOUK POLYNOMIAL, LAGUERRE POLYNOMIAL, LEAST UPPER BOUND, LEGENDRE POLYNOMIAL, LIOUVILLE POLYNOMIAL IDENTITY, LOMMEL POLYNOMIAL, LUKÁCS THEOREM, MONOMIAL, ORTHOGONAL POLYNOMIALS, PERIMETER POLYNOMIAL, POISSON-CHARLIER POLYNOMIAL, POLLACZEK POLYNOMIAL, POLYNOMIAL BAR NORM, QUARTER SQUARES RULE, RAMANUJAN 6-10-8 IDENTITY, ROOT, RUNGE-WALSH THEOREM, SCHLÄFLI POLYNOMIAL, SEPARATION THEOREM, STIELTJES-WIGERT POLYNOMIAL, TRINOMIAL, TRINOMIAL IDENTITY, WEIERSTRAẞ'S POLYNOMIAL THEOREM, ZERNIKE POLYNOMIAL

References

Barbeau, E. J. *Polynomials.* New York: Springer-Verlag, 1989.

Bini, D. and Pan, V. Y. *Polynomial and Matrix Computations, Vol. 1: Fundamental Algorithms.* Boston, MA: Birkhäuser, 1994.

Borwein, P. and Erdélyi, T. *Polynomials and Polynomial Inequalities.* New York: Springer-Verlag, 1995.

Cockle, J. "Notes on the Higher Algebra." *Quart. J. Pure Applied Math.* **4**, 49–57, 1861.

Cockle, J. "Notes on the Higher Algebra (Continued)." *Quart. J. Pure Applied Math.* **5**, 1–17, 1862.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge, England: Cambridge University Press, 1989.

Project Mathematics! *Polynomials.* Videotape (27 minutes). California Institute of Technology. Available from the Math. Assoc. Amer.

Polynomial Bar Norm

For $p = \sum a_j z^j$, define

$$\begin{array}{ll} ||P||_1 = \displaystyle\int_0^{2\pi} |P(e^{i\theta})| \frac{d\theta}{2\pi} & |P|_1 = \displaystyle\sum_j |a_j| \\ ||P||_2 = \sqrt{\displaystyle\int_0^{2\pi} |P(e^{\pi\theta})|^2 \frac{d\theta}{2\pi}} & |P|_2 = \sqrt{\displaystyle\sum_j |a_j|^2} \\ ||P||_\infty = \max_{|z|=1} |P(z)| & |P|_\infty = \max_j |a_j|, \end{array}$$

where the $||P||_i$ norms are functions on the UNIT CIRCLE and the $|P|_i$ norms refer to the COEFFICIENTS $a_0, \ldots, a_n$.

see also BOMBIERI NORM, NORM, UNIT CIRCLE

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 151, 1989.

Polynomial Bracket Norm

see BOMBIERI NORM

Polynomial Curve

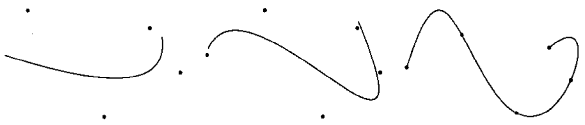

A curve obtained by fitting POLYNOMIALS to each ordinate of an ordered sequence of points. The above plots show POLYNOMIAL curves where the order of the fitting POLYNOMIAL varies from $p - 3$ to $p - 1$, where p is the number of points.

Polynomial curves have several undesirable features, including a nonintuitive variation of fitting curve with varying COEFFICIENTS, and numerical instability for high orders. SPLINES such as the BÉZIER CURVE are therefore used more commonly.

see also BÉZIER CURVE, POLYNOMIAL, SPLINE

Polynomial Factor

A FACTOR of a POLYNOMIAL $P(x)$ of degree n is a POLYNOMIAL $Q(x)$ of degree less than n which can be multiplied by another POLYNOMIAL $R(x)$ of degree less than n to yield $P(x)$, i.e., a POLYNOMIAL $Q(x)$ such that

$$P(x) = Q(x)R(x).$$

For example, since

$$x^2 - 1 = (x + 1)(x - 1),$$

both $x - 1$ and $x + 1$ are FACTORS of $x^2 - 1$. The COEFFICIENTS of factor POLYNOMIALS are often required to be REAL NUMBERS or INTEGERS but could, in general, be COMPLEX NUMBERS.

see also FACTOR, FACTORIZATION, PRIME FACTORIZATION

Polynomial Norm

see BOMBIERI NORM, MATRIX NORM, POLYNOMIAL BAR NORM, VECTOR NORM

Polynomial Remainder Theorem

If the Coefficients of the Polynomial

$$d_nx^n + d_{n-1}x^{n-1} + \ldots + d_0 = 0 \qquad (1)$$

are specified to be Integers, then integral Roots must have a Numerator which is a factor of d_0 and a Denominator which is a factor of d_n (with either sign possible). This follows since a Polynomial of Order n with k integral Roots can be expressed as

$$(a_1x+b_1)(a_2x+b_2)\cdots(a_kx+b_k)(c_{n-k}x^{n-k}+\ldots+c_0) = 0, \qquad (2)$$

where the Roots are $x_1 = -b_1/a_1$, $x_2 = -b_2/a_2$, ..., and $x_k = -b_k/a_k$. Factoring out the a_is,

$$a_1a_2\cdots a_k\left(x-\frac{b_1}{a_1}\right)\left(x-\frac{b_2}{a_2}\right)\cdots\left(x-\frac{b_k}{a_k}\right) \times(c_{n-k}x^{n-k}+\ldots+c_0) = 0. \qquad (3)$$

Now, multiplying through,

$$a_1a_2\cdots a_kc_{n-k}x^n + \ldots + b_1b_2\cdots b_kc_0 = 0, \qquad (4)$$

where we have not bothered with the other terms. Since the first and last Coefficients are d_n and d_0, all the integral roots of (1) are of the form [factors of d_0]/[factors of d_n].

Polynomial Ring

The Ring $R[x]$ of Polynomials in a variable x.

see also Polynomial, Ring

Polynomial Root

If the Coefficients of the Polynomial

$$d_nx^n + d_{n-1}x^{n-1} + \ldots + d_0 = 0 \qquad (1)$$

are specified to be Integers, then integral roots must have a Numerator which is a factor of d_0 and a Denominator which is a factor of d_n (with either sign possible). This is known as the Polynomial Remainder Theorem.

Let the Roots of the polynomial

$$P(x) \equiv a_nx^n + a_{n-1}x^{n-1} + \ldots + a_1x + a_0 \qquad (2)$$

be denoted $r_1, r_2, \ldots, r_n$. Then Newton's Relations are

$$\sum r_i = -\frac{a_{n-1}}{a_n} \qquad (3)$$

$$\sum r_ir_j = \frac{a_{n-2}}{a_n} \qquad (4)$$

$$\sum r_1r_2\cdots r_k = (-1)^k\frac{a_{n-k}}{a_n}. \qquad (5)$$

These can be derived by writing

$$(x-a)(x-b) = 0 \qquad (6)$$

$$\left(\frac{x}{a}-1\right)\left(\frac{x}{b}-1\right) = 0 \qquad (7)$$

$$\frac{1}{ab}x^2 - \left(\frac{1}{a}-\frac{1}{b}\right)x + 1 = 0. \qquad (8)$$

Similarly,

$$\left(\frac{x^2}{a^2}-1\right)\left(\frac{x^2}{b^2}=1\right) = 0 \qquad (9)$$

$$\frac{x^4}{a^2b^2} - x^2\left(\frac{1}{a^2}+\frac{1}{b^2}\right) + 1 = 0. \qquad (10)$$

Any Polynomial can be numerically factored, although different Algorithms have different strengths and weaknesses.

If there are no Negative Roots of a Polynomial (as can be determined by Descartes' Sign Rule), then the Greatest Lower Bound is 0. Otherwise, write out the Coefficients, let $n = -1$, and compute the next line. Now, if any Coefficients are 0, set them to minus the sign of the next higher Coefficient, starting with the second highest order Coefficient. If all the signs alternate, n is the greatest lower bound. If not, then subtract 1 from n, and compute another line. For example, consider the Polynomial

$$y = 2x^4 + 2x^3 - 7x^2 + x - 7. \qquad (11)$$

Performing the above Algorithm then gives

0	2	2	−7	1	−7
−1	2	0	−7	8	−15
—	2	−1	−7	8	−15
−2	2	−2	−3	7	−21
−3	2	−4	5	−14	35

so the greatest lower bound is −3.

If there are no Positive Roots of a Polynomial (as can be determined by Descartes' Sign Rule), the Least Upper Bound is 0. Otherwise, write out the Coefficients of the Polynomials, including zeros as necessary. Let $n = 1$. On the line below, write the highest order Coefficient. Starting with the second-highest Coefficient, add n times the number just written to the original second Coefficient, and write it below the second Coefficient. Continue through order zero. If all the Coefficients are Nonnegative, the least upper bound is n. If not, add one to x and repeat the process again. For example, take the Polynomial

$$y = 2x^4 - x^3 - 7x^2 + x - 7. \qquad (12)$$

Performing the above Algorithm gives

0	2	−1	−7	1	−7
1	2	1	−6	−5	−12
2	2	3	−1	−1	−9
3	2	5	8	25	68

so the LEAST UPPER BOUND is 3.

see also BAIRSTOW'S METHOD, DESCARTES' SIGN RULE, JENKINS-TRAUB METHOD, LAGUERRE'S METHOD, LEHMER-SCHUR METHOD, MAEHLY'S PROCEDURE, MULLER'S METHOD, ROOT, ZASSENHAUS-BERLEKAMP ALGORITHM

Polynomial Series

see MULTINOMIAL SERIES

Polyomino

A generalization of the DOMINO. An n-omino is defined as a collection of n squares of equal size arranged with coincident sides. FREE polyominoes can be picked up and flipped, so mirror image pieces are considered identical, whereas FIXED polyominoes are distinct if they have different chirality or orientation. FIXED polyominoes are also called LATTICE ANIMALS.

Redelmeier (1981) computed the number of FREE and FIXED polyominoes for $n \leq 24$, and Mertens (1990) gives a simple computer program. The sequence giving the number of FREE polyominoes of each order (Sloane's A000105, Ball and Coxeter 1987) is shown in the second column below, and that for FIXED polyominoes in the third column (Sloane's A014559).

n	Free	Fixed	Pos. Holes
1	1	1	0
2	1	2	0
3	2	6	0
4	5	19	0
5	12	63	0
6	35	216	0
7	108	760	1
8	369	2725	6
9	1285	9910	37
10	4655	39446	384
11	17073	135268	
12	63600	505861	
13	238591	1903890	
14	901971	7204874	
15	3426576	27394666	
16	13079255	104592937	
17	50107909	400795844	
18	192622052	1540820542	
19	742624232	5940738676	
20	2870671950	22964779660	
21	11123060678	88983512783	
22	43191857688	345532572678	
23	168047007728	1344372335524	
24	654999700403	5239988770268	

The best currently known bounds on the number of n-polyominoes are

$$3.72^n < P(n) < 4.65^n$$

(Eden 1961, Klarner 1967, Klarner and Rivest 1973, Ball and Coxeter 1987). For $n = 4$, the quartominoes are called STRAIGHT, L, T, SQUARE, and SKEW. For $n = 5$, the pentominoes are called f, I, L, N, P, T, U, V, W, X, y, and Z (Golomb 1995).

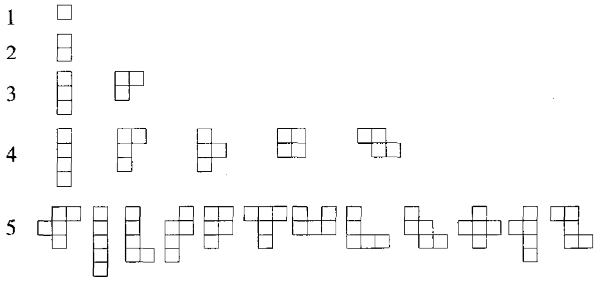

see also DOMINO, HEXOMINO, MONOMINO, PENTOMINO, POLYABOLO, POLYCUBE, POLYHEX, POLYIAMOND, POLYKING, POLYPLET, TETROMINO, TRIOMINO

References

Atkin, A. O. L. and Birch, B. J. (Eds.). *Computers in Number Theory: Proc. Sci. Research Council Atlas Symposium No. 2 Held at Oxford from 18–23 Aug., 1969.* New York: Academic Press, 1971.

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 109–113, 1987.

Beeler, M.; Gosper, R. W.; and Schroeppel, R. Item 77 in *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, pp. 48–50, Feb. 1972.

Eden, M. "A Two-Dimensional Growth Process." *Proc. Fourth Berkeley Symposium Math. Statistics and Probability, Held at the Statistical Laboratory, University of California, June 30–July 30, 1960.* Berkeley, CA: University of California Press, pp. 223–239, 1961.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/rndprc/rndprc.html`.

Gardner, M. "Polyominoes and Fault-Free Rectangles." Ch. 13 in *Martin Gardner's New Mathematical Diversions from Scientific American.* New York: Simon and Schuster, 1966.

Gardner, M. "Polyominoes and Rectification." Ch. 13 in *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, pp. 172–187, 1978.

Golomb, S. W. "Checker Boards and Polyominoes." *Amer. Math. Monthly* **61**, 675–682, 1954.

Golomb, S. W. *Polyominoes: Puzzles, Patterns, Problems, and Packings, rev. enl. 2nd ed.* Princeton, NJ: Princeton University Press, 1995.

Klarner, D. A. "Cell Growth Problems." *Can. J. Math.* **19**, 851–863, 1967.

Klarner, D. A. and Riverst, R. "A Procedure for Improving the Upper Bound for the Number of n-ominoes." *Can. J. Math.* **25**, 585–602, 1973.

Lei, A. "Bigger Polyominoes." `http://www.cs.ust.hk/~philipl/omino/bigpolyo.html`.

Lei, A. "Polyominoes." `http://www.cs.ust.hk/~philipl/omino/omino.html`.

Lunnon, W. F. "Counting Polyominoes." In *Computers in Number Theory* (Ed. A. O. L. Atkin and B. J. Brich). London: Academic Press, pp. 347–372, 1971.
Martin, G. *Polyominoes: A Guide to Puzzles and Problems in Tiling.* Washington, DC: Math. Assoc. Amer., 1991.
Mertens, S. "Lattice Animals—A Fast Enumeration Algorithm and New Perimeter Polynomials." *J. Stat. Phys.* **58**, 1095–1108, 1990.
Read, R. C. "Contributions to the Cell Growth Problem." *Canad. J. Math.* **14**, 1–20, 1962.
Redelmeier, D. H. "Counting Polyominoes: Yet Another Attack." *Discrete Math.* **36**, 191–203, 1981.
Ruskey, F. "Information on Polyominoes." `http://sue.csc.uvic.ca/~cos/inf/misc/PolyominoInfo.html`.
Sloane, N. J. A. Sequences A014559 and A000105/M1425 in "An On-Line Version of the Encyclopedia of Integer Sequences."
von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 342–343, 1993.

Polyomino Tiling

A TILING of the PLANE by specified types of POLYOMINOES. Interestingly, the FIBONACCI NUMBER F_{n+1} gives the number of ways for 2×1 dominoes to cover a $2 \times n$ checkerboard.

see also FIBONACCI NUMBER

References
Gardner, M. "Tiling with Polyominoes, Polyiamonds, and Polyhexes." Ch. 14 in *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, 1988.

Polyplet

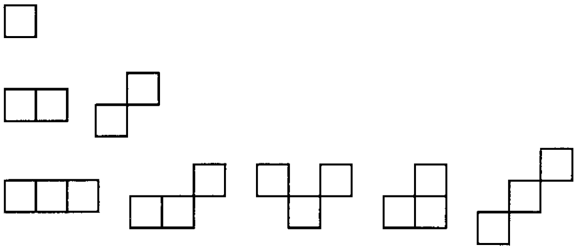

A POLYOMINO-like object made by attaching squares joined either at sides or corners. Because neighboring squares can be in relation to one another as KINGS may move on a CHESSBOARD, polyplets are sometimes also called POLYKINGS. The number of n-polyplets (with holes allowed) are 1, 2, 5, 22, 94, 524, 3031, ... (Sloane's A030222). The number of n-polyplets having bilateral symmetry are 1, 2, 4, 10, 22, 57, 131, ... (Sloane's A030234). The number of n-polyplets not having bilateral symmetry are 0, 0, 1, 12, 72, 467, 2900, ... (Sloane's A030235). The number of fixed n-polyplets are 1, 4, 20, 110, 638, 3832, ... (Sloane's A030232). The number of one-sided n-polyplets are 1, 2, 6, 34, 166, 991, ... (Sloane's A030233).

see also POLYIAMOND, POLYOMINO

References
Sloane, N. J. A. Sequences A030222, A030232, A030233, A030234, and A030235 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Polytope

A convex polytope may be defined as the CONVEX HULL of a finite set of points (which are always bounded), or as the intersection of a finite set of half-spaces. Explicitly, a d-dimensional polytope may be specified as the set of solutions to a system of linear inequalities

$$\mathsf{m}\mathbf{x} \leq \mathbf{b},$$

where m is a real $s \times d$ MATRIX and $\mathbf{b}$ is a real s-VECTOR. The positions of the vertices given by the above equations may be found using a process called VERTEX ENUMERATION.

A regular polytope is a generalization of the PLATONIC SOLIDS to an arbitrary DIMENSION. The NECESSARY condition for the figure with SCHLÄFLI SYMBOL $\{p, q, r\}$ to be a finite polytope is

$$\cos\left(\frac{\pi}{q}\right) < \sin\left(\frac{\pi}{p}\right)\sin\left(\frac{\pi}{r}\right).$$

SUFFICIENCY can be established by consideration of the six figures satisfying this condition. The table below enumerates the six regular polytopes in 4-D (Coxeter 1969, p. 414).

Name	Schläfli Symbol	N_0	N_1	N_2	N_3
regular simplex	$\{3,3,3\}$	5	10	10	5
hypercube	$\{4,3,3\}$	16	32	24	8
16-cell	$\{3,3,4\}$	8	24	32	16
24-cell	$\{3,4,3\}$	24	96	96	24
120-cell	$\{5,3,3\}$	600	1200	720	120
600-cell	$\{3,3,5\}$	120	720	1200	600

Here, N_0 is the number of VERTICES, N_1 the number of EDGES, N_2 the number of FACES, and N_3 the number of cells. These quantities satisfy the identity

$$N_0 - N_1 + N_2 - N_3 = 0,$$

which is a version of the POLYHEDRAL FORMULA.

For n-D with $n \geq 5$, there are only three regular polytopes, the MEASURE POLYTOPE, CROSS POLYTOPE, and regular SIMPLEX (which are analogs of the CUBE, OCTAHEDRON, and TETRAHEDRON).

see also 16-CELL, 24-CELL, 120-CELL, 600-CELL, CROSS POLYTOPE, EDGE (POLYTOPE), FACE, FACET, HYPERCUBE, INCIDENCE MATRIX, MEASURE POLYTOPE, RIDGE, SIMPLEX, TESSERACT, VERTEX (POLYHEDRON)

References
Coxeter, H. S. M. "Regular and Semi-Regular Polytopes I." *Math. Z.* **46**, 380–407, 1940.
Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, 1969.
Eppstein, D. "Polyhedra and Polytopes." `http://www.ics.uci.edu/~eppstein/junkyard/polytope.html`.

Poncelet's Closure Theorem

If an n-sided Poncelet Transverse constructed for two given Conic Sections is closed for one point of origin, it is closed for any position of the point of origin. Specifically, given one Ellipse inside another, if there exists one Circuminscribed (simultaneously inscribed in the outer and circumscribed on the inner) n-gon, then any point on the boundary of the outer Ellipse is the Vertex of some Circuminscribed n-gon.

References

Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, p. 193, 1965.

Poncelet's Continuity Principle

see Permanence of Mathematical Relations Principle

Poncelet-Steiner Theorem

All Euclidean Geometric Constructions can be carried out with a Straightedge alone if, in addition, one is given the Radius of a single Circle and its center. The theorem was suggested by Poncelet in 1822 and proved by Steiner in 1833. A construction using Straightedge alone is called a Steiner Construction.

see also Geometric Construction, Steiner Construction

References

Dörrie, H. "Steiner's Straight-Edge Problem." §34 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 165–170, 1965.

Steiner, J. *Geometric Constructions with a Ruler, Given a Fixed Circle with Its Center.* New York: Scripta Mathematica, 1950.

Poncelet's Theorem

see Poncelet's Closure Theorem

Poncelet Transform

see Poncelet Transverse

Poncelet Transverse

Let a Circle C_1 lie inside another Circle C_2. From any point on C_2, draw a tangent to C_1 and extend it to C_2. From the point, draw another tangent, etc. For n tangents, the result is called an n-sided Poncelet Transform.

References

Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, p. 192, 1965.

Pong Hau K'i

A Chinese Tic-Tac-Toe-like game.

see also Tic-Tac-Toe

References

Evans, R. "Pong Hau K'i." *Games and Puzzles* **53**, 19, 1976.

Straffin, P. D. Jr. "Position Graphs for Pong Hau K'i and Mu Torere." *Math. Mag.* **68**, 382–386, 1995.

Pons Asinorum

An elementary theorem in geometry whose name means "ass's bridge." The theorem states that the Angles at the base of an Isosceles Triangle (defined as a Triangle with two legs of equal length) are equal.

see also Isosceles Triangle, Pythagorean Theorem

References

Dunham, W. *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, p. 38, 1990.

Pontryagin Class

The ith Pontryagin class of a Vector Bundle is $(-1)^i$ times the ith Chern Class of the complexification of the Vector Bundle. It is also in the $4i$th cohomology group of the base Space involved.

see also Chern Class, Stiefel-Whitney Class

Pontryagin Duality

Let G be a locally compact Abelian Group. Let G^* be the group of all homeomorphisms $G \to R/Z$, in the compact open topology. Then G^* is also a locally compact Abelian Group, where the asterisk defines a contravariant equivalence of the category of locally compact Abelian groups with itself. The natural mapping $G \to (G^*)^*$, sending g to G, where $G(f) = f(g)$, is an isomorphism and a Homeomorphism. Under this equivalence, compact groups are sent to discrete groups and vice versa.

see also Abelian Group, Homeomorphism

Pontryagin Maximum Principle

A result is Control Theory. Define

$$H(\psi, x, u) \equiv (\psi, f(x, u)) \equiv \sum_{a=0}^{n} \psi_a f^a(x, u).$$

Then in order for a control $u(t)$ and a trajectory $x(t)$ to be optimal, it is Necessary that there exist Nonzero absolutely continuous vector function $\psi(t) = (\psi_0(t), \psi_1(t), \ldots, \psi_n(t))$ corresponding to the functions $u(t)$ and $x(t)$ such that

1. The function $H(\psi(t), x(t), u)$ attains its maximum at the point $u = u(t)$ almost everywhere in the interval $t_0 \leq t \leq t_1$,

$$H(\psi(t), x(t), u(t)) = \max_{u \in U} H(\psi(t), x(t), u).$$

2. At the terminal time t_1, the relations $\psi_0(t_1) \leq 0$ and $H(\psi(t_1), x(t_1), u(t_1)) = 0$ are satisfied.

References

Iyanaga, S. and Kawada, Y. (Eds.). "Pontrjagin's Maximum Principle." §88C in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 295–296, 1980.

Pontryagin Number

The Pontryagin number is defined in terms of the PONTRYAGIN CLASS of a MANIFOLD as follows. For any collection of PONTRYAGIN CLASSES such that their cup product has the same DIMENSION as the MANIFOLD, this cup product can be evaluated on the MANIFOLD'S FUNDAMENTAL CLASS. The resulting number is called the Pontryagin number for that combination of Pontryagin classes. The most important aspect of Pontryagin numbers is that they are COBORDISM invariant. Together, Pontryagin and STIEFEL-WHITNEY NUMBERS determine an oriented manifold's oriented COBORDISM class.

see also CHERN NUMBER, STIEFEL-WHITNEY NUMBER

Ponzo's Illusion

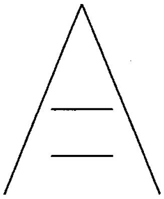

The upper HORIZONTAL line segment in the above figure appears to be longer than the lower line segment despite the fact that both are the same length.

see also ILLUSION, MÜLLER-LYER ILLUSION, POGGENDORFF ILLUSION, VERTICAL-HORIZONTAL ILLUSION

References

Fineman, M. *The Nature of Visual Illusion.* New York: Dover, p. 153, 1996.

Pop

An action which removes a single element from the top of a QUEUE or STACK, turning the LIST $(a_1, a_2, \ldots, a_n)$ into $(a_2, \ldots, a_n)$ and yielding the element a_1.

see also PUSH, STACK

Population Comparison

Let x_1 and x_2 be the number of successes in variates taken from two populations. Define

$$\hat{p}_1 \equiv \frac{x_1}{n_1} \qquad (1)$$

$$\hat{p}_2 \equiv \frac{x_2}{n_2}. \qquad (2)$$

The ESTIMATOR of the difference is then $\hat{p}_1 - \hat{p}_2$. Doing a z-TRANSFORM,

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sigma_{\hat{p}_1 - \hat{p}_2}}, \qquad (3)$$

where

$$\sigma_{\hat{p}_1 - \hat{p}_2} \equiv \sqrt{{\sigma_{\hat{p}_1}}^2 - {\sigma_{\hat{p}_2}}^2}. \qquad (4)$$

The STANDARD ERROR is

$$\text{SE}_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \qquad (5)$$

$$\text{SE}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}} \qquad (6)$$

$${s_{\text{pool}}}^2 = \frac{(n_1 - 1){s_1}^2 + (n_2 - 1){s_2}^2}{n_1 + n_2 - 2}. \qquad (7)$$

see also z-TRANSFORM

Population Growth

The differential equation describing exponential growth is

$$\frac{dN}{dt} = \frac{N}{\tau}. \qquad (1)$$

This can be integrated directly

$$\int_{N_0}^{N} \frac{dN}{N} = \int_0^t \frac{dt}{\tau} \qquad (2)$$

$$\ln\left(\frac{N}{N_0}\right) = \frac{t}{\tau}. \qquad (3)$$

Exponentiating,

$$N(t) = N_0 e^{t/\tau}. \qquad (4)$$

Defining $N(t = 1) = N_0 e^{\alpha}$ gives $\tau = 1/\alpha$ in (4), so

$$N(t) = N_0 e^{\alpha t}. \qquad (5)$$

The quantity α in this equation is sometimes known as the MALTHUSIAN PARAMETER.

Consider a more complicated growth law

$$\frac{dN}{dt} = \left(\frac{\alpha t - 1}{t}\right) N, \qquad (6)$$

where $\alpha > 1$ is a constant. This can also be integrated directly

$$\frac{dN}{N} = \left(\alpha - \frac{1}{t}\right) dt \qquad (7)$$

$$\ln N = \alpha t - \ln t + C \qquad (8)$$

$$N(t) = \frac{Ce^{\alpha t}}{t}. \qquad (9)$$

Note that this expression blows up at $t = 0$. We are given the INITIAL CONDITION that $N(t = 1) = N_0 e^{\alpha}$, so $C = N_0$.

$$N(t) = N_0 \frac{e^{\alpha t}}{t}. \qquad (10)$$

The t in the DENOMINATOR of (10) greatly suppresses the growth in the long run compared to the simple growth law.

The LOGISTIC GROWTH CURVE, defined by

$$\frac{dN}{dt} = \frac{r(K-N)}{N} \tag{11}$$

is another growth law which frequently arises in biology. It has a rather complicated solution for $N(t)$.

see also GOMPERTZ CURVE, LIFE EXPECTANCY, LOGISTIC GROWTH CURVE, LOTKA-VOLTERRA EQUATIONS, MAKEHAM CURVE, MALTHUSIAN PARAMETER, SURVIVORSHIP CURVE

Porism

An archaic type of mathematical proposition whose purpose is not entirely known.

see also AXIOM, LEMMA, POSTULATE, PRINCIPLE, STEINER'S PORISM, THEOREM

Porter's Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

The constant appearing in FORMULAS for the efficiency of the EUCLIDEAN ALGORITHM,

$$\begin{aligned} C &= \frac{6\ln 2}{\pi^2}\left[3\ln 2 + 4\gamma - \frac{24}{\pi^2}\zeta'(2) - 2\right] - \frac{1}{2} \\ &= 1.4670780794\ldots, \end{aligned}$$

where γ is the EULER-MASCHERONI CONSTANT and $\zeta(z)$ is the RIEMANN ZETA FUNCTION.

see also EUCLIDEAN ALGORITHM

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/porter/porter.html`.

Porter, J. W. "On a Theorem of Heilbronn." *Mathematika* **22**, 20–28, 1975.

Pósa's Theorem

Let G be a SIMPLE GRAPH with n VERTICES.

1. If, for every k in $1 \leq k < (n-1)/2$, the number of VERTICES of VALENCY not exceeding k is less than k, and
2. If, for n ODD, the number of VERTICES with VALENCY not exceeding $(n-1)/2$ is less than or equal to $(n-1)/2$,

then G contains a HAMILTONIAN CIRCUIT.

see also HAMILTONIAN CIRCUIT

Poset

see PARTIALLY ORDERED SET

Poset Dimension

The DIMENSION of a POSET $P = (X, \leq)$ is the size of the smallest REALIZER of P. Equivalently, it is the smallest INTEGER d such that P is ISOMORPHIC to a DOMINANCE order in $\mathbb{R}^d$.

see also DIMENSION, DOMINANCE, ISOMORPHIC POSETS, REALIZER

References

Dushnik, B. and Miller, E. W. "Partially Ordered Sets." *Amer. J. Math.* **63**, 600–610, 1941.

Trotter, W. T. *Combinatorics and Partially Ordered Sets: Dimension Theory.* Baltimore, MD: Johns Hopkins University Press, 1992.

Position Four-Vector

The CONTRAVARIANT FOUR-VECTOR arising in special and general relativity,

$$x^\mu = \begin{bmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{bmatrix} \equiv \begin{bmatrix} ct \\ x \\ y \\ z \end{bmatrix},$$

where c is the speed of light and t is time. Multiplication of two four-vectors gives the spacetime interval

$$\begin{aligned} I = g_{\mu\nu}x^\mu x^\nu &= (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 \\ &= (ct)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 \end{aligned}$$

see also FOUR-VECTOR, LORENTZ TRANSFORMATION, QUATERNION

Position Vector

see RADIUS VECTOR

Positive

A quantity $x > 0$, which may be written with an explicit PLUS SIGN for emphasis, $+x$.

see also NEGATIVE, NONNEGATIVE, PLUS SIGN, ZERO

Positive Definite Function

A POSITIVE definite FUNCTION f on a GROUP G is a FUNCTION for which the MATRIX $\{f(x_i x_j{}^{-1})\}$ is always POSITIVE SEMIDEFINITE HERMITIAN.

References

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Positive Definite Matrix

A MATRIX A is positive definite if

$$(\mathsf{A}\mathbf{v}) \cdot \mathbf{v} > 0 \tag{1}$$

for all VECTORS $\mathbf{v} \neq 0$. All EIGENVALUES of a positive definite matrix are POSITIVE (or, equivalently, the DETERMINANTS associated with *all* upper-left SUBMATRICES are POSITIVE).

The DETERMINANT of a positive definite matrix is POSITIVE, but the converse is not necessarily true (i.e., a matrix with a POSITIVE DETERMINANT is not necessarily positive definite).

A REAL SYMMETRIC MATRIX A is positive definite IFF there exists a REAL nonsingular MATRIX M such that

$$\mathsf{A} = \mathsf{M}\mathsf{M}^{\mathrm{T}}. \tag{2}$$

A 2×2 SYMMETRIC MATRIX

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} \tag{3}$$

is positive definite if

$$a{v_1}^2 + 2bv_1v_2 + c{v_2}^2 > 0 \tag{4}$$

for all $\mathbf{v} = (v_1, v_2) \neq 0$.

A HERMITIAN MATRIX A is positive definite if

1. $a_{ii} > 0$ for all i,
2. $a_{ii}a_{ij} > |a_{ij}|^2$ for $i \neq j$,
3. The element of largest modulus must lie on the leading diagonal,
4. $|\mathsf{A}| > 0$.

see also DETERMINANT, EIGENVALUE, HERMITIAN MATRIX, MATRIX, POSITIVE SEMIDEFINITE MATRIX

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1106, 1979.

Positive Definite Quadratic Form

A QUADRATIC FORM $Q(\mathbf{x})$ is said to be positive definite if $Q(\mathbf{x}) > 0$ for $\mathbf{x} \neq \mathbf{0}$. A REAL QUADRATIC FORM in n variables is positive definite IFF its canonical form is

$$Q(\mathbf{z}) = {z_1}^2 + {z_2}^2 + \ldots + {z_n}^2. \tag{1}$$

A BINARY QUADRATIC FORM

$$F(x, y) = a_{11}x^2 + 2a_{12}xy + a_{22}y^2 \tag{2}$$

of two REAL variables is positive definite if it is > 0 for any $(x, y) \neq (0, 0)$, therefore if $a_{11} > 0$ and the DISCRIMINANT $a \equiv a_{11}a_{22} - {a_{12}}^2 > 0$. A BINARY QUADRATIC FORM is positive definite if there exist NONZERO x and y such that

$$(ax^2 + 2bxy + cy^2)^2 \leq \tfrac{4}{3}|ac - b^2| \tag{3}$$

(Le Lionnais 1983).

A QUADRATIC FORM $(\mathbf{x}, \mathsf{A}\mathbf{x})$ is positive definite IFF every EIGENVALUE of A is POSITIVE. A QUADRATIC FORM $Q = (\mathbf{x}, \mathsf{A}\mathbf{x})$ with A a HERMITIAN MATRIX is positive definite if all the principal minors in the top-left corner of A are POSITIVE, in other words

$$a_{11} > 0 \tag{4}$$

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0 \tag{5}$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0. \tag{6}$$

see also INDEFINITE QUADRATIC FORM, POSITIVE SEMIDEFINITE QUADRATIC FORM

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1106, 1979.
Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 38, 1983.

Positive Definite Tensor

A TENSOR g whose discriminant satisfies

$$g \equiv g_{11}g_{22} - {g_{12}}^2 > 0.$$

Positive Integer

see $\mathbb{Z}^+$

Positive Semidefinite Matrix

A MATRIX A is positive semidefinite if

$$(\mathsf{A}\mathbf{v}) \cdot \mathbf{v} \geq 0$$

for all $\mathbf{v} \neq 0$.

see also POSITIVE DEFINITE MATRIX

Positive Semidefinite Quadratic Form

A QUADRATIC FORM $Q(\mathbf{x})$ is positive semidefinite if it is never < 0, but is 0 for some $\mathbf{x} \neq \mathbf{0}$. The QUADRATIC FORM, written in the form $(\mathbf{x}, \mathsf{A}\mathbf{x})$, is positive semidefinite IFF every EIGENVALUE of A is NONNEGATIVE.

see also INDEFINITE QUADRATIC FORM, POSITIVE DEFINITE QUADRATIC FORM

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1106, 1979.

Postage Stamp Problem

Consider a SET $A_k = \{a_1, a_2, \ldots, a_k\}$ of INTEGER denomination postage stamps with $1 = a_1 < a_2 < \ldots < a_k$. Suppose they are to be used on an envelope with room for no more than h stamps. The postage stamp problem then consists of determining the smallest INTEGER $N(h, A_k)$ which cannot be represented by a linear combination $\sum_{i=1}^{k} x_i a_i$ with $x_i \geq 0$ and $\sum_{i=1}^{k} x_i < h$.

Exact solutions exist for arbitrary A_k for $k = 2$ and 3. The $k = 2$ solution is

$$n(h, A_2) = (h + 3 - a_2)a_2 - 2$$

for $h \geq a_2 - 2$. The general problem consists of finding

$$n(h, k) = \max_{A_k} n(h, A_k).$$

It is known that

$$n(h, 2) = \left\lfloor \tfrac{1}{4}(h^2 + 6h + 1) \right\rfloor,$$

(Stöhr 1955, Guy 1994), where $\lfloor x \rfloor$ is the FLOOR FUNCTION, the first few values of which are 2, 4, 7, 10, 14, 18, 23, 28, 34, 40, ... (Sloane's A014616).

see also HARMONIOUS GRAPH, STAMP FOLDING

References

Guy, R. K. "The Postage Stamp Problem." §C12 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 123–127, 1994.

Sloane, N. J. A. Sequence A014616 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Stöhr, A. "Gelöste und ungelöste Fragen über Basen der natürlichen Zahlenreihe I, II." *J. reine angew. Math.* **194**, 111–140, 1955.

Posterior Distribution

see BAYESIAN ANALYSIS

Postnikov System

An iterated FIBRATION of EILENBERG-MAC LANE SPACES. Every TOPOLOGICAL SPACE has this HOMOTOPY type.

see also EILENBERG-MAC LANE SPACE, FIBRATION, HOMOTOPY

Postulate

A statement, also known as an AXIOM, which is taken to be true without PROOF. Postulates are the basic structure from which LEMMAS and THEOREMS are derived. The whole of EUCLIDEAN GEOMETRY, for example, is based on five postulates known as EUCLID'S POSTULATES.

see also ARCHIMEDES' POSTULATE, AXIOM, BERTRAND'S POSTULATE, CONJECTURE, EQUIDISTANCE POSTULATE, EUCLID'S FIFTH POSTULATE, EUCLID'S POSTULATES, LEMMA, PARALLEL POSTULATE, PORISM, PROOF, THEOREM, TRIANGLE POSTULATE

Potato Paradox

You buy 100 pounds of potatoes and are told that they are 99% water. After leaving them outside, you discover that they are now 98% water. The weight of the dehydrated potatoes is then a surprising 50 pounds!

References

Paulos, J. A. *A Mathematician Reads the Newspaper.* New York: BasicBooks, p. 81, 1995.

Potential Function

The term used in physics and engineering for a HARMONIC FUNCTION. Potential functions are extremely useful, for example, in electromagnetism, where they reduce the study of a 3-component VECTOR FIELD to a 1-component SCALAR FUNCTION.

see also HARMONIC FUNCTION, LAPLACE'S EQUATION, SCALAR POTENTIAL, VECTOR POTENTIAL

Potential Theory

The study of HARMONIC FUNCTIONS (also called POTENTIAL FUNCTIONS).

see also HARMONIC FUNCTION, SCALAR POTENTIAL, VECTOR POTENTIAL

References

Kellogg, O. D. *Foundations of Potential Theory.* New York: Dover, 1953.

MacMillan, W. D. *The Theory of the Potential.* New York: Dover, 1958.

Pothenot Problem

see SNELLIUS-POTHENOT PROBLEM

Poulet Number

A FERMAT PSEUDOPRIME to base 2, denoted psp(2), i.e., a COMPOSITE ODD INTEGER such that

$$2^{n-1} \equiv 1 \pmod{n}.$$

The first few Poulet numbers are 341, 561, 645, 1105, 1387, ... (Sloane's A001567). Pomerance *et al.* (1980) computed all 21,853 Poulet numbers less than 25×10^9.

Pomerance has shown that the number of Poulet numbers less than x for sufficiently large x satisfy

$$\exp[(\ln x)^{5/14}] < P_2(x) < x \exp\left(-\frac{\ln x \ln \ln \ln x}{2 \ln \ln x}\right)$$

(Guy 1994).

A Poulet number all of whose DIVISORS d satisfy $d | 2^d - 2$ is called a SUPER-POULET NUMBER. There are an infinite number of Poulet numbers which are not SUPER-POULET NUMBERS. Shanks (1993) calls *any* integer satisfying $2^{n-1} \equiv 1 \pmod{n}$ (i.e., not limited to ODD composite numbers) a FERMATIAN.

see also FERMAT PSEUDOPRIME, PSEUDOPRIME, SUPER-POULET NUMBER

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 28–29, 1994.

Pomerance, C.; Selfridge, J. L.; and Wagstaff, S. S. Jr. "The Pseudoprimes to $25 \cdot 10^9$." *Math. Comput.* **35**, 1003–1026, 1980. Available electronically from `ftp://sable.ox.ac.uk/pub/math/primes/ps2.Z`.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 115–117, 1993.

Sloane, N. J. A. Sequence A001567/M5441 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Power

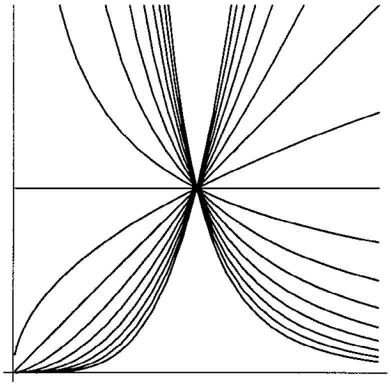

The exponent to which a given quantity is raised is known as its POWER. The expression x^a is therefore known as "x to the ath POWER." The rules for combining quantities containing powers are called the EXPONENT LAWS.

Special names given to various powers are listed in the following table.

Power	Name
1/2	square root
1/3	cube root
2	squared
3	cubed

The SUM of pth POWERS of the first n POSITIVE INTEGERS is given by FAULHABER'S FORMULA,

$$\sum_{k=1}^{n} k^p = \frac{1}{p+1} \sum_{k=1}^{p+1} (-1)^{\delta_{kp}} \binom{p+1}{k} B_{p+1-k} n^k,$$

where δ_{kp} is the KRONECKER DELTA, $\binom{n}{k}$ is a BINOMIAL COEFFICIENT, and B_k is a BERNOULLI NUMBER.

Let s_n be the largest INTEGER that is not the SUM of distinct nth powers of POSITIVE INTEGERS (Guy 1994). The first few values for $n = 2, 3, \ldots$ are 128, 12758, 5134240, 67898771, ... (Sloane's A001661).

CATALAN'S CONJECTURE states that 8 and 9 (2^3 and 3^2) are the only consecutive POWERS (excluding 0 and 1), i.e., the only solution to CATALAN'S DIOPHANTINE PROBLEM. This CONJECTURE has not yet been proved or refuted, although R. Tijdeman has proved that there can be only a finite number of exceptions should the CONJECTURE not hold. It is also known that 8 and 9 are the only consecutive CUBIC and SQUARE NUMBERS (in either order). Hyyrö and Mąkowski proved that there do not exist three consecutive POWERS (Ribenboim 1996).

Very few numbers of the form $n^p \pm 1$ are PRIME (where composite powers $p = kb$ need not be considered, since $n^(kb) \pm 1 = (n^k)^b \pm 1)$. The only PRIME NUMBERS of the form $n^p - 1$ for $n \leq 100$ and PRIME $2 \leq p \leq 10$ correspond to $n = 2$, i.e., $2^2 - 1 = 3$, $2^3 - 1 = 7$, $2^5 - 1 = 31$, ... The only PRIME NUMBERS of the form $n^p + 1$ for $n \leq 100$ and PRIME $2 \leq p \leq 10$ correspond to $p = 2$ with $n = 1, 2, 4, 6, 10, 14, 16, 20, 24, 26, \ldots$ (Sloane's A005574).

There are no nontrivial solutions to the equation

$$1^n + 2^n + \ldots + m^n = (m+1)^n$$

for $m \leq 10^{2,000,000}$ (Guy 1994, p. 153).

see also APOCALYPTIC NUMBER, BIQUADRATIC NUMBER, CATALAN'S CONJECTURE, CATALAN'S DIOPHANTINE PROBLEM, CUBE ROOT, CUBED, CUBIC NUMBER, EXPONENT, EXPONENT LAWS, FAULHABER'S FORMULA, FIGURATE NUMBER, MOESSNER'S THEOREM, NARCISSISTIC NUMBER, POWER RULE, SQUARE NUMBER, SQUARE ROOT, SQUARED, SUM, WARING'S PROBLEM

References

Barbeau, E. J. *Power Play: A Country Walk through the Magical World of Numbers.* Washington, DC: Math. Assoc. Amer., 1997.

Beyer, W. H. "Laws of Exponents." *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 158, 1987.

Guy, R. K. "Diophantine Equations." Ch. D in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 137, 139–198, and 153–154, 1994.

Ribenboim, P. "Catalan's Conjecture." *Amer. Math. Monthly* **103**, 529–538, 1996.

Sloane, N. J. A. Sequences A001661/M5393 and A005574/M1010 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Spanier, J. and Oldham, K. B. "The Integer Powers $(bx+c)^n$ and x^n" and "The Noninteger Powers x^ν." Ch. 11 and 13 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 83–90 and 99–106, 1987.

Power Center

see RADICAL CENTER

Power (Circle)

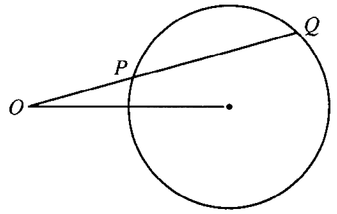

The POWER of the two points P and Q with respect to a CIRCLE is defined by

$$p \equiv OP \times PQ.$$

Let R be the RADIUS of a CIRCLE and d be the distance between a point P and the circle's center. Then the POWER of the point P relative to the circle is

$$p = d^2 - R^2.$$

If P is outside the CIRCLE, its POWER is POSITIVE and equal to the square of the length of the segment from P to the tangent to the CIRCLE through P. If P is inside the CIRCLE, then the POWER is NEGATIVE and equal to the product of the DIAMETERS through P.

The LOCUS of points having POWER k with regard to a fixed CIRCLE of RADIUS r is a CONCENTRIC CIRCLE of RADIUS $\sqrt{r^2+k}$. The CHORDAL THEOREM states that the LOCUS of points having equal POWER with respect to two given nonconcentric CIRCLES is a line called the RADICAL LINE (or CHORDAL; Dörrie 1965).

see also CHORDAL THEOREM, COAXAL CIRCLES, INVERSE POINTS, INVERSION CIRCLE, INVERSION RADIUS, INVERSIVE DISTANCE, RADICAL LINE

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 27–31, 1967.

Dixon, R. *Mathographics.* New York: Dover, p. 68, 1991.

Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, p. 153, 1965.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 28–34, 1929.

Pedoe, D. *Circles: A Mathematical View, rev. ed.* Washington, DC: Math. Assoc. Amer., pp. xxii–xxiv, 1995.

Power Curve

The curve with TRILINEAR COORDINATES $a^t : b^t : c^t$ for a given POWER t.

see also POWER

References

Kimberling, C. "Major Centers of Triangles." *Amer. Math. Monthly* **104**, 431–438, 1997.

Power Line

see RADICAL AXIS

Power Point

TRIANGLE centers with TRIANGLE CENTER FUNCTIONS of the form $\alpha = a^n$ are called nth POWER points. The 0th power point is the INCENTER, with TRIANGLE CENTER FUNCTION $\alpha = 1$.

see also INCENTER, TRIANGLE CENTER FUNCTION

References

Groenman, J. T. and Eddy, R. H. "Problem 858 and Solution." *Crux Math.* **10**, 306–307, 1984.

Kimberling, C. "Problem 865." *Crux Math.* **10**, 325–327, 1984.

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Power Rule

The DERIVATIVE of the POWER x^n is given by

$$\frac{d}{dx}(x^n) = nx^{n-1}.$$

see also CHAIN RULE, DERIVATIVE, EXPONENT LAWS, PRODUCT RULE

References

Anton, H. *Calculus with Analytic Geometry, 2nd ed.* New York: Wiley, p. 131, 1984.

Power Series

A power series in a variable z is an infinite SUM of the form

$$\sum_n^\infty a_i z^i, \tag{1}$$

where $n \geq 0$ and a_i are INTEGERS, REAL NUMBERS, COMPLEX NUMBERS, or any other quantities of a given type.

A CONJECTURE of Pólya is that if a FUNCTION has a POWER series with INTEGER COEFFICIENTS and RADIUS OF CONVERGENCE 1, then either the FUNCTION is RATIONAL or the UNIT CIRCLE is a natural boundary.

A generalized POWER sum $a(h)$ for $h = 0, 1, \ldots$ is given by

$$a(h) = \sum_{i=1}^m A_i(h){\alpha_i}^h, \tag{2}$$

with distinct NONZERO ROOTS α_i, COEFFICIENTS $A_i(h)$ which are POLYNOMIALS of degree $n_i - 1$ for POSITIVE INTEGERS n_i, and $i \in [1, m]$. The generalized POWER sum has order

$$n \equiv \sum_{i=m}^m n_i. \tag{3}$$

For any power series, one of the following is true:

1. The series converges only for $x = 0$.
2. The series converges absolutely for all x.
3. The series converges absolutely for all x in some finite open interval $(-R, R)$ and diverges if $x < -R$ or $x > R$. At the points $x = R$ and $x = -R$, the series may converge absolutely, converge conditionally, or diverge.

To determine the interval of convergence, apply the RATIO TEST for ABSOLUTE CONVERGENCE and solve for x. A POWER series may be differentiated or integrated within the interval of convergence. Convergent power series may be multiplied and divided (if there is no division by zero).

$$\sum_{k=1}^\infty k^{-p}. \tag{4}$$

CONVERGES if $p > 1$ and DIVERGES if $0 < p \leq 1$.

see also Binomial Series, Convergence Tests, Laurent Series, Maclaurin Series, Multinomial Series, p-Series, Polynomial, Power Set, Quotient-Difference Algorithm, Recurrence Sequence, Series, Series Reversion, Taylor Series

References

Arfken, G. "Power Series." §5.7 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 313–321, 1985.

Myerson, G. and van der Poorten, A. J. "Some Problems Concerning Recurrence Sequences." *Amer. Math. Monthly* **102**, 698–705, 1995.

Pólya, G. *Mathematics and Plausible Reasoning, Vol. 2: Patterns of Plausible Inference.* Princeton, NJ: Princeton University Press, p. 46, 1954.

Power Set

Given a Set S, the Power Set of S is the Set of all Subsets of S. The order of a Power set of a Set of order n is 2^n. Power sets are larger than the Sets associated with them.

see also Set, Subset

Power Spectrum

For a given signal, the power spectrum gives a plot of the portion of a signal's power (energy per unit time) falling within given frequency bins. The most common way of generating a power spectrum is by using a Fourier Transform, but other techniques such as the Maximum Entropy Method can also be used.

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Power Spectra Estimation Using the FFT" and "Power Spectrum Estimation by the Maximum Entropy (All Poles) Method." §13.4 and 13.7 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 542–551 and 565–569, 1992.

Power (Statistics)

The probability of getting a positive result for a given test which should produce a positive result.

see also Predictive Value, Sensitivity, Specificity, Statistical Test

Power Tower

$$a \uparrow\uparrow k \equiv \underbrace{a^{a^{\cdot^{\cdot^{\cdot^{a}}}}}}_{k},$$

where ↑ is Knuth's (1976) Arrow Notation.

$$a \uparrow^k n = a \uparrow^{k-1} [a \uparrow^k (n-1)].$$

The infinite power tower $x \uparrow\uparrow \infty = x^{x^{\cdot^{\cdot^{\cdot}}}}$ converges Iff $e^{-e} \leq x \leq e^{1/e}$ $(0.0659 \leq x \leq 1.4446)$.

see also Ackermann Function, Fermat Number, Mills' Constant

References

Knuth, D. E. "Mathematics and Computer Science: Coping with Finiteness. Advances in our Ability to Compute are Bringing us Substantially Closer to Ultimate Limitations." *Science* **194** 1235–1242, 1976.

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 11 and 226–229, 1991.

Power (Triangle)

The total Power of a Triangle is defined by

$$P \equiv \tfrac{1}{2}({a_1}^2 + {a_2}^2 + {a_3}^2), \tag{1}$$

where a_i are the side lengths, and the "partial power" is defined by

$$p_1 \equiv \tfrac{1}{2}({a_2}^2 + {a_3}^2 - {a_1}^2). \tag{2}$$

Then

$$P_1 = a_2 a_3 \cos\alpha_1 \tag{3}$$

$$P = p_1 + p_2 + p_3 \tag{4}$$

$$P^2 + {p_1}^2 + {p_2}^2 + {p_3}^2 = {a_1}^4 + {a_2}^4 + {a_3}^4 \tag{5}$$

$$\Delta = \tfrac{1}{2}\sqrt{p_2 p_3 + p_3 p_1 + p_3 p_1} \tag{6}$$

$$p_1 = \overline{A_1 H_2} \cdot \overline{A_1 A_3} \tag{7}$$

$$\frac{a_1 p_1}{\cos\alpha_1} = a_1 a_2 a_3 = 4\Delta R \tag{8}$$

$$p_1 \tan\alpha_1 = p_2 \tan\alpha_2 = p_3 \tan\alpha_3, \tag{9}$$

where Δ is the Area of the Triangle and H_i are the Feet of the Altitudes. Finally, if a side of the Triangle and the value of any partial power are given, then the Locus of the third Vertex is a Circle or straight line.

see also Altitude, Foot, Triangle

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 260–261, 1929.

Powerfree

see Biquadratefree, Cubefree, Prime Number, Squarefree

Powerful Number

An Integer m such that if $p|m$, then $p^2|m$, is called a powerful number. The first few are 1, 4, 8, 9, 16, 25, 27, 32, 36, 49, ... (Sloane's A001694). Powerful numbers are always of the form a^2b^3 for $a, b \geq 1$.

Not every Natural Number is the sum of two powerful numbers, but Heath-Brown (1988) has shown that every sufficiently large Natural Number is the sum of at most three powerful numbers. There are infinitely many pairs of consecutive powerful numbers, but Erdős has conjectured that there do not exist three consecutive powerful numbers. The CONJECTURE that there are no powerful number triples implies that there are infinitely many Wieferich primes (Granville 1986, Vardi 1991).

A separate usage of the term powerful number is for numbers which are the sums of the positive powers of their digits. The first few are 1, 2, 3, 4, 5, 6, 7, 8, 9, 24, 43, 63, 89, ... (Sloane's A007532).

References
Granville, A. "Powerful Numbers and Fermat's Last Theorem." *C. R. Math. Rep. Acad. Sci. Canada* **8**, 215–218, 1986.
Guy, R. K. "Powerful Numbers." §B16 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 67–73, 1994.
Heath-Brown, D. R. "Ternary Quadratic Forms and Sums of Three Square-Full Numbers." In *Séminaire de Theorie des Nombres, Paris 1986–87* (Ed. C. Goldstein). Boston, MA: Birkhäuser, pp. 137–163, 1988.
Ribenboim, P. "Catalan's Conjecture." *Amer. Math. Monthly* **103**, 529–538, 1996.
Sloane, N. J. A. Sequences A001694/M3325 and A007532/M0487 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 59–62, 1991.

Practical Number

A number n is practical if for all $k \leq n$, k is the sum of distinct proper divisors of n. Defined in 1948 by A. K. Srinivasen. All even PERFECT NUMBERS are practical. The number

$$m = 2^{n-1}(2^n - 1)$$

is practical for all $n = 2, 3, \ldots$. The first few practical numbers are 1, 2, 4, 6, 8, 12, 16, 18, 20, 24, 28, 30, 32, 36, 40, 42, 48, 54, 56, ... (Sloane's A005153). G. Melfi has computed twins, triplets, and 5-tuples of practical numbers. The first few 5-tuples are 12, 18, 30, 198, 306, 462, 1482, 2550, 4422, ...

References
Melfi, G. "On Two Conjectures About Practical Numbers." *J. Number Th.* **56**, 205–210, 1996.
Melfi, G. "Practical Numbers." `http://www.dm.unipi.it/gauss-pages/melfi/public_html/pratica.html`.
Sloane, N. J. A. Sequence A005153/M0991 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Pratt Certificate

A primality certificate based on FERMAT'S LITTLE THEOREM CONVERSE. Although the general idea had been well-established for some time, Pratt became the first to prove that the certificate tree was of polynomial size and could also be verified in polynomial time. He was also the first to observe that the tree implies that PRIMES are in the complexity class NP.

To generate a Pratt certificate, assume that n is a POSITIVE INTEGER and $\{p_i\}$ is the set of PRIME FACTORS of $n-1$. Suppose there exists an INTEGER x (called a "WITNESS") such that $x^{n-1} \equiv 1 \pmod{n}$ but $x^e \not\equiv 1 \pmod{n}$ whenever e is one of $(n-1)/p_i$. Then FERMAT'S LITTLE THEOREM CONVERSE states that n is PRIME (Wagon 1991, pp. 278–279).

By applying FERMAT'S LITTLE THEOREM CONVERSE to n and recursively to each purported factor of $n-1$, a certificate for a given PRIME NUMBER can be generated. Stated another way, the Pratt certificate gives a proof that a number a is a PRIMITIVE ROOT of the multiplicative GROUP (mod p) which, along with the fact that a has order $p-1$, proves that p is a PRIME.

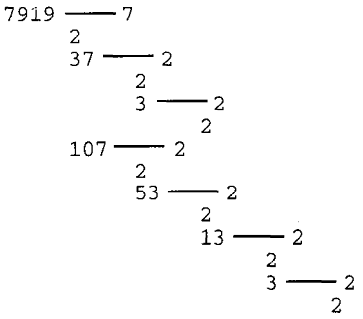

The figure above gives a certificate for the primality of $n = 7919$. The numbers to the right of the dashes are WITNESSES to the numbers to left. The set $\{p_i\}$ for $n - 1 = 7918$ is given by $\{2, 37, 107\}$. Since $7^{7918} \equiv 1 \pmod{7919}$ but $7^{7918/2}$, $7^{7918/37}$, $7^{7918/107} \not\equiv 1 \pmod{7919}$, 7 is a WITNESS for 7919. The PRIME divisors of $7918 = 7919 - 1$ are 2, 37, and 107. 2 is a so-called "self-WITNESS" (i.e., it is recognized as a PRIME without further ado), and the remainder of the witnesses are shown as a nested tree. Together, they certify that 7919 is indeed PRIME. Because it requires the FACTORIZATION of $n-1$, the METHOD of Pratt certificates is best applied to small numbers (or those numbers n known to have easily factorable $n-1$).

A Pratt certificate is quicker to generate for small numbers than are other types of primality certificates. The *Mathematica®* (Wolfram Research, Champaign, IL) task `ProvablePrime[n]` therefore generates an ATKIN-GOLDWASSER-KILIAN-MORAIN CERTIFICATE only for numbers above a certain limit (10^{10} by default), and a Pratt certificate for smaller numbers.

see also ATKIN-GOLDWASSER-KILIAN-MORAIN CERTIFICATE, FERMAT'S LITTLE THEOREM CONVERSE, PRIMALITY CERTIFICATE, WITNESS

References
Pratt, V. "Every Prime Has a Succinct Certificate." *SIAM J. Comput.* **4**, 214–220, 1975.
Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 278–285, 1991.
Wilf, H. §4.10 in *Algorithms and Complexity.* Englewood Cliffs, NJ: Prentice-Hall, 1986.

Pratt-Kasapi Theorem

see HOEHN'S THEOREM

Precedes

The relationship x precedes y is written $x \prec y$. The relation x precedes or is equal to y is written $x \preceq y$.

see also Succeeds

Precession

see Curve of Constant Precession

Precisely Unless

If A is true precisely unless B, then B implies not-A and not-B implies A. J. H. Conway has suggested the term "Unlesss" for this state of affairs, by analogy with Iff.

see also Iff, Unless

Predicate

A function whose value is either True or False.

see also And, False, Or, Predicate Calculus, True, XOR

Predicate Calculus

The branch of formal Logic dealing with representing the logical connections between statements as well as the statements themselves.

see also Gödel's Incompleteness Theorem, Logic, Predicate

Predictability

Predictability at a time τ in the future is defined by

$$\frac{R(x(t), x(t+\tau))}{H(x(t))},$$

and linear predictability by

$$\frac{L(x(t), x(t+\tau))}{H(x(t))},$$

where R and L are the Redundancy and Linear Redundancy, and H is the Entropy.

Prediction Paradox

see Unexpected Hanging Paradox

Predictive Value

The Positive predictive value is the probability that a test gives a true result for a true statistic. The negative predictive value is the probability that a test gives a false result for a false statistic.

see also Power (Statistics), Sensitivity, Specificity, Statistical Test

Predictor-Corrector Methods

A general method of integrating Ordinary Differential Equations. It proceeds by extrapolating a polynomial fit to the derivative from the previous points to the new point (the predictor step), then using this to interpolate the derivative (the corrector step). Press *et al.* (1992) opine that predictor-corrector methods have been largely supplanted by the Bulirsch-Stoer and Runge-Kutta Methods, but predictor-corrector schemes are still in common use.

see also Adams' Method, Gill's Method, Milne's Method, Runge-Kutta Method

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 896–897, 1972.

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 493–494, 1985.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Multistep, Multivalue, and Predictor-Corrector Methods." §16.7 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 740–744, 1992.

Pretzel Curve

see Knot Curve

Pretzel Knot

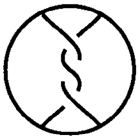

A Knot obtained from a Tangle which can be represented by a Finite sequence of Integers.

see also Tangle

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, p. 48, 1994.

Primality Certificate

A short set of data that proves the primality of a number. A certificate can, in general, be checked much more quickly than the time required to generate the certificate. Varieties of primality certificates include the Pratt Certificate and Atkin-Goldwasser-Kilian-Morain Certificate.

see also Atkin-Goldwasser-Kilian-Morain Certificate, Compositeness Certificate, Pratt Certificate

References

Wagon, S. "Prime Certificates." §8.7 in *Mathematica in Action.* New York: W. H. Freeman, pp. 277–285, 1991.

Primality Test

A test to determine whether or not a given number is PRIME. The RABIN-MILLER STRONG PSEUDOPRIME TEST is a particularly efficient ALGORITHM used by *Mathematica*® version 2.2 (Wolfram Research, Champaign, IL). Like many such algorithms, it is a probabilistic test using PSEUDOPRIMES, and can potentially (although with very small probability) falsely identify a COMPOSITE NUMBER as PRIME (although not vice versa). Unlike PRIME FACTORIZATION, primality testing is believed to be a P-PROBLEM (Wagon 1991). In order to guarantee primality, an almost certainly slower algorithm capable of generating a PRIMALITY CERTIFICATE must be used.

see also ADLEMAN-POMERANCE-RUMELY PRIMALITY TEST, FERMAT'S LITTLE THEOREM CONVERSE, FERMAT'S PRIMALITY TEST, FERMAT'S THEOREM, LUCAS-LEHMER TEST, MILLER'S PRIMALITY TEST, PÉPIN'S TEST, POCKLINGTON'S THEOREM, PROTH'S THEOREM, PSEUDOPRIME, RABIN-MILLER STRONG PSEUDOPRIME TEST, WARD'S PRIMALITY TEST, WILSON'S THEOREM

References

Beauchemin, P.; Brassard, G.; Crépeau, C.; Goutier, C.; and Pomerance, C. "The Generation of Random Numbers that are Probably Prime." *J. Crypt.* **1**, 53–64, 1988.

Brillhart, J.; Lehmer, D. H.; Selfridge, J.; Wagstaff, S. S. Jr.; and Tuckerman, B. *Factorizations of $b^n \pm 1$, $b = 2, 3, 5, 6, 7, 10, 11, 12$ Up to High Powers, rev. ed.* Providence, RI: Amer. Math. Soc., pp. lviii–lxv, 1988.

Cohen, H. and Lenstra, A. K. "Primality Testing and Jacobi Sums." *Math. Comput.* **42**, 297–330, 1984.

Knuth, D. E. *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, 1981.

Riesel, H. *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, 1994.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 15–17, 1991.

Primary

Each factor $p_i^{\alpha_i}$ in an INTEGER'S PRIME DECOMPOSITION is called a primary.

Primary Representation

Let π be a unitary REPRESENTATION of a GROUP G on a separable HILBERT SPACE, and let $R(\pi)$ be the smallest weakly closed algebra of bounded linear operators containing all $\pi(g)$ for $g \in G$. Then π is primary if the center of $R(\pi)$ consists of only scalar operations.

References

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Prime

A symbol used to distinguish one quantity x' ("x prime") from another related x. Primes are most commonly used to denote transformed coordinates, conjugate points, and DERIVATIVES.

see also PRIME ALGEBRAIC NUMBER, PRIME NUMBER

Prime Algebraic Number

An irreducible ALGEBRAIC INTEGER which has the property that, if it divides the product of two algebraic INTEGERS, then it DIVIDES at least one of the factors. 1 and -1 are the only INTEGERS which DIVIDE every INTEGER. They are therefore called the PRIME UNITS.

see also ALGEBRAIC INTEGER, PRIME UNIT

Prime Arithmetic Progression

Let the number of PRIMES of the form $mk+n$ less than x be denoted $\pi_{m,n}(x)$. Then

$$\lim_{x\to\infty} \frac{\pi_{a,b}(x)}{\mathrm{Li}(x)} = \frac{1}{\phi(a)},$$

where $\mathrm{Li}(x)$ is the LOGARITHMIC INTEGRAL and $\phi(x)$ is the TOTIENT FUNCTION.

Let P be an increasing arithmetic progression of n PRIMES with minimal difference $d > 0$. If a PRIME $p \leq n$ does not divide d, then the elements of P must assume all residues modulo p, specifically, some element of P must be divisible by p. Whereas P contains only primes, this element must be equal to p.

If $d < n\#$ (where $n\#$ is the PRIMORIAL of n), then some prime $p \leq n$ does not divide d, and that prime p is in P. Thus, in order to determine if P has $d < n\#$, we need only check a finite number of possible P (those with $d < n\#$ and containing prime $p \leq n$) to see if they contain only primes. If not, then $d \geq n\#$. If $d = n\#$, then the elements of P cannot be made to cover all residues of any prime p. The PRIME PATTERNS CONJECTURE then asserts that there are infinitely many arithmetic progressions of primes with difference d.

A computation shows that the smallest possible common difference for a set of n or more PRIMES in arithmetic progression for $n = 1, 2, 3, \ldots$ is 0, 1, 2, 6, 6, 30, 150, 210, 210, 210, 2310, 2310, 30030, 510510, ... (Sloane's A033188, Ribenboim 1989, Dubner and Nelson 1997, Wilson). The values up to $n = 13$ are rigorous, while the remainder are lower bounds which assume the validity of the PRIME PATTERNS CONJECTURE and are simply given by $p_{n-7}\#$, where p_i is the ith PRIME. The smallest first terms of arithmetic progressions of n primes *with minimal differences* are 2, 2, 3, 5, 5, 7, 7, 199, 199, 199, 60858179, 147692845283, 14933623, ... (Sloane's A033189; Wilson).

Smaller first terms are possible for nonminimal n-term progressions. Examples include the 8-term progression $11 + 1210230k$ for $k = 0, 1, \ldots, 7$, the 12-term progression $23143 + 30030k$ for $k = 0, 1, \ldots, 11$ (Golubev 1969, Guy 1994), and the 13-term arithmetic progression $766439 + 510510k$ for $k = 0, 1, \ldots, 12$ (Guy 1994).

The largest known set of primes in ARITHMETIC SEQUENCE is 22,

$$11,410,337,850,553 + 4,609,098,694,200k$$

for $k = 0, 1, \ldots, 21$ (Pritchard *et al.* 1995, UTS School of Mathematical Sciences).

The largest known sequence of *consecutive* PRIMES in ARITHMETIC PROGRESSION (i.e., all the numbers between the first and last term in the progression, except for the members themselves, are composite) is ten, given by

$$\begin{gathered} 100,996,972,469,714,247,637,786,655,587,969, \\ 840,329,509,324,689,190,041,803,603,417,758, \\ 904,341,703,348,882,159,067,229,719 + 210k \end{gathered}$$

for $k = 0, 1, \ldots, 9$, discovered by Harvey Dubner, Tony Forbes, Manfred Toplic, *et al.* on March 2, 1998. This beats the record of nine set on January 15, 1998 by the same investigators,

$$\begin{gathered} 99,679,432,066,701,086,484,490,653,695,853, \\ 561,638,982,364,080,991,618,395,774,048,585, \\ 529,071,475,461,114,799,677,694,651 + 210k \end{gathered}$$

for $k = 0, 1, \ldots, 8$ (two sequences of nine are now known), the progression of eight consecutive primes given by

$$\begin{gathered} 43,804,034,644,029,893,325,717,710,709,965, \\ 599,930,101,479,007,432,825,862,362,446,333, \\ 961,919,524,977,985,103,251,510,661 + 210k \end{gathered}$$

for $k = 0, 1, \ldots, 7$, discovered by Harvey Dubner, Tony Forbes, *et al.* on November 7, 1997 (several are now known), and the progression of seven given by

$$\begin{gathered} 1,089,533,431,247,059,310,875,780,378,922,957,732, \\ 908,036,492,993,138,195,385,213,105,561,742,150, \\ 447,308,967,213,141,717,486,151 + 210k, \end{gathered}$$

for $k = 0, 1, \ldots, 6$, discovered by H. Dubner and H. K. Nelson on Aug. 29, 1995 (Peterson 1995, Dubner and Nelson 1997). The smallest sequence of six consecutive PRIMES in arithmetic progression is

$$121,174,811 + 30k$$

for $k = 0, 1, \ldots, 5$ (Lander and Parkin 1967, Dubner and Nelson 1997). According to Dubner *et al.,* a trillion-fold increase in computer speed is needed before the search for a sequence of 11 consecutive primes is practical, so they expect the ten-primes record to stand for a long time to come.

It is conjectured that there are arbitrarily long sequences of PRIMES in ARITHMETIC PROGRESSION (Guy 1994).

see also ARITHMETIC PROGRESSION, CUNNINGHAM CHAIN, DIRICHLET'S THEOREM, LINNIK'S THEOREM, PRIME CONSTELLATION, PRIME-GENERATING POLYNOMIAL, PRIME NUMBER THEOREM, PRIME PATTERNS CONJECTURE, PRIME QUADRUPLET

References

Abel, U. and Siebert, H. "Sequences with Large Numbers of Prime Values." *Amer. Math. Monthly* **100**, 167–169, 1993.

Caldwell, C. K. "Cunningham Chain." `http://www.utm.edu/research/primes/glossary/CunninghamChain.html`.

Courant, R. and Robbins, H. "Primes in Arithmetical Progressions." §1.2b in Supplement to Ch. 1 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 26–27, 1996.

Davenport, H. "Primes in Arithmetic Progression" and "Primes in Arithmetic Progression: The General Modulus." Chs. 1 and 4 in *Multiplicative Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 1–11 and 27–34, 1980.

Dubner, H. and Nelson, H. "Seven Consecutive Primes in Arithmetic Progression." *Math. Comput.* **66**, 1743–1749, 1997.

Forbes, T. "Searching for 9 Consecutive Primes in Arithmetic Progression." `http://www.ltkz.demon.co.uk/ar2/9primes.htm`.

Forman, R. "Sequences with Many Primes." *Amer. Math. Monthly* **99**, 548–557, 1992.

Golubev, V. A. "Faktorisation der Zahlen der Form $x^3 \pm 4x^2 + 3x \pm 1$." *Anz. Österreich. Akad. Wiss. Math.-Naturwiss. Kl.* 184–191, 1969.

Guy, R. K. "Arithmetic Progressions of Primes" and "Consecutive Primes in A.P." §A5 and A6 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 15–17 and 18, 1994.

Lander, L. J. and Parkin, T. R. "Consecutive Primes in Arithmetic Progression." *Math. Comput.* **21**, 489, 1967.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 154–155, 1979.

Nelson, H. L. "There Is a Better Sequence." *J. Recr. Math.* **8**, 39–43, 1975.

Peterson, I. "Progressing to a Set of Consecutive Primes." *Sci. News* **148**, 167, Sep. 9, 1995.

Pritchard, P. A.; Moran, A.; and Thyssen, A. "Twenty-Two Primes in Arithmetic Progression." *Math. Comput.* **64**, 1337–1339, 1995.

Ramaré, O. and Rumely, R. "Primes in Arithmetic Progressions." *Math. Comput.* **65**, 397–425, 1996.

Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, p. 224, 1989.

Shanks, D. "Primes in Some Arithmetic Progressions and a General Divisibility Theorem." §104 in *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 104–109, 1993.

Sloane, N. J. A. Sequences A033188 and A033189 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Weintraub, S. "Consecutive Primes in Arithmetic Progression." *J. Recr. Math.* **25**, 169–171, 1993.

Zimmerman, P. `http://www.loria.fr/~zimmerma/records/8primes.announce`.

Prime Array

Find the $m \times n$ ARRAY of single digits which contains the maximum possible number of PRIMES, where allowable PRIMES may lie along any horizontal, vertical, or diagonal line. For $m = n = 2$, 11 PRIMES are maximal and are contained in the two distinct arrays

$$A(2,2) = \begin{bmatrix} 1 & 3 \\ 4 & 7 \end{bmatrix}, \begin{bmatrix} 1 & 3 \\ 7 & 9 \end{bmatrix},$$

giving the PRIMES (3, 7, 13, 17, 31, 37, 41, 43, 47, 71, 73) and (3, 7, 13, 17, 19, 31, 37, 71, 73, 79, 97), respectively. For the 3 × 2 array, 18 PRIMES are maximal and are contained in the arrays

$$A(3,2) = \begin{bmatrix} 1 & 1 & 3 \\ 9 & 7 & 4 \end{bmatrix}, \begin{bmatrix} 1 & 7 & 2 \\ 3 & 5 & 9 \end{bmatrix}, \begin{bmatrix} 1 & 7 & 2 \\ 4 & 3 & 9 \end{bmatrix},$$
$$\begin{bmatrix} 1 & 7 & 5 \\ 4 & 3 & 9 \end{bmatrix}, \begin{bmatrix} 1 & 7 & 9 \\ 3 & 2 & 5 \end{bmatrix}, \begin{bmatrix} 1 & 7 & 9 \\ 4 & 3 & 2 \end{bmatrix},$$
$$\begin{bmatrix} 1 & 7 & 9 \\ 4 & 3 & 4 \end{bmatrix}, \begin{bmatrix} 3 & 1 & 6 \\ 4 & 7 & 9 \end{bmatrix}, \begin{bmatrix} 3 & 7 & 6 \\ 4 & 1 & 9 \end{bmatrix}.$$

The best 3 × 3, 4 × 4, and 5 × 5 prime arrays known were found by C. Rivera and J. Ayala in 1998. They are

$$A(3,3) = \begin{bmatrix} 1 & 1 & 3 \\ 7 & 5 & 4 \\ 9 & 3 & 7 \end{bmatrix},$$

which contains 30 PRIMES,

$$A(4,4) = \begin{bmatrix} 1 & 1 & 3 & 9 \\ 6 & 4 & 5 & 1 \\ 7 & 3 & 9 & 7 \\ 3 & 9 & 2 & 9 \end{bmatrix},$$

which contains 63 PRIMES, and

$$A(5,5) = \begin{bmatrix} 1 & 1 & 9 & 3 & 3 \\ 9 & 9 & 5 & 6 & 3 \\ 8 & 9 & 4 & 1 & 7 \\ 3 & 3 & 7 & 3 & 1 \\ 3 & 2 & 9 & 3 & 9 \end{bmatrix},$$

which contains 116 PRIMES. S. C. Root found the a 6 × 6 array containing 187 primes:

$$A(6,6) = \begin{bmatrix} 3 & 1 & 7 & 3 & 3 & 3 \\ 9 & 9 & 5 & 6 & 3 & 9 \\ 1 & 1 & 8 & 1 & 4 & 2 \\ 1 & 3 & 6 & 3 & 7 & 3 \\ 3 & 4 & 9 & 1 & 9 & 9 \\ 3 & 7 & 9 & 3 & 7 & 9 \end{bmatrix}.$$

In 1998, M. Oswald found five new 6 × 6 arrays with 187 primes:

$$\begin{bmatrix} 1 & 3 & 9 & 1 & 9 & 9 \\ 3 & 1 & 7 & 2 & 3 & 4 \\ 9 & 9 & 4 & 7 & 9 & 3 \\ 9 & 1 & 5 & 7 & 1 & 3 \\ 9 & 8 & 3 & 6 & 1 & 7 \\ 9 & 1 & 7 & 3 & 3 & 3 \end{bmatrix}, \begin{bmatrix} 1 & 3 & 9 & 1 & 9 & 9 \\ 9 & 1 & 7 & 2 & 3 & 4 \\ 6 & 9 & 4 & 7 & 9 & 3 \\ 7 & 1 & 5 & 7 & 1 & 3 \\ 9 & 8 & 3 & 6 & 1 & 7 \\ 9 & 1 & 7 & 3 & 3 & 3 \end{bmatrix},$$
$$\begin{bmatrix} 3 & 1 & 7 & 3 & 3 & 3 \\ 9 & 9 & 5 & 6 & 3 & 9 \\ 1 & 1 & 8 & 1 & 4 & 2 \\ 1 & 3 & 6 & 3 & 7 & 3 \\ 3 & 4 & 9 & 1 & 9 & 9 \\ 3 & 7 & 9 & 9 & 3 & 9 \end{bmatrix}, \begin{bmatrix} 3 & 1 & 7 & 3 & 3 & 3 \\ 9 & 9 & 5 & 6 & 3 & 9 \\ 1 & 1 & 8 & 1 & 4 & 2 \\ 1 & 3 & 6 & 3 & 7 & 3 \\ 3 & 4 & 9 & 1 & 9 & 9 \\ 9 & 7 & 9 & 3 & 7 & 9 \end{bmatrix},$$
$$\begin{bmatrix} 3 & 1 & 7 & 3 & 3 & 3 \\ 9 & 9 & 5 & 6 & 3 & 9 \\ 1 & 1 & 8 & 1 & 4 & 5 \\ 1 & 3 & 6 & 3 & 7 & 3 \\ 3 & 4 & 9 & 1 & 9 & 9 \\ 9 & 9 & 9 & 2 & 3 & 3 \end{bmatrix}.$$

Rivera and Ayala conjecture that the 30-prime solution for $A(3,3)$ is maximal and unique. The following intervals have been completely searched without finding another 30-prime or better 3 × 3 array: $[1, 67 \times 10^6]$, $[100 \times 10^6, 133 \times 10^6]$, $[200 \times 10^6, 228 \times 10^6]$, $[300 \times 10^6, 325 \times 10^6]$, and $[400 \times 10^6, 418 \times 10^6]$.

Heuristic arguments by Rivera and Ayala suggest that the maximum possible number of primes in 4 × 4, 5 × 5, and 6 × 6 arrays are 58–63, 112–121, and 205–218, respectively.

see also ARRAY, PRIME ARITHMETIC PROGRESSION, PRIME CONSTELLATION, PRIME STRING

References

Dewdney, A. K. "Computer Recreations: How to Pan for Primes in Numerical Gravel." *Sci. Amer.* **259**, 120–123, July 1988.

Lee, G. "Winners and Losers." *Dragon User.* May 1984.

Lee, G. "Gordon's Paradoxically Perplexing Primesearch Puzzle." `http://www.geocities.com/MotorCity/7983/primesearch.html`.

Rivera, C. "Problems & Puzzles (Puzzles): The Gordon Lee Puzzle." `http://www.sci.net.mx/~crivera/ppp/puzz_001.htm`.

Weisstein, E. W. "Prime Arrays." `http://www.astro.virginia.edu/~eww6n/math/notebooks/PrimeArray.m`.

Prime Circle

A prime circle of order $2m$ is a CIRCULAR PERMUTATION of the numbers from 1 to $2m$ with adjacent PAIRS summing to a PRIME. The number of prime circles for $m = 1, 2, \ldots$, are 1, 1, 1, 2, 48, ...

References

Filz, A. "Problem 1046." *J. Recr. Math.* **14**, 64, 1982.

Filz, A. "Problem 1046." *J. Recr. Math.* **15**, 71, 1983.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 105–106, 1994.

Prime Cluster

see PRIME CONSTELLATION

Prime Constellation

A prime constellation, also called a PRIME k-TUPLE or PRIME k-TUPLET, is a sequence of k consecutive numbers such that the difference between the first and last is, in some sense, the least possible. More precisely, a prime k-tuplet is a sequence of consecutive PRIMES (p_1, p_2, ..., p_k) with $p_k - p_1 = s(k)$, where $s(k)$ is the smallest number s for which there exist k integers $b_1 < b_2 < \ldots < b_k$, $b_k - b_1 = s$ and, for every PRIME q, not all the residues modulo q are represented by b_1, b_2, ..., b_k (Forbes). For each k, this definition excludes a finite number of clusters at the beginning of the prime number sequence. For example, (97, 101, 103, 107, 109) satisfies the conditions of the definition of a prime 5-tuplet, but (3, 5, 7, 11, 13) does not because all three residues modulo 3 are represented (Forbes).

A prime double with $s(2) = 2$ is of the form $(p, p+2)$ and is called a pair of TWIN PRIMES. Prime doubles of the form $(p,\ p+6)$ are called SEXY PRIMES. A prime triplet has $s(3)=6$. However, the constellation $(p,\ p+2,\ p+4)$ cannot exist, since both $p+2$ and $p+4$ cannot be PRIME. However, there are several types of prime triplets which can exist: $(p,\ p+2,\ p+6)$, $(p,\ p+4,\ p+6)$, $(p,\ p+6,\ p+12)$. A PRIME QUADRUPLET is a constellation of four successive PRIMES with minimal distance $s(4)=8$, and is of the form $(p,\ p+2,\ p+6,\ p+8)$. The sequence $s(n)$ therefore begins 2, 6, 8, and continues 12, 16, 20, 26, 30, ... (Sloane's A008407). Another quadruplet constellation is $(p,\ p+6,\ p+12,\ p+18)$.

The first FIRST HARDY-LITTLEWOOD CONJECTURE states that the number of constellations $\leq x$ are asymptotically given by

$$\begin{aligned} P_x(p,p+2) &\sim 2\prod_{p\geq 3}\frac{p(p-2)}{(p-1)^2}\int_2^x\frac{dx'}{(\ln x')^2} \\ &= 1.320323632\int_2^x\frac{dx'}{(\ln x')^2} \end{aligned} \tag{1}$$

$$\begin{aligned} P_x(p,p+4) &\sim 2\prod_{p\geq 3}\frac{p(p-2)}{(p-1)^2}\int_2^x\frac{dx'}{(\ln x')^2} \\ &= 1.320323632\int_2^x\frac{dx'}{(\ln x')^2} \end{aligned} \tag{2}$$

$$\begin{aligned} P_x(p,p+6) &\sim 4\prod_{p\geq 3}\frac{p(p-2)}{(p-1)^2}\int_2^x\frac{dx'}{(\ln x')^2} \\ &= 2.640647264\int_2^x\frac{dx'}{(\ln x')^2} \end{aligned} \tag{3}$$

$$\begin{aligned} P_x(p,p+2,p+6) &\sim \frac{9}{2}\prod_{p\geq 5}\frac{p^2(p-3)}{(p-1)^3}\int_2^x\frac{dx'}{(\ln x')^3} \\ &= 2.858248596\int_2^x\frac{dx'}{(\ln x')^3} \end{aligned} \tag{4}$$

$$\begin{aligned} P_x(p,p+4,p+6) &\sim \frac{9}{2}\prod_{p\geq 5}\frac{p^2(p-3)}{(p-1)^3}\int_2^x\frac{dx'}{(\ln x')^3} \\ &= 2.858248596\int_2^x\frac{dx'}{(\ln x')^3} \end{aligned} \tag{5}$$

$$\begin{aligned} P_x(p,p+2,p+6,p+8) &\sim \frac{27}{2}\prod_{p\geq 5}\frac{p^3(p-4)}{(p-1)^4}\int_2^x\frac{dx'}{(\ln x')^4} \\ &= 4.151180864\int_2^x\frac{dx'}{(\ln x')^4} \end{aligned} \tag{6}$$

$$\begin{aligned} P_x(p,p+4,p+6,p+10) &\sim 27\prod_{p\geq 5}\frac{p^3(p-4)}{(p-1)^4}\int_2^x\frac{dx'}{(\ln x')^4} \\ &= 8.302361728\int_2^x\frac{dx'}{(\ln x')^4}. \end{aligned} \tag{7}$$

These numbers are sometimes called the HARDY-LITTLEWOOD CONSTANTS. (1) is sometimes called the extended TWIN PRIME CONJECTURE, and

$$C_{p,p+2}=2\Pi_2, \tag{8}$$

where Π_2 is the TWIN PRIMES CONSTANT. Riesel (1994) remarks that the HARDY-LITTLEWOOD CONSTANTS can be computed to arbitrary accuracy without needing the infinite sequence of primes.

The integrals above have the analytic forms

$$\int_2^x\frac{dx'}{(\ln x')^2}=\mathrm{Li}(x)+\frac{2}{\ln 2}-\frac{n}{\ln n} \tag{9}$$

$$\int_2^x\frac{dx'}{(\ln x')^3}=\tfrac{1}{2}\,\mathrm{Li}(x)-\frac{x(1+\ln x)}{(\ln x)^2}+\frac{1}{\ln 2}+\frac{1}{(\ln 2)^2} \tag{10}$$

$$\begin{aligned} \int_2^x\frac{dx'}{(\ln x')^4}=\frac{1}{6}\bigg\{\mathrm{Li}(x)&+\frac{2[2+\ln 2+(\ln 2)^2]}{(\ln 2)^3} \\ &-\frac{n[2+\ln n+(\ln n)^2]}{(\ln n)^3}\bigg\}, \end{aligned} \tag{11}$$

where $\mathrm{Li}(x)$ is the LOGARITHMIC INTEGRAL.

The following table gives the number of prime constellations $\leq 10^8$, and the second table gives the values predicted by the Hardy-Littlewood formulas.

Count	10^5	10^6	10^7	10^8
$(p,p+2)$	1224	8169	58980	440312
$(p,p+4)$	1216	8144	58622	440258
$(p,p+6)$	2447	16386	117207	879908
$(p,p+2,p+6)$	259	1393	8543	55600
$(p,p+4,p+6)$	248	1444	8677	55556
$(p,p+2,p+6,p+8)$	38	166	899	4768
$(p,p+6,p+12,p+18)$	75	325	1695	9330

Hardy-Littlewood	10^5	10^6	10^7	10^8
$(p,p+2)$	1249	8248	58754	440368
$(p,p+4)$	1249	8248	58754	440368
$(p,p+6)$	2497	16496	117508	880736
$(p,p+2,p+6)$	279	1446	8591	55491
$(p,p+4,p+6)$	279	1446	8591	55491
$(p,p+2,p+6,p+8)$	53	184	863	4735
$(p,p+6,p+12,p+18)$				

Consider prime constellations in which each term is of the form n^2+1. Hardy and Littlewood showed that the number of prime constellations of this form $< x$ is given by

$$P(x)\sim C\sqrt{x}\,(\ln x)^{-1}, \tag{12}$$

where

$$C=\prod_{\substack{p>2\\ p\ \text{prime}}}\left[1-\frac{(-1)^{(p-1)/2}}{p-1}\right]=1.3727\ldots \tag{13}$$

(Le Lionnais 1983).

Forbes gives a list of the "top ten" prime k-tuples for $2\leq k\leq 17$. The largest known 14-constellations are (11319107721272355839 + 0, 2, 8, 14, 18, 20, 24, 30, 32, 38, 42, 44, 48, 50), (10756418345074847279 + 0, 2, 8, 14, 18, 20, 24, 30, 32, 38, 42, 44, 48, 50), (6808488664768715759 + 0, 2, 8, 14, 18, 20, 24, 30, 32, 38, 42, 44, 48, 50), (6120794469172998449 + 0, 2, 8, 14, 18, 20, 24, 30, 32, 38, 42, 44, 48, 50), (5009128141636113611 + 0, 2, 6, 8, 12, 18, 20, 26, 30, 32, 36, 42, 48, 50).

The largest known prime 15-constellations are (84244343639633356306067 + 0, 2, 6, 12, 14, 20, 24, 26, 30, 36, 42, 44, 50, 54, 56), (89852089979514576043370+0, 2, 6, 12, 14, 20, 26, 30, 32, 36, 42, 44, 50, 54, 56), (35945854134669726946970 + 0, 2, 6, 12, 14, 20, 26, 30, 32, 36, 42, 44, 50, 54, 56), (35143833754615412325770+0, 2, 6, 12, 14, 20, 26, 30, 32, 36, 42, 44, 50, 54, 56), (34938645099859126094870 + 0, 2, 6, 12, 14, 20, 24, 26, 30, 36, 42, 44, 50, 54, 56).

The largest known prime 16-constellations are (3259125690557440336637+0, 2, 6, 12, 14, 20, 26, 30, 32, 36, 42, 44, 50, 54, 56, 60), (1522014304823128379267+0, 2, 6, 12, 14, 20, 26, 30, 32, 36, 42, 44, 50, 54, 56, 60), (47710850533373130107 + 0, 2, 6, 12, 14, 20, 26, 30, 32, 36, 42, 44, 50, 54, 56, 60), (13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73).

The largest known prime 17-constellations are (3259125690557440336631 + 0, 6, 8, 12, 18, 20, 26, 32, 36, 38, 42, 48, 50, 56, 60, 62, 66), (17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83) (13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79).

see also Composite Runs, Prime Arithmetic Progression, k-Tuple Conjecture, Prime k-Tuples Conjecture, Prime Quadruplet, Sexy Primes, Twin Primes

References

Forbes, T. "Prime k-tuplets." `http://www.ltkz.demon.co.uk/ktuplets.htm`.

Guy, R. K. "Patterns of Primes." §A9 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 23–25, 1994.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 38, 1983.

Riesel, H. *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, pp. 60–74, 1994.

Sloane, N. J. A. Sequence A008407 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Prime Counting Function

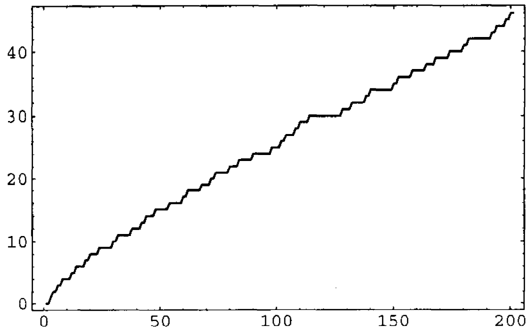

The function $\pi(n)$ giving the number of Primes less than n (Shanks 1993, p. 15). The first few values are 0, 1, 2, 2, 3, 3, 4, 4, 4, 4, 5, 5, 6, 6, 6, ... (Sloane's A000720). The following table gives the values of $\pi(n)$ for powers of 10 (Sloane's A006880; Hardy and Wright 1979, p. 4; Shanks 1993, pp. 242–243; Ribenboim 1996, p. 237). Deleglise and Rivat (1996) have computed $\pi(10^{20})$.

$$\begin{aligned}
\pi(10^3) &= 168\\
\pi(10^4) &= 1{,}229\\
\pi(10^5) &= 9{,}592\\
\pi(10^6) &= 78{,}498\\
\pi(10^7) &= 664{,}579\\
\pi(10^8) &= 5{,}761{,}455\\
\pi(10^9) &= 50{,}847{,}534\\
\pi(10^{10}) &= 455{,}052{,}511\\
\pi(10^{11}) &= 4{,}118{,}054{,}813\\
\pi(10^{12}) &= 37{,}607{,}912{,}018\\
\pi(10^{13}) &= 346{,}065{,}536{,}839\\
\pi(10^{14}) &= 3{,}204{,}941{,}750{,}802\\
\pi(10^{15}) &= 29{,}844{,}570{,}422{,}669\\
\pi(10^{16}) &= 279{,}238{,}341{,}033{,}925\\
\pi(10^{17}) &= 2{,}623{,}557{,}157{,}654{,}233\\
\pi(10^{18}) &= 24{,}739{,}954{,}287{,}740{,}860\\
\pi(10^{19}) &= 234{,}057{,}667{,}276{,}344{,}607,
\end{aligned}$$

$\pi(10^9)$ is incorrectly given as 50,847,478 in Hardy and Wright (1979). The prime counting function can be expressed by Legendre's Formula, Lehmer's Formula, Mapes' Method, or Meissel's Formula. A brief history of attempts to calculate $\pi(n)$ is given by Berndt (1994). The following table is taken from Riesel (1994).

Method	Time	Storage
Legendre	$\mathcal{O}(x)$	$\mathcal{O}(x^{1/2})$
Meissel	$\mathcal{O}(x/(\ln x)^3)$	$\mathcal{O}(x^{1/2}/\ln x)$
Lehmer	$\mathcal{O}(x/(\ln x)^4)$	$\mathcal{O}(x^{1/3}/\ln x)$
Mapes'	$\mathcal{O}(x^{0.7})$	$\mathcal{O}(x^{0.7})$
Lagarias-Miller-Odlyzko	$\mathcal{O}(x^{2/3+\epsilon})$	$\mathcal{O}(x^{1/3+\epsilon})$
Lagarias-Odlyzko 1	$\mathcal{O}(x^{3/5+\epsilon})$	$\mathcal{O}(x^{\epsilon})$
Lagarias-Odlyzko 2	$\mathcal{O}(x^{1/2+\epsilon})$	$\mathcal{O}(x^{1/4+\epsilon})$

A modified version of the prime counting function is given by

$$\pi_0(p) \equiv \begin{cases} \pi(p) & \text{for } p \text{ composite}\\ \pi(p) - \frac{1}{2} & \text{for } p \text{ prime} \end{cases}$$

$$\pi_0(p) = \sum_{n=1}^{\infty} \frac{\mu(x) f(x^{1/n})}{n},$$

where $\mu(n)$ is the Möbius Function and $f(x)$ is the Riemann-Mangoldt Function.

The notation $\pi_{a,b}$ is also used to denote the number of Primes of the form $ak + b$ (Shanks 1993, pp. 21–22).

Groups of EQUINUMEROUS values of $\pi_{a,b}$ include $(\pi_{3,1}, \pi_{3,2})$, $(\pi_{4,1}, \pi_{4,3})$, $(\pi_{5,1}, \pi_{5,2}, \pi_{5,3}, \pi_{5,4})$, $(\pi_{6,1}, \pi_{6,5})$, $(\pi_{7,1}, \pi_{7,2}, \pi_{7,3}, \pi_{7,4}, \pi_{7,5}, \pi_{7,6})$, $(\pi_{8,1}, \pi_{8,3}, \pi_{8,5}, \pi_{8,7})$, $(\pi_{9,1}, \pi_{9,2}, \pi_{9,4}, \pi_{9,5}, \pi_{9,7}, \pi_{9,8})$, and so on. The values of $\pi_{n,k}$ for small n are given in the following table for the first few powers of ten (Shanks 1993).

n	$\pi_{3,1}(n)$	$\pi_{3,2}(n)$	$\pi_{4,1}(n)$	$\pi_{4,3}(n)$
10^1	1	2	1	2
10^2	11	13	11	13
10^3	80	87	80	87
10^4	611	617	609	619
10^5	4784	4807	4783	4808
10^6	39231	39266	39175	39322
10^7	332194	332384	332180	332398

n	$\pi_{5,1}(n)$	$\pi_{5,2}(n)$	$\pi_{5,3}(n)$	$\pi_{5,4}(n)$
10^1	0	2	1	0
10^2	5	7	7	5
10^3	40	47	42	38
10^4	306	309	310	303
10^5	2387	2412	2402	2390
10^6	19617	19622	19665	19593
10^7	166104	166212	166230	166032

n	$\pi_{6,1}(n)$	$\pi_{6,5}(n)$
10^1	1	1
10^2	11	12
10^3	80	86
10^4	611	616
10^5	4784	4806
10^6	39231	39265

n	$\pi_{7,1}$	$\pi_{7,2}$	$\pi_{7,3}$	$\pi_{7,4}$	$\pi_{7,5}$	$\pi_{7,6}$
10^1	0	1	1	0	1	0
10^2	3	4	5	3	5	4
10^3	28	27	30	26	29	27
10^4	203	203	209	202	211	200
10^5	1593	1584	1613	1601	1604	1596
10^6	13063	13065	13105	13069	13105	13090

n	$\pi_{8,1}(n)$	$\pi_{8,3}(n)$	$\pi_{8,5}(n)$	$\pi_{8,7}(n)$
10^1	0	1	1	1
10^2	5	7	6	6
10^3	37	44	43	43
10^4	295	311	314	308
10^5	2384	2409	2399	2399
10^6	19552	19653	19623	19669
10^7	165976	166161	166204	166237

Note that since $\pi_{8,1}(n)$, $\pi_{8,3}(n)$, $\pi_{8,5}(n)$, and $\pi_{8,7}(n)$ are EQUINUMEROUS,

$$\pi_{4,1}(n) = \pi_{8,1}(n) + \pi_{8,5}$$
$$\pi_{4,3}(n) = \pi_{8,3}(n) + \pi_{8,7}$$

are also equinumerous.

Erdős proved that there exist at least one PRIME of the form $4k + 1$ and at least one PRIME of the form $4k + 3$ between n and $2n$ for all $n > 6$.

The smallest x such that $x \geq n\pi(x)$ for $n = 2, 3, \ldots$ are 2, 27, 96, 330, 1008, ... (Sloane's A038625), and the corresponding $\pi(x)$ are 1, 9, 24, 24, 66, 168, ... (Sloane's A038626). The number of solutions of $x \geq n\pi(x)$ for $n = 2, 3, \ldots$ are 4, 3, 3, 6, 7, 6, ... (Sloane's A038627).

see also BERTELSEN'S NUMBER, EQUINUMEROUS, PRIME ARITHMETIC PROGRESSION, PRIME NUMBER THEOREM, RIEMANN WEIGHTED PRIME-POWER COUNTING FUNCTION

References

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 134–135, 1994.

Brent, R. P. "Irregularities in the Distribution of Primes and Twin Primes." *Math. Comput.* **29**, 43–56, 1975.

Deleglise, M. and Rivat, J. "Computing $\pi(x)$: The Meissel, Lehmer, Lagarias, Miller, Odlyzko Method." *Math. Comput.* **65**, 235–245, 1996.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/hrdyltl/hrdyltl.html`.

Forbes, T. "Prime k-tuplets." `http://www.ltkz.demon.co.uk/ktuplets.htm`.

Guiasu, S. "Is There Any Regularity in the Distribution of Prime Numbers at the Beginning of the Sequence of Positive Integers?" *Math. Mag.* **68**, 110–121, 1995.

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, 1979.

Lagarias, J.; Miller, V. S.; and Odlyzko, A. "Computing $\pi(x)$: The Meissel-Lehmer Method." *Math. Comput.* **44**, 537–560, 1985.

Lagarias, J. and Odlyzko, A. "Computing $\pi(x)$: An Analytic Method." *J. Algorithms* **8**, 173–191, 1987.

Mapes, D. C. "Fast Method for Computing the Number of Primes Less than a Given Limit." *Math. Comput.* **17**, 179–185, 1963.

Meissel, E. D. F. "Über die Bestimmung der Primzahlmenge innerhalb gegebener Grenzen." *Math. Ann.* **2**, 636–642, 1870.

Ribenboim, P. *The New Book of Prime Number Records, 3rd ed.* New York: Springer-Verlag, 1996.

Riesel, H. "The Number of Primes Below x." *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, pp. 10–12, 1994.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, 1993.

Sloane, N. J. A. Sequences A038625, A038626, A038627, A000720/M2056, and A006880/M3608 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 74–76, 1991.

Wolf, M. "Unexpected Regularities in the Distribution of Prime Numbers." `http://www.ift.uni.wroc.pl/~mwolf`.

Prime Cut

Find two numbers such that $x^2 \equiv y^2 \pmod{n}$. If you know the GREATEST COMMON DIVISOR of n and $x - y$, there exists a high probability of determining a PRIME factor. Taking small numbers x which additionally give small PRIMES $x^2 \equiv p \pmod{n}$ further increases the chances of finding a PRIME factor.

Prime Decomposition

Given an INTEGER n, the prime decomposition is written

$$n = {p_1}^{\alpha_1} {p_2}^{\alpha_2} \cdots {p_n}^{\alpha_n},$$

where p_i are the n PRIME factors, each of order α_i. Each factor ${p_i}^{\alpha_i}$ is called a PRIMARY.

see also PRIMARY, PRIME FACTORIZATION ALGORITHMS, PRIME NUMBER

Prime Difference Function

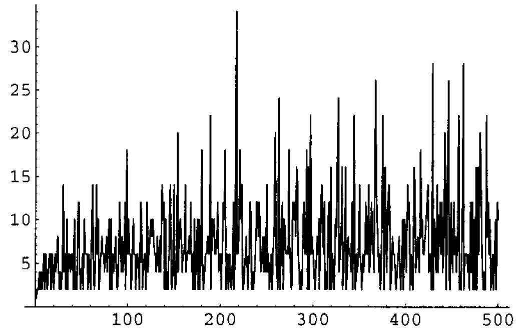

$$d_n \equiv p_{n+1} - p_n.$$

The first few values are 1, 2, 2, 4, 2, 4, 2, 4, 6, 2, 6, 4, 2, 4, 6, 6, ... (Sloane's A001223). Rankin has shown that

$$d_n > \frac{c \ln n \ln \ln n \ln \ln \ln \ln n}{(\ln \ln \ln n)^2}$$

for infinitely many n and for some constant c (Guy 1994).

An integer n is called a JUMPING CHAMPION if n is the most frequently occurring difference between consecutive primes $n \leq N$ for some N (Odlyzko *et al.*).

see also ANDRICA'S CONJECTURE, GOOD PRIME, JUMPING CHAMPION, PÓLYA CONJECTURE, PRIME GAPS, SHANKS' CONJECTURE, TWIN PEAKS

References

Bombieri, E. and Davenport, H. "Small Differences Between Prime Numbers." *Proc. Roy. Soc. A* **293**, 1–18, 1966.

Erdős, P.; and Straus, E. G. "Remarks on the Differences Between Consecutive Primes." *Elem. Math.* **35**, 115–118, 1980.

Guy, R. K. "Gaps between Primes. Twin Primes" and "Increasing and Decreasing Gaps." §A8 and A11 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 19–23 and 26–27, 1994.

Odlyzko, A.; Rubinstein, M.; and Wolf, M. "Jumping Champions." `http://www.research.att.com/~amo/doc/recent.html`.

Riesel, H. "Difference Between Consecutive Primes." *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, p. 9, 1994.

Sloane, N. J. A. Sequence A001223/M0296 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Prime Diophantine Equations

$k+2$ is PRIME IFF the 14 DIOPHANTINE EQUATIONS in 26 variables

$$wz + h + j - q = 0 \tag{1}$$

$$(gk + 2g + k + 1)(h + j) + h - z = 0 \tag{2}$$

$$16(k+1)^3(k+2)(n+1)^2 + 1 - f^2 = 0 \tag{3}$$

$$2n + p + q + z - q = 0 \tag{4}$$

$$e^3(e+2)(a+1)^2 + 1 - o^2 = 0 \tag{5}$$

$$(a^2 - 1)y^2 + 1 - x^2 = 0 \tag{6}$$

$$16r^2y^4(a^2 - 1) + 1 - u^2 = 0 \tag{7}$$

$$n + l + v - y = 0 \tag{8}$$

$$(a^2 - 1)l^2 + 1 - m^2 = 0 \tag{9}$$

$$ai + k + 1 - l - i = 0 \tag{10}$$

$$\{[a + u^2(u^2 - a)]^2 - 1\}(n + 4dy)^2 + 1 - (x + cu)^2 = 0 \tag{11}$$

$$p + l(a - n - 1) + b(2an + 2a - n^2 - 2n - 2) - m = 0 \tag{12}$$

$$q + y(a - p - 1) + s(2ap + 2a - p^2 - 2p - 2) - x = 0 \tag{13}$$

$$z + pl(a - p) + t(2ap - p^2 - 1) - pm = 0 \tag{14}$$

have a POSITIVE integral solution.

References

Riesel, H. *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, p. 39, 1994.

Prime Factorization

see FACTORIZATION, PRIME DECOMPOSITION, PRIME FACTORIZATION ALGORITHMS, PRIME FACTORS

Prime Factorization Algorithms

Many ALGORITHMS have been devised for determining the PRIME factors of a given number. They vary quite a bit in sophistication and complexity. It is very difficult to build a general-purpose algorithm for this computationally "hard" problem, so any additional information which is known about the number in question or its factors can often be used to save a large amount of time.

The simplest method of finding factors is so-called "DIRECT SEARCH FACTORIZATION" (a.k.a. TRIAL DIVISION). In this method, all possible factors are systematically tested using trial division to see if they actually DIVIDE the given number. It is practical only for very small numbers.

see also BRENT'S FACTORIZATION METHOD, CONTINUED FRACTION FACTORIZATION ALGORITHM, DIRECT SEARCH FACTORIZATION, DIXON'S FACTORIZATION METHOD, ELLIPTIC CURVE FACTORIZATION METHOD, EULER'S FACTORIZATION METHOD, EXCLUDENT FACTORIZATION METHOD, FERMAT'S FACTORIZATION METHOD, LEGENDRE'S FACTORIZATION

METHOD, LENSTRA ELLIPTIC CURVE METHOD, NUMBER FIELD SIEVE FACTORIZATION METHOD, POLLARD $p-1$ FACTORIZATION METHOD, POLLARD ρ FACTORIZATION ALGORITHM, QUADRATIC SIEVE FACTORIZATION METHOD, TRIAL DIVISION, WILLIAMS $p+1$ FACTORIZATION METHOD

References

Bressoud, D. M. *Factorization and Prime Testing.* New York: Springer-Verlag, 1989.

Brillhart, J.; Lehmer, D. H.; Selfridge, J.; Wagstaff, S. S. Jr.; and Tuckerman, B. *Factorizations of $b^n \pm 1$, $b = 2, 3, 5, 6, 7, 10, 11, 12$ Up to High Powers, rev. ed.* Providence, RI: Amer. Math. Soc., liv–lviii, 1988.

Dickson, L. E. "Methods of Factoring." Ch. 14 in *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, pp. 357–374, 1952.

Herman, P. "The Factoring Page!" `http://www.pslc.ucla.edu/~a540pau/factoring`.

Lenstra, A. K. and Lenstra, H. W. Jr. "Algorithms in Number Theory." In *Handbook of Theoretical Computer Science, Volume A: Algorithms and Complexity* (Ed. J. van Leeuwen). New York: Elsevier, pp. 673–715, 1990.

Odlyzko, A. M. "The Complexity of Computing Discrete Logarithms and Factoring Integers." §4.5 in *Open Problems in Communication and Computation* (Ed. T. M. Cover and B. Gopinath). New York: Springer-Verlag, pp. 113–116, 1987.

Odlyzko, A. M. "The Future of Integer Factorization." *CryptoBytes: The Technical Newsletter of RSA Laboratories* **1**, No. 2, 5–12, 1995.

Pomerance, C. "A Tale of Two Sieves." *Not. Amer. Math. Soc.* **43**, 1473–1485, 1996.

Riesel, H. *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, 1994.

Williams, H. C. and Shallit, J. O. "Factoring Integers Before Computers." In *Mathematics of Computation 1943–1993, Fifty Years of Computational Mathematics* (Ed. W. Gautschi). Providence, RI: Amer. Math. Soc., pp. 481–531, 1994.

Prime Factors

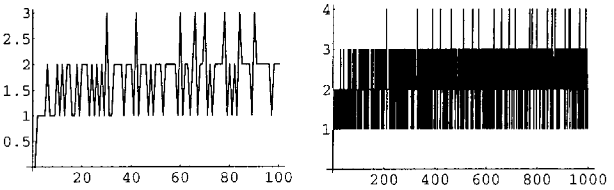

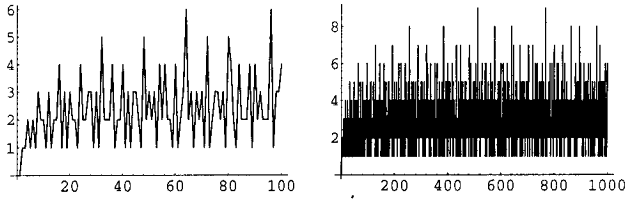

The number of DISTINCT PRIME FACTORS of a number n is denoted $\omega(n)$. The first few values for $n = 1, 2, \ldots$ are 0, 1, 1, 1, 1, 2, 1, 1, 1, 2, 1, 2, 1, 2, 2, 1, 1, 2, 1, 2, ... (Sloane's A001221; top figure). The number of not necessarily distinct prime factors of a number n is denoted $r(n)$. The first few values for $n = 1, 2, \ldots$ are 0, 1, 1, 2, 1, 2, 1, 3, 2, 2, 1, 3, 1, 2, 2, 4, 1, 3, 1, 3, ... (Sloane's A001222; bottom figure).

see also DISTINCT PRIME FACTORS, DIVISOR FUNCTION, GREATEST PRIME FACTOR, LEAST PRIME FACTOR, LIOUVILLE FUNCTION, PÓLYA CONJECTURE, PRIME FACTORIZATION ALGORITHMS

References

Sloane, N. J. A. Sequences A001222/M0094 and A001221/M0056 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Prime Field

A GALOIS FIELD $GF(p)$ where p is PRIME.

Prime Gaps

Letting

$$d_n \equiv p_{n+1} - p_n$$

be the PRIME DIFFERENCE FUNCTION, Rankin has showed that

$$d_n > \frac{c\ln n\ln\ln n\ln\ln\ln\ln n}{(\ln\ln\ln n)^2}$$

for infinitely many n are for some constant c (Guy 1994).

Let $p(d)$ be the smallest PRIME following d or more consecutive COMPOSITE NUMBERS. The largest known is

$$p(804) = 90,874,329,412,297.$$

The largest known prime gap is of length 4247, occurring following $10^{314} - 1929$ (Baugh and O'Hara 1992), although this gap is almost certainly not maximal (i.e., there probably exists a smaller number having a gap of the same length following it).

Let $c(n)$ be the smallest starting INTEGER $c(n)$ for a run of n consecutive COMPOSITE NUMBERS, also called a COMPOSITE RUN. No general method other than exhaustive searching is known for determining the first occurrence for a maximal gap, although arbitrarily large gaps exist (Nicely 1998). Cramér (1937) and Shanks (1964) conjectured that a maximal gap of length n appears at approximately $\exp(\sqrt{n})$. Wolf conjectures that the maximal gap of length n appears approximately at

$$\frac{n}{\pi(n)[2\ln\pi(n) - \ln n + \ln(2C_2)]},$$

where $\pi(n)$ is the PRIME COUNTING FUNCTION and C_2 is the TWIN PRIMES CONSTANT.

The first few $c(n)$ for $n = 1, 2, \ldots$ are 4, 8, 8, 24, 24, 90, 90, 114, ... (Sloane's A030296). The following table gives the same sequence omitting degenerate runs which are part of a run with greater n, and is a complete list of smallest maximal runs up to 10^{15}. $c(n)$ in this table is given by Sloane's A008950, and n by Sloane's A008996. The ending integers for the run corresponding to $c(n)$ are given by Sloane's A008995. Young and Potler (1989) determined the first occurrences of prime gaps up to 72,635,119,999,997, with all first occurrences found between 1 and 673. Nicely (1998) extended the list of maximal prime gaps to a length of 915, denoting gap lengths by the difference of bounding Primes, $c(n) - 1$.

n	$c(n)$	n	$c(n)$
1	4	319	2,300,942,550
3	8	335	3,842,610,774
5	24	353	4,302,407,360
7	90	381	10,726,904,660
13	114	383	20,678,048,298
17	524	393	22,367,084,960
19	888	455	25,056,082,088
21	1,130	463	42,652,618,344
33	1,328	467	127,976,334,672
35	9,552	473	182,226,896,240
43	15,684	485	241,160,024,144
51	19,610	489	297,501,075,800
71	31,398	499	303,371,455,242
85	155,922	513	304,599,508,538
95	360,654	515	416,608,695,822
111	370,262	531	461,690,510,012
113	492,114	533	614,487,453,424
117	1,349,534	539	738,832,927,928
131	1,357,202	581	1,346,294,310,750
147	2,010,734	587	1,408,695,493,610
153	4,652,354	601	1,968,188,556,461
179	17,051,708	651	2,614,941,710,599
209	20,831,324	673	7,177,162,611,713
219	47,326,694	715	13,828,048,559,701
221	122,164,748	765	19,581,334,192,423
233	189,695,660	777	42,842,283,925,352
247	191,912,784	803	90,874,329,411,493
249	387,096,134	805	171,231,342,420,521
281	436,273,010	905	218,209,405,436,543
287	1,294,268,492	915	1,189,459,969,825,483
291	1,453,168,142		

see also Jumping Champion, Prime Constellation, Prime Difference Function, Shanks' Conjecture

References

Baugh, D. and O'Hara, F. "Large Prime Gaps." *J. Recr. Math.* **24**, 186–187, 1992.

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 133–134, 1994.

Bombieri, E. and Davenport, H. "Small Differences Between Prime Numbers." *Proc. Roy. Soc. A* **293**, 1–18, 1966.

Brent, R. P. "The First Occurrence of Large Gaps Between Successive Primes." *Math. Comput.* **27**, 959–963, 1973.

Brent, R. P. "The Distribution of Small Gaps Between Successive Primes." *Math. Comput.* **28**, 315–324, 1974.

Brent, R. P. "The First Occurrence of Certain Large Prime Gaps." *Math. Comput.* **35**, 1435–1436, 1980.

Cramér, H. "On the Order of Magnitude of the Difference Between Consecutive Prime Numbers." *Acta Arith.* **2**, 23–46, 1937.

Guy, R. K. "Gaps between Primes. Twin Primes" and "Increasing and Decreasing Gaps." §A8 and A11 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 19–23 and 26–27, 1994.

Lander, L. J. and Parkin, T. R. "On First Appearance of Prime Differences." *Math. Comput.* **21**, 483–488, 1967.

Nicely, T. R. "New Maximal Prime Gaps and First Occurrences." `http://www.lynchburg.edu/public/academic/math/nicely/gaps/gaps.htm`. To Appear in *Math. Comput.*

Shanks, D. "On Maximal Gaps Between Successive Primes." *Math. Comput.* **18**, 646–651, 1964.

Sloane, N. J. A. Sequences A008950, A008995, A008996, and A030296 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Wolf, M. "First Occurrence of a Given Gap Between Consecutive Primes." `http://www.ift.uni.wroc.pl/~mwolf`.

Young, J. and Potler, A. "First Occurrence Prime Gaps." *Math. Comput.* **52**, 221–224, 1989.

Prime-Generating Polynomial

Legendre showed that there is no Rational algebraic function which always gives Primes. In 1752, Goldbach showed that no Polynomial with Integer Coefficients can give a Prime for all integral values. However, there exists a Polynomial in 10 variables with Integer Coefficients such that the set of Primes equals the set of Positive values of this Polynomial obtained as the variables run through all Nonnegative Integers, although it is really a set of Diophantine Equations in disguise (Ribenboim 1991).

$P(n)$	Range	#	Reference
$36n^2 - 810n + 2753$	[0, 44]	45	Fung and Ruby
$47n^2 - 1701n + 10181$	[0, 42]	43	Fung and Ruby
$n^2 - n + 41$	[0, 39]	40	Euler
$2n^2 + 29$	[0, 28]	29	Legendre
$n^2 + n + 17$	[0, 15]	16	Legendre
$2n^2 + 11$	[0, 10]	11	
$n^3 + n^2 + 17$	[0, 10]	11	

The above table gives some low-order polynomials which generate only Primes for the first few Nonnegative values (Mollin and Williams 1990). The best-known of these formulas is that due to Euler (Euler 1772, Ball and Coxeter 1987). Le Lionnais (1983) has christened numbers p such that the Euler-like polynomial

$$n^2 - n + p \tag{1}$$

is Prime for $p = 0, 1, \ldots, p-2$ as Lucky Numbers of Euler (where the case $p = 41$ corresponds to Euler's formula). Rabinovitch (1913) showed that for a Prime $p > 0$, Euler's polynomial represents a Prime for $n \in [0, p-2]$ (excluding the trivial case $p = 3$) Iff the Field $\mathbb{Q}(\sqrt{1-4p})$ has Class Number $h = 1$ (Rabinowitz 1913, Le Lionnais 1983, Conway and Guy 1996). As established by Stark (1967), there are only nine numbers $-d$ such that $h(-d) = 1$ (the Heegner Numbers -2, -3, -7, -11, -19, -43, -67, and -163), and of these, only 7, 11, 19, 43, 67, and 163 are of the required form. Therefore, the only Lucky Numbers of Euler are 2, 3, 5, 11, 17, and 41 (Le Lionnais 1983, Sloane's A014556), and there does not exist a better prime-generating polynomial of Euler's form.

Euler also considered quadratics of the form

$$2x^2 + p \tag{2}$$

and showed this gives PRIMES for $x \in [0, p-1]$ for PRIME $p > 0$ IFF $\mathbb{Q}(\sqrt{-2p})$ has CLASS NUMBER 2, which permits only $p = 3$, 5, 11, and 29. Baker (1971) and Stark (1971) showed that there are so such FIELDS for $p > 29$. Similar results have been found for POLYNOMIALS of the form

$$px^2 + px + n \tag{3}$$

(Hendy 1974).

see also CLASS NUMBER, HEEGNER NUMBER, LUCKY NUMBER OF EULER, PRIME ARITHMETIC PROGRESSION, PRIME DIOPHANTINE EQUATIONS, SCHINZEL'S HYPOTHESIS

References
Abel, U. and Siebert, H. "Sequences with Large Numbers of Prime Values." *Am. Math. Monthly* **100**, 167–169, 1993.
Baker, A. "Linear Forms in the Logarithms of Algebraic Numbers." *Mathematika* **13**, 204–216, 1966.
Baker, A. "Imaginary Quadratic Fields with Class Number Two." *Ann. Math.* **94**, 139–152, 1971.
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 60, 1987.
Boston, N. and Greenwood, M. L. "Quadratics Representing Primes." *Amer. Math. Monthly* **102**, 595–599, 1995.
Conway, J. H. and Guy, R. K. "The Nine Magic Discriminants." In *The Book of Numbers.* New York: Springer-Verlag, pp. 224–226, 1996.
Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 26, 1996.
Euler, L. *Nouveaux Mémoires de l'Académie royale des Sciences.* Berlin, p. 36, 1772.
Forman, R. "Sequences with Many Primes." *Amer. Math. Monthly* **99**, 548–557, 1992.
Garrison, B. "Polynomials with Large Numbers of Prime Values." *Amer. Math. Monthly* **97**, 316–317, 1990.
Hendy, M. D. "Prime Quadratics Associated with Complex Quadratic Fields of Class Number 2." *Proc. Amer. Math. Soc.* **43**, 253–260, 1974.
Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, pp. 88 and 144, 1983.
Mollin, R. A. and Williams, H. C. "Class Number Problems for Real Quadratic Fields." *Number Theory and Cryptology; LMS Lecture Notes Series* **154**, 1990.
Rabinowitz, G. "Eindeutigkeit der Zerlegung in Primzahlfaktoren in quadratischen Zahlkörpern." *Proc. Fifth Internat. Congress Math.* (Cambridge) **1**, 418–421, 1913.
Ribenboim, P. *The Little Book of Big Primes.* New York: Springer-Verlag, 1991.
Sloane, N. J. A. Sequence A014556 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Stark, H. M. "A Complete Determination of the Complex Quadratic Fields of Class Number One." *Michigan Math. J.* **14**, 1–27, 1967.
Stark, H. M. "An Explanation of Some Exotic Continued Fractions Found by Brillhart." In *Computers in Number Theory, Proc. Science Research Council Atlas Symposium No. 2 held at Oxford, from 18–23 August, 1969* (Ed. A. O. L. Atkin and B. J. Birch). London: Academic Press, 1971.
Stark, H. M. "A Transcendence Theorem for Class Number Problems." *Ann. Math.* **94**, 153–173, 1971.

Prime Group

When the ORDER h of a finite GROUP is a PRIME number, there is only one possible GROUP of ORDER h. Furthermore, the GROUP is CYCLIC.

see also p-GROUP

Prime Ideal

An IDEAL I such that if $ab \in I$, then either $a \in I$ or $b \in I$.

see also DEDEKIND RING, IDEAL, KRULL DIMENSION, MAXIMAL IDEAL, STICKELBERGER RELATION, STONE SPACE

Prime Knot

A KNOT other than the UNKNOT which cannot be expressed as a sum of two other KNOTS, neither of which is unknotted. A KNOT which is not prime is called a COMPOSITE KNOT. It is often possible to combine two prime knots to create two different COMPOSITE KNOTS, depending on the orientation of the two.

There is no known FORMULA for giving the number of distinct prime knots as a functions of number of crossings. For the first few n crossings, the numbers of prime knots are 0, 0, 1, 2, 3, 7, 21, 49, 165, 552, 2176, 9988, ... (Sloane's A002863).

see also COMPOSITE KNOT, KNOT

References
Sloane, N. J. A. Sequences A002863/M0851 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Sloane, N. J. A. and Plouffe, S. Extended entry in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.

Prime k-Tuple

see PRIME CONSTELLATION

Prime k-Tuples Conjecture

see also k-TUPLE CONJECTURE

Prime k-Tuplet

see PRIME CONSTELLATION

Prime Manifold

An n-MANIFOLD which cannot be "nontrivially" decomposed into other n-MANIFOLDS.

Prime Number

A prime number is a POSITIVE INTEGER p which has no DIVISORS other than 1 and p itself. Although the number 1 used to be considered a prime, it requires special treatment in so many definitions and applications involving primes greater than or equal to 2 that it is usually placed into a class of its own. Since 2 is the only EVEN prime, it is also somewhat special, so the set of all primes excluding 2 is called the "ODD PRIMES." The first few primes are 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, ... (Sloane's A000040, Hardy and Wright 1979, p. 3). POSITIVE INTEGERS other than 1 which are not prime are called COMPOSITE.

The function which gives the number of primes less than a number n is denoted $\pi(n)$ and is called the PRIME COUNTING FUNCTION. The theorem giving an asymptotic form for $\pi(n)$ is called the PRIME NUMBER THEOREM.

Prime numbers can be generated by sieving processes (such as the ERATOSTHENES SIEVE), and LUCKY NUMBERS, which are also generated by sieving, appear to share some interesting asymptotic properties with the primes.

Many PRIME FACTORIZATION ALGORITHMS have been devised for determining the prime factors of a given INTEGER. They vary quite a bit in sophistication and complexity. It is very difficult to build a general-purpose algorithm for this computationally "hard" problem, so any additional information which is known about the number in question or its factors can often be used to save a large amount of time. The simplest method of finding factors is so-called "DIRECT SEARCH FACTORIZATION" (a.k.a. TRIAL DIVISION). In this method, all possible factors are systematically tested using trial division to see if they actually DIVIDE the given number. It is practical only for very small numbers.

Because of their importance in encryption algorithms such as RSA ENCRYPTION, prime numbers can be important commercial commodities. In fact, Roger Schlafly has obtained U.S. Patent 5,373,560 (12/13/94) on the following two primes (expressed in hexadecimal notation):

```
98A3DF52AEAE9799325CB258D767EBD1F4630E9B
9E21732A4AFB1624BA6DF911466AD8DA960586F4
A0D5E3C36AF099660BDDC1577E54A9F402334433
ACB14BCB
```

and

```
93E8965DAFD9DFECFD00B466B68F90EA68AF5DC9
FED915278D1B3A137471E65596C37FED0C7829FF
8F8331F81A2700438ECDCC09447DC397C685F397
294F722BCC484AEDF28BED25AAAB35D35A65DB1F
D62C9D7BA55844FEB1F9401E671340933EE43C54
E4DC459400D7AD61248B83A2624835B31FFF2D95
95A5B90B276E44F9.
```

The FUNDAMENTAL THEOREM OF ARITHMETIC states that any POSITIVE INTEGER can be represented in exactly one way as a PRODUCT of primes. EUCLID'S SECOND THEOREM demonstrated that there are an infinite number of primes. However, it is not known if there are an infinite number of primes of the form x^2+1, whether there are an INFINITE number of TWIN PRIMES, or if a prime can always be found between n^2 and $(n+1)^2$.

Prime numbers satisfy many strange and wonderful properties. For example, there exists a CONSTANT $\theta \approx 1.3064$ known as MILLS' CONSTANT such that

$$\left\lfloor \theta^{3^n} \right\rfloor, \tag{1}$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION, is prime for all $n \geq 1$. However, it is not known if θ is IRRATIONAL. There also exists a CONSTANT $\omega \approx 1.9287800$ such that

$$\left\lfloor \underbrace{2^{2^{\cdot^{\cdot^{\cdot^{2^{\omega}}}}}}}_{n} \right\rfloor \tag{2}$$

(Ribenboim 1996, p. 186) is prime for every $n \geq 1$.

Explicit FORMULAS exist for the nth prime both as a function of n and in terms of the primes 2, ..., p_{n-1} (Hardy and Wright 1979, pp. 5–6; Guy 1994, pp. 36–41). Let

$$F(j) = \left\lfloor \cos^2 \left[\pi \frac{(j-1)!+1}{j} \right] \right\rfloor \tag{3}$$

for integral $j > 1$, and define $F(1) = 1$, where $\lfloor x \rfloor$ is again the FLOOR FUNCTION. Then

$$p_n = 1 + \sum_{m=1}^{2^n} \left\lfloor \left\lfloor \frac{n}{\sum_{j=1}^{m} F(j)} \right\rfloor^{1/n} \right\rfloor \tag{4}$$

$$= 1 + \sum_{m=1}^{2^n} \left\lfloor \left\lfloor \frac{n}{1+\pi(m)} \right\rfloor^{1/n} \right\rfloor, \tag{5}$$

where $\pi(m)$ is the PRIME COUNTING FUNCTION. It is also true that

$$p_{n+1} = 1 + p_n + F(p_n+1) + F(p_n+1)F(p_n+2) + \prod_{j=1}^{p} F(p_n+j) \tag{6}$$

(Ribenboim 1996, pp. 180–182). Note that the number of terms in the summation to obtain the nth prime is 2^n, so these formulas turn out not to be practical in the study of primes. An interesting INFINITE PRODUCT formula due to Euler which relates π and the nth PRIME p_n is

$$\pi = \frac{2}{\prod_{i=n}^{\infty} \left[1 + \frac{\sin(\frac{1}{2}\pi p_n)}{p_n} \right]} \tag{7}$$

$$= \frac{2}{\prod_{i=n}^{\infty} \left[1 + \frac{(-1)^{(p_n-1)/2}}{p_n} \right]} \tag{8}$$

(Blatner 1997). Conway (Guy 1983, Conway and Guy 1996, p. 147) gives an algorithm for generating primes based on 14 fractions, but it is actually just a concealed version of a SIEVE.

Some curious identities satisfied by primes p are

$$\sum_{k=1}^{p-1}\left\lfloor\frac{k^3}{p}\right\rfloor=\frac{(p-2)(p-1)(p+1)}{4} \qquad (9)$$

$$\sum_{k=1}^{(p-1)(p-2)}\left\lfloor(kp)^{1/3}\right\rfloor=\tfrac{1}{4}(3p-5)(p-2)(p-1) \qquad (10)$$

(Doster 1993),

$$\prod_{p\text{ prime}}\frac{p^2+1}{p^2-1}=\frac{5}{2} \qquad (11)$$

(Le Lionnais 1983, p. 46),

$$\sum_{k=1}^{\infty}x^k\ln k=\sum_{p\text{ prime}}\sum_{k=1}^{\infty}\frac{x^{p^k}}{1-x^{p^k}}, \qquad (12)$$

and

$$\sum_{k=1}^{\infty}(-1)^{k-1}e^{-kx}\ln k$$

$$=-\ln 2\sum_{k=1}^{\infty}\frac{1}{e^{2^kx}-1}+\sum_{\substack{p\text{ an}\\ \text{odd prime}}}\ln p\sum_{k=1}^{\infty}\frac{1}{e^{p^kx}+1} \qquad (13)$$

(Berndt 1994, p. 114).

It has been proven that the set of prime numbers is a DIOPHANTINE SET (Ribenboim 1991, pp. 106–107). Ramanujan also showed that

$$\frac{d\pi(x)}{dx}\sim\frac{1}{x\ln x}\sum_{n=1}^{\infty}\frac{\mu(n)}{n}x^{1/n}, \qquad (14)$$

where $\pi(x)$ is the PRIME COUNTING FUNCTION and $\mu(n)$ is the MÖBIUS FUNCTION (Berndt 1994, p. 117). B. M. Bredihin proved that

$$f(x,y)=x^2+y^2+1 \qquad (15)$$

takes prime values for infinitely many integral pairs (x,y) (Honsberger 1976, p. 30). In addition, the function

$$f(x,y)=\tfrac{1}{2}(y-1)\left\lfloor|B^2(x,y)-1|-(B^2(x,y)-1)\right\rfloor+2, \qquad (16)$$

where

$$B(x,y)=x(y+1)-(y!+1), \qquad (17)$$

$y!$ is the FACTORIAL, and $\lfloor x\rfloor$ is the FLOOR FUNCTION, generates only prime numbers for POSITIVE integral arguments. It not only generates every prime number, but generates ODD primes exactly once each, with all other values being 2 (Honsberger 1976, p. 33). For example,

$$f(1,2)=3 \qquad (18)$$

$$f(5,4)=5 \qquad (19)$$

$$f(103,6)=7, \qquad (20)$$

with no new primes generated for $x,y\le 1000$.

For n an INTEGER ≥ 2, n is prime IFF

$$\binom{n-1}{k}\equiv(-1)^k \pmod n \qquad (21)$$

for $k=0,1,\ldots,n-1$ (Deutsch 1996).

Cheng (1979) showed that for x sufficiently large, there always exist at least two prime factors between $(x-x^\alpha)$ and x for $\alpha\ge 0.477\ldots$ (Le Lionnais 1983, p. 26). Let $f(n)$ be the number of decompositions of n into two or more consecutive primes. Then

$$\lim_{x\to\infty}\frac{1}{x}\sum_{n=1}^{x}f(n)=\ln 2 \qquad (22)$$

(Moser 1963, Le Lionnais 1983, p. 30). Euler showed that the sum of the inverses of primes is infinite

$$\sum_{p\text{ prime}}\frac{1}{p}=\infty \qquad (23)$$

(Hardy and Wright 1979, p. 17), although it diverges very slowly. The sum exceeds 1, 2, 3, ... after 3, 59, 361139, ... (Sloane's A046024) primes, and its asymptotic equation is

$$\sum_{\substack{p=2\\ p\text{ prime}}}^{x}\frac{1}{p}=\ln\ln x+B_1+o(1), \qquad (24)$$

where B_1 is MERTENS CONSTANT (Hardy and Wright 1979, p. 351). Dirichlet showed the even stronger result that

$$\sum_{\substack{\text{prime } p\equiv b\ (\mathrm{mod}\ a)\\ (a,b)=1}}\frac{1}{p}=\infty \qquad (25)$$

(Davenport 1980, p. 34).

Despite the fact that $\sum 1/p$ diverges, Brun showed that

$$\sum_{\substack{p\\ p+2\text{ prime}}}\frac{1}{p}=B<\infty, \qquad (26)$$

where B is BRUN'S CONSTANT. The function defined by

$$P(n) \equiv \sum_p \frac{1}{p^n} \qquad (27)$$

taken over the primes converges for $n > 1$ and is a generalization of the RIEMANN ZETA FUNCTION known as the PRIME ZETA FUNCTION.

The probability that the largest prime factor of a RANDOM NUMBER x is less than $\sqrt{x}$ is $\ln 2$ (Beeler *et al.* 1972, Item 29). The probability that two INTEGERS picked at random are RELATIVELY PRIME is $[\zeta(2)]^{-1} = 6/\pi^2$, where $\zeta(x)$ is the RIEMANN ZETA FUNCTION (Cesaro and Sylvester 1883). Given three INTEGERS chosen at random, the probability that no common factor will divide them all is

$$[\zeta(3)]^{-1} \approx 1.202^{-1} = 0.832\ldots, \qquad (28)$$

where $\zeta(3)$ is APÉRY'S CONSTANT. In general, the probability that n random numbers lack a pth POWER common divisor is $[\zeta(np)]^{-1}$ (Beeler *et al.* 1972, Item 53).

Large primes include the large MERSENNE PRIMES, FERRIER'S PRIME, and $391581(2^{216193}-1)$ (Cipra 1989). The largest known prime as of 1998, is the MERSENNE PRIME $2^{3021377}-1$.

Primes consisting of consecutive DIGITS (counting 0 as coming after 9) include 2, 3, 5, 7, 23, 67, 89, 4567, 78901, ... (Sloane's A006510).

see also ADLEMAN-POMERANCE-RUMELY PRIMALITY TEST, ALMOST PRIME, ANDRICA'S CONJECTURE, BERTRAND'S POSTULATE, BROCARD'S CONJECTURE, BRUN'S CONSTANT, CARMICHAEL'S CONJECTURE, CARMICHAEL FUNCTION, CARMICHAEL NUMBER, CHEBYSHEV FUNCTION, CHEBYSHEV-SYLVESTER CONSTANT, CHEN'S THEOREM, CHINESE HYPOTHESIS, COMPOSITE NUMBER, COMPOSITE RUNS, COPELAND-ERDŐS CONSTANT, CRAMER CONJECTURE, CUNNINGHAM CHAIN, CYCLOTOMIC POLYNOMIAL, DE POLIGNAC'S CONJECTURE, DIRICHLET'S THEOREM, DIVISOR, ERDŐS-KAC THEOREM, EUCLID'S THEOREMS, FEIT-THOMPSON CONJECTURE, FERMAT NUMBER, FERMAT QUOTIENT, FERRIER'S PRIME, FORTUNATE PRIME, FUNDAMENTAL THEOREM OF ARITHMETIC, GIGANTIC PRIME, GIUGA'S CONJECTURE, GOLDBACH CONJECTURE, GOOD PRIME, GRIMM'S CONJECTURE, HARDY-RAMANUJAN THEOREM, IRREGULAR PRIME, KUMMER'S CONJECTURE, LEHMER'S PROBLEM, LINNIK'S THEOREM, LONG PRIME, MERSENNE NUMBER, MERTENS FUNCTION, MILLER'S PRIMALITY TEST, MIRIMANOFF'S CONGRUENCE, MÖBIUS FUNCTION, PALINDROMIC NUMBER, PÉPIN'S TEST, PILLAI'S CONJECTURE, POULET NUMBER, PRIMARY, PRIME ARRAY, PRIME CIRCLE, PRIME FACTORIZATION ALGORITHMS, PRIME NUMBER OF MEASUREMENT, PRIME NUMBER THEOREM, PRIME POWER SYMBOL, PRIME STRING, PRIME TRIANGLE, PRIME ZETA FUNCTION, PRIMITIVE PRIME FACTOR, PRIMORIAL, PROBABLE PRIME, PSEUDOPRIME, REGULAR PRIME, RIEMANN FUNCTION, ROTKIEWICZ THEOREM, SCHNIRELMANN'S THEOREM, SELFRIDGE'S CONJECTURE, SEMIPRIME, SHAH-WILSON CONSTANT, SIERPIŃSKI'S COMPOSITE NUMBER THEOREM, SIERPIŃSKI'S PRIME SEQUENCE THEOREM, SMOOTH NUMBER, SOLDNER'S CONSTANT, SOPHIE GERMAIN PRIME, TITANIC PRIME, TOTIENT FUNCTION, TOTIENT VALENCE FUNCTION, TWIN PRIMES, TWIN PRIMES CONSTANT, VINOGRADOV'S THEOREM, VON MANGOLDT FUNCTION, WARING'S CONJECTURE, WIEFERICH PRIME, WILSON PRIME, WILSON QUOTIENT, WILSON'S THEOREM, WITNESS, WOLSTENHOLME'S THEOREM, ZSIGMONDY THEOREM

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Berndt, B. C. "Ramanujan's Theory of Prime Numbers." Ch. 24 in *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, 1994.

Blatner, D. *The Joy of Pi.* New York: Walker, p. 110, 1997.

Caldwell, C. "Largest Primes." `http://www.utm.edu/research/primes/largest.html`.

Cheng, J. R. "On the Distribution of Almost Primes in an Interval II." *Sci. Sinica* **22**, 253–275, 1979.

Cipra, B. A. "Math Team Vaults Over Prime Record." *Science* **245**, 815, 1989.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 130, 1996.

Courant, R. and Robbins, H. "The Prime Numbers." §1 in Supplement to Ch. 1 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 21–31, 1996.

Davenport, H. *Multiplicative Number Theory, 2nd ed.* New York: Springer-Verlag, 1980.

Deutsch, E. "Problem 1494." *Math. Mag.* **69**, 143, 1996.

Dickson, L. E. "Factor Tables, Lists of Primes." Ch. 13 in *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, pp. 347–356, 1952.

Doster, D. Problem 10346. *Amer. Math. Monthly* **100**, 951, 1993.

Giblin, P. J. *Primes and Programming: Computers and Number Theory.* New York: Cambridge University Press, 1994.

Guy, R. K. "Conway's Prime Producing Machine." *Math. Mag.* **56**, 26–33, 1983.

Guy, R. K. "Prime Numbers," "Formulas for Primes," and "Products Taken Over Primes." Ch. A, §A17, and §B48 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 3–43, 36–41 and 102–103, 1994.

Hardy, G. H. Ch. 2 in *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, 1978.

Hardy, G. H. and Wright, E. M. "Prime Numbers" and "The Sequence of Primes." §1.2 and 1.4 in *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, pp. 1–4, 1979.

Honsberger, R. *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., p. 30, 1976.

Kraitchik, M. "Prime Numbers." §3.9 in *Mathematical Recreations.* New York: W. W. Norton, pp. 78–79, 1942.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, pp. 26, 30, and 46, 1983.

Moser, L. "Notes on Number Theory III. On the Sum of Consecutive Primes." *Can. Math. Bull.* **6**, 159–161, 1963.

Pappas, T. "Prime Numbers." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 100–101, 1989.
Ribenboim, P. *The Little Book of Big Primes.* New York: Springer-Verlag, 1991.
Ribenboim, P. *The New Book of Prime Number Records.* New York: Springer-Verlag, 1996.
Riesel, H. *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, 1994.
Schinzel, A. and Sierpiński, W. "Sur certains hypothèses concernant les nombres premiers." *Acta Arith.* **4**, 185–208, 1958.
Schinzel, A. and Sierpiński, W. Erratum to "Sur certains hypothèses concernant les nombres premiers." *Acta Arith.* **5**, 259, 1959.
Sloane, N. J. A. Sequences A046024, A000040/M0652, and A006510/M0679 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Wagon, S. "Primes Numbers." Ch. 1 in *Mathematica in Action.* New York: W. H. Freeman, pp. 11–37, 1991.
Zaiger, D. "The First 50 Million Prime Numbers." *Math. Intel.* **0**, 221–224, 1977.

Prime Number of Measurement

The set of numbers generated by excluding the SUMS of two or more consecutive earlier members is called the prime numbers of measurement, or sometimes the SEGMENTED NUMBERS. The first few terms are 1, 2, 4, 5, 8, 10, 14, 15, 16, 21, ... (Sloane's A002048). Excluding two *and* three terms gives the sequence 1, 2, 4, 5, 8, 10, 14, 15, 16, 19, 20, 21, ... (Sloane's A005242).

References

Guy, R. K. "MacMahon's Prime Numbers of Measurement." §E30 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 230–231, 1994.
Sloane, N. J. A. Sequence A002048/M0972 in "An On-Line Version of the Encyclopedia of Integer Sequences."0052420971

Prime Number Theorem

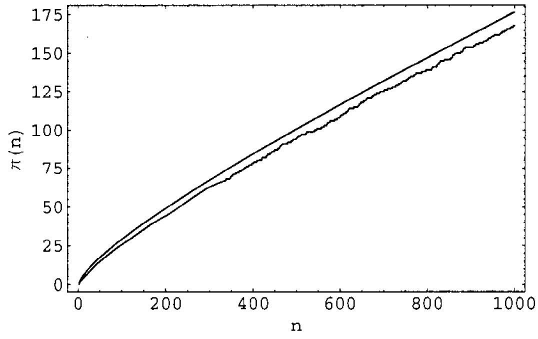

The theorem giving an asymptotic form for the PRIME COUNTING FUNCTION $\pi(n)$ for number of PRIMES less than some INTEGER n. Legendre (1808) suggested that, for large n,

$$\pi(n) \sim \frac{n}{A \ln n + B}, \tag{1}$$

with $A = 1$ and $B = -1.08366$ (where B is sometimes called LEGENDRE'S CONSTANT), a formula which is correct in the leading term only (Wagon 1991, pp. 28–29). In 1791, Gauss became the first to suggest instead

$$\pi(n) \sim \frac{n}{\ln n}. \tag{2}$$

Gauss later refined his estimate to

$$\pi(n) \sim \mathrm{Li}(n), \tag{3}$$

where $\mathrm{Li}(n)$ is the LOGARITHMIC INTEGRAL. This function has $n/\ln n$ as the leading term and has been shown to be a better estimate than $n/\ln n$ alone. The statement (3) is often known as "the" prime number theorem and was proved independently by Hadamard and Vallée Poussin in 1896. A plot of $\pi(n)$ (lower curve) and $\mathrm{Li}(n)$ is shown above for $n \leq 1000$.

For small n, it has been checked and always found that $\pi(n) < \mathrm{Li}(n)$. However, Skewes proved that the first crossing of $\pi(n) - \mathrm{Li}(n) = 0$ occurs before $10^{10^{10^{34}}}$ (the SKEWES NUMBER). The upper bound for the crossing has subsequently been reduced to 10^{371}. Littlewood (1914) proved that the INEQUALITY reverses infinitely often for sufficiently large n (Ball and Coxeter 1987). Lehman (1966) proved that at least 10^{500} reversals occur for numbers with 1166 or 1167 DECIMAL DIGITS.

Chebyshev (Rubinstein and Sarnak 1994) put limits on the RATIO

$$\frac{7}{8} < \frac{\pi(n)}{\frac{n}{\ln n}} < \frac{9}{8}, \tag{4}$$

and showed that if the LIMIT

$$\lim_{n \to \infty} \frac{\pi(n)}{\frac{n}{\ln n}} \tag{5}$$

existed, then it would be 1. This is, in fact, the prime number theorem.

Hadamard and Vallée Poussin proved the prime number theorem by showing that the RIEMANN ZETA FUNCTION $\zeta(z)$ has no zeros of the form $1 + it$ (Smith 1994, p. 128). In particular, Vallée Poussin showed that

$$\pi(x) = \mathrm{Li}(x) + \mathcal{O}\left(\frac{x}{\ln x} e^{-a\sqrt{\ln x}}\right) \tag{6}$$

for some constant a. A simplified proof was found by Selberg and Erdős (1949) (Ball and Coxeter 1987, p. 63).

Riemann estimated the PRIME COUNTING FUNCTION with

$$\pi(n) \sim \mathrm{Li}(n) - \tfrac{1}{2}\mathrm{Li}(n^{1/2}), \tag{7}$$

which is a better approximation than $\mathrm{Li}(n)$ for $n < 10^7$. Riemann (1859) also suggested the RIEMANN FUNCTION

$$R(x) = \sum_{n=1}^{\infty} \frac{\mu(n)}{n} \mathrm{Li}(x^{1/n}), \tag{8}$$

where μ is the MÖBIUS FUNCTION (Wagon 1991, p. 29). An even better approximation for small n (by a factor of 10 for $n < 10^9$) is the GRAM SERIES.

The prime number theorem is equivalent to

$$\lim_{x\to\infty} \frac{\psi(x)}{x} = 1, \tag{9}$$

where $\psi(x)$ is the SUMMATORY MANGOLDT FUNCTION.

The RIEMANN HYPOTHESIS is equivalent to the assertion that

$$|\operatorname{Li}(x) - \pi(x)| \leq c\sqrt{x}\ln x \tag{10}$$

for some value of c (Ingham 1932, Ball and Coxeter 1987). Some limits obtained without assuming the RIEMANN HYPOTHESIS are

$$\pi(x) = \operatorname{Li}(x) + \mathcal{O}[xe^{-(\ln x)^{1/2}/15}] \tag{11}$$

$$\pi(x) = \operatorname{Li}(x) + \mathcal{O}[xe^{-0.009(\ln x)^{3/5}/(\ln\ln x)^{1/5}}]. \tag{12}$$

Ramanujan showed that for sufficiently large x,

$$\pi^2(x) < \frac{ex}{\ln x}\pi\left(\frac{x}{e}\right). \tag{13}$$

The largest known PRIME for which the inequality fails is 38,358,837,677 (Berndt 1994, pp. 112–113). The related inequality

$$\operatorname{Li}^2(x) < \frac{ex}{\ln x}\operatorname{Li}\left(\frac{x}{e}\right) \tag{14}$$

is true for $x \geq 2418$ (Berndt 1994, p. 114).

see also BERTRAND'S POSTULATE, DIRICHLET'S THEOREM, GRAM SERIES, PRIME COUNTING FUNCTION, RIEMANN'S FORMULA, RIEMANN FUNCTION, RIEMANN-MANGOLDT FUNCTION, RIEMANN WEIGHTED PRIME-POWER COUNTING FUNCTION, SKEWES NUMBER

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 62–64, 1987.

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, 1994.

Courant, R. and Robbins, H. "The Prime Number Theorem." §1.2c in Supplement to Ch. 1 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 27–30, 1996.

Davenport, H. "Prime Number Theorem." Ch. 18 in *Multiplicative Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 111–114, 1980.

de la Vallée Poussin, C.-J. "Recherches analytiques la théorie des nombres premiers." *Ann. Soc. scient. Bruxelles* **20**, 183–256, 1896.

Hadamard, J. "Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques (')." *Bull. Soc. math. France* **24**, 199–220, 1896.

Hardy, G. H. and Wright, E. M. "Statement of the Prime Number Theorem." §1.8 in *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, pp. 9–10, 1979.

Ingham, A. E. *The Distribution of Prime Numbers.* London: Cambridge University Press, p. 83, 1932.

Legendre, A. M. *Essai sur la Théorie des Nombres.* Paris: Duprat, 1808.

Lehman, R. S. "On the Difference $\pi(x) - \operatorname{li}(x)$." *Acta Arith.* **11**, 397–410, 1966.

Littlewood, J. E. "Sur les distribution des nombres premiers." *C. R. Acad. Sci. Paris* **158**, 1869–1872, 1914.

Nagell, T. "The Prime Number Theorem." Ch. 8 in *Introduction to Number Theory.* New York: Wiley, 1951.

Riemann, G. F. B. "Über die Anzahl der Primzahlen unter einer gegebenen Grösse." *Monatsber. Königl. Preuss. Akad. Wiss. Berlin*, 671, 1859.

Rubinstein, M. and Sarnak, P. "Chebyshev's Bias." *Experimental Math.* **3**, 173–197, 1994.

Selberg, A. and Erdős, P. "An Elementary Proof of the Prime Number Theorem." *Ann. Math.* **50**, 305–313, 1949.

Shanks, D. "The Prime Number Theorem." §1.6 in *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 15–17, 1993.

Smith, D. E. *A Source Book in Mathematics.* New York: Dover, 1994.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 25–35, 1991.

Prime Pairs

see TWIN PRIMES

Prime Patterns Conjecture

see k-TUPLE CONJECTURE

Prime Polynomial

see PRIME-GENERATING POLYNOMIAL

Prime Power Conjecture

An Abelian planar DIFFERENCE SET of order n exists only for n a PRIME POWER. Gordon (1994) has verified it to be true for $n < 2,000,000$.

see also DIFFERENCE SET

References

Gordon, D. M. "The Prime Power Conjecture is True for $n < 2,000,000$." *Electronic J. Combinatorics* **1**, R6, 1–7, 1994. `http://www.combinatorics.org/Volume_1/volume1.html#R6`.

Prime Power Symbol

The symbol $p^e \| n$ means, for p a PRIME, that $p^e | n$, but $p^{e+1} \nmid n$.

Prime Quadratic Effect

Let $\pi_{m,n}(x)$ denote the number of PRIMES $\leq x$ which are congruent to n modulo m. Then one might expect that

$$\Delta(x) \equiv \pi_{4,3}(x) - \pi_{4,1}(x) \sim \tfrac{1}{2}\pi(x^{1/2}) > 0$$

(Berndt 1994). Although this is true for small numbers, Hardy and Littlewood showed that $\Delta(x)$ changes sign infinitely often. (The first number for which it is false is 26861.) The effect was first noted by Chebyshev in 1853, and is sometimes called the CHEBYSHEV PHENOMENON. It was subsequently studied by Shanks (1959), Hudson (1980), and Bays and Hudson (1977, 1978, 1979). The effect was also noted by Ramanujan, who incorrectly claimed that $\lim_{x\to\infty}\Delta(x)=\infty$ (Berndt 1994).

References

Bays, C. and Hudson, R. H. "The Mean Behavior of Primes in Arithmetic Progressions." *J. Reine Angew. Math.* **296**, 80–99, 1977.

Bays, C. and Hudson, R. H. "On the Fluctuations of Littlewood for Primes of the Form $4n\pm1$." *Math. Comput.* **32**, 281–286, 1978.

Bays, C. and Hudson, R. H. "Numerical and Graphical Description of All Axis Crossing Regions for the Moduli 4 and 8 which Occur Before 10^{12}." *Internat. J. Math. Math. Sci.* **2**, 111–119, 1979.

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 135–136, 1994.

Hudson, R. H. "A Common Principle Underlies Riemann's Formula, the Chebyshev Phenomenon, and Other Subtle Effects in Comparative Prime Number Theory. I." *J. Reine Angew. Math.* **313**, 133–150, 1980.

Shanks, D. "Quadratic Residues and the Distribution of Primes." *Math. Comput.* **13**, 272–284, 1959.

Prime Quadruplet

A PRIME CONSTELLATION of four successive PRIMES with minimal distance $(p, p+2, p+6, p+8)$. The quadruplet (2, 3, 5, 7) has smaller minimal distance, but it is an exceptional special case. With the exception of (5, 7, 11, 13), a prime quadruple must be of the form $(30n+11, 30n+13, 30n+17, 30n+19)$. The first few values of n which give prime quadruples are $n=0$, 3, 6, 27, 49, 62, 69, 108, 115, ... (Sloane's A014561), and the first few values of p are 5 (the exceptional case), 11, 101, 191, 821, 1481, 1871, 2081, 3251, 3461, ... The asymptotic FORMULA for the frequency of prime quadruples is analogous to that for other PRIME CONSTELLATIONS,

$$P_x(p,p+2,p+6,p+8)\sim\frac{27}{2}\prod_{p\geq5}\frac{p^3(p-4)}{(p-1)^4}\int_2^x\frac{dx}{(\ln x)^4}$$
$$=4.151180864\int_2^x\frac{dx}{(\ln x)^4},$$

where $c=4.15118\ldots$ is the Hardy-Littlewood constant for prime quadruplets. Roonguthai found the large prime quadruplets with

$$p=10^{99}+349781731$$
$$p=10^{199}+21156403891$$
$$p=10^{299}+140159459341$$
$$p=10^{399}+34993836001$$
$$p=10^{499}+883750143961$$
$$p=10^{599}+1394283756151$$
$$p=10^{699}+547634621251$$

(Roonguthai).

see also PRIME ARITHMETIC PROGRESSION, PRIME CONSTELLATION, PRIME k-TUPLES CONJECTURE, SEXY PRIMES, TWIN PRIMES

References

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* New York: Oxford University Press, 1979.

Forbes, T. "Prime k-tuplets." `http://www.ltkz.demon.co.uk/ktuplets.htm`.

Rademacher, H. *Lectures on Elementary Number Theory.* New York: Blaisdell, 1964.

Riesel, H. *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, pp. 61–62, 1994.

Roonguthai, W. "Large Prime Quadruplets." `http://www.mathsoft.com/asolve/constant/hrdyltl/roonguth.html`.

Sloane, N. J. A. Sequence A014561 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Prime Representation

Let $a\neq b$, A, and B denote POSITIVE INTEGERS satisfying

$$(a,b)=1 \qquad (A,B)=1,$$

(i.e., both pairs are RELATIVELY PRIME), and suppose every PRIME $p\equiv B \pmod{A}$ with $(p,2ab)=1$ is expressible if the form ax^2-by^2 for some INTEGERS x and y. Then every PRIME q such that $q\equiv -B \pmod{A}$ and $(q,2ab)=1$ is expressible in the form bX^2-aY^2 for some INTEGERS X and Y (Halter-Koch 1993, Williams 1991).

Prime Form	Representation
$4n+1$	x^2+y^2
$8n+1, 8n+3$	x^2+2y^2
$8n\pm1$	x^2-2y^2
$6n+1$	x^2+3y^2
$12n+1$	x^2-3y^2
$20n+1, 20n+9$	x^2+5y^2
$10n+1, 10n+9$	x^2-5y^2
$14n+1, 14n+9, 14n+25$	x^2+7y^2
$28n+1, 28n+9, 28n+25$	x^2-7y^2
$30n+1, 30n+49$	x^2+15y^2
$60n+1, 60n+49$	x^2-15y^2
$30n-7, 30n+17$	$5x^2+3y^2$
$60n-7, 60n+17$	$5x^2-3y^2$
$24n+1, 24n+7$	x^2+6y^2
$24n+1, 24n+19$	x^2-6y^2
$24n+5, 24n+11$	$2x^2+3y^2$
$24n+5, 24n-1$	$2x^2-3y^2$

References

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 70–73, 1994.

Halter-Koch, F. "A Theorem of Ramanujan Concerning Binary Quadratic Forms." *J. Number. Theory* **44**, 209–213, 1993.

Williams, K. S. "On an Assertion of Ramanujan Concerning Binary Quadratic Forms." *J. Number Th.* **38**, 118–133, 1991.

Prime Ring

A RING for which the product of any pair of IDEALS is zero only if one of the two IDEALS is zero. All SIMPLE RINGS are prime.

see also IDEAL, RING, SEMIPRIME RING, SIMPLE RING

Prime Sequence

see PRIME ARITHMETIC PROGRESSION, PRIME ARRAY, PRIME-GENERATING POLYNOMIAL, SIERPIŃSKI'S PRIME SEQUENCE THEOREM

Prime Spiral

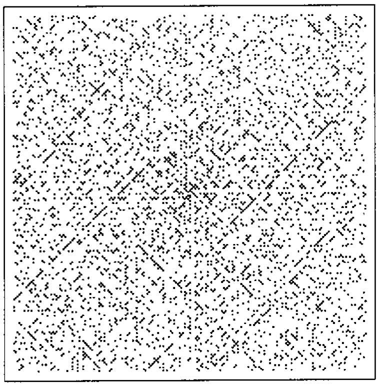

The numbers arranged in a SPIRAL

$$\begin{array}{ccc} 5 & 4 & 3 \\ 6 & 1 & 2 \\ 7 & 8 & 9 \end{array}$$

with PRIMES indicated in black, as first drawn by S. Ulam. Unexpected patterns of diagonal lines are apparent in such a plot, as illustrated in the above 199×199 grid.

References

Dewdney, A. K. "Computer Recreations: How to Pan for Primes in Numerical Gravel." *Sci. Amer.* **259**, 120–123, July 1988.

Lane, C. "Prime Spiral." `http://www.best.com/~cdl/PrimeSpiralApplet.html`.

Weisstein, E. W. "Prime Spiral." `http://www.astro.virginia.edu/~eww6n/math/notebooks/PrimeSpiral.m`.

Prime String

Call a number n a prime string from the left if n and all numbers obtained by successively removing the right-most DIGIT are PRIME. There are 83 left prime strings in base 10. The first few are 2, 3, 5, 7, 23, 29, 31, 37, 53, 59, 71, 73, 79, 233, 239, 293, 311, 313, 317, 373, 379, 593, 599, ... (Sloane's A024770), the largest being 73,939,133. Similarly, call a number n a prime string from the right if n and all numbers obtained by successively removing the left-most DIGIT are PRIME. The first few are 2, 3, 5, 7, 13, 17, 23, 37, 43, 47, 53, 67, 73, 83, 97, 103, 107, 113, 137, 167, 173, ... (Sloane's A033664). A large right prime string is 933,739,397.

see also PRIME ARRAY, PRIME NUMBER

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Rivera, C. "Problems & Puzzles (Puzzles): Prime Strings." `http://www.sci.net.mx/~crivera/ppp/puzz_002.htm`.

Sloane, N. J. A. Sequence A024770 in "An On-Line Version of the Encyclopedia of Integer Sequences."033664

Prime Sum

Let

$$\Sigma(n) \equiv \sum_{i=1}^{n} p_i$$

be the sum of the first n PRIMES. The first few terms are 2, 5, 10, 17, 28, 41, 58, 77, ... (Sloane's A007504). Bach and Shallit (1996) show that

$$\Sigma(n) \sim \frac{n^2}{2\log n},$$

and provide a general technique for estimating such sums.

see also PRIMORIAL

References

Bach, E. and Shallit, J. §2.7 in *Algorithmic Number Theory, Vol. 1: Efficient Algorithms.* Cambridge, MA: MIT Press, 1996.

Sloane, N. J. A. Sequence A007504/M1370 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Prime Theta Function

The prime theta function is defined as

$$\theta(n) \equiv \sum_{i=1}^{n} \ln p_i,$$

where p_i is the ith PRIME. As shown by Bach and Shallit (1996),

$$\theta(n) \sim n.$$

References

Bach, E. and Shallit, J. *Algorithmic Number Theory, Vol. 1: Efficient Algorithms.* Cambridge, MA: MIT Press, pp. 206 and 233, 1996.

Prime Triangle

$$\begin{array}{c} * \\ 1\ \ 2 \\ 1\ \ 2\ \ 3 \\ 1\ \ 2\ \ 3\ \ 4 \\ 1\ \ 4\ \ 3\ \ 2\ \ 5 \\ 1\ \ 4\ \ 3\ \ 2\ \ 5\ \ 6 \end{array}$$

This triangle has rows beginning with 1 and ending with n, with the SUM of each two consecutive entries being a PRIME.

see also PASCAL'S TRIANGLE

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 106, 1994.

Kenney, M. J. "Student Math Notes." *NCTM News Bulletin.* Nov. 1986.

Prime Unit

1 and −1 are the only INTEGERS which divide every INTEGER. They are therefore called the prime units.

see also INTEGER, PRIME NUMBER, UNIT

Prime Zeta Function

The prime zeta function

$$P(n) \equiv \sum_p \frac{1}{p^n}, \qquad (1)$$

where the sum is taken over PRIMES is a generalization of the RIEMANN ZETA FUNCTION

$$\zeta(n) \equiv \sum_{k=1}^{\infty} \frac{1}{k^n}, \qquad (2)$$

where the sum is over *all* integers. The prime zeta function can be expressed in terms of the RIEMANN ZETA FUNCTION by

$$\begin{aligned}\ln \zeta(n) &= -\sum_{p\geq 2} \ln(1-p^{-n}) = \sum_{p\geq 2}\sum_{k=1}^{\infty} \frac{p^{-kn}}{k} \\ &= \sum_{k=1}^{\infty} \frac{1}{k} \sum_{p\geq 2} p^{-kn} = \sum_{k=1}^{\infty} \frac{P(kn)}{k}. \qquad (3)\end{aligned}$$

Inverting then gives

$$P(n) = \sum_{k=1}^{\infty} \frac{\mu(k)}{k} \ln \zeta(kn), \qquad (4)$$

where $\mu(k)$ is the MÖBIUS FUNCTION. The values for the first few integers starting with two are

$$\begin{aligned} P(2) &\approx 0.452247 \qquad (5)\\ P(3) &\approx 0.174763 \qquad (6)\\ P(4) &\approx 0.0769931 \qquad (7)\\ P(5) &\approx 0.035755. \qquad (8)\end{aligned}$$

see also MÖBIUS FUNCTION, RIEMANN ZETA FUNCTION, ZETA FUNCTION

References

Hardy, G. H. and Weight, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Oxford University Press, pp. 355–356, 1979.

Primequad

see PRIME QUADRUPLET

Primitive Abundant Number

An ABUNDANT NUMBER for which all PROPER DIVISORS are DEFICIENT is called a primitive abundant number (Guy 1994, p. 46). The first few ODD primitive abundant numbers are 945, 1575, 2205, 3465, ... (Sloane's A006038).

see also ABUNDANT NUMBER, DEFICIENT NUMBER, HIGHLY ABUNDANT NUMBER, SUPERABUNDANT NUMBER, WEIRD NUMBER

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 46, 1994.
Sloane, N. J. A. Sequence A006038/M5486 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Primitive Function

see INTEGRAL

Primitive Irreducible Polynomial

An IRREDUCIBLE POLYNOMIAL which generates all elements of an extension field from a base field. For any PRIME or PRIME POWER q and any POSITIVE INTEGER n, there exists a primitive irreducible POLYNOMIAL of degree n over GF(q).

see also GALOIS FIELD, IRREDUCIBLE POLYNOMIAL

Primitive Polynomial Modulo 2

A special type of POLYNOMIAL of which a subclass has COEFFICIENTS of only 0 or 1. Such POLYNOMIALS define a RECURRENCE RELATION which can be used to obtain a new RANDOM bit from the n preceding ones.

Primitive Prime Factor

If $n \geq 1$ is the smallest INTEGER such that $p | a^n - b^n$ (or $a^n + b^n$), then p is a primitive prime factor.

Primitive Pseudoperfect Number

see PRIMITIVE SEMIPERFECT NUMBER

Primitive Recursive Function

For-loops (which have a fixed iteration limit) are a special case of while-loops. A function which can be implemented using only for-loops is called primitive recursive. (In contrast, a COMPUTABLE FUNCTION can be coded using a combination of for- and while-loops, or while-loops only.)

The ACKERMANN FUNCTION is the simplest example of a well-defined TOTAL FUNCTION which is COMPUTABLE but not primitive recursive, providing a counterexample to the belief in the early 1900s that every COMPUTABLE FUNCTION was also primitive recursive (Dötzel 1991).

see also ACKERMANN FUNCTION, COMPUTABLE FUNCTION, TOTAL FUNCTION

References

Dötzel, G. "A Function to End All Functions." *Algorithm: Recreational Programming* **2**, 16–17, 1991.

Primitive Root

A number g is a primitive root of m if

$$g^k \not\equiv 1 \pmod{m} \tag{1}$$

for $1 \leq k < m$ and

$$g^m \equiv 1 \pmod{m}. \tag{2}$$

Only $m = 2$, 4, p^a, and $2p^a$ have primitive roots (where $p > 2$ and a is an INTEGER). For composite m, there may be more than one primitive root (both 3 and 7 are primitive roots mod 10), but for prime p, there is only one primitive root. It is the INTEGER g satisfying $1 \leq g \leq p - 1$ such that $g \pmod{p}$ has ORDER $p - 1$.

The primitive root of m can also be defined as a cyclic generator of the multiplicative group (mod m) when m is a prime POWER or twice a PRIME POWER. Let p be any ODD PRIME $k \geq 1$, and let

$$s \equiv \sum_{j=1}^{p-1} j^k. \tag{3}$$

Then

$$s = \begin{cases} -1 \pmod{p} & \text{for } p-1 | k \\ 0 \pmod{p} & \text{for } p-1 \nmid k. \end{cases} \tag{4}$$

For numbers m with primitive roots, all y satisfying $(p, y) = 1$ are representable as

$$y \equiv g^t \pmod{m}, \tag{5}$$

where $t = 0, 1, \ldots, \phi(m) - 1$, t is known as the index, and y is an INTEGER. Kearnes showed that for any POSITIVE INTEGER m, there exist infinitely many PRIMES p such that

$$m < g_p < p - m. \tag{6}$$

Call the least primitive root g_p. Burgess (1962) proved that

$$g_p \leq C p^{1/4+\epsilon} \tag{7}$$

for C and ϵ POSITIVE constants and p sufficiently large.

The table below gives the primitive roots (for prime $m = p$; Sloane's A001918) and least primitive roots (for composite m) for the first few INTEGERS

m	g	m	g	m	g
2	1	53	2	134	7
3	2	54	5	137	3
4	3	58	3	139	2
5	2	59	2	142	7
6	5	61	2	146	5
7	3	62	3	149	2
9	2	67	2	151	6
10	3	71	7	157	5
11	2	73	5	158	3
13	2	74	5	162	5
14	3	79	3	163	2
17	3	81	2	166	5
18	5	82	7	167	5
19	2	83	2	169	2
22	7	86	3	173	2
23	5	89	3	178	3
25	2	94	5	179	2
26	7	97	5	181	2
27	2	98	3	191	19
29	2	101	2	193	5
31	3	103	5	194	5
34	3	106	3	197	2
37	2	107	2	199	3
38	3	109	6	202	3
41	6	113	3	206	5
43	3	118	11	211	2
46	5	121	2	214	5
47	5	122	7	218	11
49	3	125	2	223	3
50	3	127	3	226	3
		131	2	227	2

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Primitive Roots." §24.3.4 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 827, 1972.

Guy, R. K. "Primitive Roots." §F9 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 248–249, 1994.

Sloane, N. J. A. Sequence A001918/M0242 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Primitive Root of Unity

A number r is an nth ROOT OF UNITY if $r^n = 1$ and a primitive nth root of unity if, in addition, n is the smallest INTEGER of $k = 1, \ldots, n$ for which $r^k = 1$.

see also ROOT OF UNITY

Primitive Semiperfect Number

A SEMIPERFECT NUMBER for which none of its PROPER DIVISORS are pseudoperfect (Guy 1994, p. 46). The first few are 6, 20, 28, 88, 104, 272 ... (Sloane's A006036). Primitive pseudoperfect numbers are also called IRREDUCIBLE SEMIPERFECT NUMBERS. There are infinitely many primitive pseudoperfect numbers which are not HARMONIC DIVISOR NUMBERS, and infinitely many ODD primitive semiperfect numbers.

see also HARMONIC DIVISOR NUMBER, SEMIPERFECT NUMBER

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 46, 1994.

Sloane, N. J. A. Sequence A006036/M4133 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Primitive Sequence

A SEQUENCE in which no term DIVIDES any other.

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 202, 1994.

Primorial

For a PRIME p,

$$\text{primorial}(p_i) = p_i\# \equiv \prod_{j=1}^{i} p_j,$$

where p_i is the ith PRIME. The first few values for $p_i\#$, are 2, 6, 30, 210, 2310, 30030, 510510, ... (Sloane's A002110).

$p\# - 1$ is PRIME for PRIMES $p = 3$, 5, 11, 41, 89, 317, 337, 991, 1873, 2053, 2377, 4093, 4297, ... (Sloane's A014563; Guy 1994), or p_n for $n = 2$, 3, 5, 13, 24, 66, 68, 167, 287, 310, 352, 564, 590, ... $p\# + 1$ is known to be PRIME for the PRIMES $p = 2$, 3, 5, 7, 11, 31, 379, 1019, 1021, 2657, 3229, 4547, 4787, 11549, ... (Sloane's A005234; Guy 1994, Mudge 1997), or p_n for $n = 1$, 2, 3, 4, 5, 11, 75, 171, 172, 384, 457, 616, 643, 1391, ... Both forms have been tested to $p = 25000$ (Caldwell 1995). It is not known if there are an infinite number of PRIMES for which $p\# + 1$ is PRIME or COMPOSITE (Ribenboim 1989).

see also FACTORIAL, FORTUNATE PRIME, PRIME SUM SMARANDACHE NEAR-TO-PRIMORIAL FUNCTION, TWIN PEAKS

References

Borning, A. "Some Results for $k! + 1$ and $2 \cdot 3 \cdot 5 \cdot p + 1$." *Math. Comput.* **26**, 567–570, 1972.

Buhler, J. P.; Crandall, R. E.; and Penk, M. A. "Primes of the form $M! + 1$ and $\cdot 3 \cdot 5 \cdot p + 1$." *Math. Comput.* **38**, 639–643, 1982.

Caldwell, C. "On The Primality of $n! \pm 1$ and $2 \cdot 3 \cdot 5 \cdots p \pm 1$." *Math. Comput.* **64**, 889–890, 1995.

Dubner, H. "Factorial and Primorial Primes." *J. Rec. Math.* **19**, 197–203, 1987.

Dubner, H. "A New Primorial Prime." *J. Rec. Math.* **21**, 276, 1989.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 7–8, 1994.

Leyland, P. `ftp://sable.ox.ac.uk/pub/math/factors/primorial-.Z` and `primorial+.Z`.

Mudge, M. "Not Numerology but Numeralogy!" *Personal Computer World,* 279–280, 1997.

Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, p. 4, 1989.

Sloane, N. J. A. Sequences A014563, A002110/M1691, and A005234/M0669 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Temper, M. "On the Primality of $k! + 1$ and $\cdot 3 \cdot 5 \cdots p + 1$." *Math. Comput.* **34**, 303–304, 1980.

Prince Rupert's Cube

The largest CUBE which can be made to pass through a given CUBE. (In other words, the CUBE having a side length equal to the side length of the largest HOLE of a SQUARE CROSS-SECTION which can be cut through a unit CUBE without splitting it into two pieces.) The Prince Rupert's cube has side length $3\sqrt{2}/4 = 1.06065\ldots$, and any CUBE this size or smaller can be made to pass through the original CUBE.

see also CUBE, SQUARE

References

Cundy, H. and Rollett, A. "Prince Rupert's Cubes." §3.15.2 in *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 157–158, 1989.

Schrek, D. J. E. "Prince Rupert's Problem and Its Extension by Pieter Nieuwland." *Scripta Math.* **16**, 73–80 and 261–267, 1950.

Principal

The original amount borrowed or lent on which INTEREST is then paid or given.

see also INTEREST

Principal Curvatures

The MAXIMUM and MINIMUM of the NORMAL CURVATURE κ_1 and κ_2 at a given point on a surface are called the principal curvatures. The principal curvatures measure the MAXIMUM and MINIMUM bending of a REGULAR SURFACE at each point. The GAUSSIAN CURVATURE K and MEAN CURVATURE H are related to κ_1 and κ_2 by

$$K = \kappa_1 \kappa_2 \tag{1}$$

$$H = \tfrac{1}{2}(\kappa_1 + \kappa_2). \tag{2}$$

This can be written as a QUADRATIC EQUATION

$$\kappa^2 - 2H\kappa + K = 0, \tag{3}$$

which has solutions

$$\kappa_1 = H + \sqrt{H^2 - K} \tag{4}$$

$$\kappa_2 = H - \sqrt{H^2 - K}. \tag{5}$$

see also GAUSSIAN CURVATURE, MEAN CURVATURE, NORMAL CURVATURE, NORMAL SECTION, PRINCIPAL DIRECTION, PRINCIPAL RADIUS OF CURVATURE, RODRIGUES'S CURVATURE FORMULA

References

Geometry Center. "Principal Curvatures." `http://www.geom.umn.edu/zoo/diffgeom/surfspace/concepts/curvatures/prin-curv.html`.

Gray, A. "Normal Curvature." §14.2 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 270–273, 277, and 283, 1993.

Principal Curve

A curve $\boldsymbol{\alpha}$ on a REGULAR SURFACE M is a principal curve IFF the velocity $\boldsymbol{\alpha}'$ always points in a PRINCIPAL DIRECTION, i.e.,

$$S(\boldsymbol{\alpha}') = \kappa_i \boldsymbol{\alpha}',$$

where S is the SHAPE OPERATOR and κ_i is a PRINCIPAL CURVATURE. If a SURFACE OF REVOLUTION generated by a plane curve is a REGULAR SURFACE, then the MERIDIANS and PARALLELS are principal curves.

References

Gray, A. "Principal Curves" and "The Differential Equation for the Principal Curves." §18.1 and 21.1 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 410–413, 1993.

Principal Direction

The directions in which the PRINCIPAL CURVATURES occur.

see also PRINCIPAL DIRECTION

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 270, 1993.

Principal Ideal

An IDEAL I of a RING R is called principal if there is an element a of R such that

$$I = aR = \{ar : r \in R\}.$$

In other words, the IDEAL is generated by the element a. For example, the IDEALS $n\mathbb{Z}$ of the RING of INTEGERS $\mathbb{Z}$ are all principal, and in fact all IDEALS of $\mathbb{Z}$ are principal.

see also IDEAL, RING

Principal Normal Vector

see NORMAL VECTOR

Principal Quintic Form

A general QUINTIC EQUATION

$$a_5x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = 0 \quad (1)$$

can be reduced to one of the form

$$y^5 + b_2y^2 + b_1y + b_0 = 0, \quad (2)$$

called the principal quintic form.

NEWTON'S RELATIONS for the ROOTS y_j in terms of the b_js is a linear system in the b_j, and solving for the b_js expresses them in terms of the POWER sums $s_n(y_j)$. These POWER sums can be expressed in terms of the a_js, so the b_js can be expressed in terms of the a_js. For a quintic to have no quartic or cubic term, the sums of the ROOTS and the sums of the SQUARES of the ROOTS vanish, so

$$s_1(y_j) = 0 \quad (3)$$

$$s_2(y_j) = 0. \quad (4)$$

Assume that the ROOTS y_j of the new quintic are related to the ROOTS x_j of the original quintic by

$$y_j = {x_j}^2 + \alpha x_j + \beta. \quad (5)$$

Substituting this into (1) then yields two equations for α and β which can be multiplied out, simplified by using NEWTON'S RELATIONS for the POWER sums in the x_j, and finally solved. Therefore, α and β can be expressed using RADICALS in terms of the COEFFICIENTS a_j. Again by substitution into (4), we can calculate $s_3(y_j)$, $s_4(y_j)$ and $s_5(y_j)$ in terms of α and β and the x_j. By the previous solution for α and β and again by using NEWTON'S RELATIONS for the POWER sums in the x_j, we can ultimately express these POWER sums in terms of the a_j.

see also BRING QUINTIC FORM, NEWTON'S RELATIONS, QUINTIC EQUATION

Principal Radius of Curvature

Given a 2-D SURFACE, there are two "principal" RADII OF CURVATURE. The larger is denoted R_1, and the smaller R_2. These are PERPENDICULAR to each other, and both PERPENDICULAR to the tangent PLANE of the surface.

see also GAUSSIAN CURVATURE, MEAN CURVATURE, RADIUS OF CURVATURE

Principal Value

see CAUCHY PRINCIPAL VALUE

Principal Vector

A tangent vector $\mathbf{v_p} = v_1\mathbf{x}_u + v_2\mathbf{x}_v$ is a principal vector IFF

$$\det\begin{bmatrix} {v_2}^2 & -v_1v_2 & {v_1}^2 \\ E & F & G \\ e & f & g \end{bmatrix} = 0,$$

where e, f, and g are coefficients of the first FUNDAMENTAL FORM and E, F, G of the second FUNDAMENTAL FORM.

see also FUNDAMENTAL FORMS, PRINCIPAL CURVE

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 410, 1993.

Principal Vertex

A VERTEX x_i of a SIMPLE POLYGON P is a principal VERTEX if the diagonal $[x_{i-1}, x_{i+1}]$ intersects the boundary of P only at x_{i-1} and x_{i+1}.

see also EAR, MOUTH

References
Meisters, G. H. "Polygons Have Ears." *Amer. Math. Monthly* **82**, 648–751, 1975.
Meisters, G. H. "Principal Vertices, Exposed Points, and Ears." *Amer. Math. Monthly* **87**, 284–285, 1980.
Toussaint, G. "Anthropomorphic Polygons." *Amer. Math. Monthly* **98**, 31–35, 1991.

Principle

A loose term for a true statement which may be a POSTULATE, THEOREM, etc.

see also AREA PRINCIPLE, ARGUMENT PRINCIPLE, AXIOM, CAVALIERI'S PRINCIPLE, CONJECTURE, CONTINUITY PRINCIPLE, COUNTING GENERALIZED PRINCIPLE, DIRICHLET'S BOX PRINCIPLE, DUALITY PRINCIPLE, DUHAMEL'S CONVOLUTION PRINCIPLE, EUCLID'S PRINCIPLE, FUBINI PRINCIPLE, HASSE PRINCIPLE, INCLUSION-EXCLUSION PRINCIPLE, INDIFFERENCE PRINCIPLE, INDUCTION PRINCIPLE, INSUFFICIENT REASON PRINCIPLE, LEMMA, LOCAL-GLOBAL PRINCIPLE, MULTIPLICATION PRINCIPLE, PERMANENCE OF MATHEMATICAL RELATIONS PRINCIPLE, PONCELET'S CONTINUITY PRINCIPLE, PONTRYAGIN MAXIMUM PRINCIPLE, PORISM, POSTULATE, SCHWARZ REFLECTION PRINCIPLE, SUPERPOSITION PRINCIPLE, SYMMETRY PRINCIPLE, THEOREM, THOMSON'S PRINCIPLE, TRIANGLE TRANSFORMATION PRINCIPLE, WELL-ORDERING PRINCIPLE

Pringsheim's Theorem

Let $C^\omega(I)$ be the set of real ANALYTIC FUNCTIONS on I. Then $C^\omega(I)$ is a SUBALGEBRA of $C^\infty(I)$. A NECESSARY and SUFFICIENT condition for a function $f \in C^\infty(I)$ to belong to $C^\omega(I)$ is that

$$|f^{(n)}(x)| \leq k^n n!$$

for $n = 0, 1, \ldots$ for a suitable constant k.

see also ANALYTIC FUNCTION, SUBALGEBRA

References
Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 207, 1980.

Printer's Errors

Typesetting "errors" in which exponents or multiplication signs are omitted but the resulting expression is equivalent to the original one. Examples include

$$2^5 9^2 = 2592$$

$$3^4 425 = 34425$$

$$31^2 325 = 312325$$

$$2^5 \cdot \tfrac{25}{31} = 25\tfrac{25}{31},$$

where a whole number followed by a fraction is interpreted as addition (e.g., $1\frac{1}{2} = 1 + \frac{1}{2} = \frac{3}{2}$).

see also ANOMALOUS CANCELLATION

References
Dudeney, H. E. *Amusements in Mathematics.* New York: Dover, 1970.
Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 174–175, 1979.

Prior Distribution

see BAYESIAN ANALYSIS

Prism

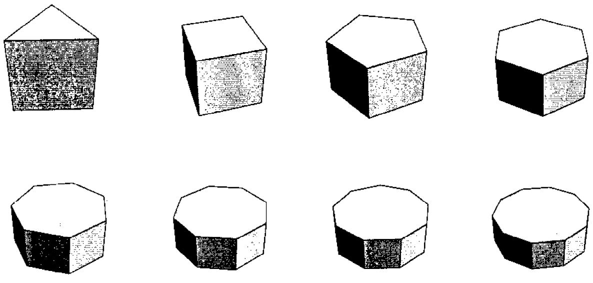

A POLYHEDRON with two congruent POLYGONAL faces and all remaining faces PARALLELOGRAMS. The 3-prism is simply the CUBE. The simple prisms and antiprisms include: decagonal antiprism, decagonal prism, hexagonal antiprism, hexagonal prism, octagonal antiprism, octagonal prism, pentagonal antiprism, pentagonal prism, square antiprism, and triangular prism. The DUAL POLYHEDRON of a simple (Archimedean) prism is a BIPYRAMID.

The triangular prism, square prism (cube), and hexagonal prism are all SPACE-FILLING POLYHEDRA.

see also ANTIPRISM, AUGMENTED HEXAGONAL PRISM, AUGMENTED PENTAGONAL PRISM, AUGMENTED TRIANGULAR PRISM, BIAUGMENTED PENTAGONAL PRISM, BIAUGMENTED TRIANGULAR PRISM, CUBE, METABIAUGMENTED HEXAGONAL PRISM, PARABIAUGMENTED HEXAGONAL PRISM, PRISMATOID, PRISMOID, TRAPEZOHEDRON, TRIAUGMENTED HEXAGONAL PRISM, TRIAUGMENTED TRIANGULAR PRISM

References
Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 127, 1987.
Cromwell, P. R. *Polyhedra.* New York: Cambridge University Press, pp. 85–86, 1997.
Weisstein, E. W. "Prisms and Antiprisms." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Prism.m`.

Prismatic Ring

A MÖBIUS STRIP with finite width.

see also MÖBIUS STRIP

References
Gardner, M. "Twisted Prismatic Rings." Ch. 5 in *Fractal Music, Hypercards, and More Mathematical Recreations from Scientific American Magazine.* New York: W. H. Freeman, 1992.

Prismatoid

A POLYHEDRON having two POLYGONS in PARALLEL planes as bases and TRIANGULAR or TRAPEZOIDAL lateral faces with one side lying in one base and the opposite VERTEX or side lying in the other base. Examples include the CUBE, PYRAMIDAL FRUSTUM, RECTANGULAR PARALLELEPIPED, PRISM, and PYRAMID. Let A_1 be the AREA of the lower base, A_2 the AREA of the upper base, M the AREA of the midsection, and h the ALTITUDE. Then

$$V = \tfrac{1}{6}h(A_1 + 4M + A_2).$$

see also GENERAL PRISMATOID, PRISMOID

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 128 and 132, 1987.

Prismoid

A PRISMATOID having planar sides and the same number of vertices in both of its parallel planes. The faces of a prismoid are therefore either TRAPEZOIDS or PARALLELOGRAMS. Ball and Coxeter (1987) use the term to describe an ANTIPRISM.

see also ANTIPRISM, PRISM, PRISMATOID

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 130, 1987.

Prisoner's Dilemma

A problem in GAME THEORY first discussed by A. Tucker. Suppose each of two prisoners A and B, who are not allowed to communicate with each other, is offered to be set free if he implicates the other. If neither implicates the other, both will receive the usual sentence. However, if the prisoners implicate each other, then both are presumed guilty and granted harsh sentences.

A DILEMMA arises in deciding the best course of action in the absence of knowledge of the other prisoner's decision. Each prisoner's best strategy would appear to be to turn the other in (since if A makes the worst-case assumption that B will turn him in, then B will walk free and A will be stuck in jail if he remains silent). However, if the prisoners turn each other in, they obtain the worst possible outcome for both.

see also DILEMMA, TIT-FOR-TAT

References

Axelrod, R. *The Evolution of Cooperation* New York: BasicBooks, 1985.

Goetz, P. "Phil's Good Enough Complexity Dictionary." `http://www.cs.buffalo.edu/~goetz/dict.html`.

Prizes

see MATHEMATICS PRIZES

Probability

Probability is the branch of mathematics which studies the possible outcomes of given events together with their relative likelihoods and distributions. In common usage, the word "probability" is used to mean the chance that a particular event (or set of events) will occur expressed on a linear scale from 0 (impossibility) to 1 (certainty), also expressed as a PERCENTAGE between 0 and 100%. The analysis of events governed by probability is called STATISTICS.

There are several competing interpretations of the actual "meaning" of probabilities. Frequentists view probability simply as a measure of the frequency of outcomes (the more conventional interpretation), while BAYESIANS treat probability more subjectively as a statistical procedure which endeavors to estimate parameters of an underlying distribution based on the observed distribution.

A properly normalized function which assigns a probability "density" to each possible outcome within some interval is called a PROBABILITY FUNCTION, and its cumulative value (integral for a continuous distribution or sum for a discrete distribution) is called a DISTRIBUTION FUNCTION.

Probabilities are defined to obey certain assumptions, called the PROBABILITY AXIOMS. Let a SAMPLE SPACE contain the UNION ($\cup$) of all possible events E_i, so

$$S \equiv \left(\bigcup_{i=1}^{N} E_i \right), \tag{1}$$

and let E and F denote subsets of S. Further, let $F' =$ not-F be the complement of F, so that

$$F \cup F' = S. \tag{2}$$

Then the set E can be written as

$$E = E \cap S = E \cap (F \cup F') = (E \cap F) \cup (E \cap F'), \tag{3}$$

where $\cap$ denotes the intersection. Then

$$\begin{aligned} P(E) &= P(E \cap F) + P(E \cap F') - P[(E \cap F) \cap (E \cap F')] \\ &= P(E \cap F) + P(E \cap F') - P[(F \cap F') \cap (E \cap E)] \\ &= P(E \cap F) + P(E \cap F') - P(\varnothing \cap E) \\ &= P(E \cap F) + P(E \cap F') - P(\varnothing) \\ &= P(E \cap F) + P(E \cap F'), \end{aligned} \tag{4}$$

where $\varnothing$ is the EMPTY SET.

Let $P(E|F)$ denote the CONDITIONAL PROBABILITY of E given that F has already occurred, then

$$\begin{aligned} P(E) &= P(E|F)P(F) + P(E|F')P(F') & (5)\\ &= P(E|F)P(F) + P(E|F')[1 - P(F)] & (6)\\ P(A \cap B) &= P(A)P(B|A) & (7)\\ &= P(B)P(A|B) & (8)\\ P(A' \cap B) &= P(A')P(B|A') & (9)\\ P(E|F) &= \frac{P(E \cap F)}{P(F)} & (10)\end{aligned}$$

A very important result states that

$$P(E \cup F) = P(E) + P(F) - P(E \cap F), \qquad (11)$$

which can be generalized to

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_i P(A_i) - \sum_{ij}{}' P(A_i \cup A_j)$$
$$+ \sum_{ijk}{}'' P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right). \qquad (12)$$

see also BAYES' FORMULA, CONDITIONAL PROBABILITY, DISTRIBUTION, DISTRIBUTION FUNCTION, LIKELIHOOD, PROBABILITY AXIOMS, PROBABILITY FUNCTION, PROBABILITY INEQUALITY, STATISTICS

Probability Axioms

Given an event E in a SAMPLE SPACE S which is either finite with N elements or countably infinite with $N = \infty$ elements, then we can write

$$S \equiv \left(\bigcup_{i=1}^{N} E_i\right),$$

and a quantity $P(E_i)$, called the PROBABILITY of event E_i, is defined such that

1. $0 \leq P(E_i) < 1$.
2. $P(S) = 1$.
3. Additivity: $P(E_1 \cup E_2) = P(E_1) + P(E_2)$, where E_1 and E_2 are mutually exclusive.
4. Countable additivity: $P\left(\cup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$ for $n = 1, 2, \ldots, N$ where $E_1, E_2, \ldots$ are mutually exclusive (i.e., $E_1 \cap E_2 = \varnothing$).

see also SAMPLE SPACE, UNION

Probability Density Function

see PROBABILITY FUNCTION

Probability Distribution Function

see PROBABILITY FUNCTION

Probability Function

The probability density function $P(x)$ (also called the PROBABILITY DENSITY FUNCTION) of a continuous distribution is defined as the derivative of the (cumulative) DISTRIBUTION FUNCTION $D(x)$,

$$D'(x) = [P(x)]_{-\infty}^{x} = P(x) - P(-\infty) = P(x), \qquad (1)$$

so

$$D(x) = P(X \leq x) \equiv \int_{-\infty}^{x} P(y)\, dy. \qquad (2)$$

A probability density function satisfies

$$P(x \in B) = \int_B P(x)\, dx \qquad (3)$$

and is constrained by the normalization condition,

$$P(-\infty < x < \infty) = \int_{-\infty}^{\infty} P(x)\, dx \equiv 1. \qquad (4)$$

Special cases are

$$\begin{aligned} P(a \leq x \leq b) &= \int_a^b P(x)\, dx & (5)\\ P(a \leq x \leq a + da) &= \int_a^{a+da} P(x)\, dx \approx P(a)\, da & (6)\\ P(x = a) &= \int_a^a P(x)\, dx = 0. & (7)\end{aligned}$$

If $u = u(x, y)$ and $v = v(x, y)$, then

$$P_{u,v}(u, v) = P_{x,y}(x, y) \left|\frac{\partial(x, y)}{\partial(u, v)}\right|. \qquad (8)$$

Given the MOMENTS of a distribution (μ, σ, and the GAMMA STATISTICS γ_r), the asymptotic probability function is given by

$$\begin{aligned} P(x) = {} & Z(x)\\ & -[\tfrac{1}{6}\gamma_1 Z^{(3)}(x)] + [\tfrac{1}{24}\gamma_2 Z^{(4)}(x) + \tfrac{1}{72}{\gamma_1}^2 Z^{(6)}(x)]\\ & -[\tfrac{1}{120}\gamma_3 Z^{(5)}(x) + \tfrac{1}{144}\gamma_1\gamma_2 Z^{(7)}(x) + \tfrac{1}{1296}{\gamma_1}^3 Z^{(9)}(x)]\\ & +[\tfrac{1}{720}\gamma_4 Z^{(6)}(x) + (\tfrac{1}{1152}{\gamma_2}^2 + \tfrac{1}{720}\gamma_1\gamma_3) Z^{(8)}(x)\\ & +\tfrac{1}{1728}{\gamma_1}^2\gamma_2 Z^{(10)}(x) + \tfrac{1}{31104}{\gamma_1}^4 Z^{(12)}(x)] + \ldots, \qquad (9)\end{aligned}$$

where

$$Z(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad (10)$$

is the NORMAL DISTRIBUTION, and

$$\gamma_r = \frac{\kappa_r}{\sigma^{r+2}} \qquad (11)$$

for $r \geq 1$ (with κ_r CUMULANTS and σ the STANDARD DEVIATION; Abramowitz and Stegun 1972, p. 935).

see also CONTINUOUS DISTRIBUTION, CORNISH-FISHER ASYMPTOTIC EXPANSION, DISCRETE DISTRIBUTION, DISTRIBUTION FUNCTION, JOINT DISTRIBUTION FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Probability Functions." Ch. 26 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 925–964, 1972.

Probability Inequality

If $B \supset A$ (B is a superset of A), then $P(A) \leq P(B)$.

Probability Integral

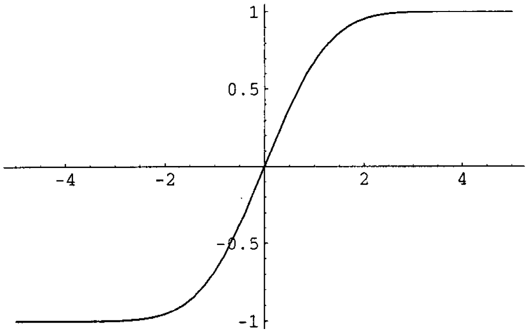

Re[ProbabilityIntegral z]

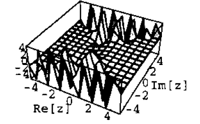

Im[ProbabilityIntegral z]

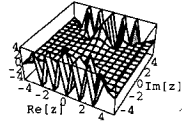

|ProbabilityIntegral z|

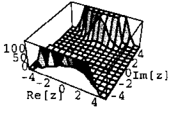

$$\alpha(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{-x}^{x} e^{-t^2/2}\, dt \tag{1}$$

$$= \sqrt{\frac{2}{\pi}} \int_{0}^{x} e^{-t^2/2}\, dt \tag{2}$$

$$= 2\Phi(x) \tag{3}$$

$$= \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right), \tag{4}$$

where $\Phi(x)$ is the NORMAL DISTRIBUTION FUNCTION and ERF is the error function.

see also ERF, NORMAL DISTRIBUTION FUNCTION

Probability Measure

Consider a PROBABILITY SPACE $(S, \mathbb{S}, P)$ where $(S, \mathbb{S})$ is a MEASURABLE SPACE and P is a MEASURE on $\mathbb{S}$ with $P(S) = 1$. Then the MEASURE P is said to be a probability measure. Equivalently, P is said to be normalized.

see also MEASURABLE SPACE, MEASURE, PROBABILITY, PROBABILITY SPACE, STATE SPACE

Probability Space

A triple $(S, \mathbb{S}, P)$, where $(S, \mathbb{S})$ is a measurable space and P is a MEASURE on $\mathbb{S}$ with $P(S) = 1$.

see also MEASURABLE SPACE, MEASURE, PROBABILITY, PROBABILITY MEASURE, RANDOM VARIABLE, STATE SPACE

Probable Error

The first QUARTILE of a standard NORMAL DISTRIBUTION occurs when

$$\int_0^t \Phi(z)\, dz = \tfrac{1}{4}.$$

The solution is $t = 0.6745\ldots$. The value of t giving 1/4 is known as the probable error of a NORMALLY DISTRIBUTED variate. However, the number δ corresponding to the 50% CONFIDENCE INTERVAL,

$$P(\delta) \equiv 1 - 2\int_0^{|\delta|} \phi(t)\, dt = \tfrac{1}{2},$$

is sometimes also called the probable error.

see also SIGNIFICANCE

Probable Prime

A number satisfying FERMAT'S LITTLE THEOREM (or some other primality test) for some nontrivial base. A probable prime which is shown to be COMPOSITE is called a PSEUDOPRIME (otherwise, of course, it is a PRIME).

see also PRIME NUMBER, PSEUDOPRIME

Problem

An exercise whose solution is desired.

see also ALHAZEN'S BILLIARD PROBLEM, ALHAZEN'S PROBLEM, ANDRÉ'S PROBLEM, APOLLONIUS' PROBLEM, APOLLONIUS PURSUIT PROBLEM, ARCHIMEDES' CATTLE PROBLEM, ARCHIMEDES' PROBLEM, BALLOT PROBLEM, BASLER PROBLEM, BERTRAND'S PROBLEM, BILLIARD TABLE PROBLEM, BIRTHDAY PROBLEM, BISHOPS PROBLEM, BOLZA PROBLEM, BOOK STACKING PROBLEM, BOUNDARY VALUE PROBLEM, BOVINUM PROBLEMA, BRACHISTOCHRONE PROBLEM, BRAHMAGUPTA'S PROBLEM, BROCARD'S PROBLEM, BUFFON-LAPLACE NEEDLE PROBLEM, BUFFON'S NEEDLE PROBLEM, BURNSIDE PROBLEM, BUSEMANN-PETTY PROBLEM, CANNONBALL PROBLEM, CASTILLON'S PROBLEM, CATALAN'S DIOPHANTINE PROBLEM, CATALAN'S PROBLEM, CATTLE PROBLEM OF ARCHIMEDES, CAUCHY PROBLEM, CHECKER-JUMPING PROBLEM, CLOSED CURVE PROBLEM, COIN PROBLEM, COLLATZ PROBLEM, CONDOM PROBLEM, CONGRUUM PROBLEM, CONSTANT PROBLEM, COUPON COLLECTOR'S PROBLEM, CROSSED LADDERS PROBLEM, CUBE DOVETAILING PROBLEM, DECISION PROBLEM, DEDEKIND'S PROBLEM, DELIAN PROBLEM, DE

Mere's Problem, Diagonals Problem, Dido's Problem, Dilemma, Dinitz Problem, Dirichlet Divisor Problem, Disk Covering Problem, Equichordal Problem, Extension Problem, Fagnano's Problem, Fejés Tóth's Problem, Fermat's Problem, Fermat's Sigma Problem, Fisher-Behrens Problem, Five Disks Problem, Four Coins Problem, Four Travelers Problem, Fuss's Problem, Gauss's Circle Problem, Gauss's Class Number Problem, Glove Problem, Guthrie's Problem, Haberdasher's Problem, Hadwiger Problem, Halting Problem, Hansen's Problem, Heesch's Problem, Heilbronn Triangle Problem, Hilbert's Problems, Illumination Problem, Indeterminate Problems, Initial Value Problem, Internal Bisectors Problem, Isoperimetric Problem, Isovolume Problem, Jeep Problem, Josephus Problem, Kakeya Needle Problem, Kakutani's Problem, Katona's Problem, Kepler Problem, Kings Problem, Kirkman's Schoolgirl Problem, Kissing Circles Problem, Knapsack Problem, Knot Problem, Königsberg Bridge Problem, Kuratowski's Closure-Component Problem, Lam's Problem, Langford's Problem, Lebesgue Measurability Problem, Lebesgue Minimal Problem, Lehmer's Problem, Lemoine's Problem, Lifting Problem, Lucas' Married Couples Problem, Malfatti's Right Triangle Problem, Malfatti's Tangent Triangle Problem, Married Couples Problem, Match Problem, Maximum Clique Problem, Ménage Problem, Metric Equivalence Problem, Mice Problem, Mikusiński's Problem, Möbius Problem, Money-Changing Problem, Monkey and Coconut Problem, Monty Hall Problem, Mortality Problem, Moser's Circle Problem, Napoleon's Problem, Navigation Problem, Nearest Neighbor Problem, NP-Complete Problem, NP-Problem, Orchard-Planting Problem, Orchard Visibility Problem, P-Problem, Party Problem, Piano Mover's Problem, Planar Bubble Problem, Plateau's Problem, Points Problem, Postage Stamp Problem, Pothenot Problem, Prouhet's Problem, Queens Problem, Railroad Track Problem, Riemann's Moduli Problem, Satisfiability Problem, Schoolgirl Problem, Schur's Problem, Schwarz's Triangle Problem, Sharing Problem, Shephard's Problem, Sinclair's Soap Film Problem, Small World Problem, Snellius-Pothenot Problem, Steenrod's Realization Problem, Steiner's Problem, Steiner's Segment Problem, Surveying Problems, Sylvester's Four-Point Problem, Sylvester's Line Problem, Sylvester's Triangle Problem, Syracuse Problem, Syzygies Problem, Tarry-Escott Problem, Tautochrone Problem, Thomson Problem, Three Jug Problem, Traveling Salesman Problem, Trawler Problem, Ulam's Problem, Utility Problem, Vibration Problem, Wallis's Problem, Waring's Problem

References

Artino, R. A.; Gaglione, A. M.; and Shell, N. *The Contest Problem Book IV: Annual High School Mathematics Examinations 1973–1982.* Washington, DC: Math. Assoc. Amer., 1982.

Alexanderson, G. L.; Klosinski, L.; and Larson, L. *The William Lowell Putnam Mathematical Competition, Problems and Solutions: 1965–1984.* Washington, DC: Math. Assoc. Amer., 1986.

Barbeau, E. J.; Moser, W. O.; and Lamkin, M. S. *Five Hundred Mathematical Challenges.* Washington, DC: Math. Assoc. Amer., 1995.

Brown, K. S. "Most Wanted List of Elementary Unsolved Problems." `http://www.seanet.com/~ksbrown/mwlist.htm`.

Chung, F. and Graham, R. *Erdős on Graphs: His Legacy of Unsolved Problems.* New York: A. K. Peters, 1998.

Cover, T. M. and Gopinath, B. (Eds.). *Open Problems in Communication and Computation.* New York: Springer-Verlag, 1987.

Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, 1965.

Dudeney, H. E. *Amusements in Mathematics.* New York: Dover, 1917.

Dudeney, H. E. *The Canterbury Puzzles and Other Curious Problems, 7th ed.* London: Thomas Nelson and Sons, 1949.

Dudeney, H. E. *536 Puzzles & Curious Problems.* New York: Scribner, 1967.

Eppstein, D. "Open Problems." `http://www.ics.uci.edu/~eppstein/junkyard/open.html`.

Erdős, P. "Some Combinatorial Problems in Geometry." In *Geometry and Differential Geometry* (Ed. R. Artzy and I. Vaisman). New York: Springer-Verlag, pp. 46–53, 1980.

Fenchel, W. (Ed.). "Problems." In *Proc. Colloquium on Convexity, 1965.* Københavns Univ. Mat. Inst., pp. 308–325, 1967.

Finch, S. "Unsolved Mathematical Problems." `http://www.mathsoft.com/asolve/`.

Gleason, A. M.; Greenwood, R. E.; and Kelly, L. M. *The William Lowell Putnam Mathematical Competition, Problems and Solutions: 1938–1964.* Washington, DC: Math. Assoc. Amer., 1980.

Graham, L. A. *Ingenious Mathematical Problems and Methods.* New York: Dover, 1959.

Graham, L. A. *The Surprise Attack in Mathematical Problems.* New York: Dover, 1968.

Greitzer, S. L. *International Mathematical Olympiads, 1959–1977.* Providence, RI: Amer. Math. Soc., 1978.

Gruber, P. M. and Schneider, R. "Problems in Geometric Convexity." In *Contributions to Geometry* (Ed. J. Tölke and J. M. Wills.) Boston, MA: Birkhäuser, pp. 255–278, 1979.

Guy, R. K. (Ed.). "Problems." In *The Geometry of Metric and Linear Spaces.* New York: Springer-Verlag, pp. 233–244, 1974.

Halmos, P. R. *Problems for Mathematicians Young and Old.* Washington, DC: Math. Assoc. Amer., 1991.

Honsberger, R. *Mathematical Gems I.* Washington, DC: Math. Assoc. Amer., 1973.

Honsberger, R. *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., 1976.

Honsberger, R. *Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., 1979.

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., 1985.

Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., 1991.

Honsberger, R. *From Erdős to Kiev.* Washington, DC: Math. Assoc. Amer., 1995.
Honsberger, R. *In Pólya's Footsteps: Miscellaneous Problems and Essays.* Washington, DC: Math. Assoc. Amer., 1997.
Honsberger, R. (Ed.). *Mathematical Plums.* Washington, DC: Math. Assoc. Amer., 1979.
Kimberling, C. "Unsolved Problems and Rewards." `http://www.evansville.edu/~ck6/integer/unsolved.html`.
Klamkin, M. S. *International Mathematical Olympiads, 1978–1985 and Forty Supplementary Problems.* Washington, DC: Math. Assoc. Amer., 1986.
Klamkin, M. S. *U.S.A. Mathematical Olympiads, 1972–1986.* Washington, DC: Math. Assoc. Amer., 1988.
Kordemsky, B. A. *The Moscow Puzzles: 359 Mathematical Recreations.* New York: Dover, 1992.
Kurschak, J. and Hajos, G. *Hungarian Problem Book, Based on the Eőtvős Competitions, Vol. 1: 1894–1905.* New York: Random House, 1963.
Kurschak, J. and Hajos, G. *Hungarian Problem Book, Based on the Eőtvős Competitions, Vol. 2: 1906–1928.* New York: Random House, 1963.
Larson, L. C. *Problem-Solving Through Problems.* New York: Springer-Verlag, 1983.
Mott-Smith, G. *Mathematical Puzzles for Beginners and Enthusiasts.* New York: Dover, 1954.
Ogilvy, C. S. *Tomorrow's Math: Unsolved Problems for the Amateur.* New York: Oxford University Press, 1962.
Ogilvy, C. S. "Some Unsolved Problems of Modern Geometry." Ch. 11 in *Excursions in Geometry.* New York: Dover, pp. 143–153, 1990.
Posamentier, A. S. and Salkind, C. T. *Challenging Problems in Algebra.* New York: Dover, 1997.
Posamentier, A. S. and Salkind, C. T. *Challenging Problems in Geometry.* New York: Dover, 1997.
Rabinowitz, S. (Ed.). *Index to Mathematical Problems 1980–1984.* Westford, MA: MathPro Press, 1992.
Salkind, C. T. *The Contest Problem Book I: Problems from the Annual High School Contests 1950–1960.* New York: Random House, 1961.
Salkind, C. T. *The Contest Problem Book II: Problems from the Annual High School Contests 1961–1965.* Washington, DC: Math. Assoc. Amer., 1966.
Salkind, C. T. and Earl, J. M. *The Contest Problem Book III: Annual High School Contests 1966–1972.* Washington, DC: Math. Assoc. Amer., 1973.
Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, 1993.
Shkliarskii, D. O.; Chentzov, N. N.; and Yaglom, I. M. *The U.S.S.R. Olympiad Problem Book: Selected Problems and Theorems of Elementary Mathematics.* New York: Dover, 1993.
Sierpiński, W. *A Selection of Problems in the Theory of Numbers.* New York: Pergamon Press, 1964. Sierpiński, W. *Problems in Elementary Number Theory.* New York: Elsevier, 1980.
Smarandache, F. *Only Problems, Not Solutions!, 4th ed.* Phoenix, AZ: Xiquan, 1993.
Steinhaus, H. *One Hundred Problems in Elementary Mathematics.* New York: Dover, 1979.
Tietze, H. *Famous Problems of Mathematics.* New York: Graylock Press, 1965.
Trigg, C. W. *Mathematical Quickies: 270 Stimulating Problems with Solutions.* New York: Dover, 1985.
Ulam, S. M. *A Collection of Mathematical Problems.* New York: Interscience Publishers, 1960.
van Mill, J. and Reed, G. M. (Eds.). *Open Problems in Topology.* New York: Elsevier, 1990.

Procedure

A specific prescription for carrying out a task or solving a problem. Also called an ALGORITHM, METHOD, or TECHNIQUE

see also BISECTION PROCEDURE, MAEHLY'S PROCEDURE

Proclus' Axiom

If a line intersects one of two parallel lines, it must intersect the other also. This AXIOM is equivalent to the PARALLEL AXIOM.

References

Dunham, W. "Hippocrates' Quadrature of the Lune." Ch. 1 in *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, p. 54, 1990.

Procrustian Stretch

see HYPERBOLIC ROTATION

Product

The term "product" refers to the result of one or more MULTIPLICATIONS. For example, the mathematical statement $a \times b = c$ would be read "a TIMES b EQUALS c," where c is the product.

The product symbol is defined by

$$\prod_{i=1}^{n} f_i \equiv f_1 \cdot f_2 \cdots f_n.$$

Useful product identities include

$$\ln\left(\prod_{i=1}^{\infty} f_i\right) = \sum_{i=1}^{\infty} \ln f_i$$

$$\prod_{i=1}^{\infty} f_i = \exp\left(\sum_{i=1}^{\infty} \ln f_i\right).$$

For $0 \leq a_i < 1$, then the products $\prod_{i=1}^{\infty}(1+a_i)$ and $\prod_{i=1}^{\infty}(1-a_i)$ converge and diverge as $\prod_{i=1}^{\infty} a_i$.

see also CROSS PRODUCT, DOT PRODUCT, INNER PRODUCT, MATRIX PRODUCT, MULTIPLICATION, NONASSOCIATIVE PRODUCT, OUTER PRODUCT, SUM, TENSOR PRODUCT, TIMES, VECTOR TRIPLE PRODUCT

References

Guy, R. K. "Products Taken over Primes." §B87 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 102–103, 1994.

Product Formula

Let α be a Nonzero Rational Number $\alpha = \pm p_1{}^{\alpha_1} p_2{}^{\alpha_2} \cdots p_L{}^{\alpha_L}$, where $p_1, \ldots, p_L$ are distinct Primes, $\alpha_l \in \mathbb{Z}$ and $\alpha_l \neq 0$. Then

$$|a| \prod_{p \text{ prime}} |\alpha|_p = p_1{}^{\alpha_1} p_2{}^{\alpha_2} \cdots p_L{}^{\alpha_L} \times p_1{}^{-\alpha_1} p_2{}^{-\alpha_2} \cdots p_L{}^{-\alpha_L} = 1.$$

References

Burger, E. B. and Struppeck, T. "Does $\sum_{n=0}^{\infty} \frac{1}{n!}$ Really Converge? Infinite Series and p-adic Analysis." *Amer. Math. Monthly* **103**, 565–577, 1996.

Product-Moment Coefficient of Correlation

see Correlation Coefficient

Product Neighborhood

see Tubular Neighborhood

Product Rule

The Derivative identity

$$\begin{aligned}\frac{d}{dx}[f(x)g(x)] &= \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\ &= \lim_{h\to 0}\left[\frac{f(x+h)g(x+h) - f(x+h)g(x)}{h} + \frac{f(x+h)g(x) - f(x)g(x)}{h}\right] \\ &= \lim_{h\to 0}\left[f(x+h)\frac{g(x+h)-g(x)}{h} + g(x)\frac{f(x+h)-f(x)}{h}\right] = f(x)g'(x) + g(x)f'(x).\end{aligned}$$

see also Chain Rule, Exponent Laws, Quotient Rule

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 11, 1972.

Product Space

A Cartesian product equipped with a "product topology" is called a product space (or product topological space, or direct product).

References

Iyanaga, S. and Kawada, Y. (Eds.). "Product Spaces." §408L *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 1281–1282, 1980.

Program

A precise sequence of instructions designed to accomplish a given task. The implementation of an Algorithm on a computer using a programming language is an example of a program.

see also Algorithm

Projection

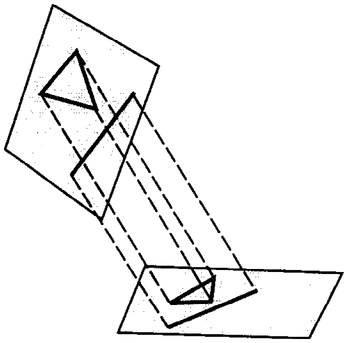

A projection is the transformation of Points and Lines in one Plane onto another Plane by connecting corresponding points on the two planes with Parallel lines. This can be visualized as shining a (point) light source (located at infinity) through a translucent sheet of paper and making an image of whatever is drawn on it on a second sheet of paper. The branch of geometry dealing with the properties and invariants of geometric figures under projection is called Projective Geometry.

The projection of a Vector $\mathbf{a}$ onto a Vector $\mathbf{u}$ is given by

$$\text{proj}_{\mathbf{u}}\mathbf{a} = \frac{\mathbf{a}\cdot\mathbf{u}}{|\mathbf{u}|^2}\mathbf{u},$$

and the length of this projection is

$$|\text{proj}_{\mathbf{u}}\mathbf{a}| = \frac{|\mathbf{a}\cdot\mathbf{u}|}{|\mathbf{u}|}.$$

General projections are considered by Foley and VanDam (1983).

see also Map Projection, Point-Plane Distance, Projective Geometry, Reflection

References

Casey, J. "Theory of Projections." Ch. 11 in *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions, with Numerous Examples, 2nd ed., rev. enl.* Dublin: Hodges, Figgis, & Co., pp. 349–367, 1893.

Foley, J. D. and VanDam, A. *Fundamentals of Interactive Computer Graphics, 2nd ed.* Reading, MA: Addison-Wesley, 1990.

Projection Operator

$$\tilde{p} \equiv |\phi_i(x)\rangle\langle\phi_i(t)|$$

$$\tilde{p}\sum_j c_j|\phi_j(t)\rangle = c_i|\phi_i(x)\rangle$$

$$\sum_i |\phi_i(x)\rangle\langle\phi_i(x)| = 1.$$

see also BRA, KET

Projective Collineation

A COLLINEATION which transforms every 1-D form projectively. Any COLLINEATION which transforms one range into a projectively related range is a projective collineation. Every PERSPECTIVE COLLINEATION is a projective collineation.

see also COLLINEATION, ELATION, HOMOLOGY (GEOMETRY), PERSPECTIVE COLLINEATION

References
Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 247–248, 1969.

Projective General Linear Group

The projective general linear group $PGL_n(q)$ is the GROUP obtained from the GENERAL LINEAR GROUP $GL_n(q)$ on factoring the scalar MATRICES contained in that group.

see also GENERAL LINEAR GROUP, PROJECTIVE GENERAL ORTHOGONAL GROUP, PROJECTIVE GENERAL UNITARY GROUP

References
Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. "The Groups $GL_n(q)$, $SL_n(q)$, $PGL_n(q)$, and $PSL_n(q) = L_n(q)$." §2.1 in *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, p. x, 1985.

Projective General Orthogonal Group

The projective general orthogonal group $PGO_n(q)$ is the GROUP obtained from the GENERAL ORTHOGONAL GROUP $GO_n(q)$ on factoring the scalar MATRICES contained in that group.

see also GENERAL ORTHOGONAL GROUP, PROJECTIVE GENERAL LINEAR GROUP, PROJECTIVE GENERAL UNITARY GROUP

References
Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. "The Groups $GO_n(q)$, $SO_n(q)$, $PGO_n(q)$, and $PSO_n(q)$, and $O_n(q)$." §2.4 in *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, pp. xi–xii, 1985.

Projective General Unitary Group

The projective general unitary group $PGU_n(q)$ is the GROUP obtained from the GENERAL UNITARY GROUP $GU_n(q)$ on factoring the scalar MATRICES contained in that group.

see also GENERAL UNITARY GROUP, PROJECTIVE GENERAL LINEAR GROUP, PROJECTIVE GENERAL ORTHOGONAL GROUP, PROJECTIVE GENERAL UNITARY GROUP

References
Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. "The Groups $GU_n(q)$, $SU_n(q)$, $PGU_n(q)$, and $PSU_n(q) = U_n(q)$." §2.2 in *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, p. x, 1985.

Projective Geometry

The branch of geometry dealing with the properties and invariants of geometric figures under PROJECTION. The most amazing result arising in projective geometry is the DUALITY PRINCIPLE, which states that a duality exists between theorems such as PASCAL'S THEOREM and BRIANCHON'S THEOREM which allows one to be instantly transformed into the other. More generally, *all* the propositions in projective geometry occur in dual pairs, which have the property that, starting from either proposition of a pair, the other can be immediately inferred by interchanging the parts played by the words "POINT" and "LINE."

The AXIOMS of projective geometry are:

1. If A and B are distinct points on a PLANE, there is at least one LINE containing both A and B.
2. If A and B are distinct points on a PLANE, there is not more than one LINE containing both A and B.
3. Any two LINES on a PLANE have at least one point of the PLANE in common.
4. There is at least one LINE on a PLANE.
5. Every LINE contains at least three points of the PLANE.
6. All the points of the PLANE do not belong to the same LINE

(Veblin and Young 1910–18, Kasner and Newman 1989).

see also COLLINEATION, DESARGUES' THEOREM, FUNDAMENTAL THEOREM OF PROJECTIVE GEOMETRY, INVOLUTION (LINE), PENCIL, PERSPECTIVITY, PROJECTIVITY, RANGE (LINE SEGMENT), SECTION (PENCIL)

References
Birkhoff, G. and Mac Lane, S. "Projective Geometry." §9.14 in *A Survey of Modern Algebra, 3rd ed.* New York: Macmillan, pp. 275–279, 1965.
Casey, J. "Theory of Projections." Ch. 11 in *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions, with Numerous Examples, 2nd ed., rev. enl.* Dublin: Hodges, Figgis, & Co., pp. 349–367, 1893.
Coxeter, H. S. M. *Projective Geometry, 2nd ed.* New York: Springer-Verlag, 1987.
Kadison, L. and Kromann, M. T. *Projective Geometry and Modern Algebra.* Boston, MA: Birkhäuser, 1996.
Kasner, E. and Newman, J. R. *Mathematics and the Imagination.* Redmond, WA: Microsoft Press, pp. 150–151, 1989.
Ogilvy, C. S. "Projective Geometry." Ch. 7 in *Excursions in Geometry.* New York: Dover, pp. 86–110, 1990.
Pappas, T. "Art & Projective Geometry." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 66–67, 1989.

Pedoe, D. and Sneddon, I. A. *An Introduction to Projective Geometry.* New York: Pergamon, 1963.
Seidenberg, A. *Lectures in Projective Geometry.* Princeton, NJ: Van Nostrand, 1962.
Struik, D. *Lectures on Projected Geometry.* Reading, MA: Addison-Wesley, 1998.
Veblen, O. and Young, J. W. *Projective Geometry, 2 vols.* Boston, MA: Ginn, 1910–18.
Whitehead, A. N. *The Axioms of Projective Geometry.* New York: Hafner, 1960.

Projective Plane

A projective plane is derived from a usual PLANE by addition of a LINE AT INFINITY. A projective plane of order n is a set of n^2+n+1 POINTS with the properties that:

1. Any two POINTS determine a LINE,
2. Any two LINES determine a POINT,
3. Every POINT has $n+1$ LINES on it, and
4. Every LINE contains $n+1$ POINTS.

(Note that some of these properties are redundant.) A projective plane is therefore a SYMMETRIC $(n^2+n+1, n+1, 1)$ BLOCK DESIGN. An AFFINE PLANE of order n exists IFF a projective plane of order n exists.

A finite projective plane exists when the order n is a POWER of a PRIME, i.e., $n = p^a$ for $a \geq 1$. It is conjectured that these are the *only* possible projective planes, but proving this remains one of the most important unsolved problems in COMBINATORICS. The first few orders which are not of this form are 6, 10, 12, 14, 15, ...

It has been proven analytically that there are no projective planes of order 6. By answering LAM'S PROBLEM in the negative using massive computer calculations on top of some mathematics, it has been proved that there are no finite projective planes of order 10 (Lam 1991). The status of the order 12 projective plane remains open. The remarkable BRUCK-RYSER-CHOWLA THEOREM says that if a projective plane of order n exists, and $n = 1$ or 2 (mod 4), then n is the sum of two SQUARES. This rules out $n = 6$.

The projective plane of order 2, also known as the FANO PLANE, is denoted PG(2, 2). It has INCIDENCE MATRIX

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}.$$

Every row and column contains 3 1s, and any pair of rows/columns has a single 1 in common.

The projective plane has EULER CHARACTERISTIC 1, and the HEAWOOD CONJECTURE therefore shows that any set of regions on it can be colored using six colors only (Saaty 1986).

see also AFFINE PLANE, BRUCK-RYSER-CHOWLA THEOREM, FANO PLANE, LAM'S PROBLEM, MAP COLORING, MOUFANG PLANE, PROJECTIVE PLANE PK^2, REAL PROJECTIVE PLANE

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 281–287, 1987.
Lam, C. W. H. "The Search for a Finite Projective Plane of Order 10." *Amer. Math. Monthly* **98**, 305–318, 1991.
Lindner, C. C. and Rodger, C. A. *Design Theory.* Boca Raton, FL: CRC Press, 1997.
Pinkall, U. "Models of the Real Projective Plane." Ch. 6 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 63–67, 1986.
Saaty, T. L. and Kainen, P. C. *The Four-Color Problem: Assaults and Conquest.* New York: Dover, p. 45, 1986.

Projective Plane PK^2

The 2-D SPACE consisting of the set of TRIPLES

$$\{(a, b, c) : a, b, c \in K,\ \text{not all zero}\},$$

where triples which are SCALAR multiples of each other are identified.

Projective Space

A SPACE which is invariant under the GROUP G of all general LINEAR homogeneous transformation in the SPACE concerned, but not under all the transformations of any GROUP containing G as a SUBGROUP.

A projective space is the space of 1-D VECTOR SUBSPACES of a given VECTOR SPACE. For REAL VECTOR SPACES, the NOTATION $\mathbb{RP}^n$ or $\mathbb{P}^n$ denotes the REAL projective space of dimension n (i.e., the SPACE of 1-D VECTOR SUBSPACES of $\mathbb{R}^{n+1}$) and $\mathbb{CP}^n$ denotes the COMPLEX projective space of COMPLEX dimension n (i.e., the space of 1-D COMPLEX VECTOR SUBSPACES of $\mathbb{C}^{n+1}$). $\mathbb{P}^n$ can also be viewed as the set consisting of $\mathbb{R}^n$ together with its POINTS AT INFINITY.

Projective Special Linear Group

The projective special linear group $PSL_n(q)$ is the GROUP obtained from the SPECIAL LINEAR GROUP $SL_n(q)$ on factoring by the SCALAR MATRICES contained in that GROUP. It is SIMPLE for $n \geq 2$ except for

$$PSL_2(2) = S_3$$
$$PSL_2(3) = A_4,$$

and is therefore also denoted $L_n(Q)$.

see also PROJECTIVE SPECIAL ORTHOGONAL GROUP, PROJECTIVE SPECIAL UNITARY GROUP, SPECIAL LINEAR GROUP

References

Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. "The Groups $GL_n(q)$, $SL_n(q)$, $PGL_n(q)$, and $PSL_n(q) = L_n(q)$." §2.1 in *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, p. x, 1985.

Projective Special Orthogonal Group

The projective special orthogonal group $PSO_n(q)$ is the GROUP obtained from the SPECIAL ORTHOGONAL GROUP $SO_n(q)$ on factoring by the SCALAR MATRICES contained in that GROUP. In general, this GROUP is not SIMPLE.

see also PROJECTIVE SPECIAL LINEAR GROUP, PROJECTIVE SPECIAL UNITARY GROUP, SPECIAL ORTHOGONAL GROUP

References

Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. "The Groups $GO_n(q)$, $SO_n(q)$, $PGO_n(q)$, and $PSO_n(q)$, and $O_n(q)$." §2.4 in *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, pp. xi–xii, 1985.

Projective Special Unitary Group

The projective special unitary group $PSU_n(q)$ is the GROUP obtained from the SPECIAL UNITARY GROUP $SU_n(q)$ on factoring by the SCALAR MATRICES contained in that GROUP. $PSU_n(q)$ is SIMPLE except for

$$\begin{aligned} PSU_2(2) &= S_3 \\ PSU_2(3) &= A_4 \\ PSU_3(2) &= 3^2 : Q_8, \end{aligned}$$

so it is given the simpler name $U_n(q)$, with $U_2(q) = L_2(q)$.

see also PROJECTIVE SPECIAL LINEAR GROUP, PROJECTIVE SPECIAL ORTHOGONAL GROUP, SPECIAL UNITARY GROUP

References

Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. "The Groups $GU_n(q)$, $SU_n(q)$, $PGU_n(q)$, and $PSU_n(q) = U_n(q)$." §2.2 in *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, p. x, 1985.

Projective Symplectic Group

The projective symplectic group $PSp_n(q)$ is the GROUP obtained from the SYMPLECTIC GROUP $Sp_n(q)$ on factoring by the SCALAR MATRICES contained in that GROUP. $PSp_{2m}(q)$ is SIMPLE except for

$$\begin{aligned} PSp_2(2) &= S_3 \\ PSp_2(3) &= A_4 \\ PSp_4(2) &= S_6, \end{aligned}$$

so it is given the simpler name $S_{2m}(q)$, with $S_2(q) = L_2(q)$.

References

Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. "The Groups $Sp_n(q)$ and $PSp_n(q) = S_n(q)$." §2.3 in *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, pp. x–xi, 1985.

Projectivity

The product of any number of PERSPECTIVITIES.

see also INVOLUTION (TRANSFORMATION), PERSPECTIVITY

Prolate Cycloid

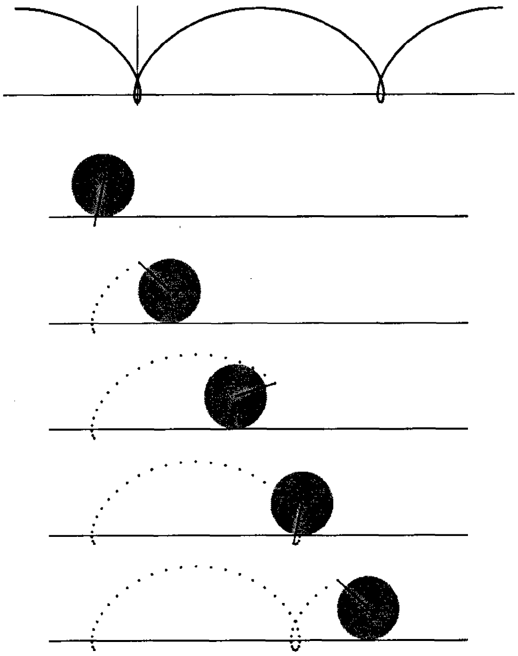

The path traced out by a fixed point at a RADIUS $b > a$, where a is the RADIUS of a rolling CIRCLE, also sometimes called an EXTENDED CYCLOID. The prolate cycloid contains loops, and has parametric equations

$$x = a\phi - b\sin\phi \tag{1}$$

$$y = a - b\cos\phi. \tag{2}$$

The ARC LENGTH from $\phi = 0$ is

$$s = 2(a+b)E(u), \tag{3}$$

where

$$\sin(\tfrac{1}{2}\phi) = \operatorname{sn} u \tag{4}$$

$$k^2 = \frac{4ab}{(a+c)^2}. \tag{5}$$

see also CURTATE CYCLOID, CYCLOID

References

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 46–50, 1991.

Prolate Cycloid Evolute

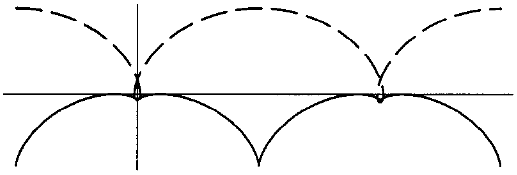

The EVOLUTE of the PROLATE CYCLOID is given by

$$x = \frac{a[-2b\phi + 2a\phi\cos\phi - 2a\sin\phi + b\sin(2\phi)]}{2(a\cos\phi - b)}$$

$$y = \frac{a(a - b\cos\phi)^2}{b(a\cos\phi - b)}.$$

Prolate Spheroid

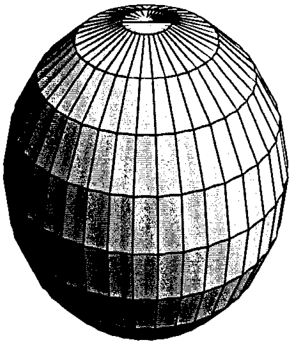

A SPHEROID which is "pointy" instead of "squashed," i.e., one for which the polar radius c is greater than the equatorial radius a, so $c > a$. A prolate spheroid has Cartesian equations

$$\frac{x^2 + y^2}{a^2} + \frac{z^2}{c^2} = 1. \tag{1}$$

The ELLIPTICITY of the prolate spheroid is defined by

$$e \equiv \sqrt{\frac{c^2 - a^2}{c^2}} = \frac{\sqrt{c^2 - a^2}}{c} = \sqrt{1 - \frac{a^2}{c^2}}, \tag{2}$$

so that

$$1 - e^2 = \frac{a^2}{c^2}. \tag{3}$$

Then

$$r = a\left(1 + \frac{e^2}{1 - e^2}\sin^2\delta\right)^{-1/2}. \tag{4}$$

The SURFACE AREA and VOLUME are

$$S = 2\pi a^2 + 2\pi\frac{ac}{e}\sin^{-1} e \tag{5}$$

$$V = \tfrac{4}{3}\pi a^2 c. \tag{6}$$

see also DARWIN-DE SITTER SPHEROID, ELLIPSOID, OBLATE SPHEROID, PROLATE SPHEROIDAL COORDINATES, SPHERE, SPHEROID

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 131, 1987.

Prolate Spheroidal Coordinates

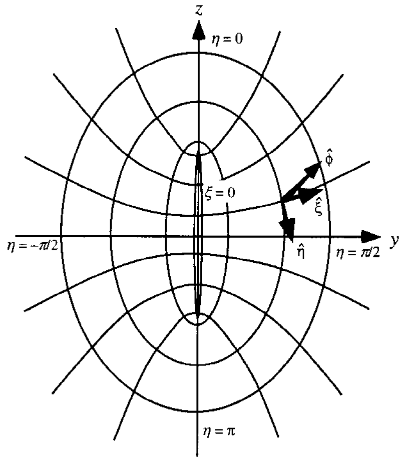

A system of CURVILINEAR COORDINATES in which two sets of coordinate surfaces are obtained by revolving the curves of the ELLIPTIC CYLINDRICAL COORDINATES about the x-AXIS, which is relabeled the z-AXIS. The third set of coordinates consists of planes passing through this axis.

$$x = a\sinh\xi\sin\eta\cos\phi \tag{1}$$

$$y = a\sinh\xi\sin\eta\sin\phi \tag{2}$$

$$z = a\cosh\xi\cos\eta, \tag{3}$$

where $\xi \in [0, \infty)$, $\eta \in [0, \pi)$, and $\phi \in [0, 2\pi)$. Arfken (1970) uses (u, v, φ) instead of (ξ, η, z). The SCALE FACTORS are

$$h_\xi = a\sqrt{\sinh^2\xi + \sin^2\eta} \tag{4}$$

$$h_\eta = a\sqrt{\sinh^2\xi + \sin^2\eta} \tag{5}$$

$$h_\phi = a\sinh\xi\sin\eta. \tag{6}$$

The LAPLACIAN is

$$\nabla^2 f = \frac{1}{a^2(\sinh^2\xi + \sin^2\eta)}\left\{\frac{1}{\sinh\xi}\frac{\partial}{\partial\xi}\left(\sinh\xi\frac{\partial f}{\partial\xi}\right) + \frac{1}{\sin\eta}\frac{\partial}{\partial\eta}\left(\sin\eta\frac{\partial f}{\partial\eta}\right) + \frac{\partial^2 f}{\partial\phi^2}\right\} \tag{7}$$

$$= \frac{1}{a^2(\sin^2\eta + \sinh^2\xi)}\left[(\csc^2\eta + \operatorname{csch}^2\xi)\frac{\partial^2}{\partial\phi^2} + \cot\eta\frac{\partial}{\partial\eta} + \left[\frac{\partial^2}{\partial\eta^2} + \coth\xi\frac{\partial}{\partial\xi} + \frac{\partial^2}{\partial\xi^2}\right]\right. . \tag{8}$$

An alternate form useful for "two-center" problems is defined by

$$\xi_1 = \cosh\xi \tag{9}$$

$$\xi_2 = \cos\eta \tag{10}$$

$$\xi_3 = \phi, \tag{11}$$

where $\xi_1 \in [1,\infty]$, $\xi_2 \in [-1,1]$, and $\xi_3 \in [0,2\pi)$ (Abramowitz and Stegun 1972). In these coordinates,

$$z = a\xi_1\xi_2 \tag{12}$$

$$x = a\sqrt{({\xi_1}^2-1)(1-{\xi_2}^2)}\cos\xi_3 \tag{13}$$

$$y = a\sqrt{({\xi_1}^2-1)(1-{\xi_2}^2)}\sin\xi_3. \tag{14}$$

In terms of the distances from the two Foci,

$$\xi_1 = \frac{r_1+r_2}{2a} \tag{15}$$

$$\xi_2 = \frac{r_1-r_2}{2a} \tag{16}$$

$$2a = r_{12}. \tag{17}$$

The Scale Factors are

$$h_{\xi_1} = a\sqrt{\frac{{\xi_1}^2-{\xi_2}^2}{{\xi_1}^2-1}} \tag{18}$$

$$h_{\xi_2} = a\sqrt{\frac{{\xi_1}^2-{\xi_2}^2}{1-{\xi_2}^2}} \tag{19}$$

$$h_{\xi_3} = a\sqrt{({\xi_1}^2-1)(1-{\xi_2}^2)}, \tag{20}$$

and the Laplacian is

$$\nabla^2 f = \frac{1}{a^2}\left\{\frac{1}{{\xi_1}^2-{\xi_2}^2}\frac{\partial}{\partial\xi_1}\left[({\xi_1}^2-1)\frac{\partial f}{\partial\xi_1}\right] + \frac{1}{{\xi_1}^2-{\xi_2}^2}\frac{\partial}{\partial\xi_2}\left[(1-{\xi_2}^2)\frac{\partial f}{\partial\xi_2}\right] + \frac{1}{({\xi_1}^2-1)(1-{\xi_2}^2)}\frac{\partial^2 f}{\partial{\xi_3}^2}\right\}. \tag{21}$$

The Helmholtz Differential Equation is separable in prolate spheroidal coordinates.

see also Helmholtz Differential Equation—Prolate Spheroidal Coordinates, Latitude, Longitude, Oblate Spheroidal Coordinates, Spherical Coordinates

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Definition of Prolate Spheroidal Coordinates." §21.2 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 752, 1972.

Arfken, G. "Prolate Spheroidal Coordinates (u, v, ϕ)." §2.10 in *Mathematical Methods for Physicists, 2nd ed.* Orlando, FL: Academic Press, pp. 103–107, 1970.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 661, 1953.

Prolate Spheroidal Wave Function

The Wave Equation in Prolate Spheroidal Coordinates is

$$\nabla^2\Phi + k^2\Phi = \frac{\partial}{\partial\xi_1}\left[({\xi_1}^2-1)\frac{\partial\Phi}{\partial\xi_1}\right] + \frac{\partial}{\partial\xi_2}\left[(1-{\xi_2}^2)\frac{\partial\Phi}{\partial\xi_2}\right] + \frac{{\xi_1}^2-{\xi_2}^2}{({\xi_1}^2-1)(1-{x_2}^2)}\frac{\partial^2\Phi}{\partial\phi^2} + c^2({\xi_1}^2-{\xi_2}^2)\Phi = 0, \tag{1}$$

where

$$c \equiv \tfrac{1}{2}ak. \tag{2}$$

Substitute in a trial solution

$$\Phi = R_{mn}(c,\xi_1)S_{mn}(c,\xi_2)\overset{\cos}{\underset{\sin}{}}(m\phi) \tag{3}$$

$$\frac{d}{d\xi_1}\left[({\xi_1}^2-1)\frac{d}{d\xi_1}R_{mn}(c,\xi_1)\right] - \left(\lambda_{mn} - c^2{\xi_1}^2 + \frac{m^2}{{\xi_1}^2-1}\right)R_{mn}(c,\xi_1) = 0. \tag{4}$$

The radial differential equation is

$$\frac{d}{d\xi_2}\left[({\xi_2}^2-1)\frac{d}{d\xi_2}S_{mn}(c,\xi_2)\right] - \left(\lambda_{mn} - c^2{\xi_2}^2 + \frac{m^2}{{\xi_2}^2-1}\right)R_{mn}(c,\xi_2) = 0, \tag{5}$$

and the angular differential equation is

$$\frac{d}{d\xi_2}\left[(1-{\xi_2}^2)\frac{d}{d\xi_2}S_{mn}(c,\xi_2)\right] - \left(\lambda_{mn} - c^2{\xi_2}^2 + \frac{m^2}{1-{\xi_2}^2}\right)R_{mn}(c,\xi_2) = 0. \tag{6}$$

Note that these are identical (except for a sign change). The prolate angular function of the first kind is given by

$$S_{mn}^{(1)} = \begin{cases} \sum_{r=1,3,\ldots}^{\infty} d_r(c)P_{m+r}^m(\eta) & \text{for } n-m \text{ odd} \\ \sum_{r=0,2,\ldots}^{\infty} d_r(c)P_{m+r}^m(\eta) & \text{for } n-m \text{ even,} \end{cases} \tag{7}$$

where $P_m^k(\eta)$ is an associated Legendre Polynomial. The prolate angular function of the second kind is given by

$$S_{mn}^{(2)} = \begin{cases} \sum\limits_{r=\ldots,-1,1,3,\ldots} d_r(c)Q_{m+r}^m(\eta) & \text{for } n-m \text{ odd} \\ \sum\limits_{r=\ldots,-2,0,2,\ldots} d_r(c)Q_{m+r}^m(\eta) & \text{for } n-m \text{ even,} \end{cases} \tag{8}$$

where $Q_k^m(\eta)$ is an associated Legendre Function of the Second Kind and the Coefficients d_r satisfy the Recurrence Relation

$$\alpha_k d_{k+2} + (\beta_k - \lambda_{mn})d_k + \gamma_k d_{k-2} = 0, \tag{9}$$

with

$$\alpha_k = \frac{(2m+k+2)(2m+k+1)c^2}{(2m+2k+3)(2m+2k+5)} \tag{10}$$

$$\beta_k = (m+k)(m+k+1) + \frac{2(m+k)(m+k+1)-2m^2-1}{(2m+2k-1)(2m+2k+3)}c^2 \tag{11}$$

$$\gamma_k = \frac{k(k-1)c^2}{(2m+2k-3)(2m+2k-1)}. \tag{12}$$

Various normalization schemes are used for the ds (Abramowitz and Stegun 1972, p. 758). Meixner and Schäfke (1954) use

$$\int_{-1}^{1} [S_{mn}(c,\eta)]^2\,d\eta = \frac{2}{2n+1}\frac{(n+m)!}{(n-m)!}. \tag{13}$$

Stratton *et al.* (1956) use

$$\frac{(n+m)!}{(n-m)!} = \begin{cases} \sum_{r=1,3,\ldots}^{\infty} \frac{(r+2m)!}{r!} d_r & \text{for } n-m \text{ odd} \\ \sum_{r=0,2,\ldots}^{\infty} \frac{(r+2m)!}{r!} d_r & \text{for } n-m \text{ even.} \end{cases} \tag{14}$$

Flammer (1957) uses

$$S_{mn}(c,0) = \begin{cases} P_n^{m+1}(0) & \text{for } n-m \text{ odd} \\ P_n^m(0) & \text{for } n-m \text{ even.} \end{cases} \tag{15}$$

see also OBLATE SPHEROIDAL WAVE FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Spheroidal Wave Functions." Ch. 21 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 751–759, 1972.

Flammer, C. *Spheroidal Wave Functions.* Stanford, CA: Stanford University Press, 1957.

Meixner, J. and Schäfke, F. W. *Mathieusche Funktionen und Sphäroidfunktionen.* Berlin: Springer-Verlag, 1954.

Stratton, J. A.; Morse, P. M.; Chu, L. J.; Little, J. D. C.; and Corbató, F. J. *Spheroidal Wave Functions.* New York: Wiley, 1956.

Pronic Number

A FIGURATE NUMBER of the form $P_n = 2T_n = n(n+1)$, where T_n is the nth TRIANGULAR NUMBER. The first few are 2, 6, 12, 20, 30, 42, 56, 72, 90, 110, ... (Sloane's A002378). The GENERATING FUNCTION of the pronic numbers is

$$\frac{2x}{(1-x)^3} = 2x + 6x^2 + 12x^3 + 20x^4 + \ldots.$$

The first few n for which P_n are PALINDROMIC are 1, 2, 16, 77, 538, 1621, ... (Sloane's A028336), and the first few PALINDROMIC NUMBERS which are pronic are 2, 6, 272, 6006, 289982, ... (Sloane's A028337).

References

De Geest, P. "Palindromic Products of Two Consecutive Integers." `http://www.ping.be/~ping6758/consec.htm`.

Sloane, N. J. A. Sequences A028336, A028337, and A002378/M1581 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Proof

A rigorous mathematical argument which unequivocally demonstrates the truth of a given PROPOSITION. A mathematical statement which has been proven is called a THEOREM.

There is some debate among mathematicians as to just what constitutes a proof. The FOUR-COLOR THEOREM is an example of this debate, since its "proof" relies on an exhaustive computer testing of many individual cases which cannot be verified "by hand." While many mathematicians regard computer-assisted proofs as valid, some purists do not.

see also PARADOX, PROPOSITION, THEOREM

References

Garnier, R. and Taylor, J. *100% Mathematical Proof.* New York: Wiley, 1996.

Solow, D. *How to Read and Do Proofs: An Introduction to Mathematical Thought Process.* New York: Wiley, 1982.

Proofreading Mistakes

If proofreader A finds a mistakes and proofreader B finds b mistakes, c of which were also found by A, how many mistakes were missed by both A and B? Assume there are a total of m mistakes, so proofreader A finds a FRACTION a/m of all mistakes, and also a FRACTION c/b of the mistakes found by B. Assuming these fractions are the same, then solving for m gives

$$m = \frac{ab}{c}.$$

The number of mistakes missed by both is therefore approximately

$$N = m - a - b + c = \frac{(a-c)(b-c)}{c}.$$

References

Pólya, G. "Probabilities in Proofreading." *Amer. Math. Monthly,* **83**, 42, 1976.

Propeller

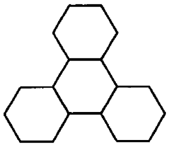

A 4-POLYHEX.

References

Gardner, M. *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, p. 147, 1978.

Proper Cover

see COVER

Proper Divisor

A DIVISOR of a number n excluding n itself.

see also ALIQUANT DIVISOR, ALIQUOT DIVISOR, DIVISOR

Proper Fraction

A FRACTION $p/q < 1$.

see also FRACTION, REDUCED FRACTION

Proper Integral

An INTEGRAL which has neither limit INFINITE and from which the INTEGRAND does not approach INFINITY at any point in the range of integration.

see also IMPROPER INTEGRAL, INTEGRAL

Proper k-Coloring

see k-COLORING

Proper Subset

A SUBSET which is not the entire SET. For example, consider a SET $\{1, 2, 3, 4, 5\}$. Then $\{1, 2, 4\}$ and $\{1\}$ are proper subsets, while $\{1, 2, 6\}$ and $\{1, 2, 3, 4, 5\}$ are not.

see also SET, SUBSET

Proper Superset

A SUPERSET which is not the entire SET.

see also SET, SUPERSET

Proportional

If a is proportional to b, then a/b is a constant. The relationship is written $a \propto b$, which implies

$$a = cb,$$

for some constant c.

Proposition

A statement which is to be proved.

Propositional Calculus

The formal basis of LOGIC dealing with the notion and usage of words such as "NOT," "OR," "AND," and "IMPLIES." Many systems of propositional calculus have been devised which attempt to achieve consistency, completeness, and independence of AXIOMS.

see also LOGIC, P-SYMBOL

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 254–255, 1989.

Nidditch, P. H. *Propositional Calculus.* New York: Free Press of Glencoe, 1962.

Prosthaphaeresis Formulas

TRIGONOMETRY formulas which convert a product of functions into a sum or difference.

Proth's Theorem

For $N = h \cdot 2^n + 1$ with ODD h and $2^n > h$, if there exists an INTEGER a such that

$$a^{(N-1)/2} \equiv -1 \pmod{N},$$

then N is PRIME.

Protractor

A ruled SEMICIRCLE used for measuring and drawing ANGLES.

Prouhet's Problem

A generalization of the TARRY-ESCOTT PROBLEM to three or more sets of INTEGERS.

see also TARRY-ESCOTT PROBLEM

References

Wright, E. M. "Prouhet's 1851 Solution of the Tarry-Escott Problem of 1910." *Amer. Math. Monthly* **102**, 199–210, 1959.

Prüfer Ring

A metric space $\hat{\mathbb{Z}}$ in which the closure of a congruence class $B(j, m)$ is the corresponding congruence class $\{x \in \hat{\mathbb{Z}} | x \equiv j \pmod{m}\}$.

References

Fried, M. D. and Jarden, M. *Field Arithmetic.* New York: Springer-Verlag, pp. 7–11, 1986.

Postnikov, A. G. *Introduction to Analytic Number Theory.* Providence, RI: Amer. Math. Soc., 1988.

Prussian Hat

A device used in the Cornwell smoothness stabilized modification of the CLEAN ALGORITHM.

see also CLEAN ALGORITHM

Pseudoanalytic Function

A pseudoanalytic function is a function defined using generalized CAUCHY-RIEMANN EQUATIONS. Pseudoanalytic functions come as close as possible to having COMPLEX derivatives and are nonsingular "quasiregular" functions.

see also ANALYTIC FUNCTION, SEMIANALYTIC, SUBANALYTIC

Pseudocrosscap

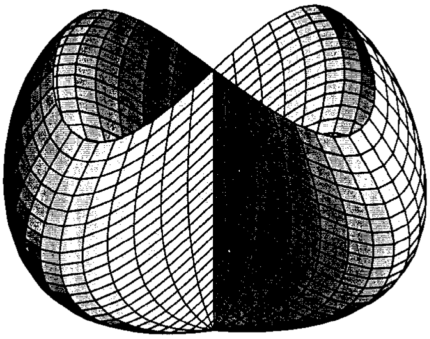

A surface constructed by placing a family of figure-eight curves into $\mathbb{R}^3$ such that the first and last curves reduce to points. The surface has parametric equations

$$\begin{aligned} x(u,v) &= (1-u^2)\sin v \\ y(u,v) &= (1-u^2)\sin(2v) \\ z(u,v) &= u. \end{aligned}$$

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 247–248, 1993.

Pseudocylindrical Projection

A projection in which latitude lines are parallel but meridians are curves.

see also Cylindrical Projection, Eckert IV Projection, Eckert VI Projection, Mollweide Projection, Robinson Projection, Sinusoidal Projection

References

Dana, P. H. "Map Projections." `http://www.utexas.edu/depts/grg/gcraft/notes/mapproj/mapproj.html`.

Pseudogroup

An algebraic structure whose elements consist of selected Homeomorphisms between open subsets of a Space, with the composition of two transformations defined on the largest possible domain. The "germs" of the elements of a pseudogroup form a Groupoid (Weinstein 1996).

see also Group, Groupoid, Inverse Semigroup

References

Weinstein, A. "Groupoids: Unifying Internal and External Symmetry." *Not. Amer. Math. Soc.* **43**, 744–752, 1996.

Pseudolemniscate Case

The case of the Weierstraß Elliptic Function with invariants $g_2 = -1$ and $g_3 = 0$.

see also Equianharmonic Case, Lemniscate Case

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Pseudo-Lemniscate Case ($g_2 = -1$, $g_3 = 0$)." §18.15 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 662–663, 1972.

Pseudoperfect Number

see Semiperfect Number

Pseudoprime

A pseudoprime is a Composite number which passes a test or sequence of tests which fail for most Composite numbers. Unfortunately, some authors drop the "Composite" requirement, calling any number which passes the specified tests a pseudoprime even if it is Prime. Pomerance, Selfridge, and Wagstaff (1980) restrict their use of "pseudoprime" to Odd Composite numbers. "Pseudoprime" used without qualification means Fermat Pseudoprime.

Carmichael Numbers are Odd Composite numbers which are pseudoprimes to every base; they are sometimes called Absolute Pseudoprimes. The following table gives the number of Fermat Pseudoprimes psp, Euler Pseudoprimes epsp, and Strong Pseudoprimes spsp to the base 2, as well as Carmichael Numbers CN which are less the first few powers of 10 (Guy 1994).

	10^3	10^4	10^5	10^6	10^7	10^8	10^9	10^{10}
psp(2)	3	22	78	245	750	2057	5597	14884
epsp(2)	1	12	36	114	375	1071	2939	7706
spsp(2)	0	5	16	46	162	488	1282	3291
CN	1	7	16	43	105	255	646	1547

see also Carmichael Number, Elliptic Pseudoprime, Euler Pseudoprime, Euler-Jacobi Pseudoprime, Extra Strong Lucas Pseudoprime, Fermat Pseudoprime, Fibonacci Pseudoprime, Frobenius Pseudoprime, Lucas Pseudoprime, Perrin Pseudoprime, Probable Prime, Somer-Lucas Pseudoprime, Strong Elliptic Pseudoprime, Strong Frobenius Pseudoprime, Strong Lucas Pseudoprime, Strong Pseudoprime

References

Grantham, J. "Frobenius Pseudoprimes." `http://www.clark.net/pub/grantham/pseudo/pseudo.ps`

Grantham, J. "Pseudoprimes/Probable Primes." `http://www.clark.net/pub/grantham/pseudo`.

Guy, R. K. "Pseudoprimes. Euler Pseudoprimes. Strong Pseudoprimes." §A12 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 27–30, 1994.

Pomerance, C.; Selfridge, J. L.; and Wagstaff, S. S. "The Pseudoprimes to $25 \cdot 10^9$." *Math. Comput.* **35**, 1003–1026, 1980. Available electronically from `ftp://sable.ox.ac.uk/pub/math/primes/ps2.Z`.

Pseudorandom Number

A slightly archaic term for a computer-generated Random Number. The prefix pseudo- is used to distinguish this type of number from a "truly" Random Number generated by a random physical process such as radioactive decay.

see also Random Number

References

Luby, M. *Pseudorandomness and Cryptographic Applications.* Princeton, NJ: Princeton University Press, 1996.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of*

Scientific Computing, 2nd ed. Cambridge, England: Cambridge University Press, p. 266, 1992.

Pseudorhombicuboctahedron

see ELONGATED SQUARE GYROBICUPOLA

Pseudoscalar

A SCALAR which reverses sign under inversion is called a pseudoscalar. The SCALAR TRIPLE PRODUCT

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C})$$

is a pseudoscalar. Given a transformation MATRIX A,

$$S' = \det|\mathsf{A}|S,$$

where det is the DETERMINANT.

see also PSEUDOTENSOR, PSEUDOVECTOR, SCALAR

References

Arfken, G. "Pseudotensors, Dual Tensors." §3.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 128–137, 1985.

Pseudosmarandache Function

The pseudosmarandache function $Z(n)$ is the smallest integer such that

$$\sum_{k=1}^{Z(n)} k = \tfrac{1}{2}Z(n)[Z(n)+1]$$

is divisible by n. The values for $n = 1, 2, \ldots$ are 1, 3, 2, 7, 4, 3, 6, 15, 8, 4, ... (Sloane's A011772).

see also SMARANDACHE FUNCTION

References

Ashbacher, C. "Problem 514." *Pentagon* **57**, 36, 1997.

Kashihara, K. "Comments and Topics on Smarandache Notions and Problems." Vail: Erhus University Press, 1996.

Sloane, N. J. A. Sequence A011772 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Pseudosphere

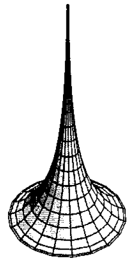

Half the SURFACE OF REVOLUTION generated by a TRACTRIX about its ASYMPTOTE to form a TRACTROID. The Cartesian parametric equations are

$$x = \operatorname{sech} u \cos v \tag{1}$$

$$y = \operatorname{sech} u \sin v \tag{2}$$

$$z = u - \tanh u \tag{3}$$

for $u \geq 0$.

It has constant NEGATIVE CURVATURE, and so is called a pseudosphere by analogy with the SPHERE, which has constant POSITIVE curvature. An equation for the GEODESICS is

$$\cosh^2 u + (v+c)^2 = k^2. \tag{4}$$

see also FUNNEL, GABRIEL'S HORN, TRACTRIX

References

Fischer, G. (Ed.). Plate 82 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 77, 1986.

Geometry Center. "The Pseudosphere." `http://www.geom.umn.edu/zoo/diffgeom/pseudosphere/`.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 383–384, 1993.

Pseudosquare

Given an ODD PRIME p, a SQUARE NUMBER n satisfies $(n/p) = 0$ or 1 for all $p < n$, where (n/p) is the LEGENDRE SYMBOL. A number $n > 2$ which satisfies this relationship but is not a SQUARE NUMBER is called a pseudosquare. The only pseudoprimes less than 10^8 are 3 and 6.

see also SQUARE NUMBER

Pseudotensor

A TENSOR-like object which reverses sign under inversion. Given a transformation MATRIX A,

$$A_{ij}{}' = \det|\mathsf{A}|a_{ik}a_{jl}A_{kl},$$

where det is the DETERMINANT. A pseudotensor is sometimes also called a TENSOR DENSITY.

see also PSEUDOSCALAR, PSEUDOVECTOR, SCALAR, TENSOR DENSITY

References

Arfken, G. "Pseudotensors, Dual Tensors." §3.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 128–137, 1985.

Pseudovector

A typical VECTOR is transformed to its NEGATIVE under inversion. A VECTOR which is invariant under inversion is called a pseudovector, also called an AXIAL VECTOR in older literature (Morse and Feshbach 1953). The CROSS PRODUCT

$$\mathbf{A} \times \mathbf{B} \tag{1}$$

is a pseudovector, whereas the VECTOR TRIPLE PRODUCT

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) \tag{2}$$

is a VECTOR.

$$[\text{pseudovector}] \times [\text{pseudovector}] = [\text{pseudovector}] \tag{3}$$

$$[\text{vector}] \times [\text{pseudovector}] = [\text{vector}]. \tag{4}$$

Given a transformation MATRIX A,

$$C_i' = \det|\mathsf{A}|a_{ij}C_j. \tag{5}$$

see also PSEUDOSCALAR, TENSOR, VECTOR

References
Arfken, G. "Pseudotensors, Dual Tensors." §3.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 128–137, 1985.
Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 46–47, 1953.

Psi Function

$$\Psi(z,s,v) \equiv \sum_{n=0}^{\infty} \frac{z^n}{(v+n)^s}$$

for $|z| < 1$ and $v \neq 0, -1, \ldots$ (Gradshteyn and Ryzhik 1980, pp. 1075–1076).

see also HURWITZ ZETA FUNCTION, RAMANUJAN PSI SUM, THETA FUNCTION

References
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, 1979.

PSLQ Algorithm

An ALGORITHM which finds INTEGER RELATIONS between real numbers $x_1, \ldots, x_n$ such that

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0,$$

with not all $a_i = 0$. This algorithm terminates after a number of iterations bounded by a polynomial in n and uses a numerically stable matrix reduction procedure (Ferguson and Bailey 1992), thus improving upon the FERGUSON-FORCADE ALGORITHM. It is based on a partial sum of squares scheme (like the PSOS ALGORITHM) implemented using LQ decomposition. A much simplified version of the algorithm was developed by Ferguson *et al.* and extended to complex numbers.

see also FERGUSON-FORCADE ALGORITHM, INTEGER RELATION, LLL ALGORITHM, PSOS ALGORITHM

References
Bailey, D. H.; Borwein, J. M.; and Girgensohn, R. "Experimental Evaluation of Euler Sums." *Exper. Math.* **3**, 17–30, 1994.
Bailey, D. and Plouffe, S. "Recognizing Numerical Constants." `http://www.cecm.sfu.ca/organics/papers/bailey`.
Ferguson, H. R. P. and Bailey, D. H. "A Polynomial Time, Numerically Stable Integer Relation Algorithm." RNR Techn. Rept. RNR-91-032, Jul. 14, 1992.
Ferguson, H. R. P.; Bailey, D. H.; and Arno, S. "Analysis of PSLQ, An Integer Relation Finding Algorithm." Unpublished manuscript.

PSOS Algorithm

An INTEGER-RELATION algorithm which is based on a partial sum of squares approach, from which the algorithm takes its name.

see also FERGUSON-FORCADE ALGORITHM, HJLS ALGORITHM, INTEGER RELATION, LLL ALGORITHM, PSLQ ALGORITHM

References
Bailey, D. H. and Ferguson, H. R. P. "Numerical Results on Relations Between Numerical Constants Using a New Algorithm." *Math. Comput.* **53**, 649–656, 1989.

Ptolemy Inequality

For a QUADRILATERAL which is not CYCLIC, PTOLEMY'S THEOREM becomes an INEQUALITY:

$$AB \times CD + BC \times DA > AC \times BD.$$

see also PTOLEMY'S THEOREM, QUADRILATERAL

Ptolemy's Theorem

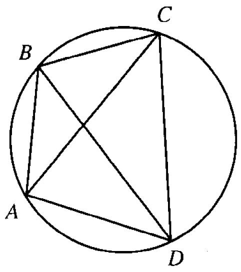

If a QUADRILATERAL is inscribed in a circle (i.e., for a cyclic quadrilateral), the sum of the products of the two pairs of opposite sides equals the product of the diagonals

$$AB \times CD + BC \times DA = AC \times BD.$$

This fact can be used to derive the TRIGONOMETRY addition formulas.

see also FUHRMANN'S THEOREM, PTOLEMY INEQUALITY

References
Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 42–43, 1967.

Public-Key Cryptography

A type of CRYPTOGRAPHY in which the encoding key is revealed without compromising the encoded message. The two best-known methods are the KNAPSACK PROBLEM and RSA ENCRYPTION.

see also KNAPSACK PROBLEM, RSA ENCRYPTION

References
Diffie, W. and Hellman, M. "New Directions in Cryptography." *IEEE Trans. Info. Th.* **22**, 644–654, 1976.
Hellman, M. E. "The Mathematics of Public-Key Cryptography." *Sci. Amer.* **241**, 130–139, Aug. 1979.
Rivest, R.; Shamir, A.; and Adleman, L. "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems." MIT Memo MIT/LCS/TM-82, 1982.
Wagon, S. "Public-Key Encryption." §1.2 in *Mathematica in Action.* New York: W. H. Freeman, pp. 20–22, 1991.

Puiseaux's Theorem

The whole neighborhood of any point y_i of an algebraic PLANE CURVE may be uniformly represented by a certain finite number of convergent developments in POWER SERIES,

$$x_i = \rho_\nu y_i + a_{\nu i1} t_\nu + a_{\nu i2} {t_\nu}^2 + \ldots.$$

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 207, 1959.

Pullback Map

A pullback is a general CATEGORICAL operation appearing in a number of mathematical contexts, sometimes going under a different name. If $T : V \to W$ is a linear transformation between VECTOR SPACES, then $T^* : W^* \to V^*$ (usually called TRANSPOSE MAP or DUAL MAP because its associated matrix is the MATRIX TRANSPOSE of T) is an example of a pullback map.

In the case of a DIFFEOMORPHISM and DIFFERENTIABLE MANIFOLD, a very explicit definition can be formulated. Given an r-form α on a MANIFOLD M_2, define the r-form $T^*(\alpha)$ on M_1 by its action on an r-tuple of tangent vectors $(X_1, \ldots, X_r)$ as the number $T^*(\alpha)(X_1, \ldots, X_r) = \alpha(T_* X_1, \ldots, T_* X_r)$. This defines a map on r-forms and is the pullback map.

see also CATEGORY

Pulse Function

see RECTANGLE FUNCTION

Purser's Theorem

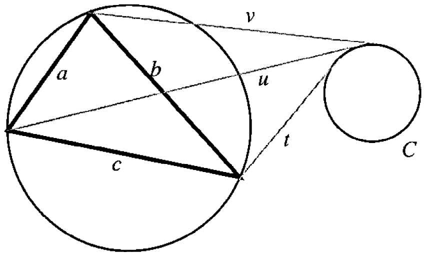

Let t, u, and v be the lengths of the tangents to a CIRCLE C from the vertices of a TRIANGLE with sides of lengths a, b, and c. Then the condition that C is tangent to the CIRCUMCIRCLE of the TRIANGLE is that

$$\pm at \pm bu \pm cv = 0.$$

The theorem was discovered by Casey prior to Purser's independent discovery.

see also CASEY'S THEOREM, CIRCUMCIRCLE

Pursuit Curve

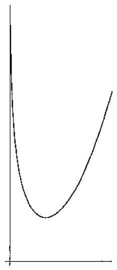

If A moves along a known curve, then P describes a pursuit curve if P is always directed toward A and A and P move with uniform velocities. These were considered in general by the French scientist Pierre Bouguer in 1732. The case restricting A to a straight line was studied by Arthur Bernhart (MacTutor Archive). It has CARTESIAN COORDINATES equation

$$y = cx - \ln x.$$

see also APOLLONIUS PURSUIT PROBLEM, MICE PROBLEM

References

Bernhart, A. "Curves of Pursuit." *Scripta Math.* **20**, 125–141, 1954.

Bernhart, A. "Curves of Pursuit–II." *Scripta Math.* **23**, 49–65, 1957.

Bernhart, A. "Polygons of Pursuit." *Scripta Math.* **24**, 23–50, 1959.

Bernhart, A. "Curves of General Pursuit." *Scripta Math.* **24**, 189–206, 1959.

MacTutor History of Mathematics Archive. "Pursuit Curve." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Pursuit.html`.

Yates, R. C. "Pursuit Curve." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 170–171, 1952.

Push

An action which adds a single element to the top of a STACK, turning the STACK (a_1, a_2, ..., a_n) into (a_0, a_1, a_2, ..., a_n).

see also POKE MOVE, POP, STACK

Puzzle

A mathematical PROBLEM, usually not requiring advanced mathematics, to which a solution is desired. Puzzles frequently require the rearrangement of existing pieces (e.g., 15 PUZZLE) or the filling in of blanks (e.g., crossword puzzle).

see also 15 PUZZLE, BAGUENAUDIER, CALIBAN PUZZLE, CONWAY PUZZLE, CRYPTARITHMETIC, DISSECTION PUZZLES, ICOSIAN GAME, PYTHAGOREAN SQUARE PUZZLE, RUBIK'S CUBE, SLOTHOUBER-GRAATSMA PUZZLE, T-PUZZLE

References

Bogomolny, A. "Interactive Mathematics Miscellany and Puzzles." `http://www.cut-the-knot.com/`.

Dudeney, H. E. *Amusements in Mathematics.* New York: Dover, 1917.
Dudeney, H. E. *The Canterbury Puzzles and Other Curious Problems, 7th ed.* London: Thomas Nelson and Sons, 1949.
Dudeney, H. E. *536 Puzzles & Curious Problems.* New York: Scribner, 1967.
Fujii, J. N. *Puzzles and Graphs.* Washington, DC: National Council of Teachers, 1966.

Pyramid

A POLYHEDRON with one face a POLYGON and all the other faces TRIANGLES with a common VERTEX. An n-gonal regular pyramid (denoted Y_n) has EQUILATERAL TRIANGLES, and is possible only for $n = 3, 4, 5$. These correspond to the TETRAHEDRON, SQUARE PYRAMID, and PENTAGONAL PYRAMID, respectively. A pyramid therefore has a single cross-sectional shape in which the length scale of the CROSS-SECTION scales linearly with height. The AREA at a height z is given by

$$A(z) = A_b \left(\frac{z^2}{h^2}\right), \tag{1}$$

where A_b is the base AREA and h is the pyramid height. The VOLUME is therefore given by

$$V = \int_0^h A(z)\,dz = A_b \int_0^h \frac{z^2}{h^2}\,dz = \frac{A_b}{h^2}(\tfrac{1}{3}h^3) = \tfrac{1}{3}A_b h. \tag{2}$$

These results also hold for the CONE, TETRAHEDRON (triangular pyramid), SQUARE PYRAMID, etc.

The CENTROID is the same as for the CONE, given by

$$\bar{z} = \tfrac{1}{4}h. \tag{3}$$

The SURFACE AREA of a pyramid is

$$S = \tfrac{1}{2}ps, \tag{4}$$

where s is the SLANT HEIGHT and p is the base PERIMETER. Joining two PYRAMIDS together at their bases gives a BIPYRAMID, also called a DIPYRAMID.

see also BIPYRAMID, ELONGATED PYRAMID, GYROELONGATED PYRAMID, PENTAGONAL PYRAMID, PYRAMID, PYRAMIDAL FRUSTUM, SQUARE PYRAMID, TETRAHEDRON, TRUNCATED SQUARE PYRAMID

References
Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 128, 1987.
Hart, G. W. "Pyramids, Dipyramids, and Trapezohedra." `http://www.li.net/~george/virtual-polyhedra/pyramids-info.html`.

Pyramidal Frustum

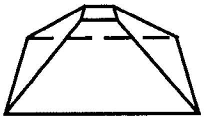

Let s be the slant height, p_i the top and bottom base PERIMETERS, and A_i the top and bottom AREAS. Then the SURFACE AREA and VOLUME of the pyramidal frustum are given by

$$S = \tfrac{1}{2}(p_1 + p_2)s$$
$$V = \tfrac{1}{3}h(A_1 + A_2 + \sqrt{A_1 A_2}).$$

see also CONICAL FRUSTUM, FRUSTUM, PYRAMID, SPHERICAL SEGMENT, TRUNCATED SQUARE PYRAMID

References
Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 128, 1987.
Dunham, W. *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 3–4, 1990.

Pyramidal Number

A FIGURATE NUMBER corresponding to a configuration of points which form a pyramid with r-sided REGULAR POLYGON bases can be thought of as a generalized pyramidal number, and has the form

$$P_n^r = \tfrac{1}{6}(n+1)(2p_n^r + n) = \tfrac{1}{6}n(n+1)[(r-2)n + (5-r)]. \tag{1}$$

The first few cases are therefore

$$P_n^3 = \tfrac{1}{6}n(n+1)(n+2) \tag{2}$$
$$P_n^4 = \tfrac{1}{6}n(n+1)(2n+1) \tag{3}$$
$$P_n^5 = \tfrac{1}{2}n^2(n+1), \tag{4}$$

so $r = 3$ corresponds to a TETRAHEDRAL NUMBER Te_n, and $r = 4$ to a SQUARE PYRAMIDAL NUMBER P_n.

The pyramidal numbers can also be generalized to 4-D and higher dimensions (Sloane and Plouffe 1995).

see also HEPTAGONAL PYRAMIDAL NUMBER, HEXAGONAL PYRAMIDAL NUMBER, PENTAGONAL PYRAMIDAL NUMBER, SQUARE PYRAMIDAL NUMBER, TETRAHEDRAL NUMBER

References
Conway, J. H. and Guy, R. K. "Tetrahedral Numbers" and "Square Pyramidal Numbers" *The Book of Numbers.* New York: Springer-Verlag, pp. 44–49, 1996.
Sloane, N. J. A. and Plouffe, S. "Pyramidal Numbers." Extended entry for sequence M3382 in *The Encyclopedia of Integer Sequences.* San Diego, CA: Academic Press, 1995.

Pyritohedron

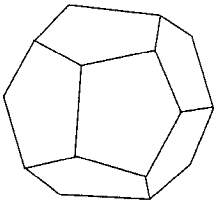

An irregular DODECAHEDRON composed of identical irregular PENTAGONS.

see also DODECAHEDRON, RHOMBIC DODECAHEDRON, TRIGONAL DODECAHEDRON

References
Cotton, F. A. *Chemical Applications of Group Theory, 3rd ed.* New York: Wiley, p. 63, 1990.

Pythagoras's Constant

The number

$$\sqrt{2} = 1.4142135623\ldots,$$

which the Pythagoreans proved to be IRRATIONAL. The Babylonians gave the impressive approximation

$$\sqrt{2} \approx 1 + \frac{24}{60} + \frac{51}{60^2} + \frac{10}{60^3} = 1.41421296296296\ldots$$

(Guy 1990, Conway and Guy 1996, pp. 181–182).

see also IRRATIONAL NUMBER, OCTAGON, PYTHAGORAS'S THEOREM, SQUARE

References
Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 25 and 181–182, 1996.
Guy, R. K. "Review: The Mathematics of Plato's Academy." *Amer. Math. Monthly* **97**, 440–443, 1990.
Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 126, 1993.

Pythagoras's Theorem

Proves that the DIAGONAL d of a SQUARE with sides of integral length s cannot be RATIONAL. Assume d/s is rational and equal to p/q where p and q are INTEGERS with no common factors. Then

$$d^2 = s^2 + s^2 = 2s^2,$$

so

$$\left(\frac{d}{s}\right)^2 = \left(\frac{p}{q}\right)^2 = 2,$$

and $p^2 = 2q^2$, so p^2 is even. But if p^2 is EVEN, then p is EVEN. Since p/q is defined to be expressed in lowest terms, q must be ODD; otherwise p and q would have the common factor 2. Since p is EVEN, we can let $p \equiv 2r$, then $4r^2 = 2q^2$. Therefore, $q^2 = 2r^2$, and q^2, so q must be EVEN. But q cannot be both EVEN and ODD, so there are no d and s such that d/s is RATIONAL, and d/s must be IRRATIONAL.

In particular, PYTHAGORAS'S CONSTANT $\sqrt{2}$ is IRRATIONAL. Conway and Guy (1996) give a proof of this fact using paper folding, as well as similar proofs for ϕ (the GOLDEN RATIO) and $\sqrt{3}$ using a PENTAGON and HEXAGON.

see also IRRATIONAL NUMBER, PYTHAGORAS'S CONSTANT, PYTHAGOREAN THEOREM

References
Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 183–186, 1996.
Pappas, T. "Irrational Numbers & the Pythagoras Theorem." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 98–99, 1989.

Pythagoras Tree

A FRACTAL with symmetric

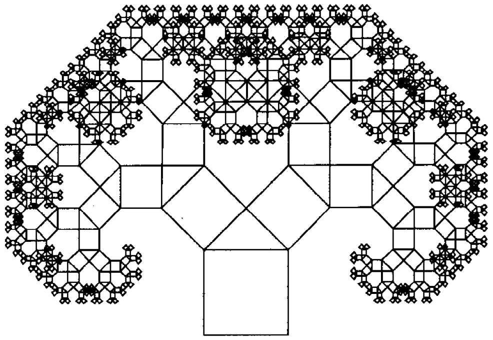

and asymmetric

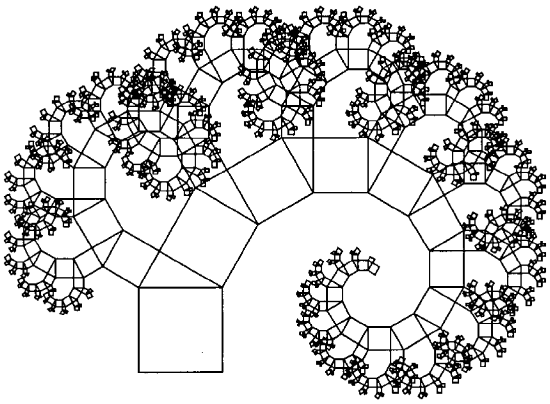

forms.

References
Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 67–77 and 111–113, 1991.
Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m`.

Pythagorean Fraction

Given a PYTHAGOREAN TRIPLE (a,b,c), the fractions a/b and b/a are called Pythagorean fractions. Diophantus showed that the Pythagorean fractions consist precisely of fractions of the form $(p^2 - q^2)/(2pq)$.

References
Conway, J. H. and Guy, R. K. "Pythagorean Fractions." In *The Book of Numbers.* New York: Springer-Verlag, pp. 171–173, 1996.

Pythagorean Quadruple

POSITIVE INTEGERS a, b, c, and d which satisfy

$$a^2 + b^2 + c^2 = d^2. \qquad (1)$$

For POSITIVE EVEN a and b, there exist such INTEGERS c and d; for POSITIVE ODD a and b, no such INTEGERS exist (Oliverio 1996). Oliverio (1996) gives the following generalization of this result. Let $S = (a_1, \ldots, a_{n-2})$, where a_i are INTEGERS, and let T be the number of ODD INTEGERS in S. Then IFF $T \not\equiv 2 \pmod 4$, there exist INTEGERS a_{n-1} and a_n such that

$${a_1}^2 + {a_2}^2 + \ldots + {a_{n-1}}^2 = {a_n}^2. \qquad (2)$$

A set of Pythagorean quadruples is given by

$$\begin{aligned} a &= 2mp && (3)\\ b &= 2np && (4)\\ c &= p^2 - (m^2 + n^2) && (5)\\ d &= p^2 + (m^2 + n^2), && (6) \end{aligned}$$

where m, n, and p are INTEGERS,

$$m + n + p \equiv 1 \pmod 2, \qquad (7)$$

and

$$(m, n, p) = 1 \qquad (8)$$

(Mordell 1969). This does not, however, generate all solutions. For instance, it excludes (36, 8, 3, 37). Another set of solutions can be obtained from

$$\begin{aligned} a &= 2mp + 2nq && (9)\\ b &= 2np - 2mq && (10)\\ c &= p^2 + q^2 - (m^2 + n^2) && (11)\\ d &= p^2 + q^2 + (m^2 + n^2) && (12) \end{aligned}$$

(Carmichael 1915).

see also EULER BRICK, PYTHAGOREAN TRIPLE

References

Carmichael, R. D. *Diophantine Analysis.* New York: Wiley, 1915.

Mordell, L. J. *Diophantine Equations.* London: Academic Press, 1969.

Oliverio, P. "Self-Generating Pythagorean Quadruples and N-tuples." *Fib. Quart.* **34**, 98–101, 1996.

Pythagorean Square Puzzle

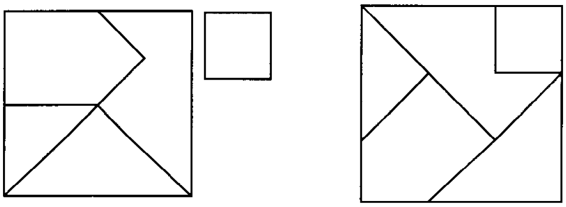

Combine the two above squares on the left into the single large square on the right.

see also DISSECTION, T-PUZZLE

Pythagorean Theorem

For a RIGHT TRIANGLE with legs a and b and HYPOTENUSE c,

$$a^2 + b^2 = c^2. \qquad (1)$$

Many different proofs exist for this most fundamental of all geometric theorems.

A clever proof by DISSECTION which reassembles two small squares into one larger one was given by the Arabian mathematician Thabit Ibn Qurra (Ogilvy 1994, Frederickson 1997).

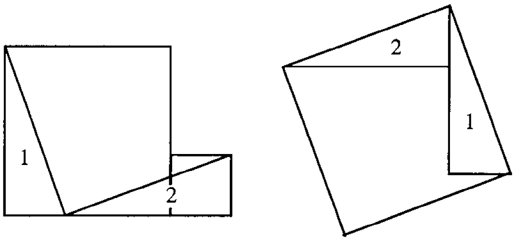

Another proof by DISSECTION is due to Perigal (Pergial 1873, Dudeney 1970, Madachy 1979, Ball and Coxeter 1987).

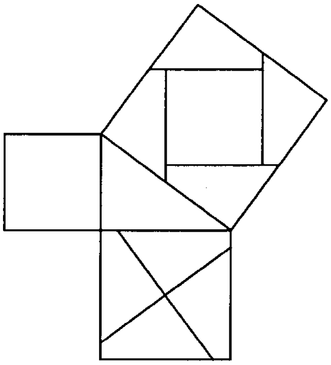

The Indian mathematician Bhaskara constructed a proof using the following figure.

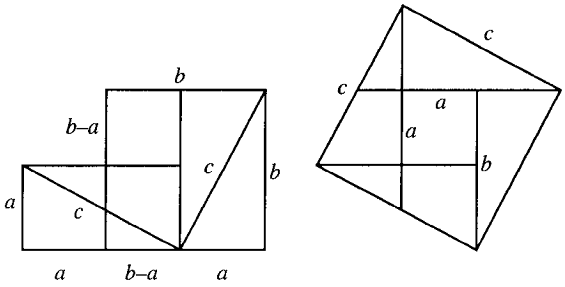

Several similar proofs are shown below.

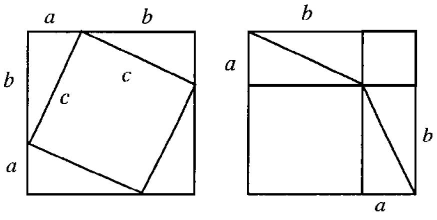

$$c^2 + 4(\tfrac{1}{2}ab) = (a + b)^2 \qquad (2)$$

$$c^2 + 2ab = a^2 + 2ab + b^2 \qquad (3)$$

$$c^2 = a^2 + b^2. \qquad (4)$$

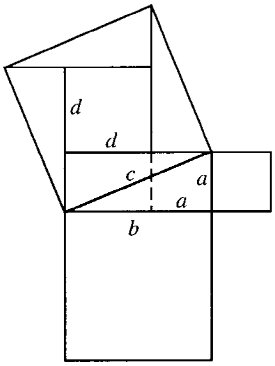

In the above figure, the AREA of the large SQUARE is four times the AREA of one of the TRIANGLES plus the AREA of the interior SQUARE. From the figure, $d = b - a$, so

$$\begin{aligned} A &= 4(\tfrac{1}{2}ab) + d^2 = 2ab + (b-a)^2 = 2ab + b^2 - 2ab + a^2 \\ &= a^2 + b^2 = c^2. \qquad (5) \end{aligned}$$

Perhaps the most famous proof of all times is Euclid's geometric proof. Euclid's proof used the figure below, which is sometimes known variously as the BRIDE'S CHAIR, PEACOCK'S TAIL, or WINDMILL.

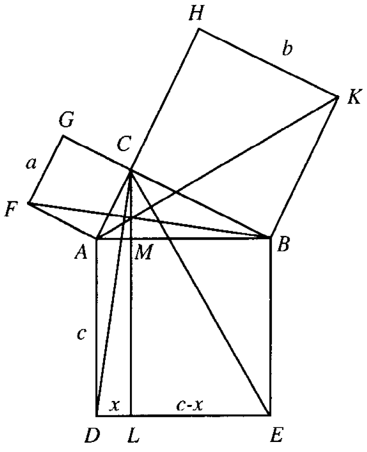

Let ΔABC be a RIGHT TRIANGLE, $\square CAFG$, $\square CBKH$, and $\square ABED$ be squares, and $CL \| BD$. The TRIANGLES ΔFAB and ΔCAD are equivalent except for rotation, so

$$2\Delta FAB = 2\Delta CAD. \qquad (6)$$

Shearing these TRIANGLES gives two more equivalent TRIANGLES

$$2\Delta CAD = \text{▭}ADLM. \qquad (7)$$

Therefore,

$$\square ACGF = \text{▭}ADLM. \qquad (8)$$

Similarly,

$$\square BC = 2\Delta ABK = 2\Delta BCE = \text{▭}BL \qquad (9)$$

so

$$a^2 + b^2 = cx + c(c - x) = c^2. \qquad (10)$$

Heron proved that AK, CL, and BF intersect in a point (Dunham 1990, pp. 48–53).

HERON'S FORMULA for the AREA of the TRIANGLE, contains the Pythagorean theorem implicitly. Using the form

$$K = \tfrac{1}{4}\sqrt{2a^2b^2 + 2a^2c^2 + ab^2c^2 - (a^4 + b^4 + c^4)} \qquad (11)$$

and equating to the AREA

$$K = \tfrac{1}{2}ab \qquad (12)$$

gives

$$\tfrac{1}{4}a^2b^2 = 2a^2b^2 + 2a^2c^2 + ab^2c^2 - (a^4 + b^4 + c^4). \qquad (13)$$

Rearranging and simplifying gives

$$a^2 + b^2 = c^2, \qquad (14)$$

the Pythagorean theorem, where K is the AREA of a TRIANGLE with sides a, b, and c (Dunham 1990, pp. 128–129).

A novel proof using a TRAPEZOID was discovered by James Garfield (1876), later president of the United States, while serving in the House of Representatives (Pappas 1989, pp. 200–201; Bogomolny).

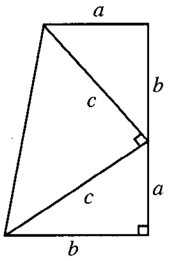

$$\begin{aligned} A_{\text{trapezoid}} &= \tfrac{1}{2}\sum[\text{bases}] \cdot [\text{altitude}] \\ &= \tfrac{1}{2}(a+b)(a+b) \\ &= \tfrac{1}{2}ab + \tfrac{1}{2}ab + \tfrac{1}{2}c^2. \qquad (15) \end{aligned}$$

Rearranging,

$$\tfrac{1}{2}(a^2 + 2ab + b^2) = ab + \tfrac{1}{2}c^2 \qquad (16)$$

$$a^2 + 2ab + b^2 = 2ab + c^2 \qquad (17)$$

$$a^2 + b^2 = c^2. \qquad (18)$$

An algebraic proof (which would not have been accepted by the Greeks) uses the EULER FORMULA. Let the sides of a TRIANGLE be a, b, and c, and the PERPENDICULAR legs of RIGHT TRIANGLE be aligned along the real and imaginary axes. Then

$$a + bi = ce^{i\theta}. \tag{19}$$

Taking the COMPLEX CONJUGATE gives

$$a - bi = ce^{-i\theta}. \tag{20}$$

Multiplying (19) by (20) gives

$$a^2 + b^2 = c^2. \tag{21}$$

Another algebraic proof proceeds by similarity.

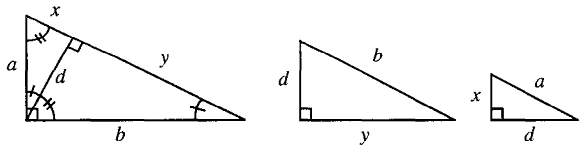

It is a property of RIGHT TRIANGLES, such as the one shown in the above left figure, that the RIGHT TRIANGLE with sides x, a, and d (small triangle in the left figure; reproduced in the right figure) is similar to the RIGHT TRIANGLE with sides d, b, and y (large triangle in the left figure; reproduced in the middle figure), giving

$$\frac{x}{a} = \frac{a}{c} \qquad \frac{y}{b} = \frac{b}{c} \tag{22}$$

$$x = \frac{a^2}{c} \qquad y = \frac{b^2}{c}, \tag{23}$$

so

$$c \equiv x + y = \frac{a^2}{c} + \frac{b^2}{c} = \frac{a^2 + b^2}{c} \tag{24}$$

$$c^2 = a^2 + b^2. \tag{25}$$

Because this proof depends on proportions of potentially IRRATIONAL NUMBERS and cannot be translated directly into a GEOMETRIC CONSTRUCTION, it was not considered valid by Euclid.

see also BRIDE'S CHAIR, COSINES LAW, PEACOCK'S TAIL, PYTHAGORAS'S THEOREM, WINDMILL

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 87–88, 1987.

Bogomolny, A. "Pythagorean Theorem." `http://www.cut-the-knot.com/pythagoras/index.html`.

Dixon, R. "The Theorem of Pythagoras." §4.1 in *Mathographics.* New York: Dover, pp. 92–95, 1991.

Dudeney, H. E. *Amusements in Mathematics.* New York: Dover, p. 32, 1958.

Dunham, W. "Euclid's Proof of the Pythagorean Theorem." Ch. 2 in *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, 1990.

Frederickson, G. *Dissections: Plane and Fancy.* New York: Cambridge University Press, pp. 28–29, 1997.

Garfield, J. A. "Pons Asinorum." *New England J. Educ.* **3**, 161, 1876.

Loomis, E. S. *The Pythagorean Proposition: Its Demonstration Analyzed and Classified and Bibliography of Sources for Data of the Four Kinds of "Proofs."* Reston, VA: National Council of Teachers of Mathematics, 1968.

Machover, M. "Euler's Theorem Implies the Pythagorean Proposition." *Amer. Math. Monthly* **103**, 351, 1996.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, p. 17, 1979.

Ogilvy, C. S. *Excursions in Mathematics.* New York: Dover, p. 52, 1994.

Pappas, T. "The Pythagorean Theorem," "A Twist to the Pythagorean Theorem," and "The Pythagorean Theorem and President Garfield." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 4, 30, and 200–201, 1989.

Perigal, H. "On Geometric Dissections and Transformations." *Messenger Math.* **2**, 103–106, 1873.

Project Mathematics! *The Theorem of Pythagoras.* Videotape (22 minutes). California Institute of Technology. Available from the Math. Assoc. Amer.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 123–127, 1993.

Yancey, B. F. and Calderhead, J. A. "New and Old Proofs of the Pythagorean Theorem." *Amer. Math. Monthly* **3**, 65–67, 110–113, 169–171, and 299–300, 1896.

Yancey, B. F. and Calderhead, J. A. "New and Old Proofs of the Pythagorean Theorem." *Amer. Math. Monthly* **4**, 11–12, 79–81, 168–170, 250–251, and 267–269, 1897.

Yancey, B. F. and Calderhead, J. A. "New and Old Proofs of the Pythagorean Theorem." *Amer. Math. Monthly* **5**, 73–74, 1898.

Yancey, B. F. and Calderhead, J. A. "New and Old Proofs of the Pythagorean Theorem." *Amer. Math. Monthly* **6**, 33–34 and 69–71, 1899.

Pythagorean Triad

see PYTHAGOREAN TRIPLE

Pythagorean Triangle

see PYTHAGOREAN TRIPLE, RIGHT TRIANGLE

Pythagorean Triple

A Pythagorean triple is a TRIPLE of POSITIVE INTEGERS a, b, and c such that a RIGHT TRIANGLE exists with legs a, b and HYPOTENUSE c. By the PYTHAGOREAN THEOREM, this is equivalent to finding POSITIVE INTEGERS a, b, and c satisfying

$$a^2 + b^2 = c^2. \tag{1}$$

The smallest and best-known Pythagorean triple is $(a, b, c) = (3, 4, 5)$.

It is usual to consider only "reduced" (or "primitive") solutions in which a and b are RELATIVELY PRIME, since other solutions can be generated trivially from the primitive ones. For primitive solutions, one of a or b must be EVEN, and the other ODD (Shanks 1993, p. 141), with c always ODD. In addition, in every primitive Pythagorean triple, one side is always DIVISIBLE by 3 and one by 5.

Given a primitive triple (a_0, b_0, c_0), three new primitive triples are obtained from

$$(a_1, b_1, c_1) = (a_0, b_0, c_0)\mathsf{U} \qquad (2)$$

$$(a_2, b_2, c_2) = (a_0, b_0, c_0)\mathsf{A} \qquad (3)$$

$$(a_3, b_3, c_3) = (a_0, b_0, c_0)\mathsf{D}, \qquad (4)$$

where

$$\mathsf{U} \equiv \begin{bmatrix} 1 & 2 & 2 \\ -2 & -1 & -2 \\ 2 & 2 & 3 \end{bmatrix} \qquad (5)$$

$$\mathsf{A} \equiv \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 3 \end{bmatrix} \qquad (6)$$

$$\mathsf{D} \equiv \begin{bmatrix} -1 & -2 & -2 \\ 2 & 1 & 2 \\ 2 & 2 & 3 \end{bmatrix}. \qquad (7)$$

Roberts (1977) proves that (a, b, c) is a primitive Pythagorean triple IFF

$$(a, b, c) = (3, 4, 5)\mathsf{M}, \qquad (8)$$

where M is a FINITE PRODUCT of the MATRICES U, A, D. It therefore follows that every primitive Pythagorean triple must be a member of the INFINITE array

$$\begin{array}{ccccccccc}
 & & & & & & (\ 7, & 24, & 25) \\
 & & & (\ 5, & 12, & 13) & (\ 55, & 48, & 73) \\
 & & & & & & (\ 45, & 28, & 53) \\
 & & & & & & (\ 39, & 80, & 89) \\
(3, & 4, & 5) & (21, & 20, & 29) & (119, & 120, & 169). \\
 & & & & & & (\ 77, & 36, & 85) \\
 & & & & & & (\ 33, & 56, & 65) \\
 & & & (15, & 8, & 17) & (\ 65, & 72, & 97) \\
 & & & & & & (\ 35, & 12, & 37)
\end{array} \qquad (9)$$

For any Pythagorean triple, the PRODUCT of the two nonhypotenuse LEGS (i.e., the two smaller numbers) is always DIVISIBLE by 12, and the PRODUCT of all three sides is DIVISIBLE by 60. It is not known if there are two distinct triples having the same PRODUCT. The existence of two such triples corresponds to a NONZERO solution to the DIOPHANTINE EQUATION

$$xy(x^4 - y^4) = zw(z^4 - w^4) \qquad (10)$$

(Guy 1994, p. 188).

Pythagoras and the Babylonians gave a formula for generating (not necessarily primitive) triples:

$$(2m, (m^2 - 1), (m^2 + 1)), \qquad (11)$$

and Plato gave

$$(2m^2, (m^2 - 1)^2, (m^2 + 1)^2). \qquad (12)$$

A general reduced solution (known to the early Greeks) is

$$(v^2 - u^2, 2uv, u^2 + v^2), \qquad (13)$$

where u and v are RELATIVELY PRIME (Shanks 1993, p. 141). Let F_n be a FIBONACCI NUMBER. Then

$$(F_n F_{n+3}, 2F_{n+1}F_{n+2}, {F_{n+1}}^2 + {F_{n+2}}^2) \qquad (14)$$

is also a Pythagorean triple.

For a Pythagorean triple $(a,\ b,\ c)$,

$$P_3(a) + P_3(b) = P_3(c), \qquad (15)$$

where P_3 is the PARTITION FUNCTION P (Garfunkel 1981, Honsberger 1985). Every three-term progression of SQUARES r^2, s^2, t^2 can be associated with a Pythagorean triple (X, Y, Z) by

$$r = X - Y \qquad (16)$$

$$s = Z \qquad (17)$$

$$t = X + Y \qquad (18)$$

(Robertson 1996).

The AREA of a TRIANGLE corresponding to the Pythagorean triple $(u^2 - v^2, 2uv, u^2 + v^2)$ is

$$A = \tfrac{1}{2}(u^2 - v^2)(2uv) = uv(u^2 - v^2). \qquad (19)$$

Fermat proved that a number of this form can never be a SQUARE NUMBER.

To find the number $L_p(s)$ of possible *primitive* TRIANGLES which may have a LEG (other than the HYPOTENUSE) of length s, factor s into the form

$$s = p_1^{\alpha_1} \cdots p_n^{\alpha_n}. \qquad (20)$$

The number of such TRIANGLES is then

$$L_p(s) = \begin{cases} 0 & \text{for } s \equiv 2 \pmod 4 \\ 2^{n-1} & \text{otherwise,} \end{cases} \qquad (21)$$

i.e., 0 for SINGLY EVEN s and 2 to the power one less than the number of distinct prime factors of s otherwise (Beiler 1966, pp. 115–116). The first few numbers for $s = 1, 2, \ldots$, are 0, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 2, 1, 0, 2, ... (Sloane's A024361). To find the number of ways $L(s)$ in which a number s can be the LEG (other than the HYPOTENUSE) of a *primitive or nonprimitive* RIGHT TRIANGLE, write the factorization of s as

$$s = 2^{a_0} {p_1}^{\alpha_1} \cdots {p_n}^{\alpha_n}. \qquad (22)$$

Then

$$L(s) = \begin{cases} \tfrac{1}{2}[(2a_1+1)(2a_2+1)\cdots(2a_n+1)-1] \\ \quad \text{for } a_0 = 0 \\ \tfrac{1}{2}[(2a_0-1)(2a_1+1)(2a_2+1)\cdots(2a_n+1)-1] \\ \quad \text{for } a_0 \geq 2 \end{cases} \tag{23}$$

(Beiler 1966, p. 116). The first few numbers for $s = 1$, 2, ... are 0, 0, 1, 1, 1, 1, 1, 2, 2, 1, 1, 4, 1, ... (Sloane's A046079).

To find the number of ways $H_p(s)$ in which a number s can be the HYPOTENUSE of a *primitive* RIGHT TRIANGLE, write its factorization as

$$s = 2^{a_0}({p_1}^{a_1}\cdots{p_n}^{a_n})({q_1}^{b_1}\cdots{q_r}^{b_r}), \tag{24}$$

where the ps are of the form $4x-1$ and the qs are of the form $4x+1$. The number of possible *primitive* RIGHT TRIANGLES is then

$$H_p(s) = \begin{cases} 2^{r-1} & \text{for } n = 0 \text{ and } a_0 = 0 \\ 0 & \text{otherwise,} \end{cases} . \tag{25}$$

The first few PRIMES of the form $4x+1$ are 5, 13, 17, 29, 37, 41, 53, 61, 73, 89, 97, 101, 109, 113, 137, ... (Sloane's A002144), so the smallest side lengths which are the hypotenuses of 1, 2, 4, 8, 16, ... primitive right triangles are 5, 65, 1105, 32045, 1185665, 48612265, ... (Sloane's A006278). The number of possible *primitive or nonprimitive* RIGHT TRIANGLES having s as a HYPOTENUSE is

$$H(s) = \tfrac{1}{2}[(2b_1+1)(2b_2+1)\cdots(2b_r+1)-1] \tag{26}$$

(Beiler 1966, p. 117). The first few numbers for $s = 1$, 2, ... are 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, ... (Sloane's A046080).

Therefore, the total number of ways in which s may be either a LEG or HYPOTENUSE of a RIGHT TRIANGLE is given by

$$T(s) = L(s) + H(s). \tag{27}$$

The values for $s = 1$, 2, ... are 0, 0, 1, 1, 2, 1, 1, 2, 2, 2, 1, 4, 2, 1, 5, 3, ... (Sloane's A046081). The smallest numbers s which may be the sides of T general RIGHT TRIANGLES for $T = 1$, 2, ... are 3, 5, 16, 12, 15, 125, 24, 40, ... (Sloane's A006593; Beiler 1966, p. 114).

There are 50 Pythagorean triples with HYPOTENUSE less than 100, the first few of which, sorted by increasing c, are $(3,4,5)$, $(6,8,10)$, $(5,12,13)$, $(9,12,15)$, $(8,15,17)$, $(12,16,20)$, $(15,20,25)$, $(7,24,25)$, $(10,24,26)$, $(20,21,29)$, $(18,24,30)$, $(16,30,34)$, $(21,28,35)$, ... (Sloane's A046083, A046084, and A046085). Of these, only 16 are primitive triplets with HYPOTENUSE less than 100: $(3,4,5)$, $(5,12,13)$, $(8,15,17)$, $(7,24,25)$, $(20,21,29)$, $(12,35,37)$, $(9,40,41)$, $(28,45,53)$, $(11,60,61)$, $(33,56,65)$, $(16,63,65)$, $(48,55,73)$, $(36,77,85)$, $(13,84,85)$, $(39,80,89)$, and $(65,72,97)$ (Sloane's A046086, A046087, and A046088). Of these 16 primitive triplets, seven are twin triplets (defined as triplets for which two members are consecutive integers). The first few twin triplets, sorted by increasing c, are $(3,4,5)$, $(5,12,13)$, $(7,24,25)$, $(20,21,29)$, $(9,40,41)$, $(11,60,61)$, $(13,84,85)$, $(15,112,113)$, ...

Let the number of triples with HYPOTENUSE less than N be denoted $\Delta(N)$, and the number of twin triplets with HYPOTENUSE less than N be denoted $\Delta_2(N)$. Then, as the following table suggests and Lehmer (1900) proved, the number of primitive solutions with HYPOTENUSE less than N satisfies

$$\lim_{N\to\infty} \frac{\Delta(N)}{N} = \frac{1}{2\pi} = 0.159155\ldots. \tag{28}$$

N	$\Delta(N)$	$\Delta(N)/N$	$\Delta_2(N)$
100	16	0.1600	7
500	80	0.1600	17
1000	158	0.1580	24
2000	319	0.1595	34
3000	477	0.1590	41
4000	639	0.1598	47
5000	792	0.1584	52
10000	1593	0.1593	74

Considering twin triplets in which the LEGS are consecutive, a closed form is available for the rth such pair. Consider the general reduced solution $(u^2-v^2, 2uv, u^2+v^2)$, then the requirement that the LEGS be consecutive integers is

$$u^2 - v^2 = 2uv \pm 1. \tag{29}$$

Rearranging gives

$$(u-v)^2 - 2v^2 = \pm 1. \tag{30}$$

Defining

$$u = x + y \tag{31}$$

$$v = y \tag{32}$$

then gives the PELL EQUATION

$$x^2 - 2y^2 = 1. \tag{33}$$

Solutions to the PELL EQUATION are given by

$$x = \frac{(1+\sqrt{2})^r + (1-\sqrt{2})^r}{2} \tag{34}$$

$$y = \frac{(1+\sqrt{2})^r - (1-\sqrt{2})^r}{2\sqrt{2}}, \tag{35}$$

so the lengths of the legs X_r and Y_r and the HYPOTENUSE Z_r are

$$\begin{aligned} X_r &= u^2 - v^2 = x^2 + 2xy \\ &= \frac{\sqrt{2}+1)^{2r+1} - (\sqrt{2}-1)^{2r+1}}{4} + \tfrac{1}{2}(-1)^r \quad (36) \\ Y_r &= 2uv = 2xy + 2y^2 \\ &= \frac{\sqrt{2}+1)^{2r+1} - (\sqrt{2}-1)^{2r+1}}{4} - \tfrac{1}{2}(-1)^r \quad (37) \\ Z_r &= u^2 + v^2 = x^2 + 2xy + 2y^2 \\ &= \frac{(\sqrt{2}+1)^{2r+1} + (\sqrt{2}-1)^{2r+1}}{2\sqrt{2}}. \quad (38) \end{aligned}$$

Denoting the length of the shortest LEG by A_r then gives

$$A_r = \frac{(\sqrt{2}+1)^{2r+1} - (\sqrt{2}-1)^{2r+1}}{4} - \frac{1}{2} \quad (39)$$

$$Z_r = \frac{(\sqrt{2}+1)^{2r+1} + (\sqrt{2}-1)^{2r+1}}{2\sqrt{2}} \quad (40)$$

(Beiler 1966, pp. 124–125 and 256–257), which cannot be solved exactly to give r as a function of Z_r. However, the approximate number of leg-leg twin triplets $\Delta_2^L(N) = r$ less than a given value of $Z_r = N$ can be found by noting that the second term in the DENOMINATOR of Z_r is a small number to the power $1 + 2r$ and can therefore be dropped, leaving

$$N = Z_r > \frac{(\sqrt{2}+1)^{1+2r}}{2\sqrt{2}} \quad (41)$$

$$N > (1+2r)\ln(\sqrt{2}+1) - \ln(2\sqrt{2}). \quad (42)$$

Solving for $r = \Delta_2^L(n)$ gives

$$\begin{aligned} \Delta_2^L(N) &< \frac{\ln N + \ln(2\sqrt{2}) - \ln(\sqrt{2}+1)}{2\ln(\sqrt{2}+1)} \\ &< \left\lfloor \frac{\ln N}{2\ln(1+\sqrt{2})} \right\rfloor \quad (43) \\ &\approx 0.567 \ln N. \quad (44) \end{aligned}$$

The first few LEG-LEG triplets are (3, 4, 5), (20, 21, 29), (119, 120, 169), (696, 697, 985), ... (Sloane's A046089, A046090, and A046091).

LEG-HYPOTENUSE twin triples $(a, b, c) = (v^2 - u^2, 2uv, u^2 + v^2)$ occur whenever

$$u^2 + v^2 = 2uv + 1 \quad (45)$$

$$(u - v)^2 = 1, \quad (46)$$

that is to say when $v = u + 1$, in which case the HYPOTENUSE exceeds the EVEN LEG by unity and the twin triplet is given by $(1 + 2u, 2u(1 + u), 1 + 2u(1 + u))$. The number of leg-hypotenuse triplets with hypotenuse less than N is therefore given by

$$\Delta_2^L(N) = \left\lfloor \tfrac{1}{2}(\sqrt{2N-1} - 1) \right\rfloor, \quad (47)$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION. The first few LEG-HYPOTENUSE triples are (3, 4, 5), (5, 12, 13), (7, 24, 25), (9, 40, 41), (11, 60, 61), (13, 84, 85), ... (Sloane's A005408, A046092, and A046093).

The total number of twin triples $\Delta_2(N)$ less than N is therefore approximately given by

$$\begin{aligned} \Delta_2(N) &= \Delta_2^H(N) + \Delta_2^L(N) - 1 \quad (48) \\ &\approx \left\lfloor \tfrac{1}{2}\sqrt{2N-1} + 0.567 \ln N - 1.5 \right\rfloor, \quad (49) \end{aligned}$$

where one has been subtracted to avoid double counting of the leg-leg-hypotenuse double-twin (3,4,5).

There is a general method for obtaining triplets of Pythagorean triangles with equal AREAS. Take the three sets of generators as

$$m_1 = r^2 + rs + s^2 \quad (50)$$

$$n_1 = r^2 - s^2 \quad (51)$$

$$m_2 = r^2 + rs + s^2 \quad (52)$$

$$n_2 = 2rs + s^2 \quad (53)$$

$$m_3 = r^2 + 2rs \quad (54)$$

$$n_3 = r^2 + rs + s^2. \quad (55)$$

Then the RIGHT TRIANGLE generated by each triple $({m_i}^2 - {n_i}^2, 2m_in_i, {m_i}^2 + {n_i}^2)$ has common AREA

$$A = rs(2r + s)(r + 2s)(r + s)(r - s)(r^2 + rs + s^2) \quad (56)$$

(Beiler 1966, pp. 126–127). The only EXTREMUM of this function occurs at $(r, s) = (0, 0)$. Since $A(r, s) = 0$ for $r = s$, the smallest AREA shared by three nonprimitive RIGHT TRIANGLES is given by $(r, s) = (1, 2)$, which results in an area of 840 and corresponds to the triplets (24, 70, 74), (40, 42, 58), and (15, 112, 113) (Beiler 1966, p. 126). The smallest AREA shared by three *primitive* RIGHT TRIANGLES is 13123110, corresponding to the triples (4485, 5852, 7373), (1380, 19019, 19069), and (3059, 8580, 9109) (Beiler 1966, p. 127).

One can also find quartets of RIGHT TRIANGLES with the same AREA. The QUARTET having smallest known area is (111, 6160, 6161), (231, 2960, 2969), (518, 1320, 1418), (280, 2442, 2458), with AREA 341,880 (Beiler 1966, p. 127). Guy (1994) gives additional information.

It is also possible to find sets of three and four Pythagorean triplets having the same PERIMETER (Beiler 1966, pp. 131–132). Lehmer (1900) showed that the number of primitive triples $N(p)$ with PERIMETER less than p is

$$\lim_{p\to\infty} N(p) = \frac{p \ln 2}{\pi^2} = 0.070230\ldots. \qquad (57)$$

In 1643, Fermat challenged Mersenne to find a Pythagorean triplet whose HYPOTENUSE and SUM of the LEGS were SQUARES. Fermat found the smallest such solution:

$$X = 4565486027761 \qquad (58)$$

$$Y = 1061652293520 \qquad (59)$$

$$Z = 4687298610289, \qquad (60)$$

with

$$Z = 2165017^2 \qquad (61)$$

$$X + Y = 2372159^2. \qquad (62)$$

A related problem is to determine if a specified INTEGER N can be the AREA of a RIGHT TRIANGLE with rational sides. 1, 2, 3, and 4 are not the AREAS of any RATIONAL-sided RIGHT TRIANGLES, but 5 is (3/2, 20/3, 41/6), as is 6 (3, 4, 5). The solution to the problem involves the ELLIPTIC CURVE

$$y^2 = x^3 - N^2 x. \qquad (63)$$

A solution (a, b, c) exists if (63) has a RATIONAL solution, in which case

$$x = \tfrac{1}{4}c^2 \qquad (64)$$

$$y = \tfrac{1}{8}(a^2 - b^2)c \qquad (65)$$

(Koblitz 1993). There is no known general method for determining if there is a solution for arbitrary N, but a technique devised by J. Tunnell in 1983 allows certain values to be ruled out (Cipra 1996).

see also HERONIAN TRIANGLE, PYTHAGOREAN QUADRUPLE, RIGHT TRIANGLE

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 57–59, 1987.

Beiler, A H. "The Eternal Triangle." Ch. 14 in *Recreations in the Theory of Numbers: The Queen of Mathematics Entertains.* New York: Dover, 1966.

Cipra, B. "A Proof to Please Pythagoras." *Science* **271**, 1669, 1996.

Courant, R. and Robbins, H. "Pythagorean Numbers and Fermat's Last Theorem." §2.3 in Supplement to Ch. 1 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 40–42, 1996.

Dickson, L. E. "Rational Right Triangles." Ch. 4 in *History of the Theory of Numbers, Vol. 2: Diophantine Analysis.* New York: Chelsea, pp. 165–190, 1952.

Dixon, R. *Mathographics.* New York: Dover, p. 94, 1991.

Garfunkel, J. *Pi Mu Epsilon J.*, p. 31, 1981.

Guy, R. K. "Triangles with Integer Sides, Medians, and Area." §D21 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 188–190, 1994.

Hindin, H. "Stars, Hexes, Triangular Numbers, and Pythagorean Triples." *J. Recr. Math.* **16**, 191–193, 1983–1984.

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., p. 47, 1985.

Koblitz, N. *Introduction to Elliptic Curves and Modular Forms, 2nd ed.* New York: Springer-Verlag, pp. 1–50, 1993.

Kraitchik, M. *Mathematical Recreations.* New York: W. W. Norton, pp. 95–104, 1942.

Kramer, K. and Tunnell, J. "Elliptic Curves and Local Epsilon Factors." *Comp. Math.* **46**, 307–352, 1982.

Lehmer, D. N. "Asymptotic Evaluation of Certain Totient Sums." *Amer. J. Math.* **22**, 294–335, 1900.

Roberts, J. *Elementary Number Theory: A Problem Oriented Approach.* Cambridge, MA: MIT Press, 1977.

Robertson, J. P. "Magic Squares of Squares." *Math. Mag.* **69**, 289–293, 1996.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 121 and 141, 1993.

Sloane, N. J. A. Sequences A006278, A046079, A002144/M3823, and A006593/M2499 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Taussky-Todd, O. "The Many Aspects of the Pythagorean Triangles." *Linear Algebra and Appl.* **43**, 285–295, 1982.

Q

ℚ

The FIELD of RATIONAL NUMBERS.

see also ℂ, ℂ*, 𝕀, ℕ, ℝ, ℤ

q-Analog

A q-analog, also called a q-EXTENSION or q-GENERALIZATION, is a mathematical expression parameterized by a quantity q which generalizes a known expression and reduces to the known expression in the limit $q \to 1$. There are q-analogs of the FACTORIAL, BINOMIAL COEFFICIENT, DERIVATIVE, INTEGRAL, FIBONACCI NUMBERS, and so on. Koornwinder, Suslov, and Bustoz, have even managed some kind of q-Fourier analysis.

The q-analog of a mathematical object is generally called the "q-object", hence q-BINOMIAL COEFFICIENT, q-FACTORIAL, etc. There are generally several q-analogs if there is one, and there is sometimes even a multibasic analog with independent q_1, q_2, ...

see also d-ANALOG, q-BETA FUNCTION, q-BINOMIAL COEFFICIENT, q-BINOMIAL THEOREM, q-COSINE, q-DERIVATIVE, q-FACTORIAL, q-GAMMA FUNCTION, q-SERIES, q-SINE, q-VANDERMONDE SUM

References

Exton, H. *q-Hypergeometric Functions and Applications.* New York: Halstead Press, 1983.

q-Beta Function

A q-ANALOG of the BETA FUNCTION

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{q-1}\,dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

where $\Gamma(z)$ is a GAMMA FUNCTION, is given by

$$B_q(a,b) \equiv \int_0^1 t^{b-1}(qt;q)_{a-1}\,d(a,t) = \frac{\Gamma_q(b)\Gamma_q(a)}{\Gamma_q(a+b)},$$

where $\Gamma_q(a)$ is a q-GAMMA FUNCTION and $(a;q)_n$ is a q-SERIES coefficient (Andrews 1986, pp. 11–12).

see also q-FACTORIAL, q-GAMMA FUNCTION

References

Andrews, G. E. *q-Series: Their Development and Application in Analysis, Number Theory, Combinatorics, Physics, and Computer Algebra.* Providence, RI: Amer. Math. Soc., 1986.

q-Binomial Coefficient

A q-ANALOG for the BINOMIAL COEFFICIENT, also called the GAUSSIAN COEFFICIENT. It is given by

$$\binom{n}{m}_q \equiv \frac{(q)_n}{(q)_m(q)_{n-m}} = \prod_{i=0}^{j-1} \frac{1-q^{k-i}}{1-q^{i+1}}, \qquad (1)$$

where

$$(q)_k \equiv \prod_{m=1}^{\infty} \frac{1-q^m}{1-q^{k+m}}. \qquad (2)$$

For example, the first few q-binomial coefficients are

$$\binom{2}{1}_q = \frac{1-q^2}{1-q} = 1+q \qquad (3)$$

$$\binom{3}{1}_q = \binom{3}{2}_q = \frac{1-q^3}{1-q} = 1+q+q^2 \qquad (4)$$

$$\binom{4}{1}_q = \binom{4}{3}_q = \frac{1-q^4}{1-q} = 1+q+q^2+q^3 \qquad (5)$$

$$\binom{4}{2}_q = \frac{(1-q^3)(1-q^4)}{(1-q)(1-q^2)} = (1+q)(1+q+q^2). \qquad (6)$$

From the definition, it follows that

$$\binom{n}{1}_q = \binom{n}{n-1}_q = \sum_{i=0}^{n-1} q^n. \qquad (7)$$

In the LIMIT $q \to 1$, the q-binomial coefficient collapses to the usual BINOMIAL COEFFICIENT.

see also CAUCHY BINOMIAL THEOREM, GAUSSIAN POLYNOMIAL

q-Binomial Theorem

The q-ANALOG of the BINOMIAL THEOREM

$$(1-z)^n = 1 - nz + \frac{n(n-1)}{1\cdot 2}z^2 - \frac{n(n-1)(n-2)}{1\cdot 2\cdot 3}z^3 + \ldots$$

is given by

$$\begin{aligned}\left(1-\frac{z}{q^n}\right)\left(1-\frac{z}{q^{n-1}}\right)\cdots\left(1-\frac{z}{q}\right) &= 1 - \frac{1-q^n}{1-q}\frac{z}{q^n} + \frac{1-q^n}{1-q}\frac{1-q^{n-1}}{1-q^2}\frac{z^2}{q^{n+(n-1)}} \\ &\quad - \ldots \pm \frac{z^n}{q^{n(n+1)/2}}.\end{aligned}$$

Written as a q-SERIES, the identity becomes

$$\sum_{n=0}^{\infty} \frac{(a;q)_n}{(q;q)_n} z^n = \frac{(az;q)_\infty}{(z;q)_\infty},$$

where

$$(a;q)_n = \prod_{m=0}^{\infty} \frac{(1-aq^m)}{(1-aq^{m+n})}$$

(Heine 1847, p. 303; Andrews 1986). The CAUCHY BINOMIAL THEOREM is a special case of this general theorem.

see also BINOMIAL SERIES, BINOMIAL THEOREM, CAUCHY BINOMIAL THEOREM, HEINE HYPERGEOMETRIC SERIES, RAMANUJAN PSI SUM

References

Andrews, G. E. *q-Series: Their Development and Application in Analysis, Number Theory, Combinatorics, Physics, and Computer Algebra.* Providence, RI: Amer. Math. Soc., p. 10, 1986.

Heine, E. "Untersuchungen über die Reihe $1+\frac{(1-q^\alpha)(1-q^\beta)}{(1-q)(1-q^\gamma)}\cdot x+\frac{(1-q^\alpha)(1-q^{\alpha+1})(1-q^\beta)(1-q^{\beta+1})}{(1-q)(1-q^2)(1-q^\gamma)(1-q^{\gamma+1})}\cdot x^2+\ldots$." *J. reine angew. Math.* **34**, 285–328, 1847.

q-Cosine

The *q*-ANALOG of the COSINE function, as advocated by R. W. Gosper, is defined by

$$\cos_q(z,q) = \frac{\vartheta_2(z,p)}{\vartheta_2(0,p)},$$

where $\vartheta_2(z,p)$ is a THETA FUNCTION and p is defined via

$$(\ln p)(\ln q) = \pi^2.$$

This is a period 2π, EVEN FUNCTION of unit amplitude with double and triple angle formulas and addition formulas which are analogous to ordinary SINE and COSINE. For example,

$$\cos_q(2z,q) = {\cos_q}^2(z,q^2) - {\sin_q}^2(z,q^2),$$

where $\sin_q(z,a)$ is the *q*-SINE, and π_q is *q*-PI. The *q*-cosine also satisfies

$$\cos_q(\pi a) = \frac{\sum_{n=-\infty}^{\infty}(-1)^n q^{(n+a)^2}}{\sum_{n=-\infty}^{\infty}(-1)^n q^{n^2}}.$$

see also *q*-FACTORIAL, *q*-SINE

References

Gosper, R. W. "Experiments and Discoveries in *q*-Trigonometry." Unpublished manuscript.

q-Derivative

The *q*-ANALOG of the DERIVATIVE, defined by

$$\left(\frac{d}{dx}\right)_q f(x) = \frac{f(x)-f(qx)}{x-qx}.$$

For example,

$$\left(\frac{d}{dx}\right)_q \sin x = \frac{\sin x - \sin(qx)}{x-qx}$$

$$\left(\frac{d}{dx}\right)_q \ln x = \frac{\ln x - \ln(qx)}{x-qx} = \frac{\ln\left(\frac{1}{q}\right)}{(1-q)x}$$

$$\left(\frac{d}{dx}\right)_q x^2 = \frac{x^2-q^2x^2}{x-qx} = (1+q)x$$

$$\left(\frac{d}{dx}\right)_q x^3 = \frac{x^3-q^3x^3}{x-qx} = (1+q+q^2)x^2.$$

In the LIMIT $q \to 1$, the *q*-derivative reduces to the usual DERIVATIVE.

see also DERIVATIVE

q-Dimension

$$D_q \equiv \frac{1}{1-q} \lim_{\epsilon\to 0} \frac{\ln I(q,\epsilon)}{\ln\left(\frac{1}{\epsilon}\right)}, \tag{1}$$

where

$$I(q,\epsilon) \equiv \sum_{i=1}^{N} {\mu_i}^q, \tag{2}$$

ϵ is the box size, and μ_i is the NATURAL MEASURE. If $q_1 > q_2$, then

$$D_{q_1} \leq D_{q_2}. \tag{3}$$

The CAPACITY DIMENSION (a.k.a. BOX COUNTING DIMENSION) is given by $q = 0$,

$$D_0 = \frac{1}{1-0}\lim_{\epsilon\to 0}\frac{\ln\left(\sum_{i=1}^{N(\epsilon)} 1\right)}{-\ln\epsilon} = -\lim_{\epsilon\to 0}\frac{\ln[N(\epsilon)]}{\ln\epsilon}. \tag{4}$$

If all μ_is are equal, then the CAPACITY DIMENSION is obtained for any q. The INFORMATION DIMENSION is defined by

$$\begin{aligned} D_1 &= \lim_{q\to 1} D_q = \lim_{q\to 1}\frac{\lim_{\epsilon\to 0}\frac{\ln\left[\sum_{i=1}^{N(\epsilon)}{\mu_i}^q\right]}{-\ln\epsilon}}{1-q} \\ &= \lim_{\epsilon\to 0}\lim_{q\to 1}\frac{\ln\left(\sum_{i=1}^{N(\epsilon)}{\mu_i}^q\right)}{\ln\epsilon(q-1)}. \end{aligned} \tag{5}$$

But

$$\lim_{q\to 1}\ln\left(\sum_{i=1}^{N(\epsilon)}{\mu_i}^q\right) = \ln\left(\sum_{i=1}^{N(\epsilon)}\mu_i\right) = \ln 1 = 0, \tag{6}$$

so use L'HOSPITAL'S RULE

$$D_1 = \lim_{\epsilon\to 0}\left(\frac{1}{\ln\epsilon}\lim_{q\to 1}\frac{\sum q{\mu_i}^{q-1}}{\sum{\mu_i}^q}\right). \tag{7}$$

Therefore,

$$D_1 = \lim_{\epsilon\to 0}\frac{\sum_{i=1}^{N(\epsilon)}\mu_i\ln\mu_i}{\ln\epsilon}. \tag{8}$$

D_2 is called the CORRELATION DIMENSION. The *q*-dimensions satisfy

$$D_{q+1} \leq D_q. \tag{9}$$

see also FRACTAL DIMENSION

Q.E.D.

An abbreviation for the Latin phrase "quod erat demonstrandum" ("that which was to be demonstrated"), a NOTATION which is often placed at the end of a mathematical proof to indicate its completion.

q-Extension

see q-ANALOG

q-Factorial

The q-ANALOG of the FACTORIAL (by analogy with the q-GAMMA FUNCTION). For a an integer, the q-factorial is defined by

$$\operatorname{faq}(a,q)=1(1+q)(1+q+q^2)\cdots(1+q+\ldots+q^{a-1}).$$

A reflection formula analogous to the GAMMA FUNCTION reflection formula is given by

$$\begin{aligned}\cos_q(\pi a)&=\sin_q[\pi(\tfrac{1}{2}-a)]\\&=\frac{\pi_q q^{(a-1/2)(a+1/2)}}{\operatorname{faq}(a-\frac{1}{2},q^2)\operatorname{faq}(-(a+\frac{1}{2}),q^2)},\end{aligned}$$

where $\cos_q(z)$ is the q-COSINE, $\sin_q(z)$ is the q-SINE, and π_q is q-PI.

see also q-BETA FUNCTION, q-COSINE, q-GAMMA FUNCTION, q-PI, q-SINE

References

Gosper, R. W. "Experiments and Discoveries in q-Trigonometry." Unpublished manuscript.

Q-Function

Let

$$q=e^{-\pi K'/K}=e^{-i\pi\tau}, \tag{1}$$

then

$$Q_0\equiv\prod_{n=1}^{\infty}(1-q^{2n}) \tag{2}$$

$$Q_1\equiv\prod_{n=1}^{\infty}(1+q^{2n}) \tag{3}$$

$$Q_2\equiv\prod_{n=1}^{\infty}(1+q^{2n-1}) \tag{4}$$

$$Q_3\equiv\prod_{n=1}^{\infty}(1-q^{2n-1}). \tag{5}$$

The Q-functions are sometimes written using a lowercase q instead of a capital Q. The Q-functions also satisfy the identities

$$Q_0Q_1=Q_0(q^2) \tag{6}$$

$$Q_0Q_3=Q_0(q^{1/2}) \tag{7}$$

$$Q_2Q_3=Q_3(q^2) \tag{8}$$

$$Q_1Q_2=Q_1(q^{1/2}). \tag{9}$$

see also JACOBI IDENTITIES, q-SERIES

References

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, pp. 55 and 63–85, 1987.

Tannery, J. and Molk, J. *Elements de la Théorie des Fonctions Elliptiques, 4 vols.* Paris: Gauthier-Villars et fils, 1893–1902.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, pp. 469–473 and 488–489, 1990.

q-Gamma Function

A q-ANALOG of the GAMMA FUNCTION defined by

$$\Gamma_q(x,q)\equiv\frac{(q;q)_\infty}{(q^x;q)_\infty}(1-q)^{1-x}, \tag{1}$$

where $(x,q)_\infty$ is a q-SERIES. The q-gamma function satisfies

$$\lim_{q\to 1^-}\Gamma_q(x)=\Gamma(x) \tag{2}$$

(Andrews 1986).

A curious identity for the functional equation

$$\begin{aligned}&f(a-b)f(a-c)f(a-d)f(a-e)-f(b)f(c)f(d)f(e)\\&\quad=q^b f(a)f(a-b-c)f(a-b-d)f(a-b-e),\end{aligned} \tag{3}$$

where

$$b+c+d+e=2a \tag{4}$$

is given by

$$f(\alpha)=\begin{cases}\sin(k\alpha) & \text{for } q=1\\ \frac{1}{\Gamma_q(\alpha)\Gamma_q(1-\alpha)} & \text{for } 0<q<1,\end{cases} \tag{5}$$

for any k.

see also q-BETA FUNCTION, q-FACTORIAL

References

Andrews, G. E. "W. Gosper's Proof that $\lim_{q\to 1^-}\Gamma_q(x)=\Gamma(x)$." Appendix A in *q-Series: Their Development and Application in Analysis, Number Theory, Combinatorics, Physics, and Computer Algebra.* Providence, RI: Amer. Math. Soc., p. 11 and 109, 1986.

Wenchang, C. Problem 10226 and Solution. "A q-Trigonometric Identity." *Amer. Math. Monthly* **103**, 175–177, 1996.

q-Generalization

see q-ANALOG

q-Hypergeometric Series

see HEINE HYPERGEOMETRIC SERIES

Q-Matrix

see FIBONACCI Q-MATRIX

Q-Number

see HOFSTADTER'S Q-SEQUENCE

q-Pi

The q-ANALOG of PI π_q can be defined by taking $a = 0$ in the q-FACTORIAL

$$\text{faq}(a,q) = 1(1+q)(1+q+q^2)\cdots(1+q+\ldots+q^{a-1}),$$

giving

$$1 = \sin_q(\tfrac{1}{2}\pi) = \frac{\pi_q}{\text{faq}^2(-\frac{1}{2}, q^2)q^{1/4}},$$

where $\sin_q(z)$ is the q-SINE. Gosper has developed an iterative algorithm for computing π based on the algebraic RECURRENCE RELATION

$$\frac{4\pi_{q^4}}{q^4+1} = \frac{(q^2+1)^2{\pi_q}^2}{\pi_{q^2}} - \frac{(q^4+1){\pi_{q^2}}^2}{\pi_{q^4}}.$$

Q-Polynomial

see BLM/HO POLYNOMIAL

q-Product

see Q-FUNCTION

q-Series

A SERIES involving coefficients of the form

$$(a)_n \equiv (a;q)_n \equiv \prod_{k=0}^{\infty} \frac{(1-aq^k)}{(1-aq^{k+n})} \tag{1}$$

$$= \prod_{k=0}^{n-1}(1-aq^k) \tag{2}$$

(Andrews 1986). The symbols

$$[n] \equiv 1+q+q^2+\ldots+q^{n-1} \tag{3}$$

$$[n]! \equiv [n][n-1]\cdots[1] \tag{4}$$

are sometimes also used when discussing q-series.

There are a great many beautiful identities involving q-series, some of which follow directly by taking the q-ANALOG of standard combinatorial identities, e.g., the q-BINOMIAL THEOREM

$$\sum_{n=0}^{\infty} \frac{(a;q)_n z^n}{(q;q)_n} = \frac{(az;q)_\infty}{(z;q)_\infty} \tag{5}$$

($|z| < 1$, $|q| < 1$; Andrews 1986, p. 10) and q-VANDERMONDE SUM

$$_2\phi_1(a, q^{-n}; c; q, q) = \frac{a^n(c/a,q)_n}{(c;q)_n}, \tag{6}$$

where $_2\phi_1(a,b;c;q,z)$ is a HEINE HYPERGEOMETRIC SERIES. Other q-series identities, e.g., the JACOBI IDENTITIES, ROGERS-RAMANUJAN IDENTITIES, and HEINE HYPERGEOMETRIC IDENTITY

$$_2\phi_1(a,b;c;q,z) = \frac{(b;q)_\infty(az;q)_\infty}{(c;q)_\infty(z;q)_\infty}\,{}_2\phi_1(c/b, a; az; q, b), \tag{7}$$

seem to arise out of the blue.

see also BORWEIN CONJECTURES, FINE'S EQUATION, GAUSSIAN COEFFICIENT, HEINE HYPERGEOMETRIC SERIES, JACKSON'S IDENTITY, JACOBI IDENTITIES, MOCK THETA FUNCTION, q-ANALOG, q-BINOMIAL THEOREM, q-COSINE, q-FACTORIAL, Q-FUNCTION, q-GAMMA FUNCTION, q-SINE, RAMANUJAN PSI SUM, RAMANUJAN THETA FUNCTIONS, ROGERS-RAMANUJAN IDENTITIES

References

Andrews, G. E. *q-Series: Their Development and Application in Analysis, Number Theory, Combinatorics, Physics, and Computer Algebra.* Providence, RI: Amer. Math. Soc., 1986.

Berndt, B. C. "q-Series." Ch. 27 in *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 261–286, 1994.

Gasper, G. and Rahman, M. *Basic Hypergeometric Series.* Cambridge, England: Cambridge University Press, 1990.

Gosper, R. W. "Experiments and Discoveries in q-Trigonometry." Unpublished manuscript.

Q-Signature

see SIGNATURE (RECURRENCE RELATION)

q-Sine

The q-ANALOG of the SINE function, as advocated by R. W. Gosper, is defined by

$$\sin_q(z,q) = \frac{\vartheta_1(z,p)}{\vartheta_1(\frac{1}{2}\pi,p)},$$

where $\vartheta_1(z,p)$ is a THETA FUNCTION and p is defined via

$$(\ln p)(\ln q) = \pi^2.$$

This is a period 2π, ODD FUNCTION of unit amplitude with double and triple angle formulas and addition formulas which are analogous to ordinary SINE and COSINE. For example,

$$\sin_q(2z,q) = (q+1)\frac{\pi_q}{p_{q^2}}\cos_q(z,q^2)\sin_q(z,q^2),$$

where $\cos_q(z,a)$ is the q-COSINE, and π_q is q-PI.

see also q-COSINE, q-FACTORIAL

References

Gosper, R. W. "Experiments and Discoveries in q-Trigonometry." Unpublished manuscript.

q-Vandermonde Sum

$$_2\phi_1(a, q^{-n}; c; q, q) = \frac{a^n (c/a, q)_n}{(c; q)_n},$$

where $_2\phi_1(a, b; c; q, z)$ is a HEINE HYPERGEOMETRIC SERIES.

see also CHU-VANDERMONDE IDENTITY, HEINE HYPERGEOMETRIC SERIES

References

Andrews, G. E. *q-Series: Their Development and Application in Analysis, Number Theory, Combinatorics, Physics, and Computer Algebra.* Providence, RI: Amer. Math. Soc., pp. 15–16, 1986.

QR Decomposition

Given a MATRIX A, its QR-decomposition is of the form

$$\mathsf{A} = \mathsf{QR},$$

where R is an upper TRIANGULAR MATRIX and Q is an ORTHOGONAL MATRIX, i.e., one satisfying

$$\mathsf{Q}^{\mathrm{T}}\mathsf{Q} = \mathsf{I},$$

where I is the IDENTITY MATRIX. This matrix decomposition can be used to solve linear systems of equations.

see also CHOLESKY DECOMPOSITION, LU DECOMPOSITION, SINGULAR VALUE DECOMPOSITION

References

Householder, A. S. *The Numerical Treatment of a Single Non-Linear Equations.* New York: McGraw-Hill, 1970.

Nash, J. C. *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation, 2nd ed.* Bristol, England: Adam Hilger, pp. 26–28, 1990.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "QR Decomposition." §2.10 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 91–95, 1992.

Stewart, G. W. "A Parallel Implementation of the QR Algorithm." *Parallel Comput.* **5**, 187–196, 1987. `ftp://thales.cs.umd.edu/pub/reports/piqra.ps`.

Quadrable

A plane figure for which QUADRATURE is possible is said to be quadrable.

Quadrangle

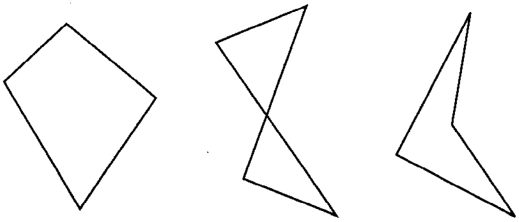

A plane figure consisting of four points, each of which is joined to two other points by a LINE SEGMENT (where the line segments may intersect). A quadrangle may therefore be CONCAVE or CONVEX; if it is CONVEX, it is called a QUADRILATERAL.

see also COMPLETE QUADRANGLE, CYCLIC QUADRANGLE, QUADRILATERAL

Quadrant

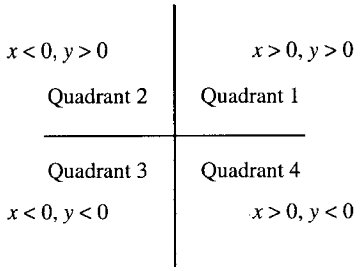

One of the four regions of the PLANE defined by the four possible combinations of SIGNS $(+,+)$, $(+,-)$, $(-,+)$, and $(-,-)$ for (x, y).

see also OCTANT, x-AXIS, y-AXIS

References

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 73, 1996.

Quadratfrei

see SQUAREFREE

Quadratic Congruence

A CONGRUENCE of the form

$$ax^2 + bx + c \equiv 0 \pmod{m},$$

where a, b, and c are INTEGERS. A general quadratic congruence can be reduced to the congruence

$$x^2 \equiv q \pmod{p}$$

and can be solved using EXCLUDENTS, although solution of the general polynomial congruence

$$a_m x^m + \ldots + a_2 x^2 + a_1 x + a_0 \equiv 0 \pmod{n}$$

is intractable.

see also CONGRUENCE, EXCLUDENT, LINEAR CONGRUENCE

Quadratic Curve

The general 2-variable quadratic equation can be written

$$ax^2 + 2bxy + cy^2 + 2dx + 2fy + g = 0. \qquad (1)$$

Define the following quantities:

$$\Delta = \begin{vmatrix} a & b & d \\ b & c & f \\ d & f & g \end{vmatrix} \qquad (2)$$

$$J = \begin{vmatrix} a & b \\ b & c \end{vmatrix} \qquad (3)$$

$$I = a + c \qquad (4)$$

$$K = \begin{vmatrix} a & d \\ d & g \end{vmatrix} + \begin{vmatrix} c & f \\ f & g \end{vmatrix}. \qquad (5)$$

Then the quadratics are classified into the types summarized in the following table (Beyer 1987). The real (nondegenerate) quadratics (the ELLIPSE, HYPERBOLA, and PARABOLA) correspond to the curves which can be created by the intersection of a PLANE with a (two-NAPPES) CONE, and are therefore known as CONIC SECTIONS.

Curve	Δ	J	Δ/I	K
coincident lines	0	0		0
ellipse (imaginary)	$\neq 0$	> 0	> 0	
ellipse (real)	$\neq 0$	> 0	< 0	
hyperbola	$\neq 0$	< 0		
intersecting lines (imaginary)	0	> 0		
intersecting lines (real)	0	< 0		
parabola	$\neq 0$	0		
parallel lines (imaginary)	0	0		> 0
parallel lines (real)	0	0		< 0

It is always possible to eliminate the xy cross term by a suitable ROTATION of the axes. To see this, consider rotation by an arbitrary angle θ. The ROTATION MATRIX is

$$\begin{aligned}\begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x' \\ y' \end{bmatrix} \\ &= \begin{bmatrix} x'\cos\theta + y'\sin\theta \\ -x'\sin\theta + y'\cos\theta \end{bmatrix}, \end{aligned} \tag{6}$$

so

$$x = x'\cos\theta + y'\sin\theta \tag{7}$$

$$y = -x'\sin\theta + y'\cos\theta \tag{8}$$

$$\begin{aligned} xy &= -x'^2\cos\theta\sin\theta + x'y'(\cos^2\theta - \sin^2\theta) \\ &\quad + y'^2\cos\theta\sin\theta \end{aligned} \tag{9}$$

$$x^2 = x'^2\cos^2\theta + 2x'y'\cos\theta\sin\theta + y'^2\sin^2\theta \tag{10}$$

$$y^2 = -x'^2\sin^2\theta - 2x'y'\sin\theta\cos\theta + y'^2\cos^2\theta. \tag{11}$$

Plugging these into (1) gives

$$\begin{aligned} &a(x'^2\cos^2\theta + 2x'y'\cos\theta + y'^2\sin^2\theta) \\ &\quad +2b(x'\cos\theta + y'\sin\theta)(-x'\sin\theta + y'\cos\theta) \\ &\quad +c(x'^2\sin^2\theta - 2x'y'\cos\theta\sin\theta + y'^2\cos^2\theta) \\ &\quad +2d(x'\cos\theta + y'\sin\theta) \\ &\quad +2f(-x'\sin\theta + y'\cos\theta) + g = 0. \end{aligned} \tag{12}$$

Rewriting,

$$\begin{aligned} &a(x'^2\cos^2\theta + 2x'y'\cos\theta + y'^2\sin^2\theta) \\ &\quad +2b(-x^2\cos^2\theta\sin\theta - xy\sin^2\theta + xy\cos^2\theta + y^2\cos\theta\sin\theta) \\ &\quad +c(x'^2\sin^2\theta - 2x'y'\cos\theta\sin\theta + y'^2\cos^2\theta) \\ &\quad +2d(x'\cos\theta + y'\sin\theta) \\ &\quad +2f(-x'\sin\theta + y'\cos\theta) + g = 0. \end{aligned} \tag{13}$$

Grouping terms,

$$\begin{aligned} &x'^2(a\cos^2\theta + c\sin^2\theta - 2b\cos\theta\sin\theta) \\ &\quad +x'y'[2a\cos\theta\sin\theta - 2c\sin\theta\cos\theta + 2b(\cos^2\theta - \sin^2\theta)] \\ &\quad +y'^2(a\sin^2\theta + c\cos^2\theta + 2b\cos\theta\sin\theta) \\ &\quad +x'(2d\cos\theta - 2f\sin\theta) + y'(-2d\sin\theta + 2f\cos\theta) \\ &\quad +g = 0. \end{aligned} \tag{14}$$

Comparing the COEFFICIENTS with (1) gives an equation of the form

$$a'x'^2 + 2b'x'y' + c'y'^2 + 2d'x' + 2f'y' + g' = 0, \tag{15}$$

where the new COEFFICIENTS are

$$a' = a\cos^2\theta - 2b\cos\theta\sin\theta + c\sin^2\theta \tag{16}$$

$$b' = b(\cos^2\theta - \sin^2\theta) + (a - c)\sin\theta\cos\theta \tag{17}$$

$$c' = a\sin^2\theta + 2b\sin\theta\cos\theta + c\cos^2\theta \tag{18}$$

$$d' = d\cos\theta - f\sin\theta \tag{19}$$

$$f' = -d\sin\theta + f\cos\theta \tag{20}$$

$$g' = g. \tag{21}$$

The cross term $2b'x'y'$ can therefore be made to vanish by setting

$$\begin{aligned} b' &= b(\cos^2\theta - \sin^2\theta) - (c - a)\sin\theta\cos\theta \\ &= b\cos(2\theta) - \tfrac{1}{2}(c - a)\sin(2\theta) = 0. \end{aligned} \tag{22}$$

For b' to be zero, it must be true that

$$\cot(2\theta) = \frac{c - a}{2b} \equiv K. \tag{23}$$

The other components are then given with the aid of the identity

$$\cos[\cot^{-1}(x)] = \frac{x}{\sqrt{1 + x^2}} \tag{24}$$

by defining

$$L \equiv \frac{K}{\sqrt{1 + K^2}}, \tag{25}$$

so

$$\sin\theta = \sqrt{\frac{1 - L}{2}} \tag{26}$$

$$\cos\theta = \sqrt{\frac{1 + L}{2}}. \tag{27}$$

Rotating by an angle

$$\theta = \tfrac{1}{2}\cot^{-1}\left(\frac{c - a}{2b}\right) \tag{28}$$

therefore transforms (1) into

$$a'x'^2 + c'y'^2 + 2d'x' + 2f'y' + g' = 0. \tag{29}$$

COMPLETING THE SQUARE,

$$a'\left(x'^2+\frac{2d'}{a'}x\right)+c'\left(y'^2+\frac{2f'}{c'}y'\right)+g'=0 \tag{30}$$

$$a'\left(x'+\frac{d'}{a'}\right)^2+c'\left(y'+\frac{f'}{c'}\right)^2=-g'+\frac{d'^2}{a'}+\frac{f'^2}{c'}. \tag{31}$$

Defining $x''\equiv x'+d'/a'$, $y''\equiv y'+f'/c'$, and $g''\equiv -g'+d'^2/a'+f'^2/c'$ gives

$$a'x''^2+c'y''^2=g''. \tag{32}$$

If $g''\neq 0$, then divide both sides by g''. Defining $a''\equiv a'/g''$ and $c''\equiv c'/g''$ then gives

$$a''x''^2+c''y''^2=1. \tag{33}$$

Therefore, in an appropriate coordinate system, the general CONIC SECTION can be written (dropping the primes) as

$$\begin{cases} ax^2+cy^2=1 & a,c,g\neq 0\\ ax^2+cy^2=0 & a,c\neq 0,\ g=0. \end{cases} \tag{34}$$

Consider an equation of the form $ax^2+2bxy+cy^2=1$ where $b\neq 0$. Re-express this using t_1 and t_2 in the form

$$ax^2+2bxy+cy^2=t_1x'^2+t_2y'^2. \tag{35}$$

Therefore, rotate the COORDINATE SYSTEM

$$\begin{bmatrix} x'\\ y' \end{bmatrix}=\begin{bmatrix} \cos\theta & \sin\theta\\ -\sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x\\ y \end{bmatrix}, \tag{36}$$

so

$$\begin{aligned} ax^2+2bxy+cy^2&=t_1x'^2+t_2y'^2\\ &=t_1(x^2\cos^2\theta+2xy\cos\theta\sin\theta+y^2\sin^2\theta)\\ &\quad+t_2(x^2\sin^2\theta-2xy\sin\theta\cos\theta+y^2\cos^2\theta)\\ &=x^2(t_1\cos^2\theta+t_2\sin^2\theta)+2xy\cos\theta\sin\theta(t_1-t_2)\\ &\quad+y^2(t_1\sin^2\theta+t_2\cos^2\theta) \end{aligned} \tag{37}$$

and

$$a=t_1\cos^2\theta+t_2\sin^2\theta \tag{38}$$

$$b=(t_1-t_2)\cos\theta\sin\theta=\tfrac{1}{2}(t_1-t_2)\sin(2\theta) \tag{39}$$

$$c=t_1\sin^2\theta+t_2\cos^2\theta. \tag{40}$$

Therefore,

$$\begin{aligned} a+c&=(t_1\cos^2\theta+t_2\sin^2\theta)+(t_1\sin^2\theta+t_2\cos^2\theta)\\ &=t_1+t_2 \end{aligned} \tag{41}$$

$$\begin{aligned} a-c&=t_1\cos^2\theta+t_2\sin^2\theta-t_1\sin^2\theta+t_2\cos^2\theta\\ &=(t_1-t_2)(\cos^2\theta-\sin^2\theta)=(t_1-t_2)\cos(2\theta). \end{aligned} \tag{42}$$

From (41) and (42),

$$\frac{a-c}{b}=\frac{(t_1-t_2)\cos(2\theta)}{\frac{1}{2}(t_1-t_2)\sin(2\theta)}=2\cot(2\theta), \tag{43}$$

the same angle as before. But

$$\begin{aligned} \cos(2\theta)&=\cos\left[\cot^{-1}\left(\frac{a-c}{2b}\right)\right]\\ &=\cos\left[\tan^{-1}\left(\frac{2b}{a-c}\right)\right]\\ &=\frac{1}{\sqrt{1+\left(\frac{2b}{a-c}\right)^2}}, \end{aligned} \tag{44}$$

so

$$a-c=\frac{t_1-t_2}{\sqrt{1+\left(\frac{2b}{a-c}\right)^2}}. \tag{45}$$

Rewriting and copying (41),

$$\begin{aligned} t_1-t_2&=(a-c)\sqrt{1+\left(\frac{2b}{a-c}\right)^2}\\ &=\sqrt{(a-c)^2+4b^2} \end{aligned} \tag{46}$$

$$t_1+t_2=a+c. \tag{47}$$

Adding (46) and (47) gives

$$t_1=\tfrac{1}{2}[a+c+\sqrt{(a-c)^2+4b^2}\,] \tag{48}$$

$$t_2=a+c-t_1=\tfrac{1}{2}[a+c-\sqrt{(a-c)^2+4b^2}\,]. \tag{49}$$

Note that these ROOTS can also be found from

$$(t-t_1)(t-t_2)=t^2-t(t_1+t_2)+t_1t_2=0 \tag{50}$$

$$\begin{aligned} t^2&-t(a+c)+\tfrac{1}{4}\{(a+c)^2-[(a-c)^2+4b^2]\}\\ &=t^2-t(a+c)\\ &\quad+\tfrac{1}{4}[a^2+2ac+c^2-a^2+2ac-c^2-4b^2]\\ &=t^2-t(a+c)+(ac-b^2)=(a-t)(c-t)-b^2\\ &=\begin{vmatrix} a-t & b\\ b & c-t \end{vmatrix}=(a-t)(c-t)-b^2=0. \end{aligned} \tag{51}$$

The original problem is therefore equivalent to looking for a solution to

$$\begin{bmatrix} a & b\\ b & c \end{bmatrix}\begin{bmatrix} x\\ y \end{bmatrix}=t\begin{bmatrix} x\\ y \end{bmatrix} \tag{52}$$

$$\begin{bmatrix} ax & bx\\ by & cy \end{bmatrix}\begin{bmatrix} x\\ y \end{bmatrix}=t\begin{bmatrix} x^2\\ y^2 \end{bmatrix}, \tag{53}$$

which gives the simultaneous equations

$$\begin{cases} ax^2+bxy=tx^2\\ bxy+cy^2=ty^2. \end{cases} \tag{54}$$

Let $\mathbf{X}$ be any point (x, y) with old coordinates and (x', y') be its new coordinates. Then

$$ax^2 + 2bxy + cy^2 = t_+ x'^2 + t_- y'^2 = 1 \qquad (55)$$

and

$$x' = \hat{\mathbf{X}}_+ \cdot \begin{bmatrix} x \\ y \end{bmatrix} \qquad (56)$$

$$y' = \hat{\mathbf{X}}_- \cdot \begin{bmatrix} x \\ y \end{bmatrix}. \qquad (57)$$

If t_+ and t_- are both > 0, the curve is an ELLIPSE. If t_+ and t_- are both < 0, the curve is empty. If t_+ and t_- have opposite SIGNS, the curve is a HYPERBOLA. If either is 0, the curve is a PARABOLA.

To find the general form of a quadratic curve in POLAR COORDINATES (as given, for example, in Moulton 1970), plug $x = r\cos\theta$ and $y = r\sin\theta$ into (1) to obtain

$$\begin{aligned} &ar^2\cos^2\theta + 2br^2\cos\theta\sin\theta + cr^2\sin^2\theta \\ &\qquad + 2dr\cos\theta + 2fr\sin\theta + g = 0 \qquad (58) \end{aligned}$$

$$\begin{aligned} &(a\cos^2\theta + 2b\cos\theta\sin\theta + c\sin^2\theta) \\ &\qquad + \frac{2}{r}(d\cos\theta + f\sin\theta) + \frac{g}{r^2} = 0. \qquad (59) \end{aligned}$$

Define $u \equiv 1/r$. For $g \neq 0$,we can divide through by $2g$,

$$\begin{aligned} &\tfrac{1}{2}u^2 + \frac{1}{g}(d\cos\theta + f\sin\theta)u \\ &\qquad + \frac{1}{2g}(a\cos^2\theta + 2b\cos\theta\sin\theta + c\sin^2\theta) = 0. \qquad (60) \end{aligned}$$

Applying the QUADRATIC FORMULA gives

$$u = -\frac{d}{g}\cos\theta - \frac{f}{g}\sin\theta \pm \sqrt{R}, \qquad (61)$$

where

$$\begin{aligned} R &\equiv \frac{(d\cos\theta + f\sin\theta)^2}{g^2} \\ &\quad - 4\left(\frac{1}{2}\right)\left(\frac{1}{2g}\right)(a\cos^2\theta + 2b\cos\theta\sin\theta + c\sin^2\theta) \\ &= \frac{d^2}{g^2}\cos^2\theta + \frac{2df}{g^2}\cos\theta\sin\theta + \frac{f^2}{g^2}\sin^2\theta \\ &\quad - \frac{1}{g}(a\cos^2\theta + 2b\cos\theta\sin\theta + c\sin^2\theta). \qquad (62) \end{aligned}$$

Using the trigonometric identities

$$\sin^2\theta = 1 - \cos^2\theta \qquad (63)$$

$$\sin(2\theta) = 2\sin\theta\cos\theta, \qquad (64)$$

it follows that

$$\begin{aligned} R &= \left(\frac{d^2}{g^2} - \frac{a}{g} - \frac{f^2}{g^2} + \frac{c}{g}\right)\cos^2\theta \\ &\quad + \left(\frac{df}{g^2} - \frac{b}{g}\right)\sin(2\theta) + \left(\frac{f^2}{g^2} - \frac{c}{g}\right) \\ &= \tfrac{1}{2}[1 + \cos(2\theta)]\frac{d^2 - ag - f^2 + cg}{g^2} \\ &\quad + \sin(2\theta)\left(\frac{df - bg}{g^2}\right) + \frac{f^2 - cg}{g^2} \\ &= \frac{d^2 - ag - f^2 + cg}{2g^2}\cos(2\theta) + \frac{df - bg}{g^2}\sin(2\theta) \\ &\quad + \frac{d^2 - ag - f^2 + cg + 2f^2 - 2cg}{2g^2}. \qquad (65) \end{aligned}$$

Defining

$$A \equiv -\frac{f}{g} \qquad (66)$$

$$B \equiv -\frac{d}{g} \qquad (67)$$

$$C \equiv \frac{df - bg}{g^2} \qquad (68)$$

$$D \equiv \frac{d^2 - f^2 + cg - ag}{2g^2} \qquad (69)$$

$$E \equiv \frac{d^2 + f^2 - ag - cg}{2g^2} \qquad (70)$$

then gives the equation

$$u \equiv \frac{1}{r} = A\sin\theta + B\cos\theta \pm \sqrt{C\sin(2\theta) + D\cos(2\theta) + E} \qquad (71)$$

(Moulton 1970). If $g = 0$, then (59) becomes instead

$$u \equiv \frac{1}{r} = -\frac{a\cos^2\theta + 2b\cos\theta\sin\theta + c\sin^2\theta}{2(d\cos\theta + f\sin\theta)}. \qquad (72)$$

Therefore, the general form of a quadratic curve in polar coordinates is given by

$$u = \begin{cases} A\sin\theta + B\cos\theta & \text{for } g \neq 0 \\ \quad \pm\sqrt{C\sin(2\theta) + D\cos(2\theta) + E} & \\ -\frac{a\cos^2\theta + 2b\cos\theta\sin\theta + c\sin^2\theta}{2(d\cos\theta + f\sin\theta)} & \text{for } g = 0. \end{cases} \qquad (73)$$

see also CONIC SECTION, DISCRIMINANT (QUADRATIC CURVE), ELLIPTIC CURVE

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 200–201, 1987.

Casey, J. "The General Equation of the Second Degree." Ch. 4 in *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions, with Numerous Examples, 2nd ed., rev. enl.* Dublin: Hodges, Figgis, & Co., pp. 151–172, 1893.

Moulton, F. R. "Law of Force in Binary Stars" and "Geometrical Interpretation of the Second Law." §58 and 59 in *An Introduction to Celestial Mechanics, 2nd rev. ed.* New York: Dover, pp. 86–89, 1970.

Quadratic Effect

see PRIME QUADRATIC EFFECT

Quadratic Equation

A quadratic equation is a second-order POLYNOMIAL

$$ax^2 + bx + c = 0, \qquad (1)$$

with $a \neq 0$. The roots x can be found by COMPLETING THE SQUARE:

$$x^2 + \frac{b}{a}x = -\frac{c}{a} \qquad (2)$$

$$\left(x + \frac{b}{2a}\right)^2 = -\frac{c}{a} + \frac{b^2}{4a^2} = \frac{b^2 - 4ac}{4a^2} \qquad (3)$$

$$x + \frac{b}{2a} = \frac{\pm\sqrt{b^2 - 4ac}}{2a}. \qquad (4)$$

Solving for x then gives

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \qquad (5)$$

This is the QUADRATIC FORMULA.

An alternate form is given by dividing (1) through by x^2:

$$a + \frac{b}{x} + \frac{c}{x^2} = 0 \qquad (6)$$

$$c\left(\frac{1}{x^2} + \frac{b}{cx}\right) + a = 0 \qquad (7)$$

$$c\left(\frac{1}{x} + \frac{b}{2c}\right)^2 = c\left(\frac{b}{2c}\right)^2 - a = \frac{b^2}{4c} - \frac{4ac}{4c} = \frac{b^2 - 4ac}{4c}. \qquad (8)$$

Therefore,

$$\frac{1}{x} + \frac{b}{2c} = \pm\frac{\sqrt{b^2 - 4ac}}{2c} \qquad (9)$$

$$\frac{1}{x} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c} \qquad (10)$$

$$x = \frac{2c}{-b \pm \sqrt{b^2 - 4ac}}. \qquad (11)$$

This form is helpful if $b^2 \gg 4ac$, in which case the usual form of the QUADRATIC FORMULA can give inaccurate numerical results for one of the ROOTS. This can be avoided by defining

$$q \equiv -\tfrac{1}{2}\left[b + \operatorname{sgn}(b)\sqrt{b^2 - 4ac}\right] \qquad (12)$$

so that b and the term under the SQUARE ROOT sign always have the same sign. Now, if $b > 0$, then

$$q = -\tfrac{1}{2}(b + \sqrt{b^2 - 4ac}) \qquad (13)$$

$$\frac{1}{q} = \frac{-2}{b + \sqrt{b^2 - 4ac}}\,\frac{b - \sqrt{b^2 - 4ac}}{b - \sqrt{b^2 - 4ac}} = \frac{-2(b - \sqrt{b^2 - 4ac})}{b^2 - (b^2 - 4ac)}$$
$$= \frac{-2(b - \sqrt{b^2 - 4ac})}{4ac} = \frac{-b + \sqrt{b^2 - 4ac}}{2ac}, \qquad (14)$$

so

$$x_1 \equiv \frac{q}{a} = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \qquad (15)$$

$$x_2 \equiv \frac{c}{q} = \frac{-b + \sqrt{b^2 - 4ac}}{2a}. \qquad (16)$$

Similarly, if $b < 0$, then

$$q = -\tfrac{1}{2}(b - \sqrt{b^2 - 4ac}) = \tfrac{1}{2}(-b + \sqrt{b^2 - 4ac}) \qquad (17)$$

$$\frac{1}{q} = \frac{2}{-b + \sqrt{b^2 - 4ac}}\,\frac{b + \sqrt{b^2 - 4ac}}{b + \sqrt{b^2 - 4ac}} = \frac{2(b + \sqrt{b^2 - 4ac})}{-b^2 + (b^2 - 4ac)}$$
$$= \frac{b + \sqrt{b^2 - 4ac}}{-2ac} = \frac{-b - \sqrt{b^2 - 4ac}}{2ac}, \qquad (18)$$

so

$$x_1 \equiv \frac{q}{a} = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \qquad (19)$$

$$x_2 \equiv \frac{c}{q} = \frac{-b - \sqrt{b^2 - 4ac}}{2a}. \qquad (20)$$

Therefore, the ROOTS are always given by $x_1 = q/a$ and $x_2 = c/q$.

see also CARLYLE CIRCLE, CONIC SECTION, CUBIC EQUATION, QUARTIC EQUATION, QUINTIC EQUATION, SEXTIC EQUATION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 17, 1972.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 9, 1987.

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 91–92, 1996.

King, R. B. *Beyond the Quartic Equation.* Boston, MA: Birkhäuser, 1996.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Quadratic and Cubic Equations." §5.6 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 178–180, 1992.

Spanier, J. and Oldham, K. B. "The Quadratic Function $ax^2 + bx + c$ and Its Reciprocal." Ch. 16 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 123–131, 1987.

Quadratic Field

An ALGEBRAIC INTEGER of the form $a + b\sqrt{D}$ where D is SQUAREFREE forms a quadratic field and is denoted $\mathbb{Q}(\sqrt{D})$. If $D > 0$, the field is called a REAL QUADRATIC FIELD, and if $D < 0$, it is called an IMAGINARY QUADRATIC FIELD. The integers in $\mathbb{Q}(\sqrt{1})$ are simply called "the" INTEGERS. The integers in $\mathbb{Q}(\sqrt{-1})$ are called GAUSSIAN INTEGERS, and the integers in $\mathbb{Q}(\sqrt{-3})$ are called EISENSTEIN INTEGERS. The ALGEBRAIC INTEGERS in an arbitrary quadratic field do not necessarily have unique factorizations. For example, the fields $\mathbb{Q}(\sqrt{-5})$ and $\mathbb{Q}(\sqrt{-6})$ are not uniquely factorable, since

$$21 = 3 \cdot 7 = (1 + 2\sqrt{-5})(1 - 2\sqrt{-5}) \quad (1)$$

$$6 = -\sqrt{6}\,(\sqrt{-6}) = 2 \cdot 3, \quad (2)$$

although the above factors are all primes within these fields. All other quadratic fields $\mathbb{Q}(\sqrt{D})$ with $|D| \leq 7$ *are* uniquely factorable.

Quadratic fields obey the identities

$$(a + b\sqrt{D}) \pm (c + d\sqrt{D}) = (a \pm c) + (b \pm d)\sqrt{D}, \quad (3)$$

$$(a + b\sqrt{D})(c + d\sqrt{D}) = (ac + bdD) + (ad + bc)\sqrt{D}, \quad (4)$$

and

$$\frac{a + b\sqrt{D}}{c + d\sqrt{D}} = \frac{ac - bdD}{c^2 - d^2 D} + \frac{bc - ad}{c^2 - d^2 D}\sqrt{D}. \quad (5)$$

The INTEGERS in the real field $\mathbb{Q}(\sqrt{D})$ are of the form $r + s\rho$, where

$$\rho = \begin{cases} \sqrt{D} & \text{for } D \equiv 2 \text{ or } D \equiv 3 \pmod 4 \\ \frac{1}{2}(-1 + \sqrt{D}) & \text{for } D \equiv 1 \pmod 4. \end{cases} \quad (6)$$

There exist 22 quadratic fields in which there is a EUCLIDEAN ALGORITHM (Inkeri 1947).

see also ALGEBRAIC INTEGER, EISENSTEIN INTEGER, GAUSSIAN INTEGER, IMAGINARY QUADRATIC FIELD, INTEGER, NUMBER FIELD, REAL QUADRATIC FIELD

References

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 153–154, 1993.

Quadratic Form

A quadratic form involving n REAL variables $x_1, x_2, \ldots, x_n$ associated with the $n \times n$ MATRIX $\mathsf{A} = a_{ij}$ is given by

$$Q(x_1, x_2, \ldots, x_n) = a_{ij} x_i x_j, \quad (1)$$

where EINSTEIN SUMMATION has been used. Letting $\mathbf{x}$ be a VECTOR made up of $x_1, \ldots, x_n$ and $\mathbf{x}^{\mathrm{T}}$ the TRANSPOSE, then

$$Q(\mathbf{x}) = \mathbf{x}^{\mathrm{T}} \mathsf{A} \mathbf{x}, \quad (2)$$

equivalent to

$$Q(\mathbf{x}) = (\mathbf{x}, \mathsf{A}\mathbf{x}) \quad (3)$$

in INNER PRODUCT notation. A BINARY QUADRATIC FORM has the form

$$Q(x, y) = a_{11} x^2 + 2a_{12} xy + a_{22} y^2. \quad (4)$$

It is always possible to express an arbitrary quadratic form

$$Q(\mathbf{x}) = \alpha_{ij} x_i x_j \quad (5)$$

in the form

$$Q(\mathbf{x}) = (\mathbf{x}, \mathsf{A}\mathbf{x}), \quad (6)$$

where $\mathsf{A} = a_{ii}$ is a SYMMETRIC MATRIX given by

$$a_{ij} = \begin{cases} \alpha_{ii} & i = j \\ \frac{1}{2}(\alpha_{ij} + \alpha_{ji}) & i \neq j. \end{cases} \quad (7)$$

Any REAL quadratic form in n variables may be reduced to the diagonal form

$$Q(\mathbf{x}) = \lambda_1 {x_1}^2 + \lambda_2 {x_2}^2 + \ldots + \lambda_n {x_n}^2 \quad (8)$$

with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ by a suitable orthogonal point-transformation. Also, two real quadratic forms are equivalent under the group of linear transformations IFF they have the same RANK and SIGNATURE.

see also DISCONNECTED FORM, INDEFINITE QUADRATIC FORM, INNER PRODUCT, INTEGER-MATRIX FORM, POSITIVE DEFINITE QUADRATIC FORM, POSITIVE SEMIDEFINITE QUADRATIC FORM, RANK (QUADRATIC FORM), SIGNATURE (QUADRATIC FORM), SYLVESTER'S INERTIA LAW

References

Buell, D. A. *Binary Quadratic Forms: Classical Theory and Modern Computations.* New York: Springer-Verlag, 1989.

Conway, J. H. and Fung, F. Y. *The Sensual (Quadratic) Form.* Washington, DC: Math. Assoc. Amer., 1998.

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 1104–106, 1979.

Lam, T. Y. *The Algebraic Theory of Quadratic Forms.* Reading, MA: W. A. Benjamin, 1973.

Quadratic Formula

The formula giving the ROOTS of a QUADRATIC EQUATION

$$ax^2 + bx + c = 0 \quad (1)$$

as

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \quad (2)$$

An alternate form is given by

$$x = \frac{2c}{-b \pm \sqrt{b^2 - 4ac}}. \quad (3)$$

see also QUADRATIC EQUATION

Quadratic Integral

To compute integral of the form

$$\int \frac{dx}{a+bx+cx^2}, \tag{1}$$

COMPLETE THE SQUARE in the DENOMINATOR to obtain

$$\int \frac{dx}{a+bx+cx^2} = \frac{1}{c}\int \frac{dx}{\left(x+\frac{b}{2c}\right)^2 + \left(\frac{a}{c}-\frac{b^2}{4c^2}\right)}. \tag{2}$$

Let $u \equiv x + b/2c$. Then define

$$-A^2 \equiv \frac{a}{c} - \frac{b^2}{4c^2} = \frac{1}{4c^2}(4ac-b^2) \equiv \frac{1}{4c^2}q, \tag{3}$$

where

$$q \equiv 4ac - b^2 \tag{4}$$

is the NEGATIVE of the DISCRIMINANT. If $q < 0$, then

$$A = \frac{1}{2c}\sqrt{-q}. \tag{5}$$

Now use PARTIAL FRACTION DECOMPOSITION,

$$\frac{1}{c}\int \frac{du}{(u+A)(u-A)} = \frac{1}{c}\int \left(\frac{A_1}{u+A} + \frac{A_2}{u-A}\right) du \tag{6}$$

$$\begin{aligned}\left(\frac{A_1}{u+A} + \frac{A_2}{u-A}\right) &= \frac{A_1(u-A)+A_2(u+A)}{u^2-A^2}\\ &= \frac{(A_1+A_2)u + A(A_2-A_1)}{u^2-A^2},\end{aligned} \tag{7}$$

so $A_2 + A_1 = 0 \Rightarrow A_2 = -A_1$ and $A(A_2 - A_1) = -2AA_1 = 1 \Rightarrow A_1 = -1/(2A)$. Plugging these in,

$$\begin{aligned}&\frac{1}{c}\int \left(-\frac{1}{2A}\frac{1}{u+A} + \frac{1}{2A}\frac{1}{u-A}\right) du\\ &= \frac{1}{2Ac}[-\ln(u+A) + \ln(u-A)]\\ &= \frac{1}{2Ac}\ln\left(\frac{u-A}{u+A}\right)\\ &= \frac{1}{2\left(\frac{1}{2c}\right)\sqrt{-q}\,c}\ln\left(\frac{x+\frac{b}{2c}-\frac{1}{2c}\sqrt{-q}}{x+\frac{b}{2c}+\frac{1}{2c}\sqrt{-q}}\right)\\ &= \frac{1}{\sqrt{-q}}\ln\left(\frac{2cx+b-\sqrt{-q}}{2cx+b+\sqrt{-q}}\right)\end{aligned} \tag{8}$$

for $q < 0$. Note that this integral is also tabulated in Gradshteyn and Ryzhik (1979, equation 2.172), where it is given with a sign flipped.

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, 1979.

Quadratic Invariant

Given the BINARY QUADRATIC FORM

$$ax^2 + 2bxy + cy^2 \tag{1}$$

with DISCRIMINANT $b^2 - ac$, let

$$x = pX + qY \tag{2}$$

$$y = rX + sY. \tag{3}$$

Then

$$\begin{aligned}a(pX+qY)^2 + 2b(pX+qY)(rX+sY) + c(rX+sY)^2\\ = AX^2 + 2BXY + CY^2,\end{aligned} \tag{4}$$

where

$$A = ap^2 + 2bpr + cr^2 \tag{5}$$

$$B = apq + b(ps+qr) + crs \tag{6}$$

$$C = aq^2 + 2bqs + cs^2, \tag{7}$$

so

$$\begin{aligned}B^2 - AC &= [a^2p^2q^2 + b^2(ps+qr)^2 + c^2r^2s^2\\ &\quad + 2abpq(ps+qr) + 2acpqrs + 2bcrs(ps+qr)]\\ &\quad - (ap^2 + 2bpr + cr^2)(aq^2 + 2bqs + cs^2)\\ &= a^2p^2q^2 + b^2p^2s^2 + 2b^2pqrs + b^2q^2r^2 + c^2r^2s^2\\ &\quad + 2abp^2qs + 2abpq^2r + 2acpqrs + 2bcprs^2 + 2bcqr^2s\\ &\quad - a^2p^2q^2 - 2abp^2qs - acp^2s^2 - 2abpq^2r - 4b^2pqrs\\ &\quad - 2bcprs^2 - acq^2r^2 - 2bcqr^2s - c^2r^2s^2\\ &= b^2p^2s^2 - 2b^2pqrs + b^2q^2r^2 + 2acpqrs - acp^2s^2\\ &\quad - acq^2r^2\\ &= p^2s^2(b^2-ac) + q^2r^2(b^2-ac) - 2pqrs(b^2-ac)\\ &= (b^2-ac)(p^2s^2 - 2pqrs + q^2r^2)\\ &= (ps-rq)^2(b^2-ac).\end{aligned} \tag{8}$$

Surprisingly, this is the same discriminant as before, but multiplied by the factor $(ps-rq)^2$. The quantity $ps-rq$ is called the MODULUS.

see also ALGEBRAIC INVARIANT

Quadratic Irrational Number

An IRRATIONAL NUMBER of the form

$$\frac{P \pm \sqrt{D}}{Q},$$

where P and Q are INTEGERS and D is a SQUAREFREE INTEGER. Quadratic irrational numbers are sometimes also called QUADRATIC SURDS. In 1770, Lagrange proved that that any quadratic irrational has a CONTINUED FRACTION which is periodic after some point.

see also CONTINUED FRACTION, QUADRATIC SURD

Quadratic Map

A 1-D MAP often called "the" quadratic map is defined by

$$x_{n+1} = {x_n}^2 + c. \tag{1}$$

This is the real version of the complex map defining the MANDELBROT SET. The quadratic map is called attracting if the JACOBIAN $J < 1$, and repelling if $J > 1$. FIXED POINTS occur when

$$x^{(1)} = [x^{(1)}]^2 + c \tag{2}$$

$$(x^{(1)})^2 - x^{(1)} + c = 0 \tag{3}$$

$$x_\pm^{(1)} = \tfrac{1}{2}(1 \pm \sqrt{1-4c}). \tag{4}$$

Period two FIXED POINTS occur when

$$\begin{aligned} x_{n+2} &= {x_{n+1}}^2 + c = ({x_n}^2 + c)^2 + c \\ &= {x_n}^4 + 2c{x_n}^2 + (c^2 + c) = x_n \end{aligned} \tag{5}$$

$$x^4 + 2x^2 - x + (cx^2 + c) = (x^2 - x + c)(x^2 + x + 1 + c) = 0 \tag{6}$$

$$x_\pm^{(2)} = \tfrac{1}{2}[1 \pm \sqrt{1 - 4(1+c)}] = \tfrac{1}{2}(1 \pm \sqrt{-3-4c}). \tag{7}$$

Period three FIXED POINTS occur when

$$\begin{aligned} x^6 + x^5 + (3c+1)x^4 + (2c+1)x^3 + (c^2 + 3c + 1)x^2 \\ +(c+1)^2 x + (c^3 + 2c^2 + c + 1) = 0. \end{aligned} \tag{8}$$

The most general second-order 2-D MAP with an elliptic fixed point at the origin has the form

$$x' = x\cos\alpha - y\sin\alpha + a_{20}x^2 + a_{11}xy + a_{02}y^2 \tag{9}$$

$$y' = x\sin\alpha + y\cos\alpha + b_{20}x^2 + b_{11}xy + b_{02}y^2. \tag{10}$$

The map must have a DETERMINANT of 1 in order to be AREA preserving, reducing the number of independent parameters from seven to three. The map can then be put in a standard form by scaling and rotating to obtain

$$x' = x\cos\alpha - y\sin\alpha + x^2\sin\alpha \tag{11}$$

$$y' = x\sin\alpha + y\cos\alpha - x^2\cos\alpha. \tag{12}$$

The inverse map is

$$x = x'\cos\alpha + y'\sin\alpha \tag{13}$$

$$y = -x'\sin\alpha + y'\cos\alpha + (x'\cos\alpha + y'\sin\alpha)^2. \tag{14}$$

The FIXED POINTS are given by

$$x_i{}^2\sin\alpha + 2x_i\cos\alpha - x_{i-1} - x_{i+1} = 0 \tag{15}$$

for $i = 0, \ldots, n-1$.

see also BOGDANOV MAP, HÉNON MAP, LOGISTIC MAP, LOZI MAP, MANDELBROT SET

Quadratic Mean

see ROOT-MEAN-SQUARE

Quadratic Reciprocity Law

see QUADRATIC RECIPROCITY THEOREM

Quadratic Reciprocity Relations

$$\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2} \tag{1}$$

$$\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8} \tag{2}$$

$$\left(\frac{q}{p}\right) = \left(\frac{p}{q}\right)(-1)^{[(p-1)/2][(q-1)/2]}, \tag{3}$$

where $\left(\frac{p}{q}\right)$ is the LEGENDRE SYMBOL.

see also QUADRATIC RECIPROCITY THEOREM

Quadratic Reciprocity Theorem

Also called the AUREUM THEOREMA (GOLDEN THEOREM) by Gauss. If p and q are distinct ODD PRIMES, then the CONGRUENCES

$$x^2 \equiv q \pmod{p}$$
$$x^2 \equiv p \pmod{q}$$

are both solvable or both unsolvable unless both p and q leave the remainder 3 when divided by 4 (in which case one of the CONGRUENCES is solvable and the other is not). Written symbolically,

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{(p-1)(q-1)/4},$$

where

$$\left(\frac{p}{q}\right) \equiv \begin{cases} 1 & \text{for } x^2 \equiv p \pmod{q} \text{ solvable for } x \\ -1 & \text{for } x^2 \equiv p \pmod{q} \text{ not solvable for } x \end{cases}$$

is known as a LEGENDRE SYMBOL. Legendre was the first to publish a proof, but it was fallacious. Gauss was the first to publish a correct proof. The quadratic reciprocity theorem was Gauss's favorite theorem from NUMBER THEORY, and he devised many proofs of it over his lifetime.

see also JACOBI SYMBOL, KRONECKER SYMBOL, LEGENDRE SYMBOL, QUADRATIC RESIDUE, RECIPROCITY THEOREM

References

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 39, 1996.

Ireland, K. and Rosen, M. "Quadratic Reciprocity." Ch. 5 in *A Classical Introduction to Modern Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 50–65, 1990.

Nagell, T. "Theory of Quadratic Residues." Ch. 4 in *Introduction to Number Theory.* New York: Wiley, 1951.

Riesel, H. "The Law of Quadratic Reciprocity." *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, pp. 279–281, 1994.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 42–49, 1993.

Quadratic Recurrence

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

A quadratic recurrence is a RECURRENCE RELATION on a SEQUENCE of numbers $\{x_n\}$ expressing x_n as a second degree polynomial in x_k with $k < n$. For example,

$$x_n = x_{n-1}x_{n-2} \tag{1}$$

is a quadratic recurrence. Another simple example is

$$x_n = (x_{n-1})^2 \tag{2}$$

with $x_0 = 2$, which has solution $x_n = 2^{2^n}$. Another example is the number of "strongly" binary trees of height $\leq n$, given by

$$y_n = (y_{n-1})^2 + 1 \tag{3}$$

with $y_0 = 1$. This has solution

$$y_n = \left\lfloor c^{2^n} \right\rfloor, \tag{4}$$

where

$$c = \exp\left[\sum_{j=0}^{\infty} 2^{-j-1} \ln(1 + {y_j}^{-2})\right] = 1.502836801\ldots \tag{5}$$

and $\lfloor x \rfloor$ is the FLOOR FUNCTION (Aho and Sloane 1973). A third example is the closest strict underapproximation of the number 1,

$$S_n = \sum_{i=1}^{n} \frac{1}{z_i}, \tag{6}$$

where $1 < z_1 < \ldots < z_n$ are integers. The solution is given by the recurrence

$$z_n = (z_{n-1})^2 - z_{n-1} + 1, \tag{7}$$

with $z_1 = 2$. This has a closed solution as

$$z_n = \left\lfloor d^{2^n} + \tfrac{1}{2} \right\rfloor \tag{8}$$

where

$$d = \tfrac{1}{2}\sqrt{6}\exp\left\{\sum_{j=1}^{\infty} 2^{-j-1} \ln[1 + (2z_j - 1)^{-2}]\right\} = 1.2640847353\ldots \tag{9}$$

(Aho and Sloane 1973). A final example is the well-known recurrence

$$c_n = (c_{n-1})^2 - \mu \tag{10}$$

with $c_0 = 0$ used to generate the MANDELBROT SET.

see also MANDELBROT SET, RECURRENCE RELATION

References

Aho, A. V. and Sloane, N. J. A. "Some Doubly Exponential Sequences." *Fib. Quart.* **11**, 429–437, 1973.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/quad/quad.html`.

Quadratic Residue

If there is an INTEGER x such that

$$x^2 \equiv q \pmod{p}, \tag{1}$$

then q is said to be a quadratic residue of x mod p. If not, q is said to be a quadratic nonresidue of x mod p. For example, $4^2 \equiv 6 \pmod{10}$, so 6 is a quadratic residue (mod 10). The entire set of quadratic residues (mod 10) are given by 1, 4, 5, 6, and 9, since

$$1^2 \equiv 1 \pmod{10} \quad 2^2 \equiv 4 \pmod{10} \quad 3^2 \equiv 9 \pmod{10}$$

$$4^2 \equiv 6 \pmod{10} \quad 5^2 \equiv 5 \pmod{10} \quad 6^2 \equiv 6 \pmod{10}$$

$$7^2 \equiv 9 \pmod{10} \quad 8^2 \equiv 4 \pmod{10} \quad 9^2 \equiv 1 \pmod{10}$$

making the numbers 2, 3, 7, and 8 the quadratic nonresidues (mod 10).

A list of quadratic residues for $p \leq 29$ is given below (Sloane's A046071), with those numbers $< p$ not in the list being quadratic nonresidues of p.

p	Quadratic Residues
1	(none)
2	1
3	1
4	1
5	1, 4
6	1, 3, 4
7	1, 2, 4
8	1, 4
9	1, 4, 7
10	1, 4, 5, 6, 9
11	1, 3, 4, 5, 9
12	1, 4, 9
13	1, 3, 4, 9, 10, 12
14	1, 2, 4, 7, 8, 9, 11
15	1, 4, 6, 9, 10
16	1, 4, 9
17	1, 2, 4, 8, 9, 13, 15, 16
18	1, 4, 7, 9, 10, 13, 16
19	1, 4, 5, 6, 7, 9, 11, 16, 17
20	1, 4, 5, 9, 16

The UNITS in the integers (mod n), $\mathbb{Z}_n$, which are SQUARES are the quadratic residues.

Given an ODD PRIME p and an INTEGER a, then the LEGENDRE SYMBOL is given by

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & \text{if } a \text{ is a quadratic residue mod } p \\ -1 & \text{otherwise.} \end{cases} \tag{2}$$

If

$$r^{(p-1)/2} \equiv \pm 1 \pmod{p}, \tag{3}$$

then r is a quadratic residue (+) or nonresidue (−). This can be seen since if r is a quadratic residue of p, then there exists a square x^2 such that $r \equiv x^2 \pmod{p}$, so

$$r^{(p-1)/2} \equiv (x^2)^{(p-1)/2} \equiv x^{p-1} \pmod{p}, \tag{4}$$

and x^{p-1} is congruent to 1 (mod p) by FERMAT'S LITTLE THEOREM. x is given by

$$\begin{cases} q^{k+1} \pmod{p} \\ \quad \text{for } p = 4k+3 \\ q^{k+1} \pmod{p} \\ \quad \text{for } p = 8k+5 \text{ and } q^{2k+1} \equiv 1 \pmod{p} \\ (4q)^{k+1} \left(\frac{p+1}{2}\right) \pmod{p} \\ \quad \text{for } p = 8k+5 \text{ and } q^{2k+1} \equiv -1 \pmod{p}. \end{cases} \tag{5}$$

More generally, let q be a quadratic residue modulo an ODD PRIME p. Choose h such that the LEGENDRE SYMBOL $(h^2 - 4q/p) = -1$. Then defining

$$V_1 = h \tag{6}$$

$$V_2 = h^2 - 2q \tag{7}$$

$$V_i = hV_{i-1} - qV_{i-2} \qquad \text{for } i \geq 3, \tag{8}$$

gives

$$V_{2i} = V_i{}^2 - 2q^i \tag{9}$$

$$V_{2i+1} = V_i V_{i+1} - hn^i, \tag{10}$$

and a solution to the quadratic CONGRUENCE is

$$x = V_{(p+1)/2} \left(\frac{p+1}{2}\right) \pmod{p}. \tag{11}$$

The following table gives the PRIMES which have a given number d as a quadratic residue.

d	Primes
−6	$24k+1, 5, 7, 11$
−5	$20k+1, 3, 7, 9$
−3	$6k+1$
−2	$8k+1, 3$
−1	$4k+1$
2	$8k \pm 1$
3	$12k \pm 1$
5	$10k \pm 1$
6	$24k \pm 1, 5$

Finding the CONTINUED FRACTION of a SQUARE ROOT $\sqrt{D}$ and using the relationship

$$Q_n = \frac{D - P_n{}^2}{Q_{n-1}} \tag{12}$$

for the nth CONVERGENT P_n/Q_n gives

$$P_n{}^2 \equiv -Q_n Q_{n-1} \pmod{D}. \tag{13}$$

Therefore, $-Q_n Q_{n-1}$ is a quadratic residue of D. But since $Q_1 = 1$, $-Q_2$ is a quadratic residue, as must be $-Q_2 Q_3$. But since $-Q_2$ is a quadratic residue, so is Q_3, and we see that $(-1)^{n-1} Q_n$ are all quadratic residues of D. This method is not guaranteed to produce all quadratic residues, but can often produce several small ones in the case of large D, enabling D to be factored.

The number of SQUARES $s(n)$ in $\mathbb{Z}_n$ is related to the number $q(n)$ of quadratic residues in $\mathbb{Z}_n$ by

$$q(p^n) = s(p^n) - s(p^{n-2}) \tag{14}$$

for $n \geq 3$ (Stangl 1996). Both q and s are MULTIPLICATIVE FUNCTIONS.

see also EULER'S CRITERION, MULTIPLICATIVE FUNCTION, QUADRATIC RECIPROCITY THEOREM, RIEMANN HYPOTHESIS

References

Burton, D. M. *Elementary Number Theory, 4th ed.* New York: McGraw-Hill, p. 201, 1997.

Courant, R. and Robbins, H. "Quadratic Residues." §2.3 in Supplement to Ch. 1 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 38–40, 1996.

Guy, R. K. "Quadratic Residues. Schur's Conjecture" and "Patterns of Quadratic Residues." §F5 and F6 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 244–248, 1994.

Niven, I. and Zuckerman, H. *An Introduction to the Theory of Numbers, 4th ed.* New York: Wiley, p. 84, 1980.

Rosen, K. H. Ch. 9 in *Elementary Number Theory and Its Applications, 3rd ed.* Reading, MA: Addison-Wesley, 1993.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 63–66, 1993.

Sloane, N. J. A. Sequence A046071 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Stangl, W. D. "Counting Squares in $\mathbb{Z}_n$." *Math. Mag.* **69**, 285–289, 1996.

Wagon, S. "Quadratic Residues." §9.2 in *Mathematica in Action.* New York: W. H. Freeman, pp. 292–296, 1991.

Quadratic Sieve Factorization Method

A procedure used in conjunction with DIXON'S FACTORIZATION METHOD to factor large numbers. The rs are chosen as

$$\lfloor \sqrt{n} \rfloor + k, \tag{1}$$

where $k = 1, 2, \ldots$ and $\lfloor x \rfloor$ is the FLOOR FUNCTION. We are then looking for factors p such that

$$n \equiv r^2 \pmod{p}, \tag{2}$$

which means that only numbers with LEGENDRE SYMBOL $(n/p) = 1$ (less than $N = \pi(d)$ for trial divisor d) need be considered. The set of PRIMES for which this is true is known as the FACTOR BASE. Next, the CONGRUENCES

$$x^2 \equiv n \pmod{p} \tag{3}$$

must be solved for each p in the FACTOR BASE. Finally, a sieve is applied to find values of $f(r) = r^2 - n$

which can be factored completely using only the Factor Base. Gaussian Elimination is then used as in Dixon's Factorization Method in order to find a product of the $f(r)$s, yielding a Perfect Square.

The method requires about $\exp(\sqrt{\log n \log\log n})$ steps, improving on the Continued Fraction Factorization Algorithm by removing the 2 under the Square Root (Pomerance 1996). The use of multiple Polynomials gives a better chance of factorization, requires a shorter sieve interval, and is well-suited to parallel processing.

see also Prime Factorization Algorithms, Smooth Number

References

Alford, W. R. and Pomerance, C. "Implementing the Self Initializing Quadratic Sieve on a Distributed Network." In *Number Theoretic and Algebraic Methods in Computer Science, Proc. Internat. Moscow Conf., June–July 1993* (Ed. A. J. van der Poorten, I. Shparlinksi, and H. G. Zimer). World Scientific, pp. 163–174, 1995.

Brent, R. P. "Parallel Algorithms for Integer Factorisation." In *Number Theory and Cryptography* (Ed. J. H. Loxton). New York: Cambridge University Press, 26–37, 1990. `ftp://nimbus.anu.edu.au/pub/Brent/115.dvi.Z`.

Bressoud, D. M. Ch. 8 in *Factorization and Prime Testing.* New York: Springer-Verlag, 1989.

Gerver, J. "Factoring Large Numbers with a Quadratic Sieve." *Math. Comput.* **41**, 287–294, 1983.

Lenstra, A. K. and Manasse, M. S. "Factoring by Electronic Mail." In *Advances in Cryptology—Eurocrypt '89* (Ed. J.-J. Quisquarter and J. Vandewalle). Berlin: Springer-Verlag, pp. 355–371, 1990.

Pomerance, C. "A Tale of Two Sieves." *Not. Amer. Math. Soc.* **43**, 1473–1485, 1996.

Pomerance, C.; Smith, J. W.; and Tuler, R. "A Pipeline Architecture for Factoring Large Integers with the Quadratic Sieve Method." *SIAM J. Comput.* **17**, 387–403, 1988.

Quadratic Surd

see Quadratic Irrational Number

Quadratic Surface

There are 17 standard-form quadratic surfaces. The general quadratic is written

$$ax^2+by^2+cz^2+2fyz+2gzx+2hxy +2px+2qy+2rz+d=0. \quad (1)$$

Define

$$e=\begin{bmatrix} a & h & g \\ h & b & f \\ g & f & c \end{bmatrix} \quad (2)$$

$$E=\begin{bmatrix} a & h & g & p \\ h & b & f & q \\ g & f & c & r \\ p & q & r & d \end{bmatrix} \quad (3)$$

$$\rho_3 = \operatorname{rank} e \quad (4)$$

$$\rho_4 = \operatorname{rank} E \quad (5)$$

$$\Delta = \det E, \quad (6)$$

and k_1, k_2, as k_3 are the roots of

$$\begin{vmatrix} a-x & h & g \\ h & b-x & f \\ g & f & c-x \end{vmatrix} = 0. \quad (7)$$

Also define

$$k \equiv \begin{cases} 1 & \text{if the signs of nonzero } k\text{s are the same} \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

Surface	Equation	ρ_3	ρ_4	sgn (Δ)	k
coincident planes	$x^2=0$	1	1		
ellipsoid ($\Im$)	$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=-1$	3	4	+	1
ellipsoid ($\Re$)	$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1$	3	4	−	1
elliptic cone ($\Im$)	$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=0$	3	3		1
elliptic cone ($\Re$)	$z^2=\frac{x^2}{a^2}+\frac{y^2}{b^2}$	3	3		0
elliptic cylinder ($\Im$)	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=-1$	2	3		1
elliptic cylinder ($\Re$)	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$	2	3		1
elliptic paraboloid	$z=\frac{x^2}{a^2}+\frac{y^2}{b^2}$	2	4	−	1
hyperbolic cylinder	$\frac{x^2}{a^2}-\frac{y^2}{b^2}=-1$	2	3		0
hyperbolic paraboloid	$z=\frac{y^2}{b^2}-\frac{x^2}{a^2}$	2	4	+	0
hyperboloid of one sheet	$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1$	3	4	+	0
hyperboloid of two sheets	$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1$	3	4	−	0
intersecting planes ($\Im$)	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=0$	2	2		1
intersecting planes ($\Re$)	$\frac{x^2}{a^2}-\frac{y^2}{b^2}=0$	2	2		0
parabolic cylinder	$x^2+2rz=0$	1	3		
parallel planes ($\Im$)	$x^2=-a^2$	1	2		
parallel planes ($\Re$)	$x^2=a^2$	1	2		

see also Cubic Surface, Ellipsoid, Elliptic Cone, Elliptic Cylinder, Elliptic Paraboloid, Hyperbolic Cylinder, Hyperbolic Paraboloid, Hyperboloid, Plane, Quartic Surface, Surface

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 210–211, 1987.

Quadratrix of Hippias

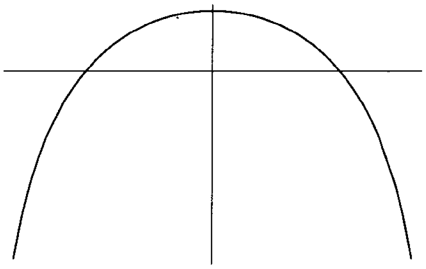

The quadratrix was discovered by Hippias of Elias in 430 BC, and later studied by Dinostratus in 350 BC (MacTutor Archive). It can be used for Angle Trisection or, more generally, division of an Angle into any integral number of equal parts, and Circle Squaring. In Polar Coordinates,

$$\pi\rho = 2r\theta\csc\theta,$$

so

$$r=\frac{\rho\pi\sin\theta}{\theta},$$

which is proportional to the COCHLEOID.

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 195 and 198, 1972.

Lee, X. "Quadratrix of Hippias." `http://www.best.com/~xah/SpecialPlaneCurves_dir/QuadratrixOfHippias_dir/quadratrixOfHippias.html`.

MacTutor History of Mathematics Archive. "Quadratrix of Hippias." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Quadratrix.html`.

Quadrature

The word quadrature has (at least) three incompatible meanings. Integration by quadrature either means solving an INTEGRAL analytically (i.e., symbolically in terms of known functions), or solving of an integral numerically (e.g., GAUSSIAN QUADRATURE, QUADRATURE FORMULAS). The word quadrature is also used to mean SQUARING: the construction of a square using only COMPASS and STRAIGHTEDGE which has the same AREA as a given geometric figure. If quadrature is possible for a PLANE figure, it is said to be QUADRABLE.

For a function tabulated at given values x_i (so the ABSCISSAS cannot be chosen at will), write the function ϕ as a sum of ORTHONORMAL FUNCTIONS p_j satisfying

$$\int_a^b p_i(x)p_j(x)W(x)\,dx = \delta_{ij} \tag{1}$$

as

$$\phi(x) = \sum_{j=0}^{\infty} a_j p_j(x), \tag{2}$$

and plug into

$$\begin{aligned}\int_a^b \phi(x)W(x)\,dx &= \int_a^b \sum_{j=1}^{m} \frac{\pi(x)W(x)}{(x-x_j)\pi'(x_j)}\,dx\, f(x_j)\\ &\equiv \sum_{j=1}^{m} w_j f(x_j),\end{aligned} \tag{3}$$

giving

$$\int_a^b \sum_{j=0}^{\infty} a_j p_j(x)W(x)\,dx = \sum_{i=1}^{n} w_i \left[\sum_{j=0}^{\infty} a_j p_j(x_i)\right]. \tag{4}$$

But we wish this to hold for all degrees of approximation, so

$$a_j \int_a^b p_j(x)W(x)\,dx = a_j \sum_{i=1}^{n} w_i p_j(x_i) \tag{5}$$

$$\int_a^b p_j(x)W(x)\,dx = \sum_{i=1}^{n} w_i p_j(x_i). \tag{6}$$

Setting $i=0$ in (1) gives

$$\int_a^b p_0(x)p_j(x)W(x)\,dx = \delta_{0j}. \tag{7}$$

The zeroth order orthonormal function can always be taken as $p_0(x)=1$, so (7) becomes

$$\int_a^b p_j(x)W(x)\,dx = \delta_{0j} \tag{8}$$

$$= \sum_{i=1}^{n} w_i p_j(x_i), \tag{9}$$

where (6) has been used in the last step. We therefore have the MATRIX equation

$$\begin{bmatrix} p_0(x_1) & \cdots & p_0(x_n)\\ p_1(x_1) & \cdots & p_1(x_n)\\ \vdots & \ddots & \vdots\\ p_{n-1}(x_1) & \cdots & p_{n-1}(x_n)\end{bmatrix}\begin{bmatrix} w_1\\ w_2\\ \vdots\\ w_n\end{bmatrix} = \begin{bmatrix}1\\ 0\\ \vdots\\ 0\end{bmatrix}, \tag{10}$$

which can be inverted to solve for the w_is (Press *et al.* 1992).

see also CALCULUS, CHEBYSHEV-GAUSS QUADRATURE, CHEBYSHEV QUADRATURE, DERIVATIVE, FUNDAMENTAL THEOREM OF GAUSSIAN QUADRATURE, GAUSS-JACOBI MECHANICAL QUADRATURE, GAUSSIAN QUADRATURE, HERMITE-GAUSS QUADRATURE, HERMITE QUADRATURE, JACOBI-GAUSS QUADRATURE, JACOBI QUADRATURE, LAGUERRE-GAUSS QUADRATURE, LAGUERRE QUADRATURE, LEGENDRE-GAUSS QUADRATURE, LEGENDRE QUADRATURE, LOBATTO QUADRATURE, MECHANICAL QUADRATURE, MEHLER QUADRATURE, NEWTON-COTES FORMULAS, NUMERICAL INTEGRATION, RADAU QUADRATURE, RECURSIVE MONOTONE STABLE QUADRATURE

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Integration." §25.4 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 885–897, 1972.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 365–366, 1992.

Quadrature Formulas

see NEWTON-COTES FORMULAS

Quadric

An equation of the form

$$\frac{x^2}{a^2+\theta} + \frac{y^2}{b^2+\theta} + \frac{z^2}{c^2+\theta} = 1,$$

where θ is said to be the parameter of the quadric.

Quadricorn

A FLEXIBLE POLYHEDRON due to C. Schwabe (with the appearance of having four horns) which flexes from one totally flat configuration to another, passing through intermediate configurations of positive VOLUME.

see also FLEXIBLE POLYHEDRON

Quadrifolium

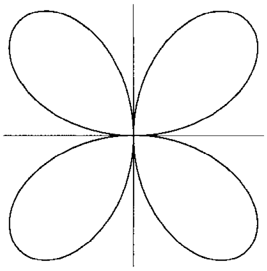

The ROSE with $n = 2$. It has polar equation

$$r = a\sin(2\theta),$$

and Cartesian form

$$(x^2 + y^2)^3 = 4a^2x^2y^2.$$

see also BIFOLIUM, FOLIUM, ROSE, TRIFOLIUM

Quadrilateral

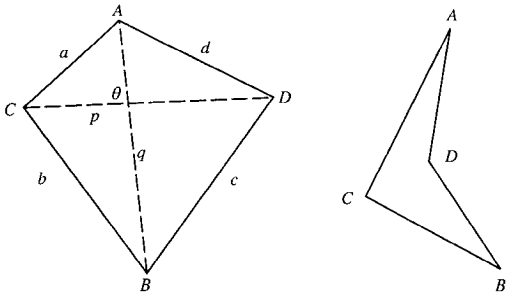

A four-sided POLYGON sometimes (but not very often) also known as a TETRAGON. If not explicitly stated, all four VERTICES are generally taken to lie in a PLANE. If the points do not lie in a PLANE, the quadrilateral is called a SKEW QUADRILATERAL.

For a planar convex quadrilateral (left figure above), let the lengths of the sides be a, b, c, and d, the SEMIPERIMETER s, and the DIAGONALS p and q. The DIAGONALS are PERPENDICULAR IFF $a^2 + c^2 = b^2 + d^2$. An equation for the sum of the squares of side lengths is

$$a^2 + b^2 + c^2 + d^2 = p^2 + q^2 + 4x^2, \tag{1}$$

where x is the length of the line joining the MIDPOINTS of the DIAGONALS. The AREA of a quadrilateral is given by

$$\begin{aligned} K &= \tfrac{1}{2}pq\sin\theta \qquad &(2)\\ &= \tfrac{1}{4}(b^2 + d^2 - a^2 - c^2)\tan\theta \qquad &(3)\\ &= \tfrac{1}{4}\sqrt{4p^2q^2 - (b^2 + d^2 - a^2 - c^2)^2} \qquad &(4)\\ &= \sqrt{(s-a)(s-b)(s-c)(s-d) - abcd\cos^2[\tfrac{1}{2}(A+B)]}, \qquad &(5) \end{aligned}$$

where (4) is known as BRETSCHNEIDER'S FORMULA (Beyer 1987).

A special type of quadrilateral is the CYCLIC QUADRILATERAL, for which a CIRCLE can be circumscribed so that it touches each VERTEX. For BICENTRIC quadrilaterals, the CIRCUMCIRCLE and INCIRCLE satisfy

$$2r^2(R^2 - s^2) = (R^2 - s^2)^2 - 4r^2s^2, \tag{6}$$

where R is the CIRCUMRADIUS, r in the INRADIUS, and s is the separation of centers. A quadrilateral with two sides PARALLEL is called a TRAPEZOID.

There is a relationship between the six distances d_{12}, d_{13}, d_{14}, d_{23}, d_{24}, and d_{34} between the four points of a quadrilateral (Weinberg 1972):

$$\begin{aligned} 0 = {} & {d_{12}}^4{d_{34}}^2 + {d_{13}}^4{d_{24}}^2 + {d_{14}}^4{d_{23}}^2 + {d_{23}}^4{d_{14}}^2\\ & + d_{24}^4d_{13}^2 + d_{34}^4d_{12}^2\\ & + d_{12}^2d_{23}^2d_{31}^2 + d_{12}^2d_{24}^2d_{41}^2 + d_{13}^2d_{34}^2d_{41}^2\\ & + d_{23}^2d_{34}^2d_{42}^2 - d_{12}^2d_{23}^2d_{34}^2 - d_{13}^2d_{32}^2d_{24}^2\\ & - d_{12}^2d_{24}^2d_{43}^2 - d_{14}^2d_{42}^2d_{23}^2 - d_{13}^2d_{34}^2d_{42}^2\\ & - d_{14}^2d_{43}^2d_{32}^2 - d_{23}^2d_{31}^2d_{14}^2 - d_{21}^2d_{13}^2d_{34}^2\\ & - d_{24}^2d_{41}^2d_{13}^2 - d_{21}^2d_{14}^2d_{43}^2 - d_{31}^2d_{12}^2d_{24}^2\\ & - d_{32}^2d_{21}^2d_{14}^2. \end{aligned} \tag{7}$$

see also BIMEDIAN, BRAHMAGUPTA'S FORMULA, BRETSCHNEIDER'S FORMULA, COMPLETE QUADRILATERAL, CYCLIC-INSCRIPTABLE QUADRILATERAL, CYCLIC QUADRILATERAL, DIAMOND, EIGHT-POINT CIRCLE THEOREM, EQUILIC QUADRILATERAL, FANO'S AXIOM, LÉON ANNE'S THEOREM, LOZENGE, ORTHOCENTRIC QUADRILATERAL, PARALLELOGRAM, PTOLEMY'S THEOREM, RATIONAL QUADRILATERAL, RHOMBUS, SKEW QUADRILATERAL, TRAPEZOID, VARIGNON'S THEOREM, VON AUBEL'S THEOREM, WITTENBAUER'S PARALLELOGRAM

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 123, 1987.

Routh, E. J. "Moment of Inertia of a Quadrilateral." *Quart. J. Pure Appl. Math.* **11**, 109–110, 1871.

Weinberg, S. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity.* New York: Wiley, p. 7, 1972.

Quadrillion

In the American system, 10^{15}.

see also LARGE NUMBER

Quadriplanar Coordinates

The analog of TRILINEAR COORDINATES for TETRAHEDRA.

see also TETRAHEDRON, TRILINEAR COORDINATES

References
Altshiller-Court, N. *Modern Pure Solid Geometry.* New York: Macmillan, 1935.
Mitrinović, D. S.; Pečarić, J. E.; and Volenec, V. Ch. 19 in *Recent Advances in Geometric Inequalities.* Dordrecht, Netherlands: Kluwer, 1989.
Woods, F. S. *Higher Geometry: An Introduction to Advanced Methods in Analytic Geometry.* New York: Dover, pp. 193–196, 1961.

Quadruple

A group of four elements, also called a QUADRUPLET or TETRAD.

see also AMICABLE QUADRUPLE, DIOPHANTINE QUADRUPLE, MONAD, PAIR, PRIME QUADRUPLET, PYTHAGOREAN QUADRUPLE, QUADRUPLET, QUINTUPLET, TETRAD, TRIAD, TRIPLE, TWINS, VECTOR QUADRUPLE PRODUCT

Quadruple Point

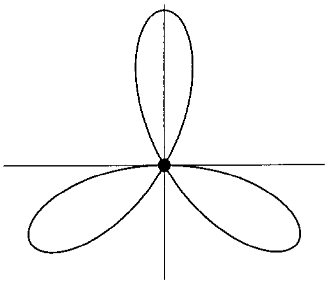

A point where a curve intersects itself along four arcs. The above plot shows the quadruple point at the ORIGIN of the QUADRIFOLIUM $(x^2+y^2)^3-4x^2y^2=0$.

see also DOUBLE POINT, TRIPLE POINT

References
Walker, R. J. *Algebraic Curves.* New York: Springer-Verlag, pp. 57–58, 1978.

Quadruplet

see QUADRUPLE

Quadtree

A TREE having four branches at each node. Quadtrees are used in the construction of some multidimensional databases (e.g., cartography, computer graphics, and image processing). For a d-D tree, the expected number of comparisons over all pairs of integers for successful and unsuccessful searches are given analytically for $d=2$ and numerically for $d \geq 3$ by Finch.

References
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/qdt/qdt.html`.
Flajolet, P.; Gonnet, G.; Puech, C.; and Robson, J. M. "Analytic Variations on Quadtrees." *Algorithmica* **10**, 473–500, 1993.
Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 11–13, 1991.

Quantic

An m-ary n-ic polynomial (i.e., a HOMOGENEOUS POLYNOMIAL with constant COEFFICIENTS of degree n in m independent variables).

see also ALGEBRAIC INVARIANT, FUNDAMENTAL SYSTEM, p-ADIC NUMBER, SYZYGIES PROBLEM

Quantifier

One of the operations EXISTS $\exists$ or FOR ALL $\forall$.

see also BOUND, EXISTS, FOR ALL, FREE

Quantization Efficiency

Quantization is a nonlinear process which generates additional frequency components (Thompson *et al.* 1986). This means that the signal is no longer band-limited, so the SAMPLING THEOREM no longer holds. If a signal is sampled at the NYQUIST FREQUENCY, information will be lost. Therefore, sampling faster than the NYQUIST FREQUENCY results in detection of more of the signal and a lower signal-to-noise ratio [SNR]. Let β be the OVERSAMPLING ratio and define

$$\eta_Q \equiv \frac{\mathrm{SNR_{quant}}}{\mathrm{SNR_{unquant}}}.$$

Then the following table gives values of η_Q for a number of parameters.

Quantization Levels	η_Q ($\beta=1$)	η_Q ($\beta=2$)
2	0.64	0.74
3	0.81	0.89
4	0.88	0.94

The Very Large Array of 27 radio telescopes in Socorro, New Mexico uses three-level quantization at $\beta=1$, so $\eta_Q=0.81$.

References
Thompson, A. R.; Moran, J. M.; and Swenson, G. W. Jr. Fig. 8.3 in *Interferometry and Synthesis in Radio Astronomy.* New York: Wiley, p. 220, 1986.

Quantum Chaos

The study of the implications of CHAOS for a system in the semiclassical (i.e., between classical and quantum mechanical) regime.

References
Ott, E. "Quantum Chaos." Ch. 10 in *Chaos in Dynamical Systems.* New York: Cambridge University Press, pp. 334–362, 1993.

Quarter

The UNIT FRACTION 1/4, also called one-fourth. It is the value of KOEBE'S CONSTANT.

see also HALF, QUARTILE

Quarter Squares Rule

$$\left(\frac{a+b}{2}\right)^2 - \left(\frac{a-b}{2}\right)^2 = ab.$$

Quartet

A SET of four, also called a TETRAD.

see also HEXAD, MONAD, QUINTET, TETRAD, TRIAD

Quartic Curve

A general plane quartic curve is a curve of the form

$$Ax^4 + By^4 + Cx^3y + Dx^2y^2 + Exy^3 + Fx^3 + Gy^3 + Hx^2y + Ixy^2 + Jx^2 + Ky^2 + Lxy + Mx + Ny + O = 0. \tag{1}$$

The incidence relations of the 28 bitangents of the general quartic curve can be put into a ONE-TO-ONE correspondence with the vertices of a particular POLYTOPE in 7-D space (Coxeter 1928, Du Val 1931). This fact is essentially similar to the discovery by Schoutte (1910) that the 27 SOLOMON'S SEAL LINES on a CUBIC SURFACE can be connected with a POLYTOPE in 6-D space (Du Val 1931). A similar but less complete relation exists between the tritangent planes of the canonical curve of genus 4 and an 8-D POLYTOPE (Du Val 1931).

The maximum number of DOUBLE POINTS for a nondegenerate quartic curve is three.

A quartic curve of the form

$$y^2 = (x-\alpha)(x-\xi)(x-\gamma)(x-\delta) \tag{2}$$

can be written

$$\left(\frac{y}{x-\alpha}\right)^2 = \left(1-\frac{\beta-\alpha}{x-\alpha}\right)\left(1-\frac{\gamma-\alpha}{x-\alpha}\right)\left(1-\frac{\delta-\alpha}{x-\alpha}\right), \tag{3}$$

and so is CUBIC in the coordinates

$$X = \frac{1}{x-\alpha} \tag{4}$$

$$Y = \frac{y}{x-\alpha^2}. \tag{5}$$

This transformation is a BIRATIONAL TRANSFORMATION.

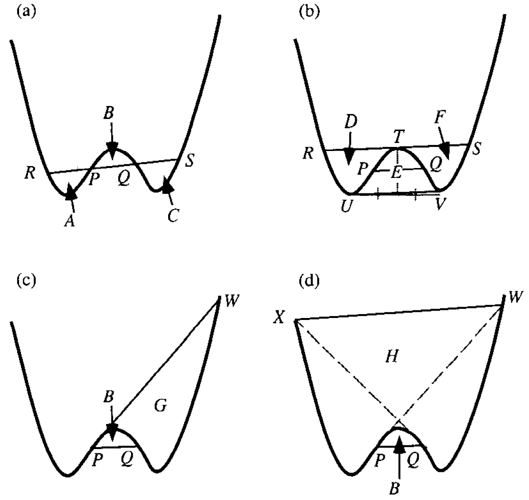

Let P and Q be the INFLECTION POINTS and R and S the intersections of the line PQ with the curve in Figure (a) above. Then

$$A = C \tag{6}$$

$$B = 2A. \tag{7}$$

In Figure (b), let UV be the double tangent, and T the point on the curve whose x coordinate is the average of the x coordinates of U and V. Then $UV \| PQ \| RS$ and

$$D = F \tag{8}$$

$$E = \sqrt{2}\,D. \tag{9}$$

In Figure (c), the tangent at P intersects the curve at W. Then

$$G = 8B. \tag{10}$$

Finally, in Figure (d), the intersections of the tangents at P and Q are W and X. Then

$$H = 27B \tag{11}$$

(Honsberger 1991).

see also CUBIC SURFACE, PEAR-SHAPED CURVE, SOLOMON'S SEAL LINES

References

Coxeter, H. S. M. "The Pure Archimedean Polytopes in Six and Seven Dimensions." *Proc. Cambridge Phil. Soc.* **24**, 7–9, 1928.

Du Val, P. "On the Directrices of a Set of Points in a Plane." *Proc. London Math. Soc. Ser. 2* **35**, 23–74, 1933.

Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., pp. 114–118, 1991.

Schoutte, P. H. "On the Relation Between the Vertices of a Definite Sixdimensional Polytope and the Lines of a Cubic Surface." *Proc. Roy. Akad. Acad. Amsterdam* **13**, 375–383, 1910.

Quartic Equation

A general quartic equation (also called a BIQUADRATIC EQUATION) is a fourth-order POLYNOMIAL of the form

$$z^4 + a_3z^3 + a_2z^2 + a_1z + a_0 = 0. \tag{1}$$

The ROOTS of this equation satisfy NEWTON'S RELATIONS:

$$x_1 + x_2 + x_3 + x_4 = -a_3 \tag{2}$$

$$x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4 = a_2 \tag{3}$$

$$x_1x_2x_3 + x_2x_3x_4 + x_1x_2x_4 + x_1x_3x_4 = -a_1 \tag{4}$$

$$x_1x_2x_3x_4 = a_0, \tag{5}$$

where the denominators on the right side are all $a_4 \equiv 1$.

Ferrari was the first to develop an algebraic technique for solving the general quartic. He applied his technique (which was stolen and published by Cardano) to the equation

$$x^4 + 6x^2 - 60x + 36 = 0 \tag{6}$$

(Smith 1994, p. 207).

The x^3 term can be eliminated from the general quartic (1) by making a substitution of the form

$$z \equiv x - \lambda, \tag{7}$$

so

$$\begin{aligned} x^4 &+ (a_3 - 4\lambda)x^3 + (a_2 - 3a_3\lambda + 6\lambda^2)x^2 \\ &+ (a_1 - 2a_2\lambda + 3a_3\lambda^2 - 4\lambda^3)x \\ &+ (a_0 - a_1\lambda + a_2\lambda^2 - a_3\lambda^3 + \lambda^4). \end{aligned} \tag{8}$$

Letting $\lambda = a_3/4$ so

$$z \equiv x - \tfrac{1}{4}\lambda \tag{9}$$

then gives

$$x^4 + px^2 + qx + r, \tag{10}$$

where

$$p \equiv a_2 - \tfrac{3}{8}a_3^2 \tag{11}$$

$$q \equiv a_1 - \tfrac{1}{2}a_2a_3 + \tfrac{1}{8}a_3^3 \tag{12}$$

$$r \equiv a_0 - \tfrac{1}{4}a_1a_3 + \tfrac{1}{16}a_2a_3^2 - \tfrac{3}{256}a_3^4. \tag{13}$$

Adding and subtracting $x^2u + u^2/4$ to (10) gives

$$x^4 + x^2u + \tfrac{1}{4}u^2 - x^2u - \tfrac{1}{4}u^2 + px^2 + qx + r = 0, \tag{14}$$

which can be rewritten

$$(x^2 - \tfrac{1}{2}u)^2 - [(u - p)x^2 - qx + (\tfrac{1}{4}u^2 - r)] = 0 \tag{15}$$

(Birkhoff and Mac Lane 1965). The first term is a perfect square P^2, and the second term is a perfect square Q^2 for those u such that

$$q^2 = 4(u - p)(\tfrac{1}{4}u^2 - r). \tag{16}$$

This is the resolvent CUBIC, and plugging a solution u_1 back in gives

$$P^2 - Q^2 = (P + Q)(P - Q), \tag{17}$$

so (15) becomes

$$(x^2 + \tfrac{1}{2}u_1 + Q)(x^2 + \tfrac{1}{2}u_1 - Q), \tag{18}$$

where

$$Q \equiv Ax - B \tag{19}$$

$$A \equiv \sqrt{u_1 - p} \tag{20}$$

$$B \equiv -\frac{q}{2A}. \tag{21}$$

Let y_1 be a REAL ROOT of the resolvent CUBIC EQUATION

$$y^3 - a_2y^2 + (a_1a_3 - 4a_0)y + (4a_2a_0 - a_1^2 - a_3^2a_0) = 0. \tag{22}$$

The four ROOTS are then given by the ROOTS of the equation

$$\begin{aligned} x^2 &+ \tfrac{1}{2}(a_3 \pm \sqrt{a_3^2 - 4a_2 + 4y_1}) \\ &+ \tfrac{1}{2}(y_1 \mp \sqrt{y_1^2 - 4a_0}\,) = 0, \end{aligned} \tag{23}$$

which are

$$z_1 = -\tfrac{1}{4}a_3 + \tfrac{1}{2}R + \tfrac{1}{2}D \tag{24}$$

$$z_2 = -\tfrac{1}{4}a_3 + \tfrac{1}{2}R - \tfrac{1}{2}D \tag{25}$$

$$z_3 = -\tfrac{1}{4}a_3 - \tfrac{1}{2}R + \tfrac{1}{2}E \tag{26}$$

$$z_4 = -\tfrac{1}{4}a_3 - \tfrac{1}{2}R - \tfrac{1}{2}E, \tag{27}$$

where

$$R \equiv \sqrt{\tfrac{1}{4}a_3^2 - a_2 + y_1} \tag{28}$$

$$D \equiv \begin{cases} \sqrt{\tfrac{3}{4}a_3^2 - R^2 - 2a_2 + \tfrac{1}{4}(4a_3a_2 - 8a_1 - a_3^3)R^{-1}} \\ R \neq 0 \\ \\ \sqrt{\tfrac{3}{4}a_3^2 - 2a_2 + 2\sqrt{y_1^2 - 4a_0}} \\ R = 0 \end{cases} \tag{29}$$

$$E \equiv \begin{cases} \sqrt{\tfrac{3}{4}a_3^2 - R^2 - 2a_2 - \tfrac{1}{4}(4a_3a_2 - 8a_1 - a_3^3)R^{-1}} \\ R \neq 0 \\ \\ \sqrt{\tfrac{3}{4}a_3^2 - 2a_2 - 2\sqrt{y_1^2 - 4a_0}} \\ R = 0. \end{cases} \tag{30}$$

Another approach to solving the quartic (10) defines

$$\alpha = (x_1 + x_2)(x_3 + x_4) = -(x_1 + x_2)^2 \tag{31}$$

$$\beta = (x_1 + x_3)(x_2 + x_4) = -(x_1 + x_3)^2 \tag{32}$$

$$\gamma = (x_1 + x_4)(x_2 + x_3) = -(x_2 + x_3)^2, \tag{33}$$

where use has been made of

$$x_1 + x_2 + x_3 + x_4 = 0 \tag{34}$$

(which follows since $a_3 = 0$), and

$$h(x) = (x - \alpha)(x - \beta)(x - \gamma) \tag{35}$$

$$= x^3 - (\alpha + \beta + \gamma)x^2 + (\alpha\beta + \alpha\gamma + \beta\gamma)x - \alpha\beta\gamma. \tag{36}$$

Comparing with

$$\begin{aligned} P(x) &= x^3 + px^2 + qx + r && (37)\\ &= (x - x_1)(x - x_2)(x - x_3)(x - x_4) && (38)\\ &= x^3 + \left(\prod_{i \neq j}^{4} x_i x_j\right) x^2 \\ &\quad + (x_1 + x_2)(x_1 + x_3)(x_2 + x_3)x \\ &\quad - x_1 x_2 x_3 (x_1 + x_2 + x_3), && (39) \end{aligned}$$

gives

$$h(x) = x^3 - 2px^2 = (p^2 - r)z + q^2. \qquad (40)$$

Solving this CUBIC EQUATION gives α, β, and γ, which can then be solved for the roots of the quartic x_i (Faucette 1996).

see also CUBIC EQUATION, DISCRIMINANT (POLYNOMIAL), QUINTIC EQUATION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 17–18, 1972.

Berger, M. §16.4.1–16.4.11.1 in *Geometry I.* New York: Springer-Verlag, 1987.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 12, 1987.

Birkhoff, G. and Mac Lane, S. *A Survey of Modern Algebra, 3rd ed.* New York: Macmillan, pp. 107–108, 1965.

Ehrlich, G. §4.16 in *Fundamental Concepts of Abstract Algebra.* Boston, MA: PWS-Kent, 1991.

Faucette, W. M. "A Geometric Interpretation of the Solution of the General Quartic Polynomial." *Amer. Math. Monthly* **103**, 51–57, 1996.

Smith, D. E. *A Source Book in Mathematics.* New York: Dover, 1994.

van der Waerden, B. L. §64 in *Algebra, Vol. 1.* New York: Springer-Verlag, 1993.

Quartic Reciprocity Theorem

Gauss stated the case $n = 4$ using the GAUSSIAN INTEGERS.

see also RECIPROCITY THEOREM

References

Ireland, K. and Rosen, M. "Cubic and Biquadratic Reciprocity." Ch. 9 in *A Classical Introduction to Modern Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 108–137, 1990.

Quartic Surface

An ALGEBRAIC SURFACE of ORDER 4. Unlike CUBIC SURFACES, quartic surfaces have not been fully classified.

see also BOHEMIAN DOME, BURKHARDT QUARTIC, CASSINI SURFACE, CUSHION, CYCLIDE, DESMIC SURFACE, KUMMER SURFACE, MITER SURFACE, PIRIFORM, ROMAN SURFACE, SYMMETROID, TETRAHEDROID, TOOTH SURFACE

References

Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, p. 9, 1986.

Fischer, G. (Ed.). Plates 40–41, 45–49, and 52–56 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 40–41, 45–49, and 52–56, 1986.

Hunt, B. "Some Quartic Surfaces." Appendix B.5 in *The Geometry of Some Special Arithmetic Quotients.* New York: Springer-Verlag, pp. 310–319, 1996.

Jessop, C. *Quartic Surfaces with Singular Points.* Cambridge, England: Cambridge University Press, 1916.

Quartile

One of the four divisions of observations which have been grouped into four equal-sized sets based on their RANK. The quartile including the top RANKED members is called the first quartile and denoted Q_1. The other quartiles are similarly denoted Q_2, Q_3, and Q_4. For N data points with N of the form $4n + 5$ (for $n = 0$, 1, ...), the HINGES are identical to the first and third quartiles.

see also HINGE, INTERQUARTILE RANGE, QUARTILE DEVIATION, QUARTILE VARIATION COEFFICIENT

Quartile Deviation

$$\mathrm{QD} = \tfrac{1}{2}(Q_3 - Q_1),$$

where Q_1 and Q_2 are INTERQUARTILE RANGES.

see also QUARTILE VARIATION COEFFICIENT

Quartile Range

see INTERQUARTILE RANGE

Quartile Skewness Coefficient

see BOWLEY SKEWNESS

Quartile Variation Coefficient

$$V \equiv 100\frac{Q_3 - Q_1}{Q_3 + Q_1},$$

where Q_1 and Q_2 are INTERQUARTILE RANGES.

Quasiamicable Pair

Let $\sigma(m)$ be the DIVISOR FUNCTION of m. Then two numbers m and n are a quasiamicable pair if

$$\sigma(m) = \sigma(n) = m + n + 1.$$

The first few are (48, 75), (140, 195), (1050, 1925), (1575, 1648), ... (Sloane's A005276). Quasiamicable numbers are sometimes called BETROTHED NUMBERS or REDUCED AMICABLE PAIRS.

see also AMICABLE PAIR

References

Beck, W. E. and Najar, R. M. "More Reduced Amicable Pairs." *Fib. Quart.* **15**, 331–332, 1977.

Guy, R. K. "Quasi-Amicable or Betrothed Numbers." §B5 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 59–60, 1994.
Hagis, P. and Lord, G. "Quasi-Amicable Numbers." *Math. Comput.* **31**, 608–611, 1977.
Sloane, N. J. A. Sequence A005276/M5291 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Quasiconformal Map

A generalized Conformal Map.

see also Beltrami Differential Equation

References
Iyanaga, S. and Kawada, Y. (Eds.). "Quasiconformal Mappings." §347 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 1086–1088, 1980.

Quasigroup

A Groupoid S such that for all $a, b \in S$, there exist unique $x, y \in S$ such that

$$\begin{aligned} ax &= b \\ ya &= b. \end{aligned}$$

No other restrictions are applied; thus a quasigroup need not have an Identity Element, not be associative, etc. Quasigroups are precisely Groupoids whose multiplication tables are Latin Squares. A quasigroup can be empty.

see also Binary Operator, Groupoid, Latin Square, Loop (Algebra), Monoid, Semigroup

References
van Lint, J. H. and Wilson, R. M. *A Course in Combinatorics.* New York: Cambridge University Press, 1992.

Quasiperfect Number

A least Abundant Number, i.e., one such that

$$\sigma(n) = 2n + 1.$$

Quasiperfect numbers are therefore the sum of their nontrivial Divisors. No quasiperfect numbers are known, although if any exist, they must be greater than 10^{35} and have seven or more Divisors. Singh (1997) called quasiperfect numbers Slightly Excessive Numbers.

see also Abundant Number, Almost Perfect Number, Perfect Number

References
Guy, R. K. "Almost Perfect, Quasi-Perfect, Pseudoperfect, Harmonic, Weird, Multiperfect and Hyperperfect Numbers." §B2 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 45–53, 1994.
Singh, S. *Fermat's Enigma: The Epic Quest to Solve the World's Greatest Mathematical Problem.* New York: Walker, p. 13, 1997.

Quasiperiodic Function

see Weierstraß Sigma Function, Weierstraß Zeta Function

Quasiperiodic Motion

The type of motion executed by a Dynamical System containing two incommensurate frequencies.

Quasirandom Sequence

A sequence of n-tuples that fills n-space more uniformly than uncorrelated random points. Such a sequence is extremely useful in computational problems where numbers are computed on a grid, but it is not known in advance how fine the grid must be to obtain accurate results. Using a quasirandom sequence allows stopping at any point where convergence is observed, whereas the usual approach of halving the interval between subsequent computations requires a huge number of computations between stopping points.

see also Pseudorandom Number, Random Number

References
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Quasi- (that is, Sub-) Random Sequences." §7.7 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 299–306, 1992.

Quasiregular Polyhedron

A quasiregular polyhedron is the solid region interior to two Dual regular polyhedra with Schläfli Symbols$\{p, q\}$ and $\{q, p\}$. Quasiregular polyhedra are denoted using a Schläfli Symbol of the form $\left\{\begin{matrix}p\\q\end{matrix}\right\}$, with

$$\left\{\begin{matrix}p\\q\end{matrix}\right\} = \left\{\begin{matrix}q\\p\end{matrix}\right\}. \tag{1}$$

Quasiregular polyhedra have two kinds of regular faces with each entirely surrounded by faces of the other kind, equal sides, and equal dihedral angles. They must satisfy the Diophantine inequality

$$\frac{1}{p} + \frac{1}{q} + \frac{1}{r} > 1. \tag{2}$$

But $p, q \geq 3$, so r must be 2. This means that the possible quasiregular polyhedra have symbols $\left\{\begin{matrix}3\\3\end{matrix}\right\}$, $\left\{\begin{matrix}3\\4\end{matrix}\right\}$, and $\left\{\begin{matrix}3\\5\end{matrix}\right\}$. Now

$$\left\{\begin{matrix}3\\3\end{matrix}\right\} = \{3, 4\} \tag{3}$$

is the Octahedron, which is a regular Platonic Solid and not considered quasiregular. This leaves only two convex quasiregular polyhedra: the Cuboctahedron $\left\{\begin{matrix}3\\4\end{matrix}\right\}$ and the Icosidodecahedron $\left\{\begin{matrix}3\\5\end{matrix}\right\}$.

If nonconvex polyhedra are allowed, then additional quasiregular polyhedra are the Great Dodecahedron $\{5, \frac{5}{2}\}$ and the Great Icosidodecahedron $\{3, \frac{5}{2}\}$ (Hart).

For faces to be equatorial $\{h\}$,

$$h = \sqrt{4N_1 + 1} - 1. \tag{4}$$

The EDGES of quasiregular polyhedra form a system of GREAT CIRCLES: the OCTAHEDRON forms three SQUARES, the CUBOCTAHEDRON four HEXAGONS, and the ICOSIDODECAHEDRON six DECAGONS. The VERTEX FIGURES of quasiregular polyhedra are RHOMBUSES (Hart). The EDGES are also all equivalent, a property shared only with the completely regular PLATONIC SOLIDS.

see also CUBOCTAHEDRON, GREAT DODECAHEDRON, GREAT ICOSIDODECAHEDRON, ICOSIDODECAHEDRON, PLATONIC SOLID

References

Coxeter, H. S. M. "Quasi-Regular Polyhedra." §2-3 in *Regular Polytopes, 3rd ed.* New York: Dover, pp. 17–20, 1973.

Hart, G. W. "Quasi-Regular Polyhedra." `http://www.li.net/~george/virtual-polyhedra/quasi-regular-info.html`.

Quasirhombicosidodecahedron

see GREAT RHOMBICOSIDODECAHEDRON (UNIFORM)

Quasirhombicuboctahedron

see GREAT RHOMBICUBOCTAHEDRON (UNIFORM)

Quasisimple Group

A FINITE GROUP L is quasisimple if $L = [L, L]$ and $L/Z(L)$ is a SIMPLE GROUP.

see also COMPONENT, FINITE GROUP, SIMPLE GROUP

Quasithin Theorem

In the classical quasithin case of the QUASI-UNIPOTENT PROBLEM, if G does not have a "strongly embedded" SUBGROUP, then G is a GROUP of LIE-TYPE in characteristic 2 of Lie RANK 2 generated by a pair of parabolic SUBGROUPS P_1 and P_2, or G is one of a short list of exceptions.

see also LIE-TYPE GROUP, QUASI-UNIPOTENT PROBLEM

Quasitruncated Cuboctahedron

see GREAT TRUNCATED CUBOCTAHEDRON

Quasitruncated Dodecadocahedron

see TRUNCATED DODECADODECAHEDRON

Quasitruncated Dodecahedron

see TRUNCATED DODECAHEDRON

Quasitruncated Great Stellated Dodecahedron

see GREAT STELLATED TRUNCATED DODECAHEDRON

Quasitruncated Hexahedron

see STELLATED TRUNCATED HEXAHEDRON

Quasitruncated Small Stellated Dodecahedron

see SMALL STELLATED TRUNCATED DODECAHEDRON

Quasi-Unipotent Group

A GROUP G is quasi-unipotent if every element of G of order p is UNIPOTENT for all PRIMES p such that G has p-RANK ≥ 3.

Quasi-Unipotent Problem

see QUASITHIN THEOREM

Quaternary

The BASE 4 method of counting in which only the DIGITS 0, 1, 2, and 3 are used. These DIGITS have the following multiplication table.

×	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	10	12
3	0	3	12	21

see also BASE (NUMBER), BINARY, DECIMAL, HEXADECIMAL, OCTAL, TERNARY

References

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 9–10, 1991.

Weisstein, E. W. "Bases." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Bases.m`.

Quaternary Tree

see QUADTREE

Quaternion

A member of a *noncommutative* DIVISION ALGEBRA first invented by William Rowan Hamilton. The quaternions are sometimes also known as HYPERCOMPLEX NUMBERS and denoted $\mathbb{H}$. While the quaternions are not commutative, they are associative.

The quaternions can be represented using complex 2×2 MATRICES

$$H = \begin{bmatrix} z & w \\ -w^* & z^* \end{bmatrix} = \begin{bmatrix} a+ib & c+id \\ -c+id & a-ib \end{bmatrix}, \tag{1}$$

where z and w are COMPLEX NUMBERS, a, b, c, and d are REAL, and z^* is the COMPLEX CONJUGATE of z. By analogy with the COMPLEX NUMBERS being representable as a sum of REAL and IMAGINARY PARTS, $a \cdot 1 + bi$, a quaternion can also be written as a linear combination

$$H = a\mathsf{U} + b\mathsf{I} + c\mathsf{J} + d\mathsf{K} \tag{2}$$

of the four matrices

$$\mathsf{U} \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{3}$$

$$\mathsf{I} \equiv \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} \tag{4}$$

$$\mathsf{J} \equiv \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{5}$$

$$\mathsf{K} \equiv \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}. \tag{6}$$

(Note that here, U is used to denote the IDENTITY MATRIX, not I.) The matrices are closely related to the PAULI SPIN MATRICES σ_x, σ_y, σ_z, combined with the IDENTITY MATRIX. From the above definitions, it follows that

$$\mathsf{I}^2 = -\mathsf{U} \tag{7}$$

$$\mathsf{J}^2 = -\mathsf{U} \tag{8}$$

$$\mathsf{K}^2 = -\mathsf{U}. \tag{9}$$

Therefore I, J, and K are three essentially different solutions of the matrix equation

$$\mathsf{X}^2 = -\mathsf{U}, \tag{10}$$

which could be considered the square roots of the negative identity matrix.

In $\mathbb{R}^4$, the basis of the quaternions can be given by

$$i \equiv \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix} \tag{11}$$

$$j \equiv \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{12}$$

$$k \equiv \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} \tag{13}$$

$$1 \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{14}$$

The quaternions satisfy the following identities, sometimes known as HAMILTON'S RULES,

$$i^2 = j^2 = k^2 = -1 \tag{15}$$

$$ij = -ji = k \tag{16}$$

$$jk = -kj = i \tag{17}$$

$$ki = -ik = j. \tag{18}$$

They have the following multiplication table.

	1	i	j	k
1	1	i	j	k
i	i	-1	k	$-j$
j	j	$-k$	-1	i
k	k	j	$-i$	-1

The quaternions ± 1, $\pm i$, $\pm j$, and $\pm k$ form a non-Abelian GROUP of order eight (with multiplication as the group operation) known as Q_8.

The quaternions can be written in the form

$$a = a_1 + a_2 i + a_3 j + a_4 k. \tag{19}$$

The conjugate quaternion is given by

$$a^* = a_1 - a_2 i - a_3 j - a_4 k. \tag{20}$$

The sum of two quaternions is then

$$a+b = (a_1+b_1)+(a_2+b_2)i+(a_3+b_3)j+(a_4+b_4)k, \tag{21}$$

and the product of two quaternions is

$$\begin{aligned} ab = {} & (a_1 b_1 - a_2 b_2 - a_3 b_3 - a_4 b_4) \\ & + (a_1 b_2 + a_2 b_1 + a_3 b_4 - a_4 b_3) i \\ & + (a_1 b_3 - a_2 b_4 + a_3 b_1 + a_4 b_2) j \\ & + (a_1 b_4 + a_2 b_3 - a_3 b_2 + a_4 b_1) k, \end{aligned} \tag{22}$$

so the norm is

$$n(a) = \sqrt{aa^*} = \sqrt{a^* a} = \sqrt{{a_1}^2 + {a_2}^2 + {a_3}^2 + {a_4}^2}. \tag{23}$$

In this notation, the quaternions are closely related to FOUR-VECTORS.

Quaternions can be interpreted as a SCALAR plus a VECTOR by writing

$$a = a_1 + a_2 i + a_3 j + a_4 k = (a_1, \mathbf{a}), \tag{24}$$

where $\mathbf{a} \equiv [a_2\, a_3\, a_4]$. In this notation, quaternion multiplication has the particularly simple form

$$\begin{aligned} q_1 q_2 &= (s_1, \mathbf{v}_1) \cdot (s_2, \mathbf{v}_2) \\ &= (s_1 s_2 - \mathbf{v}_1 \cdot \mathbf{v}_2, s_1 \mathbf{v}_2 + s_2 \mathbf{v}_1 + \mathbf{v}_1 \times \mathbf{v}_2). \end{aligned} \tag{25}$$

Division is uniquely defined (except by zero), so quaternions form a DIVISION ALGEBRA. The inverse of a quaternion is given by

$$a^{-1} = \frac{a^*}{aa^*}, \tag{26}$$

and the norm is multiplicative

$$n(ab) = n(a) n(b). \tag{27}$$

In fact, the product of two quaternion norms immediately gives the EULER FOUR-SQUARE IDENTITY.

A rotation about the UNIT VECTOR $\hat{\mathbf{n}}$ by an angle θ can be computing using the quaternion

$$q = (s, \mathbf{v}) = (\cos(\tfrac{1}{2}\theta), \hat{\mathbf{n}}\sin(\tfrac{1}{2}\theta)) \qquad (28)$$

(Arvo 1994, Hearn and Baker 1996). The components of this quaternion are called EULER PARAMETERS. After rotation, a point $p = (0, \mathbf{p})$ is then given by

$$p' = qpq^{-1} = qpq^*, \qquad (29)$$

since $n(q) = 1$. A concatenation of two rotations, first q_1 and then q_2, can be computed using the identity

$$q_2(q_1pq_1^*)q_2^* = (q_2q_1)p(q_1^*q_2^*) = (q_2q_1)p(q_2q_1)^* \qquad (30)$$

(Goldstein 1980).

see also BIQUATERNION, CAYLEY-KLEIN PARAMETERS, COMPLEX NUMBER, DIVISION ALGEBRA, EULER PARAMETERS, FOUR-VECTOR, OCTONION

References

Altmann, S. L. *Rotations, Quaternions, and Double Groups.* Oxford, England: Clarendon Press, 1986.

Arvo, J. *Graphics Gems 2.* New York: Academic Press, pp. 351–354 and 377–380, 1994.

Baker, A. L. *Quaternions as the Result of Algebraic Operations.* New York: Van Nostrand, 1911.

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Item 107, Feb. 1972.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 230–234, 1996.

Crowe, M. J. *A History of Vector Analysis: The Evolution of the Idea of a Vectorial System.* New York: Dover, 1994.

Dickson, L. E. *Algebras and Their Arithmetics.* New York: Dover, 1960.

Du Val, P. *Homographies, Quaternions, and Rotations.* Oxford, England: Oxford University Press, 1964.

Ebbinghaus, H. D.; Hirzebruch, F.; Hermes, H.; Prestel, A; Koecher, M.; Mainzer, M.; and Remmert, R. *Numbers.* New York: Springer-Verlag, 1990.

Goldstein, H. *Classical Mechanics, 2nd ed.* Reading, MA: Addison-Wesley, p. 151, 1980.

Hamilton, W. R. *Lectures on Quaternions: Containing a Systematic Statement of a New Mathematical Method.* Dublin: Hodges and Smith, 1853.

Hamilton, W. R. *Elements of Quaternions.* London: Longmans, Green, 1866.

Hamilton, W. R. *The Mathematical Papers of Sir William Rowan Hamilton.* Cambridge, England: Cambridge University Press, 1967.

Hardy, A. S. *Elements of Quaternions.* Boston, MA: Ginn, Heath, & Co., 1881.

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Cambridge, England: Clarendon Press, 1965.

Hearn, D. and Baker, M. P. *Computer Graphics: C Version, 2nd ed.* Englewood Cliffs, NJ: Prentice-Hall, pp. 419–420 and 617–618, 1996.

Joly, C. J. *A Manual of Quaternions.* London: Macmillan, 1905.

Julstrom, B. A. "Using Real Quaternions to Represent Rotations in Three Dimensions." *UMAP Modules in Undergraduate Mathematics and Its Applications, Module 652.* Lexington, MA: COMAP, Inc., 1992.

Kelland, P. and Tait, P. G. *Introduction to Quaternions, 3rd ed.* London: Macmillan, 1904.

Nicholson, W. K. *Introduction to Abstract Algebra.* Boston, MA: PWS-Kent, 1993.

Tait, P. G. *An Elementary Treatise on Quaternions, 3rd ed., enl.* Cambridge, England: Cambridge University Press, 1890.

Tait, P. G. "Quaternions." *Encyclopædia Britannica, 9th ed.* ca. 1886. `ftp://ftp.netcom.com/pub/hb/hbaker/quaternion/tait/Encyc-Brit.ps.gz.`

Quattuordecillion

In the American system, 10^{45}.

see also LARGE NUMBER

Queens Problem

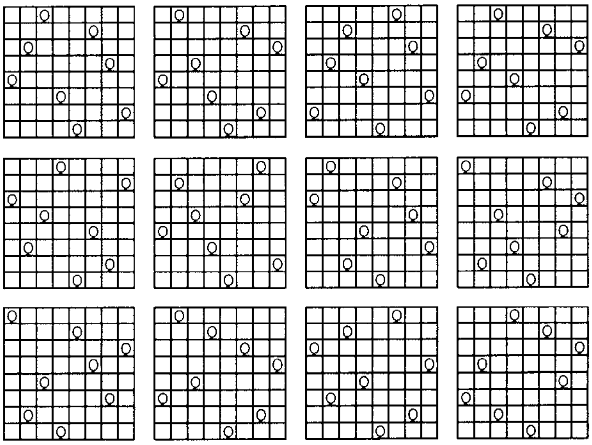

What is the maximum number of queens which can be placed on an $n \times n$ CHESSBOARD such that no two attack one another? The answer is n queens, which gives eight queens for the usual 8×8 board (Madachy 1979). The *number* of different ways the n queens can be placed on an $n \times n$ chessboard so that no two queens may attack each other for the first few n are 1, 0, 0, 2, 10, 4, 40, 92, ... (Sloane's A000170, Madachy 1979). The number of rotationally and reflectively distinct solutions are 1, 0, 0, 1, 2, 1, 6, 12, 46, 92, ... (Sloane's A002562; Dudeney 1970; p. 96). The 12 distinct solutions for $n = 8$ are illustrated above, and the remaining 80 are generated by ROTATION and REFLECTION (Madachy 1979).

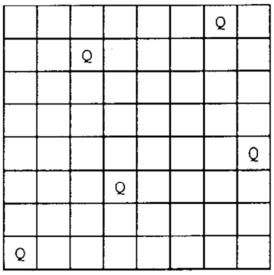

The minimum number of queens needed to occupy or attack all squares of an 8×8 board is 5. Dudeney (1970, pp. 95–96) gave the following results for the number of distinct arrangements $N_p(k,n)$ of k queens attacking or occupying every square of an $n \times n$ board for which every queen is attacked ("protected") by at least one other.

k Queens	$n \times n$	$N_p(k,n)$
2	4	3
3	5	37
3	6	1
4	7	5

Dudeney (1970, pp. 95–96) also gave the following results for the number of distinct arrangements $N_u(k,n)$ of k queens attacking or occupying every square of an $n \times n$ board for which no two queens attack one another (they are "not protected").

k Queens	$n \times n$	$N_u(k,n)$
1	2	1
1	3	1
3	4	2
3	5	2
4	6	17
4	7	1
5	8	91

Vardi (1991) generalizes the problem from a square chessboard to one with the topology of the TORUS. The number of solutions for n queens with n ODD are 1, 0, 10, 28, 0, 88, ... (Sloane's A007705). Vardi (1991) also considers the toroidal "semiqueens" problem, in which a semiqueen can move like a rook or bishop, but only on POSITIVE broken diagonals. The number of solutions to this problem for n queens with n ODD are 1, 3, 15, 133, 2025, 37851, ... (Sloane's A006717), and 0 for EVEN n.

Chow and Velucchi give the solution to the question, "How many different arrangements of k queens are possible on an order n chessboard?" as 1/8th of the COEFFICIENT of $a^k b^{n^2-k}$ in the POLYNOMIAL

$$p(a,b,n) = \begin{cases} (a+b)^{n^2} + 2(a+b)^n (a^2+b^2)^{(n^2-n)/2} \\ \quad +3(a^2+b^2)^{n^2/2} + 2(a^4+b^4)^{n^2/4} & n \text{ even} \\ (a+b)^{n^2} + 2(a+b)(a^4+b^4)^{(n^2-1)/4} \\ \quad +(a+b)(a^2+b^2)^{(n^2-1)/2} \\ \quad +4(a+b)^n (a^2+b^2)^{(n^2-n)/2} & n \text{ odd.} \end{cases}$$

Velucchi also considers the nondominating queens problem, which consists of placing n queens on an order n chessboard to leave a maximum number $U(n)$ of unattacked vacant cells. The first few values are 0, 0, 0, 1, 3, 5, 7, 11, 18, 22, 30, 36, 47, 56, 72, 82, ... (Sloane's A001366). The results can be generalized to k queens on an $n \times n$ board.

see also BISHOPS PROBLEM, CHESS, KINGS PROBLEM, KNIGHTS PROBLEM, KNIGHT'S TOUR, ROOKS PROBLEM

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 166–169, 1987.

Campbell, P. J. "Gauss and the 8-Queens Problem: A Study in the Propagation of Historical Error." *Historia Math.* **4**, 397–404, 1977.

Chow, T. and Velucchi, M. "Different Dispositions in the Chessboard." `http://www.cli.di.unipi.it/~velucchi/diff.txt`.

Dudeney, H. E. "The Eight Queens." §300 in *Amusements in Mathematics.* New York: Dover, p. 89, 1970.

Erbas, C. and Tanik, M. M. "Generating Solutions to the N-Queens Problem Using 2-Circulants." *Math. Mag.* **68**, 343–356, 1995.

Erbas, C.; Tanik, M. M.; and Aliyzaicioglu, Z. "Linear Congruence Equations for the Solutions of the N-Queens Problem." *Inform. Proc. Let.* **41**, 301–306, 1992.

Ginsburg, J. "Gauss's Arithmetization of the Problem of n Queens." *Scripta Math.* **5**, 63–66, 1939.

Guy, R. K. "The n Queens Problem." §C18 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 133–135, 1994.

Kraitchik, M. "The Problem of the Queens" and "Domination of the Chessboard." §10.3 and 10.4 in *Mathematical Recreations.* New York: W. W. Norton, pp. 247–256, 1942.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 34–36, 1979.

Pólya, G. "Über die 'doppelt-periodischen' Lösungen des n-Damen-Problems." In *Mathematische Unterhaltungen und Spiele* (Ed. W. Ahrens). 1918.

Riven, I.; Vardi, I.; and Zimmerman, P. "The n-Queens Problem." *Amer. Math. Monthly* **101**, 629–639, 1994.

Riven, I. and Zabih, R. "An Algebraic Approach to Constraint Satisfaction Problems." In *Proc. Eleventh Internat. Joint Conference on Artificial Intelligence, Vol. 1, August 20–25, 1989.* Detroit, MI: IJCAII, pp. 284–289, 1989.

Ruskey, F. "Information on the n Queens Problem." `http://sue.csc.uvic.ca/~cos/inf/misc/Queen.html`.

Sloane, N. J. A. Sequences A001366, A000170/M1958, A006717/M3005, A007705/M4691, and A002562/M0180 in "An On-Line Version of the Encyclopedia of Integer Sequences."Sloane, N. J. A. and Plouffe, S. Extended entry for M0180 in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.

Vardi, I. "The n-Queens Problems." Ch. 6 in *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 107–125, 1991.

Velucchi, M. "Non-Dominating Queens Problem." `http://www.cli.di.unipi.it/~velucchi/queens.txt`.

Queue

A queue is a special kind of LIST in which elements may only be removed from the bottom by a POP action or added to the top using a PUSH action. Examples of queues include people waiting in line, and submitted jobs waiting to be printed on a printer. The study of queues is called QUEUING THEORY.

see also LIST, QUEUING THEORY, STACK

Queuing Theory

The study of the waiting times, lengths, and other properties of QUEUES.

References

Allen, A. O. *Probability, Statistics, and Queueing Theory with Computer Science Applications.* Orlando, FL: Academic Press, 1978.

Quicksort

The fastest known SORTING ALGORITHM (on average, and for a large number of elements), requiring $\mathcal{O}(n \lg n)$ steps. Quicksort is a recursive algorithm which first partitions an array $\{a_i\}_{i=1}^n$ according to several rules (Sedgewick 1978):

1. Some key ν is in its final position in the array (i.e., if it is the jth smallest, it is in position a_j).
2. All the elements to the left of a_j are less than or equal to a_j. The elements a_1, a_2, ..., a_{j-1} are called the "left subfile."
3. All the elements to the right of a_j are greater than or equal to a_j. The elements a_{j+1}, ..., a_n are called the "right subfile."

Quicksort was invented by Hoare (1961, 1962), has undergone extensive analysis and scrutiny (Sedgewick 1975, 1977, 1978), and is known to be about twice as fast as the next fastest SORTING algorithm. In the worst case, however, quicksort is a slow n^2 algorithm (and for quicksort, "worst case" corresponds to already sorted).

see also HEAPSORT, SORTING

References

Aho, A. V.; Hopcroft, J. E.; and Ullmann, J. D. *Data Structures and Algorithms.* Reading, MA: Addison-Wesley, pp. 260–270, 1987.

Hoare, C. A. R. "Partition: Algorithm 63," "Quicksort: Algorithm 64," and "Find: Algorithm 65." *Comm. ACM* **4**, 321–322, 1961.

Hoare, C. A. R. "Quicksort." *Computer J.* **5**, 10–15, 1962.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Quicksort." §8.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 323–327, 1992.

Sedgewick, R. *Quicksort.* Ph.D. thesis. Stanford Computer Science Report STAN-CS-75-492. Stanford, CA: Stanford University, May 1975.

Sedgewick, R. "The Analysis of Quicksort Programs." *Acta Informatica* **7**, 327–355, 1977.

Sedgewick, R. "Implementing Quicksort Programs." *Comm. ACM* **21**, 847–857, 1978.

Quillen-Lichtenbaum Conjecture

A technical CONJECTURE which connects algebraic k-THEORY to Étale cohomology. The conjecture was made more precise by Dwyer and Friedlander (1982). Thomason (1985) established the first half of this conjecture, but the entire conjecture has not yet been established.

References

Dwyer, W. and Friedlander, E. "Étale K-Theory and Arithmetic." *Bull. Amer. Math. Soc.* **6**, 453–455, 1982.

Thomason, R. W. "Algebraic K-Theory and Étale Cohomology." *Ann. Sci. École Norm. Sup.* **18**, 437–552, 1985.

Weibel, C. A. "The Mathematical Enterprises of Robert Thomason." *Bull. Amer. Math. Soc.* **34**, 1–13, 1996.

Quincunx

The pattern ⁙ of dots on the "5" side of a 6-sided DIE. The word derives from the Latin words for both one and five.

see also DICE

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 9 and 22, 1996.

Quindecillion

In the American system, 10^{48}.

see also LARGE NUMBER

Quintet

A SET of five.

see also HEXAD, MONAD, QUARTET, TETRAD, TRIAD

Quintic Equation

A general quintic cannot be solved algebraically in terms of finite additions, multiplications, and root extractions, as rigorously demonstrated by Abel and Galois.

Euler reduced the general quintic to

$$x^5 - 10qx^2 - p = 0. \tag{1}$$

A quintic also can be algebraically reduced to PRINCIPAL QUINTIC FORM

$$x^5 + a_2x^2 + a_1x + a_0 = 0. \tag{2}$$

By solving a quartic, a quintic can be algebraically reduced to the BRING QUINTIC FORM

$$x^5 - x - a = 0, \tag{3}$$

as was first done by Jerrard.

Consider the quintic

$$\prod_{j=0}^{4} [x - (\omega^j u_1 + \omega^{4j} u_2)] = 0, \tag{4}$$

where $\omega = e^{2\pi i/5}$ and u_1 and u_2 are COMPLEX NUMBERS. This is called DE MOIVRE'S QUINTIC. Generalize it to

$$\prod_{j=0}^{4} [x - (\omega^j u_1 + \omega^{2j} u_2 + \omega^{3j} u_3 + \omega^{4j} u_4)] = 0. \tag{5}$$

Expanding,

$$(\omega^j u_1 + \omega^{2j} u_2 + \omega^{3j} u_3 + \omega^{4j} u_4)^5$$
$$-5U(\omega^j u_1 + \omega^{2j} u_2 + \omega^{3j} u_3 + \omega^{4j} u_4)^4$$
$$-5V(\omega^j u_1 + \omega^{2j} u_2 + \omega^{3j} u_3 + \omega^{4j} u_4)^2$$
$$+5W(\omega^j u_1 + \omega^{2j} u_2 + \omega^{3j} u_3 + \omega^{4j} u_4)$$
$$+[5(X - Y) - Z] = 0, \quad (6)$$

where

$$U = u_1 u_4 + u2u_3 \quad (7)$$
$$V = u_1 {u_2}^2 + u_2 {u_4}^2 + u_3 {u_1}^2 + u_4 {u_3}^2 \quad (8)$$
$$W = {u_1}^2 {u_4}^2 + {u_2}^2 {u_3}^2 - {u_1}^3 u_2 - {u_2}^3 u_4 - {u_3}^3 u_1 - {u_4}^3 u_3 - u_1 u_2 u_3 u_4 \quad (9)$$
$$X = {u_1}^3 u_3 u_4 + {u_2}^3 u_1 u_3 + {u_3}^3 u_2 u_4 + {u_4}^3 u_1 u_2 \quad (10)$$
$$Y = u_1 {u_3}^2 {u_4}^2 + u_2 {u_1}^2 {u_3}^2 + u_3 {u_2}^2 {u_4}^2 + u_4 {u_1}^2 {u_2}^2 \quad (11)$$
$$Z = {u_1}^5 + {u_2}^5 + {u_3}^5 + {u_4}^5. \quad (12)$$

The u_is satisfy

$$u_1 u_4 + u_2 u_3 = 0 \quad (13)$$
$$u_1 {u_2}^2 + u_2 {u_4}^2 + u_3 {u_1}^2 + u_4 {u_3}^2 = 0 \quad (14)$$
$${u_1}^2 {u_4}^2 + {u_2}^2 {u_3}^2 - {u_1}^3 u_2 - {u_2}^3 u_4 - {u_3}^3 u_1 - {u_4}^3 u_3 - u_1 u_2 u_3 u_4 = \tfrac{1}{5} a \quad (15)$$
$$5[({u_1}^3 u_3 u_4 + {u_2}^3 u_1 u_3 + {u_3}^3 u_3 u_4 + {u_4}^3 u_1 u_2) - (u_1 {u_3}^2 {u_4}^2 + u_2 {u_1}^2 {u_3}^2 + u_3 {u_2}^2 {u_4}^2 + u_4 {u_1}^2 {u_2}^2)] - ({u_1}^5 + {u_2}^5 + {u_3}^5 + {u_4}^5) = b. \quad (16)$$

Spearman and Williams (1994) show that an irreducible quintic

$$x^5 + ax + b = 0 \quad (17)$$

with RATIONAL COEFFICIENTS is solvable by radicals IFF there exist rational numbers $\epsilon = \pm 1$, $c \geq 0$, and $e \neq 0$ such that

$$a = \frac{5e^4(3 - 4\epsilon c)}{c^2 + 1} \quad (18)$$
$$b = \frac{-4e^5(11\epsilon + 2c)}{c^2 + 1}. \quad (19)$$

The ROOTS are then

$$x_j = e(\omega^j u_1 + \omega^{2j} u_2 + \omega^{3j} u_3 + \omega^{4j} u_4), \quad (20)$$

where

$$u_1 = \left(\frac{{v_1}^2 v_3}{D^2}\right)^{1/5} \quad (21)$$
$$u_2 = \left(\frac{{v_3}^2 v_4}{D^2}\right)^{1/5} \quad (22)$$
$$u_3 = \left(\frac{{v_2}^2 v_1}{D^2}\right)^{1/5} \quad (23)$$
$$u_4 = \left(\frac{{v_4}^2 v_2}{D^2}\right)^{1/5} \quad (24)$$
$$v_1 = \sqrt{D} + \sqrt{D - \epsilon\sqrt{D}} \quad (25)$$
$$v_2 = -\sqrt{D} - \sqrt{D + \epsilon\sqrt{D}} \quad (26)$$
$$v_3 = -\sqrt{D} + \sqrt{D + \epsilon\sqrt{D}} \quad (27)$$
$$v_4 = \sqrt{D} - \sqrt{D - \epsilon\sqrt{D}} \quad (28)$$
$$D = c^2 + 1. \quad (29)$$

The general quintic can be solved in terms of THETA FUNCTIONS, as was first done by Hermite in 1858. Kronecker subsequently obtained the same solution more simply, and Brioshi also derived the equation. To do so, reduce the general quintic

$$a_5 x^5 + a_4 x^4 + a_3 x^3 + a_2 x^2 + a_1 x + a_0 = 0 \quad (30)$$

into BRING QUINTIC FORM

$$x^5 - x + \rho = 0. \quad (31)$$

Then define

$$k \equiv \tan\left[\tfrac{1}{4}\sin^{-1}\left(\frac{16}{25\sqrt{5}\,\rho^2}\right)\right] \quad (32)$$
$$s \equiv \begin{cases} -\operatorname{sgn}(\Im[\rho]) & \text{for } \Re[\rho] = 0 \\ \operatorname{sgn}(\Re[\rho]) & \text{for } \Re[\rho] \neq 0 \end{cases} \quad (33)$$
$$b = \frac{s(k^2)^{1/8}}{2 \cdot 5^{3/4}\sqrt{k(1 - k^2)}} \quad (34)$$
$$q = q(k^2) = e^{i\pi K'(k^2)/K(k^2)}, \quad (35)$$

where k is the MODULUS, $m \equiv k^2$ is the PARAMETER, and q is the NOME. Solving

$$q(m) = e^{i\pi K'(m)/K(m)} \quad (36)$$

for m gives the inverse parameter

$$m(q) = \frac{{\vartheta_2}^4(q)}{{\vartheta_3}^4(q)}. \quad (37)$$

The ROOTS are then given by

$$\begin{aligned}x_1 &= (-1)^{3/4}b\{[m(e^{-2\pi i/5}q^{1/5})]^{1/8}\\ &\quad +i[m(e^{2\pi i/5}q^{1/5})]^{1/8}\}\\ &\quad \times\{[m(e^{-4\pi i/5}q^{1/5})]^{1/8}+[m(e^{4\pi i/5}q^{1/5})]^{1/8}\}\\ &\quad \times\{[m(q^{1/5})]^{1/8}+q^{5/8}(q^5)^{-1/8}[m(q^5)]^{1/8}\} \qquad (38)\end{aligned}$$

$$\begin{aligned}x_2 &= b\{-[m(q^{1/5})]^{1/8}+e^{3\pi i/4}[m(e^{2\pi i/5}q^{1/5})]^{1/8}\}\\ &\quad \times\{e^{-3\pi i/4}[m(e^{-2\pi i/5}q^{1/5})]^{1/8}+i[m(e^{4\pi i/5}q^{1/5})]^{1/8}\}\\ &\quad \times\{i[m(e^{-4\pi i/5}q^{1/5})]^{1/8}+q^{5/8}(q^5)^{-1/8}[m(q^5)]^{1/8}\}\end{aligned} \qquad (39)$$

$$\begin{aligned}x_3 &= b\{e^{-3\pi i/4}[m(e^{-2\pi i/5}q^{1/5})]^{1/8}\\ &\quad -i[m(e^{-4\pi i/5}q^{1/5})]^{1/8})\}\{-[m(q^{1/5})]^{1/8}\\ &\quad -i[m(e^{4\pi i/5}q^{1/5})]^{1/8}\}\\ &\quad \times\{e^{3\pi i/4}[m(e^{2\pi i/5}q^{1/5})]^{1/8}+q^{5/8}(q^5)^{-1/8}[m(q^5)]^{1/8}\}\end{aligned} \qquad (40)$$

$$\begin{aligned}x_4 &= b\{[m(q^{1/5})]^{1/8}-i[m(e^{-4\pi i/5}q^{1/5})]^{1/8})\}\\ &\quad \times\{-e^{3\pi i/4}[m(e^{2\pi i/5}q^{1/5})]^{1/8}-i[m(e^{4\pi i/5}q^{1/5})]^{1/8}\}\\ &\quad \times\{e^{-3\pi i/4}[m(e^{-2\pi i/5}q^{1/5})]^{1/8}\\ &\quad +q^{5/8}(q^5)^{-1/8}[m(q^5)]^{1/8}\} \qquad (41)\end{aligned}$$

$$\begin{aligned}x_5 &= b\{[m(q^{1/5})]^{1/8}-e^{-3\pi i/4}[m(e^{-2\pi i/5}q^{1/5})]^{1/8}\}\\ &\quad \times\{-e^{3\pi i/4}[m(e^{2\pi i/5}q^{1/5})]^{1/8}+i[m(e^{-4\pi i/5}q^{1/5})]^{1/8}\}\\ &\quad \times\{(-i[m(e^{4\pi i/5}q^{1/5})]^{1/8}+q^{5/8}(q^5)^{-1/8}[m(q^5)]^{1/8}\}.\end{aligned} \qquad (42)$$

Felix Klein used a TSCHIRNHAUSEN TRANSFORMATION to reduce the general quintic to the form

$$z^5+5az^2+5bz+c=0. \qquad (43)$$

He then solved the related ICOSAHEDRAL EQUATION

$$\begin{aligned}I(z,1,Z) &= z^5(-1+11z^5+z^{10})^5\\ &\quad -[1+z^{30}-10005(z^{10}+z^{20})+522(-z^5+z^{25})]^2Z=0,\end{aligned} \qquad (44)$$

where Z is a function of radicals of a, b, and c. The solution of this equation can be given in terms of HYPERGEOMETRIC FUNCTIONS as

$$\frac{Z^{-1/60}{}_2F_1(-\frac{1}{60},\frac{29}{60},\frac{4}{5},1728Z)}{Z^{11/60}{}_2F_1(\frac{11}{60},\frac{41}{60},\frac{6}{5},1728Z)}. \qquad (45)$$

Another possible approach uses a series expansion, which gives one root (the first one in the list below) of

$$t^5-t-\rho. \qquad (46)$$

All five roots can be derived using differential equations (Cockle 1860, Harley 1862). Let

$$F_1(\rho)=F_2(\rho) \qquad (47)$$

$$F_2(\rho)={}_4F_3(\tfrac{1}{5},\tfrac{2}{5},\tfrac{3}{5},\tfrac{4}{5};\tfrac{1}{2},\tfrac{3}{4},\tfrac{5}{4};\tfrac{3125}{256}\rho^4) \qquad (48)$$

$$F_3(\rho)={}_4F_3(\tfrac{9}{20},\tfrac{13}{20},\tfrac{17}{20},\tfrac{21}{20};\tfrac{3}{4},\tfrac{5}{4},\tfrac{3}{2};\tfrac{3125}{256}\rho^4) \qquad (49)$$

$$F_4(\rho)={}_4F_3(\tfrac{7}{10},\tfrac{9}{10},\tfrac{11}{10},\tfrac{13}{10};\tfrac{5}{4},\tfrac{3}{2},\tfrac{7}{4};\tfrac{3125}{256}\rho^4), \qquad (50)$$

then the ROOTS are

$$t_1=-\rho\,{}_4F_3(\tfrac{1}{5},\tfrac{2}{5},\tfrac{3}{5},\tfrac{4}{5};\tfrac{1}{2},\tfrac{3}{4},\tfrac{5}{4};\tfrac{3125}{256}\rho^4) \qquad (51)$$

$$t_2=-F_1(\rho)+\tfrac{1}{4}\rho F_2(\rho)+\tfrac{5}{32}\rho^2F_3(\rho)+\tfrac{5}{32}\rho^3F_4(\rho) \qquad (52)$$

$$t_3=-F_1(\rho)+\tfrac{1}{4}\rho F_2(\rho)-\tfrac{5}{32}\rho^2F_3(\rho)+\tfrac{5}{32}\rho^3F_4(\rho) \qquad (53)$$

$$t_4=-iF_1(\rho)+\tfrac{1}{4}\rho F_2(\rho)-\tfrac{5}{32}i\rho^2F_3(\rho)-\tfrac{5}{32}\rho^3F_4(\rho) \qquad (54)$$

$$t_5=iF_1(\rho)+\tfrac{1}{4}\rho F_2(\rho)+\tfrac{5}{32}i\rho^2F_3(\rho)-\tfrac{5}{32}\rho^3F_4(\rho). \qquad (55)$$

This technique gives closed form solutions in terms of HYPERGEOMETRIC FUNCTIONS in one variable for any POLYNOMIAL equation which can be written in the form

$$x^p+bx^q+c. \qquad (56)$$

Cadenhad, Young, and Runge showed in 1885 that all irreducible solvable quintics with COEFFICIENTS of x^4, x^3, and x^2 missing have the following form

$$x^5+\frac{5\mu^4(4\nu+3)}{\nu^2+1}x+\frac{4\mu^5(2\nu+1)(4\nu+3)}{\nu^2+1}=0, \qquad (57)$$

where μ and ν are RATIONAL.

see also BRING QUINTIC FORM, BRING-JERRARD QUINTIC FORM, CUBIC EQUATION, DE MOIVRE'S QUINTIC, PRINCIPAL QUINTIC FORM, QUADRATIC EQUATION, QUARTIC EQUATION, SEXTIC EQUATION

References

Birkhoff, G. and Mac Lane, S. *A Survey of Modern Algebra, 3rd ed.* New York: Macmillan, pp. 418–421, 1965.

Chowla, S. "On Quintic Equations Soluble by Radicals." *Math. Student* **13**, 84, 1945.

Cockle, J. "Sketch of a Theory of Transcendental Roots." *Phil. Mag.* **20**, 145–148, 1860.

Cockle, J. " On Transcendental and Algebraic Solution—Supplemental Paper." *Phil. Mag.* **13**, 135–139, 1862.

Davis, H. T. *Introduction to Nonlinear Differential and Integral Equations.* New York: Dover, p. 172, 1960.

Dummit, D. S. "Solving Solvable Quintics." *Math. Comput.* **57**, 387–401, 1991.

Glashan, J. C. "Notes on the Quintic." *Amer. J. Math.* **8**, 178–179, 1885.

Harley, R. "On the Solution of the Transcendental Solution of Algebraic Equations." *Quart. J. Pure Appl. Math.* **5**, 337–361, 1862.

Hermite, C. "Sulla risoluzione delle equazioni del quinto grado." *Annali di math. pura ed appl.* **1**, 256–259, 1858.

King, R. B. *Beyond the Quartic Equation.* Boston, MA: Birkhäuser, 1996.
King, R. B. and Cranfield, E. R. "An Algorithm for Calculating the Roots of a General Quintic Equation from Its Coefficients." *J. Math. Phys.* **32**, 823–825, 1991.
Rosen, M. I. "Niels Hendrik Abel and Equations of the Fifth Degree." *Amer. Math. Monthly* **102**, 495–505, 1995.
Shurman, J. *Geometry of the Quintic.* New York: Wiley, 1997.
Spearman, B. K. and Williams, K. S. "Characterization of Solvable Quintics $x^5 + ax + b$." *Amer. Math. Monthly* **101**, 986–992, 1994.
Wolfram Research. "Solving the Quintic." Poster. Champaign, IL: Wolfram Research, 1995. `http://www.wolfram.com/posters/quintic`.
Young, G. P. "Solution of Solvable Irreducible Quintic Equations, Without the Aid of a Resolvent Sextic." *Amer. J. Math.* **7**, 170–177, 1885.

Quintic Surface

A quintic surface is an ALGEBRAIC SURFACE of degree 5. Togliatti (1940, 1949) showed that quintic surfaces having 31 ORDINARY DOUBLE POINTS exist, although he did not explicitly derive equations for such surfaces. Beauville (1978) subsequently proved that 31 double points was the maximum possible, and quintic surfaces having 31 ORDINARY DOUBLE POINTS are therefore sometimes called TOGLIATTI SURFACES. van Straten (1993) subsequently constructed a 3-D family of solutions and in 1994, Barth derived the example known as the DERVISH.

see also ALGEBRAIC SURFACE, DERVISH, KISS SURFACE, ORDINARY DOUBLE POINT

References

Beauville, A. "Surfaces algébriques complexes." *Astérisque* **54**, 1–172, 1978.
Endraß, S. "Togliatti Surfaces." `http://www.mathematik.uni-mainz.de/Algebraische Geometrie/docs/Etogliatti.shtml`.
Hunt, B. "Algebraic Surfaces." `http://www.mathematik.uni-kl.de/~wwwagag/Galerie.html`.
Togliatti, E. G. "Una notevole superficie de 5° ordine con soli punti doppi isolati." *Vierteljschr. Naturforsch. Ges. Zürich* **85**, 127–132, 1940.
Togliatti, E. "Sulle superficie monoidi col massimo numero di punti doppi." *Ann. Mat. Pura Appl.* **30**, 201–209, 1949.
van Straten, D. "A Quintic Hypersurface in $\mathbb{P}^4$ with 130 Nodes." *Topology* **32**, 857–864, 1993.

Quintillion

In the American system, 10^{18}.

see also LARGE NUMBER

Quintuple

A group of five elements, also called a QUINTUPLET or PENTAD.

see also MONAD, PAIR, PENTAD, QUADRUPLE, QUADRUPLET, QUINTUPLET, TETRAD, TRIAD, TRIPLET, TWINS

Quintuple Product Identity

A.k.a. the WATSON QUINTUPLE PRODUCT IDENTITY.

$$\prod_{n=1}^{\infty}(1-q^n)(1-zq^n)(1-z^{-1}q^{n-1})(1-z^2q^{2n-1}) \times(1-z^{-2}q^{2n-1}) = \sum_{m=-\infty}^{\infty}(z^{3m}-z^{-3m-1})q^{m(2m+1)/2}. \quad (1)$$

It can also be written

$$\prod_{n=1}^{\infty}(1-q^{2n})(1-q^{2n-1}z)(1-q^{2n-1}z^{-1}) \times(1-q^{4n-3}z^2)(1-q^{4n-4}z^{-2}) = \sum_{n=-\infty}^{\infty} q^{3n^2-2n}[(z^{3n}+z^{-3n})-(z^{3n-2}+z^{-(3n-2)})] \quad (2)$$

or

$$\sum_{k=-\infty}^{\infty}(-1)^k q^{(3k^2-k)/2}x^{3k}(1+zq^k) = \prod_{j=1}^{\infty}(1-q^j)(1+z^{-1}q^j)(1+zq^{j-1}) \times(1+z^{-2}q^{2j-1})(1+z^2q^{2j-1}). \quad (3)$$

Using the NOTATION of the RAMANUJAN THETA FUNCTION (Berndt, p. 83),

$$f(B^3/q, q^5/B^3) - B^2 f(q/B^3, B^3q^5) = f(-q^2)\frac{f(-B^2, -q^2/B^2)}{f(Bq, q/B)}. \quad (4)$$

see also JACOBI TRIPLE PRODUCT, RAMANUJAN THETA FUNCTIONS

References

Berndt, B. C. *Ramanujan's Notebooks, Part III.* New York: Springer-Verlag, 1985.
Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, pp. 306–309, 1987.
Gasper, G. and Rahman, M. *Basic Hypergeometric Series.* Cambridge, England: Cambridge University Press, 1990.

Quintuplet

A group of five elements, also called a QUINTUPLE or PENTAD.

see also MONAD, PAIR, PENTAD, QUADRUPLE, QUADRUPLET, QUINTUPLET, TETRAD, TRIAD, TRIPLET, TWINS

Quota Rule

A RECURRENCE RELATION between the function Q arising in QUOTA SYSTEMS,

$$Q(n,r) = Q(n-1,r-1) + Q(n-1,r).$$

References

Young, S. C.; Taylor, A. D.; and Zwicker, W. S. "Counting Quota Systems: A Combinatorial Question from Social Choice Theory." *Math. Mag.* **68**, 331–342, 1995.

Quota System

A generalization of simple majority voting in which a list of quotas $\{q_0, \ldots, q_n\}$ specifies, according to the number of votes, how many votes an alternative needs to win (Taylor 1995). The quota system declares a tie unless for some k, there are exactly k tie votes in the profile and one of the alternatives has at least q_k votes, in which case the alternative is the choice.

Let $Q(n)$ be the number of quota systems for n voters and $Q(n,r)$ the number of quota systems for which $q_0 = r+1$, so

$$Q(n) = \sum_{r=\lfloor n/2 \rfloor}^{n} Q(n,r) = \binom{n+1}{\lfloor \frac{n}{2} \rfloor + 1},$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION. This produces the sequence of CENTRAL BINOMIAL COEFFICIENTS 1, 2, 3, 6, 10, 20, 35, 70, 126, ... (Sloane's A001405). It may be defined recursively by $Q(0) = 1$ and

$$Q(n+1) = \begin{cases} 2Q(n) & \text{for } n \text{ even} \\ 2Q(n) - C_{(n+1)/2} & \text{for } n \text{ odd,} \end{cases}$$

where C_k is a CATALAN NUMBER (Young *et al.* 1995). The function $Q(n,r)$ satisfies

$$Q(n,r) = \binom{n+1}{r+1} - \binom{n+1}{r+2}$$

for $r > n/2 - 1$ (Young *et al.* 1995). $Q(n,r)$ satisfies the QUOTA RULE.

see also BINOMIAL COEFFICIENT, CENTRAL BINOMIAL COEFFICIENT

References

Sloane, N. J. A. Sequence A001405/M0769 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Taylor, A. *Mathematics and Politics: Strategy, Voting, Power, and Proof.* New York: Springer-Verlag, 1995.

Young, S. C.; Taylor, A. D.; and Zwicker, W. S. "Counting Quota Systems: A Combinatorial Question from Social Choice Theory." *Math. Mag.* **68**, 331–342, 1995.

Quotient

The ratio $q = r/s$ of two quantities r and s, where $s \neq 0$.

see also DIVISION, QUOTIENT GROUP, QUOTIENT RING, QUOTIENT SPACE

Quotient-Difference Algorithm

The ALGORITHM of constructing and interpreting a QUOTIENT-DIFFERENCE TABLE which allows interconversion of CONTINUED FRACTIONS, POWER SERIES, and RATIONAL FUNCTIONS approximations.

see also QUOTIENT-DIFFERENCE TABLE

Quotient-Difference Table

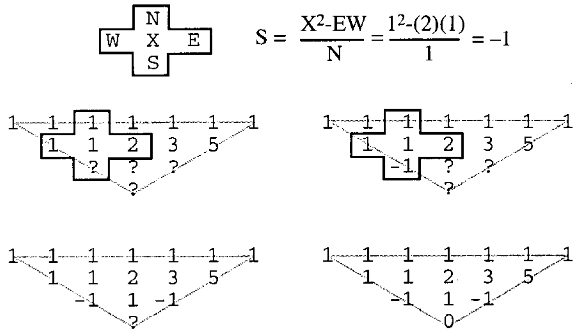

A quotient-difference table is a triangular ARRAY of numbers constructed by drawing a sequence of n numbers in a horizontal row and placing a 1 above each. An additional "1" is then placed at the beginning and end of the row of 1s, and the value of rows underneath the original row is then determined by looking at groups of adjacent numbers

$$\begin{array}{ccc} & N & \\ W & X & E \\ & S & \end{array}$$

and computing

$$S = \frac{X^2 - EW}{N}$$

for the elements falling within a triangle formed by the diagonals extended from the first and last "1," as illustrated above.

0s in quotient-difference tables form square "windows" which are bordered by GEOMETRIC PROGRESSIONS. Quotient-difference tables eventually yield a row of 0s IFF the starting sequence is defined by a linear RECURRENCE RELATION. For example, continuing the above example generated by the FIBONACCI NUMBERS

$$\begin{array}{ccccccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ & 1 & 1 & 2 & 3 & 5 & \\ & & -1 & 1 & -1 & & \\ & & & 0 & & & \end{array}$$

$$\begin{array}{cccccccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ & 1 & 1 & 2 & 3 & 5 & 8 & \\ & & -1 & 1 & -1 & 1 & & \\ & & & 0 & 0 & & & \end{array}$$

$$\begin{array}{ccccccccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ & 1 & 1 & 2 & 3 & 5 & 8 & 13 & \\ & & -1 & 1 & -1 & 1 & -1 & & \\ & & & 0 & 0 & 0 & & & \\ & & & & 0 & & & & \end{array}$$

$$\begin{array}{cccccccccc}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
 & 1 & 1 & 2 & 3 & 5 & 8 & 13 & 21 & \\
 & & -1 & 1 & -1 & 1 & -1 & 1 & & \\
 & & & 0 & 0 & 0 & 0 & & & \\
 & & & & 0 & 0 & & & &
\end{array}$$

and it can be seen that a row of 0s emerges (and furthermore that an attempt to extend the table will result in division by zero). This verifies that the FIBONACCI NUMBERS satisfy a linear recurrence, which is in fact given by the well-known formula

$$F_n = F_{n-1} + F_{n-2}.$$

However, construction of a quotient-difference table for the CATALAN NUMBERS, MOTZKIN NUMBERS, etc., does not lead to a row of zeros, suggesting that these numbers cannot be generated using a linear recurrence.

see also DIFFERENCE TABLE, FINITE DIFFERENCE

References

Conway, J. H. and Guy, R. K. In *The Book of Numbers.* New York: Springer-Verlag, pp. 85–89, 1996.

Quotient Group

The quotient group of G with respect to a SUBGROUP H is denoted G/H and is read "G modulo H." The slash NOTATION conflicts with that for a FIELD EXTENSION, but the meaning can be determined based on context.

see also ABHYANKAR'S CONJECTURE, FIELD EXTENSION, OUTER AUTOMORPHISM GROUP, SUBGROUP

Quotient Ring

The quotient ring of R with respect to a RING modulo some INTEGER n is denoted R/nR and is read "the ring R modulo n." If n is a PRIME p, then $\mathbb{Z}/p\mathbb{Z}$ is the FINITE FIELD $\mathbb{F}_p$. For COMPOSITE

$$n = \prod_{i=1}^{k} p_i$$

with distinct p_i, $\mathbb{Z}/p\mathbb{Z}$ is ISOMORPHIC to the DIRECT SUM

$$\mathbb{Z}/p\mathbb{Z} = \mathbb{F}_{p_1} \otimes \mathbb{F}_{p_2} \otimes \cdots \otimes \mathbb{F}_{p_k}.$$

see also FINITE FIELD, RING

Quotient Rule

The DERIVATIVE rule

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{g(x)f'(x) - f(x)g'(x)}{[g(x)]^2}.$$

see also CHAIN RULE, DERIVATIVE, POWER RULE, PRODUCT RULE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 11, 1972.

Quotient Space

The quotient space $X/\sim$ of a TOPOLOGICAL SPACE X and an EQUIVALENCE RELATION $\sim$ on X is the set of EQUIVALENCE CLASSES of points in X (under the EQUIVALENCE RELATION $\sim$) together with the topology given by a SUBSET U of $X/\sim$. U of $X/\sim$ is open IFF $\cup_{a \in U} a$ is open in X.

This can be stated in terms of MAPS as follows: if $q : X \to X/\sim$ denotes the MAP that sends each point to its EQUIVALENCE CLASS in $X/\sim$, the topology on $X/\sim$ can be specified by prescribing that a subset of $X/\sim$ is open IFF q^{-1}[the set] is open.

In general, quotient spaces are not well behaved, and little is known about them. However, it is known that any compact metrizable space is a quotient of the CANTOR SET, any compact connected n-dimensional MANIFOLD for $n > 0$ is a quotient of any other, and a function out of a quotient space $f : X/\sim \to Y$ is continuous IFF the function $f \circ q : X \to Y$ is continuous.

Let $\mathbb{D}^n$ be the closed n-D DISK and $\mathbb{S}^{n-1}$ its boundary, the $(n-1)$-D sphere. Then $\mathbb{D}^n/\mathbb{S}^{n-1}$ (which is homeomorphic to $\mathbb{S}^n$), provides an example of a quotient space. Here, $\mathbb{D}^n/\mathbb{S}^{n-1}$ is interpreted as the space obtained when the boundary of the n-DISK is collapsed to a point, and is formally the "quotient space by the equivalence relation generated by the relations that all points in $\mathbb{S}^{n-1}$ are equivalent."

see also EQUIVALENCE RELATION, TOPOLOGICAL SPACE

References

Munkres, J. R. *Topology: A First Course.* Englewood Cliffs, NJ: Prentice-Hall, 1975.

R

ℝ

The FIELD of REAL NUMBERS.

see also $\mathbb{C}$, $\mathbb{C}^*$, $\mathbb{I}$, $\mathbb{N}$, $\mathbb{Q}$, $\mathbb{R}^-$, $\mathbb{R}^+$, $\mathbb{Z}$

$\mathbb{R}^-$

The REAL NEGATIVE numbers.

see also $\mathbb{R}$, $\mathbb{R}^+$

$\mathbb{R}^+$

The REAL POSITIVE numbers.

see also $\mathbb{R}$, $\mathbb{R}^-$

$r_k(n)$

The number of representations of n by k squares is denoted $r_k(n)$. The *Mathematica*® (Wolfram Research, Champaign, IL) function `NumberTheory‘NumberTheoryFunctions‘SumOfSquaresR[k,n]` gives $r_k(n)$.

$r_2(n)$ is often simply written $r(n)$. Jacobi solved the problem for $k = 2$, 4, 6, and 8. The first cases $k = 2$, 4, and 6 were found by equating COEFFICIENTS of the THETA FUNCTION $\vartheta_3(x)$, $\vartheta_3{}^2(x)$, and $\vartheta_3{}^4(x)$. The solutions for $k = 10$ and 12 were found by Liouville and Eisenstein, and Glaisher (1907) gives a table of $r_k(n)$ for $k = 2s = 18$. $r_3(n)$ was found as a finite sum involving quadratic reciprocity symbols by Dirichlet. $r_5(n)$ and $r_7(n)$ were found by Eisenstein, Smith, and Minkowski.

$r(n) = r_2(n)$ is 0 whenever n has a PRIME divisor of the form $4k+3$ to an ODD POWER; it doubles upon reaching a new PRIME of the form $4k+1$. It is given explicitly by

$$r(n) = 4 \sum_{d=1,3,\ldots|n} (-1)^{(d-1)/2} = 4[d_1(n) - d_3(n)], \quad (1)$$

where $d_k(n)$ is the number of DIVISORS of n of the form $4m+k$. The first few values are 4, 4, 0, 4, 8, 0, 0, 4, 4, 8, 0, 0, 8, 0, 0, 4, 8, 4, 0, 8, 0, 0, 0, 0, 12, 8, 0, 0, ... (Sloane's A004018). The first few values of the summatory function

$$R(n) \equiv \sum_{k=1}^{n} r(n), \quad (2)$$

where are 0, 4, 8, 8, 12, 20, 20, 20, 24, 28, 36, ... (Sloane's A014198). Shanks (1993) defines instead $R'(n) = 1 + R(n)$, with $R'(0) = 1$. A LAMBERT SERIES for $r(n)$ is

$$\sum_{n=1}^{\infty} \frac{4(-1)^{n+1}x^n}{1-x^n} = \sum_{n=1}^{\infty} r(n)x^n \quad (3)$$

(Hardy and Wright 1979).

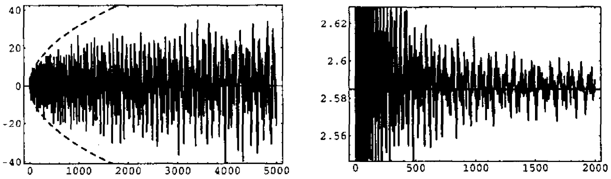

Asymptotic results include

$$\sum_{k=1}^{n} r_2(k) = \pi n + \mathcal{O}(\sqrt{n}) \quad (4)$$

$$\sum_{k=1}^{n} \frac{r_2(k)}{k} = K + \pi \ln n + \mathcal{O}(n^{-1/2}), \quad (5)$$

where K is a constant known as the SIERPIŃSKI CONSTANT. The left plot above

$$\left[\sum_{k=1}^{n} r_2(k)\right] - \pi n, \quad (6)$$

with $\pm\sqrt{n}$ illustrated by the dashed curve, and the right plot shows

$$\left[\sum_{k=1}^{n} \frac{r_2(k)}{k}\right] - \pi \ln n, \quad (7)$$

with the value of K indicated as the solid horizontal line.

The number of solutions of

$$x^2 + y^2 + z^2 = n \quad (8)$$

for a given n without restriction on the signs or relative sizes of x, y, and z is given by $r_3(n)$. If $n > 4$ is SQUAREFREE, then Gauss proved that

$$r_3(n) = \begin{cases} 24h(-n) & \text{for } n \equiv 3 \pmod 8 \\ 12h(-4n) & \text{for } n \equiv 1,2,5,6 \pmod 8 \\ 0 & \text{for } n \equiv 7 \pmod 8 \end{cases} \quad (9)$$

(Arno 1992), where $h(x)$ is the CLASS NUMBER of x.

Additional higher-order identities are given by

$$r_4(n) = 8 \sum_{d|n} d = 8\sigma(n) \quad (10)$$

$$= 24 \sum_{d=1,3,\ldots|n} d = 24\sigma_0(n) \quad (11)$$

$$r_{10}(n) = \tfrac{4}{5}[E_4(n) + 16E_4'(n) + 8\chi_4(n)] \quad (12)$$

$$r_{24}(n) = \rho_{24}(n) + \tfrac{128}{691}[(-1)^{n-1}259\tau(n) - 512\tau(\tfrac{1}{2}n)], \quad (13)$$

where

$$E_4(n) = \sum_{d=1,3,\ldots|n} (-1)^{(d-1)/2} d^4 \quad (14)$$

$$E_4'(n) = \sum_{d'=1,3,\ldots|n} (-1)^{(d'-1)/2} d^4 \quad (15)$$

$$\chi_4(n) = \tfrac{1}{4} \sum_{a^2+b^2=n} (a+bi)^4, \quad (16)$$

$d' \equiv n/d$, $d_k(n)$ is the number of divisors of n of the form $4m+k$, $\rho_{24}(n)$ is a SINGULAR SERIES, $\sigma(n)$ is the DIVISOR FUNCTION, $\sigma_0(n)$ is the DIVISOR FUNCTION of order 0 (i.e., the number of DIVISORS), and τ is the TAU FUNCTION.

Similar expressions exist for larger EVEN k, but they quickly become extremely complicated and can be written simply only in terms of expansions of modular functions.

see also CLASS NUMBER, LANDAU-RAMANUJAN CONSTANT, PRIME FACTORS, SIERPIŃSKI CONSTANT, TAU FUNCTION

References

Arno, S. "The Imaginary Quadratic Fields of Class Number 4." *Acta Arith.* **60**, 321–334, 1992.

Boulyguine. *Comptes Rendus Paris* **161**, 28–30, 1915.

New York: Chelsea, p. 317, 1952.

Glaisher, J. W. L. "On the Numbers of a Representation of a Number as a Sum of $2r$ Squares, where $2r$ Does Not Exceed 18." *Proc. London Math. Soc.* **5**, 479–490, 1907.

Grosswald, E. *Representations of Integers as Sums of Squares.* New York: Springer-Verlag, 1985.

Hardy, G. H. "The Representation of Numbers as Sums of Squares." Ch. 9 in *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, 1959.

Hardy, G. H. and Wright, E. M. "The Function $r(n)$," "Proof of the Formula for $r(n)$," "The Generating Function of $r(n)$," and "The Order of $r(n)$." §16.9, 16.10, 17.9 , and 18.7 in *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, pp. 241–243, 256–258, and 270–271, 1979.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 162–153, 1993.

Sloane, N. J. A. Sequence A004018/M3218 in "An On-Line Version of the Encyclopedia of Integer Sequences."

R-Estimate

A ROBUST ESTIMATION based on RANK tests. Examples include the statistic of the KOLMOGOROV-SMIRNOV TEST, SPEARMAN RANK CORRELATION, and WILCOXON SIGNED RANK TEST.

see also L-ESTIMATE, M-ESTIMATE, ROBUST ESTIMATION

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Robust Estimation." §15.7 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 694–700, 1992.

Raabe's Test

Given a SERIES of POSITIVE terms u_i and a SEQUENCE of POSITIVE constants $\{a_i\}$, use KUMMER'S TEST

$$\rho' \equiv \lim_{n\to\infty}\left(a_n\frac{u_n}{u_{n+1}} - a_{n+1}\right).$$

with $a_n = n$, giving

$$\begin{aligned}\rho' &\equiv \lim_{n\to\infty}\left(n\frac{u_n}{u_{n+1}} - (n+1)\right)\\ &= \lim_{n\to\infty}\left[n\left(\frac{u_n}{u_{n+1}} - 1\right) - 1\right].\end{aligned}$$

Defining

$$\rho \equiv \rho' + 1 = \lim_{n\to\infty}\left[n\left(\frac{u_n}{u_{n+1}} - 1\right)\right],$$

then gives Raabe's test:

1. If $\rho > 1$, the SERIES CONVERGES.
2. If $\rho < 1$, the SERIES DIVERGES.
3. If $\rho = 1$, the SERIES may CONVERGE or DIVERGE.

see also CONVERGENT SERIES, CONVERGENCE TESTS, DIVERGENT SERIES, KUMMER'S TEST

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 286–287, 1985.

Bromwich, T. J. I'a and MacRobert, T. M. *An Introduction to the Theory of Infinite Series, 3rd ed.* New York: Chelsea, p. 39, 1991.

Rabbit Constant

The limiting RABBIT SEQUENCE written as a BINARY FRACTION $0.1011010110110\ldots_2$ (Sloane's A005614), where b_2 denotes a BINARY number (a number in base-2). The DECIMAL value is

$$R = 0.709803442861291314\,6\ldots$$

(Sloane's A014565).

Amazingly, the rabbit constant is also given by the CONTINUED FRACTION $[0, 2^{F_0}, 2^{F_1}, 2^{F_2}, 2^{F_3}, \ldots]$, where F_n are FIBONACCI NUMBERS with F_0 taken as 0 (Gardner 1989, Schroeder 1991). Another amazing connection was discovered by S. Plouffe. Define the BEATTY SEQUENCE $\{a_i\}$ by

$$a_i \equiv \lfloor i\phi \rfloor,$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION and ϕ is the GOLDEN RATIO. The first few terms are 1, 3, 4, 6, 8, 9, 11, ... (Sloane's A000201). Then

$$R = \sum_{i=1}^{\infty} 2^{-a_i}.$$

see also RABBIT SEQUENCE, THUE CONSTANT, THUE-MORSE CONSTANT

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/cntfrc/cntfrc.html`.

Gardner, M. *Penrose Tiles and Trapdoor Ciphers... and the Return of Dr. Matrix, reissue ed.* New York: W. H. Freeman, pp. 21–22, 1989.
Plouffe, S. "The Rabbit Constant to 330 Digits." `http://lacim.uqam.ca/piDATA/rabbit.txt`.
Schroeder, M. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise.* New York: W. H. Freeman, p. 55, 1991.
Sloane, N. J. A. Sequences A005614, A014565, and A000201/M2322 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Rabbit-Duck Illusion

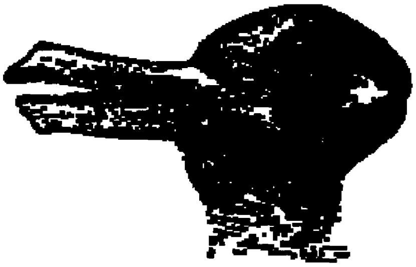

A perception ILLUSION in which the brain switches between seeing a rabbit and a duck.

see also YOUNG GIRL-OLD WOMAN ILLUSION

Rabbit Sequence

A SEQUENCE which arises in the hypothetical reproduction of a population of rabbits. Let the SUBSTITUTION MAP $0 \to 1$ correspond to young rabbits growing old, and $1 \to 10$ correspond to old rabbits producing young rabbits. Starting with 0 and iterating using STRING REWRITING gives the terms 1, 10, 101, 10110, 10110101, 1011010110110, ... The limiting sequence written as a BINARY FRACTION $0.1011010110110\ldots_2$ (Sloane's A005614), where b_2 denotes a BINARY number (a number in base-2) is called the RABBIT CONSTANT.

see also RABBIT CONSTANT, THUE-MORSE SEQUENCE

References

Schroeder, M. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise.* New York: W. H. Freeman, p. 55, 1991.
Sloane, N. J. A. Sequence A005614 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Rabdology

see NAPIER'S BONES

Rabin-Miller Strong Pseudoprime Test

A PRIMALITY TEST which provides an efficient probabilistic ALGORITHM for determining if a given number is PRIME. It is based on the properties of STRONG PSEUDOPRIMES. Given an ODD INTEGER n, let $n = 2^r s + 1$ with s ODD. Then choose a random integer a with $1 \leq a \leq n-1$. If $a^s \equiv 1 \pmod{n}$ or $a^{2^j s} \equiv -1 \pmod{n}$ for some $0 \leq j \leq r-1$, then n passes the test. A PRIME will pass the test for all a.

The test is very fast and requires no more than $(1 + o(1)) \lg n$ multiplications $(\bmod\, n)$, where LG is the LOGARITHM base 2. Unfortunately, a number which passes the test is not necessarily PRIME. Monier (1980) and Rabin (1980) have shown that a COMPOSITE NUMBER passes the test for at most 1/4 of the possible bases a.

The Rabin-Miller test (combined with a LUCAS PSEUDOPRIME test) is the PRIMALITY TEST used by *Mathematica*® versions 2.2 and later (Wolfram Research, Champaign, IL). As of 1991, the combined test had been proven correct for all $n < 2.5 \times 10^{10}$, but not beyond. The test potentially could therefore incorrectly identify a large COMPOSITE NUMBER as PRIME (but not vice versa). STRONG PSEUDOPRIME tests have been subsequently proved valid for every number up to 3.4×10^{14}.

see also LUCAS-LEHMER TEST, MILLER'S PRIMALITY TEST, PSEUDOPRIME, STRONG PSEUDOPRIME

References

Arnault, F. "Rabin-Miller Primality Test: Composite Numbers Which Pass It." *Math. Comput.* **64**, 355–361, 1995.
Miller, G. "Riemann's Hypothesis and Tests for Primality." *J. Comp. Syst. Sci.* **13**, 300–317, 1976.
Monier, L. "Evaluation and Comparison of Two Efficient Probabilistic Primality Testing Algorithms." *Theor. Comput. Sci.* **12**, 97–108, 1980.
Rabin, M. O. "Probabilistic Algorithm for Testing Primality." *J. Number Th.* **12**, 128–138, 1980.
Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 15–17, 1991.

Rabinovich-Fabrikant Equation

The 3-D MAP

$$\begin{aligned} \dot{x} &= y(z - 1 + x^2) + \gamma x \\ \dot{y} &= x(3z + 1 - x^2) + \gamma y \\ \dot{z} &= -2z(\alpha + xy) \end{aligned}$$

(Rabinovich and Fabrikant 1979). The parameters are most commonly taken as $\gamma = 0.87$ and $\alpha = 1.1$. It has a CORRELATION EXPONENT of 2.19 ± 0.01.

References

Grassberger, P. and Procaccia, I. "Measuring the Strangeness of Strange Attractors." *Physica D* **9**, 189–208, 1983.
Rabinovich, M. I. and Fabrikant, A. L. *Sov. Phys. JETP* **50**, 311–317, 1979.

Racah V-Coefficient

The Racah V-COEFFICIENTS are written

$$V(j_1 j_2 j; m_1 m_2 m) \tag{1}$$

and are sometimes expressed using the related CLEBSCH-GORDON COEFFICIENTS

$$C^j_{m_1 m_2} = (j_1 j_2 m_1 m_2 | j_1 j_2 j m), \tag{2}$$

or WIGNER $3j$-SYMBOLS. Connections among the three are

$$(j_1 j_2 m_1 m_2 | j_1 j_2 m) = (-1)^{-j_1+j_2-m} \sqrt{2j+1} \begin{pmatrix} j_1 & j_2 & j \\ m_1 & m_2 & -m \end{pmatrix} \quad (3)$$

$$(j_1 j_2 m_1 m_2 | j_1 j_2 j m) = (-1)^{j+m} \sqrt{2j+1}\, V(j_1 j_2 j; m_1 m_2 \; - m) \quad (4)$$

$$V(j_1 j_2 j; m_1 m_2 m) = (-1)^{-j_1+j_2+j} \begin{pmatrix} j_1 & j_2 & j_1 \\ m_2 & m_1 & m_2 \end{pmatrix}. \quad (5)$$

see also CLEBSCH-GORDON COEFFICIENT, RACAH W-COEFFICIENT, WIGNER $3j$-SYMBOL, WIGNER $6j$-SYMBOL, WIGNER $9j$-SYMBOL

References
Sobel'man, I. I. "Angular Momenta." Ch. 4 in *Atomic Spectra and Radiative Transitions, 2nd ed.* Berlin: Springer-Verlag, 1992.

Racah W-Coefficient

Related to the CLEBSCH-GORDON COEFFICIENTS by

$$(J_1 J_2 [J'] J_3 | J_1, J_2 J_3 [J'']) = \sqrt{(2J'+1)(2J''+1)}\, W(J_1 J_2 J J_3; J' J'')$$

and

$$(J_1 J_2 [J'] J_3 | J_1 J_3 [J''] J_2) = \sqrt{(2J'+1)(2J''+1)}\, W(J_1' J_3 J_2 J''; J J_1).$$

see also CLEBSCH-GORDON COEFFICIENT, RACAH V-COEFFICIENT, WIGNER $3j$-SYMBOL, WIGNER $6j$-SYMBOL, WIGNER $9j$-SYMBOL

References
Messiah, A. "Racah Coefficients and '6j' Symbols." Appendix C.II in *Quantum Mechanics, Vol. 2.* Amsterdam, Netherlands: North-Holland, pp. 1061–1066, 1962.
Sobel'man, I. I. "Angular Momenta." Ch. 4 in *Atomic Spectra and Radiative Transitions, 2nd ed.* Berlin: Springer-Verlag, 1992.

Radau Quadrature

A GAUSSIAN QUADRATURE-like formula for numerical estimation of integrals. It requires $m+1$ points and fits all POLYNOMIALS to degree $2m$, so it effectively fits exactly all POLYNOMIALS of degree $2m-1$. It uses a WEIGHTING FUNCTION $W(x) = 1$ in which the endpoint -1 in the interval $[-1,1]$ is included in a total of n ABSCISSAS, giving $r = n-1$ free abscissas. The general formula is

$$\int_{-1}^{1} f(x)\,dx = w_1 f(-1) + \sum_{i=2}^{n} w_i f(x_i). \quad (1)$$

The free abscissas x_i for $i = 2, \ldots, n$ are the roots of the POLYNOMIAL

$$\frac{P_{n-1}(x) + P_n(x)}{1+x}, \quad (2)$$

where $P(x)$ is a LEGENDRE POLYNOMIAL. The weights of the free abscissas are

$$w_i = \frac{1-x_i}{n^2[P_{n-1}(x_i)]^2} = \frac{1}{(1-x_i)[P'_{n-1}(x_i)]^2}, \quad (3)$$

and of the endpoint

$$w_1 = \frac{2}{n^2}. \quad (4)$$

The error term is given by

$$E = \frac{2^{2n-1} n[(n-1)!]^4}{[(2n-1)!]^3} f^{(2n-1)}(\xi), \quad (5)$$

for $\xi \in (-1,1)$.

n	x_i	w_i
2	-1	0.5
	0.333333	1.5
3	-1	0.222222
	-0.289898	1.02497
	0.689898	0.752806
4	-1	0.125
	-0.575319	0.657689
	0.181066	0.776387
	0.822824	0.440924
5	-1	0.08
	-0.72048	0.446208
	-0.167181	0.623653
	0.446314	0.562712
	0.885792	0.287427

The ABSCISSAS and weights can be computed analytically for small n.

n	x_i	w_i
2	-1	$\frac{1}{2}$
	$\frac{1}{3}$	$\frac{3}{2}$
3	-1	$\frac{2}{9}$
	$\frac{1}{5}(1-\sqrt{6})$	$\frac{1}{18}(16+\sqrt{6})$
	$\frac{1}{5}(1+\sqrt{6})$	$\frac{1}{18}(16-\sqrt{6})$

see also CHEBYSHEV QUADRATURE, LOBATTO QUADRATURE

References
Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 888, 1972.
Chandrasekhar, S. *Radiative Transfer.* New York: Dover, p. 61, 1960.
Hildebrand, F. B. *Introduction to Numerical Analysis.* New York: McGraw-Hill, pp. 338–343, 1956.

Rademacher Function

see SQUARE WAVE

Radial Curve

Let C be a curve and let O be a fixed point. Let P be on C and let Q be the CURVATURE CENTER at P. Let P_1 be the point with P_1O a line segment PARALLEL and of equal length to PQ. Then the curve traced by P_1 is the radial curve of C. It was studied by Robert Tucker in 1864. The parametric equations of a curve (f,g) with RADIAL POINT (x_0,y_0) are

$$\begin{aligned} x &= x_0 - \frac{g'(f'^2+g'^2)}{f'g''-f''g'} \\ y &= y_0 + \frac{f'(f'^2+g'^2)}{f'g''-f''g'}. \end{aligned}$$

Curve	Radial Curve
astroid	quadrifolium
catenary	kampyle of Eudoxus
cycloid	circle
deltoid	trifolium
logarithmic spiral	logarithmic spiral
tractrix	kappa curve

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 40 and 202, 1972.

Yates, R. C. "Radial Curves." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 172–174, 1952.

Radial Point

The point with respect to which a RADIAL CURVE is computed.

see also RADIANT POINT

Radian

A unit of angular measure in which the ANGLE of an entire CIRCLE is 2π radians. There are therefore 360° per 2π radians, equal to $180/\pi$ or 57.29577951°/radian. A RIGHT ANGLE is $\pi/2$ radians.

see also ANGLE, ARC MINUTE, ARC SECOND, DEGREE, GRADIAN, STERADIAN

Radiant Point

The point of illumination for a CAUSTIC.

see also CAUSTIC, RADIAL POINT

Radical

The symbol $\sqrt[n]{x}$ used to indicate a root is called a radical. The expression $\sqrt[n]{x}$ is therefore read "x radical n," or "the nth ROOT of x." $n=2$ is written $\sqrt{x}$ and is called the SQUARE ROOT of x. $n=3$ corresponds to the CUBE ROOT. The quantity under the root is called the RADICAND.

Some interesting radical identities are due to Ramanujan, and include the equivalent forms

$$(2^{1/3}+1)(2^{1/3}-1)^{1/3}=3^{1/3}$$

and

$$(2^{1/3}-1)^{1/3}=(\tfrac{1}{9})^{1/3}-(\tfrac{2}{9})^{1/3}+(\tfrac{4}{9})^{1/3}.$$

Another such identity is

$$(5^{1/3}-4^{1/3})^{1/2}=\tfrac{1}{3}(2^{1/3}+20^{1/3}-25^{1/3}).$$

see also CUBE ROOT, NESTED RADICAL, POWER, RADICAL INTEGER, RADICAND, ROOT (RADICAL), SQUARE ROOT, VINCULUM

Radical Axis

see RADICAL LINE

Radical Center

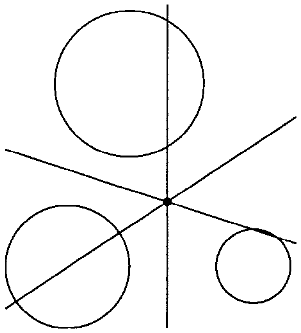

The RADICAL LINES of three CIRCLES are CONCURRENT in a point known as the radical center (also called the POWER CENTER). This theorem was originally demonstrated by Monge (Dörrie 1965, p. 153).

see also APOLLONIUS' PROBLEM, CONCURRENT, MONGE'S PROBLEM, RADICAL LINE

References

Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, 1965.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, p. 32, 1929.

Radical Integer

A radical integer is a number obtained by closing the INTEGERS under ADDITION, DIVISION, MULTIPLICATION, SUBTRACTION, and ROOT extraction. An example of such a number is $\sqrt[3]{7}+\sqrt{-2}-\sqrt{3+\sqrt[4]{1}+\sqrt{2}}$. The radical integers are a subring of the ALGEBRAIC INTEGERS. If there are ALGEBRAIC INTEGERS which are not radical integers, they must at least be cubic.

see also ALGEBRAIC INTEGER, ALGEBRAIC NUMBER, EUCLIDEAN NUMBER

Radical Line

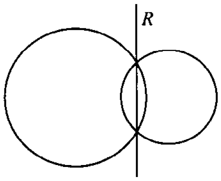

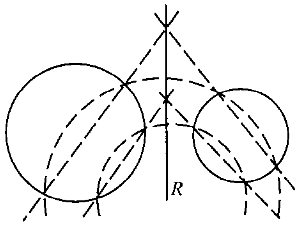

The LOCUS of points of equal POWER with respect to two nonconcentric CIRCLES which is PERPENDICULAR to the line of centers (the CHORDAL THEOREM; Dörrie 1965). Let the circles have RADII r_1 and r_2 and their centers be separated by a distance d. If the CIRCLES intersect in two points, then the radical line is the line passing through the points of intersection. If not, then draw any two CIRCLES which cut each original CIRCLE twice. Draw lines through each pair of points of intersection of each CIRCLE. The line connecting their two points of intersection is then the radical line.

The radical line is located at distances

$$d_1 = \frac{d^2 + {r_1}^2 - {r_2}^2}{2d} \tag{1}$$

$$d_2 = -\frac{d^2 + {r_2}^2 - {r_1}^2}{2d} \tag{2}$$

along the line of centers from C_1 and C_2, respectively, where

$$d \equiv d_1 - d_2. \tag{3}$$

The radical line of any two POLAR CIRCLES is the ALTITUDE from the third vertex.

see also CHORDAL THEOREM, COAXAL CIRCLES, INVERSE POINTS, INVERSION, POWER (CIRCLE), RADICAL CENTER

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 31–34, 1967.

Dixon, R. *Mathographics.* New York: Dover, p. 68, 1991.

Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, p. 153, 1965.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 28–34 and 176–177, 1929.

Radicand

The quantity under a RADICAL sign.

see also RADICAL, VINCULUM

Radius

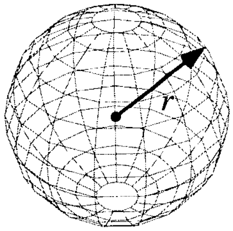

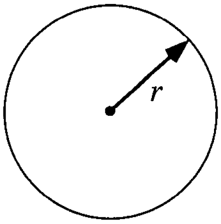

The distance from the center of a CIRCLE to its PERIMETER, or from the center of a SPHERE to its surface. The radius is equal to half the DIAMETER.

see also BERTRAND'S PROBLEM, CIRCLE, CIRCUMFERENCE, DIAMETER, EXTENT, INVERSION RADIUS, KINNEY'S SET, PI, RADIUS OF CONVERGENCE, RADIUS OF CURVATURE, RADIUS (GRAPH), RADIUS OF GYRATION, RADIUS OF TORSION, RADIUS VECTOR, SPHERE

Radius of Convergence

The RADIUS (or 1-D distance in the 1-D case) over which series expansion CONVERGES.

Radius of Curvature

The radius of curvature is given by

$$R \equiv \frac{1}{\kappa}, \tag{1}$$

where κ is the CURVATURE. At a given point on a curve, R is the radius of the OSCULATING CIRCLE. The symbol ρ is sometimes used instead of R to denote the radius of curvature.

Let x and y be given parametrically by

$$x = x(t) \tag{2}$$

$$y = y(t), \tag{3}$$

then

$$R = \frac{(x'^2 + y'^2)^{3/2}}{x'y'' - y'x''}, \tag{4}$$

where $x' = dx/dt$ and $y' = dy/dt$. Similarly, if the curve is written in the form $y = f(x)$, then the radius of curvature is given by

$$R = \frac{\left[1 + \left(\frac{dy}{dx}\right)^2\right]^{3/2}}{\frac{d^2y}{dx^2}}. \tag{5}$$

see also BEND (CURVATURE), CURVATURE, OSCULATING CIRCLE, TORSION (DIFFERENTIAL GEOMETRY)

References

Kreyszig, E. *Differential Geometry.* New York: Dover, p. 34, 1991.

Radius (Graph)

The minimum ECCENTRICITY of any VERTEX of a GRAPH.

Radius of Gyration

A function quantifying the spatial extent of the structure of a curve. It is defined by

$$R_g = \frac{\sqrt{\int_0^\infty r^2 p(r)\,dr}}{2\int_0^\infty p(r)\,dr},$$

where $p(r)$ is the LENGTH DISTRIBUTION FUNCTION. Small compact patterns have small R_g.

References

Pickover, C. A. *Keys to Infinity.* New York: W. H. Freeman, pp. 204-206, 1995.

Radius of Torsion

$$\sigma \equiv \frac{1}{\tau},$$

where τ is the TORSION. The symbol ϕ is also sometimes used instead of σ.

see also TORSION (DIFFERENTIAL GEOMETRY)

References

Kreyszig, E. *Differential Geometry.* New York: Dover, p. 39, 1991.

Radius Vector

The VECTOR $\mathbf{r}$ from the ORIGIN to the current position. It is also called the POSITION VECTOR. The derivative of $\mathbf{r}$ satisfies

$$\mathbf{r}\cdot\frac{d\mathbf{r}}{dt} = \frac{1}{3}\frac{d}{dt}(\mathbf{r}\cdot\mathbf{r}) = \frac{1}{2}\frac{d}{dt}(r^2) = r\frac{dr}{dt} = rv,$$

where v is the magnitude of the VELOCITY (i.e., the SPEED).

Radix

The BASE of a number system, i.e., 2 for BINARY, 8 for OCTAL, 10 for DECIMAL, and 16 for HEXADECIMAL. The radix is sometimes called the BASE or SCALE.

see also BASE (NUMBER)

Rado's Sigma Function

see BUSY BEAVER

Radon-Nikodym Theorem

A THEOREM which gives NECESSARY and SUFFICIENT conditions for a countably additive function of sets can be expressed as an integral over the set.

References

Doob, J. L. "The Development of Rigor in Mathematical Probability (1900–1950)." *Amer. Math. Monthly* **103**, 586–595, 1996.

Radon Transform

An INTEGRAL TRANSFORM whose inverse is used to reconstruct images from medical CT scans. A technique for using Radon transforms to reconstruct a map of a planet's polar regions using a spacecraft in a polar orbit has also been devised (Roulston and Muhleman 1997).

The Radon transform can be defined by

$$R(p,\tau)[f(x,y)] = \int_{-\infty}^{\infty} f(x,\tau+px)\,dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta[y-(\tau+px)]\,dy\,dx \equiv U(p,\tau), \quad (1)$$

where p is the SLOPE of a line and τ is its intercept. The inverse Radon transform is

$$f(x,y) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{d}{dy}H[U(p,y-px)]\,dp, \quad (2)$$

where H is a HILBERT TRANSFORM. The transform can also be defined by

$$R'(r,\alpha)[f(x,y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(r-x\cos\alpha-y\sin\alpha)\,dx\,dy, \quad (3)$$

where r is the PERPENDICULAR distance from a line to the origin and α is the ANGLE formed by the distance VECTOR.

Using the identity

$$\mathcal{F}[R[f(\omega,\alpha)]] = \mathcal{F}^2[f(u,v)], \quad (4)$$

where $\mathcal{F}$ is the FOURIER TRANSFORM, gives the inversion formula

$$f(x,y) = c\int_0^{\pi}\int_{-\infty}^{\infty}\mathcal{F}[R[f(\omega,\alpha)]]|\omega|e^{i\omega(x\cos\alpha+y\sin\alpha)}\,d\omega\,d\alpha. \quad (5)$$

The FOURIER TRANSFORM can be eliminated by writing

$$f(x,y) = \int_0^{\pi}\int_{-\infty}^{\infty} R[f(r,\alpha)]W(r,\alpha,x,y)\,dr\,d\alpha, \quad (6)$$

where W is a WEIGHTING FUNCTION such as

$$W(r,\alpha,x,y) = h(x\cos\alpha+y\sin\alpha-r) = \mathcal{F}^{-1}[|\omega|]. \quad (7)$$

Nievergelt (1986) uses the inverse formula

$$f(x,y) = \frac{1}{\pi}\lim_{c\to 0}\int_0^{\pi}\int_{-\infty}^{\infty} R[f(r+x\cos\alpha+y\sin\alpha,\alpha)]G_c(r)\,dr\,d\alpha, \quad (8)$$

where

$$G_c(r) = \begin{cases} \frac{1}{\pi c^2} & \text{for } |r| \leq c \\ \frac{1}{\pi c^2}\left(1-\frac{1}{\sqrt{1-c^2/r^2}}\right) & \text{for } |r| > c. \end{cases} \quad (9)$$

LUDWIG'S INVERSION FORMULA expresses a function in terms of its Radon transform. $R'(r,\alpha)$ and $R(p,\tau)$ are related by

$$p = \cot\alpha \qquad \tau = r\csc\alpha \quad (10)$$

$$r = \frac{\tau}{1+p^2} \qquad \alpha = \cot^{-1}p. \quad (11)$$

The Radon transform satisfies superposition

$$R(p,\tau)[f_1(x,y)+f_2(x,y)] = U_1(p,\tau)+U_2(p,\tau), \quad (12)$$

linearity

$$R(p,\tau)[af(x,y)] = aU(p,\tau), \tag{13}$$

scaling

$$R(p,\tau)\left[f\left(\frac{x}{a},\frac{y}{b}\right)\right] = |a|U\left(p\frac{a}{b},\frac{\tau}{b}\right), \tag{14}$$

ROTATION, with R_ϕ ROTATION by ANGLE ϕ

$$R(p,\tau)[R_\phi f(x,y)] = \frac{1}{|\cos\phi + p\sin\phi|} U\left(\frac{p-\tan\phi}{1+p\tan\phi}, \frac{\tau}{\cos\phi + p\sin\phi}\right), \tag{15}$$

and skewing

$$R(p,\tau)[f(ax+by,cx+dy)] = \frac{1}{|a+bp|}U\left[\frac{c+dp}{a+bp}, \tau\frac{d-b(c+bd)}{a+bp}\right] \tag{16}$$

(Durrani and Bisset 1984).

The line integral along p,τ is

$$I = \sqrt{1+p^2}\,U(p,\tau). \tag{17}$$

The analog of the 1-D CONVOLUTION THEOREM is

$$R(p,\tau)[f(x,y) * g(y)] = U(p,\tau) * g(\tau), \tag{18}$$

the analog of PLANCHEREL'S THEOREM is

$$\int_{-\infty}^{\infty} U(p,\tau)\,d\tau = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy, \tag{19}$$

and the analog of PARSEVAL'S THEOREM is

$$\int_{-\infty}^{\infty} R(p,\tau)[f(x,y)]^2\,d\tau = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^2(x,y)\,dx\,dy. \tag{20}$$

If f is a continuous function on $\mathbb{C}$, integrable with respect to a plane LEBESGUE MEASURE, and

$$\int_l f\,ds = 0 \tag{21}$$

for every (doubly) infinite line l where s is the length measure, then f must be identically zero. However, if the global integrability condition is removed, this result fails (Zalcman 1982, Goldstein 1993).

see also TOMOGRAPHY

References

Anger, B. and Portenier, C. *Radon Integrals.* Boston, MA: Birkhäuser, 1992.

Armitage, D. H. and Goldstein, M. "Nonuniqueness for the Radon Transform." *Proc. Amer. Math. Soc.* **117**, 175–178, 1993.

Deans, S. R. *The Radon Transform and Some of Its Applications.* New York: Wiley, 1983.

Durrani, T. S. and Bisset, D. "The Radon Transform and its Properties." *Geophys.* **49**, 1180–1187, 1984.

Esser, P. D. (Ed.). *Emission Computed Tomography: Current Trends.* New York: Society of Nuclear Medicine, 1983.

Gindikin, S. (Ed.). *Applied Problems of Radon Transform.* Providence, RI: Amer. Math. Soc., 1994.

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, 1979.

Helgason, S. *The Radon Transform.* Boston, MA: Birkhäuser, 1980.

Kunyansky, L. A. "Generalized and Attenuated Radon Transforms: Restorative Approach to the Numerical Inversion." *Inverse Problems* **8**, 809–819, 1992.

Nievergelt, Y. "Elementary Inversion of Radon's Transform." *SIAM Rev.* **28**, 79–84, 1986.

Rann, A. G. and Katsevich, A. I. *The Radon Transform and Local Tomography.* Boca Raton, FL: CRC Press, 1996.

Robinson, E. A. "Spectral Approach to Geophysical Inversion Problems by Lorentz, Fourier, and Radon Transforms." *Proc. Inst. Electr. Electron. Eng.* **70**, 1039–1053, 1982.

Roulston, M. S. and Muhleman, D. O. "Synthesizing Radar Maps of Polar Regions with a Doppler-Only Method." *Appl. Opt.* **36**, 3912–3919, 1997.

Shepp, L. A. and Kruskal, J. B. "Computerized Tomography: The New Medical X-Ray Technology." *Amer. Math. Monthly* **85**, 420–439, 1978.

Strichartz, R. S. "Radon Inversion—Variation on a Theme." *Amer. Math. Monthly* **89**, 377–384 and 420–423, 1982.

Zalcman, L. "Uniqueness and Nonuniqueness for the Radon Transform." *Bull. London Math. Soc.* **14**, 241–245, 1982.

Radon Transform—Cylinder

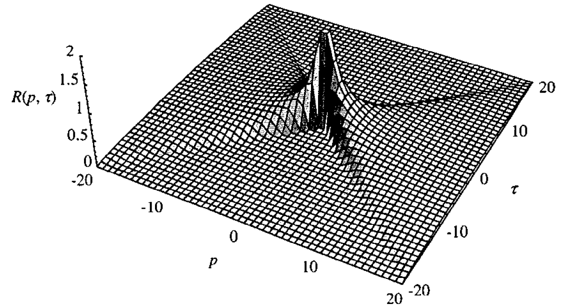

Let the 2-D cylinder function be defined by

$$f(x,y) \equiv \begin{cases} 1 & \text{for } r < R \\ 0 & \text{for } r > R. \end{cases} \tag{1}$$

Then the Radon transform is given by

$$R(p,\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta[y-(\tau+px)]\,dy\,dx, \tag{2}$$

where

$$\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ikx} \tag{3}$$

is the DELTA FUNCTION.

$$\begin{aligned} R(p,\tau) &= \frac{1}{2\pi}\int_0^{2\pi}\int_0^R\int_{-\infty}^{\infty} e^{-ik(r\sin\theta - pr\cos\theta)}r\,dr\,d\theta\,dk \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ik\tau}\int_0^{2\pi}\int_0^R e^{-ikr(\sin\theta - p\cos\theta)}r\,dr\,d\theta\,dk. \end{aligned} \tag{4}$$

Now write

$$\sin\theta - p\cos\theta = \sqrt{1+p^2}\,\cos(\theta+\phi) \equiv \sqrt{1+p^2}\,\cos\theta', \tag{5}$$

with ϕ a phase shift. Then

$$\begin{aligned}
R(p,\tau) &= \\
&\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ik\tau}\int_0^R \left(\int_0^{2\pi} e^{-ik\sqrt{1+p^2}\,r\cos\theta'}\,d\theta'\right) r\,dr\,dk \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ik\tau}\int_0^R 2\pi J_0(k\sqrt{1+p^2}\,r)r\,dr\,dk \\
&= \int_{-\infty}^{\infty} e^{ik\tau}\int_0^R J_0(k\sqrt{1+p^2}\,r)r\,dr\,dk. \qquad (6)
\end{aligned}$$

Then use

$$\int_0^z t^{n+1}J_n(t)\,dt = z^{n+1}J_{n+1}(z), \tag{7}$$

which, with $n=0$, becomes

$$\int_0^z tJ_0(t)\,dt = zJ_1(z). \tag{8}$$

Define

$$t \equiv k\sqrt{1+p^2}\,r \tag{9}$$

$$dt = k\sqrt{1+p^2}\,dr \tag{10}$$

$$r\,dr = \frac{t\,dt}{k^2(1+p^2)}, \tag{11}$$

so the inner integral is

$$\begin{aligned}
&\int_0^{R\sqrt{1+p^2}} J_0(t)\frac{t\,dt}{k^2(1+p^2)} \\
&\quad= \frac{1}{k^2(1+p^2)}kR\sqrt{1+p^2}J_1(kR\sqrt{1+p^2}\,) \\
&\quad= \frac{J_1(kR\sqrt{1+p^2}\,)}{k\sqrt{1+p^2}}R, \qquad (12)
\end{aligned}$$

and the Radon transform becomes

$$\begin{aligned}
R(p,\tau) &= \frac{R}{\sqrt{1+p^2}}\int_{-\infty}^{\infty}\frac{e^{ik\tau}J_1(kR\sqrt{1+p^2}\,)}{k}\,dk \\
&= \frac{2R}{\sqrt{1+p^2}}\int_0^{\infty}\frac{\cos(k\tau)J_1(kR\sqrt{1+p^2}\,)}{k}\,dk \\
&= \begin{cases} \frac{2}{1+p^2}\sqrt{R^2(1+p^2)-\tau^2} & \text{for } \tau^2 < R^2(1+p^2) \\ 0 & \text{for } \tau^2 \geq R^2(1+p^2). \end{cases} \qquad (13)
\end{aligned}$$

Converting to R' using $p = \cot\alpha$,

$$\begin{aligned}
R'(r,\alpha) &= \frac{2}{\sqrt{1+\cot^2\alpha}}\sqrt{(1+\cot^2\alpha)R^2 - r^2\csc^2\alpha} \\
&= \frac{2}{\csc\alpha}\sqrt{\csc^2\alpha R^2 - r^2\csc^2\alpha} \\
&= 2\sqrt{R^2-r^2}, \qquad (14)
\end{aligned}$$

which could have been derived more simply by

$$R'(r,\alpha) = \int_{-\sqrt{R^2-r^2}}^{\sqrt{R^2-r^2}} dy. \tag{15}$$

Radon Transform—Delta Function

For a DELTA FUNCTION at (x_0, y_0),

$$\begin{aligned}
R(p,\tau) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(x-x_0)\delta(y-y_0)\delta[y-(\tau+px)]\,dy\,dx \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-ik[y-(\tau+px)]}\delta(x-x_0)\delta(y-y_0) \\
&\qquad \times dk\,dy\,dx \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ik\tau}\left[\int_{-\infty}^{\infty} e^{-iky}\delta(y-y_0)\,dy \right. \\
&\qquad \left. \times \int_{-\infty}^{\infty} e^{ikpx}\delta(x-x_0)\,dx\right] dk \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ik\tau}e^{-iky_0}e^{ikpx_0}\,dk \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ik(\tau+px_0-y_0)}\,dk = \delta(\tau+px_0-y_0).
\end{aligned}$$

Radon Transform—Gaussian

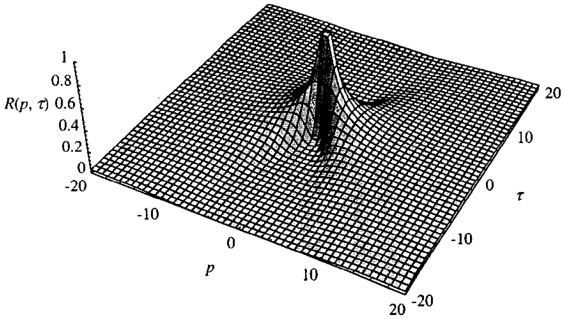

$$\begin{aligned}
R(p,\tau) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\frac{1}{\sigma\sqrt{2\pi}}e^{-(x^2+y^2)/2\sigma^2}\right] \\
&\qquad \times\delta[y-(\tau+px)]\,dy\,dx \\
&= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-[x^2+(\tau+px)^2]/2\sigma^2}\,dx \\
&= \frac{1}{\sqrt{1+p^2}}e^{-t^2/[2(1+p^2)\sigma^2]}.
\end{aligned}$$

Radon Transform—Square

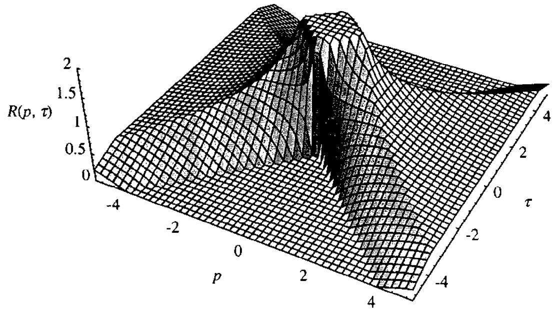

$$R(p,\tau)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\delta[y-(\tau+px)]\,dy\,dx, \quad (1)$$

where

$$f(x,y)\equiv\begin{cases}1 & \text{for } x,y\in[-a,a]\\ 0 & \text{otherwise}\end{cases} \quad (2)$$

and

$$\delta(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-ikx} \quad (3)$$

is the DELTA FUNCTION.

$$\begin{aligned}R(p,\tau)&=\frac{1}{2\pi}\int_{-a}^{a}\int_{-a}^{a}\int_{-\infty}^{\infty}e^{-ik[y-(\tau+px)]}\,dk\,dy\,dx\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{ik\tau}\left[\int_{-a}^{a}e^{-ky}\,dy\int_{-a}^{a}e^{ikpx}\,dx\right]dk\\&=\frac{1}{2\pi}e^{ik\tau}\frac{1}{-ik}[e^{-iky}]_{-a}^{a}\frac{1}{ikp}[e^{ikpx}]_{-a}^{a}\,dk\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{ik\tau}\frac{1}{k^2p}[-2i\sin(ka)][2i\sin(kpa)]\,dk\\&=\frac{2}{\pi p}\int_{-\infty}^{\infty}\frac{\sin(ka)\sin(kpa)e^{ik\tau}}{k^2}\,dk\\&=\frac{4}{\pi p}\int_{0}^{\infty}\frac{\sin(ka)\sin(kpa)\cos(k\tau)}{k^2}\,dk\\&=\frac{2}{\pi p}\int_{0}^{\infty}\frac{\sin[k(\tau+a)]-\sin[k(\tau-a)]}{k^2}\sin(kpa)\,dk\\&=\frac{2}{\pi p}\left\{\int_{0}^{\infty}\frac{\sin[k(\tau+a)]\sin(kpa)}{k^2}\,dk\right.\\&\quad\left.-\int_{0}^{\infty}\frac{\sin[k(\tau-a)]\sin(kpa)}{k^2}\,dk\right\}. \quad (4)\end{aligned}$$

From Gradshteyn and Ryzhik (1979, equation 3.741.3),

$$\int_{0}^{\infty}\frac{\sin(ax)\sin(bx)}{x^2}\,dx=\tfrac{1}{2}\pi\,\mathrm{sgn}(ab)\min(|a|,|b|), \quad (5)$$

so

$$\begin{aligned}R(p,\tau)=\frac{1}{p}\{&\mathrm{sgn}[(\tau+a)pa]\min(|\tau+a|,|pa|)\\&-\mathrm{sgn}[(\tau-a)pa]\min(|\tau-a|,|pa|)\}. \quad (6)\end{aligned}$$

References
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, 1979.

Railroad Track Problem

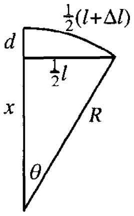

Given a straight segment of track of length l, add a small segment Δl so that the track bows into a circular ARC. Find the maximum displacement d of the bowed track. The PYTHAGOREAN THEOREM gives

$$R^2=x^2+(\tfrac{1}{2}l)^2, \quad (1)$$

But R is simply $x+d$, so

$$R^2=(x+d)^2=x^2+2xd+d^2. \quad (2)$$

Solving (1) and (2) for x gives

$$x=\frac{\frac{1}{4}l^2-d^2}{2d}. \quad (3)$$

Expressing the length of the ARC in terms of the central angle,

$$\begin{aligned}\tfrac{1}{2}(l+\Delta l)&=\theta(d+x)=\theta\left(d+\frac{\frac{1}{4}l^2-d^2}{2d}\right)\\&=\theta\left(\frac{2d^2+\frac{1}{4}l^2-d^2}{2d}\right)=\theta\left(\frac{d^2+\frac{1}{4}l^2}{2d}\right). \quad (4)\end{aligned}$$

But θ is given by

$$\tan\theta=\frac{\frac{1}{2}l}{x}=\frac{\frac{1}{2}l(2d)}{\frac{1}{4}l^2-d^2}=\frac{dl}{\frac{1}{4}l^2-d^2}, \quad (5)$$

so plugging θ in gives

$$\tfrac{1}{2}(l+\Delta l)=\left(\frac{d^2+\frac{1}{4}l^2}{2d}\right)\tan^{-1}\left(\frac{dl}{\frac{1}{4}l^2-d^2}\right) \quad (6)$$

$$d(l+\Delta l)=(d^2+\tfrac{1}{4}l^2)\tan^{-1}\left(\frac{dl}{\frac{1}{4}l^2-d^2}\right). \quad (7)$$

For $l\gg d$,

$$\frac{dl}{\frac{1}{4}l^2\left(1-\frac{d^2}{4l^2}\right)}=\frac{4d}{l}\left(1-\frac{4d^2}{l^2}\right)^{-1}\approx\frac{4d}{l}\left(1+\frac{4d}{l^2}\right). \quad (8)$$

Therefore,

$$\begin{aligned}&d(l+\Delta l)\\&\approx(d^2+\tfrac{1}{4}l^2)\left\{\frac{4d}{l}\left(1+\frac{4d^2}{l^2}\right)-\frac{1}{3}\left[\frac{4d}{l}\left(1+\frac{4d^2}{l^2}\right)\right]^3\right\}\\&\approx(d^2+\tfrac{1}{4}l^2)\left[\frac{4d}{l}+\frac{16d^3}{l^3}-\frac{1}{3}\left(\frac{4d}{l}\right)^3\left(1+3\frac{4d^2}{l^2}\right)\right]. \quad (9)\end{aligned}$$

Keeping only terms to order $(d/l)^3$,

$$dl + \Delta l \approx \frac{4d^3}{l} + dl + \frac{4d^3}{l} - \frac{16}{3}\frac{d^3}{l} \qquad (10)$$

$$\Delta l \approx \left(8 - \tfrac{16}{3}\right)\frac{d^3}{l} = \frac{24-16}{3}\frac{d^3}{l} = \frac{8}{3}\frac{d^3}{l}, \qquad (11)$$

so

$$d^2 = \tfrac{3}{8} l \Delta l \qquad (12)$$

and

$$d \approx \tfrac{1}{2}\sqrt{\tfrac{3}{2} l \Delta l} = \tfrac{1}{4}\sqrt{6 l \Delta l}. \qquad (13)$$

If we take $l = 1$ mile $= 5280$ feet and $\Delta l = 1$ foot, then $d \approx 44.450$ feet.

Ramanujan 6-10-8 Identity

Let $ad = bc$, then

$$\begin{aligned} &64[(a+b+c)^6 + (b+c+d)^6 - (c+d+a)^6 \\ &\qquad -(d+a+b)^6 + (a-d)^6 - (b-c)^6] \\ &\times[(a+b+c)^{10} + (b+c+d)^{10} - (c+d+a)^{10} \\ &\qquad -(d+a+b)^{10} + (a-d)^{10} - (b-c)^{10}] \\ &= 45[(a+b+c)^8 + (b+c+d)^8 - (c+d+a)^8 \\ &\qquad -(d+a+b)^8 + (a-d)^8 - (b-c)^8]^2. \qquad (1) \end{aligned}$$

This can also be expressed by defining

$$\begin{aligned} &F_{2m}(a,b,c,d) = (a+b+c)^{2m} + (b+c+d)^{2m} \\ &-(c+d+a)^{2m} - (d+a+b)^{2m} + (a-d)^{2m} - (b-c)^{2m} \qquad (2) \\ &f_{2m}(x,y) = (1+x+y)^{2m} + (x+y+xy)^{2m} - (y+xy+1)^{2m} \\ &\qquad -(xy+1+x)^{2m} + (1-xy)^{2m} - (x-y)^{2m}. \qquad (3) \end{aligned}$$

Then

$$F_{2m}(a,b,c,d) = a^{2m} f_{2m}(x,y), \qquad (4)$$

and identity (1) can then be written

$$64 f_6(x,y) f_{10}(x,y) = 45 {f_8}^2(x,y). \qquad (5)$$

Incidentally,

$$f_2(x,y) = 0 \qquad (6)$$

$$f_4(x,y) = 0. \qquad (7)$$

References

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 3 and 102–106, 1994.

Berndt, B. C. and Bhargava, S. "A Remarkable Identity Found in Ramanujan's Third Notebook." *Glasgow Math. J.* **34**, 341–345, 1992.

Berndt, B. C. and Bhargava, S. "Ramanujan—For Lowbrows." *Amer. Math. Monthly* **100**, 644–656, 1993.

Bhargava, S. "On a Family of Ramanujan's Formulas for Sums of Fourth Powers." *Ganita* **43**, 63–67, 1992.

Hirschhorn, M. D. "Two or Three Identities of Ramanujan." *Amer. Math. Monthly* **105**, 52–55, 1998.

Nanjundiah, T. S. "A Note on an Identity of Ramanujan." *Amer. Math. Monthly* **100**, 485–487, 1993.

Ramanujan, S. *Notebooks.* New York: Springer-Verlag, pp. 385–386, 1987.

Ramanujan Constant

The IRRATIONAL constant

$$R \equiv e^{\pi\sqrt{163}} = 262537412640768743.99999999999925\ldots$$

which is very close to an INTEGER. Numbers such as the Ramanujan constant can be found using the theory of MODULAR FUNCTIONS. A few rather spectacular examples are given by Ramanujan (1913–14), including the one above, and can be generated using some amazing properties of the j-FUNCTION.

M. Gardner (Apr. 1975) played an April Fool's joke on the readers of *Scientific American* by claiming that this number was exactly an INTEGER. He admitted the hoax a few months later (Gardner, July 1975).

see also ALMOST INTEGER, CLASS NUMBER, j-FUNCTION

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 387, 1987.

Castellanos, D. "The Ubiquitous Pi. Part I." *Math. Mag.* **61**, 67–98, 1988.

Gardner, M. "Mathematical Games: Six Sensational Discoveries that Somehow or Another have Escaped Public Attention." *Sci. Amer.* **232**, 127–131, Apr. 1975.

Gardner, M. "Mathematical Games: On Tessellating the Plane with Convex Polygons." *Sci. Amer.* **232**, 112–117, Jul. 1975.

Good, I. J. "What is the Most Amazing Approximate Integer in the Universe?" *Pi Mu Epsilon J.* **5**, 314–315, 1972.

Plouffe, S. "$e^{\pi\sqrt{163}}$, the Ramanujan Number." `http://lacim.uqam.ca/piDATA/ramanujan.txt`.

Ramanujan, S. "Modular Equations and Approximations to π." *Quart. J. Pure Appl. Math.* **45**, 350–372, 1913–1914.

Wolfram, S. *The Mathematica Book, 3rd ed.* New York: Cambridge University Press, p. 52, 1996.

Ramanujan Continued Fraction

Let $f(a,b)$ be a RAMANUJAN THETA FUNCTION. Then

$$\frac{f(-q,-q^4)}{f(-q^2,-q^3)} = \frac{1}{1+}\,\frac{q}{1+}\,\frac{q^2}{1+}\,\frac{q^3}{1+\ldots},$$

where the quantity on the right is a CONTINUED FRACTION.

see also RAMANUJAN THETA FUNCTIONS

Ramanujan Cos/Cosh Identity

$$\left[1 + 2\sum_{n=1}^{\infty}\frac{\cos(n\theta)}{\cosh(n\pi)}\right]^{-2} + \left[1 + 2\sum_{n=1}^{\infty}\frac{\cosh(n\theta)}{\cosh(n\pi)}\right]^{-2} = \frac{2\Gamma^4(\frac{3}{4})}{\pi},$$

where $\Gamma(z)$ is the GAMMA FUNCTION.

Ramanujan-Eisenstein Series

Let t be a discriminant,

$$q \equiv -e^{-\pi\sqrt{t}}, \tag{1}$$

then

$$\begin{aligned} E_2(q) \equiv L(q) &\equiv 1 - 24\sum_{k=1}^{\infty} \frac{(2k+1)q^{2k+1}}{1-q^{2k+1}} \\ &= \left(\frac{2K}{\pi}\right)^2 (1-2k^2) \end{aligned} \tag{2}$$

$$\begin{aligned} E_4(q) \equiv M(q) &\equiv 1 + 240\sum_{k=1}^{\infty} \frac{k^3 q^{2k}}{1-q^{2k}} \\ &= \left(\frac{2K}{\pi}\right)^4 (1-k^2k'^2) \end{aligned} \tag{3}$$

$$\begin{aligned} E_6(q) \equiv N(q) &\equiv 1 - 504\sum_{k=1}^{\infty} \frac{k^5 q^{2k}}{1-q^{2k}} \\ &= \left(\frac{2K}{\pi}\right)^6 (1-2k^2)(1+\tfrac{1}{2}k^2k'^2). \end{aligned} \tag{4}$$

see also KLEIN'S ABSOLUTE INVARIANT, PI

References

Borwein, J. M. and Borwein, P. B. "Class Number Three Ramanujan Type Series for $1/\pi$." *J. Comput. Appl. Math.* **46**, 281–290, 1993.

Ramanujan, S. "Modular Equations and Approximations to π." *Quart. J. Pure Appl. Math.* **45**, 350–372, 1913–1914.

Ramanujan Function

$$\phi(a,n) \equiv 1 + 2\sum_{k=1}^{n} \frac{1}{(ak)^3 - ak}$$

$$\phi(a) \equiv \lim_{n\to\infty} \phi(a,n) = 1 + 2\sum_{k=1}^{\infty} \frac{1}{(ak)^3 - ak}.$$

The values of $\phi(n)$ for $n = 2, 3, \ldots$ are

$$\begin{aligned} \phi(2) &= 2\ln 2 \\ \phi(3) &= \ln 3 \\ \phi(4) &= \tfrac{3}{2}\ln 2 \\ \phi(6) &= \tfrac{1}{2}\ln 3 + \tfrac{1}{3}\ln 4. \end{aligned}$$

Ramanujan g- and G-Functions

Following Ramanujan (1913–14), write

$$\prod_{k=1,3,5,\ldots}^{\infty} (1+e^{-k\pi\sqrt{n}}) = 2^{1/4}e^{-\pi\sqrt{n}/24}G_n \tag{1}$$

$$\prod_{k=1,3,5,\ldots}^{\infty} (1-e^{-k\pi\sqrt{n}}) = 2^{1/4}e^{-\pi\sqrt{n}/24}g_n. \tag{2}$$

These satisfy the equalities

$$g_{4n} = 2^{1/4}g_nG_n \tag{3}$$

$$G_n = G_{1/n} \tag{4}$$

$${g_n}^{-1} = g_{4/n} \tag{5}$$

$$\tfrac{1}{4} = (g_nG_n)^8({G_n}^8 - {g_n}^8). \tag{6}$$

G_n and g_n can be derived using the theory of MODULAR FUNCTIONS and can always be expressed as roots of algebraic equations when n is RATIONAL. For simplicity, Ramanujan tabulated g_n for n EVEN and G_n for n ODD. However, (6) allows G_n and g_n to be solved for in terms of g_n and G_n, giving

$$g_n = \tfrac{1}{2}\left({G_n}^8 + \sqrt{{G_n}^{16} - {G_n}^{-8}}\right)^{1/8} \tag{7}$$

$$G_n = \tfrac{1}{2}\left({g_n}^8 + \sqrt{{g_n}^{16} + G{g_n}^{-8}}\right)^{1/8}. \tag{8}$$

Using (3) and the above two equations allows g_{4n} to be computed in terms of g_n or G_n

$$g_{4n} = \begin{cases} 2^{1/8}g_n\left({g_n}^8 + \sqrt{{g_n}^{16} + {g_n}^{-8}}\right)^{1/8} & \text{for } n \text{ even} \\ 2^{1/8}G_n\left({G_n}^8 + \sqrt{{G_n}^{16} - {G_n}^{-8}}\right)^{1/8} & \text{for } n \text{ odd.} \end{cases} \tag{9}$$

In terms of the PARAMETER k and complementary PARAMETER k',

$$G_n = (2k_nk'_n)^{-1/12} \tag{10}$$

$$g_n = \left(\frac{k'^2_n}{2k}\right)^{1/12}. \tag{11}$$

Here,

$$k_n = \lambda^*(n) \tag{12}$$

is the ELLIPTIC LAMBDA FUNCTION, which gives the value of k for which

$$\frac{K'(k)}{K(k)} = \sqrt{n}. \tag{13}$$

Solving for $\lambda^*(n)$ gives

$$\lambda^*(n) = \tfrac{1}{2}[\sqrt{1+{G_n}^{-12}} - \sqrt{1-{G_n}^{-12}}] \tag{14}$$

$$\lambda^*(n) = {g_n}^6[\sqrt{{g_n}^{12} + {g_n}^{-12}} - {g_n}^6]. \tag{15}$$

Analytic values for small values of n can be found in Ramanujan (1913–1914) and Borwein and Borwein (1987), and have been compiled in Weisstein (1996). Ramanujan (1913–1914) contains a typographical error labeling G_{465} as G_{265}.

see also G-FUNCTION

References

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, pp. 139 and 298, 1987.

Ramanujan, S. "Modular Equations and Approximations to π." *Quart. J. Pure. Appl. Math.* **45**, 350–372, 1913–1914.

Weisstein, E. W. "Elliptic Singular Values." `http://www.astro.virginia.edu/~eww6n/math/notebooks/EllipticSingular`.

Ramanujan's Hypergeometric Identity

$$1-\left(\frac{1}{2}\right)^3+\left(\frac{1\cdot 3}{2\cdot 4}\right)^3+\ldots={}_3F_2\left(\begin{matrix}\frac{1}{2},\frac{1}{2},\frac{1}{2}\\ 1,1\end{matrix};-1\right)$$
$$=\left[{}_2F_1\left(\begin{matrix}\frac{1}{4},\frac{1}{4}\\ 1\end{matrix};-1\right)\right]^2=\frac{\Gamma^2(\frac{9}{8})}{\Gamma^2(\frac{5}{4})\Gamma^2(\frac{7}{8})},$$

where ${}_2F_1(a,b;c;x)$ is a HYPERGEOMETRIC FUNCTION, ${}_3F_2(a,b,c;d;e;x)$ is a GENERALIZED HYPERGEOMETRIC FUNCTION, and $\Gamma(z)$ is a GAMMA FUNCTION.

References
Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, p. 106, 1959.

Ramanujan's Hypothesis

see TAU CONJECTURE

Ramanujan's Identity

$$5\frac{\phi^5(x^5)}{\phi^6(x)}=\sum_{m=0}^{\infty}P(5m+4)x^m,$$

where

$$\phi(x)=\prod_{m=1}^{\infty}(1-x^m)$$

and $P(n)$ is the PARTITION FUNCTION P.

see also RAMANUJAN'S SUM IDENTITY

Ramanujan's Integral

$$\int_{-\infty}^{\infty}\frac{J_{\mu+\xi}(x)}{x^{\mu+\xi}}\frac{J_{\nu-\xi}(y)}{y^{\nu-\xi}}e^{it\xi}\,d\xi$$
$$=\left[\frac{2\cos\left(\frac{1}{2}t\right)}{x^2e^{-it/2}+y^2e^{it/2}}\right]^{(\mu+\nu)/2}$$
$$\times J_{\mu+\nu}\left[\sqrt{2\cos\left(\tfrac{1}{2}t\right)\left(x^2e^{-it/2}+y^2e^{it/2}\right)}\right]e^{it(\nu-\mu)/2},$$

where $J_n(z)$ is a BESSEL FUNCTION OF THE FIRST KIND.

References
Watson, G. N. *A Treatise on the Theory of Bessel Functions, 2nd ed.* Cambridge, England: Cambridge University Press, 1966.

Ramanujan's Interpolation Formula

$$\int_0^{\infty}x^{s-1}\sum_{k=0}^{\infty}(-1)^kx^k\phi(k)\,dx=\frac{\pi\phi(-s)}{\sin(s\pi)}\qquad(1)$$
$$\int_0^{\infty}x^{s-1}\sum_{k=0}^{\infty}(-1)^k\frac{x^k}{k!}\lambda(k)\,dx=\Gamma(s)\lambda(-s),\qquad(2)$$

where $\lambda(z)$ is the DIRICHLET LAMBDA FUNCTION and $\Gamma(z)$ is the GAMMA FUNCTION. Equation (2) is obtained from (1) by defining

$$\phi(u)=\frac{\lambda(u)}{\Gamma(1+u)}.\qquad(3)$$

These formulas give valid results only for certain classes of functions.

References
Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, pp. 15 and 186–195, 1959.

Ramanujan's Master Theorem

Suppose that in some NEIGHBORHOOD of $x=0$,

$$F(x)=\sum_{k=0}^{\infty}\frac{\phi(k)(-x)^k}{k!}.$$

Then

$$\int_0^{\infty}x^{n-1}F(x)\,dx=\Gamma(n)\phi(-n).$$

References
Berndt, B. C. *Ramanujan's Notebooks: Part I.* New York: Springer-Verlag, p. 298, 1985.

Ramanujan-Petersson Conjecture

A CONJECTURE for the EIGENVALUES of modular forms under HECKE OPERATORS.

Ramanujan Psi Sum

A sum which includes both the JACOBI TRIPLE PRODUCT and the q-BINOMIAL THEOREM as special cases. Ramanujan's sum is

$$\sum_{n=-\infty}^{\infty}\frac{(a)_n}{(b)_n}x^n=\frac{(ax)_\infty(q/ax)_\infty(q)_\infty(b/a)_\infty}{(x)_\infty(b/ax)_\infty(b)_\infty(q/a)_\infty},$$

where the NOTATION $(q)_k$ denotes q-SERIES. For $b=q$, this becomes the q-BINOMIAL THEOREM.

see also JACOBI TRIPLE PRODUCT, q-BINOMIAL THEOREM, q-SERIES

Ramanujan's Square Equation

It has been proved that the only solutions to the DIOPHANTINE EQUATION

$$2^n - 7 = x^2$$

are $n = 3$, 4, 5, 7, and 15 (Beeler *et al.* 1972, Item 31).

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Ramanujan's Sum

The sum

$$c_q(m) = \sum_{h^*(q)} e^{2\pi ihm/q}, \tag{1}$$

where h runs through the residues RELATIVELY PRIME to q, which is important in the representation of numbers by the sums of squares. If $(q, q') = 1$ (i.e., q and q' are RELATIVELY PRIME), then

$$c_{qq'}(m) = c_q(m)c_{q'}(m). \tag{2}$$

For argument 1,

$$c_b(1) = \mu(b), \tag{3}$$

where μ is the MÖBIUS FUNCTION, and for general m,

$$c_b(m) = \mu\left(\frac{b}{(b,m)}\right)\frac{\phi(b)}{\phi\left(\frac{b}{(b,m)}\right)}. \tag{4}$$

see also MÖBIUS FUNCTION, WEYL'S CRITERION

References

Vardi, I. *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, p. 254, 1991.

Ramanujan's Sum Identity

If

$$\frac{1 + 53x + 9x^2}{1 - 82x - 82x^2 + x^3} = \sum_{n=1}^{\infty} a_n x^n \tag{1}$$

$$\frac{2 - 26x - 12x^2}{1 - 82x - 82x^2 + x^3} = \sum_{n=0}^{\infty} b_n x^n \tag{2}$$

$$\frac{2 + 8x - 10x^2}{1 - 82x - 82x^2 + x^3} = \sum_{n=0}^{\infty} c_n x^n, \tag{3}$$

then

$$a_n{}^3 + b_n{}^3 = c_n{}^3 + (-1)^n. \tag{4}$$

Hirschhorn (1995) showed that

$$a_n = \tfrac{1}{85}[(64 + 8\sqrt{85})\alpha^n + (64 - 8\sqrt{85})\beta^n - 43(-1)^n] \tag{5}$$

$$b_n = \tfrac{1}{85}[(77 + 7\sqrt{85})\alpha^n + (77 - 7\sqrt{85})\beta^n + 16(-1)^n] \tag{6}$$

$$c_n = \tfrac{1}{85}[(93 + 9\sqrt{85})\alpha^n + (93 - 9\sqrt{85})\beta^n - 16(-1)^n], \tag{7}$$

where

$$\alpha = \tfrac{1}{2}(83 + 9\sqrt{85}) \tag{8}$$

$$\beta = \tfrac{1}{2}(83 - 9\sqrt{85}). \tag{9}$$

Hirschhorn (1996) showed that checking the first seven cases $n = 0$ to 6 is sufficient to prove the result.

References

Hirschhorn, M. D. "An Amazing Identity of Ramanujan." *Math. Mag.* **68**, 199–201, 1995.

Hirschhorn, M. D. "A Proof in the Spirit of Zeilberger of an Amazing Identity of Ramanujan." *Math. Mag.* **69**, 267–269, 1996.

Ramanujan's Tau-Dirichlet Series

see TAU-DIRICHLET SERIES

Ramanujan's Tau Function

see TAU FUNCTION

Ramanujan Theta Functions

Ramanujan's one-variable theta function is defined by

$$\varphi(x) \equiv \sum_{m=-\infty}^{\infty} x^{m^2}. \tag{1}$$

It is equal to the function in the JACOBI TRIPLE PRODUCT with $z = 1$,

$$\begin{aligned}\varphi(x) = G(1) &= \prod_{n=1}^{\infty}(1 + x^{2n-1})^2(1 - x^{2n}) \\ &= \sum_{m=-\infty}^{\infty} x^{m^2} = 1 + 2\sum_{m=0}^{\infty} x^{m^2}.\end{aligned} \tag{2}$$

Special values include

$$\varphi(e^{-\pi\sqrt{2}}) = \frac{\Gamma(\frac{9}{8})}{\Gamma(\frac{5}{4})}\sqrt{\frac{\Gamma(\frac{1}{4})}{2^{1/4}\pi}} \tag{3}$$

$$\varphi(e^{-\pi}) = \frac{\pi^{1/4}}{\Gamma(\frac{3}{4})} \tag{4}$$

$$\varphi(e^{-2\pi}) = \frac{\sqrt{2 + \sqrt{2}}}{2}\frac{\pi}{\Gamma(\frac{3}{4})}. \tag{5}$$

Ramanujan's two-variable theta function is defined by

$$f(a, b) \equiv \sum_{n=-\infty}^{\infty} a^{n(n+1)/2} b^{n(n-1)/2} \tag{6}$$

for $|ab| < 1$. It satisfies

$$f(-1, a) = 0 \tag{7}$$

$$f(a, b) = f(b, a) = (-a; ab)_\infty(-b; ab)_\infty(ab; ab)_\infty \tag{8}$$

$$f(-q) \equiv f(-q,-q^2) = \sum_{k=0}^{\infty}(-1)^k q^{k(2k-1)/2} + \sum_{k=1}^{\infty}(-1)^k q^{k(2k+1)/2} = (q;q)_\infty, \quad (9)$$

where $(q)_\infty$ are q-SERIES.

see also JACOBI TRIPLE PRODUCT, SCHRÖTER'S FORMULA, q-SERIES

Ramp Function

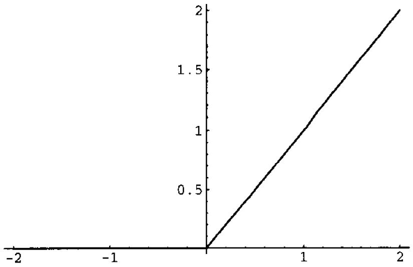

$$\begin{aligned} R(x) &\equiv xH(x) & (1)\\ &= \int_{-\infty}^{x} H(x')\,dx' & (2)\\ &= \int_{-\infty}^{\infty} H(x')H(x-x')\,dx' & (3)\\ &= H(x) * H(x), & (4)\end{aligned}$$

where $H(x)$ is the HEAVISIDE STEP FUNCTION and $*$ is the CONVOLUTION. The DERIVATIVE is

$$R'(x) = -H(x). \quad (5)$$

The FOURIER TRANSFORM of the ramp function is given by

$$\mathcal{F}[R(x)] = \int_{-\infty}^{\infty} e^{-2\pi ikx} R(x)\,dx = \pi i\delta'(2\pi k) - \frac{1}{4\pi^2 k^2}, \quad (6)$$

where $\delta(x)$ is the DELTA FUNCTION and $\delta'(x)$ its DERIVATIVE.

see also FOURIER TRANSFORM—RAMP FUNCTION, HEAVISIDE STEP FUNCTION, RECTANGLE FUNCTION, SGN, SQUARE WAVE

Ramphoid Cusp

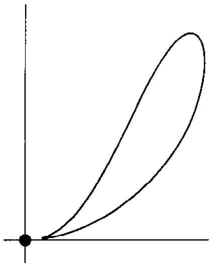

A type of CUSP as illustrated above for the curve $x^4 + x^2y^2 - 2x^2y - xy^2 + y^2 = 0$.

see also CUSP

References

Walker, R. J. *Algebraic Curves.* New York: Springer-Verlag, pp. 57–58, 1978.

Ramsey Number

The Ramsey number $R(m,n)$ gives the solution to the PARTY PROBLEM, which asks the minimum number of guests $R(m,n)$ that must be invited so that at least m will know each other (i.e., there exists a CLIQUE of order m) or at least n will not know each other (i.e., there exists an independent set of order n). By symmetry, it is true that

$$R(m,n) = R(n,m). \quad (1)$$

It also must be true that

$$R(2,m) = m. \quad (2)$$

A generalized Ramsey number is written

$$R(m_1,\ldots,m_k;n) \quad (3)$$

and is the smallest INTEGER R such that, no matter how each n-element SUBSET of an r-element SET are colored with k colors, there exists an i such that there is a SUBSET of size m_i, all of whose n-element SUBSETS are color i. The usual Ramsey numbers are then equivalent to $R(m,n) = R(m,n;2)$.

Bounds are given by

$$R(k,l) \le \begin{cases} R(k-1,l) + R(k,l-1) - 1 & \text{for } R(k-1,l) \text{ and } R(k,l-1) \text{ even} \\ R(k-1,l) + R(k,l-1) & \text{otherwise} \end{cases} \quad (4)$$

and

$$R(k,k) \le 4R(k-2,k) + 2 \quad (5)$$

(Chung and Grinstead 1983). Erdős proved that for diagonal Ramsey numbers $R(k,k)$,

$$\frac{k2^{k/2}}{e\sqrt{2}} < R(k,k). \quad (6)$$

This result was subsequently improved by a factor of 2 by Spencer (1975). $R(3,k)$ was known since 1980 to be bounded from above by $c_2k^2/\ln k$, and Griggs (1983) showed that $c_2 = 5/12$ was an acceptable limit. J.-H. Kim (Cipra 1995) subsequently bounded $R(3,k)$ by a similar expression from below, so

$$c_1\frac{k^2}{\ln k} \le R(3,k) \le c_2\frac{k^2}{\ln k}. \quad (7)$$

Burr (1983) gives Ramsey numbers for all 113 graphs with no more than 6 EDGES and no isolated points.

A summary of known results up to 1983 for $R(m,n)$ is given in Chung and Grinstead (1983). Radziszowski maintains an up-to-date list of the best current bounds, reproduced in part in the following table for $R(m,n;2)$.

m	n	$R(m,n)$
3	3	6
3	4	9
3	5	14
3	6	18
3	7	23
3	8	28
3	9	36
3	10	[40, 43]
3	11	[46, 51]
3	12	[52, 60]
3	13	[60, 69]
3	14	[66, 78]
3	15	[73, 89]
3	16	[79, ∞]
3	17	[92, ∞]
3	18	[98, ∞]
3	19	[106, ∞]
3	20	[109, ∞]
3	21	[122, ∞]
3	22	[125, ∞]
3	23	[136, ∞]

m	n	$R(m,n)$
4	4	18
4	5	25
4	6	[35, 41]
4	7	[49, 62]
4	8	[55, 85]
4	9	[69, 116]
4	10	[80, 151]
4	11	[93, 191]
4	12	[98, 238]
4	13	[112, 291]
4	14	[119, 349]
4	15	[128, 417]

m	n	$R(m,n)$
5	5	[43, 49]
5	6	[58, 87]
5	7	[80, 143]
5	8	[95, 216]
5	9	[116, 371]
5	10	[1, 445]

m	n	$R(m,n)$
6	6	[102, 165]
6	7	[109, 300]
6	8	[122, 497]
6	9	[153, 784]
6	10	[167, 1180]

m	n	$R(m,n)$
7	7	[205, 545]
7	8	[1, 1035]
7	9	[1, 1724]
7	10	[1, 2842]

m	n	$R(m,n)$
8	8	[282, 1874]
8	9	[1, 3597]
8	10	[1, 6116]

m	n	$R(m,n)$
9	9	[565, 6680]
9	10	[1, 12795]

m	n	$R(m,n)$
10	10	[798, 23981]

m	n	$R(m,n)$
11	11	[522, ∞]

Known values for generalized Ramsey numbers are given in the following table.

$R(\ldots;2)$	Bounds
$R(3,3,3;2)$	17
$R(3,3,4;2)$	[30, 32]
$R(3,3,5;2)$	[45, 59]
$R(3,4,4;2)$	[55, 81]
$R(3,4,5;2)$	≥ 80
$R(4,4,4;2)$	[128, 242]
$R(3,3,3,3;2)$	[51, 64]
$R(3,3,3,4;2)$	[87, 159]
$R(3,3,3,3,3;2)$	[162, 317]
$R(3,3,3,3,3,3;2)$	[1, 500]

$R(\ldots;3)$	Bounds
$R(4,4;3)$	[14, 15]

see also CLIQUE, COMPLETE GRAPH, EXTREMAL GRAPH, IRREDUNDANT RAMSEY NUMBER, SCHUR NUMBER

References

Burr, S. A. "Generalized Ramsey Theory for Graphs—A Survey." In *Graphs and Combinatorics* (Ed. R. A. Bari and F. Harary). New York: Springer-Verlag, pp. 52–75, 1964.

Burr, S. A. "Diagonal Ramsey Numbers for Small Graphs." *J. Graph Th.* **7**, 57–69, 1983.

Chartrand, G. "The Problem of the Eccentric Hosts: An Introduction to Ramsey Numbers." §5.1 in *Introductory Graph Theory.* New York: Dover, pp. 108–115, 1985.

Chung, F. R. K. "On the Ramsey Numbers $N(3,3,\ldots,3;2)$." *Discrete Math.* **5**, 317–321, 1973.

Chung, F. and Grinstead, C. G. "A Survey of Bounds for Classical Ramsey Numbers." *J. Graph. Th.* **7**, 25–37, 1983.

Cipra, B. "A Visit to Asymptopia Yields Insights into Set Structures." *Science* **267**, 964–965, 1995.

Exoo, G. "On Two Classical Ramsey Numbers of the Form $R(3,n)$." *SIAM J. Discrete Math.* **2**, 488–490, 1989.

Exoo, G. "Announcement: On the Ramsey Numbers $R(4,6)$, $R(5,6)$ and $R(3,12)$." *Ars Combin.* **35**, 85, 1993.

Exoo, G. "Some New Ramsey Colorings." *Electronic J. Combinatorics* **5**, No. 1, R29, 1–5, 1998. `http://www.combinatorics.org/Volume_5/v5i1toc.html`.

Folkmann, J. "Notes on the Ramsey Number $N(3,3,3,3)$." *J. Combinat. Theory. Ser. A* **16**, 371–379, 1974.

Gardner, M. "Mathematical Games: In Which Joining Sets of Points by Lines Leads into Diverse (and Diverting) Paths." *Sci. Amer.* **237**, 18–28, 1977.

Gardner, M. *Penrose Tiles and Trapdoor Ciphers... and the Return of Dr. Matrix, reissue ed.* New York: W. H. Freeman, pp. 240–241, 1989.

Giraud, G. "Une minoration du nombre de quadrangles unicolores et son application a la majoration des nombres de Ramsey binaires bicolors." *C. R. Acad. Sci. Paris A* **276**, 1173–1175, 1973.

Graham, R. L.; Rothschild, B. L.; and Spencer, J. H. *Ramsey Theory, 2nd ed.* New York: Wiley, 1990.

Graver, J. E. and Yackel, J. "Some Graph Theoretic Results Associated with Ramsey's Theorem." *J. Combin. Th.* **4**, 125–175, 1968.

Greenwood, R. E. and Gleason, A. M. "Combinatorial Relations and Chromatic Graphs." *Canad. J. Math.* **7**, 1–7, 1955.

Griggs, J. R. "An Upper Bound on the Ramsey Numbers $R(3,k)$." *J. Comb. Th. A* **35**, 145–153, 1983.

Grinstead, C. M. and Roberts, S. M. "On the Ramsey Numbers $R(3,8)$ and $R(3,9)$." *J. Combinat. Th. Ser. B* **33**, 27–51, 1982.

Guldan, F. and Tomasta, P. "New Lower Bounds of Some Diagonal Ramsey Numbers." *J. Graph. Th.* **7**, 149–151, 1983.

Hanson, D. "Sum-Free Sets and Ramsey Numbers." *Discrete Math.* **14**, 57–61, 1976.

Harary, F. "Recent Results on Generalized Ramsey Theory for Graphs." *Graph Theory and Applications* (Ed. Y. Alai, D. R. Lick, and A. T. White). New York: Springer-Verlag, pp. 125–138, 1972.

Hill, R. and Irving, R. W. "On Group Partitions Associated with Lower Bounds for Symmetric Ramsey Numbers." *European J. Combin.* **3**, 35–50, 1982.

Kalbfleisch, J. G. *Chromatic Graphs and Ramsey's Theorem.* Ph.D. thesis, University of Waterloo, January 1966.

McKay, B. D. and Min, Z. K. "The Value of the Ramsey Number $R(3,8)$." *J. Graph Th.* **16**, 99–105, 1992.

McKay, B. D. and Radziszowski, S. P. "$R(4,5) = 25$." *J. Graph. Th* **19**, 309–322, 1995.

Piwakowski, K. "Applying Tabu Search to Determine New Ramsey Numbers." *Electronic J. Combinatorics* **3**, R6, 1–4, 1996. `http://www.combinatorics.org/Volume_3/volume3.html#R6`.

Radziszowski, S. P. "Small Ramsey Numbers." *Electronic J. Combin.* **1**, DS1 1–29, Rev. Mar. 25, 1996. `http://ejc.math.gatech.edu:8080/Journal/Surveys/ds1.ps`.

Radziszowski, S. and Kreher, D. L. "Upper Bounds for Some Ramsey Numbers $R(3,k)$." *J. Combinat. Math. Combin. Comput.* **4**, 207–212, 1988.

Spencer, J. H. "Ramsey's Theorem—A New Lower Bound." *J. Combinat. Theory Ser. A* **18**, 108–115, 1975.

Wang, Q. and Wang, G. "New Lower Bounds for the Ramsey Numbers $R(3,q)$." *Beijing Daxue Xuebao* **25**, 117–121, 1989.

Whitehead, E. G. "The Ramsey Number $N(3,3,3,3;2)$." *Discrete Math.* **4**, 389–396, 1973.

Ramsey's Theorem

A generalization of DILWORTH'S LEMMA. For each $m,n \in \mathbb{N}$ with $m,n \geq 2$, there exists a least INTEGER $R(m,n)$ (the RAMSEY NUMBER) such that no matter how the COMPLETE GRAPH $K_{R(m,n)}$ is two-colored, it will contain a green SUBGRAPH K_m or a red subgroup K_n. Furthermore,

$$R(m,n) \leq R(m-1,n) + R(m,n-1)$$

if $m,n \geq 3$. The theorem can be equivalently stated that, for all $\in \mathbb{N}$, there exists an $n \in \mathbb{N}$ such that any complete DIGRAPH on n VERTICES contains a complete transitive SUBGRAPH of m VERTICES. Ramsey's theorem is a generalization of the PIGEONHOLE PRINCIPLE since

$$R(\underbrace{2,2,\ldots,2}_{t}) = t+1.$$

see also DILWORTH'S LEMMA, NATURAL INDEPENDENCE PHENOMENON, PIGEONHOLE PRINCIPLE, RAMSEY NUMBER

References

Graham, R. L.; Rothschild, B. L.; and Spencer, J. H. *Ramsey Theory, 2nd ed.* New York: Wiley, 1990.

Spencer, J. "Large Numbers and Unprovable Theorems." *Amer. Math. Monthly* **90**, 669–675, 1983.

Randelbrot Set

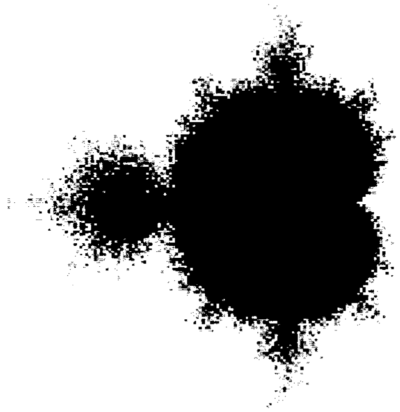

The FRACTAL-like figure obtained by performing the same iteration as for the MANDELBROT SET, but adding a random component R,

$$z_{n+1} = {z_n}^2 + c + R.$$

In the above plot, $R \equiv R_x + iR_y$, where $R_x, R_y \in [-0.05, 0.05]$.

see also MANDELBROT SET

References

Dickau, R. M. "Randelbrot Set." `http://forum.swarthmore.edu/advanced/robertd/randelbrot.html`.

Random Distribution

A DISTRIBUTION in which the variates occur with PROBABILITIES asymptotically matching their "true" underlying DISTRIBUTION is said to be random.

see also DISTRIBUTION, RANDOM NUMBER

Random Dot Stereogram

see STEREOGRAM

Random Graph

A random graph is a GRAPH in which properties such as the number of NODES, EDGES, and connections between them are determined in some random way. Erdős and Rényi showed that for many monotone-increasing properties of random graphs, graphs of a size slightly less than a certain threshold are very unlikely to have the property, whereas graphs with a few more EDGES are almost certain to have it. This is known as a PHASE TRANSITION.

see also GRAPH (GRAPH THEORY), GRAPH THEORY

References

Bollobás, B. *Random Graphs.* London: Academic Press, 1985.

Steele, J. M. "Gibbs' Measures on Combinatorial Objects and the Central Limit Theorem for an Exponential Family of Random Trees." *Prob. Eng. Inform. Sci.* **1**, 47–59, 1987.

Random Matrix

A random matrix is a MATRIX of given type and size whose entries consist of random numbers from some specified distribution.

see also MATRIX

Random Number

Computer-generated random numbers are sometimes called PSEUDORANDOM NUMBERS, while the term "random" is reserved for the output of unpredictable physical processes. It is impossible to produce an arbitrarily long string of random digits and prove it is random. Strangely, it is very difficult for humans to produce a string of random digits, and computer programs can be written which, on average, actually predict some of the digits humans will write down based on previous ones.

The LINEAR CONGRUENCE METHOD is one algorithm for generating PSEUDORANDOM NUMBERS. The initial number used as the starting point in a random number generating algorithm is known as the SEED. The goodness of random numbers generated by a given ALGORITHM can be analyzed by examining its NOISE SPHERE.

see also BAYS' SHUFFLE, CLIFF RANDOM NUMBER GENERATOR, QUASIRANDOM SEQUENCE, SCHRAGE'S ALGORITHM, STOCHASTIC

References

Bassein, S. "A Sampler of Randomness." *Amer. Math. Monthly* **103**, 483–490, 1996.

Bratley, P.; Fox, B. L.; and Schrage, E. L. *A Guide to Simulation, 2nd ed.* New York: Springer-Verlag, 1996.

Dahlquist, G. and Bjorck, A. Ch. 11 in *Numerical Methods.* Englewood Cliffs, NJ: Prentice-Hall, 1974.

Deak, I. *Random Number Generators and Simulation.* New York: State Mutual Book & Periodical Service, 1990.

Forsythe, G. E.; Malcolm, M. A.; and Moler, C. B. Ch. 10 in *Computer Methods for Mathematical Computations.* Englewood Cliffs, NJ: Prentice-Hall, 1977.

Gardner, M. "Random Numbers." Ch. 13 in *Mathematical Carnival: A New Round-Up of Tantalizers and Puzzles from Scientific American.* New York: Vintage, 1977.

James, F. "A Review of Pseudorandom Number Generators." *Computer Physics Comm.* **60**, 329–344, 1990.

Kac, M. "What is Random?" *Amer. Sci.* **71**, 405–406, 1983.

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, pp. 200–201 and 205–207, 1962.

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, pp. 151–154, 1951.

Knuth, D. E. Ch. 3 in *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, 1981.

Marsaglia, G. "A Current View of Random Number Generators." In *Computer Science and Statistics: Proceedings of the Symposium on the Interface, 16th, Atlanta, Georgia, March 1984* (Ed. L. Billard). New York: Elsevier, 1985.

Park, S. and Miller, K. "Random Number Generators: Good Ones are Hard to Find." *Comm. ACM* **31**, 1192–1201, 1988.

Peterson, I. *The Jungles of Randomness: A Mathematical Safari.* New York: Wiley, 1997.

Pickover, C. A. "Computers, Randomness, Mind, and Infinity." Ch. 31 in *Keys to Infinity.* New York: W. H. Freeman, pp. 233–247, 1995.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Random Numbers." Ch. 7 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 266–306, 1992.

Schrage, L. "A More Portable Fortran Random Number Generator." *ACM Trans. Math. Software* **5**, 132–138, 1979.

Schroeder, M. "Random Number Generators." In *Number Theory in Science and Communication, with Applications in Cryptography, Physics, Digital Information, Computing and Self-Similarity, 3rd ed.* New York: Springer-Verlag, pp. 289–295, 1990.

Random Percolation

see PERCOLATION THEORY

Random Polynomial

A POLYNOMIAL having random COEFFICIENTS.

see also KAC FORMULA

Random Variable

A random variable is a measurable function from a PROBABILITY SPACE $(S, \mathbb{S}, P)$ into a MEASURABLE SPACE $(S', \mathbb{S}')$ known as the STATE SPACE.

see also PROBABILITY SPACE, RANDOM DISTRIBUTION, RANDOM NUMBER, STATE SPACE, VARIATE

References

Gikhman, I. I. and Skorokhod, A. V. *Introduction to the Theory of Random Processes.* New York: Dover, 1997.

Random Walk

A random process consisting of a sequence of discrete steps of fixed length. The random thermal perturbations in a liquid are responsible for a random walk phenomenon known as Brownian motion, and the collisions of molecules in a gas are a random walk responsible for diffusion. Random walks have interesting mathematical properties that vary greatly depending on the dimension in which the walk occurs and whether it is confined to a lattice.

see also RANDOM WALK—1-D, RANDOM WALK—2-D, RANDOM WALK—3-D, SELF-AVOIDING WALK

References

Barber, M. N. and Ninham, B. W. *Random and Restricted Walks: Theory and Applications.* New York: Gordon and Breach, 1970.
Chandrasekhar, S. In *Selected Papers on Noise and Stochastic Processes* (Ed. N. Wax). New York: Dover, 1954.
Doyle, P. G. and Snell, J. L. *Random Walks and Electric Networks.* Washington, DC: Math. Assoc. Amer, 1984.
Dykin, E. B. and Uspenskii, V. A. *Random Walks.* New York: Heath, 1963.
Feller, W. *An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd ed.* New York: Wiley, 1968.
Gardner, M. "Random Walks." Ch. 6–7 in *Mathematical Circus: More Puzzles, Games, Paradoxes, and Other Mathematical Entertainments.* Washington, DC: Math. Assoc. Amer., 1992.
Hughes, B. D. *Random Walks and Random Environments, Vol. 1: Random Walks.* New York: Oxford University Press, 1995.
Hughes, B. D. *Random Walks and Random Environments, Vol. 2: Random Environments.* New York: Oxford University Press, 1996.
Lawler, G. F. *Intersections of Random Walks.* Boston, MA: Birkhäuser, 1996.
Spitzer, F. *Principles of Random Walk, 2nd ed.* New York: Springer-Verlag, 1976.

Random Walk—1-D

Let N steps of equal length be taken along a LINE. Let p be the probability of taking a step to the right, q the probability of taking a step to the left, n_1 the number of steps taken to the right, and n_2 the number of steps taken to the left. The quantities p, q, n_1, n_2, and N are related by

$$p+q=1 \tag{1}$$

and

$$n_1+n_2=N. \tag{2}$$

Now examine the probability of taking exactly n_1 steps out of N to the right. There are $\binom{N}{n_1}=\binom{n_1+n_2}{n_1}$ ways of taking n_1 steps to the right and n_2 to the left, where $\binom{n}{m}$ is a BINOMIAL COEFFICIENT. The probability of taking a particular ordered sequence of n_1 and n_2 steps is $p^{n_1}q^{n_2}$. Therefore,

$$P(n_1)=\frac{(n_1+n_2)!}{n_1!n_2!}p^{n_1}q^{n_2}=\frac{N!}{n_1!(N-n_1)!}p^{n_1}q^{N-n_1}, \tag{3}$$

where $n!$ is a FACTORIAL. This is a BINOMIAL DISTRIBUTION and satisfies

$$\sum_{n_1=0}^{N}P(n_1)=(p+q)^N=1^N=1. \tag{4}$$

The MEAN number of steps n_1 to the right is then

$$\langle n_1\rangle\equiv\sum_{n_1=0}^{N}n_1P(n_1)=\sum_{n_1=0}^{N}\frac{N!}{n_1!(N-n_1)!}p^{n_1}q^{N-n_1}n_1, \tag{5}$$

but

$$n_1p^{n_1}=p\frac{\partial}{\partial p}p^{n_1}, \tag{6}$$

so

$$\begin{aligned}\langle n_1\rangle&=\sum_{n_1=0}^{N}\frac{N!}{n_1!(N-n_1)!}\left(p\frac{\partial}{\partial p}p^{n_1}\right)q^{N-n_1}\\&=p\frac{\partial}{\partial p}\sum_{n_1=0}^{N}\frac{N!}{n_1!(N-n_1)!}p^{n_1}q^{N-n_1}\\&=p\frac{\partial}{\partial p}(p+q)^N=pN(p+q)^{N-1}=pN.\end{aligned} \tag{7}$$

From the BINOMIAL THEOREM,

$$\langle n_2\rangle=N-\langle n_1\rangle=N(1-p)=qN. \tag{8}$$

The VARIANCE is given by

$$\sigma_{n_1}{}^2=\left\langle n_1{}^2\right\rangle-\langle n_1\rangle^2. \tag{9}$$

But

$$\left\langle n_1{}^2\right\rangle=\sum_{n_1=0}^{N}\frac{N!}{n_1!(N-n_1)!}p^{n_1}q^{N-n_1}n_1{}^2, \tag{10}$$

so

$$\begin{aligned}n_1{}^2p^{n_1}&=n_1\left(p\frac{\partial}{\partial p}\right)p^{n_1}=\left(p\frac{\partial}{\partial p}\right)^2p^{n_1}\\&=\sum_{n_1=0}^{N}\frac{N!}{n_1!(N-n_1)!}\left(p\frac{\partial}{\partial p}\right)^2p^{n_1}q^{N-n_1}\\&=\left(p\frac{\partial}{\partial p}\right)^2\sum_{n_1=0}^{N}\frac{N!}{n_1!(N-n_1)!}p^{n_1}q^{N-n_1}\\&=\left(p\frac{\partial}{\partial p}\right)^2(p+q)^N=\frac{\partial}{\partial p}[pN(p+q)N-1]\\&=p[N(p+q)^{N-1}+pN(N-1)(p+q)^{N-2}]\\&=p[N+pN(N-1)]\\&=pN[1+pN-p]=(Np)^2+Npq\\&=\langle n_1\rangle^2+Npq.\end{aligned} \tag{11}$$

Therefore,

$$\sigma_{n_1}{}^2=\left\langle n_1{}^2\right\rangle-\langle n_1\rangle^2=Npq, \tag{12}$$

and the ROOT-MEAN-SQUARE deviation is

$$\sigma_{n_1}=\sqrt{Npq}. \tag{13}$$

For a large number of total steps N, the BINOMIAL DISTRIBUTION characterizing the distribution approaches a GAUSSIAN DISTRIBUTION.

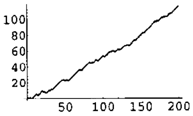

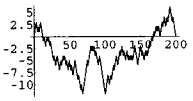

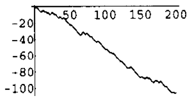

Consider now the distribution of the distances d_N traveled after a given number of steps,

$$d_N \equiv n_1 - n_2 = 2n_1 - N, \tag{14}$$

as opposed to the *number* of steps in a given direction. The above plots show $d_N(p)$ for $N = 200$ and three values $p = 0.1$, $p = 0.5$, and $p = 0.9$, respectively. Clearly, weighting the steps toward one direction or the other influences the overall trend, but there is still a great deal of random scatter, as emphasized by the plot below, which shows three random walks all with $p = 0.5$.

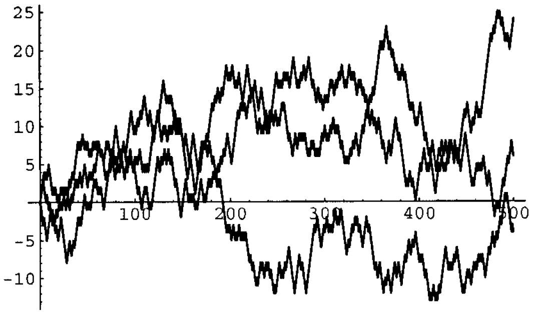

Surprisingly, the most probable number of sign changes in a walk is 0, followed by 1, then 2, etc.

For a random walk with $p = 1/2$, the probability $P_N(d)$ of traveling a given distance d after N steps is given in the following table.

steps	-5	-4	-3	-2	-1	0	1	2	3	4	5
0						1					
1					$\frac{1}{2}$	0	$\frac{1}{2}$				
2				$\frac{1}{4}$	0	$\frac{2}{4}$	0	$\frac{1}{4}$			
3			$\frac{1}{8}$	0	$\frac{3}{8}$	0	$\frac{3}{8}$	0	$\frac{1}{8}$		
4		$\frac{1}{16}$	0	$\frac{4}{16}$	0	$\frac{6}{16}$	0	$\frac{4}{16}$	0	$\frac{1}{16}$	
5	$\frac{1}{32}$	0	$\frac{5}{32}$	0	$\frac{10}{32}$	0	$\frac{10}{32}$	0	$\frac{5}{32}$	0	$\frac{1}{32}$

In this table, subsequent rows are found by adding HALF of each cell in a given row to each of the two cells diagonally below it. In fact, it is simply PASCAL'S TRIANGLE padded with intervening zeros and with each row multiplied by an additional factor of 1/2. The COEFFICIENTS in this triangle are given by

$$P_N(d) = \frac{1}{2^N}\binom{N}{\frac{d+N}{2}}. \tag{15}$$

The expectation value of the distance after N steps is therefore

$$\begin{aligned}\langle d_N\rangle &= \sum_{d=-N,-(N-2),\ldots}^{N} |d|\, P_N(d)\\ &= \frac{1}{2^N}\sum_{d=-N,-(N-2),\ldots}^{N} \frac{|d|\,N!}{\left(\frac{N+d}{2}\right)!\left(\frac{N-d}{2}\right)!}.\end{aligned} \tag{16}$$

This sum can be done symbolically by separately considering the cases N EVEN and N ODD. First, consider EVEN N so that $N \equiv 2J$. Then

$$\begin{aligned}\langle d_{2J}\rangle &= \frac{N!}{2^N}\left[\sum_{\substack{d=-2J,\\ -2(J-1),\ldots}}^{-2} \frac{|d|}{\left(\frac{2J+d}{2}\right)!\left(\frac{2J-d}{2}\right)!}\right.\\ &\qquad \left.+\sum_{d=0} \frac{|d|}{\left(\frac{2J+d}{2}\right)!\left(\frac{2J-d}{2}\right)!} + \sum_{d=2,4,\ldots}^{2J} \frac{|d|}{\left(\frac{2J+d}{2}\right)!\left(\frac{2J-d}{2}\right)!}\right]\\ &= \frac{N!}{2^N}\left[\sum_{d=-J,-(J-1),\ldots}^{-1} \frac{|2d|}{\left(\frac{2J+2d}{2}\right)!\left(\frac{2J-2d}{2}\right)!}\right.\\ &\qquad\qquad \left.+ \sum_{d=1,2,\ldots}^{J} \frac{|2d|}{\left(\frac{2J+2d}{2}\right)!\left(\frac{2J-2d}{2}\right)!}\right]\\ &= \frac{N!}{2^N}\left[2\sum_{d=1}^{J} \frac{2d}{(J+d)!(J-d)!}\right]\\ &= \frac{N!}{2^{N-2}}\sum_{d=1}^{J} \frac{d}{(J+d)!(J-d)!}.\end{aligned} \tag{17}$$

But this sum can be evaluated analytically as

$$\sum_{d=1}^{J} \frac{d}{(J+d)!(J-d)!} = \frac{J}{2\Gamma^2(1+J)}, \tag{18}$$

which, when combined with $N = 2J$ and plugged back in, gives

$$\langle d_{2J}\rangle = \frac{\Gamma(2J+1)J}{2^{2J-1}\Gamma^2(1+J)} = \frac{\Gamma(2J)}{2^{2J-2}\Gamma^2(J)}. \tag{19}$$

But the LEGENDRE DUPLICATION FORMULA gives

$$\Gamma(2J) = \frac{2^{2J-1/2}\Gamma(J)\Gamma(J+\frac{1}{2})}{\sqrt{2\pi}}, \tag{20}$$

so

$$\langle d_{2J}\rangle = \frac{\frac{1}{\sqrt{2\pi}}2^{2J-1/2}\Gamma(J)\Gamma(J+\frac{1}{2})}{2^{2J-2}\Gamma^2(J)} = \frac{2}{\sqrt{\pi}}\frac{\Gamma(J+\frac{1}{2})}{\Gamma(J)}. \tag{21}$$

Now consider N ODD, so $N \equiv 2J - 1$. Then

$$\begin{aligned}\langle d_{2J-1}\rangle &= \frac{N!}{2^N}\left[\sum_{\substack{d=-(2J-1),\\ -(2J+1),\ldots}}^{-1} \frac{|d|}{\left(\frac{2J-1+d}{2}\right)!\left(\frac{2J-1-d}{2}\right)!}\right.\\ &\quad \left.+\sum_{d=1,3,\ldots}^{2J-1} \frac{|d|}{\left(\frac{2J-1+d}{2}\right)!\left(\frac{2J-1-d}{2}\right)!}\right]\\ &= \frac{N!}{2^{N-1}}\left[\sum_{d=1,3,\ldots}^{2J-1} \frac{d}{\left(\frac{2J-1+d}{2}\right)!\left(\frac{2J-1-d}{2}\right)!}\right]\\ &= \frac{N!}{2^{N-1}}\left[\sum_{d=2,4,\ldots}^{2J} \frac{d-1}{\left(\frac{2J-2+d}{2}\right)!\left(\frac{2J-d}{2}\right)!}\right]\\ &= \frac{\Gamma(2J)}{2^{2J-2}}\left[\sum_{d=1}^{J} \frac{2d-1}{(J+d-1)!(J-d)!}\right]\\ &= \Gamma(2J)\left[\frac{1+J-{}_2F_1(1,-J;J;1)}{2^{2J-2}\Gamma(J)\Gamma(1+J)}+\frac{1}{\Gamma(2J)}\right]\\ &= \frac{\frac{2^{2J-1/2}}{\sqrt{2\pi}}\Gamma(J)\Gamma(J+1/2)}{2^{2J-2}\Gamma^2(J)J}[1+J-{}_2F_1(1,-J;J;-1)]+1\\ &= \frac{2}{\sqrt{\pi}}\frac{\Gamma(J+\frac{1}{2})}{J\Gamma(J)}[1+J-{}_2F_1(1,-J;J;-1)]+1. \end{aligned} \tag{22}$$

But the HYPERGEOMETRIC FUNCTION ${}_2F_1$ has the special value

$$ {}_2F_1(1,-J;J;-1) = \frac{\sqrt{\pi}}{2}\frac{J\Gamma(J)}{\Gamma(J+\frac{1}{2})}+1, \tag{23}$$

so

$$\langle d_{2J-1}\rangle = \frac{2}{\sqrt{\pi}}\frac{\Gamma(J+\frac{1}{2})}{\Gamma(J)}. \tag{24}$$

Summarizing the EVEN and ODD solutions,

$$\langle d_N\rangle = \frac{2}{\sqrt{\pi}}\frac{\Gamma(J+\frac{1}{2})}{\Gamma(J)}, \tag{25}$$

where

$$\begin{cases} J = \frac{1}{2}N & \text{for } N \text{ even}\\ J = \frac{1}{2}(N+1) & \text{for } N \text{ odd.}\end{cases} \tag{26}$$

Written explicitly in terms of N,

$$\langle d_N\rangle = \begin{cases} \frac{2}{\sqrt{\pi}}\frac{\Gamma(\frac{1}{2}N+\frac{1}{2})}{\Gamma(\frac{1}{2}N)} & \text{for } N \text{ even}\\ \frac{2}{\sqrt{\pi}}\frac{\Gamma(\frac{1}{2}N+1)}{\Gamma(\frac{1}{2}N+\frac{1}{2})} & \text{for } N \text{ odd.}\end{cases} \tag{27}$$

The first few values of $\langle d_N\rangle$ are then

$$\begin{aligned}\langle d_0\rangle &= 0\\ \langle d_1\rangle = \langle d_2\rangle &= 1\\ \langle d_3\rangle = \langle d_4\rangle &= \tfrac{3}{2}\\ \langle d_5\rangle = \langle d_6\rangle &= \tfrac{15}{8}\\ \langle d_7\rangle = \langle d_8\rangle &= \tfrac{35}{16}\\ \langle d_9\rangle = \langle d_{10}\rangle &= \tfrac{315}{128}\\ \langle d_{11}\rangle = \langle d_{12}\rangle &= \tfrac{693}{256}\\ \langle d_{13}\rangle = \langle d_{14}\rangle &= \tfrac{3003}{1024}.\end{aligned}$$

Now, examine the asymptotic behavior of $\langle d_N\rangle$. The asymptotic expansion of the GAMMA FUNCTION ratio is

$$\frac{\Gamma(J+\frac{1}{2})}{\Gamma(J)} = \sqrt{J}\left(1-\frac{1}{8J}+\frac{1}{128J^2}+\ldots\right) \tag{28}$$

(Graham *et al.* 1994), so plugging in the expression for $\langle d_N\rangle$ gives the asymptotic series

$$\begin{aligned}\langle d_N\rangle = \sqrt{\frac{2N}{\pi}}\biggl(1 &\mp \frac{1}{4N}+\frac{1}{32N^2}\pm\frac{5}{128N^3}\\ &-\frac{21}{2048N^4}\mp\ldots\biggr),\end{aligned} \tag{29}$$

where the top signs are taken for N EVEN and the bottom signs for N ODD. Therefore, for large N,

$$\langle d_N\rangle \sim \sqrt{\frac{2N}{\pi}}, \tag{30}$$

which is also shown in Mosteller *et al.* (1961, p. 14).

see also BINOMIAL DISTRIBUTION, CATALAN NUMBER, p-GOOD PATH, PÓLYA'S RANDOM WALK CONSTANTS, RANDOM WALK—2-D, RANDOM WALK—3-D, SELF-AVOIDING WALK

References

Chandrasekhar, S. "Stochastic Problems in Physics and Astronomy." *Rev. Modern Phys.* **15**, 1–89, 1943. Reprinted in *Noise and Stochastic Processes* (Ed. N. Wax). New York: Dover, pp. 3–91, 1954.

Feller, W. Ch. 3 in *An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd ed., rev. printing.* New York: Wiley, 1968.

Gardner, M. Chs. 6–7 in *Mathematical Carnival: A New Round-Up of Tantalizers and Puzzles from Scientific American.* New York: Vintage Books, 1977.

Graham, R. L.; Knuth, D. E.; and Patashnik, O. Answer to problem 9.60 in *Concrete Mathematics: A Foundation for Computer Science, 2nd ed.* Reading, MA: Addison-Wesley, 1994.

Hersh, R. and Griego, R. J. "Brownian Motion and Potential Theory." *Sci. Amer.* **220**, 67–74, 1969.

Kac, M. "Random Walk and the Theory of Brownian Motion." *Amer. Math. Monthly* **54**, 369–391, 1947. Reprinted in *Noise and Stochastic Processes* (Ed. N. Wax). New York: Dover, pp. 295–317, 1954.

Mosteller, F.; Rourke, R. E. K.; and Thomas, G. B. *Probability and Statistics.* Reading, MA: Addison-Wesley, 1961.

Random Walk—2-D

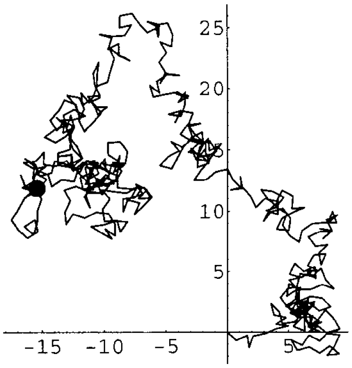

In a PLANE, consider a sum of N 2-D VECTORS with random orientations. Use PHASOR notation, and let the phase of each VECTOR be RANDOM. Assume N unit steps are taken in an arbitrary direction (i.e., with the angle θ uniformly distributed in $[0, 2\pi)$ and *not* on a LATTICE), as illustrated above. The position z in the COMPLEX PLANE after N steps is then given by

$$z = \sum_{j=1}^{N} e^{i\theta_j}, \tag{1}$$

which has ABSOLUTE SQUARE

$$\begin{aligned} |z|^2 &= \sum_{j=1}^{N} e^{i\theta_j} \sum_{k=1}^{N} e^{-i\theta_k} = \sum_{j=1}^{N}\sum_{k=1}^{N} e^{i(\theta_j-\theta_k)} \\ &= N + \sum_{\substack{j,k=1\\ k\neq j}}^{N} e^{i(\theta_j-\theta_k)}. \end{aligned} \tag{2}$$

Therefore,

$$\left\langle |z|^2 \right\rangle = N + \left\langle \sum_{\substack{j,k=1\\ k\neq j}}^{N} e^{i(\theta_j-\theta_k)} \right\rangle. \tag{3}$$

Each step is likely to be in any direction, so both θ_j and θ_k are RANDOM VARIABLES with identical MEANS of zero, and their difference is also a random variable. Averaging over this distribution, which has equally likely POSITIVE and NEGATIVE values yields an expectation value of 0, so

$$\left\langle |z|^2 \right\rangle = N. \tag{4}$$

The root-mean-square distance after N unit steps is therefore

$$|z|_{\rm rms} = \sqrt{N}, \tag{5}$$

so with a step size of l, this becomes

$$d_{\rm rms} = l\sqrt{N}. \tag{6}$$

In order to travel a distance d

$$N \approx \left(\frac{d}{l}\right)^2 \tag{7}$$

steps are therefore required.

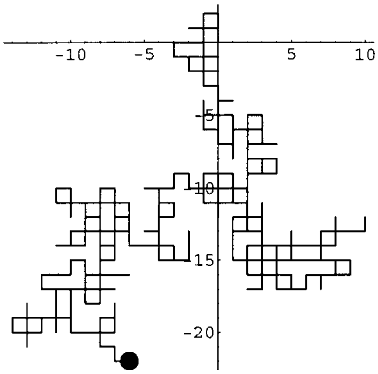

Amazingly, it has been proven that on a 2-D LATTICE, a random walk has unity probability of reaching any point (including the starting point) as the number of steps approaches INFINITY.

see also PÓLYA'S RANDOM WALK CONSTANTS, RANDOM WALK—1-D, RANDOM WALK—3-D

Random Walk—3-D

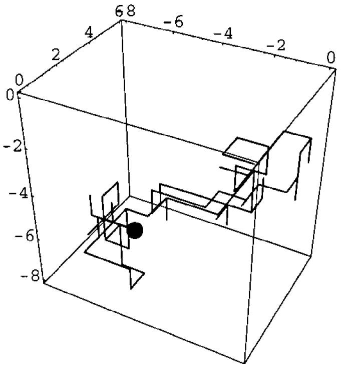

On a 3-D LATTICE, a random walk has *less than* unity probability of reaching any point (including the starting point) as the number of steps approaches infinity. The probability of reaching the starting point again is 0.3405373296.... This is one of PÓLYA'S RANDOM WALK CONSTANTS.

see also PÓLYA'S RANDOM WALK CONSTANTS, RANDOM WALK—1-D, RANDOM WALK—2-D

Range (Image)

If T is MAP over a DOMAIN D, then the range of T is defined as

$$\text{Range}(T) = T(D) = \{T(\mathbf{X}) : \mathbf{X} \in D\}.$$

The range $T(D)$ is also called the IMAGE of D under T.

see also DOMAIN, MAP

Range (Line Segment)

The set of all points on a LINE SEGMENT, also called a PENCIL.

see also PERSPECTIVITY, SECTION (PENCIL)

References

Woods, F. S. *Higher Geometry: An Introduction to Advanced Methods in Analytic Geometry.* New York: Dover, p. 8, 1961.

Range (Statistics)

$$R \equiv \max(x_i) - \min(x_i). \qquad (1)$$

For small samples, the range is a good estimator of the population STANDARD DEVIATION (Kenney and Keeping 1962, pp. 213–214). For a continuous UNIFORM DISTRIBUTION

$$P(x) = \begin{cases} \frac{1}{C} & \text{for } 0 < x < C \\ 0 & \text{for } |x| < C, \end{cases} \qquad (2)$$

the distribution of the range is given by

$$D(R) = N\left(\frac{R}{C}\right)^{N-1} - (N-1)\left(\frac{R}{C}\right)^{N}. \qquad (3)$$

Given two samples with sizes m and n and ranges R_1 and R_2, let $u \equiv R_1/R_2$. Then

$$D(u) = \begin{cases} \frac{m(m-1)n(n-1)}{(m+n)(m+n-1)(m+n-2)} \\ \times[(m+n)u^{m-2} - (m+n-2)u^{m-1}] \\ \qquad \text{for } 0 \leq u \leq 1 \\ \frac{m(m-1)n(n-1)}{(m+n)(m+n-1)(m+n-2)} \\ \times[(m+n)u^{-n} - (m+n-2)u^{-n-1}] \\ \qquad \text{for } 1 \leq u < \infty. \end{cases} \qquad (4)$$

The MEAN is

$$\mu_u = \frac{(m-1)n}{(m+1)(n-2)}, \qquad (5)$$

and the MODE is

$$\hat{u} = \begin{cases} \frac{(m-2)(m+n)}{(m-1)(m+n-2)} & \text{for } m-n \leq 2 \\ \frac{(n+1)(m+n-2)}{n(m+n)} & \text{for } m-n \geq 2. \end{cases} \qquad (6)$$

References

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, pp. 213–214, 1962.

Rank

In a total generality, the "rank" of a mathematical object is defined whenever that object is FREE. In general, the rank of a FREE object is the CARDINALITY of the FREE generating SUBSET G. The word "rank" also refers to several unrelated concepts in mathematics involving groups, quadratic forms, sequences, statistics, and tensors.

see also RANK (GROUP), RANK (QUADRATIC FORM), RANK (SEQUENCE), RANK (STATISTICS), RANK (TENSOR)

Rank (Group)

For an arbitrary finitely generated ABELIAN GROUP G, the rank of G is defined to be the rank of the FREE generating SUBSET G modulo its TORSION SUBGROUP. For a finitely generated GROUP, the rank is defined to be the rank of its "Abelianization."

see also ABELIAN GROUP, BETTI NUMBER, BURNSIDE PROBLEM, QUASITHIN THEOREM, QUASI-UNIPOTENT GROUP, TORSION (GROUP THEORY)

Rank (Quadratic Form)

For a QUADRATIC FORM Q in the canonical form

$$Q = {y_1}^2 + {y_2}^2 + \ldots + {y_p}^2 - {y_{p+1}}^2 - {y_{p+2}}^2 - \ldots - {y_r}^2,$$

the rank is the total number r of square terms (both POSITIVE and NEGATIVE).

see also SIGNATURE (QUADRATIC FORM)

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1105, 1979.

Rank (Sequence)

The position of a RATIONAL NUMBER in the SEQUENCE 1, $\frac{1}{2}$, 2, $\frac{1}{3}$, 3, $\frac{1}{4}$, $\frac{2}{3}$, $\frac{3}{2}$, 4, $\frac{1}{5}$, ..., ordered in terms of increasing NUMERATOR+DENOMINATOR.

see also ENCODING, FAREY SERIES

Rank (Statistics)

The ORDINAL NUMBER of a value in a list arranged in a specified order (usually decreasing).

see also SPEARMAN RANK CORRELATION, WILCOXON RANK SUM TEST, WILCOXON SIGNED RANK TEST, ZIPF'S LAW

Rank (Tensor)

The total number of CONTRAVARIANT and COVARIANT indices of a TENSOR. The rank of a TENSOR is independent of the number of DIMENSIONS of the SPACE.

Rank	Object
0	scalar
1	vector
≥ 2	tensor

see also CONTRAVARIANT TENSOR, COVARIANT TENSOR, SCALAR, TENSOR, VECTOR

Ranunculoid

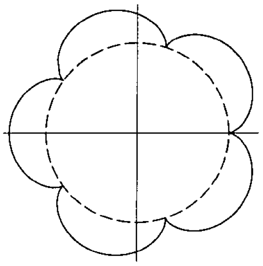

An EPICYCLOID with $n = 5$ cusps, named after the buttercup genus *Ranunculus* (Madachy 1979).

see also EPICYCLOID.

References

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, p. 223, 1979.

Pickover, C. A. *Keys to Infinity.* New York: Wiley, pp. 79–80, 1995.

RAT-Free Set

A RAT-free set is a set of points, no three of which determine a RIGHT TRIANGLE. Let $f(n)$ be the smallest RAT-free subset guaranteed to be contained in a planar set of n points, then the function $f(n)$ is bounded by

$$\sqrt{n} \leq f(n) \leq 2\sqrt{n}.$$

References

Abbott, H. L. "On a Conjecture of Erdős and Silverman in Combinatorial Geometry." *J. Combin. Th. A* **29**, 380–381, 1980.

Chan, W. K. "On the Largest RAT-FREE Subset of a Finite Set of Points." *Pi Mu Epsilon*, Spring 1987.

Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., pp. 250–251, 1991.

Seidenberg, A. "A Simple Proof of a Theorem of Erdős and Szekeres." *J. London Math. Soc.* **34**, 352, 1959.

Ratio

The ratio of two numbers r and s is written r/s, where r is the NUMERATOR and s is the DENOMINATOR. The ratio of r to s is equivalent to the QUOTIENT r/s. Betting ODDS written as $r : s$ correspond to $s/(r+s)$. A number which can be expressed as a ratio of INTEGERS is called a RATIONAL NUMBER.

see also DENOMINATOR, FRACTION, NUMERATOR, ODDS, QUOTIENT, RATIONAL NUMBER

Ratio Distribution

Given two distributions Y and X with joint probability density function $f(x,y)$, let $U = Y/X$ be the ratio distribution. Then the distribution function of u is

$$\begin{aligned} D(u) &= P(U \leq u) \\ &= P(Y \leq uX | X > 0) + P(Y \geq uX | X < 0) \\ &= \int_0^\infty \int_0^{ux} f(x,y)\,dy\,dx + \int_{-\infty}^0 \int_{ux}^0 f(x,y)\,dy\,dx. \end{aligned} \tag{1}$$

The probability function is then

$$\begin{aligned} P(u) = D'(u) &= \int_0^\infty x f(x,ux)\,dx - \int_{-\infty}^0 x f(x,ux)\,dx \\ &= \int_{-\infty}^\infty |x| f(x,ux)\,dx. \end{aligned} \tag{2}$$

For variates with a standard NORMAL DISTRIBUTION, the ratio distribution is a CAUCHY DISTRIBUTION. For a UNIFORM DISTRIBUTION

$$f(x,y) = \begin{cases} 1 & \text{for } x, y \in [0,1] \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

$$P(u) = \begin{cases} 0 & u < 0 \\ \int_0^1 x\,dx = [\tfrac{1}{2}x^2] = \tfrac{1}{2} & \text{for } 0 \leq u \leq 1 \\ \int_0^{1/u} x\,dx = [\tfrac{1}{2}x^2]_0^{1/u} = \tfrac{1}{2u^2} & \text{for } u > 1. \end{cases} \tag{4}$$

see also CAUCHY DISTRIBUTION

Ratio Test

Let u_k be a SERIES with POSITIVE terms and suppose

$$\rho \equiv \lim_{k \to \infty} \frac{u_{k+1}}{u_k}.$$

Then

1. If $\rho < 1$, the SERIES CONVERGES.
2. If $\rho > 1$ or $\rho = \infty$, the SERIES DIVERGES.
3. If $\rho = 1$, the SERIES may CONVERGE or DIVERGE.

The test is also called the CAUCHY RATIO TEST or D'ALEMBERT RATIO TEST.

see also CONVERGENCE TESTS

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 282–283, 1985.

Bromwich, T. J. I'a. and MacRobert, T. M. *An Introduction to the Theory of Infinite Series, 3rd ed.* New York: Chelsea, p. 28, 1991.

Rational Approximation

If r is any number and n is any INTEGER, then there is a RATIONAL NUMBER m/n for which

$$0 \leq r - \frac{m}{n} < \frac{1}{n}. \tag{1}$$

If r is IRRATIONAL and k is any WHOLE NUMBER, there is a FRACTION m/n with $n \leq k$ and for which

$$0 \leq r - \frac{m}{n} < \frac{1}{nk}. \tag{2}$$

Furthermore, there are an infinite number of FRACTIONS m/n for which

$$0 \leq r - \frac{m}{n} < \frac{1}{n^2}. \tag{3}$$

Hurwitz has shown that for an IRRATIONAL NUMBER ζ

$$\left| \zeta - \frac{h}{k} \right| < \frac{1}{ck^2}, \tag{4}$$

there are infinitely RATIONAL NUMBERS h/k if $0 < c \leq \sqrt{5}$, but if $c > \sqrt{5}$, there are some ζ for which this approximation holds for only finitely many h/k.

Rational Canonical Form

There is an invertible matrix Q such that

$$\mathsf{Q}^{-1}\mathsf{T}\mathsf{Q} = \operatorname{diag}[L(\psi_1), L(\psi_2), \ldots, L(\psi_s)],$$

where $L(f)$ is the companion MATRIX for any MONIC POLYNOMIAL

$$f(\lambda) = f_0 + f_1\lambda + \ldots + f_n\lambda^n$$

with $f_n = 1$. The POLYNOMIALS ψ_i are called the "invariant factors" of T, and satisfy $\psi_{i+1}|\psi_i$ for $i = s - 1$, ..., 1 (Hartwig 1996).

References

Gantmacher, F. R. *The Theory of Matrices, Vol. 1.* New York: Chelsea, 1960.

Hartwig, R. E. "Roth's Removal Rule and the Rational Canonical Form." *Amer. Math. Monthly* **103**, 332–335, 1996.

Herstein, I. N. *Topics in Algebra, 2nd ed.* New York: Springer-Verlag, p. 162, 1975.

Hoffman, K. and Kunze, K. *Linear Algebra, 3rd ed.* Englewood Cliffs, NJ: Prentice-Hall, 1996.

Lancaster, P. and Tismenetsky, M. *The Theory of Matrices, 2nd ed.* New York: Academic Press, 1985.

Turnbull, H. W. and Aitken, A. C. *An Introduction to the Theory of Canonical Matrices, 2nd impression.* New York: Blackie and Sons, 1945.

Rational Cuboid

see EULER BRICK

Rational Distances

It is possible to find six points in the PLANE, no three on a LINE and no four on a CIRCLE (i.e., none of which are COLLINEAR or CONCYCLIC), such that all the mutual distances are RATIONAL. An example is illustrated by Guy (1994, p. 185).

It is not known if a TRIANGLE with INTEGER sides, MEDIANS, and AREA exists (although there are incorrect PROOFS of the impossibility in the literature). However, R. L. Rathbun, A. Kemnitz, and R. H. Buchholz have showed that there are infinitely many triangles with RATIONAL sides (HERONIAN TRIANGLES) with *two* RATIONAL MEDIANS (Guy 1994, p. 188).

see also COLLINEAR, CONCYCLIC, CYCLIC QUADRILATERAL, EQUILATERAL TRIANGLE, EULER BRICK, HERONIAN TRIANGLE, RATIONAL QUADRILATERAL, RATIONAL TRIANGLE, SQUARE, TRIANGLE

References

Guy, R. K. "Six General Points at Rational Distances" and "Triangles with Integer Sides, Medians, and Area." §D20 and D21 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 185–190 and 188–190, 1994.

Rational Domain

see FIELD

Rational Double Point

There are nine possible types of ISOLATED SINGULARITIES on a CUBIC SURFACE, eight of them rational double points. Each type of ISOLATED SINGULARITY has an associated normal form and COXETER-DYNKIN DIAGRAM (A_1, A_2, A_3, A_4, A_5, D_4, D_5, E_6 and $\tilde{E}_6$).

The eight types of rational double points (the $\tilde{E}_6$ type being the one excluded) can occur in only 20 combinations on a CUBIC SURFACE (of which Fischer 1986 gives 19): A_1, $2A_1$, $3A_1$, $4A_1$, A_2, (A_2, A_1), $2A_2$, $(2A_2, A_1)$, $3A_2$, A_3, (A_3, A_1), $(A_3, 2A_1)$, A_4, (A_4, A_1), A_5, (A_5, A_1), D_4, D_5, and E_6 (Looijenga 1978, Bruce and Wall 1979, Fischer 1986).

In particular, on a CUBIC SURFACE, precisely those configurations of rational double points occur for which the disjoint union of the COXETER-DYNKIN DIAGRAM is a SUBGRAPH of the COXETER-DYNKIN DIAGRAM $\tilde{E}_6$. Also, a surface specializes to a more complicated one precisely when its graph is contained in the graph of the other one (Fischer 1986).

see also COXETER-DYNKIN DIAGRAM, CUBIC SURFACE, ISOLATED SINGULARITY

References

Bruce, J. and Wall, C. T. C. "On the Classification of Cubic Surfaces." *J. London Math. Soc.* **19**, 245–256, 1979.

Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, p. 13, 1986.

Fischer, G. (Ed.). Plates 14–31 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 17–31, 1986.

Looijenga, E. "On the Semi-Universal Deformation of a Simple Elliptic Hypersurface Singularity. Part II: The Discriminant." *Topology* **17**, 23–40, 1978.

Rodenberg, C. "Modelle von Flächen dritter Ordnung." In *Mathematische Abhandlungen aus dem Verlage Mathematischer Modelle von Martin Schilling.* Halle a. S., 1904.

Rational Function

A quotient of two polynomials $P(z)$ and $Q(z)$,

$$R(z) \equiv \frac{P(z)}{Q(z)},$$

is called a rational function. More generally, if P and Q are POLYNOMIALS in multiple variables, their quotient is a rational function.

see also ABEL'S CURVE THEOREM, CLOSED FORM, FUNDAMENTAL THEOREM OF SYMMETRIC FUNCTIONS, QUOTIENT-DIFFERENCE ALGORITHM, RATIONAL INTEGER, RATIONAL NUMBER, RIEMANN CURVE THEOREM

Rational Integer

A synonym for INTEGER. The word "rational" is sometimes used for emphasis to distinguish it from other types of "integers" such as CYCLOTOMIC INTEGERS, EISENSTEIN INTEGERS, and GAUSSIAN INTEGERS.

see also CYCLOTOMIC INTEGER, EISENSTEIN INTEGER, GAUSSIAN INTEGER, INTEGER, RATIONAL NUMBER

References
Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, p. 1, 1979.

Rational Number

A number that can be expressed as a Fraction p/q where p and q are Integers, is called a rational number with Numerator p and Denominator q. Numbers which are not rational are called Irrational Numbers. Any rational number is trivially also an Algebraic Number.

For a, b, and c any different rational numbers, then

$$\frac{1}{(a-b)^2}+\frac{1}{(b-c)^2}+\frac{1}{(c-a)^2}$$

is the Square of a rational number (Honsberger 1991). The probability that a random rational number has an Even Denominator is 1/3 (Beeler *et al.* 1972, Item 54).

see also Algebraic Integer, Algebraic Number, Anomalous Cancellation, Denominator, Dirichlet Function, Fraction, Integer, Irrational Number, Numerator, Quotient, Transcendental Number

References
Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.
Courant, R. and Robbins, H. "The Rational Numbers." §2.1 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 52–58,, 1996.
Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., pp. 52–53, 1991.

Rational Point

A K-rational point is a point (X, Y) on an Algebraic Curve, where X and Y are in a Field K.

The rational point may also be a Point at Infinity. For example, take the Elliptic Curve

$$Y^2 = X^3 + X + 42$$

and homogenize it by introducing a third variable Z so that each term has degree 3 as follows:

$$ZY^2 = X^3 + XZ^2 + 42Z^3.$$

Now, find the points at infinity by setting $Z = 0$, obtaining

$$0 = X^3.$$

Solving gives $X = 0$, Y equal to any value, and (by definition) $Z = 0$. Despite freedom in the choice of Y, there is only a single Point at Infinity because the two triples (X_1, Y_1, Z_1), (X_2, Y_2, Z_2) are considered to be equivalent (or identified) only if one is a scalar multiple of the other. Here, (0, 0, 0) is not considered to be a valid point. The triples (a, b, 1) correspond to the ordinary points (a, b), and the triples (a, b, 0) correspond to the Points at Infinity, usually called the Line at Infinity.

The rational points on Elliptic Curves over the Galois Field GF(q) are 5, 7, 9, 10, 13, 14, 16, ... (Sloane's A005523).

see also Elliptic Curve, Line at Infinity, Point at Infinity

References
Sloane, N. J. A. Sequence A005523/M3757 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Rational Quadrilateral

A rational quadrilateral is a Quadrilateral for which the sides, Diagonals, and Area are Rational. The simplest case has sides $a = 52$, $b = 25$, $c = 39$, and $d = 60$ and Diagonals of length $p = 63$ and $q = 56$.

see also Area, Diagonal (Polygon), Rational Quadrilateral

Rational Triangle

A rational triangle is a Triangle all of whose sides are Rational Numbers and all of whose Angles are Rational numbers of Degrees. The only such triangle is the Equilateral Triangle (Conway and Guy 1996).

see also Equilateral Triangle, Fermat's Right Triangle Theorem, Right Triangle

References
Conway, J. H. and Guy, R. K. "The Only Rational Triangle." In *The Book of Numbers.* New York: Springer-Verlag, pp. 201 and 228–239, 1996.

RATS Sequence

A sequence produced by the instructions "reverse, add, then sort the digits," where zeros are suppressed. For example, after 668 we get

$$668 + 866 = 1534,$$

so the next term is 1345. Applied to 1, the sequence gives 1, 2, 4, 8, 16, 77, 145, 668, 1345, 6677, 13444, 55778, ... (Sloane's A004000)

see also 196-Algorithm, Kaprekar Routine, Reversal, Sort-Then-Add Sequence

References
Sloane, N. J. A. Sequence A004000/M1137 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Ray

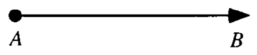

A Vector $\overrightarrow{AB}$ from a point A to a point B. In Geometry, a ray is usually taken as a half-infinite Line with one of the two points A and B taken to be at Infinity.

see also Line, Vector

Rayleigh Distribution

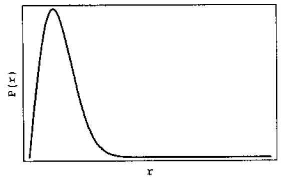

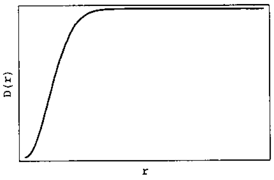

The distribution with PROBABILITY FUNCTION

$$P(r) = \frac{re^{-r^2/2s^2}}{s^2} \tag{1}$$

for $r \in [0, \infty)$. The MOMENTS about 0 are given by

$$\begin{aligned}\mu'_m &\equiv \int_0^\infty r^m P(r)\,dr = s^{-2}\int_0^\infty r^{m+1}e^{-r^2/2s^2}\,dr \\ &= s^{-2}I_{m+1}\left(\frac{1}{2s^2}\right),\end{aligned} \tag{2}$$

where $I(x)$ is a GAUSSIAN INTEGRAL. The first few of these are

$$I_1(a^{-1}) = \tfrac{1}{2}a \tag{3}$$
$$I_2(a^{-1}) = \tfrac{1}{4}a\sqrt{a\pi} \tag{4}$$
$$I_3(a^{-1}) = \tfrac{1}{2}a^2 \tag{5}$$
$$I_4(a^{-1}) = \tfrac{3}{8}a^2\sqrt{a\pi} \tag{6}$$
$$I_5(a^{-1}) = a^3, \tag{7}$$

so

$$\mu'_0 = s^{-2}\tfrac{1}{2}(2s^2) = 1 \tag{8}$$
$$\mu'_1 = s^{-2}\tfrac{1}{4}(2s^2)\sqrt{2s^2\pi} = \tfrac{1}{2}s\sqrt{2\pi} = s\sqrt{\frac{\pi}{2}} \tag{9}$$
$$\mu'_2 = s^{-2}\tfrac{1}{2}(2s^2)^2 = 2s^2 \tag{10}$$
$$\mu'_3 = s^{-2}\tfrac{3}{8}(2s^2)^2\sqrt{2s^2\pi} = \tfrac{3}{2}s^3\sqrt{2\pi} = 3s^3\sqrt{\frac{\pi}{2}} \tag{11}$$
$$\mu'_4 = s^{-2}(2s^2)^3 = 8s^4. \tag{12}$$

The MOMENTS about the MEAN are

$$\mu_2 = \mu'_2 - (\mu'_1)^2 = \frac{4-\pi}{2}s^2 \tag{13}$$
$$\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2(\mu'_1)^3 = \sqrt{\frac{\pi}{2}}(\pi-3)s^3 \tag{14}$$
$$\begin{aligned}\mu_4 &= \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2(\mu'_1)^2 - 3(\mu - 1')^4 \\ &= \frac{32-3\pi^2}{4}s^4,\end{aligned} \tag{15}$$

so the MEAN, VARIANCE, SKEWNESS, and KURTOSIS are

$$\mu = \mu'_1 = s\sqrt{\frac{\pi}{2}} \tag{16}$$
$$\sigma^2 = \mu_2 = \frac{4-\pi}{2}s^2 \tag{17}$$
$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{2(\pi-3)\sqrt{\pi}}{(4-\pi)^{3/2}} \tag{18}$$
$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{2(-3\pi^2+12\pi-8)}{(\pi-4)^2}. \tag{19}$$

Rayleigh Differential Equation

$$y'' - \mu(1 - \tfrac{1}{3}y'^2)y' + y = 0,$$

where $\mu > 0$. Differentiating and setting $y = y'$ gives the VAN DER POL EQUATION.

see also VAN DER POL EQUATION

Rayleigh's Formulas

The formulas

$$j_n(z) = z^n\left(-\frac{1}{z}\frac{d}{dz}\right)^n \frac{\sin z}{z}$$
$$y_n(z) = -z^n\left(-\frac{1}{z}\frac{d}{dz}\right)^n \frac{\cos z}{z}$$

for $n = 0, 1, 2, \ldots$, where $j_n(z)$ is a SPHERICAL BESSEL FUNCTION OF THE FIRST KIND and $y_n(z)$ is a SPHERICAL BESSEL FUNCTION OF THE SECOND KIND.

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 439, 1972.

Rayleigh-Ritz Variational Technique

A technique for computing EIGENFUNCTIONS and EIGENVALUES. It proceeds by requiring

$$J = \int_a^b [p(x){y_x}^2 - q(x)y^2]\,dx \tag{1}$$

to have a STATIONARY VALUE subject to the normalization condition

$$\int_a^b y^2 w(x)\,dx = 1 \tag{2}$$

and the boundary conditions

$$py_x y|_a^b = 0. \tag{3}$$

This leads to the STURM-LIOUVILLE EQUATION

$$\frac{d}{dx}\left(p\frac{dy}{dx}\right) + qy + \lambda wy = 0, \tag{4}$$

which gives the stationary values of

$$F[y(x)] = \frac{\int_a^b (p{y_x}^2 - qy^2)\,dx}{\int_a^b y^2 w\,dx} \tag{5}$$

as

$$F[y_n(x)] = \lambda_n, \tag{6}$$

where λ_n are the EIGENVALUES corresponding to the EIGENFUNCTION y_n.

References

Arfken, G. "Rayleigh-Ritz Variational Technique." §17.8 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 957–961, 1985.

Rayleigh's Theorem

see PARSEVAL'S THEOREM

Re-Entrant Circuit

A CYCLE in a GRAPH which terminates at the starting point.

see also CYCLE (GRAPH), EULERIAN CIRCUIT, HAMILTONIAN CYCLE

Real Analysis

That portion of mathematics dealing with functions of real variables. While this includes some portions of TOPOLOGY, it is most commonly used to distinguish that portion of CALCULUS dealing with real as opposed to COMPLEX NUMBERS.

Real Axis

see REAL LINE

Real Function

A FUNCTION whose RANGE is in the REAL NUMBERS is said to be a real function.

see also COMPLEX FUNCTION, SCALAR FUNCTION, VECTOR FUNCTION

Real Line

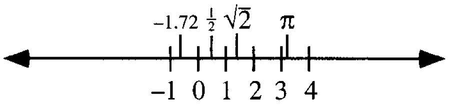

A LINE with a fixed scale so that every REAL NUMBER corresponds to a unique POINT on the LINE. The generalization of the real line to 2-D is called the COMPLEX PLANE.

see also ABSCISSA, COMPLEX PLANE

References

Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 57, 1996.

Real Matrix

A MATRIX whose elements consist entirely of REAL NUMBERS.

Real Number

The set of all RATIONAL and IRRATIONAL numbers is called the real numbers, or simply the "reals," and denoted $\mathbb{R}$. The set of real numbers is also called the CONTINUUM, denoted C.

The real numbers can be extended with the addition of the IMAGINARY NUMBER i, equal to $\sqrt{-1}$. Numbers of the form $x + iy$, where x and y are both real, are then called COMPLEX NUMBERS. Another extension which includes both the real numbers and the infinite ORDINAL NUMBERS of Georg Cantor is the SURREAL NUMBERS.

Pick two real numbers x and y at random in $(0, 1)$ with a UNIFORM DISTRIBUTION. What is the PROBABILITY P_{even} that $[x/y]$, where $[r]$ denotes NINT, the nearest INTEGER to r, is EVEN? The answer may be found as follows (Putnam Exam).

$$P\left(a < \frac{x}{y} < b\right) = \begin{cases} P(ay < x < by) \\ P\left(\frac{x}{b} < y < \frac{x}{a}\right) \end{cases}$$
$$= \begin{cases} \int_0^1 \int_{ay}^{by} dx\,dy = \frac{1}{2}(b-a) & \text{for } 0 \leq a < b < 1 \\ \int_0^1 \int_{x/b}^{x/a} dy\,dx = \frac{1}{2a} - \frac{1}{2b} & \text{for } 1 < a < b \end{cases} \tag{1}$$

$$\begin{aligned} P_{\text{even}} &= P\left(0 < \frac{x}{y} < \tfrac{1}{2}\right) + \sum_{n=1}^{\infty} P\left(2n - \tfrac{1}{2} < \frac{x}{y} < 2n + \tfrac{1}{2}\right) \\ &= \tfrac{1}{2}(\tfrac{1}{2} - 0) + \sum_{n=1}^{\infty} \left[\frac{1}{2(2n - \frac{1}{2})} - \frac{1}{2(2n + \frac{1}{2})}\right] \\ &= \frac{1}{4} + \sum_{n=1}^{\infty} \left(\frac{1}{4n-1} + \frac{1}{4n-1}\right) \\ &= \tfrac{1}{4} + (\tfrac{1}{3} - \tfrac{1}{5} + \tfrac{1}{7} - \tfrac{1}{9} + \ldots) = \tfrac{1}{4} + (1 - \tan^{-1} 1) \\ &= \frac{5}{4} - \frac{\pi}{4} = \tfrac{1}{4}(5 - \pi) \approx 46.460\%. \end{aligned} \tag{2}$$

Plouffe's "Inverse Symbolic Calculator" includes a huge database of 54 million real numbers which are algebraically related to fundamental mathematical constants and functions.

see also COMPLEX NUMBER, CONTINUUM, i, IMAGINARY NUMBER, INTEGER RELATION, RATIONAL NUMBER, REAL PART, SURREAL NUMBER

References

Plouffe, S. "Inverse Symbolic Calculator." `http://www.cecm.sfu.ca/projects/ISC/`.

Plouffe, S. "Plouffe's Inverter." `http://www.lacim.uqam.ca/pi/`.

Putnam Exam. Problem B-3 in the 54th Putnam Exam.

Real Part

The real part $\Re$ of a COMPLEX NUMBER $z = x + iy$ is the REAL NUMBER *not* multiplying i, so $\Re[x + iy] = x$. In terms of z itself,

$$\Re[z] = \tfrac{1}{2}(z + z^*),$$

where z^* is the COMPLEX CONJUGATE of z.

see also ABSOLUTE SQUARE, COMPLEX CONJUGATE, IMAGINARY PART

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 16, 1972.

Real Polynomial

A POLYNOMIAL having only REAL NUMBERS as COEFFICIENTS.

see also POLYNOMIAL

Real Projective Plane

The closed topological MANIFOLD, denoted $\mathbb{R}P^2$, which is obtained by projecting the points of a plane E from a fixed point P (not on the plane), with the addition of the LINE AT INFINITY, is called the real projective plane. There is then a one-to-one correspondence between points in E and lines through P. Since each line through P intersects the sphere $\mathbb{S}^2$ centered at P and tangent to E in two ANTIPODAL POINTS, $\mathbb{R}P^2$ can be described as a QUOTIENT SPACE of $\mathbb{S}^2$ by identifying any two such points. The real projective plane is a NONORIENTABLE SURFACE.

The BOY SURFACE, CROSS-CAP, and ROMAN SURFACE are all homeomorphic to the real projective plane and, because $\mathbb{R}P^2$ is nonorientable, these surfaces contain self-intersections (Kuiper 1961, Pinkall 1986).

see also BOY SURFACE, CROSS-CAP, NONORIENTABLE SURFACE, PROJECTIVE PLANE, ROMAN SURFACE

References

Geometry Center. "The Projective Plane." `http://www.geom.umn.edu/zoo/toptype/pplane/`.

Gray, A. "Realizations of the Real Projective Plane." §12.5 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 241–245, 1993.

Klein, F. §1.2 in *Vorlesungen über nicht-euklidische Geometrie.* Berlin, 1928.

Kuiper, N. H. "Convex Immersion of Closed Surfaces in E^3." *Comment. Math. Helv.* **35**, 85–92, 1961.

Pinkall, U. *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 64–65, 1986.

Real Quadratic Field

A QUADRATIC FIELD $\mathbb{Q}(\sqrt{D})$ with $D > 0$.

see also QUADRATIC FIELD

Realizer

A SET of R of LINEAR EXTENSIONS of a POSET $P = (X, \leq)$ is a realizer of P (and is said to realize P) provided that for all $x, y \in X$, $x \leq y$ IFF x is below y in every member of R.

see also DOMINANCE, LINEAR EXTENSION, PARTIALLY ORDERED SET, POSET DIMENSION

Rearrangement Theorem

Each row and each column in the GROUP multiplication table lists each of the GROUP elements once and only once. From this, it follows that no two elements may be in the identical location in two rows or two columns. Thus, each row and each column is a rearranged list of the GROUP elements. Stated otherwise, given a GROUP of n distinct elements $(I, a, b, c, \ldots, n)$, the set of products $(aI, a^2, ab, ac, \ldots, an)$ reproduces the n original distinct elements in a new order.

see also GROUP

Reciprocal

The reciprocal of a REAL or COMPLEX NUMBER z is its MULTIPLICATIVE INVERSE $1/z$. The reciprocal of a COMPLEX NUMBER $z = x + iy$ is given by

$$\frac{1}{x+iy} = \frac{x-iy}{x^2+y^2} = \frac{x}{x^2+y^2} - \frac{y}{x^2+y^2}i.$$

Reciprocal Difference

The reciprocal differences are closely related to the DIVIDED DIFFERENCE. The first few are explicitly given by

$$\rho(x_0, x_1) = \frac{x_0 - x_1}{f_0 - f_1} \qquad (1)$$

$$\rho_2(x_0, x_1, x_2) = \frac{x_0 - x_2}{\rho(x_0, x_1) - \rho(x_1, x_2)} + f_1 \qquad (2)$$

$$\rho_3(x_0, x_1, x_2, x_3) = \frac{x_0 - x_3}{\rho_2(x_0, x_1, x_2) - \rho_2(x_1, x_2, x_3)} + \rho(x_1, x_2) \qquad (3)$$

$$\rho_n(x_0, x_1, \ldots, x_n) = \frac{x_0 - x_n}{\rho_{n-1}(x_0, \ldots, x_{n-1}) - \rho_{n-1}(x_1, \ldots, x_n)} + \rho_{n-2}(x_1, \ldots, x_{n-1}). \qquad (4)$$

see also BACKWARD DIFFERENCE, CENTRAL DIFFERENCE, DIVIDED DIFFERENCE, FINITE DIFFERENCE, FORWARD DIFFERENCE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 878, 1972.

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 443, 1987.

Reciprocal Polyhedron

see DUAL POLYHEDRON

Reciprocating Sphere

see MIDSPHERE

Reciprocation

An incidence-preserving transformation in which points and lines are transformed into their poles and polars. A PROJECTIVE GEOMETRY-like DUALITY PRINCIPLE holds for reciprocation.

References

Coxeter, H. S. M. and Greitzer, S. L. "Reciprocation." §6.1 in *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 132–136, 1967.

Reciprocity Theorem

If there exists a RATIONAL INTEGER x such that, when n, p, and q are POSITIVE INTEGERS,

$$x^n \equiv q \pmod{p},$$

then q is the n-adic reside of p, i.e., q is an n-adic residue of p IFF $x^n \equiv q \pmod{p}$ is solvable for x.

The first case to be considered was $n = 2$ (the QUADRATIC RECIPROCITY THEOREM), of which Gauss gave the first correct proof. Gauss also solved the case $n = 3$ (CUBIC RECIPROCITY THEOREM) using INTEGERS of the form $a + b\rho$, when ρ is a root if $x^2 + x + 1 = 0$ and a, b are rational INTEGERS. Gauss stated the case $n = 4$ (QUARTIC RECIPROCITY THEOREM) using the GAUSSIAN INTEGERS.

Proof of n-adic reciprocity for PRIME n was given by Eisenstein in 1844–50 and by Kummer in 1850–61. In the 1920s, Artin formulated ARTIN'S RECIPROCITY THEOREM, a general reciprocity law for all orders.

see also ARTIN RECIPROCITY, CUBIC RECIPROCITY THEOREM, LANGLANDS RECIPROCITY, QUADRATIC RECIPROCITY THEOREM, QUARTIC RECIPROCITY THEOREM, ROOK RECIPROCITY THEOREM

Rectangle

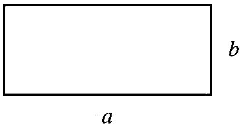

A closed planar QUADRILATERAL with opposite sides of equal lengths a and b, and with four RIGHT ANGLES. The AREA of the rectangle is

$$A = ab,$$

and its DIAGONALS are of length

$$p, q = \sqrt{a^2 + b^2}\,.$$

A SQUARE is a degenerate rectangle with $a = b$.

see also GOLDEN RECTANGLE, PERFECT RECTANGLE, SQUARE

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 122, 1987.

Eppstein, D. "Rectilinear Geometry." `http://www.ics.uci.edu/~eppstein/junkyard/rect.html`.

Rectangle Function

The rectangle function $\Pi(x)$ is a function which is 0 outside the interval $[-1, 1]$ and unity inside it. It is also called the GATE FUNCTION, PULSE FUNCTION, or WINDOW FUNCTION, and is defined by

$$\Pi(x) \equiv \begin{cases} 0 & \text{for } |x| > \frac{1}{2} \\ \frac{1}{2} & \text{for } |x| = \frac{1}{2} \\ 1 & \text{for } |x| < \frac{1}{2}. \end{cases} \tag{1}$$

The function $f(x) = h\Pi((x - c)/b)$ has height h, center c, and full-width b. Identities satisfied by the rectangle function include

$$\begin{aligned} \Pi(x) &= H(x + \tfrac{1}{2}) - H(x - \tfrac{1}{2}) && (2) \\ &= H(\tfrac{1}{2} + x) + H(\tfrac{1}{2} - x) - 1 && (3) \\ &= H(\tfrac{1}{4} - x^2) && (4) \\ &= \tfrac{1}{2}[\operatorname{sgn}(x + \tfrac{1}{2}) - \operatorname{sgn}(x - \tfrac{1}{2})], && (5) \end{aligned}$$

where $H(x)$ is the HEAVISIDE STEP FUNCTION. The FOURIER TRANSFORM of the rectangle function is given by

$$\mathcal{F}[\Pi(x)] = \int_{-\infty}^{\infty} e^{-2\pi i k x}\Pi(x)\,dx = \operatorname{sinc}(\pi k), \tag{6}$$

where $\operatorname{sinc}(x)$ is the SINC FUNCTION.

see also FOURIER TRANSFORM—RECTANGLE FUNCTION, HEAVISIDE STEP FUNCTION, RAMP FUNCTION

Rectangle Squaring

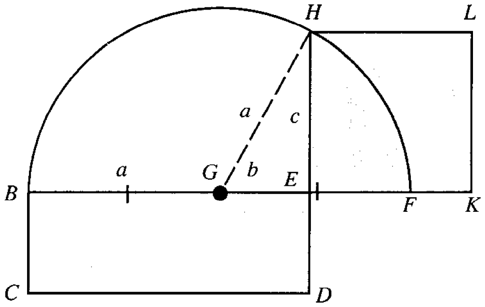

Given a RECTANGLE $\square BCDE$, draw $EF = DE$ on an extension of BE. Bisect BF and call the MIDPOINT G. Now draw a SEMICIRCLE centered at G, and construct the extension of ED which passes through the SEMICIRCLE at H. Then $\square EKLH$ has the same AREA as $\square BCDE$. This can be shown as follows:

$$\begin{aligned} A(\square BCDE) &= BE \cdot ED = BE \cdot EF \\ &= (a + b)(a - b) = a^2 - b^2 = c^2. \end{aligned}$$

References

Dunham, W. "Hippocrates' Quadrature of the Lune." Ch. 1 in *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 13–14, 1990.

Rectangular Coordinates
see CARTESIAN COORDINATES

Rectangular Distribution
see UNIFORM DISTRIBUTION

Rectangular Hyperbola

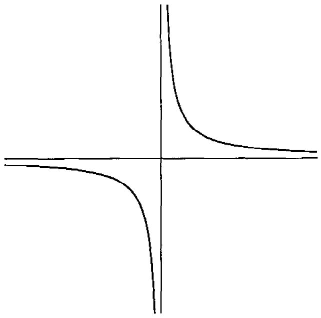

A RIGHT HYPERBOLA of the special form

$$xy = ab,$$

so that the ASYMPTOTES are the lines $x = 0$ and $y = 0$. The rectangular hyperbola is sometimes also called an EQUILATERAL HYPERBOLA.

see also HYPERBOLA, RIGHT HYPERBOLA

References
Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 76–77, 1996.

Rectangular Parallelepiped

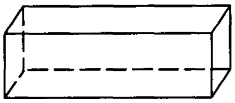

A closed box composed of 3 pairs of rectangular faces placed opposite each other and joined at RIGHT ANGLES to each other. This PARALLELEPIPED therefore corresponds to a rectangular "box." If the lengths of the sides are denoted a, b, and c, then the VOLUME is

$$V = abc, \tag{1}$$

the total SURFACE AREA is

$$A = 2(ab + bc + ca), \tag{2}$$

and the length of the "space" DIAGONAL is

$$d_{abc} = \sqrt{a^2 + b^2 + c^2}. \tag{3}$$

If $a = b = c$, then the rectangular parallelepiped is a CUBE.

see also CUBE, EULER BRICK, PARALLELEPIPED

References
Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 127, 1987.

Rectangular Projection
see EQUIRECTANGULAR PROJECTION

Rectifiable Current
The space of currents arising from rectifiable sets by integrating a differential form is called the space of 2-D rectifiable currents. For C a closed bounded rectifiable curve of a number of components in $\mathbb{R}^3$, C bounds a rectifiable current of least AREA. The theory of rectifiable currents generalizes to m-D surfaces in $\mathbb{R}^n$.

see also INTEGRAL CURRENT, REGULARITY THEOREM

References
Morgan, F. "What is a Surface?" *Amer. Math. Monthly* **103**, 369–376, 1996.

Rectifiable Set
The rectifiable sets include the image of any LIPSCHITZ FUNCTION f from planar domains into $\mathbb{R}^3$. The full set is obtained by allowing arbitrary measurable subsets of countable unions of such images of Lipschitz functions as long as the total AREA remains finite. Rectifiable sets have an "approximate" tangent plane at almost every point.

References
Morgan, F. "What is a Surface?" *Amer. Math. Monthly* **103**, 369–376, 1996.

Rectification
Rectification is the determination of the length of a curve.

see also QUADRABLE, SQUARING

Rectifying Latitude
An AUXILIARY LATITUDE which gives a sphere having correct distances along the meridians. It is denoted μ (or ω) and is given by

$$\mu = \frac{\pi M}{2M_p}. \tag{1}$$

M_p is evaluated for M at the north pole ($\phi = 90°$), and M is given by

$$\begin{aligned} M &= a(1-e^2)\int_0^\phi \frac{d\phi}{(1-e^2\sin^2\phi)^{3/2}} \\ &= a\left[\int_0^\phi \sqrt{1-e^2\sin^2\phi}\,d\phi - \frac{e^2\sin\phi\cos\phi}{\sqrt{1-e^2\sin^2\phi}}\right]. \end{aligned} \tag{2}$$

A series for M is

$$\begin{aligned} M = a[&(1 - \tfrac{1}{4}e^2 - \tfrac{3}{64}e^4 - \tfrac{5}{256}e^6 - \ldots)\phi \\ &- (\tfrac{3}{8}e^2 + \tfrac{3}{32}e^4 + \tfrac{45}{1024}e^6 + \ldots)\sin(2\phi) \\ &+ (\tfrac{15}{256}e^4 + \tfrac{45}{1024}e^6 + \ldots)\sin(4\phi) \\ &- (\tfrac{35}{3072}e^6 + \ldots)\sin(6\phi) + \ldots], \end{aligned} \tag{3}$$

and a series for μ is

$$\begin{aligned}\mu = \phi &- (\tfrac{3}{2}e_1 - \tfrac{9}{16}{e_1}^3 + \ldots)\sin(2\phi)\\ &+ (\tfrac{15}{16}{e_1}^2 - \tfrac{15}{32}{e_1}^4 + \ldots)\sin(4\phi)\\ &- (\tfrac{35}{48}{e_1}^3 - \ldots)\sin(6\phi) + (\tfrac{315}{512}{e_1}^4 - \ldots)\sin(8\phi) + \ldots,\end{aligned} \tag{4}$$

where

$$e_1 \equiv \frac{1-\sqrt{1-e^2}}{1+\sqrt{1-e^2}}. \tag{5}$$

The inverse formula is

$$\begin{aligned}\phi = \mu &+ (\tfrac{3}{2}e_1 - \tfrac{27}{32}{e_1}^3 + \ldots)\sin(2\mu)\\ &+ (\tfrac{21}{16}{e_1}^2 - \tfrac{55}{32}{e_1}^4 + \ldots)\sin(4\mu)\\ &+ (\tfrac{151}{96}{e_1}^3 - \ldots)\sin(6\mu)\\ &+ (\tfrac{1097}{512}{e_1}^4 - \ldots)\sin(8\mu) + \ldots.\end{aligned} \tag{6}$$

see also Latitude

References

Adams, O. S. "Latitude Developments Connected with Geodesy and Cartography with Tables, Including a Table for Lambert Equal-Area Meridional Projections." Spec. Pub. No. 67. U. S. Coast and Geodetic Survey, pp. 125–128, 1921.

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 16–17, 1987.

Rectifying Plane

The Plane spanned by the Tangent Vector **T** and Binormal Vector **B** .

see also Binormal Vector, Tangent Vector

Recurrence Relation

A mathematical relationship expressing f_n as some combination of f_i with $i < n$. The solutions to linear recurrence can be computed straightforwardly, but Quadratic Recurrences are not so well understood. The sequence generated by a recurrence relation is called a Recurrence Sequence. Perhaps the most famous example of a recurrence relation is the one defining the Fibonacci Numbers,

$$F_n = F_{n-2} + F_{n-1}$$

for $n \geq 3$ and with $F_1 = F_2 = 1$.

see also Argument Addition Relation, Argument Multiplication Relation, Clenshaw Recurrence Formula, Quadratic Recurrence, Recurrence Sequence, Reflection Relation, Translation Relation

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Recurrence Relations and Clenshaw's Recurrence Formula." §5.5 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 172–178, 1992.

Recurrence Sequence

A sequence of numbers generated by a Recurrence Relation is called a recurrence sequence. Perhaps the most famous recurrence sequence is the Fibonacci Numbers.

If a sequence $\{x_n\}$ with $x_1 = x_2 = 1$ is described by a two-term linear recurrence relation of the form

$$x_n = Ax_{n-1} + Bx_{n-2} \tag{1}$$

for $n \geq 3$ and A and B constants, then the closed form for x_n is given by

$$x_n = \frac{\alpha^n - \beta^n}{\alpha - \beta} \tag{2}$$

where α and β are the Roots of the Quadratic Equation

$$x^2 - Ax - B = 0, \tag{3}$$

$$\alpha = \tfrac{1}{2}(A + \sqrt{A^2 + 4B}\,) \tag{4}$$

$$\beta = \tfrac{1}{2}(A - \sqrt{A^2 + 4B}\,). \tag{5}$$

The general second-order linear recurrence

$$x_n = Ax_{n-1} + Bx_{n-2} \tag{6}$$

for constants A and B with arbitrary x_1 and x_2 has terms

$$\begin{aligned}x_1 &= x_1\\ x_2 &= x_2\\ x_3 &= Bx_1 + Ax_2\\ x_4 &= Bx_2 + ABx_1 + A^2x_2\\ x_5 &= B^2x_1 + 2ABx_2 + A^2Bx_1 + A^3x_2\\ x_6 &= B^2x_2 + 2AB^2x_1 + 3A^2Bx_2 + A^3Bx_1 + A^4x_2.\end{aligned}$$

Dropping x_1, x_2, and A, this can be written

$$\begin{array}{lllll} 1 & & & & \\ 1 & & & & \\ B & 1 & & & \\ B & B & 1 & & \\ B^2 & 2B & B & 1 & \\ B^2 & 2B^2 & 3B & B & 1, \end{array}$$

which is simply Pascal's Triangle on its side. An arbitrary term can therefore be written as

$$\begin{aligned}x_n &= \sum_{k=0}^{n-2}\binom{\lfloor\frac{1}{2}(n+k-2)\rfloor}{k}A^kB^{\lfloor(n-k-1)/2\rfloor}\\ &\qquad\times {x_1}^{[n+k \pmod 2]}{x_2}^{[n+k+1 \pmod 2]}. \qquad (7)\\ &= -(Ax_1 - x_2)\sum_{k=0}^{n-2}A^{2k-n+2}B^{-k+n-2}\binom{k}{n-k-2}\\ &\qquad + x_1\sum_{k=0}^{n-1}A^{2k-n+1}B^{-k+n-1}\binom{k}{n-k-1}. \qquad (8)\end{aligned}$$

The general linear third-order recurrence

$$x_n = Ax_{n-1} + Bx_{n-2} + Cx_{n-3} \tag{9}$$

has solution

$$\begin{aligned} x_n = x_1 & \left(\frac{\alpha^{-n}}{A + 2\alpha B + 3\alpha^2 C} + \frac{\beta^{-n}}{A + 2\beta B + 3\beta^2 C} \right. \\ & \qquad \left. + \frac{\gamma^{-n}}{A + 2\gamma B + 3\gamma^2 C} \right) \\ -(Ax_1 - x_2) & \left(\frac{\alpha^{1-n}}{A + 2\alpha B + 3\alpha^2 C} + \frac{\beta^{1-n}}{A + 2\beta B + 3\beta^2 B} \right. \\ & \qquad \left. + \frac{\gamma^{1-n}}{A + 2\gamma C + 3\gamma^2 C} \right) \\ -(Bx_1 + Ax_2 - x_3) & \left(\frac{\alpha^{2-n}}{A + 2\alpha B + 3\alpha^2 C} \right. \\ & \left. + \frac{\beta^{2-n}}{A + 2\beta B + 3\beta^2 C} + \frac{\gamma^{2-n}}{A + 2\gamma B + 3\gamma^2 C} \right), \end{aligned} \tag{10}$$

where α, β, and γ are the roots of the polynomial

$$Cx^3 + Bx^2 + Ax = 1. \tag{11}$$

A Quotient-Difference Table eventually yields a line of 0s Iff the starting sequence is defined by a linear recurrence relation.

A linear second-order recurrence

$$f_{n+1} = xf_n + yf_{n-1} \tag{12}$$

can be solved rapidly using a "rate doubling,"

$$f_{n+2} = (x^2 + 2y)f_n - y^2 f_{n-2}, \tag{13}$$

"rate tripling"

$$f_{n+3} = (x^3 + 3xy)f_n + y^3 f_{n-3}, \tag{14}$$

or in general, "rate k-tupling" formula

$$f_{n+k} = p_k f_n + q_k f_{n-k}, \tag{15}$$

where

$$p_0 = 2 \tag{16}$$

$$p_1 = x \tag{17}$$

$$p_k = 2(-y)^{k/2} T_k(x/(2i\sqrt{y})) \tag{18}$$

$$p_{k+1} = xp_k + yp_{k-1} \tag{19}$$

(here, $T_k(x)$ is a Chebyshev Polynomial of the First Kind) and

$$q_0 = -1 \tag{20}$$

$$q_1 = y \tag{21}$$

$$q_k = -(-y)^k \tag{22}$$

$$q_{k+1} = -yq_k \tag{23}$$

(Beeler *et al.* 1972, Item 14).

Let

$$s(X) = \prod_{i=1}^{m} (1 - \alpha_i X)^{n_i} = 1 - s_1 X - \ldots - s_n X^n, \tag{24}$$

where the generalized Power sum $a(h)$ for $h = 0, 1, \ldots$ is given by

$$a(h) = \sum_{i=1}^{m} A_i(h) {\alpha_i}^h, \tag{25}$$

with distinct Nonzero roots α_i, Coefficients $A_i(h)$ which are Polynomials of degree $n_i - 1$ for Positive Integers n_i, and $i \in [1, m]$. Then the sequence $\{a_h\}$ with $a_h = a(h)$ satisfies the Recurrence Relation

$$a_{h+n} = s_i a_{h+n-1} + \ldots + s_n a_h \tag{26}$$

(Meyerson and van der Poorten 1995).

The terms in a general recurrence sequence belong to a finitely generated Ring over the Integers, so it is impossible for every Rational Number to occur in any finitely generated recurrence sequence. If a recurrence sequence vanishes infinitely often, then it vanishes on an arithmetic progression with a common difference 1 that depends only on the roots. The number of values that a recurrence sequence can take on infinitely often is bounded by some Integer l that depends only on the roots. There is no recurrence sequence in which each Integer occurs infinitely often, or in which every Gaussian Integer occurs (Myerson and van der Poorten 1995).

Let $\mu(n)$ be a bound so that a nondegenerate Integer recurrence sequence of order n takes the value zero at least $\mu(n)$ times. Then $\mu(2) = 1$, $\mu(3) = 6$, and $\mu(4) \geq 9$ (Myerson and van der Poorten 1995). The maximal case for $\mu(3)$ is

$$a_{n+3} = 2a_{n+2} - 4a_{n+1} + 4a_n \tag{27}$$

with

$$a_0 = a_1 = 0 \tag{28}$$

$$a_2 = 1. \tag{29}$$

The zeros are

$$a_0 = a_1 = a_4 = a_6 = a_{13} = a_{52} = 0 \tag{30}$$

(Beukers 1991).

see also Binet Forms, Binet's Formula, Fast Fibonacci Transform, Fibonacci Sequence, Lucas Sequence, Quotient-Difference Table, Skolem-Mahler-Lerch Theorem

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Beukers, F. "The Zero-Multiplicity of Ternary Recurrences." *Composito Math.* **77**, 165–177, 1991.

Myerson, G. and van der Poorten, A. J. "Some Problems Concerning Recurrence Sequences." *Amer. Math. Monthly* **102**, 698–705, 1995.

Recurring Digital Invariant

To define a recurring digital invariant of order k, compute the sum of the kth powers of the digits of a number n. If this number n' is equal to the original number n, then $n = n'$ is called a k-Narcissistic Number. If not, compute the sums of the kth powers of the digits of n', and so on. If this process eventually leads back to the original number n, the *smallest number* in the sequence $\{n, n', n'', \ldots\}$ is said to be a k-recurring digital invariant. For example,

$$\begin{aligned} 55 &: 5^3 + 5^3 = 250 \\ 250 &: 2^3 + 5^3 + 0^3 = 133 \\ 133 &: 1^3 + 3^3 + 3^3 = 55, \end{aligned}$$

so 55 is an order 3 recurring digital invariant. The following table gives recurring digital invariants of orders 2 to 10 (Madachy 1979).

Order	RDIs	Cycle Lengths
2	4	8
3	55, 136, 160, 919	3, 2, 3, 2
4	1138, 2178	7, 2
5	244, 8294, 8299, 9044, 9045, 10933,24584, 58618, 89883	28, 10, 6, 10, 22, 4, 12, 2, 2
6	17148, 63804, 93531, 239459, 282595	30, 2, 4, 10, 3
7	80441, 86874, 253074, 376762, 922428, 982108, five more	92, 56, 27, 30, 14, 21
8	6822, 7973187, 8616804	
9	322219, 2274831, 20700388, eleven more	
10	20818070, five more	

see also 196-Algorithm, Additive Persistence, Digital Root, Digitadition, Happy Number, Kaprekar Number, Narcissistic Number, Vampire Number

References

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 163–165, 1979.

Recursion

A recursive process is one in which objects are defined in terms of other objects of the same type. Using some sort of Recurrence Relation, the entire class of objects can then be built up from a few initial values and a small number of rules. The Fibonacci Numbers are most commonly defined recursively. Care, however, must be taken to avoid Self-Recursion, in which an object is defined in terms of itself, leading to an infinite nesting.

see also Ackermann Function, Primitive Recursive Function, Recurrence Relation, Recurrence Sequence, Richardson's Theorem, Self-Recursion, Self-Similarity, TAK Function

References

Buck, R. C. "Mathematical Induction and Recursive Definitions." *Amer. Math. Monthly* **70**, 128–135, 1963.

Knuth, D. E. "Textbook Examples of Recursion." In *Artificial Intelligence and Mathematical Theory of Computation, Papers in Honor of John McCarthy* (Ed. V. Lifschitz). Boston, MA: Academic Press, pp. 207–229, 1991.

Péter, R. *Rekursive Funktionen.* Budapest: Akad. Kiado, 1951.

Recursive Function

A recursive function is a function generated by (1) Addition, (2) Multiplication, (3) selection of an element from a list, and (4) determination of the truth or falsity of the Inequality $a < b$ according to the technical rules:

1. If F and the sequence of functions $G_1, \ldots, G_n$ are recursive, then so is $F(G_1, \ldots, G_n)$.
2. If F is a recursive function such that there is an x for each a with $H(a, x) = 0$, then the smallest x can be obtained recursively.

A Turing Machine is capable of computing recursive functions.

see also Turing Machine

References

Kleene, S. C. *Introduction to Metamathematics.* Princeton, NJ: Van Nostrand, 1952.

Recursive Monotone Stable Quadrature

A Quadrature (Numerical Integration) algorithm which has a number of desirable properties.

References

Favati, P.; Lotti, G.; and Romani, F. "Interpolary Integration Formulas for Optimal Composition." *ACM Trans. Math. Software* **17**, 207–217, 1991.

Favati, P.; Lotti, G.; and Romani, F. "Algorithm 691: Improving QUADPACK Automatic Integration Routines." *ACM Trans. Math. Software* **17**, 218–232, 1991.

Red-Black Tree

An extended Binary Tree satisfying the following conditions:

1. Every node has two Children, each colored either red or black.
2. Every Leaf node is colored black.
3. Every red node has both of its Children colored black.
4. Every path from the Root to a Leaf contains the same number (the "black-height") of black nodes.

Let n be the number of internal nodes of a red-black tree. Then the number of red-black trees for $n = 1$, 2, ... is 2, 2, 3, 8, 14, 20, 35, 64, 122, ... (Sloane's A001131). The number of trees with black roots and red roots are given by Sloane's A001137 and Sloane's A001138, respectively.

Let T_h be the Generating Function for the number of red-black trees of black-height h indexed by the number of Leaves. Then

$$T_{h+1}(x) = [T_h(x)]^2 + [T_h(x)]^4, \qquad (1)$$

where $T_1(x) = x+x^2$. If $T(x)$ is the GENERATING FUNCTION for the number of red-black trees, then

$$T(x) = x + x^2 + T(x^2(1+x)^2) \qquad (2)$$

(Ruskey). Let $rb(n)$ be the number of red-black trees with n LEAVES, $r(n)$ the number of red-rooted trees, and $b(n)$ the number of black-rooted trees. All three of the quantities satisfy the RECURRENCE RELATION

$$R(n) = \sum_{n/4\leq n\leq n/2} \binom{2m}{n-2m} R(m), \qquad (3)$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT, $\text{rb}(1) = 1$, $\text{rb}(2) = 2$ for $R(n) = \text{rb}(n)$, $r(1) = r(3) = 0$, $r(2) = 1$ for $R(n) = r(n)$, and $b(1) = 1$ for $R(n) = b(n)$ (Ruskey).

References

Beyer, R. "Symmetric Binary *B*-Trees: Data Structures and Maintenance Algorithms." *Acta Informat.* **1**, 290–306, 1972.

Rivest, R. L.; Leiserson, C. E.; and Cormen, R. H. *Introduction to Algorithms.* New York: McGraw-Hill, 1990.

Ruskey, F. "Information on Red-Black Trees." `http://sue.csc.uvic.ca/~cos/inf/tree/RedBlackTree.html`.

Sloane, N. J. A. Sequences A001131, A001137, and A001138 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Red Net

The coloring red of two COMPLETE SUBGRAPHS of $n/2$ points (for EVEN n) in order to generate a BLUE-EMPTY GRAPH.

see also BLUE-EMPTY GRAPH, COMPLETE GRAPH

Reduced Amicable Pair

see QUASIAMICABLE PAIR

Reduced Fraction

A FRACTION a/b written in lowest terms, i.e., by dividing NUMERATOR and DENOMINATOR through by their GREATEST COMMON DIVISOR (a, b). For example, 2/3 is the reduced fraction of 8/12.

see also FRACTION, PROPER FRACTION

Reduced Latitude

see PARAMETRIC LATITUDE

Reducible Crossing

A crossing in a LINK projection which can be removed by rotating part of the LINK, also called REMOVABLE CROSSING.

see also ALTERNATING KNOT

Reducible Representation

see IRREDUCIBLE REPRESENTATION

Reducible Matrix

A SQUARE $n \times n$ matrix $\mathsf{A} = a_{ij}$ is called reducible if the indices 1, 2, ..., n can be divided into two disjoint nonempty sets i_1, i_2, ..., i_μ and j_1, j_2, ..., j_ν (with $\mu + \nu = n$) such that

$$a_{i_\alpha j_\beta} = 0$$

for $\alpha = 1, 2, \ldots, \mu$ and $\beta = 1, 2, \ldots, \nu$. A SQUARE MATRIX which is not reducible is said to be IRREDUCIBLE.

see also SQUARE MATRIX

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1103, 1979.

Reduction of Order

see ORDINARY DIFFERENTIAL EQUATION—SECOND-ORDER

Reduction Theorem

If a fixed point is added to each group of a special complete series, then the resulting series is complete.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 253, 1959.

Redundancy

$$R(X_1, \ldots X_n) \equiv \sum_{i=1}^n H(X_i) - H(X_1, \ldots, X_n),$$

where $H(x_i)$ is the ENTROPY and $H(X_1, \ldots, X_n)$ is the joint ENTROPY. Linear redundancy is defined as

$$L(X_1, \ldots, X_n) \equiv -\tfrac{1}{2} \sum_{i=1}^n \ln \sigma_i,$$

where σ_i are EIGENVALUES of the correlation matrix.

see also PREDICTABILITY

References

Fraser, A. M. "Reconstructing Attractors from Scalar Time Series: A Comparison of Singular System and Redundancy Criteria." *Phys. D* **34**, 391–404, 1989.

Paluš, M. "Identifying and Quantifying Chaos by Using Information-Theoretic Functionals." In *Time Series Prediction: Forecasting the Future and Understanding the Past* (Ed. A. S. Weigend and N. A. Gerschenfeld). Proc. NATO Advanced Research Workshop on Comparative Time Series Analysis held in Sante Fe, NM, May 14–17, 1992. Reading, MA: Addison-Wesley, pp. 387–413, 1994.

Reeb Foliation

The Reeb foliation of the HYPERSPHERE $\mathbb{S}^3$ is a FOLIATION constructed as the UNION of two solid TORI with common boundary.

see also FOLIATION

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 287–288, 1976.

Reef Knot

see SQUARE KNOT

Refinement

A refinement X of a COVER Y is a COVER such that every element $x \in X$ is a SUBSET of an element $y \in Y$.

see also COVER

Reflection

The operation of exchanging all points of a mathematical object with their MIRROR IMAGES (i.e., reflections in a mirror). Objects which do not change HANDEDNESS under reflection are said to be AMPHICHIRAL; those that do are said to be CHIRAL.

If the PLANE of reflection is taken as the yz-PLANE, the reflection in 2- or 3-D SPACE consists of making the transformation $x \to -x$ for each point. Consider an arbitrary point $\mathbf{x}_0$ and a PLANE specified by the equation

$$ax + by + xz + d = 0. \tag{1}$$

This PLANE has NORMAL VECTOR

$$\mathbf{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \tag{2}$$

and the POINT-PLANE DISTANCE is

$$D = \frac{|ax_0 + by_0 + cz_0 + d|}{\sqrt{a^2 + b^2 + c^2}}. \tag{3}$$

The position of the point reflected in the given plane is therefore given by

$$\begin{aligned} \mathbf{x}_0' &= \mathbf{x}_0 - 2D\hat{\mathbf{n}} \\ &= \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} - 2|ax_0 + by_0 + cz_0 + d| \begin{bmatrix} a \\ b \\ c \end{bmatrix}. \end{aligned} \tag{4}$$

see also AMPHICHIRAL, CHIRAL, DILATION, ENANTIOMER, EXPANSION, GLIDE, HANDEDNESS, IMPROPER ROTATION, INVERSION OPERATION, MIRROR IMAGE, PROJECTION, REFLECTION PROPERTY, REFLECTION RELATION, REFLEXIBLE, ROTATION, ROTOINVERSION, TRANSLATION

Reflection Property

In the plane, the reflection property can be stated as three theorems (Ogilvy 1990, pp. 73–77):

1. The LOCUS of the center of a variable CIRCLE, tangent to a fixed CIRCLE and passing through a fixed point inside that CIRCLE, is an ELLIPSE.
2. If a variable CIRCLE is tangent to a fixed CIRCLE and also passes through a fixed point outside the CIRCLE, then the LOCUS of its moving center is a HYPERBOLA.
3. If a variable CIRCLE is tangent to a fixed straight line and also passes through a fixed point not on the line, then the LOCUS of its moving center is a PARABOLA.

Let $\alpha : I \to \mathbb{R}^2$ be a smooth regular parameterized curve in $\mathbb{R}^2$ defined on an OPEN INTERVAL I, and let F_1 and F_2 be points in $\mathbb{P}^2 \backslash \alpha(I)$, where $\mathbb{P}^n$ is an n-D PROJECTIVE SPACE. Then α has a reflection property with FOCI F_1 and F_2 if, for each point $P \in \alpha(I)$,

1. Any vector normal to the curve α at P lies in the SPAN of the vectors $\overrightarrow{F_1P}$ and $\overrightarrow{F_2P}$.
2. The line normal to α at P bisects one of the pairs of opposite ANGLES formed by the intersection of the lines joining F_1 and F_2 to P.

A smooth connected plane curve has a reflection property IFF it is part of an ELLIPSE, HYPERBOLA, PARABOLA, CIRCLE, or straight LINE.

Foci	Sign	Both foci finite	One focus finite	Both foci ∞
distinct	+	confocal ellipses	confocal parabolas	|| lines
distinct	−	confocal hyperbola and ⊥ bisector of interfoci line segment	confocal parabolas	|| lines
equal		concentric circles		|| lines

Let $S \in \mathbb{R}^3$ be a smooth connected surface, and let F_1 and F_2 be points in $\mathbb{P}^3 \backslash S$, where $\mathbb{P}^n$ is an n-D PROJECTIVE SPACE. Then S has a reflection property with FOCI F_1 and F_2 if, for each point $P \in S$,

1. Any vector normal to S at P lies in the SPAN of the vectors $\overrightarrow{F_1P}$ and $\overrightarrow{F_2P}$.
2. The line normal to S at P bisects one of the pairs of opposite angles formed by the intersection of the lines joining F_1 and F_2 to P.

A smooth connected surface has a reflection property IFF it is part of an ELLIPSOID of revolution, a HYPERBOLOID of revolution, a PARABOLOID of revolution, a SPHERE, or PLANE.

Foci	Sign	Both foci finite	One focus finite	Both foci ∞
distinct	+	confocal ellipsoids	confocal paraboloids	|| planes
distinct	−	confocal hyperboloids and plane ⊥ bisector of interfoci line segment	confocal paraboloids	|| planes
equal		concentric spheres		|| planes

see also BILLIARDS

References

Drucker, D. "Euclidean Hypersurfaces with Reflective Properties." *Geometrica Dedicata* **33**, 325–329, 1990.

Drucker, D. "Reflective Euclidean Hypersurfaces." *Geometrica Dedicata* **39**, 361–362, 1991.

Drucker, D. "Reflection Properties of Curves and Surfaces." *Math. Mag.* **65**, 147–157, 1992.
Drucker, D. and Locke, P. "A Natural Classification of Curves and Surfaces with Reflection Properties." *Math. Mag.* **69**, 249–256, 1996.
Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 73–77, 1990.
Wegner, B. "Comment on 'Euclidean Hypersurfaces with Reflective Properties'." *Geometrica Dedicata* **39**, 357–359, 1991.

Reflection Relation

A mathematical relationship relating $f(-x)$ to $f(x)$.

see also ARGUMENT ADDITION RELATION, ARGUMENT MULTIPLICATION RELATION, RECURRENCE RELATION, TRANSLATION RELATION

Reflexible

An object is reflexible if it is superposable with its image in a plane mirror. Also called AMPHICHIRAL.

see also AMPHICHIRAL, CHIRAL, ENANTIOMER, HANDEDNESS, MIRROR IMAGE, REFLECTION

References

Ball, W. W. R. and Coxeter, H. S. M. "Polyhedra." Ch. 5 in *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 130, 1987.

Reflexible Map

An AUTOMORPHISM which interchanges the two vertices of a regular map at each edge without interchanging the vertices.

see also EDMONDS' MAP

Reflexive Closure

The reflexive closure of a binary RELATION R on a SET X is the minimal REFLEXIVE RELATION R' on X that contains R. Thus $aR'a$ for every element a of X and $aR'b$ for distinct elements a and b, provided that aRb.

see also REFLEXIVE REDUCTION, REFLEXIVE RELATION, RELATION, TRANSITIVE CLOSURE

Reflexive Graph

see DIRECTED GRAPH

Reflexive Reduction

The reflexive reduction of a binary RELATION R on a SET X is the minimum relation R' on X with the same REFLEXIVE CLOSURE as R. Thus $aR'b$ for any elements a and b of X, provided that a and b are distinct and aRb.

see also REFLEXIVE CLOSURE, RELATION, TRANSITIVE REDUCTION

Reflexive Relation

A RELATION R on a SET S is reflexive provided that xRx for every x in S.

see also RELATION

Reflexivity

A REFLEXIVE RELATION.

Region

An open connected set is called a region (sometimes also called a DOMAIN).

Regression

A method for fitting a curve (not necessarily a straight line) through a set of points using some goodness-of-fit criterion. The most common type of regression is LINEAR REGRESSION.

see also LEAST SQUARES FITTING, LINEAR REGRESSION, MULTIPLE REGRESSION, NONLINEAR LEAST SQUARES FITTING, REGRESSION COEFFICIENT

References

Kleinbaum, D. G. and Kupper, L. L. *Applied Regression Analysis and Other Multivariable Methods.* North Scituate, MA: Duxbury Press, 1978.

Regression Coefficient

The slope b of a line obtained using linear LEAST SQUARES FITTING is called the regression coefficient.

see also CORRELATION COEFFICIENT, LEAST SQUARES FITTING

References

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, p. 254, 1951.

Regula Falsi

see FALSE POSITION METHOD

Regular Function

see HOLOMORPHIC FUNCTION

Regular Graph

A GRAPH is said to be regular of degree r if all LOCAL DEGREES are the same number r. Then

$$E = \tfrac{1}{2}nr,$$

where E is the number of EDGES. The connected 3-regular graphs have been determined by G. Brinkman up to 24 VERTICES.

see also COMPLETE GRAPH, COMPLETELY REGULAR GRAPH, LOCAL DEGREE, SUPERREGULAR GRAPH

References

Chartrand, G. *Introductory Graph Theory.* New York: Dover, p. 29, 1985.

Regular Isotopy

The equivalence of MANIFOLDS under continuous deformation within the embedding space. KNOTS of opposite CHIRALITY have AMBIENT ISOTOPY, but not regular isotopy.

see also AMBIENT ISOTOPY

Regular Isotopy Invariant

see Bracket Polynomial

Regular Local Ring

A regular local ring is a Local Ring R with Maximal Ideal m so that m can be generated with exactly d elements where d is the Krull Dimension of the Ring R. Equivalently, R is regular if the Vector Space m/m^2 has dimension d.

see also Krull Dimension, Local Ring, Regular Ring, Ring

References

Eisenbud, D. *Commutative Algebra with a View Toward Algebraic Geometry.* New York: Springer-Verlag, p. 242, 1995.

Regular Number

A number which has a finite Decimal expansion. A number which is not regular is said to be nonregular.

see also Decimal Expansion, Repeating Decimal

Regular Parameterization

A parameterization of a Surface $\mathbf{x}(u,v)$ in u and v is regular if the Tangent Vectors

$$\frac{\partial \mathbf{x}}{\partial u} \quad \text{and} \quad \frac{\partial \mathbf{x}}{\partial v}$$

are always Linearly Independent.

Regular Patch

A regular patch is a Patch $\mathbf{x} : U \to \mathbb{R}^n$ for which the Jacobian $J(\mathbf{x})(u,v)$ has rank 2 for all $(u,v) \in U$. A Patch is said to be regular at a point $(u_0, v_0) \in U$ providing that its Jacobian has rank 2 at (u_0, v_0). For example, the points at $\phi = \pm\pi/2$ in the standard parameterization of the Sphere $(\cos\theta\sin\phi, \sin\theta\sin\phi, \cos\phi)$ are not regular.

An example of a Patch which is regular but not Injective is the Cylinder defined parametrically by $(\cos u, \sin u, v)$ with $u \in (-\infty, \infty)$ and $v \in (-2, 2)$. However, if $\mathbf{x} : U \to \mathbb{R}^n$ is an injective regular patch, then $\mathbf{x}$ maps U diffeomorphically onto $\mathbf{x}(U)$.

see also Injective Patch, Patch, Regular Surface

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 187, 1993.

Regular Point

see Ordinary Point

Regular Polygon

An n-sided Polygon in which the sides are all the same length and are symmetrically placed about a common center. The sum of Perpendiculars from any point to the sides of a regular polygon of n sides is n times the Apothem. Only certain regular polygons are "Constructible" with Ruler and Straightedge.

n	Regular Polygon
3	equilateral triangle
4	square
5	pentagon
6	hexagon
7	heptagon
8	octagon
9	nonagon
10	decagon
12	dodecagon
15	pentadecagon
16	hexadecagon
17	heptadecagon
18	octadecagon
20	icosagon
30	triacontagon

see also Constructible Polygon, Geometrography, Heptadecagon, Polygon

References

Bishop, W. "How to Construct a Regular Polygon." *Amer. Math. Monthly* **85**, 186–188, 1978.

Regular Polyhedron

A polyhedron is said to be regular if its Faces and Vertex Figures are Regular (not necessarily Convex) polygons (Coxeter 1973, p. 16). Using this definition, there are a total of nine regular polyhedra, five being the Convex Platonic Solids and four being the Concave (stellated) Kepler-Poinsot Solids. However, the term "regular polyhedra" is sometimes used to refer exclusively to the Convex Platonic Solids.

It can be proven that only nine regular solids (in the Coxeter sense) exist by noting that a possible regular polyhedron must satisfy

$$\cos^2\left(\frac{\pi}{p}\right) + \cos^2\left(\frac{\pi}{q}\right) + \cos^2\left(\frac{\pi}{r}\right) = 1.$$

Gordon showed that the only solutions to

$$1 + \cos\phi_1 + \cos\phi_2 + \cos\phi_3 = 0$$

of the form $\phi_i = \pi m_i/n_i$ are the permutations of $(\frac{2}{3}\pi, \frac{2}{3}\pi, \frac{1}{3}\pi)$ and $(\frac{2}{3}\pi, \frac{2}{5}\pi, \frac{4}{5}\pi)$. This gives three permutations of (3, 3, 4) and six of $(3,\ 5,\ \frac{5}{3})$ as possible solutions to the first equation. Plugging back in gives the Schläfli Symbols of possible regular polyhedra as $\{3,3\}$, $\{3,4\}$, $\{4,3\}$, $\{3,5\}$, $\{5,3\}$, $\{3,\frac{5}{2}\}$, $\{\frac{5}{2},3\}$, $\{5,\frac{5}{2}\}$, and $\{\frac{5}{2},5\}$ (Coxeter 1973, pp. 107–109). The first five of these are the PLATONIC SOLIDS and the remaining four the KEPLER-POINSOT SOLIDS.

Every regular polyhedron has $e + 1$ axes of symmetry, where e is the number of EDGES, and $3h/2$ PLANES of symmetry, where h is the number of sides of the corresponding PETRIE POLYGON.

see also CONVEX POLYHEDRON, KEPLER-POINSOT SOLID, PETRIE POLYGON, PLATONIC SOLID, POLYHEDRON, POLYHEDRON COMPOUND, SPONGE, VERTEX FIGURE

References
Coxeter, H. S. M. "Regular and Semi-Regular Polytopes I." *Math. Z.* **46**, 380–407, 1940.
Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, pp. 1–17, 93, and 107–112, 1973.
Cromwell, P. R. *Polyhedra.* New York: Cambridge University Press, pp. 85–86, 1997.

Regular Prime

A PRIME which does not DIVIDE the CLASS NUMBER $h(p)$ of the CYCLOTOMIC FIELD obtained by adjoining a PRIMITIVE pTH ROOT of unity to the rational FIELD. A PRIME p is regular IFF p does not divide the NUMERATORS of the BERNOULLI NUMBERS B_{10}, B_{12}, ..., B_{2p-2}. A PRIME which is not regular is said to be an IRREGULAR PRIME.

In 1915, Jensen proved that there are infinitely many IRREGULAR PRIMES. It has not yet been proven that there are an INFINITE number of regular primes (Guy 1994, p. 145). Of the 283,145 PRIMES $< 4\times10^6$, 171,548 (or 60.59%) are regular (the conjectured FRACTION is $e^{-1/2} \approx 60.65\%$). The first few are 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 41, 43, 47, ... (Sloane's A007703).

see also BERNOULLI NUMBER, FERMAT'S THEOREM, IRREGULAR PRIME

References
Buhler, J.; Crandall, R. Ernvall, R.; and Metsankyla, T. "Irregular Primes and Cyclotomic Invariants to Four Million." *Math. Comput.* **61**, 151–153, 1993.
Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 145, 1994.
Ribenboim, P. "Regular Primes." §5.1 in *The New Book of Prime Number Records.* New York: Springer-Verlag, pp. 323–329, 1996.
Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 153, 1993.
Sloane, N. J. A. Sequence A007703/M2411 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Regular Ring

In the sense of von Neumann, a regular ring is a RING R such that for all $a \in R$, there exists a $b \in R$ satisfying $a = aba$.

see also REGULAR LOCAL RING, RING

References
Jacobson, N. *Basic Algebra II, 2nd ed.* New York: W. H. Freeman, p. 196, 1989.

Regular Sequence

Let there be two PARTICULARLY WELL-BEHAVED FUNCTIONS $F(x)$ and $p_\tau(x)$. If the limit

$$\lim_{\tau\to0}\int_{-\infty}^{\infty} p_\tau(x)F(x)\,dx$$

exists, then $p_\tau(x)$ is a regular sequence of PARTICULARLY WELL-BEHAVED FUNCTIONS.

Regular Singular Point

Consider a second-order ORDINARY DIFFERENTIAL EQUATION

$$y'' + P(x)y' + Q(x)y = 0.$$

If $P(x)$ and $Q(x)$ remain FINITE at $x = x_0$, then x_0 is called an ORDINARY POINT. If either $P(x)$ or $Q(x)$ diverges as $x \to x_0$, then x_0 is called a singular point. If either $P(x)$ or $Q(x)$ diverges as $x \to x_0$ but $(x-x_0)P(x)$ and $(x-x_0)^2Q(x)$ remain FINITE as $x \to x_0$, then $x = x_0$ is called a regular singular point (or NONESSENTIAL SINGULARITY).

see also IRREGULAR SINGULARITY, SINGULAR POINT (DIFFERENTIAL EQUATION)

References
Arfken, G. "Singular Points." §8.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 451–453 and 461–463, 1985.

Regular Singularity

see REGULAR SINGULAR POINT

Regular Surface

A SUBSET $M \subset \mathbb{R}^n$ is called a regular surface if for each point $p \in M$, there exists a NEIGHBORHOOD V of p in $\mathbb{R}^n$ and a MAP $x : U \to \mathbb{R}^n$ of a OPEN SET $U \subset \mathbb{R}^2$ onto $V \cap M$ such that

1. x is differentiable,
2. $x : U \to V \cap M$ is a HOMEOMORPHISM,
3. Each map $x : U \to M$ is a REGULAR PATCH.

Any open subset of a regular surface is also a regular surface.

see also REGULAR PATCH

References
Gray, A. "The Definition of a Regular Surface in $\mathbb{R}^n$." §10.4 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 195–200, 1993.

Regular Triangle Center

A Triangle Center is regular Iff there is a Triangle Center Function which is a Polynomial in Δ, a, b, and c (where Δ is the Area of the Triangle) such that the Trilinear Coordinates of the center are

$$f(a,b,c) : f(b,c,a) : f(c,a,b).$$

The Isogonal Conjugate of a regular center is a regular center. Furthermore, given two regular centers, any two of their Harmonic Conjugate Points are also regular centers.

see also Isogonal Conjugate, Triangle Center, Triangle Center Function

Regularity Theorem

An Area-minimizing surface (Rectifiable Current) bounded by a smooth curve in $\mathbb{R}^3$ is a smooth submanifold with boundary.

see also Minimal Surface, Rectifiable Current

References

Morgan, F. "What is a Surface?" *Amer. Math. Monthly* **103**, 369–376, 1996.

Regularized Beta Function

The regularized beta function is defined by

$$I(z;a,b) = \frac{B(z;a,b)}{B(a,b)},$$

where $B(z;a,b)$ is the incomplete Beta Function and $B(a,b)$ is the complete Beta Function.

see also Beta Function, Regularized Gamma Function

Regularized Gamma Function

The regularized gamma functions are defined by

$$P(a,z) = 1 - Q(a,z) \equiv \frac{\gamma(a,z)}{\Gamma(a)}$$

and

$$Q(a,z) = 1 - P(a,z) \equiv \frac{\Gamma(a,z)}{\Gamma(a)},$$

where $\gamma(a,z)$ and $\Gamma(a,z)$ are incomplete Gamma Functions and $\Gamma(a)$ is a complete Gamma Function. Their derivatives are

$$\frac{d}{dz}P(a,z) = e^{-z}z^{a-1}$$
$$\frac{d}{dz}Q(a,z) = -e^{-z}z^{a-1}.$$

see also Gamma Function, Regularized Beta Function

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 160–161, 1992.

Regulus

The locus of lines meeting three given Skew Lines. ("Regulus" is also the name of the brightest star in the constellation Leo.)

Reidemeister Moves

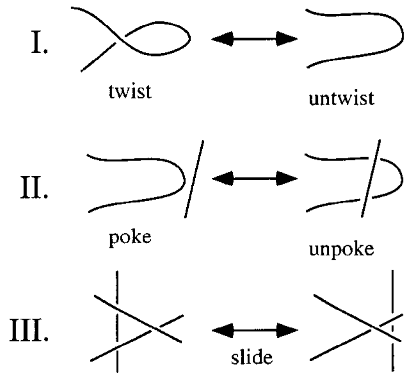

In the 1930s, Reidemeister first rigorously proved that Knots exist which are distinct from the Unknot. He did this by showing that all Knot deformations can be reduced to a sequence of three types of "moves," called the (I) Twist Move, (II) Poke Move, and (III) Slide Move.

Reidemeister's Theorem guarantees that moves I, II, and III correspond to Ambient Isotopy (moves II and III alone correspond to Regular Isotopy). He then defined the concept of Colorability, which is invariant under Reidemeister moves.

see also Ambient Isotopy, Colorable, Markov Moves, Regular Isotopy, Unknot

Reidemeister's Theorem

Two Links can be continuously deformed into each other Iff any diagram of one can be transformed into a diagram of the other by a sequence of Reidemeister Moves.

see also Reidemeister Moves

Reinhardt Domain

A Reinhardt domain with center **c** is a Domain D in C^n such that whenever D contains z_0, the Domain D also contains the closed Polydisk.

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 101, 1980.

Relation

A relation is any Subset of a Cartesian Product. For instance, a Subset of $A \times B$, called a (binary) "relation from A to B," is a collection of Ordered Pairs (a,b) with first components from A and second components from B, and, in particular, a Subset of $A \times A$ is called a "relation on A." For a binary relation R, one often writes aRb to mean that (a,b) is in R.

see also ADJACENCY RELATION, ANTISYMMETRIC RELATION, ARGUMENT ADDITION RELATION, ARGUMENT MULTIPLICATION RELATION, COVER RELATION, EQUIVALENCE RELATION, IRREFLEXIVE, PARTIAL ORDER, RECURRENCE RELATION, REFLECTION RELATION, REFLEXIVE RELATION, SYMMETRIC RELATION, TRANSITIVE, TRANSLATION RELATION

Relative Error

Let the true value of a quantity be x and the measured or inferred value x_0. Then the relative error is defined by

$$\delta x = \frac{\Delta x}{x} = \frac{x_0 - x}{x} = \frac{x_0}{x} - 1,$$

where Δx is the ABSOLUTE ERROR. The relative error of the QUOTIENT or PRODUCT of a number of quantities is less than or equal to the SUM of their relative errors. The PERCENTAGE ERROR is 100% times the relative error.

see also ABSOLUTE ERROR, ERROR PROPAGATION, PERCENTAGE ERROR

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 14, 1972.

Relative Extremum

A RELATIVE MAXIMUM or RELATIVE MINIMUM, also called a LOCAL EXTREMUM.

see also EXTREMUM, GLOBAL EXTREMUM, RELATIVE MAXIMUM, RELATIVE MINIMUM

Relative Maximum

A MAXIMUM within some NEIGHBORHOOD which need not be a GLOBAL MAXIMUM.

see also GLOBAL MAXIMUM, MAXIMUM, RELATIVE MINIMUM

Relative Minimum

A MINIMUM within some NEIGHBORHOOD which need not be a GLOBAL MINIMUM.

see also GLOBAL MINIMUM, MINIMUM, RELATIVE MAXIMUM

Relatively Prime

Two integers are relatively prime if they share no common factors (divisors) except 1. Using the notation (m, n) to denote the GREATEST COMMON DIVISOR, two integers m and n are relatively prime if $(m, n) = 1$. Relatively prime integers are sometimes also called STRANGERS or COPRIME and are denoted $m \perp n$.

The probability that two INTEGERS picked at random are relatively prime is $[\zeta(2)]^{-1} = 6/\pi^2$, where $\zeta(z)$ is the RIEMANN ZETA FUNCTION. This result is related to the fact that the GREATEST COMMON DIVISOR of m and n, $(m, n) = k$, can be interpreted as the number of LATTICE POINTS in the PLANE which lie on the straight LINE connecting the VECTORS $(0, 0)$ and (m, n) (excluding (m, n) itself). In fact $6/\pi^2$ the fractional number of LATTICE POINTS VISIBLE from the ORIGIN (Castellanos 1988, pp. 155–156).

Given three INTEGERS chosen at random, the probability that no common factor will divide them all is

$$[\zeta(3)]^{-1} \approx 1.202^{-1} = 0.832\ldots,$$

where $\zeta(3)$ is APÉRY'S CONSTANT. This generalizes to k random integers (Schoenfeld 1976).

see also DIVISOR, GREATEST COMMON DIVISOR, VISIBILITY

References

Castellanos, D. "The Ubiquitous Pi." *Math. Mag.* **61**, 67–98, 1988.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 3–4, 1994.

Schoenfeld, L. "Sharper Bounds for the Chebyshev Functions $\theta(x)$ and $\psi(x)$, II." *Math. Comput.* **30**, 337–360, 1976.

Relaxation Methods

Methods of solving an ORDINARY DIFFERENTIAL EQUATION by replacing it with a FINITE DIFFERENCE equation on a regular grid spanning the domain of interest. The finite difference equations are then solved using an n-D NEWTON'S METHOD or other similar algorithm.

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Richardson Extrapolation and the Bulirsch-Stoer Method." §17.3 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 753–763, 1992.

Remainder

In general, a remainder is a quantity "left over" after performing a particular algorithm. The term is most commonly used to refer to the number left over when two integers are divided by each other in INTEGER DIVISION. For example, $55\backslash 7 = 7$, with a remainder of 6. Of course in real division, there is no such thing as a remainder since, for example, $55/7 = 7 + 6/7$.

The term remainder is also sometimes to the RESIDUE of a CONGRUENCE.

see also DIVISION, INTEGER DIVISION, RESIDUE (CONGRUENCE)

Remainder Theorem

see POLYNOMIAL REMAINDER THEOREM

Rembs' Surfaces

A special class of ENNEPER'S SURFACES which can be given parametrically by

$$x = a(U \cos u - U' \sin u) \tag{1}$$
$$y = -a(U \sin u + U' \cos u) \tag{2}$$
$$z = v - aV', \tag{3}$$

where

$$U \equiv \frac{\cosh(u\sqrt{C})}{\sqrt{C}} \tag{4}$$
$$V \equiv \frac{\cos(v\sqrt{C+1})}{\sqrt{C+1}} \tag{5}$$
$$a \equiv \frac{2V}{(C+1)(U^2 - V^2)}. \tag{6}$$

The value of v is restricted to

$$|v| \leq v_0 \equiv \frac{\pi}{2\sqrt{C+1}} \tag{7}$$

(Reckziegel 1986), and the values $v = \pm v_0$ correspond to the ends of the cleft in the surface.

see also ENNEPER'S SURFACES, KUEN SURFACE, SIEVERT'S SURFACE

References

Fischer, G. (Ed.). Plate 88 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 84, 1986.

Reckziegel, H. "Sievert's Surface." §3.4.4.3 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 39–40, 1986.

Rembs, E. "Enneper'sche Flächen konstanter positiver Krümmung und Hazzidakissche Transformationen." *Jahrber. DMV* **39**, 278–283, 1930.

Removable Crossing

see REDUCIBLE CROSSING

Removable Singularity

A SINGULAR POINT z_0 of a FUNCTION $f(z)$ for which it is possible to assign a COMPLEX NUMBER in such a way that $f(z)$ becomes ANALYTIC. A more precise way of defining a removable singularity is as a SINGULARITY z_0 of a function $f(z)$ about which the function $f(z)$ is bounded. For example, the point $x_0 = 0$ is a removable singularity in the SINC FUNCTION $\operatorname{sinc} x = \sin x/x$, since this function satisfies $\operatorname{sinc} 0 = 1$.

Rencontres Number

see DERANGEMENT, SUBFACTORIAL

Rendezvous Values

see MAGIC GEOMETRIC CONSTANTS

Rényi's Parking Constants

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Given the CLOSED INTERVAL $[0, x]$ with $x > 1$, let 1-D "cars" of unit length be parked randomly on the interval. The MEAN number $M(x)$ of cars which can fit (without overlapping!) satisfies

$$M(x) = \begin{cases} 0 & \text{for } 0 \leq x < 1 \\ 1 + \frac{2}{x-1}\int_0^{x-1} M(y)\,dy & \text{for } x \geq 1. \end{cases} \tag{1}$$

The mean density of the cars for large x is

$$m \equiv \lim_{x\to\infty} \frac{M(x)}{x} = \int_0^\infty \exp\left(-2\int_0^x \frac{1-e^{-y}}{y}\,dy\right)dx$$
$$= 0.7475979203\ldots. \tag{2}$$

Furthermore,

$$M(x) = mx + m - 1 + \mathcal{O}(x^{-n}) \tag{3}$$

for all n (Rényi 1958), which was strengthened by Dvoretzky and Robbins (1964) to

$$M(x) = mx + m - 1 + \mathcal{O}\left[\left(\frac{2e}{x}\right)^{x-3/2}\right]. \tag{4}$$

Dvoretzky and Robbins (1964) also proved that

$$\inf_{x\leq t\leq x+1} \frac{M(t)+1}{t+1} \leq m \leq \sup_{x\leq t\leq x+1} \frac{M(t)+1}{t+1}. \tag{5}$$

Let $V(x)$ be the variance of the number of cars, then Dvoretzky and Robbins (1964) and Mannion (1964) showed that

$$v \equiv \lim_{x\to\infty} \frac{V(x)}{x}$$
$$= 2\int_0^\infty \left\{x\int_0^1 e^{-xy}R_2(y)\,dy + x^2\left[\int_0^\infty e^{-xy}R_1(y)\,dy\right]^2\right\}$$
$$\times \exp\left(-2\int_0^x \frac{1-e^{-y}}{y}\,dy\right)dx = 0.038156\ldots, \tag{6}$$

where

$$R_1(x) = M(x) - mx - m + 1 \tag{7}$$

$$R_2(x) = \begin{cases} (1-m-mx)^2 & \text{for } 0 \leq x \leq 1 \\ 4(1-m)^2 & \text{for } x = 1 \\ \frac{2}{x-1}\left[\int_0^{x-1} R_2(y)\,dy\right. & \text{for } x > 1 \\ \quad \left.+ \int_0^{x-1} R_1(y)R_1(x-y-1)\,dy\right], & \end{cases} \tag{8}$$

and the numerical value is due to Blaisdell and Solomon (1970). Dvoretzky and Robbins (1964) also proved that

$$\inf_{x\le t\le x+1}\frac{V(t)}{t+1}\le v\le \sup_{x\le t\le x+1}\frac{V(t)}{t+1}, \tag{9}$$

and that

$$V(x)=vx+v+\mathcal{O}\left[\left(\frac{4e}{x}\right)^{x-4}\right]. \tag{10}$$

Palasti (1960) conjectured that in 2-D,

$$\lim_{x,y\to\infty}\frac{M(x,y)}{xy}=m^2, \tag{11}$$

but this has not yet been proven or disproven (Finch).

References
Blaisdell, B. E. and Solomon, H. "On Random Sequential Packing in the Plane and a Conjecture of Palasti." *J. Appl. Prob.* **7**, 667–698, 1970.
Dvoretzky, A. and Robbins, H. "On the Parking Problem." *Publ. Math. Inst. Hung. Acad. Sci.* **9**, 209–224, 1964.
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/renyi/renyi.html`.
Mannion, D. "Random Space-Filling in One Dimension." *Publ. Math. Inst. Hung. Acad. Sci.* **9**, 143–154, 1964.
Palasti, I. "On Some Random Space Filling Problems." *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 353–359, 1960.
Rényi, A. "On a One-Dimensional Problem Concerning Random Space-Filling." *Publ. Math. Inst. Hung. Acad. Sci.* **3**, 109–127, 1958.
Solomon, H. and Weiner, H. J. "A Review of the Packing Problem." *Comm. Statist. Th. Meth.* **15**, 2571–2607, 1986.

Rep-Tile

A Polygon which can be divided into smaller copies of itself.

see also Dissection

References
Gardner, M. Ch. 19 in *The Unexpected Hanging and Other Mathematical Diversions.* Chicago, IL: Chicago University Press, 1991.

Repartition

see Adéle

Repdigit

A number composed of a single digit is called a repdigit. If the digits are all 1s, the repdigit is called a Repunit. The Beast Number 666 is a repdigit.

see also Keith Number, Repunit

Repeating Decimal

A number whose decimal representation eventually becomes periodic (i.e., the same sequence of digits repeats indefinitely) is called a repeating decimal. Numbers such as 0.5 can be regarded as repeating decimals since $0.5 = 0.5000\ldots = 0.4999\ldots$. All Rational Numbers have repeating decimals, e.g., $1/11 = 0.\overline{09}$. However, Transcendental Numbers, such as $\pi = 3.141592\ldots$ do not.

see also Cyclic Number, Decimal Expansion, Full Reptend Prime, Irrational Number, Midy's Theorem, Rational Number, Regular Number

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 53–54, 1987.
Courant, R. and Robbins, H. "Rational Numbers and Periodic Decimals." §2.2.4 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 66–68, 1996.

Repfigit Number

see Keith Number

Replicate

One out of a set of identical observations in a given experiment under identical conditions.

Reptend Prime

see Full Reptend Prime

Representation

The representation of a Group G on a Complex Vector Space V is a group action of G on V by linear transformations. Two finite dimensional representations π on V and π' on V' are equivalent if there is an invertible linear map $E : V \mapsto V'$ such that $\pi'(g)E = E\pi(g)$ for all $g \in G$. π is said to be irreducible if it has no proper Nonzero invariant Subspaces.

see also Character (Multiplicative), Peter-Weyl Theorem, Primary Representation, Schur's Lemma

References
Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Repunit

A (generalized) repunit to the base b is a number of the form

$$M_n^b = \frac{b^n-1}{b-1}.$$

The term "repunit" was coined by Beiler (1966), who also gave the first tabulation of known factors. Repunits $M_n = M_n^2 = 2^n - 1$ with $b = 2$ are called Mersenne

NUMBERS. If $b = 10$, the number is called a repunit (since the digits are all 1s). A number of the form

$$R_n = \frac{10^n - 1}{10 - 1} = R_n = \frac{10^n - 1}{9}$$

is therefore a (decimal) repunit of order n.

b	Sloane	b-Repunits
2	000225	1, 3, 7, 15, 31, 63, 127, ...
3	003462	1, 4, 13, 40, 121, 364, ...
4	002450	1, 5, 21, 85, 341, 1365, ...
5	003463	1, 6, 31, 156, 781, 3906, ...
6	003464	1, 7, 43, 259, 1555, 9331, ...
7	023000	1, 8, 57, 400, 2801, 19608, ...
8	023001	1, 9, 73, 585, 4681, 37449, ...
9	002452	1, 10, 91, 820, 7381, 66430, ...
10	002275	1, 11, 111, 1111, 11111, ...
11	016123	1, 12, 133, 1464, 16105, 177156, ...
12	016125	1, 13, 157, 1885, 22621, 271453, ...

Williams and Seah (1979) factored generalized repunits for $3 \leq b \leq 12$ and $2 \leq n \leq 1000$. A (base-10) repunit can be PRIME only if n is PRIME, since otherwise $10^{ab} - 1$ is a BINOMIAL NUMBER which can be factored algebraically. In fact, if $n = 2a$ is EVEN, then $10^{2a} - 1 = (10^a - 1)(10^a + 1)$. The only base-10 repunit PRIMES R_n for $n \leq 16,500$ are $n = 2$, 19, 23, 317, and 1031 (Sloane's A004023; Madachy 1979, Williams and Dubner 1986, Ball and Coxeter 1987). The number of factors for the base-10 repunits for $n = 1, 2, \ldots$ are 1, 1, 2, 2, 2, 5, 2, 4, 4, 4, 2, 7, 3, ... (Sloane's A046053).

b	Sloane	n of Prime b-Repunits
2	000043	2, 3, 5, 7, 13, 17, 19, 31, 61, 89, 107, ...
3	028491	3, 7, 13, 71, 103, 541, 1091, 1367, ...
5	004061	3, 7, 11, 13, 47, 127, 149, 181, 619, ...
6	004062	2, 3, 7, 29, 71, 127, 271, 509, 1049, ...
7	004063	5, 13, 131, 149, 1699, ...
10	004023	2, 19, 23, 317, 1031, ...
11	005808	17, 19, 73, 139, 907, 1907, 2029, 4801, ...
12	004064	2, 3, 5, 19, 97, 109, 317, 353, 701, ...

A table of the factors not obtainable algebraically for generalized repunits (a continuously updated version of Brillhart *et al.* 1988) is maintained online. These tables include factors for $10^n - 1$ (with $n \leq 209$ odd) and $10^n + 1$ (for $n \leq 210$ EVEN and ODD) in the files `ftp://sable.ox.ac.uk/pub/math/cunningham/10-` and `ftp://sable.ox.ac.uk/pub/math/cunningham/10+`. After algebraically factoring R_n, these are sufficient for complete factorizations. Yates (1982) published all the repunit factors for $n \leq 1000$, a portion of which are reproduced in the *Mathematica*® notebook by Weisstein.

A SMITH NUMBER can be constructed from every factored repunit.

see also CUNNINGHAM NUMBER, FERMAT NUMBER, MERSENNE NUMBER, REPDIGIT, SMITH NUMBER

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 66, 1987.

Beiler, A. H. "11111...111." Ch. 11 in *Recreations in the Theory of Numbers: The Queen of Mathematics Entertains.* New York: Dover, 1966.

Brillhart, J.; Lehmer, D. H.; Selfridge, J.; Wagstaff, S. S. Jr.; and Tuckerman, B. *Factorizations of $b^n \pm 1$, $b = 2, 3, 5, 6, 7, 10, 11, 12$ Up to High Powers, rev. ed.* Providence, RI: Amer. Math. Soc., 1988. Updates are available electronically from `ftp://sable.ox.ac.uk/pub/math/cunningham`.

Dubner, H. "Generalized Repunit Primes." *Math. Comput.* **61**, 927–930, 1993.

Guy, R. K. "Mersenne Primes. Repunits. Fermat Numbers. Primes of Shape $k \cdot 2^n + 2$." §A3 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 8–13, 1994.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 152–153, 1979.

Ribenboim, P. "Repunits and Similar Numbers." §5.5 in *The New Book of Prime Number Records.* New York: Springer-Verlag, pp. 350–354, 1996.

Snyder, W. M. "Factoring Repunits." *Am. Math. Monthly* **89**, 462–466, 1982.

Sorli, R. "Factorization Tables." `http://www.maths.uts.edu.au/staff/ron/fact/fact.html`.

Weisstein, E. W. "Repunits." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Repunit.m`.

Williams, H. C. and Dubner, H. "The Primality of $R1031$." *Math. Comput.* **47**, 703–711, 1986.

Williams, H. C. and Seah, E. "Some Primes of the Form $(a^n - 1)/(a - 1)$. *Math. Comput.* **33**, 1337–1342, 1979.

Yates, S. "Prime Divisors of Repunits." *J. Recr. Math.* **8**, 33–38, 1975.

Yates, S. "The Mystique of Repunits." *Math. Mag.* **51**, 22–28, 1978.

Yates, S. *Repunits and Reptends.* Delray Beach, FL: S. Yates, 1982.

Residual

The residual is the sum of deviations from a best-fit curve of arbitrary form.

$$R \equiv \sum [y_i - f(x_i, a_1, \ldots, a_n)]^2.$$

The residual should not be confused with the CORRELATION COEFFICIENT.

Residual vs. Predictor Plot

A plot of y_i vs. $e_i \equiv \hat{y}_i - y_i$. Random scatter indicates the model is probably good. A pattern indicates a problem with the model. If the spread in e_i increases as y_i increases, the errors are called HETEROSCEDASTIC.

Residue Class

The residue classes of a function $f(x)$ mod n are all possible values of the RESIDUE $f(x)$ (mod n). For example, the residue classes of x^2 (mod 6) are $\{0, 1, 3, 4\}$, since

$$\begin{aligned} 0^2 &\equiv 0 \pmod 6 \\ 1^2 &\equiv 1 \pmod 6 \\ 3^2 &\equiv 3 \pmod 6 \\ 4^2 &\equiv 4 \pmod 6 \end{aligned}$$

are all the possible residues. A COMPLETE RESIDUE SYSTEM is a set of integers containing one element from each class, so in this case, $\{0, 1, 9, 4\}$ would be a COMPLETE RESIDUE SYSTEM.

The $\phi(m)$ residue classes prime to m form a GROUP under the binary multiplication operation (mod m), where $\phi(m)$ is the TOTIENT FUNCTION (Shanks 1993) and the GROUP is classed a MODULO MULTIPLICATION GROUP.

see also COMPLETE RESIDUE SYSTEM, CONGRUENCE, CUBIC NUMBER, QUADRATIC RECIPROCITY THEOREM, QUADRATIC RESIDUE, RESIDUE (CONGRUENCE), SQUARE NUMBER

References

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 56 and 59–63, 1993.

Residue (Complex Analysis)

The constant a_{-1} in the LAURENT SERIES

$$f(z) = \sum_{n=-\infty}^{\infty} a_n(z - z_0)^n$$

of $f(z)$ is called the residue of $f(z)$. The residue is a very important property of a complex function and appears in the amazing RESIDUE THEOREM of CONTOUR INTEGRATION.

see also CONTOUR INTEGRATION, LAURENT SERIES, RESIDUE THEOREM

References

Arfken, G. "Calculus of Residues." §7.2 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 400–421, 1985.

Residue (Congruence)

The number b in the CONGRUENCE $a \equiv b \pmod m$ is called the residue of a (mod m). The residue of large numbers can be computed quickly using CONGRUENCES. For example, to find 37^{13} (mod 17), note that

$$\begin{aligned} 37 &\equiv 3 \\ 37^2 &\equiv 3^2 \equiv 9 \equiv -8 \\ 37^4 &\equiv 81 \equiv -4 \\ 37^8 &\equiv 16 \equiv -1, \end{aligned}$$

so

$$37^{13} \equiv 37^{1+4+8} \equiv 3(-4)(-1) \equiv 12 \pmod{17}.$$

see also COMMON RESIDUE, CONGRUENCE, MINIMAL RESIDUE

References

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 55–56, 1993.

Residue Index

$p - 1$ divided by the HAUPT-EXPONENT of a base b mod p for a given PRIME p.

see also HAUPT-EXPONENT

Residue Theorem (Complex Analysis)

Given a complex function $f(z)$, consider the LAURENT SERIES

$$f(z) = \sum_{n=-\infty}^{\infty} a_n(z - z_0)^n. \tag{1}$$

Integrate term by term using a closed contour γ encircling z_0,

$$\begin{aligned} \int_\gamma f(z)\,dz &= \sum_{n=-\infty}^{\infty} a_n \int_\gamma (z - z_0)^n\,dz \\ &= \sum_{n=-\infty}^{-2} a_n \int_\gamma (z - z_0)^n\,dz \\ &\quad + a_{-1} \int_\gamma \frac{dz}{z - z_0} + \sum_{n=0}^{\infty} a_n \int_\gamma (z - z_0)^n\,dz. \end{aligned} \tag{2}$$

The CAUCHY INTEGRAL THEOREM requires that the first and last terms vanish, so we have

$$\int_\gamma f(z)\,dz = a_{-1} \int_\gamma \frac{dz}{z - z_0}. \tag{3}$$

But we can evaluate this function (which has a POLE at z_0) using the CAUCHY INTEGRAL FORMULA,

$$f(z_0) = \frac{1}{2\pi i} \int_\gamma \frac{f(z)\,dz}{z - z_0}. \tag{4}$$

This equation must also hold for the constant function $f(z) = 1$, in which case it is also true that $f(z_0) = 1$, so

$$1 = \frac{1}{2\pi i} \int_\gamma \frac{dz}{z - z_0}, \tag{5}$$

$$\int_\gamma \frac{dz}{z - z_0} = 2\pi i, \tag{6}$$

and (3) becomes

$$\int_\gamma f(z)\,dz = 2\pi i a_{-1}. \tag{7}$$

The quantity a_{-1} is known as the RESIDUE of $f(z)$ at z_0. Generalizing to a curve passing through multiple poles, (7) becomes

$$\int_\gamma f(z)\,dz = 2\pi i \sum_{i=1}^{\text{poles in }\gamma} n(\gamma, z_0^{(i)})a_{-1}^{(i)}, \tag{8}$$

where n is the WINDING NUMBER and the (i) superscript denotes the quantity corresponding to POLE i.

If the path does not completely encircle the RESIDUE, take the CAUCHY PRINCIPAL VALUE to obtain

$$\int f(z)\,dz = (\theta_2 - \theta_1)ia_{-1}. \tag{9}$$

If f has only ISOLATED SINGULARITIES, then

$$\sum_{z_0^{(i)}\in\mathbb{C}^*} {a_{-1}}^{(i)} = 0. \tag{10}$$

The residues may be found without explicitly expanding into a LAURENT SERIES as follows.

$$f(z) = \sum_{n=-\infty}^{\infty} a_n(z-z_0)^n. \tag{11}$$

If $f(z)$ has a POLE of order m at z_0, then $a_n = 0$ for $n < -m$ and $a_{-m} \neq 0$. Therefore,

$$f(z) = \sum_{n=-m}^{\infty} a_n(z-z_0)^n = \sum_{n=0}^{\infty} a_{-m+n}(z-z_0)^{-m+n} \tag{12}$$

$$(z-z_0)^m f(z) = \sum_{n=0}^{\infty} a_{-m+n}(z-z_0)^n \tag{13}$$

$$\begin{aligned}\frac{d}{dz}[(z-z_0)^m f(z)] &= \sum_{n=0}^{\infty} na_{-m+n}(z-z_0)^{n-1}\\ &= \sum_{n=1}^{\infty} na_{-m+n}(z-z_0)^{n-1}\\ &= \sum_{n=0}^{\infty}(n+1)a_{-m+n+1}(z-z_0)^n \end{aligned}\tag{14}$$

$$\begin{aligned}\frac{d^2}{dz^2}[(z-z_0)^m f(z)] &= \sum_{n=0}^{\infty} n(n+1)a_{-m+n+1}(z-z_0)^{n-1}\\ &= \sum_{n=1}^{\infty} n(n+1)a_{-m+n+1}(z-z_0)^{n-1}\\ &= \sum_{n=0}^{\infty}(n+1)(n+2)a_{-m+n+2}(z-z_0)^n. \end{aligned}\tag{15}$$

Iterating,

$$\begin{aligned}&\frac{d^{m-1}}{dz^{m-1}}[(z-z_0)^m f(z)]\\ &\qquad= \sum_{n=0}^{\infty}(n+1)(n+2)(n+m-1)a_{n-1}(z-z_0)^n\\ &= (m-1)!a_{-1}\\ &+\sum_{n=1}^{\infty}(n+1)(n+2)(n+m-1)a_{n-1}(z-z_0)^{n-1}. \end{aligned}\tag{16}$$

So

$$\begin{aligned}&\lim_{z\to z_0}\frac{d^{m-1}}{dz^{m-1}}[(z-z_0)^m f(z)]\\ &\qquad= \lim_{z\to z_0}(m-1)!a_{-1} + 0 = (m-1)!a_{-1}, \end{aligned}\tag{17}$$

and the RESIDUE is

$$a_{-1} = \frac{1}{(m-1)!}\frac{d^{m-1}}{dz^{m-1}}[(z-z_0)^m f(z)]_{z=z_0}. \tag{18}$$

This amazing theorem says that the value of a CONTOUR INTEGRAL in the COMPLEX PLANE depends *only* on the properties of a few special points *inside* the contour.

see also CAUCHY INTEGRAL FORMULA, CAUCHY INTEGRAL THEOREM, CONTOUR INTEGRAL, LAURENT SERIES, POLE, RESIDUE (COMPLEX ANALYSIS)

Residue Theorem (Group)

If two groups are residual to a third, every group residual to one is residual to the other. The Gambier extension of this theorem states that if two groups are pseudoresidual to a third, then every group pseudoresidual to the first with an excess greater than or equal to the excess of the first minus the excess of the second is pseudoresidual to the second, with an excess ≥ 0.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, pp. 30–31, 1959.

Resolution

Resolution is a widely used word with many different meanings. It can refer to resolution of equations, resolution of singularities (in ALGEBRAIC GEOMETRY), resolution of modules or more sophisticated structures, etc. In a BLOCK DESIGN, a PARTITION R of a BIBD's set of blocks B into PARALLEL CLASSES, each of which in turn partitions the set V, is called a resolution (Abel and Furino 1996).

A resolution of the MODULE M over the RING R is a complex of R-modules C_i and morphisms d_i and a MORPHISM ϵ such that

$$\cdots \to C_i \to^{d_i} C_{i-1} \to \cdots \to C_0 \to^{\epsilon} M \to 0$$

satisfying the following conditions:

1. The composition of any two consecutive morphisms is the zero map,
2. For all i, $(\ker d_i)/(\operatorname{im} d_{i+1}) = 0$,
3. $C_0/(\ker \epsilon) \simeq M$,

where ker is the kernel and im is the image. Here, the quotient

$$\frac{(\ker d_i)}{(\operatorname{im} d_{i+1})}$$

is the ith Homology Group.

If all modules C_i are projective (free), then the resolution is called projective (free). There is a similar concept for resolutions "to the right" of M, which are called injective resolutions.

see also Homology Group, Module, Morphism, Ring

References

Abel, R. J. R. and Furino, S. C. "Resolvable and Near Resolvable Designs." §I.6 in *The CRC Handbook of Combinatorial Designs* (Ed. C. J. Colbourn and J. H. Dinitz). Boca Raton, FL: CRC Press, p. 4 and 87–94, 1996.

Jacobson, N. *Basic Algebra II, 2nd ed.* New York: W. H. Freeman, p. 339, 1989.

Resolution Class

see Parallel Class

Resolution Modulus

The least Positive Integer m^* with the property that $\chi(y) = 1$ whenever $y \equiv 1 \pmod{m^*}$ and $(y, m) = 1$.

Resolvable

A balanced incomplete Block Design (B, V) is called resolvable if there exists a Partition R of its set of blocks B into Parallel Classes, each of which in turn partitions the set V. The partition R is called a Resolution.

see also Block Design, Parallel Class

References

Abel, R. J. R. and Furino, S. C. "Resolvable and Near Resolvable Designs." §I.6 in *The CRC Handbook of Combinatorial Designs* (Ed. C. J. Colbourn and J. H. Dinitz). Boca Raton, FL: CRC Press, p. 4 and 87–94, 1996.

Resolving Tree

A tree of Links obtained by repeatedly choosing a crossing, applying the Skein Relationship to obtain two simpler Links, and repeating the process. The Depth of a resolving tree is the number of levels of links, not including the top. The Depth of the Link is the minimal depth for any resolving tree of that Link.

Resonance Overlap

Isolated resonances in a Dynamical System can cause considerable distortion of preserved Tori in their Neighborhood, but they do not introduce any Chaos into a system. However, when two or more resonances are simultaneously present, they will render a system nonintegrable. Furthermore, if they are sufficiently "close" to each other, they will result in the appearance of widespread (large-scale) Chaos.

To investigate this problem, Walker and Ford (1969) took the integrable Hamiltonian

$$H_0(I_1, I_2) = I_1 + I_2 - I_1^2 - 3I_1I_2 + {I_2}^2$$

and investigated the effect of adding a 2:2 resonance and a 3:2 resonance

$$H(\mathbf{I}, \theta) = H_0(\mathbf{I}) + \alpha I_1 I_2 \cos(2\theta_1 - 2\theta_2) + \beta {I_1}^{3/2} I_2 \cos(2\theta_1 - 3\theta_2).$$

At low energies, the resonant zones are well-separated. As the energy increases, the zones overlap and a "macroscopic zone of instability" appears. When the overlap starts, many higher-order resonances are also involved so fairly large areas of Phase Space have their Tori destroyed and the ensuing Chaos is "widespread" since trajectories are now free to wander between regions that previously were separated by nonresonant Tori.

Walker and Ford (1969) were able to numerically predict the energy at which the overlap of the resonances first occurred. They plotted the θ_2-axis intercepts of the inner 2:2 and the outer 2:3 separatrices as a function of total energy. The energy at which they crossed was found to be identical to that at which 2:2 and 2:3 resonance zones began to overlap.

see also Chaos, Resonance Overlap Method

References

Walker, G. H. and Ford, J. "Amplitude Instability and Ergodic Behavior for Conservative Nonlinear Oscillator Systems." *Phys. Rev.* **188**, 416–432, 1969.

Resonance Overlap Method

A method for predicting the onset of widespread Chaos.

see also Greene's Method

References

Chirikov, B. V. "A Universal Instability of Many-Dimensional Oscillator Systems." *Phys. Rep.* **52**, 264–379, 1979.

Tabor, M. *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, pp. 154–163, 1989.

Restricted Divisor Function

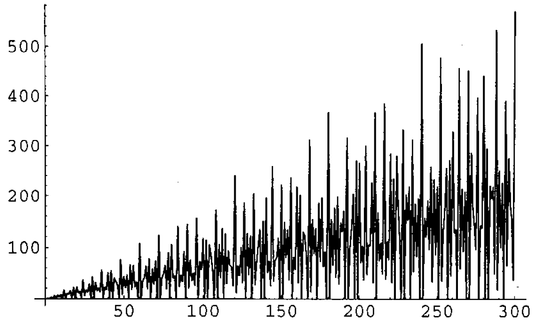

The sum of the ALIQUOT DIVISORS of n, given by

$$s(n) \equiv \sigma(n) - n,$$

where $\sigma(n)$ is the DIVISOR FUNCTION. The first few values are 0, 1, 1, 3, 1, 6, 1, 7, 4, 8, 1, 16, ... (Sloane's A001065).

see also DIVISOR FUNCTION

References
Sloane, N. J. A. Sequence A001065/M2226 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Restricted Growth Function

see RESTRICTED GROWTH STRING

Restricted Growth String

For a SET PARTITION of n elements, the n-character string $a_1 a_2 \ldots a_n$ in which each character gives the BLOCK ($\mathbf{B}_0$, $\mathbf{B}_1$, ...) into which the corresponding element belongs is called the restricted growth string (or sometimes the RESTRICTED GROWTH FUNCTION). For example, for the SET PARTITION $\{\{1\},\{2\},\{3,4\}\}$, the restricted growth string would be 0122. If the BLOCKS are "sorted" so that $a_1 = 0$, then the restricted growth string satisfies the INEQUALITY

$$a_{i+1} \leq 1 + \max\{a_1, a_2, \ldots, a_i\}$$

for $i = 1, 2, \ldots, n-1$.

References
Ruskey, F. "Info About Set Partitions." `http://sue.csc.uvic.ca/~cos/inf/setp/SetPartitions.html`.

Resultant

Given a POLYNOMIAL $p(x)$ of degree n with roots α_i, $i = 1, \ldots, n$ and a POLYNOMIAL $q(x)$ of degree m with roots β_j, $j = 1, \ldots, m$, the resultant is defined by

$$R(p,q) = \prod_{i=1}^{n} \prod_{j=1}^{m} (\beta_j - \alpha_i).$$

There exists an ALGORITHM similar to the EUCLIDEAN ALGORITHM for computing resultants (Pohst and Zassenhaus 1989). The resultant is the DETERMINANT of the corresponding SYLVESTER MATRIX. Given p and q, then

$$h(x) = R(q(t), p(x-t))$$

is a POLYNOMIAL of degree mn, having as its roots all sums of the form $\alpha_i + \beta_j$.

see also DISCRIMINANT (POLYNOMIAL), SYLVESTER MATRIX

References
Pohst, M. and Zassenhaus, H. *Algorithmic Algebraic Number Theory.* Cambridge, England: Cambridge University Press, 1989.
Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, p. 348, 1991.

Retardance

A shift in PHASE.

see also PHASE

Reuleaux Polygon

A curvilinear polygon built up of circular ARCS. The Reuleaux polygon is a generalization of the REULEAUX TRIANGLE.

see also CURVE OF CONSTANT WIDTH, REULEAUX TRIANGLE

References
Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 52–54, 1991.

Reuleaux Triangle

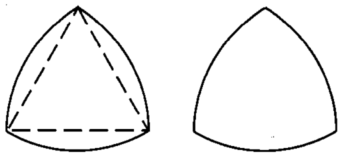

A CURVE OF CONSTANT WIDTH constructed by drawing arcs from each VERTEX of an EQUILATERAL TRIANGLE between the other two VERTICES. It is the basis for the Harry Watt square drill bit. It has the smallest AREA for a given width of any CURVE OF CONSTANT WIDTH.

The AREA of each meniscus-shaped portion is

$$A = \tfrac{1}{6}\pi r^2 - \tfrac{1}{2} r \left(\frac{\sqrt{3}}{2} r\right) = \left(\frac{\pi}{6} - \frac{\sqrt{3}}{4}\right) r^2, \quad (1)$$

where we have subtracted the AREA of the wedge from that of the EQUILATERAL TRIANGLE. The total AREA is then

$$A = 3\left(\frac{\pi}{6} - \frac{\sqrt{3}}{4}\right) r^2 + \frac{\sqrt{3}}{4} r^2 = \frac{\pi - \sqrt{3}}{2} r^2. \quad (2)$$

When rotated in a square, the fractional AREA covered is

$$A_{\text{covered}} = 2\sqrt{3} + \tfrac{1}{6}\pi = 0.9877700392\ldots. \quad (3)$$

The center does *not* stay fixed as the TRIANGLE is rotated, but moves along a curve composed of four arcs of an ELLIPSE (Wagon 1991).

see also CURVE OF CONSTANT WIDTH, FLOWER OF LIFE, PIECEWISE CIRCULAR CURVE, REULEAUX POLYGON

References

Bogomolny, A. "Shapes of Constant Width." `http://www.cut-the-knot.com/do_you_know/cwidth.html`.

Eppstein, D. "Reuleaux Triangles." `http://www.ics.uci.edu/~eppstein/junkyard/reuleaux.html`.

Reuleaux, F. *The Kinematics of Machinery.* New York: Dover, 1963.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 52–54 and 381–383, 1991.

Yaglom, I. M. and Boltyansky, B. G. *Convex Shapes.* Moscow: Nauka, 1951.

Reversal

The reversal of a decimal number $abc\cdots$ is $\cdots cba$. Ball and Coxeter (1987) consider numbers whose reversals are integral multiples of themselves. PALINDROMIC NUMBER and numbers ending with a ZERO are trivial examples. The first few nontrivial examples are 8712, 9801, 87912, 98901, 879912, 989901, 8799912, 9899901, 87128712, 87999912, 98019801, 98999901, ... (Sloane's A031877). The pattern continues for large numbers, with numbers of the form $87\underbrace{9\cdots9}12$ equal to 4 times their reversals and numbers of the form $98\underbrace{9\cdots9}01$ equal to 9 times their reversals. In addition, runs of numbers of either of these forms can be concatenated to yield numbers of the form $87\underbrace{9\cdots9}12\cdots87\underbrace{9\cdots9}12$, equal to 4 times their reversals, and $98\underbrace{9\cdots9}01\cdots98\underbrace{9\cdots9}01$, equal to 9 times their reversals.

The product of a 2-digit number and its reversal is never a SQUARE NUMBER except when the digits are the same (Ogilvy 1988). Numbers whose product is the reversal of the products of their reversals include

$$312\times221=68952$$

$$213\times122=25986$$

(Ball and Coxeter 1987, p. 14).

see also RATS SEQUENCE

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 14–15, 1987.

Ogilvy, C. S. and Anderson, J. T. *Excursions in Number Theory.* New York: Dover, pp. 88–89, 1988.

Sloane, N. J. A. Sequence A031877 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Reverse Greedy Algorithm

An algorithm for computing a UNIT FRACTION.

see also GREEDY ALGORITHM, UNIT FRACTION

Reversion of Series

see SERIES REVERSION

Reverse-Then-Add Sequence

An integer sequence produced by the 196-ALGORITHM.

see also 196-ALGORITHM, SORT-THEN-ADD SEQUENCE

Reznik's Identity

For P and Q POLYNOMIALS in n variables,

$$|P\cdot Q|_2{}^2 = \sum_{i_1,\ldots,i_n\geq0}\frac{|P^{(i_1,\ldots,i_n)}(D_1,\ldots,D_n)Q(x_1,\ldots,x_n)|_2{}^2}{i_1!\cdots i_n!},$$

where $D_i\equiv\partial/\partial x_i$, $|X|_2$ is the BOMBIERI NORM, and

$$P^{(i_1,\ldots,i_n)}=D_1^{i_1}\cdots D_n^{i_n}P.$$

BOMBIERI'S INEQUALITY follows from this identity.

see also BEAUZAMY AND DÉGOT'S IDENTITY

Rhodonea

see ROSE

Rhomb

see RHOMBUS

Rhombic Dodecahedral Number

A FIGURATE NUMBER which is constructed as a centered CUBE with a SQUARE PYRAMID appended to each face,

$$RhoDod_n=CCub_n+6P_{n-1}=(2n-1)(2n^2-2n+1),$$

where $CCub_n$ is a CENTERED CUBE NUMBER and P_n is a PYRAMIDAL NUMBER. The first few are 1, 15, 65, 175, 369, 671, ... (Sloane's A005917). The GENERATING FUNCTION of the rhombic dodecahedral numbers is

$$\frac{x(1+11x+11x^2+x^3)}{(x-1)^4}=x+15x^2+65x^3+175x^4+\ldots.$$

see also OCTAHEDRAL NUMBER

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 53–54, 1996.

Sloane, N. J. A. Sequence A005917/M4968 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Rhombic Dodecahedron

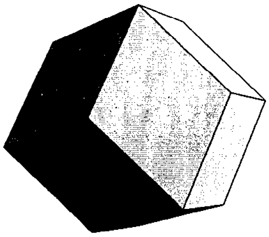

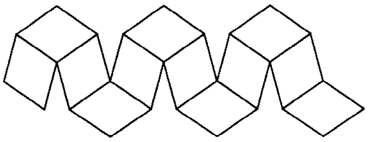

The DUAL POLYHEDRON of the CUBOCTAHEDRON, also sometimes called the RHOMBOIDAL DODECAHEDRON (Cotton 1990). Its 14 vertices are joined by 12 RHOMBUSES, and one possible way to construct it is known as the BAUSPIEL. The rhombic dodecahedron is a ZONOHEDRON and a SPACE-FILLING POLYHEDRON. The vertices are given by (±1, ±1, ±1), (±2, 0, 0), (0, ±2, 0), (0, 0, ±2).

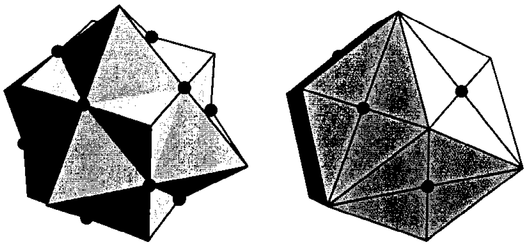

The edges of the CUBE-OCTAHEDRON COMPOUND intersecting in the points plotted above are the diagonals of RHOMBUSES, and the 12 RHOMBUSES form a rhombic dodecahedron (Ball and Coxeter 1987).

see also BAUSPIEL, CUBE-OCTAHEDRON COMPOUND, DODECAHEDRON, PYRITOHEDRON, RHOMBIC TRIACONTAHEDRON, RHOMBUS, TRIGONAL DODECAHEDRON, ZONOHEDRON

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 137, 1987.
Cotton, F. A. *Chemical Applications of Group Theory, 3rd ed.* New York: Wiley, p. 62, 1990.

Rhombic Icosahedron

A ZONOHEDRON which can be derived from the TRIACONTAHEDRON by removing any one of the zones and bringing together the two pieces into which the remainder of the surface is thereby divided.

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 143, 1987.
Bilinski, S. "Über die Rhombenisoeder." *Glasnik Mat.-Fiz. Astron. Društ̆ro Mat. Fiz. Hrvatske Ser. II* **15**, 251–263, 1960.

Rhombic Polyhedron

A POLYHEDRON with extra square faces, given by the SCHLÄFLI SYMBOL $r\left\{ {p \atop q} \right\}$.

see also RHOMBIC DODECAHEDRON, RHOMBIC ICOSAHEDRON, RHOMBIC TRIACONTAHEDRON, SNUB POLYHEDRON, TRUNCATED POLYHEDRON

Rhombic Triacontahedron

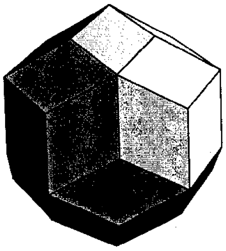

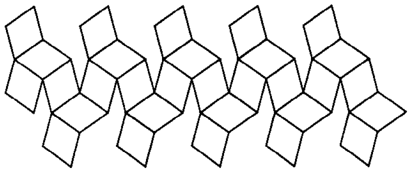

A ZONOHEDRON which is the DUAL POLYHEDRON of the ICOSIDODECAHEDRON. It is composed of 30 RHOMBUSES joined at 32 vertices. Ede (1958) enumerates 13 basic series of stellations of the rhombic triacontahedron, the total number of which is extremely large. Messer (1995) describes 226 stellations. The intersecting edges of the DODECAHEDRON-ICOSAHEDRON COMPOUND form the diagonals of 30 RHOMBUSES which comprise the TRIACONTAHEDRON. The CUBE 5-COMPOUND has the 30 facial planes of the rhombic triacontahedron (Ball and Coxeter 1987).

see also CUBE 5-COMPOUND, DODECAHEDRON-ICOSAHEDRON COMPOUND, ICOSIDODECAHEDRON, RHOMBIC DODECAHEDRON, RHOMBUS, ZONOHEDRON

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 137, 1987.
Bulatov, V.v "Stellations of Rhombic Triacontahedron." `http://www.physics.orst.edu/~bulatov/polyhedra/rtc/`.
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 127, 1989.
Ede, J. D. "Rhombic Triacontahedra." *Math. Gazette* **42**, 98–100, 1958.
Messer, P. W. "Les étoilements du rhombitricontaèdre et plus." *Structural Topology* No. 21, 25–46, 1995.

Rhombicosacron

The DUAL POLYHEDRON of the RHOMBICOSAHEDRON.

Rhombicosahedron

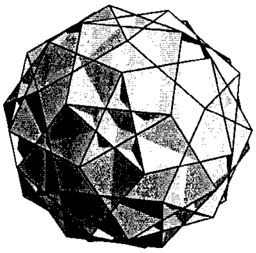

The UNIFORM POLYHEDRON U_{56} whose DUAL POLYHEDRON is the RHOMBICOSACRON. It has WYTHOFF SYMBOL $2\,3\,\genfrac{}{}{0pt}{}{\frac{5}{4}}{\frac{5}{2}}$. Its faces are $20\{6\}+30\{4\}$. The CIRCUMRADIUS for unit edge length is

$$R=\tfrac{1}{2}\sqrt{7}.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 149–150, 1971.

Rhombicosidodecahedron

see BIGYRATE DIMINISHED RHOMBICOSIDODECAHEDRON, DIMINISHED RHOMBICOSIDODECAHEDRON, GREAT RHOMBICOSIDODECAHEDRON (ARCHIMEDEAN), GREAT RHOMBICOSIDODECAHEDRON (UNIFORM), GYRATE BIDIMINISHED RHOMBICOSIDODECAHEDRON, GYRATE RHOMBICOSIDODECAHEDRON, METABIDIMINISHED RHOMBICOSIDODECAHEDRON, METABIGYRATE RHOMBICOSIDODECAHEDRON, METAGYRATE DIMINISHED RHOMBICOSIDODECAHEDRON, PARABIDIMINISHED RHOMBICOSIDODECAHEDRON, PARABIGYRATE RHOMBICOSIDODECAHEDRON, PARAGYRATE DIMINISHED RHOMBICOSIDODECAHEDRON, SMALL RHOMBICOSIDODECAHEDRON, TRIDIMINISHED RHOMBICOSIDODECAHEDRON, TRIGYRATE RHOMBICOSIDODECAHEDRON

Rhombicuboctahedron

see GREAT RHOMBICUBOCTAHEDRON (ARCHIMEDEAN), GREAT RHOMBICUBOCTAHEDRON (UNIFORM), SMALL RHOMBICUBOCTAHEDRON

Rhombidodecadodecahedron

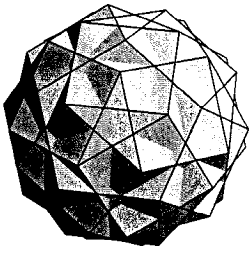

The UNIFORM POLYHEDRON U_{38} whose DUAL POLYHEDRON is the MEDIAL DELTOIDAL HEXECONTAHEDRON. It has SCHLÄFLI SYMBOL $r\left\{\begin{matrix}\frac{5}{2}\\ 5\end{matrix}\right\}$ and WYTHOFF SYMBOL $\frac{5}{2}\,5\,|\,2$. Its faces are $12\{\frac{5}{2}\}+30\{4\}+12\{5\}$. The CIRCUMRADIUS for unit edge length is

$$R=\tfrac{1}{2}\sqrt{7}.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 116–117, 1989.

Rhombihexacron

see GREAT RHOMBIHEXACRON, SMALL RHOMBIHEXACRON

Rhombihexahedron

see GREAT RHOMBIHEXAHEDRON, SMALL RHOMBIHEXAHEDRON

Rhombitruncated Cuboctahedron

see GREAT RHOMBICUBOCTAHEDRON (ARCHIMEDEAN)

Rhombitruncated Icosidodecahedron

see GREAT RHOMBICOSIDODECAHEDRON (ARCHIMEDEAN)

Rhombohedron

A PARALLELEPIPED bounded by six congruent RHOMBS.

see also PARALLELEPIPED, RHOMB

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 142 and 161, 1987.

Rhomboid

A PARALLELOGRAM in which angles are oblique and adjacent sides are of unequal length.

see also DIAMOND, LOZENGE, PARALLELOGRAM, QUADRILATERAL, RHOMBUS, SKEW QUADRILATERAL, TRAPEZIUM, TRAPEZOID

Rhomboidal Dodecahedron

see RHOMBIC DODECAHEDRON

Rhombus

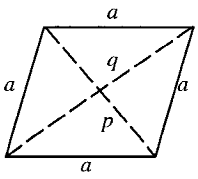

A QUADRILATERAL with both pairs of opposite sides PARALLEL and all sides the same length, i.e., an equilateral PARALLELOGRAM. The word RHOMB is sometimes used instead of rhombus. The Diagonals p and q of a rhombus satisfy

$$p^2 + q^2 = 4a^2,$$

and the Area is

$$A = \tfrac{1}{2}pq.$$

A rhombus whose Acute Angles are 45° is called a Lozenge.

see also Diamond, Lozenge, Parallelogram, Quadrilateral, Rhombic Dodecahedron, Rhombic Icosahedron, Rhombic Triacontahedron, Rhomboid, Skew Quadrilateral, Trapezium, Trapezoid

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 123, 1987.

Rhumb Line

see Loxodrome

Ribbon Knot

If the Knot K is the boundary $K = f(\mathbb{S}^1)$ of a singular disk $f : \mathbb{D} \to \mathbb{S}^3$ which has the property that each self-intersecting component is an arc $A \subset f(\mathbb{D}^2)$ for which $f^{-1}(A)$ consists of two arcs in $\mathbb{D}^2$, one of which is interior, then K is said to be a ribbon knot. Every ribbon knot is a Slice Knot, and it is conjectured that every Slice Knot is a ribbon knot.

see also Slice Knot

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 225, 1976.

Ribet's Theorem

If the Taniyama-Shimura Conjecture holds for all semistable Elliptic Curves, then Fermat's Last Theorem is true. Before its proof by Ribet in 1986, the theorem had been called the epsilon conjecture. It had its roots in a surprising result of G. Frey.

see also Elliptic Curve, Fermat's Last Theorem, Modular Form, Modular Function, Taniyama-Shimura Conjecture

Riccati-Bessel Functions

$$S_n(z) \equiv zj_n(z) = \sqrt{\frac{\pi z}{2}}\, J_{n+1/2}(z)$$

$$C_n(z) \equiv -zn_n(z) = -\sqrt{\frac{\pi z}{2}}\, N_{n+1/2}(z),$$

where $j_n(z)$ and $n_n(z)$ are Spherical Bessel Functions of the First and Second Kind.

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Riccati-Bessel Functions." §10.3 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 445, 1972.

Riccati Differential Equation

$$y' = P(z) + Q(z)y + R(z)y^2, \tag{1}$$

where $y' \equiv dy/dz$. The transformation

$$w \equiv -\frac{y'}{yR(z)} \tag{2}$$

leads to the second-order linear homogeneous equation

$$R(z)y'' - [R'(z) + Q(z)R(z)]y' + [R(z)]^2 P(z)y = 0. \tag{3}$$

Another equation sometimes called the Riccati differential equation is

$$z^2 w'' + [z^2 - n(n+1)]w = 0, \tag{4}$$

which has solutions

$$w = Azj_n(z) + Bzy_n(z). \tag{5}$$

Yet another form of "the" Riccati differential equation is

$$\frac{dy}{dz} = az^n + by^2, \tag{6}$$

which is solvable by algebraic, exponential, and logarithmic functions only when $n = -4m/(2m \pm 1)$, for $m = 0$, 1, 2, ...

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Riccati-Bessel Functions." §10.3 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 445, 1972.

Glaisher, J. W. L. "On Riccati's Equation." *Quart. J. Pure Appl. Math.* **11**, 267–273, 1871.

Ricci Curvature

The mathematical object which controls the growth rate of the volume of metric balls in a Manifold.

see also Bishop's Inequality, Milnor's Theorem

Ricci Tensor

$$R_{\mu\kappa} \equiv R^{\lambda}{}_{\mu\lambda\kappa},$$

where $R^{\lambda}{}_{\mu\lambda\kappa}$ is the Riemann Tensor.

see also Curvature Scalar, Riemann Tensor

Rice Distribution

$$P(Z) = \frac{Z}{\sigma^2} \exp\left(-\frac{Z^2 + |V|^2}{2\sigma^2}\right) I_0\left(\frac{Z|V|}{\sigma^2}\right),$$

where $I_0(z)$ is a Modified Bessel Function of the First Kind and $Z > 0$. For a derivation, see Papoulis (1962). For $|V| = 0$, this reduces to the Rayleigh Distribution.

see also Rayleigh Distribution

References

Papoulis, A. *The Fourier Integral and Its Applications.* New York: McGraw-Hill, 1962.

Richard's Paradox

It is possible to describe a set of POSITIVE INTEGERS that cannot be listed in a book containing a set of counting numbers on each consecutively numbered page.

Richardson Extrapolation

The consideration of the result of a numerical calculation as a function of an adjustable parameter (usually the step size). The function can then be fitted and evaluated at $h = 0$ to yield very accurate results. Press *et al.* (1992) describe this process as turning lead into gold. Richardson extrapolation is one of the key ideas used in the popular and robust BULIRSCH-STOER ALGORITHM of solving ORDINARY DIFFERENTIAL EQUATIONS.

see also BULIRSCH-STOER ALGORITHM

References

Acton, F. S. *Numerical Methods That Work, 2nd printing.* Washington, DC: Math. Assoc. Amer., p. 106, 1990.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Richardson Extrapolation and the Bulirsch-Stoer Method." §16.4 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 718–725, 1992.

Richardson's Theorem

Let R be the class of expressions generated by

1. The RATIONAL NUMBERS and the two REAL NUMBERS π and $\ln 2$,
2. The variable x,
3. The operations of ADDITION, MULTIPLICATION, and composition, and
4. The SINE, EXPONENTIAL, and ABSOLUTE VALUE functions.

Then if $E \in R$, the predicate "$E = 0$" is recursively UNDECIDABLE.

see also RECURSION, UNDECIDABLE

References

Caviness, B. F. "On Canonical Forms and Simplification." *J. Assoc. Comp. Mach.* **17**, 385–396, 1970.

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, 1996.

Richardson, D. "Some Unsolvable Problems Involving Elementary Functions of a Real Variable." *J. Symbolic Logic* **33**, 514–520, 1968.

Ridders' Method

A variation of the FALSE POSITION METHOD for finding ROOTS which fits the function in question with an exponential.

see also FALSE POSITION METHOD

References

Ostrowski, A. M. Ch. 12 in *Solutions of Equations and Systems of Equations, 2nd ed.* New York: Academic Press, 1966.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Secant Method, False Position Method, and Ridders' Method." §9.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 347–352, 1992.

Ralston, A. and Rabinowitz, P. §8.3 in *A First Course in Numerical Analysis, 2nd ed.* New York: McGraw-Hill, 1978.

Ridders, C. F. J. "A New Algorithm for Computing a Single Root of a Real Continuous Function." *IEEE Trans. Circuits Systems* **26**, 979–980, 1979.

Ridge

An $(n-2)$-D FACE of an n-D POLYTOPE.

see also POLYTOPE

Riemann-Christoffel Tensor

see RIEMANN TENSOR

Riemann Curve Theorem

If two algebraic plane curves with only ordinary singular points and CUSPS are related such that the coordinates of a point on either are RATIONAL FUNCTIONS of a corresponding point on the other, then the curves have the same GENUS (CURVE). This can be stated equivalently as the GENUS of a curve is unaltered by a BIRATIONAL TRANSFORMATION.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 120, 1959.

Riemann Differential Equation

see RIEMANN P-DIFFERENTIAL EQUATION

Riemann's Formula

$$J(x) = \mathrm{Li}(x) - \sum \mathrm{Li}(x^\rho) + \ln 2 \int_x^\infty \frac{dt}{t(t^2-1)\ln t},$$

where $\mathrm{Li}(x)$ is the LOGARITHMIC INTEGRAL, the sum is taken over all nontrivial zeros ρ (i.e., those other than $-2, -4, \ldots$) of the RIEMANN ZETA FUNCTION $\zeta(s)$, and $J(x)$ is RIEMANN WEIGHTED PRIME-POWER COUNTING FUNCTION.

see also LOGARITHMIC INTEGRAL, PRIME NUMBER THEOREM, RIEMANN WEIGHTED PRIME-POWER COUNTING FUNCTION, RIEMANN ZETA FUNCTION

Riemann Function

The function obtained by approximating the RIEMANN WEIGHTED PRIME-POWER COUNTING FUNCTION J_2 in

$$\pi(x) = \sum_{n=1}^\infty \frac{\mu(n)}{n} J_2(x^{1/n}) \qquad (1)$$

by the LOGARITHMIC INTEGRAL Li(x). This gives

$$R(n) \equiv 1 + \sum_{k=1}^{\infty} \frac{1}{k\zeta(k+1)} \frac{(\ln n)^k}{k!} \tag{2}$$

$$= \sum_{m=1}^{\infty} \frac{\mu(m)}{m} \operatorname{Li}(n^{1/m}), \tag{3}$$

where $\zeta(z)$ is the RIEMANN ZETA FUNCTION, $\mu(n)$ is the MÖBIUS FUNCTION, and Li(x) is the LOGARITHMIC INTEGRAL. Then

$$\pi(x) = R(x) - \sum_{\rho} R(x^{\rho}), \tag{4}$$

where π is the PRIME COUNTING FUNCTION. Ramanujan independently derived the formula for $R(n)$, but nonrigorously (Berndt 1994, p. 123).

see also MANGOLDT FUNCTION, PRIME NUMBER THEOREM, RIEMANN-MANGOLDT FUNCTION, RIEMANN ZETA FUNCTION

References

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, 1994.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 144–145, 1996.

Riesel, H. and Göhl, G. "Some Calculations Related to Riemann's Prime Number Formula." *Math. Comput.* **24**, 969–983, 1970.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 28–29 and 362–372, 1991.

Riemann Hypothesis

First published in Riemann (1859), the Riemann hypothesis states that the nontrivial ROOTS of the RIEMANN ZETA FUNCTION

$$\zeta(s) \equiv \sum_{n=1}^{\infty} \frac{1}{n^s}, \tag{1}$$

where $s \in \mathbb{C}$ (the COMPLEX NUMBERS), all lie on the "CRITICAL LINE" $\Re[s] = 1/2$, where $\Re[z]$ denotes the REAL PART of z. The Riemann hypothesis is also known as ARTIN'S CONJECTURE.

In 1914, Hardy proved that an INFINITE number of values for s can be found for which $\zeta(s) = 0$ and $\Re[s] = 1/2$. However, it is not known if *all* nontrivial roots s satisfy $\Re[s] = 1/2$, so the conjecture remains open. André Weil proved the Riemann hypothesis to be true for field functions (Weil 1948, Eichler 1966, Ball and Coxeter 1987). In 1974, Levin showed that at least 1/3 of the ROOTS must lie on the CRITICAL LINE (Le Lionnais 1983), a result which has since been sharpened to 40% (Vardi 1991, p. 142). It is known that the zeros are symmetrical placed about the line $\Im[s] = 0$.

The Riemann hypothesis is equivalent to $\Lambda \leq 0$, where Λ is the DE BRUIJN-NEWMAN CONSTANT (Csordas *et al.* 1994). It is also equivalent to the assertion that for some constant c,

$$|\operatorname{Li}(x) - \pi(x)| \leq c\sqrt{x}\ln x, \tag{2}$$

where Li(x) is the LOGARITHMIC INTEGRAL and π is the PRIME COUNTING FUNCTION (Wagon 1991).

The hypothesis was computationally tested and found to be true for the first 2×10^8 zeros by Brent *et al.* (1979), a limit subsequently extended to the first $1.5 \times 10^9 + 1$ zeros by Brent *et al.* (1979). Brent's calculation covered zeros $\sigma + it$ in the region $0 < t < 81,702,130.19$.

There is also a finite analog of the Riemann hypothesis concerning the location of zeros for function fields defined by equations such as

$$ay^l + bz^m + c = 0. \tag{3}$$

This hypothesis, developed by Weil, is analogous to the usual Riemann hypothesis. The number of solutions for the particular cases $(l, m) = (2, 2)$, (3,3), (4,4), and (2,4) were known to Gauss.

see also CRITICAL LINE, EXTENDED RIEMANN HYPOTHESIS, GRONWALL'S THEOREM, MERTENS CONJECTURE, MILLS' CONSTANT, RIEMANN ZETA FUNCTION

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 75, 1987.

Brent, R. P.; Vandelune, J.; te Riele, H. J. J.; and Winter, D. T. "On the Zeros of the Riemann Zeta Function in the Critical Strip. I." *Math. Comput.* **33**, 1361–1372, 1979.

Brent, R. P.; Vandelune, J.; te Riele, H. J. J.; and Winter, D. T. "On the Zeros of the Riemann Zeta Function in the Critical Strip. II." *Math. Comput.* **39**, 681–688, 1982. Abstract available at `ftp://nimbus.anu.edu.au/pub/Brent/rpb070a.dvi.Z`.

Csordas, G.; Smith, W.; and Varga, R. S. "Lehmer Pairs of Zeros, the de Bruijn-Newman Constant and the Riemann Hypothesis." *Constr. Approx.* **10**, 107–129, 1994.

Eichler, M. *Introduction to the Theory of Algebraic Numbers and Functions.* New York: Academic Press, 1966.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 25, 1983.

Odlyzko, A. "The 10^{20}th Zero of the Riemann Zeta Function and 70 Million of Its Neighbors."

Riemann, B. "Über die Anzahl der Primzahlen unter einer gegebenen Grösse," *Mon. Not. Berlin Akad.*, pp. 671–680, Nov. 1859.

Sloane, N. J. A. Sequence A002410/M4924 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Vandelune, J. and te Riele, H. J. J. "On The Zeros of the Riemann Zeta-Function in the Critical Strip. III." *Math. Comput.* **41**, 759–767, 1983.

Vandelune, J.; te Riele, H. J. J.; and Winter, D. T. "On the Zeros of the Riemann Zeta Function in the Critical Strip. IV." *Math. Comput.* **46**, 667–681, 1986.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, p. 33, 1991.

Weil, A. *Sur les courbes algébriques et les variét'es qui s'en déduisent.* Paris, 1948.

Riemann Integral

The Riemann integral is the INTEGRAL normally encountered in CALCULUS texts and used by physicists and engineers. Other types of integrals exist (e.g., the LEBESGUE INTEGRAL), but are unlikely to be encountered outside the confines of advanced mathematics texts.

The Riemann integral is based on the JORDAN MEASURE, and defined by taking a limit of a RIEMANN SUM,

$$\int_b^a f(x)\,dx \equiv \lim_{\max \Delta x_k \to 0} \sum_{k=1}^n f(x_k^*)\Delta x_k \tag{1}$$

$$\iint f(x,y)\,dA \equiv \lim_{\max \Delta A_k \to 0} \sum_{k=1}^n f(x_k^*, y_k^*)\Delta A_k \tag{2}$$

$$\iiint f(x,y,z)\,dV \equiv \lim_{\max \Delta V_k \to 0} \sum_{k=1}^n f(x_k^*, y_k^*, z_k^*)\Delta V_k, \tag{3}$$

where $a \leq x \leq b$ and x_k^*, y_k^*, and z_k^* are arbitrary points in the intervals Δx_k, Δy_k, and Δz_k, respectively. The value $\max \Delta x_k$ is called the MESH SIZE of a partition of the interval $[a,b]$ into subintervals Δx_k.

As an example of the application of the Riemann integral definition, find the AREA under the curve $y = x^r$ from 0 to a. Divide (a,b) into n segments, so $\Delta x_k = \frac{b-a}{n} \equiv h$, then

$$f(x_1) = f(0) = 0 \tag{4}$$

$$f(x_2) = f(\Delta x_k) = h^r \tag{5}$$

$$f(x_3) = f(2\Delta x_k) = (2h)^r. \tag{6}$$

By induction

$$f(x_k) = f([k-1]\Delta x_k) = [(k-1)h]^r = h^r(k-1)^r, \tag{7}$$

so

$$f(x_k)\Delta x_k = h^{r+1}(k-1)^r \tag{8}$$

$$\sum_{k=1}^n f(x_k)\Delta x_k = h^{r+1}\sum_{k=1}^n (k-1)^r. \tag{9}$$

For example, take $r = 2$.

$$\begin{aligned}\sum_{k=1}^n f(x_k)\Delta x_k &= h^3 \sum_{k=1}^n (k-1)^2 \\ &= h^3 \left(\sum_{k=1}^n k^2 - 2\sum_{k=1}^n k + \sum_{k=1}^n 1\right) \\ &= h^3 \left[\frac{n(n+1)(2n+1)}{6} - 2\frac{n(n+1)}{2} + n\right],\end{aligned} \tag{10}$$

so

$$\begin{aligned} I &\equiv \lim_{n\to\infty} \sum_{k=1}^n f(x_k{}^*)\Delta x_k = \lim_{n\to\infty} \sum_{k=1}^n f(x_k)\Delta x_k \\ &= \lim_{n\to\infty} h^3 \left[\frac{n(n+1)(2n+1)}{6} - 2\frac{n(n+1)}{2} + n\right] \\ &= a^3 \lim_{n\to\infty} \left[\frac{n(n+1)(2n+1)}{6n^3} - \frac{n(n+1)}{n^3} + \frac{n}{n^3}\right] \\ &= \tfrac{1}{3}a^3. \end{aligned} \tag{11}$$

see also INTEGRAL, RIEMANN SUM

References

Kestelman, H. "Riemann Integration." Ch. 2 in *Modern Theories of Integration, 2nd rev. ed.* New York: Dover, pp. 33–66, 1960.

Riemann's Integral Theorem

Associated with an irreducible curve of GENUS (CURVE) p, there are p LINEARLY INDEPENDENT integrals of the first sort. The ROOTS of the integrands are groups of the canonical series, and every such group will give rise to exactly one integral of the first sort.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 274, 1959.

Riemann-Lebesgue Lemma

Sometimes also called MERCER'S THEOREM.

$$\lim_{n\to\infty} \int_a^b K(\lambda, z) C \sin(nz)\,dz = 0$$

for arbitrarily large C and "nice" $K(\lambda, z)$. Gradshteyn and Ryzhik (1979) state the lemma as follows. If $f(x)$ is integrable on $[\pi, \pi]$, then

$$\lim_{t\to\infty} \int_{-\pi}^{\pi} f(x)\sin(tx)\,dx \to 0$$

and

$$\lim_{t\to\infty} \int_{-\pi}^{\pi} f(x)\cos(tx)\,dx \to 0.$$

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1101, 1979.

Riemann-Mangoldt Function

$$\begin{aligned} f(x) &= \sum_{n>1}^{\infty} \frac{\pi_0(x^{1/n})}{n} \\ &= \mathrm{Li}(x) - \sum_{\substack{\text{nontrivial } \rho \\ \zeta(\rho)=0}} \mathrm{ei}(\rho \ln x) - \ln 2 \\ &\quad + \int_x^\infty \frac{dt}{t(t^2-1)\ln t}, \end{aligned} \tag{1}$$

where $\zeta(z)$ is the Riemann Zeta Function, $\mathrm{Li}(x)$ is the Logarithmic Integral and $\mathrm{ei}(x)$ is the Exponential Integral. The Mangoldt Function is given by

$$\Lambda(n) = \begin{cases} \ln p & \text{if } n = p^m \text{ for } (m \geq 1) \text{ and } p \text{ prime} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$-\frac{\zeta'(x)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} \tag{3}$$

for $\Re[s] > 1$.

$$J(x) = \sum_{n \leq x} \frac{\Lambda(n)}{\ln n}. \tag{4}$$

The Summatory Riemann-Mangoldt function is defined by

$$\psi(x) = \sum_{n \leq x} \Lambda(n) = \theta(x) + \theta(x^{1/2}) + \ldots. \tag{5}$$

see also Prime Number Theorem, Riemann Function

References

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 364–365, 1991.

Riemann Mapping Theorem

Let z_0 be a point in a simply connected region $R \neq \mathbb{C}$. Then there is a unique Analytic Function $w = f(z)$ mapping R one-to-one onto the Disk $|w| < 1$ such that $f(z_0) = 0$ and $f'(z_0) = 0$. The Corollary guarantees that any two simply connected regions except $\mathbb{R}^2$ can be mapped Conformally onto each other.

Riemann's Moduli Problem

Find an Analytic parameterization of the compact Riemann Surfaces in a fixed Homomorphism class. The Ahlfors-Bers Theorem proved that Riemann's Moduli Space gives the solution.

see also Ahlfors-Bers Theorem, Riemann's Moduli Space

Riemann's Moduli Space

Riemann's moduli space R_p is the space of Analytic Equivalence Classes of Riemann Surfaces of fixed Genus p.

see also Ahlfors-Bers Theorem, Riemann's Moduli Problem, Riemann Surface

Riemann *P*-Differential Equation

$$\frac{d^2u}{dz^2} + \left[\frac{1-\alpha-\alpha'}{z-a} + \frac{1-\beta-\beta'}{z-b} + \frac{1-\gamma-\gamma'}{z-c}\right]\frac{du}{dz} + \left[\frac{\alpha\alpha'(a-b)(a-c)}{z-a} + \frac{\beta\beta'(b-c)(b-a)}{z-b} + \frac{\gamma\gamma'(c-a)(c-b)}{z-c}\right]\frac{u}{(z-a)(z-b)(z-c)} = 0,$$

where

$$\alpha + \alpha' + \beta + \beta' + \gamma + \gamma' = 1.$$

Solutions are Riemann *P*-Series (Abramowitz and Stegun 1972, pp. 564–565).

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Riemann's Differential Equation." §15.6 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 564–565, 1972.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 541–543, 1953.

Riemann *P*-Series

The solutions to the Riemann *P*-Differential Equation

$$z = P\left\{\begin{matrix} a & b & c & \\ \alpha & \beta & \gamma & ;z \\ \alpha' & \beta' & \gamma' & \end{matrix}\right\}.$$

Solutions are given in terms of the Hypergeometric Function by

$$u_1 = \left(\frac{z-a}{z-b}\right)^{\alpha}\left(\frac{z-c}{z-b}\right)^{\gamma} \times {}_2F_1(\alpha+\beta+\gamma, \alpha+\beta'+\gamma; 1+\alpha-\alpha'; \lambda)$$

$$u_2 = \left(\frac{z-a}{z-b}\right)^{\alpha'}\left(\frac{z-c}{z-b}\right)^{\gamma} \times {}_2F_1(\alpha'+\beta+\gamma, \alpha'+\beta'+\gamma; 1+\alpha'-\alpha; \lambda)$$

$$u_3 = \left(\frac{z-a}{z-b}\right)^{\alpha}\left(\frac{z-c}{z-b}\right)^{\gamma'} \times {}_2F_1(\alpha+\beta+\gamma', \alpha+\beta'+\gamma'; 1+\alpha-\alpha'; \lambda)$$

$$u_4 = \left(\frac{z-a}{z-b}\right)^{\alpha'}\left(\frac{z-c}{z-b}\right)^{\gamma'} \times {}_2F_1(\alpha'+\beta+\gamma', \alpha'+\beta'+\gamma'; 1+\alpha'-\alpha; \lambda),$$

where

$$\lambda \equiv \frac{(z-a)(c-b)}{(z-b)(c-a)}.$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Riemann's Differential Equation." §15.6 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 564–565, 1972.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 541–543, 1953.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, pp. 283–284, 1990.

Riemann-Roch Theorem

The dimension of a complete series is equal to the sum of the order and index of specialization of any group, less the Genus of the base curve

$$r = N + i + p.$$

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 261, 1959.

Riemann Series Theorem

By a suitable rearrangement of terms, a conditionally convergent SERIES may be made to converge to any desired value, or to DIVERGE.

References

Bromwich, T. J. I'a. and MacRobert, T. M. *An Introduction to the Theory of Infinite Series, 3rd ed.* New York: Chelsea, p. 74, 1991.

Riemann-Siegel Functions

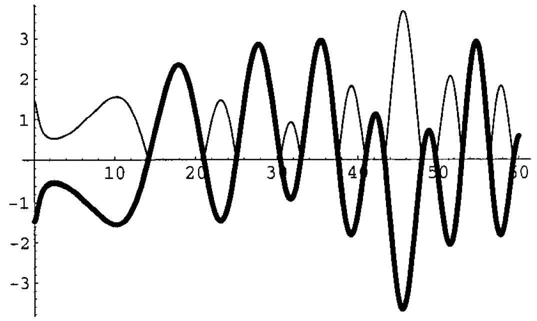

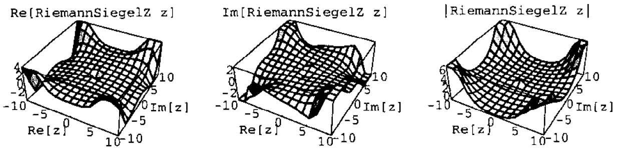

For a REAL POSITIVE t, the Riemann-Siegel Z function is defined by

$$Z(t) \equiv e^{i\vartheta(t)}\zeta(\tfrac{1}{2}+it).$$

The top plot superposes $Z(t)$ (thick line) on $|\zeta(\frac{1}{2}+it)|$, where $\zeta(z)$ is the RIEMANN ZETA FUNCTION.

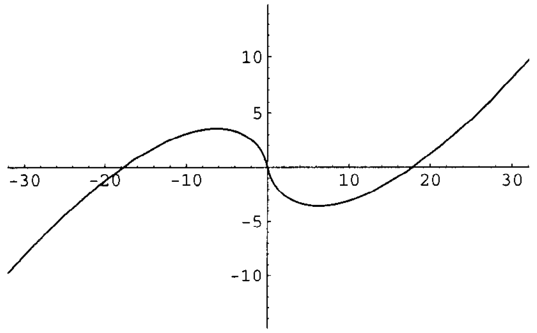

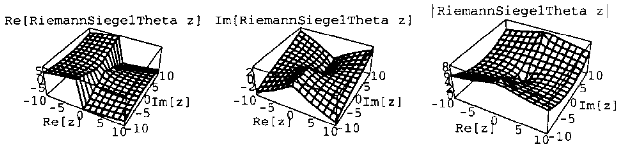

The Riemann-Siegel theta function appearing above is defined by

$$\begin{aligned}\vartheta &\equiv \Im[\ln\Gamma(\tfrac{1}{4}+\tfrac{1}{2}it)-\tfrac{1}{2}t\ln\pi]\\ &= \arg[\Gamma(\tfrac{1}{4}+\tfrac{1}{2}it)]-\tfrac{1}{2}t\ln\pi.\end{aligned}$$

These functions are implemented in *Mathematica*® (Wolfram Research, Champaign, IL) as `RiemannSiegelZ[z]` and `RiemannSiegelTheta[z]`, illustrated above.

see also RIEMANN ZETA FUNCTION

References

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 143, 1991.

Riemann Space

see METRIC SPACE

Riemann Sphere

A 1-D COMPLEX MANIFOLD $\mathbb{C}^*$, which is the one-point compactification of the COMPLEX numbers $\mathbb{C} \cup \{\infty\}$, together with two charts. For all points in the COMPLEX PLANE, the chart is the IDENTITY MAP from the SPHERE (with infinity removed) to the COMPLEX PLANE. For the point at infinity, the chart neighborhood is the sphere (with the ORIGIN removed), and the chart is given by sending infinity to 0 and all other points z to $1/z$.

Riemann-Stieltjes Integral

see STIELTJES INTEGRAL

Riemann Sum

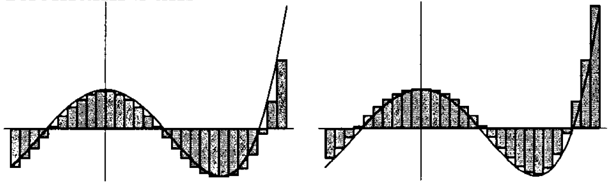

Let a CLOSED INTERVAL $[a,b]$ be partitioned by points $a < x_1 < x_2 < \ldots < x_{n-1} < b$, the lengths of the resulting intervals between the points are denoted Δx_1, Δx_2, ..., Δx_n. Then the quantity

$$\sum_{k=1}^{n} f(x_k^*)\Delta x_k$$

is called a Riemann sum for a given function $f(x)$ and partition. The value $\max \Delta x_k$ is called the MESH SIZE of the partition. If the LIMIT $\max \Delta x_k \to 0$ exists, this limit is known as the Riemann INTEGRAL of $f(x)$ over the interval $[a,b]$. The shaded areas in the above plots show the LOWER and UPPER SUMS for a constant MESH SIZE.

see also LOWER SUM, RIEMANN INTEGRAL, UPPER SUM

Riemann Surface

The Riemann surface S of the ALGEBRAIC FUNCTION FIELD K is the set of nontrivial discrete valuations on K. Here, the set S corresponds to the IDEALS of the RING A of INTEGERS of K over $\mathbb{C}(z)$. (A consists of the elements of K that are ROOTS of MONIC POLYNOMIALS over $\mathbb{C}[z]$.)

see also ALGEBRAIC FUNCTION FIELD, IDEAL, RING

References

Fischer, G. (Ed.). Plates 123–126 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 120–123, 1986.

Riemann Tensor

A TENSOR sometimes known as the RIEMANN-CHRISTOFFEL TENSOR. Let

$$\tilde{D}_s \equiv \frac{\partial}{\partial x^s} - \sum_l \left\{ \begin{matrix} s & u \\ & l \end{matrix} \right\}, \tag{1}$$

where the quantity inside the $\left\{ \begin{matrix} s & u \\ & l \end{matrix} \right\}$ is a CHRISTOFFEL SYMBOL OF THE SECOND KIND. Then

$$R_{pqrs} \equiv \tilde{D}_q \left\{ \begin{matrix} p & r \\ & s \end{matrix} \right\} - \tilde{D}_r \left\{ \begin{matrix} r & q \\ & s \end{matrix} \right\}. \tag{2}$$

Broken down into its simplest decomposition in N-D,

$$\begin{aligned} R_{\lambda\mu\nu\kappa} &= \frac{1}{N-2}(g_{\lambda\nu}R_{\mu\kappa} - g_{\lambda\kappa}R_{\mu\nu} - g_{\mu\nu}R_{\lambda\kappa} + g_{\mu\kappa}R_{\lambda\nu}) \\ &\quad - \frac{R}{(N-1)(N-2)}(g_{\lambda\nu}g_{\mu\kappa} - g_{\lambda\kappa}g_{\mu\nu}) + C_{\lambda\mu\nu\kappa}. \end{aligned} \tag{3}$$

Here, $R_{\mu\nu}$ is the RICCI TENSOR, R is the CURVATURE SCALAR, and $C_{\lambda\mu\nu\kappa}$ is the WEYL TENSOR. In terms of the JACOBI TENSOR ${J^\mu}_{\nu\alpha\beta}$,

$$R^\mu{}_{\alpha\nu\beta} = \tfrac{2}{3}(J^\mu_{\nu\alpha\beta} - J^\mu_{\beta\alpha\nu}). \tag{4}$$

The Riemann tensor is the only tensor that can be constructed from the METRIC TENSOR and its first and second derivatives,

$$R^\alpha{}_{\beta\gamma\delta} = \Gamma^\alpha_{\beta\delta,\gamma} - \Gamma^\alpha_{\beta\gamma,\delta} + \Gamma^\alpha_{\mu\gamma}\Gamma^\mu_{\beta\delta} - \Gamma^\alpha_{\mu\delta} - \Gamma^\alpha_{\beta\mu}{c_{\gamma\delta}}^\mu, \tag{5}$$

where Γ are CONNECTION COEFFICIENTS and c are COMMUTATION COEFFICIENTS. The number of independent coordinates in n-D is

$$C_n \equiv \tfrac{1}{12}n^2(n^2-1), \tag{6}$$

and the number of SCALARS which can be constructed from $R_{\lambda\mu\nu\kappa}$ and $g_{\mu\nu}$ is

$$S_n \equiv \begin{cases} 1 & \text{for } n = 2 \\ \frac{1}{12}n(n-1)(n-2)(n+3) & \text{for } n = 1, n > 2. \end{cases} \tag{7}$$

In 1-D, $R_{1111} = 0$.

n	C_n	S_n
1	0	0
2	1	1
3	6	3
4	20	14

see also BIANCHI IDENTITIES, CHRISTOFFEL SYMBOL OF THE SECOND KIND, COMMUTATION COEFFICIENT, CONNECTION COEFFICIENT, CURVATURE SCALAR, GAUSSIAN CURVATURE, JACOBI TENSOR, PETROV NOTATION, RICCI TENSOR, WEYL TENSOR

Riemann Theta Function

Let the IMAGINARY PART of a $g \times g$ MATRIX F be POSITIVE DEFINITE, and $\mathbf{m} = (m_1, \ldots, m_g)$ be a row VECTOR with coefficients in $\mathbb{Z}$. Then the Riemann theta function is defined by

$$\vartheta(u) = \sum_{\mathbf{m}} \exp[2\pi i(\mathbf{m}^{\mathrm{T}}u + \tfrac{1}{2}\mathsf{F}^{\mathrm{T}}\mathbf{m})].$$

see also RAMANUJAN THETA FUNCTIONS, THETA FUNCTION

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 9, 1980.

Riemann Weighted Prime-Power Counting Function

The Riemann weighted prime-power counting function is defined by

$$\begin{aligned} J_2(x) &\equiv \begin{cases} \pi(x) + \frac{1}{2}\pi(x^{1/2}) + \frac{1}{3}\pi(x^{1/3}) + \ldots - \frac{1}{2m} & \text{for } p^m \text{ with } p \text{ a prime} \\ \pi(x) + \frac{1}{2}\pi(x^{1/2}) + \frac{1}{3}\pi(x^{1/3}) + \ldots & \text{otherwise} \end{cases} && (1) \\ &= \lim_{t\to\infty} \frac{1}{2\pi i}\int_{2-iT}^{2+iT} \frac{x^s}{s}\ln\zeta(s)\,ds. && (2) \end{aligned}$$

The PRIME COUNTING FUNCTION is given in terms of $J_2(x)$ by

$$\pi(x) = \sum_{n=1}^\infty \frac{\mu(n)}{n} J_2(x^{1/n}). \tag{3}$$

The function also satisfies the identity

$$\frac{\ln\zeta(s)}{s} = \int_1^\infty J_2(x)x^{-s-1}\,dx. \tag{4}$$

see also MANGOLDT FUNCTION, PRIME COUNTING FUNCTION, RIEMANN'S FORMULA

Riemann Xi Function

see XI FUNCTION

Riemann Zeta Function

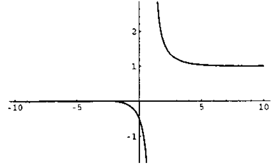

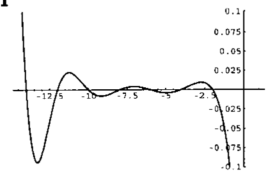

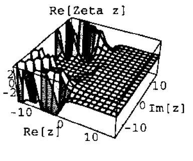

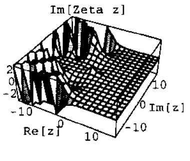

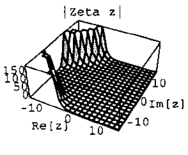

The Riemann zeta function can be defined by the integral

$$\zeta(x) \equiv \frac{1}{\Gamma(x)} \int_0^\infty \frac{u^{x-1}}{e^u - 1}\, du, \tag{1}$$

where $x > 1$. If x is an INTEGER n, then

$$\frac{u^{n-1}}{e^u - 1} = \frac{e^{-u}u^{n-1}}{1 - e^{-u}} = e^{-u}u^{n-1} \sum_{k=1}^\infty e^{-ku}u^{n-1}, \tag{2}$$

so

$$\int_0^\infty \frac{u^{n-1}}{e^u - 1}\, du = \sum_{k=1}^\infty \int_0^\infty e^{-ku}u^{n-1}\, du. \tag{3}$$

Let $y \equiv ku$, then $dy = k\, du$ and

$$\begin{aligned} \zeta(n) &= \frac{1}{\Gamma(n)} \sum_{k=1}^\infty \int_0^\infty e^{-ku}u^{n-1}\, du \\ &= \frac{1}{\Gamma(n)} \sum_{k=1}^\infty \int_0^\infty e^{-y} \left(\frac{y}{k}\right)^{n-1} \frac{dy}{k} \\ &= \frac{1}{\Gamma(n)} \sum_{k=1}^\infty \frac{1}{k^n} \int_0^\infty e^{-y}y^{n-1}\, dy, \end{aligned} \tag{4}$$

where $\Gamma(n)$ is the GAMMA FUNCTION. Integrating the final expression in (4) gives $\Gamma(n)$, which cancels the factor $1/\Gamma(n)$ and gives the most common form of the Riemann zeta function,

$$\zeta(n) = \sum_{k=1}^\infty \frac{1}{k^n}. \tag{5}$$

At $n = 1$, the zeta function reduces to the HARMONIC SERIES (which diverges), and therefore has a singularity. In the COMPLEX PLANE, trivial zeros occur at -2, -4, -6, ..., and nontrivial zeros at

$$s \equiv \sigma + it \tag{6}$$

for $0 \leq \sigma \leq 1$. The figures below show the structure of $\zeta(z)$ by plotting $|\zeta(z)|$ and $1/|\zeta(z)|$.

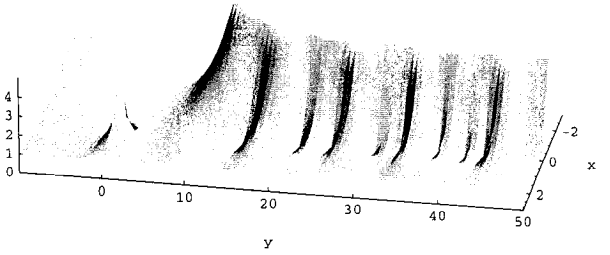

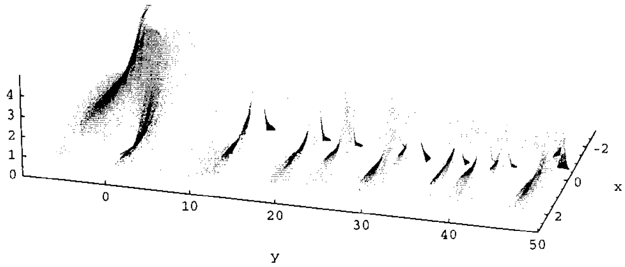

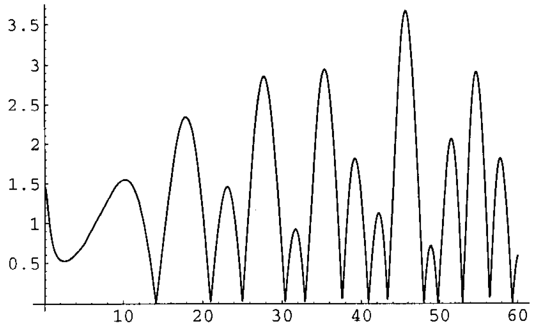

The RIEMANN HYPOTHESIS asserts that the nontrivial ROOTS of $\zeta(s)$ all have REAL PART $\sigma = \Re[s] = 1/2$, a line called the "CRITICAL STRIP." This is known to be true for the first 1.5×10^{12} roots (Brent *et al.* 1979). The above plot shows $|\zeta(1/2 + it)|$ for t between 0 and 60. As can be seen, the first few nontrivial zeros occur at $t =$ 14.134725, 21.022040, 25.010858, 30.424876, 32.935062, 37.586178, ... (Wagon 1991, pp. 361–362 and 367–368).

The Riemann zeta function can also be defined in terms of MULTIPLE INTEGRALS by

$$\zeta(n) = \underbrace{\int_0^1 \cdots \int_0^1}_{n} \frac{\prod_{i=1}^n dx_i}{1 - \prod_{i=1}^n x_i}. \tag{7}$$

The Riemann zeta function can be split up into

$$\zeta(\tfrac{1}{2} + it) = z(t)e^{-i\vartheta(t)}, \tag{8}$$

where $z(t)$ and $\vartheta(t)$ are the RIEMANN-SIEGEL FUNCTIONS. An additional identity is

$$\lim_{s \to 1} \zeta(s) - \frac{1}{s-1} = \gamma, \tag{9}$$

where γ is the EULER-MASCHERONI CONSTANT.

The Riemann zeta function is related to the DIRICHLET LAMBDA FUNCTION $\lambda(\nu)$ and DIRICHLET ETA FUNCTION $\eta(\nu)$ by

$$\frac{\zeta(\nu)}{2^\nu} = \frac{\lambda(\nu)}{2^\nu - 1} = \frac{\eta(\nu)}{2^\nu - 2} \tag{10}$$

and

$$\zeta(\nu) + \eta(\nu) = 2\lambda(\nu) \tag{11}$$

(Spanier and Oldham 1987). It is related to the LIOUVILLE FUNCTION $\lambda(n)$ by

$$\frac{\zeta(2s)}{\zeta(s)} = \sum_{n=1}^\infty \frac{\lambda(n)}{n^s} \tag{12}$$

(Lehman 1960, Hardy and Wright 1979). Furthermore,

$$\frac{\zeta^2(s)}{\zeta(2s)} = \sum_{n=1}^\infty \frac{2^{\omega(n)}}{n^s}, \tag{13}$$

where $\omega(n) = \sigma_0(n)$ is the number of different prime factors of n (Hardy and Wright 1979).

A generalized Riemann zeta function $\zeta(s,a)$ known as the Hurwitz Zeta Function can also be defined such that

$$\zeta(s) \equiv \zeta(s,0). \tag{14}$$

The Riemann zeta function may be computed analytically for Even n using either Contour Integration or Parseval's Theorem with the appropriate Fourier Series. An interesting formula involving the product of Primes was first discovered by Euler in 1737,

$$\begin{aligned}\zeta(x)(1-2^{-x}) &= \left(1+\frac{1}{2^x}+\frac{1}{3^x}+\ldots\right)\left(1-\frac{1}{2^x}\right)\\ &= \left(1+\frac{1}{2^x}+\frac{1}{3^x}+\ldots\right)-\left(\frac{1}{2^x}+\frac{1}{4^x}+\frac{1}{6^x}+\ldots\right)\end{aligned} \tag{15}$$

$$\begin{aligned}&\zeta(x)(1-2^{-x})(1-3^{-x})\\ &= \left(1+\frac{1}{3^x}+\frac{1}{5^x}+\frac{1}{7^x}+\ldots\right)-\left(\frac{1}{3^x}+\frac{1}{9^x}+\frac{1}{15^x}+\ldots\right)\end{aligned} \tag{16}$$

$$\begin{aligned}&\zeta(x)(1-2^{-x})(1-3^{-x})\cdots(1-p^{-x})\cdots\\ &\qquad = \zeta(x)\prod_{n=2}^{\infty}(1-p^{-x}) = 1.\end{aligned} \tag{17}$$

Here, each subsequent multiplication by the next Prime p leaves only terms which are Powers of p^{-x}. Therefore,

$$\zeta(x) = \left[\prod_{p=2}^{\infty}(1-p^{-x})\right]^{-1}, \tag{18}$$

where p runs over all Primes. Euler's product formula can also be written

$$\zeta(s) = (1-2^{-s})^{-1} \prod_{\substack{q\equiv 1\\ (\mathrm{mod}\ 4)}} (1-q^{-s})^{-1} \prod_{\substack{r\equiv 3\\ (\mathrm{mod}\ 4)}} (1-r^{-s})^{-1}. \tag{19}$$

A few sum identities involving $\zeta(n)$ are

$$\sum_{n=2}^{\infty}[\zeta(n)-1] = 1 \tag{20}$$

$$\sum_{n=2}^{\infty}(-1)^n[\zeta(n)-1] = \tfrac{1}{2}. \tag{21}$$

The Riemann zeta function is related to the Gamma Function $\Gamma(z)$ by

$$\Gamma\left(\frac{s}{2}\right)\pi^{-z/2}\zeta(s) = \Gamma\left(\frac{1-s}{2}\right)\pi^{-(1-s)/2}\zeta(1-s). \tag{22}$$

$\zeta(n)$ was proved to be transcendental for all even n by Euler. Apéry (1979) proved $\zeta(3)$ to be Irrational with the aid of the k^{-3} sum formula below. As a result, $\zeta(3)$ is sometimes called Apéry's Constant.

$$\zeta(2) = 3\sum_{k=1}^{\infty}\frac{1}{k^2\binom{2k}{k}} \tag{23}$$

$$\zeta(3) = \frac{5}{2}\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k^3\binom{2k}{k}} \tag{24}$$

$$\zeta(4) = \frac{36}{17}\sum_{k=1}^{\infty}\frac{1}{k^4\binom{2k}{k}} \tag{25}$$

(Guy 1994, p. 257). A relation of the form

$$\zeta(5) = Z_5\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k^5\binom{2k}{k}} \tag{26}$$

has been searched for with Z_5 a Rational or Algebraic Number, but if Z_5 is a Root of a Polynomial of degree 25 or less, then the Euclidean norm of the coefficients must be larger than 2×10^{37} (Bailey, Bailey and Plouffe). Therefore, no such sums are known for $\zeta(n)$ are known for $n \geq 5$.

The zeta function is defined for $\Re[s] > 1$, but can be analytically continued to $\Re[s] > 0$ as follows

$$\begin{aligned}\sum_{n=1}^{\infty}(-1)^n n^{-s} + \sum_{n=1}^{\infty} n^{-s} &= 2\sum_{n=2,4,\ldots}^{\infty} n^{-s}\\ &= 2\sum_{k=1}^{\infty}(2k)^{-s} = 2^{1-s}\sum_{k=1}^{\infty}k^{-s}\end{aligned} \tag{27}$$

$$\sum_{n=1}^{\infty}(-1)^n n^{-s} + \zeta(s) = 2^{1-s}\zeta(s) \tag{28}$$

$$\zeta(s) = \frac{1}{1-2^{1-s}}\sum_{n=1}^{\infty}(-1)^n n^{-s}. \tag{29}$$

The Derivative of the Riemann zeta function is defined by

$$\zeta'(s) = -s\sum_{k=1}^{\infty}k^{-s}\ln k = -\sum_{k=2}^{\infty}\frac{\ln k}{k^s}. \tag{30}$$

As $s \to 0$,

$$\zeta'(0) = -\tfrac{1}{2}\ln(2\pi). \tag{31}$$

For Even $n \equiv 2k$,

$$\zeta(n) = \frac{2^{n-1}|B_n|\pi^n}{n!}, \tag{32}$$

where B_n is a Bernoulli Number. Another intimate connection with the Bernoulli Numbers is provided by

$$B_n = (-1)^{n+1}n\zeta(1-n). \tag{33}$$

No analytic form for $\zeta(n)$ is known for ODD $n \equiv 2k+1$, but $\zeta(2k+1)$ can be expressed as the sum limit

$$\zeta(2k+1) = (\tfrac{1}{2}\pi)^{2k+1} \lim_{t\to\infty} \frac{1}{t^{2k+1}} \sum_{i=1}^{\infty} \left[\cot\left(\frac{i}{2t+1}\right)\right]^{2k+1} \quad (34)$$

(Stark 1974). The values for the first few integral arguments are

$$\zeta(0) \equiv -\tfrac{1}{2}$$
$$\zeta(1) = \infty$$
$$\zeta(2) = \frac{\pi^2}{6}$$
$$\zeta(3) = 1.2020569032\ldots$$
$$\zeta(4) = \frac{\pi^4}{90}$$
$$\zeta(5) = 1.0369277551\ldots$$
$$\zeta(6) = \frac{\pi^6}{945}$$
$$\zeta(7) = 1.0083492774\ldots$$
$$\zeta(8) = \frac{\pi^8}{9450}$$
$$\zeta(9) = 1.0020083928\ldots$$
$$\zeta(10) = \frac{\pi^{10}}{93{,}555}.$$

Euler gave $\zeta(2)$ to $\zeta(26)$ for EVEN n, and Stieltjes (1993) determined the values of $\zeta(2)$, ..., $\zeta(70)$ to 30 digits of accuracy in 1887. The denominators of $\zeta(2n)$ for $n = 1, 2, \ldots$ are 6, 90, 945, 9450, 93555, 638512875, ... (Sloane's A002432).

Using the LLL ALGORITHM, Plouffe (inspired by Zucker 1979, Zucker 1984, and Berndt 1988) has found some beautiful infinite sums for $\zeta(n)$ with ODD n. Let

$$S_\pm(n) \equiv \sum_{k=1}^{\infty} \frac{1}{k^n(e^{2\pi k} \pm 1)}, \quad (35)$$

then

$$\zeta(3) = \tfrac{7}{180}\pi^3 - 2S_-(3) \quad (36)$$
$$\zeta(5) = \tfrac{1}{294}\pi^5 - \tfrac{72}{35}S_-(5) - \tfrac{2}{35}S_+(5) \quad (37)$$
$$\zeta(7) = \tfrac{19}{56700}\pi^7 - 2S_-(7) \quad (38)$$
$$\zeta(9) = \tfrac{125}{3704778}\pi^9 - \tfrac{992}{495}S_-(9) - \tfrac{2}{495}S_+(9) \quad (39)$$
$$\zeta(11) = \tfrac{1453}{425675250}\pi^{11} - 2S_-(11) \quad (40)$$
$$\zeta(13) = \tfrac{89}{257432175}\pi^{13} - \tfrac{16512}{8255}S_-(13) - \tfrac{2}{8255}S_+(13) \quad (41)$$
$$\zeta(15) = \tfrac{13687}{390769879500}\pi^{15} - 2S_-(15) \quad (42)$$
$$\zeta(17) = \tfrac{397549}{112024529867250}\pi^{17} - \tfrac{261632}{130815}S_-(17) - \tfrac{2}{130815}S_+(17) \quad (43)$$
$$\zeta(19) = \tfrac{7708537}{21438612514068750}\pi^{19} - 2S_-(19) \quad (44)$$
$$\zeta(21) = \tfrac{68529640373}{1881063815762259253125}\pi^{21} - \tfrac{4196352}{2098175}S_-(21) - \tfrac{2}{2098175}S_+(21) \quad (45)$$

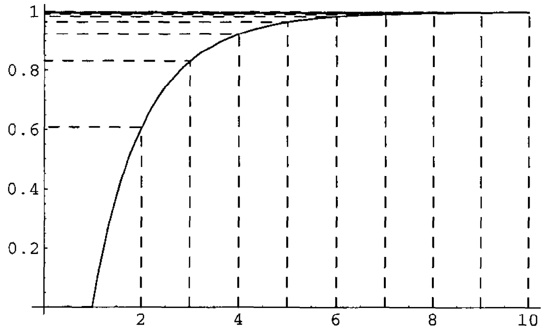

The inverse of the RIEMANN ZETA FUNCTION $1/\zeta(p)$ is the asymptotic density of pth-powerfree numbers (i.e., SQUAREFREE numbers, CUBEFREE numbers, etc.). The following table gives the number $Q_p(n)$ of pth-powerfree numbers $\leq n$ for several values of n.

p	$1/\zeta(p)$	10	100	10^3	10^4	10^5	10^6
2	0.607927	7	61	608	6083	60794	607926
3	0.831907	9	85	833	8319	83190	831910
4	0.923938	10	93	925	9240	92395	923939
5	0.964387	10	97	965	9645	96440	964388
6	0.982953	10	99	984	9831	98297	982954

The value for $\zeta(2)$ can be found using a number of different techniques (Apostol 1983, Choe 1987, Giesy 1972, Holme 1970, Kimble 1987, Knopp and Schur 1918, Kortram 1996, Matsuoka 1961, Papadimitriou 1973, Simmons 1992, Stark 1969, Stark 1970, Yaglom and Yaglom 1987). The problem of finding this value analytically is sometimes known as the BASLER PROBLEM (Castellanos 1988). Yaglom and Yaglom (1987), Holme (1970), and Papadimitrou (1973) all derive the result from DE MOIVRE'S IDENTITY or related identities.

Consider the FOURIER SERIES of $f(x) = x^{2n}$

$$f(x) = \tfrac{1}{2}a_0 + \sum_{m=1}^{\infty} a_m \cos(mx) + \sum_{m=1}^{\infty} b_m \sin(mx), \quad (46)$$

which has coefficients given by

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,dx = \frac{2}{\pi}\int_0^{\pi} x^{2n}\,dx = \frac{2}{\pi}\left[\frac{x^{2n+1}}{2n+1}\right]_0^{\pi} = \frac{2\pi^{2n}}{2n+1} \quad (47)$$

$$a_m = \frac{1}{\pi}\int_{\pi}^{\pi} x^{2n}\cos(mx)\,dx = \frac{2}{\pi}\int_0^{\pi} x^{2n}\cos(mx)\,dx \quad (48)$$

$$b_m = \frac{1}{\pi}\int_{-\pi}^{\pi} x^{2n}\sin(mx)\,dx = 0, \quad (49)$$

where the latter is true since the integrand is ODD. Therefore, the FOURIER SERIES is given explicitly by

$$x^{2n} = \frac{\pi^{2n}}{2n+1} + \sum_{m=1}^{\infty} a_m \cos(mx). \quad (50)$$

Now, a_m is given by the COSINE INTEGRAL

$$a_m = \frac{2}{\pi}(-1)^{n+1}(2n)! \left[\sin(mx) \sum_{k=0}^{n} \frac{(-1)^k}{(2k)!m^{2n-2k+1}} x^{2k} \right. \\ \left. + \cos(mx) \sum_{k=1}^{n} \frac{(-1)^{k+1}}{(2k-3)!m^{2n-2k+2}} x^{2k-1} \right]_0^\pi . \quad (51)$$

But $\cos(m\pi) = (-1)^m$, and $\sin(m\pi) = \sin 0 = 0$, so

$$a_m = \frac{2}{\pi}(-1)^{n+1}(2n)!(-1)^m \sum_{k=1}^{n} \frac{(-1)^{k+1}}{(2k-3)!m^{2n-2k+2}} \pi^{2k-1} \\ = (-1)^{m+n} 2(2n)! \sum_{k=1}^{n} \frac{(-1)^k}{(2k-3)!m^{2n-2k+2}} \pi^{2k-2}. \quad (52)$$

Now, if $n = 1$,

$$a_m = (-1)^{m+1} 2(2!) \sum_{k=1}^{1} \frac{(-1)^k}{(2k-3)!m^{4-2k}} \pi^{2k-2} \\ = 4(-1)^{m+1} \frac{(-1)}{(-1)!m^2} \pi^0 = \frac{4(-1)^m}{m^2}, \quad (53)$$

so the FOURIER SERIES is

$$x^2 = \frac{\pi^2}{3} + 4 \sum_{m=1}^{\infty} \frac{(-1)^m \cos(mx)}{m^2}. \quad (54)$$

Letting $x \equiv \pi$ gives $\cos(m\pi) = (-1)^m$, so

$$\pi^2 = \frac{\pi^2}{3} + 4 \sum_{m=1}^{\infty} \frac{1}{m^2}, \quad (55)$$

and we have

$$\zeta(2) = \sum_{m=1}^{\infty} \frac{1}{m^2} = \frac{\pi^2}{6}. \quad (56)$$

Higher values of n can be obtained by finding a_m and proceeding as above.

The value $\zeta(2)$ can also be found simply using the ROOT LINEAR COEFFICIENT THEOREM. Consider the equation $\sin z = 0$ and expand sin in a MACLAURIN SERIES

$$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} + \ldots = 0 \quad (57)$$

$$0 = 1 - \frac{z^2}{3!} + \frac{z^4}{5!} + \ldots = 1 - \frac{w}{3!} + \frac{w^2}{5!} + \ldots, \quad (58)$$

where $w \equiv z^2$. But the zeros of $\sin(z)$ occur at π, 2π, 3π, $\ldots$, so the zeros of $\sin w = \sin\sqrt{z}$ occur at π^2, $(2\pi)^2$, $\ldots$. Therefore, the sum of the roots equals the COEFFICIENT of the leading term

$$\frac{1}{\pi^2} + \frac{1}{(2\pi)^2} + \frac{1}{(3\pi^2)} + \ldots = \frac{1}{3!} = \frac{1}{6}, \quad (59)$$

which can be rearranged to yield

$$\zeta(2) = \frac{\pi^2}{6}. \quad (60)$$

Yet another derivation (Simmons 1992) evaluates the integral using the integral

$$I = \int_0^1 \int_0^1 \frac{dx\,dy}{1-xy} = \int_0^1 \int_0^1 (1 + xy + x^2y^2 + \ldots)\,dx\,dy \\ = \int_0^1 [(x + \tfrac{1}{2}x^2y + \tfrac{1}{3}x^3y^2 + \ldots)]_0^1\,dy \\ = \int_0^1 (1 + \tfrac{1}{2}y + \tfrac{1}{3}y^2 + \ldots)\,dy \\ = \left[y + \frac{y^2}{2^2} + \frac{y^3}{3^2} + \ldots\right]_0^1 = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \ldots. \quad (61)$$

To evaluate the integral, rotate the coordinate system by $\pi/4$ so

$$x = u\cos\theta - v\sin\theta = \tfrac{1}{2}\sqrt{2}\,(u-v) \quad (62)$$

$$y = u\sin\theta + v\cos\theta = \tfrac{1}{2}\sqrt{2}\,(u+v) \quad (63)$$

and

$$xy = \tfrac{1}{2}(u^2 - v^2) \quad (64)$$

$$1 - xy = \tfrac{1}{2}(2 - u^2 + v^2). \quad (65)$$

Then

$$I = 4 \int_0^{\sqrt{2}/2} \int_0^u \frac{du\,dv}{2 - u^2 + v^2} \\ + 4 \int_{\sqrt{2}/2}^{\sqrt{2}} \int_0^{\sqrt{2}-u} \frac{du\,dv}{2 - u^2 + v^2} \equiv I_1 + I_2. \quad (66)$$

Now compute the integrals I_1 and I_2.

$$I_1 = 4 \int_0^{\sqrt{2}/2} \left[\int_0^u \frac{dv}{2 - u^2 + v^2}\right] du \\ = 4 \int_0^{\sqrt{2}\,2} \left[\frac{1}{\sqrt{2-u^2}} \tan^{-1}\left(\frac{v}{\sqrt{2-u^2}}\right)\right]_0^u du \\ = 4 \int_0^{\sqrt{2}/2} \frac{1}{\sqrt{2-u^2}} \tan^{-1}\left(\frac{u}{\sqrt{2-u^2}}\right) du. \quad (67)$$

Make the substitution

$$u = \sqrt{2}\,\sin\theta \quad (68)$$

$$\sqrt{2-u^2} = \sqrt{2}\,\cos\theta \quad (69)$$

$$du = \sqrt{2}\,\cos\theta\,d\theta, \quad (70)$$

so

$$\tan^{-1}\left(\frac{u}{\sqrt{2-u^2}}\right) = \tan^{-1}\left(\frac{\sqrt{2}\,\sin\theta}{\sqrt{2}\,\cos\theta}\right) = \theta \qquad (71)$$

and

$$I_1 = 4\int_0^{\pi/6} \frac{1}{\sqrt{2}\,\cos\theta}\theta\sqrt{2}\,\cos\theta\,d\theta = 2[\theta^2]_0^{\pi/6} = \frac{\pi^2}{18}. \qquad (72)$$

I_2 can also be computed analytically,

$$\begin{aligned} I_2 &= 4\int_{\sqrt{2}/2}^{\sqrt{2}}\left[\int_0^{\sqrt{2}-u}\frac{dv}{2-u^2+v^2}\right]du \\ &= 4\int_{\sqrt{2}/2}^{\sqrt{2}}\left[\frac{1}{\sqrt{2-u^2}}\tan^{-1}\left(\frac{v}{\sqrt{2-u^2}}\right)\right]_0^{\sqrt{2}-u} du \\ &= 4\int_{\sqrt{2}/2}^{\sqrt{2}}\frac{1}{\sqrt{2-u^2}}\tan^{-1}\left(\frac{\sqrt{2}-u}{\sqrt{2-u^2}}\right)du. \qquad (73) \end{aligned}$$

But

$$\begin{aligned} \tan^{-1}\left(\frac{\sqrt{2}-u}{\sqrt{2-u^2}}\right) &= \tan^{-1}\left(\frac{\sqrt{2}-\sqrt{2}\,\sin\theta}{\sqrt{2}\,\cos\theta}\right) \\ &= \tan\left(\frac{1-\sin\theta}{\cos\theta}\right) = \tan^{-1}\left(\frac{\cos\theta}{1+\sin\theta}\right) \\ &= \tan^{-1}\left[\frac{\sin(\tfrac{1}{2}\pi-\theta)}{1+\cos(\tfrac{1}{2}\pi-\theta)}\right] \\ &= \tan^{-1}\left\{\frac{2\sin[\tfrac{1}{2}(\tfrac{1}{2}\pi-\theta)]\cos[\tfrac{1}{2}(\tfrac{1}{2}\pi-\theta)]}{2\cos^2[\tfrac{1}{2}(\tfrac{1}{2}\pi-\theta)]}\right\} \\ &= \tfrac{1}{2}(\tfrac{1}{2}\pi-\theta), \qquad (74) \end{aligned}$$

so

$$\begin{aligned} I_2 &= 4\int_{\pi/6}^{\pi/2}\frac{1}{\sqrt{2}\,\cos\theta}(\tfrac{1}{4}\pi-\tfrac{1}{2}\theta)\sqrt{2}\,\cos\theta\,d\theta \\ &= 4\left[\tfrac{1}{4}\pi\theta-\tfrac{1}{4}\theta^2\right]_{\pi/6}^{\pi/2} \\ &= 4\left[\left(\frac{\pi^2}{8}-\frac{\pi^2}{16}\right)-\left(\frac{\pi^2}{24}-\frac{\pi^2}{144}\right)\right] = \frac{\pi^2}{9}. \qquad (75) \end{aligned}$$

Combining I_1 and I_2 gives

$$\zeta(2) = I_1 + I_2 = \frac{\pi^2}{18} + \frac{\pi^2}{9} = \frac{\pi^2}{6}. \qquad (76)$$

see also Abel's Functional Equation, Debye Functions, Dirichlet Beta Function, Dirichlet Eta Function, Dirichlet Lambda Function, Harmonic Series, Hurwitz Zeta Function, Khintchine's Constant, Lehmer's Phenomenon, Psi Function, Riemann Hypothesis, Riemann P-Series, Riemann-Siegel Functions, Stieltjes Constants, Xi Function

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Riemann Zeta Function and Other Sums of Reciprocal Powers." §23.2 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 807–808, 1972.

Apéry, R. "Irrationalité de $\zeta(2)$ et $\zeta(3)$." *Astérisque* **61**, 11–13, 1979.

Apostol, T. M. "A Proof that Euler Missed: Evaluating $\zeta(2)$ the Easy Way." *Math. Intel.* **5**, 59–60, 1983.

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 332–335, 1985.

Ayoub, R. "Euler and the Zeta Function." *Amer. Math. Monthly* **71**, 1067–1086, 1974.

Bailey, D. H. "Multiprecision Translation and Execution of Fortran Programs." *ACM Trans. Math. Software.* To appear.

Bailey, D. and Plouffe, S. "Recognizing Numerical Constants." `http://www.cecm.sfu.ca/organics/papers/bailey`.

Berndt, B. C. Ch. 14 in *Ramanujan's Notebooks, Part II.* New York: Springer-Verlag, 1988.

Borwein, D. and Borwein, J. "On an Intriguing Integral and Some Series Related to $\zeta(4)$." *Proc. Amer. Math. Soc.* **123**, 1191–1198, 1995.

Brent, R. P.; van der Lune, J.; te Riele, H. J. J.; and Winter, D. T. "On the Zeros of the Riemann Zeta Function in the Critical Strip 1." *Math. Comput.* **33**, 1361–1372, 1979.

Castellanos, D. "The Ubiquitous Pi. Part I." *Math. Mag.* **61**, 67–98, 1988.

Choe, B. R. "An Elementary Proof of $\sum_{n=1}^\infty \frac{1}{n^2} = \frac{\pi^2}{6}$." *Amer. Math. Monthly* **94**, 662–663, 1987.

Davenport, H. *Multiplicative Number Theory, 2nd ed.* New York: Springer-Verlag, 1980.

Edwards, H. M. *Riemann's Zeta Function.* New York: Academic Press, 1974.

Farmer, D. W. "Counting Distinct Zeros of the Riemann Zeta-Function." *Electronic J. Combinatorics* **2**, R1, 1–5, 1995. `http://www.combinatorics.org/Volume_2/volume2.html#R1`.

Giesy, D. P. "Still Another Proof that $\sum 1/k^2 = \pi^2/6$." *Math. Mag.* **45**, 148–149, 1972.

Guy, R. K. "Series Associated with the ζ-Function." §F17 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 257–258, 1994.

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, p. 255, 1979.

Holme, F. "Ein enkel beregning av $\sum_{k=1}^\infty \frac{1}{k^2}$." *Nordisk Mat. Tidskr.* **18**, 91–92 and 120, 1970.

Ivic, A. A. *The Riemann Zeta-Function.* New York: Wiley, 1985.

Ivic, A. A. *Lectures on Mean Values of the Riemann Zeta Function.* Berlin: Springer-Verlag, 1991.

Karatsuba, A. A. and Voronin, S. M. *The Riemann Zeta-Function.* Hawthorne, NY: De Gruyter, 1992.

Katayama, K. "On Ramanujan's Formula for Values of Riemann Zeta-Function at Positive Odd Integers." *Acta Math.* **22**, 149–155, 1973.

Kimble, G. "Euler's Other Proof." *Math. Mag.* **60**, 282, 1987.

Knopp, K. and Schur, I. "Über die Herleitug der Gleichung $\sum_{n=1}^\infty \frac{1}{n^2} = \frac{\pi^2}{6}$." *Archiv der Mathematik u. Physik* **27**, 174–176, 1918.

Kortram, R. A. "Simple Proofs for $\sum_{k=1}^\infty \frac{1}{k^2} = \frac{\pi^2}{6}$ and $\sin x = x\prod_{k=1}^\infty\left(1-\frac{x^2}{k^2\pi^2}\right)$." *Math. Mag.* **69**, 122–125, 1996.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 35, 1983.

Lehman, R. S. "On Liouville's Function." *Math. Comput.* **14**, 311–320, 1960.

Matsuoka, Y. "An Elementary Proof of the Formula $\sum_{k=1}^{\infty}\frac{1}{k^2}=\frac{\pi^2}{6}$." *Amer. Math. Monthly* **68**, 486–487, 1961.

Papadimitriou, I. "A Simple Proof of the Formula $\sum_{k=1}^{\infty}\frac{1}{k^2}=\frac{\pi^2}{6}$." *Amer. Math. Monthly* **80**, 424–425, 1973.

Patterson, S. J. *An Introduction to the Theory of the Riemann Zeta-Function.* New York: Cambridge University Press, 1988.

Plouffe, S. "Identities Inspired from Ramanujan Notebooks." `http://www.lacim.uqam.ca/plouffe/identities.html`.

Simmons, G. F. "Euler's Formula $\sum_1^\infty 1/n^2=\pi^2/6$ by Double Integration." Ch. B. 24 in *Calculus Gems: Brief Lives and Memorable Mathematics.* New York: McGraw-Hill, 1992.

Sloane, N. J. A. Sequence A002432/M4283 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Spanier, J. and Oldham, K. B. "The Zeta Numbers and Related Functions." Ch. 3 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 25–33, 1987.

Stark, E. L. "Another Proof of the Formula $\sum_{k=1}^{\infty}\frac{1}{k^2}=\frac{\pi^2}{6}$." *Amer. Math. Monthly* **76**, 552–553, 1969.

Stark, E. L. "$1-\frac{1}{4}+\frac{1}{9}-\frac{1}{16}+\ldots=\frac{\pi^2}{12}$." *Praxis Math.* **12**, 1–3, 1970.

Stark, E. L. "The Series $\sum_{k=1}^{\infty}k^{-s}$ $s=2, 3, 4, \ldots$, Once More." *Math. Mag.* **47**, 197–202, 1974.

Stieltjes, T. J. *Oeuvres Complètes, Vol. 2* (Ed. G. van Dijk.) New York: Springer-Verlag, p. 100, 1993.

Titchmarsh, E. C. *The Zeta-Function of Riemann.* New York: Stechert-Hafner Service Agency, 1964.

Titchmarsh, E. C. and Heath-Brown, D. R. *The Theory of the Riemann Zeta-Function, 2nd ed.* Oxford, England: Oxford University Press, 1986.

Vardi, I. "The Riemann Zeta Function." Ch. 8 in *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 141–174, 1991.

Wagon, S. "The Evidence: Where Are the Zeros of Zeta of *s*?" *Math. Intel.* **8**, 57–62, 1986.

Wagon, S. "The Riemann Zeta Function." §10.6 in *Mathematica in Action.* New York: W. H. Freeman, pp. 353–362, 1991.

Yaglom, A. M. and Yaglom, I. M. Problem 145 in *Challenging Mathematical Problems with Elementary Solutions, Vol. 2.* New York: Dover, 1987.

Zucker, I. J. "The Summation of Series of Hyperbolic Functions." *SIAM J. Math. Anal.* **10**, 192–206, 1979.

Zucker, I. J. "Some Infinite Series of Exponential and Hyperbolic Functions." *SIAM J. Math. Anal.* **15**, 406–413, 1984.

Riemannian Geometry

The study of MANIFOLDS having a complete RIEMANNIAN METRIC. Riemannian geometry is a general space based on the LINE ELEMENT

$$ds=F(x^1,\ldots,x^n;dx^1,\ldots,dx^n),$$

with $F(x,y)>0$ for $y\neq 0$ a function on the TANGENT BUNDLE TM. In addition, F is homogeneous of degree 1 in y and of the form

$$F^2=g_{ij}(x)\,dx^i\,dx^j$$

(Chern 1996). If this restriction is dropped, the resulting geometry is called FINSLER GEOMETRY.

References

Besson, G.; Lohkamp, J.; Pansu, P.; and Petersen, P. *Riemannian Geometry.* Providence, RI: Amer. Math. Soc., 1996.

Buser, P. *Geometry and Spectra of Compact Riemann Surfaces.* Boston, MA: Birkhäuser, 1992.

Chavel, I. *Eigenvalues in Riemannian Geometry.* New York: Academic Press, 1984.

Chavel, I. *Riemannian Geometry: A Modern Introduction.* New York: Cambridge University Press, 1994.

Chern, S.-S. "Finsler Geometry is Just Riemannian Geometry without the Quadratic Restriction." *Not. Amer. Math. Soc.* **43**, 959–963, 1996.

do Carmo, M. P. *Riemannian Geometry.* Boston, MA: Birkhäuser, 1992.

Riemannian Geometry (Non-Euclidean)

see ELLIPTIC GEOMETRY

Riemannian Manifold

A MANIFOLD possessing a METRIC TENSOR. For a complete Riemannian manifold, the METRIC $d(x,y)$ is defined as the length of the shortest curve (GEODESIC) between x and y.

see also BISHOP'S INEQUALITY, CHEEGER'S FINITENESS THEOREM

Riemannian Metric

Suppose for every point x in a COMPACT MANIFOLD M, an INNER PRODUCT $\langle\cdot,\cdot\rangle_x$ is defined on a TANGENT SPACE T_xM of M at x. Then the collection of all these INNER PRODUCTS is called the Riemannian metric. In 1870, Christoffel and Lipschitz showed how to decide when two Riemannian metrics differ by only a coordinate transformation.

see also COMPACT MANIFOLD, LINE ELEMENT, METRIC TENSOR

Riesel Number

There exist infinitely many ODD INTEGERS k such that $k\cdot 2^n-1$ is COMPOSITE for every $n\geq 1$. Numbers k with this property are called RIESEL NUMBERS, and analogous numbers with the minus sign replaced by a plus are called SIERPIŃSKI NUMBERS OF THE SECOND KIND. The smallest known Riesel number is $k=509{,}203$, but there remain 963 smaller candidates (the smallest of which is 659) which generate only composite numbers for all n which have been checked (Ribenboim 1996, p. 358).

Let $a(k)$ be smallest n for which $(2k-1)\cdot 2^n-1$ is PRIME, then the first few values are 2, 0, 2, 1, 1, 2, 3, 1, 2, 1, 1, 4, 3, 1, 4, 1, 2, 2, 1, 3, 2, 7, ... (Sloane's A046069), and second smallest n are 3, 1, 4, 5, 3, 26, 7, 2, 4, 3, 2, 6, 9, 2, 16, 5, 3, 6, 2553, ... (Sloane's A046070).

see also CUNNINGHAM NUMBER, MERSENNE NUMBER, SIERPIŃSKI'S COMPOSITE NUMBER THEOREM, SIERPIŃSKI NUMBER OF THE SECOND KIND

References

Ribenboim, P. *The New Book of Prime Number Records.* New York: Springer-Verlag, p. 357, 1996.

Riesel, H. "Några stora primtal." *Elementa* **39**, 258–260, 1956.
Sloane, N. J. A. Sequence A046068 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Riesz-Fischer Theorem

A function is L_2- (square-) integrable IFF its FOURIER SERIES is L_2-convergent. The application of this theorem requires use of the LEBESGUE INTEGRAL.

see also LEBESGUE INTEGRAL

Riesz Representation Theorem

Let f be a bounded linear FUNCTIONAL on a HILBERT SPACE H. Then there exists exactly one $x_0 \in H$ such that $f(x) = \langle x, x_0 \rangle$ for all $x \in H$. Also, $\|f\| = \|x_0\|$.

see also FUNCTIONAL, HILBERT SPACE

References
Debnath, L. and Mikusiński, P. *Introduction to Hilbert Spaces with Applications.* San Diego, CA: Academic Press, 1990.

Riesz's Theorem

Every continuous linear functional $U[f]$ for $f \in C[a,b]$ can be expressed as a STIELTJES INTEGRAL

$$U[f] = \int_a^b f(x)\,dw(x),$$

where $w(x)$ is determined by U and is of bounded variation on $[a,b]$.

see also STIELTJES INTEGRAL

References
Kestelman, H. "Riesz's Theorem." §11.5 in *Modern Theories of Integration, 2nd rev. ed.* New York: Dover, pp. 265–269, 1960.

Riffle Shuffle

A SHUFFLE, also called a FARO SHUFFLE, in which a deck of $2n$ cards is divided into two HALVES which are then alternatively interleaved from the left and right hands (an "in-shuffle") or from the right and left hands (an "out-shuffle"). Using an "in-shuffle," a deck originally arranged as 1 2 3 4 5 6 7 8 would become 5 1 6 2 7 3 8 4. Using an "out-shuffle," the deck order would become 1 5 2 6 3 7 4 8. Riffle shuffles are used in card tricks (Marlo 1958ab, Adler 1973), and also in the theory of parallel processing (Stone 1971, Chen *et al.* 1981).

In general, card k moves to the position originally occupied by the $2k$th card (mod $2n+1$). Therefore, in-shuffling $2n$ cards $2n$ times (where $2n+1$ is PRIME) results in the original card order. Similarly, out-shuffling $2n$ cards $2n-2$ times (where $2n-1$ is PRIME) results in the original order (Diaconis *et al.* 1983, Conway and Guy 1996). Amazingly, this means that an ordinary deck of 52 cards is returned to its original order after 8 out-shuffles.

Morris (1994) further discusses aspects of the perfect riffle shuffle (in which the deck is cut exactly in half and cards are perfectly interlaced). Ramnath and Scully (1996) give an algorithm for the shortest sequence of in- and out-shuffles to move a card from arbitrary position i to position j. This algorithm works for any deck with an EVEN number of cards and is $\mathcal{O}(\log n)$.

see also CARDS, SHUFFLE

References
Adler, I. "Make Up Your Own Card Tricks." *J. Recr. Math.* **6**, 87–91, 1973.
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 323–325, 1987.
Chen, P. Y.; Lawrie, D. H.; Yew, P.-C.; and Padua, D. A. "Interconnection Networks Using Shuffles." *Computer* **33**, 55–64, Dec. 1981.
Conway, J. H. and Guy, R. K. "Fractions Cycle into Decimals." In *The Book of Numbers.* New York: Springer-Verlag, pp. 163–165, 1996.
Diaconis, P.; Graham, R. L.; and Kantor, W. M. "The Mathematics of Perfect Shuffles." *Adv. Appl. Math.* **4**, 175–196, 1983.
Gardner, M. *Mathematical Carnival: A New Round-Up of Tantalizers and Puzzles from Scientific American.* Washington, DC: Math. Assoc. Amer., 1989.
Herstein, I. N. and Kaplansky, I. *Matters Mathematical.* New York: Harper & Row, 1974.
Mann, B. "How Many Times Should You Shuffle a Deck of Cards." *UMAP J.* **15**, 303–332, 1994.
Marlo, E. *Faro Notes.* Chicago, IL: Ireland Magic Co., 1958a.
Marlo, E. *Faro Shuffle.* Chicago, IL: Ireland Magic Co., 1958b.
Medvedoff, S. and Morrison, K. "Groups of Perfect Shuffles." *Math. Mag.* **60**, 3–14, 1987.
Morris, S. B. and Hartwig, R. E. "The Generalized Faro Shuffle." *Discrete Math.* **15**, 333–346, 1976.
Peterson, I. *Islands of Truth: A Mathematical Mystery Cruise.* New York: W. H. Freeman, pp. 240–244, 1990.
Ramnath, S. and Scully, D. "Moving Card i to Position j with Perfect Shuffles." *Math. Mag.* **69**, 361–365, 1996.
Stone, H. S. "Parallel Processing with the Perfect Shuffle." *IEEE Trans. Comput.* **2**, 153–161, 1971.

Rigby Points

The PERSPECTIVE CENTERS of the TANGENTIAL and CONTACT TRIANGLES of the inner and outer SODDY POINTS. The Rigby points are given by

$$Ri = I + \tfrac{4}{3}Ge$$

$$Ri' = I - \tfrac{4}{3}Ge,$$

where I is the INCENTER and Ge is the GERGONNE POINT.

see also CONTACT TRIANGLE, GERGONNE POINT, GRIFFITHS POINTS, INCENTER, OLDKNOW POINTS, SODDY POINTS, TANGENTIAL TRIANGLE

References
Oldknow, A. "The Euler-Gergonne-Soddy Triangle of a Triangle." *Amer. Math. Monthly* **103**, 319–329, 1996.

Right Angle

An ANGLE equal to half the ANGLE from one end of a line segment to the other. A right angle is $\pi/2$ radians or 90°. A TRIANGLE containing a right angle is called a RIGHT TRIANGLE. However, a TRIANGLE cannot contain more than one right angle, since the sum of the two right angles plus the third angle would exceed the 180° total possessed by a TRIANGLE.

see also ACUTE ANGLE, OBLIQUE ANGLE, OBTUSE ANGLE, RIGHT TRIANGLE, SEMICIRCLE, STRAIGHT ANGLE, THALES' THEOREM

Right Conoid

A RULED SURFACE is called a right conoid if it can be generated by moving a straight LINE intersecting a fixed straight LINE such that the LINES are always PERPENDICULAR (Kreyszig 1991, p. 87). Taking the PERPENDICULAR plane as the xy-plane and the line to be the x-AXIS gives the parametric equations

$$\begin{aligned} x(u,v) &= v\cos\vartheta(u) \\ y(u,v) &= v\sin\vartheta(u) \\ z(u,v) &= h(u) \end{aligned}$$

(Gray 1993). Taking $h(u) = 2u$ and $\vartheta(u) = u$ gives the HELICOID.

see also HELICOID, PLÜCKER'S CONOID, WALLIS'S CONICAL EDGE

References

Dixon, R. *Mathographics.* New York: Dover, p. 20, 1991.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 351–352, 1993.

Kreyszig, E. *Differential Geometry.* New York: Dover, 1991.

Right Hyperbola

A HYPERBOLA for which the ASYMPTOTES are PERPENDICULAR. This occurs when the SEMIMAJOR and SEMIMINOR AXES are equal. Taking $a = b$ in the equation of a HYPERBOLA with SEMIMAJOR AXIS parallel to the x-AXIS and SEMIMINOR AXIS parallel to the y-AXIS (i.e., vertical DIRECTRIX),

$$\frac{(x-x_0)^2}{a^2} - \frac{(y-y_0)^2}{b^2} = 1$$

therefore gives

$$(x-x_0)^2 - (y-y_0)^2 = a^2.$$

A special type of right hyperbola is the so-called RECTANGULAR HYPERBOLA, which has equation $xy = ab$.

see also HYPERBOLA, RECTANGULAR HYPERBOLA

Right Line

see LINE

Right Strophoid

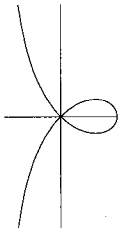

The STROPHOID of a line L with pole O not on L and fixed point O' being the point where the PERPENDICULAR from O to L cuts L is called a right strophoid. It is therefore a general STROPHOID with $a = \pi/2$.

The right strophoid is given by the Cartesian equation

$$y^2 = \frac{c-x}{c+x}x^2, \tag{1}$$

or the polar equation

$$r = c\cos(2\theta)\sec\theta. \tag{2}$$

The parametric form of the strophoid is

$$x(t) = \frac{1-t^2}{t^2+1} \tag{3}$$

$$y(t) = \frac{t(t^2-1)}{t^2+1}. \tag{4}$$

The right strophoid has CURVATURE

$$\kappa(t) = -\frac{4(1+3t^2)}{(1+6t^2+t^4)^{3/2}} \tag{5}$$

and TANGENTIAL ANGLE

$$\phi(t) = -2\tan^{-1}t - \tan^{-1}\left(\frac{2t}{1+t^2}\right). \tag{6}$$

The right strophoid first appears in work by Isaac Barrow in 1670, although Torricelli describes the curve in his letters around 1645 and Roberval found it as the LOCUS of the focus of the conic obtained when the plane cutting the CONE rotates about the tangent at its vertex (MacTutor Archive). The AREA of the loop is

$$A_{\text{loop}} = \tfrac{1}{2}c^2(4-\pi) \tag{7}$$

(MacTutor Archive).

Let C be the CIRCLE with center at the point where the right strophoid crosses the x-axis and radius the distance of that point from the origin. Then the right strophoid is invariant under inversion in the CIRCLE C and is therefore an ANALLAGMATIC CURVE.

see also STROPHOID, TRISECTRIX

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 71, 1993.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 100–104, 1972.

Lockwood, E. H. "The Right Strophoid." Ch. 10 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 90–97, 1967.

MacTutor History of Mathematics Archive. "Right Strophoid." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Right.html`.

Right Strophoid Inverse Curve

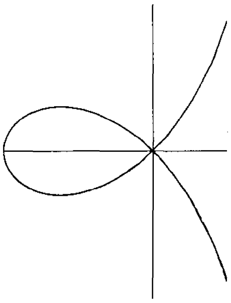

The INVERSE CURVE of a right strophoid is the same strophoid.

Right Triangle

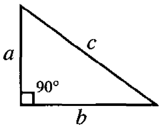

A TRIANGLE with an ANGLE of 90° ($\pi/2$ radians). The sides a, b, and c of such a TRIANGLE satisfy the PYTHAGOREAN THEOREM. The largest side is conventionally denoted c and is called the HYPOTENUSE.

For any three similar shapes on the sides of a right triangle,

$$A_1 + A_2 = A_3, \tag{1}$$

which is equivalent to the PYTHAGOREAN THEOREM. For a right triangle with sides a, b, and HYPOTENUSE c, let r be the INRADIUS. Then

$$\tfrac{1}{2}ab = \tfrac{1}{2}ra + \tfrac{1}{2}rb + \tfrac{1}{2}rc = \tfrac{1}{2}r(a+b+c). \tag{2}$$

Solving for r gives

$$r = \frac{ab}{a+b+c}. \tag{3}$$

But any PYTHAGOREAN TRIPLE can be written

$$a = m^2 - n^2 \tag{4}$$

$$b = 2mn \tag{5}$$

$$c = m^2 + n^2, \tag{6}$$

so (5) becomes

$$r = \frac{(m^2 - n^2)2mn}{m^2 - n^2 + 2mn + m^2 + n^2} = n(m-n), \tag{7}$$

which is an INTEGER when m and n are integers.

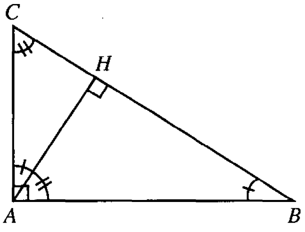

Given a right triangle ΔABC, draw the ALTITUDE AH from the RIGHT ANGLE A. Then the triangles ΔAHC and ΔBHA are similar.

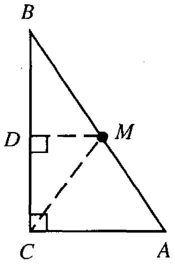

In a right triangle, the MIDPOINT of the HYPOTENUSE is equidistant from the three VERTICES (Dunham 1990). This can be proved as follows. Given ΔABC, let M be the MIDPOINT of AB (so that $AM = BM$). Draw $DM \| CA$, then since ΔBDM is similar to ΔBCA, it follows that $BD = DC$. Since both ΔBDM and ΔCDM are right triangles and the corresponding legs are equal, the HYPOTENUSES are also equal, so we have $AM = BM = CM$ and the theorem is proved.

see also ACUTE TRIANGLE, ARCHIMEDES' MIDPOINT THEOREM, BROCARD MIDPOINT, CIRCLE-POINT MIDPOINT THEOREM, FERMAT'S RIGHT TRIANGLE THEOREM, ISOSCELES TRIANGLE, MALFATTI'S RIGHT TRIANGLE PROBLEM, OBTUSE TRIANGLE, PYTHAGOREAN TRIPLE, QUADRILATERAL, RAT-FREE SET, TRIANGLE

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 121, 1987.

Dunham, W. *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 120–121, 1990.

Rigid

A FRAMEWORK is rigid IFF continuous motion of the points of the configuration maintaining the bar constraints comes from a family of motions of all EUCLIDEAN SPACE which are distance-preserving. A GRAPH G is (generically) d-rigid if, for almost all (i.e., an open dense set of) CONFIGURATIONS of p, the FRAMEWORK $G(p)$ is rigid in $\mathbb{R}^d$.

One of the first results in rigidity theory was the RIGIDITY THEOREM by Cauchy in 1813. Although rigidity problems were of immense interest to engineers, intensive mathematical study of these types of problems has occurred only relatively recently (Connelly 1993, Graver *et al.* 1993).

see also BAR (EDGE), FLEXIBLE POLYHEDRON, FRAMEWORK, LAMAN'S THEOREM, LIEBMANN'S THEOREM, RIGIDITY THEOREM

References

Connelly, R. "Rigidity." Ch. 1.7 in *Handbook of Convex Geometry, Vol. A* (Ed. P. M. Gruber and J. M. Wills). Amsterdam, Netherlands: North-Holland, pp. 223–271, 1993.

Crapo, H. and Whiteley, W. "Statics of Frameworks and Motions of Panel Structures, A Projective Geometry Introduction." *Structural Topology* **6**, 43–82, 1982.

Graver, J.; Servatius, B.; and Servatius, H. *Combinatorial Rigidity.* Providence, RI: Amer. Math. Soc., 1993.

Rigid Motion

A transformation consisting of ROTATIONS and TRANSLATIONS which leaves a given arrangement unchanged.

see also EUCLIDEAN MOTION, PLANE, ROTATION

References
Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 141, 1996.

Rigidity Theorem

If the faces of a *convex* POLYHEDRON were made of metal plates and the EDGES were replaced by hinges, the POLYHEDRON would be RIGID. The theorem was stated by Cauchy (1813), although a mistake in this paper went unnoticed for more than 50 years. An example of a *concave* "FLEXIBLE POLYHEDRON" (with 18 triangular faces) for which this is not true was given by Connelly (1978), and a FLEXIBLE POLYHEDRON with only 14 triangular faces was subsequently found by Steffen (Mackenzie 1998).

see also FLEXIBLE POLYHEDRON, RIGID

References
Cauchy, A. L. "Sur les polygons et le polyhéders." *XVIe Cahier* **IX**, 87–89, 1813.
Connelly, R. "A Flexible Sphere." *Math. Intel.* **1**, 130–131, 1978.
Graver, J.; Servatius, B.; and Servatius, H. *Combinatorial Rigidity.* Providence, RI: Amer. Math. Soc., 1993.
Mackenzie, D. "Polyhedra Can Bend But Not Breathe." *Science* **279**, 1637, 1998.

Ring

A ring is a set together with two BINARY OPERATORS $S(+, *)$ satisfying the following conditions:

1. Additive associativity: For all $a, b, c \in S$, $(a+b)+c = a + (b + c)$,
2. Additive commutativity: For all $a, b \in S$, $a + b = b + a$,
3. Additive identity: There exists an element $0 \in S$ such that for all $a \in S$, $0 + a = a + 0 = a$,
4. Additive inverse: For every $a \in S$ there exists $-ainS$ such that $a + (-a) = (-a) + a = 0$,
5. Multiplicative associativity: For all $a, b, c \in S$, $(a * b) * c = a * (b * c)$,
6. Left and right distributivity: For all $a, b, c \in S$, $a * (b+c) = (a*b)+(a*c)$ and $(b+c)*a = (b*a)+(c*a)$.

A ring is therefore an ABELIAN GROUP under addition and a SEMIGROUP under multiplication. A ring must contain at least one element, but need not contain a multiplicative identity or be commutative. The number of finite rings of n elements for $n = 1, 2, \ldots$, are 1, 2, 2, 11, 2, 4, 2, 52, 11, 4, 2, 22, 2, 4, 4, ... (Sloane's A027623 and A037234; Fletcher 1980). In general, the number of rings of order p^3 for p an ODD PRIME is $3p + 50$ and 52 for $p = 2$ (Ballieu 1947, Gilmer and Mott 1973).

A ring with a multiplicative identity is sometimes called a UNIT RING. Fraenkel (1914) gave the first abstract definition of the ring, although this work did not have much impact.

A ring that is COMMUTATIVE under multiplication, has a unit element, and has no divisors of zero is called an INTEGRAL DOMAIN. A ring which is also a COMMUTATIVE multiplication group is called a FIELD. The simplest rings are the INTEGERS $\mathbb{Z}$, POLYNOMIALS $\mathbb{R}[x]$ and $\mathbb{R}[x, y]$ in one and two variables, and SQUARE $n \times n$ REAL MATRICES.

Rings which have been investigated and found to be of interest are usually named after one or more of their investigators. This practice unfortunately leads to names which give very little insight into the relevant properties of the associated rings.

see also ABELIAN GROUP, ARTINIAN RING, CHOW RING, DEDEKIND RING, DIVISION ALGEBRA, FIELD, GORENSTEIN RING, GROUP, GROUP RING, IDEAL, INTEGRAL DOMAIN, MODULE, NILPOTENT ELEMENT, NOETHERIAN RING, NUMBER FIELD, PRIME RING, PRÜFER RING, QUOTIENT RING, REGULAR RING, RINGOID, SEMIPRIME RING, SEMIRING, SEMISIMPLE RING, SIMPLE RING, UNIT RING, ZERO DIVISOR

References
Ballieu, R. "Anneaux finis; systèmes hypercomplexes de rang trois sur un corps commutatif." *Ann. Soc. Sci. Bruxelles. Sér. I* **61**, 222–227, 1947.
Fletcher, C. R. "Rings of Small Order." *Math. Gaz.* **64**, 9–22, 1980.
Fraenkel, A. "Über die Teiler der Null und die Zerlegung von Ringen." *J. Reine Angew. Math.* **145**, 139–176, 1914.
Gilmer, R. and Mott, J. "Associative Rings of Order p^3." *Proc. Japan Acad.* **49**, 795–799, 1973.
Kleiner, I. "The Genesis of the Abstract Ring Concept." *Amer. Math. Monthly* **103**, 417–424, 1996.
Sloane, N. J. A. Sequences A027623 and A037234 in "An On-Line Version of the Encyclopedia of Integer Sequences."
van der Waerden, B. L. *A History of Algebra.* New York: Springer-Verlag, 1985.

Ring Cyclide

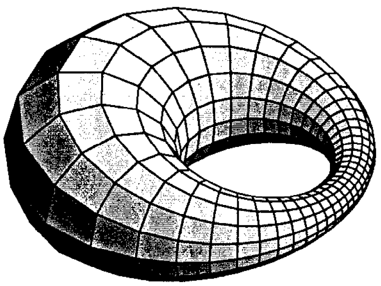

The inversion of a RING TORUS. If the inversion center lies on the torus, then the ring cyclide degenerates to a PARABOLIC RING CYCLIDE.

see also CYCLIDE, PARABOLIC CYCLIDE, RING CYCLIDE, RING TORUS, SPINDLE CYCLIDE, TORUS

Ring Function

see TOROIDAL FUNCTION

Ring Torus

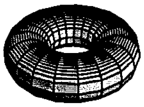

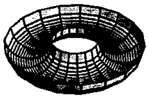

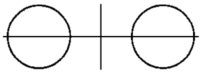

One of the three STANDARD TORI given by the parametric equations

$$\begin{aligned} x &= (c + a\cos v)\cos u \\ y &= (c + a\cos v)\sin u \\ z &= a\sin v \end{aligned}$$

with $c > a$. This is the TORUS which is generally meant when the term "torus" is used without qualification. The inversion of a ring torus is a RING CYCLIDE if the INVERSION CENTER does not lie on the torus and a PARABOLIC RING CYCLIDE if it does. The above left figure shows a ring torus, the middle a cutaway, and the right figure shows a CROSS-SECTION of the ring torus through the xz-plane.

see also CYCLIDE, HORN TORUS, PARABOLIC RING CYCLIDE, RING CYCLIDE, SPINDLE TORUS, STANDARD TORI, TORUS

References

Gray, A. "Tori." §11.4 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 218–220, 1993.

Pinkall, U. "Cyclides of Dupin." §3.3 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 28–30, 1986.

Ringoid

A ringoid R is a set $(R, +, \times)$ with two binary operators, conventionally denoted addition $(+)$ and multiplication $(\times)$, where $\times$ distributes over $+$ left and right:

$$a(b + c) = ab + acand(b + c)a = ba + ca.$$

A ringoid can be empty.

see also BINARY OPERATOR, RING, SEMIRING

References

Rosenfeld, A. *An Introduction to Algebraic Structures.* New York: Holden-Day, 1968.

Risch Algorithm

An ALGORITHM for indefinite integration.

see also INDEFINITE INTEGRAL

Rising Factorial

see POCHHAMMER SYMBOL

Rivest-Shamir-Adleman Number

see RSA NUMBER

RMS

see ROOT-MEAN-SQUARE

Robbin Constant

$$\begin{aligned} R = \tfrac{4}{105} + \tfrac{17}{105}\sqrt{2} - \tfrac{2}{35}\sqrt{3} + \tfrac{1}{5}\ln(1 + \sqrt{2}\,) \\ + \tfrac{2}{5}\ln(2 + \sqrt{3}\,) - \tfrac{1}{15}\pi = 0.661707182\ldots. \end{aligned}$$

see also TRANSFINITE DIAMETER

References

Plouffe, S. "The Robbin Constant." `http://lacim.uqam.ca/piDATA/robbin.txt`.

Robbin's Inequality

If the fourth MOMENT $\mu_4 \neq 0$, then

$$P(|\bar{x} - \mu_4| \geq \lambda) \leq \frac{\mu_4 + 3(N-1)\sigma^4}{N^3\lambda^4},$$

where σ^2 is the VARIANCE.

Robbins Algebra

Building on work of Huntington (1933), Robbins conjectured that the equations for a Robbins algebra, commutivity, associativity, and the ROBBINS EQUATION

$$n(n(x + y) + n(x + n(y))) = x,$$

imply those for a BOOLEAN ALGEBRA. The conjecture was finally proven using a computer (McCune 1997).

References

Huntington, E. V. "New Sets of Independent Postulates for the Algebra of Logic, with Special Reference to Whitehead and Russell's *Principia Mathematica.*" *Trans. Amer. Math. Soc.* **35**, 274–304, 1933.

Huntington, E. V. "Boolean Algebra. A Correction." *Trans. Amer. Math. Soc.* **35**, 557–558, 1933.

McCune, W. "Solution of the Robbins Problem." *J. Automat. Reason.* **19**, 263–276, 1997.

McCune, W. "Robbins Algebras are Boolean." `http://www.mcs.anl.gov/~mccune/papers/robbins/`.

Nelson, E. "Automated Reasoning." `http://www.math.princeton.edu/~nelson/ar.html`.

Robbins Equation

$$n(n(x + y) + n(x + n(y))) = x.$$

see also ROBBINS ALGEBRA

Robertson Condition

For the Helmholtz Differential Equation to be Separable in a coordinate system, the Scale Factors h_i in the Laplacian

$$\nabla^2 = \sum_{i=1}^{3} \frac{1}{h_1 h_2 h_3} \frac{\partial}{\partial u_i} \left(\frac{h_1 h_2 h_3}{{h_i}^2} \frac{\partial}{\partial u_i} \right) \qquad (1)$$

and the functions $f_i(u_i)$ and Φ_{ij} defined by

$$\frac{1}{f_n} \frac{\partial}{\partial u_n} \left(f_n \frac{\partial X_n}{\partial u_n} \right) + ({k_1}^2 \Phi_{n1} + {k_2}^2 \Phi_{n2} + {k_3}^2 \Phi_{n3}) X_n = 0 \qquad (2)$$

must be of the form of a Stäckel Determinant

$$S = |\Phi_{mn}| = \begin{vmatrix} \Phi_{11} & \Phi_{12} & \Phi_{13} \\ \Phi_{21} & \Phi_{22} & \Phi_{23} \\ \Phi_{31} & \Phi_{32} & \Phi_{33} \end{vmatrix} = \frac{h_1 h_2 h_3}{f_1(u_1) f_2(u_2) f_3(u_3)}. \qquad (3)$$

see also Helmholtz Differential Equation, Laplace's Equation, Separation of Variables, Stäckel Determinant

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part 1.* New York: McGraw-Hill, p. 510, 1953.

Robertson Conjecture

A conjecture due to M. S. Robertson (1936) which treats a Univalent Power Series containing only Odd powers within the Unit Disk. This conjecture Implies the Bieberbach Conjecture and follows in turn from the Milin Conjecture. de Branges' proof of the Bieberbach Conjecture proceeded by proving the Milin Conjecture, thus establishing the Robertson conjecture and hence implying the truth of the Bieberbach Conjecture.

see also Bieberbach Conjecture, Milin Conjecture

References

Stewart, I. *From Here to Infinity: A Guide to Today's Mathematics.* Oxford, England: Oxford University Press, p. 165, 1996.

Robertson-Seymour Theorem

A generalization of the Kuratowski Reduction Theorem by Robertson and Seymour, which states that the collection of finite graphs is well-quasi-ordered by minor embeddability, from which it follows that Kuratowski's "forbidden minor" embedding obstruction generalizes to higher genus surfaces.

Formally, for a fixed Integer $g \geq 0$, there is a finite list of graphs $L(g)$ with the property that a graph C embeds on a surface of genus g Iff it does not contain, as a minor, any of the graphs on the list L.

References

Fellows, M. R. "The Robertson-Seymour Theorems: A Survey of Applications." *Comtemp. Math.* **89**, 1–18, 1987.

Robin Boundary Conditions

Partial Differential Equation Boundary Conditions which, for an elliptic partial differential equation in a region Ω, specify that the sum of αu and the normal derivative of $u = f$ at all points of the boundary of Ω, α and f being prescribed.

Robin's Constant

see Transfinite Diameter

Robinson Projection

A Pseudocylindrical Map Projection which distorts shape, Area, scale, and distance to create attraction average projection properties.

References

Dana, P. H. "Map Projections." `http://www.utexas.edu/depts/grg/gcraft/notes/mapproj/mapproj.html`.

Robust Estimation

An estimation technique which is insensitive to small departures from the idealized assumptions which have been used to optimize the algorithm. Classes of such techniques include M-Estimates (which follow from maximum likelihood considerations), L-Estimates (which are linear combinations of Order Statistics), and R-Estimates (based on Rank tests).

see also L-Estimate, M-Estimate, R-Estimate

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Robust Estimation." §15.7 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 694–700, 1992.

Rodrigues's Curvature Formula

$$d\hat{\mathbf{N}} + \kappa_i \, d\mathbf{r} = \mathbf{0},$$

where $\hat{\mathbf{N}}$ is the unit Normal Vector and κ_i is one of the two Principal Curvatures.

see also Normal Vector, Principal Curvatures

Rodrigues Formula

An operator definition of a function. A Rodrigues formula may be converted into a Schläfli Integral.

see also Schläfli Integral

Rogers-Ramanujan Continued Fraction

see Ramanujan Continued Fraction

Rogers-Ramanujan Identities

For $|q| < 1$ and using the NOTATION of the RAMANUJAN THETA FUNCTION, the Rogers-Ramanujan identities are

$$\frac{f(-q^5)}{f(-q,-q^4)} = \sum_{k=0}^{\infty} \frac{q^{k^2}}{(q)_k} \tag{1}$$

$$\frac{f(-q^5)}{f(-q^2,-q^3)} = \sum_{k=0}^{\infty} \frac{q^{k(k+1)}}{(q)_k}, \tag{2}$$

where $(q)_k$ are q-SERIES. Written out explicitly (Hardy 1959, p. 13),

$$\begin{aligned} 1+\frac{q}{1-q}+\frac{q^4}{(1-q)(1-q^2)}+\frac{q^9}{(1-q)(1-q^2)(1-q^3)}+\cdots \\ = \frac{1}{(1-q)(1-q^6)\cdots(1-q^4)(1-q^9)\cdots} \end{aligned} \tag{3}$$

$$\begin{aligned} 1+\frac{q^2}{1-q}+\frac{q^6}{(1-q)(1-q^2)}+\frac{q^{12}}{(1-q)(1-q^2)(1-q^3)}+\cdots \\ = \frac{1}{(1-q^2)(1-q^7)\cdots(1-q^3)(1-q^8)\cdots}. \end{aligned} \tag{4}$$

The identities can also be written succinctly as

$$\begin{aligned} 1+\sum_{k=1}^{\infty} \frac{q^{k^2+ak}}{(1-q)(1-q^2)\cdots(1-q^k)} \\ = \prod_{j=0}^{\infty} \frac{1}{(1-q^{5j+a+1})(1-q^{5j-a+4})}, \end{aligned} \tag{5}$$

where $a = 0,\ 1$.

Other forms of the Rogers-Ramanujan identities include

$$\sum_k \frac{q^{k^2}}{(q;q)_k(q;q)_{n-k}} = \sum_k \frac{(-1)^k q^{(5k^2-k)/2}}{(q;q)_{n-k}(q;q)_{n+k}} \tag{6}$$

and

$$\sum_k \frac{2q^{k^2}}{(q;q)_k(q;q)_{n-k}} = \sum_k \frac{(-1)^k(1+q^k)q^{(5k^2-k)/2}}{(q;q)_{n-k}(q;q)_{n+k}} \tag{7}$$

(Petkovšek *et al.* 1996).

see also ANDREWS-SCHUR IDENTITY

References

Andrews, G. E. *The Theory of Partitions.* Cambridge, England: Cambridge University Press, 1985.

Andrews, G. E. *q-Series: Their Development and Application in Analysis, Number Theory, Combinatorics, Physics, and Computer Algebra.* Providence, RI: Amer. Math. Soc., pp. 17–20, 1986.

Andrews, G. E. and Baxter, R. J. "A Motivated Proof of the Rogers-Ramanujan Identities." *Amer. Math. Monthly* **96**, 401–409, 1989.

Bressoud, D. M. *Analytic and Combinatorial Generalizations of the Rogers-Ramanujan Identities.* Providence, RI: Amer. Math. Soc., 1980.

Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, p. 13, 1959.

Paule, P "Short and Easy Computer Proofs of the Rogers-Ramanujan Identities and of Identities of Similar Type." *Electronic J. Combinatorics* **1**, R10, 1–9, 1994. `http://www.combinatorics.org/Volume_1/volume1.html#R10`.

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, p. 117, 1996.

Robinson, R. M. "Comment to: 'A Motivated Proof of the Rogers-Ramanujan Identities.'" *Amer. Math. Monthly* **97**, 214–215, 1990.

Rogers, L. J. "Second Memoir on the Expansion of Certain Infinite Products." *Proc. London Math. Soc.* **25**, 318–343, 1894.

Sloane, N. J. A. Sequence A006141/M0260 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Rolle's Theorem

Let f be differentiable on (a, b) and continuous on $[a, b]$. If $f(a) = f(b) = 0$, then there is at least one point $c \in (a, b)$ where $f'(c) = 0$.

see also FIXED POINT THEOREM, MEAN-VALUE THEOREM

Roman Coefficient

A generalization of the BINOMIAL COEFFICIENT whose NOTATION was suggested by Knuth,

$$\left\lfloor \begin{matrix} n \\ k \end{matrix} \right\rceil = \frac{\lfloor n \rceil!}{\lfloor k \rceil!\,\lfloor n-k \rceil!}. \tag{1}$$

The above expression is read "Roman n choose k." Whenever the BINOMIAL COEFFICIENT is defined (i.e., $n \geq k \geq 0$ or $k \geq 0 > n$), the Roman coefficient agrees with it. However, the Roman coefficients are defined for values for which the BINOMIAL COEFFICIENTS are not, e.g.,

$$\left\lfloor \begin{matrix} n \\ -1 \end{matrix} \right\rceil = \frac{1}{\lfloor n+1 \rceil} \tag{2}$$

$$\left\lfloor \begin{matrix} 0 \\ k \end{matrix} \right\rceil = \frac{(-1)^{k+(k>0)}}{\lfloor k \rceil}, \tag{3}$$

where

$$n < 0 \equiv \begin{cases} 1 & \text{for } n < 0 \\ 0 & \text{for } n \geq 0. \end{cases} \tag{4}$$

The Roman coefficients also satisfy properties like those of the BINOMIAL COEFFICIENT,

$$\left\lfloor \begin{matrix} n \\ k \end{matrix} \right\rceil = \left\lfloor \begin{matrix} n \\ n-k \end{matrix} \right\rceil \tag{5}$$

$$\left\lfloor \begin{matrix} n \\ k \end{matrix} \right\rceil \left\lfloor \begin{matrix} k \\ r \end{matrix} \right\rceil = \left\lfloor \begin{matrix} n \\ r \end{matrix} \right\rceil \left\lfloor \begin{matrix} n-r \\ k-r \end{matrix} \right\rceil, \tag{6}$$

an analog of PASCAL'S FORMULA

$$\left\lfloor \begin{matrix} n \\ k \end{matrix} \right\rceil = \left\lfloor \begin{matrix} n-1 \\ k \end{matrix} \right\rceil + \left\lfloor \begin{matrix} n-1 \\ k-1 \end{matrix} \right\rceil, \tag{7}$$

and a curious rotation/reflection law due to Knuth

$$(-1)^{k+(k>0)} \left\lfloor \begin{matrix} -n \\ k-1 \end{matrix} \right\rceil = (-1)^{n+(n>0)} \left\lfloor \begin{matrix} -k \\ n-1 \end{matrix} \right\rceil \qquad (8)$$

(Roman 1992).

see also BINOMIAL COEFFICIENT, ROMAN FACTORIAL

References

Roman, S. "The Logarithmic Binomial Formula." *Amer. Math. Monthly* **99**, 641–648, 1992.

Roman Factorial

$$\lfloor n \rceil ! \equiv \begin{cases} n! & \text{for } n \geq 0 \\ \frac{(-1)^{-n-1}}{(-n-1)!} & \text{for } n < 0. \end{cases} \qquad (1)$$

The Roman factorial arises in the definition of the HARMONIC LOGARITHM and ROMAN COEFFICIENT. It obeys the identities

$$\lfloor n \rceil ! = \lfloor n \rceil \, \lfloor n-1 \rceil ! \qquad (2)$$

$$\frac{\lfloor n \rceil !}{\lfloor n-k \rceil !} = \lfloor n \rceil \, \lfloor n-1 \rceil \cdots \lfloor n-k+1 \rceil \qquad (3)$$

$$\lfloor n \rceil ! \, \lfloor -n-1 \rceil ! = (-1)^{n+(n<0)}, \qquad (4)$$

where

$$\lfloor n \rceil \equiv \begin{cases} n & \text{for } n \neq 0 \\ 1 & \text{for } n = 0 \end{cases} \qquad (5)$$

and

$$n < 0 \equiv \begin{cases} 1 & \text{for } n < 0 \\ 0 & \text{for } n \geq 0. \end{cases} \qquad (6)$$

see also HARMONIC LOGARITHM, HARMONIC NUMBER, ROMAN COEFFICIENT

References

Loeb, D. and Rota, G.-C. "Formal Power Series of Logarithmic Type." *Advances Math.* **75**, 1–118, 1989.

Roman, S. "The Logarithmic Binomial Formula." *Amer. Math. Monthly* **99**, 641–648, 1992.

Roman Numeral

A system of numerical notations used by the Romans. It is an additive (and subtractive) system in which letters are used to denote certain "base" numbers, and arbitrary numbers are then denoted using combinations of symbols.

Character	Numerical Value
I	1
V	5
X	10
L	50
C	100
D	500
M	1000

For example, the number 1732 would be denoted MDCCXXXII. One additional rule states that, instead of using four symbols to represent a 4, 40, 9, 90, etc., such numbers are instead denoted by preceding the symbol for 5, 50, 10, 100, etc., with a symbol indicating *subtraction.* For example, 4 is denoted IV, 9 as IX, 40 as XL, etc. However, this rule is generally *not* followed on the faces of clocks, where IIII is usually encountered instead of IV.

Roman numerals are encountered in the release year for movies and occasionally on the numerals on the faces of watches and clocks, but in few other modern instances. They do have the advantage that ADDITION can be done "symbolically" (and without worrying about the "place" of a given DIGIT) by simply combining all the symbols together, grouping, writing groups of 5 Is as V, groups of 2 Vs as X, etc.

Roman Surface

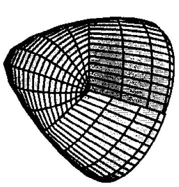

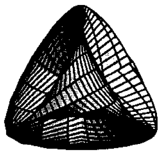

A QUARTIC NONORIENTABLE SURFACE, also known as the STEINER SURFACE. The Roman surface is one of the three possible surfaces obtained by sewing a MÖBIUS STRIP to the edge of a DISK. The other two are the BOY SURFACE and CROSS-CAP, all of which are homeomorphic to the REAL PROJECTIVE PLANE (Pinkall 1986).

The center point of the Roman surface is an ordinary TRIPLE POINT with $(\pm1, 0, 0) = (0, \pm1, 0) = (0, 0, \pm1)$, and the six endpoints of the three lines of self-intersection are singular PINCH POINTS, also known as WHITNEY SINGULARITIES. The Roman surface is essentially six CROSS-CAPS stuck together and contains a double INFINITY of CONICS.

The Roman surface can given by the equation

$$(x^2+y^2+z^2-k^2)^2 = [(z-k)^2-2x^2][(z+k)^2-2y^2]. \qquad (1)$$

Solving for z gives the pair of equations

$$z = \frac{k(y^2-x^2) \pm (x^2-y^2)\sqrt{k^2-x^2-y^2}}{2(x^2+y^2)}. \qquad (2)$$

If the surface is rotated by 45° about the z-AXIS via the ROTATION MATRIX

$$\mathsf{R}_z(45^\circ) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (3)$$

to give

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \mathsf{R}_z(45^\circ) \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \qquad (4)$$

then the simple equation

$$x^2y^2 + x^2z^2 + y^2z^2 + 2kxyz = 0 \tag{5}$$

results. The Roman surface can also be generated using the general method for NONORIENTABLE SURFACES using the polynomial function

$$\mathbf{f}(x,y,z) = (xy, yz, zx) \tag{6}$$

(Pinkall 1986). Setting

$$x = \cos u \sin v \tag{7}$$

$$y = \sin u \sin v \tag{8}$$

$$z = \cos v \tag{9}$$

in the former gives

$$x(u,v) = \tfrac{1}{2}\sin(2u)\sin^2 v \tag{10}$$

$$y(u,v) = \tfrac{1}{2}\sin u \cos(2v) \tag{11}$$

$$z(u,v) = \tfrac{1}{2}\cos u \sin(2v) \tag{12}$$

for $u \in [0, 2\pi)$ and $v \in [-\pi/2, \pi/2]$. Flipping $\sin v$ and $\cos v$ and multiplying by 2 gives the form shown by Wang.

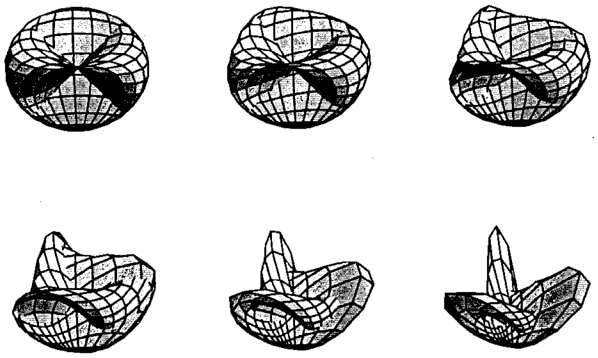

A HOMOTOPY (smooth deformation) between the Roman surface and BOY SURFACE is given by the equations

$$x(u,v) = \frac{\sqrt{2}\cos(2u)\cos^2 v + \cos u \sin(2v)}{2 - \alpha\sqrt{2}\sin(3u)\sin(2v)} \tag{13}$$

$$y(u,v) = \frac{\sqrt{2}\sin(2u)\cos^2 v - \sin u \sin(2v)}{2 - \alpha\sqrt{2}\sin(3u)\sin(2v)} \tag{14}$$

$$z(u,v) = \frac{3\cos^2 v}{2 - \alpha\sqrt{2}\sin(3u)\sin(2v)} \tag{15}$$

for $u \in [-\pi/2, \pi/2]$ and $v \in [0, \pi]$ as α varies from 0 to 1. $\alpha = 0$ corresponds to the Roman surface and $\alpha = 1$ to the BOY SURFACE (Wang).

see also BOY SURFACE, CROSS-CAP, HEPTAHEDRON, MÖBIUS STRIP, NONORIENTABLE SURFACE, QUARTIC SURFACE, STEINER SURFACE

References

Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, p. 19, 1986.

Fischer, G. (Ed.). Plates 42–44 and 108–114 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 42–44 and 108–109, 1986.

Geometry Center. "The Roman Surface." `http://www.geom.umn.edu/zoo/toptype/pplane/roman/`.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 242–243, 1993.

Nordstrand, T. "Steiner's Roman Surface." `http://www.uib.no/people/nfytn/steintxt.htm`.

Pinkall, U. *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, p. 64, 1986.

Wang, P. "Renderings." `http://www.ugcs.caltech.edu/~peterw/portfolio/renderings/`.

Roman Symbol

$$\lfloor n \rceil \equiv \begin{cases} n & \text{for } n \neq 0 \\ 1 & \text{for } n = 0. \end{cases}$$

see also ROMAN FACTORIAL, HARMONIC LOGARITHM

References

Roman, S. "The Logarithmic Binomial Formula." *Amer. Math. Monthly* **99**, 641–648, 1992.

Romberg Integration

A powerful NUMERICAL INTEGRATION technique which uses k refinements of the extended TRAPEZOIDAL RULE to remove error terms less than order $\mathcal{O}(N^{-2k})$. The routine advocated by Press *et al.* (1992) makes use of NEVILLE'S ALGORITHM.

References

Acton, F. S. *Numerical Methods That Work, 2nd printing.* Washington, DC: Math. Assoc. Amer., pp. 106–107, 1990.

Dahlquist, G. and Bjorck, A. §7.4.1–7.4.2 in *Numerical Methods.* Englewood Cliffs, NJ: Prentice-Hall, 1974.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Romberg Integration." §4.3 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 134–135, 1992.

Ralston, A. and Rabinowitz, P. §4.10 in *A First Course in Numerical Analysis, 2nd ed.* New York: McGraw-Hill, 1978.

Stoer, J.; and Burlisch, R. §3.4–3.5 in *Introduction to Numerical Analysis.* New York: Springer-Verlag, 1980.

Rook Number

The rook numbers r_n^B of an $n \times n$ BOARD B are the number of subsets of size n such that no two elements have the same first or second coordinate. In other word, it is the number of ways of placing n rooks on B such that none attack each other. The rook numbers of a board determine the rook numbers of the complementary board $\overline{B}$, defined to be $\mathbf{d} \times \mathbf{d} \backslash B$. This is known as the ROOK RECIPROCITY THEOREM. The first few rook numbers are 1, 2, 7, 23, 115, 694, 5282, 46066, … (Sloane's A000903). For an $n \times n$ board, each $n \times n$

PERMUTATION MATRIX corresponds to an allowed configuration of rooks.

see also ROOK RECIPROCITY THEOREM

References

Sloane, N. J. A. Sequence A000903/M1761 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Rook Reciprocity Theorem

$$\sum_{k=0}^{d} r_k^B (d-k)!x^k = \sum_{k=0}^{d} (-1)^k r_k^{\bar{B}} (d-k)!x^k (x+1)^{d-k}.$$

References

Chow, T. Y. "The Path-Cycle Symmetric Function of a Digraph." *Adv. Math.* **118**, 71–98, 1996.

Chow, T. "A Short Proof of the Rook Reciprocity Theorem." *Electronic J. Combinatorics* **3**, R10, 1–2, 1996. `http://www.combinatorics.org/Volume_3/volume3.html#R10`.

Goldman, J. R.; Joichi, J. T.; and White, D. E. "Rook Theory I. Rook Equivalence of Ferrers Boards." *Proc. Amer. Math. Soc.* **52**, 485–492, 1975.

Riordan, J. *An Introduction to Combinatorial Analysis.* New York: Wiley, 1958.

Rooks Problem

							R
						R	
					R		
				R			
			R				
		R					
	R						
R							

The rook is a CHESS piece which may move any number of spaces either horizontally or vertically per move. The maximum number of nonattacking rooks which may be placed on an $n \times n$ CHESSBOARD is n. This arrangement is achieved by placing the rooks along the diagonal (Madachy 1979). The total number of ways of placing n nonattacking rooks on an $n \times n$ board is $n!$ (Madachy 1979, p. 47). The number of rotationally and reflectively inequivalent ways of placing n nonattacking rooks on an $n \times n$ board are 1, 2, 7, 23, 115, 694, ... (Sloane's A000903; Dudeney 1970, p. 96; Madachy 1979, pp. 46–54).

							R
						R	
					R		
				R			
			R				
		R					
	R						
R							

The minimum number of rooks needed to occupy or attack all spaced on an 8×8 CHESSBOARD is 8, illustrated above (Madachy 1979).

Consider an $n \times n$ chessboard with the restriction that, for every subset of $\{1, \ldots, n\}$, a rook may not be put in column $s + j \pmod{n}$ when on row j, where the rows are numbered 0, 1, ..., $n - 1$. Vardi (1991) denotes the number of rook solutions so restricted as $\text{rook}(s, n)$. $\text{rook}(\{1\}, n)$ is simply the number of DERANGEMENTS on n symbols, known as a SUBFACTORIAL. The first few values are 1, 2, 9, 44, 265, 1854, ... (Sloane's A000166). $\text{rook}(\{1, 2\}, n)$ is a solution to the MARRIED COUPLES PROBLEM, sometimes known as MÉNAGE NUMBERS. The first few MÉNAGE NUMBERS are -1, 1, 0, 2, 13, 80, 579, ... (Sloane's A000179).

Although simple formulas are not known for general $\{1, \ldots, p\}$, RECURRENCE RELATIONS can be used to compute $\text{rook}(\{1, \ldots, p\}, n)$ in polynomial time for $p = 3$, ..., 6 (Metropolis *et al.* 1969, Minc 1978, Vardi 1991).

see also CHESS, MÉNAGE NUMBER, ROOK NUMBER, ROOK RECIPROCITY THEOREM

References

Dudeney, H. E. "The Eight Rooks." §295 in *Amusements in Mathematics.* New York: Dover, p. 88, 1970.

Kraitchik, M. "The Problem of the Rooks" and "Domination of the Chessboard." §10.2 and 10.4 in *Mathematical Recreations.* New York: W. W. Norton, pp. 240–247 and 255–256, 1942.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 36–37, 1979.

Metropolis, M.; Stein, M. L.; and Stein, P. R. "Permanents of Cyclic (0, 1) Matrices." *J. Combin. Th.* **7**, 291–321, 1969.

Minc, H. §3.1 in *Permanents.* Reading, MA: Addison-Wesley, 1978.

Riordan, J. Chs. 7–8 in *An Introduction to Combinatorial Analysis.* Princeton, NJ: Princeton University Press, 1978.

Sloane, N. J. A. Sequences A000903/M1761, A000166/M1937, and A000179/M2062 in "An On-Line Version of the Encyclopedia of Integer Sequences." `http://www.research.att.com/~njas/sequences/eisonline.html`.

Sloane, N. J. A. and Plouffe, S. Extended entry for M2062 in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 123–124, 1991.

Room Square

A Room square (named after T. G. Room) of order n (for n ODD) is an arrangement in an $n \times n$ SQUARE MATRIX of $n + 1$ objects such that each cell is either empty or holds exactly two different objects. Furthermore, each object appears once in each row and column and each unordered pair occupies exactly one cell. The Room square of order 2 is shown below.

$$\boxed{1, 2}$$

The Room square of order 8 is

1, 8			5, 7		3, 4	2, 6
3, 7	2, 8			6, 1		4, 5
5, 6	4, 1	3, 8			7, 2	
	6, 7	5, 2	4, 8			1, 3
2, 4		7, 1	6, 3	5 ,8		
	3, 5		1, 2	7, 4	6, 8	
		4, 6		2, 3	1, 5	7, 8

References

Dinitz, J. H. and Stinson, D. R. In *Contemporary Design Theory: A Collection of Surveys* (Ed. J. H. Dinitz and D. R. Stinson). New York: Wiley, 1992.

Gardner, M. *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, pp. 146–147 and 151–152, 1988.

Mullin, R. C. and Nemeth, E. "On Furnishing Room Squares." *J. Combin. Th.* **7**, 266–272, 1969.

Mullin, R. D. and Wallis, W. D. "The Existence of Room Squares." *Aequationes Math.* **13**, 1–7, 1975.

O'Shaughnessy, C. D. "On Room Squares of Order $6m+2$." *J. Combin. Th.* **13**, 306–314, 1972.

Room, T. G. "A New Type of Magic Square" (Note 2569). *Math. Gaz.* **39**, 307, 1955.

Wallis, W. D. "Solution of the Room Square Existence Problem." *J. Combin. Th.* **17**, 379–383, 1974.

Root

The roots of an equation

$$f(x) = 0 \tag{1}$$

are the values of x for which the equation is satisfied. The FUNDAMENTAL THEOREM OF ALGEBRA states that every POLYNOMIAL equation of degree n has exactly n roots, where some roots may have a multiplicity greater than 1 (in which case they are said to be degenerate).

To find the nth roots of a COMPLEX NUMBER, solve the equation $z^n = w$. Then

$$z^n = |z|^n[\cos(n\theta) + i\sin(n\theta)] = |w|\,(\cos\phi + i\sin\phi), \tag{2}$$

so

$$|z| = |w|^{1/n} \tag{3}$$

and

$$\arg(z) = \frac{\phi}{n}. \tag{4}$$

Rolle proved that any number has n nth roots (Boyer 1968, p. 476). Householder (1970) gives an algorithm for constructing root-finding algorithms with an arbitrary order of convergence. Special root-finding techniques can often be applied when the function in question is a POLYNOMIAL.

see also BAILEY'S METHOD, BISECTION PROCEDURE, BRENT'S METHOD, CROUT'S METHOD, DESCARTES' SIGN RULE, FALSE POSITION METHOD, FUNDAMENTAL THEOREM OF SYMMETRIC FUNCTIONS, GRAEFFE'S METHOD, HALLEY'S IRRATIONAL FORMULA, HALLEY'S METHOD, HALLEY'S RATIONAL FORMULA, HORNER'S METHOD, HOUSEHOLDER'S METHOD, HUTTON'S METHOD, ISOGRAPH, JENKINS-TRAUB METHOD, LAGUERRE'S METHOD, LAMBERT'S METHOD, LEHMER-SCHUR METHOD, LIN'S METHOD, MAEHLY'S PROCEDURE, MULLER'S METHOD, NEWTON'S METHOD, RIDDERS' METHOD, ROOT DRAGGING THEOREM, SCHRÖDER'S METHOD, POLYNOMIAL, SECANT METHOD, STURM FUNCTION, STURM THEOREM, TANGENT HYPERBOLAS METHOD, WEIERSTRAß APPROXIMATION THEOREM

References

Arfken, G. "Appendix 1: Real Zeros of a Function." *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 963–967, 1985.

Boyer, C. B. *A History of Mathematics.* New York: Wiley, 1968.

Householder, A. S. *The Numerical Treatment of a Single Nonlinear Equation.* New York: McGraw-Hill, 1970.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Roots of Polynomials." §9.5 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 362–372, 1992.

Root Dragging Theorem

If any of the ROOTS of a POLYNOMIAL are increased, then *all* of the critical points increase.

References

Anderson, B. "Polynomial Root Dragging." *Amer. Math. Monthly* **100**, 864–866, 1993.

Root Linear Coefficient Theorem

The sum of the reciprocals of ROOTS of an equation equals the NEGATIVE COEFFICIENT of the linear term in the MACLAURIN SERIES.

see also NEWTON'S RELATIONS

Root-Mean-Square

The root-mean-square (RMS) of a variate x, sometimes called the QUADRATIC MEAN, is the SQUARE ROOT of the mean squared value of x:

$$R(x) \equiv \sqrt{\langle x^2\rangle} \tag{1}$$

$$= \begin{cases} \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} & \text{for a discrete distribution} \\ \sqrt{\frac{\int P(x)x^2\,dx}{\int P(x)\,dx}} & \text{for a continuous distribution.} \end{cases} \tag{2}$$

Hoehn and Niven (1985) show that

$$R(a_1+c, a_2+c, \ldots, a_n+c) < c + R(a_1, a_2, \ldots, a_n)$$

for any POSITIVE constant c.

Physical scientists often use the term root-mean-square as a synonym for STANDARD DEVIATION when they refer to the SQUARE ROOT of the mean squared deviation of a signal from a given baseline or fit.

see also ARITHMETIC-GEOMETRIC MEAN, ARITHMETIC-HARMONIC MEAN, GENERALIZED MEAN, GEOMETRIC MEAN, HARMONIC MEAN, HARMONIC-GEOMETRIC MEAN, MEAN, MEDIAN (STATISTICS), STANDARD DEVIATION, VARIANCE

References

Hoehn, L. and Niven, I. "Averages on the Move." *Math. Mag.* **58**, 151–156, 1985.

Root (Radical)

The nth root (or "RADICAL") of a quantity z is a value r such that $z = r^n$, and therefore is the INVERSE FUNCTION to the taking of a POWER. The nth root is denoted $r = \sqrt[n]{z}$ or, using POWER notation, $r = z^{1/n}$. The special case of the SQUARE ROOT is denoted $\sqrt{z}$. The quantities for which a general FUNCTION equals 0 are also called ROOTS, or sometimes ZEROS.

see also CUBE ROOT, ROOT, SQUARE ROOT, VINCULUM

Root Test

Let u_k be a SERIES with POSITIVE terms, and let

$$\rho \equiv \lim_{k\to\infty} {u_k}^{1/k}.$$

1. If $\rho < 1$, the SERIES CONVERGES.
2. If $\rho > 1$ or $\rho = \infty$, the SERIES DIVERGES.
3. If $\rho = 1$, the SERIES may CONVERGE or DIVERGE.

This test is also called the CAUCHY ROOT TEST.

see also CONVERGENCE TESTS

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 281–282, 1985.

Bromwich, T. J. I'a and MacRobert, T. M. *An Introduction to the Theory of Infinite Series, 3rd ed.* New York: Chelsea, pp. 31–39, 1991.

Root (Tree)

A special node which is designated to turn a TREE into a ROOTED TREE. The root is sometimes also called "EVE," and each of the nodes which is one EDGE further away from a given EDGE is called a CHILD. Nodes connected to the same node are then called SIBLINGS.

see also CHILD, ROOTED TREE, SIBLING, TREE

Root of Unity

The nth ROOTS of UNITY are ROOTS $\zeta_k = e^{2\pi ik/p}$ of the CYCLOTOMIC EQUATION

$$x^p = 1,$$

which are known as the DE MOIVRE NUMBERS.

see also CYCLOTOMIC EQUATION, DE MOIVRE'S IDENTITY, DE MOIVRE NUMBER, UNITY

References

Courant, R. and Robbins, H. "De Moivre's Formula and the Roots of Unity." §5.3 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 98–100, 1996.

Rooted Tree

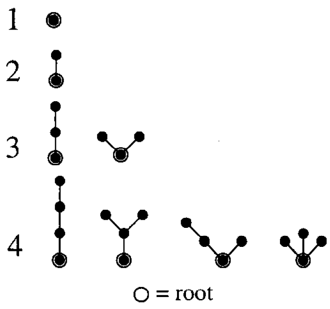

A TREE with a special node called the "ROOT" or "EVE." Denote the number of rooted trees with n nodes by T_n, then the GENERATING FUNCTION is

$$T(x) \equiv \sum_{n=0}^{\infty} T_n x^n = x + x^2 + 2x^3 + 4x^4 + 9x^5 + 20x^6 + 48x^7 + 115x^8 + 286x^9 + 719x^{10} + \ldots \quad (1)$$

(Sloane's A000081). This POWER SERIES satisfies

$$T(x) = x \exp\left[\sum_{r=1}^{\infty} \frac{1}{r} T(x^r)\right] \quad (2)$$

$$t(x) = T(x) - \tfrac{1}{2}[T^2(x) - T(x^2)], \quad (3)$$

where $t(x)$ is the GENERATING FUNCTION for unrooted TREES. A GENERATING FUNCTION for T_n can be written using a product involving *the sequence itself* as

$$x \prod_{n=1}^{\infty} \frac{1}{(1-x^n)^{T_n}} = \sum_{n=1}^{\infty} T_n x^n. \quad (4)$$

The number of rooted trees can also be calculated from the RECURRENCE RELATION

$$T_{i+1} = \frac{1}{i} \sum_{j=1}^{i} \left(\sum_{d|j} d T_d \right) T_{i-j+1}, \quad (5)$$

with $T_0 = 0$ and $T_1 = 1$, where the second sum is over all d which DIVIDE j (Finch).

see also ORDERED TREE, RED-BLACK TREE, WEAKLY BINARY TREE

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/otter/otter.html`.

Ruskey, F. "Information on Rooted Trees." `http://sue.csc.uvic.ca/~cos/inf/tree/RootedTree.html`.

Sloane, N. J. A. Sequence A000081/M1180 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Rosatti's Theorem

There is a one-to-one correspondence between the sets of equivalent correspondences (not of value 0) on an irreducible curve of GENUS (CURVE) p, and the rational COLLINEATIONS of a projective space of $2p-1$ dimensions which leave invariant a space of $p-1$ dimensions. The number of linearly independent correspondences will be that of linearly independent COLLINEATIONS.

References
Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 339, 1959.

Rose

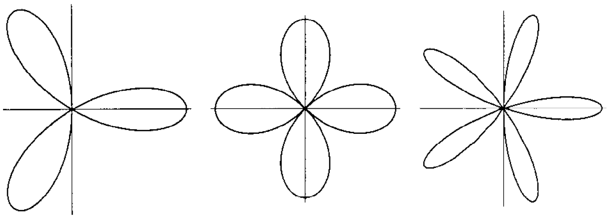

A curve which has the shape of a petalled flower. This curve was named RHODONEA by the Italian mathematician Guido Grandi between 1723 and 1728 because it resembles a rose (MacTutor Archive). The polar equation of the rose is

$$r = a\sin(n\theta),$$

or

$$r = a\cos(n\theta).$$

If n is ODD, the rose is n-petalled. If n is EVEN, the rose is $2n$-petalled. If n is IRRATIONAL, then there are an infinite number of petals.

The QUADRIFOLIUM is the rose with $n=2$. The rose is the RADIAL CURVE of the EPICYCLOID.

see also DAISY, MAURER ROSE, STARR ROSE

References
Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 175–177, 1972.
Lee, X. "Rose." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Rose_dir/rose.html`.
MacTutor History of Mathematics Archive. "Rhodonea Curves." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Rhodonea.html`.
Wagon, S. "Roses." §4.1 in *Mathematica in Action.* New York: W. H. Freeman, pp. 96–102, 1991.

Rosenbrock Methods

A generalization of the RUNGE-KUTTA METHOD for solution of ORDINARY DIFFERENTIAL EQUATIONS, also called KAPS-RENTROP METHODS.

see also RUNGE-KUTTA METHOD

References
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 730–735, 1992.

Rössler Model

The nonlinear 3-D MAP

$$\begin{aligned}\dot{X} &= -(Y+Z)\\ \dot{Y} &= X+0.2Y\\ \dot{Z} &= 0.2+XZ-cZ.\end{aligned}$$

see also LORENZ SYSTEM

References
Dickau, R. M. "Rössler Attractor." `http://www.prairienet.org/~pops/rossler.html`.
Peitgen, H.-O.; Jürgens, H.; and Saupe, D. §12.3 in *Chaos and Fractals: New Frontiers of Science.* New York: Springer-Verlag, pp. 686–696, 1992.

Rotation

The turning of an object or coordinate system by an ANGLE about a fixed point. A rotation is an ORIENTATION-PRESERVING ORTHOGONAL TRANSFORMATION. EULER'S ROTATION THEOREM states that an arbitrary rotation can be parameterized using three parameters. These parameters are commonly taken as the EULER ANGLES. Rotations can be implemented using ROTATION MATRICES.

The rotation SYMMETRY OPERATION for rotation by 360°/n is denoted "n." For periodic arrangements of points ("crystals"), the CRYSTALLOGRAPHY RESTRICTION gives the only allowable rotations as 1, 2, 3, 4, and 6.

see also DILATION, EUCLIDEAN GROUP, EULER'S ROTATION THEOREM, EXPANSION, IMPROPER ROTATION, INFINITESIMAL ROTATION, INVERSION OPERATION, MIRROR PLANE, ORIENTATION-PRESERVING, ORTHOGONAL TRANSFORMATION, REFLECTION, ROTATION FORMULA, ROTATION GROUP, ROTATION MATRIX, ROTATION OPERATOR, ROTOINVERSION, SHIFT, TRANSLATION

References
Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 211, 1987.
Yates, R. C. "Instantaneous Center of Rotation and the Construction of Some Tangents." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 119–122, 1952.

Rotation Formula

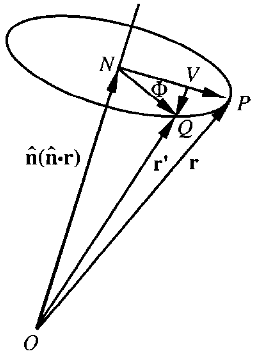

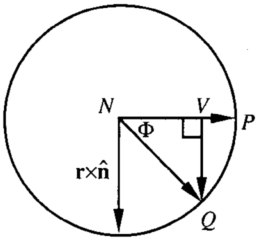

A formula which relates the VECTOR $\mathbf{r}'$ to the ANGLE Φ in the above figure (Goldstein 1980). Referring to the figure,

$$\begin{aligned}\mathbf{r}' &= \overrightarrow{ON} + \overrightarrow{NV} + \overrightarrow{VQ}\\ &= \hat{\mathbf{n}}(\hat{\mathbf{n}}\cdot\mathbf{r}) + [\mathbf{r} - \hat{\mathbf{n}}(\hat{\mathbf{n}}\cdot\mathbf{r})]\cos\Phi + (\mathbf{r}\times\hat{\mathbf{n}})\sin\Phi\\ &= \mathbf{r}\cos\Phi + \hat{\mathbf{n}}(\hat{\mathbf{n}}\cdot\mathbf{r})(1-\cos\Phi) + (\mathbf{r}\times\hat{\mathbf{n}})\sin\Phi.\end{aligned}$$

The ANGLE Φ and unit normal $\hat{\mathbf{n}}$ may also be expressed as EULER ANGLES. In terms of the EULER PARAMETERS,

$$\mathbf{r}' = \mathbf{r}({e_0}^2 - {e_1}^2 - {e_2}^2 - {e_3}^2) + 2\mathbf{e}(\mathbf{e}\cdot\mathbf{r}) + 2(\mathbf{r}\times\hat{\mathbf{n}})\sin\Phi.$$

see also EULER ANGLES, EULER PARAMETERS

References

Goldstein, H. *Classical Mechanics, 2nd ed.* Reading, MA: Addison-Wesley, 1980.

Rotation Group

There are three representations of the rotation groups, corresponding to EXPANSION/DILATION, ROTATION, and SHEAR.

Rotation Matrix

When discussing a ROTATION, there are two possible conventions: rotation of the *axes* and rotation of the *object* relative to fixed axes.

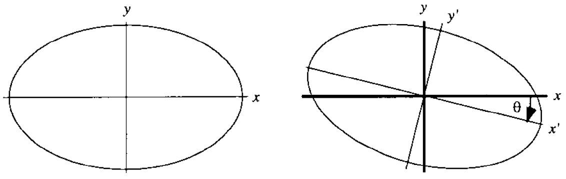

curve rotated by angle θ

In $\mathbb{R}^2$, let a curve be rotated by a clockwise ANGLE θ, so that the original axes of the curve are $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$, and the new axes of the curve are $\hat{\mathbf{x}}'$ and $\hat{\mathbf{y}}'$. The MATRIX transforming the original curve to the rotated curve, referred to the original $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ axes, is

$$\mathsf{R}_\theta = \begin{bmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{bmatrix}, \tag{1}$$

i.e.,

$$\mathbf{x} = \mathsf{R}_\theta\mathbf{x}'. \tag{2}$$

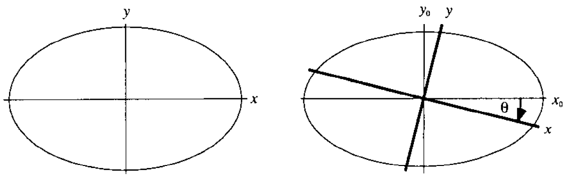

axes rotated by angle θ

On the other hand, let the *axes* with respect to which a curve is measured be rotated by a clockwise ANGLE θ, so that the original axes are $\hat{\mathbf{x}}_0$ and $\hat{\mathbf{y}}_0$, and the new axes are $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$. Then the MATRIX transforming the coordinates of the curve with respect to $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ is given by the MATRIX TRANSPOSE of the above matrix:

$$\mathsf{R}'_\theta = \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}, \tag{3}$$

i.e.,

$$\mathbf{x} = \mathsf{R}'_\theta\mathbf{x}_0. \tag{4}$$

In $\mathbb{R}^3$, rotations of the x-, y-, and z-axes give the matrices

$$\mathsf{R}_x(\alpha) = \begin{bmatrix}1 & 0 & 0\\ 0 & \cos\alpha & \sin\alpha\\ 0 & -\sin\alpha & \cos\alpha\end{bmatrix} \tag{5}$$

$$\mathsf{R}_y(\beta) = \begin{bmatrix}\cos\beta & 0 & -\sin\beta\\ 0 & 1 & 0\\ \sin\beta & 0 & \cos\beta\end{bmatrix} \tag{6}$$

$$\mathsf{R}_z(\gamma) = \begin{bmatrix}\cos\gamma & \sin\gamma & 0\\ -\sin\gamma & \cos\gamma & 0\\ 0 & 0 & 1\end{bmatrix}. \tag{7}$$

see also EULER ANGLES, EULER'S ROTATION THEOREM, ROTATION

Rotation Number

The period for a QUASIPERIODIC trajectory to pass through the same point in a SURFACE OF SECTION. If the rotation number is IRRATIONAL, the trajectory will densely fill out a curve in the SURFACE OF SECTION. If the rotation number is RATIONAL, it is called the WINDING NUMBER, and only a finite number of points in the SURFACE OF SECTION will be visited by the trajectory.

see also QUASIPERIODIC FUNCTION, SURFACE OF SECTION, WINDING NUMBER (MAP)

Rotation Operator

The rotation operator can be derived from examining an INFINITESIMAL ROTATION

$$\left(\frac{d}{dt}\right)_{\text{space}} = \left(\frac{d}{dt}\right)_{\text{body}} + \boldsymbol{\omega}\times,$$

where d/dt is the time derivative, $\boldsymbol{\omega}$ is the ANGULAR VELOCITY, and $\times$ is the CROSS PRODUCT operator.

see also ACCELERATION, ANGULAR ACCELERATION, INFINITESIMAL ROTATION

Roth's Removal Rule

If the matrices A, X, B, and C satisfy

$$\mathsf{AX} - \mathsf{XB} = \mathsf{C},$$

then

$$\begin{bmatrix}\mathsf{I} & \mathsf{X}\\ 0 & \mathsf{I}\end{bmatrix}\begin{bmatrix}\mathsf{A} & \mathsf{C}\\ 0 & \mathsf{B}\end{bmatrix}\begin{bmatrix}\mathsf{I} & -\mathsf{X}\\ 0 & \mathsf{I}\end{bmatrix} = \begin{bmatrix}\mathsf{A} & 0\\ 0 & \mathsf{B}\end{bmatrix},$$

where I is the IDENTITY MATRIX.

References

Roth, W. E. "The Equations $AX-YB=C$ and $AX-XB=C$ in Matrices." *Proc. Amer. Math. Soc.* **3**, 392–396, 1952.

Turnbull, H. W. and Aitken, A. C. *An Introduction to the Theory of Canonical Matrices.* New York: Dover, p. 422, 1961.

Roth's Theorem

For ALGEBRAIC α

$$\left|\alpha-\frac{p}{q}\right|<\frac{1}{q^{2+\epsilon}},$$

with $\epsilon>0$, has finitely many solutions. Klaus Roth received a FIELDS MEDAL for this result.

see also HURWITZ EQUATION, HURWITZ'S IRRATIONAL NUMBER THEOREM, LAGRANGE NUMBER (RATIONAL APPROXIMATION), LIOUVILLE'S RATIONAL APPROXIMATION THEOREM, LIOUVILLE-ROTH CONSTANT, MARKOV NUMBER, SEGRE'S THEOREM, THUE-SIEGEL-ROTH THEOREM

References

Davenport, H. and Roth, K. F. "Rational Approximations to Algebraic Numbers." *Mathematika* **2**, 160–167, 1955.

Roth, K. F. "Rational Approximations to Algebraic Numbers." *Mathematika* **2**, 1–20, 1955.

Roth, K. F. "Corrigendum to 'Rational Approximations to Algebraic Numbers'." *Mathematika* **2**, 168, 1955.

Rotkiewicz Theorem

If $n>19$, there exists a base-2 PSEUDOPRIME between n and n^2. The theorem was proved in 1965.

see also PSEUDOPRIME

References

Rotkiewicz, A. "Les intervalles contenants les nombres pseudoprimiers." *Rend. Circ. Mat. Palermo Ser. 2* **14**, 278–280, 1965.

Rotkiewicz, A. "Sur les nombres de Mersenne dépourvus de diviseurs carrés er sur les nombres naturels n, tel que n^2-2^n-2." *Mat. Vesnik* **2 (17)**, 78–80, 1965.

Rotkiewicz, A. "Sur les nombres pseudoprimiers carrés." *Elem. Math.* **20**, 39–40, 1965.

Rotoinversion

see IMPROPER ROTATION

Rotor

A convex figure that can be rotated inside a POLYGON (or POLYHEDRON) while always touching every side (or face). The least AREA rotor in a SQUARE is the REULEAUX TRIANGLE. The least AREA rotor in an EQUILATERAL TRIANGLE is a LENS with two 60° ARCS of CIRCLES and RADIUS equal to the TRIANGLE ALTITUDE.

There exist nonspherical rotors for the TETRAHEDRON, OCTAHEDRON, and CUBE, but not for the DODECAHEDRON and ICOSAHEDRON.

see also LENS, REULEAUX TRIANGLE

References

Gardner, M. *The Unexpected Hanging and Other Mathematical Diversions.* Chicago, IL: Chicago University Press, p. 219, 1991.

Rotunda

A class of solids whose only true member is the PENTAGONAL ROTUNDA.

see also ELONGATED ROTUNDA, GYROELONGATED ROTUNDA, PENTAGONAL ROTUNDA, TRIANGULAR HEBESPHENOROTUNDA

References

Johnson, N. W. "Convex Polyhedra with Regular Faces." *Canad. J. Math.* **18**, 169–200, 1966.

Rouché's Theorem

Given two functions f and g ANALYTIC in A with γ a simple loop HOMOTOPIC to a point in A, if $|g(z)|<|f(z)|$ for all z on γ, then f and $f+g$ have the same number of ROOTS inside γ.

References

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., p. 22, 1975.

Roulette

The curve traced by a fixed point on a closed convex curve as that curve rolls without slipping along a second curve. The roulettes described by the FOCI of CONICS when rolled upon a line are sections of MINIMAL SURFACES (i.e., they yield MINIMAL SURFACES when revolved about the line) known as UNDULOIDS.

Curve 1	Curve 2	Pole	Roulette
circle	exterior circle	on c.	epicycloid
circle	interior circle	on c.	hypocycloid
circle	line	on c.	cycloid
circle	same circle	any point	rose
circle involute	line	center	parabola
cycloid	line	center	ellipse
ellipse	line	focus	elliptic catenary
hyperbola	line	focus	hyperbolic catenary
hyperbolic spiral	line	origin	tractrix
line	any curve	on line	involute of curve
logarithmic spiral	line	any point	line
parabola	equal parabola	vertex	cissoid of Diocles
parabola	line	focus	catenary

see also GLISSETTE, UNDULOID

References

Besant, W. H. *Notes on Roulettes and Glissettes, 2nd enl. ed.* Cambridge, England: Deighton, Bell & Co., 1890.

Cundy, H. and Rollett, A. "Roulettes and Involutes." §2.6 in *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 46–55, 1989.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 56–58 and 206, 1972.

Lockwood, E. H. "Roulettes." Ch. 17 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 138–151, 1967.

Yates, R. C. "Roulettes." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 175–185, 1952.

Zwillinger, D. (Ed.). "Roulettes (Spirograph Curves)." §8.2 in *CRC Standard Mathematical Tables and Formulae, 3rd ed.* Boca Raton, FL: CRC Press, 1996. `http://www.geom.umn.edu/docs/reference/CRC-formulas/node34.html`.

Round

see NINT

Rounding

The process of approximating a quantity, usually done for convenience or, in the case of numerical computations, of necessity. If rounding is performed on each of a series of numbers in a long computation, round-off errors can become important, especially if division by a small number ever occurs.

see also SHADOWING THEOREM

References

Wilkinson, J. H. *Rounding Errors in Algebraic Processes.* New York: Dover, 1994.

Routh-Hurwitz Theorem

Consider the CHARACTERISTIC EQUATION

$$|\lambda \mathsf{I} - \mathsf{A}| = \lambda^n + b_1\lambda^{n-1} + \ldots + b_{n-1}\lambda + b_n = 0$$

determining the n EIGENVALUES λ of a REAL $n \times n$ MATRIX A, where I is the IDENTITY MATRIX. Then the EIGENVALUES λ all have NEGATIVE REAL PARTS if

$$\Delta_1 > 0, \Delta_2 > 0, \ldots, \Delta_n > 0,$$

where

$$\Delta_k = \begin{vmatrix} b_1 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ b_3 & b_2 & b_1 & 1 & 0 & 0 & \cdots & 0 \\ b_5 & b_4 & b_3 & b_2 & b_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{2k-1} & b_{2k-2} & b_{2k-3} & b_{2k-4} & b_{2k-5} & b_{2k-6} & \cdots & b_k \end{vmatrix}.$$

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1119, 1979.

Routh's Theorem

If the sides of a TRIANGLE are divided in the ratios $\lambda : 1$, $\mu : 1$, and $\nu : 1$, the CEVIANS form a central TRIANGLE whose AREA is

$$A = \frac{(\lambda\mu\nu - 1)^2}{(\lambda\mu + \lambda + 1)(\mu\nu + \mu + 1)(\nu\lambda + \nu + 1)}\Delta, \quad (1)$$

where Δ is the AREA of the original TRIANGLE. For $\lambda = \mu = \nu \equiv n$,

$$A = \frac{(n-1)^2}{n^2 + n + 1}\Delta. \quad (2)$$

For $n = 2$, 3, 4, 5, the areas are $\frac{1}{7}$, $\frac{3}{7}$, and $\frac{16}{31}$. The AREA of the TRIANGLE formed by connecting the division points on each side is

$$A' = \frac{\lambda\mu\nu}{(\lambda + 1)(\mu + 1)(\nu + 1)}\Delta. \quad (3)$$

Routh's theorem gives CEVA'S THEOREM and MENELAUS' THEOREM as special cases.

see also CEVA'S THEOREM, CEVIAN, MENELAUS' THEOREM

References

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 211–212, 1969.

RSA Encryption

A PUBLIC-KEY CRYPTOGRAPHY ALGORITHM which uses PRIME FACTORIZATION as the TRAPDOOR FUNCTION. Define

$$n \equiv pq \quad (1)$$

for p and q PRIMES. Also define a private key d and a public key e such that

$$de \equiv 1 \pmod{\phi(n)} \quad (2)$$

$$(e, \phi(n)) = 1, \quad (3)$$

where $\phi(n)$ is the TOTIENT FUNCTION.

Let the message be converted to a number M. The sender then makes n and e public and sends

$$E = M^e \pmod{n}. \quad (4)$$

To decode, the receiver (who knows d) computes

$$E^d \equiv (M^e)^d \equiv M^{ed} \equiv M^{N\phi(n)+1} \equiv M \pmod{n}, \quad (5)$$

since N is an INTEGER. In order to crack the code, d must be found. But this requires factorization of n since

$$\phi(n) = (p-1)(q-1). \quad (6)$$

Both p and q should be picked so that $p \pm 1$ and $q \pm 1$ are divisible by large PRIMES, since otherwise the POLLARD $p-1$ FACTORIZATION METHOD or WILLIAMS $p+1$ FACTORIZATION METHOD potentially factor n easily. It is also desirable to have $\phi(\phi(pq))$ large and divisible by large PRIMES.

It is possible to break the cryptosystem by repeated encryption if a unit of $\mathbb{Z}/\phi(n)\mathbb{Z}$ has small ORDER (Simmons and Norris 1977, Meijer 1996), where $\mathbb{Z}/s\mathbb{Z}$ is the RING of INTEGERS between 0 and $s-1$ under addition and multiplication (mod s). Meijer (1996) shows that "almost" every encryption exponent e is safe from breaking using repeated encryption for factors of the form

$$p = 2p_1 + 1 \qquad (7)$$
$$q = 2q_1 + 1, \qquad (8)$$

where

$$p_1 = 2p_2 + 1 \qquad (9)$$
$$q_1 = 2q_2 + 1, \qquad (10)$$

and p, p_1, p_2, q, q_1, and q_2 are all PRIMES. In this case,

$$\phi(n) = 4p_1q_1 \qquad (11)$$
$$\phi(\phi(n)) = 8p_2q_2. \qquad (12)$$

Meijer (1996) also suggests that p_2 and q_2 should be of order 10^{75}.

Using the RSA system, the identity of the sender can be identified as genuine without revealing his private code.

see also PUBLIC-KEY CRYPTOGRAPHY

References
Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 166–173, 1985.
Meijer, A. R. "Groups, Factoring, and Cryptography." *Math. Mag.* **69**, 103–109, 1996.
Rivest, R. L. "Remarks on a Proposed Cryptanalytic Attack on the MIT Public-Key Cryptosystem." *Cryptologia* **2**, 62–65, 1978.
Rivest, R.; Shamir, A.; and Adleman, L. "A Method for Obtaining Digital Signatures and Public Key Cryptosystems." *Comm. ACM* **21**, 120–126, 1978.
RSA Data Security.® A Security Dynamics Company. `http://www.rsa.com`.
Simmons, G. J. and Norris, M. J. "Preliminary Comments on the MIT Public-Key Cryptosystem." *Cryptologia* **1**, 406–414, 1977.

RSA Number

Numbers contained in the "factoring challenge" of RSA Data Security, Inc. An additional number which is not part of the actual challenge is the RSA-129 number. The RSA numbers which have been factored are RSA-100, RSA-110, RSA-120, RSA-129, and RSA-130 (Cowie *et al.* 1996).

RSA-129 is a 129-digit number used to encrypt one of the first public-key messages. This message was published by R. Rivest, A. Shamir, and L. Adleman (Gardner 1977), along with the number and a $100 reward for its decryption. Despite belief that the message encoded by RSA-129 "would take millions of years of break," RSA-129 was factored in 1994 using a distributed computation which harnessed networked computers spread around the globe performing a multiple polynomial QUADRATIC SIEVE FACTORIZATION METHOD. The effort was coordinated by P. Leylad, D. Atkins, and M. Graff. They received 112,011 full factorizations, 1,431,337 single partial factorizations, and 8,881,138 double partial factorizations out of a factor base of 524,339 PRIMES. The final MATRIX obtained was $188,346 \times 188,346$ square.

The text of the message was "The magic words are squeamish ossifrage" (an ossifrage is a rare, predatory vulture found in the mountains of Europe), and the FACTORIZATION (into a 64-DIGIT number and a 65-DIGIT number) is

$$\begin{aligned}
&114381625757888867669235779976146612010218296\cdots\\
&\cdots 721242362562561842935706935245733897830597 1\cdots\\
&\cdots 2356395870505898907514759929002687954354 1\\
&= 3490529510847650949147849619903898133417764\cdots\\
&\cdots 638493387843990820577 \cdot 32769132993266\cdots\\
&\cdots 709549961988190834461413177642967992\cdots\\
&\cdots 942539798288533
\end{aligned}$$

(Leutwyler 1994, Cipra 1995).

References
Cipra, B. "The Secret Life of Large Numbers." *What's Happening in the Mathematical Sciences, 1995–1996, Vol. 3.* Providence, RI: Amer. Math. Soc., pp. 90–99, 1996.
Cowie, J.; Dodson, B.; Elkenbracht-Huizing, R. M.; Lenstra, A. K.; Montgomery, P. L.; Zayer, J. A. "World Wide Number Field Sieve Factoring Record: On to 512 Bits." In *Advances in Cryptology—ASIACRYPT '96 (Kyongju)* (Ed. K. Kim and T. Matsumoto.) New York: Springer-Verlag, pp. 382–394, 1996.
Gardner, M. "Mathematical Games: A New Kind of Cipher that Would Take Millions of Years to Break." *Sci. Amer.* **237**, 120–124, Aug. 1977.
Klee, V. and Wagon, S. *Old and New Unsolved Problems in Plane Geometry and Number Theory, rev. ed.* Washington, DC: Math. Assoc. Amer., p. 223, 1991.
Leutwyler, K. "Superhack: Forty Quadrillion Years Early, a 129-Digit Code is Broken." *Sci. Amer.* **271**, 17–20, 1994.
Leyland, P. `ftp://sable.ox.ac.uk/pub/math/rsa129`.
RSA Data Security.® A Security Dynamics Company. `http://www.rsa.com`.
Taubes, G. "Small Army of Code-breakers Conquers a 129-Digit Giant." *Science* **264**, 776–777, 1994.
Weisstein, E. W. "RSA Numbers." `http://www.astro.virginia.edu/~eww6n/math/notebooks/RSANumbers.m`.

Rubber-Sheet Geometry

see ALGEBRAIC TOPOLOGY

Rubik's Clock

A puzzle consisting of 18 small clocks. There are 12^{18} possible configurations, although not all are realizable.

see also RUBIK'S CUBE

References
Dénes, J. and Mullen, G. L. "Rubik's Clock and Its Solution." *Math. Mag.* **68**, 378–381, 1995.
Zeilberger, D. "Doron Zeilberger's Maple Packages and Programs: RubikClock." `http://www.math.temple.edu/~zeilberg/programs.html`.

Rubik's Cube

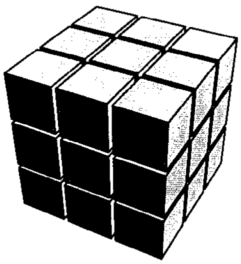

A $3\times3\times3$ CUBE in which the 26 subcubes on the outside are internally hinged in such a way that rotation (by a quarter turn in either direction or a half turn) is possible in any plane of cubes. Each of the six sides is painted a distinct color, and the goal of the puzzle is to return the cube to a state in which each side has a single color after it has been randomized by repeated rotations. The PUZZLE was invented in the 1970s by the Hungarian Erno Rubik and sold millions of copies worldwide over the next decade.

The number of possible positions of Rubik's cube is

$$\frac{8!12!3^8 2^{12}}{2\cdot3\cdot2} = 43{,}252{,}003{,}274{,}489{,}856{,}000$$

(Turner and Gold 1985). Hoey showed using the PÓLYA-BURNSIDE LEMMA that there are 901,083,404,981,813,-616 positions up to conjugacy by whole-cube symmetries.

Algorithms exist for solving a cube from an arbitrary initial position, but they are not necessarily optimal (i.e., requiring a minimum number of turns). The maximum number of turns required for an arbitrary starting position is still not known, although it is bounded from above. Michael Reid (1995) produced the best proven bound of 29 turns (or 42 "quarter-turns"). The proof involves large tables of "subroutines" generated by computer.

However, Dik Winter has produced a program based on work by Kociemba which has solved each of millions of cubes in at most 21 turns. Recently, Richard Korf (1997) has produced a different algorithm which is practical for cubes up to 18 moves away from solved. Out of 10 randomly generated cubes, one was solved in 16 moves, three required 17 moves, and six required 18 moves.

see also RUBIK'S CLOCK

References
Hofstadter, D. R. "Metamagical Themas: The Magic Cube's Cubies are Twiddled by Cubists and Solved by Cubemeisters." *Sci. Amer.* **244**, 20–39, Mar. 1981.
Larson, M. E. "Rubik's Revenge: The Group Theoretical Solution." *Amer. Math. Monthly* **92**, 381–390, 1985.
Miller, D. L. W. "Solving Rubik's Cube Using the 'Bestfast' Search Algorithm and 'Profile' Tables." `http://www.sunyit.edu/~millerd1/RUBIK.HTM`.
Schubart, M. "Rubik's Cube Resource List." `http://www.best.com/~schubart/rc/resources.html`.
Singmaster, D. *Notes on Rubik's 'Magic Cube.'* Hillside, NJ: Enslow Pub., 1981.
Taylor, D. *Mastering Rubik's Cube.* New York: Holt, Rinehart, and Winston, 1981.
Taylor, D. and Rylands, L. *Cube Games: 92 Puzzles & Solutions* New York: Holt, Rinehart, and Winston, 1981.
Turner, E. C. and Gold, K. F. "Rubik's Groups." *Amer. Math. Monthly* **92**, 617–629, 1985.

Rudin-Shapiro Sequence

The sequence of numbers given by

$$a_n = (-1)^{\sum_{i=1}^{k-1} \epsilon_i \epsilon_{i+1}}, \tag{1}$$

where n is written in binary

$$n = \epsilon_1 \epsilon_2 \ldots \epsilon_k. \tag{2}$$

It is therefore the parity of the number of pairs of consecutive 1s in the BINARY expansion of n. The SUMMATORY sequence is

$$s_n \equiv \sum_{j=0}^{n} a_j, \tag{3}$$

which gives

$$s_n = \begin{cases} 2^{k/2}+1 & \text{if } k \text{ is even} \\ 2^{(k-1)/2}+1 & \text{if } k \text{ is odd} \end{cases} \tag{4}$$

(Blecksmith and Laud 1995).

References
Blecksmith, R. and Laud, P. W. "Some Exact Number Theory Computations via Probability Mechanisms." *Amer. Math. Monthly* **102**, 893–903, 1995.
Brillhart, J.; Erdős, P.; and Morton, P. "On the Sums of the Rudin-Shapiro Coefficients II." *Pac. J. Math.* **107**, 39–69, 1983.
Brillhart, J. and Morton, P. "Über Summen von Rudin-Shapiroschen Koeffizienten." *Ill. J. Math.* **22**, 126–148, 1978.
France, M. M. and van der Poorten, A. J. "Arithmetic and Analytic Properties of Paper Folding Sequences." *Bull. Austral. Math. Soc.* **24**, 123–131, 1981.

Rudvalis Group

The SPORADIC GROUP *Ru*.

see also SPORADIC GROUP

References
Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas/Ru.html`.

Rule

A usually simple Algorithm or Identity. The term is frequently applied to specific orders of Newton-Cotes Formulas.

see also Algorithm, BAC-CAB Rule, Bode's Rule, Chain Rule, Cramer's Rule, Descartes' Sign Rule, Durand's Rule, Estimator, Euler's Rule, Euler's Totient Rule, Golden Rule, Hardy's Rule, Horner's Rule, Identity, L'Hospital's Rule, Leibniz Integral Rule, Method, Osborne's Rule, Pascal's Rule, Power Rule, Product Rule, Quarter Squares Rule, Quota Rule, Quotient Rule, Roth's Removal Rule, Rule of 72, Simpson's Rule, Slide Rule, Sum Rule, Trapezoidal Rule, Weddle's Rule, Zeuthen's Rule

Rule of 72

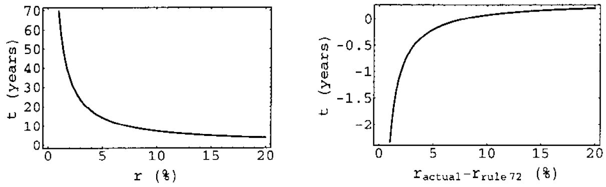

The time required for a given Principal to double (assuming $n = 1$ Conversion Period) for Compound Interest is given by solving

$$2P = P(1+r)^t, \tag{1}$$

or

$$t = \frac{\ln 2}{\ln(1+r)}, \tag{2}$$

where Ln is the Natural Logarithm. This function can be approximated by the so-called "rule of 72":

$$t \approx \frac{0.72}{r}. \tag{3}$$

The above plots show the actual doubling time t (left plot) and difference between actual and time calculated using the rule of 72 (right plot) as a function of the interest rate r.

see also Compound Interest, Interest

References

Avanzini, J. F. *Rapid Debt-Reduction Strategies.* Fort Worth, TX: HIS Pub., 1990.

Ruled Surface

A Surface which can be swept out by a moving Line in space and therefore has a parameterization of the form

$$\mathbf{x}(u,v) = \mathbf{b}(u) + v\boldsymbol{\delta}(u), \tag{1}$$

where $\mathbf{b}$ is called the Directrix (also called the Base Curve) and $\boldsymbol{\delta}$ is the Director Curve. The straight lines themselves are called Rulings. The rulings of a ruled surface are Asymptotic Curves. Furthermore, the Gaussian Curvature on a ruled Regular Surface is everywhere Nonpositive.

Examples of ruled surfaces include the elliptic Hyperboloid of one sheet (a doubly ruled surface)

$$\begin{bmatrix} a(cosu \mp v\sin u) \\ b(\sin u \pm \cos u) \\ \pm cv \end{bmatrix} = \begin{bmatrix} a\cos u \\ b\sin u \\ 0 \end{bmatrix} \pm v \begin{bmatrix} -a\cos u \\ b\sin u \\ c \end{bmatrix}, \tag{2}$$

the Hyperbolic Paraboloid (a doubly ruled surface)

$$\begin{bmatrix} a(u+v) \\ \pm bv \\ u^2+2uv \end{bmatrix} = \begin{bmatrix} au \\ 0 \\ u^2 \end{bmatrix} + v \begin{bmatrix} a \\ \pm b \\ 2u \end{bmatrix}, \tag{3}$$

Plücker's Conoid

$$\begin{bmatrix} r\cos\theta \\ r\sin\theta \\ 2\cos\theta\sin\theta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2\cos\theta\sin\theta \end{bmatrix} + r \begin{bmatrix} \cos\theta \\ \sin\theta \\ 0 \end{bmatrix}, \tag{4}$$

and the Möbius Strip

$$\begin{aligned} a \begin{bmatrix} \cos u + v\cos(\tfrac{1}{2}u)\cos u \\ \sin u + v\cos(\tfrac{1}{2}u)\sin u \\ v\sin(\tfrac{1}{2}u) \end{bmatrix} &= a \begin{bmatrix} \cos u \\ \sin u \\ 0 \end{bmatrix} + av \begin{bmatrix} \cos(\tfrac{1}{2}u)\cos u \\ \cos(\tfrac{1}{2}u)\sin u \\ \sin(\tfrac{1}{2}u) \end{bmatrix} \end{aligned} \tag{5}$$

(Gray 1993).

The only ruled Minimal Surfaces are the Plane and Helicoid (Catalan 1842, do Carmo 1986).

see also Asymptotic Curve, Cayley's Ruled Surface, Developable Surface, Director Curve, Directrix (Ruled Surface), Generalized Cone, Generalized Cylinder, Helicoid, Noncylindrical Ruled Surface, Plane, Right Conoid, Ruling

References

Catalan E. "Sur les surfaces régléess dont l'aire est un minimum." *J. Math. Pure. Appl.* **7**, 203–211, 1842.

do Carmo, M. P. "The Helicoid." §3.5B in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 44–45, 1986.

Fischer, G. (Ed.). Plates 32–33 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 32–33, 1986.

Gray, A. "Ruled Surfaces." Ch. 17 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 333–355, 1993.

Ruler

A Straightedge with markings to indicate distances. Although Geometric Constructions are sometimes said to be performed with a ruler and Compass, the term Straightedge is preferable to ruler since markings are not allowed by the classical Greek rules.

see also Coastline Paradox, Compass, Geometric Construction, Geometrography, Golomb Ruler, Perfect Ruler, Simplicity, Slide Rule, Straightedge

Ruler Function

The exponent of the largest POWER of 2 which DIVIDES a given number k. The values of the ruler function are 1, 2, 1, 3, 1, 2, 1, 4, 1, 2, ... (Sloane's A001511).

References

Guy, R. K. "Cycles and Sequences Containing All Permutations as Subsequences." §E22 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 224, 1994.

Sloane, N. J. A. Sequence A001511/M0127 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Ruling

One of the straight lines sweeping out a RULED SURFACE. The rulings on a ruled surface are ASYMPTOTIC CURVES.

see also ASYMPTOTIC CURVE, DIRECTOR CURVE, DIRECTRIX (RULED SURFACE), RULED SURFACE

Run

A run is a sequence of more than one consecutive identical outcomes, also known as a CLUMP. Given n BERNOULLI TRIALS (say, in the form of COIN TOSSINGS), the probability $P_t(n)$ of a run of t consecutive heads or tails is given by the RECURRENCE RELATION

$$P_t(n) = P_t(n-1) + 2^{-t}[1 - P_t(n-t)], \tag{1}$$

where $P_t(n) = 0$ for $n < t$ and $P_t(t) = 2^{1-t}$ (Bloom 1996).

Let $C_t(m,k)$ denote the number of sequences of m indistinguishable objects of type A and k indistinguishable objects of type B in which *no* t-run occurs. The probability that a t-run *does* occur is then given by

$$P_t(m,k) = 1 - \frac{C_t(m,k)}{\binom{m+k}{k}}, \tag{2}$$

where $\binom{a}{b}$ is a BINOMIAL COEFFICIENT. Bloom (1996) gives the following recurrence sequence for $C_t(m,k)$,

$$\begin{aligned} C_t(m,k) = \sum_{i=0}^{t-1} C_t(m-1,k-i) \\ - \sum_{i=1}^{t-1} C_t(m-t,k-i) + e_t(m,k), \end{aligned} \tag{3}$$

where

$$e_t(m,k) \equiv \begin{cases} 1 & \text{if } m = 0 \text{ and } 0 \leq k < t \\ -1 & \text{if } m = t \text{ and } 0 \leq k < t \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Another recurrence which has only a fixed number of terms is given by

$$\begin{aligned} C_t(m,k) = C_t(m-1,k) + C_t(m,k-1) - C_t(m-t,k-1) \\ - C_t(m-1,k-t) + C_t(m-t,k-t) + e_t^*(m,k), \end{aligned} \tag{5}$$

where

$$e_t^*(m,k) \equiv \begin{cases} 1 & \text{if } (m,k) = (0,0) \text{ or } (t,t) \\ -1 & \text{if } (m,k) = (0,t) \text{ or } (t,0) \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

(Goulden and Jackson 1983, Bloom 1996). These formulas disprove the assertion of Gardner (1982) that "there will almost always be a clump of six or seven CARDS of the same color" in a normal deck of cards by giving $P_6(26,26) = 0.46424$.

Given n BERNOULLI TRIALS with a probability of success (heads) p, the expected number of tails is $n(1-p)$, so the expected number of tail runs ≥ 1 is $\approx n(1-p)p$. Continuing,

$$N_R = n(1-p)p^R \tag{7}$$

is the expected number of runs $\geq R$. The longest expected run is therefore given by

$$R = \log_{1/p}[n(1-p)] \tag{8}$$

(Gordon *et al.* 1986, Schilling 1990). Given m 0s and n 1s, the number of possible arrangements with u runs is

$$f_u = \begin{cases} 2\binom{m-1}{k-1}\binom{n-1}{k-1} & u \equiv 2k \\ \binom{m-1}{k-1}\binom{n-1}{k-2} + \binom{m-1}{k-2}\binom{n-1}{k-1} & u \equiv 2k+1 \end{cases} \tag{9}$$

for k an INTEGER, where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT. Then

$$P(u \leq u') = \sum_{u=2}^{u'} \frac{f_u}{\binom{m+n}{m}}. \tag{10}$$

Bloom (1996) gives the expected number of noncontiguous t-runs in a sequence of m 0s and n 1s as

$$E(n,m,t) = \frac{(m+1)(n)_t + (n+1)(m)_t}{(m+n)_t}, \tag{11}$$

where $(a)_n$ is the POCHHAMMER SYMBOL. For $m > 10$, u has an approximately NORMAL DISTRIBUTION with MEAN and VARIANCE

$$\mu_u = 1 + \frac{2mn}{m+n} \tag{12}$$

$${\sigma_u}^2 = \frac{2mn(2mn-m-n)}{(m+n)^2(m+n-1)}. \tag{13}$$

see also COIN TOSSING, EULERIAN NUMBER, PERMUTATION, s-RUN

References

Bloom, D. M. "Probabilities of Clumps in a Binary Sequence (and How to Evaluate Them Without Knowing a Lot)." *Math. Mag.* **69**, 366–372, 1996.

Gardner, M. *Aha! Gotcha: Paradoxes to Puzzle and Delight.* New York: W. H. Freeman, p. 124, 1982.

Godbole, A. P. "On Hypergeometric and Related Distributions of Order k." *Commun. Stat.: Th. and Meth.* **19**, 1291–1301, 1990.

Godbole, A. P. and Papastavnidis, G. (Eds.). *Runs and Patterns in Probability: Selected Papers.* New York: Kluwer, 1994.

Gordon, L.; Schilling, M. F.; and Waterman, M. S. "An Extreme Value Theory for Long Head Runs." *Prob. Th. and Related Fields* **72**, 279–287, 1986.

Goulden, I. P. and Jackson, D. M. *Combinatorial Enumeration.* New York: Wiley, 1983.

Mood, A. M. "The Distribution Theory of Runs." *Ann. Math. Statistics* **11**, 367–392, 1940.

Philippou, A. N. and Makri, F. S. "Successes, Runs, and Longest Runs." *Stat. Prob. Let.* **4**, 211–215, 1986.

Schilling, M. F. "The Longest Run of Heads." *Coll. Math. J.* **21**, 196–207, 1990.

Schuster, E. F. In *Runs and Patterns in Probability: Selected Papers* (Ed. A. P. Godbole and S. Papastavridis). Boston, MA: Kluwer, pp. 91–111, 1994.

Runge-Kutta Method

A method of integrating Ordinary Differential Equations by using a trial step at the midpoint of an interval to cancel out lower-order error terms. The second-order formula is

$$\begin{aligned} k_1 &= hf(x_n, y_n) \\ k_2 &= hf(x_n + \tfrac{1}{2}h, y_n + \tfrac{1}{2}k_1) \\ y_{n+1} &= y_n + k_2 + \mathcal{O}(h^3), \end{aligned}$$

and the fourth-order formula is

$$\begin{aligned} k_1 &= hf(x_n, y_n) \\ k_2 &= hf(x_n + \tfrac{1}{2}h, y_n + \tfrac{1}{2}k_1) \\ k_3 &= hf(x_n + \tfrac{1}{2}h, y_n + \tfrac{1}{2}k_2) \\ k_4 &= hf(x_n + h, y_n + k_3) \\ y_{n+1} &= y_n + \tfrac{1}{6}k_1 + \tfrac{1}{3}k_2 + \tfrac{1}{3}k_3 + \tfrac{1}{6}k_4 + \mathcal{O}(h^5). \end{aligned}$$

(Press *et al.* 1992). This method is reasonably simple and robust and is a good general candidate for numerical solution of differential equations when combined with an intelligent adaptive step-size routine.

see also Adams' Method, Gill's Method, Milne's Method, Ordinary Differential Equation, Rosenbrock Methods

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 896–897, 1972.

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 492–493, 1985.

Cartwright, J. H. E. and Piro, O. "The Dynamics of Runge-Kutta Methods." *Int. J. Bifurcations Chaos* **2**, 427–449, 1992. `http://formentor.uib.es/~julyan/TeX/rkpaper/root/root.html`.

Lambert, J. D. and Lambert, D. Ch. 5 in *Numerical Methods for Ordinary Differential Systems: The Initial Value Problem.* New York: Wiley, 1991.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Runge-Kutta Method" and "Adaptive Step Size Control for Runge-Kutta." §16.1 and 16.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 704–716, 1992.

Runge-Walsh Theorem

Let $f(x)$ be an Analytic Function which is Regular in the interior of a Jordan Curve C and continuous in the closed Domain bounded by C. Then $f(x)$ can be approximated with an arbitrary accuracy by Polynomials.

see also Analytic Function, Jordan Curve

References

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., p. 7, 1975.

Running Knot

A Knot which tightens around an object when strained but slackens when the strain is removed. Running knots are sometimes also known as slip knots or nooses.

References

Owen, P. *Knots.* Philadelphia, PA: Courage, p. 60, 1993.

Russell's Antinomy

Let R be the set of all sets which are not members of themselves. Then R is neither a member of itself nor not a member of itself. Symbolically, let $R = \{x : x \notin x\}$. Then $R \in R$ Iff $R \notin R$.

Bertrand Russell discovered this Paradox and sent it in a letter to G. Frege just as Frege was completing *Grundlagen der Arithmetik.* This invalidated much of the rigor of the work, and Frege was forced to add a note at the end stating, "A scientist can hardly meet with anything more undesirable than to have the foundation give way just as the work is finished. I was put in this position by a letter from Mr. Bertrand Russell when the work was nearly through the press."

see also Grelling's Paradox

References

Courant, R. and Robbins, H. "The Paradoxes of the Infinite." §2.4.5 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 78, 1996.

Frege, G. *Foundations of Arithmetic.* Evanston, IL: Northwestern University Press, 1968.

Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, pp. 20–21, 1989.

Russell's Paradox

see Russell's Antinomy

Russian Multiplication

Also called ETHIOPIAN MULTIPLICATION. To multiply two numbers a and b, write $a_0 \equiv a$ and $b_0 \equiv b$ in two columns. Under a_0, write $\lfloor a_0/2 \rfloor$, where $\lfloor x \rfloor$ is the FLOOR FUNCTION, and under b_0, write $2b_0$. Continue until $a_i = 1$. Then cross out any entries in the b column which are opposite an EVEN NUMBER in the a column and add the b column. The result is the desired product. For example, for $a = 27, b = 35$

$$\begin{array}{rr} 27 & 35 \\ 13 & 70 \\ 6 & \cancel{140} \\ 3 & 280 \\ 1 & \underline{560} \\ & 945 \end{array}$$

Russian Roulette

Russian roulette is a GAME of chance in which one or more of the six chambers of a gun are filled with bullets, the magazine is rotated at random, and the gun is shot. The shooter bets on whether the chamber which rotates into place will be loaded. If it is, he loses not only his bet but his life.

A modified version is considered by Blom *et al.* (1996) and Blom (1989). In this variant, the revolver is loaded with a single bullet, and two duelists alternately spin the chamber and fire at themselves until one is killed. The probability that the first duelist is killed is then 6/11.

References

Blom, G. *Probabilities and Statistics: Theory and Applications.* New York: Springer-Verlag, p. 32, 1989.

Blom, G.; Englund, J,.-E.; and Sandell, D. "General Russian Roulette." *Math. Mag.* **69**, 293–297, 1996.

Ryser Formula

A formula for the PERMANENT of a MATRIX

$$\operatorname{perm}(a_{ij}) = (-1)^n \sum_{s \subseteq \{1,\ldots,n\}} (-1)^{|s|} \prod_{i=1}^{n} \sum_{j \in s} a_{ij},$$

where the SUM is over all SUBSETS of $\{1, \ldots, n\}$, and $|s|$ is the number of elements in s. The formula can be optimized by picking the SUBSETS so that only a single element is changed at a time (which is precisely a GRAY CODE), reducing the number of additions from n^2 to n.

It turns out that the number of disks moved after the kth step in the TOWERS OF HANOI is the same as the element which needs to be added or deleted in the kth ADDEND of the RYSER FORMULA (Gardner 1988, Vardi 1991, p. 111)

see also DETERMINANT, GRAY CODE, PERMANENT, TOWERS OF HANOI

References

Gardner, M. "The Icosian Game and the Tower of Hanoi." Ch. 6 in *The Scientific American Book of Mathematical Puzzles & Diversions.* New York: Simon and Schuster, 1959.

Knuth, D. E. *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, p. 497, 1981.

Nijenhuis, A. and Wilf, H. Chs. 7–8 in *Combinatorial Algorithms.* New York: Academic Press, 1975.

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 111, 1991.

S

s-Additive Sequence

A generalization of an ULAM SEQUENCE in which each term is the SUM of two earlier terms in exactly s ways. (s,t)-additive sequences are a further generalization in which each term has exactly s representations as the SUM of t distinct earlier numbers. It is conjectured that 0-additive sequences ultimately have periodic differences of consecutive terms (Guy 1994, p. 233).

see also GREEDY ALGORITHM, STÖHR SEQUENCE, ULAM SEQUENCE

References

Finch, S. R. "Conjectures about s-Additive Sequences." *Fib. Quart.* **29**, 209–214, 1991.

Finch, S. R. "Are 0-Additive Sequences Always Regular?" *Amer. Math. Monthly* **99**, 671–673, 1992.

Finch, S. R. "On the Regularity of Certain 1-Additive Sequences." *J. Combin. Th. Ser. A.* **60**, 123–130, 1992.

Finch, S. R. "Patterns in 1-Additive Sequences." *Experiment. Math.* **1**, 57–63, 1992.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 110 and 233, 1994.

Ulam, S. M. *Problems in Modern Mathematics.* New York: Interscience, p. ix, 1964.

s-Cluster

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let an $n \times n$ MATRIX have entries which are either 1 (with probability p) or 0 (with probability $q = 1 - p$). An s-cluster is an isolated group of s adjacent (i.e., horizontally or vertically connected) 1s. Let C_n be the total number of "SITE" clusters. Then the value

$$K_S(p) = \lim_{n\to\infty} \frac{\langle C_n\rangle}{n^2}, \qquad (1)$$

called the MEAN CLUSTER COUNT PER SITE or MEAN CLUSTER DENSITY, exists. Numerically, it is found that $K_S(1/2) \approx 0.065770\ldots$ (Ziff *et al.* 1997).

Considering instead "BOND" clusters (where numbers are assigned to the edges of a grid) and letting C_n be the total number of bond clusters, then

$$K_B(p) \equiv \lim_{n\to\infty} \frac{\langle C_n\rangle}{n^2} \qquad (2)$$

exists. The analytic value is known for $p = 1/2$,

$$K_B(\tfrac{1}{2}) = \tfrac{3}{2}\sqrt{3} - \tfrac{41}{16} \qquad (3)$$

(Ziff *et al.* 1997).

see also BOND PERCOLATION, PERCOLATION THEORY, s-RUN, SITE PERCOLATION

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/rndprc/rndprc.html`.

Temperley, H. N. V. and Lieb, E. H. "Relations Between the 'Percolation' and 'Colouring' Problem and Other Graph-Theoretical Problems Associated with Regular Planar Lattices; Some Exact Results for the 'Percolation' Problem." *Proc. Roy. Soc. London A* **322**, 251–280, 1971.

Ziff, R.; Finch, S.; and Adamchik, V. "Universality of Finite-Sized Corrections to the Number of Percolation Clusters." *Phys. Rev. Let.* To appear, 1998.

s-Run

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let $\mathbf{v}$ be a n-VECTOR whose entries are each 1 (with probability p) or 0 (with probability $q = 1 - p$). An s-run is an isolated group of s consecutive 1s. Ignoring the boundaries, the total number of runs R_n satisfies

$$K_n = \frac{\langle R_n\rangle}{n} = (1-p)^2 \sum_{s=1}^{n} p^s = p(1-p)(1-p^n),$$

so

$$K(p) \equiv \lim_{n\to\infty} K_n = p(1-p),$$

which is called the MEAN RUN COUNT PER SITE or MEAN RUN DENSITY in PERCOLATION THEORY.

see also PERCOLATION THEORY, s-CLUSTER

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/rndprc/rndprc.html`.

S-Signature

see SIGNATURE (RECURRENCE RELATION)

Saalschützian

For a GENERALIZED HYPERGEOMETRIC FUNCTION

$${}_{p+1}F_p\left[\begin{matrix}\alpha_1, \alpha_2, \ldots, \alpha_{p+1} \\ \beta_1, \beta_2, \ldots, \beta_p\end{matrix}; z\right],$$

the Saalschützian S is defined if

$$\sum \beta = \sum \alpha + 1.$$

see also GENERALIZED HYPERGEOMETRIC FUNCTION

Saalschütz's Theorem

$${}_3F_2\left[\begin{matrix}-x, -y, -z \\ n+1, -x-y-z\end{matrix}\right] = \frac{\Gamma(n+1)\Gamma(x+y+n+1)}{\Gamma(x+n+1)\Gamma(y+n+1)} \times \frac{\Gamma(y+z+n+1)\Gamma(z+x+n+1)}{\Gamma(z+n+1)\Gamma(x+y+z+n+1)},$$

where ${}_3F_2(a,b,c;d,e;z)$ is a GENERALIZED HYPERGEOMETRIC FUNCTION and $\Gamma(z)$ is the GAMMA FUNCTION.

It can be derived from the DOUGALL-RAMANUJAN IDENTITY and written in the symmetric form

$$ {}_3F_2(a,b,c;d,e;1) = \frac{(d-a)_{|c|}(d-b)_{|c|}}{d_{|c|}(d-a-b)_{|c|}} $$

for $d+e=a+b+c+1$ with c a negative integer and $(a)_n$ the POCHHAMMER SYMBOL (Petkovšek *et al.* 1996).

see also DOUGALL-RAMANUJAN IDENTITY, GENERALIZED HYPERGEOMETRIC FUNCTION

References

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, pp. 43 and 126, 1996.

Saddle

A SURFACE possessing a SADDLE POINT.

see also HYPERBOLIC PARABOLOID, MONKEY SADDLE, SADDLE POINT (FUNCTION)

Saddle-Node Bifurcation

see FOLD BIFURCATION

Saddle Point (Fixed Point)

see HYPERBOLIC FIXED POINT (DIFFERENTIAL EQUATIONS), HYPERBOLIC FIXED POINT (MAP)

Saddle Point (Function)

A POINT of a FUNCTION or SURFACE which is a STATIONARY POINT but not an EXTREMUM. An example of a 1-D FUNCTION with a saddle point is $f(x) = x^3$, which has

$$ \begin{aligned} f'(x) &= 3x^2 \\ f''(x) &= 6x \\ f'''(x) &= 6. \end{aligned} $$

This function has a saddle point at $x_0 = 0$ by the EXTREMUM TEST since $f''(x_0) = 0$ and $f'''(x_0) = 6 \neq 0$. An example of a SURFACE with a saddle point is the MONKEY SADDLE.

Saddle Point (Game)

For a general two-player ZERO-SUM GAME,

$$ \min_{i\leq m}\min_{j\leq n} a_{ij} \leq \min_{j\leq n}\max_{i\leq m} a_{ij}. $$

If the two are equal, then write

$$ \min_{i\leq m}\min_{j\leq n} a_{ij} = \min_{j\leq n}\max_{i\leq m} a_{ij} \equiv v, $$

where v is called the VALUE of the GAME. In this case, there exist optimal strategies for the first and second players.

A NECESSARY and SUFFICIENT condition for a saddle point to exist is the presence of a PAYOFF MATRIX element which is both a minimum of its row and a maximum of its column. A GAME may have more than one saddle point, but all must have the same VALUE.

see also GAME, PAYOFF MATRIX, VALUE

References

Dresher, M. "Saddle Points." §1.5 in *The Mathematics of Games of Strategy: Theory and Applications.* New York: Dover, pp. 12–14, 1981.

Llewellyn, D. C.; Tovey, C.; and Trick, M. "Finding Saddlepoints of Two-Person, Zero Sum Games." *Amer. Math. Monthly* **95**, 912–918, 1988.

Safarevich Conjecture

see SHAFAREVICH CONJECTURE

Safe

A position in a GAME is safe if the person who plays next will lose.

see also GAME, UNSAFE

Sagitta

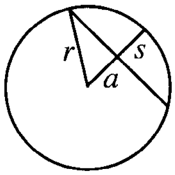

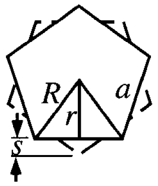

The PERPENDICULAR distance s from an ARC's MIDPOINT to the CHORD across it, equal to the RADIUS r minus the APOTHEM a,

$$ s = r - a. \tag{1} $$

For a regular POLYGON of side length a,

$$ \begin{aligned} s \equiv R - r &= \tfrac{1}{2}a\left[\csc\left(\frac{\pi}{n}\right) - \cot\left(\frac{\pi}{n}\right)\right] \\ &= \tfrac{1}{2}a\tan\left(\frac{\pi}{2n}\right) && (2) \\ &= r\tan\left(\frac{\pi}{n}\right)\tan\left(\frac{\pi}{2n}\right) && (3) \\ &= 2R\sin^2\left(\frac{\pi}{2n}\right), && (4) \end{aligned} $$

where R is the CIRCUMRADIUS, r the INRADIUS, a is the side length, and n is the number of sides.

see also APOTHEM, CHORD, SECTOR, SEGMENT

Saint Andrew's Cross

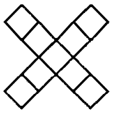

A GREEK CROSS rotated by 45°, also called the crux decussata. The MULTIPLICATION SIGN × is based on Saint Andrew's cross (Bergamini 1969).

see also CROSS, GREEK CROSS, MULTIPLICATION SIGN

References

Bergamini, D. *Mathematics.* New York: Time-Life Books, p. 11, 1969.

Saint Anthony's Cross

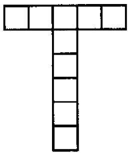

A CROSS also called the tau cross or crux commissa.

see also CROSS

Saint Petersburg Paradox

Consider a game in which a player bets on whether a given TOSS of a COIN will turn up heads or tails. If he bets $1 that heads will turn up on the first throw, $2 that heads will turn up on the second throw (if it did not turn up on the first), $4 that heads will turn up on the third throw, etc., his expected *payoff* is

$$\tfrac{1}{2}(1)+\tfrac{1}{4}(2)+\tfrac{1}{8}(4)+\ldots=\tfrac{1}{2}+\tfrac{1}{2}+\tfrac{1}{2}+\ldots=\infty.$$

Apparently, the first player can be in the hole by any amount of money and still come out ahead in the end. This PARADOX was first proposed by Daniel Bernoulli.

The paradox arises as a result of muddling the distinction between the amount of the final payoff and the net amount won in the game. It is misleading to consider the payoff without taking into account the amount lost on previous bets, as can be shown as follows. At the time the player first wins (say, on the nth toss), he will have lost

$$\sum_{k=1}^{n-1} 2^{k-1} = 2^{n-1}-1$$

dollars. In this toss, however, he wins 2^{n-1} dollars. This means that the net gain for the player is a whopping $1, no matter how many tosses it takes to finally win. As expected, the large payoff after a long run of tails is exactly balanced by the large amount that the player has to invest.

In fact, by noting that the probability of winning on the nth toss is $1/2^n$, it can be seen that the probability distribution for the number of tosses needed to win is simply a GEOMETRIC DISTRIBUTION with $p = 1/2$.

see also COIN TOSSING, GAMBLER'S RUIN, GEOMETRIC DISTRIBUTION, MARTINGALE

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 44–45, 1987.

Gardner, M. *The Scientific American Book of Mathematical Puzzles & Diversions.* New York: Simon and Schuster, pp. 51–52, 1959.

Kamke, E. *Einführung in die Wahrscheinlichkeitstheorie.* Leipzig, Germany, pp. 82–89, 1932.

Keynes, J. M. K. "The Application of Probability to Conduct." In *The World of Mathematics, Vol. 2* (Ed. K. Newman). Redmond, WA: Microsoft Press, 1988.

Kraitchik, M. "The Saint Petersburg Paradox." §6.18 in *Mathematical Recreations.* New York: W. W. Norton, pp. 138–139, 1942.

Todhunter, I. §391 in *History of the Mathematical Theory of Probability.* New York: Chelsea, p. 221, 1949.

Sal

see WALSH FUNCTION

Salamin Formula

see BRENT-SALAMIN FORMULA

Salem Constants

Each point of the PISOT-VIJAYARAGHAVAN CONSTANTS S is a LIMIT POINT from both sides of a set T known as the Salem constants (Salem 1945). The Salem constants are algebraic INTEGERS > 1 in which one or more of the conjugates is on the UNIT CIRCLE with the others inside (Le Lionnais 1983, p. 150). The smallest known Salem number was found by Lehmer (1933) as the largest REAL ROOT of

$$x^{10}+x^9-x^7-x^6-x^5-x^4-x^3+x+1=0,$$

which is

$$\sigma_1 = 1.176280818\ldots$$

(Le Lionnais 1983, p. 35). Boyd (1977) found the following table of small Salem numbers, and suggested that σ_1, σ_2, σ_3, and σ_4 are the smallest Salem numbers. The NOTATION 1 1 0 −1 −1 −1 is short for 1 1 0 −1 −1 −1 −1 −1 0 1 1, the coefficients of the above polynomial.

k	σ_k	°	polynomial
1	1.1762808183	10	1 1 0 −1 −1 −1
2	1.1883681475	18	1 −1 1 −1 0 0 −1 1 −1 1
3	1.2000265240	14	1 0 0 −1 −1 0 0 1
4	1.2026167437	14	1 0 −1 0 0 0 0 −1
5	1.2163916611	10	1 0 0 0 −1 −1
6	1.2197208590	18	1 −1 0 0 0 0 0 0 −1 1
7	1.2303914344	10	1 0 0 −1 0 −1
8	1.2326135486	20	1 −1 0 0 0 −1 1 0 0 −1 1
9	1.2356645804	22	1 0 −1 −1 0 0 0 1 1 0 −1 −1
10	1.2363179318	16	1 −1 0 0 0 0 0 0 −1
11	1.2375048212	26	1 0 −1 0 0 −1 0 0 −1 0 1 0 0 1
12	1.2407264237	12	1 −1 1 −1 0 0 −1
13	1.2527759374	18	1 0 0 0 0 0 −1 −1 −1 −1
14	1.2533306502	20	1 0 −1 0 0 −1 0 0 0 0 0
15	1.2550935168	14	1 0 −1 −1 0 1 0 −1
16	1.2562211544	18	1 −1 0 0 −1 1 0 0 0 −1
17	1.2601035404	24	1 −1 0 0 −1 1 0 −1 1 −1 0 1 −1
18	1.2602842369	22	1 −1 0 −1 1 0 0 0 −1 1 −1 1
19	1.2612309611	10	1 0 −1 0 0 −1
20	1.2630381399	26	1 −1 0 0 0 0 −1 0 0 0 0 0 0 1
21	1.2672964425	14	1 −1 0 0 0 0 −1 1
22	1.2806381563	8	1 0 0 −1 −1
23	1.2816913715	26	1 0 0 0 0 0 −1 −1 −1 −1 −1 −1 −1 −1
24	1.2824955606	20	1 −2 2 −2 2 −2 1 0 −1 1 −1
25	1.2846165509	18	1 0 0 0 −1 0 −1 −1 0 −1
26	1.2847468215	26	1 −2 1 1 −2 1 0 0 −1 1 0 −1 1 −1
27	1.2850993637	30	1 0 0 0 0 −1 −1 −1 −1 −1 −1 0 0 0 0 1
28	1.2851215202	30	1 −2 2 −2 1 0 −1 2 −2 1 0 −1 1 −1 1 −1
29	1.2851856708	30	1 −1 0 0 0 0 0 0 −1 0 0 0 −1 0 0 −1
30	1.2851967268	26	1 0 −1 −1 0 0 0 1 0 −1 −1 0 1 1
31	1.2851991792	44	1 −1 0 0 0 0 0 −1 0 0 0 −1 0 0 0 0 0 0 0 1 0 0 1
32	1.2852354362	30	1 0 −1 0 0 −1 −1 0 0 0 1 0 0 1 0 −1
33	1.2854090648	34	1 −1 0 0 −1 1 −1 0 1 −1 1 0 −1 1 −1 0 1 −1
34	1.2863959668	18	1 −2 2 −2 2 −2 2 −3 3 −3
35	1.2867301820	26	1 −1 0 0 −1 1 −1 0 1 −1 1 0 −1 1
36	1.2917414257	24	1 −1 0 0 0 0 −1 0 0 0 0 0 0
37	1.2920391602	20	1 0 −1 0 0 −1 0 0 −1 0 1
38	1.2934859531	10	1 0 −1 −1 0 1
39	1.2956753719	18	1 −1 0 0 −1 1 −1 0 1 −1

see also PISOT-VIJAYARAGHAVAN CONSTANTS

References

Boyd, D. W. "Small Salem Numbers." *Duke Math. J.* **44**, 315–328, 1977.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, 1983.

Lehmer, D. H. "Factorization of Certain Cyclotomic Functions." *Ann. Math., Ser. 2* **34**, 461–479, 1933.

Salem, R. "Power Series with Integral Coefficients." *Duke Math. J.* **12**, 153–172, 1945.

Stewart, C. L. "Algebraic Integers whose Conjugates Lie Near the Unit Circle." *Bull. Soc. Math. France* **106**, 169–176, 1978.

Salesman Problem

see TRAVELING SALESMAN PROBLEM

Salient Point

A point at which two noncrossing branches of a curve meet with different tangents.

see also CUSP

Salinon

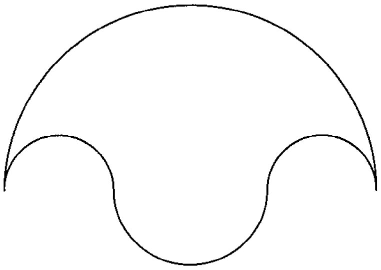

The above figure formed from four connected SEMICIRCLES. The word salinon is Greek for salt cellar, which the figure resembles.

see also ARBELOS, PIECEWISE CIRCULAR CURVE, SEMICIRCLE

Salmon's Theorem

Given a track bounded by two confocal ELLIPSES, if a ball is rolled so that its trajectory is tangent to the inner ELLIPSE, the ball's trajectory will be tangent to the inner ELLIPSE following all subsequent caroms as well.

References

Salmon, G. *A Treatise on Conic Sections.* New York: Chelsea, p. 182, 1954.

Saltus

The word saltus has two different meanings: either a jump or an oscillation of a function.

Sample Proportion

Let there be x successes out of n BERNOULLI TRIALS. The sample proportion is the fraction of samples which were successes, so

$$\hat{p} = \frac{x}{n}. \tag{1}$$

For large n, $\hat{p}$ has an approximately NORMAL DISTRIBUTION. Let RE be the RELATIVE ERROR and SE the STANDARD ERROR, then

$$\langle p \rangle = p \tag{2}$$

$$\mathrm{SE}(\hat{p}) \equiv \sigma(\hat{p}) = \sqrt{\frac{p(1-p)}{n}} \tag{3}$$

$$\mathrm{RE}(\hat{p}) = \sqrt{\frac{2\hat{p}(1-\hat{p})}{n}}\,\mathrm{erf}^{-1}(\mathrm{CI}), \tag{4}$$

where CI is the CONFIDENCE INTERVAL and $\operatorname{erf} x$ is the ERF function. The number of tries needed to determine p with RELATIVE ERROR RE and CONFIDENCE INTERVAL CI is

$$n = \frac{2[\mathrm{erf}^{-1}(\mathrm{CI})]^2\hat{p}(1-\hat{p})}{(\mathrm{RE})^2}. \tag{5}$$

Sample Space

Informally, the sample space for a given set of events is the set of all possible values the events may assume. Formally, the set of possible events for a given variate forms a SIGMA ALGEBRA, and sample space is defined as the largest set in the SIGMA ALGEBRA.

see also PROBABILITY SPACE, RANDOM VARIABLE, SIGMA ALGEBRA, STATE SPACE

Sample Variance

To estimate the population VARIANCE from a sample of N elements with a priori *unknown* MEAN (i.e., the MEAN is estimated from the sample itself), we need an unbiased ESTIMATOR for σ. This is the k-STATISTIC k_2, where

$$k_2 = \frac{N}{N-1} m_2 \tag{1}$$

and $m_2 \equiv s^2$ is the sample variance

$$s^2 \equiv \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2. \tag{2}$$

Note that some authors prefer the definition

$$s'^2 \equiv \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2, \tag{3}$$

since this makes the sample variance an UNBIASED ESTIMATOR for the population variance.

see also k-STATISTIC, VARIANCE

Sampling

For infinite precision sampling of a band-limited signal at the NYQUIST FREQUENCY, the signal-to-noise ratio after N_q samples is

$$\text{SNR} = \frac{\langle r_\infty \rangle}{\sigma_\infty} = \frac{\rho \sigma^2}{\sigma^2 {N_q}^{-1/2} \sqrt{1+\rho^2}} = \frac{\rho}{\sqrt{1+\rho^2}} \sqrt{N_q}, \tag{1}$$

where ρ is the normalized cross-correlation COEFFICIENT

$$\rho \equiv \frac{\langle x(t) \rangle \langle y(t) \rangle}{\sqrt{\langle x^2(t) \rangle \langle y^2(t) \rangle}}. \tag{2}$$

For $\rho \ll 1$,

$$\text{SNR} \approx \rho \sqrt{N_q}. \tag{3}$$

The identical result is obtained for oversampling. For undersampling, the SNR decreases (Thompson *et al.* 1986).

see also NYQUIST SAMPLING, OVERSAMPLING, QUANTIZATION EFFICIENCY, SAMPLING FUNCTION, SHANNON SAMPLING THEOREM, SINC FUNCTION

References

Feuer, A. *Sampling in Digital Signal Processing and Control.* Boston, MA: Birkhäuser, 1996.

Thompson, A. R.; Moran, J. M.; and Swenson, G. W. Jr. *Interferometry and Synthesis in Radio Astronomy.* New York: Wiley, pp. 214–216, 1986.

Sampling Function

The 1-D sampling function is given by

$$S(x) = \sum_{n=-\infty}^{\infty} \delta(x - n\Delta x),$$

where δ is the DIRAC DELTA FUNCTION. The 2-D version is

$$S(u,v) = \sum \delta(u - u_n, v - v_n),$$

which can be weighted to

$$S(u,v) = \sum R_n T_n D_n \delta(u - u_n, v - v_n),$$

where R_n is a reliability weight, D_n is a density weight (WEIGHTING FUNCTION), and T_n is a taper.

see also SHAH FUNCTION, SINC FUNCTION

Sampling Theorem

In order for a band-limited (i.e., one with a zero POWER SPECTRUM for frequencies $f > B$) baseband ($f > 0$) signal to be reconstructed fully, it must be sampled at a rate $f \geq 2B$. A signal sampled at $f = 2B$ is said to be NYQUIST SAMPLED, and $f = 2B$ is called the NYQUIST FREQUENCY. No information is lost if a signal is sampled at the NYQUIST FREQUENCY, and no additional information is gained by sampling faster than this rate.

see also ALIASING, NYQUIST FREQUENCY, NYQUIST SAMPLING, OVERSAMPLING

San Marco Fractal

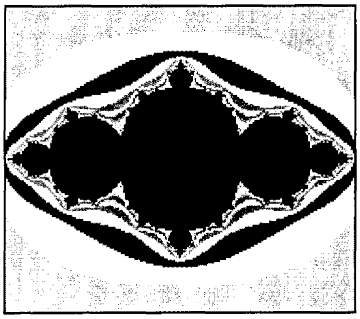

The FRACTAL $J(-3/4, 0)$, where J is the JULIA SET. It slightly resembles the MANDELBROT SET.

see also DOUADY'S RABBIT FRACTAL, JULIA SET, MANDELBROT SET

References

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, p. 173, 1991.

Sandwich Theorem

see HAM SANDWICH THEOREM, SQUEEZING THEOREM

Sard's Theorem

The set of "critical values" of a MAP $u : \mathbb{R}^n \to \mathbb{R}^n$ of CLASS C^1 has LEBESGUE MEASURE 0 in $\mathbb{R}^n$.

see also CLASS (MAP), LEBESGUE MEASURE

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 682, 1980.

Šarkovskii's Theorem

Order the NATURAL NUMBERS as follows:

$$\begin{aligned} 3 \prec 5 \prec 7 \prec 9 \prec 11 \prec 13 \prec 15 \prec \ldots \prec 2\cdot 3 \prec 2\cdot 5 \prec 2\cdot 7 \\ \prec 2\cdot 9 \prec \ldots \prec 2\cdot 2\cdot 3 \prec 2\cdot 2\cdot 5 \prec 2\cdot 2\cdot 7 \\ \prec 2\cdot 2\cdot 9 \prec \ldots \prec 2\cdot 2\cdot 2\cdot 3 \prec \ldots \\ \prec 2^5 \prec 2^4 \prec 2^3 \prec 2^2 \prec 2 \prec 1. \end{aligned}$$

Now let F be a CONTINUOUS FUNCTION from the REALS to the REALS and suppose $p \prec q$ in the above ordering. Then if F has a point of LEAST PERIOD p, then F also has a point of LEAST PERIOD q.

A special case of this general result, also known as Šarkovskii's theorem, states that if a CONTINUOUS REAL function has a PERIODIC POINT with period 3, then there is a PERIODIC POINT of period n for every INTEGER n.

A converse to Šarkovskii's theorem says that if $p \prec q$ in the above ordering, then we can find a CONTINUOUS FUNCTION which has a point of LEAST PERIOD q, but does not have any points of LEAST PERIOD p (Elaydi 1996). For example, there is a CONTINUOUS FUNCTION with no points of LEAST PERIOD 3 but having points of all other LEAST PERIODS.

see also LEAST PERIOD

References

Conway, J. H. and Guy, R. K. "Periodic Points." In *The Book of Numbers.* New York: Springer-Verlag, pp. 207–208, 1996.
Devaney, R. L. *An Introduction to Chaotic Dynamical Systems, 2nd ed.* Reading, MA: Addison-Wesley, 1989.
Elaydi, S. "On a Converse of Sharkovsky's Theorem." *Amer. Math. Monthly* **103**, 386–392, 1996.
Ott, E. *Chaos in Dynamical Systems.* New York: Cambridge University Press, p. 49, 1993.
Sharkovsky, A. N. "Co-Existence of Cycles of a Continuous Mapping of a Line onto Itself." *Ukranian Math. Z.* **16**, 61–71, 1964.
Stefan, P. "A Theorem of Sharkovsky on the Existence of Periodic Orbits of Continuous Endomorphisms of the Real Line." *Comm. Math. Phys.* **54**, 237–248, 1977.

Sárközy's Theorem

A partial solution to the ERDŐS SQUAREFREE CONJECTURE which states that the BINOMIAL COEFFICIENT $\binom{2n}{n}$ is never SQUAREFREE for all sufficiently large $n \geq n_0$. Sárközy (1985) showed that if $s(n)$ is the square part of the BINOMIAL COEFFICIENT $\binom{2n}{n}$, then

$$\ln s(n) \sim (\sqrt{2} - 2)\zeta(\tfrac{1}{2})\sqrt{n},$$

where $\zeta(z)$ is the RIEMANN ZETA FUNCTION. An upper bound on n_0 of $2^{8,000}$ has been obtained.

see also BINOMIAL COEFFICIENT, ERDŐS SQUAREFREE CONJECTURE

References

Erdős, P. and Graham, R. L. *Old and New Problems and Results in Combinatorial Number Theory.* Geneva, Switzerland: L'Enseignement Mathématique Université de Genève, Vol. 28, 1980.
Sander, J. W. "A Story of Binomial Coefficients and Primes." *Amer. Math. Monthly* **102**, 802–807, 1995.
Sárközy, A. "On the Divisors of Binomial Coefficients, I." *J. Number Th.* **20**, 70–80, 1985.
Vardi, I. "Applications to Binomial Coefficients." *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 25–28, 1991.

Sarrus Linkage

A LINKAGE which converts circular to linear motion using a hinged square.

see also HART'S INVERSOR, LINKAGE, PEAUCELLIER INVERSOR

Sarrus Number

see POULET NUMBER

SAS Theorem

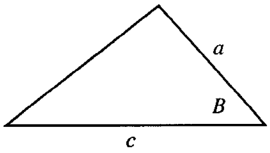

Specifying two sides and the ANGLE between them uniquely determines a TRIANGLE. Let b be the base length and h be the height. Then the AREA is

$$K = \tfrac{1}{2}ch = \tfrac{1}{2}ac\sin B. \tag{1}$$

The length of the third side is given by the LAW OF COSINES,

$$b^2 = a^2 + c^2 - 2ac\cos B,$$

so

$$b = \sqrt{a^2 + c^2 - 2ac\cos B}. \tag{2}$$

Using the LAW OF SINES

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C} \tag{3}$$

then gives the two other ANGLES as

$$A = \sin^{-1}\left(\frac{a\sin B}{\sqrt{a^2 + c^2 - 2ac\cos B}}\right) \tag{4}$$

$$C = \sin^{-1}\left(\frac{c\sin B}{\sqrt{a^2 + c^2 - 2ac\cos B}}.\right) \tag{5}$$

see also AAA THEOREM, AAS THEOREM, ASA THEOREM, ASS THEOREM, SSS THEOREM, TRIANGLE

Satellite Knot

Let K_1 be a knot inside a Torus. Now knot the Torus in the shape of a second knot (called the Companion Knot) K_2. Then the new knot resulting from K_1 is called the satellite knot K_3. Composite Knots are special cases of satellite knots. The only Knots which are not Hyperbolic Knots are Torus Knots and satellite knots (including Composite Knots). No satellite knot is an Almost Alternating Knot.

see also Almost Alternating Knot, Companion Knot, Composite Knot, Hyperbolic Knot, Torus Knot

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 115–118, 1994.

Satisfiability Problem

Deciding whether a given Boolean formula in conjunctive normal form has an assignment that makes the formula "true." In 1971, Cook showed that the problem is NP-Complete.

see also Boolean Algebra

References

Cook, S. A. and Mitchell, D. G. "Finding Hard Instances of the Satisfiability Problem: A Survey." In *Satisfiability problem: theory and applications (Piscataway, NJ, 1996). Theoret. Comput. Sci., Vol. 35.* Providence, RI: Amer. Math. Soc., pp. 1–17, 1997.

Sausage Conjecture

In n-D for $n \geq 5$ the arrangement of Hyperspheres whose Convex Hull has minimal Content is always a "sausage" (a set of Hyperspheres arranged with centers along a line), independent of the number of n-spheres. The Conjecture was proposed by Fejes Tóth, and solved for dimensions ≥ 42 by Betke *et al.* (1994) and Betke and Henk (1998).

see also Content, Convex Hull, Hypersphere, Hypersphere Packing, Sphere Packing

References

Betke, U.; Henk, M.; and Wills, J. M. "Finite and Infinite Packings." *J. Reine Angew. Math.* **453**, 165–191, 1994.

Betke, U. and Henk, M. "Finite Packings of Spheres." *Discrete Comput. Geom.* **19**, 197–227, 1998.

Croft, H. T.; Falconer, K. J.; and Guy, R. K. Problem D9 in *Unsolved Problems in Geometry.* New York: Springer-Verlag, 1991.

Fejes Tóth, L. "Research Problems." *Periodica Methematica Hungarica* **6**, 197–199, 1975.

Savitzky-Golay Filter

A low-pass filter which is useful for smoothing data.

see also Filter

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 183 and 644–645, 1992.

Savoy Knot

see Figure-of-Eight Knot

Scalar

A one-component quantity which is invariant under Rotations of the coordinate system.

see also Pseudoscalar, Scalar Field, Scalar Function, Scalar Potential, Scalar Triple Product, Tensor, Vector

Scalar Curvature

see Curvature Scalar

Scalar Field

A Map $f : \mathbb{R}^n \mapsto \mathbb{R}$ which assigns each $\mathbf{x}$ a Scalar Function $f(\mathbf{x})$.

see also Vector Field

References

Morse, P. M. and Feshbach, H. "Scalar Fields." §1.1 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 4–8, 1953.

Scalar Function

A function $f(x_1, \ldots, x_n)$ of one or more variables whose Range is one-dimensional, as compared to a Vector Function, whose Range is three-dimensional (or, in general, n-dimensional).

see also Complex Function, Real Function, Vector Function

Scalar Potential

A conservative Vector Field (for which the Curl $\nabla \times \mathbf{F} = \mathbf{0}$) may be assigned a scalar potential

$$\begin{aligned}\phi(x,y,z) - \phi(0,0,0) &\equiv -\int_C \mathbf{F} \cdot d\mathbf{s} \\ &= -\int_{(0,0,0)}^{(x,0,0)} F_1(t,0,0)\,dt + \int_{(x,0,0)}^{(x,y,0)} F_2(x,t,0)\,dt \\ &\quad + \int_{(x,y,0)}^{x,y,z} F_3(x,y,t)\,dt,\end{aligned}$$

where $\int_C \mathbf{F} \cdot d\mathbf{s}$ is a Line Integral.

see also Potential Function, Vector Potential

Scalar Triple Product

The Vector product

$$\begin{aligned}[\mathbf{A},\mathbf{B},\mathbf{C}] &\equiv \mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{B} \cdot (\mathbf{C} \times \mathbf{A}) \\ &= \mathbf{C} \cdot (\mathbf{A} \times \mathbf{B}) = \begin{vmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{vmatrix},\end{aligned}$$

which yields a Scalar (actually, a Pseudoscalar).

The VOLUME of a PARALLELEPIPED whose sides are given by the vectors **A**, **B**, and **C** is

$$V_{\text{parallelepiped}} = |\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C})|.$$

see also CROSS PRODUCT, DOT PRODUCT, PARALLELEPIPED, VECTOR TRIPLE PRODUCT

References

Arfken, G. "Triple Scalar Product, Triple Vector Product." §1.5 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 26–33, 1985.

Scale

see BASE (NUMBER)

Scale Factor

For a diagonal METRIC TENSOR $g_{ij} = g_{ii}\delta_{ij}$, where δ_{ij} is the KRONECKER DELTA, the scale factor is defined by

$$h_i \equiv \sqrt{g_{ii}}. \tag{1}$$

The LINE ELEMENT (first FUNDAMENTAL FORM) is then given by

$$\begin{aligned} ds^2 &= g_{11}\,dx_{11}{}^2 + g_{22}\,dx_{22}{}^2 + g_{33}\,dx_{33}{}^2 &(2)\\ &= h_1{}^2\,dx_{11}{}^2 + h_2{}^2\,dx_{22}{}^2 + h_3{}^2\,dx_{33}{}^2. &(3)\end{aligned}$$

The scale factor appears in vector derivatives of coordinates in CURVILINEAR COORDINATES.

see also CURVILINEAR COORDINATES, FUNDAMENTAL FORMS, LINE ELEMENT

Scalene Triangle

A TRIANGLE with three unequal sides.

see also ACUTE TRIANGLE, EQUILATERAL TRIANGLE, ISOSCELES TRIANGLE, OBTUSE TRIANGLE, TRIANGLE

Scaling

Increasing a plane figure's linear dimensions by a scale factor s increases the PERIMETER $p' \to sp$ and the AREA $A' \to s^2 A$.

see also DILATION, EXPANSION, FRACTAL, SELF-SIMILARITY

Scattering Operator

An OPERATOR relating the past asymptotic state of a DYNAMICAL SYSTEM governed by the Schrödinger equation

$$i\frac{d}{dt}\psi(t) = H\psi(t)$$

to its future asymptotic state.

see also WAVE OPERATOR

Scattering Theory

The mathematical study of the SCATTERING OPERATOR and Schrödinger equation.

see also SCATTERING OPERATOR

References

Yafaev, D. R. *Mathematical Scattering Theory: General Theory.* Providence, RI: Amer. Math. Soc., 1996.

Schaar's Identity

A generalization of the GAUSSIAN SUM. For p and q of opposite PARITY (i.e., one is EVEN and the other is ODD), Schaar's identity states

$$\frac{1}{\sqrt{q}}\sum_{r=0}^{q-1} e^{-\pi i r^2 p/q} = \frac{e^{-\pi i/4}}{\sqrt{p}}\sum_{r=0}^{p-1} e^{\pi i r^2 q/p}.$$

see also GAUSSIAN SUM

References

Evans, R. and Berndt, B. "The Determination of Gauss Sums." *Bull. Amer. Math. Soc.* **5**, 107–129, 1981.

Schanuel's Conjecture

Let $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ be linearly independent over the RATIONALS $\mathbb{Q}$, then

$$\mathbb{Q}(\lambda_1, \ldots, \lambda_n, e^{\lambda_1}, \ldots, e^{\lambda_n})$$

has TRANSCENDENCE degree at least n over $\mathbb{Q}$. Schanuel's conjecture is a generalization of the LINDEMANN-WEIERSTRAß THEOREM. If the conjecture is true, then it follows that e and π are algebraically independent. Mcintyre (1991) proved that the truth of Schanuel's conjecture also guarantees that there are no unexpected exponential-algebraic relations on the INTEGERS $\mathbb{Z}$ (Marker 1996).

see also CONSTANT PROBLEM

References

Macintyre, A. "Schanuel's Conjecture and Free Exponential Rings." *Ann. Pure Appl. Logic* **51**, 241–246, 1991.

Marker, D. "Model Theory and Exponentiation." *Not. Amer. Math. Soc.* **43**, 753–759, 1996.

Schauder Fixed Point Theorem

Let A be a closed convex subset of a BANACH SPACE and assume there exists a continuous MAP T sending A to a countably compact subset $T(A)$ of A. Then T has fixed points.

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 543, 1980.

Schauder, J. "Der Fixpunktsatz in Funktionalräumen." *Studia Math.* **2**, 171–180, 1930.

Zeidler, E. *Applied Functional Analysis: Applications to Mathematical Physics.* New York: Springer-Verlag, 1995.

Scheme

A local-ringed SPACE which is locally isomorphic to an AFFINE SCHEME.

see also AFFINE SCHEME

References

Iyanaga, S. and Kawada, Y. (Eds.). "Schemes." §18E in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 69, 1980.

Schensted Correspondence

A correspondence between a PERMUTATION and a pair of YOUNG TABLEAUX.

see also PERMUTATION, YOUNG TABLEAU

References

Knuth, D. E. *The Art of Computer Programming, Vol. 3: Sorting and Searching, 2nd ed.* Reading, MA: Addison-Wesley, 1973.

Stanton, D. W. and White, D. E. §3.6 in *Constructive Combinatorics.* New York: Springer-Verlag, pp. 85–87, 1986.

Scherk's Minimal Surfaces

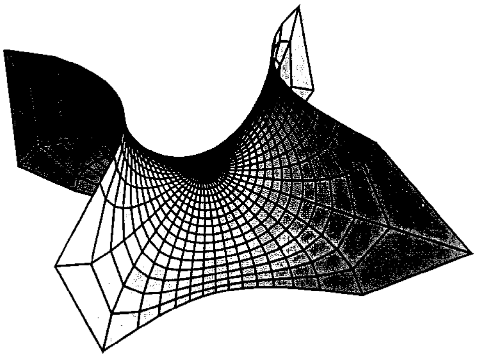

A class of MINIMAL SURFACES discovered by Scherk (1834) which were the first new surfaces discovered since Meusnier in 1776. Scherk's first surface is doubly periodic. Scherk's second surface, illustrated above, can be written parametrically as

$$\begin{aligned} x &= 2\Re[\ln(1+re^{i\theta}) - \ln(1-re^{i\theta})] \\ y &= \Re[4i\tan^{-1}(re^{i\theta})] \\ z &= \Re\left\{2i(-\ln[1-r^2e^{2i\theta}] + \ln[1+r^2e^{2i\theta}])\right\} \end{aligned}$$

for $\theta \in [0, 2\pi)$, and $r \in (0,1)$. Scherk's first surface has been observed to form in layers of block copolymers (Peterson 1988).

von Seggern (1993) calls

$$z = c\ln\left[\frac{\cos(2\pi y)}{\cos(2\pi x)}\right]$$

"Scherk's surface." Beautiful images of wood sculptures of Scherk surfaces are illustrated by Séquin.

References

Dickson, S. "Minimal Surfaces." *Mathematica J.* **1**, 38–40, 1990.

do Carmo, M. P. *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, p. 41, 1986.

Meusnier, J. B. "Mémoire sur la courbure des surfaces." *Mém. des savans étrangers* **10** (lu 1776), 477–510, 1785.

Peterson, I. "Geometry for Segregating Polymers." *Sci. News*, 151, Sep. 3, 1988.

Scherk, H. F. "Bemerkung über der kleinste Fläche innerhalb gegebener Grenzen." *J. Reine. angew. Math.* **13**, 185–208, 1834.

Thomas, E. L.; Anderson, D. M.; Henkee, C. S.; and Hoffman, D. "Periodic Area-Minimizing Surfaces in Block Copolymers." *Nature* **334**, 598–601, 1988.

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 304, 1993.

Wolfram Research "Mathematica Version 2.0 Graphics Gallery." `http://www.mathsource.com/cgi-bin/MathSource/Applications/Graphics/3D/0207-155`.

Schiffler Point

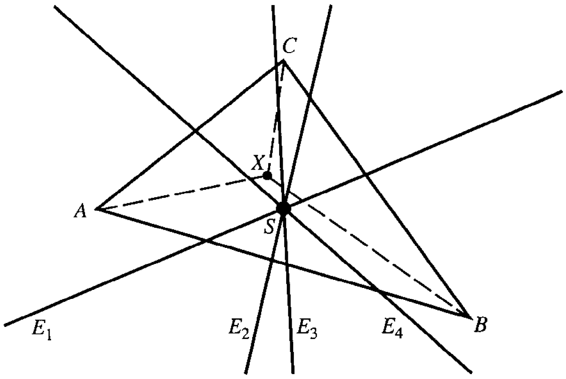

The CONCURRENCE S of the EULER LINES E_n of the TRIANGLES ΔXBC, ΔXCA, ΔXAB, and ΔABC where X is the INCENTER. The TRIANGLE CENTER FUNCTION is

$$\alpha = \frac{1}{\cos B + \cos C} = \frac{b+c-a}{b+c}.$$

References

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Kimberling, C. "Schiffler Point." `http://www.evansville.edu/~ck6/tcenters/recent/schiff.html`.

Schiffler, K.; Veldkamp, G. R.; and van der Spek, W. A. "Problem 1018 and Solution." *Crux Math.* **12**, 176–179, 1986.

Schinzel Circle

A CIRCLE having a given number of LATTICE POINTS on its CIRCUMFERENCE. The Schinzel circle halving n lattice points is given by the equation

$$\begin{cases} (x-\frac{1}{2})^2 + y^2 = \frac{1}{4}5^{k-1} & \text{for } n = 2k \text{ even} \\ (x-\frac{1}{3})^2 + y^2 = \frac{1}{9}5^{2k} & \text{for } n = 2k+1 \text{ odd.} \end{cases}$$

Note that these solutions do not necessarily have the smallest possible RADIUS. For example, while the Schinzel circle centered at (1/3, 0) and with radius 625/3 has nine lattice points on its circumference, so does the circle centered at (1/3, 0) with radius 65/3.

see also CIRCLE, CIRCLE LATTICE POINTS, KULIKOWSKI'S THEOREM, LATTICE POINT, SCHINZEL'S THEOREM, SPHERE

References

Honsberger, R. "Circles, Squares, and Lattice Points." Ch. 11 in *Mathematical Gems I.* Washington, DC: Math. Assoc. Amer., pp. 117–127, 1973.

Kulikowski, T. "Sur l'existence d'une sphère passant par un nombre donné aux coordonnées entières." *L'Enseignement Math. Ser. 2* **5**, 89–90, 1959.

Schinzel, A. "Sur l'existence d'un cercle passant par un nombre donné de points aux coordonnées entières." *L'Enseignement Math. Ser. 2* **4**, 71–72, 1958.

Sierpiński, W. "Sur quelques problèmes concernant les points aux coordonnées entières." *L'Enseignement Math. Ser. 2* **4**, 25–31, 1958.

Sierpiński, W. "Sur un problème de H. Steinhaus concernant les ensembles de points sur le plan." *Fund. Math.* **46**, 191–194, 1959.

Sierpiński, W. *A Selection of Problems in the Theory of Numbers.* New York: Pergamon Press, 1964.

Schinzel's Hypothesis

If $f_1(x), \ldots, f_s(x)$ are irreducible POLYNOMIALS with INTEGER COEFFICIENTS such that no INTEGER $n > 1$ divides $f_1(x), \ldots, f_s(x)$ for all INTEGERS x, then there should exist infinitely many x such that $f_1(x), \ldots, f_s(x)$ are simultaneous PRIME.

References

Schinzel, A. and Sierpiński, W. "Sur certaines hypothéses concernant les nombres premiers. Remarque." *Acta Arithm.* **4**, 185–208, 1958.

Schinzel's Theorem

For every POSITIVE INTEGER n, there exists a CIRCLE in the plane having exactly n LATTICE POINTS on its CIRCUMFERENCE. The theorem is based on the number $r(n)$ of integral solutions (x, y) to the equation

$$x^2 + y^2 = n, \tag{1}$$

given by

$$r(n) = 4(d_1 - d_3), \tag{2}$$

where d_1 is the number of divisors of n of the form $4k+1$ and d_3 is the number of divisors of the form $4k+3$. It explicitly identifies such circles (the SCHINZEL CIRCLES) as

$$\begin{cases} (x - \tfrac{1}{2})^2 + y^2 = \tfrac{1}{4}5^{k-1} & \text{for } n = 2k \\ (x - \tfrac{1}{3})^2 + y^2 = \tfrac{1}{9}5^{2k} & \text{for } n = 2k+1. \end{cases} \tag{3}$$

Note, however, that these solutions do not necessarily have the smallest possible radius.

see also BROWKIN'S THEOREM, KULIKOWSKI'S THEOREM, SCHINZEL CIRCLE

References

Honsberger, R. "Circles, Squares, and Lattice Points." Ch. 11 in *Mathematical Gems I.* Washington, DC: Math. Assoc. Amer., pp. 117–127, 1973.

Kulikowski, T. "Sur l'existence d'une sphère passant par un nombre donné aux coordonnées entières." *L'Enseignement Math. Ser. 2* **5**, 89–90, 1959.

Schinzel, A. "Sur l'existence d'un cercle passant par un nombre donné de points aux coordonnées entières." *L'Enseignement Math. Ser. 2* **4**, 71–72, 1958.

Sierpiński, W. "Sur quelques problèmes concernant les points aux coordonnées entières." *L'Enseignement Math. Ser. 2* **4**, 25–31, 1958.

Sierpiński, W. "Sur un problème de H. Steinhaus concernant les ensembles de points sur le plan." *Fund. Math.* **46**, 191–194, 1959.

Sierpiński, W. *A Selection of Problems in the Theory of Numbers.* New York: Pergamon Press, 1964.

Schisma

The musical interval by which eight fifths and a major third exceed five octaves,

$$\frac{(\frac{3}{2})^8(\frac{5}{4})}{2^5} = \frac{3^8 \cdot 5}{2^{15}} = \frac{32805}{32768} = 1.00112915\ldots.$$

see also COMMA OF DIDYMUS, COMMA OF PYTHAGORAS, DIESIS

Schläfli Double Six

see DOUBLE SIXES

Schläfli's Formula

For $\Re[z] > 0$,

$$J_\nu(z) = \frac{1}{\pi}\int_0^{\pi/2} \cos(z\sin t - \nu t)\,dt - \frac{\sin(\nu\pi)}{\pi}\int_0^\infty e^{-z\sinh t}e^{-\nu t}\,dt,$$

where $J_\nu(z)$ is a BESSEL FUNCTION OF THE FIRST KIND.

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1472, 1980.

Schläfli Function

The function giving the VOLUME of the spherical quadrectangular TETRAHEDRON:

$$V = \frac{\pi^2}{8} f\left(\frac{\pi}{p}, \frac{\pi}{q}, \frac{\pi}{r}\right),$$

where

$$\frac{\pi^2}{2} f\left(\frac{\pi}{2} - x, y, \frac{\pi}{2} - z\right) = \sum_{m=1}^\infty \left(\frac{D - \sin x \sin z}{D + \sin x \sin z}\right)^m \times \frac{\cos(2mx) - \cos(2my) + \cos(2mz) - 1}{m^2} - x^2 - y^2 - z^2,$$

and

$$D \equiv \sqrt{\cos^2 x \cos^2 z - \cos^2 y}.$$

see also TETRAHEDRON

Schläfli Integral

A definition of a function using a Contour Integral. Schläfli integrals may be converted into Rodrigues Formulas.

see also Rodrigues Formula

Schläfli's Modular Form

The Modular Equation of degree 5 can be written

$$\left(\frac{u}{v}\right)^3 + \left(\frac{v}{u}\right)^3 = 2\left(u^2v^2 - \frac{1}{u^2v^2}\right).$$

see also Modular Equation

Schläfli Polynomial

A polynomial given in terms of the Neumann Polynomials $O_n(x)$ by

$$S_n(x) = \frac{2xO_n(x) - 2\cos^2(\frac{1}{2}n\pi)}{n}.$$

see also Neumann Polynomial

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1477, 1980.

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 196, 1993.

Schläfli Symbol

The symbol $\{p, q\}$ is used to denote a Tessellation of regular p-gons, with q of them surrounding each Vertex. The Schläfli symbol can be used to describe Platonic Solids, and a generalized version describes Quasiregular Polyhedra and Archimedean Solids.

see also Archimedean Solid, Platonic Solid, Quasiregular Polyhedron, Tessellation

Schlegel Graph

A Graph corresponding to Polyhedra skeletons. The Polyhedral Graphs are special cases.

References

Gardner, M. *Wheels, Life, and Other Mathematical Amusements.* New York: W. H. Freeman, p. 158, 1983.

Schlömilch's Function

$$S(\nu, z) \equiv \int_0^\infty (1+t)^{-\nu} e^{-zt}\,dt = z^{\nu-1}e^z \int_z^\infty u^{-\nu}e^{-u}\,du$$
$$= z^{\nu/2-1}e^{z/2}W_{-\nu/2,(1-\nu)/2}(z),$$

where $W_{k,m}(z)$ is the Whittaker Function.

Schlömilch's Series

A Fourier Series-like expansion of a twice continuously differentiable function

$$f(x) = \tfrac{1}{2}a_0 + \sum_{n=1}^\infty a_n J_0(nx)$$

for $0 < x < \pi$, where $J_0(x)$ is a zeroth order Bessel Function of the First Kind and

$$a_0 \equiv 2f(0) + \frac{2}{\pi}\int_0^\pi du \int_0^{\pi/2} f'(u\sin\phi)\,d\phi$$
$$a_n \equiv \frac{2}{\pi}\int_0^\pi du \int_0^{\pi/2} uf'(u\sin\phi)\cos(n\pi)\,d\phi.$$

A special case gives the amazing identity

$$1 = J_0(z) + 2\sum_{n=1}^\infty J_{2n}(z) = [J_0(z)]^2 + 2\sum_{n=1}^\infty [J_n(z)]^2.$$

see also Bessel Function of the First Kind, Bessel Function Fourier Expansion, Fourier Series

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1473, 1980.

Schmitt-Conway Biprism

A Convex Polyhedron which is Space-Filling, but only aperiodically, was found by Conway in 1993.

see also Convex Polyhedron, Space-Filling Polyhedron

Schnirelmann Constant

The constant s_0 in Schnirelmann's Theorem.

see also Schnirelmann's Theorem

Schnirelmann Density

The Schnirelmann density of a sequence of natural numbers is the greatest lower bound of the fractions $A(n)/n$ where $A(n)$ is the number of terms in the sequence $\leq n$.

References

Khinchin, A. Y. "The Landau-Schnirelmann Hypothesis and Mann's Theorem." Ch. 2 in *Three Pearls of Number Theory.* New York: Dover, pp. 18–36, 1998.

Schnirelmann's Theorem

There exists a Positive Integer s such that every sufficiently large Integer is the sum of at most s Primes. It follows that there exists a Positive Integer $s_0 \geq s$ such that every Integer > 1 is a sum of at most s_0 Primes, where s_0 is the Schnirelmann Constant. The best current estimate is $s_0 = 19$.

see also Prime Number, Schnirelmann Density, Waring's Problem

References

Khinchin, A. Y. "The Landau-Schnirelmann Hypothesis and Mann's Theorem." Ch. 2 in *Three Pearls of Number Theory.* New York: Dover, pp. 18–36, 1998.

Schoenemann's Theorem

If the integral COEFFICIENTS C_0, C_1, ..., C_{N-1} of the POLYNOMIAL

$$f(x) = C_0 + C_1 x + C_2 x^2 + \ldots + C_{N-1} x^{N-1} + x^N$$

are divisible by a PRIME NUMBER p, while the free term C_0 is not divisible by p^2, then $f(x)$ is irreducible in the natural rationality domain.

see also ABEL'S IRREDUCIBILITY THEOREM, ABEL'S LEMMA, GAUSS'S POLYNOMIAL THEOREM, KRONECKER'S POLYNOMIAL THEOREM

References

Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, p. 118, 1965.

Scholz Conjecture

Let the minimal length of an ADDITION CHAIN for a number n be denoted $l(n)$. Then the Scholz conjecture states that

$$l(2^n - 1) \leq n - 1 + l(n).$$

The conjecture has been proven for a variety of special cases but not in general.

see also ADDITION CHAIN

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 111, 1994.

Schönflies Symbol

One of the set of symbols C_i, C_s, C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_{2h}, C_{3h}, C_{4h}, C_{5h}, C_{6h}, C_{2v}, C_{3v}, C_{4v}, C_{5v}, C_{6v}, $C_{\infty v}$, D_2, D_3, D_4, D_5, D_6, D_{2h}, D_{3h}, D_{4h}, D_{5h}, D_{6h}, D_{8h}, $D_{\infty h}$, D_{2d}, D_{3d}, D_{4d}, D_{5d}, D_{6d}, I, I_h, O, O_h, S_4, S_6, S_8, T, T_d, and T_h used to identify crystallographic symmetry GROUPS.

Cotton (1990), gives a table showing the translations between Schönflies symbols and HERMANN-MAUGUIN SYMBOLS. Some of the Schönflies symbols denote different sets of symmetry operations but correspond to the same abstract GROUP and so have the same CHARACTER TABLE.

see also CHARACTER TABLE, HERMANN-MAUGUIN SYMBOL, POINT GROUPS, SPACE GROUPS, SYMMETRY OPERATION

References

Cotton, F. A. *Chemical Applications of Group Theory, 3rd ed.* New York: Wiley, p. 379, 1990.

Schönflies Theorem

If J is a simple closed curve in $\mathbb{R}^2$, the closure of one of the components of $\mathbb{R}^2 - J$ is HOMEOMORPHIC with the unit 2-BALL. This theorem may be proved using the RIEMANN MAPPING THEOREM, but the easiest proof is via MORSE THEORY.

The generalization to n-D is called MAZUR'S THEOREM. It follows from the Schönflies theorem that any two KNOTS of $\mathbb{S}^1$ in $\mathbb{S}^2$ or $\mathbb{R}^2$ are equivalent.

see also JORDAN CURVE THEOREM, MAZUR'S THEOREM, RIEMANN MAPPING THEOREM

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 9, 1976.

Thomassen, C. "The Jordan-Schönflies Theorem and the Classification of Surfaces." *Amer. Math. Monthly* **99**, 116–130, 1992.

Schoolgirl Problem

see KIRKMAN'S SCHOOLGIRL PROBLEM

Schoute Coaxal System

The CIRCUMCIRCLE, BROCARD CIRCLE, LEMOINE LINE, and ISODYNAMIC POINTS belong to a COAXAL SYSTEM orthogonal to the the APOLLONIUS CIRCLES, called the Schoute coaxal system. In general, there are 12 points whose PEDAL TRIANGLES with regard to a given TRIANGLE have a given form. They lie six by six on two CIRCLES of the Schoute coaxal system.

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 297–299, 1929.

Schoute's Theorem

In any TRIANGLE, the LOCUS of a point whose PEDAL TRIANGLE has a constant BROCARD ANGLE and is described in a given direction is a CIRCLE of the SCHOUTE COAXAL SYSTEM.

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 297–299, 1929.

Schoute, P. H. *Proc. Amsterdam Acad.*, 39–62, 1887–1888.

Schrage's Algorithm

An algorithm for multiplying two 32-bit integers modulo a 32-bit constant without using any intermediates larger than 32 bits. It is also useful in certain types of RANDOM NUMBER generators.

References

Bratley, P.; Fox, B. L.; and Schrage, E. L. *A Guide to Simulation, 2nd ed.* New York: Springer-Verlag, 1996.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Random Numbers." Ch. 7 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 269, 1992.

Schrage, L. "A More Portable Fortran Random Number Generator." *ACM Trans. Math. Software* **5**, 132–138, 1979.

Schröder-Bernstein Theorem

The Schröder-Bernstein theorem for numbers states that if

$$n \le m \le n,$$

then $m = n$. For SETS, the theorem states that if there are INJECTIONS of the SET A into the SET B and of B into A, then there is a BIJECTIVE correspondence between A and B (i.e., they are EQUIPOLLENT).

see also BIJECTION, EQUIPOLLENT, INJECTION

Schröder's Equation

$$f(\lambda z) = R(z),$$

where $R(z) = \lambda x + a_2 x^2 + \ldots$, $\lambda \equiv R'(0)$, $|\lambda| = 1$, and $\lambda^n \neq 1$ for all $n \in \mathbb{N}$.

Schröder's Method

Two families of equations used to find roots of nonlinear functions of a single variable. The "B" family is more robust and can be used in the neighborhood of degenerate multiple roots while still providing a guaranteed convergence rate. Almost all other root-finding methods can be considered as special cases of Schröder's method. Householder humorously claimed that papers on root-finding could be evaluated quickly by looking for a citation of Schröder's paper; if the reference were missing, the paper probably consisted of a rediscovery of a result due to Schröder (Stewart 1993).

One version of the "A" method is obtained by applying NEWTON'S METHOD to f/f',

$$x_{n+1} = x_n - \frac{f(x_n)f'(x_n)}{[f'(x_n)]^2 - f(x_n)f''(x_n)}$$

(Scavo and Thoo 1995).

see also NEWTON'S METHOD

References

Householder, A. S. *The Numerical Treatment of a Single Nonlinear Equation.* New York: McGraw-Hill, 1970.

Scavo, T. R. and Thoo, J. B. "On the Geometry of Halley's Method." *Amer. Math. Monthly* **102**, 417–426, 1995.

Schröder, E. "Über unendlich viele Algorithmen zur Auflösung der Gleichungen." *Math. Ann.* **2**, 317–365, 1870.

Stewart, G. W. "On Infinitely Many Algorithms for Solving Equations." English translation of Schröder's original paper. College Park, MD: University of Maryland, Institute for Advanced Computer Studies, Department of Computer Science, 1993. `ftp://thales.cs.umd.edu/pub/reports/imase.ps`.

Schröder Number

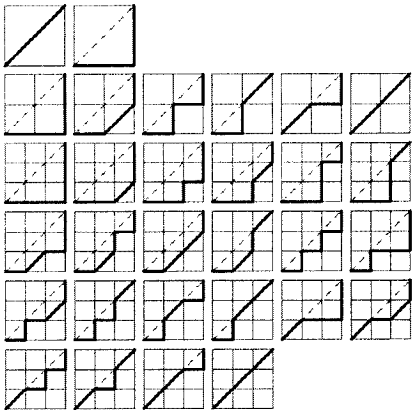

The Schröder number S_n is the number of LATTICE PATHS in the Cartesian plane that start at (0, 0), end at (n, n), contain no points above the line $y = x$, and are composed only of steps (0, 1), (1, 0), and (1, 1), i.e., $\rightarrow$, $\uparrow$, and $\nearrow$. The diagrams illustrating the paths generating S_1, S_2, and S_3 are illustrated above. The numbers S_n are given by the RECURRENCE RELATION

$$S_n = S_{n-1} + \sum_{k=0}^{n-1} S_k S_{n-1-k},$$

where $S_0 = 1$, and the first few are 2, 6, 22, 90, ... (Sloane's A006318). The Schröder Numbers bear the same relation to the DELANNOY NUMBERS as the CATALAN NUMBERS do to the BINOMIAL COEFFICIENTS.

see also BINOMIAL COEFFICIENT, CATALAN NUMBER, DELANNOY NUMBER, LATTICE PATH, MOTZKIN NUMBER, p-GOOD PATH

References

Sloane, N. J. A. Sequence A006318/M1659 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Schroeder Stairs

see PENROSE STAIRWAY

Schröter's Formula

Let a general THETA FUNCTION be defined as

$$T(x,q) \equiv \sum_{n=-\infty}^{\infty} x^n q^{n^2},$$

then

$$T(x,q^a)T(x,q^b) = \sum_{k=0}^{a+b-1} y^k q^{bk^2} T(xyq^{2bk}, q^{a+b}) T(y^q x^{-b} q^{2abk}, q^{ab(1+b)}).$$

see also BLECKSMITH-BRILLHART-GERST THEOREM, JACOBI TRIPLE PRODUCT, RAMANUJAN THETA FUNCTIONS

References

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, p. 111, 1987.

Tannery, J. and Molk, J. *Elements de la Théorie des Fonctions Elliptiques, 4 vols.* Paris: Gauthier-Villars et fils, 1893–1902.

Schur Algebra

An Auslander algebra which connects the representation theories of the symmetric group of PERMUTATIONS and the GENERAL LINEAR GROUP $GL(n, \mathbb{C})$. Schur algebras are "quasihereditary."

References

Martin, S. *Schur Algebras and Representation Theory.* New York: Cambridge University Press, 1993.

Schur Functor

A FUNCTOR which defines an equivalence of module CATEGORIES.

References

Martin, S. *Schur Algebras and Representation Theory.* New York: Cambridge University Press, 1993.

Schur's Inequalities

Let $\mathsf{A} = a_{ij}$ be an $n \times n$ MATRIX with COMPLEX (or REAL) entries and EIGENVALUES $\lambda_1, \lambda_2, \ldots, \lambda_n$, then

$$\sum_{i=1}^{n} |\lambda_i|^2 \leq \sum_{i,j=1}^{n} |a_{ij}|^2$$

$$\sum_{i=1}^{n} |\Re[\lambda_i]|^2 \leq \sum_{i,j=1}^{n} \left| \frac{a_{ij} + a_{ji}^*}{2} \right|^2$$

$$\sum_{i=1}^{n} |\Im[\lambda_i]|^2 \leq \sum_{i,j=1}^{n} \left| \frac{a_{ij} - a_{ji}^*}{2} \right|^2.$$

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1120, 1979.

Schur's Lemma

For each $k \in \mathbb{N}$ there exists a largest INTEGER $s(k)$ (known as the SCHUR NUMBER) such that no matter how the set of INTEGERS less than $\lfloor n!e \rfloor$ (where $\lfloor x \rfloor$ is the FLOOR FUNCTION) is partitioned into k classes, one class must contain INTEGERS x, y, z such that $x + y = z$, where x and y are not necessarily distinct. The upper bound has since been slightly improved to $\lfloor n!(e - 1/24) \rfloor$.

see also COMBINATORICS, SCHUR NUMBER, SCHUR'S THEOREM

References

Guy, R. K. "Schur's Problem. Partitioning Integers into Sum-Free Classes" and "The Modular Version of Schur's Problem." §E11 and E12 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 209–212, 1994.

Schur Matrix

The $p \times p$ SQUARE MATRIX formed by setting $s_{ij} = \xi^{ij}$, where ξ is an pth ROOT OF UNITY. The Schur matrix has a particularly simple DETERMINANT given by

$$\det \mathsf{S} = \epsilon_p p^{p/2},$$

where p is an ODD PRIME and

$$\epsilon_p \equiv \begin{cases} 1 & \text{if } p \equiv 1 \pmod 4 \\ i & \text{if } p \equiv 3 \pmod 4. \end{cases}$$

This determinant has been used to prove the QUADRATIC RECIPROCITY LAW (Landau 1958, Vardi 1991). The ABSOLUTE VALUES of the PERMANENTS of the Schur matrix of order $2p+1$ are given by 1, 3, 5, 105, 81, 6765, ... (Sloane's A003112, Vardi 1991).

Denote the Schur matrix S_p with the first row and first row column omitted by S'_p. Then

$$\operatorname{perm} \mathsf{S}_p = p \operatorname{perm} \mathsf{S}'_p,$$

where perm denoted the PERMANENT (Vardi 1991).

References

Graham, R. L. and Lehmer, D. H. "On the Permanent of Schur's Matrix." *J. Austral. Math. Soc.* **21**, 487–497, 1976.

Landau, E. *Elementary Number Theory.* New York: Chelsea, 1958.

Sloane, N. J. A. Sequence A003112/M2509 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 119–122 and 124, 1991.

Schur Multiplier

A property of FINITE SIMPLE GROUPS which is known for all such GROUPS.

see also FINITE GROUP, SIMPLE GROUP

Schur Number

The Schur numbers are the numbers in the partitioning of a set which are guaranteed to exist by SCHUR'S LEMMA. Schur numbers satisfy the inequality

$$s(k) \geq c(315)^{k/5}$$

for $k > 5$ and some constant c. SCHUR'S THEOREM also shows that

$$s(n) \leq R(n),$$

where $R(n)$ is a RAMSEY NUMBER. The first few Schur numbers are 1, 4, 13, 44, (≥ 157), ... (Sloane's A045652).

see also RAMSEY NUMBER, RAMSEY'S THEOREM, SCHUR'S LEMMA, SCHUR'S THEOREM

References

Frederickson, H. "Schur Numbers and the Ramsey Numbers $N(3, 3, \ldots, 3; 2)$." *J. Combin. Theory Ser. A* **27**, 376–377, 1979.

Guy, R. K. "Schur's Problem. Partitioning Integers into Sum-Free Classes" and "The Modular Version of Schur's Problem." §E11 and E12 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 209–212, 1994.

Sloane, N. J. A. Sequence A045652 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Schur's Problem

see SCHUR'S LEMMA

Schur's Representation Lemma

If π on V and π' on V' are irreducible representations and $E : V \mapsto V'$ is a linear map such that $\pi'(g)E = E\pi(g)$ for all $g \in$ and group G, then $E = 0$ or E is invertible. Furthermore, if $V = V'$, then E is a SCALAR.

References

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Schur's Theorem

As shown by Schur in 1916, the SCHUR NUMBER $s(n)$ satisfies

$$s(n) \leq R(n)$$

for $n = 1, 2, \ldots$, where $R(n)$ is a RAMSEY NUMBER.

see also RAMSEY NUMBER, SCHUR'S LEMMA, SCHUR NUMBER

Schwarz's Inequality

$$|\langle\psi_1|\psi_2\rangle|^2 \leq \langle\psi_1|\psi_1\rangle\langle\psi_2|\psi_2\rangle. \tag{1}$$

Written out explicitly

$$\left[\int_a^b \psi_1(x)\psi_2(x)\,dx\right]^2 \leq \int_a^b [\psi_1(x)]^2\,dx \int_a^b [\psi_2(x)]^2\,dx, \tag{2}$$

with equality IFF $g(x) = \alpha f(x)$ with α a constant. To derive, let $\psi(x)$ be a COMPLEX function and λ a COMPLEX constant such that $\psi(x) \equiv f(x) + \lambda g(x)$ for some f and g. Then

$$\int \psi^*\psi\,dx = \int f^* f\,dx + \lambda \int f^* g\,dx + \lambda^* \int g^* f\,dx + \lambda\lambda^* \int g^* g\,dx \geq 0, \tag{3}$$

with equality when $\psi(x) = 0$. Now, note that λ and λ^* are LINEARLY INDEPENDENT (they are ORTHOGONAL), so differentiate with respect to one of them (say λ^*) and set to zero to minimize $\int \psi^*\psi\,dx$.

$$\frac{\partial}{\partial}\int \psi^*\psi\,dx = \int g^* f\,dx + \lambda \int g^* g\,dx = 0 \tag{4}$$

$$\lambda = -\frac{\int g^* f\,dx}{\int g^* g\,dx}, \tag{5}$$

which means that

$$\lambda^* = -\frac{\int f^* g\,dx}{\int g^* g\,dx}. \tag{6}$$

Plugging back in,

$$\int \psi^*\psi\,dx = \int f^* f\,dx - \frac{\int g^* f\,dx}{\int g^* g\,dx}\int f^* g\,dx - \frac{\int f^* g\,dx}{\int g^* g\,dx}\int g^* f\,dx + \frac{\int g^* f\,dx \int f^* g\,dx}{(\int g^* g\,dx)^2}\int g^* g\,dx \geq 0. \tag{7}$$

Multiplying through by $\int g^* g\,dx$ gives

$$\int f^* f\,dx \int g^* g\,dx - \int g^* f\,dx \int f^* g\,dx - \int f^* g\,dx \int g^* f\,dx + \int g^* f\,dx \int f^* g\,dx \geq 0 \tag{8}$$

$$\int g^* f\,dx \int f^* g\,dx \leq \int f^* f\,dx \int g^* g\,dx \tag{9}$$

$$\left|\int g^* f\,dx\right| = \left|\int f^* g\,dx\right| \leq \int f^* f\,dx \int g^* g\,dx \tag{10}$$

or

$$|\langle f|g\rangle|^2 \leq \langle f|f\rangle\langle g|g\rangle. \tag{11}$$

BESSEL'S INEQUALITY can be derived from this.

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 11, 1972.

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 527–529, 1985.

Schwarz-Pick Lemma

If f is an analytic map of the DISK $\mathbb{D}$ into $\mathbb{D}$ and f preserves the hyperbolic distance between any two points, then f is a disk map and preserves all distance.

References

Busemann, H. *The Geometry of Geodesics.* New York: Academic Press, p. 41, 1955.

Schwarz Reflection Principle

Let

$$g(z) \equiv \sum_{n=0}^{\infty} (z - z_0)^n \frac{f^{(n)}(z_0)}{n!}, \qquad (1)$$

then

$$\begin{aligned} g^*(z) &= \left[\sum_{n=0}^{\infty} (z - z_0)^n \frac{f^{(n)}(z_0)}{n!} \right]^* \\ &= \sum_{n=0}^{\infty} (z^* - {z_0}^*) n \frac{f^{(n)}({z_0}^*)}{n!}. \qquad (2) \end{aligned}$$

If z_0 is pure real, then $z_0 = {z_0}^*$, so

$$g^*(z) = \sum_{n=0}^{\infty} (z^* - z_0)^n \frac{f^{(n)}(z_0)}{n!} = g(z^*). \qquad (3)$$

Therefore, if a function $f(z)$ is ANALYTIC over some region including the REAL LINE and $f(z)$ is REAL when z is real, then $f^*(z) = f(z^*)$.

Schwarz Triangle

The Schwarz triangles are SPHERICAL TRIANGLES which, by repeated reflection in their indices, lead to a set of congruent SPHERICAL TRIANGLES covering the SPHERE a finite number of times.

Schwarz triangles are specified by triples of numbers (p, q, r). There are four "families" of Schwarz triangles, and the largest triangles from each of these families are

$$(2\ 2\ n'), (\tfrac{3}{2}\ \tfrac{3}{2}\ \tfrac{3}{2}), (\tfrac{3}{2}\ \tfrac{4}{3}\ \tfrac{4}{3}), (\tfrac{5}{4}\ \tfrac{5}{4}\ \tfrac{5}{4}).$$

The others can be derived from

$$(p\ q\ r) = (p\ x\ r_1) + (x\ q\ r_2),$$

where

$$\frac{1}{r_1} + \frac{1}{r_2} = \frac{1}{r}$$

and

$$\begin{aligned} \cos\left(\frac{\pi}{x}\right) &= -\cos\left(\frac{\pi}{x'}\right) \\ &= \frac{\cos\left(\frac{\pi}{q}\right)\sin\left(\frac{\pi}{r_1}\right) - \cos\left(\frac{\pi}{p}\right)\sin\left(\frac{\pi}{r_2}\right)}{\sin\left(\frac{\pi}{r}\right)}. \end{aligned}$$

see also COLUNAR TRIANGLE, SPHERICAL TRIANGLE

References

Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, pp. 112–113 and 296, 1973.

Schwarz, H. A. "Zur Theorie der hypergeometrischen Reihe." *J. reine angew. Math.* **75**, 292–335, 1873.

Schwarz's Triangle Problem

see FAGNANO'S PROBLEM

Schwarzian Derivative

The Schwarzian derivative is defined by

$$D_{\text{Schwarzian}} \equiv \frac{f'''(x)}{f'(x)} - \frac{3}{2}\left[\frac{f''(x)}{f'(x)}\right]^2.$$

The FEIGENBAUM CONSTANT is universal for 1-D MAPS if its Schwarzian derivative is NEGATIVE in the bounded interval (Tabor 1989, p. 220).

see also FEIGENBAUM CONSTANT

References

Tabor, M. *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, 1989.

Schwenk's Formula

Let $R+B$ be the number of MONOCHROMATIC FORCED TRIANGLES (where R and B are the number of red and blue TRIANGLES) in an EXTREMAL GRAPH. Then

$$R + B = \binom{n}{3} - \left\lfloor \tfrac{1}{2} n \left\lfloor \tfrac{1}{4}(n-1)^2 \right\rfloor \right\rfloor,$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT and $\lfloor x \rfloor$ is the FLOOR FUNCTION (Schwenk 1972).

see also EXTREMAL GRAPH, MONOCHROMATIC FORCED TRIANGLE

References

Schwenk, A. J. "Acquaintance Party Problem." *Amer. Math. Monthly* **79**, 1113–1117, 1972.

Scientific Notation

Scientific notation is the expression of a number n in the form $a \times 10^p$, where

$$p \equiv \lfloor \log_{10} |n| \rfloor$$

is the FLOOR of the base-10 LOGARITHM of n (the "order of magnitude"), and

$$a \equiv \frac{n}{10^p}$$

is a REAL NUMBER satisfying $1 \leq |a| < 10$. For example, in scientific notation, the number $n = 101,325$ has order of magnitude

$$p = \lfloor \log_{10} 101,325 \rfloor = \lfloor 5.00572 \rfloor = 5,$$

so n would be written 1.01325×10^5. The special case of 0 does not have a unique representation in scientific notation, i.e., $0 = 0 \times 10^0 = 0 \times 10^1 = \ldots$.

see also CHARACTERISTIC (REAL NUMBER), FIGURES, MANTISSA, SIGNIFICANT FIGURES

Score Sequence

The score sequence of a TOURNAMENT is a monotonic nondecreasing sequence of the OUTDEGREES of the VERTICES. The score sequences for $n = 1, 2, \ldots$ are 1, 1, 2,4, 9, 22, 59, 167, ... (Sloane's A000571).

see also TOURNAMENT

References

Ruskey, F. "Information on Score Sequences." `http://sue.csc.uvic.ca/~cos/inf/nump/ScoreSequence.html`.

Ruskey, F.; Cohen, R.; Eades, P.; and Scott, A. "Alley CATs in Search of Good Homes." *Congres. Numer.* **102**, 97–110, 1994.

Sloane, N. J. A. Sequence A000571/M1189 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Screw

A TRANSLATION along a straight line L and a ROTATION about L such that the angle of ROTATION is proportional to the TRANSLATION at each instant. Also known as a TWIST.

see also DINI'S SURFACE, HELICOID, ROTATION, SCREW THEOREM, SEASHELL, TRANSLATION

Screw Theorem

Any motion of a rigid body in space at every instant is a SCREW motion. This theorem was proved by Mozzi and Cauchy.

see also SCREW

Scruple

An archaic UNIT FRACTION variously defined as 1/200 (of an hour), 1/10 or 1/12 (of an inch), 1/12 (of a celestial body's angular diameter), or 1/60 (of an hour or DEGREE).

see also CALCUS, UNCIA

Sea Horse Valley

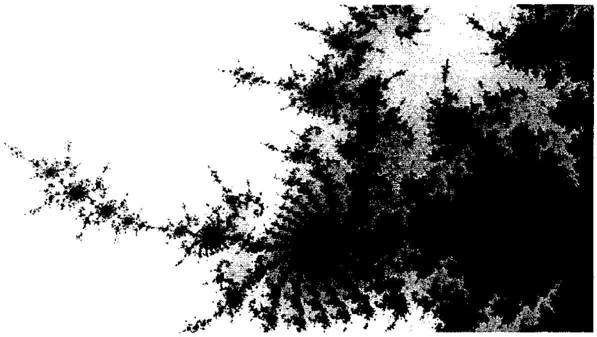

A portion of the MANDELBROT SET centered around $-1.25 + 0.047i$ with width approximately $0.009 + 0.005i$.

see also MANDELBROT SET

Searching

Searching refers to locating a given element or an element satisfying certain conditions from some (usually ordered or partially ordered) table, list, TREE, etc.

see also SORTING, TABU SEARCH, TREE SEARCHING

References

Knuth, D. E. *The Art of Computer Programming, 2nd ed, Vol. 3: Sorting and Searching.* Reading, MA: Addison-Wesley, 1973.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "How to Search an Ordered Table." §3.4 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 110–113, 1992.

Search Tree

see TREE SEARCHING

Seashell

see CONICAL SPIRAL

Secant

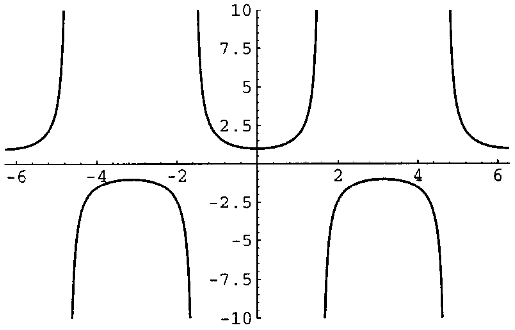

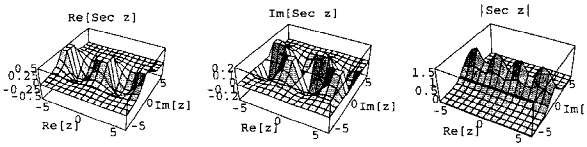

The function defined by $\sec x \equiv 1/\cos x$, where $\cos x$ is the COSINE. The MACLAURIN SERIES of the secant is

$$\begin{aligned}\sec x &= \frac{(-1)^n E_{2n}}{(2n)!} x^{2n} \\ &= 1 + \tfrac{1}{2}x^2 + \tfrac{5}{24}x^4 + \tfrac{61}{720}x^6 + \tfrac{277}{8064}x^8 + \ldots,\end{aligned}$$

where E_{2n} is an EULER NUMBER.

see also ALTERNATING PERMUTATION, COSECANT, COSINE, EULER NUMBER, EXSECANT, INVERSE SECANT

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Circular Functions." §4.3 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 71–79, 1972.

Spanier, J. and Oldham, K. B. "The Secant sec(x) and Cosecant csc(x) Functions." Ch. 33 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 311–318, 1987.

Secant Line

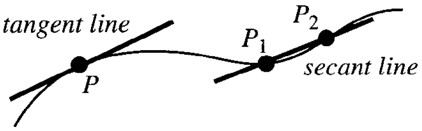

A line joining two points of a curve. In abstract mathematics, the points which a secant line connects can be either REAL or COMPLEX CONJUGATE IMAGINARY.

see also BITANGENT, TANGENT LINE, TRANSVERSAL LINE

Secant Method

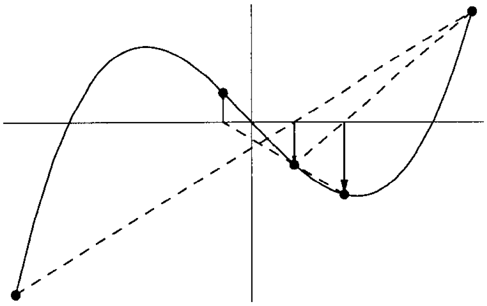

A ROOT-finding algorithm which assumes a function to be approximately linear in the region of interest. Each improvement is taken as the point where the approximating line crosses the axis. The secant method retains only the most recent estimate, so the root does not necessarily remain bracketed. When the ALGORITHM does converge, its order of convergence is

$$\lim_{k\to\infty} |\epsilon_{k+1}| \approx C|\epsilon|^{\phi}, \tag{1}$$

where C is a constant and ϕ is the GOLDEN MEAN.

$$f'(x_{n-1}) \approx \frac{f(x_{n-1}) - f(x_{n-2})}{x_{n-1} - x_{n-2}} \tag{2}$$

$$f(x_n) \approx f(x_{n-1}) + f'(x_n)(x_n - x_{n-1}) = 0 \tag{3}$$

$$f(x_{n-1}) + \frac{f(x_{n-1}) - f(x_{n-2})}{x_{n-1} - x_{n-2}}(x_n - x_{n-1}) = 0, \tag{4}$$

so

$$x_n = x_{n-1} - \frac{f(x_{n-1})(x_{n-1} - x_{n-2})}{f(x_{n-1}) - f(x_{n-2})}. \tag{5}$$

see also FALSE POSITION METHOD

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Secant Method, False Position Method, and Ridders' Method." §9.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 347–352, 1992.

Secant Number

A number, more commonly called an EULER NUMBER, giving the number of ODD ALTERNATING PERMUTATIONS. The term ZAG NUMBER is sometimes also used.

see also ALTERNATING PERMUTATION, EULER NUMBER, EULER ZIGZAG NUMBER, TANGENT NUMBER

Sech

see HYPERBOLIC SECANT

Second

see ARC SECOND

Second Curvature

see TORSION (DIFFERENTIAL GEOMETRY)

Second Derivative Test

Suppose $f(x)$ is a FUNCTION of x which is twice DIFFERENTIABLE at a STATIONARY POINT x_0.

1. If $f''(x_0) > 0$, then f has a RELATIVE MINIMUM at x_0.
2. If $f''(x_0) < 0$, then f has a RELATIVE MAXIMUM at x_0.

The EXTREMUM TEST gives slightly more general conditions under which functions with $f''(x_0) = 0$.

If $f(x, y)$ is a 2-D FUNCTION which has a RELATIVE EXTREMUM at a point (x_0, y_0) and has CONTINUOUS PARTIAL DERIVATIVES at this point, then $f_x(x_0, y_0) = 0$ and $f_y(x_0, y_0) = 0$. The second PARTIAL DERIVATIVES test classifies the point as a MAXIMUM or MINIMUM. Define the DISCRIMINANT as

$$D \equiv f_{xx}f_{yy} - f_{xy}f_{yx} = f_{xx}f_{yy} - {f_{xy}}^2.$$

1. If $D > 0$, $f_{xx}(x_0, y_0) > 0$ and $f_{xx}(x_0, y_0) + f_{yy}(x_0, y_0) > 0$, the point is a RELATIVE MINIMUM.
2. If $D > 0$, $f_{xx}(x_0, y_0) < 0$, and $f_{xx}(x_0, y_0) + f_{yy}(x_0, y_0) < 0$, the point is a RELATIVE MAXIMUM.
3. If $D < 0$, the point is a SADDLE POINT.
4. If $D = 0$, higher order tests must be used.

see also DISCRIMINANT (SECOND DERIVATIVE TEST), EXTREMUM, EXTREMUM TEST, FIRST DERIVATIVE TEST, GLOBAL MAXIMUM, GLOBAL MINIMUM, HESSIAN DETERMINANT, MAXIMUM, MINIMUM, RELATIVE MAXIMUM, RELATIVE MINIMUM, SADDLE POINT (FUNCTION)

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 14, 1972.

Second Fundamental Tensor

see WEINGARTEN MAP

Section (Graph)

A section of a GRAPH is obtained by finding its intersection with a PLANE.

Section (Pencil)

The lines of a PENCIL joining the points of a RANGE to another POINT.

see also PENCIL, RANGE (LINE SEGMENT)

Section (Tangent Bundle)

A VECTOR FIELD is a section of its TANGENT BUNDLE, meaning that to every point x in a MANIFOLD M, a VECTOR $X(x) \in T_xM$ is associated, where T_x is the TANGENT SPACE.

see also TANGENT BUNDLE, TANGENT SPACE

Sectional Curvature

The mathematical object κ which controls the rate of geodesic deviation.

see also BISHOP'S INEQUALITY, CHEEGER'S FINITENESS THEOREM, GEODESIC

Sector

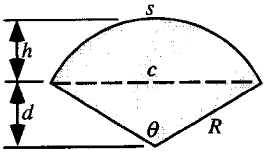

A WEDGE obtained by taking a portion of a CIRCLE with CENTRAL ANGLE $\theta < \pi$ radians (180°), illustrated above as the shaded region. A sector of π radians would be a SEMICIRCLE. Let R be the radius of the CIRCLE, c the CHORD length, s the ARC LENGTH, h the height of the arced portion, and d the height of the triangular portion. Then

$$\begin{aligned} R &= h + d && (1)\\ s &= R\theta && (2)\\ d &= R\cos(\tfrac{1}{2}\theta) && (3)\\ &= \tfrac{1}{2}c\cot(\tfrac{1}{2}\theta) && (4)\\ &= \tfrac{1}{2}\sqrt{4R^2 - c^2} && (5)\\ c &= 2R\sin(\tfrac{1}{2}\theta) && (6)\\ &= 2d\tan(\tfrac{1}{2}\theta) && (7)\\ &= 2\sqrt{R^2 - d^2} && (8)\\ &= 2\sqrt{h(2R - h)}. && (9) \end{aligned}$$

The ANGLE θ obeys the relationships

$$\begin{aligned} \theta &= \frac{s}{R} = 2\cos^{-1}\left(\frac{d}{R}\right) = 2\tan^{-1}\left(\frac{c}{2d}\right)\\ &= 2\sin^{-1}\left(\frac{c}{2R}\right). \qquad (10) \end{aligned}$$

The AREA of the sector is

$$A = \tfrac{1}{2}Rs = \tfrac{1}{2}R^2\theta \qquad (11)$$

(Beyer 1987).

see also CIRCLE-CIRCLE INTERSECTION, LENS, OBTUSE TRIANGLE, SEGMENT

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 125, 1987.

Sectorial Harmonic

A SPHERICAL HARMONIC of the form

$$\sin(m\theta)P_m^m(\cos\phi)$$

or

$$\cos(m\theta)P_m^m(\cos\phi).$$

see also SPHERICAL HARMONIC

Secular Equation

see CHARACTERISTIC EQUATION

Seed

The initial number used as the starting point in a RANDOM NUMBER generating ALGORITHM.

Seed of Life

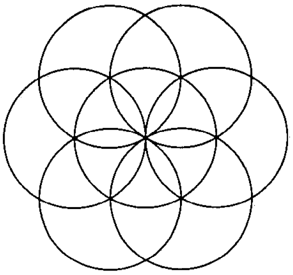

One of the beautiful arrangements of CIRCLES found at the Temple of Osiris at Abydos, Egypt (Rawles 1997). The CIRCLES are placed with 6-fold symmetry, forming a mesmerizing pattern of CIRCLES and LENSES.

see also CIRCLE, FIVE DISKS PROBLEM, FLOWER OF LIFE, VENN DIAGRAM

References

Rawles, B. *Sacred Geometry Design Sourcebook: Universal Dimensional Patterns.* Nevada City, CA: Elysian Pub., p. 15, 1997.

Weisstein, E. W. "Flower of Life." `http://www.astro.virginia.edu/~eww6n/math/notebooks/FlowerOfLife.m`.

Seek Time

see POINT-POINT DISTANCE—1-D

Segment

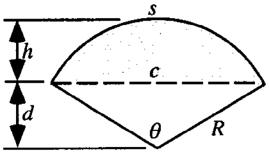

A portion of a CIRCLE whose upper boundary is a circular ARC and whose lower boundary is a CHORD making a CENTRAL ANGLE $\theta < \pi$ radians (180°), illustrated above as the shaded region. Let R be the radius of the CIRCLE, c the CHORD length, s the ARC LENGTH, h the height of the arced portion, and d the height of the triangular portion. Then

$$\begin{aligned} R &= h + d \tag{1}\\ s &= R\theta \tag{2}\\ d &= R\cos(\tfrac{1}{2}\theta) \tag{3}\\ &= \tfrac{1}{2}c\cot(\tfrac{1}{2}\theta) \tag{4}\\ &= \tfrac{1}{2}\sqrt{4R^2 - c^2} \tag{5}\\ c &= 2R\sin(\tfrac{1}{2}\theta) \tag{6}\\ &= 2d\tan(\tfrac{1}{2}\theta) \tag{7}\\ &= 2\sqrt{R^2 - d^2} \tag{8}\\ &= 2\sqrt{h(2R - h)}. \tag{9} \end{aligned}$$

The ANGLE θ obeys the relationships

$$\begin{aligned} \theta &= \frac{s}{R} = 2\cos^{-1}\left(\frac{d}{R}\right) = 2\tan^{-1}\left(\frac{c}{2d}\right)\\ &= 2\sin^{-1}\left(\frac{c}{2R}\right). \tag{10} \end{aligned}$$

The AREA of the segment is then

$$\begin{aligned} A &= A_{\text{sector}} - A_{\text{isosceles triangle}} \tag{11}\\ &= \tfrac{1}{2}R^2(\theta - \sin\theta) \tag{12}\\ &= \tfrac{1}{2}(Rs - cd) \tag{13}\\ &= R^2\cos^{-1}\left(\frac{d}{R}\right) - d\sqrt{R^2 - d^2} \tag{14}\\ &= R^2\cos^{-1}\left(\frac{R-h}{R}\right) - (R-h)\sqrt{2Rh - h^2}, \tag{15} \end{aligned}$$

where the formula for the ISOSCELES TRIANGLE in terms of the VERTEX angle has been used (Beyer 1987).

see also CHORD, CIRCLE-CIRCLE INTERSECTION, CYLINDRICAL SEGMENT, LENS, PARABOLIC SEGMENT, SAGITTA, SECTOR, SPHERICAL SEGMENT

References
Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 125, 1987.

Segmented Number

see PRIME NUMBER OF MEASUREMENT

Segner's Recurrence Formula

The recurrence FORMULA

$$E_n = E_2E_{n-1} + E_3E_{n-2} + \ldots + E_{n-1}E_2$$

which gives the solution to EULER'S POLYGON DIVISION PROBLEM.

see also CATALAN NUMBER, EULER'S POLYGON DIVISION PROBLEM

Segre's Theorem

For any REAL NUMBER $r \geq 0$, an IRRATIONAL number α can be approximated by infinitely many RATIONAL fractions p/q in such a way that

$$-\frac{1}{\sqrt{1+4r}\,q^2} < \frac{p}{q} - \alpha < \frac{r}{\sqrt{1+4r}\,q^2}.$$

If $r = 1$, this becomes HURWITZ'S IRRATIONAL NUMBER THEOREM.

see also HURWITZ'S IRRATIONAL NUMBER THEOREM

Seiberg-Witten Equations

$$\begin{aligned} D_A\psi &= 0\\ F_A^+ &= -\tau(\psi, \psi), \end{aligned}$$

where τ is the sesquilinear map $\tau : W^+ \times W^+ \to \Lambda^+ \otimes \mathbb{C}$.

see also WITTEN'S EQUATIONS

References
Donaldson, S. K. "The Seiberg-Witten Equations and 4-Manifold Topology." *Bull. Amer. Math. Soc.* **33**, 45–70, 1996.
Morgan, J. W. *The Seiberg-Witten Equations and Applications to the Topology of Smooth Four-Manifolds.* Princeton, NJ: Princeton University Press, 1996.

Seiberg-Witten Invariants

see WITTEN'S EQUATIONS

Seidel-Entringer-Arnold Triangle

The NUMBER TRIANGLE consisting of the ENTRINGER NUMBERS $E_{n,k}$ arranged in "ox-plowing" order,

$$\begin{gathered} E_{00}\\ E_{10} \to E_{11}\\ E_{22} \leftarrow E_{21} \leftarrow E_{20}\\ E_{30} \to E_{31} \to E_{32} \to E_{33}\\ E_{44} \leftarrow E_{43} \leftarrow E_{42} \leftarrow E_{41} \leftarrow E_{40} \end{gathered}$$

giving

$$\begin{gathered} 1\\ 0 \to 1\\ 1 \leftarrow 1 \leftarrow 0\\ 0 \to 1 \to 2 \to 2\\ 5 \leftarrow 5 \leftarrow 4 \leftarrow 2 \leftarrow 0 \end{gathered}$$

see also BELL NUMBER, BOUSTROPHEDON TRANSFORM, CLARK'S TRIANGLE, ENTRINGER NUMBER, EULER'S TRIANGLE, LEIBNIZ HARMONIC TRIANGLE, NUMBER TRIANGLE, PASCAL'S TRIANGLE

References
Arnold, V. I. "Bernoulli-Euler Updown Numbers Associated with Function Singularities, Their Combinatorics, and Arithmetics." *Duke Math. J.* **63**, 537–555, 1991.
Arnold, V. I. "Snake Calculus and Combinatorics of Bernoulli, Euler, and Springer Numbers for Coxeter Groups." *Russian Math. Surveys* **47**, 3–45, 1992.
Conway, J. H. and Guy, R. K. In *The Book of Numbers.* New York: Springer-Verlag, 1996.
Dumont, D. "Further Triangles of Seidel-Arnold Type and Continued Fractions Related to Euler and Springer Numbers." *Adv. Appl. Math.* **16**, 275–296, 1995.
Entringer, R. C. "A Combinatorial Interpretation of the Euler and Bernoulli Numbers." *Nieuw. Arch. Wisk.* **14**, 241–246, 1966.
Millar, J.; Sloane, N. J. A.; and Young, N. E. "A New Operation on Sequences: The Boustrophedon Transform." *J. Combin. Th. Ser. A* **76**, 44–54, 1996.
Seidel, I. "Über eine einfache Entstehungsweise der Bernoullischen Zahlen und einiger verwandten Reihen." *Sitzungsber. Münch. Akad.* **4**, 157–187, 1877.

Seifert Circle

Eliminate each knot crossing by connecting each of the strands coming into the crossing to the adjacent strand leaving the crossing. The resulting strands no longer cross but form instead a set of nonintersecting CIRCLES called Seifert circles.

References
Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, p. 96, 1994.

Seifert Conjecture

Every smooth NONZERO VECTOR FIELD on the 3-SPHERE has at least one closed orbit. The conjecture was proposed in 1950, proved true for Hopf fibrations, but proved false in general by Kuperberg (1994).

References
Kuperberg, G. "A Volume-Preserving Counterexample to the Seifert Conjecture." *Comment. Math. Helv.* **71**, 70–97, 1996.
Kuperberg, G. and Kuperberg, K. "Generalized counterexamples to the Seifert Conjecture." *Ann. Math.* **143**, 547–576, 1996.
Kuperberg, G. and Kuperberg, K. "Generalized Counterexamples to the Seifert Conjecture." *Ann. Math.* **144**, 239–268, 1996.
Kuperberg, K. "A Smooth Counterexample to the Seifert Conjecture." *Ann. Math.* **140**, 723–732, 1994.

Seifert Form

For K a given KNOT in $\mathbb{S}^3$, choose a SEIFERT SURFACE M^2 in $\mathbb{S}^3$ for K and a bicollar $\hat{M} \times [-1,1]$ in $\mathbb{S}^3 - K$. If $x \in H_1(\hat{M})$ is represented by a 1-cycle in $\hat{M}$, let x^+ denote the homology cycle carried by $x \times 1$ in the bicollar. Similarly, let x^- denote $x \times -1$. The function $f : H_1(\hat{M}) \times H_1(\hat{M}) \to Z$ defined by

$$f(x,y) = \operatorname{lk}(x, y^+),$$

where lk denotes the LINKING NUMBER, is called a Seifert form for K.

see also SEIFERT MATRIX

References
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 200–201, 1976.

Seifert Matrix

Given a SEIFERT FORM $f(x,y)$, choose a basis $e_1, \ldots, e_{2g}$ for $H_1(\hat{M})$ as a Z-module so every element is uniquely expressible as

$$n_1 e_1 + \ldots + n_{2g} e_{2g}$$

with n_i integer, define the Seifert matrix V as the $2g \times 2g$ integral MATRIX with entries

$$v_{ij} = \operatorname{lk}(e_i, e_j^+).$$

The right-hand TREFOIL KNOT has Seifert matrix

$$V = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}.$$

A Seifert matrix is not a knot invariant, but it can be used to distinguish between different SEIFERT SURFACES for a given knot.

see also ALEXANDER MATRIX

References
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 200–203, 1976.

Seifert's Spherical Spiral

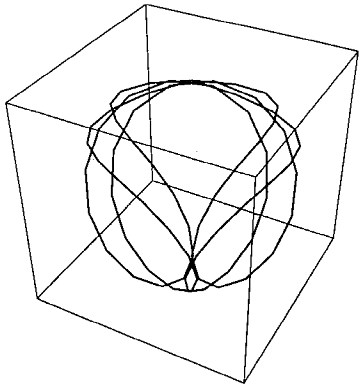

Is given by the CYLINDRICAL COORDINATES parametric equation

$$\begin{aligned} r &= \operatorname{sn}(s) \\ \theta &= ks \\ z &= \operatorname{cn}(s), \end{aligned}$$

where k is a POSITIVE constant and $\operatorname{sn}(s)$ and $\operatorname{cn}(s)$ are JACOBI ELLIPTIC FUNCTIONS (Whittaker and Watson 1990, pp. 527–528).

References

Bowman, F. *Introduction to Elliptic Functions, with Applications.* New York: Dover, p. 34, 1961.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Seifert Surface

An orientable surface with one boundary component such that the boundary component of the surface is a given KNOT K. In 1934, Seifert proved that such a surface can be constructed for any KNOT. The process of generating this surface is known as Seifert's algorithm. Applying Seifert's algorithm to an alternating projection of an alternating knot yields a Seifert surface of minimal GENUS.

There are KNOTS for which the minimal genus Seifert surface cannot be obtained by applying Seifert's algorithm to any projection of that KNOT, as proved by Morton in 1986 (Adams 1994, p. 105).

see also GENUS (KNOT), SEIFERT MATRIX

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 95–106, 1994.

Seifert, H. "Über das Geschlecht von Knotten." *Math. Ann.* **110**, 571–592, 1934.

Self-Adjoint Matrix

A MATRIX A for which

$$\mathsf{A}^\dagger \equiv (\mathsf{A}^{\mathrm{T}})^* = \mathsf{A},$$

where the ADJOINT OPERATOR is denoted $\mathsf{A}^\dagger$, A^{T} is the MATRIX TRANSPOSE, and $*$ is the COMPLEX CONJUGATE. If a MATRIX is self-adjoint, it is said to be HERMITIAN.

see also ADJOINT OPERATOR, HERMITIAN MATRIX, MATRIX TRANSPOSE

Self-Adjoint Operator

Given a differential equation

$$\tilde{\mathcal{L}}u(x) \equiv p_0 \frac{du^2}{dx^2} + p_1 \frac{du}{dx} + p_2 u, \tag{1}$$

where $p_i \equiv p_i(x)$ and $u \equiv u(x)$, the ADJOINT OPERATOR $\tilde{\mathcal{L}}^\dagger$ is defined by

$$\begin{aligned}\tilde{\mathcal{L}}^\dagger u &\equiv \frac{d}{dx^2}(p_0 u) - \frac{d}{dx}(p_1 u) + p_2 u & (2)\\ &= p_0 \frac{d^2u}{dx^2} + (2p_0{}' - p_1)\frac{du}{dx} + (p_0{}'' - p_1{}' + p_2)u. & (3)\end{aligned}$$

In order for the operator to be self-adjoint, i.e.,

$$\tilde{\mathcal{L}} = \tilde{\mathcal{L}}^\dagger, \tag{4}$$

the second terms in (1) and (3) must be equal, so

$$p_0{}'(x) = p_1(x). \tag{5}$$

This also guarantees that the third terms are equal, since

$$p_0{}'(x) = p_1(x) \Rightarrow p_0{}''(x) = p_1{}'(x), \tag{6}$$

so (3) becomes

$$\begin{aligned}\tilde{\mathcal{L}}u = \tilde{\mathcal{L}}^\dagger u &= p_0 \frac{d^2}{dx^2} + p_0{}' \frac{du}{dx} + p_2 u & (7)\\ &= \frac{d}{dx}\left(p_0 \frac{du}{dx}\right) + p_2 u = 0. & (8)\end{aligned}$$

The LEGENDRE DIFFERENTIAL EQUATION and the equation of SIMPLE HARMONIC MOTION are self-adjoint, but the LAGUERRE DIFFERENTIAL EQUATION and HERMITE DIFFERENTIAL EQUATION are not.

A nonself-adjoint second-order linear differential operator can always be transformed into a self-adjoint one using STURM-LIOUVILLE THEORY. In the special case $p_2(x) = 0$, (8) gives

$$\frac{d}{dx}\left[p_0(x)\frac{du}{dx}\right] = 0 \tag{9}$$

$$p_0(x)\frac{du}{dx} = C \tag{10}$$

$$du = C\frac{dx}{p_0(x)} \tag{11}$$

$$u = C\int \frac{dx}{p_0(x)}, \tag{12}$$

where C is a constant of integration.

A self-adjoint operator which satisfies the BOUNDARY CONDITIONS

$$v^* p U'|_{x=a} = v^* p U'|_{x=b} \tag{13}$$

is automatically a HERMITIAN OPERATOR.

see also ADJOINT OPERATOR, HERMITIAN OPERATOR, STURM-LIOUVILLE THEORY

References

Arfken, G. "Self-Adjoint Differential Equations." §9.1 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 497–509, 1985.

Self-Avoiding Walk

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let the number of RANDOM WALKS on a d-D lattice starting at the ORIGIN which never land on the same lattice point twice in n steps be denoted $c(n)$. The first few values are

$$c_d(0) = 1 \tag{1}$$

$$c_d(1) = 2d \tag{2}$$

$$c_d(2) = 2d(2d-1). \tag{3}$$

The connective constant

$$\mu_d \equiv \lim_{n\to\infty} [c_d(n)]^{1/n} \tag{4}$$

is known to exist and be FINITE. The best ranges for these constants are

$$\mu_2 \in [2.62002, 2.6939] \tag{5}$$

$$\mu_3 \in [4.572140, 4.7476] \tag{6}$$

$$\mu_4 \in [6.742945, 6.8179] \tag{7}$$

$$\mu_5 \in [8.828529, 8.8602] \tag{8}$$

$$\mu_6 \in [10.874038, 10.8886] \tag{9}$$

(Finch).

For the triangular lattice in the plane, $\mu < 4.278$ (Alm 1993), and for the hexagonal planar lattice, it is conjectured that

$$\mu = \sqrt{2+\sqrt{2}} \tag{10}$$

(Madras and Slade 1993).

The following limits are also believed to exist and to be FINITE:

$$\begin{cases} \lim_{n\to\infty} \frac{c(n)}{\mu^n n^{\gamma-1}} & \text{for } d \neq 4 \\ \lim_{n\to\infty} \frac{c(n)}{\mu^n n^{\gamma-1}(\ln n)^{1/4}} & \text{for } d = 4, \end{cases} \tag{11}$$

where the critical exponent $\gamma = 1$ for $d > 4$ (Madras and Slade 1993) and it has been conjectured that

$$\gamma = \begin{cases} \frac{43}{32} & \text{for } d = 2 \\ 1.162\ldots & \text{for } d = 3 \\ 1 & \text{for } d = 4. \end{cases} \tag{12}$$

Define the mean square displacement over all n-step self-avoiding walks ω as

$$s(n) \equiv \left\langle |\omega(n)|^2 \right\rangle = \frac{1}{c(n)} \sum_{\omega} |\omega(n)|^2. \tag{13}$$

The following limits are believed to exist and be FINITE:

$$\begin{cases} \lim_{n\to\infty} \frac{s(n)}{n^{2\nu}} & \text{for } d \neq 4 \\ \lim_{n\to\infty} \frac{s(n)}{n^{2\nu}(\ln n)^{1/4}} & \text{for } d = 4, \end{cases} \tag{14}$$

where the critical exponent $\nu = 1/2$ for $d > 4$ (Madras and Slade 1993), and it has been conjectured that

$$\nu = \begin{cases} \frac{3}{4} & \text{for } d = 2 \\ 0.59\ldots & \text{for } d = 3 \\ \frac{1}{2} & \text{for } d = 4. \end{cases} \tag{15}$$

see also RANDOM WALK

References

Alm, S. E. "Upper Bounds for the Connective Constant of Self-Avoiding Walks." *Combin. Prob. Comput.* **2**, 115–136, 1993.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/cnntv/cnntv.html`.

Madras, N. and Slade, G. *The Self-Avoiding Walk.* Boston, MA: Birkhäuser, 1993.

Self-Conjugate Subgroup

see INVARIANT SUBGROUP

Self-Descriptive Number

A 10-DIGIT number satisfying the following property. Number the DIGITS 0 to 9, and let DIGIT n be the number of ns in the number. There is exactly one such number: 6210001000.

References

Pickover, C. A. "Chaos in Ontario." Ch. 28 in *Keys to Infinity.* New York: W. H. Freeman, pp. 217–219, 1995.

Self-Homologous Point

see SIMILITUDE CENTER

Self Number

A number (usually base 10 unless specified otherwise) which has no GENERATOR. Such numbers were originally called COLUMBIAN NUMBERS (S. 1974). There are infinitely many such numbers, since an infinite sequence of self numbers can be generated from the RECURRENCE RELATION

$$C_k = 8 \cdot 10^{k-1} + C_{k-1} + 8, \tag{1}$$

for $k = 2, 3, \ldots$, where $C_1 = 9$. The first few self numbers are 1, 3, 5, 7, 9, 20, 31, 42, 53, 64, 75, 86, 97, ... (Sloane's A003052).

An infinite number of 2-self numbers (i.e., base-2 self numbers) can be generated by the sequence

$$C_k = 2^j + C_{k-1} + 1 \tag{2}$$

for $k = 1, 2, \ldots$, where $C_1 = 1$ and j is the number of digits in C_{k-1}. An infinite number of n-self numbers can be generated from the sequence

$$C_k = (n-2)n^{k-1} + C_{k-1} + (n-2) \tag{3}$$

for $k = 2, 3, \ldots$, and

$$C_1 = \begin{cases} n-1 & \text{for } n \text{ even} \\ n-2 & \text{for } n \text{ odd.} \end{cases} \tag{4}$$

Joshi (1973) proved that if k is ODD, then m is a k-self number IFF m is ODD. Patel (1991) proved that $2k$, $4k+2$, and k^2+2k+1 are k-self numbers in every EVEN base $k \geq 4$.

see also DIGITADITION

References
Cai, T. "On k-Self Numbers and Universal Generated Numbers." *Fib. Quart.* **34**, 144–146, 1996.
Gardner, M. *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, pp. 115–117, 122, 1988.
Joshi, V. S. Ph.D. dissertation. Gujarat University, Ahmadabad, 1973.
Kaprekar, D. R. *The Mathematics of New Self-Numbers.* Devaiali, pp. 19–20, 1963.
Patel, R. B. "Some Tests for k-Self Numbers." *Math. Student* **56**, 206–210, 1991.
S., B. R. Solution to Problem E 2048. *Amer. Math. Monthly* **81**, 407, 1974.
Sloane, N. J. A. Sequence A003052/M2404 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Self-Reciprocating Property

Let h be the number of sides of certain skew POLYGONS (Coxeter 1973, p. 15). Then

$$h = \frac{2(p+q+2)}{10-p-q}.$$

References
Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, 1973.

Self-Recursion

Self-recursion is a RECURSION which is defined in terms of itself, resulting in an ill-defined infinite regress.

see SELF-RECURSION

Self-Similarity

An object is said to be self-similar if it looks "roughly" the same on any scale. FRACTALS are a particularly interesting class of self-similar objects.

see also FRACTAL

References
Hutchinson, J. "Fractals and Self-Similarity." *Indiana Univ. J. Math.* **30**, 713–747, 1981.

Self-Transversality Theorem

Let j, r, and s be distinct INTEGERS (mod n), and let W_i be the point of intersection of the side or diagonal V_iV_{i+j} of the n-gon $P = [V_1, \ldots, V_n]$ with the transversal $V_{i+r}V_{i+s}$. Then a NECESSARY and SUFFICIENT condition for

$$\prod_{i=1}^{n} \left[\frac{V_iW_i}{W_iV_{i+j}}\right] = (-1)^n,$$

where $AB||CD$ and

$$\left[\frac{AB}{CD}\right],$$

is the ratio of the lengths $[A, B]$ and $[C, D]$ with a plus or minus sign depending on whether these segments have the same or opposite direction, is that

1. $n = 2m$ is EVEN with $j \equiv m \pmod{n}$ and $s \equiv r + m \pmod{n}$,
2. n is arbitrary and either $s \equiv 2r$ and $j \equiv 3r$, or
3. $r \equiv 2s \pmod{n}$ and $j \equiv 3s \pmod{n}$.

References
Grünbaum, B. and Shepard, G. C. "Ceva, Menelaus, and the Area Principle." *Math. Mag.* **68**, 254–268, 1995.

Selfridge's Conjecture

There exist infinitely many $n > 0$ with ${p_n}^2 > p_{n-i}p_{n+i}$ for all $i < n$. Also, there exist infinitely many $n > 0$ such that $2p_n < p_{n-i} + p_{n-i}$ for all $i < n$.

Selfridge-Hurwitz Residue

Let the RESIDUE from PÉPIN'S THEOREM be

$$R_n \equiv 3^{(F_n-1)/2} \pmod{F_n},$$

where F_n is a FERMAT NUMBER. Selfridge and Hurwitz use

$$R_n (\text{mod } 2^{35}-1, 2^{36}, 2^{36}-1).$$

A nonvanishing $R_n \pmod{2^{36}}$ indicates that F_n is COMPOSITE for $n > 5$.

see also FERMAT NUMBER, PÉPIN'S THEOREM

References
Crandall, R.; Doenias, J.; Norrie, C.; and Young, J. "The Twenty-Second Fermat Number is Composite." *Math. Comput.* **64**, 863–868, 1995.

Selmer Group

A GROUP which is related to the TANIYAMA-SHIMURA CONJECTURE.

see also TANIYAMA-SHIMURA CONJECTURE

Semi-Integral

An INTEGRAL of order 1/2. The semi-integral of the CONSTANT FUNCTION $f(x) = c$ is

$$\frac{d^{-1/2}c}{dx^{-1/2}} = 2c\sqrt{\frac{x}{\pi}}.$$

see also SEMIDERIVATIVE

References

Spanier, J. and Oldham, K. B. *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 8 and 14, 1987.

Semialgebraic Number

A subset of $\mathbb{R}^n$ which is a finite Boolean combination of sets of the form $\{\bar{x} = (x_1, \ldots, x_m) : f(\bar{x}) > 0\}$ and $\{\bar{x} : g(\bar{x}) = 0\}$, where $f, g \in \mathbb{R}[X_1, \ldots, X_n]$.

References

Bierstone, E. and Milman, P. "Semialgebraic and Subanalytic Sets." *IHES Pub. Math.* **67**, 5–42, 1988.

Marker, D. "Model Theory and Exponentiation." *Not. Amer. Math. Soc.* **43**, 753–759, 1996.

Semianalytic

$X \subseteq \mathbb{R}^n$ is semianalytic if, for all $x \in \mathbb{R}^n$, there is an open neighborhood U of x such that $X \cap U$ is a finite Boolean combination of sets $\{\bar{x} \in U : f(\bar{x}) = 0\}$ and $\{\bar{x} \in U : g(\bar{x}) > 0\}$, where $f, g : U \to \mathbb{R}$ are ANALYTIC.

see also ANALYTIC FUNCTION, PSEUDOANALYTIC FUNCTION, SUBANALYTIC

References

Marker, D. "Model Theory and Exponentiation." *Not. Amer. Math. Soc.* **43**, 753–759, 1996.

Semicircle

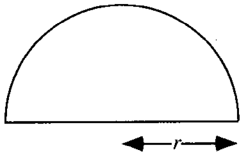

Half a CIRCLE. The PERIMETER of the semicircle of RADIUS r is

$$L = 2r + \pi r = r(2 + \pi), \tag{1}$$

and the AREA is

$$A = 2\int_0^r \sqrt{r^2 - y^2}\,dy = \tfrac{1}{2}\pi r^2. \tag{2}$$

The weighted mean of y is

$$\langle y \rangle = 2\int_0^r y\sqrt{r^2 - y^2}\,dy = \tfrac{2}{3}r^3. \tag{3}$$

The CENTROID is then given by

$$\bar{y} = \frac{\langle y \rangle}{A} = \frac{4r}{3\pi}. \tag{4}$$

The semicircle is the CROSS-SECTION of a HEMISPHERE for any PLANE through the z-AXIS.

see also ARBELOS, ARC, CIRCLE, DISK, HEMISPHERE, LENS, RIGHT ANGLE, SALINON, THALES' THEOREM, YIN-YANG

Semicolon Derivative

see COVARIANT DERIVATIVE

Semiconvergent Series

see ASYMPTOTIC SERIES

Semicubical Parabola

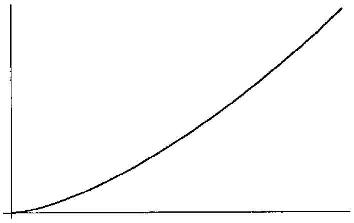

A PARABOLA-like curve with Cartesian equation

$$y = ax^{3/2}, \tag{1}$$

parametric equations

$$x = t^2 \tag{2}$$

$$y = at^3, \tag{3}$$

and POLAR COORDINATES,

$$r = \frac{\tan^2\theta\sec\theta}{a}. \tag{4}$$

The semicubical parabola is the curve along which a particle descending under gravity describes equal vertical spacings within equal times, making it an ISOCHRONOUS CURVE. The problem of finding the curve having this property was posed by Leibniz in 1687 and solved by Huygens (MacTutor Archive).

The ARC LENGTH, CURVATURE, and TANGENTIAL ANGLE are

$$s(t) = \tfrac{1}{27}(4 + 9t^2)^{3/2} - \tfrac{8}{27} \tag{5}$$

$$\kappa(t) = \frac{6}{t(4 + 9t^2)^{3/2}} \tag{6}$$

$$\phi(t) = \tan^{-1}(\tfrac{3}{2}t). \tag{7}$$

see also NEILE'S PARABOLA, PARABOLA INVOLUTE

References

Gray, A. "The Semicubical Parabola." §1.7 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 15–16, 1993.

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 85–87, 1972.

Lee, X. "Semicubic Parabola." `http://www.best.com/~xah/SpecialPlaneCurves_dir/SemicubicParabola_dir/semicubicParabola.html`.

MacTutor History of Mathematics Archive. "Neile's Parabola." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Neiles.html`.

Yates, R. C. "Semi-Cubic Parabola." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 186–187, 1952.

Semiderivative

A DERIVATIVE of order 1/2. The semiderivative of the CONSTANT FUNCTION $f(x) = c$ is

$$\frac{d^{1/2}c}{dx^{1/2}} = \frac{c}{\sqrt{\pi x}}.$$

see also DERIVATIVE, SEMI-INTEGRAL

References

Spanier, J. and Oldham, K. B. *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 8 and 14, 1987.

Semidirect Product

The "split" extension G of GROUPS N and F which contains a SUBGROUP $\bar{F}$ isomorphic to F with $G = \bar{F}\bar{N}$ and $\bar{F} \cap \bar{N} = \{e\}$.

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 613, 1980.

Semiflow

An ACTION with $G = \mathbb{R}^+$.

see also FLOW

Semigroup

A mathematical object defined for a set and a BINARY OPERATOR in which the multiplication operation is ASSOCIATIVE. No other restrictions are placed on a semigroup; thus a semigroup need not have an IDENTITY ELEMENT and its elements need not have inverses within the semigroup. A semigroup is an ASSOCIATIVE GROUPOID.

A semigroup can be empty. The total number of semigroups of order n are 1, 4, 18, 126, 1160, 15973, 836021, ... (Sloane's A001423). The number of semigroups of order n with one IDEMPOTENT are 1, 2, 5, 19, 132, 3107, 623615, ... (Sloane's A002786), and with two IDEMPOTENTS are 2, 7, 37, 216, 1780, 32652, ... (Sloane's A002787). The number $a(n)$ of semigroups having n IDEMPOTENTS are 1, 2, 6, 26, 135, 875, ... (Sloane's A002788).

see also ASSOCIATIVE, BINARY OPERATOR, FREE SEMIGROUP, GROUPOID, INVERSE SEMIGROUP, MONOID, QUASIGROUP

References

Clifford, A. H. and Preston, G. B. *The Algebraic Theory of Semigroups.* Providence, RI: Amer. Math. Soc., 1961.

Sloane, N. J. A. Sequences A001423/M3550, A002786/M1522, A002787/M1802, and A002788/M1679 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Semilatus Rectum

Given an ELLIPSE, the semilatus rectum is defined as the distance L measured from a FOCUS such that

$$\frac{1}{L} \equiv \frac{1}{2}\left(\frac{1}{r_+} + \frac{1}{r_-}\right), \tag{1}$$

where $r_+ = a(1+e)$ and $r_- = a(1-e)$ are the APOAPSIS and PERIAPSIS, and e is the ELLIPSE'S ECCENTRICITY. Plugging in for r_+ and r_- then gives

$$\begin{aligned}\frac{1}{L} &= \frac{1}{2a}\left(\frac{1}{1-e} + \frac{1}{1+e}\right) = \frac{1}{2a}\frac{(1+e)+(1-e)}{1-e^2} \\ &= \frac{1}{a}\frac{1}{1-e^2},\end{aligned} \tag{2}$$

so

$$L = a(1-e^2). \tag{3}$$

see also ECCENTRICITY, ELLIPSE, FOCUS, LATUS RECTUM, SEMIMAJOR AXIS, SEMIMINOR AXIS

Semimagic Square

A square that fails to be a MAGIC SQUARE only because one or both of the main diagonal sums do not equal the MAGIC CONSTANT is called a SEMIMAGIC SQUARE.

see also MAGIC SQUARE

Semimajor Axis

HALF the distance across an ELLIPSE along its long principal axis.

see also ELLIPSE, SEMIMINOR AXIS

Semiminor Axis

Half the distance across an ELLIPSE along its short principal axis.

see also ELLIPSE, SEMIMAJOR AXIS

Semiperfect Magic Cube

A semiperfect magic cube, also called an ANDREWS CUBE, is a MAGIC CUBE for which the cross-section diagonals do not sum to the MAGIC CONSTANT.

see also MAGIC CUBE, PERFECT MAGIC CUBE

References

Gardner, M. "Magic Squares and Cubes." Ch. 17 in *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, pp. 213–225, 1988.

Semiperfect Number

A number such as $20 = 1 + 4 + 5 + 10$ which is the SUM of some (or all) its PROPER DIVISORS. A semiperfect number which is the SUM of *all* its PROPER DIVISORS is called a PERFECT NUMBER. The first few semiperfect numbers are 6, 12, 18, 20, 24, 28, 30, 36, 40, ... (Sloane's A005835). Every multiple of a semiperfect number is semiperfect, as are all numbers $2^m p$ for $m \geq 1$ and p a PRIME between 2^m and 2^{m+1} (Guy 1994, p. 47).

A semiperfect number cannot be DEFICIENT. Rare ABUNDANT NUMBERS which are not semiperfect are called WEIRD NUMBERS. Semiperfect numbers are sometimes also called PSEUDOPERFECT NUMBERS.

see also ABUNDANT NUMBER, DEFICIENT NUMBER, PERFECT NUMBER, PRIMITIVE SEMIPERFECT NUMBER, WEIRD NUMBER

References

Guy, R. K. "Almost Perfect, Quasi-Perfect, Pseudoperfect, Harmonic, Weird, Multiperfect and Hyperperfect Numbers." §B2 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 45–53, 1994.

Sloane, N. J. A. Sequence A005835/M4094 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Zachariou, A. and Zachariou, E. "Perfect, Semi-Perfect and Ore Numbers." *Bull. Soc. Math. Grèce (New Ser.)* **13**, 12–22, 1972.

Semiperimeter

The semiperimeter on a figure is defined as

$$s \equiv \tfrac{1}{2}p, \qquad (1)$$

where p is the PERIMETER. The semiperimeter of POLYGONS appears in unexpected ways in the computation of their AREAS. The most notable cases are in the ALTITUDE, EXRADIUS, and INRADIUS of a TRIANGLE, the SODDY CIRCLES, HERON'S FORMULA for the AREA of a TRIANGLE in terms of the legs a, b, and c

$$A_\Delta = \sqrt{s(s-a)(s-b)(s-c)}, \qquad (2)$$

and BRAHMAGUPTA'S FORMULA for the AREA of a QUADRILATERAL

$$A_{\text{quadrilateral}} = \sqrt{(s-a)(s-b)(s-c)(s-d) - abcd\cos^2\left(\frac{A+B}{2}\right)}. \qquad (3)$$

The semiperimeter also appears in the beautiful L'HUILIER'S THEOREM about SPHERICAL TRIANGLES.

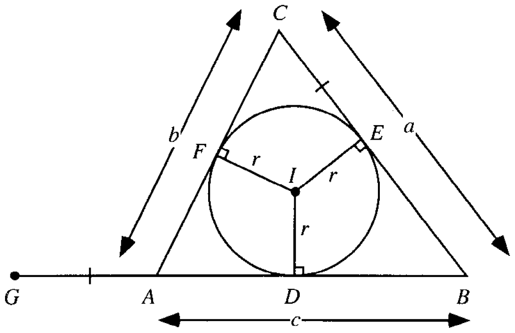

For a TRIANGLE, the following identities hold,

$$s-a = \tfrac{1}{2}(-a+b+c) \qquad (4)$$

$$s-b = \tfrac{1}{2}(a+b-c) \qquad (5)$$

$$s-c = \tfrac{1}{2}(a+b-c). \qquad (6)$$

Now consider the above figure. Let I be the INCENTER of the TRIANGLE ΔABC, with D, E, and F the tangent points of the INCIRCLE. Extend the line BA with $GA = CE$. Note that the pairs of triangles (ADI, AFI), (BDI, BEI), (CFI, CEI) are congruent. Then

$$\begin{aligned} BG &= BD + AD + AG = BD + AD + CE \\ &= \tfrac{1}{2}(2BD + 2AD + 2CE) \\ &= \tfrac{1}{2}[(BD+BE) + (AD+AF) + (CE+CF)] \\ &= \tfrac{1}{2}[(BD+AD) + (BE+CE) + (AF+CF)] \\ &= \tfrac{1}{2}(AB+BC+AC) = \tfrac{1}{2}(a+b+c) = s. \qquad (7) \end{aligned}$$

Furthermore,

$$\begin{aligned} s-a &= BG - BC \\ &= (BD+AD+AG) - (BE+CE) \\ &= (BD+AD+CE) - (BD+CE) = AC \qquad (8) \\ s-b &= BG - AC \\ &= (BD+AD+AG) - (AF+CF) \\ &= (BD+AD+CE) - (AD+CE) = BD \qquad (9) \\ s-c &= BG - AB = AG \qquad (10) \end{aligned}$$

(Dunham 1990). These equations are some of the building blocks of Heron's derivation of HERON'S FORMULA.

References

Dunham, W. "Heron's Formula for Triangular Area." Ch. 5 in *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 113–132, 1990.

Semiprime

A COMPOSITE number which is the PRODUCT of two PRIMES (possibly equal). They correspond to the 2-ALMOST PRIMES. The first few are 4, 6, 9, 10, 14, 15, 21, 22, ... (Sloane's A001358).

see also ALMOST PRIME, CHEN'S THEOREM, COMPOSITE NUMBER, PRIME NUMBER

References

Sloane, N. J. A. Sequence A001358/M3274 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Semiprime Ring

Given an IDEAL A, a semiprime ring is one for which $A^n = 0$ IMPLIES $A = 0$ for any POSITIVE n. Every PRIME RING is semiprime.

see also PRIME RING

Semiregular Polyhedron

A POLYHEDRON or plane TESSELLATION is called semiregular if its faces are all REGULAR POLYGONS and its corners are alike (Walsh 1972; Coxeter 1973, pp. 4 and 58; Holden 1991, p. 41). The usual name for a semiregular polyhedron is an ARCHIMEDEAN SOLIDS, of which there are exactly 13.

see also Archimedean Solid, Polyhedron, Tessellation

References

Coxeter, H. S. M. "Regular and Semi-Regular Polytopes I." *Math. Z.* **46**, 380–407, 1940.

Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, 1973.

Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, 1991.

Walsh, T. R. S. "Characterizing the Vertex Neighbourhoods of Semi-Regular Polyhedra." *Geometriae Dedicata* **1**, 117–123, 1972.

Semiring

A semiring is a set together with two Binary Operators $S(+,*)$ satisfying the following conditions:

1. Additive associativity: For all $a,b,c \in S$, $(a+b)+c = a+(b+c)$,
2. Additive commutativity: For all $a,b \in S$, $a+b = b+a$,
3. Multiplicative associativity: For all $a,b,c \in S$, $(a*b)*c = a*(b*c)$,
4. Left and right distributivity: For all $a,b,c \in S$, $a*(b+c) = (a*b)+(a*c)$ and $(b+c)*a = (b*a)+(c*a)$.

Thus a semiring is therefore a commutative Semigroup under addition and a Semigroup under multiplication. A semiring can be empty.

see also Binary Operator, Ring, Ringoid, Semigroup

References

Rosenfeld, A. *An Introduction to Algebraic Structures.* New York: Holden-Day, 1968.

Semisecant

see Transversal Line

Semisimple

A p-Element x of a Group G is semisimple if $E(C_G(x)) \neq 1$, where $E(H)$ is the commuting product of all components of H and $C_G(x)$ is the Centralizer of G.

see also Centralizer, p-Element

Semisimple Algebra

An Algebra with no nontrivial nilpotent Ideals. In the 1890s, Cartan, Frobenius, and Molien independently proved that any finite-dimensional semisimple algebra over the Real or Complex numbers is a finite and unique Direct Sum of Simple Algebras. This result was then extended to algebras over arbitrary fields by Wedderburn in 1907 (Kleiner 1996).

see also Ideal, Nilpotent Element, Simple Algebra

References

Kleiner, I. "The Genesis of the Abstract Ring Concept." *Amer. Math. Monthly* **103**, 417–424, 1996.

Semisimple Lie Group

A Lie Group which has a simply connected covering group Homeomorphic to $\mathbb{R}^n$. The prototype is any connected closed subgroup of upper Triangular Complex Matrices. The Heisenberg Group is such a group.

see also Heisenberg Group, Lie Group

References

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Semisimple Ring

A Semiprime Ring which is also an Artinian Ring.

see also Artinian Ring

Semistable

When a Prime l divides the Discriminant of a Elliptic Curve E, two or all three roots of E become congruent mod l. An Elliptic Curve is semistable if, for all such Primes l, only two roots become Congruent mod l (with more complicated definitions for $p = 2$ or 3).

see also Discriminant (Elliptic Curve), Elliptic Curve

Sensitivity

The probability that a Statistical Test will be positive for a true statistic.

see also Specificity, Statistical Test, Type I Error, Type II Error

Sentence

A Logic Formula with no Free variables.

Separating Edge

An Edge of a Graph is separating if a path from a point A to a point B must pass over it. Separating Edges can therefore be viewed as either bridges or dead ends.

see also Edge (Graph)

Separating Family

A Separating Family is a Set of Subsets in which each pair of adjacent elements are found separated, each in one of two disjoint subsets. The 26 letters of the alphabet can be separated by a family of 9,

$$\begin{array}{ccc}(abcdefghi) & (jklmnopqr) & (stuvwxyz)\\ (abcjklstu) & (defmnovwx) & (ghipqryz)\ .\\ (adgjmpsvy) & (behknqtwz) & (cfilorux)\end{array}$$

The minimal size of the separating family for an n-set is 0, 2, 3, 4, 5, 5, 6, 6, 6, 7, 7, 7, ... (Sloane's A007600).

see also Katona's Problem

References

Honsberger, R. "Cai Mao-Cheng's Solution to Katona's Problem on Families of Separating Subsets." Ch. 18 in *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 224–239, 1985.

Sloane, N. J. A. Sequence A007600/M0456 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Separation

Two distinct point pairs AC and BD separate each other if A, B, C, and D lie on a CIRCLE (or line) in such order that either of the arcs (or the line segment AC) contains one but not both of B and D. In addition, the point pairs separate each other if every CIRCLE through A and C intersects (or coincides with) every CIRCLE through B and D. If the point pairs separate each other, then the symbol $AC//BD$ is used.

Separation of Variables

A method of solving partial differential equations in a function Φ and variables x, y, ... by making a substitution of the form

$$\Phi(x,y,\ldots) \equiv X(x)Y(y)\cdots,$$

breaking the resulting equation into a set of independent ordinary differential equations, solving these for $X(x)$, $Y(y)$, ..., and then plugging them back into the original equation.

This technique works because if the product of functions of independent variables is a constant, each function must separately be a constant. Success requires choice of an appropriate coordinate system and may not be attainable at all depending on the equation. Separation of variables was first used by L'Hospital in 1750. It is especially useful in solving equations arising in mathematical physics, such as LAPLACE'S EQUATION, the HELMHOLTZ DIFFERENTIAL EQUATION, and the Schrödinger equation.

see also HELMHOLTZ DIFFERENTIAL EQUATION, LAPLACE'S EQUATION

References

Arfken, G. "Separation of Variables" and "Separation of Variables—Ordinary Differential Equations." §2.6 and §8.3 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 111–117 and 448–451, 1985.

Morse, P. M. and Feshbach, H. "Separable Coordinates" and "Table of Separable Coordinates in Three Dimensions." §5.1 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 464–523 and 655–666, 1953.

Separation Theorem

There exist numbers $y_1 < y_2 < \ldots < x_{n-1}$, $a < y_{n-1}$, $y_{n-1} < b$, such that

$$\lambda_\nu = \alpha(y_\nu) - \alpha(y_{\nu-1}),$$

where $\nu = 1, 2, \ldots, n$, $y_0 = a$ and $y_n = b$. Furthermore, the zeros $x_1, \ldots, x_n$, arranged in increasing order, alternate with the numbers $y_1, \ldots y_{n-1}$, so

$$x_\nu < y_\nu < x_{\nu+1}.$$

More precisely,

$$\begin{aligned}\alpha(x_\nu + \epsilon) - \alpha(a) &< \alpha(y_\nu) - \alpha(a)\\ &= \lambda_1 + \ldots + \lambda_\nu < \alpha(x_{\nu+1} - \epsilon) - \alpha(a)\end{aligned}$$

for $\nu = 1, \ldots, n-1$.

see also POINCARÉ SEPARATION THEOREM, STURMIAN SEPARATION THEOREM

References

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., p. 50, 1975.

Separatrix

A phase curve (invariant MANIFOLD) which meets a HYPERBOLIC FIXED POINT (intersection of a stable and an unstable invariant MANIFOLD). A separatrix marks a boundary between phase curves with different properties. For example, the separatrix in the equation of motion for the pendulum occurs at the angular momentum where oscillation gives way to rotation.

Septendecillion

In the American system, 10^{54}.

see also LARGE NUMBER

Septillion

In the American system, 10^{24}.

see also LARGE NUMBER

Sequence

A sequence is an ordered set of mathematical objects which is denoted using braces. For example, the symbol $\{2n\}_{n=1}^{\infty}$ denotes the infinite sequence of EVEN NUMBERS $\{2, 4, \ldots, 2n, \ldots\}$.

see also 196-ALGORITHM, A-SEQUENCE, ALCUIN'S SEQUENCE, $B2$-SEQUENCE, BEATTY SEQUENCE, CARMICHAEL SEQUENCE, CAUCHY SEQUENCE, CONVERGENT SEQUENCE, DEGREE SEQUENCE, DENSITY (SEQUENCE), FRACTAL SEQUENCE, GIUGA SEQUENCE, INFINITIVE SEQUENCE, INTEGER SEQUENCE, ITERATION SEQUENCE, LIST, NONAVERAGING SEQUENCE, PRIMITIVE SEQUENCE, REVERSE-THEN-ADD SEQUENCE, SCORE SEQUENCE, SERIES, SIGNATURE SEQUENCE, SORT-THEN-ADD SEQUENCE, ULAM SEQUENCE

Sequency

The sequency k of a WALSH FUNCTION is defined as half the number of zero crossings in the time base.

see also WALSH FUNCTION

Sequency Function

see WALSH FUNCTION

Sequential Graph

A CONNECTED GRAPH having e EDGES is said to be sequential if it is possible to label the nodes i with distinct INTEGERS f_i in$\{0, 1, 2, \ldots, e-1\}$ such that when EDGE ij is labeled $f_i + f_j$, the set of EDGE labels is a block of e consecutive integers (Grace 1983, Gallian 1990). No HARMONIOUS GRAPH is known which cannot also be labeled sequentially.

see also CONNECTED GRAPH, HARMONIOUS GRAPH

References

Gallian, J. A. "Open Problems in Grid Labeling." *Amer. Math. Monthly* **97**, 133–135, 1990.

Grace, T. "On Sequential Labelings of Graphs." *J. Graph Th.* **7**, 195–201, 1983.

Series

A series is a sum of terms specified by some rule. If each term increases by a constant amount, it is said to be an ARITHMETIC SERIES. If each term equals the previous multiplied by a constant, it is said to be a GEOMETRIC SERIES. A series usually has an INFINITE number of terms, but the phrase INFINITE SERIES is sometimes used for emphasis or clarity.

If the sum of partial sequences comprising the first few terms of the series does not converge to a LIMIT (e.g., it oscillates or approaches $\pm\infty$), it is said to diverge. An example of a convergent series is the GEOMETRIC SERIES

$$\sum_{n=0}^{\infty}(\tfrac{1}{2})^n = 2,$$

and an example of a divergent series is the HARMONIC SERIES

$$\sum_{n=1}^{\infty}\frac{1}{n} = \infty.$$

A number of methods known as CONVERGENCE TESTS can be used to determine whether a given series converges. Although terms of a series can have either sign, convergence properties can often be computed in the "worst case" of all terms being POSITIVE, and then applied to the particular series at hand. A series of terms u_n is said to be ABSOLUTELY CONVERGENT if the series formed by taking the absolute values of the u_n,

$$\sum_n |u_n|,$$

converges.

An especially strong type of convergence is called UNIFORM CONVERGENCE, and series which are uniformly convergent have particularly "nice" properties. For example, the sum of a UNIFORMLY CONVERGENT series of continuous functions is continuous. A CONVERGENT SERIES can be DIFFERENTIATED term by term, provided that the functions of the series have continuous derivatives and that the series of DERIVATIVES is UNIFORMLY CONVERGENT. Finally, a UNIFORMLY CONVERGENT series of continuous functions can be INTEGRATED term by term.

For a table listing the COEFFICIENTS for various series operations, see Abramowitz and Stegun (1972, p. 15).

While it can be difficult to calculate analytical expressions for arbitrary convergent infinite series, many algorithms can handle a variety of common series types. The program *Mathematica*® (Wolfram Research, Champaign, IL) implements many of these algorithms. General techniques also exist for computing the numerical values to any but the most pathological series (Braden 1992).

see also ALTERNATING SERIES, ARITHMETIC SERIES, ARTISTIC SERIES, ASYMPTOTIC SERIES, BIAS (SERIES), CONVERGENCE IMPROVEMENT, CONVERGENCE TESTS, EULER-MACLAURIN INTEGRATION FORMULAS, GEOMETRIC SERIES, HARMONIC SERIES, INFINITE SERIES, MELODIC SERIES, q-SERIES, RIEMANN SERIES THEOREM, SEQUENCE, SERIES EXPANSION, SERIES REVERSION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Infinite Series." §3.6 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 14, 1972.

Arfken, G. "Infinite Series." Ch. 5 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 277–351, 1985.

Boas, R. P. Jr. "Partial Sums of Infinite Series, and How They Grow." *Amer. Math. Monthly* **84**, 237–258, 1977.

Boas, R. P. Jr. "Estimating Remainders." *Math. Mag.* **51**, 83–89, 1978.

Borwein, J. M. and Borwein, P. B. "Strange Series and High Precision Fraud." *Amer. Math. Monthly* **99**, 622–640, 1992.

Braden, B. "Calculating Sums of Infinite Series." *Amer. Math. Monthly* **99**, 649–655, 1992.

Bromwich, T. J. I'a. and MacRobert, T. M. *An Introduction to the Theory of Infinite Series, 3rd ed.* New York: Chelsea, 1991.

Hansen, E. R. *A Table of Series and Products.* Englewood Cliffs, NJ: Prentice-Hall, 1975.

Hardy, G. H. *Divergent Series.* Oxford, England: Clarendon Press, 1949.

Jolley, L. B. W. *Summation of Series, 2nd rev. ed.* New York: Dover, 1961.

Knopp, K. *Theory and Application of Infinite Series.* New York: Dover, 1990.

Mangulis, V. *Handbook of Series for Scientists and Engineers.* New York: Academic Press, 1965.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Series and Their Convergence." §5.1 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 159–163, 1992.

Rainville, E. D. *Infinite Series.* New York: Macmillan, 1967.

Series Expansion

see LAURENT SERIES, MACLAURIN SERIES, POWER SERIES, SERIES REVERSION, TAYLOR SERIES

Series Inversion

see SERIES REVERSION

Series Multisection

If

$$f(x) = f_0 + f_1x + f_2x^2 + \ldots + f_nx^n + \ldots,$$

then

$$S(n,j) = f_jx^j + f_{j+n}x^{j+n} + f_{j+2n}x^{j+2n} + \ldots$$

is given by

$$S(n,j) = \frac{1}{n}\sum_{t=0}^{n-1} w^{-jt} f(w^t x),$$

where $w = e^{2\pi i/n}$.

see also SERIES REVERSION

References

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 210–214, 1985.

Series Reversion

Series reversion is the computation of the COEFFICIENTS of the inverse function given those of the forward function. For a function expressed in a series as

$$y = a_1 x + a_2 x^2 + a_3 x^3 + \ldots, \tag{1}$$

the series expansion of the inverse series is given by

$$x = A_1 y + A_2 y^2 + A_3 y^3 + \ldots. \tag{2}$$

By plugging (2) into (1), the following equation is obtained

$$\begin{aligned} y = {} & a_1 A_1 y + (a_2 {A_1}^2 + a_1 A_2) y^2 \\ & + (a_3 {A_1}^3 + 2a_2 A_1 A_2 + a_1 A_3) y^3 \\ & + (3a_3 {A_1}^2 A_2 + a_2 {A_2}^2 + a_2 A_1 A_3) + \ldots . \end{aligned} \tag{3}$$

Equating COEFFICIENTS then gives

$$A_1 = {a_1}^{-1} \tag{4}$$

$$A_2 = -\frac{a_2}{a_1}{A_1}^2 = -{a_1}^{-3} a_2 \tag{5}$$

$$A_3 = {a_1}^{-5}(2{a_2}^2 - a_1 a_3) \tag{6}$$

$$A_4 = {a_1}^{-7}(5a_1 a_2 a_3 - {a_1}^2 a_4 - 5{a_2}^3) \tag{7}$$

$$\begin{aligned} A_5 = {} & {a_1}^{-9}(6{a_1}^2 a_2 a_4 + 3{a_1}^2 a_2 a_3 + 14{a_2}^4 - {a_1}^3 a_5 \\ & - 21 a_1 {a_2}^2 a_3) \end{aligned} \tag{8}$$

$$\begin{aligned} A_6 = {} & {a_1}^{-11}(7{a_1}^3 a_2 a_5 + 7{a_1}^3 a_3 a_4 + 84 a_1 {a_2}^3 a_3 \\ & - {a_1}^4 a_6 - 28{a_1}^2 a_2 {a_3}^2 - 42{a_2}^5 - 28{a_1}^2 {a_2}^2 a_4) \end{aligned} \tag{9}$$

$$\begin{aligned} A_7 = {} & {a_1}^{-13}(8{a_1}^4 a_2 a_6 + 8{a_1}^4 a_3 a_5 + 4{a_1}^4 {a_4}^2 \\ & + 120{a_1}^2 {a_2}^3 a_4 + 180{a_1}^2 {a_2}^2 {a_3}^2 + 132{a_2}^6 \\ & - {a_1}^5 a_7 - 36{a_1}^3 {a_2}^2 a_5 - 72{a_1}^3 a_2 a_3 a_4 \\ & - 12{a_1}^3 {a_3}^3 - 330 a_1 {a_2}^4 a_3) \end{aligned} \tag{10}$$

(Dwight 1961, Abramowitz and Stegun 1972, p. 16). A derivation of the explicit formula for the nth term is given by Morse and Feshbach (1953),

$$\begin{aligned} A_n = {} & \frac{1}{n{a_1}^n} \sum_{s,t,u,\ldots} (-1)^{s+t+u+\ldots} \\ & \times \frac{n(n+1)\cdots(n-1+s+t+u+\ldots)}{s!t!u!\cdots}\left(\frac{a_2}{a_1}\right)^s \left(\frac{a_3}{a_1}\right)^t \cdots, \end{aligned} \tag{11}$$

where

$$s + 2t + 3u + \ldots = n - 1. \tag{12}$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, 1972.

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 316–317, 1985.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 297, 1987.

Dwight, H. B. *Table of Integrals and Other Mathematical Data, 4th ed.* New York: Macmillan, 1961.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 411–413, 1953.

Serpentine Curve

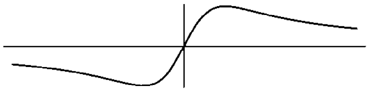

A curve named and studied by Newton in 1701 and contained in his classification of CUBIC CURVES. It had been studied earlier by L'Hospital and Huygens in 1692 (MacTutor Archive).

The curve is given by the CARTESIAN equation

$$y(x) = \frac{abx}{x^2 + a^2} \tag{1}$$

and parametric equations

$$x(t) = a \cot t \tag{2}$$

$$y(t) = b \sin t \cos t. \tag{3}$$

The curve has a MAXIMUM at $x = a$ and a MINIMUM at $x = -a$, where

$$y'(x) = \frac{ab(a-x)(a+x)}{(a^2+x^2)^2} = 0, \tag{4}$$

and inflection points at $x = \pm\sqrt{3}\, a$, where

$$y''(x) = \frac{2abx(x^2 - 3a^2)}{(x^2+a^2)^3} = 0. \tag{5}$$

The CURVATURE is given by

$$\kappa(x) = \frac{2abx(x^2 - 3a^2)}{(x^2+a^2)^3 \left[1 + \frac{(a^3 b - abx^2)^2}{(x^2+a^2)^4}\right]^{3/2}} \tag{6}$$

$$\kappa(t) = -\frac{4\sqrt{2}\, ab[2\cos(2t) - 1]\cot t \csc^2 t}{\{b^2[1+\cos(4t)] + 2a^2 \csc^4 t\}^{3/2}}. \tag{7}$$

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 111–112, 1972.

MacTutor History of Mathematics Archive. "Serpentine." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Serpentine.html`.

Serret-Frenet Formulas

see FRENET FORMULAS

Set

A set is a FINITE or INFINITE collection of objects. Older words for set include AGGREGATE and CLASS. Russell also uses the term MANIFOLD to refer to a set. The study of sets and their properties is the object of SET THEORY. Symbols used to operate on sets include $\wedge$ (which denotes the EMPTY SET $\varnothing$), $\vee =$ (which denotes the POWER SET of a set), $\cap$ (which means "and" or INTERSECTION), and $\cup$ (which means "or" or UNION).

The NOTATION A^B, where A and B are arbitrary sets, is used to denote the set of MAPS from B to A. For example, an element of $X^{\mathbb{N}}$ would be a MAP from the NATURAL NUMBERS $\mathbb{N}$ to the set X. Call such a function f, then $f(1)$, $f(2)$, etc., are elements of X, so call them x_1, x_2, etc. This now looks like a SEQUENCE of elements of X, so sequences are really just functions from $\mathbb{N}$ to X. This NOTATION is standard in mathematics and is frequently used in symbolic dynamics to denote sequence spaces.

Let E, F, and G be sets. Then operation on these sets using the $\cap$ and $\cup$ operators is COMMUTATIVE

$$E \cap F = F \cap E \tag{1}$$

$$E \cup F = F \cup E, \tag{2}$$

ASSOCIATIVE

$$(E \cap F) \cap G = E \cap (F \cap G) \tag{3}$$

$$A \cap \left(\bigcup_{i=1}^{n} B_i \right) = \bigcup_{i=1}^{n} (A \cap B_i) \tag{4}$$

$$(E \cup F) \cup G = E \cup (F \cup G), \tag{5}$$

and DISTRIBUTIVE

$$(E \cap F) \cup G = (E \cup G) \cap (F \cup G) \tag{6}$$

$$(E \cup F) \cap G = (E \cap G) \cup (F \cap G). \tag{7}$$

The proofs follow trivially using VENN DIAGRAMS.

$$P\left(\bigcup_{i=1}^{n} A_i \right) = \sum_{i=1}^{n} P(A_i). \tag{8}$$

The table below gives symbols for some common sets in mathematics.

Symbol	Set
$\mathbb{B}^n$	n-ball
$\mathbb{C}$	complex numbers
C^n, $C^{(n)}$	n-differentiable functions
$\mathbb{D}^n$	n-disk
$\mathbb{H}$	quaternions
$\mathbb{I}$	integers
$\mathbb{N}$	natural numbers
$\mathbb{Q}$	rational numbers
$\mathbb{R}^n$	real numbers in n-D
$\mathbb{S}^n$	n-sphere
$\mathbb{Z}$	integers
$\mathbb{Z}_n$	integers (mod n)
$\mathbb{Z}^-$	negative integers
$\mathbb{Z}^+$	positive integers
$\mathbb{Z}^*$	nonnegative integers

see also AGGREGATE, ANALYTIC SET, BOREL SET, $\mathbb{C}$, CLASS (SET), COANALYTIC SET, DEFINABLE SET, DERIVED SET, DOUBLE-FREE SET, EXTENSION, GROUND SET, $\mathbb{I}$, INTENSION, INTERSECTION, KINNEY'S SET, MANIFOLD, $\mathbb{N}$, PERFECT SET, POSET, $\mathbb{Q}$, $\mathbb{R}$, SET DIFFERENCE, SET THEORY, TRIPLE-FREE SET, UNION, VENN DIAGRAM, WELL-ORDERED SET, $\mathbb{Z}$, $\mathbb{Z}^-$, $\mathbb{Z}^+$

References

Courant, R. and Robbins, H. "The Algebra of Sets." Supplement to Ch. 2 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 108–116, 1996.

Set Difference

The set difference $A \backslash B$ is defined by

$$A \backslash B = \{x : x \in A \text{ and } x \notin B\}.$$

The same symbol is also used for QUOTIENT GROUPS.

Set Partition

A set partition of a SET S is a collection of disjoint SUBSETS $\mathbf{B}_0$, $\mathbf{B}_1$, ... of S whose UNION is S, where each $\mathbf{B}_i$ is called a BLOCK. The number of partitions of the SET $\{k\}_{k=1}^{n}$ is called a BELL NUMBER.

see also BELL NUMBER, BLOCK, RESTRICTED GROWTH STRING, STIRLING NUMBER OF THE SECOND KIND

References

Ruskey, F. "Info About Set Partitions." `http://sue.csc.uvic.ca/~cos/inf/setp/SetPartitions.html`.

Set Theory

The mathematical theory of SETS. Set theory is closely associated with the branch of mathematics known as LOGIC.

There are a number of different versions of set theory, each with its own rules and AXIOMS. In order of increasing CONSISTENCY STRENGTH, several versions of set theory include PEANO ARITHMETIC (ordinary ALGEBRA), second-order arithmetic (ANALYSIS), ZERMELO-FRAENKEL SET THEORY, Mahlo, weakly compact, hyper-Mahlo, ineffable, measurable, Ramsey, supercompact, huge, and n-huge set theory.

Given a set of REAL NUMBERS, there are 14 versions of set theory which can be obtained using only closure and complement (Beeler *et al.* 1972, Item 105).

see also AXIOMATIC SET THEORY, CONSISTENCY STRENGTH, CONTINUUM HYPOTHESIS, DESCRIPTIVE SET THEORY, IMPREDICATIVE, NAIVE SET THEORY, PEANO ARITHMETIC, SET, ZERMELO-FRAENKEL SET THEORY

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, pp. 36–44, Feb. 1972.

Brown, K. S. "Set Theory and Foundations." `http://www.seanet.com/~ksbrown/ifoundat.htm`.

Courant, R. and Robbins, H. "The Algebra of Sets." Supplement to Ch. 2 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 108–116, 1996.

Devlin, K. *The Joy of Sets: Fundamentals of Contemporary Set Theory, 2nd ed.* New York: Springer-Verlag, 1993.

Halmos, P. R. *Naive Set Theory.* New York: Springer-Verlag, 1974.

MacTutor History of Mathematics Archive. "The Beginnings of Set Theory." `http://www-groups.dcs.st-and.ac.uk/~history/HistTopics/Beginnings_of_set_theory.html`.

Stewart, I. *The Problems of Mathematics, 2nd ed.* Oxford: Oxford University Press, p. 96, 1987.

Sexagesimal

The base-60 notational system for representing REAL NUMBERS. A base-60 number system was used by the Babylonians and is preserved in the modern measurement of time (hours, minutes, and seconds) and ANGLES (DEGREES, ARC MINUTES, and ARC SECONDS).

see also BASE (NUMBER), BINARY, DECIMAL, HEXADECIMAL, OCTAL, QUATERNARY, SCRUPLE, TERNARY, VIGESIMAL

References

Bergamini, D. *Mathematics.* New York: Time-Life Books, pp. 16–17, 1969.

Weisstein, E. W. "Bases." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Bases.m`.

Sexdecillion

In the American system, 10^{51}.

see also LARGE NUMBER

Sextic Equation

The general sextic polynomial equation

$$x^6 + a_5x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = 0$$

can be solved in terms of HYPERGEOMETRIC FUNCTIONS in one variable using Klein's approach to solving the QUINTIC EQUATION.

see also CUBIC EQUATION, QUADRATIC EQUATION, QUARTIC EQUATION, QUINTIC EQUATION

References

Coble, A. B. "The Reduction of the Sextic Equation to the Valentiner Form—Problem." *Math. Ann.* **70**, 337–350, 1911a.

Coble, A. B. "An Application of Moore's Cross-ratio Group to the Solution of the Sextic Equation." *Trans. Amer. Math. Soc.* **12**, 311–325, 1911b.

Cole, F. N. "A Contribution to the Theory of the General Equation of the Sixth Degree." *Amer. J. Math.* **8**, 265–286, 1886.

Sextic Surface

An ALGEBRAIC SURFACE which can be represented implicitly by a polynomial of degree six in x, y, and z. Examples are the BARTH SEXTIC and BOY SURFACE.

see also ALGEBRAIC SURFACE, BARTH SEXTIC, BOY SURFACE, CUBIC SURFACE, DECIC SURFACE, QUADRATIC SURFACE, QUARTIC SURFACE

References

Catanese, F. and Ceresa, G. "Constructing Sextic Surfaces with a Given Number of Nodes." *J. Pure Appl. Algebra* **23**, 1–12, 1982.

Hunt, B. "Algebraic Surfaces." `http://www.mathematik.uni-kl.de/~wwwagag/Galerie.html`.

Sextillion

In the American system, 10^{21}.

see also LARGE NUMBER

Sexy Primes

Since a PRIME NUMBER cannot be divisible by 2 or 3, it must be true that, for a PRIME p, $p \equiv 6 \pmod{1,5}$. This motivates the definition of sexy primes as a pair of primes (p,q) such that $p-q=6$ ("sexy" since "sex" is the Latin word for "six."). The first few sexy prime pairs are (5, 11), (7, 13), (11, 17), (13, 19), (17, 23), (23, 29), (31, 37), (37, 43), (41, 47), (47, 53), ... (Sloane's A023201 and A046117).

Sexy constellations also exist. The first few sexy triplets (i.e., numbers such that each of $(p, p+6, p+12)$ is PRIME but $p+18$ is *not* PRIME) are (7, 13, 19), (17, 23, 29), (31, 37, 43), (47, 53, 59), ... (Sloane's A046118, A046119, and A046120). The first few sexy quadruplets are (11, 17, 23, 29), (41, 47, 53, 59), (61, 67, 73, 79), (251, 257, 263, 269), ... (Sloane's A046121, A046122, A046123, A046124). Sexy quadruplets can only begin with a PRIME ending in a "1." There is only a single sexy quintuplet, (5, 11, 17, 23, 29), since every fifth number of the form $6n \pm 1$ is divisible by 5, and therefore cannot be PRIME.

see also PRIME CONSTELLATION, PRIME QUADRUPLET, TWIN PRIMES

References

Sloane, N. J. A. Sequences A023201, A046117, A046118, A046119, A046120, A046121, A046122, A046123, and A046124 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Trotter, T. "Sexy Primes." `http://www.geocities.com/CapeCanaveral/Launchpad/8202/sexyprim.html`.

Seydewitz's Theorem

If a TRIANGLE is inscribed in a CONIC SECTION, any line conjugate to one side meets the other two sides in conjugate points.

see also CONIC SECTION, TRIANGLE

Sgn

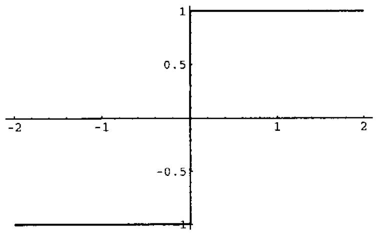

Also called SIGNUM. It can be defined as

$$\operatorname{sgn} \equiv \begin{cases} -1 & x<0 \\ 0 & x=0 \\ 1 & x>0 \end{cases} \tag{1}$$

or

$$\operatorname{sgn}(x) = 2H(x) - 1, \tag{2}$$

where $H(x)$ is the HEAVISIDE STEP FUNCTION. For $x \neq 0$, this can be written

$$\operatorname{sgn}(x) \equiv \frac{x}{|x|} \text{ for } x \neq 0. \tag{3}$$

see also HEAVISIDE STEP FUNCTION, RAMP FUNCTION

Shadow

The SURFACE corresponding to the region of obscuration when a solid is illuminated from a point light source (located at the RADIANT POINT). A DISK is the SHADOW of a SPHERE on a PLANE perpendicular to the SPHERE-RADIANT POINT line. If the PLANE is tilted, the shadow can be the interior of an ELLIPSE or a PARABOLA.

see also PROJECTIVE GEOMETRY

Shadowing Theorem

Although a numerically computed CHAOTIC trajectory diverges exponentially from the true trajectory with the same initial coordinates, there exists an errorless trajectory with a slightly different initial condition that stays near ("shadows") the numerically computed one. Therefore, the FRACTAL structure of chaotic trajectories seen in computer maps is real.

References

Ott, E. *Chaos in Dynamical Systems.* New York: Cambridge University Press, pp. 18–19, 1993.

Shafarevich Conjecture

A conjecture which implies the MORDELL CONJECTURE, as proved in 1968 by A. N. Parshin.

see also MORDELL CONJECTURE

References

Stewart, I. *The Problems of Mathematics, 2nd ed.* Oxford, England: Oxford University Press, p. 45, 1987.

Shah Function

$$\text{Ш}(x) \equiv \sum_{n=-\infty}^{\infty} \delta(x-n) \tag{1}$$

where $\delta(x)$ is the DELTA FUNCTION, so $\text{Ш}(x) = 0$ for $x \notin \mathbb{Z}$ (i.e., x not an INTEGER). The shah function obeys the identities

$$\text{Ш}(ax) = \frac{1}{a} \sum_{n=-\infty}^{\infty} \delta\left(x - \frac{n}{a}\right) \tag{2}$$

$$\text{Ш}(-x) = \text{Ш}(x) \tag{3}$$

$$\text{Ш}(x+n) = \text{Ш}(x), \tag{4}$$

for $2n \in \mathbb{Z}$ (i.e., n a half-integer).

It is normalized so that

$$\int_{n-1/2}^{n+1/2} \text{Ш}(x)\,dx = 1. \tag{5}$$

The "sampling property" is

$$\text{Ш}(x)f(x) = \sum_{n=-\infty}^{\infty} f(n)\delta(x-n) \tag{6}$$

and the "replicating property" is

$$\text{Ш}(x) * f(x) = \sum_{n=-\infty}^{\infty} f(x-n), \tag{7}$$

where $*$ denotes CONVOLUTION.

see also CONVOLUTION, DELTA FUNCTION, IMPULSE PAIR

Shah-Wilson Constant

see TWIN PRIMES CONSTANT

Shallit Constant

Define $f(x_1, x_2, \ldots, x_n)$ with x_i POSITIVE as

$$f(x_1, x_2, \ldots, x_n) \equiv \sum_{i=1}^{n} x_i + \sum_{1 \leq i \leq k \leq n} \prod_{j=i}^{k} \frac{1}{x_j}.$$

Then

$$\min f = 3n - C + o(1)$$

as n increases, where the Shallit constant is

$$C = 1.369451403937\ldots$$

(Shallit 1995). In their solution, Grosjean and De Meyer (quoted in Shallit 1995) reduced the complexity of the problem.

References

MacLeod, A. `http://www.mathsoft.com/asolve/constant/shapiro/macleod.html`.

Shallit, J. Solution by C. C. Grosjean and H. E. De Meyer. "A Minimization Problem." Problem 94-15 in *SIAM Review* **37**, 451–458, 1995.

Shallow Diagonal

see PASCAL'S TRIANGLE

Shanks' Algorithm

An ALGORITHM which finds the least NONNEGATIVE value of $\sqrt{a} \pmod{p}$ for given a and PRIME p.

Shanks' Conjecture

Let $p(g)$ be the first PRIME which follows a PRIME GAP of g between consecutive PRIMES. Shanks' conjecture holds that

$$\ln[p(g)] \sim \sqrt{g}.$$

see also PRIME DIFFERENCE FUNCTION, PRIME GAPS

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 21, 1994.

Rivera, C. "Problems & Puzzles (Conjectures): Shanks' Conjecture." `http://www.sci.net.mx/~crivera/ppp/conj_009.htm`.

Shanks, D. "On Maximal Gaps Between Successive Primes." *Math. Comput.* **18**, 646–651, 1964.

Shannon Entropy

see ENTROPY

Shannon Sampling Theorem

see SAMPLING THEOREM

Shape Operator

The negative derivative

$$S(\mathbf{v}) = -D_{\mathbf{v}}\mathbf{N} \tag{1}$$

of the unit normal $\mathbf{N}$ vector field of a SURFACE is called the shape operator (or WEINGARTEN MAP or SECOND FUNDAMENTAL TENSOR). The shape operator S is an EXTRINSIC CURVATURE, and the GAUSSIAN CURVATURE is given by the DETERMINANT of S. If $\mathbf{x} : U \to \mathbb{R}^3$ is a REGULAR PATCH, then

$$S(\mathbf{x}_u) = -\mathbf{N}_u \tag{2}$$

$$S(\mathbf{x}_v) = -\mathbf{N}_v. \tag{3}$$

At each point $\mathbf{p}$ on a REGULAR SURFACE $M \subset \mathbb{R}^3$, the shape operator is a linear map

$$S : M_{\mathbf{p}} \to M_{\mathbf{p}}. \tag{4}$$

The shape operator for a surface is given by the WEINGARTEN EQUATIONS.

see also CURVATURE, FUNDAMENTAL FORMS, WEINGARTEN EQUATIONS

References

Gray, A. "The Shape Operator," "Calculation of the Shape Operator," and "The Eigenvalues of the Shape Operator." §14.1, 14.3, and 14.4 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 268–269, 274–279, 1993.

Reckziegel, H. In *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, p. 30, 1986.

Shapiro's Cyclic Sum Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Consider the sum

$$f_n(x_1, x_2, \ldots, x_n) = \frac{x_1}{x_2 + x_3} + \frac{x_2}{x_3 + x_4} + \ldots + \frac{x_{n-1}}{x_n + x_1} + \frac{x_n}{x_1 + x_2}, \tag{1}$$

where the x_js are NONNEGATIVE and the DENOMINATORS are POSITIVE. Shapiro (1954) asked if

$$f_n(x_1, x_2, \ldots, x_n) \geq \tfrac{1}{2}n \tag{2}$$

for all n. It turns out (Mitrinovic *et al.* 1993) that this INEQUALITY is true for all EVEN $n \leq 12$ and ODD $n \leq 23$. Ranikin (1958) proved that for

$$f(n) = \inf_{x \geq 0} f_n(x_1, x_2, \ldots, x_n), \tag{3}$$

$$\lambda = \lim_{n\to\infty} \frac{f(n)}{n} = \inf_{n \geq 1} \frac{f(n)}{n} < \tfrac{1}{2} - 7 \times 10^{-8}. \tag{4}$$

λ can be computed by letting $\phi(x)$ be the CONVEX HULL of the functions

$$y_1 = e^{-x} \tag{5}$$

$$y_2 = \frac{2}{e^x + e^{x/2}}. \tag{6}$$

Then

$$\lambda = \tfrac{1}{2}\phi(0) = 0.4945668\ldots \tag{7}$$

(Drinfeljd 1971).

A modified sum was considered by Elbert (1973):

$$g_n(x_1, x_1, \ldots, x_n) = \frac{x_1 + x_3}{x_1 + x_2} + \frac{x_2 + x_4}{x_2 + x_3} + \ldots + \frac{x_{x-1} + x_1}{x_{n-1} + x_n} + \frac{x_n + x_2}{x_n + x_1}. \tag{8}$$

Consider

$$\mu = \lim_{n\to\infty} \frac{g(n)}{n}, \tag{9}$$

where

$$g(n) = \inf_{x \geq 0} g_n(x_1, x_2, \ldots, x_n), \tag{10}$$

and let $\psi(x)$ be the CONVEX HULL of

$$y_1 = \tfrac{1}{2}(1 + e^x) \tag{11}$$

$$y_2 = \frac{1 + e^x}{1 + e^{x/2}}. \tag{12}$$

Then

$$\mu = \psi(0) = 0.978012\ldots. \tag{13}$$

see also CONVEX HULL

References

Drinfeljd, V. G. "A Cyclic Inequality." *Math. Notes. Acad. Sci. USSR* **9**, 68–71, 1971.

Elbert, A. "On a Cyclic Inequality." *Period. Math. Hungar.* **4**, 163–168, 1973.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/shapiro/shapiro.html`.

Mitrinovic, D. S.; Pecaric, J. E.; and Fink, A. M. *Classical and New Inequalities in Analysis.* New York: Kluwer, 1993.

Sharing Problem

A problem also known as the Points Problem or Unfinished Game. Consider a tournament involving k players playing the same game repetitively. Each game has a single winner, and denote the number of games won by player i at some juncture w_i. The games are independent, and the probability of the ith player winning a game is p_i. The tournament is specified to continue until one player has won n games. If the tournament is discontinued before any player has won n games so that $w_i < n$ for $i = 1, \ldots, k$, how should the prize money be shared in order to distribute it proportionally to the players' chances of winning?

For player i, call the number of games left to win $r_i \equiv n - w_i > 0$ the "quota." For two players, let $p \equiv p_1$ and $q \equiv p_2 = 1 - p$ be the probabilities of winning a single game, and $a \equiv r_1 = n - w_1$ and $b \equiv r_2 = n - w_2$ be the number of games needed for each player to win the tournament. Then the stakes should be divided in the ratio $m : n$, where

$$m = p^a \left[1 + \frac{a}{1}q + \frac{a(a+1)}{2!}q^2 + \ldots + \frac{a(a+1)\cdots(a+b-2)}{(b-1)!}q^{b-1}\right] \quad (1)$$

$$n = q^b \left[1 + \frac{b}{1}p + \frac{b(b+1)}{2!}p^2 + \ldots + \frac{b(b+1)\cdots(b+a-2)}{(a-1)!}p^{a-1}\right] \quad (2)$$

(Kraitchik 1942).

If i players have equal probability of winning ("cell probability"), then the chance of player i winning for quotas $r_1, \ldots, r_k$ is

$$W_i = D_1^{k-1}(r_1, \ldots, r_{i-1}, r_{i+1}, \ldots, r_k; r_i), \quad (3)$$

where D is the Dirichlet Integral of type 2D. Similarly, the chance of player i losing is

$$L_i = C_1^{k-1}(r_1, \ldots, r_{i-1}, r_{i+1}, \ldots, r_k; r_i), \quad (4)$$

where C is the Dirichlet Integral of type 2C. If the cell quotas are not equal, the general Dirichlet integral $D_{\mathbf{a}}$ must be used, where

$$a_i = \frac{p_i}{1 - \sum_{i=1}^{k-1} p_i}. \quad (5)$$

If $r_i = r$ and $a_i = 1$, then W_i and L_i reduce to $1/k$ as they must. Let $P(r_1, \ldots, r_k)$ be the joint probability that the players would be Ranked in the order of the r_is in the argument list if the contest were completed. For $k = 3$,

$$P(r_1, r_2, r_3) = CD_1^{(1,1)}(r_1, r_2, r_3). \quad (6)$$

For $k = 4$ with quota vector $\mathbf{r} = (r_1, r_2, r_3, r_4)$ and $\Delta = p_2 + p_3 + p_4$,

$$P(\mathbf{r}) = \sum_{i=0}^{r_3-1} \sum_{j=0}^{r_4-1} \binom{r_2 - 1 + i + j}{r_2 - 1, i, j} \left(\frac{p_2}{\Delta}\right)^{r_2} \left(\frac{p_3}{\Delta}\right)^{i} \left(\frac{p_4}{\Delta}\right)^{j} \times C_{p_1/\Delta}^{(1)}(r_1, r_2 + i + j) D_{p_4/p_3}^{(1)}(r_4 - j, r_3 - i). \quad (7)$$

An expression for $k = 5$ is given by Sobel and Frankowski (1994, p. 838).

see also Dirichlet Integrals

References

Kraitchik, M. "The Unfinished Game." §6.1 in *Mathematical Recreations.* New York: W. W. Norton, pp. 117–118, 1942.

Sobel, M. and Frankowski, K. "The 500th Anniversary of the Sharing Problem (The Oldest Problem in the Theory of Probability)." *Amer. Math. Monthly* **101**, 833–847, 1994.

Sharkovsky's Theorem

see Šarkovskii's Theorem

Sharpe's Differential Equation

A generalization of the Bessel Differential Equation for functions of order 0, given by

$$zy'' + y' + (z + A)y = 0.$$

Solutions are

$$y = e^{\pm iz} {}_1F_1\left(\tfrac{1}{2} \mp \tfrac{1}{2}iA; 1; \mp 2iz\right),$$

where ${}_1F_1(a; b; x)$ is a Confluent Hypergeometric Function.

see also Bessel Differential Equation, Confluent Hypergeometric Function

Sharpe Ratio

A risk-adjusted financial measure developed by Nobel Laureate William Sharpe. It uses a fund's standard deviation and excess return to determine the reward per unit of risk. The higher a fund's Sharpe ratio, the better the fund's "risk-adjusted" performance.

see also Alpha, Beta

Sheaf (Geometry)

The set of all Planes through a Line.

see also Line, Pencil, Plane

References

Woods, F. S. *Higher Geometry: An Introduction to Advanced Methods in Analytic Geometry.* New York: Dover, p. 12, 1961.

Sheaf (Topology)

A topological Gadget related to families of Abelian Groups and Maps.

References

Iyanaga, S. and Kawada, Y. (Eds.). "Sheaves." §377 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1171–1174, 1980.

Shear

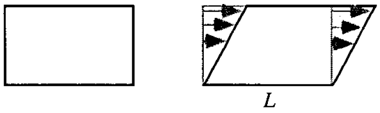

A transformation in which all points along a given LINE L remain fixed while other points are shifted parallel to L by a distance proportional to their PERPENDICULAR distance from L. Shearing a plane figure does not change its AREA. The shear can also be generalized to 3-D, in which PLANES are translated instead of lines.

Shear Matrix

The shear matrix e_{ij}^s is obtained from the IDENTITY MATRIX by inserting s at (i,j), e.g.,

$$\mathsf{e}_{12}^s = \begin{bmatrix} 1 & s & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

see also ELEMENTARY MATRIX

Shephard's Problem

Measurements of a centered convex body in Euclidean n-space (for $n \geq 3$) show that its brightness function (the volume of each projection) is smaller than that of another such body. Is it true that its VOLUME is also smaller? C. M. Petty and R. Schneider showed in 1967 that the answer is yes if the body with the larger brightness function is a projection body, but no in general for every n.

References

Gardner, R. J. "Geometric Tomography." *Not. Amer. Math. Soc.* **42**, 422–429, 1995.

Sheppard's Correction

A correction which must be applied to the MOMENTS computed from NORMALLY DISTRIBUTED data which have been binned. The corrected versions of the second, third, and fourth moments are

$$\mu_2 = {\mu_2}^{(0)} - \tfrac{1}{12}c^2 \qquad (1)$$

$$\mu_3 = {\mu_3}^{(0)} \qquad (2)$$

$$\mu_4 = {\mu_4}^{(0)} - \tfrac{1}{2}{\mu_2}^{(0)} + \tfrac{7}{240}c^2, \qquad (3)$$

where c is the CLASS INTERVAL. If κ_r' is the rth CUMULANT of an ungrouped distribution and κ_r the rth CUMULANT of the grouped distribution with CLASS INTERVAL c, the corrected cumulants (under rather restrictive conditions) are

$$\kappa_r' = \begin{cases} \kappa_r & \text{for } r \text{ odd} \\ \kappa_r - \frac{B_r}{r}c^r & \text{for } r \text{ even,} \end{cases} \qquad (4)$$

where B_r is the rth BERNOULLI NUMBER, giving

$$\kappa_1' = \kappa_1 \qquad (5)$$

$$\kappa_2' = \kappa_2 - \tfrac{1}{12}c^2 \qquad (6)$$

$$\kappa_3' = \kappa_3 \qquad (7)$$

$$\kappa_4' = \kappa_4 + \tfrac{1}{120}c^4 \qquad (8)$$

$$\kappa_5' = \kappa_5 \qquad (9)$$

$$\kappa_6' = \kappa_6 - \tfrac{1}{252}c^6. \qquad (10)$$

For a proof, see Kendall *et al.* (1987).

References

Kendall, M. G.; Stuart, A.; and Ord, J. K. *Kendall's Advanced Theory of Statistics, Vol. 1: Distribution Theory, 6th ed.* New York: Oxford University Press, 1987.

Kenney, J. F. and Keeping, E. S. "Sheppard's Correction." §4.12 in *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, pp. 80–82, 1951.

Sherman-Morrison Formula

A formula which allows the new MATRIX to be computed for a small change to a MATRIX A. If the change can be written in the form

$$\mathbf{u} \otimes \mathbf{v}$$

for two vectors $\mathbf{u}$ and $\mathbf{v}$, then the Sherman-Morrison formula is

$$(\mathsf{A} + \mathbf{u} \otimes \mathbf{v})^{-1} = \mathsf{A}^{-1} - \frac{(\mathsf{A}^{-1}\mathbf{u}) \otimes (\mathbf{v} \cdot \mathsf{A}^{-1})}{1 + \lambda},$$

where

$$\lambda \equiv \mathbf{v} \cdot \mathsf{A}^{-1}\mathbf{u}.$$

see also WOODBURY FORMULA

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Sherman-Morrison Formula." In *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 65–67, 1992.

Shi

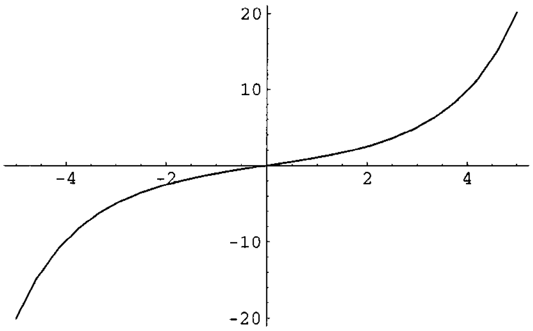

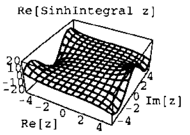
Re[SinhIntegral z]

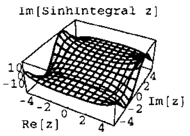
Im[SinhIntegral z]

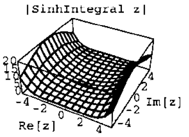
|SinhIntegral z|

$$\mathrm{Shi}(z) = \int_0^z \frac{\sinh t}{t}\,dt.$$

The function is given by the *Mathematica*® (Wolfram Research, Champaign, IL) command `SinhIntegral[z]`.

see also CHI, COSINE INTEGRAL, SINE INTEGRAL

References
Abramowitz, M. and Stegun, C. A. (Eds.). "Sine and Cosine Integrals." §5.2 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 231–233, 1972.

Shift

A TRANSLATION without ROTATION or distortion.

see also DILATION, EXPANSION, ROTATION, TRANSLATION, TWIRL

Shift Property

see DELTA FUNCTION

Shimura-Taniyama Conjecture

see TANIYAMA-SHIMURA CONJECTURE

Shimura-Taniyama-Weil Conjecture

see TANIYAMA-SHIMURA CONJECTURE

Shoe Surface

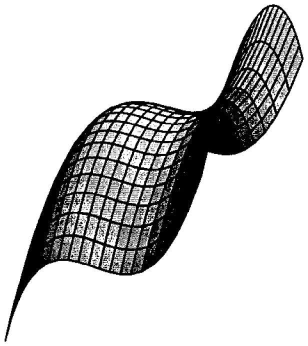

A surface given by the parametric equations

$$\begin{aligned} x(u,v) &= u \\ y(u,v) &= v \\ z(u,v) &= \tfrac{1}{3}u^3 - \tfrac{1}{2}v^2. \end{aligned}$$

References
Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 634, 1993.

Shoemaker's Knife

see ARBELOS

Shortening

A KNOT used to shorten a long rope.

see also BEND (KNOT)

References
Owen, P. *Knots.* Philadelphia, PA: Courage, p. 65, 1993.

Shuffle

The randomization of a deck of CARDS by repeated interleaving. More generally, a shuffle is a rearrangement of the elements in an ordered list. Shuffling by exactly interleaving two halves of a deck is called a RIFFLE SHUFFLE. Normal shuffling leaves gaps of different lengths between the two layers of cards and so randomizes the order of the cards.

A deck of 52 CARDS must be shuffled seven times for it to be randomized (Aldous and Diaconis 1986, Bayer and Diaconis 1992). This is intermediate between too few shuffles and the decreasing effectiveness of many shuffles. One of Bayer and Diaconis's randomness CRITERIA, however, gives $3\lg k/2$ shuffles for a k-card deck, yielding 11–12 shuffles for 52 CARDS. Keller (1995) shows that roughly $\ln k$ shuffles are needed just to randomize the bottom card.

see also BAYS' SHUFFLE, CARDS, FARO SHUFFLE, MONGE'S SHUFFLE, RIFFLE SHUFFLE

References
Aldous, D. and Diaconis, P. "Shuffling Cards and Stopping Times." *Amer. Math. Monthly* **93**, 333–348, 1986.
Bayer, D. and Diaconis, P. "Trailing the Dovetail Shuffle to Its Lair." *Ann. Appl. Probability* **2**, 294–313, 1992.
Keller, J. B. "How Many Shuffles to Mix a Deck?" *SIAM Review* **37**, 88–89, 1995.
Morris, S. B. "Practitioner's Commentary: Card Shuffling." *UMAP J.* **15**, 333–338, 1994.
Rosenthal, J. W. "Card Shuffling." *Math. Mag.* **54**, 64–67, 1981.

Siamese Dodecahedron

see SNUB DISPHENOID

Siamese Method

A method for constructing MAGIC SQUARES of ODD order, also called DE LA LOUBERE'S METHOD.

see also MAGIC SQUARE

Sibling

Two nodes connected to the same node in a ROOTED TREE are called siblings.

see also CHILD, ROOTED TREE

Sicherman Dice

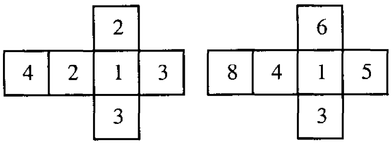

A pair of DICE which have the same ODDS for throwing every number as a normal pair of 6-sided DICE. They are the only such alternate arrangement.

see also DICE, EFRON'S DICE

Sici Spiral

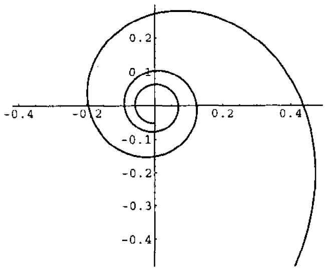

The spiral

$$x = c \operatorname{ci} t$$
$$y = c(\operatorname{si} t - \tfrac{1}{2}\pi),$$

where $\operatorname{ci}(t)$ and $\operatorname{si}(t)$ are the COSINE INTEGRAL and SINE INTEGRAL and c is a constant.

see also COSINE INTEGRAL, SINE INTEGRAL, SPIRAL

References

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 204 and 270, 1993.

Side

The edge of a POLYGON and face of a POLYHEDRON are sometimes called sides.

Sidon Sequence

see $B2$-SEQUENCE

Siegel Disk Fractal

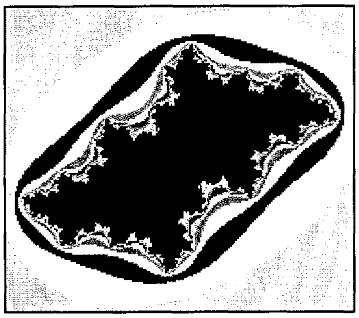

A JULIA SET with $c = -0.390541 - 0.586788i$. The FRACTAL somewhat resembles the better known MANDELBROT SET.

see also DOUADY'S RABBIT FRACTAL, JULIA SET, MANDELBROT SET, SAN MARCO FRACTAL

References

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, p. 176, 1991.

Siegel Modular Function

A Γ_n-invariant meromorphic function on the space of all $n \times n$ complex symmetric matrices with POSITIVE IMAGINARY PART. In 1984, H. Umemura expressed the ROOTS of an arbitrary POLYNOMIAL in terms of elliptic Siegel functions.

References

Iyanaga, S. and Kawada, Y. (Eds.). "Siegel Modular Functions." §34F in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 131–132, 1980.

Siegel's Paradox

If a fixed FRACTION x of a given amount of money P is lost, and then the same FRACTION x of the remaining amount is gained, the result is less than the original and equal to the final amount if a FRACTION x is first gained, then lost. This can easily be seen from the fact that

$$[P(1-x)](1+x) = P(1-x^2) < P$$
$$[P(1+x)](1-x) = P(1-x^2) < P.$$

Siegel's Theorem

An ELLIPTIC CURVE can have only a finite number of points with INTEGER coordinates.

see also ELLIPTIC CURVE

References

Davenport, H. "Siegel's Theorem." Ch. 21 in *Multiplicative Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 126–125, 1980.

Sierpiński Arrowhead Curve

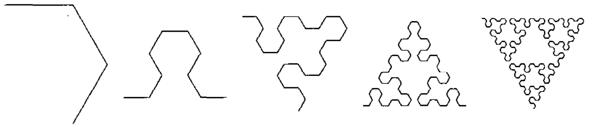

A FRACTAL which can be written as a LINDENMAYER SYSTEM with initial string `"YF"`, STRING REWRITING rules `"X" -> "YF+XF+Y"`, `"Y" -> "XF-YF-X"`, and angle 60°.

see also DRAGON CURVE, HILBERT CURVE, KOCH SNOWFLAKE, LINDENMAYER SYSTEM, PEANO CURVE, PEANO-GOSPER CURVE, SIERPIŃSKI CURVE, SIERPIŃSKI SIEVE

References

Dickau, R. M. "Two-Dimensional L-Systems." `http://forum.swarthmore.edu/advanced/robertd/lsys2d.html`.

Sierpiński Carpet

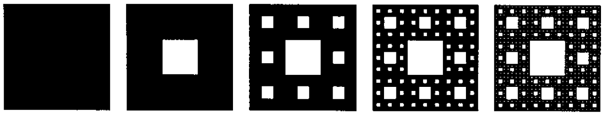

A FRACTAL which is constructed analogously to the SIERPIŃSKI SIEVE, but using squares instead of triangles. Let N_n be the number of black boxes, L_n the length of a side of a white box, and A_n the fractional AREA of black boxes after the nth iteration. Then

$$N_n = 8^n \tag{1}$$

$$L_n = (\tfrac{1}{3})^n = 3^{-n} \tag{2}$$

$$A_n = {L_n}^2 N_n = (\tfrac{8}{9})^n. \tag{3}$$

The CAPACITY DIMENSION is therefore

$$\begin{aligned} d_{\text{cap}} &= -\lim_{n\to\infty} \frac{\ln N_n}{\ln L_n} = -\lim_{n\to\infty} \frac{\ln(8^n)}{\ln(3^{-n})} = \frac{\ln 8}{\ln 3} \\ &= \frac{3\ln 2}{\ln 3} = 1.892789261\ldots. \end{aligned} \tag{4}$$

see also MENGER SPONGE, SIERPIŃSKI SIEVE

References

Dickau, R. M. "The Sierpinski Carpet." `http:// forum . swarthmore.edu/advanced/robertd/carpet.html.`

Peitgen, H.-O.; Jürgens, H.; and Saupe, D. *Chaos and Fractals: New Frontiers of Science.* New York: Springer-Verlag, pp. 112–121, 1992.

Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m.`

Sierpiński's Composite Number Theorem

There exist infinitely many ODD INTEGERS k such that $k \cdot 2^n + 1$ is COMPOSITE for every $n \geq 1$. Numbers k with this property are called SIERPIŃSKI NUMBERS OF THE SECOND KIND, and analogous numbers with the plus sign replaced by a minus are called RIESEL NUMBERS. It is conjectured that the smallest SIERPIŃSKI NUMBER OF THE SECOND KIND is $k = 78{,}557$ and the smallest RIESEL NUMBER is $k = 509{,}203$.

see also CUNNINGHAM NUMBER, SIERPIŃSKI NUMBER OF THE SECOND KIND

References

Buell, D. A. and Young, J. "Some Large Primes and the Sierpiński Problem." SRC Tech. Rep. 88004, Supercomputing Research Center, Lanham, MD, 1988.

Jaeschke, G. "On the Smallest k such that $k \cdot 2^N + 1$ are Composite." *Math. Comput.* **40**, 381–384, 1983.

Jaeschke, G. Corrigendum to "On the Smallest k such that $k \cdot 2^N + 1$ are Composite." *Math. Comput.* **45**, 637, 1985.

Keller, W. "Factors of Fermat Numbers and Large Primes of the Form $k \cdot 2^n + 1$." *Math. Comput.* **41**, 661–673, 1983.

Keller, W. "Factors of Fermat Numbers and Large Primes of the Form $k \cdot 2^n + 1$, II." In prep.

Ribenboim, P. *The New Book of Prime Number Records.* New York: Springer-Verlag, pp. 357–359, 1996.

Riesel, H. "Några stora primtal." *Elementa* **39**, 258–260, 1956.

Sierpiński, W. "Sur un problème concernant les nombres $k \cdot 2^n + 1$." *Elem. d. Math.* **15**, 73–74, 1960.

see also COMPOSITE NUMBER, SIERPIŃSKI NUMBERS OF THE SECOND KIND, SIERPIŃSKI'S PRIME SEQUENCE THEOREM

Sierpiński Constant

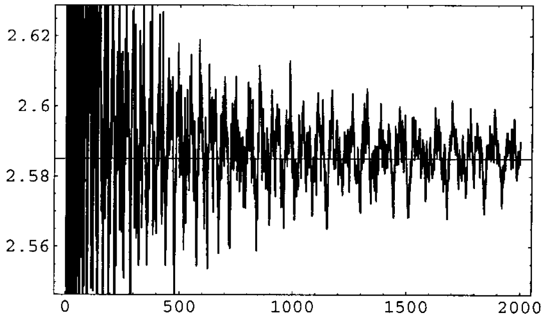

Let $r_k(n)$ denote the number of representations of n by k squares, then the SUMMATORY FUNCTION of $r_2(k)/k$ has the ASYMPTOTIC expansion

$$\sum_{k=1}^{n} \frac{r_2(k)}{k} = K + \pi \ln n + \mathcal{O}(n^{-1/2}),$$

where $K = 2.5849817596$ is the Sierpiński constant. The above plot shows

$$\left[\sum_{k=1}^{n} \frac{r_2(k)}{k}\right] - \pi \ln n,$$

with the value of K indicated as the solid horizontal line.

see also $r_k(n)$

References

Sierpiński, W. *Oeuvres Choiseies, Tome 1.* Editions Scientifiques de Pologne, 1974.

Sierpiński Curve

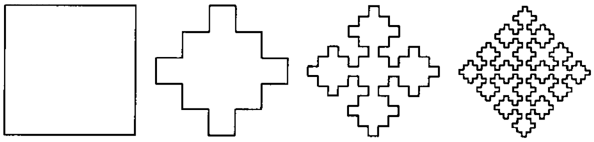

There are several FRACTAL curves associated with Sierpiński. The above curve is one example, and the SIERPIŃSKI ARROWHEAD CURVE is another. The limit of the curve illustrated above has AREA

$$A = \tfrac{5}{12}.$$

The AREA for a related curve illustrated by Cundy and Rollett (1989) is

$$A = \tfrac{1}{3}(7 - 4\sqrt{2}).$$

see also EXTERIOR SNOWFLAKE, GOSPER ISLAND, HILBERT CURVE, KOCH ANTISNOWFLAKE, KOCH SNOWFLAKE, PEANO CURVE, PEANO-GOSPER CURVE, SIERPIŃSKI ARROWHEAD CURVE

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 67–68, 1989.

Dickau, R. M. "Two-Dimensional L-Systems." `http://forum.swarthmore.edu/advanced/robertd/lsys2d.html`.

Gardner, M. *Penrose Tiles and Trapdoor Ciphers... and the Return of Dr. Matrix, reissue ed.* New York: W. H. Freeman, p. 34, 1989.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, p. 207, 1991.

Sierpiński Gasket

see SIERPIŃSKI SIEVE

Sierpiński-Menger Sponge

see MENGER SPONGE

Sierpiński Number of the First Kind

Numbers of the form $S_n \equiv n^n + 1$. The first few are 2, 5, 28, 257, 3126, 46657, 823544, 16777217, ... (Sloane's A014566). Sierpiński proved that if S_n is PRIME with $n \geq 2$, then $S_n = F_{m+2^m}$, where F_m is a FERMAT NUMBER with $m \geq 0$. The first few such numbers are $F_1 = 5$, $F_3 = 257$, F_6, F_{11}, F_{20}, and F_{37}. Of these, 5 and 257 are PRIME, and the first unknown case is $F_{37} > 10^{3\times 10^{10}}$.

see also CULLEN NUMBER, CUNNINGHAM NUMBER, FERMAT NUMBER, WOODALL NUMBER

References

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, p. 155, 1979.

Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, p. 74, 1989.

Sloane, N. J. A. Sequence A014566 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sierpiński Number of the Second Kind

A number k satisfying SIERPIŃSKI'S COMPOSITE NUMBER THEOREM, i.e., such that $k \cdot 2^n + 1$ is COMPOSITE for every $n \geq 1$. The smallest known is $k = 78,557$, but there remain 35 smaller candidates (the smallest of which is 4847) which are known to generate only composite numbers for $n \leq 18,000$ or more (Ribenboim 1996, p. 358).

Let $a(k)$ be smallest n for which $(2k - 1) \cdot 2^n + 1$ is PRIME, then the first few values are 0, 1, 1, 2, 1, 1, 2, 1, 3, 6, 1, 1, 2, 2, 1, 8, 1, 1, 2, 1, 1, 2, 2, 583, ... (Sloane's A046067). The second smallest n are given by 1, 2, 3, 4, 2, 3, 8, 2, 15, 10, 4, 9, 4, 4, 3, 60, 6, 3, 4, 2, 11, 6, 9, 1483, ... (Sloane's A046068). Quite large n can be required to obtain the first prime even for small k. For example, the smallest prime of the form $383 \cdot 2^n + 1$ is $383 \cdot 2^{6393} + 1$. There are an infinite number of Sierpiński numbers which are PRIME.

The smallest odd k such that $k + 2^n$ is COMPOSITE for all $n < k$ are 773, 2131, 2491, 4471, 5101, ...

see also MERSENNE NUMBER, RIESEL NUMBER, SIERPIŃSKI'S COMPOSITE NUMBER THEOREM

References

Buell, D. A. and Young, J. "Some Large Primes and the Sierpiński Problem." SRC Tech. Rep. 88004, Supercomputing Research Center, Lanham, MD, 1988.

Jaeschke, G. "On the Smallest k such that $k \cdot 2^N + 1$ are Composite." *Math. Comput.* **40**, 381–384, 1983.

Jaeschke, G. Corrigendum to "On the Smallest k such that $k \cdot 2^N + 1$ are Composite." *Math. Comput.* **45**, 637, 1985.

Keller, W. "Factors of Fermat Numbers and Large Primes of the Form $k \cdot 2^n + 1$." *Math. Comput.* **41**, 661–673, 1983.

Keller, W. "Factors of Fermat Numbers and Large Primes of the Form $k \cdot 2^n + 1$, II." In prep.

Ribenboim, P. *The New Book of Prime Number Records.* New York: Springer-Verlag, pp. 357–359, 1996.

Sierpiński, W. "Sur un problème concernant les nombres $k \cdot 2^n + 1$." *Elem. d. Math.* **15**, 73–74, 1960.

Sloane, N. J. A. Sequence A046067 in "An On-Line Version of the Encyclopedia of Integer Sequences."046068

Sierpiński's Prime Sequence Theorem

For any M, there exists a t' such that the sequence

$$n^2 + t',$$

where $n = 1, 2, \ldots$ contains at least M PRIMES.

see also DIRICHLET'S THEOREM, FERMAT $4n+1$ THEOREM, SIERPIŃSKI'S COMPOSITE NUMBER THEOREM

References

Abel, U. and Siebert, H. "Sequences with Large Numbers of Prime Values." *Amer. Math. Monthly* **100**, 167–169, 1993.

Ageev, A. A. "Sierpinski's Theorem is Deducible from Euler and Dirichlet." *Amer. Math. Monthly* **101**, 659–660, 1994.

Forman, R. "Sequences with Many Primes." *Amer. Math. Monthly* **99**, 548–557, 1992.

Garrison, B. "Polynomials with Large Numbers of Prime Values." *Amer. Math. Monthly* **97**, 316–317, 1990.

Sierpiński, W. "Les binômes $x^2 + n$ et les nombres premiers." *Bull. Soc. Roy. Sci. Liege* **33**, 259–260, 1964.

Sierpiński Sieve

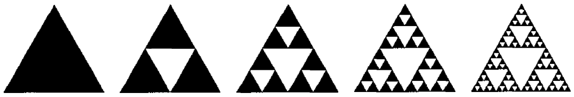

A FRACTAL described by Sierpiński in 1915. It is also called the SIERPIŃSKI GASKET or SIERPIŃSKI TRIANGLE. The curve can be written as a LINDENMAYER SYSTEM with initial string "FXF--FF--FF", STRING REWRITING rules "F" -> "FF", "X"->"--FXF++FXF++FXF--", and angle 60°.

Let N_n be the number of black triangles after iteration n, L_n the length of a side of a triangle, and A_n the fractional AREA which is black after the nth iteration. Then

$$N_n = 3^n \tag{1}$$

$$L_n = (\tfrac{1}{2})^n = 2^{-n} \tag{2}$$

$$A_n = {L_n}^2 N_n = (\tfrac{3}{4})^n. \tag{3}$$

The CAPACITY DIMENSION is therefore

$$d_{\text{cap}} = -\lim_{n\to\infty}\frac{\ln N_n}{\ln L_n} = -\lim_{n\to\infty}\frac{\ln(3^n)}{\ln(2^{-n})} = \frac{\ln 3}{\ln 2}$$
$$= 1.584962501\ldots. \tag{4}$$

In PASCAL'S TRIANGLE, coloring all ODD numbers black and EVEN numbers white produces a Sierpiński sieve.

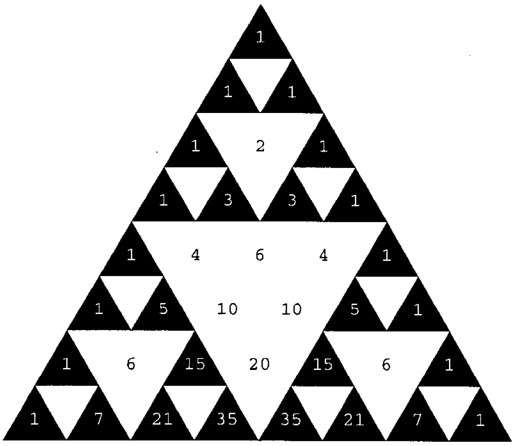

see also LINDENMAYER SYSTEM, SIERPIŃSKI ARROWHEAD CURVE, SIERPIŃSKI CARPET, TETRIX

References

Crownover, R. M. *Introduction to Fractals and Chaos.* Sudbury, MA: Jones & Bartlett, 1995.

Dickau, R. M. "Two-Dimensional L-Systems." `http://forum.swarthmore.edu/advanced/robertd/lsys2d.html`.

Dickau, R. M. "Typeset Fractals." *Mathematica J.* **7**, 15, 1997.

Dickau, R. "Sierpinski-Menger Sponge Code and Graphic." `http://www.mathsource.com/cgi-bin/MathSource/Applications/Graphics/0206-110`.

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 13–14, 1991.

Peitgen, H.-O.; Jürgens, H.; and Saupe, D. *Chaos and Fractals: New Frontiers of Science.* New York: Springer-Verlag, pp. 78–88, 1992.

Peitgen, H.-O. and Saupe, D. (Eds.). *The Science of Fractal Images.* New York: Springer-Verlag, p. 282, 1988.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 108 and 151–153, 1991.

Wang, P. "Renderings." `http://www.ugcs.caltech.edu/~peterw/portfolio/renderings/`.

Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m`.

Sierpiński Sponge

see TETRIX

Sierpiński Tetrahedron

see TETRIX

Sierpiński's Theorem

see SIERPIŃSKI'S COMPOSITE NUMBER THEOREM, SIERPIŃSKI'S PRIME SEQUENCE THEOREM

Sierpiński Triangle

see SIERPIŃSKI SIEVE

Sieve

A process of successively crossing out members of a list according to a set of rules such that only some remain. The best known sieve is the ERATOSTHENES SIEVE for generating PRIME NUMBERS. In fact, numbers generated by sieves seem to share a surprisingly large number of properties with the PRIME NUMBERS.

see also HAPPY NUMBER, NUMBER FIELD SIEVE FACTORIZATION METHOD, PRIME NUMBER, QUADRATIC SIEVE FACTORIZATION METHOD, SIERPIŃSKI SIEVE, SIEVE OF ERATOSTHENES, WALLIS SIEVE

References

Halberstam, H. and Richert, H.-E. *Sieve Methods.* New York: Academic Press, 1974.

Pomerance, C. "A Tale of Two Sieves." *Not. Amer. Math. Soc.* **43**, 1473–1485, 1996.

Sieve of Eratosthenes

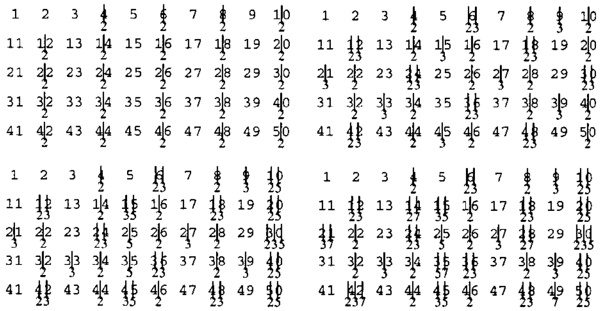

An ALGORITHM for making tables of PRIMES. Sequentially write down the INTEGERS from 2 to the highest number n you wish to include in the table. Cross out all numbers > 2 which are divisible by 2 (every second number). Find the smallest remaining number > 2. It is 3. So cross out all numbers > 3 which are divisible by 3 (every third number). Find the smallest remaining number > 3. It is 5. So cross out all numbers > 5 which are divisible by 5 (every fifth number).

Continue until you have crossed out all numbers divisible by $\lfloor\sqrt{n}\rfloor$, where $\lfloor x\rfloor$ is the FLOOR FUNCTION. The numbers remaining are PRIME. This procedure is illustrated in the above diagram which sieves up to 50, and therefore crosses out PRIMES up to $\lfloor\sqrt{50}\rfloor = 7$. If the procedure is then continued up to n, then the number of cross-outs gives the number of distinct PRIME factors of each number.

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 127–130, 1996.

Pappas, T. *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 100–101, 1989.

Ribenboim, P. *The New Book of Prime Number Records.* New York: Springer-Verlag, pp. 20–21, 1996.

Sievert Integral

The integral

$$\int_0^\theta e^{-x\sec\phi}\,d\phi.$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Sievert Integral." §27.4 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 1000–1001, 1972.

Sievert's Surface

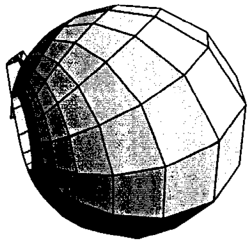

A special case of ENNEPER'S SURFACES which can be given parametrically by

$$x = r\cos\phi \qquad (1)$$

$$y = r\sin\phi \qquad (2)$$

$$z = \frac{\ln[\tan(\tfrac{1}{2}v)] + a(C+1)\cos v}{\sqrt{C}}, \qquad (3)$$

where

$$\phi \equiv -\frac{u}{\sqrt{C+1}} + \tan^{-1}(\tan u\sqrt{C+1}) \qquad (4)$$

$$a \equiv \frac{2}{C+1-C\sin^2 v\cos^2 u} \qquad (5)$$

$$r \equiv \frac{a\sqrt{(C+1)(1+C\sin^2 u)}\,\sin v}{\sqrt{C}}, \qquad (6)$$

with $|u| < \pi/2$ and $0 < v < \pi$ (Reckziegel 1986).

see also ENNEPER'S SURFACES, KUEN SURFACE, REMBS' SURFACES

References

Fischer, G. (Ed.). Plate 87 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 83, 1986.

Reckziegel, H. "Sievert's Surface." §3.4.4.3 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 38–39, 1986.

Sievert, H. *Über die Zentralflächen der Enneperschen Flächen konstanten Krümmungsmaßes.* Dissertation, Tübingen, 1886.

Sifting Property

The property

$$\int f(\mathbf{y})\delta(\mathbf{x}-\mathbf{y})\,d\mathbf{y} = f(\mathbf{x})$$

obeyed by the DELTA FUNCTION $\delta(\mathbf{x})$.

see also DELTA FUNCTION

Sigma Algebra

Let X be a SET. Then a σ-algebra F is a nonempty collection of SUBSETS of X such that the following hold:

1. The EMPTY SET is in F.
2. If A is in F, then so is the complement of A.
3. If A_n is a SEQUENCE of elements of F, then the UNION of the A_ns is in F.

If S is any collection of subsets of X, then we can always find a σ-algebra containing S, namely the POWER SET of X. By taking the INTERSECTION of all σ-algebras containing S, we obtain the smallest such σ-algebra. We call the smallest σ-algebra containing S the σ-algebra generated by S.

see also BOREL SIGMA ALGEBRA, BOREL SPACE, MEASURABLE SET, MEASURABLE SPACE, MEASURE ALGEBRA, STANDARD SPACE

Sigma Function

see DIVISOR FUNCTION

Sigmoid Curve

see SIGMOID FUNCTION

Sigmoid Function

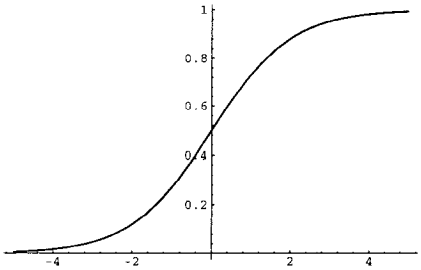

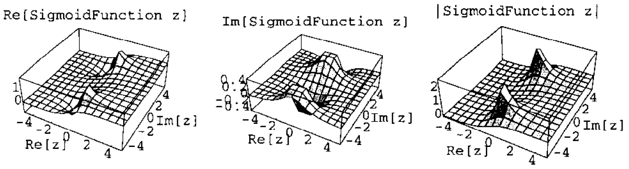

The function

$$y = \frac{1}{1+e^{-x}}$$

which is the solution to the ORDINARY DIFFERENTIAL EQUATION

$$\frac{dy}{dx} = y(1-y).$$

It has an inflection point at $x = 0$, where

$$y''(x) = -\frac{e^x(e^x-1)}{(e^x+1)^3} = 0.$$

see also EXPONENTIAL FUNCTION, EXPONENTIAL RAMP

References

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 124, 1993.

Sign

The sign of a number, also called Sgn, is -1 for a Negative number (i.e., one with a Minus Sign "$-$"), 0 for the number Zero, or $+1$ for a Positive number (i.e., one with a Plus Sign "$+$").

see also Absolute Value, Minus Sign, Negative, Plus Sign, Positive, Sgn, Zero

Signalizer Functor Theorem

$$\Theta(G;A) = \langle \theta(a) : a \in A - 1 \rangle$$

is an A-invariant solvable p'-subgroup of G.

Signature (Knot)

The signature $s(K)$ of a Knot K can be defined using the Skein Relationship

$$s(\text{unknot}) = 0$$

$$s(K_+) - s(K_-) \in \{0, 2\},$$

and

$$4 | s(K) \leftrightarrow \nabla(K)(2i) > 0,$$

where $\nabla(K)$ is the Alexander-Conway Polynomial and $\nabla(K)(2i)$ is an Odd Number.

Many Unknotting Numbers can be determined using a knot's signature.

see also Skein Relationship, Unknotting Number

References

Gordon, C. McA.; Litherland, R. A.; and Murasugi, K. "Signatures of Covering Links." *Canad. J. Math.* **33**, 381–394, 1981.

Murasugi, K. "On the Signature of Links." *Topology* **9**, 283–298, 1970.

Murasugi, K. "Signatures and Alexander Polynomials of Two-Bridge Knots." *C. R. Math. Rep. Acad. Sci. Canada* **5**, 133–136, 1983.

Murasugi, K. "On the Signature of a Graph." *C. R. Math. Rep. Acad. Sci. Canada* **10**, 107–111, 1988.

Murasugi, K. "On Invariants of Graphs with Applications to Knot Theory." *Trans. Amer. Math. Soc.* **314**, 1–49, 1989.

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, 1976.

Stoimenow, A. "Signatures." `http://www.informatik.hu-berlin.de/~stoimeno/ptab/sig10.html`.

Signature (Quadratic Form)

The signature of the Quadratic Form

$$Q = {y_1}^2 + {y_2}^2 + \ldots + {y_p}^2 - {y_{p+1}}^2 - {y_{p+2}}^2 - \ldots - {y_r}^2$$

is the number s of Positive squared terms in the reduced form. (The signature is sometimes defined as $2s - r$.)

see also p-Signature, Rank (Quadratic Form), Sylvester's Inertia Law, Sylvester's Signature

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1105, 1979.

Signature (Recurrence Relation)

Let a sequence be defined by

$$\begin{aligned} A_{-1} &= s \\ A_0 &= 3 \\ A_1 &= r \\ A_n &= rA_{n-1} - sA_{n-2} + A_{n-3}. \end{aligned}$$

Also define the associated Polynomial

$$f(x) = x^3 - rx^2 + sx + 1,$$

and let Δ be its discriminant. The Perrin Sequence is a special case corresponding to $A_n(0, -1)$. The signature mod m of an Integer n with respect to the sequence $A_k(r, s)$ is then defined as the 6-tuple (A_{-n-1}, A_{-n}, A_{-n+1}, A_{n-1}, A_n, A_{n+1}) (mod m).

1. An Integer n has an S-signature if its signature (mod n) is (A_{-2}, A_{-1}, A_0, A_1, A_2).
2. An Integer n has a Q-signature if its signature (mod n) is Congruent to (A, s, B, B, r, C) where, for some Integer a with $f(a) \equiv 0 \pmod{n}$, $A \equiv a^{-2} + 2a$, $B \equiv -ra^2 + (r^2 - s)a$, and $C \equiv a^2 + 2a^{-1}$.
3. An Integer n has an I-signature if its signature (mod n) is Congruent to (r, s, D', D, r, s), where $D' + D \equiv rs - 3$ and $(D' - D)^2 \equiv \Delta$.

see also Perrin Pseudoprime

References

Adams, W. and Shanks, D. "Strong Primality Tests that Are Not Sufficient." *Math. Comput.* **39**, 255–300, 1982.

Grantham, J. "Frobenius Pseudoprimes." `http://www.clark.net/pub/grantham/pseudo/pseudo.ps`

Signature Sequence

Let θ be an Irrational Number, define $S(\theta) = \{c + d\theta : c, d \in \mathbb{N}\}$, and let $c_n(\theta) + d_n\theta(\theta)$ be the sequence obtained by arranging the elements of $S(\theta)$ in increasing order. A sequence x is said to be a signature sequence if there Exists a Positive Irrational Number θ such that $x = \{c_n(\theta)\}$, and x is called the signature of θ.

The signature of an Irrational Number is a Fractal Sequence. Also, if x is a signature sequence, then the Lower-Trimmed Subsequence is $V(x) = x$.

References

Kimberling, C. "Fractal Sequences and Interspersions." *Ars Combin.* **45**, 157–168, 1997.

Signed Deviation

The signed deviation is defined by

$$\Delta u_i \equiv (u_i - \bar{u}),$$

so the average deviation is

$$\overline{\Delta u} = \overline{u_i - \bar{u}} = \overline{u_i} - \bar{u} = 0.$$

see also Absolute Deviation, Deviation, Dispersion (Statistics), Mean Deviation, Quartile Deviation, Standard Deviation, Variance

Significance

Let $\delta \equiv z \leq z_{\text{observed}}$. A value $0 \leq \alpha \leq 1$ such that $P(\delta) \leq \alpha$ is considered "significant" (i.e., is not simply due to chance) is known as an ALPHA VALUE. The PROBABILITY that a variate would assume a value greater than or equal to the observed value strictly by chance, $P(\delta)$, is known as a P-VALUE.

Depending on the type of data and conventional practices of a given field of study, a variety of different alpha values may be used. One commonly used terminology takes $P(\delta) \geq 5\%$ as "not significant," $1\% < P(\delta) < 5\%$, as "significant" (sometimes denoted *), and $P(\delta) < 1\%$ as "highly significant" (sometimes denoted **). Some authors use the term "almost significant" to refer to $5\% < P(\delta) < 10\%$, although this practice is not recommended.

see also ALPHA VALUE, CONFIDENCE INTERVAL, P-VALUE, PROBABLE ERROR, SIGNIFICANCE TEST, STATISTICAL TEST

Significance Test

A test for determining the probability that a given result could not have occurred by chance (its SIGNIFICANCE).

see also SIGNIFICANCE, STATISTICAL TEST

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 491–492, 1987.

Significant Digits

When a number is expressed in SCIENTIFIC NOTATION, the number of significant figures is the number of DIGITS needed to express the number to within the uncertainty of measurement. For example, if a quantity had been measured to be 1.234 ± 0.002, four figures would be significant. No more figures should be given than are allowed by the uncertainty. For example, a quantity written as 1.234 ± 0.1 is incorrect; it should really be written as 1.2 ± 0.1.

The number of significant figures of a MULTIPLICATION or DIVISION of two or more quantities is equal to the smallest number of significant figures for the quantities involved. For ADDITION or MULTIPLICATION, the number of significant figures is determined with the smallest significant figure of all the quantities involved. For example, the sum $10.234 + 5.2 + 100.3234$ is 115.7574, but should be written 115.8 (with rounding), since the quantity 5.2 is significant only to ± 0.1.

see also NINT, ROUND, TRUNCATE

Significant Figures

see SIGNIFICANT DIGITS

Signpost

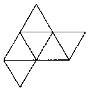

A 6-POLYIAMOND.

References

Golomb, S. W. *Polyominoes: Puzzles, Patterns, Problems, and Packings, 2nd ed.* Princeton, NJ: Princeton University Press, p. 92, 1994.

Signum

see SGN

Silver Constant

The REAL ROOT of the equation

$$x^3 - 5x^2 + 6x - 1 = 0,$$

which is 3.2469.... It is the seventh BERAHA CONSTANT.

see also BERAHA CONSTANTS

References

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, pp. 51 and 143, 1983.

Silver Mean

see SILVER RATIO

Silver Ratio

The quantity defined by the CONTINUED FRACTION

$$\delta_S \equiv [2, 2, 2, \ldots] = 2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cdots}}}.$$

It follows that

$$(\delta_S - 1)^2 = 2,$$

so

$$\delta_S = \sqrt{2} + 1 = 2.41421\ldots.$$

see also GOLDEN RATIO, GOLDEN RATIO CONJUGATE

Silverman Constant

$$\sum_{n=1}^{\infty} \frac{1}{\phi(n)\sigma(n)} = \prod_{p \text{ prime}} \left(1 + \sum_{k=1}^{\infty} \frac{1}{p^{2k} - p^{k-1}}\right)$$
$$= 1.786576459\ldots,$$

where $\phi(n)$ is the TOTIENT FUNCTION and $\sigma(n)$ is the DIVISOR FUNCTION.

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/totient/totient.html`.

Zimmerman, P. `http://www.mathsoft.com/asolve/constant/totient/zimmermn.html`.

Silverman's Sequence

Let $f(1) = 1$, and let $f(n)$ be the number of occurrences of n in a nondecreasing sequence of Integers. Then the first few values of $f(n)$ are 1, 2, 2, 3, 3, 4, 4, 4, 5, 5, 5, ... (Sloane's A001462). The asymptotic value of the nth term is $\phi^{2-\phi}n^{\phi-1}$, where ϕ is the Golden Ratio.

References

Guy, R. K. "Silverman's Sequences." §E25 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 225–226, 1994.

Sloane, N. J. A. Sequence A001462/M0257 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Similar

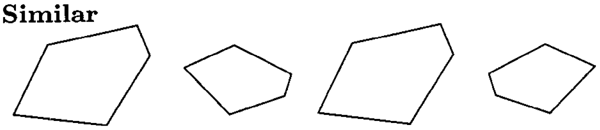

directly similar *inversely similar*

Two figures are said to be similar when all corresponding Angles are equal. Two figures are Directly Similar when all corresponding Angles are equal and described in the same rotational sense. This relationship is written $A \sim B$. (The symbol $\sim$ is also used to mean "is the same order of magnitude as" and "is Asymptotic to.") Two figures are Inversely Similar when all corresponding Angles are equal and described in the opposite rotational sense.

see also Directly Similar, Inversely Similar, Similarity Transformation

References

Project Mathematics! *Similarity.* Videotape (27 minutes). California Institute of Technology. Available from the Math. Assoc. Amer.

Similarity Axis

see d'Alembert's Theorem

Similarity Dimension

To multiply the size of a d-D object by a factor a, $c \equiv a^d$ copies are required, and the quantity

$$d = \frac{\ln c}{\ln a}$$

is called the similarity dimension.

Similarity Point

External (or positive) and internal (or negative) similarity points of two Circles with centers C and C' and Radii r and r' are the points E and I on the lines CC' such that

$$\frac{CE}{C'E} = \frac{r}{r'},$$

or

$$\frac{CI}{C'I} = -\frac{r}{r'}.$$

Similarity Transformation

An Angle-preserving transformation. A similarity transformation has a transformation Matrix A' of the form

$$\mathsf{A}' \equiv \mathsf{BAB}^{-1}.$$

If A is an Antisymmetric Matrix ($a_{ij} = -a_{ji}$) and B is an Orthogonal Matrix, then

$$\begin{aligned}(bab^{-1})_{ij} &= b_{ik}a_{kl}b_{lj}^{-1} = -b_{ik}a_{lk}b_{lj}^{-1} = -{b^\dagger}_{ki}a_{lk}{(b^\dagger)^{-1}}_{jl}\\ &= -{b^{-1}}_{ki}a_{ki}b_{jl} = b_{jl}a_{lk}b_{ki}^{-1} = -(bab^{-1})_{ji}.\end{aligned}$$

Similarity transformations and the concept of Self-Similarity are important foundations of Fractals and Iterated Function Systems.

see also Conformal Transformation

References

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 83–103, 1991.

Similitude Center

Also called a Self-Homologous Point. If two Similar figures lie in the plane but do not have parallel sides (they are not Homothetic), there exists a center of similitude which occupies the same homologous position with respect to the two figures. The Locus of similitude centers of two nonconcentric circles is another circle having the line joining the two homothetic centers as its Diameter.

There are a number of interesting theorems regarding three Circles (Johnson 1929, pp. 151–152).

1. The external similitude centers of three circles are Collinear.
2. Any two internal similitude centers are Collinear with the third external one.
3. If the center of each circle is connected with the internal similitude center of the other three [sic], the connectors are Concurrent.
4. If one center is connected with the internal similitude center of the other two, the others with the corresponding external centers, the connectors are Concurrent.

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 19–27 and 151–153, 1929.

Similitude Ratio

Two figures are Homothetic if they are related by a Dilation (a dilation is also known as a Homothecy). This means that the connectors of corresponding points are Concurrent at a point which divides each connector in the same ratio k, known as the similitude ratio.

see also Concurrent, Dilation, Homothecy, Homothetic

Simple Algebra

An ALGEBRA with no nontrivial IDEALS.

see also ALGEBRA, IDEAL, SEMISIMPLE ALGEBRA

Simple Continued Fraction

A CONTINUED FRACTION

$$\sigma = b_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{b_3 + \dots}}} \tag{1}$$

in which the b_is are all unity, leaving a continued fraction of the form

$$\sigma = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \dots}}}. \tag{2}$$

A simple continued fraction can be written in a compact abbreviated NOTATION as

$$\sigma = [a_0, a_1, a_2, a_3, \dots]. \tag{3}$$

Bach and Shallit (1996) show how to compute the JACOBI SYMBOL in terms of the simple continued fraction of a RATIONAL NUMBER a/b.

see also CONTINUED FRACTION

References

Bach, E. and Shallit, J. *Algorithmic Number Theory, Vol. 1: Efficient Algorithms.* Cambridge, MA: MIT Press, pp. 343–344, 1996.

Simple Curve

A curve is simple closed if it does not cross itself.

see also JORDAN CURVE

Simple Graph

A GRAPH for which at most one EDGE connects any two nodes.

see also ADJACENCY MATRIX, EDGE (GRAPH)

Simple Group

A simple group is a GROUP whose NORMAL SUBGROUPS (INVARIANT SUBGROUPS) are ORDER one or the whole of the original GROUP. Simple groups include ALTERNATING GROUPS, CYCLIC GROUPS, LIE-TYPE GROUPS (five varieties), and SPORADIC GROUPS (26 varieties, including the MONSTER GROUP). The CLASSIFICATION THEOREM of finite simple groups states that such groups can be classified completely into the three types:

1. CYCLIC GROUPS of PRIME ORDER,
2. ALTERNATING GROUPS of degree at least five
3. LIE-TYPE CHEVALLEY GROUPS,
4. LIE-TYPE (TWISTED CHEVALLEY GROUPS or the TITS GROUP), and
5. SPORADIC GROUPS.

BURNSIDE'S CONJECTURE states that every non-ABELIAN SIMPLE GROUP has EVEN ORDER.

see also ALTERNATING GROUP, BURNSIDE'S CONJECTURE, CHEVALLEY GROUPS, CLASSIFICATION THEOREM, CYCLIC GROUP, FEIT-THOMPSON THEOREM, FINITE GROUP, GROUP, INVARIANT SUBGROUP, LIE-TYPE GROUP, MONSTER GROUP, SCHUR MULTIPLIER, SPORADIC GROUP, TITS GROUP, TWISTED CHEVALLEY GROUPS

Simple Harmonic Motion

Simple harmonic motion refers to the periodic sinusoidal oscillation of an object or quantity. Simple harmonic motion is executed by any quantity obeying the DIFFERENTIAL EQUATION

$$\ddot{x} + {\omega_0}^2 x = 0, \tag{1}$$

where $\ddot{x}$ denotes the second DERIVATIVE of x with respect to t, and ω_0 is the angular frequency of oscillation. This ORDINARY DIFFERENTIAL EQUATION has an irregular SINGULARITY at ∞. The general solution is

$$\begin{aligned} x &= A\sin(\omega_0 t) + B\cos(\omega_0 t) && (2) \\ &= C\cos(\omega_0 t + \phi), && (3) \end{aligned}$$

where the two constants A and B (or C and ϕ) are determined from the initial conditions.

Many physical systems undergoing small displacements, including any objects obeying Hooke's law, exhibit simple harmonic motion. This equation arises, for example, in the analysis of the flow of current in an electronic CL circuit (which contains a capacitor and an inductor). If a damping force such as Friction is present, an additional term $\beta\dot{x}$ must be added to the DIFFERENTIAL EQUATION and motion dies out over time.

Adding a damping force proportional to $\dot{x}$, the first derivative of x with respect to time, the equation of motion for *damped* simple harmonic motion is

$$\ddot{x} + \beta\dot{x} + {\omega_0}^2 x = 0, \tag{4}$$

where β is the damping constant. This equation arises, for example, in the analysis of the flow of current in an electronic CLR circuit, (which contains a capacitor, an inductor, and a resistor). This ORDINARY DIFFERENTIAL EQUATION can be solved by looking for trial solutions of the form $x = e^{rt}$. Plugging this into (4) gives

$$(r^2 + \beta r + {\omega_0}^2)e^{rt} = 0 \tag{5}$$

$$r^2 + \beta r + {\omega_0}^2 = 0. \tag{6}$$

This is a QUADRATIC EQUATION with solutions

$$r = \tfrac{1}{2}(-\beta \pm \sqrt{\beta^2 - 4{\omega_0}^2}\,). \tag{7}$$

There are therefore three solution regimes depending on the SIGN of the quantity inside the SQUARE ROOT,

$$\alpha \equiv \beta^2 - 4{\omega_0}^2. \tag{8}$$

The three regimes are

1. $\alpha > 0$ is POSITIVE: overdamped,
2. $\alpha = 0$ is ZERO: critically damped,
3. $\alpha < 0$ is NEGATIVE: underdamped.

If a periodic (sinusoidal) forcing term is added at angular frequency ω, the same three solution regimes are again obtained. Surprisingly, the resulting motion is still periodic (after an initial transient response, corresponding to the solution to the unforced case, has died out), but it has an amplitude different from the forcing amplitude.

The "particular" solution $x_p(t)$ to the forced second-order nonhomogeneous ORDINARY DIFFERENTIAL EQUATION

$$\ddot{x} + p(t)\dot{x} + q(t)x = A\cos(\omega t) \tag{9}$$

due to forcing is given by the equation

$$x_p(t) = -x_1(t)\int \frac{x_2(t)g(t)}{W(t)}\,dt + x_2(t)\int \frac{x_1(t)g(t)}{W(t)}\,dt, \tag{10}$$

where x_1 and x_2 are the homogeneous solutions to the unforced equation

$$\ddot{x} + p(t)\dot{x} + q(t)x = 0 \tag{11}$$

and $W(t)$ is the WRONSKIAN of these two functions. Once the sinusoidal case of forcing is solved, it can be generalized to *any* periodic function by expressing the periodic function in a FOURIER SERIES.

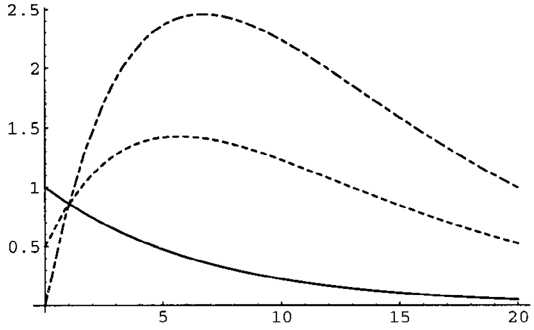

Critical damping is a special case of damped simple harmonic motion in which

$$\alpha \equiv \beta^2 - 4{\omega_0}^2 = 0, \tag{12}$$

so

$$\beta = 2\omega_0. \tag{13}$$

The above plot shows an underdamped simple harmonic oscillator with $\omega = 0.3$, $\beta = 0.15$. The solid curve is for $(A, B) = (1, 0)$, the dot-dashed for $(0, 1)$, and the dotted for $(1/2, 1/2)$. In this case, $\alpha = 0$ so the solutions of the form $x = e^{rt}$ satisfy

$$r_\pm = \tfrac{1}{2}(-\beta) = -\tfrac{1}{2}\beta = -\omega_0. \tag{14}$$

One of the solutions is therefore

$$x_1 = e^{-\omega_0 t}. \tag{15}$$

In order to find the other linearly independent solution, we can make use of the identity

$$x_2(t) = x_1(t)\int \frac{e^{-\int p(t)\,dt}}{[x_1(t)]^2}\,dt. \tag{16}$$

Since we have $p(t) = 2\omega_0$, $e^{-\int p(t)\,dt}$ simplifies to $e^{-2\omega_0 t}$. Equation (16) therefore becomes

$$x_2(t) = e^{-\omega_0 t}\int \frac{e^{-2\omega_0 t}}{[e^{-\omega_0 t}]^2}\,dt = e^{-\omega_0 t}\int dt = te^{-\omega_0 t}. \tag{17}$$

The general solution is therefore

$$x = (A + Bt)e^{-\omega_0 t}. \tag{18}$$

In terms of the constants A and B, the initial values are

$$x(0) = A \tag{19}$$

$$x'(0) = B - A\omega, \tag{20}$$

so

$$A = x(0) \tag{21}$$

$$B = x'(0) + \omega_0 x(0). \tag{22}$$

For sinusoidally forced simple harmonic motion with critical damping, the equation of motion is

$$\ddot{x} + 2\omega_0\dot{x} + {\omega_0}^2 x = A\cos(\omega t), \tag{23}$$

and the WRONSKIAN is

$$\begin{aligned} W(t) &\equiv x_1\dot{x}_2 - \dot{x}_1 x_2 \\ &= e^{-\omega_0 t}(e^{-\omega_0 t} - \omega_0 te^{-\omega_0 t}) + \omega_0 e^{-\omega_0 t}te^{-\omega_0 t} \\ &= e^{-2\omega_0 t}(1 - \omega_0 t + \omega_0 t) = e^{-2\omega_0 t}. \end{aligned} \tag{24}$$

Plugging this into the equation for the particular solution gives

$$\begin{aligned} x_p(t) &= -e^{-\omega_0 t} \int \frac{te^{-\omega_0 t} A\cos(\omega t)}{e^{-2\omega_0 t}}\,dt \\ &\quad + te^{-\omega_0 t} \int \frac{e^{-\omega_0 t} A\cos(\omega t)}{e^{-2\omega_0 t}}\,dt \\ &= Ae^{-\omega_0 t}\left[-\int te^{\omega_0 t}\cos(\omega t)\,dt + t\int e^{\omega_0 t}\cos(\omega t)\,dt\right] \\ &= Ae^{-\omega_0 t}\left\{-\frac{e^{\omega_0 t}}{(\omega^2+\omega_0{}^2)^2}[(\omega^2 + t\omega^2\omega_0 - \omega_0{}^2 + t\omega_0{}^3) \right. \\ &\quad \times \cos(\omega t) + \omega(t\omega^2 - 2\omega_0 + t\omega_0{}^2)\sin(\omega t)] \\ &\quad \left. + t\frac{e^{\omega_0 t}}{\omega^2+\omega_0{}^2}[\omega_0\cos(\omega t) + \omega\sin(\omega t)\right\} \\ &= \frac{A}{(\omega^2+\omega_0{}^2)^2}[(\omega_0{}^2 - \omega^2)\cos(\omega t) + 2\omega\omega_0\sin(\omega t)]. \end{aligned} \tag{25}$$

In order to put this in the desired form, note that we want to equate

$$\begin{aligned} C\cos\theta + S\sin\theta &= Q\cos(\theta+\delta) \\ &= Q(\cos\theta\cos\delta - \sin\theta\sin\delta). \end{aligned} \tag{26}$$

This means

$$C \equiv Q\cos\delta = \omega_0{}^2 - \omega^2 \tag{27}$$

$$S \equiv -Q\sin\delta = 2\omega\omega_0, \tag{28}$$

so

$$Q = \sqrt{C^2 + S^2} \tag{29}$$

$$\delta = \tan^{-1}\left(-\frac{S}{C}\right). \tag{30}$$

Plugging in,

$$\begin{aligned} Q &= \sqrt{\omega_0{}^4 - 2\omega_0{}^2\omega^2 + \omega^4 + 4\omega_0{}^2\omega^2} \\ &= \sqrt{\omega_0{}^4 + 2\omega_0{}^2\omega^2 + \omega^4} = \omega_0{}^2 + \omega^2. \end{aligned} \tag{31}$$

$$\delta = \tan^{-1}\left(\frac{2\omega\omega_0}{\omega^2 - \omega_0{}^2}\right). \tag{32}$$

The solution in the requested form is therefore

$$\begin{aligned} x_p &= \frac{A}{(\omega^2+\omega_0{}^2)^2}(\omega_0{}^2 + \omega^2)\cos(\omega t + \delta) \\ &= \frac{A}{\omega^2+\omega_0{}^2}\cos(\omega t+\delta), \end{aligned} \tag{33}$$

where δ is defined by (32).

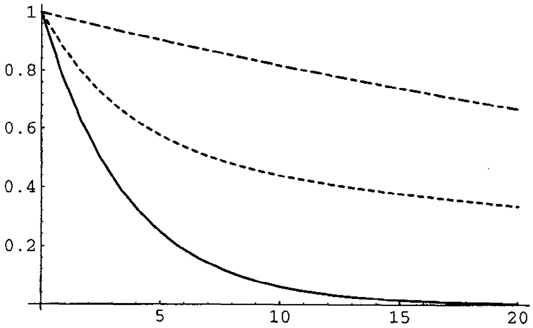

Overdamped simple harmonic motion occurs when

$$\beta^2 - 4\omega_0{}^2 > 0, \tag{34}$$

so

$$\alpha \equiv \beta^2 - 4\omega_0{}^2 > 0. \tag{35}$$

The above plot shows an overdamped simple harmonic oscillator with $\omega = 0.3$, $\beta = 0.075$. The solid curve is for $(A, B) = (1, 0)$, the dot-dashed for (0, 1), and the dotted for (1/2, 1/2). The solutions are

$$x_1 = e^{r_- t} \tag{36}$$

$$x_2 = e^{r_+ t}, \tag{37}$$

where

$$r_\pm \equiv \tfrac{1}{2}(-\beta \pm \sqrt{\beta^2 - 4\omega_0{}^2}\,). \tag{38}$$

The general solution is therefore

$$x = Ae^{r_- t} + Be^{r_+ t}, \tag{39}$$

where A and B are constants. The initial values are

$$x(0) = A + B \tag{40}$$

$$x'(0) = Ar_- + Br_+, \tag{41}$$

so

$$A = x(0) + \frac{r_+ x(0) - x'(0)}{r_- - r_+} \tag{42}$$

$$B = -\frac{r_+ x(0) - x'(0)}{r_- - r_+}. \tag{43}$$

For a cosinusoidally forced overdamped oscillator with forcing function $g(t) = C\cos(\omega t)$, the particular solutions are

$$y_1(t) = e^{r_1 t} \tag{44}$$

$$y_2(t) = e^{r_2 t}, \tag{45}$$

where

$$r_1 \equiv \tfrac{1}{2}(-\beta + \sqrt{\beta^2 - 4\omega_0{}^2}\,) \tag{46}$$

$$r_2 \equiv \tfrac{1}{2}(-\beta - \sqrt{\beta^2 - 4\omega_0{}^2}\,). \tag{47}$$

These give the identities

$$r_1 + r_2 = -\beta \tag{48}$$

$$r_1 - r_2 = \sqrt{\beta^2 - 4\omega_0{}^2} \tag{49}$$

and

$$\begin{aligned}\omega_0{}^2 &= \tfrac{1}{4}[\beta - (r_1 - r_2)^2] = \tfrac{1}{4}[(r_1 + r_2)^2 - (r_1 - r_2)^2] \\ &= \tfrac{1}{4}[2r_1r_2 + 2r_1r_2] = r_1r_2.\end{aligned} \tag{50}$$

The WRONSKIAN is

$$\begin{aligned}W(t) &\equiv y_1y_2' - y_1'y_2 = e^{r_1t}r_2e^{r_2t} - r_1e^{r_1t}e^{r_2t} \\ &= (r_2 - r_1)e^{(r_1+r_2)t}.\end{aligned} \tag{51}$$

The particular solution is

$$y_p = -y_1v_1 + y_2v_2, \tag{52}$$

where

$$v_1 \equiv \int \frac{y_2 g(t)}{W(t)} = \frac{C}{r_2 - r_1}\frac{\omega\sin(\omega t) - r_2\cos(\omega t)}{e^{r_2t}(r_2{}^2 + \omega^2)} \tag{53}$$

$$v_2 \equiv \int \frac{y_2 g(t)}{W(t)} = \frac{C}{r_2 - r_1}\frac{\omega\sin(\omega t) - r_1\cos(\omega t)}{e^{r_1t}(r_2{}^2 + \omega^2)}. \tag{54}$$

Therefore,

$$\begin{aligned}y_p &= C\frac{\cos(\omega t)(r_1r_2 - \omega^2) - \sin(\omega t)\omega(r_1 + r_2)}{(r_1{}^2 + \omega^2)(r_2{}^2 + \omega^2)} \\ &= C\frac{(\omega_0{}^2 - \omega^2)\cos(\omega t) + \beta\omega\sin(\omega t)}{\omega^2\beta^2 + (\omega^2 - \omega_0{}^2)} \\ &= \frac{C}{\omega^2\beta^2 + (\omega^2 - \omega_0{}^2)^2}\sqrt{(\omega^2 - \omega_0{}^2)^2 + \beta^2\omega^2} \\ &\quad \times \cos(\omega t + \delta) \\ &= \frac{C}{\sqrt{\beta^2\omega^2 + (\omega^2 - \omega_0{}^2)^2}}\cos(\omega t + \delta),\end{aligned} \tag{55}$$

where

$$\delta = \tan^{-1}\left(\frac{\beta\omega}{\omega^2 - \omega_0{}^2}\right). \tag{56}$$

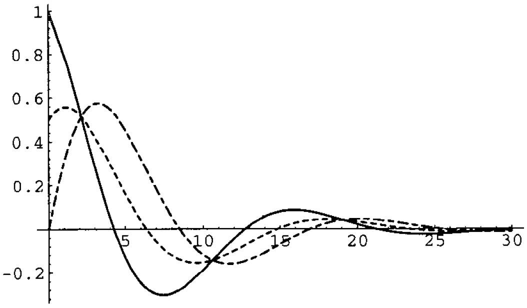

Underdamped simple harmonic motion occurs when

$$\beta^2 - 4\omega_0{}^2 < 0, \tag{57}$$

so

$$\alpha \equiv \beta^2 - 4\omega_0{}^2 < 0. \tag{58}$$

The above plot shows an underdamped simple harmonic oscillator with $\omega = 0.3$, $\beta = 0.4$. The solid curve is for $(A, B) = (1, 0)$, the dot-dashed for (0, 1), and the dotted for (1/2, 1/2). Define

$$\gamma \equiv \sqrt{-\alpha} = \tfrac{1}{2}\sqrt{4\omega_0{}^2 - \beta^2}, \tag{59}$$

then solutions satisfy

$$r_\pm = -\tfrac{1}{2}\beta \pm i\gamma, \tag{60}$$

where

$$r_\pm \equiv \tfrac{1}{2}(-\beta \pm \sqrt{\beta^2 - 4\omega_0{}^2}\,), \tag{61}$$

and are of the form

$$x = e^{-(\beta/2 \pm i\gamma)t}. \tag{62}$$

Using the EULER FORMULA

$$e^{ix} = \cos x + i\sin x, \tag{63}$$

this can be rewritten

$$x = e^{-(\beta/2)t}\left[\cos(\gamma t) \pm i\sin(\gamma t)\right]. \tag{64}$$

We are interested in the *real* solutions. Since we are dealing here with a *linear homogeneous* ODE, linear sums of LINEARLY INDEPENDENT solutions are also solutions. Since we have a sum of such solutions in (64), it follows that the IMAGINARY and REAL PARTS separately satisfy the ODE and are therefore the solutions we seek. The constant in front of the sine term is arbitrary, so we can identify the solutions as

$$x_1 = e^{-(\beta/2)t}\cos(\gamma t) \tag{65}$$

$$x_2 = e^{-(\beta/2)t}\sin(\gamma t), \tag{66}$$

so the general solution is

$$x = e^{-(\beta/2)t}[A\cos(\gamma t) + B\sin(\gamma t)]. \tag{67}$$

The initial values are

$$x(0) = A \tag{68}$$

$$x'(0) = -\tfrac{1}{2}\beta A + B,\gamma \tag{69}$$

so A and B can be expressed in terms of the initial conditions by

$$A = x(0) \tag{70}$$

$$B = \frac{\beta x(0)}{2\gamma} + \frac{x'(0)}{\gamma}. \tag{71}$$

For a cosinusoidally forced underdamped oscillator with forcing function $g(t) = C\cos(\omega t)$, use

$$\gamma \equiv \tfrac{1}{2}\sqrt{4\omega_0{}^2 - \beta^2} \tag{72}$$

$$\alpha \equiv \tfrac{1}{2}\beta \tag{73}$$

to obtain

$$4\omega_0{}^2 - \beta^2 = 4\gamma^2 \tag{74}$$

$$\omega_0{}^2 = \gamma^2 + \tfrac{1}{4}\beta^2 = \gamma^2 + \alpha^2 \tag{75}$$

$$\beta = 2\alpha. \tag{76}$$

The particular solutions are

$$y_1(t) = e^{-\alpha t}\cos(\gamma t) \tag{77}$$

$$y_2(t) = e^{-\alpha t}\sin(\gamma t). \tag{78}$$

The WRONSKIAN is

$$\begin{aligned} W(t) &\equiv y_1 y_2' - y_1' y_2 \\ &= e^{-\alpha t}\cos(\gamma t)[-\alpha e^{-\alpha t}\sin(\gamma t) + e^{-\alpha t}\gamma\cos(\gamma t)] \\ &\quad - e^{-\alpha t}\sin(\gamma t)[-\alpha e^{-\alpha t}\cos(\gamma t) - e^{-\alpha t}\gamma\sin(\gamma t)] \\ &= e^{-2\alpha t}\{\alpha[-\sin(\gamma t)\cos(\gamma t) + \sin(\gamma t)\cos(\gamma t)] \\ &\quad + \gamma[\cos^2(\gamma t) + \sin^2(\gamma t)]\} \\ &= \gamma e^{-2\alpha t}. \end{aligned} \tag{79}$$

The particular solution is given by

$$y_p = -y_1 v_1 + y_2 v_2, \tag{80}$$

where

$$v_1 \equiv \int \frac{y_2 g(t)}{W(t)} = \frac{C}{\gamma}\int e^{\alpha t}\cos(\gamma t)\cos(\omega t)\,dt \tag{81}$$

$$v_2 \equiv \int \frac{y_2 g(t)}{W(t)} = \frac{C}{\gamma}\int e^{\alpha t}\cos(\gamma t)\cos(\omega t)\,dt. \tag{82}$$

Using computer algebra to perform the algebra, the particular solution is

$$\begin{aligned} y_p(t) &= C\frac{(\alpha^2 + \gamma^2 - \omega^2)\cos(\omega t) + 2\alpha\omega\sin(\omega t)}{[\alpha^2 + (\gamma - \omega)^2][\alpha^2 + (\gamma + \omega)^2]} \\ &= C\frac{(\omega_0{}^2 - \omega^2)\cos(\omega t) + \beta\omega\sin(\omega t)}{(\alpha^2 + \gamma^2 + \omega^2)^2 - 4\gamma^2\omega^2} \\ &= C\frac{(\omega_0{}^2 - \omega^2)\cos(\omega t) + \beta\omega\sin(\omega t)}{(\omega_0{}^2 + \omega^2)^2 - 4\frac{1}{4}(4\omega_0{}^2 - \beta^2)\omega^2} \\ &= C\frac{(\omega_0{}^2 - \omega^2)\cos(\omega t) + \beta\omega\sin(\omega t)}{(\omega_0{}^2 - \omega^2)^2 - \omega^2(4\omega_0{}^2 - \beta^2)} \\ &= \frac{C\sqrt{(\omega_0{}^2 - \omega^2)^2 + \beta^2\omega^2}}{(\omega_0{}^2 - \omega^2)^2 - \omega^2(4\omega_0{}^2 - \beta^2)}\cos(\omega t + \delta) \\ &= C\frac{\sqrt{(\omega_0{}^2 - \omega^2)^2 + \beta^2\omega^2}}{(\omega_0{}^2 - \omega^2)^2 - \omega^2(4\omega_0{}^2 - \beta^2)}\cos(\omega t + \delta), \end{aligned} \tag{83}$$

where

$$\delta = \tan^{-1}\left(\frac{\beta\omega}{\omega^2 - \omega_0{}^2}\right). \tag{84}$$

If the forcing function is sinusoidal instead of cosinusoidal, then

$$\delta' = \delta - \tfrac{1}{2}\pi = \tan^{-1} x - \tfrac{1}{2}\pi = \tan^{-1}\left(-\frac{1}{x}\right), \tag{85}$$

so

$$\delta' = \tan^{-1}\left(\frac{\omega_0{}^2 - \omega^2}{\beta\omega}\right). \tag{86}$$

Simple Harmonic Motion Quadratic Perturbation

Given a simple harmonic oscillator with a quadratic perturbation ϵx^2,

$$\ddot{x} + \omega_0{}^2 x - \alpha\epsilon x^2 = 0, \tag{1}$$

find the first-order solution using a perturbation method. Write

$$x \equiv x_0 + \epsilon x_1 + \ldots, \tag{2}$$

so

$$\ddot{x} = \ddot{x}_0 + \epsilon\ddot{x}_1 + \ldots. \tag{3}$$

Plugging (2) and (3) back into (1) gives

$$(\ddot{x}_0 + \epsilon\ddot{x}_1) + (\omega_0{}^2 x_0 + \omega_0{}^2\epsilon x_1) - \alpha\epsilon(x_0 + 2x_0x_1\epsilon + \ldots) = 0. \tag{4}$$

Keeping only terms of order ϵ and lower and grouping, we obtain

$$(\ddot{x}_0 + \omega_0{}^2 x_0) + (\ddot{x}_1 + \omega_0{}^2 x_1 - \alpha x_0{}^2)\epsilon = 0. \tag{5}$$

Since this equation must hold for all POWERS of ϵ, we can separate it into the two differential equations

$$\ddot{x}_0 + \omega_0{}^2 x_0 = 0 \tag{6}$$

$$\ddot{x}_1 + \omega_0{}^2 x_1 = \alpha x_0{}^2. \tag{7}$$

The solution to (6) is just

$$x_0 = A\cos(\omega_0 t + \phi). \tag{8}$$

Setting our clock so that $\phi = 0$ gives

$$x_0 = A\cos(\omega_0 t). \tag{9}$$

Plugging this into (7) then gives

$$\ddot{x}_1 + \omega_0{}^2 x_1 = \alpha A^2\cos^2(\omega_0 t). \tag{10}$$

The two homogeneous solutions to (10) are

$$x_1 = \cos(\omega_0 t) \tag{11}$$

$$x_2 = \sin(\omega_0 t). \tag{12}$$

The particular solution to (10) is therefore given by

$$x_p(t) = -x_1(t)\int \frac{x_2(t)g(t)}{W(t)}\,dt + x_2(t)\int \frac{x_1(t)g(t)}{W(t)}\,dt, \tag{13}$$

where

$$g(t) = \alpha A^2 \cos^2(\omega_0 t), \tag{14}$$

and the WRONSKIAN is

$$\begin{aligned} W &\equiv x_1\dot{x}_2 - \dot{x}_1 x_2 \\ &= \cos(\omega_0 t)\omega_0\cos(\omega_0 t) - [-\omega_0\sin(\omega_0 t)]\sin(\omega_0 t) \\ &= \omega_0. \end{aligned} \tag{15}$$

Plugging everything into (13),

$$\begin{aligned} x_p &= \alpha A^2\left[-\cos(\omega_0 t)\int \frac{\sin(\omega_0 t)\cos^2(\omega_0 t)}{\omega_0}\,dt \right. \\ &\quad \left. + \sin(\omega_0 t)\int \frac{\cos^3(\omega_0 t)}{\omega_0}\,dt\right] \\ &= \frac{\alpha A^2}{\omega_0}\left\{\sin(\omega_0 t)\int [1-\sin^2(\omega_0 t)]\cos(\omega_0 t)\,dt \right. \\ &\quad \left. - \cos(\omega_0 t)\int \sin(\omega_0 t)\cos^2(\omega_0 t)\,dt\right\}. \end{aligned} \tag{16}$$

Now let

$$u \equiv \sin(\omega_0 t) \tag{17}$$

$$du = \omega_0\cos(\omega_0 t)\,dt \tag{18}$$

$$v \equiv \cos(\omega_0 t) \tag{19}$$

$$dv = -\omega_0\sin(\omega_0 t)\,dt. \tag{20}$$

Then

$$\begin{aligned} x_p &= \frac{\alpha A^2}{{\omega_0}^2}\left[\sin(\omega_0 t)\int (1-u^2)\,du + \cos(\omega_0 t)\int v^2\,dv\right] \\ &= \frac{\alpha A^2}{{\omega_0}^2}\left[\sin(\omega_0 t)(1-\tfrac{1}{3}u^3) + \cos(\omega_0 t)\tfrac{1}{3}v^3\right] \\ &= \frac{\alpha A^2}{{\omega_0}^2}\{\sin(\omega_0 t)[1-\tfrac{1}{3}\sin^3(\omega_0 t)] \\ &\quad + \tfrac{1}{3}\cos(\omega_0 t)\cos^3(\omega_0 t)\} \\ &= \frac{\alpha A^2}{{\omega_0}^2}\left\{\tfrac{1}{3}[\cos^4(\omega_0 t) - \sin^4(\omega_0 t)] + \sin^2(\omega_0 t)\right\} \\ &= \frac{\alpha A^2}{{\omega_0}^2}\left\{\tfrac{1}{3}[\cos^2(\omega_0 t) - \sin^2(\omega_0 t)] + \sin^2(\omega_0 t)\right\} \\ &= \frac{\alpha A^2}{{\omega_0}^2}\tfrac{1}{3}[\cos^2(\omega_0 t) + 2\sin^2(\omega_0 t)] \\ &= \frac{\alpha A^2}{3{\omega_0}^2}[2-\cos^2(\omega_0 t)] = \frac{\alpha A^2}{3{\omega_0}^2}\{2-\tfrac{1}{2}[1+\cos(2\omega_0 t)]\} \\ &= \frac{\alpha A^2}{6{\omega_0}^2}[3-\cos(2\omega_0 t)]. \end{aligned} \tag{21}$$

Plugging $x_0(t)$ and (21) into (2), we obtain the solution

$$x(t) = A\cos(\omega_0 t) - \frac{\alpha A^2}{6{\omega_0}^2}\epsilon[\cos(2\omega_0 t) - 3]. \tag{22}$$

Simple Harmonic Oscillator

see SIMPLE HARMONIC MOTION

Simple Interest

INTEREST which is paid only on the PRINCIPAL and not on the additional amount generated by previous INTEREST payments. A formula for computing simple interest is

$$a(t) = a(0)(1 + rt),$$

where $a(t)$ is the sum of PRINCIPAL and INTEREST at time t for a constant interest rate r.

see also COMPOUND INTEREST, INTEREST

References

Kellison, S. G. *Theory of Interest, 2nd ed.* Burr Ridge, IL: Richard D. Irwin, 1991.

Simple Polygon

A POLYGON P is said to be simple (or JORDAN) if the only points of the plane belonging to two EDGES of P are the VERTICES of P. Such a polygon has a well-defined interior and exterior.

see also POLYGON, REGULAR POLYGON, TWO-EARS THEOREM

References

Toussaint, G. "Anthropomorphic Polygons." *Amer. Math. Monthly* **122**, 31–35, 1991.

Simple Ring

A NONZERO RING S whose only (two-sided) IDEALS are S itself and zero. Every commutative simple ring is a FIELD. Every simple ring is a PRIME RING.

see also FIELD, IDEAL, PRIME RING, RING

Simplex

The generalization of a tetrahedral region of space to n-D. The boundary of a k-simplex has $k+1$ 0-faces (VERTICES), $k(k+1)/2$ 1-faces (EDGES), and $\binom{k+1}{i+1}$ i-faces, where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT.

The simplex in 4-D is a regular TETRAHEDRON $ABCD$ in which a point E along the fourth dimension through the center of $ABCD$ is chosen so that $EA = EB = EC = ED = AB$. The 4-D simplex has SCHLÄFLI SYMBOL $\{3,3,3\}$.

n	Simplex
0	point
1	line segment
2	equilateral triangular plane region
3	tetrahedral region
4	4-simplex

The only irreducible spherical simplexes generated by reflection are A_n ($n \geq 1$), B_n ($n \geq 4$), C_n ($n \geq 2$), D_2^p ($p \geq 5$), E_6, E_7, E_8, F_4, G_3, and G_4. The only irreducible Euclidean simplexes generated by reflection are W_2, P_m $(m \geq 3)$, Q_m $(m \geq 5)$, R_m $(m \geq 3)$, S_m $(m \geq 4)$, V_3, T_7, T_8, T_9, and U_5.

The regular simplex in n-D with $n \geq 5$ is denoted α_n and has Schläfli Symbol $\{\underbrace{3,\ldots,3}_{3^{n-1}}\}$.

see also Complex, Cross Polytope, Equilateral Triangle, Line Segment, Measure Polytope, Nerve, Point, Simplex Method, Tetrahedron

References

Eppstein, D. "Triangles and Simplices." `http://www.ics.uci.edu/~eppstein/junkyard/triangulation.html`.

Simplex Method

A method for solving problems in Linear Programming. This method, invented by G. B. Dantzig in 1947, runs along Edges of the visualization Solid to find the best answer. In 1970, Klee and Minty constructed examples in which the simplex method required an exponential number of steps, but such cases seem never to be encountered in practical applications.

A much more efficient (Polynomial-time) Algorithm was found in 1984 by N. Karmarkar. This method goes through the middle of the Solid and then transforms and warps. It offers many advantages over the simplex method (Nemirovsky and Yudin 1994).

see also Linear Programming

References

Nemirovsky, A. and Yudin, N. *Interior-Point Polynomial Methods in Convex Programming.* Philadelphia, PA: SIAM, 1994.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Downhill Simplex Method in Multidimensions" and "Linear Programming and the Simplex Method." §10.4 and 10.8 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 402–406 and 423–436, 1992.

Tokhomirov, V. M. "The Evolution of Methods of Convex Optimization." *Amer. Math. Monthly* **103**, 65–71, 1996.

Simplicial Complex

A simplicial complex is a Space with a Triangulation. Objects in the space made up of only the simplices in the triangulation of the space are called simplicial subcomplexes. When only simplicial complexes and subcomplexes are considered, defining Homology is particularly easy (and, in fact, combinatorial because of its finite/counting nature). This kind of homology is called Simplicial Homology.

see also Homology (Topology), Nerve, Simplicial Homology, Space, Triangulation

Simplicial Homology

The type of Homology which results when the spaces being studied are restricted to Simplicial Complexes and subcomplexes.

see also Simplicial Complex

Simplicity

The number of operations needed to effect a Geometric Construction as determined in Geometrography. If the number of operations of the five Geometrographic types are denoted m_1, m_2, n_1, n_2, and n_3, respectively, then the simplicity is $m_1+m_2+n_1+n_2+n_3$ and the symbol $m_1S_1 + m_2S_2 + n_1C_1 + n_2C_2 + n_3C_3$. It is apparently an unsolved problem to determine if a given Geometric Construction is of smallest possible simplicity.

see also Geometric Construction, Geometrography

References

De Temple, D. W. "Carlyle Circles and the Lemoine Simplicity of Polygonal Constructions." *Amer. Math. Monthly* **98**, 97–108, 1991.

Eves, H. *An Introduction to the History of Mathematics, 6th ed.* New York: Holt, Rinehart, and Winston, 1976.

Simply Connected

A Connected Domain is said to be simply connected (also called 1-connected) if any simple closed curve can be shrunk to a point continuously in the set. If the domain is Connected but not simply, it is said to be Multiply Connected.

A Space S is simply connected if it is 0-connected and if every Map from the 1-Sphere to S extends continuously to a Map from the 2-Disk. In other words, every loop in the Space is contractible.

see also Connected Space, Multiply Connected

Simpson's Paradox

It is not necessarily true that averaging the averages of different populations gives the average of the combined population.

References

Paulos, J. A. *A Mathematician Reads the Newspaper.* New York: BasicBooks, p. 135, 1995.

Simpson's Rule

Let $h \equiv (b-a)/n$, and assume a function $f(x)$ is defined at points $f(a+kh) = y_k$ for $k = 0, \ldots, n$. Then

$$\int_a^b f(x)\,dx = \tfrac{1}{3}h(y_1 + 4y_2 + 2y_3 + 4y_4 + \ldots + 2y_{n-2} + 4y_{n-1} + y_n) - R_n,$$

where the remainder is

$$R_n = \tfrac{1}{90}(b-a)^4 f^{(4)}(x^*)$$

for some $x^* \in [a,b]$.

see also Bode's Rule, Newton-Cotes Formulas, Simpson's 3/8 Rule, Trapezoidal Rule

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 886, 1972.

Simpson's 3/8 Rule

$$\int_{x_1}^{x_4} f(x)\,dx = \tfrac{3}{8}h(f_1+3f_2+3f_3+f_4) - \tfrac{3}{80}h^5 f^{(4)}(\xi).$$

see also BODE'S RULE, NEWTON-COTES FORMULAS, SIMPSON'S RULE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 886, 1972.

Simson Line

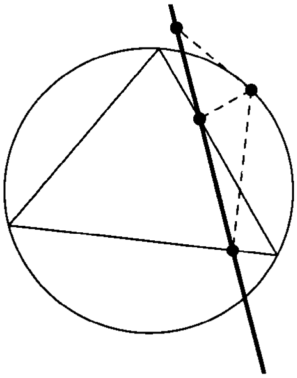

The Simson line is the LINE containing the feet of the perpendiculars from a point on the CIRCUMCIRCLE of a TRIANGLE to the sides (or their extensions) of the TRIANGLE. The Simson line is sometimes known as the WALLACE-SIMSON LINE, since it does not appear in any work of Simson (Johnson 1929, p. 137).

The ANGLE between the Simson lines of two points P and P' is half the ANGLE of the arc PP'. The Simson line of any VERTEX is the ALTITUDE through that VERTEX. The Simson line of a point opposite a VERTEX is the corresponding side. If $T_1T_2T_3$ is the Simson line of a point T of the CIRCUMCIRCLE, then the triangles TT_1T_2 and TA_2A_1 are directly similar.

see also CIRCUMCIRCLE

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 40–41 and 43–45, 1967.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 137–139, 1929.

Sinc Function

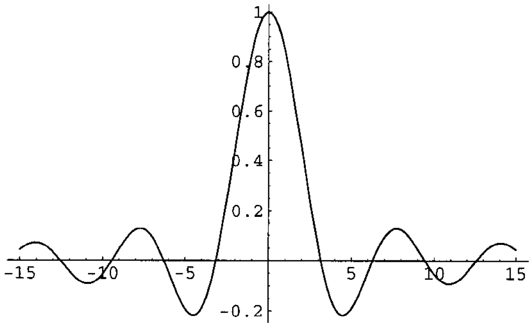

A function also called the SAMPLING FUNCTION and defined by

$$\operatorname{sinc}(x) \equiv \begin{cases} 1 & \text{for } x = 0 \\ \frac{\sin x}{x} & \text{otherwise,} \end{cases} \tag{1}$$

where $\sin x$ is the SINE function. Let $\Pi(x)$ be the RECTANGLE FUNCTION, then the FOURIER TRANSFORM of $\Pi(x)$ is the sinc function

$$\mathcal{F}[\Pi(x)] = \operatorname{sinc}(\pi k). \tag{2}$$

The sinc function therefore frequently arises in physical applications such as Fourier transform spectroscopy as the so-called INSTRUMENT FUNCTION, which gives the instrumental response to a DELTA FUNCTION input. Removing the instrument functions from the final spectrum requires use of some sort of DECONVOLUTION algorithm.

The sinc function can be written as a complex INTEGRAL by noting that

$$\begin{aligned} \operatorname{sinc}(nx) &\equiv \frac{\sin(nx)}{nx} = \frac{1}{nx}\frac{e^{inx}-e^{-inx}}{2i} \\ &= \frac{1}{2inx}[e^{itx}]_{-n}^{n} = \frac{1}{2n}\int_{-n}^{n} e^{ixt}\,dt. \end{aligned} \tag{3}$$

The sinc function can also be written as the INFINITE PRODUCT

$$\frac{\sin x}{x} = \prod_{k=1}^{\infty} \cos\left(\frac{x}{2^k}\right). \tag{4}$$

Definite integrals involving the sinc function include

$$\int_0^\infty \operatorname{sinc}(x)\,dx = \tfrac{1}{2}\pi \tag{5}$$

$$\int_0^\infty \operatorname{sinc}^2(x)\,dx = \tfrac{1}{2}\pi \tag{6}$$

$$\int_0^\infty \operatorname{sinc}^3(x)\,dx = \tfrac{3}{8}\pi \tag{7}$$

$$\int_0^\infty \operatorname{sinc}^4(x)\,dx = \tfrac{1}{3}\pi \tag{8}$$

$$\int_0^\infty \operatorname{sinc}^5(x)\,dx = \tfrac{115}{384}\pi. \tag{9}$$

These are all special cases of the amazing general result

$$\int_0^\infty \frac{\sin^a x}{x^b}\,dx = \frac{\pi^{1-c}(-1)^{\lfloor (a-b)/2\rfloor}}{2^{a-c}(b-1)!} \times \sum_{k=0}^{\lfloor a/2\rfloor - c} (-1)^k \binom{a}{k} (a-2k)^{b-1}[\ln(a-2k)]^c, \quad (10)$$

where a and b are POSITIVE integers such that $a \geq b > c$, $c \equiv a-b \pmod 2$, $\lfloor x \rfloor$ is the FLOOR FUNCTION, and 0^0 is taken to be equal to 1 (Kogan). This spectacular formula simplifies in the special case when n is a POSITIVE EVEN integer to

$$\int_0^\infty \frac{\sin^{2n} x}{x^{2n}}\,dx = \frac{\pi}{2(2n-1)!}\left\langle {2n-1 \atop n-1} \right\rangle, \quad (11)$$

where $\left\langle {n \atop k} \right\rangle$ is an EULERIAN NUMBER (Kogan). The solution of the integral can also be written in terms of the RECURRENCE RELATION for the coefficients

$$c(a,b) = \begin{cases} \frac{\pi}{2^{a+1-b}} \binom{a-1}{\frac{1}{2}(a-1)} & \text{for } b = 1 \text{ or } b = 2 \\ \frac{a}{(b-1)(b-2)}[(a-1)c(a-2,b-2) \\ \quad -a \cdot c(a,b-2)] & \text{otherwise} \end{cases} \quad (12)$$

(Zimmerman).

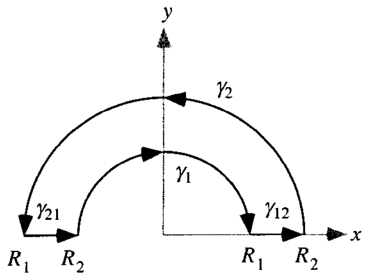

The half-infinite integral of $\operatorname{sinc}(x)$ can be derived using CONTOUR INTEGRATION. In the above figure, consider the path $\gamma \equiv \gamma_1 + \gamma_{12} + \gamma_2 + \gamma_{21}$. Now write $z = Re^{i\theta}$. On an arc, $dz = iRe^{i\theta}\,d\theta$ and on the x-AXIS, $dz = e^{i\theta}\,dR$. Write

$$\int_{-\infty}^\infty \frac{\sin x}{x}\,dx = \Im \int_\gamma \frac{e^{iz}}{z}\,dx, \quad (13)$$

where $\Im$ denotes the IMAGINARY POINT. Now define

$$\begin{aligned} I &\equiv \int_\gamma \frac{e^{iz}}{z}\,dz \\ &= \lim_{R_1\to 0} \int_\pi^0 \frac{\exp(iR_1 e^{i\theta})}{R_1 e^{i\theta}}\, i\theta R_1 e^{i\theta}\,d\theta \\ &\quad + \lim_{R_1\to 0}\lim_{R_2\to\infty} \int_{R_1}^{R_2} \frac{e^{iR}}{R}\,dR \\ &\quad + \lim_{R_2\to\infty} \int_0^\pi \frac{\exp(iz)}{z}\,dx + \lim_{R_1\to 0} \int_{R_2}^{R_1} \frac{e^{-iR}}{-R}\,(-dR), \end{aligned} \quad (14)$$

where the second and fourth terms use the identities $e^{i0} = 1$ and $e^{i\pi} = -1$. Simplifying,

$$\begin{aligned} I &= \lim_{R_1\to 0} \int_\pi^0 \exp(iR_1 e^{i\theta}) i\theta\,d\theta + \int_{0+}^\infty \frac{e^{iR}}{R}\,dR \\ &\quad + \lim_{R_2\to\infty} \int_0^\pi \frac{\exp(iz)}{z}\,dz + \int_\infty^{0^+} \frac{e^{-iR}}{-R}\,(-dR) \\ &= -\int_0^\pi i\theta\,d\theta + \int_{0+}^\infty \frac{e^{iR}}{R}\,dR + 0 + \int_{-\infty}^{0^-} \frac{e^{iR}}{R}\,dR, \end{aligned} \quad (15)$$

where the third term vanishes by JORDAN'S LEMMA. Performing the integration of the first term and combining the others yield

$$I = -i\pi + \int_{-\infty}^\infty \frac{e^{iz}}{z}\,dz = 0. \quad (16)$$

Rearranging gives

$$\int_{-\infty}^\infty \frac{e^{iz}}{z}\,dz = i\pi, \quad (17)$$

so

$$\int_{-\infty}^\infty \frac{\sin z}{z}\,dz = \pi. \quad (18)$$

The same result is arrived at using the method of RESIDUES by noting

$$\begin{aligned} I &= 0 + \tfrac{1}{2} 2\pi i \operatorname{Res}[f(z)]_{z=0} \\ &= i\pi \left[(z-0)\frac{e^{iz}}{z}\right]_{z=0} = i\pi[e^{iz}]_{z=0} \\ &= i\pi, \end{aligned} \quad (19)$$

so

$$\Im(I) = \pi. \quad (20)$$

Since the integrand is symmetric, we therefore have

$$\int_0^\infty \frac{\sin x}{x}\,dx = \tfrac{1}{2}\pi, \quad (21)$$

giving the SINE INTEGRAL evaluated at 0 as

$$\operatorname{si}(0) = -\int_0^\infty \frac{\sin x}{x}\,dx = -\tfrac{1}{2}\pi. \quad (22)$$

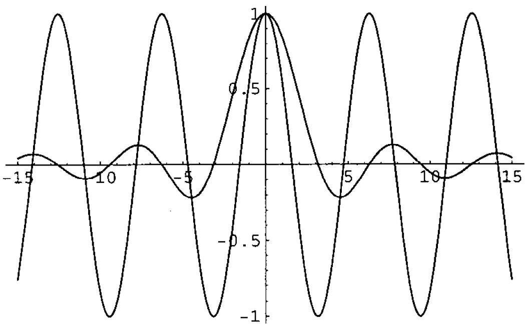

An interesting property of $\operatorname{sinc}(x)$ is that the set of LOCAL EXTREMA of $\operatorname{sinc}(x)$ corresponds to its intersections with the COSINE function $\cos(x)$, as illustrated above.

see also FOURIER TRANSFORM, FOURIER TRANSFORM—RECTANGLE FUNCTION, INSTRUMENT FUNCTION, JINC FUNCTION, SINE, SINE INTEGRAL

References

Kogan, S. "A Note on Definite Integrals Involving Trigonometric Functions." `http://www.mathsoft.com/asolve/constant/pi/sin/sin.html`.

Morrison, K. E. "Cosine Products, Fourier Transforms, and Random Sums." *Amer. Math. Monthly* **102**, 716–724, 1995.

Sinclair's Soap Film Problem

Find the shape of a soap film (i.e., MINIMAL SURFACE) which will fill two inverted conical FUNNELS facing each other is known as Sinclair's soap film problem (Bliss 1925, p. 121). The soap film will assume the shape of a CATENOID.

see also CATENOID, FUNNEL, MINIMAL SURFACE

References

Bliss, G. A. *Calculus of Variations.* Chicago, IL: Open Court, pp. 121–122, 1925.

Isenberg, C. *The Science of Soap Films and Soap Bubbles.* New York: Dover, p. 81, 1992.

Sinclair, M. E. "On the Minimum Surface of Revolution in the Case of One Variable End Point." *Ann. Math.* **8**, 177–188, 1907.

Sine

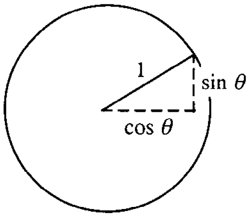

Let θ be an ANGLE measured counterclockwise from the x-AXIS along the arc of the UNIT CIRCLE. Then $\sin\theta$ is the vertical coordinate of the arc endpoint. As a result of this definition, the sine function is periodic with period 2π. By the PYTHAGOREAN THEOREM, $\sin\theta$ also obeys the identity

$$\sin^2\theta + \cos^2\theta = 1. \tag{1}$$

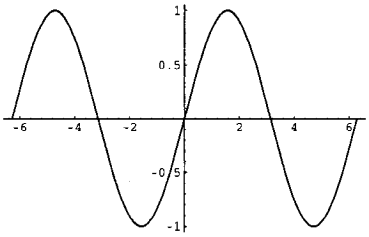

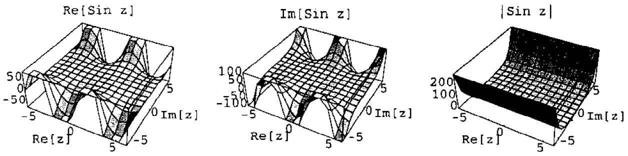

The sine function can be defined algebraically by the infinite sum

$$\sin x = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)!} x^{2n-1} \tag{2}$$

and INFINITE PRODUCT

$$\sin x = x \prod_{n=1}^{\infty} \left(1 - \frac{x^2}{n^2\pi^2}\right). \tag{3}$$

It is also given by the IMAGINARY PART of the complex exponential

$$\sin x = \Im[e^{ix}]. \tag{4}$$

The multiplicative inverse of the sine function is the COSECANT, defined as

$$\csc x \equiv \frac{1}{\sin x}. \tag{5}$$

Using the results from the EXPONENTIAL SUM FORMULAS

$$\begin{aligned} \sum_{n=0}^{\infty} \sin(nx) &= \Im\left[\sum_{n=0}^{\infty} e^{inx}\right] \\ &= \Im\left[\frac{\sin(\frac{1}{2}Nx)}{\sin(\frac{1}{2}x)} e^{i(N-1)x/2}\right] \\ &= \frac{\sin(\frac{1}{2}Nx)}{\sin(\frac{1}{2}x)} \sin[\tfrac{1}{2}x(N-1)]. \end{aligned} \tag{6}$$

Similarly,

$$\begin{aligned} \sum_{n=0}^{\infty} p^n \sin(nx) &= \Im\left[\sum_{n=0}^{\infty} p^n e^{inx}\right] \\ &= \Im\left[\frac{1-pe^{-ix}}{1-2p\cos x + p^2}\right] = \frac{p\sin x}{1-2p\cos x+p^2}. \end{aligned} \tag{7}$$

Other identities include

$$\sin(n\theta) = 2\cos\theta\sin[(n-1)\theta] - \sin[(n-2)\theta] \tag{8}$$

$$\begin{aligned} \sin(nx) = \binom{n}{1}\cos^{n-1}x\sin x - \binom{n}{3}\cos^{n-3}x\sin^3 x \\ + \binom{n}{5}\cos^{n-5}x\sin^5 x - \ldots, \end{aligned} \tag{9}$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT.

Cvijović and Klinowski (1995) show that the sum

$$S_\nu(\alpha) = \sum_{k=0}^{\infty} \frac{\sin(2k+1)\alpha}{(2k+1)^\nu} \tag{10}$$

has closed form for $\nu = 2n+1$,

$$S_{2n+1}(\alpha) = \frac{(-1)^n}{4(2n)!}\pi^{2n+1} E_{2n}\left(\frac{\alpha}{\pi}\right), \tag{11}$$

where $E_n(x)$ is an EULER POLYNOMIAL.

A CONTINUED FRACTION representation of $\sin x$ is

$$\sin x = \cfrac{x}{1 + \cfrac{x^2}{(2\cdot 3 - x^2) + \cfrac{2\cdot 3x^2}{(4\cdot 5 - x^2) + \cfrac{4\cdot 5x^2}{(6\cdot 7 - x^2) + \ldots}}}}. \tag{12}$$

The value of $\sin(2\pi/n)$ is IRRATIONAL for all n except 4 and 12, for which $\sin(\pi/2) = 1$ and $\sin(\pi/6) = 1/2$.

The FOURIER TRANSFORM of $\sin(2\pi k_0 x)$ is given by

$$\begin{aligned}\mathcal{F}[\sin(2\pi k_0 x)] &= \int_{-\infty}^{\infty} e^{-2\pi i k_0 x} \sin(2\pi k_0 x)\, dx \\ &= \tfrac{1}{2} i[\delta(k+k_0) - \delta(k-k_0)]. \end{aligned} \tag{13}$$

Definite integrals involving $\sin x$ include

$$\int_0^\infty \sin(x^2)\, dx = \tfrac{1}{4}\sqrt{2\pi} \tag{14}$$

$$\int_0^\infty \sin(x^3)\, dx = \tfrac{1}{6}\Gamma(\tfrac{1}{3}) \tag{15}$$

$$\int_0^\infty \sin(x^4)\, dx = -\cos(\tfrac{5}{8}\pi)\Gamma(\tfrac{5}{4}) \tag{16}$$

$$\int_0^\infty \sin(x^5)\, dx = \tfrac{1}{4}(\sqrt{5}-1)\Gamma(\tfrac{6}{5}), \tag{17}$$

where $\Gamma(x)$ is the GAMMA FUNCTION.

see also ANDREW'S SINE, COSECANT, COSINE, FOURIER TRANSFORM—SINE, HYPERBOLIC SINE, SINC FUNCTION, TANGENT, TRIGONOMETRY

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Circular Functions." §4.3 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 71–79, 1972.

Cvijović, D. and Klinowski, J. "Closed-Form Summation of Some Trigonometric Series." *Math. Comput.* **64**, 205–210, 1995.

Hansen, E. R. *A Table of Series and Products.* Englewood Cliffs, NJ: Prentice-Hall, 1975.

Project Mathematics! *Sines and Cosines, Parts I-III.* Videotapes (28, 30, and 30 minutes). California Institute of Technology. Available from the Math. Assoc. Amer.

Spanier, J. and Oldham, K. B. "The Sine sin(x) and Cosine cos(x) Functions." Ch. 32 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 295–310, 1987.

Sine-Gordon Equation

A PARTIAL DIFFERENTIAL EQUATION which appears in differential geometry and relativistic field theory. Its name is a pun on its similar form to the KLEIN-GORDON EQUATION. The sine-Gordon equation is

$$v_{tt} - v_{xx} + \sin v = 0, \tag{1}$$

where v_{tt} and v_{xx} are PARTIAL DERIVATIVES. The equation can be transformed by defining

$$\xi \equiv \tfrac{1}{2}(x-t) \tag{2}$$

$$\eta \equiv \tfrac{1}{2}(x+t), \tag{3}$$

giving

$$v_{\xi\eta} = \sin v. \tag{4}$$

Traveling wave analysis gives

$$z - z_0 = \sqrt{c^2-1} \int \frac{df}{\sqrt{2[d - 2\sin^2(\frac{1}{2}f)]}}. \tag{5}$$

For $d = 0$,

$$z - z_0 = \pm\sqrt{1-c^2}\, \ln[\pm\tan(\tfrac{1}{4}f)] \tag{6}$$

$$f(z) = \pm 4\tan^{-1}[e^{\pm(z-z_0)/(1-c^2)^{1/2}}]. \tag{7}$$

Letting $z \equiv \xi\eta$ then gives

$$zf'' + f' = \sin f. \tag{8}$$

Letting $g \equiv e^{if}$ gives

$$g'' - \frac{g'^2}{f} + \frac{2g' - g^2 + 1}{2z} = 0, \tag{9}$$

which is the third PAINLEVÉ TRANSCENDENT. Look for a solution of the form

$$v(x,t) = 4\tan^{-1}\left[\frac{\phi(x)}{\psi(t)}\right]. \tag{10}$$

Taking the partial derivatives gives

$$\phi_{xx} = -k^2\phi^4 + m^2\phi^2 + n^2 \tag{11}$$

$$\psi_{tt} = k^2\psi^4 + (m^2-1)\psi^2 - n^2, \tag{12}$$

which can be solved in terms of ELLIPTIC FUNCTIONS. A single SOLITON solution exists with $k = n = 0$, $m > 1$:

$$v = 4\tan^{-1}\left[\exp\left(\frac{\pm x - \beta t}{\sqrt{1-\beta^2}}\right)\right], \tag{13}$$

where

$$\beta \equiv \frac{\sqrt{m^2-1}}{m}. \tag{14}$$

A two-Soliton solution exists with $k = 0$, $m > 1$:

$$v = 4\tan^{-1}\left[\frac{\sinh(\beta m x)}{\beta\cosh(\beta m t)}\right]. \tag{15}$$

A Soliton-antisoliton solution exists with $k \neq 0$, $n = 0$, $m^2 > 1$:

$$v = -4\tan^{-1}\left[\frac{\sinh(\beta m x)}{\beta\cosh(m t)}\right]. \tag{16}$$

A "breather" solution is

$$v = -4\tan^{-1}\left[\frac{m}{\sqrt{1-m^2}}\frac{\sin(\sqrt{1-m^2}\,t)}{\cosh(m x)}\right]. \tag{17}$$

References

Infeld, E. and Rowlands, G. *Nonlinear Waves, Solitons, and Chaos.* Cambridge, England: Cambridge University Press, pp. 199–200, 1990.

Sine Integral

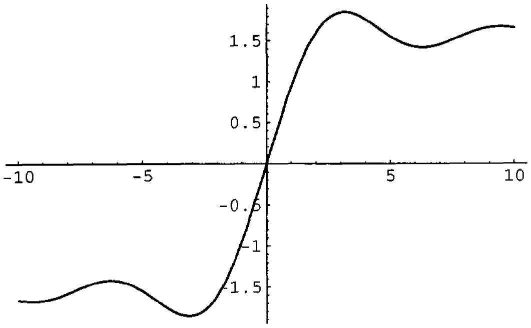

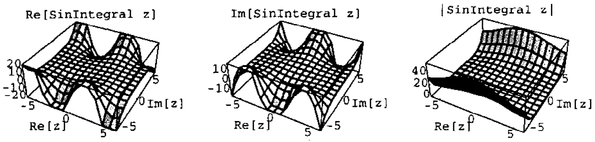

There are two types of "sine integrals" commonly defined,

$$\mathrm{Si}(x) \equiv \int_0^z \frac{\sin t}{t}\,dt \tag{1}$$

and

$$\begin{aligned}\mathrm{si}(x) &\equiv -\int_x^\infty \frac{\sin t}{t}\,dt && (2)\\ &= \frac{1}{2i}[\mathrm{ei}(ix) - \mathrm{ei}(-ix)]\\ &= \frac{1}{2i}[\mathrm{e}_1(ix) - \mathrm{e}_1(-ix)] && (3)\\ &= \mathrm{Si}(z) - \tfrac{1}{2}\pi, && (4)\end{aligned}$$

where $\mathrm{ei}(x)$ is the Exponential Integral and

$$\mathrm{e}_1(x) \equiv -\,\mathrm{ei}(-x). \tag{5}$$

$\mathrm{Si}(x)$ is the function returned by the *Mathematica*® (Wolfram Research, Champaign, IL) command `SinIntegral[x]` and displayed above. The half-infinite integral of the Sinc Function is given by

$$\mathrm{si}(0) = -\int_0^\infty \frac{\sin x}{x}\,dx = -\tfrac{1}{2}\pi. \tag{6}$$

To compute the integral of a sine function times a power

$$I \equiv \int x^{2n}\sin(mx)\,dx, \tag{7}$$

use Integration by Parts. Let

$$u = x^{2n} \qquad dv = \sin(mx)\,dx \tag{8}$$

$$du = 2nx^{2n-1}\,dx \qquad v = -\frac{1}{m}\cos(mx), \tag{9}$$

so

$$I = -\frac{1}{m}x^{2n}\cos(mx) + \frac{2n}{m}\int x^{2n-1}\cos(mx)\,dx. \tag{10}$$

Using Integration by Parts again,

$$u = x^{2n-1} \qquad dv = \cos(mx)\,dx \tag{11}$$

$$du = (2n-1)x^{2n-2}\,dx \qquad v = \frac{1}{m}\sin(mx) \tag{12}$$

$$\begin{aligned}\int x^{2n}\sin(mx)\,dx &= -\frac{1}{m}x^{2n}\cos(mx)\\ &\quad + \frac{2n}{m}\left[\frac{1}{m}x^{2n-1}\cos(mx)\right.\\ &\quad \left. - \frac{2n-1}{m}\int x^{2n-2}\sin(mx)\,dx\right]\\ &= -\frac{1}{m}x^{2n}\sin(mx) + \frac{2n}{m^2}x^{2n-1}\sin(mx)\\ &\quad - \frac{(2n)(2n-1)}{m^2}\int x^{2n-2}\sin(mx)\,dx\\ &= -\frac{1}{m}x^{2n}\cos(mx) + \frac{2n}{m^2}x^{2n-1}\sin(mx) + \ldots\\ &\quad + \frac{(2n)!}{m^{2n}}\int x^0\sin(mx)\,dx\\ &= -\frac{1}{m}x^{2n}\cos(mx) + \frac{2n}{m^2}x^{2n-1}\sin(mx) + \ldots\\ &\quad - \frac{(2n)!}{m^{2n+1}}\cos(mx)\\ &= \cos(mx)\sum_{k=0}^{n}(-1)^{k+1}\frac{(2n)!}{(2n-2k)!m^{2k+1}}x^{2n-2k}\\ &\quad + \sin(mx)\sum_{k=1}^{n}(-1)^{k+1}\frac{(2n)!}{(2k-2n-1)!m^{2k}}x^{2n-2k+1}.\end{aligned} \tag{13}$$

Letting $k' \equiv n - k$, so

$$\begin{aligned}\int x^{2n}\sin(mx)\,dx &= \cos(mx)\sum_{k=0}^{n}(-1)^{n-k+1}\frac{(2n)!}{(2k)!m^{2n-2k+1}}x^{2k}\\ &\quad+\sin(mx)\sum_{k=0}^{n-1}(-1)^{n-k+1}\frac{(2n)!}{(2k-1)!m^{2n-2k}}x^{2k+1}\\ &= (-1)^{n+1}(2n)!\left[\cos(mx)\sum_{k=0}^{n}\frac{(-1)^{k}}{(2k)!m^{2n-2k+1}}x^{2k}\right.\\ &\quad\left.+\sin(mx)\sum_{k=1}^{n}\frac{(-1)^{k+1}}{(2k-3)!m^{2n-2k+2}}x^{2k-1}\right].\end{aligned} \tag{14}$$

General integrals of the form

$$I(k,l)=\int_0^\infty \frac{\sin^k x}{x^l}\,dx \tag{15}$$

are related to the SINC FUNCTION and can be computed analytically.

see also CHI, COSINE INTEGRAL, EXPONENTIAL INTEGRAL, NIELSEN'S SPIRAL, SHI, SICI SPIRAL, SINC FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Sine and Cosine Integrals." §5.2 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 231–233, 1972.

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 342–343, 1985.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Fresnel Integrals, Cosine and Sine Integrals." §6.79 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 248–252, 1992.

Spanier, J. and Oldham, K. B. "The Cosine and Sine Integrals." Ch. 38 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 361–372, 1987.

Sine-Tangent Theorem

If

$$\frac{\sin\alpha}{\sin\beta}=\frac{m}{n},$$

then

$$\frac{\tan[\frac{1}{2}(\alpha-\beta)]}{\tan[\frac{1}{2}(\alpha+\beta)]}=\frac{m-n}{m+n}.$$

Sines Law

see LAW OF SINES

Singly Even Number

An EVEN NUMBER of the form $4n+2$ (i.e., an INTEGER which is DIVISIBLE by 2 but not by 4). The first few for $n = 0, 1, 2, \ldots$ are 2, 6, 10, 14, 18, ... (Sloane's A016825)

see also DOUBLY EVEN NUMBER, EVEN NUMBER, ODD NUMBER

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 30, 1996.

Sloane, N. J. A. Sequence A016825 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Singular Homology

The general type of HOMOLOGY which is what mathematicians generally mean when they say "homology." Singular homology is a more general version than Poincaré's original SIMPLICIAL HOMOLOGY.

see also HOMOLOGY (TOPOLOGY), SIMPLICIAL HOMOLOGY

Singular Point (Algebraic Curve)

A singular point of an ALGEBRAIC CURVE is a point where the curve has "nasty" behavior such as a CUSP or a point of self-intersection (when the underlying field K is taken as the REALS). More formally, a point (a, b) on a curve $f(x, y) = 0$ is singular if the x and y PARTIAL DERIVATIVES of f are both zero at the point (a, b). (If the field K is not the REALS or COMPLEX NUMBERS, then the PARTIAL DERIVATIVE is computed formally using the usual rules of CALCULUS.)

Consider the following two examples. For the curve

$$x^3-y^2=0,$$

the CUSP at (0, 0) is a singular point. For the curve

$$x^2+y^2=-1,$$

$(0, i)$ is a nonsingular point and this curve is nonsingular.

see also ALGEBRAIC CURVE, CUSP

Singular Point (Differential Equation)

Consider a second-order ORDINARY DIFFERENTIAL EQUATION

$$y''+P(x)y'+Q(x)y=0.$$

If $P(x)$ and $Q(x)$ remain FINITE at $x = x_0$, then x_0 is called an ORDINARY POINT. If either $P(x)$ or $Q(x)$ diverges as $x \to x_0$, then x_0 is called a singular point. Singular points are further classified as follows:

1. If either $P(x)$ or $Q(x)$ diverges as $x \to x_0$ but $(x - x_0)P(x)$ and $(x-x_0)^2Q(x)$ remain FINITE as $x \to x_0$, then $x = x_0$ is called a REGULAR SINGULAR POINT (or NONESSENTIAL SINGULARITY).

2. If $P(x)$ diverges more quickly than $1/(x-x_0)$, so $(x-x_0)P(x)$ approaches Infinity as $x \to x_0$, or $Q(x)$ diverges more quickly than $1/(x-x_0)^2 Q$ so that $(x-x_0)^2 Q(x)$ goes to Infinity as $x \to x_0$, then x_0 is called an Irregular Singularity (or Essential Singularity).

see also Irregular Singularity, Regular Singular Point, Singularity

References

Arfken, G. "Singular Points." §8.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 451–454, 1985.

Singular Point (Function)

Singular points (also simply called "singularities") are points z_0 in the Domain of a Function f where f fails to be Analytic. Isolated Singularities may be classified as Essential Singularities, Poles, or Removable Singularities.

Essential Singularities are Poles of Infinite order.

A Pole of order n is a singularity z_0 of $f(z)$ for which the function $(z-z_0)^n f(z)$ is nonsingular and for which $(z-z_0)^k f(z)$ is singular for $k = 0, 1, \ldots, n-1$.

Removable Singularities are singularities for which it is possible to assign a Complex Number in such a way that $f(z)$ becomes Analytic. For example, the function $f(z) = z^2/z$ has a Removable Singularity at 0, since $f(z) = z$ everywhere but 0, and $f(z)$ can be set equal to 0 at $z = 0$. Removable Singularities are not Poles.

The function $f(z) = \csc(1/z)$ has Poles at $z = 1/(2\pi n)$, and a nonisolated singularity at 0.

see also Essential Singularity, Irregular Singularity, Ordinary Point, Pole, Regular Singular Point, Removable Singularity, Singular Point (Differential Equation)

References

Arfken, G. "Singularities." §7.1 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 396–400, 1985.

Singular Series

$$\rho_{2s}(n) = \frac{\pi^s}{\Gamma(s)} n^{s-1} \sum_{p,q} \left(\frac{S_{p,q}}{q}\right)^{2s} e^{2np\pi i/q},$$

where $S_{p,q}$ is a Gaussian Sum, and $\Gamma(s)$ is the Gamma Function.

Singular System

A system is singular if the Condition Number is Infinite and Ill-Conditioned if it is too large.

see also Condition Number, Ill-Conditioned

Singular Value

A Modulus k_r such that

$$\frac{K'(k_r)}{K(k_r)} = \sqrt{r},$$

where $K(k)$ is a complete Elliptic Integral of the First Kind, and $K'(k_r) \equiv K(\sqrt{1-{k_r}^2})$. The Elliptic Lambda Function $\lambda^*(r)$ gives the value of k_r.

Abel (quoted in Whittaker and Watson 1990, p. 525) proved that if r is an Integer, or more generally whenever

$$\frac{K'(k)}{K(k)} = \frac{a+b\sqrt{n}}{c+d\sqrt{n}},$$

where a, b, c, d, and n are Integers, then the Modulus k is the Root of an algebraic equation with Integer Coefficients.

see also Elliptic Integral Singular Value, Elliptic Integral of the First Kind, Elliptic Lambda Function, Modulus (Elliptic Integral)

References

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, pp. 524–528, 1990.

Singular Value Decomposition

An expansion of a Real $M \times N$ Matrix by Orthogonal Outer Products according to

$$A = \sum_{k=1}^{K} s_k \mathbf{u}_k \mathbf{v}_k^{\mathrm{T}}, \qquad (1)$$

where $s_1 \geq s_2 \geq \ldots \geq 0$,

$$K \equiv \min\{M, N\} \qquad (2)$$

and

$$\mathbf{u}_k^{\mathrm{T}} \mathbf{u}_{k'} = \mathbf{v}_l^{\mathrm{T}} \mathbf{v}_{k'} = \delta_{kk'}. \qquad (3)$$

Here δ_{ij} is the Kronecker Delta and A^{T} is the Matrix Transpose.

see also Cholesky Decomposition, LU Decomposition, QR Decomposition

References

Nash, J. C. "The Singular-Value Decomposition and Its Use to Solve Least-Squares Problems." Ch. 3 in *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation, 2nd ed.* Bristol, England: Adam Hilger, pp. 30–48, 1990.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Singular Value Decomposition." §2.6 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 51–63, 1992.

Singularity

In general, a point at which an equation, surface, etc., blows up or becomes DEGENERATE.

see also ESSENTIAL SINGULARITY, ISOLATED SINGULARITY, SINGULAR POINT (ALGEBRAIC CURVE), SINGULAR POINT (DIFFERENTIAL EQUATION), SINGULAR POINT (FUNCTION), WHITNEY SINGULARITY

Sinh

see HYPERBOLIC SINE

Sink (Directed Graph)

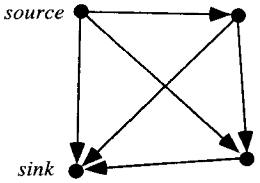

A vertex of a DIRECTED GRAPH with no exiting edges, also called a TERMINAL.

see also DIRECTED GRAPH, NETWORK, SOURCE

Sink (Map)

A stable fixed point of a MAP which, in a dissipative DYNAMICAL SYSTEM, is an ATTRACTOR.

see also ATTRACTOR, DYNAMICAL SYSTEM

Sinusoidal Projection

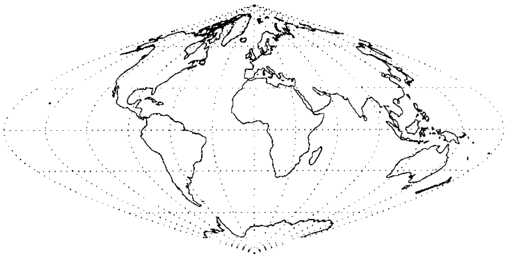

An equal AREA MAP PROJECTION.

$$x = (\lambda - \lambda_0)\cos\phi \qquad (1)$$

$$y = \phi. \qquad (2)$$

The inverse FORMULAS are

$$\phi = y \qquad (3)$$

$$\lambda = \lambda_0 + \frac{x}{\cos\phi}. \qquad (4)$$

References

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 243–248, 1987.

Sinusoidal Spiral

A curve of the form

$$r^n = a^n \cos(n\theta)$$

with n RATIONAL, which is not a true SPIRAL. Sinusoidal spirals were first studied by Maclaurin. Special cases are given in the following table.

n	Curve
-2	hyperbola
-1	line
$-\frac{1}{2}$	parabola
$-\frac{1}{3}$	Tschirnhausen cubic
0	logarithmic spiral
$\frac{1}{3}$	Cayley sextic
$\frac{1}{2}$	cardioid
1	circle
2	Bernoulli lemniscate

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, p. 184, 1972.

Lee, X. "Sinusoid." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Sinusoid_dir/sinusoid.html`.

Lockwood, E. H. *A Book of Curves.* Cambridge, England: Cambridge University Press, p. 175, 1967.

MacTutor History of Mathematics Archive. "Sinusoidal Spirals." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Sinusoidal.html`.

Sinusoidal Spiral Inverse Curve

The INVERSE CURVE of a SINUSOIDAL SPIRAL

$$r = a^{(1/n)}[\cos(nt)]^{1/n}$$

with INVERSION CENTER at the origin and inversion radius k is another SINUSOIDAL SPIRAL

$$r = ka^{(1/n)}[\cos(nt)]^{1/n}.$$

Sinusoidal Spiral Pedal Curve

The PEDAL CURVE of a SINUSOIDAL SPIRAL

$$r = a^{(1/n)}[\cos(nt)]^{1/n}$$

with PEDAL POINT at the center is another SINUSOIDAL SPIRAL

$$x = \cos^{1+1/n}(nt)\cos[(n+1)t]$$

$$y = \cos^{1+1/n}(nt)\sin[(n+1)t].$$

Sister Celine's Method

A method for finding RECURRENCE RELATIONS for hypergeometric polynomials directly from the series expansions of the polynomials. The method is effective and easily implemented, but usually slower than ZEILBERGER'S ALGORITHM. Given a sum $f(n) = \sum_k F(n,k)$, the method operates by finding a recurrence of the form

$$\sum_{i=0}^{I}\sum_{j=0}^{J} a_{ij}(n)F(n-j,k-i) = 0$$

by proceeding as follows (Petkovšek *et al.* 1996, p. 59):

1. Fix trial values of I and J.
2. Assume a recurrence formula of the above form where $a_{ij}(n)$ are to be solved for.
3. Divide each term of the assumed recurrence by $F(n,k)$ and reduce every ratio $F(n-j,k-i)/F(n,k)$ by simplifying the ratios of its constituent factorials so that only RATIONAL FUNCTIONS in n and k remain.
4. Put the resulting expression over a common DENOMINATOR, then collect the numerator as a POLYNOMIAL in k.
5. Solve the system of linear equations that results after setting the coefficients of each power of k in the NUMERATOR to 0 for the unknown coefficients a_{ij}.
6. If no solution results, start again with larger I or J.

Under suitable hypotheses, a "fundamental theorem" (Verbaten 1974, Wilf and Zeilberger 1992, Petkovšek *et al.* 1996) guarantees that this algorithm always succeeds for large enough I and J (which can be estimated in advance). The theorem also generalizes to multivariate sums and to q- and multi-q-sums (Wilf and Zeilberger 1992, Petkovšek *et al.* 1996).

see also GENERALIZED HYPERGEOMETRIC FUNCTION, GOSPER'S ALGORITHM, HYPERGEOMETRIC IDENTITY, HYPERGEOMETRIC SERIES, ZEILBERGER'S ALGORITHM

References

Fasenmyer, Sister M. C. *Some Generalized Hypergeometric Polynomials.* Ph.D. thesis. University of Michigan, Nov. 1945.

Fasenmyer, Sister M. C. "Some Generalized Hypergeometric Polynomials." *Bull. Amer. Math. Soc.* **53**, 806–812, 1947.

Fasenmyer, Sister M. C. "A Note on Pure Recurrence Relations." *Amer. Math. Monthly* **56**, 14–17, 1949.

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. "Sister Celine's Method." Ch. 4 in *A=B.* Wellesley, MA: A. K. Peters, pp. 55–72, 1996.

Rainville, E. D. Chs. 14 and 18 in *Special Functions.* New York: Chelsea, 1971.

Verbaten, P. "The Automatic Construction of Pure Recurrence Relations." *Proc. EUROSAM '74, ACM-SIGSAM Bull.* **8**, 96–98, 1974.

Wilf, H. S. and Zeilberger, D. "An Algorithmic Proof Theory for Hypergeometric (Ordinary and "q") Multisum/Integral Identities." *Invent. Math.* **108**, 575–633, 1992.

Site Percolation

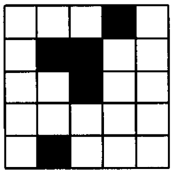

site percolation

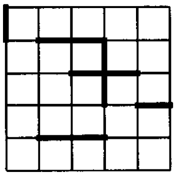

bond percolation

A PERCOLATION which considers the lattice vertices as the relevant entities (left figure).

see also BOND PERCOLATION, PERCOLATION THEORY

Siteswap

A siteswap is a sequence encountered in JUGGLING in which each term is a POSITIVE integer, encoded in BINARY. The transition rule from one term to the next consists of changing some 0 to 1, subtracting 1, and then dividing by 2, with the constraint that the DIVISION by two must be exact. Therefore, if a term is EVEN, the bit to be changed must be the units bit. In siteswaps, the number of 1-bits is a constant.

Each transition is characterized by the bit position of the toggled bit (denoted here by the numeral on top of the arrow). For example,

$$111 \xrightarrow{5} 10011 \xrightarrow{2} 1011 \xrightarrow{5} 10101 \xrightarrow{1} 1011 \xrightarrow{2} 111$$
$$\xrightarrow{6} 100011 \xrightarrow{3} 10101 \xrightarrow{3} 1110 \xrightarrow{0} 111 \xrightarrow{4} 1011 \ldots$$

The second term is given from the first as follows: 000111 with bit 5 flipped becomes 100111, or 39. Subtract 1 to obtain 38 and divide by two to obtain 19, which is 10011.

see also JUGGLING

References

Juggling Information Service. "Siteswaps." `http://www.juggling.org/help/siteswap`.

Six-Color Theorem

To color any map on the SPHERE or the PLANE requires at most six-colors. This number can be easily be reduced to five, and the FOUR-COLOR THEOREM demonstrates that the NECESSARY number is, in fact, four.

see also FOUR-COLOR THEOREM, HEAWOOD CONJECTURE, MAP COLORING

References

Franklin, P. "A Six Colour Problem." *J. Math. Phys.* **13**, 363–369, 1934.

Hoffman, I. and Soifer, A. "Another Six-Coloring of the Plane." *Disc. Math.* **150**, 427–429, 1996.

Saaty, T. L. and Kainen, P. C. *The Four-Color Problem: Assaults and Conquest.* New York: Dover, 1986.

Skein Relationship

A relationship between Knot Polynomials for links in different orientations (denoted below as L_+, L_0, and L_-). J. H. Conway was the first to realize that the Alexander Polynomial could be defined by a relationship of this type.

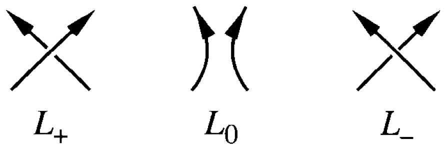

see also Alexander Polynomial, HOMFLY Polynomial, Signature (Knot)

Skeleton

The Graph obtained by collapsing a Polyhedron into the Plane. The number of topologically distinct skeletons $N(n)$ with n Vertices is given in the following table.

n	$N(n)$
4	1
5	2
6	7

References

Gardner, M. *Martin Gardner's New Mathematical Diversions from Scientific American.* New York: Simon and Schuster, p. 233, 1966.

Skeleton Division

A Long Division in which most or all of the digits are replaced by a symbol (usually asterisks) to form a Cryptarithm.

see also Cryptarithm

Skew Conic

Also known as a Gauche Conic, Space Conic, Twisted Conic, or Cubical Conic Section. A third-order Space Curve having up to three points in common with a plane and having three points in common with the plane at infinity. A skew cubic is determined by six points, with no four of them Coplanar. A line is met by up to four tangents to a skew cubic.

A line joining two points of a skew cubic (Real or conjugate imaginary) is called a Secant of the curve, and a line having one point in common with the curve is called a Semisecant or Transversal. Depending on the nature of the roots, the skew conic is classified as follows:

1. The three Roots are Real and distinct (Cubical Hyperbola).
2. One root is Real and the other two are Complex Conjugates (Cubical Ellipse).
3. Two of the Roots coincide (Cubical Parabolic Hyperbola).
4. All three Roots coincide (Cubical Parabola).

see also Conic Section, Cubical Ellipse, Cubical Hyperbola, Cubical Parabola, Cubical Parabolic Hyperbola

Skew Field

A Field in which the commutativity of multiplication is not required, more commonly called a Division Algebra.

see also Division Algebra, Field

Skew Lines

Two or more Lines which have no intersections but are not Parallel, also called Agonic Lines. Since two Lines in the Plane must intersect or be Parallel, skew lines can exist only in three or more Dimensions.

see also Gallucci's Theorem, Regulus

Skew Polyomino

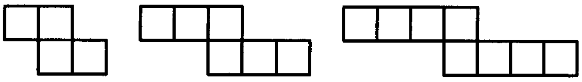

see also L-Polyomino, Square Polyomino, Straight Polyomino, T-Polyomino

Skew Quadrilateral

A four-sided Quadrilateral not contained in a plane. The problem of finding the minimum bounding surface of a skew quadrilateral was solved by Schwarz (1890) in terms of Abelian Integrals and has the shape of a Saddle. It is given by solving

$$(1+{f_y}^2)f_{xx} - 2f_x f_y f_{xy} + (1+{f_x}^2)f_{yy} = 0.$$

see also Quadrilateral

References

Isenberg, C. *The Science of Soap Films and Soap Bubbles.* New York: Dover, p. 81, 1992.

Forsyth, A. R. *Calculus of Variations.* New York: Dover, p. 503, 1960.

Schwarz, H. A. *Gesammelte Mathematische Abhandlungen, 2nd ed.* New York: Chelsea.

Skew Symmetric Matrix

A Matrix A where

$$\mathsf{A}^{\mathrm{T}} = -\mathsf{A},$$

with A^{T} denoting the Matrix Transpose.

see also Matrix Transpose, Symmetric Matrix

Skewes Number

The Skewes number (or first Skewes number) is the number Sk_1 above which $\pi(n) < \mathrm{Li}(n)$ must fail (assuming that the RIEMANN HYPOTHESIS is true), where $\pi(n)$ is the PRIME COUNTING FUNCTION and $\mathrm{Li}(n)$ is the LOGARITHMIC INTEGRAL.

$$\mathrm{Sk}_1 = e^{e^{e^{79}}} \approx 10^{10^{10^{34}}}.$$

The Skewes number has since been reduced to $e^{e^{27/4}} \approx 8.185 \times 10^{370}$ by te Riele (1987), although Conway and Guy (1996) claim that the best current limit is 10^{1167}. In 1914, Littlewood proved that the inequality must, in fact, fail infinitely often.

The second Skewes number Sk_2 is the number above which $\pi(n) < \mathrm{Li}(n)$ must fail (assuming that the RIEMANN HYPOTHESIS is false). It is much larger than the Skewes number Sk_1,

$$\mathrm{Sk}_2 = 10^{10^{10^{10^3}}}.$$

see also GRAHAM'S NUMBER, RIEMANN HYPOTHESIS

References

Asimov, I. "Skewered!" *Of Matters Great and Small.* New York: Ace Books, 1976. Originally published in *Magazine of Fantasy and Science Fiction,* Nov. 1974.

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 63, 1987.

Boas, R. P. "The Skewes Number." In *Mathematical Plums* (Ed. R. Honsberger). Washington, DC: Math. Assoc. Amer., 1979.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 61, 1996.

Lehman, R. S. "On the Difference $\pi(x) - \mathrm{li}(x)$." *Acta Arith.* **11**, 397–410, 1966.

te Riele, H. J. J. "On the Sign of the Difference $\pi(x) - \mathrm{li}(x)$." *Math. Comput.* **48**, 323–328, 1987.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, p. 30, 1991.

Skewness

The degree of asymmetry of a distribution. If the distribution has a longer tail less than the maximum, the function has NEGATIVE skewness. Otherwise, it has POSITIVE skewness. Several types of skewness are defined. The FISHER SKEWNESS is defined by

$$\gamma_1 = \frac{\mu_3}{{\mu_2}^{3/2}} = \frac{\mu_3}{\sigma^3}, \tag{1}$$

where μ_3 is the third MOMENT, and ${\mu_2}^{1/2} \equiv \sigma$ is the STANDARD DEVIATION. The PEARSON SKEWNESS is defined by

$$\beta_1 = \left(\frac{\mu_3}{\sigma^3}\right)^2 = {\gamma_1}^2. \tag{2}$$

The MOMENTAL SKEWNESS is defined by

$$\alpha^{(m)} \equiv \tfrac{1}{2}\gamma_1. \tag{3}$$

The PEARSON MODE SKEWNESS is defined by

$$\frac{[\text{mean}] - [\text{mode}]}{\sigma}. \tag{4}$$

PEARSON'S SKEWNESS COEFFICIENTS are defined by

$$\frac{3[\text{mean}] - [\text{mode}]}{s} \tag{5}$$

and

$$\frac{3[\text{mean}] - [\text{median}]}{s}. \tag{6}$$

The BOWLEY SKEWNESS (also known as QUARTILE SKEWNESS COEFFICIENT) is defined by

$$\frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{Q_1 - 2Q_2 + Q_3}{Q_3 - Q_1}, \tag{7}$$

where the Qs denote the INTERQUARTILE RANGES. The MOMENTAL SKEWNESS is

$$\alpha^{(m)} \equiv \tfrac{1}{2}\gamma = \frac{\mu_3}{2\sigma^3}. \tag{8}$$

An ESTIMATOR for the FISHER SKEWNESS γ_1 is

$$g_1 = \frac{k_3}{{k_2}^{3/2}}, \tag{9}$$

where the ks are k-STATISTICS. The STANDARD DEVIATION of g_1 is

$${\sigma_{g_1}}^2 \approx \frac{6}{N}. \tag{10}$$

see also BOWLEY SKEWNESS, FISHER SKEWNESS, GAMMA STATISTIC, KURTOSIS, MEAN, MOMENTAL SKEWNESS, PEARSON SKEWNESS, STANDARD DEVIATION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 928, 1972.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Moments of a Distribution: Mean, Variance, Skewness, and So Forth." §14.1 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 604–609, 1992.

Sklar's Theorem

Let H be a 2-D distribution function with marginal distribution functions F and G. Then there exists a COPULA C such that

$$H(x,y) = C(F(x), G(y)).$$

Conversely, for any univariate distribution functions F and G and any COPULA C, the function H is a two-dimensional distribution function with marginals F and G. Furthermore, if F and G are continuous, then C is unique.

Skolem-Mahler-Lerch Theorem

If $\{a_0, a_1, \ldots\}$ is a Recurrence Sequence, then the set of all k such that $a_k = 0$ is the union of a finite (possibly Empty) set and a finite number (possibly zero) of full arithmetical progressions, where a full arithmetic progression is a set of the form $\{r, r+d, r+2d, \ldots\}$ with $r \in [0, d)$.

References

Myerson, G. and van der Poorten, A. J. "Some Problems Concerning Recurrence Sequences." *Amer. Math. Monthly* **102**, 698–705, 1995.

Skolem Paradox

Even though Arithmetic is uncountable, it possesses a countable "model."

Skolem Sequence

A Skolem sequence of order n is a sequence $S = \{s_1, s_2, \ldots, s_{2n}\}$ of $2n$ integers such that

1. For every $k \in \{1, 2, \ldots, n\}$, there exist exactly two elements $s_i, s_j \in S$ such that $s_i = s_j = k$, and
2. If $s_i = s_j = k$ with $i < j$, then $j - i = k$.

References

Colbourn, C. J. and Dinitz, J. H. (Eds.) "Skolem Sequences." Ch. 43 in *CRC Handbook of Combinatorial Designs.* Boca Raton, FL: CRC Press, pp. 457–461, 1996.

Slant Height

The height of an object (such as a Cone) measured along a side from the edge of the base to the apex.

Slice Knot

A Knot K in $\mathbb{S}^3 = \partial\mathbb{D}^4$ is a slice knot if it bounds a Disk Δ^2 in $\mathbb{D}^4$ which has a Tubular Neighborhood $\Delta^2 \times \mathbb{D}^2$ whose intersection with $\mathbb{S}^3$ is a Tubular Neighborhood $K \times \mathbb{D}^2$ for K.

Every Ribbon Knot is a slice knot, and it is conjectured that every slice knot is a Ribbon Knot.

see also Ribbon Knot, Tubular Neighborhood

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 218, 1976.

Slide Move

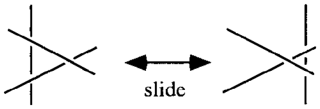

The Reidemeister Move of type III.

see also Reidemeister Moves

Slide Rule

A mechanical device consisting of a sliding portion and a fixed case, each marked with logarithmic axes. By lining up the ticks, it is possible to do Multiplication by taking advantage of the additive property of Logarithms. More complicated slide rules also allow the extraction of roots and computation of trigonometric functions. The development of the desk calculator (and subsequently pocket calculator) rendered slide rules largely obsolete beginning in the 1960s.

see also Abacus, Ruler, Straightedge

References

Electronic Teaching Laboratories. *Simplify Math: Learn to Use the Slide Rule.* New Augusta, IN: Editors and Engineers, 1966.

Saffold, R. *The Slide Rule.* Garden City, NY: Doubleday, 1962.

Slightly Defective Number

see Almost Perfect Number

Slightly Excessive Number

see Quasiperfect Number

Slip Knot

see Running Knot

Slope

A quantity which gives the inclination of a curve or line with respect to another curve or line. For a Line in the Plane making an Angle θ with the x-Axis, the Slope m is a constant given by

$$m \equiv \frac{\Delta y}{\Delta x} = \tan\theta,$$

where Δx and Δy are changes in the two coordinates over some distance. It is meaningless to talk about the slope in 3-D unless the slope *with respect to what* is specified.

Slothouber-Graatsma Puzzle

Assemble six $1 \times 2 \times 2$ blocks and three $1 \times 1 \times 1$ blocks into a $3 \times 3 \times 3$ Cube.

see also Box-Packing Theorem, Conway Puzzle, Cube Dissection, de Bruijn's Theorem, Klarner's Theorem, Polycube

References

Honsberger, R. *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., pp. 75–77, 1976.

Slutzky-Yule Effect

A Moving Average may generate an irregular oscillation even if none exists in the original data.

see also Moving Average

Sluze Pearls

see PEARLS OF SLUZE

Smale-Hirsch Theorem

The SPACE of IMMERSIONS of a MANIFOLD in another MANIFOLD is HOMOTOPICALLY equivalent to the space of bundle injections from the TANGENT SPACE of the first to the TANGENT BUNDLE of the second.

see also HOMOTOPY, IMMERSION, MANIFOLD, TANGENT BUNDLE, TANGENT SPACE

Smale Horseshoe Map

The basic topological operations for constructing an ATTRACTOR consist of stretching (which gives sensitivity to initial conditions) and folding (which gives the attraction). Since trajectories in PHASE SPACE cannot cross, the repeated stretching and folding operations result in an object of great topological complexity.

The Smale horseshoe map consists of a sequence of operations on the unit square. First, stretch by a factor of 2 in the x direction, then compress by $2a$ in the y direction. Then, fold the rectangle and fit it back into the square. Repeating this generates the horseshoe attractor. If one looks at a cross-section of the final structure, it is seen to correspond to a CANTOR SET.

see also ATTRACTOR, CANTOR SET

References
Gleick, J. *Chaos: Making a New Science.* New York: Penguin, pp. 50–51, 1988.
Rasband, S. N. *Chaotic Dynamics of Nonlinear Systems.* New York: Wiley, p. 77, 1990.
Tabor, M. *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, 1989.

Small Cubicuboctahedron

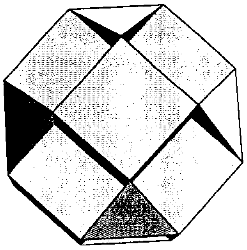

UNIFORM POLYHEDRON U_{13} whose DUAL POLYHEDRON is the SMALL HEXACRONIC ICOSITETRAHEDRON. It has WYTHOFF SYMBOL $\frac{3}{2}\,4\,|\,4$. Its faces are $8\{3\}+6\{4\}+6\{8\}$. The CIRCUMRADIUS for the solid with unit edge length is

$$R=\tfrac{1}{2}\sqrt{5+2\sqrt{2}}.$$

FACETED versions include the GREAT RHOMBICUBOCTAHEDRON (UNIFORM) and SMALL RHOMBIHEXAHEDRON.

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 104–105, 1971.

Small Ditrigonal Dodecacronic Hexecontahedron

The DUAL POLYHEDRON of the SMALL DITRIGONAL DODECICOSIDODECAHEDRON.

Small Ditrigonal Dodecicosidodecahedron

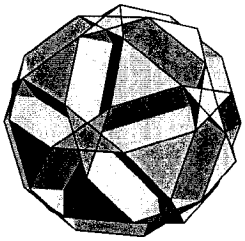

The UNIFORM POLYHEDRON U_{43} whose DUAL POLYHEDRON is the SMALL DITRIGONAL DODECACRONIC HEXECONTAHEDRON. It has WYTHOFF SYMBOL $3\,\frac{5}{3}\,|\,5$. Its faces are $20\{3\}+12\{\frac{5}{2}\}+12\{10\}$. Its CIRCUMRADIUS with $a=1$ is

$$R=\tfrac{1}{4}\sqrt{34+6\sqrt{5}}.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 126–127, 1971.

Small Ditrigonal Icosidodecahedron

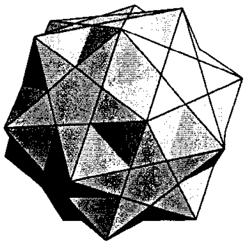

The UNIFORM POLYHEDRON U_{30} whose DUAL POLYHEDRON is the SMALL TRIAMBIC ICOSAHEDRON. It has WYTHOFF SYMBOL $3\,|\,3\,\frac{5}{2}$. Its faces are $20\{3\}+12\{\frac{5}{2}\}$. A FACETED version is the DITRIGONAL DODECADODECAHEDRON. Its CIRCUMRADIUS with $a=1$ is

$$R=\tfrac{1}{2}\sqrt{3}.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 106–107, 1971.

Small Dodecacronic Hexecontahedron

The DUAL POLYHEDRON of the SMALL DODECICOSIDODECAHEDRON.

Small Dodecahemicosacron

The DUAL POLYHEDRON of the SMALL DODECAHEMICOSAHEDRON.

Small Dodecahemicosahedron

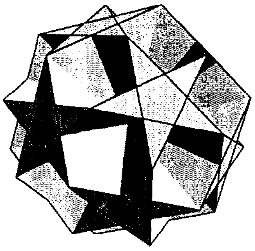

The UNIFORM POLYHEDRON U_{62} whose DUAL POLYHEDRON is the SMALL DODECAHEMICOSACRON. It has WYTHOFF SYMBOL $\frac{5}{3}\,\frac{5}{2}\,|\,3$. Its faces are $10\{6\}+12\{\frac{5}{2}\}$. It is a FACETED version of the ICOSIDODECAHEDRON. Its CIRCUMRADIUS with unit edge length is

$$R=1.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 155, 1971.

Small Dodecahemidodecacron

The DUAL POLYHEDRON of the SMALL DODECAHEMIDODECAHEDRON.

Small Dodecahemidodecahedron

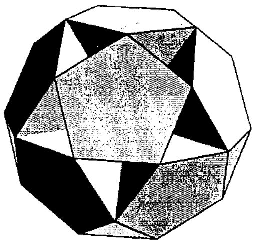

The UNIFORM POLYHEDRON U_{51} whose DUAL POLYHEDRON is the SMALL DODECAHEMIDODECACRON. It has WYTHOFF SYMBOL $2\,5\,\begin{matrix}\frac{3}{2}\\ \frac{5}{2}\end{matrix}$. Its faces are $30\{4\}+12\{10\}$. Its CIRCUMRADIUS with $a=1$ is

$$R=\tfrac{1}{2}\sqrt{11+4\sqrt{5}}\,.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 113–114, 1971.

Small Dodecicosacron

The DUAL POLYHEDRON of the SMALL DODECICOSAHEDRON.

Small Dodecicosahedron

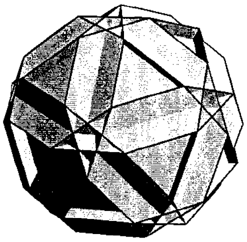

The UNIFORM POLYHEDRON U_{50} whose DUAL POLYHEDRON is the SMALL DODECICOSACRON. It has WYTHOFF SYMBOL $3\,5\,\begin{matrix}\frac{3}{2}\\ \frac{5}{4}\end{matrix}\Big|$. Its faces are $20\{6\}+12\{10\}$. Its CIRCUMRADIUS with $a=1$ is

$$R=\tfrac{1}{4}\sqrt{34+6\sqrt{5}}\,.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 141–142, 1971.

Small Dodecicosidodecahedron

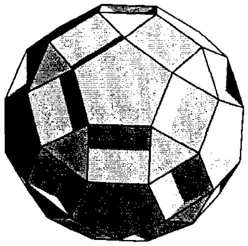

The UNIFORM POLYHEDRON U_{33} whose DUAL POLYHEDRON is the SMALL DODECACRONIC HEXECONTAHEDRON. It has WYTHOFF SYMBOL $\frac{3}{2}\,5\,|\,5$. Its faces are $20\{3\}+12\{5\}+12\{10\}$. It is a FACETED version of the SMALL RHOMBICOSIDODECAHEDRON. Its CIRCUMRADIUS with $a=1$ is

$$R=\tfrac{1}{2}\sqrt{11+4\sqrt{5}}\,.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 110–111, 1971.

Small Hexacronic Icositetrahedron

The DUAL POLYHEDRON of the SMALL CUBICUBOCTAHEDRON.

Small Hexagonal Hexecontahedron

The DUAL POLYHEDRON of the SMALL SNUB ICOSICOSIDODECAHEDRON.

Small Hexagrammic Hexecontahedron

The DUAL POLYHEDRON of the SMALL RETROSNUB ICOSICOSIDODECAHEDRON.

Small Icosacronic Hexecontahedron

The DUAL POLYHEDRON of the SMALL ICOSICOSIDODECAHEDRON.

Small Icosicosidodecahedron

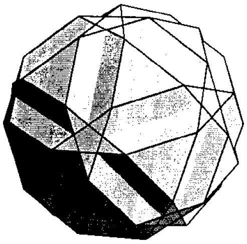

The UNIFORM POLYHEDRON U_{31} whose DUAL POLYHEDRON is the SMALL ICOSACRONIC HEXECONTAHEDRON. It has WYTHOFF SYMBOL $\frac{5}{4}\,5\,|\,5$. Its faces are $12\{5\}+6\{10\}$. Its CIRCUMRADIUS with $a=1$ is

$$R=\phi=\tfrac{1}{2}(1+\sqrt{5}),$$

where ϕ is the GOLDEN RATIO.

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 143, 1971.

Small Icosihemidodecacron

The DUAL POLYHEDRON of the SMALL ICOSIHEMIDODECAHEDRON.

Small Icosihemidodecahedron

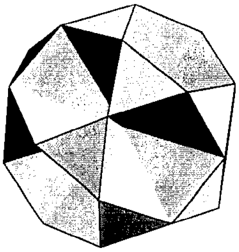

The UNIFORM POLYHEDRON U_{49} whose DUAL POLYHEDRON is the SMALL ICOSIHEMIDODECACRON. It has WYTHOFF SYMBOL $\frac{3}{2}\,3\,|\,5$. Its faces are $20\{3\}+6\{10\}$. It is a FACETED version of the ICOSIDODECAHEDRON. Its CIRCUMRADIUS with $a=1$ is

$$R=\phi=\tfrac{1}{2}(1+\sqrt{5}).$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 140, 1971.

Small Inverted Retrosnub Icosicosidodecahedron

see SMALL RETROSNUB ICOSICOSIDODECAHEDRON

Small Multiple Method

An algorithm for computing a UNIT FRACTION.

Small Number

Guy's "STRONG LAW OF SMALL NUMBERS" states that there aren't enough small numbers to meet the many demands made of them. Guy (1988) also gives several interesting and misleading facts about small numbers:

1. 10% of the first 100 numbers are SQUARE NUMBERS.
2. A QUARTER of the numbers < 100 are PRIMES.
3. All numbers less than 10, except for 6, are PRIME POWERS.
4. Half the numbers less than 10 are FIBONACCI NUMBERS.

see also LARGE NUMBER, STRONG LAW OF SMALL NUMBERS

References

Guy, R. K. "The Strong Law of Small Numbers." *Amer. Math. Monthly* **95**, 697–712, 1988.

Small Retrosnub Icosicosidodecahedron

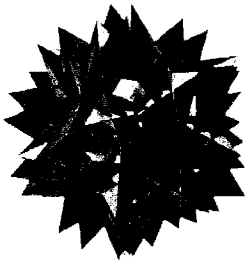

The UNIFORM POLYHEDRON U_{72} also called the SMALL INVERTED RETROSNUB ICOSICOSIDODECAHEDRON whose DUAL POLYHEDRON is the SMALL HEXAGRAMMIC HEXECONTAHEDRON. It has WYTHOFF SYMBOL $|\,\frac{3}{2}\,\frac{3}{2}\,\frac{5}{2}$. Its faces are $100\{3\}+12\{\frac{5}{2}\}$. It has CIRCUMRADIUS with $a=1$

$$R=\tfrac{1}{4}\sqrt{13+3\sqrt{5}-\sqrt{102+46\sqrt{5}}}\approx 0.580694800133921.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 194–199, 1971.

Small Rhombicosidodecahedron

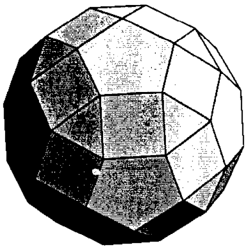

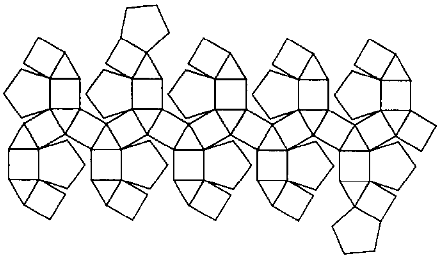

An ARCHIMEDEAN SOLID whose DUAL POLYHEDRON is the DELTOIDAL HEXECONTAHEDRON. It has SCHLÄFLI SYMBOL $r\left\{ {3 \atop 5} \right\}$. It is also UNIFORM POLYHEDRON U_{27} with WYTHOFF SYMBOL $3\,5\,|\,2$. Its faces are $20\{3\} + 30\{4\} + 12\{5\}$. The SMALL DODECICOSIDODECAHEDRON and SMALL RHOMBIDODECAHEDRON are FACETED versions. The INRADIUS, MIDRADIUS, and CIRCUMRADIUS for $a = 1$ are

$$r = \tfrac{1}{41}(15 + 2\sqrt{5})\sqrt{11 + 4\sqrt{5}} = 2.12099\ldots$$

$$\rho = \tfrac{1}{2}\sqrt{10 + 4\sqrt{5}} = 2.17625\ldots$$

$$R = \tfrac{1}{2}\sqrt{11 + 4\sqrt{5}} = 2.23295\ldots.$$

see also GREAT RHOMBICOSIDODECAHEDRON (ARCHIMEDEAN), GREAT RHOMBICOSIDODECAHEDRON (UNIFORM)

Small Rhombicuboctahedron

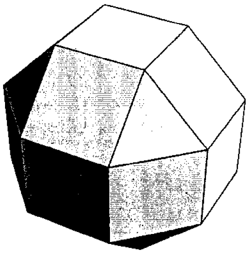

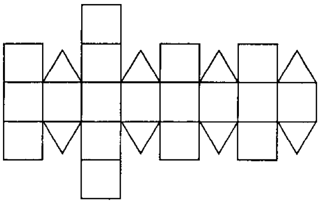

An ARCHIMEDEAN SOLID also (inappropriately) called the TRUNCATED ICOSIDODECAHEDRON. This name is inappropriate since truncation would yield rectangular instead of square faces. Its DUAL POLYHEDRON is the DELTOIDAL ICOSITETRAHEDRON, also called the TRAPEZOIDAL ICOSITETRAHEDRON. It has SCHLÄFLI SYMBOL $r\left\{ {3 \atop 4} \right\}$. It is also UNIFORM POLYHEDRON U_{10} and has WYTHOFF SYMBOL $3\,4\,|\,2$. Its INRADIUS, MIDRADIUS, and CIRCUMRADIUS for $a = 1$ are

$$r = \tfrac{1}{17}(6 + \sqrt{2})\sqrt{5 + 2\sqrt{2}} = 1.22026\ldots$$

$$\rho = \tfrac{1}{2}\sqrt{4 + 2\sqrt{2}} = 1.30656\ldots$$

$$R = \tfrac{1}{2}\sqrt{5 + 2\sqrt{2}} = 1.39897\ldots.$$

A version in which the top and bottom halves are rotated with respect to each other is known as the ELONGATED SQUARE GYROBICUPOLA.

see also ELONGATED SQUARE GYROBICUPOLA, GREAT RHOMBICUBOCTAHEDRON (ARCHIMEDEAN), GREAT RHOMBICUBOCTAHEDRON (UNIFORM)

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 137–138, 1987.

Small Rhombidodecacron

The DUAL POLYHEDRON of the SMALL RHOMBIDODECAHEDRON.

Small Rhombidodecahedron

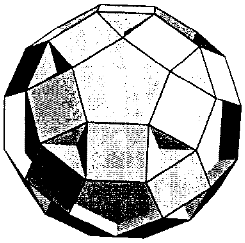

The UNIFORM POLYHEDRON U_{39} whose DUAL POLYHEDRON is the SMALL RHOMBIDODECACRON. It has WYTHOFF SYMBOL $2\,5\,\frac{\frac{3}{2}}{\frac{5}{2}}$. Its faces are $30\{4\} + 12\{10\}$. It is a FACETED version of the SMALL RHOMBICOSIDODECAHEDRON. Its CIRCUMRADIUS with $a = 1$ is

$$R = \tfrac{1}{2}\sqrt{11 + 4\sqrt{5}}.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 113–114, 1971.

Small Rhombihexacron

The DUAL POLYHEDRON of the SMALL RHOMBIHEXAHEDRON.

Small Rhombihexahedron

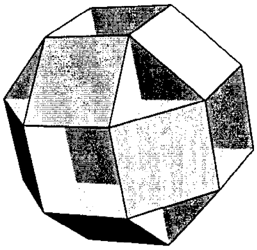

The UNIFORM POLYHEDRON U_{18} whose DUAL POLYHEDRON is the SMALL RHOMBIHEXACRON. It has WYTHOFF SYMBOL $2\,4\,\frac{\frac{3}{2}}{\frac{4}{2}}\,\Big|$. Its faces are $12\{4\} + 6\{8\}$. It is a Faceted version of the Small Rhombicuboctahedron. Its Circumradius with $a = 1$ is

$$R = \tfrac{1}{2}\sqrt{5 + 2\sqrt{2}}.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 134, 1971.

Small Snub Icosicosidodecahedron

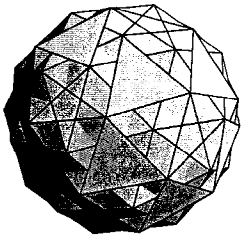

The Uniform Polyhedron U_{32} whose Dual Polyhedron is the Small Hexagonal Hexecontahedron. It has Wythoff Symbol $|\,3\,3\,\frac{5}{2}$ (Har'El 1993 gives the symbol as $|\,\frac{5}{2}\,3\,3$.) Its faces are $100\{3\} + 12\{\frac{5}{2}\}$. Its Circumradius for $a = 1$ is

$$\begin{aligned} R &= \tfrac{1}{4}\sqrt{13 + 3\sqrt{5} + \sqrt{102 + 46\sqrt{5}}} \\ &= 1.4581903307387\ldots. \end{aligned}$$

References
Har'El, Z. "Uniform Solution for Uniform Polyhedra." *Geometriae Dedicata* **47**, 57–110, 1993.
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 172–173, 1971.

Small Stellapentakis Dodecahedron

The Dual Polyhedron of the Truncated Great Dodecahedron.

Small Stellated Dodecahedron

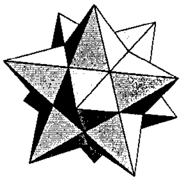

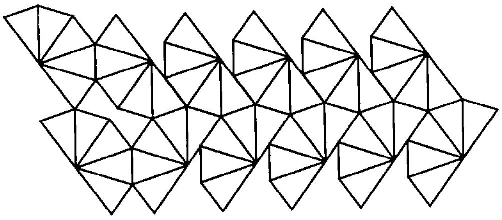

One of the Kepler-Poinsot Solids whose Dual Polyhedron is the Great Dodecahedron. Its Schläfli Symbol is $\{\frac{5}{2}, 5\}$. It is also Uniform Polyhedron U_{34} and has Wythoff Symbol $5\,|\,2\,\frac{5}{2}$. It was originally called the Urchin by Kepler. It is composed of 12 Pentagrammic faces. Its faces are $12\{\frac{5}{2}\}$. The easiest way to construct it is to build twelve pentagonal Pyramids

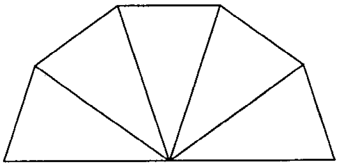

and attach them to the faces of a Dodecahedron. The Circumradius of the small stellated dodecahedron with $a = 1$ is

$$R = \tfrac{1}{2}5^{1/4}\phi^{-1/2} = \tfrac{1}{4}5^{1/4}\sqrt{2(\sqrt{5} - 1)}.$$

see also Great Dodecahedron, Great Icosahedron, Great Stellated Dodecahedron, Kepler-Poinsot Solid

References
Fischer, G. (Ed.). Plate 103 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 102, 1986.
Rawles, B. *Sacred Geometry Design Sourcebook: Universal Dimensional Patterns.* Nevada City, CA: Elysian Pub., p. 219, 1997.

Small Stellated Triacontahedron

see Medial Rhombic Triacontahedron

Small Stellated Truncated Dodecahedron

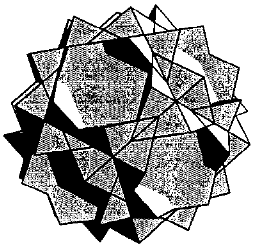

The Uniform Polyhedron U_{58} also called the Quasitruncated Small Stellated Dodecahedron whose Dual Polyhedron is the Great Pentakis Dodecahedron. It has Schläfli Symbol t'$\{\frac{5}{2}, 5\}$ and Wythoff Symbol $2\,5\,|\,\frac{5}{3}$. Its faces are $12\{5\} + 12\{\frac{10}{3}\}$. Its Circumradius with $a = 1$ is

$$R = \tfrac{1}{4}\sqrt{34 - 10\sqrt{5}}.$$

References
Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 151, 1971.

Small Triakis Octahedron

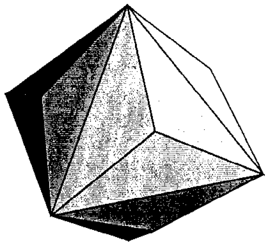

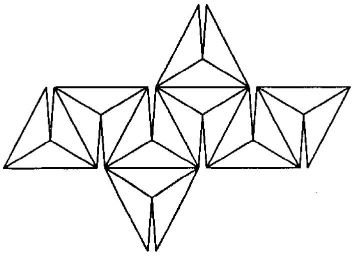

The DUAL POLYHEDRON of the TRUNCATED CUBE.

see also GREAT TRIAKIS OCTAHEDRON

Small Triambic Icosahedron

The DUAL POLYHEDRON of the SMALL DITRIGONAL ICOSIDODECAHEDRON.

Small World Problem

The small world problem asks for the probability that two people picked at random have at least one acquaintance in common.

see also BIRTHDAY PROBLEM

Smarandache Ceil Function

A SMARANDACHE-like function which is defined where $S_k(n)$ is defined as the smallest integer for which $n|S_k(n)^k$. The Smarandache $S_k(n)$ function can therefore be obtained by replacing any factors which are kth powers in n by their k roots. The functions $S_k(n)$ for $k = 2, 3, \ldots, 6$ for values such that $S_k(n) \neq n$ are tabulated by Begay (1997).

$S_1(n) = n$, so the first few values of $S_1(n)$ are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ... (Sloane's A000027). The first few values of $S_2(n)$ are 1, 2, 3, 2, 5, 6, 7, 4, 3, 10, 11, 6, 13, 14, 15, 4, 17, 6, 19, 10, ... (Sloane's A019554) The first few values of $S_3(n)$ are 1, 2, 3, 2, 5, 6, 7, 2, 3, 10, 11, 6, 13, 14, 15, 4, 17, 6, 19, 10, ... (Sloane's A019555) The first few values of $S_4(n)$ are 1, 2, 3, 2, 5, 6, 7, 2, 3, 10, 11, 6, 13, 14, 15, 2, 17, 6, 19, 10, ... (Sloane's A007947).

see also PSEUDOSMARANDACHE FUNCTION, SMARANDACHE FUNCTION, SMARANDACHE-KUREPA FUNCTION, SMARANDACHE NEAR-TO-PRIMORIAL FUNCTION, SMARANDACHE SEQUENCES, SMARANDACHE-WAGSTAFF FUNCTION, SMARANDACHE FUNCTION

References

Begay, A. "Smarandache Ceil Functions." *Bull. Pure Appl. Sci.* **16E**, 227–229, 1997.

"Functions in Number Theory." `http://www.gallup.unm.edu/~smarandache/FUNCT1.TXT`.

Sloane, N. J. A. Sequences A007947, A019554, A019555, and A0472/M000027 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Smarandache, F. *Collected Papers, Vol. 2.* Kishinev, Moldova: Kishinev University Press, 1997.

Smarandache, F. *Only Problems, Not Solutions!, 4th ed.* Phoenix, AZ: Xiquan, 1993.

Smarandache Constants

The first Smarandache constant is defined as

$$S_1 \equiv \sum_{n=2}^{\infty} \frac{1}{[S(n)]!} > 1.093111,$$

where $S(n)$ is the SMARANDACHE FUNCTION. Cojocaru and Cojocaru (1996a) prove that S_1 exists and is bounded by $0.717 < S_1 < 1.253$. The lower limit given above is obtained by taking 40,000 terms of the sum.

Cojocaru and Cojocaru (1996b) prove that the second Smarandache constant

$$S_2 \equiv \sum_{n=2}^{\infty} \frac{S(n)}{n!} \approx 1.71400629359162$$

is an IRRATIONAL NUMBER.

Cojocaru and Cojocaru (1996c) prove that the series

$$S_3 \equiv \sum_{n=2}^{\infty} \frac{1}{\prod_{i=2}^{n} S(i)} \approx 0.719960700043708$$

converges to a number $0.71 < S_3 < 1.01$, and that

$$S_4(a) \equiv \sum_{n=2}^{\infty} \frac{n^a}{\prod_{i=2}^{n} S(i)}$$

converges for a fixed REAL NUMBER $a \geq 1$. The values for small a are

$$\begin{aligned} S_4(1) &\approx 1.72875760530223 \\ S_4(2) &\approx 4.50251200619297 \\ S_4(3) &\approx 13.0111441949445 \\ S_4(4) &\approx 42.4818449849626 \\ S_4(5) &\approx 158.105463729329. \end{aligned}$$

Sandor (1997) shows that the series

$$S_5 \equiv \sum_{n=1}^{\infty} \frac{(-1)^{n-1} S(n)}{n!}$$

converges to an IRRATIONAL. Burton (1995) and Dumitrescu and Seleacu (1996) show that the series

$$S_6 \equiv \sum_{n=2}^{\infty} \frac{S(n)}{(n+1)!}$$

converges. Dumitrescu and Seleacu (1996) show that the series

$$S_7 \equiv \sum_{n=r}^{\infty} \frac{S(n)}{(n+r)!}$$

and

$$S_8 \equiv \sum_{n=r}^{\infty} \frac{S(n)}{(n-r)!}$$

converge for r a natural number (which must be nonzero in the latter case). Dumitrescu and Seleacu (1996) show that

$$S_9 \equiv \sum_{n=2}^{\infty} \frac{1}{\sum_{i=2}^{n} \frac{S(i)}{i!}}$$

converges. Burton (1995) and Dumitrescu and Seleacu (1996) show that the series

$$S_{10} \sum_{n=2}^{\infty} \frac{1}{[S(n)]^{\alpha} \sqrt{S(n)!}}$$

and

$$S_{11} \sum_{n=2}^{\infty} \frac{1}{[S(n)]^{\alpha} \sqrt{[S(n)+1]!}}$$

converge for $\alpha > 1$.

see also SMARANDACHE FUNCTION

References
Burton, E. "On Some Series Involving the Smarandache Function." *Smarandache Notions J.* **6**, 13–15, 1995.
Burton, E. "On Some Convergent Series." *Smarandache Notions J.* **7**, 7–9, 1996.
Cojocaru, I. and Cojocaru, S. "The First Constant of Smarandache." *Smarandache Notions J.* **7**, 116–118, 1996a.
Cojocaru, I. and Cojocaru, S. "The Second Constant of Smarandache." *Smarandache Notions J.* **7**, 119–120, 1996b.
Cojocaru, I. and Cojocaru, S. "The Third and Fourth Constants of Smarandache." *Smarandache Notions J.* **7**, 121–126, 1996c.
"Constants Involving the Smarandache Function." `http://www.gallup.unm.edu/~smarandache/CONSTANT.TXT`.
Dumitrescu, C. and Seleacu, V. "Numerical Series Involving the Function *S*." *The Smarandache Function.* Vail: Erhus University Press, pp. 48–61, 1996.
Ibstedt, H. *Surfing on the Ocean of Numbers—A Few Smarandache Notions and Similar Topics.* Lupton, AZ: Erhus University Press, pp. 27–30, 1997.
Sandor, J. 'On The Irrationality Of Certain Alternative Smarandache Series." *Smarandache Notions J.* **8**, 143–144, 1997.
Smarandache, F. *Collected Papers, Vol. 1.* Bucharest, Romania: Tempus, 1996.
Smarandache, F. *Collected Papers, Vol. 2.* Kishinev, Moldova: Kishinev University Press, 1997.

Smarandache Function

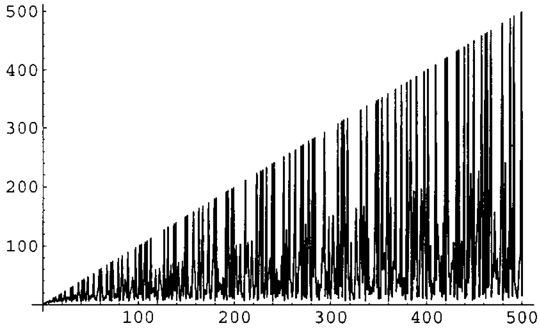

The smallest value $S(n)$ for a given n for which $n|S(n)!$ (n divides $S(n)$ FACTORIAL). For example, the number 8 does not divide 1!, 2!, 3!, but does divide $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 8 \cdot 3$, so $S(8) = 4$. For a PRIME p, $S(p) = p$, and for an EVEN PERFECT NUMBER r, $S(r)$ is PRIME (Ashbacher 1997).

The Smarandache numbers for $n = 1, 2, \ldots$ are 1, 2, 3, 4, 5, 3, 7, 4, 6, 5, 11, ... (Sloane's A002034). Letting $a(n)$ denote the smallest value of n for which $S(n) = 1, 2, \ldots$, then $a(n)$ is given by 1, 2, 3, 4, 5, 9, 7, 32, 27, 25, 11, 243, ... (Sloane's A046021). Some values of $S(n)$ first occur only for very large n, for example, $S(59{,}049) = 24$, $S(177{,}147) = 27$, $S(134{,}217{,}728) = 30$, $S(43{,}046{,}721) = 36$, and $S(9{,}765{,}625) = 45$. D. Wilson points out that if we let

$$I(n,p) = \frac{n - \Sigma(n,p)}{p-1},$$

be the power of the PRIME p in $n!$, where $\Sigma(n,p)$ is the sum of the base-p digits of n, then it follows that

$$a(n) = \min p^{I(n-1,p)+1},$$

where the minimum is taken over the PRIMES p dividing n. This minimum appears to always be achieved when p is the GREATEST PRIME FACTOR of n.

The incrementally largest values of $S(n)$ are 1, 2, 3, 4, 5, 7, 11, 13, 17, 19, 23, 29, ... (Sloane's A046022), which occur for $n = 1, 2, 3, 4, 5, 7, 11, 13, 17, 19, 23, 29, \ldots$, i.e., the values where $S(n) = n$.

Tutescu (1996) conjectures that the DIOPHANTINE EQUATION $S(n) = S(n+1)$ has no solution.

see also FACTORIAL, GREATEST PRIME FACTOR, PSEUDOSMARANDACHE FUNCTION, SMARANDACHE CEIL FUNCTION, SMARANDACHE CONSTANTS, SMARANDACHE-KUREPA FUNCTION, SMARANDACHE NEAR-TO-PRIMORIAL FUNCTION, SMARANDACHE-WAGSTAFF FUNCTION

References
Ashbacher, C. *An Introduction to the Smarandache Function.* Cedar Rapids, IA: Decisionmark, 1995.
Ashbacher, C. "Problem 4616." *School Sci. Math.* **97**, 221, 1997.

Begay, A. "Smarandache Ceil Functions." *Bulletin Pure Appl. Sci. India* **16E**, 227–229, 1997.
Dumitrescu, C. and Seleacu, V. *The Smarandache Function.* Vail, AZ: Erhus University Press, 1996.
"Functions in Number Theory." `http://www.gallup.unm.edu/~smarandache/FUNCT1.TXT`.
Ibstedt, H. *Surfing on the Ocean of Numbers—A Few Smarandache Notions and Similar Topics.* Lupton, AZ: Erhus University Press, pp. 27–30, 1997.
Sandor, J. "On Certain Inequalities Involving the Smarandache Function." *Abstracts of Papers Presented to the Amer. Math. Soc.* **17**, 583, 1996.
Sloane, N. J. A. Sequences A046021, A046022, A046023, and A002034/M0453 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Smarandache, F. *Collected Papers, Vol. 1.* Bucharest, Romania: Tempus, 1996.
Smarandache, F. *Collected Papers, Vol. 2.* Kishinev, Moldova: Kishinev University Press, 1997.
Tutescu, L. "On a Conjecture Concerning the Smarandache Function." *Abstracts of Papers Presented to the Amer. Math. Soc.* **17**, 583, 1996.

Smarandache-Kurepa Function

Given the sum-of-factorials function

$$\Sigma(n) = \sum_{k=1}^{n} k!,$$

SK(p) is the smallest integer for p PRIME such that $1 + \Sigma[\mathrm{SK}(p-1)]$ is divisible by p. The first few known values of SK(p) are 2, 4, 6, 6, 5, 7, 7, 12, 22, 16, 55, 54, 42, 24, ... for $p = 2$, 5, 7, 11, 17, 19, 23, 31, 37, 41, 61, 71, 73, 89, ... The values for $p = 3$, 13, 29, 43, 47, 53, 67, 79, 83, ..., if they are finite, must be very large (e.g., $\mathrm{SK}(3) > 100,000$).

see also PSEUDOSMARANDACHE FUNCTION, SMARANDACHE CEIL FUNCTION, SMARANDACHE FUNCTION, SMARANDACHE-WAGSTAFF FUNCTION, SMARANDACHE FUNCTION

References

Ashbacher, C. "Some Properties of the Smarandache-Kurepa and Smarandache-Wagstaff Functions." *Math. Informatics Quart.* **7**, 114–116, 1997.
Mudge, M. "Introducing the Smarandache-Kurepa and Smarandache-Wagstaff Functions." *Smarandache Notions J.* **7**, 52–53, 1996.
Mudge, M. "Introducing the Smarandache-Kurepa and Smarandache-Wagstaff Functions." *Abstracts of Papers Presented to the Amer. Math. Soc.* **17**, 583, 1996.

Smarandache Near-to-Primorial Function

$SNTP(n)$ is the smallest PRIME such that $p\# - 1$, $p\#$, or $p\# + 1$ is divisible by n, where $p\#$ is the PRIMORIAL of p. Ashbacher (1996) shows that $SNTP(n)$ only exists

1. If there are no square or higher powers in the factorization of n, or
2. If there exists a PRIME $q < p$ such that $n|(q\# \pm 1)$, where p is the smallest power contained in the factorization of n.

Therefore, $SNTP(n)$ does not exist for the SQUAREFUL numbers $n = 4$, 8, 9, 12, 16, 18, 20, 24, 25, 27, 28, ... (Sloane's A002997) The first few values of $SNTP(n)$, where defined, are 2, 2, 2, 3, 3, 3, 5, 7, ... (Sloane's A046026).

see also PRIMORIAL, SMARANDACHE FUNCTION

References

Ashbacher, C. "A Note on the Smarandache Near-To-Primordial Function." *Smarandache Notions J.* **7**, 46–49, 1996.
Mudge, M. R. "The Smarandache Near-To-Primorial Function." *Abstracts of Papers Presented to the Amer. Math. Soc.* **17**, 585, 1996.
Sloane, N. J. A. Sequence A002997 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Smarandache Paradox

Let A be some attribute (e.g., possible, present, perfect, etc.). If all is A, then the non-A must also be A. For example, "All is possible, the impossible too," and "Nothing is perfect, not even the perfect."

References

Le, C. T. "The Smarandache Class of Paradoxes." *Bull. Transylvania Univ. Brasov* **36**, 7–8, 1994.
Le, C. T. "The Smarandache Class of Paradoxes." *Bull. Pure Appl. Sci.* **14E**, 109–110, 1995.
Le, C. T. "The Smarandache Class of Paradoxes." *J. Indian Acad. Math.* **18**, 53–55, 1996.
Mitroiescu, I. *The Smarandache Class of Paradoxes.* Glendale, AZ: Erhus University Press, 1994.
Mitroiescu, I. "The Smarandache's Class of Paradoxes Applied in Computer Science." *Abstracts of Papers Presented to the Amer. Math. Soc.* **16**, 651, 1995.

Smarandache Sequences

Smarandache sequences are any of a number of simply generated INTEGER SEQUENCES resembling those considered in published works by Smarandache such as the CONSECUTIVE NUMBER SEQUENCES and EUCLID NUMBERS (Iacobescu 1997). Other Smarandache-type sequences are given below.

1. The concatenation of n copies of the INTEGER n: 1, 22, 333, 4444, 55555, ... (Sloane's A000461; Marimutha 1997),
2. The concatenation of the first n FIBONACCI NUMBERS: 1, 11, 112, 1123, 11235, ... (Sloane's A019523; Marimutha 1997),
3. The smallest number that is the sum of squares of *two* distinct earlier terms: 1, 2, 5, 26, 29, 677, ... (Sloane's A008318, Bencze 1997),
4. The smallest number that is the sum of squares of any number of distinct earlier terms: 1, 1, 2, 4, 5, 6, 16, 17, ... (Sloane's A008319, Bencze 1997),
5. The smallest number that is *not* the sum of squares of *two* distinct earlier terms: 1, 2, 3, 4, 6, 7, 8, 9, 11, ... (Sloane's A008320, Bencze 1997),
6. The smallest number that is *not* the sum of squares of any number of distinct earlier terms: 1, 2, 3, 6, 7, 8, 11, ... (Sloane's A008321, Bencze 1997),

7. The smallest number that is a sum of cubes of *two* distinct earlier terms: 1, 2, 9, 730, 737, ... (Sloane's A008322, Bencze 1997),
8. The smallest number that is a sum of cubes of any number of distinct earlier terms: 1, 1, 2, 8, 9, 512, 513, 514, ... (Sloane's A008323, Bencze 1997),
9. The smallest number that is *not* a sum of cubes of *two* of distinct earlier terms: 1, 2, 3, 4, 5, 6, 7, 8, 10, ... (Sloane's A008380, Bencze 1997),
10. The smallest number that is *not* a sum of cubes of any number of distinct earlier terms: 1, 2, 3, 4, 5, 6, 7, 10, 11, ... (Sloane's A008381, Bencze 1997),
11. The number of PARTITIONS of a number $n = 1$, 2, ... into SQUARE NUMBERS: 1, 1, 1, 1, 2, 2, 2, 2, 3, 4, 4, 4, 5, 6, 6, 6, 8, 9, 10, 10, 12, 13, ... (Sloane's A001156, Iacobescu 1997),
12. The number of PARTITIONS of a number $n = 1$, 2, ... into CUBIC NUMBERS: 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, ... (Sloane's A003108, Iacobescu 1997),
13. Two copies of the first n POSITIVE integers: 11, 1212, 123123, 12341234, ... (Sloane's A019524, Iacobescu 1997),
14. Numbers written in base of triangular numbers: 1, 2, 10, 11, 12, 100, 101, 102, 110, 1000, 1001, 1002, ... (Sloane's A000462, Iacobescu 1997),
15. Numbers written in base of double factorial numbers: 1, 10, 100, 101, 110, 200, 201, 1000, 1001, 1010, ... (Sloane's A019513, Iacobescu 1997),
16. Sequences starting with terms $\{a_1, a_2\}$ which contain no three-term arithmetic progressions starting with $\{1, 2\}$: 1, 2, 4, 5, 10, 11, 13, 14, 28, ... (Sloane's A033155, Iacobescu 1997, Mudge 1997, Weisstein),
17. Numbers of the form $(n!)^2 + 1$: 2, 5, 37, 577, 14401, 518401, 25401601, 1625702401, 131681894401, ... (Sloane's A020549, Iacobescu 1997),
18. Numbers of the form $(n!)^3 + 1$: 2, 9, 217, 13825, 1728001, 373248001, 128024064001, ... (Sloane's A019514, Iacobescu 1997),
19. Numbers of the form $1 + 1!2!3! \cdots n!$: 2, 3, 13, 289, 34561, 24883201, 125411328001, 5056584744960001, ... (Sloane's A019515, Iacobescu 1997),
20. Sequences starting with terms $\{a_1, a_2\}$ which contain no three-term geometric progressions starting with $\{1, 2\}$: 1, 2, 3, 5, 6, 7, 8, 10, 11, 13, 14, 15, 16, ... (Sloane's A000452, Iacobescu 1997),
21. Numbers repeating the digit 1 p_n times, where p_n is the nth prime: 11, 111, 11111, 1111111, ... (Sloane's A031974, Iacobescu 1997). These are a subset of the REPUNITS,
22. Integers with all 2s, 3s, 5s, and 7s (prime digits) removed: 1, 4, 6, 8, 9, 10, 11, 1, 1, 14, 1, 16, 1, 18, 19, 0, ... (Sloane's A019516, Iacobescu 1997),
23. Integers with all 0s, 1s, 4s, and 9s (square digits) removed: 2, 3, 5, 6, 7, 8, 2, 3, 5, 6, 7, 8, 2, 2, 22, 23, ... (Sloane's A031976, Iacobescu 1997).
24. (Smarandache-Fibonacci triples) Integers n such that $S(n) = S(n-1) + S(n-2)$, where $S(k)$ is the SMARANDACHE FUNCTION: 3, 11, 121, 4902, 26245, ... (Sloane's A015047; Aschbacher and Mudge 1995; Ibstedt 1997, pp. 19–23; Begay 1997). The largest known is 19,448,047,080,036,
25. (Smarandache-Radu triplets) Integers n such that there are no primes between the smaller and larger of $S(n)$ and $S(n+1)$: 224, 2057, 265225, ... (Sloane's A015048; Radu 1994/1995, Begay 1997, Ibstedt 1997). The largest known is 270,329,975,921, 205,253,634,707,051,822,848,570,391,313,
26. (Smarandache crescendo sequence): Integers obtained by concatenating strings of the first $n+1$ integers for $n = 0, 1, 2, \ldots$: 1, 1, 2, 1, 2, 3, 1, 2, 3, 4, ... (Sloane's A002260; Brown 1997, Brown and Castillo 1997). The nth term is given by $n - m(m+1)/2 + 1$, where $m = \lfloor (\sqrt{8n+1} - 1)/2 \rfloor$, with $\lfloor x \rfloor$ the FLOOR FUNCTION (Hamel 1997),
27. (Smarandache descrescendo sequence): Integers obtained by concatenating strings of the first n integers for $n = \ldots, 2, 1$: 1, 2, 1, 3, 2, 1, 4, 3, 2, 1, ... (Sloane's A004736; Smarandache 1997, Brown 1997),
28. (Smarandache crescendo pyramidal sequence): Integers obtained by concatenating strings of rising and falling integers: 1, 1, 2, 1, 1, 2, 3, 2, 1, 1, 2, 3, 4, 3, 2, 1, ... (Sloane's A004737; Brown 1997, Brown and Castillo 1997, Smarandache 1997),
29. (Smarandache descrescendo pyramidal sequence): Integers obtained by concatenating strings of falling and rising integers: 1, 2, 1, 2, 3, 2, 1, 2, 3, 4, 3, 2, 1, 2, 3, 4, ... (Brown 1997),
30. (Smarandache crescendo symmetric sequence): 1, 1, 1, 2, 2, 1, 1, 2, 3, 3, 2, 1, ... (Sloane's A004739, Brown 1997, Smarandache 1997),
31. (Smarandache descrescendo symmetric sequence): 1, 1, 2, 1, 1, 2, 3, 2, 1, 1, 2, 3, ... (Sloane's A004740; Brown 1997, Smarandache 1997),
32. (Smarandache permutation sequence): Numbers obtained by concatenating sequences of increasing length of increasing ODD NUMBERS and decreasing EVEN NUMBERS: 1, 2, 1, 3, 4, 2, 1, 3, 5, 6, 4, 2, ... (Sloane's A004741; Brown 1997, Brown and Castillo 1997),
33. (Smarandache pierced chain sequence): Numbers of the form $c(n) = 101\underbrace{0101}_{n}$ for $n = 0, 1, \ldots$: 101, 1010101, 10101010101, ... (Sloane's A031982; Ashbacher 1997). In addition, $c(n)/101$ contains no PRIMES (Ashbacher 1997),

34. (Smarandache symmetric sequence): 1, 11, 121, 1221, 12321, 123321, ... (Sloane's A007907; Smarandache 1993, Dumitrescu and Seleacu 1994, sequence 3; Mudge 1995),
35. (Smarandache square-digital sequence): square numbers all of whose digits are also squares: 1, 4, 9, 49, 100, 144, ... (Sloane's A019544; Mudge 1997),
36. (Square-digits): numbers composed of digits which are squares: 1, 4, 9, 10, 14, 19, 40, 41, 44, 49, ... (Sloane's A066030),
37. (Smarandache square-digital sequence): square-digit numbers which are themselves squares: 1, 4, 9, 49, 100, 144, ... (Sloane's A019544; Mudge 1997),
38. (Cube-digits): numbers composed of digits which are cubes: 1, 4, 10, 11, 14, 40, 41, 44, 100, 101, ... (Sloane's A046031),
39. (Smarandache cube-digital sequence): cube-digit numbers which are themselves cubes: 1, 8, 1000, 8000, 1000000, ... (Sloane's A019545; Mudge 1997),
40. (Prime-digits): numbers composed of digits which are primes: 2, 3, 5, 7, 22, 23, 25, 27, 32, 33, 35, ... (Sloane's A046034),
41. (Smarandache prime-digital sequence): prime-digit numbers which are themselves prime: 2, 3, 5, 7, 23, 37, 53, ... (Smith 1996, Mudge 1997).

see also ADDITION CHAIN, CONSECUTIVE NUMBER SEQUENCES, CUBIC NUMBER, EUCLID NUMBER, EVEN NUMBER, FIBONACCI NUMBER, INTEGER SEQUENCE, ODD NUMBER, PARTITION, SMARANDACHE FUNCTION, SQUARE NUMBER

References

Aschbacher, C. *Collection of Problems On Smarandache Notions.* Vail, AZ: Erhus University Press, 1996.

Aschbacher, C. and Mudge, M. *Personal Computer World.* pp. 302, Oct. 1995.

Begay, A. "Smarandache Ceil Functions." *Bull. Pure Appl. Sci.* **16E**, 227–229, 1997.

Bencze, M. "Smarandache Recurrence Type Sequences." *Bull. Pure Appl. Sci.* **16E**, 231–236, 1997.

Bencze, M. and Tutescu, L. (Eds.). *Some Notions and Questions in Number Theory, Vol. 2.* `http://www.gallup.unm.edu/~smarandache/SNAQINT2.TXT`.

Brown, J. "Crescendo & Descrescendo." In *Richard Henry Wilde: An Anthology in Memoriam (1789–1847)* (Ed. M. Myers). Bristol, IN: Bristol Banner Books, p. 19, 1997.

Brown, J. and Castillo, J. "Problem 4619." *School Sci. Math.* **97**, 221–222, 1997.

Dumitrescu, C. and Seleacu, V. (Ed.). *Some Notions and Questions in Number Theory, 4th ed.* Glendale, AZ: Erhus University Press, 1994. `http://www.gallup.unm.edu/~smarandache/SNAQINT.TXT`.

Dumitrescu, C. and Seleacu, V. (Ed.). *Proceedings of the First International Conference on Smarandache Type Notions in Number Theory.* Lupton, AZ: American Research Press, 1997.

Hamel, E. Solution to Problem 4619. *School Sci. Math.* **97**, 221–222, 1997.

Iacobescu, F. "Smarandache Partition Type and Other Sequences." *Bull. Pure Appl. Sci.* **16E**, 237–240, 1997.

Ibstedt, H. *Surfing on the Ocean of Numbers—A Few Smarandache Notions and Similar Topics.* Lupton, AZ: Erhus University Press, 1997.

Kashihara, K. *Comments and Topics on Smarandache Notions and Problems.*ail, AZ: Erhus University Press, 1996.

Mudge, M. "Top of the Class." *Personal Computer World,* 674–675, June 1995.

Mudge, M. "Not Numerology but Numeralogy!" *Personal Computer World,* 279–280, 1997.

Programs and the Abstracts of the First International Conference on Smarandache Notions in Number Theory. Craiova, Romania, Aug. 21–23, 1997.

Radu, I. M. *Mathematical Spectrum* **27**, 43, 1994/1995.

Sloane, N. J. A. Sequences A001156/M0221 and A003108/M0209 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Smarandache, F. "Properties of the Numbers." Tempe, AZ: Arizona State University Special Collection, 1975.

Smarandache, F. *Only Problems, Not Solutions!, 4th ed.* Phoenix, AZ: Xiquan, 1993.

Smarandache, F. *Collected Papers, Vol. 2.* Kishinev, Moldova: Kishinev University Press, 1997.

Smith, S. "A Set of Conjectures on Smarandache Sequences." *Bull. Pure Appl. Sci.* **15E**, 101–107, 1996.

Smarandache-Wagstaff Function

Given the sum-of-FACTORIALS function

$$\Sigma(n)=\sum_{k=1}^{n}k!,$$

$SW(p)$ is the smallest integer for p PRIME such that $\Sigma[SW(p)]$ is divisible by p. The first few known values are 2, 4, 5, 12, 19, 24, 32, 19, 20, 20, 20, 7, 57, 6, ... for $p=3$, 11, 17, 23, 29, 37, 41, 43, 53, 67, 73, 79, 97, ... The values for 5, 7, 13, 31, ..., if they are finite, must be very large.

see also FACTORIAL, SMARANDACHE FUNCTION

References

Ashbacher, C. "Some Properties of the Smarandache-Kurepa and Smarandache-Wagstaff Functions." *Math. Informatics Quart.* **7**, 114–116, 1997.

"Functions in Number Theory." `http://www.gallup.unm.edu/~smarandache/FUNCT1.TXT`.

Mudge, M. "Introducing the Smarandache-Kurepa and Smarandache-Wagstaff Functions." *Smarandache Notions J.* **7**, 52–53, 1996.

Mudge, M. "Introducing the Smarandache-Kurepa and Smarandache-Wagstaff Functions." *Abstracts of Papers Presented to the Amer. Math. Soc.* **17**, 583, 1996.

Smith Brothers

Consecutive SMITH NUMBERS. The first two brothers are (728, 729) and (2964, 2965).

see also SMITH NUMBER

Smith Conjecture

The set of fixed points which do not move as a knot is transformed into itself is not a KNOT. The conjecture was proved in 1978.

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 350–351, 1976.

Smith's Markov Process Theorem

Consider

$$P_2(y_1,t|y_3,t_3) = \int P_2(y_1,t_1|y_2,t_1)P_3(y_1,t_1;y_2,t_2|y_3,t_3)\,dy_2. \quad (1)$$

If the probability distribution is governed by a MARKOV PROCESS, then

$$P_3(y_1,t_1;y_2,t_2|y_3,t_3) = P_2(y_2,t_2|y_3,t_3) = P_2(y_2|y_3,t_3-t_2). \quad (2)$$

Assuming no time dependence, so $t_1 \equiv 0$,

$$P_2(y_1|y_3,t_3) = \int P_2(y_1|y_2,t_2)P_2(y_2|y_3,t_3-t_2)\,dy_2. \quad (3)$$

see also MARKOV PROCESS

Smith's Network Theorem

In a NETWORK with three EDGES at each VERTEX, the number of HAMILTONIAN CIRCUITS through a specified EDGE is 0 or EVEN.

see also EDGE (GRAPH), HAMILTONIAN CIRCUIT, NETWORK

Smith Normal Form

A form for INTEGER matrices.

Smith Number

A COMPOSITE NUMBER the SUM of whose DIGITS is the sum of the DIGITS of its PRIME factors (excluding 1). (The PRIMES are excluded since they trivially satisfy this condition). One example of a Smith number is the BEAST NUMBER

$$666 = 2 \cdot 3 \cdot 3 \cdot 37,$$

since

$$6+6+6 = 2+3+3+(3+7) = 18.$$

Another Smith number is

$$4937775 = 3 \cdot 5 \cdot 5 \cdot 65837,$$

since

$$4+9+3+7+7+7+5 = 3+5+5+(6+5+8+3+7) = 42.$$

The first few Smith numbers are 4, 22, 27, 58, 85, 94, 121, 166, 202, 265, 274, 319, 346, ... (Sloane's A006753). There are 360 Smith numbers less than 10^4 and 29,928 $\leq 10^6$. McDaniel (1987a) showed that an infinite number exist.

A generalized k-Smith number can also be defined as a number m satisfying $S_p(m) = kS(m)$, where S_p is the sum of prime factors and S is the sum of digits. There are 47 1-Smith numbers, 21 2-Smith numbers, three 3-Smith numbers, and one 7-Smith, 9-Smith, and 14-Smith number < 1000.

A Smith number can be constructed from every factored REPUNIT R_n. The largest known Smith number is

$$9 \times R_{1031}(10^{4594} + 3 \times 10^{2297} + 1)^{1476} \times 10^{3913210}.$$

see also MONICA SET, PERFECT NUMBER, REPUNIT, SMITH BROTHERS, SUZANNE SET

References

Gardner, M. *Penrose Tiles and Trapdoor Ciphers... and the Return of Dr. Matrix, reissue ed.* New York: W. H. Freeman, pp. 99–300, 1989.

Guy, R. K. "Smith Numbers." §B49 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 103–104, 1994.

McDaniel, W. L. "The Existence of Infinitely Many k-Smith Numbers." *Fib. Quart.*, **25**, 76–80, 1987a.

McDaniel, W. L. "Powerful K-Smith Numbers." *Fib. Quart.* **25**, 225-228, 1987b.

Oltikar, S. and Weiland, K. "Construction of Smith Numbers." *Math. Mag.* **56**, 36–37, 1983.

Sloane, N. J. A. Sequence A006753/M3582 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Wilansky, A. "Smith Numbers." *Two-Year College Math. J.* **13**, 21, 1982.

Yates, S. "Special Sets of Smith Numbers." *Math. Mag.* **59**, 293–296, 1986.

Yates, S. "Smith Numbers Congruent to 4 (mod 9)." *J. Recr. Math.* **19**, 139–141, 1987.

Smooth Manifold

Another word for a C^∞ (infinitely differentiable) MANIFOLD. A smooth manifold is a TOPOLOGICAL MANIFOLD together with its "functional structure" (Bredon 1995) and so differs from a TOPOLOGICAL MANIFOLD because the notion of differentiability exists on it. Every smooth manifold is a TOPOLOGICAL MANIFOLD, but not necessarily vice versa. (The first nonsmooth TOPOLOGICAL MANIFOLD occurs in 4-D.) In 1959, Milnor showed that a 7-D HYPERSPHERE can be made into a smooth manifold in 28 ways.

see also DIFFERENTIABLE MANIFOLD, HYPERSPHERE, MANIFOLD, TOPOLOGICAL MANIFOLD

References

Bredon, G. E. *Topology & Geometry.* New York: Springer-Verlag, p. 69, 1995.

Smooth Number

An INTEGER is k-smooth if it has no PRIME FACTORS $> k$. The probability that a random POSITIVE INTEGER $\leq n$ is k-smooth is $\psi(n,k)/n$, where $\psi(n,k)$ is the number of k-smooth numbers $\leq n$. This fact is important in application of Kraitchik's extension of FERMAT'S FACTORIZATION METHOD because it is related to the number of random numbers which must be examined to find a suitable subset whose product is a square.

Since about $\pi(k)$ k-smooth numbers must be found (where $\pi(k)$ is the PRIME COUNTING FUNCTION), the number of random numbers which must be examined is about $\pi(k)n/\psi(n,k)$. But because it takes about $\pi(k)$ steps to determine if a number is k-smooth using TRIAL DIVISION, the expected number of steps needed to find a subset of numbers whose product is a square is $\sim [\pi(k)]^2 n/\psi(n,k)$ (Pomerance 1996). Canfield *et al.* (1983) showed that this function is minimized when

$$k \sim \exp(\tfrac{1}{2}\sqrt{\ln n \ln\ln n}\,)$$

and that the minimum value is about

$$\exp(2\sqrt{\ln n \ln\ln n}\,).$$

In the CONTINUED FRACTION FACTORIZATION ALGORITHM, n can be taken as $2\sqrt{n}$, but in FERMAT'S FACTORIZATION METHOD, it is $n^{1/2+\epsilon}$. k is an estimate for the largest PRIME in the FACTOR BASE (Pomerance 1996).

References

Canfield, E. R.; Erdős, P.; and Pomerance, C. "On a Problem of Oppenheim Concerning 'Factorisation Numerorum.'" *J. Number Th.* **17**, 1–28, 1983.

Pomerance, C. "On the Role of Smooth Numbers in Number Theoretic Algorithms." In *Proc. Internat. Congr. Math., Zürich, Switzerland, 1994, Vol. 1* (Ed. S. D. Chatterji). Basel: Birkhäuser, pp. 411–422, 1995.

Pomerance, C. "A Tale of Two Sieves." *Not. Amer. Math. Soc.* **43**, 1473–1485, 1996.

Smooth Surface

A surface PARAMETERIZED in variables u and v is called smooth if the TANGENT VECTORS in the u and v directions satisfy

$$\mathbf{T}_u \times \mathbf{T}_v \neq \mathbf{0},$$

where $\mathbf{A} \times \mathbf{B}$ is a CROSS PRODUCT.

Snake

A simple circuit in the d-HYPERCUBE which has no chords (i.e., for which all snake edges are edges of the HYPERCUBE). Klee (1970) asked for the maximum length $s(d)$ of a d-snake. Klee (1970) gave the bounds

$$\frac{7}{4(d-1)} \leq \frac{s(d)}{2^d}\frac{1}{2} - \frac{1-12\cdots 2^{-d}}{7d(d-1)^2+2} \qquad (1)$$

for $d \geq 6$ (Danzer and Klee 1967, Douglas 1969), as well as numerous references. Abbott and Katchalski (1988) show

$$s(d) \geq 77 \cdot 2^{d-8}, \qquad (2)$$

and Snevily (1994) showed that

$$s(d) \leq 2^{n-1}\left(1 - \frac{1}{20n-41}\right) \qquad (3)$$

for $n \leq 12$, and conjectured

$$s(d) \leq 3 \cdot 2^{n-3} + 2 \qquad (4)$$

for $n \leq 5$. The first few values for $s(d)$ for $d = 1, 2, \ldots$, are 2, 4, 6, 8, 14, 26, ... (Sloane's A000937).

see also HYPERCUBE

References

Abbott, H. L. and Katchalski, M. "On the Snake in the Box Problem." *J. Combin. Th. Ser. B* **44**, 12–24, 1988.

Danzer, L. and Klee, V. "Length of Snakes in Boxes." *J. Combin. Th.* **2**, 258–265, 1967.

Douglas, R. J. "Some Results on the Maximum Length of Circuits of Spread k in the d-Cube." *J. Combin. Th.* **6**, 323–339, 1969.

Evdokimov, A. A. "Maximal Length of a Chain in a Unit n-Dimensional Cube." *Mat. Zametki* **6**, 309–319, 1969.

Guy, R. K. "Unsolved Problems Come of Age." *Amer. Math. Monthly* **96**, 903–909, 1989.

Kautz, W. H. "Unit-Distance Error-Checking Codes." *IRE Trans. Elect. Comput.* **7**, 177–180, 1958.

Klee, V. "What is the Maximum Length of a d-Dimensional Snake?" *Amer. Math. Monthly* **77**, 63–65, 1970.

Sloane, N. J. A. Sequence A000937/M0995 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Snevily, H. S. "The Snake-in-the-Box Problem: A New Upper Bound." *Disc. Math.* **133**, 307–314, 1994.

Snake Eyes

A roll of two 1s (the lowest roll possible) on a pair of six-sided DICE. The probability of rolling snake eyes is 1/36, or 2.777... %.

see also BOXCARS

Snake Oil Method

The expansion of the two sides of a sum equality in terms of POLYNOMIALS in x^m and y^k, followed by closed form summation in terms of x and y. For an example of the technique, see Bloom (1995).

References

Bloom, D. M. "A Semi-Unfriendly Identity." Problem 10206. Solution by R. J. Chapman. *Amer. Math. Monthly* **102**, 657–658, 1995.

Wilf, H. S. *Generatingfunctionology, 2nd ed.* New York: Academic Press, 1993.

Snake Polyiamond

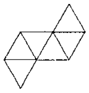

A 6-POLYIAMOND.

References

Golomb, S. W. *Polyominoes: Puzzles, Patterns, Problems, and Packings, 2nd ed.* Princeton, NJ: Princeton University Press, p. 92, 1994.

Snedecor's F-Distribution

If a random variable X has a CHI-SQUARED DISTRIBUTION with m degrees of freedom ($\chi_m{}^2$) and a random variable Y has a CHI-SQUARED DISTRIBUTION with n degrees of freedom ($\chi_n{}^2$), and X and Y are independent, then

$$F \equiv \frac{X/m}{Y/n} \qquad (1)$$

is distributed as Snedecor's F-distribution with m and n degrees of freedom

$$f(F(m,n)) = \frac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{m}{n}\right)^{m/2} F^{(m-2)/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)\left(1+\frac{m}{n}F\right)^{(m+n)/2}} \qquad (2)$$

for $0 < F < \infty$. The MOMENTS about 0 are

$$\mu_1' = \frac{n}{n-2} \qquad (3)$$

$$\mu_2' = \frac{n^2(m+2)}{m(n-2)(n-4)} \qquad (4)$$

$$\mu_3' = \frac{n^3(m+2)(m+4)}{m^2(n-2)(n-4)(n-6)} \qquad (5)$$

$$\mu_4' = \frac{n^4(m+2)(m+4)(m+6)}{m^3(n-2)(n-4)(n-6)(n-8)}, \qquad (6)$$

so the MOMENTS about the MEAN are given by

$$\mu_2 = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \qquad (7)$$

$$\mu_3 = \frac{8n^3(m+n-2)(2m+n-2)}{m^2(n-2)^3(n-4)(n-6)} \qquad (8)$$

$$\mu_4 = \frac{12n^4(m+n-2)}{m^3(n-2)^4(n-4)(n-6)(n-8)}g(m,n), \qquad (9)$$

where

$$g(m,n) = mn^2 + 4n^2 + m^2n + 8mn - 16n + 10m^2 - 20m + 16, \qquad (10)$$

and the MEAN, VARIANCE, SKEWNESS, and KURTOSIS are

$$\mu = \mu_1' = \frac{n}{n-2} \qquad (11)$$

$$\sigma^2 = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \qquad (12)$$

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = 2\sqrt{\frac{2(n-4)}{m(m+n-2)}}\,\frac{2m+n-2}{n-6} \qquad (13)$$

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{12h(m,n)}{m(m+n-2)(n-6)(n-8)}, \qquad (14)$$

where

$$h(m,n) \equiv n^3 + 5mn^2 - 8n^2 + 5m^2n - 32mn + 20n - 22m^2 + 44m - 16. \qquad (15)$$

Letting

$$w \equiv \frac{\frac{mF}{n}}{1+\frac{mF}{n}} \qquad (16)$$

gives a BETA DISTRIBUTION.

see also BETA DISTRIBUTION, CHI-SQUARED DISTRIBUTION, STUDENT'S t-DISTRIBUTION

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 536, 1987.

Snellius-Pothenot Problem

A SURVEYING PROBLEM which asks: Determine the position of an unknown accessible point P by its bearings from three inaccessible known points A, B, and C.

see also SURVEYING PROBLEMS

References

Dörrie, H. "Annex to a Survey." §40 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 193–197, 1965.

Snowflake

see EXTERIOR SNOWFLAKE, KOCH ANTISNOWFLAKE, KOCH SNOWFLAKE, PENTAFLAKE

Snub Cube

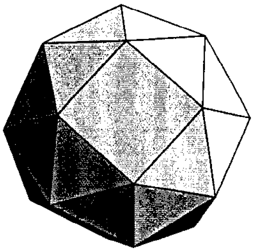

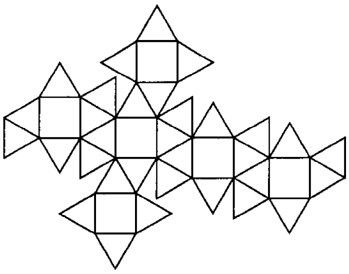

An ARCHIMEDEAN SOLID also called the SNUB CUBOCTAHEDRON whose VERTICES are the 24 points on the surface of a SPHERE for which the smallest distance between any two is as great as possible. It has two ENANTIOMERS, and its DUAL POLYHEDRON is the PENTAGONAL ICOSITETRAHEDRON. It has SCHLÄFLI SYMBOL $s\left\{\begin{smallmatrix}3\\4\end{smallmatrix}\right\}$. It is also UNIFORM POLYHEDRON U_{12} and has WYTHOFF SYMBOL $|\,2\,3\,4$. Its faces are $32\{3\} + 6\{4\}$.

The Inradius, Midradius, and Circumradius for $a = 1$ are

$$r = 1.157661791\ldots$$
$$\rho = 1.247223168\ldots$$
$$R = \tfrac{1}{2}\sqrt{\frac{x^2-8x+4}{x^2-5x+4}} = 1.3437133737446\ldots,$$

where

$$x \equiv (19 + 3\sqrt{33})^{1/3},$$

and the exact expressions for r and ρ can be computed using

$$r = \frac{R^2 - \frac{1}{4}a^2}{R}$$
$$\rho = \sqrt{R^2 - \tfrac{1}{4}a^2}.$$

see also Snub Dodecahedron

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 139, 1987.

Coxeter, H. S. M.; Longuet-Higgins, M. S.; and Miller, J. C. P. "Uniform Polyhedra." *Phil. Trans. Roy. Soc. London Ser. A* **246**, 401–450, 1954.

Snub Cuboctahedron

see Snub Cube

Snub Disphenoid

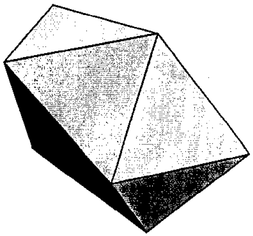

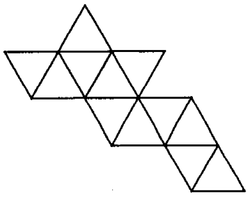

One of the convex Deltahedra also known as the Siamese Dodecahedron. It is Johnson Solid J_{84}.

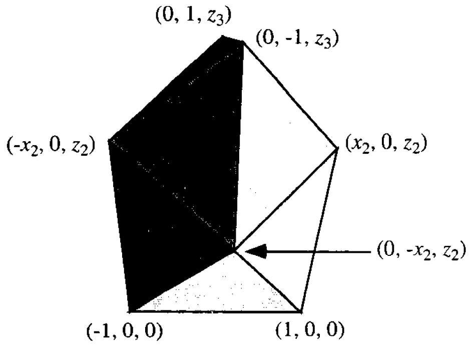

The coordinates of the Vertices may be found by solving the set of four equations

$$1 + {x_2}^2 + {z_1}^2 = 4$$
$$(x_2 - 1)^2 + (z_3 - z_1)^2 = 4$$
$${x_2}^2 + (z_3 - z_2)^2 = 4$$
$${x_2}^2 + {x_2}^2 + (z_2 - z_1)^2 = 4$$

for the four unknowns x_2, z_1, z_2, and z_3. Numerically,

$$x_2 = 1.28917$$
$$z_1 = 1.15674$$
$$z_2 = 1.97898$$
$$z_3 = 3.13572.$$

The analytic solution requires solving the Cubic Equation and gives

$$x_2 = 1 - 7\cdot 2^{-2/3}(1 - i\sqrt{3})\alpha^{-1} - \tfrac{1}{6}\cdot 2^{-1/3}(1 + i\sqrt{3})\alpha$$
$$z_1 = \tfrac{1}{3}\cdot 2^{-1/2}[-48 + 6\beta(1 + i\sqrt{3}) + \beta^2(1 - i\sqrt{3}) + 147\beta\gamma(\sqrt{3} - i) + 42\beta^2\gamma(\sqrt{3} + i)]^{1/2},$$

where

$$\alpha \equiv (12i\sqrt{237} - 54)^{1/3}$$
$$\beta \equiv 3^{1/3}(2i\sqrt{237} - 9)^{1/3}$$
$$\gamma \equiv (9i + 2\sqrt{237})^{-1}.$$

Snub Dodecadodecahedron

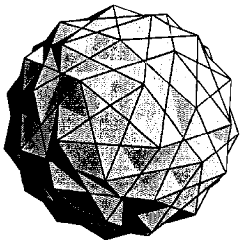

The Uniform Polyhedron U_{40} whose Dual Polyhedron is the Medial Pentagonal Hexecontahedron. It has Wythoff Symbol $|\,2\,\frac{5}{2}\,5$. Its faces are $12\{\frac{5}{2}\} + 60\{3\} + 12\{5\}$. It has Circumradius for $a = 1$ of

$$R = 1.27443994.$$

see also Snub Cube

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 139, 1987.

Coxeter, H. S. M.; Longuet-Higgins, M. S.; and Miller, J. C. P. "Uniform Polyhedra." *Phil. Trans. Roy. Soc. London Ser. A* **246**, 401–450, 1954.

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 174–176, 1971.

Snub Dodecahedron

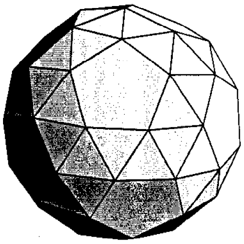

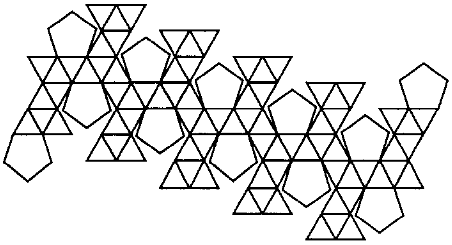

An ARCHIMEDEAN SOLID, also called the SNUB ICOSIDODECAHEDRON, whose DUAL POLYHEDRON is the PENTAGONAL HEXECONTAHEDRON. It has SCHLÄFLI SYMBOL $s\left\{\begin{smallmatrix}3\\5\end{smallmatrix}\right\}$. It is also UNIFORM POLYHEDRON U_{29} and has WYTHOFF SYMBOL $|\,2\,3\,5$. Its faces are $80\{3\}+12\{5\}$. For $a=1$, it has INRADIUS, MIDRADIUS, and CIRCUMRADIUS

$$\begin{aligned} r &= 2.039873155\ldots \\ \rho &= 2.097053835\ldots \\ R &= \frac{1}{2}\sqrt{\frac{8\cdot 2^{2/3}-16x+2^{1/3}x^2}{8\cdot 2^{2/3}-10x+2^{1/3}x^2}} \\ &= 2.15583737511564\ldots, \end{aligned}$$

where

$$x \equiv \left(49+27\sqrt{5}+3\sqrt{6}\sqrt{93+49\sqrt{5}}\right)^{1/3},$$

and the exact expressions for r and ρ can be computed using

$$\begin{aligned} r &= \frac{R^2-\frac{1}{4}a^2}{R} \\ \rho &= \sqrt{R^2-\tfrac{1}{4}a^2}. \end{aligned}$$

References

Coxeter, H. S. M.; Longuet-Higgins, M. S.; and Miller, J. C. P. "Uniform Polyhedra." *Phil. Trans. Roy. Soc. London Ser. A* **246**, 401–450, 1954.

Snub Icosidodecadodecahedron

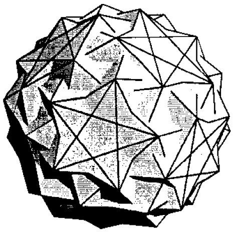

The UNIFORM POLYHEDRON U_{46} whose DUAL POLYHEDRON is the MEDIAL HEXAGONAL HEXECONTAHEDRON. It has WYTHOFF SYMBOL $|\,3\,\frac{5}{3}\,5$. Its faces are $12\{\frac{4}{2}\}+80\{3\}+12\{5\}$. It has CIRCUMRADIUS for $a=1$ of

$$\begin{aligned} R &= \frac{1}{2}\sqrt{\frac{2^{4/3}-14x+2^{2/3}x^2}{2^{4/3}-8x+2^{2/3}x^2}} \\ &= 1.12689791279994\ldots, \end{aligned}$$

where

$$x=(25+3\sqrt{69})^{1/3}.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 177–178, 1971.

Snub Icosidodecahedron

see SNUB DODECAHEDRON

Snub Polyhedron

A polyhedron with extra triangular faces, given by the SCHLÄFLI SYMBOL $s\left\{\begin{smallmatrix}p\\q\end{smallmatrix}\right\}$.

see also RHOMBIC POLYHEDRON, TRUNCATED POLYHEDRON

Snub Square Antiprism

see JOHNSON SOLID

Soap Bubble

see BUBBLE

Soccer Ball

see TRUNCATED ICOSAHEDRON

Sociable Numbers

Numbers which result in a periodic ALIQUOT SEQUENCE. If the period is 1, the number is called a PERFECT NUMBER. If the period is 2, the two numbers are called an AMICABLE PAIR. If the period is $t \geq 3$, the number is called sociable of order t. Only two sociable numbers were known prior to 1970, the sets of orders 5 and 28 discovered by Poulet (1918). In 1970, Cohen discovered nine groups of order 4.

The table below summarizes the number of sociable cycles known as given in the compilation by Moews (1995).

order	known
3	0
4	38
5	1
6	2
8	2
9	1
28	1

see also ALIQUOT SEQUENCE, PERFECT NUMBER, UNITARY SOCIABLE NUMBERS

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Item 61, Feb. 1972.

Borho, W. "Über die Fixpunkte der k-fach iterierten Teilerersummenfunktion." *Mitt. Math. Gesellsch. Hamburg* **9**, 34–48, 1969.

Cohen, H. "On Amicable and Sociable Numbers." *Math. Comput.* **24**, 423–429, 1970.

Devitt, J. S.; Guy, R. K.; and Selfridge, J. L. *Third Report on Aliquot Sequences,* Congr. Numer. XVIII, Proc. 6th Manitoba Conf. Numerical Math, pp. 177–204, 1976.

Flammenkamp, A. "New Sociable Numbers." *Math. Comput.* **56**, 871–873, 1991.

Gardner, M. "Perfect, Amicable, Sociable." Ch. 12 in *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, pp. 160–171, 1978.

Guy, R. K. "Aliquot Cycles or Sociable Numbers." §B7 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 62–63, 1994.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 145–146, 1979.

Moews, D. and Moews, P. C. "A Search for Aliquot Cycles Below 10^{10}." *Math. Comput.* **57**, 849–855, 1991.

Moews, D. and Moews, P. C. "A Search for Aliquot Cycles and Amicable Pairs." *Math. Comput.* **61**, 935–938, 1993.

Moews, D. "A List of Aliquot Cycles of Length Greater than 2." Rev. Dec. 18, 1995. `http://xraysgi.ims.uconn.edu:8080/sociable.txt`.

Poulet, P. Question 4865. *L'intermed. des Math.* **25**, 100–101, 1918.

te Riele, H. J. J. "Perfect Numbers and Aliquot Sequences." In *Computational Methods in Number Theory, Part I.* (Eds. H. W. Lenstra Jr. and R. Tijdeman). Amsterdam, Netherlands: Mathematisch Centrum, pp. 141–157, 1982.

Weisstein, E. W. "Sociable and Amicable Numbers." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Sociable.m`.

Social Choice Theory

The theory of analyzing a decision between a collection of alternatives made by a collection of n voters with separate opinions. Any choice for the entire group should reflect the desires of the individual voters to the extent possible.

Fair choice procedures usually satisfy ANONYMITY (invariance under permutation of voters), DUALITY (each alternative receives equal weight for a single vote), and MONOTONICITY (a change favorable for X does not hurt X). Simple majority vote is anonymous, dual, and monotone. MAY'S THEOREM states a stronger result.

see also ANONYMOUS, DUAL VOTING, MAY'S THEOREM, MONOTONIC VOTING, VOTING

References

Taylor, A. *Mathematics and Politics: Strategy, Voting, Power, and Proof.* New York: Springer-Verlag, 1995.

Young, S. C.; Taylor, A. D.; and Zwicker, W. S. "Counting Quota Systems: A Combinatorial Question from Social Choice Theory." *Math. Mag.* **68**, 331–342, 1995.

Socrates' Paradox

Socrates is reported to have stated: "One thing I know is that I know nothing."

see also LIAR'S PARADOX

References

Pickover, C. A. *Keys to Infinity.* New York: W. H. Freeman, p. 134, 1995.

Soddy Circles

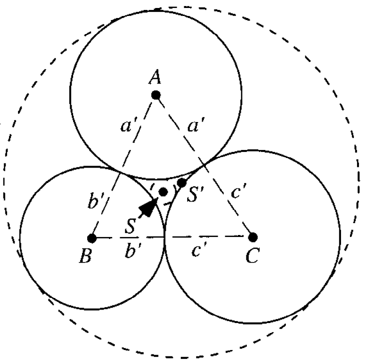

Given three distinct points A, B, and C, let three CIRCLES be drawn, one centered about each point and each one tangent to the other two. Call the RADII r_i ($r_3 = a'$, $r_1 = b'$, $r_2 = c'$). Then the CIRCLES satisfy

$$a' + b' = c \tag{1}$$

$$a' + c' = b \tag{2}$$

$$b' + c' = a, \tag{3}$$

as shown in the diagram below.

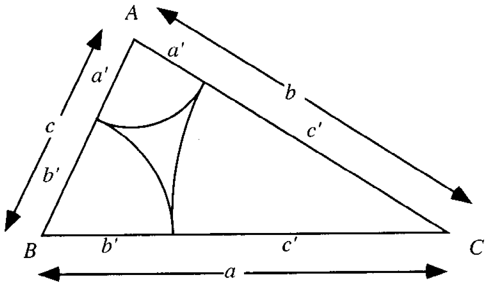

Solving for the RADII then gives

$$a' = \tfrac{1}{2}(b + c - a) \tag{4}$$

$$b' = \tfrac{1}{2}(a + c - b) \tag{5}$$

$$c' = \tfrac{1}{2}(a + b - c). \tag{6}$$

The above TRIANGLE has sides a, b, and c, and SEMIPERIMETER

$$s \equiv \tfrac{1}{2}(a + b + c). \tag{7}$$

Plugging in,

$$2s = (a' + b') + (a' + c') + (b' + c') = 2(a' + b' + c'), \tag{8}$$

giving

$$a' + b' + c' = s. \tag{9}$$

In addition,

$$a = b' + c' = a' + b' + c' - a' = s - a'. \tag{10}$$

Switching a and a' to opposite sides of the equation and noting that the above argument applies equally well to b' and c' then gives

$$a' = s - a \tag{11}$$

$$b' = s - b \tag{12}$$

$$c' = s - c. \tag{13}$$

As can be seen from the first figure, there exist exactly two nonintersecting CIRCLES which are TANGENT to all three CIRCLES. These are called the inner and outer Soddy circles (S and S', respectively), and their centers are called the inner and outer SODDY POINTS.

The inner Soddy circle is the solution to the FOUR COINS PROBLEM. The center S of the inner Soddy circle is the EQUAL DETOUR POINT, and the center of the outer Soddy circle S' is the ISOPERIMETRIC POINT (Kimberling 1994).

Frederick Soddy (1936) gave the FORMULA for finding the RADII of the Soddy circles (r_4) given the RADII r_i ($i = 1, 2, 3$) of the other three. The relationship is

$$2({\epsilon_1}^2 + {\epsilon_2}^2 + {\epsilon_3}^2 + {\epsilon_4}^2) = (\epsilon_1 + \epsilon_2 + \epsilon_3 + \epsilon_4)^2, \tag{14}$$

where $\epsilon_i \equiv \pm\kappa_i = \pm 1/r_i$ are the so-called BENDS, defined as the signed CURVATURES of the CIRCLES. If the contacts are all external, the signs are all taken as POSITIVE, whereas if one circle surrounds the other three, the sign of this circle is taken as NEGATIVE (Coxeter 1969). Using the QUADRATIC FORMULA to solve for ϵ_4, expressing in terms of radii instead of curvatures, and simplifying gives

$$r_4^\pm = \frac{r_1 r_2 r_3}{r_2 r_3 + r_1(r_2 + r_3) \pm 2\sqrt{r_1 r_2 r_3 (r_1 + r_2 + r_3)}}. \tag{15}$$

Here, the NEGATIVE solution corresponds to the outer Soddy circle and the POSITIVE one to the inner Soddy circle.

This FORMULA is called the DESCARTES CIRCLE THEOREM since it was known to Descartes. However, Soddy also extended it to SPHERES. Gosper has further extended the result to $n + 2$ mutually tangent n-D HYPERSPHERES, whose CURVATURES satisfy

$$\left(\sum_{i=0}^{n+1} \kappa_i\right)^2 - n \sum_{i=0}^{n+1} {\kappa_i}^2 = 0. \tag{16}$$

Solving for κ_{n+1} gives

$$\kappa_{n+1} = \frac{\sqrt{n}\sqrt{\left(\sum_{i=0}^n \kappa_i\right)^2 - (n-1)\sum_{i=0}^n {\kappa_i}^2} + \sum_{i=0}^n \kappa_i}{n-1}. \tag{17}$$

For (at least) $n = 2$ and 3, the RADICAL equals

$$f(n) V \kappa_0 \kappa_1 \cdots \kappa_n, \tag{18}$$

where V is the CONTENT of the SIMPLEX whose vertices are the centers of the $n+1$ independent HYPERSPHERES. The RADICAND can also become NEGATIVE, yielding an IMAGINARY κ_{n+1}. For $n = 3$, this corresponds to a sphere touching three large bowling balls and a small BB, all mutually tangent, which is an impossibility.

Bellew has derived a generalization applicable to a CIRCLE surrounded by n CIRCLES which are, in turn, circumscribed by another CIRCLE. The relationship is

$$\begin{aligned}
&[n(c_n - 1)^2 + 1] \sum_{i=1}^n {\kappa_i}^2 + n(3n{c_n}^2 - 2n - 6){c_n}^2(c_n - 1)^2 = \\
&\frac{1}{[n(c_n - 1) + 1]^2} \times \{n(c_n - 1)^2 + 1] \sum_{i=1}^n \kappa_i \\
&+ n c_n (c_n - 1)(n{c_n}^2 + (3 - n)c_n - 4])\},
\end{aligned} \tag{19}$$

where

$$c_n \equiv \csc\left(\frac{\pi}{n}\right). \tag{20}$$

For $n = 3$, this simplifies to the Soddy formula.

see also APOLLONIUS CIRCLES, APOLLONIUS' PROBLEM, ARBELOS, BEND (CURVATURE), CIRCUMCIRCLE, DESCARTES CIRCLE THEOREM, FOUR COINS PROBLEM, HART'S THEOREM, PAPPUS CHAIN, SPHERE PACKING, STEINER CHAIN

References

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 13–14, 1969.

Elkies, N. D. and Fukuta, J. "Problem E3236 and Solution." *Amer. Math. Monthly* **97**, 529–531, 1990.

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, p. 181, 1994.

"The Kiss Precise." *Nature* **139**, 62, 1937.

Soddy, F. "The Kiss Precise." *Nature* **137**, 1021, 1936.

Vandeghen, A. "Soddy's Circles and the De Longchamps Point of a Triangle." *Amer. Math. Monthly* **71**, 176–179, 1964.

Soddy's Hexlet

see HEXLET

Soddy Line

A LINE on which the INCENTER I, GERGONNE POINT Ge, and inner and outer SODDY POINTS S and S' lie (the latter two of which are the EQUAL DETOUR POINT and the ISOPERIMETRIC POINT). The Soddy line can be given parametrically by

$$I + \lambda Ge,$$

where λ is a parameter. It is also given by

$$\sum (f-e)\alpha = 0,$$

where cyclic permutations of d, e, and f are taken and the sum is over TRILINEAR COORDINATES α, β, and γ.

λ	Center
-4	outer Griffiths point *Gr'*
-2	outer Oldknow point *Ol'*
$-\frac{4}{3}$	outer Rigby point *Ri'*
-1	outer Soddy center S'
0	incenter I
1	inner Soddy center S
$\frac{4}{3}$	inner Rigby point *Ri*
2	inner Oldknow point *Ol*
4	inner Griffiths point *Gr*
∞	Gergonne point

S', I, S, and Ge are COLLINEAR and form a HARMONIC RANGE (Vandeghen 1964, Oldknow 1996). There are a total of 22 HARMONIC RANGES for sets of four points out of these 10 (Oldknow 1996).

The Soddy line intersects the EULER LINE in the DE LONGCHAMPS POINT, and the GERGONNE LINE in the FLETCHER POINT.

see also DE LONGCHAMPS POINT, EULER LINE, FLETCHER POINT, GERGONNE POINT, GRIFFITHS POINTS, HARMONIC RANGE, INCENTER, OLDKNOW POINTS, RIGBY POINTS, SODDY POINTS

References

Oldknow, A. "The Euler-Gergonne-Soddy Triangle of a Triangle." *Amer. Math. Monthly* **103**, 319–329, 1996.

Vandeghen, A. "Soddy's Circles and the De Longchamps Point of a Triangle." *Amer. Math. Monthly* **71**, 176–179, 1964.

Soddy Points

Given three mutually tangent CIRCLES, there exist exactly two nonintersecting CIRCLES TANGENT to all three CIRCLES. These are called the inner and outer SODDY CIRCLES, and their centers are called the inner and outer Soddy points. The outer Soddy circle is the solution to the FOUR COINS PROBLEM. The center S of the inner Soddy circle is the EQUAL DETOUR POINT, and the center of the outer Soddy circle S' is the ISOPERIMETRIC POINT (Kimberling 1994).

see also EQUAL DETOUR POINT, ISOPERIMETRIC POINT, SODDY CIRCLES

References

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, p. 181, 1994.

Sofa Constant

see MOVING SOFA CONSTANT

Sol Geometry

The GEOMETRY of the LIE GROUP R SEMIDIRECT PRODUCT with R^2, where R acts on R^2 by $(t,(x,y)) \to (e^t x, e^{-t} y)$.

see also THURSTON'S GEOMETRIZATION CONJECTURE

Soldner's Constant

Consider the following formulation of the PRIME NUMBER THEOREM,

$$\pi(x) = \sum \frac{\mu(m)}{m} \int_c^x \frac{dt}{\ln t},$$

where $\mu(m)$ is the MÖBIUS FUNCTION and c (sometimes also denoted μ) is Soldner's constant. Ramanujan found $c = 1.45136380\ldots$ (Hardy 1969, Le Lionnais 1983, Berndt 1994). Soldner (cited in Nielsen 1965) derived the correct value of c as $1.4513692346\ldots$, where c is the root of

$$L(x) = \lim_{\epsilon \to 0} \int_0^{1-\epsilon} \frac{dt}{\ln t} + \int_{1+\epsilon}^{\infty} \frac{dt}{\ln t}$$

(Le Lionnais 1983).

References

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 123–124, 1994.

Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, p. 45, 1959.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 39, 1983.

Nielsen, N. *Théorie des Integrallogarithms.* New York: Chelsea, p. 88, 1965.

Solenoidal Field

A solenoidal VECTOR FIELD satisfies

$$\nabla \cdot \mathbf{B} = 0 \tag{1}$$

for every VECTOR $\mathbf{B}$, where $\nabla \cdot \mathbf{B}$ is the DIVERGENCE. If this condition is satisfied, there exists a vector $\mathbf{A}$, known as the VECTOR POTENTIAL, such that

$$\mathbf{B} \equiv \nabla \times \mathbf{A}, \tag{2}$$

where $\nabla \times \mathbf{A}$ is the CURL. This follows from the vector identity

$$\nabla \cdot \mathbf{B} = \nabla \cdot (\nabla \times \mathbf{A}) = 0. \tag{3}$$

If $\mathbf{A}$ is an IRROTATIONAL FIELD, then

$$\mathbf{A} \times \mathbf{r} \tag{4}$$

is solenoidal. If $\mathbf{u}$ and $\mathbf{v}$ are irrotational, then

$$\mathbf{u} \times \mathbf{v} \tag{5}$$

is solenoidal. The quantity

$$(\nabla u) \times (\nabla v), \tag{6}$$

where ∇u is the GRADIENT, is always solenoidal. For a function ϕ satisfying LAPLACE'S EQUATION

$$\nabla^2 \phi = 0, \tag{7}$$

it follows that $\nabla \phi$ is solenoidal (and also IRROTATIONAL).

see also BELTRAMI FIELD, CURL, DIVERGENCE, DIVERGENCELESS FIELD, GRADIENT, IRROTATIONAL FIELD, LAPLACE'S EQUATION, VECTOR FIELD

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 1084, 1980.

Solid

A closed 3-D figure (which may, according to some terminology conventions, be self-intersecting). Among the simplest solids are the SPHERE, CUBE, CONE, CYLINDER, and more generally, the POLYHEDRA.

see also APPLE, ARCHIMEDEAN SOLID, CATALAN SOLID, CONE, CORK PLUG, CUBE, CUBOCTAHEDRON, CYLINDER, CYLINDRICAL HOOF, CYLINDRICAL WEDGE, DODECAHEDRON, GEODESIC DOME, GREAT DODECAHEDRON, GREAT ICOSAHEDRON, GREAT RHOMBICOSIDODECAHEDRON (ARCHIMEDEAN), GREAT RHOMBICUBOCTAHEDRON (ARCHIMEDEAN), GREAT STELLATED DODECAHEDRON, ICOSAHEDRON, ICOSIDODECAHEDRON, JOHNSON SOLID, KEPLER-POINSOT SOLID, LEMON, MÖBIUS STRIP, OCTAHEDRON, PLATONIC SOLID, POLYHEDRON, PSEUDOSPHERE, RHOMBICOSIDODECAHEDRON, RHOMBICUBOCTAHEDRON, SMALL STELLATED DODECAHEDRON, SNUB CUBE, SNUB DODECAHEDRON, SOLID OF REVOLUTION, SPHERE, STEINMETZ SOLID, STELLA OCTANGULA, SURFACE, TETRAHEDRON, TORUS, TRUNCATED CUBE, TRUNCATED DODECAHEDRON, TRUNCATED ICOSAHEDRON, TRUNCATED OCTAHEDRON, TRUNCATED TETRAHEDRON, UNIFORM POLYHEDRON, WULFF SHAPE

Solid Angle

Defined as the SURFACE AREA Ω of a UNIT SPHERE which is subtended by a given object S. Writing the SPHERICAL COORDINATES as ϕ for the COLATITUDE (angle from the pole) and θ for the LONGITUDE (azimuth),

$$\Omega \equiv A_{\text{projected}} = \iint_S \sin\phi \, d\theta \, d\phi.$$

Solid angle is measured in STERADIANS, and the solid angle corresponding to all of space being subtended is 4π STERADIANS.

see also SPHERE, STERADIAN

Solid Geometry

That portion of GEOMETRY dealing with SOLIDS, as opposed to PLANE GEOMETRY. Solid geometry is concerned with POLYHEDRA, SPHERES, 3-D SOLIDS, lines in 3-space, PLANES, and so on.

see also GEOMETRY, PLANE GEOMETRY, SPHERICAL GEOMETRY

References

Altshiller-Court, N. *Modern Pure Solid Geometry.* New York: Chelsea, 1979.

Bell, R. J. T. *An Elementary Treatise on Coordinate Geometry of Three Dimensions.* London: Macmillan, 1926.

Cohn, P. M. *Solid Geometry.* New York: Routledge, 1968.

Frost, P. *Solid Geometry, 3rd ed.* London: Macmillan, 1886.

Lines, L. *Solid Geometry.* New York: Dover, 1965.

Salmon, G. *Treatise on the Analytic Geometry of Three Dimensions, 6th ed.* London: Longmans Green, 1914.

Shute, W. G.; Shirk, W. W.; and Porter, G. F. *Solid Geometry.* New York: American Book Co., 1960.

Wentworth, G. A. and Smith, D. E. *Solid Geometry.* Boston, MA: Ginn and Company, 1913.

Solid Partition

Solid partitions are generalizations of PLANE PARTITIONS. MacMohan (1960) conjectured the GENERATING FUNCTION for the number of solid partitions was

$$f(z) = \frac{1}{(1-z)(1-z^2)^3(1-z^3)^6(1-z^4)^{10}\cdots},$$

but this was subsequently shown to disagree at $n = 6$ (Atkin *et al.* 1967). Knuth (1970) extended the tabulation of values, but was unable to find a correct generating function. The first few values are 1, 4, 10, 26, 59, 140, ... (Sloane's A000293).

References

Atkin, A. O. L.; Bratley, P.; MacDonald, I. G.; and McKay, J. K. S. "Some Computations for m-Dimensional Partitions." *Proc. Cambridge Philos. Soc.* **63**, 1097–1100, 1967.

Knuth, D. E. "A Note on Solid Partitions." *Math. Comput.* **24**, 955–961, 1970.

MacMahon, P. A. "Memoir on the Theory of the Partitions of Numbers. VI: Partitions in Two-Dimensional Space, to which is Added an Adumbration of the Theory of Partitions in Three-Dimensional Space." *Phil. Trans. Roy. Soc. London Ser. A* **211**, 345–373, 1912b.

MacMahon, P. A. *Combinatory Analysis, Vol. 2.* New York: Chelsea, pp. 75–176, 1960.

Sloane, N. J. A. Sequence A3392/M000293 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Solid of Revolution

To find the VOLUME of a solid of rotation by adding up a sequence of thin cylindrical shells, consider a region bounded above by $y = f(x)$, below by $y = g(x)$, on the left by the LINE $x = a$, and on the right by the LINE $x = b$. When the region is rotated about the y-AXIS, the resulting VOLUME is given by

$$V = 2\pi \int_b^a x[f(x) - g(x)]\, dx.$$

To find the volume of a solid of rotation by adding up a sequence of thin flat disks, consider a region bounded above by $y = f(x)$, below by $y = g(x)$, on the left by the Line $x = a$, and on the right by the Line $x = b$. When the region is rotated about the x-Axis, the resulting Volume is

$$V = \pi \int_b^a \{[f(x)]^2 - [g(x)]^2\}\, dx.$$

see also Surface of Revolution, Volume

Solidus

The diagonal slash "/" used to denote Division for in-line equations such as a/b, $1/(x-1)^2$, etc. The solidus is also called a Diagonal.

see also Division, Obelus

Solitary Number

A number which does not have any Friends. Solitary numbers include all Primes and Powers of Primes. More generally, numbers for which $(n, \sigma(n)) = 1$ are solitary, where (a, b) is the Greatest Common Divisor of a and b and $\sigma(n)$ is the Divisor Function. The first few solitary numbers are 1, 2, 3, 4, 5, 7, 8, 9, 11, 13, 16, 17, 19, 21, ... (Sloane's A014567).

see also Friend

References

Anderson, C. W. and Hickerson, D. Problem 6020. "Friendly Integers." *Amer. Math. Monthly* **84**, 65–66, 1977.

Sloane, N. J. A. Sequence A014567 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Soliton

A stable isolated (i.e., solitary) traveling wave solution to a set of equations.

see also Lax Pair, Sine-Gordon Equation

References

Bullough, R. K. and Caudrey, P. J. (Eds.). *Solitons.* Berlin: Springer-Verlag, 1980.

Dodd, R. K. *Solitons and Nonlinear Equations.* London: Academic Press, 1984.

Drazin, P. G. and Johnson, R. S. *Solitons: An Introduction.* Cambridge, England: Cambridge University Press, 1988.

Filippov, A. *The Versatile Solitons.* Boston, MA: Birkhäuser, 1996.

Gu, C. H. *Soliton Theory and Its Applications.* New York: Springer-Verlag, 1995.

Infeld, E. and Rowlands, G. *Nonlinear Waves, Solitons, and Chaos.* Cambridge, England: Cambridge University Press, 1990.

Lamb, G. L. Jr. *Elements of Soliton Theory.* New York: Wiley, 1980.

Makhankov, V. G.; Fedyann, V. K.; and Pashaev, O. K. (Eds.). *Solitons and Applications.* Singapore: World Scientific, 1990.

Newell, A. C. *Solitons in Mathematics and Physics.* Philadelphia, PA: SIAM, 1985.

Olver, P. J. and Sattinger, D. H. (Eds.). *Solitons in Physics, Mathematics, and Nonlinear Optics.* New York: Springer-Verlag, 1990.

Remoissent, M. *Waves Called Solitons, 2nd ed.* New York: Springer-Verlag, 1996.

Solomon's Seal Knot

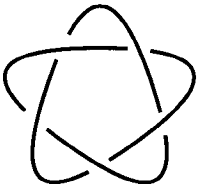

The (5,2) Torus Knot 05_{001} with Braid Word $\sigma_1{}^5$.

Solomon's Seal Lines

The 27 Real or Imaginary straight Lines which lie on the general Cubic Surface and the 45 triple tangent Planes to the surface. All are related to the 28 Bitangents of the general Quartic Curve.

Schoutte (1910) showed that the 27 lines can be put into a One-to-One correspondence with the vertices of a particular Polytope in 6-D space in such a manner that all incidence relations between the lines are mirrored in the connectivity of the Polytope and conversely (Du Val 1931). A similar correspondence can be made between the 28 bitangents and a 7-D Polytope (Coxeter 1928) and between the tritangent planes of the canonical curve of genus four and an 8-D Polytope (Du Val 1933).

see also Brianchon's Theorem, Cubic Surface, Double Sixes, Pascal's Theorem, Quartic Surface, Steiner Set

References

Bell, E. T. *The Development of Mathematics, 2nd ed.* New York: McGraw-Hill, pp. 322–325, 1945.

Coxeter, H. S. M. "The Pure Archimedean Polytopes in Six and Seven Dimensions." *Proc. Cambridge Phil. Soc.* **24**, 7–9, 1928.

Du Val, P. "On the Directrices of a Set of Points in a Plane." *Proc. London Math. Soc. Ser. 2* **35**, 23–74, 1933.

Schoutte, P. H. "On the Relation Between the Vertices of a Definite Sixdimensional Polytope and the Lines of a Cubic Surface." *Proc. Roy. Akad. Acad. Amsterdam* **13**, 375–383, 1910.

Solomon's Seal Polygon

see Hexagram

Solvable Congruence

A Congruence that has a solution.

Solvable Group

A solvable group is a group whose composition indices are all Prime Numbers. Equivalently, a solvable is a Group having a "normal series" such that each "normal factor" is Abelian. The term solvable derives from this type of group's relationship to Galois's Theorem, namely that the Symmetric Group S_n is insoluble for $n \geq 5$ while it is solvable for $n = 1$, 2, 3, and 4. As a result, the Polynomial equations of degree ≥ 5 are not solvable using finite additions, multiplications, divisions, and root extractions.

Every FINITE GROUP of order < 60, every ABELIAN GROUP, and every SUBGROUP of a solvable group is solvable.

see also ABELIAN GROUP, COMPOSITION SERIES, GALOIS'S THEOREM, SYMMETRIC GROUP

References

Lomont, J. S. *Applications of Finite Groups.* New York: Dover, p. 26, 1993.

Solvable Lie Group

The connected closed SUBGROUPS (up to an ISOMORPHISM) of COMPLEX MATRICES that are closed under conjugate transpose and have a discrete finite center. Examples include SPECIAL LINEAR GROUPS, SYMPLECTIC GROUPS, and certain isometry groups of QUADRATIC FORMS.

see also LIE GROUP

References

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Soma Cube

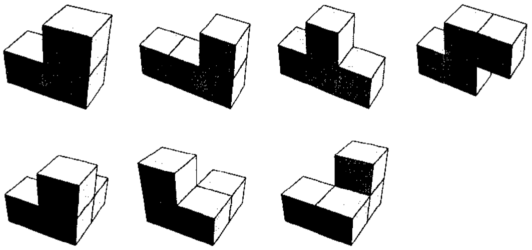

A solid DISSECTION puzzle invented by Piet Hein during a lecture on Quantum Mechanics by Werner Heisenberg. There are seven soma pieces composed of all the *irregular* face-joined cubes (POLYCUBES) with ≤ 4 cubes. The object is to assemble the pieces into a CUBE. There are 240 essentially distinct ways of doing so (Beeler *et al.* 1972, Berlekamp *et al.* 1982), as first enumerated one rainy afternoon in 1961 by J. H. Conway and Mike Guy.

A commercial version of the cube colors the pieces black, green, orange, white, red, and blue. When the 48 symmetries of the cube, three ways of assembling the black piece, and 2^5 ways of assembling the green, orange, white, red, and blue pieces are counted, the total number of solutions rises to 1,105,920.

see also CUBE DISSECTION, POLYCUBE

References

Albers, D. J. and Alexanderson, G. L. (Eds.). *Mathematical People: Profiles and Interviews.* Boston, MA: Birkhäuser, p. 43, 1985.

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 112–113, 1987.

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Item 112, Feb. 1972.

Berlekamp, E. R.; Conway, J. H.; and Guy, R. K. Ch. 24 in *Winning Ways, For Your Mathematical Plays, Vol. 2: Games in Particular.* London: Academic Press, 1982.

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 203–205, 1989.

Gardner, M. Ch. 6 in *The Second Scientific American Book of Mathematical Puzzles & Diversions: A New Selection.* New York: Simon and Schuster, pp. 65–77, 1961.

Steinhaus, H. *Mathematical Snapshots, 3rd American ed.* New York: Oxford University Press, pp. 168–169, 1983.

Somer-Lucas Pseudoprime

An ODD COMPOSITE NUMBER N is called a Somer-Lucas d-pseudoprime (with $d \geq 1$) if there EXISTS a nondegenerate LUCAS SEQUENCE $U(P,Q)$ with $U_0 = 0$, $U_1 = 1$, $D = P^2 - 4Q$, such that $(N,D) = 1$ and the rank appearance of N in the sequence $U(P,Q)$ is $(1/a)(N - (D/N))$, where (D/N) denotes the JACOBI SYMBOL.

see also LUCAS SEQUENCE, PSEUDOPRIME

References

Ribenboim, P. "Somer-Lucas Pseudoprimes." §2.X.D in *The New Book of Prime Number Records, 3rd ed.* New York: Springer-Verlag, pp. 131–132, 1996.

Sommerfeld's Formula

There are (at least) two equations known as Sommerfeld's formula. The first is

$$J_\nu(z) = \frac{1}{2\pi} \int_{-\eta+i\infty}^{2\pi-\eta+i\infty} e^{iz\cos t} e^{i\nu(t-\pi/2)}\,dt,$$

where $J_\nu(z)$ is a BESSEL FUNCTION OF THE FIRST KIND. The second states that under appropriate restrictions,

$$\int_0^\infty J_0(\tau r) e^{-|x|\sqrt{\tau^2-k^2}} \frac{\tau\,d\tau}{\sqrt{\tau^2-k^2}} = \frac{e^{ik\sqrt{\tau^2+k^2}}}{\sqrt{r^2+x^2}}.$$

see also WEYRICH'S FORMULA

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 1472 and 1474, 1980.

Somos Sequence

The Somos sequences are a set of related symmetrical RECURRENCE RELATIONS which, surprisingly, always give integers. The Somos sequence of order k is defined by

$$a_n = \frac{\sum_{j=1}^{\lfloor k/2 \rfloor} a_{n-j} a_{n-(k-j)}}{a_{n-k}},$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION and $a_j = 1$ for $j = 0$, …, $k-1$. The 2- and 3-Somos sequences consist entirely of 1s. The k-Somos sequences for $k = 4$, 5, 6, and 7 are

$$a_n = \frac{a_{n-1}a_{n-3} + {a_{n-2}}^2}{a_{n-4}}$$
$$a_n = \frac{a_{n-1}a_{n-4} + a_{n-2}a_{n-3}}{a_{n-5}}$$
$$a_n = \frac{1}{a_{n-6}}[a_{n-1}a_{n-5} + a_{n-2}a_{n-4} + {a_{n-3}}^2]$$
$$a_n = \frac{1}{a_{n-7}}[a_{n-1}a_{n-6} + a_{n-2}a_{n-5} + a_{n-3}a_{n-4}],$$

giving 1, 1, 1, 2, 3, 7, 23, 59, 314, 1529, … (Sloane's A006720), 1, 1, 1, 1, 2, 3, 5, 11, 37, 83, 274, 1217, … (Sloane's A006721), 1, 1, 1, 1, 1, 3, 5, 9, 23, 75, 421, 1103, … (Sloane's A006722), 1, 1, 1, 1, 1, 1, 3, 5, 9, 17, 41, 137, 769, … (Sloane's A006723). Gale (1991) gives simple proofs of the integer-only property of the 4-Somos and 5-Somos sequences. Hickerson proved 6-Somos generates only integers using computer algebra, and empirical evidence suggests 7-Somos is also integer-only.

However, the k-Somos sequences for $k \geq 8$ do not give integers. The values of n for which a_n first becomes nonintegral for the k-Somos sequence for $k = 8$, 9, … are 17, 19, 20, 22, 24, 27, 28, 30, 33, 34, 36, 39, 41, 42, 44, 46, 48, 51, 52, 55, 56, 58, 60, … (Sloane's A030127).

see also GÖBEL'S SEQUENCE, HERONIAN TRIANGLE

References

Buchholz, R. H. and Rathbun, R. L. "An Infinite Set of Heron Triangles with Two Rational Medians." *Amer. Math. Monthly* **104**, 107–115, 1997.

Gale, D. "Mathematical Entertainments: The Strange and Surprising Saga of the Somos Sequences." *Math. Intel.* **13**, 40–42, 1991.

Sloane, N. J. A. Sequences A030127, A006720/M0857, A006721/M0735, A006722/M2457, and A006723/M2456 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sondat's Theorem

The PERSPECTIVE AXIS bisects the line joining the two ORTHOCENTERS.

see also ORTHOCENTER, PERSPECTIVE AXIS

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, p. 259, 1929.

Sonine's Integral

$$J_m(x) = \frac{2x^{m-n}}{2^{m-n}\Gamma(m-n)} \int_0^1 J_n(xt)t^{n+1}(1-t^2)^{m-n-1}\,dt,$$

where $J_m(x)$ is a BESSEL FUNCTION OF THE FIRST KIND and $\Gamma(x)$ is the GAMMA FUNCTION.

see also HANKEL'S INTEGRAL, POISSON INTEGRAL

Sonine Polynomial

A polynomial which differs from the associated LAGUERRE POLYNOMIAL by only a normalization constant,

$$\begin{aligned} S_r^s(x) &= \frac{1}{s!}e^x x^{-r}\frac{d^s}{dx^s}(e^{-x}x^{r+s}) = \frac{(-1)^r}{(r+s)!}L_{r+s}^r(x) \\ &= \frac{x^s}{s!(r+s)!0!} - \frac{x^{s-1}}{(s-1)!(r+s-1)!1!} \\ &\quad + \frac{x^{r-2}}{(r-2)!(r+s-2)!2!} - \ldots \\ &= \frac{1}{s!(r+s)!}x^{-(r+1)/2}e^{x/2}W_{s+r/2+1/2,r/2}(x), \end{aligned}$$

where $W_{k,m}(z)$ is a WHITTAKER FUNCTION.

see also LAGUERRE POLYNOMIAL, WHITTAKER FUNCTION

Sonine-Schafheitlin Formula

$$\begin{aligned} &\int_0^\infty J_\mu(at)J_\nu(bt)t^{-\lambda}\,dt \\ &\quad = \frac{a^\mu\Gamma[(\mu+\nu-\lambda+1)/2]}{2^\lambda b^{\mu-\lambda+1}\Gamma[(-\mu+\nu+\lambda+1)/2]\Gamma(\mu+1)} \\ &\times {}_2F_1((\mu+\nu-\lambda+1)/2, (\mu-\nu-\lambda+1)/2; \mu+1; a^2/b^2), \end{aligned}$$

where $\Re[\mu+\nu-\lambda+1] > 0$, $\Re[\lambda] > -1$, $0 < a < b$, $J_\nu(x)$ is a BESSEL FUNCTION OF THE FIRST KIND, $\Gamma(x)$ is the GAMMA FUNCTION, and ${}_2F_1(a, b; c; x)$ is a HYPERGEOMETRIC FUNCTION.

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1474, 1980.

Sophie Germain Prime

A PRIME p is said to be a Sophie Germain prime if both p and $2p+1$ are PRIME. The first few Sophie Germain primes are 2, 3, 5, 11, 23, 29, 41, 53, 83, 89, 113, 131, … (Sloane's A005384).

Around 1825, Sophie Germain proved that the first case of FERMAT'S LAST THEOREM is true for such primes, i.e., if p is a Sophie Germain prime, there do not exist INTEGERS x, y, and z different from 0 and not multiples of p such that

$$x^p + y^p = z^p.$$

Sophie Germain primes p of the form $p = k \cdot 2^n - 1$ (which makes $2p+1$ a PRIME) are COMPOSITE MERSENNE NUMBERS. Since the largest known COMPOSITE MERSENNE NUMBER is M_p with $p = 39051 \times 2^{6001} - 1$, p is the largest known Sophie Germain prime.

see also CUNNINGHAM CHAIN, FERMAT'S LAST THEOREM, MERSENNE NUMBER, TWIN PRIMES

References
Dubner, H. "Large Sophie Germain Primes." *Math. Comput.* **65**, 393–396, 1996.
Ribenboim, P. "Sophie Germane Primes." §5.2 in *The New Book of Prime Number Records.* New York: Springer-Verlag, pp. 329–332, 1996.
Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 154–157, 1993.
Sloane, N. J. A. Sequence A005384 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sorites Paradox

Sorites paradoxes are a class of paradoxical arguments also known as little-by-little arguments. The name "sorites" derives from the Greek word *soros*, meaning "pile" or "heap". Sorites paradoxes are exemplified by the problem that a single grain of wheat does not comprise a heap, nor do two grains of wheat, three grains of wheat, etc. However, at some point, the collection of grains becomes large enough to be called a heap, but there is apparently no definite point where this occurs.

see also UNEXPECTED HANGING PARADOX

Sort-Then-Add Sequence

A sequence produced by sorting the digits of a number and adding them to the previous number. The algorithm terminates when a sorted number is obtained. For $n = 1, 2, \ldots$, the algorithm terminates on 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 11, 12, 13, 14, 15, 16, 17, 18, 19, 22, 33, ... (Sloane's A033862). The first few numbers not known to terminate are 316, 452, 697, 1376, 2743, 5090, ... (Sloane's A033861). The least numbers of sort-then-add persistence $n = 1, 2, \ldots$, are 1, 10, 65, 64, 175, 98, 240, 325, 302, 387, 198, 180, 550, ... (Sloane's A033863).

see also 196-ALGORITHM, RATS SEQUENCE

References
Sloane, N. J. A. Sequences A033861, A033862, and A033863 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sorting

Sorting is the rearrangement of numbers (or other orderable objects) in a list into their correct lexographic order. Alphabetization is therefore a form of sorting. Because of the extreme importance of sorting in almost all database applications, a great deal of effort has been expended in the creation and analysis of efficient sorting algorithms.

see also HEAPSORT, ORDERING, QUICKSORT

References
Knuth, D. E. *The Art of Computer Programming, Vol. 3: Sorting and Searching, 2nd ed.* Reading, MA: Addison-Wesley, 1973.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Sorting." Ch. 8 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 320–339, 1992.

Source

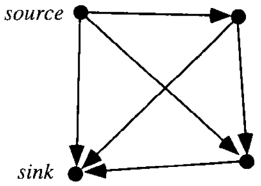

A vertex of a DIRECTED GRAPH with no entering edges.

see also DIRECTED GRAPH, NETWORK, SINK (DIRECTED GRAPH)

Sous-Double

A MULTIPERFECT NUMBER P_3. Six sous-doubles are known, and these are believed to comprise all sous-doubles.

see also MULTIPERFECT NUMBER, SOUS-TRIPLE

Souslin's Hypothesis

Every dense linear order complete set without endpoints having at most ω disjoint intervals is order isomorphic to the CONTINUUM of REAL NUMBERS, where ω is the set of NATURAL NUMBERS.

References
Iyanaga, S. and Kawada, Y. (Eds.). "Souslin's Hypothesis." §35E.4 in in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 137, 1980.

Souslin Set

The continuous image of a POLISH SPACE, also called an ANALYTIC SET.

see also ANALYTIC SET, POLISH SPACE

Sous-Triple

A MULTIPERFECT NUMBER P_4. 36 sous-triples are known, and these are believed to comprise all sous-triples.

see also MULTIPERFECT NUMBER, SOUS-DOUBLE

Space

The concept of a space is an extremely general and important mathematical construct. Members of the space obey certain addition properties. Spaces which have been investigated and found to be of interest are usually named after one or more of their investigators. This practice unfortunately leads to names which give very little insight into the relevant properties of a given space.

One of the most general type of mathematical spaces is the TOPOLOGICAL SPACE.

see also AFFINE SPACE, BAIRE SPACE, BANACH SPACE, BASE SPACE, BERGMAN SPACE, BESOV SPACE, BOREL SPACE, CALABI-YAU SPACE, CELLULAR SPACE, CHU SPACE, DODECAHEDRAL SPACE, DRINFELD'S SYMMETRIC SPACE, EILENBERG-MAC LANE SPACE, EUCLIDEAN SPACE, FIBER SPACE, FINSLER SPACE, First-Countable Space, Fréchet Space, Function Space, G-Space, Green Space, Hausdorff Space, Heisenberg Space, Hilbert Space, Hyperbolic Space, Inner Product Space, L_2-Space, Lens Space, Line Space, Linear Space, Liouville Space, Locally Convex Space, Locally Finite Space, Loop Space, Mapping Space, Measure Space, Metric Space, Minkowski Space, Müntz Space, Non-Euclidean Geometry, Normed Space, Paracompact Space, Planar Space, Polish Space, Probability Space, Projective Space, Quotient Space, Riemann's Moduli Space, Riemann Space, Sample Space, Standard Space, State Space, Stone Space, Teichmüller Space, Tensor Space, Topological Space, Topological Vector Space, Total Space, Vector Space

Space of Closed Paths

see Loop Space

Space Conic

see Skew Conic

Space Curve

A curve which may pass through any region of 3-D space, as contrasted to a Plane Curve which must lie in a single Plane. Von Staudt (1847) classified space curves geometrically by considering the curve

$$\phi : I \to \mathbb{R}^3 \qquad (1)$$

at $t_0 = 0$ and assuming that the parametric functions $\phi_i(t)$ for $i = 1, 2, 3$ are given by Power Series which converge for small t. If the curve is contained in no Plane for small t, then a coordinate transformation puts the parametric equations in the normal form

$$\phi_1(t) = t^{1+k_1} + \ldots \qquad (2)$$

$$\phi_2(t) = t^{2+k_1+k_2} + \ldots \qquad (3)$$

$$\phi_3(t) = t^{3+k_1+k_2+k_3} + \ldots \qquad (4)$$

for integers k_1, k_2, $k_3 \geq 0$, called the local numerical invariants.

see also Curve, Cyclide, Fundamental Theorem of Space Curves, Helix, Plane Curve, Seifert's Spherical Spiral, Skew Conic, Space-Filling Function, Spherical Spiral, Surface, Viviani's Curve

References

do Carmo, M.; Fischer, G.; Pinkall, U.; and Reckziegel, H. "Singularities of Space Curves." §3.1 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 24–25, 1986.

Fine, H. B. "On the Singularities of Curves of Double Curvature." *Amer. J. Math.* **8**, 156–177, 1886.

Fischer, G. (Ed.). Plates 57–64 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 58–59, 1986.

Gray, A. "Curves in $\mathbb{R}^n$" and "Curves in Space." §1.2 and Ch. 7 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 4–6 and 123–151, 1993.

Griffiths, P. and Harris, J. *Principles of Algebraic Geometry.* New York: Wiley, 1978.

Saurel, P. "On the Singularities of Tortuous Curves." *Ann. Math.* **7**, 3–9, 1905.

Staudt, C. von. *Geometrie der Lage.* Nürnberg, Germany, 1847.

Wiener, C. "Die Abhängigkeit der Rückkehrelemente der Projektion einer unebenen Curve von deren der Curve selbst." *Z. Math. & Phys.* **25**, 95–97, 1880.

Space Diagonal

The Line Segment connecting opposite Vertices (i.e., two Vertices which do not share a common face) in a Parallelepiped or other similar solid.

see also Diagonal (Polygon), Diagonal (Polyhedron), Euler Brick

Space Distance

The maximum distance in 3-D can occur no more than $2n - 2$ times. Also, there exists a fixed number c such that no distance determined by a set of n points in 3-D space occurs more than $cn^{5/3}$ times. The maximum distance can occur no more than $\left\lfloor \frac{1}{4}n^2 \right\rfloor$ times in 4-D, where $\lfloor x \rfloor$ is the Floor Function.

References

Honsberger, R. *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., pp. 122–123, 1976.

Space Division

The number of regions into which space can be divided by n Spheres is

$$N = \tfrac{1}{3}n(n^2 - 3n + 8),$$

giving 2, 4, 8, 16, 30, 52, 84, ... (Sloane's A046127).

see also Plane Division

References

Sloane, N. J. A. Sequence A046127 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Space-Filling Curve

see Space-Filling Function

Space-Filling Function

A "Curve" (i.e., a continuous map of a 1-D Interval) into a 2-D area (a Plane-Filling Function) or a 3-D volume.

see also Hilbert Curve, Peano Curve, Peano-Gosper Curve, Plane-Filling Curve, Sierpiński Curve, Space-Filling Polyhedron

References

Pappas, T. "Paradoxical Curve–Space-Filling Curve." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 208, 1989.

Platzman, L. K. and Bartholdi, J. J. "Spacefilling Curves and the Planar Travelling Salesman Problem." *J. Assoc. Comput. Mach.* **46**, 719–737, 1989.

Wagon, S. "A Spacefilling Curve." §6.3 in *Mathematica in Action.* New York: W. H. Freeman, pp. 196–209, 1991.

Space-Filling Polyhedron

A space-filling polyhedron is a POLYHEDRON which can be used to generate a TESSELLATION of space. There exists one 16-sided space-filling POLYHEDRON, but it is unknown if this is the unique 16-sided space-filler. The CUBE, RHOMBIC DODECAHEDRON, and TRUNCATED OCTAHEDRON are space-fillers, as are the ELONGATED DODECAHEDRON and hexagonal PRISM. These five solids are all "primary" PARALLELOHEDRA (Coxeter 1973).

P. Schmitt discovered a nonconvex aperiodic polyhedral space-filler around 1990, and a convex POLYHEDRON known as the SCHMITT-CONWAY BIPRISM which fills space only aperiodically was found by J. H. Conway in 1993 (Eppstein).

see also CUBE, ELONGATED DODECAHEDRON, KELLER'S CONJECTURE, PARALLELOHEDRON, PRISM, RHOMBIC DODECAHEDRON, SCHMITT-CONWAY BIPRISM, TESSELLATION, TILING, TRUNCATED OCTAHEDRON

References

Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, pp. 29–30, 1973.

Critchlow, K. *Order in Space: A Design Source Book.* New York: Viking Press, 1970.

Devlin, K. J. "An Aperiodic Convex Space-filler is Discovered." *Focus: The Newsletter of the Math. Assoc. Amer.* **13**, 1, Dec. 1993.

Eppstein, D. "Re: Aperiodic Space-Filling Tile?." `http://www.ics.uci.edu/~eppstein/junkyard/biprism.html`.

Holden, A. *Shapes, Space, and Symmetry.* New York: Dover, pp. 154–163, 1991.

Thompson, D'A. W. *On Growth and Form, 2nd ed., compl. rev. ed.* New York: Cambridge University Press, 1992.

Tutton, A. E. H. *Crystallography and Practical Crystal Measurement, 2nd ed.* London: Lubrecht & Cramer, pp. 567 and 723, 1964.

Space Groups

The space groups in 2-D are called WALLPAPER GROUPS. In 3-D, the space groups are the symmetry GROUPS possible in a crystal lattice with the translation symmetry element. There are 230 space groups in $\mathbb{R}^3$, although 11 are MIRROR IMAGES of each other. They are listed by HERMANN-MAUGUIN SYMBOL in Cotton (1990).

see also HERMANN-MAUGUIN SYMBOL, LATTICE GROUPS, POINT GROUPS, WALLPAPER GROUPS

References

Arfken, G. "Crystallographic Point and Space Groups." *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 248–249, 1985.

Buerger, M. J. *Elementary Crystallography.* New York: Wiley, 1956.

Cotton, F. A. *Chemical Applications of Group Theory, 3rd ed.* New York: Wiley, pp. 250–251, 1990.

Span (Geometry)

The largest possible distance between two points for a finite set of points.

see also JUNG'S THEOREM

Span (Link)

The span of an unoriented LINK diagram (also called the SPREAD) is the difference between the highest and lowest degrees of its BRACKET POLYNOMIAL. The span is a topological invariant of a knot. If a KNOT K has a reduced alternating projection of n crossings, then the span of K is $4n$.

see also LINK

Span (Polynomial)

The difference between the highest and lowest degrees of a POLYNOMIAL.

Span (Set)

For a SET S, the span is defined by $\max S - \min S$, where max is the MAXIMUM and min is the MINIMUM.

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 207, 1994.

Span (Vectors)

The span of SUBSPACE generated by VECTORS $\mathbf{v}_1$ and $\mathbf{v}_2 \in \mathbb{V}$ is

$$\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2) \equiv \{r\mathbf{v}_1 + s\mathbf{v}_2 : r, s \in \mathbb{R}\}.$$

Sparse Matrix

A MATRIX which has only a small number of NONZERO elements.

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Sparse Linear Systems." §2.7 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 63–82, 1992.

Spearman Rank Correlation

The Spearman rank correlation is defined by

$$r' \equiv \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} = 1 - 6\sum \frac{d^2}{N(N^2-1)}. \quad (1)$$

The VARIANCE, KURTOSIS, and higher order MOMENTS are

$$\sigma^2 = \frac{1}{N-1} \quad (2)$$

$$\gamma_2 = -\frac{114}{25N} - \frac{6}{5N^2} - \ldots \quad (3)$$

$$\gamma_3 = \gamma_5 = \ldots = 0. \quad (4)$$

Student was the first to obtain the VARIANCE. The Spearman rank correlation is an R-ESTIMATE.

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 634–637, 1992.

Special Curve

see PLANE CURVE, SPACE CURVE

Special Function

see FUNCTION

Special Linear Group

The special linear group $SL_n(q)$ is the MATRIX GROUP corresponding to the set of $n \times n$ COMPLEX MATRICES having DETERMINANT +1. It is a SUBGROUP of the GENERAL LINEAR GROUP $GL_n(q)$ and is also a LIE GROUP.

see also GENERAL LINEAR GROUP, SPECIAL ORTHOGONAL GROUP, SPECIAL UNITARY GROUP

References

Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. "The Groups $GL_n(q)$, $SL_n(q)$, $PGL_n(q)$, and $PSL_n(q) = L_n(q)$." §2.1 in *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, p. x, 1985.

Special Matrix

A matrix whose entries satisfy

$$a_{ij} = \begin{cases} 0 & \text{if } j > i+1 \\ -1 & \text{if } j = i+1 \\ 0 \text{ or } 1 & \text{if } j \leq i. \end{cases}$$

There are 2^{n-1} special MINIMAL MATRICES of size $n \times n$.

References

Knuth, D. E. "Problem 10470." *Amer. Math. Monthly* **102**, 655, 1995.

Special Orthogonal Group

The special orthogonal group $SO_n(q)$ is the SUBGROUP of the elements of GENERAL ORTHOGONAL GROUP $GO_n(q)$ with DETERMINANT 1.

see also GENERAL ORTHOGONAL GROUP, SPECIAL LINEAR GROUP, SPECIAL UNITARY GROUP

References

Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. "The Groups $GO_n(q)$, $SO_n(q)$, $PGO_n(q)$, and $PSO_n(q)$, and $O_n(q)$." §2.4 in *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, pp. xi–xii, 1985.

Special Point

A POINT which does not lie on at least one ORDINARY LINE.

see also ORDINARY POINT

References

Guy, R. K. "Unsolved Problems Come of Age." *Amer. Math. Monthly* **96**, 903–909, 1989.

Special Series Theorem

If the difference between the order and the dimension of a series is less than the GENUS (CURVE), then the series is special.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 253, 1959.

Special Unitary Group

The special unitary group $SU_n(q)$ is the set of $n \times n$ UNITARY MATRICES with DETERMINANT +1 (having $n^2 - 1$ independent parameters). $SU(2)$ is HOMEOMORPHIC with the ORTHOGONAL GROUP $O_3^+(2)$. It is also called the UNITARY UNIMODULAR GROUP and is a LIE GROUP. The special unitary group can be represented by the MATRIX

$$U(a,b) = \begin{bmatrix} a & b \\ -b^* & a^* \end{bmatrix}, \quad (1)$$

where $a^*a + b^*b = 1$ and a, b are the CAYLEY-KLEIN PARAMETERS. The special unitary group may also be represented by the MATRIX

$$U(\xi,\eta,\zeta) = \begin{bmatrix} e^{i\xi}\cos\eta & e^{i\zeta}\sin\eta \\ -e^{-i\zeta}\sin\eta & e^{-i\xi}\cos\eta \end{bmatrix}, \quad (2)$$

or the matrices

$$U_x(\tfrac{1}{2}\phi) = \begin{bmatrix} \cos(\tfrac{1}{2}\phi) & i\sin(\tfrac{1}{2}\phi) \\ i\sin(\tfrac{1}{2}\phi) & \cos(\tfrac{1}{2}\phi) \end{bmatrix} \quad (3)$$

$$U_y(\tfrac{1}{2}\beta) = \begin{bmatrix} \cos(\tfrac{1}{2}\beta) & \sin(\tfrac{1}{2}\beta) \\ -\sin(\tfrac{1}{2}\beta) & \cos(\tfrac{1}{2}\beta) \end{bmatrix} \quad (4)$$

$$U_z(\xi) = \begin{bmatrix} e^{i\xi} & 0 \\ 0 & e^{-i\xi} \end{bmatrix}. \quad (5)$$

The order $2j+1$ representation is

$$
\begin{aligned}
U_{p,q}{}^{(j)}(\alpha,\beta,\gamma) &= \sum_m \frac{(-1)^{m-q-p}\sqrt{(j+p)!(j-p)!(j+q)!(j-q)!}}{(j-p-m)!(j+q-m)!(m+p-q)!m!} \\
&\quad \times e^{iq\alpha}\cos^{2j+q-p-2m}(\tfrac{1}{2}\beta)\sin^{p+2m-q}(\tfrac{1}{2}\beta)e^{ip\gamma}. \qquad (6)
\end{aligned}
$$

The summation is terminated by putting $1/(-N)! = 0$. The CHARACTER is given by

$$
\begin{aligned}
\chi^{(j)}(\alpha) &= \begin{cases} 1+2\cos\alpha+\ldots+2\cos(j\alpha) \\ 2[\cos(\frac{1}{2}\alpha)+\cos(\frac{3}{2}\alpha)+\ldots+\cos(j\alpha)] \end{cases} \\
&= \begin{cases} \dfrac{\sin[(j+\frac{1}{2})\alpha]}{\sin(\frac{1}{2}\alpha)} & \text{for } j = 0, 1, 2, \ldots \\ \dfrac{\sin[(j+\frac{1}{2})\alpha]}{\sin(\frac{1}{2}\alpha)} & \text{for } j = \frac{1}{2}, \frac{3}{2}, \ldots. \end{cases} \qquad (7)
\end{aligned}
$$

see also ORTHOGONAL GROUP, SPECIAL LINEAR GROUP, SPECIAL ORTHOGONAL GROUP

References

Arfken, G. "Special Unitary Group, $SU(2)$ and $SU(2)$-O_3^+ Homomorphism." *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 253–259, 1985.

Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. "The Groups $GU_n(q)$, $SU_n(q)$, $PGU_n(q)$, and $PSU_n(q) = U_n(q)$." §2.2 in *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, p. x, 1985.

Species

A species of structures is a rule F which

1. Produces, for each finite set U, a finite set $F[U]$,
2. Produces, for each bijection $\sigma : U \to V$, a function

$$F[\sigma] : F[U] \to F[V].$$

The functions $F[\sigma]$ should further satisfy the following functorial properties:

1. For all bijections $\sigma : U \to V$ and $\tau : V \to W$,

$$F[\tau \circ \sigma] = F[\tau] \circ F[\sigma],$$

2. For the IDENTITY MAP $\mathrm{Id}_U : U \to U$,

$$F[\underset{U}{\mathrm{Id}}] = \underset{F[U]}{\mathrm{Id}}\ .$$

An element $\sigma \in F[U]$ is called an F-structure on U (or a structure of species F on U). The function $F[\sigma]$ is called the transport of F-structures along σ.

References

Bergeron, F.; Labelle, G.; and Leroux, P. *Combinatorial Species and Tree-Like Structures.* Cambridge, England: Cambridge University Press, p. 5, 1998.

Specificity

The probability that a STATISTICAL TEST will be negative for a negative statistic.

see also SENSITIVITY, STATISTICAL TEST, TYPE I ERROR, TYPE II ERROR

Spectral Norm

The NATURAL NORM induced by the L_2-NORM. Let $\mathsf{A}^\dagger$ be the ADJOINT of the SQUARE MATRIX A, so that $\mathsf{A}^\dagger = a^*_{ji}$, then

$$
\begin{aligned}
\|\mathsf{A}\|_2 &= (\text{maximum eigenvalue of } \mathsf{A}^\dagger\mathsf{A})^{1/2} \\
&= \max_{\|x\|_2 \neq 0} \frac{\|\mathsf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2}.
\end{aligned}
$$

see also L_2-NORM, MATRIX NORM

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 1115, 1979.

Strang, G. §6.2 and 7.2 in *Linear Algebra and Its Applications, 4th ed.* New York: Academic Press, 1980.

Spectral Power Density

$$P_y(\nu) \equiv \lim_{T\to\infty} \frac{2}{T} \left| \int_{-T/2}^{T/2} [y(t)-\bar{y}]e^{-2\pi i\nu t}\,dt \right|^2,$$

so

$$
\begin{aligned}
\int_0^\infty P_y(\nu)\,d\nu &= \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} [y(t)-\bar{y}]^2\,dt \\
&= \left\langle (y-\bar{y})^2 \right\rangle = \sigma_y{}^2.
\end{aligned}
$$

see also POWER SPECTRUM

Spectral Radius

Let A be an $n \times n$ MATRIX with COMPLEX or REAL elements with EIGENVALUES $\lambda_1, \ldots, \lambda_n$. Then the spectral radius $\rho(\mathsf{A})$ of A is

$$\rho(\mathsf{A}) = \max_{1\leq i\leq n} |\lambda_i|.$$

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 1115–1116, 1979.

Spectral Rigidity

The mean square deviation of the best local fit straight line to a staircase cumulative spectral density over a normalized energy scale.

References

Ott, E. *Chaos in Dynamical Systems.* New York: Cambridge University Press, p. 341, 1993.

Spectral Theorem

Let H be a HILBERT SPACE, $B(H)$ the set of BOUNDED linear operators from H to itself, and $\sigma(T)$ the SPECTRUM of T. Then if $T \in B(H)$ and T is normal, there exists a unique resolution of the identity E on the Borel subsets of $\sigma(T)$ which satisfies

$$T = \operatorname*{int}_{\sigma(T)} \lambda \, dE(\lambda).$$

Furthermore, every projection $E(\omega)$ COMMUTES with every $S \in B(H)$ which COMMUTES with T.

References

Rudin, W. Theorem 12.23 in *Functional Analysis, 2nd ed.* New York: McGraw-Hill, 1991.

Spectrum (Operator)

Let T be an OPERATOR on a HILBERT SPACE. The spectrum $\sigma(T)$ of T is the set of λ such that $(T - \lambda I)$ is not invertible on all of the HILBERT SPACE, where the λs are COMPLEX NUMBERS and I is the IDENTITY OPERATOR. The definition can also be stated in terms of the resolvent of an operator

$$\rho(T) = \{\lambda : (T - \lambda I) \text{ is invertible}\},$$

and then the spectrum is defined to be the complement of $\rho(T)$ in the COMPLEX PLANE. The reason for doing this is that it is easy to demonstrate that $\rho(T)$ is an OPEN SET, which shows that the spectrum is closed.

see also HILBERT SPACE

Spectrum Sequence

A spectrum sequence is a SEQUENCE formed by successive multiples of a REAL NUMBER a rounded down to the nearest INTEGER $s_n = \lfloor na \rfloor$. If a is IRRATIONAL, the spectrum is called a BEATTY SEQUENCE.

see also BEATTY SEQUENCE, LAGRANGE SPECTRUM, MARKOV SPECTRUM

Speed

The SCALAR $|\mathbf{v}|$ equal to the magnitude of the VELOCITY $\mathbf{v}$.

see also ANGULAR VELOCITY, VELOCITY

Spence's Function

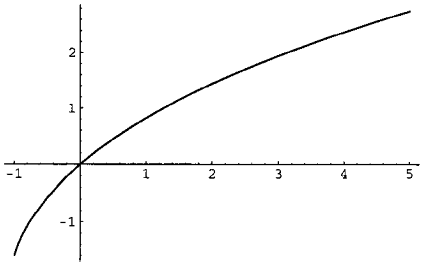

$$F(x) \equiv \int_0^x \frac{\ln(1+t)}{t}\, dt.$$

see also SPENCE'S INTEGRAL

References

Berestetskii, V. B.; Lifschitz, E. M.; and Ditaevskii, L. P. *Quantum Electrodynamics, 2nd ed.* Oxford, England: Pergamon Press, p. 596, 1982.

Spence's Integral

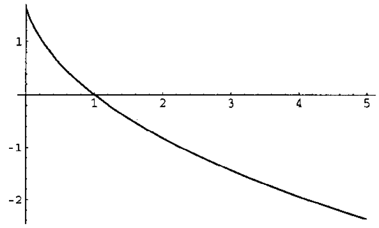

$$F(x) = \mathrm{Li}_2(1-x) = \int_{1-x}^0 \frac{\ln(1-t)}{t}\, dt,$$

where $\mathrm{Li}_2(x)$ is the DILOGARITHM.

see also SPENCE'S FUNCTION

Spencer's 15-Point Moving Average

A MOVING AVERAGE using 15 points having weights -3, -6, -5, 3, 21, 46, 67, 74, 67, 46, 21, 3, -5, -6, and -3. It is sometimes used by actuaries.

see also MOVING AVERAGE

References

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, p. 223, 1962.

Sperner's Theorem

The MAXIMUM CARDINALITY of a collection of SUBSETS of a t-element SET T, none of which contains another, is the BINOMIAL COEFFICIENT $\binom{t}{\lfloor t/2 \rfloor}$, where $\lfloor x \rfloor$ is the FLOOR FUNCTION.

see also CARDINALITY

Sphenocorona

see JOHNSON SOLID

Sphenoid

see DISPHENOID

Sphenomegacorona

see JOHNSON SOLID

Sphere

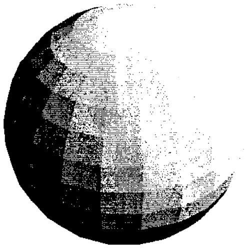

A sphere is defined as the set of all points in $\mathbb{R}^3$ which are a distance r (the "RADIUS") from a given point (the "CENTER"). Twice the RADIUS is called the DIAMETER, and pairs of points on opposite sides of a DIAMETER are called ANTIPODES. The term "sphere" technically refers to the outer surface of a "BUBBLE," which is denoted $\mathbb{S}^2$. However, in common usage, the word *sphere* is also used to mean the UNION of a sphere and its INTERIOR (a "solid sphere"), where the INTERIOR is called a BALL. The SURFACE AREA of the sphere and VOLUME of the BALL of RADIUS r are given by

$$S = 4\pi r^2 \tag{1}$$

$$V = \tfrac{4}{3}\pi r^3 \tag{2}$$

(Beyer 1987, p. 130). In *On the Sphere and Cylinder* (ca. 225 BC), Archimedes became the first to derive these equations (although he expressed π in terms of the sphere's circular cross-section). The fact that

$$\frac{V_{\text{sphere}}}{V_{\text{circumscribed cylinder}} - V_{\text{sphere}}} = \frac{\frac{4}{3}}{2-\frac{4}{3}} = \frac{\frac{4}{3}}{\frac{2}{3}} = 2 \tag{3}$$

was also known to Archimedes.

Any cross-section through a sphere is a CIRCLE (or, in the degenerate case where the slicing PLANE is tangent to the sphere, a point). The size of the CIRCLE is maximized when the PLANE defining the cross-section passes through a DIAMETER.

The equation of a sphere of RADIUS r is given in CARTESIAN COORDINATES by

$$x^2 + y^2 + z^2 = r^2, \tag{4}$$

which is a special case of the ELLIPSOID

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{5}$$

and SPHEROID

$$\frac{x^2 + y^2}{a^2} + \frac{z^2}{c^2} = 1. \tag{6}$$

A sphere may also be specified in SPHERICAL COORDINATES by

$$x = \rho\cos\theta\sin\phi \tag{7}$$

$$y = \rho\sin\theta\sin\phi \tag{8}$$

$$z = \rho\cos\phi, \tag{9}$$

where θ is an azimuthal coordinate running from 0 to 2π (LONGITUDE), ϕ is a polar coordinate running from 0 to π (COLATITUDE), and ρ is the RADIUS. Note that there are several other notations sometimes used in which the symbols for θ and ϕ are interchanged or where r is used instead of ρ. If ρ is allowed to run from 0 to a given RADIUS r, then a solid BALL is obtained. Converting to "standard" parametric variables $a = \rho$, $u = \theta$, and $v = \phi$ gives the first FUNDAMENTAL FORMS

$$E = a\sin^2 v \tag{10}$$

$$F = 0 \tag{11}$$

$$G = a, \tag{12}$$

second FUNDAMENTAL FORMS

$$e = a^2\sin^2 v \tag{13}$$

$$f = 0 \tag{14}$$

$$g = a^2, \tag{15}$$

AREA ELEMENT

$$dA = a\sin v, \tag{16}$$

GAUSSIAN CURVATURE

$$K = \frac{1}{a^2}, \tag{17}$$

and MEAN CURVATURE

$$H = \frac{1}{a}. \tag{18}$$

A sphere may also be represented parametrically by letting $u \equiv r\cos\phi$, so

$$x = \sqrt{r^2 - u^2}\cos\theta \tag{19}$$

$$y = \sqrt{r^2 - u^2}\sin\theta \tag{20}$$

$$z = u, \tag{21}$$

where θ runs from 0 to 2π and u runs from $-r$ to r.

Given two points on a sphere, the shortest path on the surface of the sphere which connects them (the SPHERE GEODESIC) is an ARC of a CIRCLE known as a GREAT CIRCLE. The equation of the sphere with points (x_1, y_1, z_1) and (x_2, y_2, z_2) lying on a DIAMETER is given by

$$(x - x_1)(x - x_2) + (y - y_1)(y - y_2) + (z - z_1)(z - z_2) = 0. \tag{22}$$

Four points are sufficient to uniquely define a sphere. Given the points (x_i, y_i, z_i) with $i = 1$, 2, 3, and 4, the sphere containing them is given by the beautiful DETERMINANT equation

$$\begin{vmatrix} x^2 + y^2 + z^2 & x & y & z & 1 \\ {x_1}^2 + {y_1}^2 + {z_1}^2 & x_1 & y_1 & z_1 & 1 \\ {x_2}^2 + {y_2}^2 + {z_2}^2 & x_2 & y_2 & z_2 & 1 \\ {x_3}^2 + {y_3}^2 + {z_3}^2 & x_3 & y_3 & z_3 & 1 \\ {x_4}^2 + {y_4}^2 + {z_4}^2 & x_4 & y_4 & z_4 & 1 \end{vmatrix} = 0 \tag{23}$$

(Beyer 1987, p. 210).

The generalization of a sphere in n dimensions is called a HYPERSPHERE. An n-D HYPERSPHERE can be specified by the equation

$$x_1{}^2 + x_2{}^2 + \ldots + x_n{}^2 = r^2. \tag{24}$$

The distribution of ANGLES for random rotation of a sphere is

$$P(\theta) = \frac{2}{\pi}\sin^2(\tfrac{1}{2}\theta), \tag{25}$$

giving a MEAN of $\pi/2 + 2/\pi$.

To pick a random point on the surface of a sphere, let u and v be random variates on $[0, 1]$. Then

$$\theta = 2\pi u \tag{26}$$

$$\phi = \cos^{-1}(2v - 1). \tag{27}$$

This works since the SOLID ANGLE is

$$d\Omega = \sin\phi\, d\theta\, d\phi = d\theta\, d(\cos\phi). \tag{28}$$

Another easy way to pick a random point on a SPHERE is to generate three gaussian random variables x, y, and z. Then the distribution of the vectors

$$\frac{1}{\sqrt{x^2 + y^2 + z^2}}\begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{29}$$

is uniform over the surface $\mathbb{S}^2$. Another method is to pick z from a UNIFORM DISTRIBUTION over $[-1, 1]$ and θ from a UNIFORM DISTRIBUTION over $[0, 2\pi)$. Then the points

$$\begin{bmatrix} \sqrt{1 - z^2}\cos\theta \\ \sqrt{1 - z^2}\sin\theta \\ z \end{bmatrix} \tag{30}$$

are uniformly distributed over $\mathbb{S}^2$.

Pick four points on a sphere. What is the probability that the TETRAHEDRON having these points as VERTICES contains the CENTER of the sphere? In the 1-D case, the probability that a second point is on the opposite side of 1/2 is 1/2. In the 2-D case, pick two points. In order for the third to form a TRIANGLE containing the CENTER, it must lie in the quadrant bisected by a LINE SEGMENT passing through the center of the CIRCLE and the bisector of the two points. This happens for one QUADRANT, so the probability is 1/4. Similarly, for a sphere the probability is one OCTANT, or 1/8.

Pick two points at random on a unit sphere. The first one can be assigned the coordinate (0, 0, 1) without loss of generality. The second point can be given the coordinates $(\sin\phi, 0\cos\phi)$ with $\theta \equiv 0$ since all points with the same ϕ are rotationally identical. The distance between the two points is then

$$r = \sqrt{\sin^2\phi + (1 - \cos\phi)^2} = \sqrt{2 - \cos\phi} = 2\sin(\tfrac{1}{2}\phi). \tag{31}$$

Because the surface AREA element is

$$d\Omega = \sin\phi\, d\theta\, d\phi, \tag{32}$$

the probability that two points are a distance r apart is

$$\begin{aligned} P_\phi(r) &= \frac{\int_0^\pi \delta(\phi - r)\sin\phi\, d\phi}{\int_0^\pi \sin\phi\, d\phi} \\ &= \tfrac{1}{2}\int_0^\pi \delta[r - 2\sin(\tfrac{1}{2}\phi)]\sin\phi\, d\phi. \end{aligned} \tag{33}$$

The DELTA FUNCTION contributes when

$$\tfrac{1}{2}r = \sin(\tfrac{1}{2}\phi) \tag{34}$$

$$\phi = 2\sin^{-1}(\tfrac{1}{2}r), \tag{35}$$

so

$$\begin{aligned} P_\phi(r) &= \tfrac{1}{2}\sin[2\sin^{-1}(\tfrac{1}{2}r)] = \sin[\sin^{-1}(\tfrac{1}{2}r)]\cos[\sin^{-1}(\tfrac{1}{2}r)] \\ &= \tfrac{1}{2}r\sqrt{1 - (\tfrac{1}{2}r)^2} = \tfrac{1}{4}r\sqrt{4 - r^2}. \end{aligned} \tag{36}$$

However, we need

$$P_r(r)\, dr = P_\phi(r)\frac{d\phi}{dr}\, dr, \tag{37}$$

and

$$\begin{aligned} \tfrac{1}{2}dr &= \tfrac{1}{2}\cos(\tfrac{1}{2}\phi)\, d\phi = \tfrac{1}{2}\sqrt{1 - \sin^2(\tfrac{1}{2}\phi)}\, d\phi \\ &= \tfrac{1}{2}\sqrt{1 - (\tfrac{1}{2}r)^2}\, d\phi = \tfrac{1}{4}\sqrt{4 - r^2}\, d\phi \end{aligned} \tag{38}$$

so

$$\frac{d\phi}{dr} = \frac{2}{\sqrt{4 - r^2}}, \tag{39}$$

and

$$P_r(r) = \tfrac{1}{4}r\sqrt{4 - r^2}\frac{2}{\sqrt{4 - r^2}} = \tfrac{1}{2}r \tag{40}$$

for $r \in [0, 2]$. Somewhat surprisingly, the largest distances are the most common, contrary to most people's intuition. A plot of 15 random lines is shown below.

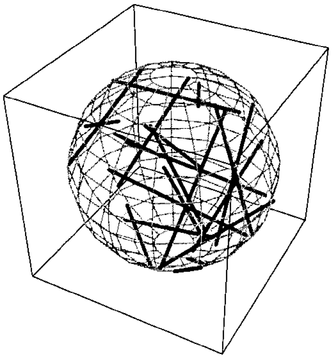

The MOMENTS about zero are

$$\mu_n' = \langle r^n \rangle = \int_0^2 r^n \, dr = \frac{2^{n+1}}{2+n}, \tag{41}$$

giving the first few as

$$\mu_1' = \tfrac{4}{3} \tag{42}$$
$$\mu_2' = 2 \tag{43}$$
$$\mu_3' = \tfrac{16}{5} \tag{44}$$
$$\mu_4' = \tfrac{16}{3}. \tag{45}$$

Moments about the MEAN are

$$\mu = \tfrac{4}{3} \tag{46}$$
$$\mu_2 = \sigma^2 = \tfrac{2}{9} \tag{47}$$
$$\mu_3 = -\tfrac{8}{135} \tag{48}$$
$$\mu_4 = \tfrac{16}{135}, \tag{49}$$

so the SKEWNESS and KURTOSIS are

$$\gamma_1 = \tfrac{4}{5}\sqrt{2} \tag{50}$$
$$\gamma_2 = -\tfrac{5}{3}. \tag{51}$$

see also BALL, BING'S THEOREM, BUBBLE, CIRCLE, DANDELIN SPHERES, DIAMETER, ELLIPSOID, EXOTIC SPHERE, FEJES TÓTH'S PROBLEM, HYPERSPHERE, LIEBMANN'S THEOREM, LIOUVILLE'S SPHERE-PRESERVING THEOREM, MIKUSIŃSKI'S PROBLEM, NOISE SPHERE, OBLATE SPHEROID, OSCULATING SPHERE, PARALLELIZABLE, PROLATE SPHEROID, RADIUS, SPACE DIVISION, SPHERE PACKING, TENNIS BALL THEOREM

References

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, 1987.

Eppstein, D. "Circles and Spheres." `http://www.ics.uci.edu/~eppstein/junkyard/sphere.html`.

Geometry Center. "The Sphere." `http://www.geom.umn.edu/zoo/toptype/sphere/`.

Sphere-Cylinder Intersection

see CYLINDER-SPHERE INTERSECTION

Sphere Embedding

A 4-sphere has POSITIVE CURVATURE, with

$$R^2 = x^2 + y^2 + z^2 + w^2 \tag{1}$$

$$2x\frac{dx}{dw} + 2y\frac{dy}{dw} + 2z\frac{dz}{dw} + 2w = 0. \tag{2}$$

Since

$$\mathbf{r} \equiv x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}, \tag{3}$$

$$dw = -\frac{x\,dx + y\,dy + z\,dz}{w} = -\frac{\mathbf{r}\cdot d\mathbf{r}}{\sqrt{R^2 - r^2}}. \tag{4}$$

To stay on the surface of the sphere,

$$\begin{aligned} ds^2 &= dx^2 + dy^2 + dz^2 + dw^2 \\ &= dx^2 + dy^2 + dz^2 + \frac{r^2\,dr^2}{R^2 - r^2} \\ &= dr^2 + r^2\,d\Omega^2 + \frac{dr^2}{\frac{R^2}{r^2} - 1} \\ &= dr^2\left(1 + \frac{1}{\frac{R^2}{r^2} - 1}\right) + r^2\,d\Omega^2 \\ &= dr^2\left(\frac{\frac{R^2}{r^2}}{\frac{R^2}{r^2} - 1}\right) + r^2\,d\Omega^2 \\ &= \frac{dr^2}{1 - \frac{r^2}{R^2}} + r^2\,d\Omega^2. \end{aligned} \tag{5}$$

With the addition of the so-called expansion parameter, this is the Robertson-Walker line element.

Sphere Eversion

Smale (1958) proved that it is mathematically possible to turn a SPHERE inside-out without introducing a sharp crease at any point. This means there is a regular homotopy from the standard embedding of the 2-SPHERE in EUCLIDEAN 3-space to the mirror-reflection embedding such that at every stage in the homotopy, the sphere is being IMMERSED in EUCLIDEAN SPACE. This result is so counterintuitive and the proof so technical that the result remained controversial for a number of years.

In 1961, Arnold Shapiro devised an explicit eversion but did not publicize it. Phillips (1966) heard of the result and, in trying to reproduce it, actually devised an independent method of his own. Yet another eversion was devised by Morin, which became the basis for the movie by Max (1977). Morin's eversion also produced explicit algebraic equations describing the process. The original method of Shapiro was subsequently published by Francis and Morin (1979).

see also EVERSION, SPHERE

References

Francis, G. K. Ch. 6 in *A Topological Picturebook.* New York: Springer-Verlag, 1987.

Francis, G. K. and Morin, B. "Arnold Shapiro's Eversion of the Sphere." *Math. Intell.* **2**, 200–203, 1979.

Levy, S. *Making Waves: A Guide to the Ideas Behind Outside In.* Wellesley, MA: A. K. Peters, 1995.

Levy, S. "A Brief History of Sphere Eversions." `http://www.geom.umn.edu/docs/outreach/oi/history.html`.

Levy, S.; Maxwell, D.; and Munzner, T. *Outside-In.* 22 minute videotape. `http://www.geom.umn.edu/docs/outreach/oi/`.

Max, N. "Turning a Sphere Inside Out." Videotape. Chicago, IL: International Film Bureau, 1977.

Peterson, I. *Islands of Truth: A Mathematical Mystery Cruise.* New York: W. H. Freeman, pp. 240–244, 1990.

Petersen, I. "Forging Links Between Mathematics and Art." *Science News* **141**, 404–405, June 20, 1992.

Phillips, A. "Turning a Surface Inside Out." *Sci. Amer.* **214**, 112–120, Jan. 1966.

Smale, S. "A Classification of Immersions of the Two-Sphere." *Trans. Amer. Math. Soc.* **90**, 281–290, 1958.

Sphere Geodesic

see GREAT CIRCLE

Sphere Packing

Let η denote the PACKING DENSITY, which is the fraction of a VOLUME filled by identical packed SPHERES. In 2-D (CIRCLE PACKING), there are two periodic packings for identical CIRCLES: square lattice and hexagonal lattice. Fejes Tóth (1940) proved that the hexagonal lattice is indeed the densest of *all* possible plane packings (Conway and Sloane 1993, pp. 8–9).

In 3-D, there are three periodic packings for identical spheres: cubic lattice, face-centered cubic lattice, and hexagonal lattice. It was hypothesized by Kepler in 1611 that close packing (cubic or hexagonal) is the densest possible (has the greatest η), and this assertion is known as the KEPLER CONJECTURE. The problem of finding the densest packing of spheres (not necessarily periodic) is therefore known as the KEPLER PROBLEM. The KEPLER CONJECTURE is intuitively obvious, but the proof still remains elusive. However, Gauss (1831) did prove that the face-centered cubic is the densest *lattice* packing in 3-D (Conway and Sloane 1993, p. 9). This result has since been extended to HYPERSPHERE PACKING.

In 3-D, face-centered cubic close packing and hexagonal close packing (which is distinct from hexagonal lattice), both give

$$\eta = \frac{\pi}{3\sqrt{2}} \approx 74.048\%. \qquad (1)$$

For packings in 3-D, C. A. Rogers (1958) showed that

$$\eta < \sqrt{18}\,(\cos^{-1}\tfrac{1}{3} - \tfrac{1}{3}\pi) \approx 77.96355700\% \qquad (2)$$

(Le Lionnais 1983). This was subsequently improved to 77.844% (Lindsey 1986), then 77.836% (Muder 1988). However, Rogers (1958) remarks that "many mathematicians believe, and all physicists know" that the actual answer is 74.05% (Conway and Sloane 1993, p. 3).

"Random" close packing in 3-D gives only $\eta \approx 64\%$ (Jaeger and Nagel 1992).

The PACKING DENSITIES for several packing types are summarized in the following table.

Packing	η (exact)	η (approx.)
square lattice (2-D)	$\frac{\pi}{4}$	0.7854
hexagonal lattice (2-D)	$\frac{\pi}{2\sqrt{3}}$	0.9069
cubic lattice	$\frac{\pi}{6}$	0.5236
hexagonal lattice	$\frac{\pi}{3\sqrt{3}}$	0.6046
face-centered cubic lattice	$\frac{\pi}{3\sqrt{2}}$	0.7405
random	—	0.6400

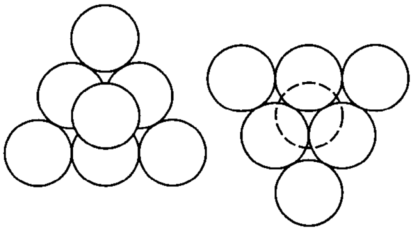

For cubic close packing, pack six SPHERES together in the shape of an EQUILATERAL TRIANGLE and place another SPHERE on top to create a TRIANGULAR PYRAMID. Now create another such grouping of seven SPHERES and place the two PYRAMIDS together facing in opposite directions. A CUBE emerges. Consider a face of the CUBE, illustrated below.

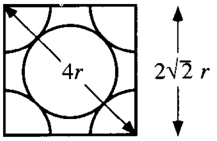

The "unit cell" cube contains eight 1/8-spheres (one at each VERTEX) and six HEMISPHERES. The total VOLUME of SPHERES in the unit cell is

$$\begin{aligned} V_{\text{spheres in unit cell}} &= (8\cdot\tfrac{1}{8} + 6\cdot\tfrac{1}{2})\frac{4\pi}{3}r^3 \\ &= 4\cdot\frac{4\pi}{3}r^3 = \tfrac{16}{3}\pi r^3. \end{aligned} \qquad (3)$$

The diagonal of the face is $4r$, so each side is $2\sqrt{2}\,r$. The VOLUME of the unit cell is therefore

$$V_{\text{unit cell}} = (2\sqrt{2}\,r)^3 = 16\sqrt{2}\,r^3. \qquad (4)$$

The PACKING DENSITY is therefore

$$\eta_{\text{CCP}} = \frac{\frac{16}{3}\pi r^2}{16\sqrt{2}r^3} = \frac{\pi}{3\sqrt{2}} \qquad (5)$$

(Conway and Sloane 1993, p. 2).

Hexagonal close packing must give the same values, since sliding one sheet of SPHERES cannot affect the volume they occupy. To verify this, construct a 3-D diagram containing a hexagonal unit cell with three layers. Both the top and the bottom contain six 1/6-SPHERES and one HEMISPHERE. The total number of spheres in these two rows is therefore

$$2(6\tfrac{1}{6} + 1\tfrac{1}{2}) = 3. \qquad (6)$$

The VOLUME of SPHERES in the middle row cannot be simply computed using geometry. However, symmetry requires that the piece of the SPHERE which is cut off is exactly balanced by an extra piece on the other side. There are therefore three SPHERES in the middle layer, for a total of six, and a total VOLUME

$$V_{\text{spheres in unit cell}} = 6\cdot\frac{4\pi}{3}r^3(3+3) = 8\pi r^3. \qquad (7)$$

The base of the HEXAGON is made up of 6 EQUILATERAL TRIANGLES with side lengths $2r$. The unit cell base AREA is therefore

$$A_{\text{unit cell}} = 6[\tfrac{1}{2}(2r)(\sqrt{3}\,r)] = 6\sqrt{3}\,r^2. \tag{8}$$

The height is the same as that of two TETRAHEDRA length $2r$ on a side, so

$$h_{\text{unit cell}} = 2\left(2r\sqrt{\frac{2}{3}}\right), \tag{9}$$

giving

$$\eta_{\text{HCP}} = \frac{8\pi r^3}{(6\sqrt{3}\,r^2)\left(4r\sqrt{\frac{2}{3}}\right)} = \frac{\pi}{3\sqrt{2}} \tag{10}$$

(Conway and Sloane 1993, pp. 7 and 9).

If we had actually wanted to compute the VOLUME of SPHERE inside and outside the HEXAGONAL PRISM, we could use the SPHERICAL CAP equation to obtain

$$\begin{aligned} V_{\subset} &= \tfrac{1}{3}\pi h^2(3r-h) = \frac{1}{3}\pi r^3 \frac{1}{3}\left(3 - \frac{1}{\sqrt{3}}\right) \\ &= \frac{1}{9}\pi r^3\left(3 - \frac{\sqrt{3}}{3}\right) = \tfrac{1}{27}\pi r^3(9-\sqrt{3}) \qquad (11) \\ V_{\supset} &= \pi r^3\left[\tfrac{4}{3} - \tfrac{1}{27}(9-\sqrt{3})\right] = \tfrac{1}{27}\pi r^3(36 - 9 + \sqrt{3}) \\ &= \tfrac{1}{27}\pi r^3(27+\sqrt{3}). \qquad (12) \end{aligned}$$

The rigid packing with *lowest* density known has $\eta \approx 0.0555$ (Gardner 1966). To be RIGID, each SPHERE must touch at least four others, and the four contact points cannot be in a single HEMISPHERE or all on one equator.

If spheres packed in a cubic lattice, face-centered cubic lattice, and hexagonal lattice are allowed to expand, they form cubes, hexagonal prisms, and rhombic dodecahedra. Compressing a random packing gives polyhedra with an average of 13.3 faces (Coxeter 1958, 1961).

For sphere packing inside a CUBE, see Goldberg (1971) and Schaer (1966).

see also CANNONBALL PROBLEM, CIRCLE PACKING, DODECAHEDRAL CONJECTURE, HEMISPHERE, HERMITE CONSTANTS, HYPERSPHERE, HYPERSPHERE PACKING, KEPLER CONJECTURE, KEPLER PROBLEM, KISSING NUMBER, LOCAL DENSITY, LOCAL DENSITY CONJECTURE, SPHERE

References
Conway, J. H. and Sloane, N. J. A. *Sphere Packings, Lattices, and Groups, 2nd ed.* New York: Springer-Verlag, 1993.
Coxeter, H. S. M. "Close-Packing and so Forth." *Illinois J. Math.* **2**, 746–758, 1958.
Coxeter, H. S. M. "Close Packing of Equal Spheres." Section 22.4 in *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 405–411, 1961.
Coxeter, H. S. M. "The Problem of Packing a Number of Equal Nonoverlapping Circles on a Sphere." *Trans. New York Acad. Sci.* **24**, 320–331, 1962.
Critchlow, K. *Order in Space: A Design Source Book.* New York: Viking Press, 1970.
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 195–197, 1989.
Eppstein, D. "Covering and Packing." `http://www.ics.uci.edu/~eppstein/junkyard/cover.html`.
Fejes Tóth, G. "Über einen geometrischen Satz." *Math. Z.* **46**, 78–83, 1940.
Fejes Tóth, G. *Lagerungen in der Ebene, auf der Kugel und in Raum, 2nd ed.* Berlin: Springer-Verlag, 1972.
Gardner, M. "Packing Spheres." Ch. 7 in *Martin Gardner's New Mathematical Diversions from Scientific American.* New York: Simon and Schuster, 1966.
Gauss, C. F. "Besprechung des Buchs von L. A. Seeber: Intersuchungen über die Eigenschaften der positiven ternären quadratischen Formen usw." *Göttingsche Gelehrte Anzeigen (1831, July 9)* **2**, 188–196, 1876.
Goldberg, M. "On the Densest Packing of Equal Spheres in a Cube." *Math. Mag.* **44**, 199–208, 1971.
Hales, T. C. "The Sphere Packing Problem." *J. Comput. Appl. Math* **44**, 41–76, 1992.
Jaeger, H. M. and Nagel, S. R. "Physics of Granular States." *Science* **255**, 1524, 1992.
Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 31, 1983.
Lindsey, J. H. II. "Sphere Packing in $\mathbb{R}^3$." *Math.* **33**, 137–147, 1986.
Muder, D. J. "Putting the the Best Face of a Voronoi Polyhedron." *Proc. London Math. Soc.* **56**, 329–348, 1988.
Rogers, C. A. "The Packing of Equal Spheres." *Proc. London Math. Soc.* **8**, 609–620, 1958.
Rogers, C. A. *Packing and Covering.* Cambridge, England: Cambridge University Press, 1964.
Schaer, J. "On the Densest Packing of Spheres in a Cube." *Can. Math. Bul.* **9**, 265–270, 1966.
Sloane, N. J. A. "The Packing of Spheres." *Sci. Amer.* **250**, 116–125, 1984.
Stewart, I. *The Problems of Mathematics, 2nd ed.* Oxford, England: Oxford University Press, pp. 69–82, 1987.
Thompson, T. M. *From Error-Correcting Codes Through Sphere Packings to Simple Groups.* Washington, DC: Math. Assoc. Amer., 1984.

Sphere Point Picking

see FEJES TÓTH'S PROBLEM

Sphere-Sphere Intersection

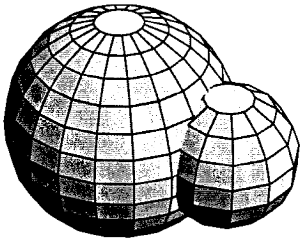

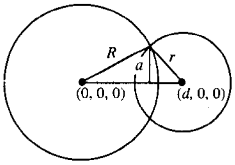

Let two spheres of RADII R and r be located along the x-AXIS centered at $(0,0,0)$ and $(d,0,0)$, respectively. Not surprisingly, the analysis is very similar to the case of the CIRCLE-CIRCLE INTERSECTION. The equations of the two SPHERES are

$$x^2+y^2+z^2=R^2 \qquad (1)$$
$$(x-d)^2+y^2+z^2=r^2. \qquad (2)$$

Combining (1) and (2) gives

$$(x-d)^2+(R^2-x^2)=r^2. \qquad (3)$$

Multiplying through and rearranging give

$$x^2-2dx+d^2-x^2=r^2-R^2. \qquad (4)$$

Solving for x gives

$$x=\frac{d^2-r^2+R^2}{2d}. \qquad (5)$$

The intersection of the SPHERES is therefore a curve lying in a PLANE parallel to the yz-plane at a single x-coordinate. Plugging this back into (1) gives

$$\begin{aligned} y^2+z^2=R^2-x^2&=R^2-\left(\frac{d^2-r^2+R^2}{2d}\right)^2\\ &=\frac{4d^2R^2-(d^2-r^2+R^2)^2}{4d^2}, \qquad (6)\end{aligned}$$

which is a CIRCLE with RADIUS

$$\begin{aligned} a&=\frac{1}{2d}\sqrt{4d^2R^2-(d^2-r^2+R^2)^2}\\ &=\frac{1}{2d}[(-d+r-R)(-d-r+R)\\ &\quad\times[(-d+r+R)(d+r+R)]^{1/2}. \qquad (7)\end{aligned}$$

The VOLUME of the 3-D LENS common to the two spheres can be found by adding the two SPHERICAL CAPS. The distances from the SPHERES' centers to the bases of the caps are

$$d_1=x \qquad (8)$$
$$d_2=d-x, \qquad (9)$$

so the heights of the caps are

$$h_1=R-d_1=\frac{(r-R+d)(r+R-d)}{2d} \qquad (10)$$
$$h_2=r-d_2=\frac{(R-r+d)(R+r-d)}{2d}. \qquad (11)$$

The VOLUME of a SPHERICAL CAP of height h' for a SPHERE of RADIUS R' is

$$V(R',h')=\tfrac{1}{3}\pi h'^2(3R'-h'). \qquad (12)$$

Letting $R_1=R$ and $R_2=r$ and summing the two caps gives

$$\begin{aligned} V&=V(R_1,h_1)+V(R_2,h_2)\\ &=\frac{\pi(R+r-d)^2(d^2+2dr-3r^2+2dR+6rR-3R^2)}{12d}. \qquad (13)\end{aligned}$$

This expression gives $V=0$ for $d=r+R$ as it must. In the special case $r=R$, the VOLUME simplifies to

$$V=\tfrac{1}{12}\pi(4R+d)(2R-d)^2. \qquad (14)$$

see also APPLE, CIRCLE-CIRCLE INTERSECTION, DOUBLE BUBBLE, LENS, SPHERE

Sphere with Tunnel

Find the tunnel between two points A and B on a gravitating SPHERE which gives the shortest transit time under the force of gravity. Assume the SPHERE to be nonrotating, of RADIUS a, and with uniform density ρ. Then the standard form EULER-LAGRANGE DIFFERENTIAL EQUATION in polar coordinates is

$$r_{\phi\phi}(r^3-ra^2)+{r_\phi}^2(2a^2-r^2)+a^2r^2=0, \qquad (1)$$

along with the boundary conditions $r(\phi=0)=r_0$, $r_\phi(\phi=0)=0$, $r(\phi=\phi_A)=a$, and $r(\phi=\phi_B)=a$. Integrating once gives

$${r_\phi}^2=\frac{a^2r^2}{{r_0}^2}\frac{r^2-{r_0}^2}{a^2-r^2}. \qquad (2)$$

But this is the equation of a HYPOCYCLOID generated by a CIRCLE of RADIUS $\frac{1}{2}(a-r_0)$ rolling inside the CIRCLE of RADIUS a, so the tunnel is shaped like an arc of a HYPOCYCLOID. The transit time from point A to point B is

$$T=\pi\sqrt{\frac{a^2-{r_0}^2}{ag}}, \qquad (3)$$

where

$$g=\frac{GM}{a^2}=\tfrac{4}{3}\pi\rho Ga \qquad (4)$$

is the surface gravity with G the universal gravitational constant.

Spherical Bessel Differential Equation

Take the HELMHOLTZ DIFFERENTIAL EQUATION

$$\nabla^2F+k^2F=0 \qquad (1)$$

in SPHERICAL COORDINATES. This is just LAPLACE'S EQUATION in SPHERICAL COORDINATES with an additional term,

$$\begin{aligned}&\frac{d^2R}{dr^2}\Phi\Theta+\frac{2}{r}\frac{dR}{dr}+\frac{1}{r^2\sin^2\phi}\frac{d^2\Theta}{d\theta^2}\Phi R\\ &\quad+\frac{\cos\phi}{r^2\sin\phi}\frac{d\Phi}{d\phi}\Theta R+\frac{1}{r^2}\frac{d^2\Phi}{d\phi^2}+k^2R\Phi\Theta=0. \qquad (2)\end{aligned}$$

Multiply through by $r^2/R\Phi\Theta$,

$$\frac{r^2}{R}\frac{d^2R}{dr^2}+\frac{2r}{R}\frac{dR}{dr}+k^2r^2+\frac{1}{\Theta\sin^2\phi}\frac{d^2\Theta}{d\theta^2}+\frac{\cos\phi}{\Phi\sin\phi}\frac{d\Phi}{d\phi}+\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2}=0. \tag{3}$$

This equation is separable in R. Call the separation constant $n(n+1)$,

$$\frac{r^2}{R}\frac{d^2R}{dr^2}+\frac{2r}{R}\frac{dR}{dr}+k^2r^2=n(n+1). \tag{4}$$

Now multiply through by R,

$$r^2\frac{d^2R}{dr^2}+2r\frac{dR}{dr}+[k^2r^2-n(n+1)]R=0. \tag{5}$$

This is the SPHERICAL BESSEL DIFFERENTIAL EQUATION. It can be transformed by letting $x\equiv kr$, then

$$r\frac{dR(r)}{dr}=kr\frac{dR(r)}{k\,dr}=kr\frac{dR(r)}{d(kr)}=x\frac{dR(r)}{dx}. \tag{6}$$

Similarly,

$$r^2\frac{d^2R(r)}{dr^2}=x^2\frac{d^2R(r)}{dx^2}, \tag{7}$$

so the equation becomes

$$x^2\frac{d^2R}{dx^2}+2x\frac{dR}{dx}+[x^2-n(n+1)]R=0. \tag{8}$$

Now look for a solution of the form $R(r)=Z(x)x^{-1/2}$, denoting a derivative with respect to x by a prime,

$$R'=Z'x^{-1/2}-\tfrac{1}{2}Zx^{-3/2} \tag{9}$$

$$\begin{aligned}R''&=Z''x^{-1/2}-\tfrac{1}{2}Z'x^{-3/2}-\tfrac{1}{2}Z'x^{-3/2}\\&\quad-\tfrac{1}{2}(-\tfrac{3}{2})Zx^{-5/2}\\&=Z''x^{-1/2}-Z'x^{-3/2}+\tfrac{3}{4}Zx^{-5/2},\end{aligned} \tag{10}$$

so

$$\begin{aligned}&x^2(Z''x^{-1/2}-Z'x^{-3/2}+\tfrac{3}{4}Zx^{-5/2})\\&+2x(Z'x^{-1/2}-\tfrac{1}{2}Zx^{-3/2})+[x^2-n(n+1)]Zx^{-1/2}=0\end{aligned} \tag{11}$$

$$x^2(Z''-Z'x^{-1}+\tfrac{3}{4}Zx^{-2})+2x(Z'-\tfrac{1}{2}Zx^{-1})+[x^2-n(n+1)]Z=0 \tag{12}$$

$$x^2Z''+(-x+2x)Z'+[\tfrac{3}{4}-1+x^2-n(n+1)]Z=0 \tag{13}$$

$$\begin{aligned}&x^2Z''+xZ'+[x^2-(n^2+n+\tfrac{1}{4})]Z=0\\&x^2Z''+xZ'+[x^2-(n+\tfrac{1}{2})^2]Z=0.\end{aligned} \tag{14}$$

But the solutions to this equation are BESSEL FUNCTIONS of half integral order, so the normalized solutions to the original equation are

$$R(r)\equiv A\frac{J_{n+1/2}(kr)}{\sqrt{kr}}+B\frac{Y_{n+1/2}(kr)}{\sqrt{kr}} \tag{15}$$

which are known as SPHERICAL BESSEL FUNCTIONS. The two types of solutions are denoted $j_n(x)$ (SPHERICAL BESSEL FUNCTION OF THE FIRST KIND) or $n_n(x)$ (SPHERICAL BESSEL FUNCTION OF THE SECOND KIND), and the general solution is written

$$R(r)=A'j_n(kr)+B'n_n(kr), \tag{16}$$

where

$$j_n(z)\equiv\sqrt{\frac{\pi}{2}}\frac{J_{n+1/2}(z)}{\sqrt{z}} \tag{17}$$

$$n_n(z)\equiv\sqrt{\frac{\pi}{2}}\frac{Y_{n+1/2}(z)}{\sqrt{z}}. \tag{18}$$

see also SPHERICAL BESSEL FUNCTION, SPHERICAL BESSEL FUNCTION OF THE FIRST KIND, SPHERICAL BESSEL FUNCTION OF THE SECOND KIND

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 437, 1972.

Spherical Bessel Function

A solution to the SPHERICAL BESSEL DIFFERENTIAL EQUATION. The two types of solutions are denoted $j_n(x)$ (SPHERICAL BESSEL FUNCTION OF THE FIRST KIND) or $n_n(x)$ (SPHERICAL BESSEL FUNCTION OF THE SECOND KIND).

see also SPHERICAL BESSEL FUNCTION OF THE FIRST KIND, SPHERICAL BESSEL FUNCTION OF THE SECOND KIND

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Spherical Bessel Functions." §10.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 437–442, 1972.

Arfken, G. "Spherical Bessel Functions." §11.7 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 622–636, 1985.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Bessel Functions of Fractional Order, Airy Functions, Spherical Bessel Functions." §6.7 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 234–245, 1992.

Spherical Bessel Function of the First Kind

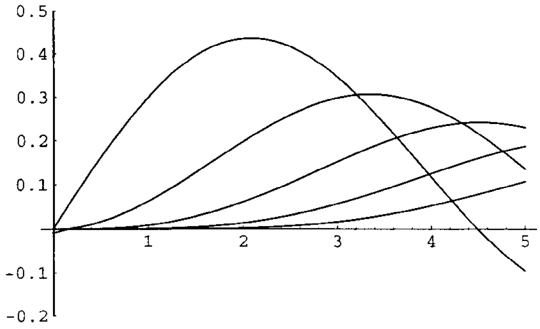

$$\begin{aligned} j_n(x) &\equiv \sqrt{\frac{\pi}{2x}} J_{n+1/2}(x) \\ &= 2^n x^n \sum_{s=0}^{\infty} \frac{(-1)^s (s-n)!}{s!(2s+2n+1)!} x^{2s} \\ &= \frac{x^n}{(2n+1)!!} \left[1 - \frac{\frac{1}{2}x^2}{1!(2n+3)} \right. \\ &\quad \left. + \frac{\left(\frac{1}{2}x^2\right)^2}{2!(2n+3)(2n+5)} + \ldots \right] \\ &= (-1)^n x^n \left(\frac{d}{x\,dx} \right)^n \frac{\sin x}{x}. \end{aligned}$$

The first few functions are

$$\begin{aligned} j_0(x) &= \frac{\sin x}{x} \\ j_1(x) &= \frac{\sin x}{x^2} - \frac{\cos x}{x} \\ j_2(x) &= \left(\frac{3}{x^3} - \frac{1}{x} \right) \sin x - \frac{3}{x^2} \cos x. \end{aligned}$$

see also POISSON INTEGRAL REPRESENTATION, RAYLEIGH'S FORMULAS

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Spherical Bessel Functions." §10.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 437–442, 1972.

Arfken, G. "Spherical Bessel Functions." §11.7 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 622–636, 1985.

Spherical Bessel Function of the Second Kind

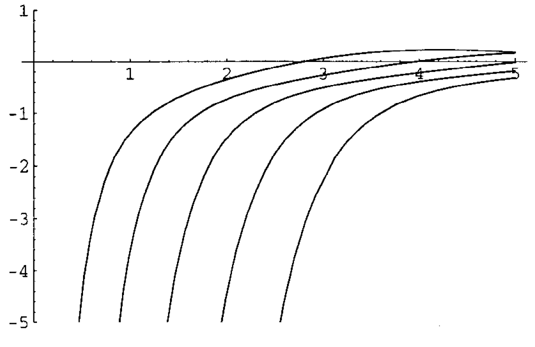

$$\begin{aligned} n_n(x) &\equiv \sqrt{\frac{\pi}{2x}} Y_{n+1/2}(x) \\ &= \frac{(-1)^{n+1}}{2^n x^{n+1}} \sum_{n=0}^{\infty} \frac{(-1)^s (s-n)!}{s!(2s-2n)!} x^{2s} \\ &= -\frac{(2n-1)!!}{x^{n+1}} \left[1 - \frac{\frac{1}{2}x^2}{1!(1-2n)} \right. \\ &\quad \left. + \frac{\left(\frac{1}{2}x^2\right)^2}{2!(1-2n)(3-2n)} + \ldots \right] \\ &= (-1)^{n+1} \sqrt{\frac{\pi}{2x}}\, J_{-n-1/2}(x). \end{aligned}$$

The first few functions are

$$\begin{aligned} n_0(x) &= -\frac{\cos x}{x} \\ n_1(x) &= -\frac{\cos x}{x^2} - \frac{\sin x}{x} \\ n_2(x) &= -\left(\frac{3}{x^3} - \frac{1}{x} \right) \cos x - \frac{3}{x^2} \sin x. \end{aligned}$$

see also RAYLEIGH'S FORMULAS

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Spherical Bessel Functions." §10.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 437–442, 1972.

Arfken, G. "Spherical Bessel Functions." §11.7 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 622–636, 1985.

Spherical Bessel Function of the Third Kind

see SPHERICAL HANKEL FUNCTION OF THE FIRST KIND, SPHERICAL HANKEL FUNCTION OF THE SECOND KIND

Spherical Cap

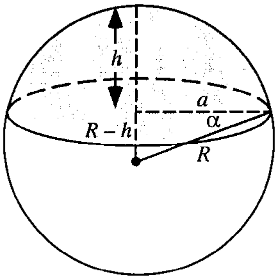

A spherical cap is the region of a SPHERE which lies above (or below) a given PLANE. If the PLANE passes through the CENTER of the SPHERE, the cap is a HEMISPHERE. Let the SPHERE have RADIUS R, then the VOLUME of a spherical cap of height h and base RADIUS a is given by the equation of a SPHERICAL SEGMENT (which is a spherical cut by a second PLANE)

$$V_{\text{spherical segment}} = \tfrac{1}{6}\pi h(3a^2 + 3b^2 + h^2) \tag{1}$$

with $b = 0$, giving

$$V_{\text{cap}} = \tfrac{1}{6}\pi h(3a^2 + h^2). \tag{2}$$

Using the PYTHAGOREAN THEOREM gives

$$(R-h)^2 + a^2 = R^2, \tag{3}$$

which can be solved for a^2 as

$$a^2 = 2Rh - h^2, \tag{4}$$

and plugging this in gives the equivalent formula

$$V_{\text{cap}} = \tfrac{1}{3}\pi h^2(3R - h). \tag{5}$$

In terms of the so-called CONTACT ANGLE (the angle between the normal to the sphere at the bottom of the cap and the base plane)

$$R - h = R\sin\theta \tag{6}$$

$$\alpha \equiv \sin^{-1}\left(\frac{R-h}{R}\right), \tag{7}$$

so

$$V_{\text{cap}} = \tfrac{1}{3}\pi R^3(2 - 3\sin\alpha + \sin^3\alpha). \tag{8}$$

Consider a cylindrical box enclosing the cap so that the top of the box is tangent to the top of the SPHERE. Then the enclosing box has VOLUME

$$\begin{aligned} V_{\text{box}} = \pi a^2 h &= \pi(R\cos\alpha)[R(1-\sin\alpha)] \\ &= \pi R^3(1 - \sin\alpha - \sin^2\alpha + \sin^3\alpha), \end{aligned} \tag{9}$$

so the hollow volume between the cap and box is given by

$$V_{\text{box}} - V_{\text{cap}} = \tfrac{1}{3}\pi R^3(1 - 3\sin^2\alpha + 2\sin^3\alpha). \tag{10}$$

If a second PLANE cuts the cap, the resulting SPHERICAL FRUSTUM is called a SPHERICAL SEGMENT. The SURFACE AREA of the spherical cap is given by the same equation as for a general ZONE:

$$S_{\text{cap}} = 2\pi Rh. \tag{11}$$

see also CONTACT ANGLE, DOME, FRUSTUM, HEMISPHERE, SOLID OF REVOLUTION, SPHERE, SPHERICAL SEGMENT, TORISPHERICAL DOME, ZONE

Spherical Coordinates

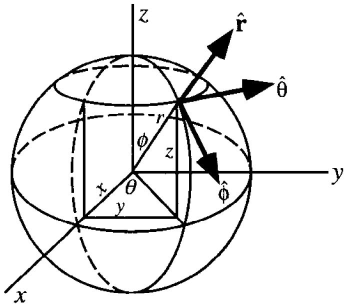

A system of CURVILINEAR COORDINATES which is natural for describing positions on a SPHERE or SPHEROID. Define θ to be the azimuthal ANGLE in the xy-PLANE from the x-AXIS with $0 \leq \theta < 2\pi$ (denoted λ when referred to as the LONGITUDE), ϕ to be the polar ANGLE from the z-AXIS with $0 \leq \phi \leq \pi$ (COLATITUDE, equal to $\phi = 90° - \delta$ where δ is the LATITUDE), and r to be distance (RADIUS) from a point to the ORIGIN.

Unfortunately, the convention in which the symbols θ and ϕ are reversed is frequently used, especially in physics, leading to unnecessary confusion. The symbol ρ is sometimes also used in place of r. Arfken (1985) uses (r, ϕ, θ), whereas Beyer (1987) uses (ρ, θ, ϕ). Be very careful when consulting the literature.

In this work, the symbols for the azimuthal, polar, and radial coordinates are taken as θ, ϕ, and r, respectively. Note that this definition provides a logical extension of the usual POLAR COORDINATES notation, with θ remaining the ANGLE in the xy-PLANE and ϕ becoming the ANGLE out of the PLANE.

$$r = \sqrt{x^2 + y^2 + z^2} \tag{1}$$

$$\theta = \tan^{-1}\left(\frac{y}{x}\right) \tag{2}$$

$$\phi = \sin^{-1}\left(\frac{\sqrt{x^2+y^2}}{r}\right) = \cos^{-1}\left(\frac{z}{r}\right), \tag{3}$$

where $r \in [0, \infty)$, $\theta \in [0, 2\pi)$, and $\phi \in [0, \pi]$. In terms of CARTESIAN COORDINATES,

$$x = r\cos\theta\sin\phi \tag{4}$$

$$y = r\sin\theta\sin\phi \tag{5}$$

$$z = r\cos\phi. \tag{6}$$

The SCALE FACTORS are

$$h_r = 1 \tag{7}$$

$$h_\theta = r\sin\phi \tag{8}$$

$$h_\phi = r, \tag{9}$$

so the METRIC COEFFICIENTS are

$$g_{rr} = 1 \qquad (10)$$

$$g_{\theta\theta} = r^2 \sin^2 \phi \qquad (11)$$

$$g_{\phi\phi} = r^2. \qquad (12)$$

The LINE ELEMENT is

$$d\mathbf{s} = dr\,\hat{\mathbf{r}} + r\,d\phi\,\hat{\boldsymbol{\phi}} + r\sin\phi\,d\theta\,\hat{\boldsymbol{\theta}}, \qquad (13)$$

the AREA element

$$d\mathbf{a} = r^2 \sin\phi\,d\theta\,d\phi\,\hat{\mathbf{r}}, \qquad (14)$$

and the VOLUME ELEMENT

$$dV = r^2 \sin\phi\,d\theta\,d\phi\,dr. \qquad (15)$$

The JACOBIAN is

$$\left|\frac{\partial(x,y,z)}{\partial(r,\theta,\phi)}\right| = r^2|\sin\phi|. \qquad (16)$$

The POSITION VECTOR is

$$\mathbf{r} \equiv \begin{bmatrix} r\cos\theta\sin\phi \\ r\sin\theta\sin\phi \\ r\cos\phi \end{bmatrix}, \qquad (17)$$

so the UNIT VECTORS are

$$\hat{\mathbf{r}} \equiv \frac{\frac{d\mathbf{r}}{dr}}{\left|\frac{d\mathbf{r}}{dr}\right|} = \begin{bmatrix} \cos\theta\sin\phi \\ \sin\theta\sin\phi \\ \cos\phi \end{bmatrix} \qquad (18)$$

$$\hat{\boldsymbol{\theta}} \equiv \frac{\frac{d\mathbf{r}}{d\theta}}{\left|\frac{d\mathbf{r}}{d\theta}\right|} = \begin{bmatrix} -\sin\theta \\ \cos\theta \\ 0 \end{bmatrix} \qquad (19)$$

$$\hat{\boldsymbol{\phi}} \equiv \frac{\frac{d\mathbf{r}}{d\phi}}{\left|\frac{d\mathbf{r}}{d\phi}\right|} = \begin{bmatrix} \cos\theta\cos\phi \\ \sin\theta\cos\phi \\ -\sin\phi \end{bmatrix}. \qquad (20)$$

Derivatives of the UNIT VECTORS are

$$\frac{\partial\hat{\mathbf{r}}}{\partial r} = \mathbf{0} \qquad (21)$$

$$\frac{\partial\hat{\boldsymbol{\theta}}}{\partial r} = \mathbf{0} \qquad (22)$$

$$\frac{\partial\hat{\boldsymbol{\phi}}}{\partial r} = \mathbf{0} \qquad (23)$$

$$\frac{\partial\hat{\mathbf{r}}}{\partial\theta} = \begin{bmatrix} -\sin\theta\sin\phi \\ \cos\theta\sin\phi \\ 0 \end{bmatrix} = \sin\phi\,\hat{\boldsymbol{\theta}} \qquad (24)$$

$$\frac{\partial\hat{\boldsymbol{\theta}}}{\partial\theta} = \begin{bmatrix} -\cos\theta \\ -\sin\theta \\ 0 \end{bmatrix} = -\cos\phi\,\hat{\boldsymbol{\phi}} - \sin\phi\,\hat{\mathbf{r}} \qquad (25)$$

$$\frac{\partial\hat{\boldsymbol{\phi}}}{\partial\theta} = \begin{bmatrix} -\sin\theta\cos\phi \\ \cos\theta\cos\phi \\ 0 \end{bmatrix} = \cos\phi\,\hat{\boldsymbol{\theta}} \qquad (26)$$

$$\frac{\partial\hat{\mathbf{r}}}{\partial\phi} = \begin{bmatrix} \cos\theta \\ \sin\theta\cos\phi \\ -\sin\phi \end{bmatrix} = \hat{\boldsymbol{\phi}} \qquad (27)$$

$$\frac{\partial\hat{\boldsymbol{\theta}}}{\partial\phi} = \mathbf{0} \qquad (28)$$

$$\frac{\partial\hat{\boldsymbol{\phi}}}{\partial\phi} = \begin{bmatrix} -\cos\theta\sin\phi \\ -\sin\theta\sin\phi \\ -\cos\phi \end{bmatrix} = -\hat{\mathbf{r}}. \qquad (29)$$

The GRADIENT is

$$\nabla = \hat{\mathbf{r}}\frac{\partial}{\partial r} + \frac{1}{r}\hat{\boldsymbol{\phi}}\frac{\partial}{\partial\phi} + \frac{1}{r\sin\phi}\hat{\boldsymbol{\theta}}\frac{\partial}{\partial\theta}, \qquad (30)$$

so

$$\nabla_r\hat{\mathbf{r}} = \mathbf{0} \qquad (31)$$

$$\nabla_r\hat{\boldsymbol{\theta}} = \mathbf{0} \qquad (32)$$

$$\nabla_r\hat{\boldsymbol{\phi}} = \hat{\mathbf{0}} \qquad (33)$$

$$\nabla_\theta\hat{\mathbf{r}} = \frac{\sin\phi\,\hat{\boldsymbol{\theta}}}{r\sin\phi} = \frac{1}{r}\hat{\boldsymbol{\theta}} \qquad (34)$$

$$\nabla_\theta\hat{\boldsymbol{\theta}} = -\frac{\cos\phi\,\hat{\boldsymbol{\phi}} + \sin\phi\,\hat{\mathbf{r}}}{r\sin\phi} = -\frac{\cot\phi}{r}\hat{\boldsymbol{\phi}} - \frac{1}{r}\hat{\mathbf{r}} \qquad (35)$$

$$\nabla_\theta\hat{\boldsymbol{\phi}} = \frac{\cos\phi\,\hat{\boldsymbol{\phi}}}{r\sin\phi} = \frac{1}{r}\cot\phi\,\hat{\boldsymbol{\theta}}. \qquad (36)$$

Now, since the CONNECTION COEFFICIENTS are given by $\Gamma^i_{jk} = \hat{\mathbf{x}}_i \cdot (\nabla_k \hat{\mathbf{x}}_j)$,

$$\Gamma^\theta = \begin{bmatrix} 0 & \frac{1}{r} & 0 \\ 0 & 0 & 0 \\ 0 & \frac{\cot\phi}{r} & 0 \end{bmatrix} \qquad (37)$$

$$\Gamma^\phi = \begin{bmatrix} 0 & 0 & \frac{1}{r} \\ 0 & -\frac{\cot\phi}{r} & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad (38)$$

$$\Gamma^r = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -\frac{1}{r} & 0 \\ 0 & 0 & -\frac{1}{r} \end{bmatrix}. \qquad (39)$$

The DIVERGENCE is

$$\begin{aligned}
\nabla \cdot \mathbf{F} &= A^k{}_{,k} + \Gamma^k_{jk}A^j \\
&= [A^r{}_{,r} + (\Gamma^r_{rr}A^r + \Gamma^r_{\theta r}A^\theta + \Gamma^r_{\phi r}A^\phi] \\
&\quad + [A^\theta{}_{,\theta} + (\Gamma^\theta_{r\theta}A^r + \Gamma^\theta_{\theta\theta}A^\theta + \Gamma^\theta_{\phi\theta}A^\phi)] \\
&\quad + [A^\phi{}_{,\phi} + (\Gamma^\phi_{r\phi}A^r + \Gamma^\phi_{\theta\phi}A^\theta + \Gamma^\phi_{\phi\phi}A^\phi)] \\
&= \frac{1}{g_r}\frac{\partial A^r}{\partial r} + \frac{1}{g_\theta}\frac{\partial A^\theta}{\partial\theta} + \frac{1}{g_\phi}\frac{\partial A^\phi}{\partial\phi} + (0+0+0) \\
&\quad + \left(\frac{1}{r}A^r + 0 + \frac{\cot\phi}{r}A^\phi\right) + \left(\frac{1}{r}A^r + 0 + 0\right) \\
&= \frac{\partial}{\partial r}A^r + \frac{2}{r}A^r + \frac{1}{r\sin\phi}\frac{\partial}{\partial\theta}A^\theta + \frac{1}{r}\frac{\partial}{\partial\phi}A^\phi + \frac{\cot\phi}{r}A^\phi,
\end{aligned} \qquad (40)$$

or, in VECTOR notation,

$$\begin{aligned}
\nabla \cdot \mathbf{F} &= \left(\frac{2}{r} + \frac{\partial}{\partial r}\right)F_r + \left(\frac{1}{r}\frac{\partial}{\partial\phi} + \frac{\cot\phi}{r}\right)F_\phi + \frac{1}{\sin\phi}\frac{\partial F_\theta}{\partial\theta} \\
&= \frac{1}{r^2}\frac{\partial}{\partial r}(r^2F_r) + \frac{1}{r\sin\phi}\frac{\partial}{\partial\phi}(\sin\phi F_\phi) + \frac{1}{r\sin\phi}\frac{\partial F_\theta}{\partial\theta}.
\end{aligned} \qquad (41)$$

The COVARIANT DERIVATIVES are given by

$$A_{j;k} = \frac{1}{g_{kk}}\frac{\partial A_j}{\partial x_k} - \Gamma^i_{jk}A_i, \qquad (42)$$

so

$$\begin{aligned}
A_{r;r} &= \frac{\partial A_r}{\partial r} - \Gamma^i_{rr} A_i = \frac{\partial A_r}{\partial r} && (43)\\
A_{r;\theta} &= \frac{1}{r\sin\phi}\frac{\partial A_r}{\partial\theta} - \Gamma^i_{r\theta} = \frac{1}{r\sin\phi}\frac{\partial A_r}{\partial\theta} - \Gamma_{r\theta}A_\theta\\
&= \frac{1}{r\sin\phi}\frac{\partial A_r}{\partial\phi} - \frac{A_\theta}{r} && (44)\\
A_{r;\phi} &= \frac{1}{r}\frac{\partial A_r}{\partial\phi} - \Gamma^i_{r\phi}A_i = \frac{1}{r}\frac{\partial A_r}{\partial\phi} - \Gamma^\phi_{r\phi}A_\phi\\
&= \frac{1}{r}\left(\frac{\partial A_r}{\partial\phi} - A_\phi\right) && (45)\\
A_{\theta;r} &= \frac{\partial A_\theta}{\partial r} - \Gamma^i_{\theta r}A_i = \frac{\partial A_\theta}{\partial r} && (46)\\
A_{\theta;\theta} &= \frac{1}{r\sin\phi}\frac{\partial A_\phi}{\partial\theta} - \Gamma^i_{\theta\theta}A_i\\
&= \frac{1}{r\sin\phi}\partial A_\theta \partial\theta - \Gamma^\phi_{\theta\theta}A_\phi - \Gamma^r_{\theta\theta}A_r\\
&= \frac{1}{r\sin\phi}\frac{\partial A_\theta}{\partial\theta} + \frac{\cot\phi}{r}A_\phi + \frac{A_r}{r} && (47)\\
A_{\theta;\phi} &= \frac{1}{r}\frac{\partial A_\theta}{\partial r} - \Gamma^i_{\phi r}A_i\frac{\partial A_\theta}{\partial\phi} && (48)\\
A_{\phi;r} &= \frac{\partial A_\phi}{\partial r} - \Gamma^i_{\phi r}A_i = \frac{\partial A_\phi}{r} && (49)\\
A_{\phi;\theta} &= \frac{1}{r\sin\phi}\frac{\partial A_\phi}{\partial\theta} - \Gamma^i_{\phi\theta}A_i = \frac{1}{r\sin\phi}\frac{\partial A_\phi}{\partial\theta} - \Gamma^\theta_{\phi\theta}\\
&= \frac{1}{r\sin\phi}\frac{\partial A_\phi}{\partial\theta} - \frac{\cot\phi}{r}A_\theta && (50)\\
A_{\phi;\phi} &= \frac{1}{r}\frac{\partial A_\phi}{\partial\phi} - \Gamma^i_{\phi\phi}A_i = \frac{1}{r}\frac{\partial A_\phi}{\partial\phi} - \Gamma^r_{\phi\phi}A_r\\
&= \frac{1}{r}\frac{\partial A_\phi}{\partial\phi} + \frac{A_r}{r}. && (51)
\end{aligned}$$

The Commutation Coefficients are given by

$$c^\mu_{\alpha\beta}\vec{e}_\mu = [\vec{e}_\alpha, \vec{e}_\beta] = \nabla_\alpha\vec{e}_\beta - \nabla_\beta\vec{e}_\alpha \qquad (52)$$

$$[\hat{\mathbf{r}}, \hat{\mathbf{r}}] = [\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}] = [\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\phi}}] = \mathbf{0}, \qquad (53)$$

so $c^\alpha_{rr} = c^\alpha_{\theta\theta} = c^\alpha_{\phi\phi} = 0$, where $\alpha = r, \theta, \phi$.

$$[\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}] = -[\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}] = \nabla_r\hat{\boldsymbol{\theta}} - \nabla_\theta\hat{\mathbf{r}} = \mathbf{0} - \frac{1}{r}\hat{\boldsymbol{\theta}} = -\frac{1}{r}\hat{\boldsymbol{\theta}}, \qquad (54)$$

so $c^\theta_{r\theta} = -c^\theta_{\theta r} = -\frac{1}{r}$, $c^r_{r\theta} = c^\phi_{r\theta} = 0$.

$$[\hat{\mathbf{r}}, \hat{\boldsymbol{\phi}}] = -[\hat{\boldsymbol{\phi}}, \hat{\mathbf{r}}] = \mathbf{0} - \frac{1}{r}\hat{\boldsymbol{\phi}} = -\frac{1}{r}\hat{\boldsymbol{\phi}}, \qquad (55)$$

so $c^\phi_{r\phi} = -c^\phi_{\phi r} = \frac{1}{r}$.

$$[\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}}] = -[\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}] = \frac{1}{r}\cot\phi\hat{\boldsymbol{\theta}} - \mathbf{0} = \frac{1}{r}\cot\phi\hat{\boldsymbol{\theta}}, \qquad (56)$$

so

$$c^\theta_{\theta\phi} = -c^\theta_{\phi\theta} = \frac{1}{r}\cot\phi. \qquad (57)$$

Summarizing,

$$c^r = \begin{bmatrix} 0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0 \end{bmatrix} \qquad (58)$$

$$c^\theta = \begin{bmatrix} 0 & -\frac{1}{r} & 0\\ \frac{1}{r} & 0 & \frac{1}{r}\cot\phi\\ 0 & -\frac{1}{r}\cot\phi & 0 \end{bmatrix} \qquad (59)$$

$$c^\phi = \begin{bmatrix} 0 & 0 & -\frac{1}{r}\\ 0 & 0 & 0\\ \frac{1}{r} & 0 & 0 \end{bmatrix}. \qquad (60)$$

Time derivatives of the Position Vector are

$$\begin{aligned}
\dot{\mathbf{r}} &= \begin{bmatrix} \cos\theta\sin\phi\,\dot{r} - r\sin\theta\sin\phi\,\dot{\theta} + r\cos\theta\cos\phi\,\dot{\phi}\\ \sin\theta\sin\phi\,\dot{r} + r\cos\theta\sin\phi\,\dot{\theta} + r\sin\theta\cos\phi\,\dot{\phi}\\ \cos\phi\,\dot{r} - r\sin\phi\,\dot{\phi} \end{bmatrix}\\
&= \begin{bmatrix} \cos\theta\sin\phi\\ \sin\theta\sin\phi\\ \cos\phi \end{bmatrix}\dot{r} + r\sin\phi\begin{bmatrix} -\sin\theta\\ \cos\theta\\ 0 \end{bmatrix}\dot{\theta}\\
&\quad + r\begin{bmatrix} \cos\theta\cos\phi\\ \sin\theta\cos\phi\\ -\sin\phi \end{bmatrix}\dot{\phi}\\
&= \dot{r}\,\hat{\mathbf{r}} + r\sin\phi\,\dot{\theta}\,\hat{\boldsymbol{\theta}} + r\,\dot{\phi}\,\hat{\boldsymbol{\phi}}. && (61)
\end{aligned}$$

The Speed is therefore given by

$$v \equiv |\dot{\mathbf{r}}| = \sqrt{\dot{r}^2 + r^2\sin^2\phi\dot{\theta}^2 + r^2\dot{\phi}^2}. \qquad (62)$$

The Acceleration is

$$\begin{aligned}
\ddot{x} &= (-\sin\theta\sin\phi\dot{\theta}\dot{r} + \cos\theta\cos\phi\dot{r}\dot{\phi} + \cos\theta\sin\phi\ddot{r})\\
&\quad - (\sin\theta\sin\phi\dot{r}\dot{\theta} + r\cos\theta\sin\phi\dot{\theta}^2 + r\sin\theta\cos\phi\dot{\theta}\dot{\phi}\\
&\quad + r\sin\theta\sin\phi\ddot{\theta}) + (\cos\theta\cos\phi\dot{r}\dot{\phi} - r\sin\theta\cos\dot{\theta}\dot{\phi}\\
&\quad - r\cos\theta\sin\phi\dot{\phi}^2 + r\cos\theta\cos\phi\ddot{\phi})\\
&= -2\sin\theta\sin\phi\dot{\theta}\dot{r} + 2\cos\theta\cos\phi\dot{r}\dot{\phi} - 2r\sin\theta\cos\phi\dot{\theta}\dot{\phi}\\
&\quad + \cos\theta\sin\phi\ddot{r} - r\sin\theta\sin\phi\ddot{\theta} + r\cos\theta\cos\phi\ddot{\phi}\\
&\quad - r\cos\theta\sin\phi(\dot{\theta}^2 + \dot{\phi}^2) && (63)\\
\ddot{y} &= (\sin\theta\sin\phi\ddot{r} + r\cos\theta\sin\phi\dot{\theta} + r\cos\phi\sin\theta\dot{\phi})\\
&\quad + (\cos\theta\sin\phi\dot{r}\dot{\theta} - r\sin\theta\sin\phi\dot{\theta}^2 + r\cos\theta\cos\phi\dot{\theta}\dot{\phi}\\
&\quad + r\cos\theta\sin\phi\ddot{\theta}) + (\sin\theta\cos\phi\dot{r}\dot{\phi} + r\cos\theta\cos\phi\dot{\theta}\dot{\phi}\\
&\quad - r\sin\theta\sin\phi\dot{\phi}^2 + r\sin\theta\cos\phi\ddot{\phi})\\
&= 2\cos\theta\sin\phi\dot{\theta}\dot{r} + 2\sin\theta\cos\phi\dot{r}\dot{\phi} + 2r\cos\theta\cos\phi\dot{\theta}\dot{\phi}\\
&\quad + \sin\theta\sin\phi\ddot{r} + r\cos\theta\sin\phi\ddot{\theta} + r\sin\theta\cos\phi\ddot{\phi}\\
&\quad - r\sin\theta\sin\phi(\dot{\theta}^2 + \dot{\phi}^2) && (64)\\
\ddot{z} &= (\cos\phi\ddot{r} - \sin\phi\dot{r}\dot{\phi}) - (\dot{r}\sin\phi\dot{\phi} + r\cos\phi\dot{\phi}^2 + r\sin\phi\ddot{\phi})\\
&= -r\cos\phi\dot{\phi}^2 + \cos\phi\ddot{r} - 2\sin\phi\dot{\phi}\dot{r} - r\sin\phi\ddot{\phi}. && (65)
\end{aligned}$$

Plugging these in gives

$$\ddot{\mathbf{r}} = (\ddot{r} - r\dot{\phi}^2)\begin{bmatrix}\cos\theta\sin\phi\\ \sin\theta\sin\phi\\ \cos\phi\end{bmatrix} + (2r\cos\phi\dot{\theta}\dot{\phi} + r\sin\phi\ddot{\theta})\begin{bmatrix}-\sin\theta\\ \cos\theta\\ 0\end{bmatrix} + (2\dot{r}\dot{\phi} + r\ddot{\phi})\begin{bmatrix}\cos\theta\cos\phi\\ \sin\theta\cos\phi\\ -\sin\phi\end{bmatrix} - r\sin\phi\dot{\theta}^2\begin{bmatrix}\cos\theta\\ \sin\theta\\ 0\end{bmatrix}, \tag{66}$$

but

$$\sin\phi\hat{\mathbf{r}} + \cos\phi\hat{\boldsymbol{\phi}} = \begin{bmatrix}\cos\theta\sin^2\phi + \cos\theta\cos^2\phi\\ \sin\theta\sin^2\phi + \sin\theta\cos^2\phi\\ 0\end{bmatrix} = \begin{bmatrix}\cos\theta\\ \sin\theta\\ 0\end{bmatrix}, \tag{67}$$

so

$$\begin{aligned}\ddot{\mathbf{r}} &= (\ddot{r} - r\dot{\phi}^2)\hat{\mathbf{r}} + (2r\cos\phi\dot{\theta}\dot{\phi} + 2\sin\phi\dot{\theta}\dot{r} + r\sin\phi\ddot{\theta})\hat{\boldsymbol{\theta}}\\ &\quad + (2\dot{r}\dot{\phi} + r\ddot{\phi})\hat{\boldsymbol{\phi}} - r\sin\phi\dot{\theta}^2(\sin\phi\hat{\mathbf{r}} + \cos\phi\hat{\boldsymbol{\phi}})\\ &= (\ddot{r} - r\dot{\phi}^2 - r\sin^2\phi\dot{\theta}^2)\hat{\mathbf{r}}\\ &\quad + (2\sin\phi\dot{\theta}\dot{r} + 2r\cos\phi\dot{\theta}\dot{\phi} + r\sin\phi\ddot{\theta})\hat{\boldsymbol{\theta}}\\ &\quad + (2\dot{r}\dot{\phi} + r\ddot{\phi} - r\sin\phi\cos\phi\dot{\theta}^2)\hat{\boldsymbol{\phi}}.\end{aligned} \tag{68}$$

Time DERIVATIVES of the UNIT VECTORS are

$$\dot{\hat{\mathbf{r}}} = \begin{bmatrix}-\sin\theta\sin\phi\,\dot{\theta} + \cos\theta\cos\phi\,\dot{\phi}\\ \cos\theta\sin\phi\,\dot{\theta} + \sin\theta\cos\phi\,\dot{\phi}\\ -\sin\phi\,\dot{\phi}\end{bmatrix} = \sin\phi\,\dot{\theta}\hat{\boldsymbol{\theta}} + \dot{\phi}\hat{\boldsymbol{\phi}} \tag{69}$$

$$\dot{\hat{\boldsymbol{\theta}}} = \begin{bmatrix}-\cos\theta\dot{\theta}\\ -\sin\theta\dot{\theta}\\ 0\end{bmatrix} = -\dot{\theta}\begin{bmatrix}\cos\theta\\ \sin\theta\\ 0\end{bmatrix} = -\dot{\theta}(\sin\phi\hat{\mathbf{r}} + \cos\phi\hat{\boldsymbol{\phi}}) \tag{70}$$

$$\dot{\hat{\boldsymbol{\phi}}} = \begin{bmatrix}-\sin\theta\cos\phi\,\dot{\theta} - \cos\theta\sin\phi\,\dot{\phi}\\ \cos\theta\cos\phi\,\dot{\theta} - \sin\theta\sin\phi\,\dot{\phi}\\ -\cos\phi\,\dot{\phi}\end{bmatrix} = -\dot{\phi}\hat{\mathbf{r}} + \cos\phi\dot{\theta}\hat{\boldsymbol{\theta}}. \tag{71}$$

The CURL is

$$\nabla\times\mathbf{F} = \frac{1}{r\sin\phi}\left[\frac{\partial}{\partial\phi}(\sin\phi F_\theta) - \frac{\partial F_\phi}{\partial\theta}\right]\hat{\mathbf{r}} + \frac{1}{r}\left[\frac{1}{\sin\phi}\frac{\partial F_r}{\partial\theta} - \frac{\partial}{\partial r}(rF_\theta)\right]\hat{\boldsymbol{\phi}} + \frac{1}{r}\left[\frac{\partial}{\partial r}(rF_\phi) - \frac{\partial F_r}{\partial\phi}\right]\hat{\boldsymbol{\theta}}. \tag{72}$$

The LAPLACIAN is

$$\begin{aligned}\nabla^2 &\equiv \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin^2\phi}\frac{\partial^2}{\partial\theta^2}\\ &\quad + \frac{1}{r^2\sin\phi}\frac{\partial}{\partial\phi}\left(\sin\phi\frac{\partial}{\partial\phi}\right)\\ &= \frac{1}{r^2}\left(r^2\frac{\partial^2}{\partial r^2} + 2r\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin^2\phi}\frac{\partial^2}{\partial\theta^2}\\ &\quad + \frac{1}{r^2\sin\phi}\left(\cos\phi\frac{\partial}{\partial\phi} + \sin\phi\frac{\partial^2}{\partial\phi^2}\right)\\ &= \frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} + \frac{1}{r^2\sin^2\phi}\frac{\partial^2}{\partial\theta^2}\\ &\quad + \frac{\cos\phi}{r^2\sin\phi}\frac{\partial}{\partial\phi} + \frac{1}{r^2}\frac{\partial^2}{\partial\phi^2}.\end{aligned} \tag{73}$$

The vector LAPLACIAN is

$$\nabla^2\mathbf{v} = \begin{bmatrix}\frac{1}{r}\frac{\partial^2(rv_r)}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 v_r}{\partial\theta^2} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 v_r}{\partial\phi^2} + \frac{\cot\theta}{r^2}\frac{\partial v_r}{\theta}\\ \quad - \frac{2}{r^2}\frac{\partial v_\theta}{\partial\theta} - \frac{2}{r^2\sin\theta}\frac{\partial v_\phi}{\partial\phi} - \frac{2v_r}{r^2} - \frac{2\cot\theta}{r^2}v_\theta\\ \frac{1}{r}\frac{\partial^2(rv_\theta)}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 v_\theta}{\partial\theta^2} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 v_\theta}{\partial\phi^2} + \frac{\cot\theta}{r^2}\frac{\partial v_\theta}{\partial\theta}\\ \quad - \frac{2}{r^2}\frac{2\cot\theta}{r^2\sin\theta}\frac{\partial v_\phi}{\partial\phi} + \frac{2}{r^2}\frac{\partial v_r}{\partial\theta} - \frac{v_\theta}{r^2\sin^2\theta}\\ \frac{1}{r}\frac{\partial^2(rv_\phi)}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 v_\phi a}{\partial\theta^2} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 v_\phi}{\partial\phi^2} + \frac{\cot\theta}{r^2}\frac{\partial v_\phi}{\partial\theta}\\ \quad + \frac{2}{r^2}\frac{\partial v_r}{\partial\phi} + \frac{2\cot\theta}{r^2\sin\theta}\frac{\partial v_\theta}{\partial\phi} - \frac{v_\phi}{r^2\sin^2\theta}\end{bmatrix}. \tag{74}$$

To express PARTIAL DERIVATIVES with respect to Cartesian axes in terms of PARTIAL DERIVATIVES of the spherical coordinates,

$$\begin{bmatrix}x\\ y\\ z\end{bmatrix} = \begin{bmatrix}r\cos\theta\sin\phi\\ r\sin\theta\sin\phi\\ r\cos\phi\end{bmatrix} \tag{75}$$

$$\begin{bmatrix}dx\\ dy\\ dz\end{bmatrix} = \begin{bmatrix}\cos\theta\sin\phi\,dr - r\sin\theta\sin\phi\,d\theta + r\cos\theta\cos\phi\,d\phi\\ \sin\theta\sin\phi\,dr + r\sin\phi\cos\theta\,d\theta + r\sin\theta\cos\phi\,d\phi\\ \cos\phi\,dr - r\sin\phi\,d\phi\end{bmatrix} = \begin{bmatrix}\cos\theta\sin\phi & -r\sin\theta\sin\phi & r\cos\theta\cos\phi\\ \sin\theta\sin\phi & r\sin\phi\cos\theta & r\sin\theta\cos\phi\\ \cos\phi & 0 & -r\sin\phi\end{bmatrix}\begin{bmatrix}dr\\ d\theta\\ d\phi\end{bmatrix}. \tag{76}$$

Upon inversion, the result is

$$\begin{bmatrix}dr\\ d\theta\\ d\phi\end{bmatrix} = \begin{bmatrix}\cos\theta\sin\phi & \sin\theta\sin\phi & \cos\phi\\ -\frac{\sin\theta}{r\sin\phi} & \frac{\cos\theta}{r\sin\phi} & 0\\ \frac{\cos\theta\cos\phi}{r} & \frac{\sin\theta\cos\phi}{r} & -\frac{\sin\phi}{r}\end{bmatrix}\begin{bmatrix}dx\\ dy\\ dz\end{bmatrix}. \tag{77}$$

The Cartesian PARTIAL DERIVATIVES in spherical coordinates are therefore

$$\begin{aligned}\frac{\partial}{\partial x} &= \frac{\partial r}{\partial x}\frac{\partial}{\partial r} + \frac{\partial \theta}{\partial x}\frac{\partial}{\partial \theta} + \frac{\partial \phi}{\partial x}\frac{\partial}{\partial \phi} \\ &= \cos\theta\sin\phi\frac{\partial}{\partial r} - \frac{\sin\theta}{r\sin\phi}\frac{\partial}{\partial\theta} + \frac{\cos\theta\cos\phi}{r}\frac{\partial}{\partial\phi}\end{aligned} \tag{78}$$

$$\begin{aligned}\frac{\partial}{\partial y} &= \frac{\partial r}{\partial y}\frac{\partial}{\partial r} + \frac{\partial \theta}{\partial y}\frac{\partial}{\partial \theta} + \frac{\partial \phi}{\partial y}\frac{\partial}{\partial \phi} \\ &= \sin\theta\sin\phi\frac{\partial}{\partial r} + \frac{\cos\theta}{r\sin\phi}\frac{\partial}{\partial\theta} + \frac{\sin\theta\cos\phi}{r}\frac{\partial}{\partial\phi}\end{aligned} \tag{79}$$

$$\begin{aligned}\frac{\partial}{\partial z} &= \frac{\partial r}{\partial z}\frac{\partial}{\partial r} + \frac{\partial \theta}{\partial z}\frac{\partial}{\partial \theta} + \frac{\partial \phi}{\partial z}\frac{\partial}{\partial \phi} \\ &= \cos\phi\frac{\partial}{\partial r} - \frac{\sin\phi}{r}\frac{\partial}{\partial\phi}\end{aligned} \tag{80}$$

(Gasiorowicz 1974, pp. 167–168).

The HELMHOLTZ DIFFERENTIAL EQUATION is separable in spherical coordinates.

see also COLATITUDE, GREAT CIRCLE, HELMHOLTZ DIFFERENTIAL EQUATION—SPHERICAL COORDINATES, LATITUDE, LONGITUDE, OBLATE SPHEROIDAL COORDINATES, PROLATE SPHEROIDAL COORDINATES

References

Arfken, G. "Spherical Polar Coordinates." §2.5 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 102–111, 1985.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 212, 1987.

Gasiorowicz, S. *Quantum Physics.* New York: Wiley, 1974.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 658, 1953.

Spherical Design

X is a spherical t-design in E IFF it is possible to exactly determine the average value on E of any POLYNOMIAL f of degree at most t by sampling f at the points of X. In other words,

$$\frac{1}{\text{volume } E}\int_E f(\xi)\,d\xi = \frac{1}{|X|}\sum_{x\in X} f(x).$$

References

Colbourn, C. J. and Dinitz, J. H. (Eds.) "Spherical t-Designs." Ch. 44 in *CRC Handbook of Combinatorial Designs.* Boca Raton, FL: CRC Press, pp. 462–466, 1996.

Spherical Excess

The difference between the sum of the angles of a SPHERICAL TRIANGLE and 180°.

see also ANGULAR DEFECT, DESCARTES TOTAL ANGULAR DEFECT, GIRARD'S SPHERICAL EXCESS FORMULA, L'HUILIER'S THEOREM, SPHERICAL TRIANGLE

Spherical Frustum

see SPHERICAL SEGMENT

Spherical Geometry

The study of figures on the surface of a SPHERE (such as the SPHERICAL TRIANGLE and SPHERICAL POLYGON), as opposed to the type of geometry studied in PLANE GEOMETRY or SOLID GEOMETRY.

see also PLANE GEOMETRY, SOLID GEOMETRY, SPHERICAL TRIGONOMETRY, THURSTON'S GEOMETRIZATION CONJECTURE

Spherical Hankel Function of the First Kind

$$h_n^{(1)}(x) \equiv \sqrt{\frac{\pi}{2x}}\,H_{n+1/2}^{(1)}(x) = j_n(x) + in_n(x),$$

where $H^{(1)}(x)$ is the HANKEL FUNCTION OF THE FIRST KIND and $j_n(x)$ and $n_n(x)$ are the SPHERICAL BESSEL FUNCTIONS OF THE FIRST and SECOND KINDS. Explicitly, the first few are

$$\begin{aligned}h_0^{(1)}(x) &= \frac{1}{x}(\sin x - i\cos x) = -\frac{i}{x}e^{ix} \\ h_1^{(1)}(x) &= e^{ix}\left(-\frac{1}{x} - \frac{i}{x^2}\right) \\ h_2^{(1)}(x) &= e^{ix}\left(\frac{i}{x} - \frac{3}{x^2} - \frac{3i}{x^3}\right).\end{aligned}$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Spherical Bessel Functions." §10.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 437–442, 1972.

Spherical Hankel Function of the Second Kind

$$h_n^{(2)}(x) \equiv \sqrt{\frac{\pi}{2x}}\,H_{n+1/2}^{(2)}(x) = j_n(x) - in_n(x),$$

where $H^{(2)}(x)$ is the HANKEL FUNCTION OF THE SECOND KIND and $j_n(x)$ and $n_n(x)$ are the SPHERICAL BESSEL FUNCTIONS OF THE FIRST and SECOND KINDS. Explicitly, the first is

$$h_0^{(2)}(x) = \frac{1}{x}(\sin x + i\cos x) = \frac{i}{x}e^{-ix}.$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Spherical Bessel Functions." §10.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 437–442, 1972.

Spherical Harmonic

The spherical harmonics $Y_l^m(\theta,\phi)$ are the angular portion of the solution to LAPLACE'S EQUATION in SPHERICAL COORDINATES where azimuthal symmetry is not present. Some care must be taken in identifying the notational convention being used. In the below equations, θ is taken as the azimuthal (longitudinal) coordinate, and ϕ as the polar (latitudinal) coordinate (opposite the notation of Arfken 1985).

$$Y_l^m(\theta,\phi) \equiv \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\phi)e^{im\theta}, \qquad (1)$$

where $m = -l, -1+1, \ldots, 0, \ldots, l$ and the normalization is chosen such that

$$\int_0^{2\pi}\int_0^{\pi} Y_l^m Y_{l'}^{m'\,*}\sin\phi\,d\phi\,d\theta = \int_0^{2\pi}\int_{-1}^{1} Y_l^m Y_{l'}^{m'\,*}\,d(\cos\phi)\,d\theta = \delta_{mm'}\delta_{ll'}, \qquad (2)$$

where δ_{mn} is the KRONECKER DELTA. Sometimes, the CONDON-SHORTLEY PHASE $(-1)^m$ is prepended to the definition of the spherical harmonics.

Integrals of the spherical harmonics are given by

$$\int Y_{l_1}^{m_1}Y_{l_2}^{m_2}Y_{l_3}^{m_3}\,d\Omega = \sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}} \times \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix}, \qquad (3)$$

where $\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$ is a WIGNER 3j-SYMBOL (which is related to the CLEBSCH-GORDON COEFFICIENTS). The spherical harmonics obey

$$Y_l^{-l} = \frac{1}{2^l l!}\sqrt{\frac{(2l+1)!}{4\pi}}\sin^l\phi e^{-il\theta} \qquad (4)$$

$$Y_l^0 = \sqrt{\frac{2l+1}{4\pi}}P_l(\cos\phi) \qquad (5)$$

$$Y_l^{-m} = (-1)^m Y_l^{m*}, \qquad (6)$$

where $P_l(x)$ is a LEGENDRE POLYNOMIAL.

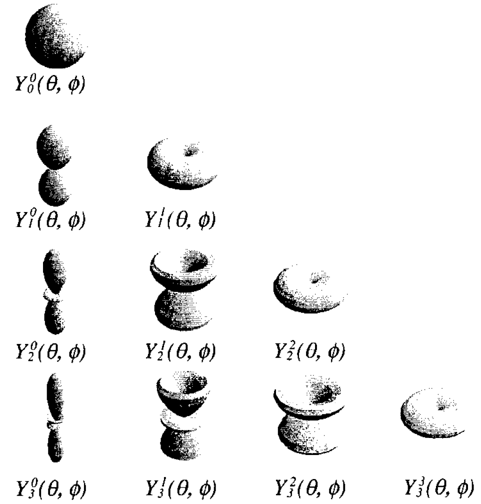

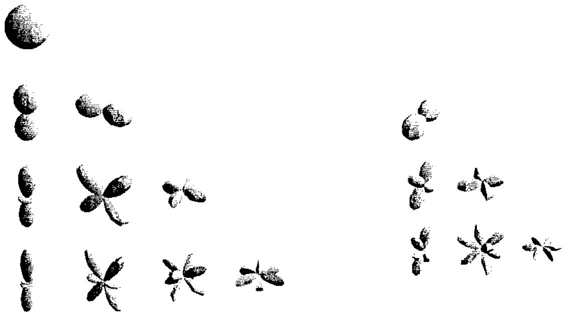

The above illustrations show $|Y_l^m(\theta,\phi)|$ (top) and $\Re[Y_l^m(\theta,\phi)]$ and $\Im[Y_l^m(\theta,\phi)]$ (bottom). The first few spherical harmonics are

$$Y_0^0 = \frac{1}{2}\frac{1}{\sqrt{\pi}}$$

$$Y_1^{-1} = \frac{1}{2}\sqrt{\frac{3}{2\pi}}\sin\phi\,e^{-i\theta}$$

$$Y_1^0 = \frac{1}{2}\sqrt{\frac{3}{\pi}}\cos\phi$$

$$Y_1^1 = -\frac{1}{2}\sqrt{\frac{3}{2\pi}}\sin\phi\,e^{i\theta}$$

$$Y_2^{-2} = \frac{1}{4}\sqrt{\frac{15}{2\pi}}\sin^2\phi\,e^{-2i\theta}$$

$$Y_2^{-1} = \frac{1}{2}\sqrt{\frac{15}{2\pi}}\sin\phi\cos\phi\,e^{-i\theta}$$

$$Y_2^0 = \frac{1}{4}\sqrt{\frac{5}{\pi}}(3\cos^2\phi - 1)$$

$$Y_2^1 = -\frac{1}{2}\sqrt{\frac{15}{2\pi}}\sin\phi\cos\phi\,e^{i\theta}$$

$$Y_2^2 = \frac{1}{4}\sqrt{\frac{15}{2\pi}}\sin^2\phi\,e^{2i\theta}$$

$$Y_3^{-3} = \frac{1}{8}\sqrt{\frac{35}{\pi}}\sin^3\phi\,e^{-3i\theta}$$

$$Y_3^{-2} = \frac{1}{4}\sqrt{\frac{105}{2\pi}}\sin^2\phi\cos\phi\,e^{-2i\theta}$$

$$Y_3^{-1} = \frac{1}{8}\sqrt{\frac{21}{\pi}}\sin\phi(5\cos^2\phi - 1)e^{-i\theta}$$

$$Y_3^0 = \frac{1}{4}\sqrt{\frac{7}{\pi}}(5\cos^3\phi - 3\cos\phi)$$

$$Y_3^1 = -\frac{1}{8}\sqrt{\frac{21}{\pi}}\sin\phi(5\cos^2\phi - 1)e^{i\theta}$$

$$Y_3^2 = \frac{1}{4}\sqrt{\frac{105}{2\pi}}\sin^2\phi\cos\phi\,e^{2i\theta}$$

$$Y_3^3 = -\frac{1}{8}\sqrt{\frac{35}{\pi}}\sin^3\phi\,e^{3i\theta}.$$

Written in terms of CARTESIAN COORDINATES,

$$e^{i\theta} = \frac{x+iy}{\sqrt{x^2+y^2}} \tag{7}$$

$$\phi = \sin^{-1}\left(\sqrt{\frac{x^2+y^2}{x^2+y^2+z^2}}\right) \tag{8}$$

$$= \cos^{-1}\left(\frac{z}{\sqrt{x^2+y^2+z^2}}\right), \tag{9}$$

so

$$Y_0^0 = \frac{1}{2}\frac{1}{\sqrt{\pi}} \tag{10}$$

$$Y_1^0 = \frac{1}{2}\sqrt{\frac{3}{\pi}}\frac{z}{\sqrt{x^2+y^2+z^2}} \tag{11}$$

$$Y_1^1 = -\frac{1}{2}\sqrt{\frac{3}{2\pi}}\frac{x+iy}{\sqrt{x^2+y^2+z^2}} \tag{12}$$

$$Y_2^0 = \frac{1}{4}\sqrt{\frac{5}{\pi}}\left(\frac{3z^2}{x^2+y^2+z^2}-1\right) \tag{13}$$

$$Y_2^1 = -\frac{1}{2}\sqrt{\frac{15}{2\pi}}\frac{z(x+iy)}{x^2+y^2+z^2} \tag{14}$$

$$Y_2^2 = \frac{1}{4}\sqrt{\frac{15}{2\pi}}\frac{(x+iy)^2}{x^2+y^2+z^2}. \tag{15}$$

These can be separated into their REAL and IMAGINARY PARTS

$$Y_l^{m\,s}(\theta,\phi) \equiv P_l^m(\cos\phi)\sin(m\theta) \tag{16}$$

$$Y_l^{m\,c}(\theta,\phi) \equiv P_l^m(\cos\phi)\cos(m\theta). \tag{17}$$

The ZONAL HARMONICS are defined to be those of the form

$$P_n^m(\cos\theta). \tag{18}$$

The TESSERAL HARMONICS are those of the form

$$\sin(m\phi)P_n^m(\cos\theta) \tag{19}$$

$$\cos(m\phi)P_n^m(\cos\theta) \tag{20}$$

for $n \neq m$. The SECTORIAL HARMONICS are of the form

$$\sin(m\phi)P_m^m(\cos\theta) \tag{21}$$

$$\cos(m\phi)P_m^m(\cos\theta). \tag{22}$$

The spherical harmonics form a COMPLETE ORTHONORMAL BASIS, so an arbitrary REAL function $f(\theta,\phi)$ can be expanded in terms of COMPLEX spherical harmonics

$$f(\theta,\phi) \equiv \sum_{l=0}^{\infty}\sum_{m=-1}^{l} A_l^m Y_l^m(\theta,\phi), \tag{23}$$

or REAL spherical harmonics

$$f(\theta,\phi) \equiv \sum_{l=0}^{\infty}\sum_{m=0}^{l}[C_l^m Y_l^{m\,c}(\theta,\phi)\sin(m\theta) + S_l^m Y_l^{m\,s}(\theta,\phi)]. \tag{24}$$

see also CORRELATION COEFFICIENT, SPHERICAL HARMONIC ADDITION THEOREM, SPHERICAL HARMONIC CLOSURE RELATIONS, SPHERICAL VECTOR HARMONIC

References

Arfken, G. "Spherical Harmonics." §12.6 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 680–685, 1985.

Ferrers, N. M. *An Elementary Treatise on Spherical Harmonics and Subjects Connected with Them.* London: Macmillan, 1877.

Groemer, H. *Geometric Applications of Fourier Series and Spherical Harmonics.* New York: Cambridge University Press, 1996.

Hobson, E. W. *The Theory of Spherical and Ellipsoidal Harmonics.* New York: Chelsea, 1955.

MacRobert, T. M. and Sneddon, I. N. *Spherical Harmonics: An Elementary Treatise on Harmonic Functions, with Applications, 3rd ed. rev.* Oxford, England: Pergamon Press, 1967.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Spherical Harmonics." §6.8 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 246–248, 1992.

Sansone, G. "Harmonic Polynomials and Spherical Harmonics," "Integral Properties of Spherical Harmonics and the Addition Theorem for Legendre Polynomials," and "Completeness of Spherical Harmonics with Respect to Square Integrable Functions." §3.18–3.20 in *Orthogonal Functions, rev. English ed.* New York: Dover, pp. 253–272, 1991.

Sternberg, W. and Smith, T. L. *The Theory of Potential and Spherical Harmonics, 2nd ed.* Toronto: University of Toronto Press, 1946.

Spherical Harmonic Addition Theorem

A FORMULA also known as the LEGENDRE ADDITION THEOREM which is derived by finding GREEN'S FUNCTIONS for the SPHERICAL HARMONIC expansion and equating them to the generating function for LEGENDRE POLYNOMIALS. When γ is defined by

$$\cos\gamma \equiv \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2\cos\phi_1 - \phi_2,$$

$$P_n(\cos\gamma) = \frac{4\pi}{2n+1}\sum_{m=-n}^{n}(-1)^m Y_m^n(\theta_1,\phi_1)Y_{-m}^n(\theta_2,\phi_2)$$

$$= \frac{4\pi}{2n+1}\sum_{m=-n}^{n} Y_m^n(\theta_1,\phi_1)Y_m^{n\,*}(\theta_2,\phi_2)$$

$$= P_n(\cos\theta_1)P_n(\cos\theta_2)$$

$$+2\sum_{m=-n}^{n}\frac{(n-m)!}{(n+m)!}P_m^n(\cos\theta_1)P_m^n(\cos\theta_2)\cos[m(\phi_1-\phi_2)].$$

References

Arfken, G. "The Addition Theorem for Spherical Harmonics." §12.8 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 693–695, 1985.

Spherical Harmonic Closure Relations

The sum of the absolute squares of the SPHERICAL HARMONICS $Y_l^m(\theta,\phi)$ over all values of m is

$$\sum_{m=-l}^{l} |Y_l^m(\theta,\phi)|^2 = \frac{2l+1}{4\pi}.$$

The double sum over m and l is given by

$$\begin{aligned}\sum_{l=0}^{\infty}\sum_{m=-l}^{l} Y_l^m(\theta_1,\phi_1)Y_l^{m*}(\theta_2,\phi_2)\\ &= \frac{1}{\sin\theta_1}\delta(\theta_1-\theta_2)\delta(\phi_1-\phi_2)\\ &= \delta(\cos\theta_1-\cos\theta_2)\delta(\phi_1-\phi_2),\end{aligned}$$

where $\delta(x)$ is the DELTA FUNCTION.

Spherical Harmonic Tensor

A tensor defined in terms of the TENSORS which satisfy the DOUBLE CONTRACTION RELATION.

see also DOUBLE CONTRACTION RELATION, SPHERICAL HARMONIC

Spherical Helix

The TANGENT INDICATRIX of a CURVE OF CONSTANT PRECESSION is a spherical helix. The equation of a spherical helix on a SPHERE with RADIUS r making an ANGLE θ with the z-axis is

$$\begin{aligned}x(\psi) &= \tfrac{1}{2}r(1+\cos\theta)\cos\psi\\ &\quad -\tfrac{1}{2}r(1-\cos\theta)\cos\left(\frac{1+\cos\theta}{1-\cos\theta}\psi\right) \qquad (1)\\ y(\psi) &= \tfrac{1}{2}r(1+\cos\theta)\sin\psi\\ &\quad -\tfrac{1}{2}r(1-\cos\theta)\sin\left(\frac{1+\cos\theta}{1-\cos\theta}\psi\right) \qquad (2)\\ z(\psi) &= r\sin\theta\cos\left(\frac{\cos\theta}{1-\cos\theta}\psi\right). \qquad (3)\end{aligned}$$

The projection on the xy-plane is an EPICYCLOID with RADII

$$a = r\cos\theta \qquad (4)$$

$$b = r\sin^2(\tfrac{1}{2}\theta). \qquad (5)$$

see also HELIX, LOXODROME, SPHERICAL SPIRAL

References

Scofield, P. D. "Curves of Constant Precession." *Amer. Math. Monthly* **102**, 531–537, 1995.

Spherical Point System

How can n points be distributed on a SPHERE such that they maximize the minimum distance between any pair of points? This is FEJES TÓTH'S PROBLEM.

see also FEJES TÓTH'S PROBLEM

Spherical Polygon

A closed geometric figure on the surface of a SPHERE which is formed by the ARCS of GREAT CIRCLES. The spherical polygon is a generalization of the SPHERICAL TRIANGLE. If θ is the sum of the RADIAN ANGLES of a spherical polygon on a SPHERE of RADIUS r, then the AREA is

$$S = [\theta-(n-2)\pi]r^2.$$

see also GREAT CIRCLE, SPHERICAL TRIANGLE

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 131, 1987.

Spherical Ring

A SPHERE with a CYLINDRICAL HOLE cut so that the centers of the CYLINDER and SPHERE coincide, also called a NAPKIN RING.

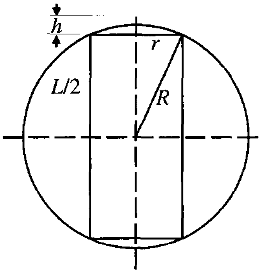

The volume of the entire CYLINDER is

$$V_{\text{cyl}} = \pi LR^2, \qquad (1)$$

and the VOLUME of the upper segment is

$$V_{\text{seg}} = \tfrac{1}{6}\pi h(3R^2+h^2), \qquad (2)$$

where

$$R = \sqrt{r^2-\tfrac{1}{4}L^2} \qquad (3)$$

$$h = r-\tfrac{1}{2}L, \qquad (4)$$

so the VOLUME removed upon drilling of a CYLINDRICAL hole is

$$\begin{aligned}V_{\text{rem}} &= V_{\text{cyl}}+2V_{\text{seg}} = \pi[LR^2+\tfrac{1}{3}h(3R^2+h^2)]\\ &= \pi(LR^2+hR^2+\tfrac{1}{3}h^3)\\ &= \pi[L(r^2-\tfrac{1}{4}L^2)+(r-\tfrac{1}{2}L)(r^2-\tfrac{1}{4}L^2)\\ &\quad +\tfrac{1}{3}(r-\tfrac{1}{2}L)^3]\\ &= \pi[Lr^2-\tfrac{1}{4}L^3+(r^3-\tfrac{1}{2}r^2L-\tfrac{1}{4}RL^2+\tfrac{1}{8}L^3)\\ &\quad +\tfrac{1}{3}(r^3-\tfrac{3}{2}r^2L+\tfrac{3}{4}rL^2-\tfrac{1}{8}L^3)]\\ &= \pi[\tfrac{4}{3}r^3+(1-\tfrac{1}{2}-\tfrac{1}{2})r^2L+(-\tfrac{1}{4}+\tfrac{1}{4})RL^2\\ &\quad +L^3(-\tfrac{1}{4}+\tfrac{1}{8}-\tfrac{1}{24})]\\ &= \tfrac{4}{3}\pi r^3-\tfrac{1}{6}\pi L^3 = \tfrac{1}{6}\pi(8r^3-L^3), \qquad (5)\end{aligned}$$

so

$$V_{\text{left}} = V_{\text{sphere}} - V_{\text{rem}} = \tfrac{4}{3}\pi r^3 - (\tfrac{4}{3}\pi r^3 - \tfrac{1}{6}\pi L^3) = \tfrac{1}{6}\pi L^3. \tag{6}$$

Spherical Sector

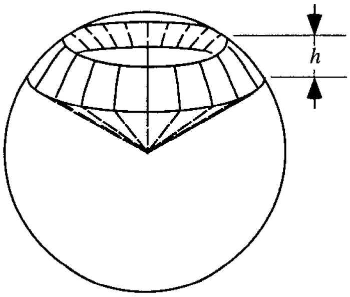

The VOLUME of a spherical sector, depicted above, is given by

$$V = \tfrac{2}{3}\pi R^2 h,$$

where h is the vertical height of the upper and lower curves.

see also CYLINDRICAL SEGMENT, SPHERE, SPHERICAL CAP, SPHERICAL SEGMENT, ZONE

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 131, 1987.

Spherical Segment

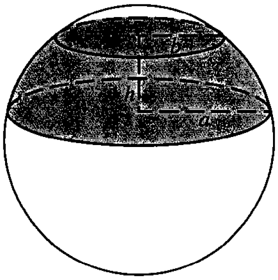

A spherical segment is the solid defined by cutting a SPHERE with a pair of PARALLEL PLANES. It can be thought of as a SPHERICAL CAP with the top truncated, and so it corresponds to a SPHERICAL FRUSTUM. The surface of the spherical segment (excluding the bases) is called a ZONE.

Call the RADIUS of the SPHERE R and the height of the segment (the distance from the plane to the top of SPHERE) h. Let the RADII of the lower and upper bases be denoted a and b, respectively. Call the distance from the center to the start of the segment d, and the height from the bottom to the top of the segment h. Call the RADIUS parallel to the segment r, and the height above the center y. Then $r^2 = R^2 - y^2$,

$$\begin{aligned} V &= \int_d^{d+h} \pi r^2\,dy = \pi \int_d^{d+h} (R^2 - y^2)\,dy \\ &= \pi\left[R^2 y - \tfrac{1}{3}y^3\right]_d^{d+h} = \pi\{R^2 h - \tfrac{1}{3}[(d+h)^3 - d^3]\} \\ &= \pi[R^2 h - \tfrac{1}{3}(d^3 + 3d^2 h + 3h^2 d + h^3 - d^3)] \\ &= \pi(R^2 h - d^2 h - h^2 d - \tfrac{1}{3}h^3) \\ &= \pi h(R^2 - d^2 - hd - \tfrac{1}{3}h^2). \end{aligned} \tag{1}$$

Using

$$a^2 = R^2 - d^2 \tag{2}$$

$$b^2 = R^2 - (d+h)^2 = R^2 - d^2 - 2dh - h^2, \tag{3}$$

gives

$$a^2 + b^2 = 2R^2 - 2d^2 - 2dh - h^2 \tag{4}$$

$$R^2 - d^2 - dh = \tfrac{1}{2}(a^2 + b^2 + h^2), \tag{5}$$

so

$$\begin{aligned} V &= \pi h[\tfrac{1}{2}(a^2 + b^2 + h^2) - \tfrac{1}{3}h^2] = \pi h(\tfrac{1}{2}a^2 + \tfrac{1}{2}b^2 + \tfrac{1}{6}h^2) \\ &= \tfrac{1}{6}\pi h(3a^2 + 3b^2 + h^2). \end{aligned} \tag{6}$$

The surface area of the ZONE (which excludes the top and bottom bases) is given by

$$S = 2\pi R h. \tag{7}$$

see also ARCHIMEDES' PROBLEM, FRUSTUM, HEMISPHERE, SPHERE, SPHERICAL CAP, SPHERICAL SECTOR, SURFACE OF REVOLUTION, ZONE

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 130, 1987.

Spherical Shell

A generalization of an ANNULUS to 3-D. A spherical shell is the intersection of two concentric BALLS of differing RADII.

see also ANNULUS, BALL, CHORD, SPHERE, SPHERICAL HELIX

Spherical Spiral

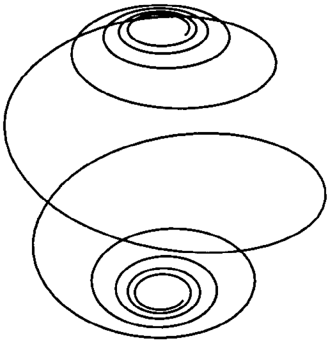

The path taken by a ship which travels from the south pole to the north pole of a SPHERE while keeping a fixed (but not RIGHT) ANGLE with respect to the meridians. The curve has an infinite number of loops since the separation of consecutive revolutions gets smaller and smaller near the poles. It is given by the parametric equations

$$\begin{aligned} x &= \cos t \cos c \\ y &= \sin t \cos c \\ z &= -\sin c, \end{aligned}$$

where

$$c \equiv \tan^{-1}(at)$$

and a is a constant.

see also MERCATOR PROJECTION, SEIFERT'S SPHERICAL SPIRAL

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 162, 1993.

Lauwerier, H. "Spherical Spiral." In *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 64–66, 1991.

Spherical Symmetry

Let **A** and **B** be constant VECTORS. Define

$$Q \equiv 3(\mathbf{A} \cdot \hat{\mathbf{r}})(\mathbf{B} \cdot \hat{\mathbf{r}}) - \mathbf{A} \cdot \mathbf{B}.$$

Then the average of Q over a spherically symmetric surface or volume is

$$\langle Q \rangle = \left\langle 3\cos^2\theta - 1 \right\rangle (\mathbf{A} \cdot \mathbf{B}) = 0,$$

since $\left\langle 3\cos^2\theta - 1 \right\rangle = 0$ over the sphere.

Spherical Tessellation

see TRIANGULAR SYMMETRY GROUP

Spherical Triangle

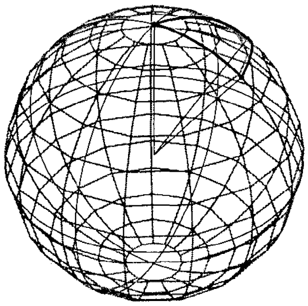

A spherical triangle is a figure formed on the surface of a sphere by three great circular arcs intersecting pairwise in three vertices. The spherical triangle is the spherical analog of the planar TRIANGLE. Let a spherical triangle have ANGLES α, β, and γ and RADIUS r. Then the AREA of the spherical triangle is

$$K = r^2[(\alpha + \beta + \gamma) - \pi].$$

The sum of the angles of a spherical triangle is between 180° and 540°. The amount by which it exceeds 180° is called the SPHERICAL EXCESS and is denoted E or Δ.

The study of angles and distances of figures on a sphere is known as SPHERICAL TRIGONOMETRY.

see also COLUNAR TRIANGLE, GIRARD'S SPHERICAL EXCESS FORMULA, L'HUILIER'S THEOREM, SPHERICAL POLYGON, SPHERICAL TRIGONOMETRY

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 79, 1972.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 131 and 147–150, 1987.

Spherical Trigonometry

Define a SPHERICAL TRIANGLE on the surface of a unit SPHERE, centered at a point O, with vertices A, B, and C. Define ANGLES $a \equiv \angle BOC$, $b \equiv \angle COA$, and $c \equiv \angle AOB$. Let the ANGLE between PLANES AOB and AOC be α, the ANGLE between PLANES BOC and AOB be β, and the ANGLE between PLANES BOC and AOC be γ. Define the VECTORS

$$\mathbf{a} \equiv \overrightarrow{OA} \tag{1}$$

$$\mathbf{b} \equiv \overrightarrow{OB} \tag{2}$$

$$\mathbf{c} \equiv \overrightarrow{OC}. \tag{3}$$

Then

$$\begin{aligned} (\hat{\mathbf{a}} \times \hat{\mathbf{b}}) \cdot (\hat{\mathbf{a}} \times \hat{\mathbf{c}}) &= (|\hat{\mathbf{a}}|\,|\hat{\mathbf{b}}| \sin c)(|\hat{\mathbf{a}}|\,|\hat{\mathbf{c}}| \sin b) \cos\alpha \\ &= \sin b \sin c \cos\alpha. \end{aligned} \tag{4}$$

Equivalently,

$$\begin{aligned} (\hat{\mathbf{a}} \times \hat{\mathbf{b}}) \cdot (\hat{\mathbf{a}} \times \hat{\mathbf{c}}) &= \hat{\mathbf{a}} \cdot [\hat{\mathbf{b}} \times (\hat{\mathbf{a}} \times \hat{\mathbf{c}})] \\ &= \hat{\mathbf{a}} \cdot [\hat{\mathbf{a}}(\hat{\mathbf{b}} \cdot \hat{\mathbf{c}}) - \hat{\mathbf{c}}(\hat{\mathbf{a}} \cdot \hat{\mathbf{b}})] \\ &= (\hat{\mathbf{b}} \cdot \hat{\mathbf{c}}) - (\hat{\mathbf{a}} \cdot \hat{\mathbf{c}})(\hat{\mathbf{a}} \cdot \hat{\mathbf{b}}) \\ &= \cos a - \cos c \cos b. \end{aligned} \tag{5}$$

Since these two expressions are equal, we obtain the identity

$$\cos a = \cos b \cos c + \sin b \sin c \cos\alpha \tag{6}$$

The identity

$$\begin{aligned} \sin\alpha &= \frac{|(\hat{\mathbf{a}} \times \hat{\mathbf{b}}) \times (\hat{\mathbf{a}} \times \hat{\mathbf{c}})|}{|\hat{\mathbf{a}} \times \hat{\mathbf{b}}||\hat{\mathbf{a}} \times \hat{\mathbf{c}}|} = -\frac{|\hat{\mathbf{a}}[\hat{\mathbf{b}}, \hat{\mathbf{a}}, \hat{\mathbf{c}}] + \hat{\mathbf{b}}[\hat{\mathbf{a}}, \hat{\mathbf{a}}, \hat{\mathbf{c}}]|}{\sin b \sin c} \\ &= \frac{[\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}]}{\sin b \sin c}, \end{aligned} \tag{7}$$

where $[\mathbf{a},\mathbf{b},\mathbf{c}]$ is the SCALAR TRIPLE PRODUCT, gives a spherical analog of the LAW OF SINES,

$$\frac{\sin\alpha}{\sin a}=\frac{\sin\beta}{\sin b}=\frac{\sin\gamma}{\sin c}=\frac{6\operatorname{Vol}(OABC)}{\sin a\sin b\sin c}, \tag{8}$$

where $\operatorname{Vol}(OABC)$ is the VOLUME of the TETRAHEDRON. From (7) and (8), it follows that

$$\sin a\cos\beta=\cos b\sin c-\sin b\cos c\cos\alpha \tag{9}$$

$$\cos a\cos\gamma=\sin a\cot b-\sin\gamma\cot\beta. \tag{10}$$

These are the fundamental equalities of spherical trigonometry.

There are also spherical analogs of the LAW OF COSINES for the sides of a spherical triangle,

$$\cos a=\cos b\cos c+\sin b\sin c\cos A \tag{11}$$

$$\cos b=\cos c\cos a+\sin c\sin a\cos B \tag{12}$$

$$\cos c=\cos a\cos b+\sin a\sin b\cos C, \tag{13}$$

and the angles of a spherical triangle,

$$\cos A=-\cos B\cos C+\sin B\sin C\cos a \tag{14}$$

$$\cos B=-\cos C\cos A+\sin C\sin A\cos b \tag{15}$$

$$\cos C=-\cos A\cos B+\sin A\sin B\cos c \tag{16}$$

(Beyer 1987), as well as the LAW OF TANGENTS

$$\frac{\tan[\frac{1}{2}(a-b)]}{\tan[\frac{1}{2}(a+b)]}=\frac{\tan[\frac{1}{2}(A-B)]}{\tan[\frac{1}{2}(A+B)]}. \tag{17}$$

Let

$$s\equiv\tfrac{1}{2}(a+b+c) \tag{18}$$

$$S\equiv\tfrac{1}{2}(A+B+C), \tag{19}$$

then the half-angle formulas are

$$\tan(\tfrac{1}{2}A)=\frac{k}{\sin(s-a)} \tag{20}$$

$$\tan(\tfrac{1}{2}B)=\frac{k}{\sin(s-b)} \tag{21}$$

$$\tan(\tfrac{1}{2}C)=\frac{k}{\sin(s-c)}, \tag{22}$$

where

$$k^2=\frac{\sin(s-a)\sin(s-b)\sin(s-c)}{\sin s}=\tan^2 r, \tag{23}$$

and the half-side formulas are

$$\tan(\tfrac{1}{2}a)=K\cos(S-A) \tag{24}$$

$$\tan(\tfrac{1}{2}b)=K\cos(S-B) \tag{25}$$

$$\tan(\tfrac{1}{2}c)=K\cos(S-C), \tag{26}$$

where

$$K^2=-\frac{\cos S}{\cos(S-A)\cos(S-B)\cos(S-C)}=\tan^2 R, \tag{27}$$

where R is the RADIUS of the SPHERE on which the spherical triangle lies.

Additional formulas include the HAVERSINE formulas

$$\operatorname{hav}a=\operatorname{hav}(b-c)+\sin b\sin c\sin(s-c) \tag{28}$$

$$\operatorname{hav}A=\frac{\sin(s-b)\sin(s-c)}{\sin b\sin c} \tag{29}$$

$$=\frac{\operatorname{hav}a-\operatorname{hav}(b-c)}{\sin b\sin c} \tag{30}$$

$$=\operatorname{hav}[\pi-(B+C)]+\sin B\sin C\operatorname{hav}a, \tag{31}$$

GAUSS'S FORMULAS

$$\frac{\sin[\frac{1}{2}(a-b)]}{\sin(\frac{1}{2}c)}=\frac{\sin[\frac{1}{2}(A-B)]}{\cos(\frac{1}{2}C)} \tag{32}$$

$$\frac{\sin[\frac{1}{2}(a+b)]}{\sin(\frac{1}{2}c)}=\frac{\cos[\frac{1}{2}(A-B)]}{\sin(\frac{1}{2}C)} \tag{33}$$

$$\frac{\cos[\frac{1}{2}(a-b)]}{\cos(\frac{1}{2}c)}=\frac{\sin[\frac{1}{2}(A+B)]}{\cos(\frac{1}{2}C)} \tag{34}$$

$$\frac{\cos[\frac{1}{2}(a+b)]}{\cos(\frac{1}{2}c)}=\frac{\cos[\frac{1}{2}(A+B)]}{\sin(\frac{1}{2}C)}, \tag{35}$$

and NAPIER'S ANALOGIES

$$\frac{\sin[\frac{1}{2}(A-B)]}{\sin[\frac{1}{2}(A+B)]}=\frac{\tan[\frac{1}{2}(a-b)]}{\tan(\frac{1}{2}c)} \tag{36}$$

$$\frac{\cos[\frac{1}{2}(A-B)]}{\cos[\frac{1}{2}(A+B)]}=\frac{\tan[\frac{1}{2}(a+b)]}{\tan(\frac{1}{2}c)} \tag{37}$$

$$\frac{\sin[\frac{1}{2}(a-b)]}{\sin[\frac{1}{2}(a+b)]}=\frac{\tan[\frac{1}{2}(A-B)]}{\cot(\frac{1}{2}C)} \tag{38}$$

$$\frac{\cos[\frac{1}{2}(a-b)]}{\cos[\frac{1}{2}(a+b)]}=\frac{\tan[\frac{1}{2}(A+B)]}{\cot(\frac{1}{2}C)} \tag{39}$$

(Beyer 1987).

see also ANGULAR DEFECT, DESCARTES TOTAL ANGULAR DEFECT, GAUSS'S FORMULAS, GIRARD'S SPHERICAL EXCESS FORMULA, LAW OF COSINES, LAW OF SINES, LAW OF TANGENTS, L'HUILIER'S THEOREM, NAPIER'S ANALOGIES, SPHERICAL EXCESS, SPHERICAL GEOMETRY, SPHERICAL POLYGON, SPHERICAL TRIANGLE

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 131 and 147–150, 1987.

Danby, J. M. *Fundamentals of Celestial Mechanics, 2nd ed., rev. ed.* Richmond, VA: Willmann-Bell, 1988.

Smart, W. M. *Text-Book on Spherical Astronomy, 6th ed.* Cambridge, England: Cambridge University Press, 1960.

Spherical Vector Harmonic

see VECTOR SPHERICAL HARMONIC

Spheroid

A spheroid is an ELLIPSOID

$$\frac{r^2\cos^2\theta\sin^2\phi}{a^2}+\frac{r^2\sin^2\theta\sin^2\phi}{b^2}+\frac{r^2\cos^2\phi}{c^2}=1 \quad (1)$$

with two SEMIMAJOR AXES equal. Orient the ELLIPSE so that the a and b axes are equal, then

$$\frac{r^2\cos^2\theta\sin^2\phi}{a^2}+\frac{r^2\sin^2\theta\sin^2\phi}{a^2}+\frac{r^2\cos^2\phi}{c^2}=1 \quad (2)$$

$$\frac{r^2\sin^2\phi}{a^2}+\frac{r^2\cos^2\phi}{c^2}=1, \quad (3)$$

where a is the equatorial RADIUS and c is the polar RADIUS. Here ϕ is the colatitude, so take $\delta\equiv\pi/2-\phi$ to express in terms of latitude.

$$\frac{r^2\cos^2\delta}{a^2}+\frac{r^2\sin^2\delta}{c^2}=1. \quad (4)$$

Rewriting $\cos^2\delta=1-\sin^2\delta$ gives

$$\frac{r^2}{a^2}+r^2\sin^2\delta\left(\frac{1}{c^2}-\frac{1}{a^2}\right)=1 \quad (5)$$

$$r^2\left(1+a^2\sin^2\delta\frac{a^2-c^2}{c^2a^2}\right)$$
$$=r^2\left(1+\sin^2\delta\frac{a^2-c^2}{c^2}\right)=a^2, \quad (6)$$

so

$$r=a\left(1+\sin^2\delta\frac{a^2-c^2}{c^2}\right)^{-1/2}. \quad (7)$$

If $a>c$, the spheroid is OBLATE. If $a<c$, the spheroid is PROLATE. If $a=c$, the spheroid degenerates to a SPHERE.

see also DARWIN-DE SITTER SPHEROID, ELLIPSOID, OBLATE SPHEROID, PROLATE SPHEROID

Spheroidal Harmonic

A spheroidal harmonic is a special case of the ELLIPSOIDAL HARMONIC which satisfies the differential equation

$$\frac{d}{dx}\left[(1-x^2)\frac{dS}{dx}\right]+\left(\lambda-c^2x^2-\frac{m^2}{1-x^2}\right)S=0$$

on the interval $-1\le x\le 1$.

see also ELLIPSOIDAL HARMONIC

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "A Worked Example: Spheroidal Harmonics." §17.4 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 764–773, 1992.

Spheroidal Wavefunction

Whittaker and Watson (1990, p. 403) define the internal and external spheroidal wavefunctions as

$$S_{mn}^{(1)}=2\pi\frac{(n-m)!}{(n+m)!}P_n^m(ir)P_n^m(\cos\theta)\begin{matrix}\cos\\ \sin\end{matrix}(m\phi)$$

$$S_{mn}^{(2)}=2\pi\frac{(n-m)!}{(n+m)!}Q_n^m(ir)Q_n^m(\cos\theta)\begin{matrix}\cos\\ \sin\end{matrix}(m\phi).$$

see also ELLIPSOIDAL HARMONIC, OBLATE SPHEROIDAL WAVE FUNCTION, PROLATE SPHEROIDAL WAVE FUNCTION, SPHERICAL HARMONIC

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Spheroidal Wave Functions." Ch. 21 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 751–759, 1972.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 642–644, 1953.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Sphinx

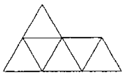

A 6-POLYIAMOND named for its resemblance to the Great Sphinx of Egypt.

References

Golomb, S. W. *Polyominoes: Puzzles, Patterns, Problems, and Packings, 2nd ed.* Princeton, NJ: Princeton University Press, p. 92, 1994.

Spider and Fly Problem

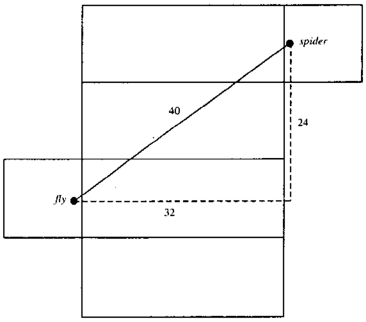

In a rectangular room (a CUBOID) with dimensions 30′ × 12′ × 12′, a spider is located in the middle of one 12′ × 12′ wall one foot away from the ceiling. A fly is in the middle of the opposite wall one foot away from the floor. If the fly remains stationary, what is the shortest distance the spider must crawl to capture the fly? The answer, 40′, can be obtained by "flattening" the walls as illustrated above.

References

Pappas, T. "The Spider & the Fly Problem." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 218 and 233, 1989.

Spider Lines

see EPITROCHOID

Spiegeldrieck

see FUHRMANN TRIANGLE

Spieker Center

The center of the SPIEKER CIRCLE. It is the CENTROID of the PERIMETER of the original TRIANGLE. The third BROCARD POINT is COLLINEAR with the Spieker center and the ISOTOMIC CONJUGATE POINT of its INCENTER.

see also BROCARD POINTS, CENTROID (TRIANGLE), INCENTER, ISOTOMIC CONJUGATE POINT, PERIMETER, SPIEKER CIRCLE, TAYLOR CENTER

References

Casey, J. *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions, with Numerous Examples, 2nd ed., rev. enl.* Dublin: Hodges, Figgis, & Co., p. 81, 1893.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 226–229 and 249, 1929.

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Spieker Circle

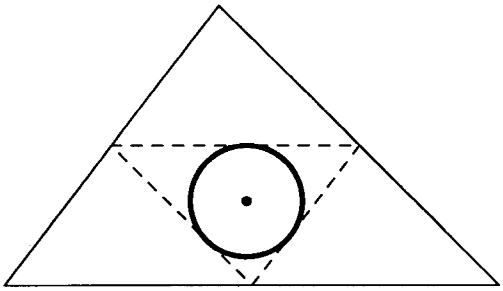

The INCIRCLE of the MEDIAL TRIANGLE. The center of the Spieker circle is called the SPIEKER CENTER.

see also INCIRCLE, MEDIAL TRIANGLE, SPIEKER CENTER

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 226–228, 1929.

Spigot Algorithm

An ALGORITHM which generates digits of a quantity one at a time without using or requiring previously computed digits. Amazingly, spigot ALGORITHMS are known for both PI and e.

Spijker's Lemma

The image on the RIEMANN SPHERE of any CIRCLE under a COMPLEX rational mapping with NUMERATOR and DENOMINATOR having degrees no more than n has length no longer than $2n\pi$.

References

Edelman, A. and Kostlan, E. "How Many Zeros of a Random Polynomial are Real?" *Bull. Amer. Math. Soc.* **32**, 1–37, 1995.

Spindle Cyclide

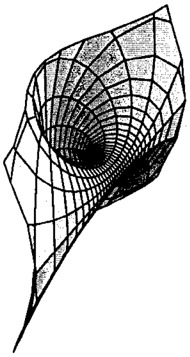

The inversion of a SPINDLE TORUS. If the inversion center lies on the torus, then the spindle cyclide degenerates to a PARABOLIC SPINDLE CYCLIDE.

see also CYCLIDE, HORN CYCLIDE, PARABOLIC CYCLIDE, RING CYCLIDE, SPINDLE TORUS, TORUS

Spindle Torus

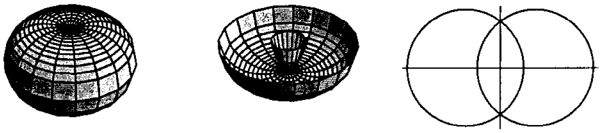

One of the three STANDARD TORI given by the parametric equations

$$\begin{aligned} x &= (c + a\cos v)\cos u \\ y &= (c + a\cos v)\sin u \\ z &= a\sin v \end{aligned}$$

with $c < a$. The exterior surface is called an APPLE and the interior surface a LEMON. The above left figure shows a spindle torus, the middle a cutaway, and the right figure shows a cross-section of the spindle torus through the xz-plane.

see also APPLE, CYCLIDE, HORN TORUS, LEMON, PARABOLIC SPINDLE CYCLIDE, RING TORUS, SPINDLE CYCLIDE, STANDARD TORI, TORUS

References

Gray, A. "Tori." §11.4 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 218–220, 1993.

Pinkall, U. "Cyclides of Dupin." §3.3 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 28–30, 1986.

Spinode

see also ACNODE, CRUNODE, CUSP, TACNODE

Spinor

A two-component COMPLEX column VECTOR. Spinors are used in physics to represent particles with half-integral spin (i.e., Fermions).

References

Lounesto, P. "Counterexamples to Theorems Published and Proved in Recent Literature on Clifford Algebras, Spinors, Spin Groups, and the Exterior Algebra." `http://www.hit.fi/~lounesto/counterexamples.htm`.

Morse, P. M. and Feshbach, H. "The Lorentz Transformation, Four-Vectors, Spinors." §1.7 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 93–107, 1953.

Spira Mirabilis

see LOGARITHMIC SPIRAL

Spiral

In general, a spiral is a curve with $\tau(s)/\kappa(s)$ equal to a constant for all s, where τ is the TORSION and κ is the CURVATURE.

see also ARCHIMEDES' SPIRAL, CIRCLE INVOLUTE, CONICAL SPIRAL, CORNU SPIRAL, COTES' SPIRAL, DAISY, EPISPIRAL, FERMAT'S SPIRAL, HYPERBOLIC SPIRAL, LOGARITHMIC SPIRAL, MICE PROBLEM, NIELSEN'S SPIRAL, PHYLLOTAXIS, POINSOT'S SPIRALS, POLYGONAL SPIRAL, SPHERICAL SPIRAL

References

Eppstein, D. "Spirals." `http://www.ics.uci.edu/~eppstein/junkyard/spiral.html`.

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 54–66, 1991.

Lockwood, E. H. "Spirals." Ch. 22 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 172–175, 1967.

Yates, R. C. "Spirals." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 206–216, 1952.

Spiral Point

A FIXED POINT for which the EIGENVALUES are COMPLEX CONJUGATES.

see also STABLE SPIRAL POINT, UNSTABLE SPIRAL POINT

References

Tabor, M. "Classification of Fixed Points." §1.4.b in *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, pp. 22–25, 1989.

Spiric Section

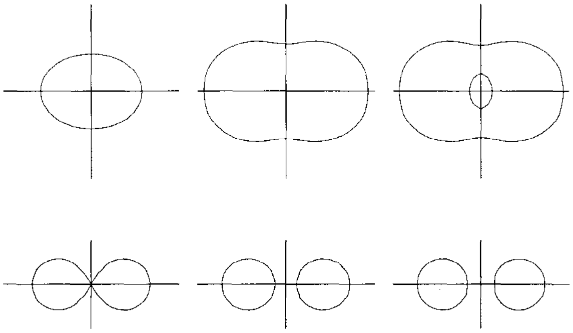

A curve with Cartesian equation

$$(r^2 - a^2 + c^2 + x^2 + y^2) = 4r^2(x^2 + c^2).$$

Around 150 BC, Menaechmus constructed CONIC SECTIONS by cutting a CONE by a PLANE. Two hundred years later, the Greek mathematician Perseus investigated the curves obtained by cutting a TORUS by a PLANE which is PARALLEL to the line through the center of the HOLE of the TORUS (MacTutor).

In the FORMULA of the curve given above, the TORUS is formed from a CIRCLE of RADIUS a whose center is rotated along a CIRCLE of RADIUS r. The value of c gives the distance of the cutting PLANE from the center of the TORUS.

When $c = 0$, the curve consists of two CIRCLES of RADIUS a whose centers are at $(r, 0)$ and $(-r, 0)$. If $c = r + a$, the curve consists of one point (the origin), while if $c > r + a$, no point lies on the curve. The above curves have $(a, b, r) = (3, 4, 2)$, (3, 1, 2) (3, 0.8, 2), (3, 1, 4), (3, 1, 4.5), and (3, 0, 4.5).

References

MacTutor History of Mathematics Archive. "Spiric Sections." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Spiric.html`.

Spirograph

A HYPOTROCHOID generated by a fixed point on a CIRCLE rolling inside a fixed CIRCLE. It has parametric equations,

$$x = (R + r)\cos\theta - (r + \rho)\cos\left(\frac{R + r}{r}\theta\right) \qquad (1)$$

$$y = (R + r)\sin\theta - (r + \rho)\sin\left(\frac{R + r}{r}\theta\right), \qquad (2)$$

where R is the radius of the fixed circle, r is the radius of the rotating circle, and ρ is the offset of the edge of the rotating circle. The figure closes only if R, r, and ρ are RATIONAL. The equations can also be written

$$x = x_0[m\cos t + a\cos(nt)] - y_0[m\sin t - a\sin(nt)] \qquad (3)$$

$$y = y_0[m\cos t + a\cos(nt)] + x_0[m\sin t - a\sin(nt)], \qquad (4)$$

where the outer wheel has radius 1, the inner wheel a radius p/q, the pen is placed a units from the center, the beginning is at θ radians above the x-axis, and

$$m \equiv \frac{q-p}{q} \qquad (5)$$

$$n \equiv \frac{q-p}{p} \qquad (6)$$

$$x_0 \equiv \cos\theta \qquad (7)$$

$$y_0 \equiv \sin\theta. \qquad (8)$$

The following curves are for $a = i/10$, with $i = 1, 2, \ldots, 10$, and $\theta = 0$.

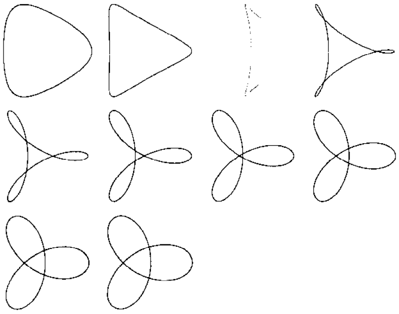

$(p,q) = (1,3)$

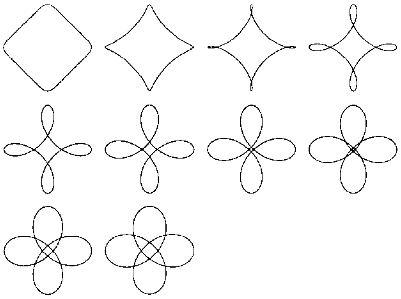

$(p,q) = (1,4)$

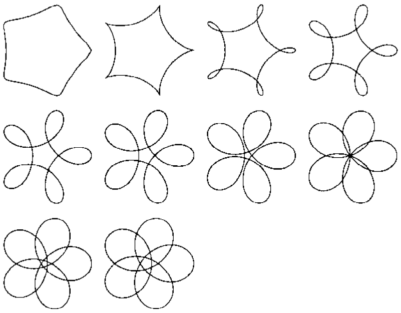

$(p,q) = (1,5)$

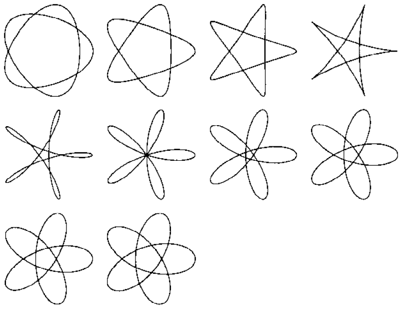

$(p,q) = (2,5)$

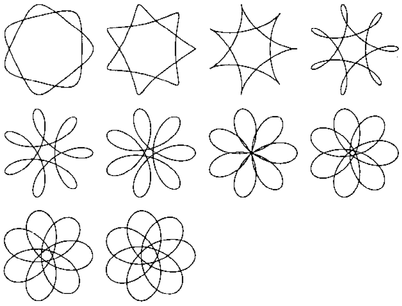

$(p,q) = (2,7)$

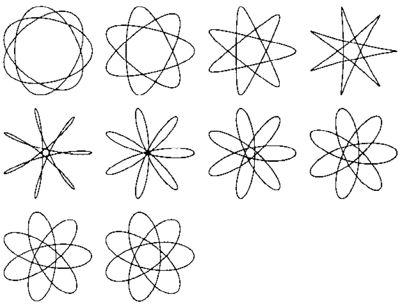

$(p,q) = (3,7)$

Additional attractive designs such as the following can also be made by superposing individual spirographs.

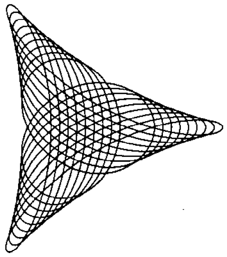

see also EPITROCHOID, HYPOTROCHOID, MAURER ROSE, SPIROLATERAL

Spirolateral

A figure formed by taking a series of steps of length 1, 2, ..., n, with an angle θ turn after each step. The symbol for a spirolateral is $^{a_1,\ldots,a_k}n_\theta$, where the a_is indicate that turns are in the $-\theta$ direction for these steps.

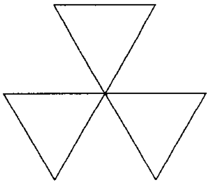

2_{60}

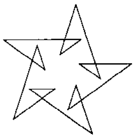

3_{36}

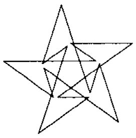

$^13_{36}$

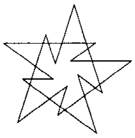

$^23_{36}$

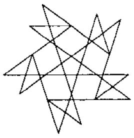

$^33_{36}$

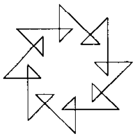

3_{45}

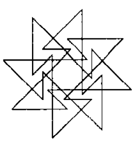

$^13_{45}$

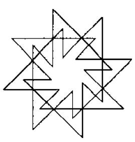

$^23_{45}$

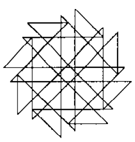

$^33_{45}$

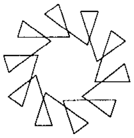

3_{72}

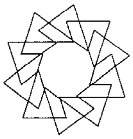

$^13_{72}$

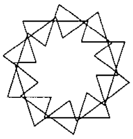

$^23_{72}$

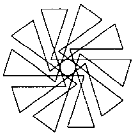

$^33_{72}$

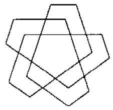

3_{108}

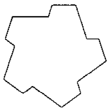

$^13_{108}$

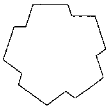

$^23_{108}$

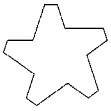

$^33_{108}$

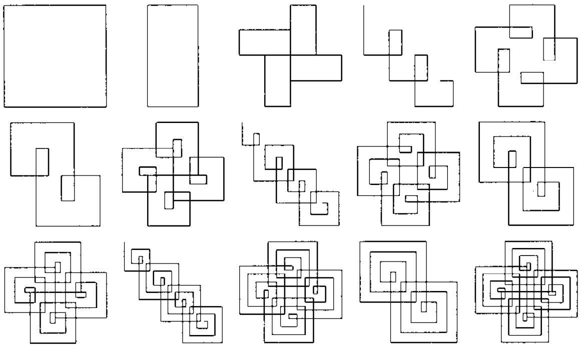

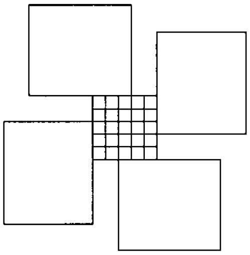

$^69_{90}$

see also MAURER ROSE, SPIROGRAPH

References

Gardner, M. "Worm Paths." Ch. 17 in *Knotted Doughnuts and Other Mathematical Entertainments.* New York: W. H. Freeman, 1986.

Odds, F. C. "Spirolaterals." *Math. Teacher* **66**, 121–124, 1973.

Spline

An interpolating POLYNOMIAL which uses information from neighboring points to obtain a degree of global smoothness.

see also B-SPLINE, BÉZIER SPLINE, CUBIC SPLINE, NURBS CURVE

References

Bartels, R. H.; Beatty, J. C.; and Barsky, B. A. *An Introduction to Splines for Use in Computer Graphics and Geometric Modelling.* San Francisco, CA: Morgan Kaufmann, 1987.

de Boor, C. *A Practical Guide to Splines.* New York: Springer-Verlag, 1978.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Interpolation and Extrapolation." Ch. 3 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 99–122, 1992.

Späth, H. *One Dimensional Spline Interpolation Algorithms.* Wellesley, MA: A. K. Peters, 1995.

Splitting

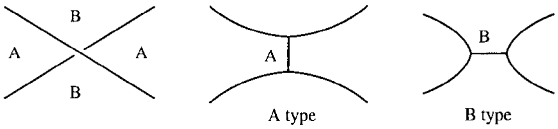

Splitting Algorithm

A method for computing a UNIT FRACTION. This method always terminates (Beeckmans 1993).

References

Beeckmans, L. "The Splitting Algorithm for Egyptian Fractions." *J. Number Th.* **43**, 173–185, 1993.

Sponge

A sponge is a solid which can be parameterized by INTEGERS p, q, and n which satisfy the equation

$$2\sin\left(\frac{\pi}{p}\right)\sin\left(\frac{\pi}{q}\right)=\cos\left(\frac{\pi}{k}\right).$$

The possible sponges are $\{p,q|k\} = \{6,6|3\}$, $\{6,4|4\}$, $\{4,6|4\}$, $\{3,6|6\}$, and $\{4,4|\infty\}$ (Ball and Coxeter 1987).

see also HONEYCOMB, MENGER SPONGE, SIERPIŃSKI SPONGE, TETRIX

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 152, 1987.

Cromwell, P. R. *Polyhedra.* New York: Cambridge University Press, p. 79, 1997.

Sporadic Group

One of the 26 finite SIMPLE GROUPS. The most complicated is the MONSTER GROUP. A summary, as given by Conway *et al.* (1985), is given below.

Sym	Name	Order	M	A
M_{11}	Mathieu	$2^4 \cdot 3^2 \cdot 5 \cdot 11$	1	1
M_{12}	Mathieu	$2^6 \cdot 3^3 \cdot 5 \cdot 11$	2	2
M_{22}	Mathieu	$2^7 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11$	12	2
M_{23}	Mathieu	$2^7 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11 \cdot 23$	1	1
M_{24}	Mathieu	$2^{10} \cdot 3^3 \cdot 5 \cdot 7 \cdot 11 \cdot 23$	1	1
$J_2 = HJ$	Janko	$2^7 \cdot 3^3 \cdot 5^2 \cdot 7$	2	2
Suz	Suzuki	$2^{13} \cdot 3^7 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13$	6	2
HS	Higman-Sims	$2^9 \cdot 3^2 \cdot 5^3 \cdot 7 \cdot 11$	2	2
McL	McLaughlin	$2^7 \cdot 3^6 \cdot 5^3 \cdot 7 \cdot 11$	3	2
Co_3	Conway	$2^{10} \cdot 3^7 \cdot 5^3 \cdot 7 \cdot 11 \cdot 23$	1	1
Co_2	Conway	$2^{18} \cdot 3^6 \cdot 5^3 \cdot 7 \cdot 11 \cdot 23$	1	1
Co_1	Conway	$2^{21} \cdot 3^9 \cdot 5^4 \cdot 7^2 \cdot 11 \cdot 13 \cdot 23$	2	1
He	Held	$2^{10} \cdot 3^3 \cdot 5^2 \cdot 7^3 \cdot 17$	1	2
Fi_{22}	Fischer	$2^{17} \cdot 3^9 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13$	6	2
Fi_{23}	Fischer	$2^{18} \cdot 3^{13} \cdot 5^2 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 23$	1	1
Fi'_{24}	Fischer	$2^{21} \cdot 3^{16} \cdot 5^2 \cdot 7^3 \cdot 11 \cdot 13 \cdot 17$ $\cdot 23 \cdot 29$	3	2
HN	Harada-Norton	$2^{14} \cdot 3^6 \cdot 5^6 \cdot 7 \cdot 11 \cdot 19$	1	2
Th	Thompson	$2^{15} \cdot 3^{10} \cdot 5^3 \cdot 7^2 \cdot 13 \cdot 19 \cdot 31$	1	1
B	Baby Monster	$2^{41} \cdot 3^{13} \cdot 5^6 \cdot 7^2 \cdot 11 \cdot 13 \cdot 17 \cdot 19$ $\cdot 23 \cdot 31 \cdot 47$	2	1
M	Monster	$2^{46} \cdot 3^{20} \cdot 5^9 \cdot 7^6 \cdot 11^2 \cdot 13^3 \cdot 17 \cdot 19$ $\cdot 23 \cdot \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71$	1	1
J_1	Janko	$2^3 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 19$	1	1
$O'N$	O'Nan	$2^9 \cdot 3^4 \cdot 7^3 \cdot 5 \cdot 11 \cdot 19 \cdot 31$	3	2
J_3	Janko	$2^7 \cdot 3^5 \cdot 5 \cdot 17 \cdot 19$	3	2
Ly	Lyons	$2^8 \cdot 3^7 \cdot 5^6 \cdot 7 \cdot 11 \cdot 31 \cdot 37 \cdot 67$	1	1
Ru	Rudvalis	$2^{14} \cdot 3^3 \cdot 5^3 \cdot 7 \cdot 13 \cdot 29$	2	1
J_4	Janko	$2^{21} \cdot 3^3 \cdot 5 \cdot 7 \cdot 11^3 \cdot 23 \cdot 29 \cdot 31$ $\cdot 37 \cdot 43$	1	1

see also BABY MONSTER GROUP, CONWAY GROUPS, FISCHER GROUPS, HARADA-NORTON GROUP, HELD GROUP, HIGMAN-SIMS GROUP, JANKO GROUPS, LYONS GROUP, MATHIEU GROUPS, MCLAUGHLIN GROUP, MONSTER GROUP, O'NAN GROUP, RUDVALIS GROUP, SUZUKI GROUP, THOMPSON GROUP

References

Aschbacher, M. *Sporadic Groups.* New York: Cambridge University Press, 1994.

Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, p. viii, 1985.

Math. Intell. Cover of volume **2**, 1980.

Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas#spo`.

Sports

see also BASEBALL, BOWLING, CHECKERS, CHESS, GO

Sprague-Grundy Function

see NIM-VALUE

Sprague-Grundy Number

see NIM-VALUE

Sprague-Grundy Value

see NIM-VALUE

Spread (Link)

see SPAN (LINK)

Spread (Tree)

A TREE having an infinite number of branches and whose nodes are sequences generated by a set of rules.

see also FAN

Spun Knot

A 3-D KNOT spun about a plane in 4-D. Unlike SUSPENDED KNOTS, spun knots are smoothly embedded at the poles.

see also SUSPENDED KNOT, TWIST-SPUN KNOT

Squarable

An object which can be constructed by SQUARING is called squarable.

Square

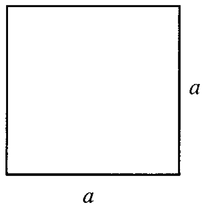

The term square is sometimes used to mean SQUARE NUMBER. When used in reference to a geometric figure, however, it means a convex QUADRILATERAL with four equal sides at RIGHT ANGLES to each other, illustrated above.

The PERIMETER of a square with side length a is

$$L = 4a \tag{1}$$

and the AREA is

$$A = a^2. \tag{2}$$

The INRADIUS r, CIRCUMRADIUS R, and AREA A can be computed directly from the formulas for a general regular POLYGON with side length a and $n = 4$ sides,

$$r = \tfrac{1}{2}a\cot\left(\frac{\pi}{4}\right) = \tfrac{1}{2}a \tag{3}$$

$$R = \tfrac{1}{2}a\csc\left(\frac{\pi}{4}\right) = \tfrac{1}{2}\sqrt{2}\,a \tag{4}$$

$$A = \tfrac{1}{4}na^2\cot\left(\frac{\pi}{4}\right) = a^2. \tag{5}$$

The length of the DIAGONAL of the UNIT SQUARE is $\sqrt{2}$, sometimes known as PYTHAGORAS'S CONSTANT.

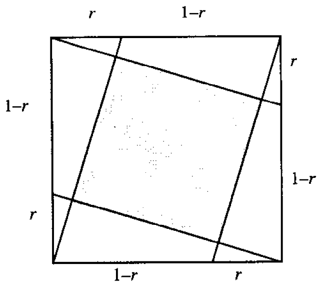

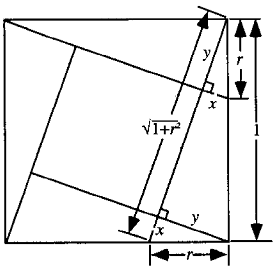

The AREA of a square inscribed inside a UNIT SQUARE as shown in the above diagram can be found as follows. Label x and y as shown, then

$$x^2 + y^2 = r^2 \tag{6}$$

$$(\sqrt{1+r^2} - x)^2 + y^2 = 1. \tag{7}$$

Plugging (6) into (7) gives

$$(\sqrt{1+r^2} - x)^2 + (r^2 - x^2) = 1. \tag{8}$$

Expanding

$$x^2 - 2x\sqrt{1+r^2} + 1 + r^2 + r^2 - x^2 = 1 \tag{9}$$

and solving for x gives

$$x = \frac{r^2}{\sqrt{1+r^2}}. \tag{10}$$

Plugging in for y yields

$$y = \sqrt{r^2 - x^2} = \frac{r}{\sqrt{1+r^2}}. \tag{11}$$

The area of the shaded square is then

$$A = (\sqrt{1+r^2} - x - y)^2 = \frac{(1-r)^2}{1+r^2} \tag{12}$$

(Detemple and Harold 1996).

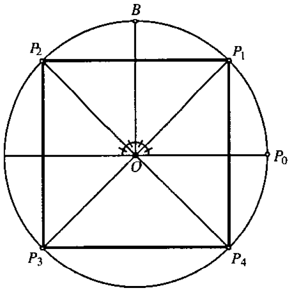

Square

The STRAIGHTEDGE and COMPASS construction of the square is simple. Draw the line OP_0 and construct a circle having OP_0 as a radius. Then construct the perpendicular OB through O. Bisect P_0OB and $P_0'OB$ to locate P_1 and P_2, where P_0' is opposite P_0. Similarly, construct P_3 and P_4 on the other SEMICIRCLE. Connecting $P_1P_2P_3P_4$ then gives a square.

As shown by Schnirelmann, a square can be INSCRIBED in any closed convex planar curve (Steinhaus 1983). A square can also be CIRCUMSCRIBED about any closed curve (Steinhaus 1983).

An infinity of points in the interior of a square are known whose distances from three of the corners of a square are RATIONAL NUMBERS. Calling the distances a, b, and c where s is the side length of the square, these solutions satisfy

$$(s^2 + b^2 - a^2)^2 + (s^2 + b^2 - c^2)^2 = (2bs)^2 \tag{13}$$

(Guy 1994). In this problem, one of a, b, c, and s is DIVISIBLE by 3, one by 4, and one by 5. It is not known if there are points having distances from *all four* corners RATIONAL, but such a solution requires the additional condition

$$a^2 + c^2 = b^2 + d^2. \tag{14}$$

In this problem, s is DIVISIBLE by 4 and a, b, c, and d are ODD. If s is not DIVISIBLE by 3 (5), then two of a, b, c, and d are DIVISIBLE by 3 (5) (Guy 1994).

see also BROWKIN'S THEOREM, DISSECTION, DOUGLAS-NEUMANN THEOREM, FINSLER-HADWIGER THEOREM, LOZENGE, PERFECT SQUARE DISSECTION, PYTHAGORAS'S CONSTANT, PYTHAGOREAN SQUARE PUZZLE, RECTANGLE, SQUARE CUTTING, SQUARE NUMBER, SQUARE PACKING, SQUARE QUADRANTS, UNIT SQUARE, VON AUBEL'S THEOREM

References

Detemple, D. and Harold, S. "A Round-Up of Square Problems." *Math. Mag.* **69**, 15–27, 1996.

Dixon, R. *Mathographics.* New York: Dover, p. 16, 1991.

Eppstein, D. "Rectilinear Geometry." `http://www.ics.uci.edu/~eppstein/junkyard/rect.html.`

Guy, R. K. "Rational Distances from the Corners of a Square." §D19 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 181–185, 1994.

Steinhaus, H. *Mathematical Snapshots, 3rd American ed.* New York: Oxford University Press, p. 104, 1983.

Square Bracket Polynomial

A POLYNOMIAL which is not necessarily an invariant of a LINK. It is related to the DICHROIC POLYNOMIAL. It is defined by the SKEIN RELATIONSHIP

$$B_{L_+} = q^{-1/2} v B_{L_0} + B_{L_\infty}, \tag{1}$$

and satisfies

$$B_{\text{unknot}} = q^{1/2} \tag{2}$$

and

$$B_{L\cup\text{unknot}} = q^{1/2} B_L. \tag{3}$$

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 235–241, 1994.

Square Cupola

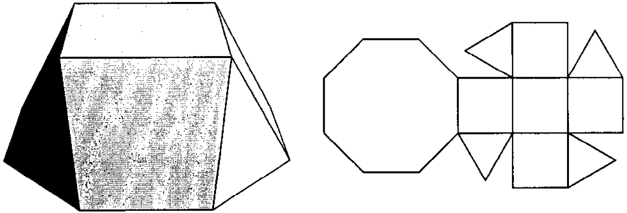

JOHNSON SOLID J_4. The bottom eight VERTICES are

$$(\pm\tfrac{1}{2}(1+\sqrt{2}),\pm\tfrac{1}{2},0),(\pm\tfrac{1}{2},\pm\tfrac{1}{2}(1+\sqrt{2}),0),$$

and the top four VERTICES are

$$\left(\pm\frac{1}{\sqrt{2}},0,\frac{1}{\sqrt{2}}\right),\left(0,\pm\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right).$$

Square Curve

see SIERPIŃSKI CURVE

Square Cutting

The average number of regions into which N lines divide a SQUARE is

$$\tfrac{1}{16}N(N-1)\pi+N+1$$

(Santaló 1976).

see also CIRCLE CUTTING

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/geom/geom.html`.

Santaló, L. A. *Integral Geometry and Geometric Probability.* Reading, MA: Addison-Wesley, 1976.

Square-Free

see SQUAREFREE

Square Gyrobicupola

see JOHNSON SOLID

Square Integrable

A function $f(x)$ is said to be square integrable if

$$\int_{-\infty}^{\infty}|f(x)|^2\,dx$$

is finite.

see also INTEGRABLE, L_2-NORM, TITCHMARSH THEOREM

References

Sansone, G. "Square Integrable Functions." §1.1 in *Orthogonal Functions, rev. English ed.* New York: Dover, pp. 1–2, 1991.

Square Knot

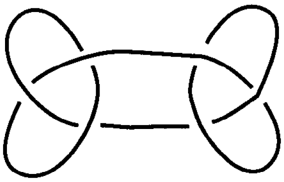

A composite KNOT of six crossings consisting of a KNOT SUM of a TREFOIL KNOT and its MIRROR IMAGE. The GRANNY KNOT has the same ALEXANDER POLYNOMIAL $(x^2-x+1)^2$ as the square knot. The square knot is also called the REEF KNOT.

see also GRANNY KNOT, MIRROR IMAGE, TREFOIL KNOT

References

Owen, P. *Knots.* Philadelphia, PA: Courage, p. 50, 1993.

Square Matrix

A MATRIX for which horizontal and vertical dimensions are the same (i.e., an $n \times n$ MATRIX).

see also MATRIX

Square Number

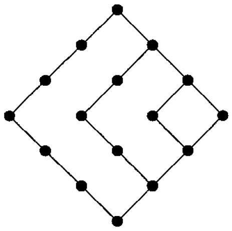

A FIGURATE NUMBER of the form $m = n^2$, where n is an INTEGER. A square number is also called a PERFECT SQUARE. The first few square numbers are 1, 4, 9, 25, 36, 49, ... (Sloane's A000290). The GENERATING FUNCTION giving the square numbers is

$$\frac{x(x+1)}{(1-x)^3}=x+4x^2+9x^3+16x^4+\ldots. \tag{1}$$

The kth nonsquare number a_k is given by

$$a_n=n+\left\lfloor\tfrac{1}{2}+\sqrt{n}\right\rfloor, \tag{2}$$

where $\lfloor x\rfloor$ is the FLOOR FUNCTION, and the first few are 2, 3, 5, 6, 7, 8, 10, 11, ... (Sloane's A000037).

The only numbers which are simultaneously square and PYRAMIDAL (the CANNONBALL PROBLEM) are $P_1 = 1$ and $P_{24} = 4900$, corresponding to $S_1 = 1$ and $S_{70} = 4900$ (Dickson 1952, p. 25; Ball and Coxeter 1987, p. 59; Ogilvy 1988), as conjectured by Lucas (1875, 1876) and proved by Watson (1918). The CANNONBALL PROBLEM is equivalent to solving the DIOPHANTINE EQUATION

$$y^2=\tfrac{1}{6}x(x+1)(2x+1) \tag{3}$$

(Guy 1994, p. 147).

The only numbers which are square and TETRAHEDRAL are $Te_1 = 1$, $Te_2 = 4$, and $Te_{48} = 19600$ (giving $S_1 = 1$, $S_2 = 4$, and $S_{140} = 19600$), as proved by Meyl (1878; cited in Dickson 1952, p. 25; Guy 1994, p. 147). In general, proving that only certain numbers are simultaneously figurate in two different ways is far from elementary.

To find the possible last digits for a square number, write $n = 10a+b$ for the number written in decimal NOTATION as ab_{10} ($a, b = 0, 1, \ldots, 9$). Then

$$n^2 = 100a^2 + 20ab + b^2, \tag{4}$$

so the last digit of n^2 is the same as the last digit of b^2. The following table gives the last digit of b^2 for $b = 0, 1, \ldots, 9$. As can be seen, the last digit can be only 0, 1, 4, 5, 6, or 9.

0	1	2	3	4	5	6	7	8	9
0	1	4	9	_6	_5	_6	_9	_4	_1

We can similarly examine the allowable last two digits by writing abc_{10} as

$$n = 100a + 10b + c, \tag{5}$$

so

$$\begin{aligned} n^2 &= (100a + 10b + c)^2 \\ &= 10^4a^2 + 2(1000ab + 100ac + 10bc) + 100b^2 + c^2 \\ &= (10^4a^2 + 2000ab + 100ac + 100b^2) + 20bc + c^2, \end{aligned} \tag{6}$$

so the last two digits are given by $20bc + c^2 = c(20b+c)$. But since the last digit must be 0, 1, 4, 5, 6, or 9, the following table exhausts all possible last two digits.

c	*b*									
	0	1	2	3	4	5	6	7	8	9
1	01	21	41	61	81	_01	_21	_41	_61	_81
4	16	96	_76	_56	_36	_16	_96	_76	_56	_36
5	25	_25	_25	_25	_25	_25	_25	_25	_25	_25
6	36	_56	_76	_96	_16	_36	_56	_76	_96	_16
9	81	_61	_41	_21	_01	_81	_61	_41	_21	_01

The only possibilities are 00, 01, 04, 09, 16, 21, 24, 25, 29, 36, 41, 44, 49, 56, 61, 64, 69, 76, 81, 84, 89, and 96, which can be summarized succinctly as 00, *e*1, *e*4, 25, *o*6, and *e*9, where *e* stands for an EVEN NUMBER and *o* for an ODD NUMBER. Additionally, unless the sum of the digits of a number is 1, 4, 7, or 9, it cannot be a square number.

The following table gives the possible residues mod n for square numbers for $n = 1$ to 20. The quantity $s(n)$ gives the number of distinct residues for a given n.

n	$s(n)$	$x^2 \pmod{n}$
2	2	0, 1
3	2	0, 1
4	2	0, 1
5	3	0, 1, 4
6	4	0, 1, 3, 4
7	4	0, 1, 2, 4
8	3	0, 1, 4
9	4	0, 1, 4, 7
10	6	0, 1, 4, 5, 6, 9
11	6	0, 1, 3, 4, 5, 9
12	4	0, 1, 4, 9
13	7	0, 1, 3, 4, 9, 10, 12
14	8	0, 1, 2, 4, 7, 8, 9, 11
15	6	0, 1, 4, 6, 9, 10
16	4	0, 1, 4, 9
17	9	0, 1, 2, 4, 8, 9, 13, 15, 16
18	8	0, 1, 4, 7, 9, 10, 13, 16
19	10	0, 1, 4, 5, 6, 7, 9, 11, 16, 17
20	6	0, 1, 4, 5, 9, 16

In general, the ODD squares are congruent to 1 (mod 8) (Conway and Guy 1996). Stangl (1996) gives an explicit formula by which the number of squares $s(n)$ in $\mathbb{Z}_n$ (i.e., mod n) can be calculated. Let p be an ODD PRIME. Then $s(n)$ is the MULTIPLICATIVE FUNCTION given by

$$s(2) = 2 \tag{7}$$

$$s(p) = \tfrac{1}{2}(p+1) \qquad (p \neq 2) \tag{8}$$

$$s(p^2) = \tfrac{1}{2}(p^2 - p + 2) \qquad (p \neq 2) \tag{9}$$

$$s(2^n) = \begin{cases} \frac{1}{3}(2^{n-1} + 4) & \text{for } n \text{ even} \\ \frac{1}{3}(2^{n-1} + 5) & \text{for } n \text{ odd} \end{cases} \tag{10}$$

$$s(p^n) = \begin{cases} \frac{p^{n+1}+p+2}{2(p+1)} & \text{for } n \geq 3 \text{ even} \\ \frac{p^{n+1}+2p+1}{2(p+1)} & \text{for } n \geq 3 \text{ odd.} \end{cases} \tag{11}$$

$s(n)$ is related to the number $q(n)$ of QUADRATIC RESIDUES in $\mathbb{Z}_n$ by

$$q(p^n) = s(p^n) - s(p^{n-2}) \tag{12}$$

for $n \geq 3$ (Stangl 1996).

For a perfect square n, $(n/p) = 0$ or 1 for all ODD PRIMES $p < n$ where (n/p) is the LEGENDRE SYMBOL. A number n which is not a perfect square but which satisfies this relationship is called a PSEUDOSQUARE.

The minimum number of squares needed to represent the numbers 1, 2, 3, ... are 1, 2, 3, 1, 2, 3, 4, 2, 1, 2, ... (Sloane's A002828), and the number of distinct ways to represent the numbers 1, 2, 3, ... in terms of squares are 1, 1, 1, 2, 2, 2, 2, 3, 4, 4, ... (Sloane's A001156). A brute-force algorithm for enumerating the square permutations of n is repeated application of the GREEDY ALGORITHM. However, this approach rapidly becomes impractical since the number of representations grows extremely rapidly with n, as shown in the following table.

n	Square Partitions
10	4
50	104
100	1116
150	6521
200	27482

Every POSITIVE integer is expressible as a SUM of (at most) $g(2) = 4$ square numbers (WARING'S PROBLEM). (Actually, the basis set is $\{0, 1, 4, 9, 16, 25, 36, 64, 81, 100, \ldots\}$, so 49 need never be used.) Furthermore, an infinite number of n require four squares to represent them, so the related quantity $G(2)$ (the least INTEGER n such that every POSITIVE INTEGER beyond a certain point requires $G(2)$ squares) is given by $G(2) = 4$.

Numbers expressible as the sum of two squares are those whose PRIME FACTORS are of the form $4k - 1$ taken to an EVEN POWER. Numbers expressible as the sum of three squares are those not of the form $4^k(8l + 7)$ for $k, l \geq 0$. The following table gives the first few numbers which require $N = 1$, 2, 3, and 4 squares to represent them as a sum.

N	Sloane	Numbers
1	000290	1, 4, 9, 16, 25, 36, 49, 64, 81, ...
2	000415	2, 5, 8, 10, 13, 17, 18, 20, 26, 29, ...
3	000419	3, 6, 11, 12, 14, 19, 21, 22, 24, 27, ...
4	004215	7, 15, 23, 28, 31, 39, 47, 55, 60, 63, ...

The FERMAT $4n + 1$ THEOREM guarantees that every PRIME of the form $4n+1$ is a sum of two SQUARE NUMBERS in only one way.

There are only 31 numbers which cannot be expressed as the sum of *distinct* squares: 2, 3, 6, 7, 8, 11, 12, 15, 18, 19, 22, 23, 24, 27, 28, 31, 32, 33, 43, 44, 47, 48, 60, 67, 72, 76, 92, 96, 108, 112, 128 (Sloane's A001422; Guy 1994). All numbers > 188 can be expressed as the sum of at most five distinct squares, and only

$$124 = 1 + 4 + 9 + 25 + 36 + 49 \tag{13}$$

and

$$188 = 1 + 4 + 9 + 25 + 49 + 100 \tag{14}$$

require six distinct squares (Bohman *et al.* 1979; Guy 1994, p. 136). In fact, 188 can also be represented using seven distinct squares:

$$188 = 1 + 4 + 9 + 25 + 36 + 49 + 64. \tag{15}$$

The following table gives the numbers which can be represented in W different ways as a sum of S squares. For example,

$$50 = 1^2 + 7^2 = 5^2 + 5^2$$

can be represented in two ways ($W = 2$) by two squares ($S = 2$).

S	W	Sloane	Numbers
1	1	000290	1, 4, 9, 16, 25, 36, 49, 64, 81, 100, ...
2	1	025284	2, 5, 8, 10, 13, 17, 18, 20, 25, 26, 29, ...
2	2	025285	50, 65, 85, 125, 130, 145, 170, 185, ...
3	1	025321	3, 6, 9, 11, 12, 14, 17, 18, 19, 21, 22, ...
3	2	025322	27, 33, 38, 41, 51, 57, 59, 62, 69, 74, ...
3	3	025323	54, 66, 81, 86, 89, 99, 101, 110, 114, ...
3	4	025324	129, 134, 146, 153, 161, 171, 189, ...
4	1	025357	4, 7, 10, 12, 13, 15, 16, 18, 19, 20, ...
4	2	025358	31, 34, 36, 37, 39, 43, 45, 47, 49, ...
4	3	025359	28, 42, 55, 60, 66, 67, 73, 75, 78, ...
4	4	025360	52, 58, 63, 70, 76, 84, 87, 91, 93, ...

The number of INTEGERS $< x$ which are squares or sums of two squares is

$$N(x) \sim kx(\ln x)^{-1/2}, \tag{16}$$

where

$$k = \sqrt{\frac{1}{2} \prod_{\substack{r=4n+3 \\ r \text{ prime}}} (1 - r^{-2})^{-1}} \tag{17}$$

(Landau 1908; Le Lionnais 1983, p. 31). The product of four distinct NONZERO INTEGERS in ARITHMETIC PROGRESSION is square only for $(-3, -1, 1, 3)$, giving $(-3)(-1)(1)(3) = 9$ (Le Lionnais 1983, p. 53). It is possible to have three squares in ARITHMETIC PROGRESSION, but not four (Dickson 1952, pp. 435–440). If these numbers are r^2, s^2, and t^2, there are POSITIVE INTEGERS p and q such that

$$r = |p^2 - 2pq - q^2| \tag{18}$$

$$s = p^2 + q^2 \tag{19}$$

$$t = p^2 + 2pq - q^2, \tag{20}$$

where $(p, q) = 1$ and one of r, s, or t is EVEN (Dickson 1952, pp. 437–438). Every three-term progression of squares can be associated with a PYTHAGOREAN TRIPLE (X, Y, Z) by

$$X = \tfrac{1}{2}(r + t) \tag{21}$$

$$Y = \tfrac{1}{2}(t - r) \tag{22}$$

$$Z = s \tag{23}$$

(Robertson 1996).

CATALAN'S CONJECTURE states that 8 and 9 (2^3 and 3^2) are the only consecutive POWERS (excluding 0 and 1), i.e., the only solution to CATALAN'S DIOPHANTINE PROBLEM. This CONJECTURE has not yet been proved or refuted, although R. Tijdeman has proved that there can be only a finite number of exceptions should the CONJECTURE not hold. It is also known that 8 and 9 are the only consecutive CUBIC and square numbers (in either order).

A square number can be the concatenation of two squares, as in the case $16 = 4^2$ and $9 = 3^2$ giving $169 = 13^2$.

It is conjectured that, other than 10^{2n}, 4×10^{2n} and 9×10^{2n}, there are only a FINITE number of squares n^2 having exactly two distinct NONZERO DIGITS (Guy 1994, p. 262). The first few such n are 4, 5, 6, 7, 8, 9, 11, 12, 15, 21, ... (Sloane's A016070), corresponding to n^2 of 16, 25, 36, 49, 64, 81, 121, ... (Sloane's A016069).

The following table gives the first few numbers which, when squared, give numbers composed of only certain digits. The only known square number composed only of the digits 7, 8, and 9 is 9. Vardi (1991) considers numbers composed only of the square digits: 1, 4, and 9.

Digits	Sloane	n, n^2
1, 2, 3	030175	1, 11, 111, 36361, 363639, ...
	030174	1, 121, 12321, 1322122321, ...
1, 4, 6	027677	1, 2, 4, 8, 12, 31, 38, 108, ...
	027676	1, 4, 16, 64, 144, 441, 1444, ...
1, 4, 9	027675	1, 2, 3, 7, 12, 21, 38, 107, ...
	006716	1, 4, 9, 49, 144, 441, 1444, 11449, ...
2, 4, 8	027679	2, 22, 168, 478, 2878, 210912978, ...
	027678	4, 484, 28224, 228484, 8282884, ...
4, 5, 6	030177	2, 8, 216, 238, 258, 738, 6742, ...
	030176	4, 64, 46656, 56644, 66564, ...

BROWN NUMBERS are pairs (m, n) of INTEGERS satisfying the condition of BROCARD'S PROBLEM, i.e., such that

$$n! + 1 = m^2, \tag{24}$$

where $n!$ is a FACTORIAL. Only three such numbers are known: (5,4), (11,5), (71,7). Erdős conjectured that these are the only three such pairs.

Either $5x^2 + 4 = y^2$ or $5x^2 - 4 = y^2$ has a solution in POSITIVE INTEGERS IFF, for some n, $(x, y) = (F_n, L_n)$, where F_n is a FIBONACCI NUMBER and L_n is a LUCAS NUMBER (Honsberger 1985, pp. 114–118).

The smallest and largest square numbers containing the digits 1 to 9 are

$$11,826^2 = 139,854,276, \tag{25}$$

$$30,384^2 = 923,187,456. \tag{26}$$

The smallest and largest square numbers containing the digits 0 to 9 are

$$32,043^2 = 1,026,753,849, \tag{27}$$

$$99,066^2 = 9,814,072,356 \tag{28}$$

(Madachy 1979, p. 159). The smallest and largest square numbers containing the digits 1 to 9 twice each are

$$335,180,136^2 = 112,345,723,568,978,496 \tag{29}$$

$$999,390,432^2 = 998,781,235,573,146,624, \tag{30}$$

and the smallest and largest containing 1 to 9 three times are

$$\begin{aligned} &10,546,200,195,312^2 \\ &\quad = 111,222,338,559,598,866,946,777,344 \quad (31) \\ &31,621,017,808,182^2 \\ &\quad = 999,888,767,225,363,175,346,145,124 \quad (32) \end{aligned}$$

(Madachy 1979, p. 159).

Madachy (1979, p. 165) also considers number which are equal to the sum of the squares of their two "halves" such as

$$1233 = 12^2 + 33^2 \tag{33}$$

$$8833 = 88^2 + 33^2 \tag{34}$$

$$10100 = 10^2 + 100^2 \tag{35}$$

$$5882353 = 588^2 + 2353^2, \tag{36}$$

in addition to a number of others.

see also ANTISQUARE NUMBER, BIQUADRATIC NUMBER, BROCARD'S PROBLEM, BROWN NUMBERS, CANNONBALL PROBLEM, CATALAN'S CONJECTURE, CENTERED SQUARE NUMBER, CLARK'S TRIANGLE, CUBIC NUMBER, DIOPHANTINE EQUATION, FERMAT $4n + 1$ THEOREM, GREEDY ALGORITHM, GROSS, LAGRANGE'S FOUR-SQUARE THEOREM, LANDAU-RAMANUJAN CONSTANT, PSEUDOSQUARE, PYRAMIDAL NUMBER, $r_k(n)$, SQUAREFREE, SQUARE TRIANGULAR NUMBER, WARING'S PROBLEM

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 59, 1987.

Bohman, J.; Fröberg, C.-E.; and Riesel, H. "Partitions in Squares." *BIT* **19**, 297–301, 1979.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 30–32, 1996.

Dickson, L. E. *History of the Theory of Numbers, Vol. 2: Diophantine Analysis.* New York: Chelsea, 1952.

Grosswald, E. *Representations of Integers as Sums of Squares.* New York: Springer-Verlag, 1985.

Guy, R. K. "Sums of Squares" and "Squares with Just Two Different Decimal Digits." §C20 and F24 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 136–138 and 262, 1994.

Honsberger, R. "A Second Look at the Fibonacci and Lucas Numbers." Ch. 8 in *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., 1985.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, 1983.

Lucas, É. Question 1180. *Nouv. Ann. Math. Ser. 2* **14**, 336, 1875.

Lucas, É. Solution de Question 1180. *Nouv. Ann. Math. Ser. 2* **15**, 429–432, 1876.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 159 and 165, 1979.

Meyl, A.-J.-J. Solution de Question 1194. *Nouv. Ann. Math.* **17**, 464–467, 1878.

Ogilvy, C. S. and Anderson, J. T. *Excursions in Number Theory.* New York: Dover, pp. 77 and 152, 1988.

Pappas, T. "Triangular, Square & Pentagonal Numbers." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 214, 1989.
Pietenpol, J. L. "Square Triangular Numbers." *Amer. Math. Monthly* **69**, 168–169, 1962.
Robertson, J. P. "Magic Squares of Squares." *Math. Mag.* **69**, 289-293, 1996.
Stangl, W. D. "Counting Squares in $\mathbb{Z}_n$." *Math. Mag.* **69**, 285–289, 1996.
Taussky-Todd, O. "Sums of Squares." *Amer. Math. Monthly* **77**, 805–830, 1970.
Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 20 and 234–237, 1991.
Watson, G. N. "The Problem of the Square Pyramid." *Messenger. Math.* **48**, 1–22, 1918.

Square Orthobicupola

see JOHNSON SOLID

Square Packing

Find the minimum size SQUARE capable of bounding n equal SQUARES arranged in any configuration. The only packings which have been proven optimal are 2, 3, 5, and SQUARE NUMBERS (4, 9, ...). If $n = a^2 - a$ for some a, it is CONJECTURED that the size of the minimum bounding square is a for small n. The smallest n for which the CONJECTURE is known to be violated is 1560. The size is known to scale as k^b, where

$$\tfrac{1}{2}(3-\sqrt{3}) < b < \tfrac{1}{2}.$$

n	Exact	Decimal
1	1	1
2	2	2
3	2	2
4	2	2
5	$2+\frac{1}{2}\sqrt{2}$	2.707...
6	3	3
7	3	3
8	3	3
9	3	3
10	$3+\frac{1}{2}\sqrt{2}$	3.707...
11		3.877...
12	4	4
13	4	4
14	4	4
15	4	4
16	4	4
17	$4+\frac{1}{2}\sqrt{2}$	4.707...
18	$2(7+\sqrt{7})$	4.822...
19	$3+\frac{4}{3}\sqrt{2}$	4.885...
20	5	5
21	5	5
22	5	5
23	5	5
24	5	5
25	5	5
26		5.650...

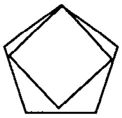

The best packing of a SQUARE inside a PENTAGON, illustrated above, is 1.0673....

References

Erdős, P. and Graham, R. L. "On Packing Squares with Equal Squares." *J. Combin. Th. Ser. A* **19**, 119–123, 1975.
Friedman, E. "Packing Unit Squares in Squares." *Elec. J. Combin.* DS7, 1–24, Mar. 5, 1998. `http://www.combinatorics.org/Surveys/`.
Gardner, M. "Packing Squares." Ch. 20 in *Fractal Music, Hypercards, and More Mathematical Recreations from Scientific American Magazine.* New York: W. H. Freeman, 1992.
Göbel, F. "Geometrical Packing and Covering Problems." In *Packing and Covering in Combinatorics* (Ed. A. Schrijver). Amsterdam: Tweede Boerhaavestraat, 1979.
Roth, L. F. and Vaughan, R. C. "Inefficiency in Packing Squares with Unit Squares." *J. Combin. Th. Ser. A* **24**, 170–186, 1978.

Square Polyomino

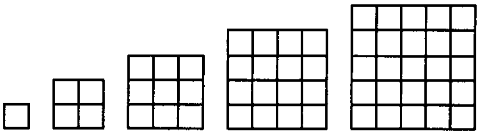

see also L-POLYOMINO, SKEW POLYOMINO, STRAIGHT POLYOMINO, T-POLYOMINO

Square Pyramid

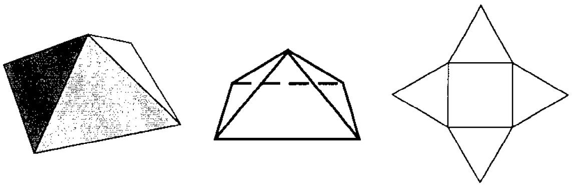

A square pyramid is a PYRAMID with a SQUARE base. If the top of the pyramid is cut off by a PLANE, a square PYRAMIDAL FRUSTUM is obtained. If the four TRIANGLES of the square pyramid are EQUILATERAL, the square pyramid is the "regular" POLYHEDRON known as JOHNSON SOLID J_1 and, for side length a, has height

$$h = \tfrac{1}{2}\sqrt{2}\,a. \tag{1}$$

Using the equation for a general PYRAMID, the VOLUME of the "regular" is therefore

$$V = \tfrac{1}{3}hA_b = \tfrac{1}{6}\sqrt{2}\,a^3. \tag{2}$$

If the apex of the pyramid does not lie atop the center of the base, then the SLANT HEIGHT is given by

$$s = \sqrt{h^2 + \tfrac{1}{2}a^2}, \tag{3}$$

where h is the height and a is the length of a side of the base.

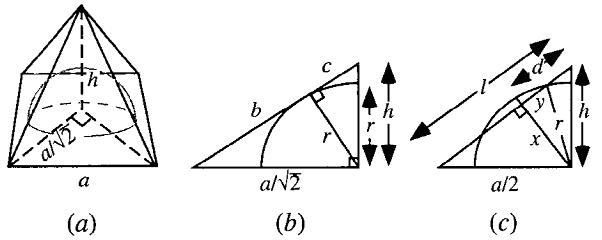

Consider a HEMISPHERE placed on the base of a square pyramid (having side lengths a and height h). Further, let the hemisphere be tangent to the four apex edges. Then what is the volume of the HEMISPHERE which is interior the pyramid (Cipra 1993)?

From Fig. (a), the CIRCUMRADIUS of the base is $a/\sqrt{2}$. Now find h in terms of r and a. Fig. (b) shows a CROSS-SECTION cut by the plane through the pyramid's apex, one of the base's vertices, and the base center. This figure gives

$$b = \sqrt{\tfrac12 a^2 - r^2} \tag{4}$$

$$c = \sqrt{h^2 - r^2}, \tag{5}$$

so the SLANT HEIGHT is

$$s = \sqrt{h^2 + \tfrac12 a^2} = b + c = \sqrt{\tfrac12 a^2 - r^2} + \sqrt{h^2 - r^2}. \tag{6}$$

Solving for h gives

$$h = \frac{ra}{\sqrt{a^2 - 2r^2}}. \tag{7}$$

We know, however, that the HEMISPHERE must be tangent to the sides, so $r = a/2$, and

$$h = \frac{\frac12 a}{\sqrt{a^2 - \frac12 a^2}} a = \frac{\frac12}{\sqrt{\frac12}} a = \tfrac12\sqrt{2}\, a. \tag{8}$$

Fig. (c) shows a CROSS-SECTION through the center, apex, and midpoints of opposite sides. The PYTHAGOREAN THEOREM once again gives

$$l = \sqrt{\tfrac14 a^2 + h^2} = \sqrt{\tfrac14 a^2 + \tfrac12 a^2} = \tfrac12\sqrt{3}\, a. \tag{9}$$

We now need to find x and y.

$$\sqrt{\tfrac14 a^2 - x^2} + d = l. \tag{10}$$

But we know l and h, and d is given by

$$d = \sqrt{h^2 - x^2}, \tag{11}$$

so

$$\sqrt{\tfrac14 a^2 - x^2} + \sqrt{\tfrac12 a^2 - x^2} = \tfrac12\sqrt{3}\, a. \tag{12}$$

Solving gives

$$x = \tfrac16\sqrt{6}\, a, \tag{13}$$

so

$$y = \sqrt{r^2 - x^2} = \sqrt{\tfrac14 - \tfrac16}\, a = \sqrt{\frac{3-2}{12}}\, a = \frac{a}{2\sqrt{3}}. \tag{14}$$

We can now find the AREA of the SPHERICAL CAP as

$$V_{\text{cap}} = \tfrac16 \pi H(3A^2 + H^2), \tag{15}$$

where

$$A \equiv y = \frac{a}{2\sqrt{3}} \tag{16}$$

$$H \equiv r - x = \tfrac12 a - \frac{a}{\sqrt{6}} = a\left(\frac12 - \frac{1}{\sqrt{6}}\right), \tag{17}$$

so

$$\begin{aligned} V_{\text{cap}} &= \tfrac16 \pi a^3 \left[3\left(\frac{1}{12}\right) + \left(\frac12 - \frac{1}{\sqrt{6}}\right)^2\right]\left(\frac12 - \frac{1}{\sqrt{6}}\right) \\ &= \tfrac16 \pi a^3 \left[\frac14 + \left(\frac14 + \frac16 - \frac{1}{\sqrt{6}}\right)\right]\left(\frac12 - \frac{1}{\sqrt{6}}\right) \\ &= \tfrac16 \pi a^3 \left(\frac23 - \frac{1}{\sqrt{6}}\right)\left(\frac12 - \frac{1}{\sqrt{6}}\right) \\ &= \tfrac16 \pi a^3 \left(\frac13 - \frac{1}{2\sqrt{6}} - \frac{2}{3\sqrt{6}} + \frac16\right) \\ &= \tfrac16 \pi a^3 \left(\frac12 - \frac{7}{6\sqrt{6}}\right). \end{aligned} \tag{18}$$

Therefore, the volume within the pyramid is

$$\begin{aligned} V_{\text{inside}} &= \tfrac23 \pi r^3 - 4V_{\text{cap}} = \tfrac23 \pi \tfrac18 a^3 - \tfrac23 \pi a^3 \left(\frac12 - \frac{7}{6\sqrt{6}}\right) \\ &= \tfrac23 \pi a^3 \left(\frac18 - \frac12 + \frac{7}{6\sqrt{6}}\right) = \tfrac23 \pi a^3 \left(\frac{7}{6\sqrt{6}} - \frac38\right) \\ &= \pi a^3 \left(\frac{7}{9\sqrt{6}} - \frac14\right). \end{aligned} \tag{19}$$

This problem appeared in the Japanese scholastic aptitude test (Cipra 1993).

see also SQUARE PYRAMIDAL NUMBER

References

Cipra, B. "An Awesome Look at Japan Math SAT." *Science* **259**, 22, 1993.

Square Pyramidal Number

A FIGURATE NUMBER of the form

$$P_n = \tfrac{1}{6}n(n+1)(2n+1), \tag{1}$$

corresponding to a configuration of points which form a SQUARE PYRAMID, is called a square pyramidal number (or sometimes, simply a PYRAMIDAL NUMBER). The first few are 1, 5, 14, 30, 55, 91, 140, 204, ... (Sloane's A000330). They are sums of consecutive pairs of TETRAHEDRAL NUMBERS and satisfy

$$P_n = \tfrac{1}{3}(2n+1)T_n, \tag{2}$$

where T_n is the nth TRIANGULAR NUMBER.

The only numbers which are simultaneously SQUARE and pyramidal (the CANNONBALL PROBLEM) are $P_1 = 1$ and $P_{24} = 4900$, corresponding to $S_1 = 1$ and $S_{70} = 4900$ (Dickson 1952, p. 25; Ball and Coxeter 1987, p. 59; Ogilvy 1988), as conjectured by Lucas (1875, 1876) and proved by Watson (1918). The proof is far from elementary, and is equivalent to solving the DIOPHANTINE EQUATION

$$y^2 = \tfrac{1}{6}x(x+1)(2x+1) \tag{3}$$

(Guy 1994, p. 147). However, an elementary proof has also been given by a number of authors.

Numbers which are simultaneously TRIANGULAR and square pyramidal satisfy the DIOPHANTINE EQUATION

$$3(2y+1)^2 = 8x^3 + 12x^2 + 4x + 3. \tag{4}$$

The only solutions are $x = -1$, 0, 1, 5, 6, and 85 (Guy 1994, p. 147). Beukers (1988) has studied the problem of finding numbers which are simultaneously TETRAHEDRAL and square pyramidal via INTEGER points on an ELLIPTIC CURVE. He finds that the only solution is the trivial $Te_1 = P_1 = 1$.

see also TETRAHEDRAL NUMBER

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 59, 1987.

Beukers, F. "On Oranges and Integral Points on Certain Plane Cubic Curves." *Nieuw Arch. Wisk.* **6**, 203–210, 1988.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 47–50, 1996.

Dickson, L. E. *History of the Theory of Numbers, Vol. 2: Diophantine Analysis.* New York: Chelsea, 1952.

Guy, R. K. "Figurate Numbers." §D3 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 147–150, 1994.

Lucas, É. Question 1180. *Nouvelles Ann. Math. Ser. 2* **14**, 336, 1875.

Lucas, É. Solution de Question 1180. *Nouvelles Ann. Math. Ser. 2* **15**, 429–432, 1876.

Ogilvy, C. S. and Anderson, J. T. *Excursions in Number Theory.* New York: Dover, pp. 77 and 152, 1988.

Sloane, N. J. A. Sequence A000330/M3844 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Watson, G. N. "The Problem of the Square Pyramid." *Messenger. Math.* **48**, 1–22, 1918.

Square Quadrants

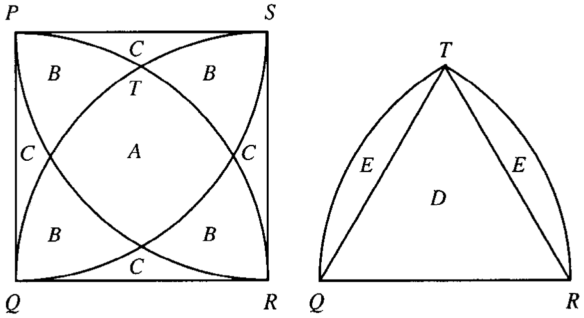

The areas of the regions illustrated above can be found from the equations

$$A + 4B + 4C = 1 \tag{1}$$

$$A + 3B + 2C = \tfrac{1}{4}\pi. \tag{2}$$

Since we want to solve for three variables, we need a third equation. This can be taken as

$$A + 2B + C = 2E + D, \tag{3}$$

where

$$D = \tfrac{1}{4}\sqrt{3} \tag{4}$$

$$D + E = \tfrac{1}{6}\pi, \tag{5}$$

leading to

$$A+2B+C = D+2E = 2(D+E)-D = \tfrac{1}{3}\pi - \tfrac{1}{4}\sqrt{3}. \tag{6}$$

Combining the equations (1), (2), and (6) gives the matrix equation

$$\begin{bmatrix} 1 & 4 & 4 \\ 1 & 3 & 2 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{1}{4}\pi \\ \frac{1}{3}\pi - \frac{1}{4}\sqrt{3} \end{bmatrix}, \tag{7}$$

which can be inverted to yield

$$A = 1 - \sqrt{3} - \tfrac{1}{3}\pi \tag{8}$$

$$B = -1 + \tfrac{1}{2}\sqrt{3} + \tfrac{1}{12}\pi \tag{9}$$

$$C = 1 - \tfrac{1}{4}\sqrt{3} + \tfrac{1}{6}\pi. \tag{10}$$

References

Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., pp. 67–69, 1991.

Square Root

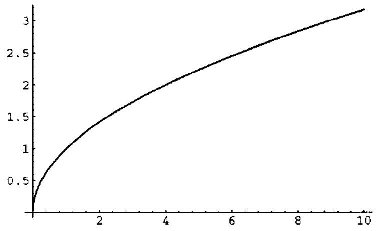

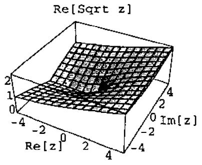

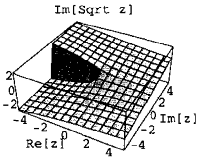

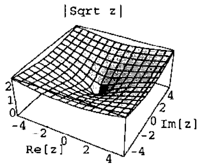

A square root of x is a number r such that $r^2 = x$. This is written $r = x^{1/2}$ (x to the 1/2 POWER) or $r = \sqrt{x}$. The square root function $f(x) = \sqrt{x}$ is the INVERSE FUNCTION of $f(x) = x^2$. Square roots are also called RADICALS or SURDS. A general COMPLEX NUMBER z has *two* square roots. For example, for the real POSITIVE number $x = 9$, the two square roots are $\sqrt{9} = \pm 3$, since $3^2 = (-3)^2 = 9$. Similarly, for the real NEGATIVE number $x = -9$, the two square roots are $\sqrt{-9} = \pm 3i$, where i is the IMAGINARY NUMBER defined by $i^2 = -1$. In common usage, unless otherwise specified, "the" square root is generally taken to mean the POSITIVE square root.

The square root of 2 is the IRRATIONAL NUMBER $\sqrt{2} \approx 1.41421356$ (Sloane's A002193), which has the simple periodic CONTINUED FRACTION 1, 2, 2, 2, 2, 2, ... The square root of 3 is the IRRATIONAL NUMBER $\sqrt{3} \approx 1.73205081$ (Sloane's A002194), which has the simple periodic CONTINUED FRACTION 1, 1, 2, 1, 2, 1, 2, ... In general, the CONTINUED FRACTIONS of the square roots of all POSITIVE integers are periodic.

The square roots of a COMPLEX NUMBER are given by

$$\sqrt{x+iy} = \pm\sqrt{x^2+y^2}\left\{\cos\left[\frac{1}{2}\tan^{-1}\left(\frac{y}{x}\right)\right] + i\sin\left[\frac{1}{2}\tan^{-1}\left(\frac{y}{x}\right)\right]\right\}. \quad (1)$$

As can be seen in the above figure, the IMAGINARY PART of the complex square root function has a BRANCH CUT along the NEGATIVE real axis.

A NESTED RADICAL of the form $\sqrt{a \pm b\sqrt{c}}$ can sometimes be simplified into a simple square root by equating

$$\sqrt{a \pm b\sqrt{c}} = \sqrt{d} \pm \sqrt{e}. \quad (2)$$

Squaring gives

$$a \pm b\sqrt{c} = d + e \pm 2\sqrt{de}, \quad (3)$$

so

$$a = d + e \quad (4)$$

$$b^2 c = 4de. \quad (5)$$

Solving for d and e gives

$$d, e = \frac{a \pm \sqrt{a^2 - b^2 c}}{2}. \quad (6)$$

A sequence of approximations a/b to $\sqrt{n}$ can be derived by factoring

$$a^2 - nb^2 = \pm 1 \quad (7)$$

(where -1 is possible only if -1 is a QUADRATIC RESIDUE of n). Then

$$(a + b\sqrt{n})(a - b\sqrt{n}) = \pm 1 \quad (8)$$

$$(a + b\sqrt{n})^k (a - b\sqrt{n})^k = (\pm 1)^k = \pm 1, \quad (9)$$

and

$$(1 + \sqrt{n})^1 = 1 + \sqrt{n} \quad (10)$$

$$(1 + \sqrt{n})^2 = (1 + n) + 2\sqrt{n} \quad (11)$$

$$(1 + \sqrt{n})(a + b\sqrt{n}) = (a + bn) + \sqrt{n}(a + b). \quad (12)$$

Therefore, a and b are given by the RECURRENCE RELATIONS

$$a_i = a_{i-1} + b_{i-1} n \quad (13)$$

$$b_i = a_{i-1} + b_{i-1} \quad (14)$$

with $a_1 = b_1 = 1$. The error obtained using this method is

$$\left|\frac{a}{b} - \sqrt{n}\right| = \frac{1}{b(a + b\sqrt{n})} < \frac{1}{2b^2}. \quad (15)$$

The first few approximants to $\sqrt{n}$ are therefore given by

$$1, \tfrac{1}{2}(1+n), \frac{1+3n}{3+n}, \frac{1+6n+n^2}{4(n+1)}, \frac{1+10n+5n^2}{5+10n+n^2}, \ldots. \quad (16)$$

This ALGORITHM is sometimes known as the BHASKARA-BROUCKNER ALGORITHM. For the case $n = 2$, this gives the convergents to $\sqrt{2}$ as 1, 3/2, 7/5, 17/12, 41/29, 99/70, ...

Another general technique for deriving this sequence, known as NEWTON'S ITERATION, is obtained by letting $x = \sqrt{n}$. Then $x = n/x$, so the SEQUENCE

$$x_k = \frac{1}{2}\left(x_{k-1} + \frac{n}{x_{k-1}}\right) \quad (17)$$

converges quadratically to the root. The first few approximants to $\sqrt{n}$ are therefore given by

$$1, \tfrac{1}{2}(1+n), \frac{1+6n+n^2}{4(n+1)}, \frac{1+26n+70n^2+28n^3+n^4}{8(1+n)(1+6n+n^2)}, \ldots. \quad (18)$$

For $\sqrt{2}$, this gives the convergents 1, 3/2, 17/12, 577/408, 665857/470832, ...

see also CONTINUED SQUARE ROOT, CUBE ROOT, NESTED RADICAL, NEWTON'S ITERATION, QUADRATIC SURD, ROOT OF UNITY, SQUARE NUMBER, SQUARE TRIANGULAR NUMBER, SURD

References

Sloane, N. J. A. Sequences A002193/M3195 and A002194/M4326 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Spanier, J. and Oldham, K. B. "The Square-Root Function $\sqrt{bx+c}$ and Its Reciprocal," "The $b\sqrt{a^2-x^2}$ Function and Its Reciprocal," and "The $b\sqrt{x^2+a}$ Function." Chs. 12, 14, and 15 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 91–99, 107–115, and 115–122, 1987.

Williams, H. C. "A Numerical Investigation into the Length of the Period of the Continued Fraction Expansion of $\sqrt{D}$." *Math. Comp.* **36**, 593–601, 1981.

Square Root Inequality

$$2\sqrt{n+1}-2\sqrt{n}<\frac{1}{\sqrt{n}}<2\sqrt{n}-2\sqrt{n-1}\,.$$

Square Root Method

The square root method is an algorithm which solves the MATRIX EQUATION

$$\mathsf{A}\mathbf{u}=\mathbf{g} \tag{1}$$

for $\mathbf{u}$, with A a $p\times p$ SYMMETRIC MATRIX and $\mathbf{g}$ a given VECTOR. Convert A to a TRIANGULAR MATRIX such that

$$\mathsf{T}^{\mathrm{T}}\mathsf{T}=\mathsf{A}, \tag{2}$$

where T^{T} is the MATRIX TRANSPOSE. Then

$$\mathsf{T}^{\mathrm{T}}\mathbf{k}=\mathbf{g} \tag{3}$$

$$\mathsf{T}\mathbf{u}=\mathbf{k}, \tag{4}$$

so

$$\mathsf{T}=\begin{bmatrix} s_{11} & s_{12} & \cdots & \cdots \\ 0 & s_{22} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_{pp} \end{bmatrix}, \tag{5}$$

giving the equations

$$\begin{aligned} {s_{11}}^2 &= a_{11} \\ s_{11}s_{12} &= a_{12} \\ {s_{12}}^2+{s_{22}}^2 &= a_{22} \\ {s_{1j}}^2+{s_{2j}}^2+\ldots+{s_{jj}}^2 &= a_{jj} \\ s_{1j}+s_{2j}s_{2k}+\ldots+s_{jj}s_{jk} &= a_{jk}. \end{aligned} \tag{6}$$

These give

$$\begin{aligned} s_{11} &= \sqrt{a_{11}} \\ s_{12} &= \frac{a_{12}}{s_{11}} \\ s_{22} &= \sqrt{a_{22}-{s_{12}}^2} \\ s_{jj} &= \sqrt{a_{jj}-{s_{ij}}^2-{s_{2j}}^2-\ldots-{s_{j-1,j}}^2} \\ s_{jk} &= \frac{a_{jk}-s_{1j}s_{1k}-s_{2j}s_{2k}-\ldots-s_{j-1,j}s_{j-1,k}}{s_{jj}}, \end{aligned} \tag{7}$$

giving T from A. Now solve for $\mathbf{k}$ in terms of the s_{ij}s and $\mathbf{g}$,

$$\begin{aligned} s_{11}k_1 &= g_1 \\ s_{12}k_1+s_{22}k_2 &= g_2 \\ s_{1j}k_1+s_{2j}k_2+\ldots+s_{jj}k_j &= g_j, \end{aligned} \tag{8}$$

which gives

$$\begin{aligned} k_1 &= \frac{g_1}{s_{11}} \\ k_2 &= \frac{g_2-s_{12}k_1}{s_{22}} \\ k_j &= \frac{g_j-s_{1j}k_1-s_{2j}k_2-\ldots-s_{j-1,j}k_{j-1}}{s_{jj}}. \end{aligned} \tag{9}$$

Finally, find $\mathbf{u}$ from the s_{ij}s and $\mathbf{k}$,

$$\begin{aligned} s_{11}u_1+s_{12}u_2\ldots+s_{1p}u_p &= k_1 \\ s_{22}u_2+\ldots+s_{2p}u_p &= k_2 \\ s_{pp}u_p &= k_p, \end{aligned} \tag{10}$$

giving the desired solution,

$$\begin{aligned} u_p &= \frac{k_p}{s_{pp}} \\ u_{p-1} &= \frac{k_{p-1}-s_{p-1,p}u_p}{s_{p-1,p-1}} \\ u_j &= \frac{k_j-s_{j,j+1}u_{j+1}-s_{j,j+2}u_{j+2}-\ldots-s_{jp}u_p}{s_{jj}}. \end{aligned} \tag{11}$$

see also LU DECOMPOSITION

References

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, pp. 298–300, 1951.

Square Triangular Number

A number which is simultaneously SQUARE and TRIANGULAR. The first few are 1, 36, 1225, 41616, 1413721, 48024900, ... (Sloane's A001110), corresponding to $T_1=S_1$, $T_8=S_6$, $T_{49}=S_{35}$, $T_{288}=S_{204}$, $T_{1681}=S_{1189}$, ... (Pietenpol 1962), but there are an infinite number, as first shown by Euler in 1730 (Dickson 1952).

The general FORMULA for a square triangular number ST_n is b^2c^2, where b/c is the nth convergent to the CONTINUED FRACTION of $\sqrt{2}$ (Ball and Coxeter 1987, p. 59; Conway and Guy 1996). The first few are

$$\frac{1}{1},\frac{3}{2},\frac{7}{5},\frac{17}{12},\frac{41}{29},\frac{99}{70},\frac{239}{169},\ldots. \tag{1}$$

The NUMERATORS and DENOMINATORS give solutions to the PELL EQUATION

$$x^2-2y^2=\pm1, \tag{2}$$

but can also be obtained by doubling the previous FRACTION and adding to the FRACTION before that. The connection with the PELL EQUATION can be seen by letting N denote the Nth TRIANGULAR NUMBER and M the Mth SQUARE NUMBER, then

$$\tfrac{1}{2}N(N+1) = M^2. \tag{3}$$

Defining

$$x \equiv 2N+1 \tag{4}$$
$$y \equiv 2M \tag{5}$$

then gives the equation

$$x^2 - 2y^2 = 1 \tag{6}$$

(Conway and Guy 1996). Numbers which are simultaneously TRIANGULAR and SQUARE PYRAMIDAL also satisfy the DIOPHANTINE EQUATION

$$3(2y+1)^2 = 8x^3 + 12x^2 + 4x + 3. \tag{7}$$

The only solutions are $x = -1$, 0, 1, 5, 6, and 85 (Guy 1994, p. 147).

A general FORMULA for square triangular numbers is

$$ST_n = \left[\frac{(1+\sqrt{2})^{2n} - (1-\sqrt{2})^{2n}}{4\sqrt{2}}\right]^2 \tag{8}$$
$$= \tfrac{1}{32}[(17+12\sqrt{2})^n + (17-12\sqrt{2})^n - 2]. \tag{9}$$

The square triangular numbers also satisfy the RECURRENCE RELATION

$$ST_n = 34ST_{n-1} - ST_{n-2} + 2 \tag{10}$$
$$u_{n+2} = 6u_{n+1} - u_n, \tag{11}$$

with $u_0 = 0$, $u_1 = 1$, where $ST_n \equiv {u_n}^2$. A curious product formula for ST_n is given by

$$ST_n = 2^{2n-5}\prod_{k=1}^{2n}\left[3+\cos\left(\frac{k\pi}{n}\right)\right]. \tag{12}$$

An amazing GENERATING FUNCTION is

$$f(x) = \frac{1+x}{(1-x)(1-34x+x^2)} = 1 + 36x + 1225x^2 + \ldots \tag{13}$$

(Sloane and Plouffe 1995).

see also SQUARE NUMBER, SQUARE ROOT, TRIANGULAR NUMBER

References

Allen, B. M. "Squares as Triangular Numbers." *Scripta Math.* **20**, 213–214, 1954.

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, 1987.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 203–205, 1996.

Dickson, L. E. *A History of the Theory of Numbers, Vol. 2: Diophantine Analysis.* New York: Chelsea, pp. 10, 16, and 27, 1952.

Guy, R. K. "Sums of Squares" and "Figurate Numbers." §C20 and §D3 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 136–138 and 147–150, 1994.

Khatri, M. N. "Triangular Numbers Which are Also Squares." *Math. Student* **27**, 55–56, 1959.

Pietenpol, J. L. "Square Triangular Numbers." Problem E 1473. *Amer. Math. Monthly* **69**, 168–169, 1962.

Sierpiński, W. *Teoria Liczb, 3rd ed.* Warsaw, Poland: Monografie Matematyczne t. 19, p. 517, 1950.

Sierpiński, W. "Sur les nombres triangulaires carrés." *Pub. Faculté d'Électrotechnique l'Université Belgrade*, No. 65, 1–4, 1961.

Sierpiński, W. "Sur les nombres triangulaires carrés." *Bull. Soc. Royale Sciences Liège*, 30 ann., 189–194, 1961.

Sloane, N. J. A. Sequence A001110/M5259 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Walker, G. W. "Triangular Squares." Problem E 954. *Amer. Math. Monthly* **58**, 568, 1951.

Square Wave

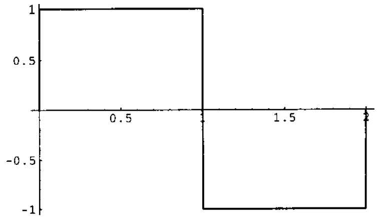

The square wave is a periodic waveform consisting of instantaneous transitions between two levels which can be denoted ±1. The square wave is sometimes also called the RADEMACHER FUNCTION. Let the square wave have period $2L$. The square wave function is ODD, so the FOURIER SERIES has $a_0 = a_n = 0$ and

$$b_n = \frac{2}{L}\int_0^L \sin\left(\frac{n\pi x}{L}\right)\,dx$$
$$= \frac{4}{n\pi}\sin^2(\tfrac{1}{2}n\pi) = \frac{4}{n\pi}\begin{cases} 0 & n \text{ even} \\ 1 & n \text{ odd.}\end{cases}$$

The FOURIER SERIES for the square wave is therefore

$$f(x) = \frac{4}{\pi}\sum_{n=1,3,5,\ldots}^{\infty}\frac{1}{n}\sin\left(\frac{n\pi x}{L}\right).$$

see also HADAMARD MATRIX, WALSH FUNCTION

References

Thompson, A. R.; Moran, J. M.; and Swenson, G. W. Jr. *Interferometry and Synthesis in Radio Astronomy.* New York: Wiley, p. 203, 1986.

Squared

A number to the POWER 2 is said to be squared, so that x^2 is called "x squared."

see also CUBED, SQUARE ROOT

Squared Square

see PERFECT SQUARE DISSECTION

Squarefree

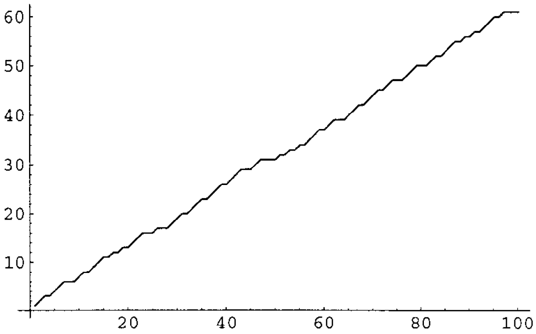

A number is said to be squarefree (or sometimes QUADRATFREI; Shanks 1993) if its PRIME decomposition contains no repeated factors. All PRIMES are therefore trivially squarefree. The squarefree numbers are 1, 2, 3, 5, 6, 7, 10, 11, 13, 14, 15, ... (Sloane's A005117). The SQUAREFUL numbers (i.e., those that contain at least one square) are 4, 8, 9, 12, 16, 18, 20, 24, 25, ... (Sloane's A013929).

The asymptotic number $Q(n)$ of squarefree numbers $\leq n$ is given by

$$Q(n) = \frac{6n}{\pi^2} + \mathcal{O}(\sqrt{n}) \qquad (1)$$

(Hardy and Wright 1979, pp. 269–270). $Q(n)$ for $n = 10, 100, 1000, \ldots$ are 7, 61, 608, 6083, 60794, 607926, ..., while the asymptotic density is $1/\zeta(2) = 6/\pi^2 \approx 0.607927$, where $\zeta(n)$ is the RIEMANN ZETA FUNCTION.

The MÖBIUS FUNCTION is given by

$$\mu(n) \equiv \begin{cases} 0 & \text{if } n \text{ has one or more repeated prime factors} \\ 1 & \text{if } n = 1 \\ (-1)^k & \text{if } n \text{ is product of } k \text{ distinct primes,} \end{cases} \qquad (2)$$

so $\mu(n) \neq 0$ indicates that n is squarefree. The asymptotic formula for $Q(x)$ is equivalent to the formula

$$\sum_{n=1}^{x} |\mu(n)| = \frac{6x}{\pi^2} + \mathcal{O}(\sqrt{x}) \qquad (3)$$

(Hardy and Wright 1979, p. 270)

There is no known polynomial-time algorithm for recognizing squarefree INTEGERS or for computing the squarefree part of an INTEGER. In fact, this problem may be no easier than the general problem of integer factorization (obviously, if an integer n can be factored completely, n is squarefree IFF it contains no duplicated factors). This problem is an important unsolved problem in NUMBER THEORY because computing the RING of integers of an algebraic number field is reducible to computing the squarefree part of an INTEGER (Lenstra 1992, Pohst and Zassenhaus 1997). The *Mathematica®* (Wolfram Research, Champaign, IL) function `NumberTheory'NumberTheoryFunctions'` `SquareFreeQ[n]` determines whether a number is squarefree.

The largest known SQUAREFUL FIBONACCI NUMBER is F_{336}, and no SQUAREFUL FIBONACCI NUMBERS F_p are known with p PRIME. All numbers less than 2.5×10^{15} in SYLVESTER'S SEQUENCE are squarefree, and no SQUAREFUL numbers in this sequence are known (Vardi 1991). Every CARMICHAEL NUMBER is squarefree. The BINOMIAL COEFFICIENTS $\binom{2n-1}{n}$ are squarefree only for $n = 2, 3, 4, 6, 9, 10, 12, 36, \ldots$, with no others less than $n = 1500$. The CENTRAL BINOMIAL COEFFICIENTS are SQUAREFREE only for $n = 1, 2, 3, 4, 5, 7, 8, 11, 17, 19, 23, 71, \ldots$ (Sloane's A046098), with no others less than 1500.

see also BINOMIAL COEFFICIENT, BIQUADRATEFREE, COMPOSITE NUMBER, CUBEFREE, ERDŐS SQUAREFREE CONJECTURE, FIBONACCI NUMBER, KORSELT'S CRITERION, MÖBIUS FUNCTION, PRIME NUMBER, RIEMANN ZETA FUNCTION, SÁRKÖZY'S THEOREM, SQUARE NUMBER, SQUAREFUL, SYLVESTER'S SEQUENCE

References

Bellman, R. and Shapiro, H. N. "The Distribution of Squarefree Integers in Small Intervals." *Duke Math. J.* **21**, 629–637, 1954.

Hardy, G. H. and Wright, E. M. "The Number of Squarefree Numbers." §18.6 in *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, pp. 269–270, 1979.

Lenstra, H. W. Jr. "Algorithms in Algebraic Number Theory." *Bull. Amer. Math. Soc.* **26**, 211–244, 1992.

Pohst, M. and Zassenhaus, H. *Algorithmic Algebraic Number Theory.* Cambridge, England: Cambridge University Press, p. 429, 1997.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 114, 1993.

Sloane, N. J. A. Sequences A013929 and A005117/M0617 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Vardi, I. "Are All Euclid Numbers Squarefree?" §5.1 in *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 7–8, 82–85, and 223–224, 1991.

Squareful

A number is squareful, also called NONSQUAREFREE, if it contains at least one SQUARE in its prime factorization. Such a number is also called SQUAREFUL. The first few are 4, 8, 9, 12, 16, 18, 20, 24, 25, ... (Sloane's A013929). The greatest multiple prime factors for the squareful integers are 2, 2, 3, 2, 2, 3, 2, 2, 5, 3, 2, 2, 3, ... (Sloane's A046028). The least multiple prime factors for squareful integers are 2, 2, 3, 2, 2, 3, 2, 2, 5, 3, 2, 2, 2, ... (Sloane's A046027).

see also GREATEST PRIME FACTOR, LEAST PRIME FACTOR, SMARANDACHE NEAR-TO-PRIMORIAL FUNCTION, SQUAREFREE

References

Sloane, N. J. A. Sequences A013929, A046027, and A046028 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Squaring

Squaring is the GEOMETRIC CONSTRUCTION, using only COMPASS and STRAIGHTEDGE, of a SQUARE which has the same area as a given geometric figure. Squaring is also called QUADRATURE. An object which can be constructed by squaring is called SQUARABLE.

see also CIRCLE SQUARING, COMPASS, CONSTRUCTIBLE NUMBER, GEOMETRIC CONSTRUCTION, RECTANGLE SQUARING, STRAIGHTEDGE, TRIANGLE SQUARING

Squeezing Theorem

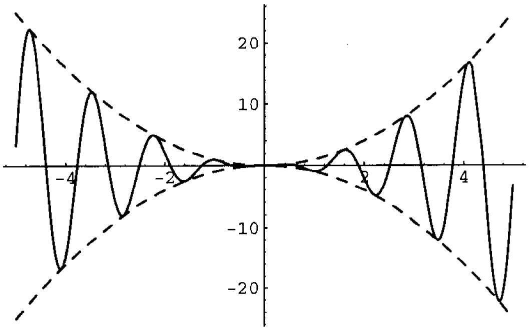

Let there be two functions $f_-(x)$ and $f_+(x)$ such that $f(x)$ is "squeezed" between the two,

$$f_-(x) \leq f(x) \leq f_+(x).$$

If

$$r = \lim_{x \to a} f_-(x) = \lim_{x \to a} f_+(x),$$

then $\lim_{x \to a} f(x) = r$. In the above diagram the functions $f_-(x) = -x^2$ and $f_+(x) = x^2$ "squeeze" $x^2 \sin(cx)$ at 0, so $\lim_{x \to 0} x^2 \sin(cx) = 0$. The squeezing theorem is also called the SANDWICH THEOREM.

SSS Theorem

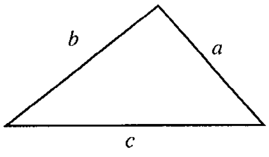

Specifying three sides uniquely determines a TRIANGLE whose AREA is given by HERON'S FORMULA,

$$A = \sqrt{s(s-a)(s-b)(s-c)}, \tag{1}$$

where

$$s \equiv \tfrac{1}{2}(a+b+c) \tag{2}$$

is the SEMIPERIMETER of the TRIANGLE. Let R be the CIRCUMRADIUS, then

$$A = \frac{abc}{4R}. \tag{3}$$

Using the LAW OF COSINES

$$a^2 = b^2 + c^2 - 2bc \cos A \tag{4}$$

$$b^2 = a^2 + c^2 - 2ac \cos B \tag{5}$$

$$c^2 = a^2 + b^2 - 2ab \cos C \tag{6}$$

gives the three ANGLES as

$$A = \cos^{-1}\left(\frac{b^2 + c^2 - a^2}{2bc}\right) \tag{7}$$

$$B = \cos^{-1}\left(\frac{b^2 + c^2 - b^2}{2ac}\right) \tag{8}$$

$$C = \cos^{-1}\left(\frac{a^2 + b^2 - c^2}{2ab}\right). \tag{9}$$

see also AAA THEOREM, AAS THEOREM, ASA THEOREM, ASS THEOREM, HERON'S FORMULA, SAS THEOREM, SEMIPERIMETER, TRIANGLE

Stability

The robustness of a given outcome to small changes in initial conditions or small random fluctuations. CHAOS is an example of a process which is not stable.

see also STABILITY MATRIX

Stability Matrix

Given a system of two ordinary differential equations

$$\dot{x} = f(x, y) \tag{1}$$

$$\dot{y} = g(x, y), \tag{2}$$

let x_0 and y_0 denote FIXED POINTS with $\dot{x} = \dot{y} = 0$, so

$$f(x_0, y_0) = 0 \tag{3}$$

$$g(x_0, y_0) = 0. \tag{4}$$

Then expand about (x_0, y_0) so

$$\begin{aligned}\delta\dot{x} &= f_x(x_0, y_0)\delta x + f_y(x_0, y_0)\delta y \\ &\quad + f_{xy}(x_0, y_0)\delta x \delta y + \ldots \end{aligned} \tag{5}$$

$$\begin{aligned}\delta\dot{y} &= g_x(x_0, y_0)\delta x + g_y(x_0, y_0)\delta y \\ &\quad + g_{xy}(x_0, y_0)\delta x \delta y + \ldots. \end{aligned} \tag{6}$$

To first-order, this gives

$$\frac{d}{dt}\begin{bmatrix} \delta x \\ \delta y \end{bmatrix} = \begin{bmatrix} f_x(x_0, y_0) & f_y(x_0, y_0) \\ g_x(x_0, y_0) & g_y(x_0, y_0) \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \end{bmatrix}, \tag{7}$$

where the 2×2 MATRIX, or its generalization to higher dimension, is called the stability matrix. Analysis of the EIGENVALUES (and EIGENVECTORS) of the stability matrix characterizes the type of FIXED POINT.

see also ELLIPTIC FIXED POINT (DIFFERENTIAL EQUATIONS), FIXED POINT, HYPERBOLIC FIXED POINT

(Differential Equations), Linear Stability, Stable Improper Node, Stable Node, Stable Spiral Point, Stable Star, Unstable Improper Node, Unstable Node, Unstable Spiral Point, Unstable Star

References

Tabor, M. "Linear Stability Analysis." §1.4 in *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, pp. 20–31, 1989.

Stabilization

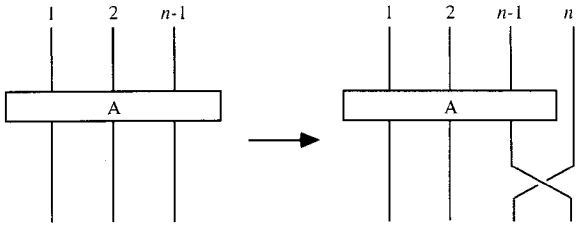

A type II Markov Move.

see also Markov Moves

Stable Equivalence

Two Vector Bundles are stably equivalent Iff Isomorphic Vector Bundles are obtained upon Whitney Summing each Vector Bundle with a trivial Vector Bundle.

see also Vector Bundle, Whitney Sum

Stable Improper Node

A Fixed Point for which the Stability Matrix has equal Negative Eigenvalues.

see also Elliptic Fixed Point (Differential Equations), Fixed Point, Hyperbolic Fixed Point (Differential Equations), Stable Node, Stable Spiral Point, Unstable Improper Node, Unstable Node, Unstable Spiral Point, Unstable Star

References

Tabor, M. "Classification of Fixed Points." §1.4.b in *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, pp. 22–25, 1989.

Stable Node

A Fixed Point for which the Stability Matrix has both Eigenvalues Negative, so $\lambda_1 < \lambda_2 < 0$.

see also Elliptic Fixed Point (Differential Equations), Fixed Point, Hyperbolic Fixed Point (Differential Equations), Stable Improper Node, Stable Spiral Point, Stable Star, Unstable Improper Node, Unstable Node, Unstable Spiral Point, Unstable Star

References

Tabor, M. "Classification of Fixed Points." §1.4.b in *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, pp. 22–25, 1989.

Stable Spiral Point

A Fixed Point for which the Stability Matrix has Eigenvalues of the form $\lambda_\pm = -\alpha \pm i\beta$ (with $\alpha, \beta > 0$).

see also Elliptic Fixed Point (Differential Equations), Fixed Point, Hyperbolic Fixed Point (Differential Equations), Stable Improper Node, Stable Node, Stable Star, Unstable Improper Node, Unstable Node, Unstable Spiral Point, Unstable Star

References

Tabor, M. "Classification of Fixed Points." §1.4.b in *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, pp. 22–25, 1989.

Stable Star

A Fixed Point for which the Stability Matrix has one zero Eigenvector with Negative Eigenvalue $\lambda < 0$.

see also Elliptic Fixed Point (Differential Equations), Fixed Point, Hyperbolic Fixed Point (Differential Equations), Stable Improper Node, Stable Node, Stable Spiral Point, Unstable Improper Node, Unstable Node, Unstable Spiral Point, Unstable Star

References

Tabor, M. "Classification of Fixed Points." §1.4.b in *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, pp. 22–25, 1989.

Stable Type

A Polynomial equation whose Roots all have Negative Real Parts. For a Real Quadratic Equation

$$z^2 + Bz + C = 0,$$

the stability conditions are $B, C > 0$. For a Real Cubic Equation

$$z^3 + Az^2 + Bz + C = 0,$$

the stability conditions are $A, B, C > 0$ and $AB > C$.

References

Birkhoff, G. and Mac Lane, S. *A Survey of Modern Algebra, 3rd ed.* New York: Macmillan, pp. 108–109, 1965.

Stack

A Data Structure which is a special kind of List in which elements may be added to or removed from the top only. These actions are called a Push or a Pop, respectively. Actions may be taken by popping one or more values, operating on them, and then pushing the result back onto the stack.

Stacks are used as the basis for computer languages such as FORTH, PostScript® (Adobe Systems), and the RPN language used in Hewlett-Packard® programmable calculators.

see also List, Pop, Push, Queue

Stäckel Determinant

A DETERMINANT used to determine in which coordinate systems the HELMHOLTZ DIFFERENTIAL EQUATION is separable (Morse and Feshbach 1953). A determinant

$$S = |\Phi_{mn}| = \begin{vmatrix} \Phi_{11} & \Phi_{12} & \Phi_{13} \\ \Phi_{21} & \Phi_{22} & \Phi_{23} \\ \Phi_{31} & \Phi_{32} & \Phi_{33} \end{vmatrix} \tag{1}$$

in which Φ_{ni} are functions of u_i alone is called a Stäckel determinant. A coordinate system is separable if it obeys the ROBERTSON CONDITION, namely that the SCALE FACTORS h_i in the LAPLACIAN

$$\nabla^2 = \sum_{i=1}^{3} \frac{1}{h_1h_2h_3}\frac{\partial}{\partial u_i}\left(\frac{h_1h_2h_3}{{h_i}^2}\frac{\partial}{\partial u_i}\right) \tag{2}$$

can be rewritten in terms of functions $f_i(u_i)$ defined by

$$\begin{aligned} \frac{1}{h_1h_2h_3}\frac{\partial}{\partial u_i}\left(\frac{h_1h_2h_3}{{h_i}^2}\frac{\partial}{\partial u_i}\right) &= \frac{g(u_{i+1},u_{i+2})}{h_1h_2h_3}\frac{\partial}{\partial u_i}\left[f_i(u_i)\frac{\partial}{\partial u_i}\right] \\ &= \frac{1}{{h_i}^2 f_i}\frac{\partial}{\partial u_i}\left(f_i\frac{\partial}{\partial u_i}\right) \end{aligned} \tag{3}$$

such that S can be written

$$S = \frac{h_1h_2h_3}{f_1(u_1)f_2(u_2)f_3(u_3)}. \tag{4}$$

When this is true, the separated equations are of the form

$$\frac{1}{f_n}\frac{\partial}{\partial u_n}\left(f_n\frac{\partial X_n}{\partial u_n}\right) + ({k_1}^2\Phi_{n1} + {k_2}^2\Phi_{n2} + {k_3}^2\Phi_{n3})X_n = 0 \tag{5}$$

The Φ_{ij}s obey the minor equations

$$M_1 = \Phi_{22}\Phi_{33} - \Phi_{23}\Phi_{32} = \frac{S}{h_1^2} \tag{6}$$

$$M_2 = \Phi_{13}\Phi_{31} - \Phi_{12}\Phi_{33} = \frac{S}{h_2^2} \tag{7}$$

$$M_3 = \Phi_{12}\Phi_{23} - \Phi_{13}\Phi_{22} = \frac{S}{h_3^2}, \tag{8}$$

which are equivalent to

$$M_1\Phi_{11} + M_2\Phi_{21} + M_3\Phi_{31} = S \tag{9}$$

$$M_1\Phi_{12} + M_2\Phi_{22} + M_3\Phi_{32} = 0 \tag{10}$$

$$M_1\Phi_{13} + M_2\Phi_{23} + M_3\Phi_{33} = 0. \tag{11}$$

This gives a total of four equations in nine unknowns. Morse and Feshbach (1953, pp. 655–666) give not only the Stäckel determinants for common coordinate systems, but also the elements of the determinant (although it is not clear how these are derived).

see also HELMHOLTZ DIFFERENTIAL EQUATION, LAPLACE'S EQUATION, POISSON'S EQUATION, ROBERTSON CONDITION, SEPARATION OF VARIABLES

References

Morse, P. M. and Feshbach, H. "Tables of Separable Coordinates in Three Dimensions." *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 509–511 and 655–666, 1953.

Stamp Folding

The number of ways of folding a strip of stamps has several possible variants. Considering only positions of the hinges for unlabeled stamps without regard to orientation of the stamps, the number of foldings is denoted $U(n)$. If the stamps are labelled and orientation is taken into account, the number of foldings is denoted $N(n)$. Finally, the number of symmetric foldings is denoted $S(n)$. The following table summarizes these values for the first n.

n	$S(n)$	$U(n)$	$N(n)$
1	1	1	1
2	1	1	1
3	2	2	6
4	4	5	16
5	6	14	50
6	8	39	144
7	18	120	462
8	20	358	1392
9	56	1176	4536
10		3572	

see also MAP FOLDING

References

Gardner, M. "The Combinatorics of Paper-Folding." In *Wheels, Life, and Other Mathematical Amusements.* New York: W. H. Freeman, pp. 60–73, 1983.

Ruskey, F. "Information of Stamp Folding." `http://sue.csc.uvic.ca/~cos/inf/perm/StampFolding.html`.

Sloane, N. J. A. *A Handbook of Integer Sequences.* Boston, MA: Academic Press, p. 22, 1973.

Standard Deviation

The standard deviation is defined as the SQUARE ROOT of the VARIANCE,

$$\sigma = \sqrt{\langle x^2\rangle - \langle x\rangle^2} = \sqrt{\mu_2' - \mu^2}, \tag{1}$$

where $\mu = \langle x\rangle$ is the MEAN and $\mu_2' = \langle x^2\rangle$ is the second MOMENT about 0. The variance σ^2 is equal to the second MOMENT about the MEAN,

$$\sigma^2 = \mu_2. \tag{2}$$

The square root of the SAMPLE VARIANCE is the "sample" standard deviation,

$$s_N = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}. \tag{3}$$

It is a Biased Estimator of the population standard deviation. As unbiased Estimator is given by

$$s_{N-1} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}. \qquad (4)$$

Physical scientists often use the term Root-Mean-Square as a synonym for standard deviation when they refer to the Square Root of the mean squared deviation of a signal from a given baseline or fit.

see also Mean, Moment, Root-Mean-Square, Sample Variance, Standard Error, Variance

Standard Error

The square root of the Estimated Variance of a quantity. The standard error is also sometimes used to mean

$$\mathrm{var}(\bar{x}) = \sum_{i=1}^{n}\left(\frac{1}{n}\right)^2 {\sigma_i}^2 = \sum_{i=1}^{n}\left(\frac{1}{n}\right)^2 \sigma^2 = \frac{\sigma^2}{n}.$$

see also Standard Deviation

Standard Map

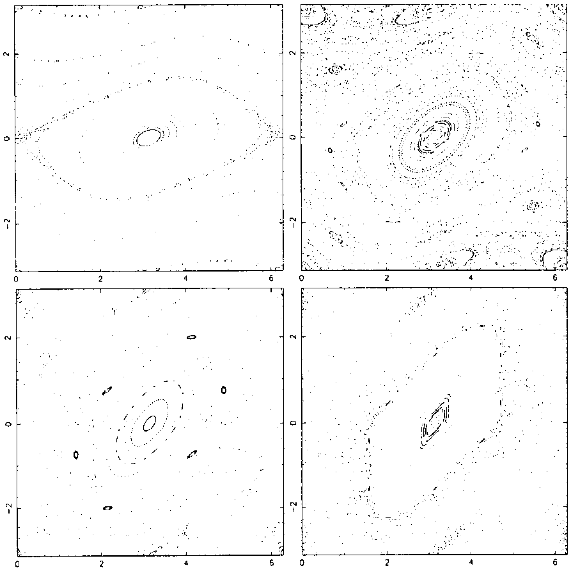

A 2-D Map, also called the Taylor-Greene-Chirikov Map in some of the older literature.

$$I_{n+1} = I_n + K\sin\theta_n \qquad (1)$$
$$\theta_{n+1} = \theta_n + I_{n+1} = I_n + \theta_n + K\sin\theta_n, \qquad (2)$$

where I and θ are computed mod 2π and K is a Positive constant. An analytic estimate of the width of the Chaotic zone (Chirikov 1979) finds

$$\delta I = Be^{-AK^{-1/2}}. \qquad (3)$$

Numerical experiments give $A \approx 5.26$ and $B \approx 240$. The value of K at which global Chaos occurs has been bounded by various authors. Greene's Method is the most accurate method so far devised.

Author	Bound	Fraction	Decimal
Hermann	$>$	$\frac{1}{34}$	0.029411764
Italians	$>$	–	0.65
Greene	$\approx$	–	0.971635406
MacKay and Pearson	$<$	$\frac{63}{64}$	0.984375000
Mather	$<$	$\frac{4}{3}$	1.333333333

Fixed Points are found by requiring that

$$I_{n+1} = I_n \qquad (4)$$
$$\theta_{n+1} = \theta_n. \qquad (5)$$

The first gives $K\sin\theta_n = 0$, so $\sin\theta_n = 0$ and

$$\theta_n = 0, \pi. \qquad (6)$$

The second requirement gives

$$I_n + K\sin\theta_n = I_n = 0. \qquad (7)$$

The Fixed Points are therefore $(I,\theta) = (0,0)$ and $(0,\pi)$. In order to perform a Linear Stability analysis, take differentials of the variables

$$dI_{n+1} = dI_n + K\cos\theta_n\, d\theta_n \qquad (8)$$
$$d\theta_{n+1} = dI_n + (1 + K\cos\theta_n)\, d\theta_n. \qquad (9)$$

In Matrix form,

$$\begin{bmatrix}\delta I_{n+1}\\ \delta\theta_{n+1}\end{bmatrix} = \begin{bmatrix}1 & K\cos\theta_n\\ 1 & 1+K\cos\theta_n\end{bmatrix}\begin{bmatrix}\delta I_n\\ \delta\theta_n\end{bmatrix}. \qquad (10)$$

The Eigenvalues are found by solving the Characteristic Equation

$$\begin{vmatrix}1-\lambda & K\cos\theta_n\\ 1 & 1+K\cos\theta_n-\lambda\end{vmatrix} = 0, \qquad (11)$$

so

$$\lambda^2 - \lambda(K\cos\theta_n + 2) + 1 = 0 \qquad (12)$$
$$\lambda_\pm = \tfrac{1}{2}[K\cos\theta_n + 2 \pm \sqrt{(K\cos\theta_n + 2)^2 - 4}\,]. \qquad (13)$$

For the Fixed Point $(0,\pi)$,

$$\lambda_\pm^{(0,\pi)} = \tfrac{1}{2}[2 - K \pm \sqrt{(2-K)^2 - 4}\,]$$
$$= \tfrac{1}{2}(2 - K \pm \sqrt{K^2 - 4K}\,). \qquad (14)$$

The Fixed Point will be stable if $|\Re(\lambda^{(0,\pi)})| < 2$. Here, that means

$$\tfrac{1}{2}|2-K| < 1 \qquad (15)$$
$$|2-K| < 2 \qquad (16)$$
$$-2 < 2-K < 2 \qquad (17)$$
$$-4 < -K < 0 \qquad (18)$$

so $K \in [0,4)$. For the FIXED POINT (0, 0), the EIGENVALUES are

$$\begin{aligned}\lambda_{\pm}^{(0,0)} &= \tfrac{1}{2}[2+K\pm\sqrt{(K+2)^2-4}\,]\\ &= \tfrac{1}{2}(2+K\pm\sqrt{K^2+4K}\,). \qquad (19)\end{aligned}$$

If the map is unstable for the larger EIGENVALUE, it is unstable. Therefore, examine $\lambda_{+}^{(0,0)}$. We have

$$\frac{1}{2}\left|2+K+\sqrt{K^2+4K}\right| < 1, \qquad (20)$$

so

$$-2 < 2+K+\sqrt{K^2+4K} < 2 \qquad (21)$$

$$-4-K < \sqrt{K^2+4K} < -K. \qquad (22)$$

But $K > 0$, so the second part of the inequality cannot be true. Therefore, the map is unstable at the FIXED POINT (0, 0).

References

Chirikov, B. V. "A Universal Instability of Many-Dimensional Oscillator Systems." *Phys. Rep.* **52**, 264–379, 1979.

Standard Normal Distribution

A NORMAL DISTRIBUTION with zero MEAN ($\mu = 0$) and unity STANDARD DEVIATION ($\sigma^2 = 1$).

see also NORMAL DISTRIBUTION

Standard Space

A SPACE which is ISOMORPHIC to a BOREL SUBSET B of a POLISH SPACE equipped with its SIGMA ALGEBRA of BOREL SETS.

see also BOREL SET, POLISH SPACE, SIGMA ALGEBRA

Standard Tori

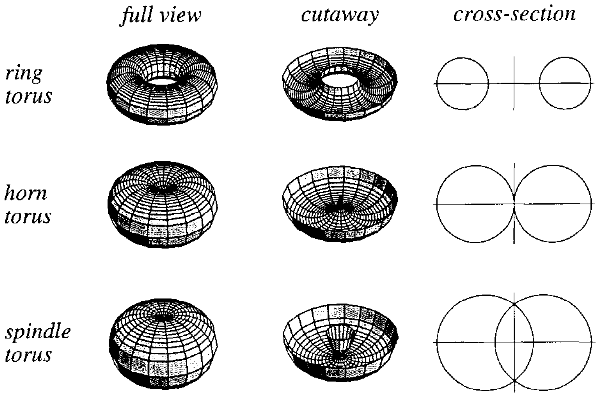

One of the three classes of TORI illustrated above and given by the parametric equations

$$x = (c + a\cos v)\cos u \qquad (1)$$

$$y = (c + a\cos v)\sin u \qquad (2)$$

$$z = a\sin v. \qquad (3)$$

The three different classes of standard tori arise from the three possible relative sizes of a and c. $c > a$ corresponds to the RING TORUS shown above, $c = a$ corresponds to a HORN TORUS which touches itself at the point (0, 0, 0), and $c < a$ corresponds to a self-intersecting SPINDLE TORUS (Pinkall 1986). If no specification is made, "torus" is taken to mean RING TORUS.

The standard tori and their inversions are CYCLIDES.

see also APPLE, CYCLIDE, HORN TORUS, LEMON, RING TORUS, SPINDLE TORUS, TORUS

References

Pinkall, U. "Cyclides of Dupin." §3.3 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 28–30, 1986.

Standardized Moment

Defined for samples x_i, $i = 1, \ldots, N$ by

$$\alpha_r \equiv \frac{1}{N}\sum_{i=1}^{N} {z_i}^r = \frac{\mu_r}{\sigma^r}, \qquad (1)$$

where

$$z_i \equiv \frac{x_i - \bar{x}}{s_x}. \qquad (2)$$

The first few are

$$\alpha_1 = 0 \qquad (3)$$

$$\alpha_2 = 1 \qquad (4)$$

$$\alpha_3 = \frac{\mu_3}{s^3} \qquad (5)$$

$$\alpha_4 = \frac{\mu_4}{s^4}. \qquad (6)$$

see also KURTOSIS, MOMENT, SKEWNESS

Standardized Score

see z-SCORE

Stanley's Theorem

The total number of 1s that occur among all unordered PARTITIONS of a POSITIVE INTEGER is equal to the sum of the numbers of distinct parts of (i.e., numbers in) those PARTITIONS.

see also ELDER'S THEOREM, PARTITION

References

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer, pp. 6–8, 1985.

Star

In formal geometry, a star is a set of $2n$ VECTORS $\pm\mathbf{a}_1$, ..., $\pm\mathbf{a}_n$ which form a fixed center in EUCLIDEAN 3-SPACE. In common usage, a star is a STAR POLYGON (i.e., regular convex polygon) such as the PENTAGRAM or HEXAGRAM

see also CROSS, EUTACTIC STAR, STAR OF GOLIATH, STAR POLYGON

Star of David

see HEXAGRAM

Star Figure

A STAR POLYGON-like figure $\left\{\frac{p}{q}\right\}$ for which p and q are not RELATIVELY PRIME.

see also STAR POLYGON

Star (Fixed Point)

A FIXED POINT which has one zero EIGENVECTOR.

see STABLE STAR, UNSTABLE STAR

Star Fractal

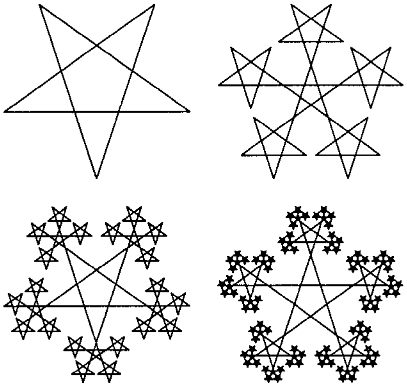

A FRACTAL composed of repeated copies of a PENTAGRAM or other polygon.

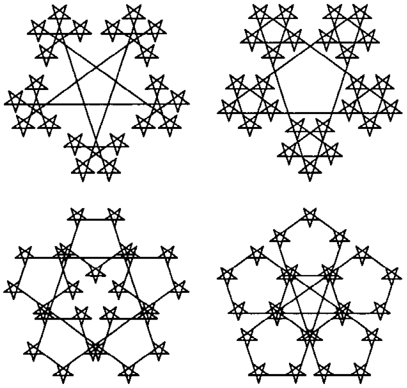

The above figure shows a generalization to different offsets from the center.

References

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 72–77, 1991.

Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m`.

Star of Goliath

see NONAGRAM

Star Graph

The k-star graph is a TREE on $k+1$ nodes with one node having valency k and the others having valency 1. Star graphs S_n are always GRACEFUL.

Star of Lakshmi

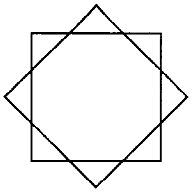

The STAR FIGURE $\{8/2\}$, which is used by Hindus to symbolize *Ashtalakshmi,* the eight forms of wealth. This symbol appears prominently in the Lugash national museum portrayed in the fictional film *Return of the Pink Panther.*

see also DISSECTION, HEXAGRAM, PENTAGRAM, STAR FIGURE, STAR POLYGON

References

Savio, D. Y. and Suryanaroyan, E. R. "Chebyshev Polynomials and Regular Polygons." *Amer. Math. Monthly* **100**, 657–661, 1993.

Star Number

The number of cells in a generalized Chinese checkers board (or "centered" HEXAGRAM).

$$S_n = 6n(n+1)+1 = S_{n-1}+12(n-1). \quad (1)$$

The first few are 1, 13, 37, 73, 121, ... (Sloane's A003154). Every star number has DIGITAL ROOT 1 or 4, and the final digits must be one of: 01, 21, 41, 61, 81, 13, 33, 53, 73, 93, or 37.

The first TRIANGULAR star numbers are 1, 253, 49141, 9533161, ... (Sloane's A006060), and can be computed using

$$\begin{aligned} TS_n &= \frac{3[(7+4\sqrt{3})^{2n-1}+(7-4\sqrt{3})^{2n-1}]-10}{32} \\ &= 194TS_{n-1}+60-TS_{n-2}. \end{aligned} \quad (2)$$

The first few SQUARE star numbers are 1, 121, 11881, 1164241, 114083761, ... (Sloane's A006061). SQUARE star numbers are obtained by solving the DIOPHANTINE EQUATION

$$2x^2+1=3y^2 \quad (3)$$

and can be computed using

$$SS_n = \frac{[(5+2\sqrt{6})^n(\sqrt{6}-2)-(5-2\sqrt{6})^n(\sqrt{6}+2)]^2}{4}. \tag{4}$$

see also HEX NUMBER, SQUARE NUMBER, TRIANGULAR NUMBER

References

Gardner, M. "Hexes and Stars." Ch. 2 in *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, 1988.

Hindin, H. "Stars, Hexes, Triangular Numbers, and Pythagorean Triples." *J. Recr. Math.* **16**, 191–193, 1983–1984.

Sloane, N. J. A. Sequences A003154/M4893, A006060/M5425, and A006061/M5385 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Star Polygon

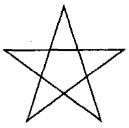
$\{\frac{5}{2}\}$

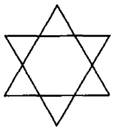
$\{\frac{6}{2}\}$

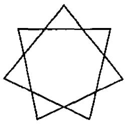
$\{\frac{7}{2}\}$

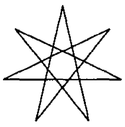
$\{\frac{7}{3}\}$

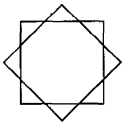
$\{\frac{8}{2}\}$

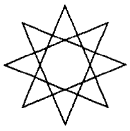
$\{\frac{8}{3}\}$

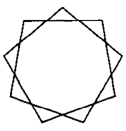
$\{\frac{9}{2}\}$

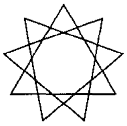
$\{\frac{9}{3}\}$

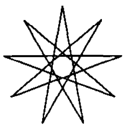
$\{\frac{9}{4}\}$

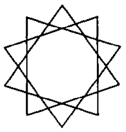
$\{\frac{10}{3}\}$

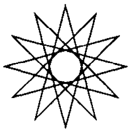
$\{\frac{12}{5}\}$

A star polygon $\{p/q\}$, with p, q POSITIVE INTEGERS, is a figure formed by connecting with straight lines every qth point out of p regularly spaced points lying on a CIRCUMFERENCE. The number q is called the DENSITY of the star polygon. Without loss of generality, take $q < p/2$.

The usual definition (Coxeter 1969) requires p and q to be RELATIVELY PRIME. However, the star polygon can also be generalized to the STAR FIGURE (or "improper" star polygon) when p and q share a common divisor (Savio and Suryanaroyan 1993). For such a figure, if all points are not connected after the first pass, i.e., if $(p,q) \neq 1$, then start with the first unconnected point and repeat the procedure. Repeat until all points are connected. For $(p,q) \neq 1$, the $\{p/q\}$ symbol can be factored as

$$\left\{\frac{p}{q}\right\} = n\left\{\frac{p}{q'}\right\}, \tag{1}$$

where

$$p' \equiv \frac{p}{n} \tag{2}$$

$$q' \equiv \frac{q}{n}, \tag{3}$$

to give n $\{p'/q'\}$ figures, each rotated by $2\pi/p$ radians, or $360°/p$.

If $q = 1$, a REGULAR POLYGON $\{p\}$ is obtained. Special cases of $\{p/q\}$ include $\{5/2\}$ (the PENTAGRAM), $\{6/2\}$ (the HEXAGRAM, or STAR OF DAVID), $\{8/2\}$ (the STAR OF LAKSHMI), $\{8/3\}$ (the OCTAGRAM), $\{10/3\}$ (the DECAGRAM), and $\{12/5\}$ (the DODECAGRAM).

The star polygons were first systematically studied by Thomas Bradwardine.

see also DECAGRAM, HEXAGRAM, NONAGRAM, OCTAGRAM, PENTAGRAM, REGULAR POLYGON, STAR OF LAKSHMI, STELLATED POLYHEDRON

References

Coxeter, H. S. M. "Star Polygons." §2.8 in *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 36–38, 1969.

Frederickson, G. "Stardom." Ch. 16 in *Dissections: Plane and Fancy.* New York: Cambridge University Press, pp. 172–186, 1997.

Savio, D. Y. and Suryanaroyan, E. R. "Chebyshev Polynomials and Regular Polygons." *Amer. Math. Monthly* **100**, 657–661, 1993.

Star Polyhedron

see KEPLER-POINSOT SOLID

Starr Rose

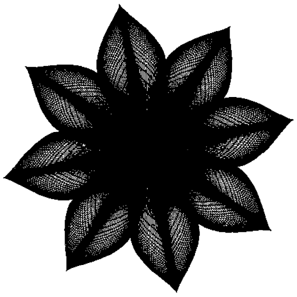
$a = 8, b = 16, c = 16$

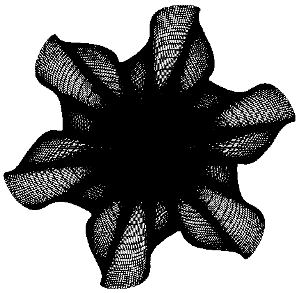
$a = 6, b = 18, c = 18$

see also MAURER ROSE

References

Wagon, S. "Variations of Circular Motion." §4.5 in *Mathematica in Action.* New York: W. H. Freeman, pp. 137–140, 1991.

State Space

The measurable space $(S', \mathbb{S}')$ into which a RANDOM VARIABLE from a PROBABILITY SPACE is a measurable function.

see also PROBABILITY SPACE, RANDOM VARIABLE

Stationary Point

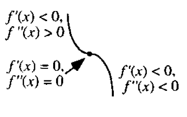

stationary point

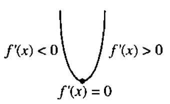

minimum

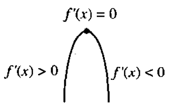

maximum

A point x_0 at which the DERIVATIVE of a FUNCTION $f(x)$ vanishes,

$$f'(x_0) = 0.$$

A stationary point may be a MINIMUM, MAXIMUM, or INFLECTION POINT.

see also CRITICAL POINT, DERIVATIVE, EXTREMUM, FIRST DERIVATIVE TEST, INFLECTION POINT, MAXIMUM, MINIMUM, SECOND DERIVATIVE TEST

Stationary Tangent

see INFLECTION POINT

Stationary Value

The value at a STATIONARY POINT.

Statistic

A function of one or more random variables.

see also ANDERSON-DARLING STATISTIC, KUIPER STATISTIC, VARIATE

Statistical Test

A test used to determine the statistical SIGNIFICANCE of an observation. Two main types of error can occur:

1. A TYPE I ERROR occurs when a false negative result is obtained in terms of the NULL HYPOTHESIS by obtaining a *false positive measurement.*
2. A TYPE II ERROR occurs when a false positive result is obtained in terms of the NULL HYPOTHESIS by obtaining a *false negative measurement.*

The probability that a statistical test will be positive for a true statistic is sometimes called the test's SENSITIVITY, and the probability that a test will be negative for a negative statistic is sometimes called the SPECIFICITY. The following table summarizes the names given to the various combinations of the actual state of affairs and observed test results.

result	name
true positive result	sensitivity
false negative result	1 − sensitivity
true negative result	specificity
false positive result	1 − specificity

Multiple-comparison corrections to statistical tests are used when several statistical tests are being performed simultaneously. For example, let's suppose you were measuring leg length in eight different lizard species and wanted to see whether the MEANS of any pair were different. Now, there are $8!/2!6! = 28$ pairwise comparisons possible, so even if all of the *population* means are equal, it's quite likely that at least one pair of sample means would differ significantly at the 5% level. An ALPHA VALUE of 0.05 is therefore appropriate for each individual comparison, but not for the set of *all* comparisons.

In order to avoid a lot of spurious positives, the ALPHA VALUE therefore needs to be lowered to account for the number of comparisons being performed. This is a correction for multiple comparisons. There are *many* different ways to do this. The simplest, and the most conservative, is the BONFERRONI CORRECTION. In practice, more people are more willing to accept false positives (false rejection of NULL HYPOTHESIS) than false negatives (false acceptance of NULL HYPOTHESIS), so less conservative comparisons are usually used.

see also ANOVA, BONFERRONI CORRECTION, CHI-SQUARED TEST, FISHER'S EXACT TEST, FISHER SIGN TEST, KOLMOGOROV-SMIRNOV TEST, LIKELIHOOD RATIO, LOG LIKELIHOOD PROCEDURE, NEGATIVE LIKELIHOOD RATIO, PAIRED t-TEST, PARAMETRIC TEST, PREDICTIVE VALUE, SENSITIVITY, SIGNIFICANCE TEST, SPECIFICITY, TYPE I ERROR, TYPE II ERROR, WILCOXON RANK SUM TEST, WILCOXON SIGNED RANK TEST

Statistics

The mathematical study of the LIKELIHOOD and PROBABILITY of events occurring based on known information and inferred by taking a limited number of samples. Statistics plays an extremely important role in many aspects of economics and science, allowing educated guesses to be made with a minimum of expensive or difficult-to-obtain data.

see also BOX-AND-WHISKER PLOT, BUFFON-LAPLACE NEEDLE PROBLEM, BUFFON'S NEEDLE PROBLEM, CHERNOFF FACE, COIN FLIPPING, DE MERE'S PROBLEM, DICE, DISTRIBUTION, GAMBLER'S RUIN, INDEX, LIKELIHOOD, MOVING AVERAGE, P-VALUE, POPULATION COMPARISON, POWER (STATISTICS), PROBABILITY, RESIDUAL VS. PREDICTOR PLOT, RUN, SHARING PROBLEM, STATISTICAL TEST, TAIL PROBABILITY

References

Brown, K. S. "Probability." `http://www.seanet.com/~ksbrown/iprobabi.htm`.

Babu, G. and Feigelson, E. *Astrostatistics.* New York: Chapman & Hall, 1996.

Dixon, W. J. and Massey, F. J. *Introduction to Statistical Analysis, 4th ed.* New York: McGraw-Hill, 1983.

Doob, J. L. *Stochastic Processes.* New York: Wiley, 1953.

Feller, W. *An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd ed.* New York: Wiley, 1968.

Feller, W. *An Introduction to Probability Theory and Its Applications, Vol. 2, 2nd ed.* New York: Wiley, 1968.

Fisher, N. I.; Lewis, T.; and Embleton, B. J. J. *Statistical Analysis of Spherical Data.* Cambridge, England: Cambridge University Press, 1987.

Fisher, R. A. and Prance, G. T. *The Design of Experiments, 9th ed. rev.* New York: Hafner, 1974.

Fisher, R. A. *Statistical Methods for Research Workers, 14th ed., rev. and enl.* Darien, CO: Hafner, 1970.

Goldberg, S. *Probability: An Introduction.* New York: Dover, 1986.

Gonick, L. and Smith, W. *The Cartoon Guide to Statistics.* New York: Harper Perennial, 1993.

Goulden, C. H. *Methods of Statistical Analysis, 2nd ed.* New York: Wiley, 1956.

Hoel, P. G.; Port, S. C.; and Stone, C. J. *Introduction to Statistical Theory.* New York: Houghton Mifflin, 1971.

Hogg, R. V. and Tanis, E. A. *Probability and Statistical Inference, 3rd ed.* New York: Macmillan, 1988.

Keeping, E. S. *Introduction to Statistical Inference.* New York: Dover, 1995.

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, 1962.

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, 1951.

Kendall, M. G.; Stuart, A.; and Ord, J. K. *Kendall's Advanced Theory of Statistics, Vol. 1: Distribution Theory, 6th ed.*0340614307 New York: Oxford University Press, 1987.

Kendall, M. G.; Stuart, A.; and Ord, J. K. *Kendall's Advanced Theory of Statistics, Vol. 2A: 5th ed.* New York: Oxford University Press, 1987.

Kendall, M. G.; Stuart, A.; and Ord, J. K. *Kendall's Advanced Theory of Statistics, Vol. 2B: Bayesian Inference.* New York: Oxford University Press, 1987.

Keynes, J. M. *A Treatise on Probability.* London: Macmillan, 1921.

Mises, R. von *Mathematical Theory of Probability and Statistics.* New York: Academic Press, 1964.

Mises, R. von *Probability, Statistics, and Truth, 2nd rev. English ed.* New York: Dover, 1981.

Mood, A. M. *Introduction to the Theory of Statistics.* New York: McGraw-Hill, 1950.

Mosteller, F. *Fifty Challenging Problems in Probability with Solutions.* New York: Dover, 1987.

Mosteller, F.; Rourke, R. E. K.; and Thomas, G. B. *Probability: A First Course, 2nd ed.* Reading, MA: Addison-Wesley, 1970.

Neyman, J. *First Course in Probability and Statistics.* New York: Holt, 1950.

Ostle, B. *Statistics in Research: Basic Concepts and Techniques for Research Workers, 4th ed.* Ames, IA: Iowa State University Press, 1988.

Papoulis, A. *Probability, Random Variables, and Stochastic Processes, 2nd ed.* New York: McGraw-Hill, 1984.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Statistical Description of Data." Ch. 14 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 603–649, 1992.

Pugh, E. M. and Winslow, G. H. *The Analysis of Physical Measurements.* Reading, MA: Addison-Wesley, 1966.

Rényi, A. *Foundations of Probability.* San Francisco, CA: Holden-Day, 1970.

Robbins, H. and van Ryzin, J. *Introduction to Statistics.* Chicago, IL: Science Research Associates, 1975.

Ross, S. M. *A First Course in Probability.* New York: Macmillan, 1976.

Ross, S. M. *Introduction to Probability and Statistics for Engineers and Scientists.* New York: Wiley, 1987.

Ross, S. M. *Applied Probability Models with Optimization Applications.* New York: Dover, 1992.

Ross, S. M. *Introduction to Probability Models, 5th ed.* New York: Academic Press, 1993.

Snedecor, G. W. *Statistical Methods Applied to Experiments in Agriculture and Biology, 5th ed.* Ames, IA: State College Press, 1956.

Tippett, L. H. C. *The Methods of Statistics: An Introduction Mainly for Experimentalists, 3rd rev. ed.* London: Williams and Norgate, 1941.

Todhunter, I. *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace.* New York: Chelsea, 1949.

Tukey, J. W. *Explanatory Data Analysis.* Reading, MA: Addison-Wesley, 1977.

Uspensky, J. V. *Introduction to Mathematical Probability.* New York: McGraw-Hill, 1937.

Weaver, W. *Lady Luck: The Theory of Probability.* New York: Dover, 1963.

Whittaker, E. T. and Robinson, G. *The Calculus of Observations: A Treatise on Numerical Mathematics, 4th ed.* New York: Dover, 1967.

Young, H. D. *Statistical Treatment of Experimental Data.* New York: McGraw-Hill, 1962.

Yule, G. U. and Kendall, M. G. *An Introduction to the Theory of Statistics, 14th ed., rev. and enl.* New York: Hafner, 1950.

Staudt-Clausen Theorem

see VON STAUDT-CLAUSEN THEOREM

Steenrod Algebra

The Steenrod algebra has to do with the COHOMOLOGY operations in singular COHOMOLOGY with INTEGER mod 2 COEFFICIENTS. For every $n \in \mathbb{Z}$ and $i \in \{0,1,2,3,\ldots\}$ there are natural transformations of FUNCTORS

$$Sq^i : H^n(\bullet;\mathbb{Z}_2) \to H^{n+i}(\bullet;\mathbb{Z}_2)$$

satisfying:

1. $Sq^i = 0$ for $i > n$.
2. $Sq^n(x) = x \smile x$ for all $x \in H^n(X,A;\mathbb{Z}_2)$ and all pairs (X,A).
3. $Sq^0 = id_{H^n(\bullet;\mathbb{Z}_2)}$.
4. The Sq^i maps commute with the coboundary maps in the long exact sequence of a pair. In other words,

   $$Sq^i : H^*(\bullet;\mathbb{Z}_2) \to H^{*+i}(\bullet;\mathbb{Z}_2)$$

   is a degree i transformation of cohomology theories.
5. (CARTAN RELATION)

   $$Sq^i(x \smile y) = \Sigma_{j+k=i} Sq^j(x) \smile Sq^k(y).$$

6. (ADEM RELATIONS) For $i < 2j$,

   $$Sq^i \circ Sq^j(x) = \Sigma_{k=0}^{\lfloor i \rfloor} \binom{j-k-1}{i-2k} Sq^{i+j-k} \circ Sq^k(x).$$

7. $Sq^i \circ \Sigma = \Sigma \circ Sq^i$ where Σ is the cohomology suspension isomorphism.

The existence of these cohomology operations endows the cohomology ring with the structure of a MODULE over the Steenrod algebra $\mathcal{A}$, defined to be $T(F_{\mathbb{Z}_2}\{Sq^i : i \in \{0,1,2,3,\ldots\}\})/R$, where $F_{\mathbb{Z}_2}(\bullet)$ is the free module functor that takes any set and sends it to the free $\mathbb{Z}_2$ module over that set. We think of $F_{\mathbb{Z}_2}\{Sq^i : i \in \{0,1,2,\ldots\}\}$ as being a graded $\mathbb{Z}_2$ module, where the i-th gradation is given by $\mathbb{Z}_2 \cdot Sq^i$. This makes the tensor algebra $T(F_{\mathbb{Z}_2}\{Sq^i : i \in \{0,1,2,3,\ldots\}\})$ into a GRADED ALGEBRA over $\mathbb{Z}_2$. R is the IDEAL generated by the elements $Sq^i Sq^j + \Sigma_{k=0}^{\lfloor i \rfloor} \binom{j-k-1}{i-2k} Sq^{i+j-k} Sq^k$ and

$1 + Sq^0$ for $0 < i < 2j$. This makes $\mathcal{A}$ into a graded $\mathbb{Z}_2$ algebra.

By the definition of the Steenrod algebra, for any SPACE (X, A), $H^*(X, A; \mathbb{Z}_2)$ is a MODULE over the Steenrod algebra $\mathcal{A}$, with multiplication induced by $Sq^i \cdot x \equiv Sq^i(x)$. With the above definitions, cohomology with COEFFICIENTS in the RING $\mathbb{Z}_2$, $H^*(\bullet; \mathbb{Z}_2)$ is a FUNCTOR from the category of pairs of TOPOLOGICAL SPACES to graded modules over $\mathcal{A}$.

see also ADEM RELATIONS, CARTAN RELATION, COHOMOLOGY, GRADED ALGEBRA, IDEAL, MODULE, TOPOLOGICAL SPACE

Steenrod-Eilenberg Axioms

see EILENBERG-STEENROD AXIOMS

Steenrod's Realization Problem

When can homology classes be realized as the image of fundamental classes of MANIFOLDS? The answer is known, and singular BORDISM GROUPS provide insight into this problem.

see also BORDISM GROUP, MANIFOLD

Steepest Descent Method

An ALGORITHM for calculating the GRADIENT $\nabla f(\mathbf{P})$ of a function at an n-D point $\mathbf{P}$. The steepest descent method starts at a point $\mathbf{P}_0$ and, as many times as needed, moves from $\mathbf{P}_i$ to $\mathbf{P}_{i+1}$ by minimizing along the line extending from $\mathbf{P}_i$ in the direction of $-\nabla f(\mathbf{P}_i)$, the local downhill gradient. This method has the severe drawback of requiring a great many iterations for functions which have long, narrow valley structures. In such cases, a CONJUGATE GRADIENT METHOD is preferable.

see also CONJUGATE GRADIENT METHOD, GRADIENT

References

Arfken, G. "The Method of Steepest Descents." §7.4 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 428–436, 1985.

Menzel, D. (Ed.). *Fundamental Formulas of Physics, Vol. 2, 2nd ed.* New York: Dover, p. 80, 1960.

Morse, P. M. and Feshbach, H. "Asymptotic Series; Method of Steepest Descent." §4.6 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 434–443, 1953.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 414, 1992.

Steffenson's Formula

$$\begin{aligned} f_p = f_0 &+ \tfrac{1}{2}p(p+1)\delta_{1/2} - \tfrac{1}{2}(p-1)p\delta_{-1/2} \\ &+ (S_3 + S_4)\delta^3_{1/2} + (S_3 - S_4)\delta^3_{-1/2} + \ldots, \quad (1) \end{aligned}$$

for $p \in [-\frac{1}{2}, \frac{1}{2}]$, where δ is the CENTRAL DIFFERENCE and

$$S_{2n+1} = \frac{1}{2}\binom{p+n}{2n+1} \quad (2)$$

$$S_{2n+2} = \frac{p}{2n+2}\binom{p+n}{2n+1} \quad (3)$$

$$S_{2n+1} - S_{2n+2} = \binom{p+n+1}{2n+2} \quad (4)$$

$$S_{2n+1} - S_{2n+2} = -\binom{p+n}{2n+2}, \quad (5)$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT.

see also CENTRAL DIFFERENCE, STIRLING'S FINITE DIFFERENCE FORMULA

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 433, 1987.

Steffensen's Inequality

Let $f(x)$ be a NONNEGATIVE and monotonic decreasing function in $[a, b]$ and $g(x)$ satisfy such that $0 \leq g(x) \leq 1$ in $[a, b]$, then

$$\int_{b-k}^{b} f(x)\,dx \leq \int_a^b f(x)g(x)\,dx \leq \int_a^{a+k} f(x)\,dx,$$

where

$$k = \int_a^b g(x)\,dx.$$

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1099, 1979.

Steinbach Screw

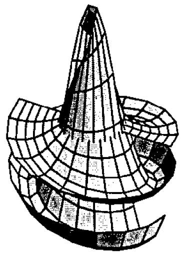

A SURFACE generated by the parametric equations

$$x(u, v) = u \cos v$$

$$y(u, v) = u \sin v$$

$$z(u, v) = v \cos u.$$

The above image uses $u \in [-4, 4]$ and $v \in [0, 6.25]$.

References

Naylor, B. "Steinbach Screw 1." `http://www.garlic.com/~bnaylor/rtstein1.html`.

Pickover, C. A. *Mazes for the Mind: Computers and the Unexpected.* New York: St. Martin's Press, 1992.

Wang, P. "Renderings." `http://www.ugcs.caltech.edu/~peterw/portfolio/renderings/`.

Steiner Chain

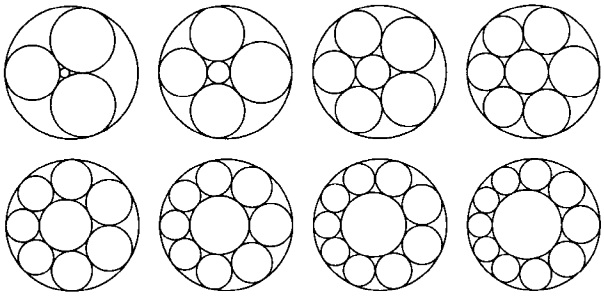

Given two nonconcentric CIRCLES with one interior to the other, if small TANGENT CIRCLES can be inscribed around the region between the two CIRCLES such that the final CIRCLE is TANGENT to the first, the CIRCLES form a Steiner chain.

The simplest way to construct a Steiner chain is to perform an INVERSION on a symmetrical arrangement on n circles packed between a central circle of radius b and an outer concentric circle of radius a. In this arrangement,

$$\sin\left(\frac{\pi}{n}\right)=\frac{a-b}{a+b}, \tag{1}$$

so the ratio of the radii for the small and large circles is

$$\frac{b}{a}=\frac{1-\sin\left(\frac{\pi}{n}\right)}{1+\sin\left(\frac{\pi}{n}\right)}. \tag{2}$$

To transform the symmetrical arrangement into a Steiner chain, find an INVERSION CENTER which transforms two centers initially offset by a fixed distance c to the same point. This can be done by equating

$$\frac{k^2x}{x^2-a^2}=\frac{k^2(x-c)}{(x-c)^2-b^2}, \tag{3}$$

giving the offset of the inversion center from the large circle's center as

$$x=\frac{a^2-b^2+c^2\pm\sqrt{(a^2-b^2+c^2)^2-4a^2c^2}}{2c}. \tag{4}$$

Plugging in a fixed value of a fixes b, which therefore determines x for a given c. Equivalently, a Steiner chain results whenever the INVERSIVE DISTANCE between the two original circles is given by

$$\delta=2\ln\left[\sec\left(\frac{\pi}{n}\right)+\tan\left(\frac{\pi}{n}\right)\right] \tag{5}$$

$$=2\ln\left[\tan\left(\frac{\pi}{4}+\frac{\pi}{2n}\right)\right] \tag{6}$$

(Coxeter and Greitzer 1967). The centers of the circles in a Steiner chain lie on an ELLIPSE (Ogilvy 1990, p. 57).

STEINER'S PORISM states that if a Steiner chain is formed from one starting circle, then a Steiner chain is also formed from any other starting circle.

see also ARBELOS, COXETER'S LOXODROMIC SEQUENCE OF TANGENT CIRCLES, HEXLET, PAPPUS CHAIN, STEINER'S PORISM

References
Coxeter, H. S. M. "Interlocking Rings of Spheres." *Scripta Math.* **18**, 113–121, 1952.
Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, p. 87, 1969.
Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 124–126, 1967.
Forder, H. G. *Geometry, 2nd ed.* London: Hutchinson's University Library, p. 23, 1960.
Gardner, M. "Mathematical Games: The Diverse Pleasures of Circles that Are Tangent to One Another." *Sci. Amer.* **240**, 18–28, Jan. 1979.
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 113–115, 1929.
Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 51–54, 1990.
Weisstein, E. W. "Plane Geometry." `http://www.astro.virginia.edu/~eww6n/math/notebooks/PlaneGeometry.m`.

Steiner Construction

A construction done using only a STRAIGHTEDGE. The PONCELET-STEINER THEOREM proves that all constructions possible using a COMPASS and STRAIGHTEDGE are possible using a STRAIGHTEDGE alone, as long as a fixed CIRCLE and its center, two intersecting CIRCLES without their centers, or three nonintersecting CIRCLES are drawn beforehand.

see also GEOMETRIC CONSTRUCTION, MASCHERONI CONSTRUCTION, PONCELET-STEINER THEOREM, STRAIGHTEDGE

References
Dörrie, H. "Steiner's Straight-Edge Problem." §34 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 165–170, 1965.
Steiner, J. *Geometric Constructions with a Ruler, Given a Fixed Circle with Its Center.* Translated from the first German ed. (1833). New York: Scripta Mathematica, 1950.

Steiner's Ellipse

Let $\alpha' : \beta' : \gamma'$ be the ISOTOMIC CONJUGATE POINT of a point with TRILINEAR COORDINATES $\alpha : \beta : \gamma$. The isotomic conjugate of the LINE AT INFINITY having trilinear equation

$$a\alpha+b\beta+c\gamma=0$$

is

$$\frac{\beta'\gamma'}{a}+\frac{\gamma'\alpha'}{b}+\frac{\alpha'\beta'}{c}=0,$$

known as Steiner's ellipse (Vandeghen 1965).

see also ISOTOMIC CONJUGATE POINT, LINE AT INFINITY

References
Vandeghen, A. "Some Remarks on the Isogonal and Cevian Transforms. Alignments of Remarkable Points of a Triangle." *Amer. Math. Monthly* **72**, 1091–1094, 1965.

Steiner's Hypocycloid

see DELTOID

Steiner-Lehmus Theorem

Any TRIANGLE that has two equal ANGLE BISECTORS (each measured from a VERTEX to the opposite sides) is an ISOSCELES TRIANGLE. This theorem is also called the INTERNAL BISECTORS PROBLEM and LEHMUS' THEOREM.

see also ISOSCELES TRIANGLE

References

Altshiller-Court, N. *College Geometry: A Second Course in Plane Geometry for Colleges and Normal Schools, 2nd ed., rev. enl.* New York: Barnes and Noble, p. 72, 1952.

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, p. 9, 1969.

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 14–16, 1967.

Gardner, M. *Martin Gardner's New Mathematical Diversions from Scientific American.* New York: Simon and Schuster, pp. 198–199 and 206–207, 1966.

Henderson, A. "The Lehmus-Steiner-Terquem Problem in Global Survey." *Scripta Math.* **21**, 223–232 and 309-312, 1955.

Hunter, J. A. H. and Madachy, J. S. *Mathematical Diversions.* New York: Dover, pp. 72–73, 1975.

Steiner Points

There are two different types of points known as Steiner points.

The point of CONCURRENCE of the three lines drawn through the VERTICES of a TRIANGLE PARALLEL to the corresponding sides of the first BROCARD TRIANGLE. It lies on the CIRCUMCIRCLE opposite the TARRY POINT and has TRIANGLE CENTER FUNCTION

$$\alpha = bc(a^2 - b^2)(a^2 - c^2).$$

The BRIANCHON POINT for KIEPERT'S PARABOLA is the Steiner point. The LEMOINE POINT K is the Steiner point of the first BROCARD TRIANGLE. The SIMSON LINE of the Steiner point is PARALLEL to the line OK, when O is the CIRCUMCENTER and K is the LEMOINE POINT.

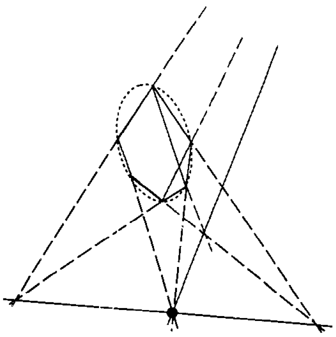

If triplets of opposites sides on a CONIC SECTION in PASCAL'S THEOREM are extended for all permutations of VERTICES, 60 PASCAL LINES are produced. The 20 points of their 3 by 3 intersections are called Steiner points.

see also BRIANCHON POINT, BROCARD TRIANGLES, CIRCUMCIRCLE, CONIC SECTION, KIEPERT'S PARABOLA, LEMOINE POINT, PASCAL LINE, PASCAL'S THEOREM, STEINER SET, STEINER TRIPLE SYSTEM, TARRY POINT

References

Casey, J. *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions, with Numerous Examples, 2nd ed., rev. enl.* Dublin: Hodges, Figgis, & Co., p. 66, 1893.

Gallatly, W. *The Modern Geometry of the Triangle, 2nd ed.* London: Hodgson, p. 102, 1913.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 281–282, 1929.

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Steiner's Porism

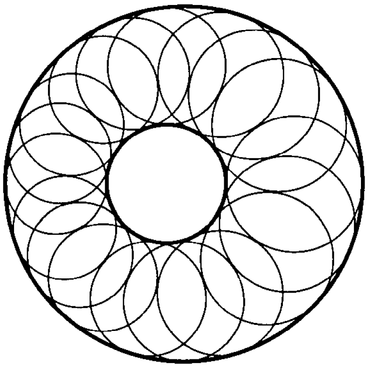

If a STEINER CHAIN is formed from one starting circle, then a STEINER CHAIN is formed from any other starting circle. In other words, given two nonconcentric CIRCLES, draw CIRCLES successively touching them and each other. If the last touches the first, this will also happen for any position of the first CIRCLE.

see also HEXLET, STEINER CHAIN

References

Coxeter, H. S. M. "Interlocking Rings of Spheres." *Scripta Math.* **18**, 113–121, 1952.

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, p. 87, 1969.

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 124–126, 1967.

Forder, H. G. *Geometry, 2nd ed.* London: Hutchinson's University Library, p. 23, 1960.

Gardner, M. "Mathematical Games: The Diverse Pleasures of Circles that Are Tangent to One Another." *Sci. Amer.* **240**, 18–28, Jan. 1979.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 113–115, 1929.

Ogilvy, C. S. *Excursions in Geometry.* New York: Dover, pp. 53–54, 1990.

Steiner's Problem

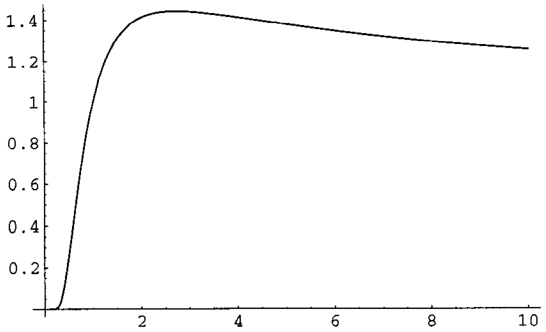

For what value of x is $f(x) = x^{1/x}$ a MAXIMUM? The maximum occurs at $x = e$, where

$$f'(x) = x^{-2+1/x}(1 - \ln x) = 0,$$

which gives a maximum of

$$e^{1/e} = 1.444667861\ldots.$$

The function has an inflection point at $x = 0.581933\ldots$, where

$$f''(x) = x^{-4+1/x}[1 - 3x + (\ln x)(2x - 2 + \ln x)] = 0.$$

see also FERMAT'S PROBLEM

Steiner Quadruple System

A Steiner quadruple system is a STEINER SYSTEM $S(t = 3, k = 4, v)$, where S is a v-set and B is a collection of k-sets of S such that every t-subset of S is contained in exactly one member of B. Barrau (1908) established the uniqueness of $S(3,4,8)$,

1	2	4	8	3	5	6	7
2	3	5	8	1	4	6	7
3	4	6	8	1	2	5	7
4	5	7	8	1	2	3	6
1	5	6	8	2	3	4	7
2	6	7	8	1	3	4	5
1	3	7	8	2	4	5	6

and $S(3,4,10)$

1	2	4	5	1	2	3	7	1	3	5	8
2	3	5	6	2	3	4	8	2	4	6	9
3	4	6	7	3	4	5	9	3	5	7	0
4	5	7	8	4	5	6	0	1	4	6	8
5	6	8	9	1	5	6	7	2	5	7	9
6	7	9	0	2	6	7	8	3	6	8	0
1	7	8	0	3	7	8	9	1	4	7	9
1	2	8	9	4	8	9	0	2	5	8	0
2	3	9	0	1	5	9	0	1	3	6	9
1	3	4	0	1	2	6	0	2	4	7	0

Fitting (1915) subsequently constructed the cyclic systems $S(3,4,26)$ and $S(3,4,34)$, and Bays and de Weck (1935) showed the existence of at least one $S(3,4,14)$. Hanani (1960) proved that a NECESSARY and SUFFICIENT condition for the existence of an $S(3,4,v)$ is that $v \equiv 2$ or $4 \pmod 6$.

The number of nonisomorphic steiner quadruple systems of orders 8, 10, 14, and 16 are 1, 1, 4 (Mendelsohn and Hung 1972), and at least 31,021 (Lindner and Rosa 1976).

see also STEINER SYSTEM, STEINER TRIPLE SYSTEM

References

Barrau, J. A. "On the Combinatory Problem of Steiner." *K. Akad. Wet. Amsterdam Proc. Sect. Sci.* **11**, 352–360, 1908.

Bays, S. and de Weck, E. "Sur les systémes de quadruples." *Comment. Math. Helv.* **7**, 222–241, 1935.

Fitting, F. "Zyklische Lösungen des Steiner'schen Problems." *Nieuw. Arch. Wisk.* **11**, 140–148, 1915.

Hanani, M. "On Quadruple Systems." *Canad. J. Math.* **12**, 145–157, 1960.

Lindner, C. L. and Rosa, A. "There are at Least 31,021 Nonisomorphic Steiner Quadruple Systems of Order 16." *Utilitas Math.* **10**, 61–64, 1976.

Lindner, C. L. and Rosa, A. "Steiner Quadruple Systems—A Survey." *Disc. Math.* **22**, 147–181, 1978.

Mendelsohn, N. S. and Hung, S. H. Y. "On the Steiner Systems $S(3,4,14)$ and $S(4,5,15)$." *Utilitas Math.* **1**, 5–95, 1972.

Steiner's Segment Problem

Given n points, find the line segments with the shortest possible total length which connect the points. The segments need not necessarily be straight from one point to another.

For three points, if all ANGLES are less than 120°, then the line segments are those connecting the three points to a central point P which makes the ANGLES $\langle A \rangle\, PB$, $\langle B \rangle\, PC$, and $\langle C \rangle\, PA$ all 120°. If one ANGLE is greater that 120°, then P coincides with the offending ANGLE.

For four points, P is the intersection of the two diagonals, but the required minimum segments are not necessarily these diagonals.

A modified version of the problem is, given two points, to find the segments with the shortest total length connecting the points such that each branch point may be connected to only three segments. There is no general solution to this version of the problem.

Steiner Set

Three sets of three LINES such that each line is incident with two from both other sets.

see also SOLOMON'S SEAL LINES, STEINER POINTS, STEINER TRIPLE SYSTEM

Steiner Surface

A projection of the Veronese Surface into 3-D (which must contain singularities) is called a Steiner surface. A classification of Steiner surfaces allowing complex parameters and projective transformations was accomplished in the 19th century. The surfaces obtained by restricting to real parameters and transformations were classified into 10 types by Coffman *et al.* (1996). Examples of Steiner surfaces include the Roman Surface (Coffman type 1) and Cross-Cap (type 3).

The Steiner surface of type 2 is given by the implicit equation

$$x^2y^2 - x^2z^2 + y^2z^2 - xyz = 0,$$

and can be transformed into the Roman Surface or Cross-Cap by a complex projective change of coordinates (but not by a real transformation). It has two pinch points and three double lines and, unlike the Roman Surface or Cross-Cap, is not compact in any affine neighborhood.

The Steiner surface of type 4 has the implicit equation

$$y^2 - 2xy^2 - xz^2 + x^2y^2 + x^2z^2 - z^4 = 0,$$

and two of the three double lines of surface 2 coincide along a line where the two noncompact "components" are tangent.

see also Cross-Cap, Roman Surface, Veronese Variety

References

Coffman, A. "Steiner Surfaces." `http://www.ipfw.edu/math/Coffman/steinersurface.html`.

Coffman, A.; Schwartz, A.; and Stanton, C. "The Algebra and Geometry of Steiner and Other Quadratically Parametrizable Surfaces." *Computer Aided Geom. Design* **13**, 257–286, 1996.

Nordstrand, T. "Steiner Relative." `http://www.uib.no/people/nfytn/stmtxt.htm`.

Nordstrand, T. "Steiner Relative [2]." `http://www.uib.no/people/nfytn/stm2txt.htm`.

Steiner System

A Steiner system is a set X of v points, and a collection of subsets of X of size k (called blocks), such that any t points of X are in exactly one of the blocks. The special case $t = 2$ and $k = 3$ corresponds to a so-called Steiner Triple System. For a Projective Plane, $v = n^2 + n + 1$, $k = n + 1$, $t = 2$, and the blocks are simply lines.

see also Steiner Quadruple System, Steiner Triple System.

References

Colbourn, C. J. and Dinitz, J. H. (Eds.) *CRC Handbook of Combinatorial Designs.* Boca Raton, FL: CRC Press, 1996.

Woolhouse, W. S. B. "Prize Question 1733." *Lady's and Gentleman's Diary.* 1844.

Steiner's Theorem

Let Lines x and y join a variable point on a Conic Section to two fixed points on the same Conic Section. Then x and y are Projectively related.

see also Conic Section, Projection

Steiner Triple System

Let X be a set of $v \geq 3$ elements together with a set B of 3-subset (triples) of X such that every 2-Subset of X occurs in exactly one triple of B. Then B is called a Steiner triple system and is a special case of a Steiner System with $t = 2$ and $k = 3$. A Steiner triple system $S(v) = S(v, k = 3, \lambda = 1)$ of order v exists Iff $v \equiv 1, 3 \pmod 6$ (Kirkman 1847). In addition, if Steiner triple systems S_1 and S_2 of orders v_1 and v_2 exist, then so does a Steiner triple system S of order v_1v_2 (Ryser 1963, p. 101).

Examples of Steiner triple systems $S(v)$ of small orders v are

$$\begin{aligned}
S_3 &= \{\{1,2,3\}\} \\
S_7 &= \{\{1,2,4\},\{2,3,5\},\{3,4,6\},\{4,5,7\}, \\
&\quad \{5,6,1\},\{6,7,2\},\{7,1,3\}\} \\
S_9 &= \{\{1,2,3\},\{4,5,6\},\{7,8,9\},\{1,4,7\}, \\
&\quad \{2,5,8\},\{3,6,9\},\{1,5,9\},\{2,6,7\}\}.
\end{aligned}$$

The number of nonisomorphic Steiner triple systems $S(v)$ of orders $v = 7, 9, 13, 15, 19, \ldots$ (i.e., $6k + 1, 3$) are 1, 1, 20, 80, $> 1.1 \times 10^9$, … (Colbourn and Dinitz 1996, pp. 14–15; Sloane's A030129). $S(7)$ is the same as the finite Projective Plane of order 2. $S(9)$ is a finite Affine Plane which can be constructed from the array

$$\begin{array}{ccc} a & b & c \\ d & e & f. \\ g & h & i \end{array}$$

One of the two $S(13)$s is a finite Hyperbolic Plane. The 80 Steiner triple systems $S(15)$ have been studied by Tonchev and Weishaar (1997). There are more than 1.1×10^9 Steiner triple systems of order 19 (Stinson and Ferch 1985; Colbourn and Dinitz 1996, p. 15).

see also Hadamard Matrix, Kirkman Triple System, Steiner Quadruple System, Steiner System

References

Colbourn, C. J. and Dinitz, J. H. (Eds.) "Steiner Triple Systems." §4.5 in *CRC Handbook of Combinatorial Designs.* Boca Raton, FL: CRC Press, pp.14–15 and 70, 1996.

Kirkman, T. P. "On a Problem in Combinatorics." *Cambridge Dublin Math. J.* **2**, 191–204, 1847.

Lindner, C. C. and Rodger, C. A. *Design Theory.* Boca Raton, FL: CRC Press, 1997.

Ryser, H. J. *Combinatorial Mathematics.* Buffalo, NY: Math. Assoc. Amer., pp. 99–102, 1963.

Sloane, N. J. A. Sequence A030129 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Stinson, D. R. and Ferch, H. "2000000 Steiner Triple Systems of Order 19." *Math. Comput.* **44**, 533–535, 1985.

Tonchev, V. D. and Weishaar, R. S. "Steiner Triple Systems of Order 15 and Their Codes." *J. Stat. Plan. Inference* **58**, 207–216, 1997.

Steinerian Curve

The LOCUS of points whose first POLARS with regard to the curves of a linear net have a common point. It is also the LOCUS of points of CONCURRENCE of line POLARS of points of the JACOBIAN CURVE. It passes through all points common to all curves of the system and is of order $3(n-1)^2$.

see also CAYLEYIAN CURVE, JACOBIAN CURVE

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 150, 1959.

Steinhaus-Moser Notation

A NOTATION for LARGE NUMBERS defined by Steinhaus (1983, pp. 28–29). In this notation, $\bigtriangleup\!\!\!\!\!{\scriptstyle n}$ denotes n^n, $\boxed{n}$ denotes "n in n TRIANGLES," and ⓝ denotes "n in n SQUARES." A modified version due to Moser eliminates the circle notation, continuing instead with POLYGONS of ever increasing size, so n in a PENTAGON is n with n SQUARES around it, etc.

see also CIRCLE NOTATION, LARGE NUMBER, MEGA, MOSER

References

Steinhaus, H. *Mathematical Snapshots, 3rd American ed.* New York: Oxford University Press, 1983.

Steinitz's Theorem

A GRAPH G is the edge graph of a POLYHEDRON IFF G is a SIMPLE, PLANAR GRAPH which is 3-connected.

see also PLANAR GRAPH, SIMPLE GRAPH

Steinmetz Solid

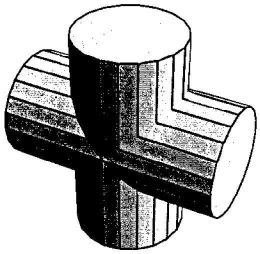

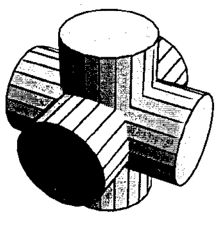

The solid common to two (or three) right circular CYLINDERS of equal RADII intersecting at RIGHT ANGLES is called the Steinmetz solid. (Two CYLINDERS intersecting at RIGHT ANGLES are sometimes called a BICYLINDER, and three intersecting CYLINDERS a TRICYLINDER.)

The VOLUME common to two intersecting right CYLINDERS of RADIUS r is

$$V_2(r,r) = \tfrac{16}{3}r^3. \qquad (1)$$

If the two right CYLINDERS are of *different* RADII a and b with $a > b$, then the VOLUME common to them is

$$V_2(a,b) = \tfrac{8}{3}a[(a^2+b^2)E(k) - (a^2-b^2)K(k)], \qquad (2)$$

where $K(k)$ is the complete ELLIPTIC INTEGRAL OF THE FIRST KIND, $E(k)$ is the complete ELLIPTIC INTEGRAL OF THE SECOND KIND, and $k \equiv b/a$ is the MODULUS.

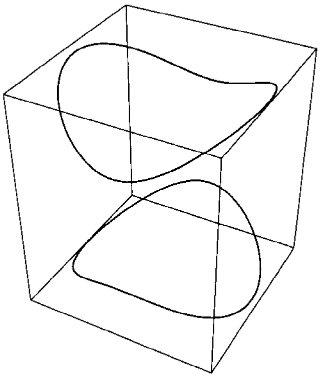

The curves of intersection of two cylinders of RADII a and b, shown above, are given by the parametric equations

$$x(t) = a\cos t \qquad (3)$$

$$y(t) = a\sin t \qquad (4)$$

$$z(t) = \pm\sqrt{b^2 - a^2\sin^2 t} \qquad (5)$$

(Gray 1993).

The VOLUME common to two ELLIPTIC CYLINDERS

$$\frac{x^2}{a^2} + \frac{z^2}{c^2} = 1 \qquad \frac{y^2}{b^2} + \frac{z^2}{c'^2} = 1 \qquad (6)$$

with $c < c'$ is

$$V_2(a,c;b,c') = \frac{8ab}{3c}[(c'^2+c^2)E(k) - (c'^2-c^2)K(k)], \quad (7)$$

where $k = c/c'$ (Bowman 1961, p. 34).

For three CYLINDERS of RADII r intersecting at RIGHT ANGLES, the VOLUME of intersection is

$$V_3(r,r,r) = 8(2-\sqrt{2})r^3. \qquad (8)$$

see also BICYLINDER, CYLINDER

References

Bowman, F. *Introduction to Elliptic Functions, with Applications.* New York: Dover, 1961.

Gardner, M. *The Unexpected Hanging and Other Mathematical Diversions.* Chicago, IL: Chicago University Press, pp. 183–185, 1991.

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 149–150, 1993.

Wells, D. G. #555 in *The Penguin Book of Curious and Interesting Puzzles.* London: Penguin Books, 1992.

Stella Octangula

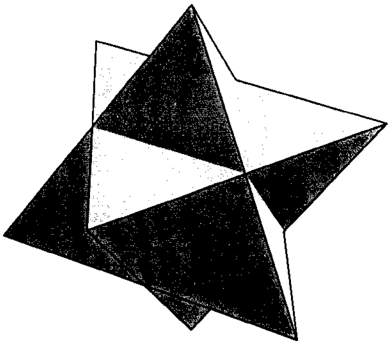

A POLYHEDRON COMPOUND composed of a TETRAHEDRON and its RECIPROCAL (a second TETRAHEDRON rotated 180° with respect to the first). The stella octangula is also called a STELLATED TETRAHEDRON. It can be constructed using the following NET by cutting along the solid lines, folding back along the plain lines, and folding forward along the dotted lines.

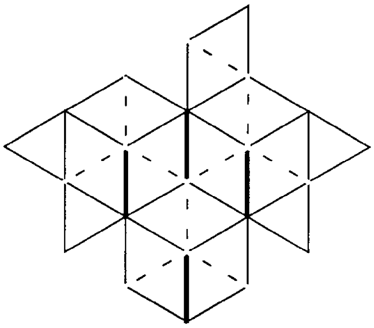

Another construction builds a single TETRAHEDRON, then attaches four tetrahedral caps, one to each face.

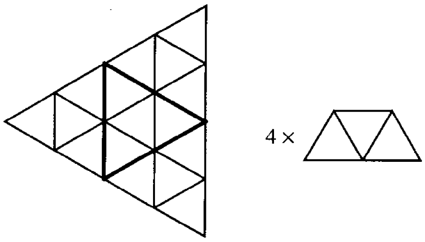

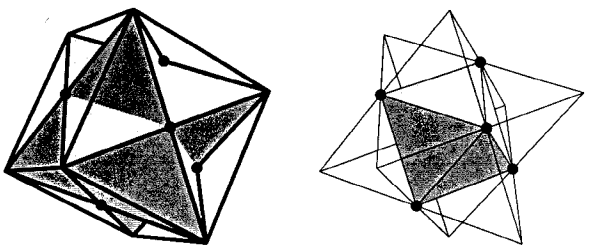

The edges of the two tetrahedra form the 12 DIAGONALS of a CUBE. The solid common to both tetrahedra is an OCTAHEDRON (Ball and Coxeter 1987).

see also CUBE, OCTAHEDRON, POLYHEDRON COMPOUND, TETRAHEDRON

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 135–137, 1987.

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, p. 158, 1969.

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 129, 1989.

Stella Octangula Number

A FIGURATE NUMBER of the form,

$$StOct_n = O_n + 8T_{n-1} = n(2n^2 - 1).$$

The first few are 1, 14, 51, 124, 245, ... (Sloane's A007588). The GENERATING FUNCTION for the stella octangula numbers is

$$\frac{x(x^2 + 10x + 1)}{(x-1)^4} = x + 14x^2 + 51x^3 + 124x^4 + \ldots.$$

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 51, 1996.

Sloane, N. J. A. Sequence A007588/M4932 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Stellated Polyhedron

A convex regular POLYHEDRON. Stellated polyhedra include the KEPLER-POINSOT SOLIDS, which consist of three DODECAHEDRON STELLATIONS and one of the ICOSAHEDRON STELLATIONS. Coxeter (1982) shows that 59 ICOSAHEDRON STELLATIONS exist. The CUBE and the TETRAHEDRON cannot be stellated. The OCTAHEDRON has only one stellation, the STELLA OCTANGULA which is a compound of two TETRAHEDRA.

There are therefore a total of $3 + 1 + (59 - 1) + 1 = 63$ stellated POLYHEDRA, although some are COMPOUND POLYHEDRA and therefore not UNIFORM POLYHEDRA. The set of all possible EDGES of the stellations can be obtained by finding all intersections on the facial planes.

see also ARCHIMEDEAN SOLID STELLATION, DODECAHEDRON STELLATIONS, ICOSAHEDRON STELLATIONS, KEPLER-POINSOT SOLID, POLYHEDRON, STELLA OCTANGULA, STELLATED TRUNCATED HEXAHEDRON, STELLATION, UNIFORM POLYHEDRON

References

Coxeter, H. S. M. *The Fifty-Nine Icosahedra.* New York: Springer-Verlag, 1982.

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Publications, 1989.

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: University Press, 1974.

Stellated Tetrahedron

see STELLA OCTANGULA

Stellated Truncated Hexahedron

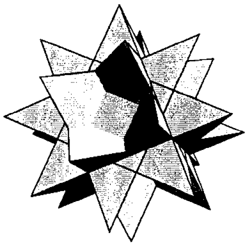

The UNIFORM POLYHEDRON U_{19}, also called the QUASITRUNCATED HEXAHEDRON, whose DUAL POLYHEDRON is the GREAT TRIAKIS OCTAHEDRON. It has SCHLÄFLI SYMBOL $t'\{4,3\}$ and WYTHOFF SYMBOL $2\,3\,|\,\frac{4}{3}$. Its faces are $8\{3\}+6\{\frac{8}{3}\}$. For $a=1$, its CIRCUMRADIUS is

$$R=\tfrac{1}{2}\sqrt{7-4\sqrt{2}}.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 144, 1989.

Stellation

The process of constructing POLYHEDRA by extending the facial PLANES past the EDGES of a given POLYHEDRON.

see also ARCHIMEDEAN SOLID STELLATION, DODECAHEDRON STELLATIONS, FACETING, ICOSAHEDRON STELLATIONS, KEPLER-POINSOT SOLID, POLYHEDRON, STELLA OCTANGULA, STELLATED POLYHEDRON, STELLATED TRUNCATED HEXAHEDRON, STELLATION TRUNCATION, UNIFORM POLYHEDRON

References

Fleurent, G. M. "Symmetry and Polyhedral Stellation Ia and Ib. Symmetry 2: Unifying Human Understanding, Part 1." *Comput. Math. Appl.* **17**, 167–193, 1989.

Messer, P. W. "Les étoilements du rhombitricontaèdre et plus." *Structural Topology* **21**, 25–46, 1995.

Messer, P. W. and Wenninger, M. J. "Symmetry and Polyhedral Stellation. II. Symmetry 2: Unifying Human Understanding, Part 1." *Comput. Math. Appl.* **17**, 195–201, 1989.

Stem-and-Leaf Diagram

The "stem" is a column of the data with the last digit removed. The final digits of each column are placed next to each other in a row next to the appropriate column. Then each row is sorted in numerical order. This diagram was invented by John Tukey.

References

Tukey, J. W. *Explanatory Data Analysis.* Reading, MA: Addison-Wesley, pp. 7–16, 1977.

Step

1.5 times the H-SPREAD.

see also FENCE, H-SPREAD

References

Tukey, J. W. *Explanatory Data Analysis.* Reading, MA: Addison-Wesley, p. 44, 1977.

Step Function

A function on the REALS $\mathbb{R}$ is a step function if it can be written as a finite linear combination of semi-open intervals $[a,b)\subseteq\mathbb{R}$. Therefore, a step function f can be written as

$$f(x)=\alpha_1 f_1(x)+\cdots+\alpha_n f_n(x),$$

where $\alpha_i\in\mathbb{R}$, $f_i(x)=1$ if $x\in[a_i,b_i)$ and 0 otherwise, for $i=1,\ldots,n$.

see also HEAVISIDE STEP FUNCTION

Step Polynomial

see HERMITE'S INTERPOLATING FUNDAMENTAL POLYNOMIAL

Steradian

The unit of SOLID ANGLE. The SOLID ANGLE corresponding to all of space being subtended is 4π steradian.

see also RADIAN, SOLID ANGLE

Stereogram

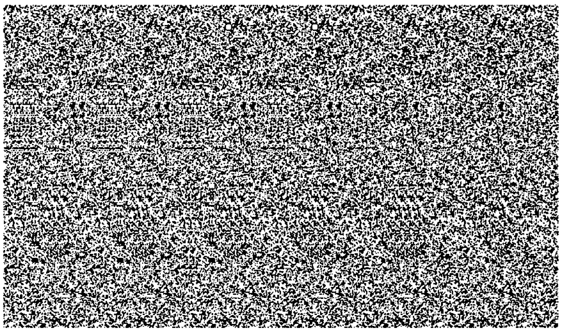

A plane image or pair of 2-D images which, when appropriately viewed using both eyes, produces an image which appears to be three-dimensional. By taking a pair of photographs from slightly different angles and then allowing one eye to view each image, a stereogram is not difficult to produce.

Amazingly, it turns out that the 3-D effect can be produced by both eyes looking at a *single* image by defocusing the eyes at a certain distance. Such stereograms are called "random-dot stereograms."

References

Bar-Natan, D. "Random-Dot Stereograms." *Math. J.* **1**, 69–71, 1991.

Fineman, M. *The Nature of Visual Illusion.* New York: Dover, pp. 89–93, 1996.

Julesz, B. *Foundations of Cyclopean Perception.* Chicago, IL: University of Chicago Press, 1971.

Julesz, B. "Stereoscopic Vision." *Vision Res.* **26**, 1601–1611, 1986.

Terrell, M. S. and Terrell, R. E. "Behind the Scenes of a Random Dot Stereogram." *Amer. Math. Monthly* **101**, 715–724, 1994.

Tyler, C. "Sensory Processing of Binocular Disparity." In *Vergence Eye Movements: Basic and Clinical Aspects.* Boston, MA: Butterworth, pp. 199–295, 1983.

Stereographic Projection

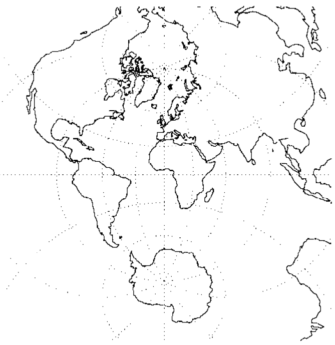

A MAP PROJECTION in which GREAT CIRCLES are CIRCLES and LOXODROMES are LOGARITHMIC SPIRALS.

$$x = k\cos\phi\sin(\lambda - \lambda_0) \tag{1}$$
$$y = k[\cos\phi_1\sin\phi - \sin\phi_1\cos\phi\cos(\lambda - \lambda_0)], \tag{2}$$

where

$$k = \frac{2}{1 + \sin\phi_1\sin\phi + \cos\phi_1\cos\phi\cos(\lambda - \lambda_0)}. \tag{3}$$

The inverse FORMULAS are given by

$$\phi = \sin^{-1}\left(\cos c\sin\phi_1 + \frac{y\sin c\cos\phi_1}{\rho}\right) \tag{4}$$
$$\lambda = \lambda_0 + \tan^{-1}\left(\frac{x\sin c}{\rho\cos\phi_1\cos c - y\sin\phi_1\sin c}\right), \tag{5}$$

where

$$\rho = \sqrt{x^2 + y^2} \tag{6}$$
$$c = 2\tan^{-1}(\tfrac{1}{2}\rho). \tag{7}$$

see also GALL'S STEREOGRAPHIC PROJECTION

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 150–153, 1967.

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 154–163, 1987.

Stereology

The exploration of 3-D space from 2-D sections of PROJECTIONS of solid bodies.

see also AXONOMETRY, CORK PLUG, CROSS-SECTION, PROJECTION, TRIP-LET

Stern-Brocot Tree

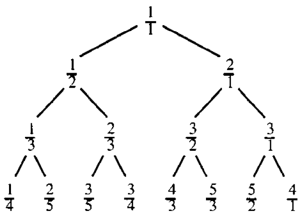

A special type of BINARY TREE obtained by starting with the fractions $\frac{0}{1}$ and $\frac{1}{0}$ and iteratively inserting $(m+m')/(n+n')$ between each two adjacent fractions m/n and m'/n'. The result can be arranged in tree form as illustrated above. The FAREY SEQUENCE F_n defines a subtree of the Stern-Brocot tree obtained by pruning off unwanted branches (Vardi 1991, Graham *et al.* 1994).

see also BINARY TREE, FAREY SEQUENCE, FORD CIRCLE

References

Brocot, A. "Calcul des rouages par approximation, nouvelle méthode." *Revue Chonométrique* **6**, 186–194, 1860.

Graham, R. L.; Knuth, D. E.; and Patashnik, O. *Concrete Mathematics: A Foundation for Computer Science, 2nd ed.* Reading, MA: Addison-Wesley, pp. 116–117, 1994.

Stern, M. A. "Über eine zahlentheoretische Funktion." *J. reine angew. Math.* **55**, 193–220, 1858.

Vardi, I. *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, p. 253, 1991.

Stevedore's Knot

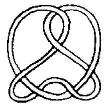

The 6-crossing KNOT 06_{001} having CONWAY-ALEXANDER POLYNOMIAL

$$\Delta(t) = 2t^2 - 5t + 2.$$

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 225, 1976.

Stewart's Theorem

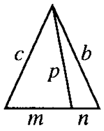

$$a(p^2 + mn) = b^2m + c^2n,$$

where

$$a \equiv m + n.$$

References

Altshiller-Court, N. "Stewart's Theorem." §6B in *College Geometry: A Second Course in Plane Geometry for Colleges and Normal Schools, 2nd ed., rev. enl.* New York: Barnes and Noble, pp. 152–153, 1952.

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., p. 6, 1967.

Stick Number

Let the stick number $s(K)$ of a KNOT K be the least number of straight sticks needed to make a KNOT K. The smallest stick number of any KNOT is $s(T) = 6$, where T is the TREFOIL KNOT. If J and K are KNOTS, then

$$s(J + K) \leq s(J) + s(K) + 1.$$

For a nontrivial KNOT K, let $c(K)$ be the CROSSING NUMBER (i.e., the least number of crossings in any projection of K). Then

$$\tfrac{1}{2}[5+\sqrt{25+8(c(K)-2)}] \leq s(K) \leq 2c(K).$$

The following table gives the stick number for some common knots.

Knot	s
trefoil knot	6
Whitehead link	8

see also CROSSING NUMBER (LINK), TRIANGLE COUNTING

References
Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 27–30, 1994.

Stickelberger Relation

Let P be a PRIME IDEAL in D_m not containing m. Then

$$(\Phi(P)) = P^{\sum t\sigma_t^{-1}},$$

where the sum is over all $1 \leq t < m$ which are RELATIVELY PRIME to m. Here D_m is the RING of integers in $\mathbb{Q}(\zeta_m)$, $\Phi(P) = g(P)^m$, and other quantities are defined by Ireland and Rosen (1990).

see also PRIME IDEAL

References
Ireland, K. and Rosen, M. "The Stickelberger Relation and the Eisenstein Reciprocity Law." Ch. 14 in *A Classical Introduction to Modern Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 203–227, 1990.

Stiefel Manifold

The Stiefel manifold of ORTHONORMAL k-frames in $\mathbb{R}^n$ is the collection of vectors $(v_1, \ldots, v_k)$ where v_i is in $\mathbb{R}^n$ for all i, and the k-tuple $(v_1, \ldots, v_k)$ is ORTHONORMAL. This is a submanifold of $\mathbb{R}^{nk}$, having DIMENSION $nk - (k+1)k/2$.

Sometimes the "orthonormal" condition is dropped in favor of the mildly weaker condition that the k-tuple $(v_1, \ldots, v_k)$ is linearly independent. Usually, this does not affect the applications since Stiefel manifolds are usually considered only during HOMOTOPY THEORETIC considerations. With respect to HOMOTOPY THEORY, the two definitions are more or less equivalent since GRAM-SCHMIDT ORTHONORMALIZATION gives rise to a smooth deformation retraction of the second type of Stiefel manifold onto the first.

see also GRASSMANN MANIFOLD

Stiefel-Whitney Class

The ith Stiefel-Whitney class of a REAL VECTOR BUNDLE (or TANGENT BUNDLE or a REAL MANIFOLD) is in the ith cohomology group of the base SPACE involved. It is an OBSTRUCTION to the existence of $(n-i+1)$ REAL linearly independent VECTOR FIELDS on that VECTOR BUNDLE, where n is the dimension of the FIBER. Here, OBSTRUCTION means that the ith Stiefel-Whitney class being NONZERO implies that there do *not* exist $(n-i+1)$ everywhere linearly dependent VECTOR FIELDS (although the Stiefel-Whitney classes are not always the OBSTRUCTION).

In particular, the nth Stiefel-Whitney class is the obstruction to the existence of an everywhere NONZERO VECTOR FIELD, and the first Stiefel-Whitney class of a MANIFOLD is the obstruction to orientability.

see also CHERN CLASS, OBSTRUCTION, PONTRYAGIN CLASS, STIEFEL-WHITNEY NUMBER

Stiefel-Whitney Number

The Stiefel-Whitney number is defined in terms of the STIEFEL-WHITNEY CLASS of a MANIFOLD as follows. For any collection of STIEFEL-WHITNEY CLASSES such that their cup product has the same DIMENSION as the MANIFOLD, this cup product can be evaluated on the MANIFOLD's FUNDAMENTAL CLASS. The resulting number is called the PONTRYAGIN NUMBER for that combination of Pontryagin classes.

The most important aspect of Stiefel-Whitney numbers is that they are COBORDISM invariant. Together, PONTRYAGIN and Stiefel-Whitney numbers determine an oriented MANIFOLD's COBORDISM class.

see also CHERN NUMBER, PONTRYAGIN NUMBER, STIEFEL-WHITNEY CLASS

Stieltjes Constants

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Expanding the RIEMANN ZETA FUNCTION about $z = 1$ gives

$$\zeta(z) = \frac{1}{z-1} + \sum_{n=0}^{\infty} \frac{(-1)^n}{n!}\gamma_n(z-1)^n, \qquad (1)$$

where

$$\gamma_n \equiv \lim_{m\to\infty}\left[\sum_{k=1}^{m}\frac{(\ln k)^n}{k} - \frac{(\ln m)^{n+1}}{n+1}\right]. \qquad (2)$$

An alternative definition is given by

$$\gamma_n' \equiv \frac{(-1)^n}{n!}\gamma_n. \qquad (3)$$

The case $n = 0$ gives the EULER-MASCHERONI CONSTANT γ. The first few numerical values are given in the following table.

n	γ_n
0	0.5772156649
1	−0.07281584548
2	−0.009690363192
3	0.002053834420
4	0.002325370065
5	0.0007933238173

Briggs (1955–1956) proved that there infinitely many γ_n of each SIGN. Berndt (1972) gave upper bounds of

$$|\gamma_n| < \begin{cases} \frac{4(n-1)!}{\pi^n} & \text{for } n \text{ even} \\ \frac{2(n-1)!}{\pi^n} & \text{for } n \text{ odd.} \end{cases} \tag{4}$$

Vacca (1910) proves that the EULER-MASCHERONI CONSTANT may be expressed as

$$\gamma = \sum_{k=1}^{\infty} \frac{(-1)^k}{k} \lfloor \lg k \rfloor, \tag{5}$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION. Hardy (1912) gave the FORMULA

$$\frac{2\gamma_1}{\ln 2} = \sum_{k=1}^{\infty} \frac{(-1)^k}{k} [2 \lg k - \lfloor \lg(2k) \rfloor] \lfloor \lg k \rfloor. \tag{6}$$

Kluyver (1927) gave similar series for γ_n with $n > 1$.

A set of constants related to γ_n is

$$\delta_n \equiv \lim_{m\to\infty} \left[\sum_{k=1}^{m} (\ln k)^n - \int_1^m (\ln x)^n \, dx - \tfrac{1}{2}(\ln m)^n \right] \tag{7}$$

(Sitaramachandrarao 1986, Lehmer 1988).

References

Berndt, B. C. "On the Hurwitz Zeta-Function." *Rocky Mountain J. Math.* **2**, 151–157, 1972.

Bohman, J. and Fröberg, C.-E. "The Stieltjes Function—Definitions and Properties." *Math. Comput.* **51**, 281–289, 1988.

Briggs, W. E. "Some Constants Associated with the Riemann Zeta-Function." *Mich. Math. J.* **3**, 117–121, 1955–1956.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/stltjs/stltjs.html`.

Hardy, G. H. "Note on Dr. Vacca's Series for γ." *Quart. J. Pure Appl. Math.* **43**, 215-216, 1912.

Kluyver, J. C. "On Certain Series of Mr. Hardy." *Quart. J. Pure Appl. Math.* **50**, 185–192, 1927.

Knopfmacher, J. "Generalised Euler Constants." *Proc. Edinburgh Math. Soc.* **21**, 25–32, 1978.

Lehmer, D. H. "The Sum of Like Powers of the Zeros of the Riemann Zeta Function." *Math. Comput.* **50**, 265–273, 1988.

Liang, J. J. Y. and Todd, J. "The Stieltjes Constants." *J. Res. Nat. Bur. Standards—Math. Sci.* **76B**, 161–178, 1972.

Sitaramachandrarao, R. "Maclaurin Coefficients of the Riemann Zeta Function." *Abstracts Amer. Math. Soc.* **7**, 280, 1986.

Vacca, G. "A New Series for the Eulerian Constant." *Quart. J. Pure Appl. Math.* **41**, 363–368, 1910.

Stieltjes Integral

The Stieltjes integral is a generalization of the RIEMANN INTEGRAL. Let $f(x)$ and $\alpha(x)$ be real-values bounded functions defined on a CLOSED INTERVAL $[a, b]$. Take a partition of the INTERVAL

$$a = x_0 < x_1 < x_2, \ldots < x_{n-1} < x_n = b, \tag{1}$$

and consider the Riemann sum

$$\sum_{i=0}^{n-1} f(\xi_i)[\alpha(x_{i+1}) - \alpha(x_i)] \tag{2}$$

with $\xi_i \in [x_i, x_{i+1}]$. If the sum tends to a fixed number I as $\max(x_{i+1} - x_i) \to 0$, then I is called the Stieltjes integral, or sometimes the RIEMANN-STIELTJES INTEGRAL. The Stieltjes integral of P with respect to F is denoted

$$\int P(x)\, dF(x), \tag{3}$$

where

$$\int P(x)\, dF(x) = \begin{cases} \int f(x)\, dx & \text{for } x \text{ continuous} \\ \sum_x f(x) & \text{for } x \text{ discrete.} \end{cases} \tag{4}$$

If P and F have a common point of discontinuity, then the integral does not exist. However, if the Stieltjes integral exists and F has a derivative F', then

$$\int P(x)\, dF(x) = \int P(x) F'(x)\, dx. \tag{5}$$

For enumeration of many of the integral's properties, see Dresher (1981, p. 105).

see also RIEMANN INTEGRAL

References

Dresher, M. *The Mathematics of Games of Strategy: Theory and Applications.* New York: Dover, 1981.

Hardy, G. H.; Littlewood, J. E.; and Pólya, G. *Inequalities, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 152–155, 1988.

Kestelman, H. "Riemann-Stieltjes Integration." Ch. 11 in *Modern Theories of Integration, 2nd rev. ed.* New York: Dover, pp. 247–269, 1960.

Stieltjes' Theorem

The $m + 1$ ELLIPSOIDAL HARMONICS when κ_1, κ_2, and κ_3 are given can be arranged in such a way that the rth function has $r - 1$ zeros between $-a^2$ and $-b^2$ and the remaining $m + r - 1$ zeros between $-b^2$ and $-c^2$ (Whittaker and Watson 1990).

see also ELLIPSOIDAL HARMONIC

References

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, pp. 560–562, 1990.

Stieltjes-Wigert Polynomial

Orthogonal POLYNOMIALS associated with WEIGHTING FUNCTION

$$w(x) = \pi^{-1/2} k \exp(-k^2 \ln^2 x) = \pi^{-1/2} k x^{-k^2 \ln x} \quad (1)$$

for $x \in (0, \infty)$ and $k > 0$. Using

$$\begin{bmatrix} n \\ \nu \end{bmatrix} = \frac{(1-q^n)(1-q^{n-1})\cdots(1-q^{n-\nu+1})}{(1-q)(1-q^2)\cdots(1-q^\nu)} \quad (2)$$

where $0 < \nu < n$,

$$\begin{bmatrix} n \\ 0 \end{bmatrix} = \begin{bmatrix} n \\ n \end{bmatrix} = 1, \quad (3)$$

and

$$q = \exp[-(2k^2)^{-1}]. \quad (4)$$

Then

$$p_n(x) = (-1)^n q^{n/2+1/4}[(1-q)(1-q^2)\cdots(1-q^n)]^{-1/2} \times \sum_{\nu=0}^{n} \begin{bmatrix} n \\ \nu \end{bmatrix} q^{\nu^2}(-q^{1/2}x)^\nu \quad (5)$$

for $n > 0$ and

$$p_0(x) = q^{1/4}. \quad (6)$$

References

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., p. 33, 1975.

Stirling's Approximation

Stirling's approximation gives an approximate value for the FACTORIAL function $n!$ or the GAMMA FUNCTION $\Gamma(n)$ for $n \gg 1$. The approximation can most simply be derived for n an INTEGER by approximating the sum over the terms of the FACTORIAL with an INTEGRAL, so that

$$\ln n! = \ln 1 + \ln 2 + \ldots + \ln n = \sum_{k=1}^{n} \ln k \approx \int_1^n \ln x \, dx = [x \ln x - x]_1^n = n \ln n - n + 1 \approx n \ln n - n. \quad (1)$$

The equation can also be derived using the integral definition of the FACTORIAL,

$$n! = \int_0^\infty e^{-x} x^n \, dx. \quad (2)$$

Note that the derivative of the LOGARITHM of the integrand can be written

$$\frac{d}{dx} \ln(e^{-x} x^n) = \frac{d}{dx}(n \ln x - x) = \frac{n}{x} - 1. \quad (3)$$

The integrand is sharply peaked with the contribution important only near $x = n$. Therefore, let $x \equiv n + \xi$ where $\xi \ll n$, and write

$$\ln(x^n e^{-x}) = n \ln x - x = n \ln(n + \xi) - (n + \xi). \quad (4)$$

Now,

$$\ln(n+\xi) = \ln\left[n\left(1+\frac{\xi}{n}\right)\right] = \ln n + \ln\left(1+\frac{\xi}{n}\right) = \ln n + \frac{\xi}{n} - \frac{1}{2}\frac{\xi^2}{n^2} + \ldots, \quad (5)$$

so

$$\ln(x^n e^{-x}) = n \ln(n+\xi) - (n+\xi) = n \ln n + \xi - \frac{1}{2}\frac{\xi^2}{n} - n - \xi + \ldots = n \ln n - n - \frac{\xi^2}{2n} + \ldots. \quad (6)$$

Taking the EXPONENTIAL of each side then gives

$$x^n e^{-x} \approx e^{n \ln n} e^{-n} e^{-\xi^2/2n} = n^n e^{-n} e^{-\xi^2/2n}. \quad (7)$$

Plugging into the integral expression for $n!$ then gives

$$n! \approx \int_{-n}^{\infty} n^n e^{-n} e^{-\xi^2/2n} \, d\xi \approx n^n e^{-n} \int_{-\infty}^{\infty} e^{-\xi^2/2n} \, d\xi = n^n. \quad (8)$$

Evaluating the integral gives

$$n! \approx n^n e^{-n} \sqrt{2\pi n}, \quad (9)$$

$$\approx \sqrt{2\pi}\, n^{n+1/2} e^{-n}. \quad (10)$$

Taking the LOGARITHM of both sides then gives

$$\ln n! \approx n \ln n - n + \tfrac{1}{2}\ln(2\pi n) = (n + \tfrac{1}{2}) \ln n - n + \tfrac{1}{2}\ln(2\pi). \quad (11)$$

This is STIRLING'S SERIES with only the first term retained and, for large n, it reduces to Stirling's approximation

$$\ln n! \approx n \ln n - n. \quad (12)$$

Gosper notes that a better approximation to $n!$ (i.e., one which approximates the terms in STIRLING'S SERIES instead of truncating them) is given by

$$n! \approx \sqrt{(2n + \tfrac{1}{3})\pi}\, n^n e^{-n}. \quad (13)$$

This also gives a much closer approximation to the FACTORIAL of 0, $0! = 1$, yielding $\sqrt{\pi/3} \approx 1.02333$ instead of 0 obtained with the conventional Stirling approximation.

see also STIRLING'S SERIES

Stirling Cycle Number

see STIRLING NUMBER OF THE FIRST KIND

Stirling's Finite Difference Formula

$$f_p = f_0 + \tfrac{1}{2}p(\delta_{1/2} + \delta_{-1/2}) + \tfrac{1}{2}p^2\delta_0^2 + S_3(\delta_{1/2}^2 + \delta_{-1/2}^2) + S_4\delta_0^4 + \ldots,$$

for $p \in [-1/2, 1/2]$, where δ is the CENTRAL DIFFERENCE and

$$S_{2n+1} = \frac{1}{2}\binom{p+n}{2n+1}$$

$$S_{2n+2} = \frac{p}{2n+2}\binom{p+n}{2n+1},$$

with $\binom{n}{k}$ a BINOMIAL COEFFICIENT.

see also CENTRAL DIFFERENCE, STEFFENSON'S FORMULA

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 433, 1987.

Stirling's Formula

see STIRLING'S SERIES

Stirling Number of the First Kind

The definition of the (signed) Stirling number of the first kind is a number $S_n^{(m)}$ such that the number of permutations of n elements which contain exactly m CYCLES is

$$(-1)^{n-m} S_n^{(m)}. \qquad (1)$$

This means that $S_n^{(m)} = 0$ for $m > n$ and $S_n^{(n)} = 1$. The GENERATING FUNCTION is

$$x(x-1)\cdots(x-n+1) = \sum_{m=0}^{n} S_n^{(m)} x^m. \qquad (2)$$

This is the Stirling number of the first kind returned by the *Mathematica*® (Wolfram Research, Champaign, IL) command `StirlingS1[n,m]`. The triangle of signed Stirling numbers of the first kind is

$$\begin{array}{c} 1 \\ -1 \quad 1 \\ 2 \quad -3 \quad 1 \\ -6 \quad 11 \quad 6 \quad 1 \\ 24 \quad -50 \quad 35 \quad -10 \quad 1 \end{array}$$

(Sloane's A008275).

The NONNEGATIVE version simply gives the number of PERMUTATIONS of n objects having m CYCLES (with cycles in opposite directions counted as distinct) and is obtained by taking the ABSOLUTE VALUE of the signed version. The nonnegative Stirling number of the first kind is denoted $S_1(n,m) = |S_n^{(m)}|$ or $\begin{bmatrix} n \\ m \end{bmatrix}$. Diagrams illustrating $S_1(5,1) = 24$, $S_1(5,3) = 35$, $S_1(5,4) = 10$, and $S_1(5,5) = 1$ (Dickau) are shown below.

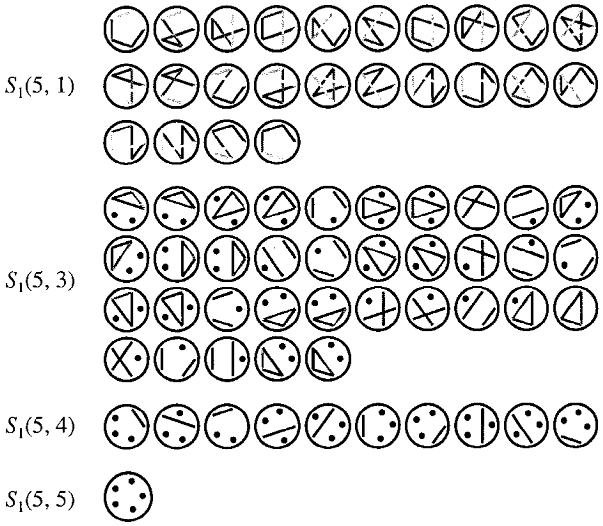

The nonnegative Stirling numbers of the first kind satisfy the curious identity

$$\sum_{n=1}^{\infty}\left[\sum_{k=0}^{n-2}\frac{(e^x - x - 1)^{k+1} S_1(n, n-k)}{(k+1)!}\right] e^{-xn} = \ln(x+1) \qquad (3)$$

(Gosper) and have the GENERATING FUNCTION

$$(1+x)(1+2x)\cdots(1+nx) = \sum_{k=1}^{n} S_1(n,m)x^k \qquad (4)$$

and satisfy

$$_1(n+1,k) = nS_1(n,k) + S_1(n,k-1). \qquad (5)$$

The Stirling numbers can be generalized to nonintegral arguments (a sort of "Stirling polynomial") using the identity

$$\frac{\Gamma(j+h)}{j^h\Gamma(j)} = \sum_{k=0}^{\infty}\frac{S_1(h,h-k)}{j^k} = 1 + \frac{(h-1)h}{2j} + \frac{(h-2)(3h-1)(h-1)h}{24j^2} + \frac{(h-3)(h-2)(h-1)^2h^2}{48j^3} + \ldots, \qquad (6)$$

which is a generalization of an ASYMPTOTIC SERIES for a ratio of GAMMA FUNCTIONS $\Gamma(j+1/2)/\Gamma(j)$ (Gosper).

see also CYCLE (PERMUTATION), HARMONIC NUMBER, PERMUTATION, STIRLING NUMBER OF THE SECOND KIND

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Stirling Numbers of the First Kind." §24.1.3 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 824, 1972.

Adamchik, V. "On Stirling Numbers and Euler Sums." *J. Comput. Appl. Math.* **79**, 119–130, 1997. `http://www.wolfram.com/~victor/articles/stirling.html`.

Conway, J. H. and Guy, R. K. In *The Book of Numbers.* New York: Springer-Verlag, pp. 91–92, 1996.

Dickau, R. M. "Stirling Numbers of the First Kind." `http://forum.swarthmore.edu/advanced/robertd/stirling1.html`.

Knuth, D. E. "Two Notes on Notation." *Amer. Math. Monthly* **99**, 403–422, 1992.

Sloane, N. J. A. Sequence A008275 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Stirling Number of the Second Kind

The number of ways of partitioning a set of n elements into m nonempty SETS (i.e., m BLOCKS), also called a STIRLING SET NUMBER. For example, the SET $\{1,2,3\}$ can be partitioned into three SUBSETS in one way: $\{\{1\},\{2\},\{3\}\}$; into two SUBSETS in three ways: $\{\{1,2\},\{3\}\}$, $\{\{1,3\},\{2\}\}$, and $\{\{1\},\{2,3\}\}$; and into one SUBSET in one way: $\{\{1,2,3\}\}$.

The Stirling numbers of the second kind are denoted $s_n^{(m)}$, $S_2(n,m)$, $s(n,m)$, or $\left\{ n \atop m \right\}$, so the Stirling numbers of the second kind for three elements are

$$s(3,1) = 1 \qquad (1)$$

$$s(3,2) = 3 \qquad (2)$$

$$s(3,3) = 1. \qquad (3)$$

Since a set of n elements can only be partitioned in a single way into 1 or n SUBSETS,

$$s(n,1) = s(n,n) = 1. \qquad (4)$$

The triangle of Stirling numbers of the second kind is

$$\begin{array}{ccccccccccc} &&&&& 1 &&&&& \\ &&&& 1 && 1 &&&& \\ &&& 1 && 3 && 1 &&& \\ && 1 && 7 && 6 && 1 && \\ & 1 && 15 && 25 && 10 && 1 & \\ 1 && 31 && 90 && 65 && 15 && 1 \end{array}$$

(Sloane's A008277).

The Stirling numbers of the second kind can be computed from the sum

$$s(n,k) = \frac{1}{k!}\sum_{i=0}^{k-1}(-1)^i\binom{k}{i}(k-i)^n, \qquad (5)$$

with $\binom{n}{k}$ a BINOMIAL COEFFICIENT, or the GENERATING FUNCTIONS

$$x^n = \sum_{m=0}^{n} s(n,m)x(x-1)\cdots(x-m+1), \qquad (6)$$

$$\sum_{n\geq k} s(n,k)\frac{x^n}{n!} = \frac{1}{k!}(e^x-1)^k, \qquad (7)$$

and

$$\frac{1}{(1-x)(1-2x)\cdots(1-kx)} = \sum_{n=1}^{k} s(n,k)x^n. \qquad (8)$$

The following diagrams (Dickau) illustrate the definition of the Stirling numbers of the second kind $s(n,m)$ for $n=3$ and 4.

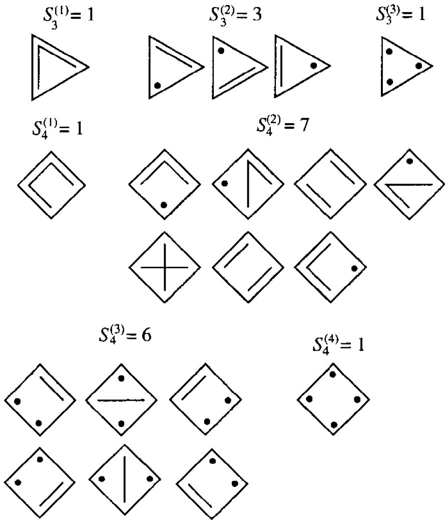

Stirling numbers of the second kind obey the RECURRENCE RELATIONS

$$s(n,k) = s(n-1,k-1) + ks(n-1,k). \qquad (9)$$

An identity involving Stirling numbers of the second kind is

$$f(m,n) \equiv \sum_{k=1}^{\infty} k^n\left(\frac{m}{m+1}\right)^l = (m+1)\sum_{k=1}^{m} k!s(n,k)m^k. \qquad (10)$$

It turns out that $f(1,n)$ can have only 0, 2, or 6 as a last DIGIT (Riskin 1995).

see also BELL NUMBER, COMBINATION LOCK, LENGYEL'S CONSTANT, MINIMAL COVER, STIRLING NUMBER OF THE FIRST KIND

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Stirling Numbers of the Second Kind." §24.1.4 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 824–825, 1972.

Comtet, L. *Advanced Combinatorics.* Boston, MA: Reidel, 1974.

Conway, J. H. and Guy, R. K. In *The Book of Numbers.* New York: Springer-Verlag, pp. 91–92, 1996.

Dickau, R. M. "Stirling Numbers of the Second Kind." `http://forum.swarthmore.edu/advanced/robertd/stirling2.html`.

Graham, R. L.; Knuth, D. E.; and Patashnik, O. *Concrete Mathematics: A Foundation for Computer Science, 2nd ed.* Reading, MA: Addison-Wesley, 1994.

Knuth, D. E. "Two Notes on Notation." *Amer. Math. Monthly* **99**, 403–422, 1992.

Riordan, J. *An Introduction to Combinatorial Analysis.* New York: Wiley, 1958.

Riordan, J. *Combinatorial Identities.* New York: Wiley, 1968.

Riskin, A. "Problem 10231." *Amer. Math. Monthly* **102**, 175–176, 1995.

Sloane, N. J. A. Sequence A008277 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Stanley, R. P. *Enumerative Combinatorics, Vol. 1.* Cambridge, England: Cambridge University Press, 1997.

Stirling's Series

The ASYMPTOTIC SERIES for the GAMMA FUNCTION is given by

$$\Gamma(z)=e^{-z}z^{z-1/2}\sqrt{2\pi}\left(1+\frac{1}{12z}+\frac{1}{288z^2}-\frac{139}{51840z^3}-\frac{571}{2488320z^4}+\ldots\right)\quad(1)$$

(Sloane's A001163 and A001164). The series for $z!$ is obtained by adding an additional factor of z,

$$z!=\Gamma(z+1)=e^{-z}z^{z+1/2}\sqrt{2\pi}\left(1+\frac{1}{12z}+\frac{1}{288z^2}-\frac{139}{51840z^3}-\frac{571}{2488320z^4}+\ldots\right).\quad(2)$$

The expansion of $\ln\Gamma(z)$ is what is usually called Stirling's series. It is given by the simple analytic expression

$$\ln\Gamma(z)=\sum_{n=1}^{\infty}\frac{B_{2n}}{2n(2n-1)z^{2n-1}}\quad(3)$$

$$=\tfrac{1}{2}\ln(2\pi)+(z+\tfrac{1}{2})\ln z-z+\frac{1}{12z}-\frac{1}{360z^3}+\frac{1}{1260z^5}-\ldots,\quad(4)$$

where B_n is a BERNOULLI NUMBER.

see also BERNOULLI NUMBER, K-FUNCTION, STIRLING'S APPROXIMATION

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 257, 1972.

Arfken, G. "Stirling's Series." §10.3 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 555–559, 1985.

Conway, J. H. and Guy, R. K. "Stirling's Formula." In *The Book of Numbers.* New York: Springer-Verlag, pp. 260–261, 1996.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 443, 1953.

Sloane, N. J. A. Sequences A001163/M5400 and A001164/M4878 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Stirling Set Number

see STIRLING NUMBER OF THE SECOND KIND

Stirrup Curve

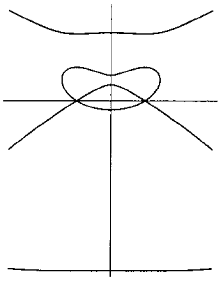

A plane curve given by the equation

$$(x^2-1)^2=y^2(y-1)(y-2)(y+5).$$

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 72, 1989.

Stochastic

see RANDOM VARIABLE

Stochastic Calculus of Variations

see MALLIAVIN CALCULUS

Stochastic Group

The GROUP of all nonsingular $n\times n$ STOCHASTIC MATRICES over a FIELD F. It is denoted $S(n,F)$. If p is PRIME and F is the GALOIS FIELD of ORDER $q=p^m$, $S(n,q)$ is written instead of $S(n,F)$. Particular examples include

$$\begin{aligned}S(2,2)&=\mathbb{Z}_2\\ S(2,3)&=S_3\\ S(2,4)&=A_4\\ S(3,2)&=S_4\\ S(2,5)&=\mathbb{Z}_4\times_\theta\mathbb{Z}_5,\end{aligned}$$

where $\mathbb{Z}_2$ is an ABELIAN GROUP, S_n are SYMMETRIC GROUPS on n elements, and $\times_\theta$ denotes the semidirect product with $\theta:\mathbb{Z}_4\to\mathrm{Aut}(\mathbb{Z}_5)$ (Poole 1995).

see also STOCHASTIC MATRIX

References

Poole, D. G. "The Stochastic Group." *Amer. Math. Monthly* **102**, 798–801, 1995.

Stochastic Matrix

A Stochastic matrix is the transition matrix for a finite MARKOV CHAIN, also called a MARKOV MATRIX. Elements of the matrix must be REAL NUMBERS in the CLOSED INTERVAL [0, 1].

A completely independent type of stochastic matrix is defined as a SQUARE MATRIX with entries in a FIELD F such that the sum of elements in each column equals 1. There are two nonsingular 2×2 STOCHASTIC MATRICES over $\mathbb{Z}_2$ (i.e., the integers mod 2),

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

There are six nonsingular stochastic 3×3 MATRICES over $\mathbb{Z}_3$,

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 2 & 0 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}.$$

In fact, the set S of all nonsingular stochastic $n \times n$ matrices over a FIELD F forms a GROUP under MATRIX MULTIPLICATION. This GROUP is called the STOCHASTIC GROUP.

see also MARKOV CHAIN, STOCHASTIC GROUP

References

Poole, D. G. "The Stochastic Group." *Amer. Math. Monthly* **102**, 798–801, 1995.

Stochastic Process

A stochastic process is a family of RANDOM VARIABLES $\{x(t, \bullet), t \in \mathcal{J}\}$ from some PROBABILITY SPACE $(S, \mathbb{S}, P)$ into a STATE SPACE $(S', \mathbb{S}')$. Here, $\mathcal{J}$ is the INDEX SET of the process.

see also INDEX SET, PROBABILITY SPACE, RANDOM VARIABLE, STATE SPACE

References

Doob, J. L. "The Development of Rigor in Mathematical Probability (1900–1950)." *Amer. Math. Monthly* **103**, 586–595, 1996.

Stochastic Resonance

A stochastic resonance is a phenomenon in which a nonlinear system is subjected to a periodic modulated signal so weak as to be normally undetectable, but it becomes detectable due to resonance between the weak deterministic signal and stochastic NOISE. The earliest definition of stochastic resonance was the maximum of the output signal strength as a function of NOISE (Bulsara and Gammaitoni 1996).

see also KRAMERS RATE, NOISE

References

Benzi, R.; Sutera, A.; and Vulpiani, A. "The Mechanism of Stochastic Resonance." *J. Phys. A* **14**, L453–L457, 1981.

Bulsara, A. R. and Gammaitoni, L. "Tuning in to Noise." *Phys. Today* **49**, 39–45, March 1996.

Stöhr Sequence

Let $a_1 = 1$ and define a_{n+1} to be the least INTEGER greater than a_n for $n \geq k$ which cannot be written as the SUM of at most h addends among the terms a_1, a_2, $\ldots$, a_n.

see also GREEDY ALGORITHM, s-ADDITIVE SEQUENCE, ULAM SEQUENCE

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 233, 1994.

Stokes Phenomenon

The asymptotic expansion of the AIRY FUNCTION $\mathrm{Ai}(z)$ (and other similar functions) has a different form in different sectors of the COMPLEX PLANE.

see also AIRY FUNCTIONS

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 609–611, 1953.

Stokes' Theorem

For w a DIFFERENTIAL $(n-1)$-FORM with compact support on an oriented n-dimensional MANIFOLD M,

$$\int_M dw = \int_{\partial M} w, \tag{1}$$

where dw is the EXTERIOR DERIVATIVE of the differential form w. This connects to the "standard" GRADIENT, CURL, and DIVERGENCE THEOREMS by the following relations. If f is a function on $\mathbb{R}^3$,

$$\mathrm{grad}(f) = c^{-1}\, df, \tag{2}$$

where $c : \mathbb{R}^3 \to \mathbb{R}^{3*}$ (the dual space) is the duality isomorphism between a VECTOR SPACE and its dual, given by the Euclidean INNER PRODUCT on $\mathbb{R}^3$. If f is a VECTOR FIELD on a $\mathbb{R}^3$,

$$\mathrm{div}(f) = {}^*d^*c(f), \tag{3}$$

where $*$ is the HODGE STAR operator. If f is a VECTOR FIELD on $\mathbb{R}^3$,

$$\mathrm{curl}(f) = c^{-1*}dc(f). \tag{4}$$

With these three identities in mind, the above Stokes' theorem in the three instances is transformed into the GRADIENT, CURL, and DIVERGENCE THEOREMS respectively as follows. If f is a function on $\mathbb{R}^3$ and γ is a curve in $\mathbb{R}^3$, then

$$\int_\gamma \mathrm{grad}(f) \cdot dl = \int_\gamma df = f(\gamma(1)) - f(\gamma(0)), \tag{5}$$

which is the GRADIENT THEOREM. If $f:\mathbb{R}^3\to\mathbb{R}^3$ is a VECTOR FIELD and M an embedded compact 3-manifold with boundary in $\mathbb{R}^3$, then

$$\int_{\partial M} f\cdot dA = \int_{\partial M} {}^*cf = \int_M d{*}cf = \int_M \operatorname{div}(f)\,dV, \quad (6)$$

which is the DIVERGENCE THEOREM. If f is a VECTOR FIELD and M is an oriented, embedded, compact 2-MANIFOLD with boundary in $\mathbb{R}^3$, then

$$\int_{\partial M} f\,dl = \int_{\partial M} cf = \int_M dc(f) = \int_M \operatorname{curl}(f)\cdot dA, \quad (7)$$

which is the CURL THEOREM.

Physicists generally refer to the CURL THEOREM

$$\int_S (\nabla\times\mathbf{F})\cdot d\mathbf{a} = \int_{\partial S} \mathbf{F}\cdot d\mathbf{s} \qquad (8)$$

as Stokes' theorem.

see also CURL THEOREM, DIVERGENCE THEOREM, GRADIENT THEOREM

Stolarsky Array

A INTERSPERSION array given by

1	2	3	5	8	13	21	34	55	⋯
4	6	10	16	26	42	68	110	178	⋯
7	11	18	29	47	76	123	199	322	⋯
9	15	24	39	63	102	165	267	432	⋯
12	19	31	50	81	131	212	343	555	⋯
14	23	37	60	97	157	254	411	665	⋯
17	28	45	73	118	191	309	500	809	⋯
20	32	52	84	136	220	356	576	932	⋯
22	36	58	94	152	246	398	644	1042	⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱

the first row of which is the FIBONACCI NUMBERS.

see also INTERSPERSION, WYTHOFF ARRAY

References

Kimberling, C. "Interspersions and Dispersions." *Proc. Amer. Math. Soc.* **117**, 313–321, 1993.

Morrison, D. R. "A Stolarsky Array and Wythoff Pairs." In *A Collection of Manuscripts Related to the Fibonacci Sequence.* Santa Clara, CA: Fibonacci Assoc., pp. 134–136, 1980.

Stolarsky-Harborth Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let $b(k)$ be the number of 1s in the BINARY expression of k. Then the number of ODD BINOMIAL COEFFICIENTS $\binom{k}{j}$ where $0\leq j\leq k$ is $2^{b(k)}$ (Glaisher 1899, Fine 1947). The number of ODD elements in the first n rows of PASCAL'S TRIANGLE is

$$f(n)=\sum_{k=0}^{n-1}2^{b(k)}. \qquad (1)$$

This function is well approximated by n^θ, where

$$\theta\equiv\frac{\ln 3}{\ln 2}=1.58496\ldots. \qquad (2)$$

Stolarsky and Harborth showed that

$$0.812556\leq\liminf_{n\to\infty}\frac{f(n)}{n^\theta} < 0.812557 < \limsup_{n\to\infty}\frac{f(n)}{n^\theta}=1. \qquad (3)$$

The value

$$SH=\liminf_{n\to\infty}\frac{f(n)}{n^\theta} \qquad (4)$$

is called the Stolarsky-Harborth constant.

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/stlrsky/stlrsky.html`.

Fine, N. J. "Binomial Coefficients Modulo a Prime." *Amer. Math. Monthly* **54**, 589–592, 1947.

Wolfram, S. "Geometry of Binomial Coefficients." *Amer. Math. Monthly* **91**, 566–571, 1984.

Stolarsky's Inequality

If $0\leq g(x)\leq 1$ and g is nonincreasing on the INTERVAL [0,1], then for all possible values of a and b,

$$\int_0^1 g(x^{1/(a+b)})\,dx \geq \int_0^1 g(x^{1/a})\,dx \int_0^1 g(x^{1/b})\,dx.$$

Stomachion

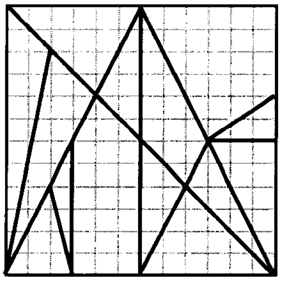

A DISSECTION game similar to TANGRAMS described in fragmentary manuscripts attributed to Archimedes and was referred to as the LOCULUS OF ARCHIMEDES (Archimedes' box) in Latin texts. The word Stomachion has as its root the Greek word for stomach. The game consisted of 14 flat pieces of various shapes arranged in the shape of a square. Like TANGRAMS, the object is to rearrange the pieces to form interesting shapes.

see also DISSECTION, TANGRAM

References
Rorres, C. "Stomachion Introduction." `http://www.mcs.drexel.edu/~crorres/Archimedes/Stomachion/intro.html`.
Rorres, C. "Stomachion Construction." `http://www.mcs.drexel.edu/~crorres/Archimedes/Stomachion/construction.html`.

Stone Space

Let $P(L)$ be the set of all PRIME IDEALS of L, and define $r(a) = \{P | a \notin P\}$. Then the Stone space of L is the TOPOLOGICAL SPACE defined on $P(L)$ by postulating that the sets of the form $r(a)$ are a subbase for the open sets.

see also PRIME IDEAL, TOPOLOGICAL SPACE

References
Grätzer, G. *Lattice Theory: First Concepts and Distributive Lattices.* San Francisco, CA: W. H. Freeman, p. 119, 1971.

Stone-von Neumann Theorem

A theorem which specifies the structure of the generic unitary representation of the Weyl relations and thus establishes the equivalence of Heisenberg's matrix mechanics and Schrödinger's wave mechanics formulations of quantum mechanics.

References
Neumann, J. von. "Die Eindeutigkeit der Schrödingerschen Operationen." *Math. Ann.* **104**, 570–578, 1931.

Stopper Knot

A KNOT used to prevent the end of a string from slipping through a hole.

References
Owen, P. *Knots.* Philadelphia, PA: Courage, p. 11, 1993.

Størmer Number

A Størmer number is a POSITIVE INTEGER n for which the largest PRIME factor p of n^2+1 is at least $2n$. Every GREGORY NUMBER t_x can be expressed uniquely as a sum of t_ns where the ns are Størmer numbers. Conway and Guy (1996) give a table of Størmer numbers reproduced below (Sloane's A005529). In a paper on INVERSE TANGENT relations, Todd (1949) gives a similar compilation.

n	p	n	p	n	p	n	p	n	p
1	2	10	101	19	181	26	617	35	613
2	5	11	61	20	401	27	73	36	1297
4	17	12	29	22	97	28	157	37	137
5	13	14	197	23	53	29	421	39	761
6	37	15	113	24	577	33	109	40	1601
9	41	16	257	25	313	34	89	42	353

see also GREGORY NUMBER, INVERSE TANGENT

References
Conway, J. H. and Guy, R. K. "Størmer's Numbers." *The Book of Numbers.* New York: Springer-Verlag, pp. 245–248, 1996.
Sloane, N. J. A. Sequence A005529/M1505 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Todd, J. "A Problem on Arc Tangent Relations." *Amer. Math. Monthly* **56**, 517–528, 1949.

Straight Angle

An ANGLE of $180° = \pi$ RADIANS.

see also DIGON, RIGHT ANGLE

Straight Line

see LINE

Straight Polyomino

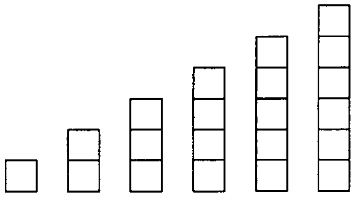

The straight polyomino of order n is the n-POLYOMINO in which all squares are placed along a line.

see also L-POLYOMINO, SKEW POLYOMINO, SQUARE POLYOMINO, T-POLYOMINO

Straightedge

An idealized mathematical object having a rigorously straight edge which can be used to draw a LINE SEGMENT. Although GEOMETRIC CONSTRUCTIONS are sometimes said to be performed with a RULER and COMPASS, the term straightedge is preferable to RULER since markings on the straightedge (usually assumed to be present on a RULER) are not allowed by the classical Greek rules.

see also COMPASS, GEOMETRIC CONSTRUCTION, GEOMETROGRAPHY, MASCHERONI CONSTANT, POLYGON, PONCELET-STEINER THEOREM, RULER, SIMPLICITY, STEINER CONSTRUCTION

Strange Attractor

An attracting set that has zero MEASURE in the embedding PHASE SPACE and has FRACTAL dimension. Trajectories within a strange attractor appear to skip around randomly.

see also CORRELATION EXPONENT, FRACTAL

References
Benmizrachi, A.; Procaccia, I.; and Grassberger, P. "Characterization of Experimental (Noisy) Strange Attractors." *Phys. Rev. A* **29**, 975–977, 1984.
Grassberger, P. "On the Hausdorff Dimension of Fractal Attractors." *J. Stat. Phys.* **26**, 173–179, 1981.
Grassberger, P. and Procaccia, I. "Measuring the Strangeness of Strange Attractors." *Physica D* **9**, 189–208, 1983a.
Grassberger, P. and Procaccia, I. "Characterization of Strange Attractors." *Phys. Rev. Let.* **50**, 346–349, 1983b.

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 137–138, 1991.
Sprott, J. C. *Strange Attractors: Creating Patterns in Chaos.* New York: Henry Holt, 1993.

Strange Loop

A phenomenon in which, whenever movement is made upwards or downwards through the levels of some heirarchial system, the system unexpectedly arrives back where it started. Hofstadter (1987) uses the strange loop as a paradigm in which to interpret paradoxes in logic (such as GRELLING'S PARADOX and RUSSELL'S PARADOX) and calls a system in which a strange loop appears a TANGLED HIERARCHY.

see also GRELLING'S PARADOX, RUSSELL'S PARADOX, TANGLED HIERARCHY

References
Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, p. 10, 1989.

Strangers

Two numbers which are RELATIVELY PRIME.

References
Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 145, 1983.

Strassen Formulas

The usual number of scalar operations (i.e., the total number of additions and multiplications) required to perform $n \times n$ MATRIX MULTIPLICATION is

$$M(n) = 2n^3 - n^2 \tag{1}$$

(i.e., n^3 multiplications and $n^3 - n^2$ additions). However, Strassen (1969) discovered how to multiply two MATRICES in

$$S(n) = 7 \cdot 7^{\lg n} - 6 \cdot 4^{\lg n} \tag{2}$$

scalar operations, where lg is the LOGARITHM to base 2, which is less than $M(n)$ for $n > 654$. For n a power of two ($n = 2^k$), the two parts of (2) can be written

$$7 \cdot 7^{\lg n} = 7 \cdot 7^{\lg 2^k} = 7 \cdot 7^k = 7 \cdot 2^{k \lg 7} = 7(2^k)^{\lg 7} = 7n^{\lg 7} \tag{3}$$

$$6 \cdot 4^{\lg n} = 6 \cdot 4^{\lg 2^k} = 6 \cdot 4^{k \lg 2} = 6 \cdot 4^k = 6(2^k)^2 = 6n^2, \tag{4}$$

so (2) becomes

$$S(2^k) = 7n^{\lg 7} - 6n^2. \tag{5}$$

Two 2×2 matrices can therefore be multiplied

$$\mathsf{C} = \mathsf{AB} \tag{6}$$

$$\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \tag{7}$$

with only

$$S(2) = 7 \cdot 2^{\lg 7} - 6 \cdot 2^2 = 49 - 24 = 25 \tag{8}$$

scalar operations (as it turns out, seven of them are multiplications and 18 are additions). Define the seven products (involving a total of 10 additions) as

$$Q_1 \equiv (a_{11} + a_{22})(b_{11} + b_{22}) \tag{9}$$
$$Q_2 \equiv (a_{21} + a_{22})b_{11} \tag{10}$$
$$Q_3 \equiv a_{11}(b_{12} - b_{22}) \tag{11}$$
$$Q_4 \equiv a_{22}(-b_{11} + b_{21}) \tag{12}$$
$$Q_5 \equiv (a_{11} + a_{12})b_{22} \tag{13}$$
$$Q_6 \equiv (-a_{11} + a_{12})(b_{11} + b_{12}) \tag{14}$$
$$Q_7 \equiv (a_{12} - a_{22})(b_{21} + b_{22}). \tag{15}$$

Then the matrix product is given using the remaining eight additions as

$$c_{11} = Q_1 + Q_4 - Q_5 + Q_7 \tag{16}$$
$$c_{21} = Q_2 + Q_4 \tag{17}$$
$$c_{12} = Q_3 + Q_5 \tag{18}$$
$$c_{22} = Q_1 + Q_3 - Q_2 + Q_6 \tag{19}$$

(Strassen 1969, Press *et al.* 1989).

Matrix inversion of a 2×2 matrix A to yield $\mathsf{C} = \mathsf{A}^{-1}$ can also be done in fewer operations than expected using the formulas

$$R_1 \equiv {a_{11}}^{-1} \tag{20}$$
$$R_2 \equiv a_{21}R_1 \tag{21}$$
$$R_3 \equiv R_1 a_{12} \tag{22}$$
$$R_4 \equiv a_{21}R_3 \tag{23}$$
$$R_5 \equiv R_4 - a_{22} \tag{24}$$
$$R_6 \equiv {R_5}^{-1} \tag{25}$$
$$c_{12} = R_3 R_6 \tag{26}$$
$$c_{21} = R_6 R_2 \tag{27}$$
$$R_7 = R_3 c_{21} \tag{28}$$
$$c_{11} = R_1 - R_7 \tag{29}$$
$$c_{22} = -R_6 \tag{30}$$

(Strassen 1969, Press *et al.* 1989). The leading exponent for Strassen's algorithm for a POWER of 2 is $\lg 7 \approx 2.808$. The best leading exponent currently known is 2.376 (Coppersmith and Winograd 1990). It has been shown that the exponent must be at least 2.

see also COMPLEX MULTIPLICATION, KARATSUBA MULTIPLICATION

References
Coppersmith, D. and Winograd, S. "Matrix Multiplication via Arithmetic Programming." *J. Symb. Comput.* **9**, 251–280, 1990.
Pan, V. *How to Multiply Matrices Faster.* New York: Springer-Verlag, 1982.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Is Matrix Inversion an N^3 Process?" §2.11 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 95–98, 1989.
Strassen, V. "Gaussian Elimination is Not Optimal." *Numerische Mathematik* **13**, 354–356, 1969.

Strassman's Theorem

Let $(K, |\cdot|)$ be a complete non-ARCHIMEDEAN VALUATED FIELD, with VALUATION RING R, and let $f(X)$ be a POWER series with COEFFICIENTS in R. Suppose at least one of the COEFFICIENTS is NONZERO (so that f is not identically zero) and the sequence of COEFFICIENTS converges to 0 with respect to $|\cdot|$. Then $f(X)$ has only finitely many zeros in R.

see also ARCHIMEDEAN VALUATION, MAHLER-LECH THEOREM, VALUATION, VALUATION RING

Strassnitzky's Formula

The MACHIN-LIKE FORMULA

$$\tfrac{1}{4}\pi = \cot^{-1} 2 + \cot^{-1} 5 + \cot^{-1} 8.$$

see also MACHIN'S FORMULA, MACHIN-LIKE FORMULAS

Strategy

A set of moves which a player plans to follow while playing a GAME.

see also GAME, MIXED STRATEGY

Stratified Manifold

A set that is a smooth embedded 2-D MANIFOLD except for a subset that consists of smooth embedded curves, except for a set of ISOLATED POINTS.

References
Morgan, F. "What is a Surface?" *Amer. Math. Monthly* **103**, 369–376, 1996.

Strehl Identity

The sum identity

$$\sum_{j=0}^{\infty} \binom{n}{j}^3 = \sum_{k=0}^{\infty} \binom{n}{k}^2 \binom{2(n-k)}{n},$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT.

see also BINOMIAL COEFFICIENT

Striction Curve

A NONCYLINDRICAL RULED SURFACE always has a parameterization of the form

$$\mathbf{x}(u,v) = \boldsymbol{\sigma}(u) + v\boldsymbol{\delta}(u), \tag{1}$$

where $|\boldsymbol{\delta}| = 1$, $\boldsymbol{\sigma}' \cdot \boldsymbol{\delta}' = 0$, and $\boldsymbol{\sigma}$ is called the striction curve of $\mathbf{x}$. Furthermore, the striction curve does not depend on the choice of the base curve. The striction and DIRECTOR CURVES of the HELICOID

$$\mathbf{x}(u,v) = \begin{bmatrix} 0 \\ 0 \\ bu \end{bmatrix} + av \begin{bmatrix} \cos u \\ \sin u \\ 0 \end{bmatrix} \tag{2}$$

are

$$\boldsymbol{\sigma}(u) = \begin{bmatrix} 0 \\ 0 \\ bu \end{bmatrix} \tag{3}$$

$$\boldsymbol{\delta}(u) = \begin{bmatrix} a\cos u \\ a\sin u \\ 0 \end{bmatrix}. \tag{4}$$

For the HYPERBOLIC PARABOLOID

$$\mathbf{x}(u,v) = \begin{bmatrix} u \\ 0 \\ 0 \end{bmatrix} + v \begin{bmatrix} 0 \\ 1 \\ u \end{bmatrix}, \tag{5}$$

the striction and DIRECTOR CURVES are

$$\boldsymbol{\sigma}(u) = \begin{bmatrix} u \\ 0 \\ 0 \end{bmatrix} \tag{6}$$

$$\boldsymbol{\delta}(u) = \begin{bmatrix} 0 \\ 1 \\ u \end{bmatrix}. \tag{7}$$

see also DIRECTOR CURVE, DISTRIBUTION PARAMETER, NONCYLINDRICAL RULED SURFACE, RULED SURFACE,

References
Gray, A. "Noncylindrical Ruled Surfaces" and "Examples of Striction Curves of Noncylindrical Ruled Surfaces." §17.3 and 17.4 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 345–350, 1993.

String Rewriting

A SUBSTITUTION MAP in which rules are used to operate on a string consisting of letters of a certain alphabet. String rewriting is a particularly useful technique for generating successive iterations of certain types of FRACTALS, such as the BOX FRACTAL, CANTOR DUST, CANTOR SQUARE FRACTAL, and SIERPIŃSKI CARPET.

see also RABBIT SEQUENCE, SUBSTITUTION MAP

References
Peitgen, H.-O. and Saupe, D. (Eds.). "String Rewriting Systems." §C.1 in *The Science of Fractal Images.* New York: Springer-Verlag, pp. 273–275, 1988.
Wagon, S. "Recursion via String Rewriting." §6.2 in *Mathematica in Action.* New York: W. H. Freeman, pp. 190–196, 1991.

Strip

see CRITICAL STRIP, MÖBIUS STRIP

Strong Convergence

Strong convergence is the type of convergence usually associated with convergence of a SEQUENCE. More formally, a SEQUENCE $\{x_n\}$ of VECTORS in an INNER PRODUCT SPACE E is called convergent to a VECTOR x in E if

$$\|x_n - x\| \to 0 \quad \text{as } n \to \infty.$$

see also CONVERGENT SEQUENCE, INNER PRODUCT SPACE, WEAK CONVERGENCE

Strong Elliptic Pseudoprime

Let n be an ELLIPTIC PSEUDOPRIME associated with (E, P), and let $n+1 = 2^s k$ with k ODD and $s \geq 0$. Then n is a strong elliptic pseudoprime when either $kP \equiv 0 \pmod{n}$ or $2^r kP \equiv 0 \pmod{n}$ for some r with $1 \leq r < s$.

see also ELLIPTIC PSEUDOPRIME

References

Ribenboim, P. *The New Book of Prime Number Records, 3rd ed.* New York: Springer-Verlag, pp. 132–134, 1996.

Strong Frobenius Pseudoprime

A PSEUDOPRIME which obeys an additional restriction beyond that required for a FROBENIUS PSEUDOPRIME. A number n with $(n, 2a) = 1$ is a strong Frobenius pseudoprime with respect to $x - a$ IFF n is a STRONG PSEUDOPRIME with respect to $f(x)$. Every strong Frobenius pseudoprime with respect to $x - a$ is an EULER PSEUDOPRIME to the base a.

Every strong Frobenius pseudoprime with respect to $f(x) = x^2 - bx - c$ such that $((b^2 + 4c)/n) = -1$ is a STRONG LUCAS PSEUDOPRIME with parameters (b, c). Every strong Frobenius pseudoprime n with respect to $x^2 - bx + 1$ is an EXTRA STRONG LUCAS PSEUDOPRIME to the base b.

see also FROBENIUS PSEUDOPRIME

References

Grantham, J. "Frobenius Pseudoprimes." 1996. `http://www.clark.net/pub/grantham/pseudo/pseudo.ps`

Strong Law of Large Numbers

For a set of random variates x_i from a distribution having unit MEAN,

$$P\left(\lim_{n\to\infty} \frac{x_1 + \ldots + x_n}{n}\right) = P\left(\lim_{n\to\infty} \langle x\rangle\right) = 1.$$

This result is due to Kolmogorov.

see also LAW OF TRULY LARGE NUMBERS, STRONG LAW OF SMALL NUMBERS, WEAK LAW OF LARGE NUMBERS

Strong Law of Small Numbers

There aren't enough small numbers to meet the many demands made of them.

References

Gardner, M. "Patterns in Primes are a Clue to the Strong Law of Small Numbers." *Sci. Amer.* **243**, 18–28, Dec. 1980.

Guy, R. K. "The Strong Law of Small Numbers." *Amer. Math. Monthly* **95**, 697–712, 1988.

Strong Lucas Pseudoprime

Let $U(P, Q)$ and $V(P, Q)$ be LUCAS SEQUENCES generated by P and Q, and define

$$D \equiv P^2 - 4Q.$$

Let n be an ODD COMPOSITE NUMBER with $(n, D) = 1$, and $n - (D/n) = 2^s d$ with d ODD and $s \geq 0$, where (a/b) is the LEGENDRE SYMBOL. If

$$U_d \equiv 0 \pmod{n}$$

or

$$V_{2^r d} \equiv 0 \pmod{n}$$

for some r with $0 \leq r < s$, then n is called a strong Lucas pseudoprime with parameters (P, Q).

A strong Lucas pseudoprime is a LUCAS PSEUDOPRIME to the same base. Arnault (1997) showed that any COMPOSITE NUMBER n is a strong Lucas pseudoprime for at most 4/15 of possible bases (unless n is the PRODUCT of TWIN PRIMES having certain properties).

see also EXTRA STRONG LUCAS PSEUDOPRIME, LUCAS PSEUDOPRIME

References

Arnault, F. "The Rabin-Monier Theorem for Lucas Pseudoprimes." *Math. Comput.* **66**, 869–881, 1997.

Ribenboim, P. "Euler-Lucas Pseudoprimes (elpsp(P, Q)) and Strong Lucas Pseudoprimes (slpsp(P, Q))." §2.X.C in *The New Book of Prime Number Records, 3rd ed.* New York: Springer-Verlag, pp. 130–131, 1996.

Strong Pseudoprime

A strong pseudoprime to a base a is an ODD COMPOSITE NUMBER n with $n - 1 = d \cdot 2^s$ (for d ODD) for which either

$$a^d \equiv 1 \pmod{n} \tag{1}$$

or

$$a^{d \cdot 2^r} \equiv -1 \pmod{n} \tag{2}$$

for some $r \in [0, s)$.

The definition is motivated by the fact that a FERMAT PSEUDOPRIME n to the base b satisfies

$$b^{n-1} - 1 \equiv 0 \pmod{n}. \tag{3}$$

But since n is ODD, it can be written $n = 2m + 1$, and

$$b^{2m} - 1 = (b^m - 1)(b^m + 1) \equiv 0 \pmod{n}. \tag{4}$$

If n is PRIME, it must DIVIDE at least one of the FACTORS, but can't DIVIDE both because it would then DIVIDE their difference

$$(b^m + 1) - (b^m - 1) = 2. \tag{5}$$

Therefore,

$$b^m \equiv \pm 1 \pmod{n}, \tag{6}$$

so write $n = 2^a t + 1$ to obtain

$$b^{n-1} - 1 = (b^t - 1)(b^t + 1)(b^{2t} + 1) \cdots (b^{2^{a-1}t} + 1). \tag{7}$$

If n DIVIDES exactly one of these FACTORS but is COMPOSITE, it is a strong pseudoprime. A COMPOSITE number is a strong pseudoprime to at most 1/4 of all bases less than itself (Monier 1980, Rabin 1980). The strong pseudoprimes provide the basis for MILLER'S PRIMALITY TEST and RABIN-MILLER STRONG PSEUDOPRIME TEST.

A strong pseudoprime to the base a is also an EULER PSEUDOPRIME to the base a (Pomerance *et al.* 1980). The strong pseudoprimes include some EULER PSEUDOPRIMES, FERMAT PSEUDOPRIMES, and CARMICHAEL NUMBERS.

There are 4842 strong psp(2) less than 2.5×10^{10}, where a psp(2) is also known as a POULET NUMBER. The strong k-pseudoprime test for $k = 2, 3, 5$ correctly identifies all PRIMES below 2.5×10^{10} with only 13 exceptions, and if 7 is added, then the only exception less than 2.5×10^{10} is 315031751. Jaeschke (1993) showed that there are only 101 strong pseudoprimes for the bases 2, 3, and 5 less than 10^{12}, nine if 7 is added, and none if 11 is added. Also, the bases 2, 13, 23, and 1662803 have no exceptions up to 10^{12}.

If n is COMPOSITE, then there is a base for which n is not a strong pseudoprime. There are therefore no "strong CARMICHAEL NUMBERS." Let ψ_k denote the smallest strong pseudoprime to all of the first k PRIMES taken as bases (i.e, the smallest ODD NUMBER for which the RABIN-MILLER STRONG PSEUDOPRIME TEST on bases less than or equal to k fails). Jaeschke (1993) computed ψ_k from $k = 5$ to 8 and gave upper bounds for $k = 9$ to 11.

$$\begin{aligned}
\psi_1 &= 2047\\
\psi_2 &= 1373653\\
\psi_3 &= 25326001\\
\psi_4 &= 3215031751\\
\psi_5 &= 2152302898747\\
\psi_6 &= 3474749660383\\
\psi_7 &= 34155071728321\\
\psi_8 &= 34155071728321\\
\psi_9 &\leq 41234316135705689041\\
\psi_{10} &\leq 1553360566073143205541002401\\
\psi_{11} &\leq 56897193526942024370326972321
\end{aligned}$$

(Sloane's A014233). A seven-step test utilizing these results (Riesel 1994) allows all numbers less than 3.4×10^{14} to be tested.

Pomerance *et al.* (1980) have proposed a test based on a combination of STRONG PSEUDOPRIMES and LUCAS PSEUDOPRIMES. They offer a \$620 reward for discovery of a COMPOSITE NUMBER which passes their test (Guy 1994, p. 28).

see also CARMICHAEL NUMBER, MILLER'S PRIMALITY TEST, POULET NUMBER, RABIN-MILLER STRONG PSEUDOPRIME TEST, ROTKIEWICZ THEOREM, STRONG ELLIPTIC PSEUDOPRIME, STRONG LUCAS PSEUDOPRIME

References

Baillie, R. and Wagstaff, S. "Lucas Pseudoprimes." *Math. Comput.* **35**, 1391–1417, 1980.

Guy, R. K. "Pseudoprimes. Euler Pseudoprimes. Strong Pseudoprimes." §A12 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 27–30, 1994.

Jaeschke, G. "On Strong Pseudoprimes to Several Bases." *Math. Comput.* **61**, 915–926, 1993.

Monier, L. "Evaluation and Comparison of Two Efficient Probabilistic Primality Testing Algorithms." *Theor. Comput. Sci.* **12**, 97–108, 1980.

Pomerance, C.; Selfridge, J. L.; and Wagstaff, S. S. Jr. "The Pseudoprimes to $25 \cdot 10^9$." *Math. Comput.* **35**, 1003–1026, 1980. Available electronically from `ftp://sable.ox.ac.uk/pub/math/primes/ps2.Z`.

Rabin, M. O. "Probabilistic Algorithm for Testing Primality." *J. Number Th.* **12**, 128–138, 1980.

Riesel, H. *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Basel: Birkhäuser, p. 92, 1994.

Sloane, N. J. A. Sequence A014233 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Strong Pseudoprime Test

see RABIN-MILLER STRONG PSEUDOPRIME TEST

Strong Subadditivity Inequality

$$\phi(A) + \phi(B) - \phi(A \cup B) \geq \phi(A \cap B).$$

References

Doob, J. L. "The Development of Rigor in Mathematical Probability (1900–1950)." *Amer. Math. Monthly* **103**, 586–595, 1996.

Strong Triangle Inequality

$$|x + y|_p \leq \max(|x|_p, |y|_p)$$

for all x and y.

see also p-ADIC NUMBER, TRIANGLE INEQUALITY

Strongly Connected Component

A maximal subgraph of a DIRECTED GRAPH such that for every pair of vertices u, v in the SUBGRAPH, there is a directed path from u to v and a directed path from v to u.

see also BI-CONNECTED COMPONENT

Strongly Embedded Theorem

The strongly embedded theorem identifies all SIMPLE GROUPS with a strongly 2-embedded SUBGROUP. In particular, it asserts that no SIMPLE GROUP has a strongly 2-embedded 2'-local SUBGROUP.

see also SIMPLE GROUP, SUBGROUP

Strongly Independent

An infinite sequence $\{a_i\}$ of POSITIVE INTEGERS is called strongly independent if any relation $\sum \epsilon_i a_i$, with $\epsilon_i = 0$, ± 1, or ± 2 and $\epsilon_i = 0$ except finitely often, IMPLIES $\epsilon_i = 0$ for all i.

see also WEAKLY INDEPENDENT

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 136, 1994.

Strongly Triple-Free Set

see TRIPLE-FREE SET

Strophoid

Let C be a curve, let O be a fixed point (the POLE), and let O' be a second fixed point. Let P and P' be points on a line through O meeting C at Q such that $P'Q = QP = QO'$. The LOCUS of P and P' is called the strophoid of C with respect to the POLE O and fixed point O'. Let C be represented parametrically by $(f(t), g(t))$, and let $O = (x_0, y_0)$ and $O' = (x_1, y_1)$. Then the equation of the strophoid is

$$x = f \pm \sqrt{\frac{(x_1 - f)^2 + (y_1 - g)^2}{1 + m^2}} \quad (1)$$

$$y = g \pm \sqrt{\frac{(x_1 - f)^2 + (y_1 - g)^2}{1 + m^2}}, \quad (2)$$

where

$$m \equiv \frac{g - y_0}{f - x_0}. \quad (3)$$

The name strophoid means "belt with a twist," and was proposed by Montucci in 1846 (MacTutor Archive). The polar form for a general strophoid is

$$r = \frac{b \sin(a - 2\theta)}{\sin(a - \theta)}. \quad (4)$$

If $a = \pi/2$, the curve is a RIGHT STROPHOID. The following table gives the strophoids of some common curves.

Curve	Pole	Fixed Point	Strophoid
line	not on line	on line	oblique strophoid
line	not on line	foot of ⊥ origin to line	right strophoid
circle	center	on circumf.	Freeth's nephroid

see also RIGHT STROPHOID

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 51–53 and 205, 1972.

Lockwood, E. H. "Strophoids." Ch. 16 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 134–137, 1967.

MacTutor History of Mathematics Archive. "Right." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Right.html`.

Yates, R. C. "Strophoid." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 217–220, 1952.

Structurally Stable

A MAP $\phi : M \to M$ where M is a MANIFOLD is C^r structurally stable if any C^r perturbation is TOPOLOGICALLY CONJUGATE to ϕ. Here, C^r perturbation means a FUNCTION ψ such that ψ is close to ϕ and the first r derivatives of ψ are close to those of ϕ.

see also TOPOLOGICALLY CONJUGATE

Structure

see LATTICE

Structure Constant

The structure constant is defined as $i\epsilon_{ijk}$, where ϵ_{ijk} is the PERMUTATION SYMBOL. The structure constant forms the starting point for the development of LIE ALGEBRA.

see also LIE ALGEBRA, PERMUTATION SYMBOL

Structure Factor

The structure factor S_Γ of a discrete set Γ is the FOURIER TRANSFORM of δ-scatterers of equal strengths on all points of Γ,

$$S_\Gamma(k) = \int \sum_{x \in \Gamma} \delta(x' - x) e^{-2\pi i k x'}\, dx' = \sum_{x \in \Gamma} e^{-2\pi i k x}.$$

References

Baake, M.; Grimm, U.; and Warrington, D. H. "Some Remarks on the Visible Points of a Lattice." *J. Phys. A: Math. General* **27**, 2669–2674, 1994.

Struve Differential Equation

The ordinary differential equation

$$z^2 y'' + z y' + (z^2 - \nu^2) y = \frac{4(\frac{1}{2}z)^{\nu+1}}{\sqrt{\pi}\,\Gamma(\nu + \frac{1}{2})},$$

where $\Gamma(z)$ is the GAMMA FUNCTION. The solution is

$$y = a J_\nu(z) + b Y_\nu(z) + \mathcal{H}_\nu(z),$$

where $J_\nu(z)$ and $Y_\nu(z)$ are BESSEL FUNCTIONS OF THE FIRST and SECOND KINDS, and $\mathcal{H}_\nu(z)$ is a STRUVE FUNCTION (Abramowitz and Stegun 1972).

see also Bessel Function of the First Kind, Bessel Function of the Second Kind, Struve Function

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 496, 1972.

Struve Function

Abramowitz and Stegun (1972, pp. 496–499) define the Struve function as

$$\mathcal{H}_\nu(z) = (\tfrac{1}{2}z)^{\nu+1} \sum_{k=0}^{\infty} \frac{(-1)^k (\frac{1}{2}z)^{2k}}{\Gamma(k+\frac{3}{2})\Gamma(k+\nu+\frac{3}{2})}, \qquad (1)$$

where $\Gamma(z)$ is the Gamma Function. Watson (1966, p. 338) defines the Struve function as

$$\mathcal{H}_\nu(z) \equiv \frac{2(\frac{1}{2}z)^\nu}{\Gamma(\nu+\frac{1}{2})\Gamma(\frac{1}{2})} \int_0^1 (1-t^2)^{\nu-1/2} \sin(zt)\,dt. \quad (2)$$

The series expansion is

$$\mathcal{H}_\nu(z) = \sum_{m=0}^{\infty} (-1)^m \frac{(\frac{1}{2}z)^{2m+\nu+1}}{\Gamma(m+\frac{3}{2})\Gamma(\nu+m+\frac{3}{2})}. \qquad (3)$$

For half integral orders,

$$\begin{aligned} &\mathcal{H}_{n+1/2}(z) \\ &\quad = Y_{n+1/2}(z) + \frac{1}{\pi} \sum_{m=0}^{n} \frac{\Gamma(m+\frac{1}{2})(\frac{1}{2}z)^{-2m+n-1/2}}{\Gamma(n+1-m)} &(4) \\ &\mathcal{H}_{-(n+1/2)}(z) = (-1)^n J_{n+1/2}(z). &(5) \end{aligned}$$

The Struve function and its derivatives satisfy

$$\mathcal{H}_{\nu-1}(z) - \mathcal{H}_{\nu+1}(z) = 2\mathcal{H}'_\nu(z) - \frac{(\frac{1}{2}z)^\nu}{\sqrt{\pi}\,\Gamma(\nu+\frac{3}{2})}. \qquad (6)$$

see also Anger Function, Bessel Function, Modified Struve Function, Weber Functions

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Struve Function $\mathbf{H}_\nu(x)$." §12.1 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 496–498, 1972.

Spanier, J. and Oldham, K. B. "The Struve Function." Ch. 57 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 563–571, 1987.

Watson, G. N. *A Treatise on the Theory of Bessel Functions, 2nd ed.* Cambridge, England: Cambridge University Press, 1966.

Student's *t*-Distribution

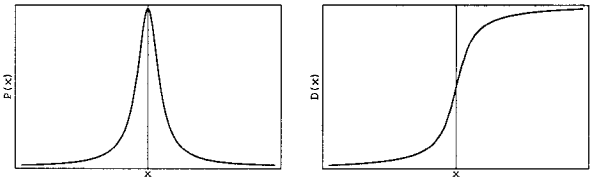

A Distribution published by William Gosset in 1908. His employer, Guinness Breweries, required him to publish under a pseudonym, so he chose "Student." Given n independent measurements x_i, let

$$t \equiv \frac{\bar{x}-\mu}{s/\sqrt{n}}, \qquad (1)$$

where μ is the population Mean, $\bar{x}$ is the sample Mean, and s is the Estimator for population Standard Deviation (i.e., the Sample Variance) defined by

$$s^2 \equiv \frac{1}{N-1} \sum_{i=1}^{n} (x_i - \bar{x})^2. \qquad (2)$$

Student's t-distribution is defined as the distribution of the random variable t which is (very loosely) the "best" that we can do not knowing σ. If $\sigma = s$, $t = z$ and the distribution becomes the Normal Distribution. As N increases, Student's t-distribution approaches the Normal Distribution.

Student's t-distribution is arrived at by transforming to Student's z-Distribution with

$$z \equiv \frac{\bar{x}-\mu}{s}. \qquad (3)$$

Then define

$$t \equiv z\sqrt{n-1}. \qquad (4)$$

The resulting probability and cumulative distribution functions are

$$\begin{aligned} f_r(t) &= \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{r\pi}\,\Gamma\left(\frac{r}{2}\right)\left(1+\frac{t^2}{r}\right)^{(r+1)/2}} \\ &= \frac{\left(\frac{r}{r+t^2}\right)^{(1+r)/2}}{\sqrt{r}\,B(\frac{1}{2}r,\frac{1}{2})} &(5) \\ F_r(t) &= \int_{-\infty}^{t} \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{r\pi}\,\Gamma\left(\frac{r}{2}\right)\left(1+\frac{t^2}{r}\right)^{(r+1)/2}}\,dt \\ &= \frac{1}{\sqrt{n}B\left(\frac{r}{2},\frac{1}{2}\right)\left(1+\frac{t^2}{r}\right)^{(r+1)/2}} \\ &= \frac{1}{2} + \frac{1}{2}\left[I(1;\tfrac{1}{2}n,\tfrac{1}{2}) - I\left(\frac{r}{r+t^2},\tfrac{1}{2}n,\tfrac{1}{2}\right)\right], \\ & &(6) \end{aligned}$$

where

$$r \equiv n-1 \qquad (7)$$

is the number of Degrees of Freedom, $-\infty < t < \infty$, $\Gamma(z)$ is the Gamma Function, $B(a,b)$ is the Beta Function, and $I(z;a,b)$ is the Regularized Beta Function defined by

$$I(z;a,b) = \frac{B(z;a,b)}{B(a,b)}. \tag{8}$$

The Mean, Variance, Skewness, and Kurtosis of Student's t-distribution are

$$\mu = 0 \tag{9}$$

$$\sigma^2 = \frac{r}{r-2} \tag{10}$$

$$\gamma_1 = 0 \tag{11}$$

$$\gamma_2 = \frac{6}{r-4}. \tag{12}$$

Beyer (1987, p. 514) gives 60%, 70%, 90%, 95%, 97.5%, 99%, 99.5%, and 99.95% confidence intervals, and Goulden (1956) gives 50%, 90%, 95%, 98%, 99%, and 99.9% confidence intervals. A partial table is given below for small r and several common confidence intervals.

r	80%	90%	95%	99%
1	3.08	6.31	12.71	63.66
2	1.89	2.92	4.30	9.92
3	1.64	2.35	3.18	5.84
4	1.53	2.13	2.78	4.60
5	1.48	2.01	2.57	4.03
10	1.37	1.81	2.23	4.14
30	1.31	1.70	2.04	2.75
100	1.29	1.66	1.98	2.63
∞	1.28	1.65	1.96	2.58

The so-called $A(t|n)$ distribution is useful for testing if two observed distributions have the same Mean. $A(t|n)$ gives the probability that the difference in two observed Means for a certain statistic t with n Degrees of Freedom would be smaller than the observed value purely by chance:

$$A(t|n) = \frac{1}{\sqrt{n}\,B(\tfrac{1}{2},\tfrac{1}{2}n)} \int_{-t}^{t} \left(1+\frac{x^2}{n}\right)^{-(1+n)/2} dx. \tag{13}$$

Let X be a Normally Distributed random variable with Mean 0 and Variance σ^2, let Y^2/σ^2 have a Chi-Squared Distribution with n Degrees of Freedom, and let X and Y be independent. Then

$$t \equiv \frac{X\sqrt{n}}{Y} \tag{14}$$

is distributed as Student's t with n Degrees of Freedom.

The noncentral Student's t-distribution is given by

$$\begin{aligned} P(x) = {} & \frac{n^{n/2}n!}{2^n e^{\lambda^2/2}\Gamma(\frac{1}{2}n)} \\ & \times \left\{ \frac{\sqrt{2}\,\lambda x(n+x^2)^{-(1+n/2)}\,{}_1F_1\left(1+\frac{1}{2}n;\frac{3}{2};\frac{\lambda^2x^2}{2(n+x^2)}\right)}{\Gamma[\frac{1}{2}(1+n)]} \right. \\ & \left. \frac{+e^{(\lambda^2x^2)/[2(n+x^2)]}\sqrt{\pi}(n+x^2)^{-(n+1)/2}L_{n/2}^{-1/2}\left(-\frac{\lambda^2x^2}{2(n+x^2)}\right)}{\Gamma[\frac{1}{2}(1+n)]} \right\}, \end{aligned} \tag{15}$$

where $\Gamma(z)$ is the Gamma Function, ${}_1F_1(a;b;z)$ is a Confluent Hypergeometric Function, and $L_n^m(x)$ is an associated Laguerre Polynomial.

see also Paired t-Test, Student's z-Distribution

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 948–949, 1972.

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 536, 1987.

Fisher, R. A. "Applications of 'Student's' Distribution." *Metron* **5**, 3–17, 1925.

Fisher, R. A. "Expansion of 'Student's' Integral in Powers of $n-1$." *Metron* **5**, 22–32, 1925.

Fisher, R. A. *Statistical Methods for Research Workers, 10th ed.* Edinburgh: Oliver and Boyd, 1948.

Goulden, C. H. Table A–3 in *Methods of Statistical Analysis, 2nd ed.* New York: Wiley, p. 443, 1956.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Incomplete Beta Function, Student's Distribution, F-Distribution, Cumulative Binomial Distribution." §6.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 219–223, 1992.

Spiegel, M. R. *Theory and Problems of Probability and Statistics.* New York: McGraw-Hill, pp. 116–117, 1992.

Student. "The Probable Error of a Mean." *Biometrika* **6**, 1–25, 1908.

Student's z-Distribution

The probability density function and cumulative distribution functions for Student's z-distribution are given by

$$f(z) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)}(1+z^2)^{-n/2} \tag{1}$$

$$D(z) = \frac{-z^{1-n}\Gamma(\frac{1}{2}n)\,{}_2F_1(\frac{1}{2}(n-1),\frac{1}{2}n;\frac{1}{2}(n+1);-z^{-2})}{2\sqrt{\pi}\,\Gamma[\frac{1}{2}(n+1)]}. \tag{2}$$

The Mean is 0, so the Moments are

$$\mu_1 = 0 \tag{3}$$

$$\mu_2 = \frac{1}{n-3} \tag{4}$$

$$\mu_3 = 0 \tag{5}$$

$$\mu_4 = \frac{3}{(n-3)(n-5)}. \tag{6}$$

The MEAN, VARIANCE, SKEWNESS, and KURTOSIS are

$$\mu = 0 \tag{7}$$

$$\sigma^2 = \frac{1}{n-3} \tag{8}$$

$$\gamma_1 = 0 \tag{9}$$

$$\gamma_2 = \frac{6}{n-5}. \tag{10}$$

Letting

$$z \equiv \frac{(\bar{x}-\mu)}{s}, \tag{11}$$

where x is the sample MEAN and μ is the population MEAN gives STUDENT'S t-DISTRIBUTION.

see also STUDENT'S t-DISTRIBUTION

Study's Theorem

Given three curves ϕ_1, ϕ_2, ϕ_3 with the common group of ordinary points G (which may be empty), let their remaining groups of intersections g_{23}, g_{31}, and g_{12} also be ordinary points. If ϕ_1' is any other curve through g_{23}, then there exist two other curves ϕ_2', ϕ_3' such that the three combined curves $\phi_i\phi_i'$ are of the same order and LINEARLY DEPENDENT, each curve ϕ_k' contains the corresponding group g_{ij}, and every intersection of ϕ_i or ϕ_i' with ϕ_j or ϕ_j' lies on ϕ_k or ϕ_k'.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 34, 1959.

Sturm Chain

The series of STURM FUNCTIONS arising in application of the STURM THEOREM.

see also STURM FUNCTION, STURM THEOREM

Sturm Function

Given a function $f(x) \equiv f_0(x)$, write $f_1 \equiv f'(x)$ and define the Sturm functions by

$$f_n(x) = -\left\{ f_{n-2}(x) - f_{n-1}(x)\left[\frac{f_{n-2}(x)}{f_{n-1}(x)}\right]\right\}, \tag{1}$$

where $[P(x)/Q(x)]$ is a polynomial quotient. Then construct the following chain of Sturm functions,

$$\begin{aligned} f_0 &= q_0 f_1 - f_2 \\ f_1 &= q_1 f_2 - f_3 \\ f_2 &= q_2 f_3 - f_4 \\ &\vdots \\ f_{s-2} &= q_{s-2} f_{s-1} - f_s, \end{aligned} \tag{2}$$

known as a STURM CHAIN. The chain is terminated when a constant $-f_s(x)$ is obtained.

Sturm functions provide a convenient way for finding the number of real roots of an algebraic equation with real coefficients over a given interval. Specifically, the difference in the number of sign changes between the Sturm functions evaluated at two points $x = a$ and $x = b$ gives the number of real roots in the interval (a, b). This powerful result is known as the STURM THEOREM.

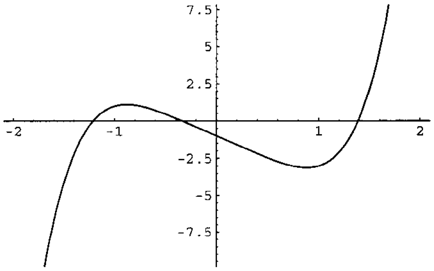

As a specific application of Sturm functions toward finding POLYNOMIAL ROOTS, consider the function $f_0(x) = x^5 - 3x - 1$, plotted above, which has roots -1.21465, -0.334734, $0.0802951 \pm 1.32836i$, and 1.38879 (three of which are real). The DERIVATIVE is given by $f'(x) = 5x^4 - 3$, and the STURM CHAIN is then given by

$$f_0 = x^5 - 3x - 1 \tag{3}$$

$$f_1 = 5x^4 - 3 \tag{4}$$

$$f_2 = \tfrac{1}{5}(12x + 5) \tag{5}$$

$$f_3 = \tfrac{59083}{20736}. \tag{6}$$

The following table shows the signs of f_i and the number of sign changes Δ obtained for points separated by $\Delta x = 2$.

x	f_0	f_1	f_2	f_3	Δ
-2	-1	1	-1	1	3
0	-1	-1	1	1	1
2	1	1	1	1	0

This shows that $3 - 1 = 2$ real roots lie in $(-2, 0)$, and $1 - 0 = 1$ real root lies in $(0, 2)$. Reducing the spacing to $\Delta x = 0.5$ gives the following table.

x	f_0	f_1	f_2	f_3	Δ
-2.0	-1	1	-1	1	3
-1.5	-1	1	-1	1	3
-1.0	1	1	-1	1	2
-0.5	1	-1	-1	1	2
0.0	-1	-1	1	1	1
0.5	-1	-1	1	1	1
1.0	-1	1	1	1	1
1.5	1	1	1	1	0
2.0	1	1	1	1	0

This table isolates the three real roots and shows that they lie in the intervals $(-1.5, -1.0)$, $(-0.5, 0.0)$, and $(1.0, 1.5)$. If desired, the intervals in which the roots fall could be further reduced.

The Sturm functions satisfy the following conditions:

1. Two neighboring functions do not vanish simultaneously at any point in the interval.
2. At a null point of a Sturm function, its two neighboring functions are of different signs.
3. Within a sufficiently small AREA surrounding a zero point of $f_0(x)$, $f_1(x)$ is everywhere greater than zero or everywhere smaller than zero.

see also DESCARTES' SIGN RULE, STURM CHAIN, STURM THEOREM

References
Acton, F. S. *Numerical Methods That Work, 2nd printing.* Washington, DC: Math. Assoc. Amer., p. 334, 1990.
Dörrie, H. "Sturm's Problem of the Number of Roots." §24 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 112–116, 1965.
Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 469, 1992.
Rusin, D. "Known Math." `http://www.math.niu.edu./~rusin/known-math/polynomials/sturm`.
Sturm, C. "Mémoire sur la résolution des équations numériques." *Bull. des sciences de Férussac* **11**, 1929.

Sturm-Liouville Equation

A second-order ORDINARY DIFFERENTIAL EQUATION

$$\frac{d}{dx}\left[p(x)\frac{dy}{dx}\right]+[\lambda w(x)-q(x)]y=0,$$

where λ is a constant and $w(x)$ is a known function called either the density or WEIGHTING FUNCTION. The solutions (with appropriate boundary conditions) of λ are called EIGENVALUES and the corresponding $u_\lambda(x)$ EIGENFUNCTIONS. The solutions of this equation satisfy important mathematical properties under appropriate boundary conditions (Arfken 1985).

see also ADJOINT OPERATOR, SELF-ADJOINT OPERATOR

References
Arfken, G. "Sturm-Liouville Theory—Orthogonal Functions." Ch. 9 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 497–538, 1985.

Sturm-Liouville Theory

see STURM-LIOUVILLE EQUATION

Sturm Theorem

The number of REAL ROOTS of an algebraic equation with REAL COEFFICIENTS whose REAL ROOTS are simple over an interval, the endpoints of which are not ROOTS, is equal to the difference between the number of sign changes of the STURM CHAINS formed for the interval ends.

see also STURM CHAIN, STURM FUNCTION

References
Dörrie, H. "Sturm's Problem of the Number of Roots." §24 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 112–116, 1965.
Rusin, D. "Known Math." `http://www.math.niu.edu./~rusin/known-math/polynomials/sturm`.

Sturmian Separation Theorem

Let $\mathsf{A}_r = a_{ij}$ be a SEQUENCE of N SYMMETRIC MATRICES of increasing order with $i, j = 1, 2, \ldots, r$ and $r = 1, 2, \ldots, N$. Let $\lambda_k(\mathsf{A}_r)$ be the kth EIGENVALUE of A_r for $k = 1, 2, \ldots, r$, where the ordering is given by

$$\lambda_1(\mathsf{A}_r)\geq\lambda_2(\mathsf{A}_r)\geq\ldots\geq\lambda_r(\mathsf{A}_r).$$

Then it follows that

$$\lambda_{k+1}(\mathsf{A}_{i+1})\leq\lambda_k(\mathsf{A}_i)\leq\lambda_k(\mathsf{A}_{i+1}).$$

References
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1121, 1979.

Sturmian Sequence

If a SEQUENCE has the property that the BLOCK GROWTH function $B(n) = n + 1$ for all n, then it is said to have minimal block growth, and the sequence is called a Sturmian sequence. An example of this is the sequence arising from the SUBSTITUTION MAP

$$\begin{aligned}0&\to 01\\ 1&\to 0,\end{aligned}$$

yielding $0 \to 01 \to 010 \to 01001 \to 01001010 \to \ldots$, which gives us the Sturmian sequence 01001010....

STURM FUNCTIONS are sometimes also said to form a Sturmian sequence.

see also STURM FUNCTION, STURM THEOREM

Subalgebra

An ALGEBRA S' which is part of a large ALGEBRA S and shares its properties.

see also ALGEBRA

Subanalytic

$X \subseteq \mathbb{R}^n$ is subanalytic if, for all $x \in \mathbb{R}^n$, there is an open U and $Y \subset \mathbb{R}^{n+m}$ a bounded SEMIANALYTIC set such that $X \cap U$ is the projection of Y into U.

see also SEMIANALYTIC

References
Bierstone, E. and Milman, P. "Semialgebraic and Subanalytic Sets." *IHES Pub. Math.* **67**, 5–42, 1988.
Marker, D. "Model Theory and Exponentiation." *Not. Amer. Math. Soc.* **43**, 753–759, 1996.

Subfactorial

The number of Permutations of n objects in which no object appear in its natural place (i.e., so-called "Derangements").

$$!n \equiv n! \sum_{k=0}^{n} \frac{(-1)^k}{k!} \tag{1}$$

or

$$!n \equiv \left[\frac{n!}{e}\right], \tag{2}$$

where $k!$ is the usual Factorial and $[x]$ is the Nint function. The first few values are $!1 = 0$, $!2 = 1$, $!3 = 2$, $!4 = 9$, $!5 = 44$, $!6 = 265$, $!7 = 1854$, $!8 = 14833$, ... (Sloane's A000166). For example, the only Derangements of $\{1,2,3\}$ are $\{2,3,1\}$ and $\{3,1,2\}$, so $!3 = 2$. Similarly, the Derangements of $\{1,2,3,4\}$ are $\{2,1,4,3\}$, $\{2,3,4,1\}$, $\{2,4,1,3\}$, $\{3,1,4,2\}$, $\{3,4,1,2\}$, $\{3,4,2,1\}$, $\{4,1,2,3\}$, $\{4,3,1,2\}$, and $\{4,3,2,1\}$, so $!4 = 9$.

The subfactorials are also called the Rencontres Numbers and satisfy the Recurrence Relations

$$!n = n{\cdot}!(n-1) + (-1)^n \tag{3}$$

$$!(n+1) = n[!n+!(n-1)]. \tag{4}$$

The subfactorial can be considered a special case of a restricted Rooks Problem.

The only number equal to the sum of subfactorials of its digits is

$$148,349 = !1+!4+!8+!3+!4+!9 \tag{5}$$

(Madachy 1979).

see also Derangement, Factorial, Married Couples Problem, Rooks Problem, Superfactorial

References

Dörrie, H. §6 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 19–21, 1965.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, p. 167, 1979.

Sloane, N. J. A. Sequences A000166/M1937 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sloane, N. J. A. and Plouffe, S. Extended entry in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.

Stanley, R. P. *Enumerative Combinatorics, Vol. 1.* Cambridge, England: Cambridge University Press, p. 67, 1997.

Subfield

If a subset S of the elements of a Field F satisfies the Field Axioms with the same operations of F, then S is called a subfield of F. Let F be a Finite Field of order p^n, then there exists a subfield of Order p^m for Prime p Iff m Divides n.

see also Field, Submanifold, Subspace

Subgraph

A Graph G' whose Vertices and Edges form subsets of the Vertices and Edges of a given Graph G. If G' is a subgraph of G, then G is said to be a Supergraph of G'.

see also Graph (Graph Theory), Supergraph

Subgroup

A subset of Group elements which satisfies the four Group requirements. The Order of any subgroup of a Group Order h must be a Divisor of h.

see also Cartan Subgroup, Composition Series, Fitting Subgroup, Group

Sublime Number

Let $\tau(n)$ and $\sigma(n)$ denote the number and sum of the divisors of n, respectively (i.e., the zeroth- and first-order Divisor Functions). A number N is called sublime if $\tau(N)$ and $\sigma(N)$ are both Perfect Numbers. The only two known sublime numbers are 12 and

$$60865556702383789896703717342431696\cdots$$
$$\cdots 22657830773351885970528324860512791691264.$$

It is not known if any Odd sublime number exists.

see also Divisor Function, Perfect Number

Submanifold

A C^∞ (infinitely differentiable) Manifold is said to be a submanifold of a C^∞ Manifold M' if M is a Subset of M' and the Identity Map of M into M' is an embedding.

see also Manifold, Subfield, Subspace

Submatrix

An $p \times q$ submatrix of an $m \times n$ Matrix (with $p \leq m$, $n \leq q$) is a $p \times q$ Matrix formed by taking a block of the entries of this size from the original matrix.

see also Matrix

Subnormal

L is a subnormal Subgroup of H if there is a a "normal series" (in the sense of Jordan-Holder) from L to H.

Subordinate Norm

see Natural Norm

Subscript

A quantity displayed below the normal line of text (and generally in a smaller point size), as the "i" in a_i, is called a subscript. Subscripts are commonly used to indicate indices (a_{ij} is the entry in the ith row and jth column of a Matrix A), partial differentiation (y_x is an abbreviation for $\partial y/\partial x$), and a host of other operations and notations in mathematics.

see also Superscript

Subsequence

A subsequence of a SEQUENCE $S = \{x_i\}_{i=1}^{n}$ is a derived sequence $\{y_i\}_{i=1}^{N} = \{x_{i+j}\}$ for some $j \geq 0$ and $N \leq n - j$. More generally, the word subsequence is sometimes used to mean a sequence derived from a sequence S by discarding some of its terms.

see also LOWER-TRIMMED SUBSEQUENCE, UPPER-TRIMMED SUBSEQUENCE

Subset

A portion of a SET. B is a subset of A (written $B \subseteq A$) IFF every member of B is a member of A. If B is a PROPER SUBSET of A (i.e., a subset other than the set itself), this is written $B \subset A$.

A SET of n elements has 2^n subsets (including the set itself and the EMPTY SET). For sets of $n = 1, 2, \ldots$ elements, the numbers of subsets are therefore 2, 4, 8, 16, 32, 64, ... (Sloane's A000079). For example, the set $\{1\}$ has the two subsets $\varnothing$ and $\{1\}$. Similarly, the set $\{1, 2\}$ has subsets $\varnothing$ (the EMPTY SET, $\{1\}$, $\{2\}$, and $\{1, 2\}$.

see also EMPTY SET, IMPLIES, k-SUBSET, PROPER SUBSET, SUPERSET, VENN DIAGRAM

References
Courant, R. and Robbins, H. *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, p. 109, 1996.
Ruskey, F. "Information of Subsets of a Set." `http://sue.csc.uvic.ca/~cos/inf/comb/SubsetInfo.html`.
Sloane, N. J. A. Sequence A000079/M1129 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Subspace

Let $\mathbb{V}$ be a REAL VECTOR SPACE (e.g., the real continuous functions $C(I)$ on a CLOSED INTERVAL I, 2-D EUCLIDEAN SPACE $\mathbb{R}^2$, the twice differentiable real functions $C^{(2)}(I)$ on I, etc.). Then $\mathbb{W}$ is a real SUBSPACE of $\mathbb{V}$ if $\mathbb{W}$ is a SUBSET of $\mathbb{V}$ and, for every $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{W}$ and $t \in \mathbb{R}$ (the REALS), $\mathbf{w}_1 + \mathbf{w}_2 \in \mathbb{W}$ and $t\mathbf{w}_1 \in \mathbb{W}$. Let (H) be a homogeneous system of linear equations in $x_1, \ldots, x_n$. Then the SUBSET S of $\mathbb{R}^n$ which consists of all solutions of the system (H) is a subspace of $\mathbb{R}^n$.

More generally, let F_q be a FIELD with $q = p^\alpha$, where p is PRIME, and let $F_{q,n}$ denote the n-D VECTOR SPACE over F_q. The number of k-D linear subspaces of $F_{q,n}$ is

$$N(F_{q,n}) = \binom{n}{k}_q,$$

where this is the q-BINOMIAL COEFFICIENT (Aigner 1979, Exton 1983). The asymptotic limit is

$$N(F_{q,n}) = \begin{cases} c_e q^{n^2/4}[1 + o(1)] & \text{for } n \text{ even} \\ c_o q^{n^2/4}[1 + o(1)] & \text{for } n \text{ odd,} \end{cases}$$

where

$$c_e = \frac{\sum_{k=-\infty}^{\infty} q^{-k^2}}{\prod_{j=1}^{\infty}(1 - q^{-j})}$$

$$c_o = \frac{\sum_{k=-\infty}^{\infty} q^{-(k+1/2)^2}}{\prod_{j=1}^{\infty}(1 - q^{-j})}$$

(Finch). The case $q = 2$ gives the q-ANALOG of the WALLIS FORMULA.

see also q-BINOMIAL COEFFICIENT, SUBFIELD, SUBMANIFOLD

References
Aigner, M. *Combinatorial Theory.* New York: Springer-Verlag, 1979.
Exton, H. *q-Hypergeometric Functions and Applications.* New York: Halstead Press, 1983.
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/dig/dig.html`.

Substitution Group

see PERMUTATION GROUP

Substitution Map

A MAP which uses a set of rules to transform elements of a sequence into a new sequence using a set of rules which "translate" from the original sequence to its transformation. For example, the substitution map $\{1 \to 0, 0 \to 11\}$ would take 10 to 011.

see also GOLDEN RATIO, MORSE-THUE SEQUENCE, STRING REWRITING, THUE CONSTANT

Subtend

Given a geometric object O in the PLANE and a point P, let A be the ANGLE from one edge of O to the other with VERTEX at P. Then O is said to subtend an ANGLE A from P.

see also ANGLE, VERTEX ANGLE

Subtraction

Subtraction is the operation of taking the DIFFERENCE $x - y$ of two numbers x and y. Here, the symbol between the x and y is called the MINUS SIGN and $x - y$ is read "x MINUS y."

see also ADDITION, DIVISION, MINUS, MINUS SIGN, MULTIPLICATION

Succeeds

The relationship x succeeds (or FOLLOWS) y is written $x \succ y$. The relation x succeeds or is equal to y is written $x \succeq y$.

see also PRECEDES

Successes

see DIFFERENCE OF SUCCESSES

Sufficient

A CONDITION which, if true, guarantees that a result is also true. (However, the result may also be true if the CONDITION is not met.) If a CONDITION is both NECESSARY and SUFFICIENT, then the result is said to be true IFF ("if and only if") the CONDITION holds.

For example, the condition that a decimal number n end in the DIGIT 2 is a sufficient but not NECESSARY condition that n be EVEN.

see also IFF, IMPLIES, NECESSARY

Suitable Number

see IDONEAL NUMBER

Sum

A sum is the result of an ADDITION. For example, adding 1, 2, 3, and 4 gives the sum 10, written

$$1+2+3+4=10. \tag{1}$$

The numbers being summed are called ADDENDS, or sometimes SUMMANDS. The summation operation can also be indicated using a capital sigma with upper and lower limits written above and below, and the index indicated below. For example, the above sum could be written

$$\sum_{k=1}^{4} k = 10. \tag{2}$$

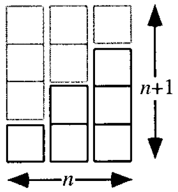

A simple graphical proof of the sum $\sum_{k=1}^{n} k = n(n+1)/2$ can also be given. Construct a sequence of stacks of boxes, each 1 unit across and k units high, where $k = 1$, 2, ..., n. Now add a rotated copy on top, as in the above figure. Note that the resulting figure has WIDTH n and HEIGHT $n+1$, and so has AREA $n(n+1)$. The desired sum is half this, so the AREA of the boxes in the sum is $n(n+1)/2$. Since the boxes are of unit width, this is also the value of the sum.

The sum can also be computed using the first EULER-MACLAURIN INTEGRATION FORMULA

$$\sum_{k=1}^{n} f(k) = \int_1^n f(x)\,dx + \tfrac{1}{2}f(1) + \tfrac{1}{2}f(n) + \tfrac{1}{2!}B_2[f'(n) - f'(1)] + \ldots \tag{3}$$

with $f(k) = k$. Then

$$\sum_{k=1}^{n} k = \int_1^n x\,dx + \tfrac{1}{2}\cdot 1 + \tfrac{1}{2}\cdot n + \tfrac{1}{6}(1-1) + \ldots = \tfrac{1}{2}(n^2-1) - \tfrac{1}{2} + h + \tfrac{1}{2}n = \tfrac{1}{2}n(n+1). \tag{4}$$

The general finite sum of integral POWERS can be given by the expression

$$\sum_{k=1}^{n} k^p = \frac{(B+n+1)^{[p+1]} - B^{[p+1]}}{p+1}, \tag{5}$$

where the NOTATION $B^{[k]}$ means the quantity in question is raised to the appropriate POWER k and all terms of the form B^m are replaced with the corresponding BERNOULLI NUMBERS B_m. It is also true that the COEFFICIENTS of the terms in such an expansion sum to 1, as stated by Bernoulli without proof (Boyer 1943).

An analytic solution for a sum of POWERS of integers is

$$\sum_{k=1}^{n} k^p = \zeta(-p) - \zeta(-p, 1+n), \tag{6}$$

where $\zeta(z)$ is the RIEMANN ZETA FUNCTION and $\zeta(z;a)$ is the HURWITZ ZETA FUNCTION. For the special case of p a POSITIVE integer, FAULHABER'S FORMULA gives the SUM explicitly as

$$\sum_{k=1}^{n} k^p = \frac{1}{p+1}\sum_{k=1}^{p+1}(-1)^{\delta_{kp}}\binom{p+1}{k}B_{p+1-k}n^k, \tag{7}$$

where δ_{kp} is the KRONECKER DELTA, $\binom{n}{k}$ is a BINOMIAL COEFFICIENT, and B_k is a BERNOULLI NUMBER. Written explicitly in terms of a sum of POWERS,

$$\sum_{k=1}^{n} k^p = \frac{B_k p!}{k!(p-k+1)!}n^{p-k+1}. \tag{8}$$

Computing the sums for $p = 1, \ldots, 10$ gives

$$\sum_{k=1}^{n} k = \tfrac{1}{2}(n^2+n) \tag{9}$$

$$\sum_{k=1}^{n} k^2 = \tfrac{1}{6}(2n^3+3n^2+n) \tag{10}$$

$$\sum_{k=1}^{n} k^3 = \tfrac{1}{4}(n^4+2n^3+n^2) \tag{11}$$

$$\sum_{k=1}^{n} k^4 = \tfrac{1}{30}(6n^5+15n^4+10n^3-n) \tag{12}$$

$$\sum_{k=1}^{n} k^5 = \tfrac{1}{12}(2n^6+6n^5+5n^4-n^2) \tag{13}$$

$$\sum_{k=1}^{n} k^6 = \tfrac{1}{42}(6n^7 + 21n^6 + 21n^5 - 7n^3 + n) \quad (14)$$

$$\sum_{k=1}^{n} k^7 = \tfrac{1}{24}(3n^8 + 12n^7 + 14n^6 - 7n^4 + 2n^2) \quad (15)$$

$$\sum_{k=1}^{n} k^8 = \tfrac{1}{90}(10n^9 + 45n^8 + 60n^7 - 42n^5 + 20n^3 - 3n) \quad (16)$$

$$\sum_{k=1}^{n} k^9 = \tfrac{1}{20}(2n^{10} + 10n^9 + 15n^8 - 14n^6 + 10n^4 - 3n^2) \quad (17)$$

$$\sum_{k=1}^{n} k^{10} = \tfrac{1}{66}(6n^{11} + 33n^{10} + 55n^9 - 66n^7 + 66n^5 - 33n^3 + 5n). \quad (18)$$

Factoring the above equations results in

$$\sum_{k=1}^{n} k = \tfrac{1}{2}n(n+1) \quad (19)$$

$$\sum_{k=1}^{n} k^2 = \tfrac{1}{6}n(n+1)(2n+1) \quad (20)$$

$$\sum_{k=1}^{n} k^3 = \tfrac{1}{4}n^2(n+1)^2 \quad (21)$$

$$\sum_{k=1}^{n} k^4 = \tfrac{1}{30}n(n+1)(2n+1)(3n^2+3n-1) \quad (22)$$

$$\sum_{k=1}^{n} k^5 = \tfrac{1}{12}n^2(n+1)^2(2n^2+2n-1) \quad (23)$$

$$\sum_{k=1}^{n} k^6 = \tfrac{1}{42}n(n+1)(2n+1)(3n^4+6n^3-3n+1) \quad (24)$$

$$\sum_{k=1}^{n} k^7 = \tfrac{1}{24}n^2(n+1)^2(3n^4+6n^3-n^2-4n+2) \quad (25)$$

$$\sum_{k=1}^{n} k^8 = \tfrac{1}{90}n(n+1)(2n+1)(5n^6+15n^5+5n^4 -15n^3-n^2+9n-3) \quad (26)$$

$$\sum_{k=1}^{n} k^9 = \tfrac{1}{20}n^2(n+1)^2(n^2+n-1) \times(2n^4+4n^3-n^2-3n+3) \quad (27)$$

$$\sum_{k=1}^{n} k^{10} = \tfrac{1}{66}n(n+1)(2n+1)(n^2+n-1) \times(3n^6+9n^5+2n^4-11n^3+3n^2+10n-5). \quad (28)$$

From the above, note the interesting identity

$$\sum_{k=1}^{n} k^3 = \left(\sum_{k=1}^{n} k\right)^2. \quad (29)$$

Sums of the following type can also be done analytically.

$$\left(\sum_{k=0}^{\infty} x^k\right)^2 = \sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} 1\right) x^n = \sum_{n=0}^{\infty}(n+1)x^n \quad (30)$$

$$\begin{aligned}\left(\sum_{k=0}^{\infty} x^k\right)^3 &= \sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} k\right) x^n \\ &= \frac{1}{2}\sum_{n=0}^{\infty}(n+1)(n+2)x^n \quad (31)\end{aligned}$$

$$\begin{aligned}\left(\sum_{k=0}^{\infty} x^k\right)^4 &= \sum_{n=0}^{\infty}\left[\sum_{k=0}^{n} \tfrac{1}{2}(k+1)(k+2)\right] x^n \\ &= \frac{1}{2}\sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} k^2+3k+2\right) x^n \\ &= \frac{1}{2}\sum_{n=0}^{\infty}[\tfrac{1}{6}n(n+1)(2n+1) \\ &\quad + 3\tfrac{1}{2}n(n+1) + 2(n+1)]x^n \\ &= \frac{1}{12}\sum_{n=0}^{\infty}(n+1)[n(2n+1)+9n+12]x^n \\ &= \frac{1}{12}\sum_{n=0}^{\infty}(n+1)(2n^2+10n+12)x^n \\ &= \frac{1}{6}\sum_{n=0}^{\infty}(n+1)(n+2)(n+3)x^n. \quad (32)\end{aligned}$$

By INDUCTION, the sum for an arbitrary POWER p is

$$\left(\sum_{k=0}^{\infty} x^k\right)^p = \frac{1}{(p-1)!}\sum_{n=0}^{\infty}\frac{(n+p-1)!}{n!}x^n. \quad (33)$$

Other analytic sums include

$$\left(\sum_{k=0}^{n} x^k\right)^2 = \frac{1}{(p-1)!}\sum_{k=0}^{2n}\frac{(n-|n-k|+p-1)!}{(n-|n-k|)!}x^k \quad (34)$$

$$\left(\sum_{n=0}^{\infty} a_n x^n\right)^2 = \sum_{n=0}^{\infty} {a_n}^2 x^{2n} + 2\sum_{\substack{n=1\\ i+j=n\\ i<j}}^{\infty} a_i a_j x^n. \quad (35)$$

$$\begin{aligned}\sum xy &= x_1y_1 + x_1y_2 + \ldots + x_2y_1 + x_2y_2 + \ldots \\ &= (x_1+x_2+\ldots)y_1 + (x_1+x_2+\ldots)y_2 \\ &= \left(\sum x\right)(y_1+y_2+\ldots) = \sum x \sum y, \quad (36)\end{aligned}$$

so

$$\sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j = \left(\sum_{i=1}^{m} x_i\right)\left(\sum_{j=1}^{n} y_j\right). \tag{37}$$

$$\sum_{j=0}^{n} j x^j = \frac{n x^{n+2} - (n+1)x^{n+1} + x}{(x-1)^2} \tag{38}$$

$$\sum_{j=1}^{n} \frac{{x_j}^r}{\prod_{\substack{k=1\\k\neq j}}^{n} (x_j - x_k)} = \begin{cases} 0 & \text{for } 0 \leq r < n-1 \\ 1 & \text{for } r = n-1 \\ \sum_{j=1}^{n} x_j & \text{for } r = n \end{cases} \tag{39}$$

$$\sum_{k=1}^{n} \frac{\prod_{\substack{r=1\\r\neq k}}^{n} (x+k-r)}{\prod_{\substack{r=1\\r\neq k}}^{n} (k-r)} = 1 \tag{40}$$

$$(n+1)\sum_{m=1}^{n} m^k = \sum_{m=1}^{n}\left[m^{k+1} + \sum_{p=1}^{n}\left(\sum_{m=1}^{p} m^k\right)\right]. \tag{41}$$

To minimize the sum of a set of squares of numbers $\{x_i\}$ about a given number x_0

$$S \equiv \sum_i (x_i - x_0)^2 = \sum_i {x_i}^2 - 2x_0 \sum x_i + N{x_0}^2, \tag{42}$$

take the DERIVATIVE.

$$\frac{d}{dx_0} S = -2\sum_i x_i + 2Nx_0 = 0. \tag{43}$$

Solving for x_0 gives

$$x_0 \equiv \bar{x} = \frac{1}{N}\sum_i x_i, \tag{44}$$

so S is maximized when x_0 is set to the MEAN.

see also ARITHMETIC SERIES, BERNOULLI NUMBER, CLARK'S TRIANGLE, CONVERGENCE IMPROVEMENT, DEDEKIND SUM, DOUBLE SUM, EULER SUM, FACTORIAL SUM, FAULHABER'S FORMULA, GABRIEL'S STAIRCASE, GAUSSIAN SUM, GEOMETRIC SERIES, GOSPER'S METHOD, HURWITZ ZETA FUNCTION, INFINITE PRODUCT, KLOOSTERMAN'S SUM, LEGENDRE SUM, LERCH TRANSCENDENT, PASCAL'S TRIANGLE, PRODUCT, RAMANUJAN'S SUM, RIEMANN ZETA FUNCTION, WHITNEY SUM

References

Boyer, C. B. "Pascal's Formula for the Sums of the Powers of the Integers." *Scripta Math.* **9**, 237–244, 1943.

Courant, R. and Robbins, H. "The Sum of the First n Squares." §1.4 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 14–15, 1996.

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. *A=B.* Wellesley, MA: A. K. Peters, 1996.

Sum-Product Number

A sum-product number is a number n such that the sum of n's digits times the product of n's digit is n itself, for example

$$135 = (1+3+5)(1 \cdot 3 \cdot 5).$$

The only sum-product numbers less than 10^7 are 1, 135, and 144.

see also AMENABLE NUMBER

Sum Rule

$$\frac{d}{dx}[f(x) + g(x)] = f'(x) + g'(x),$$

where d/dx denotes a derivative and $f'(x)$ and $g'(x)$ are the derivatives of f and g, respectively.

see also DERIVATIVE

Summand

see ADDEND

Summatory Function

For an discrete function $f(n)$, the summatory function is defined by

$$F(n) \equiv \sum_{k \in D}^{n} f(k),$$

where D is the DOMAIN of the function.

see also DIVISOR FUNCTION, MANGOLDT FUNCTION, MERTENS FUNCTION, RUDIN-SHAPIRO SEQUENCE, TAU FUNCTION, TOTIENT FUNCTION

Sup

see SUPREMUM, SUPREMUM LIMIT

Super-3 Number

An INTEGER n such that $3n^3$ contains three consecutive 3s in its DECIMAL representation. The first few super-3 numbers are 261, 462, 471, 481, 558, 753, 1036, ... (Sloane's A014569). A. Anderson has conjectured that all numbers ending in 471, 4710, or 47100 are super-3 (Pickover 1995).

For a digit d, super-3 numbers can be generalized to super-d numbers n such that dn^d contains d ds in its DECIMAL representation. The following table gives the first few super-d numbers for small d.

d	Sloane	Super-d numbers
2	032743	19, 31, 69, 81, 105, 106, 107, 119, ...
3	014569	261, 462, 471, 481, 558, 753, 1036, ...
4	032744	1168, 4972, 7423, 7752, 8431, 10267, ...
5	032745	4602, 5517, 7539, 12955, 14555, 20137, ...
6	032746	27257, 272570, 302693, 323576, ...
7	032747	140997, 490996, 1184321, 1259609, ...
8	032748	185423, 641519, 1551728, 1854230, ...
9	032749	17546133, 32613656, 93568867, ...

References

Pickover, C. A. *Keys to Infinity.* New York: Wiley, p. 7, 1995.

Sloane, N. J. A. Sequence A014569 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Super Catalan Number

While the CATALAN NUMBERS are the number of p-GOOD PATHS from (n,n) to (0,0) which do not cross the diagonal line, the super Catalan numbers count the number of LATTICE PATHS with diagonal steps from (n,n) to (0,0) which do not touch the diagonal line $x = y$.

The super Catalan numbers are given by the RECURRENCE RELATION

$$S(n) = \frac{3(2n-3)S(n-1)-(n-3)S(n-2)}{n}$$

(Comtet 1974), with $S(1) = S(2) = 1$. (Note that the expression in Vardi (1991, p. 198) contains *two* errors.) A closed form expression in terms of LEGENDRE POLYNOMIALS $P_n(x)$ is

$$S(n) = \frac{3P_{n-1}(3) - P_{n-2}(3)}{4n}$$

(Vardi 1991, p. 199). The first few super Catalan numbers are 1, 1, 3, 11, 45, 197, ... (Sloane's A001003).

see also CATALAN NUMBER

References

Comtet, L. *Advanced Combinatorics.* Dordrecht, Netherlands: Reidel, p. 56, 1974.

Graham, R. L.; Knuth, D. E.; and Patashnik, O. Exercise 7.50 in *Concrete Mathematics: A Foundation for Computer Science, 2nd ed.* Reading, MA: Addison-Wesley, 1994.

Motzkin, T. "Relations Between Hypersurface Cross Ratios and a Combinatorial Formula for Partitions of a Polygon for Permanent Preponderance and for Non-Associative Products." *Bull. Amer. Math. Soc.* **54**, 352–360, 1948.

Schröder, E. "Vier combinatorische Probleme." *Z. Math. Phys.* **15**, 361–376, 1870.

Sloane, N. J. A. Sequence A001003/M2898 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 198–199, 1991.

Super-Poulet Number

A POULET NUMBER whose DIVISORS d all satisfy $d|2^d - 2$.

see also POULET NUMBER

Superabundant Number

see HIGHLY COMPOSITE NUMBER

Superegg

A superegg is a solid described by the equation

$$\left|\sqrt{\frac{x^2+y^2}{a^2}}\right|^n + \left|\frac{z}{b}\right|^n = 1.$$

Supereggs will balance on either end for any a, b, and n.

see also EGG, SUPERELLIPSE

References

Gardner, M. "Pier Hein's Superellipse." Ch. 18 in *Mathematical Carnival: A New Round-Up of Tantalizers and Puzzles from Scientific American.* New York: Vintage, 1977.

Superellipse

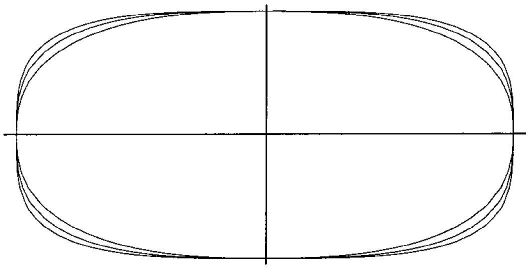

A curve of the form

$$\left|\frac{x}{a}\right|^r + \left|\frac{y}{b}\right|^r = 1.$$

where $r > 2$. "The" superellipse is sometimes taken as the curve of the above form with $r = 5/2$. Superellipses with $a = b$ are also known as LAMÉ CURVES. The above curves are for $a = 1$, $b = 2$, and $r = 2.5$, 3.0, and 3.5.

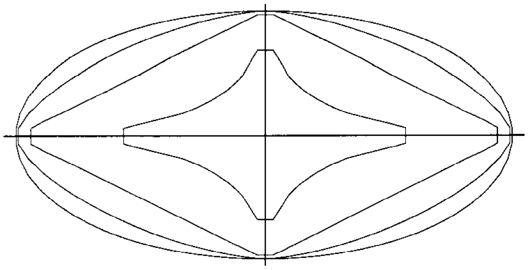

A degenerate superellipse is a superellipse with $r \leq 2$. The above curves are for $a = 1$, $b = 2$, and $r = 0.5$, 1.0, 1.5, and 2.0.

see also ELLIPSE, LAMÉ CURVE, SUPEREGG

References

Gardner, M. "Piet Hein's Superellipse." Ch. 18 in *Mathematical Carnival: A New Round-Up of Tantalizers and Puzzles from Scientific American.* New York: Vintage, 1977.

Superfactorial

The superfactorial of n is defined by Pickover (1995) as

$$n\$ \equiv \underbrace{n!^{n!^{\cdot^{\cdot^{\cdot^{n!}}}}}}_{n!}.$$

The first two values are 1 and 4, but subsequently grow so rapidly that 3$ already has a huge number of digits.

Sloane and Plouffe (1995) define the superfactorial by

$$n\$ \equiv \prod_{i=1}^{n} i!,$$

which is equivalent to the integral values of the G-FUNCTION. The first few values are 1, 1, 2, 12, 288, 34560, ... (Sloane's A000178).

see also FACTORIAL, G-FUNCTION, LARGE NUMBER, SUBFACTORIAL

References

Pickover, C. A. *Keys to Infinity.* New York: Wiley, p. 102, 1995.

Sloane, N. J. A. Sequence A000178/M2049 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Supergraph

If G' is a SUBGRAPH of G, then G is said to be a supergraph of G'.

see also GRAPH (GRAPH THEORY), SUBGRAPH

Supernormal

Trials for which the LEXIS RATIO

$$L \equiv \frac{\sigma}{\sigma_B},$$

satisfies $L > 1$, where σ is the VARIANCE in a set of s LEXIS TRIALS and σ_B is the VARIANCE assuming BERNOULLI TRIALS.

see also BERNOULLI TRIAL, LEXIS TRIALS, SUBNORMAL

Superperfect Number

A number n such that

$$\sigma^2(n) = \sigma(\sigma(n)) = 2n,$$

where $\sigma(n)$ is the DIVISOR FUNCTION. EVEN superperfect numbers are just 2^{p-1}, where $M_p = 2^p - 1$ is a MERSENNE PRIME. If any ODD superperfect numbers exist, they are SQUARE NUMBERS and either n or $\sigma(n)$ is DIVISIBLE by at least three distinct PRIMES.

More generally, an m-superperfect number is a number for which $\sigma^m(n) = 2n$. For $m \geq 3$, there are no EVEN m-superperfect numbers.

see also MERSENNE NUMBER

References

Guy, R. K. "Superperfect Numbers." §B9 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 65–66, 1994.

Kanold, H.-J. "Über 'Super Perfect Numbers.'" *Elem. Math.* **24**, 61–62, 1969.

Lord, G. "Even Perfect and Superperfect Numbers." *Elem. Math.* **30**, 87–88, 1975.

Suryanarayana, D. "Super Perfect Numbers." *Elem. Math.* **20**, 16–17, 1969.

Suryanarayana, D. "There is No Odd Super Perfect Number of the Form $p^{2\alpha}$." *Elem. Math.* **24**, 148–150, 1973.

Superposition Principle

For a linear homogeneous ORDINARY DIFFERENTIAL EQUATION, if $y_1(x)$ and $y_2(x)$ are solutions, then so is $y_1(x) + y_2(x)$.

Superregular Graph

For a VERTEX x of a GRAPH, let Γ_x and Δ_x denote the SUBGRAPHS of $\Gamma - x$ induced by the VERTICES adjacent to and nonadjacent to x, respectively. The empty graph is defined to be superregular, and Γ is said to be superregular if Γ is a REGULAR GRAPH and both Γ_x and Δ_x are superregular for all x.

The superregular graphs are precisely C_5, mK_n $(m, n \geq 1)$, G_n $(n \geq 1)$, and the complements of these graphs, where C_n is a CYCLIC GRAPH, K_n is a COMPLETE GRAPH and mK_n is m disjoint copies of K_n, and G_n is the Cartesian product of K_n with itself (the graph whose VERTEX set consists of n^2 VERTICES arranged in an $n \times n$ square with two VERTICES adjacent IFF they are in the same row or column).

see also COMPLETE GRAPH, CYCLIC GRAPH, REGULAR GRAPH

References

Vince, A. "The Superregular Graph." Problem 6617. *Amer. Math. Monthly* **103**, 600–603, 1996.

West, D. B. "The Superregular Graphs." *J. Graph Th.* **23**, 289–295, 1996.

Superscript

A quantity displayed above the normal line of text (and generally in a smaller point size), as the "i" in x^i, is called a superscript. Superscripts are commonly used to indicate raising to a POWER (x^3 means $x \cdot x \cdot x$ or x CUBED), multiple differentiation ($f^{(3)}(x)$ is an abbreviation for $f'''(x) = d^3f/dx^3$), and a host of other operations and notations in mathematics.

see also SUBSCRIPT

Superset

A SET containing all elements of a smaller SET. If B is a SUBSET of A, then A is a superset of B, written $A \supseteq B$. If A is a PROPER SUPERSET of B, this is written $A \supset B$.

see also PROPER SUBSET, PROPER SUPERSET, SUBSET

Supplementary Angle

Two ANGLES α and $\pi - \alpha$ which together form a STRAIGHT ANGLE are said to be supplementary.

see also ANGLE, COMPLEMENTARY ANGLE, DIGON, STRAIGHT ANGLE

Support

The CLOSURE of the SET of arguments of a FUNCTION f for which f is not zero.

see also CLOSURE

Support Function

Let M be an oriented REGULAR SURFACE in $\mathbb{R}^3$ with normal $\mathbf{N}$. Then the support function of M is the function $h : M \to \mathbb{R}$ defined by

$$h(\mathbf{p}) = \mathbf{p} \cdot \mathbf{N}(\mathbf{p}).$$

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 293, 1993.

Supremum

The supremum of a set is the least upper bound of the set. It is denoted

$$\sup_S .$$

On the REAL LINE, the supremum of a set is the same as the supremum of its CLOSURE.

see also INFIMUM, SUPREMUM LIMIT

Supremum Limit

The limit supremum is used for sequences and nets (as opposed to sets) and is denoted

$$\limsup_S .$$

see also SUPREMUM

Surd

An archaic term for a SQUARE ROOT.

see also QUADRATIC SURD, SQUARE ROOT

Surface

The word "surface" is an important term in mathematics and is used in many ways. The most common and straightforward use of the word is to denote a 2-D SUBMANIFOLD of 3-D EUCLIDEAN SPACE. Surfaces can range from the very complicated (e.g., FRACTALS such as the MANDELBROT SET) to the very simple (such as the PLANE). More generally, the word "surface" can be used to denote an $(n-1)$-D SUBMANIFOLD of an n-D MANIFOLD, or in general, any co-dimension 1 subobject in an object (like a BANACH SPACE or an infinite-dimensional MANIFOLD).

Even simple surfaces can display surprisingly counterintuitive properties. For example, the SURFACE OF REVOLUTION of $y = 1/x$ around the x-AXIS for $x \geq 1$ (called GABRIEL'S HORN) has FINITE VOLUME but INFINITE SURFACE AREA.

see also ALGEBRAIC SURFACE, BARTH DECIC, BARTH SEXTIC, BERNSTEIN MINIMAL SURFACE THEOREM, BOHEMIAN DOME, BOY SURFACE, CATALAN'S SURFACE, CAYLEY'S RULED SURFACE, CHAIR, CLEBSCH DIAGONAL CUBIC, COMPACT SURFACE, CONE, CONICAL WEDGE, CONOCUNEUS OF WALLIS, CORK PLUG, CORKSCREW SURFACE, CORNUCOPIA, COSTA MINIMAL SURFACE, CROSS-CAP, CROSSED TROUGH, CUBIC SURFACE, CYCLIDE, CYLINDER, CYLINDROID, DARWIN-DE SITTER SPHEROID, DECIC SURFACE, DEL PEZZO SURFACE, DERVISH, DESMIC SURFACE, DEVELOPABLE SURFACE, DINI'S SURFACE, EIGHT SURFACE, ELLIPSOID, ELLIPTIC CONE, ELLIPTIC CYLINDER, ELLIPTIC HELICOID, ELLIPTIC HYPERBOLOID, ELLIPTIC PARABOLOID, ELLIPTIC TORUS, ENNEPER'S SURFACES, ENRIQUES SURFACES, ETRUSCAN VENUS SURFACE, FLAT SURFACE, FRESNEL'S ELASTICITY SURFACE, GABRIEL'S HORN, HANDKERCHIEF SURFACE, HELICOID, HENNEBERG'S MINIMAL SURFACE, HOFFMAN'S MINIMAL SURFACE, HORN CYCLIDE, HORN TORUS, HUNT'S SURFACE, HYPERBOLIC CYLINDER, HYPERBOLIC PARABOLOID, HYPERBOLOID, IDA SURFACE, IMMERSED MINIMAL SURFACE, KISS SURFACE, KLEIN BOTTLE, KUEN SURFACE, KUMMER SURFACE, LICHTENFELS SURFACE, MAEDER'S OWL MINIMAL SURFACE, MANIFOLD, MENN'S SURFACE, MINIMAL SURFACE, MITER SURFACE, MÖBIUS STRIP, MONGE'S FORM, MONKEY SADDLE, NONORIENTABLE SURFACE, NORDSTRAND'S WEIRD SURFACE, NURBS SURFACE, OBLATE SPHEROID, OCTIC SURFACE, ORIENTABLE SURFACE, PARABOLIC CYLINDER, PARABOLIC HORN CYCLIDE, PARABOLIC RING CYCLIDE, PARABOLIC SPINDLE CYCLIDE, PARABOLOID, PEANO SURFACE, PIRIFORM, PLANE, PLÜCKER'S CONOID, POLYHEDRON, PRISM, PRISMATOID, PROLATE SPHEROID, PSEUDOCROSSCAP, QUADRATIC SURFACE, QUARTIC SURFACE, QUINTIC SURFACE, REGULAR SURFACE, REMBS' SURFACES, RIEMANN SURFACE, RING CYCLIDE, RING TORUS, ROMAN SURFACE, RULED SURFACE, SCHERK'S MINIMAL SURFACES, SEIFERT SURFACE, SEXTIC SURFACE, SHOE SURFACE, SIEVERT'S SURFACE, SMOOTH SURFACE, SOLID, SPHERE, SPHEROID, SPINDLE CYCLIDE, SPINDLE TORUS, STEINBACH SCREW, STEINER SURFACE, SWALLOWTAIL CATASTROPHE, SYMMETROID, TANGLECUBE, TETRAHEDRAL SURFACE, TOGLIATTI SURFACE, TOOTH SURFACE, TRINOID, UNDULOID, VERONESE SURFACE, VERONESE VARIETY, WALLIS'S CONICAL EDGE, WAVE SURFACE, WEDGE, WHITNEY UMBRELLA

References

Endraß, S. "Home Page of S. Endraß." `http://www.mathematik.uni-mainz.de/~endrass/`.

Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, 1986.
Francis, G. K. *A Topological Picturebook.* New York: Springer-Verlag, 1987.
Geometry Center. "The Topological Zoo." `http://www.geom.umn.edu/zoo/`.
Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, 1993.
Hunt, B. "Algebraic Surfaces." `http://www.mathematik.uni-kl.de/~wwwagag/Galerie.html`.
Morgan, F. "What is a Surface?" *Amer. Math. Monthly* **103**, 369–376, 1996.
Nordstrand, T. "Gallery." `http://www.uib.no/people/nfytn/mathgal.htm`.
Nordstrand, T. "Surfaces." `http://www.uib.no/people/nfytn/surfaces.htm`.
von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, 1993.
Wagon, S. "Surfaces." Ch. 3 in *Mathematica in Action.* New York: W. H. Freeman, pp. 67–91, 1991.
Yamaguchi, F. *Curves and Surfaces in Computer Aided Geometric Design.* New York: Springer-Verlag, 1988.

Surface Area

Surface area is the AREA of a given surface. Roughly speaking, it is the "amount" of a surface, and has units of distance squares. It is commonly denoted S for a surface in 3-D, or A for a region of the plane (in which case it is simply called "the" AREA).

If the surface is PARAMETERIZED using u and v, then

$$S = \int_S |\mathbf{T}_u \times \mathbf{T}_v|\, du\, dv, \tag{1}$$

where $\mathbf{T}_u$ and $\hat{\mathbf{T}}_v$ are tangent vectors and $\mathbf{a} \times \mathbf{b}$ is the CROSS PRODUCT.

The surface area given by rotating the curve $y = f(x)$ from $x = a$ to $x = b$ about the x-axis is

$$S = \int_b^a 2\pi f(x)\sqrt{1 + [f'(x)]^2}\, dx. \tag{2}$$

If $z = f(x, y)$ is defined over a region R, then

$$S = \iint_R \sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2 + 1}\, dA, \tag{3}$$

where the integral is taken over the entire surface.

The following tables gives surface areas for some common SURFACES. In the first table, S denotes the lateral surface, and in the second, T denotes the total surface. In both tables, r denotes the RADIUS, h the height, p the base PERIMETER, and s the SLANT HEIGHT (Beyer 1987).

Surface	S
cone	$\pi r\sqrt{r^2 + h^2}$
conical frustum	$\pi(R_1 + R_2)\sqrt{(R_1 - R_2)^2 + h^2}$
cube	$6a^2$
cylinder	$2\pi rh$
lune	$2r^2\theta$
oblate spheroid	$2\pi a^2 + \frac{\pi b^2}{e}\ln\left(\frac{1+e}{1-e}\right)$
prolate spheroid	$2\pi b^2 + \frac{2\pi ab}{e}\sin^{-1} e$
pyramid	$\frac{1}{2}ps$
pyramidal frustum	$\frac{1}{2}ps$
sphere	$4\pi r^2$
torus	$4\pi^2 Rr$
zone	$2\pi rh$

Surface	T
cone	$\pi r(r + \sqrt{r^2 + h^2})$
conical frustum	$\pi[{R_1}^2 + {R_2}^2 + (R_1 + R_2)\sqrt{(R_1 - R_2)^2 + h^2}]$
cylinder	$2\pi r(r + h)$

Even simple surfaces can display surprisingly counterintuitive properties. For instance, the surface of revolution of $y = 1/x$ around the x-AXIS for $x \geq 1$ is called GABRIEL'S HORN, and has FINITE VOLUME but INFINITE surface AREA.

see also AREA, SURFACE INTEGRAL, SURFACE OF REVOLUTION, VOLUME

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 127–132, 1987.

Surface Integral

For a SCALAR FUNCTION f over a surface parameterized by u and v, the surface integral is given by

$$\Phi = \int_S f\, da = \int_S f(u, v)\, |\mathbf{T}_u \times \mathbf{T}_v|\, du\, dv, \tag{1}$$

where $\mathbf{T}_u$ and $\hat{\mathbf{T}}_v$ are tangent vectors and $\mathbf{a} \times \mathbf{b}$ is the CROSS PRODUCT.

For a VECTOR FUNCTION over a surface, the surface integral is given by

$$\Phi = \int_S \mathbf{F} \cdot d\mathbf{a} = \int_S (\mathbf{F} \cdot \hat{\mathbf{n}})\, da \tag{2}$$

$$= \int_S f_x\, dy\, dz + f_y\, dz\, dx + f_z\, dx\, dy, \tag{3}$$

where $\mathbf{a} \cdot \mathbf{b}$ is a DOT PRODUCT and $\hat{\mathbf{n}}$ is a unit NORMAL VECTOR. If $z = f(x, y)$, then $d\mathbf{a}$ is given explicitly by

$$d\mathbf{a} = \pm\left(-\frac{\partial z}{\partial x}\hat{\mathbf{x}} - \frac{\partial z}{\partial y}\hat{\mathbf{y}} + \hat{\mathbf{z}}\right) dx\, dy. \tag{4}$$

If the surface is SURFACE PARAMETERIZED using u and v, then

$$\Phi = \int_S \mathbf{F} \cdot (\mathbf{T}_u \times \mathbf{T}_v)\, du\, dv. \tag{5}$$

see also SURFACE PARAMETERIZATION

Surface Parameterization

A surface in 3-Space can be parameterized by two variables (or coordinates) u and v such that

$$x = x(u, v) \tag{1}$$

$$y = y(u, v) \tag{2}$$

$$z = z(u, v). \tag{3}$$

If a surface is parameterized as above, then the tangent Vectors

$$\mathbf{T}_u = \frac{\partial x}{\partial u}\hat{\mathbf{x}} + \frac{\partial y}{\partial u}\hat{\mathbf{y}} + \frac{\partial z}{\partial u}\hat{\mathbf{z}} \tag{4}$$

$$\mathbf{T}_v = \frac{\partial x}{\partial v}\hat{\mathbf{x}} + \frac{\partial y}{\partial v}\hat{\mathbf{y}} + \frac{\partial z}{\partial v}\hat{\mathbf{z}} \tag{5}$$

are useful in computing the Surface Area and Surface Integral.

see also Smooth Surface, Surface Area, Surface Integral

Surface of Revolution

A surface of revolution is a Surface generated by rotating a 2-D Curve about an axis. The resulting surface therefore always has azimuthal symmetry. Examples of surfaces of revolution include the Apple, Cone (excluding the base), Conical Frustum (excluding the ends), Cylinder (excluding the ends), Darwin-de Sitter Spheroid, Gabriel's Horn, Hyperboloid, Lemon, Oblate Spheroid, Paraboloid, Prolate Spheroid, Pseudosphere, Sphere, Spheroid, and Torus (and its generalization, the Toroid).

The standard parameterization of a surface of revolution is given by

$$x(u, v) = \phi(v)\cos u \tag{1}$$

$$y(u, v) = \phi(v)\sin u \tag{2}$$

$$z(u, v) = \psi(v). \tag{3}$$

For a curve so parameterized, the first Fundamental Form has

$$E = \psi^2 \tag{4}$$

$$F = 0 \tag{5}$$

$$G = \phi'^2 + \psi'^2. \tag{6}$$

Wherever ϕ and $\phi'^2 + \psi'^2$ are nonzero, then the surface is regular and the second Fundamental Form has

$$e = -\frac{|\phi|\psi'}{\sqrt{\phi'^2 + \psi'^2}} \tag{7}$$

$$f = 0 \tag{8}$$

$$g = \frac{\operatorname{sgn}(\phi)(\phi''\psi' - \phi'\psi'')}{\sqrt{\phi'^2 + \psi'^2}}. \tag{9}$$

Furthermore, the unit Normal Vector is

$$\hat{\mathbf{N}}(u, v) = \frac{\operatorname{sgn}(\phi)}{\sqrt{\phi'^2 + \psi'^2}} \begin{bmatrix} \phi'\cos u \\ \psi'\sin u \\ \phi' \end{bmatrix}, \tag{10}$$

and the Principal Curvatures are

$$\kappa_1 = \frac{g}{G} = \frac{\operatorname{sgn}(\phi)(\phi''\psi' - \phi'\psi'')}{(\phi'^2 + \psi'^2)^{3/2}} \tag{11}$$

$$\kappa_2 = \frac{e}{E} = -\frac{\psi'}{|\phi|\sqrt{\phi'^2 + \psi'^2}}. \tag{12}$$

The Gaussian and Mean Curvatures are

$$K = \frac{-\psi'^2\phi'' + \phi'\psi'\psi''}{\phi(\phi'^2 + \psi'^2)^2} \tag{13}$$

$$H = \frac{\phi(\phi''\psi' - \phi'\psi'') - \psi'(\phi'^2 + \psi'^2)}{2|\phi|(\phi'^2 + \psi'^2)^{3/2}} \tag{14}$$

(Gray 1993).

Pappus's Centroid Theorem gives the Volume of a solid of rotation as the cross-sectional Area times the distance traveled by the centroid as it is rotated.

Calculus of Variations can be used to find the curve from a point (x_1, y_1) to a point (x_2, y_2) which, when revolved around the x-Axis, yields a surface of smallest Surface Area A (i.e., the Minimal Surface). This is equivalent to finding the Minimal Surface passing through two circular wire frames. The Area element is

$$dA = 2\pi y\, ds = 2\pi y\sqrt{1 + y'^2}\, dx, \tag{15}$$

so the Surface Area is

$$A = 2\pi \int y\sqrt{1 + y'^2}\, dx, \tag{16}$$

and the quantity we are minimizing is

$$f = y\sqrt{1 + y'^2}. \tag{17}$$

This equation has $f_x = 0$, so we can use the Beltrami Identity

$$f - y_x \frac{\partial f}{\partial y_x} = a \tag{18}$$

to obtain

$$y\sqrt{1 + y'^2} - y'\frac{yy'}{\sqrt{1 + y'^2}} = a \tag{19}$$

$$y(1 + y'^2) - yy'^2 = a\sqrt{1 + y'^2} \tag{20}$$

$$y = a\sqrt{1 + y'^2} \tag{21}$$

$$\frac{y}{\sqrt{1 + y'^2}} = a \tag{22}$$

$$\frac{y^2}{a} - 1 = y'^2 \tag{23}$$

$$\frac{dx}{dy} = \frac{1}{y'} = \frac{a}{\sqrt{y^2 - a^2}} \tag{24}$$

$$x = a \int \frac{dy}{\sqrt{y^2 - a^2}} = a \cosh^{-1}\left(\frac{y}{a}\right) + b \tag{25}$$

$$y = a \cosh\left(\frac{x - b}{a}\right), \tag{26}$$

which is called a CATENARY, and the surface generated by rotating it is called a CATENOID. The two constants a and b are determined from the two implicit equations

$$y_1 = a \cosh\left(\frac{x_1 - b}{a}\right) \tag{27}$$

$$y_2 = a \cosh\left(\frac{x_2 - b}{a}\right), \tag{28}$$

which cannot be solved analytically.

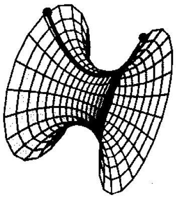

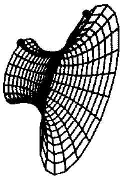

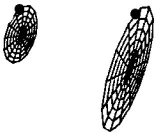

The general case is somewhat more complicated than this solution suggests. To see this, consider the MINIMAL SURFACE between two rings of equal RADIUS y_0. Without loss of generality, take the origin at the midpoint of the two rings. Then the two endpoints are located at $(-x_0, y_0)$ and (x_0, y_0), and

$$y_0 = a \cosh\left(\frac{-x_0 - b}{a}\right) = a \cosh\left(\frac{x_0 - b}{a}\right). \tag{29}$$

But $\cosh(-x) = \cosh(x)$, so

$$\cosh\left(\frac{-x_0 - b}{a}\right) = \cosh\left(\frac{-x_0 + b}{a}\right). \tag{30}$$

Inverting each side

$$-x_0 - b = -x_0 + b, \tag{31}$$

so $b = 0$ (as it must by symmetry, since we have chosen the origin between the two rings), and the equation of the MINIMAL SURFACE reduces to

$$y = a \cosh\left(\frac{x}{a}\right). \tag{32}$$

At the endpoints

$$y_0 = a \cosh\left(\frac{x_0}{a}\right), \tag{33}$$

but for certain values of x_0 and y_0, this equation has no solutions. The physical interpretation of this fact is that the surface breaks and forms circular disks in each ring to minimize AREA. CALCULUS OF VARIATIONS cannot be used to find such discontinuous solutions (known in this case as GOLDSCHMIDT SOLUTIONS). The minimal surfaces for several choices of endpoints are shown above. The first two cases are CATENOIDS, while the third case is a GOLDSCHMIDT SOLUTION.

To find the maximum value of x_0/y_0 at which CATENARY solutions can be obtained, let $p \equiv 1/a$. Then (31) gives

$$y_0 p = \cosh(p x_0). \tag{34}$$

Now, denote the maximum value of x_0 as x_0^*. Then it will be true that $dx_0/dp = 0$. Take d/dp of (34),

$$y_0 = \sinh(p x_0)\left(x_0 + p\frac{dx_0}{dp}\right). \tag{35}$$

Now set $dx_0/dp = 0$

$$y_0 = x_0 \sinh(p x_0^*). \tag{36}$$

From (34),

$$p y_0{}^* = \cosh(p x_0{}^*). \tag{37}$$

Take (37) ÷ (36),

$$p x_0^* = \coth(p x_0^*). \tag{38}$$

Defining $u \equiv p x_0{}^*$,

$$u = \coth u. \tag{39}$$

This has solution $u = 1.1996789403\ldots$. From (36), $y_0 p = \cosh u$. Divide this by (39) to obtain $y_0/x_0 = \sinh u$, so the maximum possible value of x_0/y_0 is

$$\frac{x_0}{y_0} = \operatorname{csch} u = 0.6627434193\ldots. \tag{40}$$

Therefore, only Goldschmidt ring solutions exist for $x_0/y_0 > 0.6627\ldots$.

The SURFACE AREA of the minimal CATENOID surface is given by

$$A = 2(2\pi)\int_0^{x_0} y\sqrt{1 + y'^2}\, dx, \tag{41}$$

but since

$$y = \sqrt{1 + y'^2}\, a \tag{42}$$

$$y = a \cosh\left(\frac{x}{a}\right), \tag{43}$$

$$\begin{aligned}A &= \frac{4\pi}{a}\int_0^{x_0} y^2\,dx = 4\pi a\int_0^{x_0}\cosh^2\left(\frac{x}{a}\right)dx\\ &= 4\pi a\int_0^{x_0}\tfrac{1}{2}\left[\cosh\left(\frac{2x}{a}\right)+1\right]dx\\ &= 2\pi a\left[\int_0^{x_0}\cosh\left(\frac{2x}{a}\right)dx+\int_0^{x_0}dx\right]\\ &= 2\pi a\left[\frac{a}{2}\sinh\left(\frac{2x}{a}\right)+x\right]_0^{x_0}\\ &= \pi a^2\left[\sinh\left(\frac{2x}{a}\right)+\frac{2x}{a}\right]_0^{x_0}\\ &= \pi a^2\left[\sinh\left(\frac{2x_0}{a}\right)+\frac{2x_0}{a}\right].\end{aligned}\tag{44}$$

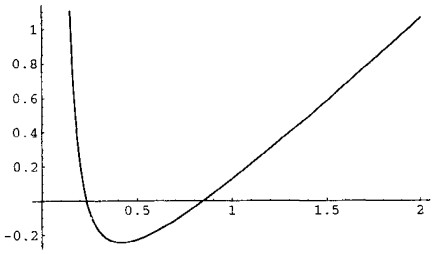

Some caution is needed in solving (33) for a. If we take $x_0 = 1/2$ and $y_0 = 1$ then (33) becomes

$$1 = a\cosh\left(\frac{1}{2a}\right),\tag{45}$$

which has *two* solutions: $a_1 = 0.2350\ldots$ ("deep"), and $a_2 = 0.8483\ldots$ ("flat"). However, upon plugging these into (44) with $x_0 = 1/2$, we find $A_1 = 6.8456\ldots$ and $A_2 = 5.9917\ldots$. So A_1 is *not,* in fact, a local minimum, and A_2 is the only *true* minimal solution.

The SURFACE AREA of the CATENOID solution equals that of the GOLDSCHMIDT SOLUTION when (44) equals the AREA of two disks,

$$\pi a^2\left[\sinh\left(\frac{2x_0}{a}\right)+\frac{2x_0}{a}\right] = 2\pi{y_0}^2\tag{46}$$

$$a^2\left[2\sinh\left(\frac{x_0}{a}\right)\cosh\left(\frac{x_0}{a}\right)+\frac{2x_0}{a}\right]-2{y_0}^2 = 0\tag{47}$$

$$a^2\left[\cosh\left(\frac{x_0}{a}\right)\sqrt{\cosh^2\left(\frac{x_0}{a}\right)-1}+\frac{x_0}{a}\right]-{y_0}^2 = 0.\tag{48}$$

Plugging in

$$\frac{y_0}{a} = \cosh\left(\frac{x_0}{a}\right),\tag{49}$$

$$\frac{y_0}{a}\sqrt{\left(\frac{y_0}{a}\right)^2-1}+\cosh^{-1}\left(\frac{y_0}{a}\right)-\left(\frac{y_0}{a}\right)^2 = 0.\tag{50}$$

Defining

$$u \equiv \frac{y_0}{a}\tag{51}$$

gives

$$u\sqrt{u^2-1}+\cosh^{-1}u-u^2 = 0.\tag{52}$$

This has a solution $u = 1.2113614259$. The value of x_0/y_0 for which

$$A_{\text{catenary}} = A_{2\text{ disks}}\tag{53}$$

is therefore

$$\frac{x_0}{y_0} = \frac{\frac{x_0}{a}}{\frac{y_0}{a}} = \frac{\cosh^{-1}\left(\frac{y_0}{a}\right)}{\frac{y_0}{a}} = \frac{\cosh^{-1}u}{u} = 0.5276973967.\tag{54}$$

For $x_0/y_0 \in (0.52770, 0.6627)$, the CATENARY solution has larger AREA than the two disks, so it exists only as a RELATIVE MINIMUM.

There also exist solutions with a disk (of radius r) between the rings supported by two CATENOIDS of revolution. The AREA is larger than that for a simple CATENOID, but it is a RELATIVE MINIMUM. The equation of the POSITIVE half of this curve is

$$y = c_1\cosh\left(\frac{x}{c_1}+c_3\right).\tag{55}$$

At $(0, r)$,

$$r = c_1\cosh(c_3).\tag{56}$$

At (x_0, y_0),

$$y_0 = c_1\cosh\left(\frac{x_0}{c_1}+c_3\right).\tag{57}$$

The AREA of the two CATENOIDS is

$$\begin{aligned}A_{\text{catenoids}} &= 2(2\pi)\int_0^{x_0} y\sqrt{1+y'^2}\,dx = \frac{4\pi}{c_1}\int_0^{x_0} y^2\,dx\\ &= 4\pi c_1\int_0^{x_0}\cosh^2\left(\frac{x}{c_1}+c_3\right)dx.\end{aligned}\tag{58}$$

Now let $u \equiv x/c_1 + c_3$, so $du = dx/c_1$

$$\begin{aligned}A &= 4\pi{c_1}^2\int_{c_3}^{x_0/x_1+c_3}\cosh^2 u\,du\\ &= 4\pi{c_1}^2\tfrac{1}{2}\int_{c_3}^{x_0/x_1+c_3}[\cosh(2u)+1]\,du\\ &= 2\pi{c_1}^2\left[\tfrac{1}{2}\sinh(2u)+u\right]_{c_3}^{x_0/x_1+c_3}\\ &= 2\pi{c_1}^2\left\{\tfrac{1}{2}\sinh\left[2\left(\frac{x_0}{c_1}+c_3\right)\right]-\tfrac{1}{2}\sinh(2c_3)+\frac{x_0}{c_1}\right\}\\ &= \pi{c_1}^2\left\{\sinh\left[2\left(\frac{x_0}{c_1}+c_3\right)\right]-\sinh(2c_3)+\frac{2x_0}{c_1}\right\}.\end{aligned}\tag{59}$$

The AREA of the central DISK is

$$A_{\text{disk}} = \pi r^2 = \pi{c_1}^2\cosh^2 c_3,\tag{60}$$

so the total AREA is

$$\begin{aligned} A = \pi {c_1}^2 \biggl\{ & \sinh\left[2\left(\frac{x_0}{c_1}+c_3\right)\right] \\ & + \left[\cosh^2 c_3 - \sinh(2c_3)\right] + \frac{2x_0}{c_1}\biggr\}. \end{aligned} \tag{61}$$

By PLATEAU'S LAWS, the CATENOIDS meet at an ANGLE of 120°, so

$$\begin{aligned} \tan 30^\circ &= \left[\frac{dy}{dx}\right]_{x=0} = \left[\sinh\left(\frac{x}{c_1}+c_3\right)\right]_{x=0} \\ &= \sinh c_3 = \frac{1}{\sqrt{3}} \end{aligned} \tag{62}$$

and

$$c_3 = \sinh^{-1}\left(\frac{1}{\sqrt{3}}\right). \tag{63}$$

This means that

$$\begin{aligned} & \cosh^2 c_3 - \sinh(2c_3) \\ &\quad = [1+\sinh^2 c_3] - 2\sinh c_3 \sqrt{1+\sinh^2 c_3} \\ &\quad = (1+\tfrac{1}{3}) - 2\left(\frac{1}{\sqrt{3}}\right)\sqrt{1+\tfrac{1}{3}} \\ &\quad = \frac{4}{3} - \frac{2}{\sqrt{3}}\frac{2}{\sqrt{3}} = 0, \end{aligned} \tag{64}$$

so

$$A = \pi {c_1}^2 \left\{\sinh\left[2\left(\frac{x_0}{c_1}+c_3\right)\right] + \frac{2x_0}{c_1}\right\}. \tag{65}$$

Now examine x_0/y_0,

$$\frac{x_0}{y_0} = \frac{\frac{x_0}{c_1}}{\frac{y_0}{c_1}} = \frac{\frac{x_0}{c_1}}{\cosh\left(\frac{x_0}{c_1}+c_3\right)} = u \operatorname{sech}(u+c_3), \tag{66}$$

where $u \equiv x_0/c_1$. Finding the maximum ratio of x_0/y_0 gives

$$\frac{d}{du}\left(\frac{x_0}{y_0}\right) = \operatorname{sech}(u+c_3) - u\tanh(u+c_3)\operatorname{sech}(u+c_3) = 0 \tag{67}$$

$$u\tanh(u+c_3) = 1, \tag{68}$$

with $c_3 = \sinh^{-1}(1/\sqrt{3})$ as given above. The solution is $u = 1.0799632187$, so the maximum value of x_0/y_0 for two CATENOIDS with a central disk is $y_0 = 0.4078241702$.

If we are interested instead in finding the curve from a point (x_1, y_1) to a point (x_2, y_2) which, when revolved around the y-AXIS (as opposed to the x-AXIS), yields a surface of smallest SURFACE AREA A, we proceed as above. Note that the solution is physically equivalent to that for rotation about the x-AXIS, but takes on a different mathematical form. The AREA element is

$$dA = 2\pi x\, ds = 2\pi x\sqrt{1+y'^2}\, dx \tag{69}$$

$$A = 2\pi \int x\sqrt{1+y'^2}\, dx, \tag{70}$$

and the quantity we are minimizing is

$$f = x\sqrt{1+y'^2}. \tag{71}$$

Taking the derivatives gives

$$\frac{\partial f}{\partial y} = 0 \tag{72}$$

$$\frac{d}{dx}\frac{\partial f}{\partial y'} = \frac{d}{dx}\left(\frac{xy'}{\sqrt{1+y'^2}}\right), \tag{73}$$

so the EULER-LAGRANGE DIFFERENTIAL EQUATION becomes

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'} = \frac{d}{dx}\left(\frac{xy'}{\sqrt{1+y'^2}}\right) = 0. \tag{74}$$

$$\frac{xy'}{\sqrt{1+y'^2}} = a \tag{75}$$

$$x^2 y'^2 = a^2(1+y'^2) \tag{76}$$

$$y'^2(x^2-a^2) = a^2 \tag{77}$$

$$\frac{dy}{dx} = \frac{a}{\sqrt{x^2-a^2}} \tag{78}$$

$$y = a\int \frac{dx}{\sqrt{x^2-a^2}} + b = a\cosh^{-1}\left(\frac{x}{a}\right) + b. \tag{79}$$

Solving for x then gives

$$x = a\cosh\left(\frac{y-b}{a}\right), \tag{80}$$

which is the equation for a CATENARY. The SURFACE AREA of the CATENOID product by rotation is

$$\begin{aligned} A &= 2\pi \int x\sqrt{1+y'^2}\, dx = 2\pi \int x\sqrt{1+\frac{a^2}{x^2-a^2}}\, dx \\ &= 2\pi \int \frac{x}{\sqrt{x^2-a^2}}\sqrt{(x^2-a^2)+a^2}\, dx \\ &= 2\pi \int \frac{x^2\, dx}{\sqrt{x^2-a^2}} \\ &= \left[\frac{x}{2}\sqrt{x^2-a^2} + \frac{a^2}{2}\ln\left(x+\sqrt{x^2-a^2}\right)\right]_{x_1}^{x_2} \\ &= \frac{1}{2}\biggl[x_2\sqrt{{x_2}^2-a^2} - x_1\sqrt{{x_1}^2-a^2} \\ &\quad + a^2\ln\left(\frac{x_2+\sqrt{{x_2}^2-a^2}}{x_1+\sqrt{{x_1}^2-a^2}}\right)\biggr]. \end{aligned} \tag{81}$$

Isenberg (1992, p. 80) discusses finding the MINIMAL SURFACE passing through two rings with axes offset from each other.

see also APPLE, CATENOID, CONE CONICAL FRUSTUM, CYLINDER, DARWIN-DE SITTER SPHEROID, EIGHT SURFACE, GABRIEL'S HORN, HYPERBOLOID, LEMON, MERIDIAN, OBLATE SPHEROID, PAPPUS'S CENTROID THEOREM, PARABOLOID, PARALLEL (SURFACE OF REVOLUTION), PROLATE SPHEROID, PSEUDOSPHERE, SINCLAIR'S SOAP FILM PROBLEM, SOLID OF REVOLUTION, SPHERE, SPHEROID, TOROID, TORUS

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 931–937, 1985.

Goldstein, H. *Classical Mechanics, 2nd ed.* Reading, MA: Addison-Wesley, p. 42, 1980.

Gray, A. "Surfaces of Revolution." Ch. 18 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 357–375, 1993.

Isenberg, C. *The Science of Soap Films and Soap Bubbles.* New York: Dover, pp. 79–80 and Appendix III, 1992.

Surface of Section

A surface (or "space") of section is a way of presenting a trajectory in n-D PHASE SPACE in an $(n-1)$-D SPACE. By picking one phase element constant and plotting the values of the other elements each time the selected element has the desired value, an intersection surface is obtained. If the equations of motion can be formulated as a MAP in which an explicit FORMULA gives the values of the other elements at successive passages through the selected element value, the time required to compute the surface of section is greatly reduced.

see also PHASE SPACE

Surgery

In the process of attaching a k-HANDLE to a MANIFOLD M, the BOUNDARY of M is modified by a process called $(k-1)$-surgery. Surgery consists of the removal of a TUBULAR NEIGHBORHOOD of a $(k-1)$-SPHERE $\mathbb{S}^{k-1}$ from the BOUNDARIES of M and the $\dim(M)-1$ standard SPHERE, and the gluing together of these two scarred-up objects along their common BOUNDARIES.

see also BOUNDARY, DEHN SURGERY, HANDLE, MANIFOLD, SPHERE, TUBULAR NEIGHBORHOOD

Surjection

An ONTO (SURJECTIVE) MAP.

see also BIJECTION, INJECTION, ONTO

Surjective

see ONTO

Surprise Examination Paradox

see UNEXPECTED HANGING PARADOX

Surreal Number

The most natural collection of numbers which includes both the REAL NUMBERS and the infinite ORDINAL NUMBERS of Georg Cantor. They were invented by John H. Conway in 1969. Every REAL NUMBER is surrounded by surreals, which are closer to it than any REAL NUMBER. Knuth (1974) describes the surreal numbers in a work of fiction.

The surreal numbers are written using the NOTATION $\{a|b\}$, where $\{|\}=0$, $\{0|\}=1$ is the simplest number greater than 0, $\{1|\}=2$ is the simplest number greater than 1, etc. Similarly, $\{|0\}=-1$ is the simplest number less than 1, etc. However, 2 can also be represented by $\{1|3\}$, $\{3/2|4\}$, $\{1|\omega\}$, etc.

see also OMNIFIC INTEGER, ORDINAL NUMBER, REAL NUMBER

References

Berlekamp, E. R.; Conway, J. H.; and Guy, R. K. *Winning Ways, For Your Mathematical Plays, Vol. 1: Games in General.* London: Academic Press, 1982.

Conway, J. H. *On Numbers and Games.* New York: Academic Press, 1976.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 283–284, 1996.

Conway, J. H. and Jackson, A. "Budding Mathematician Wins Westinghouse Competition." *Not. Amer. Math. Soc.* **43**, 776–779, 1996.

Gonshor, H. *An Introduction to Surreal Numbers.* Cambridge: Cambridge University Press, 1986.

Knuth, D. *Surreal Numbers: How Two Ex-Students Turned on to Pure Mathematics and Found Total Happiness.* Reading, MA: Addison-Wesley, 1974. `http://www-cs-faculty.stanford.edu/~knuth/sn.html`.

Surrogate

Surrogate data are artificially generated data which mimic statistical properties of real data. Isospectral surrogates have identical POWER SPECTRA as real data but with randomized phases. Scrambled surrogates have the same probability distribution as real data, but with white noise POWER SPECTRA.

see also POWER SPECTRUM

Surveying Problems

see HANSEN'S PROBLEM, SNELLIUS-POTHENOT PROBLEM

Survivorship Curve

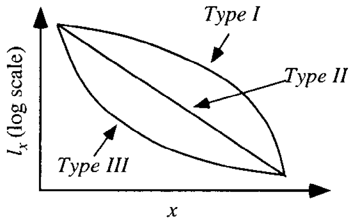

Plotting l_x from a LIFE EXPECTANCY table on a logarithmic scale versus x gives a curve known as a survivorship curve. There are three general classes of survivorship curves, illustrated above.

1. Type I curves are typical of populations in which most mortality occurs among the elderly (e.g., humans in developed countries).
2. Type II curves occur when mortality is not dependent on age (e.g., many species of large birds and fish). For an infinite type II population, $e_0 = e_1 = \ldots$, but this cannot hold for a finite population.
3. Type III curves occur when juvenile mortality is extremely high (e.g., plant and animal species producing many offspring of which few survive). In type III populations, it is often true that $e_{i+1} > e_i$ for small i. In other words, life expectancy increases for individuals who survive their risky juvenile period.

see also LIFE EXPECTANCY

Suslin's Theorem

A SET in a POLISH SPACE is a BOREL SET IFF it is both ANALYTIC and COANALYTIC. For subsets of w, a set is δ_1^1 IFF it is "hyperarithmetic."

see also ANALYTIC SET, BOREL SET, COANALYTIC SET, POLISH SPACE

Suspended Knot

An ordinary KNOT in 3-D suspended in 4-D to create a knotted 2-sphere. Suspended knots are not smooth at the poles.

see also SPUN KNOT, TWIST-SPUN KNOT

Suspension

The JOIN of a TOPOLOGICAL SPACE X and a pair of points S^0, $\Sigma(X) = X * S^0$.

see also JOIN (SPACES), TOPOLOGICAL SPACE

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 6, 1976.

Suzanne Set

The nth Suzanne set S_n is defined as the set of COMPOSITE NUMBERS x for which $n|S(x)$ and $n|S_p(x)$, where

$$x = a_0 + a_1(10^1) + \ldots + a_d(10^d) = p_1p_2\cdots p_n,$$

and

$$S(x) = \sum_{j=0}^{d} a_j$$

$$S_p(x) = \sum_{i=1}^{m} S(p_i).$$

Every Suzanne set has an infinite number of elements. The Suzanne set S_n is a superset of the MONICA SET M_n.

see also MONICA SET

References

Smith, M. "Cousins of Smith Numbers: Monica and Suzanne Sets." *Fib. Quart.* **34**, 102–104, 1996.

Suzuki Group

The SPORADIC GROUP *Suz.*

References

Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas/Suz.html`.

Swallowtail Catastrophe

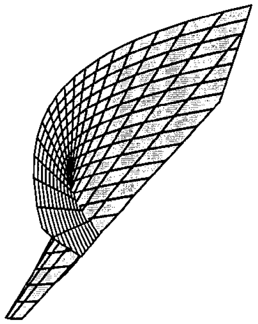

A CATASTROPHE which can occur for three control factors and one behavior axis. The equations

$$\begin{aligned} x &= uv^2 + 3v^4 \\ y &= -2uv - 4v^3 \\ z &= u \end{aligned}$$

display such a catastrophe (von Seggern 1993, Nordstrand). The above surface uses $u \in [-2, 2]$ and $v \in [-0.8, 0.8]$.

References

Nordstrand, T. "Swallowtail." `http://www.uib.no/people/nfytn/stltxt.htm`.

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 94, 1993.

Swastika

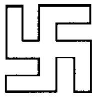

An irregular ICOSAGON, also called the gammadion or fylfot, which symbolized good luck in ancient Arabic and Indian cultures. In more recent times, it was adopted as the symbol of the Nazi Party in Hitler's Germany and has thence come to symbolize anti-Semitism.

see also CROSS, DISSECTION

Swastika Curve

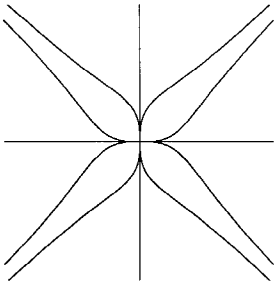

The plane curve with Cartesian equation

$$y^4 - x^4 = xy$$

and polar equation

$$r^2 = \frac{\sin\theta\cos\theta}{\sin^4\theta - \cos^4\theta}.$$

References
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 71, 1989.

Sweep Signal

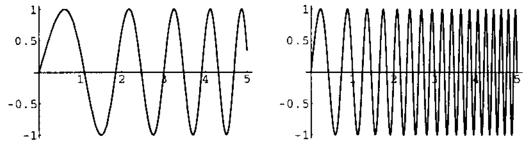

The general function

$$y(a,b,c,d) = c\sin\left\{\frac{\pi}{b-a}\left[\left((b-a)\frac{x}{d}+a\right)^2 - a^2\right]\right\}.$$

References
von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 160, 1993.

Swinnerton-Dyer Conjecture

In the early 1960s, B. Birch and H. P. F. Swinnerton-Dyer conjectured that if a given ELLIPTIC CURVE has an infinite number of solutions, then the associated L-function has value 0 at a certain fixed point. In 1976, Coates and Wiles showed that elliptic curves with COMPLEX multiplication having an infinite number of solutions have L-functions which are zero at the relevant fixed point (COATES-WILES THEOREM), but they were unable to prove the converse. V. Kolyvagin extended this result to modular curves.

see also COATES-WILES THEOREM, ELLIPTIC CURVE

References
Cipra, B. "Fermat Prover Points to Next Challenges." *Science* **271**, 1668–1669, 1996.
Ireland, K. and Rosen, M. "New Results on the Birch–Swinnerton-Dyer Conjecture." §20.5 in *A Classical Introduction to Modern Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 353–357, 1990.
Mazur, B. and Stevens, G. (Eds.). *p-Adic Monodromy and the Birch and Swinnerton-Dyer Conjecture.* Providence, RI: Amer. Math. Soc., 1994.

Swinnerton-Dyer Polynomial

The minimal POLYNOMIAL $S_n(x)$ whose ROOTS are sums and differences of the SQUARE ROOTS of the first n PRIMES,

$$S_n(x) = \prod(x \pm \sqrt{2} \pm \sqrt{3} \pm \sqrt{5} \pm \ldots \pm \sqrt{p_n}).$$

References
Vardi, I. *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 11 and 225–226, 1991.

Swirl

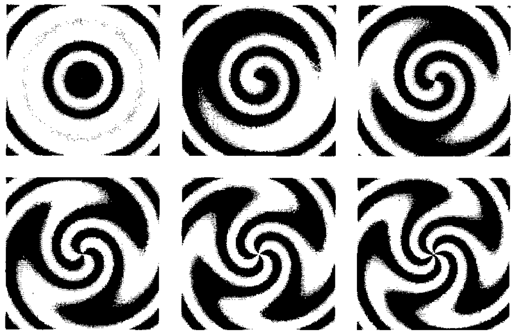

A swirl is a generic word to describe a function having arcs which double back swirl around each other. The plots above correspond to the function

$$f(r,\theta) = \sin(6\cos r - n\theta)$$

for $n = 0, 1, \ldots, 5$.

see also DAISY, WHIRL

Sylow p-Subgroup

If p^k is the highest POWER of a PRIME p dividing the ORDER of a finite GROUP G, then a SUBGROUP of G of ORDER p^k is called a Sylow p-subgroup of G.

see also ABHYANKAR'S CONJECTURE, SUBGROUP, SYLOW THEOREMS

Sylow Theorems

Let p be a PRIME NUMBER, G a GROUP, and $|G|$ the order of G.

1. If p divides $|G|$, then G has a SYLOW p-SUBGROUP.
2. In a FINITE GROUP, all the SYLOW p-SUBGROUPS are isomorphic for some fixed p.
3. The number of SYLOW p-SUBGROUPS for a fixed p is CONGRUENT to 1 (mod p).

Sylvester Cyclotomic Number

Given a LUCAS SEQUENCE with parameters P and Q, discriminant $D \neq 0$, and roots α and β, the Sylvester cyclotomic numbers are

$$Q_n = \prod_r (\alpha - \zeta^r \beta),$$

where

$$\zeta \equiv \cos\left(\frac{2\pi}{n}\right) + i \sin\left(\frac{2\pi}{n}\right)$$

is a PRIMITIVE ROOT OF UNITY and the product is over all exponents r RELATIVELY PRIME to n such that $r \in [1, n)$.

see also LUCAS SEQUENCE

References

Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, p. 69, 1989.

Sylvester's Determinant Identity

$$|A|\,|A_{rs,pq}| = |A_{r,p}|\,|A_{s,q}| - |A_{r,q}|\,|A_{s,p}|,$$

where $A_{u,w}$ is the submatrix of A formed by the intersection of the subset w of columns and u of rows.

Sylvester's Four-Point Problem

Let $q(R)$ be the probability that four points chosen at random in a region R have a CONVEX HULL which is a QUADRILATERAL. For an open, convex subset of the PLANE of finite AREA,

$$0.667 \approx \tfrac{2}{3} \leq q(R) \leq 1 - \frac{35}{12\pi^2} \approx 0.704.$$

References

Schneinerman, E. and Wilf, H. S. "The Rectilinear Crossing Number of a Complete Graph and Sylvester's 'Four Point' Problem of Geometric Probability." *Amer. Math. Monthly* **101**, 939–943, 1994.

Sylvester Graph

The Sylvester graph of a configuration is the set of ORDINARY POINTS and ORDINARY LINES.

see also ORDINARY LINE, ORDINARY POINT

References

Guy, R. K. "Monthly Unsolved Problems, 1969–1987." *Amer. Math. Monthly* **94**, 961–970, 1987.

Guy, R. K. "Unsolved Problems Come of Age." *Amer. Math. Monthly* **96**, 903–909, 1989.

Sylvester's Inertia Law

The numbers of EIGENVALUES that are POSITIVE, NEGATIVE, or 0 do not change under a congruence transformation. Gradshteyn and Ryzhik (1979) state it as follows: when a QUADRATIC FORM Q in n variables is reduced by a nonsingular linear transformation to the form

$$Q = {y_1}^2 + {y_2}^2 + \ldots + {y_p}^2 - {p_{p+1}}^2 - {y_{p2}}^2 - \ldots - {y_r}^2,$$

the number p of POSITIVE SQUARES appearing in the reduction is an invariant of the QUADRATIC FORM Q and does not depend on the method of reduction.

see also EIGENVALUE, QUADRATIC FORM

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1105, 1979.

Sylvester's Line Problem

It is not possible to arrange a finite number of points so that a LINE through every two of them passes through a third, unless they are all on a single LINE.

see also COLLINEAR, SYLVESTER'S FOUR-POINT PROBLEM

Sylvester Matrix

For POLYNOMIALS of degree m and n, the Sylvester matrix is an $(m+n) \times (m+n)$ matrix whose DETERMINANT is the RESULTANT of the two POLYNOMIALS.

see also RESULTANT

Sylvester's Sequence

The sequence defined by $e_0 = 2$ and the RECURRENCE RELATION

$$e_n = 1 + \prod_{i=0}^{n-1} e_i = {e_{n-1}}^2 - e_{n-1} + 1. \qquad (1)$$

This sequence arises in Euclid's proof that there are an INFINITE number of PRIMES. The proof proceeds by constructing a sequence of PRIMES using the RECURRENCE RELATION

$$e_{n+1} = e_0 e_1 \cdots e_n + 1 \qquad (2)$$

(Vardi 1991). Amazingly, there is a constant

$$E \approx 1.264084735306 \qquad (3)$$

such that

$$e_n = \left\lfloor E^{2^{n+1}} + \tfrac{1}{2} \right\rfloor \qquad (4)$$

(Vardi 1991, Graham *et al.* 1994). The first few numbers in Sylvester's sequence are 2, 3, 7, 43, 1807, 3263443, 10650056950807, ... (Sloane's A000058). The e_n satisfy

$$\sum_{n=0}^{\infty} \frac{1}{e_n} = 1. \tag{5}$$

In addition, if $0 < x < 1$ is an IRRATIONAL NUMBER, then the nth term of an infinite sum of unit fractions used to represent x as computed using the GREEDY ALGORITHM must be smaller than $1/e_n$.

The n of the first few PRIME e_n are 0, 1, 2, 3, 5, ... Vardi (1991) gives a lists of factors less than 5×10^7 of e_n for $n \leq 200$ and shows that e_n is COMPOSITE for $6 \leq n \leq 17$. Furthermore, all numbers less than 2.5×10^{15} in Sylvester's sequence are SQUAREFREE, and no SQUAREFUL numbers in this sequence are known (Vardi 1991).

see also EUCLID'S THEOREMS, GREEDY ALGORITHM, SQUAREFREE, SQUAREFUL

References
Graham, R. L.; Knuth, D. E.; and Patashnik, O. Research problem 4.65 in *Concrete Mathematics: A Foundation for Computer Science, 2nd ed.* Reading, MA: Addison-Wesley, 1994.
Sloane, N. J. A. Sequence A000058/M0865 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Vardi, I. "Are All Euclid Numbers Squarefree?" and "`PowerMod` to the Rescue." §5.1 and 5.2 in *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 82–89, 1991.

Sylvester's Signature

Diagonalize a form over the RATIONALS to

$$\operatorname{diag}[p^a \cdot A, p^b \cdot B, \ldots],$$

where all the entries are INTEGERS and A, B, ... are RELATIVELY PRIME to p. Then Sylvester's signature is the sum of the -1-parts of the entries.

see also p-SIGNATURE

Sylvester's Triangle Problem

The resultant of the vectors represented by the three RADII from the center of a TRIANGLE'S CIRCUMCIRCLE to its VERTICES is the segment extending from the CIRCUMCENTER to the ORTHOCENTER.

see also CIRCUMCENTER, CIRCUMCIRCLE, ORTHOCENTER, TRIANGLE

References
Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, p. 142, 1965.

Symbolic Logic

The study of the meaning and relationships of statements used to represent precise mathematical ideas. Symbolic logic is also called FORMAL LOGIC.

see also FORMAL LOGIC, LOGIC, METAMATHEMATICS

References
Carnap, R. *Introduction to Symbolic Logic and Its Applications.* New York: Dover, 1958.

Symmedian Line

The lines ISOGONAL to the MEDIANS of a TRIANGLE are called the triangle's symmedian lines. The symmedian lines are concurrent in a point called the LEMOINE POINT.

see also ISOGONAL CONJUGATE, LEMOINE POINT, MEDIAN (TRIANGLE)

References
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 213–218, 1929.

Symmedian Point

see LEMOINE POINT

Symmetric

A quantity which remains unchanged in SIGN when indices are reversed. For example, $A_{ij} \equiv a_i + a_j$ is symmetric since $A_{ij} = A_{ji}$.

see also ANTISYMMETRIC

Symmetric Block Design

A symmetric design is a BLOCK DESIGN (v, k, λ, r, b) with the same number of blocks as points, so $b = v$ (or, equivalently, $r = k$). An example of a symmetric block design is a PROJECTIVE PLANE.

see also BLOCK DESIGN, PROJECTIVE PLANE

References
Dinitz, J. H. and Stinson, D. R. "A Brief Introduction to Design Theory." Ch. 1 in *Contemporary Design Theory: A Collection of Surveys* (Ed. J. H. Dinitz and D. R. Stinson). New York: Wiley, pp. 1–12, 1992.

Symmetric Design

see SYMMETRIC BLOCK DESIGN

Symmetric Function

A symmetric function on n variables x_1, ..., x_n is a function that is unchanged by any PERMUTATION of its variables. In most contexts, the term "symmetric function" refers to a polynomial on n variables with this feature (more properly called a "symmetric polynomial"). Another type of symmetric functions is symmetric rational functions, which are the RATIONAL FUNCTIONS that are unchanged by PERMUTATION of variables.

The symmetric polynomials (respectively, symmetric rational functions) can be expressed as polynomials (respectively, rational functions) in the ELEMENTARY SYMMETRIC FUNCTIONS. This is called the FUNDAMENTAL THEOREM OF SYMMETRIC FUNCTIONS.

A function $f(x)$ is sometimes said to be symmetric about the y-AXIS if $f(-x) = f(x)$. Examples of such functions include $|x|$ (the ABSOLUTE VALUE) and x^2 (the PARABOLA).

see also ELEMENTARY SYMMETRIC FUNCTION, FUNDAMENTAL THEOREM OF SYMMETRIC FUNCTIONS, RATIONAL FUNCTION

References

Macdonald, I. G. *Symmetric Functions and Hall Polynomials, 2nd ed.* Oxford, England: Oxford University Press, 1995.

Macdonald, I. G. *Symmetric Funtions and Orthogonal Polynomials.* Providence, RI: Amer. Math. Soc., 1997.

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. "Symmetric Function Identities." §1.7 in *A=B.* Wellesley, MA: A. K. Peters, pp. 12–13, 1996.

Symmetric Group

The symmetric group S_n of DEGREE n is the GROUP of all PERMUTATIONS on n symbols. S_n is therefore of ORDER $n!$ and contains as SUBGROUPS every GROUP of ORDER n. The number of CONJUGACY CLASSES of S_n is given by the PARTITION FUNCTION P.

NETTO'S CONJECTURE states that the probability that two elements P_1 and P_2 of a symmetric group generate the entire group tends to 3/4 as $n \to \infty$. This was proven by Dixon in 1967.

see also ALTERNATING GROUP, CONJUGACY CLASS, FINITE GROUP, NETTO'S CONJECTURE, PARTITION FUNCTION P, SIMPLE GROUP

References

Lomont, J. S. "Symmetric Groups." Ch. 7 in *Applications of Finite Groups.* New York: Dover, pp. 258–273, 1987.

Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas#alt`.

Symmetric Matrix

A symmetric matrix is a SQUARE MATRIX which satisfies $\mathsf{A}^{\mathrm{T}} = \mathsf{A}$ where A^{T} denotes the TRANSPOSE, so $a_{ij} = a_{ji}$. This also implies

$$\mathsf{A}^{-1}\mathsf{A}^{\mathrm{T}} = \mathsf{I}, \tag{1}$$

where I is the IDENTITY MATRIX. Written explicitly,

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}. \tag{2}$$

The symmetric part of any MATRIX may be obtained from

$$\mathsf{A}_s = \tfrac{1}{2}(\mathsf{A} + \mathsf{A}^{\mathrm{T}}). \tag{3}$$

A MATRIX A is symmetric if it can be expressed in the form

$$\mathsf{A} = \mathsf{QDQ}^{\mathrm{T}}, \tag{4}$$

where Q is an ORTHOGONAL MATRIX and D is a DIAGONAL MATRIX. This is equivalent to the MATRIX equation

$$\mathsf{AQ} = \mathsf{QD}, \tag{5}$$

which is equivalent to

$$\mathsf{A}\mathbf{Q}_n = \lambda_n \mathbf{Q}_n \tag{6}$$

for all n, where $\lambda_n = D_{nn}$. Therefore, the diagonal elements of D are the EIGENVALUES of A, and the columns of Q are the corresponding EIGENVECTORS.

see also ANTISYMMETRIC MATRIX, SKEW SYMMETRIC MATRIX

References

Nash, J. C. "Real Symmetric Matrices." Ch. 10 in *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation, 2nd ed.* Bristol, England: Adam Hilger, pp. 119–134, 1990.

Symmetric Points

Two points z and $z^S \in \mathbb{C}^*$ are symmetric with respect to a CIRCLE or straight LINE L if all CIRCLES and straight LINES passing through z and z^S are orthogonal to L. MÖBIUS TRANSFORMATIONS preserve symmetry. Let a straight line be given by a point z_0 and a unit VECTOR $e^{i\theta}$, then

$$z^S = e^{2i\theta}(z - z_0)^* + z_0.$$

Let a CIRCLE be given by center z_0 and RADIUS r, then

$$z^S = z_0 + \frac{r^2}{(z - z_0)^*}.$$

see also MÖBIUS TRANSFORMATION

Symmetric Relation

A RELATION R on a SET S is symmetric provided that for every x and y in S we have xRy IFF yRx.

see also RELATION

Symmetric Tensor

A second-RANK symmetric TENSOR is defined as a TENSOR A for which

$$A^{mn} = A^{nm}. \tag{1}$$

Any TENSOR can be written as a sum of symmetric and ANTISYMMETRIC parts

$$\begin{aligned} A^{mn} &= \tfrac{1}{2}(A^{mn} + A^{nm}) + \tfrac{1}{2}(A^{mn} - A^{nm}) \\ &= \tfrac{1}{2}({B_S}^{mn} + {B_A}^{mn}). \end{aligned} \tag{2}$$

The symmetric part of a TENSOR is denoted by parentheses as follows:

$$T_{(a,b)} \equiv \tfrac{1}{2}(T_{ab} + T_{ba}) \tag{3}$$

$$T_{(a_1,a_2,\ldots,a_n)} \equiv \frac{1}{n!} \sum_{\text{permutations}} T_{a_1a_2\cdots a_n}. \tag{4}$$

The product of a symmetric and an ANTISYMMETRIC TENSOR is 0. This can be seen as follows. Let $a^{\alpha\beta}$ be ANTISYMMETRIC, so

$$a^{11} = a^{22} = 0 \tag{5}$$

$$a^{21} = -a^{12}. \tag{6}$$

Let $b_{\alpha\beta}$ be symmetric, so

$$b_{12} = b_{21}. \tag{7}$$

Then

$$\begin{aligned} a^{\alpha\beta}b_{\alpha\beta} &= a^{11}b_{11} + a^{12}b_{12} + a^{21}b_{21} + a^{22}b_{22} \\ &= 0 + a^{12}b_{12} - a^{12}b_{12} + 0 = 0. \end{aligned} \tag{8}$$

A symmetric second-RANK TENSOR A_{mn} has SCALAR invariants

$$s_1 = A_{11} + A_{22} + A_{22} \tag{9}$$

$$\begin{aligned} s_2 = A_{22}A_{33} + A_{33}A_{11} + A_{11}A_{22} - {A_{23}}^2 \\ - {A_{31}}^2 - {A_{12}}^2. \end{aligned} \tag{10}$$

Symmetroid

A QUARTIC SURFACE which is the locus of zeros of the DETERMINANT of a SYMMETRIC 4 × 4 matrix of linear forms. A general symmetroid has 10 ORDINARY DOUBLE POINTS (Jessop 1916, Hunt 1996).

References

Hunt, B. "Algebraic Surfaces." `http://www.mathematik.uni-kl.de/~wwwagag/Galerie.html`.

Hunt, B. "Symmetroids and Weddle Surfaces." §B.5.3 in *The Geometry of Some Special Arithmetic Quotients.* New York: Springer-Verlag, pp. 315–319, 1996.

Jessop, C. *Quartic Surfaces with Singular Points.* Cambridge, England: Cambridge University Press, p. 166, 1916.

Symmetry

An intrinsic property of a mathematical object which causes it to remain invariant under certain classes of transformations (such as ROTATION, REFLECTION, INVERSION, or more abstract operations). The mathematical study of symmetry is systematized and formalized in the extremely powerful and beautiful AREA of mathematics called GROUP THEORY.

Symmetry can be present in the form of coefficients of equations as well as in the physical arrangement of objects. By classifying the symmetry of polynomial equations using the machinery of GROUP THEORY, for example, it is possible to prove the unsolvability of the general QUINTIC EQUATION.

In physics, an extremely powerful theorem of Noether states that each symmetry of a system leads to a physically conserved quantity. Symmetry under TRANSLATION corresponds to momentum conservation, symmetry under ROTATION to angular momentum conservation, symmetry in time to energy conservation, etc.

see also GROUP THEORY

References

Eppstein, D. "Symmetry and Group Theory." `http://www.ics.uci.edu/~eppstein/junkyard/sym.html`.

Farmer, D. *Groups and Symmetry.* Providence, RI: Amer. Math. Soc., 1995.

Pappas, T. "Art & Dynamic Symmetry." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 154–155, 1989.

Rosen, J. *Symmetry in Science: An Introduction to the General Theory.* New York: Springer-Verlag, 1995.

Schattschneider, D. *Visions of Symmetry: Notebooks, Periodic Drawings, and Related Work of M. C. Escher.* New York: W. H. Freeman, 1990.

Stewart, I. and Golubitsky, M. *Fearful Symmetry.* New York: Viking Penguin, 1993.

Symmetry Group

see GROUP

Symmetry Operation

Symmetry operations include the IMPROPER ROTATION, INVERSION OPERATION, MIRROR PLANE, and ROTATION. Together, these operations create 32 crystal classes corresponding to the 32 POINT GROUPS.

The INVERSION OPERATION takes

$$(x, y, z) \to (-x, -y, -z)$$

and is denoted i. When used in conjunction with a ROTATION, it becomes an IMPROPER ROTATION. An IMPROPER ROTATION by 360°$/n$ is denoted $\bar{n}$ (or S_n). For periodic crystals, the CRYSTALLOGRAPHY RESTRICTION allows only the IMPROPER ROTATIONS $\bar{1}$, $\bar{2}$, $\bar{3}$, $\bar{4}$, and $\bar{6}$.

The MIRROR PLANE symmetry operation takes

$$(x, y, z) \to (x, y, -z), (x, -y, z) \to (x, -y, z),$$

etc., which is equivalent to $\bar{2}$. Invariance under reflection can be denoted $n\sigma_v$ or $n\sigma_h$. The ROTATION symmetry operation for 360°$/n$ is denoted n (or C_n). For periodic crystals, CRYSTALLOGRAPHY RESTRICTION allows only 1, 2, 3, 4, and 6.

Symmetry operations can be indicated with symbols such as C_n, S_n, E, i, $n\sigma_v$, and $n\sigma_h$.

1. C_n indicates ROTATION about an n-fold symmetry axis.
2. S_n indicates IMPROPER ROTATION about an n-fold symmetry axis.
3. E (or I) indicates invariance under TRANSLATION.
4. i indicates a center of symmetry under INVERSION.

5. $n\sigma_v$ indicates invariance under n vertical REFLECTIONS.
6. $n\sigma_h$ indicates invariance under n horizontal REFLECTIONS.

see also CRYSTALLOGRAPHY RESTRICTION, IMPROPER ROTATION, INVERSION OPERATION, MIRROR PLANE, POINT GROUPS, ROTATION, SYMMETRY

Symmetry Principle

SYMMETRIC POINTS are preserved under a MÖBIUS TRANSFORMATION.

see also MÖBIUS TRANSFORMATION, SYMMETRIC POINTS

Symplectic Diffeomorphism

A MAP $T : (M_1, \omega_1) \to (M_2, \omega_2)$ between the SYMPLECTIC MANIFOLDS (M_1, ω_1) and (M_2, ω_2) which is a DIFFEOMORPHISM and $T^*(\omega_2) = \omega_1$ (where T^* is the PULLBACK MAP induced by T, i.e., the derivative of the DIFFEOMORPHISM T acting on tangent vectors). A symplectic diffeomorphism is also known as a SYMPLECTOMORPHISM or CANONICAL TRANSFORMATION.

see also DIFFEOMORPHISM, PULLBACK MAP, SYMPLECTIC MANIFOLD

References

Guillemin, V. and Sternberg, S. *Symplectic Techniques in Physics.* New York: Cambridge University Press, p. 34, 1984.

Symplectic Form

A symplectic form on a SMOOTH MANIFOLD M is a smooth closed 2-FORM ω on M which is nondegenerate such that at every point m, the alternating bilinear form ω_m on the TANGENT SPACE $T_m M$ is nondegenerate.

A symplectic form on a VECTOR SPACE V over F_q is a function $f(x, y)$ (defined for all $x, y \in V$ and taking values in F_q) which satisfies

$$f(\lambda_1 x_1 + \lambda_2 x_2, y) = \lambda_1 f(x_1, y) + \lambda_2 f(x_2, y),$$

$$f(y, x) = -f(x, y),$$

and

$$f(x, x) = 0.$$

Symplectic forms can exist on M (or V) only if M (or V) is EVEN-dimensional.

Symplectic Group

The symplectic group $Sp_n(q)$ for n EVEN is the GROUP of elements of the GENERAL LINEAR GROUP GL_n that preserve a given nonsingular SYMPLECTIC FORM. Any such MATRIX has DETERMINANT 1.

see also GENERAL LINEAR GROUP, LIE-TYPE GROUP, PROJECTIVE SYMPLECTIC GROUP, SYMPLECTIC FORM

References

Conway, J. H.; Curtis, R. T.; Norton, S. P.; Parker, R. A.; and Wilson, R. A. "The Groups $Sp_n(q)$ and $PSp_n(q) = S_n(q)$." §2.3 in *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups.* Oxford, England: Clarendon Press, pp. x–xi, 1985.

Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas#symp`.

Symplectic Manifold

A pair (M, ω), where M is a MANIFOLD and ω is a SYMPLECTIC FORM on M. The PHASE SPACE $\mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n$ is a symplectic manifold. Near every point on a symplectic manifold, it is possible to find a set of local "Darboux coordinates" in which the SYMPLECTIC FORM has the simple form

$$\omega = \sum_k dq_k \wedge dp_k$$

(Sjamaar 1996), where $dq_k \wedge dp_k$ is a WEDGE PRODUCT.

see also MANIFOLD, SYMPLECTIC DIFFEOMORPHISM, SYMPLECTIC FORM

References

Sjamaar, R. "Symplectic Reduction and Riemann-Roch Formulas for Multiplicities." *Bull. Amer. Math. Soc.* **33**, 327–338, 1996.

Symplectic Map

A MAP which preserves the sum of AREAS projected onto the set of (p_i, q_i) planes. It is the generalization of an AREA-PRESERVING MAP.

see also AREA-PRESERVING MAP, LIOUVILLE'S PHASE SPACE THEOREM

Symplectomorphism

see SYMPLECTIC DIFFEOMORPHISM

Synclastic

A surface on which the GAUSSIAN CURVATURE K is everywhere POSITIVE. When K is everywhere NEGATIVE, a surface is called ANTICLASTIC. A point at which the GAUSSIAN CURVATURE is POSITIVE is called an ELLIPTIC POINT.

see also ANTICLASTIC, ELLIPTIC POINT, GAUSSIAN QUADRATURE, HYPERBOLIC POINT, PARABOLIC POINT, PLANAR POINT

Synergetics

Synergetics deals with systems composed of many subsystems which may each be of a very different nature. In particular, synergetics treats systems in which cooperation among subsystems creates organized structure on macroscopic scales (Haken 1993). Examples of problems treated by synergetics include BIFURCATIONS, phase transitions in physics, convective instabilities, coherent oscillations in lasers, nonlinear oscillations in electrical circuits, population dynamics, etc.

see also Bifurcation, Chaos, Dynamical System

References

Haken, H. *Synergetics, an Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry, and Biology, 3rd rev. enl. ed.* New York: Springer-Verlag, 1983.

Haken, H. *Advanced Synergetics: Instability Hierarchies of Self-Organizing Systems and Devices.* New York: Springer-Verlag, 1993.

Mikhailov, A. S. *Foundations of Synergetics: Distributed Active Systems, 2nd ed.* New York: Springer-Verlag, 1994.

Mikhailov, A. S. and Loskutov, A. Y. *Foundations of Synergetics II: Complex Patterns, 2nd ed., enl. rev.* New York: Springer-Verlag, 1996.

Synthesized Beam

see Dirty Beam

Syntonic Comma

see Comma of Didymus

Syracuse Algorithm

see Collatz Problem

Syracuse Problem

see Collatz Problem

System of Differential Equations

see Ordinary Differential Equation

System of Equations

Let a linear system of equations be denoted

$$\mathsf{A}\mathbf{X} = \mathbf{Y}, \qquad (1)$$

where A is a Matrix and $\mathbf{X}$ and $\mathbf{Y}$ are Vectors. As shown by Cramer's Rule, there is a unique solution if A has a Matrix Inverse A^{-1}. In this case,

$$\mathbf{X} = \mathsf{A}^{-1}\mathbf{Y}. \qquad (2)$$

If $\mathbf{Y} = \mathbf{0}$, then the solution is $\mathbf{X} = \mathbf{0}$. If A has no Matrix Inverse, then the solution Subspace is either a Line or the Empty Set. If two equations are multiples of each other, solutions are of the form

$$\mathbf{X} = \mathbf{A} + t\mathbf{B}, \qquad (3)$$

for t a Real Number.

see also Cramer's Rule, Matrix Inverse

Syzygies Problem

The problem of finding all independent irreducible algebraic relations among any finite set of Quantics.

see also Quantic

Syzygy

A technical mathematical object defined in terms of a Polynomial Ring of n variables over a Field k.

see also Fundamental System, Hilbert Basis Theorem, Syzygies Problem

References

Hilbert, D. "Über die Theorie der algebraischen Formen." *Math. Ann.* **36**, 473–534, 1890.

Iyanaga, S. and Kawada, Y. (Eds.). "Syzygy Theory." §364F in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1140, 1980.

Szilassi Polyhedron

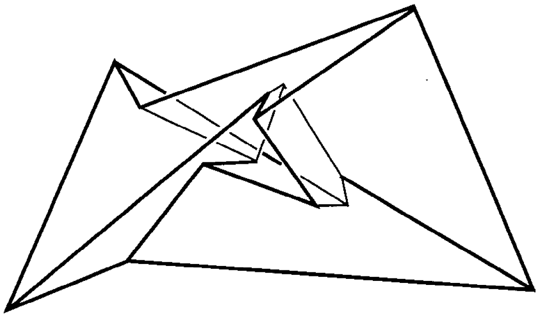

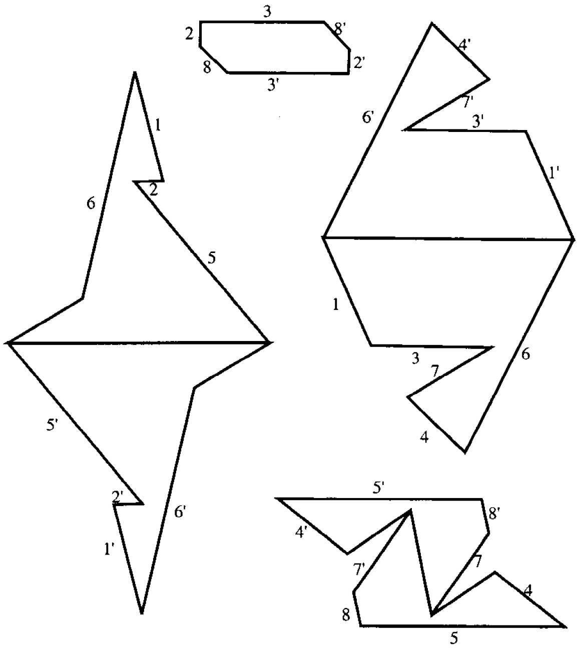

A Polyhedron which is topologically equivalent to a Torus and for which every pair of faces has an Edge in common. This polyhedron was discovered by L. Szilassi in 1977. Its Skeleton is equivalent to the seven-color torus map illustrated below.

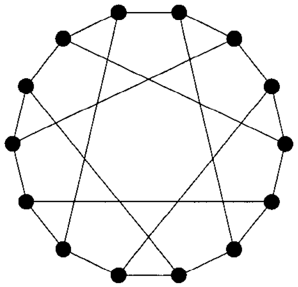

The Szilassi polyhedron has 14 VERTICES, seven faces, and 21 EDGES, and is the DUAL POLYHEDRON of the CSÁSZÁR POLYHEDRON.

see also CSÁSZÁR POLYHEDRON, TOROIDAL POLYHEDRON

References
Eppstein, D. "Polyhedra and Polytopes." `http://www.ics.uci.edu/~eppstein/junkyard/polytope.html`.
Gardner, M. *Fractal Music, Hypercards, and More Mathematical Recreations from Scientific American Magazine.* New York: W. H. Freeman, pp. 118–120, 1992.
Hart, G. "Toroidal Polyhedra." `http://www.li.net/~george/virtual-polyhedra/toroidal.html`.

Szpiro's Conjecture

A conjecture which relates the minimal DISCRIMINANT of an ELLIPTIC CURVE to the CONDUCTOR. If true, it would imply FERMAT'S LAST THEOREM for sufficiently large exponents.

see also CONDUCTOR, DISCRIMINANT (ELLIPTIC CURVE), ELLIPTIC CURVE

References
Cox, D. A. "Introduction to Fermat's Last Theorem." *Amer. Math. Monthly* **101**, 3–14, 1994.

T

t-Distribution

see STUDENT'S *t*-DISTRIBUTION

T-Polyomino

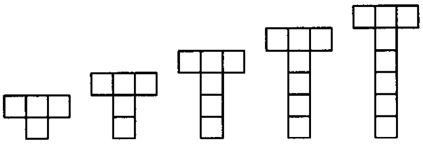

The order n T-polyomino consists of a vertical line of $n-3$ squares capped by a horizontal line of three squares centered on the line.

see also L-POLYOMINO, SKEW POLYOMINO, SQUARE POLYOMINO, STRAIGHT POLYOMINO

T-Puzzle

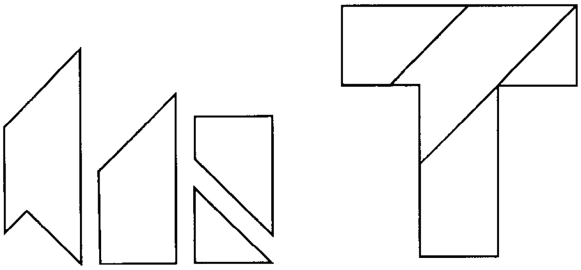

The DISSECTION of the four pieces shown at left into the capital letter "T" shown at right.

see also DISSECTION

References

Pappas, T. "The T Problem." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 35 and 230, 1989.

T2-Separation Axiom

Finite SUBSETS are CLOSED.

see also CLOSURE

Tableau

see YOUNG TABLEAU

Tabu Search

A heuristic procedure which has proven efficient at solving COMBINATORIAL optimization problems.

References

Glover, F.; Taillard, E.; and De Werra, D. "A User's Guide to Tabu Search." *Ann. Oper. Res.* **41**, 3–28, 1993.

Piwakowski, K. "Applying Tabu Search to Determine New Ramsey Numbers." *Electronic J. Combinatorics* **3**, R6, 1–4, 1996. `http://www.combinatorics.org/Volume_3/volume3.html#R6`.

Tacnode

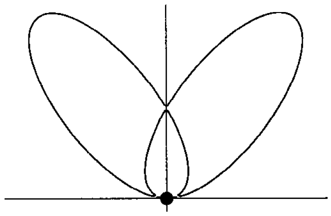

A DOUBLE POINT at which two OSCULATING CURVES are tangent. The above plot shows the tacnode of the curve $2x^4 - 3x^2y + y^2 - 2y^3 + y^4 = 0$. The LINKS CURVE also has a tacnode at the origin.

see also ACNODE, CRUNODE, DOUBLE POINT SPINODE

References

Walker, R. J. *Algebraic Curves.* New York: Springer-Verlag, pp. 57–58, 1978.

Tacpoint

A tangent point of two similar curves.

Tactix

see NIM

Tail Probability

Define T as the set of all points t with probabilities $P(x)$ such that $a > t \Rightarrow P(a \leq x \leq a + da) < P_0$ or $a < t \Rightarrow P(a \leq x \leq a + da < P_0$, where P_0 is a POINT PROBABILITY (often, the likelihood of an observed event). Then the associated tail probability is given by $\int_T P(x)\,dx$.

see also P-VALUE, POINT PROBABILITY

Tait Coloring

A 3-coloring of GRAPH EDGES so that no two EDGES of the same color meet at a VERTEX (Ball and Coxeter 1987, pp. 265–266).

see also EDGE (GRAPH), TAIT CYCLE, VERTEX (GRAPH)

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, 1987.

Tait Cycle

A set of circuits going along the EDGES of a GRAPH, each with an EVEN number of EDGES, such that just one of the circuits passes through each VERTEX (Ball and Coxeter 1987, pp. 265–266).

see also EDGE (GRAPH), EULERIAN CYCLE, HAMILTONIAN CYCLE, TAIT COLORING, VERTEX (GRAPH)

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, 1987.

Tait Flyping Conjecture

see FLYPING CONJECTURE

Tait's Hamiltonian Graph Conjecture

Every 3-connected cubic Graph (each Vertex has Valency 3) has a Hamiltonian Circuit. Proposed by Tait in 1880 and refuted by W. T. Tutte in 1946 with a counterexample, Tutte's Graph. If it had been true, it would have implied the Four-Color Theorem. A simpler counterexample was later given by Kozyrev and Grinberg.

see also Hamiltonian Circuit, Tutte's Graph, Vertex (Graph)

References
Honsberger, R. *Mathematical Gems I.* Washington, DC: Math. Assoc. Amer., pp. 82–89, 1973.

Tait's Knot Conjectures

P. G. Tait undertook a study of Knots in response to Kelvin's conjecture that the atoms were composed of knotted vortex tubes of ether (Thomson 1869). He categorized Knots in terms of the number of crossings in a plane projection. He also made some conjectures which remained unproven until the discovery of Jones Polynomials.

Tait's Flyping Conjecture states that the number of crossings is the same for any diagram of an Alternating Knot. This was proved true in 1986.

see also Alternating Knot, Flyping Conjecture, Jones Polynomial, Knot

References
Tait, P. G. "On Knots I, II, III." *Scientific Papers, Vol. 1.* London: Cambridge University Press, pp. 273–347, 1900.
Thomson, W. H. "On Vortex Motion." *Trans. Roy. Soc. Edinburgh* **25**, 217–260, 1869.

TAK Function

A Recursive Function devised by I. Takeuchi. For Integers x, y, and z, and a function h, it is

$$\mathrm{TAK}_h(x,y,z) = \begin{cases} h(x,y,z) & \text{for } x \leq y \\ h(h(x-1,y,z), h(y-1,z,x), & \text{for } x > y. \\ \quad h(z-1,x,y)) & \end{cases}$$

The number of function calls $F_0(a,b)$ required to compute $\mathrm{TAK}_0(a,b,0)$ for $a > b > 0$ is

$$F_0(a,b) = 4\sum_{k=0}^{b} \frac{a-b}{a+b-2k}\binom{a+b-2k}{b-k} - 3$$
$$= 1 + 4\sum_{k=0}^{b-1} \frac{a-b}{a+b-2k}\binom{a+b-2k}{b-k}$$

(Vardi 1991).

The TAK function is also connected with the Ballot Problem (Vardi 1991).

see also Ackermann Function, Ballot Problem

References
Gabriel, R. P. *Performance and Implementation of Lisp Systems.* Cambridge, MA: MIT Press, 1985.
Knuth, D. E. *Textbook Examples of Recursion.* Preprint 1990.
Vardi, I. "The Running Time of TAK." Ch. 9 in *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 179–199, 1991.

Takagi Fractal Curve

see Blancmange Function

Take-Away Game

see Nim-Heap

Takeuchi Function

see TAK Function

Talbot's Curve

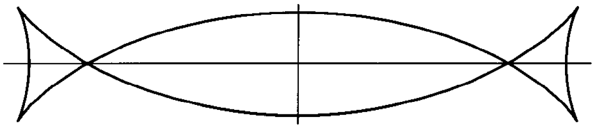

A curve investigated by Talbot which is the Negative Pedal Curve of an Ellipse with respect to its center. It has four Cusps and two Nodes, provided the Eccentricity of the Ellipse is greater than $1/\sqrt{2}$. Its Cartesian Equation is

$$x = \frac{(a^2 + f^2 \sin^2 t)\cos t}{a}$$
$$y = \frac{(a^2 - 2f^2 + f^2 \sin^2 t)\sin t}{b},$$

where f is a constant.

References
Lockwood, E. H. *A Book of Curves.* Cambridge, England: Cambridge University Press, p. 157, 1967.
MacTutor History of Mathematics Archive. "Talbot's Curve." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Talbots.html`.

Talisman Hexagon

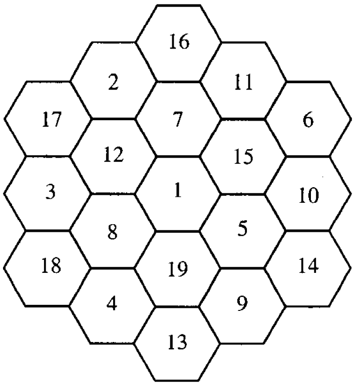

An (n,k)-talisman hexagon is an arrangement of nested hexagons containing the integers $1, 2, \ldots, H_n = 3n(n-$

1) + 1, where H_n is the nth HEX NUMBER, such that the difference between all adjacent hexagons is at least as large as k. The hexagon illustrated above is a (3, 5)-talisman hexagon.

see also HEX NUMBER, MAGIC SQUARE, TALISMAN SQUARE

References
Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 111–112, 1979.

Talisman Square

1	5	3	7
9	11	13	15
2	6	4	8
10	12	14	16

5	15	9	12
10	1	6	3
13	16	11	14
2	8	4	7

15	1	12	4	9
20	7	22	18	24
16	2	13	5	10
21	8	23	19	25
17	3	14	6	11

28	10	31	13	34	16
19	1	22	4	25	7
29	11	32	14	35	17
20	2	23	5	26	8
30	12	33	15	36	18
21	3	24	6	27	9

An $n \times n$ ARRAY of the integers from 1 to n^2 such that the difference between any one integer and its neighbor (horizontally, vertically, or diagonally, without wrapping around) is greater than or equal to some value k is called a (n, k)-talisman square. The above illustrations show (4, 2)-, (4, 3)-, (5, 4)-, and (6, 8)-talisman squares.

see also ANTIMAGIC SQUARE, HETEROSQUARE, MAGIC SQUARE, TALISMAN HEXAGON

References
Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 110–113, 1979.
Weisstein, E. W. "Magic Squares." `http://www.astro.virginia.edu/~eww6n/math/notebooks/MagicSquares.m`.

Tame Algebra

Let A denote an $\mathbb{R}$-algebra, so that A is a VECTOR SPACE over R and

$$A \times A \to A$$

$$(x, y) \mapsto x \cdot y,$$

where $x \cdot y$ is vector multiplication which is assumed to be BILINEAR. Now define

$$Z \equiv \{x \in a : x \cdot y = 0 \text{ for some nonzero } y \in A\},$$

where $0 \in Z$. A is said to be tame if Z is a finite union of SUBSPACES of A. A 2-D 0-ASSOCIATIVE algebra is tame, but a 4-D 4-ASSOCIATIVE algebra and a 3-D 1-ASSOCIATIVE algebra need not be tame. It is conjectured that a 3-D 2-ASSOCIATIVE algebra is tame, and proven that a 3-D 3-ASSOCIATIVE algebra is tame if it possesses a multiplicative IDENTITY ELEMENT.

References
Finch, S. "Zero Structures in Real Algebras." `http://www.mathsoft.com/asolve/zerodiv/zerodiv.html`.

Tame Knot

A KNOT equivalent to a POLYGONAL KNOT. Knots which are not tame are called WILD KNOTS.

References
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 49, 1976.

Tangency Theorem

The external (internal) SIMILARITY POINT of two fixed CIRCLES is the point at which all the CIRCLES homogeneously (nonhomogeneously) tangent to the fixed CIRCLES have the same POWER and at which all the tangency secants intersect.

References
Dörrie, H. *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, p. 157, 1965.

Tangent

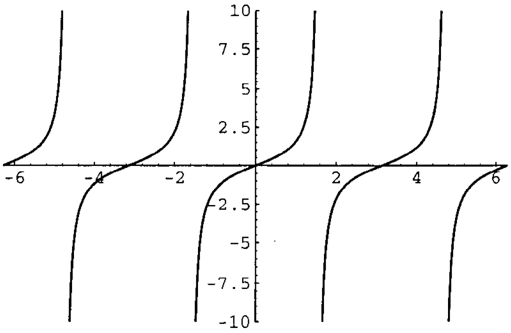

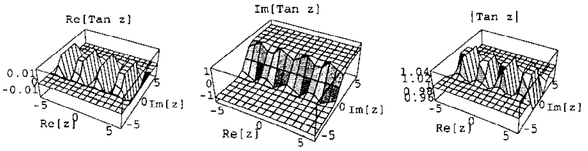

The tangent function is defined by

$$\tan\theta \equiv \frac{\sin\theta}{\cos\theta}, \tag{1}$$

where $\sin x$ is the SINE function and $\cos x$ is the COSINE function. The word "tangent," however, also has an important related meaning as a LINE or PLANE which touches a given curve or solid at a single point. These geometrical objects are then called a TANGENT LINE or TANGENT PLANE, respectively.

The MACLAURIN SERIES for the tangent function is

$$\tan x = \sum \frac{(-1)^{n-1}2^{2n}(2^{2n}-1)B_{2n}}{(2n)!}x^{2n-1} + \ldots$$
$$= x + \tfrac{1}{3}x^3 + \tfrac{2}{15}x^5 + \tfrac{17}{315}x^7 + \tfrac{62}{2835}x^9 + \ldots, \quad (2)$$

where B_n is a BERNOULLI NUMBER.

$\tan x$ is IRRATIONAL for any RATIONAL $x \neq 0$, which can be proved by writing $\tan x$ as a CONTINUED FRACTION

$$\tan x = \cfrac{x}{1 - \cfrac{x^2}{3 - \cfrac{x^2}{5 - \cfrac{x^2}{7 - \ldots}}}}. \quad (3)$$

Lambert derived another CONTINUED FRACTION expression for the tangent,

$$\tan x = \cfrac{1}{\frac{1}{x} - \cfrac{1}{\frac{3}{x} - \cfrac{1}{\frac{5}{x} - \cfrac{1}{\frac{7}{x} - \ldots}}}}. \quad (4)$$

An interesting identity involving the PRODUCT of tangents is

$$\prod_{k=1}^{\lfloor (n-1)/2 \rfloor} \tan\left(\frac{k\pi}{n}\right) = \begin{cases} \sqrt{n} & \text{for } n \text{ odd} \\ 1 & \text{for } n \text{ even,} \end{cases} \quad (5)$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION. Another tangent identity is

$$\tan(n \tan^{-1} x) = \frac{1}{i}\frac{(1+ix)^n - (1-ix)^n}{(1+ix)^n + (1-ix)^m} \quad (6)$$

(Beeler *et al.* 1972, Item 16).

see also ALTERNATING PERMUTATION, COSINE, COTANGENT, INVERSE TANGENT, MORRIE'S LAW, SINE, TANGENT LINE, TANGENT PLANE

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Circular Functions." §4.3 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 71–79, 1972.

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.

Spanier, J. and Oldham, K. B. "The Tangent tan(x) and Cotangent cot(x) Functions." Ch. 34 in *An Atlas of Functions.* Washington, DC: Hemisphere, pp. 319–330, 1987.

Tangent Bifurcation

see FOLD BIFURCATION

Tangent Bundle

The tangent bundle TM of a SMOOTH MANIFOLD M is the SPACE of TANGENT VECTORS to points in the manifold, i.e., it is the set (x, v) where $x \in M$ and v is tangent to $x \in M$. For example, the tangent bundle to the CIRCLE is the CYLINDER.

see also COTANGENT BUNDLE, TANGENT VECTOR

Tangent Developable

A RULED SURFACE M is a tangent developable of a curve $\mathbf{y}$ if M can be parameterized by $\mathbf{x}(u,v) = \mathbf{y}(u) + v\mathbf{y}'(u)$. A tangent developable is a FLAT SURFACE.

see also BINORMAL DEVELOPABLE, NORMAL DEVELOPABLE

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 341–343, 1993.

Tangent Hyperbolas Method

see HALLEY'S METHOD

Tangent Indicatrix

Let the SPEED $\boldsymbol{\sigma}$ of a closed curve on the unit sphere S^2 never vanish. Then the tangent indicatrix

$$\boldsymbol{\tau} \equiv \frac{\dot{\boldsymbol{\sigma}}}{|\dot{\boldsymbol{\sigma}}|}$$

is another closed curve on S^2. It is sometimes called the TANTRIX. If $\boldsymbol{\sigma}$ IMMERSES in S^2, then so will $\boldsymbol{\tau}$.

References

Solomon, B. "Tantrices of Spherical Curves." *Amer. Math. Monthly* **103**, 30–39, 1996.

Tangent Line

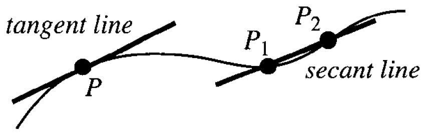

A tangent line is a LINE which meets a given curve at a single POINT.

see also CIRCLE TANGENTS, SECANT LINE, TANGENT, TANGENT PLANE, TANGENT SPACE, TANGENT VECTOR

References

Yates, R. C. "Instantaneous Center of Rotation and the Construction of Some Tangents." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 119–122, 1952.

Tangent Map

If $f : M \to N$, then the tangent map Tf associated to f is a VECTOR BUNDLE HOMEOMORPHISM $Tf : TM \to TN$ (i.e., a MAP between the TANGENT BUNDLES of M and N respectively). The tangent map corresponds to DIFFERENTIATION by the formula

$$Tf(v) = (f \circ \phi)'(0), \tag{1}$$

where $\phi'(0) = v$ (i.e., ϕ is a curve passing through the base point to v in TM at time 0 with velocity v). In this case, if $f : M \to N$ and $g : N \to O$, then the CHAIN RULE is expressed as

$$T(f \circ g) = Tf \circ Tg. \tag{2}$$

In other words, with this way of formalizing differentiation, the CHAIN RULE can be remembered by saying that "the process of taking the tangent map of a map is functorial." To a topologist, the form

$$(f \circ g)'(a) = f'(g(a)) \circ g'(a), \tag{3}$$

for all a, is more intuitive than the usual form of the CHAIN RULE.

see also DIFFEOMORPHISM

References
Gray, A. "Tangent Maps." §9.3 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 168–171, 1993.

Tangent Number

A number also called a ZAG NUMBER giving the number of EVEN ALTERNATING PERMUTATIONS. The first few are 1, 2, 16, 272, 7936, ... (Sloane's A000182).

see also ALTERNATING PERMUTATION, EULER ZIGZAG NUMBER, SECANT NUMBER

References
Knuth, D. E. and Buckholtz, T. J. "Computation of Tangent, Euler, and Bernoulli Numbers." *Math. Comput.* **21**, 663–688, 1967.
Sloane, N. J. A. Sequence A000182/M2096 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Tangent Plane

A tangent plane is a PLANE which meets a given SURFACE at a single POINT. Let (x_0, y_0) be any point of a surface function $z = f(x, y)$. The surface has a nonvertical tangent plane at (x_0, y_0) with equation

$$z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

see also NORMAL VECTOR, TANGENT, TANGENT LINE, TANGENT SPACE, TANGENT VECTOR

Tangent Space

Let x be a point in an n-dimensional COMPACT MANIFOLD M, and attach at x a copy of $\mathbb{R}^n$ tangential to M. The resulting structure is called the TANGENT SPACE of M at x and is denoted T_xM. If γ is a smooth curve passing through x, then the derivative of γ at x is a VECTOR in T_xM.

see also TANGENT, TANGENT BUNDLE, TANGENT PLANE, TANGENT VECTOR

Tangent Vector

For a curve with POSITION VECTOR $\mathbf{r}(t)$, the unit tangent vector $\hat{\mathbf{T}}(t)$ is defined by

$$\hat{\mathbf{T}}(t) \equiv \frac{\mathbf{r}'(t)}{|\mathbf{r}'(t)|} = \frac{\frac{d\mathbf{r}}{dt}}{\left|\frac{d\mathbf{r}}{dt}\right|} \tag{1}$$

$$= \frac{\frac{d\mathbf{r}}{dt}}{\frac{ds}{dt}} \tag{2}$$

$$= \frac{d\mathbf{r}}{ds}, \tag{3}$$

where t is a parameterization variable and s is the ARC LENGTH. For a function given parametrically by $(f(t), g(t))$, the tangent vector relative to the point $(f(t), g(t))$ is therefore given by

$$x(t) = \frac{f'}{\sqrt{f'^2 + g'^2}} \tag{4}$$

$$y(t) = \frac{g'}{\sqrt{f'^2 + g'^2}}. \tag{5}$$

To actually place the vector tangent to the curve, it must be displaced by $(f(t), g(t))$. It is also true that

$$\frac{d\hat{\mathbf{T}}}{ds} = \kappa\hat{\mathbf{N}} \tag{6}$$

$$\frac{d\hat{\mathbf{T}}}{dt} = \kappa\frac{ds}{dt}\hat{\mathbf{N}} \tag{7}$$

$$[\dot{\mathbf{T}}, \ddot{\mathbf{T}}, \dddot{\mathbf{T}}] = \kappa^5 \frac{d}{ds}\left(\frac{\tau}{\kappa}\right), \tag{8}$$

where $\mathbf{N}$ is the NORMAL VECTOR, κ is the CURVATURE, and τ is the TORSION.

see also CURVATURE, NORMAL VECTOR, TANGENT, TANGENT BUNDLE, TANGENT PLANE, TANGENT SPACE, TORSION (DIFFERENTIAL GEOMETRY)

References
Gray, A. "Tangent and Normal Lines to Plane Curves." §5.5 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 85–90, 1993.

Tangential Angle

For a PLANE CURVE, the tangential angle ϕ is defined by

$$\rho\, d\phi = ds, \tag{1}$$

where s is the ARC LENGTH and ρ is the RADIUS OF CURVATURE. The tangential angle is therefore given by

$$\phi = \int_0^t s'(t)\kappa(t)\, dt, \tag{2}$$

where $\kappa(t)$ is the CURVATURE. For a plane curve $\mathbf{r}(t)$, the tangential angle $\phi(t)$ can also be defined by

$$\frac{\mathbf{r}'(t)}{|\mathbf{r}'(t)|} = \begin{bmatrix} \cos[\phi(t)] \\ \sin[\phi(t)] \end{bmatrix}. \tag{3}$$

Gray (1993) calls ϕ the TURNING ANGLE instead of the tangential angle.

see also ARC LENGTH, CURVATURE, PLANE CURVE, RADIUS OF CURVATURE, TORSION (DIFFERENTIAL GEOMETRY)

References

Gray, A. "The Turning Angle." §1.6 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 13–14, 1993.

Tangential Triangle

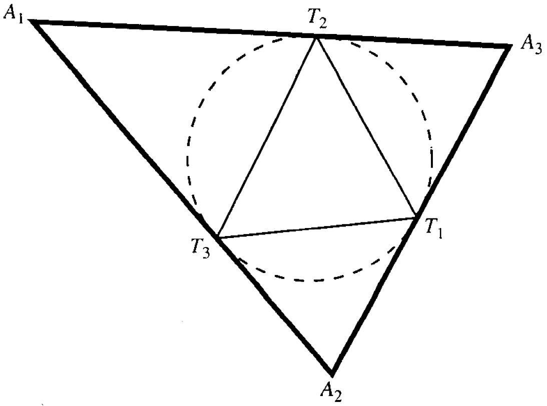

The TRIANGLE $\Delta T_1T_2T_3$ formed by the lines tangent to the CIRCUMCIRCLE of a given TRIANGLE $\Delta A_1A_2A_3$ at its VERTICES. It is the PEDAL TRIANGLE of $\Delta A_1A_2A_3$ with the CIRCUMCENTER as the PEDAL POINT. The TRILINEAR COORDINATES of the VERTICES of the tangential triangle are

$$\begin{aligned} A' &= -a : b : c \\ B' &= a : -b : c \\ C' &= a : b : -c. \end{aligned}$$

The CONTACT TRIANGLE and tangential triangle are perspective from the GERGONNE POINT.

see also CIRCUMCIRCLE, CONTACT TRIANGLE, GERGONNE POINT, PEDAL TRIANGLE, PERSPECTIVE

Tangential Triangle Circumcenter

A POINT with TRIANGLE CENTER FUNCTION

$$\alpha = a[b^2\cos(2B) + c^2\cos(2C) - a^2\cos(2A)].$$

It lies on the EULER LINE.

References

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Tangents Law

see LAW OF TANGENTS

Tangle

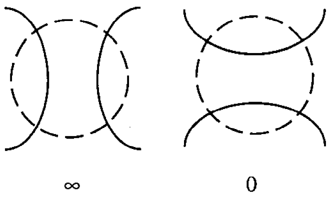

A region in a KNOT or LINK projection plane surrounded by a CIRCLE such that the KNOT or LINK crosses the circle exactly four times. Two tangles are equivalent if a sequence of REIDEMEISTER MOVES can be used to transform one into the other while keeping the four string endpoints fixed and not allowing strings to pass outside the CIRCLE.

The simplest tangles are the ∞-tangle and 0-tangle, shown above. A tangle with n left-handed twists is called an n-tangle, and one with n right-handed twists is called a $-n$-tangle. By placing tangles side by side, more complicated tangles can be built up such as $(-2, 3, 2)$, etc. The link created by connecting the ends of the tangles is now described by the sequence of tangle symbols, known as CONWAY'S KNOT NOTATION. If tangles are multiplied by 0 and then added, the resulting tangle symbols are separated by commas. Additional symbols which are used are the period, colon, and asterisk.

Amazingly enough, two tangles described in this NOTATION are equivalent IFF the CONTINUED FRACTIONS of the form

$$2 + \cfrac{1}{3 + \cfrac{1}{-2}}$$

are equal (Burde and Zieschang 1985)! An ALGEBRAIC TANGLE is any tangle obtained by ADDITIONS and MULTIPLICATIONS of rational tangles (Adams 1994). Not all tangles are ALGEBRAIC.

see also ALGEBRAIC LINK, FLYPE, PRETZEL KNOT

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman pp. 41–51, 1994.

Burde, G. and Zieschang, H. *Knots.* Berlin: de Gruyter, 1985.

Tanglecube

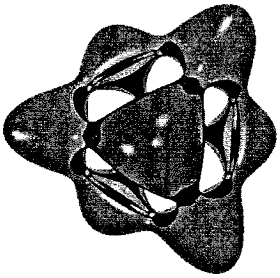

A QUARTIC SURFACE given by the implicit equation

$$x^4 - 5x^2 + y^4 - 5y^2 + z^4 - 5z^2 + 11.8 = 0.$$

References

Banchoff, T. "The Best Homework Ever?" `http://www.brown.edu/Administration/Brown_Alumni_Monthly/12-96/features/homework.html`.

Nordstrand, T. "Tangle." `http://www.uib.no/people/nfytn/tangltxt.htm`.

Tangled Hierarchy

A system in which a STRANGE LOOP appears.

see also STRANGE LOOP

References

Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, p. 10, 1989.

Tangram

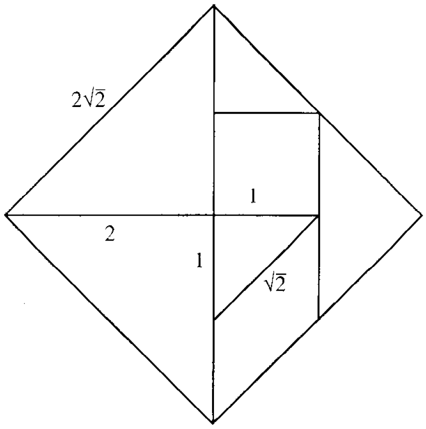

A combination of the above plane polygonal pieces such that the EDGES are coincident. There are 13 convex tangrams (where a "convex tangram" is a set of tangram pieces arranged into a CONVEX POLYGON).

see also ORIGAMI, STOMACHION

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 19–20, 1989.

Gardner, M. "Tangrams, Parts 1 and 2." Ch. 3-4 in *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, 1988.

Johnston, S. *Fun with Tangrams Kit: 120 Puzzles with Two Complete Sets of Tangram Pieces.* New York: Dover, 1977.

Pappas, T. "Tangram Puzzle." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 212, 1989.

Tanh

see HYPERBOLIC TANGENT

Taniyama Conjecture

see TANIYAMA-SHIMURA CONJECTURE

Taniyama-Shimura Conjecture

A conjecture which arose from several problems proposed by Taniyama in an international mathematics symposium in 1955. Let E be an ELLIPTIC CURVE whose equation has INTEGER COEFFICIENTS, let N be the CONDUCTOR of E and, for each n, let a_n be the number appearing in the L-function of E. Then there exists a MODULAR FORM of weight two and level N which is an eigenform under the HECKE OPERATORS and has a FOURIER SERIES $\sum a_n q^n$.

The conjecture says, in effect, that every rational ELLIPTIC CURVE is a MODULAR FORM in disguise. Stated formally, the conjecture suggests that, for every ELLIPTIC CURVE $y^2 = Ax^3 + Bx^2 + Cx + D$ over the RATIONALS, there exist nonconstant MODULAR FUNCTIONS $f(z)$ and $g(z)$ of the same level N such that

$$[f(z)]^2 = A[g(z)]^2 + Cg(z) + D.$$

Equivalently, for every ELLIPTIC CURVE, there is a MODULAR FORM with the same DIRICHLET L-SERIES.

In 1985, starting with a fictitious solution to FERMAT'S LAST THEOREM, G. Frey showed that he could create an unusual ELLIPTIC CURVE which appeared not to be modular. If the curve were not modular, then this would show that if FERMAT'S LAST THEOREM were false, then the Taniyama-Shimura conjecture would also be false. Furthermore, if the Taniyama-Shimura conjecture were true, then so would be FERMAT'S LAST THEOREM!

However, Frey did not actually prove whether his curve was modular. The conjecture that Frey's curve *was* modular came to be called the "epsilon conjecture," and was quickly proved by Ribet (RIBET'S THEOREM) in 1986, establishing a very close link between two mathematical structures (the Taniyama-Shimura conjecture and FERMAT'S LAST THEOREM) which appeared previously to be completely unrelated.

As of the early 1990s, most mathematicians believed that the Taniyama-Shimura conjecture was not accessible to proof. However, A. Wiles was not one of these. He attempted to establish the correspondence between the set of ELLIPTIC CURVES and the set of modular elliptic curves by showing that the number of each was the same. Wiles accomplished this by "counting" Galois representations and comparing them with the number of modular forms. In 1993, after a monumental seven-year effort, Wiles (almost) proved the Taniyama-Shimura conjecture for special classes of curves called SEMISTABLE ELLIPTIC CURVES.

Wiles had tried to use horizontal Iwasawa theory to create a so-called CLASS NUMBER formula, but was initially unsuccessful and therefore used instead an extension of a result of Flach based on ideas from Kolyvagin. However, there was a problem with this extension which was discovered during review of Wiles' manuscript in September 1993. Former student Richard Taylor came to Princeton in early 1994 to help Wiles patch up this error. After additional effort, Wiles discovered the reason that the Flach/Kolyvagin approach was failing, and also discovered that it was precisely what had prevented Iwasawa theory from working.

With this additional insight, he was able to successfully complete the erroneous portion of the proof using Iwasawa theory, proving the SEMISTABLE case of the Taniyama-Shimura conjecture (Taylor and Wiles 1995, Wiles 1995) and, at the same time, establishing FERMAT'S LAST THEOREM as a true theorem.

see also ELLIPTIC CURVE, FERMAT'S LAST THEOREM, MODULAR FORM, MODULAR FUNCTION, RIBET'S THEOREM

References
Lang, S. "Some History of the Shimura-Taniyama Conjecture." *Not. Amer. Math. Soc.* **42**, 1301-1307, 1995.
Taylor, R. and Wiles, A. "Ring-Theoretic Properties of Certain Hecke Algebras." *Ann. Math.* **141**, 553–572, 1995.
Wiles, A. "Modular Elliptic-Curves and Fermat's Last Theorem." *Ann. Math.* **141**, 443–551, 1995.

Tank

see CYLINDRICAL SEGMENT

Tantrix

see TANGENT INDICATRIX

Tapering Function

see APODIZATION FUNCTION

Tarry-Escott Problem

For each POSITIVE INTEGER l, there exists a POSITIVE INTEGER n and a PARTITION of $\{1, \ldots, n\}$ as a disjoint union of two sets A and B, such that for $1 \leq i \leq l$,

$$\sum_{a\in A} a^i = \sum_{b\in B} b^i.$$

The results extended to three or more sets of INTEGERS are called PROUHET'S PROBLEM.

see also PROUHET'S PROBLEM

References
Dickson, L. E. *History of the Theory of Numbers, Vol. 2: Diophantine Analysis.* New York: Chelsea, pp. 709–710, 1971.
Hahn, L. "The Tarry-Escott Problem." Problem 10284. *Amer. Math. Monthly* **102**, 843–844, 1995.

Tarry Point

The point at which the lines through the VERTICES of a TRIANGLE PERPENDICULAR to the corresponding sides of the first BROCARD TRIANGLE, are CONCURRENT. The Tarry point lies on the CIRCUMCIRCLE opposite the STEINER POINT. It has TRIANGLE CENTER FUNCTION

$$\alpha = \frac{bc}{b^4 + c^4 - a^2b^2 - a^2c^2} = \sec(A + \omega),$$

where ω is the BROCARD ANGLE. The SIMSON LINE of the Tarry point is PERPENDICULAR to the line OK, when O is the CIRCUMCENTER and K is the LEMOINE POINT.

see also BROCARD ANGLE, BROCARD TRIANGLES, CIRCUMCIRCLE, LEMOINE POINT, SIMSON LINE, STEINER POINTS

References
Gallatly, W. *The Modern Geometry of the Triangle, 2nd ed.* London: Hodgson, p. 102, 1913.
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 281–282, 1929.
Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Tarski's Theorem

Tarski's theorem states that the first-order theory of the FIELD of REAL NUMBERS is DECIDABLE. However, the best-known ALGORITHM for eliminating QUANTIFIERS is doubly exponential in the number of QUANTIFIER blocks (Heintz *et al.* 1989).

References
Heintz, J.; Roy, R.-F.; and Solerno, P. "Complexité du principe de Tarski-Seidenberg." *C. R. Acad. Sci. Paris Sér. I Math.* **309**, 825–830, 1989.
Marker, D. "Model Theory and Exponentiation." *Not. Amer. Math. Soc.* **43**, 753–759, 1996.
Tarski, A. "Sur les ensembles définissables de nombres réels." *Fund. Math.* **17**, 210–239, 1931.
Tarski, A. "A Decision Method for Elementary Algebra and Geometry." RAND Corp. monograph, 1948.

Tau Conjecture

Also known as RAMANUJAN'S HYPOTHESIS. Ramanujan proposed that

$$\tau(n) \sim \mathcal{O}(n^{11/2+\epsilon}),$$

where $\tau(n)$ is the TAU FUNCTION, defined by

$$\sum_{n=1}^{\infty} \tau(n)x^n = x(1 - 3x + 5x^3 - 7x^6 + \ldots)^8.$$

This was proven by Deligne (1974), who was subsequently awarded the FIELDS MEDAL for his proof.

see also TAU FUNCTION

References
Deligne, P. "La conjecture de Weil. I." *Inst. Hautes Études Sci. Publ. Math.* **43**, 273–307, 1974.

Deligne, P. "La conjecture de Weil. II." *Inst. Hautes Études Sci. Publ. Math.* **52**, 137–252, 1980.
Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, p. 169, 1959.

Tau-Dirichlet Series

$$\tau_{DS}(s) \equiv \sum_{n=1}^{\infty} \frac{\tau(n)}{n^s},$$

where $\tau(n)$ is the TAU FUNCTION. Ramanujan conjectured that all nontrivial zeros of $f(z)$ lie on the line $\Re[s] = 6$, where

$$f(s) \equiv \sum_{n=1}^{\infty} \tau(n) n^{-z}$$

and $\tau(n)$ is the TAU FUNCTION.

see also TAU FUNCTION

References

Spira, R. "Calculation of the Ramanujan Tau-Dirichlet Series." *Math. Comput.* **27**, 379–385, 1973.
Yoshida, H. "On Calculations of Zeros of L-Functions Related with Ramanujan's Discriminant Function on the Critical Line." *J. Ramanujan Math. Soc.* **3**, 87–95, 1988.

Tau Function

A function $\tau(n)$ related to the DIVISOR FUNCTION $\sigma_k(n)$, also sometimes called RAMANUJAN'S TAU FUNCTION. It is given by the GENERATING FUNCTION

$$\sum_{n=1}^{\infty} \tau(n) x^n = \prod_{n=1}^{\infty} (1 - x^n)^{24}, \tag{1}$$

and the first few values are 1, −24, 252, −1472, 4380, ... (Sloane's A000594). $\tau(n)$ is also given by

$$g(-x) = \sum_{n=1}^{\infty} (-1)^n \tau(n) x^n \tag{2}$$

$$g(x^2) = \sum_{n=1}^{\infty} \tau(\tfrac{1}{2}n) x^n \tag{3}$$

$$\sum_{n=1}^{\infty} \tau(n) x^n = x(1 - 3x + 5x^3 - 7x^6 + \ldots)^8. \tag{4}$$

In ORE'S CONJECTURE, the tau function appears as the number of DIVISORS of n. Ramanujan conjectured and Mordell proved that if (n, n'), then

$$\tau(nn') = \tau(n)\tau(n'). \tag{5}$$

Ramanujan conjectured and Watson proved that $\tau(n)$ is divisible by 691 for almost all n. If

$$\tau(p) \equiv 0 \pmod{p}, \tag{6}$$

then

$$\tau(pn) \equiv 0 \pmod{p}. \tag{7}$$

Values of p for which the first equation holds are $p = 2$, 3, 5, 7, 23.

Ramanujan also studied

$$f(x) \equiv \sum_{n=1}^{\infty} \tau(n) n^{-s}, \tag{8}$$

which has properties analogous to the RIEMANN ZETA FUNCTION. It satisfies

$$\frac{f(s)\Gamma(s)}{(2\pi)^s} = \frac{f(12-s)}{(2\pi)^{12-s}}, \tag{9}$$

and Ramanujan's TAU-DIRICHLET SERIES conjecture alleges that all nontrivial zeros of $f(s)$ lie on the line $\Re[s] = 6$. f can be split up into

$$f(6 + it) = z(t) e^{-i\theta(t)}, \tag{10}$$

where

$$z(t) = \Gamma(6 + it) f(6 + it) (2\pi)^{-it} \times \sqrt{\frac{\sinh(\pi t)}{\pi t(1 + t^2)(4 + t^2)(9 + t^2)(16 + t^2)(25 + t^2)}} \tag{11}$$

$$\theta(t) = -\tfrac{1}{2} i \ln \left[\frac{\Gamma(6 + it)}{\Gamma(6 - it)}\right] - t \ln(2\pi). \tag{12}$$

The SUMMATORY tau function is given by

$$T(n) = \sum_{n \leq x}{}' \tau(n). \tag{13}$$

Here, the prime indicates that when x is an INTEGER, the last term $\tau(x)$ should be replaced by $\frac{1}{2}\tau(x)$.

Ramanujan's tau theta function $Z(t)$ is a REAL function for REAL t and is analogous to the RIEMANN-SIEGEL FUNCTION Z. The number of zeros in the critical strip from $t = 0$ to T is given by

$$N(t) = \frac{\Theta(T) + \Im\{\ln[\tau_{DS}(6 + iT)]\}}{\pi}, \tag{14}$$

where Θ is the RIEMANN THETA FUNCTION and τ_{DS} is the TAU-DIRICHLET SERIES, defined by

$$\tau_{DS}(s) \equiv \sum_{n=1}^{\infty} \frac{\tau(n)}{n^s}. \tag{15}$$

Ramanujan conjectured that the nontrivial zeros of the function are all real.

Ramanujan's τ_z function is defined by

$$\tau_z(t) = \frac{\Gamma(6+it)(2\pi)^{-it}}{\tau_{DS}(6+it)\sqrt{\frac{\sinh(\pi t)}{\pi t \prod_{k=1}^{5} k^2+t^2}}}, \tag{16}$$

where $\tau_{DS}(z)$ is the Tau-Dirichlet Series.

see also Ore's Conjecture, Tau Conjecture, Tau-Dirichlet Series

References

Hardy, G. H. "Ramanujan's Function $\tau(n)$." Ch. 10 in *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, 1959.

Sloane, N. J. A. Sequence A000594/M5153 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Tauberian Theorem

A Tauberian theorem is a theorem which deduces the convergence of an Infinite Series on the basis of the properties of the function it defines and any kind of auxiliary Hypothesis which prevents the general term of the series from converging to zero too slowly.

see also Hardy-Littlewood Tauberian Theorem

Tautochrone Problem

Find the curve down which a bead placed anywhere will fall to the bottom in the same amount of time. The solution is a Cycloid, a fact first discovered and published by Huygens in *Horologium oscillatorium* (1673). Huygens also constructed the first pendulum clock with a device to ensure that the pendulum was isochronous by forcing the pendulum to swing in an arc of a Cycloid.

The parametric equations of the Cycloid are

$$x = a(\theta - \sin\theta) \tag{1}$$

$$y = a(1 - \cos\theta). \tag{2}$$

To see that the Cycloid satisfies the tautochrone property, consider the derivatives

$$x' = a(1 - \cos\theta) \tag{3}$$

$$y' = a\sin\theta, \tag{4}$$

and

$$\begin{aligned} x'^2 + y'^2 &= a^2[(1 - 2\cos\theta + \cos^2\theta) + \sin^2\theta] \\ &= 2a^2(1 - \cos\theta). \end{aligned} \tag{5}$$

Now

$$\tfrac{1}{2}mv^2 = mgy \tag{6}$$

$$v = \frac{ds}{dt} = \sqrt{2gy} \tag{7}$$

$$\begin{aligned} dt = \frac{ds}{\sqrt{2gy}} &= \frac{\sqrt{dx^2 + dy^2}}{\sqrt{2gy}} \\ &= \frac{a\sqrt{2(1-\cos\theta)}\,d\theta}{\sqrt{2ga(1-\cos\theta)}} = \sqrt{\frac{a}{g}}\,d\theta, \end{aligned} \tag{8}$$

so the time required to travel from the top of the Cycloid to the bottom is

$$T = \int_0^\pi dt = \sqrt{\frac{a}{g}}\,\pi. \tag{9}$$

However, from an intermediate point θ_0,

$$v = \frac{ds}{dt} = \sqrt{2g(y - y_0)}, \tag{10}$$

so

$$\begin{aligned} T &= \int_{\theta_0}^{\pi} \frac{\sqrt{2a^2(1-\cos\theta)}}{2ag(\cos\theta_0 - \cos\theta)}\,d\theta \\ &= \sqrt{\frac{a}{g}} \int_{\theta_0}^{\pi} \sqrt{\frac{1-\cos\theta}{\cos\theta_0 - \cos\theta}}\,d\theta \\ &= \sqrt{\frac{a}{g}} \int_{\theta_0}^{\pi} \frac{\sin(\frac{1}{2}\theta)\,d\theta}{\sqrt{\cos^2(\frac{1}{2}\theta_0) - \cos^2(\frac{1}{2}\theta)}}. \end{aligned} \tag{11}$$

Now let

$$u = \frac{\cos(\frac{1}{2}\theta)}{\cos(\frac{1}{2}\theta_0)} \tag{12}$$

$$du = -\frac{\sin(\frac{1}{2}\theta)d\theta}{2\cos(\theta_0)}, \tag{13}$$

so

$$T = -2\sqrt{\frac{a}{g}} \int_1^0 \frac{du}{\sqrt{1-u^2}} = 2\sqrt{\frac{a}{g}}\,[\sin^{-1} u]_0^1 = \pi\sqrt{\frac{a}{g}}, \tag{14}$$

and the amount of time is the same from any point!

see also Brachistochrone Problem, Cycloid

References

Muterspaugh, J.; Driver, T.; and Dick, J. E. "The Cycloid and Tautochronism." `http://ezinfo.ucs.indiana.edu/~jedick/project/intro.html`.

Muterspaugh, J.; Driver, T.; and Dick, J. E. "P221 Tautochrone Problem." `http://ezinfo.ucs.indiana.edu/~jedick/project/project.html`.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 54–60 and 384–385, 1991.

Tautology

A logical statement in which the conclusion is equivalent to the premise. If p is a tautology, it is written $\models p$.

Taxicab Number

The nth taxicab number $\mathrm{Ta}(n)$ is the smallest number representable in n ways as a sum of POSITIVE CUBES. The numbers derive their name from the HARDY-RAMANUJAN NUMBER

$$\begin{aligned}\mathrm{Ta}(2) &= 1729 \\ &= 1^3 + 12^3 \\ &= 9^3 + 10^3,\end{aligned} \tag{1}$$

which is associated with the following story told about Ramanujan by G. H. Hardy. "Once, in the taxi from London, Hardy noticed its number, 1729. He must have thought about it a little because he entered the room where Ramanujan lay in bed and, with scarcely a hello, blurted out his disappointment with it. It was, he declared, 'rather a dull number,' adding that he hoped that wasn't a bad omen. 'No, Hardy,' said Ramanujan, 'it is a very interesting number. It is the smallest number expressible as the sum of two [POSITIVE] cubes in two different ways'" (Hofstadter 1989, Kanigel 1991, Snow 1993).

However, this property was also known as early as 1657 by F. de Bessy (Berndt and Bhargava 1993, Guy 1994). Leech (1957) found

$$\begin{aligned}\mathrm{Ta}(3) &= 87539319 \\ &= 167^3 + 436^3 \\ &= 228^3 + 423^3 \\ &= 255^3 + 414^3.\end{aligned} \tag{2}$$

Rosenstiel *et al.* (1991) recently found

$$\begin{aligned}\mathrm{Ta}(4) &= 6963472309248 \\ &= 2421^3 + 19083^3 \\ &= 5436^3 + 18948^3 \\ &= 10200^3 + 18072^3 \\ &= 13322^3 + 16630^3.\end{aligned} \tag{3}$$

D. Wilson found

$$\begin{aligned}\mathrm{Ta}(5) &= 48988659276962496 \\ &= 38787^3 + 365757^3 \\ &= 107839^3 + 362753^3 \\ &= 205292^3 + 342952^3 \\ &= 221424^3 + 336588^3 \\ &= 231518^3 + 331954^3.\end{aligned} \tag{4}$$

The first few taxicab numbers are therefore 2, 1729, 87539319, 6963472309248, ... (Sloane's A011541).

Hardy and Wright (Theorem 412, 1979) show that the number of such sums can be made arbitrarily large but, updating Guy (1994) with Wilson's result, the least example is not known for six or more equal sums.

Sloane defines a slightly different type of taxicab numbers, namely numbers which are sums of two cubes in two or more ways, the first few of which are 1729, 4104, 13832, 20683, 32832, 39312, 40033, 46683, 64232, ... (Sloane's A001235).

see also DIOPHANTINE EQUATION—CUBIC, HARDY-RAMANUJAN NUMBER

References

Berndt, B. C. and Bhargava, S. "Ramanujan—For Lowbrows." *Am. Math. Monthly* **100**, 645–656, 1993.
Guy, R. K. "Sums of Like Powers. Euler's Conjecture." §D1 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 139–144, 1994.
Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, p. 68, 1959.
Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, 1979.
Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, p. 564, 1989.
Kanigel, R. *The Man Who Knew Infinity: A Life of the Genius Ramanujan.* New York: Washington Square Press, p. 312, 1991.
Leech, J. "Some Solutions of Diophantine Equations." *Proc. Cambridge Phil. Soc.* **53**, 778–780, 1957.
Plouffe, S. "Taxicab Numbers." `http://www.lacim.uqam.ca/pi/problem.html`.
Rosenstiel, E.; Dardis, J. A.; and Rosenstiel, C. R. "The Four Least Solutions in Distinct Positive Integers of the Diophantine Equation $s = x^3 + y^3 = z^3 + w^3 = u^3 + v^3 = m^3 + n^3$." *Bull. Inst. Math. Appl.* **27**, 155–157, 1991.
Silverman, J. H. "Taxicabs and Sums of Two Cubes." *Amer. Math. Monthly* **100**, 331–340, 1993.
Sloane, N. J. A. Sequences A001235 and A011541 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Snow, C. P. Foreword to *A Mathematician's Apology, reprinted with a foreword by C. P. Snow* (by G. H. Hardy). New York: Cambridge University Press, p. 37, 1993.
Wooley, T. D. "Sums of Two Cubes." *Internat. Math. Res. Not.*, 181–184, 1995.

Taylor Center

The center of the TAYLOR CIRCLE, which is the SPIEKER CENTER of $\Delta H_1H_2H_3$, where H_i are the ALTITUDES.

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, p. 277, 1929.

Taylor Circle

From the feet of each ALTITUDE of a TRIANGLE, draw lines PERPENDICULAR to the adjacent sides. Then the feet of these perpendiculars lie on a CIRCLE called the TAYLOR CIRCLE.

see also TUCKER CIRCLES

References

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, p. 277, 1929.

Taylor's Condition

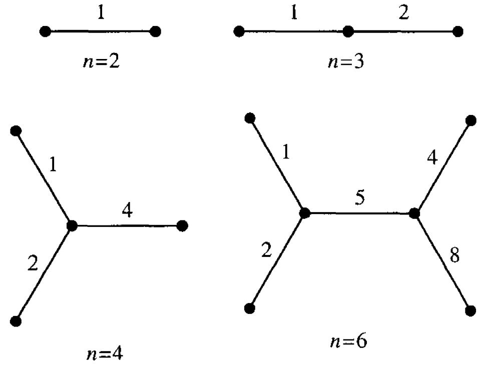

For a given POSITIVE INTEGER n, does there exist a WEIGHTED TREE with n VERTICES whose paths have weights 1, 2, ..., $\binom{n}{2}$, where $\binom{n}{2}$ is a BINOMIAL COEFFICIENT? Taylor showed that no such TREE can exist unless it is a PERFECT SQUARE or a PERFECT SQUARE plus 2. No such TREES are known except $n = 2$, 3, 4, and 6.

see also GOLOMB RULER, PERFECT DIFFERENCE SET

References

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 56–60, 1985.

Leech, J. "Another Tree Labeling Problem." *Amer. Math. Monthly* **82**, 923–925, 1975.

Taylor, H. "Odd Path Sums in an Edge-Labeled Tree." *Math. Mag.* **50**, 258–259, 1977.

Taylor Expansion

see TAYLOR SERIES

Taylor-Greene-Chirikov Map

see STANDARD MAP

Taylor Polynomial

see TAYLOR SERIES

Taylor Series

A Taylor series is a series expansion of a FUNCTION about a point. A 1-D Taylor series is an expansion of a SCALAR FUNCTION $f(x)$ about a point $x = a$. If $a = 0$, the expansion is known as a MACLAURIN SERIES.

$$\int_a^x f^{(n)}(x)\,dx = [f^{(n-1)}(x)]_a^x = f^{(n-1)}(x) - f^{(n-1)}(a) \tag{1}$$

$$\int_a^x \left[\int_a^x f^{(n)}(x)\,dx\right] dx = \int_a^x [f^{(n-1)}(x) - f^{(n-1)}(a)]\,dx$$
$$= f^{(n-2)}(x) - f^{(n-2)}(a) - (x-a)f^{(n-1)}(a). \tag{2}$$

Continuing,

$$\iiint_a^x f^{(n)}(x)\,(dx)^3 = f^{(n-3)}(a) - (x-a)f^{(n-2)}(a)$$
$$-\tfrac{1}{2!}(x-a)^2 f^{(n-1)}(a) \tag{3}$$

$$\underbrace{\int\cdots\int_a^x}_{n} f^{(n)}(x)\,(dx)^n = f(x) - f(a) - (x-a)f'(a)$$
$$-\tfrac{1}{2!}(x-a)^2 f''(a) - \ldots - \tfrac{1}{(n-1)!}(x-a)^{n-1} f^{(n-1)}(a). \tag{4}$$

Therefore, we obtain the 1-D Taylor series

$$f(x) = f(a) + (x-a)f'(a) + \tfrac{1}{2!}(x-a)^2 f''(a) + \ldots$$
$$+ \tfrac{1}{(n-1)!}(x-a)^{n-1} f^{(n-1)}(a) + R_n, \tag{5}$$

where R_n is a remainder term defined by

$$R_n = \underbrace{\int\cdots\int_a^x}_{n} f^{(n)}(x)\,(dx)^n. \tag{6}$$

Using the MEAN-VALUE THEOREM for a function g, it must be true that

$$\int_a^x g(x)\,dx = (x-a)g(x^*) \tag{7}$$

for some $x^* \in [a, x]$. Therefore, integrating n times gives the result

$$R_n = \frac{(x-a)^n}{n!} f^{(n)}(x). \tag{8}$$

The maximum error is then the maximum value of (8) for all possible $x^* \in [a, x]$.

An alternative form of the 1-D Taylor series may be obtained by letting

$$x - a \equiv \Delta x \tag{9}$$

so that

$$x = a + \Delta x \equiv x_0 + \Delta x. \tag{10}$$

Substitute this result into (5) to give

$$f(x_0 + \Delta x) = f(x_0) + \Delta x f'(x_0) + \tfrac{1}{2!}(\Delta x)^2 f''(x_0) + \ldots. \tag{11}$$

A Taylor series of a FUNCTION in two variables $f(x, y)$ is given by

$$f(x+\Delta x, y+\Delta y) = f(x,y) + [f_x(x,y)\Delta x + f_y(x,y)\Delta y]$$
$$+\tfrac{1}{2!}[(\Delta x)^2 f_{xx}(x,y) + 2\Delta x\Delta y f_{xy}(x,y) + (\Delta y)^2 f_{yy}(x,y)]$$
$$+\tfrac{1}{3!}[(\Delta x)^3 f_{xxx}(x,y) + 3(\Delta x)^2\Delta y f_{xxy}(x,y)$$
$$+3\Delta x(\Delta y)^2 f_{xyy}(x,y) + (\Delta y)^3 f_{yyy}(x,y)] + \ldots. \tag{12}$$

This can be further generalized for a FUNCTION in n variables,

$$f(x_1, \ldots, x_n)$$
$$= \sum_{j=0}^{\infty} \left\{ \frac{1}{j!} \left[\sum_{k=1}^{n} (x'_k - a_k) \frac{\partial}{\partial x'_k} \right]^j f(x'_1, \ldots, x'_n) \right\}_{x'_1=a_1,\ldots,x'_n=a_n}. \tag{13}$$

Rewriting,

$$f(x_1+a_1,\dots,x_n+a_n) = \sum_{j=0}^{\infty}\left\{\frac{1}{j!}\left(\sum_{k=1}^{n} a_k \frac{\partial}{\partial x'_k}\right)^j f(x'_1,\dots,x'_n)\right\}_{x'_1=x_1,\dots,x'_n=x_n}. \tag{14}$$

Taking $n=2$ in (13) gives

$$\begin{aligned} f(x_1,x_2) &= \sum_{j=0}^{\infty}\left\{\frac{1}{j!}\left[(x'_1-a_1)\frac{\partial}{\partial x'_1} + (x'_2-a_2)\frac{\partial}{\partial x'_2}\right]^j f(x'_1,x'_2)\right\}_{x'_1=x_1,x'_2=x_2} \\ &= f(a_1,a_2) + \left[(x_1-a_1)\frac{\partial f}{\partial x_1} + (x_2-a_2)\frac{\partial f}{\partial x_2}\right] \\ &\quad + \frac{1}{2!}\left[(x_1-a_1)^2\frac{\partial^2 f}{\partial {x_1}^2} + 2(x_1-a_1)(x_2-a_2)\frac{\partial^2 f}{\partial x_1 \partial x_2} + (x_2-a_2)^2\frac{\partial^2 f}{\partial {x_2}^2}\right] + \dots \end{aligned} \tag{15}$$

Taking $n=3$ in (14) gives

$$\begin{aligned} f(x_1+a_1,x_2+a_2,x_3+a_3) &= \sum_{j=0}^{\infty}\left\{\frac{1}{j!}\left(a_1\frac{\partial}{\partial x'_1} + a_2\frac{\partial}{\partial x'_2} + a_3\frac{\partial}{\partial x'_3}\right)^j \right. \\ &\quad \left. \times f(x'_1,x'_2,x'_3)\right\}_{x'_1=x_1,x'_2=x_2,x'_3=x_3}, \end{aligned} \tag{16}$$

or, in VECTOR form

$$f(\mathbf{r}+\mathbf{a}) = \sum_{j=0}^{\infty}\left[\frac{1}{j!}(\mathbf{a}\cdot\nabla_{\mathbf{r}'})^j f(\mathbf{r}')\right]_{\mathbf{r}'=\mathbf{r}}. \tag{17}$$

The zeroth- and first-order terms are

$$f(\mathbf{r}) \tag{18}$$

and

$$(\mathbf{a}\cdot\nabla_{\mathbf{r}'})f(\mathbf{r}')|_{\mathbf{r}'=\mathbf{r}}, \tag{19}$$

respectively. The second-order term is

$$\begin{aligned} \tfrac{1}{2}(\mathbf{a}\cdot\nabla_{\mathbf{r}'})(\mathbf{a}\cdot\nabla_{\mathbf{r}'})f(\mathbf{r}')|_{\mathbf{r}'=\mathbf{r}} &= \tfrac{1}{2}\mathbf{a}\cdot\nabla_{\mathbf{r}'}[\mathbf{a}\cdot(\nabla f(\mathbf{r}'))]_{\mathbf{r}'=\mathbf{r}} \\ &= \tfrac{1}{2}\mathbf{a}\cdot[\mathbf{a}\cdot\nabla_{\mathbf{r}'}(\nabla_{\mathbf{r}'} f(\mathbf{r}'))]|_{\mathbf{r}'=\mathbf{r}}, \end{aligned} \tag{20}$$

so the first few terms of the expansion are

$$\begin{aligned} f(\mathbf{r}+\mathbf{a}) &= f(\mathbf{r}) + (\mathbf{a}\cdot\nabla_{\mathbf{r}'})f(\mathbf{r}')|_{\mathbf{r}'=\mathbf{r}} \\ &\quad + \tfrac{1}{2}\mathbf{a}\cdot[\mathbf{a}\cdot\nabla_{\mathbf{r}'}(\nabla_{\mathbf{r}'} f(\mathbf{r}'))]|_{\mathbf{r}'=\mathbf{r}}. \end{aligned} \tag{21}$$

Taylor series can also be defined for functions of a COMPLEX variable. By the CAUCHY INTEGRAL FORMULA,

$$\begin{aligned} f(z) &= \frac{1}{2\pi i}\int_C \frac{f(z')\,dz}{z'-z} = \frac{1}{2\pi i}\int_C \frac{f(z')\,dz'}{(z'-z_0)-(z-z_0)} \\ &= \frac{1}{2\pi i}\int_C \frac{f(z')\,dz'}{(z'-z_0)\left(1-\frac{z-z_0}{z'-z_0}\right)}. \end{aligned} \tag{22}$$

In the interior of C,

$$\frac{|z-z_0|}{|z'-z_0|} < 1 \tag{23}$$

so, using

$$\frac{1}{1-t} = \sum_{n=0}^{\infty} t^n, \tag{24}$$

it follows that

$$\begin{aligned} f(z) &= \frac{1}{2\pi i}\int_C \sum_{n=0}^{\infty}\frac{(z-z_0)^n f(z')\,dz'}{(z'-z_0)^{n+1}} \\ &= \frac{1}{2\pi i}\sum_{n=0}^{\infty}(z-z_0)^n \int_C \frac{f(z')\,dz}{(z'-z_0)^{n+1}}. \end{aligned} \tag{25}$$

Using the the CAUCHY INTEGRAL FORMULA for derivatives,

$$f(z) = \sum_{n=0}^{\infty}(z-z_0)^n \frac{f^{(n)}(z_0)}{n!}. \tag{26}$$

see also CAUCHY REMAINDER FORM, LAGRANGE EXPANSION, LAURENT SERIES, LEGENDRE SERIES, MACLAURIN SERIES, NEWTON'S FORWARD DIFFERENCE FORMULA

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 880, 1972.

Arfken, G. "Taylor's Expansion." §5.6 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 303–313, 1985.

Morse, P. M. and Feshbach, H. "Derivatives of Analytic Functions, Taylor and Laurent Series." §4.3 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 374–398, 1953.

Tchebycheff

see CHEBYSHEV APPROXIMATION FORMULA, CHEBYSHEV CONSTANTS, CHEBYSHEV DEVIATION, CHEBYSHEV DIFFERENTIAL EQUATION, CHEBYSHEV FUNCTION, CHEBYSHEV-GAUSS QUADRATURE, CHEBYSHEV INEQUALITY, CHEBYSHEV INEQUALITY, CHEBYSHEV INTEGRAL, CHEBYSHEV PHENOMENON, CHEBYSHEV POLYNOMIAL OF THE FIRST KIND, CHEBYSHEV POLYNOMIAL OF THE SECOND KIND, CHEBYSHEV QUADRATURE, CHEBYSHEV-RADAU QUADRATURE, CHEBYSHEV-SYLVESTER CONSTANT

Teardrop Curve

A plane curve given by the parametric equations

$$x = \cos t$$
$$y = \sin t \sin^m(\tfrac{1}{2}t).$$

see also PEAR-SHAPED CURVE

References

von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 174, 1993.

Technique

A specific method of performing an operation. The terms ALGORITHM, METHOD, and PROCEDURE are also used interchangeably.

see also ALGORITHM, METHOD, PROCEDURE

Teichmüller Space

TEICHMÜLLER'S THEOREM asserts the EXISTENCE and UNIQUENESS of the extremal quasiconformal map between two compact RIEMANN SURFACES of the same GENUS modulo an EQUIVALENCE RELATION. The equivalence classes form the Teichmüller space T_p of compact RIEMANN SURFACES of GENUS p.

see also RIEMANN'S MODULI PROBLEM

Teichmüller's Theorem

Asserts the EXISTENCE and UNIQUENESS of the extremal quasiconformal map between two compact RIEMANN SURFACES of the same GENUS modulo an EQUIVALENCE RELATION.

see also TEICHMÜLLER SPACE

Telescoping Sum

A sum in which subsequent terms cancel each other, leaving only initial and final terms. For example,

$$\begin{aligned} S &= \sum_{i=1}^{n-1}\left(\frac{1}{a_i} - \frac{1}{a_{i+1}}\right) \\ &= \left(\frac{1}{a_1} - \frac{1}{a_2}\right) + \left(\frac{1}{a_2} - \frac{1}{a_3}\right) + \ldots \\ &\quad + \left(\frac{1}{a_{n-2}} - \frac{1}{a_{n-1}}\right) + \left(\frac{1}{a_{n-1}} - \frac{1}{a_n}\right) \\ &= \frac{1}{a_1} - \frac{1}{a_n} \end{aligned}$$

is a telescoping sum.

see also ZEILBERGER'S ALGORITHM

Temperature

The "temperature" of a curve Γ is defined as

$$T \equiv \frac{1}{\ln\left(\frac{2l}{2l-h}\right)},$$

where l is the length of Γ and h is the length of the PERIMETER of the CONVEX HULL. The temperature of a curve is 0 only if the curve is a straight line, and increases as the curve becomes more "wiggly."

see also CURLICUE FRACTAL

References

Pickover, C. A. *Keys to Infinity.* New York: W. H. Freeman, pp. 164–165, 1995.

Templar Magic Square

S	A	T	O	R
A	R	E	P	O
T	E	N	E	T
O	P	E	R	A
R	O	T	A	S

A MAGIC SQUARE-type arrangement of the words in the Latin sentence "Sator Arepo tenet opera rotas" ("the farmer Arepo keeps the world rolling"). This square has been found in excavations of ancient Pompeii.

see also MAGIC SQUARE

References

Bouisson, S. M. *La Magie: Ses Grands Rites, Son Histoire.* Paris, pp. 147–148, 1958.

Grosser, F. "Ein neuer Vorschlag zur Deutung der Sator-Formel." *Archiv. f. Relig.* **29**, 165–169, 1926.

Heietala, H. "The Templar Magic Square." `http://www.trantex.fi/staff/heikkih/knights/pubsator.htm`.

Hocke, G. R. *Manierismus in der Literatur: Sprach-Alchimie und esoterische Kombinationskunst.* Hamburg, Germany: Rowohlt, p. 24, 1967.

Tennis Ball Theorem

A closed simple smooth spherical curve dividing the SPHERE into two parts of equal areas has at least four inflection points.

see also BALL, BASEBALL COVER

References

Arnold, V. I. *Topological Invariants of Plane Curves and Caustics.* Providence, RI: Amer. Math. Soc., 1994.

Martinez-Maure, Y. "A Note on the Tennis Ball Theorem." *Amer. Math. Monthly* **103**, 338–340, 1996.

Tensor

An nth-RANK tensor of order m is a mathematical object in m-dimensional space which has n indices and m^n components and obeys certain transformation rules. Each index of a tensor ranges over the number of dimensions of SPACE. If the components of any tensor of any RANK vanish in one particular coordinate system, they vanish in all coordinate systems.

Zeroth-RANK tensors are called SCALARS, and first-RANK tensors are called VECTORS. In tensor notation, a vector $\mathbf{v}$ would be written v_i, where $i = 1, \ldots, m$. Tensor notation can provide a very concise way of writing vector and more general identities. For example, in tensor notation, the DOT PRODUCT $\mathbf{u} \cdot \mathbf{v}$ is simply written

$$\mathbf{u} \cdot \mathbf{v} = u_i v_i, \tag{1}$$

where repeated indices are summed over (EINSTEIN SUMMATION) so that $u_i v_i$ stands for $u_1 v_1 + \ldots + u_m v_m$. Similarly, the CROSS PRODUCT can be concisely written as

$$\mathbf{u} \times \mathbf{v} = \epsilon_{ijk} u^j v^k, \tag{2}$$

where ϵ_{ijk} is the LEVI-CIVITA TENSOR.

Second-RANK tensors resemble square MATRICES. CONTRAVARIANT second-RANK tensors are objects which transform as

$$A'^{ij} = \frac{\partial x'_i}{\partial x_k} \frac{\partial x'_j}{\partial x_l} A^{kl}. \tag{3}$$

COVARIANT second-RANK tensors are objects which transform as

$$C'_{ij} = \frac{\partial x_k}{\partial x'_i} \frac{\partial x_l}{\partial x'_i} C_{kl}. \tag{4}$$

MIXED second-RANK tensors are objects which transform as

$$B'^{j}_{i} = \frac{\partial x'_i}{\partial x_k} \frac{\partial x_l}{\partial x'_j} B^k_l. \tag{5}$$

If two tensors A and B have the same RANK and the same COVARIANT and CONTRAVARIANT indices, then

$$A^{ij} + B^{ij} = C^{ij} \tag{6}$$

$$A_{ij} + B_{ij} = C_{ij} \tag{7}$$

$$A^i_j + B^i_j = C^i_j. \tag{8}$$

A transformation of the variables of a tensor changes the tensor into another whose components are linear HOMOGENEOUS FUNCTIONS of the components of the original tensor.

see also ANTISYMMETRIC TENSOR, CURL, DIVERGENCE, GRADIENT, IRREDUCIBLE TENSOR, ISOTROPIC TENSOR, JACOBI TENSOR, RICCI TENSOR, RIEMANN TENSOR, SCALAR, SYMMETRIC TENSOR, TORSION TENSOR, VECTOR, WEYL TENSOR

References

Abraham, R.; Marsden, J. E.; and Ratiu, T. S. *Manifolds, Tensor Analysis, and Applications.* New York: Springer-Verlag, 1991.

Akivis, M. A. and Goldberg, V. V. *An Introduction to Linear Algebra and Tensors.* New York: Dover, 1972.

Arfken, G. "Tensor Analysis." Ch. 3 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 118–167, 1985.

Aris, R. *Vectors, Tensors, and the Basic Equations of Fluid Mechanics.* New York: Dover, 1989.

Bishop, R. and Goldberg, S. *Tensor Analysis on Manifolds.* New York: Dover, 1980.

Jeffreys, H. *Cartesian Tensors.* Cambridge, England: Cambridge University Press, 1931.

Joshi, A. W. *Matrices and Tensors in Physics, 3rd ed.* New York: Wiley, 1995.

Lass, H. *Vector and Tensor Analysis.* New York: McGraw-Hill, 1950.

Lawden, D. F. *An Introduction to Tensor Calculus, Relativity, and Cosmology, 3rd ed.* Chichester, England: Wiley, 1982.

McConnell, A. J. *Applications of Tensor Analysis.* New York: Dover, 1947.

Morse, P. M. and Feshbach, H. "Vector and Tensor Formalism." §1.5 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 44–54, 1953.

Simmonds, J. G. *A Brief on Tensor Analysis, 2nd ed.* New York: Springer-Verlag, 1994.

Sokolnikoff, I. S. *Tensor Analysis—Theory and Applications, 2nd ed.* New York: Wiley, 1964.

Synge, J. L. and Schild, A. *Tensor Calculus.* New York: Dover, 1978.

Wrede, R. C. *Introduction to Vector and Tensor Analysis.* New York: Wiley, 1963.

Tensor Calculus

The set of rules for manipulating and calculating with TENSORS.

Tensor Density

A quantity which transforms like a TENSOR except for a scalar factor of a JACOBIAN.

Tensor Dual

see DUAL TENSOR

Tensor Product

see DIRECT PRODUCT (TENSOR)

Tensor Space

Let E be a linear space over a FIELD K. Then the DIRECT PRODUCT $\bigotimes_{\lambda=1}^{k} E$ is called a tensor space of degree k.

References

Yokonuma, T. *Tensor Spaces and Exterior Algebra.* Providence, RI: Amer. Math. Soc., 1992.

Tensor Spherical Harmonic

see DOUBLE CONTRACTION RELATION

Tensor Transpose

see TRANSPOSE

Tent Map

A piecewise linear, 1-D MAP on the interval $[0,1]$ exhibiting CHAOTIC dynamics and given by

$$x_{n+1} = \mu(1 - 2|x_n - \tfrac{1}{2}|).$$

The case $\mu = 1$ is equivalent to the LOGISTIC EQUATION WITH $r = 4$, so the NATURAL INVARIANT in this case is

$$\rho(x) = \frac{1}{\pi\sqrt{x(1-x)}}.$$

see also $2x$ MOD 1 MAP, LOGISTIC EQUATION, LOGISTIC EQUATION WITH $r = 4$

Terminal

see SINK (DIRECTED GRAPH)

Ternary

The BASE 3 method of counting in which only the digits 0, 1, and 2 are used. These digits have the following multiplication table.

×	0	1	2
0	0	0	0
1	0	1	2
2	0	2	11

Erdős and Graham (1980) conjectured that no POWER of 2, 2^n, is a SUM of distinct powers of 3 for $n > 8$. This is equivalent to the requirement that the ternary expansion of 2^n always contains a 2. This has been verified by Vardi (1991) up to $n = 2 \cdot 3^{20}$. N. J. A. Sloane has conjectured that any POWER of 2 has a 0 in its ternary expansion (Vardi 1991, p. 28).

see also BASE (NUMBER), BINARY, DECIMAL, HEXADECIMAL, OCTAL, QUATERNARY

References

Erdős, P. and Graham, R. L. *Old and New Problems and Results in Combinatorial Number Theory.* Geneva, Switzerland: L'Enseignement Mathématique Université de Genève, Vol. 28, 1980.

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, pp. 10–11, 1991.

Vardi, I. "The Digits of 2^n in Base Three." *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 20–25, 1991.

Weisstein, E. W. "Bases." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Bases.m`.

Tessellation

A regular TILING of POLYGONS (in 2-D), POLYHEDRA (3-D), or POLYTOPES (n-D) is called a tessellation. Tessellations can be specified using a SCHLÄFLI SYMBOL.

Consider a 2-D tessellation with q regular p-gons at each VERTEX. In the PLANE,

$$\left(1 - \frac{2}{p}\right)\pi = \frac{2\pi}{q} \tag{1}$$

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{2}, \tag{2}$$

so

$$(p-2)(q-2) = 4 \tag{3}$$

(Ball and Coxeter 1987), and the only factorizations are

$$\begin{aligned} 4 &= 4 \cdot 1 = (6-2)(3-2) \Rightarrow \{6,3\} && (4)\\ &= 2 \cdot 2 = (4-2)(4-2) \Rightarrow \{4,4\} && (5)\\ &= 1 \cdot 4 = (3-2)(6-2) \Rightarrow \{3,6\}. && (6) \end{aligned}$$

Therefore, there are only three regular tessellations (composed of the HEXAGON, SQUARE, and TRIANGLE), illustrated as follows.

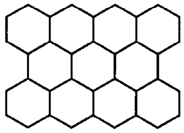

$\{6,3\}$

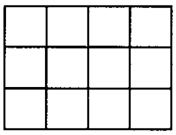

$\{4,4\}$

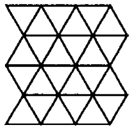

$\{3,6\}$

There do not exist any regular STAR POLYGON tessellations in the PLANE. Regular tessellations of the SPHERE by SPHERICAL TRIANGLES are called TRIANGULAR SYMMETRY GROUPS.

Regular tilings of the plane by *two or more* convex regular POLYGONS such that the same POLYGONS in the same order surround each VERTEX are called semiregular tilings. In the plane, there are eight such tessellations, illustrated below.

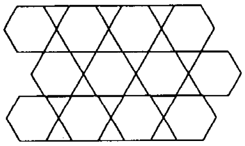

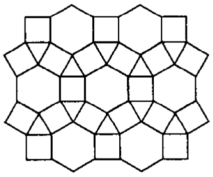

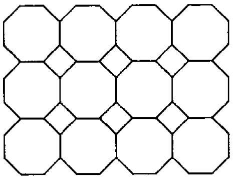

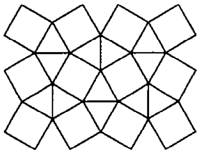

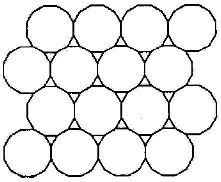

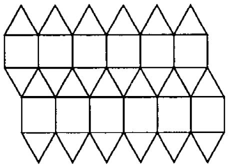

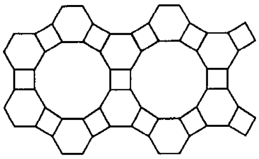

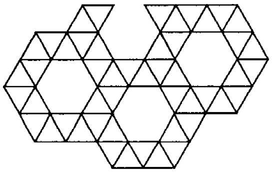

In 3-D, a POLYHEDRON which is capable of tessellating space is called a SPACE-FILLING POLYHEDRON. Examples include the CUBE, RHOMBIC DODECAHEDRON, and TRUNCATED OCTAHEDRON. There is also a 16-sided space-filler and a convex POLYHEDRON known as the SCHMITT-CONWAY BIPRISM which fills space only aperiodically.

A tessellation of n-D polytopes is called a HONEYCOMB.

see also ARCHIMEDEAN SOLID, CELL, HONEYCOMB, SCHLÄFLI SYMBOL, SEMIREGULAR POLYHEDRON, SPACE-FILLING POLYHEDRON, TILING, TRIANGULAR SYMMETRY GROUP

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 105–107, 1987.

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 60–63, 1989.

Gardner, M. *Martin Gardner's New Mathematical Diversions from Scientific American.* New York: Simon and Schuster, pp. 201–203, 1966.

Gardner, M. "Tilings with Convex Polygons." Ch. 13 in *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, pp. 162–176, 1988.

Kraitchik, M. "Mosaics." §8.2 in *Mathematical Recreations.* New York: W. W. Norton, pp. 199–207, 1942.

Lines, L. *Solid Geometry.* New York: Dover, pp. 199 and 204–207 1965.

Pappas, T. "Tessellations." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 120–122, 1989.

Peterson, I. *The Mathematical Tourist: Snapshots of Modern Mathematics.* New York: W. H. Freeman, p. 75, 1988.

Rawles, B. *Sacred Geometry Design Sourcebook: Universal Dimensional Patterns.* Nevada City, CA: Elysian Pub., 1997.

Walsh, T. R. S. "Characterizing the Vertex Neighbourhoods of Semi-Regular Polyhedra." *Geometriae Dedicata* **1**, 117–123, 1972.

Tesseract

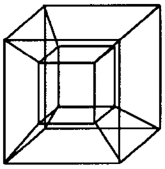

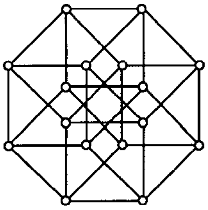

The HYPERCUBE in $\mathbb{R}^4$ is called a tesseract. It has the SCHLÄFLI SYMBOL $\{4,3,3\}$, and VERTICES $(\pm 1, \pm 1, \pm 1, \pm 1)$. The above figures show two visualizations of the TESSERACT. The figure on the left is a projection of the TESSERACT in 3-space (Gardner 1977), and the figure on the right is the GRAPH of the TESSERACT symmetrically projected into the PLANE (Coxeter 1973). A TESSERACT has 16 VERTICES, 32 EDGES, 4 SQUARES, and 8 CUBES.

see also HYPERCUBE, POLYTOPE

References

Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, p. 123, 1973.

Gardner, M. "Hypercubes." Ch. 4 in *Mathematical Carnival: A New Round-Up of Tantalizers and Puzzles from Scientific American.* New York: Vintage Books, 1977.

Geometry Center. "The Tesseract (or Hypercube)." `http://www.geom.umn.edu/docs/outreach/4-cube/`.

Tesseral Harmonic

A SPHERICAL HARMONIC which is expressible as products of factors linear in x^2, y^2, and z^2 multiplied by one of 1, x, y, z, yz, zx, xy, and xyz.

see also ZONAL HARMONIC

Tethered Bull Problem

Let a bull be tethered to a silo whose horizontal CROSS-SECTION is a CIRCLE of RADIUS R by a leash of length L. Then the AREA which the bull can graze if $L \leq R\pi$ is

$$A = \frac{\pi L^2}{2} + \frac{L^3}{3R}.$$

References

Hoffman, M. E. "The Bull and the Silo: An Application of Curvature." *Amer. Math. Monthly* **105**, 55–58, 1998.

Tetrabolo

A 4-POLYABOLO.

Tetrachoric Function

The function defined by

$$T_n \equiv \frac{(-1)^{n-1}}{\sqrt{n!}} Z^{(n-1)}(x),$$

where

$$Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

see also NORMAL DISTRIBUTION

References

Kenney, J. F. and Keeping, E. S. "Tetrachoric Correlation." §8.5 in *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, pp. 205–207, 1951.

Tetracontagon

A 40-sided POLYGON.

Tetracuspid

see HYPOCYCLOID—4-CUSPED

Tetrad

A SET of four, also called a QUARTET.

see also HEXAD, MONAD, PAIR, QUARTET, QUINTET, TRIAD, TRIPLE, TWINS

Tetradecagon

A 14-sided POLYGON, sometimes called a TETRAKAIDECAGON.

Tetradecahedron

A 14-sided POLYHEDRON, sometimes called a TETRAKAIDECAHEDRON.

see also CUBOCTAHEDRON, TRUNCATED OCTAHEDRON

References

Ghyka, M. *The Geometry of Art and Life.* New York: Dover, p. 54, 1977.

Tetradic

Tetradics transform DYADICS in much the same way that DYADICS transform VECTORS. They are represented using Hebrew characters and have 81 components (Morse and Feshbach 1953, pp. 72–73). The use of tetradics is archaic, since TENSORS perform the same function but are notationally simpler.

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Vol. 1.* New York: McGraw-Hill, 1953.

Tetradyakis Hexahedron

The DUAL POLYHEDRON of the CUBITRUNCATED CUBOCTAHEDRON.

Tetraflexagon

A FLEXAGON made with SQUARE faces. Gardner (1961) shows how to construct a tri-tetraflexagon,

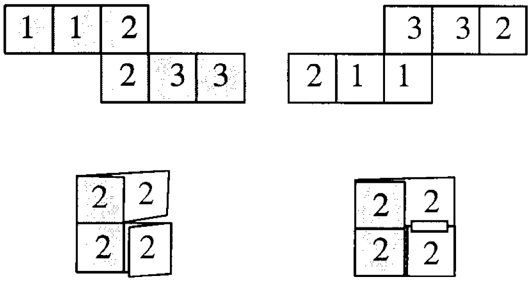

tetra-tetraflexagon,

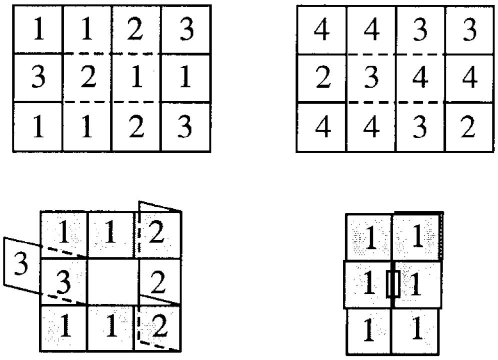

and hexa-tetraflexagon.

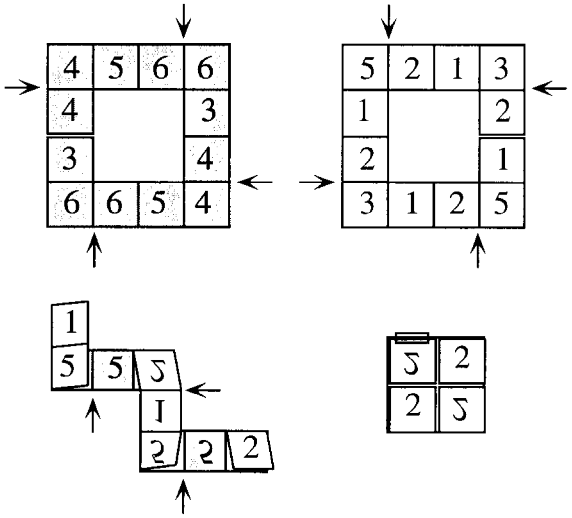

see also FLEXAGON, FLEXATUBE, HEXAFLEXAGON

References

Chapman, P. B. "Square Flexagons." *Math. Gaz.* **45**, 192–194, 1961.

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 207, 1989.

Gardner, M. Ch. 1 in *The Scientific American Book of Mathematical Puzzles & Diversions.* New York: Simon and Schuster, 1959.

Gardner, M. Ch. 2 in *The Second Scientific American Book of Mathematical Puzzles & Diversions: A New Selection.* New York: Simon and Schuster, 1961.

Pappas, T. "Making a Tri-Tetra Flexagon." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 107, 1989.

Tetragon

see QUADRILATERAL

Tetrahedral Coordinates

Coordinates useful for plotting projective 3-D curves of the form $f(x_0, x_1, x_2, x_3) = 0$ which are defined by

$$\begin{aligned} x_0 &= 1 - z - \sqrt{2}\,x \\ x_1 &= 1 - z + \sqrt{2}\,x \\ x_2 &= 1 + z + \sqrt{2}\,y \\ x_3 &= 1 + z - \sqrt{2}\,y. \end{aligned}$$

see also CAYLEY CUBIC, KUMMER SURFACE

Tetrahedral Graph

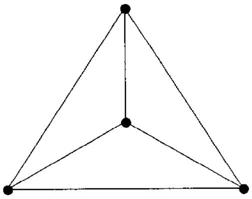

A POLYHEDRAL GRAPH which is also the COMPLETE GRAPH K_4.

see also CUBICAL GRAPH, DODECAHEDRAL GRAPH, ICOSAHEDRAL GRAPH, OCTAHEDRAL GRAPH, TETRAHEDRON

Tetrahedral Group

The POINT GROUP of symmetries of the TETRAHEDRON, denoted T_d. The tetrahedral group has symmetry operations E, $8C_3$, $3C_2$, $6S_4$, and $6\sigma_d$ (Cotton 1990).

see also ICOSAHEDRAL GROUP, OCTAHEDRAL GROUP, POINT GROUPS, TETRAHEDRON

References

Cotton, F. A. *Chemical Applications of Group Theory, 3rd ed.* New York: Wiley, p. 47, 1990.

Lomont, J. S. "Icosahedral Group." §3.10.C in *Applications of Finite Groups.* New York: Dover, p. 81, 1987.

Tetrahedral Number

A FIGURATE NUMBER Te_n of the form

$$Te_n = \sum_{i=1}^{n} T_n = \tfrac{1}{6}n(n+1)(n+2) = \binom{n+2}{3}, \qquad (1)$$

where T_n is the nth TRIANGULAR NUMBER and $\binom{n}{m}$ is a BINOMIAL COEFFICIENT. These numbers correspond to placing discrete points in the configuration of a TETRAHEDRON (triangular base pyramid). Tetrahedral numbers are PYRAMIDAL NUMBERS with $r = 3$, and are the sum of consecutive TRIANGULAR NUMBERS. The first few are 1, 4, 10, 20, 35, 56, 84, 120, ... (Sloane's A000292). The GENERATING FUNCTION of the tetrahedral numbers is

$$\frac{x}{(x-1)^4} = x + 4x^2 + 10x^3 + 20x^4 + \ldots. \qquad (2)$$

Tetrahedral numbers are EVEN, except for every fourth tetrahedral number, which is ODD (Conway and Guy 1996).

The only numbers which are simultaneously SQUARE and TETRAHEDRAL are $Te_1 = 1$, $Te_2 = 4$, and $Te_{48} = 19600$ (giving $S_1 = 1$, $S_2 = 4$, and $S_{140} = 19600$), as proved by Meyl (1878; cited in Dickson 1952, p. 25). Numbers which are simultaneously TRIANGULAR and tetrahedral satisfy the BINOMIAL COEFFICIENT equation

$$\binom{n}{2} = \binom{m}{3}, \qquad (3)$$

the only solutions of which are $(m, n) = (10, 16)$, (22, 56), and (36, 120) (Guy 1994, p. 147). Beukers (1988) has studied the problem of finding numbers which are simultaneously tetrahedral and PYRAMIDAL via INTEGER points on an ELLIPTIC CURVE, and finds that the only solution is the trivial $Te_1 = P_1 = 1$.

see also PYRAMIDAL NUMBER, TRUNCATED TETRAHEDRAL NUMBER

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 59, 1987.

Beukers, F. "On Oranges and Integral Points on Certain Plane Cubic Curves." *Nieuw Arch. Wisk.* **6**, 203–210, 1988.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 44–46, 1996.

Dickson, L. E. *History of the Theory of Numbers, Vol. 2: Diophantine Analysis.* New York: Chelsea, 1952.

Guy, R. K. "Figurate Numbers." §D3 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 147–150, 1994.

Meyl, A.-J.-J. "Solution de Question 1194." *Nouv. Ann. Math.* **17**, 464–467, 1878.

Sloane, N. J. A. Sequence A000292/M3382 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Tetrahedral Surface

A SURFACE given by the parametric equations

$$\begin{aligned} x &= A(u-a)^m (v-a)^n \\ y &= B(u-b)^m (v-b)^n \\ z &= C(u-c)^m (v-c)^n. \end{aligned}$$

References

Eisenhart, L. P. *A Treatise on the Differential Geometry of Curves and Surfaces.* New York: Dover, p. 267, 1960.

Tetrahedroid

A special case of a quartic KUMMER SURFACE.

References

Fischer, G. (Ed.). *Mathematical Models from the Collections of Universities and Museums.* Braunschweig, Germany: Vieweg, pp. 17–19, 1986.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 183, 1994.

Tetrahedron

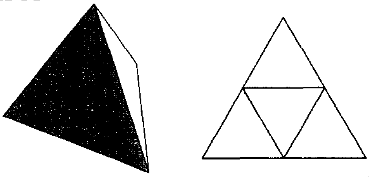

The regular tetrahedron, often simply called "the" tetrahedron, is the PLATONIC SOLID P_1 with four VERTICES, six EDGES, and four equivalent EQUILATERAL TRIANGULAR faces ($4\{3\}$). It is also UNIFORM POLYHEDRON U_1. It is described by the SCHLÄFLI SYMBOL $\{3,3\}$ and the WYTHOFF SYMBOL is $3\,|\,2\,3$. It is the prototype of the TETRAHEDRAL GROUP T_d,

The tetrahedron is its own DUAL POLYHEDRON. It is the only simple POLYHEDRON with no DIAGONALS, and cannot be STELLATED. The VERTICES of a tetrahedron are given by $(0,0,\sqrt{3})$, $(0,\frac{2}{3}\sqrt{6},-\frac{1}{3}\sqrt{3})$, $(-\sqrt{2},-\frac{1}{3}\sqrt{6},-\frac{1}{3}\sqrt{3})$, and $(\sqrt{2},-\frac{1}{3}\sqrt{6},-\frac{1}{3}\sqrt{3})$, or by (0, 0, 0), (0, 1, 1), (1, 0, 1), (1, 1, 0). In the latter case, the face planes are

$$x+y+z=2 \tag{1}$$

$$x-y-z=0 \tag{2}$$

$$-x+y-z=0 \tag{3}$$

$$x+y-z=0. \tag{4}$$

Let a tetrahedron be length a on a side. The VERTICES are located at $(x, 0, 0)$, $(-d, \pm a/2, 0)$, and $(0, 0, h)$. From the figure,

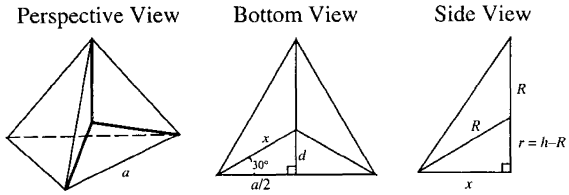

$$x=\frac{\frac{a}{2}}{\cos(\frac{\pi}{6})}=\frac{1}{2}\frac{2}{\sqrt{3}}a=\tfrac{1}{3}\sqrt{3}\,a. \tag{5}$$

d is then

$$d=\sqrt{x^2-(\tfrac{1}{2}a)^2}=a\sqrt{\tfrac{1}{3}-\tfrac{1}{4}}=a\sqrt{\frac{4-3}{12}}=\frac{a}{\sqrt{12}}$$
$$=\tfrac{1}{6}\sqrt{3}\,a. \tag{6}$$

This gives the AREA of the base as

$$\begin{aligned}A&=\tfrac{1}{2}a(R+x)=\tfrac{1}{2}a\left(\frac{\sqrt{3}}{6}a+\frac{1}{\sqrt{3}}a\right)\\&=\tfrac{1}{2}a^2\left(\frac{\sqrt{3}}{6}+\frac{2\sqrt{3}}{6}\right)\\&=\tfrac{1}{2}a^2\frac{3\sqrt{3}}{6}=\tfrac{1}{4}\sqrt{3}\,a^2.\end{aligned} \tag{7}$$

The height is

$$h=\sqrt{a^2-x^2}=a\sqrt{1-\tfrac{1}{3}}=\tfrac{1}{3}\sqrt{6}\,a. \tag{8}$$

The CIRCUMRADIUS R is found from

$$x^2+(h-R)^2=R^2 \tag{9}$$

$$x^2+h^2-2hR+R^2=R^2. \tag{10}$$

Solving gives

$$R=\frac{x^2+h^2}{2h}=\frac{\frac{1}{3}+\frac{2}{3}}{2\sqrt{\frac{2}{3}}}=\frac{1}{2}\sqrt{\frac{3}{2}}=\tfrac{1}{4}\sqrt{6}\,a\approx 0.61237a. \tag{11}$$

The INRADIUS r is

$$r\equiv h-R=\sqrt{\frac{2}{3}}\,a-\frac{\sqrt{3}}{6}\,a=\tfrac{1}{12}\sqrt{6}\,a\approx 0.20412a, \tag{12}$$

which is also

$$r=\tfrac{1}{4}h=\tfrac{1}{3}R. \tag{13}$$

The MIDRADIUS is

$$\rho=\sqrt{r^2+d^2}=a\sqrt{\tfrac{6}{144}+\tfrac{3}{36}}=\sqrt{\tfrac{1}{8}}\,a=\tfrac{1}{4}\sqrt{2}\,a$$
$$\approx 0.35355a. \tag{14}$$

Plugging in for the VERTICES gives

$$(a\sqrt{3},0,0),(-\tfrac{1}{6}\sqrt{3}\,a,\pm\tfrac{1}{2}a,0),\text{ and }(0,0,\tfrac{1}{3}\sqrt{6}\,a). \tag{15}$$

Since a tetrahedron is a PYRAMID with a triangular base, $V=\frac{1}{3}A_bh$, and

$$V=\tfrac{1}{3}\left(\tfrac{1}{4}\sqrt{3}\,a^2\right)\left(\sqrt{\frac{2}{3}}\,a\right)=\tfrac{1}{12}\sqrt{2}\,a^3. \tag{16}$$

The DIHEDRAL ANGLE is

$$\theta=\tan^{-1}(2\sqrt{2})=2\sin^{-1}(\tfrac{1}{3}\sqrt{6})=\cos^{-1}(\tfrac{1}{3}). \tag{17}$$

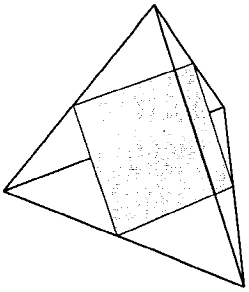

By slicing a tetrahedron as shown above, a SQUARE can be obtained. This cut divides the tetrahedron into two congruent solids rotated by 90°.

Now consider a general (not necessarily regular) tetrahedron, defined as a convex POLYHEDRON consisting of four (not necessarily identical) TRIANGULAR faces. Let the tetrahedron be specified by its VERTICES at (x_i, y_i) where $i = 1, \ldots, 4$. Then the VOLUME is given by

$$V = \frac{1}{3!}\begin{vmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \end{vmatrix}. \tag{18}$$

Specifying the tetrahedron by the three EDGE vectors **a**, **b**, and **c** from a given VERTEX, the VOLUME is

$$V = \tfrac{1}{3!}|\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})|. \tag{19}$$

If the faces are congruent and the sides have lengths a, b, and c, then

$$V = \sqrt{\frac{(a^2+b^2-c^2)(a^2+c^2-b^2)(b^2+c^2-a^2)}{72}} \tag{20}$$

(Klee and Wagon 1991, p. 205). Let a, b, c, and d be the areas of the four faces, and define

$$B \equiv \angle cd \tag{21}$$

$$C \equiv \angle bd \tag{22}$$

$$D \equiv \angle bc, \tag{23}$$

where $\angle jk$ means here the ANGLE between the PLANES formed by the FACES j and k, with VERTEX along their intersecting EDGE. Then

$$a^2 = b^2+c^2+d^2-2cd\cos B-2bd\cos C-2bc\cos D. \tag{24}$$

The analog of GAUSS'S CIRCLE PROBLEM can be asked for tetrahedra: how many LATTICE POINTS lie within a tetrahedron centered at the ORIGIN with a given INRADIUS (Lehmer 1940, Granville 1991, Xu and Yau 1992, Guy 1994).

see also AUGMENTED TRUNCATED TETRAHEDRON, BANG'S THEOREM, EHRHART POLYNOMIAL, HERONIAN TETRAHEDRON, HILBERT'S 3RD PROBLEM, ISOSCELES TETRAHEDRON, SIERPIŃSKI TETRAHEDRON, STELLA OCTANGULA, TETRAHEDRON 5-COMPOUND, TETRAHEDRON 10-COMPOUND, TRUNCATED TETRAHEDRON

References
Davie, T. "The Tetrahedron." `http://www.dcs.st-and.ac.uk/~ad/mathrecs/polyhedra/tetrahedron.html`.
Granville, A. "The Lattice Points of an n-Dimensional Tetrahedron." *Aequationes Math.* **41**, 234–241, 1991.
Guy, R. K. "Gauß's Lattice Point Problem." §F1 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 240–241, 1994.
Klee, V. and Wagon, S. *Old and New Unsolved Problems in Plane Geometry and Number Theory, rev. ed.* Washington, DC: Math. Assoc. Amer., 1991.
Lehmer, D. H. "The Lattice Points of an n-Dimensional Tetrahedron." *Duke Math. J.* **7**, 341–353, 1940.
Xu, Y. and Yau, S. "A Sharp Estimate of the Number of Integral Points in a Tetrahedron." *J. reine angew. Math.* **423**, 199–219, 1992.

Tetrahedron 5-Compound

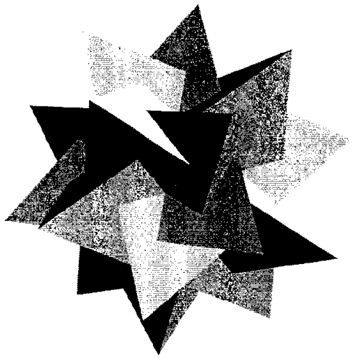

A POLYHEDRON COMPOUND composed of 5 TETRAHEDRA. Two tetrahedron 5-compounds of opposite CHIRALITY combine to make a TETRAHEDRON 10-COMPOUND. The following diagram shows pieces which can be assembled to form a tetrahedron 5-compound (Cundy and Rollett 1989).

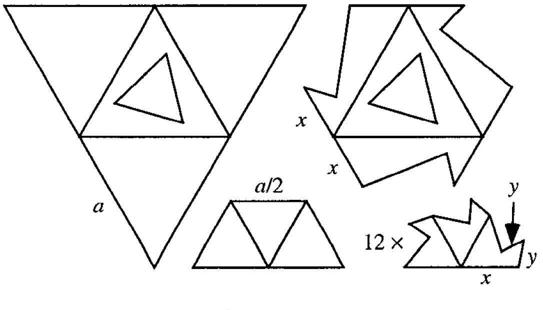

see also POLYHEDRON COMPOUND, TETRAHEDRON 10-COMPOUND

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 135, 1987.
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 139–141, 1989.
Wang, P. "Renderings." `http://www.ugcs.caltech.edu/~peterw/portfolio/renderings/`.
Wenninger, M. J. *Polyhedron Models.* New York: Cambridge University Press, p. 44, 1989.

Tetrahedron 10-Compound

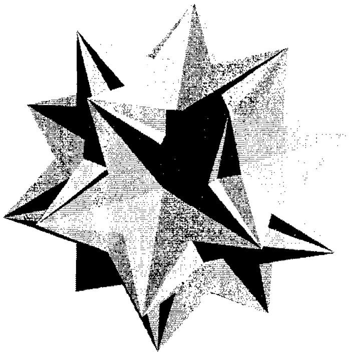

Two TETRAHEDRON 5-COMPOUNDS of opposite CHIRALITY combined.

see also POLYHEDRON COMPOUND, TETRAHEDRON 5-COMPOUND

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 135, 1987.

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 141–142, 1989.

Wenninger, M. J. *Polyhedron Models.* New York: Cambridge University Press, p. 45, 1989.

Tetrahedron Inscribing

Pick four points at random on the surface of a unit SPHERE. Find the distribution of possible volumes of (nonregular) TETRAHEDRA. Without loss of generality, the first point can be chosen as (1, 0, 0). Designate the other points **a**, **b**, and **c**. Then the distances from the first VERTEX are

$$\mathbf{a} = \begin{bmatrix} \cos\theta_1 - 1 \\ \sin\theta_1 \\ 0 \end{bmatrix} \tag{1}$$

$$\mathbf{b} = \begin{bmatrix} \cos\theta_2 \sin\phi_2 - 1 \\ \sin\theta_2 \sin\phi_2 \\ \cos\phi_2 \end{bmatrix} \tag{2}$$

$$\mathbf{c} = \begin{bmatrix} \cos\theta_3 \sin\phi_3 - 1 \\ \sin\theta_3 \sin\phi_3 \\ \cos\phi_3 \end{bmatrix}. \tag{3}$$

The average volume is then

$$\bar{V} = \frac{1}{C}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}\int_{-\pi/2}^{\pi/2}\int_{-\pi/2}^{\pi/2} \tfrac{1}{3!}|\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})| \, d\phi_3\, d\phi_2\, d\theta_3\, d\theta_2\, d\theta_1, \tag{4}$$

where

$$C = \int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi}\int_{-\pi/2}^{\pi/2} d\phi_3\, d\phi_2\, d\theta_3\, d\theta_2\, d\theta_1 = 8\pi^5 \tag{5}$$

and

$$\begin{aligned} \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = {} & -\cos\phi_2 \sin\theta_1 + \cos\phi_3 \sin\theta_1 \\ & -\cos\phi_3 \cos\theta_2 \sin\phi_2 \sin\theta_1 + \cos\phi_2 \cos\theta_3 \sin\phi_3 \sin\theta_1 \\ & -\cos\phi_3 \sin\phi_2 \sin\theta_2 + \cos\phi_3 \cos\theta_1 \sin\phi_2 \sin\theta_2 \\ & +\cos\phi_2 \sin\phi_3 \sin\theta_3 - \cos\phi_2 \cos\theta_1 \sin\phi_3 \sin\theta_3. \end{aligned} \tag{6}$$

The integrals are difficult to compute analytically, but 10^7 computer TRIALS give

$$\langle V\rangle \approx 0.1080 \tag{7}$$

$$\langle V^2\rangle \approx 0.02128 \tag{8}$$

$$\sigma_V{}^2 = \langle V^2\rangle - \langle V\rangle^2 \approx 0.009937. \tag{9}$$

see also POINT-POINT DISTANCE—1-D, TRIANGLE INSCRIBING IN A CIRCLE, TRIANGLE INSCRIBING IN AN ELLIPSE

References

Buchta, C. "A Note on the Volume of a Random Polytope in a Tetrahedron." *Ill. J. Math.* **30**, 653–659, 1986.

Tetrahemihexacron

The DUAL POLYHEDRON of the TETRAHEMIHEXAHEDRON.

Tetrahemihexahedron

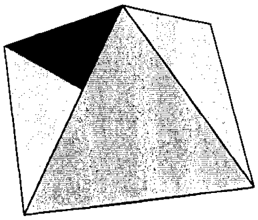

The UNIFORM POLYHEDRON U_4 whose DUAL POLYHEDRON is the TETRAHEMIHEXACRON. It has SCHLÄFLI SYMBOL $r'\{{3 \atop 3}\}$ and WYTHOFF SYMBOL $\frac{3}{2}\,3\,|\,2$. Its faces are $4\{3\} + 3\{4\}$. It is a faceted form of the OCTAHEDRON. Its CIRCUMRADIUS is

$$R = \tfrac{1}{2}\sqrt{2}.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 101–102, 1971.

Tetrakaidecagon

see TETRADECAGON

Tetrakaidecahedron

see TETRADECAHEDRON

Tetrakis Hexahedron

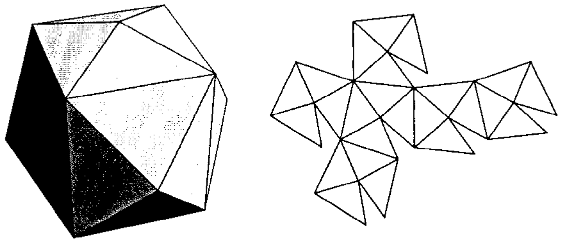

The DUAL POLYHEDRON of the TRUNCATED OCTAHEDRON.

Tetranacci Number

The tetranacci numbers are a generalization of the FIBONACCI NUMBERS defined by $T_0 = 0$, $T_1 = 1$, $T_2 = 1$, $T_3 = 2$, and the RECURRENCE RELATION

$$T_n = T_{n-1} + T_{n-2} + T_{n-3} + T_{n-4}$$

for $n \geq 4$. They represent the $n = 4$ case of the FIBONACCI n-STEP NUMBERS. The first few terms are 1, 1, 2, 4, 8, 15, 29, 56, 108, 208, ... (Sloane's A000078). The ratio of adjacent terms tends to 1.92756, which is the REAL ROOT of $x^5 - 2x^4 + 1 = 0$.

see also FIBONACCI n-STEP NUMBER, FIBONACCI NUMBER, TRIBONACCI NUMBER

References

Sloane, N. J. A. Sequence A000078/M1108 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Tetrix

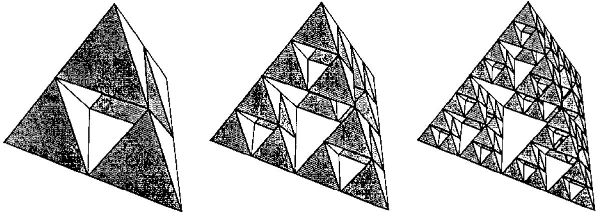

The 3-D analog of the SIERPIŃSKI SIEVE illustrated above, also called the SIERPIŃSKI SPONGE or SIERPIŃSKI TETRAHEDRON. Let N_n be the number of tetrahedra, L_n the length of a side, and A_n the fractional VOLUME of tetrahedra after the nth iteration. Then

$$N_n = 4^n \quad (1)$$

$$L_n = (\tfrac{1}{2})^n = 2^{-n} \quad (2)$$

$$A_n = {L_n}^3 N_n = (\tfrac{1}{2})^n. \quad (3)$$

The CAPACITY DIMENSION is therefore

$$\begin{aligned} d_{\text{cap}} &= -\lim_{n\to\infty} \frac{\ln N_n}{\ln L_n} = -\lim_{n\to\infty} \frac{\ln(4^n)}{\ln(2^{-n})} \\ &= \frac{\ln 4}{\ln 2} = \frac{2\ln 2}{\ln 2} = 2, \end{aligned} \quad (4)$$

so the tetrix has an INTEGRAL CAPACITY DIMENSION (albeit one less than the DIMENSION of the 3-D TETRAHEDRA from which it is built), despite the fact that it is a FRACTAL.

The following illustration demonstrates how this counterintuitive fact can be true by showing three stages of the rotation of a tetrix, viewed along one of its edges. In the last frame, the tetrix "looks" like the 2-D PLANE.

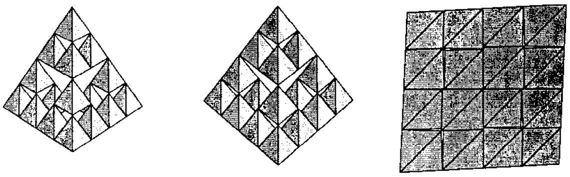

see also MENGER SPONGE, SIERPIŃSKI SIEVE

References

Dickau, R. M. "Sierpinski Tetrahedron." `http://forum.swarthmore.edu/advanced/robertd/tetrahedron.html`.

Eppstein, D. "Sierpinski Tetrahedra and Other Fractal Sponges." `http://www.ics.uci.edu/~eppstein/junkyard/sierpinski.html`.

Tetromino

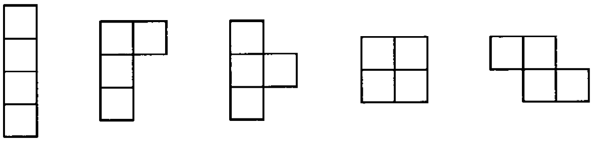

The five 4-POLYOMINOES, known as STRAIGHT, L-, T-, SQUARE, and SKEW.

References

Gardner, M. "Polyominoes." Ch. 13 in *The Scientific American Book of Mathematical Puzzles & Diversions.* New York: Simon and Schuster, pp. 124–140, 1959.

Hunter, J. A. H. and Madachy, J. S. *Mathematical Diversions.* New York: Dover, pp. 80–81, 1975.

Thales' Theorem

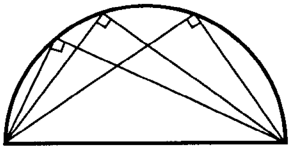

An ANGLE inscribed in a SEMICIRCLE is a RIGHT ANGLE.

see also RIGHT ANGLE, SEMICIRCLE

Theorem

A statement which can be demonstrated to be true by accepted mathematical operations and arguments. In general, a theorem is an embodiment of some general principle that makes it part of a larger theory.

According to the Nobel Prize-winning physicist Richard Feynman (1985), any theorem, no matter how difficult to prove in the first place, is viewed as "TRIVIAL" by mathematicians once it has been proven. Therefore, there are exactly two types of mathematical objects: TRIVIAL ones, and those which have not yet been proven.

see also AXIOM, AXIOMATIC SYSTEM, COROLLARY, DEEP THEOREM, PORISM, LEMMA, POSTULATE, PRINCIPLE, PROPOSITION

References

Feynman, R. P. and Leighton, R. *Surely You're Joking, Mr. Feynman!* New York: Bantam Books, 1985.

Theorema Egregium

see GAUSS'S THEOREMA EGREGIUM

Theta Function

The theta functions are the elliptic analogs of the EXPONENTIAL FUNCTION, and may be used to express the JACOBI ELLIPTIC FUNCTIONS. Let t be a constant COMPLEX NUMBER with $\Im[t] > 0$. Define the NOME

$$q \equiv e^{i\pi t} = e^{\pi K'(k)/K(k)}, \tag{1}$$

where

$$t \equiv -i\frac{K'(k)}{K(k)}, \tag{2}$$

and $K(k)$ is a complete ELLIPTIC INTEGRAL OF THE FIRST KIND, k is the MODULUS, and k' is the complementary MODULUS. Then the theta functions are, in the NOTATION of Whittaker and Watson,

$$\begin{aligned}\vartheta_1(z,q) &\equiv 2\sum_{n=0}^{\infty}(-1)^n q^{(n+1/2)^2}\sin[(2n+1)z]\\ &= zq^{1/4}\sum_{n=0}^{\infty}(-1)^n q^{n(n+1)}\sin[(2n+1)z] \qquad (3)\end{aligned}$$

$$\begin{aligned}\vartheta_2(z,q) &\equiv 2\sum_{n=0}^{\infty} q^{(n+1/2)^2}\cos[(2n+1)z]\\ &= 2q^{1/4}\sum_{n=0}^{\infty} q^{n(n+1)}\cos[(2n+1)z] \qquad (4)\end{aligned}$$

$$\vartheta_3(z,q) \equiv 1+2\sum_{n=1}^{\infty} q^{n^2}\cos(2nz) \tag{5}$$

$$\begin{aligned}\vartheta_4(z,q) &\equiv \sum_{n=-\infty}^{\infty}(-1)^n q^{n^2} e^{2niz}\\ &= 1+2\sum_{n=1}^{\infty}(-1)^n q^{n^2}\cos(2nz). \qquad (6)\end{aligned}$$

Written in terms of t,

$$\vartheta_2(t,q) = \sum_{n=-\infty}^{\infty} q^{(n+1/2)^2} e^{\Im[t]} \tag{7}$$

$$\vartheta_3(t,q) = \sum_{n=-\infty}^{\infty} q^{n^2} e^{\Im[t]}. \tag{8}$$

These functions are sometimes denoted Θ_i or θ_i, and a number of indexing conventions have been used. For a summary of these notations, see Whittaker and Watson (1990). The theta functions are quasidoubly periodic, as illustrated in the following table.

ϑ_i	$\vartheta_i(z+\pi)/\vartheta_i(z)$	$\vartheta_i(z+t\pi)/\vartheta_i(z)$
ϑ_1	-1	$-N$
ϑ_2	-1	N
ϑ_3	1	N
ϑ_4	1	$-N$

Here,

$$N \equiv q^{-1}e^{-2iz}. \tag{9}$$

The quasiperiodicity can be established as follows for the specific case of ϑ_4,

$$\begin{aligned}\vartheta_4(z+\pi,q) &= \sum_{n=-\infty}^{\infty}(-1)^n q^{n^2} e^{2niz} e^{2ni\pi}\\ &= \sum_{n=-\infty}^{\infty}(-1)^n q^{n^2} e^{2niz} = \vartheta_4(z,q) \qquad (10)\end{aligned}$$

$$\begin{aligned}\vartheta_4(z+\pi t,q) &= \sum_{n=-\infty}^{\infty}(-1)^n q^{n^2} e^{2ni\pi t} e^{2niz}\\ &= \sum_{n=-\infty}^{\infty}(-1)^n q^{n^2} q^{2n} e^{2niz}\\ &= -q^{-1}e^{-2iz}\sum_{n=-\infty}^{\infty}(-1)^{n+1} q^{(n+1)^2} q^{2(n+1)iz}\\ &= -q^{-1}e^{-2iz}\sum_{n=-\infty}^{\infty}(-1)^n q^{n^2} q^{2niz}\\ &= -q^{-1}e^{-2iz}\vartheta_4(z,q). \qquad (11)\end{aligned}$$

The theta functions can be written in terms of each other:

$$\vartheta_1(z,q) = -ie^{iz+\pi it/4}\vartheta_4(z+\tfrac{1}{4}\pi t,q) \tag{12}$$

$$\vartheta_2(z,q) = \vartheta_1(z+\tfrac{1}{2}\pi,q) \tag{13}$$

$$\vartheta_3(z,q) = \vartheta_4(z+\tfrac{1}{2}\pi,q). \tag{14}$$

Any theta function of given arguments can be expressed in terms of any other two theta functions with the same arguments.

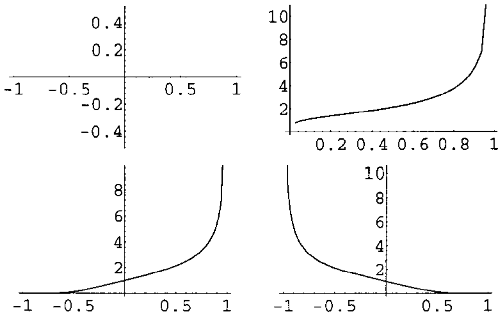

Define

$$\vartheta_i \equiv \vartheta_i(z=0), \tag{15}$$

which are plotted above. Then we have the identities

$$\vartheta_1{}^2(z)\vartheta_4{}^2 = \vartheta_3{}^2(z)\vartheta_2{}^2 - \vartheta_2{}^2(z)\vartheta_3{}^2 \tag{16}$$

$$\vartheta_2{}^2(z)\vartheta_4{}^2 = \vartheta_4{}^2(z)\vartheta_2{}^2 - \vartheta_1{}^2(z)\vartheta_3{}^2 \tag{17}$$

$$\vartheta_3{}^2(z)\vartheta_4{}^2 = \vartheta_4{}^2(z)\vartheta_3{}^2 - \vartheta_1{}^2(z)\vartheta_2{}^2 \tag{18}$$

$$\vartheta_4{}^2(z)\vartheta_4{}^2 = \vartheta_3{}^2(z)\vartheta_3{}^2 - \vartheta_2{}^2(z)\vartheta_2{}^2. \tag{19}$$

Taking $z = 0$ in the last gives the special case

$$\vartheta_4{}^4 = \vartheta_3{}^4 - \vartheta_2{}^4. \tag{20}$$

In addition,

$$\vartheta_3(x) = \sum_{n=-\infty}^{\infty} x^{n^2} = 1 + 2x + 2x^4 + 2x^9 + \ldots \tag{21}$$

$$\vartheta_3{}^2(x) = 1 + 4\left(\frac{x}{1-x} - \frac{x^3}{1-x^3} + \frac{x^5}{1-x^5} - \frac{x^7}{1-x^7} + \ldots\right) \tag{22}$$

$$\vartheta_3{}^4(x) = 1 + 8\left(\frac{x}{1-x} + \frac{2x^2}{1+x^2} + \frac{3x^3}{1-x^3} + \frac{4x^4}{1+x^4} + \ldots\right). \tag{23}$$

The theta functions obey addition rules such as

$$\vartheta_3(z+y)\vartheta_3(z-y)\vartheta_3{}^2 = \vartheta_3{}^2(y)\vartheta_3{}^2(z) + \vartheta_1{}^2(y)\vartheta_1{}^2(z). \tag{24}$$

Letting $y = z$ gives a duplication FORMULA

$$\vartheta_3(2z)\vartheta_3{}^3 = \vartheta_3{}^4(z) + \vartheta_1{}^4(z). \tag{25}$$

For more addition FORMULAS, see Whittaker and Watson (1990, pp. 487–488). Ratios of theta function derivatives to the functions themselves have the simple forms

$$\frac{\vartheta_1'(z)}{\vartheta_1(z)} = \cot z + 4\sum_{n=1}^{\infty} \frac{q^{2n}}{1-q^{2n}} \sin(2nz) \tag{26}$$

$$\frac{\vartheta_2'(z)}{\vartheta_2(z)} = -\tan z + 4\sum_{n=1}^{\infty} (-1)^n \frac{q^{2n}}{1-q^{2n}} \sin(2nz) \tag{27}$$

$$\frac{\vartheta_3'(z)}{\vartheta_3(z)} = 4\sum_{n=1}^{\infty} (-1)^n \frac{q^n}{1-q^{2n}} \sin(2nz) \tag{28}$$

$$\begin{aligned}\frac{\vartheta_4'(z)}{\vartheta_4(z)} &= \sum_{n=1}^{\infty} \frac{q^{2n-1}\sin(2z)}{1-2q^{2n-1}\cos(2z)+q^{4n-2}} \\ &= \sum_{n=1}^{\infty} \frac{4q^n \sin(2nz)}{1-q^{2n}}.\end{aligned} \tag{29}$$

The theta functions can be expressed as products instead of sums by

$$\vartheta_1(z) = 2Gq^{1/4}\sin z \prod_{n=1}^{\infty}[1 - 2q^{2n}\cos(2z) + q^{4n}] \tag{30}$$

$$\vartheta_2(z) = 2Gq^{1/4}\cos z \prod_{n=1}^{\infty}[1 + 2q^{2n}\cos(2z) + q^{4n}] \tag{31}$$

$$\vartheta_3(z) = G\prod_{n=1}^{\infty}[1 + 2q^{2n-1}\cos(2z) + q^{4n-2}] \tag{32}$$

$$\vartheta_4(z) = G\prod_{n=1}^{\infty}[1 - 2q^{2n-1}\cos(2z) + q^{4n-2}], \tag{33}$$

where

$$G \equiv \prod_{n=1}^{\infty}(1 - q^{2n}) \tag{34}$$

(Whittaker and Watson 1990, pp. 469–470).

The theta functions satisfy the PARTIAL DIFFERENTIAL EQUATION

$$\tfrac{1}{4}\pi i\frac{\partial^2 y}{\partial z^2} + \frac{\partial y}{\partial t} = 0, \tag{35}$$

where $y \equiv \vartheta_j(z|t)$. Ratios of the theta functions with ϑ_4 in the DENOMINATOR also satisfy differential equations

$$\frac{d}{dz}\left[\frac{\vartheta_1(z)}{\vartheta_4(z)}\right] = \vartheta_4{}^2\frac{\vartheta_2(z)\vartheta_3(z)}{\vartheta_4{}^2(z)} \tag{36}$$

$$\frac{d}{dz}\left[\frac{\vartheta_2(z)}{\vartheta_4(z)}\right] = -\vartheta_3{}^2\frac{\vartheta_1(z)\vartheta_3(z)}{\vartheta_4{}^2(z)} \tag{37}$$

$$\frac{d}{dz}\left[\frac{\vartheta_3(z)}{\vartheta_4(z)}\right] = -\vartheta_2{}^2\frac{\vartheta_1(z)\vartheta_2(z)}{\vartheta_4{}^2(z)}. \tag{38}$$

Some additional remarkable identities are

$$\vartheta_1' = \vartheta_2\vartheta_3\vartheta_4 \tag{39}$$

$$\vartheta_3(z,t) = -(it)^{1/2}e^{z^2/\pi it}\vartheta_3\left(\frac{2}{t}, -\frac{1}{t}\right), \tag{40}$$

which were discovered by Poisson in 1827 and are equivalent to

$$\sum_{n=-\infty}^{\infty} e^{-t(x+n)^2} = \sqrt{\frac{\pi}{t}}\sum_{k=-\infty}^{\infty} 2^{2\pi ikx-(\pi^2k^2/t)}. \tag{41}$$

Another amazing identity is

$$\begin{aligned}&2\vartheta_1[\tfrac{1}{2}(-b+c+d+e)]\vartheta_2[\tfrac{1}{2}(b-c+d+e)]\vartheta_3[\tfrac{1}{2}(b+c-d+e)] \\ &\quad\times\vartheta_4[\tfrac{1}{2}(b+c+d-e)] = \vartheta_3(b)\vartheta_4(c)\vartheta_1(d)\vartheta_2(e) \\ &\quad+\vartheta_2(b)\vartheta_1(c)\vartheta_4(d)\vartheta_3(e) - \vartheta_1(b)\vartheta_2(c)\vartheta_3(d)\vartheta_4(e) \\ &\quad+\vartheta_4(b)\vartheta_3(c)\vartheta_2(d)\vartheta_1(e)\end{aligned} \tag{42}$$

(Whittaker and Watson 1990, p. 469).

The complete ELLIPTIC INTEGRALS OF THE FIRST and SECOND KINDS can be expressed using theta functions. Let

$$\xi \equiv \frac{\vartheta_1(z)}{\vartheta_4(z)}, \tag{43}$$

and plug into (36)

$$\left(\frac{d\xi}{dz}\right)^2 = ({\vartheta_2}^2 - \xi^2{\vartheta_3}^2)({\vartheta_3}^2 - \xi^2{\vartheta_2}^2). \tag{44}$$

Now write

$$\xi\frac{\vartheta_3}{\vartheta_2} \equiv y \tag{45}$$

and

$$z{\vartheta_3}^2 \equiv u. \tag{46}$$

Then

$$\left(\frac{dy}{du}\right)^2 = (1-y^2)(1-k^2y^2), \tag{47}$$

where the MODULUS is defined by

$$k = k(q) = \frac{{\vartheta_2}^2(q)}{{\vartheta_3}^2(q)}. \tag{48}$$

Define also the complementary MODULUS

$$k' = k'(q) = \frac{{\vartheta_4}^2(-q)}{{\vartheta_3}^2(q)}. \tag{49}$$

Now, since

$${\vartheta_2}^4 + {\vartheta_4}^4 = {\vartheta_3}^4, \tag{50}$$

we have shown

$$k^2 + k'^2 = 1. \tag{51}$$

The solution to the equation is

$$y = \frac{\vartheta_3}{\vartheta_2}\frac{\vartheta_1(u{\vartheta_3}^{-2}|t)}{\vartheta_4(u{\vartheta_3}^{-2}|t)} \equiv \mathrm{sn}(u,k), \tag{52}$$

which is a JACOBI ELLIPTIC FUNCTION with periods

$$4K(k) = 2\pi{\vartheta_3}^2(q) \tag{53}$$

and

$$2iK'(k) = \pi t{\vartheta_3}^2(q). \tag{54}$$

Here, K is the complete ELLIPTIC INTEGRAL OF THE FIRST KIND,

$$K(k) = \tfrac{1}{2}\pi{\vartheta_3}^2(q). \tag{55}$$

see also BLECKSMITH-BRILLHART-GERST THEOREM, ELLIPTIC FUNCTION, ETA FUNCTION, EULER'S PENTAGONAL NUMBER THEOREM, JACOBI ELLIPTIC FUNCTIONS, JACOBI TRIPLE PRODUCT, LANDEN'S FORMULA, MOCK THETA FUNCTION, MODULAR EQUATION, MODULAR TRANSFORMATION, MORDELL INTEGRAL, NEVILLE THETA FUNCTION, NOME, POINCARÉ-FUCHS-KLEIN AUTOMORPHIC FUNCTION, PRIME THETA FUNCTION, QUINTUPLE PRODUCT IDENTITY, RAMANUJAN THETA FUNCTIONS, SCHRÖTER'S FORMULA, WEBER FUNCTIONS

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 577, 1972.

Bellman, R. E. *A Brief Introduction to Theta Functions.* New York: Holt, Rinehart and Winston, 1961.

Berndt, B. C. "Theta-Functions and Modular Equations." Ch. 25 in *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 138–244, 1994.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 430–432, 1953.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Theta Operator

In the NOTATION of Watson (1966),

$$\vartheta \equiv z\frac{d}{dz}.$$

References

Watson, G. N. *A Treatise on the Theory of Bessel Functions, 2nd ed.* Cambridge, England: Cambridge University Press, 1966.

Theta Subgroup

see LAMBDA GROUP

Thiele's Interpolation Formula

Let ρ be a RECIPROCAL DIFFERENCE. Then Thiele's interpolation formula is the CONTINUED FRACTION

$$f(x) = f(x_1) + \frac{x-x_1}{\rho(x_1,x_2)+}\,\frac{x-x_2}{\rho_2(x_1,x_2,x_3)-f(x_1)+}\\ \frac{x-x_3}{\rho_3(x_1,x_2,x_3,x_4)-\rho(x_1,x_2)+\ldots}.$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 881, 1972.

Milne-Thomson, L. M. *The Calculus of Finite Differences.* London: Macmillan, 1951.

Thiessen Polytope

see VORONOI POLYGON

Third Curvature

Also known as the TOTAL CURVATURE. The linear element of the INDICATRIX

$$ds_P = \sqrt{{ds_T}^2 + {ds_B}^2}.$$

see also LANCRET EQUATION

Thirteenth

see FRIDAY THE THIRTEENTH

Thom's Eggs

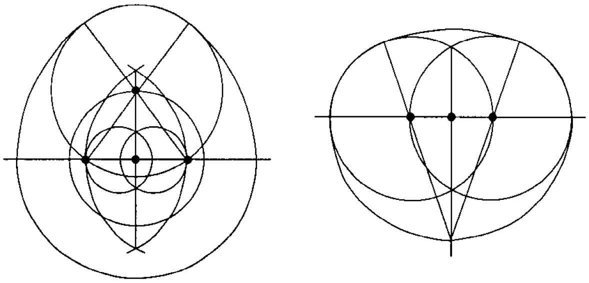

EGG-shaped curves constructed using multiple CIRCLES which Thom (1967) used to model Megalithic stone rings in Britain.

see also EGG, OVAL

References
Dixon, R. *Mathographics.* New York: Dover, p. 6, 1991.
Thom, A. "Mathematical Background." Ch. 4 in *Megalithic Sites in Britain.* Oxford, England: Oxford University Press, pp. 27–33, 1967.

Thomae's Theorem

$$\frac{\Gamma(x+y+s+1)}{\Gamma(x+s+1)\Gamma(y+s+1)}\,{}_3F_2\left(\begin{matrix}-a,-b,x+y+s+1\\ x+s+1,y+s+1\end{matrix};1\right) = \frac{\Gamma(a+b+s+1)}{\Gamma(a+s+1)\Gamma(b+s+1)}\,{}_3F_2\left(\begin{matrix}-x,-y,a+b+s+1\\ a+s+1,b+s+1\end{matrix};1\right),$$

where $\Gamma(z)$ is the GAMMA FUNCTION and the function ${}_3F_2(a,b,c;d,e;z)$ is a GENERALIZED HYPERGEOMETRIC FUNCTION.

see also GENERALIZED HYPERGEOMETRIC FUNCTION

References
Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, pp. 104–105, 1959.

Thomassen Graph

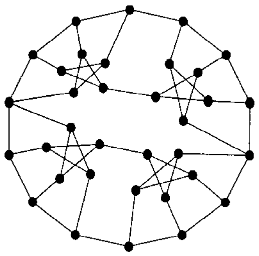

The GRAPH illustrated above.

see also THOMSEN GRAPH

Thompson's Functions

see BEI, BER, KELVIN FUNCTIONS

Thompson Group

The SPORADIC GROUP *Th*.

References
Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas/Th.html`.

Thomsen's Figure

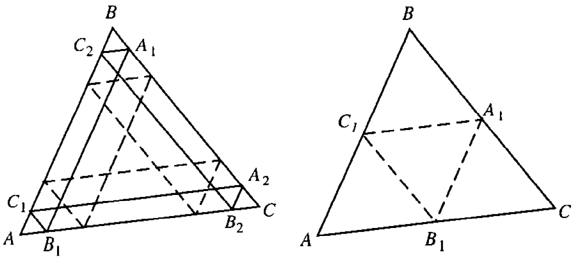

Take any TRIANGLE with VERTICES A, B, and C. Pick a point A_1 on the side opposite A, and draw a line PARALLEL to AB. Upon reaching the side AC at B_1, draw the line PARALLEL to BC. Continue (left figure). Then $A_3 = A_1$ for any TRIANGLE. If A_1 is the MIDPOINT of BC, then $A_2 = A_1$ (right figure).

see also MIDPOINT, TRIANGLE

References
Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 234, 1979.

Thomsen Graph

The COMPLETE BIPARTITE GRAPH $K_{3,3}$, which is equivalent to the UTILITY GRAPH. It has a CROSSING NUMBER 1.

see also COMPLETE BIPARTITE GRAPH, CROSSING NUMBER (GRAPH), THOMASSEN GRAPH, UTILITY GRAPH

Thomson Lamp Paradox

A lamp is turned on for 1/2 minute, off for 1/4 minute, on for 1/8 minute, etc. At the end of one minute, the lamp switch will have been moved $\aleph_0$ times, where $\aleph_0$ is ALEPH-0. Will the lamp be on or off? This PARADOX is actually nonsensical, since it is equivalent to asking if the "last" INTEGER is EVEN or ODD.

References
Pickover, C. A. *Keys to Infinity.* New York: Wiley, pp. 19–23, 1995.

Thomson's Principle

see DIRICHLET'S PRINCIPLE

Thomson Problem

Determine the stable equilibrium positions of N classical electrons constrained to move on the surface of a SPHERE and repelling each other by an inverse square law. Exact solutions for $N = 2$ to 8 are known, but $N = 9$ and 11 are still unknown.

In reality, Earnshaw's theorem guarantees that no system of discrete electric charges can be held in stable equilibrium under the influence of their electrical interaction alone (Aspden 1987).

see also FEJES TÓTH'S PROBLEM

References

Altschuler, E. L.; Williams, T. J.; Ratner, E. R.; Dowla, F.; and Wooten, F. "Method of Constrained Global Optimization." *Phys. Rev. Let.* **72**, 2671–2674, 1994.

Altschuler, E. L.; Williams, T. J.; Ratner, E. R.; Dowla, F.; and Wooten, F. "Method of Constrained Global Optimization—Reply." *Phys. Rev. Let.* **74**, 1483, 1995.

Ashby, N. and Brittin, W. E. "Thomson's Problem." *Amer. J. Phys.* **54**, 776–777, 1986.

Aspden, H. "Earnshaw's Theorem." *Amer. J. Phys.* **55**, 199–200, 1987.

Berezin, A. A. "Spontaneous Symmetry Breaking in Classical Systems." *Amer. J. Phys.* **53**, 1037, 1985.

Calkin, M. G.; Kiang, D.; and Tindall, D. A. "Minimum Energy Configurations." *Nature* **319**, 454, 1986.

Erber, T. and Hockney, G. M. "Comment on 'Method of Constrained Global Optimization.'" *Phys. Rev. Let.* **74**, 1482–1483, 1995.

Marx, E. "Five Charges on a Sphere." *J. Franklin Inst.* **290**, 71–74, Jul. 1970.

Melnyk, T. W.; Knop, O.; and Smith, W. R. "Extremal Arrangements of Points and Unit Charges on a Sphere: Equilibrium Configurations Revisited." *Canad. J. Chem.* **55**, 1745–1761, 1977.

Whyte, L. L. "Unique Arrangement of Points on a Sphere." *Amer. Math. Monthly* **59**, 606–611, 1952.

Thousand

1,000 = 10^3. The word "thousand" appears in common expressions in a number of languages, for example, "a thousand pardons" in English and "tusen takk" ("a thousand thanks") in Norwegian.

see also HUNDRED, LARGE NUMBER, MILLION

Three

see 3

Three-Colorable

see COLORABLE

Three-In-A-Row

see TIC-TAC-TOE

Three Jug Problem

Given three jugs with x pints in the first, y in the second, and z in the third, obtain a desired amount in one of the vessels by completely filling up and/or emptying vessels into others. This problem can be solved with the aid of TRILINEAR COORDINATES.

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 89–93, 1967.

Three-Valued Logic

A logical structure which does not assume the EXCLUDED MIDDLE LAW. Three possible truth values are possible: true, false, or undecided. There are 3072 such logics.

see also EXCLUDED MIDDLE LAW, FUZZY LOGIC, LOGIC

Threefoil Knot

see TREFOIL KNOT

Thue Constant

The base-2 TRANSCENDENTAL NUMBER

$$0.11011011111011011111\ldots_2,$$

where the nth bit is 1 if n is not divisible by 3 and is the complement of the $(n/3)$th bit if n is divisible by 3. It is also given by the SUBSTITUTION MAP

$$\begin{aligned} 0 &\to 111 \\ 1 &\to 110. \end{aligned}$$

In decimal, the Thue constant equals 0.8590997969....

see also RABBIT CONSTANT, THUE-MORSE CONSTANT

References

Thue-Morse Constant

The constant also called the PARITY CONSTANT and defined by

$$P \equiv \tfrac{1}{2}\sum_{n=0}^{\infty} P(n)2^{-n} = 0.4124540336401075977\ldots \quad (1)$$

(Sloane's A014571), where $P(n)$ is the PARITY of n. Dekking (1977) proved that the Thue-Morse constant is TRANSCENDENTAL, and Allouche and Shallit give a complete proof correcting a minor error of Dekking.

The Thue-Morse constant can be written in base 2 by stages by taking the previous iteration a_n, taking the complement $\overline{a_n}$, and appending, producing

$$\begin{aligned} a_0 &= 0.0_2 \\ a_1 &= 0.01_2 \\ a_2 &= 0.0110_2 \\ a_3 &= 0.01101001_2 \\ a_4 &= 0.0110100110010110_2. \end{aligned} \quad (2)$$

This can be written symbolically as

$$a_{n+1} = a_n + \overline{a_n} \cdot 2^{-2^n} \quad (3)$$

with $a_0 = 0$. Here, the complement is the number $\overline{a_n}$ such that $a_n + \overline{a_n} = 0.11\ldots_2$, which can be found from

$$a_n + \overline{a_n} = \sum_{k=1}^{2^n} (\tfrac{1}{2})^k = \frac{1 - (\frac{1}{2})^{2^n}}{1 - \frac{1}{2}} - 1 = 1 - 2^{-2^n}. \quad (4)$$

Therefore,

$$\overline{a_n} = 1 - a_n - 2^{-2^n}, \tag{5}$$

and

$$a_{n+1} = a_n + (1 - 2^{-2^n} - a_n)2^{-2^n}. \tag{6}$$

The regular CONTINUED FRACTION for the Thue-Morse constant is [0 2 2 2 1 4 3 5 2 1 4 2 1 5 44 1 4 1 2 4 1 1 1 5 14 1 50 15 5 1 1 1 4 2 1 4 1 43 1 4 1 2 1 3 16 1 2 1 2 1 50 1 2 424 1 2 5 2 1 1 1 5 5 2 22 5 1 1 1 1274 3 5 2 1 1 1 4 1 1 15 154 7 2 1 2 2 1 2 1 1 50 1 4 1 2 867374 1 1 1 5 5 1 1 6 1 2 7 2 1650 23 3 1 1 1 2 5 3 84 1 1 1 1284 ...] (Sloane's A014572), and seems to continue with sporadic large terms in suspicious-looking patterns. A nonregular CONTINUED FRACTION is

$$P = \cfrac{1}{3 - \cfrac{1}{2 - \cfrac{1}{4 - \cfrac{3}{16 - \cfrac{15}{256 - \cfrac{255}{65536 - \ldots}}}}}}. \tag{7}$$

A related infinite product is

$$4P = 2 - \frac{1 \cdot 3 \cdot 15 \cdot 255 \cdot 65535 \cdots}{2 \cdot 4 \cdot 16 \cdot 256 \cdot 65536 \cdots}. \tag{8}$$

The SEQUENCE $a_\infty = 0110100110010110100101100\ldots$ (Sloane's A010060) is known as the THUE-MORSE SEQUENCE.

see also RABBIT CONSTANT, THUE CONSTANT

References

Allouche, J. P.; Arnold, A.; Berstel, J.; Brlek, S.; Jockusch, W.; Plouffe, S.; and Sagan, B. "A Relative of the Thue-Morse Sequence." *Discr. Math.* **139**, 455–461, 1995.

Allouche, J. P. and Shallit, J. In preparation.

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Item 122, Feb. 1972.

Dekking, F. M. "Transcendence du nombre de Thue-Morse." *Comptes Rendus de l'Academie des Sciences de Paris* **285**, 157–160, 1977.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/cntfrc/cntfrc.html`.

Sloane, N. J. A. Sequences A010060, A014571, and A014572 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Thue-Morse Sequence

The INTEGER SEQUENCE (also called the MORSE-THUE SEQUENCE)

$$0110100110010110100101100110 1001\ldots \tag{1}$$

(Sloane's A010060) which arises in the THUE-MORSE CONSTANT. It can be generated from the SUBSTITUTION MAP

$$0 \to 01 \tag{2}$$

$$1 \to 10 \tag{3}$$

starting with 0 as follows:

$$0 \to 01 \to 0110 \to 01101001 \to \ldots. \tag{4}$$

Writing the sequence as a POWER SERIES over the GALOIS FIELD GF(2),

$$F(x) = 0 + 1x + 1x^2 + 0x^3 + 1x^4 + \ldots, \tag{5}$$

then F satisfies the quadratic equation

$$(1+x)F^2 + F = \frac{x}{1+x^2} \pmod{2}. \tag{6}$$

This equation has two solutions, F and F', where F' is the complement of F, i.e.,

$$F + F' = 1 + x + x^2 + x^3 + \ldots = \frac{1}{1+x}, \tag{7}$$

which is consistent with the formula for the sum of the roots of a quadratic. The equality (6) can be demonstrated as follows. Let $(abcdef\ldots)$ be a shorthand for the POWER series

$$a + bx + cx^2 + dx^3 + \ldots, \tag{8}$$

so $F(x)$ is $(0110100110010110\ldots)$. To get F^2, simply use the rule for squaring POWER SERIES over GF(2)

$$(A+B)^2 = A^2 + B^2 \pmod{2}, \tag{9}$$

which extends to the simple rule for squaring a POWER SERIES

$$(a_0 + a_1x + a_2x^2 + \ldots)^2 = a_0 + a_1x^2 + a_2x^4 + \ldots \pmod{2}, \tag{10}$$

i.e., space the series out by a factor of 2, (0 1 1 0 1 0 0 1 ...), and insert zeros in the ODD places to get

$$F^2 = (0010100010000010\ldots). \tag{11}$$

Then multiply by x (which just adds a zero at the front) to get

$$xF^2 = (00010100010000010\ldots). \tag{12}$$

Adding to F^2 gives

$$(1+x)F^2 = (0011110011000011\ldots). \tag{13}$$

This is the first term of the quadratic equation, which is the Thue-Morse sequence with each term doubled up. The next term is F, so we have

$$(1+x)F^2 = (0011110011000011\ldots) \tag{14}$$

$$F = (0110100110010110\ldots). \tag{15}$$

The sum is the above two sequences XORed together (there are no CARRIES because we're working over GF(2)), giving

$$(1+x)F^2+F=(0101010101010101\ldots). \tag{16}$$

We therefore have

$$\begin{aligned}(1+x)F^2+F&=\frac{x}{1+x^2}\\&=x+x^3+x^5+x^7+x^9+x^{11}+\ldots\pmod 2.\end{aligned} \tag{17}$$

The Thue-Morse sequence is an example of a cube-free sequence on two symbols (Morse and Hedlund 1944), i.e., it contains no substrings of the form WWW, where W is any WORD. For example, it does not contain the WORDS 000, 010101 or 010010010. In fact, the following stronger statement is true: the Thue-Morse sequence does not contain any substrings of the form WWa, where a is the first symbol of W. We can obtain a SQUAREFREE sequence on three symbols by doing the following: take the Thue-Morse sequence 0110100110010110... and look at the sequence of WORDS of length 2 that appear: 01 11 10 01 10 00 01 11 10 .... Replace 01 by 0, 10 by 1, 00 by 2 and 11 by 2 to get the following: 021012021.... Then this SEQUENCE is SQUAREFREE (Morse and Hedlund 1944).

The Thue-Morse sequence has important connections with the GRAY CODE. Kindermann generates fractal music using the SELF-SIMILARITY of the Thue-Morse sequence.

see also GRAY CODE, PARITY CONSTANT, RABBIT SEQUENCE, THUE SEQUENCE

References

Kindermann, L. "MusiNum—The Music in the Numbers." `http://www.forwiss.uni-erlangen.de/~kinderma/musinum/`.

Morse, M. and Hedlund, G. A. "Unending Chess, Symbolic Dynamics, and a Problem in Semigroups." *Duke Math. J.* **11**, 1–7, 1944.

Schroeder, M. R. *Fractals, Chaos, and Power Laws: Minutes from an Infinite Paradise.* New York: W. H. Freeman, 1991.

Sloane, N. J. A. Sequence A010060 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Thue Sequence

The SEQUENCE of BINARY DIGITS of the THUE CONSTANT, $0.110110111110110111110110110\ldots_2$ (Sloane's A014578).

see also RABBIT CONSTANT, THUE CONSTANT

References

Guy, R. K. "Thue Sequences." §E21 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 223–224, 1994.

Sloane, N. J. A. Sequence A014578 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Thue-Siegel-Roth Theorem

If α is a TRANSCENDENTAL NUMBER, it can be approximated by infinitely many RATIONAL NUMBERS m/n to within n^{-r}, where r is any POSITIVE number.

see also LIOUVILLE'S RATIONAL APPROXIMATION THEOREM, LIOUVILLE-ROTH CONSTANT, ROTH'S THEOREM

Thue-Siegel-Schneider-Roth Theorem

see THUE-SIEGEL-ROTH THEOREM

Thue's Theorem

If $n>1$, $(a,n)=1$ (i.e., a and n are RELATIVELY PRIME), and m is the least integer $>\sqrt{n}$, then there exist an x and y such that

$$ay\equiv\pm x\pmod n$$

where $0<x<m$ and $0<y<m$.

References

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 161, 1993.

Thurston's Geometrization Conjecture

Thurston's conjecture has to do with geometric structures on 3-D MANIFOLDS. Before stating Thurston's conjecture, some background information is useful. 3-dimensional MANIFOLDS possess what is known as a standard 2-level DECOMPOSITION. First, there is the CONNECTED SUM DECOMPOSITION, which says that every COMPACT 3-MANIFOLD is the CONNECTED SUM of a unique collection of PRIME 3-MANIFOLDS.

The second DECOMPOSITION is the JACO-SHALEN-JOHANNSON TORUS DECOMPOSITION, which states that irreducible orientable COMPACT 3-MANIFOLDS have a canonical (up to ISOTOPY) minimal collection of disjointly EMBEDDED incompressible TORI such that each component of the 3-MANIFOLD removed by the TORI is either "atoroidal" or "Seifert-fibered."

Thurston's conjecture is that, after you split a 3-MANIFOLD into its CONNECTED SUM and then JACO-SHALEN-JOHANNSON TORUS DECOMPOSITION, the remaining components each admit exactly one of the following geometries:

1. EUCLIDEAN GEOMETRY,
2. HYPERBOLIC GEOMETRY,
3. SPHERICAL GEOMETRY,
4. the GEOMETRY of $\mathbb{S}^2\times\mathbb{R}$,
5. the GEOMETRY of $\mathbb{H}^2\times\mathbb{R}$,
6. the GEOMETRY of SL_2R,
7. NIL GEOMETRY, or
8. SOL GEOMETRY.

Here, $\mathbb{S}^2$ is the 2-Sphere and $\mathbb{H}^2$ is the Hyperbolic Plane. If Thurston's conjecture is true, the truth of the Poincaré Conjecture immediately follows.

see also Connected Sum Decomposition, Euclidean Geometry, Hyperbolic Geometry, Jaco-Shalen-Johannson Torus Decomposition, Nil Geometry, Poincaré Conjecture, Sol Geometry, Spherical Geometry

Thwaites Conjecture

see Collatz Problem

Tic-Tac-Toe

The usual game of tic-tac-toe (also called Ticktacktoe) is 3-in-a-row on a 3 × 3 board. However, a generalized n-in-a-Row on an $n \times m$ board can also be considered. For $n = 1$ and 2 the first player can always win. If the board is at least 3 × 4, the first player can win for $n = 3$.

However, for Tic-Tac-Toe which uses a 3 × 3 board, a draw can always be obtained. If the board is at least 4 × 30, the first player can win for $n = 4$. For $n = 5$, a draw can always be obtained on a 5 × 5 board, but the first player can win if the board is at least 15 × 15. The cases $n = 6$ and 7 have not yet been fully analyzed for an $n \times n$ board, although draws can always be forced for $n = 8$ and 9. On an $\infty \times \infty$ board, the first player can win for $n = 1$, 2, 3, and 4, but a tie can always be forced for $n \geq 8$. For $3 \times 3 \times 3$ and $4 \times 4 \times 4$, the first player can always win (Gardner 1979).

see also Pong Hau K'i

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 103–104, 1987.

de Fouquières, B. Ch. 18 in *Les Jeux des Anciens, 2nd ed..* Paris, 1873.

Gardner, M. "Mathematical Games: The Diverse Pleasures of Circles that Are Tangent to One Another." *Sci. Amer.* **240**, 18–28, Jan. 1979a.

Gardner, M. "Ticktacktoe Games." Ch. 9 in *Wheels, Life, and Other Mathematical Amusements.* New York: W. H. Freeman, 1983.

Stewart, I. "A Shepherd Takes A Sheep Shot." *Sci. Amer.* **269**, 154–156, 1993.

Ticktacktoe

see Tic-Tac-Toe

Tight Closure

The application of characteristic p methods in Commutative Algebra, which is a synthesis of some areas of Commutative Algebra and Algebraic Geometry.

see also Algebraic Geometry, Commutative Algebra

References

Bruns, W. "Tight Closure." *Bull. Amer. Math. Soc.* **33**, 447–457, 1996.

Huneke, C. "An Algebraist Commuting in Berkeley." *Math. Intell.* **11**, 40–52, 1989.

Tightly Embedded

Q is said to be tightly embedded if $|Q \cap Q^g|$ is Odd for all $g \in G - N_G(Q)$, where $N_G(Q)$ is the Normalizer of Q in G.

Tiling

A plane-filling arrangement of plane figures or its generalization to higher dimensions. Formally, a tiling is a collection of disjoint open sets, the closures of which cover the plane. Given a single tile, the so-called first Corona is the set of all tiles that have a common boundary point with the tile (including the original tile itself).

Wang's Conjecture (1961) stated that if a set of tiles tiled the plane, then they could always be arranged to do so periodically. A periodic tiling of the Plane by Polygons or Space by Polyhedra is called a Tessellation. The conjecture was refuted in 1966 when R. Berger showed that an aperiodic set of 20,426 tiles exists. By 1971, R. Robinson had reduced the number to six and, in 1974, R. Penrose discovered an aperiodic set (when color-matching rules are included) of two tiles: the so-called Penrose Tiles. (Penrose also sued the Kimberly Clark Corporation over their quilted toilet paper, which allegedly resembles a Penrose aperiodic tiling; Mirsky 1997.)

It is not known if there is a single aperiodic tile.

n-gon	tilings
3	any
4	any
5	14
6	3

The number of tilings possible for convex irregular Polygons are given in the above table. Any Triangle or convex Quadrilateral tiles the plane. There are at least 14 classes of convex Pentagonal tilings. There are at least three aperiodic tilings of Hexagons, given by the following types:

$$\begin{array}{cc} A+B+C=360^\circ & a=d \\ A+B+D=360^\circ & a=d, c=e \\ A=C=E & a=b, c=d, e=f \end{array} \tag{1}$$

(Gardner 1988). Note that the periodic hexagonal Tessellation is a degenerate case of all three tilings with

$$A=B=C=D=E=F \qquad a=b=c=d=e=f. \tag{2}$$

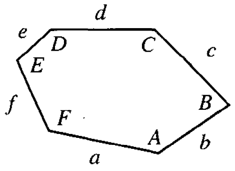

There are no tilings for convex n-gons for $n \geq 7$.

see also ANISOHEDRAL TILING, CORONA (TILING), GOSPER ISLAND, HEESCH'S PROBLEM, ISOHEDRAL TILING, KOCH SNOWFLAKE, MONOHEDRAL TILING, PENROSE TILES, POLYOMINO TILING, SPACE-FILLING POLYHEDRON, TILING THEOREM, TRIANGLE TILING

References

Eppstein, D. "Tiling." `http://www.ics.uci.edu/~eppstein/junkyard/tiling.html`.

Gardner, M. "Tilings with Convex Polygons." Ch. 13 in *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, pp. 162–176, 1988.

Gardner, M. Chs. 1–2 in *Penrose Tiles to Trapdoor Ciphers... and the Return of Dr. Matrix, reissue ed.* Washington, DC: Math. Assoc. Amer.

Grünbaum, B. and Shepard, G. C. "Some Problems on Plane Tilings." In *The Mathematical Gardner* (Ed. D. Klarner). Boston, MA: Prindle, Weber, and Schmidt, pp. 167–196, 1981.

Grünbaum, B. and Sheppard, G. C. *Tilings and Patterns.* New York: W. H. Freeman, 1986.

Lee, X. "Visual Symmetry." `http://www.best.com/~xah/MathGraphicsGallery_dir/Tiling_dir/tiling.html`.

Mirsky, S. "The Emperor's New Toilet Paper." *Sci. Amer.* **277**, 24, July 1997.

Pappas, T. "Mathematics & Moslem Art." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 178, 1989.

Peterson, I. *The Mathematical Tourist: Snapshots of Modern Mathematics.* New York: W. H. Freeman, pp. 82–85, 1988.

Rawles, B. *Sacred Geometry Design Sourcebook: Universal Dimensional Patterns.* Nevada City, CA: Elysian Pub., 1997. `http://www.oro.net/~elysian/bruce_rawles_books.html`.

Schattschneider, D. "In Praise of Amateurs." In *The Mathematical Gardner* (Ed. D. Klarner). Boston, MA: Prindle, Weber, and Schmidt, pp. 140–166, 1981.

Seyd, J. A. and Salman, A. S. *Symmetries of Islamic Geometrical Patterns.* River Edge, NJ: World Scientific, 1995.

Stein, S. and Szabó, S. *Algebra and Tiling.* Washington, DC: Math. Assoc. Amer., 1994.

Tiling Theorem

Due to Lebesgue and Brouwer. If an n-D figure is covered in any way by sufficiently small subregions, then there will exist points which belong to at least $n+1$ of these subareas. Moreover, it is always possible to find a covering by arbitrarily small regions for which no point will belong to more than $n+1$ regions.

see also TESSELLATION, TILING

Times

The operation of MULTIPLICATION, i.e., a times b. Various notations are $a \times b$, $a \cdot b$, ab, and $(a)(b)$. The "multiplication sign" × is based on SAINT ANDREW'S CROSS (Bergamini 1969). Floating point MULTIPLICATION is sometimes denoted $\otimes$.

see also CROSS PRODUCT, DOT PRODUCT, MINUS, MULTIPLICATION, PLUS, PRODUCT

References

Bergamini, D. *Mathematics.* New York: Time-Life Books, p. 11, 1969.

Tit-for-Tat

A strategy for the iterated PRISONER'S DILEMMA in which a prisoner cooperates on the first move, and thereafter copies the previous move of the other prisoner. Any better strategy has more complicated rules.

see also PRISONER'S DILEMMA

References

Goetz, P. "Phil's Good Enough Complexity Dictionary." `http://www.cs.buffalo.edu/~goetz/dict.html`.

Titanic Prime

A PRIME with $\geq$ 1000 DIGITS. As of 1990, there were more than 1400 known (Ribenboim 1990). The table below gives the number of known titanic primes as a function of year end.

Year	Titanic Primes
1992	2254
1993	9166
1994	9779
1995	12391

References

Caldwell, C. "The Ten Largest Known Primes." `http://www.utm.edu/research/primes/largest.html#largest`.

Morain, F. "Elliptic Curves, Primality Proving and Some Titanic Primes." *Astérique* **198–200**, 245–251, 1992.

Ribenboim, P. *The Little Book of Big Primes.* Berlin: Springer-Verlag, p. 97, 1990.

Yates, S. "Titanic Primes." *J. Recr. Math.* **16**, 250–262, 1983-84.

Yates, S. "Sinkers of the Titanics." *J. Recr. Math.* **17**, 268–274, 1984–85.

Titchmarsh Theorem

If $f(\omega)$ is SQUARE INTEGRABLE over the REAL ω axis, then any one of the following implies the other two:

1. The FOURIER TRANSFORM of $f(\omega)$ is 0 for $t < 0$.
2. Replacing ω by z, the function $f(z)$ is analytic in the COMPLEX PLANE z for $y > 0$ and approaches $f(x)$ almost everywhere as $y \to 0$. Furthermore, $\int_{-\infty}^{\infty} |f(x+iy)|^2\,dx < k$ for some number k and $y > 0$ (i.e., the integral is bounded).
3. The REAL and IMAGINARY PARTS of $f(z)$ are HILBERT TRANSFORMS of each other.

Tits Group

A finite SIMPLE GROUP which is a SUBGROUP of the TWISTED CHEVALLEY GROUP $^2F_4(2)$.

Toeplitz Matrix

Given $2N-1$ numbers r_k where $k = -N+1, \ldots, -1, 0, 1, \ldots, N-1$, a MATRIX of the form

$$\begin{bmatrix} r_0 & r_{-1} & r_{-2} & \cdots & r_{-n+1} \\ r_1 & r_0 & r_{-1} & \cdots & r_{-n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{n-1} & r_{n-2} & r_{n-3} & \cdots & r_0 \end{bmatrix}$$

is called a Toeplitz matrix. MATRIX equations of the form

$$\sum_{j=1}^{N} r_{i-j}x_j = y_i$$

can be solved with $\mathcal{O}(N^2)$ operations.

see also VANDERMONDE MATRIX

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Vandermonde Matrices and Toeplitz Matrices." §2.8 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 82–89, 1992.

Togliatti Surface

Togliatti (1940, 1949) showed that QUINTIC SURFACES having 31 ORDINARY DOUBLE POINTS exist, although he did not explicitly derive equations for such surfaces. Beauville (1978) subsequently proved that 31 double points are the maximum possible, and quintic surfaces having 31 ORDINARY DOUBLE POINTS are therefore sometimes called Togliatti surfaces. van Straten (1993) subsequently constructed a 3-D family of solutions and in 1994, Barth derived the example known as the DERVISH.

see also DERVISH, ORDINARY DOUBLE POINT, QUINTIC SURFACE

References

Beauville, A. "Surfaces algébriques complexes." *Astérisque* **54**, 1–172, 1978.

Endraß, S. "Togliatti Surfaces." `http://www.mathematik.uni-mainz.de/Algebraische Geometrie/docs/Etogliatti.shtml`.

Hunt, B. "Algebraic Surfaces." `http://www.mathematik.uni-kl.de/~wwwagag/Galerie.html`.

Togliatti, E. G. "Una notevole superficie de 5° ordine con soli punti doppi isolati." *Vierteljschr. Naturforsch. Ges. Zürich* **85**, 127–132, 1940.

Togliatti, E. "Sulle superficie monoidi col massimo numero di punti doppi." *Ann. Mat. Pura Appl.* **30**, 201–209, 1949.

van Straten, D. "A Quintic Hypersurface in $\mathbb{P}^4$ with 130 Nodes." *Topology* **32**, 857–864, 1993.

Tomography

Tomography is the study of the reconstruction of 2- and 3-dimensional objects from 1-dimensional slices. The RADON TRANSFORM is an important tool in tomography.

Rather surprisingly, there exist certain sets of four directions in Euclidean n-space such that X-rays of a convex body in these directions distinguish it from all other convex bodies.

see also ALEKSANDROV'S UNIQUENESS THEOREM, BRUNN-MINKOWSKI INEQUALITY, BUSEMANN-PETTY PROBLEM, DVORETZKY'S THEOREM, RADON TRANSFORM, STEREOLOGY

References

Gardner, R. J. "Geometric Tomography." *Not. Amer. Math. Soc.* **42**, 422–429, 1995.

Gardner, R. J. *Geometric Tomography.* New York: Cambridge University Press, 1995.

Tooth Surface

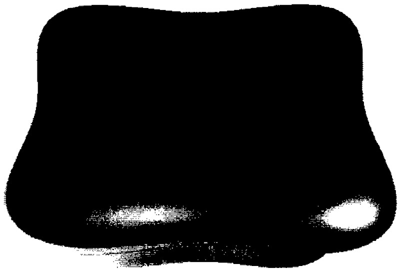

The QUARTIC SURFACE given by the equation

$$x^4 + y^4 + z^4 - (x^2 + y^2 + z^2) = 0.$$

References

Nordstrand, T. "Surfaces." `http://www.uib.no/people/nfytn/surfaces.htm`.

Topological Basis

A topological basis is a SUBSET B of a SET T in which all other OPEN SETS can be written as UNIONS or finite INTERSECTIONS of B. For the REAL NUMBERS, the SET of all OPEN INTERVALS is a basis.

Topological Completion

The topological completion C of a FIELD F with respect to the ABSOLUTE VALUE $|\cdot|$ is the smallest FIELD containing F for which all CAUCHY SEQUENCES or rationals converge.

References

Burger, E. B. and Struppeck, T. "Does $\sum_{n=0}^{\infty} \frac{1}{n!}$ Really Converge? Infinite Series and p-adic Analysis." *Amer. Math. Monthly* **103**, 565–577, 1996.

Topologically Conjugate

Two MAPS $\phi, \psi : M \to M$ are said to be topologically conjugate if there EXISTS a HOMEOMORPHISM $h : M \to M$ such that $\phi \circ h = h \circ \psi$, i.e., h maps ψ-orbits onto ϕ-orbits. Two maps which are topologically conjugate cannot be distinguished topologically.

see also ANOSOV DIFFEOMORPHISM, STRUCTURALLY STABLE

Topological Dimension

see LEBESGUE COVERING DIMENSION

Topological Entropy

The topological entropy of a MAP M is defined as

$$h_T(M) = \sup_{\{W_i\}} h(M, \{W_i\}),$$

where $\{W_i\}$ is a partition of a bounded region W containing a probability measure which is invariant under M, and sup is the SUPREMUM.

References

Ott, E. *Chaos in Dynamical Systems.* New York: Cambridge University Press, pp. 143–144, 1993.

Topological Groupoid

A topological groupoid over B is a GROUPOID G such that B and G are TOPOLOGICAL SPACES and α, β, and multiplication are continuous maps. Here, α and β are maps from G onto $\mathbb{R}^2$ with $\alpha : (x, \gamma, y) \mapsto x$ and $\beta : (x, \gamma, y) \mapsto y$.

References
Weinstein, A. "Groupoids: Unifying Internal and External Symmetry." *Not. Amer. Math. Soc.* **43**, 744–752, 1996.

Topological Manifold

A TOPOLOGICAL SPACE M satisfying some separability (i.e., it is a HAUSDORFF SPACE) and countability (i.e., it is a PARACOMPACT SPACE) conditions such that every point $p \in M$ has a NEIGHBORHOOD homeomorphic to an OPEN SET in $\mathbb{R}^n$ for some $n \geq 0$. Every SMOOTH MANIFOLD is a topological manifold, but not necessarily vice versa. The first nonsmooth topological manifold occurs in 4-D.

Nonparacompact manifolds are of little use in mathematics, but non-Hausdorff manifolds do occasionally arise in research (Hawking and Ellis 1975). For manifolds, Hausdorff and second countable are equivalent to Hausdorff and paracompact, and both are equivalent to the manifold being embeddable in some large-dimensional Euclidean space.

see also HAUSDORFF SPACE, MANIFOLD, PARACOMPACT SPACE, SMOOTH MANIFOLD, TOPOLOGICAL SPACE

References
Hawking, S. W. and Ellis, G. F. R. *The Large Scale Structure of Space-Time.* New York: Cambridge University Press, 1975.

Topological Space

A SET X for which a TOPOLOGY T has been specified is called a topological space (Munkres 1975, p. 76).

see also KURATOWSKI'S CLOSURE-COMPONENT PROBLEM, OPEN SET, TOPOLOGICAL VECTOR SPACE

References
Berge, C. *Topological Spaces Including a Treatment of Multi-Valued Functions, Vector Spaces and Convexity.* New York: Dover, 1997.
Munkres, J. R. *Topology: A First Course.* Englewood Cliffs, NJ: Prentice-Hall, 1975.

Topological Vector Space

A TOPOLOGICAL SPACE such that the two algebraic operations of VECTOR SPACE are continuous in the topology.

References
Köthe, G. *Topological Vector Spaces.* New York: Springer-Verlag, 1979.

Topologically Transitive

A FUNCTION f is topologically transitive if, given any two intervals U and V, there is some POSITIVE INTEGER k such that $f^k(U) \cap V = \varnothing$. Vaguely, this means that neighborhoods of points eventually get flung out to "big" sets so that they don't necessarily stick together in one localized clump.

see also CHAOS

Topology

Topology is the mathematical study of properties of objects which are preserved through deformations, twistings, and stretchings. (Tearing, however, is not allowed.) A CIRCLE is topologically equivalent to an ELLIPSE (into which it can be deformed by stretching) and a SPHERE is equivalent to an ELLIPSOID. Continuing along these lines, the SPACE of all positions of the minute hand on a clock is topologically equivalent to a CIRCLE (where SPACE of all positions means "the collection of all positions"). Similarly, the SPACE of all positions of the minute and hour hands is equivalent to a TORUS. The SPACE of all positions of the hour, minute and second hands form a 4-D object that cannot be visualized quite as simply as the former objects since it cannot be placed in our 3-D world, although it can be visualized by other means.

There is more to topology, though. Topology began with the study of curves, surfaces, and other objects in the plane and 3-space. One of the central ideas in topology is that spatial objects like CIRCLES and SPHERES can be treated as objects in their own right, and knowledge of objects is independent of how they are "represented" or "embedded" in space. For example, the statement "if you remove a point from a CIRCLE, you get a line segment" applies just as well to the CIRCLE as to an ELLIPSE, and even to tangled or knotted CIRCLES, since the statement involves only topological properties.

Topology has to do with the study of spatial objects such as curves, surfaces, the space we call our universe, the space-time of general relativity, fractals, knots, manifolds (objects with some of the same basic spatial properties as our universe), phase spaces that are encountered in physics (such as the space of hand-positions of a clock), symmetry groups like the collection of ways of rotating a top, etc.

The "objects" of topology are often formally defined as TOPOLOGICAL SPACES. If two objects have the same topological properties, they are said to be HOMEOMORPHIC (although, strictly speaking, properties that are not destroyed by stretching and distorting an object are really properties preserved by ISOTOPY, not HOMEOMORPHISM; ISOTOPY has to do with distorting embedded objects, while HOMEOMORPHISM is intrinsic).

Topology is divided into ALGEBRAIC TOPOLOGY (also called COMBINATORIAL TOPOLOGY), DIFFERENTIAL TOPOLOGY, and LOW-DIMENSIONAL TOPOLOGY.

There is also a formal definition for a topology defined in terms of set operations. A SET X along with a collection T of SUBSETS of it is said to be a topology if the SUBSETS in T obey the following properties:

1. The (trivial) subsets X and the EMPTY SET $\varnothing$ are in T.
2. Whenever sets A and B are in T, then so is $A \cap B$.
3. Whenever two or more sets are in T, then so is their UNION

(Bishop and Goldberg 1980).

A SET X for which a topology T has been specified is called a TOPOLOGICAL SPACE (Munkres 1975, p. 76). For example, the SET $X = \{0, 1, 2, 3\}$ together with the SUBSETS $T = \{0\}, \{1, 2, 3\}, \varnothing, \{0, 1, 2, 3\}\}$ comprises a topology, and X is a TOPOLOGICAL SPACE.

Topologies can be built up from TOPOLOGICAL BASES. For the REAL NUMBERS, the topology is the UNION of OPEN INTERVALS.

see also ALGEBRAIC TOPOLOGY, DIFFERENTIAL TOPOLOGY, GENUS, KLEIN BOTTLE, KURATOWSKI REDUCTION THEOREM, LEFSHETZ TRACE FORMULA, LOW-DIMENSIONAL TOPOLOGY, POINT-SET TOPOLOGY, ZARISKI TOPOLOGY

References

Adamson, I. *A General Topology Workbook.* Boston, MA: Birkhäuser, 1996.

Armstrong, M. A. *Basic Topology, rev.* New York: Springer-Verlag, 1997.

Barr, S. *Experiments in Topology.* New York: Dover, 1964.

Berge, C. *Topological Spaces Including a Treatment of Multi-Valued Functions, Vector Spaces and Convexity.* New York: Dover, 1997.

Bishop, R. and Goldberg, S. *Tensor Analysis on Manifolds.* New York: Dover, 1980.

Blackett, D. W. *Elementary Topology: A Combinatorial and Algebraic Approach.* New York: Academic Press, 1967.

Bloch, E. *A First Course in Geometric Topology and Differential Geometry.* Boston, MA: Birkhäuser, 1996.

Chinn, W. G. and Steenrod, N. E. *First Concepts of Topology: The Geometry of Mappings of Segments, Curves, Circles, and Disks.* Washington, DC: Math. Assoc. Amer., 1966.

Eppstein, D. "Geometric Topology." `http://www.ics.uci.edu/~eppstein/junkyard/topo.html`.

Francis, G. K. *A Topological Picturebook.* New York: Springer-Verlag, 1987.

Gemignani, M. C. *Elementary Topology.* New York: Dover, 1990.

Greever, J. *Theory and Examples of Point-Set Topology.* Belmont, CA: Brooks/Cole, 1967.

Hirsch, M. W. *Differential Topology.* New York: Springer-Verlag, 1988.

Hocking, J. G. and Young, G. S. *Topology.* New York: Dover, 1988.

Kahn, D. W. *Topology: An Introduction to the Point-Set and Algebraic Areas.* New York: Dover, 1995.

Kelley, J. L. *General Topology.* New York: Springer-Verlag, 1975.

Kinsey, L. C. *Topology of Surfaces.* New York: Springer-Verlag, 1993.

Lipschutz, S. *Theory and Problems of General Topology.* New York: Schaum, 1965.

Mendelson, B. *Introduction to Topology.* New York: Dover, 1990.

Munkres, J. R. *Elementary Differential Topology.* Princeton, NJ: Princeton University Press, 1963.

Munkres, J. R. *Topology: A First Course.* Englewood Cliffs, NJ: Prentice-Hall, 1975.

Praslov, V. V. and Sossinsky, A. B. *Knots, Links, Braids and 3-Manifolds: An Introduction to the New Invariants in Low-Dimensional Topology.* Providence, RI: Amer. Math. Soc., 1996.

Shakhmatv, D. and Watson, S. "Topology Atlas." `http://www.unipissing.ca/topology/`.

Steen, L. A. and Seebach, J. A. Jr. *Counterexamples in Topology.* New York: Dover, 1996.

Thurston, W. P. *Three-Dimensional Geometry and Topology, Vol. 1.* Princeton, NJ: Princeton University Press, 1997.

van Mill, J. and Reed, G. M. (Eds.). *Open Problems in Topology.* New York: Elsevier, 1990.

Veblen, O. *Analysis Situs, 2nd ed.* New York: Amer. Math. Soc., 1946.

Topos

A CATEGORY modeled after the properties of the CATEGORY of sets.

see also CATEGORY, LOGOS

References

Freyd, P. J. and Scedrov, A. *Categories, Allegories.* Amsterdam, Netherlands: North-Holland, 1990.

McLarty, C. *Elementary Categories, Elementary Toposes.* New York: Oxford University Press, 1992.

Toric Variety

Let $m_1, m_2, \ldots, m_n$ be distinct primitive elements of a 2-D LATTICE M such that $\det(m_i, m_{i+1}) > 0$ for $i = 1, \ldots, n$. Each collection $\Gamma = \{m_1, m_2, \ldots, m_n\}$ then forms a set of rays of a unique complete fan in M, and therefore determines a 2-D toric variety X_Γ.

References

Danilov, V. I. "The Geometry of Toric Varieties." *Russ. Math. Surv.* **33**, 97–154, 1978.

Fulton, W. *Introduction to Toric Varieties.* Princeton, NJ: Princeton University Press, 1993.

Morelli, R. "Pick's Theorem and the Todd Class of a Toric Variety." *Adv. Math.* **100**, 183–231, 1993.

Oda, T. *Convex Bodies and Algebraic Geometry.* New York: Springer-Verlag, 1987.

Pommersheim, J. E. "Toric Varieties, Lattice Points, and Dedekind Sums." *Math. Ann.* **295**, 1–24, 1993.

Torispherical Dome

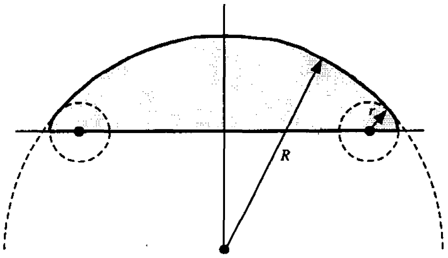

A torispherical dome is the surface obtained from the intersection of a SPHERICAL CAP with a tangent TORUS, as illustrated above. The radius of the sphere R is called the "crown radius," and the radius of the torus is called the "knuckle radius." Torispherical domes are used to construct pressure vessels.

see also DOME, SPHERICAL CAP

Torn Square Fractal

see CESÀRO FRACTAL

Toroid

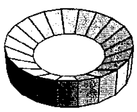

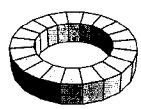

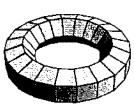

A SURFACE OF REVOLUTION obtained by rotating a closed PLANE CURVE about an axis parallel to the plane which does not intersect the curve. The simplest toroid is the TORUS.

see also PAPPUS'S CENTROID THEOREM, SURFACE OF REVOLUTION, TORUS

Toroidal Coordinates

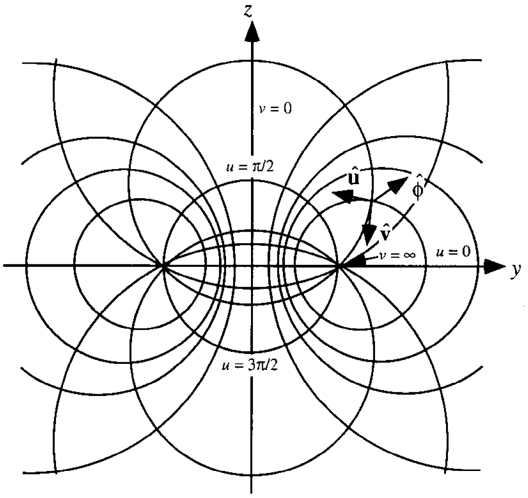

A system of CURVILINEAR COORDINATES for which several different notations are commonly used. In this work $(u,\ v,\ \phi)$ is used, whereas Arfken (1970) uses $(\xi,\ \eta,\ \varphi)$. The toroidal coordinates are defined by

$$x = \frac{a\sinh v\cos\phi}{\cosh v - \cos u} \tag{1}$$

$$y = \frac{a\sinh v\sin\phi}{\cosh v - \cos u} \tag{2}$$

$$z = \frac{a\sin u}{\cosh v - \cos u}, \tag{3}$$

where $\sinh z$ is the HYPERBOLIC SINE and $\cosh z$ is the HYPERBOLIC COSINE. The SCALE FACTORS are

$$h_u = \frac{a}{\cosh v - \cos u} \tag{4}$$

$$h_v = \frac{a}{\cosh v - \cos u} \tag{5}$$

$$h_\phi = \frac{a\sinh v}{\cosh v - \cos u}. \tag{6}$$

The LAPLACIAN is

$$\begin{aligned}\nabla^2 f = {} & \frac{(\cosh v - \cos u)^3}{a^2}\frac{\partial}{\partial u}\left(\frac{1}{\cosh v - \cos u}\frac{\partial f}{\partial u}\right) \\ & + \frac{(\cosh v - \cos u)^3}{a^2\sinh v}\frac{\partial}{\partial v}\left(\frac{\sinh v}{\cosh v - \cos u}\frac{\partial f}{\partial v}\right) \\ & + \frac{(\cosh v - \cos u)^2}{a^2\sinh v}\frac{\partial^2 f}{\partial\phi^2} \qquad (7) \\ = {} & \left(\frac{-3\cos\coth^2 v + \cosh v\coth^2 v}{\cosh v - \cos u}\right. \\ & \left. + \frac{+3\cos^2 u\coth v\operatorname{csch} v - \cos^3 u\operatorname{csch}^2 v}{\cosh v - \cos u}\right)\frac{\partial^2}{\partial\phi^2} \\ & + (\cos u - \cosh v)\sin u\frac{\partial}{\partial u} + (\cosh v - \cos u)^2\frac{\partial^2}{\partial u^2} \\ & + (\cosh v - \cos u)(\cosh v\coth v - \sinh v \\ & - \cos u\coth v)\frac{\partial}{\partial v} + (\cosh^2 v - \cos u)^2\frac{\partial^2}{\partial v^2}. \qquad (8)\end{aligned}$$

The HELMHOLTZ DIFFERENTIAL EQUATION is not separable in toroidal coordinates, but LAPLACE'S EQUATION is.

see also BISPHERICAL COORDINATES, LAPLACE'S EQUATION—TOROIDAL COORDINATES

References

Arfken, G. "Toroidal Coordinates $(\xi,\ \eta,\ \phi)$." §2.13 in *Mathematical Methods for Physicists, 2nd ed.* Orlando, FL: Academic Press, pp. 112–115, 1970.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, p. 666, 1953.

Toroidal Field

A VECTOR FIELD resembling a TORUS which is purely circular about the z-AXIS of a SPHERE (i.e., follows lines of LATITUDE). A toroidal field takes the form

$$\mathbf{T} = \begin{bmatrix} 0 \\ \frac{1}{\sin\theta}\frac{\partial T}{\partial\phi} \\ -\frac{\partial T}{\partial\theta} \end{bmatrix}.$$

see also DIVERGENCELESS FIELD, POLOIDAL FIELD

References

Stacey, F. D. *Physics of the Earth, 2nd ed.* New York: Wiley, p. 239, 1977.

Toroidal Function

A class of functions also called Ring Functions which appear in systems having toroidal symmetry. Toroidal functions can be expressed in terms of the Legendre Functions and Second Kinds (Abramowitz and Stegun 1972, p. 336):

$$P^{\mu}_{\nu-1/2}(\cosh\eta) = [\Gamma(1-\mu)]^{-1}2^{2\mu}(1-e^{-2\eta})^{-\mu}e^{-(\nu+1/2)\eta}$$
$$\times {}_2F_1(\tfrac{1}{2}-\mu, \tfrac{1}{2}+\nu-\mu; 1-2\mu; 1-e^{-2\eta})$$
$$P^{m}_{n-1/2}(\cosh\eta) = \frac{\Gamma(n+m+\frac{1}{2})(\sinh\eta)^m}{\Gamma(n-m+\frac{1}{2})2^m\sqrt{\pi}\,\Gamma(m+\frac{1}{2})}$$
$$\times \int_0^\pi \frac{sin^{2m}\phi\,d\phi}{(\cosh\eta+\cos\phi\sinh\eta)^{n+m+1/2}}$$
$$Q^{\mu}_{\nu-1/2}(\cosh\eta) = [\Gamma(1+\nu)]^{-1}\sqrt{\pi}\,e^{i\mu\pi}\Gamma(\tfrac{1}{2}+\nu+\mu)$$
$$\times (1-e^{-2\eta})^{\mu}e^{-(\nu+1/2)\eta}{}_2F_1(\tfrac{1}{2}-\mu, \tfrac{1}{2}+\nu+\mu; 1+\mu; 1-e^{-2\eta})$$
$$Q^{m}_{n-1/2}(\cosh\eta) = \frac{(-1)^m\Gamma(n+\frac{1}{2})}{\Gamma(n-m+\frac{1}{2})}$$
$$\times \int_0^\infty \frac{\cosh(mt)\,dt}{(\cosh\eta+\cosh t\sinh\eta)^{n+1/2}}$$

for $n > m$. Byerly (1959) identifies

$$\frac{1}{i^{n/2}}P^n_m(\coth x) = \operatorname{csch}^n x\frac{d^nP_m(\coth x)}{d(\coth x)^n}$$

as a Toroidal Harmonic.

see also Conical Function

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Toroidal Functions (or Ring Functions)." §8.11 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 336, 1972.

Byerly, W. E. *An Elementary Treatise on Fourier's Series, and Spherical, Cylindrical, and Ellipsoidal Harmonics, with Applications to Problems in Mathematical Physics.* New York: Dover, p. 266, 1959.

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1468, 1980.

Toroidal Harmonic

see Toroidal Function

Toroidal Polyhedron

A toroidal polyhedron is a Polyhedron with Genus $g \geq 1$ (i.e., having one or more Holes). Examples of toroidal polyhedra include the Császár Polyhedron and Szilassi Polyhedron, both of which have Genus 1 (i.e., the Topology of a Torus).

The only known Toroidal Polyhedron with no Diagonals is the Császár Polyhedron. If another exists, it must have 12 or more Vertices and Genus $g \geq 6$. The smallest known single-hole toroidal Polyhedron made up of only Equilateral Triangles is composed of 48 of them.

see also Császár Polyhedron, Szilassi Polyhedron

References

Gardner, M. *Time Travel and Other Mathematical Bewilderments.* New York: W. H. Freeman, p. 141, 1988.

Hart, G. "Toroidal Polyhedra." `http://www.li.net/~george/virtual-polyhedra/toroidal.html`.

Stewart, B. M. *Adventures Among the Toroids, 2nd rev. ed.* Okemos, MI: B. M. Stewart, 1984.

Toronto Function

$$T(m,n,r) \equiv \frac{\Gamma(\frac{1}{2}m+\frac{1}{2})}{n!r^{-2n+m-1}}\,{}_1F_1(\tfrac{1}{2}; i+n; r^2),$$

where ${}_1F_1(a;b;z)$ is a Confluent Hypergeometric Function and $\Gamma(z)$ is the Gamma Function (Abramowitz and Stegun 1972).

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 509, 1972.

Torricelli Point

see Fermat Point

Torsion (Differential Geometry)

The rate of change of the Osculating Plane of a Space Curve. The torsion τ is Positive for a right-handed curve, and Negative for a left-handed curve. A curve with Curvature $\kappa \neq 0$ is planar Iff $\tau = 0$.

The torsion can be defined by

$$\tau \equiv -\mathbf{N}\cdot\mathbf{B}',$$

where $\mathbf{N}$ is the unit Normal Vector and $\mathbf{B}$ is the unit Binormal Vector. Written explicitly in terms of a parameterized Vector Function $\mathbf{x}$,

$$\tau = \frac{|\dot{\mathbf{x}}\,\ddot{\mathbf{x}}\,\dddot{\mathbf{x}}|}{\ddot{\mathbf{x}}\cdot\ddot{\mathbf{x}}} = \rho^2|\dot{\mathbf{x}}\,\ddot{\mathbf{x}}\,\dddot{\mathbf{x}}|,$$

where $|\mathbf{a}\,\mathbf{b}\,\mathbf{c}|$ denotes a Scalar Triple Product and ρ is the Radius of Curvature. The quantity $1/\tau$ is called the Radius of Torsion and is denoted σ or ϕ.

see also Curvature, Radius of Curvature, Radius of Torsion

References

Gray, A. "Drawing Space Curves with Assigned Curvature." §7.8 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 145–147, 1993.

Kreyszig, E. "Torsion." §14 in *Differential Geometry.* New York: Dover, pp. 37–40, 1991.

Torsion (Group Theory)

If G is a GROUP, then the torsion elements Tor(G) of G (also called the torsion of G) are defined to be the set of elements g in G such that $g^n = e$ for some NATURAL NUMBER n, where e is the IDENTITY ELEMENT of the GROUP G.

In the case that G is ABELIAN, Tor(G) is a SUBGROUP and is called the torsion subgroup of G. If Tor(G) consists only of the IDENTITY ELEMENT, the GROUP G is called torsion-free.

see also ABELIAN GROUP, GROUP, IDENTITY ELEMENT

Torsion Number

One of a set of numbers defined in terms of an invariant generated by the finite cyclic covering spaces of a KNOT complement. The torsion numbers for KNOTS up to 9 crossings were cataloged by Reidemeister (1948).

References

Reidemeister, K. *Knotentheorie.* New York: Chelsea, 1948.

Rolfsen, D. "Torsion Numbers." §6A in *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 145–146, 1976.

Torsion Tensor

The TENSOR defined by

$$T^l{}_{jk} \equiv -(\Gamma^l{}_{jk} - \Gamma^l{}_{kj}),$$

where $\Gamma^l{}_{jk}$ are CONNECTION COEFFICIENTS.

see also CONNECTION COEFFICIENT

Torus

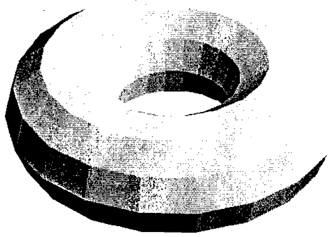

A torus is a surface having GENUS 1, and therefore possessing a single "HOLE." The usual torus in 3-D space is shaped like a donut, but the concept of the torus is extremely useful in higher dimensional space as well. One of the more common uses of n-D tori is in DYNAMICAL SYSTEMS. A fundamental result states that the PHASE SPACE trajectories of a HAMILTONIAN SYSTEM with n DEGREES OF FREEDOM and possessing n INTEGRALS OF MOTION lie on an n-D MANIFOLD which is topologically equivalent to an n-torus (Tabor 1989).

The usual 3-D "ring" torus is known in older literature as an "ANCHOR RING." Let the radius from the center of the hole to the center of the torus tube be c, and the radius of the tube be a. Then the equation in CARTESIAN COORDINATES is

$$(c - \sqrt{x^2 + y^2}\,)^2 + z^2 = a^2. \tag{1}$$

The parametric equations of a torus are

$$x = (c + a\cos v)\cos u \tag{2}$$

$$y = (c + a\cos v)\sin u \tag{3}$$

$$z = a\sin v \tag{4}$$

for $u, v \in [0, 2\pi)$. Three types of torus, known as the STANDARD TORI, are possible, depending on the relative sizes of a and c. $c > a$ corresponds to the RING TORUS (shown above), $c = a$ corresponds to a HORN TORUS which is tangent to itself at the point (0, 0, 0), and $c < a$ corresponds to a self-intersecting SPINDLE TORUS (Pinkall 1986).

If no specification is made, "torus" is taken to mean RING TORUS. The three STANDARD TORI are illustrated below, where the first image shows the full torus, the second a cut-away of the bottom half, and the third a CROSS-SECTION of a plane passing through the z-AXIS.

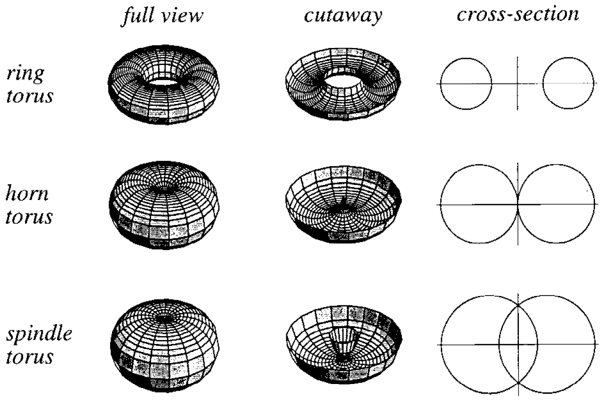

The STANDARD TORI and their inversions are CYCLIDES. If the coefficient of $\sin v$ in the formula for z is changed to $b \neq a$, an ELLIPTIC TORUS results.

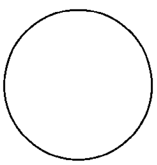

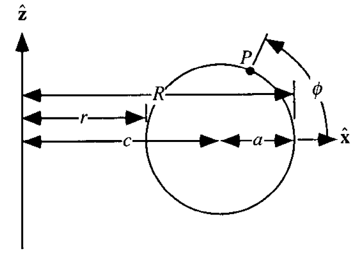

To compute the metric properties of the ring torus, define the inner and outer radii by

$$r \equiv c - a \tag{5}$$

$$R \equiv c + a. \tag{6}$$

Solving for a and c gives

$$a = \tfrac{1}{2}(R - r) \tag{7}$$

$$c = \tfrac{1}{2}(R + r). \tag{8}$$

Then the SURFACE AREA of this torus is

$$S = (2\pi a)(2\pi c) = 4\pi^2 ac \tag{9}$$

$$= \pi^2 (R + r)(R - r), \tag{10}$$

and the VOLUME can be computed from PAPPUS'S CENTROID THEOREM

$$V = (\pi a^2)^2 \pi c = 2\pi^2 a^2 c \tag{11}$$

$$= \tfrac{1}{4}\pi^2 (R + r)(R - r)^2. \tag{12}$$

The coefficients of the first and second FUNDAMENTAL FORMS of the torus are given by

$$e = -(c + a\cos v)\cos v \tag{13}$$

$$f = 0 \tag{14}$$

$$g = -a \tag{15}$$

$$E = (c + a\cos v)^2 \tag{16}$$

$$F = 0 \tag{17}$$

$$G = a^2, \tag{18}$$

giving RIEMANNIAN METRIC

$$ds^2 = (c + a\cos v)^2\, du^2 + a^2\, dv^2, \tag{19}$$

AREA ELEMENT

$$dA = a(c + a\cos v)\, du \wedge dv \tag{20}$$

(where $du \wedge dv$ is a WEDGE PRODUCT), and GAUSSIAN and MEAN CURVATURES as

$$K = \frac{\cos v}{a(c + a\cos v)} \tag{21}$$

$$H = -\frac{c + 2a\cos v}{2a(c + a\cos v)} \tag{22}$$

(Gray 1993, pp. 289–291).

A torus with a HOLE in *its surface* can be turned inside out to yield an identical torus. A torus can be knotted externally or internally, but not both. These two cases are AMBIENT ISOTOPIES, but not REGULAR ISOTOPIES. There are therefore three possible ways of embedding a torus with zero or one KNOT.

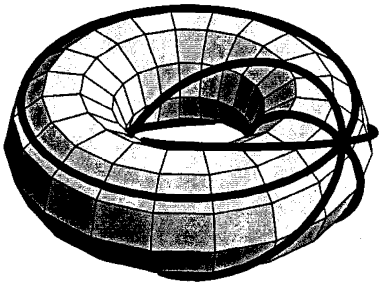

An arbitrary point P on a torus (not lying in the xy-plane) can have four CIRCLES drawn through it. The first circle is in the plane of the torus and the second is PERPENDICULAR to it. The third and fourth CIRCLES are called VILLARCEAU CIRCLES (Villarceau 1848, Schmidt 1950, Coxeter 1969, Melnick 1983).

To see that two additional CIRCLES exist, consider a coordinate system with origin at the center of torus, with $\hat{\mathbf{z}}$ pointing up. Specify the position of P by its ANGLE ϕ measured around the tube of the torus. Define $\phi = 0$ for the circle of points farthest away from the center of the torus (i.e., the points with $x^2 + y^2 = R^2$), and draw the x-AXIS as the intersection of a plane through the z-axis and passing through P with the xy-plane. Rotate about the y-AXIS by an ANGLE θ, where

$$\theta = \sin^{-1}\left(\frac{a}{c}\right). \tag{23}$$

In terms of the old coordinates, the new coordinates are

$$x = x_1\cos\theta - z_1\sin\theta \tag{24}$$

$$z = x_1\sin\theta + z_1\cos\theta. \tag{25}$$

So in (x_1, y_1, z_1) coordinates, equation (1) of the torus becomes

$$[\sqrt{(x_1\cos\theta - z_1\sin\theta)^2 + y_1{}^2} - c]^2 + (x_1\sin\theta + z_1\cos\theta)^2 = a^2. \tag{26}$$

Squaring both sides gives

$$(x_1\cos\theta - z_1\sin\theta)^2 + y_1{}^2 + c^2 - 2c\sqrt{(x_1\cos\theta - z_1\sin\theta)^2 + y_1{}^2} + (x_1\sin\theta + z_1\cos\theta)^2 = a^2. \tag{27}$$

But

$$(x_1\cos\theta - z_1\sin\theta)^2 + (x_1\sin\theta + z_1\cos\theta)^2 = x_1{}^2 + z_1{}^2, \tag{28}$$

so

$$x_1{}^2 + y_1{}^2 + z_1{}^2 + c^2 - 2c\sqrt{(x_1\cos\theta - z_1\sin\theta)^2 + y_1{}^2} = a^2. \tag{29}$$

In the $z_1 = 0$ plane, plugging in (23) and factoring gives

$$[x_1{}^2 + (y_1 - a)^2 - c^2][x_1{}^2 + (y_1 + a) - c^2] = 0. \tag{30}$$

This gives the CIRCLES

$$x_1{}^2 + (y_1 - a)^2 = c^2 \tag{31}$$

and

$$x_1{}^2 + (y_1 + a)^2 = c^2 \tag{32}$$

in the z_1 plane. Written in MATRIX form with parameter $t \in [0, 2\pi)$, these are

$$C_1 = \begin{bmatrix} c\cos t \\ c\sin t + a \\ 0 \end{bmatrix} \tag{33}$$

$$C_2 = \begin{bmatrix} c\cos t \\ c\sin t - a \\ 0 \end{bmatrix}. \tag{34}$$

In the original (x, y, z) coordinates,

$$\begin{aligned} C_1 &= \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} c\cos t \\ c\sin t + a \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} c\cos\theta\cos t \\ c\sin t + a \\ -c\sin\theta\cos t \end{bmatrix} \end{aligned} \tag{35}$$

$$\begin{aligned} C_2 &= \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} c\cos t \\ c\sin t - a \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} c\cos\theta\cos t \\ c\sin t - a \\ -c\sin\theta\cos t \end{bmatrix}. \end{aligned} \tag{36}$$

The point P must satisfy

$$z = a\sin\phi = c\sin\theta\cos t, \tag{37}$$

so

$$\cos t = \frac{a\sin\phi}{c\sin\theta}. \tag{38}$$

Plugging this in for x_1 and y_1 gives the ANGLE ψ by which the CIRCLE must be rotated about the z-AXIS in order to make it pass through P,

$$\psi = \tan^{-1}\left(\frac{y}{x}\right) = \frac{c\sin t + a}{c\cos\theta\cos t} = \frac{c\sqrt{1-\cos^2 t} + a}{c\cos\theta\cos t}. \tag{39}$$

The four CIRCLES passing through P are therefore

$$C_1 = \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c\cos\theta\cos t \\ c\sin t + a \\ -c\sin\theta\cos t \end{bmatrix} \tag{40}$$

$$C_2 = \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c\cos\theta\cos t \\ c\sin t - a \\ -c\sin\theta\cos t \end{bmatrix} \tag{41}$$

$$C_3 = \begin{bmatrix} (c + a\cos\phi)\cos t \\ (c + a\cos\phi)\sin t \\ a\sin\phi \end{bmatrix} \tag{42}$$

$$C_4 = \begin{bmatrix} c + a\cos t \\ 0 \\ a\sin t \end{bmatrix}. \tag{43}$$

see also APPLE, CYCLIDE, ELLIPTIC TORUS, GENUS (SURFACE), HORN TORUS, KLEIN QUARTIC, LEMON, RING TORUS, SPINDLE TORUS, SPIRIC SECTION, STANDARD TORI, TOROID, TORUS COLORING, TORUS CUTTING

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 131–132, 1987.

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, pp. 132–133, 1969.

Geometry Center. "The Torus." `http://www.geom.umn.edu/zoo/toptype/torus/`.

Gray, A. "Tori." §11.4 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 218–220 and 289–290, 1993.

Melzak, Z. A. *Invitation to Geometry.* New York: Wiley, pp. 63–72, 1983.

Pinkall, U. "Cyclides of Dupin." §3.3 in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 28–30, 1986.

Schmidt, H. *Die Inversion und ihre Anwendungen.* Munich: Oldenbourg, p. 82, 1950.

Tabor, M. *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, pp. 71–74, 1989.

Villarceau, M. "Théorème sur le tore." *Nouv. Ann. Math.* **7**, 345–347, 1848.

Torus Coloring

The number of colors SUFFICIENT for MAP COLORING on a surface of GENUS g is given by the HEAWOOD CONJECTURE,

$$\chi(g) = \left\lfloor \tfrac{1}{2}\left(7 + \sqrt{48g + 1}\right) \right\rfloor,$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION. The fact that $\chi(g)$ (which is called the CHROMATIC NUMBER) is also NECESSARY was proved by Ringel and Youngs (1968) with two exceptions: the SPHERE (which requires the same number of colors as the PLANE) and the KLEIN BOTTLE. A g-holed TORUS therefore requires $\chi(g)$ colors. For $g = 0, 1, \ldots$, the first few values of $\chi(g)$ are 4, 7, 8, 9, 10, 11, 12, 12, 13, 13, 14, 15, 15, 16, ... (Sloane's A000934).

see also CHROMATIC NUMBER, FOUR-COLOR THEOREM, HEAWOOD CONJECTURE, KLEIN BOTTLE, MAP COLORING

References

Gardner, M. "Mathematical Games: The Celebrated Four-Color Map Problem of Topology." *Sci. Amer.* **203**, 218–222, Sep. 1960.

Ringel, G. *Map Color Theorem.* New York: Springer-Verlag, 1974.

Ringel, G. and Youngs, J. W. T. "Solution of the Heawood Map-Coloring Problem." *Proc. Nat. Acad. Sci. USA* **60**, 438–445, 1968.

Sloane, N. J. A. Sequence A000934/M3292 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Wagon, S. "Map Coloring on a Torus." §7.5 in *Mathematica in Action.* New York: W. H. Freeman, pp. 232–237, 1991.

Torus Cutting

With n cuts of a Torus of Genus 1, the maximum number of pieces which can be obtained is

$$N(n) = \tfrac{1}{6}(n^3 + 3n^2 + 8n).$$

The first few terms are 2, 6, 13, 24, 40, 62, 91, 128, 174, 230, ... (Sloane's A003600).

see also Cake Cutting, Circle Cutting, Cylinder Cutting, Pancake Cutting, Plane Cutting, Pie Cutting, Square Cutting

References

Gardner, M. *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, pp. 149–150, 1978.

Sloane, N. J. A. Sequence A003600/M1594 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Torus Knot

A (p,q)-torus Knot is obtained by looping a string through the Hole of a Torus p times with q revolutions before joining its ends, where p and q are Relatively Prime. A (p,q)-torus knot is equivalent to a (q,p)-torus knot. The Crossing Number of a (p,q)-torus knot is

$$c = \min\{p(q-1), q(p-1)\} \tag{1}$$

(Murasugi 1991). The Unknotting Number of a (p,q)-torus knot is

$$u = \tfrac{1}{2}(p-1)(q-1) \tag{2}$$

(Adams 1991).

Torus knots with fewer than 11 crossings are the Trefoil Knot 03_{001} (3, 2), Solomon's Seal Knot 05_{001} (5, 2), 07_{001} (7, 2), 08_{019} (4, 3), 09_{001} (9, 2), and 10_{124} (5, 3) (Adams *et al.* 1991). The only Knots which are not Hyperbolic Knots are torus knots and Satellite Knots (including Composite Knots). The $(2,q)$, $(3,4)$, and $(3,5)$-torus knots are Almost Alternating Knots.

The Jones Polynomial of an (m,n)-Torus Knot is

$$\frac{t^{(m-1)(n-1)/2}(1-t^{m+1}-t^{n+1}+t^{m+n})}{1-t^2}. \tag{3}$$

The Bracket Polynomial for the torus knot $K_n = (2,n)$ is given by the Recurrence Relation

$$\langle K_n \rangle = A\langle K_{n-1}\rangle + (-1)^{n-1}A^{-3n+2}, \tag{4}$$

where

$$\langle K_1 \rangle = -A^3. \tag{5}$$

see also Almost Alternating Knot, Hyperbolic Knot, Knot, Satellite Knot, Solomon's Seal Knot, Trefoil Knot

References

Adams, C.; Hildebrand, M.; and Weeks, J. "Hyperbolic Invariants of Knots and Links." *Trans. Amer. Math. Soc.* **326**, 1–56, 1991.

Gray, A. "Torus Knots." §8.2 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 155–161, 1993.

Murasugi, K. "On the Braid Index of Alternating Links." *Trans. Amer. Math. Soc.* **326**, 237–260, 1991.

Total Angular Defect

see Descartes Total Angular Defect

Total Curvature

The total curvature of a curve is the quantity $\sqrt{\tau^2+\kappa^2}$, where τ is the Torsion and κ is the Curvature. The total curvature is also called the Third Curvature.

see also Curvature, Torsion (Differential Geometry)

Total Differential

see Exact Differential

Total Function

A Function defined for all possible input values.

Total Intersection Theorem

If one part of the total intersection group of a curve of order n with a curve of order $n_1 + n_2$ constitutes the total intersection with a curve of order n_1, then the other part will constitute the total intersection with a curve of order n_2.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 32, 1959.

Total Order

A total order satisfies the conditions for a Partial Order plus the comparability condition. A Relation $\leq$ is a partial order on a Set S if

1. Reflexivity: $a \leq a$ for all $a \in S$
2. Antisymmetry: $a \leq b$ and $b \leq a$ implies $a = b$
3. Transitivity: $a \leq b$ and $b \leq c$ implies $a \leq c$,

and is a total order if, in addition,

4. Comparability: For any $a, b \in S$, either $a \leq b$ or $b \leq a$.

see also Partial Order, Relation

Total Space

The Space E of a Fiber Bundle given by the Map $f: E \to B$, where B is the Base Space of the Fiber Bundle.

see also Base Space, Fiber Bundle, Space

Totative

A POSITIVE INTEGER less than or equal to a number n which is also RELATIVELY PRIME to n, where 1 is counted as being RELATIVELY PRIME to all numbers. The number of totatives of n is the value of the TOTIENT FUNCTION $\phi(n)$.

see also RELATIVELY PRIME, TOTIENT FUNCTION

Totient Function

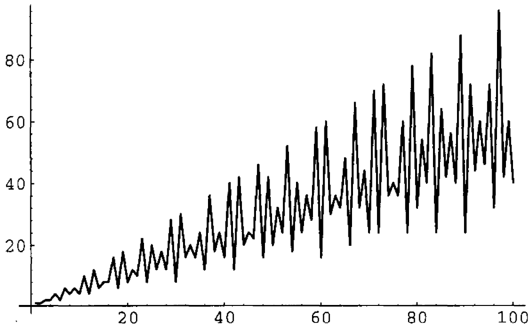

The totient function $\phi(n)$, also called Euler's totient function, is defined as the number of POSITIVE INTEGERS $\leq n$ which are RELATIVELY PRIME to (i.e., do not contain any factor in common with) n, where 1 is counted as being RELATIVELY PRIME to all numbers. Since a number less than or equal to and RELATIVELY PRIME to a given number is called a TOTATIVE, the totient function $\phi(n)$ can be simply defined as the number of TOTATIVES of n. For example, there are eight TOTATIVES of 24 (1, 5, 7, 11, 13, 17, 19, and 23), so $\phi(24) = 8$.

By convention, $\phi(0) = 1$. The first few values of $\phi(n)$ for $n = 1, 2, \ldots$ are 1, 1, 2, 2, 4, 2, 6, 4, 6, 4, 10, ... (Sloane's A000010). $\phi(n)$ is plotted above for small n.

For a PRIME p,

$$\phi(p) = p - 1, \tag{1}$$

since all numbers less than p are RELATIVELY PRIME to p. If $m = p^\alpha$ is a POWER of a PRIME, then the numbers which have a common factor with m are the multiples of p: p, $2p$, ..., $(p^{\alpha-1})p$. There are $p^{\alpha-1}$ of these multiples, so the number of factors RELATIVELY PRIME to p^α is

$$\phi(p^\alpha) = p^\alpha - p^{\alpha-1} = p^{\alpha-1}(p-1) = p^\alpha\left(1 - \frac{1}{p}\right). \tag{2}$$

Now take a general m divisible by p. Let $\phi_p(m)$ be the number of POSITIVE INTEGERS $\leq m$ not DIVISIBLE by p. As before, p, $2p$, ..., $(m/p)p$ have common factors, so

$$\phi_p(m) = m - \frac{m}{p} = m\left(1 - \frac{1}{p}\right). \tag{3}$$

Now let q be some other PRIME dividing m. The INTEGERS divisible by q are q, $2q$, ..., $(m/q)q$. But these duplicate pq, $2pq$, ..., $(m/pq)pq$. So the number of terms which must be subtracted from ϕ_p to obtain ϕ_{pq} is

$$\Delta\phi_q(m) = \frac{m}{q} - \frac{m}{pq} = \frac{m}{q}\left(1 - \frac{1}{p}\right), \tag{4}$$

and

$$\begin{aligned}\phi_{pq}(m) &\equiv \phi_q(m) - \Delta\phi_q(m)\\ &= m\left(1 - \frac{1}{p}\right) - \frac{m}{q}\left(1 - \frac{1}{p}\right)\\ &= m\left(1 - \frac{1}{p}\right)\left(1 - \frac{1}{q}\right).\end{aligned} \tag{5}$$

By induction, the general case is then

$$\phi(n) = n\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right)\cdots\left(1 - \frac{1}{p_r}\right). \tag{6}$$

An interesting identity relates $\phi(n^2)$ to $\phi(n)$,

$$\phi(n^2) = n\phi(n). \tag{7}$$

Another identity relates the DIVISORS d of n to n via

$$\sum_d \phi(d) = n. \tag{8}$$

The DIVISOR FUNCTION satisfies the CONGRUENCE

$$n\sigma(n) \equiv 2 \pmod{\phi(n)} \tag{9}$$

for all PRIMES and no COMPOSITE with the exceptions of 4, 6, and 22 (Subbarao 1974), where $\sigma(n)$ is the DIVISOR FUNCTION. No COMPOSITE solution is currently known to

$$n - 1 \equiv 0 \pmod{\phi(n)} \tag{10}$$

(Honsberger 1976, p. 35).

Walfisz (1963), building on the work of others, showed that

$$\sum_{n=1}^N \phi(n) = \frac{3N^2}{\pi^2} + \mathcal{O}[N(\ln N)^{2/3}(\ln\ln N)^{4/3}], \tag{11}$$

and Landau (1900, quoted in Dickson 1952) showed that

$$\sum_{n=1}^N \frac{1}{\phi(n)} = A\ln N + B + \mathcal{O}\left(\frac{\ln N}{N}\right), \tag{12}$$

where

$$\begin{aligned}A &= \sum_{k=1}^\infty \frac{[\mu(k)]^2}{k\phi(k)} = \frac{\zeta(2)\zeta(3)}{\zeta(6)} = \frac{315}{2\pi^4}\zeta(3)\\ &= 1.9435964368\ldots\end{aligned} \tag{13}$$

$$\begin{aligned}B &= \gamma\frac{315}{2\pi^4}\zeta(3) - \sum_{k=1}^\infty \frac{[\mu(k)]^2\ln k}{k\phi(k)}\\ &= -0.0595536246\ldots,\end{aligned} \tag{14}$$

$\mu(k)$ is the MÖBIUS FUNCTION, $\zeta(z)$ is the RIEMANN ZETA FUNCTION, and γ is the EULER-MASCHERONI CONSTANT (Dickson). A can also be written

$$A = \prod_{k=1}^{\infty} \frac{1 - p_k{}^6}{(1 - p_k{}^{-2})(1 - p_k{}^{-3})} = \prod_{k=1}^{\infty} \left[1 + \frac{1}{p_k(p_k - 1)}\right]. \tag{15}$$

Note that this constant is similar to ARTIN'S CONSTANT.

If the GOLDBACH CONJECTURE is true, then for every number m, there are PRIMES p and q such that

$$\phi(p) + \phi(q) = 2m \tag{16}$$

(Guy 1994, p. 105).

Curious equalities of consecutive values include

$$\phi(5186) = \phi(5187) = \phi(5188) = 2^5 3^4 \tag{17}$$

$$\phi(25930) = \phi(25935) = \phi(25940) = \phi(25942) = 2^7 3^4 \tag{18}$$

$$\phi(404471) = \phi(404473) = \phi(404477) = 2^8 3^2 5^2 7 \tag{19}$$

(Guy 1994, p. 91).

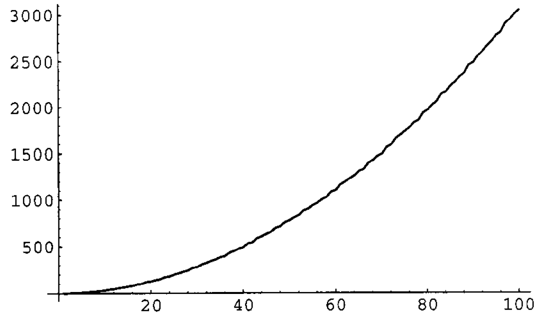

The SUMMATORY totient function, plotted above, is defined by

$$\Phi(n) \equiv \sum_{k=1}^{n} \phi(k) \tag{20}$$

and has the asymptotic series

$$\Phi(x) \sim \frac{1}{2\zeta(2)} x^2 + (x \ln x) \tag{21}$$

$$\sim \frac{3}{\pi^2} x^2 + \mathcal{O}(x \ln x), \tag{22}$$

where $\zeta(z)$ is the RIEMANN ZETA FUNCTION (Perrot 1881). The first values of $\Phi(n)$ are 1, 2, 4, 6, 10, 12, 18, 22, 28, ... (Sloane's A002088).

see also DEDEKIND FUNCTION, EULER'S TOTIENT RULE, FERMAT'S LITTLE THEOREM, LEHMER'S PROBLEM, LEUDESDORF THEOREM, NONCOTOTIENT, NONTOTIENT, SILVERMAN CONSTANT, TOTATIVE, TOTIENT VALENCE FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "The Euler Totient Function." §24.3.2 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 826, 1972.

Beiler, A. H. Ch. 12 in *Recreations in the Theory of Numbers: The Queen of Mathematics Entertains.* New York: Dover, 1966.

Conway, J. H. and Guy, R. K. "Euler's Totient Numbers." *The Book of Numbers.* New York: Springer-Verlag, pp. 154–156, 1996.

Courant, R. and Robbins, H. "Euler's φ Function. Fermat's Theorem Again." §2.4.3 in Supplement to Ch. 1 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 48–49, 1996.

DeKoninck, J.-M. and Ivic, A. *Topics in Arithmetical Functions: Asymptotic Formulae for Sums of Reciprocals of Arithmetical Functions and Related Fields.* Amsterdam, Netherlands: North-Holland, 1980.

Dickson, L. E. *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, pp. 113–158, 1952.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/totient/totient.html`.

Guy, R. K. "Euler's Totient Function," "Does $\phi(n)$ Properly Divide $n-1$," "Solutions of $\phi(m) = \sigma(n)$," "Carmichael's Conjecture," "Gaps Between Totatives," "Iterations of ϕ and σ," "Behavior of $\phi(\sigma(n))$ and $\sigma(\phi(n))$." §B36–B42 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 90-99, 1994.

Halberstam, H. and Richert, H.-E. *Sieve Methods.* New York: Academic Press, 1974.

Honsberger, R. *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., p. 35, 1976.

Perrot, J. 1811. Quoted in Dickson, L. E. *History of the Theory of Numbers, Vol. 1: Divisibility and Primality.* New York: Chelsea, p. 126, 1952.

Shanks, D. "Euler's ϕ Function." §2.27 in *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 68–71, 1993.

Sloane, N. J. A. Sequences A000010/M0299 and A002088/M1008 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Subbarao, M. V. "On Two Congruences for Primality." *Pacific J. Math.* **52**, 261–268, 1974.

Totient Function Constants

see SILVERMAN CONSTANT, TOTIENT FUNCTION

Totient Valence Function

$N_\phi(m)$ is the number of INTEGERS n for which $\phi(n) = m$, also called the MULTIPLICITY of m (Guy 1994). The table below lists values for $\phi(N) \leq 50$.

$\phi(N)$	m	N
1	2	1, 2
2	3	3, 4, 6
4	4	5, 8, 10, 12
6	4	7, 9, 14, 18
8	5	15, 16, 20, 24, 30
10	2	11, 22
12	6	13, 21, 26, 28, 36, 42
16	6	17, 32, 34, 40, 48, 60
18	4	19, 27, 38, 54
20	5	25, 33, 44, 50, 66
22	2	23, 46
24	10	35, 39, 45, 52, 56, 70, 72, 78, 84, 90
28	2	29, 58
30	2	31, 62
32	7	51, 64, 68, 80, 96, 102, 120
36	8	37, 57, 63, 74, 76, 108, 114, 126
40	9	41, 55, 75, 82, 88, 100, 110, 132, 150
42	4	43, 49, 86, 98
44	3	69, 92, 138
46	2	47, 94
48	11	65, 104, 105, 112, 130, 140, 144, 156, 168, 180, 210

A table listing the first value of $\phi(N)$ with multiplicities up to 100 follows (Sloane's A014573).

M	ϕ	M	ϕ	M	ϕ	M	ϕ
0	3	26	2560	51	4992	76	21840
2	1	27	384	52	17640	77	9072
3	2	28	288	53	2016	78	38640
4	4	29	1320	54	1152	79	9360
5	8	30	3696	55	6000	80	81216
6	12	31	240	56	12288	81	4032
7	32	32	768	57	4752	82	5280
8	36	33	9000	58	2688	83	4800
9	40	34	432	59	3024	84	4608
10	24	35	7128	60	13680	85	16896
11	48	36	4200	61	9984	86	3456
12	160	37	480	62	1728	87	3840
13	396	38	576	63	1920	88	10800
14	2268	39	1296	64	2400	89	9504
15	704	40	1200	65	7560	90	18000
16	312	41	15936	66	2304	91	23520
17	72	42	3312	67	22848	92	39936
18	336	43	3072	68	8400	93	5040
19	216	44	3240	69	29160	94	26208
20	936	45	864	70	5376	95	27360
21	144	46	3120	71	3360	96	6480
22	624	47	7344	72	1440	97	9216
23	1056	48	3888	73	13248	98	2880
24	1760	49	720	74	11040	99	26496
25	360	50	1680	75	27720	100	34272

It is thought that $N_\phi(m) \geq 2$ (i.e., the totient valence function never takes on the value 1), but this has not been proven. This assertion is called CARMICHAEL'S TOTIENT FUNCTION CONJECTURE and is equivalent to the statement that for all n, there exists $m \neq n$ such that $\phi(n) = \phi(m)$ (Ribenboim 1996, pp. 39–40). Any counterexample must have more than 10,000,000 DIGITS (Schlafly and Wagon 1994, Conway and Guy 1996).

see also CARMICHAEL'S TOTIENT FUNCTION CONJECTURE, TOTIENT FUNCTION

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 155, 1996.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 94, 1994.

Ribenboim, P. *The New Book of Prime Number Records.* New York: Springer-Verlag, 1996.

Schlafly, A. and Wagon, S. "Carmichael's Conjecture on the Euler Function is Valid Below $10^{10,000,000}$." *Math. Comput.* **63**, 415–419, 1994.

Sloane, N. J. A. Sequence A014573 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Touchard's Congruence

$$B_{p+k} \equiv B_k + B_{k+1} \pmod{p},$$

when p is PRIME and B_n is a BELL NUMBER.

see also BELL NUMBER

Tour

A sequence of moves on a chessboard by a CHESS piece in which each square of a CHESSBOARD is visited exactly once.

see also CHESS, KNIGHT'S TOUR, MAGIC TOUR, TRAVELING SALESMAN CONSTANTS

Tournament

A COMPLETE DIRECTED GRAPH. A so-called SCORE SEQUENCE can be associated with every tournament. Every tournament contains a HAMILTONIAN PATH.

see also COMPLETE GRAPH, DIRECTED GRAPH, HAMILTONIAN PATH, SCORE SEQUENCE

References

Chartrand, G. "Tournaments." §27.2 in *Introductory Graph Theory.* New York: Dover, pp. 155–161, 1985.

Moon, J. W. *Topics on Tournaments.* New York: Holt, Rinehart, and Winston, 1968.

Ruskey, F. "Information on Score Sequences." `http://sue.csc.uvic.ca/~cos/inf/nump/ScoreSequence.html`.

Tournament Matrix

A matrix for a round-robin tournament involving n players competing in $n(n-1)/2$ matches (no ties allowed) having entries

$$a_{ij} = \begin{cases} 1 & \text{if player } i \text{ defeats player } j \\ -1 & \text{if player } i \text{ loses to player } j \\ 0 & \text{if } i = j. \end{cases}$$

The MATRIX satisfies

$$\mathsf{A} + \mathsf{A}^{\mathrm{T}} + \mathsf{I} = \mathsf{J},$$

where I is the IDENTITY MATRIX, J is an $n \times n$ MATRIX of all 1s, and A^{T} is the MATRIX TRANSPOSE of A.

The tournament matrix for n players has zero DETERMINANT IFF n is ODD (McCarthy and Benjamin 1996). The dimension of the NULLSPACE of an n-player tournament matrix is

$$\dim[\text{nullspace}] = \begin{cases} 0 & \text{for } n \text{ even} \\ 1 & \text{for } n \text{ odd} \end{cases}$$

(McCarthy 1996).

References

McCarthy, C. A. and Benjamin, A. T. "Determinants of the Tournaments." *Math. Mag.* **69**, 133–135, 1996.

Michael, T. S. "The Ranks of Tournament Matrices." *Amer. Math. Monthly* **102**, 637–639, 1995.

Tower of Power

see POWER TOWER

Towers of Hanoi

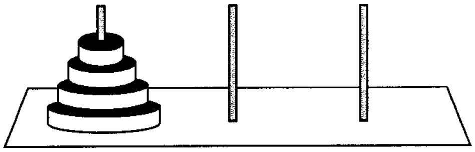

A PUZZLE invented by E. Lucas in 1883. Given a stack of n disks arranged from largest on the bottom to smallest on top placed on a rod, together with two empty rods, the towers of Hanoi puzzle asks for the minimum number of moves required to reverse the order of the stack (where moves are allowed only if they place smaller disks on top of larger disks). The problem is ISOMORPHIC to finding a HAMILTONIAN PATH on an n-HYPERCUBE.

For n disks, the number of moves h_n required is given by the RECURRENCE RELATION

$$h_n = 2h_{n-1} + 1.$$

Solving gives

$$h_n = 2^n - 1.$$

The number of disks moved after the kth step is the same as the element which needs to be added or deleted in the kth ADDEND of the RYSER FORMULA (Gardner 1988, Vardi 1991).

A HANOI GRAPH can be constructed whose VERTICES correspond to legal configurations of n towers of Hanoi, where the VERTICES are adjacent if the corresponding configurations can be obtained by a legal move. It can be solved using a binary GRAY CODE.

Poole (1994) gives *Mathematica*® (Wolfram Research, Champaign, IL) routines for solving an arbitrary disk configuration in the fewest possible moves. The proof of minimality is achieved using the LUCAS CORRESPONDENCE which relates PASCAL'S TRIANGLE to the HANOI GRAPH. ALGORITHMS are known for transferring disks for four pegs, but none has been proved minimal. For additional references, see Poole (1994).

see also GRAY CODE, RYSER FORMULA

References

Bogomolny, A. "Towers of Hanoi." `http://www.cut-the-knot.com/recurrence/hanoi.html`.

Chartrand, G. "The Tower of Hanoi Puzzle." §6.3 in *Introductory Graph Theory.* New York: Dover, pp. 135–139, 1985.

Dubrovsky, V. "Nesting Puzzles, Part I: Moving Oriental Towers." *Quantum* **6**, 53–57 (Jan.) and 49–51 (Feb.), 1996.

Gardner, M. "The Icosian Game and the Tower of Hanoi." Ch. 6 in *The Scientific American Book of Mathematical Puzzles & Diversions.* New York: Simon and Schuster, 1959.

Kasner, E. and Newman, J. R. *Mathematics and the Imagination.* Redmond, WA: Tempus Books, pp. 169–171, 1989.

Kolar, M. "Towers of Hanoi." `http://www.pangea.ca/kolar/javascript/Hanoi/Hanoi.html`.

Poole, D. G. "The Towers and Triangles of Professor Claus (or, Pascal Knows Hanoi)." *Math. Mag.* **67**, 323–344, 1994.

Poole, D. G. "Towers of Hanoi." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Hanoi.m`.

Ruskey, F. "Towers of Hanoi." `http://sue.csc.uvic.ca/~cos/inf/comb/SubsetInfo.html#Hanoi`.

Schoutte, P. H. "De Ringen van Brahma." *Eigen Haard* **22**, 274–276, 1884.

Kraitchik, M. "The Tower of Hanoi." §3.12.4 in *Mathematical Recreations.* New York: W. W. Norton, pp. 91–93, 1942.

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 111–112, 1991.

Trace (Complex)

The image of the path γ in $\mathbb{C}$ under the FUNCTION f is called the trace. This term is unrelated to that applied to MATRICES and TENSORS.

Trace (Group)

see CHARACTER (GROUP)

Trace (Map)

Let a PATCH be given by the map $\mathbf{x} : U \to \mathbb{R}^n$, where U is an open subset of $\mathbb{R}^2$, or more generally by $\mathbf{x} : A \to \mathbb{R}^n$, where A is any SUBSET of $\mathbb{R}^2$. Then $\mathbf{x}(U)$ (or more generally, $\mathbf{x}(A)$) is called the trace of $\mathbf{x}$.

see also PATCH

References

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 183–184, 1993.

Trace (Matrix)

The trace of an $n \times n$ SQUARE MATRIX A is defined by

$$\operatorname{Tr}(\mathsf{A}) \equiv a_{ii}, \tag{1}$$

where EINSTEIN SUMMATION is used (i.e., the a_{ii} is summed over $i = 1, \ldots, n$). For SQUARE MATRICES A and B, it is true that

$$\operatorname{Tr}(\mathsf{A}) = \operatorname{Tr}(\mathsf{A}^{\mathrm{T}}) \tag{2}$$

$$\operatorname{Tr}(\mathsf{A} + \mathsf{B}) = \operatorname{Tr}(\mathsf{A}) + \operatorname{Tr}(\mathsf{B}) \tag{3}$$

$$\operatorname{Tr}(\alpha\mathsf{A}) = \alpha \operatorname{Tr}(\mathsf{A}) \tag{4}$$

(Lange 1987, p. 40). The trace is invariant under a SIMILARITY TRANSFORMATION

$$\mathsf{A}' \equiv \mathsf{BAB}^{-1} \tag{5}$$

(Lange 1987, p. 64). Since

$$(bab^{-1})_{ij} = b_{il}a_{lk}b^{-1}_{kj}, \tag{6}$$

$$\begin{aligned}\mathrm{Tr}(\mathsf{BAB}^{-1}) &= b_{il}a_{lk}b^{-1}{}_{ki} \\ &= (b^{-1}b)_{kl}a_{lk} = \delta_{kl}a_{lk} \\ &= a_{kk} = \mathrm{Tr}(\mathsf{A}),\end{aligned} \tag{7}$$

where δ_{ij} is the KRONECKER DELTA.

The trace of a product of square matrices is independent of the order of the multiplication since

$$\begin{aligned}\mathrm{Tr}(\mathsf{AB}) &= (ab)_{ii} = a_{ij}b_{ji} = b_{ji}a_{ij} \\ &= (ba)_{jj} = \mathrm{Tr}(\mathsf{BA}).\end{aligned} \tag{8}$$

Therefore, the trace of the COMMUTATOR of A and B is given by

$$\mathrm{Tr}([\mathsf{A},\mathsf{B}]) \equiv \mathrm{Tr}(\mathsf{AB}) - \mathrm{Tr}(\mathsf{BA}) = 0. \tag{9}$$

The product of a SYMMETRIC and an ANTISYMMETRIC MATRIX has zero trace,

$$\mathrm{Tr}(A_S B_A) = 0. \tag{10}$$

The value of the trace can be found using the fact that the matrix can always be transformed to a coordinate system where the z-AXIS lies along the axis of rotation. In the new coordinate system, the MATRIX is

$$\mathsf{A}' = \begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{11}$$

so the trace is

$$\mathrm{Tr}(\mathsf{A}') = \mathrm{Tr}(\mathsf{A}) \equiv a_{ii} = 1 + 2\cos\phi. \tag{12}$$

References

Lang, S. *Linear Algebra, 3rd ed.* New York: Springer-Verlag, pp. 40 and 64, 1987.

Trace (Tensor)

The trace of a second-RANK TENSOR T is a SCALAR given by the CONTRACTED mixed TENSOR equal to T_i^i. The trace satisfies

$$\mathrm{Tr}\left[M^{-1}(x)\frac{\partial}{\partial x^\lambda}M(x)\right] = \frac{\partial}{\partial x^\lambda}\ln[\det(x)],$$

and

$$\begin{aligned}\delta\ln[\det M] &= \ln[\det(M+\delta M)] - \ln(\det M) \\ &= \ln\left[\frac{\det(M+\delta M)}{\det M}\right] \\ &= \ln[\det M^{-1}(M+\delta M)] \\ &= \ln[\det(1+M^{-1}\delta M)] \\ &\approx \ln[1+\mathrm{Tr}(M^{-1}\delta M)] \\ &\approx \mathrm{Tr}(M^{-1}\delta M).\end{aligned}$$

see also CONTRACTION (TENSOR)

Tractory

see TRACTRIX

Tractrix

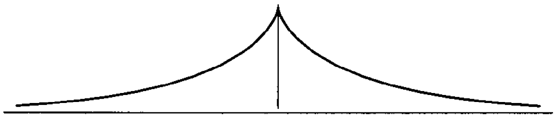

The tractrix is the CATENARY INVOLUTE described by a point initially on the vertex. It has a constant NEGATIVE CURVATURE and is sometimes called the TRACTORY or EQUITANGENTIAL CURVE. The tractrix was first studied by Huygens in 1692, who gave it the name "tractrix." Later, Leibniz, Johann Bernoulli, and others studied the curve.

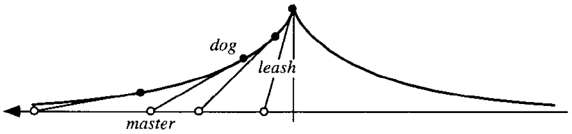

The tractrix arises from the following problem posed to Leibniz: What is the path of an object starting off with a vertical offset when it is dragged along by a string of constant length being pulled along a straight horizontal line? By associating the object with a dog, the string with a leash, and the pull along a horizontal line with the dog's master, the curve has the descriptive name HUNDKURVE (hound curve) in German. Leibniz found the curve using the fact that the axis is an asymptote to the tractrix (MacTutor Archive).

In CARTESIAN COORDINATES the tractrix has equation

$$x = a\,\mathrm{sech}^{-1}\left(\frac{y}{a}\right) - \sqrt{a^2-y^2}. \tag{1}$$

One parametric form is

$$x(t) = a(t - \tanh t) \tag{2}$$

$$y(t) = a \operatorname{sech} t. \tag{3}$$

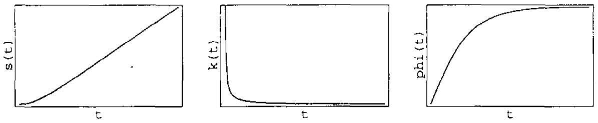

The Arc Length, Curvature, and Tangential Angle are

$$s(t) = \ln(\cosh t) \tag{4}$$

$$\kappa(t) = \operatorname{csch} t \tag{5}$$

$$\phi(t) = 2\tan^{-1}[\tanh(\tfrac{1}{2}t)]. \tag{6}$$

A second parametric form in terms of the Angle ϕ of the straight line tangent to the tractrix is

$$x = a\{\ln[\tan(\tfrac{1}{2}\phi)] + \cos\phi\} \tag{7}$$

$$y = a\sin\phi \tag{8}$$

(Gray 1993). This parameterization has Curvature

$$\kappa(\phi) = |\tan\phi|. \tag{9}$$

A parameterization which traverses the tractrix with constant speed a is given by

$$x(t) = \begin{cases} ae^{-v/a} & \text{for } v \in [0,\infty) \\ ae^{v/a} & \text{for } v \in (-\infty, 0] \end{cases} \tag{10}$$

$$y(t) = \begin{cases} a[\tanh^{-1}(\sqrt{1-e^{-2v/a}}\,) - \sqrt{1-e^{-2v/a}}] & \text{for } v \in [0,\infty) \\ a[-\tanh^{-1}(\sqrt{1-e^{2v/a}}\,) + \sqrt{1-e^{2v/a}}] & \text{for } v \in (-\infty, 0]. \end{cases} \tag{11}$$

When a tractrix is rotated around its asymptote, a Pseudosphere results. This is a surface of constant Negative Curvature. For a tractrix, the length of a Tangent from its point of contact to an asymptote is constant. The Area between the tractrix and its asymptote is finite.

see also Curvature, Dini's Surface, Mice Problem, Pseudosphere, Pursuit Curve, Tractroid

References
Geometry Center. "The Tractrix." `http://www.geom.umn.edu/zoo/diffgeom/pseudosphere/tractrix.html`.
Gray, A. "The Tractrix" and "The Evolute of a Tractrix is a Catenary." §3.5 and 5.3 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 46–50 and 80–81, 1993.
Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 199–200, 1972.
Lee, X. "Tractrix." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Tractrix_dir/tractrix.html`.
Lockwood, E. H. "The Tractrix and Catenary." Ch. 13 in *A Book of Curves.* Cambridge, England: Cambridge University Press, pp. 118–124, 1967.
MacTutor History of Mathematics Archive. "Tractrix." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Tractrix.html`.
Yates, R. C. "Tractrix." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 221–224, 1952.

Tractrix Evolute

The Evolute of the Tractrix is the Catenary.

Tractrix Radial Curve

The Radial Curve of the Tractrix is the Kappa Curve.

Tractroid

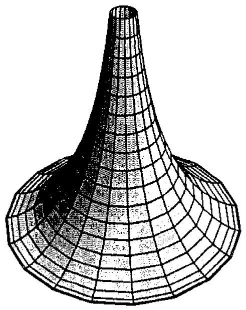

The Surface of Revolution produced by revolving the Tractrix

$$x = \operatorname{sech} u \tag{1}$$

$$z = u - \tanh u \tag{2}$$

about the z-Axis is a tractroid given by

$$x = \operatorname{sech} u \cos v \tag{3}$$

$$y = \operatorname{sech} u \sin v \tag{4}$$

$$z = u - \tanh u. \tag{5}$$

see also Pseudosphere, Surface of Revolution, Tractrix

Transcendental Curve

A curve which intersects some straight line in an infinity of points (but for which not every point lies on this curve).

References
Borwein, J. M.; Borwein, P. B.; and Bailey, D. H. "Ramanujan, Modular Equations, and Approximations to Pi or How to Compute One Billion Digits of Pi." *Amer. Math. Monthly* **96**, 201–219, 1989.

Transcendental Equation

An equation or formula involving Transcendental Functions.

Transcendental Function

A function which "transcends," i.e., cannot be expressed in terms of, the usual Elementary Functions. Define

$$\begin{aligned} l_1(z) &\equiv l(z) \equiv \ln(z) \\ e_1(z) &\equiv e(z) \equiv e^z \\ \varsigma_1 f(z) &\equiv \varsigma f(z) \equiv \int f(z)\,dz, \end{aligned}$$

and let $l_2 \equiv l(l(z))$, etc. These are called the "elementary" transcendental functions (Watson 1966, p. 111).

see also Algebraic Function, Elementary Function

References

Watson, G. N. *A Treatise on the Theory of Bessel Functions, 2nd ed.* Cambridge, England: Cambridge University Press, 1966.

Transcendental Number

A number which is not the Root of *any* Polynomial equation with Integer Coefficients, meaning that it not an Algebraic Number of any degree, is said to be transcendental. This definition guarantees that every transcendental number must also be Irrational, since a Rational Number is, by definition, an Algebraic Number of degree one.

Transcendental numbers are important in the history of mathematics because their investigation provided the first proof that Circle Squaring, one of the Geometric Problems of Antiquity which had baffled mathematicians for more than 2000 years was, in fact, insoluble. Specifically, in order for a number to be produced by a Geometric Construction using the ancient Greek rules, it must be either Rational or a very special kind of Algebraic Number known as a Euclidean Number. Because the number π is transcendental, the construction cannot be done according to the Greek rules.

Georg Cantor was the first to prove the Existence of transcendental numbers. Liouville subsequently showed how to construct special cases (such as Liouville's Constant) using Liouville's Rational Approximation Theorem. In particular, he showed that any number which has a rapidly converging sequence of rational approximations must be transcendental. For many years, it was only known how to determine if special classes of numbers were transcendental. The determination of the status of more general numbers was considered an important enough unsolved problem that it was one of Hilbert's Problems.

Great progress was subsequently made by Gelfond's Theorem, which gives a general rule for determining if special cases of numbers of the form α^β are transcendental. Baker produced a further revolution by proving the transcendence of sums of numbers of the form $\alpha \ln \beta$ for Algebraic Numbers α and β.

The number e was proven to be transcendental by Hermite in 1873, and Pi (π) by Lindemann in 1882. e^π is transcendental by Gelfond's Theorem since

$$(-1)^{-i} = (e^{i\pi})^{-i} = e^\pi.$$

The Gelfond-Schneider Constant $2^{\sqrt{2}}$ is also transcendental. Other known transcendentals are $\sin 1$ where $\sin x$ is the Sine function, $J_0(1)$ where $J_0(x)$ is a Bessel Function of the First Kind (Hardy and Wright 1985), $\ln 2$, $\ln 3/\ln 2$, the first zero $x_0 = 2.4048255\ldots$ of the Bessel Function $J_0(x_0)$ (Le Lionnais 1983, p. 46), $\pi + \ln 2 + \sqrt{2}\ln 3$ (Borwein *et al.* 1989), the Thue-Morse Constant $P = 0.4124540336\ldots$ (Dekking 1977, Allouche and Shallit), the Champernowne Constant $0.1234567891011\ldots$, the Thue Constant

$$0.110110111110110111110110110\ldots,$$

$\Gamma(\frac{1}{3})$ (Le Lionnais 1983, p. 46), $\Gamma(\frac{1}{4})\pi^{-1/4}$ (Davis 1959), and $\Gamma(\frac{1}{4})$ (Chudnovsky, Waldschmidt), where $\Gamma(x)$ is the Gamma Function. At least one of πe and $\pi + e$ (and probably both) are transcendental, but transcendence has not been proven for either number on its own.

It is not known if e^e, π^π, π^e, γ (the Euler-Mascheroni Constant), $I_0(2)$, or $I_1(2)$ (where $I_n(x)$ is a Modified Bessel Function of the First Kind) are transcendental.

The "degree" of transcendence of a number can be characterized by a so-called Liouville-Roth Constant. There are still many fundamental and outstanding problems in transcendental number theory, including the Constant Problem and Schanuel's Conjecture.

see also Algebraic Number, Constant Problem, Gelfond's Theorem, Irrational Number, Lindemann-Weierstraß Theorem, Liouville-Roth Constant, Roth's Theorem, Schanuel's Conjecture, Thue-Siegel-Roth Theorem

References

Allouche, J. P. and Shallit, J. In preparation.

Baker, A. "Approximations to the Logarithm of Certain Rational Numbers." *Acta Arith.* **10**, 315–323, 1964.

Baker, A. "Linear Forms in the Logarithms of Algebraic Numbers I." *Mathematika* **13**, 204–216, 1966.

Baker, A. "Linear Forms in the Logarithms of Algebraic Numbers II." *Mathematika* **14**, 102–107, 1966.

Baker, A. "Linear Forms in the Logarithms of Algebraic Numbers III." *Mathematika* **14**, 220–228, 1966.

Baker, A. "Linear Forms in the Logarithms of Algebraic Numbers IV." *Mathematika* **15**, 204–216, 1966.

Borwein, J. M.; Borwein, P. B.; and Bailey, D. H. "Ramanujan, Modular Equations, and Approximations to Pi or How to Compute One Billion Digits of Pi." *Amer. Math. Monthly* **96**, 201–219, 1989.

Chudnovsky, G. V. *Contributions to the Theory of Transcendental Numbers.* Providence, RI: Amer. Math. Soc., 1984.

Courant, R. and Robbins, H. "Algebraic and Transcendental Numbers." §2.6 in *What is Mathematics?: An Elementary*

Approach to Ideas and Methods, 2nd ed. Oxford, England: Oxford University Press, pp. 103–107, 1996.

Davis, P. J. "Leonhard Euler's Integral: A Historical Profile of the Gamma Function." *Amer. Math. Monthly* **66**, 849–869, 1959.

Dekking, F. M. "Transcendence du nombre de Thue-Morse." *Comptes Rendus de l'Academie des Sciences de Paris* **285**, 157–160, 1977.

Gray, R. "Georg Cantor and Transcendental Numbers." *Amer. Math. Monthly* **101**, 819–832, 1994.

Hardy, G. H. and Wright, E. M. *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Oxford University Press, 1985.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 46, 1983.

Siegel, C. L. *Transcendental Numbers.* New York: Chelsea, 1965.

Transcritical Bifurcation

Let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a one-parameter family of C^2 maps satisfying

$$f(0,\mu) = 0 \tag{1}$$

$$\left[\frac{\partial f}{\partial x}\right]_{\mu=0,x=0} = 1 \tag{2}$$

$$\left[\frac{\partial f}{\partial x}\right]_{\mu,x} = \left[\frac{\partial f}{\partial x}\right]_{\mu=0,x=\mu} \tag{3}$$

$$\left[\frac{\partial^2 f}{\partial x \partial \mu}\right]_{0,0} > 0 \tag{4}$$

$$\left[\frac{\partial^2 f}{\partial \mu^2}\right]_{\mu=0,x=0} > 0. \tag{5}$$

Then there are two branches, one stable and one unstable. This BIFURCATION is called a transcritical bifurcation. An example of an equation displaying a transcritical bifurcation is

$$\dot{x} = \mu x - x^2. \tag{6}$$

(Guckenheimer and Holmes 1997, p. 145).

see also BIFURCATION

References

Guckenheimer, J. and Holmes, P. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields, 3rd ed.* New York: Springer-Verlag, pp. 145 and 149–150, 1997.

Rasband, S. N. *Chaotic Dynamics of Nonlinear Systems.* New York: Wiley, pp. 27–28, 1990.

Transfer Function

The engineering terminology for one use of FOURIER TRANSFORMS. By breaking up a wave pulse into its frequency spectrum

$$f_\nu = F(\nu)e^{2\pi i\nu t}, \tag{1}$$

the entire signal can be written as a sum of contributions from each frequency,

$$f(t) = \int_{-\infty}^{\infty} f_\nu \, d\nu = \int_{-\infty}^{\infty} F(\nu)e^{2\pi i\nu t} \, d\nu. \tag{2}$$

If the signal is modified in some way, it will become

$$g_\nu(t) = \phi(\nu)f_\nu(t) = \phi(\nu)F(\nu)e^{2\pi i\nu t} \tag{3}$$

$$g(t) = \int_{-\infty}^{\infty} g_\nu(t)\, dt = \int_{-\infty}^{\infty} \phi(\nu)F(\nu)e^{2\pi i\nu t} \, d\nu, \tag{4}$$

where $\phi(\nu)$ is known as the "transfer function." FOURIER TRANSFORMING ϕ and F,

$$\phi(\nu) = \int_{-\infty}^{\infty} \Phi(t)e^{-2\pi i\nu t} \, dt \tag{5}$$

$$F(\nu) = \int_{-\infty}^{\infty} f(t)e^{-2\pi i\nu t} \, dt. \tag{6}$$

From the CONVOLUTION THEOREM,

$$g(t) = f(t) * \Phi(t) = \int_{-\infty}^{\infty} f(t)\Phi(t-\tau)\, d\tau. \tag{7}$$

see also CONVOLUTION THEOREM, FOURIER TRANSFORM

Transfinite Diameter

Let

$$\phi(z) = cz + c_0 + c_1 z^{-1} + c_2 z^{-2} + \ldots$$

be an ANALYTIC FUNCTION, REGULAR and UNIVALENT for $|z| > 1$, which maps $|z| > 1$ CONFORMALLY onto the region T preserving the POINT AT INFINITY and its direction. Then the function $\phi(z)$ is uniquely determined and c is called the transfinite diameter, sometimes also known as ROBIN'S CONSTANT or the CAPACITY of $\phi(z)$.

see also ANALYTIC FUNCTION, REGULAR FUNCTION, UNIVALENT FUNCTION

Transfinite Number

One of Cantor's ORDINAL NUMBERS ω, $\omega+1$, $\omega+2$, ..., $\omega+\omega$, $\omega+\omega+1$, ... which is "larger" than any WHOLE NUMBER.

see also $\aleph_0$, $\aleph_1$, CARDINAL NUMBER, CONTINUUM, ORDINAL NUMBER, WHOLE NUMBER

References

Pappas, T. "Transfinite Numbers." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 156–158, 1989.

Transform

A shortened term for INTEGRAL TRANSFORM.

Geometrically, if S and T are two transformations, then the SIMILARITY TRANSFORMATION TST^{-1} is sometimes called the transform (Woods 1961).

see also Abel Transform, Boustrophedon Transform, Discrete Fourier Transform, Fast Fourier Transform, Fourier Transform, Fractional Fourier Transform, Hankel Transform, Hartley Transform, Hilbert Transform, Laplace-Stieltjes Transform, Laplace Transform, Mellin Transform, Number Theoretic Transform, Poncelet Transform, Radon Transform, Wavelet Transform, z-Transform, Z-Transform

References
Woods, F. S. *Higher Geometry: An Introduction to Advanced Methods in Analytic Geometry.* New York: Dover, p. 5, 1961.

Transformation

see Function, Map

Transitive

A Relation R on a Set S is transitive provided that for all x, y and z in S such that xRy and yRz, we also have xRz.

see also Associative, Commutative, Relation

Transitive Closure

The transitive closure of a binary Relation R on a Set X is the minimal Transitive relation R' on X that contains R. Thus $aR'b$ for any elements a and b of X, provided either that aRb or that there exists some element c of X such that aRc and cRb.

see also Reflexive Closure, Transitive Reduction

Transitive Reduction

The transitive reduction of a binary Relation R on a Set X is the minimum relation R' on X with the same Transitive Closure as R. Thus $aR'b$ for any elements a and b of X, provided that aRb and there exists no element c of X such that aRc and cRb.

see also Reflexive Reduction, Transitive Closure

Transitivity Class

Let $S(T)$ be the group of symmetries which map a Monohedral Tiling T onto itself. The Transitivity Class of a given tile T is then the collection of all tiles to which T can be mapped by one of the symmetries of $S(T)$.

see also Monohedral Tiling

References
Berglund, J. "Is There a k-Anisohedral Tile for $k \geq 5$?" *Amer. Math. Monthly* **100**, 585–588, 1993.

Translation

A transformation consisting of a constant offset with no Rotation or distortion. In n-D Euclidean Space, a translation may be specified simply as a Vector giving the offset in each of the n coordinates.

see also Affine Group, Dilation, Euclidean Group, Expansion, Glide, Improper Rotation, Inversion Operation, Mirror Image, Reflection, Rotation

References
Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 211, 1987.

Translation Relation

A mathematical relationship transforming a function $f(x)$ to the form $f(x+a)$.

see also Argument Addition Relation, Argument Multiplication Relation, Recurrence Relation, Reflection Relation

Transpose

The object obtained by replacing all elements a_{ij} with a_{ji}. For a second-Rank Tensor a_{ij}, the tensor transpose is simply a_{ji}. The matrix transpose, written A^{T}, is the Matrix obtained by exchanging A's rows and columns, and satisfies the identity

$$(\mathsf{A}^{\mathrm{T}})^{-1} = (\mathsf{A}^{-1})^{\mathrm{T}}.$$

The product of two transposes satisfies

$$(\mathsf{B}^{\mathrm{T}}\mathsf{A}^{\mathrm{T}})_{ij} = (b^{\mathrm{T}})_{ik}(a^{\mathrm{T}})_{kj} = b_{ki}a_{jk} = a_{jk}b_{ki} = (\mathsf{AB})_{ji}$$
$$= (\mathsf{AB})^{\mathrm{T}}_{ij}.$$

Therefore,

$$(\mathsf{AB})^{\mathrm{T}} = \mathsf{B}^{\mathrm{T}}\mathsf{A}^{\mathrm{T}}.$$

Transpose Map

see Pullback Map

Transposition

An exchange of two elements of a Set with all others staying the same. A transposition is therefore a Permutation of two elements. For example, the swapping of 2 and 5 to take the list 123456 to 153426 is a transposition.

see also Permutation, Transposition Order

Transposition Group

A Permutation Group in which the Permutations are limited to Transpositions.

see also Permutation Group

Transposition Order

An ordering of PERMUTATIONS in which each two adjacent permutations differ by the TRANSPOSITION of two elements. For the permutations of $\{1, 2, 3\}$ there are two listings which are in transposition order. One is 123, 132, 312, 321, 231, 213, and the other is 123, 321, 312, 213, 231, 132.

see also LEXICOGRAPHIC ORDER, PERMUTATION

References
Ruskey, F. "Information on Combinations of a Set." `http://sue.csc.uvic.ca/~cos/inf/comb/CombinationsInfo.html`.

Transversal Array

A set of n cells in an $n \times n$ SQUARE such that no two come from the same row and no two come from the same column. The number of transversals of an $n \times n$ SQUARE is $n!$ (n FACTORIAL).

Transversal Design

A transversal design $\mathrm{TD}_\lambda(k, n)$ of order n, block size k, and index λ is a triple (V, G, B) such that

1. V is a set of kn elements,
2. G is a partition of V into k classes, each of size n (the "groups"),
3. B is a collection of k-subsets of V (the "blocks"), and
4. Every unordered pair of elements from V is contained in either exactly one group or in exactly λ blocks, but not both.

References
Colbourn, C. J. and Dinitz, J. H. (Eds.) *CRC Handbook of Combinatorial Designs.* Boca Raton, FL: CRC Press, p. 112, 1996.

Transversal Line

A transversal line is a LINE which intersects each of a given set of other lines. It is also called a SEMISECANT.

see also LINE

Transylvania Lottery

A lottery in which three numbers are picked at random from the INTEGERS 1–14.

see also FANO PLANE

Trapdoor Function

An easily computed function whose inverse is extremely difficult to compute. An example is the multiplication of two large PRIMES. Finding and verifying two large PRIMES is easy, as is their multiplication. But factorization of the resultant product is very difficult.

see also RSA ENCRYPTION

References
Gardner, M. Chs. 13–14 in *Penrose Tiles and Trapdoor Ciphers... and the Return of Dr. Matrix, reissue ed.* New York: W. H. Freeman, pp. 299–300, 1989.

Trapezium

There are two common definitions of the trapezium. The American definition is a QUADRILATERAL with no PARALLEL sides. The British definition for a trapezium is a QUADRILATERAL *with* two sides PARALLEL. Such a trapezium is equivalent to a TRAPEZOID and therefore has AREA

$$A = \tfrac{1}{2}(a + b)h.$$

see also DIAMOND, LOZENGE, PARALLELOGRAM, QUADRILATERAL, RHOMBOID, RHOMBUS, SKEW QUADRILATERAL, TRAPEZOID

Trapezohedron

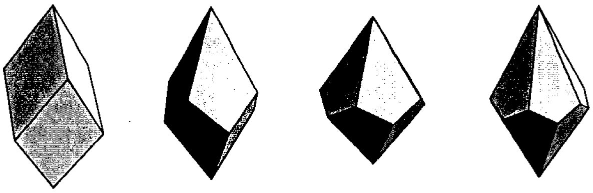

The trapezohedra are the DUAL POLYHEDRA of the Archimedean ANTIPRISMS. However, their faces are not TRAPEZOIDS.

see also ANTIPRISM, DIPYRAMID, HEXAGONAL SCALENOHEDRON, PRISM, TRAPEZOID

References
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 117, 1989.

Trapezoid

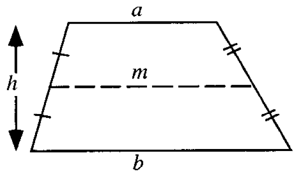

A QUADRILATERAL with two sides PARALLEL. The trapezoid depicted above satisfies

$$m = \tfrac{1}{2}(a + b)$$

and has AREA

$$A = \tfrac{1}{2}(a + b)h = mh.$$

The trapezoid is equivalent to the British definition of TRAPEZIUM.

see also PYRAMIDAL FRUSTUM, TRAPEZIUM

References
Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 123, 1987.

Trapezoidal Hexecontahedron

see DELTOIDAL HEXECONTAHEDRON

Trapezoidal Icositetrahedron

see DELTOIDAL ICOSITETRAHEDRON

Trapezoidal Rule

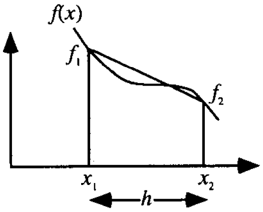

The 2-point NEWTON-COTES FORMULA

$$\int_{x_1}^{x_2} f(x)\,dx = \tfrac{1}{2}h(f_1+f_2) - \tfrac{1}{2}h^3 f''(\xi),$$

where $f_i \equiv f(x_i)$, h is the separation between the points, and ξ is a point satisfying $x_1 \le \xi \le x_2$. Picking ξ to maximize $f''(\xi)$ gives an upper bound for the error in the trapezoidal approximation to the INTEGRAL.

see also BODE'S RULE, HARDY'S RULE, NEWTON-COTES FORMULAS, SIMPSON'S 3/8 RULE, SIMPSON'S RULE, WEDDLE'S RULE

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 885, 1972.

Traveling Salesman Constants

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let $L(n,d)$ be the smallest TOUR length for n points in a d-D HYPERCUBE. Then there exists a smallest constant $\alpha(d)$ such that for all optimal TOURS in the HYPERCUBE,

$$\limsup_{n\to\infty} \frac{L(n,d)}{n^{(d-1)/d}\sqrt{d}} \le \alpha(d), \tag{1}$$

and a constant $\beta(d)$ such that for *almost* all optimal tours in the HYPERCUBE,

$$\lim_{n\to\infty} \frac{L(n,d)}{n^{(d-1)/d}\sqrt{d}} = \beta(d). \tag{2}$$

These constants satisfy the inequalities

$$\begin{aligned} 0.44194 < \gamma_2 = \tfrac{5}{16}\sqrt{2} \le \beta(2) \\ \le \delta < 0.6508 < 0.75983 < 3^{-1/4} \le \alpha(2) \\ \le \phi < 0.98398 \end{aligned} \tag{3}$$

$$\begin{aligned} 0.37313 < \gamma_3 \le \beta(3) \le 12^{1/6}6^{-1/2} < 0.61772 < 0.64805 \\ < 2^{1/6}3^{-1/2} \le \alpha(3) \le 0.90422 \end{aligned} \tag{4}$$

$$\begin{aligned} 0.34207 < \gamma_4 \le \beta(4) \le 12^{1/8}6^{-1/2} < 0.55696 \\ < 0.59460 < 2^{-3/4} \le \alpha(4) \le 0.8364 \end{aligned} \tag{5}$$

(Fejes Tóth 1940, Verblunsky 1951, Few 1955, Beardwood *et al.* 1959), where

$$\gamma_d \equiv \frac{\Gamma\left(3+\frac{1}{d}\right)\left[\Gamma(\frac{1}{2}d+1)\right]^{1/d}}{2\sqrt{\pi}(d^{1/2}+d^{-1/2})}, \tag{6}$$

$\Gamma(z)$ is the GAMMA FUNCTION, δ is an expression involving STRUVE FUNCTIONS and NEUMANN FUNCTIONS,

$$\phi \equiv \frac{280(3-\sqrt{3})}{840-280\sqrt{3}+4\sqrt{5}-\sqrt{10}} \tag{7}$$

(Karloff 1989), and

$$\psi \equiv \tfrac{1}{2}3^{-2/3}(4+\ln 3)^{2/3} \tag{8}$$

(Goddyn 1990). In the LIMIT $d\to\infty$,

$$\begin{aligned} 0.24197 < \lim_{d\to\infty}\gamma_d = \frac{1}{\sqrt{2\pi e}} \le \liminf_{d\to\infty}\beta(d) \\ \le \limsup_{d\to\infty}\beta(d) \le \lim_{d\to\infty} 12^{1/(2d)}6^{-1/2} \\ = \frac{1}{\sqrt{6}} < 0.40825 \end{aligned} \tag{9}$$

and

$$\begin{aligned} 0.24197 < \frac{1}{\sqrt{2\pi e}} \le \lim_{d\to\infty}\alpha(d) \\ \le \frac{2(3-\sqrt{3})\theta}{\sqrt{2\pi e}} < 0.4052, \end{aligned} \tag{10}$$

where

$$\tfrac{1}{2} \le \theta = \lim_{d\to\infty}[\theta(d)]^{1/d} \le 0.6602, \tag{11}$$

and $\theta(d)$ is the best SPHERE PACKING density in d-D space (Goddyn 1990, Moran 1984, Kabatyanskii and Levenshtein 1978). Steele and Snyder (1989) proved that the limit $\alpha(d)$ exists.

Now consider the constant

$$\kappa \equiv \lim_{n\to\infty}\frac{L(n,2)}{\sqrt{n}} = \beta(2)\sqrt{2}, \tag{12}$$

so

$$\tfrac{5}{8} = \gamma_2\sqrt{2} \le \kappa \le \delta\sqrt{2} < 0.9204. \tag{13}$$

The best current estimate is $\kappa \approx 0.7124$.

A certain self-avoiding SPACE-FILLING CURVE is an optimal TOUR through a set of n points, where n can be arbitrarily large. It has length

$$\lambda \equiv \lim_{m\to\infty}\frac{L_m}{\sqrt{n_m}} = \frac{4(1+2\sqrt{2})\sqrt{51}}{153} = 0.7147827\ldots, \tag{14}$$

where L_m is the length of the curve at the mth iteration and n_m is the point-set size (Moscato and Norman).

References
Beardwood, J.; Halton, J. H.; and Hammersley, J. M. "The Shortest Path Through Many Points." *Proc. Cambridge Phil. Soc.* **55**, 299–327, 1959.
Chartrand, G. "The Salesman's Problem: An Introduction to Hamiltonian Graphs." §3.2 in *Introductory Graph Theory.* New York: Dover, pp. 67–76, 1985.
Fejes Tóth, L. "Über einen geometrischen Satz." *Math. Zeit.* **46**, 83–85, 1940.
Few, L. "The Shortest Path and the Shortest Road Through *n* Points." *Mathematika* **2**, 141–144, 1955.
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/sales/sales.html`.
Flood, M. "The Travelling Salesman Problem." *Operations Res.* **4**, 61–75, 1956.
Goddyn, L. A. "Quantizers and the Worst Case Euclidean Traveling Salesman Problem." *J. Combin. Th. Ser. B* **50**, 65–81, 1990.
Kabatyanskii, G. A. and Levenshtein, V. I. "Bounds for Packing on a Sphere and in Space." *Problems Inform. Transm.* **14**, 1–17, 1978.
Karloff, H. J. "How Long Can a Euclidean Traveling Salesman Tour Be?" *SIAM J. Disc. Math.* **2**, 91–99, 1989.
Moran, S. "On the Length of Optimal TSP Circuits in Sets of Bounded Diameter." *J. Combin. Th. Ser. B* **37**, 113–141, 1984.
Moscato, P. "Fractal Instances of the Traveling Salesman Constant." `http://www.ing.unlp.edu.ar/cetad/mos/FRACTAL_TSP_home.html`
Steele, J. M. and Snyder, T. L. "Worst-Case Growth Rates of Some Classical Problems of Combinatorial Optimization." *SIAM J. Comput.* **18**, 278–287, 1989.
Verblunsky, S. "On the Shortest Path Through a Number of Points." *Proc. Amer. Math. Soc.* **2**, 904–913, 1951.

Traveling Salesman Problem

A problem in GRAPH THEORY requiring the most efficient (i.e., least total distance) TOUR (i.e., closed path) a salesman can take through each of n cities. No general method of solution is known, and the problem is NP-HARD.

see also TRAVELING SALESMAN CONSTANTS

References
Platzman, L. K. and Bartholdi, J. J. "Spacefilling Curves and the Planar Travelling Salesman Problem." *J. Assoc. Comput. Mach.* **46**, 719–737, 1989.

Trawler Problem

A fast boat is overtaking a slower one when fog suddenly sets in. At this point, the boat being pursued changes course, but not speed. How should the pursuing vessel proceed in order to be sure of catching the other boat?

The amazing answer is that the pursuing boat should continue to the point where the slow boat would be if it had set its course directly for the pursuing boat when the fog set in. If the boat is not there, it should proceed in a SPIRAL whose origin is the point where the slow boat was when the fog set in. The SPIRAL can be constructed in such a way that the two boats will intersect before a complete turn is made.

References
Ogilvy, C. S. *Excursions in Mathematics.* New York: Dover, pp. 84 and 148, 1994.

Trebly Magic Square

see TRIMAGIC SQUARE

Tredecillion

In the American system, 10^{42}.

see also LARGE NUMBER

Tree

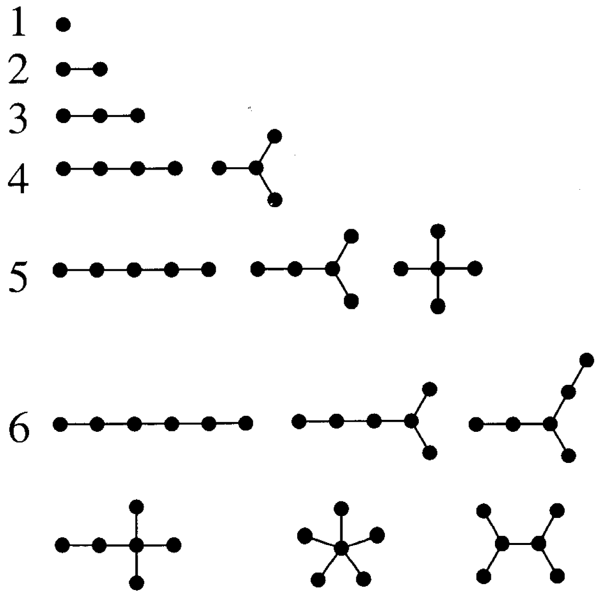

A tree is a mathematical structure which can be viewed as either a GRAPH or as a DATA STRUCTURE. The two views are equivalent, since a tree DATA STRUCTURE contains not only a set of elements, but also connections between elements, giving a tree graph.

A tree graph is a set of straight line segments connected at their ends containing no closed loops (cycles). A tree with n nodes has $n-1$ EDGES. The points of connection are known as FORKS and the segments as BRANCHES. Final segments and the nodes at their ends are called LEAVES. A tree with two BRANCHES at each FORK and with one or two LEAVES at the end of each branch is called a BINARY TREE.

When a special node is designated to turn a tree into a ROOTED TREE, it is called the ROOT (or sometimes "EVE.") In such a tree, each of the nodes which is one EDGE further away from a given EDGE is called a CHILD, and nodes connected to the same node are then called SIBLINGS.

Note that two BRANCHES placed end-to-end are equivalent to a single BRANCH which means, for example, that there is only *one* tree of order 3. The number $t(n)$ of nonisomorphic trees of order $n = 1, 2, \ldots$ (where trees of orders 1, 2, ..., 6 are illustrated above), are 1, 1, 1, 2, 3, 6, 11, 23, 47, 106, 235, ... (Sloane's A000055).

Otter showed that

$$\lim_{n\to\infty} \frac{t(n)n^{5/2}}{\alpha^n} = \beta, \qquad (1)$$

(Otter 1948, Harary and Palmer 1973, Knuth 1969), where the constants α and β are sometimes called OTTER'S TREE ENUMERATION CONSTANTS. Write the GENERATING FUNCTION for ROOTED TREES as

$$f(z) = \sum_{i=0}^{\infty} f_i z^i, \qquad (2)$$

where the COEFFICIENTS are

$$f_{i+1} = \frac{1}{i}\sum_{j=1}^{i}\left(\sum_{d|j} d f_d\right) f_{i-j+1}, \qquad (3)$$

with $f_0 = 0$ and $f_1 = 1$. Then

$$\alpha = 2.955765\ldots \qquad (4)$$

is the unique POSITIVE ROOT of

$$f\left(\frac{1}{x}\right) = 1, \qquad (5)$$

and

$$\beta = \frac{1}{\sqrt{2\pi}}\left[1 + \sum_{k=2}^{\infty} f'\left(\frac{1}{\alpha_k}\right)\frac{1}{\alpha_k}\right]^{3/2} = 0.5349485\ldots. \qquad (6)$$

see also B-TREE, BINARY TREE, CATERPILLAR GRAPH, CAYLEY TREE, CHILD, DIJKSTRA TREE, EVE, FOREST, KRUSKAL'S ALGORITHM, KRUSKAL'S TREE THEOREM, LEAF (TREE), ORCHARD-PLANTING PROBLEM, ORDERED TREE, PATH GRAPH, PLANTED PLANAR TREE, PÓLYA ENUMERATION THEOREM, QUADTREE, RED-BLACK TREE, ROOT (TREE), ROOTED TREE, SIBLING, STAR GRAPH, STERN-BROCOT TREE, WEAKLY BINARY TREE, WEIGHTED TREE

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/otter/otter.html`.

Chauvin, B.; Cohen, S.; and Rouault, A. (Eds.). *Trees: Workshop in Versailles, June 14–16, 1995.* Basel, Switzerland: Birkhäuser, 1996.

Gardner, M. "Trees." Ch. 17 in *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, pp. 240–250, 1978.

Harary, F. *Graph Theory.* Reading, MA: Addison-Wesley, 1994.

Harary, F. and Manvel, B. "Trees." *Scripta Math.* **28**, 327–333, 1970.

Harary, F. and Palmer, E. M. *Graphical Enumeration.* New York: Academic Press, 1973.

Knuth, D. E. *The Art of Computer Programming, Vol. 1: Fundamental Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, 1973.

Otter, R. "The Number of Trees." *Ann. Math.* **49**, 583–599, 1948.

Sloane, N. J. A. Sequences A000055/M0791 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Sloane, N. J. A. and Plouffe, S. Extended entry in *The Encyclopedia of Integer Sequences.* San Diego: Academic Press, 1995.

Tree-Planting Problem

see ORCHARD-PLANTING PROBLEM

Tree Searching

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

In database structures, two quantities are generally of interest: the average number of comparisons required to

1. Find an existing random record, and
2. Insert a new random record into a data structure.

Some constants which arise in the theory of digital tree searching are

$$\alpha \equiv \sum_{k=1}^{\infty} \frac{1}{2^k - 1} = 1.6066951524\ldots \qquad (1)$$

$$\beta \equiv \sum_{n=1}^{\infty} \frac{1}{(2^n - 1)^2} = 1.1373387363\ldots. \qquad (2)$$

Erdős (1948) proved that α is IRRATIONAL. The expected number of comparisons for a successful search is

$$\begin{aligned} E &= \frac{\ln n}{\ln 2} + \frac{\gamma - 1}{\ln 2} - \alpha + \tfrac{3}{2} + \delta(n) + \mathcal{O}(n^{-1/2}) \qquad (3)\\ &\sim \lg n - 0.716644\ldots + \delta(n), \qquad (4) \end{aligned}$$

and for an unsuccessful search is

$$\begin{aligned} E &= \frac{\ln n}{\ln 2} + \frac{\gamma}{\ln 2} - \alpha + \tfrac{1}{2} + \delta(n) + \mathcal{O}(n^{-1/2}) \qquad (5)\\ &\sim \lg n - 0.273948\ldots + \delta(n). \qquad (6) \end{aligned}$$

Here $\delta(n)$, $\epsilon(s)$, and $\rho(n)$ are small-amplitude periodic functions, and LG is the base 2 LOGARITHM. The VARIANCE for searching is

$$V \sim \frac{1}{12} + \frac{\pi^2 + 6}{6(\ln 2)^2} - \alpha - \beta + \epsilon(s) \sim 2.844383\ldots + \epsilon(s) \qquad (7)$$

and for inserting is

$$V \sim \frac{1}{12} + \frac{\pi^2}{6(\ln 2)^2} - \alpha - \beta + \epsilon(s) \sim 0.763014\ldots + \epsilon(s). \qquad (8)$$

The expected number of pairs of twin vacancies in a digital search tree is

$$\langle A_n\rangle = \left[\theta + 1 - \frac{1}{Q}\left(\frac{1}{\ln 2} + \alpha^2 - \alpha\right) + \rho(n)\right] n + \mathcal{O}(\sqrt{n}), \tag{9}$$

where

$$Q \equiv \prod_{k=1}^{\infty}\left(1 - \frac{1}{2^k}\right) = 0.2887880950\ldots \tag{10}$$

$$= \frac{1}{3} - \frac{1}{3\cdot 7} + \frac{1}{3\cdot 5\cdot 15} - \frac{1}{3\cdot 5\cdot 15\cdot 21} + \ldots \tag{11}$$

$$= \exp\left[-\sum_{n=1}^{\infty}\frac{1}{n(2^n - 1)}\right] \tag{12}$$

$$= \sqrt{\frac{2\pi}{\ln 2}}\exp\left(\frac{\ln 2}{24} - \frac{\pi^2}{6\ln 2}\right) \times \prod_{n=1}^{\infty}\left[1 - \exp\left(-\frac{4\pi^2 n}{\ln 2}\right)\right] \tag{13}$$

and

$$\theta = \sum_{k=1}^{\infty}\frac{k2^{k+1}}{1\cdot 3\cdot 7\cdot 16\cdots(2^k - 1)}\sum_{j=1}^{k}\frac{1}{2^j - 1} = 7.7431319855\ldots. \tag{14}$$

(Flajolet and Sedgewick 1986). The linear COEFFICIENT of $\langle A_n\rangle$ fluctuates around

$$c = \theta + 1 - \frac{1}{Q}\left(\frac{1}{\ln 2} + \alpha^2 - \alpha\right) = 0.3720486812\ldots, \tag{15}$$

which can also be written

$$c = \frac{1}{\ln 2}\int_0^{\infty}\frac{x}{1+x} \times \frac{dx}{(1+x)(1+\frac{1}{2}x)(1+\frac{1}{4}x)(1+\frac{1}{8}x)\cdots}. \tag{16}$$

(Flajolet and Richmond 1992).

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/bin/bin.html`.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/dig/dig.html`.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/qdt/qdt.html`.

Flajolet, P. and Richmond, B. "Generalized Digital Trees and their Difference-Differential Equations." *Random Structures and Algorithms* **3**, 305–320, 1992.

Flajolet, P. and Sedgewick, R. "Digital Search Trees Revisited." *SIAM Review* **15**, 748–767, 1986.

Knuth, D. E. *The Art of Computer Programming, Vol. 3: Sorting and Searching, 2nd ed.* Reading, MA: Addison-Wesley, pp. 21, 134, 156, 493–499, and 580, 1973.

Trefoil Curve

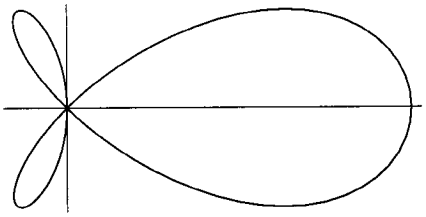

The plane curve given by the equation

$$x^4 + x^2y^2 + y^4 = x(x^2 - y^2).$$

Trefoil Knot

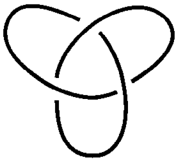

The knot 03_{001}, also called the THREEFOIL KNOT, which is the unique PRIME KNOT of three crossings. It has BRAID WORD ${\sigma_1}^3$. The trefoil and its MIRROR IMAGE are not equivalent. The trefoil has ALEXANDER POLYNOMIAL $-x^2 + x - 1$ and is a (3, 2)-TORUS KNOT. The BRACKET POLYNOMIAL can be computed as follows.

$$\begin{aligned}\langle L\rangle &= A^3d^{2-1} + A^2Bd^{1-1} + A^2Bd^{1-1} + AB^2d^{2-1} \\ &\quad + A^2Bd^{1-1} + AB^2d^{2-1} + AB^2d^{2-1} + B^3d^{3-1} \\ &= A^3d^1 + 3A^2Bd^0 + 3AB^2d^1 + B^3d^2.\end{aligned}$$

Plugging in

$$\begin{aligned}B &= A^{-1} \\ d &= -A^2 - A^{-2}\end{aligned}$$

gives

$$\langle L\rangle = A^{-7} - A^{-3} - A^5.$$

The normalized one-variable KAUFFMAN POLYNOMIAL X is then given by

$$\begin{aligned}X_L &= (-A^3)^{-w(L)}\langle L\rangle = (-A^3)^{-3}(A^{-7} - A^{-3} - A^5) \\ &= A^{-4} + A^{-12} - A^{-16},\end{aligned}$$

where the WRITHE $w(L) = 3$. The JONES POLYNOMIAL is therefore

$$V(t) = L(A = t^{-1/4}) = t + t^3 - t^4 = t(1 + t^2 - t^3).$$

Since $V(t^{-1}) \neq V(t)$, we have shown that the mirror images are not equivalent.

References

Claremont High School. "Trefoil_Knot Movie." Binary encoded QuickTime movie. `ftp://chs.cusd.claremont.edu/pub/knot/trefoil.cptbin`.

Crandall, R. E. *Mathematica for the Sciences.* Redwood City, CA: Addison-Wesley, 1993.

Kauffman, L. H. *Knots and Physics.* Singapore: World Scientific, pp. 29–35, 1991.

Nordstrand, T. "Threefoil Knot." `http://www.uib.no/people/nfytn/tknottxt.htm`.

Pappas, T. "The Trefoil Knot." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 96, 1989.

Trench Diggers' Constant

see BEAM DETECTOR

Triabolo

A 3-POLYABOLO.

Triacontagon

A 30-sided POLYGON.

Triacontahedron

A 30-sided POLYHEDRON such as the RHOMBIC TRIACONTAHEDRON.

Triad

A SET with three elements.

see also HEXAD, MONAD, QUARTET, QUINTET, TETRAD

Triakis Icosahedron

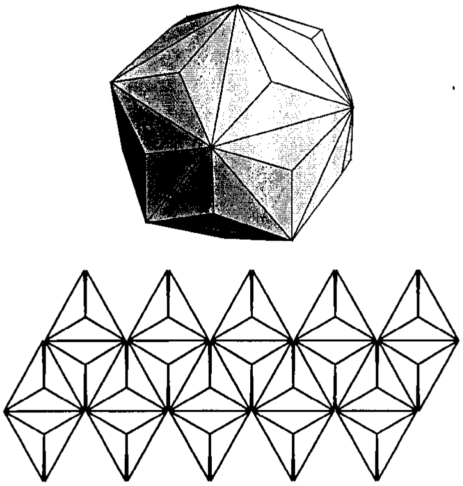

The DUAL POLYHEDRON of the TRUNCATED DODECAHEDRON ARCHIMEDEAN SOLID. The triakis icosahedron is also ICOSAHEDRON STELLATION #2.

References

Wenninger, M. J. *Polyhedron Models.* New York: Cambridge University Press, p. 46, 1989.

Triakis Octahedron

see GREAT TRIAKIS OCTAHEDRON, SMALL TRIAKIS OCTAHEDRON

Triakis Tetrahedron

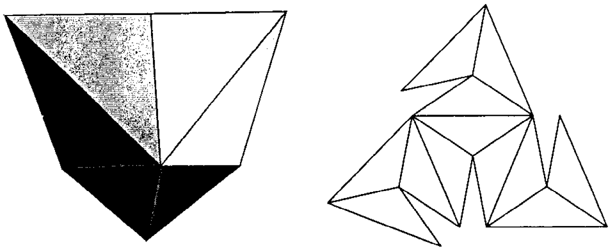

The DUAL POLYHEDRON of the TRUNCATED TETRAHEDRON ARCHIMEDEAN SOLID.

Trial

In statistics, a trial is a single measurable random event, such as the flipping of a COIN, the generation of a RANDOM NUMBER, the dropping of a ball down the apex of a triangular lattice and having it fall into a single bin at the bottom, etc.

see also BERNOULLI TRIAL, LEXIS TRIALS, POISSON TRIALS

Trial Division

A brute-force method of finding a DIVISOR of an INTEGER n by simply plugging in one or a set of INTEGERS and seeing if they DIVIDE n. Repeated application of trial division to obtain the complete PRIME FACTORIZATION of a number is called DIRECT SEARCH FACTORIZATION. An individual integer being tested is called a TRIAL DIVISOR.

see also DIRECT SEARCH FACTORIZATION, DIVISION, PRIME FACTORIZATION

Trial Divisor

An INTEGER n which is tested to see if it divides a given number.

see also TRIAL DIVISION

Triamond

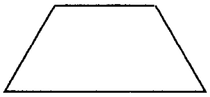

The unique 3-POLYIAMOND, illustrated above.

see also POLYIAMOND, TRAPEZOID

Triangle

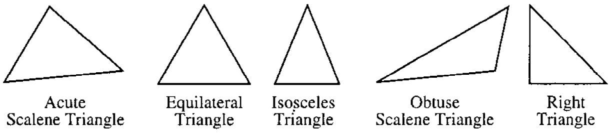

A triangle is a 3-sided POLYGON sometimes (but not very commonly) called the TRIGON. All triangles are convex. An ACUTE TRIANGLE is a triangle whose three angles are all ACUTE. A triangle with all sides equal is called EQUILATERAL. A triangle with two sides equal is called ISOSCELES. A triangle having an OBTUSE ANGLE is called an OBTUSE TRIANGLE. A triangle with a RIGHT ANGLE is called RIGHT. A triangle with all sides a different length is called SCALENE.

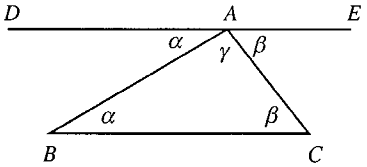

The sum of ANGLES in a triangle is 180°. This can be established as follows. Let $DAE||BC$ (DAE be PARALLEL to BC) in the above diagram, then the angles α and β

satisfy $\alpha = \angle DAB = \angle ABC$ and $\beta = \angle EAC = \angle BCE$, as indicated. Adding γ, it follows that

$$\alpha + \beta + \gamma = 180^\circ, \tag{1}$$

since the sum of angles for the line segment must equal two RIGHT ANGLES. Therefore, the sum of angles in the triangle is also 180°.

Let S stand for a triangle side and A for an angle, and let a set of Ss and As be concatenated such that adjacent letters correspond to adjacent sides and angles in a triangle. Triangles are uniquely determined by specifying three sides (SSS THEOREM), two angles and a side (AAS THEOREM), or two sides with an adjacent angle (SAS THEOREM). In each of these cases, the unknown three quantities (there are three sides and three angles total) can be uniquely determined. Other combinations of sides and angles do not uniquely determine a triangle: three angles specify a triangle only modulo a scale size (AAA THEOREM), and one angle and two sides not containing it may specify one, two, or no triangles (ASS THEOREM).

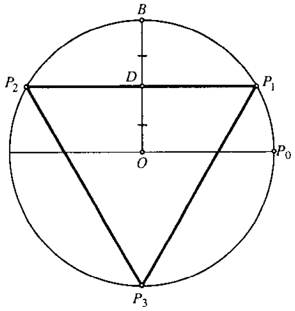

Equilateral Triangle

The RULER and COMPASS construction of the triangle can be accomplished as follows. In the above figure, take OP_0 as a RADIUS and draw $OB \perp OP_0$. Then bisect OB and construct $P_2P_1 \| OP_0$. Extending BO to locate P_3 then gives the EQUILATERAL TRIANGLE $\Delta P_1P_2P_3$.

In Proposition IV.4 of the *Elements*, Euclid showed how to inscribe a CIRCLE (the INCIRCLE) in a given triangle by locating the CENTER as the point of intersection of ANGLE BISECTORS. In Proposition IV.5, he showed how to circumscribe a CIRCLE (the CIRCUMCIRCLE) about a given triangle by locating the CENTER as the point of intersection of the perpendicular bisectors.

If the coordinates of the triangle VERTICES are given by (x_i, y_i) where $i = 1, 2, 3$, then the AREA Δ is given by the DETERMINANT

$$\Delta = \frac{1}{2!}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}. \tag{2}$$

If the coordinates of the triangle VERTICES are given in 3-D by (x_i, y_i, z_i) where $i = 1, 2, 3$, then

$$\Delta = \frac{1}{2}\sqrt{\begin{vmatrix} y_1 & z_1 & 1 \\ y_2 & z_2 & 1 \\ y_3 & z_3 & 1 \end{vmatrix}^2 + \begin{vmatrix} z_1 & x_1 & 1 \\ z_2 & x_2 & 1 \\ z_3 & x_3 & 1 \end{vmatrix}^2 + \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}^2}. \tag{3}$$

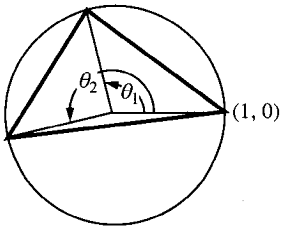

In the above figure, let the CIRCUMCIRCLE passing through a triangle's VERTICES have RADIUS r, and denote the CENTRAL ANGLES from the first point to the second θ_1, and to the third point by θ_2. Then the AREA of the triangle is given by

$$\Delta = 2r^2 \left|\sin(\tfrac{1}{2}\theta_1)\sin(\tfrac{1}{2}\theta_2)\sin[\tfrac{1}{2}(\theta_1 - \theta_2)]\right|. \tag{4}$$

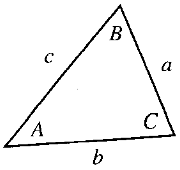

If a triangle has sides a, b, c, call the angles opposite these sides A, B, and C, respectively. Also define the SEMIPERIMETER s as HALF the PERIMETER:

$$s \equiv \tfrac{1}{2}p = \tfrac{1}{2}(a + b + c). \tag{5}$$

The AREA of a triangle is then given by HERON'S FORMULA

$$\Delta = \sqrt{s(s-a)(s-b)(s-c)}, \tag{6}$$

as well by the FORMULAS

$$\Delta = \tfrac{1}{4}\sqrt{(a+b+c)(b+c-a)(c+a-b)(a+b-c)} \tag{7}$$

$$= \tfrac{1}{4}\sqrt{2(a^2b^2 + a^2c^2 + b^2c^2) - (a^4 + b^4 + c^4)} \tag{8}$$

$$= \tfrac{1}{4}\sqrt{[(a+b)^2 - c^2][c^2 - (a-b)^2]} \tag{9}$$

$$= \tfrac{1}{4}\sqrt{p(p-2a)(p-2b)(p-2c)}, \tag{10}$$

$$= 2R^2 \sin A \sin B \sin C \tag{11}$$

$$= \frac{abc}{4R} = rs \tag{12}$$

$$= \tfrac{1}{2}ah_a \tag{13}$$

$$= \tfrac{1}{2}bc\sin A. \tag{14}$$

In the above formulas, h_i is the Altitude on side i, R is the Circumradius, and r is the Inradius (Johnson 1929, p. 11). Expressing the side lengths a, b, and c in terms of the radii a', b', and c' of the mutually tangent circles centered on the Triangle vertices (which define the Soddy Circles),

$$a = b' + c' \tag{15}$$

$$b = a' + c' \tag{16}$$

$$c = a' + b', \tag{17}$$

gives the particularly pretty form

$$\Delta = \sqrt{a'b'c'(a' + b' + c')}. \tag{18}$$

For additional Formulas, see Beyer (1987) and Baker (1884), who gives 110 Formulas for the Area of a triangle.

The Angles of a triangle satisfy

$$\cot A = \frac{b^2 + c^2 - a^2}{4\Delta} \tag{19}$$

where Δ is the Area (Johnson 1929, p. 11, with missing squared symbol added). This gives the pretty identity

$$\cot A + \cot B + \cot C = \frac{a^2 + b^2 + c^2}{4\Delta}. \tag{20}$$

Let a triangle have Angles A, B, and C. Then

$$\sin A \sin B \sin C \leq kABC, \tag{21}$$

where

$$k = \left(\frac{3\sqrt{3}}{2\pi}\right)^3 \tag{22}$$

(Abi-Khuzam 1974, Le Lionnais 1983). This can be used to prove that

$$8\omega^3 < ABC, \tag{23}$$

where ω is the Brocard Angle.

Trigonometric Functions of half angles can be expressed in terms of the triangle sides:

$$\cos(\tfrac{1}{2}A) = \sqrt{\frac{s(s-a)}{bc}} \tag{24}$$

$$\sin(\tfrac{1}{2}A) = \sqrt{\frac{(s-b)(s-c)}{bc}} \tag{25}$$

$$\tan(\tfrac{1}{2}A) = \sqrt{\frac{(s-b)(s-c)}{s(s-a)}}, \tag{26}$$

where s is the Semiperimeter.

The number of different triangles which have Integral sides and Perimeter n is

$$\begin{aligned} T(n) &= P_3(n) - \sum_{1 \leq j \leq \lfloor n/2 \rfloor} P_2(j) \\ &= \left[\frac{n^2}{12}\right] - \left\lfloor \frac{n}{4} \right\rfloor \left\lfloor \frac{n+2}{4} \right\rfloor \\ &= \begin{cases} \left[\frac{n^2}{48}\right] & \text{for } n \text{ even} \\ \left[\frac{(n+3)^2}{48}\right] & \text{for } n \text{ odd,} \end{cases} \end{aligned} \tag{27}$$

where P_2 and P_3 are Partition Functions P, $[x]$ is the Nint function, and $\lfloor x \rfloor$ is the Floor Function (Jordan *et al.* 1979, Andrews 1979, Honsberger 1985). The values of $T(n)$ for $n = 1, 2, \ldots$ are 0, 0, 1, 0, 1, 1, 2, 1, 3, 2, 4, 3, 5, 4, 7, 5, 8, 7, 10, 8, 12, 10, 14, 12, 16, ... (Sloane's A005044), which is also Alcuin's Sequence padded with two initial 0s. $T(n)$ also satisfies

$$T(2n) = T(2n-3) = P_3(n). \tag{28}$$

It is not known if a triangle with Integer sides, Medians, and Area exists (although there are incorrect Proofs of the impossibility in the literature). However, R. L. Rathbun, A. Kemnitz, and R. H. Buchholz have shown that there are infinitely many triangles with Rational sides (Heronian Triangles) with *two* Rational Medians (Guy 1994).

In the following paragraph, assume the specified sides and angles are adjacent to each other. Specifying three Angles does not uniquely define a triangle, but any two triangles with the same Angles are similar (the AAA Theorem). Specifying two Angles A and B and a side a uniquely determines a triangle with Area

$$\Delta = \frac{a^2 \sin B \sin C}{2 \sin A} = \frac{a^2 \sin B \sin(\pi - A - B)}{2 \sin A} \tag{29}$$

(the AAS Theorem). Specifying an Angle A, a side c, and an Angle B uniquely specifies a triangle with Area

$$\Delta = \frac{c^2}{2(\cot A + \cot B)} \tag{30}$$

(the ASA Theorem). Given a triangle with two sides, a the smaller and c the larger, and one known Angle A, Acute and opposite a, if $\sin A < a/c$, there are two possible triangles. If $\sin A = a/c$, there is one possible triangle. If $\sin A > a/c$, there are no possible triangles. This is the ASS Theorem. Let a be the base length and h be the height. Then

$$\Delta = \tfrac{1}{2}ah = \tfrac{1}{2}ac \sin B \tag{31}$$

(the SAS Theorem). Finally, if all three sides are specified, a unique triangle is determined with Area given by Heron's Formula or by

$$\Delta = \frac{abc}{4R}, \tag{32}$$

where R is the Circumradius. This is the SSS Theorem.

There are four Circles which are tangent to the sides of a triangle, one internal and the rest external. Their centers are the points of intersection of the Angle Bisectors of the triangle.

Any triangle can be positioned such that its shadow under an orthogonal projection is Equilateral.

see also AAA Theorem, AAS Theorem, Acute Triangle, Alcuin's Sequence, Altitude, Angle Bisector, Anticevian Triangle, Anticomplementary Triangle, Antipedal Triangle, ASS Theorem, Bell Triangle, Brianchon Point, Brocard Angle, Brocard Circle, Brocard Midpoint, Brocard Points, Butterfly Theorem, Centroid (Triangle), Ceva's Theorem, Cevian, Cevian Triangle, Chasles's Theorem, Circumcenter, Circumcircle, Circumradius, Contact Triangle, Crossed Ladders Problem, Crucial Point, D-Triangle, de Longchamps Point, Desargues' Theorem, Dissection, Elkies Point, Equal Detour Point, Equilateral Triangle, Euler Line, Euler's Triangle, Euler Triangle Formula, Excenter, Excentral Triangle, Excircle, Exeter Point, Exmedian, Exmedian Point, Exradius, Exterior Angle Theorem, Fagnano's Problem, Far-Out Point, Fermat Point, Fermat's Problem, Feuerbach Point, Feuerbach's Theorem, Fuhrmann Triangle, Gergonne Point, Grebe Point, Griffiths Points, Griffiths' Theorem, Harmonic Conjugate Points, Heilbronn Triangle Problem, Heron's Formula, Heronian Triangle, Hofstadter Triangle, Homothetic Triangles, Incenter, Incircle, Inradius, Isodynamic Points, Isogonal Conjugate, Isogonic Centers, Isoperimetric Point, Isosceles Triangle, Kabon Triangles, Kanizsa Triangle, Kiepert's Hyperbola, Kiepert's Parabola, Law of Cosines, Law of Sines, Law of Tangents, Leibniz Harmonic Triangle, Lemoine Circle, Lemoine Point, Line at Infinity, Malfatti Points, Medial Triangle, Median (Triangle), Median Triangle, Menelaus' Theorem, Mid-Arc Points, Mittenpunkt, Mollweide's Formulas, Morley Centers, Morley's Theorem, Nagel Point, Napoleon's Theorem, Napoleon Triangles, Newton's Formulas, Nine-Point Circle, Number Triangle, Obtuse Triangle, Orthic Triangle, Orthocenter, Orthologic, Paralogic Triangles, Pascal's Triangle, Pasch's Axiom, Pedal Triangle, Perpendicular Bisector, Perspective Triangles, Petersen-Shoute Theorem, Pivot Theorem, Power Point, Power (Triangle), Prime Triangle, Purser's Theorem, Quadrilateral, Rational Triangle, Routh's Theorem, SAS Theorem, Scalene Triangle, Schiffler Point, Schwarz Triangle, Schwarz's Triangle Problem, Seidel-Entringer-Arnold Triangle, Seydewitz's Theorem, Simson Line, Spieker Center, SSS Theorem, Steiner-Lehmus Theorem, Steiner Points, Stewart's Theorem, Symmedian Point, Tangential Triangle, Tangential Triangle Circumcenter, Tarry Point, Thomsen's Figure, Torricelli Point, Triangle Tiling, Triangle Transformation Principle, Yff Points, Yff Triangles

References

Abi-Khuzam, F. "Proof of Yff's Conjecture on the Brocard Angle of a Triangle." *Elem. Math.* **29**, 141–142, 1974.

Andrews, G. "A Note on Partitions and Triangles with Integer Sides." *Amer. Math. Monthly* **86**, 477, 1979.

Baker, M. "A Collection of Formulæ for the Area of a Plane Triangle." *Ann. Math.* **1**, 134–138, 1884.

Berkhan, G. and Meyer, W. F. "Neuere Dreiecksgeometrie." In *Encyklopaedie der Mathematischen Wissenschaften, Vol. 3AB 10* (Ed. F. Klein). Leipzig: Teubner, pp. 1173–1276, 1914.

Beyer, W. H. (Ed.) *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 123–124, 1987.

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, 1969.

Davis, P. "The Rise, Fall, and Possible Transfiguration of Triangle Geometry: A Mini-History." *Amer. Math. Monthly* **102**, 204–214, 1995.

Eppstein, D. "Triangles and Simplices." `http://www.ics.uci.edu/~eppstein/junkyard/triangulation.html`.

Feuerbach, K. W. *Eigenschaften einiger merkwürdingen Punkte des geradlinigen Dreiecks, und mehrerer durch die bestimmten Linien und Figuren.* Nürnberg, Germany, 1822.

Guy, R. K. "Triangles with Integer Sides, Medians, and Area." §D21 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 188–190, 1994.

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 39–47, 1985.

Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, 1929.

Jordan, J. H.; Walch, R.; and Wisner, R. J. "Triangles with Integer Sides." *Amer. Math. Monthly* **86**, 686–689, 1979.

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Kimberling, C. "Triangle Centers and Central Triangles." *Congr. Numer.* **129**, 1–295, 1998.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 28, 1983.

Schroeder. *Das Dreieck und seine Beruhungskreise.*

Sloane, N. J. A. Sequence A005044/M0146 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Vandeghen, A. "Some Remarks on the Isogonal and Cevian Transforms. Alignments of Remarkable Points of a Triangle." *Amer. Math. Monthly* **72**, 1091–1094, 1965.

Weisstein, E. W. "Plane Geometry." `http://www.astro.virginia.edu/~eww6n/math/notebooks/PlaneGeometry.m`.

Triangle Arcs

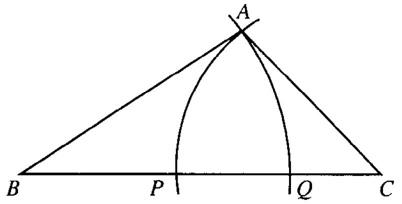

In the above figure, the curves are arcs of a CIRCLE and

$$a = BC \tag{1}$$
$$b = CA = CP \tag{2}$$
$$c = BA = BQ. \tag{3}$$

Then

$$PQ^2 = 2BP \cdot QC. \tag{4}$$

The figure also yields the algebraic identity

$$(b + c - \sqrt{b^2 + c^2}\,)^2 = 2(\sqrt{b^2 + c^2} - b)(\sqrt{b^2 + c^2} - c). \tag{5}$$

see also ARC

References

Berndt, B. C. *Ramanujan's Notebooks, Part IV.* New York: Springer-Verlag, pp. 8–9, 1994.

Dharmarajan, T. and Srinivasan, P. K. *An Introduction to Creativity of Ramanujan, Part III.* Madras: Assoc. Math. Teachers, pp. 11–13, 1987.

Triangle Center

A triangle center is a point whose TRILINEAR COORDINATES are defined in terms of the side lengths and angles of a TRIANGLE. The function giving the coordinates $\alpha : \beta : \gamma$ is called the TRIANGLE CENTER FUNCTION. The four ancient centers are the CENTROID, INCENTER, CIRCUMCENTER, and ORTHOCENTER. For a listing of these and other triangle centers, see Kimberling (1994).

A triangle center is said to be REGULAR IFF there is a TRIANGLE CENTER FUNCTION which is a POLYNOMIAL in Δ, a, b, and c (where Δ is the AREA of the TRIANGLE) such that the TRILINEAR COORDINATES of the center are

$$f(a, b, c) : f(b, c, a) : f(c, a, b).$$

A triangle center is said to be a MAJOR TRIANGLE CENTER if the TRIANGLE CENTER FUNCTION α is a function of ANGLE A alone, and therefore β and γ of B and C alone, respectively.

see also MAJOR TRIANGLE CENTER, REGULAR TRIANGLE CENTER, TRIANGLE, TRIANGLE CENTER FUNCTION, TRILINEAR COORDINATES, TRILINEAR POLAR

References

Davis, P. J. "The Rise, Fall, and Possible Transfiguration of Triangle Geometry: A Mini-History." *Amer. Math. Monthly* **102**, 204–214, 1995.

Dixon, R. "The Eight Centres of a Triangle." §1.5 in *Mathographics.* New York: Dover, pp. 55–61, 1991.

Gale, D. "From Euclid to Descartes to Mathematica to Oblivion?" *Math. Intell.* **14**, 68–69, 1992.

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–167, 1994.

Kimberling, C. "Triangle Centers and Central Triangles." *Congr. Numer.* **129**, 1–295, 1998.

Triangle Center Function

A HOMOGENEOUS FUNCTION $f(a, b, c)$, i.e., a function f such that

$$f(ta, tb, tc) = t^n f(a, b, c),$$

which gives the TRILINEAR COORDINATES of a TRIANGLE CENTER as

$$\alpha : \beta : \gamma = f(a, b, c) : f(b, c, a) : f(c, a, b).$$

The variables may correspond to angles (A, B, C) or side lengths (a, b, c), since these can be interconverted using the LAW OF COSINES.

see also MAJOR TRIANGLE CENTER, REGULAR TRIANGLE CENTER, TRIANGLE CENTER, TRILINEAR COORDINATES

References

Kimberling, C. "Triangle Centers as Functions." *Rocky Mtn. J. Math.* **23**, 1269–1286, 1993.

Kimberling, C. "Triangle Centers." `http://www.evansville.edu/~ck6/tcenters/`.

Kimberling, C. "Triangle Centers and Central Triangles." *Congr. Numer.* **129**, 1–295, 1998.

Triangle Coefficient

A function of three variables written $\Delta(abc) \equiv \Delta(a, b, c)$ and defined by

$$\Delta(abc) \equiv \sqrt{\frac{(a + b - c)!(a - b + c)!(-a + b + c)!}{(a + b + c + 1)!}}.$$

References

Shore, B. W. and Menzel, D. H. *Principles of Atomic Spectra.* New York: Wiley, p. 273, 1968.

Triangle Condition

The condition that j takes on the values

$$j = j_1 + j_2, j_1 + j_2 - 1, \ldots, |j_1 - j_2|,$$

denoted $\Delta(j_1 j_2 j)$.

References

Sobelman, I. I. *Atomic Spectra and Radiative Transitions, 2nd ed.* Berlin: Springer-Verlag, p. 60, 1992.

Triangle Counting

Given rods of length 1, 2, ..., n, how many distinct triangles $T(n)$ can be made? Lengths for which

$$l_i = l_j + l_k$$

obviously do not give triangles, but all other combinations of three rods do. The answer is

$$T(n) = \begin{cases} \frac{1}{24}n(n-2)(2n-5) & \text{for } n \text{ even} \\ \frac{1}{24}(n-1)(n-3)(2n-1) & \text{for } n \text{ odd.} \end{cases}$$

The values for $n = 1, 2, \ldots$ are 0, 0, 0, 1, 3, 7, 13, 22, 34, 50, ... (Sloane's A002623). Somewhat surprisingly, this sequence is also given by the GENERATING FUNCTION

$$f(x) = \frac{x^4}{(1-x)^3(1-x^2)} = x^4 + 3x^5 + 7x^6 + 13x^7 + \ldots.$$

References

Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., pp. 278–282, 1991.

Sloane, N. J. A. Sequence A002623/M2640 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Triangle of Figurate Numbers

see FIGURATE NUMBER TRIANGLE

Triangle Function

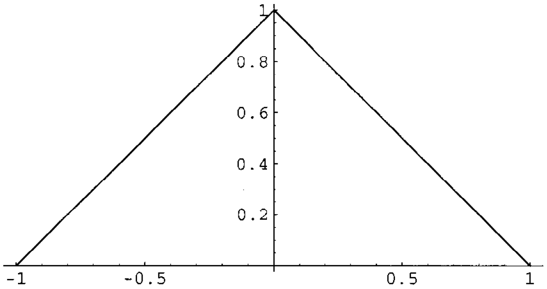

$$\begin{aligned} \Lambda(x) &\equiv \begin{cases} 0 & |x| > 1 \\ 1 - |x| & |x| < 1 \end{cases} & (1) \\ &= \Pi(x) * \Pi(x) & (2) \\ &= \Pi(x) * H(x + \tfrac{1}{2}) - \Pi(x) * H(x - \tfrac{1}{2}), & (3) \end{aligned}$$

where Π is the RECTANGLE FUNCTION and H is the HEAVISIDE STEP FUNCTION. An obvious generalization used as an APODIZATION FUNCTION goes by the name of the BARTLETT FUNCTION.

There is also a three-argument function known as the triangle function:

$$\lambda(x, y, z) \equiv x^2 + y^2 + z^2 - 2xy - 2xz - 2yz. \quad (4)$$

It follows that

$$\lambda(a^2, b^2, c^2) = (a+b+c)(a+b-c)(a-b+c)(a-b-c). \quad (5)$$

see also ABSOLUTE VALUE, BARTLETT FUNCTION, HEAVISIDE STEP FUNCTION, RAMP FUNCTION, SGN, TRIANGLE COEFFICIENT

Triangle Inequality

Let $\mathbf{x}$ and $\mathbf{y}$ be vectors

$$|\mathbf{x}| - |\mathbf{y}| \leq |\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|. \quad (1)$$

Equivalently, for COMPLEX NUMBERS z_1 and z_2,

$$|z_1| - |z_2| \leq |z_1 + z_2| \leq |z_1| + |z_2|. \quad (2)$$

A generalization is

$$\left| \sum_{k=1}^{n} a_k \right| \leq \sum_{k=1}^{n} |a_k|. \quad (3)$$

see also p-ADIC NUMBER, STRONG TRIANGLE INEQUALITY

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 11, 1972.

Triangle Inscribing in a Circle

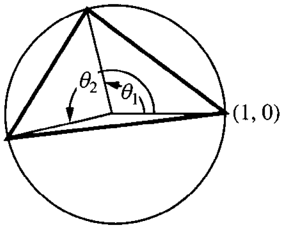

Select three points at random on a unit CIRCLE. Find the distribution of possible areas. The first point can be assigned coordinates (1, 0) without loss of generality. Call the central angles from the first point to the second and third θ_1 and θ_2. The range of θ_1 can be restricted to $[0, \pi]$ because of symmetry, but θ_2 can range from $[0, 2\pi)$. Then

$$A(\theta_1, \theta_2) = 2\left|\sin(\tfrac{1}{2}\theta_1)\sin(\tfrac{1}{2}\theta_2)\sin[\tfrac{1}{2}(\theta_1 - \theta_2)]\right|, \quad (1)$$

so

$$\bar{A} = \frac{\int_0^\pi \int_0^{2\pi} A(\theta_1, \theta_2)\, d\theta_2\, d\theta_1}{C}, \quad (2)$$

where

$$C \equiv \int_0^\pi \int_0^{2\pi} d\theta_2\, d\theta_1 = 2\pi^2. \quad (3)$$

Therefore,

$$\begin{aligned}
\bar{A} &= \frac{2}{2\pi^2}\int_0^{\pi}\int_0^{2\pi}\left|\sin(\tfrac{1}{2}\theta_1)\sin(\tfrac{1}{2}\theta_2)\sin[\tfrac{1}{2}(\theta_1-\theta_2)]\right|\,d\theta_2\,d\theta_1\\
&= \frac{1}{\pi^2}\int_0^{\pi}\sin(\tfrac{1}{2}\theta_1)\left[\int_0^{2\pi}\sin(\tfrac{1}{2}\theta_2)\left|\sin[\tfrac{1}{2}(\theta_2-\theta_1)]\right|\,d\theta_2\right]d\theta_1\\
&= \frac{1}{\pi^2}\underset{\theta_2-\theta_1>0}{\int_0^{\pi}\int_0^{2\pi}}\sin(\tfrac{1}{2}\theta_1)\sin(\tfrac{1}{2}\theta_2)\sin[\tfrac{1}{2}(\theta_1-\theta_2)]\,d\theta_2\,d\theta_1\\
&\quad+\frac{1}{\pi^2}\underset{\theta_2-\theta_1<0}{\int_0^{\pi}\int_0^{2\pi}}\sin(\tfrac{1}{2}\theta_1)\sin(\tfrac{1}{2}\theta_2)\sin[\tfrac{1}{2}(\theta_1-\theta_2)]\,d\theta_2\,d\theta_1\\
&= \frac{1}{\pi^2}\int_0^{\pi}\sin(\tfrac{1}{2}\theta_1)\left[\int_{\theta_1}^{2\pi}\sin(\tfrac{1}{2}\theta_2)\sin[\tfrac{1}{2}(\theta_2-\theta_1)]\,d\theta_2\right]d\theta_1\\
&\quad+\frac{1}{\pi^2}\int_0^{\pi}\sin(\tfrac{1}{2}\theta_1)\left[\int_0^{\theta_1}\sin(\tfrac{1}{2}\theta_2)\sin[\tfrac{1}{2}(\theta_2-\theta_1)]\,d\theta_2\right]d\theta_1.
\end{aligned}\tag{4}$$

But

$$\begin{aligned}
&\int(\tfrac{1}{2}\theta_2)\sin[\tfrac{1}{2}(\theta_2-\theta_1)]\,d\theta_2\\
&= \int\sin(\tfrac{1}{2}\theta_2)\left[\sin(\tfrac{1}{2}\theta_2)\cos(\tfrac{1}{2}\theta_2)-\sin(\tfrac{1}{2}\theta_1)\cos(\tfrac{1}{2}\theta_2)\right]\,d\theta_2\\
&= \cos(\tfrac{1}{2}\theta_1)\int\sin^2(\tfrac{1}{2}\theta_2)\,d\theta_2-\sin(\tfrac{1}{2}\theta_1)\int\sin(\tfrac{1}{2}\theta_1)\cos(\tfrac{1}{2}\theta_2)\,d\theta_2\\
&= \tfrac{1}{2}\cos(\tfrac{1}{2}\theta_1)\int(1-\cos\theta_2)\,d\theta_2-\tfrac{1}{2}\sin(\tfrac{1}{2}\theta_2)\int\sin\theta_2\,d\theta_2\\
&= \tfrac{1}{2}\cos(\tfrac{1}{2}\theta_1)(\theta_2-\sin\theta_2)+\tfrac{1}{2}\sin(\tfrac{1}{2}\theta_1)\cos(\theta_2).
\end{aligned}\tag{5}$$

Write (4) as

$$\bar{A}=\frac{1}{\pi^2}\left[\int_0^{\pi}\sin(\tfrac{1}{2}\theta_1)I_1\,d\theta_1+\int_0^{\pi}\sin(\tfrac{1}{2}\theta_1)I_2\,d\theta_1\right],\tag{6}$$

then

$$I_1\equiv\int_{\theta_1}^{2\pi}\sin(\tfrac{1}{2}\theta_2)\sin[\tfrac{1}{2}(\theta_2-\theta_1)]\,d\theta_2,\tag{7}$$

and

$$I_2\equiv\int_0^{\theta_1}\sin(\tfrac{1}{2}\theta_2)\sin[\tfrac{1}{2}(\theta_1-\theta_2)]\,d\theta_2.\tag{8}$$

From (6),

$$\begin{aligned}
I_1 &= \tfrac{1}{2}\cos(\tfrac{1}{2}\theta_2)[\theta_2-\sin\theta_2]_{\theta_1}^{2\pi}+\tfrac{1}{2}\sin(\tfrac{1}{2}\theta_1)[\cos\theta_2]_{\theta_1}^{2\pi}\\
&= \tfrac{1}{2}\cos(\tfrac{1}{2}\theta_1)(2\pi-\theta_1+\sin\theta_1)\\
&\quad+\tfrac{1}{2}\sin(\tfrac{1}{2}\theta_1)(1-\cos\theta_1)\\
&= \pi\cos(\tfrac{1}{2}\theta_1)-\tfrac{1}{2}\theta_1\cos(\tfrac{1}{2}\theta_1)+\tfrac{1}{2}[\cos(\tfrac{1}{2}\theta_1)\sin\theta_1\\
&\quad-\cos\theta_1\sin(\tfrac{1}{2}\theta_1)]+\tfrac{1}{2}\sin(\tfrac{1}{2}\theta_1)\\
&= \pi\cos(\tfrac{1}{2}\theta_1)-\tfrac{1}{2}\theta_1\cos(\tfrac{1}{2}\theta_1)+\tfrac{1}{2}+\tfrac{1}{2}\sin(\theta_1-\tfrac{1}{2}\theta_1)\\
&\quad+\tfrac{1}{2}\sin(\tfrac{1}{2}\theta_1)\\
&= \pi\cos(\tfrac{1}{2}\theta_1)-\tfrac{1}{2}\theta_1\cos(\tfrac{1}{2}\theta_1)+\sin(\tfrac{1}{2}\theta_1),
\end{aligned}\tag{9}$$

so

$$\int_0^{\pi}I_1\sin(\tfrac{1}{2}\theta_1)\,d\theta_1=\tfrac{5}{4}\pi.\tag{10}$$

Also,

$$\begin{aligned}
I_2 &= \tfrac{1}{2}\cos(\tfrac{1}{2}\theta_1)[\sin\theta_2-\theta_2]_0^{\theta_1}-\tfrac{1}{2}\sin(\tfrac{1}{2}\theta_1)[\cos\theta_2]_0^{\theta_1}\\
&= \tfrac{1}{2}\cos(\tfrac{1}{2}\theta_2)(\sin\theta_1-\theta_1)-\tfrac{1}{2}\sin(\tfrac{1}{2}\theta_1)(\cos\theta_1-1)\\
&= -\tfrac{1}{2}\theta_1\cos(\tfrac{1}{2}\theta_1)\\
&\quad+\tfrac{1}{2}[\sin\theta_1\cos(\tfrac{1}{2}\theta_1)-\cos\theta_1\sin(\tfrac{1}{2}\theta_2)]\\
&\quad+\tfrac{1}{2}\sin(\tfrac{1}{2}\theta_1)\\
&= -\tfrac{1}{2}\theta_1\cos(\tfrac{1}{2}\theta_1)+\sin(\tfrac{1}{2}\theta_1),
\end{aligned}\tag{11}$$

so

$$\int_0^{\pi}I_2\sin(\tfrac{1}{2}\theta_1)\,d\theta_1=\tfrac{1}{4}\pi.\tag{12}$$

Combining (10) and (12) gives

$$\bar{A}=\frac{1}{\pi^2}\left(\frac{5\pi}{4}+\frac{\pi}{4}\right)=\frac{3}{2\pi}\approx 0.4775.\tag{13}$$

The VARIANCE is

$$\begin{aligned}
{\sigma_A}^2 &= \frac{1}{2\pi^2}\int_0^{\pi}\int_0^{2\pi}[A(\theta_1,\theta_2)-\tfrac{3}{2\pi}]^2\,d\theta_2\,d\theta_1\\
&= \frac{1}{2\pi^2}\int_0^{\pi}\int_0^{2\pi}\Big[2\left|\sin(\tfrac{1}{2}\theta_1)\sin(\tfrac{1}{2}\theta_2)\sin[\tfrac{1}{2}(\theta_1-\theta_2)]\right|\\
&\quad-\frac{3}{2\pi}\Big]^2\,d\theta_2\,d\theta_1\\
&= \frac{1}{2\pi^2}\int_0^{\pi}\int_0^{2\pi}\Big\{4\sin^2(\tfrac{1}{2}\theta_1)\sin^2(\tfrac{1}{2}\theta_2)\sin^2[\tfrac{1}{2}(\theta_2-\theta_1)]\\
&\quad-\frac{6}{\pi}\left|\sin(\tfrac{1}{2}\theta_1)\sin(\tfrac{1}{2}\theta_2)\sin[\tfrac{1}{2}(\theta_1-\theta_2)]\right|+\frac{9}{4\pi^2}\Big\}\,d\theta_2\,d\theta_1\\
&= \frac{1}{2\pi^2}\Big[\int_0^{\pi}\pi(2+\theta_1)\sin^2(\tfrac{1}{2}\theta_1)\,d\theta_1\\
&\quad-\frac{6}{\pi}\left(\frac{5\pi}{4}+\frac{\pi}{4}\right)+\frac{9}{4\pi^2}(2\pi^2)\Big]\\
&= \frac{1}{2\pi^2}\left(\frac{3\pi^2}{4}-9+\frac{9}{2}\right)=\frac{1}{2\pi^2}\left(\frac{3\pi^2}{4}-\frac{9}{2}\right)\\
&= \frac{3(\pi^2-6)}{8\pi^2}\approx 0.1470.
\end{aligned}\tag{14}$$

see also POINT-POINT DISTANCE—1-D, TETRAHEDRON INSCRIBING

Triangle Inscribing in an Ellipse

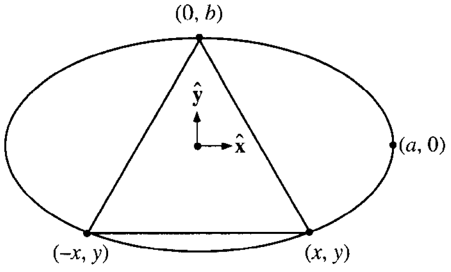

To inscribe an EQUILATERAL TRIANGLE in an ELLIPSE, place the top VERTEX at $(0, b)$, then solve to find the (x, y) coordinate of the other two VERTICES.

$$\sqrt{x^2 + (b-y)^2} = 2x \tag{1}$$

$$x^2 + (b-y)^2 = 4x^2 \tag{2}$$

$$3x^2 = (b-y)^2. \tag{3}$$

Now plugging in the equation of the ELLIPSE

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \tag{4}$$

gives

$$3a^2\left(1 - \frac{y^2}{b^2}\right) = b^2 - 2by + y^2 \tag{5}$$

$$y^2\left(1 + 3\frac{a^2}{b^2}\right) - 2by + (b^2 - 3a^2) = 0 \tag{6}$$

$$\begin{aligned} y &= \frac{2b - \sqrt{4b^2 - 4(b^2-3a^2)\left(1+3\frac{a^2}{b^2}\right)}}{2\left(1+3\frac{a^2}{b^2}\right)} \\ &= \frac{1 - \sqrt{1 - \left(1-3\frac{a^2}{b^2}\right)\left(1+3\frac{a^2}{b^2}\right)}}{1+3\frac{a^2}{b^2}}\,b, \end{aligned} \tag{7}$$

and

$$x = \pm a\sqrt{1 - \frac{y^2}{b^2}}. \tag{8}$$

Triangle Postulate

The sum of the ANGLES of a TRIANGLE is two RIGHT ANGLES. This POSTULATE is equivalent to the PARALLEL AXIOM.

References
Dunham, W. "Hippocrates' Quadrature of the Lune." Ch. 1 in *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, p. 54, 1990.

Triangle Squaring

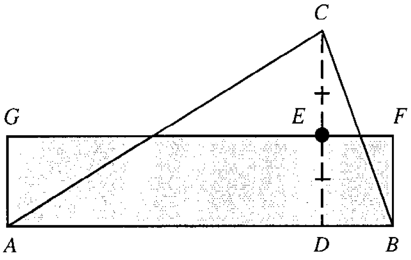

Let CD be the ALTITUDE of a TRIANGLE ΔABC and let E be its MIDPOINT. Then

$$\text{area}(\Delta ABC) = \tfrac{1}{2}AB \cdot CD = AB \cdot DE,$$

and $\square ABFG$ can be squared by RECTANGLE SQUARING. The general POLYGON can be treated by drawing diagonals, squaring the constituent triangles, and then combining the squares together using the PYTHAGOREAN THEOREM.

see also PYTHAGOREAN THEOREM, RECTANGLE SQUARING

References
Dunham, W. "Hippocrates' Quadrature of the Lune." Ch. 1 in *Journey Through Genius: The Great Theorems of Mathematics.* New York: Wiley, pp. 14–15, 1990.

Triangle Tiling

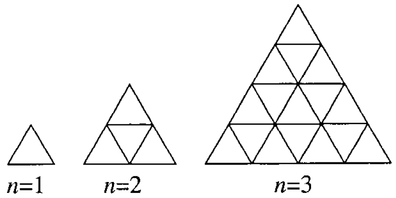

The total number of triangle (including inverted ones) in the above figures is given by

$$N(n) = \begin{cases} \frac{1}{8}n(n+2)(2n+1) & \text{for } n \text{ even} \\ \frac{1}{8}[n(n+2)(2n+1) - 1] & \text{for } n \text{ odd.} \end{cases}$$

The first few values are 1, 5, 13, 27, 48, 78, 118, 170, 235, 315, 411, 525, 658, 812, 988, 1188, 1413, 1665, ... (Sloane's A002717).

References
Conway, J. H. and Guy, R. K. "How Many Triangles." In *The Book of Numbers.* New York: Springer-Verlag, pp. 83–84, 1996.
Sloane, N. J. A. Sequence A002717/M3827 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Triangle Transformation Principle

The triangle transformation principle gives rules for transforming equations involving an INCIRCLE to equations about EXCIRCLES.

see also EXCIRCLE, INCIRCLE

References
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 191–192, 1929.

Triangular Cupola

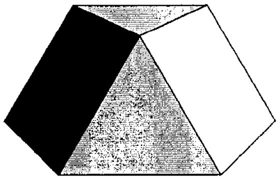

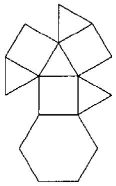

JOHNSON SOLID J_3. The bottom six VERTICES are

$$(\pm\tfrac{1}{2}\sqrt{3}, \pm\tfrac{1}{2}, 0), (0, \pm 1, 0),$$

and the top three VERTICES are

$$\left(\frac{1}{\sqrt{3}},0,\sqrt{\frac{2}{3}}\right),-\left(\frac{1}{2\sqrt{3}},\pm\frac{1}{2},\sqrt{\frac{2}{3}}\right).$$

see also JOHNSON SOLID

Triangular Dipyramid

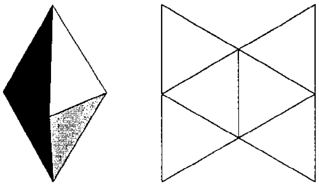

The triangular (or TRIGONAL) dipyramid is one of the convex DELTAHEDRA, and JOHNSON SOLID J_{12}.

see also DELTAHEDRON, DIPYRAMID, JOHNSON SOLID, PENTAGONAL DIPYRAMID

Triangular Graph

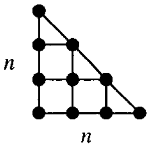

The triangular graph with n nodes on a side is denoted $T(n)$. Tutte (1970) showed that the CHROMATIC POLYNOMIALS of planar triangular graphs possess a ROOT close to $\phi^2 = 2.618033\ldots$, where ϕ is the GOLDEN MEAN. More precisely, if n is the number of VERTICES of G, then

$$P_G(\phi^2) \leq \phi^{5-n}$$

(Le Lionnais 1983, p. 46). Every planar triangular graph possesses a VERTEX of degree 3, 4, or 5 (Le Lionnais 1983, pp. 49 and 53).

see also LATTICE GRAPH

References

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, 1983.

Tutte, W. T. "On Chromatic Polynomials and the Golden Ratio." *J. Combin. Theory* **9**, 289–296, 1970.

Triangular Hebesphenorotunda

see JOHNSON SOLID

Triangular Matrix

An upper triangular MATRIX U is defined by

$$U_{ij} = \begin{cases} a_{ij} & \text{for } i \leq j \\ 0 & \text{for } i > j. \end{cases} \tag{1}$$

Written explicitly,

$$\mathsf{U} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}. \tag{2}$$

A lower triangular MATRIX L is defined by

$$L_{ij} = \begin{cases} a_{ij} & \text{for } i \geq j \\ 0 & \text{for } i < j. \end{cases} \tag{3}$$

Written explicitly,

$$\mathsf{L} = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}. \tag{4}$$

see also HESSENBERG MATRIX, HILBERT MATRIX, MATRIX, VANDERMONDE MATRIX

Triangular Number

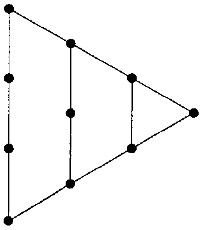

A FIGURATE NUMBER of the form $T_n \equiv n(n+1)/2$ obtained by building up regular triangles out of dots. The first few triangle numbers are 1, 3, 6, 10, 15, 21, ... (Sloane's A000217). $T_4 = 10$ gives the number and arrangement of BOWLING pins, while $T_5 = 15$ gives the number and arrangement of balls in BILLIARDS. Triangular numbers satisfy the RECURRENCE RELATION

$${T_{n+1}}^2 - {T_n}^2 = (n+1)^3, \tag{1}$$

as well as

$$3T_n + T_{n-1} = T_{2n} \tag{2}$$

$$3T_n + T_{n+1} = T_{2n+1} \tag{3}$$

$$1+3+5+\ldots+(2n-1) = T_n + T_{n-1} \tag{4}$$

and

$$(2n+1)^2 = 8T+1 = T_{n-1} + 6T_n + T_{n+1} \tag{5}$$

(Conway and Guy 1996). They have the simple GENERATING FUNCTION

$$f(x) = \frac{x}{(1-x)^3} = x+3x^2+6x^3+10x^4+15x^5+\ldots. \tag{6}$$

Every triangular number is also a HEXAGONAL NUMBER, since

$$\tfrac{1}{2}r(r+1) = \begin{cases} \left(\frac{r+1}{2}\right)\left[2\left(\frac{r+1}{2}\right)-1\right] & \text{for } r \text{ odd} \\ \left(-\frac{r}{2}\right)\left[2\left(-\frac{r}{2}\right)-1\right] & \text{for } r \text{ even.} \end{cases} \tag{7}$$

Also, every PENTAGONAL NUMBER is 1/3 of a triangular number. The sum of consecutive triangular numbers is a SQUARE NUMBER, since

$$\begin{aligned} T_r + T_{r-1} &= \tfrac{1}{2}r(r+1) + \tfrac{1}{2}(r-1)r \\ &= \tfrac{1}{2}r[(r+1)+(r-1)] = r^2. \end{aligned} \tag{8}$$

Interesting identities involving triangular numbers and SQUARE NUMBERS are

$$\sum_{k=1}^{2n-1} (-1)^{k+1} T_k = n^2 \tag{9}$$

$$T_n{}^2 = \sum_{k=1}^{n} k^3 = \tfrac{1}{4}n^2(n+1)^2 \tag{10}$$

$$\sum_{k=1,3,\ldots,q} k^3 = T_n \tag{11}$$

for q ODD and

$$n = \tfrac{1}{2}(q^2 + 2q - 1). \tag{12}$$

All EVEN PERFECT NUMBERS are triangular T_p with PRIME p. Furthermore, every EVEN PERFECT NUMBER $P > 6$ is of the form

$$P = 1 + 9T_n = T_{3n+1}, \tag{13}$$

where T_n is a triangular number with $n = 8j + 2$ (Eaton 1995, 1996). Therefore, the nested expression

$$9(9\cdots(9(9(9(9T_n+1)+1)+1)+1)\ldots+1)+1 \tag{14}$$

generates triangular numbers for any T_n. An INTEGER k is a triangular number IFF $8k + 1$ is a SQUARE NUMBER > 1.

The numbers 1, 36, 1225, 41616, 1413721, 48024900, ... (Sloane's A001110) are SQUARE TRIANGULAR NUMBERS, i.e., numbers which are simultaneously triangular and SQUARE (Pietenpol 1962). Numbers which are simultaneously triangular and TETRAHEDRAL satisfy the BINOMIAL COEFFICIENT equation

$$\binom{n}{2} = \binom{m}{3}, \tag{15}$$

the only solutions of which are $(m, n) = (10, 16)$, (22, 56), and (36, 120) (Guy 1994, p. 147).

The smallest of two INTEGERS for which $n^3 - 13$ is four times a triangular number is 5 (Cesaro 1886; Le Lionnais 1983, p. 56). The only FIBONACCI NUMBERS which are triangular are 1, 3, 21, and 55 (Ming 1989), and the only PELL NUMBER which is triangular is 1 (McDaniel 1996). The BEAST NUMBER 666 is triangular, since

$$T_{6\cdot 6} = T_{36} = 666. \tag{16}$$

In fact, it is the largest REPDIGIT triangular number (Bellew and Weger 1975–76).

FERMAT'S POLYGONAL NUMBER THEOREM states that every POSITIVE INTEGER is a sum of *most* three TRIANGULAR NUMBERS, four SQUARE NUMBERS, five PENTAGONAL NUMBERS, and n n-POLYGONAL NUMBERS. Gauss proved the triangular case, and noted the event in his diary on July 10, 1796, with the notation

$$**\mathit{E\Upsilon RHKA} \qquad \mathit{num} = \Delta + \Delta + \Delta. \tag{17}$$

This case is equivalent to the statement that every number of the form $8m + 3$ is a sum of three ODD SQUARES (Duke 1997). Dirichlet derived the number of ways in which an INTEGER m can be expressed as the sum of three triangular numbers (Duke 1997). The result is particularly simple for a PRIME of the form $8m + 3$, in which case it is the number of squares mod $8m+3$ minus the number of nonsquares mod $8m + 3$ in the INTERVAL $4m + 1$ (Deligne 1973).

The only triangular numbers which are the PRODUCT of three consecutive INTEGERS are 6, 120, 210, 990, 185136, 258474216 (Guy 1994, p. 148).

see also FIGURATE NUMBER, PRONIC NUMBER, SQUARE TRIANGULAR NUMBER

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 59, 1987.

Bellew, D. W. and Weger, R. C. "Repdigit Triangular Numbers." *J. Recr. Math.* **8**, 96–97, 1975–76.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 33–38, 1996.

Deligne, P. "La Conjecture de Weil." *Inst. Hautes Études Sci. Pub. Math.* **43**, 273–308, 1973.

Dudeney, H. E. *Amusements in Mathematics.* New York: Dover, pp. 67 and 167, 1970.

Duke, W. "Some Old Problems and New Results about Quadratic Forms." *Not. Amer. Math. Soc.* **44**, 190–196, 1997.

Eaton, C. F. "Problem 1482." *Math. Mag.* **68**, 307, 1995.

Eaton, C. F. "Perfect Number in Terms of Triangular Numbers." Solution to Problem 1482. *Math. Mag.* **69**, 308–309, 1996.

Guy, R. K. "Sums of Squares" and "Figurate Numbers." §C20 and §D3 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 136–138 and 147–150, 1994.

Hindin, H. "Stars, Hexes, Triangular Numbers and Pythagorean Triples." *J. Recr. Math.* **16**, 191–193, 1983–1984.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 56, 1983.

McDaniel, W. L. "Triangular Numbers in the Pell Sequence." *Fib. Quart.* **34**, 105–107, 1996.

Ming, L. "On Triangular Fibonacci Numbers." *Fib. Quart.* **27**, 98–108, 1989.

Pappas, T. "Triangular, Square & Pentagonal Numbers." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 214, 1989.

Pietenpol, J. L "Square Triangular Numbers." *Amer. Math. Monthly* **169**, 168–169, 1962.
Satyanarayana, U. V. "On the Representation of Numbers as the Sum of Triangular Numbers." *Math. Gaz.* **45**, 40–43, 1961.
Sloane, N. J. A. Sequences A000217/M2535 and A001110/M5259 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Triangular Orthobicupola

see JOHNSON SOLID

Triangular Pyramid

see TETRAHEDRON

Triangular Square Number

see SQUARE TRIANGULAR NUMBER

Triangular Symmetry Group

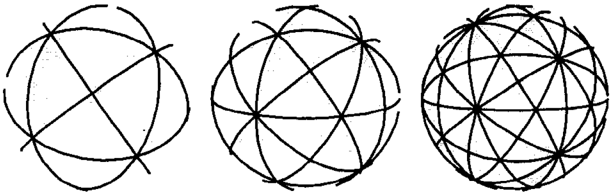
(2, 3, 3) (2, 3, 4) (2, 3, 5)

Given a TRIANGLE with angles (π/p, π/q, π/r), the resulting symmetry GROUP is called a (p, q, r) triangle group (also known as a SPHERICAL TESSELLATION). In 3-D, such GROUPS must satisfy

$$\frac{1}{p}+\frac{1}{q}+\frac{1}{r}>1,$$

and so the only solutions are $(2,2,n)$, $(2,3,3)$, $(2,3,4)$, and $(2,3,5)$ (Ball and Coxeter 1987). The group $(2,3,6)$ gives rise to the semiregular planar TESSELLATIONS of types 1, 2, 5, and 7. The group $(2,3,7)$ gives hyperbolic tessellations.

see also GEODESIC DOME

References
Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 155–161, 1987.
Coxeter, H. S. M. "The Partition of a Sphere According to the Icosahedral Group." *Scripta Math* **4**, 156–157, 1936.
Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, 1973.
Kraitchik, M. "A Mosaic on the Sphere." §7.3 in *Mathematical Recreations.* New York: W. W. Norton, pp. 208–209, 1942.

Triangulation

Triangulation is the division of a surface into a set of TRIANGLES, usually with the restriction that each TRIANGLE side is entirely shared by two adjacent TRIANGLES. It was proved in 1930 that every surface has a triangulation, but it might require an infinite number of TRIANGLES. A surface with a finite number of triangles in its triangulation is called COMPACT. B. Chazelle showed that an arbitrary SIMPLE POLYGON can be triangulated in linear time.

see also COMPACT SURFACE, DELAUNAY TRIANGULATION, JAPANESE TRIANGULATION THEOREM, SIMPLE POLYGON

Triaugmented Dodecahedron

see JOHNSON SOLID

Triaugmented Hexagonal Prism

see JOHNSON SOLID

Triaugmented Triangular Prism

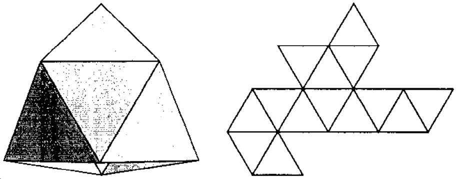

One of the convex DELTAHEDRA and JOHNSON SOLID J_{51}. The VERTICES are $(\pm 1, \pm 1, 0)$, $(0, 0, \sqrt{2})$, $(0, \pm 1, -\sqrt{3})$, $(\pm(1+\sqrt{6})/2, 0, -(\sqrt{2}+\sqrt{3})/2)$, where the x and z coordinates of the last are found by solving

$$\begin{gathered} x^2+1^2+(z+\sqrt{3})^2=2^2 \\ (x-1)^2+1^2+z^2=2^2. \end{gathered}$$

see also DELTAHEDRON, JOHNSON SOLID

Triaugmented Truncated Dodecahedron

see JOHNSON SOLID

Triaxial Ellipsoid

see ELLIPSOID

Tribar

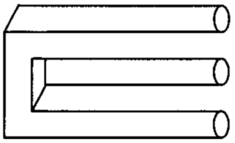

An IMPOSSIBLE FIGURE published by R. Penrose (1958). It also exists as a TRIBOX.

References
Draper, S. W. "The Penrose Triangle and a Family of Related Figures." *Perception* **7**, 283–296, 1978.
Fineman, M. *The Nature of Visual Illusion.* New York: Dover, p. 119, 1996.

Jablan, S. "Set of Modular Elements 'Space Tiles'." `http://members.tripod.com/~modularity/space.htm`.
Pappas, T. "The Impossible Tribar." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 13, 1989.
Penrose, R. "Impossible Objects: A Special Type of Visual Illusion." *Brit. J. Psychology* **49**, 31–33, 1958.

Tribox

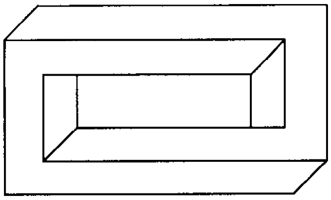

An IMPOSSIBLE FIGURE.

see also IMPOSSIBLE FIGURE, TRIBAR

References

Jablan, S. "Are Impossible Figures Possible?" `http://members.tripod.com/~modularity/kulpa.htm`.

Tribonacci Number

The tribonacci numbers are a generalization of the FIBONACCI NUMBERS defined by $T_1 = 1$, $T_2 = 1$, $T_3 = 2$, and the RECURRENCE RELATION

$$T_n = T_{n-1} + T_{n-2} + T_{n-3} \qquad (1)$$

for $n \geq 4$. The represent the $n = 3$ case of the FIBONACCI n-STEP NUMBERS. The first few terms are 1, 1, 2, 4, 7, 13, 24, 44, 81, 149, ... (Sloane's A000073). The ratio of adjacent terms tends to 1.83929, which is the REAL ROOT of $x^4 - 2x^3 + 1 = 0$. The Tribonacci numbers can also be computed using the GENERATING FUNCTION

$$\begin{aligned}\frac{1}{1-z-z^2-z^3} &= 1 + z + 2z^2 + 4z^3 + 7z^4 \\ &\quad + 13z^5 + 24z^6 + 44z^7 + 81z^8 + 149z^9 + \ldots. \qquad (2)\end{aligned}$$

An explicit FORMULA for T_n is also given by

$$\left[3\frac{\{\frac{1}{3}(19+3\sqrt{33})^{1/3}+\frac{1}{3}(19-3\sqrt{33})^{1/3}+\frac{1}{3}\}^n(586+102\sqrt{33})^{1/3}}{(586+102\sqrt{33})^{2/3}+4-2(586+102\sqrt{33})^{1/3}}\right], \qquad (3)$$

where $[x]$ denotes the NINT function (Plouffe). The first part of a NUMERATOR is related to the REAL root of $x^3 - x^2 - x - 1$, but determination of the DENOMINATOR requires an application of the LLL ALGORITHM. The numbers increase asymptotically to

$$T_n \sim c^n, \qquad (4)$$

where

$$\begin{aligned}c &= (\tfrac{19}{27} + \tfrac{1}{9}\sqrt{33}\,)^{1/3} + \tfrac{4}{9}(\tfrac{19}{27} + \tfrac{1}{9}\sqrt{33}\,)^{-1/3} + \tfrac{1}{3} \\ &= 1.83928675521\ldots \qquad (5)\end{aligned}$$

(Plouffe).

see also FIBONACCI n-STEP NUMBER, FIBONACCI NUMBER, TETRANACCI NUMBER

References

Plouffe, S. "Tribonacci Constant." `http://lacim.uqam.ca/piDATA/tribo.txt`.
Sloane, N. J. A. Sequence A000073/M1074 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Trichotomy Law

Every REAL NUMBER is NEGATIVE, 0, or POSITIVE.

Tricolorable

A projection of a LINK is tricolorable if each of the strands in the projection can be colored in one of three different colors such that, at each crossing, all three colors come together or only one does and at least two different colors are used. The TREFOIL KNOT and trivial 2-link are tricolorable, but the UNKNOT, WHITEHEAD LINK, and FIGURE-OF-EIGHT KNOT are not.

If the projection of a knot is tricolorable, then REIDEMEISTER MOVES on the knot preserve tricolorability, so either every projection of a knot is tricolorable or none is.

Tricomi Function

see CONFLUENT HYPERGEOMETRIC FUNCTION OF THE SECOND KIND, GORDON FUNCTION

Tricuspoid

see DELTOID

Tricylinder

see STEINMETZ SOLID

Tridecagon

A 13-sided POLYGON, sometimes also called the TRISKAIDECAGON.

Trident

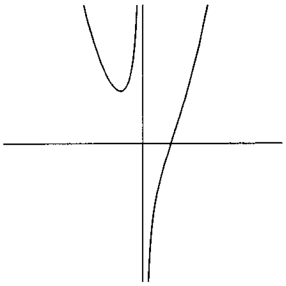

The plane curve given by the equation

$$xy = x^3 - a^3.$$

see also TRIDENT OF DESCARTES, TRIDENT OF NEWTON

Trident of Descartes

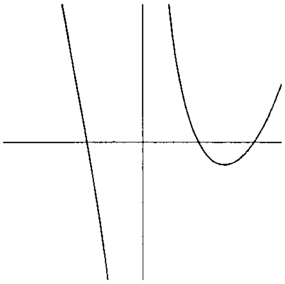

The plane curve given by the equation

$$(a+x)(a-x)(2a-x) = x^3 - 2ax^2 - a^2x + 2a^3 = axy$$

$$y = \frac{(a+x)(a-x)(2a-x)}{ax}$$

The above plot has $a = 2$.

Trident of Newton

The CUBIC CURVE defined by

$$ax^3 + bx^2 + cx + d = xy$$

with $a \neq 0$. The curve cuts the axis in either one or three points. It was the 66th curve in Newton's classification of CUBICS. Newton stated that the curve has four infinite legs and that the y-axis is an ASYMPTOTE to two tending toward contrary parts.

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 109–110, 1972.

MacTutor History of Mathematics Archive. "Trident of Newton." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Trident.html`.

Tridiagonal Matrix

A MATRIX with NONZERO elements only on the diagonal and slots horizontally or vertically adjacent the diagonal. A general 4 × 4 tridiagonal MATRIX has the form

$$\begin{bmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ 0 & a_{32} & a_{33} & a_{34} \\ 0 & 0 & a_{43} & a_{44} \end{bmatrix}.$$

Inversion of such a matrix requires only n (as opposed to n^3) arithmetic operations (Acton 1990).

see also DIAGONAL MATRIX, JACOBI ALGORITHM

References

Acton, F. S. *Numerical Methods That Work, 2nd printing.* Washington, DC: Math. Assoc. Amer., p. 103, 1990.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Tridiagonal and Band Diagonal Systems of Equations." §2.4 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 42–47, 1992.

Tridiminished Icosahedron

see JOHNSON SOLID

Tridiminished Rhombicosidodecahedron

see JOHNSON SOLID

Tridyakis Icosahedron

The DUAL POLYHEDRON of the ICOSITRUNCATED DODECADODECAHEDRON.

Trifolium

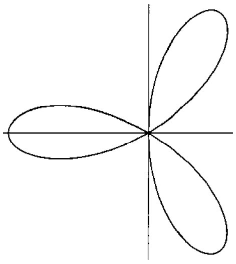

Lawrence (1972) defines a trifolium as a FOLIUM with $b \in (0, 4a)$. However, the term "the" trifolium is sometimes applied to the FOLIUM with $b = a$, which is then the 3-petalled ROSE with Cartesian equation

$$(x^2 + y^2)[y^2 + x(x+a)] = 4axy^2$$

and polar equation

$$r = a\cos\theta(4\sin^2\theta - 1) = -a\cos(3\theta).$$

The trifolium with $b = a$ is the RADIAL CURVE of the DELTOID.

see also BIFOLIUM, FOLIUM, QUADRIFOLIUM

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 152–153, 1972.

MacTutor History of Mathematics Archive. "Trifolium." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Trifolium.html`.

Trigon

see TRIANGLE

Trigonal Dipyramid

see TRIANGULAR DIPYRAMID

Trigonal Dodecahedron

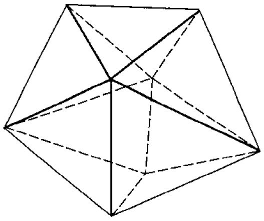

An irregular DODECAHEDRON.

see also DODECAHEDRON, PYRITOHEDRON, RHOMBIC DODECAHEDRON

References
Cotton, F. A. *Chemical Applications of Group Theory, 3rd ed.* New York: Wiley, p. 62, 1990.

Trigonometric Functions

see TRIGONOMETRY

Trigonometric Series

$$\begin{aligned}&A\sin(2\phi)+B\sin(4\phi)+C\sin(6\phi)+D\sin(8\phi)\\&=\sin(2\phi)(A'+\cos(2\phi)(B'+\cos(2\phi)(C'+D'\cos(2\phi)))),\end{aligned}$$

where

$$\begin{aligned}A'&\equiv A-C\\B'&\equiv 2B-4D\\C'&\equiv 4C\\D'&\equiv 8D.\end{aligned}$$

$$\begin{aligned}&A\sin\phi+B\sin(3\phi)+C\sin(5\phi)+D\sin(7\phi)\\&\qquad=\sin\phi(A'+\sin^2\phi(B'+\sin^2\phi(C'+D'\sin^2\phi))),\end{aligned}$$

where

$$\begin{aligned}A'&\equiv A+3B+5C+7D\\B'&\equiv -4B-20C-56D\\C'&\equiv 16C+112D\\D'&\equiv -64D.\end{aligned}$$

$$\begin{aligned}&A+B\cos(2\phi)+C\cos(4\phi)+D\cos(6\phi)+E\cos(8\phi)\\&\qquad=A'+\cos(2\phi)(B'+\cos(2\phi)(C'+\cos(2\phi)\\&\qquad\qquad\times(D'+E'\cos(2\phi)))),\end{aligned}$$

where

$$\begin{aligned}A'&\equiv A-C+E\\B'&\equiv B-3D\\C'&\equiv 2C-8E\\D'&\equiv 4D\\E'&\equiv 8E.\end{aligned}$$

References
Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, p. 19, 1987.

Trigonometric Substitution

INTEGRALS of the form

$$\int f(\cos\theta,\sin\theta)\,d\theta \tag{1}$$

can be solved by making the substitution $z=e^{i\theta}$ so that $dz=ie^{i\theta}\,d\theta$ and expressing

$$\cos\theta=\frac{e^{i\theta}+e^{-i\theta}}{2}=\frac{z+z^{-1}}{2} \tag{2}$$

$$\sin\theta=\frac{e^{i\theta}-e^{-i\theta}}{2i}=\frac{z-z^{-1}}{2i}. \tag{3}$$

The integral can then be solved by CONTOUR INTEGRATION.

Alternatively, making the substitution $t\equiv\tan(\theta/2)$ transforms (1) into

$$\int f\left(\frac{2t}{1+t^2},\frac{1-t^2}{1+t^2}\right)\frac{2\,dt}{1+t^2}. \tag{4}$$

The following table gives trigonometric substitutions which can be used to transform integrals involving square roots.

Form	Substitution
$\sqrt{a^2-x^2}$	$x=a\sin\theta$
$\sqrt{a^2+x^2}$	$x=a\tan\theta$
$\sqrt{x^2-a^2}$	$x=a\sec\theta$

see also HYPERBOLIC SUBSTITUTION

Trigonometry

The study of ANGLES and of the angular relationships of planar and 3-D figures is known as trigonometry. The trigonometric functions (also called the CIRCULAR FUNCTIONS) comprising trigonometry are the COSECANT $\csc x$, COSINE $\cos x$, COTANGENT $\cot x$, SECANT $\sec x$, SINE $\sin x$, and TANGENT $\tan x$. The inverses of these functions are denoted $\csc^{-1}x$, $\cos^{-1}x$, $\cot^{-1}x$, $\sec^{-1}x$, $\sin^{-1}x$, and $\tan^{-1}x$. Note that the f^{-1} NOTATION here means INVERSE FUNCTION, *not* f to the -1 POWER.

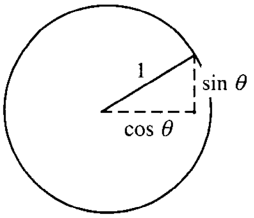

The trigonometric functions are most simply defined using the UNIT CIRCLE. Let θ be an ANGLE measured counterclockwise from the x-AXIS along an ARC of the CIRCLE. Then $\cos\theta$ is the horizontal coordinate of the ARC endpoint, and $\sin\theta$ is the vertical component. The RATIO $\sin\theta/\cos\theta$ is defined as $\tan\theta$. As a result of this definition, the trigonometric functions are periodic with period 2π, so

$$\text{func}(2\pi n + \theta) = \text{func}(\theta), \tag{1}$$

where n is an INTEGER and func is a trigonometric function.

From the PYTHAGOREAN THEOREM,

$$\sin^2\theta + \cos^2\theta = 1. \tag{2}$$

Therefore, it is also true that

$$\tan^2\theta + 1 = \sec^2\theta \tag{3}$$

$$1 + \cot^2\theta = \csc^2\theta. \tag{4}$$

The trigonometric functions can be defined algebraically in terms of COMPLEX EXPONENTIALS (i.e., using the EULER FORMULA) as

$$\sin z \equiv \frac{e^{iz} - e^{-iz}}{2i} \tag{5}$$

$$\csc z \equiv \frac{1}{\sin z} = \frac{2i}{e^{iz} - e^{-iz}} \tag{6}$$

$$\cos z \equiv \frac{e^{iz} + e^{-iz}}{2} \tag{7}$$

$$\sec z \equiv \frac{1}{\cos z} = \frac{2}{e^{iz} + e^{-iz}} \tag{8}$$

$$\tan z \equiv \frac{\sin z}{\cos z} = \frac{e^{iz} - e^{-iz}}{i(e^{iz} + e^{-iz})} \tag{9}$$

$$\cot z \equiv \frac{1}{\tan z} = \frac{i(e^{iz} + e^{-iz})}{e^{iz} - e^{-iz}} = \frac{i(1 + e^{-2iz})}{1 - e^{-2iz}}. \tag{10}$$

OSBORNE'S RULE gives a prescription for converting trigonometric identities to analogous identities for HYPERBOLIC FUNCTIONS.

The ANGLES $n\pi/m$ (with m, n integers) for which the trigonometric function may be expressed in terms of finite root extraction of *real numbers* are limited to values of m which are precisely those which produce constructible POLYGONS. Gauss showed these to be of the form

$$m = 2^k p_1 p_2 \cdots p_s, \tag{11}$$

where k is an INTEGER ≥ 0 and the p_i are distinct FERMAT PRIMES. The first few values are m = 1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 16, 17, 20, ... (Sloane's A003401). Although formulas for trigonometric functions may be found analytically for other m as well, the expressions involve ROOTS of COMPLEX NUMBERS obtained by solving a CUBIC, QUARTIC, or higher order equation. The cases $m = 7$ and $m = 9$ involve the CUBIC EQUATION and QUARTIC EQUATION, respectively. A partial table of the analytic values of SINE, COSINE, and TANGENT for arguments π/m is given below. Derivations of these formulas appear in the following entries.

°	rad	$\sin x$	$\cos x$	$\tan x$
0.0	0	0	1	0
15.0	$\frac{1}{12}\pi$	$\frac{1}{4}(\sqrt{6} - \sqrt{2})$	$\frac{1}{4}(\sqrt{6} + \sqrt{2})$	$2 - \sqrt{3}$
18.0	$\frac{1}{10}\pi$	$\frac{1}{4}(\sqrt{5} - 1)$	$\frac{1}{4}\sqrt{10 + 2\sqrt{5}}$	$\frac{1}{5}\sqrt{25 - 10\sqrt{5}}$
22.5	$\frac{1}{8}\pi$	$\frac{1}{2}\sqrt{2 - \sqrt{2}}$	$\frac{1}{2}\sqrt{2 + \sqrt{2}}$	$\sqrt{2} - 1$
30.0	$\frac{1}{6}\pi$	$\frac{1}{2}$	$\frac{1}{2}\sqrt{3}$	$\frac{1}{3}\sqrt{3}$
36.0	$\frac{1}{5}\pi$	$\frac{1}{4}\sqrt{10 - 2\sqrt{5}}$	$\frac{1}{4}(1 + \sqrt{5})$	$\sqrt{5 - 2\sqrt{5}}$
45.0	$\frac{1}{4}\pi$	$\frac{1}{2}\sqrt{2}$	$\frac{1}{2}\sqrt{2}$	1
60.0	$\frac{1}{3}\pi$	$\frac{1}{2}\sqrt{3}$	$\frac{1}{2}$	$\sqrt{3}$
90.0	$\frac{1}{2}\pi$	1	0	∞
180.0	π	0	-1	0

The INVERSE TRIGONOMETRIC FUNCTIONS are generally defined on the following domains.

Function	Domain
$\sin^{-1} x$	$-\frac{1}{2}\pi \leq y \leq \frac{1}{2}\pi$
$\cos^{-1} x$	$0 \leq y \leq \pi$
$\tan^{-1} x$	$-\frac{1}{2}\pi < y < \frac{1}{2}\pi$
$\csc^{-1} x$	$0 \leq y \leq \frac{1}{2}\pi$ or $\pi \leq y \leq \frac{3\pi}{2}$
$\sec^{-1} x$	$0 \leq y \leq \pi$
$\cot^{-1} x$	$0 \leq y \leq \frac{1}{2}\pi$ or $-\pi \leq y \leq -\frac{1}{2}\pi$

Inverse-forward identities are

$$\tan^{-1}(\cot x) = \tfrac{1}{2}\pi - x \tag{12}$$

$$\sin^{-1}(\cos x) = \tfrac{1}{2}\pi - x \tag{13}$$

$$\sec^{-1}(\csc x) = \tfrac{1}{2}\pi - x, \tag{14}$$

and forward-inverse identities are

$$\cos(\sin^{-1} x) = \sqrt{1 - x^2} \tag{15}$$

$$\cos(\tan^{-1} x) = \frac{1}{\sqrt{1 + x^2}} \tag{16}$$

$$\sin(\cos^{-1} x) = \sqrt{1 - x^2}$$

$$\sin(\tan^{-1} x) = \frac{x}{\sqrt{1 + x^2}} \tag{17}$$

$$\tan(\cos^{-1} x) = \frac{\sqrt{1 - x^2}}{x} \tag{18}$$

$$\tan(\sin^{-1} x) = \frac{x}{\sqrt{1 - x^2}}. \tag{19}$$

Inverse sum identities include

$$\sin^{-1} x + \cos^{-1} x = \tfrac{1}{2}\pi \tag{20}$$

$$\tan^{-1} x + \cot^{-1} x = \tfrac{1}{2}\pi \tag{21}$$

$$\sec^{-1} x + \csc^{-1} x = \tfrac{1}{2}\pi, \tag{22}$$

where (20) follows from

$$x = \sin(\sin^{-1} x) = \cos(\tfrac{1}{2}\pi - \sin^{-1} x). \tag{23}$$

Complex inverse identities in terms of LOGARITHMS include

$$\begin{aligned}
\sin^{-1}(z) &= -i\ln(iz \pm \sqrt{1-z^2}\,) && (24)\\
\cos^{-1}(z) &= -i\ln(z \pm i\sqrt{1-z^2}\,) && (25)\\
\tan^{-1}(z) &= -i\ln\left(\frac{1+iz}{\sqrt{1+z^2}}\right) && (26)\\
&= \tfrac{1}{2}i\ln\left(\frac{1-iz}{1+iz}\right). && (27)
\end{aligned}$$

For IMAGINARY arguments,

$$\begin{aligned}
\sin(iz) &= i\sinh z && (28)\\
\cos(iz) &= \cosh z. && (29)
\end{aligned}$$

For COMPLEX arguments,

$$\begin{aligned}
\sin(x+iy) &= \sin x\cosh y + i\cos x\sinh y && (30)\\
\cos(x+iy) &= \cos x\cosh y - i\sin x\sinh y. && (31)
\end{aligned}$$

For the ABSOLUTE SQUARE of COMPLEX arguments $z = x+iy$,

$$\begin{aligned}
|\sin(x+iy)|^2 &= \sin^2 x + \sinh^2 y && (32)\\
|\cos(x+iy)|^2 &= \cos^2 x + \sinh^2 y. && (33)
\end{aligned}$$

The MODULUS also satisfies the curious identity

$$|\sin(x+iy)| = |\sin x + \sin(iy)|. \tag{34}$$

The only functions satisfying identities of this form,

$$|f(x+iy)| = |f(x) + f(iy)| \tag{35}$$

are $f(z) = Az$, $f(z) = A\sin(bz)$, and $f(z) = A\sinh(bz)$ (Robinson 1957).

Trigonometric product formulas can be derived using the following figure (Kung 1996).

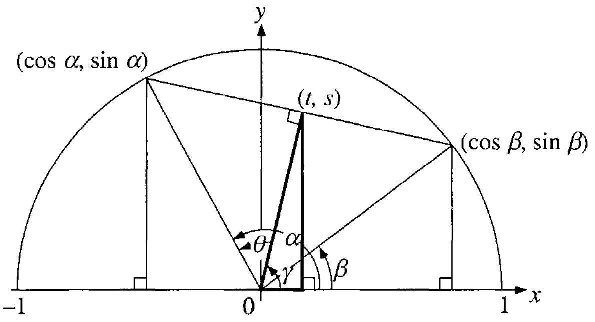

In the figure,

$$\begin{aligned}
\theta &= \tfrac{1}{2}(\alpha-\beta) && (36)\\
\gamma &= \tfrac{1}{2}(\alpha+\beta), && (37)
\end{aligned}$$

so

$$s = \tfrac{1}{2}(\sin\alpha + \sin\beta) = \cos[\tfrac{1}{2}(\alpha-\beta)]\sin[\tfrac{1}{2}(\alpha+\beta)] \tag{38}$$

$$t = \tfrac{1}{2}(\cos\alpha + \cos\beta) = \cos[\tfrac{1}{2}(\alpha-\beta)]\cos[\tfrac{1}{2}(\alpha+\beta)]. \tag{39}$$

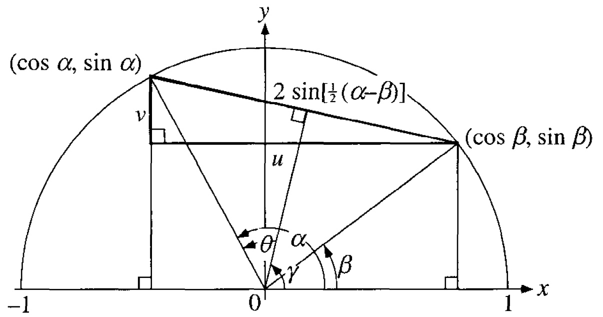

With θ and γ as previously defined, the above figure (Kung 1996) gives

$$u = \cos\beta - \cos\alpha = 2\sin[\tfrac{1}{2}(\alpha-\beta)]\sin[\tfrac{1}{2}(\alpha+\beta)] \tag{40}$$

$$v = \sin\alpha - \sin\beta = 2\sin[\tfrac{1}{2}(\alpha-\beta)]\cos[\tfrac{1}{2}(\alpha+\beta)]. \tag{41}$$

Angle addition FORMULAS express trigonometric functions of sums of angles $\alpha \pm \beta$ in terms of functions of α and β. They can be simply derived used COMPLEX exponentials and the EULER FORMULA,

$$\begin{aligned}
\sin(\alpha+\beta) &= \frac{e^{i(\alpha+\beta)} - e^{-i(\alpha+\beta)}}{2i} = \frac{e^{i\alpha}e^{i\beta} - e^{-i\alpha}e^{-i\beta}}{2i}\\
&= \frac{(\cos\alpha + i\sin\alpha)(\cos\beta + i\sin\beta)}{2i}\\
&\quad - \frac{(\cos\alpha - i\sin\alpha)(\cos\beta - i\sin\beta)}{2i}\\
&= \frac{\cos\alpha\cos\beta + i\sin\beta\cos\alpha + i\sin\alpha\cos\beta - \sin\alpha\sin\beta}{2i}\\
&\quad + \frac{-\cos\alpha\cos\beta + i\cos\alpha\sin\beta + i\sin\alpha\cos\beta + \sin\alpha\sin\beta}{2i}\\
&= \sin\alpha\cos\beta + \sin\beta\cos\alpha
\end{aligned} \tag{42}$$

$$\begin{aligned}
\cos(\alpha+\beta) &= \frac{e^{i(\alpha+\beta)} + e^{-i(\alpha+\beta)}}{2} = \frac{e^{i\alpha}e^{i\beta} + e^{-i\alpha}e^{-i\beta}}{2}\\
&= \frac{(\cos\alpha + i\sin\alpha)(\cos\beta + i\sin\beta)}{2}\\
&\quad + \frac{(\cos\alpha - i\sin\alpha)(\cos\beta - i\sin\beta)}{2}\\
&= \frac{\cos\alpha\cos\beta + i\cos\alpha\sin\beta + i\sin\alpha\cos\beta - \sin\alpha\sin\beta}{2}\\
&\quad + \frac{\cos\alpha\cos\beta - i\cos\alpha\sin\beta - i\sin\alpha\cos\beta - \sin\alpha\sin\beta}{2}\\
&= \cos\alpha\cos\beta - \sin\alpha\sin\beta.
\end{aligned} \tag{43}$$

Taking the ratio gives the tangent angle addition FORMULA

$$\tan(\alpha+\beta) \equiv \frac{\sin(\alpha+\beta)}{\cos(\alpha+\beta)} = \frac{\sin\alpha\cos\beta+\sin\beta\cos\alpha}{\cos\alpha\cos\beta-\sin\alpha\sin\beta}$$

$$= \frac{\frac{\sin\alpha}{\cos\alpha}+\frac{\sin\beta}{\cos\beta}}{1-\frac{\sin\alpha\sin\beta}{\cos\alpha\cos\alpha\beta}} = \frac{\tan\alpha+\tan\beta}{1-\tan\alpha\tan\beta}. \qquad (44)$$

The angle addition FORMULAS can also be derived purely algebraically without the use of COMPLEX NUMBERS. Consider the following figure.

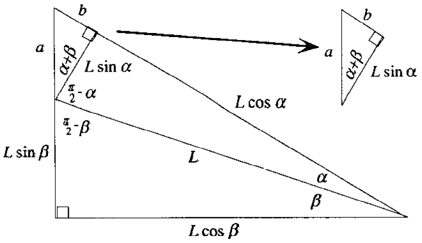

From the large RIGHT TRIANGLE,

$$\sin(\alpha+\beta) = \frac{L\sin\beta+a}{L\cos\alpha+b} \qquad (45)$$

$$\cos(\alpha+\beta) = \frac{L\cos\beta}{L\cos\alpha+b}. \qquad (46)$$

But, from the small triangle (inset at upper right),

$$a = \frac{L\sin\alpha}{\cos(\alpha+\beta)} \qquad (47)$$

$$b = L\sin\alpha\tan(\alpha+\beta). \qquad (48)$$

Plugging a and b from (47) and (48) into (45) and (46) gives

$$\sin(\alpha+\beta) = \frac{L\sin\beta+\frac{L\sin\alpha}{\cos(\alpha+\beta)}}{L\cos\alpha+\frac{L\sin\alpha\sin(\alpha+\beta)}{\cos(\alpha+\beta)}}$$

$$= \frac{\sin\beta\cos(\alpha+\beta)+\sin\alpha}{\cos\alpha\cos(\alpha+\beta)+\sin\alpha\sin(\alpha+\beta)}, \qquad (49)$$

and

$$\cos(\alpha+\beta) = \frac{L\cos\beta}{L\cos\alpha+\frac{L\sin\alpha\sin(\alpha+\beta)}{\cos(\alpha+\beta)}}$$

$$= \frac{\cos\beta}{\cos\alpha+\frac{\sin\alpha\sin(\alpha+\beta)}{\cos(\alpha+\beta)}}. \qquad (50)$$

Now solve (50) for $\cos(\alpha+\beta)$,

$$\cos(\alpha+\beta)\cos\alpha+\sin\alpha\sin(\alpha+\beta) = \cos\beta \qquad (51)$$

to obtain

$$\cos(\alpha+\beta) = \frac{\cos\beta-\sin\alpha\sin(\alpha+\beta)}{\cos\alpha}. \qquad (52)$$

Plugging (52) into (49) gives

$$\sin(\alpha+\beta) = \frac{\sin\beta\left[\frac{\cos\beta-\sin\alpha\sin(\alpha+\beta)}{\cos\alpha}\right]+\sin\alpha}{\cos\alpha\left[\frac{\cos\beta-\sin\alpha\sin(\alpha+\beta)}{\cos\alpha}\right]+\sin\alpha\sin(\alpha+\beta)}$$

$$= \frac{\sin\beta\cos\beta-\sin\alpha\sin\beta\sin(\alpha+\beta)+\sin\alpha\cos\alpha}{\cos\alpha\cos\beta-\sin\alpha\cos\alpha\sin(\alpha+\beta)+\sin\alpha\cos\alpha\sin(\alpha+\beta)}$$

$$= \frac{\sin\beta\cos\beta-\sin\alpha\sin\beta\sin(\alpha+\beta)+\sin\alpha\cos\alpha}{\cos\alpha\cos\beta}$$

$$= \frac{\sin\alpha\cos\alpha+\sin\beta\cos\beta}{\cos\alpha\cos\beta}-\frac{\sin\alpha\sin\beta}{\cos\alpha\cos\beta}\sin(\alpha+\beta), \qquad (53)$$

so

$$\sin(\alpha+\beta)\left(1+\frac{\sin\alpha\sin\beta}{\cos\alpha\cos\beta}\right) = \frac{\sin\alpha\cos\alpha+\sin\beta\cos\beta}{\cos\alpha\cos\beta} \qquad (54)$$

$$\sin(\alpha+\beta)(\cos\alpha\cos\beta+\sin\alpha\sin\beta) = \sin\alpha\cos\alpha+\sin\beta\cos\beta, \qquad (55)$$

and

$$\sin(\alpha+\beta) = \frac{\sin\alpha\cos\alpha+\sin\beta\cos\beta}{\sin\alpha\sin\beta+\cos\alpha\cos\beta}$$

$$= \frac{\sin\alpha\cos\alpha+\sin\beta\cos\beta}{\sin\alpha\sin\beta+\cos\alpha\cos\beta}\,\frac{\sin\alpha\cos\beta+\sin\beta\cos\alpha}{\sin\alpha\cos\beta+\sin\beta\cos\alpha}. \qquad (56)$$

Multiplying out the DENOMINATOR gives

$$\begin{aligned}(\cos\alpha\cos\beta+\sin\alpha\sin\beta)(\sin\alpha\cos\beta+\sin\beta\cos\alpha) &= \sin\alpha\cos\alpha\cos^2\beta+\cos^2\alpha\sin\beta\cos\beta \\ &\quad+\sin^2\alpha\sin\beta\cos\beta+\sin\alpha\cos\alpha\sin^2\beta \\ &= \sin\alpha\cos\alpha+\sin\beta\cos\beta, \end{aligned} \qquad (57)$$

so

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta+\sin\beta\cos\alpha. \qquad (58)$$

Multiplying out (50),

$$\cos(\alpha+\beta)\cos\alpha+\sin\alpha\sin(\alpha+\beta) = \cos\beta \qquad (59)$$

$$\begin{aligned}\cos(\alpha+\beta) &= \frac{\cos\beta-\sin\alpha\sin(\alpha+\beta)}{\cos\alpha} \\ &= \frac{\cos\beta-\sin\alpha(\sin\alpha\cos\beta+\sin\beta\cos\alpha)}{\cos\alpha} \\ &= \frac{\cos\beta(1-\sin^2\alpha)+\sin\alpha\cos\alpha\sin\beta}{\cos\alpha} \\ &= \frac{\cos^2\alpha\cos\beta+\sin\alpha\cos\alpha\sin\beta}{\cos\alpha} \\ &= \cos\alpha\cos\beta+\sin\alpha\sin\beta. \end{aligned} \qquad (60)$$

Summarizing,

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \sin\beta\cos\alpha \quad (61)$$

$$\sin(\alpha-\beta) = \sin\alpha\cos\beta - \sin\beta\cos\alpha \quad (62)$$

$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta \quad (63)$$

$$\cos(\alpha-\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta \quad (64)$$

$$\tan(\alpha+\beta) = \frac{\tan\alpha+\tan\beta}{1-\tan\alpha\tan\beta} \quad (65)$$

$$\tan(\alpha-\beta) = \frac{\tan\alpha-\tan\beta}{1+\tan\alpha\tan\beta}. \quad (66)$$

The sine and cosine angle addition identities can be summarized by the MATRIX EQUATION

$$\begin{bmatrix} \cos x & \sin x \\ -\sin x & \cos x \end{bmatrix} \begin{bmatrix} \cos y & \sin y \\ -\sin y & \cos y \end{bmatrix} = \begin{bmatrix} \cos(x+y) & \sin(x+y) \\ -\sin(x+y) & \cos(x+y) \end{bmatrix}. \quad (67)$$

The double angle formulas are

$$\sin(2\alpha) = 2\sin\alpha\cos\alpha \quad (68)$$

$$\cos(2\alpha) = \cos^2\alpha - \sin^2\alpha \quad (69)$$

$$= 2\cos^2\alpha - 1 \quad (70)$$

$$= 1 - 2\sin^2\alpha \quad (71)$$

$$\tan(2\alpha) = \frac{2\tan\alpha}{1-\tan^2\alpha}. \quad (72)$$

General multiple angle formulas are

$$\sin(n\alpha) = 2\sin[(n-1)\alpha]\cos\alpha - \sin[(n-2)\alpha] \quad (73)$$

$$\sin(nx) = n\cos^{n-1}x\sin x - \frac{n(n-1)(n-2)}{1\cdot2\cdot3}\cos^{n-3}x\sin^3 x + \ldots \quad (74)$$

$$\cos(n\alpha) = 2\cos[(n-1)\alpha]\cos\alpha - \cos[(n-2)\alpha] \quad (75)$$

$$\cos(nx) = \cos^n x - \frac{n(n-1)}{1\cdot2}\cos^{n-2}x\sin^2 x + \frac{n(n-1)(n-2)(n-3)}{1\cdot2\cdot3\cdot4}\cos^{n-4}x\sin^4 x - \ldots \quad (76)$$

$$\tan(n\alpha) = \frac{\tan[(n-1)\alpha]+\tan\alpha}{1-\tan[(n-1)\alpha]\tan\alpha}. \quad (77)$$

Therefore, any trigonometric function of a sum can be broken up into a sum of trigonometric functions with $\sin\alpha\cos\alpha$ cross terms. Particular cases for multiple angle formulas up to $n = 4$ are given below.

$$\sin(3\alpha) = 3\sin\alpha - 4\sin^3\alpha \quad (78)$$

$$\cos(3\alpha) = 4\cos^3\alpha - 3\cos\alpha \quad (79)$$

$$\tan(3\alpha) = \frac{3\tan\alpha - \tan^3\alpha}{1-3\tan^2\alpha} \quad (80)$$

$$\sin(4\alpha) = 4\sin\alpha\cos\alpha - 8\sin^3\alpha\cos\alpha \quad (81)$$

$$\cos(4\alpha) = 8\cos^4\alpha - 8\cos^2\alpha + 1 \quad (82)$$

$$\tan(4\alpha) = \frac{4\tan\alpha - 4\tan^3\alpha}{1-6\tan^2\alpha+\tan^4\alpha}. \quad (83)$$

Beyer (1987, p. 139) gives formulas up to $n = 6$.

Sum identities include

$$\frac{\tan(\alpha-\beta)}{\tan(\alpha+\beta)} = \frac{\sin(\alpha-\beta)\cos(\alpha+\beta)}{\cos(\alpha-\beta)\sin(\alpha+\beta)} = \frac{(\sin\alpha\cos\beta-\sin\beta\cos\alpha)(\cos\alpha\cos\beta-\sin\alpha\sin\beta)}{(\cos\alpha\cos\beta+\sin\alpha\sin\beta)(\sin\alpha\cos\beta+\sin\beta\cos\alpha)} = \frac{\sin\alpha\cos\alpha-\sin\beta\cos\beta}{\sin\alpha\cos\alpha+\sin\beta\cos\beta}. \quad (84)$$

Infinite sum identities include

$$\sum_{k=1,\ 3,\ 5,\ldots}^{\infty} \frac{e^{-kx}\sin(ky)}{k} = \frac{1}{2}\tan^{-1}\left(\frac{\sin y}{\sinh x}\right). \quad (85)$$

Trigonometric half-angle formulas include

$$\sin\left(\frac{\alpha}{2}\right) = \sqrt{\frac{1-\cos\alpha}{2}} \quad (86)$$

$$\cos\left(\frac{\alpha}{2}\right) = \sqrt{\frac{1+\cos\alpha}{2}} \quad (87)$$

$$\tan\left(\frac{\alpha}{2}\right) = \frac{\sin\alpha}{1+\cos\alpha} \quad (88)$$

$$= \frac{1-\cos\alpha}{\sin\alpha} \quad (89)$$

$$= \frac{1\pm\sqrt{1+\tan^2\alpha}}{\tan\alpha} \quad (90)$$

$$= \frac{\tan\alpha\sin\alpha}{\tan\alpha+\sin\alpha}. \quad (91)$$

The PROSTHAPHAERESIS FORMULAS are

$$\sin\alpha+\sin\beta = 2\sin[\tfrac{1}{2}(\alpha+\beta)]\cos[\tfrac{1}{2}(\alpha-\beta)] \quad (92)$$

$$\sin\alpha-\sin\beta = 2\cos[\tfrac{1}{2}(\alpha+\beta)]\sin[\tfrac{1}{2}(\alpha-\beta)] \quad (93)$$

$$\sin\alpha+\cos\beta = 2\cos[\tfrac{1}{2}(\alpha+\beta)]\cos[\tfrac{1}{2}(\alpha-\beta)] \quad (94)$$

$$\cos\alpha-\cos\beta = -2\sin[\tfrac{1}{2}(\alpha+\beta)]\sin[\tfrac{1}{2}(\alpha-\beta)]. \quad (95)$$

Related formulas are

$$\sin\alpha\cos\beta = \tfrac{1}{2}[\sin(\alpha-\beta)+\sin(\alpha+\beta)] \quad (96)$$

$$\cos\alpha\cos\beta = \tfrac{1}{2}[\cos(\alpha-\beta)+\cos(\alpha+\beta)] \quad (97)$$

$$\cos\alpha\sin\beta = \tfrac{1}{2}[\sin(\alpha+\beta)-\sin(\alpha-\beta)] \quad (98)$$

$$\sin\alpha\sin\beta = \tfrac{1}{2}[\cos(\alpha-\beta)-\cos(\alpha+\beta)]. \quad (99)$$

Multiplying both sides by 2 gives the equations sometimes known as the WERNER FORMULAS.

Trigonometric product/sum formulas are

$$\sin(\alpha+\beta)\sin(\alpha-\beta) = \sin^2\alpha - \sin^2\beta = \cos^2\beta - \cos^2\alpha \quad (100)$$

$$\cos(\alpha+\beta)\cos(\alpha-\beta) = \cos^2\alpha - \sin^2\beta = \cos^2\beta - \sin^2\alpha. \quad (101)$$

Power formulas include

$$\sin^2 x = \tfrac{1}{2}[1-\cos(2x)] \quad (102)$$
$$\sin^3 x = \tfrac{1}{4}[3\sin x - \sin(3x)] \quad (103)$$
$$\sin^4 x = \tfrac{1}{8}[3 - 4\cos(2x) + \cos(4x)] \quad (104)$$

and

$$\cos^2 x = \tfrac{1}{2}[1+\cos(2x)] \quad (105)$$
$$\cos^3 x = \tfrac{1}{4}[3\cos x + \cos(3x)] \quad (106)$$
$$\cos^4 x = \tfrac{1}{8}[3 + 4\cos(2x) + \cos(4x)] \quad (107)$$

(Beyer 1987, p. 140). Formulas of these types can also be given analytically as

$$\sin^{2n} x = \frac{1}{2^{2n}}\binom{2n}{n} + \frac{(-1)^n}{2^{2n-1}}\sum_{k=0}^{n-1}(-1)^k\binom{2n}{k}\cos[2(n-k)x] \quad (108)$$

$$\sin^{2n+1} = \frac{(-1)^n}{4^n}\sum_{k=0}^{n}(-1)^k\binom{2n+1}{k}\sin[(2n+1-2k)x] \quad (109)$$

$$\cos^{2n} x = \frac{(-1)^n}{2^{2n}}\binom{2n}{n} + \frac{1}{2^{2n-1}}\sum_{k=0}^{n-1}\binom{2n}{k}\cos[2(n-k)x] \quad (110)$$

$$\cos^{2n+1} x = \frac{1}{4^n}\sum_{k=0}^{n}\binom{2n+1}{k}\cos[(2n+1-2k)x] \quad (111)$$

(Kogan), where $\binom{n}{m}$ is a BINOMIAL COEFFICIENT.

see also COSECANT, COSINE, COTANGENT, EUCLIDEAN NUMBER, INVERSE COSECANT, INVERSE COSINE, INVERSE COTANGENT, INVERSE SECANT, INVERSE SINE, INVERSE TANGENT, INVERSE TRIGONOMETRIC FUNCTIONS, OSBORNE'S RULE, POLYGON, SECANT, SINE, TANGENT, TRIGONOMETRY VALUES: π, π/2, π/3, π/4, π/5, π/6, π/7, π/8, π/9, π/10, π/11, π/12, π/15, π/16, π/17, π/18, π/20, 0, WERNER FORMULAS

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Circular Functions." §4.3 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 71–79, 1972.

Bahm, L. B. *The New Trigonometry on Your Own.* Patterson, NJ: Littlefield, Adams & Co., 1964.

Beyer, W. H. "Trigonometry." *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 134–152, 1987.

Dixon, R. "The Story of Sine and Cosine." § 4.4 in *Mathographics.* New York: Dover, pp. 102–106, 1991.

Hobson, E. W. *A Treatise on Plane Trigonometry.* London: Cambridge University Press, 1925.

Kells, L. M.; Kern, W. F.; and Bland, J. R. *Plane and Spherical Trigonometry.* New York: McGraw-Hill, 1940.

Kogan, S. "A Note on Definite Integrals Involving Trigonometric Functions." `http://www.mathsoft.com/asolve/constant/pi/sin/sin.html`.

Kung, S. H. "Proof Without Words: The Difference-Product Identities" and "Proof Without Words: The Sum-Product Identities." *Math. Mag.* **69**, 269, 1996.

Maor, E. *Trigonometric Delights.* Princeton, NJ: Princeton University Press, 1998.

Morrill, W. K. *Plane Trigonometry, rev. ed.* Dubuque, IA: Wm. C. Brown, 1964.

Robinson, R. M. "A Curious Mathematical Identity." *Amer. Math. Monthly* **64**, 83–85, 1957.

Sloane, N. J. A. Sequence A003401/M0505 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Thompson, J. E. *Trigonometry for the Practical Man.* Princeton, NJ: Van Nostrand.

❁ Weisstein, E. W. "Exact Values of Trigonometric Functions." `http://www.astro.virginia.edu/~eww6n/math/notebooks/TrigExact.m`.

Yates, R. C. "Trigonometric Functions." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 225–232, 1952.

Zill, D. G. and Dewar, J. M. *Trigonometry.* New York: McGraw-Hill 1990.

Trigonometry Values—π

By the definition of the trigonometric functions,

$$\sin\pi = 0 \quad (1)$$
$$\cos\pi = -1 \quad (2)$$
$$\tan\pi = 0 \quad (3)$$
$$\csc\pi = \infty \quad (4)$$
$$\sec\pi = -1 \quad (5)$$
$$\cot\pi = \infty. \quad (6)$$

Trigonometry Values—π/2

By the definition of the trigonometric functions,

$$\sin\left(\frac{\pi}{2}\right) = 1 \quad (1)$$
$$\cos\left(\frac{\pi}{2}\right) = 0 \quad (2)$$
$$\tan\left(\frac{\pi}{2}\right) = \infty \quad (3)$$
$$\csc\left(\frac{\pi}{2}\right) = 1 \quad (4)$$
$$\sec\left(\frac{\pi}{2}\right) = \infty \quad (5)$$
$$\cot\left(\frac{\pi}{2}\right) = 0. \quad (6)$$

see also DIGON

Trigonometry Values—$\pi/3$

From TRIGONOMETRY VALUES: $\pi/6$

$$\sin\left(\frac{\pi}{6}\right) = \tfrac{1}{2} \tag{1}$$

$$\cos\left(\frac{\pi}{6}\right) = \tfrac{1}{2}\sqrt{3} \tag{2}$$

together with the trigonometric identity

$$\sin(2\alpha) = 2\sin\alpha\cos\alpha, \tag{3}$$

the identity

$$\sin\left(\frac{\pi}{3}\right) = 2\sin\left(\frac{\pi}{6}\right)\cos\left(\frac{\pi}{6}\right) = 2(\tfrac{1}{2})(\tfrac{1}{2}\sqrt{3}) = \tfrac{1}{2}\sqrt{3} \tag{4}$$

is obtained. Using the identity

$$\cos(2\alpha) = 1 - 2\sin^2\alpha, \tag{5}$$

then gives

$$\cos\left(\frac{\pi}{3}\right) = 1 - 2\sin^2\left(\frac{\pi}{6}\right) = 1 - 2(\tfrac{1}{2})^2 = \tfrac{1}{2}. \tag{6}$$

Summarizing,

$$\sin\left(\frac{\pi}{3}\right) = \tfrac{1}{2}\sqrt{3} \tag{7}$$

$$\cos\left(\frac{\pi}{3}\right) = \tfrac{1}{2} \tag{8}$$

$$\tan\left(\frac{\pi}{3}\right) = \sqrt{3}. \tag{9}$$

see also EQUILATERAL TRIANGLE

Trigonometry Values—$\pi/4$

For a RIGHT ISOSCELES TRIANGLE, symmetry requires that the angle at each VERTEX be given by

$$\tfrac{1}{2}\pi + 2\alpha = \pi, \tag{1}$$

so $\alpha = \pi/4$. The sides are equal, so

$$\sin^2\alpha + \cos^2\alpha = 2\sin^2\alpha = 1. \tag{2}$$

Solving,

$$\sin\left(\frac{\pi}{4}\right) = \tfrac{1}{2}\sqrt{2} \tag{3}$$

$$\cos\left(\frac{\pi}{4}\right) = \tfrac{1}{2}\sqrt{2} \tag{4}$$

$$\tan\left(\frac{\pi}{4}\right) = 1. \tag{5}$$

see also SQUARE

Trigonometry Values—$\pi/5$

Use the identity

$$\sin(5\alpha) = 5\sin\alpha - 20\sin^3\alpha + 16\sin^5\alpha. \tag{1}$$

Now, let $\alpha \equiv \pi/5$ and $x \equiv \sin\alpha$. Then

$$\sin\pi = 0 = 5x - 20x^3 + 16x^5 \tag{2}$$

$$16x^4 - 20x^2 + 5 = 0. \tag{3}$$

Solving the QUADRATIC EQUATION for x^2 gives

$$\begin{aligned}\sin^2\left(\frac{\pi}{5}\right) = x^2 &= \frac{20 \pm \sqrt{(-20)^2 - 4\cdot 16\cdot 5}}{2\cdot 16}\\ &= \frac{20 \pm \sqrt{80}}{32} = \tfrac{1}{8}(5 \pm \sqrt{5}\,).\end{aligned} \tag{4}$$

Now, $\sin(\pi/5)$ must be less than

$$\sin\left(\frac{\pi}{4}\right) = \tfrac{1}{2}\sqrt{2}, \tag{5}$$

so taking the MINUS SIGN and simplifying gives

$$\sin\left(\frac{\pi}{5}\right) = \sqrt{\frac{5-\sqrt{5}}{8}} = \tfrac{1}{4}\sqrt{10 - 2\sqrt{5}}. \tag{6}$$

$\cos(\pi/5)$ can be computed from

$$\cos\left(\frac{\pi}{5}\right) = \sqrt{1 - \sin^2\left(\frac{\pi}{5}\right)} = \tfrac{1}{4}(1 + \sqrt{5}). \tag{7}$$

Summarizing,

$$\sin\left(\frac{\pi}{5}\right) = \tfrac{1}{4}\sqrt{10 - 2\sqrt{5}} \tag{8}$$

$$\sin\left(\frac{2\pi}{5}\right) = \tfrac{1}{4}\sqrt{10 + 2\sqrt{5}} \tag{9}$$

$$\sin\left(\frac{3\pi}{5}\right) = \tfrac{1}{4}\sqrt{10 + 2\sqrt{5}} \tag{10}$$

$$\sin\left(\frac{4\pi}{5}\right) = \tfrac{1}{4}\sqrt{10 - 2\sqrt{5}} \tag{11}$$

$$\cos\left(\frac{\pi}{5}\right) = \tfrac{1}{4}(1 + \sqrt{5}) \tag{12}$$

$$\cos\left(\frac{2\pi}{5}\right) = \tfrac{1}{4}(-1 + \sqrt{5}\,) \tag{13}$$

$$\cos\left(\frac{3\pi}{5}\right) = \tfrac{1}{4}(1 - \sqrt{5}\,) \tag{14}$$

$$\cos\left(\frac{4\pi}{5}\right) = -\tfrac{1}{4}(1 + \sqrt{5}\,) \tag{15}$$

$$\tan\left(\frac{\pi}{5}\right) = \sqrt{5 - 2\sqrt{5}} \tag{16}$$

$$\tan\left(\frac{2\pi}{5}\right) = \sqrt{5 + 2\sqrt{5}} \tag{17}$$

$$\tan\left(\frac{3\pi}{5}\right) = -\sqrt{5 + 2\sqrt{5}} \tag{18}$$

$$\tan\left(\frac{4\pi}{5}\right) = -\sqrt{5 - 2\sqrt{5}}\,. \tag{19}$$

see also DODECAHEDRON, ICOSAHEDRON, PENTAGON, PENTAGRAM

Trigonometry Values—$\pi/6$

Given a RIGHT TRIANGLE with angles defined to be α and 2α, it must be true that

$$\alpha + 2\alpha + \tfrac{1}{2}\pi = \pi, \qquad (1)$$

so $\alpha = \pi/6$. Define the hypotenuse to have length 1 and the side opposite α to have length x, then the side opposite 2α has length $\sqrt{1-x^2}$. This gives $\sin\alpha \equiv x$ and

$$\sin(2\alpha) = \sqrt{1-x^2}. \qquad (2)$$

But

$$\sin(2\alpha) = 2\sin\alpha\cos\alpha = 2x\sqrt{1-x^2}, \qquad (3)$$

so we have

$$\sqrt{1-x^2} = 2x\sqrt{1-x^2}. \qquad (4)$$

This gives $2x = 1$, or

$$\sin\left(\frac{\pi}{6}\right) = \tfrac{1}{2}. \qquad (5)$$

$\cos(\pi/6)$ is then computed from

$$\cos\left(\frac{\pi}{6}\right) = \sqrt{1-\sin^2\left(\frac{\pi}{6}\right)} = \sqrt{1-(\tfrac{1}{2})^2} = \tfrac{1}{2}\sqrt{3}. \quad (6)$$

Summarizing,

$$\sin\left(\frac{\pi}{6}\right) = \tfrac{1}{2} \qquad (7)$$

$$\cos\left(\frac{\pi}{6}\right) = \tfrac{1}{2}\sqrt{3} \qquad (8)$$

$$\tan\left(\frac{\pi}{6}\right) = \tfrac{1}{3}\sqrt{3}. \qquad (9)$$

see also HEXAGON, HEXAGRAM

Trigonometry Values—$\pi/7$

Trigonometric functions of $n\pi/7$ for n an integer cannot be expressed in terms of sums, products, and finite root extractions on *real* rational numbers because 7 is not a FERMAT PRIME. This also means that the HEPTAGON is not a CONSTRUCTIBLE POLYGON.

However, exact expressions involving roots of *complex* numbers can still be derived using the trigonometric identity

$$\sin(n\alpha) = 2\sin[(n-1)\alpha]\cos\alpha - \sin[(n-2)\alpha]. \qquad (1)$$

The case $n = 7$ gives

$$\begin{aligned}\sin(7\alpha) &= 2\sin(6\alpha)\cos\alpha - \sin(5\alpha)\\ &= 2(32\cos^5\alpha\sin\alpha - 32\cos^3\alpha\sin\alpha + 6\cos\alpha\sin\alpha)\cos\alpha\\ &\quad -(5\sin\alpha - 20\sin^3\alpha + 16\sin^5\alpha)\\ &= 64\cos^6\alpha\sin\alpha - 64\cos^4\alpha\sin\alpha + 12\cos^2\alpha\sin\alpha\\ &\quad -5\sin\alpha + 20(1-\cos^2\alpha)\sin\alpha\\ &\quad -16(1-2\cos^2\alpha+\cos^4\alpha)\sin\alpha\\ &= \sin\alpha(64\cos^6\alpha - 80\cos^4\alpha + 24\cos^2\alpha - 1). \qquad (2)\end{aligned}$$

Rewrite this using the identity $\cos^2\alpha = 1 - \sin^2\alpha$,

$$\begin{aligned}\sin\left(\frac{\pi}{7}\right) &= \sin\alpha(7 - 56\sin^2\alpha + 112\sin^4\alpha - 64\sin^6\alpha)\\ &= -64\sin\alpha(\sin^6\alpha - \tfrac{112}{64}\sin^4\alpha + \tfrac{56}{64}\sin^2\alpha - \tfrac{7}{64}). \qquad (3)\end{aligned}$$

Now, let $\alpha \equiv \pi/7$ and $x \equiv \sin^2\alpha$, then

$$\sin(\pi) = 0 = x^3 - \tfrac{7}{4}x^2 + \tfrac{7}{8}x - \tfrac{7}{64}, \qquad (4)$$

which is a CUBIC EQUATION in x. The ROOTS are numerically found to be $x \approx 0.188255, 0.611260\ldots, 0.950484\ldots$. But $\sin\alpha = \sqrt{x}$, so these ROOTS correspond to $\sin\alpha \approx 0.4338$, $\sin(2\alpha) \approx 0.7817$, $\sin(3\alpha) \approx 0.9749$. By NEWTON'S RELATION

$$\prod_i r_i = -a_0, \qquad (5)$$

we have

$$x_1x_2x_3 = \tfrac{7}{64}, \qquad (6)$$

or

$$\sin\left(\frac{\pi}{7}\right)\sin\left(\frac{2\pi}{7}\right)\sin\left(\frac{3\pi}{7}\right) = \sqrt{\frac{7}{64}} = \tfrac{1}{8}\sqrt{7}. \qquad (7)$$

Similarly,

$$\cos\left(\frac{\pi}{7}\right)\cos\left(\frac{2\pi}{7}\right)\cos\left(\frac{3\pi}{7}\right) = \frac{1}{8}. \qquad (8)$$

The constants of the CUBIC EQUATION are given by

$$Q \equiv \tfrac{1}{9}(3a_1 - a_2{}^2) = \tfrac{1}{9}[3\cdot\tfrac{7}{8} - (-\tfrac{7}{4})^2] = -\tfrac{7}{144} \qquad (9)$$

$$\begin{aligned}R &\equiv \tfrac{1}{54}(9a_2a_1 - 2a_2^3 - 27a_0)\\ &= \tfrac{1}{54}[9(-\tfrac{7}{4})(\tfrac{1}{7}8) - 2(-\tfrac{7}{4})^3 - 27(-\tfrac{7}{64})]\\ &= -\tfrac{7}{3456}. \qquad (10)\end{aligned}$$

The DISCRIMINANT is then

$$\begin{aligned}D &\equiv Q^3 + R^2 = -\tfrac{343}{2{,}985{,}984} + \tfrac{49}{11{,}943{,}936}\\ &= -\tfrac{49}{442{,}368} < 0, \qquad (11)\end{aligned}$$

so there are three distinct REAL ROOTS. Finding the first one,

$$x = \sqrt[3]{R+\sqrt{D}} + \sqrt[3]{R-\sqrt{D}} - \tfrac{1}{3}a_2. \qquad (12)$$

Writing

$$\sqrt{D} = 3^{-3/2}\tfrac{7}{128}i, \qquad (13)$$

plugging in from above, and anticipating that the solution we have picked corresponds to $\sin(3\pi/7)$,

$$\sin\left(\frac{3\pi}{7}\right) = \sqrt{x} =$$

$$\sqrt{\sqrt[3]{-\frac{7}{3456} + 3^{-3/2}\frac{7}{128}i} + \sqrt[3]{-\frac{7}{3456} - 3^{-3/2}\frac{7}{128}i} - \frac{1}{3}\left(-\frac{7}{4}\right)}$$
$$= \sqrt{\sqrt[3]{-\frac{7}{3456} + 3^{-3/2}\frac{7}{128}i} + \sqrt[3]{-\frac{7}{3456} - 3^{-3/2}\frac{7}{128}i} + \frac{7}{12}}$$
$$= \sqrt{\sqrt[3]{\frac{7}{3456}\left(-1 + 3^{3/2}i\right)} - \sqrt[3]{\frac{7}{3456}\left(1 + 3^{3/2}i\right)} + \frac{7}{12}}$$
$$= \sqrt{\frac{1}{12}\left[\sqrt[3]{\frac{7}{2}\left(-1 + 3^{3/2}i\right)} - \sqrt[3]{\frac{7}{2}\left(1 + 3^{3/2}i\right)} + 7\right]}. \qquad (14)$$

see also HEPTAGON

Trigonometry Values—π/8

$$\sin\left(\frac{\pi}{8}\right) = \sin\left(\frac{1}{2}\cdot\frac{\pi}{4}\right) = \sqrt{\frac{1}{2}\left(1 - \cos\frac{\pi}{4}\right)}$$
$$= \sqrt{\tfrac{1}{2}\left(1 - \tfrac{1}{2}\sqrt{2}\right)} = \tfrac{1}{2}\sqrt{2 - \sqrt{2}}. \qquad (1)$$

Now, checking to see if the SQUARE ROOT can be simplified gives

$$a^2 - b^2c = 2^2 - 1^2 \cdot 2 = 4 - 2 = 2, \qquad (2)$$

which is not a PERFECT SQUARE, so the above expression cannot be simplified. Similarly,

$$\cos\left(\frac{\pi}{8}\right) = \cos\left(\frac{1}{2}\frac{\pi}{4}\right) = \sqrt{\frac{1}{2}\left(1 + \cos\frac{\pi}{4}\right)}$$
$$= \sqrt{\frac{1}{2}\left(1 + \frac{\sqrt{2}}{2}\right)} = \tfrac{1}{2}\sqrt{2 + \sqrt{2}} \qquad (3)$$

$$\tan\left(\frac{\pi}{8}\right) = \sqrt{\frac{2 - \sqrt{2}}{2 + \sqrt{2}}} = \sqrt{\frac{(2 - \sqrt{2})^2}{4 - 2}} = \sqrt{\frac{4 + 2 - 4\sqrt{2}}{2}}$$
$$= \sqrt{\frac{6 - 4\sqrt{2}}{2}} = \sqrt{3 - 2\sqrt{2}}. \qquad (4)$$

But

$$a^2 - b^2c = 3^2 - 2^2 2 = 9 - 8 = 1 \qquad (5)$$

is a PERFECT SQUARE, so we can find

$$d = \tfrac{1}{2}(3 \pm 1) = 1, 2.$$

Rewrite the above as

$$\tan\left(\frac{\pi}{8}\right) = \sqrt{2} - 1 \qquad (6)$$
$$\cot\left(\frac{\pi}{8}\right) = \frac{1}{\sqrt{2} - 1} = \frac{\sqrt{2} + 1}{2 - 1} = \sqrt{2} + 1. \qquad (7)$$

Summarizing,

$$\sin\left(\frac{\pi}{8}\right) = \tfrac{1}{2}\sqrt{2 - \sqrt{2}} \qquad (8)$$
$$\sin\left(\frac{3\pi}{8}\right) = \tfrac{1}{2}\sqrt{2 + \sqrt{2}} \qquad (9)$$
$$\cos\left(\frac{\pi}{8}\right) = \tfrac{1}{2}\sqrt{2 + \sqrt{2}} \qquad (10)$$
$$\cos\left(\frac{3\pi}{8}\right) = \tfrac{1}{2}\sqrt{2 - \sqrt{2}} \qquad (11)$$
$$\tan\left(\frac{\pi}{8}\right) = \sqrt{2} - 1 \qquad (12)$$
$$\tan\left(\frac{3\pi}{8}\right) = \sqrt{2} + 1. \qquad (13)$$

see also OCTAGON

Trigonometry Values—π/9

Trigonometric functions of $n\pi/9$ radians for n an integer not divisible by 3 (e.g., 40° and 80°) cannot be expressed in terms of sums, products, and finite root extractions on *real* rational numbers because 9 is not a product of distinct FERMAT PRIMES. This also means that the NONAGON is not a CONSTRUCTIBLE POLYGON.

However, exact expressions involving roots of *complex* numbers can still be derived using the trigonometric identity

$$\sin(3\alpha) = 3\sin\alpha - 4\sin^3\alpha. \qquad (1)$$

Let $\alpha \equiv \pi/9$ and $x \equiv \sin\alpha$. Then the above identity gives the CUBIC EQUATION

$$4x^3 - 3x + \tfrac{1}{2}\sqrt{3} = 0 \qquad (2)$$

$$x^3 - \tfrac{3}{4}x = -\tfrac{1}{8}\sqrt{3}. \qquad (3)$$

This cubic is of the form

$$x^3 + px = q, \qquad (4)$$

where

$$p = -\tfrac{3}{4} \qquad (5)$$
$$q = -\tfrac{1}{8}\sqrt{3}. \qquad (6)$$

The DISCRIMINANT is then

$$D \equiv \left(\frac{p}{3}\right)^3 + \left(\frac{q}{2}\right)^2$$
$$= \left(-\frac{1}{4}\right)^3 + \left(\frac{\sqrt{3}}{16}\right)^2 = -\frac{1}{16\cdot 4} + \frac{3}{16\cdot 16} = \frac{-4 + 3}{256}$$
$$= -\frac{1}{256} < 0. \qquad (7)$$

There are therefore three REAL distinct roots, which are approximately −0.9848, 0.3240, and 0.6428. We want the one in the first QUADRANT, which is 0.3240.

$$\begin{aligned}\sin\left(\frac{\pi}{9}\right) &= \sqrt[3]{-\frac{\sqrt{3}}{16}+\sqrt{-\frac{1}{256}}}+\sqrt[3]{-\frac{\sqrt{3}}{16}-\sqrt{-\frac{1}{256}}}\\ &= \sqrt[3]{-\frac{\sqrt{3}}{16}+\frac{1}{16}i}-\sqrt[3]{\frac{\sqrt{3}}{16}+\frac{1}{16}i}\\ &= 2^{-4/3}(\sqrt[3]{i-\sqrt{3}}-\sqrt[3]{i+\sqrt{3}})\\ &\approx 0.3240\ldots. \end{aligned} \tag{8}$$

Similarly,

$$\begin{aligned}\cos\left(\frac{\pi}{9}\right) &= 2^{-4/3}(\sqrt[3]{1+i\sqrt{3}}+\sqrt[3]{1-i\sqrt{3}})\\ &\approx 0.7660\ldots. \end{aligned} \tag{9}$$

Because of the NEWTON'S RELATIONS, we have the identities

$$\sin\left(\frac{\pi}{9}\right)\sin\left(\frac{2\pi}{9}\right)\sin\left(\frac{4\pi}{9}\right)=\tfrac{1}{8} \tag{10}$$

$$\cos\left(\frac{\pi}{9}\right)\cos\left(\frac{2\pi}{9}\right)\cos\left(\frac{4\pi}{9}\right)=\tfrac{1}{8}\sqrt{3} \tag{11}$$

$$\tan\left(\frac{\pi}{9}\right)\tan\left(\frac{2\pi}{9}\right)\tan\left(\frac{4\pi}{9}\right)=\sqrt{3}. \tag{12}$$

see also NONAGON, STAR OF GOLIATH

Trigonometry Values—$\pi/10$

$$\begin{aligned}\sin\left(\frac{\pi}{10}\right) &= \sin\left(\frac{1}{2}\cdot\frac{\pi}{5}\right)=\sqrt{\tfrac{1}{2}\left[1-\cos\left(\frac{\pi}{5}\right)\right]}\\ &= \sqrt{\tfrac{1}{2}[1-\tfrac{1}{4}(1+\sqrt{5})]}=\tfrac{1}{4}(\sqrt{5}-1). \end{aligned} \tag{1}$$

So we have

$$\begin{aligned}\cos\left(\frac{\pi}{10}\right) &= \cos\left(\frac{1}{2}\cdot\frac{\pi}{5}\right)=\sqrt{\frac{1}{2}\left[1+\cos\left(\frac{\pi}{5}\right)\right]}\\ &= \sqrt{\tfrac{1}{2}[1+\tfrac{1}{4}(1+\sqrt{5})]}\\ &= \tfrac{1}{4}\sqrt{10+2\sqrt{5}}, \end{aligned} \tag{2}$$

and

$$\tan\left(\frac{\pi}{10}\right)=\sqrt{\frac{3-\sqrt{5}}{5+\sqrt{5}}}=\tfrac{1}{5}\sqrt{25-10\sqrt{5}}. \tag{3}$$

Summarizing,

$$\sin\left(\frac{\pi}{10}\right)=\tfrac{1}{4}(\sqrt{5}-1) \tag{4}$$

$$\cos\left(\frac{\pi}{10}\right)=\tfrac{1}{4}\sqrt{10+2\sqrt{5}} \tag{5}$$

$$\tan\left(\frac{\pi}{10}\right)=\tfrac{1}{5}\sqrt{25-10\sqrt{5}} \tag{6}$$

$$\sin\left(\frac{3\pi}{10}\right)=\tfrac{1}{4}(1+\sqrt{5}) \tag{7}$$

$$\cos\left(\frac{3\pi}{10}\right)=\tfrac{1}{4}(10-2\sqrt{5}) \tag{8}$$

$$\tan\left(\frac{3\pi}{10}\right)=\tfrac{1}{5}\sqrt{25+10\sqrt{5}}. \tag{9}$$

An interesting near-identity is given by

$$\frac{1}{4}\left[\cos(\tfrac{1}{10})+\cosh(\tfrac{1}{10})+2\cos(\tfrac{1}{20}\sqrt{2})\cosh(\tfrac{1}{20}\sqrt{2})\right]\approx 1. \tag{10}$$

In fact, the left-hand side is approximately equal to $1+2.480\times 10^{-13}$.

see also DECAGON, DECAGRAM

Trigonometry Values—$\pi/11$

Trigonometric functions of $n\pi/11$ for n an integer cannot be expressed in terms of sums, products, and finite root extractions on *real* rational numbers because 11 is not a FERMAT PRIME. This also means that the UNDECAGON is not a CONSTRUCTIBLE POLYGON.

However, exact expressions involving roots of *complex* numbers can still be derived using the trigonometric identity

$$\begin{aligned}\sin(11\alpha) &= \sin(12\alpha-\alpha)\cos\alpha-\cos(12\alpha)\sin\alpha\\ &= 2\sin(6\alpha)\cos(6\alpha)\cos\alpha-[1-2\sin^2(6\alpha)]\sin\alpha. \end{aligned} \tag{1}$$

Using the identities from Beyer (1987, p. 139),

$$\sin(6\alpha)=\sin\alpha\cos\alpha[32\cos^4\alpha-32\cos^2\alpha+6] \tag{2}$$

$$\cos(6\alpha)=32\cos^6\alpha-48\cos^4\alpha+18\cos^2\alpha-1 \tag{3}$$

gives

$$\begin{aligned}\sin(11\alpha) &= 2\cos^2\alpha\sin\alpha(32\cos^4\alpha-32\cos^2\alpha+6)\\ &\quad\times(32\cos^6\alpha-48\cos^4\alpha+18\cos^2\alpha-1)\\ &\quad-\sin\alpha[1-2\sin^2\alpha\cos^2\alpha(32\cos^4\alpha-32\cos^2\alpha+6)^2]\\ &= \sin\alpha(11-220\sin^2\alpha+1232\sin^4\alpha\alpha\\ &\quad-2816\sin^6\alpha+2816\sin^8-1024\sin^{10}\alpha). \end{aligned} \tag{4}$$

Now, let $\alpha\equiv\pi/11$ and $x\equiv\sin^2\alpha$, then

$$\begin{aligned}\sin\pi &= 0\\ &= 11-220x+1232x^2-2816x^3+2816x^4-1024x^5. \end{aligned} \tag{5}$$

This equation is an irreducible QUINTIC EQUATION, so an analytic solution involving FINITE ROOT extractions does not exist. The numerical ROOTS are $x = 0.07937$, 0.29229, 0.57115, 0.82743, 0.97974. So $\sin\alpha = 0.2817$, $\sin(2\alpha) = 0.5406$, $\sin(3\alpha) = 0.7557$, $\sin(4\alpha) = 0.9096$, $\sin(5\alpha) = 0.9898$. From one of NEWTON'S IDENTITIES,

$$\sin\left(\frac{\pi}{11}\right)\sin\left(\frac{2\pi}{11}\right)\sin\left(\frac{3\pi}{11}\right)\sin\left(\frac{4\pi}{11}\right)\sin\left(\frac{5\pi}{11}\right) = \sqrt{\frac{11}{1024}} = \frac{\sqrt{11}}{32} \quad (6)$$

$$\cos\left(\frac{\pi}{11}\right)\cos\left(\frac{2\pi}{11}\right)\cos\left(\frac{3\pi}{11}\right)\cos\left(\frac{4\pi}{11}\right)\cos\left(\frac{5\pi}{11}\right) = \frac{1}{32} \quad (7)$$

$$\tan\left(\frac{\pi}{11}\right)\tan\left(\frac{2\pi}{11}\right)\tan\left(\frac{3\pi}{11}\right)\tan\left(\frac{4\pi}{11}\right)\tan\left(\frac{5\pi}{11}\right) = \sqrt{11}. \quad (8)$$

The trigonometric functions of $\pi/11$ also obey the identity

$$\tan\left(\frac{3\pi}{11}\right) + 4\sin\left(\frac{2\pi}{11}\right) = \sqrt{11}. \quad (9)$$

see also UNDECAGON

References

Beyer, W. H. "Trigonometry." *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, 1987.

Trigonometry Values—$\pi/12$

$$\begin{aligned}\sin\left(\frac{\pi}{12}\right) &= \sin\left(\frac{\pi}{3} - \frac{\pi}{4}\right)\\ &= -\sin\left(\frac{\pi}{4}\right)\cos\left(\frac{\pi}{3}\right) + \sin\left(\frac{\pi}{3}\right)\cos\left(\frac{\pi}{4}\right)\\ &= -\tfrac{1}{2}\sqrt{2}(\tfrac{1}{2}) + \tfrac{1}{2}\sqrt{3}(\tfrac{1}{2}\sqrt{2})\\ &= \tfrac{1}{4}(\sqrt{6} - \sqrt{2}). \quad (1)\end{aligned}$$

Similarly,

$$\begin{aligned}\cos\left(\frac{\pi}{12}\right) &= \cos\left(\frac{\pi}{3} - \frac{\pi}{4}\right)\\ &= \cos\left(\frac{\pi}{4}\right)\cos\left(\frac{\pi}{3}\right) - \sin\left(\frac{\pi}{3}\right)\sin\left(\frac{\pi}{4}\right)\\ &= \tfrac{1}{2}(\tfrac{1}{2}\sqrt{2}) + \tfrac{1}{2}\sqrt{3}(-\tfrac{1}{2}\sqrt{2})\\ &= \tfrac{1}{4}(\sqrt{6} + \sqrt{2}). \quad (2)\end{aligned}$$

Summarizing,

$$\sin\left(\frac{\pi}{12}\right) = \tfrac{1}{4}(\sqrt{6} - \sqrt{2}) \approx 0.25881 \quad (3)$$

$$\cos\left(\frac{\pi}{12}\right) = \tfrac{1}{4}(\sqrt{6} + \sqrt{2}) \approx 0.96592 \quad (4)$$

$$\tan\left(\frac{\pi}{12}\right) = 2 - \sqrt{3} \approx 0.26794 \quad (5)$$

$$\csc\left(\frac{\pi}{12}\right) = \sqrt{6} + \sqrt{2} \approx 3.86370 \quad (6)$$

$$\sec\left(\frac{\pi}{12}\right) = \sqrt{6} - \sqrt{2} \approx 1.03527 \quad (7)$$

$$\cot\left(\frac{\pi}{12}\right) = 2 + \sqrt{3} \approx 3.73205. \quad (8)$$

Trigonometry Values—$\pi/15$

$$\begin{aligned}\sin\left(\frac{\pi}{15}\right) &= \sin\left(\frac{\pi}{6} - \frac{\pi}{10}\right)\\ &= \sin\left(\frac{\pi}{6}\right)\cos\left(\frac{\pi}{10}\right) - \sin\left(\frac{\pi}{10}\right)\cos\left(\frac{\pi}{6}\right)\\ &= \frac{1}{2}\sqrt{\frac{1}{8}(5+\sqrt{5})} - \frac{\sqrt{3}}{2}\frac{1}{4}(\sqrt{5} - 1)\\ &= \tfrac{1}{16}(2\sqrt{3} - 2\sqrt{15} + \sqrt{40 + 8\sqrt{5}}) \quad (1)\end{aligned}$$

and

$$\begin{aligned}\cos\left(\frac{\pi}{15}\right) &= \cos\left(\frac{\pi}{6} - \frac{\pi}{10}\right)\\ &= \cos\left(\frac{\pi}{6}\right)\cos\left(\frac{\pi}{10}\right) + \sin\left(\frac{\pi}{6}\right)\sin\left(\frac{\pi}{10}\right)\\ &= \frac{\sqrt{3}}{2}\sqrt{\frac{1}{8}(5+\sqrt{5})} + \frac{1}{2}\frac{1}{4}(\sqrt{5} - 1)\\ &= \tfrac{1}{8}(\sqrt{30 + 6\sqrt{5}} + \sqrt{5} - 1). \quad (2)\end{aligned}$$

Summarizing,

$$\begin{aligned}\sin\left(\frac{\pi}{15}\right) &= \tfrac{1}{16}(2\sqrt{3} - 2\sqrt{15} + \sqrt{40 + 8\sqrt{5}})\\ &\approx 0.20791 \quad (3)\end{aligned}$$

$$\begin{aligned}\sin\left(\frac{2\pi}{15}\right) &= \tfrac{1}{8}(\sqrt{3} + \sqrt{15} - \sqrt{10 - 2\sqrt{5}})\\ &\approx 0.40673 \quad (4)\end{aligned}$$

$$\cos\left(\frac{\pi}{15}\right) = \tfrac{1}{8}(\sqrt{30 + 6\sqrt{5}} + \sqrt{5} - 1) \approx 0.97814 \quad (5)$$

$$\cos\left(\frac{2\pi}{15}\right) = \tfrac{1}{8}(\sqrt{30 - 6\sqrt{5}} + 1) \approx 0.91354 \quad (6)$$

$$\begin{aligned}\tan\left(\frac{\pi}{15}\right) &= \tfrac{1}{2}(3\sqrt{3} - \sqrt{15} - \sqrt{50 - 22\sqrt{5}})\\ &\approx 0.21255. \quad (7)\end{aligned}$$

Trigonometry Values—$\pi/16$

$$\begin{aligned}\sin\left(\frac{\pi}{16}\right) &= \sin\left(\frac{1}{2}\cdot\frac{\pi}{8}\right)\\ &= \sqrt{\tfrac{1}{2}\left(1 - \cos\frac{\pi}{8}\right)} = \sqrt{\tfrac{1}{2}\left(1 - \tfrac{1}{2}\sqrt{2+\sqrt{2}}\right)}\\ &= \sqrt{\tfrac{1}{2} - \tfrac{1}{4}\sqrt{2+\sqrt{2}}} = \tfrac{1}{2}\sqrt{2 - \sqrt{2+\sqrt{2}}} \quad (1)\end{aligned}$$

$$\begin{aligned}\cos\left(\frac{\pi}{16}\right) &= \cos\left(\frac{1}{2}\cdot\frac{\pi}{8}\right)\\ &= \sqrt{\frac{1}{2}\left(1 + \cos\frac{\pi}{8}\right)} = \sqrt{\tfrac{1}{2}\left(1 + \tfrac{1}{2}\sqrt{2+\sqrt{2}}\right)}\\ &= \sqrt{\tfrac{1}{2} + \tfrac{1}{4}\sqrt{2+\sqrt{2}}} = \tfrac{1}{2}\sqrt{2 + \sqrt{2+\sqrt{2}}} \quad (2)\end{aligned}$$

$$\begin{aligned}\tan\left(\frac{\pi}{16}\right) &= \sqrt{\frac{2 - \sqrt{2+\sqrt{2}}}{2 + \sqrt{2+\sqrt{2}}}}\\ &= \sqrt{4 + 2\sqrt{2}} - \sqrt{2} - 1. \quad (3)\end{aligned}$$

Summarizing,

$$\sin\left(\frac{\pi}{16}\right) = \tfrac{1}{2}\sqrt{2-\sqrt{2+\sqrt{2}}} \approx 0.19509 \tag{4}$$

$$\sin\left(\frac{3\pi}{16}\right) = \tfrac{1}{2}\sqrt{2-\sqrt{2-\sqrt{2}}} \approx 0.55557 \tag{5}$$

$$\cos\left(\frac{\pi}{16}\right) = \tfrac{1}{2}\sqrt{2+\sqrt{2+\sqrt{2}}} \approx 0.98079 \tag{6}$$

$$\cos\left(\frac{3\pi}{16}\right) = \tfrac{1}{2}\sqrt{2+\sqrt{2-\sqrt{2}}} \approx 0.83147 \tag{7}$$

$$\tan\left(\frac{\pi}{16}\right) = \sqrt{4+2\sqrt{2}} - \sqrt{2} - 1 \approx 0.19891. \tag{8}$$

Trigonometry Values—$\pi/17$

Rather surprisingly, trigonometric functions of $n\pi/17$ for n an integer can be expressed in terms of sums, products, and finite root extractions because 17 is a FERMAT PRIME. This makes the HEPTADECAGON a CONSTRUCTIBLE, as first proved by Gauss. Although Gauss did not actually explicitly provide a construction, he did derive the trigonometric formulas below using a series of intermediate variables from which the final expressions were then built up.

Let

$$\epsilon \equiv \sqrt{17+\sqrt{17}}$$

$$\epsilon^* \equiv \sqrt{17-\sqrt{17}}$$

$$\alpha \equiv \sqrt{\sqrt{34}+6\sqrt{17}+(\sqrt{34}-\sqrt{2}\,)\epsilon^* - 8\sqrt{2}\,\epsilon}$$

$$\beta \equiv 2\sqrt{17+3\sqrt{17}-2\sqrt{2}\,\epsilon-\sqrt{2}\,\epsilon^*}\,.$$

Then

$$\sin\left(\frac{\pi}{17}\right) = \tfrac{1}{8}[34 - 2\sqrt{17} - 2\sqrt{2}\,\epsilon^* - 2\sqrt{68+12\sqrt{17}+2\sqrt{2}(\sqrt{17}-1)\epsilon^* - 16\sqrt{2}\,\epsilon}\,]^{1/2} \approx 0.18375$$

$$\cos\left(\frac{\pi}{17}\right) = \tfrac{1}{8}[30 + 2\sqrt{17} + 2\sqrt{2}\,\epsilon^* + 2\sqrt{68+12\sqrt{17}+2\sqrt{2}(\sqrt{17}-1)\epsilon^* - 16\sqrt{2}\,\epsilon}\,]^{1/2} \approx 0.98297$$

$$\sin\left(\frac{2\pi}{17}\right) = \tfrac{1}{16}[136 - 8\sqrt{17} + 4\sqrt{2}(1-\sqrt{17})\epsilon^* + 16\sqrt{2}\,\epsilon + 2(\sqrt{2}-\sqrt{34}-2\epsilon^*)\sqrt{34+6\sqrt{17}+(\sqrt{34}-\sqrt{2})\epsilon^* - 8\sqrt{2}\epsilon}\,]^{1/2} \approx 0.36124$$

$$\cos\left(\frac{2\pi}{17}\right) = \tfrac{1}{16}[-1 + \sqrt{17} + \sqrt{2}\epsilon^* + \sqrt{68+12\sqrt{17}-2\sqrt{2}(1-\sqrt{17}\,)\epsilon^* - 16\sqrt{2}\,\epsilon}\,] \approx 0.0.93247$$

$$\sin\left(\frac{4\pi}{17}\right) = \tfrac{1}{128}(-\sqrt{2}+\sqrt{34}+2\epsilon^*+2\alpha) \times\sqrt{68-4\sqrt{17}-2(\sqrt{34}-\sqrt{2}\,)\epsilon^* + 8\sqrt{2}\,\epsilon + \alpha(\sqrt{2}-\sqrt{34}-2\epsilon^*)} \approx 0.0.67370$$

$$\sin\left(\frac{8\pi}{17}\right) = \tfrac{1}{16}[136 - 8\sqrt{17} + 8\sqrt{2}\epsilon - 2(\sqrt{34}-3\sqrt{2})\epsilon^* - 2\beta(1-\sqrt{17}-\sqrt{2}\epsilon^*)\,]^{1/2} \approx 0.99573$$

$$\cos\left(\frac{8\pi}{17}\right) = \tfrac{1}{16}(-1+\sqrt{17}+\sqrt{2}\,\epsilon^* - 2\sqrt{17+3\sqrt{17}-\sqrt{2}\,\epsilon^*-2\sqrt{2}\,\epsilon}\,). \approx 0.09227$$

There are some interesting analytic formulas involving the trigonometric functions of $n\pi/17$. Define

$$P(x) \equiv (x-1)(x-2)(x^2+1)$$

$$g_1(x) \equiv \frac{2+\sqrt{P(x)}}{1-x}$$

$$g_4(x) \equiv \frac{2-\sqrt{P(x)}}{1-x}$$

$$f_i(x) \equiv \tfrac{1}{4}[g_i(x)-1]$$

$$a \equiv \tfrac{1}{4}\tan^{-1}4,$$

where $i = 1$ or 4. Then

$$f_1(\tan a) = \cos\left(\frac{2\pi}{17}\right)$$

$$f_4(\tan a) = \cos\left(\frac{8\pi}{17}\right).$$

see also CONSTRUCTIBLE POLYGON, FERMAT PRIME, HEPTADECAGON

References

Casey, J. *Plane Trigonometry.* Dublin: Hodges, Figgis, & Co., p. 220, 1888.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 192–194 and 229–230, 1996.

Dörrie, H. "The Regular Heptadecagon." §37 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 177–184, 1965.

Ore, Ø. *Number Theory and Its History.* New York: Dover, 1988.

Smith, D. E. *A Source Book in Mathematics.* New York: Dover, p. 348, 1994.

Trigonometry Values—$\pi/18$

The exact values of $\cos(\pi/18)$ and $\sin(\pi/18)$ are given by infinite NESTED RADICALS.

$$\sin\left(\frac{\pi}{18}\right) = \tfrac{1}{2}\sqrt{2-\sqrt{2+\sqrt{2+\sqrt{2-\ldots}}}} \approx 0.17365,$$

where the sequence of signs $+$, $+$, $-$ repeats with period 3, and

$$\cos\left(\frac{\pi}{18}\right) = \tfrac{1}{6}\sqrt{3}\left(\sqrt{8-\sqrt{8-\sqrt{8+\sqrt{8-\ldots}}}}+1\right) \approx 0.98481,$$

where the sequence of signs $-$, $-$, $+$ repeats with period 3.

Trigonometry Values—$\pi/20$

$$\begin{aligned}\sin\left(\frac{\pi}{20}\right) &= \sin\left(\frac{1}{2}\frac{\pi}{10}\right) = \sqrt{\frac{1}{2}\left(1-\cos\frac{\pi}{10}\right)}\\ &= \tfrac{1}{4}\sqrt{8-2\sqrt{10+2\sqrt{5}}}\\ &\approx 0.15643\ldots \qquad (1)\\ \cos\left(\frac{\pi}{20}\right) &= \cos\left(\frac{1}{2}\frac{\pi}{10}\right) = \sqrt{\frac{1}{2}\left(1+\cos\frac{\pi}{10}\right)}\\ &= \tfrac{1}{4}\sqrt{8+2\sqrt{10+2\sqrt{5}}}\\ &\approx 0.98768\ldots \qquad (2)\\ \tan\left(\frac{\pi}{20}\right) &= 1+\sqrt{5}-\sqrt{5+2\sqrt{5}}\\ &\approx 0.15838. \qquad (3)\end{aligned}$$

An interesting near-identity is given by

$$\frac{1}{4}\left[\cos(\tfrac{1}{10})+\cosh(\tfrac{1}{10})+2\cos(\tfrac{1}{20}\sqrt{2})\cosh(\tfrac{1}{20}\sqrt{2})\right]\approx 1. \qquad (4)$$

In fact, the left-hand side is approximately equal to $1 + 2.480 \times 10^{-13}$.

Trigonometry Values—0

By the definition of the trigonometric functions,

$$\begin{aligned}\sin 0 &= 0\\ \cos 0 &= 1\\ \tan 0 &= 0\\ \csc 0 &= \infty\\ \sec 0 &= 1\\ \cot 0 &= \infty.\end{aligned}$$

Trigyrate Rhombicosidodecahedron

see JOHNSON SOLID

Trihedron

The TRIPLE of unit ORTHOGONAL VECTORS **T**, **N**, and **B** (TANGENT VECTOR, NORMAL VECTOR, and BINORMAL VECTOR).

see also BINORMAL VECTOR, NORMAL VECTOR, TANGENT VECTOR

Trilinear Coordinates

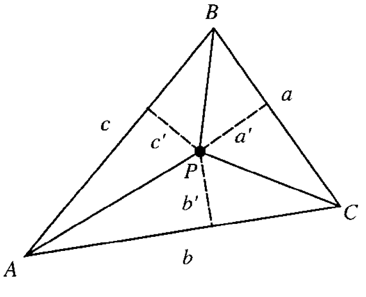

Given a TRIANGLE ΔABC, the trilinear coordinates of a point P with respect to ΔABC are an ordered TRIPLE of numbers, each of which is PROPORTIONAL to the directed distance from P to one of the side lines. Trilinear coordinates are denoted $\alpha : \beta : \gamma$ or (α, β, γ) and also are known as BARYCENTRIC COORDINATES, HOMOGENEOUS COORDINATES, or "trilinears."

In trilinear coordinates, the three VERTICES A, B, and C are given by $1:0:0$, $0:1:0$, and $0:0:1$. Let the point P in the above diagram have trilinear coordinates $\alpha : \beta : \gamma$ and lie at distances a', b', and c' from the sides BC, AC, and AB, respectively. Then the distances $a' = k\alpha$, $b' = k\beta$, and $c' = k\gamma$ can be found by writing Δ_a for the AREA of ΔBPC, and similarly for Δ_b and Δ_c. We then have

$$\begin{aligned}\Delta &= \Delta_a + \Delta_b + \Delta_c = \tfrac{1}{2}aa' + \tfrac{1}{2}bb' + \tfrac{1}{2}cc'\\ &= \tfrac{1}{2}(ak\alpha + bk\beta + ck\gamma) = \tfrac{1}{2}k(a\alpha + b\beta + c\gamma). \qquad (1)\end{aligned}$$

so

$$k \equiv \frac{2\Delta}{a\alpha + b\beta + c\gamma}, \qquad (2)$$

where Δ is the AREA of ΔABC and a, b, and c are the lengths of its sides. When the values of the coordinates are taken as the actual lengths (i.e., the trilinears are chosen so that $k = 1$), the coordinates are known as EXACT TRILINEAR COORDINATES.

Trilinear coordinates are unchanged when each is multiplied by any constant μ, so

$$t_1 : t_2 : t_3 = \mu t_1 : \mu t_2 : \mu t_3. \qquad (3)$$

When normalized so that

$$t_1 + t_2 + t_3 = 1, \qquad (4)$$

trilinear coordinates are called AREAL COORDINATES. The trilinear coordinates of the line

$$ux + vy + wz = 0 \qquad (5)$$

are

$$u : v : w = ad_A : bd_B : cd_C, \qquad (6)$$

where d_i is the POINT-LINE DISTANCE from VERTEX A to the LINE.

Trilinear coordinates for some common POINTS are summarized in the following table, where A, B, and C are the angles at the corresponding vertices and a, b, and c are the opposite side lengths.

Point	Triangle Center Function
centroid M	$\csc A$, $1/a$
circumcenter O	$\cos A$
de Longchamps point	$\cos A - \cos B \cos C$
equal detour point	$\sec(\frac{1}{2}A)\cos(\frac{1}{2}B)\cos(\frac{1}{2}C)+1$
Feuerbach point F	$1-\cos(B-C)$
incenter I	1
isoperimetric point	$\sec(\frac{1}{2}A)\cos(\frac{1}{2}B)\cos(\frac{1}{2}C)-1$
Lemoine point	a
nine-point center N	$\cos(B-C)$
orthocenter H	$\cos B \cos C$
vertex A	$1:0:0$
vertex B	$0:1:0$
vertex C	$0:0:1$

To convert trilinear coordinates to a vector position for a given triangle specified by the x- and y-coordinates of its axes, pick two UNIT VECTORS along the sides. For instance, pick

$$\hat{\mathbf{a}} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{7}$$

$$\hat{\mathbf{c}} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}, \tag{8}$$

where these are the UNIT VECTORS BC and AB. Assume the TRIANGLE has been labeled such that $A = \mathbf{x}_1$ is the lower rightmost VERTEX and $C = \mathbf{x}_2$. Then the VECTORS obtained by traveling l_a and l_c along the sides and then inward PERPENDICULAR to them must meet

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + l_c \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} - k\gamma \begin{bmatrix} c_2 \\ -c_1 \end{bmatrix} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} + l_a \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} - k\alpha \begin{bmatrix} a_2 \\ -a_1 \end{bmatrix}. \tag{9}$$

Solving the two equations

$$x_1 + l_c c_1 - k\gamma c_2 = x_2 l_a a_1 - k\alpha a_2 \tag{10}$$

$$y_1 + l_c c_2 + k\gamma c_1 = y_2 l_a a_2 + k\alpha a_1, \tag{11}$$

gives

$$l_c = \frac{k\alpha(a_1c_1+a_2c_2) - \gamma k({c_1}^2+{c_2}^2) + c_2(x_1-x_2) + c_1(y_2-y_1)}{a_1c_2 - a_2c_1} \tag{12}$$

$$l_a = \frac{k\alpha({a_1}^2+{a_2}^2) - \gamma k(a_1c_1+a_2c_2) + a_2(x_1-x_2) + a_1(y_2-y_1)}{a_1c_2 - a_2c_1}. \tag{13}$$

But $\hat{\mathbf{a}}$ and $\hat{\mathbf{c}}$ are UNIT VECTORS, so

$$l_c = \frac{k\alpha(a_1c_1+a_2c_2) - \gamma k + c_2(x_1-x_2) + c_1(y_2-y_1)}{a_1c_2 - a_2c_1} \tag{14}$$

$$l_a = \frac{k\alpha - \gamma k(a_1c_1+a_2c_2) + a_2(x_1-x_2) + a_1(y_2-y_1)}{a_1c_2 - a_2c_1}. \tag{15}$$

And the VECTOR coordinates of the point $\alpha : \beta : \gamma$ are then

$$\mathbf{x} = \mathbf{x}_1 + l_c \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} - k\gamma \begin{bmatrix} c_2 \\ -c_1 \end{bmatrix}. \tag{16}$$

see also AREAL COORDINATES, EXACT TRILINEAR COORDINATES, ORTHOCENTRIC COORDINATES, POWER CURVE, QUADRIPLANAR COORDINATES, TRIANGLE, TRILINEAR POLAR

References

Boyer, C. B. *History of Analytic Geometry.* New York: Yeshiva University, 1956.

Casey, J. "The General Equation—Trilinear Co-Ordinates." Ch. 10 in *A Treatise on the Analytical Geometry of the Point, Line, Circle, and Conic Sections, Containing an Account of Its Most Recent Extensions, with Numerous Examples, 2nd ed., rev. enl.* Dublin: Hodges, Figgis, & Co., pp. 333–348, 1893.

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, pp. 67–71, 1959.

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, 1969.

Coxeter, H. S. M. "Some Applications of Trilinear Coordinates." *Linear Algebra Appl.* **226–228**, 375–388, 1995.

Kimberling, C. "Triangle Centers and Central Triangles." *Congr. Numer.* **129**, 1–295, 1998.

Trilinear Line

A LINE is given in TRILINEAR COORDINATES by

$$l\alpha + m\beta + n\gamma = 0.$$

see also LINE, TRILINEAR COORDINATES

Trilinear Polar

Given a TRIANGLE CENTER $X = l : m : n$, the line

$$l\alpha + m\beta + n\gamma = 0$$

is called the trilinear polar of X^{-1} and is denoted L.

see also CHASLES'S POLARS THEOREM

Trillion

The word trillion denotes different numbers in American and British usage. In the American system, one trillion equals 10^{12}. In the British, French, and German systems, one trillion equals 10^{18}.

see also BILLION, LARGE NUMBER, MILLION

Trimagic Square

If replacing each number by its square or cube in a MAGIC SQUARE produces another MAGIC SQUARE, the square is said to be a trimagic square. Trimagic squares of order 32, 64, 81, and 128 are known. Tarry gave a method for constructing a trimagic square of order 128, Cazalas a method for trimagic squares of orders 64 and 81, and R. V. Heath a method for constructing an order 64 trimagic square which is different from Cazalas's (Kraitchik 1942).

Trimagic squares are also called TREBLY MAGIC SQUARES, and are 3-MULTIMAGIC SQUARES.

see also BIMAGIC SQUARE, MAGIC SQUARE, MULTIMAGIC SQUARE

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 212–213, 1987.

Kraitchik, M. "Multimagic Squares." §7.10 in *Mathematical Recreations.* New York: W. W. Norton, pp. 176–178, 1942.

Trimean

The trimean is defined to be

$$TM \equiv \tfrac{1}{4}(H_1 + 2M + H_2),$$

where H_i are the HINGES and M is the MEDIAN. Press *et al.* (1992) call this TUKEY'S TRIMEAN. It is an L-ESTIMATE.

see also HINGE, L-ESTIMATE, MEAN, MEDIAN (STATISTICS)

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 694, 1992.

Tukey, J. W. *Explanatory Data Analysis.* Reading, MA: Addison-Wesley, pp. 46–47, 1977.

Trimorphic Number

A number n such that the last digits of n^3 are the same as n. 49 is trimorphic since $49^3 = 117649$ (Wells 1986, p. 124). The first few are 1, 4, 5, 6, 9, 24, 25, 49, 51, 75, 76, 99, 125, 249, 251, 375, 376, 499, ...

see also AUTOMORPHIC NUMBER, NARCISSISTIC NUMBER, SUPER-3 NUMBER

References

Wells, D. *The Penguin Dictionary of Curious and Interesting Numbers.* Middlesex, England: Penguin Books, 1986.

Trinoid

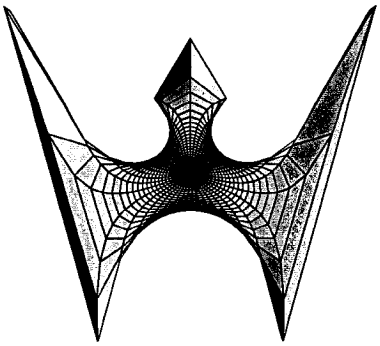

A MINIMAL SURFACE discovered by L. P. M. Jorge and W. Meeks III in 1983 with ENNEPER-WEIERSTRAß PARAMETERIZATION

$$f = \frac{1}{(\zeta^3 - 1)^2} \tag{1}$$

$$g = \zeta^2 \tag{2}$$

(Dickson 1990). Explicitly, it is given by

$$x = \Re\left[\frac{re^{i\theta}}{3(1 + re^{i\theta} + r^2e^{2i\theta})} - \frac{4\ln(re^{i\theta} - 1)}{9} + \frac{2\ln(1 + re^{i\theta} + r^2e^{2i\theta})}{9}\right] \tag{3}$$

$$y = -\tfrac{1}{9}\Im\left[\frac{-3re^{i\theta}(1 + re^{i\theta})}{r^3e^{3i\theta} - 1} + \frac{4\sqrt{3}\,(r^3e^{3i\theta} - 1)\tan^{-1}\left(\frac{1+2re^{i\theta}}{\sqrt{3}}\right)}{r^3e^{3i\theta} - 1}\right] \tag{4}$$

$$z = \Re\left[-\frac{2}{3} - \frac{2}{3(r^3e^{3i\theta} - 1)}\right], \tag{5}$$

for $\theta \in [0, 2\pi)$ and $r \in [0, 4]$.

see also MINIMAL SURFACE

References

Dickson, S. "Minimal Surfaces." *Mathematica J.* **1**, 38–40, 1990.

Wolfram Research "Mathematica Version 2.0 Graphics Gallery." `http://www.mathsource.com/cgi-bin/MathSource/Applications/Graphics/3D/0207-155`.

Trinomial

A POLYNOMIAL with three terms.

see also BINOMIAL, MONOMIAL, POLYNOMIAL

Trinomial Identity

$$(x^2 + axy + by^2)(t^2 + atu + bu^2) = r^2 + ars + bs^2, \tag{1}$$

where

$$r = xt - byu \tag{2}$$

$$s = yt + xu + ayu. \tag{3}$$

Trinomial Triangle

The NUMBER TRIANGLE obtained by starting with a row containing a single "1" and the next row containing three 1s and then letting subsequent row elements be computed by summing the elements above to the left, directly above, and above to the right:

$$\begin{array}{ccccccccc} & & & & 1 & & & & \\ & & & 1 & 1 & 1 & & & \\ & & 1 & 2 & 3 & 2 & 1 & & \\ & 1 & 3 & 6 & 7 & 6 & 3 & 1 & \\ 1 & 4 & 10 & 16 & 19 & 16 & 10 & 4 & 1 \end{array}$$

(Sloane's A027907). The nth row can also be obtained by expanding $(1+x+x^2)^n$ and taking coefficients:

$$\begin{aligned} (1+x+x^2)^0 &= 1 \\ (1+x+x^2)^1 &= 1+x+x^2 \\ (1+x+x^2)^2 &= 1+2x+3x^2+2x^3+x^4 \end{aligned}$$

and so on.

see also PASCAL'S TRIANGLE

References

Sloane, N. J. A. Sequence A027907 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Triomino

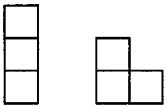

The two 3-POLYOMINOES are called triominoes, and are also known as the TROMINOES. The left triomino above is "STRAIGHT," while the right triomino is called "right" or L-.

see also L-POLYOMINO, POLYOMINO, STRAIGHT POLYOMINO

References

Gardner, M. "Polyominoes." Ch. 13 in *The Scientific American Book of Mathematical Puzzles & Diversions.* New York: Simon and Schuster, pp. 124–140, 1959.

Hunter, J. A. H. and Madachy, J. S. *Mathematical Diversions.* New York: Dover, pp. 80–81, 1975.

Lei, A. "Tromino." `http://www.cs.ust.hk/~philipl/omino/tromino.html`

Trip-Let

A 3-dimensional solid which is shaped in such a way that its projections along three mutually perpendicular axes are three different letters of the alphabet. Hofstadter (1989) has constructed such a solid for the letters G, E, and B.

see also CORK PLUG

References

Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, cover and pp. xiv, 1, and 273, 1989.

Triple

A group of three elements, also called a TRIAD.

see also AMICABLE TRIPLE, MONAD, PAIR, PYTHAGOREAN TRIPLE, QUADRUPLET, QUINTUPLET, TETRAD, TRIAD, TWINS

Triple-Free Set

A SET of POSITIVE integers is called weakly triple-free if, for any integer x, the SET $\{x, 2x, 3x\} \not\subset S$. It is called strongly triple-free if $x \in S$ IMPLIES $2x \notin S$ and $3x \notin S$. Define

$$\begin{aligned} p(n) &= \max\{|S| : S \subset \{1, 2, \ldots, n\} \\ &\qquad \text{is weakly triple-free}\} \\ q(n) &= \max\{|S| : S \subset \{1, 2, \ldots, n\} \\ &\qquad \text{is strongly triple-free}\}, \end{aligned}$$

where $|S|$ denotes the CARDINALITY of S, then

$$\lim_{n\to\infty} \frac{p(n)}{n} \geq \tfrac{4}{5}$$

and

$$\lim_{n\to\infty} \frac{q(n)}{n} = 0.6134752692\ldots$$

(Finch).

see also DOUBLE-FREE SET

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/triple/triple.html`.

Triple Jacobi Product

see JACOBI TRIPLE PRODUCT

Triple Point

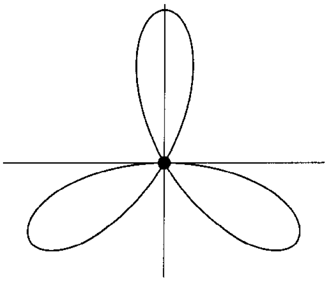

A point where a curve intersects itself along three arcs. The above plot shows the triple point at the ORIGIN of the TRIFOLIUM $(x^2+y^2)^2+3x^2y-y^3=0$.

see also DOUBLE POINT, QUADRUPLE POINT

References

Walker, R. J. *Algebraic Curves.* New York: Springer-Verlag, pp. 57–58, 1978.

Triple Scalar Product

see SCALAR TRIPLE PRODUCT

Triple Vector Product

see VECTOR TRIPLE PRODUCT

Triplet

see TRIPLE

Triplicate-Ratio Circle

see LEMOINE CIRCLE

Trisected Perimeter Point

A triangle center which has a TRIANGLE CENTER FUNCTION

$$\alpha = bc(v - c + a)(v - a + b),$$

where v is the unique REAL ROOT of

$$2x^3 - 3(a+b+c)x^2 + (a^2+b^2+c^2+8bc+8ca+8ab)x \\ -(b^2c+c^2a+a^2b+5bc^2+5ca^2+5ab^2+9abc) = 0.$$

References

Kimberling, C. "Central Points and Central Lines in the Plane of a Triangle." *Math. Mag.* **67**, 163–187, 1994.

Trisection

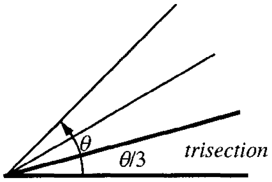

Angle trisection is the division of an *arbitrary* ANGLE into three equal ANGLES. It was one of the three GEOMETRIC PROBLEMS OF ANTIQUITY for which solutions using only COMPASS and STRAIGHTEDGE were sought. The problem was algebraically proved impossible by Wantzel (1836).

Although trisection is not possible for a general ANGLE using a Greek construction, there are some specific angles, such as $\pi/2$ and π radians (90° and 180°, respectively), which *can* be trisected. Furthermore, some ANGLES are geometrically trisectable, but cannot be constructed in the first place, such as $3\pi/7$ (Honsberger 1991). In addition, trisection of an arbitrary angle *can* be accomplished using a *marked* RULER (a NEUSIS CONSTRUCTION) as illustrated below (Courant and Robbins 1996).

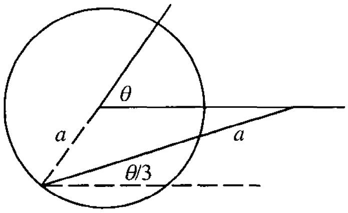

An ANGLE can also be divided into three (or any WHOLE NUMBER) of equal parts using the QUADRATRIX OF HIPPIAS or TRISECTRIX.

see also ANGLE BISECTOR, MACLAURIN TRISECTRIX, QUADRATRIX OF HIPPIAS, TRISECTRIX

References

Bogomolny, A. "Angle Trisection." `http://www.cut-the-knot.com/pythagoras/archi.html`.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 190–191, 1996.

Courant, R. and Robbins, H. "Trisecting the Angle." §3.3.3 in *What is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd ed.* Oxford, England: Oxford University Press, pp. 137–138, 1996.

Coxeter, H. S.M. "Angle Trisection." §2.2 in *Introduction to Geometry, 2nd ed.* New York: Wiley, p. 28, 1969.

Dixon, R. *Mathographics.* New York: Dover, pp. 50–51, 1991.

Dörrie, H. "Trisection of an Angle." §36 in *100 Great Problems of Elementary Mathematics: Their History and Solutions.* New York: Dover, pp. 172–177, 1965.

Dudley, U. *The Trisectors.* Washington, DC: Math. Assoc. Amer., 1994.

Geometry Center. "Angle Trisection." `http://www.geom.umn.edu:80/docs/forum/angtri/`.

Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., pp. 25–26, 1991.

Ogilvy, C. S. "Angle Trisection." *Excursions in Geometry.* New York: Dover, pp. 135–141, 1990.

Wantzel, P. L. "Recherches sur les moyens de reconnaître si un Problème de Géométrie peut se résoudre avec la règle et le compas." *J. Math. pures appliq.* **1**, 366–372, 1836.

Trisectrix

see CATALAN'S TRISECTRIX, MACLAURIN TRISECTRIX

Trisectrix of Catalan

see CATALAN'S TRISECTRIX

Trisectrix of Maclaurin

see MACLAURIN TRISECTRIX

Triskaidecagon

see TRIDECAGON

Triskaidekaphobia

The number 13 is traditionally associated with bad luck. This superstition leads some people to fear or avoid anything involving this number, a condition known as triskaidekaphobia. Triskaidekaphobia leads to interesting practices such as the numbering of floors as 1, 2, ..., 11, 12, 14, 15, ..., *omitting the number 13,* in many high-rise hotels.

see also 13, BAKER'S DOZEN, FRIDAY THE THIRTEENTH, TRISKAIDEKAPHOBIA

Tritangent

The tritangent of a CUBIC SURFACE is a PLANE which intersects the surface in three mutually intersecting lines. Each intersection of two lines is then a tangent point of the surface.

see also CUBIC SURFACE

References

Hunt, B. "Algebraic Surfaces." `http://www.mathematik.uni-kl.de/~wwwagag/Galerie.html`.

Tritangent Triangle

see EXCENTRAL TRIANGLE

Trivial

According to the Nobel Prize-winning physicist Richard Feynman (1985), mathematicians designate any THEOREM as "trivial" once a proof has been obtained—no matter how difficult the theorem was to prove in the first place. There are therefore exactly two types of true mathematical propositions: trivial ones, and those which have not yet been proven.

see also PROOF, THEOREM

References

Feynman, R. P. and Leighton, R. *Surely You're Joking, Mr. Feynman!* New York: Bantam Books, 1985.

Trivialization

In the definition of a FIBER BUNDLE $f : E \to B$, the homeomorphisms $g_U : f^{-1}(U) \to U \times F$ that commute with projection are called local trivializations for the FIBER BUNDLE f.

see also FIBER BUNDLE

Trochoid

The curve described by a point at a distance b from the center of a rolling CIRCLE of RADIUS a.

$$x = a\phi - b\sin\phi$$
$$y = a - b\cos\phi.$$

If $b < a$, the curve is a CURTATE CYCLOID. If $b = a$, the curve is a CYCLOID. If $b > a$, the curve is a PROLATE CYCLOID.

see also CURTATE CYCLOID, CYCLOID, PROLATE CYCLOID

References

Lee, X. "Trochoid." `http://www.best.com/~xah/SpecialPlaneCurves_dir/Trochoid_dir/trochoid.html`.

Wagon, S. *Mathematica in Action.* New York: W. H. Freeman, pp. 46–50, 1991.

Yates, R. C. "Trochoids." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 233–236, 1952.

Tromino

see TRIOMINO

True

A statement which is rigorously known to be correct. A statement which is not true is called FALSE, although certain statements can be proved to be rigorously UNDECIDABLE within the confines of a given set of assumptions and definitions. Regular two-valued LOGIC allows statements to be only true or FALSE, but FUZZY LOGIC treats "truth" as a continuum which can have any value between 0 and 1.

see also ALETHIC, FALSE, FUZZY LOGIC, LOGIC, TRUTH TABLE, UNDECIDABLE

Truncate

To truncate a REAL NUMBER is to remove its nonintegral part. Truncation of a number x therefore corresponds to taking the FLOOR FUNCTION $\lfloor x \rfloor$.

see also CEILING FUNCTION, FLOOR FUNCTION, NINT, ROUND

Truncated Cone

see CONICAL FRUSTUM

Truncated Cube

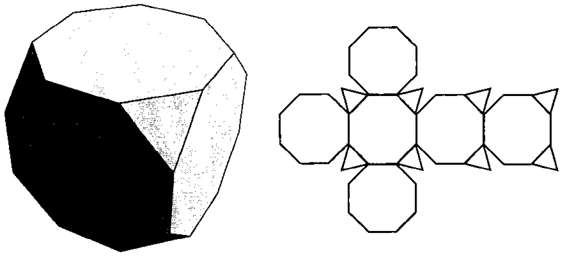

An ARCHIMEDEAN SOLID whose DUAL POLYHEDRON is the TRIAKIS OCTAHEDRON. It has SCHLÄFLI SYMBOL t{4,3}. It is also UNIFORM POLYHEDRON U_9 and has WYTHOFF SYMBOL $2\,3\,|\,4$. Its faces are $8\{3\}+6\{8\}$. The INRADIUS, MIDRADIUS, and CIRCUMRADIUS for $a = 1$ are

$$r = \tfrac{1}{17}(5 + 2\sqrt{2})\sqrt{7 + 4\sqrt{2}} \approx 1.63828$$
$$\rho = \tfrac{1}{2}(2 + \sqrt{2}) \approx 1.70711$$
$$R = \tfrac{1}{2}\sqrt{7 + 4\sqrt{2}} \approx 1.77882.$$

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 138, 1987.

Truncated Cuboctahedron

see GREAT RHOMBICUBOCTAHEDRON (ARCHIMEDEAN)

Truncated Dodecadodecahedron

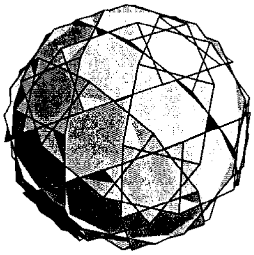

The UNIFORM POLYHEDRON U_{59}, also called the QUASITRUNCATED DODECAHEDRON, whose DUAL POLYHEDRON is the MEDIAL DISDYAKIS TRIACONTAHEDRON. It has SCHLÄFLI SYMBOL $t'\left\{\begin{matrix}\frac{5}{2}\\ 5\end{matrix}\right\}$ and WYTHOFF SYMBOL $2\,\frac{5}{3}\,|\,5$. Its faces are $12\{10\} + 30\{4\} + 12\{\frac{10}{3}\}$. Its CIRCUMRADIUS for $a = 1$ is

$$R = \tfrac{1}{2}\sqrt{11}.$$

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, pp. 152–153, 1989.

Truncated Dodecahedron

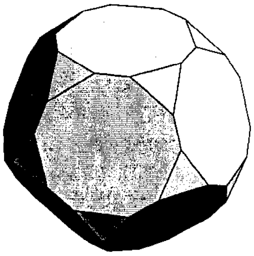

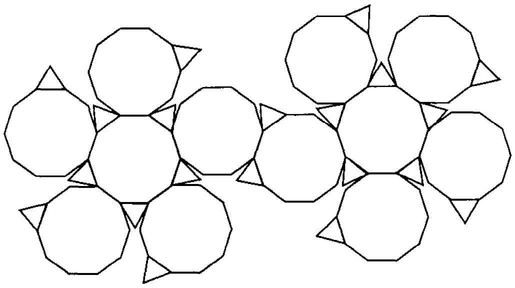

An ARCHIMEDEAN SOLID whose DUAL POLYHEDRON is the TRIAKIS ICOSAHEDRON. It has SCHLÄFLI SYMBOL t{5, 3}. It is also UNIFORM POLYHEDRON U_{26} and has WYTHOFF SYMBOL 2 3 | 5. Its faces are 20{3} + 12{10}. The INRADIUS, MIDRADIUS, and CIRCUMRADIUS for $a = 1$ are

$$r = \tfrac{5}{488}(17\sqrt{2} + 3\sqrt{10})\sqrt{37 + 15\sqrt{5}} \approx 2.88526$$
$$\rho = \tfrac{1}{4}(5 + 3\sqrt{5}) \approx 2.92705$$
$$R = \tfrac{1}{4}\sqrt{74 + 30\sqrt{5}} \approx 2.96945.$$

Truncated Great Dodecahedron

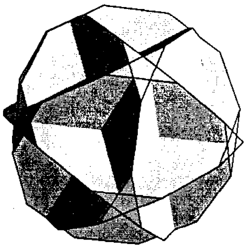

The UNIFORM POLYHEDRON U_{37} whose DUAL POLYHEDRON is the SMALL STELLAPENTAKIS DODECAHEDRON. It has SCHLÄFLI SYMBOL $t\{5, \frac{5}{2}\}$. It has WYTHOFF SYMBOL $2\ \frac{5}{2}\ 5$. Its faces are $12\{\frac{5}{2}\} + 12\{10\}$. Its CIRCUMRADIUS for $a = 1$ is

$$R = \tfrac{1}{4}\sqrt{34 + 10\sqrt{5}}.$$

see also GREAT ICOSAHEDRON

References

Wenninger, M. J. *Polyhedron Models.* Cambridge, England: Cambridge University Press, p. 115, 1971.

Truncated Great Icosahedron

see GREAT TRUNCATED ICOSAHEDRON

Truncated Hexahedron

see TRUNCATED CUBE

Truncated Icosahedron

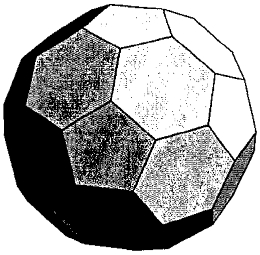

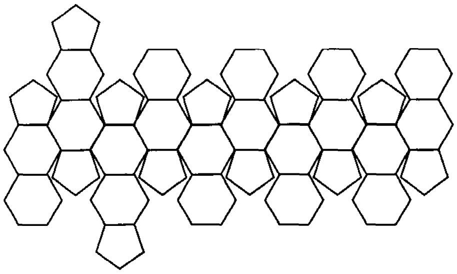

An ARCHIMEDEAN SOLID used in the construction of SOCCER BALLS. Its DUAL POLYHEDRON is the PENTAKIS DODECAHEDRON. It has SCHLÄFLI SYMBOL t{3, 5}. It is also UNIFORM POLYHEDRON U_{25} and has WYTHOFF SYMBOL 2 5 | 3. Its faces are 20{6} + 12{5}. The INRADIUS, MIDRADIUS, and CIRCUMRADIUS for $a = 1$ are

$$r = \tfrac{9}{872}(21 + \sqrt{5})\sqrt{58 + 18\sqrt{5}} \approx 2.37713$$
$$\rho = \tfrac{3}{4}(1 + \sqrt{5}) \approx 2.42705$$
$$R = \tfrac{1}{4}\sqrt{58 + 18\sqrt{5}} \approx 2.47802.$$

Truncated Icosidodecahedron

see GREAT RHOMBICOSIDODECAHEDRON (ARCHIMEDEAN)

Truncated Octahedral Number

A FIGURATE NUMBER which is constructed as an OCTAHEDRAL NUMBER with a SQUARE PYRAMID removed from each of the six VERTICES,

$$TO_n = O_{3n-2} - 6P_{n-1} = \tfrac{1}{3}(3n-2)[2(3n-2)^2 + 1],$$

where O_n is an OCTAHEDRAL NUMBER and P_n is a PYRAMIDAL NUMBER. The first few are 1, 38, 201, 586, ... (Sloane's A005910). The GENERATING FUNCTION for the truncated octahedral numbers is

$$\frac{x(6x^3 + 55x^2 + 34x + 1)}{(x-1)^4} = x + 38x^2 + 201x^3 + \ldots.$$

see also OCTAHEDRAL NUMBER

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 52, 1996.

Sloane, N. J. A. Sequence A005910/M5266 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Truncated Octahedron

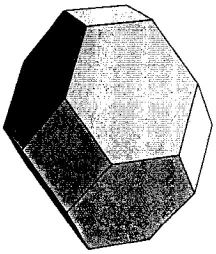

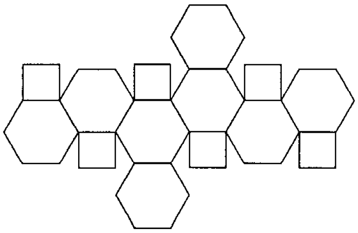

An ARCHIMEDEAN SOLID, also known as the MECON, whose DUAL POLYHEDRON is the TETRAKIS HEXAHEDRON. It is also UNIFORM POLYHEDRON U_8 and has SCHLÄFLI SYMBOL t{3, 4} and WYTHOFF SYMBOL 2 4 | 3. The faces of the truncated octahedron are 8{6}+6{4}. The truncated octahedron has the O_h OCTAHEDRAL GROUP of symmetries.

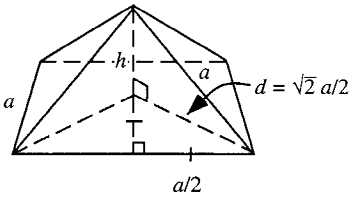

The solid can be formed from an OCTAHEDRON via TRUNCATION by removing six SQUARE PYRAMIDS, each with edge slant height $a = s/3$ and height h, where s is the side length of the original OCTAHEDRON. From the above diagram, the height and base area of the SQUARE PYRAMID are

$$h = \sqrt{a^2 - d^2} = \tfrac{1}{2}\sqrt{2}\,a \tag{1}$$

$$A_b = a^2. \tag{2}$$

The VOLUME of the truncated octahedron is then given by the VOLUME of the OCTAHEDRON

$$V_{\text{octahedron}} = \tfrac{1}{3}\sqrt{2}\,s^3 = 9\sqrt{2}\,a^3 \tag{3}$$

minus six times the volume of the SQUARE PYRAMID,

$$V = V_{\text{octahedron}} - 6(\tfrac{1}{3}A_b h) = (9\sqrt{2} - \sqrt{2})a^3 = 8\sqrt{2}\,a^3. \tag{4}$$

The truncated octahedron is a SPACE-FILLING POLYHEDRON. The INRADIUS, MIDRADIUS, and CIRCUMRADIUS for $a = 1$ are

$$r = \tfrac{9}{20}\sqrt{10} \approx 1.42302 \tag{5}$$

$$\rho = \tfrac{3}{2} = 1.5 \tag{6}$$

$$R = \tfrac{1}{2}\sqrt{10} \approx 1.58114. \tag{7}$$

see also OCTAHEDRON, SQUARE PYRAMID, TRUNCATION

References

Coxeter, H. S. M. *Regular Polytopes, 3rd ed.* New York: Dover, pp. 29–30 and 257, 1973.

Truncated Polyhedron

A polyhedron with truncated faces, given by the SCHLÄFLI SYMBOL $\mathrm{t}\left\{\begin{smallmatrix} p \\ q \end{smallmatrix}\right\}$.

see also RHOMBIC POLYHEDRON, SNUB POLYHEDRON

Truncated Pyramid

see PYRAMIDAL FRUSTUM

Truncated Square Pyramid

The truncated square pyramid is a special case of a PYRAMIDAL FRUSTUM for a SQUARE PYRAMID. Let the base and top side lengths of the truncated pyramid be a and b, and let the height be h. Then the VOLUME of the solid is

$$V = \tfrac{1}{3}(a^2 + ab + b^2)h.$$

This FORMULA was known to the Egyptians ca. 1850 BC. The Egyptians cannot have proved it without calculus, however, since Dehn showed in 1900 that no proof of this equation exists which does not rely on the concept of continuity (and therefore some form of INTEGRATION).

see also FRUSTUM, PYRAMID, PYRAMIDAL FRUSTUM, SQUARE PYRAMID

Truncated Tetrahedral Number

A FIGURATE NUMBER constructed by taking the $(3n-2)$th TETRAHEDRAL NUMBER and removing the $(n-1)$th TETRAHEDRAL NUMBER from each of the four corners,

$$\mathrm{Ttet}_n \equiv \mathrm{Te}_{3n-3} - 4\mathrm{Te}_{n-1} = \tfrac{1}{6}n(23n^2 - 27n + 10).$$

The first few are 1, 16, 68, 180, 375, ... (Sloane's A005906). The GENERATING FUNCTION for the truncated tetrahedral numbers is

$$\frac{x(10x^2 + 12x + 1)}{(x-1)^4} = x + 16x^2 + 89x^3 + 180x^4 + \ldots.$$

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 46–47, 1996.

Sloane, N. J. A. Sequence A005906/M5002 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Truncated Tetrahedron

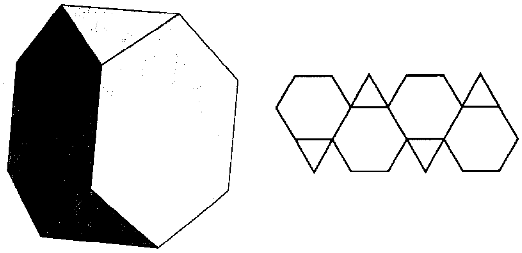

An ARCHIMEDEAN SOLID whose dual is the TRIAKIS TETRAHEDRON. It has SCHLÄFLI SYMBOL t{3,3}. It is also UNIFORM POLYHEDRON U_2 and has WYTHOFF SYMBOL 2 3 | 3. Its faces are 4{3} + 4{6}. The INRADIUS, MIDRADIUS, and CIRCUMRADIUS for a truncated tetrahedron with $a = 1$ are

$$\begin{aligned} r &= \tfrac{9}{44}\sqrt{22} \approx 0.95940 \\ \rho &= \tfrac{3}{4}\sqrt{2} \approx 1.06066 \\ R &= \tfrac{1}{4}\sqrt{22} \approx 1.17260. \end{aligned}$$

Truncation

The removal of portions of SOLIDS falling outside a set of symmetrically placed planes. The five PLATONIC SOLIDS belong to one of the following three truncation series (which, in the first two cases, carry the solid to its DUAL POLYHEDRON).

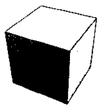
Cube

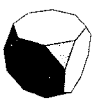
Truncated Cube

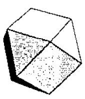
Cuboctahedron

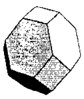
Truncated Octahedron

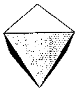
Octahedron

Icosahedron

Truncated Icosahedron

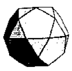
Icosidodecahedron

Truncated Dodecahedron

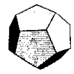
Dodecahedron

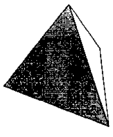
Tetrahedron

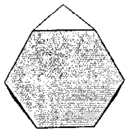
Truncated Tetrahedron

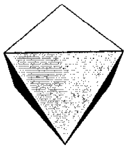
Octahedron

see also STELLATION, TRUNCATED CUBE, TRUNCATED DODECAHEDRON, TRUNCATED ICOSAHEDRON, TRUNCATED OCTAHEDRON, TRUNCATED TETRAHEDRON, VERTEX FIGURE

Truth Table

A truth table is a 2-D array with $n + 1$ columns. The first n columns correspond to the possible values of n inputs, and the last column to the operation being performed. The rows list all possible combinations of inputs together with the corresponding outputs. For example, the following truth table shows the result of the binary AND operator acting on two inputs A and B, each of which may be true or false.

A	B	$A \wedge B$
F	F	F
F	T	F
T	F	F
T	T	T

see also AND, MULTIPLICATION TABLE, OR, XOR

Tschebyshev

An alternative spelling of the name "Chebyshev."

see also CHEBYSHEV APPROXIMATION FORMULA, CHEBYSHEV CONSTANTS, CHEBYSHEV DEVIATION, CHEBYSHEV DIFFERENTIAL EQUATION, CHEBYSHEV FUNCTION, CHEBYSHEV-GAUSS QUADRATURE, CHEBYSHEV INEQUALITY, CHEBYSHEV INTEGRAL, CHEBYSHEV PHENOMENON, CHEBYSHEV POLYNOMIAL OF THE FIRST KIND, CHEBYSHEV POLYNOMIAL OF THE SECOND KIND, CHEBYSHEV QUADRATURE, CHEBYSHEV-RADAU QUADRATURE, CHEBYSHEV-SYLVESTER CONSTANT

Tschirnhausen Cubic

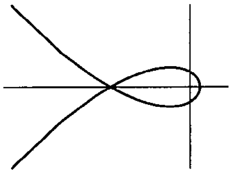

The Tschirnhausen cubic is a plane curve given by

$$a = r\cos^3\left(\tfrac{1}{3}\theta\right),$$

and is also known as CATALAN'S TRISECTRIX and L'HOSPITAL'S CUBIC. The name Tschirnhaus's cubic is given in R. C. Archibald's 1900 paper attempting to classify curves (MacTutor Archive). Tschirnhaus's cubic is the NEGATIVE PEDAL CURVE of a PARABOLA with respect to the FOCUS.

References

Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 87–90, 1972.

MacTutor History of Mathematics Archive. "Tschirnhaus's Cubic." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Tschirnhaus.html`.

Tschirnhausen Cubic Caustic

The CAUSTIC of the TSCHIRNHAUSEN CUBIC taking the RADIANT POINT as the pole is NEILE'S PARABOLA.

Tschirnhausen Cubic Pedal Curve

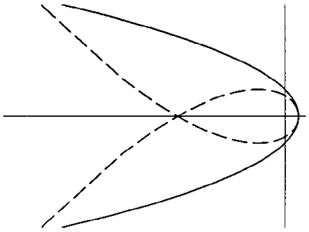

The PEDAL CURVE to the TSCHIRNHAUSEN CUBIC for PEDAL POINT at the origin is the PARABOLA

$$\begin{aligned} x &= 1 - t^2 \\ y &= 2t. \end{aligned}$$

see also PARABOLA, PEDAL CURVE, PEDAL POINT, TSCHIRNHAUSEN CUBIC

Tschirnhausen Transformation

A transformation of a POLYNOMIAL equation $f(x) = 0$ which is of the form $y = g(x)/h(x)$ where g and h are POLYNOMIALS and $h(x)$ does not vanish at a root of $f(x) = 0$. The CUBIC EQUATION is a special case of such a transformation. Tschirnhaus (1683) showed that a POLYNOMIAL of degree $n > 2$ can be reduced to a form in which the x^{n-1} and x^{n-2} terms have 0 COEFFICIENTS. In 1786, E. S. Bring showed that a general QUINTIC EQUATION can be reduced to the form

$$x^5 + px + q = 0.$$

In 1834, G. B. Jerrard showed that a Tschirnhaus transformation can be used to eliminate the x^{n-1}, x^{n-2}, *and* x^{n-3} terms for a general POLYNOMIAL equation of degree $n > 3$.

see also BRING QUINTIC FORM, CUBIC EQUATION

References
Boyer, C. B. *A History of Mathematics.* New York: Wiley, pp. 472–473, 1968.
Tschirnhaus. *Acta Eruditorum.* 1683.

Tubular Neighborhood

The tubular embedding of a SUBMANIFOLD $M^m \subset N^n$ of another MANIFOLD N^n is an EMBEDDING $t : M \times \mathbb{B}^{n-m} \to N$ such that $t(x, 0) = x$ whenever $x \in M$, where $\mathbb{B}^{n-m}$ is the unit BALL in $\mathbb{R}^{n-m}$ centered at 0. The tubular neighborhood is also called the PRODUCT NEIGHBORHOOD.

see also BALL, EMBEDDING, PRODUCT NEIGHBORHOOD

References
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 34–35, 1976.

Tucker Circles

Let three equal lines P_1Q_1, P_2Q_2, and P_3Q_3 be drawn ANTIPARALLEL to the sides of a triangle so that two (say P_2Q_2 and P_3Q_3) are on the same side of the third line as $A_2P_2Q_3A_3$. Then $P_2Q_3P_3Q_2$ is an isosceles TRAPEZOID, i.e., P_3Q_2, P_1Q_3, and P_2Q_1 are parallel to the respective sides. The MIDPOINTS C_1, C_2, and C_3 of the antiparallels are on the respective symmedians and divide them proportionally.

If T divides KO in the same ratio, TC_1, TC_2, TC_3 are parallel to the radii OA_1, OA_2, and OA_3 and equal. Since the antiparallels are perpendicular to the symmedians, they are equal chords of a circle with center T which passes through the six given points. This circle is called the Tucker circle.

If

$$c \equiv \frac{\overline{KC_1}}{\overline{KA_1}} = \frac{\overline{KC_2}}{\overline{KA_2}} = \frac{\overline{KC_3}}{\overline{KA_3}} = \frac{\overline{KT}}{\overline{KO}},$$

then the radius of the Tucker circle is

$$R\sqrt{c^2 + (1-c)^2 \tan\omega},$$

where ω is the BROCARD ANGLE.

The COSINE CIRCLE, LEMOINE CIRCLE, and TAYLOR CIRCLE are Tucker circles.

see also ANTIPARALLEL, BROCARD ANGLE, COSINE CIRCLE, LEMOINE CIRCLE, TAYLOR CIRCLE

References
Johnson, R. A. *Modern Geometry: An Elementary Treatise on the Geometry of the Triangle and the Circle.* Boston, MA: Houghton Mifflin, pp. 271–277 and 300–301, 1929.

Tukey's Biweight

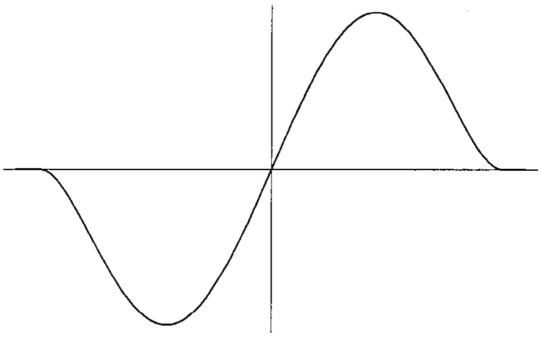

The function

$$\psi(z) = \begin{cases} z\left(1 - \frac{z^2}{c^2}\right)^2 & \text{for } |z| < c \\ 0 & \text{for } |z| > c \end{cases}$$

sometimes used in ROBUST ESTIMATION. It has a minimum at $z = -c/\sqrt{3}$ and a maximum at $z = c/\sqrt{3}$, where

$$\psi'(z) = 1 - \frac{3x^2}{c^2} = 0,$$

and an inflection point at $z = 0$, where

$$\psi''(z) = -\frac{6z}{c^2} = 0.$$

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 697, 1992.

Tukey's Trimean

see TRIMEAN

Tunnel Number

Let a KNOT K be n-EMBEDDABLE. Then its tunnel number is a KNOT invariant which is related to n.

see also EMBEDDABLE KNOT

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, p. 114, 1994.

Turán Graph

The (n, k)-Turán graph is the EXTREMAL GRAPH on n VERTICES which contains no k-CLIQUE. In other words, the Turán graph has the maximum possible number of EDGES of any n-vertex graph not containing a COMPLETE GRAPH K_k. TURÁN'S THEOREM gives the maximum number of edges $t(n,k)$ for the (n,k)-Turán graph. For $k = 3$,

$$t(n,3) = \tfrac{1}{4}n^4,$$

so the Turán graph is given by the COMPLETE BIPARTITE GRAPHS

$$\begin{cases} K_{n/2,n/2} & n \text{ even} \\ K_{(n-1)/2,(n+1)/2} & n \text{ odd.} \end{cases}$$

see also CLIQUE, COMPLETE BIPARTITE GRAPH, TURÁN'S THEOREM

References

Aigner, M. "Turán's Graph Theorem." *Amer. Math. Monthly* **102**, 808–816, 1995.

Turán's Inequalities

For a set of POSITIVE γ_k, $k = 0, 1, 2\ldots$, Turán's inequalities are given by

$$\gamma_k{}^2 - \gamma_{k-1}\gamma_{k+1} \geq 0$$

for $k = 1, 2, \ldots$.

see also JENSEN POLYNOMIAL

References

Csordas, G.; Varga, R. S.; and Vincze, I. "Jensen Polynomials with Applications to the Riemann ζ-Function." *J. Math. Anal. Appl.* **153**, 112–135, 1990.
Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., p. 388, 1975.

Turán's Theorem

Let $G(V, E)$ be a GRAPH with VERTICES V and EDGES E on n VERTICES without a k-CLIQUE. Then

$$t(n,k) \leq \frac{(k-2)n^2}{2(k-1)},$$

where $t(n,k) = |E|$ is the EDGE NUMBER. More precisely, the K-GRAPH $K_{n_1,\ldots,n_{k-1}}$ with $|n_i - n_j| \leq 1$ for $i \neq j$ is the unique GRAPH without a k-CLIQUE with the maximal number of EDGES $t(n,k)$.

see also CLIQUE, K-GRAPH, TURÁN GRAPH

References

Aigner, M. "Turán's Graph Theorem." *Amer. Math. Monthly* **102**, 808–816, 1995.

Turbine

A VECTOR FIELD on a CIRCLE in which the directions of the VECTORS are all at the same ANGLE to the CIRCLE.

see also CIRCLE, VECTOR FIELD

Turing Machine

A theoretical computing machine which consists of an infinitely long magnetic tape on which instructions can be written and erased, a single-bit register of memory, and a processor capable of carrying out the following instructions: move the tape right, move the tape left, change the state of the register based on its current value and a value on the tape, and write or erase a value on the tape. The machine keeps processing instructions until it reaches a particular state, causing it to halt. Determining whether a Turing machine will halt for a given input and set of rules is called the HALTING PROBLEM.

see also BUSY BEAVER, CELLULAR AUTOMATON, CHAITIN'S OMEGA, CHURCH-TURING THESIS, COMPUTABLE NUMBER, HALTING PROBLEM, UNIVERSAL TURING MACHINE

References

Penrose, R. "Algorithms and Turning Machines." Ch. 2 in *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics.* Oxford, England: Oxford University Press, pp. 30–73, 1989.
Turing, A. M. "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proc. London Math. Soc. Ser. 2* **42**, 230–265, 1937.
Turing, A. M. "Correction to: On Computable Numbers, with an Application to the Entscheidungsproblem." *Proc. London Math. Soc. Ser. 2* **43**, 544–546, 1938.

Turning Angle

see TANGENTIAL ANGLE

Tutte's Graph

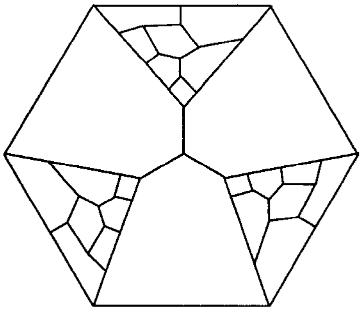

A counterexample to Tait's Hamiltonian Graph Conjecture given by Tutte (1946). A simpler counterexample was later given by Kozyrev and Grinberg.

see also Hamiltonian Circuit, Tait's Hamiltonian Graph Conjecture

References

Honsberger, R. *Mathematical Gems I.* Washington, DC: Math. Assoc. Amer., pp. 82–89, 1973.

Saaty, T. L. and Kainen, P. C. *The Four-Color Problem: Assaults and Conquest.* New York: Dover, p. 112, 1986.

Tutte, W. T. "On Hamiltonian Circuits." *J. London Math. Soc.* **21**, 98–101, 1946.

Tutte Polynomial

Let G be a Graph, and let $\operatorname{ea}(T)$ denote the cardinality of the set of externally active edges of a spanning tree T of G and $\operatorname{ia}(T)$ denote the cardinality of the set of internally active edges of T. Then

$$t_G(x,y)=\sum_{T\subseteq G} x^{\operatorname{ia}(T)}y^{\operatorname{ea}(T)}.$$

References

Gessel, I. M. and Sagan, B. E. "The Tutte Polynomial of a Graph, Depth-First Search, and Simplicial Complex Partitions." *Electronic J. Combinatorics* **3**, No. 2, R9, 1–36, 1996. `http://www.combinatorics.org/Volume_3/volume3_2.html#R9.`

Tutte, W. T. "A Contribution to the Theory of Chromatic Polynomials." *Canad. J. Math.* **6**, 80–91, 1953.

Tutte's Theorem

Let G be a Graph and S a Subgraph of G. Let the number of Odd components in $G-S$ be denoted S', and $|S|$ the number of Vertices of S. The condition $|S|\geq S'$ for every Subset of Vertices is Necessary and Sufficient for G to have a 1-Factor.

see also Factor (Graph)

References

Honsberger, R. "Lovász' Proof of a Theorem of Tutte." Ch. 14 in *Mathematical Gems II.* Washington, DC: Math. Assoc. Amer., pp. 147–157, 1976.

Tutte, W. T. "The Factorization of Linear Graphs." *J. London Math. Soc.* **22**, 107–111, 1947.

Twin Peaks

For an Integer $n\geq 2$, let $\operatorname{lpf}(x)$ denote the Least Prime Factor of n. A Pair of Integers (x,y) is called a twin peak if

1. $x<y$,
2. $\operatorname{lpf}(x)=\operatorname{lpf}(y)$,
3. For all z, $x<z<y$ Implies $\operatorname{lpf}(z)<\operatorname{lpf}(x)$.

A broken-line graph of the least prime factor function resembles a jagged terrain of mountains. In terms of this terrain, a twin peak consists of two mountains of equal height with no mountain of equal or greater height between them. Denote the height of twin peak (x,y) by $p=\operatorname{lpf}(x)=\operatorname{lpf}(y)$. By definition of the Least Prime Factor function, p must be Prime.

Call the distance between two twin peaks (x,y)

$$s\equiv y-x.$$

Then s must be an Even multiple of p; that is, $s=kp$ where k is Even. A twin peak with $s=kp$ is called a kp-twin peak. Thus we can speak of $2p$-twin peaks, $4p$-twin peaks, etc. A kp-twin peak is fully specified by k, p, and x, from which we can easily compute $y\equiv x+kp$.

The set of kp-twin peaks is periodic with period $q=p\#$, where $p\#$ is the Primorial of p. That is, if (x,y) is a kp-twin peak, then so is $(x+q,y+q)$. A fundamental kp-twin peak is a twin peak having x in the fundamental period $[0,q)$. The set of fundamental kp-twin peaks is symmetric with respect to the fundamental period; that is, if (x,y) is a twin peak on $[0,q)$, then so is $(q-y,q-x)$.

The question of the Existence of twin peaks was first raised by David Wilson in the math-fun mailing list on Feb. 10, 1997. Wilson already had privately showed the Existence of twin peaks of height $p\leq 13$ to be unlikely, but was unable to rule them out altogether. Later that same day, John H. Conway, Johan de Jong, Derek Smith, and Manjul Bhargava collaborated to discover the first twin peak. Two hours at the blackboard revealed that $p=113$ admits the $2p$-twin peak

$$x=126972592296404970720882679404584182254788131,$$

which settled the Existence question. Immediately thereafter, Fred Helenius found the smaller $2p$-twin peak with $p=89$ and

$$x=9503844926749390990454854843625839.$$

The effort now shifted to finding the least Prime p admitting a $2p$-twin peak. On Feb. 12, 1997, Fred Helenius found $p=71$, which admits 240 fundamental $2p$-twin peaks, the least being

$$x=7310131732015251470110369.$$

Helenius's results were confirmed by Dan Hoey, who also computed the least $2p$-twin peak $L(2p)$ and number of fundamental $2p$-twin peaks $N(2p)$ for $p = 73$, 79, and 83. His results are summarized in the following table.

p	$L(2p)$	$N(2p)$
71	7310131732015251470110369	240
73	2061519317176132799110061	40296
79	3756800873017263196139951	164440
83	6316254452384500173544921	6625240

The $2p$-twin peak of height $p = 73$ is the smallest known twin peak. Wilson found the smallest known $4p$-twin peak with $p = 1327$, as well as another very large $4p$-twin peak with $p = 3203$. Richard Schroeppel noted that the latter twin peak is at the high end of its fundamental period and that its reflection within the fundamental period $[0, p\#)$ is smaller.

Many open questions remain concerning twin peaks, e.g.,

1. What is the smallest twin peak (smallest n)?
2. What is the least PRIME p admitting a $4p$-twin peak?
3. Do $6p$-twin peaks exist?
4. Is there, as Conway has argued, an upper bound on the span of twin peaks?
5. Let $p < q < r$ be PRIME. If p and r each admit kp-twin peaks, does q then necessarily admit a kp-twin peak?

see also ANDRICA'S CONJECTURE, DIVISOR FUNCTION, LEAST COMMON MULTIPLE, LEAST PRIME FACTOR

Twin Prime Conjecture

Adding a correction proportional to $1/\ln p$ to a computation of BRUN'S CONSTANT ending with $\ldots + 1/p + 1/(p+2)$ will give an estimate with error less than $c(\sqrt{p}\,\ln p)^{-1}$. An extended form of the conjecture states that

$$P_x(p, p+2) \sim 2\Pi_2 \int_2^x \frac{dx}{(\ln x)^2},$$

where Π_2 is the TWIN PRIMES CONSTANT. The twin prime conjecture is a special case of the more general PRIME PATTERNS CONJECTURE corresponding to the set $S = \{0, 2\}$.

see also BRUN'S CONSTANT, PRIME ARITHMETIC PROGRESSION, PRIME CONSTELLATION, PRIME PATTERNS CONJECTURE, TWIN PRIMES

Twin Primes

Twin primes are PRIMES (p, q) such that $p - q = 2$. The first few twin primes are $n \pm 1$ for $n = 4$, 6, 12, 18, 30, 42, 60, 72, 102, 108, 138, 150, 180, 192, 198, 228, 240, 270, 282, ... (Sloane's A014574). Explicitly, these are (3, 5), (5, 7), (11, 13), (17, 19), (29, 31), (41, 43), ... (Sloane's A001359 and A006512).

Let $\pi_2(n)$ be the number of twin primes p and $p+2$ such that $p \leq n$. It is not known if there are an infinite number of such PRIMES (Shanks 1993), but all twin primes except (3, 5) are of the form $6n \pm 1$. J. R. Chen has shown there exists an INFINITE number of PRIMES p such that $p + 2$ has at most two factors (Le Lionnais 1983, p. 49). Bruns proved that there exists a computable INTEGER x_0 such that if $x \geq x_0$, then

$$\pi_2(x) < \frac{100x}{(\ln x)^2} \tag{1}$$

(Ribenboim 1989, p. 201). It has been shown that

$$\pi_2(x) \leq c \prod_{p>2} \left[1 - \frac{1}{(p-1)^2}\right] \frac{x}{(\ln x)^2} \left[1 + \mathcal{O}\left(\frac{\ln \ln x}{\ln x}\right)\right], \tag{2}$$

where c has been reduced to $68/9 \approx 7.5556$ (Fouvry and Iwaniec 1983), $128/17 \approx 7.5294$ (Fouvry 1984), 7 (Bombieri *et al.* 1986), 6.9075 (Fouvry and Grupp 1986), and 6.8354 (Wu 1990). The bound on c is further reduced to 6.8324107886 in a forthcoming thesis by Haugland (1998). This calculation involved evaluation of 7-fold integrals and fitting of three different parameters. Hardy and Littlewood conjectured that $c = 2$ (Ribenboim 1989, p. 202).

Define

$$E \equiv \liminf_{n\to\infty} \frac{p_{n+1} - p_n}{\ln p_n}. \tag{3}$$

If there are an infinite number of twin primes, then $E = 0$. The best upper limit to date is $E \leq \frac{1}{4} + \pi/16 = 0.44634\ldots$ (Huxley 1973, 1977). The best previous values were 15/16 (Ricci), $(2 + \sqrt{3})/8 = 0.46650\ldots$ (Bombieri and Davenport 1966), and $(2\sqrt{2} - 1)/4 = 0.45706\ldots$ (Pil'Tai 1972), as quoted in Le Lionnais (1983, p. 26).

Some large twin primes are $10{,}006{,}428 \pm 1$, $1{,}706{,}595 \times 2^{11235} \pm 1$, and $571{,}305 \times 2^{7701} \pm 1$. An up-to-date table of known twin primes with 2000 or more digits follows. An extensive list is maintained by `Caldwell`.

$(p, p+1)$	dig.	Reference
$260{,}497{,}545 \times 2^{6625} \pm 1$	2003	Atkin & Rickert 1984
$43{,}690{,}485{,}351{,}513 \times 10^{1995} \pm 1$	2009	Dubner, Atkin 1985
$2{,}846!!!! \pm 1$	2151	Dubner 1992
$10{,}757{,}0463 \times 10^{2250} \pm 1$	2259	Dubner, Atkin 1985
$663{,}777 \times 2^{7650} \pm 1$	2309	Brown *et al.* 1989
$75{,}188{,}117{,}004 \times 10^{2298} \pm 1$	2309	Dubner 1989
$571305 \times 2^{7701} \pm 1$	2324	Brown *et al.* 1989
$1{,}171{,}452{,}282 \times 10^{2490} \pm 1$	2500	Dubner 1991
$459 \cdot 2^{8529} \pm 1$	2571	Dubner 1993
$1{,}706{,}595 \cdot 2^{11235} \pm 1$	3389	Noll *et al.* 1989
$4{,}655{,}478{,}828 \cdot 10^{3429} \pm 1$	3439	Dubner 1993
$1{,}692{,}923{,}232 \cdot 10^{4020} \pm 1$	4030	Dubner 1993
$6{,}797{,}727 \cdot 2^{15328} \pm 1$	4622	Forbes 1995
$697{,}053{,}813\dot{2}^{16352} \pm 1$	4932	Indlekofer & Ja'rai 1994
$570{,}918{,}348 \cdot 10^{5120} \pm 1$	5129	Dubner 1995
$242{,}206{,}083 \cdot 2^{38880} \pm 1$	11713	Indlekofer & Ja'rai 1995

The last of these is the largest known twin prime pair. In 1995, Nicely discovered a flaw in the Intel® Pentium™ microprocessor by computing the reciprocals of 824,633,702,441 and 824,633,702,443, which should have been accurate to 19 decimal places but were incorrect from the tenth decimal place on (Cipra 1995, 1996; Nicely 1996).

If $n \geq 2$, the INTEGERS n and $n+2$ form a pair of twin primes IFF

$$4[(n-1)!+1]+n \equiv 0 \pmod{n(n+2)}. \qquad (4)$$

$n = pp'$ where (p, p') is a pair of twin primes IFF

$$\phi(n)\sigma(n) = (n-3)(n+1) \qquad (5)$$

(Ribenboim 1989). The values of $\pi_2(n)$ were found by Brent (1976) up to $n = 10^{11}$. T. Nicely calculated them up to 10^{14} in his calculation of BRUN'S CONSTANT. The following table gives the number less than increasing powers of 10 (Sloane's A007508).

n	$\pi_2(n)$
10^3	35
10^4	205
10^5	1224
10^6	8,169
10^7	58,980
10^8	440,312
10^9	3,424,506
10^{10}	27,412,679
10^{11}	224,376,048
10^{12}	1,870,585,220
10^{13}	15,834,664,872
10^{14}	135,780,321,665

see also BRUN'S CONSTANT, DE POLIGNAC'S CONJECTURE PRIME CONSTELLATION, SEXY PRIMES, TWIN PRIME CONJECTURE, TWIN PRIMES CONSTANT

References

Bombieri, E. and Davenport, H. "Small Differences Between Prime Numbers." *Proc. Roy. Soc. Ser. A* **293**, 1–8, 1966.
Bombieri, E.; Friedlander, J. B.; and Iwaniec, H. "Primes in Arithmetic Progression to Large Moduli." *Acta Math.* **156**, 203–251, 1986.
Bradley, C. J. "The Location of Twin Primes." *Math. Gaz.* **67**, 292–294, 1983.
Brent, R. P. "Irregularities in the Distribution of Primes and Twin Primes." *Math. Comput.* **29**, 43–56, 1975.
Brent, R. P. "UMT 4." *Math. Comput.* **29**, 221, 1975.
Brent, R. P. "Tables Concerning Irregularities in the Distribution of Primes and Twin Primes to 10^{11}." *Math. Comput.* **30**, 379, 1976.
Caldwell, C. `http://www.utm.edu/cgi-bin/caldwell/primes.cgi/twin`.
Cipra, B. "How Number Theory Got the Best of the Pentium Chip." *Science* **267**, 175, 1995.
Cipra, B. "Divide and Conquer." *What's Happening in the Mathematical Sciences, 1995–1996, Vol. 3.* Providence, RI: Amer. Math. Soc., pp. 38–47, 1996.
Fouvry, É. "Autour du theoreme de Bombieri-Vinogradov." *Acta. Math.* **152**, 219–244, 1984.
Fouvry, É. and Grupp, F. "On the Switching Principle in Sieve Theory." *J. Reine Angew. Math.* **370**, 101–126, 1986.
Fouvey, É. and Iwaniec, H. "Primes in Arithmetic Progression." *Acta Arith.* **42**, 197–218, 1983.
Guy, R. K. "Gaps between Primes. Twin Primes." §A8 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 19–23, 1994.
Haugland, J. K. *Topics in Analytic Number Theory.* Ph.D. thesis. Oxford, England: Oxford University, Oct. 1998.
Huxley, M. N. "Small Differences between Consecutive Primes." *Mathematica* **20**, 229–232, 1973.
Huxley, M. N. "Small Differences between Consecutive Primes. II." *Mathematica* **24**, 142–152, 1977.
Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, 1983.
Nicely, T. R. "The Pentium Bug.' `http://www.lynchburg.edu/public/academic/math/nicely/pentbug/pentbug.htm`.
Nicely, T. "Enumeration to 10^{14} of the Twin Primes and Brun's Constant." *Virginia J. Sci.* **46**, 195–204, 1996. `http://www.lynchburg.edu/public/academic/math/nicely/twins/twins.htm`.
Parady, B. K.; Smith, J. F.; and Zarantonello, S. E. "Largest Known Twin Primes." *Math. Comput.* **55**, 381–382, 1990.
Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, pp. 199–204, 1989.
Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 30, 1993.
Sloane, N. J. A. Sequences A014574, A001359/M2476, A006512/M3763, and A007508/M1855 in "An On-Line Version of the Encyclopedia of Integer Sequences."
Weintraub, S. "A Prime Gap of 864." *J. Recr. Math.* **25**, 42–43, 1993.
Wu, J. "Sur la suite des nombres premiers jumeaux." *Acta. Arith.* **55**, 365–394, 1990.

Twin Primes Constant

The twin primes constant Π_2 is defined by

$$\Pi_2 \equiv \prod_{\substack{p>2\\ p\text{ prime}}} \left[1 - \frac{1}{(p-1)^2}\right] \qquad (1)$$

$$\begin{aligned}\ln(\tfrac{1}{2}\Pi_2) &= \sum_{\substack{p\geq 3\\ p\text{ prime}}} \ln\left[\frac{p(p-2)}{(p-1)^2}\right]\\ &= \sum_{\substack{p\geq 3\\ p\text{ prime}}} \left[\ln\left(1-\frac{2}{p}\right) - 2\ln\left(1-\frac{1}{p}\right)\right]\\ &= -\sum_{j=2}^{\infty} \frac{2^j-2}{j} \sum_{\substack{p\geq 3\\ p\text{ prime}}} p^{-j}, \qquad (2)\end{aligned}$$

where the ps in sums and products are taken over PRIMES only. Flajolet and Vardi (1996) give series with accelerated convergence

$$\begin{aligned}\Pi_2 &= \prod_{n=2}^{\infty} [\zeta(n)(1-2^{-n})]^{-I_n} \qquad (3)\\ &= \tfrac{3}{4}\tfrac{15}{16}\tfrac{35}{36} \prod_{n=2}^{\infty} [\zeta(n)(1-2^{-n})(1-3^{-n})(1-5^{-n})\\ &\qquad \times(1-7^{-n})]^{-I_n}, \qquad (4)\end{aligned}$$

with

$$I_n \equiv \frac{1}{n}\sum_{d|n}\mu(d)2^{n/d}, \qquad (5)$$

where $\mu(x)$ is the MÖBIUS FUNCTION. (4) has convergence like $\sim (11/2)^{-n}$.

The most accurately known value of Π_2 is

$$\Pi_2 = 0.6601618158\ldots. \qquad (6)$$

Le Lionnais (1983, p. 30) calls C_2 the SHAH-WILSON CONSTANT, and $2C_2$ the twin prime constant (Le Lionnais 1983, p. 37).

see also BRUN'S CONSTANT, GOLDBACH CONJECTURE, MERTENS CONSTANT

References
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/hrdyltl/hrdyltl.html`.
Flajolet, P. and Vardi, I. "Zeta Function Expansions of Classical Constants." Unpublished manuscript. 1996. `http://pauillac.inria.fr/algo/flajolet/Publications/landau.ps`.
Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, 1983.
Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, p. 202, 1989.
Ribenboim, P. *The Little Book of Big Primes.* New York: Springer-Verlag, p. 147, 1991.
Riesel, H. *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, pp. 61–66, 1994.
Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 30, 1993.
Wrench, J. W. "Evaluation of Artin's Constant and the Twin Prime Constant." *Math. Comput.* **15**, 396–398, 1961.

Twins

see BROTHERS, PAIR

Twirl

A ROTATION combined with an EXPANSION or DILATION.

see also SCREW, SHIFT

Twist

The twist of a ribbon measures how much it twists around its axis and is defined as the integral of the incremental twist around the ribbon. Letting Lk be the linking number of the two components of a ribbon, Tw be the twist, and Wr be the WRITHE, then

$$\mathrm{Lk}(R) = \mathrm{Tw}(R) + \mathrm{Wr}(R)$$

(Adams 1994, p. 187).

see also SCREW, WRITHE

References
Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, 1994.

Twist Map

A class of AREA-PRESERVING MAPS of the form

$$\begin{aligned}\theta_{i+1} &= \theta_i + 2\pi\alpha(r_i)\\ r_{i+1} &= r_i,\end{aligned}$$

which maps CIRCLES into CIRCLES but with a twist resulting from the $\alpha = \alpha(r_i)$ term.

Twist Move

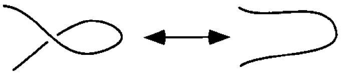

The REIDEMEISTER MOVE of type II.

see also REIDEMEISTER MOVES

Twist Number

see WRITHE

Twist-Spun Knot

A generalization of SPUN KNOTS due to Zeeman. This method produces 4-D KNOT types that cannot be produced by ordinary spinning.

see also SPUN KNOT

Twisted Chevalley Groups

FINITE SIMPLE GROUPS OF LIE-TYPE of ORDERS 14, 52, 78, 133, and 248. They are denoted $^3D_4(q)$, $E_6(q)$, $E_7(q)$, $E_8(q)$, $F_4(q)$, $^2F_4(2^n)'$, $G_2(q)$,$^2G_2(3^n)$, $^2B(2^n)$.

see also CHEVALLEY GROUPS, FINITE GROUP, SIMPLE GROUP, TITS GROUP

References
Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas#twi`.

Twisted Conic

see SKEW CONIC

Twisted Sphere

see CORKSCREW SURFACE

Two

see 2

Two-Ears Theorem

Except for TRIANGLES, every SIMPLE POLYGON has at least two nonoverlapping EARS.

see also EAR, ONE-MOUTH THEOREM, PRINCIPAL VERTEX

References
Meisters, G. H. "Principal Vertices, Exposed Points, and Ears." *Amer. Math. Monthly* **87**, 284–285, 1980.
Toussaint, G. "Anthropomorphic Polygons." *Amer. Math. Monthly* **122**, 31–35, 1991.

Two-Point Distance

see POINT-POINT DISTANCE—1-D, POINT-POINT DISTANCE—2-D, POINT-POINT DISTANCE—3-D

Two Triangle Theorem

see DESARGUES' THEOREM

Tychonof Compactness Theorem

The topological product of any number of COMPACT SPACES is COMPACT.

Type

Whitehead and Russell (1927) devised a hierarchy of "types" in order to eliminate self-referential statements from *Principia Mathematica*, which purported to derive all of mathematics from logic. A set of the lowest type contained only objects (not sets), a set of the next higher type could contain only objects or sets of the lower type, and so on. Unfortunately, GÖDEL'S INCOMPLETENESS THEOREM showed that both *Principia Mathematica* and all consistent formal systems must be incomplete.

see also GÖDEL'S INCOMPLETENESS THEOREM

References

Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books, pp. 21–22, 1989.

Whitehead, A. N. and Russell, B. *Principia Mathematica.* New York: Cambridge University Press, 1927.

Type I Error

An error in a STATISTICAL TEST which occurs when a true hypothesis is rejected (a false negative in terms of the NULL HYPOTHESIS).

see also NULL HYPOTHESIS, SENSITIVITY, SPECIFICITY, STATISTICAL TEST, TYPE II ERROR

Type II Error

An error in a STATISTICAL TEST which occurs when a false hypothesis is accepted (a false positive in terms of the NULL HYPOTHESIS).

see also NULL HYPOTHESIS, SENSITIVITY, SPECIFICITY, STATISTICAL TEST, TYPE I ERROR

U

U-Number

see ULAM SEQUENCE

Ulam Map

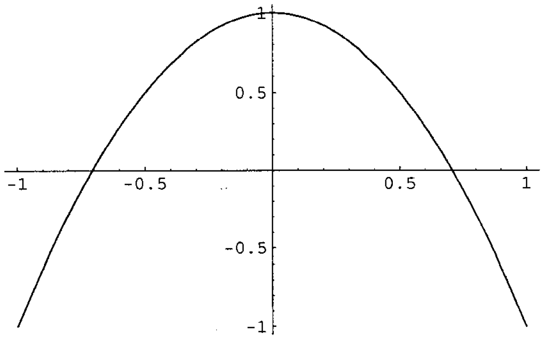

$$f(x) = 1 - 2x^2$$

for $x \in [-1, 1]$. Fixed points occur at $x = -1$, $1/2$, and order 2 fixed points at $x = (1 \pm \sqrt{5})/4$. The INVARIANT DENSITY of the map is

$$\rho(y) = \frac{1}{\pi\sqrt{1-y^2}}.$$

References

Beck, C. and Schlögl, F. *Thermodynamics of Chaotic Systems: An Introduction.* Cambridge, England: Cambridge University Press, p. 194, 1995.

Ulam Number

see ULAM SEQUENCE

Ulam's Problem

see COLLATZ PROBLEM

Ulam Sequence

The Ulam sequence $\{a_i\} = (u, v)$ is defined by $a_1 = u$, $a_2 = v$, with the general term a_n for $n > 2$ given by the least INTEGER expressible uniquely as the SUM of two distinct earlier terms. The numbers so produced are sometimes called U-NUMBERS or ULAM NUMBERS.

The first few numbers in the (1, 2) Ulam sequence are 1, 2, 3, 4, 6, 8, 11, 13, 16, ... (Sloane's A002858). Here, the first term after the initial 1, 2 is obviously 3 since $3 = 1 + 2$. The next term is $4 = 1 + 3$. (We don't have to worry about $4 = 2 + 2$ since it is a sum of a single term instead of *unique* terms.) 5 is not a member of the sequence since it is representable in *two* ways, $5 = 1 + 4 = 2 + 3$, but $6 = 2 + 4$ is a member. Proceeding in the manner, we can generate Ulam sequences for any (u, v), examples of which are given below.

$$\begin{aligned}
(1,2) &= \{1,2,3,4,6,8,11,13,16,18,\ldots\}\\
(1,3) &= \{1,3,4,5,6,8,10,12,17,21,\ldots\}\\
(1,4) &= \{1,4,5,6,7,8,10,16,18,19,\ldots\}\\
(1,5) &= \{1,5,6,7,8,9,10,12,20,22,\ldots\}\\
(2,3) &= \{2,3,5,7,8,9,13,14,18,19,\ldots\}\\
(2,4) &= \{2,4,6,8,12,16,22,26,32,36,\ldots\}\\
(2,5) &= \{2,5,7,9,11,12,13,15,19,23,\ldots\}.
\end{aligned}$$

Schmerl and Spiegel (1994) proved that Ulam sequences $(2, v)$ for ODD $v \geq 5$ have exactly two EVEN terms. Ulam sequences with only finitely many EVEN terms eventually must have periodic successive differences (Finch 1991, 1992abc). Cassaigne and Finch (1995) proved that the Ulam sequences $(4, v)$ for $5 \leq v \equiv 1 \pmod 4$ have exactly three EVEN terms.

The Ulam sequence can be generalized by the s-ADDITIVE SEQUENCE.

see also GREEDY ALGORITHM, s-ADDITIVE SEQUENCE, STÖHR SEQUENCE

References

Cassaigne, J. and Finch, S. "A Class of 1-Additive Sequences and Quadratic Recurrences." *Exper. Math* **4**, 49–60, 1995.

Finch, S. "Conjectures About 1-Additive Sequences." *Fib. Quart.* **29**, 209–214, 1991.

Finch, S. "Are 0-Additive Sequences Always Regular?" *Amer. Math. Monthly* **99**, 671–673, 1992a.

Finch, S. "On the Regularity of Certain 1-Additive Sequences." *J. Combin. Th. Ser. A* **60**, 123–130, 1992b.

Finch, S. "Patterns in 1-Additive Sequences." *Exper. Math.* **1**, 57–63, 1992c.

Finch, S. "Ulam s-Additive Sequences." `http://www.mathsoft.com/asolve/sadd/sadd.html`.

Guy, R. K. "A Quarter Century of *Monthly* Unsolved Problems, 1969–1993." *Amer. Math. Monthly* **100**, 945–949, 1993.

Guy, R. K. "Ulam Numbers." §C4 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 109–110, 1994.

Guy, R. K. and Nowakowski, R. J. "*Monthly* Unsolved Problems, 1969–1995." *Amer. Math. Monthly* **102**, 921–926, 1995.

Recaman, B. "Questions on a Sequence of Ulam." *Amer. Math. Monthly* **80**, 919–920, 1973.

Schmerl, J. and Spiegel, E. "The Regularity of Some 1-Additive Sequences." *J. Combin. Theory Ser. A* **66**, 172–175, 1994.

Sloane, N. J. A. Sequence A002858/M0557 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Ultrametric

An ultrametric is a METRIC which satisfies the following strengthened version of the TRIANGLE INEQUALITY,

$$d(x,z) \leq \max(d(x,y), d(y,z))$$

for all x, y, z. At least two of $d(x,y)$, $d(y,z)$, and $d(x,z)$ are the same.

Let X be a Set, and let $X^{\mathbb{N}}$ (where $\mathbb{N}$ is the Set of Natural Numbers) denote the collection of sequences of elements of X (i.e., all the possible sequences x_1, x_2, x_3, …). For sequences $a = (a_1, a_2, \ldots)$, $b = (b_1, b_2, \ldots)$, let n be the number of initial places where the sequences agree, i.e., $a_1 = b_1$, $a_2 = b_2$, …, $a_n = b_n$, but $a_{n+1} \neq b_{n+1}$. Take $n = 0$ if $a_1 \neq b_1$. Then defining $d(a,b) = 2^{-n}$ gives an ultrametric.

The p-adic Number metric is another example of an ultrametric.

see also Metric, p-adic Number

Ultraradical

A symbol which can be used to express solutions not obtainable by finite Root extraction. The solution to the irreducible Quintic Equation

$$x^5 + x = a$$

is written $\smallint\!\overline{a}$.

see also Radical

Ultraspherical Differential Equation

$$(1-x^2)y'' - (2\alpha+1)xy' + n(n+2\alpha)y = 0. \quad (1)$$

Alternate forms are

$$(1-x^2)Y'' + (2\lambda-3)xY' + (n+1)(n+2\lambda-1)Y = 0, \quad (2)$$

where

$$Y = (1-x^2)^{\lambda-1/2} P_n^{(\lambda)}(x), \quad (3)$$

$$\frac{d^2u}{dx^2} + \left[\frac{(n+\lambda)^2}{1-x^2} + \frac{\frac{1}{2} + \lambda - \lambda^2 + \frac{1}{4}x^2}{(1-x^2)^2}\right] u = 0, \quad (4)$$

where

$$u = (1-x^2)^{\lambda/2+1/4} P_n^{(\lambda)}(x), \quad (5)$$

and

$$\frac{d^2u}{d\theta^2} + \left[(n+\lambda)^2 + \frac{\lambda(1-\lambda)}{\sin^2\theta}\right] u = 0, \quad (6)$$

where

$$u = \sin^\lambda \theta \, P_n^{(\lambda)}(\cos\theta). \quad (7)$$

The solutions are the Ultraspherical Functions $P_n^{(\lambda)}(x)$. For integral n with $\alpha < 1/2$, the function converges to the Ultraspherical Polynomials $C_n^{(\alpha)}(x)$.

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 547–549, 1953.

Ultraspherical Function

A function defined by a Power Series whose coefficients satisfy the Recurrence Relation

$$a_{j+2} = a_j \frac{(k+j)(k+j+2\alpha) - n(n+2\alpha)}{(k+j+1)(k+j+2)}.$$

For $x \neq -1$, the function converges for $\alpha < 1/2$ and diverges for $\alpha > 1/2$.

Ultraspherical Polynomial

The ultraspherical polynomials are solutions $P_n^{(\lambda)}(x)$ to the Ultraspherical Differential Equation for Integer n and $\alpha < 1/2$. They are generalizations of Legendre Polynomials to $(n+2)$-D space and are proportional to (or, depending on the normalization, equal to) the Gegenbauer Polynomials $C_n^{(\lambda)}(x)$, denoted in *Mathematica*® (Wolfram Research, Champaign, IL) `GegenbauerC[n,lambda,x]`. The ultraspherical polynomials are also Jacobi Polynomials with $\alpha = \beta$. They are given by the Generating Function

$$\frac{1}{(1-2xt+t^2)^\lambda} = \sum_{n=0}^{\infty} P_n^{(\lambda)}(x) t^n, \quad (1)$$

and can be given explicitly by

$$P_n^{(\lambda)}(x) = \frac{\Gamma(\lambda+\frac{1}{2})}{\Gamma(2\lambda)} \frac{\Gamma(n+2\lambda)}{\Gamma(n+\lambda+\frac{1}{2})} P_n^{(\lambda-1/2,\lambda-1/2)}(x), \quad (2)$$

where $P_n^{(\lambda-1/2,\lambda-1/2)}$ is a Jacobi Polynomial (Szegő 1975, p. 80). The first few ultraspherical polynomials are

$$P_0^{(\lambda)}(x) = 1 \quad (3)$$

$$P_1^{(\lambda)}(x) = 2\lambda x \quad (4)$$

$$P_2^{(\lambda)}(x) = -\lambda + 2\lambda(1+\lambda)x^2 \quad (5)$$

$$P_3^{(\lambda)}(x) = -2\lambda(1+\lambda)x + \tfrac{4}{3}\lambda(1+\lambda)(2+\lambda)x^3. \quad (6)$$

In terms of the Hypergeometric Functions,

$$P_n^{(\lambda)}(x) = \binom{n+2\lambda-1}{n} \times {}_2F_1(-n, n+2\lambda; \lambda+\tfrac{1}{2}; \tfrac{1}{2}(1-x)) \quad (7)$$

$$= 2^n \binom{n+\lambda-1}{n} (x-1)^n \times {}_2F_1\left(-n, -n-\lambda+\tfrac{1}{2}; -2n-2\lambda+1; \frac{2}{1-x}\right) \quad (8)$$

$$= \binom{n+2\lambda+1}{n} \left(\frac{x+1}{2}\right)^n \times {}_2F_1\left(-n, -n-\lambda+\tfrac{1}{2}; \lambda+\tfrac{1}{2}; \frac{x-1}{x+1}\right). \quad (9)$$

They are normalized by

$$\int_{-1}^{1}(1-x^2)^{\lambda-1/2}[P_n^{(\lambda)}]^2\,dx = 2^{1-2\lambda}\pi\frac{\Gamma(n+2\lambda)}{(n+\lambda)\Gamma^2(\lambda)\Gamma(n+1)}. \tag{10}$$

Derivative identities include

$$\frac{d}{dx}P_n^{(\lambda)}(x) = 2\lambda P_{n-1}^{(\lambda+1)}(x) \tag{11}$$

$$(1-x^2)\frac{d}{dx}[P_n^{(\lambda)}] = [2(n+\lambda)]^{-1}[(n+2\lambda-1) \times(n+2\lambda)P_{n-1}^{(\lambda)}(x) - n(n+1)P_{n+1}^{(\lambda)}(x)] \tag{12}$$

$$= -nxP_n^{(\lambda)}(x) + (n+2\lambda-1)P_{n-1}^{(\lambda)}(x) \tag{13}$$

$$= (n+2\lambda)xP_n^{(\lambda)}(x) - (n+1)P_{n+1}^{(\lambda)}(x) \tag{14}$$

$$nP_n^{(\lambda)}(x) = x\frac{d}{dx}[P_n^{(\lambda)}(x)] - \frac{d}{dx}[P_{n-1}^{(\lambda)}(x)] \tag{15}$$

$$(n+2\lambda)P_n^{(\lambda)}(x) = \frac{d}{dx}[P_{n+1}^{(\lambda)}(x)] - x\frac{d}{dx}[P_n^{(\lambda)}(x)] \tag{16}$$

$$\frac{d}{dx}[P_{n+1}^{(\lambda)}(x) - P_{n-1}^{(\lambda)}(x)] = 2(n+\lambda)P_n^{(\lambda)}P_n^{(\lambda)}(x) \tag{17}$$

$$= 2\lambda[P_n^{(\lambda+1)}(x) - P_{n-2}^{(\lambda+1)}(x)] \tag{18}$$

(Szegő 1975, pp. 80–83).

A Recurrence Relation is

$$nP_n^{(\lambda)}(x) = 2(n+\lambda-1)xP_{n-1}^{(\lambda)}(x) - (n+2\lambda-2)P_{n-2}^{(\lambda)}(x) \tag{19}$$

for $n = 2, 3, \ldots$.

Special double-ν Formulas also exist

$$P_{2\nu}^{(\lambda)}(x) = \binom{2\nu+2\lambda-1}{2\nu}{}_2F_1(-\nu,\nu+\lambda;\lambda+\tfrac{1}{2};1-x^2) \tag{20}$$

$$= (-1)^\nu\binom{\nu+\lambda-1}{\nu}{}_2F_1(-\nu,\nu+\lambda;\tfrac{1}{2};x^2) \tag{21}$$

$$P_{2\nu+1}^{(\lambda)}(x) = \binom{2\nu+2\lambda}{2\nu+1}x\,{}_2F_1(-\nu,\nu+\lambda+1;\lambda+\tfrac{1}{2};1-x^2) \tag{22}$$

$$= (-1)^\nu 2\lambda\binom{\nu+\lambda}{\nu}x\,{}_2F_1(-\nu,\nu+\lambda+1;\tfrac{3}{2};x^2). \tag{23}$$

Special values are given in the following table.

λ	Special Polynomial
$\frac{1}{2}$	Legendre
1	Chebyshev polynomial of the second kind

Koschmieder (1920) gives representations in terms of Elliptic Functions for $\alpha = -3/4$ and $\alpha = -2/3$.

see also Birthday Problem, Chebyshev Polynomial of the Second Kind, Elliptic Function, Hypergeometric Function, Jacobi Polynomial

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Orthogonal Polynomials." Ch. 22 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 771–802, 1972.

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 643, 1985.

Iyanaga, S. and Kawada, Y. (Eds.). "Gegenbauer Polynomials (Gegenbauer Functions)." Appendix A, Table 20.I in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 1477–1478, 1980.

Koschmieder, L. "Über besondere Jacobische Polynome." *Math. Zeitschrift* **8**, 123–137, 1920.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 547–549 and 600–604, 1953.

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., 1975.

Umbilic Point

A point on a surface at which the Curvature is the same in any direction.

Umbral Calculus

The study of certain properties of Finite Differences. The term was coined by Sylvester from the word "umbra" (meaning "shadow" in Latin), and reflects the fact that for many types of identities involving sequences of polynomials with Powers a^n, "shadow" identities are obtained when the polynomials are changed to discrete values and the exponent in a^n is changed to the Pochhammer Symbol $(a)_n \equiv a(a-1)\cdots(a-n+1)$.

For example, Newton's Forward Difference Formula written in the form

$$f(x+a) = \sum_{n=0}^{\infty}\frac{(a)_n\Delta^n f(x)}{n!} \tag{1}$$

with $f(x+a) \equiv f_{x+a}$ looks suspiciously like a finite analog of the Taylor Series expansion

$$f(x+a) = \sum_{n=0}^{\infty}\frac{a^n\tilde{D}^n f(x)}{n!}, \tag{2}$$

where $\tilde{D}$ is the Differential Operator. Similarly, the Chu-Vandermonde Identity

$$(x+a)_n = \sum_{k=0}^{\infty}\binom{n}{k}(a)_k(x)_{n-k} \tag{3}$$

with $\binom{n}{k}$ a Binomial Coefficient, looks suspiciously like an analog of the Binomial Theorem

$$(x+a)^n = \sum_{k=0}^{\infty}\binom{n}{k}a^k x^{n-k} \tag{4}$$

(Di Bucchianico and Loeb).

see also BINOMIAL THEOREM, CHU-VANDERMONDE IDENTITY, FINITE DIFFERENCE

References
Roman, S. and Rota, G.-C. "The Umbral Calculus." *Adv. Math.* **27**, 95–188, 1978.
Roman, S. *The Umbral Calculus.* New York: Academic Press, 1984.

Umbrella

see WHITNEY UMBRELLA

Unambiguous

see WELL-DEFINED

Unbiased

A quantity which does not exhibit BIAS. An ESTIMATOR $\hat{\theta}$ is an UNBIASED ESTIMATOR of θ if

$$\langle\hat{\theta}\rangle = \theta.$$

see also BIAS (ESTIMATOR), ESTIMATOR

Uncia

$$1 \text{ uncia} \equiv \tfrac{1}{12}.$$

The word *uncia* was Latin for a unit equal to 1/12 of another unit called the *as*. The words "inch" (1/12 of a foot) and "ounce" (originally 1/12 of a pound and still 1/12 of a "Troy pound," now used primarily to weigh precious metals) are derived from the word *uncia*.

see also CALCUS, HALF, QUARTER, SCRUPLE, UNIT FRACTION

References
Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 4, 1996.

Uncorrelated

Variables x_i and x_j are said to be uncorrelated if their COVARIANCE is zero:

$$\operatorname{cov}(x_i, x_j) = 0.$$

INDEPENDENT STATISTICS are always uncorrelated, but the converse is not necessarily true.

see also COVARIANCE, INDEPENDENT STATISTICS

Uncountable Set

see UNCOUNTABLY INFINITE SET

Uncountably Infinite Set

An INFINITE SET which is not a COUNTABLY INFINITE SET.

see also ALEPH-0, ALEPH-1, COUNTABLE SET, COUNTABLY INFINITE SET, FINITE, INFINITE

Undecagon

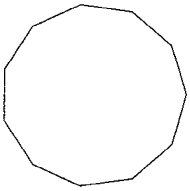

The unconstructible 11-sided POLYGON with SCHLÄFLI SYMBOL $\{11\}$.

see also DECAGON, DODECAGON, TRIGONOMETRY VALUES—$\pi/11$

Undecidable

Not DECIDABLE as a result of being neither formally provable nor unprovable.

see also GÖDEL'S INCOMPLETENESS THEOREM, RICHARDSON'S THEOREM

Undecillion

In the American system, 10^{36}.

see also LARGE NUMBER

Undetermined Coefficients Method

Given a nonhomogeneous ORDINARY DIFFERENTIAL EQUATION, select a differential operator which will annihilate the right side, and apply it to both sides. Find the solution to the homogeneous equation, plug it into the left side of the original equation, and solve for constants by setting it equal to the right side. The solution is then obtained by plugging the determined constants into the homogeneous equation.

see also ORDINARY DIFFERENTIAL EQUATION

Undulating Number

A number of the form $aba\cdots$, $abab\cdots$, etc. The first few nontrivial undulants (with the stipulation that $a \neq b$) are 101, 121, 131, 141, 151, 161, 171, 181, 191, 202, 212, ... (Sloane's A046075). Including the trivial 1- and 2-digit undulants and dropping the requirement that $a \neq b$ gives Sloane's A033619.

The first few undulating SQUARES are 121, 484, 676, 69696, ... (Sloane's A016073), with no larger such numbers of fewer than a million digits (Pickover 1995). Several tricks can be used to speed the search for square undulating numbers, especially by examining the possible patterns of ending digits. For example, the only possible sets of four trailing digits for undulating SQUARES are 0404, 1616, 2121, 2929, 3636, 6161, 6464, 6969, 8484, and 9696.

The only undulating POWER $n^p = aba\cdots$ for $3 \leq p \leq 31$ and up to 100 digits is $7^3 = 343$ (Pickover 1995). A large undulating prime is given by $7 + 720(100^{49} - 1)/99$ (Pickover 1995).

A binary undulant is a POWER of 2 whose base-10 representation contains one or both of the sequences $010\cdots$ and $101\cdots$. The first few are 2^n for $n = 103$, 107, 138, 159, 179, 187, 192, 199, 205, ... (Sloane's A046076). The smallest n for which an undulating sequence of *exactly d*-digit occurs for $d = 3$, 4, ... are $n = 103$, 138, 875, 949, 6617, 1802, 14545, ... (Sloane's A046077). An undulating binary sequence of length 10 occurs for $n = 1,748,219$ (Pickover 1995).

References
Pickover, C. A. "Is There a Double Smoothly Undulating Integer?" In *Computers, Pattern, Chaos and Beauty.* New York: St. Martin's Press, 1990.
Pickover, C. A. "The Undulation of the Monks." Ch. 20 in *Keys to Infinity.* New York: W. H. Freeman, pp. 159–161 1995.
Sloane, N. J. A. Sequences A016073, A033619, A046075, A046076, and A046077 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Unduloid

A SURFACE OF REVOLUTION with constant NONZERO MEAN CURVATURE also called an ONDULOID. It is a ROULETTE obtained from the path described by the FOCI of a CONIC SECTION when rolled on a LINE. This curve then generates an unduloid when revolved about the LINE. These curves are special cases of the shapes assumed by soap film spanning the gap between prescribed boundaries. The unduloid of a PARABOLA gives a CATENOID.

see also CALCULUS OF VARIATIONS, CATENOID, ROULETTE

References
Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., p. 48, 1989.
Delaunay, C. "Sur la surface de révolution dont la courbure moyenne est constante." *J. math. pures appl.* **6**, 309–320, 1841.
do Carmo, M. P. "The Onduloid." §3.5G in *Mathematical Models from the Collections of Universities and Museums* (Ed. G. Fischer). Braunschweig, Germany: Vieweg, pp. 47–48, 1986.
Fischer, G. (Ed.). Plate 97 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, p. 93, 1986.
Thompson, D'A. W. *On Growth and Form, 2nd ed., compl. rev. ed.* New York: Cambridge University Press, 1992.
Yates, R. C. *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, p. 184, 1952.

Unexpected Hanging Paradox

A PARADOX also known as the SURPRISE EXAMINATION PARADOX or PREDICTION PARADOX.

A prisoner is told that he will be hanged on some day between Monday and Friday, but that he will not know on which day the hanging will occur before it happens. He cannot be hanged on Friday, because if he were still alive on Thursday, he would know that the hanging will occur on Friday, but he has been told he will not know the day of his hanging in advance. He cannot be hanged Thursday for the same reason, and the same argument shows that he cannot be hanged on any other day. Nevertheless, the executioner unexpectedly arrives on some day other than Friday, surprising the prisoner.

This PARADOX is similar to that in Robert Louis Stevenson's "The Imp in the Bottle," in which you are offered the opportunity to buy, for whatever price you wish, a bottle containing a genie who will fulfill your every desire. The only catch is that the bottle must thereafter be resold for a price smaller than what you paid for it, or you will be condemned to live out the rest of your days in excrutiating torment. Obviously, no one would buy the bottle for 1¢ since he would have to give the bottle away, but no one would accept the bottle knowing he would be unable to get rid of it. Similarly, no one would buy it for 2¢, and so on. However, for some reasonably large amount, it will always be possible to find a next buyer, so the bottle will be bought (Paulos 1995).

see also SORITES PARADOX

References
Chow, T. Y. "The Surprise Examination or Unexpected Hanging Paradox." *Amer. Math. Monthly* **105**, 41–51, 1998.
Clark, D. "How Expected is the Unexpected Hanging?" *Math. Mag.* **67**, 55–58, 1994.
Gardner, M. "The Paradox of the Unexpected Hanging." Ch. 1 in *The Unexpected Hanging and Other Mathematical Diversions.* Chicago, IL: Chicago University Press, 1991.
Margalit, A. and Bar-Hillel, M. "Expecting the Unexpected." *Philosophia* **13**, 263–288, 1983.
Pappas, T. "The Paradox of the Unexpected Exam." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 147, 1989.
Paulos, J. A. *A Mathematician Reads the Newspaper.* New York: BasicBooks, p. 97, 1995.
Quine, W. V. O. "On a So-Called Paradox." *Mind* **62**, 65–67, 1953.

Unfinished Game

see SHARING PROBLEM

Unhappy Number

A number which is not HAPPY is said to be unhappy.

see also HAPPY NUMBER

Unicursal Circuit

A CIRCUIT in which an entire GRAPH is traversed in one route. An example of a curve which can be traced unicursally is the MOHAMMED SIGN.

Uniform Apodization Function

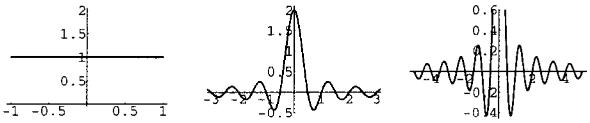

An APODIZATION FUNCTION

$$f(x) = 1, \tag{1}$$

having INSTRUMENT FUNCTION

$$I(x) = \int_{-a}^{a} e^{-2\pi ikx}\,dx = -\frac{1}{2\pi ik}(e^{-2\pi ika} - e^{2\pi ikx})$$
$$= \frac{\sin(2\pi ka)}{\pi k} = 2a\,\mathrm{sinc}(2\pi ka). \qquad (2)$$

The peak (in units of a) is 2. The extrema are given by letting $\beta \equiv 2\pi ka$ and solving

$$\frac{d}{d\beta}(\beta\sin\beta) = \frac{\sin\beta - \beta\cos\beta}{\beta^2} = 0 \qquad (3)$$

$$\sin\beta - \beta\cos\beta = 0 \qquad (4)$$

$$\tan\beta = \beta. \qquad (5)$$

Solving this numerically gives $\beta_0 = 0$, $\beta_1 = 4.49341$, $\beta_2 = 7.72525$, ... for the first few solutions. The second of these is the peak POSITIVE sidelobe, and the third is the peak NEGATIVE sidelobe. As a fraction of the peak, they are 0.128375 and −0.217234. The FULL WIDTH AT HALF MAXIMUM is found by setting $I(x) = 1$

$$\mathrm{sinc}(x) = \tfrac{1}{2}, \qquad (6)$$

and solving for $x_{1/2}$, yielding

$$x_{1/2} = 2\pi k_{1/2}a = 1.89549. \qquad (7)$$

Therefore, with $L \equiv 2a$,

$$\mathrm{FWHM} = 2k_{1/2} = \frac{0.603353}{a} = \frac{1.20671}{L}. \qquad (8)$$

see also APODIZATION FUNCTION

Uniform Boundedness Principle

If a "pointwise-bounded" family of continuous linear OPERATORS from a BANACH SPACE to a NORMED SPACE is "uniformly bounded." Symbolically, if $\sup ||T_i(x)||$ is FINITE for each x in the unit BALL, then $\sup ||T_i||$ is FINITE. The theorem is also called the BANACH-STEINHAUS THEOREM.

References

Zeidler, E. *Applied Functional Analysis: Applications to Mathematical Physics.* New York: Springer-Verlag, 1995.

Uniform Convergence

A SERIES $\sum_{n=1}^{\infty} u_n(x)$ is uniformly convergent to $S(x)$ for a set E of values of x if, for each $\epsilon > 0$, an INTEGER N can be found such that

$$|S_n(x) - S(x)| < \epsilon \qquad (1)$$

for $n \geq N$ and all $x \in E$. To test for uniform convergence, use ABEL'S UNIFORM CONVERGENCE TEST or the WEIERSTRAß M-TEST. If individual terms $u_n(x)$ of a uniformly converging series are continuous, then

1. The series sum

$$f(x) = \sum_{n=1}^{\infty} u_n(x) \qquad (2)$$

is continuous,

2. The series may be integrated term by term

$$\int_a^b f(x)\,dx = \sum_{n=1}^{\infty}\int_a^b u_n(x)\,dx, \qquad (3)$$

and

3. The series may be differentiated term by term

$$\frac{d}{dx}f(x) = \sum_{n=1}^{\infty} n\frac{d}{dx}u_n(x). \qquad (4)$$

see also ABEL'S THEOREM, ABEL'S UNIFORM CONVERGENCE TEST, WEIERSTRAß M-TEST

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 299–301, 1985.

Uniform Distribution

A distribution which has constant probability is called a uniform distribution, sometimes also called a RECTANGULAR DISTRIBUTION. The probability density function and cumulative distribution function for a *continuous* uniform distribution are

$$P(x) = \begin{cases} \frac{1}{b-a} & \text{for } a < x < b \\ 0 & \text{for } x < a,\ x > b \end{cases} \qquad (1)$$

$$D(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a \leq x < b \\ 1 & \text{for } x \geq b. \end{cases} \qquad (2)$$

With $a = 0$ and $b = 1$, these can be written

$$P(x) = \tfrac{1}{2}\,\mathrm{sgn}(x) - \mathrm{sgn}(x-1) \qquad (3)$$
$$D(x) = \tfrac{1}{2}[1 - (1-x)^2\,\mathrm{sgn}(1-x) + x\,\mathrm{sgn}(x)]. \qquad (4)$$

The CHARACTERISTIC FUNCTION is

$$\phi(t) = \frac{2}{ht}\sin(\tfrac{1}{2}ht)e^{imt}, \qquad (5)$$

where

$$a = m - \tfrac{1}{2}h \qquad (6)$$
$$b = m + \tfrac{1}{2}h. \qquad (7)$$

The MOMENT-GENERATING FUNCTION is

$$M(t) = \langle e^{xt}\rangle = \int_a^b \frac{e^{xt}}{b-a}dx = \left[\frac{e^{xt}}{t(b-a)}\right]_a^b, \qquad (8)$$

so

$$M(t) = \begin{cases} \frac{e^{tb}-e^{ta}}{t(b-a)} & \text{for } t \neq 0 \\ 0 & \text{for } t = 0, \end{cases} \tag{9}$$

and

$$\begin{aligned} M'(t) &= \frac{1}{b-a}\left[\frac{1}{t}(be^{bt} - ae^{at}) - \frac{1}{t^2}(e^{bt} - e^{at})\right] \\ &= \frac{e^{bt}(bt-1) - e^{at}(at-1)}{(b-a)t^2}. \end{aligned} \tag{10}$$

The function is not differentiable at zero, so the MOMENTS cannot be found using the standard technique. They can, however, be found by direct integration. The MOMENTS about 0 are

$$\mu'_1 = \tfrac{1}{2}(a+b) \tag{11}$$
$$\mu'_2 = \tfrac{1}{3}(a^2 + ab + b^2) \tag{12}$$
$$\mu'_3 = \tfrac{1}{4}(a+b)(a^2+b^2) \tag{13}$$
$$\mu'_4 = \tfrac{1}{5}(a^4 + a^3b + a^2b^2 + ab^3 + b^4). \tag{14}$$

The MOMENTS about the MEAN are

$$\mu_1 = 0 \tag{15}$$
$$\mu_2 = \tfrac{1}{12}(b-a)^2 \tag{16}$$
$$\mu_3 = 0 \tag{17}$$
$$\mu_4 = \tfrac{1}{80}(b-a)^4, \tag{18}$$

so the MEAN, VARIANCE, SKEWNESS, and KURTOSIS are

$$\mu = \tfrac{1}{2}(a+b) \tag{19}$$
$$\sigma^2 = \mu_2 = \tfrac{1}{12}(b-a)^2 \tag{20}$$
$$\gamma_1 = \frac{\mu_3}{\sigma^{3/2}} = 0 \tag{21}$$
$$\gamma_2 = -\tfrac{6}{5}. \tag{22}$$

The probability distribution function and cumulative distributions function for a *discrete* uniform distribution are

$$P(n) = \frac{1}{N} \tag{23}$$
$$D(n) = \frac{n}{N} \tag{24}$$

for $n = 1, \ldots, N$. The MOMENT-GENERATING FUNCTION is

$$\begin{aligned} M(t) = \langle e^{nt} \rangle &= \sum_{n=1}^{N} \frac{1}{N} e^{nt} = \frac{1}{N}\frac{e^t - e^{t(N+1)}}{1-e^t} \\ &= \frac{e^t(1-e^{Nt})}{N(1-e^t)}. \end{aligned} \tag{25}$$

The MOMENTS about 0 are

$$\mu'_m = \frac{1}{N}\sum_{n=1}^{N} n^m, \tag{26}$$

so

$$\mu'_1 = \tfrac{1}{2}(N+1) \tag{27}$$
$$\mu'_2 = \tfrac{1}{6}(N+1)(2N+1) \tag{28}$$
$$\mu'_3 = \tfrac{1}{4}N(N+1)^2 \tag{29}$$
$$\mu'_4 = \tfrac{1}{30}(N+1)(2N+1)(3N^2+3N-1), \tag{30}$$

and the MOMENTS about the MEAN are

$$\mu_2 = \tfrac{1}{12}(N-1)(N+1) \tag{31}$$
$$\mu_3 = 0 \tag{32}$$
$$\mu_4 = \tfrac{1}{240}(N-1)(N+1)(3N^2-7). \tag{33}$$

The MEAN, VARIANCE, SKEWNESS, and KURTOSIS are

$$\mu = \tfrac{1}{2}(N+1) \tag{34}$$
$$\sigma^2 = \mu_2 = \tfrac{1}{12}(N-1)(N+1) \tag{35}$$
$$\gamma_1 = \frac{\mu_3}{\sigma^{3/2}} = 0 \tag{36}$$
$$\gamma_2 = \frac{6(N^2+1)}{5(N-1)(N+1)}. \tag{37}$$

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 531 and 533, 1987.

Uniform Polyhedron

The uniform polyhedra are POLYHEDRA with identical VERTICES. Coxeter *et al.* (1954) conjectured that there are 75 such polyhedra in which only two faces are allowed to meet at an EDGE, and this was subsequently proven. (However, when any EVEN number of faces may meet, there are 76 polyhedra.) If the five pentagonal PRISMS are included, the number rises to 80.

The VERTICES of a uniform polyhedron all lie on a SPHERE whose center is their CENTROID. The VERTICES joined to another VERTEX lie on a CIRCLE.

Source code and binary programs for generating and viewing the uniform polyhedra are also available at `http://www.math.technion.ac.il/~rl/kaleido/`. The following depictions of the polyhedra were produced by R. Maeder's `UniformPolyhedra.m` package for *Mathematica*® (Wolfram Research, Champaign, IL). Due to a limitation in *Mathematica*'s renderer, uniform polyhedra 69, 72, 74, and 75 cannot be displayed using this package.

n	Name/Dual
1	tetrahedron
	tetrahedron
2	truncated tetrahedron
	triakis tetrahedron
3	octahemioctahedron
	octahemioctacron
4	tetrahemihexahedron
	tetrahemihexacron
5	octahedron
	cube
6	cube
	octahedron
7	cuboctahedron
	rhombic dodecahedron
8	truncated octahedron
	tetrakis hexahedron
9	truncated cube
	triakis octahedron
10	small rhombicuboctahedron
	deltoidal icositetrahedron
11	truncated cuboctahedron
	disdyakis dodecahedron
12	snub cube
	pentagonal icositetrahedron
13	small cubicuboctahedron
	small hexacronic icositetrahedron
14	great cubicuboctahedron
	great hexacronic icositetrahedron
15	cubohemioctahedron
	hexahemioctahedron
16	cubitruncated cuboctahedron
	tetradyakis hexahedron
17	great rhombicuboctahedron
	great deltoidal icositetrahedron
18	small rhombihexahedron
	small rhombihexacron
19	stellated truncated hexahedron
	great triakis octahedron
20	great truncated cuboctahedron
	great disdyakis dodecahedron
21	great rhombihexahedron
	great rhombihexacron
22	icosahedron
	dodecahedron
23	dodecahedron
	icosahedron
24	icosidodecahedron
	rhombic triacontahedron
25	truncated icosahedron
	pentakis dodecahedron
26	truncated dodecahedron
	triakis icosahedron
27	small rhombicosidodecahedron
	deltoidal hexecontahedron
28	truncated icosidodecahedron
	disdyakis triacontahedron
29	snub dodecahedron
	pentagonal hexecontahedron
30	small ditrigonal icosidodecahedron
	small triambic icosahedron
31	small icosicosidodecahedron
	small icosacronic hexecontahedron
32	small snub icosicosidodecahedron
	small hexagonal hexecontahedron
33	small dodecicosidodecahedron
	small dodecacronic hexecontahedron
34	small stellated dodecahedron
	great dodecahedron
35	great dodecahedron
	small stellated dodecahedron
36	dodecadodecahedron
	medial rhombic triacontahedron
37	truncated great dodecahedron
	small stellapentakis dodecahedron
38	rhombidodecadodecahedron
	medial deltoidal hexecontahedron
39	small rhombidodecahedron
	small rhombidodecacron
40	snub dodecadodecahedron
	medial pentagonal hexecontahedron
41	ditrigonal dodecadodecahedron
	medial triambic icosahedron
42	great ditrigonal dodecicosidodecahedron
	great ditrigonal dodecacronic hexecontahedron
43	small ditrigonal dodecicosidodecahedron
	small ditrigonal dodecacronic hexecontahedron
44	icosidodecadodecahedron
	medial icosacronic hexecontahedron
45	icositruncated dodecadodecahedron
	tridyakis icosahedron
46	snub icosidodecadodecahedron
	medial hexagonal hexecontahedron
47	great ditrigonal icosidodecahedron
	great triambic icosahedron
48	great icosicosidodecahedron
	great icosacronic hexecontahedron
49	small icosihemidodecahedron
	small icosihemidodecacron
50	small dodecicosahedron
	small dodecicosacron

n	Name/Dual
51	small dodecahemidodecahedron
	small dodecahemidodecacron
52	great stellated dodecahedron
	great icosahedron
53	great icosahedron
	great stellated dodecahedron
54	great icosidodecahedron
	great rhombic triacontahedron
55	great truncated icosahedron
	great stellapentakis dodecahedron
56	rhombicosahedron
	rhombicosacron
57	great snub icosidodecahedron
	great pentagonal hexecontahedron
58	small stellated truncated dodecahedron
	great pentakis dodecahedron
59	truncated dodecadodecahedron
	medial disdyakis triacontahedron
60	inverted snub dodecadodecahedron
	medial inverted pentagonal hexecontahedron
61	great dodecicosidodecahedron
	great dodecacronic hexecontahedron
62	small dodecahemicosahedron
	small dodecahemicosacron
63	great dodecicosahedron
	great dodecicosacron
64	great snub dodecicosidodecahedron
	great hexagonal hexecontahedron
65	great dodecahemicosahedron
	great dodecahemicosacron
66	great stellated truncated dodecahedron
	great triakis icosahedron
67	great rhombicosidodecahedron
	great deltoidal hexecontahedron
68	great truncated icosidodecahedron
	great disdyakis triacontahedron
69	great inverted snub icosidodecahedron
	great inverted pentagonal hexecontahedron
70	great dodecahemidodecahedron
	great dodecahemidodecacron
71	great icosihemidodecahedron
	great icosihemidodecacron
72	small retrosnub icosicosidodecahedron
	small hexagrammic hexecontahedron
73	great rhombidodecahedron
	great rhombidodecacron
74	great retrosnub icosidodecahedron
	great pentagrammic hexecontahedron
75	great dirhombicosidodecahedron
	great dirhombicosidodecacron

n	Name/Dual
76	pentagonal prism
	pentagonal dipyramid
77	pentagonal antiprism
	pentagonal deltahedron
78	pentagrammic prism
	pentagrammic dipyramid
79	pentagrammic antiprism
	pentagrammic deltahedron
80	pentagrammic crossed antiprism
	pentagrammic concave deltahedron

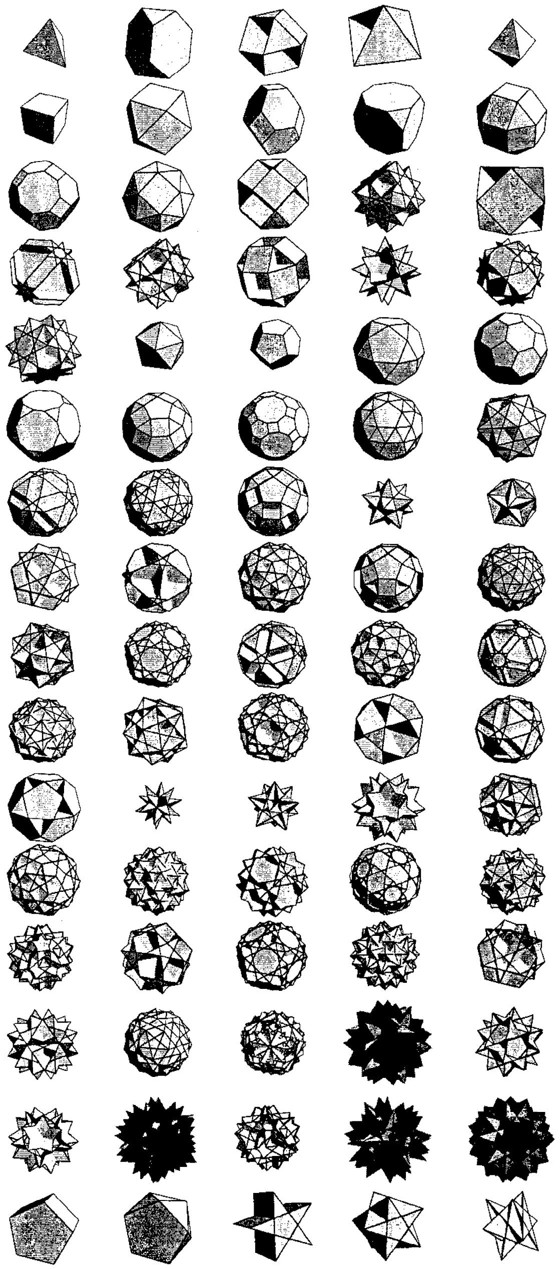

see also Archimedean Solid, Augmented Polyhedron, Johnson Solid, Kepler-Poinsot Solid, Platonic Solid, Polyhedron, Vertex Figure, Wythoff Symbol

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 136, 1987.

Bulatov, V.v "Compounds of Uniform Polyhedra." `http://www.physics.orst.edu/~bulatov/polyhedra/uniform_compounds/`.

Bulatov, V. "Dual Uniform Polyhedra." `http://www.physics.orst.edu/~bulatov/polyhedra/dual/`.

Bulatov, V. "Uniform Polyhedra." `http://www.physics.orst.edu/~bulatov/polyhedra/uniform/`.

Coxeter, H. S. M.; Longuet-Higgins, M. S.; and Miller, J. C. P. "Uniform Polyhedra." *Phil. Trans. Roy. Soc. London Ser. A* **246**, 401–450, 1954.

Har'El, Z. "Uniform Solution for Uniform Polyhedra." *Geometriae Dedicata* **47**, 57–110, 1993.

Har'El, Z. "Kaleido." `http://www.math.technion.ac.il/~rl/kaleido/`.

Har'El, Z. "Eighty Dual Polyhedra Generated by Kaleido." `http://www.math.technion.ac.il/~rl/kaleido/dual.html`.

Har'El, Z. "Eighty Uniform Polyhedra Generated by Kaleido." `http://www.math.technion.ac.il/~rl/kaleido/poly.html`.

Hume, A. "Exact Descriptions of Regular and Semi-Regular Polyhedra and Their Duals." Computing Science Tech. Rept. No. 130. Murray Hill, NJ: AT&T Bell Lab., 1986.

Hume, A. Information files on polyhedra. `http://netlib.bell-labs.com/netlib/polyhedra/`.

Johnson, N. W. "Convex Polyhedra with Regular Faces." *Canad. J. Math.* **18**, 169–200, 1966.

Maeder, R. E. "Uniform Polyhedra." *Mathematica J.* **3**, 1993. `ftp://ftp.inf.ethz.ch/doc/papers/ti/scs/unipoly.ps.gz`.

Maeder, R. E. `Polyhedra.m` and `PolyhedraExamples` *Mathematica®* notebooks. `http://www.inf.ethz.ch/department/TI/rm/programs.html`.

Maeder, R. E. "The Uniform Polyhedra." `http://www.inf.ethz.ch/department/TI/rm/unipoly/`.

Skilling, J. "The Complete Set of Uniform Polyhedron." *Phil. Trans. Roy. Soc. London, Ser. A* **278**, 111–136, 1975.

Virtual Image. "The Uniform Polyhedra CD-ROM." `http://ourworld.compuserve.com/homepages/vir_image/html/uniformpolyhedra.html`.

Wenninger, M. J. *Polyhedron Models.* New York: Cambridge University Press, pp. 1-10 and 98, 1989.

Zalgaller, V. *Convex Polyhedra with Regular Faces.* New York: Consultants Bureau, 1969.

Ziegler, G. M. *Lectures on Polytopes.* Berlin: Springer-Verlag, 1995.

Uniform Variate

A RANDOM NUMBER which lies within a specified range (which can, without loss of generality, be taken as [0, 1]), with a UNIFORM DISTRIBUTION.

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Uniform Deviates." §7.1 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 267–277, 1992.

Unimodal Distribution

A DISTRIBUTION such as the GAUSSIAN DISTRIBUTION which has a single "peak."

see also BIMODAL DISTRIBUTION

Unimodal Sequence

A finite SEQUENCE which first increases and then decreases. A SEQUENCE $\{s_1, s_2, \ldots, s_n\}$ is unimodal if there exists a t such that

$$s_1 \leq s_2 \leq \ldots \leq s_t$$

and

$$s_t \geq s_{t+1} \geq \ldots \geq s_n.$$

Unimodular Group

A group whose left HAAR MEASURE equals its right HAAR MEASURE.

see also HAAR MEASURE

References

Knapp, A. W. "Group Representations and Harmonic Analysis, Part II." *Not. Amer. Math. Soc.* **43**, 537–549, 1996.

Unimodular Matrix

A MATRIX A with INTEGER elements and DETERMINANT $\det(\mathsf{A}) = \pm 1$, also called a UNIT MATRIX.

The inverse of a unimodular matrix is another unimodular matrix. A POSITIVE unimodular matrix has $\det(\mathsf{A}) = +1$. The nth POWER of a POSITIVE UNIMODULAR MATRIX

$$\mathsf{M} \equiv \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \tag{1}$$

is

$$\mathsf{M}^n = \begin{bmatrix} m_{11}U_{n-1}(a) - U_{n-2}(a) & m_{12}U_{n-1}(a) \\ m_{21}U_{n-1}(a) & m_{22}U_{n-1}(a) - U_{n-2}(a) \end{bmatrix}, \tag{2}$$

where

$$a \equiv \tfrac{1}{2}(m_{11} + m_{22}) \tag{3}$$

and the U_n are CHEBYSHEV POLYNOMIALS OF THE SECOND KIND,

$$U_m(x) = \frac{\sin[(m+1)\cos^{-1} x]}{\sqrt{1-x^2}}. \tag{4}$$

see also CHEBYSHEV POLYNOMIAL OF THE SECOND KIND

References

Born, M. and Wolf, E. *Principles of Optics: Electromagnetic Theory of Propagation, Interference, and Diffraction of Light, 6th ed.* New York: Pergamon Press, p. 67, 1980.

Goldstein, H. *Classical Mechanics, 2nd ed.* Reading, MA: Addison-Wesley, p. 149, 1980.

Unimodular Transformation

A transformation $\mathbf{x}' = \mathsf{A}\mathbf{x}$ is unimodular if the DETERMINANT of the MATRIX A satisfies

$$\det(\mathsf{A}) = \pm 1.$$

A NECESSARY and SUFFICIENT condition that a linear transformation transform a lattice to itself is that the transformation be unimodular.

Union

The union of two sets A and B is the set obtained by combining the members of each. This is written $A \cup B$, and is pronounced "A union B" or "A cup B." The union of sets A_1 through A_n is written $\bigcup_{i=1}^n A_i$.

Let A, B, C, ... be sets, and let $P(S)$ denote the probability of S. Then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{1}$$

Similarly,

$$\begin{aligned} &P(A \cup B \cup C) = P[A \cup (B \cup C)] \\ &= P(A) + P(B \cup C) - P[A \cap (B \cup C)] \\ &= P(A) + [P(B) + P(C) - P(B \cap C)] \\ &\quad -P[(A \cap B) \cup (A \cap C)] \\ &= P(A) + P(B) + P(C) - P(B \cap C) \\ &\quad -\{P(A \cap B) + P(A \cap C) - P[(A \cap B) \cap (A \cap C)]\} \\ &= P(A) + P(B) + P(C) - P(A \cap B) \\ &\quad -P(A \cap C) - P(B \cap C) + P(A \cap B \cap C). \end{aligned} \tag{2}$$

If A and B are DISJOINT, by definition $P(A \cap B) = 0$, so

$$P(A \cup B) = P(A) + P(B). \tag{3}$$

Continuing, for a set of n disjoint elements E_1, E_2, ..., E_n

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i), \tag{4}$$

which is the COUNTABLE ADDITIVITY PROBABILITY AXIOM. Now let

$$E_i \equiv A \cap B_i, \tag{5}$$

then

$$P\left(\bigcup_{i=1}^{n} E \cap B_i\right) = \sum_{i=1}^{n} P(E \cap B_i). \tag{6}$$

see also INTERSECTION, OR

Uniplanar Double Point

see ISOLATED SINGULARITY

Unipotent

A p-ELEMENT x of a GROUP G is unipotent if $F^*(C_G(x))$ is a p-GROUP, where F^* is the generalized FITTING SUBGROUP.

see also FITTING SUBGROUP, p-ELEMENT, p-GROUP

Unique

The property of being the only possible solution (perhaps modulo a constant, class of transformation, etc.).

see also ALEKSANDROV'S UNIQUENESS THEOREM, EXISTENCE, MAY-THOMASON UNIQUENESS THEOREM

Unique Factorization Theorem

see FUNDAMENTAL THEOREM OF ARITHMETIC

Unit

A unit is an element in a RING that has a multiplicative inverse. If n is an ALGEBRAIC INTEGER which divides every ALGEBRAIC INTEGER in the FIELD, n is called a unit in that FIELD. A given FIELD may contain an infinity of units. The units of $\mathbb{Z}_n$ are the elements RELATIVELY PRIME to n. The units in $\mathbb{Z}_n$ which are SQUARES are called QUADRATIC RESIDUES.

see also EISENSTEIN UNIT, FUNDAMENTAL UNIT, PRIME UNIT, QUADRATIC RESIDUE

Unit Circle

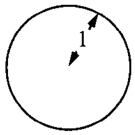

A CIRCLE of RADIUS 1, such as the one used to defined the functions of TRIGONOMETRY.

see also UNIT DISK, UNIT SQUARE

Unit Disk

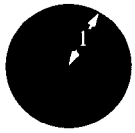

A DISK with RADIUS 1.

see also FIVE DISKS PROBLEM, UNIT CIRCLE, UNIT SQUARE

Unit Fraction

A unit fraction is a FRACTION with NUMERATOR 1, also known as an EGYPTIAN FRACTION. Any RATIONAL NUMBER has infinitely many unit fraction representations, although only finitely many have a given fixed number of terms. Each FRACTION x/y with y ODD has a unit fraction representation in which each DENOMINATOR is ODD (Breusch 1954; Guy 1994, p. 160). Every x/y has a t-term representation where $t = \mathcal{O}(\sqrt{\log y})$ (Vose 1985).

There are a number of ALGORITHMS (including the BINARY REMAINDER METHOD, CONTINUED FRACTION UNIT FRACTION ALGORITHM, GENERALIZED REMAINDER METHOD, GREEDY ALGORITHM, REVERSE GREEDY ALGORITHM, SMALL MULTIPLE METHOD, and SPLITTING ALGORITHM) for decomposing an arbitrary FRACTION into unit fractions.

see also CALCUS, HALF, QUARTER, SCRUPLE, UNCIA

References

Beck, A.; Bleicher, M. N.; and Crowe, D. W. *Excursions into Mathematics.* New York: Worth Publishers, 1970.

Beeckmans, L. "The Splitting Algorithm for Egyptian Fractions." *J. Number Th.* **43**, 173–185, 1993.

Bleicher, M. N. "A New Algorithm for the Expansion of Continued Fractions." *J. Number Th.* **4**, 342–382, 1972.

Breusch, R. "A Special Case of Egyptian Fractions." Solution to advanced problem 4512. *Amer. Math. Monthly* **61**, 200–201, 1954.

Brown, K. S. "Egyptian Unit Fractions." `http://www.seanet.com/~ksbrown`.

Eppstein, D. "Ten Algorithms for Egyptian Fractions." *Math. Edu. Res.* **4**, 5–15, 1995.

Eppstein, D. "Egyptian Fractions." `http://www.ics.uci.edu/~eppstein/numth/egypt/`.

Eppstein, D. `Egypt.ma` Mathematica notebook. `http://www.ics.uci.edu/~eppstein/numth/egypt/egypt.ma`.

Graham, R. "On Finite Sums of Unit Fractions." *Proc. London Math. Soc.* **14**, 193–207, 1964.

Guy, R. K. "Egyptian Fractions." §D11 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 87–93 and 158–166, 1994.

Klee, V. and Wagon, S. *Old and New Unsolved Problems in Plane Geometry and Number Theory.* Washington, DC: Math. Assoc. Amer., pp. 175–177 and 206–208, 1991.

Niven, I. and Zuckerman, H. S. *An Introduction to the Theory of Numbers, 5th ed.* New York: Wiley, p. 200, 1991.

Stewart, I. "The Riddle of the Vanishing Camel." *Sci. Amer.*, 122–124, June 1992.

Tenenbaum, G. and Yokota, H. "Length and Denominators of Egyptian Fractions." *J. Number Th.* **35**, 150–156, 1990.

Vose, M. "Egyptian Fractions." *Bull. London Math. Soc.* **17**, 21, 1985.

Wagon, S. "Egyptian Fractions." §8.6 in *Mathematica in Action.* New York: W. H. Freeman, pp. 271–277, 1991.

Unit Matrix

see UNIMODULAR MATRIX

Unit Point

The point in the PLANE with Cartesian coordinates (1, 1).

References

Woods, F. S. *Higher Geometry: An Introduction to Advanced Methods in Analytic Geometry.* New York: Dover, p. 9, 1961.

Unit Ring

A unit ring is a set together with two BINARY OPERATORS $S(+,*)$ satisfying the following conditions:

1. Additive associativity: For all $a, b, c \in S$, $(a+b)+c = a+(b+c)$,
2. Additive commutativity: For all $a, b \in S$, $a+b = b+a$,
3. Additive identity: There exists an element $0 \in S$ such that for all $a \in S: 0+a = a+0 = a$,
4. Additive inverse: For every $a \in S$, there exists a $-a \in S$ such that $a+(-a) = (-a)+a = 0$,
5. Multiplicative associativity: For all $a, b, c \in S$, $(a*b)*c = a*(b*c)$,
6. Multiplicative identity: There exists an element $1 \in S$ such that for all $a \in S$, $1*a = a*1 = a$,
7. Left and right distributivity: For all $a, b, c \in S$, $a*(b+c) = (a*b)+(a*c)$ and $(b+c)*a = (b*a)+(c*a)$.

Thus, a unit ring is a RING with a multiplicative identity.

see also BINARY OPERATOR, RING

References

Rosenfeld, A. *An Introduction to Algebraic Structures.* New York: Holden-Day, 1968.

Unit Sphere

A SPHERE of RADIUS 1.

see also SPHERE, UNIT CIRCLE

Unit Square

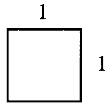

A SQUARE with side lengths 1. *The* unit square usually means the one with coordinates (0, 0), (1, 0), (1, 1), (0, 1) in the real plane, or 0, 1, $1+i$, and i in the COMPLEX PLANE.

see also HEILBRONN TRIANGLE PROBLEM, UNIT CIRCLE, UNIT DISK

Unit Step

see HEAVISIDE STEP FUNCTION

Unit Vector

A VECTOR of unit length. The unit vector $\hat{\mathbf{v}}$ having the same direction as a given (nonzero) vector $\mathbf{v}$ is defined by

$$\hat{\mathbf{v}} \equiv \frac{\mathbf{v}}{|\mathbf{v}|},$$

where $|\mathbf{v}|$ denotes the NORM of $\mathbf{v}$, is the unit vector in the same direction as the (finite) VECTOR $\mathbf{v}$. A unit VECTOR in the $\mathbf{x}_n$ direction is given by

$$\hat{\mathbf{x}}_n \equiv \frac{\frac{\partial \mathbf{r}}{\partial x_n}}{\left|\frac{\partial \mathbf{r}}{\partial x_n}\right|},$$

where $\mathbf{r}$ is the RADIUS VECTOR.

see also NORM, RADIUS VECTOR, VECTOR

Unital

A BLOCK DESIGN of the form $(q^3+1,\ q+1,\ 1)$.

References

Dinitz, J. H. and Stinson, D. R. "A Brief Introduction to Design Theory." Ch. 1 in *Contemporary Design Theory: A Collection of Surveys* (Ed. J. H. Dinitz and D. R. Stinson). New York: Wiley, pp. 1–12, 1992.

Unitary Aliquot Sequence

An ALIQUOT SEQUENCE computed using the analog of the RESTRICTED DIVISOR FUNCTION $s^*(n)$ in which only UNITARY DIVISORS are included.

see also ALIQUOT SEQUENCE, UNITARY SOCIABLE NUMBERS

References

Guy, R. K. "Unitary Aliquot Sequences." §B8 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 63–65, 1994.

Unitary Amicable Pair

A PAIR of numbers m and n such that

$$\sigma^*(m) = \sigma^*(n) = m + n,$$

where $\sigma^*(n)$ is the sum of UNITARY DIVISORS. Hagis (1971) and García (1987) give 82 such pairs. The first few are (114, 126), (1140, 1260), (18018, 22302), (32130, 40446), ... (Sloane's A002952 and A002953).

References

García, M. "New Unitary Amicable Couples." *J. Recr. Math.* **19**, 12–14, 1987.

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 57, 1994.

Hagis, P. "Relatively Prime Amicable Numbers of Opposite Parity." *Math. Comput.* **25**, 915–918, 1971.

Sloane, N. J. A. Sequences A002952/M5372 and A002953/M5389 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Unitary Divisor

A DIVISOR d of c for which

$$\mathrm{GCD}(d, c/d) = 1,$$

where GCD is the GREATEST COMMON DIVISOR.

see also DIVISOR, GREATEST COMMON DIVISOR, UNITARY PERFECT NUMBER

References

Guy, R. K. "Unitary Perfect Numbers." §B3 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 53–59, 1994.

Unitary Group

The unitary group $U_n(q)$ is the set of $n \times n$ UNITARY MATRICES.

see also LIE-TYPE GROUP, UNITARY MATRIX

References

Wilson, R. A. "ATLAS of Finite Group Representation." `http://for.mat.bham.ac.uk/atlas#unit`.

Unitary Matrix

A unitary matrix is a MATRIX U for which

$$\mathsf{U}^\dagger = \mathsf{U}^{-1}, \tag{1}$$

where † denotes the ADJOINT OPERATOR. This guarantees that

$$\mathsf{U}^\dagger\mathsf{U} = 1. \tag{2}$$

Unitary matrices leave the length of a COMPLEX vector unchanged. The product of two unitary matrices is itself unitary. If U is unitary, then so is U^{-1}. A SIMILARITY TRANSFORMATION of a HERMITIAN MATRIX with a unitary matrix gives

$$\begin{aligned}(uau^{-1})^\dagger &= [(ua)(u^{-1})]^\dagger = (u^{-1})^\dagger(ua)^\dagger = (u^\dagger)^\dagger(a^\dagger u^\dagger)\\ &= uau^\dagger = uau^{-1}.\end{aligned} \tag{3}$$

For REAL MATRICES, HERMITIAN is the same as ORTHOGONAL. Unitary matrices are NORMAL MATRICES.

If M is a unitary matrix, then the PERMANENT

$$|\operatorname{perm}(\mathsf{M})| \leq 1 \tag{4}$$

(Minc 1978, p. 25, Vardi 1991).

see also ADJOINT OPERATOR, HERMITIAN MATRIX, NORMAL MATRIX, ORTHOGONAL MATRIX, PERMANENT

References

Arfken, G. "Hermitian Matrices, Unitary Matrices." §4.5 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 209–217, 1985.

Minc, H. *Permanents.* Reading, MA: Addison-Wesley, 1978.

Vardi, I. "Permanents." §6.1 in *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 108 and 110–112, 1991.

Unitary Multiperfect Number

A number n which is an INTEGER multiple k of the SUM of its UNITARY DIVISORS $\sigma^*(n)$ is called a unitary k-multiperfect number. There are no ODD unitary multiperfect numbers.

References

Guy, R. K. "Unitary Perfect Numbers." §B3 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 53–59, 1994.

Unitary Multiplicative Character

A MULTIPLICATIVE CHARACTER is called unitary if it has ABSOLUTE VALUE 1 everywhere.

see also CHARACTER (MULTIPLICATIVE)

Unitary Perfect Number

A number n which is the sum of its UNITARY DIVISORS with the exception of n itself. There are no ODD unitary perfect numbers, and it has been conjectured that there are only a FINITE number of EVEN ones. The first few are 6, 60, 90, 87360, 146361946186458562560000, ... (Sloane's A002827).

References

Guy, R. K. "Unitary Perfect Numbers." §B3 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 53–59, 1994.

Sloane, N. J. A. Sequence A002827/M4268 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Wall, C. R. "On the Largest Odd Component of a Unitary Perfect Number." *Fib. Quart.* **25**, 312–316, 1987.

Unitary Sociable Numbers

SOCIABLE NUMBERS computed using the analog of the RESTRICTED DIVISOR FUNCTION $s^*(n)$ in which only UNITARY DIVISORS are included.

see also SOCIABLE NUMBERS

References

Guy, R. K. "Unitary Aliquot Sequences." §B8 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 63–65, 1994.

Unitary Transformation

A transformation of the form

$$\mathsf{A}' = \mathsf{UAU}^\dagger,$$

where † denotes the ADJOINT OPERATOR.

see also ADJOINT OPERATOR, TRANSFORMATION

Unitary Unimodular Group

see SPECIAL UNITARY GROUP

Unity

The number 1. There are n nth ROOTS OF UNITY, known as the DE MOIVRE NUMBERS.

see also 1, PRIMITIVE ROOT OF UNITY

Univalent Function

A function or transformation f in which $f(z)$ does not overlap z.

Univariate Function

A FUNCTION of a single variable (e.g., $f(x)$, $g(z)$, $\theta(\xi)$, etc.).

see also MULTIVARIATE FUNCTION

Univariate Polynomial

A POLYNOMIAL in a single variable. In common usage, univariate POLYNOMIALS are sometimes simply called "POLYNOMIALS."

see also POLYNOMIAL

Universal Graph

see COMPLETE GRAPH

Universal Statement

A universal statement S is a FORMULA whose FREE variables are all in the scope of universal quantifiers.

Universal Turing Machine

A TURING MACHINE which, by appropriate programming using a finite length of input tape, can act as *any* TURING MACHINE whatsoever.

see CHAITIN'S CONSTANT, HALTING PROBLEM, TURING MACHINE

References

Penrose, R. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics.* Oxford: Oxford University Press, pp. 51–57, 1989.

Unknot

A closed loop which is not KNOTTED. In the 1930s, by making use of REIDEMEISTER MOVES, Reidemeister first proved that KNOTS exist which are distinct from the unknot. He proved this by COLORING each part of a knot diagram with one of three colors.

The KNOT SUM of two unknots is another unknot.

The JONES POLYNOMIAL of the unknot is defined to give the normalization

$$V(t) = 1.$$

Haken (1961) devised an ALGORITHM to tell if a knot projection is the unknot. The ALGORITHM is so complicated, however, that it has never been implemented. Although it is not immediately obvious, the unknot is a PRIME KNOT.

see also COLORABLE, KNOT, KNOT THEORY, LINK, REIDEMEISTER MOVES, UNKNOTTING NUMBER

References

Haken, W. "Theorie der Normalflachen." *Acta Math.* **105**, 245–375, 1961.

Unknotting Number

The smallest number of times a KNOT must be passed through itself to untie it. Lower bounds can be computed using relatively straightforward techniques, but it is in general difficult to determine exact values. Many unknotting numbers can be determined from a knot's SIGNATURE. A KNOT with unknotting number 1 is a PRIME KNOT (Scharlemann 1985). It is not always true that the unknotting number is achieved in a projection with the minimal number of crossings.

The following table is from Kirby (1997, pp. 88–89), with the values for 10_{139} and 10_{152} taken from Kawamura. The unknotting numbers for 10_{154} and 10_{161} can be found using MENASCO'S THEOREM (Stoimenow 1998).

3_1	1	8_9	1	9_{10}	2 or 3	9_{32}	1 or 2
4_1	1	8_{10}	1 or 2	9_{11}	2	9_{33}	1
5_1	2	8_{11}	1	9_{12}	1	9_{34}	1
5_2	1	8_{12}	2	9_{13}	2 or 3	9_{35}	2 or 3
6_1	1	8_{13}	1	9_{14}	1	9_{36}	2
6_2	1	8_{14}	1	9_{15}	2	9_{37}	2
6_3	1	8_{15}	2	9_{16}	3	9_{38}	2 or 3
7_1	3	8_{16}	2	9_{17}	2	9_{39}	1
7_2	1	8_{17}	1	9_{18}	2	9_{40}	2
7_3	2	8_{18}	2	9_{19}	1	9_{41}	2
7_4	2	8_{19}	3	9_{20}	2	9_{42}	1
7_5	2	8_{20}	1	9_{21}	1	9_{43}	2
7_6	1	8_{21}	1	9_{22}	1	9_{44}	1
7_7	1	9_1	4	9_{23}	2	9_{45}	1
8_1	1	9_2	1	9_{24}	1	9_{46}	2
8_2	2	9_3	3	9_{25}	2	9_{47}	2
8_3	2	9_4	2	9_{26}	1	9_{48}	2
8_4	2	9_5	2	9_{27}	1	9_{49}	2 or 3
8_5	2	9_6	3	9_{28}	1	10_{139}	4
8_6	2	9_7	2	9_{29}	1	10_{152}	4
8_7	1	9_8	2	9_{30}	1	10_{154}	3
8_8	2	9_9	3	9_{31}	2	10_{161}	3

see also BENNEQUIN'S CONJECTURE, MENASCO'S THEOREM, MILNOR'S CONJECTURE, SIGNATURE (KNOT)

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 57–64, 1994.

Cipra, B. "From Knot to Unknot." *What's Happening in the Mathematical Sciences, Vol. 2.* Providence, RI: Amer. Math. Soc., pp. 8–13, 1994.

Kawamura, T. "The Unknotting Numbers of 10_{139} and 10_{152} are 4." To appear in *Osaka J. Math.* `http://ms421sun.ms.u-tokyo.ac.jp/~kawamura/worke.html`.

Kirby, R. (Ed.) "Problems in Low-Dimensional Topology." *AMS/IP Stud. Adv. Math., 2.2, Geometric Topology (Athens, GA, 1993).* Providence, RI: Amer. Math. Soc., pp. 35–473, 1997.

Scharlemann, M. "Unknotting Number One Knots are Prime." *Invent. Math.* **82**, 37–55, 1985.

Stoimenow, A. "Positive Knots, Closed Braids and the Jones Polynomial." Rev. May, 1997. `http://www.informatik.hu-berlin.de/~stoimeno/pos.ps.gz`.

Weisstein, E. W. "Knots and Links." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Knots.m`.

Unless

If A is true unless B, then not-B implies A, but B does not necessarily imply not-A.

see also PRECISELY UNLESS

Unlesss

see PRECISELY UNLESS

Unmixed

A homogeneous IDEAL defining a projective ALGEBRAIC VARIETY is unmixed if it has no embedded PRIME divisors.

Unpoke Move

see POKE MOVE

Unsafe

A position in a GAME is unsafe if the person who plays next can win. Every unsafe position can be made SAFE by at least one move.

see also GAME, SAFE

Unsolved Problem

see PROBLEM

Unstable Improper Node

A FIXED POINT for which the STABILITY MATRIX has equal POSITIVE EIGENVALUES.

see also ELLIPTIC FIXED POINT (DIFFERENTIAL EQUATIONS), FIXED POINT, HYPERBOLIC FIXED POINT (DIFFERENTIAL EQUATIONS), STABLE IMPROPER NODE, STABLE NODE, STABLE SPIRAL POINT, UNSTABLE NODE, UNSTABLE SPIRAL POINT, UNSTABLE STAR

References

Tabor, M. "Classification of Fixed Points." §1.4.b in *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, pp. 22–25, 1989.

Unstable Node

A FIXED POINT for which the STABILITY MATRIX has both EIGENVALUES POSITIVE, so $\lambda_1 > \lambda_2 > 0$.

see also ELLIPTIC FIXED POINT (DIFFERENTIAL EQUATIONS), FIXED POINT, HYPERBOLIC FIXED POINT (DIFFERENTIAL EQUATIONS), STABLE IMPROPER NODE, STABLE NODE, STABLE SPIRAL POINT, STABLE STAR, UNSTABLE IMPROPER NODE, UNSTABLE SPIRAL POINT, UNSTABLE STAR

References

Tabor, M. "Classification of Fixed Points." §1.4.b in *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, pp. 22–25, 1989.

Unstable Spiral Point

A FIXED POINT for which the STABILITY MATRIX has EIGENVALUES of the form $\lambda_\pm = \alpha \pm i\beta$ (with $\alpha, \beta > 0$).

see also ELLIPTIC FIXED POINT (DIFFERENTIAL EQUATIONS), FIXED POINT, HYPERBOLIC FIXED POINT (DIFFERENTIAL EQUATIONS), STABLE IMPROPER NODE, STABLE NODE, STABLE SPIRAL POINT, STABLE STAR, UNSTABLE IMPROPER NODE, UNSTABLE NODE, UNSTABLE STAR

References

Tabor, M. "Classification of Fixed Points." §1.4.b in *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, pp. 22–25, 1989.

Unstable Star

A FIXED POINT for which the STABILITY MATRIX has one zero EIGENVECTOR with POSITIVE EIGENVALUE $\lambda > 0$.

see also ELLIPTIC FIXED POINT (DIFFERENTIAL EQUATIONS), FIXED POINT, HYPERBOLIC FIXED POINT (DIFFERENTIAL EQUATIONS), STABLE IMPROPER NODE, STABLE NODE, STABLE SPIRAL POINT, STABLE STAR, UNSTABLE IMPROPER NODE, UNSTABLE NODE, UNSTABLE SPIRAL POINT

References

Tabor, M. "Classification of Fixed Points." §1.4.b in *Chaos and Integrability in Nonlinear Dynamics: An Introduction.* New York: Wiley, pp. 22–25, 1989.

Untouchable Number

An untouchable number is an INTEGER which is not the sum of the PROPER DIVISORS of any other number. The first few are 2, 5, 52, 88, 96, 120, 124, 146, ... (Sloane's A005114). Erdős has proven that there are infinitely many. It is thought that 5 is the only ODD untouchable number.

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 840, 1972.

Guy, R. K. "Untouchable Numbers." §B10 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 66–67, 1994.

Sloane, N. J. A. Sequence A005114/M1552 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Upper Bound

see LEAST UPPER BOUND

Upper Integral

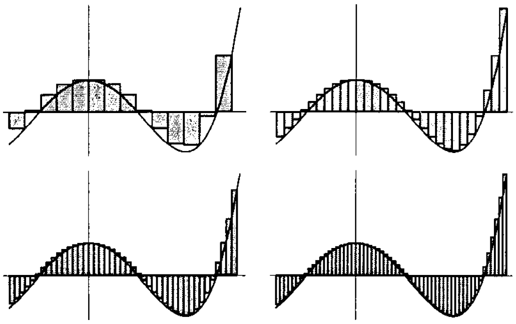

The limit of an UPPER SUM, when it exists, as the MESH SIZE approaches 0.

see also LOWER INTEGRAL, RIEMANN INTEGRAL, UPPER SUM

Upper Limit

Let the greatest term H of a SEQUENCE be a term which is greater than all but a finite number of the terms which are equal to H. Then H is called the upper limit of the SEQUENCE.

An upper limit of a SERIES

$$\text{upper} \lim_{n\to\infty} S_n = \overline{\lim_{n\to\infty}} S_n = k$$

is said to exist if, for every $\epsilon > 0$, $|S_n - k| < \epsilon$ for infinitely many values of n and if no number larger than k has this property.

see also LIMIT, LOWER LIMIT

References

Bromwich, T. J. I'a and MacRobert, T. M. "Upper and Lower Limits of a Sequence." §5.1 in *An Introduction to the Theory of Infinite Series, 3rd ed.* New York: Chelsea, p. 40, 1991.

Upper Sum

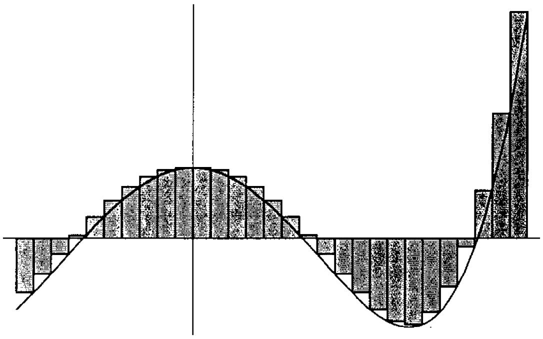

For a given function $f(x)$ over a partition of a given interval, the upper sum is the sum of box areas $f(x_k^*)\Delta x_k$ using the greatest value of the function $f(x_k^*)$ in each subinterval Δx_k.

see also LOWER SUM, RIEMANN INTEGRAL, UPPER INTEGRAL

Upper-Trimmed Subsequence

The upper-trimmed subsequence of $x = \{x_n\}$ is the sequence $\lambda(x)$ obtained by dropping the first occurrence of n for each n. If x is a FRACTAL SEQUENCE, then $\lambda(x) = x$.

see also LOWER-TRIMMED SUBSEQUENCE

References

Kimberling, C. "Fractal Sequences and Interspersions." *Ars Combin.* **45**, 157–168, 1997.

Upward Drawing

see HASSE DIAGRAM

Urchin

Kepler's original name for the SMALL STELLATED DODECAHEDRON.

Utility Graph

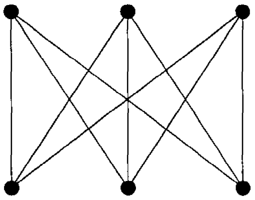

The utility problem asks, "Can a PLANAR GRAPH be constructed from each of three nodes ('house owners') to each of three other nodes ('wells')?" The answer is no, and the proof can be effected using the JORDAN CURVE THEOREM, while a more general result encompassing this one is the KURATOWSKI REDUCTION THEOREM. The utility graph UG is the graph showing the relationships described above. It is identical to the THOMSEN GRAPH and, in the more formal parlance of GRAPH THEORY, is known as the COMPLETE BIPARTITE GRAPH $K_{3,3}$.

see also COMPLETE BIPARTITE GRAPH, KURATOWSKI REDUCTION THEOREM, PLANAR GRAPH, THOMSEN GRAPH

References

Chartrand, G. "The Three Houses and Three Utilities Problem: An Introduction to Planar Graphs." §9.1 in *Introductory Graph Theory.* New York: Dover, pp. 191–202, 1985.

Ore, Ø. *Graphs and Their Uses.* New York: Random House, pp. 14–17, 1963.

Pappas, T. "Wood, Water, Grain Problem." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 175 and 233, 1989.

Utility Problem

see UTILITY GRAPH

V

Valence

see VALENCY

Valency

The number of EDGES at a GRAPH VERTEX.

Valuation

A generalization of the p-ADIC NUMBERS first proposed by Kürschák in 1913. A valuation $|\cdot|$ on a FIELD K is a FUNCTION from K to the REAL NUMBERS $\mathbb{R}$ such that the following properties hold for all $x, y \in K$:

1. $|x| \geq 0$,
2. $|x| = 0$ IFF $x = 0$,
3. $|xy| = |x|\,|y|$,
4. $|x| \leq 1$ IMPLIES $|1+x| \leq C$ for some constant $C \geq 1$ (independent of x).

If (4) is satisfied for $C = 2$, then $|\cdot|$ satisfies the TRIANGLE INEQUALITY,

4a. $|x+y| \leq |x| + |y|$ for all $x, y \in K$.

If (4) is satisfied for $C = 1$ then $|\cdot|$ satisfies the stronger TRIANGLE INEQUALITY

4b. $|x+y| \leq \max(|x|, |y|)$.

The simplest valuation is the ABSOLUTE VALUE for REAL NUMBERS. A valuation satisfying (4b) is called non-ARCHIMEDEAN VALUATION; otherwise, it is called ARCHIMEDEAN.

If $|\cdot|_1$ is a valuation on K and $\lambda \geq 1$, then we can define a new valuation $|\cdot|_2$ by

$$|x|_2 = |x|_1^\lambda. \qquad (1)$$

This does indeed give a valuation, but possibly with a different constant C in AXIOM 4. If two valuations are related in this way, they are said to be equivalent, and this gives an equivalence relation on the collection of all valuations on K. Any valuation is equivalent to one which satisfies the triangle inequality (4a). In view of this, we need only to study valuations satisfying (4a), and we often view axioms (4) and (4a) as interchangeable (although this is not strictly true).

If two valuations are equivalent, then they are both non-ARCHIMEDEAN or both ARCHIMEDEAN. $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ with the usual Euclidean norms are Archimedean valuated fields. For any PRIME p, the p-ADIC NUMBERS $\mathbb{Q}_p$ with the p-adic valuation $|\cdot|_p$ is a non-Archimedean valuated field.

If K is any FIELD, we can define the trivial valuation on K by $|x| = 1$ for all $x \neq 0$ and $|0| = 0$, which is a non-Archimedean valuation. If K is a FINITE FIELD, then the only possible valuation over K is the trivial one. It can be shown that any valuation on $\mathbb{Q}$ is equivalent to one of the following: the trivial valuation, Euclidean absolute norm $|\cdot|$, or p-adic valuation $|\cdot|_p$.

The equivalence of any nontrivial valuation of $\mathbb{Q}$ to either the usual ABSOLUTE VALUE or to a p-ADIC NUMBER absolute value was proved by Ostrowski (1935). Equivalent valuations give rise to the same topology. Conversely, if two valuations have the same topology, then they are equivalent. A stronger result is the following: Let $|\cdot|_1$, $|\cdot|_2$, ..., $|\cdot|_k$ be valuations over K which are pairwise inequivalent and let a_1, a_2, ..., a_k be elements of K. Then there exists an infinite sequence $(x_1, x_2, \ldots)$ of elements of K such that

$$\lim_{n\to\infty \text{ w.r.t. } |\cdot|_1} x_n = a_1 \qquad (2)$$

$$\lim_{n\to\infty \text{ w.r.t. } |\cdot|_2} x_n = a_2, \qquad (3)$$

etc. This says that inequivalent valuations are, in some sense, completely independent of each other. For example, consider the rationals $\mathbb{Q}$ with the 3-adic and 5-adic valuations $|\cdot|_3$ and $|\cdot|_5$, and consider the sequence of numbers given by

$$x_n = \frac{43 \cdot 5^n + 92 \cdot 3^n}{3^n + 5^n}. \qquad (4)$$

Then $x_n \to 43$ as $n \to \infty$ with respect to $|\cdot|_3$, but $x_n \to 92$ as $n \to \infty$ with respect to $|\cdot|_5$, illustrating that a sequence of numbers can tend to two different limits under two different valuations.

A discrete valuation is a valuation for which the VALUATION GROUP is a discrete subset of the REAL NUMBERS $\mathbb{R}$. Equivalently, a valuation (on a FIELD K) is discrete if there exists a REAL NUMBER $\epsilon > 0$ such that

$$|x| \in (1-\epsilon, 1+\epsilon) \Rightarrow |x| = 1 \text{ for all } x \in K. \qquad (5)$$

The p-adic valuation on $\mathbb{Q}$ is discrete, but the ordinary absolute valuation is not.

If $|\cdot|$ is a valuation on K, then it induces a metric

$$d(x,y) = |x-y| \qquad (6)$$

on K, which in turn induces a TOPOLOGY on K. If $|\cdot|$ satisfies (4b) then the metric is an ULTRAMETRIC. We say that $(K, |\cdot|)$ is a complete valuated field if the METRIC SPACE is complete.

see also ABSOLUTE VALUE, LOCAL FIELD, METRIC SPACE, p-ADIC NUMBER, STRASSMAN'S THEOREM, ULTRAMETRIC, VALUATION GROUP

References

Cassels, J. W. S. *Local Fields*. Cambridge, England: Cambridge University Press, 1986.

Ostrowski, A. "Untersuchungen zur aritmetischen Theorie der Körper." *Math. Zeit.* **39**, 269–404, 1935.

Valuation Group

Let $(K, |\cdot|)$ be a valuated field. The valuation group G is defined to be the set

$$G = \{|x| : x \in K, x \neq 0\},$$

with the group operation being multiplication. It is a SUBGROUP of the POSITIVE REAL NUMBERS, under multiplication.

Valuation Ring

Let $(K, |\cdot|)$ be a non-Archimedean valuated field. Its valuation ring R is defined to be

$$R = \{x \in K : |x| \leq 1\}.$$

The valuation ring has maximal IDEAL

$$M = \{x \in K : |x| < 1\},$$

and the field R/M is called the residue field, class field, or field of digits. For example, if $K = \mathbb{Q}_p$ (p-adic numbers), then $R = Z_p$ (p-adic integers), $M = pZ_p$ (p-adic integers congruent to 0 mod p), and $R/M = \mathrm{GF}(p)$, the FINITE FIELD of order p.

Valuation Theory

The study of VALUATIONS which simplifies class field theory and the theory of algebraic function fields.

see also VALUATION

References

Iyanaga, S. and Kawada, Y. (Eds.). "Valuations." §425 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 1350–1353, 1980.

Value

The quantity which a FUNCTION f takes upon application to a given quantity.

see also VALUE (GAME)

Value (Game)

The solution to a GAME in GAME THEORY. When a SADDLE POINT is present

$$\min_{i\leq m} \min_{j\leq n} a_{ij} = \min_{j\leq n} \max_{i\leq m} a_{ij} \equiv v,$$

and v is the value for pure strategies.

see also ABSOLUTE VALUE, GAME THEORY, MINIMAX THEOREM, VALUATION

Vampire Number

A number $v = xy$ with an EVEN number n of DIGITS formed by multiplying a pair of $n/2$-DIGIT numbers (where the DIGITS are taken from the original number in any order) x and y together. Pairs of trailing zeros are not allowed. If v is a vampire number, then x and y are called its "fangs." Examples of vampire numbers include

$$\begin{aligned}1260 &= 21 \times 60\\ 1395 &= 15 \times 93\\ 1435 &= 35 \times 41\\ 1530 &= 30 \times 51\\ 1827 &= 21 \times 87\\ 2187 &= 27 \times 81\\ 6880 &= 80 \times 86\end{aligned}$$

(Sloane's A014575). There are seven 4-digit vampires, 155 6-digit vampires, and 3382 8-digit vampires. General formulas can be constructed for special classes of vampires, such as the fangs

$$\begin{aligned}x &= 25 \cdot 10^k + 1\\ y &= 100(10^{k+1} + 52)/25,\end{aligned}$$

giving the vampire

$$\begin{aligned}v = xy &= (10^{k+1} + 52)10^{k+2} + 100(10^{k+1} + 52)/25\\ &= x^* \cdot 10^{k+2} + t\\ &= 8(26 + 5 \cdot 10^k)(1 + 25 \cdot 10^k),\end{aligned}$$

where x^* denotes x with the DIGITS reversed (Roushe and Rogers).

Pickover (1995) also defines pseudovampire numbers, in which the multiplicands have different number of digits.

References

Pickover, C. A. "Vampire Numbers." Ch. 30 in *Keys to Infinity.* New York: W. H. Freeman, pp. 227–231, 1995.

Pickover, C. A. "Vampire Numbers." *Theta* **9**, 11–13, Spring 1995.

Pickover, C. A. "Interview with a Number." *Discover* **16**, 136, June 1995.

Roushe, F. W. and Rogers, D. G. "Tame Vampires." Undated manuscript.

Sloane, N. J. A. Sequence A014575 in "An On-Line Version of the Encyclopedia of Integer Sequences."

van der Grinten Projection

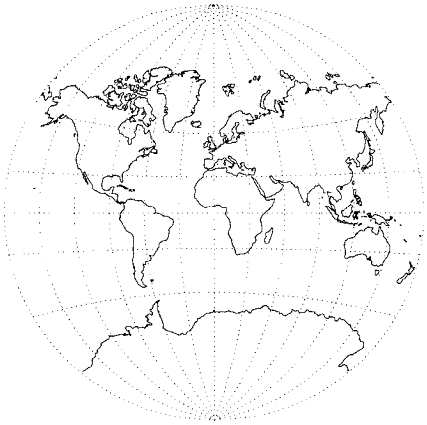

A MAP PROJECTION given by the transformation

$$x = \operatorname{sgn}(\lambda - \lambda_0)$$
$$\times \frac{\pi|A(G-P^2) - \sqrt{A^2(G-P^2)^2 - (P^2+A^2)(G^2-P^2)}|}{P^2+A^2} \quad (1)$$

$$y = \operatorname{sgn}(\phi) \frac{\pi|PQ - A\sqrt{(A^2+1)(P^2+A^2) - Q^2}}{P^2+A^2}, \quad (2)$$

where

$$A = \frac{1}{2}\left|\frac{\pi}{\lambda-\lambda_0} - \frac{\lambda-\lambda_0}{\pi}\right| \quad (3)$$

$$G = \frac{\cos\theta}{\sin\theta + \cos\theta - 1} \quad (4)$$

$$P = G\left(\frac{2}{\sin\theta} - 1\right) \quad (5)$$

$$\theta = \sin^{-1}\left|\frac{2\phi}{\pi}\right| \quad (6)$$

$$Q = A^2 + G. \quad (7)$$

The inverse FORMULAS are

$$\phi = \operatorname{sgn}(y)\pi\left[-m_1\cos(\theta_1 + \tfrac{1}{3}\pi) - \frac{c_2}{3c_3}\right] \quad (8)$$

$$\lambda = \frac{\pi|X^2+Y^2-1+\sqrt{1+2(X^2-Y^2)+(X^2+Y^2)^2}|}{2X} + \lambda_0, \quad (9)$$

where

$$X = \frac{x}{\pi} \quad (10)$$

$$Y = \frac{y}{\pi} \quad (11)$$

$$c_1 = -|Y|(1+X^2+Y^2) \quad (12)$$

$$c_2 = c_1 - 2Y^2 + X^2 \quad (13)$$

$$c_3 = -2c_1 + 1 + 2Y^2 + (X^2+Y^2)^2 \quad (14)$$

$$d = \frac{Y^2}{c_3} + \frac{1}{27}\left(\frac{2{c_2}^3}{{c_3}^3} - \frac{9c_1c_2}{{c_3}^2}\right) \quad (15)$$

$$a_1 = \frac{1}{c_3}\left(c_1 - \frac{{c_2}^2}{3c_3}\right) \quad (16)$$

$$m_1 = 2\sqrt{-\tfrac{1}{3}a_1} \quad (17)$$

$$\theta_1 = \tfrac{1}{3}\cos^{-1}\left(\frac{3d}{a_1m_1}\right). \quad (18)$$

References

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 239–242, 1987.

van der Pol Equation

An ORDINARY DIFFERENTIAL EQUATION which can be derived from the RAYLEIGH DIFFERENTIAL EQUATION by differentiating and setting $y = y'$. It is an equation describing self-sustaining oscillations in which energy is fed into small oscillations and removed from large oscillations. This equation arises in the study of circuits containing vacuum tubes and is given by

$$y'' - \mu(1-y^2)y' + y = 0.$$

see also RAYLEIGH DIFFERENTIAL EQUATION

References

Kreyszig, E. *Advanced Engineering Mathematics, 6th ed.* New York: Wiley, pp. 165–166, 1988.

van der Waerden Number

The threshold numbers proven to exist by VAN DER WAERDEN'S THEOREM. The first few are 1, 3, 9, 35, 178, ... (Sloane's A005346).

References

Goodman, J. E. and O'Rourke, J. (Eds.). *Handbook of Discrete & Computational Geometry.* Boca Raton, FL: CRC Press, p. 159, 1997.

Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., p. 29, 1991.

Sloane, N. J. A. Sequence A005346/M2819 in "An On-Line Version of the Encyclopedia of Integer Sequences."

van der Waerden's Theorem

For any given POSITIVE INTEGERS k and r, there exists a threshold number $n(k,r)$ (known as a VAN DER WAERDEN NUMBER) such that no matter how the numbers 1, 2, ..., n are partitioned into k classes, at least one of the classes contains an ARITHMETIC PROGRESSION of length at least r. However, no FORMULA for $n(k,r)$ is known.

see also ARITHMETIC PROGRESSION

References

Honsberger, R. *More Mathematical Morsels.* Washington, DC: Math. Assoc. Amer., p. 29, 1991.

Khinchin, A. Y. "Van der Waerden's Theorem on Arithmetic Progressions." Ch. 1 in *Three Pearls of Number Theory.* New York: Dover, pp. 11–17, 1998.

van der Waerden, B. L. "Beweis einer Baudetschen Vermutung." *Nieuw Arch. Wiskunde* **15**, 212–216, 1927.

van Kampen's Theorem

In the usual diagram of inclusion homeomorphisms, if the upper two maps are injective, then so are the other two.

References

Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, pp. 74–75 and 369–373, 1976.

van Wijngaarden-Deker-Brent Method

see BRENT'S METHOD

Vandermonde Determinant

$$\begin{aligned}\Delta(x_1,\ldots,x_n) &\equiv \begin{vmatrix} 1 & x_1 & {x_1}^2 & \cdots & {x_1}^{n-1} \\ 1 & x_2 & {x_2}^2 & \cdots & {x_2}^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & {x_n}^2 & \cdots & {x_n}^{n-1} \end{vmatrix} \\ &= \prod_{\substack{i,j \\ i>j}} (x_i - x_j)\end{aligned}$$

(Sharpe 1987). For INTEGERS $a_1, \ldots, a_n$, $\Delta(a_1, \ldots, a_n)$ is divisible by $\prod_{i=1}^{n}(i-1)!$ (Chapman 1996).

see also VANDERMONDE MATRIX

References

Chapman, R. "A Polynomial Taking Integer Values." *Math. Mag.* **69**, 121, 1996.

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1111, 1979.

Sharpe, D. §2.9 in *Rings and Factorization.* Cambridge, England: Cambridge University Press, 1987.

Vandermonde Identity

see CHU-VANDERMONDE IDENTITY

Vandermonde Matrix

A type of matrix which arises in the LEAST SQUARES FITTING of POLYNOMIALS and the reconstruction of a DISTRIBUTION from the distribution's MOMENTS. The solution of an $n \times n$ Vandermonde matrix equation requires $\mathcal{O}(n^2)$ operations. A Vandermonde matrix of order n is of the form

$$\begin{bmatrix} 1 & x_1 & {x_1}^2 & \cdots & {x_1}^{n-1} \\ 1 & x_2 & {x_2}^2 & \cdots & {x_2}^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & {x_n}^2 & \cdots & {x_n}^{n-1} \end{bmatrix}.$$

see also TOEPLITZ MATRIX, TRIDIAGONAL MATRIX, VANDERMONDE DETERMINANT

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Vandermonde Matrices and Toeplitz Matrices." §2.8 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 82–89, 1992.

Vandermonde's Sum

see CHU-VANDERMONDE IDENTITY

Vandermonde Theorem

A special case of GAUSS'S THEOREM with a a NEGATIVE INTEGER $-n$:

$${}_2F_1(-n, b; c; 1) = \frac{(c-b)_n}{(c)_n},$$

where ${}_2F_1(a, b; c; z)$ is a HYPERGEOMETRIC FUNCTION and $(a)_n$ is a POCHHAMMER SYMBOL (Bailey 1935, p. 3).

see also GAUSS'S THEOREM

References

Bailey, W. N. *Generalised Hypergeometric Series.* Cambridge, England: Cambridge University Press, 1935.

Vandiver's Criteria

Let p be a IRREGULAR PRIME, and let $P = rp + 1$ be a PRIME with $P < p^2 - p$. Also let t be an INTEGER such that $t^3 \not\equiv 1 \pmod{P}$. For an IRREGULAR PAIR $(p, 2k)$, form the product

$$Q_{2k} = t^{-rd/2} \prod_{b=1}^{m} (t^{rb} - 1)^{b^{p-1-2k}},$$

where

$$\begin{aligned} m &= \tfrac{1}{2}(pl - 1) \\ d &= \sum_{n=1}^{m} n^{p-2k}. \end{aligned}$$

If ${Q_{2k}}^r \not\equiv 1 \pmod{P}$ for all such IRREGULAR PAIRS, then FERMAT'S LAST THEOREM holds for exponent p.

see also FERMAT'S LAST THEOREM, IRREGULAR PAIR, IRREGULAR PRIME

References

Johnson, W. "Irregular Primes and Cyclotomic Invariants." *Math. Comput.* **29**, 113–120, 1975.

Vanishing Point

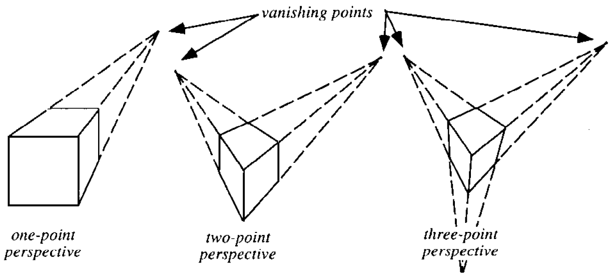

The point or points to which the extensions of PARALLEL lines appear to converge in a PERSPECTIVE drawing.

see also PERSPECTIVE, PROJECTIVE GEOMETRY

References

Dixon, R. "Perspective Drawings." Ch. 3 in *Mathographics.* New York: Dover, pp. 79–88, 1991.

Varga's Constant

$$V \equiv \frac{1}{\Lambda} = 9.2890254919\ldots,$$

where Λ is the ONE-NINTH CONSTANT.

see also ONE-NINTH CONSTANT

Variance

For N samples of a variate having a distribution with *known* MEAN μ, the "population variance" (usually called "variance" for short, although the word "population" should be added when needed to distinguish it from the SAMPLE VARIANCE) is defined by

$$\begin{aligned} \operatorname{var}(x) &\equiv \frac{1}{N}\sum (x-\mu)^2 = \left\langle x^2 - 2\mu x + \mu^2 \right\rangle \\ &= \left\langle x^2 \right\rangle - \langle 2\mu x \rangle + \left\langle \mu^2 \right\rangle \\ &= \left\langle x^2 \right\rangle - 2\mu \langle x \rangle + \mu^2, \end{aligned} \tag{1}$$

where

$$\langle x \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{2}$$

But since $\langle x \rangle$ is an UNBIASED ESTIMATOR for the MEAN

$$\mu \equiv \langle x \rangle, \tag{3}$$

it follows that the variance

$$\sigma^2 \equiv \operatorname{var}(x) = \left\langle x^2 \right\rangle - \mu^2. \tag{4}$$

The population STANDARD DEVIATION is then defined as

$$\sigma \equiv \sqrt{\operatorname{var}(x)} = \sqrt{\langle x^2 \rangle - \mu^2}. \tag{5}$$

A useful identity involving the variance is

$$\operatorname{var}(f(x)+g(x)) = \operatorname{var}(f(x)) + \operatorname{var}(g(x)). \tag{6}$$

Therefore,

$$\begin{aligned} \operatorname{var}(ax+b) &= \left\langle [(ax+b) - \langle ax+b \rangle]^2 \right\rangle \\ &= \left\langle (ax+b-a\langle x \rangle - b)^2 \right\rangle \\ &= \left\langle (ax-a\mu)^2 \right\rangle = \left\langle a^2(x-\mu)^2 \right\rangle \\ &= a^2 \left\langle (x-\mu)^2 \right\rangle = a^2 \operatorname{var}(x) \qquad (7) \\ \operatorname{var}(b) &= 0. \qquad (8) \end{aligned}$$

If the population MEAN is not known, using the sample mean $\bar{x}$ instead of the population mean μ to compute

$$s^2 \equiv \hat{\sigma}_N^2 \equiv \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{9}$$

gives a BIASED ESTIMATOR of the population variance. In such cases, it is appropriate to use a STUDENT'S t-DISTRIBUTION instead of a GAUSSIAN DISTRIBUTION. However, it turns out (as discussed below) that an UNBIASED ESTIMATOR for the population variance is given by

$$s'^2 \equiv \hat{\sigma}_N'^2 \equiv \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2. \tag{10}$$

The MEAN and VARIANCE of the sample standard deviation for a distribution with population mean μ and VARIANCE are

$$\mu_{s_N{}^2} = \frac{N-1}{N} s^2 \tag{11}$$

$$\sigma_{s_N{}^2}{}^2 = \frac{N-1}{N^3}[(N-1)\mu_4 - (N-3)\mu_2{}^2]. \tag{12}$$

The quantity $N s_N{}^2/\sigma^2$ has a CHI-SQUARED DISTRIBUTION.

For multiple variables, the variance is given using the definition of COVARIANCE,

$$\begin{aligned} \operatorname{var}\left(\sum_{i=1}^{n} x_i\right) &= \operatorname{cov}\left(\sum_{i=1}^{n} x_i, \sum_{j=1}^{m} x_j\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} \operatorname{cov}(x_i, x_j) \\ &= \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j=i}}^{m} \operatorname{cov}(x_i, x_j) + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{m} \operatorname{cov}(x_i, x_j) \\ &= \sum_{i=1}^{n} \operatorname{cov}(x_i, x_j) + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{m} \operatorname{cov}(x_i, x_j) \\ &= \sum_{i=1}^{n} \operatorname{var}(x_i) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{m} \operatorname{cov}(x_i, x_j). \end{aligned} \tag{13}$$

A linear sum has a similar form:

$$\begin{aligned} \operatorname{var}\left(\sum_{i=1}^{n} a_i x_i\right) &= \operatorname{cov}\left(\sum_{i=1}^{n} a_i x_i, \sum_{j=1}^{m} a_j x_j\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i a_j \operatorname{cov}(x_i, x_j) \\ &= \sum_{i=1}^{n} a_i{}^2 \operatorname{var}(x_i) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{m} a_i a_j \operatorname{cov}(x_i, x_j). \end{aligned} \tag{14}$$

These equations can be expressed using the COVARIANCE MATRIX.

To estimate the population VARIANCE from a sample of N elements with a priori *unknown* MEAN (i.e., the MEAN is estimated from the sample itself), we need an

UNBIASED ESTIMATOR for σ. This is given by the k-STATISTIC k_2, where

$$k_2 = \frac{N}{N-1} m_2 \tag{15}$$

and $m_2 \equiv s^2$ is the SAMPLE VARIANCE

$$s^2 \equiv \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2. \tag{16}$$

Note that some authors prefer the definition

$$s'^2 \equiv \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2, \tag{17}$$

since this makes the sample variance an UNBIASED ESTIMATOR for the population variance.

When computing numerically, the MEAN must be computed before s^2 can be determined. This requires storing the set of sample values. It is possible to calculate s'^2 using a recursion relationship involving only the last sample as follows. Here, use μ_j to denote μ calculated from the first j samples (*not* the jth MOMENT)

$$\mu_j \equiv \frac{\sum_{i=1}^{j} x_i}{j}, \tag{18}$$

and $s_j{}^2$ denotes the value for the sample variance s'^2 calculated from the first j samples. The first few values calculated for the MEAN are

$$\mu_1 = x_1 \tag{19}$$

$$\mu_2 = \frac{1 \cdot \mu_1 + x_2}{2} \tag{20}$$

$$\mu_3 = \frac{2\mu_2 + x_3}{3}. \tag{21}$$

Therefore, for $j = 2, 3$ it is true that

$$\mu_j = \frac{(j-1)\mu_{j-1} + x_j}{j}. \tag{22}$$

Therefore, by induction,

$$\begin{aligned} \mu_{j+1} &= \frac{[(j+1)-1]\mu_{(j+1)-1} + x_{j+1}}{j+1} \\ &= \frac{j\mu_j + x_{j+1}}{j+1} \qquad (23) \\ \mu_{j+1}(j+1) &= (j+1)\mu_j + (x_{j+1} - \mu_j) \qquad (24) \\ \mu_{j+1} &= \mu_j + \frac{x_{j+1} - \mu_j}{j+1}, \qquad (25) \end{aligned}$$

and

$$s_j{}^2 = \frac{\sum_{i=1}^{j} (x_i - \mu_j)^2}{j-1} \tag{26}$$

for $j \geq 2$, so

$$\begin{aligned} j s_{j+1}{}^2 &= j \frac{\sum_{i=1}^{j+1} (x_i - \mu_{j+1})^2}{j} = \sum_{i=1}^{j+1} (x_i - \mu_{j+1})^2 \\ &= \sum_{i=1}^{j+1} [(x_i - \mu_j)(\mu_j - \mu_{j+1})]^2 \\ &= \sum_{i=1}^{j+1} (x_i - \mu_j)^2 + \sum_{i=1}^{j+1} (\mu_j - \mu_{j+1})^2 \\ &\quad + 2 \sum_{i=1}^{j+1} (x_i - \mu_j)(\mu_j - \mu_{j+1}). \qquad (27) \end{aligned}$$

Working on the first term,

$$\begin{aligned} \sum_{i=1}^{j+1} (x_i - \mu_j)^2 &= \sum_{i=1}^{j} (x_i - \mu_j)^2 + (x_{j+1} - \mu_j)^2 \\ &= (j-1) s_j{}^2 + (x_{j+1} - \mu_j)^2. \qquad (28) \end{aligned}$$

Use (24) to write

$$x_{j+1} - \mu_j = (j+1)(\mu_{j+1} - \mu_j), \tag{29}$$

so

$$\sum_{i=1}^{j+1} (x_i - \mu_j)^2 = (j-1) s_j{}^2 + (j+1)^2 (\mu_{j+1} - \mu_j)^2. \tag{30}$$

Now work on the second term in (27),

$$\sum_{i=1}^{j+1} (\mu_j - \mu_{j+1})^2 = (j+1)(\mu_j - \mu_{j+1})^2. \tag{31}$$

Considering the third term in (27),

$$\begin{aligned} \sum_{i=1}^{j+1} (x_i - \mu_j)(\mu_j - \mu_{j+1}) &= (\mu_j - \mu_{j+1}) \sum_{i=1}^{j+1} (x_i - \mu_j) \\ &= (\mu_j - \mu_{j+1}) \left[\sum_{i=1}^{j} (x_i - \mu_j) + (x_{j+1} - \mu_j) \right] \\ &= (\mu_j - \mu_{j+1}) \left(x_{j+1} - \mu_j - j\mu_j + \sum_{i=1}^{j} x_i \right). \qquad (32) \end{aligned}$$

But

$$\sum_{i=1}^{j} x_i = j\mu_j, \tag{33}$$

so

$$\begin{aligned} &\sum_{i=1}^{j+1} (\mu_j - \mu_{j+1})(x_{j+1} - \mu_j) \\ &\quad = \sum_{i=1}^{j+1} (\mu_j - \mu_{j+1})(j+1)(\mu_{j+1} - \mu_j) \\ &\quad = -(j+1)(\mu_j - \mu_{j+1})^2. \qquad (34) \end{aligned}$$

Plugging (30), (31), and (34) into (27),

$$\begin{aligned} j{s_{j+1}}^2 &= [(j-1){s_j}^2 + (j+1)^2(\mu_{j+1}-\mu_j)^2] \\ &\quad + [(j+1)(\mu_j-\mu_{j+1}) + 2[-(j+1)(\mu_j-\mu_{j+1})] \\ &= (j-1){s_j}^2 + (j+1)^2(\mu_{j+1}-\mu_j)^2 \\ &\quad - (j+1)(\mu_j-\mu_{j+1})^2 \\ &= (j-1){s_j}^2 + (j+1)[(j+1)-1](\mu_{j+1}-\mu_j)^2 \\ &= (j-1){s_j}^2 + j(j+1)(\mu_{j+1}-\mu_j)^2, \qquad (35) \end{aligned}$$

so

$${s_{j+1}}^2 = \left(1-\frac{1}{j}\right){s_j}^2 + (j+1)(\mu_{j+1}-\mu_j)^2. \qquad (36)$$

To find the variance of s^2 itself, remember that

$$\operatorname{var}(s^2) \equiv \left\langle s^4\right\rangle - \left\langle s^2\right\rangle^2, \qquad (37)$$

and

$$\left\langle s^2\right\rangle = \frac{N-1}{N}\mu_2. \qquad (38)$$

Now find $\left\langle s^4\right\rangle$.

$$\begin{aligned} \left\langle s^4\right\rangle &= \left\langle (s^2)^2\right\rangle = \left\langle (\langle x^2\rangle - \langle x\rangle^2)^2\right\rangle \\ &= \left\langle \left[\frac{1}{N}\sum {x_i}^2 - \left(\frac{1}{N}\sum x_i\right)^2\right]^2\right\rangle \\ &= \frac{1}{N^2}\left\langle \left(\sum x_i\right)^2\right\rangle - \frac{2}{N^3}\left\langle \sum {x_i}^2\left(\sum x_i\right)^2\right\rangle \\ &\quad + \frac{1}{N^4}\left\langle \left(\sum x_i\right)^4\right\rangle. \qquad (39) \end{aligned}$$

Working on the first term of (39),

$$\begin{aligned} \left\langle \left(\sum {x_i}^2\right)^2\right\rangle &= \left\langle \sum {x_i}^4 + \sum {x_i}^2{x_j}^2\right\rangle \\ &= \left\langle \sum {x_i}^4\right\rangle + \left\langle \sum {x_i}^2{x_j}^2\right\rangle \\ &= N\left\langle {x_i}^4\right\rangle + N(N-1)\left\langle {x_i}^2\right\rangle\left\langle {x_j}^2\right\rangle \\ &= N\mu_4' + N(N-1){\mu_2'}^2. \qquad (40) \end{aligned}$$

The second term of (39) is known from k-STATISTICS,

$$\left\langle \sum {x_i}^2\left(\sum x_j\right)^2\right\rangle = N\mu_4' + N(N-1){\mu_2'}^2, \qquad (41)$$

as is the third term,

$$\begin{aligned} \left\langle \left(\sum x_i\right)^4\right\rangle &= N\left\langle \sum {x_i}^4\right\rangle + 3N(N-1)\left\langle \sum {x_i}^2{x_j}^2\right\rangle \\ &= N\mu_4' + 3N(N-1){\mu_2'}^2. \qquad (42) \end{aligned}$$

Combining (39)-(42) gives

$$\begin{aligned} \left\langle s^4\right\rangle &= \frac{1}{N^2}[N\mu_4' + N(N-1){\mu_2'}^2] \\ &\quad - \frac{2}{N^3}[N\mu_4' + N(N-1){\mu_2'}^2] \\ &\quad + \frac{1}{N^4}[N\mu_4' + 3N(N-1){\mu_2'}^2] \\ &= \left(\frac{1}{N} - \frac{2}{N^2} + \frac{1}{N^3}\right)\mu_4' \\ &\quad + \left[\frac{N-1}{N} - \frac{2(N-1)}{N^2} + \frac{3(N-1)}{N^3}\right]{\mu_2'}^2 \\ &= \left(\frac{N^2-2N+1}{N^3}\right)\mu_4' \\ &\quad + \frac{(N-1)(N^2-2N+3)}{N^3}{\mu_2'}^2 \\ &= \frac{(N-1)[(N-1)\mu_4' + (N^2-2N+3){\mu_2'}^2]}{N^3}, \end{aligned} \qquad (43)$$

so plugging in (38) and (43) gives

$$\begin{aligned} \operatorname{var}(s^2) &= \left\langle s^4\right\rangle - \left\langle s^2\right\rangle^2 \\ &= \frac{(N-1)[(N-1)\mu_4' + (N^2-2N+3){\mu_2'}^2]}{N^3} \\ &\quad - \frac{(N-1)^2N}{N^3}{\mu_2'}^2 \\ &= \frac{N-1}{N^3}\{(N-1)\mu_4' + [(N^2-2N+3) \\ &\quad - N(N-1)]{\mu_2'}^2\} \\ &= \frac{(N-1)[(N-1)\mu_4' - (N-3){\mu_2'}^2]}{N^3}. \qquad (44) \end{aligned}$$

Student calculated the SKEWNESS and KURTOSIS of the distribution of s^2 as

$$\gamma_1 = \sqrt{\frac{8}{N-1}} \qquad (45)$$

$$\gamma_2 = \frac{12}{N-1} \qquad (46)$$

and conjectured that the true distribution is PEARSON TYPE III DISTRIBUTION

$$f(s^2) = C(s^2)^{(N-3)/2}e^{-Ns^2/2\sigma^2}, \qquad (47)$$

where

$$\sigma^2 = \frac{Ns^2}{N-1} \qquad (48)$$

$$C = \frac{\left(\frac{N}{2\sigma^2}\right)^{(N-1)/2}}{\Gamma\left(\frac{N-1}{2}\right)}. \qquad (49)$$

This was proven by R. A. Fisher.

The distribution of s itself is given by

$$f(s) = 2\frac{\left(\frac{N}{2\sigma^2}\right)^{(N-1)/2}}{\Gamma\left(\frac{N-1}{2}\right)} e^{-ns^2/2\sigma^2} s^{N-2} \qquad (50)$$

$$\langle s\rangle = \sqrt{\frac{2}{N}}\frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)}\sigma \equiv b(N)\sigma, \qquad (51)$$

where

$$b(N) \equiv \sqrt{\frac{2}{N}}\frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)}. \qquad (52)$$

The MOMENTS are given by

$$\mu_r = \left(\frac{2}{N}\right)^{r/2}\frac{\Gamma\left(\frac{N-1+r}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)}\sigma^r, \qquad (53)$$

and the variance is

$$\begin{aligned}\operatorname{var}(s) = \nu_2 - {\nu_1}^2 &= \frac{N-1}{N}\sigma^2 - [b(N)\sigma]^2 \\ &= \frac{1}{N}\left[N-1-\frac{2\Gamma^2\left(\frac{N}{2}\right)}{\Gamma^2\left(\frac{N-1}{2}\right)}\sigma^2\right]. \qquad (54)\end{aligned}$$

An UNBIASED ESTIMATOR of σ is $s/b(N)$. Romanovsky showed that

$$b(N) = 1 - \frac{3}{4N} - \frac{7}{32N^2} - \frac{139}{51849N^3} + \ldots. \qquad (55)$$

see also CORRELATION (STATISTICAL), COVARIANCE, COVARIANCE MATRIX, k-STATISTIC, MEAN, SAMPLE VARIANCE

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Moments of a Distribution: Mean, Variance, Skewness, and So Forth." §14.1 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 604–609, 1992.

Variate

A RANDOM VARIABLE in statistics.

Variation

The Δ-variation is a variation in which the varied path over which an integral is evaluated may end at different times than the correct path, and there may be variation in the coordinates at the endpoints.

The δ-variation is a variation in which the varied path in configuration space terminates at the endpoints representing the system configuration at the same time t_1 and t_2 as the correct path; i.e., the varied path always returns to the same endpoints in configuration space, so

$$\delta q_i(t_1) = \delta q_i(t_2) = 0.$$

see also CALCULUS OF VARIATIONS, VARIATION OF ARGUMENT, VARIATION OF PARAMETERS

Variation of Argument

Let $[\arg f(z)]$ denote the change in argument of a function $f(z)$ around a closed loop γ. Also let N denote the number of ROOTS of $f(z)$ in γ and P denote the number of POLES of $f(z)$ in γ. Then

$$[\arg f(z)] = \frac{1}{2\pi}(N-P). \qquad (1)$$

To find $[\arg f(z)]$ in a given region R, break R into paths and find $[\arg f(z)]$ for each path. On a circular ARC

$$z = Re^{i\theta}, \qquad (2)$$

let $f(z)$ be a POLYNOMIAL $P(z)$ of degree n. Then

$$\begin{aligned}[\arg P(z)] &= \left[\arg\left(z^n\frac{P(z)}{z^n}\right)\right] \\ &= [\arg z^n] + \left[\arg\left(\frac{P(z)}{z^n}\right)\right]. \qquad (3)\end{aligned}$$

Plugging in $z = Re^{i\theta}$ gives

$$[\arg P(z)] = [\arg Re^{i\theta n}] + \left[\arg\frac{P(Re^{i\theta})}{Re^{i\theta n}}\right] \qquad (4)$$

$$\lim_{R\to\infty}\frac{P(Re^{i\theta})}{Re^{i\theta n}} = [\text{constant}], \qquad (5)$$

so

$$\left[\frac{P(Re^{i\theta})}{Re^{i\theta n}}\right] = 0, \qquad (6)$$

and

$$[\arg P(z)] = [\arg e^{i\theta n}] = n(\theta_2 - \theta_1). \qquad (7)$$

For a REAL segment $z = x$,

$$[\arg f(x)] = \tan^{-1}\left[\frac{0}{f(x)}\right] = 0. \qquad (8)$$

For an IMAGINARY segment $z = iy$,

$$[\arg f(iy)] = \left\{\tan^{-1}\frac{\Im[P(iy)]}{\Re[P(iy)]}\right\}_{\theta_1}^{\theta_2}. \qquad (9)$$

Note that the ARGUMENT must change continuously, so "jumps" occur across inverse tangent asymptotes.

Variation Coefficient

If s_x is the STANDARD DEVIATION of a set of samples x_i and $\bar{x}$ its MEAN, then

$$V \equiv \frac{s_x}{\bar{x}}.$$

Variation of Parameters

For a second-order ORDINARY DIFFERENTIAL EQUATION,

$$y'' + p(x)y' + q(x)y = g(x). \tag{1}$$

Assume that linearly independent solutions $y_1(x)$ and $y_2(x)$ are known. Find v_1 and v_2 such that

$$y^*(x) = v_1(x)y_1(x) + v_2(x)y_2(x) \tag{2}$$
$$y^{*\prime}(x) = (v_1' + v_2'y_2) + (v_1y_1' + v_2y_2'). \tag{3}$$

Now, impose the additional condition that

$$v_1'y_1 + v_2'y_2 = 0 \tag{4}$$

so that

$$y^{*\prime}(x) = (v_1y_1' + v_2y_2') \tag{5}$$
$$y^{*\prime\prime}(x) = v_1'y_1' + v_2'y_2' + v_1y_1'' + v_2y_2'. \tag{6}$$

Plug y^*, $y^{*\prime}$, and $y^{*\prime\prime}$ back into the original equation to obtain

$$v_1(y_1''+py_1'+qy_1)+v_2(y_2''+py_2'+qy_2)+v_1'y_1'+v_2'y_2' = g(x) \tag{7}$$
$$v_1'y_1' + v_2'y_2' = g(x). \tag{8}$$

Therefore,

$$v_1'y_1 + v_2'y_2 = 0 \tag{9}$$
$$v_1'y_1' + v_2'y_2' = g(x). \tag{10}$$

Generalizing to an nth degree ODE, let y_1, ..., y_n be the solutions to the homogeneous ODE and let $v_1'(x)$, ..., $v_n'(x)$ be chosen such that

$$\begin{cases} y_1v_1' + y_2v_2' + \ldots + y_nv_n' = 0 \\ y_1'v_1' + y_2'v_2' + \ldots + y_n'v_n' = 0 \\ \vdots \\ y_1^{(n-1)}v_1' + y_2^{(n-1)}v_2' + \ldots + y_n^{(n-1)}v_n' = g(x). \end{cases} \tag{11}$$

Then the particular solution is

$$y^*(x) = v_1(x)y_1(x) + \ldots + v_n(x)y_n(x). \tag{12}$$

Variety

see ALGEBRAIC VARIETY

Varignon Parallelogram

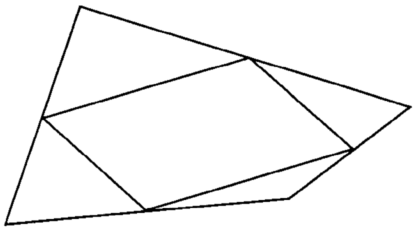

The figure formed when the BIMEDIANS (MIDPOINTS of the sides) of a convex QUADRILATERAL are joined. VARIGNON'S THEOREM demonstrated that this figure is a PARALLELOGRAM. The center of the Varignon parallelogram is the CENTROID if four point masses are placed on the VERTICES of the QUADRILATERAL.

see also MIDPOINT, PARALLELOGRAM, QUADRILATERAL, VARIGNON'S THEOREM

Varignon's Theorem

The figure formed when the BIMEDIANS (MIDPOINTS of the sides) of a convex QUADRILATERAL are joined in order is a PARALLELOGRAM. Equivalently, the BIMEDIANS bisect each other. The AREA of this VARIGNON PARALLELOGRAM is half that of the QUADRILATERAL. The PERIMETER is equal to the sum of the diagonals of the original QUADRILATERAL.

see also BIMEDIAN, MIDPOINT, QUADRILATERAL, VARIGNON PARALLELOGRAM

References

Coxeter, H. S. M. and Greitzer, S. L. *Geometry Revisited.* Washington, DC: Math. Assoc. Amer., pp. 51–56, 1967.

Vassiliev Polynomial

Vassiliev (1990) introduced a radically new way of looking at KNOTS by considering a multidimensional space in which each point represents a possible 3-D knot configuration. If two KNOTS are equivalent, a path then exists in this space from one to the other. The paths can be associated with polynomial invariants.

Birman and Lin (1993) subsequently found a way to translate this scheme into a set of rules and list of potential starting points, which makes analysis of Vassiliev polynomials much simpler. Bar-Natan (1995) and Birman and Lin (1993) proved that JONES POLYNOMIALS and several related expressions are directly connected (Peterson 1992). In fact, substituting the POWER series for e^x as the variable in the JONES POLYNOMIAL yields a POWER SERIES whose COEFFICIENTS are Vassiliev polynomials (Birman and Lin 1993). Bar-Natan (1995) also discovered a link with Feynman diagrams (Peterson 1992).

References

Bar-Natan, D. "On the Vassiliev Knot Invariants." *Topology* **34**, 423–472, 1995.

Birman, J. S. "New Points of View in Knot Theory." *Bull. Amer. Math. Soc.* **28**, 253–287, 1993.

Birman, J. S. and Lin, X.-S. "Knot Polynomials and Vassiliev's Invariants." *Invent. Math.* **111**, 225–270, 1993.

Peterson, I. "Knotty Views: Tying Together Different Ways of Looking at Knots." *Sci. News* **141**, 186–187, 1992.

Praslov, V. V. and Sossinsky, A. B. *Knots, Links, Braids and 3-Manifolds: An Introduction to the New Invariants in Low-Dimensional Topology.* Providence, RI: Amer. Math. Soc., 1996.

Stoimenow, A. "Degree-3 Vassiliev Invariants." `http://www.informatik.hu-berlin.de/~stoimeno/vas3.html`.

Vassiliev, V. A. "Cohomology of Knot Spaces." In *Theory of Singularities and Its Applications* (Ed. V. I. Arnold). Providence, RI: Amer. Math. Soc., pp. 23–69, 1990.

Vassiliev, V. A. *Complements of Discriminants of Smooth Maps: Topology and Applications.* Providence, RI: Amer. Math. Soc., 1992.

Vault

Let a vault consist of two equal half-CYLINDERS of length and diameter $2a$ which intersect at RIGHT ANGLES so that the lines of their intersections (the "groins") terminate in the VERTICES of a SQUARE. Then the SURFACE AREA of the vault is given by

$$A = 4(\pi - 2)a^2.$$

see also DOME

References

Lines, L. *Solid Geometry.* New York: Dover, pp. 112–113, 1965.

Vector

A vector is a set of numbers $A_0, \ldots, A_n$ that transform as

$$A_i' = a_{ij} A_j. \tag{1}$$

This makes a vector a TENSOR of RANK 1. Vectors are invariant under TRANSLATION, and they reverse sign upon inversion.

A vector is uniquely specified by giving its DIVERGENCE and CURL within a region and its normal component over the boundary, a result known as HELMHOLTZ'S THEOREM (Arfken 1985, p. 79). A vector from a point A to a point B is denoted $\overrightarrow{AB}$, and a vector v may be denoted $\vec{v}$, or more commonly, $\mathbf{v}$.

A vector with unit length is called a UNIT VECTOR and is denoted with a HAT. An arbitrary vector may be converted to a UNIT VECTOR by dividing by its NORM, i.e.,

$$\hat{\mathbf{v}} \equiv \frac{\mathbf{v}}{|\mathbf{v}|}. \tag{2}$$

Let $\hat{\mathbf{n}}$ be the UNIT VECTOR defined by

$$\hat{\mathbf{n}} \equiv \begin{bmatrix} \cos\theta \sin\phi \\ \sin\theta \sin\phi \\ \cos\phi \end{bmatrix}. \tag{3}$$

Then the vectors $\hat{\mathbf{n}}$, $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ satisfy the identities

$$\begin{aligned}
\langle n_x \rangle &= \int_0^{2\pi} \int_0^{\pi} (\cos\theta \sin\phi) \sin\phi \, d\theta \, d\phi \\
&= [\sin\theta]_0^{2\pi} \int_0^{2\pi} \sin^2\phi \, d\phi = 0 && (4) \\
\langle n_i \rangle &= 0 && (5) \\
\langle n_i n_j \rangle &= \tfrac{1}{3}\delta_{ij} && (6) \\
\langle n_i n_k n_k \rangle &= 0 && (7) \\
\langle n_i n_k n_l n_m \rangle &= \tfrac{1}{15}(\delta_{ik}\delta_{lm} + \delta_{il}\delta_{km} + \delta_{im}\delta_{kl}) && (8) \\
\left\langle (\mathbf{a}\cdot\hat{\mathbf{n}})^2 \right\rangle &= \tfrac{1}{3}a^2 && (9) \\
\langle (\mathbf{a}\cdot\hat{\mathbf{n}})(\mathbf{b}\cdot\hat{\mathbf{n}}) \rangle &= \tfrac{1}{3}\mathbf{a}\cdot\mathbf{b} && (10) \\
\langle (\mathbf{a}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}} \rangle &= \tfrac{1}{3}a && (11) \\
\left\langle (\mathbf{a}\times\hat{\mathbf{n}})^2 \right\rangle &= \tfrac{2}{3}a^2 && (12) \\
\langle (\mathbf{a}\times\hat{\mathbf{n}})\cdot(\mathbf{b}\times\hat{\mathbf{n}}) \rangle &= \tfrac{2}{3}\mathbf{a}\cdot\mathbf{b}, && (13)
\end{aligned}$$

and

$$\begin{aligned}
&\langle (\mathbf{a}\cdot\hat{\mathbf{n}})(\mathbf{b}\cdot\hat{\mathbf{n}})(\mathbf{c}\cdot\hat{\mathbf{n}})(\mathbf{d}\cdot\hat{\mathbf{n}}) \rangle \\
&= \tfrac{1}{15}[(bfa\cdot\mathbf{b})(bfc\cdot\mathbf{d}) + (bfa\cdot\mathbf{c})(bfb\cdot\mathbf{d}) + (bfa\cdot\mathbf{d})(bfb\cdot\mathbf{c})] \qquad (14)
\end{aligned}$$

where δ_{ij} is the KRONECKER DELTA, $\mathbf{a}\cdot\mathbf{b}$ is a DOT PRODUCT, and EINSTEIN SUMMATION has been used.

see also FOUR-VECTOR, HELMHOLTZ'S THEOREM, NORM, PSEUDOVECTOR, SCALAR, TENSOR, UNIT VECTOR, VECTOR FIELD

References

Arfken, G. "Vector Analysis." Ch. 1 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 1–84, 1985.

Aris, R. *Vectors, Tensors, and the Basic Equations of Fluid Mechanics.* New York: Dover, 1989.

Crowe, M. J. *A History of Vector Analysis: The Evolution of the Idea of a Vectorial System.* New York: Dover, 1985.

Gibbs, J. W. and Wilson, E. B. *Vector Analysis: A Text-Book for the Use of Students of Mathematics and Physics, Founded Upon the Lectures of J. Willard Gibbs.* New York: Dover, 1960.

Marsden, J. E. and Tromba, A. J. *Vector Calculus, 4th ed.* New York: W. H. Freeman, 1996.

Morse, P. M. and Feshbach, H. "Vector and Tensor Formalism." §1.5 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 44–54, 1953.

Schey, H. M. *Div, Grad, Curl, and All That: An Informal Text on Vector Calculus.* New York: Norton, 1973.

Schwartz, M.; Green, S.; and Rutledge, W. A. *Vector Analysis with Applications to Geometry and Physics.* New York: Harper Brothers, 1960.

Spiegel, M. R. *Theory and Problems of Vector Analysis.* New York: Schaum, 1959.

Vector Bundle

A special class of FIBER BUNDLE in which the FIBER is a VECTOR SPACE. Technically, a little more is required; namely, if $f : E \to B$ is a BUNDLE with FIBER $\mathbb{R}^n$, to be a vector bundle, all of the FIBERS $f^{-1}(x)$ for

$x \in B$ need to have a coherent VECTOR SPACE structure. One way to say this is that the "trivializations" $h : f^{-1}(U) \to U \times \mathbb{R}^n$, are FIBER-for-FIBER VECTOR SPACE ISOMORPHISMS.

see also BUNDLE, FIBER, FIBER BUNDLE, LIE ALGEBROID, STABLE EQUIVALENCE, TANGENT MAP, VECTOR SPACE, WHITNEY SUM

Vector Derivative

The basic types of derivatives operating on a VECTOR FIELD are the CURL $\nabla\times$, DIVERGENCE $\nabla\cdot$, and GRADIENT ∇.

Vector derivative identities involving the CURL include

$$\nabla \times (k\mathbf{A}) = k\nabla \times \mathbf{A} \tag{1}$$

$$\nabla \times (f\mathbf{A}) = f(\nabla \times \mathbf{A}) + (\nabla f) \times \mathbf{A} \tag{2}$$

$$\begin{aligned}\nabla \times (\mathbf{A} \times \mathbf{B}) = (\mathbf{B} \cdot \nabla)\mathbf{A} - (\mathbf{A} \cdot \nabla)\mathbf{B} \\ +\mathbf{A}(\nabla \cdot \mathbf{B}) - \mathbf{B}(\nabla \cdot \mathbf{A})\end{aligned} \tag{3}$$

$$\nabla \times \left(\frac{\mathbf{A}}{f}\right) = \frac{f(\nabla \times \mathbf{A}) + \mathbf{A} \times (\nabla f)}{f^2} \tag{4}$$

$$\nabla \times (\mathbf{A} + \mathbf{B}) = \nabla \times \mathbf{A} + \nabla \times \mathbf{B}. \tag{5}$$

In SPHERICAL COORDINATES,

$$\nabla \times \mathbf{r} = \mathbf{0} \tag{6}$$

$$\nabla \times \hat{\mathbf{r}} = \mathbf{0} \tag{7}$$

$$\begin{aligned}\nabla \times [\mathbf{r}f(r)] &= f(r)(\nabla \times \mathbf{r}) + [\nabla f(r)] \times \mathbf{r} \\ &= f(r)(\mathbf{0}) + \frac{df}{dr}\hat{\mathbf{r}} \times \mathbf{r} = \mathbf{0} + \mathbf{0} = \mathbf{0}.\end{aligned} \tag{8}$$

Vector derivative identities involving the DIVERGENCE include

$$\nabla \cdot (k\mathbf{A}) = k\nabla \cdot \mathbf{A} \tag{9}$$

$$\nabla \cdot (f\mathbf{A}) = f(\nabla \cdot \mathbf{A}) + (\nabla f) \cdot \mathbf{A} \tag{10}$$

$$\nabla \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot (\nabla \times \mathbf{B}) \tag{11}$$

$$\nabla \cdot \left(\frac{\mathbf{A}}{f}\right) = \frac{f(\nabla \cdot \mathbf{A}) - (\nabla f) \cdot \mathbf{A}}{f^2} \tag{12}$$

$$\nabla \cdot (\mathbf{A} + \mathbf{B}) = \nabla \cdot \mathbf{A} + \nabla \cdot \mathbf{B} \tag{13}$$

$$\nabla(\mathbf{uv}) = \mathbf{u}\nabla \cdot \mathbf{v} + (\nabla \mathbf{u}) \cdot \mathbf{v}. \tag{14}$$

In SPHERICAL COORDINATES,

$$\nabla \cdot \mathbf{r} = 3 \tag{15}$$

$$\nabla \cdot \hat{\mathbf{r}} = \frac{2}{r} \tag{16}$$

$$\nabla \cdot [\mathbf{r}f(r)] = \frac{\partial}{\partial x}[xf(r)] + \frac{\partial}{\partial y}[yf(r)] + \frac{\partial}{\partial z}[zf(r)] \tag{17}$$

$$\frac{\partial}{\partial x}[xf(r)] = x\frac{\partial f}{\partial x} + f = x\frac{\partial f}{\partial r}\frac{\partial r}{\partial x} + f \tag{18}$$

$$\frac{\partial r}{\partial x} = \frac{\partial}{\partial x}(x^2+y^2+z^2)^{1/2} = x(x^2+y^2+z^2)^{-1/2} = \frac{x}{r} \tag{19}$$

$$\frac{\partial}{\partial x}[xf(r)] = \frac{x^2}{r}\frac{df}{dr} + f. \tag{20}$$

By symmetry,

$$\nabla \cdot [\mathbf{r}f(r)] = 3f(r) + \frac{1}{r}(x^2+y^2+z^2)\frac{df}{dr} = 3f(r) + r\frac{df}{dr} \tag{21}$$

$$\nabla \cdot (\hat{\mathbf{r}}f(r)) = \frac{3}{r}f(r) + \frac{df}{dr} \tag{22}$$

$$\nabla \cdot (\hat{\mathbf{r}}r^n) = 3r^{n-1} + (n-1)r^{n-1} = (n+2)r^{n-1}. \tag{23}$$

Vector derivative identities involving the GRADIENT include

$$\nabla(kf) = k\nabla f \tag{24}$$

$$\nabla(fg) = f\nabla g + g\nabla f \tag{25}$$

$$\begin{aligned}\nabla(\mathbf{A} \cdot \mathbf{B}) = \mathbf{A} \times (\nabla \times \mathbf{B}) + \mathbf{B} \times (\nabla \times \mathbf{A}) \\ +(\mathbf{A} \cdot \nabla)\mathbf{B} + (\mathbf{B} \cdot \nabla)\mathbf{A}\end{aligned} \tag{26}$$

$$\begin{aligned}\nabla(\mathbf{A} \cdot \nabla f) &= \mathbf{A} \times (\nabla \times \nabla f) + \nabla f \times (\nabla \times \mathbf{A}) \\ &\quad +\mathbf{A} \cdot \nabla(\nabla f) + \nabla f \cdot \nabla \mathbf{A} \\ &= \nabla f \times (\nabla \times \mathbf{A}) + \mathbf{A} \cdot \nabla(\nabla f) + \nabla f \cdot \nabla \mathbf{A}\end{aligned} \tag{27}$$

$$\nabla\left(\frac{f}{g}\right) = \frac{g\nabla f - f\nabla g}{g^2} \tag{28}$$

$$\nabla(f+g) = \nabla f + \nabla g \tag{29}$$

$$\nabla(\mathbf{A} \cdot \mathbf{A}) = 2\mathbf{A} \times (\nabla \times \mathbf{A}) + 2(\mathbf{A} \cdot \nabla)\mathbf{A} \tag{30}$$

$$(\mathbf{A} \cdot \nabla)\mathbf{A} = \nabla(\tfrac{1}{2}\mathbf{A}^2) - \mathbf{A} \times (\nabla \times \mathbf{A}). \tag{31}$$

Vector second derivative identities include

$$\nabla^2 t \equiv \nabla \cdot (\nabla t) = \frac{\partial^2 t}{\partial x^2} + \frac{\partial^2 t}{\partial y^2} + \frac{\partial^2 t}{\partial z^2} \tag{32}$$

$$\nabla^2 \mathbf{A} = \nabla(\nabla \cdot \mathbf{A}) - \nabla \times (\nabla \times \mathbf{A}). \tag{33}$$

This very important second derivative is known as the LAPLACIAN.

$$\nabla \times (\nabla t) = \mathbf{0} \tag{34}$$

$$\nabla(\nabla \cdot \mathbf{A}) = \nabla^2\mathbf{A} + \nabla \times (\nabla \times \mathbf{A}) \tag{35}$$

$$\nabla \cdot (\nabla \times \mathbf{A}) = 0 \tag{36}$$

$$\begin{aligned}&\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2\mathbf{A} \\ &\nabla \times (\nabla^2\mathbf{A}) = \nabla \times [\nabla(\nabla \cdot \mathbf{A})] - \nabla \times [\nabla \times (\nabla \times \mathbf{A})] \\ &\quad = -\nabla \times [\nabla \times (\nabla \times \mathbf{A})] \\ &\quad = -\{\nabla[\nabla \cdot (\nabla \times \mathbf{A})] - \nabla^2(\nabla \times \mathbf{A})]\} \\ &\quad = \nabla^2(\nabla \times \mathbf{A})\end{aligned} \tag{37}$$

$$\begin{aligned}\nabla^2(\nabla \cdot \mathbf{A}) &= \nabla \cdot [\nabla(\nabla \cdot \mathbf{A})] \\ &= \nabla \cdot [\nabla^2\mathbf{A} + \nabla \times (\nabla \times \mathbf{A})] = \nabla \cdot (\nabla^2\mathbf{A})\end{aligned} \tag{38}$$

$$\begin{aligned}\nabla^2[\nabla \times (\nabla \times \mathbf{A})] &= \nabla^2[\nabla(\nabla \cdot \mathbf{A}) - \nabla^2\mathbf{A}] \\ &= \nabla^2[\nabla(\nabla \cdot \mathbf{A})] - \nabla^4\mathbf{A}\end{aligned} \tag{39}$$

$$\nabla \times [\nabla^2(\nabla \times \mathbf{A})] = \nabla^2[\nabla(\nabla \cdot \mathbf{A})] - \nabla^4\mathbf{A} \tag{40}$$

$$\begin{aligned}\nabla^4\mathbf{A} &= -\nabla^2[\nabla \times (\nabla \times \mathbf{A})] + \nabla^2[\nabla(\nabla \cdot \mathbf{A})] \\ &= \nabla \times [\nabla^2(\nabla \times \mathbf{A})] - \nabla^2[\nabla \times (\nabla \times \mathbf{A})].\end{aligned} \tag{41}$$

Combination identities include

$$\mathbf{A} \times (\nabla \mathbf{A}) = \tfrac{1}{2}\nabla(\mathbf{A} \cdot \mathbf{A}) - (\mathbf{A} \cdot \nabla)\mathbf{A} \tag{42}$$

$$\nabla \times (\phi \nabla \phi) = \phi \nabla \times (\nabla \phi) + (\nabla \phi) \times (\nabla \phi) = \mathbf{0} \tag{43}$$

$$(\mathbf{A} \cdot \nabla)\hat{\mathbf{r}} = \frac{\mathbf{A} - \hat{\mathbf{r}}(\mathbf{A} \cdot \hat{\mathbf{r}})}{r} \tag{44}$$

$$\nabla f \cdot \mathbf{A} = \nabla \cdot (f\mathbf{A}) - f(\nabla \cdot \mathbf{A}) \tag{45}$$

$$f(\nabla \cdot \mathbf{A}) = \nabla \cdot (f\mathbf{A}) - \mathbf{A}\nabla f, \tag{46}$$

where (45) and (46) follow from divergence rule (2).

see also CURL, DIVERGENCE, GRADIENT, LAPLACIAN, VECTOR INTEGRAL, VECTOR QUADRUPLE PRODUCT, VECTOR TRIPLE PRODUCT

References

Gradshteyn, I. S. and Ryzhik, I. M. "Vector Field Theorem." Ch. 10 in *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, pp. 1081–1092, 1980.

Morse, P. M. and Feshbach, H. "Table of Useful Vector and Dyadic Equations." *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 50–54 and 114–115, 1953.

Vector Direct Product

Given VECTORS $\mathbf{u}$ and $\mathbf{v}$, the vector direct product is

$$\mathbf{uv} \equiv \mathbf{u} \otimes \mathbf{v}^{\mathrm{T}},$$

where $\otimes$ is the MATRIX DIRECT PRODUCT and $\mathbf{v}^{\mathrm{T}}$ is the matrix TRANSPOSE. For 3×3 vectors

$$\mathbf{uv} = \begin{bmatrix} u_1\mathbf{v}^{\mathrm{T}} \\ u_2\mathbf{v}^{\mathrm{T}} \\ u_3\mathbf{v}^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} u_1v_1 & u_1v_2 & u_1v_3 \\ u_2v_1 & u_2v_2 & u_2v_3 \\ u_3v_1 & u_3v_2 & u_3v_3 \end{bmatrix}.$$

Note that if $\mathbf{u} = \hat{\mathbf{x}}_i$, then $u_j = \delta_{ij}$, where δ_{ij} is the KRONECKER DELTA.

see also MATRIX DIRECT PRODUCT, SHERMAN-MORRISON FORMULA, WOODBURY FORMULA

Vector Division

There is no unique solution A to the MATRIX equation $\mathbf{y} = \mathsf{A}\mathbf{x}$ unless $\mathbf{x}$ is PARALLEL to $\mathbf{y}$, in which case A is a SCALAR. Therefore, vector division is not defined.

see also MATRIX, SCALAR

Vector Field

A MAP $\mathbf{f} : \mathbb{R}^n \mapsto \mathbb{R}^n$ which assigns each $\mathbf{x}$ a VECTOR FUNCTION $\mathbf{f}(\mathbf{x})$. FLOWS are generated by vector fields and vice versa. A vector field is a SECTION of its TANGENT BUNDLE.

see also FLOW, SCALAR FIELD, SEIFERT CONJECTURE, TANGENT BUNDLE, VECTOR, WILSON PLUG

References

Gray, A. "Vector Fields $\mathbb{R}^n$" and "Derivatives of Vector Fields $\mathbb{R}^n$." §9.4–9.5 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 171–174 and 175–178, 1993.

Morse, P. M. and Feshbach, H. "Vector Fields." §1.2 in *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 8-21, 1953.

Vector Function

A function of one or more variables whose RANGE is 3-dimensional, as compared to a SCALAR FUNCTION, whose RANGE is 1-dimensional.

see also COMPLEX FUNCTION, REAL FUNCTION, SCALAR FUNCTION

Vector Harmonic

see VECTOR SPHERICAL HARMONIC

Vector Integral

The following vector integrals are related to the CURL THEOREM. If

$$\mathbf{F} \equiv \mathbf{c} \times \mathbf{P}(x, y, z), \tag{1}$$

then

$$\int_C d\mathbf{s} \times \mathbf{P} = \int_S (d\mathbf{a} \times \nabla) \times \mathbf{P}. \tag{2}$$

If

$$\mathbf{F} \equiv \mathbf{c}F, \tag{3}$$

then

$$\int_C F\,d\mathbf{s} = \int_S d\mathbf{a} \times \nabla \mathbf{F}. \tag{4}$$

The following are related to the DIVERGENCE THEOREM. If

$$\mathbf{F} \equiv \mathbf{c} \times \mathbf{P}(x, y, z), \tag{5}$$

then

$$\int_V \nabla \times \mathbf{F}\,dV = \int_S d\mathbf{a} \times \mathbf{F}. \tag{6}$$

Finally, if

$$\mathbf{F} \equiv \mathbf{c}F, \tag{7}$$

then

$$\int_V \nabla F\,dV = \int_S F\,d\mathbf{a}. \tag{8}$$

see also CURL THEOREM, DIVERGENCE THEOREM, GRADIENT THEOREM, GREEN'S FIRST IDENTITY, GREEN'S SECOND IDENTITY, LINE INTEGRAL, SURFACE INTEGRAL, VECTOR DERIVATIVE, VOLUME INTEGRAL

Vector Norm

Given an n-D VECTOR

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

a vector norm $||\mathbf{x}||$ (sometimes written simply $|\mathbf{x}|$) is a NONNEGATIVE number satisfying

1. $||x|| > 0$ when $\mathbf{x} \neq \mathbf{0}$ and $||\mathbf{x}|| = 0$ IFF $\mathbf{x} = \mathbf{0}$,
2. $||k\mathbf{x}|| = |k|\,||\mathbf{x}||$ for any SCALAR k,
3. $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$.

see also COMPATIBLE, MATRIX NORM, NATURAL NORM, NORM

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1114, 1980.

Vector Ordering

If the first NONZERO component of the vector difference $\mathbf{A}-\mathbf{B}$ is >0, then $\mathbf{A}\succ\mathbf{B}$. If the first NONZERO component of $\mathbf{A}-\mathbf{B}$ is <0, then $\mathbf{A}\prec\mathbf{B}$.

see also PRECEDES, SUCCEEDS

Vector Potential

A function $\mathbf{A}$ such that

$$\mathbf{B}\equiv\nabla\times\mathbf{A}.$$

The most common use of a vector potential is the representation of a magnetic field. If a VECTOR FIELD has zero DIVERGENCE, it may be represented by a vector potential.

see also DIVERGENCE, HELMHOLTZ'S THEOREM, POTENTIAL FUNCTION, SOLENOIDAL FIELD, VECTOR FIELD

Vector Quadruple Product

$$(\mathbf{A}\times\mathbf{B})\cdot(\mathbf{C}\times\mathbf{D})=(\mathbf{A}\cdot\mathbf{C})(\mathbf{B}\cdot\mathbf{D})-(\mathbf{A}\cdot\mathbf{D})(\mathbf{B}\cdot\mathbf{C}) \tag{1}$$

$$\begin{aligned}(\mathbf{A}\times\mathbf{B})^2&\equiv(\mathbf{A}\times\mathbf{B})\cdot(\mathbf{A}\times\mathbf{B})\\&=(\mathbf{A}\cdot\mathbf{A})(\mathbf{B}\cdot\mathbf{B})-(\mathbf{A}\cdot\mathbf{B})(\mathbf{B}\cdot\mathbf{A})\\&=\mathbf{A}^2\mathbf{B}^2-(\mathbf{A}\cdot\mathbf{B})^2\end{aligned} \tag{2}$$

$$\mathbf{A}\times(\mathbf{B}\times(\mathbf{C}\times\mathbf{D}))=\mathbf{B}(\mathbf{A}\cdot(\mathbf{C}\times\mathbf{D}))-(\mathbf{A}\cdot\mathbf{B})(\mathbf{C}\times\mathbf{D}) \tag{3}$$

$$\begin{aligned}&(\mathbf{A}\times\mathbf{B})\times(\mathbf{C}\times\mathbf{D})=[\mathbf{A},\mathbf{B},\mathbf{D}]\mathbf{C}-[\mathbf{A},\mathbf{B},\mathbf{C}]\mathbf{D}\\&=(\mathbf{C}\times\mathbf{D})\times(\mathbf{B}\times\mathbf{A})=[\mathbf{C},\mathbf{D},\mathbf{A}]\mathbf{D}-[\mathbf{C},\mathbf{D},\mathbf{B}]\mathbf{A},\end{aligned} \tag{4}$$

where $[\mathbf{A},\mathbf{B},\mathbf{D}]$ denotes the VECTOR TRIPLE PRODUCT. Equation (1) is known as LAGRANGE'S IDENTITY.

see also LAGRANGE'S IDENTITY, VECTOR TRIPLE PRODUCT

Vector Space

A vector space over $\mathbb{R}^n$ is a set of VECTORS for which any VECTORS $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}\in\mathbb{R}^n$ and any SCALARS r, $s\in\mathbb{R}$ have the following properties:

1. COMMUTATIVITY:

$$\mathbf{X}+\mathbf{Y}=\mathbf{Y}+\mathbf{X}.$$

2. ASSOCIATIVITY of vector addition:

$$(\mathbf{X}+\mathbf{Y})+\mathbf{Z}=\mathbf{X}+(\mathbf{Y}+\mathbf{Z}).$$

3. Additive identity: For all $\mathbf{X}$,

$$\mathbf{0}+\mathbf{X}=\mathbf{X}+\mathbf{0}=\mathbf{X}.$$

4. Existence of additive inverse: For any $\mathbf{X}$, there exists a $-\mathbf{X}$ such that

$$\mathbf{X}+(-\mathbf{X})=\mathbf{0}.$$

5. ASSOCIATIVITY of scalar multiplication:

$$r(s\mathbf{X})=(rs)\mathbf{X}.$$

6. DISTRIBUTIVITY of scalar sums:

$$(r+s)\mathbf{X}=r\mathbf{X}+s\mathbf{X}.$$

7. DISTRIBUTIVITY of vector sums:

$$r(\mathbf{X}+\mathbf{Y})=r\mathbf{X}+r\mathbf{Y}.$$

8. Scalar multiplication identity:

$$1\mathbf{X}=\mathbf{X}.$$

An n-D vector space of characteristic two has

$$S(k,n)=(2^n-2^0)(2^n-2^1)\cdots(2^n-2^{k-1})$$

distinct SUBSPACES of DIMENSION k.

A MODULE is abstractly similar to a vector space, but it uses a RING to define COEFFICIENTS instead of the FIELD used for vector spaces. MODULES have COEFFICIENTS in much more general algebraic objects.

see also BANACH SPACE, FIELD, FUNCTION SPACE, HILBERT SPACE, INNER PRODUCT SPACE, MODULE, RING, TOPOLOGICAL VECTOR SPACE

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 530–534, 1985.

Vector Spherical Harmonic

The SPHERICAL HARMONICS can be generalized to vector spherical harmonics by looking for a SCALAR FUNCTION ψ and a constant VECTOR $\mathbf{c}$ such that

$$\begin{aligned}\mathbf{M}&\equiv\nabla\times(\mathbf{c}\psi)=\psi(\nabla\times\mathbf{c})+(\nabla\psi)\times\mathbf{c}\\&=(\nabla\psi)\times\mathbf{c}=-\mathbf{c}\times\nabla,\psi\end{aligned} \tag{1}$$

so

$$\nabla\cdot\mathbf{M}=0. \tag{2}$$

Now use the vector identities

$$\begin{aligned}\nabla^2\mathbf{M}&=\nabla^2(\nabla\times\mathbf{M})=\nabla\times(\nabla^2\mathbf{M})\\&=\nabla\times(\nabla^2\mathbf{c}\psi)=\nabla\times(\mathbf{c}\nabla^2\psi)\end{aligned} \tag{3}$$

$$k^2\mathbf{M}=k^2\nabla\times(\mathbf{c}\psi)=\nabla\times(\mathbf{c}\nabla^2\psi), \tag{4}$$

so

$$\nabla^2\mathbf{M} + k^2\mathbf{M} = \nabla \times [\mathbf{c}(\nabla^2\psi + k^2\psi)], \tag{5}$$

and $\mathbf{M}$ satisfies the vector HELMHOLTZ DIFFERENTIAL EQUATION if ψ satisfies the scalar HELMHOLTZ DIFFERENTIAL EQUATION

$$\nabla^2\psi + k^2\psi = 0. \tag{6}$$

Construct another vector function

$$\mathbf{N} \equiv \frac{\nabla \times \mathbf{M}}{k}, \tag{7}$$

which also satisfies the vector HELMHOLTZ DIFFERENTIAL EQUATION since

$$\begin{aligned}\nabla^2\mathbf{N} &= \frac{1}{k}\nabla^2(\nabla \times \mathbf{M}) = \frac{1}{k}\nabla \times (\nabla^2\mathbf{M})\\ &= \frac{1}{k}\nabla \times (-k^2\mathbf{M}) = -k\nabla \times \mathbf{M} = -k^2\mathbf{N},\end{aligned} \tag{8}$$

which gives

$$\nabla^2\mathbf{N} + k^2\mathbf{N} = 0. \tag{9}$$

We have the additional identity

$$\begin{aligned}\nabla \times \mathbf{N} &= \frac{1}{k}\nabla \times (\nabla \times \mathbf{M}) = \frac{1}{k}\nabla(\nabla \cdot \mathbf{M})\\ &= \frac{1}{k}\nabla^2\mathbf{M} - \frac{1}{k}\nabla^2\mathbf{M} = \frac{-\nabla^2\mathbf{M}}{k} = k\mathbf{M}.\end{aligned} \tag{10}$$

In this formalism, ψ is called the generating function and $\mathbf{c}$ is called the PILOT VECTOR. The choice of generating function is determined by the symmetry of the scalar equation, i.e., it is chosen to solve the desired scalar differential equation. If $\mathbf{M}$ is taken as

$$\mathbf{M} = \nabla \times (\mathbf{r}\psi), \tag{11}$$

where $\mathbf{r}$ is the radius vector, then $\mathbf{M}$ is a solution to the vector wave equation in spherical coordinates. If we want vector solutions which are tangential to the radius vector,

$$\mathbf{M} \cdot \mathbf{r} = \mathbf{r} \cdot (\nabla\psi \times \mathbf{c}) = (\nabla\psi)(\mathbf{c} \times \mathbf{r}) = 0, \tag{12}$$

so

$$\mathbf{c} \times \mathbf{r} = \mathbf{0} \tag{13}$$

and we may take

$$\mathbf{c} = \mathbf{r} \tag{14}$$

(Arfken 1985, pp. 707–711; Bohren and Huffman 1983, p. 88).

A number of conventions are in use. Hill (1954) defines

$$\begin{aligned}\mathbf{V}_l^m &\equiv -\sqrt{\frac{l+1}{2l+1}}Y_l^m\hat{\mathbf{r}} + \frac{1}{\sqrt{(l+1)(2l+1)}}\frac{\partial Y_l^m}{\partial\theta}\hat{\boldsymbol{\theta}}\\ &\quad + iM\sqrt{(l+1)(2l+1)}\sin\theta Y_l^m\hat{\boldsymbol{\phi}}\end{aligned} \tag{15}$$

$$\begin{aligned}\mathbf{W}_l^m &= \sqrt{\frac{l}{2l+1}}Y_l^m\hat{\mathbf{r}} + \frac{1}{\sqrt{l(2l+1)}}\frac{\partial Y_l^m}{\partial\theta}\hat{\boldsymbol{\theta}}\\ &\quad + \frac{iM}{\sqrt{l(2l+1)}\sin\theta}Y_l^m\hat{\boldsymbol{\phi}}\end{aligned} \tag{16}$$

$$\mathbf{X}_l^m = -\frac{M}{\sqrt{l(l+1)}\sin\theta}Y_l^m\hat{\boldsymbol{\theta}} - \frac{i}{\sqrt{l(l+1)}}\frac{\partial Y_l^m}{\partial\theta}\hat{\boldsymbol{\phi}}. \tag{17}$$

Morse and Feshbach (1953) define vector harmonics called $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{P}$ using rather complicated expressions.

References

Arfken, G. "Vector Spherical Harmonics." §12.11 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 707–711, 1985.

Blatt, J. M. and Weisskopf, V. "Vector Spherical Harmonics." Appendix B, §1 in *Theoretical Nuclear Physics.* New York: Wiley, pp. 796–799, 1952.

Bohren, C. F. and Huffman, D. R. *Absorption and Scattering of Light by Small Particles.* New York: Wiley, 1983.

Hill, E. H. "The Theory of Vector Spherical Harmonics." *Amer. J. Phys.* **22**, 211–214, 1954.

Jackson, J. D. *Classical Electrodynamics, 2nd ed.* New York: Wiley, pp. 744–755, 1975.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part II.* New York: McGraw-Hill, pp. 1898–1901, 1953.

Vector Transformation Law

The set of n quantities v_j are components of an n-D VECTOR $\mathbf{v}$ IFF, under ROTATION,

$$v_i' = a_{ij}v_j$$

for $i = 1, 2, \ldots, n$. The DIRECTION COSINES between x_i' and x_j are

$$a_{ij} \equiv \frac{\partial x_i'}{\partial x_j} = \frac{\partial x_j}{\partial x_i'}.$$

They satisfy the orthogonality condition

$$a_{ij}a_{ik} = \frac{\partial x_j}{\partial x_i'}\frac{\partial x_i'}{\partial x_k} = \frac{\partial x_j}{\partial x_k} = \delta_{jk},$$

where δ_{jk} is the KRONECKER DELTA.

see also TENSOR, VECTOR

Vector Triple Product

The triple product can be written in terms of the Levi-Civita Symbol ϵ_{ijk} as

$$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})=\epsilon_{ijk}A^iB^jC^k. \tag{1}$$

The BAC-CAB Rule can be written in the form

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C})=\mathbf{B}(\mathbf{A}\cdot\mathbf{C})-\mathbf{C}(\mathbf{A}\cdot\mathbf{B}) \tag{2}$$

$$\begin{aligned}(\mathbf{A}\times\mathbf{B})\times\mathbf{C}&=-\mathbf{C}\times(\mathbf{A}\times\mathbf{B})\\&=-\mathbf{A}(\mathbf{B}\cdot\mathbf{C})+\mathbf{B}(\mathbf{A}\cdot\mathbf{C}).\end{aligned} \tag{3}$$

Addition identities are

$$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})=\mathbf{B}\cdot(\mathbf{C}\times\mathbf{A})=\mathbf{C}\cdot(\mathbf{A}\times\mathbf{B}) \tag{4}$$

$$[\mathbf{A},\mathbf{B},\mathbf{C}]\mathbf{D}=[\mathbf{D},\mathbf{B},\mathbf{C}]\mathbf{A}+[\mathbf{A},\mathbf{D},\mathbf{C}]\mathbf{B}+[\mathbf{A},\mathbf{B},\mathbf{D}]\mathbf{C} \tag{5}$$

$$[\mathbf{q},\mathbf{q}',\mathbf{q}''][\mathbf{r},\mathbf{r}',\mathbf{r}'']=\begin{vmatrix}\mathbf{q}\cdot\mathbf{r} & \mathbf{q}\cdot\mathbf{r}' & \mathbf{q}\cdot\mathbf{r}''\\ \mathbf{q}'\cdot\mathbf{r} & \mathbf{q}'\cdot\mathbf{r}' & \mathbf{q}'\cdot\mathbf{r}''\\ \mathbf{q}''\cdot\mathbf{r} & \mathbf{q}''\cdot\mathbf{r}' & \mathbf{q}''\cdot\mathbf{r}''\end{vmatrix}. \tag{6}$$

see also BAC-CAB Rule, Cross Product, Dot Product, Levi-Civita Symbol, Scalar Triple Product, Vector Quadruple Product

References

Arfken, G. "Triple Scalar Product, Triple Vector Product." §1.5 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 26–33, 1985.

Vee

The symbol $\vee$ variously means "disjunction" (in Logic) or "join" (for a Lattice).

see also Wedge

Velocity

$$\mathbf{v}\equiv\frac{d\mathbf{r}}{dt},$$

where $\mathbf{r}$ is the Position Vector and d/dt is the derivative with respect to time. Expressed in terms of the Arc Length,

$$\mathbf{v}=\frac{ds}{dt}\hat{\mathbf{T}},$$

where $\hat{\mathbf{T}}$ is the unit Tangent Vector, so the Speed (which is the magnitude of the velocity) is

$$v\equiv|\mathbf{v}|=\frac{ds}{dt}=|\mathbf{r}'(t)|.$$

see also Angular Velocity, Position Vector, Speed

Venn Diagram

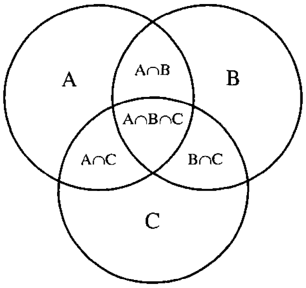

The simplest Venn diagram consists of three symmetrically placed mutually intersecting Circles. It is used in Logic theory to represent collections of sets. The region of intersection of the three Circles $A\cap B\cap C$, in the special case of the center of each being located at the intersection of the other two, is called a Reuleaux Triangle.

In general, an order n Venn diagram is a collection of n simple closed curves in the Plane such that

1. The curves partition the Plane into 2^n connected regions, and
2. Each Subset S of $\{1, 2, \ldots, n\}$ corresponds to a unique region formed by the intersection of the interiors of the curves in S (Ruskey).

see also Circle, Flower of Life, Lens, Magic Circles, Reuleaux Triangle, Seed of Life

References

Cundy, H. and Rollett, A. *Mathematical Models, 3rd ed.* Stradbroke, England: Tarquin Pub., pp. 255–256, 1989.

Ruskey, F. "A Survey of Venn Diagrams." *Elec. J. Combin.* 4, DS#5, 1997. `http://www.combinatorics.org/Surveys/ds5/VennEJC.html`.

Ruskey, F. "Venn Diagrams." `http://sue.csc.uvic.ca/~cos/inf/comb/SubsetInfo.html#Venn`.

Verging Construction

see Neusis Construction

Verhulst Model

see Logistic Map

Veronese Surface

A smooth 2-D surface given by embedding the Projective Plane into projective 5-space by the homogeneous parametric equations

$$v(x,y,z)=(x^2,y^2,z^2,xy,xz,yz).$$

The surface can be projected smoothly into 4-space, but all 3-D projections have singularities (Coffman). The projections of these surfaces in 3-D are called Steiner Surfaces. The Volume of the Veronese surface is $2\pi^2$.

see also Steiner Surface

References

Coffman, A. "Steiner Surfaces." `http://www.ipfw.edu/math/Coffman/steinersurface.html`.

Veronese Variety

see Veronese Surface

Versed Sine

see Versine

Versiera

see Witch of Agnesi

Versine

$$\operatorname{vers}(z) \equiv 1 - \cos z,$$

where $\cos z$ is the Cosine. Using a trigonometric identity, the versine is equal to

$$\operatorname{vers}(z) = 2\sin^2(\tfrac{1}{2}z).$$

see also Cosine, Coversine, Exsecant, Haversine

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 78, 1972.

Vertex Angle

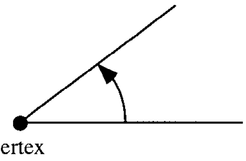

The point about which an Angle is measured is called the angle's vertex, and the angle associated with a given vertex is called the vertex angle.

see also Angle

Vertex Coloring

Brelaz's Heuristic Algorithm can be used to find a good, but not necessarily minimal, Vertex coloring of a Graph.

see also Brelaz's Heuristic Algorithm, Coloring

Vertex Connectivity

The minimum number of Vertices whose deletion from a Graph disconnects it.

see also Edge Connectivity

Vertex Cover

see Hitting Set

Vertex Degree

The degree of a Vertex of a Graph is the number of Edges which touch the Vertex, also called the Local Degree. The Vertex degree of a point A in a Graph, denoted $\rho(A)$, satisfies

$$\sum_{i=1}^{n} \rho(A_i) = 2E,$$

where E is the total number of Edges. Directed Graphs have two types of degrees, known as the Indegree and the Outdegree.

see also Directed Graph, Indegree, Local Degree, Outdegree

Vertex Enumeration

A Convex Polyhedron is defined as the set of solutions to a system of linear inequalities

$$\mathsf{m}\mathbf{x} \leq \mathbf{b},$$

where m is a Real $s \times d$ Matrix and $\mathbf{b}$ is a Real s-Vector. Given m and $\mathbf{b}$, vertex enumeration is the determination of the polyhedron's Vertices.

see also Convex Polyhedron, Polyhedron

References

Avis, D. and Fukuda, K. "A Pivoting Algorithm for Convex Hulls and Vertex Enumeration of Arrangements and Polyhedra." In *Proceedings of the 7th ACM Symposium on Computational Geometry, North Conway, NH, 1991,* pp. 98–104, 1991.

Fukada, K. and Mizukosh, I. "Vertex Enumeration Package for Convex Polytopes and Arrangements, Version 0.41 Beta." `http://www.mathsource.com/cgi-bin/MathSource/Applications/Mathematics/0202-633`.

Vertex Figure

The line joining the Midpoints of adjacent sides in a Polygon is called the polygon's vertex figure. For a regular n-gon with side length s,

$$v = s\cos\left(\frac{\pi}{n}\right).$$

For a Polyhedron, the faces that join at a Vertex form a solid angle whose section by the plane is the vertex figure.

see also Truncation

Vertex (Graph)

A point of a Graph, also called a Node.

see also Edge (Graph), Null Graph, Tait Coloring, Tait Cycle, Tait's Hamiltonian Graph Conjecture, Vertex (Polygon)

Vertex (Parabola)

For a Parabola oriented vertically and opening upwards, the vertex is the point where the curve reaches a minimum.

Vertex (Polygon)

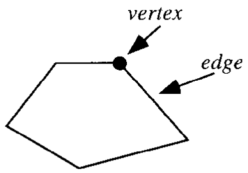

A point at which two EDGES of a POLYGON meet.

see also PRINCIPAL VERTEX, VERTEX (GRAPH), VERTEX (POLYHEDRON)

Vertex (Polyhedron)

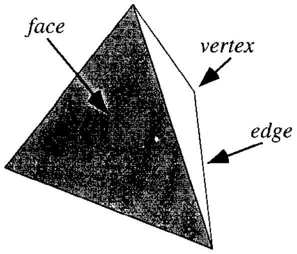

A point at which three of more EDGES of a POLYHEDRON meet. The concept can also be generalized to a POLYTOPE.

see also VERTEX (GRAPH), VERTEX (POLYGON)

Vertex (Polytope)

The vertex of a POLYTOPE is a point where edges of the POLYTOPE meet.

Vertical

Oriented in an up-down position.

see also HORIZONTAL

Vertical-Horizontal Illusion

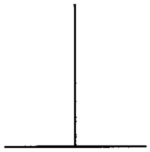

The HORIZONTAL line segment in the above figure appears to be shorter than the VERTICAL line segment, despite the fact that it has the same length.

see also ILLUSION, MÜLLER-LYER ILLUSION, POGGENDORFF ILLUSION, PONZO'S ILLUSION

References

Fineman, M. *The Nature of Visual Illusion.* New York: Dover, p. 153, 1996.

Vertical Perspective Projection

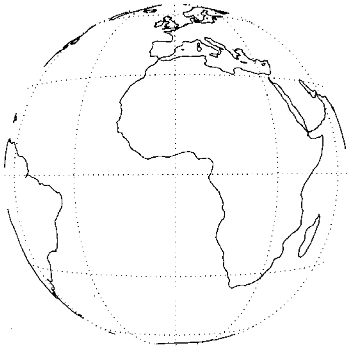

A MAP PROJECTION given by the transformation equations

$$x = k' \cos\phi \sin(\lambda - \lambda_0) \tag{1}$$

$$y = k'[\cos\phi_1 \sin\phi - \sin\phi_1 \cos\phi \cos(\lambda - \lambda_0)], \tag{2}$$

where P is the distance of the point of perspective in units of SPHERE RADII and

$$k' = \frac{P-1}{P - \cos c} \tag{3}$$

$$\cos c = \sin\phi_1 \sin\phi + \cos\phi_1 \cos\phi \cos(\lambda - \lambda_0). \tag{4}$$

References

Snyder, J. P. *Map Projections—A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 173–178, 1987.

Vertical Tangent

A function $f(x)$ has a vertical tangent line at x_0 if f is continuous at x_0 and

$$\lim_{x \to x_0} f'(x) = \pm\infty.$$

Vesica Piscis

see LENS

Vibration Problem

Solution of a system of second-order homogeneous ordinary differential equations with constant COEFFICIENTS of the form

$$\frac{d^2\mathbf{x}}{dt^2} + \mathsf{B}\mathbf{x} = 0,$$

where B is a POSITIVE DEFINITE MATRIX. To solve the vibration problem,

1. Solve the CHARACTERISTIC EQUATION of B to get EIGENVALUES $\lambda_1, \ldots, \lambda_n$. Define $\omega_i \equiv \sqrt{\lambda_i}$.
2. Compute the corresponding EIGENVECTORS $\mathbf{e}_1, \ldots, \mathbf{e}_n$.
3. The normal modes of oscillation are given by $\mathbf{x}_1 = A_1 \sin(\omega_1 t + \alpha_1)\mathbf{e}_1, \ldots, \mathbf{x}_n = A_n \sin(\omega_n t + \alpha_n)\mathbf{e}_n$, where $A_1, \ldots, A_n$ and $\alpha_1, \ldots, \alpha_n$ are arbitrary constants.
4. The general solution is $\mathbf{x} = \sum_{i=1}^n \mathbf{x}_i$.

Vickery Auction

An AUCTION in which the highest bidder wins but pays only the second-highest bid. This variation over the normal bidding procedure is supposed to encourage bidders to bid the largest amount they are willing to pay.

see also AUCTION

Viergruppe

The mathematical group $Z_4 \otimes Z_4$, also denoted D_2. Its multiplication table is

V	I	V_1	V_2	V_3
I	V_1	V_2	V_3	V_4
V_1	V_1	I	V_3	V_2
V_2	V_2	V_3	I	V_1
V_3	V_3	V_2	V_1	I

see also DIHEDRAL GROUP, FINITE GROUP—Z_4

Vieta's Substitution

The substitution of

$$x = w - \frac{p}{3w}$$

into the standard form CUBIC EQUATION

$$x^3 + px = q,$$

which reduces the cubic to a QUADRATIC EQUATION in w^3,

$$(w^3)^2 - \tfrac{1}{27}p^3(w^3) - q = 0.$$

see also CUBIC EQUATION

Vigesimal

The base-20 notational system for representing REAL NUMBERS. The digits used to represent numbers using vigesimal NOTATION are 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F, G, H, I, and J. A base-20 number system was used by the Aztecs and Mayans. The Mayans compiled extensive observations of planetary positions in base-20 notation.

see also BASE (NUMBER), BINARY, DECIMAL, HEXADECIMAL, OCTAL, QUATERNARY, TERNARY

References

⊛ Weisstein, E. W. "Bases." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Bases.m`.

Vigintillion

In the American system, 10^{63}.

see also LARGE NUMBER

Villarceau Circles

Given an arbitrary point on a TORUS, four CIRCLES can be drawn through it. The first is in the plane of the torus and the second is PERPENDICULAR to it. The third and fourth CIRCLES are called Villarceau circles.

see also TORUS

References

Melzak, Z. A. *Invitation to Geometry.* New York: Wiley, pp. 63–72, 1983.

Villarceau, M. "Théorème sur le tore." *Nouv. Ann. Math.* **7**, 345–347, 1848.

Vinculum

A horizontal line placed above multiple quantities to indicate that they form a unit. It is most commonly used to denote ROOTS ($\sqrt{12345}$) and repeating decimals ($0.\overline{111}$).

Vinogradov's Theorem

Every sufficiently large ODD number is a sum of three PRIMES. Proved in 1937.

see also GOLDBACH CONJECTURE

Virtual Group

see GROUPOID

Visibility

see VISIBLE POINT

Visible Point

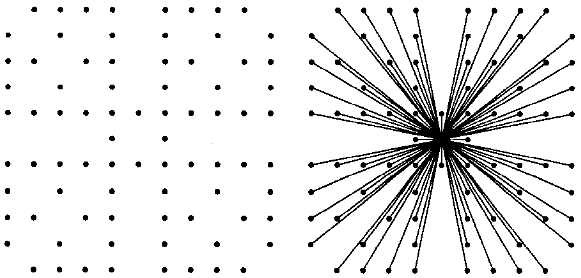

Two LATTICE POINTS (x, y) and (x', y') are mutually visible if the line segment joining them contains no further LATTICE POINTS. This corresponds to the requirement that $(x' - x, y' - y) = 1$, where (m, n) denotes the GREATEST COMMON DIVISOR. The plots above show the first few points visible from the ORIGIN.

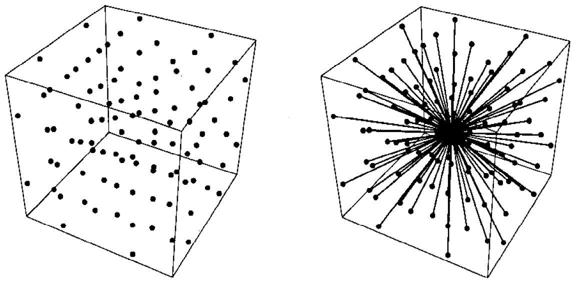

If a LATTICE POINT is selected at random in 2-D, the probability that it is visible from the origin is $6/\pi^2$. This is also the probability that two INTEGERS picked at random are RELATIVELY PRIME. If a LATTICE POINT is picked at random in n-D, the probability that it is visible from the ORIGIN is $1/\zeta(n)$, where $\zeta(n)$ is the RIEMANN ZETA FUNCTION.

An invisible figure is a POLYGON all of whose corners are invisible. There are invisible sets of every finite shape. The lower left-hand corner of the invisible squares with smallest x coordinate of AREAS 2 and 3 are (14, 20) and (104, 6200).

see also LATTICE POINT, ORCHARD VISIBILITY PROBLEM, RIEMANN ZETA FUNCTION

References
Apostol, T. §3.8 in *Introduction to Analytic Number Theory.* New York: Springer-Verlag, 1976.
Baake, M.; Grimm, U.; and Warrington, D. H. "Some Remarks on the Visible Points of a Lattice." *J. Phys. A: Math. General* **27**, 2669–2674, 1994.
Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Feb. 1972.
Herzog, F. and Stewart, B. M. "Patterns of Visible and Nonvisible Lattice Points." *Amer. Math. Monthly* **78**, 487–496, 1971.
Mosseri, R. "Visible Points in a Lattice." *J. Phys. A: Math. Gen.* **25**, L25–L29, 1992.
Schroeder, M. R. "A Simple Function and Its Fourier Transform." *Math. Intell.* **4**, 158–161, 1982.
Schroeder, M. R. *Number Theory in Science and Communication, 2nd ed.* New York: Springer-Verlag, 1990

Visible Point Vector Identity

A set of identities involving n-D visible lattice points was discovered by Campbell (1994). Examples include

$$\prod_{\substack{(a,b)=1\\ a\geq 0,b\leq 1}} (1-y^a z^b)^{-1/b} = (1-z)^{-1/(1-y)}$$

for $|yz|, |z| < 1$ and

$$\prod_{\substack{(a,b,c)=1\\ a,b\geq 0,c\leq 1}} (1-x^a y^b z^c)^{-1/c} = (1-z)^{-1/[(1-x)(1-y)]}$$

for $|xyz|, |xz|, |yz|, |z| < 1$.

References
Campbell, G. B. "Infinite Products Over Visible Lattice Points." *Internat. J. Math. Math. Sci.* **17**, 637–654, 1994.
Campbell, G. B. "Visible Point Vector Identities." `http://www.geocities.com/CapeCanaveral/Launchpad/9416/vpv.html`.

Vitali's Convergence Theorem

Let $f_n(z)$ be a sequence of functions, each regular in a region D, let $|f_n(z)| \leq M$ for every n and z in D, and let $f_n(z)$ tend to a limit as $n \to \infty$ at a set of points having a LIMIT POINT inside D. Then $f_n(z)$ tends uniformly to a limit in any region bounded by a contour interior to D, the limit therefore being an analytic function of z.

see also MONTEL'S THEOREM

References
Titchmarsh, E. C. *The Theory of Functions, 2nd ed.* Oxford, England: Oxford University Press, p. 168, 1960.

Viviani's Curve

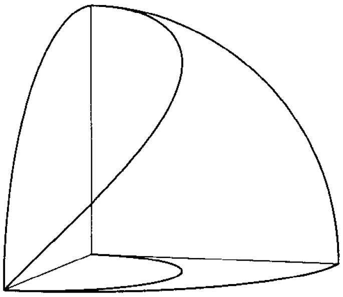

The SPACE CURVE giving the intersection of the CYLINDER

$$(x-a)^2 + y^2 = a^2 \tag{1}$$

and the SPHERE

$$x^2 + y^2 + z^2 = 4^2. \tag{2}$$

It is given by the parametric equations

$$x = a(1+\cos t) \tag{3}$$
$$y = a\sin t \tag{4}$$
$$z = 2a\sin(\tfrac{1}{2}t). \tag{5}$$

The CURVATURE and TORSION are given by

$$\kappa(t) = \frac{\sqrt{13+3\cos t}}{a(3+\cos t)^{3/2}} \tag{6}$$
$$\tau(t) = \frac{6\cos(\frac{1}{2}t)}{a(13+3\cos t)}. \tag{7}$$

see also CYLINDER, SPHERE, STEINMETZ SOLID

References
Gray, A. "Viviani's Curve." §7.6 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 140–142, 1993.
von Seggern, D. *CRC Standard Curves and Surfaces.* Boca Raton, FL: CRC Press, p. 270, 1993.

Viviani's Theorem

For a point P inside an EQUILATERAL TRIANGLE ΔABC, the sum of the perpendiculars p_i from P to the sides of the TRIANGLE is equal to the ALTITUDE h. This result is simply proved as follows,

$$\Delta ABC = \Delta PBC + \Delta PCA + \Delta PAB. \tag{1}$$

With s the side length,

$$\tfrac{1}{2}sh = \tfrac{1}{2}sp_a + \tfrac{1}{2}sp_b + \tfrac{1}{2}sp_c, \tag{2}$$

so

$$h = p_a + p_b + p_c. \tag{3}$$

see also ALTITUDE, EQUILATERAL TRIANGLE

Vojta's Conjecture

A conjecture which treats the heights of points relative to a canonical class of a curve defined over the INTEGERS.

References

Cox, D. A. "Introduction to Fermat's Last Theorem." *Amer. Math. Monthly* **101**, 3–14, 1994.

Volterra Integral Equation of the First Kind

An INTEGRAL EQUATION of the form

$$f(x)=\int_a^x k(x,t)\phi(t)\,dt.$$

see also FREDHOLM INTEGRAL EQUATION OF THE FIRST KIND, FREDHOLM INTEGRAL EQUATION OF THE SECOND KIND, INTEGRAL EQUATION, VOLTERRA INTEGRAL EQUATION OF THE SECOND KIND

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 865, 1985.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Volterra Equations." §18.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 786–788, 1992.

Volterra Integral Equation of the Second Kind

An INTEGRAL EQUATION of the form

$$\phi(x)=f(x)+\int_a^x k(x,t)\phi(t)\,dt.$$

see also FREDHOLM INTEGRAL EQUATION OF THE FIRST KIND, FREDHOLM INTEGRAL EQUATION OF THE SECOND KIND, INTEGRAL EQUATION, VOLTERRA INTEGRAL EQUATION OF THE FIRST KIND

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, p. 865, 1985.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Volterra Equations." §18.2 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 786–788, 1992.

Volume

The volume of a solid body is the amount of "space" it occupies. Volume has units of LENGTH cubed (i.e., cm^3, m^3, in^3, etc.) For example, the volume of a box (RECTANGULAR PARALLELEPIPED) of LENGTH L, WIDTH W, and HEIGHT H is given by

$$V=L\times W\times H.$$

The volume can also be computed for irregularly-shaped and curved solids such as the CYLINDER and CUBE. The volume of a SURFACE OF REVOLUTION is particularly simple to compute due to its symmetry.

The following table gives volumes for some common SURFACES. Here r denotes the RADIUS, h the height, A the base AREA, and s the SLANT HEIGHT (Beyer 1987).

Surface	V
cone	$\frac{1}{3}\pi r^2 h$
conical frustum	$\frac{1}{3}\pi h({R_1}^2+{R_2}^2+R_1R_2)$
cube	a^3
cylinder	$\pi r^2 h$
ellipsoid	$\frac{4}{3}\pi abc$
oblate spheroid	$\frac{4}{3}\pi a^2 b$
prolate spheroid	$\frac{4}{3}\pi ab^2$
pyramid	$\frac{1}{3}Ah$
pyramidal frustum	$\frac{1}{3}h(A_1+A_2+\sqrt{A_1A_2})$
sphere	$\frac{4}{3}\pi r^3$
spherical sector	$\frac{2}{3}\pi r^2 h$
spherical segment	$\frac{1}{3}\pi h^2 r(3r-h)$
torus	$2\pi^2 Rr^2$

Even simple SURFACES can display surprisingly counterintuitive properties. For instance, the SURFACE OF REVOLUTION of $y=1/x$ around the x-axis for $x\geq 1$ is called GABRIEL'S HORN, and has finite volume, but infinite SURFACE AREA.

The generalization of volume to n DIMENSIONS for $n\geq 4$ is known as CONTENT.

see also ARC LENGTH, AREA, CONTENT, HEIGHT, LENGTH (SIZE), SURFACE AREA, SURFACE OF REVOLUTION, VOLUME ELEMENT, WIDTH (SIZE)

References

Beyer, W. H. *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 127–132, 1987.

Volume Element

A volume element is the differential element dV whose VOLUME INTEGRAL over some range in a given coordinate system gives the VOLUME of a solid,

$$V=\iiint_G dx\,dy\,dz. \tag{1}$$

In $\mathbb{R}^n$, the volume of the infinitesimal n-HYPERCUBE bounded by dx_1, ..., dx_n has volume given by the WEDGE PRODUCT

$$dV=dx_1\wedge\ldots\wedge dx_n \tag{2}$$

(Gray 1993).

The use of the antisymmetric WEDGE PRODUCT instead of the symmetric product $dx_1\ldots dx_n$ is a technical refinement often omitted in informal usage. Dropping the wedges, the volume element for CURVILINEAR COORDINATES in $\mathbb{R}^3$ is given by

$$\begin{aligned} dV &= |(h_1\hat{\mathbf{u}}_1\,du_1)\cdot(h_2\hat{\mathbf{u}}_2\,du_2)\times(h_3\hat{\mathbf{u}}_3\,du_3)| && (3)\\ &= h_1h_2h_3\,du_1\,du_2\,du_3 && (4)\\ &= \left|\frac{\partial\mathbf{r}}{\partial u_1}\cdot\frac{\partial\mathbf{r}}{\partial u_2}\times\frac{\partial\mathbf{r}}{\partial u_3}\right| du_1\,du_2\,du_3 && (5)\\ &= \begin{vmatrix} \frac{\partial x}{\partial u_1} & \frac{\partial x}{\partial u_2} & \frac{\partial x}{\partial u_3}\\ \frac{\partial y}{\partial u_1} & \frac{\partial y}{\partial u_2} & \frac{\partial y}{\partial u_3}\\ \frac{\partial z}{\partial u_1} & \frac{\partial z}{\partial u_2} & \frac{\partial z}{\partial u_3} \end{vmatrix} du_1\,du_2\,du_3 && (6)\\ &= \left|\frac{\partial(x,y,z)}{\partial(u_1,u_2,u_3)}\right| du_1\,du_2\,du_3, && (7) \end{aligned}$$

where the latter is the JACOBIAN and the h_i are SCALE FACTORS.

see also AREA ELEMENT, JACOBIAN, LINE ELEMENT, RIEMANNIAN METRIC, SCALE FACTOR, SURFACE INTEGRAL, VOLUME INTEGRAL

References

Gray, A. "Isometries of Surfaces." §13.2 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 255–258, 1993.

Volume Integral

A triple integral over three coordinates giving the VOLUME within some region R,

$$V = \iiint_G dx\,dy\,dz.$$

see also INTEGRAL, LINE INTEGRAL, MULTIPLE INTEGRAL, SURFACE INTEGRAL, VOLUME, VOLUME ELEMENT

von Aubel's Theorem

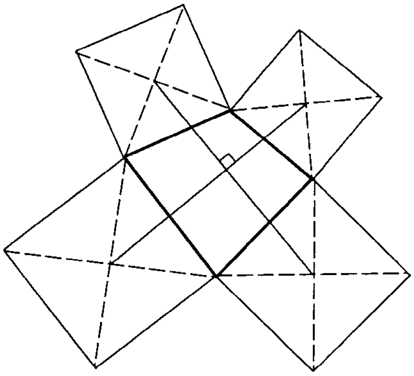

Given an arbitrary QUADRILATERAL, place a SQUARE outwardly on each side, and connect the centers of opposite SQUARES. Then the two lines are of equal length and cross at a RIGHT ANGLE.

see also QUADRILATERAL, RIGHT ANGLE, SQUARE

References

Kitchen, E. "Dörrie Tiles and Related Miniatures." *Math. Mag.* **67**, 128–130, 1994.

von Dyck's Theorem

Let a GROUP G have a presentation

$$G = (x_1,\ldots,x_n | r_j(x_1,\ldots,x_n), j \in J)$$

so that $G = F/R$, where F is the FREE GROUP with basis $\{x_1,\ldots,x_n\}$ and R is the NORMAL SUBGROUP generated by the r_j. If H is a GROUP with $H = \langle y_1,\ldots,y_n\rangle$ and if $r_j(y_1,\ldots,y_n) = 1$ for all j, then there is a surjective homomorphism $G \to H$ with $x_i \mapsto y_i$ for all i.

see also FREE GROUP, NORMAL SUBGROUP

References

Rotman, J. J. *An Introduction to the Theory of Groups, 4th ed.* New York: Springer-Verlag, p. 346, 1995.

von Mangoldt Function

see MANGOLDT FUNCTION

von Neumann Algebra

A GROUP "with bells and whistles." It was while studying von Neumann algebras that Jones discovered the amazing and highly unexpected connections with KNOT THEORY which led to the formulation of the JONES POLYNOMIAL.

References

Iyanaga, S. and Kawada, Y. (Eds.). "Von Neumann Algebras." §430 in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 1358–1363, 1980.

von Staudt-Clausen Theorem

$$B_{2n} = A_n - \sum_{\substack{p_k\\ p_k-1|2n}} \frac{1}{p_k},$$

where B_{2n} is a BERNOULLI NUMBER, A_n is an INTEGER, and the p_ks are the PRIMES satisfying $p_k - 1|2k$. For example, for $k = 1$, the primes included in the sum are 2 and 3, since $(2-1)|2$ and $(3-1)|2$. Similarly, for $k = 6$, the included primes are (2, 3, 5, 7, 13), since (1, 2, 3, 6, 12) divide $12 = 2 \cdot 6$. The first few values of A_n for $n = 1, 2, \ldots$ are 1, 1, 1, 1, 1, 1, 2, -6, 56, -528, ... (Sloane's A000164).

The theorem was rediscovered by Ramanujan (Hardy 1959, p. 11) and can be proved using p-ADIC NUMBERS.

see also BERNOULLI NUMBER, p-ADIC NUMBER

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, p. 109, 1996.

Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by His Life and Work, 3rd ed.* New York: Chelsea, 1959.

Hardy, G. H. and Wright, E. M. "The Theorem of von Staudt" and "Proof of von Staudt's Theorem." §7.9–7.10 in *An Introduction to the Theory of Numbers, 5th ed.* Oxford, England: Clarendon Press, pp. 90–93, 1979.

Sloane, N. J. A. Sequence A000146/M1717 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Staudt. "Beweis eines Lehrsatzes, die Bernoullischen Zahlen betreffend." *J. reine angew. Math.* **21**, 372–374, 1840.

von Staudt Theorem

see VON STAUDT-CLAUSEN THEOREM

Voronoi Cell

The generalization of a VORONOI POLYGON to n-D, for $n > 2$.

Voronoi Diagram

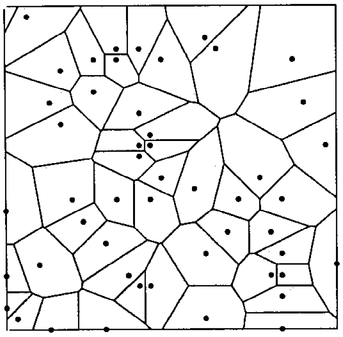

The partitioning of a plane with n points into n convex POLYGONS such that each POLYGON contains exactly one point and every point in a given POLYGON is closer to its central point than to any other. A Voronoi diagram is sometimes also known as a DIRICHLET TESSELLATION. The cells are called DIRICHLET REGIONS, THIESSEN POLYTOPES, or VORONOI POLYGONS.

see also DELAUNAY TRIANGULATION, MEDIAL AXIS, VORONOI POLYGON

References

Eppstein, D. "Nearest Neighbors and Voronoi Diagrams." `http://www.ics.uci.edu/~eppstein/junkyard/nn.html`.

Voronoi Polygon

A POLYGON whose interior consists of all points in the plane which are closer to a particular LATTICE POINT than to any other. The generalization to n-D is called a DIRICHLET REGION, THIESSEN POLYTOPE, or VORONOI CELL.

Voting

It is possible to conduct a secret ballot even if the votes are sent in to a central polling station (Lipton and Widgerson, Honsberger 1985).

see also ARROW'S PARADOX, BALLOT PROBLEM, MAY'S THEOREM, QUOTA SYSTEM, SOCIAL CHOICE THEORY

References

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 157–162, 1985.

Lipton, R. G.; and Widgerson, A. "Multi-Party Cryptographic Protocols."

VR Number

A "visual representation" number which is a sum of some simple function of its digits. For example,

$$1233 = 12^2 + 33^2$$

$$2661653 = 1653^2 - 266^2$$

$$221859 = 22^3 + 18^3 + 59^3$$

$$40585 + 4! + 0! + 5! + 8! + 5!$$

$$148349 = !1+!4+!8+!3+!4+!9$$

$$4913 = (4 + 9 + 1 + 3)^3$$

are all VR numbers given by Madachy (1979).

References

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 165–171, 1979.

Vulgar Series

see FAREY SERIES

W

W2-Constant

$$W_2 = 1.529954037\ldots.$$

References
Plouffe, S. "W2 Constant." `http://lacim.uqam.ca/piDATA/w2.txt`.

W-Function

see LAMBERT'S W-FUNCTION

Wada Basin

A BASIN OF ATTRACTION in which every point on the common boundary of that basin and another basin is also a boundary of a third basin. In other words, no matter how closely a boundary point is zoomed into, all three basins appear in the picture.

see also BASIN OF ATTRACTION

References
Nusse, H. E. and Yorke, J. A. "Basins of Attraction." *Science* **271**, 1376–1380, 1996.

Walk

A sequence of VERTICES and EDGES such that the VERTICES and EDGES are adjacent. A walk is therefore equivalent to a graph CYCLE, but with the VERTICES along the walk enumerated as well as the EDGES.

see also CIRCUIT, CYCLE (GRAPH), PATH, RANDOM WALK

Wallace-Bolyai-Gerwein Theorem

Two POLYGONS are congruent by DISSECTION IFF they have the same AREA. In particular, any POLYGON is congruent by DISSECTION to a SQUARE of the same AREA. Laczkovich (1988) also proved that a CIRCLE is congruent by DISSECTION to a SQUARE (furthermore, the DISSECTION can be accomplished using TRANSLATIONS only).

see also DISSECTION

References
Klee, V. and Wagon, S. *Old and New Unsolved Problems in Plane Geometry and Number Theory.* Washington, DC: Math. Assoc. Amer., pp. 50–51, 1991.
Laczkovich, M. "Von Neumann's Paradox with Translation." *Fund. Math.* **131**, 1–12, 1988.

Wallace-Simson Line

see SIMSON LINE

Wallis's Conical Edge

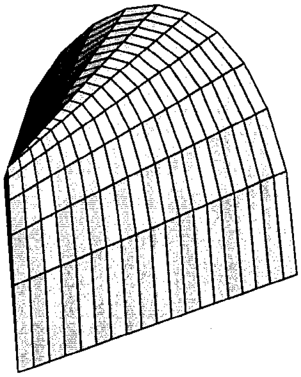

The RIGHT CONOID surface given by the parametric equations

$$\begin{aligned} x(u,v) &= v\cos u \\ y(u,v) &= v\sin u \\ z(u,v) &= c\sqrt{a^2 - b^2\cos^2 u}. \end{aligned}$$

see also RIGHT CONOID

References
Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 354–355, 1993.

Wallis Cosine Formula

$$\int_0^{\pi/2} \cos^n x\,dx = \begin{cases} \frac{\pi}{2}\frac{1\cdot3\cdot5\cdots(n-1)}{2\cdot4\cdot6\cdots n} & \text{for } n = 2, 4, \ldots \\ \frac{2\cdot4\cdot6\cdots(n-1)}{1\cdot3\cdot5\cdots n} & \text{for } n = 3, 5, \ldots. \end{cases}$$

see also WALLIS FORMULA, WALLIS SINE FORMULA

Wallis Formula

The Wallis formula follows from the INFINITE PRODUCT representation of the SINE

$$\sin x = x\prod_{n=1}^\infty \left(1 - \frac{x^2}{\pi^2 n^2}\right). \tag{1}$$

Taking $x = \pi/2$ gives

$$1 = \frac{\pi}{2}\prod_{n=1}^\infty \left[1 - \frac{1}{(2n)^2}\right] = \frac{\pi}{2}\prod_{n=1}^\infty \left[\frac{(2n)^2 - 1}{(2n)^2}\right], \tag{2}$$

so

$$\frac{\pi}{2} = \prod_{n=1}^\infty \left[\frac{(2n)^2}{(2n-1)(2n+1)}\right] = \frac{2\cdot2}{1\cdot3}\frac{4\cdot4}{3\cdot5}\frac{6\cdot6}{5\cdot7}\cdots. \tag{3}$$

A derivation due to Y. L. Yung uses the RIEMANN ZETA FUNCTION. Define

$$F(s) \equiv -\operatorname{Li}_s(-1) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n^s}$$

$$= (1-2^{1-s})\zeta(s) \qquad (4)$$

$$F'(s) = \sum_{n=1}^{\infty} \frac{(-1)^n \ln n}{n^s}, \qquad (5)$$

so

$$F'(0) = \sum_{n=1}^{\infty} (-1)^n \ln n = -\ln 1 + \ln 2 - \ln 3 + \ldots$$

$$= \ln\left(\frac{2\cdot 4\cdot 6\cdots}{1\cdot 3\cdot 5\cdots}\right). \qquad (6)$$

Taking the derivative of the zeta function expression gives

$$\frac{d}{ds}(1-2^{1-s})\zeta(s) = 2^{1-s}(\ln 2)\zeta(s) + (1-2^{1-s})\zeta'(s) \quad (7)$$

$$\left[\frac{d}{ds}(1-2^{1-s})\zeta(s)\right]_{s=0} = -\ln 2 - \zeta'(0)$$

$$= -\ln 2 + \tfrac{1}{2}\ln(2\pi) = \ln\left(\frac{\sqrt{2\pi}}{2}\right) = \ln\left(\sqrt{\frac{\pi}{2}}\right). \quad (8)$$

Equating and squaring then gives the Wallis formula, which can also be expressed

$$\frac{\pi}{2} = \left[4^{\zeta(0)} e^{-\zeta'(0)}\right]^2. \qquad (9)$$

The q-ANALOG of the Wallis formula for $q = 2$ is

$$\prod_{k=1}^{\infty} (1-q^{-k})^{-1} = 3.4627466194\ldots \qquad (10)$$

(Finch).

see also WALLIS COSINE FORMULA, WALLIS SINE FORMULA

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 258, 1972.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/dig/dig.html`.

Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 2, 2nd ed.* Princeton, NJ: Van Nostrand, pp. 63–64, 1951.

Wallis's Problem

Find solutions to $\sigma(x^2) = \sigma(y^2)$ other than $(x, y) = (4, 5)$, where σ is the DIVISOR FUNCTION.

see also FERMAT'S SIGMA PROBLEM

Wallis Sieve

A compact set W_∞ with AREA

$$\mu(W_\infty) = \frac{8}{9}\frac{24}{25}\frac{48}{49}\cdots = \frac{\pi}{4}$$

created by punching a square hole of length 1/3 in the center of a square. In each of the eight squares remaining, punch out another hole of length $1/(3\cdot 5)$, and so on.

Wallis Sine Formula

$$\int_0^{\pi/2} \sin^n x\,dx = \begin{cases} \frac{\pi}{2}\frac{1\cdot 3\cdot 5\cdots(n-1)}{2\cdot 4\cdot 6\cdots n} & \text{for } n = 2,\ 4,\ \ldots \\ \frac{2\cdot 4\cdot 6\cdots(n-1)}{1\cdot 3\cdot 5\cdots n} & \text{for } n = 3,\ 5,\ \ldots. \end{cases}$$

see also WALLIS COSINE FORMULA, WALLIS FORMULA

Wallpaper Groups

The 17 PLANE SYMMETRY GROUPS. Their symbols are p1, p2, pm, pg, cm, pmm, pmg, pgg, cmm, p4, p4m, p4g, p3, p31m, p3m1, p6, and p6m. For a description of the symmetry elements present in each space group, see Coxeter (1969, p. 413).

References

Coxeter, H. S. M. *Introduction to Geometry, 2nd ed.* New York: Wiley, 1969.

Hilbert, D. and Cohn-Vossen, S. *Geometry and the Imagination.* New York: Chelsea, 1952.

Joyce, D. E. "Wallpaper Groups (Plane Symmetry Groups)." `http://aleph0.clarku.edu/~djoyce/wallpaper/`.

Lee, X. "The Discontinuous Groups of Rotation and Translation in the Plane." `http://www.best.com/~xah/Wallpaper_dir/c0_WallPaper.html`.

Schattschneider, D. "The Plane Symmetry Groups: Their Recognition and Notation." *Amer. Math. Monthly* **85**, 439–450, 1978.

Weyl, H. *Symmetry.* Princeton, NJ: Princeton University Press, 1952.

Walsh Function

Functions consisting of a number of fixed-amplitude square pulses interposed with zeros. Following Harmuth (1969), designate those with EVEN symmetry $\operatorname{Cal}(k, t)$ and those with ODD symmetry $\operatorname{Sal}(k, t)$. Define the SEQUENCY k as half the number of zero crossings in the time base. Walsh functions with nonidentical SEQUENCIES are ORTHOGONAL, as are the functions $\operatorname{Cal}(k, t)$ and $\operatorname{Sal}(k, t)$. The product of two Walsh functions is also a Walsh function. The Walsh functions

$$\operatorname{Wal}(k, t) = \begin{cases} \operatorname{Cal}(k/2, t) & \text{for } k = 0,\ 2,\ 4,\ \ldots \\ \operatorname{Sal}((k+1)/2, t) & \text{for } k = 1,\ 3,\ 5,\ \ldots. \end{cases}$$

The Walsh functions $\operatorname{Cal}(k, t)$ for $k = 0, 1, \ldots, n/2 - 1$ and $\operatorname{Sal}(k, t)$ for $k = 1, 2, \ldots, n/2$ are given by the rows of the HADAMARD MATRIX H_n.

see also HADAMARD MATRIX, SEQUENCY

References
Beauchamp, K. G. *Walsh Functions and Their Applications.* London: Academic Press, 1975.
Harmuth, H. F. "Applications of Walsh Functions in Communications." *IEEE Spectrum* **6**, 82–91, 1969.
Thompson, A. R.; Moran, J. M.; and Swenson, G. W. Jr. *Interferometry and Synthesis in Radio Astronomy.* New York: Wiley, p. 204, 1986.
Tzafestas, S. G. *Walsh Functions in Signal and Systems Analysis and Design.* New York: Van Nostrand Reinhold, 1985.
Walsh, J. L. "A Closed Set of Normal Orthogonal Functions." *Amer. J. Math.* **45**, 5–24, 1923.

Walsh Index

The statistical INDEX

$$P_W \equiv \frac{\sum \sqrt{q_0 q_n}\, p_n}{\sum \sqrt{q_0 q_n}\, p_0},$$

where p_n is the price per unit in period n and q_n is the quantity produced in period n.

see also INDEX

References
Kenney, J. F. and Keeping, E. S. *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, p. 66, 1962.

Wang's Conjecture

Wang's conjecture states that if a set of tiles can tile the plane, then they can always be arranged to do so periodically (Wang 1961). The CONJECTURE was refuted when Berger (1966) showed that an aperiodic set of tiles existed. Berger used 20,426 tiles, but the number has subsequently been greatly reduced.

see also TILING

References
Adler, A. and Holroyd, F. C. "Some Results on One-Dimensional Tilings." *Geom. Dedicata* **10**, 49–58, 1981.
Berger, R. "The Undecidability of the Domino Problem." *Mem. Amer. Math. Soc. No.* **66**, 1–72, 1966.
Grünbaum, B. and Sheppard, G. C. *Tilings and Patterns.* New York: W. H. Freeman, 1986.
Hanf, W. "Nonrecursive Tilings of the Plane. I." *J. Symbolic Logic* **39**, 283–285, 1974.
Mozes, S. "Tilings, Substitution Systems, and Dynamical Systems Generated by Them." *J. Analyse Math.* **53**, 139–186, 1989.
Myers, D. "Nonrecursive Tilings of the Plane. II." *J. Symbolic Logic* **39**, 286–294, 1974.
Robinson, R. M. "Undecidability and Nonperiodicity for Tilings of the Plane." *Invent. Math.* **12**, 177–209, 1971.
Wang, H. *Bell Systems Tech. J.* **40**, 1–41, 1961.

Ward's Primality Test

Let N be an ODD INTEGER, and assume there exists a LUCAS SEQUENCE $\{U_n\}$ with associated SYLVESTER CYCLOTOMIC NUMBERS $\{Q_n\}$ such that there is an $n > \sqrt{N}$ (with n and N RELATIVELY PRIME) for which N DIVIDES Q_n. Then N is a PRIME unless it has one of the following two forms:

1. $N = (n-1)^2$, with $n-1$ PRIME and $n > 4$, or
2. $N = n^2 - 1$, with $n-1$ and $n+1$ PRIME.

see also LUCAS SEQUENCE, SYLVESTER CYCLOTOMIC NUMBER

References
Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, pp. 69–70, 1989.

Waring's Conjecture

see WARING'S PRIME CONJECTURE, WARING'S SUM CONJECTURE

Waring Formula

$$A^n + B^n = \sum_{j=0}^{\lfloor n/2 \rfloor} (-1)^j \frac{n}{n-j} \binom{n-j}{j} (AB)^j (A+B)^{n-2j},$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION and $\binom{n}{k}$ is a BINOMIAL COEFFICIENT.

see also FERMAT'S LAST THEOREM

Waring's Prime Conjecture

Every ODD INTEGER is a PRIME or the sum of three PRIMES.

Waring's Problem

Waring proposed a generalization of LAGRANGE'S FOUR-SQUARE THEOREM, stating that every RATIONAL INTEGER is the sum of a fixed number $g(n)$ of nth POWERS of INTEGERS, where n is any given POSITIVE INTEGER and $g(n)$ depends only on n. Waring originally speculated that $g(2) = 4$, $g(3) = 9$, and $g(4) = 19$. In 1909, Hilbert proved the general conjecture using an identity in 25-fold multiple integrals (Rademacher and Toeplitz 1957, pp. 52–61).

In LAGRANGE'S FOUR-SQUARE THEOREM, Lagrange proved that $g(2) = 4$, where 4 may be reduced to 3 except for numbers of the form $4^n(8k+7)$ (as proved by Legendre). In the early twentieth century, Dickson, Pillai, and Niven proved that $g(3) = 9$. Hilbert, Hardy, and Vinogradov proved $g(4) \leq 21$, and this was subsequently reduced to $g(4) = 19$ by Balasubramanian *et al.* (1986). Liouville proved (using LAGRANGE'S FOUR-SQUARE THEOREM and LIOUVILLE POLYNOMIAL IDENTITY) that $g(5) \leq 53$, and this was improved to 47, 45, 41, 39, 38, and finally $g(5) \leq 37$ by Wieferich. See Rademacher and Toeplitz (1957, p. 56) for a simple proof. J.-J. Chen (1964) proved that $g(5) = 37$.

Dickson, Pillai, and Niven also conjectured an explicit formula for $g(s)$ for $s > 6$ (Bell 1945), based on the relationship

$$\left(\frac{3}{2}\right)^n - \left\lfloor \left(\frac{3}{2}\right)^n \right\rfloor = 1 - \left(\frac{1}{2}\right)^n \left\{ \left\lfloor \left(\frac{3}{2}\right)^n + 2 \right\rfloor \right\}. \quad (1)$$

If the DIOPHANTINE (i.e., n is restricted to being an INTEGER) inequality

$$\left\{\left(\frac{3}{2}\right)^n\right\} \le 1 - \left(\frac{3}{4}\right)^n \tag{2}$$

is true, then

$$g(n) = 2^n + \left\lfloor\left(\frac{3}{2}\right)^n\right\rfloor - 2. \tag{3}$$

This was given as a lower bound by Euler, and has been verified to be correct for $6 \le n \le 200,000$. Since 1957, it has been known that at most a FINITE number of k exceed Euler's lower bound.

There is also a related problem of finding the least INTEGER n such that every POSITIVE INTEGER beyond a certain point (i.e., all but a FINITE number) is the SUM of $G(n)$ nth POWERS. From 1920–1928, Hardy and Littlewood showed that

$$G(n) \le (n-2)2^{n-1} + 5 \tag{4}$$

and conjectured that

$$G(k) < \begin{cases} 2k+1 & \text{for } k \text{ not a power of 2} \\ 4k & \text{for } k \text{ a power of 2.} \end{cases} \tag{5}$$

The best currently known bound is

$$G(k) < ck \ln k \tag{6}$$

for some constant c. Heilbronn (1936) improved Vinogradov's results to obtain

$$G(n) \le 6n \ln n + \left[4 + 3 \ln\left(3 + \frac{2}{n}\right)\right] n + 3. \tag{7}$$

It has long been known that $G(2) = 4$. Dickson and Landau proved that the only INTEGERS requiring nine CUBES are 23 and 239, thus establishing $G(3) \le 8$. Wieferich proved that only 15 INTEGERS require eight CUBES: 15, 22, 50, 114, 167, 175, 186, 212, 213, 238, 303, 364, 420, 428, and 454, establishing $G(3) \le 7$. The largest number known requiring seven CUBES is 8042. In 1933, Hardy and Littlewood showed that $G(4) \le 19$, but this was improved in 1936 to 16 or 17, and shown to be exactly 16 by Davenport (1939b). Vaughan (1986) greatly improved on the method of Hardy and Littlewood, obtaining improved results for $n \ge 5$. These results were then further improved by Brüdern (1990), who gave $G(5) \le 18$, and Wooley (1992), who gave $G(n)$ for $n = 6$ to 20. Vaughan and Wooley (1993) showed $G(8) \le 42$.

Let $G^+(n)$ denote the smallest number such that *almost all* sufficiently large INTEGERS are the sum of $G^+(n)$ nth POWERS. Then $G^+(3) = 4$ (Davenport 1939a), $G^+(4) = 15$ (Hardy and Littlewood 1925), $G^+(8) = 32$ (Vaughan 1986), and $G^+(16) = 64$ (Wooley 1992). If the negatives of POWERS are permitted in addition to the powers themselves, the largest number of nth POWERS needed to represent an aribtrary integer are denoted $eg(n)$ and $EG(n)$ (Wright 1934, Hunter 1941, Gardner 1986). In general, these values are much harder to calculate than are $g(n)$ and $G(n)$.

The following table gives $g(n)$, $G(n)$, $G^+(n)$, $eg(n)$, and $EG(n)$ for $n \le 20$. The sequence of $g(n)$ is Sloane's A002804.

n	$g(n)$	$G(n)$	$G^+(n)$	$eg(n)$	$EG(n)$
2	4	4		3	3
3	9	≤ 7	≤ 4	[4, 5]	
4	19	16	≤ 15	[9, 10]	
5	37	≤ 18			
6	73	≤ 27			
7	143	≤ 36			
8	279	≤ 42	≤ 32		
9	548	≤ 55			
10	1079	≤ 63			
11	2132	≤ 70			
12	4223	≤ 79			
13	8384	≤ 87			
14	16673	≤ 95			
15	33203	≤ 103			
16	66190	≤ 112	≤ 64		
17	132055	≤ 120			
18	263619	≤ 129			
19	526502	≤ 138			
20	1051899	≤ 146			

see also EULER'S CONJECTURE, SCHNIRELMANN'S THEOREM, VINOGRADOV'S THEOREM

References

Balasubramanian, R.; Deshouillers, J.-M.; and Dress, F. "Problème de Waring por les bicarrés 1, 2." *C. R. Acad. Sci. Paris Sér. I Math.* **303**, 85–88 and 161–163, 1986.

Bell, E. T. *The Development of Mathematics, 2nd ed.* New York: McGraw-Hill, p. 318, 1945.

Brüdern, J. "On Waring's Problem for Fifth Powers and Some Related Topics." *Proc. London Math. Soc.* **61**, 457–479, 1990.

Davenport, H. "On Waring's Problem for Cubes." *Acta Math.* **71**, 123–143, 1939a.

Davenport, H. "On Waring's Problem for Fourth Powers." *Ann. Math.* **40**, 731–747, 1939b.

Dickson, L. E. "Waring's Problem and Related Results." Ch. 25 in *History of the Theory of Numbers, Vol. 2: Diophantine Analysis.* New York: Chelsea, pp. 717–729, 1952.

Gardner, M. "Waring's Problems." Ch. 18 in *Knotted Doughnuts and Other Mathematical Entertainments.* New York: W. H. Freeman, 1986.

Guy, R. K. "Sums of Squares." §C20 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 136–138, 1994.

Hardy, G. H. and Littlewood, J. E. "Some Problems of Partitio Numerorum (VI): Further Researches in Waring's Problem." *Math. Z.* **23**, 1–37, 1925.

Hunter, W. "The Representation of Numbers by Sums of Fourth Powers." *J. London Math. Soc.* **16**, 177–179, 1941.

Khinchin, A. Y. "An Elementary Solution of Waring's Problem." Ch. 3 in *Three Pearls of Number Theory.* New York: Dover, pp. 37–64, 1998.

Rademacher, H. and Toeplitz, O. *The Enjoyment of Mathematics: Selections from Mathematics for the Amateur.* Princeton, NJ: Princeton University Press, 1957.
Stewart, I. "The Waring Experience." *Nature* **323**, 674, 1986.
Vaughan, R. C. "On Waring's Problem for Smaller Exponents." *Proc. London Math. Soc.* **52**, 445–463, 1986.
Vaughan, R. C. and Wooley, T. D. "On Waring's Problem: Some Refinements." *Proc. London Math. Soc.* **63**, 35–68, 1991.
Vaughan, R. C. and Wooley, T. D. "Further Improvements in Waring's Problem." *Phil. Trans. Roy. Soc. London A* **345**, 363–376, 1993a.
Vaughan, R. C. and Wooley, T. D. "Further Improvements in Waring's Problem III. Eighth Powers." *Phil. Trans. Roy. Soc. London A* **345**, 385–396, 1993b.
Wooley, T. D. "Large Improvements in Waring's Problem." *Ann. Math.* **135**, 131–164, 1992.
Wright, E. M. "An Easier Waring's Problem." *J. London Math. Soc.* **9**, 267–272, 1934.

Waring's Sum Conjecture

see WARING'S PROBLEM

Waring's Theorem

If each of two curves meets the LINE AT INFINITY in distinct, nonsingular points, and if all their intersections are finite, then if to each common point there is attached a weight equal to the number of intersections absorbed therein, the CENTER OF MASS of these points is the center of gravity of the intersections of the asymptotes.

References
Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 166, 1959.

Watchman Theorem

see ART GALLERY THEOREM

Watson's Formula

Let $J_\nu(z)$ be a BESSEL FUNCTION OF THE FIRST KIND, $Y_\nu(z)$ a BESSEL FUNCTION OF THE SECOND KIND, and $K_\nu(z)$ a MODIFIED BESSEL FUNCTION OF THE FIRST KIND. Also let $\Re[z] > 0$ and require $\Re[\mu - \nu] < 1$. Then

$$\begin{aligned} J_\mu(z)Y_\nu(z) - J_\nu(z)Y_\mu(z) \\ = \frac{4\sin[(\mu-\nu)\pi]}{\pi^2} \int_0^\infty K_{\nu-\mu}(2z\sinh t)e^{-(\mu+\nu)t}\,dt. \end{aligned}$$

The fourth edition of Gradshteyn and Ryzhik (1979), Iyanaga and Kawada (1980), and Ito (1987) erroneously give the exponential with a PLUS SIGN. A related integral is given by

$$\begin{aligned} J_\nu(z)\frac{\partial Y_\nu(z)}{\partial \nu} - Y_\nu(z)\frac{\partial J_\nu(z)}{\partial \nu} \\ = -\frac{4}{\pi}\int_0^\infty K_0(2z\sinh t)e^{-2\nu t}\,dt \end{aligned}$$

for $\Re[z] > 0$.

see also DIXON-FERRAR FORMULA, NICHOLSON'S FORMULA

References
Gradshteyn, I. S. and Ryzhik, I. M. Eqns. 6.617.1 and 6.617.2 in *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 710, 1979.
Ito, K. (Ed.). *Encyclopedic Dictionary of Mathematics, 2nd ed.* Cambridge, MA: MIT Press, p. 1806, 1987.
Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1476, 1980.

Watson-Nicholson Formula

Let $H_\nu^{(\iota)}$ be a HANKEL FUNCTION OF THE FIRST or SECOND KIND, let $x, \nu > 0$, and define

$$w = \sqrt{\left(\frac{x}{\nu}\right)^2 - 1}.$$

Then

$$\begin{aligned} H_\nu^{(\iota)}(x) = 3^{-1/2} w \exp\{(-1)^{\iota+1} i[\pi/6 + \nu(w - \tfrac{1}{3}w^3 \\ - \tan^{-1} w)]\} H_{1/3}^{(\iota)}(\tfrac{1}{3}\nu w) + \mathcal{O}|\nu^{-1}|. \end{aligned}$$

References
Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1475, 1980.

Watson Quintuple Product Identity

see QUINTUPLE PRODUCT IDENTITY

Watson's Theorem

$$\begin{aligned} {}_3F_2\left[\begin{matrix} a, b, c \\ \frac{1}{2}(a+b+1), c \end{matrix}\right] \\ = \frac{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}+c)\Gamma[\frac{1}{2}(1+a+b)]\Gamma(\frac{1}{2}-\frac{1}{2}a-\frac{1}{2}b+c)}{\Gamma[\frac{1}{2}(1+a)]\Gamma[\frac{1}{2}(1+b)]\Gamma(\frac{1}{2}-\frac{1}{2}a+c)\Gamma(\frac{1}{2}-\frac{1}{2}b+c)}, \end{aligned}$$

where ${}_3F_2(a, b, c; d, e; z)$ is a GENERALIZED HYPERGEOMETRIC FUNCTION and $\Gamma(z)$ is the GAMMA FUNCTION.

Watt's Curve

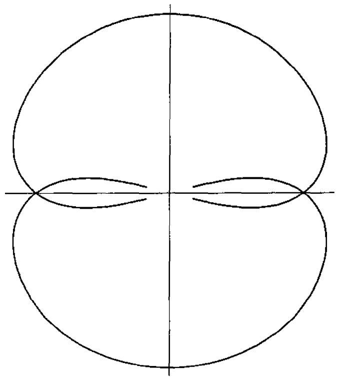

A curve named after James Watt (1736–1819), the Scottish engineer who developed the steam engine (MacTutor Archive). The curve is produced by a LINKAGE of rods connecting two wheels of equal diameter. Let the two wheels have RADIUS b and let their centers be located a distance $2a$ apart. Further suppose that a rod of length $2c$ is fixed at each end to the CIRCUMFERENCE of the two wheels. Let P be the MIDPOINT of the rod. Then Watt's curve C is the LOCUS of P.

The POLAR equation of Watt's curve is

$$r^2 = b^2 - (a\sin\theta \pm \sqrt{c^2 - a^2\cos^2\theta}\,)^2.$$

If $a = c$, then C is a CIRCLE of RADIUS b with a figure of eight inside it.

References
Lockwood, E. H. *A Book of Curves.* Cambridge, England: Cambridge University Press, p. 162, 1967.
MacTutor History of Mathematics Archive. "Watt's Curve." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Watts.html`.

Watt's Parallelogram

A LINKAGE used in the original steam engine to turn back-and-forth motion into approximately straight-line motion.

see also LINKAGE

References
Rademacher, H. and Toeplitz, O. *The Enjoyment of Mathematics: Selections from Mathematics for the Amateur.* Princeton, NJ: Princeton University Press, pp. 119–121, 1957.

Wave

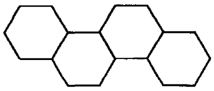

A 4-POLYHEX.

References
Gardner, M. *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, p. 147, 1978.

Wave Equation

The wave equation is

$$\nabla^2\psi = \frac{1}{v^2}\frac{\partial^2\psi}{\partial t^2}, \tag{1}$$

where ∇^2 is the LAPLACIAN.

The 1-D wave equation is

$$\frac{\partial^2\psi}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2\psi}{\partial t^2}. \tag{2}$$

In order to specify a wave, the equation is subject to boundary conditions

$$\psi(0,t) = 0 \tag{3}$$

$$\psi(L,t) = 0, \tag{4}$$

and initial conditions

$$\psi(x,0) = f(x) \tag{5}$$

$$\frac{\partial\psi}{\partial t}(x,0) = g(x). \tag{6}$$

The wave equation can be solved using the so-called d'Alembert's solution, a FOURIER TRANSFORM method, or SEPARATION OF VARIABLES.

d'Alembert devised his solution in 1746, and Euler subsequently expanded the method in 1748. Let

$$\xi \equiv x - at \tag{7}$$

$$\eta \equiv x + at. \tag{8}$$

By the CHAIN RULE,

$$\frac{\partial^2\psi}{\partial x^2} = \frac{\partial^2\psi}{\partial\xi^2} + 2\frac{\partial^2\psi}{\partial\xi\partial\eta} + \frac{\partial 2\psi}{\partial\eta^2} \tag{9}$$

$$\frac{1}{v^2}\frac{\partial^2\psi}{\partial t^2} = \frac{\partial^2\psi}{\partial\xi^2} - 2\frac{\partial^2\psi}{\partial\xi\partial\eta} + \frac{\partial^2\psi}{\partial\eta^2}. \tag{10}$$

The wave equation then becomes

$$\frac{\partial^2\psi}{\partial\xi\partial\eta} = 0. \tag{11}$$

Any solution of this equation is of the form

$$\psi(\xi,\eta) = f(\eta) + g(\xi) = f(x+vt) + g(x-vt), \tag{12}$$

where f and g are *any* functions. They represent two waveforms traveling in opposite directions, f in the NEGATIVE x direction and g in the POSITIVE x direction.

The 1-D wave equation can also be solved by applying a FOURIER TRANSFORM to each side,

$$\int_{-\infty}^{\infty}\frac{\partial^2\psi(x,t)}{\partial x^2}e^{-2\pi ikx}\,dx = \frac{1}{v^2}\int_{-\infty}^{\infty}\frac{\partial^2\psi(x,t)}{\partial t^2}e^{-2\pi ikx}\,dx, \tag{13}$$

which is given, with the help of the FOURIER TRANSFORM DERIVATIVE identity, by

$$(2\pi ik)^2\Psi(k,t) = \frac{1}{v^2}\frac{\partial^2\Psi(k,t)}{\partial t^2}, \tag{14}$$

where

$$\Psi(k,t) \equiv \mathcal{F}[\psi(x,t)] = \int_{-\infty}^{\infty}\psi(x,t)e^{-2\pi ikx}\,dx. \tag{15}$$

This has solution

$$\Psi(k,t) = A(k)e^{2\pi ikvt} + B(k)e^{-2\pi ikvt}. \tag{16}$$

Taking the inverse FOURIER TRANSFORM gives

$$\begin{aligned}\psi(x,t) &\equiv \int_{-\infty}^{\infty} \Psi(k,t)e^{2\pi ikx}\,dx \\ &= \int_{-\infty}^{\infty} [A(k)e^{2\pi ikvt} \\ &\quad + B(k)e^{-2\pi ikvt}]e^{-2\pi ikx}\,dk \\ &= \int_{-\infty}^{\infty} A(k)e^{-2\pi ik(x-vt)}\,dk \\ &\quad + \int_{-\infty}^{\infty} B(k)e^{-2\pi ik(x+vt)}\,dk \\ &= f_1(x-vt) + b(k)f_2(x+vt),\end{aligned} \tag{17}$$

where

$$f_1(u) \equiv \mathcal{F}[A(k)] = \int_{-\infty}^{\infty} A(k)e^{-2\pi iku}\,dk \tag{18}$$

$$f_2(u) \equiv \mathcal{F}[B(k)] = \int_{-\infty}^{\infty} B(k)e^{-2\pi iku}\,dk. \tag{19}$$

This solution is still subject to all other initial and boundary conditions.

The 1-D wave equation can be solved by SEPARATION OF VARIABLES using a trial solution

$$\psi(x,t) = X(x)T(t). \tag{20}$$

This gives

$$T\frac{d^2X}{dx^2} = \frac{1}{v^2}X\frac{d^2T}{dt^2} \tag{21}$$

$$\frac{1}{X}\frac{d^2X}{dx^2} = \frac{1}{v^2}\frac{1}{T}\frac{d^2T}{dt^2} = -k^2. \tag{22}$$

So the solution for X is

$$X(x) = C\cos(kx) + D\sin(kx). \tag{23}$$

Rewriting (22) gives

$$\frac{1}{T}\frac{d^2T}{dt^2} = -v^2k^2 \equiv -\omega^2, \tag{24}$$

so the solution for T is

$$T(t) = E\cos(\omega t) + F\sin(\omega t), \tag{25}$$

where $v \equiv \omega/k$. Applying the boundary conditions $\psi(0,t) = \psi(L,t) = 0$ to (23) gives

$$C = 0 \qquad kL = m\pi, \tag{26}$$

where m is an INTEGER. Plugging (23), (25) and (26) back in for ψ in (21) gives, for a particular value of m,

$$\begin{aligned}\psi_m(x,t) &= [E_m\sin(\omega_m t) + F_m\cos(\omega_m t)]D_m\sin\left(\frac{m\pi x}{L}\right) \\ &\equiv [A_m\cos(\omega_m t) + B_m\sin(\omega_m t)]\sin\left(\frac{m\pi x}{L}\right).\end{aligned} \tag{27}$$

The initial condition $\dot\psi(x,0) = 0$ then gives $B_m = 0$, so (27) becomes

$$\psi_m(x,t) = A_m\cos(\omega_m t)\sin\left(\frac{m\pi x}{L}\right). \tag{28}$$

The general solution is a sum over all possible values of m, so

$$\psi(x,t) = \sum_{m=1}^{\infty} A_m\cos(\omega_m t)\sin\left(\frac{m\pi x}{L}\right). \tag{29}$$

Using ORTHOGONALITY of sines again,

$$\int_0^L \sin\left(\frac{l\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)dx = \tfrac{1}{2}L\delta_{lm}, \tag{30}$$

where δ_{lm} is the KRONECKER DELTA defined by

$$\delta_{mn} \equiv \begin{cases}1 & m = n \\ 0 & m \neq n\end{cases}, \tag{31}$$

gives

$$\begin{aligned}\int_0^L &\psi(x,0)\sin\left(\frac{m\pi x}{L}\right)dx \\ &= \sum_{l=1}^{\infty} A_l\sin\left(\frac{l\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)dx \\ &\qquad = \sum_{l=1}^{\infty} A_l\tfrac{1}{2}L\delta_{lm} = \tfrac{1}{2}LA_m,\end{aligned} \tag{32}$$

so we have

$$A_m = \frac{2}{L}\int_0^L \psi(x,0)\sin\left(\frac{m\pi x}{L}\right)dx. \tag{33}$$

The computation of A_ms for specific initial distortions is derived in the FOURIER SINE SERIES section. We already have found that $B_m = 0$, so the equation of motion for the string (29), with

$$\omega_m \equiv vk_m = \frac{vm\pi}{L}, \tag{34}$$

is

$$\psi(x,t) = \sum_{m=1}^{\infty} A_m\cos\left(\frac{vm\pi t}{L}\right)\sin\left(\frac{m\pi x}{L}\right), \tag{35}$$

where the A_m COEFFICIENTS are given by (33).

A damped 1-D wave

$$\frac{\partial^2\psi}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2\psi}{\partial t^2} + b\frac{\partial\psi}{\partial t}, \tag{36}$$

given boundary conditions

$$\psi(0,t) = 0 \tag{37}$$

$$\psi(L,t) = 0, \tag{38}$$

initial conditions

$$\psi(x,0) = f(x) \tag{39}$$

$$\frac{\partial\psi}{\partial t}(x,0) = g(x), \tag{40}$$

and the additional constraint

$$0 < b < \frac{2\pi}{Lv}, \tag{41}$$

can also be solved as a FOURIER SERIES.

$$\psi(x,t) = \sum_{n=1}^{\infty} \sin\left(\frac{n\pi x}{L}\right) e^{-v^2bt/2}[a_n \sin(\mu_n t) + b_n \cos(\mu_n t)], \tag{42}$$

where

$$\mu_n \equiv \frac{\sqrt{4v^2n^2\pi^2 - b^2L^2v^4}}{2L} = \frac{v\sqrt{4n^2\pi^2 - b^2L^2v^2}}{2L} \tag{43}$$

$$b_n = \frac{2}{L}\int_0^L \sin\left(\frac{n\pi x}{L}\right) f(x)\,dx \tag{44}$$

$$a_n = \frac{2}{L\mu_n}\left\{\int_0^L \sin\left(\frac{n\pi x}{L}\right)\left[g(x) + \frac{v^2b}{2}f(x)\right]\right\} dx. \tag{45}$$

To find the motion of a rectangular membrane with sides of length L_x and L_y (in the absence of gravity), use the 2-D wave equation

$$\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = \frac{1}{v^2}\frac{\partial^2 z}{\partial t^2}, \tag{46}$$

where $z(x,y,t)$ is the vertical displacement of a point on the membrane at position (x,y) and time t. Use SEPARATION OF VARIABLES to look for solutions of the form

$$z(x,y,t) = X(x)Y(y)T(t). \tag{47}$$

Plugging (47) into (46) gives

$$YT\frac{d^2X}{dx^2} + XT\frac{d^2Y}{dy^2} = \frac{1}{v^2}XY\frac{d^2T}{dt^2}, \tag{48}$$

where the partial derivatives have now become complete derivatives. Multiplying (48) by v^2/XYT gives

$$\frac{v^2}{X}\frac{d^2X}{dx^2} + \frac{v^2}{Y}\frac{d^2Y}{dy^2} = \frac{1}{T}\frac{d^2T}{dt^2}. \tag{49}$$

The left and right sides must both be equal to a constant, so we can separate the equation by writing the right side as

$$\frac{1}{T}\frac{d^2T}{dt^2} = -\omega^2. \tag{50}$$

This has solution

$$T(t) = C_\omega \cos(\omega t) + D_\omega \sin(\omega t). \tag{51}$$

Plugging (50) back into (49),

$$\frac{v^2}{X}\frac{d^2X}{dx^2} + \frac{v^2}{Y}\frac{d^2Y}{dy^2} = -\omega^2, \tag{52}$$

which we can rewrite as

$$\frac{1}{X}\frac{d^2X}{dx^2} = -\frac{1}{Y}\frac{d^2Y}{dy^2} - \frac{\omega^2}{v^2} = -{k_x}^2 \tag{53}$$

since the left and right sides again must both be equal to a constant. We can now separate out the $Y(y)$ equation

$$\frac{1}{Y}\frac{d^2Y}{dy^2} = {k_x}^2 - \frac{\omega^2}{v^2} \equiv -{k_y}^2, \tag{54}$$

where we have defined a new constant k_y satisfying

$${k_x}^2 + {k_y}^2 = \frac{\omega^2}{v^2}. \tag{55}$$

Equations (53) and (54) have solutions

$$X(x) = E\cos(k_x x) + F\sin(k_x x) \tag{56}$$

$$Y(y) = G\cos(k_y y) + H\sin(k_y y). \tag{57}$$

We now apply the boundary conditions to (56) and (57). The conditions $z(0,y,t) = 0$ and $z(x,0,t) = 0$ mean that

$$E = 0 \qquad G = 0. \tag{58}$$

Similarly, the conditions $z(L_x,y,t) = 0$ and $z(x,L_y,t) = 0$ give $\sin(k_xL_x) = 0$ and $\sin(k_yL_y) = 0$, so $L_xk_x = p\pi$ and $L_yk_y = q\pi$, where p and q are INTEGERS. Solving for the allowed values of k_x and k_y then gives

$$k_x = \frac{p\pi}{L_x} \qquad k_y = \frac{q\pi}{L_y}. \tag{59}$$

Plugging (52), (56), (57), (58), and (59) back into (22) gives the solution for particular values of p and q,

$$z_{pq}(x,y,t) = [C_\omega \cos(\omega t) + D_\omega \sin(\omega t)] \times \left[F_p \sin\left(\frac{p\pi x}{L_x}\right)\right]\left[H_q \sin\left(\frac{q\pi y}{L_y}\right)\right]. \tag{60}$$

Lumping the constants together by writing $A_{pq} \equiv C_\omega F_p H_q$ (we can do this since ω is a function of p and q, so C_ω can be written as C_{pq}) and $B_{pq} \equiv D_\omega F_p H_q$, we obtain

$$z_{pq}(x,y,t) = [A_{pq}\cos(\omega_{pq}t) + B_{pq}\sin(\omega_{pq}t)] \times \sin\left(\frac{p\pi x}{L_x}\right)\sin\left(\frac{q\pi y}{L_y}\right). \quad (61)$$

Plots of the spatial part for modes (1, 1), (1, 2), (2, 1), and (2, 2) follow.

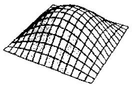
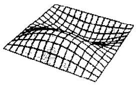
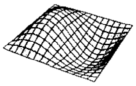
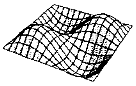

The general solution is a sum over all possible values of p and q, so the final solution is

$$z(x,y,t) = \sum_{p=1}^{\infty}\sum_{q=1}^{\infty}[A_{pq}\cos(\omega_{pq}t) + B_{pq}\sin(\omega_{pq}t)] \times \sin\left(\frac{p\pi x}{L_x}\right)\sin\left(\frac{q\pi y}{L_y}\right), \quad (62)$$

where ω is defined by combining (55) and (59) to yield

$$\omega_{pq} \equiv \pi v\sqrt{\left(\frac{p}{L_x}\right)^2 + \left(\frac{q}{L_y}\right)^2}. \quad (63)$$

Given the initial conditions $z(x,y,0)$ and $\frac{\partial z}{\partial t}(x,y,0)$, we can compute the A_{pq}s and B_{pq}s explicitly. To accomplish this, we make use of the orthogonality of the SINE function in the form

$$I \equiv \int_0^L \sin\left(\frac{m\pi x}{L}\right)\sin\left(\frac{n\pi x}{L}\right)dx = \tfrac{1}{2}L\delta_{mn}, \quad (64)$$

where δ_{mn} is the KRONECKER DELTA. This can be demonstrated by direct INTEGRATION. Let $u \equiv \pi x/L$ so $du = (\pi/L)\,dx$ in (64), then

$$I = \frac{L}{\pi}\int_0^\pi \sin(mu)\sin(nu)\,du. \quad (65)$$

Now use the trigonometric identity

$$\sin\alpha\sin\beta = \tfrac{1}{2}[\cos(\alpha-\beta) - \cos(\alpha+\beta)] \quad (66)$$

to write

$$I = \frac{L}{2\pi}\int_0^\pi \cos[(m-n)u]\,du + \int_0^\pi \cos[(m+n)u]\,du. \quad (67)$$

Note that for an INTEGER $l \neq 0$, the following INTEGRAL vanishes

$$\int_0^\pi \cos(lu)\,du = \frac{1}{l}[\sin(lu)]_0^\pi = \frac{1}{l}[\sin(l\pi) - \sin 0] = \frac{1}{l}\sin(l\pi) = 0, \quad (68)$$

since $\sin(l\pi) = 0$ when l is an INTEGER. Therefore, $I = 0$ when $l \equiv m - n \neq 0$. However, I does *not* vanish when $l = 0$, since

$$\int_0^\pi \cos(0\cdot u)\,du = \int_0^\pi du = \pi. \quad (69)$$

We therefore have that $I = L\delta_{mn}/2$, so we have derived (64). Now we multiply $z(x,y,0)$ by two sine terms and integrate between 0 and L_x and between 0 and L_y,

$$I = \int_0^{L_y}\left[\int_0^{L_x} z(x,y,0)\sin\left(\frac{p\pi x}{L_x}\right)dx\right] \times \sin\left(\frac{q\pi y}{L_y}\right)dy. \quad (70)$$

Now plug in $z(x,y,t)$, set $t = 0$, and prime the indices to distinguish them from the p and q in (70),

$$I = \sum_{q'=1}^{\infty}\int_0^{L_y}\left[\sum_{p'=1}^{\infty}A_{p'q'}\int_0^{L_x}\sin\left(\frac{p\pi x}{L_x}\right)\sin\left(\frac{p'\pi x}{L_x}\right)dx\right] \times \sin\left(\frac{q\pi y}{L_y}\right)\sin\left(\frac{q'\pi y}{L_y}\right)dy. \quad (71)$$

Making use of (64) in (71),

$$I = \sum_{q'=1}^{\infty}\int_0^{L_y}\sum_{p'=1}^{\infty}A_{p'q'}\frac{L_x}{2}\delta_{p,p'} \times \sin\left(\frac{q\pi y}{L_y}\right)\sin\left(\frac{q'\pi y}{L_y}\right)dy, \quad (72)$$

so the sums over p' and q' collapse to a single term

$$I = \frac{L_x}{2}\sum_{q=1}^{\infty}A_{pq'}\frac{L_y}{2}\delta_{q,q'} = \frac{L_xL_y}{4}A_{pq}. \quad (73)$$

Equating (72) and (73) and solving for A_{pq} then gives

$$A_{pq} = \frac{4}{L_xL_y}\int_0^{L_y}\left[\int_0^{L_x} z(x,y,0)\sin\left(\frac{p\pi x}{L_x}\right)dx\right] \times \sin\left(\frac{q\pi y}{L_y}\right)dy. \quad (74)$$

An analogous derivation gives the B_{pq}s as

$$B_{pq} = \frac{4}{\omega_{pq} L_x L_y} \int_0^{L_y} \left[\int_0^{L_x} \frac{\partial z}{\partial t}(x,y,0) \sin\left(\frac{p\pi x}{L_x}\right) dx \right] \times \sin\left(\frac{q\pi y}{L_y}\right) dy. \quad (75)$$

The equation of motion for a membrane shaped as a RIGHT ISOSCELES TRIANGLE of length c on a side and with the sides oriented along the POSITIVE x and y axes is given by

$$\psi(x,y,t) = [C_{pq}\cos(\omega_{pq}t) + D_{pq}\sin(\omega_{pq}t)] \times \left[\sin\left(\frac{p\pi x}{c}\right)\sin\left(\frac{q\pi y}{c}\right) - \sin\left(\frac{q\pi x}{c}\right)\sin\left(\frac{p\pi y}{c}\right)\right], \quad (76)$$

where

$$\omega_{pq} = \frac{\pi v}{c}\sqrt{p^2+q^2} \quad (77)$$

and p, q INTEGERS with $p > q$. This solution can be obtained by subtracting two wave solutions for a square membrane with the indices reversed. Since points on the diagonal which are equidistant from the center must have the same wave equation solution (by symmetry), this procedure gives a wavefunction which will vanish along the diagonal *as long as* p *and* q *are both* EVEN *or* ODD. We must further restrict the modes since those with $p < q$ give wavefunctions which are just the NEGATIVE of (q,p) and (p,p) give an identically zero wavefunction. The following plots show (3, 1), (4, 2), (5, 1), and (5,3).

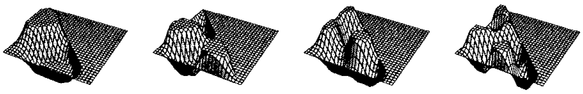

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Wave Equation in Prolate and Oblate Spheroidal Coordinates." §21.5 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 752–753, 1972.

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 124–125, 1953.

Wave Operator

An OPERATOR relating the asymptotic state of a DYNAMICAL SYSTEM governed by the Schrödinger equation

$$i\frac{d}{dt}\psi(t) = H\psi(t)$$

to its original asymptotic state.

see also SCATTERING OPERATOR

Wave Surface

A SURFACE represented parametrically by ELLIPTIC FUNCTIONS.

Wavelet

Wavelets are a class of a functions used to localize a given function in both space and scaling. A family of wavelets can be constructed from a function $\psi(x)$, sometimes known as a "mother wavelet," which is confined in a finite interval. "Daughter wavelets" $\psi^{a,b}(x)$ are then formed by translation (b) and contraction (a). Wavelets are especially useful for compressing image data, since a WAVELET TRANSFORM has properties which are in some ways superior to a conventional FOURIER TRANSFORM.

An individual wavelet can be defined by

$$\psi^{a,b}(x) = |a|^{-1/2}\psi\left(\frac{x-b}{a}\right). \quad (1)$$

Then

$$W_\psi(f)(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} f(t)\psi\left(\frac{t-b}{a}\right) dt, \quad (2)$$

and CALDERÓN'S FORMULA gives

$$f(x) = C_\psi \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left\langle f, \psi^{a,b}\right\rangle \psi^{a,b}(x) a^{-2}\, da\, db. \quad (3)$$

A common type of wavelet is defined using HAAR FUNCTIONS.

see also FOURIER TRANSFORM, HAAR FUNCTION, LEMARIÉ'S WAVELET, WAVELET TRANSFORM

References

Benedetto, J. J. and Frazier, M. (Eds.). *Wavelets: Mathematics and Applications.* Boca Raton, FL: CRC Press, 1994.

Chui, C. K. *An Introduction to Wavelets.* San Diego, CA: Academic Press, 1992.

Chui, C. K. (Ed.). *Wavelets: A Tutorial in Theory and Applications.* San Diego, CA: Academic Press, 1992.

Chui, C. K.; Montefusco, L.; and Puccio, L. (Eds.). *Wavelets: Theory, Algorithms, and Applications.* San Diego, CA: Academic Press, 1994.

Daubechies, I. *Ten Lectures on Wavelets.* Philadelphia, PA: Society for Industrial and Applied Mathematics, 1992.

Erlebacher, G. H.; Hussaini, M. Y.; and Jameson, L. M. (Eds.). *Wavelets: Theory and Applications.* New York: Oxford University Press, 1996.

Foufoula-Georgiou, E. and Kumar, P. (Eds.). *Wavelets in Geophysics.* San Diego, CA: Academic Press, 1994.

Hernández, E. and Weiss, G. *A First Course on Wavelets.* Boca Raton, FL: CRC Press, 1996.

Hubbard, B. B. *The World According to Wavelets: The Story of a Mathematical Technique in the Making.* New York: A. K. Peters, 1995.

Jawerth, B. and Sweldens, W. "An Overview of Wavelet Based Multiresolution Analysis." *SIAM Rev.* **36**, 377–412, 1994.

Kaiser, G. *A Friendly Guide to Wavelets.* Cambridge, MA: Birkhäuser, 1994.

Massopust, P. R. *Fractal Functions, Fractal Surfaces, and Wavelets.* San Diego, CA: Academic Press, 1994.

Meyer, Y. *Wavelets: Algorithms and Applications.* Philadelphia, PA: SIAM Press, 1993.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Wavelet Transforms." §13.10 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 584–599, 1992.

Schumaker, L. L. and Webb, G. (Eds.). *Recent Advances in Wavelet Analysis.* San Diego, CA: Academic Press, 1993.

Stollnitz, E. J.; DeRose, T. D.; and Salesin, D. H. "Wavelets for Computer Graphics: A Primer, Part 1." *IEEE Computer Graphics and Appl.* **15**, No. 3, 76–84, 1995.

Stollnitz, E. J.; DeRose, T. D.; and Salesin, D. H. "Wavelets for Computer Graphics: A Primer, Part 2." *IEEE Computer Graphics and Appl.* **15**, No. 4, 75–85, 1995.

Strang, G. "Wavelets and Dilation Equations: A Brief Introduction." *SIAM Rev.* **31**, 614–627, 1989.

Strang, G. "Wavelets." *Amer. Sci.* **82**, 250–255, 1994.

Taswell, C. *Handbook of Wavelet Transform Algorithms.* Boston, MA: Birkhäuser, 1996.

Teolis, A. *Computational Signal Processing with Wavelets.* Boston, MA: Birkhäuser, 1997.

Walter, G. G. *Wavelets and Other Orthogonal Systems with Applications.* Boca Raton, FL: CRC Press, 1994.

"Wavelet Digest." `http://www.math.sc.edu/~wavelet/`.

Wickerhauser, M. V. *Adapted Wavelet Analysis from Theory to Software.* Wellesley, MA: Peters, 1994.

Wavelet Matrix

A MATRIX composed of HAAR FUNCTIONS which is used in the WAVELET TRANSFORM. The fourth-order wavelet matrix is given by

$$\begin{aligned} W_4 &= \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & 0 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & & \\ 1 & -1 & & \\ & & 1 & 1 \\ & & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & & 1 & \\ & 1 & & \\ & & & 0 \end{bmatrix} \\ &\quad \times \begin{bmatrix} 1 & 1 & & \\ 1 & -1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}. \end{aligned}$$

A wavelet matrix can be computed in $\mathcal{O}(n)$ steps, compared to $\mathcal{O}(n \lg 2)$ for the FOURIER MATRIX.

see also FOURIER MATRIX, WAVELET, WAVELET TRANSFORM

Wavelet Transform

A transform which localizes a function both in space and scaling and has some desirable properties compared to the FOURIER TRANSFORM. The transform is based on a WAVELET MATRIX, which can be computed more quickly than the analogous FOURIER MATRIX.

see also DAUBECHIES WAVELET FILTER, LEMARIE'S WAVELET

References

Blair, D. and MathSoft, Inc. "Wavelet Resources." `http://www.mathsoft.com/wavelets.html`.

Daubechies, I. *Ten Lectures on Wavelets.* Philadelphia, PA: SIAM, 1992.

DeVore, R.; Jawerth, B.; and Lucier, B. "Images Compression through Wavelet Transform Coding." *IEEE Trans. Information Th.* **38**, 719–746, 1992.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Wavelet Transforms." §13.10 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 584–599, 1992.

Strang, G. "Wavelet Transforms Versus Fourier Transforms." *Bull. Amer. Math. Soc.* **28**, 288–305, 1993.

Weak Convergence

Weak convergence is usually either denoted $x_n \overset{w}{\to} x$ or $x_n \rightharpoonup x$. A SEQUENCE $\{x_n\}$ of VECTORS in an INNER PRODUCT SPACE E is called weakly convergent to a VECTOR in E if

$$\langle x_n, y\rangle \to \langle x, y\rangle \quad \text{as } n \to \infty, \quad \text{for all } y \in E.$$

Every STRONGLY CONVERGENT sequence is also weakly convergent (but the opposite does not usually hold). This can be seen as follows. Consider the sequence $\{x_n\}$ that converges strongly to x, i.e., $\|x_n - x\| \to 0$ as $n \to \infty$. SCHWARZ'S INEQUALITY now gives

$$|\langle x_n - x, y\rangle| \leq \|x_n - x\| \, \|y\| \quad \text{as } n \to \infty.$$

The definition of weak convergence is therefore satisfied.

see also INNER PRODUCT SPACE, SCHWARZ'S INEQUALITY, STRONG CONVERGENCE

Weak Law of Large Numbers

Also known as BERNOULLI'S THEOREM. Let $x_1, \ldots, x_n$ be a sequence of independent and identically distributed random variables, each having a MEAN $\langle xi\rangle = \mu$ and STANDARD DEVIATION σ. Define a new variable

$$x \equiv \frac{x_1 + \ldots + x_n}{n}. \tag{1}$$

Then, as $n \to \infty$, the sample mean $\langle x\rangle$ equals the population MEAN μ of each variable.

$$\langle x\rangle = \left\langle \frac{x_1 + \ldots + x_n}{n} \right\rangle = \frac{1}{n}(\langle x_1\rangle + \ldots + \langle x_n\rangle) = \frac{n\mu}{n} = \mu \tag{2}$$

$$\begin{aligned} \operatorname{var}(x) &= \operatorname{var}\left(\frac{x_1 + \ldots + x_2}{n}\right) \\ &= \operatorname{var}\left(\frac{x_1}{n}\right) + \ldots + \operatorname{var}\left(\frac{x_n}{n}\right) \\ &= \frac{\sigma^2}{n^2} + \ldots + \frac{\sigma^2}{n^2} = \frac{\sigma^2}{n}. \end{aligned} \tag{3}$$

Therefore, by the CHEBYSHEV INEQUALITY, for all $\epsilon > 0$,

$$P(|x - \mu| \geq \epsilon) \leq \frac{\operatorname{var}(x)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}. \tag{4}$$

As $n \to \infty$, it then follows that

$$\lim_{n\to\infty} P(|x-\mu| \geq \epsilon) = 0 \tag{5}$$

for ϵ arbitrarily small; i.e., as $n \to \infty$, the sample MEAN is the same as the population MEAN.

Stated another way, if an event occurs x times in s TRIALS and if p is the probability of success in a single TRIAL, then the probability that the relative frequency of successes is x/s differs from p by less than any arbitrary POSITIVE quantity ϵ which approaches 1 as $s \to \infty$.

see also LAW OF TRULY LARGE NUMBERS, STRONG LAW OF LARGE NUMBERS

Weakly Binary Tree

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

A ROOTED TREE for which the ROOT is adjacent to at most two VERTICES, and all nonroot VERTICES are adjacent to at most three VERTICES. Let $b(n)$ be the number of weakly binary trees of order n, then $b(5) = 6$. Let

$$g(z) = \sum_{i=0}^{\infty} g_i z^i, \tag{1}$$

where

$$g_0 = 0 \tag{2}$$

$$g_1 = g_2 = g_3 = 1 \tag{3}$$

$$g_{2i+1} = \sum_{j=1}^{i} g_{2i+1-j} g_j \tag{4}$$

$$g_{2i} = \tfrac{1}{2} g_i(g_i+1) + \sum_{j=1}^{i-1} g_{2i-j} g_j. \tag{5}$$

Otter (Otter 1948, Harary and Palmer 1973, Knuth 1969) showed that

$$\lim_{n\to\infty} \frac{b(n) n^{3/2}}{\xi^n} = \eta, \tag{6}$$

where

$$\xi = 2.48325\ldots \tag{7}$$

is the unique POSITIVE ROOT of

$$g\left(\frac{1}{x}\right) = 1,$$

and

$$\eta = 0.7916032\ldots. \tag{8}$$

ξ is also given by

$$\xi = \lim_{n\to\infty} (c_n)^{2^{-n}}, \tag{9}$$

where c_n is given by

$$c_0 = 2 \tag{10}$$

$$c_n = (c_{n-1})^2 + 2, \tag{11}$$

giving

$$\eta = \frac{1}{2}\sqrt{\frac{\xi}{\pi}}\sqrt{3 + \frac{1}{c_1} + \frac{1}{c_1c_2} + \frac{1}{c_1c_2c_3} + \ldots}. \tag{12}$$

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/otter/otter.html`.

Harary, F. *Graph Theory.* Reading, MA: Addison-Wesley, 1969.

Harary, F. and Palmer, E. M. *Graphical Enumeration.* New York: Academic Press, 1973.

Knuth, D. E. *The Art of Computer Programming, Vol. 1: Fundamental Algorithms, 2nd ed.* Reading, MA: Addison-Wesley, 1973.

Otter, R. "The Number of Trees." *Ann. Math.* **49**, 583–599, 1948.

Weakly Complete Sequence

A SEQUENCE of numbers $V = \{\nu_n\}$ is said to be weakly complete if every POSITIVE INTEGER n beyond a certain point N is the sum of some SUBSEQUENCE of V (Honsberger 1985). Dropping two terms from the FIBONACCI NUMBERS produces a SEQUENCE which is not even weakly complete. However, the SEQUENCE

$$F_n' \equiv F_n - (-1)^n$$

is weakly complete, even with any finite subsequence deleted (Graham 1964).

see also COMPLETE SEQUENCE

References

Graham, R. "A Property of Fibonacci Numbers." *Fib. Quart.* **2**, 1–10, 1964.

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., p. 128, 1985.

Weakly Independent

An infinite sequence $\{a_i\}$ of POSITIVE INTEGERS is called weakly independent if any relation $\sum \epsilon_i a_i$ with $\epsilon_i = 0$ or ± 1 and $\epsilon_i = 0$, except finitely often, IMPLIES $\epsilon_i = 0$ for all i.

see also STRONGLY INDEPENDENT

References

Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 136, 1994.

Weakly Triple-Free Set

see TRIPLE-FREE SET

Web Graph

A graph formed by connecting several concentric WHEEL GRAPHS along spokes.

see also WHEEL GRAPH

Weber Differential Equations

Consider the differential equation satisfied by

$$w = z^{-1/2} W_{k,-1/4}(\tfrac{1}{2}z^2), \tag{1}$$

where W is a WHITTAKER FUNCTION.

$$\frac{d}{z\,dz}\left[\frac{d(wz^{1/2})}{z\,dz}\right] + \left(-\frac{1}{4} + \frac{2k}{z^2} + \frac{3}{4z^4}\right) wz^{1/2} = 0 \tag{2}$$

$$\frac{d^2 w}{dz^2} + (2k - \tfrac{1}{4}z^2)w = 0. \tag{3}$$

This is usually rewritten

$$\frac{d^2 D_n(z)}{dz^2} + (n + \tfrac{1}{2} - \tfrac{1}{4}z^2)D_n(z) = 0. \tag{4}$$

The solutions are PARABOLIC CYLINDER FUNCTIONS.

The equations

$$\frac{d^2 U}{du^2} - (c + k^2 u^2)U = 0 \tag{5}$$

$$\frac{d^2 V}{dv^2} + (c - k^2 v^2)V = 0, \tag{6}$$

which arise by separating variables in LAPLACE'S EQUATION in PARABOLIC CYLINDRICAL COORDINATES, are also known as the Weber differential equations. As above, the solutions are known as PARABOLIC CYLINDER FUNCTIONS.

Weber's Discontinuous Integrals

$$\int_0^\infty J_0(ax)\cos(cx)\,dx = \begin{cases} 0 & a < c \\ \frac{1}{\sqrt{a^2-c^2}} & a > c \end{cases}$$

$$\int_0^\infty J_0(ax)\sin(cx)\,dx = \begin{cases} \frac{1}{\sqrt{c^2-a^2}} & a < c \\ 0 & a > c, \end{cases}$$

where $J_0(z)$ is a zeroth order BESSEL FUNCTION OF THE FIRST KIND.

References

Bowman, F. *Introduction to Bessel Functions.* New York: Dover, pp. 59–60, 1958.

Weber's Formula

$$\frac{1}{2p^2} e^{-(a^2+b^2)/(4p^2)} I_\nu\left(\frac{ab}{2p^2}\right) = \int_0^\infty e^{-p^2t^2} J_\nu(at) J_\nu(bt) t\,dt,$$

where $\Re[\nu] > -1$, $|\arg p| < \pi/4$, and a, $b > 0$, $J_\nu(z)$ is a BESSEL FUNCTION OF THE FIRST KIND, and $I_\nu(z)$ is a MODIFIED BESSEL FUNCTION OF THE FIRST KIND.

see also BESSEL FUNCTION OF THE FIRST KIND, MODIFIED BESSEL FUNCTION OF THE FIRST KIND

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1476, 1980.

Weber Functions

Although BESSEL FUNCTIONS OF THE SECOND KIND are sometimes called Weber functions, Abramowitz and Stegun (1972) define a separate Weber function as

$$\mathcal{E}_\nu(z) = \frac{1}{\pi}\int_0^\pi \sin(\nu\theta - z\sin\theta)\,d\theta. \tag{1}$$

Letting $\zeta_n = e^{2\pi i/m}$ be a ROOT OF UNITY, another set of Weber functions is defined as

$$f(z) = \frac{\eta(\frac{1}{2}(z+1))}{\zeta_{48}\eta(z)} \tag{2}$$

$$f_1(z) = \frac{\eta(\frac{1}{2}z)}{\eta(z)} \tag{3}$$

$$f_2(z) = \sqrt{2}\frac{\eta(2z)}{\eta(z)} \tag{4}$$

$$\gamma_2 = \frac{f^{24}(z) - 16}{f^8(z)} \tag{5}$$

$$\gamma_3 = \frac{[f^{24}(z) + 8][f_1{}^8(z) - f_2{}^8(z)]}{f^8(z)} \tag{6}$$

(Weber 1902, Atkin and Morain 1993), where $\eta(z)$ is the DEDEKIND ETA FUNCTION. The Weber functions satisfy the identities

$$f(z+1) = \frac{f_1(z)}{\zeta_{48}} \tag{7}$$

$$f_1(z+1) = \frac{f(z)}{\zeta_{48}} \tag{8}$$

$$f_2(z+1) = \zeta_{24} f_2(z) \tag{9}$$

$$f\left(-\frac{1}{z}\right) = f(z) \tag{10}$$

$$f_1\left(-\frac{1}{z}\right) = f_2(z) \tag{11}$$

$$f_2\left(-\frac{1}{z}\right) = f_1(z) \tag{12}$$

(Weber 1902, Atkin and Morain 1993).

see also ANGER FUNCTION, BESSEL FUNCTION OF THE SECOND KIND, DEDEKIND ETA FUNCTION, j-FUNCTION, JACOBI IDENTITIES, JACOBI TRIPLE PRODUCT, MODIFIED STRUVE FUNCTION, Q-FUNCTION, STRUVE FUNCTION

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Anger and Weber Functions." §12.3 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 498–499, 1972.

Atkin, A. O. L. and Morain, F. "Elliptic Curves and Primality Proving." *Math. Comput.* **61**, 29–68, 1993.

Borwein, J. M. and Borwein, P. B. *Pi & the AGM: A Study in Analytic Number Theory and Computational Complexity.* New York: Wiley, pp. 68–69, 1987.

Weber, H. *Lehrbuch der Algebra, Vols. I–II.* New York: Chelsea, pp. 113–114, 1902.

Weber-Sonine Formula

For $\Re[\mu + nu] > 0$, $|\arg p| < \pi/4$, and $a > 0$,

$$\int_0^\infty J_\nu(at)e^{-p^2t^2}t^{\mu-1}\,dt$$

$$\left(\frac{a}{2p}\right)^\nu \frac{\Gamma[\frac{1}{2}(\nu+\mu)]}{2p^\mu\Gamma(\nu+1)}\,{}_1F_1\left(\tfrac{1}{2}(\nu+\mu);\nu+1;-\frac{a^2}{2p^2}\right),$$

where $J_\nu(z)$ is a BESSEL FUNCTION OF THE FIRST KIND, $\Gamma(z)$ is the GAMMA FUNCTION, and ${}_1F_1(a;b;z)$ is a CONFLUENT HYPERGEOMETRIC FUNCTION.

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1474, 1980.

Weber's Theorem

If two curves of the same GENUS (CURVE) > 1 are in rational correspondence, then that correspondence is BIRATIONAL.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 135, 1959.

Wedderburn's Theorem

A FINITE DIVISION RING is a FIELD.

Weddle's Rule

$$\int_{x_1}^{x_{6n}} f(x)\,dx = \tfrac{3}{10}h(f_1 + 5f_2 + f_3$$
$$+6f_4 + 5f_5 + f_6 + \ldots + 5f_{6n-1} + f_{6n})$$

see also BODE'S RULE, HARDY'S RULE, NEWTON-COTES FORMULAS, SIMPSON'S 3/8 RULE, SIMPSON'S RULE, TRAPEZOIDAL RULE, WEDDLE'S RULE

Wedge

A right triangular PRISM turned so that it rests on one of its lateral faces.

see also CONICAL WEDGE, CYLINDRICAL WEDGE, PRISM

Wedge Product

An antisymmetric operation on DIFFERENTIAL FORMS (also called the EXTERIOR DERIVATIVE)

$$dx_i \wedge dx_j = -dx_j \wedge dx_i, \tag{1}$$

which IMPLIES

$$\begin{aligned} dx_i \wedge dx_i &= 0 && (2)\\ b_i \wedge dx_j &= dx_j \wedge b_i = b_i\,dx_j && (3)\\ dx_i \wedge (b_i\,dx_j) &= b_i\,dx_i \wedge dx_j && (4)\\ \theta_1 \wedge \theta_2 &= (b_1\,dx_1 + b_2\,dx_2) \wedge (c_1\,dx_1 + c_2\,dx_2)\\ &= (b_1c_2 - b_2c_1)\,dx_1 \wedge dx_2\\ &= -\theta_2 \wedge \theta_1. && (5) \end{aligned}$$

The wedge product is ASSOCIATIVE

$$(s \wedge t) \wedge u = s \wedge (t \wedge u), \tag{6}$$

and BILINEAR

$$(\alpha_1 s_1 + \alpha_2 s_2) \wedge t = \alpha_1(s_1 \wedge t) + \alpha_2(s_2 \wedge t) \tag{7}$$

$$s \wedge (\alpha_1 t_1 + \alpha_2 t_2) = \alpha_1(s \wedge t_1) + \alpha_2(s \wedge t_2), \tag{8}$$

but not (in general) COMMUTATIVE

$$s \wedge t = (-1)^{pq}(t \wedge s), \tag{9}$$

where s is a p-form and t is a q-form. For a 0-form s and 1-form t,

$$(s \wedge t)_\mu = st_\mu. \tag{10}$$

For a 1-form s and 1-form t,

$$(s \wedge t)_{\mu\nu} = \tfrac{1}{2}(s_\mu t_\nu - s_\nu t_\mu). \tag{11}$$

The wedge product is the "correct" type of product to use in computing a VOLUME ELEMENT

$$dV = dx_1 \wedge \ldots \wedge dx_n. \tag{12}$$

see also DIFFERENTIAL FORM, EXTERIOR DERIVATIVE, INNER PRODUCT, VOLUME ELEMENT

Weekday

The day of the week W for a given day of the month D, month M, and year $100C + Y$ can be determined from the simple equation

$$W \equiv D + \lfloor 2.6M - 0.2 \rfloor + \left\lfloor \tfrac{1}{4}Y \right\rfloor + \left\lfloor \tfrac{1}{4}C \right\rfloor - 2C \pmod 7,$$

where *months are numbered beginning with March* and $W = 0$ for Sunday, $W = 1$ for Monday, etc. (Uspensky and Heaslet 1939, Vardi 1991).

A more complicated form is given by

$$W \equiv D + M + C + Y \pmod 7,$$

where $W = 1$ for Sunday, $W = 2$ for Monday, etc. and the numbers assigned to months, centuries, and years are given in the tables below (Kraitchik 1942, pp. 110–111).

Month	M
January	1
February	4
March	3
April	6
May	1
June	4
July	6
August	2
September	5
October	0
November	3
December	5

Gregorian Century	C
15, 19, 23	1
16, 20, 24	0
17, 21, 25	5
18, 22, 26	3

Julian Century	C
00, 07, 14	5
01, 08, 15	4
02, 09, 16	3
03, 10, 17	2
04, 11, 18	1
05, 12, 19	0
06, 13, 20	6

Year									Y
00	06		17	23	28	34		45	0
01	07	12	18		29	35	40	46	1
02		13	19	24	30		41	47	2
03	08	14		25	31	36	42		3
	09	15	20	26		37	43	48	4
04	10		21	27	32	38		49	5
05	11	16	22		33	39	44	50	6
51	56	62		73	79	84	90		0
	57	63	68	74		85	91	96	1
52	58		69	75	80	86		97	2
53	59	64	70		81	87	92	98	3
54		65	71	76	82		93	99	4
55	60	66		77	83	88	94		5
	61	67	72	78		89	95		6

see also FRIDAY THE THIRTEENTH

References

Kraitchik, M. "The Calendar." Ch. 5 in *Mathematical Recreations.* New York: W. W. Norton, pp. 109–116, 1942.

Uspensky, J. V. and Heaslet, M. A. *Elementary Number Theory.* New York: McGraw-Hill, pp. 206–211, 1939.

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 237–238, 1991.

Weibull Distribution

The Weibull distribution is given by

$$P(x) = \alpha\beta^{-\alpha}x^{\alpha-1}e^{-(x/\beta)^\alpha} \tag{1}$$

$$D(x) = 1 - e^{-(x/\beta)^\alpha} \tag{2}$$

for $x \in [0, \infty)$ (*Mathematica*® `Statistics'ContinuousDistributions'WeibullDistribution[a,b]`, Wolfram Research, Champaign, IL). The MEAN, VARIANCE, SKEWNESS, and KURTOSIS of this distribution are

$$\mu = \beta\Gamma(1+\alpha^{-1}) \tag{3}$$

$$\sigma^2 = \beta^2[\Gamma(1+2\alpha^{-1}) - \Gamma^2(1+\alpha^{-1})] \tag{4}$$

$$\gamma_1 = \frac{2\Gamma^3(1+\alpha^{-1}) - 3\Gamma(1+\alpha^{-1})\Gamma(1+2\alpha^{-1})}{[\Gamma(1+2\alpha^{-1}) - \Gamma^2(1+\alpha^{-1})]^{3/2}} + \frac{\Gamma(1+3\alpha^{-1})}{[\Gamma(1+2\alpha^{-1}) - \Gamma^2(1+\alpha^{-1})]^{3/2}} \tag{5}$$

$$\gamma_2 = \frac{f(a)}{[\Gamma(1+2\alpha^{-1}) - \Gamma^2(1+\alpha^{-1})]^2}, \tag{6}$$

where $\Gamma(z)$ is the GAMMA FUNCTION and

$$\begin{aligned} f(a) \equiv -6\Gamma^4(1+\alpha^{-1}) + 12\Gamma^2(1+\alpha^{-1})\Gamma(1+2\alpha^{-1}) \\ -3\Gamma^2(1+2\alpha^{-1}) - 4\Gamma(1+\alpha^{-1})\Gamma(1+3\alpha^{-1}) \\ +\Gamma(1+4\alpha^{-1}). \end{aligned} \tag{7}$$

A slightly different form of the distribution is

$$P(x) = \frac{\alpha}{\beta}x^{\alpha-1}e^{-x^\alpha/\beta} \tag{8}$$

$$D(x) = 1 - e^{-x^\alpha/\beta} \tag{9}$$

(Mendenhall and Sincich 1995). The MEAN and VARIANCE for this form are

$$\mu = \beta^{1/\alpha}\Gamma(1+\alpha^{-1}) \tag{10}$$

$$\sigma^2 = \beta^{2/\alpha}[\Gamma(1+2\alpha^{-1}) - \Gamma^2(1+\alpha^{-1})]. \tag{11}$$

The Weibull distribution gives the distribution of lifetimes of objects. It was originally proposed to quantify fatigue data, but it is also used in analysis of systems involving a "weakest link."

see also FISHER-TIPPETT DISTRIBUTION

References

Mendenhall, W. and Sincich, T. *Statistics for Engineering and the Sciences, 4th ed.* Englewood Cliffs, NJ: Prentice Hall, 1995.

Spiegel, M. R. *Theory and Problems of Probability and Statistics.* New York: McGraw-Hill, p. 119, 1992.

Weierstraß Approximation Theorem

If f is continuous on $[a, b]$, then there exists a POLYNOMIAL p on $[a, b]$ such that

$$|f(x) - P(x)| < \epsilon$$

for all $x \in [a, b]$ and $\epsilon > 0$. In words, any continuous function on a closed and bounded interval can be uniformly approximated on that interval by POLYNOMIALS to any degree of accuracy.

see also MÜNTZ'S THEOREM

Weierstraß-Casorati Theorem

An ANALYTIC FUNCTION approaches any given value arbitrarily closely in any ϵ-NEIGHBORHOOD of an ESSENTIAL SINGULARITY.

Weierstraß Constant

$$\sigma(\tfrac{1}{2}) = \tfrac{1}{2} \prod_{\substack{(m,n)\neq\\(0,0)}} \left[1 - \frac{1}{2(m+ni)}\right] \times e^{1/[2(m+ni)]+1/[8(m+ni)^2]}$$
$$= \frac{2^{5/4}\sqrt{\pi}\, e^{\pi/8}}{\Gamma^2(\frac{1}{4})} = 0.4749493799\ldots.$$

References

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, p. 62, 1983.

Plouffe, S. "Weierstrass Constant." `http://lacim.uqam.ca/piDATA/weier.txt`.

Waldschmidt, M. "Fonctions entières et nombres transcendants." *Cong. Nat. Soc. Sav. Nancy* **5**, 1978.

Waldschmidt, M. "Nombres transcendants et fonctions sigma de Weierstrass." *C. R. Math. Rep. Acad. Sci. Canada* **1**, 111–114, 1978/79.

Weierstraß Elliptic Function

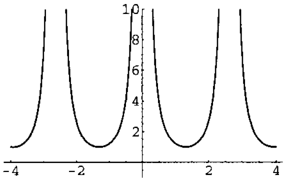

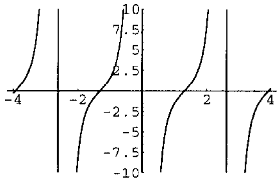

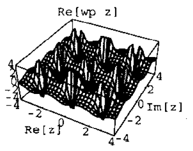

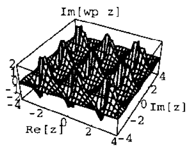

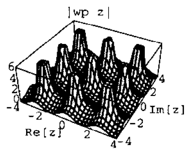

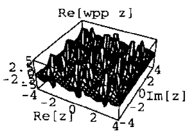

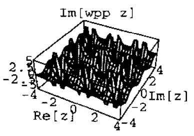

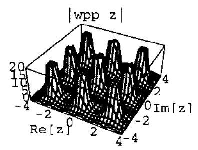

The Weierstraß elliptic functions are elliptic functions which, unlike the JACOBI ELLIPTIC FUNCTIONS, have a second-order POLE at $z = 0$. The above plots show the Weierstraß elliptic function $\wp(z)$ and its derivative $\wp'(z)$ for invariants (defined below) of $g_2 = 0$ and $g_3 = 0$. Weierstraß elliptic functions are denoted $\wp(z)$ and can be defined by

$$\wp(z) = \frac{1}{z^2} + \sum_{m,n=-\infty}^{\infty}{}' \left[\frac{1}{(z - 2m\omega_1 - 2n\omega_2)^2} - \frac{1}{(2m\omega_1 + 2n\omega_2)^2}\right]. \quad (1)$$

Write $\Omega_{mn} \equiv 2m\omega_1 + 2n\omega_2$. Then this can be written

$$\wp(z) = z^{-2} + \sum_{m,n}{}' [(z - \Omega_{mn})^{-2} - \Omega_{mn}^{-2}]. \quad (2)$$

An equivalent definition which converges more rapidly is

$$\wp(z) = \left(\frac{\pi}{2\omega_1}\right)^2 \left[-\frac{1}{3} + \sum_{-\infty}^{\infty} \csc^2\left(\frac{z - 2n\omega_2}{2\omega_1}\pi\right) - \sum_{n=-\infty}^{\infty}{}' \csc^2\left(\frac{n\omega_2}{n_1}\pi\right)\right]. \quad (3)$$

$\wp(z)$ is an EVEN FUNCTION since $\wp(-z)$ gives the same terms in a different order. To specify $\wp$ completely, its periods or invariants, written $\wp(z|\omega_1, \omega_2)$ and $\wp(z; g_2, g_3)$, respectively, must also be specified.

The differential equation from which Weierstraß elliptic functions arise can be found by expanding about the origin the function $f(z) \equiv \wp(z) - z^{-2}$.

$$\wp(z) - z^{-2} = f(0) + f'(0)z + \tfrac{1}{2!}f''(0)z^2 + \tfrac{1}{3!}f'''(0)z^3 + \tfrac{1}{4}f^{(4)}(0)z^4 + \ldots. \quad (4)$$

But $f(0) = 0$ and the function is even, so $f'(0) = f'''(0) = 0$ and

$$f(z) = \wp(z) - z^{-2} = \tfrac{1}{2!}f''(0)z^2 + \tfrac{1}{4}f^{(4)}(0)z^4 + \ldots. \quad (5)$$

Taking the derivatives

$$f' = -2\Sigma'[(z - \Omega_{mn})^{-3}] \quad (6)$$
$$f'' = 6\Sigma'(z - \Omega_{mn})^{-4} \quad (7)$$
$$f''' = -24\Sigma'(z - \Omega_{mn})^{-5} \quad (8)$$
$$f^{(4)} = 120\Sigma'(z - \Omega_{mn})^{-6}. \quad (9)$$

So

$$f''(0) = 6\Sigma'\Omega_{mn}^{-4} \quad (10)$$
$$f^{(4)}(0) = 120\Sigma'\Omega_{mn}^{-6}. \quad (11)$$

Plugging in,

$$\wp(z) - z^{-2} = 3\Sigma'\Omega_{mn}^{-4}z^2 + 5\Sigma'\Omega_{mn}^{-6}z^4 + \mathcal{O}(z^6). \quad (12)$$

Define the INVARIANTS

$$g_2 \equiv 60\Sigma'\Omega_{mn}^{-4} \quad (13)$$
$$g_3 \equiv 140\Sigma'\Omega_{mn}^{-6}, \quad (14)$$

then

$$\wp(z) = z^{-2} + \tfrac{1}{20}g_2z^2 + \tfrac{1}{28}g_3z^4 + \mathcal{O}(z^6) \quad (15)$$

$$\wp'(z) = -2z^{-3} + \tfrac{1}{10}g_2z + \tfrac{1}{7}g_3z^3 + \mathcal{O}(z^5). \quad (16)$$

Now cube (15) and square (16)

$$\wp^3(z) = z^{-6} + \tfrac{3}{20}g_2z^{-2} + \tfrac{3}{28}g_3 + \mathcal{O}(z^2) \quad (17)$$

$$\wp'^2(z) = 4z^{-6} - \tfrac{2}{5} g_2 z^{-2} - \tfrac{4}{7} g_3 + \mathcal{O}(z^2). \tag{18}$$

Taking (18) – 4 × (17) cancels out the z^{-6} term, giving

$$\begin{aligned} \wp'^2(z) - 4\wp^3(z) & \\ &= \left(-\tfrac{2}{5} - \tfrac{3}{5}\right) g_2 z^{-2} + \left(-\tfrac{4}{7} - \tfrac{3}{7}\right) g_3 + \mathcal{O}(z^2) \\ &= -g_2 z^{-2} - g_3 + \mathcal{O}(z^2) \end{aligned} \tag{19}$$

$$\wp'^2(z) - 4\wp^3(z) + g_2 z^{-2} + g_3 = \mathcal{O}(z^2). \tag{20}$$

But, from (5)

$$\wp(z) = z^{-2} + \tfrac{1}{2!} f''(0) z^2 + \tfrac{1}{4} f^{(4)}(0) z^4 + \ldots, \tag{21}$$

so $\wp(z) = z^{-2} + \mathcal{O}(z^2)$ and (20) can be written

$$\wp'^2(z) - 4\wp^3(z) + g_2 \wp(z) + g_3 = \mathcal{O}(z^2). \tag{22}$$

The Weierstraß elliptic function is analytic at the origin and therefore at all points congruent to the origin. There are no other places where a singularity can occur, so this function is an ELLIPTIC FUNCTION with no SINGULARITIES. By LIOUVILLE'S ELLIPTIC FUNCTION THEOREM, it is therefore a constant. But as $z \to 0$, $\mathcal{O}(z^2) \to 0$, so

$$\wp'^2(z) = 4\wp^3(z) - g_2 \wp(z) - g_3. \tag{23}$$

The solution to the differential equation

$$y'^2 = 4y^3 - g_2 y - g_3 \tag{24}$$

is therefore given by $y = \wp(z + \alpha)$, providing that numbers ω_1 and ω_2 exist which satisfy the equations defining the INVARIANTS. Writing the differential equation in terms of its roots e_1, e_2, and e_3,

$$y'^2 = 4y^3 - g_2 y - g_3 = 4(y - e_1)(y - e_2)(y - e_3) \tag{25}$$

$$2\ln(y') = \ln 4 + \sum_{r=1}^{3} \ln(y - e_r) \tag{26}$$

$$\frac{2y''}{y'} = y' \sum_{r=1}^{3} (y - e_r)^{-1} \tag{27}$$

$$\frac{2y''}{y'^2} = \sum_{r=1}^{3} (y - e_r)^{-1} \tag{28}$$

$$2\frac{y'^2 y''' - y''(2y'y'')}{y'^4} = -y' \sum_{r=1}^{3} (y - e_r)^{-2} \tag{29}$$

$$\frac{2y'''}{y'^3} - \frac{4y''^2}{y'^4} = -\sum_{r=1}^{3} (y - e_r)^{-2}. \tag{30}$$

Now take $(30)/4 + [(30)/4]^2$,

$$\begin{aligned} \left[\frac{y'''}{2y'^3} - \frac{y''^2}{y'^4}\right] + \left[\frac{y''^2}{4y'^4}\right] & \\ &= -\frac{1}{4} \sum_{r=1}^{3} (y - e_r)^{-2} + \frac{1}{16} \left[\sum_{r=1}^{3} (y - e_r)^{-1}\right]^2 \end{aligned} \tag{31}$$

$$\frac{3y''^2}{4y'^4} - \frac{y'''}{2y'^3} = \tfrac{3}{16} \sum_{r=1}^{3} (y - e_r)^{-2} - \tfrac{3}{8} y \prod_{r=1}^{3} (y - e_r)^{-1}. \tag{32}$$

The term on the right is half the SCHWARZIAN DERIVATIVE.

The DERIVATIVE of the Weierstraß elliptic function is given by

$$\begin{aligned} \wp'(z) &= \frac{d}{dz} \wp(z) = -2 \sum_{m,n} \frac{1}{(z - \Omega_{mn})^3} \\ &= -2z^{-3} - 2 \sum_{m,n}{}' (z - \Omega_{mn})^{-3}. \end{aligned} \tag{33}$$

This is an ODD FUNCTION which is itself an elliptic function with pole of order 3 at $z = 0$. The INTEGRAL is given by

$$z = \int_{\wp(z)}^{\infty} (4t^3 - g_2 t - g_3)^{-1/2} \, dt. \tag{34}$$

A duplication formula is obtained as follows.

$$\begin{aligned} \wp(2z) &= \lim_{y \to z} \wp(y + z) = \frac{1}{4} \lim_{y \to z} \left[\frac{\wp'(z) - \wp'(y)}{\wp(z) - \wp(y)}\right]^2 \\ &\quad - \wp(z) - \lim_{y \to z} \wp(y) \\ &= \frac{1}{4} \lim_{h \to 0} \left[\frac{\wp(z) - \wp'(z + h)}{\wp(z) - \wp(z + h)}\right]^2 - 2\wp(z) \\ &= \frac{1}{4} \left\{ \left[\lim_{h \to 0} \frac{\wp'(z) - \wp'(z + h)}{h}\right] \left[\lim_{h \to 0} \frac{h}{\wp(z) - \wp(z + h)}\right] \right\}^2 \\ &\quad - 2\wp(z) \\ &= \frac{1}{4} \left[\frac{\wp''(z)}{\wp'(z)}\right]^2 - 2\wp(z). \end{aligned} \tag{35}$$

A general addition theorem is obtained as follows. Given

$$\wp'(z) = A\wp(z) + B \tag{36}$$

$$\wp'(y) = A\wp(y) + B \tag{37}$$

with zero y and z where $z \not\equiv \pm y \pmod{2\omega_1, 2\omega_2}$, find the third zero ζ. Consider $\wp'(\zeta) - A\wp(\zeta) - B$. This has a pole of order three at $\zeta = 0$, but the sum of zeros (= 0) equals the sum of poles for an ELLIPTIC FUNCTION, so $z + y + \zeta = 0$ and $\zeta = -z - y$.

$$\wp'(-z - y) = A\wp(-z - y) + B \tag{38}$$

$$-\wp'(z+y) = A\wp(z+y)+B. \tag{39}$$

Combining (36), (37), and (39) gives

$$\begin{bmatrix} \wp(z) & \wp'(z) & 1 \\ \wp(y) & \wp'(y) & 1 \\ \wp(z+y) & -\wp(z+y) & 1 \end{bmatrix} \begin{bmatrix} A \\ -1 \\ B \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \tag{40}$$

so

$$\begin{vmatrix} \wp(z) & \wp'(z) & 1 \\ \wp(y) & \wp'(y) & 1 \\ \wp(z+y) & -\wp(z+y) & 1 \end{vmatrix} = 0. \tag{41}$$

Defining $u+v+w=0$ where $u \equiv z$ and $v \equiv y$ gives the symmetric form

$$\begin{vmatrix} \wp(u) & \wp'(u) & 1 \\ \wp(v) & \wp'(v) & 1 \\ \wp(w) & \wp(w) & 1 \end{vmatrix} = 0. \tag{42}$$

To get the expression explicitly, start again with

$$\wp'(\zeta) - A\wp(\zeta) - B = 0, \tag{43}$$

where $\zeta = z, y, -z-y$.

$$\wp'^2(\zeta) - [A\wp(\zeta)+B]^2 = 0. \tag{44}$$

But $\wp^2(\zeta) = 4\wp^4(\zeta) - g_2\wp(\zeta) - g_3$, so

$$4\wp^3(\zeta) - A^2\wp^2(\zeta) - (2AB+g_2)\wp(\zeta) - (B^2+g_3) = 0. \tag{45}$$

The solutions $\wp(\zeta) \equiv z$ are given by

$$4z^3 - A^2z^2 - (2AB+g_2)z - (B^2+g_3) = 0. \tag{46}$$

But the sum of roots equals the COEFFICIENT of the squared term, so

$$\wp(z)+\wp(y)+\wp(z+y) = \tfrac{1}{4}A^2 \tag{47}$$

$$\wp'(z) - \wp'(y) = A[\wp(z)-\wp(y)] \tag{48}$$

$$A = \frac{\wp'(z)-\wp'(y)}{\wp(z)-\wp(y)} \tag{49}$$

$$\wp(z+y) = \frac{1}{4}\left[\frac{\wp(z)-\wp'(y)}{\wp(z)-\wp(y)}\right]^2 - \wp(z) - \wp(y). \tag{50}$$

Half-period identities include

$$\begin{aligned} x \equiv \wp(\tfrac{1}{2}\omega_1) = \wp(-h\omega_1+\omega_1) &= e_1 + \frac{(e_1-e_2)(e_1-e_3)}{\wp(-\tfrac{1}{2}\omega_1)-e_1} \\ &= e_1 + \frac{(e_1-e_2)(e_1-e_3)}{x-e_1}. \end{aligned} \tag{51}$$

Multiplying through,

$$x^2 - e_1x = e_1x - e_1{}^2 + (e_1-e_2)(e_1-e_3) \tag{52}$$

$$x^2 - 2e_1 + [e_1{}^2 - (e_1-e_2)(e_1-e_3)] = 0, \tag{53}$$

which gives

$$\begin{aligned} \wp(\tfrac{1}{2}\omega_1) &= \tfrac{1}{2}\left\{2e_1 \pm \sqrt{4e_1{}^2 - 4[e_1{}^2-(e_1-e_2)(e_1-e_3)]}\right\} \\ &= e_1 \pm \sqrt{(e_1-e_2)(e_1-e_3)}. \end{aligned} \tag{54}$$

From Whittaker and Watson (1990, p. 445),

$$\begin{aligned} \wp'(\tfrac{1}{2}\omega_1) = -2\sqrt{(e_1-e_2)(e_1-e_3)} \\ \times(\sqrt{e_1-e_2}+\sqrt{e_1-e_3}). \end{aligned} \tag{55}$$

The function is HOMOGENEOUS,

$$\wp(\lambda z|\lambda\omega_1, \lambda\omega_2) = \lambda^{-2}\wp(z|\omega_1,\omega_2) \tag{56}$$

$$\wp(\lambda z; \lambda^{-4}g_2, \lambda^{-6}g_3) = \lambda^{-2}\wp(z;g_2,g_3). \tag{57}$$

To invert the function, find $2\omega_1$ and $2\omega_2$ of $\wp(z|\omega_1,\omega_2)$ when given $\wp(z;g_2,g_3)$. Let e_1, e_2, and e_3 be the roots such that $(e_1-e_2)/(e_1-e_3)$ is not a REAL NUMBER > 1 or < 0. Determine the PARAMETER τ from

$$\frac{e_1-e_2}{e_1-e_3} = \frac{\vartheta_4{}^4(0|\tau)}{\vartheta_3{}^4(0|\tau)}. \tag{58}$$

Now pick

$$A \equiv \frac{\sqrt{e_1-e_2}}{\vartheta_4{}^2(0|\tau)}. \tag{59}$$

As long as $g_2{}^3 \neq 27g_3$, the periods are then

$$2\omega_1 = \pi A \tag{60}$$

$$2\omega_2 = \frac{\pi\tau}{A}. \tag{61}$$

Weierstraß elliptic functions can be expressed in terms of JACOBI ELLIPTIC FUNCTIONS by

$$\begin{aligned} \wp(u;g_2,g_3) = e_3 + (e_1-e_3) \\ \times \mathrm{ns}^2\left(u\sqrt{e_1-e_3}, \sqrt{\frac{e_2-e_3}{e_1-e_3}}\right), \end{aligned} \tag{62}$$

where

$$\wp(\omega_1) = e_1 \tag{63}$$

$$\wp(\omega_2) = e_2 \tag{64}$$

$$\wp(\omega_3) = -\wp(-\omega_1-\omega_2) = e_3, \tag{65}$$

and the INVARIANTS are

$$g_2 \equiv 60\Sigma'\Omega_{mn}^{-4} \tag{66}$$

$$g_3 \equiv 140\Sigma'\Omega_{mn}^{-6}. \tag{67}$$

Here, $\Omega_{mn} \equiv 2m\omega_1 - 2n\omega_2$.

An addition formula for the Weierstraß elliptic function can be derived as follows.

$$\wp(z+\omega_1)+\wp(z)+\wp(\omega_1)$$
$$=\frac{1}{4}\left[\frac{\wp'(z)-\wp'(\omega_1)}{\wp(z)-\wp(\omega_1)}\right]^2=\frac{1}{4}\frac{\wp'^2(z)}{[\wp(z)-e_1]^2}. \quad (68)$$

Use

$$\wp'(z)=4\prod_{r=1}^{3}[\wp(z)-e_r], \quad (69)$$

so

$$\begin{aligned}\wp(z+\omega_1)&=-\wp(z)-e_1+\frac{1}{4}\frac{4\prod_{r=1}^{3}[\wp(z)-e_r]}{[\wp(z)-e_1]^2}\\&=-\wp(z)-e_1+\frac{[\wp(z)-e_2][\wp(z)-e_3]}{\wp(z)-e_1}.\end{aligned} \quad (70)$$

Use $\sum_{r=1}^{3}e_r=0$,

$$\begin{aligned}\wp(z+\omega_1)&=e_1+\frac{[-2e_1-\wp(z)][\wp(z)-e_1]}{\wp(z)-e_1}\\&\quad+\frac{\wp^2(z)-\wp(z)(e_2+e_3)+e_2e_3}{\wp(z)-e_1}\\&=e_1+\frac{-\wp(z)(e_1+e_2+e_3)+e_2e_3+2{e_1}^2}{\wp(z)-e_1}.\end{aligned} \quad (71)$$

But $\sum_{r=1}^{3}e_r=0$ and

$$2{e_1}^2+e_2e_3={e_1}^2-e_1(e_2+e_3)+e_2e_3=(e_1-e_2)(e_1-e_3), \quad (72)$$

so

$$\wp(z+\omega_1)=e_1+\frac{(e_1-e_2)(e_1-e_3)}{\wp(z)-e_1}. \quad (73)$$

The periods of the Weierstraß elliptic function are given as follows. When g_2 and g_3 are REAL and ${g_2}^3-27{g_3}^2>0$, then e_1, e_2, and e_3 are REAL and defined such that $e_1>e_2>e_3$.

$$\omega_1=\int_{e_1}^{\infty}(4t^3-g_2t-g_3)^{-1/2}\,dt \quad (74)$$

$$\omega_3=-i\int_{-\infty}^{e_3}(g_3+g_2t-4t^3)^{-1/2}\,dt \quad (75)$$

$$\omega_2=-\omega_1-\omega_3. \quad (76)$$

The roots of the Weierstraß elliptic function satisfy

$$e_1=\wp(\omega_1) \quad (77)$$

$$e_2=\wp(\omega_2) \quad (78)$$

$$e_3=\wp(\omega_3), \quad (79)$$

where $\omega_3\equiv-\omega_1-\omega_2$. The e_is are ROOTS of $4t^3-g_2t-g_3$ and are unequal so that $e_1\neq e_2\neq e_3$. They can be found from the relationships

$$e_1+e_2+e_3=-a_2=0 \quad (80)$$

$$e_2e_3+e_3e_1+e_1e_2=a_1=-\tfrac{1}{4}g_2 \quad (81)$$

$$e_1e_2e_3=-a_0=\tfrac{1}{4}g_3. \quad (82)$$

see also EQUIANHARMONIC CASE, LEMNISCATE CASE, PSEUDOLEMNISCATE CASE

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Weierstrass Elliptic and Related Functions." Ch. 18 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 627–671, 1972.

Fischer, G. (Ed.). Plates 129–131 in *Mathematische Modelle/Mathematical Models, Bildband/Photograph Volume.* Braunschweig, Germany: Vieweg, pp. 126–128, 1986.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Weierstraß-Erdman Corner Condition

In the CALCULUS OF VARIATIONS, the condition

$$f_{y'}(x,y,y'(x_-))=f_{y'}(x,y,y'(x_+))$$

must hold at a corner (x,y) of a minimizing arc E_{12}.

Weierstraß Extreme Value Theorem

see EXTREME VALUE THEOREM

Weierstraß Form

A general form into which an ELLIPTIC CURVE over any FIELD K can be transformed is called the Weierstraß form, and is given by

$$y^2+ay=x^3+bx^2+cxy+dx+e,$$

where a, b, c, d, and e are elements of K.

Weierstraß Function

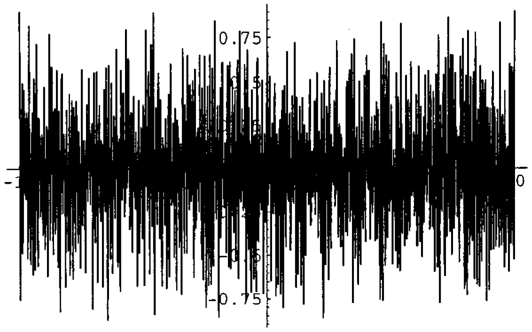

A CONTINUOUS FUNCTION which is nowhere DIFFERENTIABLE. It is given by

$$f(x) = \sum_{n=1}^{\infty} b^n \cos(a^n \pi x)$$

where n is an ODD INTEGER, $b \in (0,1)$, and $ab > 1 + 3\pi/2$. The above plot is for $a = 10$ and $b = 1/2$.

see also BLANCMANGE FUNCTION, CONTINUOUS FUNCTION, DIFFERENTIABLE

References

Darboux, G. "Mémoir sur les fonctions discontinues." *Ann. lÉcole Normale, Ser. 2* **4**, 57–112, 1875.

Darboux, G. "Mémoir sur les fonctions discontinues." *Ann. lÉcole Normale, Ser. 2* **8**, 195–202, 1879.

du Bois-Reymond, P. "Versuch einer Klassification der willkürlichen Functionen reeller Argumente nach ihren Änderungen in den kleinsten Intervallen." *J. für Math.* **79**, 21–37, 1875.

Faber, G. "Einfaches Beispiel einer stetigen nirgends differentiierbaren Funktion." *Jahresber. Deutschen Math. Verein.* **16** 538–540, 1907.

Hardy, G. H. "Weierstrass's Non-Differentiable Function." *Trans. Amer. Math. Soc.* **17**, 301–325, 1916.

Landsberg, G. "Über Differentzierbarkeit stetiger Funktionen." *Jahresber. Deutschen Math. Verein.* **17**, 46–51, 1908.

Lerch, M. "Über die Nichtdifferentiirbarkeit gewisser Functionen." *J. reine angew. Math.* **13**, 126–138, 1888.

Pickover, C. A. *Keys to Infinity.* New York: W. H. Freeman, p. 190, 1995.

Weierstraß, K. *Abhandlungen aus der Functionenlehre.* Berlin: J. Springer, p. 97, 1886.

Weierstraß's Gap Theorem

Given a succession of nonsingular points which are on a nonhyperelliptic curve of GENUS p, but are not a group of the canonical series, the number of groups of the first k which cannot constitute the group of simple POLES of a RATIONAL FUNCTION is p. If points next to each other are taken, then the theorem becomes: Given a nonsingular point of a nonhyperelliptic curve of GENUS p, then the orders which it cannot possess as the single pole of a RATIONAL FUNCTION are p in number.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 290, 1959.

Weierstraß Intermediate Value Theorem

If a continuous function defined on an interval is sometimes POSITIVE and sometimes NEGATIVE, it must be 0 at some point.

Weierstraß M-Test

Let $\sum_{k=1}^{\infty} u_n(x)$ be a SERIES of functions all defined for a set E of values of x. If there is a CONVERGENT series of constants

$$\sum_{n=1}^{\infty} M_n,$$

such that

$$|u_n(x)| \leq M_n$$

for all $x \in E$, then the series exhibits ABSOLUTE CONVERGENCE for each $x \in E$ as well as UNIFORM CONVERGENCE in E.

see also ABSOLUTE CONVERGENCE, UNIFORM CONVERGENCE

References

Arfken, G. *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 301–303, 1985.

Weierstraß Point

A POLE of multiplicity less than $p + 1$.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, pp. 290–291, 1959.

Weierstraß's Polynomial Theorem

A function, continuous in a finite close interval, can be approximated with a preassigned accuracy by POLYNOMIALS. A function of a REAL variable which is continuous and has period 2π can be approximated by trigonometric POLYNOMIALS.

References

Szegő, G. *Orthogonal Polynomials, 4th ed.* Providence, RI: Amer. Math. Soc., p. 5, 1975.

Weierstraß Product Inequality

If $0 \leq a, b, c, d \leq 1$, then

$$(1-a)(1-b)(1-c)(1-d) + a + b + c + d \geq 1.$$

References

Honsberger, R. *Mathematical Gems III.* Washington, DC: Math. Assoc. Amer., pp. 244–245, 1985.

Weierstraß Sigma Function

The QUASIPERIODIC FUNCTION defined by

$$\frac{d}{dz} \ln \sigma(z) = \zeta(z), \tag{1}$$

where $\zeta(z)$ is the WEIERSTRAß ZETA FUNCTION and

$$\lim_{z \to 0} \frac{\sigma(z)}{z} = 1. \tag{2}$$

Then

$$\sigma(z) = z \prod_{mn}{}' \left[\left(1 - \frac{z}{\Omega_{mn}}\right) \exp\left(\frac{z}{\Omega_{mn}} + \frac{z^2}{2\Omega_{mn}^2}\right) \right] \tag{3}$$

$$\sigma(z + 2\omega_1) = -e^{2\eta_1(z+\omega_1)} \sigma(z) \tag{4}$$

$$\sigma(z + 2\omega_2) = -e^{2\eta_2(z+\omega_2)} \sigma(z) \tag{5}$$

$$\sigma_r(z) = \frac{e^{-\eta_r z}\sigma(z+\omega_r)}{\sigma(\omega_r)} \tag{6}$$

for $r = 1, 2, 3$.

$$\sigma(z|\omega_1,\omega_2) = \frac{2\omega_1}{\pi\vartheta_1'}\exp\left(-\frac{\nu^2\vartheta_1'''}{6\vartheta_1'}\right)\vartheta_1\left(\nu\,\middle|\,\frac{\omega_2}{\omega_1}\right), \tag{7}$$

where $\nu \equiv \pi z/(2\omega_1)$, and

$$\eta_1 = -\frac{\pi^2\vartheta_1'''}{12\omega_1\vartheta_1'} \tag{8}$$

$$\eta_2 = -\frac{\pi^2\omega_2\vartheta_1'''}{12{\omega_1}^2\vartheta_1'} - \frac{\pi i}{2\omega_1}. \tag{9}$$

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Weierstrass Elliptic and Related Functions." Ch. 18 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 627–671, 1972.

Weierstraß's Theorem

The only hypercomplex number systems with commutative multiplication and addition are the algebra with one unit such that $e = e^2$ and the GAUSSIAN INTEGERS.

see also GAUSSIAN INTEGER, PEIRCE'S THEOREM

Weierstraß Zeta Function

The QUASIPERIODIC FUNCTION defined by

$$\frac{d\zeta(z)}{dz} \equiv -\wp(z) \tag{1}$$

with

$$\lim_{z\to 0}[\zeta(z) - z^{-1}] = 0. \tag{2}$$

Then

$$\begin{aligned}\zeta(z) - z^{-1} &= -\int_0^z [\wp(z) - z^{-2}]\,dz\\ &= -\Sigma'\int_0^z [(z-\Omega_{mn})^{-2} - \Omega_{mn}^{-2}]\,dz \end{aligned} \tag{3}$$

$$\zeta(z) = z^{-1} + \sum_{m,n=-\infty}^{\infty}{}' [(z-\Omega_{mn})^{-1} + \Omega_{mn}^{-1} + z\Omega_{mn}^{-2}] \tag{4}$$

so $\zeta(z)$ is an ODD FUNCTION. Integrating $\wp(z+2\omega_1) = \wp(z)$ gives

$$\zeta(z+2\omega_1) = \zeta(z) + 2\eta_1. \tag{5}$$

Letting $z = -\omega_1$ gives $\zeta(-\omega_1) + 2\eta_1 = -\zeta(\omega_1) + 2\eta_1$, so $\eta_1 = \zeta(\omega_1)$. Similarly, $\eta_2 = \zeta(\omega_2)$. From Whittaker and Watson (1990),

$$\eta_1\omega_2 - \eta_2\omega_1 = \tfrac{1}{2}\pi i. \tag{6}$$

If $x + y + z = 0$, then

$$[\zeta(x) + \zeta(y) + \zeta(z)]^2 + \zeta'(x) + \zeta'(y)\zeta'(z) = 0. \tag{7}$$

Also,

$$2\frac{\begin{vmatrix} 1 & \wp(x) & \wp^2(x)\\ 1 & \wp(y) & \wp^2(y)\\ 1 & \wp(z) & \wp^2(z)\end{vmatrix}}{\begin{vmatrix} 1 & \wp(x) & \wp'(x)\\ 1 & \wp(y) & \wp'(y)\\ 1 & \wp(z) & \wp'(z)\end{vmatrix}} = \zeta(x+y+z) - \zeta(x) - \zeta(y) - \zeta(z) \tag{8}$$

(Whittaker and Watson 1990, p. 446).

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Weierstrass Elliptic and Related Functions." Ch. 18 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 627–671, 1972.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Weighings

n weighings are SUFFICIENT to find a bad COIN among $(3^n - 1)/2$ COINS. vos Savant (1993) gives an algorithm for finding a bad ball among 12 balls in three weighings (which, in addition, determines if the bad ball is heavier or lighter than the other 11).

Bachet's weights problem asks for the *minimum number* of weights (which can be placed in *either* pan of a two-arm balance) required to weigh any integral number of pounds from 1 to 40. The solution is 1, 3, 9, and 27: 1, $2 = -1 + 3$, 3, $4 = 1 + 3$, $5 = -1 - 3 + 9$, $6 = -3 + 9$, $7 = 1 - 3 + 9$, $8 = -1 + 9$, 9, $10 = 1 + 9$, $11 = -1 + 3 + 9$, $12 = 3 + 9$, $13 = 1 + 3 + 9$, $14 = -1 - 3 - 9 + 27$, $15 = -3 - 9 + 27$, $16 = 1 - 3 - 9 + 27$, $17 = -1 - 9 + 27$, and so on.

see also GOLOMB RULER, PERFECT DIFFERENCE SET, THREE JUG PROBLEM

References

Bachet, C. G. Problem 5, Appendix in *Problèmes plaisans et délectables, 2nd ed.* p. 215, 1624.

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 50–52, 1987.

Kraitchik, M. *Mathematical Recreations.* New York: W. W. Norton, pp. 52–55, 1942.

Pappas, T. "Counterfeit Coin Puzzle." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 181, 1989.

Tartaglia. Book 1, Ch. 16, §32 in *Trattato de' numeri e misure, Vol. 2.* Venice, 1556.

vos Savant, M. *The World's Most Famous Math Problem.* New York: St. Martin's Press, pp. 39–42, 1993.

Weight

The word weight has many uses in mathematics. It can refer to a function $w(x)$ (also called a WEIGHTING FUNCTION or WEIGHT FUNCTION) used to normalize ORTHONORMAL FUNCTIONS. It can also be used to indicate one of a set of a multiplicative constants placed in front of terms in a MOVING AVERAGE, NEWTON-COTES FORMULAS, edge or vertex of a GRAPH or TREE, etc.

see also WEIGHTED TREE, WEIGHTING FUNCTION

Weight Function

see WEIGHTING FUNCTION

Weighted Tree

A TREE in which each branch is given a numerical WEIGHT (i.e., a labelled TREE).

see also LABELLED GRAPH, TAYLOR'S CONDITION, TREE

Weighting Function

A function $w(x)$ used to normalize ORTHONORMAL FUNCTIONS

$$\int [f_n(x)]^2\, w(x)\, dx = N_n.$$

see also WEIGHT

Weingarten Equations

The Weingarten equations express the derivatives of the NORMAL using derivatives of the position vector. Let $\mathbf{x} : U \to \mathbb{R}^3$ a REGULAR PATCH, then the SHAPE OPERATOR S of $\mathbf{x}$ is given in terms of the basis $\{\mathbf{x}_u, \mathbf{x}_v\}$ by

$$-S(\mathbf{x}_u) = \mathbf{N}_u = \frac{fF - eG}{EG - F^2}\mathbf{x}_u + \frac{eF - fE}{EG - F^2}\mathbf{x}_v \qquad (1)$$

$$-S(\mathbf{x}_v) = \mathbf{N}_v = \frac{gF - fG}{EG - F^2}\mathbf{x}_u + \frac{fF - gE}{EG - F^2}\mathbf{x}_v, \qquad (2)$$

where $\mathbf{N}$ is the NORMAL VECTOR, E, F, and G the coefficients of the first FUNDAMENTAL FORM

$$ds^2 = E\, du^2 + 2F\, du\, dv + G\, dv^2, \qquad (3)$$

and e, f, and g the coefficients of the second FUNDAMENTAL FORM given by

$$e = -\mathbf{N}_u \cdot \mathbf{x}_u = \mathbf{N} \cdot \mathbf{x}_{uu} \qquad (4)$$

$$\begin{aligned} f &= -\mathbf{N}_v \cdot \mathbf{x}_u = \mathbf{N} \cdot \mathbf{x}_{uv} \\ &= \mathbf{N}_{vu} \cdot \mathbf{x}_{vu} = -\mathbf{N}_u \cdot \mathbf{x}_v \qquad (5) \end{aligned}$$

$$g = -\mathbf{N}_v \cdot \mathbf{x}_v = \mathbf{N} \cdot \mathbf{x}_{vv}. \qquad (6)$$

see also FUNDAMENTAL FORMS, SHAPE OPERATOR

References

Gray, A. "Calculation of the Shape Operator." §14.3 in *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 274–277, 1993.

Weingarten Map

see SHAPE OPERATOR

Weird Number

A number which is ABUNDANT without being SEMIPERFECT. (A SEMIPERFECT NUMBER is the sum of any set of its own DIVISORS.) The first few weird numbers are 70, 836, 4030, 5830, 7192, 7912, 9272, 10430, ... (Sloane's A006037). No ODD weird numbers are known, but an infinite number of weird numbers are known to exist. The SEQUENCE of weird numbers has POSITIVE SCHNIRELMANN DENSITY.

see also ABUNDANT NUMBER, SCHNIRELMANN DENSITY, SEMIPERFECT NUMBER

References

Benkoski, S. "Are All Weird Numbers Even?" *Amer. Math. Monthly* **79**, 774, 1972.

Benkoski, S. J. and Erdős, P. "On Weird and Pseudoperfect Numbers." *Math. Comput.* **28**, 617–623, 1974.

Guy, R. K. "Almost Perfect, Quasi-Perfect, Pseudoperfect, Harmonic, Weird, Multiperfect and Hyperperfect Numbers." §B2 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 45–53, 1994.

Sloane, N. J. A. Sequence A006037/M5339 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Welch Apodization Function

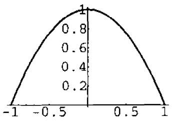

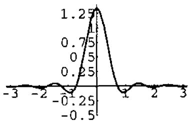

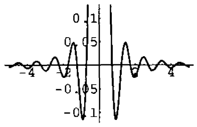

The APODIZATION FUNCTION

$$A(x) = 1 - \frac{x^2}{a^2}.$$

Its FULL WIDTH AT HALF MAXIMUM is $\sqrt{2}\,a$. Its INSTRUMENT FUNCTION is

$$\begin{aligned} I(k) &= a2\sqrt{2\pi}\,\frac{J_{3/2}(2\pi ka)}{(2\pi ka)^{3/2}} \\ &= a\frac{\sin(2\pi ka) - 2\pi ak\cos(2\pi ak)}{2a^3k^3\pi^3}, \end{aligned}$$

where $J_\nu(z)$ is a BESSEL FUNCTION OF THE FIRST KIND. It has a width of 1.59044, a maximum of $\frac{4}{3}$, maximum NEGATIVE sidelobe of -0.0861713 times the peak, and maximum POSITIVE sidelobe of 0.356044 times the peak.

see also APODIZATION FUNCTION, INSTRUMENT FUNCTION

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, p. 547, 1992.

Well-Defined

An expression is called well-defined (or UNAMBIGUOUS) if its definition assigns it a unique interpretation or value. Otherwise, the expression is said to not be well defined or to be AMBIGUOUS.

For example, the expression *abc* (the PRODUCT) is well-defined if a, b, and c are integers. Because integers are ASSOCIATIVE, *abc* has the same value whether it is interpreted to mean $(ab)c$ or $a(bc)$. However, if a, b, and c are MATRICES or CAYLEY NUMBERS, then the expression *abc* is *not* well-defined, since MATRICES and CAYLEY NUMBER are not, in general, ASSOCIATIVE, so that the two interpretations $(ab)c$ and $a(bc)$ can be different.

Sometimes, ambiguities are implicitly resolved by notational convention. For example, the conventional interpretation of $a \wedge b \wedge c = a^{b^c}$ is $a^{(b^c)}$, never $(a^b)^c$, so that the expression $a \wedge b \wedge c$ is well-defined even though exponentiation is nonassociative.

Well-Ordered Set

A SET having the property that every nonempty SUBSET has a least member.

see also AXIOM OF CHOICE, HILBERT'S PROBLEMS, SUBSET, WELL-ORDERING PRINCIPLE

Well-Ordering Principle

Every nonempty set of POSITIVE integers contains a smallest member.

see also WELL-ORDERED SET

References

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, p. 149, 1993.

Werner Formulas

$$2\sin\alpha\cos\beta = \sin(\alpha-\beta) + \sin(\alpha+\beta) \qquad (1)$$

$$2\cos\alpha\cos\beta = \cos(\alpha-\beta) + \cos(\alpha+\beta) \qquad (2)$$

$$2\cos\alpha\sin\beta = \sin(\alpha+\beta) - \sin(\alpha-\beta) \qquad (3)$$

$$2\sin\alpha\sin\beta = \cos(\alpha-\beta) - \cos(\alpha+\beta). \qquad (4)$$

see also TRIGONOMETRY

Weyl's Criterion

A SEQUENCE $\{x_1, x_2, \ldots\}$ is EQUIDISTRIBUTED IFF

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n<N} e^{2\pi i m x_n} = 0$$

for each $m = 1, 2, \ldots$.

see also EQUIDISTRIBUTED SEQUENCE, RAMANUJAN'S SUM

References

Pólya, G. and Szegő, G. *Problems and Theorems in Analysis I.* New York: Springer-Verlag, 1972.

Vardi, I. *Computational Recreations in Mathematica.* Redwood City, CA: Addison-Wesley, pp. 155–156 and 254, 1991.

Weyl Tensor

The TENSOR

$$C^{ij}{}_{kl} = R^i j_{kl} - 2\delta^{[i}{}_{[k} R^{j]}{}_{l]} + \tfrac{1}{3}\delta^{[i}{}_{[k}\delta^{j]}{}_{l]} R,$$

where $R^i j_{kl}$ is the RIEMANN TENSOR and R is the CURVATURE SCALAR. The Weyl tensor is defined so that every CONTRACTION between indices gives 0. In particular, $C^\lambda{}_{\mu\lambda\kappa} = 0$. The number of independent components for a Weyl tensor in N-D is given by

$$C_N = \tfrac{1}{12} N(N+1)(N+2)(N-3).$$

see also CURVATURE SCALAR, RIEMANN TENSOR

References

Weinberg, S. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity.* New York: Wiley, p. 146, 1972.

Weyrich's Formula

Under appropriate constraints,

$$\frac{1}{2} i \int_{-\infty}^{\infty} H_0^{(1)}(r\sqrt{k^2-r^2}\,) e^{i\tau x}\, d\tau = \frac{e^{ik\sqrt{r^2+k^2}}}{\sqrt{r^2+x^2}},$$

where $H_0^{(1)}(z)$ is a HANKEL FUNCTION OF THE FIRST KIND.

References

Iyanaga, S. and Kawada, Y. (Eds.). *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 1474, 1980.

Wheat and Chessboard Problem

Let one grain of wheat be placed on the first square of a CHESSBOARD, two on the second, three on the third, etc. How many grains total are placed on an 8 × 8 CHESSBOARD? Since this is a GEOMETRIC SERIES, the answer for n squares is

$$\sum_{i=0}^{n-1} 2^i = 2^n - 1.$$

Plugging in $n = 8 \times 8 = 64$ then gives $2^{64} - 1 = 18446744073709551615$.

References

Pappas, T. "The Wheat and & Chessboard." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 17, 1989.

Wheel

see ARISTOTLE'S WHEEL PARADOX, BENHAM'S WHEEL, WHEEL GRAPH

Wheel Graph

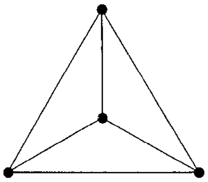

W_4

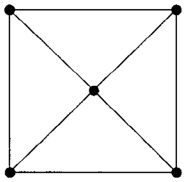

W_5

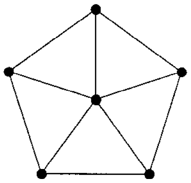

W_6

A Graph W_n of order n which contains a Cycle of order $n-1$, and for which every Node in the cycle is connected to one other Node (known as the Hub). In a wheel graph, the Hub has Degree $n-1$, and other nodes have degree 3. $W_4 = K_4$, where K_4 is the Complete Graph of order four.

see also Complete Graph, Gear Graph, Hub, Web Graph

Wheel Paradox

see Aristotle's Wheel Paradox

Whewell Equation

An Intrinsic Equation which expresses a curve in terms of its Arc Length s and Tangential Angle ϕ.

see also Arc Length, Cesàro Equation, Intrinsic Equation, Natural Equation, Tangential Angle

References

Yates, R. C. "Intrinsic Equations." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 123–126, 1952.

Whipple's Transformation

$$
{}_7F_6\left[\begin{matrix} a, 1+\tfrac{1}{2}a, b, c, d, e, -m \\ \tfrac{1}{2}a, 1+a-b, 1+a-c, \\ 1+a-d, 1+a-e, 1+a+m \end{matrix}\right]
= \frac{(1+a)_m(1+a-d-e)_m}{(1+a-d)_m(1+a-e)_m}
\times {}_4F_3\left[\begin{matrix} 1+a-b-c, d, e, -m \\ 1+a-b, 1+a-c, d+e-a-m \end{matrix}\right],
$$

where ${}_7F_6$ and ${}_4F_3$ are Generalized Hypergeometric Functions and $\Gamma(z)$ is the Gamma Function.

see also Generalized Hypergeometric Function

Whirl

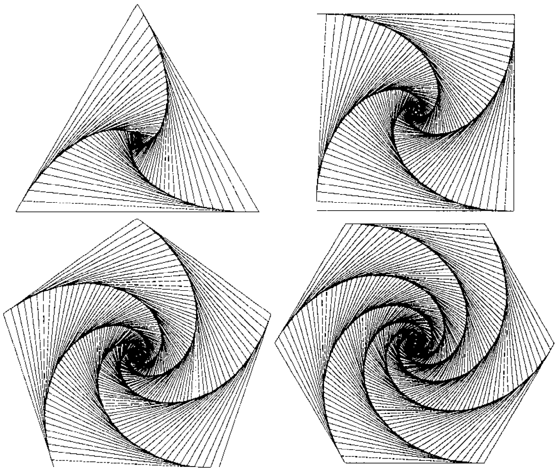

Whirls are figures constructed by nesting a sequence of polygons (each having the same number of sides), each slightly smaller and rotated relative to the previous one.

see also Daisy, Swirl

References

Lauwerier, H. *Fractals: Endlessly Repeated Geometric Figures.* Princeton, NJ: Princeton University Press, p. 66, 1991.

Pappas, T. "Spider & Spirals." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 228, 1989.

Weisstein, E. W. "Fractals." `http://www.astro.virginia.edu/~eww6n/math/notebooks/Fractal.m`.

Whisker Plot

see Box-and-Whisker Plot

Whitehead Double

The Satellite Knot of an Unknot twisted inside a Torus.

see also Satellite Knot, Torus, Unknot

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, pp. 115–116, 1994.

Whitehead Link

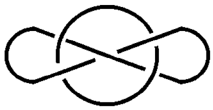

The Link 5_{01}^{02}, illustrated above, with Braid Word ${\sigma_1}^2{\sigma_2}^2{\sigma_1}^{-1}{\sigma_2}^{-2}$ and Jones Polynomial

$$V(t) = t^{-3/2}(-1 + t - 2t^2 + t^3 - 2t^4 + t^5).$$

The Whitehead link has Linking Number 0.

Whitehead Manifold

An open 3-MANIFOLD which is simply connected but is topologically distinct from Euclidean 3-space.

References
Rolfsen, D. *Knots and Links.* Wilmington, DE: Publish or Perish Press, p. 82, 1976.

Whitehead's Theorem

MAPS between CW-COMPLEXES that induce ISOMORPHISMS on all HOMOTOPY GROUPS are actually HOMOTOPY equivalences.

see also CW-COMPLEX, HOMOTOPY GROUP, ISOMORPHISM

Whitney-Graustein Theorem

A 1937 theorem which classified planar regular closed curves up to regular HOMOTOPY by their WINDING NUMBERS. In his thesis, S. Smale generalized this result to regular closed curves on an n-MANIFOLD.

Whitney-Mikhlin Extension Constants

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let $B_n(r)$ be the n-D closed BALL of RADIUS $r > 1$ centered at the ORIGIN. A function which is defined on $B(r)$ is called an extension to $B(r)$ of a function f defined on $B(1)$ if

$$F(x) = f(x) \forall\ x \in B(1). \tag{1}$$

Given 2 BANACH SPACES of functions defined on $B(1)$ and $B(r)$, find the extension operator from one to the other of minimal norm. Mikhlin (1986) found the best constants χ such that this condition, corresponding to the Sobolev $W(1,2)$ integral norm, is satisfied,

$$\sqrt{\int_{B(1)} \left[[f(x)]^2 + \sum_{j=1}^n \left(\frac{\partial f}{\partial x_j}\right)^2 \right] dx} \leq \chi \sqrt{\int_{B(r)} \left[[F(x)]^2 + \sum_{j=1}^n \left(\frac{\partial F}{\partial x_j}\right)^2 \right] dx}. \tag{2}$$

$\chi(1,r) = 1$. Let

$$\nu = \tfrac{1}{2}(n-2), \tag{3}$$

then for $n > 2$,

$$\chi(n,r) = \sqrt{1 + \frac{I_\nu(1)}{I_{\nu+1}(1)} \frac{I_\nu(r)K_{\nu+1}(1) + K_\nu(r)I_{\nu+1}(1)}{I_\nu(r)K_\nu(1) - K_\nu(r)I_\nu(1)}}, \tag{4}$$

where $I_\nu(z)$ is a MODIFIED BESSEL FUNCTION OF THE FIRST KIND and $K_\nu(z)$ is a MODIFIED BESSEL FUNCTION OF THE SECOND KIND. For $n = 2$,

$$\chi(2,r) = \max \left\{ \sqrt{1 + \frac{I_\nu(1)}{I_{\nu+1}(1)} \frac{I_\nu(r)K_{\nu+1}(1) + K_\nu(r)I_{\nu+1}(1)}{I_\nu(r)K_\nu(1) - K_\nu(r)I_\nu(1)}} \sqrt{1 + \frac{I_1(1)}{I_1(1) + I_2(1)} \left[1 + \frac{I_1(r)K_0(1) + K_1(r)I_0(1)}{I_1(r)K_1(1) - K_1(r)I_1(1)} \right]} \right\}. \tag{5}$$

For $r \to \infty$,

$$\chi(n,\infty) = \sqrt{1 + \frac{I_\nu(1)}{I_{\nu+1}(1)} \frac{K_\nu(1)}{K_\nu(1)}}, \tag{6}$$

which is bounded by

$$n - 1 < \chi(n,\infty) < \sqrt{(n-1)^2 + 4}. \tag{7}$$

For ODD n, the RECURRENCE RELATIONS

$$a_{k+1} = a_{k-1} - (2k-1)a_k \tag{8}$$

$$b_{k+1} = b_{k-1} + (2k-1)b_k \tag{9}$$

with

$$a_0 = e + e^{-1} \tag{10}$$

$$a_1 = e - e^{-1} \tag{11}$$

$$b_0 = e^{-1} \tag{12}$$

$$b_1 = e^{-1} \tag{13}$$

where e is the constant 2.71828..., give

$$\chi(2k+1,\infty) = \sqrt{1 + \frac{a_k}{a_{k+1}} \frac{b_{k+1}}{b_k}}. \tag{14}$$

The first few are

$$\chi(3,\infty) = e \tag{15}$$

$$\chi(5,\infty) = \sqrt{\frac{e^2}{e^2 - 7}} \tag{16}$$

$$\chi(7,\infty) = \sqrt{\frac{2}{7}} \sqrt{\frac{e^2}{37 - 5e^2}} \tag{17}$$

$$\chi(9,\infty) = \frac{1}{\sqrt{37}} \sqrt{\frac{e^2}{18e^2 - 133}} \tag{18}$$

$$\chi(11,\infty) = \frac{1}{\sqrt{133}} \sqrt{\frac{e^2}{2431 - 329e^2}} \tag{19}$$

$$\chi(13,\infty) = \sqrt{\frac{2}{2431}} \sqrt{\frac{e^2}{3655e^2 - 27007}}. \tag{20}$$

Similar formulas can be given for even n in terms of $I_0(1)$, $I_1(1)$, $K_0(1)$, $K_1(1)$.

References
Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/mkhln/mkhln.html`.
Mikhlin, S. G. *Constants in Some Inequalities of Analysis.* New York: Wiley, 1986.

Whitney Singularity

see PINCH POINT

Whitney Sum

An operation that takes two VECTOR BUNDLES over a fixed SPACE and produces a new VECTOR BUNDLE over the same SPACE. If E_1 and E_2 are VECTOR BUNDLES over B, then the Whitney sum $E_1 \oplus E_2$ is the VECTOR BUNDLE over B such that each FIBER over B is naturally the direct sum of the E_1 and E_2 FIBERS over B.

The Whitney sum is therefore the FIBER for FIBER direct sum of the two BUNDLES E_1 and E_2. An easy formal definition of the Whitney sum is that $E_1 \oplus E_2$ is the pull-back BUNDLE of the diagonal map from B to $B \times B$, where the BUNDLE over $B \times B$ is $E_1 \times E_2$.

see also BUNDLE, FIBER, VECTOR BUNDLE

Whitney Umbrella

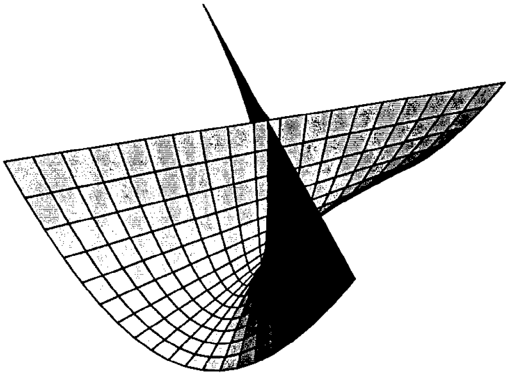

A surface which can be interpreted as a self-intersecting RECTANGLE in 3-D. It is given by the parametric equations

$$x = uv \qquad (1)$$

$$y = u \qquad (2)$$

$$z = v^2 \qquad (3)$$

for $u, v \in [-1, 1]$. The center of the "plus" shape which is the end of the line of self-intersection is a PINCH POINT. The coefficients of the first FUNDAMENTAL FORM are

$$E = 0 \qquad (4)$$

$$F = \frac{2v}{\sqrt{u^2 + 4v^2 + 4v^4}} \qquad (5)$$

$$G = -\frac{2u}{\sqrt{u^2 + 4v^2 + 4v^4}}, \qquad (6)$$

and the coefficients of the second FUNDAMENTAL FORM are

$$e = 1 + v^2 \qquad (7)$$

$$f = uv \qquad (8)$$

$$g = u^2 + 4v^2, \qquad (9)$$

giving GAUSSIAN CURVATURE and MEAN CURVATURE

$$K = -\frac{4v^2}{(u^2 + 4v^2 + 4v^4)^2} \qquad (10)$$

$$H = -\frac{u(1 + 3v^2)}{(u^2 + 4v^2 + 4v^4)^{3/2}}. \qquad (11)$$

References

Francis, G. K. *A Topological Picturebook.* New York: Springer-Verlag, pp. 8–9, 1987.

Geometry Center. "Whitney's Umbrella." `http://www.geom.umn.edu/zoo/features/whitney/.`

Gray, A. *Modern Differential Geometry of Curves and Surfaces.* Boca Raton, FL: CRC Press, pp. 225 and 309–310, 1993.

Whittaker Differential Equation

$$\frac{d^2u}{dz^2} + \frac{du}{dz} + \left(\frac{k}{z} + \frac{\frac{1}{4} - m^2}{z^2}\right) u = 0. \qquad (1)$$

Let $u \equiv e^{-z/2} W_{k,m}(z)$, where $W_{k,m}(z)$ denotes a WHITTAKER FUNCTION. Then (1) becomes

$$\frac{d}{dz}(-\tfrac{1}{2}e^{-z/2}W + e^{-z/2}W') + (-\tfrac{1}{2}e^{-z/2}W + e^{-z/2}W') + \left(\frac{k}{z} + \frac{\frac{1}{4} - m^2}{z^2}\right) e^{-z/2}W = 0. \qquad (2)$$

Rearranging,

$$(\tfrac{1}{4}e^{-z/2}W - \tfrac{1}{2}e^{-z/2}W' - \tfrac{1}{2}e^{-z/2}W' + e^{-z/2}W'')p + (-\tfrac{1}{2}e^{-z/2}W + e^{-z/2}W') + \left(\frac{k}{z} + \frac{\frac{1}{4} - m^2}{z^2}\right) e^{-z/2}W = 0 \qquad (3)$$

$$-\tfrac{1}{4}e^{-z/2}W + e^{-z/2}W'' + \left(\frac{k}{z} + \frac{\frac{1}{4} - m^2}{z^2}\right) e^{-z/2}W = 0, \qquad (4)$$

so

$$W'' + \left(-\frac{1}{4} + \frac{k}{z} + \frac{\frac{1}{4} - m^2}{z^2}\right) W = 0, \qquad (5)$$

where $W' \equiv dW/dz$. The solutions are known as WHITTAKER FUNCTIONS.

References

Abramowitz, M. and Stegun, C. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, p. 505, 1972.

Whittaker Function

Solutions to the Whittaker Differential Equation. The linearly independent solutions are

$$M_{k,m}(z) \equiv z^{1/2+m}e^{-z/2} \times \left(1+\frac{\frac{1}{2}+m-k}{1!(2m+1)}+\frac{(\frac{1}{2}+m-k)(\frac{3}{2}+m-k)}{2!(2m+1)(2m+2)}z^2+\ldots\right), \quad (1)$$

and $M_{k,-m}(z)$, where $M_{k,m}(z)$ is a Confluent Hypergeometric Function. In terms of Confluent Hypergeometric Functions, the Whittaker functions are

$$M_{k,m}(z) = e^{-z/2}z^{m+1/2}{}_1F_1(\tfrac{1}{2}+m-k, 1+2m; z) \quad (2)$$

$$W_{k,m}(z) = e^{-z/2}z^{m+1/2}U(\tfrac{1}{2}+m-k, 1+2m; z) \quad (3)$$

(see Whittaker and Watson 1990, pp. 339–351). However, the Confluent Hypergeometric Function disappears when $2m$ is an Integer, so Whittaker functions are often defined instead. The Whittaker functions are related to the Parabolic Cylinder Functions. When $|\arg z| < 3\pi/2$ and $2m$ is not an Integer,

$$W_{k,m}(z) = \frac{\Gamma(-2m)}{\Gamma(\frac{1}{2}-m-k)}M_{k,m}(z) + \frac{\Gamma(2m)}{\Gamma(\frac{1}{2}+m-k)}M_{k,-m}(z). \quad (4)$$

When $|\arg(-z)| < 3\pi/2$ and $2m$ is not an Integer,

$$W_{-k,m}(-z) = \frac{\Gamma(-2m)}{\Gamma(\frac{1}{2}-m-k)}M_{-k,m}(-z) + \frac{\Gamma(2m)}{\Gamma(\frac{1}{2}+m+k)}M_{-k,-m}(-z). \quad (5)$$

Whittaker functions satisfy the Recurrence Relations

$$W_{k,m}(z) = z^{1/2}W_{k-1/2,m-1/2}(z)+(\tfrac{1}{2}-k+m)W_{k-1,m}(z) \quad (6)$$

$$W_{k,m}(z) = z^{1/2}W_{k-1/2,m+1/2}(z)+(\tfrac{1}{2}-k-m)W_{k-1,m}(z) \quad (7)$$

$$zW'_{k,m}(z) = (k-\tfrac{1}{2}z)W_{k,m}(z)-[m^2-(k-\tfrac{1}{2})^2]W_{k-1,m}(z). \quad (8)$$

see also Confluent Hypergeometric Function, Kummer's Formulas, Pearson-Cunningham Function, Schlömilch's Function, Sonine Polynomial

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Confluent Hypergeometric Functions." Ch. 13 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 503–515, 1972.

Iyanaga, S. and Kawada, Y. (Eds.). "Whittaker Functions." Appendix A, Table 19.II in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, pp. 1469–1471, 1980.

Whittaker, E. T. and Watson, G. N. *A Course in Modern Analysis, 4th ed.* Cambridge, England: Cambridge University Press, 1990.

Whole Number

One of the numbers 1, 2, 3, ... (Sloane's A000027), also called the Counting Numbers or Natural Numbers. 0 is sometimes included in the list of "whole" numbers (Bourbaki 1968, Halmos 1974), but there seems to be no general agreement. Some authors also interpret "whole number" to mean "a number having Fractional Part of zero," making the whole numbers equivalent to the integers.

Due to lack of standard terminology, the following terms are recommended in preference to "Counting Number," "Natural Number," and "whole number."

Set	Name	Symbol
..., −2, −1, 0, 1, 2, ...	integers	$\mathbb{Z}$
1, 2, 3, 4, ...	positive integers	$\mathbb{Z}^+$
0, 1, 2, 3, 4 ...	nonnegative integers	$\mathbb{Z}^*$
−1, −2, −3, −4, ...	negative integers	$\mathbb{Z}^-$

see also Counting Number, Fractional Part, Integer, $\mathbb{N}$, Natural Number, $\mathbb{Z}$, $\mathbb{Z}^+$, $\mathbb{Z}^+$, $\mathbb{Z}^*$

References

Bourbaki, N. *Elements of Mathematics: Theory of Sets.* Paris, France: Hermann, 1968.

Halmos, P. R. *Naive Set Theory.* New York: Springer-Verlag, 1974.

Sloane, N. J. A. Sequence A000027/M0472 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Width (Partial Order)

For a Partial Order, the size of the longest Antichain is called the width.

see also Antichain, Length (Partial Order), Partial Order

Width (Size)

The width of a box is the horizontal distance from side to side (usually defined to be greater than the Depth, the horizontal distance from front to back).

see also Depth (Size), Height

References

Eppstein, D. "Width, Diameter, and Geometric Inequalities." `http://www.ics.uci.edu/~eppstein/junkyard/diam.html`.

Wiedersehen Manifold

The only Wiedersehen manifolds are the standard round spheres, as was established by proof of the Blaschke Conjecture.

see also Blaschke Conjecture

Wieferich Prime

A Wieferich prime is a Prime p which is a solution to the Congruence equation

$$2^{p-1} \equiv 1 \pmod{p^2}.$$

Note the similarity of this expression to the special case of FERMAT'S LITTLE THEOREM

$$2^{p-1} \equiv 1 \pmod{p},$$

which holds for *all* ODD PRIMES. However, the only Wieferich primes less than 4×10^{12} are $p = 1093$ and 3511 (Lehmer 1981, Crandall 1986, Crandall *et al.* 1997). Interestingly, one less than these numbers have suggestive periodic BINARY representations

$$\begin{aligned} 1092 &= 10001000100_2 \\ 3510 &= 110110110110_2. \end{aligned}$$

A PRIME factor p of a MERSENNE NUMBER $M_q = 2^q - 1$ is a Wieferich prime IFF $p^2 | 2^q - 1$. Therefore, MERSENNE PRIMES are *not* Wieferich primes.

If the first case of FERMAT'S LAST THEOREM is false for exponent p, then p must be a Wieferich prime (Wieferich 1909). If $p | 2^n \pm 1$ with p and n RELATIVELY PRIME, then p is a Wieferich prime IFF p^2 also divides $2^n \pm 1$. The CONJECTURE that there are no three POWERFUL NUMBERS implies that there are infinitely many Wieferich primes (Granville 1986, Vardi 1991). In addition, the ABC CONJECTURE implies that there are at least $C \ln x$ Wieferich primes $\leq x$ for some constant C (Silverman 1988, Vardi 1991).

see also ABC CONJECTURE, FERMAT'S LAST THEOREM, FERMAT QUOTIENT, MERSENNE NUMBER, MIRIMANOFF'S CONGRUENCE, POWERFUL NUMBER

References

Brillhart, J.; Tonascia, J.; and Winberger, P. "On the Fermat Quotient." In *Computers and Number Theory* (Ed. A. O. L. Atkin and B. J. Birch). New York: Academic Press, pp. 213–222, 1971.

Crandall, R. *Projects in Scientific Computation.* New York: Springer-Verlag, 1986.

Crandall, R.; Dilcher, K; and Pomerance, C. "A search for Wieferich and Wilson Primes." *Math. Comput.* **66**, 433–449, 1997.

Granville, A. "Powerful Numbers and Fermat's Last Theorem." *C. R. Math. Rep. Acad. Sci. Canada* **8**, 215–218, 1986.

Lehmer, D. H. "On Fermat's Quotient, Base Two." *Math. Comput.* **36**, 289–290, 1981.

Ribenboim, P. "Wieferich Primes." §5.3 in *The New Book of Prime Number Records.* New York: Springer-Verlag, pp. 333–346, 1996.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 116 and 157, 1993.

Silverman, J. "Wieferich's Criterion and the abc Conjecture." *J. Number Th.* **30**, 226–237, 1988.

Vardi, I. "Wieferich." §5.4 in *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, pp. 59–62 and 96–103, 1991.

Wieferich, A. "Zum letzten Fermat'schen Theorem." *J. reine angew. Math.* **136**, 293–302, 1909.

Wielandt's Theorem

Let the $n \times n$ MATRIX A satisfy the conditions of the PERRON-FROBENIUS THEOREM and the $n \times n$ MATRIX $\mathsf{C} = c_{ij}$ satisfy

$$|c_{ij}| \leq a_{ij}$$

for $i, j = 1, 2, \ldots, n$. Then any EIGENVALUE λ_0 of C satisfies the inequality $|\lambda_0| \leq R$ with the equality sign holding only when there exists an $n \times n$ MATRIX $\mathsf{D} = \delta_{ij}$ (where δ_{ij} is the KRONECKER DELTA) and

$$\mathsf{C} = \frac{\lambda_0}{R} \mathsf{DAD}^{-1}.$$

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1121, 1979.

Wiener Filter

An optimal FILTER used for the removal of noise from a signal which is corrupted by the measuring process itself.

see also FILTER

References

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. "Optimal (Wiener) Filtering with the FFT." §13.3 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 539–542, 1992.

Wiener Function

see BROWN FUNCTION

Wiener-Khintchine Theorem

Recall the definition of the AUTOCORRELATION function $C(t)$ of a function $E(t)$,

$$C(t) \equiv \int_{-\infty}^{\infty} E^*(\tau) E(t+\tau)\, d\tau. \qquad (1)$$

Also recall that the FOURIER TRANSFORM of $E(t)$ is defined by

$$E(\tau) = \int_{-\infty}^{\infty} E_\nu e^{-2\pi i \nu \tau}\, d\nu, \qquad (2)$$

giving a COMPLEX CONJUGATE of

$$E^*(\tau) = \int_{-\infty}^{\infty} E_\nu^* e^{2\pi i \nu \tau}\, d\nu. \qquad (3)$$

Plugging $E^*(\tau)$ and $E(t+\tau)$ into the AUTOCORRELATION function therefore gives

$$\begin{aligned} C(t) &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} E_\nu^* e^{2\pi i\nu\tau}\,d\nu\right] \\ &\quad\times\left[\int_{-\infty}^{\infty} E_{\nu'} e^{-2\pi i\nu'(t+\tau)}\,d\nu'\right] d\tau \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_\nu^* E_{\nu'} e^{-2\pi i\tau(\nu'-\nu)} e^{-2\pi i\nu' t}\,d\tau\,d\nu\,d\nu' \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_\nu^* E_{\nu'}\delta(\nu'-\nu) e^{-2\pi i\nu' t}\,d\nu\,d\nu' \\ &= \int_{-\infty}^{\infty} E_\nu^* E_\nu e^{-2\pi i\nu t}\,d\nu \\ &= \int_{-\infty}^{\infty} |E_\nu|^2 e^{-2\pi i\nu t}\,d\nu \\ &= \mathcal{F}[\,|E_\nu|^2], \end{aligned} \tag{4}$$

so, amazingly, the AUTOCORRELATION is simply given by the FOURIER TRANSFORM of the ABSOLUTE SQUARE of $E(\nu)$,

$$C(t) = \mathcal{F}[|E(\nu)|^2]. \tag{5}$$

The Wiener-Khintchine theorem is a special case of the CROSS-CORRELATION THEOREM with $f = g$.

see also AUTOCORRELATION, CROSS-CORRELATION THEOREM, FOURIER TRANSFORM

Wiener Measure

The distribution which arises whenever a central limit scaling procedure is carried out on path-space valued random variables.

Wiener Space

see MALLIAVIN CALCULUS

Wigner 3j-Symbol

The Wigner $3j$ symbols are written

$$\begin{pmatrix} j_1 & j_2 & j \\ m_1 & m_2 & m \end{pmatrix} \tag{1}$$

and are sometimes expressed using the related CLEBSCH-GORDON COEFFICIENTS

$$C^j_{m_1 m_2} = (j_1 j_2 m_1 m_2 | j_1 j_2 j m) \tag{2}$$

(Condon and Shortley 1951, pp. 74–75; Wigner 1959, p. 206), or RACAH V-COEFFICIENTS

$$V(j_1 j_2 j; m_1 m_2 m). \tag{3}$$

Connections among the three are

$$(j_1 j_2 m_1 m_2 | j_1 j_2 m) = (-1)^{-j_1+j_2-m}\sqrt{2j+1}\begin{pmatrix} j_1 & j_2 & j \\ m_1 & m_2 & -m \end{pmatrix} \tag{4}$$

$$(j_1 j_2 m_1 m_2 | j_1 j_2 j m) = (-1)^{j+m}\sqrt{2j+1}V(j_1 j_2 j; m_1 m_2\ -m) \tag{5}$$

$$V(j_1 j_2 j; m_1 m_2 m) = (-1)^{-j_1+j_2+j}\begin{pmatrix} j_1 & j_2 & j_1 \\ m_2 & m_1 & m_2 \end{pmatrix}. \tag{6}$$

The Wigner $3j$-symbols have the symmetries

$$\begin{aligned} \begin{pmatrix} j_1 & j_2 & j \\ m_1 & m_2 & m \end{pmatrix} &= \begin{pmatrix} j_2 & j & j_1 \\ m_2 & m & m_1 \end{pmatrix} \\ &= \begin{pmatrix} j & j_1 & j_2 \\ m & m_1 & m_2 \end{pmatrix} = (-1)^{j_1+j_2+j}\begin{pmatrix} j_2 & j_1 & j \\ m_2 & m_1 & m \end{pmatrix} \\ &= (-1)^{j_1+j_2+j}\begin{pmatrix} j_1 & j & j_2 \\ m_1 & m & m_2 \end{pmatrix} \\ &= (-1)^{j_1+j_2+j}\begin{pmatrix} j & j_2 & j_1 \\ m & m_2 & m_1 \end{pmatrix} \\ &= (-1)^{j_1+j_2+j}\begin{pmatrix} j_1 & j_2 & j \\ -m_1 & -m_2 & -m \end{pmatrix}. \end{aligned} \tag{7}$$

The symbols obey the orthogonality relations

$$\sum_{j,m}(2j+1)\begin{pmatrix} j_1 & j_2 & j \\ m_1 & m_2 & m \end{pmatrix}\begin{pmatrix} j_1 & j_2 & j \\ m_1' & m_2' & m \end{pmatrix} = \delta_{m_1 m_1'}\delta_{m_2 m_2'} \tag{8}$$

$$\sum_{m_1, m_2}\begin{pmatrix} j_1 & j_2 & j \\ m_1 & m_2 & m \end{pmatrix}\begin{pmatrix} j_1 & j_2 & j' \\ m_1 & m_2 & m' \end{pmatrix} = \delta_{jj'}\delta_{m_1 m_1'}, \tag{9}$$

where δ_{ij} is the KRONECKER DELTA.

General formulas are very complicated, but some specific cases are

$$\begin{aligned} \begin{pmatrix} j_1 & j_2 & j_1+j_2 \\ m_1 & m_2 & -m_1-m_2 \end{pmatrix} &= (-1)^{j_1-j_2+m_1+m_2} \\ &\quad\times\left[\frac{(2j_1)!(2j_2)!}{(2j_1+2j_2+1)!(j_1+m_1)!}\right. \\ &\quad\left.\times\frac{(j_1+j_2+m_1+m_2)!(j_1+j_2-m_1-m_2)!}{(j_1-m_1)!(j_2+m_2)!(j_2-m_2)!}\right]^{1/2} \end{aligned} \tag{10}$$

$$\begin{aligned} \begin{pmatrix} j_1 & j_2 & j \\ j_1 & -j_1- & m \end{pmatrix} &= (-1)^{-j_1+j_2+m} \\ &\quad\times\left[\frac{(2j_1)!(-j_1+j_2+j)!}{(j_1+j_2+j+1)!(j_1-j_2+j)!}\right. \\ &\quad\left.\frac{(j_1+j_2+m)!(j-m)!}{(j_1+j_2-j)!(-j_1+j_2-m)!(j+m)!}\right]^{1/2} \end{aligned} \tag{11}$$

$$\begin{pmatrix} j_1 & j_2 & j \\ 0 & 0 & 0 \end{pmatrix} = \begin{cases} (-1)^g \sqrt{\frac{(2g-2j_1)!(2g-2j_2)!(2g-2j)!}{(2g+1)!}} \frac{g!}{(g-j_1)!(g-j_2)!(g-j)!} \\ \quad \text{if } J = 2g \\ 0 \\ \quad \text{if } J = 2g+1, \end{cases} \tag{12}$$

for $J \equiv j_1 + j_2 + j$.

For SPHERICAL HARMONICS $Y_{lm}(\theta, \phi)$,

$$Y_{l_1 m_1}(\theta,\phi) Y_{l_2 m_2}(\theta,\phi) = \sum_{l,m} \sqrt{\frac{(2l_1+1)(2l_2+1)(2l+1)}{4\pi}} \begin{pmatrix} l_1 & l_2 & l \\ m_1 & m_2 & m \end{pmatrix} \times Y_{lm}^*(\theta,\psi) \begin{pmatrix} l_1 & l_2 & l \\ 0 & 0 & 0 \end{pmatrix}. \tag{13}$$

For values of l_3 obeying the TRIANGLE CONDITION $\Delta(l_1 l_2 l_3)$,

$$\int Y_{l_1 m_1}(\theta,\phi) Y_{l_2 m_2}(\theta,\phi) Y_{l_3 m_3}(\theta,\phi) \sin\theta \, d\theta \, d\phi = \sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}} \times \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \tag{14}$$

and

$$\frac{1}{2} \int P_{l_1}(\cos\theta) P_{l_2}(\cos\theta) P_{l_3}(\cos\theta) \sin\theta \, d\theta = \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}^2. \tag{15}$$

see also CLEBSCH-GORDON COEFFICIENT, RACAH V-COEFFICIENT, RACAH W-COEFFICIENT, WIGNER $6j$-SYMBOL, WIGNER $9j$-SYMBOL

References

Abramowitz, M. and Stegun, C. A. (Eds.). "Vector-Addition Coefficients." §27.9 in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 1006–1010, 1972.

Condon, E. U. and Shortley, G. *The Theory of Atomic Spectra.* Cambridge, England: Cambridge University Press, 1951.

de Shalit, A. and Talmi, I. *Nuclear Shell Theory.* New York: Academic Press, 1963.

Gordy, W. and Cook, R. L. *Microwave Molecular Spectra, 3rd ed.* New York: Wiley, pp. 804–811, 1984.

Messiah, A. "Clebsch-Gordon (C.-G.) Coefficients and '$3j$' Symbols." Appendix C.I in *Quantum Mechanics, Vol. 2.* Amsterdam, Netherlands: North-Holland, pp. 1054–1060, 1962.

Rotenberg, M.; Bivens, R.; Metropolis, N.; and Wooten, J. K. *The 3j and 6j Symbols.* Cambridge, MA: MIT Press, 1959.

Shore, B. W. and Menzel, D. H. *Principles of Atomic Spectra.* New York: Wiley, pp. 275–276, 1968.

Sobel'man, I. I. "Angular Momenta." Ch. 4 in *Atomic Spectra and Radiative Transitions, 2nd ed.* Berlin: Springer-Verlag, 1992.

Wigner, E. P. *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra, expanded and improved ed.* New York: Academic Press, 1959.

Wigner $6j$-Symbol

A generalization of CLEBSCH-GORDON COEFFICIENTS and WIGNER $3j$-SYMBOL which arises in the coupling of three angular momenta. Let tensor operators $T^{(k)}$ and $U^{(k)}$ act, respectively, on subsystems 1 and 2 of a system, with subsystem 1 characterized by angular momentum $\mathbf{j}_1$ and subsystem 2 by the angular momentum $\mathbf{j}_2$. Then the matrix elements of the scalar product of these two tensor operators in the coupled basis $\mathbf{J} = \mathbf{j}_1 + \mathbf{j}_2$ are given by

$$(\tau_1' j_1' \tau_2' j_2' J' M' | T^{(k)} \cdot U^{(k)} | \tau_1 j_1 \tau_2 j_2 J M) = \delta_{JJ'} \delta_{MM'} (-1)^{j_1 + j_2' + J} \begin{Bmatrix} J & j_2' & j_1' \\ k & j_1 & j_2 \end{Bmatrix} \times (\tau_1' j_1' || T^{(k)} || \tau_1 j_1)(\tau_2' j_2' || U^{(k)} || \tau_2 j_2), \tag{1}$$

where $\begin{Bmatrix} J & j_2' & j_1' \\ k & j_1 & j_2 \end{Bmatrix}$ is the Wigner $6j$-symbol and τ_1 and τ_2 represent additional pertinent quantum numbers characterizing subsystems 1 and 2 (Gordy and Cook 1984).

Edmonds (1968) gives analytic forms of the $6j$-symbol for simple cases, and Shore and Menzel (1968) and Gordy and Cook (1984) give

$$\begin{Bmatrix} a & b & c \\ 0 & c & b \end{Bmatrix} = \frac{(-1)^s}{\sqrt{(2b+1)(2c+1)}} \tag{2}$$

$$\begin{Bmatrix} a & b & c \\ 1 & c & b \end{Bmatrix} = \frac{2(-1)^{s+1} X}{\sqrt{2b(2b+1)(2b+2)2c(2c+1)(2c+2)}} \tag{3}$$

$$\begin{Bmatrix} a & b & c \\ 2 & c & b \end{Bmatrix} = \frac{2(-1)^s [3X(X-1) - 4b(b+1)c(c+1)]}{\sqrt{(2b-1)2b(2b+1)(2b+2)(2b+3)}} \times \frac{1}{\sqrt{(2c-1)2c(2c+1)(2c+2)(2c+3)}}, \tag{4}$$

where

$$s \equiv a + b + c \tag{5}$$

$$X \equiv b(b+1) + c(c+1) - a(a+1). \tag{6}$$

see also CLEBSCH-GORDON COEFFICIENT, RACAH V-COEFFICIENT, RACAH W-COEFFICIENT, WIGNER $3j$-SYMBOL, WIGNER $9j$-SYMBOL

References

Carter, J. S.; Flath, D. E.; and Saito, M. *The Classical and Quantum 6j-Symbols.* Princeton, NJ: Princeton University Press, 1995.

Edmonds, A. R. *Angular Momentum in Quantum Mechanics, 2nd ed., rev. printing.* Princeton, NJ: Princeton University Press, 1968.
Gordy, W. and Cook, R. L. *Microwave Molecular Spectra, 3rd ed.* New York: Wiley, pp. 807–809, 1984.
Messiah, A. "Racah Coefficients and '6j' Symbols." Appendix C.II in *Quantum Mechanics, Vol. 2.* Amsterdam, Netherlands: North-Holland, pp. 567–569 and 1061–1066, 1962.
Rotenberg, M.; Bivens, R.; Metropolis, N.; and Wooten, J. K. *The* 3j *and* 6j *Symbols.* Cambridge, MA: MIT Press, 1959.
Shore, B. W. and Menzel, D. H. *Principles of Atomic Spectra.* New York: Wiley, pp. 279–284, 1968.

Wigner $9j$-Symbol

A generalization of CLEBSCH-GORDON COEFFICIENTS and WIGNER $3j$- and $6j$-SYMBOLS which arises in the coupling of four angular momenta and can be written in terms of the WIGNER $3j$- and $6j$-SYMBOLS. Let tensor operators $T^{(k_1)}$ and $U^{(k_2)}$ act, respectively, on subsystems 1 and 2. Then the reduced matrix element of the product $T^{(k_1)} \times U^{(k_2)}$ of these two irreducible operators in the coupled representation is given in terms of the reduced matrix elements of the individual operators in the uncoupled representation by

$$\begin{aligned}(\tau'\tau_1'j_1'\tau_2'j_2'J'||[T^{(k_1)} \times U^{(k_2)}]^{(k)}||\tau\tau_1j_1\tau_2j_2J) \\ = \sqrt{(2J+1)(2J'+1)(2k+1)}\sum_{\tau''}\begin{Bmatrix} j_1' & j_1 & k_1 \\ j_2' & j_2 & k_2 \\ J' & J & k \end{Bmatrix} \\ (\tau'\tau_1'j_1'||T^{(k_1)}||\tau''\tau_1j_1)(\tau''\tau_2'j_2'||U^{(k_2)}||\tau\tau_2j_2), \quad (1)\end{aligned}$$

where $\begin{Bmatrix} j_1' & j_1 & k_1 \\ j_2' & j_2 & k_2 \\ J' & J & k \end{Bmatrix}$ is a Wigner $9j$-symbol (Gordy and Cook 1984).

Shore and Menzel (1968) give the explicit formulas

$$\begin{Bmatrix} a & b & C \\ d & e & F \\ G & H & J \end{Bmatrix} = \sum_x (-1)^{2x}(2x+1) \times \begin{Bmatrix} a & b & C \\ F & J & x \end{Bmatrix}\begin{Bmatrix} d & e & F \\ b & x & H \end{Bmatrix}\begin{Bmatrix} G & H & J \\ x & a & d \end{Bmatrix} \quad (2)$$

$$\begin{Bmatrix} a & b & J \\ c & d & J \\ K & K & 0 \end{Bmatrix} = \frac{(-1)^{b+c+J+K}}{\sqrt{(2J+1)(2K+1)}}\begin{Bmatrix} a & b & J \\ d & c & K \end{Bmatrix} \quad (3)$$

$$\begin{Bmatrix} S & S & 1 \\ L & L & 2 \\ J & J & 1 \end{Bmatrix} = \frac{\begin{Bmatrix} S & L & J \\ L & S & 1 \end{Bmatrix}\begin{Bmatrix} J & L & S \\ L & J & 1 \end{Bmatrix}}{5\begin{Bmatrix} 2 & L & L \\ L & 1 & 1 \end{Bmatrix}} + \frac{(-1)^{S+L+J+1}}{15(2L+1)}\frac{\begin{Bmatrix} S & J & L \\ J & S & 1 \end{Bmatrix}}{\begin{Bmatrix} 2 & L & L \\ L & 1 & 1 \end{Bmatrix}}. \quad (4)$$

see also CLEBSCH-GORDON COEFFICIENT, RACAH V-COEFFICIENT, RACAH W-COEFFICIENT, WIGNER $3j$-SYMBOL, WIGNER $6j$-SYMBOL

References
Gordy, W. and Cook, R. L. *Microwave Molecular Spectra, 3rd ed.* New York: Wiley, pp. 807–809, 1984.
Messiah, A. "'9j' Symbols." Appendix C.III in *Quantum Mechanics, Vol. 2.* Amsterdam, Netherlands: North-Holland, pp. 567–569 and 1066–1068, 1962.
Shore, B. W. and Menzel, D. H. *Principles of Atomic Spectra.* New York: Wiley, pp. 279–284, 1968.

Wigner-Eckart Theorem

A theorem of fundamental importance in spectroscopy and angular momentum theory which provides both (1) an explicit form for the dependence of all matrix elements of irreducible tensors on the projection quantum numbers and (2) a formal expression of the conservation laws of angular momentum (Rose 1995).

The theorem states that the dependence of the matrix element $(j'm'|T_{LM}|jm)$ on the projection quantum numbers is entirely contained in the WIGNER $3j$-SYMBOL (or, equivalently, the CLEBSCH-GORDON COEFFICIENT), given by

$$(j'm'|T_{LM}|jm) = C(jLj'; mMm')(j'||T_L||j),$$

where $C(jLj'; mMm')$ is a CLEBSCH-GORDON COEFFICIENT and T_{LM} is a set of tensor operators (Rose 1995, p. 85).

see also CLEBSCH-GORDON COEFFICIENT, WIGNER $3j$-SYMBOL

References
Cohen-Tannoudji, C.; Diu, B.; and Laloë, F. "Vector Operators: The WIgner-Eckart Theorem." Complement D_X in *Quantum Mechanics, Vol. 2.* New York: Wiley, pp. 1048–1058, 1977.
Edmonds, A. R. *Angular Momentum in Quantum Mechanics, 2nd ed., rev. printing.* Princeton, NJ: Princeton University Press, 1968.
Gordy, W. and Cook, R. L. *Microwave Molecular Spectra, 3rd ed.* New York: Wiley, p. 807, 1984.
Messiah, A. "Representation of Irreducible Tensor Operators: Wigner-Eckart Theorem." §32 in *Quantum Mechanics, Vol. 2.* Amsterdam, Netherlands: North-Holland, pp. 573–575, 1962.
Rose, M. E. "The Wigner-Eckart Theorem." §19 in *Elementary Theory of Angular Momentum.* New York: Dover, pp. 85–94, 1995.
Shore, B. W. and Menzel, D. H. "Tensor Operators and the Wigner-Eckart Theorem." §6.4 in *Principles of Atomic Spectra.* New York: Wiley, pp. 285–294, 1968.
Wigner, E. P. *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra, expanded and improved ed.* New York: Academic Press, 1959.
Wybourne, B. G. *Symmetry Principles and Atomic Spectroscopy.* New York: Wiley, pp. 89 and 93–96, 1970.

Wilbraham-Gibbs Constant

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

Let a piecewise smooth function f with only finitely many discontinuities (which are all jumps) be defined on $[-\pi, \pi]$ with FOURIER SERIES

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos(kt)\,dt \tag{1}$$

$$b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin(kt)\,dt, \tag{2}$$

$$S_n(f,x) = \tfrac{1}{2}a_0 + \left\{\sum_{k=1}^{n}[a_k\cos(kx) + b_k\sin(kx)]\right\}. \tag{3}$$

Let a discontinuity be at $x = c$, with

$$\lim_{x\to c^-} f(x) > \lim_{x\to c^+} f(x), \tag{4}$$

so

$$D \equiv \left[\lim_{x\to c^-} f(x)\right] - \left[\lim_{x\to c^+} f(x)\right] > 0. \tag{5}$$

Define

$$\phi(c) = \frac{1}{2}\left[\lim_{x\to c^-} f(x) + \lim_{x\to c^+} f(x)\right], \tag{6}$$

and let $x = x_n < c$ be the first local minimum and $x = \xi_n > c$ the first local maximum of $S_n(f,x)$ on either side of x_n. Then

$$\lim_{n\to\infty} S_n(f, x_n) = \phi(c) + \frac{D}{\pi}G' \tag{7}$$

$$\lim_{n\to\infty} S_n(f, \xi_n) = \phi(c) - \frac{D}{\pi}G', \tag{8}$$

where

$$G' \equiv \int_0^{\pi} \operatorname{sinc}\theta\,d\theta = 1.851937052\ldots. \tag{9}$$

Here, $\operatorname{sinc} x \equiv \sin x/x$ is the SINC FUNCTION. The FOURIER SERIES of $y = x$ therefore does not converge to $-\pi$ and π at the ends, but to $-2G'$ and $2G'$. This phenomenon was observed by Wilbraham (1848) and Gibbs (1899). Although Wilbraham was the first to note the phenomenon, the constant G' is frequently (and unfairly) credited to Gibbs and known as the GIBBS CONSTANT. A related constant sometimes also called the GIBBS CONSTANT is

$$G \equiv \frac{2}{\pi}G' = \frac{2}{\pi}\int_0^{\pi} \operatorname{sinc} x\,dx = 1.17897974447216727\ldots \tag{10}$$

(Le Lionnais 1983).

References

Carslaw, H. S. *Introduction to the Theory of Fourier's Series and Integrals, 3rd ed.* New York: Dover, 1930.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/gibbs/gibbs.html`.

Le Lionnais, F. *Les nombres remarquables.* Paris: Hermann, pp. 36 and 43, 1983.

Zygmund, A. G. *Trigonometric Series 1, 2nd ed.* Cambridge, England: Cambridge University Press, 1959.

Wilcoxon Rank Sum Test

A nonparametric alternative to the two-sample t-test.

see also PAIRED t-TEST, PARAMETRIC TEST

Wilcoxon Signed Rank Test

A nonparametric alternative to the PAIRED t-TEST which is similar to the FISHER SIGN TEST. This test assumes that there is information in the magnitudes of the differences between paired observations, as well as the signs. Take the paired observations, calculate the differences, and rank them from smallest to largest by ABSOLUTE VALUE. Add all the ranks associated with POSITIVE differences, giving the T_+ statistic. Finally, the P-VALUE associated with this statistic is found from an appropriate table. The Wilcoxon test is an R-ESTIMATE.

see also FISHER SIGN TEST, HYPOTHESIS TESTING, PAIRED t-TEST, PARAMETRIC TEST

Wild Knot

A KNOT which is not a TAME KNOT.

see also TAME KNOT

References

Milnor, J. "Most Knots are Wild." *Fund. Math.* **54**, 335–338, 1964.

Wild Point

For any point P on the boundary of an ordinary BALL, find a NEIGHBORHOOD of P in which the intersection with the BALL's boundary cuts the NEIGHBORHOOD into two parts, each HOMEOMORPHIC to a BALL. A wild point is a point on the boundary that has no such NEIGHBORHOOD.

see also BALL, HOMEOMORPHIC, NEIGHBORHOOD

Wilf-Zeilberger Pair

A pair of CLOSED FORM functions (F, G) is said to be a Wilf-Zeilberger pair if

$$F(n+1,k) - F(n,k) = G(n,k+1) - G(n,k). \tag{1}$$

The Wilf-Zeilberger formalism provides succinct proofs of *known* identities and allows new identities to be discovered whenever it succeeds in finding a proof certificate for a known identity. However, if the starting point is an unknown hypergeometric sum, then the Wilf-Zeilberger method cannot discover a closed form solution, while ZEILBERGER'S ALGORITHM can.

Wilf-Zeilberger pairs are very useful in proving HYPERGEOMETRIC IDENTITIES of the form

$$\sum_k t(n,k) = \mathrm{rhs}(n) \tag{2}$$

for which the SUMMAND $t(n,k)$ vanishes for all k outside some finite interval. Now divide by the right-hand side to obtain

$$\sum_k F(n,k) = 1, \tag{3}$$

where

$$F(n,k) \equiv \frac{t(n,k)}{\mathrm{rhs}(n)}. \quad (4)$$

Now use a RATIONAL FUNCTION $R(n,k)$ provided by ZEILBERGER'S ALGORITHM, define

$$G(n,k) \equiv R(n,k)F(n,k). \quad (5)$$

The identity (1) then results. Summing the relation over all integers then telescopes the right side to 0, giving

$$\sum_k F(n+1,k) = \sum_k F(n,k). \quad (6)$$

Therefore, $\sum_k F(n,k)$ is independent of n, and so must be a constant. If F is properly normalized, then it will be true that $\sum_k F(0,k) = 1$.

For example, consider the BINOMIAL COEFFICIENT identity

$$\sum_k = \sum_{k=0}^{n} \binom{n}{k} = 2^n, \quad (7)$$

the function $R(n,k)$ returned by ZEILBERGER'S ALGORITHM is

$$R(n,k) = \frac{k}{2(k-n-1)}. \quad (8)$$

Therefore,

$$F(n,k) = \binom{n}{k} 2^{-n} \quad (9)$$

and

$$\begin{aligned} G(n,k) \equiv R(n,k)F(n,k) &= \frac{k}{2(k-n-1)}\binom{n}{k}2^{-n} \\ &= -\frac{kn!2^{-n}}{2(n+1-k)k!(n-k)!} = -\binom{n}{k-1}2^{-n-1}. \end{aligned} \quad (10)$$

Taking

$$F(n+1,k) - F(n,k) = G(n,k+1) - G(n,k) \quad (11)$$

then gives the alleged identity

$$\begin{aligned} &\binom{n+1}{k}2^{-n-1} - \binom{n}{k}2^{-n} \\ &\quad = -\binom{n}{k}2^{-n-1} + \binom{n}{k-1}2^{-n-1}? \end{aligned} \quad (12)$$

Expanding and evaluating shows that the identity does actually hold, and it can also be verified that

$$F(0,k) = \binom{0}{k} = \begin{cases} 1 & \text{for } k=0 \\ 0 & \text{otherwise,} \end{cases} \quad (13)$$

so $\sum_k F(0,k) = 1$ (Petkovšek *et al.* 1996, pp. 25–27).

For any Wilf-Zeilberger pair (F,G),

$$\sum_{n=0}^{\infty} G(n,0) = \sum_{n=1}^{\infty}[F(n,n-1) + G(n-1,n-1)] \quad (14)$$

whenever either side converges (Zeilberger 1993). In addition,

$$\sum_{n=0}^{\infty} G(n,0) = \sum_{n=0}^{\infty}\left[F(s(n+1),n) + \sum_{i=0}^{s-1} G(sn+i,n)\right], \quad (15)$$

$$\sum_{k=0}^{\infty} F(0,k) = \sum_{n=0}^{\infty} G(n,0), \quad (16)$$

and

$$\begin{aligned} \sum_{n=0}^{\infty} G(n,0) = \sum_{n=0}^{\infty}\Bigg[&\sum_{j=0}^{t-1} F(s(n+1),tn+j) \\ &+ \sum_{i=0}^{s-1} G(sn+i,tn)\Bigg], \end{aligned} \quad (17)$$

where

$$F_{s,t}(n,k) = \sum_{j=0}^{t-1} F(sn,tk+j) \quad (18)$$

$$G_{s,t}(n,k) = \sum_{i=0}^{s-1} G(sn+i,tk) \quad (19)$$

(Amdeberhan and Zeilberger 1997). The latter identity has been used to compute APÉRY'S CONSTANT to a large number of decimal places (Plouffe).

see also APÉRY'S CONSTANT, CONVERGENCE IMPROVEMENT, ZEILBERGER'S ALGORITHM

References

Amdeberhan, T. and Zeilberger, D. "Hypergeometric Series Acceleration via the WZ Method." *Electronic J. Combinatorics* **4**, No. 2, R3, 1–3, 1997. `http://www.combinatorics.org/Volume_4/wilftoc.html#R03`. Also available at `http://www.math.temple.edu/~zeilberg/mamarim/mamarimhtml/accel.html`.

Cipra, B. A. "How the Grinch Stole Mathematics." *Science* **245**, 595, 1989.

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. "The WZ Phenomenon." Ch. 7 in *A=B*. Wellesley, MA: A. K. Peters, pp. 121–140, 1996.

Plouffe, S. "32,000,279 Digits of Zeta(3)." `http://lacim.uqam.ca/piDATA/Zeta3.txt`.

Wilf, H. S. and Zeilberger, D. "Rational Functions Certify Combinatorial Identities." *J. Amer. Math. Soc.* **3**, 147–158, 1990.

Zeilberger, D. "The Method of Creative Telescoping." *J. Symb. Comput.* **11**, 195–204, 1991.

Zeilberger, D. "Closed Form (Pun Intended!)." *Contemporary Math.* **143**, 579–607, 1993.

Wilkie's Theorem

Let $\phi(x_1, \ldots, x_n)$ be an $\mathcal{L}_{\text{exp}}$ formula, where $\mathcal{L}_{\text{exp}} \equiv \mathcal{L} \cup \{e^x\}$ and $\mathcal{L}$ is the language of ordered rings $\mathcal{L} = \{+, -, \cdot, <, 0, 1\}$. Then there are $n \geq m$ and $f_1, \ldots, f_s \in \mathbb{Z}[x_1, \ldots, x_n, e^{x_1}, \ldots, e^{x_n}]$ such that $\phi(x_1, \ldots, x_n)$ is equivalent to

$$\exists x_{m+1} \cdots \exists x_n f_1(x_1, \ldots, x_n, e^{x_1}, \ldots, e^{x_n}) = \ldots$$
$$= f_s(x_1, \ldots, x_n, e^{x_1}, \ldots, e^{x_n}) = 0$$

(Wilkie 1996). In other words, every formula is equivalent to an existential formula and every definable set is the projection of an exponential variety (Marker 1996).

References

Marker, D. "Model Theory and Exponentiation." *Not. Amer. Math. Soc.* **43**, 753–759, 1996.

Wilkie, A. J. "Model Completeness Results for Expansions of the Ordered Field of Real Numbers by Restricted Pfaffian Functions and the Exponential Function." *J. Amer. Math. Soc.* **9**, 1051–1094, 1996.

Williams $p+1$ Factorization Method

A variant of the POLLARD $p-1$ METHOD which uses LUCAS SEQUENCES to achieve rapid factorization if some factor p of N has a decomposition of $p+1$ in small PRIME factors.

see also LUCAS SEQUENCE, POLLARD $p-1$ METHOD, PRIME FACTORIZATION ALGORITHMS

References

Riesel, H. *Prime Numbers and Computer Methods for Factorization, 2nd ed.* Boston, MA: Birkhäuser, p. 177, 1994.

Williams, H. C. "A $p+1$ Method of Factoring." *Math. Comput.* **39**, 225–234, 1982.

Wilson Plug

A 3-D surface with constant VECTOR FIELD on its boundary which traps at least one trajectory which enters it.

see also VECTOR FIELD

Wilson's Primality Test

see WILSON'S THEOREM

Wilson Prime

A PRIME satisfying

$$W(p) \equiv 0 \pmod{p},$$

where $W(p)$ is the WILSON QUOTIENT, or equivalently,

$$(p-1)! \equiv -1 \pmod{p^2}.$$

5, 13, and 563 are the only Wilson primes less than 5×10^8 (Crandall *et al.* 1997).

References

Crandall, R.; Dilcher, K; and Pomerance, C. "A search for Wieferich and Wilson Primes." *Math. Comput.* **66**, 433–449, 1997.

Ribenboim, P. "Wilson Primes." §5.4 in *The New Book of Prime Number Records.* New York: Springer-Verlag, pp. 346–350, 1996.

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 73, 1991.

Wilson Quotient

$$W(p) \equiv \frac{(p-1)! - 1}{p}.$$

Wilson's Theorem

IFF p is a PRIME, then $(p-1)! + 1$ is a multiple of p, that is

$$(p-1)! \equiv -1 \pmod{p}.$$

This theorem was proposed by John Wilson in 1770 and proved by Lagrange in 1773. Unlike FERMAT'S LITTLE THEOREM, Wilson's theorem is both NECESSARY and SUFFICIENT for primality. For a COMPOSITE NUMBER, $(n-1)! \equiv 0 \pmod{n}$ except when $n = 4$.

see also FERMAT'S LITTLE THEOREM, WILSON'S THEOREM COROLLARY, WILSON'S THEOREM (GAUSS'S GENERALIZATION)

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, p. 61, 1987.

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 142–143 and 168–169, 1996.

Ore, Ø. *Number Theory and Its History.* New York: Dover, pp. 259–261, 1988.

Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, pp. 37–38, 1993.

Wilson's Theorem Corollary

Iff a PRIME p is of the form $4x+1$, then

$$[(2x)!]^2 \equiv -1 \pmod{p}.$$

Wilson's Theorem (Gauss's Generalization)

Let P be the product of INTEGERS less than or equal to n and RELATIVELY PRIME to n. Then

$$P \equiv \prod_{\substack{k=2 \\ k \nmid n}}^{n} = \begin{cases} -1 \pmod{n} & \text{for } n = 4, p^\alpha, 2p^\alpha \\ 1 \pmod{n} & \text{otherwise.} \end{cases}$$

When $m = 2$, this reduces to $P \equiv 1 \pmod{2}$ which is equivalent to $P \equiv -1 \pmod{2}$.

see also WILSON'S THEOREM, WILSON'S THEOREM COROLLARY

Winding Number (Contour)

Denoted $n(\gamma, z_0)$ and defined as the number of times a path γ curve passes around a point.

$$n(\gamma, a) = \frac{1}{2\pi i} \int_\gamma \frac{dz}{z-a}.$$

The contour winding number was part of the inspiration for the idea of the DEGREE of a MAP between two COMPACT, oriented MANIFOLDS of the same DIMENSION. In the language of the DEGREE of a MAP, if $\gamma : [0,1] \to \mathbb{C}$

is a closed curve (i.e., $\gamma(0) = \gamma(1)$), then it can be considered as a FUNCTION from $\mathbb{S}^1$ to $\mathbb{C}$. In that context, the winding number of γ around a point p in $\mathbb{C}$ is given by the degree of the MAP

$$\frac{\gamma - p}{|\gamma - p|}$$

from the CIRCLE to the CIRCLE.

Winding Number (Map)

The winding number of a map is defined by

$$W \equiv \lim_{n\to\infty} \frac{f^n(\theta) - \theta}{n},$$

which represents the average increase in the angle θ per unit time (average frequency). A system with a RATIONAL winding number $W = p/q$ is MODE-LOCKED, whereas a system with an IRRATIONAL winding number is QUASIPERIODIC. Note that since the RATIONALS are a set of zero MEASURE on any finite interval, almost all winding numbers will be irrational, so almost all maps will be QUASIPERIODIC.

Windmill

One name for the figure used by Euclid to prove the PYTHAGOREAN THEOREM.

see BRIDE'S CHAIR, PEACOCK'S TAIL

Window Function

see RECTANGLE FUNCTION

Winkler Conditions

Conditions arising in the study of the ROBBINS EQUATION and its connection with BOOLEAN ALGEBRA. Winkler studied Boolean conditions (such as idempotence or existence of a zero) which would make a ROBBINS ALGEBRA become a BOOLEAN ALGEBRA. Winkler showed that each of the conditions

$$\exists C, \exists D, C + D = C$$

$$\exists C, \exists D, n(C + D) = n(C),$$

known as the first and second Winkler conditions, SUFFICES. A computer proof demonstrated that every ROBBINS ALGEBRA satisfies the second Winkler condition, from which it follows immediately that all ROBBINS ALGEBRAS are BOOLEAN.

References

McCune, W. "Robbins Algebras are Boolean." `http://www.mcs.anl.gov/home/mccune/ar/robbins/`.

Winkler, S. "Robbins Algebra: Conditions that Make a Near-Boolean Algebra Boolean." *J. Automated Reasoning* **6**, 465–489, 1990.

Winkler, S. "Absorption and Idempotency Criteria for a Problem in Near-Boolean Algebra." *J. Algebra* **153**, 414–423, 1992.

Winograd Transform

A discrete FAST FOURIER TRANSFORM ALGORITHM which can be implemented for $N = 2, 3, 4, 5, 7, 8, 11, 13$, and 16 points.

see also FAST FOURIER TRANSFORM

Wirtinger's Inequality

If y has period 2π, y' is L^2, and

$$\int_0^{2\pi} y\,dx = 0,$$

then

$$\int_0^{2\pi} y^2\,dx < \int_0^{2\pi} y'^2\,dx$$

unless

$$y = A\cos x + B\sin x.$$

References

Hardy, G. H.; Littlewood, J. E.; and Pólya, G. *Inequalities, 2nd ed.* Cambridge, England: Cambridge University Press, pp. 184–187, 1988.

Wirtinger-Sobolev Isoperimetric Constants

Constants γ such that

$$\left[\int_\Omega |f|^q\,dx\right]^{1/q} \leq \gamma \left[\int_\Omega \sum_{i=1}^N \left|\frac{\partial f}{\partial x_i}\right|^p\,dx\right]^{1/p},$$

where f is a real-valued smooth function on a region Ω satisfying some BOUNDARY CONDITIONS.

References

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/ws/ws.html`.

Witch of Agnesi

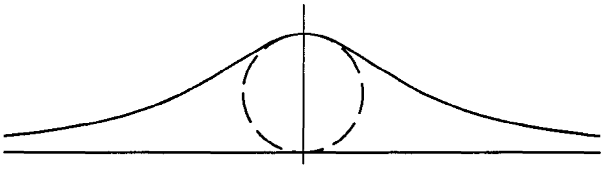

A curve studied and named "versiera" (Italian for "she-devil" or "witch") by Maria Agnesi in 1748 in her book *Istituzioni Analitiche* (MacTutor Archive). It is also known as CUBIQUE D'AGNESI or AGNÉSIENNE. Some suggest that Agnesi confused an old Italian word meaning "free to move" with another meaning "witch." The curve had been studied earlier by Fermat and Guido Grandi in 1703.

It is the curve obtained by drawing a line from the origin through the CIRCLE of radius $2a$ (OB), then picking the point with the y coordinate of the intersection with the circle and the x coordinate of the intersection of the extension of line OB with the line $y = 2a$. The curve has INFLECTION POINTS at $y = 3a/2$. The line $y = 0$ is an ASYMPTOTE to the curve.

In parametric form,

$$x = 2a \cot \theta \tag{1}$$
$$y = a[1 - \cos(2\theta)], \tag{2}$$

or

$$x = 2at \tag{3}$$
$$y = \frac{2a}{1+t^2}. \tag{4}$$

In rectangular coordinates,

$$y = \frac{8a^3}{x^2 + 4a^2}. \tag{5}$$

see also LAMÉ CURVE

References
Lawrence, J. D. *A Catalog of Special Plane Curves.* New York: Dover, pp. 90–93, 1972.
Lee, X. "Witch of Agnesi." `http://www.best.com/~xah/SpecialPlaneCurves_dir/WitchOfAgnesi_dir/witchOfAgnesi.html`.
MacTutor History of Mathematics Archive. "Witch of Agnesi." `http://www-groups.dcs.st-and.ac.uk/~history/Curves/Witch.html`.
Yates, R. C. "Witch of Agnesi." *A Handbook on Curves and Their Properties.* Ann Arbor, MI: J. W. Edwards, pp. 237–238, 1952.

Witness

A witness is a number which, as a result of its number theoretic properties, guarantees either the compositeness or primality of a number n. Witnesses are most commonly used in connection with FERMAT'S LITTLE THEOREM CONVERSE. A PRATT CERTIFICATE uses witnesses to prove primality, and MILLER'S PRIMALITY TEST uses witnesses to prove compositeness.

see also ADLEMAN-POMERANCE-RUMELY PRIMALITY TEST, FERMAT'S LITTLE THEOREM CONVERSE, MILLER'S PRIMALITY TEST, PRATT CERTIFICATE, PRIMALITY CERTIFICATE

Witten's Equations

Also called the SEIBERG-WITTEN INVARIANTS. For a connection A and a POSITIVE SPINOR $\phi \in \Gamma(V_+)$,

$$D_A\phi = 0$$
$$F_+^A = i\sigma(\phi, \phi).$$

The solutions are called monopoles and are the minima of the functional

$$\int_X (|F_+^A - i\sigma(\phi,\phi)|^2 + |D_A\phi|^2).$$

see also LICHNEROWICZ FORMULA, LICHNEROWICZ-WEITZENBOCK FORMULA, SEIBERG-WITTEN EQUATIONS

References
Cipra, B. "A Tale of Two Theories." *What's Happening in the Mathematical Sciences, 1995-1996, Vol. 3.* Providence, RI: Amer. Math. Soc., pp. 14–25, 1996.
Donaldson, S. K. "The Seiberg-Witten Equations and 4-Manifold Topology." *Bull. Amer. Math. Soc.* **33**, 45–70, 1996.
Kotschick, D. "Gauge Theory is Dead!—Long Live Gauge Theory!" *Not. Amer. Math. Soc.* **42**, 335–338, 1995.
Seiberg, N. and Witten, E. "Monopoles, Duality, and Chiral Symmetry Breaking in $N = 2$ Supersymmetric QCD." *Nucl. Phys. B* **431**, 581–640, 1994.
Witten, E. "Monopoles and 4-Manifolds." *Math. Res. Let.* **1**, 769–796, 1994.

Wittenbauer's Parallelogram

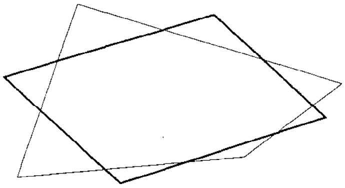

Divide the sides of a QUADRILATERAL into three equal parts. The figure formed by connecting and extending adjacent points on either side of a VERTEX is a PARALLELOGRAM known as Wittenbauer's parallelogram.

see also QUADRILATERAL, WITTENBAUER'S THEOREM

Wittenbauer's Theorem

The CENTROID of a QUADRILATERAL LAMINA is the center of its WITTENBAUER'S PARALLELOGRAM.

see also CENTROID (GEOMETRIC), LAMINA, QUADRILATERAL, WITTENBAUER'S PARALLELOGRAM

Wolstenholme's Theorem

If p is a PRIME > 3, then the NUMERATOR of

$$1 + \tfrac{1}{2} + \tfrac{1}{3} + \ldots + \tfrac{1}{p-1}$$

is divisible by p^2 and the NUMERATOR of

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \ldots + \frac{1}{(p-1)^2}$$

is divisible by p. These imply that if $p \geq 5$ is PRIME, then

$$\binom{2p-1}{p-1} \equiv 1 \pmod{p^3}.$$

References
Guy, R. K. *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 85, 1994.
Ribenboim, P. *The Book of Prime Number Records, 2nd ed.* New York: Springer-Verlag, p. 21, 1989.

Woodall Number

Numbers of the form

$$W_n = 2^n n - 1.$$

The first few are 1, 7, 23, 63, 159, 383, ... (Sloane's A003261). The only Woodall numbers W_n for $n < 100,000$ which are PRIME are for $n = 5312$, 7755, 9531, 12379, 15822, 18885, 22971, 23005, 98726, ... (Sloane's A014617; Ballinger).

see also CULLEN NUMBER, CUNNINGHAM NUMBER, FERMAT NUMBER, MERSENNE NUMBER, SIERPIŃSKI NUMBER OF THE FIRST KIND

References

Ballinger, R. "Cullen Primes: Definition and Status." `http://ballingerr.xray.ufl.edu/proths/cullen.html`.

Guy, R. K. "Cullen Numbers." §B20 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 77, 1994.

Leyland, P. `ftp://sable.ox.ac.uk/pub/math/factors/woodall`.

Ribenboim, P. *The New Book of Prime Number Records.* New York: Springer-Verlag, pp. 360–361, 1996.

Sloane, N. J. A. Sequences A014617 and A003261/M4379 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Woodbury Formula

$$(\mathsf{A}+\mathsf{U}\mathsf{V}^{\mathrm{T}})^{-1} = \mathsf{A}^{-1} - [\mathsf{A}^{-1}\mathsf{U}(1+\mathsf{V}^{\mathrm{T}}\mathsf{A}^{-1}\mathsf{U})^{-1}\mathsf{V}^{\mathrm{T}}\mathsf{A}^{-1}].$$

Word

N.B. A detailed on-line essay by S. Finch was the starting point for this entry.

A finite sequence of n letters from some ALPHABET is said to be an n-ary word. A "square" word consists of two identical subwords (for example, *acbacb*). A squarefree word contains *no* square words as subwords (for example, *abcacbabcb*). The only squarefree binary words are *a*, *b*, *ab*, *ba*, *aba*, and *bab*. However, there are arbitrarily long ternary squarefree words. The number of ternary squarefree words of length n is bounded by

$$6 \cdot 1.032^n \leq s(n) \leq 6 \cdot 1.379^n \tag{1}$$

(Brandenburg 1983). In addition,

$$S \equiv \lim_{n\to\infty} [s(n)]^{1/n} = 1.302\ldots \tag{2}$$

(Brinkhuis 1983, Noonan and Zeilberger). Binary cubefree words satisfy

$$2 \cdot 1.080^n \leq c(n) \leq 2 \cdot 1.522^n. \tag{3}$$

A word is said to be overlapfree if it has no subwords of the form *xyxyx*. A squarefree word is overlapfree, and an overlapfree word is cubefree. The number $t(n)$ of binary overlapfree words of length n satisfies

$$p \cdot n^{1.155} \leq t(n) \leq q \cdot n^{1.587} \tag{4}$$

for some constants p and q (Restivo and Selemi 1985, Kobayashi 1988). In addition, while

$$\lim_{n\to\infty} \frac{\ln t(n)}{\ln n} \tag{5}$$

does not exist,

$$1.155 < T_L < 1.276 < 1.332 < T_U < 1.587, \tag{6}$$

where

$$T_L \equiv \liminf_{n\to\infty} \frac{\ln t(n)}{\ln n} \tag{7}$$

$$T_U \equiv \limsup_{n\to\infty} \frac{\ln t(n)}{\ln n} \tag{8}$$

(Cassaigne 1993).

see also ALPHABET

References

Brandenburg, F.-J. "Uniformly Growing *k*th Power-Free Homomorphisms." *Theor. Comput. Sci.* **23**, 69–82, 1983.

Brinkhuis, J. "Non-Repetitive Sequences on Three Symbols." *Quart. J. Math. Oxford Ser. 2* **34**, 145–149, 1983.

Cassaigne, J. "Counting Overlap-Free Binary Words." *STACS '93: Tenth Annual Symposium on Theoretical Aspects of Computer Science, Würzburg, Germany, February 25–27, 1993 Proceedings* (Ed. G. Goos, J. Hartmanis, A. Finkel, P. Enjalbert, K. W. Wagner). New York: Springer-Verlag, pp. 216–225, 1993.

Finch, S. "Favorite Mathematical Constants." `http://www.mathsoft.com/asolve/constant/words/words.html`.

Kobayashi, Y. "Enumeration of Irreducible Binary Words." *Discrete Appl. Math.* **20**, 221–232, 1988.

Noonan, J. and Zeilberger, D. "The Goulden-Jackson Cluster Method: Extensions, Applications, and Implementations." Submitted.

World Line

The path of an object through PHASE SPACE.

Worm

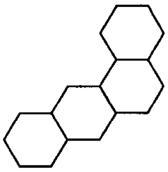

A 4-POLYHEX.

References

Gardner, M. *Mathematical Magic Show: More Puzzles, Games, Diversions, Illusions and Other Mathematical Sleight-of-Mind from Scientific American.* New York: Vintage, p. 147, 1978.

Worpitzky's Identity

$$x^n = \sum_{k=0}^{n-1} \left\langle {n \atop k} \right\rangle \binom{n+k}{n},$$

where $\left\langle {n \atop k} \right\rangle$ is an EULERIAN NUMBER and $\binom{n}{k}$ is a BINOMIAL COEFFICIENT.

Writhe

Also called the TWIST NUMBER. The sum of crossings p of a LINK L,

$$w(L) = \sum_{p \in C(L)} \epsilon(p),$$

where $\epsilon(p)$ defined to be ± 1 if the overpass slants from top left to bottom right or bottom left to top right and $C(L)$ is the set of crossings of an oriented LINK. If a KNOT K is AMPHICHIRAL, then $w(K) = 0$ (Thistlethwaite). Letting Lk be the LINKING NUMBER of the two components of a ribbon, Tw be the TWIST, and Wr be the writhe, then

$$\mathrm{Lk}(K) = \mathrm{Tw}(K) + \mathrm{Wr}(K).$$

(Adams 1994, p. 187).

see also SCREW, TWIST

References

Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: W. H. Freeman, 1994.

Wronskian

$$W(\phi_1, \ldots, \phi_n) \equiv \begin{vmatrix} \phi_1 & \phi_2 & \cdots & \phi_n \\ \phi_1' & \phi_2' & \cdots & \phi_n' \\ \vdots & \vdots & \ddots & \vdots \\ {\phi_1}^{(n-1)} & {\phi_2}^{(n-1)} & \cdots & {\phi_n}^{(n-1)}. \end{vmatrix}$$

If the Wronskian is NONZERO in some region, the functions ϕ_i are LINEARLY INDEPENDENT. If $W = 0$ over some range, the functions are linearly dependent somewhere in the range.

see also ABEL'S IDENTITY, GRAM DETERMINANT, LINEARLY DEPENDENT FUNCTIONS

References

Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 524–525, 1953.

Wulff Shape

An equilibrium MINIMAL SURFACE for a crystal which has the least anisotropic surface energy for a given volume. It is the anisotropic analog of a SPHERE.

see also SPHERE

Wynn's Epsilon Method

A method for numerical evaluation of SUMS and PRODUCTS which samples a number of additional terms in the series and then tries to fit them to a POLYNOMIAL multiplied by a decaying exponential.

see also EULER-MACLAURIN INTEGRATION FORMULAS

Wythoff Array

A INTERSPERSION array given by

$$\begin{array}{cccccccccc} 1 & 2 & 3 & 5 & 8 & 13 & 21 & 34 & 55 & \cdots \\ 4 & 7 & 11 & 18 & 29 & 47 & 76 & 123 & 199 & \cdots \\ 6 & 10 & 16 & 26 & 42 & 68 & 110 & 178 & 288 & \cdots \\ 9 & 15 & 24 & 39 & 63 & 102 & 165 & 267 & 432 & \cdots \\ 12 & 20 & 32 & 52 & 84 & 136 & 220 & 356 & 576 & \cdots \\ 14 & 23 & 37 & 60 & 97 & 157 & 254 & 411 & 665 & \cdots \\ 17 & 28 & 45 & 73 & 118 & 191 & 309 & 500 & 809 & \cdots \\ 19 & 31 & 50 & 81 & 131 & 212 & 343 & 555 & 898 & \cdots \\ 22 & 36 & 58 & 94 & 152 & 246 & 398 & 644 & 1042 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \end{array}$$

the first row of which is the FIBONACCI NUMBERS.

see also FIBONACCI NUMBER, INTERSPERSION, STOLARSKY ARRAY

References

Kimberling, C. "Fractal Sequences and Interspersions." *Ars Combin.* **45**, 157–168, 1997.

Wythoff Construction

A method of constructing UNIFORM POLYHEDRA.

see also UNIFORM POLYHEDRON

References

Har'El, Z. "Uniform Solution for Uniform Polyhedra." *Geometriae Dedicata* **47**, 57–110, 1993.

Wythoff's Game

A game played with two heaps of counters in which a player may take any number from either heap or the same number from both. The player taking the last counter wins. The rth SAFE combination is $(x, x+r)$, where $x = \lfloor \phi r \rfloor$, with ϕ the GOLDEN RATIO and $\lfloor x \rfloor$ the FLOOR FUNCTION. It is also true that $x + r = \lfloor \phi^2 r \rfloor$. The first few SAFE combinations are (1, 2), (3, 5), (4, 7), (6, 10), ...

see also NIM, SAFE

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 39–40, 1987.

Coxeter, H. S. M. "The Golden Section, Phyllotaxis, and Wythoff's Game." *Scripta Math.* **19**, 135–143, 1953.

O'Beirne, T. H. *Puzzles and Paradoxes.* Oxford, England: Oxford University Press, pp. 109 and 134–138, 1965.

Wythoff Symbol

A symbol used to describe UNIFORM POLYHEDRA. For example, the Wythoff symbol for the TETRAHEDRON is $3 \,|\, 2\,3$. There are three types of Wythoff symbols $p \,|\, q\, r$, $p\, q \,|\, r$ and $p\, q\, r \,|$, and one exceptional symbol $|\, \frac{3}{2}\, \frac{5}{3}\, 3\, \frac{5}{2}$ used for the GREAT DIRHOMBICOSIDODECAHEDRON. Some special cases in terms of SCHLÄFLI SYMBOLS are

$$p \,|\, q\, 2 = p \,|\, 2\, q = \{q, p\}$$

$$2 \,|\, p\, q = \begin{Bmatrix} p \\ q \end{Bmatrix}$$

$$p\, q \,|\, 2 = r \begin{Bmatrix} p \\ q \end{Bmatrix}$$

$$2\, q \,|\, p = \mathrm{t}\{p, q\}$$

$$2\, p\, q \,| = \mathrm{t} \begin{Bmatrix} p \\ q \end{Bmatrix}$$

$$|\, 2\, p\, q = \mathrm{s} \begin{Bmatrix} p \\ q \end{Bmatrix}.$$

For the symbol $p\, q\, r \,|$, permuting the letters gives the same POLYHEDRON.

see also UNIFORM POLYHEDRON

References

Har'El, Z. "Uniform Solution for Uniform Polyhedra." *Geometriae Dedicata* **47**, 57–110, 1993.

X

x-Axis

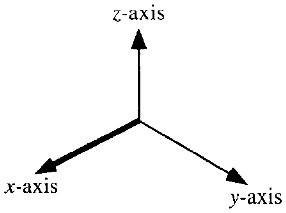

The horizontal axis of a 2-D plot in CARTESIAN COORDINATES, also called the ABSCISSA.

see also ABSCISSA, ORDINATE, y-AXIS, z-AXIS

x-Intercept

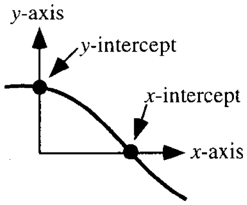

The point at which a curve or function crosses the x-AXIS (i.e., when $y = 0$ in 2-D).

see also LINE, y-INTERCEPT

Xi Function

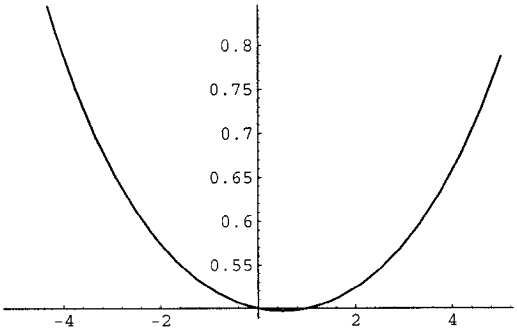

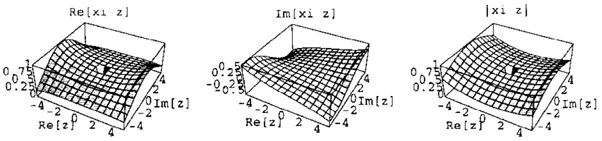

$$\xi(z) \equiv \tfrac{1}{2}z(z-1)\frac{\Gamma(\frac{1}{2}z)}{\pi^{z/2}}\zeta(z) = \frac{(z-1)\Gamma(\frac{1}{2}z+1)\zeta(z)}{\sqrt{\pi^z}}, \tag{1}$$

where $\zeta(z)$ is the RIEMANN ZETA FUNCTION and $\Gamma(z)$ is the GAMMA FUNCTION (Gradshteyn and Ryzhik 1980, p. 1076). The ξ function satisfies the identity

$$\xi(1-z) = \xi(z). \tag{2}$$

The zeros of $\xi(z)$ and of its DERIVATIVES are all located on the CRITICAL STRIP $z = \sigma + it$, where $0 < \sigma < 1$. Therefore, the nontrivial zeros of the RIEMANN ZETA FUNCTION exactly correspond to those of $\xi(z)$. The function $\xi(z)$ is related to what Gradshteyn and Ryzhik (1980, p. 1074) call $\Xi(t)$ by

$$\Xi(t) \equiv \xi(z), \tag{3}$$

where $z \equiv \frac{1}{2} + it$. This function can also be defined as

$$\Xi(it) \equiv \tfrac{1}{2}(t^2 - \tfrac{1}{4})\pi^{-t/2-1/4}\Gamma(\tfrac{1}{2}t + \tfrac{1}{4})\zeta(t + \tfrac{1}{2}), \tag{4}$$

giving

$$\Xi(t) = -\tfrac{1}{2}(t^2 + \tfrac{1}{4})\pi^{it/2-1/4}\Gamma(\tfrac{1}{4} - \tfrac{1}{2}it)\zeta(\tfrac{1}{2} - it). \tag{5}$$

The DE BRUIJN-NEWMAN CONSTANT is defined in terms of the $\Xi(t)$ function.

see also DE BRUIJN-NEWMAN CONSTANT

References

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, corr. enl. 4th ed.* San Diego, CA: Academic Press, 1980.

XOR

An operation in LOGIC known as EXCLUSIVE OR. It yields true if exactly one (but not both) of two conditions is true. The BINARY XOR operator has the following TRUTH TABLE.

A	B	A XOR B
F	F	F
F	T	T
T	F	T
T	T	F

The BINOMIAL COEFFICIENT $\binom{m}{n}$ mod 2 can be computed using the XOR operation n XOR m, making PASCAL'S TRIANGLE mod 2 very easy to construct.

see also AND, BINARY OPERATOR, BOOLEAN ALGEBRA, LOGIC, NOT, OR, PASCAL'S TRIANGLE, TRUTH TABLE

Y

y-Axis

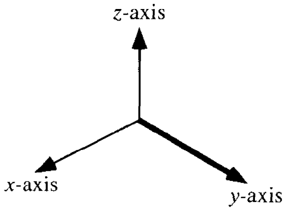

The vertical axis of a 2-D plot in CARTESIAN COORDINATES, also called the ORDINATE.

see also ABSCISSA, ORDINATE, x-AXIS, z-AXIS

y-Intercept

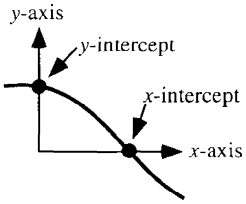

The point at which a curve or function crosses the y-AXIS (i.e., when $x = 0$ in 2-D).

see also LINE, x-INTERCEPT

Yacht

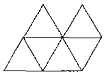

A 6-POLYIAMOND.

References

Golomb, S. W. *Polyominoes: Puzzles, Patterns, Problems, and Packings, 2nd ed.* Princeton, NJ: Princeton University Press, p. 92, 1994.

Yanghui Triangle

see PASCAL'S TRIANGLE

Yff Center of Congruence

Let three ISOSCELIZERS, one for each side, be constructed on a TRIANGLE such that the four interior triangles they determine are congruent. Now parallel-displace these ISOSCELIZERS until they concur in a single point. This point is called the Yff center of congruence and has TRIANGLE CENTER FUNCTION

$$\alpha = \sec(\tfrac{1}{2}A).$$

see also CONGRUENT ISOSCELIZERS POINT, ISOSCELIZER

References

Kimberling, C. "Yff Center of Congruence." `http://www.evansville.edu/~ck6/tcenters/recent/yffcc.html`.

Yff Points

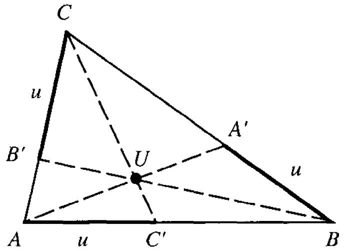

Let points A', B', and C' be marked off some fixed distance x along each of the sides BC, CA, and AB. Then the lines AA', BB', and CC' concur in a point U known as the first Yff point if

$$x^3 = (a - x)(b - x)(c - x). \tag{1}$$

This equation has a single real root u, which can by obtained by solving the CUBIC EQUATION

$$f(x) = 2x^3 - px^2 + qx - r = 0, \tag{2}$$

where

$$p = a + b + c \tag{3}$$

$$q = ab + ac + bc \tag{4}$$

$$r = abc. \tag{5}$$

The ISOTOMIC CONJUGATE POINT U' is called the second Yff point. The TRIANGLE CENTER FUNCTIONS of the first and second points are given by

$$\alpha = \frac{1}{a}\left(\frac{c-u}{b-u}\right)^{1/3} \tag{6}$$

and

$$\alpha' = \frac{1}{a}\left(\frac{b-u}{c-u}\right)^{1/3}, \tag{7}$$

respectively. Analogous to the inequality $\omega \leq \pi/6$ for the BROCARD ANGLE ω, $u \leq p/6$ holds for the Yff points, with equality in the case of an EQUILATERAL TRIANGLE. Analogous to

$$\omega < \alpha_i < \pi - 3\omega \tag{8}$$

for $i = 1, 2, 3$, the Yff points satisfy

$$u < a_i < p - 3u. \tag{9}$$

Yff (1963) gives a number of other interesting properties. The line UU' is PERPENDICULAR to the line containing the INCENTER I and CIRCUMCENTER O, and its length is given by

$$\overline{UU'} = \frac{4u\overline{IO}\Delta}{u^3 + abc}, \tag{10}$$

where Δ is the AREA of the TRIANGLE.

see also BROCARD POINTS, YFF TRIANGLES

References

Yff, P. "An Analog of the Brocard Points." *Amer. Math. Monthly* **70**, 495–501, 1963.

Yff Triangles

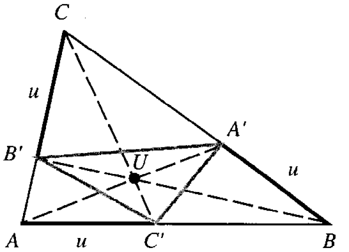

The TRIANGLE $\Delta A'B'C'$ formed by connecting the points used to construct the YFF POINTS is called the first Yff triangle. The AREA of the triangle is

$$\Delta = \frac{u^3}{2R},$$

where R is the CIRCUMRADIUS of the original TRIANGLE ΔABC. The second Yff triangle is formed by connecting the ISOTOMIC CONJUGATE POINTS of A', B', and C'.

see also YFF POINTS

References
Yff, P. "An Analog of the Brocard Points." *Amer. Math. Monthly* **70**, 495–501, 1963.

Yin-Yang

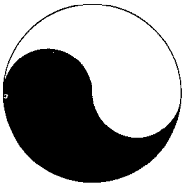

A figure used in many Asian cultures to symbolize the unity of the two "opposite" male and female elements, the "yin" and "yang." The solid and hollow parts composing the symbol are similar and combine to make a CIRCLE. Each part consists of two equal oppositely oriented SEMICIRCLES of radius 1/2 joined at their edges, plus a SEMICIRCLE of radius 1 joining the other edges.

see also BASEBALL COVER, CIRCLE, PIECEWISE CIRCULAR CURVE, SEMICIRCLE

References
Dixon, R. *Mathographics.* New York: Dover, p. 11, 1991.
Gardner, M. "Mathematical Games: A New Collection of 'Brain-Teasers.'" *Sci. Amer.* **203**, 172–180, Oct. 1960.
Gardner, M. "Mathematical Games: More About the Shapes that Can Be Made with Complex Dominoes." *Sci. Amer.* **203**, 186–198, Nov. 1960.

Young Diagram

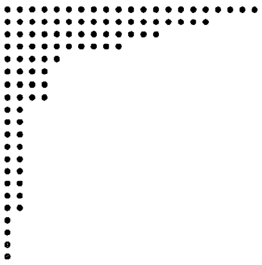

A Young diagram, also called a FERRERS DIAGRAM, represents PARTITIONS as patterns of dots, with the nth row having the same number of dots as the nth term in the PARTITION. A Young diagram of the PARTITION

$$n = a + b + \ldots + c,$$

for a list $a, b, \ldots, c$ of k POSITIVE INTEGERS with $a \geq b \geq \ldots \geq c$ is therefore the arrangement of n dots or square boxes in k rows, such that the dots or boxes are left-justified, the first row is of length a, the second row is of length b, and so on, with the kth row of length c. The above diagram corresponds to one of the possible partitions of 100.

see also DURFEE SQUARE, HOOK LENGTH FORMULA, PARTITION, PARTITION FUNCTION P, YOUNG TABLEAU

References
Messiah, A. Appendix D in *Quantum Mechanics, 2 vols.* Amsterdam, Netherlands: North-Holland, p. 1113, 1961–62.

Young Girl-Old Woman Illusion

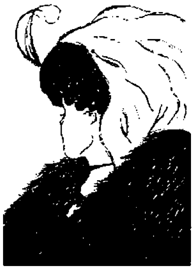

A perceptual ILLUSION in which the brain switches between seeing a young girl and an old woman.

see also RABBIT-DUCK ILLUSION

References
Pappas, T. *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 173, 1989.

Young Inequality

For $0 < p < 1$,

$$ab \leq \frac{a^p}{p} + \left(1 - \frac{1}{p}\right) b^{1/(1-1/p)}.$$

Young's Integral

Let $f(x)$ be a REAL continuous monotonic strictly increasing function on the interval $[0, a]$ with $f(0) = 0$ and $b \leq f(a)$, then

$$ab \leq \int_0^a f(x)\,dx + \int_0^b f^{-1}(y)\,dy,$$

where $f^{-1}(y)$ is the INVERSE FUNCTION. Equality holds IFF $b = f(a)$.

References
Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic Press, p. 1099, 1979.

Young Tableau

The YOUNG TABLEAU of a YOUNG DIAGRAM is obtained by placing the numbers 1, ..., n in the n boxes of the diagram. A "standard" Young tableau is a Young tableau in which the numbers form a nondecreasing sequence along each line and along each column. The standard Young tableaux of size three are given by $\{\{1,2,3\}\}$, $\{\{1,3\},\{2\}\}$, $\{\{1,2\},\{3\}\}$, and $\{\{1\},\{2\},\{3\}\}$. The number of standard Young tableaux of size 1, 2, 3, ... are 1, 2, 4, 10, 26, 76, 232, 764, 2620, 9496, ... (Sloane's A000085). These numbers can be generated by the RECURRENCE RELATION

$$a(n) = a(n-1) + (n-1)a(n-2)$$

with $a(1) = 1$ and $a(2) = 2$.

There is a correspondence between a PERMUTATION and a pair of Young tableaux, known as the SCHENSTED CORRESPONDENCE. The number of all standard Young tableaux with a given shape (corresponding to a given YOUNG DIAGRAM) is calculated with the HOOK LENGTH FORMULA. The BUMPING ALGORITHM is used to construct a standard Young tableau from a permutation of $\{1, \ldots, n\}$.

see also BUMPING ALGORITHM, HOOK LENGTH FORMULA, INVOLUTION (SET), SCHENSTED CORRESPONDENCE, YOUNG DIAGRAM

References

Fulton, W. *Young Tableaux with Applications to Representation Theory and Geometry.* New York: Cambridge University Press, 1996.

Ruskey, F. "Information on Permutations." `http://sue.csc.uvic.ca/~cos/inf/perm/PermInfo.html#Tableau`.

Skiena, S. S. *The Algorithm Design Manual.* New York: Springer-Verlag, pp. 254–255, 1997.

Sloane, N. J. A. Sequence A000085/M1221 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Z

$\mathbb{Z}$

The RING of INTEGERS ..., −2, −1, 0, 1, 2, ..., also denoted $\mathbb{I}$.

see also $\mathbb{C}$, $\mathbb{C}^*$, COUNTING NUMBER, $\mathbb{I}$, $\mathbb{N}$, NATURAL NUMBER, $\mathbb{Q}$, $\mathbb{R}$, WHOLE NUMBER, $\mathbb{Z}^-$, $\mathbb{Z}^-$

$\mathbb{Z}^-$

The NEGATIVE INTEGERS ..., −3, −2, −1.

see also COUNTING NUMBER, NATURAL NUMBER, NEGATIVE, WHOLE NUMBER, $\mathbb{Z}$, $\mathbb{Z}^+$, $\mathbb{Z}^*$

$\mathbb{Z}^+$

The POSITIVE INTEGERS 1, 2, 3, ..., equivalent to $\mathbb{N}$.

see also COUNTING NUMBER, $\mathbb{N}$, NATURAL NUMBER, POSITIVE, WHOLE NUMBER, $\mathbb{Z}$, $\mathbb{Z}^-$, $\mathbb{Z}^*$

$\mathbb{Z}^*$

The NONNEGATIVE INTEGERS 0, 1, 2, ...

see also COUNTING NUMBER, NATURAL NUMBER, NONNEGATIVE, WHOLE NUMBER $\mathbb{Z}$, $\mathbb{Z}^-$, $\mathbb{Z}^+$

z-Axis

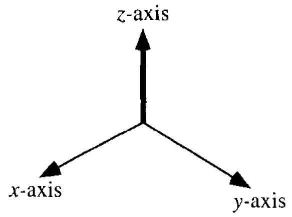

The axis in 3-D CARTESIAN COORDINATES which is usually oriented vertically. CYLINDRICAL COORDINATES are defined such that the *z*-axis is the axis about which the azimuthal coordinate θ is measured.

see also AXIS, *x*-AXIS, *y*-AXIS

z-Distribution

see FISHER'S *z*-DISTRIBUTION, STUDENT'S *z*-DISTRIBUTION

Z-Number

A *Z*-number is a REAL NUMBER ξ such that

$$0 \le \operatorname{frac}\left[\left(\frac{3}{2}\right)^k \xi\right] < \tfrac{1}{2}$$

for all $k = 1, 2, \ldots$, where $\operatorname{frac}(x)$ is the fractional part of x. Mahler (1968) showed that there is at most one *Z*-number in each interval $[n, n+1)$ for integral n. Mahler (1968) therefore concluded that it is unlikely that any *Z*-numbers exist. The *Z*-numbers arise in the analysis of the COLLATZ PROBLEM.

see also COLLATZ PROBLEM

References

Flatto, L. "*Z*-Numbers and β-Transformations." *Symbolic Dynamics and its Applications, Contemporary Math.* **135**, 181–201, 1992.

Guy, R. K. "Mahler's *Z*-Numbers." §E18 in *Unsolved Problems in Number Theory, 2nd ed.* New York: Springer-Verlag, p. 220, 1994.

Lagarias, J. C. "The $3x+1$ Problem and its Generalizations." *Amer. Math. Monthly* **92**, 3–23, 1985. `http://www.cecm.sfu.ca/organics/papers/lagarias/`.

Mahler, K. "An Unsolved Problem on the Powers of 3/2." *Austral. Math. Soc.* **8**, 313–321, 1968.

Tijdman, R. "Note on Mahler's $\frac{3}{2}$-Problem." *Kongel. Norske Vidensk Selsk. Skr.* **16**, 1–4, 1972.

z-Score

The *z*-score associated with the *i*th observation of a random variable x is given by

$$z_i \equiv \frac{x_i - \bar{x}}{\sigma},$$

where $\bar{x}$ is the MEAN and σ the STANDARD DEVIATION of all observations $x_1, \ldots, x_n$.

z-Transform

The discrete *z*-transform is defined as

$$\mathcal{Z}[a] = \sum_{n=0}^{N-1} a_n z^{kn}. \tag{1}$$

The DISCRETE FOURIER TRANSFORM is a special case of the *z*-transform with

$$z \equiv e^{-2\pi i/N}. \tag{2}$$

A *z*-transform with

$$z \equiv e^{-2\pi i\alpha/N} \tag{3}$$

for $\alpha \neq \pm 1$ is called a FRACTIONAL FOURIER TRANSFORM.

see also DISCRETE FOURIER TRANSFORM, FRACTIONAL FOURIER TRANSFORM

References

Arndt, J. "The *z*-Transform (ZT)." Ch. 3 in "Remarks on FFT Algorithms." `http://www.jjj.de/fxt/`.

z-Transform (Population)

see POPULATION COMPARISON

Z-Transform

The *Z*-transform of $F(t)$ is defined by

$$Z[F(t)] = \mathcal{L}[F^*(t)], \tag{1}$$

where

$$F^*(t) = F(t)\delta_T(t) = \sum_{n=0}^{\infty} F(nT)\delta(t - nT), \tag{2}$$

$\delta(t)$ is the DELTA FUNCTION, T is the sampling period, and $\mathcal{L}$ is the LAPLACE TRANSFORM. An alternative definition is

$$Z[F(t)] = \sum_{\text{residues}} \left(\frac{1}{1-e^{Tz}z^{-1}}\right) f(z), \tag{3}$$

where

$$f(z) = \sum_{n=0}^{\infty} F(nT)z^{-n}. \tag{4}$$

The inverse Z-transform is

$$Z^{-1}[f(z)] = F^{*}(t) = \frac{1}{2\pi i}\oint f(z)z^{n-1}\,dz. \tag{5}$$

It satisfies

$$Z[aF(t)+bG(t)] = aZ[F(t)]+bZ[F(t)] \tag{6}$$
$$Z[F(t+T)] = zZ[F(t)] - zF(0) \tag{7}$$
$$Z[F(t+2T)] = z^2Z[F(t)] - z^2F(0) - zF(t) \tag{8}$$
$$Z[F(t+mT)] = z^mZ[F(t)] - \sum_{r=0}^{m-1} z^{m-r}F(rt) \tag{9}$$
$$Z[F(t-mT)] = z^{-m}Z[F(t)] \tag{10}$$
$$Z[e^{at}F(t)] = Z[e^{-aT}z] \tag{11}$$
$$Z[e^{-at}F(t)] = Z[e^{aT}z] \tag{12}$$
$$tF(t) = -Tz\frac{d}{dz}Z[F(t)] \tag{13}$$
$$t^{-1}F(t) = -\frac{1}{T}\int_0^z \frac{f(z)}{z}\,dz. \tag{14}$$

Transforms of special functions (Beyer 1987, pp. 426–427) include

$$Z[\delta(t)] = 1 \tag{15}$$
$$Z[\delta(t-mT)] = z^{-m} \tag{16}$$
$$Z[H(t)] = \frac{z}{z-1} \tag{17}$$
$$Z[H(t-mT)] = \frac{z}{z^m(z-1)} \tag{18}$$
$$Z[t] = \frac{Tz}{(z-1)^2} \tag{19}$$
$$Z[t^2] = \frac{T^2z(z+1)}{(z-1)^3} \tag{20}$$
$$Z[t^3] = \frac{T^3z(z^2+4z+1)}{(z-1)^4} \tag{21}$$
$$Z[a^{\omega t}] = \frac{z}{z-a^{\omega T}} \tag{22}$$
$$Z[\cos(\omega t)] = \frac{z\sin(\omega T)}{z^2-2z\cos(\omega T)+1} \tag{23}$$
$$Z[\sin(\omega t)] = \frac{z[z-\cos(\omega T)]}{z^2-2z\cos(\omega T)+1}, \tag{24}$$

where $H(t)$ is the HEAVISIDE STEP FUNCTION. In general,

$$Z[t^n] = (-1)^n \lim_{x\to 0}\frac{\partial^n}{\partial x^n}\left(\frac{z}{z-e^{-xT}}\right) \tag{25}$$
$$= \frac{T^n z \sum_{k=1}^{n} \left\langle {n \atop k} \right\rangle z^{k-1}}{(z-1)^{n+1}}, \tag{26}$$

where the $\left\langle {n \atop k} \right\rangle$ are EULERIAN NUMBERS. Amazingly, the Z-transforms of t^n are therefore generators for EULER'S TRIANGLE.

see also EULER'S TRIANGLE, EULERIAN NUMBER

References

Beyer, W. H. (Ed.). *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, pp. 424–428, 1987.

Bracewell, R. *The Fourier Transform and Its Applications.* New York: McGraw-Hill, pp. 257–262, 1965.

Zag Number

An EVEN ALTERNATING PERMUTATION number, more commonly called a TANGENT NUMBER.

see also ALTERNATING PERMUTATION, TANGENT NUMBER, ZIG NUMBER

Zarankiewicz's Conjecture

The CROSSING NUMBER for a COMPLETE BIGRAPH is

$$\left\lfloor\frac{n}{2}\right\rfloor \left\lfloor\frac{n-1}{2}\right\rfloor \left\lfloor\frac{m}{2}\right\rfloor \left\lfloor\frac{m-1}{2}\right\rfloor,$$

where $\lfloor x \rfloor$ is the FLOOR FUNCTION. This has been shown to be true for all $m, n \leq 7$. Zarankiewicz has shown that, in general, the FORMULA provides an upper bound to the actual number.

see also COMPLETE BIGRAPH, CROSSING NUMBER (GRAPH)

Zariski Topology

A TOPOLOGY of an infinite set whose OPEN SETS have finite complements.

Zaslavskii Map

The 2-D map

$$x_{n+1} = [x_n + \nu(1+\mu y_n) + \epsilon\nu\mu\cos(2\pi x_n)] \pmod 1$$
$$y_{n+1} = e^{-\Gamma}[y_n + \epsilon\cos(2\pi x_n)],$$

where

$$\mu \equiv \frac{1-e^{-\Gamma}}{\Gamma}$$

(Zaslavskii 1978). It has CORRELATION EXPONENT $\nu \approx 1.5$ (Grassberger and Procaccia 1983) and CAPACITY DIMENSION 1.39 (Russell *et al.* 1980).

References

Grassberger, P. and Procaccia, I. "Measuring the Strangeness of Strange Attractors." *Physica D* **9**, 189–208, 1983.

Russell, D. A.; Hanson, J. D.; and Ott, E. "Dimension of Strange Attractors." *Phys. Rev. Let.* **45**, 1175–1178, 1980.

Zaslavskii, G. M. "The Simplest Case of a Strange Attractor." *Phys. Let.* **69A**, 145–147, 1978.

Zassenhaus-Berlekamp Algorithm

A method for factoring POLYNOMIALS.

Zeckendorf Representation

A number written as a sum of nonconsecutive FIBONACCI NUMBERS,

$$n=\sum_{k=0}^{L}\epsilon_k F_k,$$

where ϵ_k are 0 or 1 and

$$\epsilon_k\epsilon_{k+1}=0.$$

Every POSITIVE INTEGER can be written uniquely in such a form.

see also ZECKENDORF'S THEOREM

References

Grabner, P. J.; Tichy, R. F.; Nemes, I.; and Pethő, A. "On the Least Significant Digit of Zeckendorf Expansions." *Fib. Quart.* **34**, 147–151, 1996.

Vardi, I. *Computational Recreations in Mathematica.* Reading, MA: Addison-Wesley, p. 40, 1991.

Zeckendorf, E. "Répresentation des nombres naturels par une somme des nombres de Fibonacci ou de nombres de Lucas." *Bull. Soc. Roy. Sci. Liège* **41**, 179–182, 1972.

Zeckendorf's Theorem

The SEQUENCE $\{F_n-1\}$ is COMPLETE even if restricted to subsequences which contain no two consecutive terms, where F_n is a FIBONACCI NUMBER.

see also FIBONACCI DUAL THEOREM, ZECKENDORF REPRESENTATION

References

Brown, J. L. Jr. "Zeckendorf's Theorem and Some Applications." *Fib. Quart.* **2**, 163–168, 1964.

Keller, T. J. "Generalizations of Zeckendorf's Theorem." *Fib. Quart.* **10**, 95–112, 1972.

Lekkerkerker, C. G. "Voorstelling van natuurlÿke getallen door een som van Fibonacci." *Simon Stevin* **29**, 190–195, 1951–52.

Zeeman's Paradox

There is only one point in front of a PERSPECTIVE drawing where its three mutually PERPENDICULAR VANISHING POINTS appear in mutually PERPENDICULAR directions, but such a drawing nonetheless appears realistic from a variety of distances and angles.

see also LEONARDO'S PARADOX, PERSPECTIVE, VANISHING POINT

References

Dixon, R. *Mathographics.* New York: Dover, p. 82, 1991.

Zeilberger's Algorithm

An ALGORITHM which finds a POLYNOMIAL recurrence for a terminating HYPERGEOMETRIC IDENTITIES of the form

$$\sum_k \binom{n}{k}\frac{\prod_{i=1}^{A}(a_in+a_i'k+a_i'')!}{\prod_{i=1}^{B}(b_in+b_i'k+b_i'')!}z_k \\ =C\frac{\prod_{i=1}^{\bar A}(\bar a_in+\bar a_i')!}{\prod_{i=1}^{\bar B}(\bar b_in+\bar b_i')}\bar x^n,$$

where $\binom{n}{k}$ is a BINOMIAL COEFFICIENT, a_i, a_i', $\bar a_i$, b_i, b_i', $\bar b_i$ are constant integers and a_i'', $\bar a_i'$, b_i'', $\bar b_i'$, C, x, and z are complex numbers (Zeilberger 1990). The method was called CREATIVE TELESCOPING by van der Poorten (1979), and led to the development of the amazing machinery of WILF-ZEILBERGER PAIRS.

see also BINOMIAL SERIES, GOSPER'S ALGORITHM, HYPERGEOMETRIC IDENTITY, SISTER CELINE'S METHOD, WILF-ZEILBERGER PAIR

References

Krattenthaler, C. "HYP and HYPQ: The Mathematica Package HYP." `http://radon.mat.univie.ac.at/People/kratt/hyp_hypq/hyp.html`.

Paule, P. "The Paule-Schorn Implementation of Gosper's and Zeilberger's Algorithms." `http://www.risc.uni-linz.ac.at/research/combinat/risc/software/PauleSchorn/`.

Paule, P. and Riese, A. "A *Mathematica* q-Analogue of Zeilberger's Algorithm Based on an Algebraically Motivated Approach to q-Hypergeometric Telescoping." In *Special Functions, q-Series and Related Topics, Fields Institute Communications* **14**, 179–210, 1997.

Paule, P. and Schorn, M. "A Mathematica Version of Zeilberger's Algorithm for Proving Binomial Coefficient Identities." *J. Symb. Comput.* **20**, 673–698, 1995.

Petkovšek, M.; Wilf, H. S.; and Zeilberger, D. "Zeilberger's Algorithm." Ch. 6 in *A=B.* Wellesley, MA: A. K. Peters, pp. 101–119, 1996.

Riese, A. "A Generalization of Gosper's Algorithm to Bibasic Hypergeometric Summation." *Electronic J. Combinatorics* **1**, R19, 1–16, 1996. `http://www.combinatorics.org/Volume_1/volume1.html#R19`.

van der Poorten, A. "A Proof that Euler Missed... Apéry's Proof of the Irrationality of $\zeta(3)$." *Math. Intel.* **1**, 196–203, 1979.

Wegschaider, K. *Computer Generated Proofs of Binomial Multi-Sum Identities.* Diploma Thesis, RISC. Linz, Austria: J. Kepler University, May 1997. `http://www.risc.uni-linz.ac.at/research/combinat/risc/software/MultiSum/`.

Zeilberger, D. "Doron Zeilberger's Maple Packages and Programs: EKHAD." `http://www.math.temple.edu/~zeilberg/programs.html`.

Zeilberger, D. "A Fast Algorithm for Proving Terminating Hypergeometric Series Identities." *Discrete Math.* **80**, 207–211, 1990.

Zeilberger, D. "A Holonomic Systems Approach to Special Function Identities." *J. Comput. Appl. Math.* **32**, 321–368, 1990.

Zeilberger, D. "The Method of Creative Telescoping." *J. Symb. Comput.* **11**, 195–204, 1991.

Zeisel Number

A number $N = p_1p_2\cdots p_k$ (where the p_is are distinct PRIMES) such that

$$p_n = Ap_{n-1} + B,$$

with A and B constants and $p_0 \equiv 1$. For example, $1885 = 1\cdot 5\cdot 13\cdot 29$ and $114985 = 1\cdot 5\cdot 13\cdot 29\cdot 61$ are Zeisel numbers with $(A, B) = (2, 3)$.

References

Brown, K. S. "Zeisel Numbers." `http://www.seanet.com/~ksbrown/kmath015.htm`.

Zeno's Paradoxes

A set of four PARADOXES dealing with counterintuitive aspects of continuous space and time.

1. Dichotomy paradox: Before an object can travel a given distance d, it must travel a distance $d/2$. In order to travel $d/2$, it must travel $d/4$, etc. Since this sequence goes on forever, it therefore appears that the distance d cannot be traveled. The resolution of the paradox awaited CALCULUS and the proof that infinite GEOMETRIC SERIES such as $\sum_{i=1}^{\infty}(1/2)^i = 1$ can converge, so that the infinite number of "half-steps" needed is balanced by the increasingly short amount of time needed to traverse the distances.
2. Achilles and the tortoise paradox: A fleet-of-foot Achilles is unable to catch a plodding tortoise which has been given a head start, since during the time it takes Achilles to catch up to a given position, the tortoise has moved forward some distance. But this is obviously fallacious since Achilles will clearly pass the tortoise! The resolution is similar to that of the dichotomy paradox.
3. Arrow paradox: An arrow in flight has an instantaneous position at a given instant of time. At that instant, however, it is indistinguishable from a motionless arrow in the same position, so how is the motion of the arrow perceived?
4. Stade paradox: A paradox arising from the assumption that space and time can be divided only by a definite amount.

References

Pappas, T. "Zeno's Paradox—Achilles & the Tortoise." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, pp. 116–117, 1989.

Russell, B. *Our Knowledge and the External World as a Field for Scientific Method in Philosophy.* New York: Routledge, 1993.

Salmon, W. (Ed.). *Zeno's Paradoxes.* New York: Bobs-Merrill, 1970.

Stewart, I. "Objections from Elea." In *From Here to Infinity: A Guide to Today's Mathematics.* Oxford, England: Oxford University Press, p. 72, 1996.

vos Savant, M. *The World's Most Famous Math Problem.* New York: St. Martin's Press, pp. 50–55, 1993.

Zermelo's Axiom of Choice

see AXIOM OF CHOICE

Zermelo-Fraenkel Axioms

The Zermelo-Fraenkel axioms are the basis for ZERMELO-FRAENKEL SET THEORY. In the following, $\exists$ stands for EXISTS, $\in$ for "is an element of," $\forall$ for FOR ALL, $\Rightarrow$ for IMPLIES, $\neg$ for NOT (NEGATION), $\land$ for AND, $\lor$ for OR, $\rightleftharpoons$ for "is EQUIVALENT to," and $\mathcal{S}$ denotes the union y of all the sets that are the elements of x.

1. Existence of the empty set: $\exists x\forall u\neg(u\in x)$.
2. Extensionality axiom: $\forall x\forall y(\forall u(u\in x \rightleftharpoons u\in y)\to x=y)$.
3. Unordered pair axiom: $\forall x\forall y\exists z\forall u(u\in z\rightleftharpoons u=x\lor u=y)$.
4. Union (or "sum-set") axiom: $\forall x\exists y\forall u(u\in y\rightleftharpoons \exists v(u\in v\land v\in x))$.
5. Subset axiom: $\forall x\exists y\forall u(u\in y\rightleftharpoons\forall v(v\in u\to v\in x))$.
6. Replacement axiom: For any set-theoretic formula $A(u,v)$,
$$\begin{aligned}&\forall u\forall v\forall w(A(u,v)\land A(u,w)\to v=w)\\&\qquad\to\forall x\exists y\forall v(v\in y\rightleftharpoons\exists u(u\in x\land A(u,v))).\end{aligned}$$
7. Regularity axiom: For any set-theoretic formula $A(u)$, $\exists x A(x)\to\exists x(A(x)\land\neg\exists y(A(y)\land y\in x))$.
8. AXIOM OF CHOICE:
$$\begin{aligned}&\forall x[\forall u(u\in x\to\exists v(v\in u))\\&\qquad\land\forall u\forall v((u\in x\land v\in x\land\neg u=v)\\&\quad\to\neg\exists w(w\in u\land w\in v))\to\exists y\{y\subset\mathcal{S}(x)\\&\qquad\land\forall u(u\in x\to\exists z(z\in u\land z\in y\\&\qquad\qquad\land\forall w(w\in u\land w\in y\to w=z)))\}]\end{aligned}$$
9. Infinity axiom: $\exists x(\exists u(u\in x)\land\forall u(u\in x\to\exists v(v\in x\land u\subset v\land\neg v=u)))$.

If Axiom 6 is replaced by

6'. Axiom of subsets: for any set-theoretic formula $A(u)$, $\forall x\exists y\forall u(u\in y\rightleftharpoons u\in x\land A(u))$,

which can be deduced from Axiom 6, then the set theory is called ZERMELO SET THEORY instead of ZERMELO-FRAENKEL SET THEORY.

Abian (1969) proved CONSISTENCY and independence of four of the Zermelo-Fraenkel axioms.

see also ZERMELO-FRAENKEL SET THEORY

References

Abian, A. "On the Independence of Set Theoretical Axioms." *Amer. Math. Monthly* **76**, 787–790, 1969.

Iyanaga, S. and Kawada, Y. (Eds.). "Zermelo-Fraenkel Set Theory." §35B in *Encyclopedic Dictionary of Mathematics, Vol. 1.* Cambridge, MA: MIT Press, pp. 134–135, 1980.

Zermelo-Fraenkel Set Theory

A version of SET THEORY which is a formal system expressed in first-order predicate LOGIC. Zermelo-Fraenkel set theory is based on the ZERMELO-FRAENKEL AXIOMS.

see also LOGIC, SET THEORY, ZERMELO-FRAENKEL AXIOMS, ZERMELO SET THEORY

Zermelo Set Theory

The version of set theory obtained if Axiom 6 of ZERMELO-FRAENKEL SET THEORY is replaced by

6'. Axiom of subsets: for any set-theoretic formula $A(u)$, $\forall x \exists y \forall u (u \in y \rightleftharpoons u \in x \wedge A(u))$,

which can be deduced from Axiom 6.

see also ZERMELO-FRAENKEL SET THEORY

References

Iyanaga, S. and Kawada, Y. (Eds.). "Zermelo-Fraenkel Set Theory." §35B in *Encyclopedic Dictionary of Mathematics.* Cambridge, MA: MIT Press, p. 135, 1980.

Zernike Polynomial

ORTHOGONAL POLYNOMIALS which arise in the expansion of a wavefront function for optical systems with circular pupils. The ODD and EVEN Zernike polynomials are given by

$$\begin{matrix} {}^oU_n^m(\rho,\phi) \\ {}^eU_n^m(\rho,\phi) \end{matrix} = R_n^m(\rho) \begin{matrix} \sin \\ \cos \end{matrix} (m\phi) \tag{1}$$

with radial function

$$R_n^m(\rho) = \sum_{l=0}^{(n-m)/2} \frac{(-1)^l (n-l)!}{l![\frac{1}{2}(n+m)-l]![\frac{1}{2}(n-m)-l]!} \rho^{n-2l} \tag{2}$$

for n and m integers with $n \geq m \geq 0$ and $n-m$ EVEN. Otherwise,

$$R_n^m(\rho) = 0. \tag{3}$$

Here, ϕ is the azimuthal angle with $0 \leq \phi < 2\pi$ and ρ is the radial distance with $0 \leq \rho \leq 1$ (Prata and Rusch 1989). The radial functions satisfy the orthogonality relation

$$\int_0^1 R_n^m(\rho) R_{n'}^m(\rho) \rho \, d\rho = \frac{1}{2(n+1)} \delta_{nn'}, \tag{4}$$

where δ_{ij} is the KRONECKER DELTA, and are related to the BESSEL FUNCTION OF THE FIRST KIND by

$$\int_0^1 R_n^m(\rho) J_m(v\rho) \rho \, d\rho = (-1)^{(n-m)/2} \frac{J_{n+1}(v)}{v} \tag{5}$$

(Born and Wolf 1989, p. 466). The radial Zernike polynomials have the GENERATING FUNCTION

$$\frac{[1+z-\sqrt{1-2z(1-2\rho^2)+z^2}\,]^m}{(2z\rho)^m \sqrt{1-2z(1-2\rho^2)+z^2}} = \sum_{s=0}^{\infty} z^s R_{m+2s}^{\pm m}(\rho), \tag{6}$$

and are normalized so that

$$R_n^{\pm m}(1) = 1 \tag{7}$$

(Born and Wolf 1989, p. 465). The first few NONZERO radial polynomials are

$$\begin{aligned}
R_0^0(\rho) &= 1 \\
R_1^1(\rho) &= \rho \\
R_2^0(\rho) &= 2\rho^2 - 1 \\
R_2^2(\rho) &= \rho^2 \\
R_3^1(\rho) &= 3\rho^3 - 2\rho \\
R_3^3(\rho) &= \rho^3 \\
R_4^0(\rho) &= 6\rho^4 - 6\rho^2 + 1 \\
R_4^2(\rho) &= 4\rho^4 - 3\rho^2 \\
R_4^4(\rho) &= \rho^4
\end{aligned}$$

(Born and Wolf 1989, p. 465).

The Zernike polynomial is a special case of the JACOBI POLYNOMIAL with

$$P_{n'}^{(\alpha,\beta)}(x) = (-1)^{n'} \frac{R_n^m(\rho)}{\rho^\alpha} \tag{8}$$

and

$$x = 1 - 2\rho^2 \tag{9}$$

$$\beta = 0 \tag{10}$$

$$\alpha = m \tag{11}$$

$$n' = \tfrac{1}{2}(n-m). \tag{12}$$

The Zernike polynomials also satisfy the RECURRENCE RELATIONS

$$\rho R_n^m(\rho) = \frac{1}{2(n+1)}[(n+m+2) R_{n+1}^{m+1}(\rho) + (n-m) R_{n-1}^{m+1}(\rho)] \tag{13}$$

$$R_{n+2}^m(\rho) = \frac{n+2}{(n+2)^2 - m^2} \left\{ \left[4(n+1)\rho^2 - \frac{(n+m)^2}{n} - \frac{(n-m+2)^2}{n+2} \right] R_n^m(\rho) - \frac{n^2 - m^2}{n} R_{n-2}^m(\rho) \right. \tag{14}$$

$$R_n^m(\rho) + R_n^{m+2}(\rho) = \frac{1}{n+1} \frac{d[R_{n+1}^{m+1}(\rho) - R_{n-1}^{m+1}(\rho)]}{d\rho} \tag{15}$$

(Prata and Rusch 1989). The coefficients A_n^m and B_n^m in the expansion of an arbitrary radial function $F(\rho,\phi)$ in terms of Zernike polynomials

$$F(\rho,\phi) = \sum_{m=0}^{\infty} \sum_{n=m}^{\infty} [A_n^m \, {}^oU_n^m(\rho,\phi) + B_n^m \, {}^eU_n^m(\rho,\phi)] \tag{16}$$

are given by

$$\begin{matrix} A_n^m \\ B_n^m \end{matrix} = \frac{(n+1)}{\epsilon_{mn}{}^2\pi} \int_0^1 \int_0^{2\pi} F(\rho,\phi) \begin{matrix} {}^oU_n^m(\rho,\phi) \\ {}^eU_n^m(\rho,\phi) \end{matrix} \rho\, d\phi\, d\rho, \tag{17}$$

where

$$\epsilon_{mn} \equiv \begin{cases} \epsilon \equiv \frac{1}{\sqrt{2}} & \text{for } m=0,\ n \neq 0 \\ 1 & \text{otherwise} \end{cases} \tag{18}$$

Let a "primary" aberration be given by

$$\Phi = a'_{lmn} Y_1^{2l+m^*}(\theta,\phi)\rho^n \cos^m\theta \tag{19}$$

with $2l+m+n=4$ and where Y^* is the COMPLEX CONJUGATE of Y, and define

$$A'_{lmn} = a'_{lmn} Y_1^{2l+m^*}(\theta,\phi), \tag{20}$$

giving

$$\Phi = \frac{1}{\epsilon_{nm}{}^2} A_{lmn} R_n^m(\rho)\cos(m\theta). \tag{21}$$

Then the types of primary aberrations are given in the following table (Born and Wolf 1989, p. 470).

Aberration	l	m	n	A	A'
spherical aberration	0	4	0	$A'_{040}\rho^4$	$\epsilon A_{040} R_4^0(\rho)$
coma	0	3	1	$A'_{031}\rho^3\cos\theta$	$A_{031} R_3^1(\rho)\cos\theta$
astigmatism	0	2	2	$A'_{022}\rho^2\cos^2\theta$	$A_{022} R_2^2(\rho)\cos(2\theta)$
field curvature	1	2	0	$A'_{120}\rho^2$	$\epsilon A_{120} R_2^0(\rho)$
distortion	1	1	1	$A'_{111}\rho\cos\theta$	$A_{111} R_1^1(\rho)\cos\theta$

see also JACOBI POLYNOMIAL

References

Bezdidko, S. N. "The Use of Zernike Polynomials in Optics." *Sov. J. Opt. Techn.* **41**, 425, 1974.

Bhatia, A. B. and Wolf, E. "On the Circle Polynomials of Zernike and Related Orthogonal Sets." *Proc. Cambridge Phil. Soc.* **50**, 40, 1954.

Born, M. and Wolf, E. "The Diffraction Theory of Aberrations." Ch. 9 in *Principles of Optics: Electromagnetic Theory of Propagation, Interference, and Diffraction of Light, 6th ed.* New York: Pergamon Press, pp. 459–490, 1989.

Mahajan, V. N. "Zernike Circle Polynomials and Optical Aberrations of Systems with Circular Pupils." In *Engineering and Lab. Notes* **17** (Ed. R. R. Shannon), p. S-21, Aug. 1994.

Prata, A. and Rusch, W. V. T. "Algorithm for Computation of Zernike Polynomials Expansion Coefficients." *Appl. Opt.* **28**, 749–754, 1989.

Wang, J. Y. and Silva, D. E. "Wave-Front Interpretation with Zernike Polynomials." *Appl. Opt.* **19**, 1510–1518, 1980.

Zernike, F. "Beugungstheorie des Schneidenverfahrens und seiner verbesserten Form, der Phasenkontrastmethode." *Physica* **1**, 689–704, 1934.

Zhang, S. and Shannon, R. R. "Catalog of Spot Diagrams." Ch. 4 in *Applied Optics and Optical Engineering, Vol. 11.* New York: Academic Press, p. 201, 1992.

Zero

The INTEGER denoted 0 which, when used as a counting number, means that no objects are present. It is the only INTEGER (and, in fact, the only REAL NUMBER) which is neither NEGATIVE nor POSITIVE. A number which is not zero is said to be NONZERO.

Because the number of PERMUTATIONS of 0 elements is 1, 0! (zero FACTORIAL) is often defined as 1. This definition is useful in expressing many mathematical identities in simple form. A number *other than 0* taken to the POWER 0 is defined to be 1. 0^0 is undefined, but defining $0^0 = 1$ allows concise statement of the beautiful analytical formula for the integral of the generalized SINC FUNCTION

$$\int_0^\infty \frac{\sin^a x}{x^b}\,dx = \frac{\pi^{1-c}(-1)^{\lfloor (a-b)/2 \rfloor}}{2^{a-c}(b-1)!} \times \sum_{k=0}^{\lfloor a/2 \rfloor - c} (-1)^k \binom{a}{k} (a-2k)^{b-1}[\ln(a-2k)]^c$$

given by Kogan, where $a \geq b > c$, $c \equiv a - b \pmod 2$, and $\lfloor x \rfloor$ is the FLOOR FUNCTION.

The following table gives the first few numbers n such that n^k contains no zeros, for small k. The largest known n for which 2^n contain no zeros is 86 (Madachy 1979), with no other $n \leq 4.6 \times 10^7$ (M. Cook), improving the 3.0739×10^7 limit obtained by Beeler *et al.* (1972). The values $a(n)$ such that the positions of the right-most zero in $2^{a(n)}$ increases are 10, 20, 30, 40, 46, 68, 93, 95, 129, 176, 229, 700, 1757, 1958, 7931, 57356, 269518, ... (Sloane's A031140). The positions in which the right-most zeros occur are 2, 5, 8, 11, 12, 13, 14, 23, 36, 38, 54, 57, 59, 93, 115, 119, 120, 121, 136, 138, 164, ... (Sloane's A031141). The right-most zero of $2^{781,717,865}$ occurs at the 217th decimal place, the farthest over for powers up to 2.5×10^9.

k	Sloane	n such that n^k contains no 0s
2	007377	1, 2, 3, 4, 5, 6, 7, 8, 9, 13, 14, 15, 16, ...
3	030700	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, ...
4	030701	1, 2, 3, 4, 7, 8, 9, 12, 14, 16, 17, 18, ...
5	008839	1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 17, 18, 30, ...
6	030702	1, 2, 3, 4, 5, 6, 7, 8, 12, 17, 24, 29, 44, ...
7	030703	1, 2, 3, 6, 7, 10, 11, 19, 35, ...
8	030704	1, 2, 3, 5, 6, 8, 9, 11, 12, 13, 17, 24, 27, ...
9	030705	1, 2, 3, 4, 6, 7, 12, 13, 14, 17, 34, ...
11	030706	1, 2, 3, 4, 6, 7, 8, 9, 12, 13, 14, 15, 16, ...

While it has not been proven that the numbers listed above are the only ones without zeros for a given base, the probability that any additional ones exist is vanishingly small. Under this assumption, the sequence of largest n such that k^n contains no zeros for $k = 2$, 3, ... is then given by 86, 68, 43, 58, 44, 35, 27, 34, 0, 41, ... (Sloane's A020665).

see also 10, NAUGHT, NEGATIVE, NONNEGATIVE, NONZERO, ONE, POSITIVE, TWO

References

Beeler, M.; Gosper, R. W.; and Schroeppel, R. *HAKMEM.* Cambridge, MA: MIT Artificial Intelligence Laboratory, Memo AIM-239, Item 57, Feb. 1972.

Kogan, S. "A Note on Definite Integrals Involving Trigonometric Functions." `http://www.mathsoft.com/asolve/constant/pi/sin/sin.html`.

Madachy, J. S. *Madachy's Mathematical Recreations.* New York: Dover, pp. 127–128, 1979.

Pappas, T. "Zero–Where & When." *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 162, 1989.

Sloane, N. J. A. Sequence A007377/M0485 in "An On-Line Version of the Encyclopedia of Integer Sequences."

Zero Divisor

A NONZERO element x of a RING for which $x \cdot y = 0$, where y is some other NONZERO element and the vector multiplication $x \cdot y$ is assumed to be BILINEAR. A RING with no zero divisors is known as an INTEGRAL DOMAIN. Let A denote an $\mathbb{R}$-algebra, so that A is a VECTOR SPACE over R and

$$A \times A \to A$$

$$(x, y) \mapsto x \cdot y.$$

Now define

$$Z \equiv \{x \in A : x \cdot y = 0 \text{ for some NONZERO } y \in A\},$$

where $0 \in Z$. A is said to be m-ASSOCIATIVE if there exists an m-dimensional SUBSPACE S of A such that $(y \cdot x) \cdot z = y \cdot (x \cdot z)$ for all $y, z \in A$ and $x \in S$. A is said to be TAME if Z is a finite union of SUBSPACES of A.

References

Finch, S. "Zero Structures in Real Algebras." `http://www.mathsoft.com/asolve/zerodiv/zerodiv.html`.

Zero (Root)

see ROOT

Zero-Sum Game

A GAME in which players make payments only to each other. One player's loss is the other player's gain, so the total amount of "money" available remains constant.

see also FINITE GAME, GAME

References

Dresher, M. *The Mathematics of Games of Strategy: Theory and Applications.* New York: Dover, p. 2, 1981.

Zeta Fuchsian

A class of functions discovered by Poincaré which are related to the AUTOMORPHIC FUNCTIONS.

see also AUTOMORPHIC FUNCTION

Zeta Function

A function satisfying certain properties which is computed as an INFINITE SUM of NEGATIVE POWERS. The most commonly encountered zeta function is the RIEMANN ZETA FUNCTION,

$$\zeta(n) \equiv \sum_{k=1}^{\infty} \frac{1}{k^n}.$$

see also DEDEKIND FUNCTION, DIRICHLET BETA FUNCTION, DIRICHLET ETA FUNCTION, DIRICHLET L-SERIES, DIRICHLET LAMBDA FUNCTION, EPSTEIN ZETA FUNCTION, JACOBI ZETA FUNCTION, NINT ZETA FUNCTION, PRIME ZETA FUNCTION, RIEMANN ZETA FUNCTION

References

Ireland, K. and Rosen, M. "The Zeta Function." Ch. 11 in *A Classical Introduction to Modern Number Theory, 2nd ed.* New York: Springer-Verlag, pp. 151–171, 1990.

Zeuthen's Rule

On an ALGEBRAIC CURVE, the sum of the number of coincidences at a noncuspidal point C is the sum of the orders of the infinitesimal distances from a nearby point P to the corresponding points when the distance PC is taken as the principal infinitesimal.

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 131, 1959.

Zeuthen's Theorem

If there is a (ν, ν') correspondence between two curves of GENUS p and p' and the number of BRANCH POINTS properly counted are β and β', then

$$\beta + 2\nu'(p-1) = \beta' + 2\nu(p'-1).$$

see also CHASLES-CAYLEY-BRILL FORMULA

References

Coolidge, J. L. *A Treatise on Algebraic Plane Curves.* New York: Dover, p. 246, 1959.

Zig Number

An ODD ALTERNATING PERMUTATION number, more commonly called an EULER NUMBER or SECANT NUMBER.

see also ALTERNATING PERMUTATION, EULER NUMBER, ZAG NUMBER

Zig-Zag Triangle

see also SEIDEL-ENTRINGER-ARNOLD TRIANGLE

Zigzag Permutation

see ALTERNATING PERMUTATION

Zillion

A generic word for a very LARGE NUMBER. The term has no well-defined mathematical meaning. Conway and Guy (1996) define the nth zillion as 10^{3n+3} in the American system (million $= 10^6$, billion $= 10^9$, trillion $= 10^{12}$, ...) and 10^{6n} in the British system (million $= 10^6$, billion $= 10^{12}$, trillion $= 10^{18}$, ...). Conway and Guy (1996) also define the words n-PLEX and n-MINEX for 10^n and 10^{-n}, respectively.

see also LARGE NUMBER

References

Conway, J. H. and Guy, R. K. *The Book of Numbers.* New York: Springer-Verlag, pp. 13–16, 1996.

Zipf's Law

In the English language, the probability of encountering the rth most common word is given roughly by $P(r) = 0.1/r$ for r up to 1000 or so. The law breaks down for less frequent words, since the HARMONIC SERIES diverges. Pierce's (1980, p. 87) statement that $\sum P(r) > 1$ for $r = 8727$ is incorrect. Goetz states the law as follows: The frequency of a word is inversely proportional to its RANK r such that

$$P(r) \approx \frac{1}{r\ln(1.78R)},$$

where R is the number of different words.

see also HARMONIC SERIES, RANK (STATISTICS)

References

Goetz, P. "Phil's Good Enough Complexity Dictionary." `http://www.cs.buffalo.edu/~goetz/dict.html`.

Pierce, J. R. *Introduction to Information Theory: Symbols, Signals, and Noise, 2nd rev. ed.* New York: Dover, pp. 86–87 and 238–239, 1980.

Zollner's Illusion

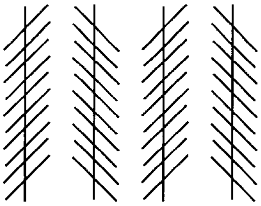

In this ILLUSION, the VERTICAL lines in the above figure are PARALLEL, but appear to be tilted at an angle.

see also ILLUSION

References

Jablan, S. "Some Visual Illusions Occurring in Interrupted Systems." `http://members.tripod.com/~modularity/interr.htm`.

Pappas, T. *The Joy of Mathematics.* San Carlos, CA: Wide World Publ./Tetra, p. 172, 1989.

Zonal Harmonic

A SPHERICAL HARMONIC which is a product of factors linear in x^2, y^2, and z^2, with the product multiplied by z when n is ODD.

see also TESSERAL HARMONIC

Zone

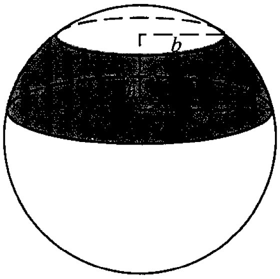

The SURFACE AREA of a SPHERICAL SEGMENT. Call the RADIUS of the SPHERE R, the upper and lower RADII b and a, respectively, and the height of the SPHERICAL SEGMENT h. The zone is a SURFACE OF REVOLUTION about the z-AXIS, so the SURFACE AREA is given by

$$S = 2\pi \int x\sqrt{1+x'^2}\,dz. \tag{1}$$

In the xz-plane, the equation of the zone is simply that of a CIRCLE,

$$x = \sqrt{R^2 - z^2}, \tag{2}$$

so

$$x' = -z(R^2 - z^2)^{-1/2} \tag{3}$$

$$x'^2 = \frac{z^2}{R^2 - z^2}, \tag{4}$$

and

$$\begin{aligned} S &= 2\pi \int_{\sqrt{R^2-a^2}}^{\sqrt{R^2-b^2}} \sqrt{R^2 - z^2}\sqrt{1 + \frac{z^2}{R^2 - z^2}}\,dz \\ &= 2\pi R \int_{\sqrt{R^2-a^2}}^{\sqrt{R^2-b^2}} dz = 2\pi R(\sqrt{R^2 - b^2} - \sqrt{R^2 - a^2}\,) \\ &= 2\pi Rh. \end{aligned} \tag{5}$$

This result is somewhat surprising since it depends only on the *height* of the zone, not its vertical position with respect to the SPHERE.

see also SPHERE, SPHERICAL CAP, SPHERICAL SEGMENT, ZONOHEDRON

References

Beyer, W. H. (Ed.). *CRC Standard Mathematical Tables, 28th ed.* Boca Raton, FL: CRC Press, p. 130, 1987.

Zonohedron

A convex POLYHEDRON whose faces are PARALLEL-sided $2m$-gons. There exist $n(n-1)$ PARALLELOGRAMS in a nonsingular zonohedron, where n is the number of different directions in which EDGES occur (Ball and Coxeter

1987, pp. 141–144). Zonohedra include the CUBE, ENNEACONTAHEDRON, GREAT RHOMBIC TRIACONTAHEDRON, MEDIAL RHOMBIC TRIACONTAHEDRON, RHOMBIC DODECAHEDRON, RHOMBIC ICOSAHEDRON, RHOMBIC TRIACONTAHEDRON, RHOMBOHEDRON, and TRUNCATED CUBOCTAHEDRON, as well as the entire class of PARALLELEPIPEDS.

Regular zonohedra have bands of PARALLELOGRAMS which form equators and are called "ZONES." Every convex polyhedron bounded solely by PARALLELOGRAMS is a zonohedron (Coxeter 1973, p. 27). Plate II (following p. 32 of Coxeter 1973) illustrates some equilateral zonohedra. Equilateral zonohedra can be regarded as 3-dimensional projections of n-D HYPERCUBES (Ball and Coxeter 1987).

see also HYPERCUBE

References

Ball, W. W. R. and Coxeter, H. S. M. *Mathematical Recreations and Essays, 13th ed.* New York: Dover, pp. 141–144, 1987.

Coxeter, H. S. M. "Zonohedra." §2.8 in *Regular Polytopes, 3rd ed.* New York: Dover, pp. 27–30, 1973.

Coxeter, H. S. M. Ch. 4 in *Twelve Geometric Essays.* Carbondale, IL: Southern Illinois University Press, 1968.

Eppstein, D. "Ukrainian Easter Egg." `http://www.ics.uci.edu/~eppstein/junkyard/ukraine.`

Fedorov, E. S. *Zeitschr. Krystallographie und Mineralogie* **21**, 689, 1893.

Fedorov, E.W. *Nachala Ucheniya o Figurakh.* Leningrad, 1953.

Hart, G. W. "Zonohedra." `http://www.li.net/~george/virtual-polyhedra/zonohedra-info.html.`

Zonotype

The MINKOWSKI SUM of line segments.

Zorn's Lemma

If S is any nonempty PARTIALLY ORDERED SET in which every CHAIN has an upper bound, then S has a maximal element. This statement is equivalent to the AXIOM OF CHOICE.

see also AXIOM OF CHOICE

Zsigmondy Theorem

If $1 \leq b < a$ and $(a, b) = 1$ (i.e., a and b are RELATIVELY PRIME), then $a^n - b^n$ has a PRIMITIVE PRIME FACTOR with the following two possible exceptions:

1. $2^6 - 1^6$.
2. $n = 2$ and $a + b$ is a POWER of 2.

Similarly, if $a > b \geq 1$, then $a^n + b^n$ has a PRIMITIVE PRIME FACTOR with the exception $2^3 + 1^3 = 9$.

References

Ribenboim, P. *The Little Book of Big Primes.* New York: Springer-Verlag, p. 27, 1991.

digital electronics

DeMYSTiFieD

A SELF-TEACHING GUIDE

- DIGITAL LOGIC and interfacing made EASY
- Loaded with circuit design "PROBLEMS" to work out
- EXPERIMENTS you can DO for less than $10
- LOTS of quizzes and FINAL EXAMS to map your understanding and progress

0010
001
10010
1
0
1

Mc Graw Hill

Myke Predko

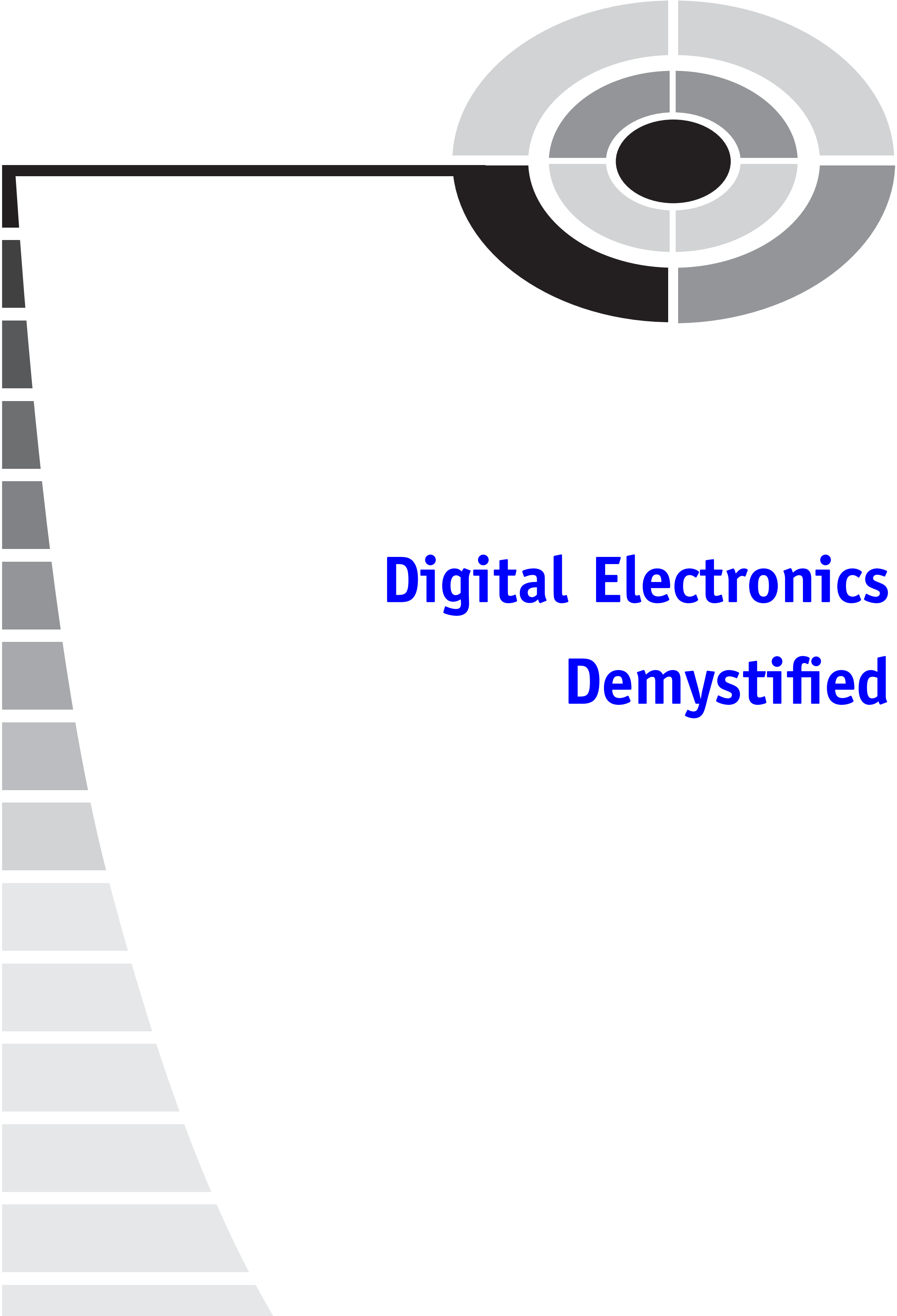

Digital Electronics Demystified

Demystified Series

Advanced Statistics Demystified
Algebra Demystified
Anatomy Demystified
Astronomy Demystified
Biology Demystified
Business Statistics Demystified
Calculus Demystified
Chemistry Demystified
College Algebra Demystified
Differential Equations Demystified
Earth Science Demystified
Electronics Demystified
Everyday Math Demystified
Geometry Demystified
Math Word Problems Demystified
Physics Demystified
Physiology Demystified
Pre-Algebra Demystified
Pre-Calculus Demystified
Project Management Demystified
Robotics Demystified
Statistics Demystified
Trigonometry Demystified

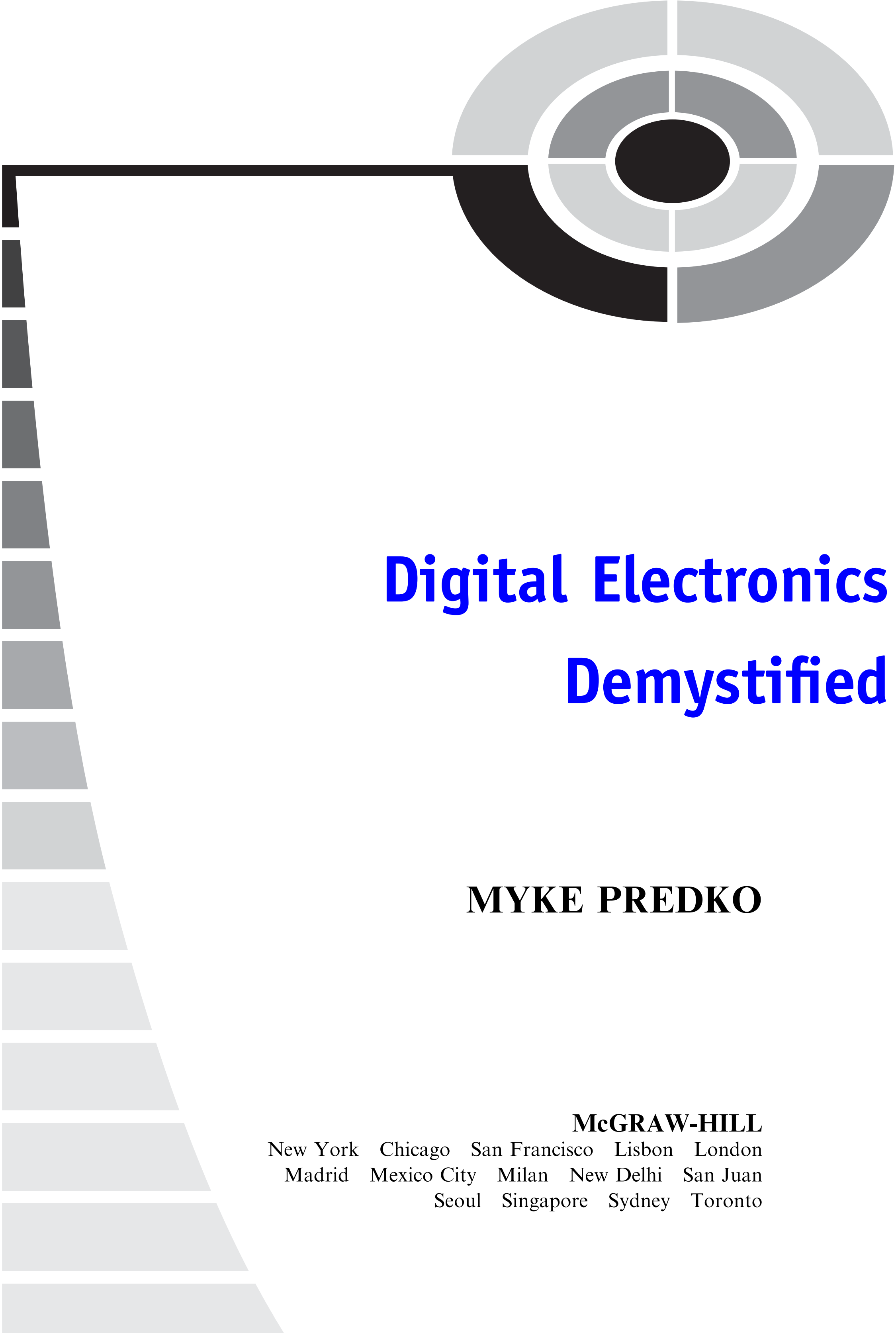

Digital Electronics Demystified

MYKE PREDKO

McGRAW-HILL
New York Chicago San Francisco Lisbon London
Madrid Mexico City Milan New Delhi San Juan
Seoul Singapore Sydney Toronto

The McGraw·Hill Companies

Copyright © 2005 by The McGraw-Hill Companies, Inc. All rights reserved. Manufactured in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

0-07-147124-3

The material in this eBook also appears in the print version of this title: 0-07-144141-7.

All trademarks are trademarks of their respective owners. Rather than put a trademark symbol after every occurrence of a trademarked name, we use names in an editorial fashion only, and to the benefit of the trademark owner, with no intention of infringement of the trademark. Where such designations appear in this book, they have been printed with initial caps.

McGraw-Hill eBooks are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. For more information, please contact George Hoare, Special Sales, at george_hoare@mcgraw-hill.com or (212) 904-4069.

TERMS OF USE

This is a copyrighted work and The McGraw-Hill Companies, Inc. ("McGraw-Hill") and its licensors reserve all rights in and to the work. Use of this work is subject to these terms. Except as permitted under the Copyright Act of 1976 and the right to store and retrieve one copy of the work, you may not decompile, disassemble, reverse engineer, reproduce, modify, create derivative works based upon, transmit, distribute, disseminate, sell, publish or sublicense the work or any part of it without McGraw-Hill's prior consent. You may use the work for your own noncommercial and personal use; any other use of the work is strictly prohibited. Your right to use the work may be terminated if you fail to comply with these terms.

THE WORK IS PROVIDED "AS IS." McGRAW-HILL AND ITS LICENSORS MAKE NO GUARANTEES OR WARRANTIES AS TO THE ACCURACY, ADEQUACY OR COMPLETENESS OF OR RESULTS TO BE OBTAINED FROM USING THE WORK, INCLUDING ANY INFORMATION THAT CAN BE ACCESSED THROUGH THE WORK VIA HYPERLINK OR OTHERWISE, AND EXPRESSLY DISCLAIM ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. McGraw-Hill and its licensors do not warrant or guarantee that the functions contained in the work will meet your requirements or that its operation will be uninterrupted or error free. Neither McGraw-Hill nor its licensors shall be liable to you or anyone else for any inaccuracy, error or omission, regardless of cause, in the work or for any damages resulting therefrom. McGraw-Hill has no responsibility for the content of any information accessed through the work. Under no circumstances shall McGraw-Hill and/or its licensors be liable for any indirect, incidental, special, punitive, consequential or similar damages that result from the use of or inability to use the work, even if any of them has been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause arises in contract, tort or otherwise.

DOI: 10.1036/0071441417

Professional

Want to learn more?

We hope you enjoy this McGraw-Hill eBook! If you'd like more information about this book, its author, or related books and websites, please click here.

For more information about this title, click here

CONTENTS

Preface ix

Acknowledgment xiii

PART ONE **Introduction to Digital Electronics** 1

CHAPTER 1 **The Underpinnings of Digital Electronics** 3
- Boolean Arithmetic, Truth Tables and Gates 6
- The Six Elementary Logic Operations 10
- Combinatorial Logic Circuits: Combining Logic Gates 14
- Sum of Products and Product of Sums 17
- Waveform Diagrams 21
- Quiz 25

CHAPTER 2 **Effectively Optimizing Combinatorial Circuits** 29
- Truth Table Function Reduction 32
- Karnaugh Maps 36
- Boolean Arithmetic Laws 40
- Optimizing for Technology 45
- Quiz 50

CHAPTER 3 **Creating Digital Electronic Circuits** **53**
Basic Electronic Laws 57
Capacitors 64
Semiconductor Operation 69
Logic Gate Input and Output 78
Simple Digital Logic Circuit Development 87
Testing a Simple TTL Inverter 91
Quiz 97

CHAPTER 4 **Number Systems** **100**
Base 16 or Hexadecimal Numbers 104
Binary Coded Decimal 107
Gray Codes 111
Quiz 115

CHAPTER 5 **Binary Arithmetic Using Digital Electronics** **118**
Adders 122
Subtraction and Negative Numbers 129
Magnitude Comparators and Bus Nomenclature 135
Multiplication and Division 138
Quiz 144

CHAPTER 6 **Practical Combinatorial Circuit Implementation** **146**
Race Conditions and Timing Analysis 150
Quick and Dirty Logic Gates 153
Dotted AND and Tri-State Logic Drivers 156
Combining Functions on a Net 159
Quiz 161

CHAPTER 7 **Feedback and Sequential Circuits** **164**
Flip Flops (RS and JK) 168
Edge Triggered Flip Flops 173
Latches Versus Registers 178
Reset 180
Quiz 184

Test for Part One **186**

PART TWO **Digital Electronics Applications** **197**

CHAPTER 8 **Oscillators** **199**
Transistor Astable Oscillators 202
Ring Oscillators 204
Relaxation Oscillators 206
Crystals and Ceramic Resonators 208
555 Timer Chip 211
Delay Circuits 219
Quiz 223

CHAPTER 9 **Complex Sequential Circuits** **225**
Counters 228
Shift Registers 231
Linear Feedback Shift Registers 234
Hardware State Machines 237
Quiz 240

CHAPTER 10 **Circuit Interfaces** **243**
Address and Data Decoders 247
Multi-Segment LEDs 250
Pulse Width Modulation 253
Button ''Debouncing'' 258
Switch Matrix Keypad Interfacing 262
Quiz 265

CHAPTER 11 **Reading Datasheets** **268**
Chip Operating Characteristics 270
IEEE Logic Symbols 272
Power Usage and Fanouts 274
Quiz 275

CHAPTER 12 **Computer Processors and Support** **277**
IEEE754 Floating Point Numbers 281
Memory Types 283
Power Supplies 287
Programmable Logic Devices 298
Quiz 301

CHAPTER 13 **PC Interfacing Basics** **303**
The Parallel (Printer) Port 306
Video Output 309
Synchronous Serial Interfaces (SPI, I2C, Microwire) 314
Asynchronous Serial Interfaces 317
RS-232 Electrical Standards 322
Quiz 325

Test for Part Two **327**

Final Exam **336**

APPENDIX **Answers to Quiz, Tests, and Final Exam** **357**

Index **361**

PREFACE

Philosophy is sometimes described as the study of what people take for granted. It examines the reasons why people make assumptions about things in their lives by understanding the relationships between the basic ‘‘truths’’ that are used to come up with these assumptions. This analysis takes a very precise logical path that is scientific in nature. For example, the following statement can be broken down into a set of simple truths and the relationships between them plotted out and understood to allow philosophers to carry on the natural thought process (such as what is a body that has three ‘‘extensions’’ with a ‘‘thinking substance’’).

> Thus, extension in length, breadth and depth, constitutes the nature of corporeal substance; and thought constitutes the nature of thinking substance. For all else that may be attributed to body presupposes extension, and is but a mode of this extended thing; as everything that we find in mind is but so many diverse forms of thinking.
>
> Descartes

Surprisingly enough, the rules that were developed for understanding philosophic statements like the one above were applied in the 1930s and 1940s to help define how electrical circuits could be designed that would be used in the first electronic computers. One of the elements of the success of this effort was to reduce the electronic logic ‘‘truths’’ into two simple electrical states.

These two electrical states are often represented as two numbers that can be manipulated using ‘‘binary arithmetic.’’ Binary arithmetic was formally described by the English mathematician George Boole in the middle of the 19th century and is often referred to as ‘‘Boolean arithmetic’’ or ‘‘Boolean algebra’’ as a way to perform mathematical operations on numbers that only have two values (‘‘0’’ or ‘‘1’’). These two values are manipulated within electronic computers and other devices built from ‘‘digital electronics.’’

Over the past 60-plus years, digital logic circuits, processing binary signals have been miniaturized, sped up and integrated together to create the fantastic electronic gadgets that we take for granted. Despite their

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

complexity, they operate using the basic rules and circuits that are explained in this book. After working through this book, not only will you understand how these products are designed but you will also have some experience in designing and working through the problems of implementing them on your own.

This book was written for people that would like to learn about digital electronics without taking a formal course. After working through this book, along with a reasonably good understanding of the subject as well as some of the background material needed to create electronic circuits, it can also serve as a supplemental text in a classroom, tutored or home-schooling environment. The book should also be useful for career changers who need to refresh their knowledge in electronics and would like to better understand what are the different facets of current digital electronic products.

This introductory work contains an abundance of practice quiz, test and exam questions. They are all multiple-choice and are similar to the sorts of questions used in standardized texts. There is a short quiz at the end of every chapter. The quizzes are ‘‘open-book.’’ You may (and should) refer to the chapter texts when taking them. When you think you’re ready, take the quiz, write down the answers and then give your list of answers to a friend. Have the friend tell you the score, but not which questions you got wrong. The answers are listed in the back of the book. Stick with a chapter until you get most of the answers correct.

This book is divided into two parts. At the end of each part is a multiple-choice test. Take these tests when you’ve completed with the respective sections and have taken all the chapter quizzes. The section tests are ‘‘closed-book’’, but the questions are not as difficult as those in the quizzes. A satisfactory score is three-quarters of the answers correct. Again, answers are in the back of the book.

There is a final exam at the end of this course. It contains questions drawn uniformly from all the chapters in the book. Take it when you have finished both sections, both section tests and all of the chapter quizzes. A satisfactory score is at least 75% correct answers.

With the section tests and the final exam, as with the quizzes, have a friend tell you your score without letting you know which questions you missed. That way, you will not subconsciously memorize the answers. You can check to see where your knowledge is strong and where it is not.

I recommend that you complete one chapter a week. An hour or two daily ought to be enough time for this. As part of this work, you should notice that I have given a number of suggestions on how you could implement the described circuits to see exactly how they work. When you’ve worked through this material, you can use this book as a permanent reference.

Now, work hard, but be sure to have fun and look to see where you can use the information provided here to help you to understand how the complex electronic devices of modern society are implemented using digital logic devices that are just capable of following simple rules of logic.

myke

This page intentionally left blank

ACKNOWLEDGMENT

I would like to thank my wife, Patience, for her love and support and willingness to become the first person to have worked through the material in this book. Without her support, suggestions, love, and willingness to understand what ‘‘fanout’’ means, this book and its material would never have been possible.

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

This page intentionally left blank

PART ONE

Introduction to Digital Electronics

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

This page intentionally left blank

CHAPTER 1

The Underpinnings of Digital Electronics

If you were asked to define what a bit is, chances are you would probably do a pretty good job, saying something like:

> A bit is something that can only have two values: on or off.

Instead of "on or off", you might have used terms for two values like "one or zero", "high or low voltage", "up or down", "empty or full" or (if you fancy yourself as being erudite) "dominant or recessive". All of these terms are correct and imply that the two values are at opposite extremes and are easily differentiated.

When you think of "bits", you are probably thinking of something in a wire or an electronic device contained within a computer, but when the concept of binary (two state) logic was first invented, the values that were

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

applied were tests to see if a statement was ‘‘true’’ or ‘‘false’’. Examples of true and false statements are:

- The sun always rises in the East. (true)
- Dogs live at the bottom of the ocean like fish. (false)

Looking at these simple statements determining if they are true or false seems to reduce the information within to an extreme degree. The truthfulness of a statement can be combined with another statement to help determine if a more complex postulate is true. If you consider the following ‘‘true’’ statements:

- A dog has fur over its body.
- A dog has four legs.
- Animals have four legs and fur.
- Humans have two legs.
- A snake has scales on its body.
- A reptile’s body has scales or smooth skin.

and combining them together, you can make some surprisingly complex ‘‘assertions’’ from these data using three basic operations. These three basic operations consist of ‘‘AND’’ which is true if all the statements combined together by the AND are true, ‘‘OR’’ which is true if any of the combined statements are true and ‘‘NOT’’ which is true if a single statement is false. To differentiate these three operations from their prose synonyms, I will capitalize them (as well as other basic logic operations) throughout the book. These operations are often called ‘‘logic operations’’ because they were first used to understand the logic of complex philosophical statements.

From the seven statements above and using these three basic operations, you can make the following true assertions:

- Humans are not dogs.
- A dog is an animal.
- A snake is a reptile.

The first statement is true because we know that a human has two legs (violating a condition that is required for the definition of a dog to be true). This is an example of the ‘‘negation’’ or ‘‘NOT’’ operation; the assertion is true if the single input statement is false:

> The room is dark because the lights are not on.

The NOT function is often called an ‘‘Inverter’’ because it changes the value of the input from high to low and vice versa.

The second assertion, “A dog is an animal”, is true because both of the two statements concerning animals are true when applied to dogs (which have four legs and fur). This is an example of the “AND” operation; the assertion is true if and only if the input statements are true. The AND operation has two or more input statements. In trying to differentiate bicycles and motorcycles from cars, you might make the assertion which uses the AND operation:

> A car has four wheels and a motor.

The last assertion, “A snake is a reptile”, is true because one of the two statements giving the necessary characteristics for a reptile is true. This is an example of an “inclusive or” (usually referred to as just “OR”) operation; the assertion is true if any of the input statements are true. Like the “and” operation, OR can have two or more input statements. If you’re a parent, you will be familiar with the assertion:

> During the course of a day, a baby eats, sleeps, cries or poops.

I use this example to illustrate an important point about the “OR” operation that is often lost when it is used in colloquial speech: if more than one input statement is true, the entire assertion is still true. As incredible as it sounds to someone who has not had children yet, a baby is very capable of performing all four actions listed above simultaneously (and seemingly constantly).

I’m making this point because when we speak, we usually use the “exclusive or” instead of “inclusive or” to indicate that only one of two actions can be true. An example statement in which an “exclusive or” is used in everyday speech could be:

> Tom is at a restaurant or the movies.

This is an example of “exclusive OR” because Tom can only be at one of the two places at any given time. I will discuss the “exclusive or” operation in more detail later in this chapter, but for now try to remember that an assertion using the “OR” operation will be true if one or more of the input statements are true.

So far I have been working with “bits” of “binary” information contained in “statements” and “assertions”. You are probably wondering why a term like “bit electronics” or “binary electronics” is used instead of “digital electronics”. “Digital” comes from the Latin word for “fingers” and indicates that there are many discrete signals that are at one of two values. Naming the circuitry “bit electronics” or “binary electronics” would imply that it can only work with one piece of information; digital electronic circuits can process many bits of information simultaneously, either as separate pieces of information or collections of large amounts of data.

In the first few pages of this book, I have introduced you to the concept of the ‘‘bit’’, the ‘‘digit’’, the ‘‘NOT’’, ‘‘AND’’ and ‘‘OR’’ operations along with the ‘‘exclusive OR’’. Different combinations of these concepts are the basis for the majority of the material presented through the remainder of this book and any course in digital electronics. I suggest that you read over this chapter and make sure you are completely comfortable with the terms and how they work before going on.

Boolean Arithmetic, Truth Tables and Gates

In the introduction to this chapter, I demonstrated the operation of the three operations ‘‘AND’’, ‘‘OR’’ and ‘‘NOT’’, which can be used to test input values (which are in the form of two state ‘‘bits’’) and produce assertions based on the state of the input bits. The verbose method I used could be used with digital electronics, but you will find that it is cumbersome and not intuitively obvious when you are working with electronic circuits. Fortunately, a number of different tools have been developed to simplify working with logic operations.

The first tool that simplifies how logic operations are expressed is known as ‘‘Boolean arithmetic’’ (or sometimes as ‘‘Boolean logic’’), a branch of mathematics where a mathematical expression is used to express how bit inputs can be transformed into an output using the three operations presented in the introduction. Boolean arithmetic was first described by the English mathematician Charles Lutwidge Dodgson, whom you may be familiar with by his nom de plume Lewis Carroll, and expanded upon by George Boole, in the mid 19th century, as a way of understanding, proving or disproving complex philosophical statements. Boole demonstrated that a statement, involving bits of data and the AND, OR or NOT operations could be written in the form:

```
Result = Data1 operation Data2  {operation Data3...}
```

The braces (‘‘{‘‘ and ’’}’’) are often used to indicate that what’s inside them is optional and the three periods (‘‘. . .’’) indicate that the previous text can be repeated. Using these conventions you can see that a Boolean arithmetic statement is not limited to just one operation with two input bits – they can actually be very lengthy and complex with many bit inputs and multiple operations.

To demonstrate how a Boolean arithmetic statement could be articulated, I can write the proof that a dog is an animal in the form:

`Result = (Does Dog have 4 Legs)   AND   (Does Dog have Fur)`

If both statements within the parentheses are true, then the “Result” will be true.

This method of writing out assertions and the logic behind them is quite a bit simpler and much easier to understand, but we can do better. Instead of writing out the true or false statement as a condition, it can be expressed in terms of a simple “variable” (like “X”). So, if we assign “A” as the result of testing if dogs have four legs and “B” as the result of testing if dogs have fur, we can write out the Boolean arithmetic equation above as:

`Result = A AND B`

To further simplify how a logic operation is written out, the basic characters “.”, “+” and “!” can be used instead of AND, OR and NOT, respectively. AND behaves like a binary multiplication, so along with the “·” character, you may see an “x” or “*”. The OR operation may be represented as “|”. The ampersand (“&”) for AND and “|” for OR are often used because they are the same symbols as are used in most computer programming languages. When I write out Boolean arithmetic equations throughout the book, I will be using the “.”, “+” and “!” characters for the three basic logic operations instead of the full words.

An important advantage of converting a statement into a simple equation is that it more clearly shows how the logic operation works. If the variables “A” and “B” were just given the values of “true” or “false”, the “Result” of the equation above could be written out in the form shown in Table 1-1. This is known as a “truth table” and it is a very effective way of expressing how a Boolean operator works. The truth table is not limited to just three inputs,

Table 1-1 “AND” operation truth table using Gray code inputs.

Input “A”	Input “B”	“AND” Output
False	False	False
False	True	False
True	True	True
True	False	False

and a function with more than one Boolean operator can be modeled in this way. Functions with more than one output can be expressed using the truth table, but I don't recommend doing this because relationships between inputs and outputs (which I will discuss in greater detail later in the book) can be obscured.

One other thing to notice about the truth table is that I have expressed the inputs as a "Gray code", rather than incrementing inputs. Gray codes are a technique for sequencing multiple bits in such a way that only one bit changes from one state to the next. Incrementing inputs behave as if the inputs were bits of a larger binary number and the value of this number is increased by one when moving from one state to the next. The truth table above, for the "AND" gate could be written out using incrementing inputs as Table 1-2.

In many cases, truth tables are introduced with incrementing inputs, but I would like to discourage this. Incrementing inputs can obscure relationships between inputs that become obvious when you use Gray codes. This advantage will become more obvious as you work through more complex logic operations and are looking for ways to simplify the expression.

The OR operation's truth table is given in Table 1-3, while the NOT operation's truth table is shown in Table 1-4.

The OR operation would be written in Boolean arithmetic, using the "+" character to represent the OR operation as:

$$\texttt{Output} = \texttt{A} + \texttt{B}$$

and the NOT operation (using the "!" character) is written out in Boolean arithmetic as:

$$\texttt{Output} = \texttt{!A}$$

Table 1-2 "AND" operation truth table using incrementing inputs.

Input "A"	Input "B"	"AND" Output
False	False	False
False	True	False
True	False	False
True	True	True

Table 1-3 "OR" operation truth table using Gray code inputs.

Input "A"	Input "B"	"OR" Output
False	False	False
False	True	True
True	True	True
True	False	True

Table 1-4 "NOT" operation truth table using Gray code inputs.

Input	"NOT" Output
False	True
True	False

Sometimes, when a signal is NOTted, its symbol is given either a minus sign ("–") or an underscore ("_") as its first character to indicate that it has been inverted by a NOT operation.

The final way of expressing the three basic logic operations is graphically with the inputs flowing through lines into a symbol representing each operation and the output flowing out of the line. Figures 1-1 through 1-3 show the graphical representations of the AND, OR and NOT gates, respectively.

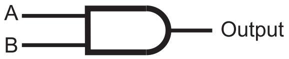

Fig. 1-1. "AND" gate.

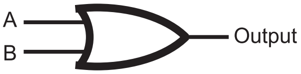

Fig. 1-2. "OR" gate.

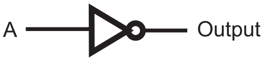

Fig. 1-3. "NOT" gate.

The graphical representation of the logic operations is a very effective way of describing and documenting complex functions and is the most popular way of representing logic operations in digital electronic circuits. When graphics are used to represent the logic operations, they are most often referred to as "gates", because the TRUE is held back until its requirements are met, at which point it is allowed out by opening the gate. "Gate" is the term I will use most often when describing Boolean arithmetic operations in this book.

If you were to replace the lines leading to each gate with a wire and the symbol with an electronic circuit, you can transfer a Boolean arithmetic design to a digital electronic circuit directly.

The Six Elementary Logic Operations

When you look at a catalog of digital electronics chips, you are going to discover that they are built from ANDs, ORs and NOTs as well as three other elementary gates. Two of these gates are critically important to understand because they are actually the basis of digital logic while the third is required for adding numbers together.

TTL logic is based on the "NAND" gate which can be considered a "NOTted AND" – the output of an AND gate is passed through a NOT gate as shown in Fig. 1-4. Instead of drawing the NAND gate as an AND gate and NOT gate connected together as in Fig. 1-4, they are contracted into the one symbol shown in Fig. 1-5. It's truth table is in Table 1-5.

When writing out the NAND function in Boolean arithmetic, it is normally in the form:

$$\texttt{Output} = !(\texttt{A} \cdot \texttt{B})$$

which is a literal translation of the operation – the inputs are ANDed together and the result is NOTted before it is passed to the Output.

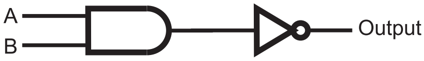

Fig. 1-4. "NAND" gate made from "AND" and "OR" gates.

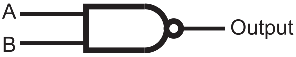

Fig. 1-5. "NAND" gate.

Table 1-5 “NAND” operation truth table.

Input “A”	Input “B”	“NAND” Output
False	False	True
False	True	True
True	True	False
True	False	True

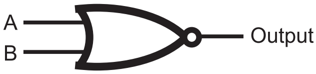

Fig. 1-6 “NOR” gate.

You will see the small circuit on various parts in different electronic devices, both on inputs and outputs. The small circle on the NAND gate is the conventional shorthand symbol to indicate that the input or output of a gate is NOTted.

In case you didn’t note the point above, the NAND gate is the basis for TTL logic, as I will explain later in the book. Being very comfortable with NAND gates is very important to being able to design and use TTL electronics. This is a point that I find is not stressed enough in most electronics courses and by having a strong knowledge of how NAND gates work as well as how they are implemented you will better understand what is happening within your digital electronics circuits.

If you are going to be working with CMOS logic, in the same way you should be comfortable with the NAND gate for TTL, you should be familiar with the “NOR” gate (Fig. 1-6). The NOR gate can be considered to be a contraction of the OR and NOT gates (as evidenced by the circle on the output of the OR gate) and operates in the opposite manner as the OR gate, as shown in Table 1-6. When using NOR operations in Boolean arithmetic, a similar format to the NAND gate is used:

$$\texttt{Output} = !\,(\texttt{A} + \texttt{B})$$

The last elementary logic gate that you will have to work with is the “Exclusive OR” (Fig. 1-7) with Table 1-7 being its truth table. The

Table 1-6 "NOR" operation truth table.

Input "A"	**Input "B"**	**"NAND" Output**
False	False	True
False	True	False
True	True	False
True	False	False

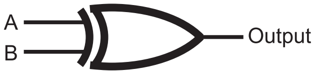

Fig. 1-7. "XOR" gate.

Table 1-7 "Exclusive OR" operation truth table.

Input "A"	**Input "B"**	**"Exclusive OR" Output**
False	False	False
False	True	True
True	True	False
True	False	True

Exclusive OR (also referred to as "Ex-OR" or "XOR") only returns a True output if only one of its inputs is true. If both inputs are the same, then the Exclusive OR outputs False. The Boolean arithmetic symbol for Exclusive OR is often a caret ("^") as is used in computer programming languages or a circle with an "x" character in it $\otimes$. Writing a Boolean statement with the Exclusive OR would be in the format:

`Output = A ^ B`

Table 1-8 summarizes the six elementary gates along with their Boolean arithmetic symbols and formats, graphical symbols and truth tables.

Table 1-8 Summary of the six elementary logic operations.

Gate	Boolean arithmetic symbols	Boolean arithmetic equation	Graphic symbol	Truth table
AND	·, &, *, x	Out = A · B	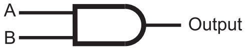	A B | Out −− −− + −− F F | F F T | F T T | T T F | F
OR	+, |	Out = A + B	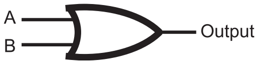	A B | Out −− −− + −− F F | F F T | T T T | T T F | T
NOT	!, _, -	Out = ! A	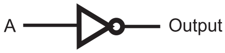	A | Out −− + −− F | T T | F
NAND	! ·	Out = ! (A · B)	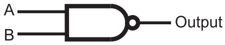	A B | Out −− −− + −− F F | T F T | T T T | F T F | T
NOR	! +	Out = ! (A + B)	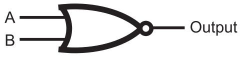	A B | Out −− −− + −− F F | T F T | F T T | F T F | F
Exclusive OR	ˆ, ⊙	Out = A ˆ B	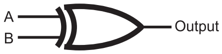	A B | Out −− −− + −− F F | F F T | T T T | F T F | T

Combinatorial Logic Circuits: Combining Logic Gates

As I hinted at in the previous section, multiple gate functions can be combined to form more complex or different Boolean logic functions. Wiring together multiple gates are used to build a complex logic function that only outputs a specific value when a specific combination of True and False inputs are passed to it is known as "combinatorial logic". The output of a combinatorial logic circuit is dependent on its input; if the input changes then the output will change as well.

When I wrote the preceding paragraph, I originally noted that combinatorial logic circuits produce a "True" output for a given set of inputs. This is incorrect, as there will be some cases where you will require a False output in your application. I made the definition a bit more ambiguous so that you do not feel like the output has to be a single, specific value when the input consists of the required inputs. It is also important to note that in a combinatorial logic circuit, data flows in one direction and outputs in logic gates cannot be used as inputs to gates which output back to themselves. These two points may seem subtle now, but they are actually critically important to the definition of combinatorial logic circuits and using them in applications.

An example of a combinatorial circuit is shown in Fig. 1-8. In this circuit, I have combined three AND gates, a NOR gate, a NOT gate and an XOR gate to produce the following logic function:

$$\texttt{Output} = ((\texttt{A} \cdot \texttt{B}) \cdot !(!(\texttt{A} + \texttt{C}))) \mathbin{\hat{}} (!(\texttt{A} + \texttt{C}) \cdot \texttt{B})$$

This combinatorial circuit follows the convention that inputs to a gate (or a chip or other electronic component) are passed into the left and outputs

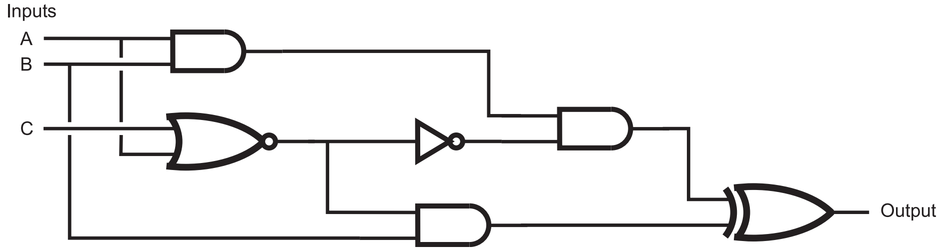

Fig. 1-8 Combinatorial circuit built from multiple logic gates.

exit from the right. This will help you "read" the circuit from left to right, something that should be familiar to you.

While seeing a series of logic gates, like the one in Fig. 1-8, seems to be overwhelming, you already have the tools to be able to work through it and understand how it works. In the previous section, I noted that gates could be connected by passing the output of one into an input of another; a combinatorial circuit (like Fig. 1-8) is simply an extension of this concept and, being an extension, you can use the same tools you used to understand single gates to understand the multiple gate operation.

I should point out that the two broken lines on the left side of Fig. 1-8 (leading down from "A" and "B") indicate that these lines are not connected to the lines that they intersect with. You will find that it can be very difficult to design logic circuits without connected and separate lines from becoming confused. In Fig. 1-9, I have shown a couple of the conventional ways of drawing intersecting lines, depending on whether or not they connect or bypass each other. Probably the most intuitively obvious way of drawing connecting and bypassing lines is to use the dot and arc, respectively. I tend not to because they add extra time to the logic (and circuit) diagram drawing process. As you see more circuit diagrams, you will see the different conventions used and you should be comfortable with recognizing what each means.

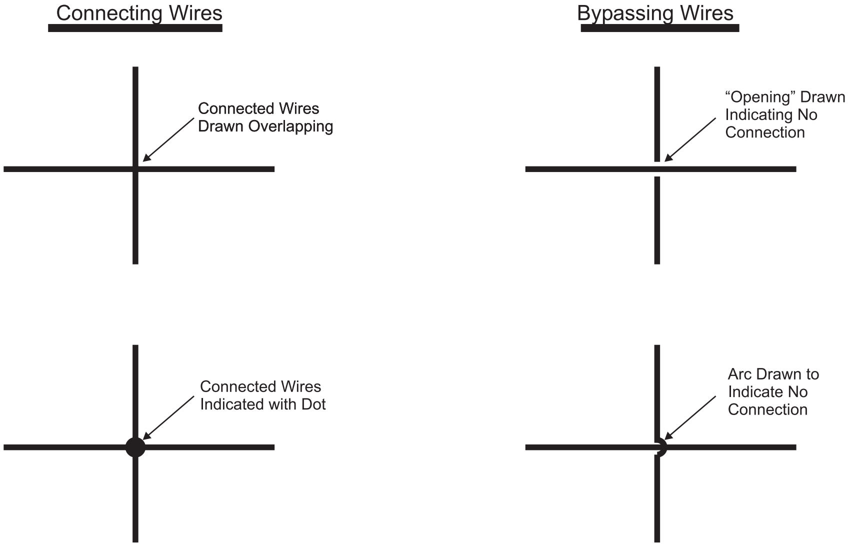
Connecting Wires
Bypassing Wires
Connected Wires Drawn Overlapping
"Opening" Drawn Indicating No Connection
Connected Wires Indicated with Dot
Arc Drawn to Indicate No Connection

Fig. 1-9. Different representations for wires that connect or bypass.

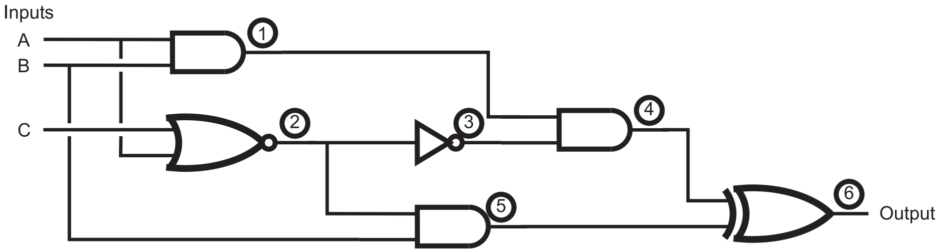

Fig. 1-10. Combinatorial circuit with logic gate outputs marked.

When I am faced with a complex combinatorial circuit, the first thing I do is to mark the outputs of each of the gates (Fig. 1-10) and then list them according to their immediate inputs:

$$\texttt{Output 1} = \texttt{A} \cdot \texttt{B}$$

$$\texttt{Output 2} = !(\texttt{A} + \texttt{C})$$

$$\texttt{Output 3} = !\texttt{2}$$

$$\texttt{Output 4} = \texttt{1} \cdot \texttt{3}$$

$$\texttt{Output 5} = \texttt{B} \cdot \texttt{2}$$

$$\texttt{Output 6} = \texttt{4}\,\hat{}\,\texttt{5}$$

After listing them, I then work through a truth table, passing the outputs of each gate along until I have the final outputs of the complete function (Table 1-9). In keeping with my comments of the previous section, I have used a three bit Gray code for the inputs to this circuit.

Before going on, there are two points that I would like you to keep in the back of your mind. First, this is actually quite an efficient methodology for decoding combinatorial circuits that you are given the design for. Designing a logic gate circuit that responds in a specific manner is actually quite a different process and I will be devoting the rest of this chapter as well as the next to explaining the design and optimization of combinatorial circuits. Once you have finished with Chapter 2, you might want to revisit the example circuit in Fig. 1-8 and see how effectively you can reduce its complexity and the number of logic gates needed to implement it.

The second point that you should be aware of is the example circuit that I used in this section is actually quite unwieldy and does not conform to the typical methods used to design most practical combinatorial digital electronic circuits. In the next section, I will present you with the conventional methods for specifying and documenting combinatorial circuits.

Table 1-9 Decoding the response of the combinatorial circuit in Fig. 1-8.

Inputs			1 = A · B	2 = !(A + C)	3 = !2	4 = 1 · 3	5 = B · 2	6 = 4 ˆ 5
A	B	C						
F	F	F	False	True	False	False	False	False
F	F	T	False	False	True	False	False	False
F	T	T	False	False	True	False	False	False
F	T	F	False	True	False	False	True	True
T	T	F	True	False	True	True	False	True
T	T	T	True	False	True	True	False	True
T	F	T	False	False	True	False	False	False
T	F	F	False	False	True	False	False	False

Sum of Products and Product of Sums

Presenting combinatorial circuits as a collection of gates wired together almost randomly, like the circuit shown in Fig. 1-8, is sub-optimal from a variety of perspectives. The first is, the function provided by the combinatorial circuit is not obvious. Secondly, using a variety of different gates can make your parts planning process difficult, with only one gate out of many available in a chip being used. Lastly, the arrangement of gates will be difficult for automated tools to combine on a printed circuit board (“PCB”) or within a logic chip. What is needed is a conventional way of drawing combinatorial logic circuits.

The most widely used format is known as “sum of products”. Earlier in the chapter, I presented the concept that the AND operation was analogous to multiplication just as the OR operation is to addition. Using this background, you can assume that a “sum of products” circuit is built from AND and OR gates. Going further, you might also guess that the final output is the “OR” (because addition produces a “sum”) with the gates that

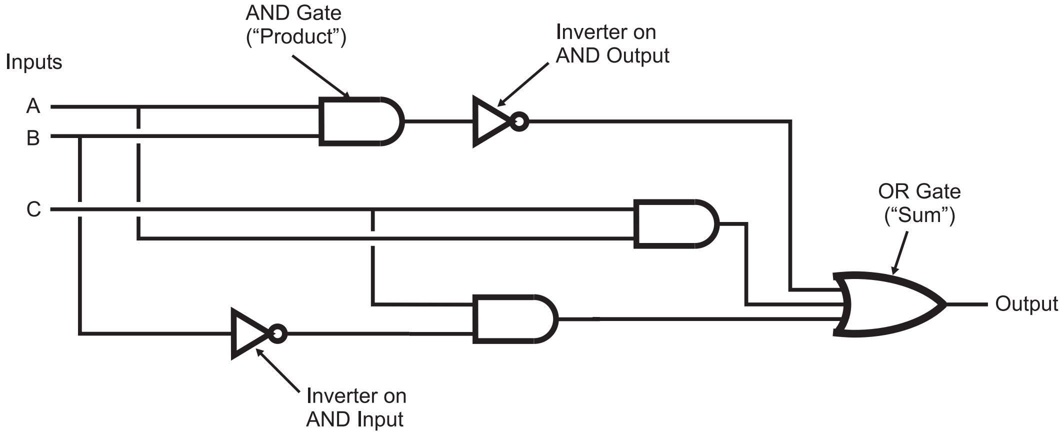

Fig. 1-11. Example "sum of products" combinatorial logic circuit.

convert the inputs being "AND" gates (a "product" is the result of a multiplication operation). An example "sum of products" combinatorial logic circuit is shown in Fig. 1-11.

In this circuit, the inputs are ANDed together and the result is passed to an OR gate. In this circuit, the output will be "True" if any of the inputs to the OR gate (which are the outputs from the AND gates) are true. In some cases, to make sure that the inputs and outputs of the AND gates are in the correct state, they will be inverted using NOT gates, as I have shown in Fig. 1-11.

Figure 1-11 has one feature that I have not introduced to you yet and that is the three input OR gate on the right side of the diagram. So far, I have only discussed two input gates, but I should point out that three input gates can be built from multiple two input gates, as I show in Fig. 1-12, in which two, two input AND gates are combined to form a single three input AND gate. A three input OR gate could be built exactly the same way.

A three input NAND or NOR gate is a bit trickier, as Fig. 1-13 shows. For this case, the output of the NAND gate processing "A" and "B" must be inverted (which can be accomplished with a NAND gate and both inputs tied together as I show in Fig. 1-13) to make its output the same as an "AND". The NAND gate's function is to first AND its inputs together and then invert them before driving out the output signal. As I will explain in greater detail in the next chapter, an inverted output, when it is inverted, becomes a "positive" output and I use this rule to produce the three input NAND gate. A three input NOR gate would be built in exactly the same way as a three input NAND gate.

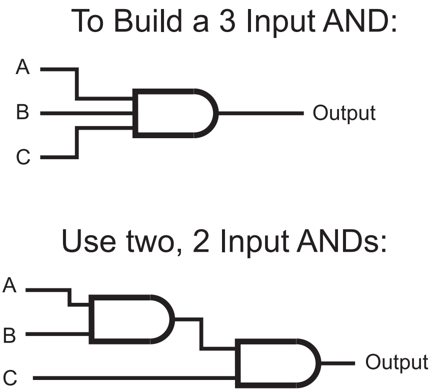
To Build a 3 Input AND:
A
B
C
Output
Use two, 2 Input ANDs:
A
B
C
Output

Fig. 1-12. 3 Input AND gate built from 2 input AND gates.

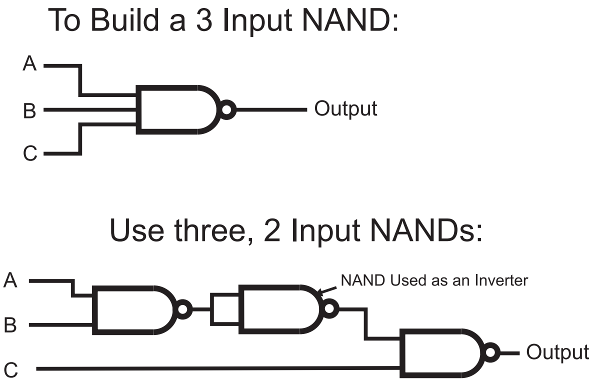
To Build a 3 Input NAND:
A
B
C
Output
Use three, 2 Input NANDs:
A
B
C
NAND Used as an Inverter
Output

Fig. 1-13. 3 Input AND gate built from 2 input AND gates.

Along with having a “sum of products” combinatorial logic circuit that outputs a True when one of the intermediate AND gates outputs True, there is the complementary “product of sums” (Fig. 1-14), which outputs False when one of its intermediate OR gates outputs False.

While product of sums combinatorial circuits can produce the same functions as sum of product combinatorial circuits, you will not see as many product of sum combinatorial circuits in various designs because they rely on what I call “negative logic”. Most people cannot easily visualize something happening because the inputs do not meet an expected case, which is exactly what happens in a product of sums combinatorial logic circuit.

To demonstrate how a sum of product combinatorial logic circuit is designed, consider the messy combinatorial logic circuit I presented

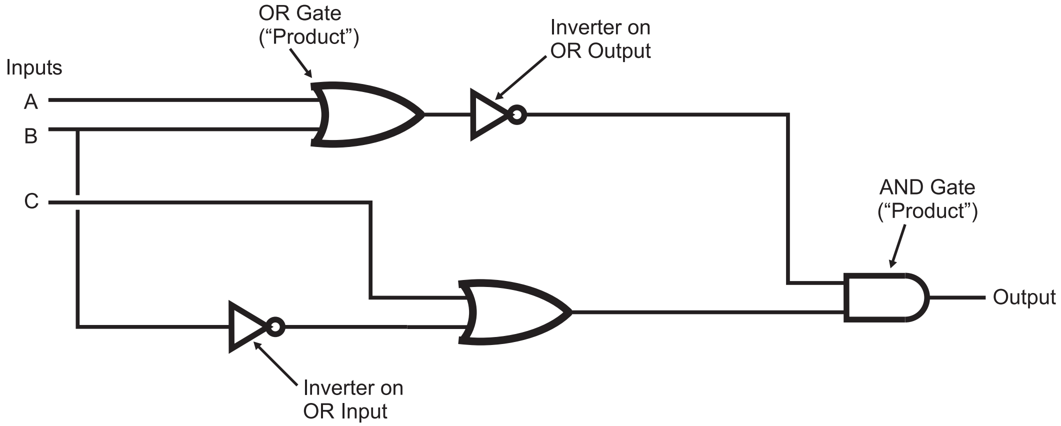

Fig. 1-14. Example "product of sums" combinatorial logic circuit.

in the previous section (see Fig. 1-8). To understand the operation of this circuit, I created a large truth table (Table 1-9) and listed the outputs of each of the intermediate gates and finally discovered that the function output True in three cases that can be directly translated into AND operations by assuming that in each case the output was true and the input conditions were all true. To make all three inputs True to the AND gates when the input is False, I invert them and came up with the three statements below:

```
A . B · !C
A · B · C
!A · B · !C
```

These three AND statements can be placed directly into a sum of products combinatorial circuit, as shown in Fig. 1-15.

Looking at Fig. 1-15, you'll probably notice that this circuit has the same total number of gates as the original circuit – and, assuming that each three input gate is made up of two, two input AND gates, it probably requires four *more* gates than the original circuit shown in Fig. 1-8. The only apparent advantage to the sum of product format for combinatorial logic circuit is that it is easier to follow through and see that the output is True for the three input conditions I listed above.

In the following chapters, I will introduce you to combinatorial logic circuit optimization as well as explain in more detail how digital electronic gates are actually built. It will probably be surprising to discover that the sum

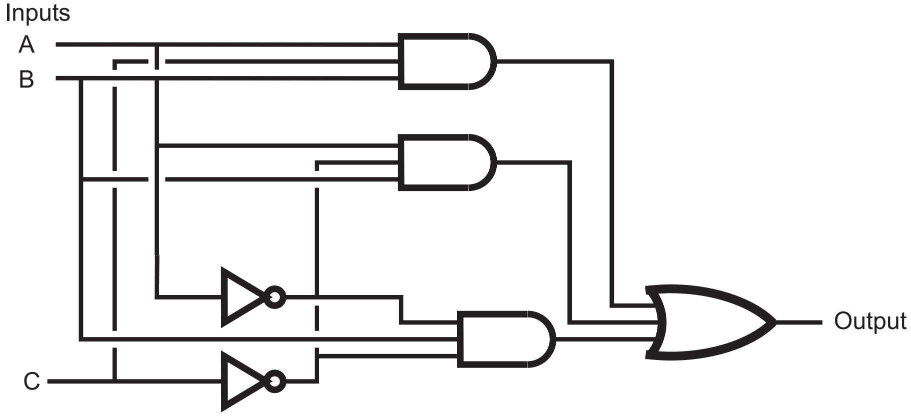

Fig. 1-15. Original combinatorial circuit built in "sum of products" format.

of product combinatorial logic circuit format leads to applications that are more efficient (in terms of total gate or transistor count along with speed and heat dissipation) than ones using less conventional design methodologies.

Waveform Diagrams

So far in this chapter, I have shown how logic functions can be presented as English text, mathematical equations (Boolean arithmetic), truth tables and graphical circuit diagrams. There are actually two more ways in which the logic data can be presented that you should be familiar with. The first method is not one that you will see a lot of except when debugging microprocessor instructions from a very low level, while the second is one that you will have to become very familiar with, especially when the digital electronic signals pass from the combinatorial logic shown here to more complex circuits that have the ability to "store" information.

The first method, the "state list" consists of a list of text columns for each state of the circuit. The state list is really a compressed form of the truth table and is best suited for displaying a fairly large amount of numerical data. Going back to the example circuit of Fig. 1-8, and Table 1-9, I could express the truth table as the series of columns below. Note that I have used the numeric values "1" for True and "0" for False because they are easier to differentiate than "T" and "F" over a number of rows.

```
ABC123456   < — Signals being displayed
000010000
001001000
011001000
010010010
110101101
111101101
101001000
100001000
```

As I said, not a lot of information is obvious from the state list. Some formatting could be done to make the inputs and outputs better differentiated, but for the most part, I don't recommend using state lists for most digital electronics applications. Where the state list is useful is in debugging state machine or microcontroller applications in which you have added hardware to the data, address and control busses to record how the device responds to specific inputs.

The state list is not ideal for this type of application, but it's better than nothing.

The other method, which is not only recommended as a circuit analysis and design tool but is also one you should be intimately familiar with is the "waveform diagram". Waveforms are electrical signals that have been plotted over time. The original waveform display tool was the oscilloscope; a drawing of a typical oscilloscope's display is shown in Fig. 1-16.

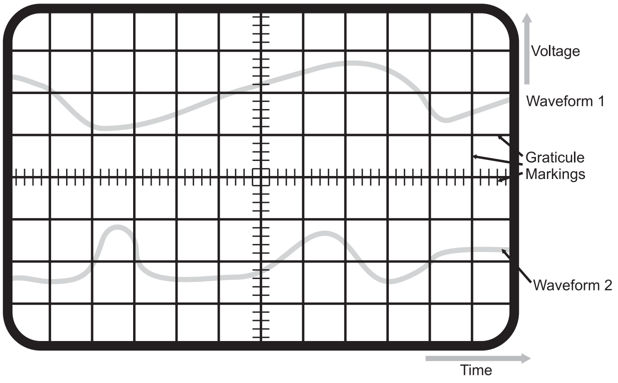

Fig. 1-16. Basics of an oscilloscope screen.

The features of the two ‘‘waveforms’’ displayed on the oscilloscope screen can be measured by placing them against the ‘‘graticule markings’’ on the display. These markings (usually just referred to as ‘‘graticules’’ and etched onto the glass screen of the oscilloscope) are indicators of a specific passage of time or change in voltage. Along with the ‘‘gross’’ graticules, most oscilloscopes have finer markings, to allow more accurate measurements by moving the waveforms over them.

Oscilloscopes are very useful tools for a variety of different applications, which contain varying voltage levels (which are known as ‘‘analog’’ voltage levels). They can be (and often are) used for digital logic applications but they are often not the best tool because digital waveforms only have two levels, when applied to electronics: digital signals are either a high voltage or a low voltage. The timing of the changes of these two voltage levels is more interesting to the designer.

So instead of thinking of digital waveforms in terms of voltage over time, in digital electronics, we prefer to think of them as states (High/Low, True/False, 1/0) over time and display them using a waveform diagram like the one shown in Fig. 1-17. When designing your digital electronics circuit, you will create a waveform diagram to help you understand how the logic states will be passed through the circuit; later, when you are debugging the circuit, you will be comparing what you actually see with this diagram to see if your assumptions about how the circuit would operate are correct. The different signals shown in Fig. 1-17 are samples of what you will see when you are designing your own application circuit.

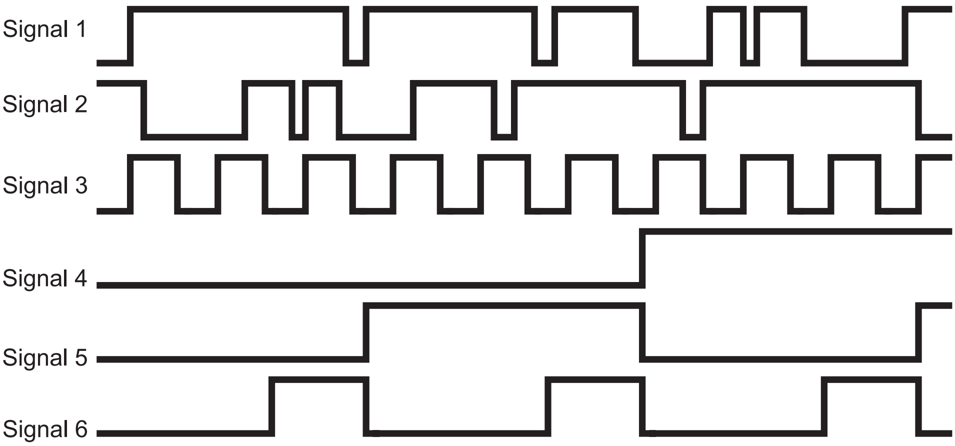

Fig. 1-17. Digital waveforms.

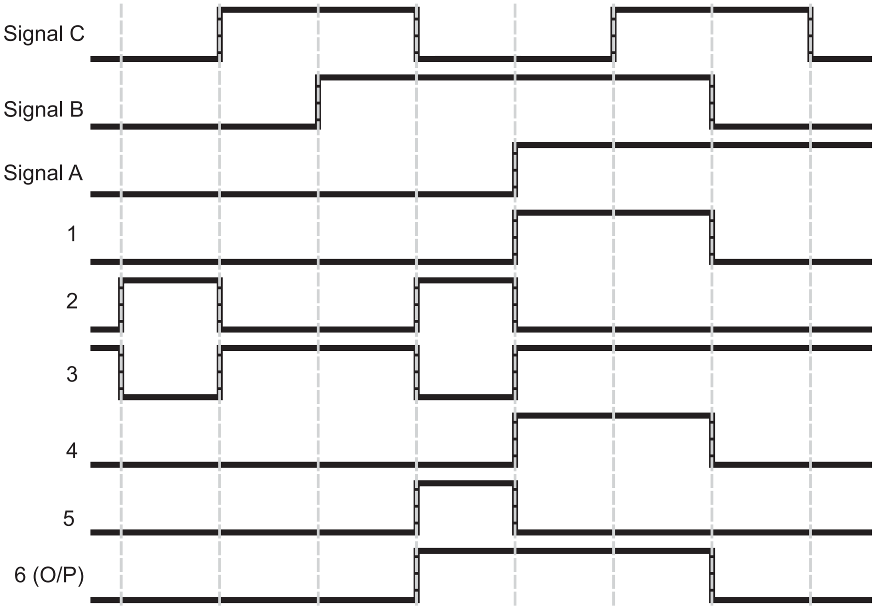

Fig. 1-18. Digital waveforms of Fig. 1-8 example combinatorial logic circuit.

The waveform diagram is the first tool that will help you optimize your circuit. Before writing up this section, I was planning on the diagrams I wanted to include with it and one was a waveform representation of the first example combinatorial logic circuit's operation from Table 1-9. The thin vertical lines indicate the edges of each state.

After drawing out Fig. 1-18, it was obvious that signals "1" and "4" (from the marked circuit diagram Fig. 1-8) were redundant. Looking back at the diagram for the circuit, I realized that the AND gate with output 4 and inverter with output 3 could be completely eliminated from the circuit – the output of AND gate 1 could be passed directly to the XOR gate (with output 6).

The waveform diagram shown in Fig. 1-18 is what I call an "idealized waveform diagram" and does not encompass what is actually happening in a physical circuit. Looking at Fig. 1-18, you will see that I have assumed that the switching time for the gates is instantaneous. In real components, switching time is finite, measurable and can have a significant impact to your application's ability to work. This is discussed in more detail in later chapters. Finally, this circuit does not allow for basic errors in understanding, such as what happens when multiple gate outputs are passed to a single gate input – your assumption of this type of case may not match what really happens in an actual circuit.

In this chapter, I have introduced you to the basic concepts of combinatorial logic circuits and the parts that make them up. In the next chapter, I will start working through some of the practical aspects of designing efficient digital electronic circuits.

Quiz

1. Which of the following statements is true?
   (a) Negative logic is the same as reverse psychology. You get somebody to do something by telling them to do what you don't want them to do
   (b) Using the logic definition, "A dog has four legs and fur", a cat could be accurately described as a dog
   (c) "High" and "Higher" are valid logic states
   (d) Assertions are the same as logic operations
2. Boolean arithmetic is a:
   (a) way to express logic statements in a traditional mathematical equation format
   (b) terrible fraud perpetrated by philosophers to disprove things they don't agree with
   (c) very difficult calculation used in astronomy
   (d) fast way to solve problems around the house
3. The truth table using "incrementing input" for the OR gate is correctly represented as:
   (a)

Input "A"	Input "B"	"OR" Output
False	True	True
True	False	True
False	False	False
True	True	True

(b)

Input “A”	Input “B”	“OR” Output
False	False	False
False	True	True
True	False	True
True	True	False

(c)

Input “A”	Input “B”	“OR” Output
False	False	False
False	True	True
True	False	True
True	True	True

(d)

Input “A”	Input “B”	“OR” Output
False	False	False
False	True	False
True	False	False
True	True	True

4. When writing a logic equation, which symbols are typically used to represent optional operations?
   (a) {and}
   (b) < and >
   (c) (and)
   (d) [and]

5. If the output of an Exclusive OR gate was passed to a NOT gate's input, the NOT gate output would be "True" if:
   (a) Input "A" was True and input "B" is False
   (b) There is only one input and the output would be True if the input was False
   (c) A dot was placed on the output of the Exclusive OR symbol
   (d) Both inputs were at the same state (either True or False)

6. Boolean arithmetic statements are similar to:
   (a) Verbal descriptions of what the logic is to do
   (b) HTML, the language used to program an internet web page
   (c) Simple mathematical equations
   (d) The holes punched into computer cards

7. When decoding a combinatorial logic circuit diagram, you
   (a) Write out the Boolean arithmetic equation for the function and list the output for each possible input
   (b) Start slamming your forehead on your desk
   (c) Give each gate's output a unique label and list their outputs for each changing circuit input as well as outputs for other gates in the circuit
   (d) Rearrange the gates in the diagram to help you understand what the function is doing

8. "Sum of product" combinatorial logic circuits are popular because:
   (a) They are the most efficient way of designing circuitry
   (b) Their operation can be quickly seen by looking at the circuit diagram
   (c) They dissipate less heat than other design methodologies
   (d) They are more robust and allow for more reliable product designs

9. When trying to debug a digital clock circuit, what tool is not recommended?
   (a) Truth tables
   (b) Boolean arithmetic
   (c) State lists
   (d) Graphical circuit diagrams

10. Waveform diagrams display:
    - (a) Logic state changes over time
    - (b) Switching times of digital electronic gates
    - (c) Problems with line impedance
    - (d) Voltage variances in a logic signal over time

CHAPTER 2

Effectively Optimizing Combinatorial Circuits

In the first chapter, I introduced you to the basic theory behind digital electronics: binary data is manipulated by six different simple operations. With this knowledge, you actually have enough information to be able to design very complex operations, taking a number of different bits as input. The problem with these circuits is that they will probably not be ‘‘optimized’’ in order to minimize the number of gates, the speed which the digital electronic circuit responds to the inputs and finally, whether or not the circuit is optimized for the technology that it will be implemented in.

These three parameters are the basic measurements used to determine whether or not a circuit is effectively optimized. The number of gates should be an obvious one and you should realize that the more gates, the higher the chip count and cost of implementing the circuit as well as the increased complexity in wiring it. Right now, connections between logic gates are just black lines on paper to you – but when you start trying to wire circuits that

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

you have designed, you will discover first hand that simplifying the wiring of a circuit often reduces costs more than reducing the number of chips would indicate. Small improvements in the complexity of a circuit can have surprising cost ramifications when you look at the overall cost of the application. You may find that eliminating 1% of the gates in an application will result in as much as a 10–20% overall reductions in product cost. These savings are a result of being able to build the circuit on a smaller PCB or one which requires fewer layers (which can reduce the overall product cost dramatically). If the application is going to use a programmable logic technology, you may find that with the optimized circuit, lower cost chips can be substituted into the design. Fewer gates in an application also results in less power being dissipated by the circuit, requiring less cooling and a smaller power supply.

The speed that signals pass through gates is not infinite; standard TTL requires 8 billionths of a second (called a ‘‘nanosecond’’ and uses the abbreviation ‘‘ns’’) to pass a signal through a ‘‘NAND’’ gate. The term given to this time is known as the ‘‘gate delay’’. Halving the number of gates a signal has to pass through (which is halving the number of gate delays) will double the speed in which it can respond to a changing input. As you work with more complex circuits, you will discover that you will often have to optimize a circuit for speed or else you may have to use a faster (and generally more expensive) technology.

The last parameter, what I call ‘‘technology optimization’’, on the surface may seem more intangible than the other two parameters as well as have its measurements use the other two parameters, but when working with physical devices, it is the most important factor in properly optimizing your application. Before moving on and considering your circuit ‘‘done’’, you should look at how it will actually be implemented in the technology that you are using and look for optimizations that will reduce the actual number of gates and gate delays required by the application.

You can consider logic optimization to be a recursive operation, repeatedly optimizing all the different parameters and measurements. Once you have specified the required logic functions, you should look at how it will be implemented in the actual circuit. Once you have converted it to the actual circuit, you will then go back and look for opportunities for decreasing the number of gates, speeding up the time the signal passes through the gates and again look for technology optimizations. This will continue until you are satisfied with the final result.

To illustrate what I mean, in this chapter, I will look at a practical example, a simple home burglar alarm. In Fig. 2-1, I have drawn a very basic house, which has two windows, a door and power running to it. Sensors, on

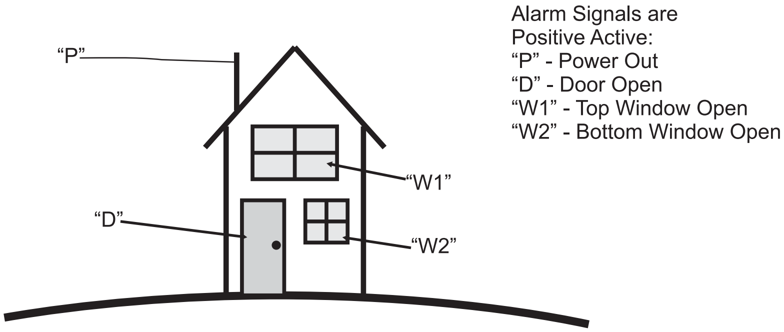

Fig. 2-1. Home alarm logic.

Table 2-1 Home alarm truth table.

P	D	W1	W2	Alarm Response
0	0	0	0	
0	0	0	1	
0	0	1	0	Sound Alarm
0	0	1	1	
0	1	0	0	
0	1	0	1	
0	1	1	0	Sound Alarm
0	1	1	1	
1	0	0	0	
1	0	0	1	Sound Alarm
1	0	1	0	Sound Alarm
1	0	1	1	Sound Alarm
1	1	0	0	
1	1	0	1	Sound Alarm
1	1	1	0	Sound Alarm
1	1	1	1	Sound Alarm

the windows, door and power are passed to an alarm system. When the alarm system was designed a table of the different possible combinations of inputs was generated (Table 2-1), with the combinations that would cause the alarm to sound indicated. As I have noted in Fig. 2-1, the alarm inputs are positive active, which means I can represent them as being active with a "1".

In this fictional house, I assumed that the upper window ("W1") should never be opened – if it were opened, then the alarm would sound. Along with this, I decided that if the power failed and either of the windows were opened, then the alarm failed; this would be the case where the power to the house was cut and somebody forced open the window. Table 2-1 shows the cases where the alarm should sound and you will notice that the cases where the alarm should sound are either a single event in the table, or a case where three are grouped together.

After building the table, you should also create a sum of products equation for the function:

$$
\begin{aligned}
\texttt{Alarm State} = {} & (\texttt{!P} \cdot \texttt{!D} \cdot \texttt{W1} \cdot \texttt{!W2}) \\
& + (\texttt{!P} \cdot \texttt{D} \cdot \texttt{W1} \cdot \texttt{!W2}) \\
& + (\texttt{P} \cdot \texttt{!D} \cdot \texttt{!W1} \cdot \texttt{W2}) \\
& + (\texttt{P} \cdot \texttt{!D} \cdot \texttt{W1} \cdot \texttt{!W2}) \\
& + (\texttt{P} \cdot \texttt{!D} \cdot \texttt{W1} \cdot \texttt{W2}) \\
& + (\texttt{P} \cdot \texttt{D} \cdot \texttt{!W1} \cdot \texttt{W2}) \\
& + (\texttt{P} \cdot \texttt{D} \cdot \texttt{W1} \cdot \texttt{!W2}) \\
& + (\texttt{P} \cdot \texttt{D} \cdot \texttt{W1} \cdot \texttt{W2})
\end{aligned}
$$

You could also draw a logic diagram using the gate symbols that I introduced in the first chapter. I found that this diagram was very complex and very difficult to follow. If you were to try it yourself, you would discover that the logic diagram would consist of 12 NOTs, 24 two input ANDs (knowing that a single four input AND can be produced from three two input ANDs) and seven two input OR gates with the maximum gate delay being eleven (the number of basic TTL gates the signal has to pass through). At first take, this alarm function is quite complex.

Looking at Table 2-1 and the sum of products equation, you will be hard pressed to believe that this home alarm circuit can be significantly optimized, but in this chapter, I will show how these four alarm inputs and eight alarm events can be reduced to fit in the most basic TTL chip there is.

Truth Table Function Reduction

I like to tell new circuit designers to approach optimizing a logic circuit by first looking for opportunities in its truth table. This may not seem like a useful tool (especially in light of Table 2-1), but it can be as effective a tool as any of the others presented in this chapter. It can also be used as a useful verification tool for making sure that an optimized logic circuit will perform the desired function. The drawback to the truth table function reduction is that it tends to be the most demanding in terms of the amount of rote effort that you will have to put into it.

Table 2-2 Gray code home alarm truth table.

P	D	W1	W2	Alarm Response
0	0	0	0	
0	0	0	1	
0	0	1	1	
0	0	1	0	Sound Alarm
0	1	1	0	Sound Alarm
0	1	1	1	
0	1	0	1	
0	1	0	0	
1	1	0	0	
1	1	0	1	Sound Alarm
1	1	1	1	Sound Alarm
1	1	1	0	Sound Alarm
1	0	1	0	Sound Alarm
1	0	1	1	Sound Alarm
1	0	0	1	Sound Alarm
1	0	0	0	

In the introduction to this chapter, the initial truth table I came up with didn't seem very helpful. The reason for this is something that I will harp upon throughout this book – listing logic responses to binary input is not very effective, because of the large number of states that can change at any given time. If you look at Table 2-1, you will see that going from the state where $P=0$, $D=W1=W2=1$ to $P=1$, $D=W1=W2=0$ involves the changing of four bits. While this is a natural progression of binary numbers and probably an intuitive way of coming up with a number of different input states, it is not an effective way to look at how a logic circuit responds to varying inputs.

A much better method is to list the output responses in a truth table that is ordered using Gray codes, as I have shown in Table 2-2. Gray codes are a numbering system in which only one bit changes at a time: they are explained in detail along with how they are generated in Chapter 4. When you are listing data, regardless of the situation, you should *always* default to using Gray code inputs instead of incrementing binary inputs, as I have shown in Table 2-1.

Taking this advice, I recreated the home alarm system truth table using Gray codes in Table 2-2. When you look at Table 2-2, you should notice that the "discontinuities" of Table 2-1 have disappeared. The bit patterns which "Sound Alarm" group together quite nicely.

Looking at each value which "Sound Alarm", you'll notice that each pair has three bits in common. To illustrate this, in Table 2-3, I have circled the bit which is different between each of the four pairs. In each of these

Table 2-3 Uncommon bits in “Sound Alarm” pairs.

P	D	W1	W2	Alarm Response
0	0	0	0	
0	0	0	1	
0	0	1	1	
0	0	1	0	Sound Alarm
0	1	1	0	Sound Alarm
0	1	1	1	
0	1	0	1	
0	1	0	0	
1	1	0	0	
1	1	0	1	Sound Alarm
1	1	1	1	Sound Alarm
1	1	1	0	Sound Alarm
1	0	1	0	Sound Alarm
1	0	1	1	Sound Alarm
1	0	0	1	Sound Alarm
1	0	0	0	

Table 2-4 Home alarm truth table with don’t care bits replaced with an “x”.

P	D	W1	W2	Alarm Response
0	0	0	0	
0	0	0	1	
0	0	1	1	
0	x	1	0	Sound Alarm
0	1	1	1	
0	1	0	1	
0	1	0	0	
1	1	0	0	
1	1	x	1	Sound Alarm
1	x	1	0	Sound Alarm
1	0	x	1	Sound Alarm
1	0	0	0	

pairs, to sound the alarm we have very specific requirements for three bits, but the fourth bit can be in either state.

Another way of saying this is: for the alarm to sound, we don’t care what the fourth bit is and it can be ignored when we are determining the sum of products equation for the logic function. To indicate the “don’t care” bit, in Table 2-4, I have combined the bit pairs and changed the previously circled bits with an “x”. This “x” indicates that the bit can be in either state for the output to be true. By replacing the two truth table entries with a single one with the don’t care bit indicated by an “x” you should see that something magical is starting to happen.

The obvious observation is that the table is shorter, but you should notice that the number of events which “Sound Alarm” has been halved and they are less complex than the eight original events. The sum of products equation for the bits shown in Table 2-4 is:

$$\begin{aligned}\texttt{Alarm State} = {} & (!\texttt{P}\cdot\texttt{W1}\cdot!\texttt{W2})\\ & + (\texttt{P}\cdot\texttt{D}\cdot\texttt{W2})\\ & + (\texttt{P}\cdot\texttt{W1}\cdot!\texttt{W2})\\ & + (\texttt{P}\cdot!\texttt{D}\cdot\texttt{W2})\end{aligned}$$

This sum of products expression will require four NOT gates, eight AND gates and three OR gates and the maximum gate delay will be nine. This has reduced the total gate count to less than 50% of the original total and this logic equation will operate somewhat faster than the original.

This is pretty good improvement in the logic circuit. You should be asking yourself if we can do better. To see if we can do better, I rearranged the data in Table 2-4 so that the “Sound Alarm” events with common don’t care bits were put together and came up with Table 2-5.

When I put the “Sound Alarm” events that had the same don’t care bits together, I noticed that in each of these cases, two of the remaining bits were in common and one bit changed in the two events (which I circled in Table 2.5).

In Table 2-5, you may have noticed that the single changing bit of the original Gray code input sequence has been lost; this is not a problem. The Gray code sequence has served its purpose – it has indicated the initial input patterns which are common with its neighbors. In complex truth tables, you may have to rearrange bit patterns multiple times to find different

Table 2-5 Optimized home alarm truth table rearranged with don’t care event bits moved together.

P	D	W1	W2	Alarm Response
0	0	0	0	
0	0	0	1	
0	0	1	1	
0	x	1	0	Sound Alarm
1	x	1	0	Sound Alarm
0	1	1	1	
0	1	0	1	
0	1	0	0	
1	1	0	0	
1	1	x	1	Sound Alarm
1	0	x	1	Sound Alarm
1	0	0	0	

Table 2-6 Reoptimized home alarm truth table with "don't care" bits replaced with an "x".

P	D	W1	W2	Alarm Response
0	0	0	0	
0	0	0	1	
0	0	1	1	
x	x	1	0	Sound Alarm
0	1	1	1	
0	1	0	1	
0	1	0	0	
1	1	0	0	
1	x	x	1	Sound Alarm
1	0	0	0	

commonalities. When you do this, don't worry about "loosing data"; the important bit patterns are still saved in the active bit patterns.

Table 2-6 shows what happens when the second don't care bit is indicated. Since the two events which "Sound Alarm" do not have common don't care bits, we can't repeat this process any more. The two events from Table 2-6 can be written out as the sum of products:

$$\begin{aligned}\texttt{Alarm State} &= (\texttt{W1} \cdot !\texttt{W2}) \\ &\quad + (\texttt{P} \cdot \texttt{W2})\end{aligned}$$

This optimized "Alarm State" truth table has reduced our component count to one NOT gate, two AND gates and one OR gate and executes in five gate delays – quite an improvement from the original 43 gates and 11 gate delays!

Depending on how cynical you are, you might think that I "cooked up" this example to come up with such a dramatic improvement. Actually, the application shown here was my first attempt at coming up with a logic circuit to demonstrate how optimization operations of a logic circuit are performed; you will find similar improvements as this one when you start with a basic logic circuit and want to see how much you can reduce it.

Karnaugh Maps

Using truth tables is an effective but not very efficient method of optimizing digital logic circuits. A very clever French mathematician, Maurice Karnaugh (pronounced "carno") came up with a way to simplify the truth table optimization process by splitting the truth table inputs down the middle and arranging the two halves perpendicularly in order to display the relationships between bits more effectively. These modified truth tables are called "Karnaugh Maps" and are best suited for single bit output functions with three to six input bits.

My description of what a Karnaugh map is may sound cursory, but it is actually very accurate. A standard truth table can be considered to be a single dimensional presentation of a logic function and when it is properly manipulated, relationship between active outputs can be observed as I showed in the previous section. The problem with this method is that it is fairly labor intensive and will burn up a lot of paper. Karnaugh maps present the data in a two-dimensional "field" which allows for quick scanning of active output bits against their inputs, to find basic relationships between them.

An example of converting a three input logic function from a truth table to a Karnaugh map is shown in Fig. 2-2. The initial logic function would be:

$$\begin{aligned}\mathtt{Output} = {} & (!\mathtt{A}\cdot !\mathtt{B}\cdot \mathtt{C}) + (!\mathtt{A}\cdot \mathtt{B}\cdot !\mathtt{C}) \\ & + (\mathtt{A}\cdot \mathtt{B}\cdot !\mathtt{C}) + (\mathtt{A}\cdot \mathtt{B}\cdot \mathtt{C}) \\ & + (\mathtt{A}\cdot !\mathtt{B}\cdot \mathtt{C})\end{aligned}$$

To create the Karnaugh map, I created a two by four matrix, with the rows being given the two different values for "A" and the columns given the four different values for "B" and "C". Note that the columns are listed as a two bit Gray code – this is an important feature of the Karnaugh map and, as I have pointed out, an important tool to being able to optimize a function.

Once the two axes of the Karnaugh map are chosen, the outputs from the truth table are *carefully* transferred from the truth table to the Karnaugh map. When transferring the outputs, treat the Karnaugh map as a two-dimensional array, with the "X" dimension being the inputs which

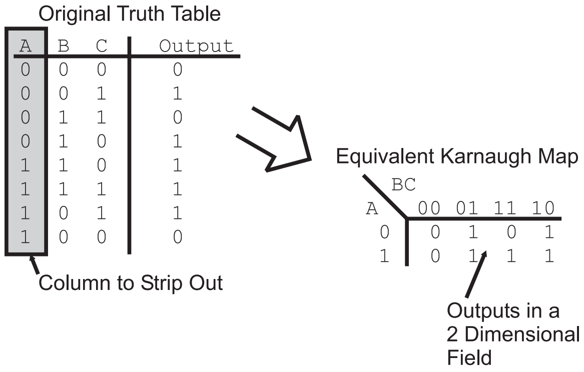

A	B	C	Output
0	0	0	0
0	0	1	1
0	1	1	0
0	1	0	1
1	1	0	1
1	1	1	1
1	0	1	1
1	0	0	0

A \ BC	00	01	11	10
0	0	1	0	1
1	0	1	1	1

Fig. 2-2. Converting a truth table to a Karnaugh map.

weren't stripped out and the "Y" dimension being the inputs which were stripped out from the truth table. When you are first starting out, this will be an operation in which you will tend to make mistakes because it is unfamiliar to you. To make sure you understand the process, it is a good idea to go back and convert your Karnaugh map into a truth table and compare it to your original truth table.

When you have created the Karnaugh map for your function, it is a good idea to either photocopy it or write it out in pen before going on. I am suggesting this action because, just as you did with the truth table, you are going to circle outputs which have the same unchanging bit patterns. As you circle the outputs, chances are you are not going to see the most effective groups of bits to circle together, or you will find that you have made a mistake in circling the bits. A photocopy or list in ink will allow you to try again without having to redraw the Karnaugh map.

For the example shown in Fig. 2-2, the Karnaugh map has three circles put on it, as shown in Fig. 2-3. Each circle should result in combining two input sets together and making at least one bit into a "don't care".

Correctly circling bits can be difficult to understand, but there are a few rules that can be applied to it. First, each circle *must* be around a power of two number of bits – you cannot circle three bits (as shown in Fig. 2-4 for this example). Secondly, it is not a problem if circles overlap over specific bits. I should point out that there is the case for redundant circles (Fig. 2-5). If a circle is drawn and all the circled bits are enclosed in another circle, then the enclosed circle is redundant. Thirdly, remember that when you are circling bits that you want to circle a power of two number of bits, not just two. In Fig. 2-6, I have modified the three bit Karnaugh map with the outputs at A = 0 and B = C = 1 and A = 1 and B = C = 0 being a "1"

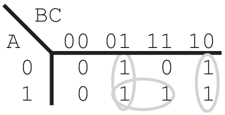

Fig. 2-3. Circling like bits in the example Karnaugh map.

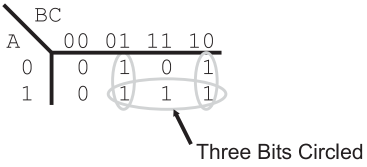

Fig. 2-4. Incorrectly circling an odd number of bits in the example Karnaugh map.

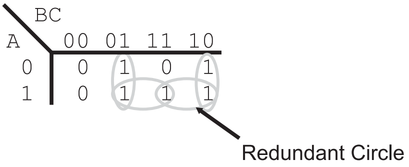
BC
A
00 01 11 10
0 0 1 0 1
1 0 1 1 1
Redundant Circle

Fig. 2-5. Redundant circles on the example Karnaugh map.

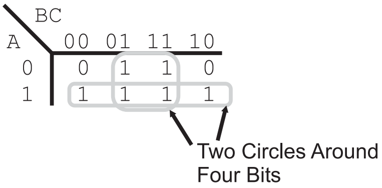
BC
A
00 01 11 10
0 0 1 1 0
1 1 1 1 1
Two Circles Around
Four Bits

Fig. 2-6. Karnaugh map showing that more than two bits can be circled at the same time.

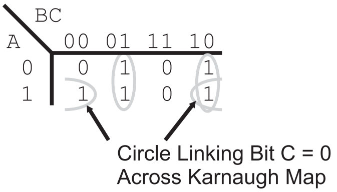
BC
A
00 01 11 10
0 0 1 0 1
1 1 1 0 1
Circle Linking Bit C = 0
Across Karnaugh Map

Fig. 2-7. Circle extending outside the apparent boundaries of the Karnaugh map.

and found that I could circle two groups of four bits. In each of these cases, I have made two bits "don't care".

Finally, saying that a Karnaugh map is like a two-dimensional array is inaccurate – it is actually a continuum unto itself, with the tops and sides being connected. When you draw out your Karnaugh map, you may find that the bits which can be circled (meaning ones with similar patterns) are on opposite ends of the Karnaugh map. This is not a problem as long as there are matching bits.

Once you have the outputs circled, you can now start writing out the optimized equation. As an exercise, you might want to look at the example Karnaugh maps in Figs. 2-3, 2-6 and 2-7. The output equations for these figures are:

$$\mathtt{Output}_{2.03} = (!\mathtt{B} \cdot \mathtt{C}) + (\mathtt{B} \cdot !\mathtt{C}) + (\mathtt{A} \cdot \mathtt{C})$$
$$\mathtt{Output}_{2.06} = \mathtt{A} \cdot \mathtt{C}$$
$$\mathtt{Output}_{2.07} = (!\mathtt{B} \cdot \mathtt{C}) + (\mathtt{B} \cdot !\mathtt{C}) + (\mathtt{A} \cdot !\mathtt{C})$$

P	D	W1	W2	Alarm Response
0	0	0	0	
0	0	0	1	
0	0	1	1	
0	0	1	0	Sound Alarm
0	1	1	0	Sound Alarm
0	1	1	1	
0	1	0	1	
0	1	0	0	
1	1	0	0	
1	1	0	1	Sound Alarm
1	1	1	1	Sound Alarm
1	1	1	0	Sound Alarm
1	0	1	0	Sound Alarm
1	0	1	1	Sound Alarm
1	0	0	1	Sound Alarm
1	0	0	0	

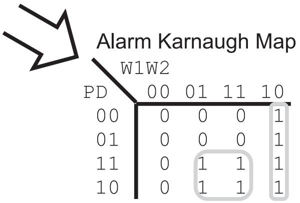
Alarm Karnaugh Map

PD \ W1W2	00	01	11	10
00	0	0	0	1
01	0	0	0	1
11	0	1	1	1
10	0	1	1	1

Fig. 2-8. Home alarm truth table to Karnaugh map.

In this chapter, I wanted to show how the different optimizing tools are used for the home alarm system presented in the chapter introduction. The alarm system's functions can be optimized using the Karnaugh map shown in Fig. 2-8. In Fig. 2-8, I have drawn the circles around the two groups of four active output bits which are in common and result in the logic equation

$$\texttt{Alarm Response} = (\texttt{P} \cdot \texttt{W2}) + (\texttt{W1} \cdot \texttt{!W2})$$

which is identical to the equation produced by the truth table reduction and a lot less work.

Before going on, I want to just say that once you are comfortable with Karnaugh maps, you will find them to be a fast and efficient method of optimizing simple logic functions. Becoming comfortable and being able to accurately convert the information from a truth table to a Karnaugh map will take some time, as will correctly circling active outputs to produce the optimized sum of products circuit. Once you have mastered this skill, you will find that you can go directly to the Karnaugh map from the requirements without the initial step of writing out the truth table.

Boolean Arithmetic Laws

One of the ways of optimizing circuits is look through their output equations and try to find relationships that you can take advantage of using the rules and laws in Table 2-7. These rules should be committed to memory as quickly as possible (or at least written down on a crib sheet) to help you with

Table 2-7 Boolean arithmetic laws and rules.

Rule/law	Boolean arithmetic example
AND identity function	$A \cdot 1 = A$
OR identity function	$A + 0 = A$
Output reset	$A \cdot 0 = 0$
Output set	$A + 1 = 1$
Identity law	$A = A$
AND complementary law	$A \cdot !A = 0$
OR complementary law	$A + !A = 1$
AND idempotent law	$A \cdot A = A$
OR idempotent law	$A + A = A$
AND commutative law	$A \cdot B = B \cdot A$
OR commutative law	$A + B = B + A$
AND associative law	$(A \cdot B) \cdot C = A \cdot (B \cdot C) = A \cdot B \cdot C$
OR associative law	$(A + B) + C = A + (B + C) = A + B + C$
AND distributive law	$A \cdot (B + C) = (A \cdot B) + (A \cdot C)$
OR distributive law	$A + (B \cdot C) = (A + B) \cdot (A + C)$
De Morgan's NOR theorem	$!(A + B) = !A \cdot !B$
De Morgan's NAND theorem	$!(A \cdot B) = !A + !B$

optimizing logic equations without the need of truth tables or Karnaugh maps. Many of these rules and laws will seem self-evident, but when you are working at optimizing a logic equation in an exam, it is amazing what you will forget or won't seem that obvious to you.

When I talk about using the laws and rules in Table 2-7 to simplify a logic equation, I normally use the term "reduce" instead of "optimize". The reason for thinking of these operations as a reduction is due to how much the logic equation shrinks as you work through it, trying to find the most efficient sum of products expression.

The two identity functions are used to indicate the conditions where an input value can pass unchanged through an AND or OR gate. The output set, reset and complementary laws are used to output a specific state when a value is passing through an AND or OR gate. The idempotent laws can be summarized by saying that if an input passes through a non-inverting gate, its value is not changed.

The remaining laws – commutative, associative and distributive – and De Morgan's theorems are not as trivial and are extremely powerful tools when you have a logic equation to optimize. The commutative laws state that the inputs to AND and OR gates can be reversed, which may seem obvious, but when you have a long logic equation that is written in an arbitrary format (not necessarily in sum of product format), you can get confused very easily as to what is happening. It's useful to have a law like this in your back pocket to change the logic equation into something that you can more easily manipulate.

To demonstrate the operation of these laws, we can go back to some of the logic circuits described in the Karnaugh map examples of the previous section. Looking at Fig. 2-3, the initial sum of products logic equation would be:

$$\begin{aligned}\texttt{Output} = {} & (!\texttt{A} \cdot !\texttt{B} \cdot \texttt{C}) + (!\texttt{A} \cdot \texttt{B} \cdot !\texttt{C}) + \\ & (\texttt{A} \cdot !\texttt{B} \cdot \texttt{C}) + (\texttt{A} \cdot \texttt{B} \cdot \texttt{C}) + \\ & (\texttt{A} \cdot \texttt{B} \cdot !\texttt{C})\end{aligned}$$

Using the AND associative law, I can rewrite this equation with the A term separate from the B and C terms to see if there are any cases where the B and C terms are identical.

$$\begin{aligned}\texttt{Output} = {} & !\texttt{A} \cdot (!\texttt{B} \cdot \texttt{C}) + !\texttt{A} \cdot (\texttt{B} \cdot !\texttt{C}) + \\ & \texttt{A} \cdot (!\texttt{B} \cdot \texttt{C}) + \texttt{A} \cdot (\texttt{B} \cdot \texttt{C}) + \\ & \texttt{A} \cdot (\texttt{B} \cdot !\texttt{C})\end{aligned}$$

By doing this, I can see that the inside terms of the first and third products are identical. Along with this, I can see that the second and fifth products are also identical. Using the OR distributive law, I can combine the first and third terms like:

$$\begin{aligned}&!\mathtt{A}\cdot(!\mathtt{B}\cdot\mathtt{C})+\mathtt{A}\cdot(!\mathtt{B}\cdot\mathtt{C})\\&\quad=(!\mathtt{A}+\mathtt{A})\cdot(!\mathtt{B}\cdot\mathtt{C})\end{aligned}$$

Using the OR complementary law, I know that A OR !A will always be true. This is actually a clear and graphic example of the "don't care" bit; regardless of the value of this bit, the output will be true so it can be ignored. The partial equation of the two terms reduces to:

$$\begin{aligned}&(!\mathtt{A}\cdot!\mathtt{B}\cdot\mathtt{C})+\mathtt{A}\cdot(!\mathtt{B}\cdot\mathtt{C})\\&\quad=\mathtt{1}\cdot(!\mathtt{B}\cdot\mathtt{C})\end{aligned}$$

The 1 ANDed with !B AND C can be further reduced using the AND identity law (1 AND A equals A):

$$\begin{aligned}&!\mathtt{A}\cdot(!\mathtt{B}\cdot\mathtt{C})+\mathtt{A}\cdot(!\mathtt{B}\cdot\mathtt{C})\\&\quad=(!\mathtt{B}\cdot\mathtt{C})\end{aligned}$$

This can be repeated for the second and fifth terms:

$$\begin{aligned}&(!\mathtt{A}\cdot\mathtt{B}\cdot!\mathtt{C})+(\mathtt{A}\cdot\mathtt{B}\cdot!\mathtt{C})\\&\quad=(\mathtt{B}\cdot\mathtt{C})\end{aligned}$$

If you go back to the original logic equation, you will see that the fourth term $(A\cdot B\cdot C)$ has not been reduced by combining it with another term. It can actually be paired with the third term $(A\cdot!B\cdot C)$ by rearranging the two terms (using the AND commutative law) so that part of the terms operating on two bits are in common $(A\cdot C)$. Once this is done, the third and fourth terms can be reduced as:

$$\begin{aligned}&(\mathtt{A}\cdot!\mathtt{B}\cdot\mathtt{C})+(\mathtt{A}\cdot\mathtt{B}\cdot!\mathtt{C})\\&\quad=(\mathtt{A}\cdot!\mathtt{C})\end{aligned}$$

After doing this work, the optimized or reduced sum of product logic equation for this function is

$$\mathtt{Output}=(!\mathtt{B}\cdot\mathtt{C})+(\mathtt{B}\cdot!\mathtt{C})+(\mathtt{A}\cdot\mathtt{C})$$

which is identical to what was found using the Karnaugh map.

Looking at the reduced logic equation, you should have noticed that there are two terms that will output a "1" at the same time $((!B\cdot C)$ and $(A\cdot C)$ with $A=1$, $B=0$ and $C=1$). This is not a problem because the OR gate (even though the symbol that I use is a "+") will only output a 1, regardless of how many true inputs it has. This was mentioned when the

Karnaugh maps were presented, but I wanted to reinforce that the same issue is present when you are reducing logic equations.

Before moving on, let's go back to the home alarm logic equation and see if it can be reduced in the same way as the example above. Starting with the sum of products logic equation:

$$
\begin{aligned}
\texttt{Alarm State} = {} & (\texttt{!P} \cdot \texttt{!D} \cdot \texttt{W1} \cdot \texttt{!W2}) \\
& + (\texttt{!P} \cdot \texttt{D} \cdot \texttt{W1} \cdot \texttt{!W2}) \\
& + (\texttt{P} \cdot \texttt{!D} \cdot \texttt{!W1} \cdot \texttt{W2}) \\
& + (\texttt{P} \cdot \texttt{!D} \cdot \texttt{W1} \cdot \texttt{!W2}) \\
& + (\texttt{P} \cdot \texttt{!D} \cdot \texttt{W1} \cdot \texttt{W2}) \\
& + (\texttt{P} \cdot \texttt{D} \cdot \texttt{!W1} \cdot \texttt{W2}) \\
& + (\texttt{P} \cdot \texttt{D} \cdot \texttt{W1} \cdot \texttt{!W2}) \\
& + (\texttt{P} \cdot \texttt{D} \cdot \texttt{W1} \cdot \texttt{W2})
\end{aligned}
$$

We can bring out the "P" values from the products and look for similarities in the remaining bracketed values and combine them using the associative, distributive, complementary and AND identity laws. I can see that the first and fourth, second and seventh can be combined, resulting in the logic equation:

$$
\begin{aligned}
\texttt{Alarm State} = {} & (\texttt{!D} \cdot \texttt{W1} \cdot \texttt{!W2}) \\
& + (\texttt{D} \cdot \texttt{W1} \cdot \texttt{!W2}) \\
& + (\texttt{P} \cdot \texttt{!D} \cdot \texttt{!W1} \cdot \texttt{W2}) \\
& + (\texttt{P} \cdot \texttt{!D} \cdot \texttt{W1} \cdot \texttt{W2}) \\
& + (\texttt{P} \cdot \texttt{D} \cdot \texttt{!W1} \cdot \texttt{W2}) \\
& + (\texttt{P} \cdot \texttt{D} \cdot \texttt{W1} \cdot \texttt{W2})
\end{aligned}
$$

Bringing "W1" to the forefront allows the combination of the third and fourth and fifth and sixth terms of the logic equation above, resulting in the new equation:

$$
\begin{aligned}
\texttt{Alarm State} = {} & (\texttt{!D} \cdot \texttt{W1} \cdot \texttt{!W2}) \\
& + (\texttt{D} \cdot \texttt{W1} \cdot \texttt{!W2}) \\
& + (\texttt{P} \cdot \texttt{!D} \cdot \texttt{W2}) \\
& + (\texttt{P} \cdot \texttt{D} \cdot \texttt{W2})
\end{aligned}
$$

We have eliminated half the terms and, of those remaining, they are 25% smaller. Looking at the new logic equation, we can see that by combining the first and second terms (making "D" a don't care bit in the process)

Table 2-8 Testing optimized home alarm logic equation.

P	D	W1	W2	Alarm	W1 ● !W2	P ● W2	OR
0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0
0	0	1	1	0	0	0	0
0	0	1	0	1	1	0	1
0	1	1	0	1	1	0	1
0	1	1	1	0	0	0	0
0	1	0	1	0	0	0	0
0	1	0	0	0	0	0	0
1	1	0	0	0	0	0	0
1	1	0	1	1	0	1	1
1	1	1	1	1	0	1	1
1	1	1	0	1	1	0	1
1	0	1	0	1	1	0	1
1	0	1	1	1	0	1	0
1	0	0	1	1	0	1	0
1	0	0	0	0	0	0	0

and combining the third and fourth terms ("D" again is the don't care bit) we end up with:

$$\texttt{Alarm State} = (\texttt{W1} \cdot \texttt{!W2}) + (\texttt{P} \cdot \texttt{W2})$$

which is, again, the logic equation found by optimizing the function using truth tables or Karnaugh maps.

Personally, I tend to optimize logic equations using the Boolean arithmetic laws and rules listed in Table 2-7. Once a reduced sum of products equation has been produced, I then go back and compare its outputs in a truth table with the required outputs. In doing this, I present the values for each product (AND) and the final sum (OR) in separate columns, as shown in Table 2-8.

Optimizing for Technology

If you review the laws in Table 2-7 and correlate them to the text in the previous section, you'll see that I missed the last two (De Morgan's theorem). These two laws are not typically used during basic logic equation reduction because they typically involve converting part of an equation into an NAND or NOR gate, which is important when finally implementing a logic function in actual electronics. Another important aspect of optimizing for technology is adding functions out of the leftover gates in your circuit; by looking at how differently a logic circuit could be implemented, you may be able to add functionality to your circuit, without adding any cost to it.

Table 2-9 XOR gate truth table.

A	B	A ˆ B
0	0	0
0	1	1
1	1	0
1	0	1

So far in the book, I haven't discussed the "Exclusive OR" (XOR) gate in a lot of detail, but it is vital for implementing binary adders, as I will show you later in the book. In the first chapter, I presented the XOR gate with the truth table shown in Table 2-9.

You should probably be able to create the logic equation for the XOR table as:

$$\texttt{Output} = (!\texttt{A} \cdot \texttt{B}) + (\texttt{A} \cdot !\texttt{B})$$

which does not seem like a very likely candidate for optimization. Similarly, you probably would have a hard time believing that the following logic equation would perform the same function:

$$\texttt{Output} = !((\texttt{A} \cdot \texttt{B}) + !(\texttt{A} + \texttt{B}))$$

But, using De Morgan's theorem as well as the other rules and laws from Table 2-7, I can go through the manipulations shown in Table 2-10 to show that they are equal, as well as count out the gates required by intermediate steps to give you a list of different implementations of the XOR gate. Each intermediate step in Table 2-10 is an implementation of the XOR gate that you could implement using the number of gates listed to the right of the terms.

It's interesting to note that a total of five gates is required for each implementation – this is not something that you can count on when you are working at optimizing a circuit.

The basic gate used in TTL is the "NAND" gate: this means that the three basic gates (AND, OR and NOT) are built from multiples of it, as I've shown in Fig. 2-9. The basic gate for CMOS is the NOR gate, and Fig. 2-10 shows how the three basic gates are implemented for it. The three gate NAND and NOR equivalencies for the OR and AND gates, respectively, are perfect examples of De Morgan's theorem in operation. These implementations

Table 2-10 Different implementations of the XOR gate.

Terms	NOTs	ANDs	ORs	NANDs	NORs
$(!A \cdot B) + (A \cdot !B)$	2	2	1	0	0
$!(!(!A \cdot B) \cdot !(A \cdot !B))$	2	0	0	3	0
$!((A + !B) \cdot (!A + B))$	2	0	2	1	0
$!((A \cdot B) + (!A \cdot !B))$	2	2	0	0	1
$!((A \cdot B) + !(A + B))$	0	1	0	0	2

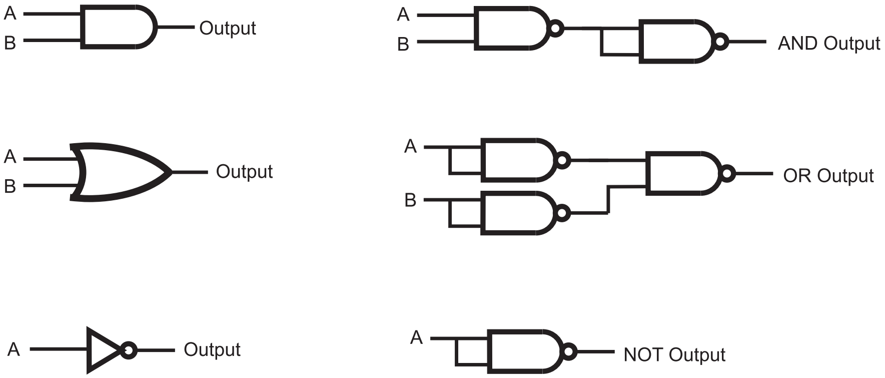

Fig. 2-9. Implementing the three basic gates using NAND gates.

can be checked against De Morgan's theorem and the rules and laws presented in Table 2-7.

By understanding how gates are implemented in chips, we can now look at how to optimize the gates to provide the fastest possible operation of the logic function. Using the example of the XOR gate, we can graphically show how the gate is implemented using ANDs, ORs and NOTs and how these gates are implemented as NAND gates in TTL chips (Fig. 2-11).

Looking at the bottom logic diagram of Fig. 2-11, you can see that there are two sets of NAND gates wired as inverters together. Going back to Table 2-7, we can see that a doubly inverted signal is the same signal, so we can eliminate these two sets of NOT gates, as shown in Fig. 2-12. The resulting XOR circuit will pass signals through three NAND gates, which

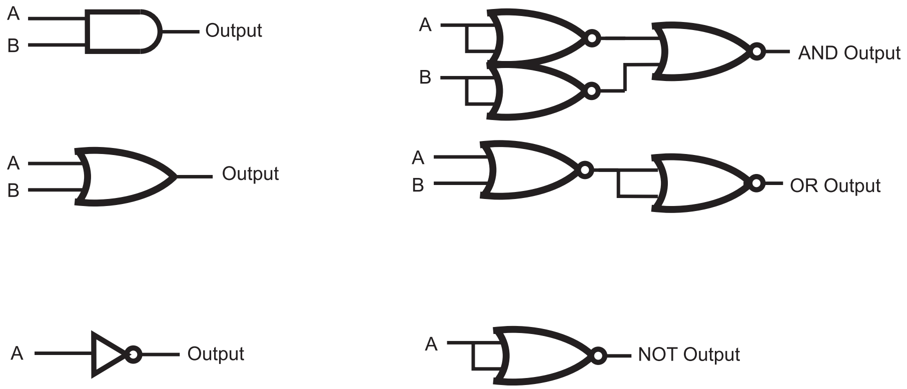

Fig. 2-10. Implementing the three basic gates using NOR gates.

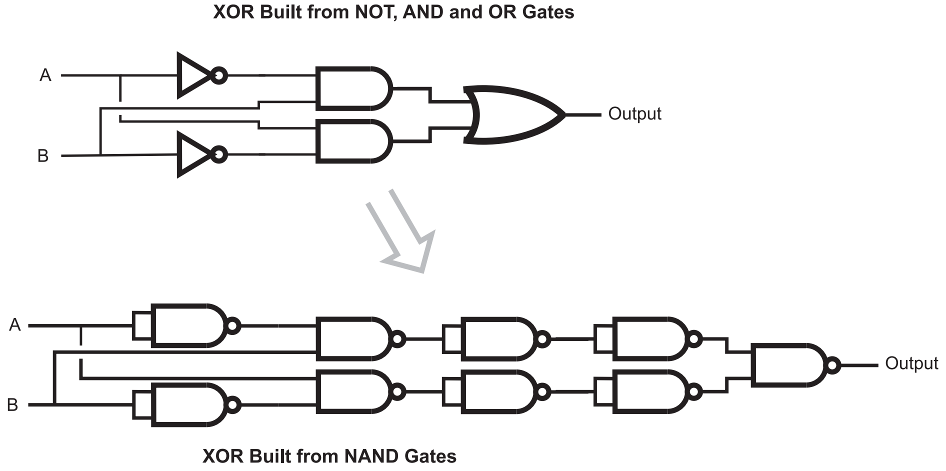
XOR Built from NOT, AND and OR Gates

XOR Built from NAND Gates

Fig. 2-11. XOR gate built from sum of products equation and converted to NAND gates.

counts as three “gate delays”. This is an example of what I call “technology optimization”: the logic circuit has been reduced to its bare minimum, taking advantage of the operation of the basic logic gates that make up the technology that it is implemented in.

Before moving on, I want to take one more look at the home alarm circuit that has been discussed throughout this chapter. I made a pretty bold statement at the start of the chapter, saying that it could be reduced to fit into the most basic TTL chip available – let’s see how honest I was being.

The (repeatedly) optimized logic equation for the home alarm system was:

$$\texttt{Alarm State} = (\texttt{W1} \cdot \texttt{!W2}) + (\texttt{P} \cdot \texttt{W2})$$

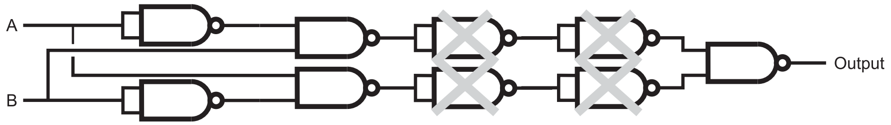

XOR Built from NAND Gates with Redundant Gates
Marked and Removed Below

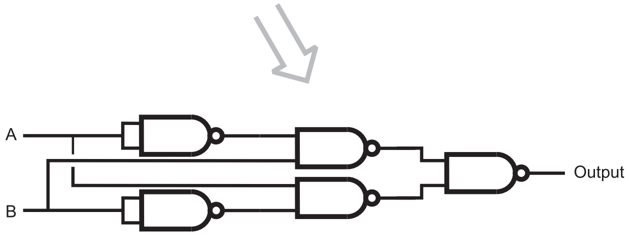

Fig. 2-12. Optimized XOR gate built from NAND gates.

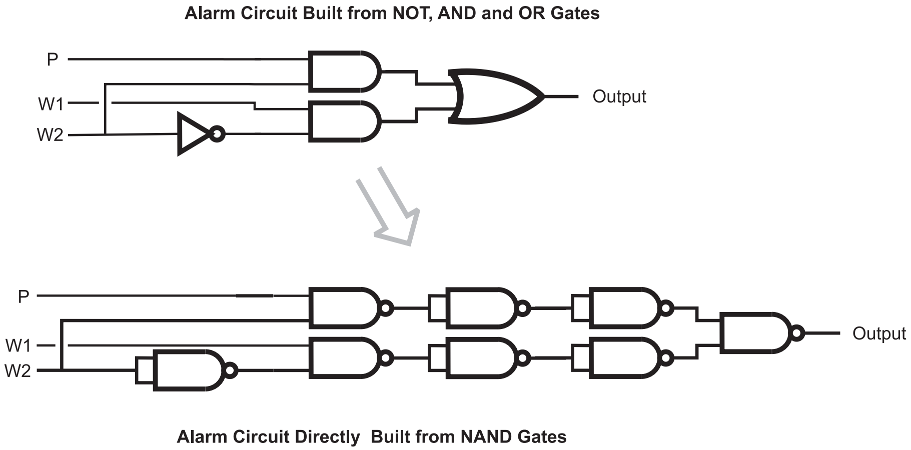
Alarm Circuit Built from NOT, AND and OR Gates

Alarm Circuit Directly Built from NAND Gates

Fig. 2-13. Home alarm logic circuit built using AND, OR and NOT gates and converted to NAND gates.

which could be first implemented in two AND, one OR and one NOT gate, as shown in Fig. 2-13 and converted to just NAND gates. You may have noted in Fig. 2-13 the remarkable similarity between the home alarm logic diagram and the XOR logic diagram – as I've shown in Fig. 2-14, the logic function reduces to just four NAND gates (one less than the XOR gate built out of NAND gates).

The final home alarm logic function requires four two input NAND gates – which is just what the 7400, the most basic TTL chip, provides. Every TTL chip, except for this one and a derivative revision, has more than four

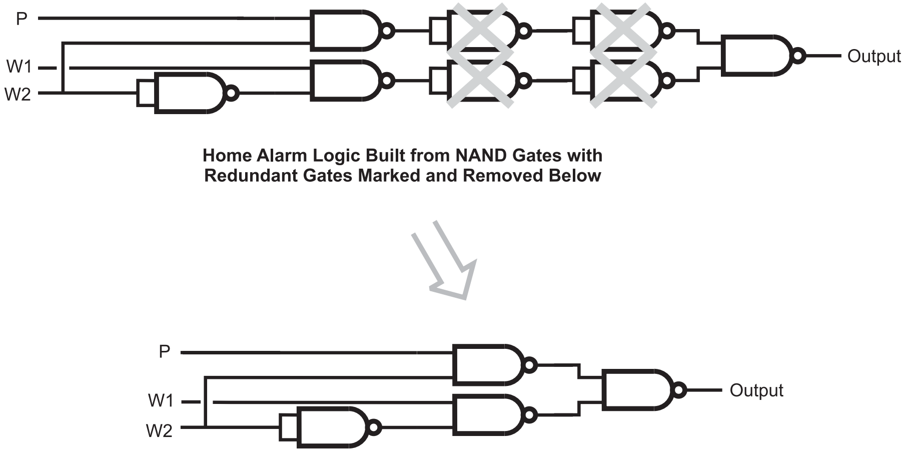
P
W1
W2
Output
Home Alarm Logic Built from NAND Gates with
Redundant Gates Marked and Removed Below
P
W1
W2
Output

Fig. 2-14. Optimized alarm circuit built from NAND gates.

gates built into them because they provide additional functions requiring multiple NAND gates. I was not exaggerating when I said that the home alarm logic function could be reduced to the most basic TTL chip available. In the next chapter, I will introduce you to the operation of TTL chips that provide the basis for digital electronic logic functions.

Quiz

1. The three parameters that are used to measure the optimization of a digital electronic circuit are:
   (a) Cost, speed and complexity
   (b) Gate delay, gate count and technology optimization
   (c) Gate count, number of gate delays a signal must pass through and technology optimization
   (d) Gate count, number of connections a signal must pass through and technology optimization
2. If TTL logic has a gate delay of 8 ns and the signal passing through an XOR gate built from NAND gates has to go through 9 gates and the shortest path is five gate delays, the time required for a signal to pass through the gates is:
   (a) 40 ns
   (b) 8 ns

(c) indeterminate
(d) 24 ns

3. When writing out a truth table, the inputs should be listed:
   (a) Using a "Gray code"
   (b) Using a "binary progression"
   (c) In alphabetical order
   (d) In order of importance

4. The "don't care" bit in a truth table is:
   (a) Indicated by a "dc" and replaces the common bits in two true sets of inputs
   (b) Indicated by an "x" and replaces the common bits in two true sets of inputs
   (c) Indicated by a "dc" and replaces the uncommon bits in two true sets of inputs
   (d) Indicated by an "x" and replaces the uncommon bits in two true sets of inputs

5. When optimizing a logic function you can expect:
   (a) That the number of chips that are required is reduced from the initial design
   (b) That the optimized function runs faster than the initial design
   (c) Cheaper chips can be used than in the initial design
   (d) Answers (a) through (c) are all possible and it might not be able to optimize the circuit from the initial sum of products equation

6. Karnaugh maps are:
   (a) Tools designed to help you find your way around a digital electronic circuit
   (b) A tool that will help you optimize a logic function
   (c) The most efficient method of optimizing logic fuctions
   (d) Hard to understand but must be used in every logic function design

7. The sum of products logic equation

   $$\texttt{Output} = (\texttt{A} \cdot \texttt{!B} \cdot \texttt{C}) + (\texttt{!A} \cdot \texttt{!B} \cdot \texttt{C})$$

   can be reduced to:
   (a) A · C
   (b) !A · !B
   (c) C · !B
   (d) C

8. Which of the following pairs of Boolean arithmetic laws cannot be used together?
   - (a) Identity and De Morgan's theorem
   - (b) Associative and idempotent
   - (c) Complementary and commutative
   - (d) All the laws and rules can be used together

9. The NAND equivalent to an AND gate is:
   - (a) Built from two NAND gates and requires two gate delays for a signal to pass through
   - (b) Built from three NAND gates and requires two gate delays for a signal to pass through
   - (c) Built from three NAND gates and requires three gate delays for a signal to pass through
   - (d) Built from one NAND gate as well as a NOT gate and requires two gate delays for a signal to pass through

10. Technology optimization is defined as:
    - (a) Designing the circuit which uses the fewest number of chips and signals pass through it as fast as possible
    - (b) Implementing logic functions to take advantage of the base logic of the logic technology used as well as using any leftover gates
    - (c) Finding the most efficient digital electronic technology to use for the application
    - (d) Designing circuitry that dissipates the least amount of heat to perform a desired function

CHAPTER 3

Creating Digital Electronic Circuits

In the previous chapters, I introduced you to the basic Boolean arithmetic theory behind decoding and design combinatorial circuits; binary data is manipulated by simple operations to produce a desired output. Before going on and showing you how these basic operations are extended to create complicated functions and products, I want to take a step back and look at basic electrical theory and semiconductor operation and how they are applied to digital electronics. While digital electronics work with ‘‘ones and zeros’’, it is still built from the basic electronic devices that are outlined in the beginning of this chapter. It is impossible to work successfully with digital electronics without understanding basic electrical theory and how simple electronic devices work.

For many people, this chapter will be a review, but I still urge you to read through this chapter and answer the quiz at the end of it. While you may be familiar with electrical rules and device operation, you may not be so comfortable understanding how they are used to create digital electronics.

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

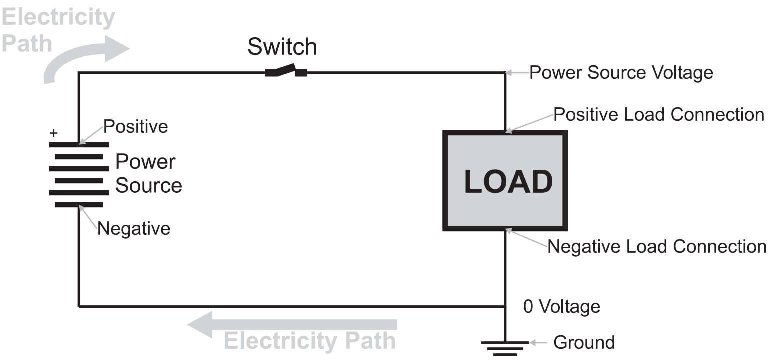

Fig. 3-1. Basic circuit diagram.

The most basic rule of electricity is that it can only move in a ‘‘closed circuit’’ (Fig. 3-1) in which a ‘‘power source’’ passes electricity to and then pulls it from a *load*. The power source has two connections that are marked with a ‘‘+’’ (‘‘positive’’) and ‘‘−’’ (‘‘negative’’) markings to indicate the ‘‘polarity’’ of the power source and the power source symbol consists of a number of pairs of parallel lines with the longer line in each pair representing the positive connection. The black lines connecting the power source to the load represent wires. When basic electricity is presented, this ‘‘load’’ is most often a lightbulb, because it turns on when electricity passes through it. As well as being a lightbulb, the load can be electrical motors, heater elements or digital electronic chips or any combination of these devices.

In the ‘‘electrical circuit’’ (or ‘‘schematic diagram’’) shown in Fig. 3-1 you can see that I have included a *switch,* which will *open* or *close* the circuit. When the switch is closed, electricity will flow through from the power source, to the load and back. If the switch is open or the wires connecting the power source to the load are broken, then electricity will not flow through the load.

As you are probably aware, electricity consists of electrons moving from the power source through the wires to the load and back to the power source. There are actually two properties of electricity that you should be aware of and they are analogous to the two properties of water flowing through a pipe. *Voltage* is the term given to the *pressure* placed on the electrons to move and *current* is the number of electrons passing by a point at a given time.

In the early days of electrical experimentation, it was Benjamin Franklin who postulated that electricity was a fluid, similar to water. As part of this supposition, he suggested that the electrical current flowed from the positive power supply connection to the negative. By suggesting that electrical current flowed from positive to negative, he started drawing electrical wiring diagrams or *schematics* (like the one in Fig. 3-1) with the electrical energy at the positive power supply connection being at the highest state. As the electrical current ‘‘flowed down’’ the page to the negative connection of the power supply, the energy of the electricity decreased. This method of drawing electrical circuits is clever and intuitive and caught on because it described what was happening in it.

Unfortunately, Franklin’s suggestion that electrical current flowed from the positive to negative connections of the power source through the load was wrong. As we now know, electrons that make up electricity flow from the negative to positive connections of the power supply. This discovery was made about 150 years after his kite in a lightning storm experiment, so the notion that electrical current flowed from positive to negative was widely accepted and was never really challenged. For this reason, you should keep in mind that ‘‘electrical current flow’’ takes place in the opposite direction to ‘‘electron flow’’ in electrical circuits. This point trips many people new to electronics and I should state emphatically that the direction of current flow follows Franklin’s convention.

Looking at the bottom right hand corner of Fig. 3-1, you will see a funny set of lines attached to the wiring lines – this is the circuit’s ‘‘ground’’ connection. The circuit ground is another invention of Benjamin Franklin. If there is ever a large amount of electricity that finds its way into the circuit, it will have an ‘‘escape route’’ to prevent damage to the circuit’s components or hurting anybody working with the circuit. The ground connection was literally a metal spike driven into the ground and connected to a home or barn’s lightning rod. In modern structures, the ‘‘ground’’ is a connection to the metal pipe bringing in water.

Another term commonly used for a circuit’s wire connections or wiring lines is ‘‘nets’’. The term originated when circuit analysis was first done on complex *networks* of wiring. It is used to describe the individual wiring connections in a circuit. I will use this term along with ‘‘wiring’’ and ‘‘lines’’ in this book interchangeably.

Like power supplies, many load devices also have connections that are marked with a positive (‘‘+’’) and negative (‘‘–’’) connections. When discussing the positive and negative connections of a basic two-wire load device, I like to use the terms, *anode* and *cathode* to describe the positive and negative connections of the load, respectively. The load’s anode must always be connected to the positive terminal of the power supply and the load’s cathode must always be connected to the negative terminal of the power supply. Reversing these connections may result in the device not working or even going so far as literally ‘‘burning out’’. To keep the terms anode and cathode straight, I remember that a “cathode ray tube” (i.e. your TV set) involves firing electrons, which are negative, at a phosphorus screen.

More complex load devices, like logic chips, also have positive and negative connections, but these connections are normally called *Vcc* or *Vdd* for the positive connection or *Gnd* and *Vss* for the negative (ground) connections.

When working with most basic digital electronic circuits, the binary value “1” is applied to a high, positive voltage (usually close to the voltage applied to the Vcc or Vdd pin of the chip). The binary value “0” is applied to low voltage (very close to the ground voltage level of the chip). This is generally considered intuitively obvious and can be easily remembered that a “1” input is the same as connecting an input to the power supply and a “0” input is the same as connecting an input to ground (resulting in “0” voltage). Similarly for outputs, when a “1” is output, you can assume that the chip can turn on a light. These conventions are true for virtually all basic electronic logic technologies; when you get into some advanced, very high speed logic, you may find that chips are designed with different operating conditions.

To simplify wiring diagrams, you will see many cases where the positive power connection and negative power connection are passed to terminal symbols to simplify the diagram and avoid the additional complexity of power and ground lines passing over the page and getting confused with the circuit “signal” lines.

When you are wondering how to connect an electronic device to its power supply, you can use Table 3-1 as a quick reference.

Table 3-1 Power wiring reference.

Positive (“+”) connection	**Negative (“−”) connection**	**Comments**
Red wire	Black wire	Wires connected to and between devices
Anode	Cathode	Diodes and capacitors
Vcc	Gnd	TTL
Vdd	Vss	CMOS

Basic Electronic Laws

Before starting to build your own digital electronics circuits, you should make sure that you are very familiar with the basic direct current electricity laws that govern how electricity flows through them. Don't worry if you have not yet been exposed to any direct current electrical theory, it's actually pretty simple and in the introduction to this chapter, I gave you a quick run down of how direct current circuits operate. I'm sure you were able to get through that without too many problems.

To make sure that you are clear on what *direct current* (also known as "DC") is, it consists of electricity running in a single direction without any changes. *Alternating current* ("AC") continuously changes from positive to negative (as shown in Fig. 3-2). AC is primarily used for high-power circuitry and not for any kind of digital electronics, except as something that is controlled by it. Digital electronics is powered by direct current, which consists of a fixed voltage which does not change level or polarity, as AC does.

As I indicated in the introduction, there are two components to electricity: *voltage* is the "pressure" applied to the electrons and *current* is the number of electrons that flow past a point or a set amount of time. I use the terms "pressure" and "flow" to help you visualize electricity moving in a wire as being the same as water flowing through a pipe. Using a water/pipe analogy can help you visualize how electricity moves and changes according to the conditions it is subjected to.

It should be obvious that the more pressure you apply to water in a pipe, the more water will pass through it. You can demonstrate this with a garden hose and a tap. By partially closing the tap, you are restricting the flow of the water coming from it, and the stream will not go very far from the end of the hose and very little water will flow out. When you completely open the tap, the water will spray out considerably further and a lot more water will be passing out the end of the hose. Instead of saying that you are closing the tap,

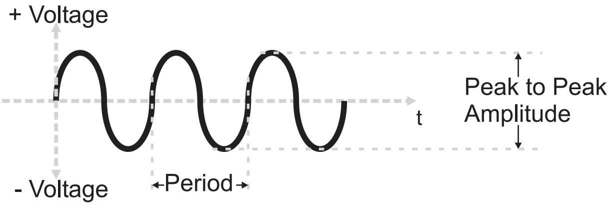

Fig. 3-2. Alternating current.

why don't you think of the closing tap as resisting the flow of water through the pipe and into the hose? This is exactly analogous to the *load* in a circuit converting electrical energy into something else. Electricity coming out of the load will be at a lower pressure (or voltage) than the electricity going into the load and the amount of current will be reduced as well.

When you visualized the pipe/tap/hose analogy, you probably considered that all the resistance in the circuit was provided by the tap – the pipe and the hose did not impede the water's flow in any way. This is also how we model how electricity flows in wires; the wires do not cause a drop in voltage and do not restrict the amount of current that is flowing in them. If you think about it for a moment, you will probably realize that this assumption means that the wires are ''superconductors''; any amount of electricity and at any voltage could be carried in the wires without any loss.

The wires that you use are certainly not superconductors, but the assumption that the wires do not impede the flow of electricity is a good one as their resistance in most circuits is usually negligible. By assuming that the wires are superconductors, you can apply some simple rules to understand the behavior of electricity in a circuit.

Going back to the original schematic diagram in this chapter (see Fig. 3-1), we can relate it to the pipe/tap/hose example of this section. The circuit's power supply is analogous to the pipe supplying water to the tap (which itself is analogous to the electrical circuit's load). The hose provides the same function as the wires bringing the electrical current back to the power supply.

In the pipe/tap/hose example, you should be able to visualize that the amount of water coming through the hose is dependent on how much the tap impedes the water flow from the pipe. It should be obvious that the less the tap impedes the water flow, the more water will come out the hose. Exactly the same thing happens in an electrical circuit; the ''load'' will impede or ''resist'' the flow of electricity through it and, in the process, take energy from the electricity to do something with it.

The most basic load that can be present in a circuit is known as the ''resistor'' (Fig. 3-3), which provides a specified amount of resistance,

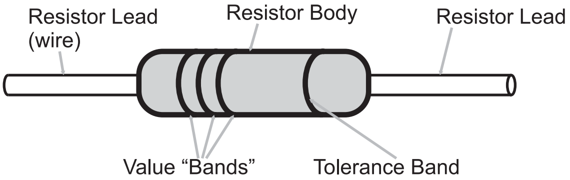

Fig. 3-3. Basic resistor.

measured in ‘‘ohms’’, to electricity. The ‘‘schematic symbol’’ is the jagged line you will see in various schematic diagrams in this book and in other sources. The schematic symbol is the graphic representation of the component and can be used along with the graphic symbol for a gate in a schematic diagram.

In traditional resistors, the amount of resistance is specified by a number of colored bands that are painted on its sides – the values specified by these bands are calculated using the formula below and the values for each of the colors listed in Table 3-2.

$$\texttt{Resistance} = ((\texttt{Band 1 Color Value} \times 10) + (\texttt{Band 2 Color Value})) \times 10^{\texttt{Band 3 Color Value}} \texttt{Ohms}$$

In the introduction to the chapter, I stated that power supplies provide electrons with a specific ‘‘pressure’’ called voltage. Knowing the voltage applied

Table 3-2 Resistor color code values.

Color	Band color value	Tolerance
Black	0	N/A
Brown	1	1%
Red	2	2%
Orange	3	N/A
Yellow	4	N/A
Green	5	0.5%
Blue	6	0.25%
Violet	7	0.1%
Gray	8	0.05%
White	9	N/A
Gold	N/A	5%
Silver	N/A	10%

to a load (or resistor), you can calculate the electrical current using Ohm's law which states:

> The voltage applied to a load is equal to the product of its resistance and the current passing through it.

This can be expressed mathematically as:

$$V = i \times R$$

where "V" is voltage, "R" is resistance and "i" is current. The letter "i" is used to represent current instead of the more obvious "C" because this character was already for specifying capacitance, as I will explain below. Voltage is measured in "volts", resistance in "ohms" and current in "amperes". For the work done in this book, you can assume that ohms have the units of volts/amperes and is given the symbol Ω; you can look up how these values are derived, but for now just take them for what I've presented here. With a bit of basic algebra, once you know two of the values used in Ohm's law, you can calculate the third.

Voltage, current, resistance, and, indeed, all the electrical values that you will see are part of the "SI" (Système Internationale), and its values are governed by SI standards. Each time a unit deviates by three orders of magnitude from the base value, the units are given a prefix that indicates the magnitude multiplier and these multipliers are listed in Table 3-3. For example, one thousandth of a volt is known as a "millivolt". The actual component values are normally given a single letter symbol that indicates its value. Most electronic devices, like resistors are given a two digit value that is multiplied by the power of ten which the symbol indicates. For example,

Table 3-3 Système Internationale magnitude of prefixes and symbols.

Power multiplier	**Prefix**	**Symbol**	**Power multiplier**	**Prefix**	**Symbol**
10^3	kilo	k	10^{-3}	milli	m
10^6	mega	M	10^{-6}	micro	μ
10^9	Giga	G	10^{-9}	nano	n
10^{12}	tera	T	10^{-12}	pico	p
10^{15}	peta	P	10^{-15}	femto	f

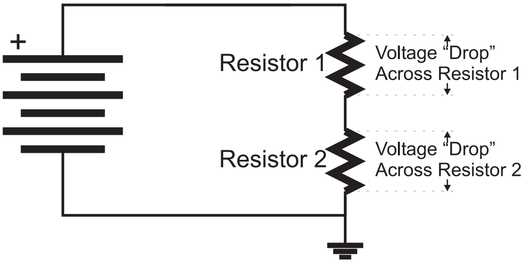
+
Resistor 1
Voltage "Drop" Across Resistor 1
Resistor 2
Voltage "Drop" Across Resistor 2

Fig. 3-4. Electrical circuit with two resistors in series.

thousands of units are given the prefix "k", so a resistor having a value of 10,000 ohms is usually referred to as having a value of "10 kohms", or most popularly "10 k".

Looking at more complex circuits, such as the two resistor "series" circuit shown in Fig. 3-4, you must remember that individual measurements must be taken across each resistor's two terminals; you do *NOT* make measurements relative to a common point. The reason for making this statement is to point out that the voltage across a resistor, which is also known as the "voltage drop", is dependent on the current flowing through it.

Using this knowledge, you can understand how electricity flows through the two series resistors in Fig. 3-4. The voltage applied to the circuit causes current to flow through both of the resistors and the amount of current is equal to the current passing through a single resistor value which is the sum of the two resistors. Knowing this current, and an individual resistor's value, you can calculate the voltage drop across each one. If you do the calculations, you will discover that the voltage drop across each resistor is equal to the applied voltage.

This may be a bit hard to understand, but go back to the pipe/tap/hose example and think about the situation where you had a pipe/tape/pipe/tap/hose. In this case, there would be a pressure drop across the first tap and then another pressure drop across the second tap. This is exactly what happens in Fig. 3-4: some voltage "drops" across Resistor 1 and the rest drops across Resistor 2. The amount of the drop across each resistor is proportional to its value relative to the total resistance in the circuit.

To demonstrate this, consider the case where Resistor 1 in Fig. 3-4 is 5 ohms and Resistor 2 is 8 ohms. Current has to flow through Resistor 1 followed by Resistor 2, which means that the total resistance it experiences is equivalent to the sum of the two resistances (13 ohms). The current through the two resistors could be calculated using Ohm's law, as voltage applied divided by Resistor 1 plus Resistor 2. The general formula for calculating equivalent the resistance of a series circuit is the sum of the resistances, which is written out as:

$$R_e = R_1 + R_2 + \ldots$$

Knowing the resistor values, the voltage drop across each resistor can be calculated as its fraction of the total resistance; the voltage across Resistor 1 would be 5/13ths of the applied voltage while the voltage across Resistor 2 would be 8/13ths of the applied voltage. Dividing the resistor values into the individual resistor voltage drops will yield the same current as dividing the applied voltage by the total resistance of the circuit.

Adding the two resistor voltage drops together, you will see that they total the applied voltage. This is a useful test to remember when you are checking your calculations, to make sure they are correct.

The properties of series resistance circuits are summed up quite well as Kirchoff's voltage law, which states that "the sum of the voltage drops in a series circuit is equivalent to the applied voltage and current is the same at all points in the circuit."

Along with being able to calculate the amount of current passing through a series resistor circuit and the voltage drop across each resistor, you can also calculate the voltage across each resistor in a parallel resistor circuit like Fig. 3-5 as well as the current through all the resistors. To do this, you have to remember Kirchoff's current law, which states that "the sum of the currents through each resistance is equivalent to the total current drawn by the circuit and the voltage drops across each resistor is the same as the applied voltage."

With each resistor in parallel, it should be fairly obvious that the voltage drop across each one is the same as the applied voltage, and the current flowing through each one can be calculated using Ohm's law. It should also

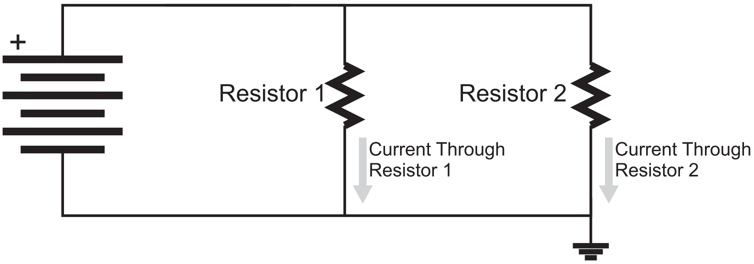

Fig. 3-5. Electrical circuit with two resistors in parallel.

be obvious that the current drawn from the power source is equivalent to the sum of the currents passing through each resistor.

If you were to calculate some different current values for different resistances, you would discover that the general formula for the equivalent resistance was:

$$R_e = 1/(1/R_1) + (1/R_2 + \ldots)$$

For the simple case of two resistors in parallel, the equivalent resistance can be expressed using the formula:

$$R_e = (R_1 \times R_2)/(R_1 + R_2)$$

Complex resistor circuits, made up of resistors wired in both series and parallel, like the one shown in Fig. 3-6, can be simplified to a single equivalent resistor by applying the series and resistor formulas that I have presented so far in this section. When doing this, I recommend first finding the equivalent to the series resistances and then the equivalent to the parallel resistances until you are left with one single equivalent resistance.

The last piece of basic electrical theory that I would like to leave you with is how to calculate the power dissipated by a resistor. When you took Newtonian physics, you were told that power was the product of the rate at which something was moving and the force applied to it. In electrical circuits, we have both these quantities, voltage being the force applied to the electrons and current being the rate of movement. To find the power being dissipated (in watts), you can use the simple formula:

$$P = V \times i$$

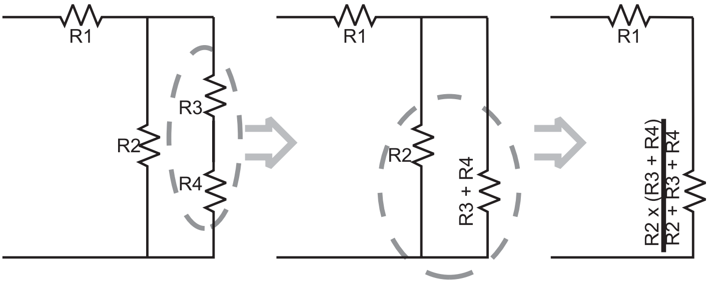

Step 1.
Combine Series
Resistances into a
Single Equivalent
Resistor

Step 2.
Combine Parallel
Resistances into a
Single Equivalent
Resistor

Repeat Step 1 and Step 2
until there is a Single
Resistance Equivalent
to the Resistance of the
"Network"

Fig. 3-6. Reducing the multiple resistor "network" into a single equivalent resistor.

or, if you don't know one of the two input quantities, you can apply Ohm's law and the formula becomes:

$$
\begin{aligned}
P &= V^2/R \\
&= i^2 \times R
\end{aligned}
$$

I must point out that when you are working with digital electronics, most currents in the circuits are measured somewhere between 100 μA to 20 mA. This seemingly small amount of current minimizes the amount of power that is dissipated (or used) in the digital electronic circuits. I'm pointing this out because if you were to get a book on basic electronics you would discover that the examples and questions will usually involve full amperes of current – not thousands or tens of thousands as I have noted here. The reason why basic electronics books work with full amps is because it is easier for students to do the calculations and they don't have to worry about working with different orders of magnitude.

So far in these few initial pages of this chapter, I have gone through the same amount of material that is presented in multiple courses in electrical theory. Much of the background material has been left out as well as derivations of the various formulas. For the purposes of working with digital electronics, you should be familiar with the following concepts:

1. Electricity flows like water in a closed circuit.
2. The amount of current flow in a circuit is proportional to the amount of resistance it encounters.
3. Voltage across a load or resistance is measured at its two terminals.
4. Voltage is current times resistance (Ohm's law).
5. Power is simply voltage times current in a DC circuit.

The other rules are derivations of these basic concepts and while I don't recommend trying to work them out in an exam, what you do remember can be checked against the basic concepts listed above.

Capacitors

When working with digital electronic circuits, it is very important for you to understand the purpose and operation of the capacitor. Many people shy away from working at understanding the role of capacitors in digital electronics because the formulas that define their response to an applied voltage do not seem to be intuitive and many of them are quite complex. Further reducing the attractiveness of understanding capacitors is that they do not seem to be a basic component of digital electronics, and when they are used their value and wiring seems to be simply specified by a datasheet or an application note. I must confess that these criteria used to apply to me and I never understood the importance of capacitors in digital electronics until I was reviewing failure analysis of a 4 MB memory chip. As I will show, a dynamic RAM memory element (along with a MOSFET transistor) is essentially a capacitor, and the failure analysis of the chips showed how the differences in these capacitors affected their operation. One of the major conclusions of the failure analysis was that the memory chip wasn't so much a digital electronic device as a massive array of four million capacitors. This example is meant to show the importance of understanding the operation of capacitors and how they influence digital electronic circuits – being comfortable with the information in this section is more than good enough to use and successfully specify capacitors in digital electronic circuits.

The capacitor itself is a very simple energy storage device; two metal plates (as shown in the leftmost capacitor symbol in Fig. 3-7) are physically separated by a "dielectric" which prevents current from flowing between them. The dielectric is an insulator ("dielectric" is a synonym for "insulator") material which enhances the metal plates' ability to store an electric charge.

The capacitor is specified by the amount of charge it is able to store. The amount of charge stored in a capacitor (which has the symbol "C") is measured in "farads" which are "coulombs" per volt. One coulomb of electrons is a very large number (roughly 6.2×10^{18}) and you will find that for the most part you will only be working with capacitors that can store a very small fraction of a coulomb of electrons.

Knowing that farads are in the units of coulombs per volt, you can find the amount of charge (which has the symbol "Q") in a capacitor by using the formula:

$$Q = C \times V$$

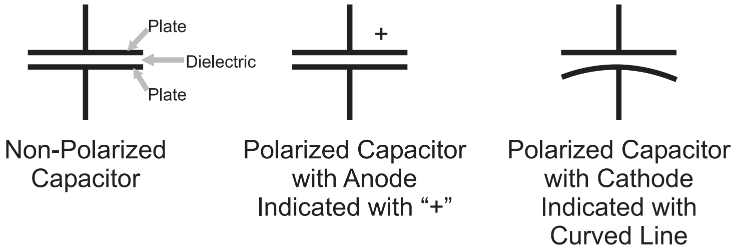
Plate
Dielectric
Plate
Non-Polarized Capacitor
+
Polarized Capacitor with Anode Indicated with "+"
Polarized Capacitor with Cathode Indicated with Curved Line

Fig. 3-7. Capacitor symbol.

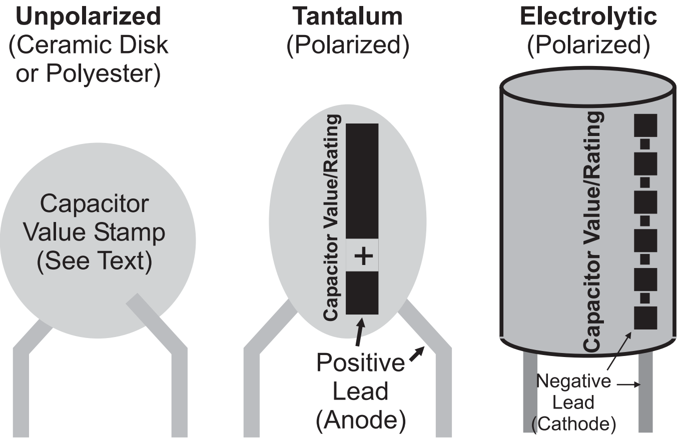
Unpolarized
(Ceramic Disk
or Polyester)
Capacitor
Value Stamp
(See Text)
Tantalum
(Polarized)
Capacitor Value/Rating
+
Positive
Lead
(Anode)
Electrolytic
(Polarized)
Capacitor Value/Rating
Negative
Lead
(Cathode)

Fig. 3-8. Capacitor appearance and markings.

The fraction of a coulomb that is stored in a capacitor is so small, that the most popularly used capacitors are rated in millionth's ("microfarads" or "µF") or trillionth's ("picofarads" or "pF") of farads. Microfarads are commonly referred to as "mikes" and picofarads are often known by the term "puffs". Using standard materials (such as mica, polyester and ceramics), it is possible to build capacitors of a reasonable size of 1 microfarad (one millionth of a farad) but more exotic materials are required for larger value capacitors. For larger capacitors, the dielectric is often a liquid and the capacitor must be wired according to parameter markings stamped on it, as I have indicated in Fig. 3-8. These are known as "polarized" capacitors and either a "+" marking or a curved plate (as shown in Fig. 3-7) is used to indicate how the capacitor is wired in the schematic. Like other polarized components, the positive connection is called an "anode" and the negative a "cathode". Along with the markings, you should remember that the anode of a polarized two lead component is always longer than the cathode. The different lead lengths allow automated assembly equipment to distinguish between the two leads and determine the component's polarity.

Capacitors have two primary purposes in digital electronic circuits. The first is as a voltage "filter" (Fig. 3-9), reducing "spikes" and other problems on a wire carrying current. This use is similar to the use of a water tower in a city; the water tower is filled due to the pressure of the water being pumped into the community. Water is continually pumped to both houses and the water tower, but in times of high usage (like during the day when people are watering their lawns and washing their cars), water from the tower supplements the pumped water to keep the pressure constant. During the

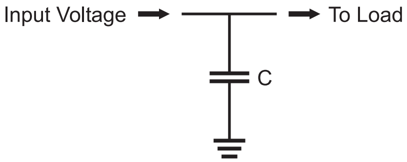

Fig. 3-9. Power supply filter using capacitor.

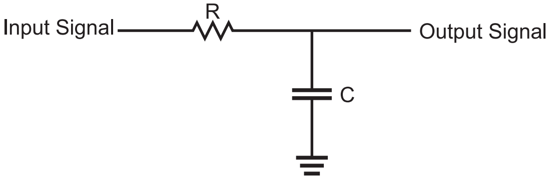

Fig. 3-10. Low-pass filter built from resistor and capacitor.

night, when few people are using water, the pumped water is stored in the water tower, in preparation for the next day's requirements.

When you look at digital electronic circuits, you will see two types of capacitors used for power filtering. At the connectors to the power supply, you will see a high value capacitor (10 µF or more) filtering out any "ripples" or "spikes" from the incoming power. "Decoupling" capacitors of 0.047 µF to 0.1 µF are placed close to the digital electronic chips to eliminate small spikes caused when the gates within the chips change state.

Large capacitors will filter out low-frequency (long-duration) problems on the power line while the small capacitors will filter out high-frequency (short-duration) spikes on the power line. The combination of the two will keep the power line "clean" and constant, regardless of the changes in current demand from the chips in the circuit.

The capacitor's ability to filter signals is based on its ability to accept or lose charge when the voltage across it changes. This capability allows voltage signals to be transformed using nothing more than a resistor and a capacitor, as in the "low-pass filter" shown in Fig. 3-10. This circuit is known as a low-pass filter because it will pass low-frequency alternating current signals more readily than high-frequency alternating current signals.

In digital electronics, we are not so much concerned with how a capacitor affects an alternating current as how it affects a changing direct current. Figure 3-11 shows the response, across Fig. 3-10's low-pass filter's capacitor and resistor, to a digital signal that starts off with a low voltage "steps" up to "V" and then has a falling step back to 0 V.

In Fig. 3-11, I have listed formulas defining the voltage response across the resistor and capacitor to the rising and falling step inputs. These formulas are

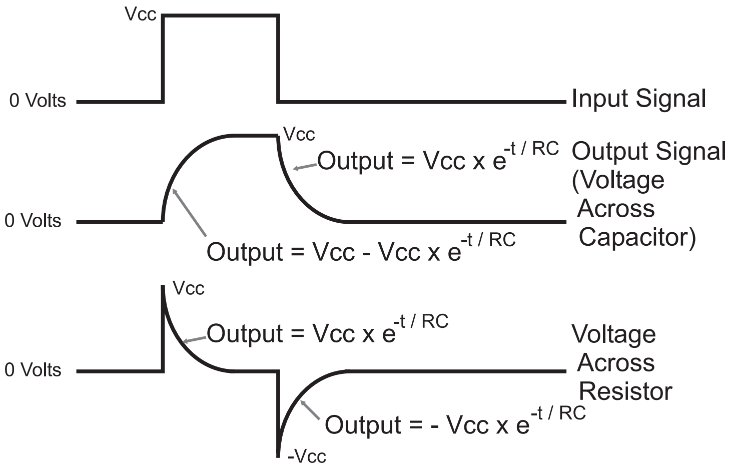

Fig. 3-11. Resistor/capacitor circuit response to changing input.

found within introductory college electricity courses by knowing that the voltage across the capacitor can be defined by using the formula:

$$V_C(t) = Q_C(t)/C$$

which simply states that the voltage across a capacitor at some point in time is a function of the charge within the capacitor at that point of time. The charge within the capacitor is supplied by the current passing through the resistor and the resistor limits the amount of current that can pass through it. As the voltage in the capacitor increases, the voltage across the resistor falls and as the voltage across the resistor falls, the amount of current that is available to charge the capacitor falls. It is a good exercise in calculus to derive these formulas, but understanding how this derivation works is not necessary for working with digital electronics.

There are two things I want to bring out from the discussion of low-pass filters. The first is that the response of the low-pass filter is a function of the product of the resistance and capacitance in the circuit. This product is known as the “RC time constant” and is given the Greek letter “tau” (τ) as its symbol. Looking at the formulas, you should see that by increasing the value of τ (either by using a larger value resistor or capacitor) the response time of the low-pass filter is increased.

This has two ramifications for digital electronics. The first should be obvious: to minimize the time signals take to pass between gates, the resistance and capacitance of the connection should be minimized. The second is more subtle: the resistor–capacitor response can be used to delay a signal in a circuit. This second issue with resistor–capacitor circuits is actually very useful in digital electronics for a number of different applications that I will discuss later in the book.

This is a very short introduction to capacitors and their operation in (digital) electronic circuits. Before going on, I would like to reinforce what I've said about their importance and recommend that you follow up this section's material by working through a book devoted to analog electronics.

Semiconductor Operation

Over the past 100 years, we have refined our ability to control the electrical properties of materials in ways that have made radios, TVs and, of course, digital electronic circuits possible. These materials have the ability to change their conductance, allowing current to pass through them under varying conditions. This ability to change from being an insulator to a conductor has resulted in these materials being called ''semiconductors'', and without them many of the basic devices we take for granted would be impossible.

The most basic electronic semiconductor device is the ''diode''. The electrical symbol and a sketch of the actual part is shown in Fig. 3-12. Diodes are a ''one-way'' switch for electricity; current will pass easily in one direction and not in the other. If you were to cut a silicon diode in half and look at its operation at a molecular level, you would see that one-half of the silicon was ''doped'' (infused with atoms) with an element which can easily give up electrons, which is known as an ''N-type'' semiconductor. On the other side of the diode, the silicon has been doped with an element that can easily accept electrons, a ''P-type'' semiconductor.

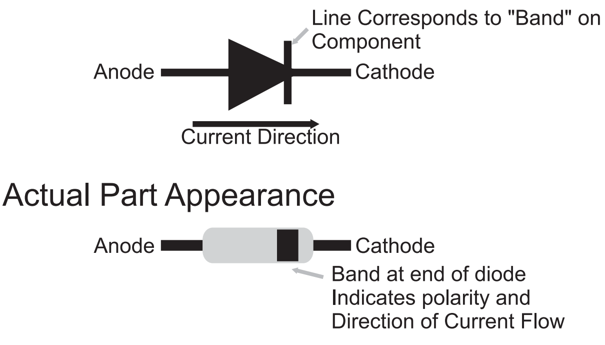
Schematic Symbol
Line Corresponds to "Band" on Component
Anode
Cathode
Current Direction
Actual Part Appearance
Anode
Cathode
Band at end of diode Indicates polarity and Direction of Current Flow

Fig. 3-12. Diode symbol.

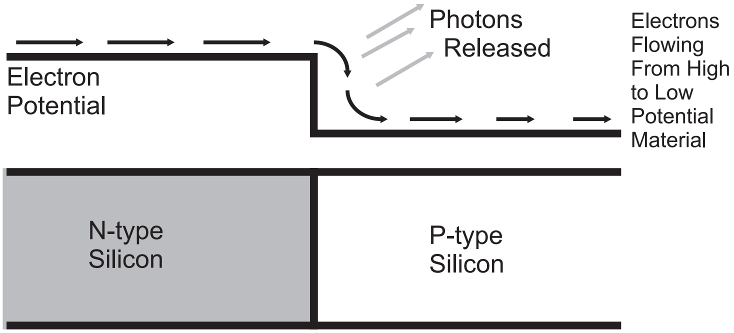

Fig. 3-13. Diode operation.

When a voltage is applied to the diode, causing electrons to travel from the atoms of the N-type semiconductor to the atoms of the P-type, the electrons "fall" in energy from their orbits in the N-type to the accepting orbit spaces in the P-type, as shown in Fig. 3-13. This drop in energy by the electron is accompanied by a release in energy by the atoms in the form of photons. The "quanta" of photon energy released is specific to the materials used in the diode – for silicon diodes, the photons are in the far infrared.

The voltage polarity applied to the diode is known as "bias". When the voltage is applied in the direction the diode conducts in, it is known as "forward biased". As you might expect, when the voltage is applied in the direction the diode blocks current flow, it is known as "reverse biased". This is an important point to remember, both for communicating with others about your designs and for understanding the operation of transistors, as explained below.

To keep the thermodynamic books balanced, the release in energy in terms of photons is accompanied by a corresponding voltage drop across the diode. For silicon diodes, this drop is normally 0.7 volts. The power equation I gave earlier ($P = V \times i$) applies to diodes. When large currents are passed through the diode and this is multiplied by 0.7 V, quite a bit of power can be dissipated within the diode.

If voltage is applied in the opposite direction (i.e. injecting electrons into the P-type side of the diode), the electrons normally do not have enough energy to rise up the slope and leave the orbits of the P-type atoms and enter the electron-filled orbits of the N-type atoms. If enough voltage is applied, the diode will "break down" and electrons will jump up the energy slope. The break down voltage for a typical silicon diode is 100 V or more – it is quite substantial.

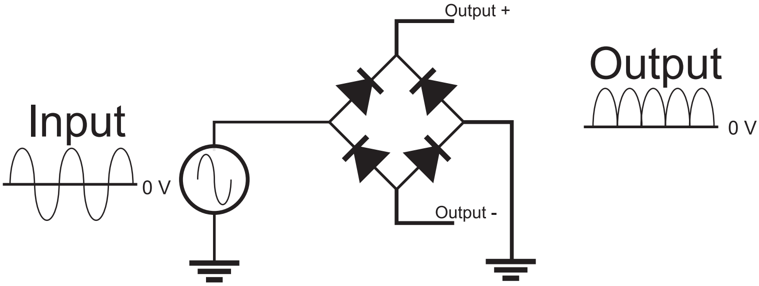

Fig. 3-14. "Full wave rectifier" using four diodes.

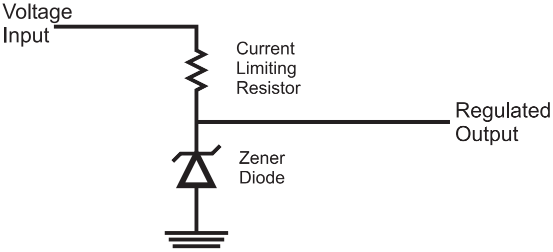

Fig. 3-15. Zener diode voltage regulator.

A typical use for a diode is to "rectify" AC to DC, as shown in Fig. 3-14, in which a positive and negative alternating current is converted using the four diodes to a "lobed" positive voltage signal, which can be filtered using capacitors, as discussed in the previous section.

Along with the simple silicon diode discussed above, there are two other types of diodes that you should be aware of. The first is the "Zener" diode which will break down at a low, predetermined voltage. The typical uses for the Zener diode is for accurate voltage references (Zener diodes are typically built with 1% tolerances) or for low-current power supplies like the one shown in Fig. 3-15. The symbol for the Zener diode is the diode symbol with the bent current bar shown in Fig. 3-15.

Building a power supply using this circuit is actually quite simple: the Zener diode's break down voltage rating will be the "regulated output" and the "voltage input" should be something greater than it. The value of the current limiting resistor is specified by the formula

$$\mathtt{R}_{\mathtt{iLimit}} = (\mathtt{V}_{\mathtt{in}} - \mathtt{V}_{\mathtt{Zener}})/\mathtt{i}_{\mathtt{app}}$$

where "i_{app}" is the current expected to be drawn (plus a couple of tens of percent margin). The power rating of the Zener diode should take into account the power dissipated if i_{app} was passing through it.

As I will discuss later in this chapter, there are a lot of inexpensive power regulators that are a lot more efficient than the Zener diode one shown in Fig. 3-15. If you do the math for a typical application (say 9 volts in, 5.1 volt Zener diode and a 20 mA current draw), you will find that at best it is 60% efficient (which is to say 60% of the power drawn by the Zener regulator circuit and the application is passed to the application, and can often be as low as 25%). The reason for using the Zener diode regulator is its low cost, very small form factor and extreme robustness. Most practical applications will use a linear regulator chip.

The other type of diode that I want to mention in this section is one that you are already very familiar with – the light-emitting diode or LED. As its name implies, this diode emits light (like a light bulb) when a current passes through it. In Fig. 3-16, note that the LED symbol is the same as the diode's symbol, but with light rays coming from it. The most common package for the LED is also shown in Fig. 3-16 and it consists of a rounded cylinder (somewhat like "R2D2" from Star Wars) with a raised edge at its base with one side flattened to indicate the LED's cathode (negative voltage connection).

There are a few points that you should be aware of with regard to LEDs. In the past few years, LEDs producing virtually every color of the rainbow (including white) have become available. I must point out that LEDs can only produce one color because of the chemistry of the semiconductors used to build them. You may see advertisements for two or three color LEDs, but these devices consist of two or three LEDs placed in the same plastic package and wired so that when current passes through its pins in a certain direction, a specific LED turns on.

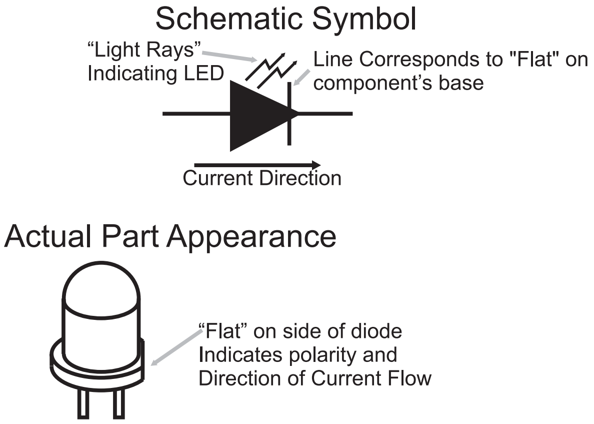
Schematic Symbol
"Light Rays" Indicating LED
Line Corresponds to "Flat" on component's base
Current Direction
Actual Part Appearance
"Flat" on side of diode Indicates polarity and Direction of Current Flow

Fig. 3-16. LED symbol.

The brightness of a LED cannot be controlled reliably by varying the current passing through it, as you would with a light bulb. LEDs are designed to provide a set amount of light with current usually in the range of 5 to 10 mA. Reducing the current below 5 mA may dim its output or it may turn it off completely. A much better way to dim a LED is to use "pulse wave modulation" (PWM), in which the current being passed to the LED is turned on, and faster than the human eye can perceive, with varying amounts of on and off time to set the LED's brightness. I will discuss PWMs later in the book.

Finally, when I first introduced diodes, I noted that silicon diodes output photons of light in the far infrared and have a 0.7 volt drop when current passes through them. To produce visible light, LEDs are not made out of silicon, they are made from other semiconductor materials in which the energy drop from the N-type semiconductor to the P-type semiconductor produces light in the visible spectrum. This change in material means that LEDs do not have silicon's 0.7 V drop; instead, they typically have a 2.0 V drop. This is an important point because it will affect the value of the current limiting resistor that you put in series to make sure the LED's current limit rating is not exceeded or that it does not allow too much current in the circuit to pass through it, resulting in an unnecessary current drain.

It is always a source of amazement to me how many people do not understand how transistors work. For the rest of this section, I will introduce you to the two most common types of transistors and explain how they work as well as what applications they are best suited for. Understanding the characteristics of the two types of transistors is critical to understanding how digital logic is implemented and how you can interface it to different technologies.

As I explain the operation of the "bipolar" transistor, I will endeavor to keep to the "high level" and avoid trying to explain transistor operation using tools like the "small signal model", which is intimidating and obfuscates the actual operation of the device. Instead, I want to introduce you straight to the "NPN bipolar transistor" by its symbol and typical package and pinout for a small scale (low-power) device in Fig. 3-17.

As you have probably heard, a bipolar transistor can be considered a simple switch or a voltage amplifier, but you are probably mistaken on how it is controlled and how it actually works. The transistor is not voltage controlled (as you may have been led to expect); it is actually current controlled. The amount of current passing through the "base" to the "emitter" controls the amount of current that can pass from the "collector" to the emitter. The amount of current that can be passed through the collector is a multiple (called "beta" and given the symbol "β" or h_{FE}) of the

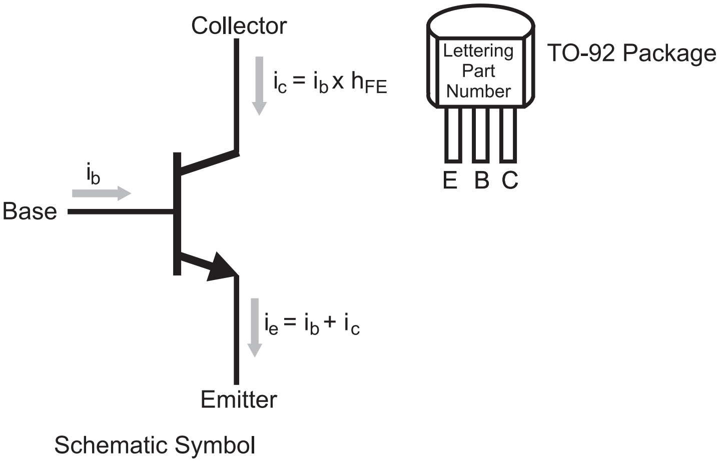

Fig. 3-17. NPN transistor symbol with parameters.

current flowing through the base; the bipolar transistor is actually an amplifier – a small amount of current allows a greater amount to flow. The simple formulas for the relationship between the base and collector currents are listed in Fig. 3-17.

I must point out that these formulas apply while the maximum collector current is in the ‘‘small signal’’ or ‘‘linear’’ operating range. As a physical device, a transistor can only allow so much current to flow through it; as it reaches this limit, increases in the transistor’s base current will not result in a proportional increase in collector current. This operating region is known as the ‘‘non-linear’’ or ‘‘saturation’’ region and what happens in this situation can be easily understood by looking at what happens in a cross section of a transistor (Fig. 3-18).

A bipolar transistor consists of a P-type semiconductor sandwiched between two N-type semiconductors. This structure forms a reverse biased diode and no current can flow through it. With no current being injected into the NPN bipolar transistor, the P-type semiconductor is known as the ‘‘depletion region’’ because it does not have any electrons. When current is passed to the device, electrons are drawn through the P-type semiconductor via the emitter N-type semiconductor. As electrons are drawn into the P-type semiconductor, the properties of the P-type semiconductor change and take on the characteristics of the N-type semiconductors surrounding it and becomes known as the ‘‘conduction region’’. The more electrons that are drawn from the P-type semiconductor, the larger the conduction region bridging the two pieces of N-type semiconductor and the greater amount of current that can pass from the collector to the emitter. As more electrons are

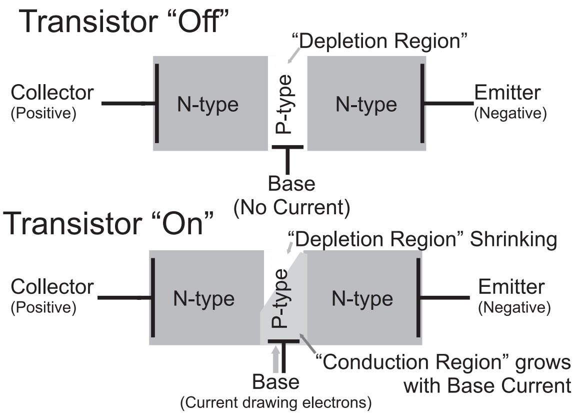
Transistor "Off"
"Depletion Region"
Collector
(Positive)
N-type
P-type
N-type
Emitter
(Negative)
Base
(No Current)
Transistor "On"
"Depletion Region" Shrinking
Collector
(Positive)
N-type
P-type
N-type
Emitter
(Negative)
"Conduction Region" grows with Base Current
Base
(Current drawing electrons)

Fig. 3-18. NPN transistor operation.

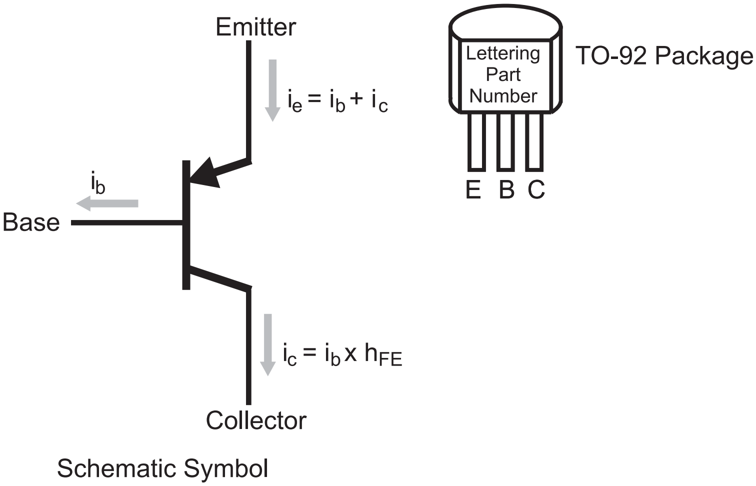
Emitter
$i_e = i_b + i_c$
i_b
Base
$i_c = i_b \times h_{FE}$
Collector
Schematic Symbol
Lettering
Part
Number
TO-92 Package
E B C

Fig. 3-19. PNP transistor symbol with parameters.

drawn from the P-type semiconductor, the conduction region grows until the entire P-type semiconductor of the transistor becomes "saturated".

The PNP bipolar transistor (Fig. 3-19) operates in the complete opposite way to the NPN transistor. It is built from an N-type semiconductor between two P-type semiconductors and to create a conduction region, electrons are injected into the base instead of being withdrawn, as in the case of the NPN bipolar transistor. As in the NPN bipolar transistor, the amount of collector current is a multiple of the base current (and that multiple is also called β or h_{FE}).

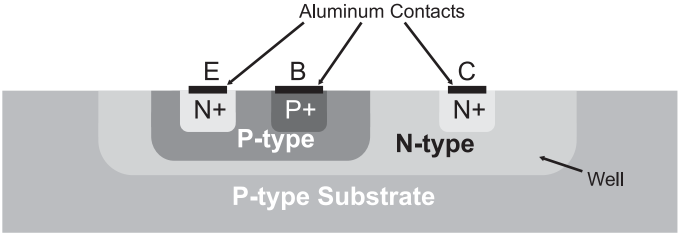

Fig. 3-20. Side view cross section of NPN bipolar transistor in an integrated circuit.

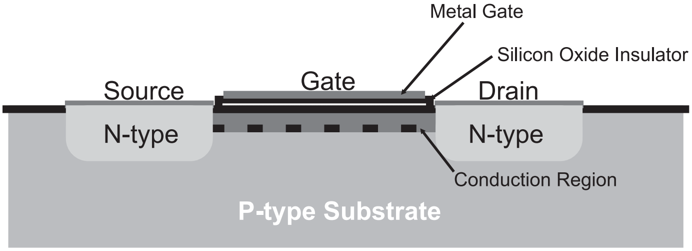

Fig. 3-21. N-channel MOSFET side view showing features during operation.

Bipolar transistor h_{FE} values can range anywhere from 50 to 500 and the amount of collector current they can handle ranges from a few tens of milliamps to tens of amps. As well as discrete (single) devices being inexpensive, they respond to changes in inputs in extremely short time intervals. You may think they are perfect for use in digital electronics, but they have two faults that make them less than desirable. First, the base current is actually a source of power dissipation in the device, which is usually not an issue when single transistors are used, but is of major concern when thousands or millions are used together in a highly complex digital electronic system.

Secondly, they take up a lot of chip "real estate" and are very expensive to manufacture. Figure 3.20 shows the side view of an NPN bipolar transistor built on a silicon chip. Instead of butting together different types of semiconductor, it is manufactured as a series of "wells", which are doped with the chemicals to produce the desired type of semiconductor by repeated operations. As many as 35 process steps are required to produce a bipolar transistor.

The N-channel enhancement "metal oxide silicon field effect transistor" (MOSFET) does not have these faults – it is built using a much simpler process (the side view of the transistor is shown in Fig. 3-21) that only requires one doping of the base silicon along with the same bonding of aluminum contacts as the bipolar transistor. N-channel MOSFETs (as they are most popularly known) require nine manufacturing processes and take a fraction of the chip real estate used by bipolar transistors.

The N-channel MOSFET is not a current-controlled device, like the bipolar transistor, but a voltage-controlled one. To ‘‘turn on’’ the MOSFET (allow current to flow from the ‘‘source’’ to the ‘‘drain’’ pins), a voltage is applied to the ‘‘gate’’. The gate is a metal plate separated from the P-type silicon semiconductor substrate by a layer of silicon dioxide (most popularly known as ‘‘glass’’). When there is no voltage applied to the gate, the P-type silicon substrate forms a reverse biased diode and does not allow current flow from the source to the drain. When a positive voltage is applied to the gate of the N-channel MOSFET, electrons are drawn to the substrate immediately beneath it, forming a temporary N-type semiconductor ‘‘conduction region’’, which provides a low-resistance path from the source to the drain. MOSFET transistors are normally characterized by the amount of current that can pass from the source to the drain along with the resistance of the source/drain current path.

The symbol for the N-channel MOSFET, along with its complementary device, the P-channel MOSFET are shown in Fig. 3-22. The P-channel MOSFET creates a conduction region when a negative voltage is applied to its gate. MOSFET transistors come in a variety of packages and some can handle tens of amps of current, but they tend to be very expensive.

MOSFETs do not have the issues of bipolar transistors; their gate widths (the measurement used to characterize the size of MOSFET devices) are, at the time of this writing, as small as 57 nm in high-performance microprocessors and memory chips. The voltage-controlled operation of MOSFETs eliminates the wasted current and power of the bipolar transistor’s base, but while MOSFETs do not have the disadvantages of bipolar transistors, they do not have their advantages.

MOSFET transistors do not have a small signal/linear operating region; they tend to change from completely off to completely on (conducting) with a

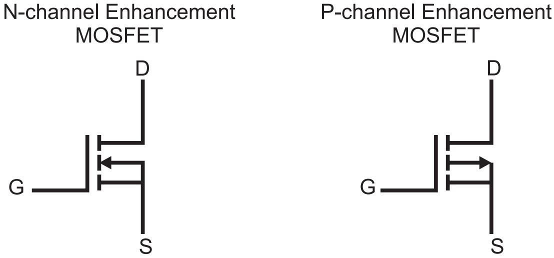

Fig. 3-22. MOSFET schematic symbols.

very small intermediate range. MOSFETs also tend to operate at slower speeds than bipolar devices because the gates become capacitors and "slow down" the signals, as I showed in the previous section. This point has become somewhat moot as modern MOSFET designs are continually increasing in speed, providing us with extremely high-speed PCs and other electronic devices. Finally, it is difficult to manufacture MOSFETs with high current capabilities; while high current MOSFETs are available, they are surprisingly expensive.

The characteristics of the two types of transistors give way to the conclusion that bipolar transistors are best suited to situations where a few high current devices are required. MOSFET transistors are best suited for applications where large numbers of transistors are placed on a single chip.

Today, for the most part, digital electronic designs follow these guidelines, but we are left with an interesting legacy. Despite being much simpler structurally and cheaper to manufacture, MOSFET transistors were only perfected in the late 1960s, whereas bipolar technology had already been around for 20 years and it was able to become entrenched as the basis for many digital electronic devices and chips. For this reason, you must be cognizant of the operating characteristics of bipolar transistors as well as those of MOSFET transistors. In the next section, many of these differences will become apparent.

Logic Gate Input and Output

If you have worked with digital electronics before, you probably have made a few assumptions about how the circuitry works and how you can demonstrate how digital electronic devices work. Chances are many of these assumptions are with regard to how gate and chip inputs and outputs work as well as how to properly interface them together and to different electronic devices. These assumptions are generally made on the evidence of by what somebody has seen with a voltmeter or logic probe and do not look at the underlying circuitry and how it works. In this section, I will give you a detailed introduction to the input and output pins on digital electronics and how they should be wired.

When we talk about digital electronics, we should identify the different technologies used. "Transistor to transistor logic" (TTL) is based on NPN bipolar transistors. TTL chips have the part number prefix "74" (i.e. a chip with four, two input NAND gates known as the "7400"). There are actually quite a few different technology chip families based on the 74xx "standard" pinout and operation and the technology is indicated by letter codes following the "74"; a chip marked with "74LS00" is a low-power, Shotkey four two-input NAND gate chips. Many of these technologies used with the 7400 series of chips are based on bipolar transistors, but some are based on MOSFET technology. These MOSFET technology based chips have the 74 prefix and a technology letter code containing a "C" (i.e. "C", "HC", "HCT"). Along with being used in 7400 series form factors, MOSFET devices are used in the "4000" series of logic chips. Understanding which type of transistor is used in a logic chip is critical to being able to successfully interface it to other chips or input/output devices.

When the term "TTL" is used, it is referring to bipolar transistor logic in the 7400 series. "CMOS" indicates MOSFET transistor logic used in the 74C00 and 4000 chip logic series.

Probably the biggest erroneous assumption that people have about digital logic is that TTL circuitry is voltage controlled. In the previous section, I emphasized the notion that bipolar transistors are *current* controlled and not voltage controlled. I'm sure that many people will argue with me and say that when they put a voltage meter to the input of a TTL gate, they saw a high voltage when a "1" was being input and a low voltage when a "0" was input. I won't argue with what they have seen; although I will state that the conclusion that TTL logic is voltage controlled made from these observations is incorrect.

The standard TTL input consists of an NPN bipolar transistor wired in the unusual configuration shown in Fig. 3-23. On the left side of this diagram, I have drawn a two input TTL gate which is implemented with a two emitter

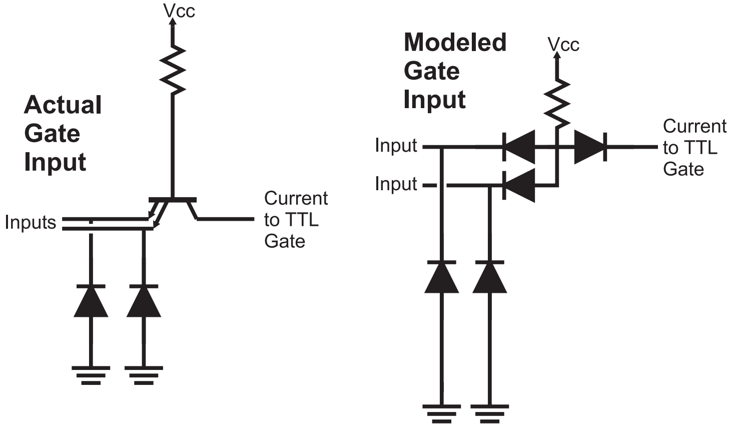

Fig. 3-23. Actual and model TTL input circuits.

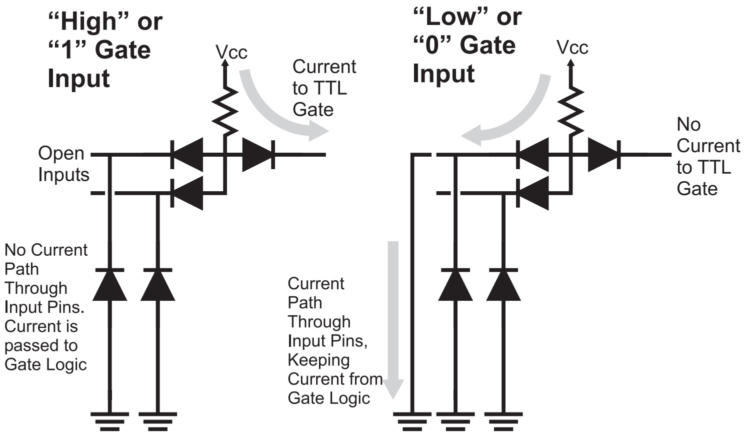
"High" or "1" Gate Input
Vcc
Current to TTL Gate
Open Inputs
No Current Path Through Input Pins. Current is passed to Gate Logic
"Low" or "0" Gate Input
Vcc
No Current to TTL Gate
Current Path Through Input Pins, Keeping Current from Gate Logic

Fig. 3-24. Current control operation of TTL inputs.

NPN transistor – as unusual as this type of transistor sounds, they really do exist. To understand how the input works, I replaced the two emitter NPN transistor with the three diode equivalent "model" on the right side of Fig. 3-23.

Normally, an NPN transistor passes current from its base to the emitter, but when wired in the TTL input configuration, the base current does not have a path through the transistor's emitters and passes through the transistor's collector to the gate logic. Figure 3.24 shows this situation along the other case where one of the input transistor's emitter's is tied to ground and the base current passes through the emitter and not the collector. The logic connected to the input NPN transistor's collector responds depending on whether or not current is available from the collector.

Obviously a simple switch, connected to ground, will allow current to pass through the emitter but you are probably wondering how other logic devices can control this device. A typical logic device output looks like Fig. 3-25 and consists of two transistors: one that will connect the output to the device power and one that will connect the output to the device ground. This transistor path to ground will provide the emitter current path of the chip. When the output is a high voltage (the top transistor is on and the bottom one is off), no current will flow into the TTL input gate because of the reverse diode nature of the emitter input pin.

The TTL output shown in Fig. 3-25 is known as a "totem pole" output because of its resemblance to its namesake. If you were to connect a totem pole output to a TTL input and measured the voltage at the input or output pins, you would see a high voltage, which the gate connected to the input

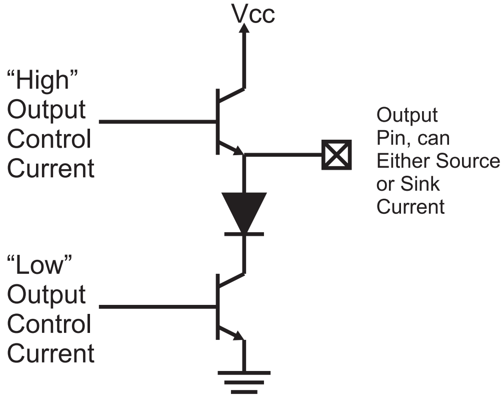
Vcc
"High" Output Control Current
Output Pin, can Either Source or Sink Current
"Low" Output Control Current

Fig. 3-25. TTL "totem pole" output.

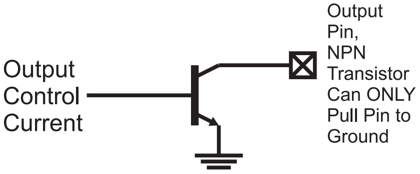
Output Control Current
Output Pin, NPN Transistor Can ONLY Pull Pin to Ground

Fig. 3-26. TTL open collector output.

would respond to as a "1". When a low voltage is output, the TTL gate will respond as if a "0" was input. What you are not measuring is the current flow between the two pins.

There are two terms used in Fig. 3-25 that I should explain. When a transistor is connected to the power supply of a chip and is turned on, it is said to be "sourcing" current. When a transistor is connected to ground and is turned on, the transistor is said to be a current "sink". I will use these terms throughout the book and you will see in other books and references any time a device is either supplying ("sourcing") or taking away ("sinking") current.

There is another type of output which does not source any current and is known as the open collector output (Fig. 3-26). This output typically has two uses. The first is it can pull down voltages which are greater than the positive voltage applied to the chip. Normally these voltages are less than 15 V and can only source 10 to 20 mA. For higher currents and voltages, discrete transistors must be used.

By not sourcing any current, these outputs can be "ganged" together in parallel, as I have shown in Fig. 3-27. This circuit is known as a "dotted AND" because it only outputs a 1 if all the outputs are "high" and each

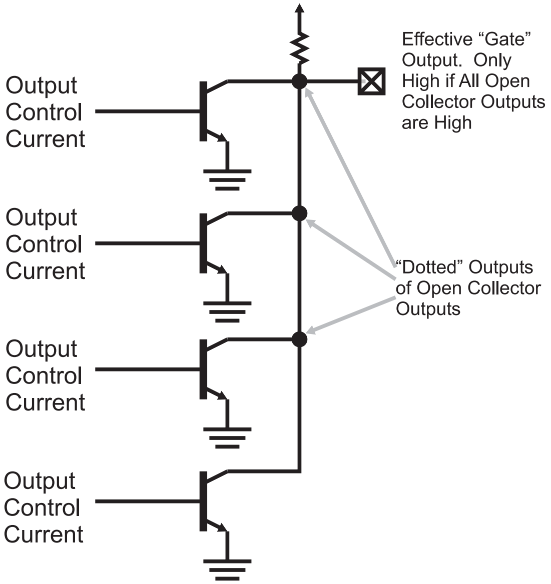
Output
Control
Current
Output
Control
Current
Output
Control
Current
Output
Control
Current
Effective "Gate" Output. Only High if All Open Collector Outputs are High
"Dotted" Outputs of Open Collector Outputs

Fig. 3-27. Multiple ‘‘open collector’’ outputs combined to form a ‘‘dotted AND’’ gate.

transistor is ‘‘off’’ and not pulling the common output line to ground. Note that there must be a pull up resistor connected to the output to provide a high-voltage, low-current source. Dotted AND gates are useful in a variety of different situations ranging from circuits where an arbitrary number of outputs can control one line or where digital outputs and buttons are combined. (I will discuss this in more detail later in the book.)

Totem pole outputs are the recommended default gate output because you can easily check voltage levels between intermediate gates in a logic string. As I will show later in this chapter, you cannot use a voltmeter or logic probe to check the logic levels if a TTL gate is driven by an open collector output. Along with this, a CMOS input is connected to an open collector (or open drain, as I will discuss below) output. Then there will be no high voltage for the gate to operate. The only cases where an open collector/open drain output should be used is when you are wiring a dotted AND gate or are switching an input that is operating at a voltage different from the gate's power.

TTL output pins are internally limited to only sink or source around 20 mA of current, which limits the number of inputs that it can drive. If you were to do the math, you would discover that when a TTL input is pulled low, 1.075 mA of current is passed through the output pin (this was found by assuming the base/emitter voltage of a transistor is 0.7 volts and the current limiting resistor connected to the input transistor's base is 4 k, which is typical for TTL inputs.

Along with the totem pole and the open collector outputs, there is also the "tri-state driver" output, which cannot only source or sink current but can be turned "off" to electrically isolate itself from the circuit that it is connected to. I will discuss tri-state drivers later in the book, when I present busses and multiple devices on the same line.

Knowing that each TTL input requires a current sink of just over 1 mA and most TTL outputs can sink up 20 mA, you might expect the maximum number of TTL inputs driven by a single output (which is called "fanout") to be 18 or 19. The actual maximum fanout is 8 to ensure that there is a comfortable margin in the output to be able to pull down each output in a timely manner. Practically, I would recommend that you try to keep the number of inputs driven by an output to two and never exceed four. Some different technologies that you work with, do not have the same electrical drive characteristics and may not be designed to pull down eight inputs of another technology; so, to be on the safe side, always be very conservative with the number of inputs you drive with a single output.

Re-reading the last sentence of the previous paragraph, you might wonder if any potential low-drive situations could be improved by wiring multiple outputs together. This must be avoided because of the danger that the gates will switch at different times, resulting in large currents passing through the gate output circuitry, and not through the net the outputs are connected to.

The CMOS logic gate input (Fig. 3-28) is quite a bit simpler than the TTL gate input and much easier to understand. The CMOS input and, as I will explain, the output, consist of a balanced P-channel MOSFET and an N-channel MOSFET wired as a very high gain amplifier. The slightest positive or negative voltage applied to this input circuit will cause the

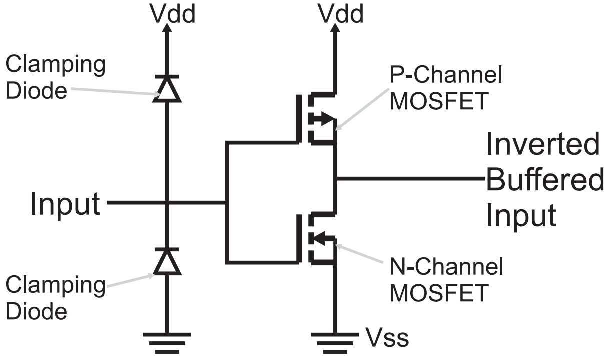

Fig. 3-28. Basic CMOS input/output circuit.

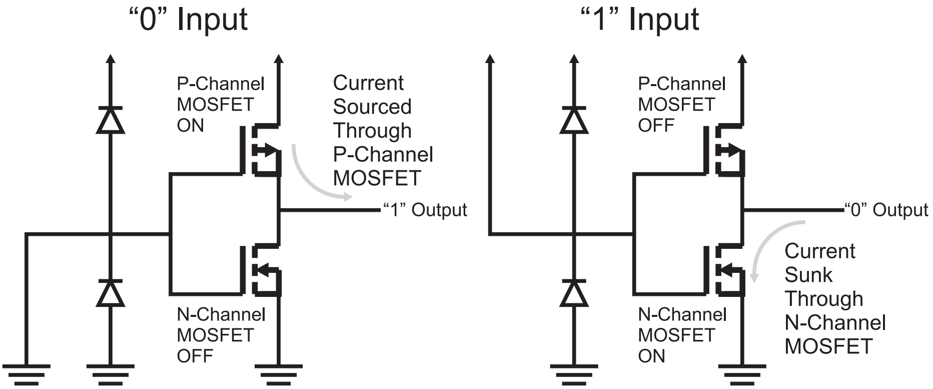

Fig. 3-29. CMOS gate response for different inputs.

appropriate transistor to turn on and either source current (in the case where a negative voltage is applied and the P-channel MOSFET turns on) or sink current (a positive voltage will turn on the N-channel MOSFET). This operation can be seen in Fig. 3-29.

One interesting aspect of the two MOSFET transistors that I have shown wired as an inverter is that they not only provide the ability to sense and respond to voltage inputs, but as the voltage controls transistor switches, they are also effective totem pole output circuits as well. Not only are MOSFET transistors much easier to place on a piece of silicon semiconductor and can be placed in a smaller amount of surface area but also gates built from them are also much simpler than their TTL counterparts.

When the P-channel MOSFET is removed from the output of a CMOS gate, its output is said to be ‘‘open drain’’. This term refers to the drain of the N-channel MOSFET that is not connected to a transistor which can source current in just the same way as an ‘‘open collector’’ TTL output transistor and does not have a transistor which can source current. The CMOS logic open drain output works exactly the same way as the TTL open collector output.

The two ‘‘clamping diodes’’ are placed in the circuit to hold the voltages to within Vdd (power input) and Vss (ground) and are primarily there to protect the P-channel and N-channel MOSFETs from damage from static electricity. These diodes also provide you with the ability to power a CMOS chip through its input pins; when no voltage is applied to Vdd but there is a high-voltage input to one or more input pins, the clamping diodes will allow current to pass to the internal MOSFETs and power the circuit. This is usually an undesirable side effect and one that you should watch for.

The clamping diode function is provided in TTL by the diode and the bipolar transistor emitter that makes up a TTL gate input. Whereas CMOS logic requires additional diodes built into the circuitry, TTL has this function built in.

Unlike TTL, CMOS logic is voltage controlled; there is no path for current to enter or leave the MOSFET's gate circuitry. This has some interesting side effects that you should be aware of. The first is that while at first glance of the inverter operation in Fig. 3-29 it appears that there is no current flow if the output of the CMOS input transistors was another CMOS gate, there actually is a very small amount of change passed to the gates of the transistor from Vdd when the P-channel MOSFET is turned on and this charge is sunk to Vss when the N-channel MOSFET is turned on. This transfer of charge grows with the number of CMOS gates as well as the speed that the gates switch; the faster they switch the more charge that is transferred over time. As I discussed at the start of this chapter, the measurement of charge movement over time is current.

Earlier in the book, I said that the basic gate used in CMOS logic circuits is the NOR gate (just as the NAND gate is the basic gate used in TTL). Before leaving this chapter, I would like to show you the circuit used by a CMOS NOR gate (Fig. 3-30). If you trace through the operation of the four MOSFETs that make up this circuit, you will discover that the only time both P-channel MOSFETs are on (and voltage/current from Vdd is passed to the "Output") is when the two inputs are low, which matches the expected operation of the NOR gate.

The reason why the NOR gate was selected for use as the basic CMOS logic gate has to do with how MOSFETs and other circuits are put

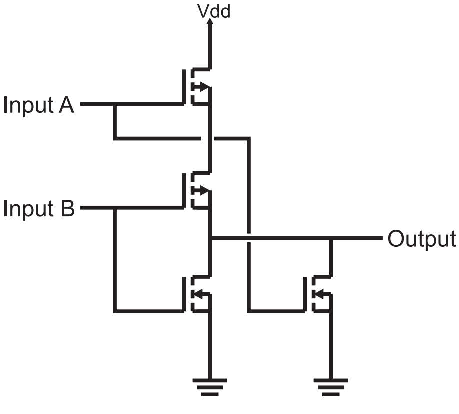

Fig. 3-30. CMOS NOR gate.

down on a silicon semiconductor. The NOR gate is the most efficient while the NAND (which would make the basic building blocks of TTL and CMOS logic the same) cannot be accomplished as easily and in as small amount of space.

The last point I want to make about inputs and outputs is how to wire them when you want to hold them at a specific state (high/"1" or low/"0"). While you could connect the pins directly to power (for a high input) and ground (for a low input), I want to show you the recommended way of doing this and explain why you should go through the extra effort. Connecting the input to high value is accomplished using a 10 k resistor (called a "pull up"), as I show in Fig. 3-31. This circuit will allow input to be temporarily wired to ground (for testing or circuit debug), without causing a short circuit (a low-resistance path between positive and negative power voltage).

Providing a "pull down" (connection to ground) is not quite so simple; the single resistor pull up of Fig. 3-31 is input into an inverter, as shown in Fig. 3-32. This circuit allows the pull up to be connected to ground for testing and debug (changing the input of the gate to a high from a solid low) just as in the pull up case.

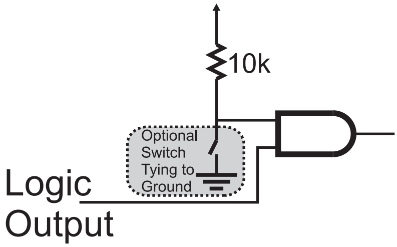
10k
Optional
Switch
Tying to
Ground
Logic
Output

Fig. 3-31. The best method of implementing a "pull up".

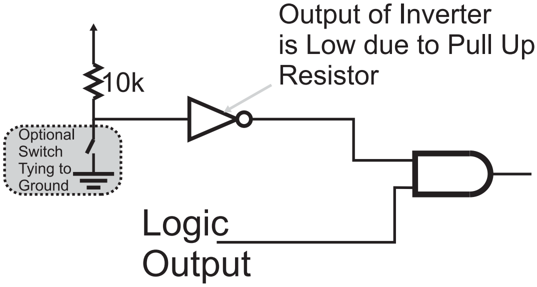
Output of Inverter
is Low due to Pull Up
Resistor
10k
Optional
Switch
Tying to
Ground
Logic
Output

Fig. 3-32. Recommended way to "pull down" a logic input.

If you have followed the gate explanations up to this point, you might be feeling like these methods of tying the gates to pull ups and pull downs is "overkill". I admit that these methods may seem more complex than just wiring the inputs to positive or negative power, but there are a number of reasons for specifying that pull ups and pull downs are wired in this way. For TTL, to make an input high all the time it can be simply left unconnected and to pull it down it can be pulled directly to ground; the 1 mA of current that will flow through the gate to ground should not be an excessive amount of current. For CMOS logic, the input pin can be tied directly to Vdd (positive power) for a high input and Vss (negative power) for a low input – there will be no current flow in either case. It is important to understand the three reasons why I recommend using the pull up resistor or the pull up resistor and inverter.

First, as I said above, it allows you to temporarily change the input value by connecting the resistor voltage to negative voltage without worrying about damaging any part of the circuit. Secondly, it allows simple test equipment to change the state of the input pin for testing without potentially overloading the circuit or the tester. This is a very important consideration when you are designing a product for mass production. Finally, this method can be used for both TTL and CMOS logic without regard to what type of logic is being used. I realize that going through the rigor of following these recommendations increases the complexity of a circuit as well as increasing the number of gates required, its cost and power consumption. In many cases, you will not feel that it is necessary, but if you decide to forgo using pull ups and inverted pull ups, make sure you understand what are the tradeoffs and the risks of the decision.

Simple Digital Logic Circuit Development

Many people do not realize that it is quite easy to build sample digital electronic logic circuits that demonstrate the concepts that have been presented to you as well as let you try out your own simple experiments. If you have, or are taking, a course in digital electronics, it probably includes a well-equipped laboratory in which you worked through a number of experiments. You do not need to replicate this laboratory at home if you wish to experiment with digital electronics. As I will show in this chapter, you can come up with a very capable digital logic circuit test kit for less than $20 and use parts available in modest electronics stores (like "Radio Shack").

Chances are, you are familiar with a variety of different electrical power sources: the ones that comes to mind first are batteries. There are a confusing number of different batteries that you can choose from, ranging from simple ‘‘AA’’ batteries that cost a few cents to the batteries used in the International Space Station that weigh (on Earth) 1200 pounds and cost over $200,000 each. Along with batteries, electricity can also be produced by generators, solar cells and fuel cells. Within your home you can access electrical power very conveniently through outlets in the walls, although this power is *alternating current* (‘‘AC’’) and not the *direct current* (‘‘DC’’) required for digital logic. AC power coming from the sockets in your home will have to be reduced and rectified into DC.

When you are experimenting with simple electronics, I think it’s best to use a power source that is definitely ‘‘low end’’; ‘‘alkaline’’ and rechargeable nickel–metal hydride (‘‘NiMH’’) batteries are widely available to power your experiments. TTL digital electronic chips generally operate between 4.5 and 5.5 volts – you could come up with a combination of batteries that will provide 5 volts to your circuit, or convert a 9 volt radio battery output to 5 volts using a ‘‘regulator’’. Rather than going through this effort and potential expense for TTL, I am going to recommend that you use CMOS digital logic chips that can be powered by 9 volts directly.

A 9 volt battery ‘‘clip’’ (Fig. 3-33) will cost you just a few cents and a bag of them can be bought for a dollar or so. For the purposes of the digital logic circuit test kit, you should look for a 9 volt battery clip that either has wire’s individual strands soldered together (the ends of the wires will look silver, shiny and attached together) or has a single strand. The wires will be covered

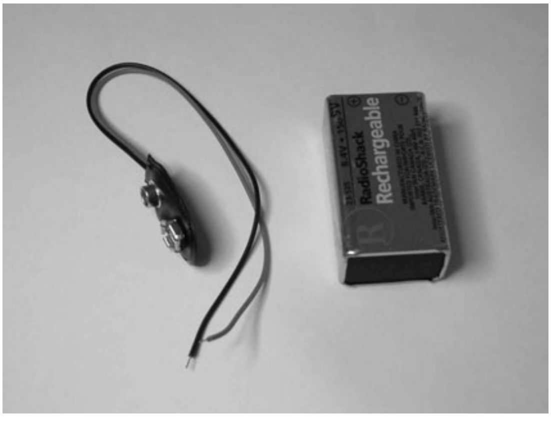

Fig. 3-33. 9 volt battery clip with red (positive) and black (negative) wires attached to it.

in a red and black plastic insulation and the strands will poke out the ends for a 1/4 inch or so.

Make sure the strands of the 9 volt battery clip wires are either soldered together or the wires consist of a single strand, because the wires from the battery clip will be pushed into holes and clamped by copper springs to provide power for the test circuits. Loose, individual strands break easily, can short with other loose wires or become a tangled mess, none of which are good things.

The battery clip is only one part of the wiring that will be used with the digital logic circuit test kit. By itself, the battery clip brings power out of the 9 volt battery conveniently, but is difficult to work with when you are working with chips and even moderately complex circuitry. The ‘‘breadboard’’ and wiring kit (Fig. 3-34) provide a customizable platform in which chips and other electronic components can be inserted into and easily wired together.

‘‘Breadboards’’ allow you to simply and quickly wire up your own prototyping circuits. From the top, a breadboard looks like a sea of holes, but if you were to ‘‘peel back’’ the top (Fig. 3-35), you would see that the holes are actually interconnected, with the central groups of holes connected outwards and the outermost two sets of holes connected along the length of the breadboard.

The central holes are spaced so that DIP chips can be placed in the breadboard and wired into the circuit easily. The outside two rows

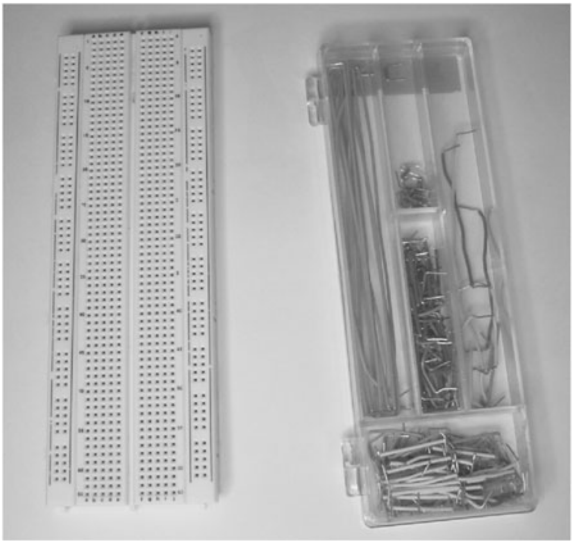

Fig. 3-34. Breadboard with a wiring kit.

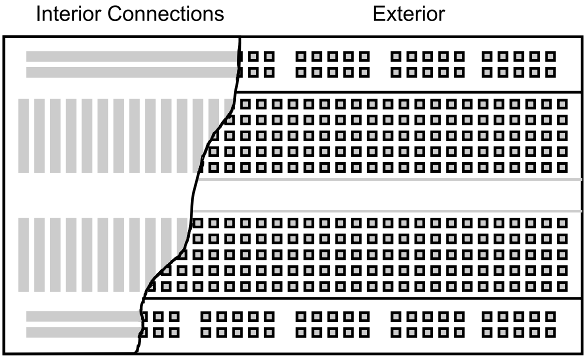

Fig. 3-35. “Breadboard” with interior connections shown.

of holes, I use as power “buss bars” and connect the power source to them directly.

Along with the breadboard, you can either buy a pre-cut and stripped wiring kit (shown in Fig. 3-34) or a roll of 24-gauge solid core wire and some needle nose pliers, wire clippers and maybe some wire strippers. For convenience, I usually go with the wiring kit as it costs just a few dollars.

Along with buying the battery clip, breadboard and wiring kit, you should also buy:

1. 5 or so LEDs in a 5 mm package
2. 10 or so 1k, 1/4 watt resistors
3. 10 or so 0.01 µF ceramic capacitors
4. One 555 oscillator/monostable chip
5. 5 or so SPDT switches, that can be inserted into the breadboard
6. One 74C00 quad two-input NAND gates chips
7. One 74C02 quad two-input NOR gates chips
8. One 74C04 hex inverter chip
9. One 74C08 quad two-input AND gates chip
10. One 74C32 quad two-input OR gates chip
11. One 74C74 dual D-flip flop chip.

All these parts should cost you less than $20 and are available at a fairly wide variety of sources including:

- Radio Shack (http://www.radioshack.com)
- Digi-Key (http://www.digikey.com)
- Mouser Electronics (http://www.mouser.com)
- Active Components (http://www.active-electronics.com).

You will not require any test equipment (such as a Digital Multi-Meter) for this kit and the sample circuits that I will present in this book.

Testing a Simple TTL Inverter

So far I have used the term ‘‘load’’ when I’ve described the electronic devices that are to be used in a circuit, but before going on, I want to familiarize you with the basic, ‘‘dual in-line package’’ ‘‘chip’’ (Fig. 3-36). The ‘‘chip’’ consists of a rectangular plastic box which has a series of metal pins (or connections) coming out from the two long sides. These pins are the electrical connections that are to be made to make up the digital logic circuits as well as provide power to the chip. As I have shown in Fig. 3-36, there can be one or two ‘‘pin 1’’ indicators on each chip (not all chips have both indicators) and the pins are numbered by going counterclockwise around the top of the chip.

Before leaving this chapter, I would like to show both how easy it is to create a simple circuit to test out ideas and parts of applications as well as demonstrate how the TTL gate works. You should have a pretty good idea of how to wire in the chip, but you probably have some questions on how to create useful inputs and outputs to see what’s happening. The output will simply consist of a resistor and a LED – when the chip’s output is high, the LED will be on. Providing the same function for the input, a LED that is on

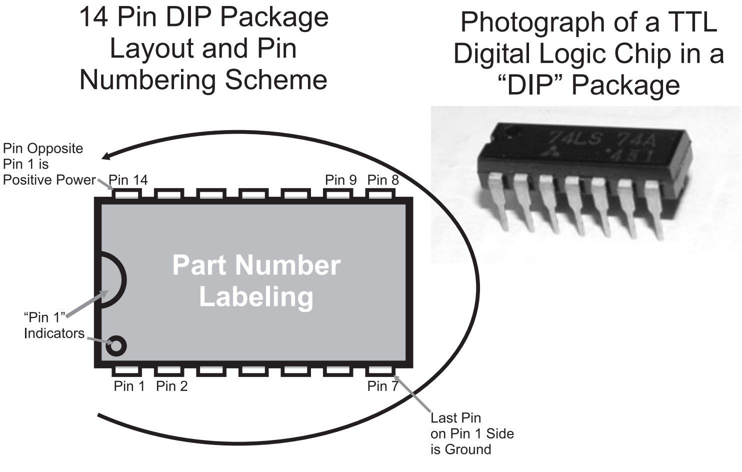
14 Pin DIP Package Layout and Pin Numbering Scheme
Photograph of a TTL Digital Logic Chip in a “DIP” Package
Pin Opposite Pin 1 is Positive Power
Pin 14
Pin 9
Pin 8
Part Number Labeling
“Pin 1” Indicators
Pin 1
Pin 2
Pin 7
Last Pin on Pin 1 Side is Ground
74LS 74A
431

Fig. 3-36. Integrated circuit ‘‘dual in-line package’’ (‘‘DIP’’) – aka a ‘‘chip.’’

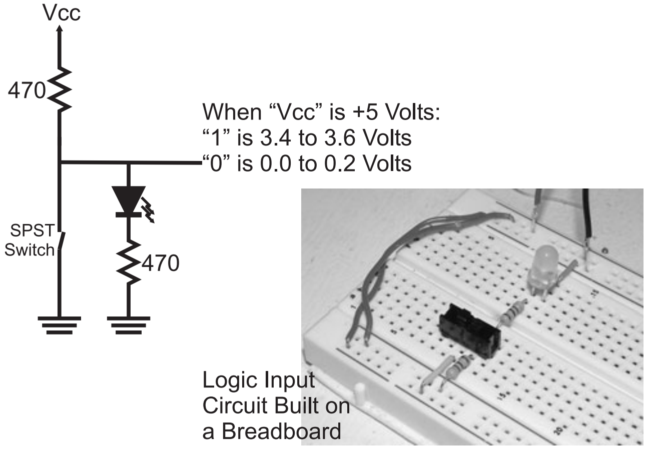

Fig. 3-37. Switch input circuit with LED.

when the input is high is a bit more difficult and uses the circuit shown in Fig. 3-37.

This input circuit probably seems to be much more complex than I have led you to believe is necessary, but there are some requirements that were important for this circuit to meet so that it could be used in a variety of different situations. The first requirement was that it had to work for both TTL (using 5 volt power) as well as CMOS logic (powered from 5 to 9 volts). By providing a direct path to ground, the low voltage requirement of CMOS logic and the current path to ground for TTL was provided. Next, it had to light a LED when the input was high and turn it off when the input was low; the switch will provide a zero impedance current path for the current from the positive power to bypass the LED. Finally, it had to be easy for you to wire and check over in case it doesn't seem to be working properly.

In Fig. 3-37, along with the logic input circuit schematic, I have included a photograph of the completed circuit built on a breadboard. In the photograph, notice that I have clipped the LED and resistor leads to keep the circuit as neat as possible on the breadboard. I strongly recommend that you keep components as close to the surface of the breadboard as possible to minimize your confusion when you are starting to build more complex circuits.

To demonstrate the operation of the inverter, you can build the circuit shown in the left side of Fig. 3-38 on your breadboard using the wiring diagram on the right side of Fig. 3-38. When the input LED is on, the output LED will be off and vice versa. If one or the other LED does not light, then first check your wiring followed by the polarity of the LEDs – the flat side of

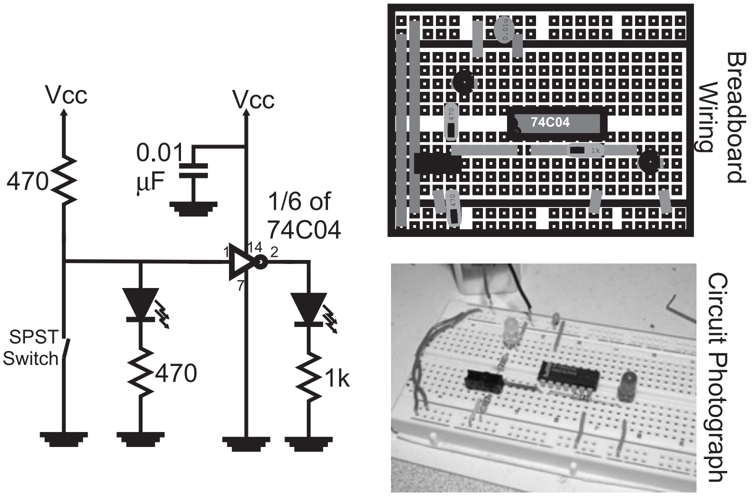

Fig. 3-38. Putting the input with an inverter.

the LED must be connected to the negative voltage (Vss) connection of your circuit.

To build the inverter test circuit, you will need the following parts:

- Breadboard
- 9 volt battery
- 9 volt battery clip
- 74C04 CMOS hex inverter chip
- Two 5 mm LEDs
- Two 470 Ω 1/4 watt resistors
- 1 k 1/4 watt resistor
- 0.01 µF capacitor (any type)
- Breadboard mountable switch (Digi-Key EG1903 suggested).

The only part that you might have some problems finding is the breadboard mountable switch (the EG1903 is a single-pole, double-throw switch with three posts 0.100 inch apart). This part is fairly unique and if you don't want to go through the trouble of ordering the part from Digi-Key, you can either add wires to another switch or simply connect the circuit to the Vss connection to simulate the switch closing (in this case, the LED will go off indicating a low input, just as if a switch were in circuit).

The 74Cxx family of chips are CMOS logic that are pin and output current compatible with 74LSxx TTL chips. The 74C04 used in the circuit shown in Fig. 3-38 demonstrates the operation of the NOT gate (or inverter) to quite good effect. The 74C04 does not demonstrate the operation of a TTL gate all

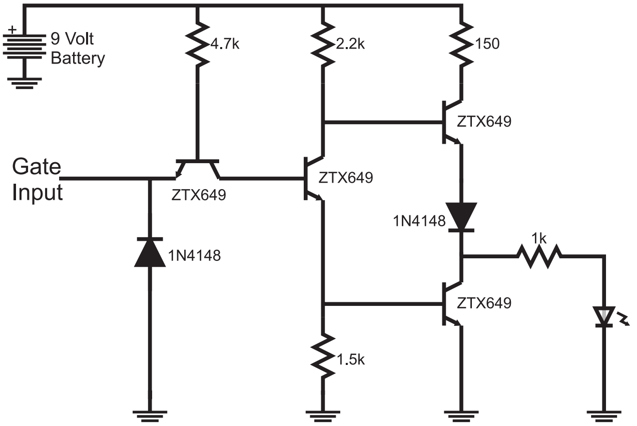
9 Volt
Battery
4.7k
2.2k
150
ZTX649
Gate
Input
ZTX649
ZTX649
1N4148
1N4148
1k
ZTX649
1.5k

Fig. 3-39. Circuit to demonstrate inverter operation.

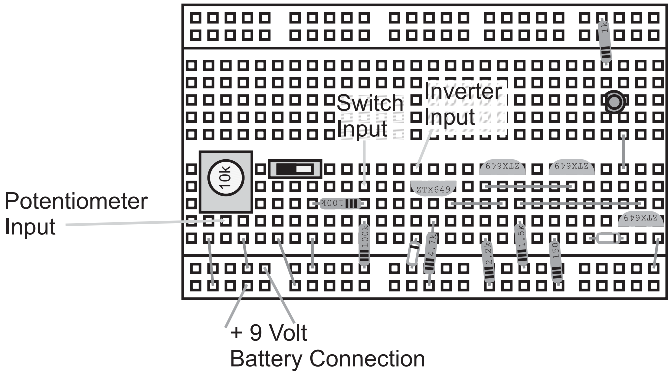
Switch
Input
Inverter
Input
Potentiometer
Input
10k
ZTX649
ZTX649
ZTX649
ZTX649
100k
100k
1k
4.7k
2.2k
1.5k
150
+ 9 Volt
Battery Connection

Fig. 3-40. Breadboard wiring diagram for the TTL inverter circuit.

that well, so if you have a few moments, I suggest that you build the circuit shown in Fig. 3-39 (wired according to Fig. 3-40) and test it out – externally, it will seem to work identically to the 74C04 circuit shown in Fig. 3-38, but there are a few differences that you can experiment with.

The parts that you will need for this circuit are:

- Breadboard
- 9 volt battery
- 9 volt battery clip
- Four 2N3904 NPN bipolar transistors
- Two 1N914 (or equivalent) silicon diodes
- Two 5 mm LEDs
- 150 Ω 1/4 watt resistor

- Two 470 Ω 1/4 watt resistors
- 1 k 1/4 watt resistor
- 1.5 k 1/4 watt resistor
- 2.2 k 1/4 watt resistor
- 4.7 k 1/4 watt resistor
- 100 k 1/4 watt resistor
- 10 k potentiometer
- Breadboard mountable switch (Digi-Key EG1903 suggested).

Going through the circuit, you can see that current flows through the circuit in two different directions, as shown in Figs. 3-41 and 3-42. When the input is

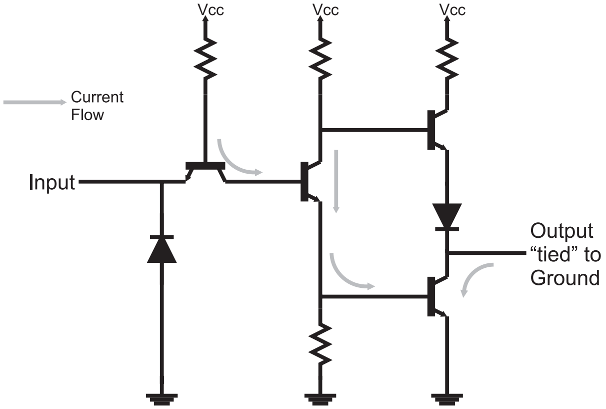
Vcc
Vcc
Vcc
Current
Flow
Input
Output
"tied" to
Ground

Fig. 3-41. TTL inverter with a "1" or floating input.

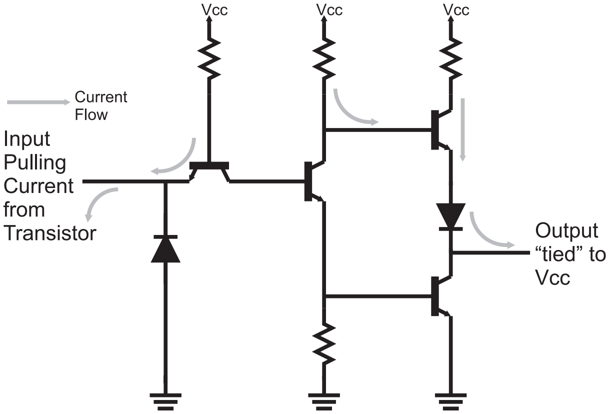
Vcc
Vcc
Vcc
Current
Flow
Input
Pulling
Current
from
Transistor
Output
"tied" to
Vcc

Fig. 3-42. TTL inverter with a "0" input.

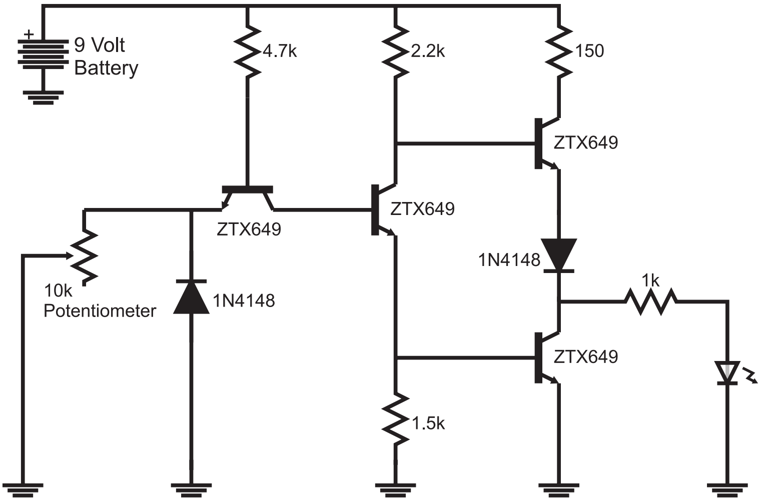

Fig. 3-43. Circuit to test current draw for TTL operation.

"high" (LED on) and you follow the current path, you will see that the current will ultimately turn on the bottom right transistor, connecting the gate's output pin to ground ("low" voltage output). When current is drawn from the TTL input pin (Fig. 3-42), the current that ultimately turned on the bottom right transistor is taken away, resulting in a different path for currents within the gate. This change in current flow ultimately turns on the top right transistor, effectively tying the output to power and driving out a "high" voltage.

Once you have built the circuit and tested it, you can now look at the operating aspects of it by putting a potentiometer in the circuit, as I have shown in Fig. 3-43, and adjust it until the LED either flashes on and off or dims. If you have a digital multi-meter (DMM), you will find that the *threshold* current is about 1 mA, with a voltage across the potentiometer of around 0.5 volts.

The final aspect of this experiment is to wire the inverter's input as shown in Fig. 3-44 and alternatively connect the input (passing through the 100 k resistor) to the power in or ground. You will find that the LED never turns on regardless of the switch position. If you were to measure the voltage at the 100 k resistor, you would see that it is connected directly to the power and ground connections, but the circuit seems to ignore the ground connection. The 100 k resistor prevents the 1 mA of current passing through to ground, resulting in the LED being turned on. If you were to repeat this experiment with the 74C04, you would see the LED turning on and off according to the voltage at the 100 k resistor.

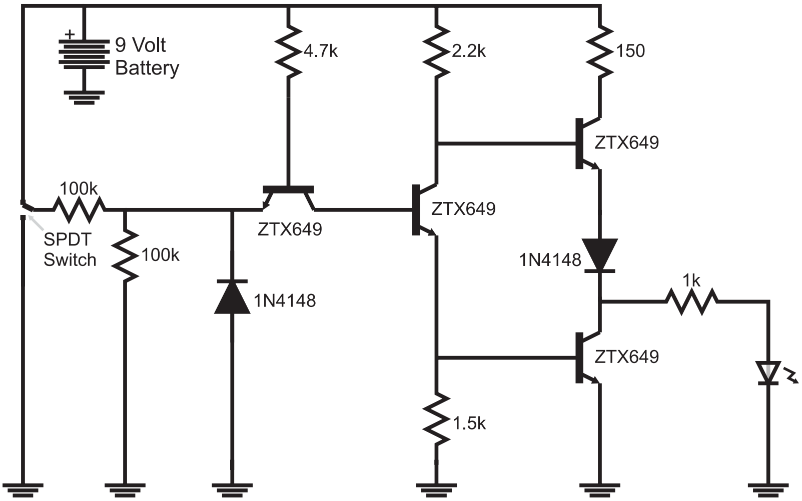

Fig. 3-44. Circuit to test voltage control of TTL operation.

In this chapter, I have given you a brief tutorial in basic electronics, an introduction to semiconductors and a method that you can use to build test circuits to experiment with digital electronics. In these few pages, I have covered the material included in several high school and college courses. It was not my intention to overwhelm you, but provide you with enough information to understand what is happening in a digital electronic circuit as well as give you a few basic rules to help you avoid problems, or if things aren't working as you would expect, to have some ideas on where to look for the problems.

Quiz

1. Electricity must:
   (a) Change polarity 60 times a second
   (b) Flow between the planets
   (c) Be equal in all parts of a circuit
   (d) Flow in a closed, continuous loop
2. Every electrical circuit has three parts:
   (a) Breadboards, batteries and electronic parts
   (b) Power source, load and conductors

(c) Intelligence, compassion and a sense of humor
(d) Speed, power (or torque) and corporeal form

3. In the water pipe/tap/hose example, if you were to partially close the tap:
   (a) Water would stream out faster from the hose
   (b) The tap would get hot in your hand from the friction of the water passing through it
   (c) The amount of water leaving the hose would decrease
   (d) The water leaving the hose would stream further

4. In a single resistor circuit, if you apply 9 volts and measure 100 mA flowing through it, the resistance value is:
   (a) 9 ohms
   (b) 900 ohms
   (c) 90 ohms
   (d) 1,111 ohms

5. The equivalent resistance of a 10 ohm and 20 ohm resistor in parallel:
   (a) Is always zero
   (b) 30 ohms
   (c) 7.5 ohms
   (d) 6.7 ohms

6. A diode is said to be "forward biased" when:
   (a) A positive voltage is applied to the "bar" painted on the side of the diode
   (b) Electrons are injected into the P-type semiconductor of the diode
   (c) Current flows into the diode through the end which doesn't have a band painted on it
   (d) More than 0.7 volts is applied to it

7. If a bipolar transistor with an h_{FE} of 150 had a "small signal operating region" base current of 1 μA to 1 mA, what base current would be required to allow 10 mA collector current?
   (a) This is impossible to answer because 10 mA collector current is greater than 1 mA.
   (b) 1 mA
   (c) 67 μA
   (d) 667 μA

8. The basic TTL gate is:
   (a) The NOT gate
   (b) The AND gate
   (c) The NOR gate
   (d) The NAND gate

9. Totem pole outputs are best used:
   (a) When there are multiple outputs tied together as a ‘‘dotted AND’’
   (b) To drive electric motors
   (c) As the default output type used in digital electronic circuits
   (d) When high-speed operation of the digital electronic circuit is required

10. The dual in-line package:
    (a) Is a standard method for packaging digital electronic chips
    (b) Is used because part numbers cannot be stamped on bare chips
    (c) Allows for an easy visual check to see whether or not the part was damaged by heat
    (d) Facilitates effective cooling to the chip inside

4

CHAPTER

Number Systems

Working through the book to this point, you should be comfortable with combining multiple single bit values together in a variety of different ways to perform different combinatorial circuit functions. Along with being able to meet the basic requirements, you should be able to optimize the circuit to the fewest number of gates that is available within the technology that you are going to use. This skill is very useful in itself, but it is only scratching the surface of what can be done with digital electronics; most data consists of more than a single bit (which can have only two values) to process, and working with multiple single bits of data can be cumbersome. What is needed is a methodology for combining bits together so they can represent larger values that can be simply expressed.

The solution to this issue is to combine bits in exactly the same way as a 10-value character is combined to produce the decimal numbers that you are familiar with. While on the surface, combining bits does not seem to be directly analogous to decimal numbers, by using the same method that decimal numbers are produced, multi-bit numbers (which are most often described as "binary") numbers can be produced.

In primary school, you learned that the four-digit number "1,234" was built out of four digits, any of which could have the 10 values "0", "1", "2",

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

"3", "4", "5", "6", "7", "8" and "9". When listing the different values for a digit, zero is stated because the number "10" is actually a two digit number. The number of different values for each digit is referred to as its "base" or "radix". It is important to note that the first value is always zero and the last value is the base minus one.

When expressing each digit, its value was stated by the "column" it was in ("ones", "tens", "hundreds", "thousands", etc.). For example, the second column of "1,234" is the "hundreds" column and in 1234, there are two hundreds.

In high school, you would have been introduced to the concept of exponents and instead of expressing each digit in the number by the column, you would express it by the digit multiplier. So, 1,234 could now be written out as:

$$
\begin{aligned}
&\texttt{1 Thousand} = 1 \times 1{,}000 = 1 \times 10^3 \\
&\texttt{2 Hundreds} = 2 \times 100 = 2 \times 10^2 \\
&\texttt{3 Tens} \quad\;\; = 3 \times 10 = 3 \times 10^1 \\
&\texttt{4 Ones} \quad\;\; = 4 \times 1 = 4 \times 10^0
\end{aligned}
$$

The beauty of expressing a number in this way is that each digit's multiplier is mathematically defined as a power of the base. Using this format, it is possible to create a numbering system using single bits to represent "binary" numbers.

For example, four bits could be put together with the bit containing the least significant digit labelled "Bit_0", the second least significant as "Bit_1", the second most significant as "Bit_2" and the most significant as "Bit_3". The term significance when applied to bits is used to express the magnitude of the bit's multiplier. For example, Bit_0, which is multiplied by 2^0 or 1, has less significance than Bit_3 which is multiplied by 2^3 or 8.

Using the same exponent format as was used to define the decimal number 1,234, the four-bit binary number could be defined as:

$$
\begin{aligned}
&\texttt{Bit}_3 \times 2^3 \\
&\texttt{Bit}_2 \times 2^2 \\
&\texttt{Bit}_1 \times 2^1 \\
&\texttt{Bit}_0 \times 2^0
\end{aligned}
$$

and written out in a similar format to a decimal number. Collectively, the number is written out as a series of ones and zeros, in a similar manner to that of a decimal number.

Many books go into great length trying to explain how to convert a decimal number to a binary number. I won't go into the same amount of detail because the algorithm to do this is really quite simple: you simply start at some most power of two and work your way down, writing out a "1" for each time the subtraction the power of two results in a positive number or zero and a "0" when the difference is negative.

Written out as part of a "C" program, converting a decimal number to a character four-bit binary number is accomplished by the following statements:

```
for (i = 4; i! = 0; i − −)
    if ((DecVal − (2 ∗ ∗(i − 1)))> = 0)
    {                            //Can Take Away Digit Value
        DecVal = DecVal − (2 ∗ ∗(i − 1));
        Bit[i − 1] = `1´;
    } else                       //Result of subtraction is negative
        Bit[i − 1] = `0´;  //Can't take away value
```

Note that I start at "4" and subtract one for the actual bit value in the example code above.

Demonstrating the algorithm, consider the case where you wanted to express the decimal number "11" as a four-bit decimal. In Table 4-1, I have listed each step of the program with the variable values at each step.

Converting binary numbers to decimal is very easy because the power of two of each digit that has a value of "1" are summed together. The "C" code

Table 4-1 Converting decimal 11 to binary 1011.

'Bit' variable	i	DecVal	DecVal − ((2 ** (i − 1))
xxxx	4	11	3 (Positive)
1xxx	3	3	−1
10xx	2	3	1 (Positive)
101x	1	1	0 (Positive)
1011	0	0	N/A

Table 4-2 Converting binary 0110 to decimal 6.

'Bit' variable	i	Bit[i − 1] != 0?	DecVal
0110	4	No	0
0**1**10	3	Yes	4 (DecValue + (2 ** (3 – 1)))
01**1**0	2	Yes	6 (DecValue + (2 ** (2 – 1)))
011**0**	1	No	6
0110	0	N/A	6

to convert a value in "Bit" to a decimal value is:

```
DecVal = 0;                    // Initialize the Decimal
                                  Value Variable
for (I = 4; i> = 0; i − −)     // Repeat for four bits
   if(Bit[i − 1] == "1")      // Add Digit Value if Digit Not 0
     DecValue = DecValue + (2 ∗ ∗(i − 1));
```

In Table 4-2, I have listed the process of converting the binary number 0110 to decimal and you should note that I have highlighted the bit that is being tested.

Before going on, I would like to point out that there can be a lot of confusion with regards to using binary numbers with decimal numbers or numbers of different bases. To eliminate the confusion, you should always identify the binary numbers by placing a percentage sign ("%") or surrounding it with the letter "B" and two single quotes ("' "). Using these conventions, the bit pattern converted in Table 4-1 would be written out as %0110 or B'0110'. The % character put before a binary number is a common assembly language programming convention. The letter "B" and the single quotes around the number is the format used in "C" programming and will be the convention that I use in this book.

Another area of confusion with regards to binary numbers is how they are broken up for easier reading. Each group of three digits in a decimal number is usually separated from other groups of digits by use of a comma ("," in North America and a period or dot (".") in Europe and other parts of the world). When working with binary numbers, instead of separating each three digit group with a punctuation character, it is customary to use a blank to separate four digit groups. Using the conventions outlined here, the eight bit number 10111101 would be written out as:

`B'1011 1101'`

This is the binary number format convention that I will use for the rest of the book.

Base 16 or Hexadecimal Numbers

As I will show in this and the next section, having programming experience is a two-edged sword – it will help you understand certain concepts (such as the "bit" and some data structures like the ones presented in this and the next section), but it will blind you to other opportunities. The goal of these sections is to illustrate how bits can be grouped together to make your design efforts more efficient as well as making it easier for you to both see possibilities for the design and articulate them to other people.

Creating binary numbers from groups of bits, as I demonstrated in the introduction to this chapter, is quite easy to do, but can be very cumbersome to write out as well as transfer correctly. You may also have difficulty in figuring out exactly how to express the number, asking should it be passed along starting from the most significant or least significant bit. At the end of this chapter's introduction, I left you with the number B'1011 1101' and you should agree that telling somebody its value is quite cumbersome; for example, you might say something like, "The eight bit, binary number, starting with the most significant bit is one, zero, one, one, one, one, zero and one."

It is much more efficient to combine multiple bits together into a single entity or digit.

The most popular way of doing this is to combine four bits together as a "hexadecimal" digit which has 16 different values. This numbering system has a base of 16. If you are familiar with programming, chances are you are familiar with hexadecimal digits (which is often contracted to the term "hex"), which I have listed out with their decimal and binary equivalents in Table 4-3.

To create a way of expressing the 16 values, the first 10 hexadecimal values are the same as the 10 decimal number values, with the following six being given letter codes. This is why I included the "phonetic" values for the hexadecimal values greater than 9; the letter names "B", "C" and "D" can be easily confused, but their phonetic representations are much clearer.

Table 4-3 Hexadecimal digits with binary, decimal equivalents and phonetic values.

Decimal	Binary	Hex	Phonetic	Decimal	Binary	Hex	Phonetic
0	B'0000'	0	Zero	8	B'1000'	8	Eight
1	B'0001'	1	One	9	B'1001'	9	Nine
2	B'0010'	2	Two	10	B'1010'	A	Able
3	B'0011'	3	Three	11	B'1011'	B	Baker
4	B'0100'	4	Four	12	B'1100'	C	Charlie
5	B'0101'	5	Five	13	B'1101'	D	Dog
6	B'0110'	6	Six	14	B'1110'	E	Easy
7	B'0111'	7	Seven	15	B'1111'	F	Fox

I tend to place a lot of importance to using conventions when expressing letters. You may be tempted to make up your own letter codes or use the aviation phonetic alphabet (Table 4-4) when communicating hexadecimal values to other people ("AF" could be "Apple-Frank" or "Alpha-Foxtrot" instead of "Able-Fox"). I would like to discourage this for two reasons: the first is that the person you are talking to will have to mentally convert your words into letters and then hex digits – this process is complicated when unexpected words are used. Secondly, I prefer using the phonetic codes in Table 4-3 for hex values and the aviation phonetic codes for letter codes.

Multi-digit hexadecimal numbers are written out in a similar way as decimal or binary numbers with each digit multiplied by 16 to the power of the number of value's position. For a 16 bit number (four hexadecimal digits), the digit multipliers are listed below:

$$\mathtt{HexDigit}_3 \times 16^3 = \mathtt{HexDigit}_3 \times 4{,}096$$

$$\mathtt{HexDigit}_2 \times 16^2 = \mathtt{HexDigit}_2 \times 256$$

$$\mathtt{HexDigit}_1 \times 16^1 = \mathtt{HexDigit}_1 \times 16$$

$$\mathtt{HexDigit}_0 \times 16^0 = \mathtt{HexDigit}_0 \times 1$$

To indicate a hex number, you should use one of the programming conventions, such as putting the prefix "0x0" or "$" at the start of the hexadecimal

Table 4-4 Aviation phonetic codes.

Letter	Phonetic	Letter	Phonetic	Letter	Phonetic
A	Alpha	J	Juliet	S	Sierra
B	Beta	K	Kilo	T	Tango
C	Charlie	L	Lima	U	Uniform
D	Delta	M	Mike	V	Victor
E	Echo	N	November	W	Whiskey
F	Foxtrot	O	Oscar	X	X-Ray
G	Gulf	P	Papa	Y	Yankee
H	Hotel	Q	Quebec	Z	Zulu
I	India	R	Romeo		

value. The same formatting convention used with binary numbers (X'##', where "##" are the hex digits) could also be used. For this book, I will be expressing hexadecimal numbers in the format 0x0## which is visually very different from binary numbers, which should help to immediately differentiate them.

To convert a decimal number to a character 16 bit hexadecimal number, you can use the "C" algorithm shown below. Note that I have used the C modulo ("%") operation which returns the remainder from an integer division operation and not its dividend.

```
for (i = 16; i! = 0; i = i − 4)
{
  if ((DecVal/(16 ∗ (1<<(i − 8)))) > 9)
    Hex[(i/4) − 1] = (DecVal/(16 ∗ (1<<(i − 8)))) − 10 + `A´;
  else
    Hex[(i/4) − 1] = (DecVal/(16 ∗ (1<<(i − 8)))) + `0´;
  DecValue = DecValue%(16 ∗ (1<<(i − 8));
} //  rof
```

Going the other way, to convert a four hexadecimal digit number to decimal you can use the algorithm:

```
DecVal = 0;
for(i = 4; i! = 0; i − −)
   if((Hex[i − 1] >= “A”)&&(Hex[i − 1] <= “F”))
      DecVal = (DecValue ∗ 16) + (Hex[i − 1] − `A´ + 10);
   else
      DecVal = (DecValue ∗ 16) + (Hex[I − 1] − `0´);
```

Many books provide a conversion table between binary, hexadecimal and decimal numbers, but I would like you to be familiar with the conversion algorithms written out above as well as buy yourself an inexpensive scientific calculator which has ability to convert between base systems. The ability to convert between the base systems is actually quite simple and available in many basic scientific calculators which cost $10 or less. Understanding how to convert between base systems and having an inexpensive calculator will enable you to perform the conversions faster and with more flexibility than using a table, which is limited in the number of different values it can present.

If you are familiar with numbers in different languages, then you will know that the prefix ‘‘hex’’ actually refers to the number ‘‘six’’ and not ‘‘16’’. The actual prefix for 16 is the term ‘‘sex’’ and in the early days of computers, this was (obviously) a source of some amusement. When IBM introduced the System/360, in the early 1960s, the company was uncomfortable with releasing something that was programmed in ‘‘sexadecimal’’, fearing that it might upset some users. To avoid any controversy, all documentation for the System/360 was written using the 16 bit ‘‘hexadecimal’’ numbering system presented here. The System/360 was a wild success, becoming the first ‘‘computer for the masses’’ and many people’s first experience in programming and electronics. The term ‘‘hexadecimal’’ became the popular term for 16 bit numbers and displaced the more correct ‘‘sexadecimal.’’

Binary Coded Decimal

In the early days of programming, data structures were often the result of a curious blend of trying to come up with a data format that best suited the programmer and what best suited the current hardware. One of the more

Table 4-5 Decimal digits with binary and BCD decimal equivalents.

Decimal	Binary	BCD	Decimal	Binary	BCD
0	B’0000’	0	8	B’1000’	8
1	B’0001’	1	9	B’1001’	9
2	B’0010’	2	10	B’1010’	Invalid
3	B’0011’	3	11	B’1011’	Invalid
4	B’0100’	4	12	B’1100’	Invalid
5	B’0101’	5	13	B’1101’	Invalid
6	B’0110’	6	14	B’1110’	Invalid
7	B’0111’	7	15	B’1111’	Invalid

enduring structures that came from this time is the ‘‘binary coded decimal’’ (most often referred to by its acronym ‘‘BCD’’) which used four bits, like hexadecimal values, but only allowed the values of zero through nine rather than the full 16 values that were possible (as shown in Table 4-5). The reason for using this data structure has largely disappeared in computer systems, but it is still a viable and useful method of handling data in digital electronics and one that you should keep in your ‘‘hip pocket’’ when you design circuits.

The original reason for using the BCD data format in computer programming was its elimination of the need to add code to the program to convert a binary or hex number into decimal. The code storage required for the conversion was expensive and the processors were nowhere near as possible as what is available today. Using decimal values was actually an optimal way of processing data in these old systems.

The lasting legacy of this is the number of standard chips that can process BCD values just as easily as other standard chips can process hexadecimal values and will allow you to design circuitry that works with decimal values just as easily as if you were working with hex values.

While this is getting a bit ahead of things, I want to give the example of designing a delay that holds back a signal for 100 seconds. Using traditional binary logic, which only works with bits that are a power of two, you would have to design a circuit that compares a counter value and indicates when the value “100” was reached and reset itself. When using digital electronic chips that are designed for BCD values, the comparator function is not required, as each BCD digit cannot be greater than “9” and, cascaded together, they can only count to a maximum value of “99” to “00”.

This may seem like a trivial example, but you will find a number of cases like this one where you will have to create circuits that work on base 10 data and by using chips which are designed for BCD values, the complexity of your work will be greatly reduced.

Going back to Table 4-5, the production of the “invalid” indication is worthy of some discussion as it provides a good example of how gate optimization is not always as straightforward as you might expect.

In most BCD chips, if the value of 10 or more is passed in the binary bits, then the value is converted to zero and a carry indication is output. Using the tools presented in Chapter 2, you should be able to derive the sum of products formula for the positive active “invalid” indicator as:

$$\texttt{Invalid} = (\texttt{A3} \cdot \texttt{A2}) + (\texttt{A3} \cdot \texttt{A2})$$

and using the conversion formulas of Chapter 2, you would simplify the “invalid” formula above to:

$$\texttt{Invalid} = \texttt{A3} \cdot (\texttt{A2} + \texttt{A1})$$

Figure 4-1 shows the AND/OR gates for this function along with the “NAND equivalent” function beneath it. The NAND equivalent was chosen by assuming that the function would be implemented in TTL. While this circuit looks a bit complex, if you follow it through, you will find that it provides the same function as the AND/OR combination above it.

It will probably surprise you to find out that this circuit is not optimal by any measurement: you can do better in terms of the number of gates, the time

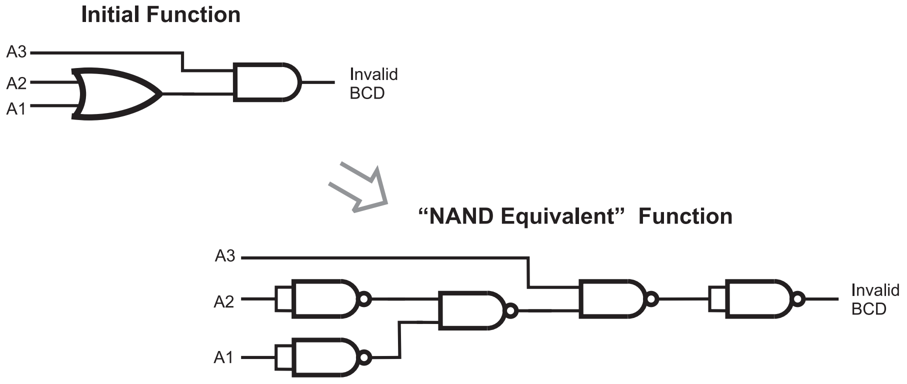

Fig. 4-1. Basic circuit for detecting values over “9” with reduction.

it takes a signal to pass through the gates and in providing a constantly timed output. The circuit at the bottom half of Fig. 4-1 will respond in two gate delays if A3 changes and in four gate delays if A2 changes. For many circuits, this is not a problem, but when you are working with high-performance designs, a variable output delay can result in the application not working correctly and being almost impossible to debug.

A much better approach to optimizing the circuit is to work at converting it to the basic gate used by the technology that you are working with and then optimizing this. Going back to the original "Invalid" equation:

$$\mathtt{Invalid} = (\mathtt{A3} \cdot \mathtt{A2}) + (\mathtt{A3} \cdot \mathtt{A2})$$

I can convert the OR to a NAND, by inverting its two parameters (according to De Morgan's theorem), ending up with:

$$\mathtt{Invalid} = \mathtt{!}(\mathtt{!}(\mathtt{A3} \cdot \mathtt{A2}) \cdot \mathtt{!}(\mathtt{A3} \cdot \mathtt{A2}))$$

It is probably astounding to see that the function provided by the mess of NAND gates in Fig. 4-1 can be reduced to the three simple gates required by the formula above. Along with reducing the number of gates, you should also notice that the maximum number of gate delays is two, regardless of which bit changes.

Looking at the NAND circuits in both diagrams, you are probably at a loss as to how you could reduce the NAND circuit in Fig. 4-1 to the three

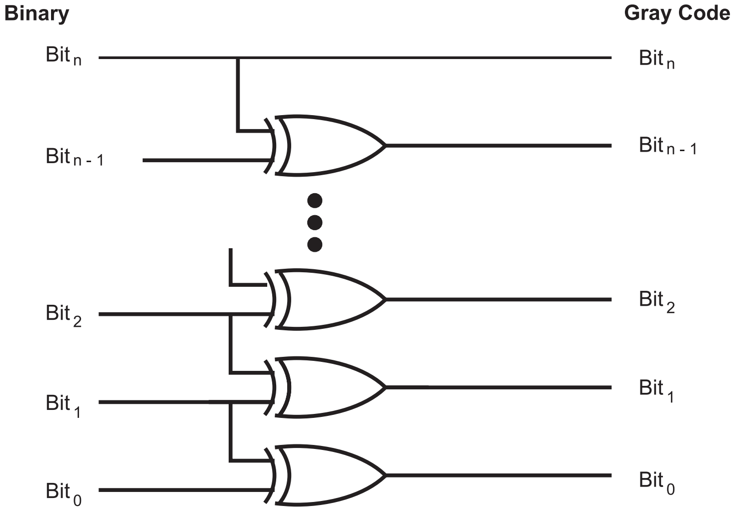

Fig. 4-2. Simple digital electronic circuit to convert a binary number to a Gray code.

gates of the optimized circuit. Personally, I would be surprised if you could; when I look at the two circuits, they look like they provide completely different functions.

What I want to leave you with is an example of how looking at a logic function from different perspectives can result in radically different circuits with surprisingly different parameters. In the first case, I reduced three gates to two, to end up with six NAND gates, while in the second, I avoided reducing the basic function and converted it directly to a much more efficient three NAND gate circuit.

In going through this exercise to produce the “invalid” output for BCD, I hope that you can apply this knowledge for creating circuits that work with different base systems than just a power of two. In some cases, you may have to work with numbers that are base 9 or 13 and using the example here, you should have some idea of how to keep the values within certain “bounds”.

Gray Codes

I hope I have convinced you of the usefulness of using Gray codes for inputs when you are illustrating how digital electronic logic functions respond to inputs. I must point out, however, that Gray codes were originally created for a much different function – they were designed for use in position sensors as the single changing bit allowed hardware to be designed to respond to a single changing bit and not the potentially several bits of a binary sequence. By only changing one bit at a time, absolutely precise positioning of marking sensors (causing all changing bits being sensed at the exact same instant) was not required.

Gray codes were invented by Frank Gray of Bell Labs in the mid 1950s and has a “hamming value” of 1. The hamming value is the number of bits that change between one value and the next. A four bit binary number can have all four bits change as it increments or decrements; a Gray code never has more than one bit change during incrementing or decrementing operations.

Chances are, you would not have any trouble coming up with a two bit Gray code (b’00’, b’01’, b’11’ and b’10’) and in a pinch, you would be able to come up with a three bit Gray code (b’000’, b’001’, b’011’, b’010’, b’110’, b’111’, b’101’ and b’100’). I suspect that if you were given the task of coming up with any more bits than this, you would be stumped.

In trying to come up with a way of explaining how Gray codes worked, I noticed that when a new most significant bit was set, the previous values were ORed with this bit, but written out in reverse order. In some texts, this property is recognized by calling Gray codes a *binary reflected code*. Looking at Table 4-6, you can see that I created a four bit

Table 4-6 Building Gray codes from previous value.

Binary value	Gray code	Comments
B'0000'	B'0000'	1 bit Gray code, Gray code = binary value
B'0001'	B'0001'	1 bit Gray code, Gray code = binary value
B'0010'	B'0011'	2 bit Gray code, last value of bit 1 (B'1') list ORed
B'0011'	B'0010'	2 bit Gray code, first value of bit 1 (B'0') list Used
B'0100'	B'0110'	3 bit Gray code, last value of bit 2 (B'10') list ORed
B'0101'	B'0111'	3 bit Gray code, next last value of bit 2 bit (B'11') list ORed
B'0110'	B'0101'	3 bit Gray code, next last value of bit 2 (B'01') list ORed
B'0111'	B'0100'	3 bit Gray code, first value of bit 2 (B'00') list ORed
B'1000'	B'1100'	4 bit Gray code, last value of bit 3 (B'100') list ORed
B'1001'	B'1101'	4 bit Gray code, next last value of bit 3 (B'101') list ORed
B'1010'	B'1111'	4 bit Gray code, next last value of bit 3 (B'111') list ORed
B'1011'	B'1110'	4 bit Gray code, next last value of bit 3 (B'110') list ORed
B'1100'	B'1010'	4 bit Gray code, next last value of bit 3 (B'010') list ORed
B'1101'	B'1011'	4 bit Gray code, next last value of bit 3 (B'011') list ORed
B'1110'	B'1001'	4 bit Gray code, next last value of bit 3 (B'001') list ORed
B'1111'	B'1000'	4 bit Gray code, first value of bit 3 (B'000') list ORed

Gray code by taking the eight values of the three bit code, reversing them and setting bit 3.

This could be written out as a computer program algorithm as:

```
for  (i = 1; i < #Bits; i + +)  //Loop Around for Each Code
     if(1 == i)                  // − Trivial Initial Case
     {
      GrayCode[0] = 0;  GrayCode[1] = 1;
     } else {                    //Copy Previous Codes with
                                   //New MSB Set
         for(j = (1 << (i − 1)),k = ((1 << (i − 1)) − 1);
          j < (1 << i); j + +,k − −)
            GrayCode[j] = (1 << (i − 1)) + GrayCode[k];
     }        //fi
```

This code demonstrates how Gray codes are produced, but is not the optimal method for producing Gray codes (it is actually an ‘‘order n_2’’ algorithm, which means that every time the number of bits is doubled, the amount of time required to produce the values is quadrupled). Along with this, it is not easy to create digital logic hardware that will create these codes.

Fortunately, individual binary codes can be converted to Gray codes using the circuit shown in Fig. 4-2, which simply implements the formula:

```
Gray Code = Binary ^ (Binary >> 1)
```

Going the other way (from Gray code to binary) is a bit more complex and while it uses n − 1 (where ‘‘n’’ is the number of bits) XOR gates, like converting binary codes to Gray codes, the output of each XOR gate is required as an input to the next least significant bit, as shown in Fig. 4-3. The output of the circuit is not correct until the most significant bit has passed through each of the XOR gates to the least significant bit.

To perform the data conversion a simple formula cannot be used. Instead the following algorithm is required:

```
GrayCode = Binary; Shift = 1;
While((GrayCode >> Shift)! = 0)
```

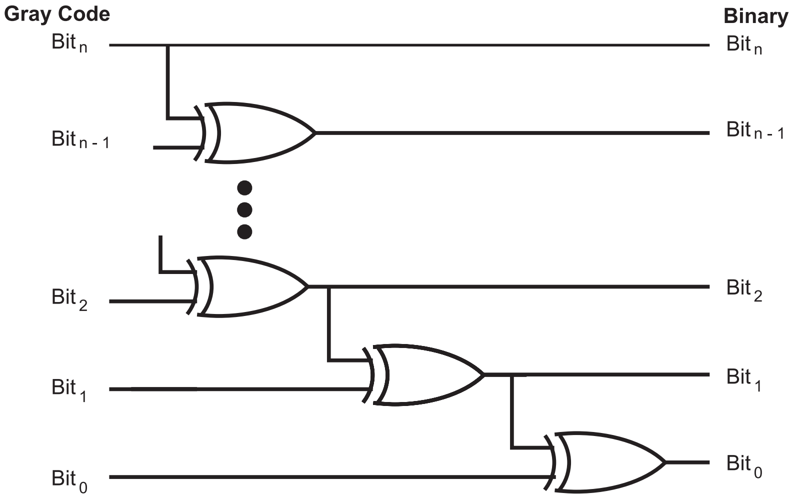

Fig. 4-3. Digital electronic circuit to convert a Gray code to a binary number.

```
  {
      GrayCode = GrayCode ^ (GrayCode >> Shift);
      Shift = Shift * 2;
  } //  elihw
```

I find it very difficult to explain exactly how this code works, except to say that with each iteration of the while loop, the "Gray code" value gets shifted down more and more to move the most significant bits into position for XORing with the less significant bits. To convince yourself that the algorithm works, you might want to perform a "thought experiment" on it and list the changing value of "Gray code" as I have done in Table 4-7.

In this chapter, more than anywhere else in the book, I have used sample computer programs to show how different values can be produced. This is a somewhat different approach to explaining how multi-bit binary data conversions are implemented and one that takes advantage of the ubiquity of the personal computer and the ability of most technical students to perform even rudimentary programming.

Using computer code to help demonstrate how the conversions are done should also give you another method for processing binary values as well as of testing formulas and optimizations. I always find it useful to have a number of different ways to solve a problem, or test a potential solution,

Table 4-7 Working through the shifting values of the Gray code convention algorithm.

Initial bit values	Shift = 1 bit values	Shift = 2 bit values	Shift = 4 bit values
B7	B7	B7	B7
B6	B6 ˆ B7	B6 ˆ B7	B6 ˆ B7
B5	B5 ˆ B6	B5 ˆ B6 ˆ B7	B5 ˆ B6 ˆ B7
B4	B4 ˆ B5	B4 ˆ B5 ˆ B6 ˆ B7	B4 ˆ B5 ˆ B6 ˆ B7
B3	B3 ˆ B4	B3 ˆ B4 ˆ B5 ˆ B6	B3 ˆ B4 ˆ B5 ˆ B6 ˆ B7
B2	B2 ˆ B3	B2 ˆ B3 ˆ B4 ˆ B5	B2 ˆ B3 ˆ B4 ˆ B5 ˆ B6 ˆ B7
B1	B1 ˆ B2	B1 ˆ B2 ˆ B3 ˆ B4	B1 ˆ B2 ˆ B3 ˆ B4 ˆ B5 ˆ B6 ˆ B7
B0	B0 ˆ B1	B0 ˆ B1 ˆ B2 ˆ B3	B0 ˆ B1 ˆ B2 ˆ B3 ˆ B4 ˆ B5 ˆ B6 ˆ B7

and I suggest that along with the various tools and computer algorithms presented in this book that you try to come up with methods for yourself that will help you design and test digital electronic circuits more efficiently.

Quiz

1. If you had a number system that was base 5, the most significant value in a digit would be:
   (a) 6
   (b) 10
   (c) 4
   (d) 5
2. The eight bit binary equivalent to decimal 47 is:
   (a) 0010 1111
   (b) B'0010 1111'

   (c) 101111
   (d) 1011 11

3. The third most significant digit in the decimal number "1234" is:
   (a) The hundreds column
   (b) 3
   (c) 1
   (d) No digit can be the third most significant
4. To verbally tell somebody the hex number value 0x04AC you would say:
   (a) "Four-Able-Charlie"
   (b) "Hexadecimal Four-Eh-See"
   (c) "Hexadecimal Four-Apple-Charlie"
   (d) "Hexadecimal Four-Able-Charlie"
5. The decimal number "123" in hexadecimal is:
   (a) 0x0123
   (b) B'0111 1011'
   (c) 7B
   (d) 0x07B
6. The four bit hexadecimal number 0x01234 expressed in decimal is:
   (a) 1,234
   (b) 4,660
   (c) B'0001 0010 0011 0100'
   (d) 0x04D2
7. Binary coded decimal is defined as:
   (a) Ten bits providing ten different values
   (b) Four bits providing ten numeric values and six control codes
   (c) Four bits providing ten numeric values
   (d) Five bits with each bit providing two values for a total of 10
8. BCD should:
   (a) Never be used
   (b) Used with circuits that operate with base 10 numbers
   (c) Only be used when you've run out of binary chips
   (d) Used when values are not expected to exceed 9
9. B'0110' in binary, using the formula $\text{Gray code} = \text{binary}\ \hat{}\ (\text{binary} \gg 1)$ can be converted to the Gray code:
   (a) B'1010'
   (b) B'0110'

(c) B’0101’
(d) B’0111’

10. The Gray code B’0010’ corresponds to the binary value:
(a) B’0011’
(b) Unknown because more data is required
(c) B’1101’
(d) B’0010’

5

CHAPTER

Binary Arithmetic Using Digital Electronics

Before going into showing how basic binary arithmetic operations are performed in digital electronic circuits, I thought it would be useful to review how you would perform basic arithmetic operations. Before discussing how many binary arithmetic operations there are, some different characteristics of binary numbers should be discussed. I realize that much of the material in this chapter introduction is a review of work that you first did in grade school, but often when confronted with situations that require you to develop binary arithmetic operations in digital electronics, this basic information can easily be forgotten and standard devices that provide this function are often overlooked.

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

Table 5-1 Necessary decimal addition pairs to memorize.

0 + 0									
0 + 1	1 + 1								
0 + 2	1 + 2	2 + 2							
0 + 3	1 + 3	2 + 3	3 + 3						
0 + 4	1 + 4	2 + 4	3 + 4	4 + 4					
0 + 5	1 + 5	2 + 5	3 + 5	4 + 5	5 + 5				
0 + 6	1 + 6	2 + 6	3 + 6	4 + 6	5 + 6	6 + 6			
0 + 7	1 + 7	2 + 7	3 + 7	4 + 7	5 + 7	6 + 7	7 + 7		
0 + 8	1 + 8	2 + 8	3 + 8	4 + 8	5 + 8	6 + 8	7 + 8	8 + 8	
0 + 9	1 + 9	2 + 9	3 + 9	4 + 9	5 + 9	6 + 9	7 + 9	8 + 9	9 + 9

When you first learned to add decimal numbers together, you probably were required to memorize all 100 different combinations of single digit parameters when only 55 are really required. In Table 5-1, I have listed the 55 pairs which have to be memorized; the remaining 45 pairs do not have to be memorized because of the commutative law which states:

$$A + B = B + A$$

and means that the number pairs like "4 + 7" and "7 + 4" are equivalent.

The result of adding each of these two parameters produces either a single digit or double digit sum. The double digit sum indicates that the value of the result is greater than could be represented in a single digit of the number base. For decimal numbers, the maximum value that can be represented by a single digit is "9". Looking at the general case, the maximum value that can be represented by a number system is the base minus one. So, for the binary number system (base 2), the maximum value is "1"; for hexadecimal (base 16), the maximum value is "15" (or "0x0F").

The leftmost digit of a double digit number is known as the "carry" digit.

In Chapter 4, I showed how multi-digit numbers are made up of single digit values multiplied by powers of their base. Knowing the sums of the 55 addition pairs of Table 5-1, multi-digit numbers can be added together

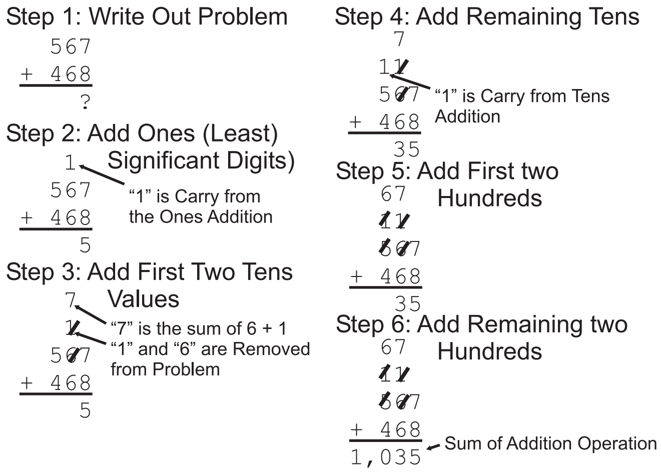
Step 1: Write Out Problem
567
+ 468
?
Step 2: Add Ones (Least) Significant Digits)
1
567
+ 468
5
"1" is Carry from the Ones Addition
Step 3: Add First Two Tens Values
7
1
567
+ 468
5
"7" is the sum of 6 + 1
"1" and "6" are Removed from Problem
Step 4: Add Remaining Tens
7
11
567
+ 468
35
"1" is Carry from Tens Addition
Step 5: Add First two Hundreds
67
11
567
+ 468
35
Step 6: Add Remaining two Hundreds
67
11
567
+ 468
1,035
Sum of Addition Operation

Fig. 5-1. The process steps used to add two three digit numbers together.

by working through pairs of numbers, as I show in Fig. 5-1. This is a rather pedantic way of showing addition and I'm sure that when you add two multi-digit numbers together, you are much more efficient, but when you were learning, this was probably the process that you went through.

While saying that you are much more efficient, it really comes down to the idea that you are able to recognize that one plus another number is the same as incrementing the other number. You are still only adding one digit at a time and the carry is "rippling" to the next significant digit. Carry "ripple" is an important concept that will be discussed in more detail in the next section.

Subtraction has many of the same issues as addition, but with some additional complexities. The first being that you cannot simplify your memorization of the 100 pairs of subtracted parameters as you did for addition; the commutative law does not apply to subtraction as it did for addition. For example,

$$6 - 4 \neq 4 - 6$$

Next, if the number being taken away is greater than the original number, the result (or "difference") could be less than 0 or "negative". There is a very big question on how to represent that negative number. Typically, it is represented as a value with a "minus" or "subtraction" sign in front of it, e.g. "-2".

The minus sign is only used when the digit cannot "borrow" from the next significant digit, as shown in Fig. 5-2. The result of 25 minus 9 is 16, with the ones borrowing 10 from the tens column (the next significant digit) to allow the operation to proceed without a negative result.

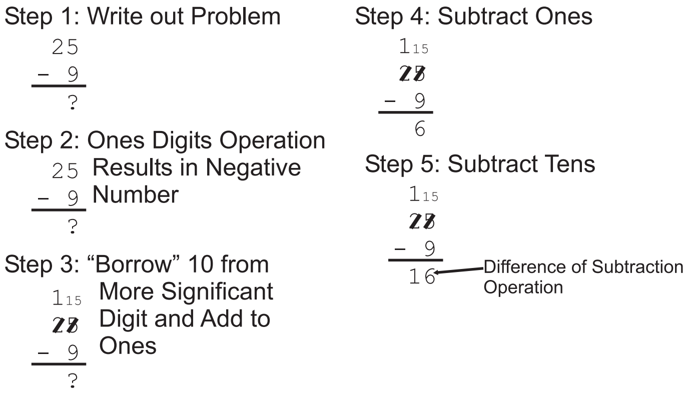
Step 1: Write out Problem
25
- 9
?
Step 2: Ones Digits Operation Results in Negative Number
25
- 9
?
Step 3: "Borrow" 10 from More Significant Digit and Add to Ones
1 15
25
- 9
?
Step 4: Subtract Ones
1 15
25
- 9
6
Step 5: Subtract Tens
1 15
25
- 9
16
Difference of Subtraction Operation

Fig. 5-2. Multi-digit subtraction with borrowing.

Subtraction can also be expressed as adding a negative value and can be written out as:

$$A - B = A + (-B)$$

This should not be a surprise to you unless you consider the following philosophical question: What would happen if infinity was arbitrarily defined as one million (1,000,000)? Instead of adding a minus sign to our value to make it negative, we could subtract it from "infinity".

For example, if we had the problem:

$$8 - 5 = 8 + (-5)$$

we could define "−5" as one million subtract 5 or "999,995". Now, going back to the addition of the negative number and substituting in "999,995" for "−5" we get:

$$\begin{aligned} 8 - 5 &= 8 + (-5) \\ &= 8 + 999{,}995 \\ &= 1{,}000{,}003 \end{aligned}$$

Since a million is defined as infinity and has no meaning, it can be taken away from the result, leaving us with the difference of 8 minus 5 being "3". This method may seem to be overly complex, but I will show you how this applies to digital electronics later in the chapter.

Like addition, the method presented here for subtraction is carried out a single digit at a time with the need to borrow from the next more significant digit being similar to passing the carry digit in addition. Like the "ripple carry" in addition, the "borrow" in subtraction can also be thought of as a "ripple" operation.

Multiplication and division have, not surprisingly, many of the same issues and when I discuss them later in this chapter, I will review them with you. Before reading the section discussing multiplication and division, I suggest that you review these operations and try to think of how they can be accomplished using digital electronics.

Adders

The circuit shown in Fig. 5-3 will add two bits together and output the sum ("S") bit along with a carry ("C") bit, if both inputs are "1" and the sum is "2", which is greater than the maximum number that can be represented by the number base (which is 1 for binary). Table 5-2 is a truth table, showing the output of each bit for different input values. You should be able to see that the "sum" bit is 1 when one or the other (but not both) of the two input bits is 1 and the "carry" bit is 1 only when both input bits are 1.

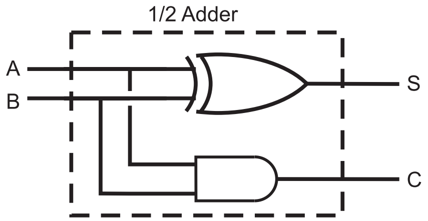

Fig. 5-3. Half adder circuit.

Table 5-2 Half adder operation truth table.

"A" input	**"B" input**	**"Sum" bit**	**"Carry" bit**
0	0	0	0
0	1	1	0
1	1	0	1
1	0	1	0

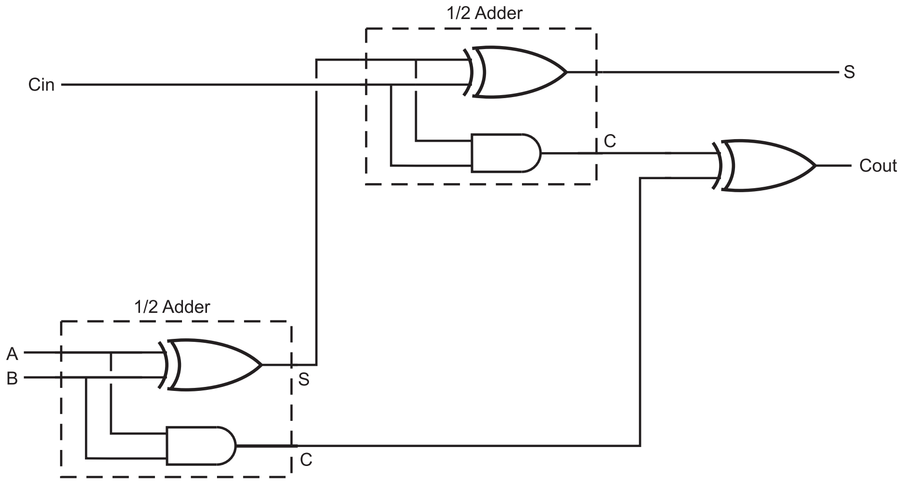

Fig. 5-4. Full adder circuit.

The adder is the first practical use most people have for the XOR gate and its function can be seen very clearly in Table 5-2 for the sum bit. Along with the XOR gate providing the function for the sum bit, you should also recognize that the carry bit is the output of a simple AND gate.

This simple digital electronic circuit is known as a ‘‘half adder’’ because it will handle half the operations required of the general case addition circuit. The ‘‘full adder’’ (Fig. 5-4) starts with a half adder and adds another bit (which is the less significant bit’s ‘‘carry’’ output) to its sum. Put another way, the full adder adds three individual bits together (two bits being the digit inputs and the third bit assumed to be carry from the addition of the next least-significant bit addition operation, known as ‘‘C_{in}’’).

You can analyze the operation of the full adder to check on its operation. The sum bit is 1 only if one or three of the input bits is 1. In the half adder, I showed that the sum bit could be written out as:

$$\mathtt{Sum} = \mathtt{A} \otimes \mathtt{B}$$

and should only be ‘‘1’’ if only one of the two inputs was 1. To understand the logic required to produce the sum bit for the three bit full adder, I created Table 5-3 in which the XOR output of the A and B inputs was given a single column entry. From the data presented in this table, you can see that the sum could be expressed as:

$$\mathtt{Sum} = (\mathtt{A} \otimes \mathtt{B}) \otimes \mathtt{C}$$

Table 5-3 Full adder sum bit operation truth table.

“A ⊗ B”	**“C_{in}” bit**	**“Sum” bit**
0	0	0
0	1	1
1	1	0
1	0	1

which, if you look back at Fig. 5-4, is exactly how it is implemented in the “full adder”.

The carry output bit is 1 if two or three of the input bits are 1. As an exercise, you may wish to create a truth table and reduce it down to see if you can match the carry gate logic of Fig. 5-3, but you can write out and reduce a sum of products equation quite easily:

$$
\begin{aligned}
\mathtt{Carry} &= (A \cdot B \cdot !C_{in}) + (!A \cdot B \cdot C_{in}) \\
&\quad + (A \cdot !B \cdot C_{in}) + (A \cdot B \cdot C_{in}) \\
&= (!C_{in} \cdot (A \cdot B)) + (C_{in} \cdot (A \cdot B)) \\
&\quad + (C_{in} \cdot (A \cdot !B) + (!C_{in} \cdot (!A \cdot B) \\
&= ((!C_{in} + C_{in}) \cdot (A \cdot B)) + (C_{in} \cdot (A \cdot !B)) \\
&\quad + (C_{in} \cdot (!A \cdot B)) \\
&= (A \cdot B) + (C_{in} \cdot ((!A \cdot B) + (A \cdot !B))
\end{aligned}
$$

Knowing that

$$((!A \cdot B) + (A \cdot !B)) = A \otimes B$$

the equation for the carry output of the full adder can be written out as:

$$\mathtt{Carry} = (A \cdot B) + ((!A \otimes B) \cdot C_{in})$$

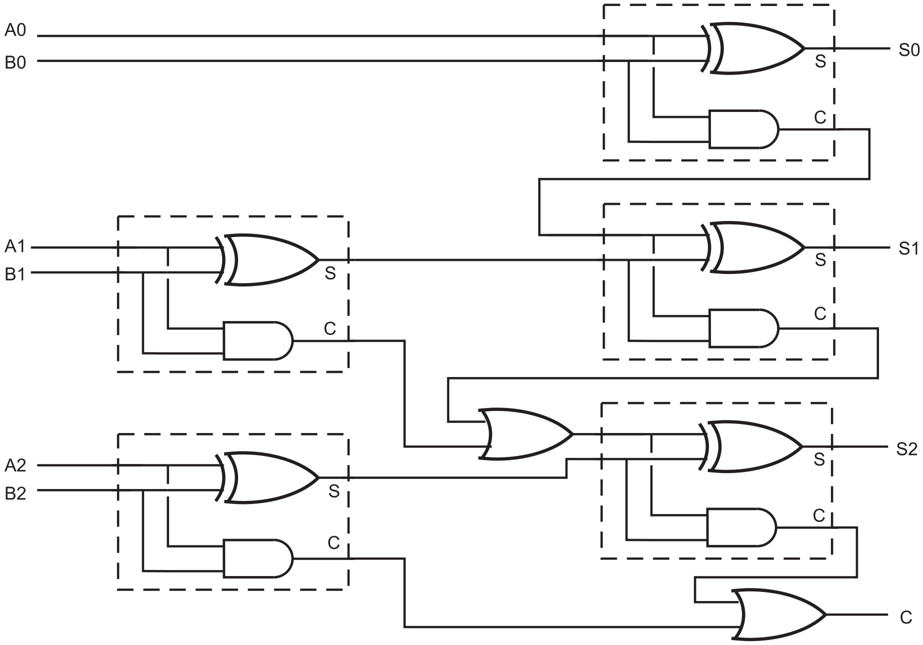

Fig. 5-5. Three bit "ripple adder" circuit.

which is *exactly* the carry logic circuit shown in Fig. 5-2. This type of analysis is useful to do when you are trying to puzzle out what a circuit is doing or to confirm that it is doing exactly what you expect it to do. It is also good practice of using the logic equation optimization skills first presented in Chapter 2.

Multiple full adders can be chained together (like in Fig. 5-5) to produce a multi-bit adder in which the carry results for each bit "ripples" through the various adder circuits. For most applications, this "ripple carry adder" can be used safely, but in something like your PC's processor, where quite a few bits are required and the adder is expected to execute quickly, the time required for the carry to ripple through the adders is prohibitive.

The solution to this problem is the "carry look-ahead" adder in which each bit takes not only the appropriate bits for input but also all the least significant bits that can affect the bit. The length of time the carry look-ahead adder needs to produce a sum is generally independent of the number of bits in the operation (unlike the time the ripple adder requires to produce a sum which is a function of the number of bits). Table 5-4 lists the different inputs and expected outputs for a three bit carry look-ahead adder. Reducing the

Table 5-4 Carry look-ahead adder input/output truth table.

A2	B2	A1	B1	A0	B0		S0	S1	S2	Carry
0	0	0	0	0	0		0	0	0	0
0	0	0	0	0	1		1	0	0	0
0	0	0	0	1	1		0	1	0	0
0	0	0	0	1	0		1	0	0	0
0	0	0	1	1	0		1	1	0	0
0	0	0	1	1	1		0	0	1	0
0	0	0	1	0	1		1	1	0	0
0	0	0	1	0	0		0	1	0	0
0	0	1	1	0	0		0	0	1	0
0	0	1	1	0	1		1	0	1	0
0	0	1	1	1	1		0	1	1	0
0	0	1	1	1	0		1	0	1	0
0	0	1	0	1	0		1	1	0	0
0	0	1	0	1	1		0	0	1	0
0	0	1	0	0	1		1	1	0	0
0	0	1	0	0	0		0	1	0	0
0	1	1	0	0	0		0	1	1	0
0	1	1	0	0	1		1	1	1	0
0	1	1	0	1	1		0	0	0	1
0	1	1	0	1	0		1	1	1	0
0	1	1	1	1	0		1	0	0	1

(*continued*)

Table 5-4 Continued.

A2	B2	A1	B1	A0	B0		S0	S1	S2	Carry
0	1	1	1	1	1		0	1	0	1
0	1	1	1	0	1		1	0	0	1
0	1	1	1	0	0		0	0	0	1
0	1	0	1	0	0		0	1	1	0
0	1	0	1	0	1		1	1	1	0
0	1	0	1	1	1		0	0	0	1
0	1	0	1	1	0		1	1	1	0
0	1	0	0	1	0		1	0	1	0
0	1	0	0	1	1		0	1	1	0
0	1	0	0	0	1		1	0	1	0
0	1	0	0	0	0		0	0	1	0
1	1	0	0	0	0		0	0	0	1
1	1	0	0	0	1		1	0	0	1
1	1	0	0	1	1		0	1	0	1
1	1	0	0	1	0		1	0	0	1
1	1	0	1	1	0		1	1	0	1
1	1	0	1	1	1		0	0	1	1
1	1	0	1	0	1		1	1	0	1
1	1	0	1	0	0		0	1	0	1
1	1	1	1	0	0		0	0	1	1
1	1	1	1	0	1		1	0	1	1

(*continued*)

Table 5-4 Continued.

A2	B2	A1	B1	A0	B0		S0	S1	S2	Carry
1	1	1	1	1	1		0	1	1	1
1	1	1	1	1	0		1	0	1	1
1	1	1	0	1	0		1	1	0	1
1	1	1	0	1	1		0	0	1	1
1	1	1	0	0	1		1	1	0	1
1	1	1	0	0	0		0	1	0	1
1	0	1	0	0	0		0	1	1	0
1	0	1	0	0	1		1	1	1	0
1	0	1	0	1	1		0	0	0	1
1	0	1	0	1	0		1	1	1	0
1	0	1	1	1	0		1	0	0	1
1	0	1	1	1	1		0	1	0	1
1	0	1	1	0	1		1	0	0	1
1	0	1	1	0	0		0	0	0	1
1	0	0	1	0	0		0	1	1	0
1	0	0	1	0	1		1	1	1	0
1	0	0	1	1	1		0	0	0	1
1	0	0	1	1	0		1	1	1	0
1	0	0	0	1	0		1	0	1	0
1	0	0	0	1	1		0	1	1	0
1	0	0	0	0	1		1	0	1	0
1	0	0	0	0	0		0	0	1	0

information from this table, I have listed the equations for the three sum bits and the carry bit below:

$$\begin{aligned}
\texttt{S0} &= (\texttt{A0}\otimes\texttt{B0})\\
\texttt{S1} &= ((\texttt{A0}\otimes\texttt{B0})\cdot(\texttt{A1}\otimes\texttt{B1}))\\
&\quad+((\texttt{A1}\otimes\texttt{B1})\cdot!(\texttt{A0}+\texttt{B0}))\\
&\quad+(!(\texttt{A1}\otimes\texttt{B1})\cdot(\texttt{A0}\cdot\texttt{B0}))\\
\texttt{S2} &= ((\texttt{A2}\otimes\texttt{B2})\cdot!(\texttt{A2}\cdot\texttt{B2}))\\
&\quad+(!(\texttt{A2}\otimes\texttt{B2})+(\texttt{A1}\cdot\texttt{B1}))+(!(\texttt{A2}\cdot\texttt{B2})\\
&\quad+(\texttt{A1}\otimes\texttt{B1})\cdot(\texttt{A0}\cdot\texttt{B0}))\\
\texttt{Carry} &= (\texttt{A2}\cdot\texttt{B2})+((\texttt{A2}\otimes\texttt{B2})\cdot(\texttt{A1}\cdot\texttt{B1}))\\
&\quad+(\texttt{A2}\otimes\texttt{B2})\cdot(\texttt{A1}\otimes\texttt{B1})\cdot(\texttt{A0}\cdot\texttt{B0}))
\end{aligned}$$

It was a major effort on my part to reduce the equations for each sum bit and the carry bit. To do this, I used the truth table reduction method discussed in Chapter 2. To reduce the number of terms in the resulting sum of products equations, I first deleted all the instances where the specific bit was not "1"–in every case, this reduced the number of instances by half. Next, I worked at combining instances that were similar and found that rather than combining "don't care" bits, I found a number of places where two bits were XORed together. In the resulting equations, I kept the "XOR" terms in, even though when the "technology optimization" stage of the development effort is completed, these gates will be reduced to the technology's basic gates.

If you read through the equations and try to understand them, you will find that they do make a kind of sense. Obviously, as more bits are added to the carry look-ahead adder, the circuit becomes much more complex. Despite this complexity, the carry look-ahead is the most efficient way to provide an adder circuit for large bit words in fast applications.

Subtraction and Negative Numbers

As you might expect, binary subtraction has many of the same issues as addition, along with a few complexities that can make it harder to work with. In this section, I will introduce some of the issues in implementing a practical "subtracter" as well as look at some ways in which subtraction can be implemented easily with an existing addition circuit.

To make sure we're talking the same language, I want to define the terms that I will be using to describe the different parts of the subtraction operation. The "horizontal" arithmetic equation:

```
Result = A - B
```

can be written out "vertically" as:

```
      A  <- Minuend
     -B  <- Subtrahend
 ------
 Result  <- Difference
```

The "minuend" and "subtrahend" terms are probably something that you forgot that you learned from grade school. I use them here because they are less awkward than referring to them as the "value to be subtracted from" and the "value subtracted". The term "difference" as being the result of a subtraction operation is generally well understood and accepted.

When you carry out subtraction operations, you do it in a manner that is very similar to how you carry out addition; each digit is done individually and if the digit result is less than zero, the base value is "borrowed" from the next significant digit. With the assumption that subtraction works the same way as addition, you could create a "half subtracter", which is analogous to the half adder and could be defined by the truth table shown in Table 5-5

The "difference" bit is simply the minuend and subtrahend XORed together, while the "borrow" bit (decrementing the next significant digit) is only true if the minuend is 0 and the subtrahend is 1. The borrow bit can be defined as the inverted minuend ANDed with the subtrahend. The equations

Table 5-5 "Half subtracter" defining truth table.

Minuend	Subtrahend	Difference	Borrow
0	0	0	0
0	1	1	1
1	1	0	0
1	0	1	0

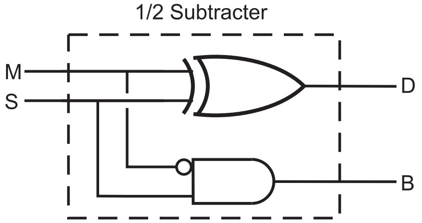

Fig. 5-6. Half subtracter circuit.

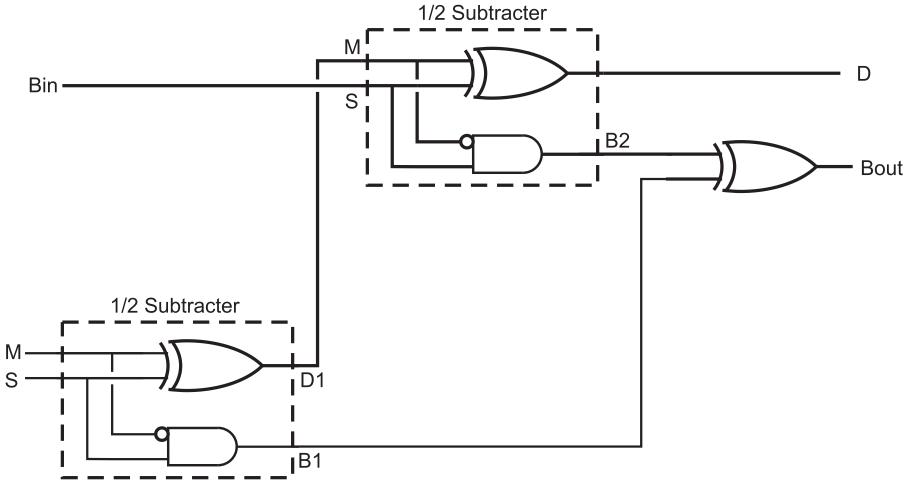

Fig. 5-7. Full subtracter circuit.

for the half subtracter are listed below and the subtracter building block is shown in Fig. 5-6.

$$\texttt{Difference} = \texttt{Minuend} \otimes \texttt{Subtrahend}$$
$$\texttt{Borrow} = !\texttt{Minuend} \cdot \texttt{Subtrahend}$$

The small circle on the single input indicates that the value is inverted before being passed into the gate. This convention avoids the need for putting a full inverter symbol in the wiring diagram of a digital circuit and is often used in chip datasheets to indicate inverted inputs to complex internal functions.

Two half subtracters can be combined into a "full subtracter", just as two half adders can be combined to form a full adder (Fig. 5-7). In Fig. 5-7, I have labeled the two half subtracters, so that their operation can be listed in Table 5-6, to test the operation of the full subtracter.

Table 5-6 Full subtracter operation truth table.

Bin	Minuend ("M")	Subtrahend ("S")		D1	B1	D	B2	Bout
0	0	0		0	0	0	0	0
0	0	1		1	1	1	0	1
0	1	1		0	0	0	0	0
0	1	0		1	0	1	0	0
1	1	0		1	0	0	0	0
1	1	1		0	0	1	1	1
1	0	1		1	1	0	0	1
1	0	0		0	0	1	1	1

Like the ripple adder, full subtracters can be chained together to create a multi-bit subtracter circuit (Fig. 5-8) and a "borrow look-ahead" (to coin a phrase) subtracter could be designed, but instead of going through the pain of designing one, there is another option and that is to add the negative of the subtrahend to the minuend.

In the introduction to this chapter, I introduced the idea of negative numbers as being the value being subtracted from an arbitrary large number and showed an example that produced "−5" in a universe where infinity was equal to one million. When you first went through this example, you might have thought that this was an interesting mathematical diversion and an illustration as to how negative and positive numbers converge when they approach infinity. This concept, while seemingly having little application in the "real world", is very useful in the digital domain.

In the digital domain, the term "infinity" can be replaced with "word size" and if the most significant bit of the word is considered to be the "sign" bit, positive and negative numbers can be expressed easily. In Table 5-7, I have listed the decimal, hex as well as the positive and negative values which take into account that a negative number can be written as:

$$-\mathtt{A} = \mathtt{Word\ Size} - \mathtt{A}$$

This negative value is known as a "two's complement" negative number and is the most commonly used negative bit format used. There is a "one's

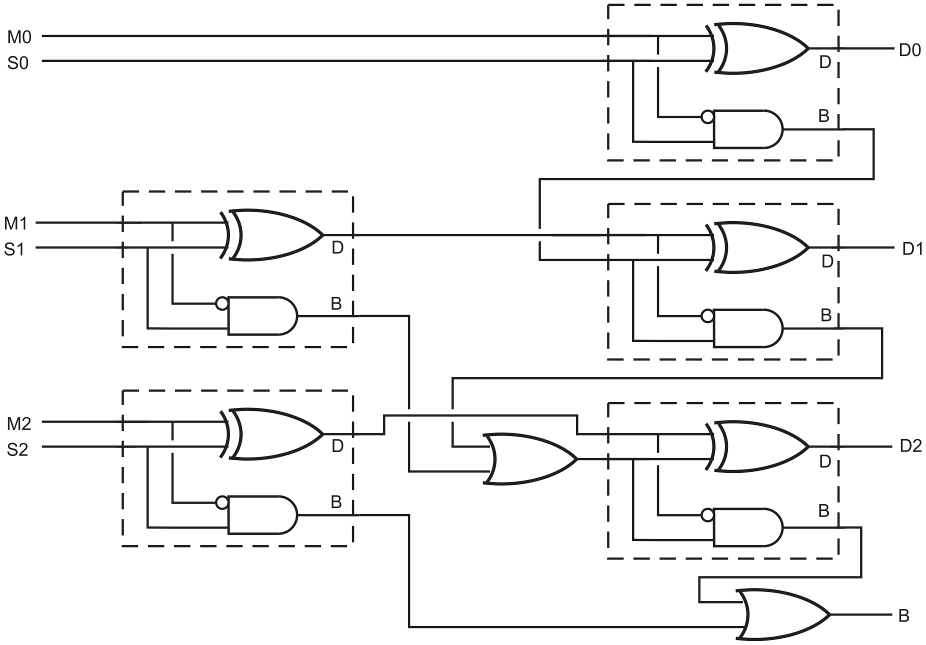

Fig. 5-8. Three bit "ripple subtracter" circuit.

complement" number format, but it does not avail itself as efficiently as two's complement to allow for easier subtraction and addition of negative numbers.

Looking at the formula above, you are probably confused as to why it would be used because it requires both a subtraction operation as well as an addition operation to carry out one subtraction operation. Negating a number in two's complement format does not actually require a subtraction operation; it can be done by inverting each bit (XORing each bit with 1) and then incrementing the result. Using the values of Table 5-7, you can demonstrate how a positive value is negated.

For example, to negate the value "5", the following steps are used:

1. Each bit of the number is XORed with "1". B'0101' becomes B'1010'.
2. The XORed result is incremented. B'1010' becomes B'1011', which is "−5".

The opposite is also true: the individual bits can be inverted and the result incremented to convert a negative two's complement value to a positive.

Once the value has been negated, it can be simply added to the other parameter, as I show in Fig. 5-9. There are three things that you should be

Table 5-7 Different ways of representing a four bit number.

Binary value	Decimal value	Hex value	Two's complement value
B'0000'	0	0x00	0
B'0001'	1	0x01	1
B'0010'	2	0x02	2
B'0011'	3	0x03	3
B'0100'	4	0x04	4
B'0101'	5	0x05	5
B'0110'	6	0x06	6
B'0111'	7	0x07	7
B'1000'	8	0x08	−8
B'1001'	9	0x09	−7
B'1010'	10	0x0A	−6
B'1011'	11	0x0B	−5
B'1100'	12	0x0C	−4
B'1101'	13	0x0D	−3
B'1110'	14	0x0E	−2
B'1111'	15	0x0F	−1

aware of before leaving this section. The first is the use of the "V" shaped mathematical function symbols in Fig. 5-9; these symbols indicate that two parameters are brought together to form one output. I use this symbol when a group of bits (not just one) are passing through the same operation.

You might be wondering why instead of simply inverting the individual bits of the value to be converted to a negative two's complement value, I XOR the bits with the value 1. The reason for doing this is in the interests

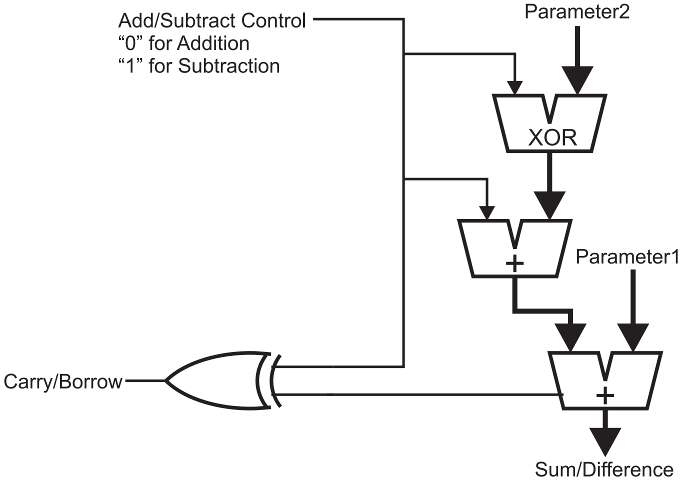

Fig. 5-9. Integrated adder/subtracter circuit.

of practicality and looking ahead. In Fig. 5-9, I show a circuit in which two parameters can be added together or one can be subtracted by the other – the "switch" control for which operation is selected. If a 1 is passed to the "Parameter2" circuitry, each bit of Parameter2 is XORed with 1, inverting each bit and a 1 is passed to the Parameter2 adder, which increments the value. If a zero is passed to the Parameter2 circuitry, the bits of Parameter2 are not inverted and zero is added to the output of the XOR function, resulting in an unchanged value of Parameter2 being passed to the adder with Parameter1. To net it out, if a "1" is passed to this circuit, Parameter2 is subtracted from Parameter1; if a "0" is passed to the circuit, the two parameters are added together.

The last point to note is that the "carry" output of the final adder is a negated "borrow" output when the subtraction operation is taking place. To integrate the operation of the "carry/borrow" bit with the add/subtract switch bit, this bit is set when a carry or borrow of the next significant word is required, regardless of the operation.

Magnitude Comparators and Bus Nomenclature

Along with being able to add and subtract binary values, you will find the need to compare binary values to determine whether or not a value is less than, equal to or greater than another value. Just as if this were a programming requirement, to test two binary values together, you would subtract one from the other and look at the results. An important issue when comparing a value made up of multiple bits is specifying how it is to be represented in logic drawings and schematic diagrams. In the previous section I touched on both these issues, in this section, I want to expand upon them and help you to understand a bit more about them.

When you are comparing two binary values, you are comparing the *magnitude* of the values, which is where the term “magnitude comparator” comes from. The typical magnitude comparator consists of two subtracters which either subtract one value from another and vice versa or subtract one value from another and then compare the result to zero. In either case, the magnitude comparator outputs values indicating which value is greater than the other or if they are equal.

Figure 5.10 shows a basic comparator, which consists of two subtracters utilizing the negative addition discussed in the previous section. The differences are discarded, but the !borrow outputs are used to determine if the negative value is greater. If the !borrow outputs from the two

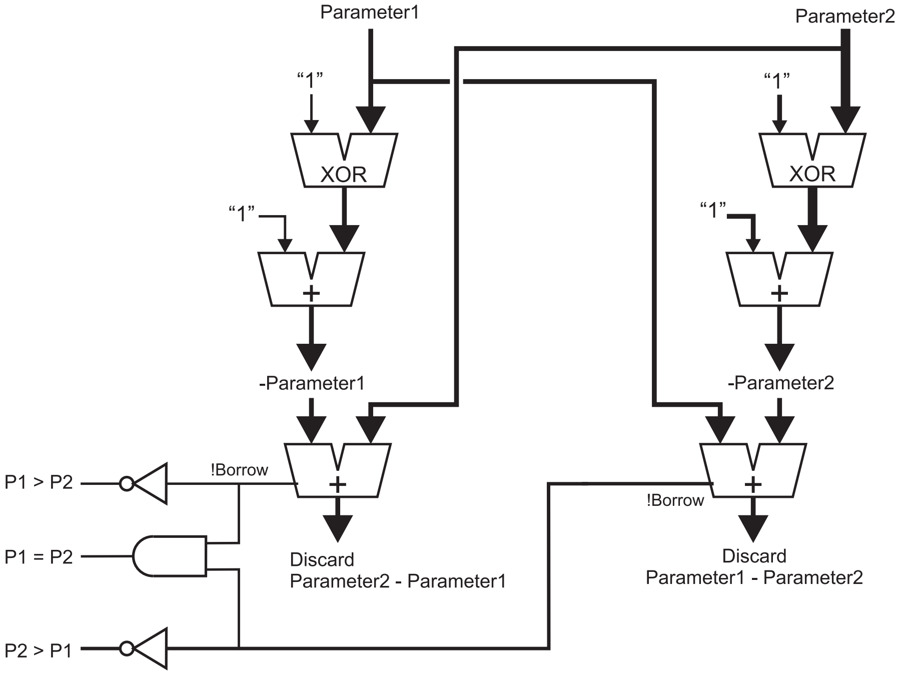
Parameter1
Parameter2
“1”
XOR
“1”
+
-Parameter1
“1”
XOR
“1”
+
-Parameter2
P1 > P2
!Borrow
+
Discard
Parameter2 - Parameter1
P1 = P2
!Borrow
+
Discard
Parameter1 - Parameter2
P2 > P1

Fig. 5-10. Comparator built from two subtracters.

subtracters are both equal to “1”, then it can be assumed that the two values are equal.

If one value is subtracted from the other to determine if one is lower than the other and if the value is not lower (i.e. !borrow is not zero), the result can then be compared to zero to see if the value is greater than or equal to the other. This method is probably less desirable because it tends to take longer to get a valid result and the result outputs will be valid at different times. Ideally, when multiple outputs are being produced by a circuit, they should all be available at approximately the same time (which is the advantage of the two subtracter circuit shown in Fig. 5-10 over this one).

If you are working with TTL and require a magnitude comparator, you will probably turn to the 7485, which is a four bit magnitude comparator consisting of two borrow look-ahead subtracters to ensure that the outputs are available in a minimum amount of time and are all valid at approximately the same time.

In Fig. 5-10 (as well as the multi-bit subtracter shown in the previous section), I contained related multiple bits in a single, thick line. This very common method of indicating multiple related bits is often known as a “bus”. Other methods include using a line of a different color or style. The advantage of grouping multiple bits that function together like this should be obvious: the diagram is simpler and it is easier to see the path that related bits take.

When I use the term “related bits”, I should point out that this does not only include the multiple bits of a binary value. You may have situations where busses are made up of bits which are not a binary value, but perform a similar function within the circuit. For example, the memory control lines for a microprocessor are often grouped together as a bus even though each function is provided by a single bit (memory read, memory write, etc.) and they are active at different times.

As well as indicating a complete set of related bits, a bus may be broken up into subsets, as shown in Fig. 5-11. In this diagram, I have shown how two four bit magnitude comparators can be “ganged” together to provide a comparison function on eight bits. The least significant four bits are passed to the first magnitude comparator and the most significant four bits are passed to the second magnitude comparator. The bits are typically listed as shown in Fig. 5-11, with the most significant bit listed first and separated from the least significant bit by a colon. In very few cases will you see the width of the bus reduced to indicate a subset of bits as I have done in Fig. 5-11; most design systems will keep the same width for a bus, even if one bit is being used in it.

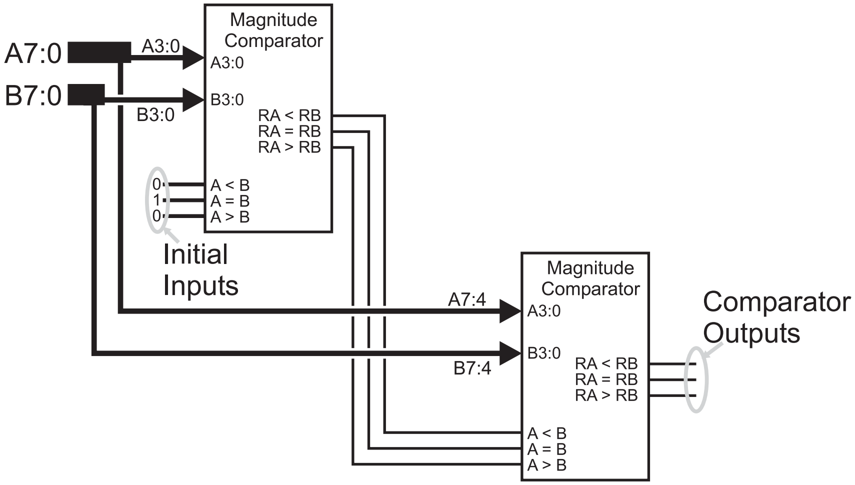
A7:0
B7:0
A3:0
B3:0
Magnitude Comparator
A3:0
B3:0
RA < RB
RA = RB
RA > RB
0
1
0
A < B
A = B
A > B
Initial Inputs
A7:4
B7:4
Magnitude Comparator
A3:0
B3:0
RA < RB
RA = RB
RA > RB
A < B
A = B
A > B
Comparator Outputs

Fig. 5-11. Cascading two four bit comparators.

Before going on, I want to make some comments about Fig. 5-11 as it provides a function that is often required when more bits must be operated on than is available by basic TTL or CMOS logic chips. To carry out the magnitude comparison operation on eight bits, I used two four bit magnitude comparator chips (modeled on the 7485) with the initial state inputs (marked ‘‘Initial Inputs’’ on Fig. 5-11) to start the chip off in the ‘‘neutral’’ state as if everything ‘‘upstream’’ (before) was equal to each other and the chip’s bits as well as any ‘‘downstream’’ (after) bits will determine which value is greater or if the two values are equal. This is a typical method for combining multiple chips to provide the capability to process more bits than one chip is able to.

Multiplication and Division

As you’ve read through this chapter, you should have noticed that there are usually many different ways of implementing different digital electronics functions. Each of the different implementation methods have their own advantages and tradeoffs – it is up to the application designer to understand what are the important ones. Nowhere is this more true than when you start discussing multiplication and division; there are a number of different methods of performing these arithmetic operations, each with their own characteristics.

Off the top of my head, I can come up with five different ways to multiply two binary numbers together. Before listing the different methods, I should make sure that I have defined the terms used in multiplication. The "multiplicand" is the value that is multiplied by the "multiplier" and typically remains static. The "multiplier" is the number of times that the multiplicand is added together to form the result or "product".

It should go without saying that if you had to multiply by a power of 2 (i.e. 1, 2, 4, 8, 16, etc.) a true multiplication operation is not required at all; the operation can be accomplished by shifting the multiplicand. For example to multiply a value by 4, you would simply shift the value to the left two times. Division is the same, except the value is shifted to the right.

Understanding the basic terms "multiplier" and "multiplicand" leads to a second method to implement a multiplication function in software – the multiplicand is added multiplier number of times to the product. It can be written out in "C" as:

```
Product = 0;                      //Clear result accumulator
for (i = 0; i < Multiplier; i++)
                                 //Repeatedly add Multiplicand
    Product = Product + Multiplicand;
```

This method is painfully slow (especially for large multiplier values) and is difficult to implement in combinatorial digital logic. It is also different from the method which you were taught in school in which the multiplicand is shifted up by the radix for each digit of the multiplier. Using this method, "123" decimal is multiplied by "24" decimal in the format:

```
     123
×     45
    ----
     615
+   492   < -Sum Products of each digit
    ----
    5535
```

In the first line of the solution, I multiplied the multiplicand "123" by the least significant digit of the multiplier followed by multiplying the multiplicand by 10 (the radix) followed by the next significant digit of the multiplier. Once the multiplicand has been multiplied by each digit of the multiplier (along with the appropriate multiplication of the digit position), each product is added together to get the final result.

This method lends itself extremely well to working within binary systems. Rather than multiplying the multiplicand repeated by the radix for each digit, the multiplicand is simply shifted to the right (which is multiplying by two) for each bit of the multiplier. If the multiplier bit is zero, then the value added to the product is zero. The binary multiplication operation for 123 by 45 is:

```
            B′01111011′
          × B′00101101′
           B′01111011′      < – Bit 0 of Multiplier is not Zero
          B′00000000′       < –Bit 1 of Multiplier is Zero
         B′01111011′        < –Bit 2
        B′01111011′         < –Bit 3
       B′00000000′          < –Bit 4
      B′01111011′           < –Bit 5
     B′00000000′            < –Bit 6
 +  B′00000000′             < –Bit 7
  B′0001010110011111′       < – Product (5,535 Decimal)
```

This is much more efficient than the previous version in terms of execution time and not significantly more complex than the other version. The “C” code that implements it is:

```
Product = 0;                              //Clear result accumulator
for (i = 0; i < log 2 (Multiplier); i ++)
{
     if ((Multiplier & 1) == 1)       //If LS Bit Set, Add Current
     Product = Product + Multiplicand;           //to Product
     Multiplier = Multiplier >>1; //Shift down Multiplier
     Multiplicand = Multiplicand << 1;    //Shift up
                                                 //Multiplicand
} //  rof
```

The first version is known as “Order n” because it loops once for each value of the multiplier. The shift and add version shown directly above is known as “Order $\log_2$” because it executes the $\log_2$ of the word size of the multiplier. For the eight bit multiplication example shown here, for the first method, up to 255 loops (and addition operations) may have to be executed. For the second example, a maximum of 8 loops (one for each bit, which is a simple way to calculate $\log_2$ of a number) is required.

The final method of multiplying two numbers together is known as “Booth’s algorithm” and looks for opportunities to reduce the number of addition operations that are required by rounding the multiplier up and then adjusting the product by subtracting the multiplier’s zero bits that were ignored. For the example given in this section (123 multiplied by 45), Booth’s algorithm would recognize that 45, B’00101101’ rounds up to 64 (B’01000000’). Multiplying a binary number by 64 is the same as shifting left six times.

To adjust the product, the basic multiplicand (multiplied by 1) along with all the instances where the multiplier has bit value of zero (in this case, bits one and four) have to be taken away from the rounded up value. For this example, the multiplication operation would look like:

$$
\begin{array}{rr}
\multicolumn{2}{l}{(\texttt{B}'\texttt{01111011}' << 6) - \texttt{B}'\texttt{01111011}' - (\texttt{B}'\texttt{01111011}' << 1)} \\
\multicolumn{2}{l}{\quad - (\texttt{B}'\texttt{01111011}' << 4)} \\
= & \texttt{B}'\texttt{01111011000000}' \\
- & \texttt{B}'\texttt{01111011}' \\
- & \texttt{B}'\texttt{011110110}' \\
- & \texttt{B}'\texttt{011110110000}' \\
\hline
 & \texttt{B}'\texttt{01010110011111}'
\end{array}
$$

which is the same result as we got above for a bit less work. Booth’s algorithm can produce a product in fewer, the same or more addition/subtraction operations as the previous method, so care must be taken to make sure that it is only used when it is going to provide an advantage.

Each of the three methods presented so far requires the ability to “loop” through multiple iterations. This is a problem for most digital electronic circuits, as it not only requires a “clock” to synchronize the operations but it will also most likely take up more time than is desired. When a digital logic multiplier is designed, it typically looks something like Fig. 5-12. This circuit is wired to add all the multiplicand bits together for each possible multiplier values.

The multiplier bits are taken into account by ANDing them with each of the multiplicand bits. If a multiplier bit is zero, then the shifted up multiplicand bits are not passed to the multi-bit adders.

There are a couple points about the multi-bit adders that you should be aware of. The first is that the maximum number of input bits for the adders used in the multiplier circuit is the number of bits in the multiplicand plus the $\log_2$ value of the multiplier. Secondly, as drawn, the adders are connected in a “ripple” configuration – a commercial circuit would probably wire the

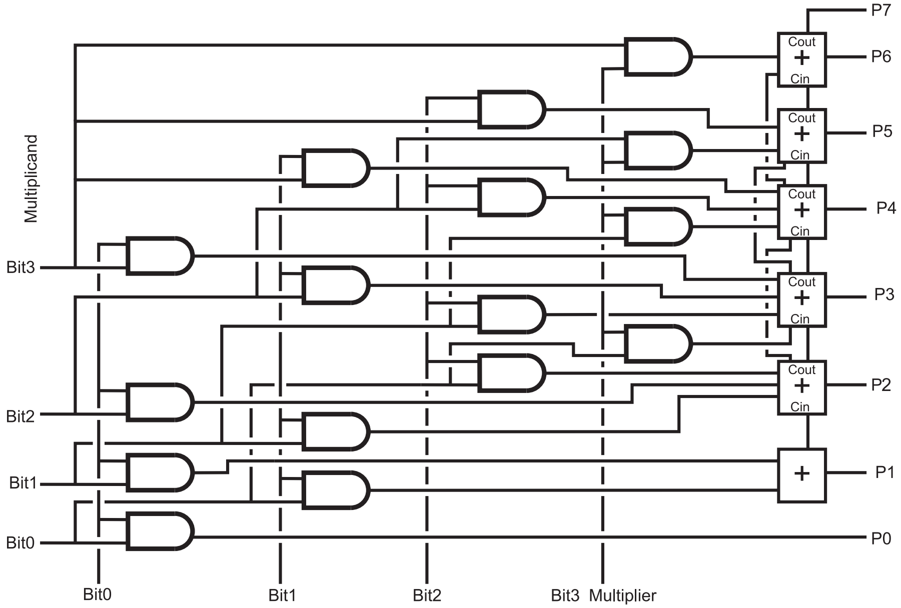

Fig. 5-12. Four bit multiplier circuit.

adders together as a carry look-ahead to minimize the time required for the multiplication operation to take place.

Before leaving the topic of multiplication, I should point out that all the methods presented here will handle multiplication of two's complement negative numbers "natively". This is to say that no additional gates must be added to support the multiplicand or multiplier being negative.

Division is significantly more difficult to implement and is very rarely implemented in low-cost devices. Handling negative values considerably complicates the division operation and in this section, as well as most commercial solutions, negative values cannot be used for the dividend or divisor. To avoid the hardware complexities of division, software intensive solutions are normally done such as a repeated subtraction:

```
Quotient = 0;                      //Clear result accumulator
while (Dividend >= Divisor)        //Repeat Subtraction
{
    Dividend = Dividend - Divisor;
    Quotient = Quotient + 1;
}  //  elihw
```

The bit shifting method shown for multiplication can also be used, but before comparisons can start, the divisor should be shifted up the word size of the dividend. To follow the bit shifting division code listed below you might want to do a thought experiment and single step through it with arbitrary values to see exactly how it works:

```
Quotient = 0;                  //Clear result accumulator
Divisor = Divisor << DividendWordSize;
DivisorPos = 1<<DivendendWordSize;
                               //Current Divisor Bit Shift
while (DivisorPos != 0)
                            //Repeat While Divisor Shifted
{
if (Dividend >= Divisor)    //Can you take away current
{
   Dividend = Dividend - Divisor;
   Quotient = Quotient + DivisorPos;
} //  fi
Divisor = Divisor >> 1;          //Shift Down Divisor
DivisorPos = DivisorPos >> 1;
} //  elihw
```

At the end of both these division algorithms, “Quotient” contains the quotient of the division operation and “Dividend” contains the remainder.

The bit shifting division algorithm could be implemented using digital electronic gates as I demonstrated for the bit shifting multiplication algorithm, but you will find that it is quite a bit more complex than the bit shifting multiplication application in Fig. 5-12. This does not mean that there are some tricks you cannot implement if a division operation is required.

For example, if you wanted to divide a value by a known, constant value, you could multiply it by its fraction of 256 (rather than the typical 1) and then divide by 256 (which is accomplished by shifting right by eight bits). For example, dividing 123 by 5 would be accomplished by multiplying 123 by 51 (256 divided by 5) and shifting the product (6,273) to the right by 8 bits to get the result 24. While this method seems like it’s complex, it is actually quite easy to implement in digital electronics.

Quiz

1. 6 − 5 is the same as:
   (a) 6 + (−5)
   (b) 5 − 6
   (c) 999999
   (d) B'1111 1111'
2. In a universe where infinity (the highest possible number) is one million (1,000,000); "−11" could be represented as:
   (a) Only −11
   (b) 999,988
   (c) 999,989
   (d) 89
3. A "half adder":
   (a) Can perform an addition operation on two bits
   (b) Can add half the bits together of an addition operation
   (c) Combines the "carry" outputs of a "full adder" to produce the correct result
   (d) Is built from half the number of gates as a full adder
4. A "ripple adder" is not used in a PC or workstation processor because:
   (a) Its complexity can affect the operation of other arithmetic functions
   (b) The result is often wrong by one or two bits
   (c) The delay required for the signal to pass through the gates can be unacceptably long
   (d) It cannot handle the 32 or 64 bit addition operations required
5. B'10' − B'01' passed through two full subtracters produces the result:
   (a) Cannot be done because a borrow operation is required
   (b) B'01' with borrow = 1
   (c) B'10' with borrow = 0
   (d) B'01' with borrow = 0
6. Converting the four bit, two's complement value "−4" to a positive number by inverting each bit and incrementing the result produces the bit pattern:
   (a) B'0100'

   (b) Which is five bits long and is invalid
   (c) B'0011'
   (d) B'1100'

7. Busses are made up of:
   (a) Multiple bits of a single value
   (b) Multiple bits passing to the same subsystem of the application
   (c) The highest speed signals in the application
   (d) Related bits

8. Multiplying two four bit numbers by repeated addition:
   (a) Will require up to 4 addition operations
   (b) Will require up to 15 addition operations
   (c) Cannot be implemented in digital electronics
   (d) Is the fastest way of performing binary multiplication

9. Multiplying a binary number by 16 can be accomplished by:
   (a) Clearing the least significant four bits
   (b) Shifting left four bits
   (c) Shifting right four bits
   (d) Setting the least significant four bits

10. Dividing an eight bit value by the constant value 6 is best accomplished by:
    (a) Using the repeated subtraction method
    (b) Using the bit shifting method
    (c) Shifting the value to the right by two and then shifting the value to the right by 1 and adding the two values.
    (d) Multiplying by 256/6 (42) and shifting the product to the right by 8 bits

6

CHAPTER

Practical Combinatorial Circuit Implementation

When you are designing your first application that is built from digital electronics, you will probably feel like you have just joined a never-ending role-playing game in which all the other players know more than you do. Later in the book, I will present some ideas on how to read a datasheet and what to look for in it, but for now, I would like to discuss a number of the options that you should be aware of and are thinking about when you first start designing your application.

Using the role-playing game analogy with digital electronics may seem to be facetious, but there are actually a lot of similarities that you should be aware of. First and foremost, each digital electronic chip that you can choose from has a number of characteristics that you will have to be aware of and

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

Table 6-1 Important characteristics of a digital electronic chip.

Characteristic	Comments
Function	Gate type, chip function
# Bits	The number of bits per gate input or number of bits used by the function
# Gates/functions	What does the chip do and how many are there
Technology	Electronic standard chip is implemented
Output type	Gate output type
Dependencies	Issues to be aware of
Manufacturer	Who makes the chip/where can it be purchased

choose from when you are specifying the parts used in your application. When choosing between the parts, it might be a good idea to come up with a card, similar to the cards used in role-playing games to explain the different characters, characteristics and strengths and weaknesses. A sample card for a digital electronic device might look something like Table 6-1.

The "function" of the chip is a brief description of the gates provided by the chip or the digital logic function provided by the chip (such as an adder or a magnitude comparator). At this point in the book, you might feel that it is sufficient to specify a chip for a needed function, but the following characteristics are critical for you to understand that you need to be able to select the right chip for the right application.

When I have presented simple gates, they have all (with the exception of the NOT gate) two inputs. Along with two inputs, there are a number of different inputs for a variety of different types of gates. For example, in standard TTL, you can get NAND gates with two, three, four and eight inputs. Four and eight bit adders as well as different chips with different bit counts are also available. When selecting a chip for an application, you should be cognizant of the bit options that are available to minimize the number of chips required.

The basic TTL chips have four two input gates and six one input gates, but if the number of bits changes, then the number of gates within the chip changes (or the plastic package type and the number of pins changes). As surprising as it seems, many complex functions can have more than one built into the package. Like the number of bits, the number of functions within the chip will help you plan out how many chips you will need in the application.

So far in the book, I have really just indicated that there are two types of technology used for standard logic devices. In actuality, there are dozens and in Table 6-2, I have listed the most popular ones with their input, output and operating characteristics. For the different varieties of "TTL", "C", "AC"

Table 6-2 Digital logic technologies with electrical characteristics.

Chip type	Power supply	Gate delay	Input threshold	"0" output	"1" output	Output sink
TTL	Vcc = 4.5 to 5.5 V	8 ns	N/A	0.3 V	3.3 V	12 mA
L TTL	Vcc = 4.5 to 5.5 V	15 ns	N/A	0.3 V	3.4 V	5 mA
LS TTL	Vcc = 4.5 to 5.5 V	10 ns	N/A	0.3 V	3.4 V	8 mA
S TTL	Vcc = 4.75 to 5.25 V	5 ns	N/A	0.5 V	3.4 V	40 mA
AS TTL	Vcc = 4.5 to 5.5 V	2 ns	N/A	0.3 V	Vcc – 2 V	20 mA
ALS TTL	Vcc = 4.5 to 5.5 V	4 ns	N/A	0.3 V	Vcc – 2 V	8 mA
F TTL	Vcc = 4.5 to 5.5 V	2 ns	N/A	0.3 V	3.5 V	20 mA
C CMOS	Vcc = 3 to 15 V	50 ns	0.7 Vcc	0.1 Vcc	0.9 Vcc	3.3 mA*
AC CMOS	Vcc = 2 to 6 V	8 ns	0.7 Vcc	0.1 V	Vcc – 0.1 V	50 mA
HC/HCT	Vcc = 2 to 6 V	9 ns	0.7 Vcc	0.1 V	Vcc – 0.1 V	25 mA
4000	Vdd = 3 to 15 V	30 ns	0.5 Vdd	0.1 V	Vdd – 0.1 V	0.8 mA*

and “HC/HCT” logic families, the part number starts with “74” and for the “4000” series of CMOS chips, it has a four digit part number, starting with “4”. Table 6-2 lists the different aspects of the different types of logic chips that you will want to work with.

The “output sink” currents are specified for a power voltage of 5 volts. If you increase the power supply voltage of the indicated (with a “*”) CMOS parts, you will also increase their output current source and sink capabilities considerably.

In Table 6-2, I marked TTL input threshold voltage as “not applicable” (N/A) because, as you know, TTL is current driven rather than voltage driven. You should assume that the current drawn from the TTL input is 1 mA for a “0” or “low” input. CMOS logic is voltage driven, so the input voltage threshold specification is an appropriate parameter.

The output current source capability is not specified because many early chips were just able to sink current. This was all that was required for TTL and it allowed external devices, such as LEDs, to be driven from the logic gate’s output without any additional hardware and it simplified the design of the first MOSFET-based logic chips. The asterisk (“*”) indicates that the sink current specification is for 5 volts power; changing the power supply voltage will change the maximum current sink capability as well.

There are three basic output types: totem pole, open collector and the tri-state driver (which is presented later in this chapter). In cases where multiple outputs are combined, different output types should *never* be combined due to possible bus contention.

Virtually all of the electrical dependencies that you should be aware of are listed in Table 6-2, but you may have a number of operating dependencies (such as making sure CMOS inputs are tied high or low) or physical design issues that you should be aware of. “Physical design” is the process of designing a printed circuit board with internal connections built into it that have the chips soldered onto it. The primary chip dependencies that you should be aware of when designing a printed circuit board are the location and type of the chip’s pins as well as any heat removal (i.e. heat sinks) requirements that the chip may have.

Finally, you should know who makes the part and where you can purchase it. This point is often overlooked, but you will find many manufacturers that advertise parts that are seemingly designed just for your application. The first problem that you encounter is that your company has a policy of only buying parts that are available from multiple sources or you may discover that the manufacturer is not considered to be reliable and production quantity parts are difficult to come by. For your first designs, it is a good rule to only use parts that are easily obtainable and, ideally, built by multiple sources.

This may make the design operation a bit more difficult and the final product larger than it could have been, but chances are the product will go through manufacturing very smoothly and with few difficult ‘‘hiccups’’.

Race Conditions and Timing Analysis

As you begin to create digital electronic circuits that are more and more complex and run faster and faster, you are going to discover that they are going to stop working or they are going to start working unpredictably. In trying to find the problem, you will probably look at different parts of the circuit, ranging from the power supply to the wiring and maybe rebuilding it several times to see if it is being affected by other electrical devices running near the application. At some point you will give up and build as well as redesign the circuit, only to discover that the problem is still there.

So what’s the problem? Chances are you have encountered a ‘‘race condition’’, which is normally defined as ‘‘A condition in digital electronics where two or more signals do not always arrive in the same order.’’ Personally, I use a slightly different definition for race condition which states that ‘‘A race condition occurs in any digital electronic circuit where the output to input response time changes according to the inputs passed to it.’’ My definition is a bit more specific and should give you some ideas on where to look for the problem.

Simply put, a race condition is a case where an expected event does not occur.

To illustrate the issue, assume that the application consists of a circuit that is designed to respond to an internal value at a specific time. If the digital electronics used to produce this internal value does not always complete within the specified time, what happens in the circuit that uses this value for input? Chances are the circuit will respond incorrectly, resulting in the problem that you are trying to debug.

An example circuit that has the capability of producing a race condition is shown in Fig. 6-1.

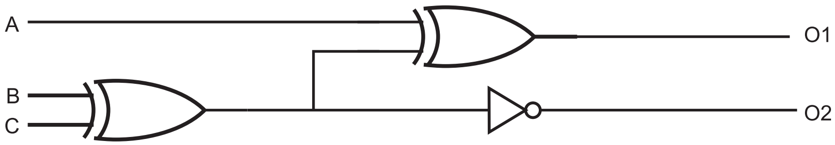

Fig. 6-1. Example race condition circuit.

Figure 6-2 is the waveform output of this circuit to a three bit incrementing signal and, in it, I have indicated the output bits (“O1” and “O2”) and indicated where the operation of the circuit is “correct” (O2 is valid after O1) as well as a possible race condition (when O1 is valid after O2). I have also indicated times, using a shaded block, when both of the XOR inputs are changing and there could be a “glitch” caused by both inputs changing state simultaneously, at which time the output of the XOR gate is unknown.

The glitch produced by the XOR gate is an excellent example of a race condition. As I presented earlier in the book, the XOR gate is typically made up of five NAND gates in the configuration shown in Fig. 6-3. If one input changes, then the output will change state to either “1” or “0” without any glitches, but what happens when the two inputs change state simultaneously?

Quickly thinking about it, you might think that the output doesn’t change state, but consider what happens at the NAND gate level of the XOR gate.

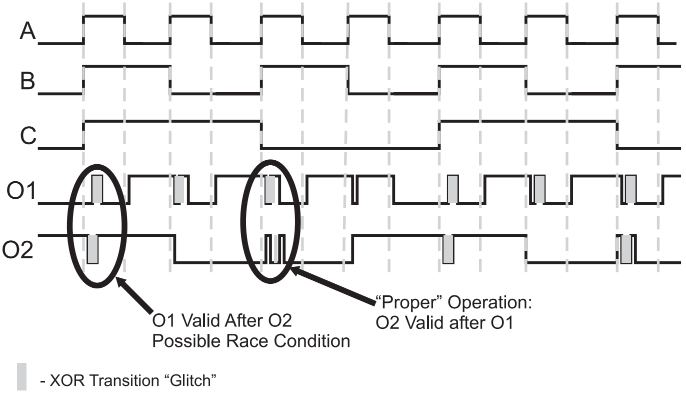
A
B
C
O1
O2
O1 Valid After O2
Possible Race Condition
“Proper” Operation:
O2 Valid after O1
▌ - XOR Transition “Glitch”

Fig. 6-2. Example race circuit wave forms.

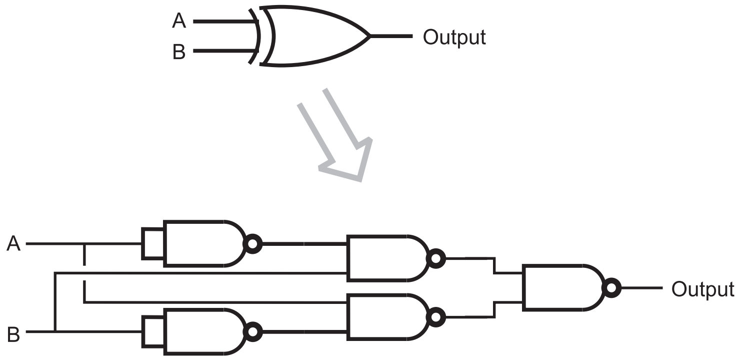
A
B
Output
A
B
Output

Fig. 6-3. XOR gate TTL implementation.

Table 6.3 XOR gate input change study.

Gate delay	Input "A"	Input "B"	G1	G2	G3	G4	Output
–1	1	0	0	1	1	0	1
0	0	1	0	1	1	0	1
1	0	1	1	0	1	1	0
2	0	1	1	0	0	1	1

Table 6-3 lists the NAND gate outputs for the different gates as I've marked them in Fig. 6-3. To help illustrate what's happening, I use ‘‘gate delays’’ as the time increments of this study. In Table 6-3, the initial conditions are one gate delay before the two inputs change value. The inputs change value at gate delay ‘‘0’’.

According to this study, at gate delay 1, the output will be a ‘‘0’’ because the direct inputs from A and B to G3 and G4 have changed at gate delay 0, *but* the inverted inputs from G1 and G2 have not. It won't be until gate delay 2 that the inputs to NAND gates G3 and G4 have stabilized. Thus, the time from gate delay 1 to gate delay 2 will result in generally unknown logic levels, which are normally characterized by the term ‘‘glitch’’.

Going back to Fig. 6-2, you can probably observe what I mean by the race condition, but I'm sure it seems very subtle. Actually, this is the point that I want to make: race conditions are very subtle and are very difficult to observe. For this section, I spent quite a bit of time with a 74C85 (quad XOR gate), a PIC16F627A (Microchip PIC Microcontroller to produce the ‘‘A’’, ‘‘B’’ and ‘‘C’’ inputs to the circuit) and an oscilloscope trying to capture the events shown in Fig. 6-2. I gave up after about 5 hours of trying to capture the event on the oscilloscope in a way that it would be easily seen.

Race conditions are dependent on part mix, applied voltage and ambient conditions. You may find some sample circuits which never have the problem while others will never seem to work right. Finding the actual event is extremely difficult and only after doing a thorough timing analysis of the circuit will you find the opportunity for a race condition to occur. The prevention for this problem is quite simple – figure out what your worst case gate delay is through the circuit and only sample data after this time (even add a 10% margin to make sure there is no chance of marginal components causing problems).

Avoiding the opportunity is why chip designers work at making sure multiple outputs are active at the same time for changing inputs. Looking at the ‘‘A’’/‘‘B’’/‘‘C’’ waveform of Fig. 6-2, you might have thought that it is impossible to achieve the signals at precisely the same time, but it is very likely that if a single gate is producing incrementing outputs, the ‘‘edge’’ of each output bit will be precisely aligned with each other and will cause the glitch on the output of the XOR gate.

The process of determining what is the worst case gate delay is the same process I used for finding the ‘‘glitch’’ in the XOR gate and is known as ‘‘timing analysis’’. It is unusual for somebody to work through this analysis by hand as I have done, except for very simple circuits. When timing analysis is done on a commercial product, it is normally done using a logic simulator, which can find the longest delays and report on any problems.

Quick and Dirty Logic Gates

One of the most frustrating aspects of designing digital electronic circuits is that when you are almost finished, you often discover that you are a gate or two short and you are left with the question of whether or not you should add another chip to the circuit. The major problem with adding another chip to the circuit is the requirement for additional space to place the chip in the circuit. Along with the need for additional space, adding another chip will add to the costs of the application and the difficulty in assembling it. In Chapter 2, I discussed that by using the Boolean arithmetic laws and rules, you could produce various functions using different gates than the ones that are ‘‘best suited’’ for the requirements. In the cases where there are no leftover gates available, a gate can be ‘‘cobbled’’ together with a few resistors, diodes and maybe a transistor. These simple gates are often referred to as being ‘‘MML’’ or ‘‘Mickey Mouse logic’’ technology because they can generally be used in most situations and with different logic families when a quick and dirty solution is required.

To be used successfully, they must be matched to the inputs and outputs of the different logic families that you are using and should not result in long switching times, which will affect the operation of the application, or large current draws, which could damage other components. As a rule of thumb, do not use one of the simple gates presented here between differing

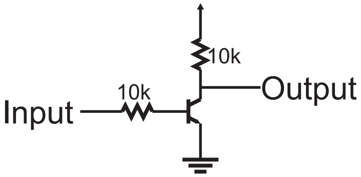

Fig. 6-4. RTL inverter.

technology gates; you will find that operation of different technologies can often be incompatible when you are adding resistors, diodes and transistors like the ones used in the sample gates presented here. Another rule of thumb is to make sure that each MML gate only drives one input – you can get into trouble with input fan-outs and multiple gate current sinking requirements very quickly. Along with trying to satisfy these requirements, there are cases where you will find that the MML gate will require at least as many pins as adding another chip and will be more difficult to wire. Generally speaking, adding MML gates to your application should be considered a ‘‘last’’ resort, not something you design in right from the start.

The most basic MML gate is the ‘‘inverter’’ and should not be a surprise. Figure 6-4 shows the circuit for the MML inverter, built out of two 10 k resistors and an NPN transistor. This inverter is actually a basic ‘‘RTL’’ (resistor–transistor logic) technology device and outputs a high voltage, when it is not being driven by any current. When current is passed to the gate, the transistor turns on and the output is pulled to ground (with good current sinking capability).

This circuit (as well as the other MML gates I discuss in this section) cannot handle high voltage or current inputs and outputs as well as commercially available logic gates and need to be ‘‘buffered’’. The need for buffering the MML’s gate inputs and output is an important point to note when considering using an MML gate in an application. As a rule, MML gates must be placed in the middle of a logic ‘‘string’’ rather than at the input or output ends to ensure that if you are expecting certain characteristics (such as the ability to drive a LED), standard TTL or CMOS technology gates will provide you with it.

The inverter circuit can be simply modified by adding another transistor and resistor, as shown in Fig. 6-5, to create an RTL NOR gate. The RTL NAND gate is shown in Fig. 6-6. The NOR gate is considered the basis of RTL technology.

Implementing an AND or OR gate in MML is a bit more complex and requires a good understanding of the input/output parameters

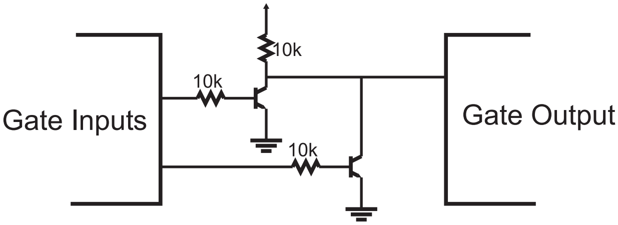

Fig. 6-5. RTL NOR gate.

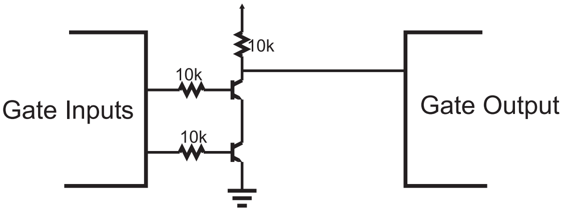

Fig. 6-6. RTL NAND gate.

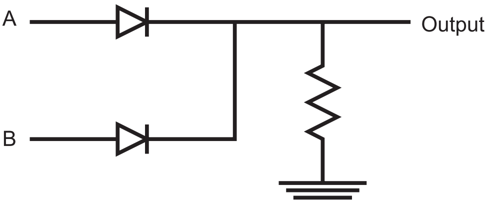

Fig. 6-7. MML OR gate.

of the logic families. In Fig. 6-7, I have shown a sample design for an OR using two diodes and a resistor. The use of a 470 ohm resistor is probably surprising, but it was chosen to allow the gate to be used with both CMOS and TTL logic. In this case, if neither input has a high voltage, then the output will pull the input to ground. If the input is a CMOS gate, then the input will behave as if it were tied to ground. The 470 ohm resistor will allow the TTL input current to pass through ground and it will behave as if the input was at a low logic level. When the resistor is connected to a CMOS input, it will be effectively tying the input to ground, even though no current is flowing through it. In either case, when one of the inputs is driven high, the input pin will be held high and the gate connected to the output of the OR gate will behave as if a high logic level was applied to it.

An MML AND gate (Fig. 6-8) is the simplest in terms of the number of components. The diode and resistor work together to provide a high voltage

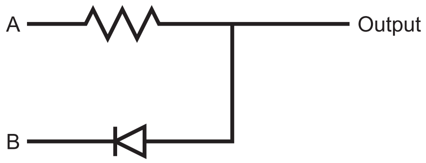

Fig. 6-8. MML AND gate.

when both inputs are high, but when one of them is pulled low, the voltage level will be pulled down and current drawn from the input gate it is connected to.

While the MML AND presented in Fig. 6-8 will work in virtually any application, you may find that you will want to use a 470 ohm resistor in the circuit and a 10 k one in CMOS logic applications. The reason for doing this is to minimize the current drawn by the application; with a 470 ohm resistor, roughly 10 mA will be drawn when the output of the gate is low. This current draw decreases to 100 µA when a 10 k resistor is used instead for the resistors in these two gates.

Dotted AND and Tri-State Logic Drivers

You may feel constrained by the rule that you can only have one driver on a single line (or net). In Chapter 3, I introduced you to the concept of the "dotted AND" bus in which there was a common pull up on the net along with a number of transistor switches, each one of which could "pull" the net to a low voltage/logic level (and draw the current from any TTL gates inputs connected to the dotted AND). The dotted AND works reasonably well and has the advantage that it can control output voltages greater than the power applied to the logic chips. Some more subtle advantages are that more than one output can be active (tying the net to ground) and the operation of the bus will not be affected and TTL open collector and CMOS logic open drain outputs can be placed on the bus along with mechanical switches and other devices which can pull the bus to ground.

The dotted AND bus's main disadvantage is its inability to source significant amounts of current. Smaller value pull up resistors can be used, but this increases the amount of current passed to ground when one of the open collector transistors is on. The dotted AND can be considered to be quite inefficient if it is low for a long period of time, because it is passing current directly to ground. The inability to source large amounts of current

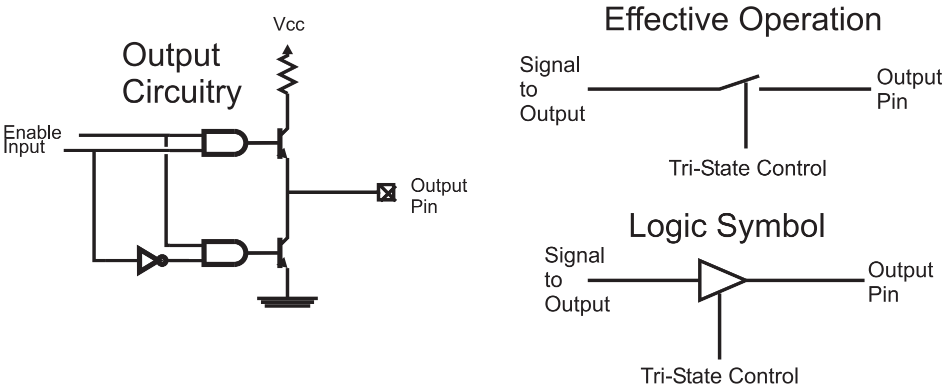

Fig. 6-9. Tri-state driver circuit diagram and operation.

is a drawback when high-speed signals are involved is the major disadvantage of the dotted AND bus. Changing an input from a high to a low, especially when there are some relatively large capacitances on the net, the switching time can become unreasonably slow.

A common error made by new circuit designers when they are adding a dotted AND bus to their designs is forgetting to add the pull up resistor. If the resistor has been forgotten, then the bus will never have a ‘‘high’’ voltage (although it will have a ‘‘low’’ voltage that can be detected). You will find that TTL inputs connected to a dotted AND but not having a pull up will work correctly, but CMOS logic inputs will not.

Another solution to the problem of wanting to have multiple drivers on the same net is to use ‘‘tri-state’’ drivers (Fig. 6-9). These drivers can ‘‘turn off’’ the transistors as effectively as if a switch were opened (the diagram marked ‘‘Effective Operation’’ in Fig. 6-9).

The left-hand side of circuit diagram of Fig. 6-9 shows how the tri-state driver works. If the tri-state control bit is inactive, the outputs of the two AND gates will always be low and the NPN output transistors can never be turned on. This ‘‘inactive’’ state is also known as the ‘‘high impedance state’’. When the tri-state control bit is active, then a high to either the top or bottom NPN transistor will allow the output to behave as an ordinary TTL output.

This ability to ‘‘turn off’’ allows multiple drivers, such as I have shown in Fig. 6-10, to be wired together. In this case, if data was to be placed on the net from Driver ‘‘B’’, the ‘‘Ctrl A’’ line would become inactive (the ‘‘high impedance state’’), followed by the ‘‘Ctrl B’’ line becoming active. At this point, the bus would be driven with the data coming from Driver ‘‘B’’.

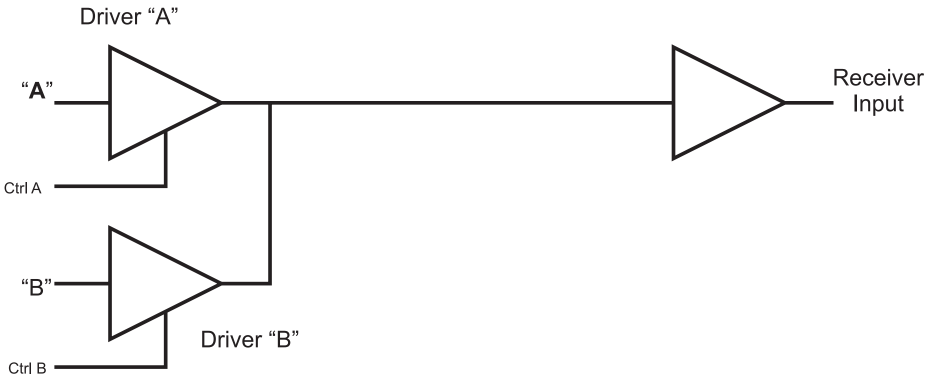

Fig. 6-10. Multiple tri-state drivers on a single net.

It is important that only one tri-state driver is active at one time because the voltage on the common net will be indeterminate, as will be the logic level. You may be thinking that the term ''indeterminate'' to be the case when two drivers are active and are attempting to drive the net at different levels. This is true, but it is often also the case if two different drivers are driving at the same level: CMOS logic and TTL drivers will attempt to drive the net to different voltage levels and even TTL will not give repeating answers when you are trying to understand what is happening. The technical term for the situation where two tri-state drivers are active at the same time is ''bus contention'' and it should be avoided at all costs – only one driver should be active on the net at any one time.

At the start of this section, I noted that there could be more than one output active on a net at the same time. Note that when I say ''multiple active outputs'', I mean more than one driver pulling the net low. I do not recommend this to be part of the design, however; multiple active outputs are impossible to differentiate and you will have problems figuring out which bits are active and what signal is being sent (with multiple outputs active, state changes from one output will most likely be masked by the active operation of others).

Before leaving this section, I do want to point out that tri-state drivers *can* be used on a dotted AND bus. This is probably surprising, considering the dire warnings I have put in regarding bus contention. The trick to adding a tri-state driver to a dotted AND bus is that it is normally disabled and only a low voltage can be put on the net by the tri-state driver. High values are output by simply disabling the tri-state driver and letting the net's pull up provide the high voltage.

Combining Functions on a Net

As a purely intellectual exercise, it can be interesting to see how many functions you can build into a single digital electronics net. From a practical point of view, cramming multiple functions on a single line will minimize the amount of effort that must be expended to build a prototype application. Many products carry out multiple functions on a single line; generally, this is done to allow the manufacture and sale of simpler products. Whatever the motivation, ''stretching'' a logic technology to allow multiple functions on a single net requires a strong knowledge of the technology's electrical parameters and the technology's normal operating conditions. The most important thing to remember is that the input/output devices attached to the net must be properly coordinated to make sure that data is read and written at the right times.

The most obvious ways of connecting two drivers together is to use dotted AND and tri-state drivers on a ''bus'', as I discussed in the previous section. These methods work well and should be considered as the primary method of implementing multiple devices on the net. The other methods discussed here work best for specific situations; but there is no reason why you can't modify your design to take advantage of these specific instances.

When interfacing the bi-directional digital I/O pin to a CMOS driver and a CMOS receiver (such as a memory with separate output and input pins), a resistor can be used to avoid bus contention at any of the pins, as is shown in Fig. 6-11.

Using this wiring, when the bi-directional I/O pin is driving an output, it will be driving the ''Data In'' pin register, regardless of the output of the ''Data Out'' pin. If the bi-directional and ''Data Out'' pins are driving different logic levels, the resistor will limit the current flowing between the bi-directional and the memory ''Data Out'' pin. The value received on the ''Data In'' pin will be the bi-directional device's output.

When the bi-directional digital I/O is receiving data from the memory, the I/O pin will be put in ''input'' (or ''high impedance'') mode and the

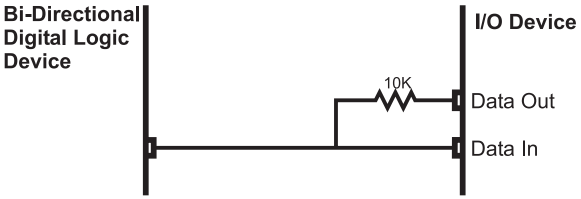

Fig. 6-11. Combining CMOS input and output pins to create a bidirectional bus.

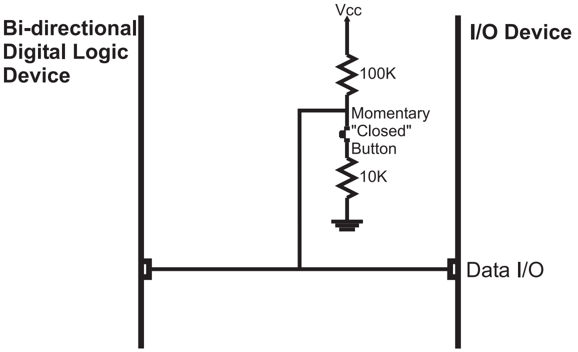
Bi-directional Digital Logic Device
Vcc
100K
Momentary "Closed" Button
10K
I/O Device
Data I/O

Fig. 6-12. Adding a button to a CMOS bidirectional net.

“Data Out” pin will drive its value to not only the bi-directional device’s I/O pin, but the “Data In” pin, as I noted above. In this situation, the “Data In” pin should not be latching any data in; the simplest way to ensure this is to make the digital I/O pin part of the I/O control circuitry. This is an important point because it defines how this circuit works. A common use for this method of connection data in and data out pins is used in memory chips that have separate data input and output pins.

User buttons can be placed on the same net as logic signals as Fig. 6-12 shows.

When the button is open or closed, the bi-directional logic device can drive data to the input device, the 100 k and 10 k resistors will limit the current flow between Vcc and ground. If the bi-directional logic device is going to read the button “high” (switch open) or “low” (switch closed) it will be driven on the bus at low currents when the pin is in “Input Mode”. If the button switch is open, then the 100 k resistor acts like a “pull up” and a “1” is returned. When the button switch is closed, there will be approximately a half volt across the 10 k resistor, which will be read as a “0”.

The button with the two resistors tying the circuit to power and ground is like a low-current driver and the voltage produced is easily “overpowered” by active drivers. Like the first method, the external input device cannot receive data except when the bi-directional device is driving the circuit. A separate clock or enable should be used to ensure that input data is received when the bi-directional device is driving the line.

This method of adding a button to a net can be extrapolated to work with a switch matrix keyboard (presented later in the book), although the circuit and interface operation will become quite complex. Secondly, a resistor/capacitor network for “debouncing” the button cannot be used with this circuit as it will “slow down” the response of the bi-directional device driving the data input pin and will cause problems with the correct value being accepted.

For both of these methods of providing multiple features to a single net, you should only use CMOS logic as it is voltage controlled and not current controlled, like TTL. You *may* be able to use TTL drivers with these circuits, but they may be unreliable. To avoid problems with invalid currents being available to TTL receivers, just use the latter two circuits with CMOS digital logic.

Designing a circuit in which multiple functions are provided on a single net for an application is not always possible or even desirable. Like any design feature implemented in an application, before trying to combine multiple functions on a single net, you should understand the benefits as well as the costs. When it is possible, you can see some pretty spectacular results; my personal record was for a LCD driver in which I was able to combine five functions on a single net – LCD Data Write, LCD Data Read, Data In Strobe, Data Ready Poll and configuration switch poll.

Quiz

1. What parameter is not listed in the chip characteristic card?
   (a) Input fanout
   (b) Number of gates built into the chip
   (c) Electrical dependencies
   (d) Maximum operating speed
2. What is not a typical digital electronic output pin type?
   (a) Totem pole
   (b) Open collector
   (c) High-current
   (d) Tri-state driver
3. Other than the XOR gate, are any other of the six basic I/O gates capable of producing race conditions just by themselves?
   (a) Each one is capable of producing a race condition under certain circumstances
   (b) The NOR Gate in TTL
   (c) The AND Gate in CMOS Logic
   (d) No

4. What is not a factor in determining if a marginal circuit and component will produce a race condition?
   (a) Ambient temperature
   (b) Net length
   (c) Power voltage
   (d) The phases of the moon

5. Mickey Mouse logic should be used:
   (a) Never
   (b) When you are in a hurry to get the application finished
   (c) When you have board space, cost and available gate constraints that preclude adding a standard chip
   (d) When there is a need to pass a CMOS output to a TTL input

6. Each item is an advantage of a dotted AND bus except:
   (a) The dotted AND bus can have tri-state drivers on it as well as mechanical switches
   (b) The dotted AND bus can control voltages greater than the chip's Vdd/Vss
   (c) The dotted AND bus is cheaper than one manufactured with tri-state drivers
   (d) The dotted AND bus can consist of CMOS logic as well as TTL drivers

7. When tri-state drivers are inactive, another term that is used to describe the state is:
   (a) High resistance
   (b) High impedance
   (c) Low current output
   (d) Driver isolation

8. When should multiple tri-state drivers be active?
   (a) When more current is required on the net
   (b) When more speed is required on the net
   (c) When the receiver detects an ambiguous logic level
   (d) Never

9. When adding a push button to a net, can the 100 k resistor connected to positive power and the 10 k resistor connected to positive power be swapped?
   (a) Yes
   (b) No

(c) Only if TTL receivers and drivers are used.
(d) Yes, if you can ensure that the signals passing between the digital devices are still within specified operating margins.

10. When putting a receiver and driver on the same net, can the current limiting resistor be wired between the bi-directional logic device and the "Data In" pin, leaving a direct connection between the bi-directional logic device and "Data Out"?
(a) Yes. There aren't any cases where it wouldn't work
(b) Yes, if the resistor value is within 1 k and 10 k
(c) Yes, for certain technologies of CMOS logic
(d) No. This will cause bus contention

7

CHAPTER

Feedback and Sequential Circuits

This chapter's title probably seems like a bit of a misnomer; you are probably wondering what feedback has to do with digital electronics? When I use the term "feedback", I am using it in the most literal sense, past state data are used to maintain current state data. These circuits built from the theory that I am going to provide in this chapter are commonly known as "memory devices". For digital electronic circuits to store information, that information will continually move through the circuit and is used to determine what the future value of the circuit is. Feedback is critical to provide digital electronics with the ability to "remember" previous states and data.

When you first hear the term "feedback", you probably think of an amplifier with its microphone input brought close to its speaker output (Fig. 7-1). You also probably involuntarily wince at the thought of the term "feedback" because it brings back the memory of the horrible sound the amplifier made when the microphone was too close. This type of feedback

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

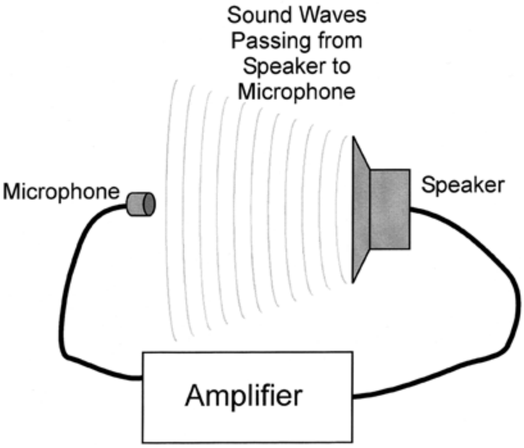
Sound Waves
Passing from
Speaker to
Microphone
Microphone
Speaker
Amplifier

Fig. 7-1. Analog feedback.

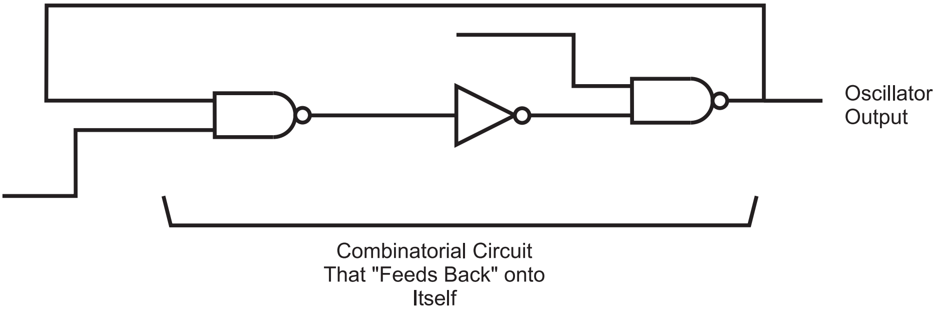
Oscillator
Output
Combinatorial Circuit
That "Feeds Back" onto
Itself

Fig. 7-2. Digital feedback circuit.

cannot save information; the uncontrolled amplification of the signal distorts and destroys information in a very short order.

When I introduced combinatorial circuits at the start of the book, I noted that an important part of combinatorial circuits was that data could only travel in one path; no outputs were passed back to earlier inputs in a logic chain, like the one shown in Fig. 7-2. The reason for specifying that outputs were not to be passed to inputs was to make sure that an inadvertent oscillator, known as a ‘‘ring oscillator’’ (Fig. 7-3), was not created.

You should know that the inverter in Fig. 7-3 is going to invert its input. The problem arises when the input is tied to the output, as it is in this case. When the input is passed to the inverter, it outputs the inverted value, which is then immediately passed back to the input and the gate inverts the value again, and again, and so on.

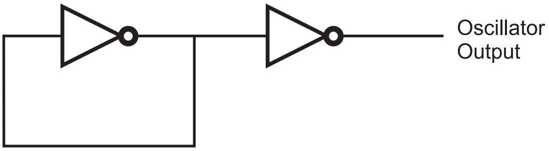
Oscillator
Output

Fig. 7-3. Ring oscillator.

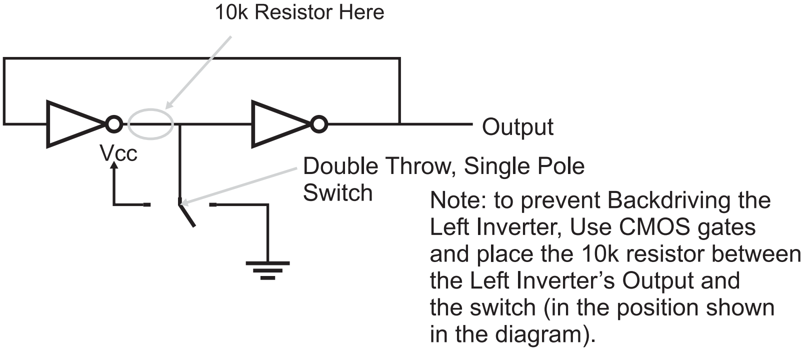
10k Resistor Here
Output
Vcc
Double Throw, Single Pole
Switch
Note: to prevent Backdriving the
Left Inverter, Use CMOS gates
and place the 10k resistor between
the Left Inverter's Output and
the switch (in the position shown
in the diagram).

Fig. 7-4. Single switch debounce circuit.

The ring oscillator is probably the simplest oscillator that you can build and the period of the oscillation runs at literally the speed of the technology's gate delay times the number of gate delays. If the ring oscillator shown in Fig. 7-3 was built from TTL (which has a gate delay of 8 ns) you would see a "square wave" with a period of 62.5 MHz. One of the functions that ring oscillators perform is the measurement of a logic technology's gate delay.

Extrapolating from what has been discussed here, you could build a simple memory circuit using the two inverters and a double throw switch, wired as shown in Fig. 7-4. This circuit is used to "debounce" a switch input. As I will discuss in later chapters, when a mechanical switch is thrown, the physical contacts within the switch literally bounce against each other before a hard, stable contact is made. This bouncing can cause quite a bit of grief when you are trying to respond to a single switch movement.

The circuit in Fig. 7-4 will pass a signal continuously between the two inverters (the output of the two inverters is the same as the input, so there is no chance for a ring oscillator) until the switch comes in contact with a connection that forces the state to change. If the switch was originally at the ground position, the signal coming from the inverter to its left would be a "0". When the switch was moved to the "Vcc" position, the signal going to the inverter to the right would be changed and its output would change.

The beauty of this circuit is that when the switch is in between contacts, the output state of the circuit remains constant.

When the switch is not touching either contact, the two inverters are maintaining the previous bit value and the circuit behaves essentially as a memory device.

There is a downside to the button debounce circuit in Fig. 7-4 and that is when the switch is thrown, it connects the left inverter output to the opposite power supply that it is driving out. This is known as ‘‘backdriving’’ and it should always be avoided.

Backdriving a gate will shorten its life in the best case and could burn it out in very short order. As noted in Fig. 7-4, you should only use CMOS inverters (which are voltage, rather than current controlled) and place the 10 k resistor between the switch and the output of the left inverter. By using this circuit, there will be no chance that the left inverter's output is tied directly to power or ground (which will be the opposite value that it's at) and the 10 k resistor will limit the amount of current that is passed. I did not put the resistor into Fig. 7-4 as it is a basic circuit that I have seen in a number of references and I wanted to point out that it does backdrive a gate output and there are ways of avoiding this problem.

The other term used in this chapter's title, ‘‘sequential circuits’’, is used to identify the class of digital electronic circuits that have memory devices within them and use their data, along with combinatorial circuits, to produce applications. A digital clock (Fig. 7-5) is an excellent example of a sequential circuit. The data output from the memory circuits of the clock are passed to

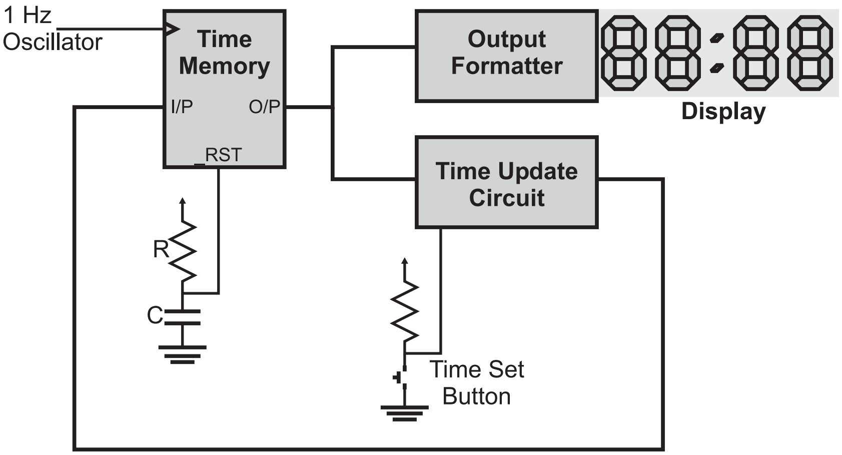

Fig. 7-5. Digital clock circuit.

combinatorial logic circuits and the outputs of the combinatorial circuits are passed back to the inputs of the ‘‘time memory’’ circuits.

Any time memory circuits, like the ones presented in this chapter, are used in a digital electronics application, the circuit is called a ‘‘sequential circuit’’.

Flip Flops (RS and JK)

The best analogy I can find for a simple, one bit ‘‘memory device’’ is the two coiled relay of Fig. 7-6. The relay coil does not have a return spring that only one coil pulls against; when the relay’s wiper is placed in a position it stays there. This memory device is set to one of two states, depending on which relay coil was last energized, pulling the wiper contact into connection with it. Once electricity to the coil is stopped, the memory device will stay in this state until the other coil is energized and the wiper is pulled towards it. This device works very similarly to the most basic electronic memory device that you will work with, the ‘‘reset-set’’ (RS) ‘‘flip flop’’.

The term ‘‘flip flop’’ is indicative of the operation of the memory device: it is either ‘‘flipped’’ to one value or ‘‘flopped’’ to another. Where the relay device relies on friction to keep the saved value constant, the electronic memory unit takes advantage of feedback to store the value. Digital feedback can only be one of two values, so its use in circuits probably seems like it is much more limited than that of analog feedback. This is true, except when it is used as a method to store a result in a circuit like the ‘‘NOR flip flop’’, shown in Fig. 7-7. Normally, the two inputs are at low voltage levels, except to change its state, in which case one of the inputs is raised to a high logic level.

If you are looking at this circuit for the first time, then it probably seems like an improbable device, one that will potentially oscillate because if the output value of one gate is passed to the other and that output is passed to

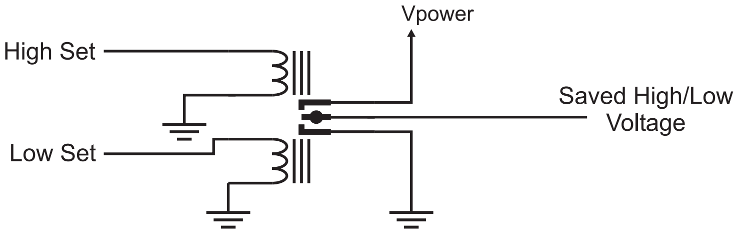

Fig. 7-6. Relay-based ‘‘flip flop’’ circuit.

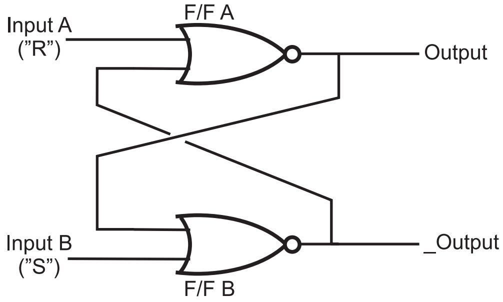
F/F A
Input A
("R")
Output
Input B
("S")
_Output
F/F B

Fig. 7-7. NOR-based RS flip flop.

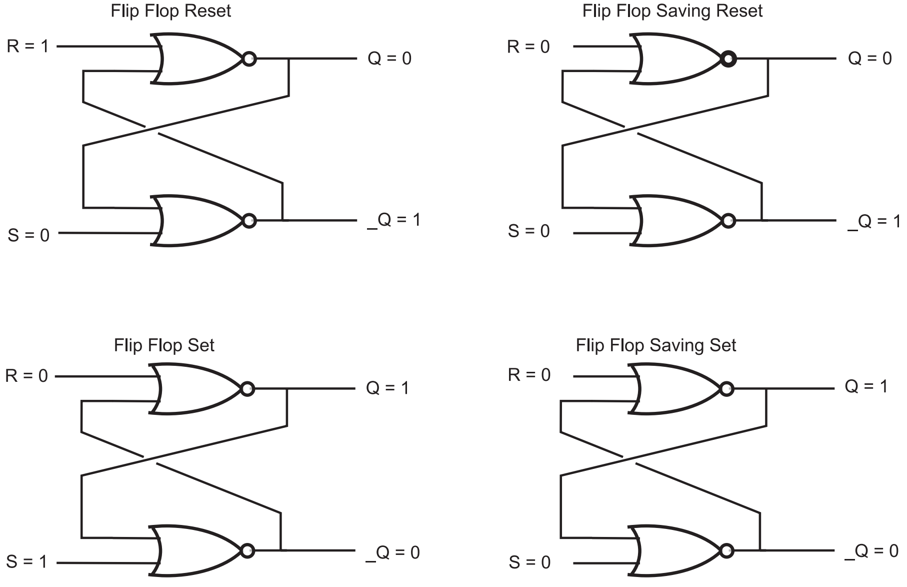
Flip Flop Reset
R = 1
Q = 0
S = 0
_Q = 1
Flip Flop Saving Reset
R = 0
Q = 0
S = 0
_Q = 1
Flip Flop Set
R = 0
Q = 1
S = 1
_Q = 0
Flip Flop Saving Set
R = 0
Q = 1
S = 0
_Q = 0

Fig. 7-8. RS flip flop state changes.

the original, it seems logical that a changing value will loop between the two gates. Fortunately, this is not the case; instead, once a value is placed in this circuit, it will stay there until it is changed or power to the circuit is taken away. Figure 7-8 shows how by raising one pin at a time, the output values of the two NOR gates are changed.

When the "R" and "S" inputs to their respective NOR gates are low, there is only one signal left that will affect the output of the NOR gates and that is output of the other NOR gate. When "Q" is low, then a low voltage will be passed to the other NOR gate. The other NOR gate outputs a high voltage because its other input is low. This high signal is passed to the original

NOR gate and causes it to output a low voltage level, which is passed to the other NOR gate and so on. . . .

The outputs of the flip flop are labeled as “Q” and “_Q”. “Q” is the positive output while “_Q” is the negative value of “Q” – exactly the same as if it were passed through an inverter. The underscore character (“_”) in front of the output label (“Q”) indicates that the signal is inverted (the same as if an exclamation mark (“!”) is used for an inverter’s output). When you look at some chip diagrams, you will see some inputs and outputs that have the underscore before or on the line above the pin label.

The “R” and “S” input pins of the flip flop are known as the “reset” and “set” pins, respectively. When the “R” input is driven high the “Q” output will be low and when “S” is high the “Q” output will be driven high. These values for “Q” will be saved when “R” and “S” are returned to the normal low voltage levels. “Q_0” and “$\_Q_0$” are the conventional shorthand to indicate the previous values for the two bits and indicates that the current values of “Q” and “_Q” are the same as the previous values. Truth tables are often used to describe the operation of flip flops and the truth table for the NOR RS flip flop is given in Table 7-1.

In Table 7-1, I have marked that if both “R” and “S” were high, while the outputs are both low, the inputs were invalid. The reason why they are considered invalid is because of what happens when R and S are driven low. If one line is driven slower than the other, then the flip flop will store its state. If both R and S are driven low at exactly the same time (not a trivial feat), then the flip flop will be in a “metastable” state, Q being neither high nor low, but anything that disturbs this balance will cause the flip flop to change to that state. The metastable state, while seemingly useless and undesirable is actually very effective as a “charge amplifier” – it can be used to detect very

Table 7-1 NOR RS flip flop truth table.

R	**S**	**Q**	**_Q**	**Comments**
0	0	Q_0	$\_Q_0$	Store current value
1	0	0	1	Reset flip flop
0	1	1	0	Set flip flop
1	1	0	0	Invalid input condition

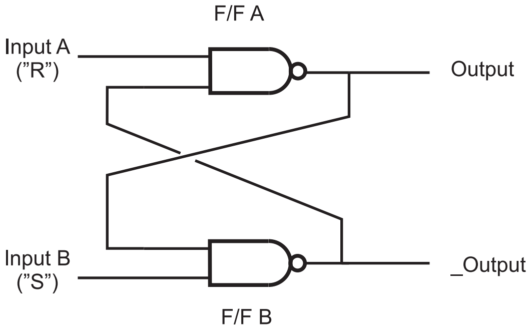

Fig. 7-9. NAND-based RS flip flop.

Table 7-2 NAND RS flip flop truth table.

R	S	Q	_Q	Comments
0	0	1	1	Metastable input state
0	1	0	1	Reset flip flop
1	0	1	0	Set flip flop
1	1	Q_0	$\_Q_0$	Save current value

small charges in capacitors. This is an important mode of operation that is taken advantage of for DRAM and SDRAM memories.

Along with building a flip flop out of NOR gates, you can also build one out of NAND gates (Fig. 7-9). This circuit works similarly to the NOR gate, except that its metastable state occurs when both inputs are low, and the inputs are active at low voltage levels, as I have shown in Table 7-2, which is the NAND RS flip flop's truth table.

You can build your own NOR RS flip flop, which has its state set by two switches as I show in Fig. 7-10 and is wired according to Fig. 7-11. I suggest that you test out the circuit in as many different ways as possible – especially investigating the metastable and post-metastable states. Unless you were to wire the R and S inputs to one switch, you will find it impossible to achieve the metastable state. The parts that are needed to build the RS flip flop are listed in Table 7-3.

Before going on, there is one additional point about flip flops that may not be immediately obvious but will be something that you will have to consider

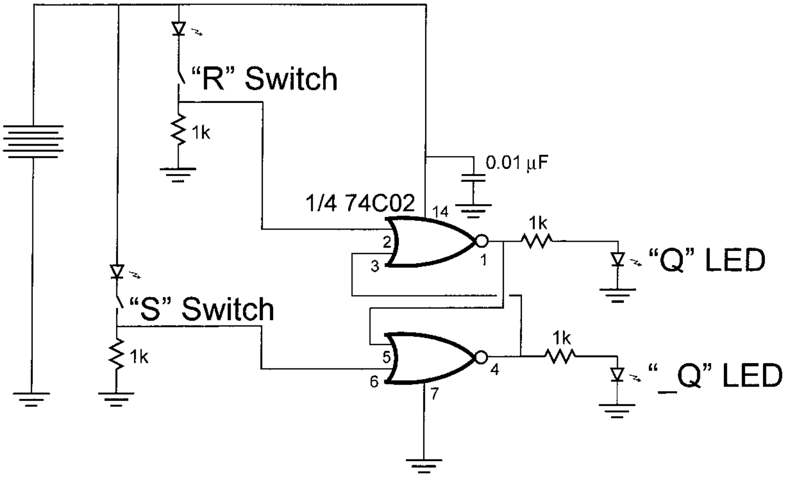
"R" Switch
1k
0.01 µF
1/4 74C02
14
2
3
1
1k
"Q" LED
"S" Switch
1k
5
6
4
7
1k
"_Q" LED

Fig. 7-10. NOR RS flip flop test circuit.

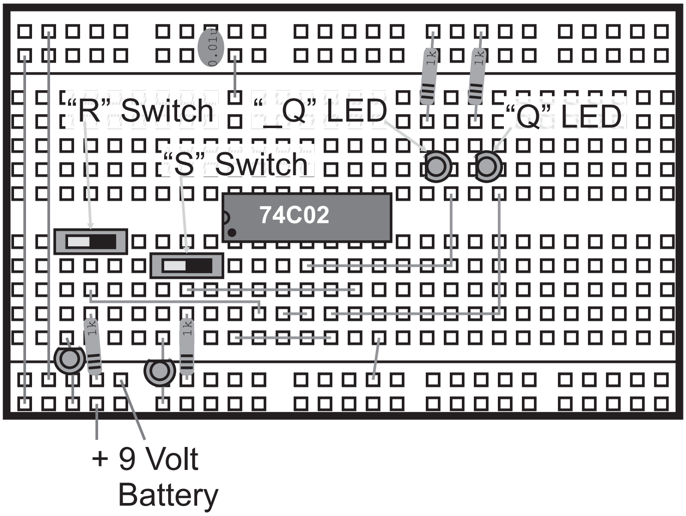
0.01u
1k
1k
"R" Switch
"_Q" LED
"Q" LED
"S" Switch
74C02
1k
1k
+ 9 Volt
Battery

Fig. 7-11. NOR RS flip flop test circuit wiring.

in your career as a designer of digital electronic devices; when power is removed, the flip flops will lose the bit information contained within them. The term used to describe this phenomena is ‘‘volatility’’; flip flops are considered ‘‘volatile’’ devices. Flash memory (like the flash used in your PC) does not lose its information when power is shut off and is known as ‘‘non-volatile’’ memory.

Table 7-3 Parts needed to build NOR gate based RS flip flop.

Part	Description
74C02	CMOS quad, 2 input NOR gate
4 × LEDs	Any color
4 × 1 k resistors	1/4 watt
2 × breadboard mountable switches	Digikey EG1903-ND
0.01 μF capacitor	Any type
Breadboard	
9 volt battery with clip	

Edge Triggered Flip Flops

The RS flip flop is useful for many ad hoc types of sequential circuits in which the flip flop state is changed asynchronously (or whenever the appropriate inputs are active). For most advanced sequential circuits (like a microprocessor), the RS flip flop is a challenge to work with and is very rarely used. Instead, most circuits use an "edge triggered" flip flop which only stores a bit when it is required. You will probably discover the edge triggered flip flop (which may also be known as a "clocked latch") to be very useful in your own applications and easier to design with than a simple RS flip flop.

The most basic type of edge triggered flip flop is the "JK" (Fig. 7-12), which provides a similar function to the RS flip flop except that it changes state when the "clock" input is "rising" (changing from "0" to "1"), as shown in the waveform diagram of Fig. 7-13.

There are a few points about Fig. 7-13 that should be discussed. I have assumed that in the initial state for this example, the output value "Q" is "1". When the first rising edge of the clock ("Clk") is encountered, both J and K are 1, so Q "toggles" or changes state. Next, when the rising edge of the clock is encountered, J is 1 and $K = 0$, so Q becomes 1 and the opposite is true for the next rising edge. In the final rising edge, both J and K are 0 and the value of Q remains the same. There is no metastable state for the JK flip flop. The operation of the JK flip flop is outlined in Table 7-4.

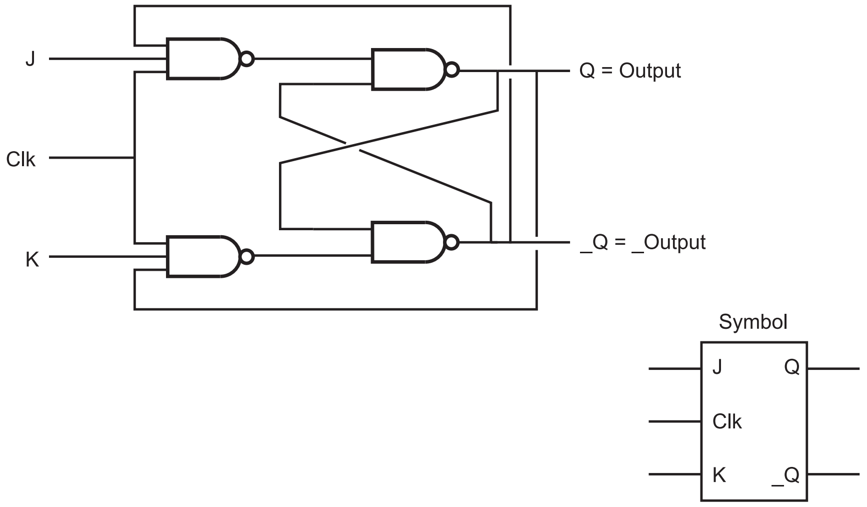

Fig. 7-12. Edge triggered JK flip flop.

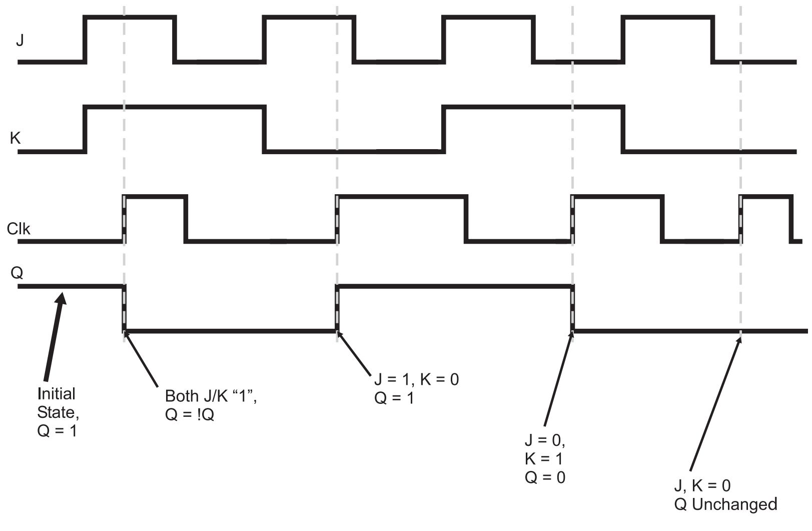

Fig. 7-13. Edge triggered JK flip flop operating waveform.

Just as a small circle on an input or an output of a logic gate indicates that the value is inverted, the clock pin on some chip diagrams is indicated by a small triangle. This convention helps minimize the clutter present in a logic diagram.

The JK flip flop is useful in general digital electronics applications, but it does not provide the necessary function for a computer register. Ideally,

Table 7-4 JK flip flop truth table. "Q" and "_Q" change when a rising clock edge is received.

J	K	Q	_Q	Comments
0	0	Q_0	$\_Q_0$	No change
0	1	0	1	Reset flip flop
1	0	1	0	Set flip flop
1	1	$\_Q_0$	Q_0	JK flip flop toggles

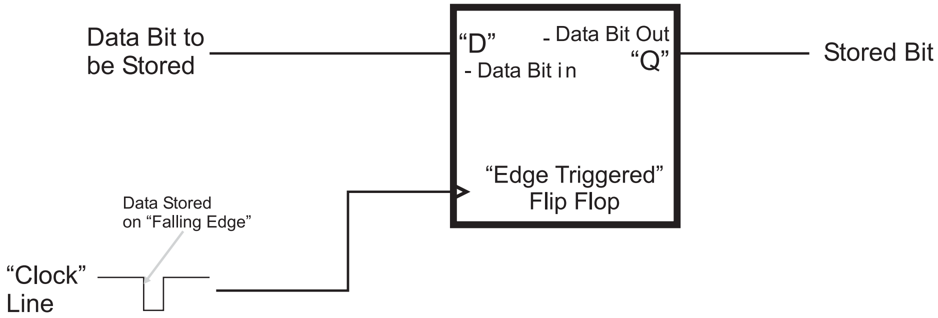

Fig. 7-14. "D" flip flop.

a clocked register's block diagram is quite simple (Fig. 7-14), consisting of a data line passed to the flip flop along with a "clock" line. While the data line stays constant, the contents of the flip flop doesn't change. When the clock line goes from high to low, the data is stored in the flip flop – this is known as a "falling edge clocked flip flop" or a "falling edge clocked register" and it is probably the most common type of flip flop that you will work with.

The edge triggered flip flop (Fig. 7-15) is based on the RS flip flop. Instead of always calling this circuit a "falling edge triggered flip flop" or "clocked register", this circuit is normally known as a "D flip flop". The organization of the flip flops used in this circuit may seem complex, but their operation is actually quite simple: the two "input" flip flops "condition" the clock and data lines and only pass a changing signal when the clock is falling, as I show in Fig. 7-16. To try and make it easier for you to understand, I have marked the outputs of the RS flip flops in Fig. 7-15 and showed the waveforms at these points.

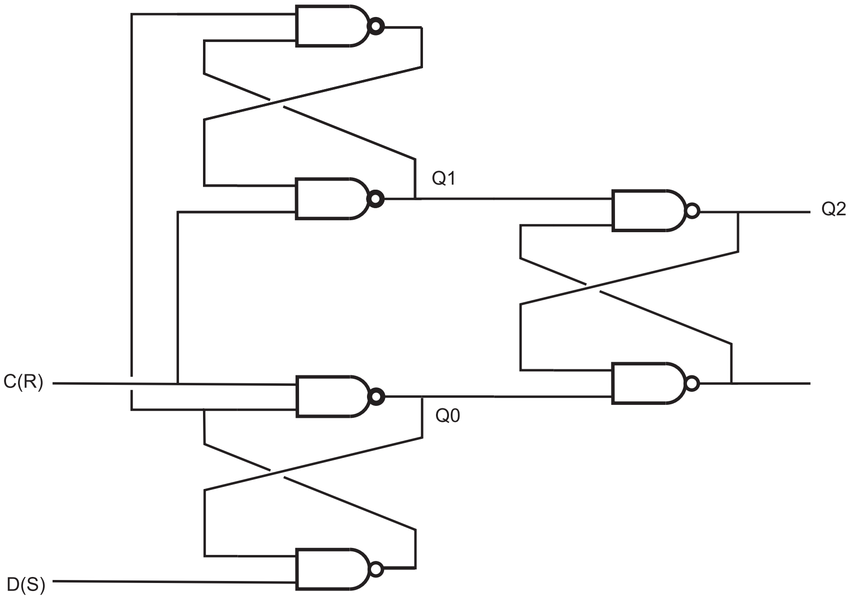

Fig. 7-15. D flip flop logic diagram.

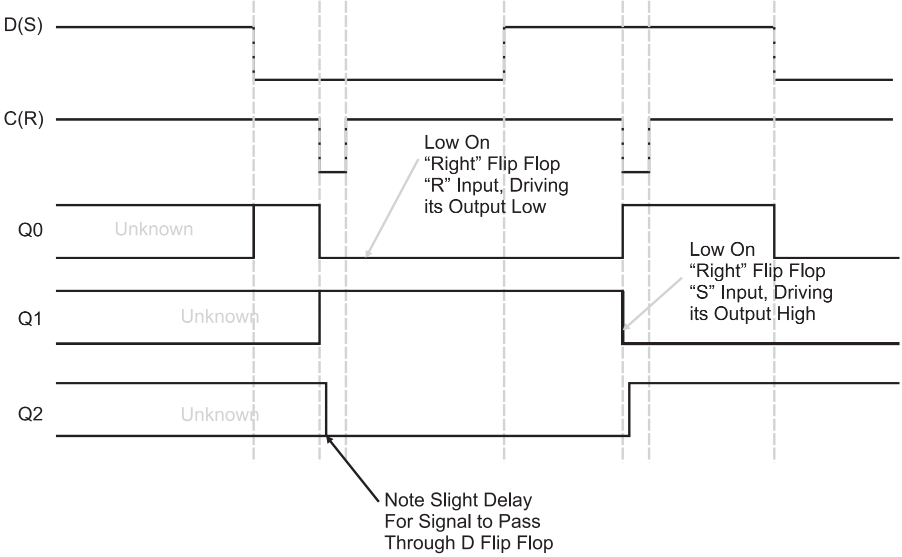

Fig. 7-16. D flip flop operating waveform.

Note that in Fig. 7-16, I have marked the flip flop states *before* the first clock pulse as being ‘‘unknown’’ (in Fig. 7-14, the initial state was assumed). This is actually a very important point and one that you will have to keep in mind when you are designing your own circuits. You cannot expect a flip flop to be at a specific state unless it is set there by some kind of ‘‘reset’’ circuit (which is discussed in the next section). The output of the edge triggered flip flop stays ‘‘unknown’’ until some value is written in it. If you look at the signals being passed to the right flip flop (Output Q0), you will see that the inputs are unknown until the ‘‘data’’ line becomes low, at which point the two inputs to the right flip flop become high and the ‘‘unknown’’ bit value is stored properly in the flip flop.

The first value written into the D flip flop is ‘‘zero’’, the ‘‘data’’ line’s value for the write is changed before the ‘‘clock’’ line goes negative. When the ‘‘clock’’ line goes low, it forces out a ‘‘1’’ to be passed to the ‘‘right’’ flip flop, keeping it in its current state. The operation of the edge triggered flip flop should become very obvious if you were to build it (it would require two 74C00s).

I find the D flip flop to be the flip flop that I build into my circuits most often. It is simple to work with and can interface to microcontrollers and microprocessors very easily. It is, however, quite awkward to wire, especially when you want to work with the ‘‘full circuit’’, which is shown in Fig. 7-17. This circuit not only has data stored on the rising edge of the clock line but also two other lines ‘‘_Clr’’ and ‘‘_Pre’’ will force the flip flop’s output to a ‘‘0’’ (low voltage) or a ‘‘1’’ (high voltage), respectively, when they are pulled low. This allows for a number of different options for using the D flip flop in your circuit that can allow you to pull off some amazing feats of digital logic.

If you want to experiment with this circuit using two input NANDs (74C00s), I must warn you that it will be quite difficult and complex for you to wire. If you were to use three gates to produce one three input NAND gate, 18 NAND gates would be required to implement the full D flip flop function, which would require four and a half 7400 chips. To demonstrate the operation of the circuit, you could build it out of two 7410 (three, three input NAND gates) or be lazy like I am and just use one 74LS74 (Fig. 7-18) to experiment with the different functions of the full D flip flop.

The 7474 chip consists of two D flip flops with both the ‘‘Q’’ and ‘‘_Q’’ outputs passed to the chip pins. All four inputs shown in Fig. 7-18 (Data and Clock as well as two pins that provide you with the ability to set or reset the state of the flip flop without the use of the data or clock pins) are provided for each of the two flip flops built into the chip. The 7474 is a very versatile chip and can be used for a wide range of applications.

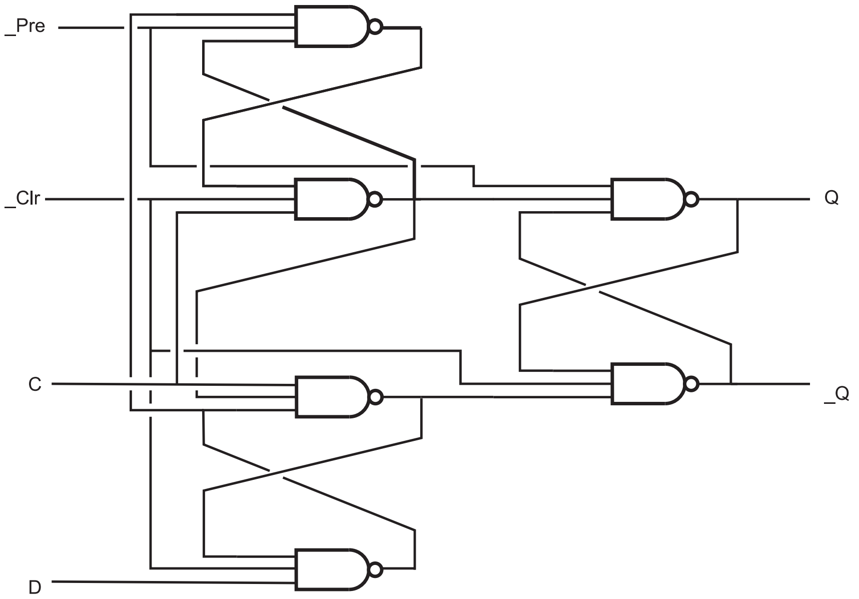
_Pre
_Clr
C
D
Q
_Q

Fig. 7-17. Logic diagram of full D flip flop with negative active preset and negative active clear pins.

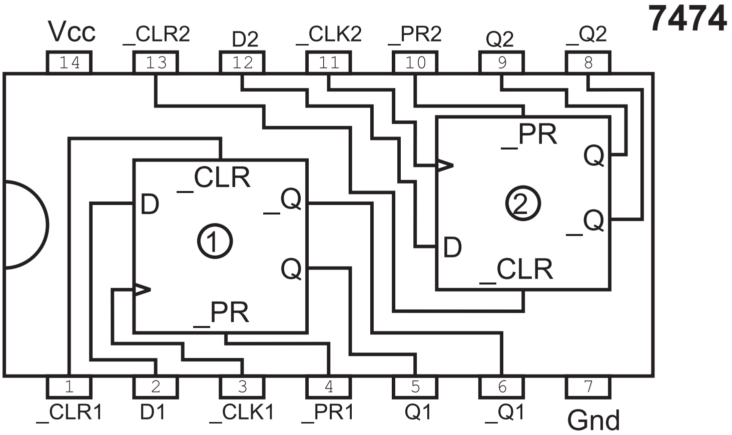
7474
Vcc
_CLR2
D2
_CLK2
_PR2
Q2
_Q2
14
13
12
11
10
9
8
_PR
Q
②
_Q
D
_CLR
_CLR
D
_Q
①
Q
_PR
1
2
3
4
5
6
7
_CLR1
D1
_CLK1
_PR1
Q1
_Q1
Gnd

Fig. 7-18. 7474 dual D flip flop chip pinout.

Latches Versus Registers

Two terms that are often used interchangeably are ‘‘register’’ and ‘‘latch’’. In the previous section, I introduced you to the ‘‘register’’, which is another term for an edge triggered flip flop. When you look at parts lists and datasheets, you will see parts that are identified as "registers" and others as "latches" and these parts will have the same pinouts with no obvious differentiation in operation between the devices. Furthermore, I have found many chip manufacturers that have labeled their parts as "latches" when in fact they were "registers" and vice versa.

Quite simply put, "registers" are flip flops that store data when the rising (low to high or 0 to 1) or falling (high to low or 1 to 0) edge (whichever is used by the device) is received on the "clock" (or, my abbreviation, "Clk") pin. Registers are aptly named because they are normally used as simple data storage devices for microprocessor memory. Latches are often used in microprocessor applications to save an address on a multi-purpose bus.

The best analogy for the "latch" that I can think of is a latch on a barn door: when the latch is not engaged, animals and whatever can wander in. Once the latch is closed, what is in the barn stays in. The "latch" flip flop works similarly to this; with one state of the clock line, the input data is passed to the output directly and can be changed at any time (i.e. there isn't any storage) but once the clock line changes state, the last value of the data is stored in the latch until the clock changes value.

In the previous section, I introduced you to the edge triggered D flip flop "registers". The D flip flop "latch" is actually quite a bit simpler (Fig. 7-19), but what is interesting about it is that it doesn't work anything like its edge triggered cousin. In Fig. 7-20, I have drawn a data input along with a clock and the "Q" (output pin) values for an edge-triggered D flip flop register and a D flip flop latch.

You will probably be surprised to see that waveforms for the two memory devices are completely different. The edge triggered D flip flop register stores data in a very consistent and logical way – every time the clock pin

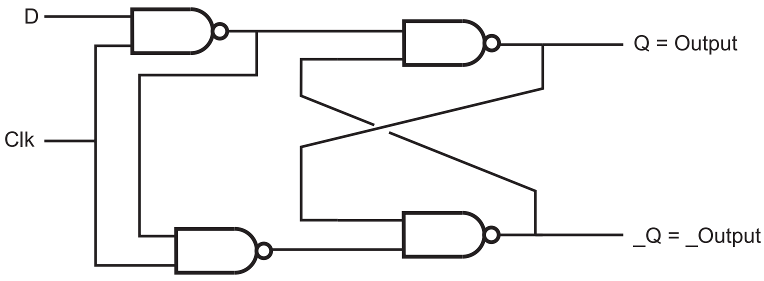

Fig. 7-19. D flip flop latch.

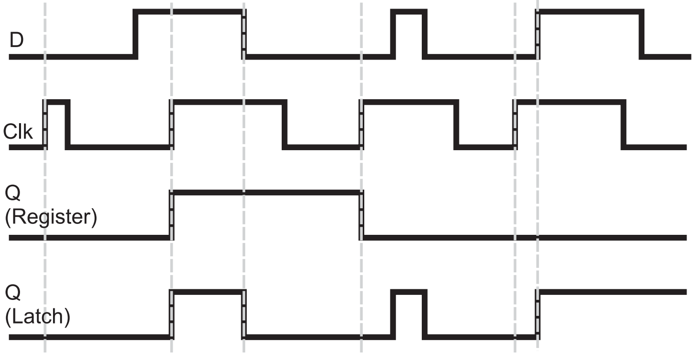

Fig. 7-20. D flip flop latch operating wave form.

rises, the value of ‘‘D’’ is stored in the flip flop and nothing changes until the next rising edge of the clock pin.

The latch, on the other hand, seems to operate more like an AND gate than a memory storage device. The storage function tends to be obfuscated in the example of Fig. 7-20 because in many cases, I show the D pin changing state before the clock line returns low. This is an important point because many people consider the two devices to be interchangeable and this is simply not the case. Latches and registers have different applications and it is critical for you to understand what they are. You cannot put a latch chip in place of a register simply because they are pin compatible; you must make sure that the incoming data does not change state until the clock goes low.

Interestingly enough, latches do not need as much time to save data as a register; there are 9 fewer gate levels for a signal to pass through and even though I show the data save operations being instantaneous in Fig. 7-20, they are not. The latch can take as little as one-third of the time to save data as a register and only requires two gate delays before passing the data along (after which the data can be stored). This makes the latch an important chip for working with microprocessors with a ‘‘multiplexed’’ address bus.

Reset

If you cycle the power to any flip flop, you will have noticed that the initial ‘‘state’’ (or value) can be either ‘‘0’’ (LED off) or ‘‘1’’ (LED on), with no way of predicting which value it will be. This is normal because when power is applied to the flip flop it will start executing in the metastable state, and for any kind of imbalance in the circuit (e.g. residual charge or induced voltage) on the inputs of either NAND/NOR gate, the flip flop will respond and this will be its initial state. Often, this random initial state is not desired – instead, the circuitry should power up into a specific known state for it to work properly. This is why throughout this chapter I have taken pains to note that the initial state of a flip flop is not known. You may find that an application with one flip flop usually powers up the same way; if you were to do a statistical analysis of the power up values, you might even find that a single power up state approaches 100%, but you cannot guarantee this for all occurrences of the chip, or even that all other similar chips in the same application circuit will power up the same way.

Specifying the state when the circuit is powered up is known as "initialization" (just as it is for programming) and is required for more than just sequential logic circuits. Initialization normally takes place when the application is "reset", or waiting to start executing. To avoid confusion later, I should point out there are two types of "reset" described in this book when I talk about digital circuits. Earlier, when I was talking about simple combinatorial circuits, I also called a "low" or "0" voltage level "reset" (and "high" or "1" as "set"). Now, when the term "reset" is used, I am describing the state when the circuit is first powered up or stopped to restart it from the beginning. When you read the term "reset" later in the book (as well as in other books), remember that if a single bit or pin is being described, the term "reset" means that it is "0" or at a low level. If a sequential circuit (like a microcontroller) is "held reset" or "powering up from reset", I mean that it is being allowed to execute from a known state.

The "_CLR" pin on the full D flip flop (like the 7474) is known as a "negative active control" and is active when the input is at a "0" logic level. To make this pin active during power up, yet allow the chip to function normally, a resistor/capacitor network on the TTL input pin delays the rise of the pin (as shown in Fig. 7-21) so that the pin is active low while power is good. When the signal on "_CLR" goes high, and the clear function is no longer active, the chip can operate normally, with it being in an initial known state.

The time for the RC (resistor/capacitor) network to reach the threshold voltage can be approximated using the equation:

$$\texttt{Delay time} = 2.2 \times \texttt{R} \times \texttt{C}$$

When you work with microprocessors and microcontrollers, you will want to implement a more sophisticated reset circuit. Many microprocessor manufacturers recommend an analog comparator based reset circuit like the one shown in Fig. 7-22. This circuit controls an open collector (or open drain)

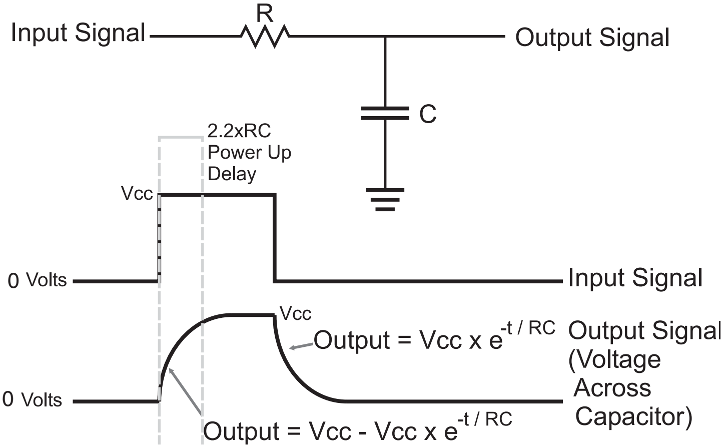
Input Signal
R
Output Signal
C
2.2xRC
Power Up
Delay
Vcc
0 Volts
Input Signal
Vcc
Output = Vcc x e-t / RC
Output Signal
(Voltage
Across
Capacitor)
0 Volts
Output = Vcc - Vcc x e-t / RC

Fig. 7-21. RC network operation.

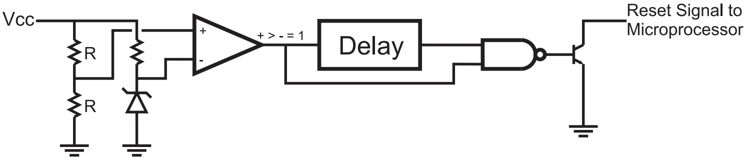
Vcc
R
R
+
-
+ > - = 1
Delay
Reset Signal to
Microprocessor

Fig. 7-22. Block diagram of commercial reset control chip.

transistor output pin that will pull down a negative active reset pin when power dips below some threshold value. This circuit is often available as a "processor reset control" chip and is put into the same black plastic package as a small transistor (known as a TO-92).

Processor reset control chips are available for a very wide variety of different "cut off" voltages, ranging from 2.2 volts and upwards. Figure 7-23 shows the operation of the internal parts of the processor reset control chip when the input voltage drops below the set value; the comparator stops outputting a "1" and a delay line is activated. This delay line is used to filter out any subsequent "glitches" in the power line and makes sure that the power line is stable before allowing the processor to return from reset and continue executing. When the comparator outputs a low value or the delay line is continuing to output a low value, the output of the NAND gate they are connected to is high and it turns on the open collector output transistor, pulling the circuit to ground.

The Panasonic MN1381 line of chips is a very popular processor reset control and can be used to control a sequential circuits reset using

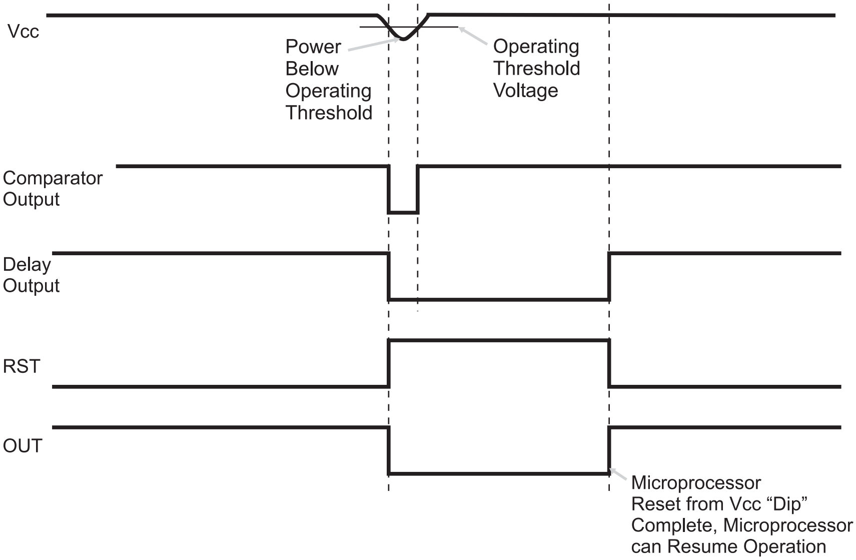
Vcc
Power Below Operating Threshold
Operating Threshold Voltage
Comparator Output
Delay Output
RST
OUT
Microprocessor Reset from Vcc "Dip" Complete, Microprocessor can Resume Operation

Fig. 7-23. Reset circuit operation.

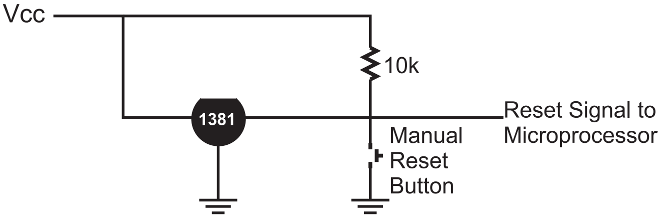
Vcc
10k
1381
Reset Signal to Microprocessor
Manual Reset Button

Fig. 7-24. Practical microprocessor reset circuit.

a circuit similar to the one shown in Fig. 7-24. This circuit will take advantage of the RC network delaying the rise of the control signal, provide you with the ability to reset or stop the operation of the microprocessor and halt the operation of the robot if the battery falls below a safe minimum.

If you power on and off a circuit quickly, you may find that it does not power up properly. This is due to the capacitor in the reset circuit not discharging fully – it may take as much as 10 seconds for it to discharge completely. This was actually an issue with the original IBM PC; if you had a situation where the PC ‘‘hung’’, you would have to power down and wait at least 15 seconds to make sure that the reset circuit would allow the computer to power up properly.

Quiz

1. Feedback in digital electronics:
   (a) Is built into every gate
   (b) Must always be avoided
   (c) Can be used to store bit data
   (d) Is only used in radio interface circuitry

2. Ring oscillators can be used:
   (a) In digital watches
   (b) To measure the gate delay of a logic technology
   (c) To test the operation of a combinatorial circuit
   (d) Only when current limiting resistors are in place to protect gate outputs

3. What do the letters "R" and "S" stand for in the RS flip flop?
   (a) "Recessive" and "Static"
   (b) "Reset" and "Set"
   (c) "Rothchild" and "Stanislav"
   (d) "Receive" and "Send"

4. What is the "metastable state" of a flip flop?
   (a) When it has started to oscillate
   (b) The time between when the inputs change the output is correct
   (c) The state in which the outputs of a flip flop are half way between "0" and "1" and can be easily "pushed" into a specific state
   (d) The state in which "Q_0" is unknown

5. "Toggling" a bit means:
   (a) Setting (making the output a 1) of a bit
   (b) Leaving the bit in its current state
   (c) Inverting the bit's state
   (d) Resetting (making the output a 0) of a bit

6. A "Register" can be used in:
   (a) Nowhere, it is a thought experiment used to show feedback in a digital application
   (b) Just computer processors
   (c) Just sequential digital electronics application
   (d) Just about any digital electronics application

7. The "_Pre" pin of a D flip flop will:
   (a) Set the bit
   (b) Reset the bit
   (c) Nothing
   (d) Toggle the state of the bit

8. Which formula specifies the RC network response to a sudden voltage input?
   (a) $V = 2.2 \times R \times C$
   (b) $V = Vcc - Vcc \times e^{-t/RC}$
   (c) $V = Vcc \times e^{-t/RC}$
   (d) $V = i \times R$

9. Why are latches like barn doors?
   (a) They provide a secure environment for what's inside them
   (b) They are both relatively heavy
   (c) They allow free passage until the latch is engaged
   (d) They are the fastest method for passing things in and out

10. Which application is a latch best suited for?
   (a) Main memory in a computer system
   (b) Bicycle lock combinations
   (c) Stopping and saving data mid-stream
   (d) Temporary storage of data in a microprocessor

Test for Part One

Do not refer to the text when taking this test. You may draw diagrams or use a calculator if necessary. A good score is at least 38 correct answers. Answers are in the back of the book. It's best to have a friend check your score the first time so you won't memorize the answers if you want to take the test again.

1. The assertion "John is going to go out with the boys tonight or date Mary" is an example of:
    (a) Negative logic
    (b) The AND operation
    (c) The inclusive OR operation
    (d) The exclusive OR operation

2. Which symbol does not represent AND, OR or NOT?
    (a) "*"
    (b) %
    (c) "+"
    (d) "!"

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

3. How would you create a three input NOR gate from two input NOR gates?
   (a) Invert each input by passing them to a two input NOR gate and then combining it like the three input AND gate
   (b) Pass two inputs to a NOR gate and pass this input to a second NOR gate along with the remaining input and invert the final result
   (c) Pass two inputs to a NOR gate, use a second NOR gate to invert this NOR gate's output and pass this result, along with the third input to a third NOR gate
   (d) Pass two inputs to a NOR gate and pass this input to a second NOR gate along with the remaining input
4. "Product of sums" combinatorial logic circuits are not as common as "sum of products" because:
   (a) They rely on "negative logic", which makes their operation more difficult to understand by simply looking at the circuit
   (b) They are not as fast as product of sums combinatorial circuits
   (c) Automated design tools are typically not programmed to work with product of sums circuits
   (d) Product of sums combinatorial logic circuits cannot produce the same functions as sum of product combinatorial logic circuits
5. Which one of the following statements is false?
   (a) Combinatorial logic circuits are drawn with inputs entering the gates from the left and exiting from the right
   (b) Outputs from some of the gates in the combinatorial are passed back so that they are part of their own inputs
   (c) Combinatorial logic circuits can be designed to have true or false outputs for given inputs
   (d) The function of the individual gates in a combinatorial logic circuit does not change, even if the gates are used to provide a function which is radically different
6. Idealized waveform diagrams do not show:
   (a) Potential "glitches" caused by gates changing state
   (b) Delays in gates, responding to changes in inputs
   (c) What happens with wiring problems such as when multiple outputs are connected to the same input
   (d) All of the above

7. If an application has a critical speed requirement, you should design your circuit:
   (a) To be as simple as possible, as this will minimize the delay a signal has passing through the circuit
   (b) With as few gate delays as possible, while keeping an eye on the number of gates required as well as whether or not it can be efficiently implemented in the technology that you are using
   (c) Using the fastest technology available
   (d) Using computerized design systems

8. The NOR equivalent to an AND gate is:
   (a) Built from two NOR gates and requires two gate delays for a signal to pass through
   (b) Built from three NOR gates and requires three gate delays for a signal to pass through
   (c) Built from three NOR gates and requires two gate delays for a signal to pass through
   (d) Built from one NOR gate as well as a NOT gate and requires two gate delays for a signal to pass through

9. When circling “1” outputs in a Karnaugh map:
   (a) A maximum of two bits can only be circled at one time
   (b) No bits can be circled more than once
   (c) Each circle should be around a power of two number of bits
   (d) Single bits on one side of a the map cannot be circled with bits on the other side

10. The four bit Karnaugh map

	AB			
	00	01	11	10
CD – 00	0	0	1	0
01	1	1	1	1
11	0	1	1	0
10	1	0	1	1

has the optimized sum of product equation:
   (a) $\text{Output} = (!B \cdot !D) + (!A \cdot !B) + (C \cdot !D) + (B \cdot !C \cdot D)$
   (b) $\text{Output} = (A \cdot B) + (!C \cdot D) + (B \cdot D) + (!B \cdot C \cdot !D)$

(c) Output = $(!C \cdot !D) + (A \cdot B) + (B \cdot D) + (!B \cdot C \cdot !D)$
(d) Output = $(A \cdot B) + (!C \cdot D) + (B \cdot !D) + (!B \cdot C \cdot !D)$

11. Benjamin Franklin postulated:
(a) Electricity flows from positive to negative
(b) Lightning is dangerous
(c) Keys had to be charged before they would open doors
(d) Thomas Edison unfairly copied his work

12. The term "net" is used for:
(a) Wires in a circuit and lines on a circuit diagram
(b) Indicating the active signal lines of an ethernet cable
(c) A search tool through which information is passed through and relevant "hits" sticks to
(d) Nylon webbing used to protect a circuit against falling metal components.

13. Knowing Ohm's law and the resistance of a load and the voltage of a battery powering it, you can determine:
(a) The current passing through it
(b) The amount of water coming through an analogous pipe/tap/hose
(c) Its equivalent parallel resistance
(d) The Thevenin equivalent circuit

14. Using the SI numbering methodology and symbols, 10,000,000 volts would be written out as:
(a) 10 million V
(b) 10 MV
(c) 10 μV
(d) 10,000,000 V

15. The voltage drop across a resistor in a series circuit:
(a) Cannot be calculated
(b) Is proportional to the power dissipated in the circuit
(c) Is always zero
(d) Is proportional to the resistor's value relative to the total resistance in the circuit multiplied by the applied voltage

16. If a 10 ohm and 20 ohm resistor are in series, the equivalent resistance:
(a) Cannot be calculated without knowing the voltage applied
(b) 30 ohms
(c) 7.5 ohms
(d) 6.7 ohms

17. A 0.01 μF capacitor is most often used in:
    (a) Radio applications; it has no use in digital electronics
    (b) Decoupling digital electronic chips
    (c) Filtering power supply "noise"
    (d) Ballast in fluorescent lighting

18. The value at a given time for the capacitor voltage in a resistor–capacitor low-pass filter circuit responding to a rising step input is:
    (a) Infinite
    (b) Defined by the formula $V(t) = V - V \times e^{-t/\tau}$
    (c) Defined by the formula $V(t) = V \times e^{-t/\tau}$
    (d) Zero; the capacitor has no voltage drop across it

19. The NOR gate was chosen as the basic CMOS logic gate because:
    (a) It can be built most efficiently using MOSFET transistors
    (b) It provides the fastest logic functions in CMOS logic
    (c) It helps the circuit designer differentiate the functions provided by TTL and CMOS logic circuitry
    (d) The NOR gate minimizes the power lost in the chip

20. Which statement is not a reason cited for using resistor pull ups and resistor/NOT gates for pull downs?
    (a) The resistor can be connected to negative voltage without damaging the circuit
    (b) Test equipment can easily change the state of logic pin inputs
    (c) TTL and CMOS logic operate optimally with these circuits
    (d) The resistor pull ups and resistor/NOT gate pull downs will work for both TTL and CMOS logic

21. If a silicon diode was passing 2 A of current, it would be dissipating:
    (a) 14 watts of power
    (b) 0.2 watts of power
    (c) 0 watts of power
    (d) 1.4 watts of power

22. In a 5 volt powered circuit, you have two LEDs in series and want to pass approximately 5 mA through them. What is the best current limiting resistor value should you use?
    (a) 47 ohms
    (b) 100 k ohms
    (c) 5 ohms
    (d) 1 k ohms

23. A bipolar transistor is best suited for:
    (a) Radios and high-fidelity sound systems
    (b) Small, high-density chips
    (c) Memory circuits
    (d) Low-power, high-density chips

24. The basic CMOS logic gate is:
    (a) The NOT gate
    (b) The AND gate
    (c) The NOR gate
    (d) The NAND gate

25. TTL is:
    (a) Sound controlled
    (b) Resistor controlled
    (c) Current controlled
    (d) Voltage controlled

26. When a TTL input is low:
    (a) Current is being drawn from it
    (b) A low voltage is being applied to it
    (c) A "0" is being passed to it
    (d) Electrons are being drawn from the emitter of the input gate's NPN transistor

27. TTL/CMOS logic outputs:
    (a) Can be used to drive neon lamps
    (b) Can source/sink roughly 20 mA
    (c) Cannot be used with different technology inputs
    (d) Are limited to driving inputs less than 20 m away

28. "Fanout" is the term applied to:
    (a) The number of outputs that can be driven by one input
    (b) The number of fans required to cool a set number of chips
    (c) The speed a signal travels through multiple paths of a logic chain
    (d) The number of inputs that can be driven by one output

29. CMOS logic has the following characteristics:
    (a) They are low speed, low power
    (b) Require just about no power, regardless of the speed they operate at

(c) The current required is a function of the speed of operation
(d) Require less power than TTL because MOSFETs cannot be packed as tightly as bipolar transistors

30. LEDs are used in beginner digital electronic circuits:
(a) To indicate analog voltage levels
(b) To indicate a part is overheating
(c) To indicate input and output binary values
(d) To communicate with other circuits

31. The difference between 74Cxx and 74xx chips is:
(a) The 74Cxx is built from CMOS logic while the 74xx is TTL
(b) Signals in the 74Cxx propagate at the speed of light (as indicated by the ‘‘C’’ in the part number)
(c) The 74xx can work from 5 to 9 volts while the 74Cxx can only work with 5 volts
(d) The 74Cxx is built with a ‘‘compacted’’ chip

32. Gray codes were invented:
(a) To make your life miserable
(b) For simplifying Boolean logic statements
(c) For simplifying the task of determining the position of a device
(d) As a method of counting that was faster than binary

33. Adding 6 to 5 and getting the result 11 is the same as:
(a) Adding 7 to 4 and getting the result 11
(b) Adding 3 to 4 and getting the result 7 because in both cases, a prime number is produced
(c) Adding 5 to 6 using the commutative law and getting the result 11
(d) Adding 6 to 5, writing down ‘‘1’’ and then ‘‘10 × 1’’ because a carry digit is produced

34. The term ‘‘ripple’’ as applied to addition and subtraction is:
(a) The carry and borrow bits
(b) The result of the two single digit operation passed to the next significant digit
(c) The affect the operation has on its surrounding digits
(d) The oscillations caused by the need to carry and borrow data

35. Using the negated addition for subtraction, the borrow (negated carry) bit for the operation 5 – 6 is:
(a) Not required
(b) 1

(c) 0
(d) Indeterminate

36. A small circle on a gate's input indicates:
    (a) That the signal can only be used for output.
    (b) That the signal is inverted before being passed to the gate
    (c) Only open collector drivers can be used with this input
    (d) The I/O can be used for monitoring the passage of the signal output in the gate

37. Magnitude comparators are based on:
    (a) Three initial input state values
    (b) Two four bit inputs
    (c) Two subtracters
    (d) One subtracter and one adder

38. Cascading chips is usually required because:
    (a) Faster speed is required than a single chip can provide
    (b) More bits must be processed than a single chip can handle
    (c) The only chips that can provide all the necessary function require too much power
    (d) It minimizes the cost of a circuit

39. Dividing a binary number by 8 can be accomplished by:
    (a) Clearing the least significant three bits
    (b) Shifting left three bits
    (c) Shifting right three bits
    (d) Setting the least significant three bits

40. Mickey Mouse logic solutions should be placed in the circuit:
    (a) In the middle of a logic string
    (b) On the inputs of a logic string
    (c) On the outputs of a logic string
    (d) Where high-current I/O is required

41. The resistor used in the Mickey Mouse logic AND gate shown in Fig. Test 1-1 should be:
    (a) 10 k for TTL applications
    (b) 10 k for CMOS applications
    (c) The complementary one specified by the diode's manufacturer
    (d) Power rated for the load current of the application

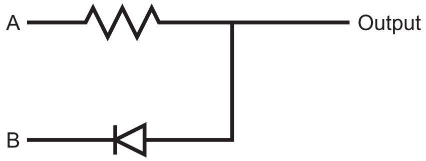

Fig. Test 1-1.

42. Each item below is a disadvantage of a dotted AND bus except:
    (a) High power consumption when the output is low
    (b) The dotted AND bus has a slower response than tri-state buffers
    (c) The dotted AND bus is cheaper than one manufactured with tri-state drivers
    (d) It is very difficult to find open collector output chips

43. When providing multiple functions to a net, what logic technology/ technologies should be used?
    (a) Just CMOS
    (b) Just TTL
    (c) CMOS receivers and TTL drivers
    (d) TTL receivers and CMOS drivers

44. Sequential circuits contain:
    (a) Memory devices
    (b) Power supplies
    (c) Input and output devices
    (d) CMOS logic

45. Backdriving gates can:
    (a) Simplify your application design
    (b) Speed up gate operation
    (c) Change the input of a downstream device
    (d) Burn out the gate's output transistors

46. What is the difference between "Q_0" and "_Q_0"?
    (a) There is no difference
    (b) "Q_0" is correct earlier than "_Q_0"
    (c) "Q_0" is current state of the flip flop and "_Q_0" is the previous
    (d) "_Q_0" is the inverted value of "Q_0"

47. The "_Clr" pin of a D flip flop will:
    (a) Set the bit
    (b) Reset the bit
    (c) Nothing

(d) Toggle the state of the bit

48. Which full D flip flop input pin is typically connected to the RC delay circuitry?
(a) D
(b) Clk
(c) _Clr
(d) _Pre

49. Which application is a register best suited for?
(a) Main memory in a computer system
(b) Permanently storing access passwords
(c) LED output states
(d) Temporary storage of data in a microprocessor

50. "Volatile memory" means:
(a) The contents of the memory device will not be lost when power is taken away
(b) The contents of the memory device will be lost when power is taken away
(c) The memory device is made up of a liquid which will evaporate if the chip package is broken
(d) Data is stored as patterns of a condensed gas

This page intentionally left blank

PART TWO

Digital Electronics Applications

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

This page intentionally left blank

CHAPTER 8

Oscillators

In the introduction to the previous chapter of this book, I presented you with a simple block diagram of a "complete" digital electronic device: the digital clock. This device has all three of the necessary components for a functional, "stand alone" device: combinatorial logic for converting binary data as required; a memory function which "remembers" the last state it was in; and a clock or "oscillator" which synchronizes the functions together. The science of oscillator design is extremely rich and, as I will show in this chapter, there are a lot of options that you can choose from to make sure your application operates most efficiently.

It could also be argued that there is a fourth component to producing a complete, stand alone digital device – the power supply. Power supply design is a facet of electronics which is just as rich and sophisticated as digital electronics or any other major study in electronics. While I introduce you to some of the basic types of power supplies that are available to you later in the book, this does little more than just scratch the surface of this complex topic.

The application's "clock" is a set of repeating pulses (ideally with the same "on" and "off" time) which is input into the sequential circuits of an application to carry out the operations within them. Figure 8-1 shows an ideal digital electronic clock waveform with the important features marked

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

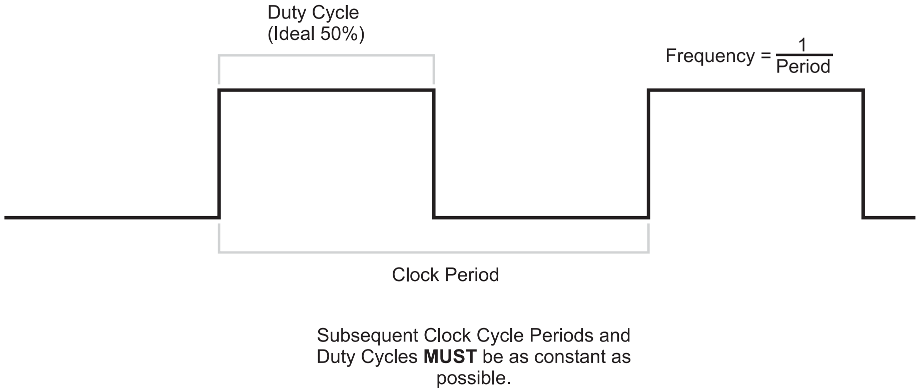
Duty Cycle
(Ideal 50%)
Frequency = $\frac{1}{\text{Period}}$
Clock Period
Subsequent Clock Cycle Periods and Duty Cycles **MUST** be as constant as possible.

Fig. 8-1. Features of a clocking signal.

on it. The frequency output of a clock is measured in "hertz" (cycles per second), which is the reciprocal of the time the pulse is on added to the time the pulse is off:

$$\texttt{Frequency} = 1/(\texttt{Time "On"} + \texttt{Time "Off"})$$

A rather obvious example of this would be a 1 pulse per second signal that is used to drive the sequential circuits of a timekeeping clock. The term "clock" is probably confusing, as the digital "clock" that I have defined is only very rarely used to tell the time. The clock signal that I am discussing in this section is used to drive the digital counters of the timepiece.

The "clock" in a digital circuit is driven from an "oscillator" that uses some form of feedback to toggle the clock line in a consistent manner. In this book, you will find that I use the terms "clock" and "oscillator" interchangeably, with the clock or oscillator signal being responsible for the operation of the digital circuit.

In Fig. 8-1, I noted the important features of a clock signal to be its constant period as well as its 50% duty cycle. If the clock is not constant and you were to look at the operation of the clock and the digital circuit on an oscilloscope, you would see the signals blur as they "jittered" back and forth. The shortened or lengthened features of a clock waveform are called "jitter" and are shown in Fig. 8-2. Jitter is a problem from a couple of perspectives. First, it makes it very difficult to observe the operation of a circuit using an oscilloscope, making it hard to debug an application. Secondly, the combinatorial logic parts of sequential circuits are often designed to have completed their bit data processing by the time the clock has completed; a shortened clock cycle, like the one shown in Fig. 8-2 could result in incorrect data being stored in a circuit.

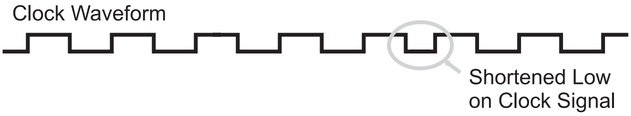

Fig. 8-2. Clock waveform with ‘‘jitter.’’

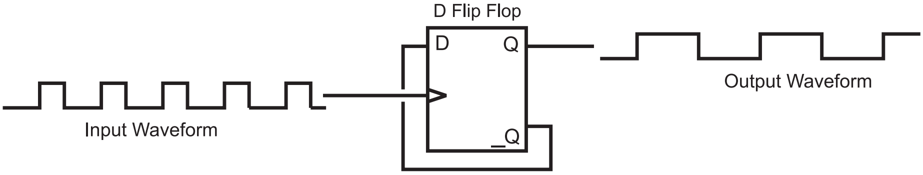

Fig. 8-3. Clock divide by 2 conditioning circuit.

Incorrect logic operation caused by jitter is extremely difficult to capture on an oscilloscope or a logic analyzer. If you have errors and have a clock with jitter, you would be well advised to assume that the jitter is the problem and look towards another clocking scheme. Jitter is often caused by voltage transients caused by different parts of a digital electronic circuit changing state; often it can be reduced or eliminated by providing better noise filtering between the oscillator circuit and the digital logic devices.

In Fig. 8-1, I indicated that the ideal ‘‘duty cycle’’ of a clock signal is 50%. I will discuss duty cycles in more detail later in the book when I present PWMs, but for now, you should understand that the duty cycle is the percentage of the period in which the waveform is high. Ideally, the clock should have a 50% duty cycle (be high for half the period) to minimize harmonics, simplify using an oscilloscope or logic analyzer to observe the operation of the circuit and, because some digital electronic devices (most notably microprocessors) poll both the high and low of the clock signal, to help speed up its operation.

In some oscillator circuit designs, the output does not come close to having a 50% duty cycle and in these cases, some kind of signal ‘‘conditioning’’ is required. The most basic way of ensuring the clock signal has a 50% duty cycle is to use an edge triggered D flip flop as I show in Fig. 8-3. The D flip flop (called a ‘‘toggle flip flop’’ when wired this way) will only change its output when a triggering edge on the input clock has been received. The only drawback to this circuit is that it halves the frequency of the clock, so, in some cases, to use this circuit you will have to double the clock output frequency.

There are many different designs of oscillators that you can choose from. In this chapter I will introduce you to many of the most common ones along with their characteristics and the formulas required to work with them.

Transistor Astable Oscillators

When I was growing up, all educational and hobbyist circuits were built up from individual transistors – it wasn't until the mid to late 1970s that ''building block'' chips such as the 555, LM339, LM386, and LM741 started to be commonly used in circuits. These chips are all very configurable, but none offer the range of operation and low cost of discrete transistors. The term ''astable'' indicates that the oscillator circuit is never stable; its output will continue to switch from high to low and back again. The science of oscillators can be thought of as ''taming'' the oscillator in terms of frequency, duty cycle and jitter.

A very common and relatively simple oscillator circuit that I am going to examine is the basic ''relaxation oscillator'' circuit shown in Fig. 8-4. Included in Fig. 8-4 are the defining formulas for the time that the output is high and low as well as an important formula indicating that the value of R1 and R2 (the time defining resistors) must be the transistor h_{FE} multiplied by the value of the pull up resistors (Rpu). If R1 or R2 is less than this product, then you will find that the oscillator will not start reliably and not run at a constant frequency.

The operation of the relaxation oscillator is illustrated in Figs. 8.5 through 8.7. In Fig. 8-5, I show an initial condition where one transistor is on and the other is off. In this case, the capacitor by the on transistor is charging because its cathode is being pulled to ground by the ''on'' transistor. The other capacitor is unable to be charged because the transistor connected to its cathode is off, holding the voltage at the cathode at the same voltage as the anode.

In Fig. 8-6, the capacitor that was charging in Fig. 8-5 has finished and any current passing through the resistor is passed to the other transistor, turning it on. By turning on this transistor, the capacitor's cathode connected to its collector is now tied to ground and it is able to be charged. With this capacitor now charging, the current that was once available to the

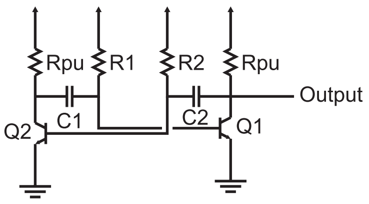

$\text{Output}_{High} = \text{Q1}_{Off} = 0.69 \times \text{R1} \times \text{C1}$

$\text{Output}_{Low} = \text{Q1}_{On} = 0.69 \times \text{R2} \times \text{C2}$

$\text{R1} = \text{R2} = h_{FE} \times \text{Rpu}$

Fig. 8-4. Basic transistor relaxation oscillator.

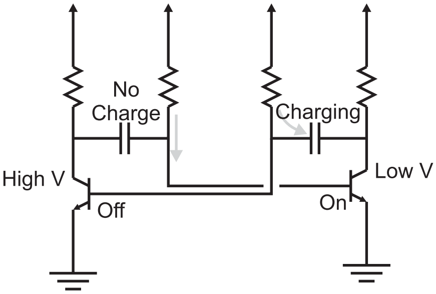
No
Charge
Charging
High V
Off
Low V
On

Fig. 8-5. Relaxation oscillator charging right-side capacitor.

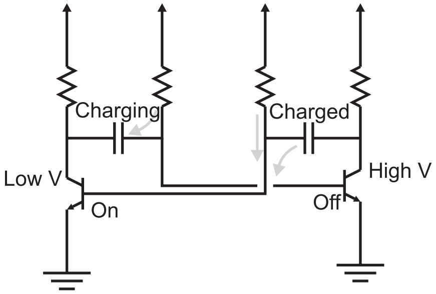
Charging
Charged
Low V
On
High V
Off

Fig. 8-6. Relaxation oscillator discharging right side.

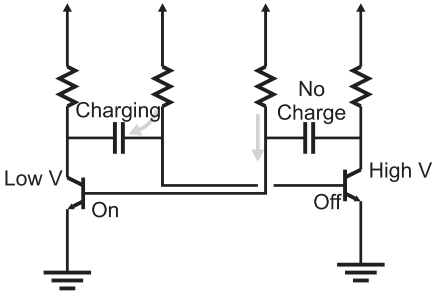
Charging
No
Charge
Low V
On
High V
Off

Fig. 8-7. Relaxation oscillator charging left side.

transistor's base to turn it on is no longer available and the transistor turns off. This raises the collector of this transistor to the applied voltage and, along with this, the cathode of the capacitor connected to it. This places the charge in the capacitor tied to the collector of the transistor just turned off at a voltage higher than the applied voltage, so its charge is now passed to the transistor that was just turned on. In the final case (Fig. 8-7), the operation of

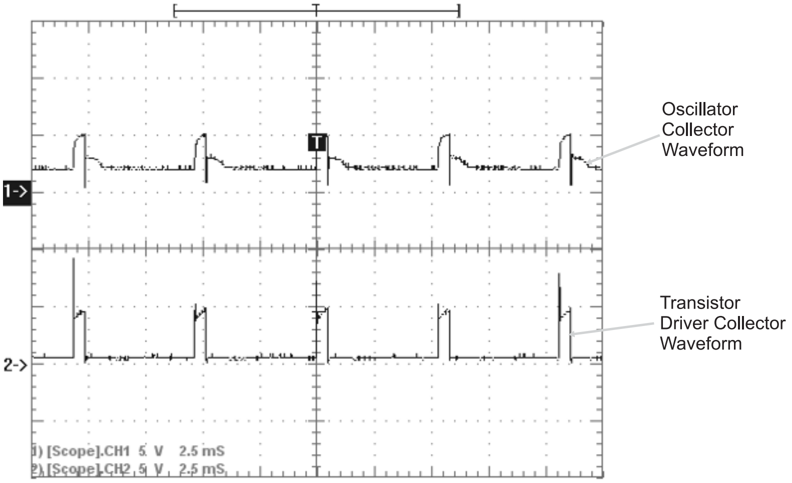

Fig. 8-8. Relaxation oscillator waveforms.

the oscillator circuit is the mirror image of the initial conditions shown in Fig. 8-5. When the charging capacitor is finished, its current is passed to the transistor that is currently turned off and the process repeats itself.

The output of this oscillator is probably nothing that you would expect – Fig. 8-8 is an oscilloscope display of a sample NPN transistor relaxation oscillator output as well as the collector voltage. The duty cycle of the waveform is nowhere close to 50% (which means it will have to be conditioned by some kind of circuit, like the one in Fig. 8-3. However, this circuit is quite good in terms of accuracy, with very little jitter.

The drawbacks to using a transistor oscillator like the one presented in this section include the unusual waveform output and the use of discrete analog components for timing the oscillator. The unusual waveform output makes the need for some kind of signal condition mandatory when working with digital electronics and the use of analog components makes the frequency output quite imprecise. The characteristics of the transistor-based oscillator make it best suited for low-cost applications where clocking accurate to 20% is acceptable.

Ring Oscillators

In the previous chapter, I introduced the concept of “ring oscillators” as being a digital electronic device in which an inverted output signal is fed back to the input of a combinatorial circuit and showed that it could be created inadvertently (Fig. 8-9) or purposely using a single logic inverter (Fig. 8-10). One of the useful characteristics of the ring oscillator is that it will always produce a 50% duty cycle and its output is literally the maximum speed of the technology.

I should say that the ring oscillator's maximum output is the maximum speed of the technology. In Fig. 8-10, I have drawn two ring oscillators, the first outputting the signal from a single inverter – the period of the output of this circuit will be 1 gate delay. In the lower diagram of Fig. 8-10, I show that you are not limited to just running at the technology's maximum speed; by adding an even number of additional inverters to the ring oscillator, the output signal's period can be lengthened.

The ring oscillator's actual frequency output can be ''tuned'' by varying the number of inverters in the ring oscillator along with the technology used in the inverters. Knowing this, along with a couple of operating rules, provides you with an inexpensive, high-speed oscillator that is quite reliable and robust.

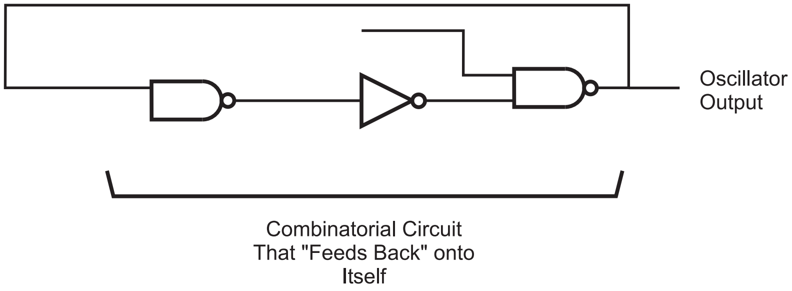
Oscillator
Output
Combinatorial Circuit
That "Feeds Back" onto
Itself

Fig. 8-9. Inadvertent ring oscillator.

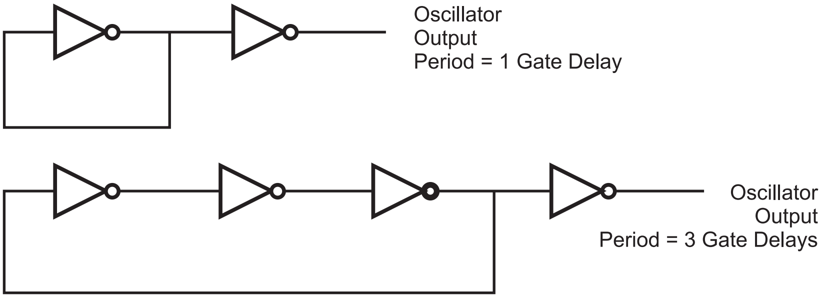
Oscillator
Output
Period = 1 Gate Delay
Oscillator
Output
Period = 3 Gate Delays

Fig. 8-10. Ring oscillators using gate delays for frequency determination.

The first of the operating rules should not be surprising because I have alluded to it in the text above: the number of inverters in a ring oscillator should always be odd. If an even number of inverters is used in a ring oscillator, there will be no signal which cannot be resolved (which is the cause of the "astable" operation of an oscillator) and the circuit will not oscillate. The second operating rule is that no other functions should be used by the leftover gates in the chip and the chip's power pins should have both small (0.1 μF or less) and large (1.0 μF or greater) decoupling and filter capacitors on its power supply. The oscillating gates within the chip are experiencing significant transients which could affect the operation of other devices in the application.

While I have only used ring oscillators a couple of times over the years, they are fascinating circuits to build and watch executing. The ring oscillator is what I consider to be a "hip pocket" circuit: something to be pulled out only when nothing else seems to work or have the characteristics that you require.

Relaxation Oscillators

The most basic type of logic chip based oscillator is the "relaxation" oscillator which feeds back the output of an inverter through a "resistor/capacitor" ("RC") network to delay the switching of the oscillator. The basic circuit and its defining output equation is given in Fig. 8-11.

In this circuit, the R1 and C network are driven by the first inverter and the characteristic "RC" response is fed back to the first inverter's input. When the voltage on the capacitor reaches the threshold voltage of the left inverter input, the inverter changes state and drives a new output voltage. This voltage is again passed through the R1, C network and delayed until the threshold voltage is reached again.

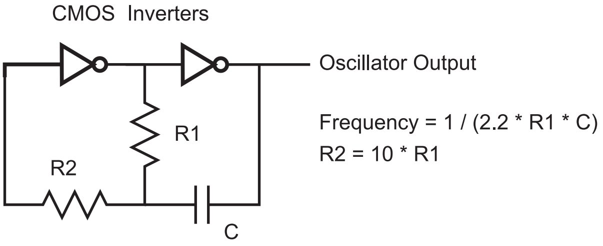

Fig. 8-11. CMOS logic technology relaxation oscillator.

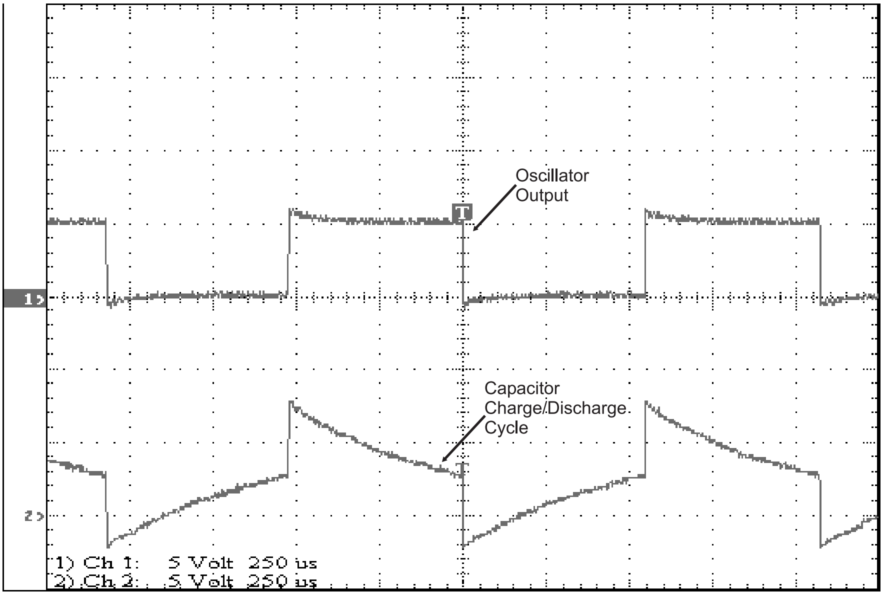

Fig. 8-12. CMOS relaxation oscillator waveform.

In Fig. 8-12, I have shown the voltage waveforms at the R1, R2 and C junction of this circuit as well as the output voltage signal. Note that the R1, R2 and C junction voltage exceeds the Vcc/Gnd (+5 volts and 0.0 volt) limits. This is due to the capacitor being connected to the output driver.

Having the capacitor wired to the output driver "moves" the charge (and the capacitor voltage) by 5 volts each time the state changes. Observing the circuit's operation from the capacitor, the output value changes the charge within the capacitor until it is back at the threshold voltage for the CMOS inverter, which is 2.5 volts (one-half applied power). You can see in Fig. 8-12 that the transitions take place every time the voltage across the capacitor is at 2.5 volts relative to Gnd.

CMOS inverters are used in this circuit because they are voltage controlled rather than current controlled and this makes the oscillator's operation easier to understand. A TTL inverter cannot be used in this circuit because of the current drain operation required by the input when a "0" is input will affect the operation of the oscillator. A Schmidt trigger input device (i.e. the 74HC14) could be used, but it is not necessary because the reference voltage of the capacitor is changing with every transition.

You may want to test out this circuit with a 74C04 or 74HC04 with a 4.7 k resistor, a 47 k resistor and 0.1 μF capacitor to create an oscillator that produces a clock signal of approximately 1 kHz. I say that the output is "approximately" 1 kHz because of the tolerances of the parts used in the circuit. For the circuit used to produce the signal shown in Fig. 8-12, I used a 0.1 µF tantalum capacitor for the "C" in the relaxation oscillator circuit. This is probably not a "correct" use of a tantalum capacitor, as they can have tolerances approaching 30% of their rated value – I only used it because I have a lot of them around. Along with the tolerance of the capacitor, there are also the tolerances of the resistors in the circuit to consider as well. These tolerances result in the opportunity for the actual clock signal to be "out" by 40% or more.

Your immediate response may be to add a potentiometer (variable resistor) into the circuit and "tune" it to the exact frequency that you want. Personally, I would discourage this practice as it involves a lot of work (especially if production parts are involved), which will drive up the cost of the product. If you are using a simple RC relaxation oscillator in your application, then additional costs are something that you would want to avoid. The relaxation oscillator is adequate for many applications where a low-cost oscillator of an approximate value is required. Like the NPN transistor astable oscillator, I recommend that the circuit should not be used in any applications where any kind of precision is required.

Another aspect of this circuit that you must be aware of is the potential for large current transients within the chip that are produced to change and discharge the capacitor. These transients are similar to the transients discussed in the ring oscillator. For most circuits, this is not a problem, but if you have other sensitive circuits built into an application, you will want to keep the relaxation oscillator (as well as any other oscillators in the circuit) as electrically removed as possible from the other chips by using both large and small decoupling and filtering capacitors. Also like in the ring oscillator, as a rule of thumb, no other gates should be used in a chip if it is being used as an oscillator.

Crystals and Ceramic Resonators

For the best clock accuracy, a *quartz crystal* should be used in an oscillator circuit like the one shown in Fig. 8-13. A quartz crystal is a *piezo-electric* device that provides a constant delay between one side of the piece of quartz within the part to the other. The term piezo-electric refers to the property of quartz (and some other compounds) to mechanically deform when a current is applied to it or produce a voltage potential when it is mechanically deformed.

In an oscillator, the quartz crystal will have a voltage applied to one end of it and this will cause the quartz crystal to deform. The rate at which this deformation takes place is known and will cause a voltage potential to be produced at the other end of the quartz crystal after a known delay. This voltage is used as a feedback value to an inverter built into the oscillator circuit. The NPN bipolar transistor-based inverter can be seen in Fig. 8-13.

The circuit in Fig. 8-13 is somewhat ''fiddly'' to build and to get working reliably. There are some formulas that can be used to specify the different resistor, capacitor and inductor values, but, personally, I would never use this circuit in my own applications. This is why I did not put in any component values on the diagram; instead, I would use the inverter-based oscillator shown in Fig. 8-14. In this circuit, instead of understanding a circuit well enough to specify the correct different analog values, you can simply put a crystal across the input and output of a CMOS inverter.

The capacitors and resistors are necessary to ensure that the oscillator runs reliably and there are not any large over- or under-voltage spikes (caused by the operation of the piezo-electric producing its own voltage output). For most MHz range oscillator circuits, 15–33 pF capacitors are

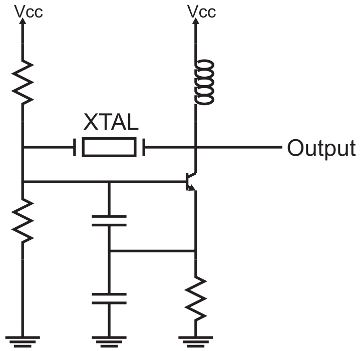

Fig. 8-13. Crystal oscillator circuit.

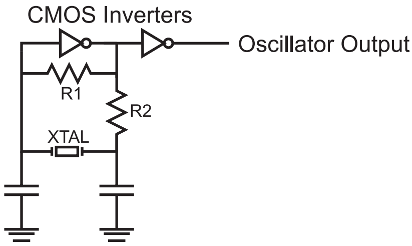

Fig. 8-14. CMOS logic gate crystal oscillator circuit.

adequate, 1–10 MΩ for R1 and 100 to 100 kΩ for R2 is appropriate. You may find that depending on the frequency of the crystal that you choose and how it is wired, you may have to vary these parts.

The second inverter in the circuit is not ‘‘strictly’’ required, but I like to have it in place to ensure the crystal is not ‘‘loaded down’’ by other devices and the operation of the oscillator doesn’t change. Any large loads on the output side of the oscillator’s primary inverter will affect the amount of current/voltage available to the crystal to pass the signal to the other side (and the oscillator’s frequency will drop or the oscillator won’t work at all).

Changing the capacitance on the inverter output side of the primary inverter results in small (1–2%) changes to the output and to help ensure the absolutely correct frequency output is produced a variable capacitor is used in place of the fixed capacitors. I do not feel this is practical and the nominal 0.01% or less error rate of the crystal should be accepted. I realize that there are applications (like digital clocks) where these changes are critical, but for the most part you should not have to ‘‘tune’’ the oscillator for the application.

Crystals work quite well, although there are two drawbacks that you should be aware of. Crystals are relatively expensive parts (especially compared to the RC network relaxation oscillator). You can pay up to $10 for a crystal (although you can pay less than 1$ for common frequencies). In addition, the oscillator is somewhat ‘‘fragile’’ and can be easily damaged by rough handling. A relatively new device that can be used in place of a crystal and does not have these shortcomings is the ‘‘ceramic resonator’’. A ceramic resonator is used in a very similar way to a quartz crystal (Fig. 8-15), but it is usually much less expensive and very rugged. I use ceramic resonators almost exclusively for clocking all my microcontroller applications. Despite the somewhat poorer accuracy of the parts (they are usually accurate to 0.5%), they really are the part of choice for most applications.

Many designers eschew the use of oscillators built from discrete parts as I have shown in this section. These circuits are rarely used because of the difficulty in specifying the correct parts for an application, the cost of the crystals and the potentially large amount of ‘‘real estate’’ that they can take up. Instead, oscillators are usually implemented using some kind of ‘‘canned’’

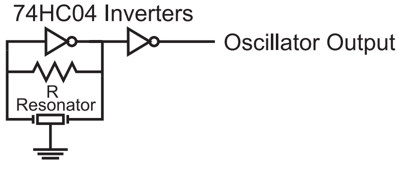

Fig. 8-15. Ceramic resonator oscillator.

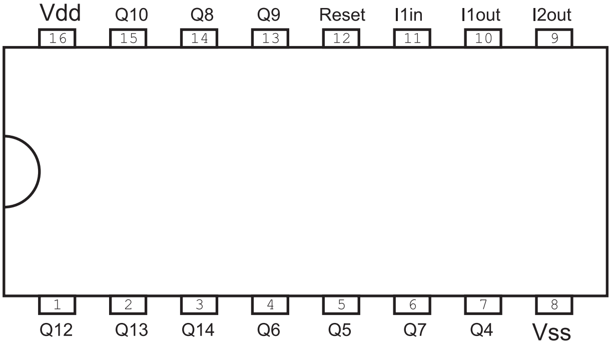

Fig. 8-16. CMOS 4060 chip pinout.

solution. These parts are designed to take up the same "footprint" as an eight or 14 pin "DIP" package and normally have four pins – power, ground, oscillator output and oscillator enable.

One of the most common chips used in digital logic applications is the CMOS 4060 (shown in Fig. 8-16). This chip can be used with the different oscillator types listed in this section and the "divide by" outputs are very handy in many circuits (often eliminating the need for separate counters). The crystal, ceramic resonator and relaxation oscillator circuits that I have shown in this chapter can be used with this part. The "Q4" through "Q13" outputs are divided by counters (i.e. "Q4" is the clock divided by 2 to the 4 or 16 times).

When using the 4060, note that Pin 11 is the input to the inverter used in the oscillator circuits shown in this chapter, Pin 10 is output of the first inverter and input of the second inverter while Pin 9 is the output of the second inverter. If Pin 12 ("Reset") is pulled high, the oscillator is stopped and the counters in the chip are reset.

555 Timer Chip

The 555 timer chip is probably the most versatile non-programmable part I have ever seen. Over the past 40 years, many people have created at least hundreds probably thousands of applications that have used this chip in ways I'm sure the original designer never would have thought possible; the original function of the chip was to provide a regular train of pulses. In this section, I will show how the chip is used in a circuit, along with some of the tricks that can be performed with it.

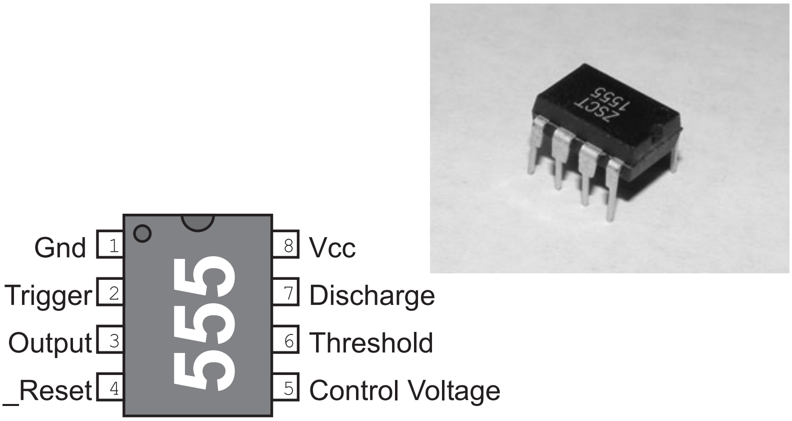

Fig. 8-17. 555 pinout.

In the previous sections, I have shown you the "pinout" of a number of different components – each one of them having a unique form factor. The 555 is usually built into an eight pin "dual in-line package" that is commonly used for chips. In Fig. 8-17, I have put in an "overhead" view of the 555, along with a photograph of an actual 555 chip.

Looking at the labels for each of the pins, most of them do not make a lot of sense. What should jump out at you is the "Gnd" (ground) at Pin 1 and the "Vcc" (positive power) at Pin 8. These two pins are used to provide power for the part; they match the power pins I've presented elsewhere for digital devices elsewhere in the book.

To try and get a better understanding of a chip, one of the first things I do is look for its block diagram and try to understand it. In Fig. 8-18, I have drawn out the block diagram for the 555 timer.

There should be two parts to the block diagram that you should recognize immediately. The first is the transistor at the bottom middle of the diagram. This transistor is wired in an open collector configuration and is acting as a switch that will pass current to ground. The next piece that you should recognize is the voltage divider running along the left side of the block diagram that I have separated out into Fig. 8-19. If you were to work out the voltages at "Vcontrol" and "Vtrig", you would discover that they are at 2/3 Vcc and 1/3 Vcc, respectively. This is actually an important clue as to how the chip works.

One aspect of the 555's voltage divider circuit that you may find confusing is its connection to an outside pin called "Control Voltage". As I have shown in Fig. 8-20, this connection allows the circuit designer to change the voltage

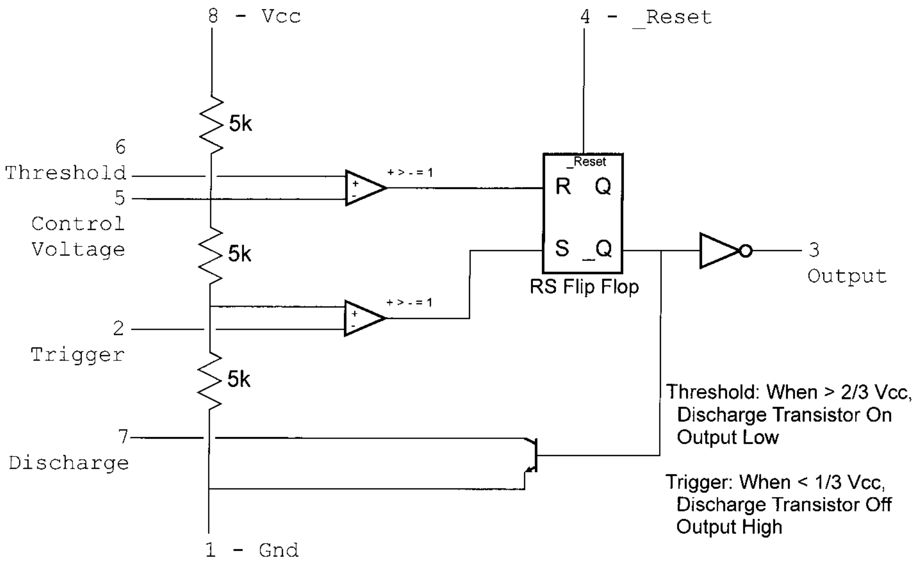
8 - Vcc
4 - _Reset
5k
6
Threshold
5
Control
Voltage
5k
2
Trigger
5k
7
Discharge
1 - Gnd
+>-=1
+>-=1
_Reset
R Q
S _Q
RS Flip Flop
3
Output
Threshold: When > 2/3 Vcc,
Discharge Transistor On
Output Low
Trigger: When < 1/3 Vcc,
Discharge Transistor Off
Output High

Fig. 8-18. 555 block diagram.

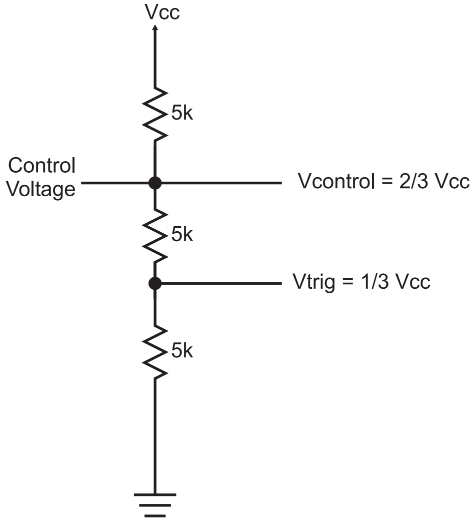
Vcc
5k
Control
Voltage
Vcontrol = 2/3 Vcc
5k
Vtrig = 1/3 Vcc
5k

Fig. 8-19. 555 voltage reference/voltage divider.

levels of the voltage divider circuit. Rather than "Vcontrol" being 2/3 Vcc, it can now be any value (less than Vcc) that the designer would like. Changing "Vcontrol" also changes "Vtrig" to 1/2 "Vcontrol", as I have shown in Fig. 8-20.

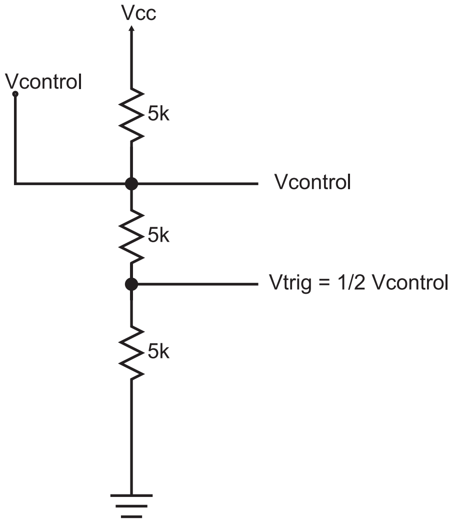
Vcc
Vcontrol
5k
Vcontrol
5k
Vtrig = 1/2 Vcontrol
5k

Fig. 8-20. Modified 555 operating voltage threshold.

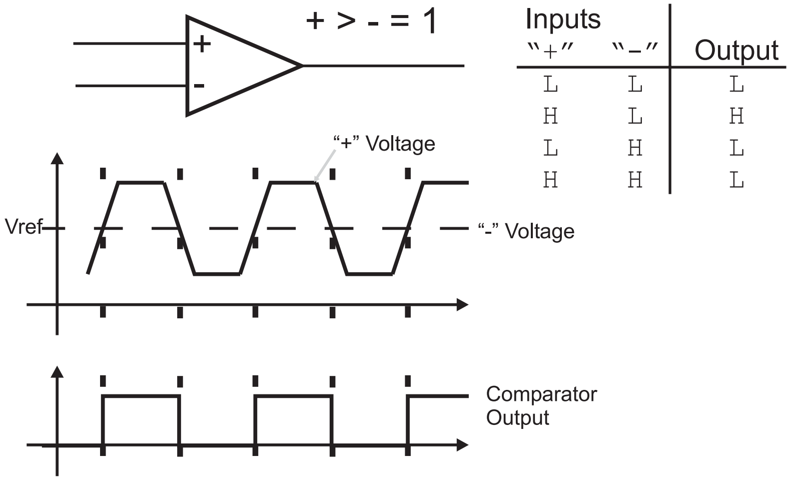
+
-
+ > - = 1
"+" Voltage
Vref
"-" Voltage
Comparator Output

Inputs "+"	Inputs "−"	Output
L	L	L
H	L	H
L	H	L
H	H	L

Fig. 8-21. Voltage comparator operation.

The voltages at the "Vcontrol" and "Vtrig" are passed to two triangular boxes with a "+" and "−" along with a funny looking equation. These boxes are representations for voltage "comparators" and, as I have shown in Fig. 8-21, the comparators output a high voltage level when the voltage at the "+" input is greater than the voltage at the "−" input. The 555 uses the two comparators to continuously compare two external voltage levels to "Vcontrol" and "Vtrig" and pass the results to a box labeled "RS flip flop".

The 555's RS flip flop saves an indication of which comparator last passed a high voltage to it. If the comparator connected to the "threshold" pin of the 555 and "Vcontrol" of the voltage divider output a high voltage, then the flip flop will output a high voltage at "_Q", which turns on the transistor at the bottom of the block diagram. If the other comparator passes a high voltage to the "RS flip/flop", then the voltage at "_Q" is driven low and the transistor is turned off. This is a fairly complete explanation of how the 555 works and I'm sure that you are at least as confused as you were when I first showed you the block diagram of the chip. The individual parts are quite easy to understand, but I'm sure you're mystified how they work together.

When I described the operation of the 555 chip, I neglected to take into account the components that would be wired to it. The timing delay that is integral to the operation of the 555 is produced by resistors and capacitors wired in the "RC networks" that I have described earlier in the book. What I didn't go into detail on in the previous sections of the book is that as you change the value of the resistor or capacitor in the circuit, you will change the delay produced by the two components (Fig. 8-22).

In Fig. 8-23, I have drawn a 555 oscillator circuit; when this circuit starts running, the 555 will be an "astable" oscillator with the output toggling,

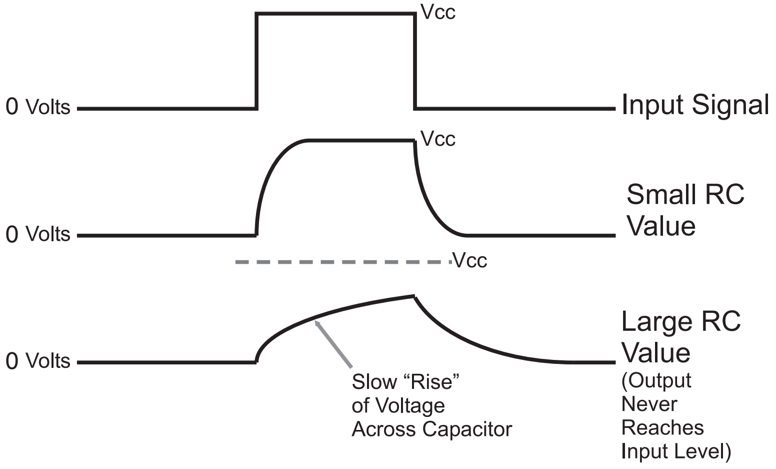
Vcc
0 Volts
Input Signal
Vcc
Small RC Value
0 Volts
Vcc
Large RC Value
(Output Never Reaches Input Level)
0 Volts
Slow "Rise" of Voltage Across Capacitor

Fig. 8-22. RC network operation for ranging values of RC.

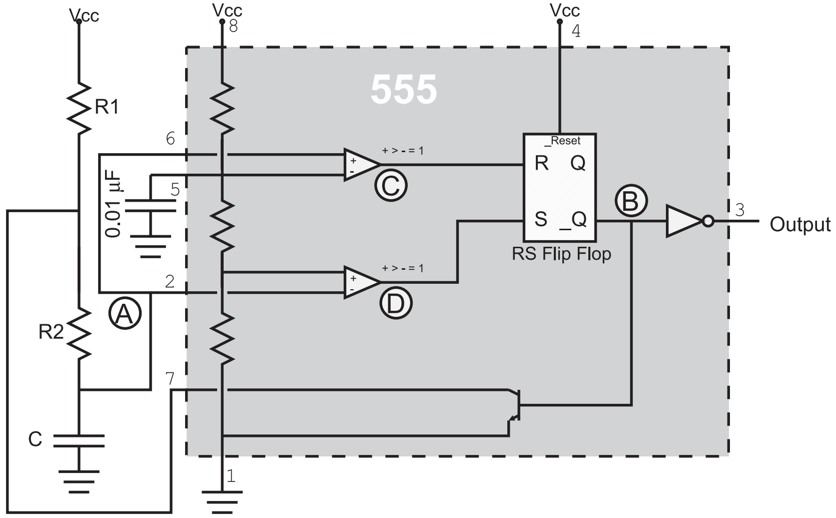

Fig. 8-23. 555 astable oscillator.

according to the values R1, R2 and C. The 555's output times are defined by the formulas:

$$T_{high} = 0.693 \text{ x } C \text{ x } (R1 + R2)$$

$$T_{low} = 0.693 \text{ x } C \text{ x } R2$$

$$\text{Period} = 0.693 \text{ x } C \text{ x } (R1 + 2R2)$$

$$\text{Frequency} = 1.443/(C \text{ x } (R1 + 2R2))$$

When choosing the resistor and component values for working with the 555 timer, you should only use components that are within the ranges listed below:

$$10k \leq R \leq 14M$$

$$100\,pF \leq C \leq 1{,}000\mu F$$

The 0.01 µF capacitor wired to the "control voltage" pin of the 555 (in Fig. 8-23) is used as a "filter" for the internal voltages. This capacitor works very similarly to the logic circuit's decoupling capacitor; if the input voltage changes, the capacitor will absorb or release charge to keep the voltage as even as possible.

To get a better idea of how the 555 timer works as an oscillator, in Fig. 8-23, I labeled the RC voltage ("A"), the RS flip flop output ("B" – which is in the inverted 555 output), the "threshold" comparator voltage

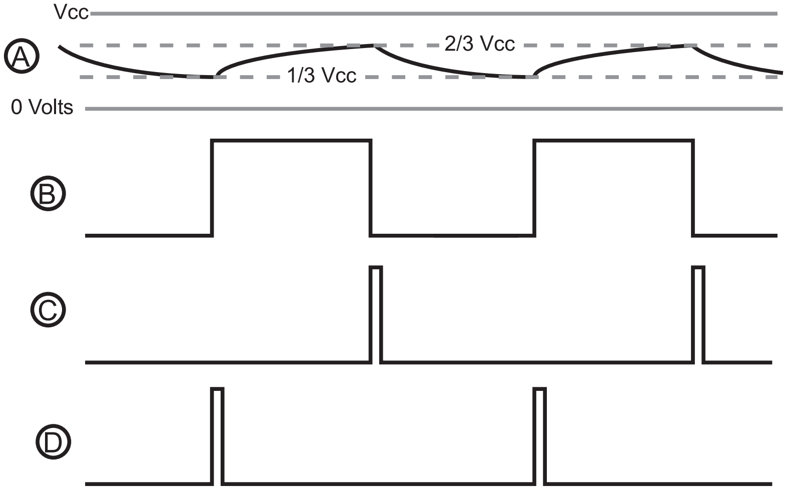

Fig. 8-24. 555 astable operation.

("C") and the "trigger" comparator voltage ("D"). Figure 8-24 shows the waveforms for each of these parts marked in Fig. 8-23, so you can see the changing RC waveform, the output from the two comparators and the action of the RS flip flop.

Before going on, I want to share with you some of the more clever and useful circuits that have been created using the 555 timer. The few I will show in this chapter are just a small fraction of the number that is possible or has already been developed. If you are in a used book store, you should look for a copy of Don Lancaster's "555 Timer Cookbook"; it will really open your eyes to the incredible variety of applications this chip can help implement.

When stretching the envelope and using the 555 timer in a way that it wasn't originally designed for, you generally look at the different input pins and see how they can be given a completely different function. If you wanted to make a circuit that drove out a tone for a set amount of time, you could use two cascaded 555 timers (or a single "556", which consists of two 555 timer chips in a single package) or the circuit shown in Fig. 8-25.

When this circuit is first turned on, the voltage at Pin 4 (the RS flip flop reset pin) is low, stopping the circuit from oscillating. When the momentary on button is pressed, current is passed to the 10 μF capacitor, charging it and driving up the voltage on Pin 4. This happens quite quickly and when Pin 4 reaches the logic threshold to stop holding the RS flip flop reset, then the 555 will start oscillating, driving out a signal that oscillates at 464 times per second.

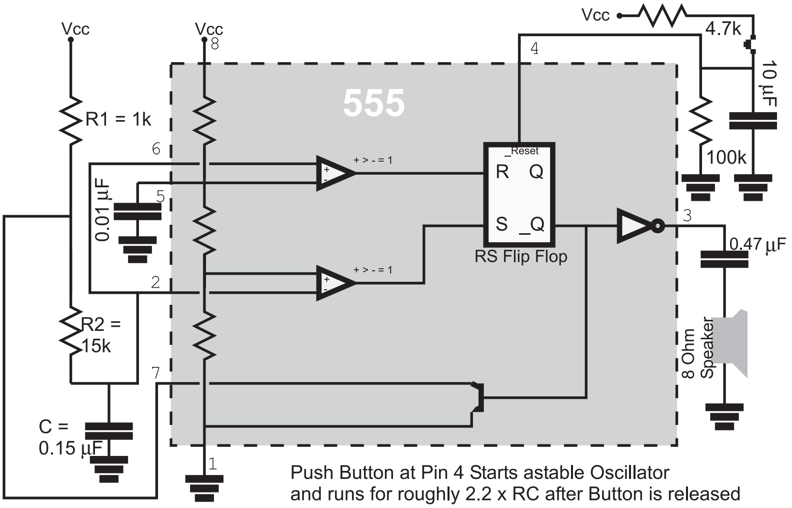

Push Button at Pin 4 Starts astable Oscillator
and runs for roughly 2.2 x RC after Button is released

Fig. 8-25. 555 tone output.

This tone will stay on as long as the button is pressed and then will continue until the 10 µF capacitor discharges through the 100 k resistor. By varying the values for these two parts, you can vary the length of time the 555 continues to oscillate. Remember the rule of thumb that is used to approximate the time for an RC network, like this one, to discharge:

$$\mathtt{T_{discharge}} = \mathtt{2.2\ x\ R\ x\ C}$$

I take advantage of the ability of the human ear to distinguish between different sounds and categorize them with a circuit design for a continuity tester that you will probably find is a lot more useful than the simple instrument built into your digital multi-meter.

This continuity tester circuit shown in Fig. 8-26, is useful in a variety of different situations. Instead of just driving out a simple tone when an electrical path between the two probes has been found, it provides you with different tones and sounds, based on the resistance between the probes as well as an indication of whether or not a diode is between the probes. Because the circuit is self-powered, you can use it with circuits that are already working, without worrying about having a valid ground connection.

When there is no connection between the leads, the 5.1k resistor is part of the RC network that provides the delay for the 555 timer wired as an astable oscillator. With no connections, you will find that the oscillator outputs a tone that is at about 440 cycles per second (“hertz” or “Hz”). If there is a direct connection (or short circuit) between the two probes the 5.1 resistor is

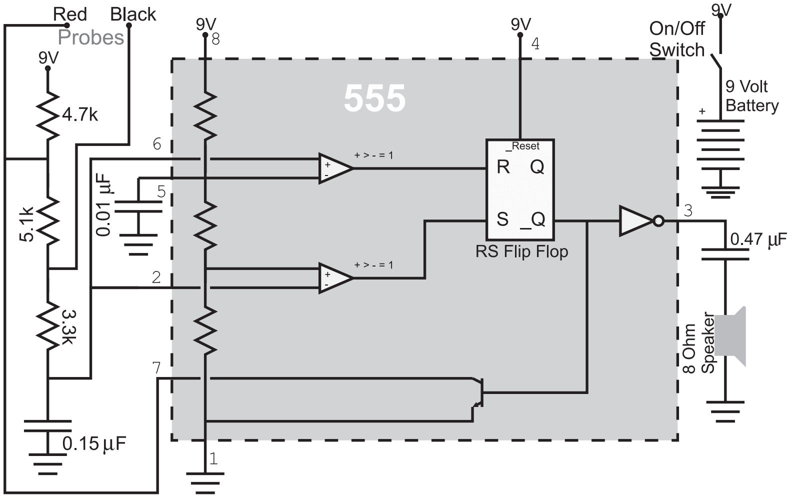

Fig. 8-26. 555 based continuity tester.

not used in the delay and the frequency output is around 880 hertz, there is a full octave difference between the two signals.

What I like about this circuit is that you get a different frequency based on the resistance across the probes – the probe resistance is in parallel with the 5.1 k resistor, changing its frequency and the frequency output from it. This can be useful in finding "almost" short circuits such as a "just touching" connection rather than a hard soldered connection or a "high impedance short", when you expect no connection at all. In addition, try out a diode in a forward biased and reversed bias connection; you will be able to hear a noticeable difference here as well.

Delay Circuits

Another basic function of the 555 is use as a "monostable". In the previous section, I alluded to this function and noted that the 555 could do more than just be part of the "astable" oscillator which will run for ever – the monostable, on the other hand, will only execute once and requires triggering. The monostable is very useful for a variety of different applications and works similarly to digital logic chips that can provide a similar delay. Figure 8-27 shows the 555 wired as a monostable generator and Fig. 8-28 shows the response waveforms to the input pulse.

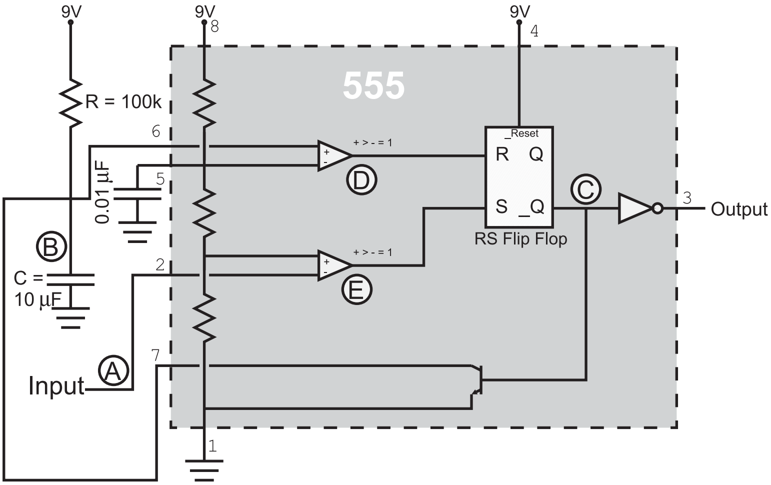
9V
9V
8
9V
4
555
R = 100k
6
0.01 µF
5
+ > - = 1
D
Reset
R Q
S _Q
RS Flip Flop
C
3
Output
B
C =
10 µF
2
+ > - = 1
E
7
Input
A
1

Fig. 8-27. 555 monostable circuit.

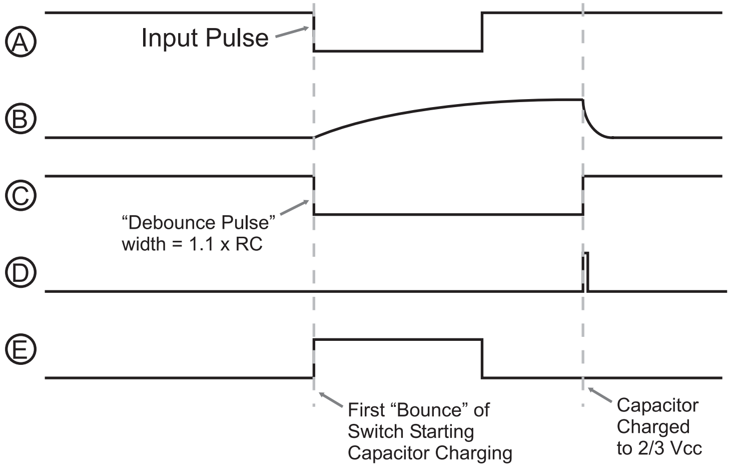

Fig. 8-28. 555 monostable operation waveforms.

When the input goes low, the pulse output from the 555 timer is determined using the formula:

$$T_{\text{pulse}} = 1.1 \times R \times C$$
$$= 1.1 \times 100\,\text{k} \times 10\,\mu\text{F} = 1.1 \text{ seconds}$$

The 555's RS flip flop is initially "reset", and the transistor that passes capacitor charge to ground is turned on. When the input ("A") goes low and the "trigger" input receives a low voltage input, its comparator signal ("E") goes high, changing the state of the RS flip flop (“C”) and changing the state of the output pin. When the RS flip flop state changes again, the transistor is turned off and the capacitor charges through the resistor. The capacitor charges according to the formula:

$$\mathrm{Output} = \mathrm{Vcc} - \mathrm{Vcc} \times e^{-t/RC}$$

until its voltage reaches 2/3 Vcc. When it reaches 2/3 Vcc, the “threshold” comparator (“D”) goes high and the RS flip flop changes state again, changing the value of the output pin (“C”). At this point, the 555 is back to its original state.

This circuit works quite well except for one point – the input pulse must *always* be shorter than the calculated output pulse. If the input pulse is longer than the calculated output pulse duration, you will find that the output will stay active, but it will pulse periodically. When the capacitor charges to 2/3 Vcc and the input is low down, both of the comparators will be driving a high voltage to the RS flip flop. This is an invalid condition for the RS flip flop and the output from the flip flop is “indeterminate”, resulting in the transistor tying the capacitor to ground periodically. To avoid this behavior, you should always make sure that the length of time for the pulse output from the 555 is longer than the expected input.

Instead of using the 555 timer as a monostable delay generator, there are a number of logic chips that perform the same function using a resistor and capacitor. The 74123 incorporates two monostable delays that are programmed using a resistor and capacitor. The 74123 and other logic family chips have the advantage that the voltage level transitions do not cause as much disruption to the surrounding circuitry.

In some applications, there is a delay that is either more precise, shorter or longer than can be practically created using the RC-controlled monostables that I have presented so far. In the next chapter, I will be introducing you to “counters” which will either count continuously or stop when a specific value has been reached. The counter, driven by one of the oscillators presented in this chapter, is used to produce either a very long or very precise delay.

For shorter delays, there are two methods that you can consider. The first method is to use a “canned” delay line. These components usually consist of an inverting buffer, driving a long copper line. At different points along the line, inverting “taps” are put in place to drive out the signal. Figure 8-29 shows how these components are used.

When you see an actual “delay line” component, you will probably refer to it as a “chip”. I hesitate to do so because I consider a chip to be simply a silicon chip bonded to a “lead frame” and “encapsulated” in some manner. In a delay line component (or “module”), the wire delay is wound in a coil

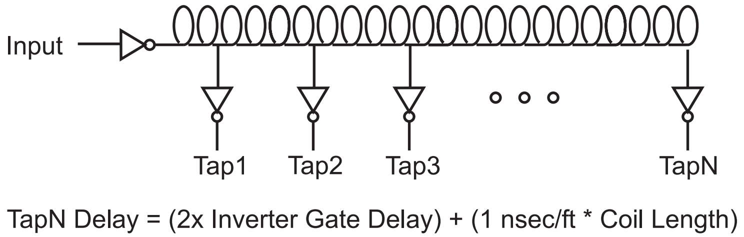
Input
Tap1
Tap2
Tap3
TapN
TapN Delay = (2x Inverter Gate Delay) + (1 nsec/ft * Coil Length)

Fig. 8-29. Delay circuit.

with the tap inverter inputs soldered to it at different intervals. The actual device requires very high precision mechanical assembly (more than a standard plastic encapsulated chip) and any errors in assembly or encapsulation will result in a useless part.

The wire in the part passes the digital signal to the various taps within the part, with a delay of roughly 1 nanosecond per foot (30 cm) of wire. When you design high-speed applications, this rule of thumb is very important when you are designing high-speed digital electronic circuits. Chances are you will ask that the traces on PCBs are "routed" with 0.1 inch (2.54 mm) precision to ensure that parallel signals all "show up" at the same place at the same time in the high-speed application circuit.

The advantage of the delay line module is that timing delay can be very precise. Custom-made delay line modules are available (the manufacturer solders the taps at specified points in the coil rather than at standard positions), which can be critical in some applications. This high level of assembly/encapsulation precision has a price that you will have to pay. If you can buy a 74LS04 for less than a quarter in single units, you should not be surprised to discover that a delay line module will cost you over $10.00. The delay line provides you with the best control over different delays required in a circuit, but at quite a significant cost. Delay line modules should only be considered if no other options are available to you when you are designing a circuit.

Another method of delaying signals is to take advantage of the natural delays of digital electronic gates and simply "chain" a number of them together to get a needed delay. In Fig. 8-30, I have shown a 20, 40 and 60 nsec delay built out of a 74LS04 TTL chip.

The advantage of this method is that it is quite low cost and reasonable precision can be built into the circuit. When you are designing delays for your applications, you should consult with the technology operational characteristics chart that I provided earlier in the book (see Table 6-2).

Working with different technologies, you should be able to get quite accurate delays quite inexpensively. The disadvantage of this method is that

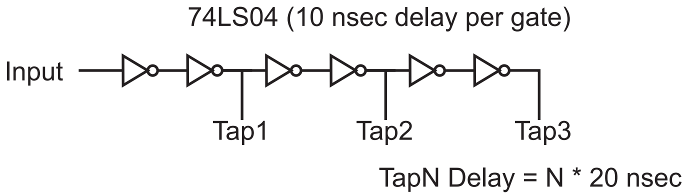

Fig. 8-30. Using logic gates for specific delays.

it can take up a lot of space on a board (at which point the canned delay line may have to be considered).

Quiz

1. An ideal digital electronic clock waveform has:
   (a) A constant period with a 50% duty cycle
   (b) A selectable speed range
   (c) A period that is less than the gate delay of the logic technology being used with it
   (d) A varying period that takes advantage of the operation of the combinatorial circuitry in the sequential circuit
2. Each of the following are important characteristics of astable oscillators except for
   (a) Period
   (b) Jitter
   (c) Duty cycle
   (d) Power required
3. If the R1 or R2 resistor values are less than the product of h_{FE} and Rpu, the NPN transistor relaxation oscillator will:
   (a) Not be reliable and may not start up
   (b) Not work correctly
   (c) Get very hot because the transistors are continuously in saturation
   (d) Produce a perfect, 50% duty cycle output
4. The practice of putting ring oscillators in leftover gates:
   (a) Helps minimize the cost of a digital electronics application
   (b) Helps synchronize other gates in the chip

(c) Should only be done if there is an odd number of inverting gates left in the chip
(d) Should be strenuously discouraged

5. A relaxation oscillator has an R1 value of 10 k, C of 0.1 μF and R2 equal to 1 k. What frequency will it oscillate at?
(a) It won't oscillate
(b) 4.54 kHz
(c) 4.54 MHz
(d) 4.54 Hz

6. What type of application is the relaxation oscillator best suited for?
(a) High power
(b) High cost, not requiring accurate operation
(c) Low cost, high accuracy
(d) Low cost, not requiring accurate operation

7. Canned oscillators are used because:
(a) They are more accurate than discrete component solutions
(b) They are cheaper than discrete oscillator solutions
(c) They are simple and accurate
(d) They are more reliable than discrete component solutions

8. The only type of oscillator that the 4060 cannot implement is:
(a) Ring oscillator
(b) Relaxation oscillator
(c) Crystal/ceramic resonator oscillator
(d) NPN transistor relaxation oscillator

9. Increasing the value of a resistor or capacitor in a 555 astable oscillator will:
(a) Lower its operating frequency
(b) Raise its operating frequency
(c) Increase the output voltage
(d) Lower the output voltage

10. Connecting six 74AS inverters (gate delay 2 ns) end to end will produce a delay circuit that is:
(a) 24 ns
(b) 12 ns
(c) 72 ns
(d) 48 ns

CHAPTER 9

Complex Sequential Circuits

As the saying goes: "You now know enough to be dangerous." You should be fairly comfortable with working with logic functions and equations, have an understanding of electronics and how to interface logic chips (of different families together), understand the basics of memory and have gone through a number of clocking schemes. I'm sure that you now feel you are ready to start bringing these pieces together into some interesting applications. I'm sure that you have some ideas of things you would like to have your hand at designing. Before being set free to wreak havoc on an unsuspecting world, I want to spend some time presenting you with some chips and tools that will make your plans for world domination much easier. In this chapter, I want to go through some of the subsystems that are available in chips that will make your design work easier.

In Chapter 7, I introduced you to the digital clock block diagram shown in Fig. 9-1. There shouldn't be any part of this diagram that is a surprise to you; the "time memory" consists of a number of flip flop registers that are reset

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

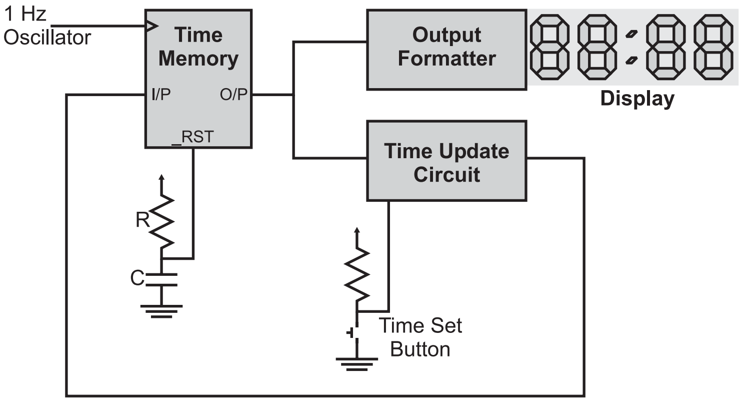

Fig. 9-1. Digital clock block diagram.

upon power up. The "time update circuit" consists of an adder along with logic and a user push button to determine what the *next* time will be. The "output formatter" consists of logic to decode the time memory and display it on a set of LEDs. Finally, the 1 Hz clock can be produced a number of different ways that are covered in the previous chapter.

While I must admit that it would be cool to see a digital clock designed using two input logic gates, I want to point out that there are a number of commonly available chips that provide major subsystems needed for such an endeavor. For the rest of this book I will be focusing on these chips and how they are interconnected to form "real world" applications.

Virtually all of these chips are sequential circuits in their own right; consider a "counter" chip that increments its internal memory devices each time a rising clock edge is received. The counter chip consists of several flip flop bits and combinatorial logic that processes input data, provides the value increment and outputs the data in a specific format. These functions are very similar to that provided by the digital clock in Fig. 9-1.

To show what I mean, consider the sequential circuit block diagram in Fig. 9-2. Superficially, the diagram has a very strong resemblance to Fig. 9-1 because many of the same basic functions and capabilities are required in both instances.

The "state memory" is the current operating state of the chip. The term "state" simply means at what operating point the chip is at. For a counter or other arithmetic function chip, the term "state" probably seems somewhat grandiose, but it is an accurate way of describing the current value in a counter. For a microprocessor, the term "state memory" is almost an understatement, as it includes not only the program counter (which points to the next instruction to execute) but also data and status register information.

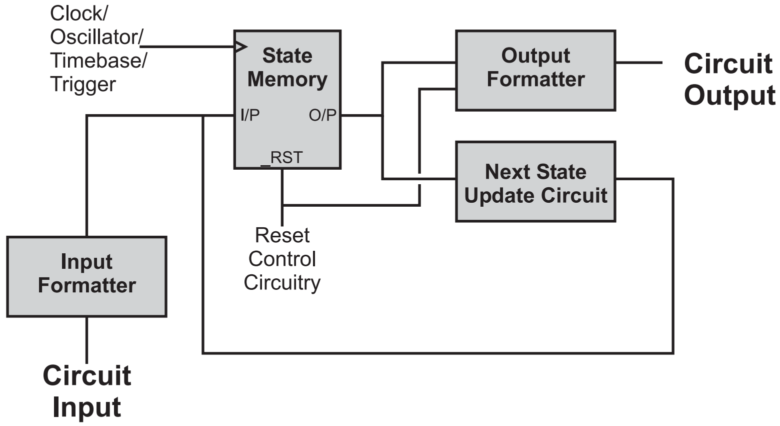

Fig. 9-2. General case sequential circuit.

The state memory can be reset (as shown in the previous chapter) and, more importantly, it is usually the only point in which the clock/oscillator input is received. The philosophy behind most sequential circuits is that the combinatorial logic processing input, output and the next state information, regardless of the circumstances, will be available in time for the next active clock cycle. The term normally used to describe this methodology of design is ‘‘synchronous’’ because a central clock is keeping track of the operations within the chip.

This philosophy is currently under challenge from scientists interested in investigating ‘‘asynchronous’’ digital logic design. This effort involves designing sequential circuits that are not ‘‘paced’’ by a central clock, but the length of time of each operation. For example, moving data from one register to another should take much less time than an instruction which stores data in the main memory. Asynchronous digital logic design holds the promise of faster computers that use much less power because the only active circuitry are the required gates and flip flops of the current time – nothing else needs to be active, nor do other circuits need to be clocked.

The ‘‘next state update circuit’’ and ‘‘input formatter’’ blocks process the current bit data and any relevant input for storage in the state memory. In Fig. 9-1, I combined both of these functions into the ‘‘time update circuit’’ because the only input required for this clock is whether or not the ‘‘time set button’’ is pressed – if it is, then the time update circuit will increment the hours, minutes and seconds stored in the digital clock’s ‘‘time memory’’.

The ‘‘input formatter’’ circuitry can be processing different inputs controlling what the next state is going to be – this is why I link it to the ‘‘next state update circuit’’. For a counter, this information could be the direction the counter executes in or whether or not the counter counts in binary or BCD.

The output formatter converts data into the required output and provides appropriate drivers for the function. Note that I have drawn a link to this box from the ''reset control circuitry'' despite my statement earlier that only the state memory could be reset. The reason for drawing in this link is to indicate that the chip may have tri-state drivers and these are held in a high impedance (''off'') condition while reset is active.

As you work through the material in the rest of the book, try to see how the described chip functions fit in with this model. You might have a better model to work with that makes more sense to you and if this is the case use it. The model that's presented here allows *me* to visualize what is happening in an application, and it would be arrogant of me to assume that it works for everyone else.

Counters

One of the most useful functions that you will use when you develop digital electronic circuits is the counter. The counter is actually a smaller piece of many complex chips, as it provides a basic way of maintaining the current operating state along with a method of progressing to the next one. The basic counter circuit consists of a set of flip flops that drive into and are driven from an adder. A counter circuit is shown in Fig. 9-3.

The use of ''edge triggered'' flip flops is a very important aspect of the circuit shown in Fig. 9-3 and one that you should keep in mind. When the ''counter clock'' changes state, the output value of the adder (which is the D flip flop value plus 1) is presented to the inputs of the D flip flop register bits as the next value to be saved.

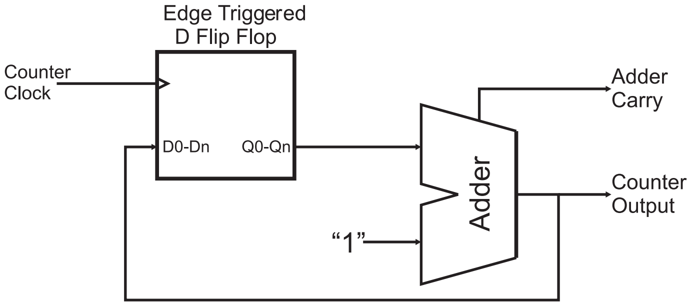

Fig. 9-3. Basic counter circuit.

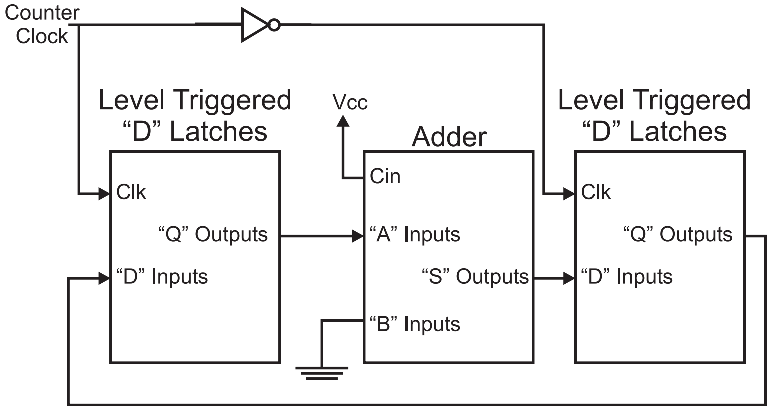

Fig. 9-4. Counter circuit built from level triggered flip flops.

If an edge triggered flip flop register wasn't used in the circuit, then you would have to use latches and design the counter something like the one shown in Fig. 9-4. In this circuit, I have put in two latches, each one "out of phase" with each other. This is to say that when the clock is high, one latch is storing the data while the other is passing through the value presented at its inputs. When the clock changes value, the latches change from passing data to storing and vice versa. This method of implementing a counter is unnecessarily complex and potentially very slow – the extra set of flip flops will slow down the performance of the counter and limit its maximum speed.

The counter circuit of Fig. 9-3 can be built using a 74C174 hex D flip flop and a 74C283 four bit adder circuit. The circuit shown in Fig. 9-5 will demonstrate how the counter works. When the term "floating" is used with respect to pins, it means that the pins are left unconnected.

When you try out this circuit, the first thing that you will probably notice is that when you press the button, the LEDs will not "increment" by 1, but by 2, 3 or even 4. The reason for this is known as "switch bounce". Earlier in the book, I showed a two inverter circuit for eliminating switch bounce, and later I will discuss a number of other strategies for minimizing the problem. For now, if you wire a 0.1 μF tantalum capacitor as shown in Fig. 9-5, you should minimize this problem (although you will probably not eliminate it).

The counter circuit should work well for you. As with the previous projects, a single chip can be used where multiple required. The counter chip that I usually work with is the 74LS193 (Fig. 9-6) which combines a four bit D flip flop register and adder along with the ability to decrement the result. Later in the book, I will show how this chip can be used with others to "cascade" from a 4 bit counter to an 8 and 16 bit counter.

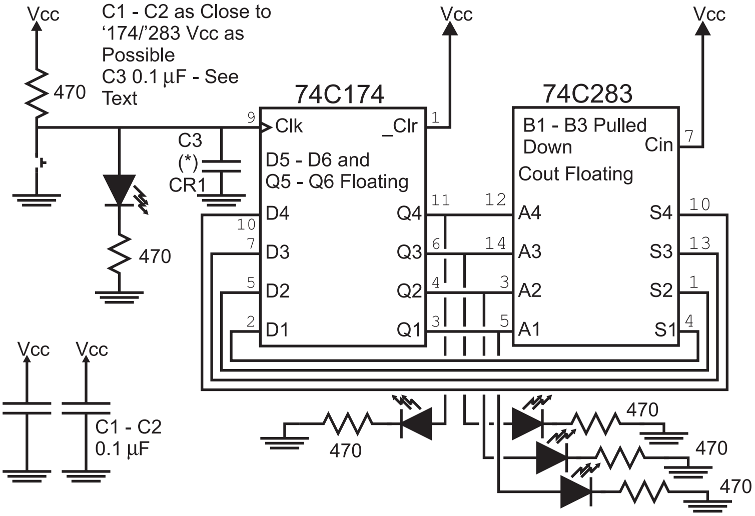
Vcc
C1 - C2 as Close to
'174/'283 Vcc as
Possible
C3 0.1 µF - See
Text
470
C3
(*)
CR1
470
74C174
9
Clk
_Clr
1
D5 - D6 and
Q5 - Q6 Floating
D4
Q4
10
D3
Q3
7
D2
Q2
5
D1
Q1
2
11
6
4
3
Vcc
74C283
B1 - B3 Pulled
Down
Cin
7
Cout Floating
12
A4
S4
10
14
A3
S3
13
3
A2
S2
1
5
A1
S1
4
Vcc
Vcc
Vcc
C1 - C2
0.1 µF
470
470
470
470

Fig. 9-5. Discrete component counter.

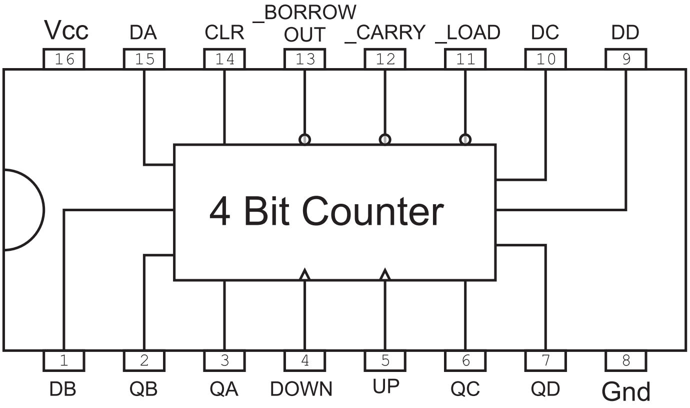
Vcc
DA
CLR
_BORROW
OUT
_CARRY
_LOAD
DC
DD
16
15
14
13
12
11
10
9
4 Bit Counter
1
2
3
4
5
6
7
8
DB
QB
QA
DOWN
UP
QC
QD
Gnd

_BORROW OUT/_CARRY - Active when Output 0000 or 1111
Counting According to Active
Clock ("DOWN" or "UP")

Fig. 9-6. 74193 counter chip pinout.

The "_Carry" out bit of the 74193 can be passed from one counter to the clock input of another to provide the ability to count more bits, as I show in Fig. 9-7. The carry bit can be thought of as an overflow to the more significant counter, indicating that it should increment its value. If by looking at this circuit you recognize it as being similar to the "ripple" adder presented earlier

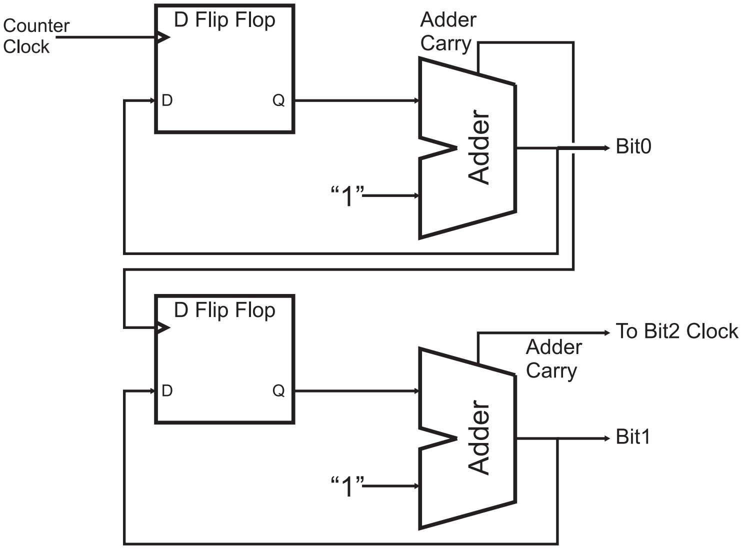

Fig. 9-7. Ripple counter block diagram.

in the book, go to the head of the class. This circuit is known as a ''ripple counter'' and does not have the same high speed as a counter built from look-ahead carry adders.

Along with the 74193, you might want to consider the 74161 counter, which can only count up and changes to the count value must be clocked in (the 74193 allows changes to the count value asynchronously, which is to say without the clock). The 74160 and 74192 chips are identical to the 74161 and 74193, respectively, but only count up to 9 and are known as ''decade'' counters. The 74160 and 74192 are useful in circuits in which the digits 0 through 9 are required for counting.

Shift Registers

Most intersystem (or intercomputer) communications are done serially. This means that a byte of data is sent over a single wire, one bit at a time, with the timing coordinated between the sender and the receiver. So far in this book, if you were to transfer a number of bits at the same time, you would send them in ''parallel'', one connection for each bit. The basis for serial communications is the ''shift register'', which converts a number of ''parallel'' bits into a time-dependent single string of bits and converts these strings of bits back

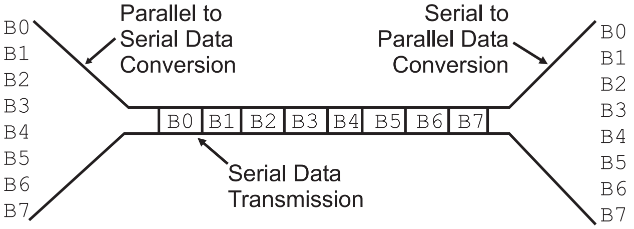
B0
B1
B2
B3
B4
B5
B6
B7
Parallel to Serial Data Conversion
Serial to Parallel Data Conversion
B0 B1 B2 B3 B4 B5 B6 B7
Serial Data Transmission
B0
B1
B2
B3
B4
B5
B6
B7

Fig. 9-8. Parallel to serial and back to parallel conversions.

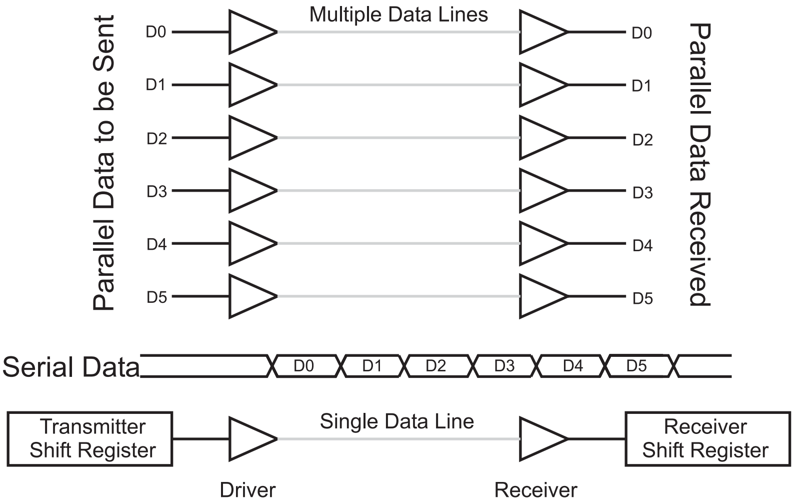
Parallel Data to be Sent
D0
D1
D2
D3
D4
D5
Multiple Data Lines
D0
D1
D2
D3
D4
D5
Parallel Data Received
Serial Data
D0 D1 D2 D3 D4 D5
Transmitter Shift Register
Single Data Line
Receiver Shift Register
Driver
Receiver

Fig. 9-9. Parallel vs. serial transmission hardware block diagrams.

into a set of parallel bits. Figure 9-8 shows this process with eight parallel data bits being converted into a bit stream and transmitted to a receiver, which "recreates" the eight bits back into their parallel data format.

The differences between serial and parallel data transfers are shown in Fig. 9-9. To send six bits in parallel, a half dozen transmitting "drivers" and an equal number of "receivers" are required. To send six bits serially, just a single driver and receiver is required, but the sending circuit must have a "shift register transmitter" and the receiving circuit must have a "shift register receiver". The parallel data can be sent in the time required for just one bit while the serial data requires enough time to send each of the six bits individually.

It probably looks like transmitting data serially requires a lot of overhead and it slows down the data transfer. There are a number of factors to consider before making this assumption. The first is that most chips are not made out of individual logic gates as the simple chips presented here so far; they are

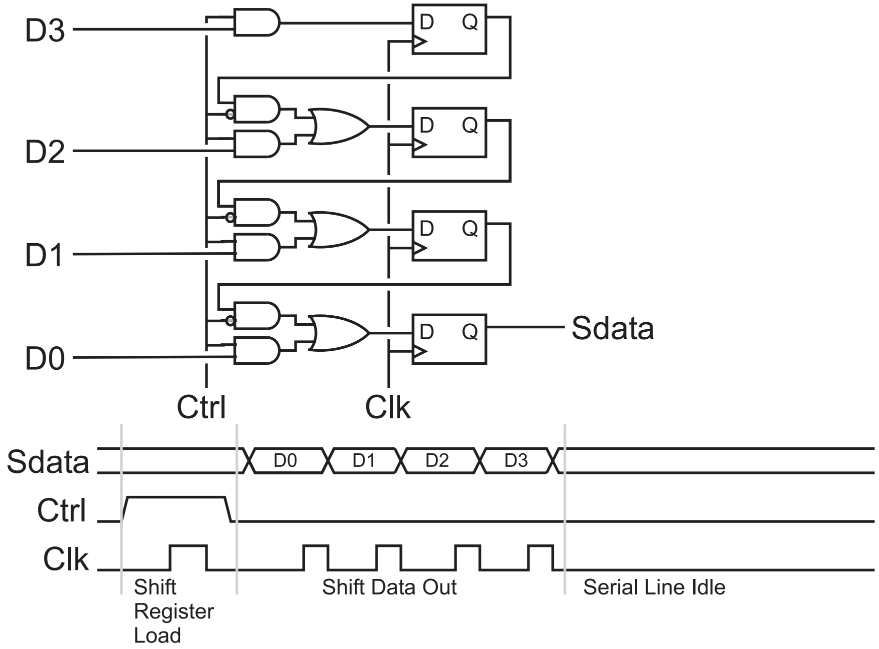

Fig. 9-10. Parallel to serial conversion hardware.

usually very dense circuits consisting of thousands of gates, with the impact of adding serial shift registers being very minimal. Another issue to consider is that it can be very difficult to synchronize all the parallel bits to ''arrive'' at the receiver at the same time in high-speed circuits. Finally, multiple wires can take up a lot of space and be quite expensive; if chips or subsystems could have shift registers built into them, then it often makes sense (both practical and economic) that data be transferred serially.

The circuit that converts the parallel data into the serial stream is quite simple. Figure 9-10 shows a circuit along with a waveform showing how the circuit works. Four bits are first loaded in parallel into a series of four flip flops. These four flip flops can be driven with data either from an external source or from the next significant bit depending on the ''Ctrl'' bit state. If ''Ctrl'' is high, when the ''Clk'' (''clock'') is cycled, the data in the D3:0 bits are stored in the four flip flops. If ''Ctrl'' is low, when ''Clk'' is cycled, each bit is updated with its next significant bit and data is shifted out, least significant bit first.

The process of each bit of data ''passing'' through each of the flip flops is known as ''shifting''. As can be seen in Fig. 9-10 that four data bits are ''shifted'' out on the ''Sdata'' line in ascending order, with the ''Clk'' line specifying when a new bit is to be shifted out. If this method was used to

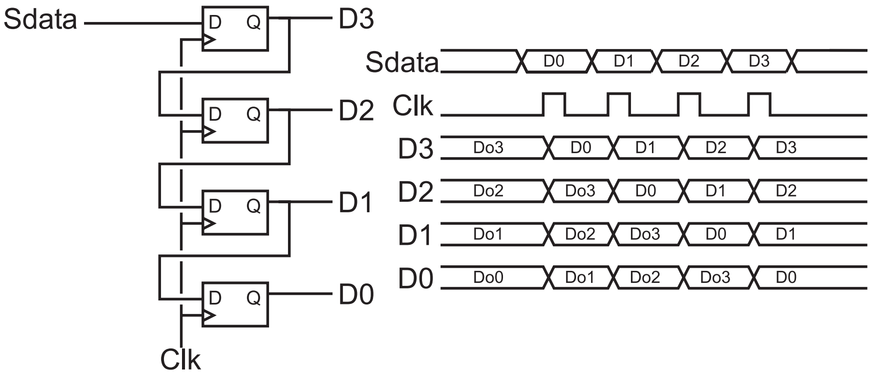

Fig. 9-11. Serial to parallel conversion hardware.

transmit data between two digital devices, it would be known as "synchronous serial data transmission".

Receiving "Sdata" is accomplished by simply using four flip flops wired with their outputs wired to the next input, as I've shown in Fig. 9-11. The same clock that is used to shift out the data from the transmitter should be used to shift in the data in the receiver. Along with the circuit used to shift in the data, I have included a waveform diagram for you to take a look at in Fig. 9-11. One potentially confusing aspect of the waveforms is my use of the "DoX" convention to indicate the previous values within the receiver. These bits will be shifted out in a similar manner as to how the data was shifted in.

There are a number of very common synchronous data protocols that are used in computer systems to provide simple interfaces to common peripherals. These interfaces, which include "Microwire", "SPI" and "I2C", are very easy and relatively fast ways of adding peripherals such as analog to digital converters and external memory to microcontrollers and complete computer systems. In fact, your PC has an I2C processor peripheral bus for controlling power supplies and monitoring the processor's chip temperature.

Linear Feedback Shift Registers

One of the most interesting logic devices you can work with is the "linear feedback shift register" ("LFSR"). It is built from a shift register along with two or more XOR gates modifying the contents of the register as shown in Fig. 9-12. This circuit can be used to "pseudo-randomize" data, encrypt and

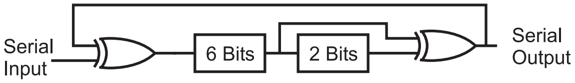

Fig. 9-12. Basic linear feedback shift register (LFSR).

decrypt serial data and provide very good serial data integrity checking. You may have heard the term ''cyclical redundancy check (''CRC'') when applied to data transmission; this is a type of linear feedback shift register. Linear feedback shift registers can also be implemented fairly easily in software with a microcontroller or microprocessor, although it is in hardware where the device is the most efficient.

The simple LFSR illustrated in Fig. 9-12 feeds back bits 5 and 7 of the shift register through XOR gates to the input. This changes the bit values in the shift register according to the formula:

$$\texttt{Bit}_0 = \texttt{Bit}_{\texttt{in}}\ \texttt{XOR}\ (\texttt{Bit}_5\texttt{XOR}\ \texttt{Bit}_7)$$

The LFSR is typically used for three purposes:

1. Creating a ''checksum'' value known as a cyclical redundancy check (CRC), which is a unique value or ''signature'' for a string of bits. Both the transmitter and receiver will pass the data through LFSRs and, at the end of the process, the CRC produced by the transmitter will be compared to the CRC produced by the receiver. If there is a difference in the CRCs, then the receiver will request that the transmitter resend the data.
2. Encrypting a string of bits. LFSRs can be used as an encryption/decryption tool with part of the encryption being the initial value in the LFSR. The value output from the LFSR is dependent on the initial value loaded into the LFSR. Decrypting data is also accomplished by using an LFSR, but configured as the complementary function.
3. Producing ''pseudo-random'' numbers. One of the most challenging computer tasks that you will be given is to come up with a series of random numbers. Computers are designed to be ''deterministic'', which means that what they are doing at any given time can be calculated mathematically. This property is important for most applications (nobody wants a computer to boot differently each time or to have a word processing program that responds randomly to keystrokes), but it is a problem for many applications which rely on the pseudo-random numbers for animated displays or ''lifelike'' responses to user input.

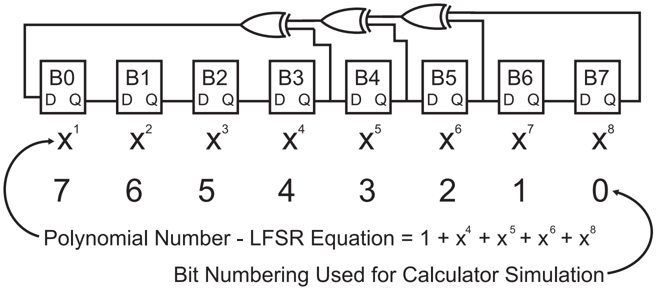
B0
B1
B2
B3
B4
B5
B6
B7
D Q
x1
x2
x3
x4
x5
x6
x7
x8
7
6
5
4
3
2
1
0
Polynomial Number - LFSR Equation = 1 + x4 + x5 + x6 + x8
Bit Numbering Used for Calculator Simulation

Fig. 9-13. Eight bit LFSR with defining polynomial expression.

In all of these applications, the LFSR is an ideal choice as a solution because it can be built very simply from just a few gates (meaning low cost and fast operation). The LFSR can also be implemented in software, as I will show below.

If you were going to express this LFSR to somebody else, you could send a graphic something like Fig. 9-13, or you could express it in the "polynomial" format like:

$$f(x) = 1 + x^4 + x^5 + x^6 + x^8$$

The polynomial format is the traditional way of expressing how an LFSR works and is used by mathematicians to evaluate an LFSR operation.

There are a few important facts about LFSRs that you should be aware of:

1. The LFSR can *never* have the value zero in it. If it contains zero, then none of the internal bits will ever become set.
2. The ideal LFSR implementation will be able to produce $2^n - 1$ different values. It should be obvious that the one value that cannot be produced is zero.
3. A poorly specified LFSR may have the situation where it ends out with a value of zero.

The operation of a single shift of the 8 bit LFSR in Fig. 9-13 can be modeled using the "C" function:

```
int SingleLFSRShift (intx)

{                                   //Shift the "CurByte" Value

//  with the polynomial 1 + x**4 + x**5 + x**6 + x**8

int LowBit;                         //XOR'd Low Bit
LowBit = x >> 7;                    //“x**8″ Term
LowBit = LowBit ^ (x >> 5);         //“x**6″ Term
LowBit = LowBit ^ (x >> 4);         //“x**5″ Term
LowBit = LowBit ^ ((x >> 3)&1);     //“x**4″ Term
return((x << 1) + LowBit) &0x0FF;   Return shifted 8 Bit value
} //  End SingleLFSRShift
```

Hardware State Machines

The hardware state machine circuit was originally designed to allow designers to create a complex application using a simple, single ROM, a few register bits and some basic logic gates instead of a complex sequential circuit design or a processor-based solution. State machines are not widely used in modern applications because the costs of the parts needed to make up the circuit can very easily exceed that of a microcontroller. Almost ironically, hardware state machines are used as the control mechanism for most modern computer systems because they are fairly easy to design, program and debug. The use of hardware state machines (which are typically referred to as just “state machines”) as the control mechanism for computer processors has given a new importance to the understanding of state machines.

The typical “state machine” is shown in Fig. 9-14. This circuit consists of an ROM (usually EPROM) which has part of its output data fed back as a “state address”. Other address lines are used as circuit inputs and the state machine changes its state address based on these inputs.

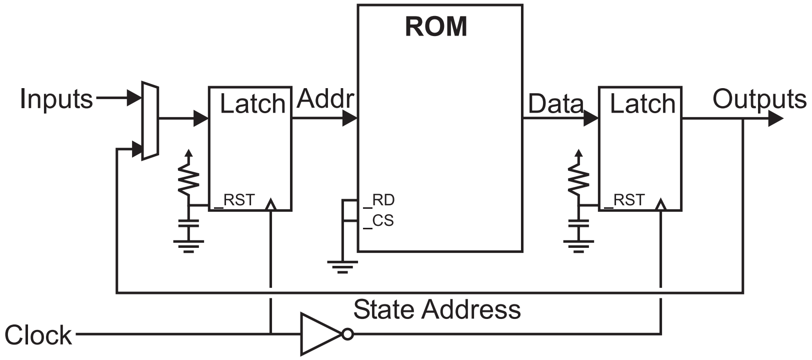

Fig. 9-14. General case hardware state machine.

The clock is used to pass the new address to the ROM and then pass the output from the ROM to the output and input state circuits. The two latches are operated 180° out of phase to prevent ‘‘glitches’’ from the ROM changing state from invalidly affecting any output circuits. A single edge triggered register is not typically used with the state machine because toggling inputs while the ROM is being accessed could result in invalid data being passed into the latches.

As few output bits are used as the ‘‘state address’’ as possible. The reason for this is to maximize the number of outputs and minimize the number of states which have to be programmed. Each state requires two to the number of inputs to function. Each state responds differently according to the inputs it receives.

A typical application for state machines is a traffic light. If a press-button crossing light, as shown in Fig. 9-15, is considered, a state machine circuit, like that shown in Fig. 9-16 could be used.

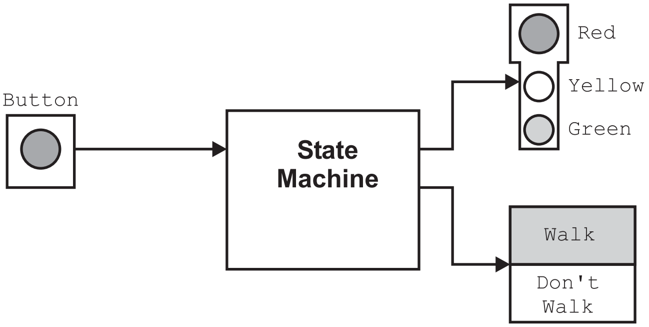
Button
State Machine
Red
Yellow
Green
Walk
Don't Walk

Fig. 9-15. Traffic light state machine block diagram.

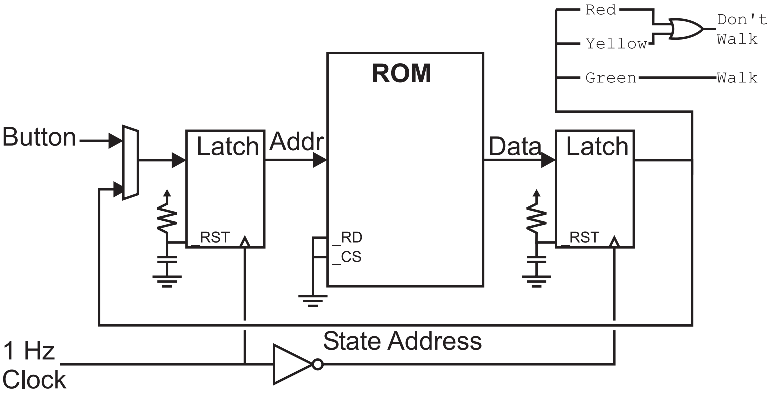
Red
Yellow
Green
Don't Walk
Walk
ROM
Button
Latch
Addr
Data
Latch
_RST
_RD
_CS
_RST
State Address
1 Hz Clock

Fig. 9-16. Traffic light state machine circuit.

In normal operation (which is known as "state 0"), the green light is on and the button is not pressed. If the button is pressed, then execution jumps to state 1, which turns on the yellow light for 5 seconds (states 2, 3, 4 and 5), after which the red light is put on for 26 seconds (states 6–31). If the button is pressed during states 7–31, then execution jumps to state 6 to reset the timer.

Table 9-1 ROM programming for simple traffic light state machine.

State	Button	New state	Green	Yellow	Red	Comments
B'00000'	1	B'00000'	1	0	0	Power up
B'00000'	0	B'00001'	0	1	0	Power up/button press
B'00001'	x	B'00010'	0	1	0	Yellow LED on
B'00010'	x	B'00011'	0	1	0	Yellow LED on
B'00011'	x	B'00100'	0	1	0	Yellow LED on
B'00100'	x	B'00101'	0	1	0	Yellow LED on
B'00101'	x	B'00110'	0	1	0	Yellow LED on
B'00110'	x	B'00111'	0	1	0	Yellow LED on
B'00111'	x	B'01000'	0	0	1	Red LED on
B'01000'	1	B'01001'	0	0	1	Red LED on
B'01000'	0	B'00111'	0	0	1	Red LED on/reset Ctr
B'01001'	1	B'01010'	0	0	1	Red LED on
B'01001'	0	B'00111'	0	0	1	Red LED on/reset Ctr
⋮		⋮				⋮
B'11111'	1	B'00000'	0	0	1	Return to green
B'11111'	0	B'00111'	0	0	1	Red LED on/reset Ctr

To keep the circuit simple, I want to use an eight bit data bus ROM with six inputs (five state, one button). This means that 2**6 (or 64) states are required in the ROM. These states are listed in Table 9-1. The reset on the input address latch is used to reset the state to 0 on the power up. The button is assumed to be "pressed" if a "0" is returned.

Table 9-1 would then be converted into bits and burned into the ROM. An "x" means both input states have the same result on outputs.

This application is reasonable to code and build, but a problem arises with very complex state machines (ones that require tens of inputs and hundreds of different states). These state machines are normally hard coded into a custom chip rather than built out of discrete parts like I have shown for this application. The reason for placing it within a chip is to give more outputs as well as more states in a custom application. The depth and the width of the data in "real" applications is better suited to custom chips which can have non-custom memories added much more easily than in the situation where only commercial chips are used.

In the example above, I have used a state machine with a one second clock. Obviously in this situation there can be problems (such as the missed input if the button is pressed for less than 1 second and it isn't released after it is pressed). This function makes state machines unattractive for rapidly changing inputs and any kind of sophisticated real-time processing of inputs is simply not economical to do with the state machine. When I say "not economical", I am thinking in terms of the memory and properly programming the many states.

Quiz

1. In the sequential circuit block diagram where is the clock signal passed to?
   (a) To the "state memory" and "output formatter" blocks
   (b) To just the "state memory" block
   (c) To the "input formatter" block
   (d) To the "reset control circuitry" block
2. Asynchronous digital logic design is being pursued because:
   (a) It will result in simpler chip designs
   (b) Circuitry designed under this philosophy will be easier to interface to

(c) The end of performance gains using traditional design methodologies is in sight
(d) It offers faster operations with less power usage

3. Why are edge triggered registers used for counters instead of latches?
(a) It will result in simpler circuit designs
(b) Circuitry designed under this philosophy will be easier to interface to
(c) Less power is required
(d) It offers faster operations with less power usage

4. Ripple counters are:
(a) Always the fastest way to implement counters
(b) Usually more complex electronically than other counter designs
(c) Always the slowest way to implement counters
(d) Similar to ripple adders in operation

5. What are advantages of serial data transmission over parallel data transmission?
(a) Reduced number of drivers and receivers
(b) Faster data transmission
(c) Lower product costs
(d) Higher product quality

6. Where is serial data transmission not used?
(a) The internet
(b) Broadcasting stations to TVs/radios
(c) Keyboard to PC interface
(d) PCI bus interfaces

7. Linear feedback shift registers are built from:
(a) The system architectural drawings
(b) The high-speed circuits to support communications
(c) Shift registers and XOR gates
(d) The basic system serial interface

8. When the value of a linear feedback shift register equals zero:
(a) The operation has completed
(b) Either the initial and input values are zero or there is a problem with the LFSR design
(c) There was an error in encrypting a message
(d) Power has been removed from the circuit

9. Hardware state machines are rarely used except in:
    - (a) Computer processors
    - (b) Military and space applications
    - (c) High-performance custom logic applications
    - (d) Situations where old ROMs are easily available

10. State machines are normally built:
    - (a) Out of discrete chips and ROM chips
    - (b) In complex custom chips
    - (c) On specially designed carrier PCBs
    - (d) With the checksum of the ROM printed on them

CHAPTER 10

Circuit Interfaces

For a sequential digital electronic circuit to be effective, it has to interface with something. This something could be a person or it could be other digital electronic circuits. If you were to look at different interfaces for either case (human or machine), you will discover that as the function of the circuit increases in sophistication, so does the interface. The reasons for this increase in interface complexity can be attributed to an increased amount of data to present as well as an increased number of operating parameters to choose from and select. The challenge is to come up with a way of adding these user and device interfaces simply, effectively and not affect the operation of the central sequential circuit.

Simple logic level switches and individual LEDs for each bit are perfect examples of the types of interfaces that I am talking about; to add these devices to your application, you generally don't require any types of busses nor do you need to have any special communications protocols for communicating with the devices. These interfaces are simple to add and modify to an application.

The problem with simple logic level switches and LEDs connected to each bit is that they cannot be very descriptive; nor are they very efficient methods of transferring data. An eight bit system is quite manageable, but it becomes

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

very difficult when there are tens, hundreds or even thousands of bits to control and monitor. Early computers started out using simple switches and lights for input and output, respectively, but quickly outgrew them and began using printers, teletypes and punch cards to get state information from the computer. Today's computer systems have very sophisticated input and output capabilities, requiring the power of a processor that would have been identified as a "supercomputer" 10 years ago or less.

An example of a complex interface that you would be hard pressed (if it were possible at all) to create digital electronics for is the Hitachi 44780 based LCD module (Fig. 10-1). The controller hardware is fairly complex and must be accurately timed. The LCD module works like a "teletype" or a single line TV display – as you write characters to it, a "cursor" will move to the right, to prepare for the next character. The character interface consists of the eight data bits and three I/O pins listed in Table 10-1.

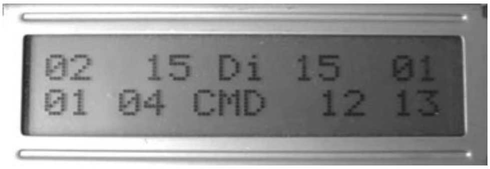

Fig. 10-1. Sample LCD output.

Table 10-1 Hitachi 44780 pin interface.

Pins	Description/function
1	Ground
2	Vcc
3	Contrast voltage
4	"RS" – _instruction/register select
5	"RW" – _write/read select
6	"E" clock
7–14	Data I/O pins

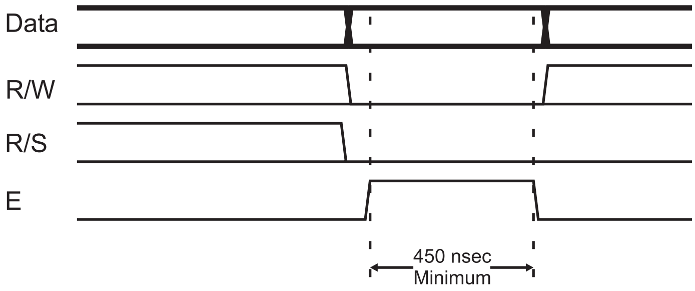

Fig. 10-2. LCD write waveform.

I typically attach a series of pins to the 14 connector pins so that the LCD can be easily mounted on a breadboard. In some LCDs, you may discover that there are 16 connector holes with the extra two holes used for backlighting. Some other LCD modules have two rows of seven or eight pins. For the ease of creating the experiments in this book and wiring them to the breadboard you should just use LCD modules that have a single row of pins.

Wiring the LCD to the hardware is quite straightforward as you will see in the waveform diagram (Fig. 10-2). The only unexpected aspect of the interface circuit is a potentiometer used to set the ‘‘contrast voltage’’ used by the LCD. The potentiometer is wired as a voltage divider, with the contrast voltage pin connected to the wiper of the potentiometer. Depending on the type of LCD that you are using, you will find that the voltage producing the best contrast will either be high or low, depending on the technology used in the LCD.

To communicate with the LCD, you will have to send the data words listed in Table 10-2 via the LCD interface. These bytes are commands that set the operating mode of the LCD or command it to perform some other operation. In Table 10-2, I have listed the different commands, along with the ‘‘RS’’ and ‘‘RW’’ lines that are used to control them. To clock in the command, the ‘‘E’’ bit must have a high value (‘‘1’’) written to it and then a low value (‘‘0’’).

Data displayed on the LCD is, for the most part, ASCII and you can pass ASCII characters directly from the hardware to the LCD. I say that the LCD can display ASCII ‘‘for the most part’’ because you will find that some characters are not supported (such as the backslash, ‘‘\’’) and if you go outside the normal ASCII character limits, you will see Japanese characters on the display. If you were to send a carriage return, line feed or any of the other ASCII terminal command characters, you would discover that they

Table 10-2 Hitachi 44780 based LCD command set.

RS	RW	D7	D6	D5	D4	D3	D2	D1	D0	Instruction/description
4	5	14	13	12	11	10	9	8	7	LCD I/O Pins
0	0	0	0	0	0	0	0	0	1	Clear Display (Takes up to 5 ms)
0	0	0	0	0	0	0	0	1	*	Move Cursor to "Home" (5 ms)
0	0	0	0	0	0	0	1	ID	S	ID = 1, Increment Cursor after Write S = 1, Shift Display after Write
0	0	0	0	0	0	1	D	C	B	D = 1, Turn on Display C = 1, Cursor On B = 1, Cursor Blink
0	0	0	0	0	1	SC	RL	*	*	SC = 1, Shift Display after Write RL = 1, Shift Display to Right
0	0	0	0	1	DL	N	F	*	*	Reset the 44780 Interface Length DL = 1,8 Bits/ DL = 0,4 Bits N = 1, Two Display Lines F = 1, 5x10 Font (Normally 0)
0	0	0	1	A	A	A	A	A	A	Move Cursor to Graphic RAM Address B'AAAAAA'
0	0	1	A	A	A	A	A	A	A	Move Cursor to LCD Position B'AAAAAAA'
0	1	BF	*	*	*	*	*	*	*	Poll LCD "Busy Flag" (Active "1")
1	0	D	D	D	D	D	D	D	D	Write Data to the LCD
1	1	D	D	D	D	D	D	D	D	Read Data from LCD at Current Cursor

*, "Don't Care".

result in a strange character being displayed. If you want to provide more "terminal"-like functions to the 44780 based LCD you will have to write them yourself and add them to your application.

Most commands execute in 160 µs or less with the display clear and move cursor to home commands can take up to 5 ms. The initialization process for the LCD is:

1. Wait more than 15 ms after power is applied.
2. Write 0x030 to LCD and wait 5 ms for the instruction to complete.
3. Write 0x030 to LCD and wait 160 µs for instruction to complete.
4. Write 0x030 AGAIN to LCD and wait 160 µs or Poll the Busy Flag.
5. Set the Operating Characteristics of the LCD:

- Write "Set Interface Length"
- Write 0x010 to disable display shifting
- Write 0x001 to clear the display
- Write "Set Cursor Move Direction" setting cursor behavior bits
- Write "Enable Display/Cursor" & "Enable display and optional cursor".

The LCD could be controlled by a state machine, but there would be a significant amount of work to do this (and the state machine would be quite large). Along with the eight bit interface, the LCD can also be controlled by a four bit interface; each character and eight bit instruction is passed in four bit blocks through the D7:4 pins, but this interface would probably be even more difficult to create for the LCD module.

Address and Data Decoders

When you have decided upon the interfaces to your application, you will probably have to determine the best method of selecting which device is active at any time. The method that would make the most sense is to use the same method that a microprocessor uses: output a bit value, selecting the device and one control bit to activate the interface device. Depending on the resources available, the section bits may consist of a number of bits, each one passed to a different interface device, or a binary value, which is decoded into a specific control bit.

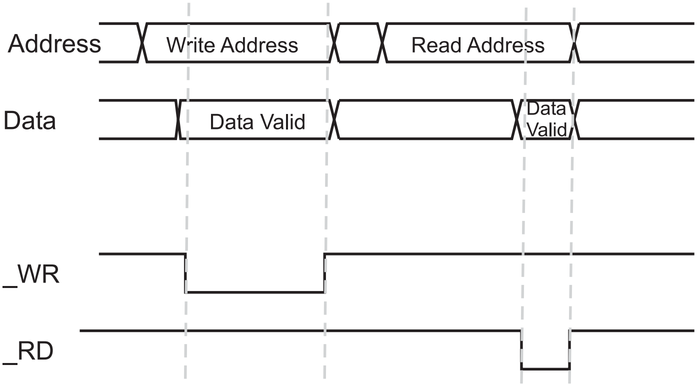

Fig. 10-3. Memory bus reads and writes.

Ideally, the signal being passed to the interface device would look something like Fig. 10-3 – an ‘‘Address’’ value is passed to the device and after a data ‘‘set up’’ time, a Read (‘‘_RD’’) or Write (‘‘_WR’’) line becomes active. A ‘‘Read’’ action polls the interface device and returns the value to the sequential circuit. A ‘‘Write’’ action does the exact opposite: it sends a value from the sequential circuit to the interface device.

You should notice that the timing of the read and write operations are quite a bit different. The short ‘‘Read’’ pulse is indicative of the expected operation of the device being accessed; once it receives the ‘‘Read Address’’ which selects the device, it takes some time to prepare the data before it can be read out. Similarly, when writing data, the _WR line is active for a surprisingly long period of time to allow the interface device to pass the data internally and prepare the interface circuitry to correctly store the data.

The interface read and write operations are good examples of situations where the latches rather than registers are used. When the _RD and _WR signals become active, data should be passed through them as quickly as possible rather than being held on a rising or falling value of the signal edge. For most applications, this need for taking advantage of every possible picosecond of time for data transfer is not needed, but you will find that it’s a good idea to work with a standard design interface that will work in all situations.

Following this philosophy, rather than providing an individual bit to each interface device, how about a binary ‘‘address’’ that can be decoded to an individual address using a ‘‘decoder’’ like the 74139 (Fig. 10-4) that converts a two bit value into four individual active low outputs. The 74139 contains

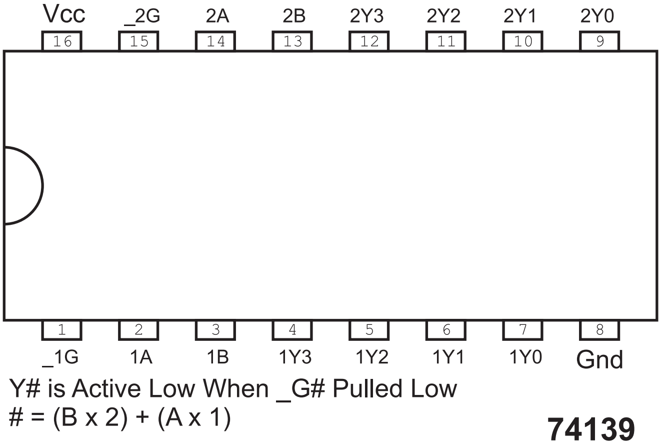
Vcc
_2G
2A
2B
2Y3
2Y2
2Y1
2Y0
16
15
14
13
12
11
10
9
1
2
3
4
5
6
7
8
_1G
1A
1B
1Y3
1Y2
1Y1
1Y0
Gnd
Y# is Active Low When _G# Pulled Low
# = (B x 2) + (A x 1)
74139

Fig. 10-4. 74139 dual 2 to 4 decoder chip pinout.

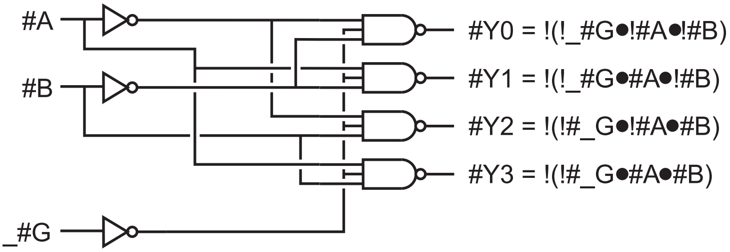
#A
#B
_#G
#Y0 = !(!_#G●!#A●!#B)
#Y1 = !(!_#G●#A●!#B)
#Y2 = !(!#_G●!#A●#B)
#Y3 = !(!#_G●#A●#B)

Fig. 10-5. 2 to 4 decoder logic diagram.

two “two to four” decoders like the ones shown in Fig. 10-5. Decoders (also known as “demultiplexors”) convert binary values into individual output lines and are primarily used to decode memory addresses to individual chips.

The resulting “interface” is essentially a computer “bus” and will allow you to add standard computer interfaces to your circuit, probably with no modification. You might have been wondering why I went through the effort of providing an interface select function along with the _RD and _WR signals – you may be thinking that a single pin that both selects the device as well as initiate the read or write would be enough; the philosophy of interfacing like in a computer is only so useful. This is true if you are going to only work with custom-designed interfaces, but there are a lot of standard interfaces (including the various ones presented in this chapter) designed for being accessed by computer systems that you will want to take advantage of.

Multi-Segment LEDs

I think you would be hard pressed to find somebody in the industrialized world that has never seen a ‘‘seven-segment’’ LED display (Fig. 10-6) before. It first became popular in the 1970s and is used in almost literally *everything* from digital clocks to car instruments. Seven-segment LEDs can be found virtually everywhere, being used not only in digital watches but also in kitchen appliances, cars, instruments and, of course, in videocassette recorders (VCRs). The flashing ‘‘12:00’’ on a clock or VCR created using seven-segment LEDs is the symbol of a person’s inability to handle the latest in technology.

In Fig. 10-6, I have shown the appearance of the seven-segment LED display – it can be put in the same ‘‘footprint’’ as a 0.300 inch’’ wide 14 pin DIP package, but some of the pins (‘‘N/C’’ for ‘‘no connect’’) are not present. The ‘‘DP’’ LED stands for the ‘‘decimal point’’.

The seven-segment LED display can be wired as either a ‘‘common anode’’ or ‘‘common ‘‘cathode’’; in this experiment we will be using ‘‘common anode’’, wired as shown in Fig. 10-7. For this part, the two ‘‘common’’ pins are connected to all (and occasionally some) of the anodes of the eight LEDs built into the display. This simplifies the wiring you will have to do somewhat and makes working with multiple displays a bit easier, as I will show in a later experiment.

Despite its commonality, the seven-segment LED display is not trivial to work with. There are a number of chips on the market that make the component easier to work with in some applications, but when you are working with your own sequential circuits, you will find that these ‘‘canned’’ functions never quite do what you hope for.

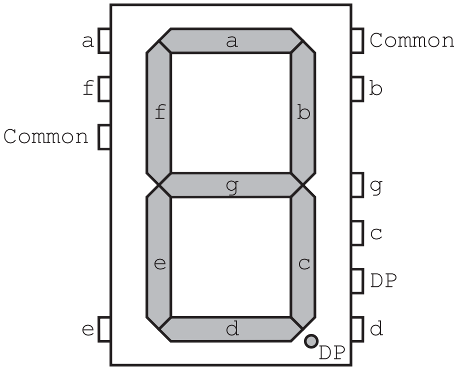

Fig. 10-6. Seven-segment LED display with pinout.

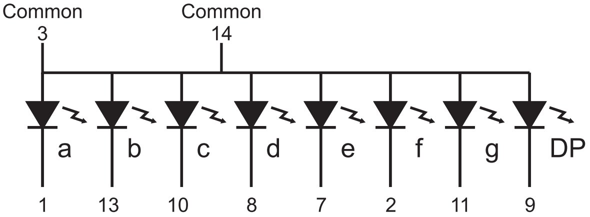

Fig. 10-7. Common anode seven-segment LED display internal wiring.

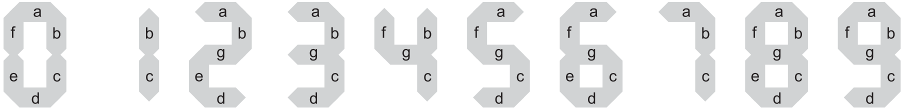

Fig. 10-8. Lit seven-segment LED display elements for different numbers.

As you are probably aware, by turning on each of the different LEDs with their unique values individually, you can create a multiple-digit display. Figure 10.8 shows how by turning on different LED segments, the display can be used to display the 10 numeric characters. Along with the 10 numbers, there are a number of letters that can be displayed, although only a few of them look exactly like the characters they are supposed to represent. If you want to display letters as well as numbers, then you will have to use a LED with more segments – these are available as either 16-segment displays or as matrixes of LEDs that display the character as a multi-dot "font" like on your computer screen.

Each LED in the display can be wired conventionally to control whether or not they are turned on or off. Controlling individual LEDs in a single display is quite easy. It gets quite a bit more difficult when you have to display different values on multiple LED displays. To convert incoming bits to meaningful characters on the display, you will have to pass the bit values through a combinatorial circuit, like the one shown in Table 10-3 for the first four decimal digits.

In Table 10-3 in the "Comments" column, you can see that I have noted any commonalities between different equations and noted that segment "B" is always active for all four digits. I should point out that coming up with the equations for each of the segments is good practice for working with Boolean arithmetic equations, but it is much easier and simpler to buy a seven-segment LED driver. The 7447 chip is commonly used for decoding the incoming bits and driving the LEDs and is an excellent solution when there is

Table 10-3 Seven-segment LED display combinatorial specification for the first four decimal characters.

Segment	"0"	"1"	"2"	"3"	Terms	Comments
A	1		1	1	$(!A0 \cdot !A1) + A1$	Same as "d"
B	1	1	1	1	1	Always on
C	1	1		1	$!A1 + (A0 \cdot A1)$	
D	1		1	1	$(!A0 \cdot\ !A1) + A1$	Same as "a"
E	1		1		$!A0$	
F	1				$(!A0 \cdot !A1)$	Uses AND from "a" & "d'
G			1	1	$A1$	

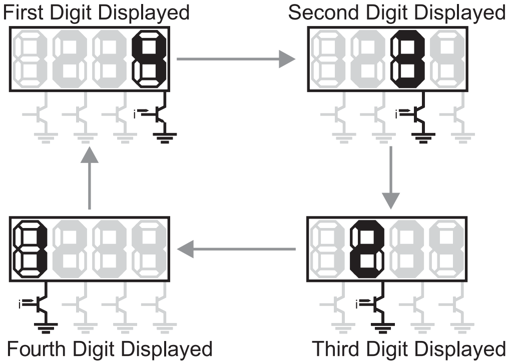

Fig. 10-9. Scanning through four seven-segment LED displays.

only one seven-segment LED display outputting information from the application.

When multiple LED displays are required, instead of providing multiple-digit drivers, a single-digit driver is used and different values are passed to it for different displays very quickly. Figure 10-9 shows a four-digit LED display with each digit having a different value. Each digit is turned on momentarily to display its value and then switched off for the next digit. The eye's visual persistence ignores the flickering if the sequencing is done fast

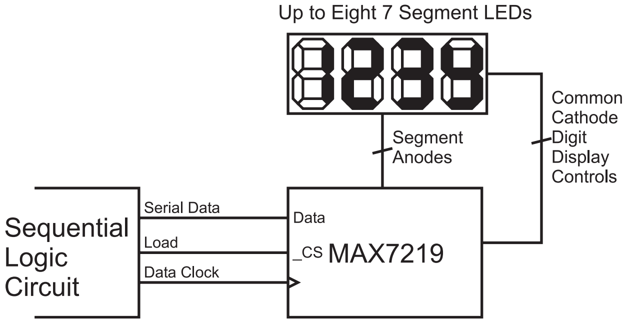

Fig. 10-10. Using a Maxim MAX7219 seven-segment LED controller chip.

enough and it appears that all the digits are on simultaneously, even though they are displaying different values.

As a rule of thumb, each display should be active 50 or more times per second. The slower each display is flashed on and off, the more likely the human eye will pick up the flashing. A flashing multi-character display is not attractive and could cause headaches in some people (especially if the displays are very bright). The time each display is turned on must be as equal as possible. If one display is on for a longer period of time than the others, then it will appear brighter and, conversely, a display active for a shorter period of time will appear dimmer. When working with multiple displays, in order to meet the 50 times per second guideline, you are actually going to have to loop through your individual display action 50 times per second multiplied by the number of displays. So, for a four-digit display, you will have to loop 200 times per second and each digit will be on for 5 ms at a time. There are some chips, such as the very popular Maxim MAX7219 (Fig. 10-10), which can control multiple seven-segment LED displays. This chip takes care of all the driving and timing requirements for the displays; the only catch is that you must shift in the desired value for the display.

Pulse Width Modulation

Despite showing how logic gates and other digital devices are built from simple analog components, they do not handle working with analog voltages very well. There are some circuits that will produce a valid analog (an arbitrary voltage, not just logic ‘‘high’’ and ‘‘low’’) voltage but they do not work very well if the circuit has to drive a high current device. Instead of varying the voltage level to provide varying levels of power, I produce a

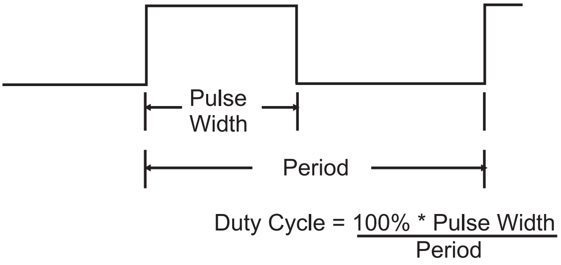
Pulse
Width
Period
Duty Cycle = 100% * Pulse Width / Period

Fig. 10-11. PWM waveform features.

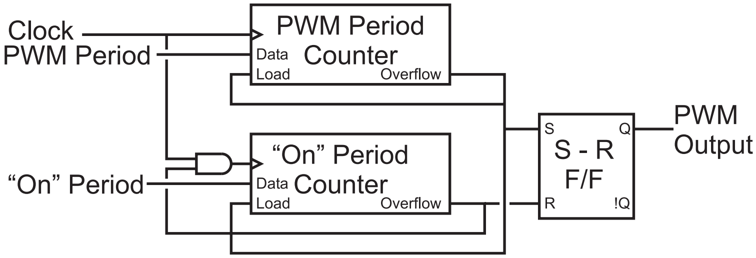
Clock
PWM Period
PWM Period Counter
Data
Load
Overflow
"On" Period
"On" Period Counter
Data
Load
Overflow
S
Q
S - R
F/F
R
!Q
PWM
Output

Fig. 10-12. PWM generator block diagram.

string of timed pulses known as a ‘‘pulse width modulated’’ (‘‘PWM’’) signal (Fig. 10-11). A PWM signal is a repeating signal that is ‘‘on’’ for a set period of time that is proportional to the voltage being output. I call the ‘‘on time’’ the ‘‘pulse width’’ in Fig. 10-11 and the ‘‘duty cycle’’ is the percentage of time the ‘‘on time’’ is relative to the PWM signal’s ‘‘period’’.

To output a PWM signal, there are several possible methods. One way is to use two counters that have a common clock. When one counter overflows, it resets itself and the second counter. Until the second counter overflows, the output of the circuit is set to ‘‘1’’. When the second counter overflows, the output of the circuit is reset until the first counter overflows and the process is repeated. Figure 10-12 shows how this type of circuit could be implemented.

This PWM generator circuit uses counters that are reloaded (from the ‘‘Data’’ pins) upon an ‘‘Overflow’’ positive pulse. The ‘‘PWM Period Counter’’ (the ‘‘first counter’’) runs continuously and when it overflows (reaches the final count), it resets and reloads the count value for not only itself but also for the second counter (the ‘‘‘On’’ Period Counter’).

When the PWM Period Counter resets, it ‘‘Sets’’ the S-R flip flop, driving the ‘‘PWM Output’’ high for the start of the PWM signal output. The ‘‘On’’ Period Counter is reset and reloaded by the PWM Period Counter and runs until it overflows. When the ‘‘On’’ Period Counter Overflows, the PWM

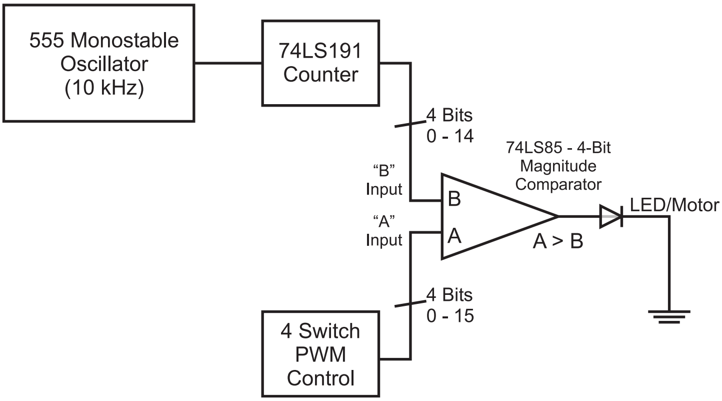

Fig. 10-13. Another PWM generator circuit design block diagram.

Output is halted and it also stops running until the PWM Period Counter reloads it, which resets the "Overflow" output and allows the Counter to drive the "On" Period Counter once more.

Another type of PWM generator is shown in Fig. 10-13. The counter output will be continuously compared against a bit value and when the bit value is greater than the counter value, a "1" will be output. The block diagram for the circuit that I envisioned is shown in Fig. 10-13 and can be built quite easily as I show in this section.

When you study Fig. 10-13, there will probably be one point that won't make sense to you: I show that the counter ranges from 0 to 14 and not 0 to 15, as you would expect for the typical four-bit counter. I wanted the counter to reset itself at 14 rather than 15 so that when the binary values were compared, a 100% duty cycle could be produced as well as a 0% duty cycle by outputting a "1" when the set value was greater than the counter value. If the counter ran from 0 to 15 then the circuit would not be able to produce a PWM with a 100% duty cycle.

To produce the bit range from 0 to 14, I used the 74 × 191 chip counting down and tying the "_LOAD" pin to the "_RIPPLE" pin and driving the inputs to 14. The "_R" ("Ripple" Output) pin becomes active when the chip is "rolling over" from one extreme to another and the "_LD" pin moves the value at the input pins into the counter's latches when it is active. Normally, when a four-bit counter is "rolling over" as it counts down it goes from 0 to 15, but by tying the "_R" pin to the "_LD" (negative active "Load") pin of the 74 × 191, you can load in a new value when the counter reaches 0 and is about to roll over. This feature is ideal for this application as it ensures the count stays in the range of 0 to 14.

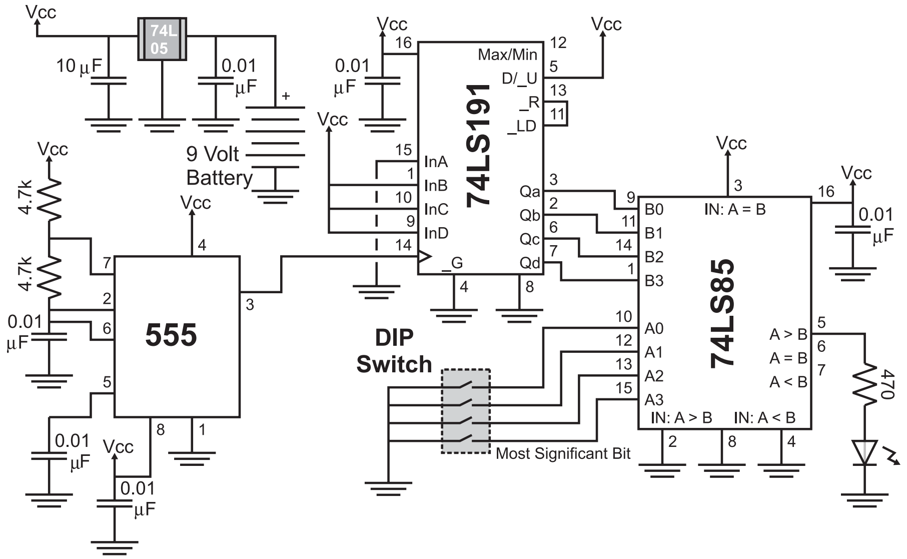
Vcc
74L05
10 µF
0.01 µF
+
9 Volt Battery
Vcc
4.7k
4.7k
0.01 µF
555
7
2
6
5
4
3
8
1
0.01 µF
Vcc
0.01 µF
Vcc
16
0.01 µF
Vcc
74LS191
Max/Min
D/_U
_R
_LD
12
5
13
11
Vcc
15
1
10
9
14
InA
InB
InC
InD
_G
4
8
Qa
Qb
Qc
Qd
3
2
6
7
DIP Switch
Most Significant Bit
Vcc
3
IN: A = B
16
0.01 µF
9
11
14
1
B0
B1
B2
B3
10
12
13
15
A0
A1
A2
A3
74LS85
A > B
A = B
A < B
5
6
7
IN: A > B
IN: A < B
2
8
4
470

Fig. 10-14. Sample PWM generator circuit schematic.

Converting the block diagram to a schematic one is very straightforward (Fig. 10-14) and wiring it onto the PCB's breadboard is tight but not really a challenge (Fig. 10-15). The PWM output value is specified by the four-position DIP switch. I placed 0.01 µF decoupling capacitors on all of the power inputs of each of the chips. These decoupling capacitors are very important when working with the standard (not CMOS) 555 because it can place large transients on the power line.

I used TTL chips (powered by 5 volts from the 78L05 regulator) rather than CMOS chips because I found that it is difficult locating 74C85 chips. An advantage of using TTL instead of CMOS for this circuit was that I could simply pull the comparator inputs to ground without the need of a pull up resistor. If you build this circuit with CMOS chips, make sure that you have 10 k pull up resistors on the DIP switch to ensure a high voltage is passed to the comparator.

Once you have built the circuit, you will find that the LED's brightness will be dependent on the value on the DIP switch. It will be confusing, as the value on the DIP switch will seem to be the opposite to the behavior of the PWM. When all the switches are "on", the LED will be off and, for what seems to be a "large" value, the LED will be dim. When all the switches are "off" the LED will be full on. This confusion is a result of the "on" marking indicating when the switches are closed, not when the signal is a "1" or "high" (which is often extrapolated to being "on") – when the switches

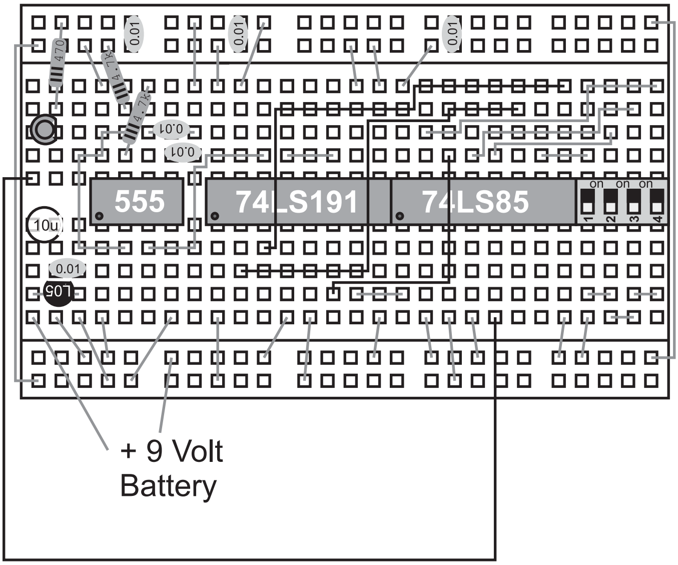

Fig. 10-15. Sample PWM generator circuit wiring diagram.

are closed (“on”), the comparator input is pulled to ground and has the value “0”.

For all PWM circuits (not just the two I’ve shown here), you must remember that the effective frequency is the input clock frequency divided by the counter value. For the example circuit shown here, the 10 kHz signal is divided by 15 (how many cycles the 74LS191 counts before resetting) so the resulting output signal frequency is 667 Hz, which is still faster than the human eye can perceive a flashing LED, but much lower than required for some DC motors. PWMs are commonly used to control the speed of electric motors and if the PWM frequency is within the audible range of human hearing, you will hear a definite “whine” from the motors. The solution to this problem is to either run the PWM at frequencies above human hearing (greater than 18 kHz) or below the range of human hearing (60 Hz or below).

The lower PWM frequencies should not be an issue to produce, but the higher ones can be a challenge, especially if more bits are used in the counter. For example, to create a 20 kHz PWM output signal, you will have to provide a 300 kHz clock for a 15-value PWM and 5.1 MHz for a 255-value PWM! You may find that to get a practical circuit, you will have to find a compromise between the number of bits used in the PWM for the signal level and the speed of the oscillator that is going to be used with it.

An interesting feature of a PWM is how it can save you power. If you were to run the PWM with a 75% duty cycle, what do you think the average power output would be? If you answered 75%, then you didn't go back in the book to look up the power formula. Power is defined by the formula:

$$P = V \times i$$

and substituting in values from Ohm's law, it can be also expressed as:

$$P = V^2/R = i^2 \times R$$

From these formulas, it should be obvious that if the voltage is high only three-quarters of the time, the power dissipated by the device being driven by the PWM is nine-sixteenths or 56% of the total power used by PWMs running with a 100% duty cycle. This means that, along with providing the ability to ''throttle'' direct current devices, a PWM can also result in significant power savings as well.

Finally, you might be confused that I gave you two quite different implementations of the PWM circuit; I did this to show you that there is almost always more than one solution to any problem. I normally recommend that new designers come up with three solutions to a problem before going ahead and implementing *something*. Having three solutions to choose from will allow you to compare features and drawbacks and choose the solution that is best for the application.

Button ''Debouncing''

I consider the issue of debouncing switches and buttons to be one of the most important and vexing problems that you will have to deal with when you are developing applications that work with operator input. Most people think that electrical connections happen instantaneously; you might be surprised to discover that the contacts within a switch actually bounce a few times before the switch makes a constant contact. This is shown in the oscilloscope picture in Fig. 10-16.

Earlier in the book, I showed you a simple method of debouncing a switch input by creating a small memory device from two inverters. A major drawback of this circuit is that it ''backdrives'' the outputs of one of the inverters, but this problem can be eliminated through the use of CMOS inverters and a 10 k current limiting resistor. Even with this fix in place, there is another problem to consider when deciding whether or not to use this circuit – finding double throw push buttons can be difficult. This circuit is well suited for double throw switches but, from the practical difficulty of finding double throw buttons, it becomes impractical.

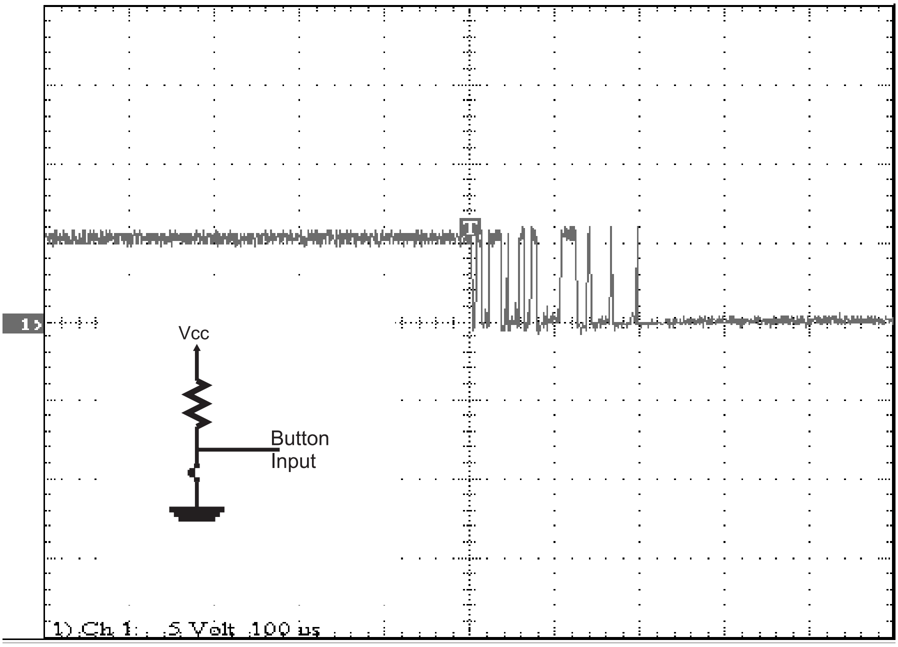

Fig. 10-16. Button bounce waveform.

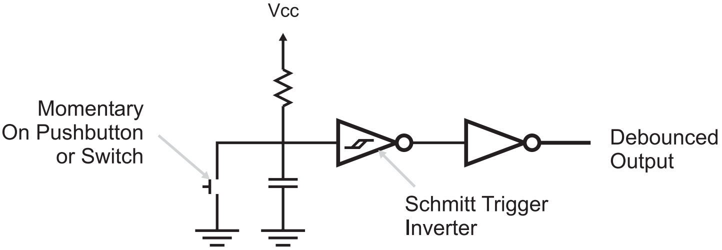
Vcc
Momentary
On Pushbutton
or Switch
Schmitt Trigger
Inverter
Debounced
Output

Fig. 10-17. Simple button debounce circuit.

The debounce circuit that I recommend you use is shown in Fig. 10-17. This circuit consists of a resistor–capacitor network that charges over a given amount of time or discharges quickly through a closed switch or button. Figure 10-18 shows the filtering of the bouncing; it is not perfect, but it is much better than what we started with.

The inverter with the funny symbol in Fig. 10-17 is called a “Schmitt Trigger Input Inverter” and provides an extra measure of filtering of the button input. Schmitt trigger inputs are designed to change state on the rising or falling edge of a signal with “hysteresis”, as shown in Fig. 10-19.

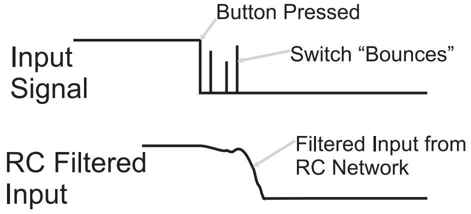
Button Pressed
Input Signal
Switch "Bounces"
RC Filtered Input
Filtered Input from RC Network

Fig. 10-18. RC network button debounce operation.

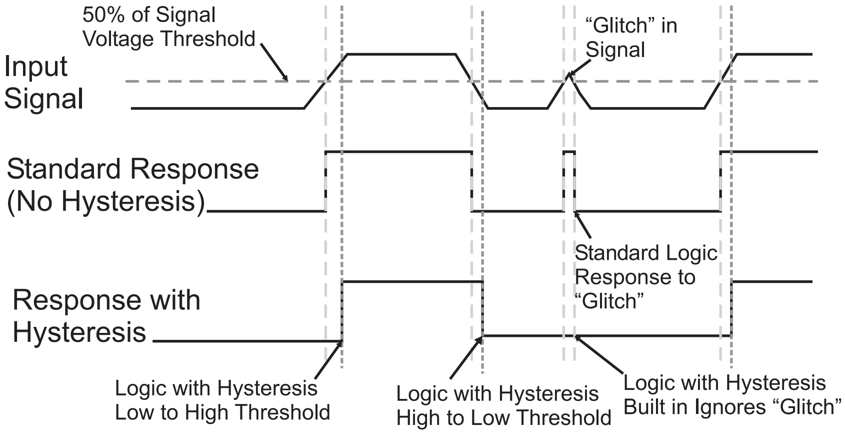
50% of Signal Voltage Threshold
Input Signal
"Glitch" in Signal
Standard Response (No Hysteresis)
Standard Logic Response to "Glitch"
Response with Hysteresis
Logic with Hysteresis Low to High Threshold
Logic with Hysteresis High to Low Threshold
Logic with Hysteresis Built in Ignores "Glitch"

Fig. 10-19. Schmitt trigger input operation.

''Hysteresis'' is the property of the Schmitt trigger inputs in which the threshold point for the rising edge of the signal is different than the falling edge. Looking at Fig. 10-19, you can see that the rising edge threshold is above the ''normal gate voltage threshold'', while the falling edge threshold is less.

These changing threshold values are the reason for the strange symbol on the inverters, indicating Schmitt trigger inputs. Figure 10.20 shows the input versus the gate response on an ''X-Y'' chart. The ''X'' axis is the input voltage with rising voltages to the right and the ''Y'' axis represents the response of the Schmitt trigger input. By following the numbers, you can see the response of the input and that it forms the same symbol that I put on the inverter gates. For comparison, a traditional logic gate does not use this symbol – the response threshold is the same for rising and falling edge signals.

Another method of debouncing button inputs is to use a 555 or monostable circuit. In Fig. 10-21, I show a 555 wired as a monostable, driving out a pulse from a button press. The internal waveforms of the circuit

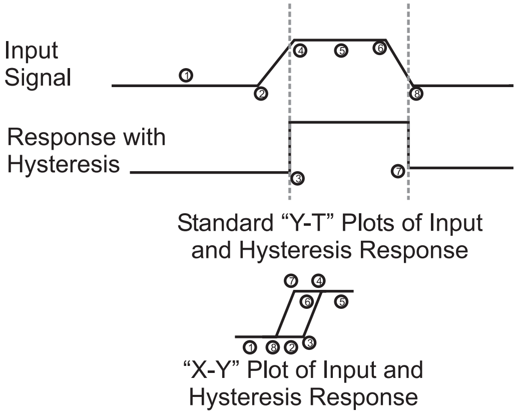

Fig. 10-20. Schmitt input hysteresis.

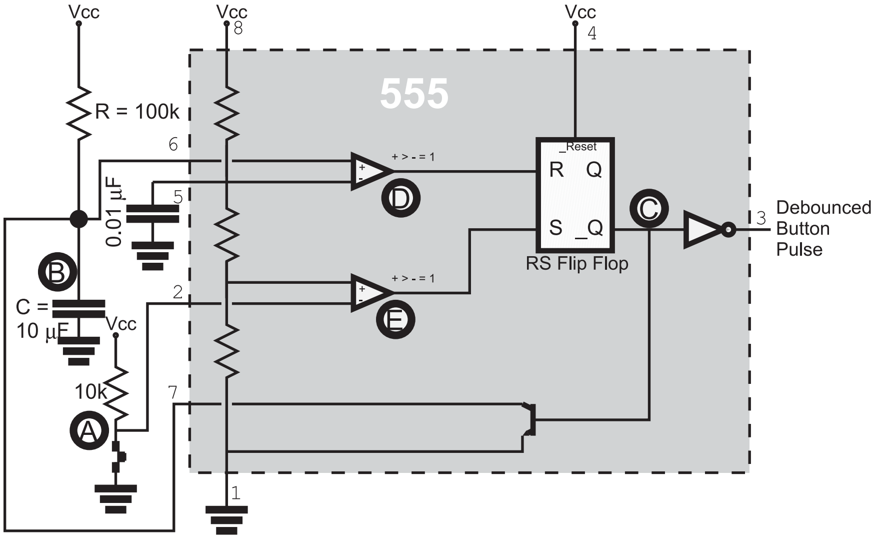
Vcc
Vcc
8
Vcc
4
555
R = 100k
6
5
0.01 µF
_Reset
R Q
S _Q
RS Flip Flop
D
C
E
B
A
C =
10 µF
Vcc
10k
2
7
1
3
Debounced
Button
Pulse

Fig. 10-21. 555 button debounce circuit.

are shown in Fig. 10-22, which shows that any subsequent bouncing of the button after it makes its first connection are ignored by the circuit as the pulse is being output. If you work out the pulse time from R and C, you'll discover that the pulse time is roughly 1 second in length. This should be long enough for a single button press to be registered and the user to remove his fingers. Obviously, this delay is too long to implement multiple buttons or even any kind of data entry functions in the circuit. To do this, you should consider the next section.

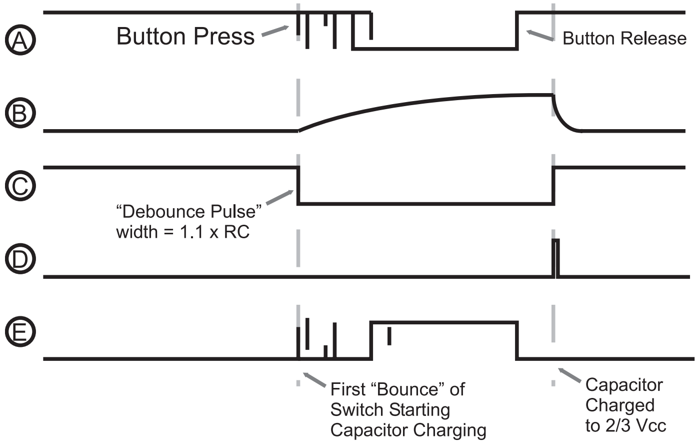

Fig. 10-22. 555 button debounce circuit waveform.

Switch Matrix Keypad Interfacing

As I ended off the previous section, you cannot use simple button debouncing techniques to implement a large number of buttons or even a keyboard for data entry. Just so there's no confusion, I consider a "large number of buttons" to be four or more; providing individual debounce circuits for anything more than a couple of buttons is expensive and time consuming. Along with the cost and time involved, you will also have to come up with some way of prioritizing the button inputs and recognizing non-standard keys like "shift" and "control".

The keys and buttons in PC keyboards and numeric keypads are arranged in "rows" and "columns" and they can be drawn out in such a way that they look like a "matrix". A "momentary on" switch is placed at the intersection of each row and column, as shown in Fig. 10-23. This "switch matrix" provides the ability to "scan" a large number of button inputs with a relatively small number of lines. Your PC's 104/105 keyboard usually has a 22 by 7 matrix connection to a microcontroller, which scans through the keys and reports any key presses using the algorithms presented in this section. Keyboards with a 100 keys or more are an extension of the four-button key matrix shown in Fig. 10-23 and have the same concerns and issues to watch out for.

You probably cannot see immediately how the individual keys or buttons of the switch matrix shown in Fig. 10-23 can be polled, but the operation will probably become clearer when you see the resistors and transistors I've added to the switch matrix in Fig. 10-24.

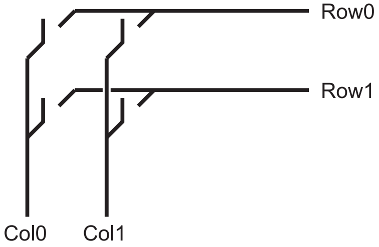

Fig. 10-23. Simple switch matrix.

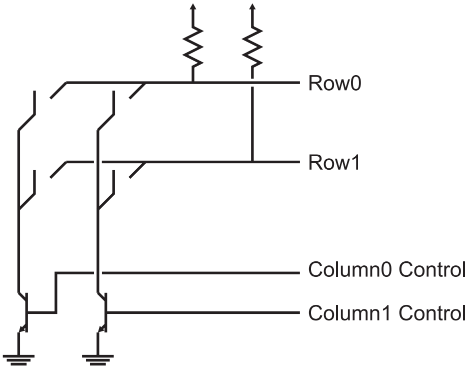

Fig. 10-24. Switch matrix with I/O circuitry.

In this case, by connecting one of the columns to ground, if a switch is closed, the pull down on the row will connect the line to ground. When the row is polled by an I/O pin, a "0" or low voltage will be returned instead of a "1" (which is what will be returned if the switch in the row that is connected to the ground is open due to the pull up on it). To scan the keyboard, the column transistors are turned on, one at a time, and while the column transistor is on and the column is pulled to ground, the rows are compared to a logic level of "0", which would indicate that the button is pressed.

This methodology for handling switch matrix keypad scans I've outlined here probably seems pretty simple. Depending on your familiarity with programming and different microprocessors and microcontrollers, you will probably realize that implementing these functions could be done even simply in assembly language programming or "C". You should also realize that this code would be quite difficult to implement just using logic chips.

To avoid the complexities of trying to develop TTL logic that will carry out the functions described in the pseudo-code presented above, I normally

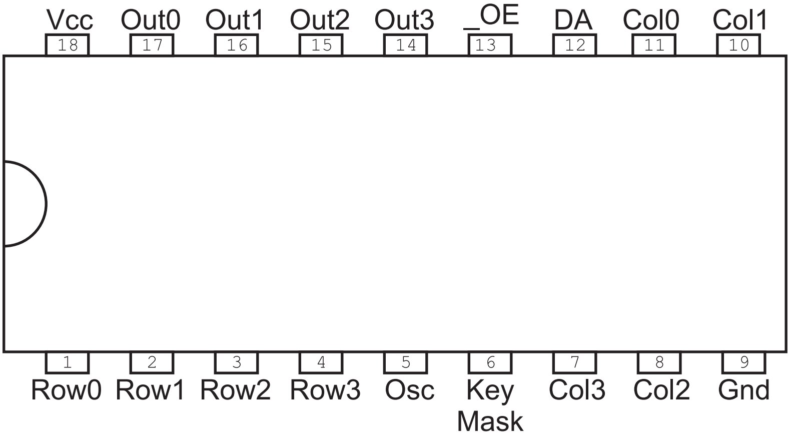
Vcc
18
Out0
17
Out1
16
Out2
15
Out3
14
_OE
13
DA
12
Col0
11
Col1
10
1
Row0
2
Row1
3
Row2
4
Row3
5
Osc
6
Key Mask
7
Col3
8
Col2
9
Gnd

Fig. 10-25. 74C922 switch matrix keypad decoder chip pinout.

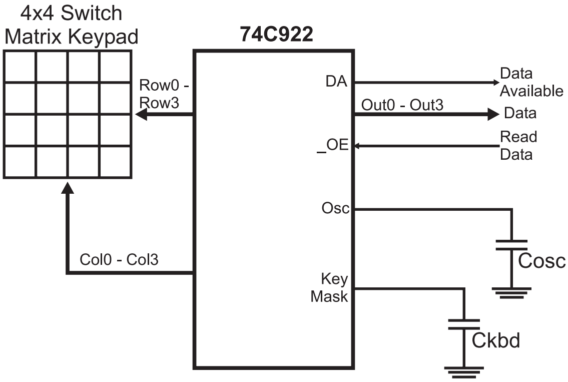
4x4 Switch Matrix Keypad
74C922
Row0 - Row3
DA
Data Available
Out0 - Out3
Data
_OE
Read Data
Osc
Cosc
Col0 - Col3
Key Mask
Ckbd

Fig. 10-26. Basic 74C922 wiring.

use the 74C922 keypad decoder chip (Fig. 10-25). This chip can be used to debounce and encode up to 32 buttons (although 16 is the normal maximum) and carries out button debouncing internally as well as keeping track of two currently held-down keys when new keys are pressed. The 74C922 is quite easy to wire to a four by four (16 button) switch matrix keypad, as shown in Fig. 10-26. By "doubling up" rows of sensors of the 74C922, you can add a number of additional keys to the application. In the next section, I will show how this is done to create a 20 button input device (with up to 32 possible).

The two capacitors are used to create a relaxation oscillator within the chip that is used to "scan" through the buttons as well as provide a "debounce" delay count for the application. The two capacitor values are calculated as:

$$\texttt{Cosc} = \texttt{Scan Rate/10 K}$$
$$\texttt{Ckbd} = \texttt{10}^{*}\,\texttt{Cosc}$$

I like a debounce interval of 20 ms: plugging this into the formulas above, I get a value of 2 μF and 20 μF. When I build my own applications, I tend to have a lot of 10 μF (for power filtering) and 1 μF (for MAX232 RS-232 level converters) electrolytic capacitors on hand. I have not found any problems with using these components and I would recommend that you use them as well to avoid having to stock multiple capacitor values for different applications.

Quiz

1. What are ideal sequential circuit interfaces?
   (a) LCDs
   (b) Individual switches and LEDs
   (c) Keyboards
   (d) USB flash disks
2. What is the suggested digital electronics interface to an Hitachi 44780 controlled LCD display?
   (a) Hardware state machine
   (b) Microcontroller
   (c) Sequential circuit
   (d) Combinatorial circuit
3. Is a bus "Read" or "Write" faster?
   (a) Write is faster
   (b) Read is faster
   (c) Using a synchronous clocked circuit, they take the same amount of time
   (d) Read is slower due to the need to retrieve data from the interface device
4. Seven-segment LED displays have a common:
   (a) LED anode or cathode
   (b) Segment pins

- (c) LED drivers built into the package
- (d) Pin interface that is used by all devices, regardless of the number of digits in the package

5. Multiple seven-segment LED displays show different values by using:
   - (a) Linear feedback shift registers that have encoded the bit patterns
   - (b) Turning on each individual digit with its unique value periodically
   - (c) Multiple LED driver circuits that drive the value for its respective digit to the LED display
   - (d) Multiple bits of memory, one for each segment, which are loaded according to the display value

6. For a PWM circuit running its logic at 5 volts and a duty cycle of 67%, what is the "on" voltage level of the output signal?
   - (a) 0.67 volts
   - (b) 5 volts
   - (c) 2/3 volts
   - (d) 3.35 volts

7. The power dissipated by a PWM running with a 20% duty cycle will be what fraction of a 100% duty cycle?
   - (a) 0.04
   - (b) 40%
   - (c) 0.4
   - (d) 400%

8. A 555 monostable with $R = 100\,\text{k}$ and C of $4.7\,\mu\text{F}$ will output a pulse of:
   - (a) Approximately 0.5 s
   - (b) Approximately 4.7 s
   - (c) Approximately 1.1 s
   - (d) Insufficient data given to determine the pulse width

9. Rows and columns in a switch matrix keypad have what connected to them?
   - (a) The rows have a transistor connected to ground and the columns have a capacitor
   - (b) The columns have transistors connected to ground and the rows are left open

(c) The columns are left open and the rows have a transistor connected to ground
(d) The columns have transistors connected to ground and the rows have pull up resistors

10. The 74C922 reads a switch matrix keypad by:
(a) Pulling the columns of a switch matrix keypad to ground and scanning the rows for pulled down bits
(b) Driving a 1 kHz square wave on the rows and polling the columns for the signal
(c) Using an internal microprocessor
(d) Measuring the capacitances of individual lines and looking for changes

11

CHAPTER

Reading Datasheets

I've never understood why college and university courses do not give an introductory course in reading digital electronic device datasheets. Despite how prepared you are for them, you will feel quite overwhelmed the first time you have to look through a number of datasheets trying to find a part that meets your requirements. When I first started working with electronics, datasheets were generally quite poor, with only a few standout companies providing good documentation for their chips. Fortunately, this has changed over the past 10 years, the Internet and the capability of downloading good-quality datasheets being almost a marketing tool to help engineers select the parts they are going to use in their designs.

Personally, I find it more daunting to look at datasheets over the Internet because they are generally encoded as Adobe Acrobat pdfs that take a while to load and you can never flip the pages on the screen as fast as you would like. To make matters worse, it can be very difficult to put multiple datasheets up on a computer display to allow you to compare the features of the different chips. This difficulty gives rise to the most important recommendation that I can make about downloading datasheets from the Internet – print them out! I have several binders of printed out datasheets for parts that I often use. By printing them out, I have immediate access to them and I can

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

flip back and forth between pages effortlessly. It is my opinion that documentation shouldn't be "paperless".

The first sheet of the datasheet is usually a one page description of the part. It normally contains:

1. Part number
2. High-level part description
3. Part pinout
4. Common/related/pin compatible parts
5. Important chip features
6. Basic operations truth table

When looking at a datasheet, you should first check out the part number of the datasheet versus what you are interested in. This means that you should be checking not only the numeric code for the device but also the high-level identifier and the technology identifier. For example, if you were looking for a low-power TTL dual input NAND gate and looked up the datasheet based on a web search for "TTL dual input NAND", you could see such diverse part numbers as:

```
74LS00
74C00
54LS00
74W00
74ALS03
```

with the question being: Which is the one that you want?

For these parts the high-level identifier is the "74" or "54". In this book, I have focused on the 74 series of logic – but if you look at "54" series logic you will see that its operation is identical and may decide to go ahead and order the parts. This could be a big problem because "54" series parts are military-grade chips and they tend to cost 10 times that of standard "74" series logic and do not necessarily have the same pinout as "74" series chips. The technology identifier is the letter code between the high-level identifier and the part number. In the list of five chips above, I have presented traditional TTL low-power logic ("LS"), CMOS logic ("C"), advanced Shottkey low power ("ALS") and single gate CMOS ("W"). The danger of not reading the datasheet's part number is that you could end up ordering the wrong part number, resulting in higher than expected costs and lost time looking up and reordering the correct part.

Always read through the datasheet's first page high-level part number description. This can range from a single sentence to four or five bulleted items. Like the part number check, this should just be a filter operation, resulting in you making sure the part will do essentially what you want it to do.

The part pinout is something that is critical to know when you are wiring a circuit. Except for the wiring experiments presented in this book, I have listed only a few part pinouts because a part's pinout may vary between manufacturers of the same part and they may vary according to the packaging type. The part pinout may also change according to packaging technology. It isn't unusual to see a pin through hole (PTH) packaged chip with a specific pinout but its surface mount technology (SMT) sibling having extra pins or different connections to different pin numbers. I'm sure that both of these statements are a bit hard to understand; you might be thinking that the part numbers are standard. I wish I could tell you how many times I have been bitten by these two little traps. Circuit design systems also make assumptions about part pinouts based on the pinouts from specific manufacturers and don't bother checking the pinouts from others.

The important chip features listed on the front page of the datasheet will not list the features of the chip to the lowest possible level, but it will give you some ideas about how the chip works and if there are any issues that could be a problem with you using the chip in your application.

A lot of times you will discover that a part will not have exactly the functions that you want but, by checking the datasheet, it may list related parts that provide a similar function that you can take a look at. Finally, for very simple chips, the front page of the datasheet will present you with truth tables describing the operation of the chip or the different parts of the chip.

The front page of a chip's datasheet can be incredibly useful to you and by spending a few minutes familiarizing yourself with it, you can decide whether or not the chip is appropriate for your application without having to delve into the minutia of the following pages.

Chip Operating Characteristics

An important feature of the datasheet is the "operating characteristics" for the chip. This section of the datasheet explains such operating parameters as:

1. Input pin voltage thresholds and currents
2. Output voltages along with current source and sinking capabilities

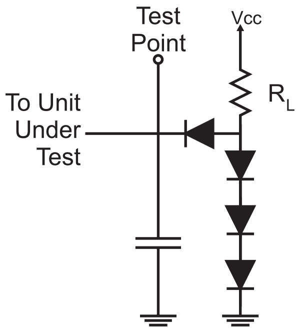

Fig. 11-1. Sample chip test circuit.

3. Gate delay timing
4. Expected input and output pin line impedances
5. Miscellaneous operating information.

Each chip datasheet lists the logic thresholds, along with their characteristics when subjected to different parameters. Often you will see a two-axis graph, with a curve showing the chip characteristic response to the changing input parameters. This part of the drawing should not be difficult to understand but what can be confusing is the small schematic marked as a ‘‘test circuit’’ that often accompanies the graph. An example of such a schematic is shown in Fig. 11-1; it shows the circuit that was connected to the pin while the test was taking place. These test circuits simulate other circuits connected to the chip, helping to ensure that the chip is operating as it would in a typical application.

The output voltage and current characteristics are really a function of the logic technology used and not unique to the individual chip’s pins. If you were to read other datasheets of chips built from the same technology you would discover that the output parameters are the same between the two chips and, by extrapolation, all the chips built from this technology. If you were to search the manufacturer’s web site, you would discover that this information has been published for all parts in the technology family and the information in the chip’s datasheet is really redundant. The reason why the information is repeated in the individual chip’s datasheets is to minimize the amount of cross-checking that you will have to do.

The previous comment could be made about gate delay timing but there is a wrinkle in the specification in the datasheet. Is the quoted ‘‘gate delay’’ for the chip function or for the individual basic technology gate (i.e. the ‘‘NAND’’ gate for TTL)? Normally, the datasheet will list the chip function gate delay instead of the basic technology gate delay because the actual gate delay is probably less than the product of the basic gate delay time multiplied by the number of gates the signal has to pass through.

As you learn more about electronics, you learn that not only do wire connections have resistance but they also have capacitance and impedance. All these factors affect the transmission of data signals and are known by the term "characteristic impedance". Printed circuit boards (PCBs) have a characteristic impedance of 55 Ω (the coax cable that sends signals to your TV has a characteristic impedance of 75 Ω). The input and output pins must be designed to match with the 55 Ω PCB characteristic impedance to ensure that signals pass between pins as efficiently as possible.

All these chip characteristics and any miscellaneous data that the chip manufacturer feels important enough to include should be read through and understood in order to best wire a chip into your application circuit.

IEEE Logic Symbols

When you look at some datasheets, you will see the function of the chip described using a graphical system that is different from the one that I have used in this book. Instead of unique shapes for each gate, they are represented as a rectangular block like the one in Fig. 11-2. These blocks are part of the "IEEE Standard Graphic Symbols for Logic Functions". This standard is often used to describe the operation of a chip instead of the graphical symbols that I have used in the book.

The IEEE gate definition contains a single character to indicate what the function is. Table 11-1 lists the basic characters and their functions. For negated outputs and inputs, the gate pin modifiers presented in Fig. 11-3 are used. Note in Table 11-1, only the four unique gate functions are listed – NAND and NOR gates are represented with the gate modifiers shown in Fig. 11-3.

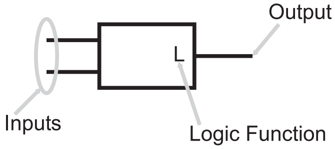

Fig. 11-2. Basic IEEE logic gate symbol.

Table 11-1 Four basic IEEE logic functions.

IEEE gate symbol	Gate type
&	AND
|	OR
!	NOT
ˆ	XOR

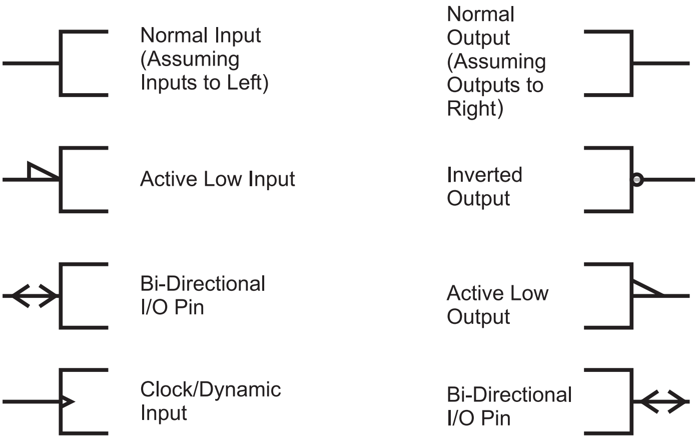
Normal Input (Assuming Inputs to Left)
Normal Output (Assuming Outputs to Right)
Active Low Input
Inverted Output
Bi-Directional I/O Pin
Active Low Output
Clock/Dynamic Input
Bi-Directional I/O Pin

Fig. 11-3. IEEE logic gate pin modifiers.

For the complete IEEE logic symbol definition, I suggest that you download and printout ANSI/IEEE Std 91a-1991 from:

http://www.ee.ic.ac.uk/pcheung/teaching/ee1_digital/logic symbols.pdf

This document outlines the conventions used to identify each of the different functions used to describe different logic functions in the IEEE standard format. This method of presenting chip functions may seem to be rather difficult to decode when you first see it, but after you've worked with it a while, it will become second nature to you.

Having said this, I would suggest that you avoid working with these symbols until you are very familiar with working with digital electronics. The standard graphic symbols have been well thought out and are immediately recognizable when you are first learning to work with digital electronics and logic gates. The IEEE symbols can be difficult to distinguish when you are first starting out and you can very easily get yourself into trouble if you misread a symbol or forget the purpose of a pin modifier.

Power Usage and Fanouts

An important consideration for selecting a chip is the amount of power dissipated. This information is necessary not only for the individual gate but also for the complete application. The sum of all the power is the total power needed by the application and this value will dictate the power methodology used as well as the cooling requirements for the final product. While not explicitly a correlation, you will find that the more power a logic technology uses, the more external inputs that can be driven (this is the technology's ''fanout'').

When you look at a chip's current (which is related to its power) consumption, remember to look at not only the current required to power the gate but also at maximum input sinking and output sourcing or sinking. These currents should all be added together to get the worst-case power consumed by the chip. I have seen a number of products where the designer expressed the power consumption by what he thought was a ''typical case'' and found out that the actual current consumption is somewhat higher and the specified power supply did not have sufficient margin for the product to work reliably.

Along with the current consumption, the datasheet should also specify the number of input pins the chip's output pins can drive. It is important to note that the number of input pins quoted is the *same technology* as the chip. When you are mixing technology, you will have to understand the input current requirements of the input pins and, as ''a rule of thumb'', make sure that the total current drawn by the input pins does not exceed 50% of the total sinking current capability of the output pin. This will ensure that the logic functions will be at the correct levels regardless of the circumstances.

Actually, I would recommend that for your first applications, you *never* drive more than three inputs from a single output and strive to drive no more than two outputs in the design. Marginal signals due to overloaded output pins are very difficult to recognize from the failure symptoms and difficult to confirm when the problem is suspected.

Quiz

1. What isn't on the first page of the datasheet?
   (a) Part number
   (b) Part pinout
   (c) Chip cost
   (d) Important chip features
2. What is the technology identifier in the part number "74S174"?
   (a) 174
   (b) 74
   (c) 74S
   (d) S
3. What parameter isn't a chip operating characteristic?
   (a) Gate delay timing
   (b) Chip logic function
   (c) Input pin voltage thresholds and currents
   (d) Expected input and output pin line impedances
4. The chip gate delay specification
   (a) Is for ideal conditions
   (b) Is for the basic technology gate delay
   (c) Is for the chip function
   (d) Is for the NAND gate delay
5. IEEE symbols
   (a) Will replace the standard graphical symbols
   (b) Represent negative output functions by placing a symbol on the output pin
   (c) Is used to define all chips
   (d) Are only used for basic logic gates
6. Starting out using IEEE symbols
   (a) Is a bad idea as the symbols are not immediately recognizable
   (b) Is the recommended way to learn about digital electronics
   (c) Will help you design highly optimized digital electronics circuits
   (d) Will encourage you to buy from manufacturers that properly document their products

7. When planning for the current consumption of a product, which current specification should be ignored?
   (a) Standby current
   (b) Output low current sink
   (c) Input low current drain
   (d) None

8. The maximum number of inputs a single output can drive is:
   (a) Determined by the total current drawn by the inputs
   (b) Three in all cases
   (c) The output sink current specification divided by the average input current drain
   (d) Infinite

CHAPTER

12

Computer Processors and Support

I'm sure that you realize that computer processors are really just a great big sequential circuit, but I'm sure that you have no idea where to start understanding how they work. Traditional computer processors are designed using a selection of six or so basic design philosophies that give them different characteristics. In this chapter, I will introduce you to the different issues that have to be confronted in designing computer processors, along with some of the technologies that have been developed to support them.

From a high level, computer processor architects choose from making the processors "RISC" ("Reduced Instruction Set Computers" – pronounced "risk") based or "CISC" ("Complex Instruction Set Computers") based. CISC processors tend to have a large number of instructions, each carrying out a different permutation of the same operation (accessing data directly, through index registers, etc.) with instructions perceived to be useful by the processor's designer while RISC systems minimize the instruction set, but give them as much flexibility and access as much of the memory in the system

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

as possible. CISC processors also have the same requirement, but by definition, they are designed to simplify the amount of manipulation that is required by the programmer. Both computer types have their advantages and disadvantages – the RISC tends to be easier to design and executes instructions faster while the CISC tends to be easier to program but may be cumbersome in implementing some functions.

The second option processor designs have came from a competition between Harvard and Princeton universities to come up with a computer architecture that could be used to compute tables of naval artillery shell distances for varying elevations and environmental conditions. Princeton's response was for a computer that had common memory for storing the control program as well as variables and other data structures. It was best known by the chief scientist's name "John Von Neumann". Figure 12-1 is a block diagram of the Princeton architecture. The "Memory Interface Unit" is responsible for arbitrating access to the memory space between reading instructions (based upon the current Program Counter) and passing data back and forth with the processor and its internal registers. In contrast, Harvard's response (Fig. 12-2) was a design that used separate memory banks for program storage, the processor stack and variable RAM. By separating the data and program memories and avoiding the need to arbitrate data movements between them, there was an opportunity for programs to execute faster in Harvard's computer.

It may at first seem that the Memory Interface Unit of the Princeton architecture is a bottleneck between the processor and the variable/RAM space – especially with the requirement for fetching instructions at the same time. In many Princeton architected processors, this is not the case because of the time required to execute an instruction is normally used

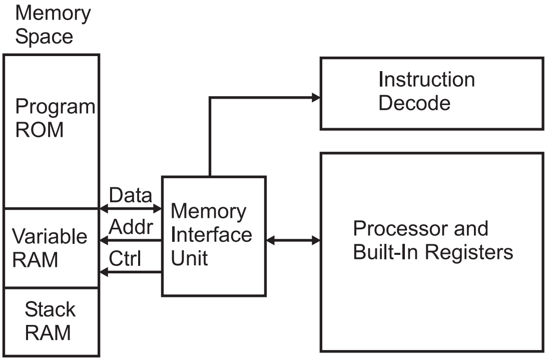

Fig. 12-1. Princeton processor architecture.

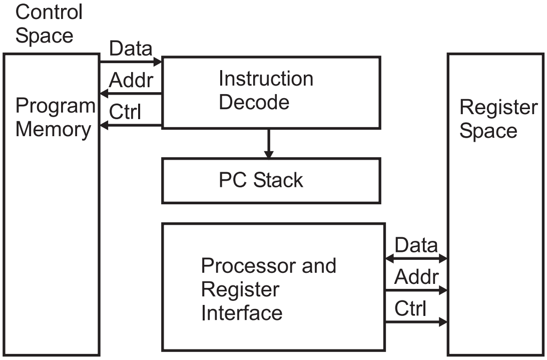

Fig. 12-2. Harvard processor architecture.

to fetch the next instruction (this is known as "pre-fetching"). Other processors (most notably the processor in your PC) have separate program and data "caches" that can be accessed directly while other address accesses are taking place.

The Princeton architecture won the competition because it was better suited to the technology of the time. Using one memory was preferable because of the unreliability of then current electronics (this was before transistors were in widespread general use): a single memory and associated interface would have fewer things that could fail. The Harvard architecture is really best for processor applications that do not process large amounts of memory from different sources (which is what the Von Neumann architecture is best at) and be able to access this small amount of memory very quickly.

Once the processor's instruction set philosophy and architecture have been decided upon, the design of the processor is then passed to the engineers responsible for implementing the design in silicon. Most of these details are left "under the covers" and do not affect how the application designer interfaces with the application. There is one detail that can have a big effect on how applications execute, and that is whether or not the processor is a "hardcoded" or "microcoded" device. Each processor instruction is in fact a series of instructions that are executed to carry out the instruction. For example, to load the accumulator in a processor, the following steps could be taken:

1. Output Address in Instruction to the Data Memory Address Bus Drivers.
2. Configure Internal Bus for Data Memory value to be stored in Accumulator.
3. Enable Bus Read.

4. Compare Data read in to zero or any other important conditions and set bits in the "STATUS" Register.
5. Disable Bus Read.

A microcoded processor is really a computer processor within a processor. In a microcoded processor, a "state machine" executes each different instruction as the address to a subroutine of instructions. When an instruction is loaded into the "Instruction Holding Register", certain bits of the instruction are used to point to the start of the instruction routine (or microcode) and the "uCode Instruction Decode and Processor" Logic executes the microcode instructions until an "instruction end" is encountered. This is shown in Fig. 12-3.

A "hardwired" processor uses the bit pattern of the instruction to access specific logic gates (possibly unique to the instruction) which are executed as a combinatorial circuit to carry out the instruction. Figure 12-4 shows how

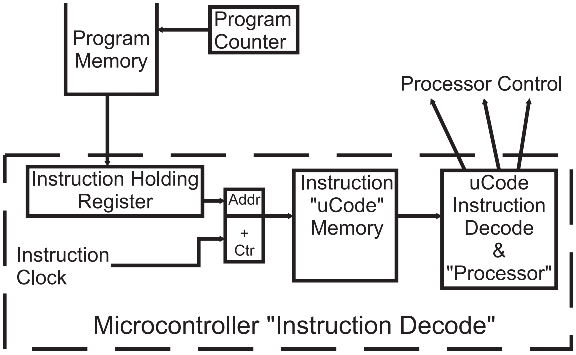
Program Memory
Program Counter
Processor Control
Instruction Holding Register
Addr
+ Ctr
Instruction "uCode" Memory
uCode Instruction Decode & "Processor"
Instruction Clock
Microcontroller "Instruction Decode"

Fig. 12-3. Microcoded or state machine processor instruction decode circuitry.

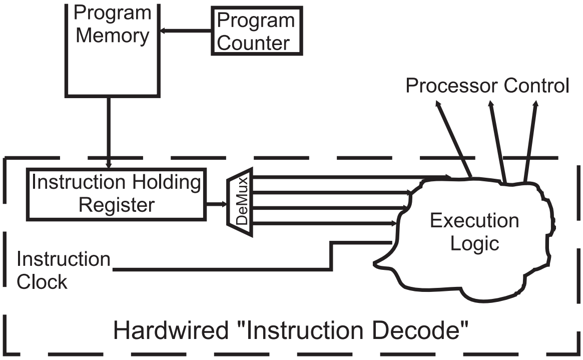
Program Memory
Program Counter
Processor Control
Instruction Holding Register
DeMux
Execution Logic
Instruction Clock
Hardwired "Instruction Decode"

Fig. 12-4. Hardwired processor instruction decode circuitry.

the instruction loaded into the Instruction Holding Register is used to initiate a specific portion of the "Execution Logic" which carries out all the functions of the instruction.

Each of the two methods offers advantages over the other. A microcoded processor is usually simpler than a hardwired one to design and can be implemented faster with less chance of having problems at specific conditions. If problems are found, revised "steppings" of the silicon can be made with a relatively small amount of design effort. The hardwired processor tends to execute instructions much faster but is much harder to modify.

IEEE754 Floating Point Numbers

When I introduced binary numbers earlier in the book, I discussed binary integers, but I did not discuss how binary "real" numbers were produced or how they were manipulated in workstation processors. It should not be a surprise to discover that binary floating point numbers are analogous to decimal floating point numbers.

For example, if you were going to convert decimal 7.80 to binary, you would first convert the value equal to or greater than one to binary. Decimal 7 becomes B'0111', leaving decimal 0.80 to convert. This is accomplished by knowing that decimal fraction digits are multiplied by *negative* exponents of the base 10. The same methodology can be used for binary numbers.

To convert decimal 0.80 to a binary fraction, I will start with the exponent "−1" which is equal to 0.5 decimal and test to see if it can be removed from the fraction. Since it can, my binary number becomes B'0111.1' with a remainder of 0.30. Going to the next negative exponent ("−2"), I discover that I can subtract this value, giving me the binary value B'0111.11'. Continuing this on for another four bits, the binary value is B'0111.110011'. It's interesting to see that the binary number is irrational; the bit pattern will change the smaller the fraction that is calculated even though the decimal number ends at the first digit after the decimal point.

This method can be expressed as the "C" function, which converts the floating point number to a binary string:

```
BinaryFraction(float Value, char* BinaryValue)
{                            // Convert the Floating Point Value to
                              //  a string pointed to by BinaryValue
int i, k;
int j = 0;
      BinaryValue(j + +) = `B´;      BinaryValue(j + +) = `\´´;
                                            // Initialie Binary String
    for(i = 0;  (2<<i) < = Value;  i + +);
                                            //Find Most Significant Bit
                                             // of Value´s Integer Value
    for (; i ! = −1;  i − −)
      if ((Value − (2<<i))> = 0)
         {                                     // Take Away Binary Value
        Value = Value − (2<<i);
          BinaryValue (j + +) = `1´;
        } else                                // Can´t take anything away
        BinaryValue (j + +) = `0´;
          BinaryValue (j + +) = `.´;    // Find fractional value
                                                          // to 4 digits
    for (I = 1, k = 0;  k<4;  i + +, k + +)
      if (Value − (1 / (2<<i)) > = 0)
      {                                 // Can take away binary fraction
        Value = Value − (1 / (2<<i));
        BinaryValue (j + +) = `1´;
      } else
        BinaryValue(j + +) = `0´;
    BinaryValue(j + +) = `\´´;    //Close Binary String
    BinaryValue(j + +) = `\0´;
} // End BinaryFraction
```

This operation can be performed within high-performance processors (like the Intel Pentium), but instead of producing a string of characters representing binary data, they generally put them into the IEEE754 format, which stores the floating binary value in a format which is similar to that of "Scientific Notation":

(Sign) Mantissa × 2** (Exponent Sign) Exponent

The "Mantissa" is multiplied by the signed exponent to get values less than or greater than one.

Table 12-1 IEEE754 data sizes.

Data format	Data range	Data size
Long word	+/− 9.2(10**18)	8 Bytes
Packed BCD	+/− 10**17	8 Bytes
Single precision real	+/− 10**38 to +/− 10**−38	4 Bytes
Double precision real	+/− 10**308 to +/− 10**−308	8 Bytes
Extended precision real	+/− 10**4932 to +/− 10**−4932	10 Bytes

Table 12-1 lists the data formats supported by the Intel Pentium. These different formats give you a lot of flexibility to work with a wide range of numbers in different applications. All the number formats can be processed together, with the final result being in the most accurate format (i.e. a “word” and “single precision” combined together will have a result as a single precision number).

Memory Types

A number of different memory types are currently available. In this introduction, I will first show you three different technologies and discuss where (and why) they are used in a computer system. The boot up, non-volatile memory used in a computer system is based on ultraviolet light “Erasable PROM” (“EPROM”) program memory (Fig. 12-5) and was first introduced in the late 1960s. An EPROM memory cell consists of a transistor

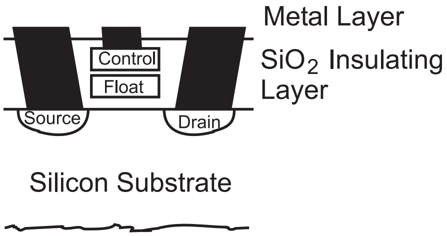

Fig. 12-5. MOSFET EPROM memory cell.

that can be set to be always ‘‘on’’ or ‘‘off’’. Figure 12-5 shows the side view of the EPROM transistor.

The EPROM transistor is a MOSFET-like transistor with a ‘‘floating’’ gate surrounded by silicon dioxide above the substrate of the device. ‘‘Silicon dioxide’’ is best known as ‘‘glass’’ and is a very good insulator. To program the floating gate, the ‘‘Control’’ gate above the floating gate is raised to a high enough voltage potential to have the silicon dioxide surrounding it to ‘‘break down’’ and allow a charge to pass into the floating gate. With a charge in the floating gate, the transistor is turned ‘‘on’’ at all times, until the charge escapes (which will take a very long time that is usually measured in tens of years).

An improvement over UV erasable EPROM technology is ‘‘Electrically Erasable PROM’’ (‘‘EEPROM’’). This non-volatile memory is built with the same technology as EPROM, but the floating gate’s charge can be removed by circuits on the chip and no UV light is required. There are two types of EEPROM available. The first type is simply known as ‘‘EEPROM’’ and allows each bit (and byte) in the program memory array to be reprogrammed without affecting any other cells in the array. This type of memory first became available in the early 1980s.

In the late 1980s, Intel introduced a modification to EEPROM that was called ‘‘Flash’’. The difference between Flash and EEPROM is Flash’s use of a bussed circuit for erasing the cells’ floating gates rather than making each cell independent. This reduced the cost of the EEPROM memory and speeded up the time required to program a device (rather than having to erase each cell in the EEPROM individually, in Flash the erase cycle, which takes as long for one byte, erases all the memory in the array).

For high-speed storage, data is saved in ‘‘Static Random Access Memory’’ (‘‘SRAM’’) which will retain the current contents as long as power is applied to it and is known as ‘‘volatile’’ memory. This is in contrast to the ‘‘EPROM’’ or Flash, which does not loose its contents when power is taken away but cannot have its contents changed as easily as ‘‘SRAM’’. Each bit in a SRAM memory array is made up of the six transistor memory cell, as shown in Fig. 12-6. This memory cell will stay in one state until the ‘‘Write Enable’’ transistor is enabled and the write data is used to set the state of the SRAM cell.

The SRAM cell could be modeled as the two inverters shown in Fig. 12-7. Once a value has been set in the inverters’ feedback loop it will stay there until changed. Reading data is accomplished by asserting the read enable line and inverting the value output (because the ‘‘read’’ side contains the inverted ‘‘write’’ side’s data). The driver to the SRAM cell must be able to ‘‘overpower’’ the output of the inverter in order for it to change state.

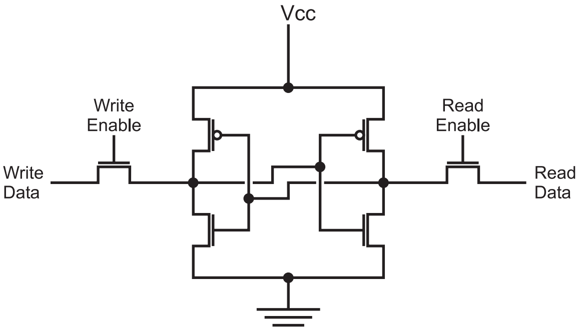
Vcc
Write
Enable
Read
Enable
Write
Data
Read
Data

Fig. 12-6. MOSFET SRAM cell.

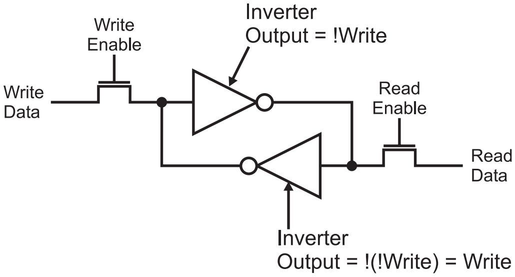
Write
Enable
Inverter
Output = !Write
Write
Data
Read
Enable
Read
Data
Inverter
Output = !(!Write) = Write

Fig. 12-7. Equivalent circuit to SRAM cell.

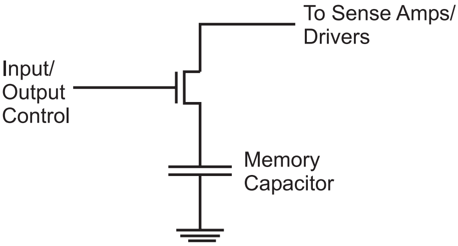
To Sense Amps/
Drivers
Input/
Output
Control
Memory
Capacitor

Fig. 12-8. DRAM memory cell.

The outputs of the inverters are usually current limited to avoid any backdriving concerns.

Large, inexpensive and reasonably high-speed memory can be built from "Dynamic Random Access Memory" ("DRAM") cells. You may have heard the term "Single Transistor Memory Cells" for descriptions of DRAM and that's actually a pretty good description of each cell, which is shown in Fig. 12-8.

In DRAM memory cell, the transistor is used as a switch to allow a charge to be moved into or out of the capacitor. For a write, the transistor is turned on and a charge is either pushed into or pulled out of the capacitor. When the transistor is turned off, the charge is trapped in the capacitor and cannot change until the transistor is turned on again.

A DRAM read is accomplished by turning on the transistor and any charge that is in the capacitor will leak out and will be detected and amplified by a "Sense Amplifier". The "Sense Amp" is a metastable flip flop that will be set to the state of the capacitor when the transistor switch is closed. Before the transistor is turned on when writing to the cell, the sense amp will be set to a specific state to load the correct charge into the capacitor.

In a DRAM memory chip, the cells are arranged in rows and columns, as shown in Fig. 12-9. To address each cell within the chip, a row/column address for the element in the array has to be provided. Usually, to save pins on the DRAM chips, the row and column address lines are shared (multiplexed) together so that during a read or a write, first the "Row" is selected and then the "Column".

The row is selected first so that if a write is taking place the sense amp for the row can be set to the specific value. All the other sense amps are set in their metastable state. When the Column address is latched in, the transistors for the array row are turned on. Next, when the row address is available, the "Input"/"Output Control" transistor is turned on, then the read/write takes

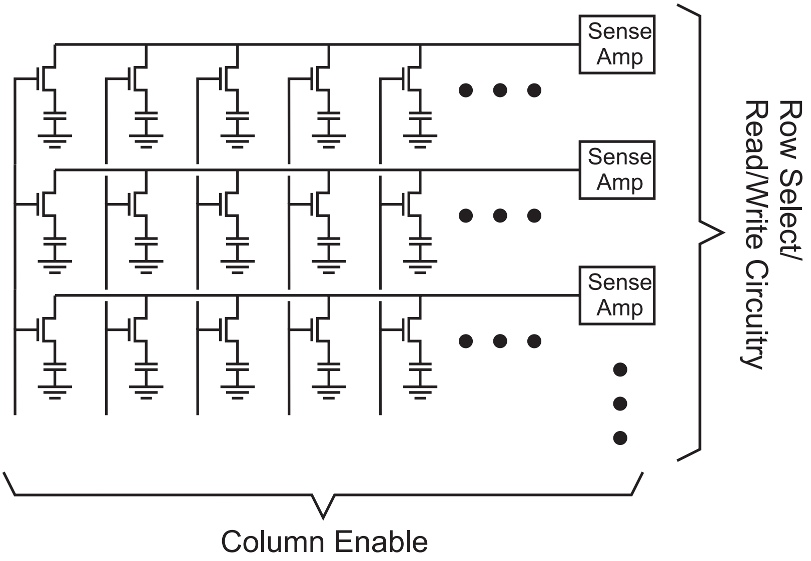

Fig. 12-9. DRAM cell array.

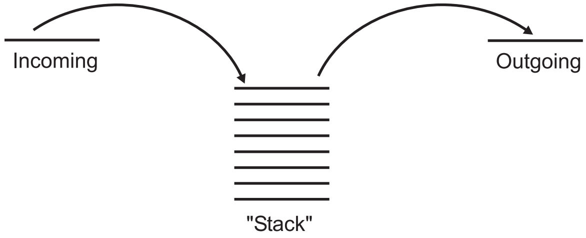

Fig. 12-10. Data stack operation.

place. For the cells not being written to, the sense amps will not only read what the charge is in the capacitor but will also ‘‘refresh’’ it as well.

This is a very good thing in DRAM because the capacitor is actually a MOS transistor built into the chip, acting as a capacitor. Over time, any charge in this capacitor will leak away into the silicon substrate. By periodically ‘‘refreshing’’ the charge by performing a read (which will cause the sense amps to amplify the charge), the contents of the memory will never be lost.

Refreshing is typically done by enabling all the transistors in a column (without first specifying a row) and letting the sense amps do their thing. A read of an incrementing column address is usually implemented in the DRAM support hardware and is known as a ‘‘CAS (‘‘Column Address Strobe’’) Only Refresh’’. In the original PC, 5% of the processor bandwidth was lost due to DRAM refresh requirements. To overcome these potential deficiencies, PC designers have come up with a few hardware features.

Along with being arranged as simple, single-dimensional arrays of data, memory can also be built into ‘‘stacks’’ (Fig. 12-10). Processor stacks are a simple and fast way of saving data during program execution. Stacks save data in a processor the same way you save papers on your desk and is known as ‘‘last in/first out’’ (‘‘LIFO’’) memory. As you are working, the work piles up in front of you and you do the task that is at the top of the pile.

Power Supplies

It is surprising to many people that you can add a simple voltage regulator to power your projects for just a few dollars; cheaper than a set of rechargeable batteries. Voltage regulators, powered by an AC/DC ‘‘Wall wart’’ power converter, will convert one DC voltage to another that can be used by the electronics in your circuit and, more importantly, will be tolerant of changes in the AC supply and the current load. In this section, I will introduce you to some simple power supply circuits that have the following characteristics:

1. They are safe for their users and designers.
2. They are relatively efficient in terms of the amount of power that is lost converting voltage levels.
3. They provide very accurate voltage levels, independent of the voltage input or the current required by the application.
4. They are inexpensive.
5. Their design can be optimized for the application that they are providing power for.
6. These supplies source up to 1 amp of current.

The power supply ideas presented here are very appropriate for the simple circuits discussed in this book; the 250 watt power supply used for your PC requires methodologies and circuits for producing this much power that are quite a bit different than what is required for the simple power supplies presented here. Advanced degrees are normally required for properly designing high current power supplies that work at high efficiencies.

There are some semiconductor-based circuits, like Zener diode power supplies (Fig. 12-11) that do lend themselves to being modeled using water analogs. The Zener diode power supply works as a shunt regulator – applying a specified amount of current to a circuit at a rated voltage and shunting the rest away as wasted power.

When the term "shunt" is used, it is simply saying that excess voltage and current is turned away from the circuit. This concept can be illustrated with a water pressure regulator created from a catch basin with a hole at the bottom; water coming out of the hole is at a pressure which is determined by the depth of water in the basin. To maintain this depth (and bottom pressure), even though water is being drawn from the hole at the bottom, "source" water is continually poured into the basin. More water is pouring in than is expected to exit through the hole in the bottom, with the excess leaking out over the side. This is exactly how the Zener diode works, except

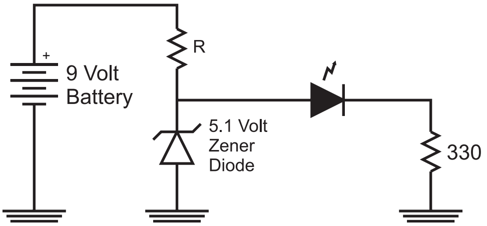

Fig. 12-11. Zener 5.1 volt regulated power supply to LED.

that extra current does not "leak out over the side" but is passed (or "shunted") through the diode. The diode itself is expected to be reverse biased when it is wired into the circuit and it will pass current through it to maintain a set voltage level at its anode (positive terminal). This property is known as "breakdown" and it is not unique to the Zener diode. All diodes will "breakdown" when a high enough reverse bias voltage is applied to them. The breakdown voltage for a Zener diode is usually specified to be in the range of 1.5–25 volts where the breakdown voltage for a typical diode (say the 1N4148/1N914 that I usually use) is 75–100 volts.

Specifying a Zener diode for use as a power supply in an application isn't very difficult but it will require you to understand what your incoming power specifications are as well as what the required current is for the circuit being powered. The powered circuit's voltage should be the same as the rating of the Zener diode. For 5 volt circuits, I use a Zener diode rated at 5.1 volts. Specifying the resistor that is to be used with the Zener diode as well as the Zener diode's power rating can be somewhat complex. Care must be taken to ensure that the circuit has enough current to be powered in all circumstances, including if the input power "sags" (if it is powered by a battery that is discharging). To do this, some kind of "margins" must be designed into the circuit.

For this experiment, I would like to use a 5.1 volt Zener diode to act as a power supply for a LED circuit requiring approximately 10 mA to light the LED. The circuit is shown in Fig. 12-11 and, before it can be assembled, the value for the Zener diode's current limiting resistor "R" must be determined. For a Zener diode power supply to be 100% efficient in terms of current (no current is shunted through the Zener diode), "R" must be chosen so that the voltage drop through it will allow the same amount of current as the powered circuit uses to pass through it. In this application, I am going to assume that the LED has a 2 volt drop, so using the basic electrical formulas, I can determine the current through the LED:

$$\begin{aligned} i &= V/R \\ &= (5.1\,V - 2\,V)/330\,\Omega \\ &= 9.39\text{ mA} \end{aligned}$$

Assuming that the battery produces an even 9 V, the value of R can be calculated:

$$\begin{aligned} R &= V/i \\ &= (9\,V - 5.1\,V)/9.39\text{ mA} \\ &= 415\,\Omega \end{aligned}$$

There are no standard 415 Ω resistors available, but I can make a 420 Ω resistor using a 200 Ω and a 220 Ω in series. This will result in a current of 9.29 mA (a difference of about 1% from the targeted value).

When I described the Zener diode regulator as acting like a basin of water in which the unused current was simply lost, I'm sure that many people grimaced because they knew of devices which are much better at regulating fluid pressure. If this book was written in the 1980s (or earlier) just about everybody would know about the commonly used fluid regulator that is used in older cars called a carburetor (Fig. 12-12). Virtually all cars built in the past 15 years have utilized some form of computer-controlled "fuel injection" which relies on active, rather than passive, control of the fuel being passed to the engine.

The carburetor is a very clever device that only provides fuel on demand. In Fig. 12-12, I have drawn the situation where no fuel is being drawn from the carburetor – a "float" is connected to a simple valve that closes when the fuel in the bowl that the float is in is full. When fuel is drawn from the bowl, the fuel level within the bowl drops (along with the float) and the valve opens, allowing more fuel into the bowl (Fig. 12-13). The carburetor is quite efficient and very simple in operation.

The carburetor acts as a regulator, just providing the volume of fuel (current) as required and the shallow bowl will result in lower pressure (pressure regulation) than what was available from the high-pressure source (the fuel pump). An electrical version of the carburetor would look

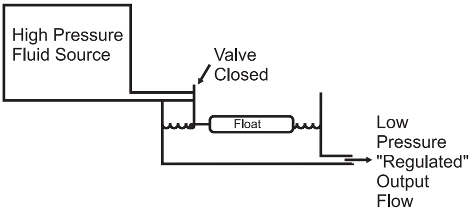

Fig. 12-12. Carburetor with no fluid being drawn from the bowl.

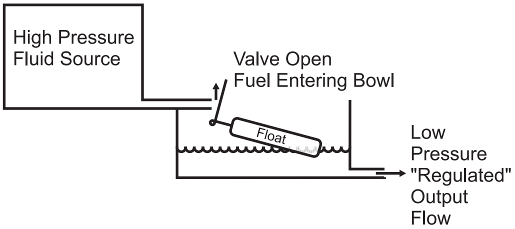

Fig. 12-13. Carburetor in operation.

like Fig. 12-14; current from the high voltage source is switched through a PNP bipolar transistor, with the control of the transistor being the output of the comparator. The comparator's inputs are the current voltage level of the regulator's output and the specific ‘‘output’’ voltage which comes from some kind of voltage reference. The voltage reference is usually a Zener diode that has a miniscule amount of current passing through it; the comparator does not need a lot of current to operate.

Adding the current and temperature ‘‘crowbar’’ sensors is implemented something like in Fig. 12-15. When either the current output or temperature exceeds the preset limits, the reference voltage is pulled to ground using an NPN transistor (remember that the voltage reference is very low current so this can be done safely). In some regulators, if the current or temperature parameters are exceeded, they ‘‘latch’’ the failing state until power is removed and the crowbar conditions are reset. The need for the current sense and shut down should be pretty obvious to you; if the current drawn exceeds the

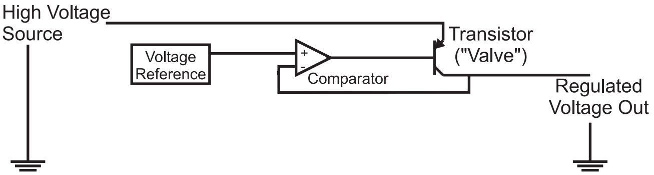
High Voltage Source
Voltage Reference
+
-
Comparator
Transistor ("Valve")
Regulated Voltage Out

Fig. 12-14. Linear voltage regulator block diagram.

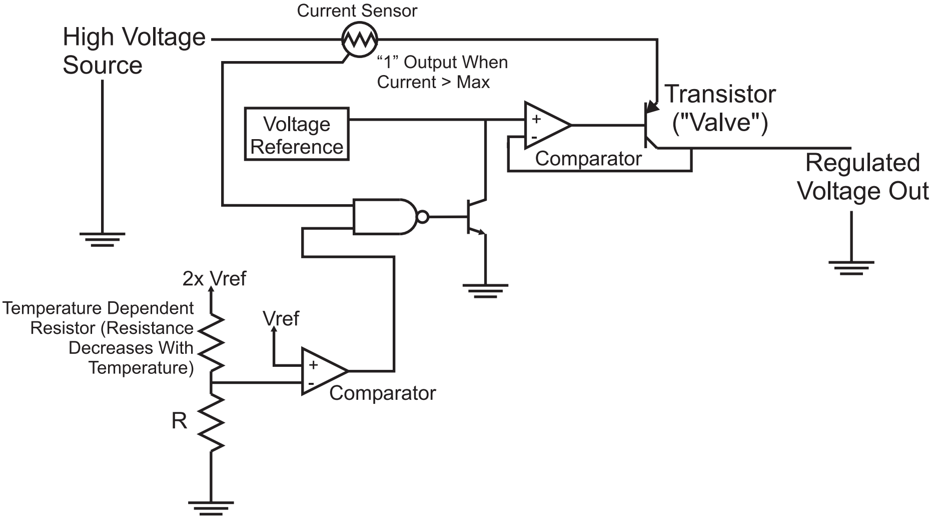
Current Sensor
High Voltage Source
“1” Output When Current > Max
Voltage Reference
+
-
Comparator
Transistor ("Valve")
Regulated Voltage Out
2x Vref
Temperature Dependent Resistor (Resistance Decreases With Temperature)
Vref
+
-
Comparator
R

Fig. 12-15. Complete linear voltage regulator circuit.

maximum rating for the PNP transistor, it could be damaged. The temperature sensor may be a bit more unexpected but shouldn't be surprising when you consider what is happening in the regulator when it is transforming a high voltage into a lower one. The difference between the input voltage and the regulated voltage multiplied by the current being drawn by the circuit being regulated is the power dissipated by the regulator. For example, if you had a 12 volt voltage source and a 5 volt regulator providing 200 mA current, the power being dissipated by the regulator would be 1.4 watts. This level of power dissipation could damage the internal circuitry of the regulator or, at the very least, raise the temperature of the part so that it does not work as designed.

The most popular linear voltage regulators that provide the crowbar features are the 78xx and 78Lxx series. The 78xx (or the LM2940 series of regulators which have the same pinout and package) shown in Fig. 12-16 ("xx" standing for the voltage, so a 5 volt regulator is a "7805") can normally source up to 500 mA and up to 1 A with heat sinking. The heat sink is used to dissipate the power and keep the temperature within the regulator less than 125°C, which is the crowbar temperature. For lower current applications (up to 100 mA), the 78Lxx (Fig. 12-17) can be used. For either device, the input voltage should be at least 2 volts above the regulated output voltage. When wiring the regulator in circuit, you should include at least 10 μF of capacitance on the input and a 0.1 μF capacitor on the output.

While the Zener diode and linear power supplies presented so far in this chapter are useful and easy to work with, they do have two concerns that can make them problematic when they are being used in a battery-powered application. First off, they require a higher voltage than the regulated output; this can be an issue when you want to use very simple power like two AA cells

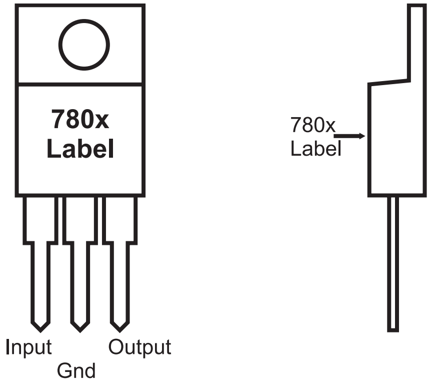

Fig. 12-16. 780x voltage regulator chip pinouts.

for a digital electronics circuit. Secondly, they are not terribly efficient. It isn't unusual for 80% or more of the power input to the Zener diode power supply to be lost and 40% or more lost in the linear power supply. What is required is a power supply circuit that is very efficient and will ''step up'' voltages.

While these two requirements seem impossible, they can actually be achieved very easily through the use of the ''switch mode power supply'' (SMPS). The basic SMPS circuit (Fig. 12-18) is quite simple and relies on the energy storing characteristic of the inductor or ''coil''. While the capacitor stores energy in the form of charge, the coil stores energy in the form of a magnetic field which is maintained by current running through the coil. When this current is shut off, the magnetic field produces a voltage ''spike'' (which I called ''kickback'' when discussing magnetic devices) that can be used as the basis for an output voltage.

Using the circled letters in Fig. 12-18, I have drawn the waveforms (Fig. 12-19) that you can expect to see in the SMPS. The ''Control'' signal is a PWM produced by a ''voltage controlled oscillator'' (VCO). A voltage controlled oscillator oscillates at a different frequency based on the voltage at an input. The input to the VCO used in the SMPS is the output voltage

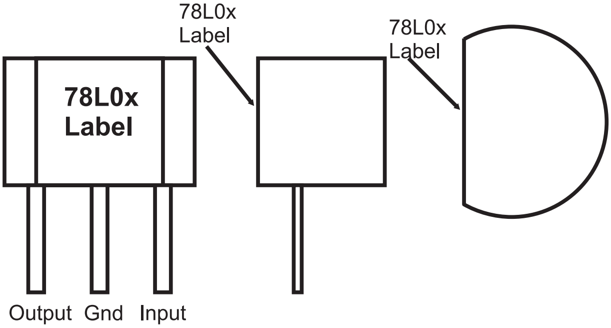

Fig. 12-17. 78L0x voltage regulator chip pinouts.

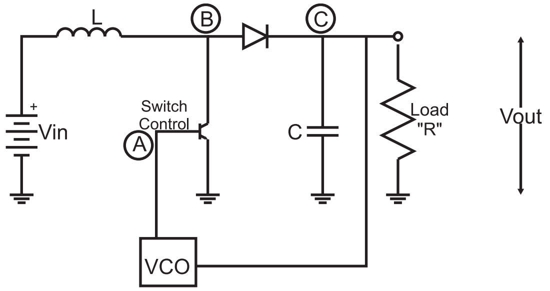

Fig. 12-18. Basic switch mode power supply (SMPS).

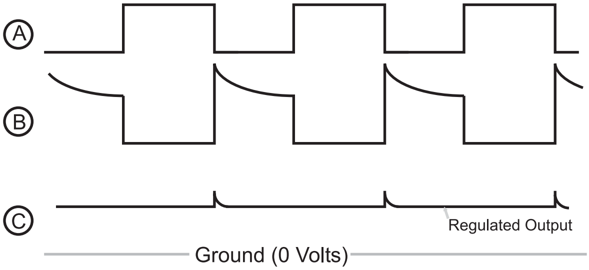

Fig. 12-19. Switch mode power supply operating waveforms.

of the power supply; the VCO frequency will change according to the power supply output to ensure the output stays as stable as possible at the required voltage. The output of the VCO is the base of a transistor that periodically pulls one side of the coil to ground, allowing current to flow through it. When the transistor connected to the coil is turned off, current flow through the coil stops and the magnetic field ‘‘kicks back’’, producing a higher voltage.

The operation of the VCO PWM output along with the coil’s response and the output voltage is shown in Fig. 12-19. When the VCO is turning on the transistor, the coil (symbol ‘‘L’’) is tied to ground and current flows through it. When the transistor is off, the coil kickback can be seen and any voltage greater than the current voltage output from the supply passes through the diode and is stored in the output capacitor. As I said above, if the output voltage is more or less than the target voltage, the VCO frequency changes along with the transistor control PWM, bringing the output voltage into line.

To determine the correct coil value as well as the PWM parameters, the following three formulas are used once the output voltage (‘‘V_{out}’’) is known along with the expected output current draw (‘‘I_{out}’’) and the input voltage (‘‘V_{in}’’). These formulas are used repeatedly until the values for ‘‘L’’ (the coil value), ‘‘T_{on}’’ (time the transistor is on) and ‘‘T_{off}’’ (time the transistor is off) are values that can be produced by reasonable hardware.

$$\begin{aligned} I_{peak} &= 2 \times I_{out} \times (V_{out}/V_{in}) \\ T_{off} &= L \times I_{peak}/(V_{out} - V_{in}) \\ T_{on} &= (V_{out}/V_{in}) - 1 \end{aligned}$$

Designing an SMPS is not a trivial exercise. While you may think you can do it using something like a 555 timer, I’m going to recommend that you use a commercially available chip that provides the function for you, like the LT1173-5. This chip can be used to create 5 volts (neccesary TTL and

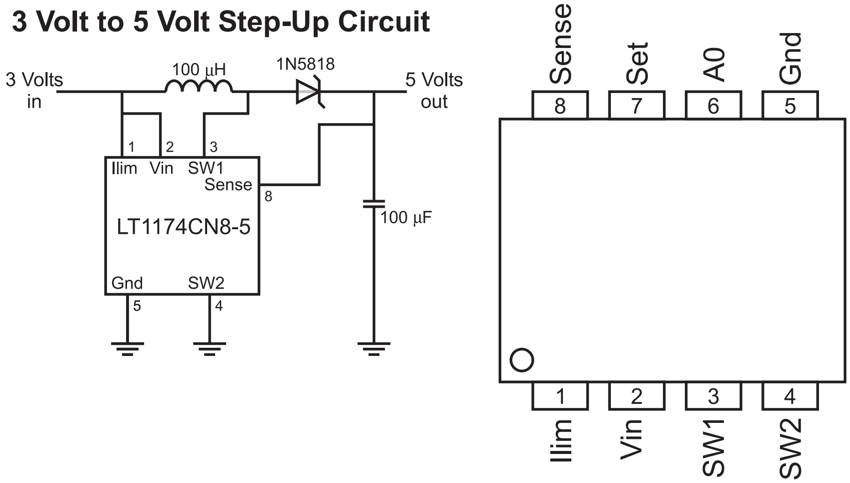

Fig. 12-20. LT1174CN8-5 switch mode power supply controller chip pinout and sample circuit.

many CMOS logic chips) from 3 volts, as shown in the basic circuit in Fig. 12-20.

With the appropriate regulator selected, you now have to find a source of DC current to power the application and regulator. By far the most popular way of providing power to an electronic device is by simply plugging it into a wall socket.

> I must caution you that the power coming out of your wall socket can conceivably destroy your application, cause a fire or hurt you (e.g. burns or electrocution). Despite the fact that it is commonly used for appliances, light and electronic devices in the home, electricity is not to be trifled with.
>
> The circuits provided below may not be appropriate for where you live. The information provided here is strictly ‘‘rule of thumb’’ and is primarily written for use in North America. If you are going to design a power supply for a specific country’s use, make sure you understand what are the characteristics of the local power supply, along with any laws or regulations that are appropriate when connecting to it.

Power coming from your wall sockets (‘‘the mains’’), comes in as either a 110 or 220 volts ‘‘peak-to-peak’’ as a ‘‘sine wave’’ with a frequency of 50 or 60 cycles per second (or ‘‘hertz’’ (‘‘Hz’’)). In North America, power is provided at 110–120 volts peak-to-peak voltage (typically 115 volts) at 60 Hz. Different countries around the world will use different peak-to-peak voltage levels and operating frequencies.

This power coming in is normally provided by a ‘‘socket’’, which is built into your walls. Figure 12-21 shows the layout of the socket and labels the individual connections. “Live” or “Hot” is the incoming alternating voltage sine wave shown in Fig. 12-21. “Neutral” is the return path for this current, while “Ground” is a shunt to “earth ground” if the circuit is damaged and the live voltage is passed to the neutral connection. If these three signals are being wired manually by convention, “Live” is black, “Neutral” is white and “Ground” is Green.

Because the AC voltage coming from the “mains” is so high and has positive and negative voltage components, it has to be converted into a lower DC voltage for the electronics. This is done in three stages. The first is reducing the voltage from more than 100 volts to 15 volts or less using a “transformer”. A transformer (Fig. 12-22) is a device made up of two coils that share their magnetic field. When current is passed through one coil, the second coil will produce an “inducted” voltage and current, which can be used to power the circuit. Figure 12-22 also gives the relationship between the voltage and current on the secondary side coil based on the number of turns for each coil.

Note that the current is inversely proportional to the turns ratio. In North America (which has 110 volts AC), an 8:1 transformer is often used. This means that with 110 volts in, there will be 14 volts out. For 220 volts, a 16:1 transformer should be used for the same voltage output.

While the voltage has been lowered, it is still “AC” and it is still going positive and negative. This voltage has to be “rectified” into a straight

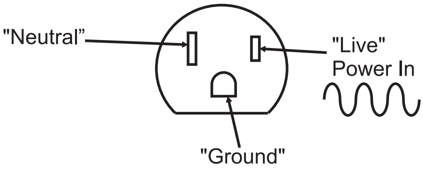

Fig. 12-21. North American wall plug.

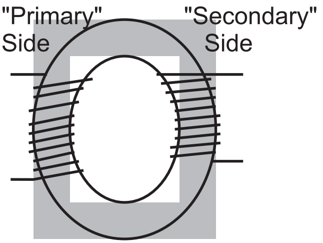

Electrical Characteristics:

$$\frac{\#\ \text{Turns Primary}}{\#\ \text{Turns Secondary}} = \frac{\text{Voltage Primary}}{\text{Voltage Secondary}} = \frac{\text{Secondary Current}}{\text{Primary Current}}$$

Fig. 12-22. AC voltage transformer operation.

DC voltage. This is done using diodes in either a "half wave" or "full wave" rectifier. Full wave rectifiers transform the positive and negative "lobes" of the AC circuit into a positive voltage, whereas the half wave rectifier "clips" the negative wave (providing half the total power available to the circuit). Inputting the rectified signal directly from the diodes into a voltage regulator should not be attempted; instead, a filtering electrolytic capacitor or a few tens of μF should be used. The filtered signal output from the full wave rectifier was shown in Chapter 3.

As long as the rectified signal does not drop below the minimum voltage of the Voltage Regulator, the regulated DC voltage output will be constant. The filtering cap should be a minimum of 10 μF with a good rule of thumb being that for digital circuits; a 20 μF capacitor is required for each Amp of current drawn. For DC electric motors, this value increases to 100 μF per amp drawn to help prevent inductive "kickback" "spikes" from being driven back through the transformer to the mains circuit.

Using the transformer, full wave rectifier, an electrolytic filter capacitor and a 7805 voltage regulator, a +5 volt 0.5 amp power supply for digital logic applications could be created, as shown in Fig. 12-23. The *voltage regulator* converts the rectified transformer-reduced AC voltage into a voltage that can be used by the digital logic.

There are a few things to note in Fig. 12-23. The first is that the mains ground is connected to the case and not to the "digital ground". In any DC-powered circuit, the negative terminal of the full wave rectifier can be called "digital ground", but should be left "floating" relative to "earth ground", which is provided by the AC plug. In this case, "digital ground" is simply a common negative terminal for the circuit. I have put a "fuse" in the power line, which will cut out in high current draw situations (like short circuits).

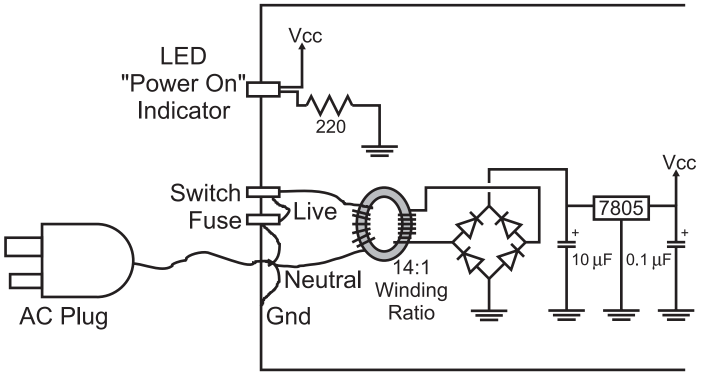

Fig. 12-23. Sample AC to 5 volt DC power supply for digital electronics.

It is rated at 0.1 amps, which may seem low, but remember that current output is inversely proportional to the turns ratio of the transformer: 0.1 amps at 110 volts translates into 1.4 amps at 8 volts at the output. Without this fuse, very large (and very dangerous) currents could build up inside the circuit. For example, 2 amps at 110 volts translates to 28 amps at 8 volts or 224 watts of power. Along with the fuse, the ‘‘Switch’’ in the circuit should be one that is certified for switching AC voltages. AC switches usually have a mechanical assembly inside them that ‘‘snaps’’ the switch contacts on and off. This minimizes ‘‘arcing’’ within the switch. This may seem hard to believe, but if you look inside an AC switch while it is opening or closing, you will see a blue spark and sometimes hear a ‘‘pop’’. This is caused by high inductive voltages produced by the transformer coils that ‘‘kickback’’ when the AC power is shut off.

If you do build mains power supply circuits, like this one, I recommend that you use 14-gauge stranded wire for all connections. Connections should consist of soldered connections (not household ‘‘Marette’’ connectors) for safety. ‘‘Heat shrink tubing’’ should be placed over all solder joints and bare wire. As well, only UL/CSA (or the local country testing organization) approved plugs, wires, switches, fuse holders and transformers should be used in a properly grounded metal case.

> If any of these terms are unfamiliar to you or you doubt your ability to build the circuit safely, then don’t build it!

Programmable Logic Devices

Programmable logic devices (‘‘PLDs’’) are chips which have logic gates and flip flops built in, but are not interconnected. The application designer will specify how the gates and flip flops are interconnected in order to create a portion of the application’s circuit. Most people feel that programmable logic devices are a relatively new invention, but they have been around for many years. It has only been quite recently (in the last 10 years or so) that reusable chip technology (i.e. EPROM and flash)-based PLDs have been available at prices hobbyists and small companies could afford.

There are several types of PLDs. The first is the simple array of logic gates and devices that are built of this type are known as ‘‘PALs’’ and ‘‘GALs’’ (I generically refer to them as ‘‘PALs’’). The chips themselves are quite simple and relatively easy to design circuits for. These circuits are normally arranged as a ‘‘sum of products’’ in which signals on the chip can be easily

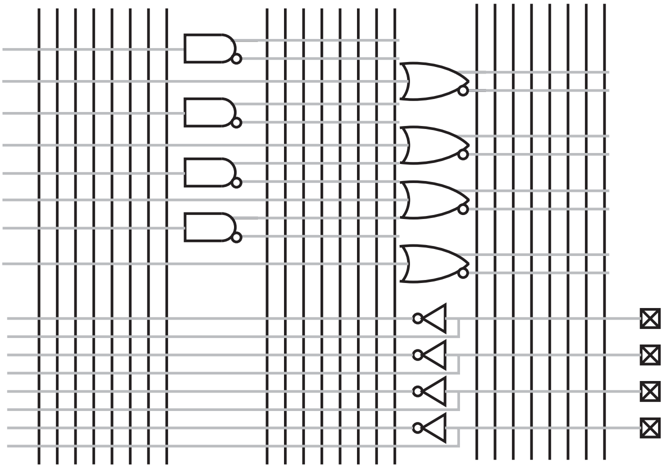

Fig. 12-24. Simple unprogrammed programmable logic device (PLD).

interconnected to form more complex logic functions. The chips are normally blocked out as a series of inputs and outputs, as shown in Fig. 12-24.

The vertical lines or "busses" in Fig. 12-24 are referenced to the gates and I/O pins they are connected to.

To form logic functions, the "sum of products" is used. In Fig. 12-24, a simple 4 I/O, 12 gate PAL is shown. Every output is driven on a bus in both positive as well as negative format. Connections are made between the gates and the busses to create logic functions.

For example, the "XOR" gate which is characterized by:

```
A XOR B = A ^ B
        = _A AND B OR A AND _B
        = (_A * B) + (A *_ B)
```

which is not often available in standard logic. Taking Fig. 12-24 and connecting the busses to the different I/O pins and gates within the PAL, I can implement the XOR gate, as shown in Fig. 12-25.

Note in Fig. 12-25, that an I/O pin changes from an input to an output by simply connecting it directly to a gate output. This feature allows the pins to be used as either input or output.

Options for PALs include varying numbers of inputs to the internal AND and OR gates. For the PLD shown in Fig. 12-25, I have left open the option that any of the pins can be used for any purpose. This is a bit unusual and, normally in PALs, the number of inputs to a gate is restricted. Another

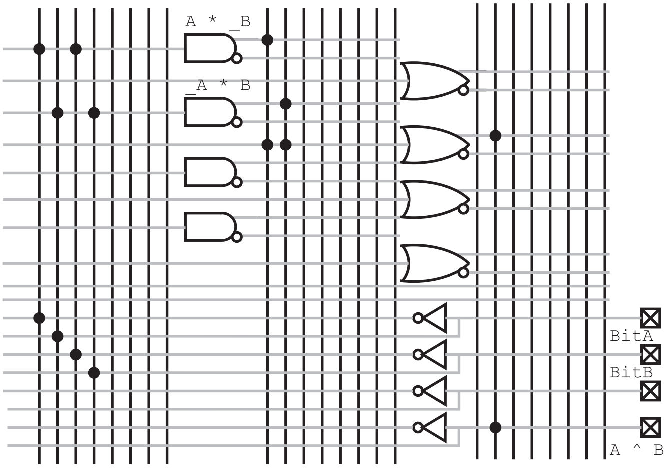

Fig. 12-25. PLD programmed as an XOR gate.

option is to include built in flip flops to store states and turn the PAL from a combinatorial circuit into a sequential one.

PALs may seem simple, but they can result in large decreases in the chip count for an application. In some cases, PALs may be more expensive than the chips they replace, but they reduce application power and board space chip requirements. These savings could result in all-over product savings. It is not unusual for 10 TTL chips to be replaced by a single PAL, resulting in huge PCB and power supply cost savings.

At the high end of the programmable logic device family range, some devices are virtually "ASICs" ("Application Specific Integrated Circuits") and use the same programming language ("VHDL") and development tools as ASICs. These complex parts generally have their functions broken up into "macros". An ASIC/PLD macro can be an AND, or XOR not, logic gate, flip flops or collections of functions (such as multiplexors and arithmetic logic units) which simplify the task of circuit development and eliminate the need for wiring individual gates into basic functions.

The high-end programmable logic device's programming information is often directly transferable to the technology. This allows initial production to use programmable logic devices that require little cost to program and, when the design is qualified, ASICs can be built at a chip foundry for reduced per unit costs.

Programmable logic devices have the advantage of being able to implement fast (less than 10 ns) logic switching, but they do not have the ability to store more than a few bits of data.

Often, programmable logic devices are used in proprietary circuits because their functions cannot be easily traced and decoded.

Programmable logic devices and ASIC development tools are generally function text based as opposed to graphically based applications (like a schematic drawing). This means that a text format, like the "XOR" definition above, must be used to define the functions. Most "compilers" for these statements are intelligent enough to pick the best gates within the device to work with and pick the best paths without your intervention. They are typically much more sophisticated (and expensive) than the compilers used for converting high-level program statements into instructions for a processor.

Quiz

1. The choices in processor design are:
   (a) Intel vs AMD vs PowerPC
   (b) CISC vs RISC, Princeton architecture vs Harvard architecture, hardcoded instruction execution vs state machine instruction execution
   (c) TTL vs CMOS logic
   (d) Speed vs minimal power consumption

2. Microcoded instructions are:
   (a) Short instructions which take less time to execute
   (b) Coded instructions that cannot be read by spies
   (c) Instructions that are specific to a microprocessor
   (d) State machine instructions outlining the steps needed to execute an instruction

3. Decimal 47.123 in binary is:
   (a) Invalid; you cannot perform this conversion
   (b) B'101111.00011111'
   (c) B'11111.101111'
   (d) 0x02F.1F7

4. Which statement is false? "Flash" memory cells:
   (a) Are designed from EPROM memory cells
   (b) Can be erased by applying an electrical voltage
   (c) Are built from flip flops
   (d) Are limited to 256 bits in size

5. DRAM Memory is:
   - (a) Faster than SRAM memory
   - (b) Less expensive per bit than SRAM memory
   - (c) More reliable than SRAM memory due to ‘‘refreshing’’
   - (d) More expensive per bit than SRAM memory
6. What is not a feature of the DC/DC power regulators presented in the book?
   - (a) They will convert AC to DC directly
   - (b) They have current limiting capability
   - (c) They have temperature limiting capability
   - (d) They have voltage ‘‘Brown out Reset’’ capability
7. With 12 volts coming in, the current limiting resistor for a 5 volt, 200 mA Zener diode:
   - (a) Insufficient information to calculate the resistor’s parameters
   - (b) 3.5 Ω, 10 watt
   - (c) 350 Ω, 1 watt
   - (d) 35 Ω, 2 watt
8. A switch mode regulator needs the following components to work:
   - (a) Capacitors, diodes and inductor
   - (b) Capacitor, diode and PWM driver
   - (c) Capacitors, diode, transistor, PWM driver and inductor
   - (d) Comparator, PWM driver, transistor and inductor
9. ‘‘PAL’’ is the acronym for:
   - (a) Pound and lever
   - (b) Peripheral aspect light
   - (c) Partial AND logic
   - (d) Programmable array logic
10. VHDL is used for:
    - (a) Defining PLD electrical parameters
    - (b) Defining PLD gate requirements
    - (c) Defining PLD gate operations
    - (d) Defining PLD speed parameters

CHAPTER 13

PC Interfacing Basics

From a practical point of view, chances are you will be designing an interface or enhancement to your PC. Most modern commercial devices utilize USB ports, but you can still do a lot of interesting projects with the ‘‘legacy’’ interfaces built into the PC. Along with this, a basic understanding of your PC will help you understand how commercial products are designed and may give you some ideas as to how you can design your own complex applications.

The PC ‘‘core’’ circuitry consists of the microprocessor, memory, the interrupt controller and a DMA controller, as shown in Fig. 13-1. This set of hardware can run any program or interface with hardware attached to this ‘‘local bus’’. While you may think of processor memory in terms of the megabytes that were advertised when you bought the PC, there are actually three different types of memory that are accessed.

The term ‘‘local memory’’ is kind of a loosely defined term that I use to describe memory on the PC’s motherboard and not on external cards or subsystems. There are a number of different kinds of memory used on the motherboard, each with a different set of characteristic features. The first type, ‘‘ROM’’ (‘‘Read Only Memory’’) is fairly slow and can only be read; it contains the PC start up (‘‘boot’’) code. Next, there is what I will call

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

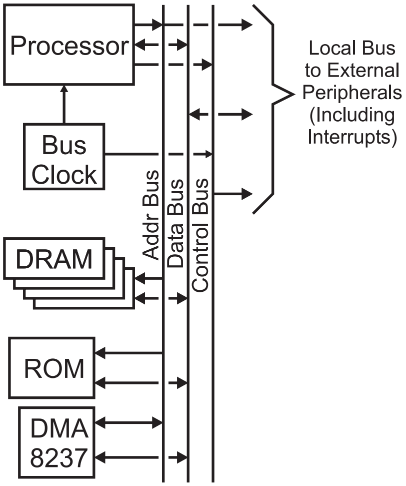
Processor
Bus Clock
DRAM
ROM
DMA 8237
Addr Bus
Data Bus
Control Bus
Local Bus to External Peripherals (Including Interrupts)

Fig. 13-1. PC major device block diagram.

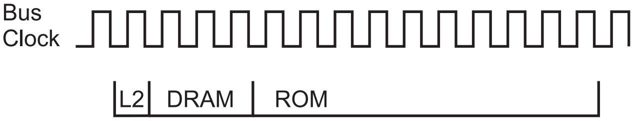
Bus Clock
L2
DRAM
ROM

Fig. 13-2. PC memory access time comparison.

“main memory”, which is measured in tens or hundreds of megabytes (and was prominently advertised when you bought your PC) and is moderately fast.

The PC’s processor itself has “cache memory” which runs at the processor’s speed and, ideally, all execution takes place from this area as this will help speed up the operation of the PC. Figure 13-2 shows the number of clock cycles for the different cases of the PC’s processor reading from cache memory (at full local bus clock speed), main memory (which has 50 ns access time DRAMs) and ROM (150 ns access time). The importance of being able to run entire applications from cache memory should be obvious to you.

“DMA” stands for “Direct Memory Access” and consists of a hardware device that can be programmed to create addressing and control signals to move data between devices within the system without involving the processor. DMA is most typically used in the PC for moving data to and from the disk drives.

Interrupts are hardware events, passed to the “8237 interrupt controller” chip that requests the PC’s processor to stop what it is currently doing and respond to the external request. This request does not have to be responded to immediately and it is up to the programmer's discretion to decide how to respond to the request.

All PC systems have multiple busses for system expansion and improved communications. Before going too much further, I should list the characteristics of what a computer bus is. A computer bus is defined as having:

1. A method for the controlling microprocessor to provide addresses bus hardware for both memory and I/O access.
2. The ability of the cards to request interrupt and DMA processing by the controlling microprocessor.
3. The ability to enhance the operation of the PC.

In a modern PC, there are usually three primary busses, each one delineated by the access speed they are capable of running at and the devices normally attached to them (see Fig. 13-3). The three busses have evolved over time to provide data rates consistent for the needs of the different devices. You might consider that there is a problem with cause and effect, but the busses and devices attached to them have sorted themselves out over the years.

The "Front Side Bus" (usually referred to by its acronym "FSB") runs at a very high speed specific to the processor. This is used by the processor to access DRAM memory directly as well as provide an interface to the system peripherals.

The "PCI" ("Personal Computer Interface") Bus. This bus is not only a staple of all modern PCs but is also available on many other system architectures as well. This allows PCI bus cards to be used across a number of

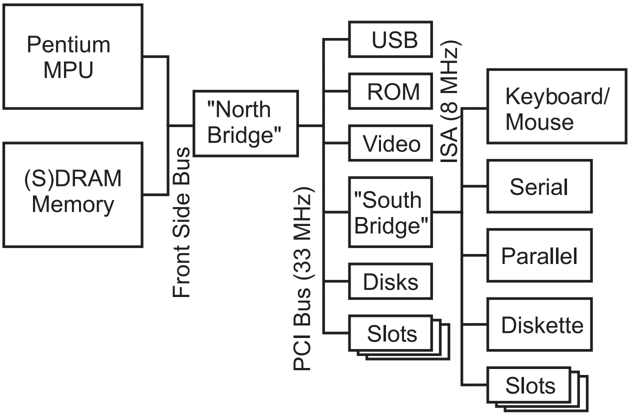

Fig. 13-3. Modern PC block diagram.

different systems and eliminates the need for designers and manufacturers to replicate their products for different platforms. PCI is somewhat of a "hybrid" bus, with some internal features of the PC (notably video and hard file controllers) using the PCI busses built into the copper traces on the motherboard as well as providing access to adapter cards in "slots". PCI is notoriously difficult to create expansion cards for. Along with the fairly high data rate speed (33 and 66 MHz), there are data transfer protocols which generally require an ASIC to decode and process bus requests.

The last bus interfaces the "legacy" interfaces of the PC together and is known as the "Industry Standard Architecture" ("ISA") bus. This bus typically has a data transfer speed of 125 ns – the same as the original PC/AT. When you are interfacing digital electronic devices to your PC, chances are the interface will be connected to this bus.

The Parallel (Printer) Port

When hardware is to be interfaced with the PC, often the first method chosen is the "Parallel" (Printer) port. If I was being introduced to the PC for the first time, I would probably look at this method first as well, but as I have learned more about the PC, using the parallel port would actually be one of the *last* methods that I would look at. The parallel port is really the most difficult interface in the PC to use because it is really device (printer) specific, has a limited number of I/O pins and is very difficult to time operations with accuracy. The only reason why I would see applications using a parallel port for interfacing is because, electronically, it does not need a level translator and can connect directly to TTL/CMOS logic.

In the 1970s, the Centronics Corporation developed a 25 pin "D-Shell" connector standard for wiring a computer to their printers. The early printer interfaces in the PC, after sending a byte, would wait for a handshaking line to indicate that the printer was ready for the next character before sending the next one. As you could imagine, this method was very slow and took up all the PC's cycles in printing a file.

This "pinout" very quickly became a standard in the computer industry. This connector consisted of an eight-pin data output port with strobe, four control lines passed to the printer and five status lines from the printer, as shown in Fig. 13-4. By the time the PC was being developed, this interface was the de facto industry standard and was chosen by IBM to help make the PC an "Industry" standard. When the printer port for the PC was being

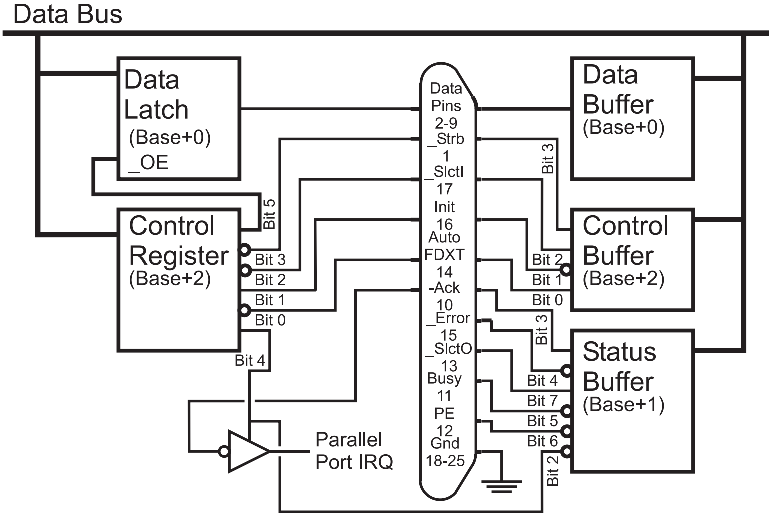
Data Bus
Data Latch
(Base+0)
_OE
Control Register
(Base+2)
Bit 5
Bit 3
Bit 2
Bit 1
Bit 0
Bit 4
Parallel Port IRQ
Data Pins 2-9
_Strb 1
_Slctl 17
Init 16
Auto FDXT 14
-Ack 10
_Error 15
_SlctO 13
Busy 11
PE 12
Gnd 18-25
Data Buffer
(Base+0)
Bit 3
Control Buffer
(Base+2)
Bit 2
Bit 1
Bit 0
Bit 3
Status Buffer
(Base+1)
Bit 4
Bit 7
Bit 5
Bit 6
Bit 2

Fig. 13-4. PC parallel port wiring and pinout.

developed, little thought was given to enhancing the port beyond the then current standard level of functionality.

To improve the functionality of the printer port, IBM, when designing the PC/AT, changed the eight data bit output only circuitry to allow data bits to be output as well as read back. This was accomplished by using pulled up open collector outputs that were passed from the ‘‘Data Latch’’ of Fig. 13-4, to the 25 pin D-Shell connector and on to the ‘‘Data Buffer’’, which could read the eight bits and see if anything was changed.

When a read of the eight data bits takes place, the eight output bits should all be ‘‘High’’ to allow the parallel port’s pull ups and the open collector outputs of the peripheral device to change the state of the pins. The bi-directional data capability in the PC/AT’s parallel port allowed hardware to interface to the PC to be very easily designed. Further enhancing the ease in which devices could be designed to work with the parallel port was the standard timing provided by the PC/AT’s ISA bus and BIOS.

Surprisingly, given this background, the biggest problem today with interfacing to the parallel port is timing signals properly. Depending on the different PCs that I have experimented with, I have seen parallel I/O port read and write timings from about 700 ns to less than 100 ns. When you design your interface to the IBM PC, you can either ignore the actual timings of the parallel port and execute signals that are many milliseconds long, ‘‘tune’’ the code to the specific PC you are working with or add a time delay function, like a 555 wired as a monostable that the PC can poll from the printer port to properly time its I/O operations.

There are three registers that you will have to access from your PC program to control the parallel port. The "Control Register", provides an operating control to the external device as well as to the internal interrupt function of the parallel port. The "Status Buffer" is read to poll the status bits from the printer. The last register, called "Data", allows eight bits to be written or read back from the parallel port. Starting at a parallel port (known as "LPT" in the PC), the registers are addressed as shown in Table 13-1 relative to a "Base" address which is at 0x0378 and 0x0278 for "LPT1" and "LPT2", respectively.

When the parallel port passes data to a printer, the I/O pins create the basic waveform shown in Fig. 13-5. It is important to note that the Printer BIOS routines will not send a new character to the printer until it is no longer "Busy". When "Busy" is no longer active, the "Ack" line is pulsed active, which can be used to request an interrupt to indicate to data output code that

Table 13-1 PC LPT port registers.

Register	Address	Function
Data	Base + 0	Read/Write 8 Data Bits
Status	Base + 1	Poll the Printer/Peripheral Status
Control	Base + 2	Write Control Bits to Printer/Peripheral and LPT Port Hardware

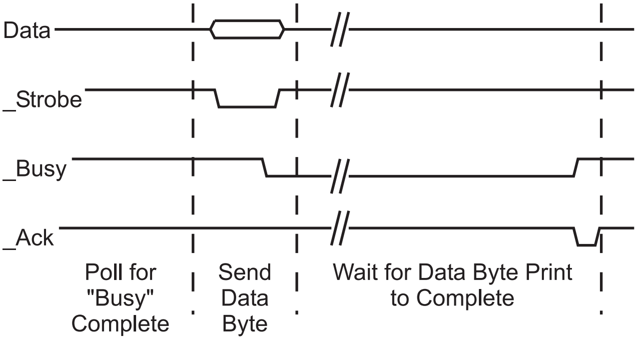

Fig. 13-5. PC parallel port write operation waveform.

the printer is ready to accept another character. The timing of the circuit for printer applications is quite simple, with 0.5 μs minimum delays needed between edges on the waveforms in Fig. 13-5.

Video Output

The first computer CRT display was the "vector" display in which the "X-Y" deflection plates moved the electron beam to a desired location on the screen and then drew a line to the next location. This was repeated until the entire image was drawn on the screen (at which time the process started over). This was popular in early computers because only a modest amount of processing power and video output hardware was needed to draw simple graphics. Because of the way the vector displays operated, a complex image was often darker than and could "flash" more than a "raster" (TV-like) display because a simple image would require fewer vector "strokes" and could be refreshed more often. Vector displays enjoyed some popularity in early computer video displays (including video games), but really haven't been used at all for over 10 years.

Today, a more popular method of outputting data from a computer is to use a "raster" display, in which an electron beam is drawn across the cathode ray tube in a regular, left to right, up and down, pattern. When some detail of the image is to be drawn on the screen, the intensity of the beam is increased when the beam passes a particular location on the screen, which causes the phosphors to glow more brightly, as shown in Fig. 13-6. If you have an old

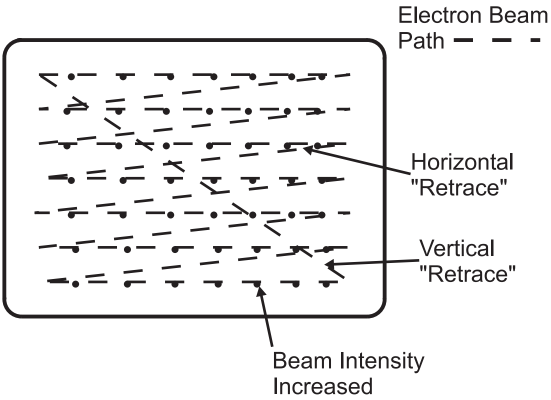

Fig. 13-6. Video raster.

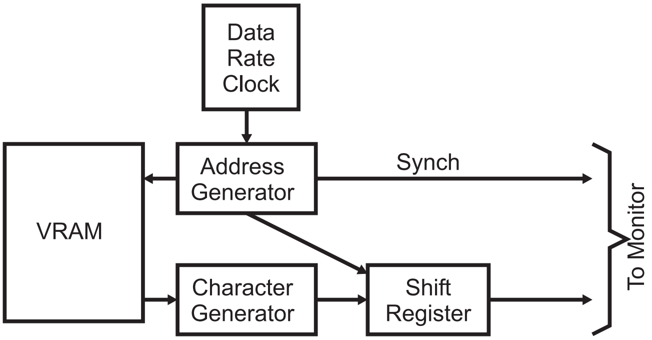

Fig. 13-7. Basic computer video driver circuit.

black and white TV (or monochrome computer monitor), chances are you can increase the “brightness”/“contrast” controls on a dark signal to see the different features I’ve shown above.

All raster computer output hardware consists of a shift register fed by data from a “Video RAM” (“VRAM”). In Fig. 13-7, the data to be output is read from the VRAM and then passed to a character generator, which converts the data from the Video RAM into a series of dots appropriate for the character. This series of dots is then shifted out to the hardware that drives the raster on the video monitor. If graphical data is output, then the “Character Generator” is not needed in the circuit and the output from the VRAM is passed directly to the shift register. The shift register may also be connected to a “DAC” (“Digital Analog Converter”), which converts the digital data into a series of analog voltages that display different intensities of color on the display.

The addresses for each byte’s data to be transferred to the display are accomplished by using an “Address Generator”, which controls the operation of the display (Fig. 13-8). The Address Generator divides the “data rate” (the number of “pixels” displayed per second on the screen) into character-sized “chunks” for shifting out, resets address counters when the end of the line is encountered and also outputs “Synch” information for the monitor to properly display data on the screen.

The circuit in Fig. 13-8 shows an example address generator for an 80 character by 25-row display. This circuit resets the counters when the end of the line and “field” (end of the display) is reached. At the start of the line and “field”, horizontal and vertical synch information is sent to the monitor so the new field will line up with the previous one.

I have presented this information on standard video inputs to give you some background on how “NTSC” televisions work and what their input

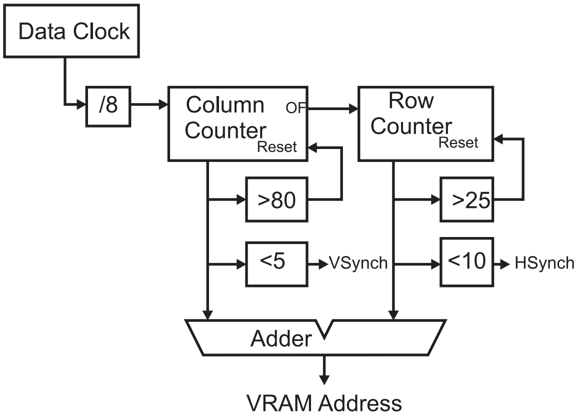
Data Clock
/8
Column Counter
OF
Reset
Row Counter
Reset
>80
>25
<5
VSynch
<10
HSynch
Adder
VRAM Address

Fig. 13-8. Video RAM address generator.

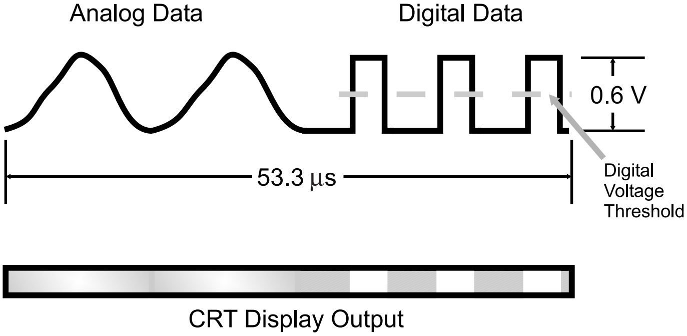
Analog Data
Digital Data
0.6 V
53.3 µs
Digital Voltage Threshold
CRT Display Output

Fig. 13-9. Analog and digital video data output.

signals look like. "NTSC" is the television "composite video" standard for North America and is an acronym for the "National Television Standards Council", the governing body put in place for television in the United States in the 1940s (today, television broadcasting technical standards are controlled by the Federal Communications Commission or FCC).

"Composite NTSC" is the signal typically received by your TV set and consists of both the data input streams along with the formatting information (called synchronization or synch pulses) that are used by the TV set to recognize where the incoming signal is to be placed on the screen. The incoming data, shown in Fig. 13-9 consists of voltage levels (from 0 to 0.6 volts) varying over time. This data is wiped across the CRT from left to right as a single line raster.

On the left-hand side of the incoming serial data shown in Fig. 13-9, I have drawn the data as a typical, analog video stream. The analog voltage levels are used to control the output level of the electron ‘‘gun’’ in the CRT: the higher the input voltage level, the brighter the output on the current raster of the CRT display, as shown at the bottom of the diagram. To the right of the analog data of Fig. 13-9, I have drawn some signals that jump from 0 volts to 0.6 volts and back again. This section of the raster is labeled as ‘‘Digital Data’’. I call this digital data because it is either all on or all off and it can be read as data by simply comparing whether the data is above or below the ‘‘threshold’’ marked on Fig. 13-9.

The start of each new line is specified by the inclusion of a ‘‘horizontal synch pulse’’ (‘‘hsynch’’), as shown in Fig. 13-10. This pulse is differentiated from the serial data by being at a voltage level less than the ‘‘black’’ level – this is where the term ‘‘blacker than black’’ comes from. Before and after the synch pulse, there are two slightly negative voltage regions called the ‘‘front porch’’ and ‘‘back porch’’, respectively. The front porch is 1.4 µs long, the back porch and horizontal synch pulse are 4.4 µs each. The entire line is 63.5 µs in length, and when you take away the 10.2 µs of horizontal synch information, you are left with 53.3 µs per line for the analog or digital data.

There are 262 or 263 lines of data sent to the TV, with the first 12 normally being at the −0.08 volt ‘‘porch’’ level and called the ‘‘vertical blanking interval’’. I say it is normally at the −0.08 volt porch level, except when ‘‘closed captioning’’ information is available. Closed captioning, as I’m sure you are aware, is the text that comes on the screen to allow hearing impaired

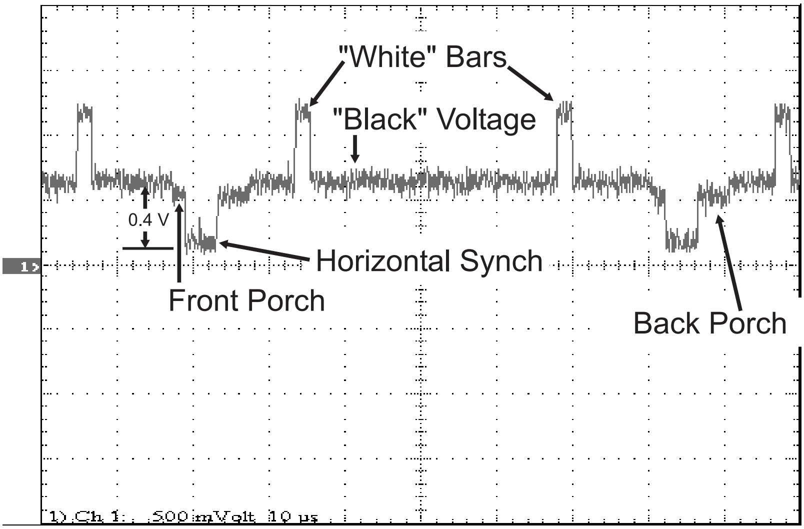

Fig. 13-10. Full NTSC composite video line.

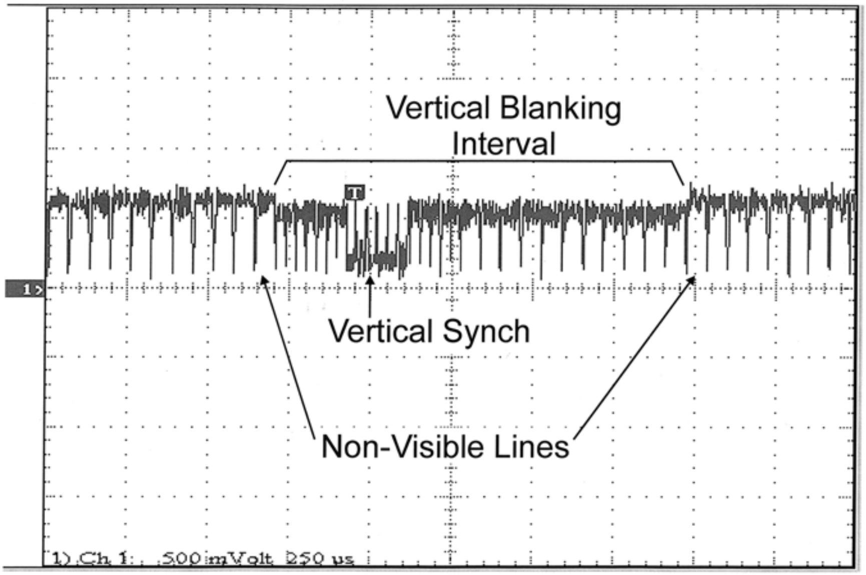

Fig. 13-11. Composite NTSC video vertical synch.

people to watch TV. When closed captioning data is included in the vertical blanking interval, digital pulses, similar to the ones shown in Fig. 13-11 are built into each line and decoded by circuitry built into the TV.

The TV set does not have a counter built into it to indicate when the CRT's gun should reset and go back to the top of the screen. Instead the incoming signal indicates when a vertical retrace must take place by its inclusion of a ‘‘vertical synchronization pulse’’ (‘‘vsynch’’). The oscilloscope picture (Fig. 13-11) shows what the vertical synch pulse looks like: it is actually a series of half-sized lines followed by a number of inverted half-sized lines.

So far, I have described NTSC video as having 262 or 263 lines, but if you look at the specifications for a TV set, you will see that it can accept 525 lines of NTSC data. A full NTSC screen consists of two ‘‘fields’’ of data: one 262 lines in size and the other 263 lines. The fields are offset by one half the width of a line – after the first field is displayed, the second fills in the spaces between the lines displayed by the first field. This is called ‘‘interlacing’’ the two fields together into a single ‘‘frame’’.

With the information contained within this section, you have enough information to create a monochrome video output circuit. I've done it a number of times with basic microcontrollers; as long as you keep your field width, vertical and horizontal synch pulses timed consistently, you will find that you can build your own composite video output display that you can send to an old TV. Even though with the information contained here, you should be able to design an NTSC composite video output perfectly, I still don't recommend trying it out on the family's good TV.

Synchronous Serial Interfaces (SPI, I2C, Microwire)

There are two very common synchronous data protocols, Microwire and SPI, from which a variety of standard devices (memory and peripheral functions) have been designed around, including serial EEPROMs, sensors and other I/O functions. I consider these protocols to be methods of transferring synchronous serial data rather than intelligent network protocols because each device is individually addressed (even though the clock/data lines can be common between multiple devices). If the chip select for the device is not asserted, the device ignores the clock and data lines. With these protocols, only a single "Master" can be on the bus. Normally, just eight bits of data are sent out at a time. For protocols like Microwire where a "Start bit" is initially sent, the "Start bit" is sent using direct reads and writes to the I/O pins. To receive data, a similar circuit would be used, but data would be shifted into the shift register and then read by the microcontroller.

The Microwire protocol is capable of transferring data at up to one megabit per second. Sixteen bits are only transferred at a time. After selecting a chip and sending a start bit, the clock strobes out an eight bit command byte, followed by (optionally) a 16 bit address word transmitted and then another 16 bit word either written or read by the microcontroller.

With a one megabit per second maximum speed, the clock is both high and low for 500 ns. Transmitted bits should be sent 100 ns before the rising edge of the clock. When reading a bit, it should be checked 100 ns before the falling edge of the clock. While these timings will work for most devices, you should make sure you understand the requirements of the device being interfaced to.

The SPI protocol is similar to Microwire, but with a few differences:

1. SPI is capable of up to 3 megabits per second data transfer rate.
2. The SPI data "word" size is eight bits.
3. SPI has a "Hold" which allows the transmitter to suspend data transfer.

4. Data in SPI can be transferred as multiple bytes known as ‘‘Blocks’’ or ‘‘Pages’’.

Like Microwire, SPI first sends a byte instruction to the receiving device. After the byte is sent, a 16 bit address is optionally sent followed by eight bits of I/O. As noted above, SPI does allow for multiple byte transfers. An SPI data transfer is shown in Fig. 13-12. The SPI clock is symmetrical (an equal low and high time). Output data should be available at least 30 ns before the clock line goes high and read 30 ns before the falling edge of the clock.

When wiring up a Microwire or SPI device, one trick that you can do to simplify the microcontroller connection is to combine the ‘‘DI’’ and ‘‘DO’’ lines into one pin, as was shown earlier in the book.

The most popular form of microcontroller network is ‘‘I2C’’, which stands for ‘‘Inter-Intercomputer Communications’’. This standard was originally developed by Philips in the late 1970s as a method to provide an interface between microprocessors and peripheral devices without wiring full address, data and control busses between devices. I2C also allows sharing of network resources between processors (which is known as ‘‘Multi-Mastering’’). Your PC probably has several I2C busses built into it for controlling system control and monitoring functions.

The I2C bus consists of two lines: a clock line (‘‘SCL’’) which is used to strobe data (from the ‘‘SDA’’ line) from or to the master that currently has control over the bus. Both these bus lines are pulled up (to allow multiple devices to drive them). A I2C controlled stereo system might be wired as shown in Fig. 13-13.

The two bus lines are used to indicate that a data transmission is about to begin as well as pass the data on the bus. To begin a data transfer, a ‘‘Master’’ puts a ‘‘Start Condition’’ on the bus. Normally, when the bus is in

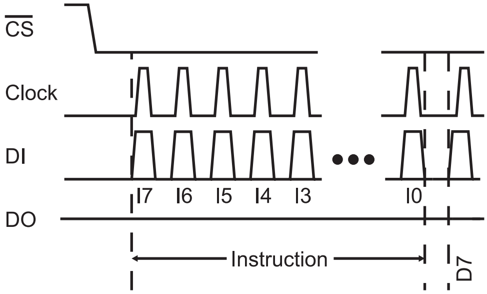

Fig. 13-12. SPI data transfer waveform.

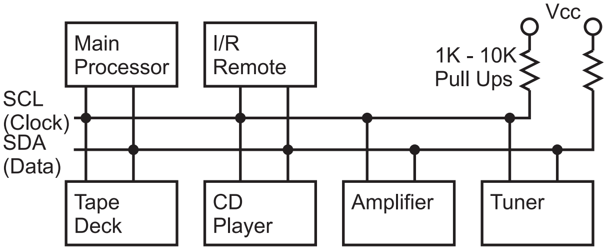
Main Processor
I/R Remote
Vcc
1K - 10K Pull Ups
SCL (Clock)
SDA (Data)
Tape Deck
CD Player
Amplifier
Tuner

Fig. 13-13. Example I2C application wiring.

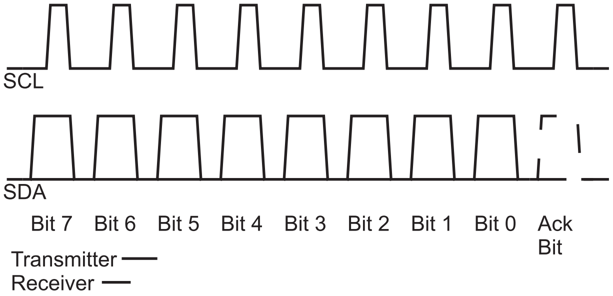
SCL
SDA
Bit 7
Bit 6
Bit 5
Bit 4
Bit 3
Bit 2
Bit 1
Bit 0
Ack Bit
Transmitter
Receiver

Fig. 13-14. I2C data packet waveform.

the "Idle State", both the clock and data lines are not being driven (and are pulled high). To initiate a data transfer, the master requesting the bus pulls down the SDA bus line followed by the SCL bus line. During data transmission this is an invalid condition (because the data line is changing while the clock line is active/high). Each bit is then transmitted to or from the "Slave" (the device the message is being communicated with by the "Master") with the negative clock edge being used to latch in the data. To end data transmission, the reverse is executed; the clock line is allowed to go high, which is followed by the data line.

Data is transmitted in a synchronous (clocked) fashion. The most significant bit is sent first and after eight bits are sent, the master allows the data line to float (it doesn't drive it low) while strobing the clock to allow the receiving device to pull the data line low as an acknowledgment that the data was received. After the acknowledge bit, both the clock and data lines are pulled low in preparation for the next byte to be transmitted or a Stop/Start Condition is put on the bus. Fig. 13-14 shows the data waveform.

Sometimes, the acknowledge bit will be allowed to float high, even though the data transfer has been completed successfully. This is done to indicate that the data transfer has completed and the receiver (which is usually a "slave device" or a "Master" which is unable to initiate data transfer) can prepare for the next data request.

There are two maximum speeds for I2C (because the clock is produced by a master, there really is no minimum speed). "Standard Mode" runs at up to 100 kbps and "Fast Mode" can transfer data at up to 400 kbps.

Asynchronous Serial Interfaces

Understanding how asynchronous serial communications works is easy; implementing them on a PC is frustrating and hard and used to be especially true for casual PC owners. This situation is improving with newer high-speed interface connection protocols (USB and Firewire), but actually sitting down and getting RS-232 communication between two devices working can be a pain. Much of the information I will present on RS-232 in this section will seem quite obvious, but I urge you to read through this material as there are some tricks and information contained here that could probably get you out of a bind or two.

Asynchronous long distance communications came about as a result of the Baudot "teletype". This device mechanically (and, later, electronically) sent a string of electrical signals (which we would call "bits"), like the one shown in Fig. 13-15, to a receiving printer. This data packet format is still used today for the electrical asynchronous transmission protocols described in the following sections. With the invention of the teletype, data could be sent and retrieved automatically without having an operator having to sit by the teletype all night unless an urgent message was expected.

The Baud Rate specified for the data transmission is the maximum number of possible data bit transitions per second (measured in "bits per second", abbreviated to "bps"). This includes the "Start", "Parity" and "Stop" bits at the ends of the data "packet" shown in Fig. 13-15 as well as the five data bits

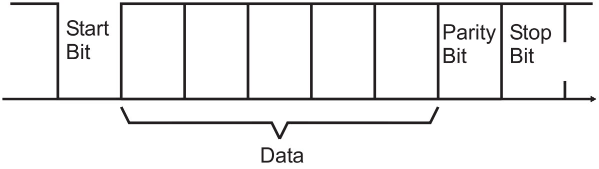

Fig. 13-15. Five-bit Baudot data packet.

in the middle. I use the term "packet" because we are including more than just data (there is also some additional information in there as well), so "character" or "byte" (if there were eight bits of data) are not appropriate terms. This means that for every five data bits transmitted, eight bits in total are transmitted (which means that nearly 40% of the data transmission bandwidth is lost in teletype asynchronous serial communications).

The "Data Rate" is the number of data bits that are transmitted per second. For this example, if you were transmitting at 110 baud, the actual data rate is 68.75 bits per second (or, assuming five bits per character, 13.75 characters per second).

With only five data bits, the Baudot code could only transmit up to 32 distinct characters. To handle a complete character set, a specific five-digit code was used to notify the receiving teletype that the next five-bit character would be an extended character. With the alphabet and most common punctuation characters in the "primary" 32, this second data packet wasn't required very often. In the data packet shown in Fig. 13-15, there are three control bits. The "Start Bit" is used to synchronize the receiver to the incoming data. In most hardware circuits designed to read an asynchronous serial packet, there is an overspeed clock running at 16 times the incoming bit speed which samples the incoming data and verifies whether or not the data is valid. When waiting for a character, the receiver hardware polls the line repeatedly at 1/16 bit period intervals until a "0" ("Space") is detected (down arrow at the left of Fig. 13-16). The receiver then waits half a cycle before

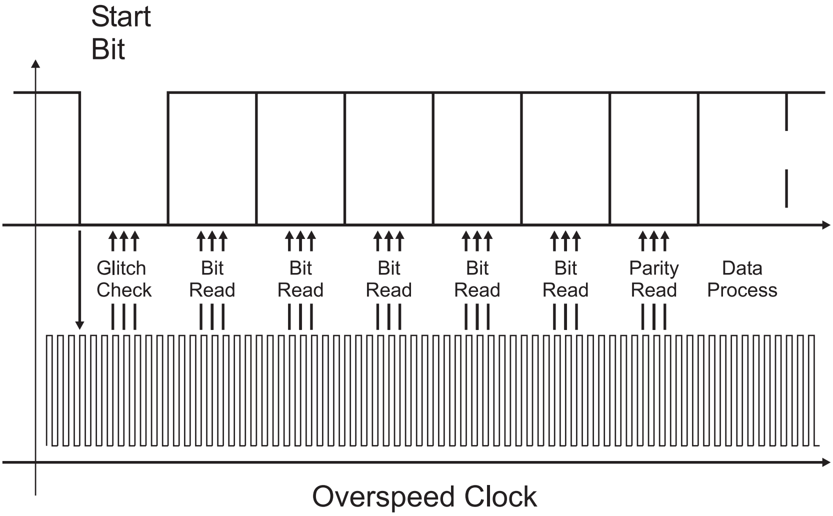

Fig. 13-16. Asynchronous serial data read.

polling the line again to see if a ‘‘glitch’’ was detected and not a Start bit. This polling takes place in the middle of each bit to avoid problems with bit transitions (or if the transmitter’s clock is slightly different from the receivers, the chance of misreading a bit will be minimized).

Once the Start bit is validated, the receiver hardware polls the incoming data once every bit period multiple times (again to ensure that glitches are not read as incorrect data).

The ‘‘Stop’’ bit was originally provided to give both the receiver and the transmitter some time before the next packet is transferred (in early computers, the serial data stream was created and processed by the computers and not custom hardware, as in modern computers).

The ‘‘Parity’’ bit is a crude method of error detection that was first brought in with teletypes. The purpose of the parity bit is to indicate whether the data was received correctly. An ‘‘odd’’ parity meant that if all the data bits and parity bits set to a ‘‘Mark’’ were counted, then the result would be an odd number. ‘‘Even’’ parity is checking all the data and parity bits and seeing if the number of ‘‘Mark’’ bits is an odd number. Along with even and odd parity, there are ‘‘Mark’’, ‘‘Space’’ and ‘‘No’’ parity. ‘‘Mark’’ parity means that the parity bit is always set to a ‘‘1’’, ‘‘Space’’ parity is always having a ‘‘0’’ for the parity bit and ‘‘No’’ parity is eliminating the parity bit all together. I said that parity bits are a ‘‘crude’’ form of error detection because they can only detect one bit error (i.e. if two bits are in error, the parity check will not detect the problem). If you are working in a high induced noise environment, you may want to consider using a data protocol that can detect (and, ideally, correct) multiple bit errors.

The most common form of asynchronous serial data packet is ‘‘8-N-1’’, which means eight data bits, no parity and one stop bit. This reflects the capabilities of modern computers to handle the maximum amount of data with the minimum amount of overhead and with a very high degree of confidence that the data will be correct.

Having reviewed the data protocol for asynchronous serial communications, let’s go on and look at the electrical specifications, starting with the connectors that you will find on the back of your PC (Fig. 13-17). Either a male 25 pin or male 9 pin connector is available on the back of the PC for each serial port – chances are you will have a 9 pin connector because it takes up the least amount of space.

Working with MS-DOS in early systems, only a maximum of four serial ports could be addressed by the PC, and of these probably only two were useable for connecting external devices to the PC due to conflicts with other hardware devices. In modern systems, which have ‘‘Plug and Play’’ capabilities and the Windows operating system, which can allocate resources

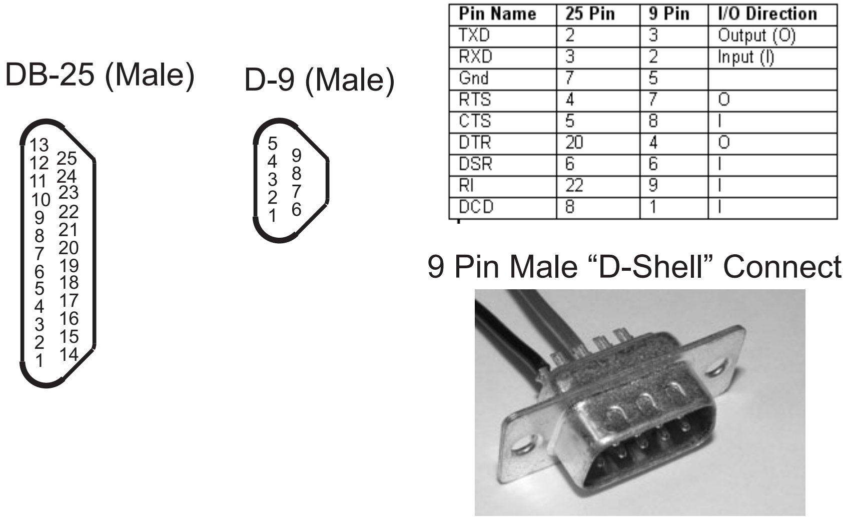

Pin Name	25 Pin	9 Pin	I/O Direction
TXD	2	3	Output (O)
RXD	3	2	Input (I)
Gnd	7	5	
RTS	4	7	O
CTS	5	8	I
DTR	20	4	O
DSR	6	6	I
RI	22	9	I
DCD	8	1	I

Fig. 13-17. PC asynchronous serial port connectors and pinouts.

throughout the system, the number of standard serial ports has been greatly expanded. My personal record is 64 serial devices for environmental chamber status test software.

The standard RS-232 data rates available to you in the PC are listed in Table 13-2. As an interesting exercise, I suggest that you find the reciprocal of the data rates listed in Table 13-2, multiply by 1,000,000 and then divide by 13 – what you will find is that the results for the data rates starting at 1200 bps and going higher will have a very small fraction. In fact, the error will be much less than 0.1%! This is a good trick to keep in your hip pocket when you have to implement an RS-232 interface and you don't have any crystals that have been cut specifically to provide a multiple of these data rates.

Asynchronous communications based on the Baudot teletype/RS-232 is known as "Non-Return to Zero" ("NRZ") asynchronous communications, because at the end of each data packet the serial line is high. There are other methods of sending asynchronous serial data, with one of the most popular being "Manchester encoding". In this type of data transfer, each bit is synchronized to a "start" bit and the following "Data Bits" are read with the "space" dependent on the value of the bit.

Manchester encoding is unique in that the "Start Bit" of a packet is quantitatively different from a "1" or a "0" (shown in Fig. 13-18). This allows a receiver to determine whether or not the data packet being received is actually at the start of the packet or somewhere in the middle (and should be ignored until a start bit is encountered).

Table 13-2 Standard RS-232 data rates.

Data rate
110 bps
150 bps
300 bps
600 bps
1200 bps
2400 bps
4800 bps
9600 bps
19200 bps
38400 bps
57600 bps
115200 bps

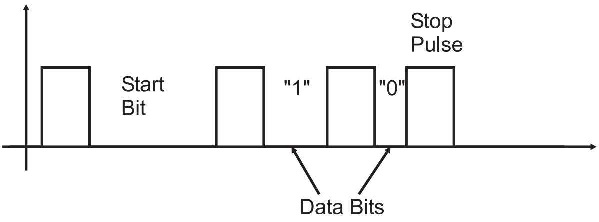

Fig. 13-18. Manchester data encoding.

Manchester encoding is well suited for situations where data can be easily interrupted or there is a conflict in the middle of data reception. Because of this, it is the primary method of data transmission for infrared control (such as used in your TV's remote control).

RS-232 Electrical Standards

In the previous section, I didn't tell the whole story about RS-232 asynchronous serial communications. I left out one very important point – signals do not travel at the same voltage levels as what we've discussed so far in the book. When RS-232 was first developed into a standard, computers and the electronics that drive them were still very primitive and unreliable. Because of that, we've got a couple of legacies to deal with and this can complicate connecting digital electronics circuits to another device using RS-232.

The biggest concern is the voltage levels of the data. A "Mark" ("1") is specified to be -3 volts to -12 volts and a "Space" ("0") is $+3$ volts to $+12$ volts. This means that there must be some kind of voltage level conversion when passing RS-232 to or from digital electronic devices. There are a number of ways of doing this.

Before working with the voltage levels, I just want to say a few words about the "handshaking" signals built into RS-232 (these are all the RS-232 connections other than RX, TX and GND in Fig. 13-17). These six additional lines (which are at the same logic levels as the transmit/receive lines) are used to interface between devices and control the flow of information between computers.

The "Request To Send" ("RTS") and "Clear To Send" ("CTS") lines are used to control data flow between the computer ("DCE") and the modem ("DTE" device). The "Data Transmitter Ready" ("DTR") and "Data Set Ready" ("DSR") lines are used to establish communications. There are two more handshaking lines that are available in the RS-232 standard that you should be aware of, even though chances are you will never connect anything to them. The first is the "Data Carrier Detect" ("DCD"), which is asserted when the modem has connected with another device (i.e. the other device has "picked up the phone"). The "Ring Indicator" ("RI") is used to indicate to a PC whether or not the phone on the other end of the line is ringing or if it is busy. Few of these lines are used in modern RS-232 applications and, as shown in Fig. 13-19, the DTR/DSR and CTS/RTS pairs are often simply shorted together to avoid any "hardware handshaking" issues with the PC.

There is a common ground connection between the DCE and DTE devices. This connection is critical for the RS-232 level converters to determine the actual incoming voltages. The ground pin should never be connected to a chassis or shield ground (to avoid large current flows or be shifted and prevent accurate reading of incoming voltage signals). Incorrect grounding of an application can result in the computer or the device it is

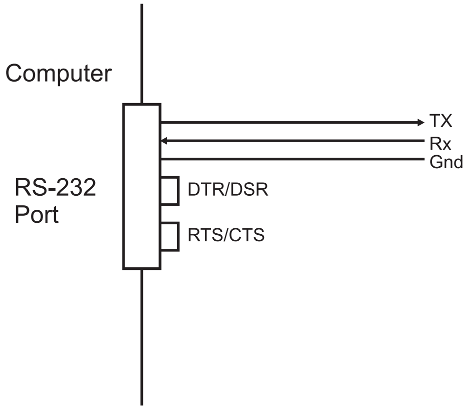
Computer
RS-232
Port
TX
Rx
Gnd
DTR/DSR
RTS/CTS

Fig. 13-19. Three wire RS-232 connection.

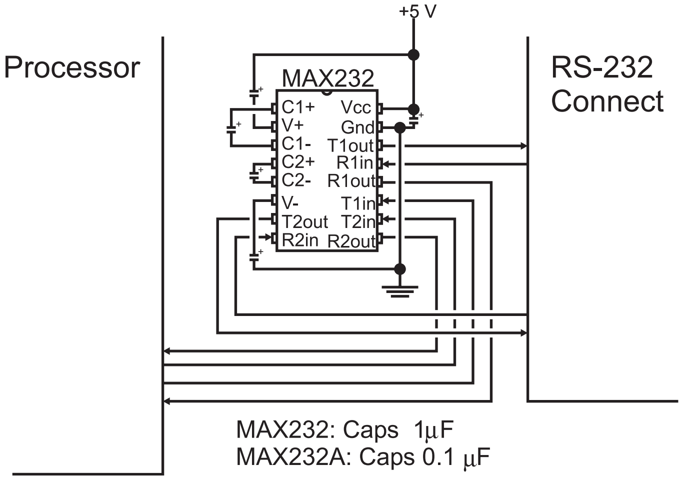
+5 V
Processor
MAX232
RS-232
Connect
C1+
V+
C1-
C2+
C2-
V-
T2out
R2in
Vcc
Gnd
T1out
R1in
R1out
T1in
T2in
R2out
MAX232: Caps 1μF
MAX232A: Caps 0.1 μF

Fig. 13-20. MAX232 RS-232 level shirting chip and schematic diagram.

interfacing having to reset or have they power supplies blow a fuse or burn out. The latter consequences are unlikely, but I have seen it happen in a few cases. To avoid these problems make sure that chassis and signal grounds are separate or connected by a high value (hundreds of k Ω) resistor.

The most popular method for converting TTL/CMOS Logic signals to RS-232 levels is to use the MAXIM MAX232 (see Fig. 13-20,) which has a built in charge-pump voltage generator. This chip is ideal for implementing three-wire RS-232 interfaces (or to add a simple DTR/DSR or RTS/CTS handshaking interface). Ground for the incoming signal is normally connected to the digital electronics ground. Along with the MAX232, MAXIM and some other chip vendors have a number of other RS-232 charge-pump equipped devices that will allow you to handle more RS-232 lines (to include the handshaking lines). Some charge-pump devices that are also available do not require the external capacitors that the MAX232 chip does, which will simplify the layout of your circuit (although these chips do cost quite a bit more).

The next method of translating RS-232 and TTL/CMOS voltage levels is to use the transmitter's negative voltage as the circuit, as Fig. 13-21 shows. This circuit relies off of the RS-232 communications, only running in ''Half-Duplex'' mode (i.e. only one device can transmit at a given time). When the external device wants to transmit to the PC, it sends the data either as a ''Mark'' (leaving the voltage being returned to the PC as a negative value) or as a ''Space'' by turning on the transistor and enabling the positive voltage output to the PC's receivers. If you go back to the RS-232 voltage specification drawing, you'll see that +5 volts is within the valid voltage range for RS-232 ''Spaces''. This method works very well (consuming just about no power) and is obviously a very cheap way to implement a three-wire RS-232 bi-directional interface. There is a chip, the Dallas Semiconductor DS275, which incorporates the circuit above (with a built-in inverter) into the single package shown, making the RS-232 very simple.

Before finishing, I want to make a final point on three wire RS-232 connections. The first is that it cannot be implemented blindly; in about 20% of the RS-232 applications that I have had to do over the years, I have had to implement some subset of the total seven wire (transmit, receive, ground and four handshaking lines) protocol lines. Interestingly enough, I have never had to implement the full hardware protocol. This still means that four out of five times if you wire the connection as shown in Fig. 13-19, the application would have worked.

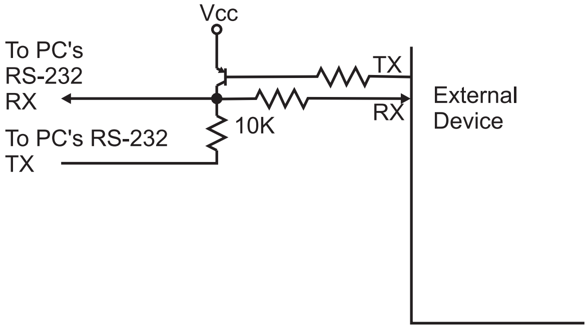

Fig. 13-21. ''Voltage stealing'' RS-232 interface.

Quiz

1. Which type of memory responds in the least amount of time?
   (a) Processor cache
   (b) All three types respond equally fast
   (c) Main Memory (S)DRAM
   (d) Processor ROM

2. What is not a characteristic of a computer bus?
   (a) Controlled by a central processor
   (b) The ability of peripherals on the bus to request information
   (c) Data can be read/written to peripheral busses
   (d) Keeping a log of the processor bus accesses

3. The minimum timing between signal edges in the parallel port is:
   (a) 0.5 µs
   (b) 0.5 ms
   (c) 0.5 ns
   (d) There are no delays; the port works completely asynchronously

4. The raster display on a video display looks like:
   (a) A series of parallel lines going across the display
   (b) Parallel lines that loop back like plough lines in a farmer's field
   (c) A series of small dots drawn on the video display
   (d) A series of brush strokes on the video display

5. "Horizontal synch" in a video display:
   (a) Moves the electron beam in the CRT to the left
   (b) Moves the electron beam in the CRT to the top
   (c) Starts the raster going side to side
   (d) Is needed for CRTs laid down on their side

6. Devices on a synchronous serial bus are normally referred to as:
   (a) "Dominant" and "Recessive"
   (b) "Master" or "Slave"
   (c) "1" or "0"
   (d) "Transmitter" or "Receiver"

7. The maximum speed of the "SPI" synchronous serial protocol is:
   (a) 100 kpbs (thousands of bits per second)
   (b) 1 Mbps (millions of bits per second)

(c) 2 Mbps
(d) 3 Mbps

8. To connect to a modern PC's RS-232 connector, you will need:
   (a) Fiber optic cable
   (b) 75 Ohm Coax and BNC connector
   (c) RJ-45 cable and crimper
   (d) A 9 pin male D-Shell connector

9. The RS-232 voltage levels of −12 volts for a "1" and +12 volts for "0" means that:
   (a) You can connect digital electronics devices directly to RS-232 connections
   (b) You can connect TTL/CMOS Logic drivers directly to RS-232
   (c) You can connect TTL/CMOS Logic inputs directly to RS-232
   (d) You cannot connect digital electronics devices directly to RS-232 connections

10. The MAX232:
   (a) Uses a charge pump to power the circuit
   (b) Requires +12 volts and −12 volts to convert TTL/CMOS Logic signals
   (c) Produces correct RS-232 voltages and converts TTL/CMOS Logic signals
   (d) Will allow a full RS-232 connection with only one chip

Test for Part Two

Do not refer to the text when taking this test. You may draw diagrams or use a calculator if necessary. A good score is at least 38 correct answers. Answers are in the back of the book. It's best to have a friend check your score the first time so you won't memorize the answers if you want to take the test again.

1. A signal with a 100 ns time "On" and a 230 ns time "Off" has a frequency of:
    (a) 30.3%
    (b) 3 GHz
    (c) 3 MHz
    (d) 3 kHz
2. A toggle flip flop will divide a clock frequency by:
    (a) The value selected in its inputs
    (b) 2
    (c) 1.5, but the output duty cycle is 50%
    (d) The toggle flip flop is used for debouncing switch inputs

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

3. The NPN transistor relaxation oscillator is well suited for:
   (a) High-speed computer applications
   (b) Only audio applications
   (c) Low-cost applications where accuracy isn't important
   (d) Power supply duty cycle generators

4. If you wanted an oscillator that produced a period of 4 gate delays, you would:
   (a) Use a standard ring oscillator
   (b) Use something other than a ring oscillator
   (c) Use two three inverter ring oscillators and XOR their outputs
   (d) Use a two inverter ring oscillator and divide the output frequency by two using a toggle D flip flop

5. The duty cycle of a ring oscillator is:
   (a) Always 50%
   (b) Equal to the number of inverters used in the ring oscillator multiplied by the gate delay
   (c) Dependent on the technology used in the inverters
   (d) Always 100%

6. Why are TTL inverters not recommended for relaxation oscillators?
   (a) They are too costly
   (b) They do not operate at a fast enough speed
   (c) Their current controlled inputs will affect the operation of the oscillator
   (d) They may short out internally when used as part of an oscillator.

7. A relaxation oscillator has an R1 value of 1k, C of 0.1 μF and R2 equal to 10 k. What frequency will it oscillate at?
   (a) It won't oscillate
   (b) 4.54 kHz
   (c) 4.54 MHz
   (d) 4.54 Hz

8. Crystals used in oscillators are:
   (a) Relatively high cost but very accurate
   (b) Very robust and immune to shock damage
   (c) Low cost with an accuracy that is similar to that of a "canned oscillator"
   (d) Mined in South America

9. The crystal inverter circuit requires the following parts:
   (a) Crystal and inverter
   (b) Crystal, inverter, two capacitors and two resistors
   (c) Crystal, inductor, three resistors, two capacitors and an NPN transistor
   (d) Crystal, diode, two capacitors, hand wound coil, two diodes

10. Changing the voltage at the "Vcontrol" pin of a 555 astable oscillator:
    (a) Will change the period of the oscillator output
    (b) Won't change anything
    (c) Will lower the amount of power consumed
    (d) Will add unnecessary costs to the circuit

11. A 555 astable oscillator with $R1 = 47\,\text{k}$, $R2 = 100\,\text{k}$ and $C = 4.7\,\mu\text{F}$ will oscillate at:
    (a) 805 Hz
    (b) 0.805 Hz
    (c) 1.24 Hz
    (d) 1.24 kHz

12. Decreasing value of a resistor or capacitor in a 555 monostable will:
    (a) Lower its operating frequency
    (b) Raise its operating frequency
    (c) Increase the pulse width
    (d) Decrease the pulse width

13. A chip with two 555 timers built in has the part number:
    (a) 75555
    (b) 556
    (c) No such chip exists
    (d) 2x555

14. The delay in a "canned delay line" is produced by:
    (a) Elves
    (b) A piece of quartz, cut to provide a specified delay
    (c) A long piece of copper wire, formed as a coil
    (d) A large array of simple gates

15. Why is reset passed to the "Output Formatter" in the sequential circuit block diagram?
    (a) To control the operation of tri-state drivers built into the "Output Formatter"

(b) To synchronize outputs with the operation of the chip
(c) To minimize power usage when the chip is reset
(d) To speed up operation of the sequential circuit

16. To create a cascaded counter from multiple chips, what signal(s) are passed between the chips?
    (a) The most significant two bits
    (b) The "Carry" from the least significant counter and the "Borrow" from the more significant counter
    (c) Just the carry from the least significant counter to the clock input of the more significant counter
    (d) The clock and chip enable signals to be shared between the two counters

17. What are the disadvantages of serial data transmission over parallel data transmission?
    (a) Increased product chip count
    (b) Slower data transmission
    (c) Increased product power requirements
    (d) Increased data error rate

18. The linear feedback shift register polynomial defines:
    (a) The XOR "taps" used to modify the incoming data
    (b) The execution order of magnitude
    (c) The value needed to decode any value run through the LFSR
    (d) The relationship between the number of bits and speed of the LFSR's operation

19. The number of states and inputs in a hardware state machine are a function of:
    (a) The fanout capabilities of the chips used
    (b) The number of discrete states required
    (c) The requirements of the application
    (d) The size of the ROM used

20. Since commands to a Hitachi 44780 controlled LCD require 160 μs or more to execute, what is an effective way of waiting the suggested delay before sending the next command?
    (a) Using a 555 timer wired as a monostable
    (b) Passing requests to the LCD through an RC delay
    (c) Polling the busy flag
    (d) Using a clock with a 5 ms period

21. What signal is not included in a typical sequential logic circuit interface bus?
    (a) _Read
    (b) Data Read Ready
    (c) _Write
    (d) Address Bit 0

22. A "three to eight" decoder:
    (a) Cannot be implemented in TTL or CMOS Logic
    (b) Converts three data selects into eight device read requests
    (c) Converts three clock cycles into eight state machine addresses
    (d) Converts three binary bits into eight interface device selects

23. The segments to be turned on in a seven-segment LED display is determined by:
    (a) A circuit which enables the common cathodes or common anodes in the LED displays
    (b) A microprocessor built into each seven-segment LED display
    (c) A combinatorial circuit performing Boolean arithmetic operations on the display bits
    (d) A sequential circuit that compares the display bits to expected values

24. A PWM with a period of 10 μs and a "high" signal of 4 μs has a duty cycle of:
    (a) Not enough information is available to determine the duty cycle
    (b) 60%
    (c) 0.4
    (d) 40%

25. To produce a PWM output signal of 20 kHz and have 6 bits of accuracy, what input clock frequency should be used?
    (a) 20 kHz
    (b) 64 Hz
    (c) 1.28 MHz
    (d) 5.1 MHz

26. The two inverter debounce circuit is best suited for what kind of input devices?
    (a) Push buttons
    (b) 40%

(c) 0.4
(d) Double throw switches

27. Hysteresis causes logic inputs to be registered:
(a) Faster than normal inputs
(b) At lower thresholds than devices with standard inputs
(c) At more extreme thresholds than devices with standard inputs
(d) When there is an inductor on the line

28. A switch matrix keypad is:
(a) Used by Trinity and Neo to access the Matrix
(b) A two-dimensional array of switches used to provide multiple button inputs
(c) A series of switches that change the output signal of a sequential circuit
(d) The natural evolution of a single button to multiple buttons

29. The Princeton architecture was criticized because:
(a) The single memory interface was felt to be a performance bottleneck
(b) It relied too heavily on vacuum tube technology
(c) The design was taken from other, earlier computer designers
(d) The amount of space required for it seemed to be prohibitive

30. The first digit to the right of the decimal point in a binary floating point number has a value of:
(a) 0.1 decimal
(b) 0.25 decimal
(c) 0.5 decimal
(d) 1.0 decimal

31. Floating point numbers are stored in data size except:
(a) 2 bytes
(b) 4 bytes
(c) 8 bytes
(d) 10 bytes

32. SRAM memory is best suited for applications which require:
(a) Data stacks
(b) Gigabytes of memory
(c) Mass storage of data
(d) Fast data access for a relatively large memory area

33. Stacks cannot be modeled by:
    (a) Trays in a lunchroom
    (b) Pieces of paper on a person's desk
    (c) Stacks of paper produced by a computer processor
    (d) A jelly bean jar

34. The operation of the Zener regulator is analogous to:
    (a) A car's carburetor
    (b) A bowl with a hole cut in the bottom
    (c) A PWM output driven by a comparator
    (d) A comparator output driven by sensor

35. The operation of the linear regulator is analogous to:
    (a) A car's carburetor
    (b) A bowl with a hole cut in the bottom
    (c) A PWM output driven by a comparator
    (d) A comparator output driven by sensor

36. A 10:1 transformer has an input of 110 volts AC. What is its output?
    (a) 11 volts DC
    (b) 11 volts AC
    (c) 110 volts DC
    (d) 11 amps AC

37. "PAL" I/O pins are:
    (a) Input and Output
    (b) Programmable Input or Output
    (c) Input only
    (d) Output only

38. What is not a "local memory" device:
    (a) Processor cache
    (b) PCI status registers
    (c) Main memory (S)DRAM
    (d) Processor ROM

39. The Logic levels for the PC's parallel port are?
    (a) 3.3 volt CMOS Logic compatible
    (b) 5 volt TTL/CMOS Logic compatible
    (c) 5 volt "HCT" Logic compatible
    (d) Current controlled, voltage levels are not considered

40. The inventor of the parallel port pinout standard was:
    (a) IBM
    (b) Apple Computers
    (c) The Centronics Corporation
    (d) Xerox FARC

41. The ‘‘Data’’ bits of the PC’s parallel port have what kind of outputs?
    (a) Discrete transistors
    (b) CMOS totem pole outputs
    (c) Pulled up Open Collectors that can be changed by external devices
    (d) TTL tri-state buffers

42. ‘‘Vertical synch’’ in a video display:
    (a) Moves the electron beam in the CRT to the left
    (b) Moves the electron beam in the CRT to the top
    (c) Starts the raster going up and down
    (d) Is needed for CRTs pointing upwards

43. The ‘‘NTSC’’ video standard was first developed for:
    (a) Military radar screens
    (b) Portable news gathering services
    (c) Electronic instruments
    (d) Televisions

44. The ‘‘Microwire’’ synchronous serial protocol sends:
    (a) A word consisting of 16 bits of data
    (b) A word consisting of 8 bits of data
    (c) Status Information before Data
    (d) Command Information before Data

45. I2C’s ‘‘Acknowledge’’ bit:
    (a) Is used as a ‘‘Parity Bit’’ for Received Data
    (b) Indicates the Received Data was Valid
    (c) Is a placeholder in case 9 bits of data is required
    (d) Is not currently implemented, but will be in the future

46. Baudot asynchronous serial communications was developed for:
    (a) Networking
    (b) Televisions
    (c) Teletypes
    (d) Telegraphs

47. What are the differences between Baudot asynchronous serial communications and NRZ serial communications?
    (a) None
    (b) The number of bits transmitted
    (c) The speed at which data is transmitted
    (d) The number of "overhead" bits

48. The primary difference between Manchester Encoding and NRZ is:
    (a) The media used to transmit the data
    (b) The bit timing for "0" and "1" change in Manchester Encoding
    (c) The data rates used to transmit the data
    (d) The companies that hold the original patents on the technologies

49. To implement a three wire RS-232 connection, you will have to tie together:
    (a) RTS/CTS and DSR/DTR
    (b) RI/CTS and DSR/RTS
    (c) DCD/DSR and RTS/CTS
    (d) DCD/CTS and RI/DSR

50. The RS-232 to TTL/CMOS level translator shown in Fig. Test 2.1:
    (a) Is an inexpensive, full capability RS-232 interface
    (b) Sends data in "half duplex" mode only
    (c) Can be used for handshaking lines as well as RX and TX
    (d) Is limited by the speed it can operate at

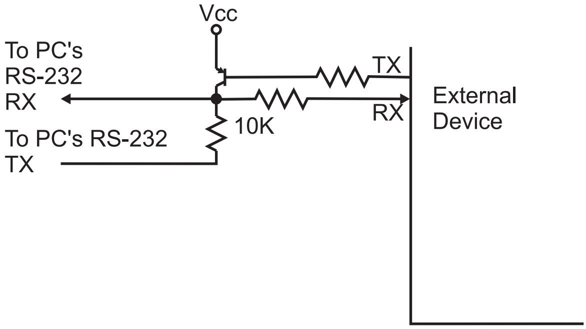

Fig. Test 2-1.

Final Exam

Do not refer to the text when taking this test. You may draw diagrams or use a calculator if necessary. A good score is at least 75 correct answers. Answers are in the back of the book. It’s best to have a friend check your score the first time so you won’t memorize the answers if you want to take the test again.

1. The NAND Gate is the base technology for:
   (a) TTL logic circuits
   (b) Your TV set
   (c) Computer chips
   (d) Symbolic logic

2. What is not an indicator for an Inverted bit?
   (a) “!”
   (b) “_”
   (c) “^”
   (d) “-”

3. Combinatorial logic is used:
   (a) In a DJ’s mixing table to combine audio inputs together with special effects
   (b) To produce a desired output from a specific set of inputs
   (c) To express the mathematics of cryptography
   (d) To collate data from different sources

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

4. The following truth table represents the following basic gate:

Input "A"	Input "B"	Output
0	0	0
0	1	1
1	1	0
1	0	0

(a) AND
(b) OR
(c) NOT
(d) None

5. The following truth table can be represented as the sum of products:

Input "A"	Input "B"	Input "C"	Output
0	0	0	0
0	0	1	0
0	1	1	1
0	1	0	0
1	1	0	0
1	1	1	0
1	0	1	0
1	0	0	1

(a) $(A \cdot B \cdot C) + (!A + !B + !C)$
(b) $(!A \cdot B \cdot C) + (!A + !B + !C)$
(c) $(!A \cdot B \cdot C) + (A + !B + !C)$
(d) $(A \cdot !B \cdot C) + (A + !B + !C)$

6. The following truth table can be represented as the product of sums:

Input "A"	Input "B"	Input "C"	Output
0	0	0	0
0	0	1	1
0	1	1	1
0	1	0	1
1	1	0	1
1	1	1	0
1	0	1	1
1	0	0	1

(a) $(!A + !B + !C) \cdot (A + B + C)$
(b) $(!A + !B + C) \cdot (A + B + C)$
(c) $(!A + !B + !C) \cdot (A + !B + C)$
(d) $(!A + B + !C) \cdot (A + B + C)$

7. Oscilloscopes are used for:
(a) Testing the impedance of a circuit trace
(b) Display the state of multiple signals
(c) Displaying the current voltage of individual signals
(d) Displaying the analog voltage of individual signals

8. According to the AND associative law, which equation is true?
(a) $(A + B) + C = A + (B + C) = A + B + C$
(b) $A \cdot (B + C) = (A \cdot B) + (A \cdot C)$
(c) $A + (B \cdot C) = (A + B) \cdot (A + C)$
(d) $!(A + B) = !A \cdot !B$

9. A CMOS logic technology has a gate delay of 25 ns and an AND statement built from it requires two gate delays for the signal to pass through it. How long will it take a signal to pass through the AND gate?
(a) 0 ns
(b) 25 ns

(c) 50 ns
(d) Indeterminate

10. The first step in reducing the following truth table is:

Input "A"	Input "B"	Input "C"	Output
0	0	0	0
0	0	1	0
0	1	1	1
0	1	0	1
1	1	0	0
1	1	1	1
1	0	1	1
1	0	0	0

(a)

Input "A"	Input "B"	Input "C"	Output
0	0	0	0
0	0	1	0
0	1	x	1
1	1	0	0
1	x	1	1
1	0	0	0

(b)

Input “A”	Input “B”	Input “C”	Output
0	0	0	0
0	0	1	0
0	1	x	1
1	1	0	0
1	1	1	1
1	0	1	1
1	0	0	0

(c)

Input “A”	Input “B”	Input “C”	Output
0	0	0	0
0	0	1	0
0	1	1	1
x	1	0	1
1	1	1	1
1	0	1	1
1	0	0	0

(d)

Input "A"	Input "B"	Input "C"	Output
0	0	0	0
0	0	1	0
0	1	1	1
0	1	0	1
1	1	0	0
x	1	1	1
1	0	0	0

11. Using De Morgan's NAND Theorem. "!(!A · !B)" is equal to:
    (a) A + B
    (b) !A + !B
    (c) !(A · B)
    (d) A · B

12. The sum of products logic equation

    `Output` $= ($`!A · B · !C`$) + ($`!A · !B · !C`$)$

    can be reduced to:

    (a) A · C
    (b) !A · !C
    (c) !C · CB
    (d) C

13. A Karnaugh map has three bits equal to 1, can one circle (and one AND statement) be used to represent the logic function?
    (a) Yes
    (b) Yes, if inverters can be used
    (c) Yes, if an XOR gate could be used
    (d) No

14. The four bit Karnaugh map

	AB			
	00	01	11	10
CD-00	0	1	1	1
01	0	1	0	0
11	0	1	0	0
10	0	1	1	1

has the optimized sum of product equation:

(a) Output = (!B · !D) + (!A · !B)
(b) Output = (A · B) + (A · D)
(c) Output = (!B · A) + (B · D)
(d) Output = (!A · B) + (A · !D)

15. The NAND equivalent to an OR gate is:
(a) Built from two NAND gates and requires three gate delays for a signal to pass through
(b) Built from two NAND gates and requires two gate delays for a signal to pass through
(c) Built from three NAND gates and requires two gate delays for a signal to pass through
(d) Built from one NOR gate as well as a NOT gate and requires two gate delays for a signal to pass through

16. The voltage drop across a resistor in a parallel circuit:
(a) Is the same as the applied voltage
(b) Is proportional to the power dissipated in the circuit
(c) Is always zero
(d) Is proportional to the resistor's value relative to the total resistance in the circuit multiplied by the applied voltage

17. A 47μF capacitor:
(a) Has a value of 47 trillionths of a farad.
(b) Weighs 47 micrograms
(c) Probably uses polyester as a dielectric
(d) Has a value of 0.000047 farads

18. The equivalent resistance of a 10 ohm resistor in parallel with 10 ohm and 20 ohm resistors in series is:
    (a) 40 ohms
    (b) 30 ohms
    (c) 7.5 ohms
    (d) 6.7 ohms

19. In a PNP bipolar transistor, collector current:
    (a) Flows into the emitter through to the collector
    (b) Is dependent on the emitter current
    (c) Is negative with respect to the power source
    (d) Like in the NPN bipolar transistor, does not include the base current

20. CMOS Logic is:
    (a) Sound controlled
    (b) Resistor controlled
    (c) Current controlled
    (d) Voltage controlled

21. "Breadboards" allow:
    (a) Delivery of bagels and other breakfast foods without any garbage
    (b) Simple creation of electronic products that can be delivered to consumers
    (c) The method that provides the fastest signal paths for digital electronics circuits
    (d) Simple creation of prototype circuits

22. The holes in a breadboard:
    (a) Allow effective cooling of the breadboard circuitry
    (b) Allow components to be mounted and interconnected
    (c) Are there simply for aesthetics
    (d) Minimize the resistance encountered by signals

23. A circuit that would provide valid TTL and CMOS logic inputs would:
    (a) Both source current as well as provide high and low voltages
    (b) Both provide high voltage as well as sink and source current
    (c) Both provide low voltage as well as sink and source current
    (d) Both sink current as well as provide high and low voltages

24. The four bit binary number B’1010 1101’ expressed as a hexadecimal number is:
(a) 171
(b) 0x0AB
(c) 0x02231
(d) 0x0171

25. The most significant digit in the decimal number “1234” is:
(a) 1
(b) 2
(c) 3
(d) 4

26. The decimal number “42” in hexadecimal is:
(a) 0x042
(b) B’0010 0100’
(c) 2A
(d) 0x02A

27. The four bit hexadecimal number 0x0F4EA expressed in decimal is:
(a) −2,838
(b) 62,698
(c) B’1111 0100 1110 1010’
(d) 29.930

28. B’0101’ in binary, using the formula Gray code $=$ Binary ˆ (Binary$\gg$1) can be converted to the Gray code:
(a) B’1010’
(b) B’0110’
(c) B’0111’
(d) B’1110’

29. The Gray code B’1010’ corresponds to the binary value
(a) B’1111’
(b) B’1110’
(c) B’1101’
(d) B’1010’

30. A “Carry Look-Ahead” adder:
(a) Is the only way addition should be carried out in digital electronics
(b) Is often needlessly complex for most applications
(c) Produces its result in an amount of time independent of the number of bits

(d) Produces its result in an amount of time dependent on the number of bits

31. In a universe where infinity (the highest possible number) is one thousand (1,000); "−34" could be represented as:
    (a) Only −34
    (b) 966
    (c) 976
    (d) 34

32. Converting the eight bit, two's complement value "−124" to a positive number by inverting each bit and incrementing the result produces the bit pattern:
    (a) B'0111 1100'
    (b) Which is seven bits long and is invalid
    (c) B'1000 0100'
    (d) B'01111 110'

33. Multiplying two eight bit numbers by repeated addition:
    (a) Will require up to 8 addition operations
    (b) Will require up to 255 addition operations
    (c) Cannot be implemented in digital electronics
    (d) Is the fastest way of performing binary multiplication

34. Multiplying a binary number by 64 can be accomplished by:
    (a) Clearing the least significant six bits
    (b) Shifting left six bits
    (c) Shifting right six bits
    (d) Setting the least significant six bits

35. What is the best method of multiplying binary number B'0110 1101' by B'0011 1110'?
    (a) Boothe's algorithm: shift B'0110 1101' to the left by 7 and subtract B'0110 1101 from the result
    (b) Shifting left seven bits according to Boothe's algorithm
    (c) Shifting right six bits according to Boothe's algorithm
    (d) Boothe's algorithm: shift B'0110 1101' to the right by 7 and subtract B'0110 1101 from the result

36. If you looked up the "gate delay" specification for the technology used to build an AND gate chip, what would the chip's gate delay be?
    (a) Exactly what is listed for the technology used by the AND gate chip
    (b) Dependent on whether or not the chip technology is NOR or NAND gate based
    (c) Twice the listed gate delay
    (d) Three times the listed gate delay

37. What tools should *not* be used to find a possible race condition in a circuit?
    (a) An oscilloscope
    (b) A logic analyzer
    (c) A logic simulator
    (d) A time delay reflectometer

38. A chip fanout of 6 means:
    (a) The chip's output can drive six inputs
    (b) The chip provides six outputs for a single input
    (c) There can be six chips connected to the various outputs
    (d) The chip provides six outputs, each one a variation on the inputs

39. When working with TTL, the Mickey Mouse "AND" circuit should have a resistor value of:
    (a) The Mickey Mouse AND gate will not work with TTL
    (b) 4 watts
    (c) 10 k
    (d) 470 Ω

40. Can a pulled up button (which pulls to ground) be attached to a dotted AND bus?
    (a) Yes, if the signal passes through an open collector transistor
    (b) Yes, if the signal passes through a gate with an open collector driver
    (c) Unconditionally, yes
    (d) No

41. Can a totem pole driver be attached to a dotted AND bus?
    (a) Yes, if the power to the chip with the totem pole driver is the same as the chips on the dotted AND bus
    (b) Yes, if care is taken to ensure grounds are kept at the same voltage level

(c) Unconditionally, yes
(d) No

42. When the TTL input is pulled low, how much current flows out of the pin?
(a) 1 μA
(b) 100 μA
(c) 1 mA
(d) No current at all

43. The most popular output driver type found in logic chips is:
(a) The totem pole driver
(b) Tri-state driver
(c) Open collector
(d) Open drain

44. A commercial microprocessor reset chip will:
(a) Pull the microprocessor's _Reset line low at power up and when its input voltage falls below a set value
(b) Drive the microcontroller's _Reset line high at power up and when its input voltage falls below a set value
(c) Pull the microprocessor's _Reset line low at power up.
(d) Turn on a LED when its input voltage falls below a set value to indicate a power supply problem

45. A small triangle on the input of a sequential device indicates:
(a) The pin is buffered before being processed within the chip
(b) The signal is inverted before being processed within the chip
(c) The signal is a clock to the chip's internal flip flops
(d) The circuitry connected to the chip was either not completed or qualified and it should be used with caution

46. Which formula approximates the time required for an RC network to change state due to a sudden voltage?
(a) $t = 2.2 \times R \times C$
(b) $t = -RC \times \ln(Vcc)$
(c) $t = -RC \times \ln(Vcc - (Vcc \times e^{-1/RC}))$
(d) $t = V \times i$

47. A switch bounces:
(a) Never
(b) Most of the time, but this can be filtered out with a capacitor
(c) Only if it is not feeding the input of a flip flop
(d) Always

48. The NOR "RS" flip flop saves data when the inputs are:
    (a) R = 0, S = 0
    (b) R = 0, S = 1
    (c) R = 1, S = 1
    (d) R = 1, S = 0

49. The NAND "RS" flip flop sets the output (drives out "1") when the inputs are:
    (a) R = 0, S = 0
    (b) R = 0, S = 1
    (c) R = 1, S = 1
    (d) R = 1, S = 0

50. Reset circuitry for sequential circuits should be used:
    (a) Always
    (b) Only for digital clocks to avoid the clock starting up at a random time
    (c) In situations where the start up values are critical
    (d) Never

51. A signal with a 100 ns time "On" and a 230 ns time "Off" has a frequency of:
    (a) 30.3%
    (b) 3 GHz
    (c) 3 MHz
    (d) 3 kHz

52. A toggle flip flop will divide a clock frequency by:
    (a) The value selected in its inputs
    (b) 2
    (c) 1.5, but the output duty cycle is 50%
    (d) The toggle flip flop is used for debouncing switch inputs

53. The NPN transistor relaxation oscillator is well suited for:
    (a) High-speed computer applications
    (b) Only audio applications
    (c) Low-cost applications where accuracy isn't important
    (d) Power supply duty cycle generators

54. If you wanted an oscillator that produced a period of 4 gate delays, you would:
    (a) Use a standard ring oscillator
    (b) Use something other than a ring oscillator
    (c) Use a two three inverter ring oscillators and XOR their outputs

(d) Use a two inverter ring oscillator and divide the output frequency by two using a toggle D flip flop

55. The duty cycle of a ring oscillator is:
    (a) Always 50%
    (b) Equal to the number of inverters used in the ring oscillator multiplied by the gate delay
    (c) Dependent on the technology used in the inverters
    (d) Always 100%

56. Why are TTL inverters not recommended for relaxation oscillators?
    (a) They are too costly
    (b) They do not operate at a fast enough speed
    (c) Their current controlled inputs will affect the operation of the oscillator
    (d) They may short out internally when used as part of an oscillator

57. A relaxation oscillator has an R1 value of 1 k, C of 0.1 μF and R2 equal to 10 k. What frequency will it oscillate at?
    (a) It won't oscillate
    (b) 4.54 kHz
    (c) 4.54 MHz
    (d) 4.54 Hz

58. Crystals used in oscillators are:
    (a) Relatively high cost but very accurate
    (b) Very robust and immune to shock damage
    (c) Low cost with an accuracy that is similar to that of a "canned oscillator"
    (d) Mined in South America

59. The crystal inverter circuit requires the following parts:
    (a) Crystal and inverter
    (b) Crystal, inverter, two capacitors and two resistors
    (c) Crystal, Inductor, three resistors, two capacitors and an NPN transistor
    (d) Crystal, diode, two capacitors, hand wound coil, two diodes

60. Changing the voltage at the "Vcontrol" pin of a 555 astable oscillator:
    (a) Will change the period of the oscillator output
    (b) Won't change anything

(c) Will lower the amount of power consumed
(d) Will add unnecessary costs to the circuit

61. A 555 astable oscillator with $R1 = 47\,k$, $R2 = 100\,k$ and $C = 4.7\,\mu F$ will oscillate at:
    (a) 805 Hz
    (b) 0.805 Hz
    (c) 1.24 Hz
    (d) 1.24 kHz

62. Decreasing value of a resistor or capacitor in a 555 monostable will:
    (a) Lower its operating frequency
    (b) Raise its operating frequency
    (c) Increase the pulse width
    (d) Decrease the pulse width

63. A chip with two 555 timers built in has the part number:
    (a) 75555
    (b) 556
    (c) No such chip exists
    (d) 2 × 555

64. The delay in a "canned delay line" is produced by:
    (a) Elves
    (b) A piece of quartz, cut to provide a specified delay
    (c) A long piece of copper wire, formed as a coil
    (d) A large array of simple gates

65. Why is reset passed to the "Output Formatter" in the sequential circuit block diagram?
    (a) To control the operation of tri-state drivers built into the "Output Formatter"
    (b) To synchronize outputs with the operation of the chip
    (c) To minimize power usage when the chip is reset
    (d) To speed up operation of the sequential circuit

66. To create a cascaded counter from multiple chips, what signal(s) are passed between the chips?
    (a) The most significant two bits
    (b) The "Carry" from the least significant counter and the "Borrow" from the more significant counter
    (c) Just the carry from the least significant counter to the clock input of the more significant counter

(d) The clock and chip enable signals to be shared between the two counters

67. What are the disadvantages of serial data transmission over parallel data transmission?
    (a) Increased product chip count
    (b) Slower data transmission
    (c) Increased product power requirements
    (d) Increased data error rate

68. The linear feedback shift register polynomial defines:
    (a) The XOR "taps" used to modify the incoming data
    (b) The execution order of magnitude
    (c) The value needed to decode any value run through the LFSR
    (d) The relationship between the number of bits and speed of the LFSR's operation

69. The number of states and inputs in a hardware state machine are a function of:
    (a) The fanout capabilities of the chips used
    (b) The number of discrete states required
    (c) The requirements of the application
    (d) The size of the ROM used

70. A constant delay of 100 µs can be used for instruction and data writes to the 44780 controlled LCD except when:
    (a) Writing the "Cursor Move" instruction
    (b) When scrolling the LCD upwards
    (c) When writing the "Screen Clear" and "Cursor Home" commands
    (d) Backspacing from the start of the current line to the start of the next

71. What signal is not included in a typical sequential logic circuit interface bus?
    (a) _Read
    (b) Data Read Ready
    (c) _Write
    (d) Address Bit 0

72. A "three to eight" decoder:
    (a) Cannot be implemented in TTL or CMOS Logic
    (b) Converts three data selects into eight device read requests
    (c) Converts three clock cycles into eight state machine addresses
    (d) Converts three binary bits into eight interface device selects

73. The segments to be turned on in a seven-segment LED display is determined by:
    (a) A circuit which enables the common cathodes or common anodes in the LED displays
    (b) A microprocessor built into each seven-segment LED display
    (c) A combinatorial circuit performing Boolean arithmetic operations on the display bits
    (d) A sequential circuit that compares the display bits to expected values

74. A PWM with a period of 10 µs and a "high" signal of 4 µs has a duty cycle of:
    (a) Not enough information is available to determine the duty cycle
    (b) 60%
    (c) 0.4
    (d) 40%

75. To produce a PWM output signal of 20 kHz and have 6 bits of accuracy, what input clock frequency should be used?
    (a) 20 kHz
    (b) 64 Hz
    (c) 1.28 MHz
    (d) 5.1 MHz

76. The two inverter debounce circuit is best suited for what kind of input devices?
    (a) Push buttons
    (b) 40%
    (c) 0.4
    (d) Double throw switches

77. Hysteresis causes logic inputs to be registered:
    (a) Faster than normal inputs
    (b) At lower thresholds than devices with standard inputs
    (c) At more extreme thresholds than devices with standard inputs
    (d) When there is an inductor on the line

78. A switch matrix keypad is:
    (a) Used by Trinity and Neo
    (b) A two-dimensional array of switches used to provide multiple button inputs

(c) A series of switches that change the output signal of a sequential circuit
(d) The natural evolution of a single button to multiple buttons

79. The Princeton architecture was criticized because:
   (a) The single memory interface was felt to be a performance bottleneck
   (b) It relied too heavily on vacuum tube technology
   (c) The design was taken from other, earlier computer designers
   (d) The amount of space required for it seemed to be prohibitive

80. The first digit to the right of the decimal point in a binary floating point number has a value of:
   (a) 0.1 decimal
   (b) 0.25 decimal
   (c) 0.5 decimal
   (d) 1.0 decimal

81. Floating point numbers are stored in data size except:
   (a) 2 bytes
   (b) 4 bytes
   (c) 8 bytes
   (d) 10 bytes

82. SRAM memory is best suited for applications which require:
   (a) Data stacks
   (b) Gigabytes of memory
   (c) Mass storage of data
   (d) Fast data access for a relatively large memory area

83. Stacks cannot be modeled by:
   (a) Trays in a lunchroom
   (b) Pieces of paper on a person's desk
   (c) Stacks of paper produced by a computer processor
   (d) A jelly bean jar

84. The operation of the Zener regulator is analogous to:
   (a) A car's carburetor
   (b) A bowl with a hole cut in the bottom
   (c) A PWM output driven by a comparator
   (d) A comparator output driven by sensor

85. The operation of the linear regulator is analogous to:
   (a) A car's carburetor

(b) A bowl with a hole cut in the bottom
(c) A PWM output driven by a comparator
(d) A comparator output driven by sensor

86. A 10:1 transformer has an input of 110 volts AC. What is its output?
(a) 11 volts DC
(b) 11 volts AC
(c) 110 volts DC
(d) 11 amps AC

87. "PAL" I/O pins are:
(a) Input and Output
(b) Programmable Input or Output
(c) Input only
(d) Output only

88. What is not a "local memory" device:
(a) Processor cache
(b) PCI status registers
(c) Main memory (S)DRAM
(d) Processor ROM

89. The Logic levels for the PC's parallel port are?
(a) 3.3 volt CMOS Logic compatible
(b) 5 volt TTL/CMOS Logic compatible
(c) 5 volt "HCT" Logic compatible
(d) Current controlled, voltage levels are not considered

90. The inventor of the parallel port pinout standard was:
(a) IBM
(b) Apple Computers
(c) The Centronics Corporation
(d) Xerox PARC

91. The "Data" bits of the PC's parallel port have what kind of outputs?
(a) Discrete transistors
(b) CMOS totem pole outputs
(c) Pulled up Open Collectors that can be changed by external devices
(d) TTL tri-state buffers

92. "Vertical synch" in a video display:
    (a) Moves the electron beam in the CRT to the left
    (b) Moves the electron beam in the CRT to the top
    (c) Starts the raster going up and down
    (d) Is needed for CRTs pointing upwards

93. The "NTSC" video standard was first developed for:
    (a) Military radar screens
    (b) Portable news gathering services
    (c) Electronic instruments
    (d) Televisions

94. The "Microwire" synchronous serial protocol sends:
    (a) A word consisting of 16 bits of data
    (b) A word consisting of 8 bits of data
    (c) Status Information before Data
    (d) Command Information before Data

95. I2C's "Acknowledge" bit:
    (a) Is used as a "Parity Bit" for Received Data
    (b) Indicates the Received Data was Valid
    (c) Is a placeholder in case 9 bits of data is required
    (d) Is not currently implemented, but will be in the future

96. Baudot asynchronous serial communications was developed for:
    (a) Networking
    (b) Televisions
    (c) Teletypes
    (d) Telegraphs

97. What are the differences between Baudot asynchronous serial communications and NRZ serial communications?
    (a) None
    (b) The number of bits transmitted
    (c) The speed at which data is transmitted
    (d) The number of "overhead" bits

98. The primary difference between Manchester Encoding and NRZ is:
    (a) The media used to transmit the data
    (b) The bit timing for "0" and "1" change in Manchester Encoding
    (c) The data rates used to transmit the data

99. To implement a three wire RS-232 connection, you will have to tie together:
    (a) RTS/CTS and DSR/DTR
    (b) RI/CTS and DSR/RTS
    (c) DCD/DSR and RTS/CTS
    (d) DCD/CTS and RI/DSR

100. The RS-232 to TTL/CMOS level translator shown in Figure Final Exam 1:
    (a) Is an inexpensive, full capability RS-232 interface.
    (b) Send data in ''half duplex'' mode only
    (c) Can be used for handshaking lines as well as RX and TX
    (d) Is limited by the speed it can operate at

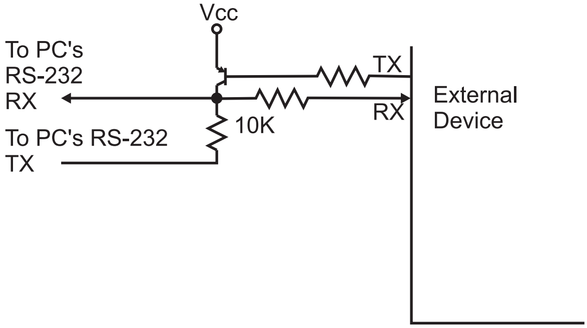
Vcc
To PC's RS-232 RX
To PC's RS-232 TX
10K
TX
RX
External Device

Fig. Final Exam-1.

Appendix: Answers to Quiz, Tests, and Final Exam

Chapter 1:
1. b 2. a 3. c 4. a 5. d
6. c 7. c 8. b 9. c 10. a

Chapter 2:
1. c 2. a 3. a 4. d 5. d
6. b 7. c 8. d 9. a 10. b

Chapter 3:
1. d 2. b 3. c 4. c 5. d
6. c 7. c 8. d 9. c 10. a

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

Chapter 4:
1. c 2. b 3. b 4. d 5. d
6. b 7. c 8. b 9. c 10. a

Chapter 5:
1. a 2 .c 3. a 4. c 5. d
6. a 7. d 8. b 9. b 10. d

Chapter 6:
1. a 2. c 3. d 4. d 5. c
6. c 7. b 8. d 9. b 10. d

Chapter 7:
1. c 2. b 3. b 4. c 5. c
6. d 7. a 8. b 9. c 10. c

Test for Part One
1. d 2. b 3. c 4. a 5. b 6. d 7. b 8. c
9. c 10. b 11. a 12. a 13. a 14. b 15. d 16. b
17. b 18. b 19. a 20. c 21. d 22. a 23. a 24. c
25. c 26. a 27. b 28. d 29. c 30. c 31. a 32. c
33. c 34. b 35. c 36. b 37. c 38. b 39. c 40. a
41. b 42. d 43. a 44. a 45. d 46. d 47. b 48. c
49. d 50. b

Chapter 8:
1. a 2. d 3. a 4. d 5. a
6. d 7. c 8. d 9. a 10. b

Chapter 9:
1. b 2. d 3. a 4. d 5. a
6. d 7. c 8. b 9. a 10. b

Chapter 10:
1. b 2. a 3. c 4. a 5. b
6. b 7. a 8. a 9. d 10. a

Chapter 11:
1. c 2. d 3. b 4. c
5. b 6. a 7. d 8. a

Chapter 12:
1. b 2. d 3. b 4. c 5. c
6. b 7. a 8. d 9. c 10. d

Chapter 13:
1. a 2. d 3. a 4. a 5. a
6. b 7. d 8. d 9. d 10. c

Test for Part Two
1. c 2. b 3. c 4. b 5. a 6. c 7. b 8. a
9. b 10. a 11. c 12. d 13. b 14. c 15. a 16. c
17. b 18. a 19. d 20. c 21. b 22. d 23. c 24. c
25. c 26. d 27. c 28. b 29. a 30. c 31. a 32. d
33. c 34. b 35. a 36. b 37. a 38. b 39. b 40. c
41. c 42. b 43. d 44. a 45. b 46. c 47. a 48. b
49. a 50. b

Final Exam:
1. a 2. c 3. b 4. d 5. c 6. a 7. d 8. b
9. c 10. a 11. a 12. b 13. d 14. d 15. c 16. a
17. d 18. c 19. a 20. d 21. d 22. b 23. d 24. b
25. a 26. d 27. b 28. c 29. a 30. c 31. b 32. a
33. b 34. b 35. a 36. b 37. c 38. a 39. d 40. c
41. d 42. c 43. a 44. a 45. c 46. a 47. d 48. a
49. d 50. a 51. c 52. b 53. c 54. b 55. a 56. c
57. b 58. a 59. b 60. a 61. c 62. d 63. b 64. c
65. a 66. c 67. b 68. a 69. d 70. c 71. b 72. d
73. c 74. c 75. c 76. d 77. c 78. b 79. a 80. c
81. a 82. d 83. c 84. b 85. a 86. b 87. a 88. b
89. b 90. c 91. c 92. b 93. d 94. a 95. b 96. c
97. a 98. b 99. a 100. b

This page intentionally left blank

INDEX

Page numbers for illustrations are shown in **bold** type

adder in binary arithmetic, 122–9
address and data decoders, 247–9 *see also* circuit interfaces
Adobe Acrobat, 268
American Standard Code for Information Interchange *see* ASCII
AND logic operation, 4–8, **9**, 45, 123
  associative law, 42
  identity law, 44
ASCII, display by LCD, 245
ASIC (Application Specific Integrated Circuit), 300, 301
asynchronous communications, 320–1

Baud Rate, **317**
Baudot Teletype, 317, 320
BCD (binary coded decimal), 107–11, 228
Bell Laboratories, 111
binary, defined, 5
binary arithmetic, 118–44
  adder, 122–9
    carry bit, 122, 123, 124, 129
    carry look-ahead adder input/output truth table, 125–9
    full adder circuit, **123**
    half adder circuit, **122**
    sum bit, 122, 123, 124, 129
    three bit ripple adder circuit, **125**
    XOR gate, 123
  addition, 119, **120**
  floating point numbers, 281–3
  integrated adder/subtracter circuit, **135**
  magnitude comparators, 135–8
    bus to group multiple bits, 137
    cascading two 4 bit comparators, **138**
    comparator built from two subtracters, **136**
  multiplication and division, 138–44
    division, 143–4
    four bit multiplier circuit, **142**
    multplicand and multiplier defined, 139
    quotient and dividend defined, 143
    techniques for multiplication
      Booth's algorithm, 141
      order log and order n, 140
  subtracter, 129–35
    full and half subtracter circuits, **131**
    Minuend, 130, 131
    representing a four bit number, 134
    subtrahend, 130, 131
    three bit ripple subtractor, **133**
    truth table, 132
  subtraction, 120, **121**
binary reflected code *see* Gray code
BIOS (basic input/output system), 307, 308
bit, defined, 3, 6
Boole, George, ix, 6

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

- Boolean arithmetic, ix, 6–10
  - equations, 21
  - laws and rules, 40–5
  - statements, 6, 7
  - summary of laws and rules, 141
- Booth's algorithm in binary arithmetic, 141
- button debouncing, 258–62 *see also* circuit interfaces

- capacitors, 64–9 *see also* digital electronic circuits
- Carroll, Lewis, 6
- Carry bit in binary arithmetic, 122, 123, 124, 129, 230
- CAS (column address strobe), 287
- cathode ray tube *see* CRT
- Centronics Corporation, 306
- checksum value *see* CRC
- chip numbering system, 269
- chip types:
  - CMOS *see* CMOS (complementary metal oxide semiconductor)
  - Maxim, 323
  - MOSFET *see* MOSFET (metal oxide semiconductor field effect transistor)
  - PTH (pin through hole), 270
  - Shottkey, 269
  - SMT (surface mount technology), 270
  - TTL *see* TTL (transistor–transistor logic)
- circuit interfaces, 243–65 *see also* complex sequential circuits
  - adding LEDs, 243
  - address and data decoders, 247–9
    - 74139 decoder, 248, **249**
    - Address, Read, and Write operations, **248**
  - button debouncing, 258–62
    - chart of input *versus* gate response, **261**
    - debounce circuit, **259**
    - filtering of the bouncing, **260**
    - hysteresis property of Schmitt Trigger inputs, **260**
    - monostable 555 timer circuit, **261**
    - oscilloscope picture of bounce, **259**
    - Schmitt Trigger Input Inverter, 259–60
    - use of CMOS inverter, 258
    - waveforms of timer circuit, **262**
  - Hitachi 44780 LCD, **244**
    - command set, 246
    - initialization process, 247
    - waveform diagram, **245**
  - multi-segment LEDs, 250–3
    - 7447 chip as decoder, 251
    - appearance of display, **250**, **251**
    - combinatorial specification, 252
    - Maxim MAX7219 chip for multiple display, **253**
    - wiring diagram, **251**
  - PWM (pulse width modulation), 253–8
    - duty cycle, **254**
    - effective frequency, 257
    - generator circuit, **255**
    - installation on breadboard, **257**
    - period counter circuit, **254**
    - power saving, 257–8
    - schematic diagram, **256**
  - switch matrix keypad interfacing, 262–5
    - 74C922 keypad decoder chip, **264**
    - PC keyboards, 262–3
    - switch matrix circuit, **263**
    - use of capacitors, 264–5
- CISC (Complex Instruction Set Computers), 277
- clocked register, 175
- CMOS (complementary metal oxide semiconductor), 11
  - 74C85 chip in PWM circuit, 256
  - input/output circuit, **83**, **84**, **85**
  - inverter, 207, 209, 258
  - and MOSFET, 79
  - NOR gate, 46
  - numbering of chips, 149
- column address strobe *see* CAS
- combinatorial logic circuits, 14–7, 29–50, 146–61 *see also* digital electronic circuits
  - circuit using multiple gates, **14**, 15
    - decoded response of the circuit, 17
  - circuit with marked outputs, **16**
  - home alarm logic, **31**, 44, **49**, **50**
    - testing, 45
    - truth table, 32
    - truth table function reduction, 32–36
      - using Gray Codes, 33, 35

implementation, 146–61
  chip characteristics, 147
  combining functions on a net, 159–61
    dotted AND and tri-state drivers bus, 156–7, 158, **159**, **160**
    use of CMOS, 159–61
  logic technologies, 148–9
  quick and dirty logic gates, 153–6
    inverter circuit, **154**, **155**
    MML AND gate, **156**
    MML (Mickey Mouse logic), 153–6
    RTL (resistor–transistor logic), 154
  race conditions, 150–3
    definition, 150
    examples of circuits, **150**, **151**
    timing analysis, 153
    waveform output, **151**
    XOR gate glitch, 151–3
  tri-state drivers, **157**, **158**
optimisation, 29–50
  Boolean arithmetic laws, 40–5
    associative law, 42
    commutative law, 42
    complementary law, 43
    De Morgan theorem, 42
    distributive law, 43
    identity law, 44
    summary of laws and rules, 41
  gate delay, 30
  of home alarm, 31–6, **49**, **50**
  Karnaugh maps, 36–40
    circling bits, **38**, **39**
    converting a truth table, **37**, 38
    description, 37
  number of gates, 29
  optimising for technology, 30, 45–50
    De Morgan theorem, 45, 46, 47
    exclusive OR gate truth table, **46**
    home alarm logic circuit, **49**, **50**
    implementing gates using NAND or NOR, **47**, **48**
    optimised XOR gate built from NAND gates, **49**
    XOR gate converted to NAND gates, **48**
wires that connect or bypass, **15**
complementary metal oxide semiconductor *see* CMOS
Complex Instruction Set Computers *see* CISC
complex sequential circuits, 225–40 *see also* feedback and sequential circuits
  asynchronous digital logic design, 227
  counters, 228–31
    circuit using
      carry bit, 230, **231**
      D flip flop, 229, **230**
      edge triggered flip flop, **228**, 229
      latches, **229**
      74LS193, 229, **230**
    defined, 228
    use of different chips, 231
  digital clock block diagram, **226**
  hardware state machines, 237–40
    application to traffic lights, 238–40
      circuit, **238**
      ROM programming, 239
    circuit based on ROM, **237**
  shift registers, 231–7
    circuit converting parallel data into serial stream, **233**
    definition, 231–2
    linear feedback shift register (LFSR), 234–7
      circuit, **236**
      creation of checksum value (CRC), 235
      features, 236
      producing pseudo-random numbers, 235
      shift register with XOR gates, 234, **235**
      use for encryption, 235
    serial and parallel data transfers, **232**
    synchronous data protocols, 234
    use of four flip flops to receive data, **234**
  state memory block diagram, 226, **227**
  time update circuit, 226, **227**
computer processors, 277–301
  CISC basis, 277–88
  floating point numbers, 281–3
    binary numbers, 281
    IEEE 754 format, 282
    Intel Pentium data format, **283**
  hardwired processor, **280**, 281

computer processors (*continued*)
  Harvard computer, 278, **279**
  Memory Interface Unit, 278
  memory types, 283–7
    arrangement of cells, **286**
    CAS (column address strobe), 287
    DRAM (dynamic random access memory), **285**, **286**, 287
    EEPROM (electrically eraseable programmable read-only memory), 284
    EPROM (eraseable programmable read-only memory), **283**, 284
    LIFO (last in/first out), 287
    memory stacks, **287**
    SRAM (static random access memory), 284, **285**
  microcoded processor, 279, **280**
  PLD (programmable logic devices), 298–301
    advantages, 300–1
    ASIC (Application Specific Integrated Circuits), 300, 301
    defined, 298
    GAL (generic array logic), 298
    input/output circuit, **299**
    PAL (programmable array logic), 298
    VHDL (Very High Speed Integrated Circuit Hardware Description Language), 300
    XOR gate, 299, **300**, 301
  power supplies, 287–98
    to LED circuit, 289–90
    linear voltage regulators, **292**, **293**, 297
    power supply circuit, **297**, 298
    rectifier, 297
    SMPS (switch mode power supply) circuit, **293**
      chip circuit, **295**
      PWM (pulse width modulation), 293, 294
      VCO (voltage controlled oscillator), 293–4
      waveforms, **294**
    socket layout, **296**
    transformer, **296**
    voltage regulation, **290**, **291**
    Zener diode, **288**, 289, 291
  Princeton architecture, **278**, 279
  RISC basis, 277–88
counters, 228–31 *see also* complex sequential circuits
CRC (checksum value or cyclic redundancy check), 235
CRT (cathode ray tube), 309, 311, 312
crystals and ceramic resonators, 208–11 *see also* oscillators
cyclic redundancy check *see* CRC

D flip flop, 177, **178**, 229, **230**
DAC (digital analog converter), 309
De Morgan theorem, 42, 45, 46, 47, 110
decoders, 247–9
demultiplexors *see* decoders
digit, defined, 5–6
digital clock, **167**, 168, 199–201
digital electronic circuits, 53–97 *see also* combinatorial logic circuits
  basic circuit diagram, **54**
  capacitors, 64–9
    appearance and markings, **66**
    capacitor symbol, **65**
    electrical units, 65–6
    as low-pass filter, **67**, 68
    polarized and unpolarized, 66
    RC time constant, 68–9
    resistor/capacitor circuit, **68**
  components of electrical circuits, 54–6
  direct and alternating current, **57**, 58
  electrical load, 58
  electrical power, 63–4
  inverter testing, 91–7
    circuit to test current draw, **96**
    circuit to test voltage control, **97**
    circuits with ‘1’ or ‘0’ input, **95**
    DIP package for TTL chip, **91**
    input to inverter circuit, **93**
    inverter operation, 92, **94**
    parts needed, 93, 94–5
    switch input circuit with LED, **92**
    wiring diagram for TTL inverter, **94**
  Kirchoff’s Laws, 62–3
  logic circuit test kit, 87–91
    9 volt battery clip, **88**
    breadboard with a wiring kit, **89**
    interior connections of breadboard, **90**

    - necessary parts, 90
    - power supplies, 88
  - logic gate input and output, 78–87
    - CMOS
      - gate response for different inputs, **84**
      - logic gate input, **83**
      - NOR gate, **85**
    - MOSFET chips, 79, 83, 84, 85
    - pull up and pull down, **86**, 87
    - TTL chip
      - current control of input, **80**
      - definition, 78–9
      - dotted AND gate, 81, **82**
      - input circuits, **79**
      - open collector output, **81**
      - totem pole output, 80, **81**
      - tri-state driver, 83
  - Ohm's Law, 60, 62, 64
  - positive and negative connections, 56
  - resistors, **58**, 59, **61**, **62**, **63**
    - equivalent resistance, **63**
    - in parallel, **62**, 63
    - in series, **61**, 62
  - semiconductors *see* semiconductor operation
  - SI (Système Internationale) units, 60
- *Digital Electronics Demystified* (Predko), overview, x
- digital multi-meter *see* DMM
- digital waveforms, **23**, **24**
- DIP (dual-in-line package), **91**
- DMA (direct memory access), 303
- DMM (Digital Multi-Meter), 91, 96
- Dodgson, Charles Lutwidge, 6
- dotted AND gate, 81, **82**, 156–7, 159–61
- DRAM (dynamic random access memory), 171, **285**, 286, 306

- edge triggered flip flop, 175, **176**, 228–229
- EEPROM (electrically eraseable programmable read-only memory), 284
- encryption using linear feedback shift register (LFSR), 235
- EPROM (eraseable programmable read-only memory), 237, **283**, 284, 298
- Ex-OR gate *see* Exclusive OR gate
- Exclusive OR gate, 11, **12**, 113, 123, 129, 135, 151–3, 234, 299
- exclusive or operation *see* OR

- falling edge clocked register/triggered flip flop, 175
- FCC (Federal Communications Commission), 311
- feedback and sequential circuits, 164–83 *see also* complex sequential circuits
  - backdriving, 167
  - feedback defined, 164, **165**
  - flip flops *see* flip flops
  - latches and registers, 178–80
    - analogy for latch, 179
    - D flip flop latch, **179**
    - registers defined, 179
    - waveforms for register and latch, **180**
  - memory circuit, **166**
  - reset, 180–3
    - analogue comparator circuit, 181, **182**
    - initialization, 181
    - processor reset control chip, 182, **183**
    - resistor/capacitor network, **182**
  - ring oscillator, 165, **166**
  - sequential circuits
    - defined, 167
    - digital clock circuit, **167**, 168
- Firewire protocol, 317
- 555 timer chip, 211–19 *see also* oscillators
- *555 Timer Cookbook* (Lancaster), 217
- flip flops, 168–78
  - defined, 168, 173
  - edge triggered, 173–8
    - clocked register block diagram, **175**
    - D flip flop
      - 7474 chip, 177, **178**, 229, **230**
      - circuit and waveforms, 175, **176**
      - full circuit, 177, **178**
    - JK flip flop
      - circuit and waveform, **174**
      - truth table, 175
  - NAND flip flop
    - circuit, **171**
    - truth table, 171
  - NOR flip flop
    - circuits, **169**, **172**
    - parts list, 173

flip flops (*continued*)
    truth table, 170
  RS (reset-set) flip flop, **168**
  volatility, 172
format for floating point numbers, 282
Franklin, Benjamin, 55
FSB (front side bus), 305

GAL (generic array logic), 298
gate *see also* logic operations
  defined, 10
  delay, 30
  output True or False, 12, 13, 14
graticule markings, 23
Gray, Frank, 111
Gray Code, 7, 8, 9, 16, 33, 35, 111–5

hardware state machines, 237–40 *see also* complex sequential circuits
Harvard University, 278, 279
Hitachi 44780 LCD, **244**
home alarm logic circuit, **31**, 32–6, 45, **49**, **50** *see also* combinatorial logic circuits

IBM (International Business Machines), 306
I2C (Inter-Intercomputer Communications), 316, 317, 318, 319
idealised waveform diagram, **24**
idempotent laws, 42
IEEE (Institute of Electrical and Electronic Engineers) logic symbols, 272, **273**, 274
IEEE Standard Graphic Symbols for Logic Functions, 272
implementation of combinatorial logic circuits, 146–61 *see also* combinatorial logic circuits
Institute of Electrical and Electronic Engineers *see* IEEE
Intel Pentium data format, **283**
Internet, 268
inverters, 91–7 *see also* digital electronic circuits
ISA (industry standard architecture) bus, 308

Karnaugh, Maurice, 36
Karnaugh maps, 36–40, 42, 43, 45
  circling bits, **38**, **39**
  converting a truth table, **37**, 38
  description, 37
Kirchoff's Laws, 62–3

Lancaster, Don, 217
LCD (liquid crystal display), 244–7
LED (light emitting diode), **72**, 91–3, 243, 250–3, 289
LFSR (linear feedback shift register), 234–7 *see also* complex sequential circuits
LIFO (last in/first out), 287
light emitting diode *see* LED
linear feedback shift register *see* LFSR
liquid crystal display *see* LCD
logic circuit test kit, 87–91
logic operations, 4–6, 10–4
  AND, 4–8, **9**
  Exclusive OR gate, 11, **12**
  gates
    defined, 10
    output True or False, 12, 13, 14
  Gray Code, 7, 8, 9, 16
  NAND gate, **10**
  NOR gate, **11**
  NOT, 4–8, **9**
  OR, 4–8, **9**
  six elementary gates, 13
  state list, 21–2
  truth table, 7, 8, 9, 11, 12

magnitude comparators in binary arithmetic, 135–8
Manchester encoding in asynchronous communications, 320–321
Maxim:
  MAX232 chip, **323**
  MAX7219 chip, **253**
Memory Interface Unit, 278
Microwire protocol, 314
MML (Mickey Mouse logic), 153–6
MOSFET (metal oxide semiconductor field effect transistor), 65, **76**, 77–8, 79, 84, 85, 149, 284

multiplication and division in binary arithmetic, 138–44
NAND gate, **10**, 30, 45–50, 78
nanosecond, defined, 30
NOR gate, **11**, 45–50
NOT logic operation, 4–8, **9**
NPN bipolar transistors, 73–6, 78, 79, 80, 204, 208, 209, 291
NRZ (Non-return to Zero) asynchronous communications, 320
NTSC (National Television Standards Council), 310–311, 313, 314
number systems, 100–15
  binary coded decimal, 107–11
    advantages, 108
    circuit for 'invalid' formula, **109**
    conversion of binary number to Gray code, **110**
  decimal and binary numbers, 100–4
  Gray codes, 111–5
    circuit to convert Gray code to binary number, **114**, 115
    conversion from binary codes, 113
    hamming value, 111
  hexadecimal numbers, 104–7
    use of phonetic codes, 105, 106

Ohm's Law, 60, 62, 64
optimisation of combinatorial logic circuits, 29–50 *see also* combinatorial logic circuits
OR logic operation, 4–8, **9**, 45
  complementary law, 43
  distributive law, 43
oscillators, 199–223
  crystals and ceramic resonators, 208–11
    ceramic resonator oscillator circuit, **210**
    CMOS 4060 chip, **211**
    CMOS inverter oscillator circuit, **209**
    disadvantages of crystals, 210
    quartz crystal oscillator circuit, **209**
  delay circuits, 219–23
    alternative chips to the 555 timer, 221
    canned delay line, 221, **222**
    chained delay line, 222, **223**
    monostable generator, 219, **220**
    RS flip flop operation, 220, 221
    waveforms, **220**
  digital clock, 199–201
    duty cycle and pulse width modulation, 201
    jitter, 200, **201**
    waveform, **200**
  555 timer chip, 211–9
    block diagrams, 212, **213**
    change delay, **215**
    continuity tester circuit, 218, **219**
    555 Timer Cookbook (Lancaster), 217
    oscillator circuit, **216**
    RS flip flop, 215
    two cascaded timer circuits, 217, **218**
    view, **212**
    voltage divider circuit, 212, 213, **214**
    waveforms, **217**
  relaxation oscillators, 206–8
    disadvantages, 208
    resistor/capacitor circuit, **206**
    waveform, **207**
  ring oscillators, 204–6
    combinatorial circuits, **205**
    operating rules, 205–6
  transistor astable oscillators, 202–4
    circuit, **202**
    defined, 202
    operation, **203**
    waveform, **204**
oscilloscope screen, **22**

PAL (programmable array logic), 298–301
PC (personal computer), 125, 183, 234, 305–26
  asynchronous serial interfaces, 317–322
    Baud rate and data rate, **317**, 318, 320
    data packets, 319
    Manchester encoding, **320**, 321
    NRZ (Non-return to Zero), 320
    overspeed clock, **318**
  cache memory, **304**
  computer bus
    defined, **305**
    FSB (front side bus), 305
    ISA (industry standard architecture) bus, 306

PC (personal computer) (*continued*)
      PCI (personal computer interface), 305–6
   core circuitry, **304**
   DMA (direct memory access), 304–5
   parallel (printer) port, 306–9
      connector diagram, **307**
      control register, 308
      LPT port registers, 308
      status buffer, 308
      timing problems, 307–308
      waveform, **308**
   PC connectors, **320**
   RS-232 (recommended standard), 317–24
      disadvantages, 322–4
      handshaking, 322, **324**
      MAX232 chip, **323**
      voltage translation, **323**
   synchronous serial interfaces, 314–317
      I2C (Inter-Intercomputer Communications) network, 314, 315, 316, 317
         clock line to strobe data, 315
         stereo system circuit, **316**
         waveform, **316**
      Microwire protocol, 314, 315
      SPI (serial peripheral interface) protocol, 314, **315**
   video output, 311–6
      address generator, **311**
      CRT (cathode ray tube), 311, 312, 313
      DAC (digital analog converter), 309
      hsynch (horizontal synch pulse), **312**
      NTSC (National Television Standards Council), 310–311, 313, 314
      raster display, **309**, 311
      VRAM (video random access memory) schematic, **310**
      vsynch (vertical synch pulse), **313**
PCB (printed circuit board), 17, 30, 222, 256, 272
PCI (personal computer interface), 307–8
philosophy and binary arithmetic, ix
piezo-electric effect, 208
PLD (programmable logic devices), 298
PNP bipolar transistor, 75, 76, 291, 292
Princeton University, 278, 279
printed circuit board *see* PCB
programmable array logic *see* PAL
pseudo-random numbers and linear feedback shift register (LFSR), 235
PTH (pin through hole) chip, 270
PWM (pulse width modulation), 73, 201, 253–8, 293

RAM (random access memory), 65, 278
RC (resistor/capacitor) network, **182**, 183, 206
read-only memory *see* ROM
reading datasheets, 268–75
   chip datasheet, 271–2
   chip operating characteristics, 270–2
   IEEE logic symbols, 272–4
      gate pin modifiers, **273**
      representation of gates, **272**
   part numbers, 269–70
   power usage, 274
   schematic test circuit, **271**
Reduced Instruction Set Computers *see* RISC
reduction of logic circuits *see* optimisation of combinatorial logic circuits
resistors, 58–63 *see also* digital electronic circuits
RISC (Reduced Instruct Set Computers), 277
ROM (read-only memory), 237–40, 305
RS-232 (recommended standard), 317–325 *see also* PC (personal computer)
RTL logic (resistor-transistor logic), 154

Schmitt Trigger Input Inverter, 259–60
SDRAM (synchronous dynamic random access memory), 171
semiconductor operation, 69–78 *see also* digital electronic circuits
   diode operation, **70**
      N-type and P-type, 69–70
      types of bias, 70
   diode symbol, **69**
   transistors
      N-channel MOSFET, 76, 77–8

- operation, **76**
- symbols, **77**
- NPN bipolar, 73–6, 291
  - cross-section, **76**
  - operation, **75**
  - symbol, **74**
- P-channel MOSFET, 77
  - symbol, **77**
- PNP bipolar, 75, 76, 291, 292
  - symbol, **75**
- uses for diodes, 71–3
  - full wave rectifier, **71**
  - LED (light emitting diode), **72**
    - use of PWM (pulse width modulation), 73
  - Zener voltage regulator, **71**, 72

7474 chip, 177, **178**, 222, 229
shift registers, 231–7 *see also* complex sequential circuits
Shottkey chips, 269
SI (Système Internationale) units, 60
single transistor memory cells, 285
six elementary gates, 13
SMPS (switch mode power supply) circuit, **293**
SMT (surface mount technology) chip, 270
SPI (serial peripheral interface) protocol, 314, **315**
SRAM (static random access memory), 284
state list, 21–2
subtracter in binary arithmetic, 129–35
Sum bit in binary arithmetic, 122, 123, 124, 129
sum of products format for circuits, 17–21
- derivation, 17–8
- examples of sum of products logic circuit, **18**, **21**
- inputs/outputs True/False, 19, 20
- product of sums circuit, 19, **20**
- three input gates, 18, **19**

switch matrix keypad interfacing, 262–5 *see also* circuit interfaces

transistor–transistor logic *see* TTL
transistors, 73, **74**, **75**, **76**, 77–8
tri-state drivers, **157**, **158**, 159–61
True and False:
- inputs and outputs, 14
- statements, 4, 7–10

truth table, 7, 8, 9, 11, 12
- adder input/output, 125–9
- conversion, **37**, 38
- exclusive OR gate, **46**
- flip flops, 170, 171, 175
- function reduction for home alarm, 32–6
- for subtracter, 132

TTL (transistor–transistor logic), 11
- bipolar transistors in 7400 series, 49, 78, 79, 177, 178, 222, 229, 253
- current control of input, **80**
- definition, 78–9, 91
- in a DIP package, **91**
- dotted AND gate, 81, **82**
- gate delay, 30
- input circuits, **79**
- inverter, **94**, 207
- open collector output, **81**
- in PWM circuit, 256
- 74LS04, 222, 229
- totem pole output, 80, **81**
- tri-state driver, 83
- use of NAND gate, 46

USB (universal serial bus), 303, 317

VCO (voltage controlled oscillator), 293–4
VCR (videocassette recorder), 250
VHDL (Very High Speed Integrated Circuit Hardware Description Language), 300
Von Neumann, John, 278, 279
VRAM (video random access memory), **312**

waveform diagrams, 21–5
- defined, 22
- digital waveforms, **23**, **24**
- flip flops, **174**, 175, **176**
- graticule markings, 23
- I2C network, **316**

waveform diagrams (*continued*)
  idealised waveform diagram, **24**
  oscillators, **200**, **204**, **207**
  oscilloscope screen, **22**
  parallel port, **308**
  race conditions, **151**
  register and latch, **180**
  SMPS circuit, **294**
  timer circuit, **262**

XOR gate *see* Exclusive OR gate

Zener diode for power supply, **71**, **288**

ABOUT THE AUTHOR

Myke Predko is New Technologies Test Engineer at Celestica in Toronto, Canada. He is the author of McGraw-Hill's *Programming and Customizing PICMicro Microcontrollers*, Second Edition, and is a co-designer of both *TAB Electronics Build Your Own Robot Kits*.

Copyright © 2005 by The McGraw-Hill Companies, Inc. Click here for terms of use.

Donald Lee Smith

Don Smith is elderly and has suffered several serious strokes in the last few years. He is no longer able to respond to e-mails and his web site has been taken over by people who appear very keen to ensure that his work and information can no longer be accessed by members of the public.

At this point in time there is very little information available on Don's achievements. What I have been able to locate is four video recordings of lectures which he gave, plus a copy of his pdf document entitled "Resonance Energy Systems". There is also a web site with limited information. This site appears to have been set up by Don's son whom I suspect does not fully understand how his father's devices work. Don has produced at least forty eight different devices which draw energy from what Don prefers to call "the ambient background". His devices are capable of supplying kilowatts of excess energy and in most cases they do not require any input energy to be supplied by the user.

It is said that 40,000 copies of Don's pdf document have been sold worldwide, but it appears that it is no longer possible to buy a copy, and for that reason, a copy is contained within this document. Please be aware that Don states quite openly that he does not disclose all of the details on any of his devices in his public discussions. This is because the rights to each device have been assigned to a different company in which Don has a financial interest, and so it is not in his interests to disclose the full details. However, ha says that he discloses enough for somebody who is experienced in radio-frequency electronics to be able to deduce the things which he does not disclose and so build a device for his own use. If that is the case, then anybody who has succeeded in doing so has kept very quiet about it afterwards (which is understandable). Having said that, Tariel Kapanadze of Georgia appears to have replicated one of the designs although it is quite likely that Tariel deduced the operating principles for himself. These principles are clearly based on the work of Nikola Tesla.

In Don's work, he refers to the educational kit entitled "Resonant Circuits" No 10 - 416 from The Science Source at www.thesciencesiource.com. That kit now costs US$100 and they don't mention the fact that you need a signal generator capable of producing a 10 volt output at up to 1 MHz in order to be able to do resonance tests, and the kit now uses an LED rather than a neon bulb.

There is a "book" of Don's entitled "An Answer to America's Energy Deficit" but, being an earlier work than his pdf document, it does not appear to contain anything extra of any significance. Here is Don's pdf document, in it's original American-language wording with just minor clerical corrections:

RESONANCE ENERGY METHODS

Donald L. Smith
TransWorld Energy, CEO
September 23, 2002

Fax/Phone 281-370-4547 and e-mail donsm1@earthlink.net

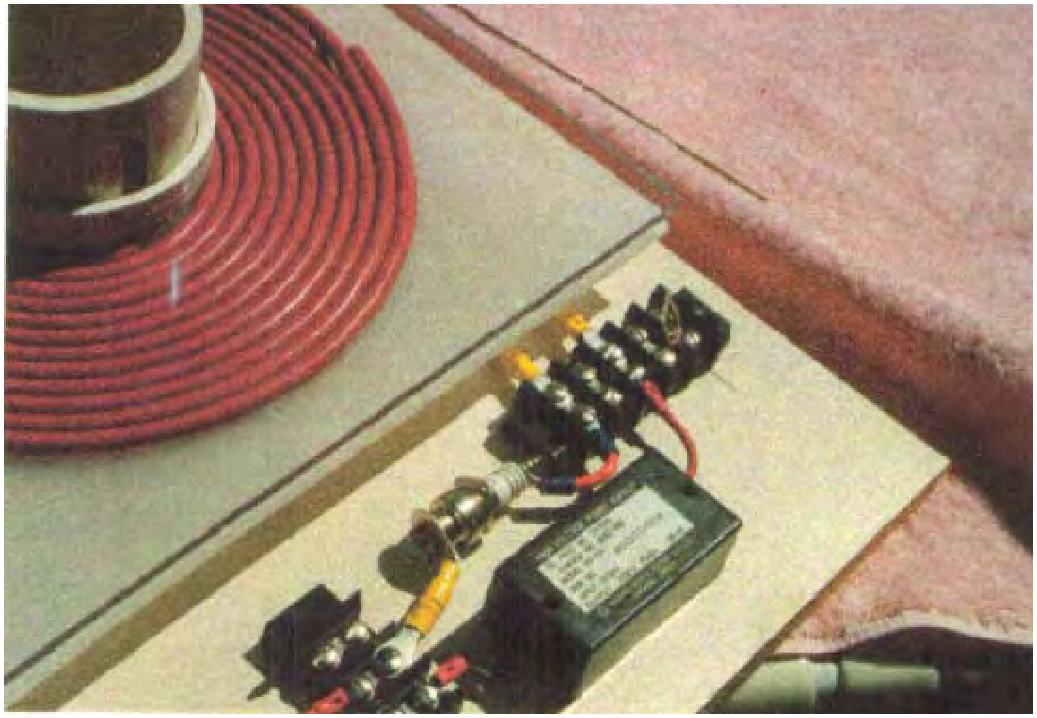

DANGER
HIGH VOLTAGE

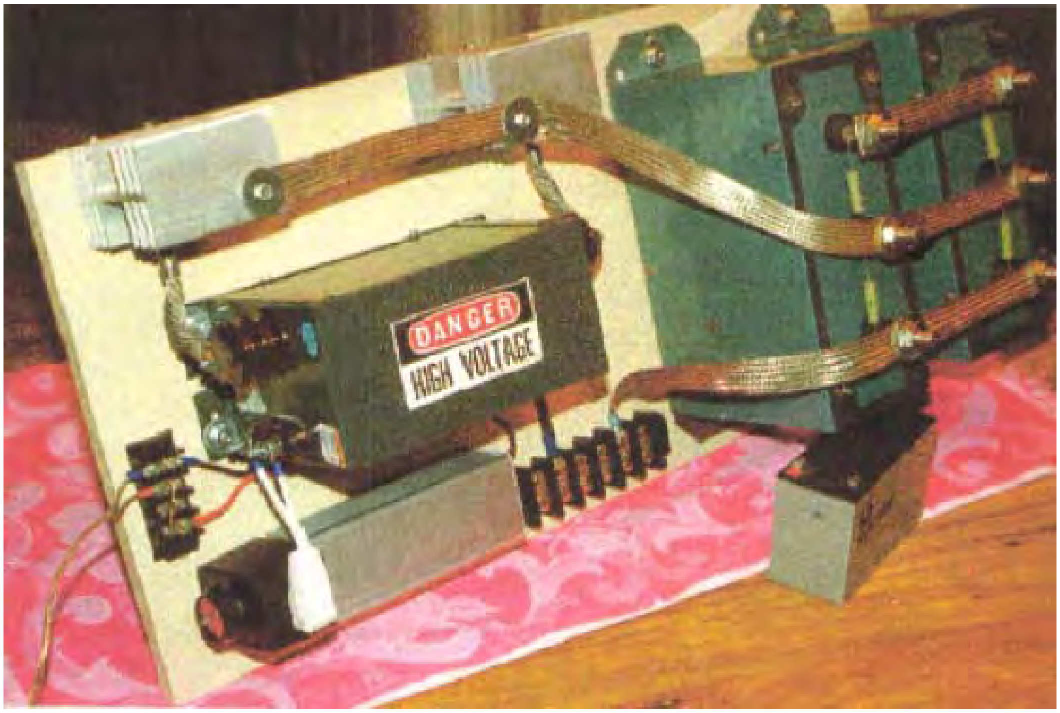
DANGER
HIGH VOLTAGE

DRI-VAR
capacitor
Dry metallized
Self healing
V - 60 Hz - 3 Ph
Outdoor type
Insulation level 5 / 15 kV
Contains discharge resistors
Made in Finland
Before use read instruction leaflet
LISTED
DRI-VAR
capacitor
Dry metallized
Self healing
V - 60 Hz - 3 Ph
Outdoor type
Insulation level 5 / 15 kV
Contains discharge resistors
Made in Finland
Before use read instruction leaflet
LISTED

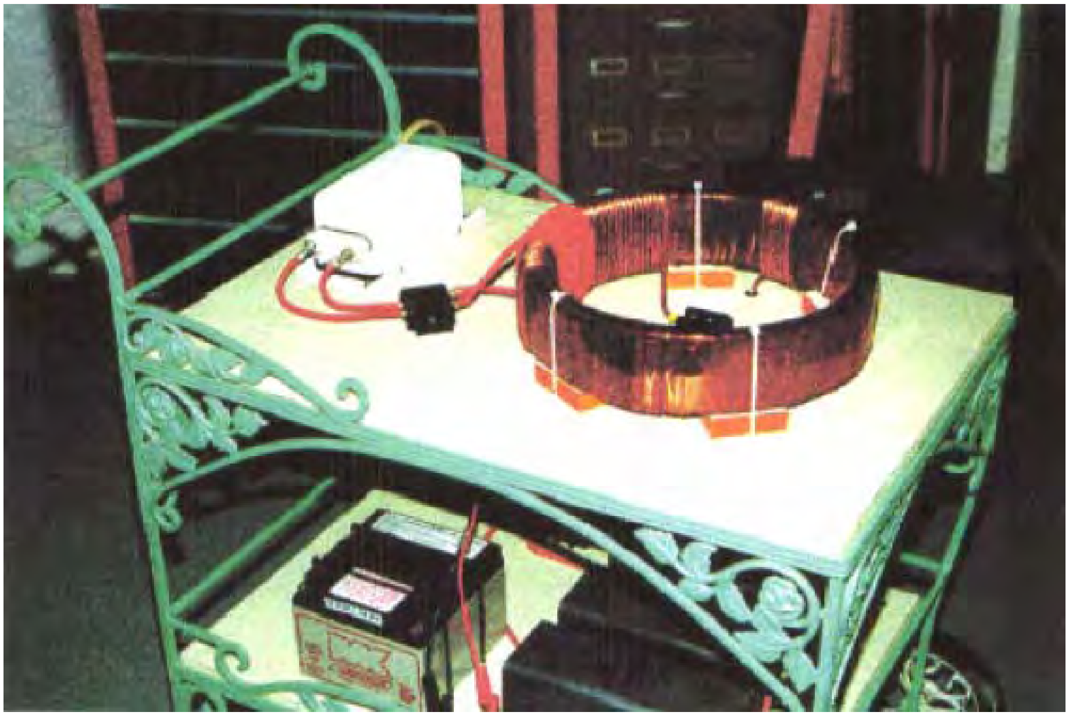

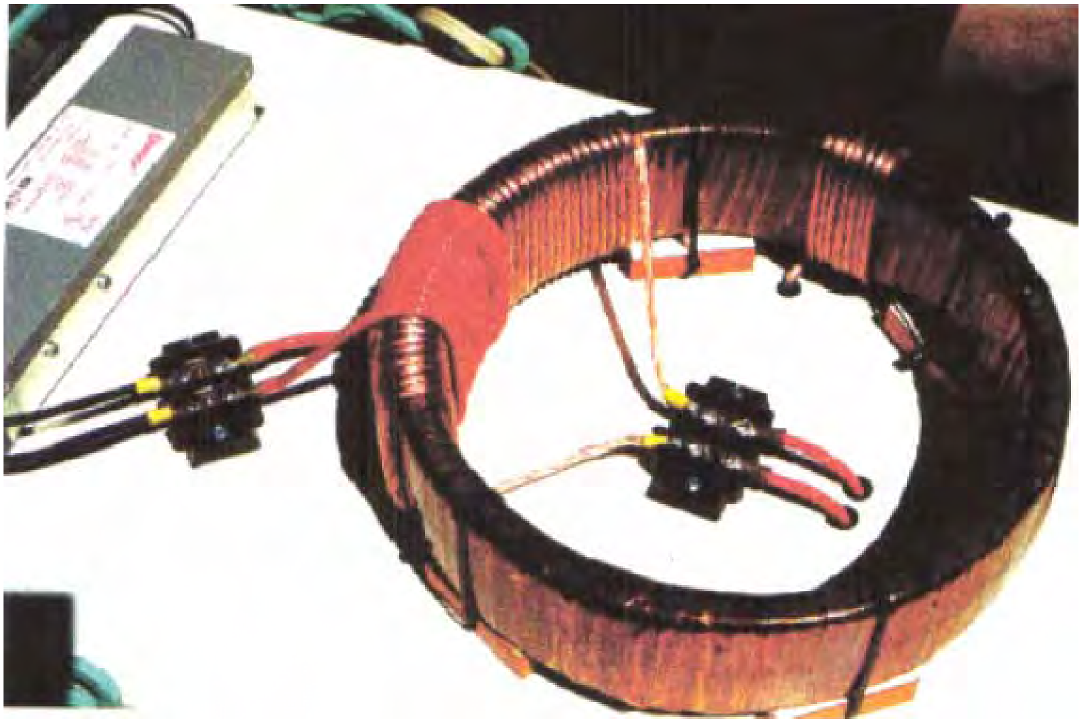

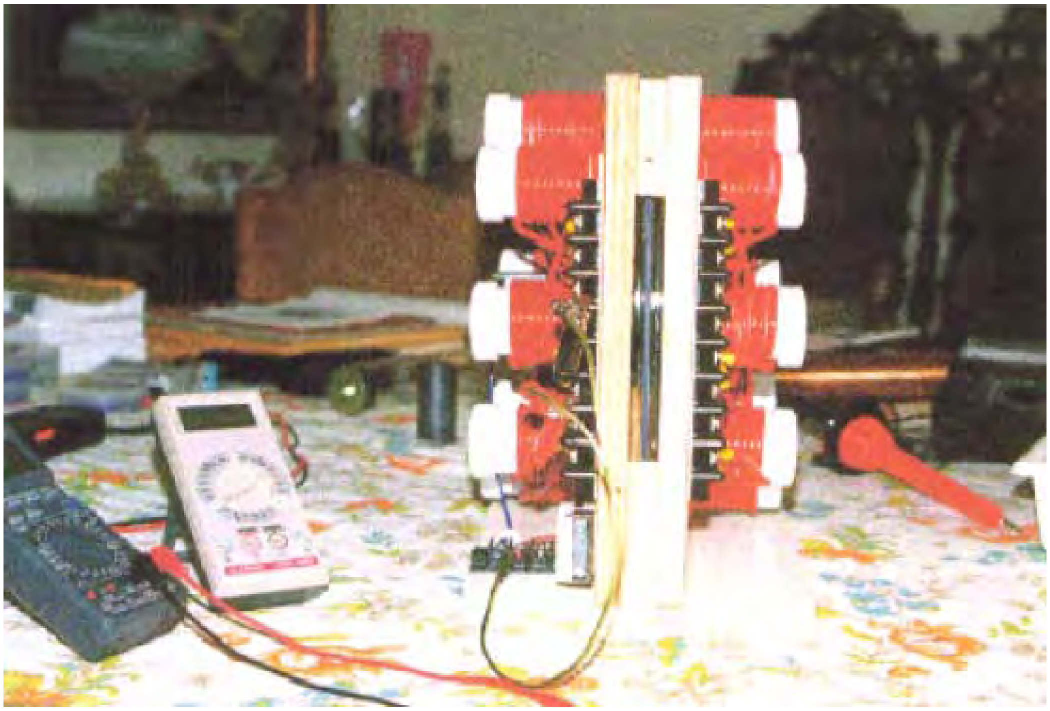

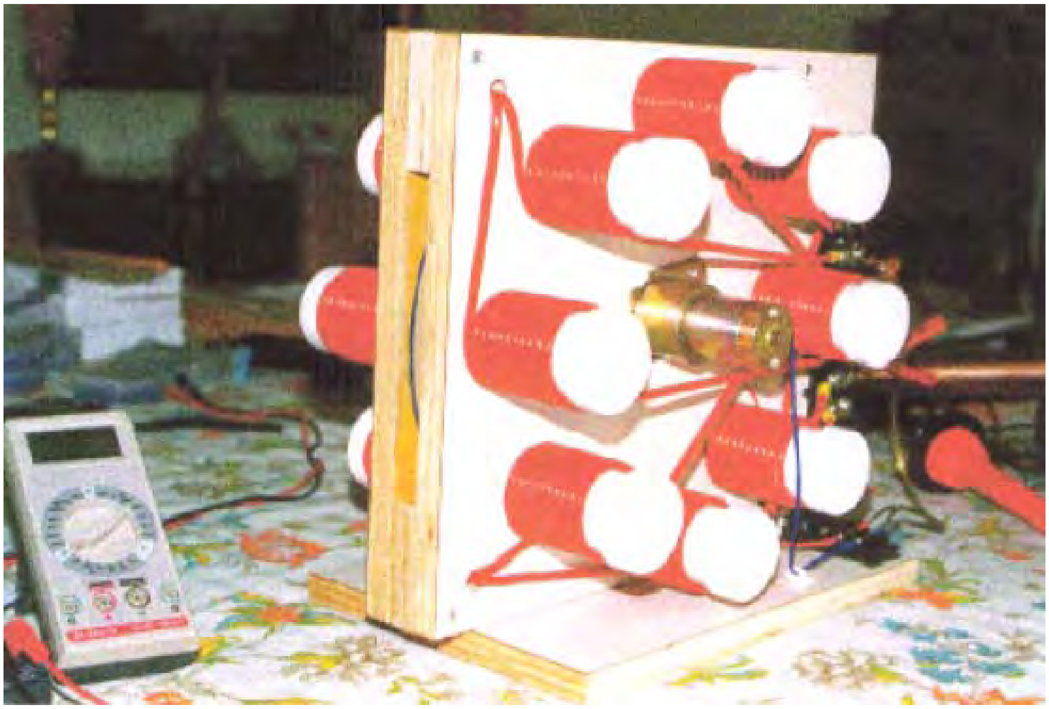

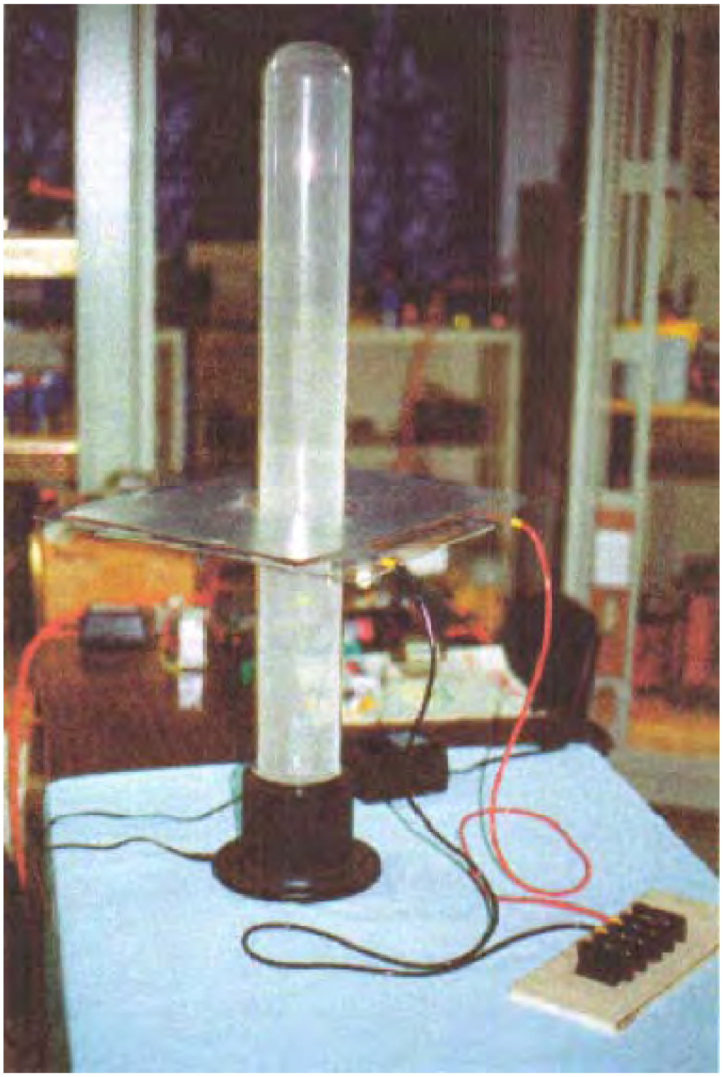

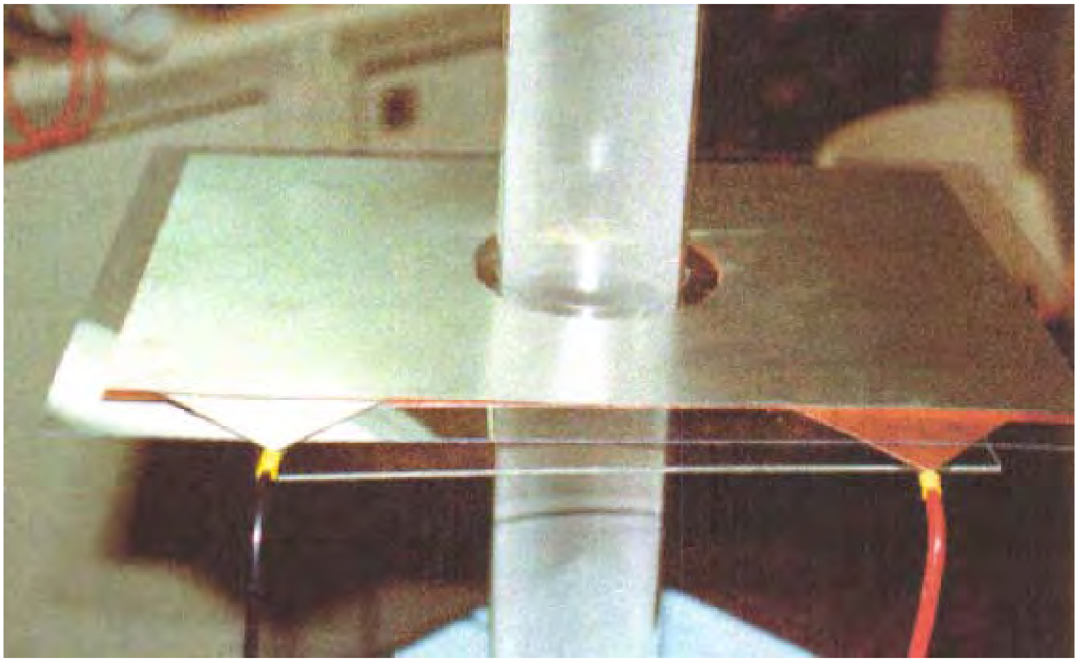

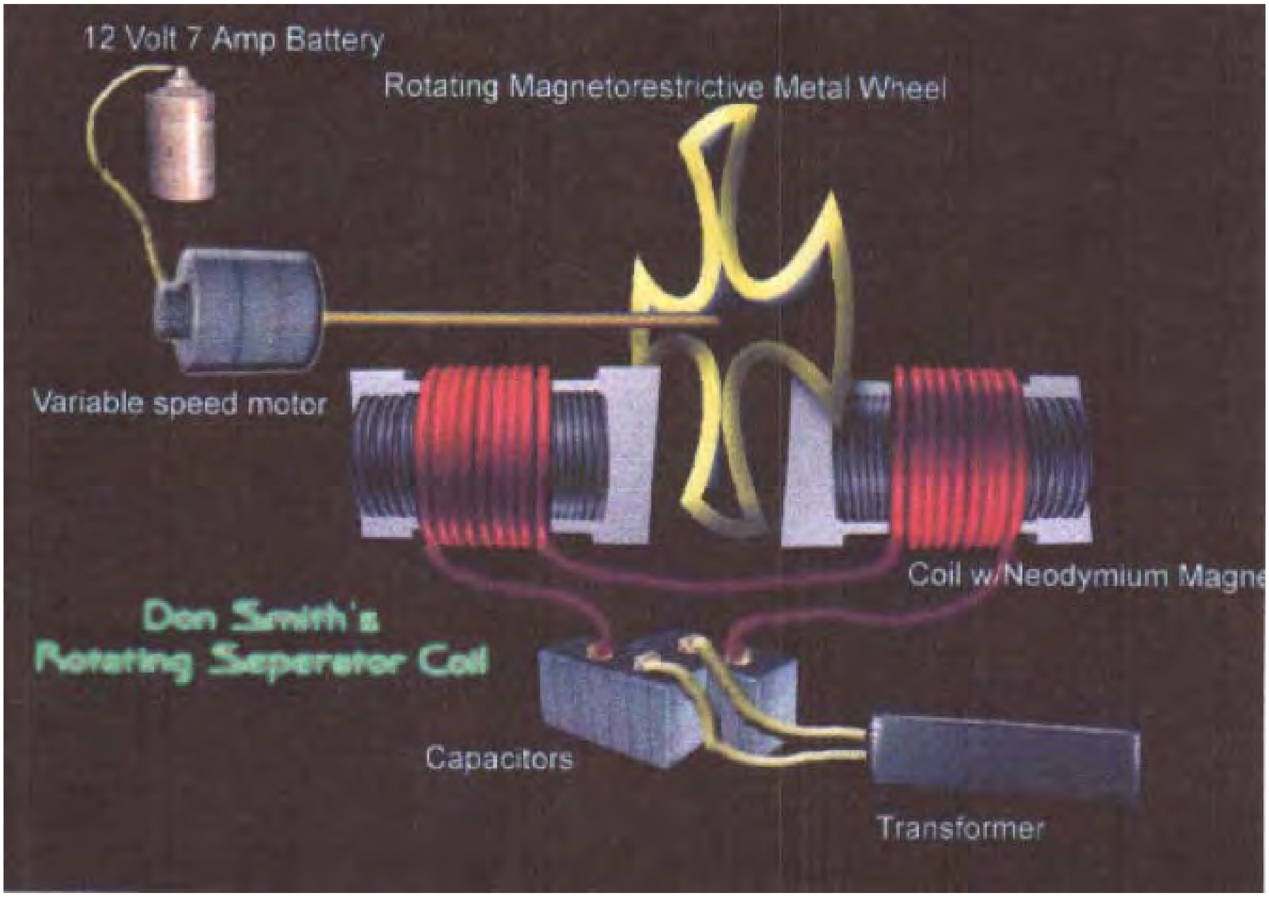
12 Volt 7 Amp Battery
Rotating Magnetorestrictive Metal Wheel
Variable speed motor
Coil w/Neodymium Magne
Don Smith's
Rotating Seperator Coil
Capacitors
Transformer

DIPOLE TRANSFORMER GENERATOR
DESCRIPTION

TECHNICAL FIELD:

The Invention relates to loaded Dipole Antenna Systems and their Electromagnetic radiation. When used as a transformer with an appropriate energy collector system it becomes a transformer generator. The invention collects and converts energy which, with conventional devices, is radiated and wasted.

BACKGROUND ART:

An International search of Patent Databases for closely related methods did not reveal any prior Art with an Interest in conserving radiated and wasted magnetic waves as useful energy.

DISCLOSURE OF INVENTION:

The Invention is a new and useful departure from transformer generator construction, such that radiated and wasted magnetic energy changes into useful electrical energy. Gauss Meters show that much energy from conventional electromagnetic devices is radiated back into the ambient background and wasted. In the case of conventional transformer generators, a radical change in the physical construction, allows better access to the energy available. It is found that creating a dipole and Inserting capacitor plates at right angle to the current flow, allows magnetic waves to change back to useful electrical (coulombs) energy. Magnetic waves passing through the capacitor plates do not degrade and the full impact of the available energy is accessed. One, or many sets of capacitor plates, may be used as desired. Each set of plates makes an exact copy of the full force and effect of the energy present in the magnetic waves. The originating source is not depleted or degraded as is common in conventional transformers.

BRIEF DESCRIPTION OF THE DRAWINGS:

The Dipole at right angle allows the magnetic flux surrounding it to intercept the capacitor plate, or plates, at right angles. The electrons present are spun in such a way that the electrical component of the electrons is collected by the capacitor plates. Essential parts are the South and North component of an active Dipole. Examples presented here, exist as fully functional prototypes, and were engineer constructed and fully tested for utility by the Inventor. Corresponding parts are utilized in each of the three examples as shown in the Drawings.

DRAWING 1 OF 4: VIEW OF THE METHOD

N = North and S = South of the Dipole

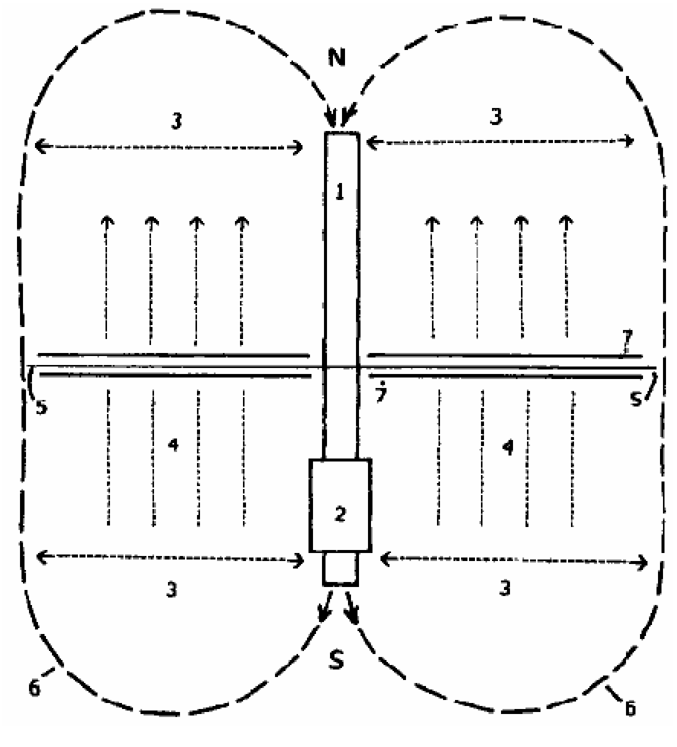

1. North and South component of the Dipole.
2. Resonate High Voltage induction coil.
3. Dipole's electromagnetic wave emission.
4. Heaviside current component.
5. Dielectric separator for the capacitor plates
6. For purposes of the drawing, a virtual limit of the electromagnetic wave energy.
7. Capacitor plates, with dielectric in between.

DRAWING 2 OF 4: COMPONENTS, 2A and 2B

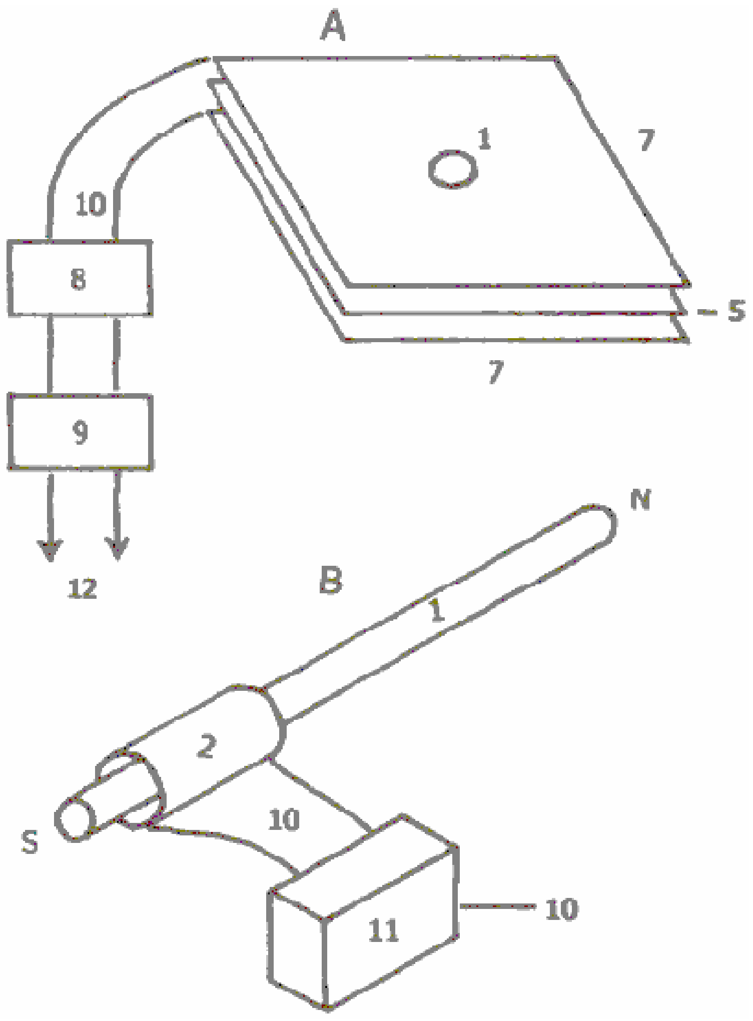

Fig.2-A:
1. Hole for mounting Dipole **B-1**.
2. Resonate high voltage induction coil.
5. Dielectric separator, a thin sheet of plastic separating the capacitor plates.
7. Capacitor plates, upper plate is aluminum and lower plate is copper.
8. Battery system, deep cycle.
9. Inverter. Input: Direct Current, output: 120 Volts at 60 cycles per second.
10. Connector wires.
12. Output to point of use being the load.

Fig.2-B N = North and S = South component of the Dipole
1. Metal rod, being soft magnetic metal such as iron.
2. Resonate high voltage induction coil.
10. Connector wires.
11. High Voltage input energy source such as a neon tube transformer.

DRAWING 3 OF 4 : Proof of Principle Device, using a Plasma Tube as an active Dipole.
N = North and S = South Components of the active Dipole.

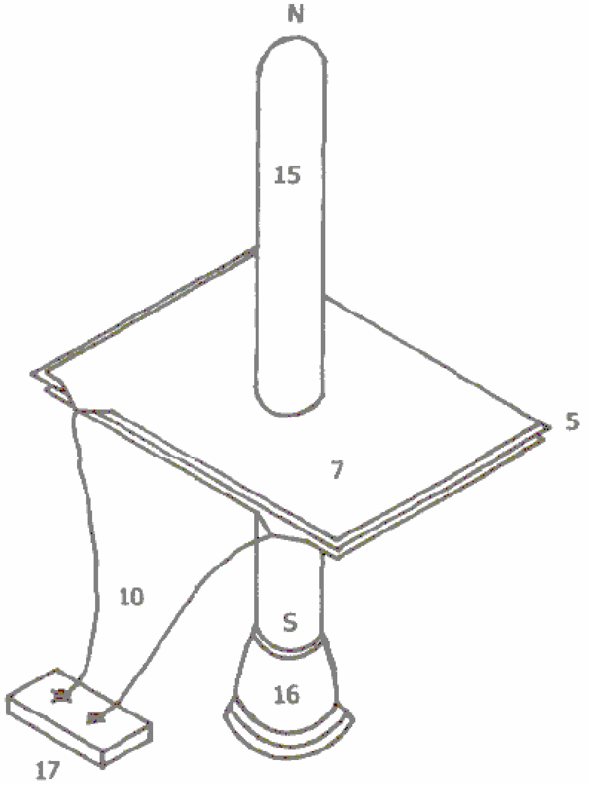

5. Dielectric separator of the capacitor plates.
7. Upper capacitor plate: upper plate is aluminum and lower plate is copper.
10. Connector wires.
15. Plasma Tube, 4 feet long and 6 inches in diameter.
16. High Voltage Energy source for the active Plasma Dipole.
17. Connector block: outlet for testing and use.

DRAWING 4 OF 4: Manufactures Prototype, Constructed and fully tested.

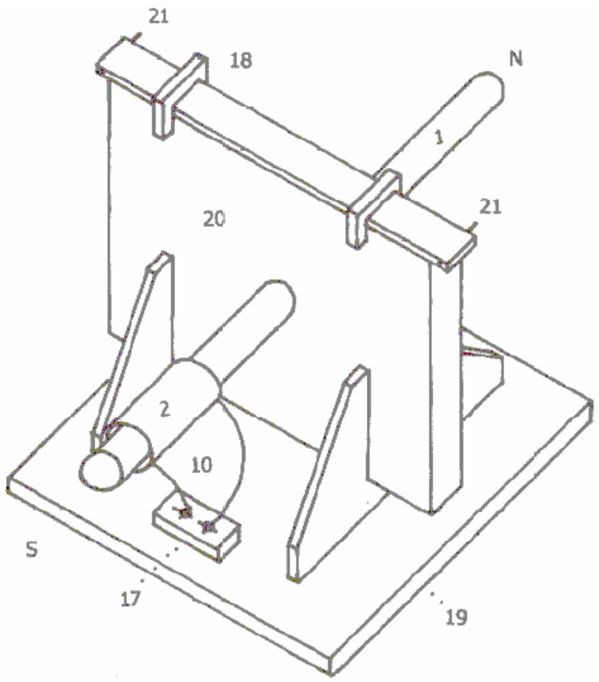

1. Metal Dipole rod.
2. Resonate High Voltage induction coil.
10. Connector wires.
17. Connector block for Input from high voltage energy source.
18. Clamps for upper edge of capacitor packet.
19. Support Device for The Dipole Transformer Generator.
20. Packet of Capacitor Plates.
21. Output connectors of the capacitor, producing energy into a deep cycle battery which then powers the inverter.

BEST METHOD OF CARRYING OUT THE INVENTION:

The Invention is applicable to any and all electrical energy requirements. The small size and high efficiency makes It an attractive option. It is particularly attractive for remote areas, homes, office buildings, factories, shopping centers, public places, transportation, water systems, electric trains, boats, ships and all things small or great Construction materials are readily available and the skill level required is moderate.

CLAIMS:

1. Radiated magnetic flux from the Dipole, when intercepted by capacitor plates at right angles, changes to useful electrical energy.
2. A Device and method for converting for use, normally wasted electromagnetic energy.
3. The Dipole of the Invention is any resonating substance such as Metal Rods, Coils and Plasma Tubes which have interacting Positive and Negative Components.
4. The Resulting Heaviside current component Is changed to useful electrical energy.

ABSTRACT

An Electromagnetic Dipole Device and Method, wherein, radiated and wasted energy is transformed into useful energy. A Dipole as seen in Antenna Systems is adapted for use with capacitor plates such that the Heaviside Current Component becomes a useful source of electrical energy.

TransWorld Energy
227 W. Airtex Blvd.
Houston, Tx. 77090

September 23.2002

Phone: (281)876-9200
Fax: (281)876-9216

Web-site: altenergy-pro.com
E-mail: donsm1@earthlink.net

Dear Reader:

TransWorld Energy is dedicated to improving the Human Condition in the Field of Energy which, at the same time, makes possible Healthy Water and increases the food Supply. A never-ending source of energy found throughout the universe is easily accessed with the minimum of effort and cost. The technology for doing this has been around since the 1820s. Selfish special interests have made sure that the technology remains discredited. People who control the Energy Sources control the World.

Extensive research and development by TransWorld and Associates has been progressing for more than 15 years. Numerous successful Energy Producing Devices have been produced and demonstrated throughout the World. Some of these can be viewed by the Web Site located using any major search engine (such as Lycos, Yahoo, Altavista, NorthenLight and more than 2,000 others throughout the World).

The Book which You are viewing has more than 40,000 copies in circulation. It has been translated and distributed in all major languages including Japanese, Arabic, Portuguese, French, Italian, Russian, Chinese, German, Spanish and many more. There are seven editions in circulation. An enormous interest is evident in the subject matter. An average of about fifty e-mails per day are received from the ends of the Earth (that is about 1,500 per month).

Once the Web Site and the book are viewed, it will become evident that abundant, self-sustainable energy is available everywhere for the taking. This is natural energy which does not harm the environment or those using it. The proper Device for Collecting is all that's required.

The Good News is that the problem is solved and with assistance, an ultimate source of energy which is environmentally benign, abundant throughout the universe and inexpensive to capture, is there for the taking.

Thank You for your consideration

Donald L. Smith, CEO

MEM

Electrical Energy Generating System

Description and Function:

The Generation of Electrical Power requires the presence of electrons with various methods of stimulation, yielding magnetic and electrical impulses, collectively resulting in Electrical Energy (Power). In place of the mechanical - coils and magnet system, present in conventional electrical power generation, visible moving parts are replaced by resonate magnetic induction, using radio frequency. Transfer of energy by resonate induction is related to the ratio of the square of the cycles per second.

The Energy System, presented here, operates at millions of cycles per second verses the conventional 60 C.P.S. This tells us that it has a frequency advantage over conventional methods. This same advantage applies to the amount of electrical energy output. Therefore the Device is small in size and produces large amounts of Electrical Energy. The Electrons acquired, are from the surrounding Air and Earth Groundings, being the same source as in conventional methods. This is accomplished by magnetic resonate radio induction.

Applications:

This Electrical System adapts nicely to all Energy Requirements. It is a direct replacement for all existing Energy Systems. This includes such things as Manufacturing, Agricultural, Home Usage, Office Complexes, Shopping Centers, Rail Transportation, Automobiles, Electrical Power Grids, Municipalities, Subdivisions, and Remote Areas. Briefly, the only limiting factor is the imagination.

Economic Possibilities:

No Historical Reference Point exists for a comparison of the Possibilities of this System. One can see from the impacted applications listed above, that the magnitude exceeds any known invention, presently a part of the Human Experience.

Present and Future Plans:

The Energy System has been in the developmental stage during the past seven years. It is Patent Pending # 08/100,074 with the Patent Office. No prior art exists according to the Patent Office's response. The System is presently being introduced into the World Market.

Useful energy occurs as the result of imbalances in the ambient background energy, which is a transient phenomena. In the electrical field, it is a closed system subject to heat death, which severely limits it's utility. The flip side of the electron, produces magnetic waves which are an open system, not subject to heat death. These waves, being unrestricted, are the universal source of energy when unlimited resonate duplicates from this one source are available. Therefore, the key to unlimited energy, is Magnetic Resonance. In order to understand this, requires putting a stake through the Heart of Antique Physics. Non-linear and Open Systems are universally available in Magnetic Resonance Systems, Explosions of any sort [including Atomic Explosions] and Combustibles of any type. Mechanical equivalents would be levers, pulleys and hydraulics. A highly obvious example is the Piano where the Key impacts the one note giving one sound level, which resonates with it's two side keys providing a much higher sound level. Magnetic Resonance Energy clearly amplifies itself, demonstrating more energy out, than in.

Ohmic resistance does not apply to Magnetic Resonance which travels unrestricted for great distances, therefore multitudes of electrons are disturbed, and their back-spin translates magnetic into usable electric energy. The right angle component which the magnetic flux provides, translates into useful electrical energy. Taken at right angles, the Magnetic Dipole provides an unlimited source of electrical energy. The writer is recognized world-wide for his knowledge and experience. See his Web Site at altenergy-pro.com .

Gravity is a function of spin phenomenon as observed in gravity separation of liquids. When spun, milk and cream separate. Therefore, relative specific gravity is function of mass versus spin. Magnetics and gravity are both spin related. In part, a top levitates when spun. Therefore, spinning magnetic fields are a functional motor source as in flying saucers.

ABSTRACT: Technology of New Energy:

Developments in the understanding of Electricity, along with Materials which were not previously available, allows the construction of Devices which collect energy in large quantities, from the Earth's Ambient Electrical Background. This Energy is naturally occurring, environmentally benign and is available everywhere. It is available wherever and whenever it is required. New Devices use Resonate Magnetic Waves which replicate upon spinning the locally present electrons, providing multiple duplicate copies of the Energy Present. Each electron when spun yields both magnetic and electric waves in equal proportion. The electrical component is a closed system limited by Ohms Law. The magnetic component is an open system not limited and it replicates multiple copies of the energy present.

Special materials and recent developments allow the magnetic energy to reproduce, through resonance, unlimited duplicate copies acquired from the ambient background. These Devices harvest the energy that has been, and is always present universally. Conventional methods consist of coils and magnets systems. Upon moving past each other, the magnetic flux field disturbs electrons which yield electricity, which is collected by the coils system. This is accomplished electronically with the new technology, without any moving parts and the energy is multiplied such that the Device becomes self-sustaining once it is started. This Technology, already presented Worldwide, will be shown at the Conference.

Dr. Smith
www.altenergy-pro.com
e-mail donsml@earthlink.net

"Putting a stake through the Heart" and thus removing the mental block created by antique physics is required. Conditions wherein this becomes necessary are non-linearity, resonance and explosions of any sort. Combustibles of any sort such gasoline and atomic explosions are good examples wherein more energy out than in, is obvious. You can add to that the non-linearity found in pulleys, hydraulics, steam power and suchlike. Magnetic resonance is a highly obvious source for multiplying energy output. The sound system present in the piano, demonstrates this very clearly. Energy amplification clearly present in the above, demonstrates the silliness attested to by many Physicists.

Ohmic resistance does not apply to magnetic resonance which travels unrestricted for great distances, therefore multitudes of electrons are disturbed, and their back-spin converts from magnetic energy to usable electric energy. These same electrons have been around from the beginning of time and they are undiminished and will remain so until the end of time.

ELECTRICAL ENERGY SYSTEMS PREFACE

Useful Electrical Energy is obtained directly from electron spin induced by incoming magnetic waves, or indirectly through mechanical exchange as in dynamo type devices. Simply put, electron spin converts from magnetic to electrical energy and vice versa. Nature provides grand scale magnetic wave induction throughout the universe, for free. In Electrical Systems, movement is at right angles to the direction of current movement. This explains the rotary movement of the Earth and other related Systems. The rate of Spin for the Earth is known as well as the mass (5.98 x 10^{24} Kg - "Physics for Scientist and Engineers", by Raymond A. Serway, Saunders College Publishing, 2nd Ed. page 288, Table 14.2), therefore the amount of incoming Electrical Energy which produces this action can be calculated.

It can be seen quite easily, that the incoming magnetic wave energy is Vast and Continuous. As an accretion mass, the Earth is an Energy Sink, getting it's energy from elsewhere, being Cosmic, Galactic and Solar. Conversion of incoming magnetic waves into electrical energy provides an unending, inexpensive and environmentally friendly source available to all. Cosmic and Galactic Energy is available twenty four hours per day. Large amounts of this Energy accumulates in the Earth's radiation belts. This Giant Energy Storage, when properly understood, provides a major source of free unending electrical energy. Each of My Inventions plugs into this vast energy source.

A perverse, Intentional Ignorance on the part of the Establishment, prevents recognition of the importance of the Energy Systems shown here. Any new system which is favorable towards the masses, is considered as disruptive, and therefore not allowed. Those who have the (Gold) Energy Rule (Golden rule) Mandated Destruction of all Humanity is not a consideration.

This Presenter will remove some of the Fog placed with the intention of preventing the recognition of this unending, environmentally clean, electrical energy Source, which is present everywhere throughout the Universe. The Cost of Harvesting and Using this Free Energy is a function of Human Stupidity.

RESONANCE CIRCUITS DEMO

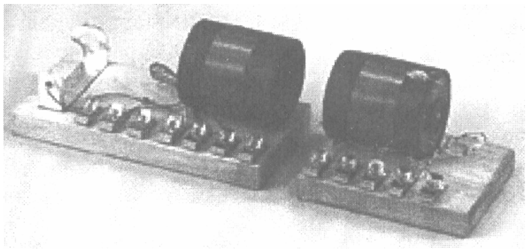

Used to demonstrate electromagnetic radiation between two UC circuits - one a transmitter and the other a receiver. When the 1.5 volt power transmitter is pulsed, the radiated signal is picked up by the remote receiver circuit which then lights up a 70 volt neon lamp.

With this apparatus, the student quickly understands some basic principles governing wireless communication, broadcasting, etc.

Kit: #10-416 $49.95

THE SCIENCE SOURCE
WALDOBORO, MAINE 04572
P.O. BOX 727

Tel. 1-800-299-5469
info@thesciencesource.com

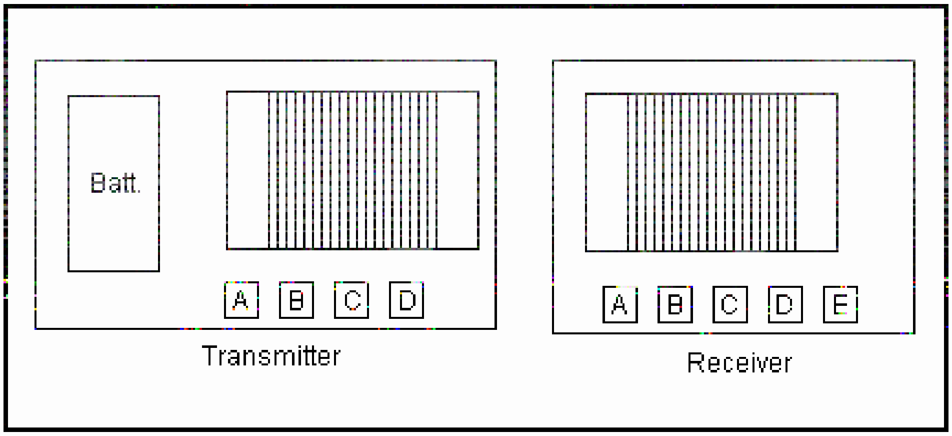

Diagram of transmitter and receiver coils

ULTIMATE ENERGY SOURCES

A human is a speck of dust on Earth, the Solar System is a speck of dust in the Galaxy and in turn, the Galaxy is a speck of dust in the Universe (Cosmos). All of these respectively represent vast ambient energy reservoirs. Awareness of the Sun, opens doors into other energy sources. Electromagnetic Energy which is present everywhere throughout the Universe, is accessed by catalytic activity, directly as in Solar Cells or indirectly as by mechanical means. Resonate, Magnetic Waves (Faraday's "Action at a Distance") allow Energy Activation Transfer to remote points of usage. The method of capture and use of this Energy is optional, and therefore it's cost is a function of Human Stupidity (Free-Energy).

Direct access is more desirable, and technology transfer from Solar Cell-type Devices provides the Catalyst. Enormously high Ambient Energy Levels are not detected by instruments that use the Ambient Background as a Reference Plane. A spoonful of water lifted from the Ocean does not define the Ocean. Incoming magnetic waves are reflected, Deflected or absorbed. Deflected Magnetic Waves spin electrons sideways producing useful Electrical Energy. Absorbed Wave Energy produces heat, therefore a hot interior for the Earth. In Electrical Systems physical movement is in the direction of current flow, frictional drag from inflow current defines gravity. Accretion masses resulting from Energy Sinks, provide all solid entities with their respective gravitational effect.

Increasing the tolerance level for Intellectual Awakening opens Doors of Reality. These doors blink into, and out of existence, and upon recognition, benefit Mankind. Opening some of these Doors, which at the present time are seen through a deep fog, is our purpose. Exploring Unrecognized Energy Sources, which are a Part of the Ambient Background, is another goal. Our Available Instruments do not use reference planes which allow recognition of this energy, as we shall see, vast Energy Sources that totally surround us are available through Technology Transfer. They are inexpensive (Free), fully self-renewable and environmentally benign.

Incoming Magnetic Wave Energy with Faraday's "action at a distance" will be looked at closely. Particle Physics will be left for the Astrophysics. Excited Electrons at point "A" the Sun (including the Galaxy and Cosmos) do not travel to point "B" the Earth, however a corresponding action occurs at point "B". The Electrons being disturbed at the Central Power Plant, in the same manner excite the Electrons at Your House, upon switching into an Earth grounding (known as "flipping the switch"). Correspondingly, there are Four Major Power Sources providing enormous amounts of Ambient Background Magnetic Wave Energy. They are The Cosmic, Galactic, Solar and Earth's Ambient Electromagnetic Backgrounds. The Earth's Electromagnetic Field comes from reflection, deflection and absorption as a result of action at a distance from the above.

Prescription Physics mandates that the Earth's background is of little interest. When we have Considered the evidence herein, it will become obvious that Special Interest's effort at keeping the People ignorant has, until now, largely succeeded.

Information for the entire World is available regarding the Magnetic Flux Background of the Earth's Surface (United State's Geological Survey, Colorado, USA, Office). When examined and properly understood, these Maps yield important information regarding reflection, deflection and absorption of incoming Magnetic Waves, plus action at a distance. When properly understood, these Maps reveal a very large Ambient Electromagnetic Energy Source. This is the Part of the Earth's Energy System that relates to the Bird on the High Voltage Line. When deflected, magnetic flux from electrons changes to electrical flux, providing the Motor System that spins or rotates the Earth. Physical movement by electrical systems is from inflow current movement. What level of current movement is required to spin the Earth? The Earth's Mass is 5.98×10^{24}. From this Information, the Watts of Electricity Required may be calculated! Absorbed microwave flux energy heats from the inside out, therefore a hot interior of the Earth results. Water is strongly diamagnetic, and on windless days, ocean waves provide visible Proof of the overhead incoming magnetic flux. From the information above, the Earth's weight and rate of spin allows the calculation of the amount of incoming ambient background energy required. As You can see, it is not inconsequential as Prescription Physics mandates.

Astrophysicist are concerned with charged particles that whiz by, once every one hundred years, rather than Wave Phenomenon associated with action at a distance. This highly Active Wave Energy translates into Electrical Energy at point "B". The Galaxy is alive With Energy which is billions of times greater than that of the Sun. Visible Light is a very tiny part of the Electromagnetic Energy Spectrum. Frequencies present in the Galaxy and Cosmos allow Radio Telescope photographs of their existence and magnitude. One such 408 MHz photograph of the Electromagnetic Energy Spectrum shows that the Earth is a tiny speck of dust in this Enormous Ocean of Energy, and can be seen near the left end of the Central High Energy Area.

This Energy extends in all directions. Accretion and formation of Planets, Suns and Galaxies are results of energy sinks and variable sized black holes. Mass retains heat, and is cooked from the inside out by the microwave background energy provided by the Universe. Flux movement into energy sinks, provides the frictional force know as gravity. Spinning mass in the presence of incoming flux amplifies the gravitational effect.

At present, only Solar Energy is recognized. It is inconsistent, flaky and a very small Part of the Magnetic Wave Energy Present. Technology Transfer from Solar Power provides uncomplicated and inexpensive, direct access to the Other Greater Energy Sources. All Electromagnetic Energy harvesting methods include a Catalyst, a Collector and a Pump. Catalysts include sensitization through doping with certain elements, air and earth groundings. Collectors include temporary storage as in Capacitors, Coils and Transformers. The Pump System includes induced movement onward to the point of use. Conventional rotating coils and magnet systems activate electrons present, such that action at a distance can occur, therefore it is an energy activation pump. In Direct Access Systems such as Solar Cells, the same occurs without mechanical action. Direct access occurs when Magnetic Waves impact a catalyst, spinning the local electrons sideways, producing useful electrical energy.

Indirect acquisition of electrical energy by mechanical means is wasteful, troublesome, expensive and degrades the environment. The dynamo is a combination collector and pump of energy which is collected from the Earth's Ambient Energy Background. Generators do not. make electricity, they collect it from the Ambient Background and forward it, as in Faraday's "action at a distance". Energy Conservation Laws relating to these systems, relate to gray areas, and when understood, are excluded because of the existence of External forces, open and non-linear systems as per Einstein. The Knowledge Base just viewed, provides a Direct Understanding of the Requirements for Harvesting of unending, fully renewable, environmentally benign Sources of Electrical Energy.

Magnetic Resonance Power System
Suggestions for Construction

This is the Basic Sonar Power System which permits submarines to see approximately 50 miles distance. What is not commonly known is that it works better at higher frequencies in the Gigahertz range. Any Device that can radiate 50 miles plus, is producing an enormous electromagnetic disturbance from a small input into a rod of magnetostrictive material. Disturbing the Earth's Ambient Background plus the strong dipole being produced, turns the magnetostrictive rod into a combination of a receiving antenna and a vastly superior output transformer.

The Drawing is only the Key Unit. A power input module and an output inverter circuit (diode bridge plus output transformer) are also required. The metal core and the wire size of the output transformer, plus adjusting the Earth Grounding of the Load, will determine the Amperage.

The Ideal rod material is Terfenol-D (check the internet). However a 1.5" diameter 10" long rod, costs over $5,000 each. Less expensive alternatives are obvious. When constructing, use PVC tubing with removable caps. Wind the coils on it and insert the experimental rod. Use only magnetostrictive material. When you get it right, you will have exactly what the Doctor ordered:

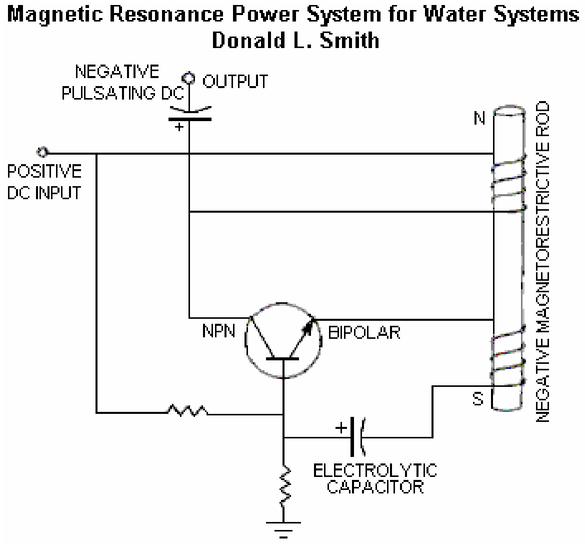

Magnetostriction oscillators work by magnetic resonance in a rod of magnetostriction material. This rod serves two purposes: It vibrates at the frequency of resonance oscillation, and it becomes the feedback transformer. Frequency is determined by items 4, 5, 6 and 8. The diameter, length and volume of the rod and output windings, determines the output. Item 2 provides feedback into the system. The negative magnetic character of item 8 plus the windings 2, in reaction to the magnetic flux field provided by 9, increases (amplifies or magnifies) the output. Magnetic permeability is the counterpart of negative resistance. Resonating with negative magnetic resistance, it pumps energy from the Earth's ambient background. Magnetic permeability is the ratio of flux density (Earth's B field) to the magnetizing force (H) in oersteds.

Magnetostrictive materials are piezoelectric in character, and have a very high resistance to electrical current flow. Examples are:

1. Permealloy Negative Magnetic Permeability	> 80,000
2. Sendust Negative Magnetic Permeability	30,000 -120,000
3. Metglas Negative Magnetic Permeability	> 200,000
4. Iron with (34%) Cobalt Magnetic Permeability	13,000
5. New Technology Magnetic Permeability	> 1,000,000

ELECTRICAL ENERGY SYSTEMS METHODS

1. **DIRECT -** Faraday's "Action at a Distance" incoming magnetic wave conversion to useful electrical energy. This includes Cosmic, Galactic, Solar and Magnets. Technology Transfer is from Solar Cell Technology.

2. **INDIRECT -** Electron Stimulation-Induced Electron Spin Systems, Electron Avalanche Pumping Systems

   Primitive, Indirect Conversion from another form of energy. Coils and Magnet as in Dynamo Systems (Closed Systems). Chemical Systems, Atomic, Pons & Fleischman and etc.

   Advanced, Direct Conversion, Magnetic Wave (Open Systems).
   Ambient Sources
   Air Core Coil Systems
   Gaseous Tube Systems,
   Solid State Marx Generator Avalanche Type Systems.
   Leyden Bottle Capacitor Types inserted in Lakes and other.
   Magnet Systems
   Electron Beam Antenna Systems

3. **TRANSFER MECHANISMS**

   Solids - as in metal conductors

   Gaseous as in radio wave transmission, a form of ionization.

   Sensitizing of Systems by use of Trace Doping with Radioactive elements, includes metal surfaces.

   Open Systems, non-linear with external forces. Albert Einstein in a direct quote from his biography states that these are excluded from the conservation of energy laws.

   Closed Systems Maxwellian Type Systems. Mathematics are predictable requiring deductive reasoning. Ohm's Law is King, and Establishment Intellectuals being comfortable with this, brand all else as a violation of the Laws of Nature by obtaining something for nothing. This is Dishonesty grand mal.

AMBIENT ENERGY SOURCES

Radiation System	*Diffusion Method*	*Magnetic Wave Energy*
1. Cosmic	**Reflection, Deflection and Absorption**	**Ultraviolet**
2. Galactic	**Reflection, Deflection and Absorption**	**Infrared**
3. Solar	**Reflection, Deflection and Absorption**	**Visible Light**
4. Earth	**Reflection, Deflection, Absorption,** **Faraday's "Action at a Distance"** **also, a Composite of all of the above**	**Earth's Electrical**

A deep fog pervades the entire Scientific Community with regards to the Significance of the Above Energy Sources. Magnetic Waves convert directly into Electrical Waves (useful electricity). Two sides of the electromagnetic system are always present and never separate. Local electron spin provides (action at a distance) the flip side of the incoming magnetic wave energy.

Enormous amounts of incoming magnetic wave energy becomes a part of the Ambient Background, and as such, cannot be measured directly. Reconstruction from indirect information, allows us to establish the actual energy levels which are present. Instruments provided by the Scientific Community measure only point "A" to "B", and when both are ambient, no potential energy is shown. This is the "bird sitting on the million volt power line and sensing nothing" approach. The Earth's actual ambient background has as it's Energy level multi-billions of Volts, which are conveniently and obliviously ignored by the scientific community. When properly understood, this enormous, never-ending source of environmentally-friendly energy becomes available.

ELECTRICAL ENERGY GENERATING SYSTEM

Patent Pending 08 / 100,074

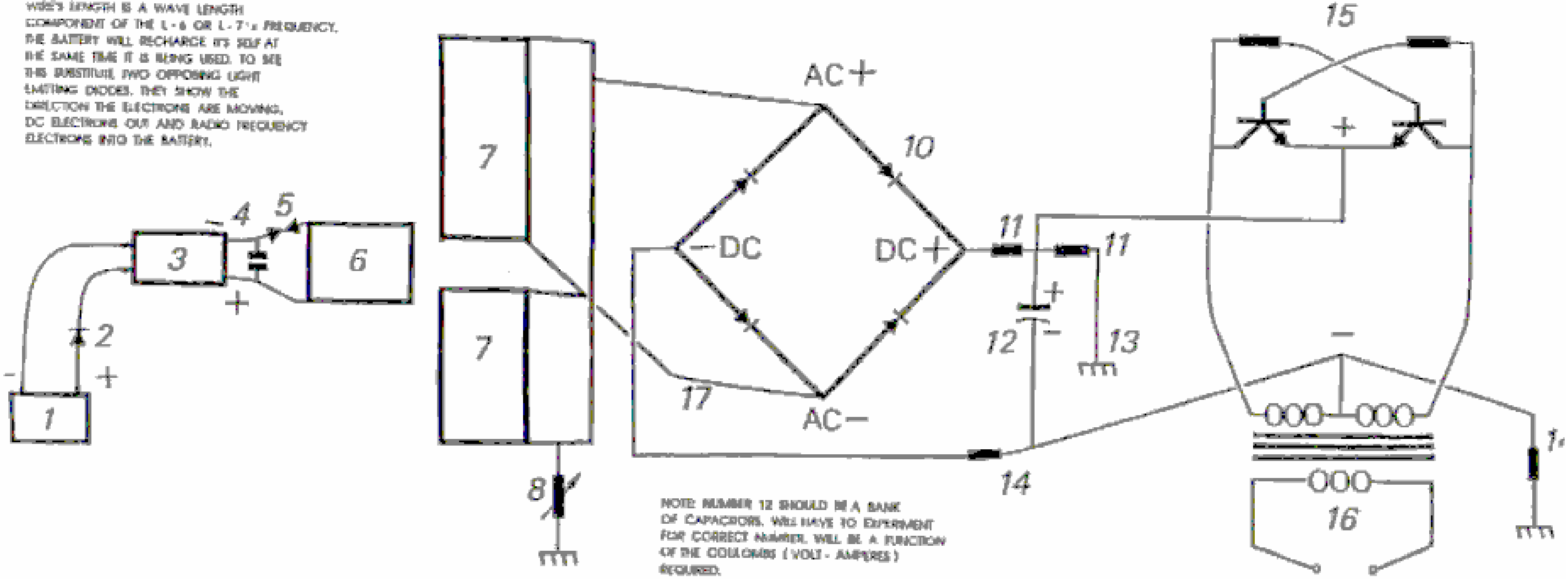

NOTE: IF THE PLUG CONNECTOR (2) WIRE'S LENGTH IS A WAVE LENGTH COMPONENT OF THE L-6 OR L-7's FREQUENCY, THE BATTERY WILL RECHARGE IT'S SELF AT THE SAME TIME IT IS BEING USED. TO SEE THIS SUBSTITUE TWO OPPOSING LIGHT EMITTING DIODES. THEY SHOW THE DIRECTION THE ELECTRONS ARE MOVING, DC ELECTRONS OUT AND RADIO FREQUENCY ELECTRONS INTO THE BATTERY.

NOTE: NUMBER 12 SHOULD BE A BANK OF CAPACITORS. WILL HAVE TO EXPERIMENT FOR CORRECT NUMBER. WILL BE A FUNCTION OF THE COULOMBS (VOLT - AMPERES) REQUIRED.

1. Gelcel, 6 or 12 Volt,
2. Diode, Pass. use a Varactor,
3. High Voltage Module, Constituting the L-1 and L-2 Coils.
4. Capacitor, TDK 10.9 Pf., 30 KV,
5. Spark Gap, Small Engine Spark Plug, Gap = .0025 In.
6. Induction Transfer Coil L-3.
7. Induction Receiving Coil L-4.
8. Voltage Control Shunt.
9. Frequency Adjustor, prevents derating by Diode Bridge
10. Diode Bridge, 200 Nanosecond, R.F., > 100 KV .
11. Voltage Divider Circuit, corrects voltage for next stage.
12. Capacitor, electrolytic, smooths out DC + ripple effect.
13. Earth Ground.
14. Voltage Divider Curcit, corrects voltage for Transformer
15. Inverter Circuit, DC + in and 60 CPS to Transformer
16. Output from Transformer to Load (Work).
17. Center Tap

17 JANUARY, 1997

Electrical Energy System

Don L. Smith, Energy Consultant

At a meeting between J.P. Morgan, Edison and Tesla, Tesla proposed an Electrical Energy System which could he connected into directly, without using a meter. Tesla's Idea of "Free Energy" was not compatible with their thinking. Courtesy of Morgan and Edison, from that day foreword, a complete and total bastardization of the Idea has been in progress. Agents for Morgan and Friends include the U.S. Patent Office and Academia. Academia's bad habit of incestuous quoting of each other, eliminates them as a possibility in cleaning up the mess. This selective ignorance, permeates throughout the study of electricity.

Many people, otherwise known as "intellectuals", have a total blackout and become jabbering idiots when "free-energy" is mentioned. The term has been amended to say, "something which was never there is being harvested and that this violates the laws of physics". For the selectively ignorant, this seems the way to run. Those who choose Morgan's drum beat, have severely limited the possibilities built into electricity.

This paper will be an exercise in creative understanding, in placing updated knowledge at your disposal. Whether it becomes a useful tool or is selectively ignored is your choice.

Electrons are defined as being the practical source of electrical and magnetic energy. The electron as a particle, was postulated by professor J. Thompson in early 1900's. It is now universally accepted that the electron exists and that it is the source of electricity. When the electron is agitated it produces magnetic and negative electrical energy. Physics as it exists today, cannot explain why the electron remains intact and is not diminished by the energy it releases. This is a part of the built-in ignorance provided by the Morgan and Edison Camp.

One volts worth of electrons, when cycled, yields one volts worth of electricity. This can be repeated continuously forever and it never deplete or diminishes the electrons in question. They simply return to their air and/or earth source, waiting to do the whole thing again and again. Therefore, electrical energy is available, anywhere and everywhere humans go. People who intercede for profit, set the cost of electrical energy. Otherwise, all electrical energy is free, Morgan and Edison be damned.

Improving upon Professor Thompson's postulation, other obvious characteristics can be seen to further define the electron. It has both magnetic and electrical emanations resulting from a right-hand and left-hand spin. Since magnetism and amperage come as one package, this suggest, that electrons in their natural non-ionic state, exist as doublets. When pushed apart by agitation one spins and supplies electricity and the other spins and provides magnetic (amperage) energy. When they reunite, we have Volts x Amperage = Watts. This Idea, until now, has been totally absent from the knowledge base.

The number of times that an electron is cycled, sets the collective energy potential present. The electrical equivalent of $E = mC^2$ is $E = (\text{Volts x Amperes}) \text{ x } (\text{Cycles Per Second})^2$. Those who choose, are now free to head for the bushes and make their usual contribution to humanity.

Prior to Tesla, there was a large group of people in Europe, who were building resonant coil systems for medical use. Amperage was dangerous in their coil systems. The Tesla Coil is only the Voltage half of their coil system, as will be demonstrated.

A short list of those (from 1860 onwards) active in resonate high frequency coil systems include; the Curies, Roentgen, Ruhmkoff, Oudin, Hertz, Levassor, Dumont, D'Arsonval and many others.

Peugeot, Panhard-Levassor, Bollee, Renault and others had successful electric automobiles in production using A C. motors. Various electrically-powered airships, including the Dirigible "France" were in service.

D'Arsonval, Professor of Experimental Medicine at the College of France, invented the electrocardiograph, oscilloscope, amp and volt meters, thermography and numerous other medical applications of high frequency electricity. As early as 1860, he was building high frequency coil systems, which he used in his experimental work. There is a strong connection between the work of Tesla and the people mentioned above.

Electric vehicles of all sorts, dominated until the 1920s, when the electric starter motor made the internal combustion engine practical. Prior to that, upon cranking, it frequently would break the owner's arm. At that point the use of batteries as a source of power was replaced by oil.

The establishment's carpet has some rather large lumps under it. Coulomb's and Newton's inverse square law is politely ignored and it's opposite is allowed to have only the most abstract status. Without opposites we have no definition.

The source value of a remote flux reading, requires the squaring of the distance, times the remote reading, to obtain the original value. The opposite of this, being the derivations relate to Energy equals Mass times the Velocity constant squared. The electrical equivalent, being Energy equal capacitance times voltage squared and Energy equals induction times amperes squared. Flux lines increase as the law of squares and then activate electron energy which was not previously a part of the sum. The cumulative capacitance and inductance increase as the outer ends of a Tesla coil are approached, and this results in output energy being greater than the input energy present. This Energy is real. It can be safely measured by magnetic flux methods and electrostatic voltmeters, based on the inverse square law

As seen above, flux lines result both from induction-henrys-amperage and capacitance-coulombs-volts, and define electrical energy. The non-linearity of this system does not obey Ohm's law, which is replaced with impedance and reactance for alternating current systems. Impedance is the sum of the system's resistance to AC current flow, and this becomes zero at resonance. In resonant induction systems, a cycles-per-second increase, invokes a second round for the law of squares.

The degree to which flux lines are present, disturbs an equal amount of electrons, upsetting the ambient background energy, resulting in useful electrical energy being obtained. The frequency at which the disturbance occurs, increases the useful energy available, and it obeys the law of squares. Two square-law components, *flux density* and *frequency* are involved. Enter resonance which cancels the resistive effect.

Only the electrical energy which is either above or below the ambient level is useful. For the Central U.S. going east to west, ambient as approximated by electro-static voltmeters and flux methods is 200,000 volts on a solar-quiet day. At night time, the ambient energy level drops to about one half of the daytime value. On a solar-active day, it may reach more than five times that of a solar-quiet day. Ambient background energy at the polar regions, is approximately 500,000 volts on a solar-quiet day. The background energy varies as it relates to the North-South component and the East-West continuum.

This leaves us with an interesting problem. Electrons, when disturbed, first produce magnetic flux and then produce electrical flux when they spin back to their normal position. Therefore any electron movement produces above ambient energy, being over unity.

ELECTRICAL ENERGY WITH ASSOCIATED PHENOMENA

1. Current-amperes results from the unequal distribution of negativity (electrons).

2. Electron spin causes electrical current and magnetic lines of force.

3. Magnetic imbalance causes the gravitational effect. This is evidenced in electric motors by magneto-gravitational displacement of mess, which causes the motor to rotate.

ENERGY LINES QF FLUX (FORCE)
FIELDS & WAVES *

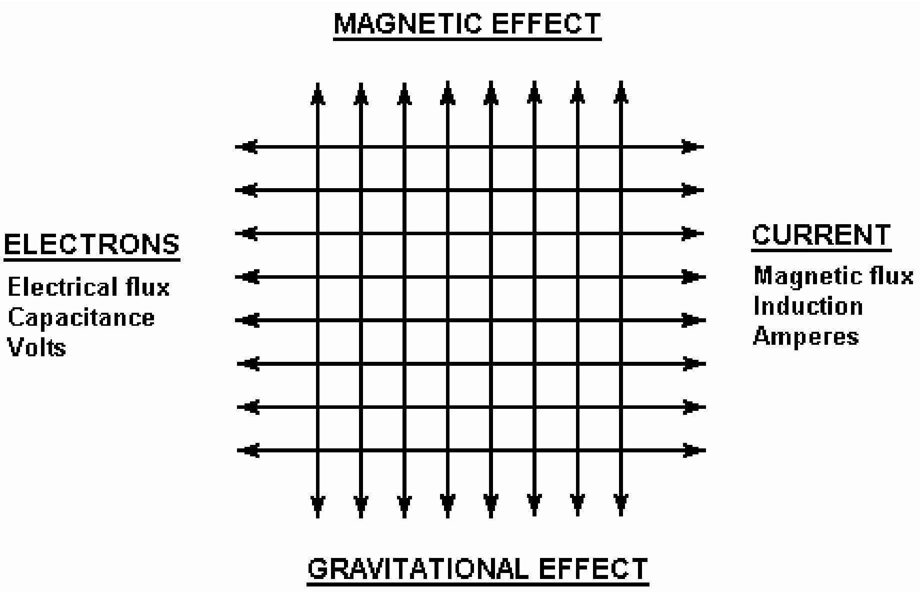
MAGNETIC EFFECT
ELECTRONS
Electrical flux
Capacitance
Volts
CURRENT
Magnetic flux
Induction
Amperes
GRAVITATIONAL EFFECT

* Below 20,000 Cycles Per Second = **Fields**
Above 20,000 Cycles Per Second = **Waves** (Radio Frequency)

Derivation of Magnetic and Electrical Power

Analogous Relationships:

1. Potential Power is present in a bar magnet as shown:

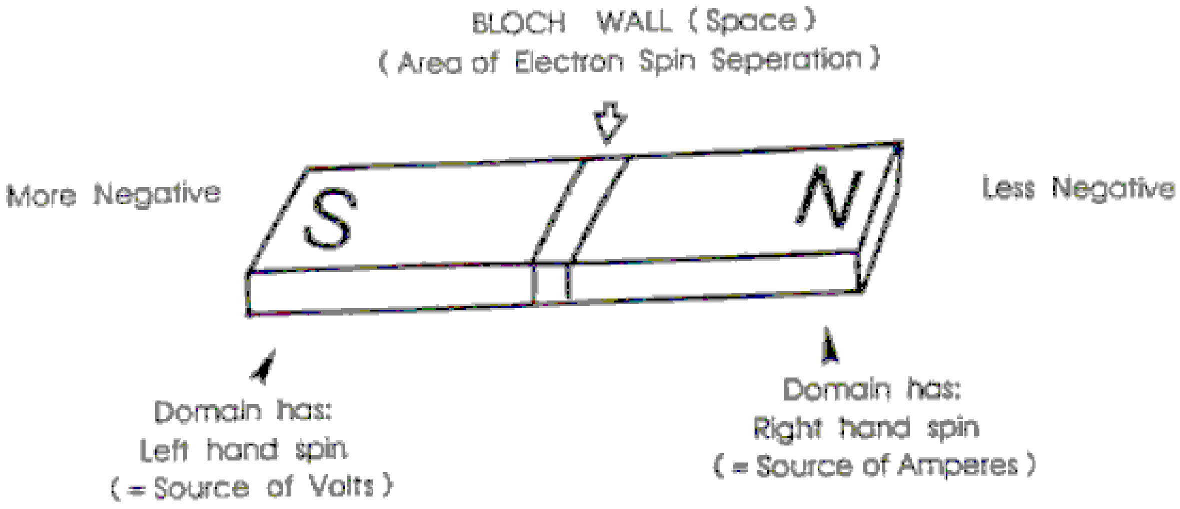
BLOCH WALL (Space)
(Area of Electron Spin Seperation)
More Negative
S
N
Less Negative
Domain has:
Left hand spin
(= Source of Volts)
Domain has:
Right hand spin
(= Source of Amperes)

2. The Source of these Electrons being from the Solar Plasma, are non-ionic and occupy all Free Space. They are commonly obtained from Earth and Air Groundings. They exist in Doublet Pairs, one being more negative than the other. The more negative one has a Left Hand Spin. The less negative one has a Right Hand Spin.

3. Resonate Electrical Coil Systems (Tesla) are Analogous to the System observed in the Bar Magnet (above). The Bloch Wall Area is Located at the base of the L-2 Coil. The Left Spin portion of the coil (Voltage Only) Coil predominates. The Right hand Spin portion of the coil (Magnetic-Amperage) is mostly absent.

Tesla Coil Geometry *

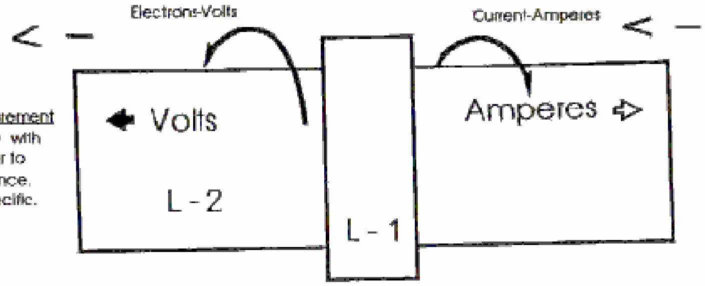

Correct Method of Measurement
Use the inverse square law with an Electrostatic Volt Meter to measure from a safe distance, they are frequency nonspecific.

Correct Method of Measurement
Use the inverse square law with Hal-Effect devices to measure from a safe distance, they are normally frequency nonspecific. The magnetic field they measure converts directly to Amperage

Volts dominate if the L-1 Coil is far right
Volts and Amperes about equal if L-1 Coil centered

This End Has Greater Voltage
Distributive Capacitance at Maximum
Capacitance-Coulombs-Voltage
Voltage has Electrons spinning to Left

This End Has Greater Amperage
Distributive Inductance at Maximum
Induction-Henrys-Amperage
Amperage has Electrons spinning to Right

Volts represent the more negative and Amperes the less negative. Therefore the electrical charge of more negative must seek the less negative to regain ambient.

* Contains proprietary information related to Patent Procedure
Geometry - properties of lines, surfaces and solids

Donald L. Smith
2 November, 1995

Induced Electrical Energy System

Collection and transfer of energy requires temporary storage, which occurs as capacitors and coils of a resonant circuit are cycled, on and off. The frequency at which the capacitors and coils are pumped, determines the amount of electrical energy that moves onward.

The amount of Energy transferred relates directly to the density of lines of flux present. The Kinetic Energy Formula is helpful in establishing the amount of energy present. This formula squares the velocity times mass. In the case of electrical energy, the intensity of voltage and amperes multiplied by the cycles per second, replace the velocity component.

Note that the "acceleration" of the Voltage "E" and Amperage "I", which increase as non-linear components, then obeys the Law of Squares.

Each unit of increase, causes a squaring of the flux lines present. The amount of energy transfer caused by this increase in flux lines is demonstrated below.

Increase in Flux Lines Present Symbolized

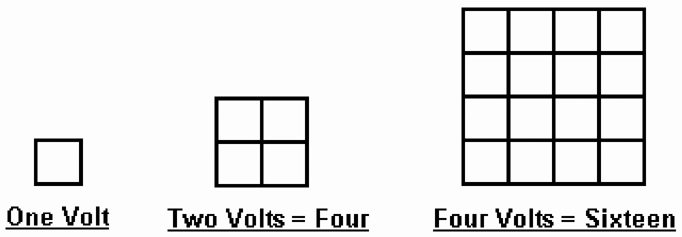

In resonant air-core coil energy transfer, the increase in flux lines present disturbs more electrons than previously, resulting in over-unity energy being present and available.

Energy stored, times the cycles per second, represents the energy being pumped by the system. Capacitors and inductors store electrons temporarily.

Capacitor formula: **W = 0.5 x C x E x Cycles per second** where:

W = energy in Joules (Watt Seconds)
C = capacitance in farads
E = applied potential in volts squared.

Inductor (Coil) formula: **W = 0.5 x L x I x Cycles per second** where:

W = energy in Joules (Watt Seconds)
L = inductance in henrys
I = current in amperes squared

Both one henry, and one farad, equal one volt. The higher the cycles per second, including the squaring of the flux lines, cause a large increase in the amount of energy being produced.

The above combined with a resonant energy induction system (where all electrons are moving in the same direction at the same time), make the next move into over-unity practical.

The dampening process of conventional electrical power generation, has all the available electrons bouncing randomly, mostly cancelling out each other. In that System, the useful energy available is a very small percentage of the energy which is present.

In the resonant induction system, a very high percentage of the energy present is useful. At resonance, (ohms-impedance-Z) becomes zero and all of the energy present is not degraded and becomes available to do useful work. "Ohms" is load or wasted energy, and "amperes" is the rate of that wasting of energy.

Using the previous information, if we now apply it to an air-core coil, resonant transformer energy system. L-1 and L-2 coils are now present. L-1 has a smaller number of turns and is several times the diameter of L-2. Input from a 12 volt high-voltage laser driver source, produces 8,000 volts with a low level of wasted energy, pushing amperage into, say, 4 turns of coil L-1. Each turn of the L-1 coil then acquires 2,000 volts of resonant potential. Consequently, each turn of L-2 is then exposed to the electric flux of 2,000 volts. Each turn at the bottom end of L-2 acquires 2,000 volts. The flux lines are squared and are additive as the voltage and amperage progresses towards the top end of L-2's large number of turns.

A huge number of additional flux lines which were not previously present become present at the top end of the L-2 coil. These flux lines excite the nearby electrons in it's earth and air and groundings. This high level of excitement above the ambient, causes a large number of electrons which were not previously a part of the energy present, to become available for use. At this point over-unity is present in large amounts.

The "bubble gum between the ears" response to this is: "this must be lots of volts but no amperes". Please recall that amperage is wasted energy, and that until that wasting occurs, there are no amperes. A good way to demonstrate this, would be to let the bubble gum crowd put their hands on the high-voltage end of the device while standing on wet ground (a people zapper). **Note: don't do this.**

This over-unity device produces energy at radio frequencies which range into the megahertz band. This allows the device to be small in size, and yet produce large amounts of energy. A megawatt-sized unit will sit comfortably on a breakfast table. This energy is changed to Direct Current and then switched to produce the desired working frequency AC.

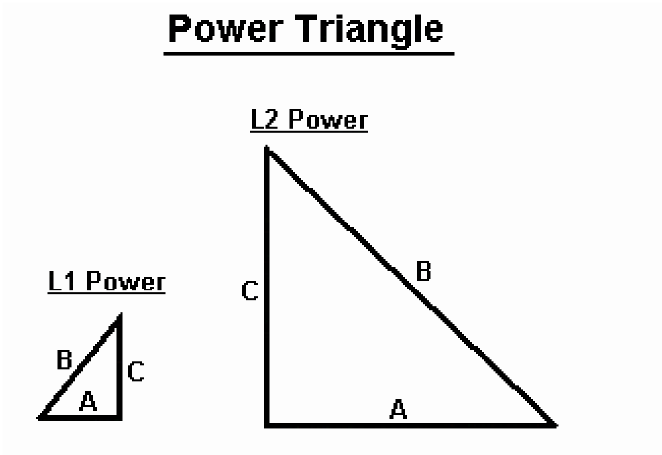

A: Volts x Amperes (the Available Power)

B: Volts x Ampres x Time (the Used Power)

C: Volts x Amperes x Recative (the Resinant Power)

1. Random movement of electrons in "A" and "B", mostly cancel each other out. This dampening, or wasteful concept of energy, is a source of much pleasure for the establishment.

2. "C" (Volt, Amperes, Reactive "V.A.R."), is the situation where all of the electrons move in the same direction at the same time. This results in near-unity energy output by resonant induction transfer.

3. Resonant induction transfer from one isolated power system, allows other resonant induction systems to duplicate the original source, which in no way diminishes the original source. Air-core coils (isolation-transformers) confirm this when they are a part of one of these functioning systems. A less perfect illustration would be the fact that the number of radio sets tuned to a particular radio transmission, does not alter the power required at the radio transmitter.

4. Resonant induction transfer, disturbs a large number of adjacent electrons which were not a part of the original input power source. The pulsating-pumping effect then draws in the newly available additional electrons into the on-going energy generation system. A near unity energy system of resonant air-core coils and the extra acquired electron-energy source constitute an over-unity system.

Electrical Power Generation / Points of Reference

Useful Electrical Power is Generated when Electrons from Earth and Air Groundings are disturbed by the movement of coils and magnets with reference to each other. The resulting electrical and magnetic energy is then changed to joules [watt-seconds: Volts x Amps x Seconds]. Each forward electron movement results in a magnetic impulse and each return movement causes an electrical impulse. The composite of the electrical energy impulses from these electrons yields useful energy [Power].

Let the above electron movement be represented by a room full of ping pong balls bouncing randomly. Most of the energy present cancels out by random impacts. This is the Classic Under-Unity approach to Electrical Power Generation, sanctioned by the Establishment.

In contrast to that, in the Electrical Energy Generation System presented here, the resonant Electrons are all moving in the same direction at the same time. This allows Near-Unity Electrical Power to Develop. This is the room-temperature equivalent of super conductivity.

The Energy System presented here, consists of a properly-adjusted and functional resonant air-core coil tank. The magnetic energy is stored in the coil system and the Electrical Energy is stored in capacitors. From Maxwell and others, we know that electrical-related energy has an equal amount of magnetic energy associated with it.

"The formula which establishes the Useful Energy of the System":

Joules = 0.5 x **C** x **V** squared x **C.P.S.** squared
units:
Joules (Volts x Amps x Seconds) Watt Seconds

C = Capacitance in microfarads
V = Potential in Volts
C.P.S. = Cycles per second

The transfer of Electrical Power by Resonant Induction is a direct function of the squaring of the cycles per second. For example, square 60 C.P.S. and then square the radio frequency C.P.S.s of the System here presented. Obviously, One Million Cycles per Second transfers more energy than Sixty Cycles per second. The Sanctioned Method of Electrical Power Generation uses the 60 C.P.S. Method. Using 60 C.P.S. and the random scattering of the Electrons System, assures the Establishment of it's desired Under-Unity Goal.

This random bouncing of the Electrons is the Ohms of Ohm's Law and is used to establish the rate of dissipation and/or Load [Work].

In the Resonant Tank Induction Energy Transfer System presented here, Impedance [system resistance] replaces the conventional ohm's usage. At Resonance, impedance becomes zero and the full force and effect of the Energy Transfer occurs. This is superconducting conditions at room temperature. At radio frequency the Electrons do not pass through the conductor as they do at lower frequencies. Instead, these Electrons encircle the conductor and are free of the conductor's resistance.

Let the Establishments Power Generation System be called 'A" and the System presented here be called "B".

With "A": Given 60 C.P.S. at 120 Volts using a 10 microfarad Capacitor:

Joules = [0.5 x 0.000010 x 120 squared] x C.P.S.s squared

(120 x 120 = 14,400)
[0.000010 x 14,400 = 0.144]
[0.144 x 0.5 = .072]
(0.072 x 3,600 = **259.2**]

Using the Inventor's Resonant Induction System, the Electrical Power available would then be 259.2 Joules [Watt-Seconds]. Using the Establishment's method only permits less than 10 Watt-Seconds of Useful Electrical Energy.

"B". Given One Million Cycles per second at 100,000 Volts, using a 10 microfarad Capacitor.

Joules = [0.5 x 0.000010 x 100,000 squared] x C.P.S.s squared

(100,000 x 100,000 = 10,000,000,000)
[0.000010 x 10,000,000,000 = 100,000]
(100,000 x 0.5 = 50,000]
(50,000 x One Million squared = **50,000,000,000,000,000**)

The useful Electrical Energy available is greater than 50 Mega Watts. Since the Resonant Electrons are non-impacting, all of the Energy is available for direct usage.

Benefits of the Inventor's System

1. Induction Energy transfer is enhanced by the squaring of the cycles per second produced by the System.

2. Induction Energy transfer is enhanced by the squaring the input voltage and amperage.

3. The increase of the flux lines occurring from the above, disturbing more electrons, causes more electrical energy to become available.

4. Resonant Induction has all of the Electrons moving unimpeded, resulting in superconductor conditions at room temperature.

5. A smaller amount of energy is used to disturb a larger number of Electrons. Electrons not originally a part of the System then contribute their energy, resulting in a net gain in available usable power.

6. The physical size of the System [Device] is small. The Device described in "B" sits comfortable on a breakfast table.

7. A small energy source is used to start the device and that source remains fully charged at all times by the System.

The Evidence Against Under Unity

1. Use of Logarithmic Scales on electrical measurement instruments. Linear measurement works fine where Ohm's Law applies (direct current). In alternating current, ohms are replaced by impedance and the measurements become non-linear.

2. Infinite "Q" at resonance confirms that voltage and amperage is squared, as in the kinetic energy formula. See the formulas of this report.

3. Square waves are clipped infinite "Q"s.

4. Maxwell and others show that magnetic-inductance-amperage and electrical-capacitance-voltage are two sides of the same coin. Magnetic-inductance is directly equal to amperage. Both obey the Law of Squares, which has over-unity built in.

5. Magnetic and electrical flux are present in enormous amounts at the outer ends of an operating Tesla Coil.

6. Ignorance of how to measure and relate magnetic and electrical flux, is the chief weapon of the under-unity gaggle.

7. The Cumulative inductance and capacitance of the Tesla Coil grounds itself out, if not properly utilized. See this report for the temporary energy storage accessible, when properly managed.

8. The Patent Office refers devices related to over-unity to their metering group, which is a sure indication that they are aware and accept the logarithmic measuring devices. This is direct and absolute evidence that they accept the square law as it relates to kinetic energy. This also indicates they are aware that over-unity exists. Since their bureaucratic brain is improperly motivated they continue to badger inventors who are working in the over-unity arena. Their level of intellectual dishonesty is sanctioned by, and is a real part of doing business with, a government which prides itself in being a hooliganistic bureaucracy.

Reading List

1. Alston, L.L. (Ed.), 1968, High-Voltage Technology, Oxford Univ. Press, London
2. Beck, E. 1954, Lightning Protection for Electric Systems, McGraw-Hill, N Y.
3. Bowdler, GAV. 1973, Measurements in High-voltage Test Circuits Pergamom, Oxford. L C. 72-86488
4. Brewley, L.V., 1951, Traveling Waves in Transmission Lines Pub: John Wiley. N.Y.
5. Chapman, S. and Barteis, J . 1940, Geomagnetism, two volumes, 1,000 plus pages, Oxford at the Clarendon press, England
6. Craggs. J.D. and Meek, J.M. 1954, High-Voltage laboratory Technique., Butterworth, London
7. EHV Transmission Line Reference Book, 1968 Edison Elcctric Institute, N.Y.
8. Farno, R.M., Chu. L.J., and Adler, R.B., 1968, Electromagnetic Fields, Energy and Forces, Pub: M.I.T Press, Cambridge, Mass.
9. Feinberg, R. (Ed.) 1979, Modern Power Transformer Practice, McMillan. London
10. Frungel, F., High Speed Pulse Technology, Vols 1 and 2, Pub: Academic Press. 1965, N.Y., and London
11. Gallagher, T.J. and Pearmain, AJ. 1983, ISBN 0-471-90096-6 High Voltage Measurement and Design, John Wiley, N.Y.
12. Hague, B., Alternating-Current Bridge Methods, 5 th. ed. Pub: Sir lsaac Pitman and Sons, 1959, London.
13. Hawley, W.G. 1959, Impulse-voltage Testing, Chapman and Hill, London.
14. Hayashi, Ch., Non-linear Oscillations in Physical Systems, Pub: McGraw-Hill, 1964. N.Y.
15. Henny, K. 1933-1959, Editor-in-Chief, Radio Engineering Handbook, five different copyrights, L.C. 58-11174, McGraw-Hill. N.Y.
16. Hudlestone, R.H. and Leonard, S.L., Plasma Diagnostics Techniques, Pub: Academic Press. N.Y.
17. Jacobs, J A., Editor, Geomagnetism, (a massive work) 3 large volumes, Pub: Academic Press, London.
18. Jeans, J.H. 1925, The Mathematics of Electricity and Magnetism, 5 th. Ed., Cambridge University Press,
19. Jones, B. 1972, New Approaches to the Design and Economics of EHV Transmission Plant, Pergamom, London.
20. Kind, D. 1978, An Introduction to High-voltage Experimental Technique, Viewieg, Braunschweig ISBN 3-528-08383-2
21. Knoepfel. H. 1970. Pulsed High Magnetic Fields, North-Holland, Amsterdam.
22. Kreuger, F H. 1964, Discharge Detection in High-Voltage Equipment, Temple Press, Heywood, London.
23. Kuffel, E. and Zacngl. W S. 1984, High-Voltage Engineering, Pergamom, ;London. ISBN 0-08-024213-8
24. Kupfmuller, K. 1957, Introduction to the Scientific Basis of Electrical Engineering, Pub Spinger. Berlin.
25. Lemon. H.B. and Ference, M. Jr., Analytical Experimental Physics " a major work" from The Ryerson Physical Laboratory at the University of Chicago, numerous copyrights 1933-1944. Printed as a text book by the University of Chicago Press
26. Lewis, I.A.D., and Well, F.H., Milli-microsecond Pulse Techniques Pub: Pergamom Press, N.Y. and London.
27. Malan, D.J. 1963, Physics of Lightning, English University Press, London.
28. Martin, T.L Jr., Physical Basis for Electrical Engineering, Prentice-Hall, N.J.
29. Matsusta, S. and Campbell, W.H . 1967 Physics of Geomagnetic Phenomena, Massive work in two volumes of 700 pages each. Academic Press, London.
30. National Physical laboratory, 1956, Notes on Applied Science ii 17, High Voltage Impulse Testing. HMSO.
31. Rather, H.. 1961, The Electron Avalanche and it's Generation, Pub: Vo. 33, Springer, Berlin.
32. Rokityansky, I.I., 1982. Geo-electromagnetic Investigation of the Earth's Crust and Mantle, Pub: Springer-Verlag, Berlin.
33. Salge, J., Pcier, D , Brilka, R., Schneider, D„ 1970, Applications of Inductive Energy Storage for the Production of Intense Magnetic Fields, Pub: Procedure of 6 th. Symp. on Fusion Technology, Aachen
34. Schulz, E.H., Anderson, L.T., and Leger, R.M., Experiments in Electronics and Communication Engineering, Copyrights, various, from 1943 thru 1954 LC 55-5033 Harper and Brothers, N.Y.
35. Schwab, AJ 1972, High-Voltage Measurement Techniques, The M I T. Press, Cambridge. Mass., ISBN 0262-19096-6
36. Sevin, L., 1965, Field Effect Transistors, McGraw-Hill, London.
37. Silsbee, F.B., 1942, Static Electricity, U.S. Department of Commerce, Washington. D C , U.S. Government Printing Office, (N.B.S. Circular C 438)
38. Smith, A. A., 1977, Coupling of External Electromagnetic Fields to Transmission Lines, ISBN 0-471-01995-X
39. Terman. F.E., Electronics and Radio Engineering, 4 th. ed., Copyrights 1932, 1937, 1947 and 1955 LC 55-6174 McGraw-Hill Book Co., N.Y
40. Thomas, R.T.,1970 High-Impulse Current and Voltage Measurement, Trans. I.E.E.E. EV1-19, pages 102-107
41. Thompson, Professor W, (titled Lord Kelvin) Reprint of Papers or. Electrostatics and Magnetism (original 1872) London.
42. Traister, R., 1983, The Experimenter's Guide to Solid State Diodes, Prentice-Hall, Inc. ISBN 0-13-295444
43. Valley, G.E, and Wallman. II . 1948, Vacuum Tube Amplifiers, McGraw-Hill, N.Y.
44. Weedy, B.M. 1980, Underground Transmission of Electrical Power, Wiley, London,
45. Weeks, W L. 1981, Transmission and Distribution of Electrical Energy, Harper and Row, N.Y.
46. Winch, R_P. Electricity and Magnetism (several printings) Composed in Copenhagen, Denmark, Prentice-Hall Physics Series
47. Zijlstra, H., 1957, Experimental Methods in Magnetism, two volumes, the North-Holland Publishing Co.. Amsterdam.

An Answer to America's Energy Deficit

Donald L. Smith

Energy Consultant

Energy, energy everywhere and not a Joule to Jounce. Conventional wisdom, when properly tuned will appreciate the nature of energy, as here presented. The basic unit of electricity (the electron) upon encountering a moving magnetic field (or wave) spins, giving off an electric impulse. When this impulse collapses, it spins back to it's natural position, giving off a magnetic impulse. Therefore, magnetic and electric are two sides of the same coin. When the magnetic side is pulsed, it yields electricity and conversely, pulsing of the electrical side yields a magnetic field. Moving one in relation to the other produces useful energy. When done consecutively, each cycle pushes (current) forward, while pulling electrons into the system... in much the same way as a water pump moves water. These electrons are obtained from Earth and air grounding.

The word "electric" comes from the Latin word electron "amber". When rubbed, amber develops an electrical charge, which can be transferred to a dissimilar substance. During the seventeenth and eighteenth centuries, a great deal of attention was centered on this attribute of amber. Amber was used to differentiate the non-metals. Carbon-related substances and other non-metals, when subjected to friction, give up negative electrical charges. On the other hand, metals when subjected to friction, simply conduct the charge. It is important to note that approximately 70% of the Earth's exposed crustal portions (surface) consist of silicone-related non-metals (electron donors) and become a direct source of electrical energy when properly agitated.

Useful electrical energy can be obtained by grounding into the Earth's non-metal crust and into it's atmosphere as a natural source of electrons. These electrons have accumulated from the solar plasma during the aging of the Earth for more than 4.5 billion years, at a rate exceeding 3.9 exajoules per year This indicates that the Earth's electrical field contains in excess of 17.6×10^{18} of cumulative exajoules of energy. One exajoule is the approximate energy equivalent of 125 million barrels of oil. The electrical energy in one display of lightning is approximately ten trillion joules. During each 24 hour period, the land portions of Earth's surface yields in excess of 200,000 emissions, which involves more than 2,000 quadrillion watts.

C.F. Gauss (1777-1855) and H.C. Oersted (1777-1851) were each separately trying to define the Earth's electrical field with all external influences removed. These external influences being solar-quiet periods and being remote from the land's surface. The air electricity background which they measured varies with latitude. Their European measurements correspond to approximately the latitude of Washington. D C. They were measuring magnetic field flux as an indicator of negative electron energy active and present. A related family of measurement are now presented. Units of measurement used to define flux fields include Gauss (one unit = 100,000 volts), Oersted (one unit = 50,000 volts), Tesla (one unit = 10,000 Gauss) and Gamma (one unit = 1/10,000 of a Gauss). Much confusion exists in electrical related publications about these units. As presented here they are correct with values taken from their original definitions.

The entire surface of the Earth has been surveyed by aerial magnetometer, in most cases using gamma or nano teslas. One gamma is the magnetic flux equivalent of 10 active volts of electricity. When the data is corrected for flight height it becomes obvious that there are numerous areas where the gamma readings exceed one trillion gammas. Lightning strikes from the ground up are in that energy range. With knowledge of these electron enriched areas, the quality of Earth grounding, becomes enhanced. The correction necessary for land surface data when acquired from aerial magnetometer maps (using Coulomb's law) requires that the remote distance be squared and then multiplied by the remote reading. As an example, if the remote reading is 1,600 gammas and the flight height being 1,000 feet. Take 1,000 x 1,000 = 1,000,000 x 1,600 gammas = 1.6 trillion gammas x 10 volts = 16 trillion volts equivalent for land surface data. Present day methodology requires mechanical energy in exchange for electrical energy. Once obtained, this energy is subject to Ohm's Law. Present Methodology obtains it's electrical energy from it's non-metal and air groundings.

This same energy can be obtained without the wasteful mechanical approach and at a much, much lower cost. Any required amount of electricity is available by resonant induction transfer from the

Earth's magnetic and electrical fields. The major difference is in the functioning of Ohm's Law in relation to resonant circuits. In the resonant induction system suggested here, system resistance (Z) becomes zero at resonance.

Therefore, Volts and Amperes are equal (V.A.R.) until work (load) is introduced.

Each cycling of this resonant induction system pulls in additional electrons from the Earth's electrical field, generating electrical energy in any required amount. In this system, a small amount of electrical energy is used to activate and pull a much larger amount of energy into the system.

This electrical advantage corresponds to the pulley and lever of the mechanical world. The electrical system presented here is extremely efficient. Using present methodology as a basis for comparison, with it's 60 cycles per second system. The resonant induction system, cycling at 60 million times per second produces one million times the energy which is produced by the present energy systems. A single small size unit of the resonant induction system has more usable electrical output than a major conventional unit. The radio frequency energy produced is easily changed to Direct Current, and then to the present 60 cycles per second system in preparation for commercial usage.

The Patent Pending on this system is #08/100,074, "Electrical Energy Generating System", dated
4 February, 1992.

Definitions: One **Joule** is one watt for one second
One **Watt** is one volt ampere
V.A.R. is Volt Amperes Reactive

Additional Reading:

Electricity and Magnetism by B.I. Bleany and B. Bleany Oxford University Press 1991 ISBN. 0-19-851172-8

Engineering Electromagnetics by W. H. Hayt. Jr. McGraw-Hill 1989 ISBN. 0-07-027406-1

Energy Methods in Electromagnetism by P. Hammond Oxford University Press 1986 ISBN. 0-19-859368-6

Energy in Electromagnetism by H. G. Booker Institution of Electrical Engineers by Peter Peregrinus Ltd. 1982 ISBN 0-900040-59-1

The American Radio Relay League Handbook for 1992 and 1993. 69th and 70th editions. Published by The American Radio Relay League. (For V.A.R. information) ISBN. 0-87259-169-7

Electron Paramagnetic Resonance, Techniques & Applications by R. S. Alger, U. S. Naval Radiological Defence Laboratory, San Francisco, California. Pub. Interscience Lib. Congress #67-20255

Geomagnetic Diagnosis of the Magnetosphere by A. Nishida, University of Tokyo 1978 Pub: Springer-Verlag ISBN. 0-387-08297-2

Energy and The Missing Resource by I. Dostrovsky. Pub: Cambridge University Press 1988 ISBN. 0-521-26592-4

High Voltage Measurement Techniques By A. Schwab, M.I.T.. 1971 I S.B.N. 0-262-19096

Environmental Magnetism by R. Thompson & F. Oldfield Pub: Allen & Unwin, London 1986 ISBN. 0-04-538003-1

Geo-electromagnetic Investigation of the Earth's Crust and Mantle. Translated from Russian, By I. I. Rokityansky. Institute of Geophysics, Kiev, U.S.S R. Pub: Springer-Verlag 1082. ISBN. 3-540-10630-8

Electron Paramagnetic Resonance of Transition Ions by A. Abragam and B. Bleaney Dover Publications, New York, N.Y. 1986

The Electromagnetic Field by A. Shadowitz, Dover Publications, New York, N.Y.

Geomagnetism, Several Volumes, Pub. Periodically by J.A. Jacobs, Institute of Earth Studies, Dyfed, U.K. Pub: Academic Press 1989-1990's.

Geomagnetism by S. Chapman and J. Bartels, 3 Volumes Oxford University Press, 1940

Physics of Geomagnetic Phenomena, Several Volumes by S. Matsushita and W. H. Campbell National Center for Atmospheric Research, Boulder, Colorado Library of Congress #67-23168 Pub: Academic Press, 1967

Physics Problems and How to Solve Them by C. E. Bennett, Professor Emeritus of Physics. University of Main (Particularly the sections on Electricity and Magnetism, and Units of Measure). Pub: Harper & Row ISBN. 0-06-460203-6

Units and Standards for Electromagnetics By P. Vigoureux, National Physical Laboratory Pub: Springer-Verlag 1071 ISBN. 0-387-91077-8

Surveyor's Guide to Electromagnetic Distance Measurement. Edited by J. J. Saastamoinen, Canada Pub: University of Toronto Press

Electromagnetic Distance Measurement by C. D. Burnside Pub: Granada, London 1971 ISBN. 0-258-96793-5

Der Magnetische Kreis "The Magnetic Circuit" By Von Heinz Rieger of Siemens AG. 1970 Berlin and Munchen, Germany I.S.B.N. 3-8009-4719 6

Electronic Modelling of Power Electronic Converters by J. A Ferreira Pub: Kluwer Academic 1989 33 AH Dordrecht, The Netherlands I.S.B.N. 0-7923-9034-2

ALTERNATING IMPACT ENERGY TRANSFER ANALOGY

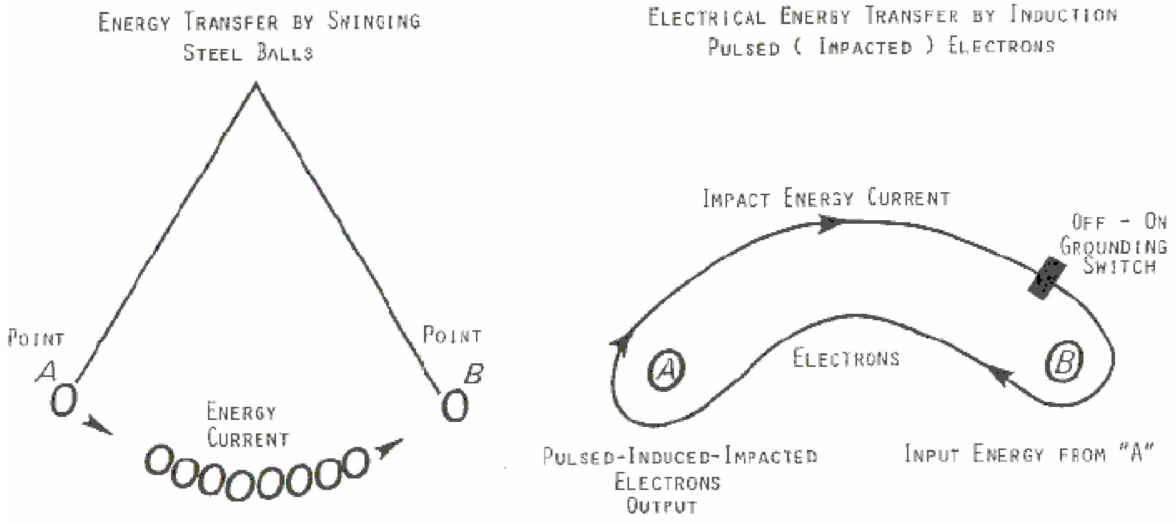
ENERGY TRANSFER BY SWINGING STEEL BALLS
POINT A
POINT B
ENERGY CURRENT
ELECTRICAL ENERGY TRANSFER BY INDUCTION PULSED (IMPACTED) ELECTRONS
IMPACT ENERGY CURRENT
OFF - ON GROUNDING SWITCH
A
B
ELECTRONS
PULSED-INDUCED-IMPACTED ELECTRONS OUTPUT
INPUT ENERGY FROM "A"

IN ALTERNATING ELECTRICAL ENERGY SYSTEMS, ONLY THE POTENTIAL-IMPACT LEVEL REPLICATES IT'S SELF AT POINT "B". ELECTRONS DO NOT TRAVEL FROM POINT "A" TO POINT "B".

E.E.S. II, BACKGROUND INFORMATION & CONCEPT

With alternating electrical current, electrons do not move from point "A" to point "B" as commonly envisioned! Electrical potential (oscillating electrons) at point "A", results in harmonic electron activity at point "B", when the grounding switch (circuit) is closed. That is to say, point "B" supplies it's own electrons and mirrors the activity of point "A". Impulsing (turbulence) by magnetic induction causes electrons to be pulled into the system, which then oscillates. When the magnetic field collapses (becomes absent) the electrical potential returns to it's natural background level.

Several major flaws are present in the conventional 60 cycles per second method of electrical power generation and it's iron core transformer system. ***This system is handcuffed by the inverse relationship of volts to amperes.*** This represents a stodgy, inflexible inheritance, courtesy of Mr. T.A. Edison and his concept of electrical power generation.

Nikola Tesla stood, almost alone, against Edison and managed to prevail with his Alternating Current system. Without the alternating current system, electronic things in the modern sense would not exist.

This report will be concerned with some of the extensions and benefits of the alternating current electrical system. ***This study will limit it's scope to air core coil transformers at radio frequency and upwards. The electrical power produced by this method is inverted to Direct Current and then to Alternating Current as required for popular usage.*** There are several important advantages of this system over conventional power generation.

Start with two coils (separate-apart), one being a reactor coil (L-1) and a second coil (L-2), being the reactant coil. Magnetic field fluxing (off-on of the electrical source) causes inductive reactance of L-1 which replicates by induction in L-2. Pulsing of the magnetic field (from L-1) in the presence of L-2, generates electrical potential. For example, should the L-1 coil have ten turns, with an imposed AC. potential of 1,200 volts. This results in each turn of L-1 acquiring 120 volts of potential. This induced magnetic field, then replicates itself in each turn of the L-2 coil. The L-2 coil may have one or many hundreds of turns. Modern encapsulation techniques makes high frequency and high energy controllable.

Let's take another important step in this air-core transformer process. For purpose of discussion, let the value of inductive reactance at 60 cycles per second, equal one. Each time the frequency is doubled, the effectiveness of induction is squared. At about 20,000 Hz, ***when radio frequency is achieved, the electrons begin spinning free, outside of the inductor and they become increasingly free of the inverse relationship of volt-amperes.*** From this point on, they replicate by the inductive process as V.A.R.. *That is to say,* ***volts and amperes are equal, until resistance (work) is introduced. Therefore, additional, not previously available electrons become incorporated for a very large net gain in potential. This gain is real !***

The quality of the grounding system determines the effectiveness of this method of producing electricity. A handy reference to locate the negative grounding areas for power generation can be found in the Aeromagnetic Map Studies of the US Geological Survey. They provide an excellent method for locating the best sites for optimum negative grounding areas.

When this method is combined with the induction coil system, already described, it provides an electrical power generating system millions of times more efficient than any known conventional method.

This new system ("E.E.S. **II**") is uncomplicated, physically small and it is inexpensive to build. The technology required for it's construction already exists. Maintenance is near zero, as there are no moving parts. Once operating, this system could last forever.

Small mobile E.E.S. **II** units are already available as replacements for the batteries used in electric automobiles. Larger E.E.S. **II** units can be provided as a replacement source of power for hotels, office buildings, subdivisions, electric trains, manufacturing, heavy equipment, ships, and generally speaking, any present day application of electrical power.

Earth Electrical System II, Modular Units

The system consists of three separate modules. Reverse engineering is used in matching the modules to the desired usage.

HIGH VOLTAGE INDUCTION TRANSFORMER MODULE:

1. Preferably an off-the-shelf-unit similar to a TV flyback and/or automobile ignition type related coil (transformer).
2. Ratio of input to output may be from less than 1:100 to greater than 1:1,000 A voltage tripler may then be used.
3. A connection allowing the high voltage output to pass onward through the induction coil L-1 and then to it's grounding.

AN AIR CORE INDUCTION COIL TRANSFORMER MODULE:

1. There are two coils: the reactor coil L-1 and the reactant coil L-2. L-1 has a high voltage radio frequency capacitor between it and it's grounding.
2. Input into the L-1 inductor is divided by the number of turns in it. The magnetic flux field provided from each turn of L-1 replicates itself as an electrical potential in each turn of L-2.
3. L-2 may have one turn or many hundreds of turns. The net gain depends upon the number of turns in L-2. Output from L-2 is in V.A.R. ***With this type of output, volts and amperes are the same until work(resistivity) is introduced.***

THE INVERTER MODULE:

1. Inverts to direct current (D C.)
2. Inverts to alternating current (A C), as desired.
3. Provides customized output of electrical power ready for designated usage

Earth Electrical System II, Domestic Use Range Module
Up to Two Million Volt-Amperes-Reactive Output *
Plan "A", With Variable Controls
Not to Scale

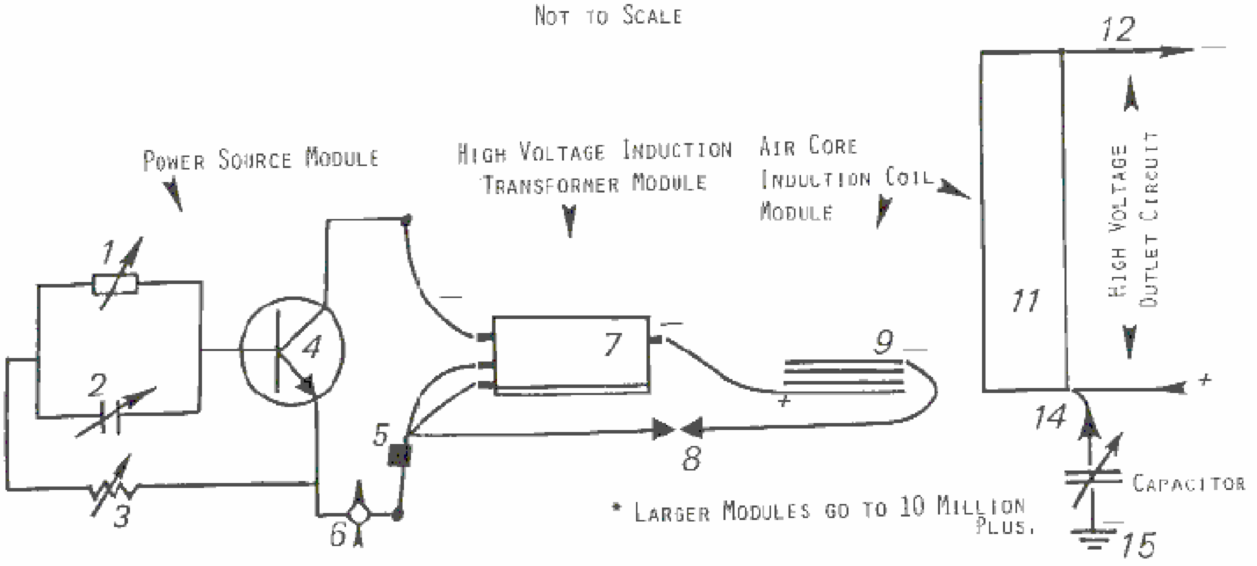
12
Power Source Module
High Voltage Induction Transformer Module
Air Core Induction Coil Module
High Voltage Outlet Circuit
1
11
2
4
7
9
+
5
8
14
3
6
Capacitor
15

* Larger Modules go to 10 Million Plus.

Parts: 1. Coil, Variable, 2. Capacitor, Variable, 3. Resistor, Variable, 4. Transistor, R.F., 5. Battery, Rechargeable, 6. Off-On Switch, Variable, 7. High Voltage Transformer, 8. Feed Back with Spark Gap, 9. Reactor, Induction Coil, 10. Feed Back with Spark Gap, 11, Reactant Coil, 12. Output for # 11, 13. Input for eleven, 14. Grounding for eleven.

Earth Electrical System II, Domestic Use Range Module
Up to Two Million Volt-Amperes-Reactive Output *
Plan "B", Electrical Automobile Energy Source

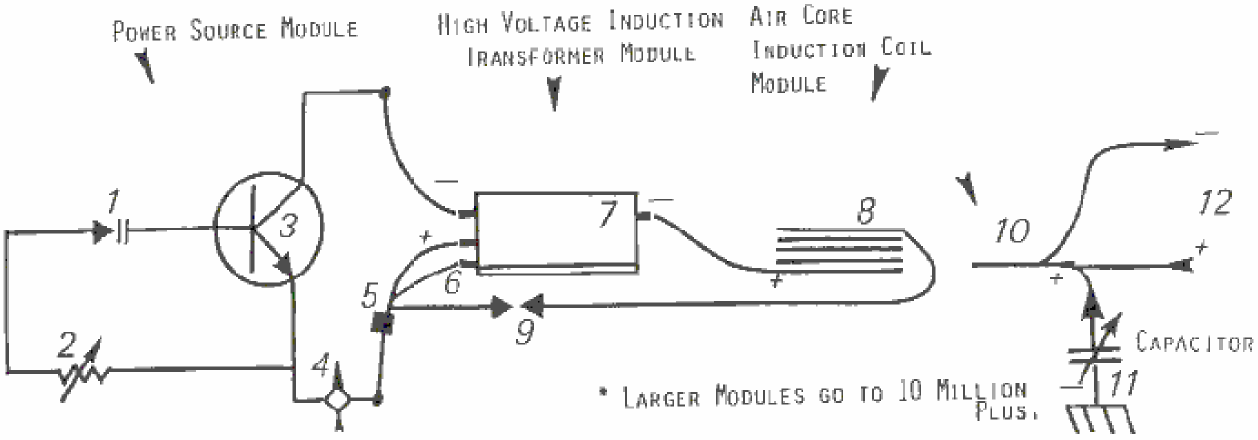
Power Source Module
High Voltage Induction Transformer Module
Air Core Induction Coil Module
1
3
7
8
10
12
+
5
6
9
2
4
Capacitor
11

* Larger Modules go to 10 Million Plus.

Parts: 1. Varactor, Radio Frequency, 2. Resistor, 3. Transistor, Radio Frequency, 4. Off-On Switch, Multi-Position, 5. Battery, Rechargeable, 6. Transformer Grounding, 7. High Voltage Induction Transformer, 8. Reactor, Induction Coil, 9. Feed Back with Spark Gap, 10. Reactant, Induction Coil, 11. Grounding for # 10, Out-put Circuit, In Volt-Amperes-Reactive.

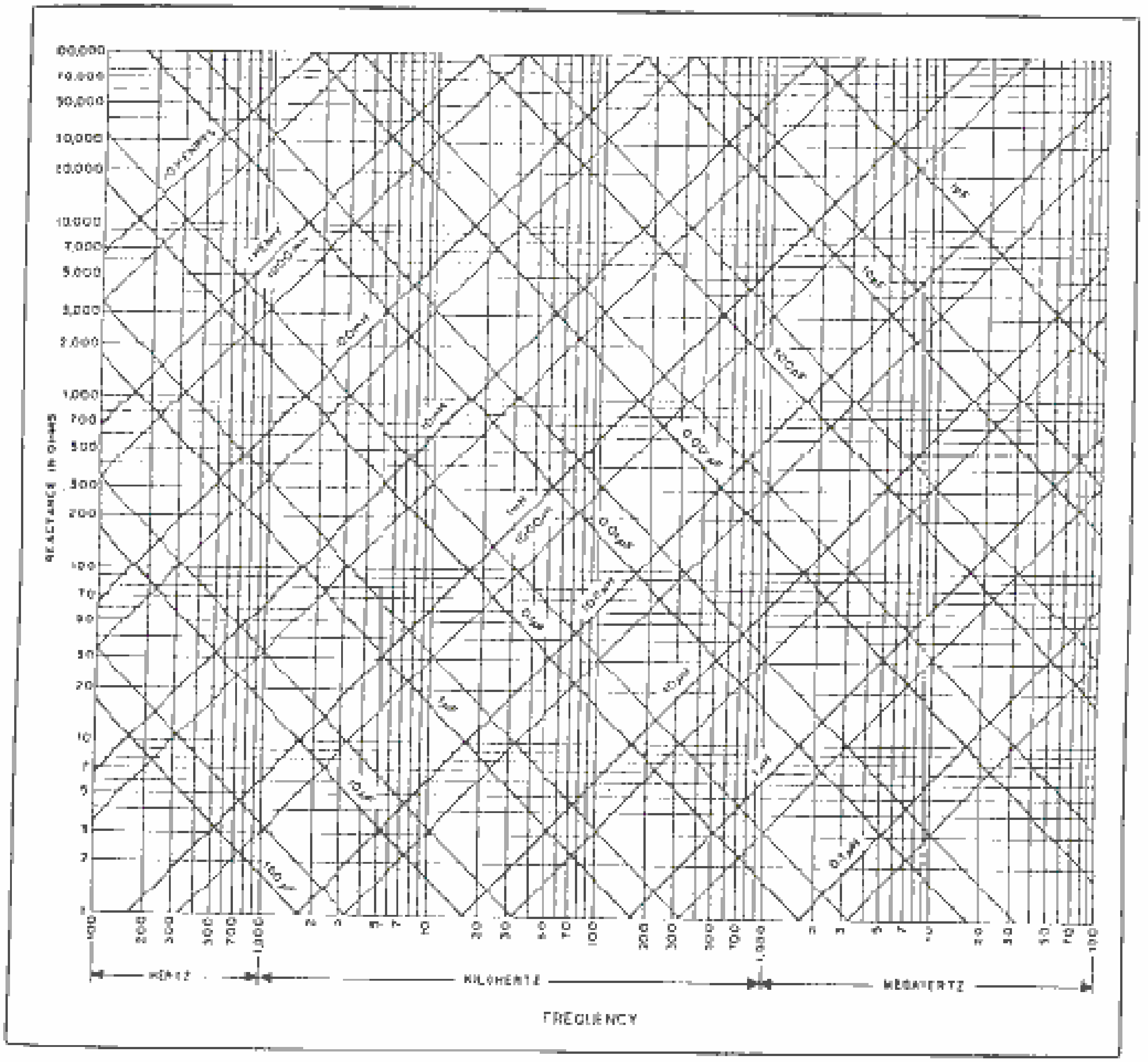

Fig. 44 — Inductive and capacitive reactance vs. frequency. Heavy lines represent multiples of 10, intermediate light lines multiples of five. For example, the light line between 10 µH and 100 µH represents 50 µH; the light line between 0.1 µF and 1 µF represents 0.5 µF, and so on. Intermediate values can be [illegible] within the chart range. For example, the reactance of 10 henrys at 60 Hz can be found by taking the reactance of 10 henrys at 600 Hz and dividing by 10 for the 10 × times decrease in frequency

CHART, COURTESY OF A.R.R.L., 1992 ED.

Speech presented the evening of 23 July. 1994 at the International Tesla Society Convention at Colorado Springs. Colorado.

DONALD L. SMITH
ENERGY CONSULTANT
8110 BENT OAK LANE
SPRING. TEXAS 77379

ELECTRICAL ENERGY REFERENCE POINTS

Electrical Energy Generating System
Patent Pending # 08/100.074. 2/4/92

The word "electric" comes from the Latin word electron "amber". When rubbed, amber develops an electrical charge, which can be transferred to a dissimilar substance. During the seventeenth and eighteenth centuries, a great deal of attention was centered on this attribute of amber. Amber was used to differentiate the non-metals. Carbon-related substances and other non-metals, when subjected to friction, give up negative electrical charges. On the other hand, metals when subjected to friction, simply conduct the charge. It is important to note that approximately 70% of the Earth's exposed crustal portions (surface) consist of silicone related non-metals (electron donors) and therefore becomes a direct source of electrical energy when properly agitated.

Useful electrical energy is obtained by grounding into the Earth's non-metal crust and into it's atmosphere as a natural source of electrons. These electrons have accumulated from the solar plasma during the aging of the Earth for more than 4.5 billion years, at a rate exceeding 3.9 exajoules per year. This indicates that the Earth's electrical field contains in excess of 17.6 x 10^{18} power of cumulative exajoules of energy. One exajoule is the approximate energy equivalent of 125 million barrels of oil. The electrical energy in one display of lightning is approximately ten trillion joules. During each 24 hour period, the land portions of the Earth's surface yields in excess of 200,000 emissions, which involves more than 2,000 quadrillion watt-seconds of active energy on display.

This physical phenomenon indicates that the Earth's crust is an unending source of electrical energy. The surface area involved is a very small portion of the Earth's crust.

J.C. Maxwell (1891) suggested that an active electron field gives rise to an associated magnetic field. Therefore, both are present with pulsating current. Early studies, involving observation of compass needles by microscopy, revealed that the needle vibrates as with alternating current. More recent studies by A. Nishida and others, confirm that alternating current is common in the Earth's crust.

C.F. Gauss (1777-1855) and H.C. Oersted (1777-1851), both were separately trying to define the Earth's electrical field with all external influences removed. These external influences being solar-quiet periods and being remote from the land's surface. The air electricity background which they measured varies with latitude. Their European measurements correspond to approximately the latitude of Washington, D.C. They were measuring magnetic field flux as an indicator of negative electron energy active and present.

A related family of measurement is now presented. Units of measurement used to define flux fields include Gauss (one unit = 100,000 volts), Oersted (one unit - 50,000 volts), Tesla (one unit = 10,000 Gauss) and Gamma (one unit = 1/10,000 th of a Gauss). Much confusion exists in electrical related publications about these units. As presented here, they are correct with values taken from their original definitions.

The entire surface of the Earth has been surveyed by aerial magnetometer, in most cases using gamma or nano teslas. One gamma is the magnetic flux equivalent of 10 active volts of electricity. When this data is corrected for flight height, it becomes obvious that there are numerous areas where the gamma readings exceed one trillion gammas. Lightning strikes from the ground up are in that energy range. With knowledge of these electron enriched areas, the quality of Earth grounding, becomes enhanced.

The correction necessary for land surface data when acquired from aerial magnetometer maps (using the inverse square law) requires that the remote distance be squared and then multiplied by the remote reading. For example, if the reading is 1,600 gammas and the flight height is 1,000 feet. Take 1,000 x 1,000 = 1,000,000 x 1,600 gammas = 1.6 trillion gammas x 10 volts = 16 trillion volts equivalent for land surface data.

Present day methodology requires mechanical energy to be expended in exchange for electrical energy. Any required amount of electricity is available by resonant induction transfer from the Earth's magnetic and electrical fields. Each cycling of this resonant induction system pulls in additional electrons, generating energy in any required amount. A small amount of electrical energy is used to activate and pull into the system a much larger amount of energy.

ENERGY VERSUS MASS

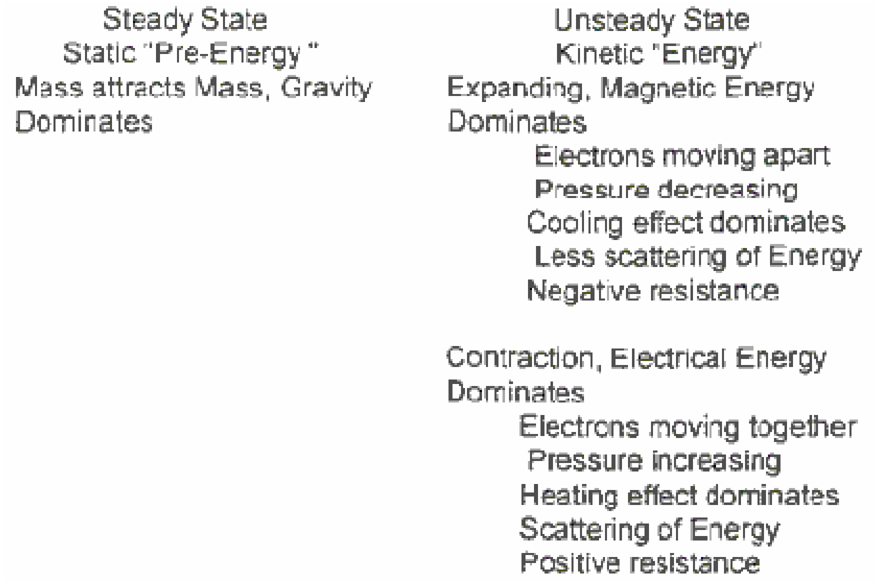

Steady State Static "Pre-Energy " Mass attracts Mass, Gravity Dominates	Unsteady State Kinetic "Energy"
	Expanding, Magnetic Energy Dominates Electrons moving apart Pressure decreasing Cooling effect dominates Less scattering of Energy Negative resistance
	Contraction, Electrical Energy Dominates Electrons moving together Pressure increasing Heating effect dominates Scattering of Energy Positive resistance

Functions of active Electrons

Electrons become active when placed inside the critical distance allowed by their negativity.

Active Electrons provide:

1. Electricity
2. Magnetics
3. Gravitational thrust as in Electric Motors
4. The source of Visible Light
5. It's charge is Negative

They move in a closed loop as seen in the Icon for infinity, not in a circle as shown in many books. One half of the loop consist of a magnetic impulse and the return half consist of the electrical impulse. This is seen as the classic sine wave of alternating electrical energy.

A flash of light occurs when two electrons suddenly find they are too close together. Daylight results from the impingement of Electrons in the Earth's atmosphere with the Electrons of the Solar Plasma.

My Concept of the Forces of Nature differs from the conventional. It consist of a weak and a strong force, each being additionally composed of electrical, magnetic and gravitational (fields and waves). Any two of the three constitute the third member; Gravity "B" of the weak force competes with humans on a daily basis. Gravity "A" of the strong force is the force that holds the Solar System and the Universe in place. Energy from the Electrons represent the weak force. Energy inside the Atom represents the strong force "A". Controlled resonant induction of any two of the three, changes into the third and is the motor that runs the Universe. We see this in the electrically-induced magnetic thrust against gravity in electric motors.

Weak force is required to dislodge electrons and strong force (atomic) to dislodge protons. Unless dislodged, these particles are of little value in producing Conventional Electrical Energy.

Therefore, in conventional electrical energy production, the particle of importance is the negative electron. Electrons have a "grudging" relationship with other electrons. They like each other, especially at arms length. Like potentials repel each other, and unlike potentials attract. To demonstrate this, take two batteries of the same type, but of a different charge level (unequal potentials). Put the plus and minus ends facing the same direction. Then with a volt meter, measure the electrical potential between the two negative ends and then the two positive ends. It is obvious that the "more negative" moves to the "less negative" is the correct concept for electrical energy generation. Electrical Energy flow consist of a higher concentration of electrons moving to an area of lesser concentration.

OHM'S LAW WITH CORRECTIONS:

A major obstruction in reference to the correct function of electrical energy is the establishment's incorrect interpretation of Ohm's Law. The corrected version is:

Volts = Energy Available (Potential)

Ohm = Scattering, dissipation of Energy (Load)

Ampere = the rate of, dissipation / scattering of energy

It is important to note that Ohm and Ampere are after the fact, and are not decisive except for the dissipation factor. High Voltage at low amperage simply means that the High Voltage is still intact for future usage. In no way is the potential diminished by low amperage.

EXAMPLES OF OVERUNITY

Dominos did not exist in England when the Laws of Conservation were originally put in place. Otherwise they might have been very different. For example, let us take a long row of upright dominos, (many thousands) and flip number one. The Energy required to flip the first domino must now be added with that of thousands more in order to have a correct assessment.

The Electron itself is an excellent example of over-unity. The electron provides various forms of energy continuously throughout eternity and is in no way diminished. It simply cycles through the system and is available thereafter.

In Electrical Systems, Electrons active at point "A" are not the same Electrons active at point "B". That is to say, the Electrons activated at the Central Electrical Energy Station are not the ones used at your house. When you ground your system by flipping the wall switch, you use your own electrons. In closed energy systems, electrons communicate with and replicate the activity of the overbalanced potential, when provided with Earth and or Air Groundings.

The number of Radio sets and Television sets running at any one time do not diminish, in any way the electrical output of the source station.

For example, let now use an Air Coil Resonant Induction System for the purpose of flipping some electrons. The flipping device (reactor coil L-1) is pulsed, which then provides a resonant induction pulse. In turn, this flips the electrons present at the (reactant L-2) Coil. The energy input in L-1 is divided by the number of turns present. The induced magnetic pulsing in turn flips the electrons in each turn of L-2. If more turns are present in L-2 than L-1, there is a net gain in the Energy present, as demonstrated by the dominos above. The farads and henrys of the resonant system provide the resonant frequency when pulsed by an external energy system. A system shunt in the resonant circuit sets the containment level for energy potential.

The Induction Process itself provides an excellent example of over-unity. When comparing rate of induction, the cycles per second must be squared and then compared to the square of the second System. Let us then compare the 60 c.p.s. System with my 220 MHz Device. Energy produced at radio frequency has several major advantages over the conventional system. Ohm's Law does not apply to a resonant air-core radio frequency system.

For example: When the system is resonant, the following is true:

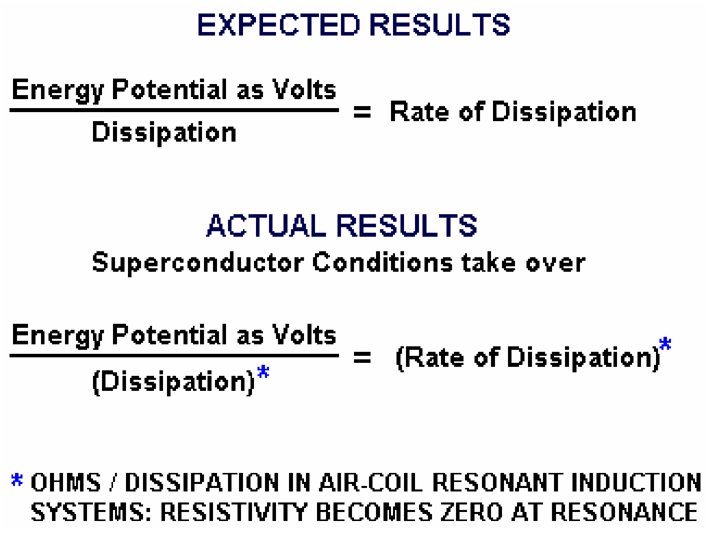

EXPECTED RESULTS

$$\frac{\text{Energy Potential as Volts}}{\text{Dissipation}} = \text{Rate of Dissipation}$$

ACTUAL RESULTS

Superconductor Conditions take over

$$\frac{\text{Energy Potential as Volts}}{(\text{Dissipation})^{*}} = (\text{Rate of Dissipation})^{*}$$

* OHMS / DISSIPATION IN AIR-COIL RESONANT INDUCTION SYSTEMS: RESISTIVITY BECOMES ZERO AT RESONANCE

This is named the V.A.R. (Volt Amperes Reactive) System.

When compared to the Conventional Under-Unity iron-core transformer system, the results are over-unity. It is strange that mechanical advantage as in pulleys, gears, levers and others which correspond to the electrical advantage above mentioned, are not considered over-unity devices.

Let us take a closer look at resonant induction. As an example, let a room full of ping pong balls randomly bouncing at a high speed represent the Conventional method of under-unity energy generation. Suppose that by resonant induction the balls all move in the same direction at the same time. When this occurs a huge amount of energy not previously available is present. The resonant air-core coil system lines up the electrons in such a manner that the energy factor is nearly 100 % , and not the 2% or 3% of Conventional under-unity devices sanction by the establishment.

Some other devices where overunity is common would be resonant induction circuits present in conventional radio tubes (high plate voltage), negative-feedback systems found in Op-Amps and possibly others.

SUMMARY

Useful electrical energy is achieved when the electron density at point "A" becomes greater than at point "B", (being the more-negative moving to the less-negative concept). Coils moving through a magnetic field or vice versa causes this imbalance.

The mindset of the professional Electrical Engineer is restricted to non-resonant and iron-core coil resonant systems. Ohm's Law, when applied to resonant air-core induction systems, becomes, system resistivity (impedance, Z). "Z" becomes zero at resonance. Therefore, in this system, volts and amperes are equal until load (resistivity) is introduced. This is called the Volt Ampere Reactive (V.A.R.) System. With impedance being zero, the System grounding is coupled directly into the Earth's immense electrical potential. Efficiency of induction relates to the square of the cycles per second. Compare the ratio of the conventional 60 c.p.s. System and the 220 million plus cycles of my Earth Electrical System II.

Electrons which cycle through this system, after being used, are returned intact to their former state for future usage.

Electron spin causes electrical current and magnetic lines of force

The effect of current, results from the unequal distribution of negativity (electrons).

Magnetic imbalance causes the gravitational effect. This is evidenced in electric motors by magnet-gravitational displacement of mass which causes the motor to rotate.

The System is an extension of present technology.

The System and it's source utilizes magnetometer studies.

This System (Earth Electrical System II. "EES.II") utilizes a fully renewable energy source.

This System utilizes a non-polluting energy source.

This System utilizes an universally available energy source.

Endorsement and Certification of The System can be anticipated by States with pollution problems.

AIR CORE INDUCTION COIL BUILDERS GUIDE

DONALD L. SMITH

Energy Consultant

1. Decide frequency. Considerations are: (economy of size)

- **a.** Use radio frequency upward (above 20,000 Hz).
- **b.** Use natural frequency (coils have both capacitance and inductance), that is match the wire length of the wire in the coil to the desired frequency.
- **c.** Wire length is either one quarter, one half or full wave length.
- **d.** To obtain the wire length (in feet) use the following: If using one quarter wave length divide 247 by the desired frequency (megahertz range is desirable). If using one half wave length divide 494 by the desired frequency. If using full wave length divide 998 by the desired frequency.

2. Decide number of turns, ratio of increase in number of turns sets the function. In the case of the L-1 coil, each turn divides the input voltage by the number of turns. In the case of L-2 coil, the resulting voltage in each turn of L-1 is induced into each turn of L-2, adding up with each turn. For example if the input into L-1 from a high voltage, low amperage module is 2,400 volts, and L-1, for example, has 10 turns, then each turn of L-1 will have 240 volts of magnetic induction which transfers 240 volts of electricity to each turn of L-2. L-2 may be one turn or many turns, such as 100 to 500 or more turns. At 100 turns, 24,000 volts would be produced. At 500 turns, 120,000 volts would be produced.

3. Decide the height and diameter of the coil system. The larger the diameter of the coil, the fewer turns are required, and the coil has a lesser height. In the case of L-2 this results in lowering the amplification of the induced voltage from L-1.

4. For example, if 24.7 MHz is the desired frequency output from L-2. One quarter wave length would be 247 divided by 24.7 which equals 10 feet of wire. The number of turns will be the amplification factor. The coil may be wound on standard size P.V.C. or purchased from a supplier. The supplier is normally a ham radio supply source. Once the length is determined and the number of turns decided, move to the next step. For example, let each turn of L-1 have 24 volts and desired output of L-2 be 640 volts. Therefore L-2 needs 26.67 turns. It has been determined that the wire length for one quarter wave length is 10 feet. The number of inches in 10 feet is 120. Using Chart "A" supplied look for next higher number of turns showing (being between 20 and 30 turns with a 2" diameter coil). This tells us to use a 2" coil. If ready-made as in the case of Barker and Williamson, 10 Canal Street, Bristol, Penna., 215-788-5581, the coils come in standard sizes of 4, 6 and 10 turns per inch. For higher "Q" use wider spacing of the turns. These coils come in a ready-made length of 10 inches. Select from the coil 30 turns and put input clamps on the base of the coil and at 30 turns. For exact determination of the correct position of the output clamp, use an externally grounded voltage probe. The node of maximum intensity, being the natural resonant point. Off the shelf multimeters are not radio frequency responsive. The easiest way to accomplish the above is to get from the hardware store or Radio Shack a voltage detector having a neon bulb system (Radio Shack Cat. No. 272-1100b, NE2-Neon Lamps) will work. With your hand as a ground, move the wire extension of the neon lamp along the coil surface until the neon is brightest. This is the desired point of resonance and it is the optimum connection point.

5. The input power now needs consideration. A 2,400 High Voltage module has been previously selected. This module can be made from a diode bridge or any combination of voltage amplifiers. The one used here is an off-the-shelf type, similar to those used for laser technology.

6. Construction of the input L-1 coil. It has already been decided that there will be 10 turns. The length of the wire here is not critical. Since the L-2 coil is 2-inches in diameter, the next off-the-shelf larger size may be used for L-1. Use a 3" diameter off-the-shelf coil which has 10 turns to the inch. Remove (cut) a 10 turn portion from the larger coil. Use an L.C.R. meter and measure the natural farads (capacitance) and henrys (inductance) values of the L-2 coil. Now do the same for the L-1 coil. It will be necessary to put a capacitor across the voltage input of L-1 in order to match the L-1 coil to the L-2 coil. A spark gap across L1 is also required to deal with the return voltage from L-1. A tuneable capacitor of the pad ("trimmer") type for L-1 is desirable.

7. The performance of the L-2 coil can be further enhanced by having an Earth grounding from the base of the coil. The maximum voltage output will be between the base and the top of the L-2 coil. Lesser voltages can be obtained at intermediate points along the length of the L-2 coil.

SUPPLY SOURCES

1. HAM RADIO SUPPLY STORES

2. COILS, AIR INDUCTOR IN HOUSTON

3. BAKER AND WILLIAMSON (READY MADE), BRISTOL. PENNA.
ALSO R.F. DUMMY LOADS AND WATTMETERS.

NOTES

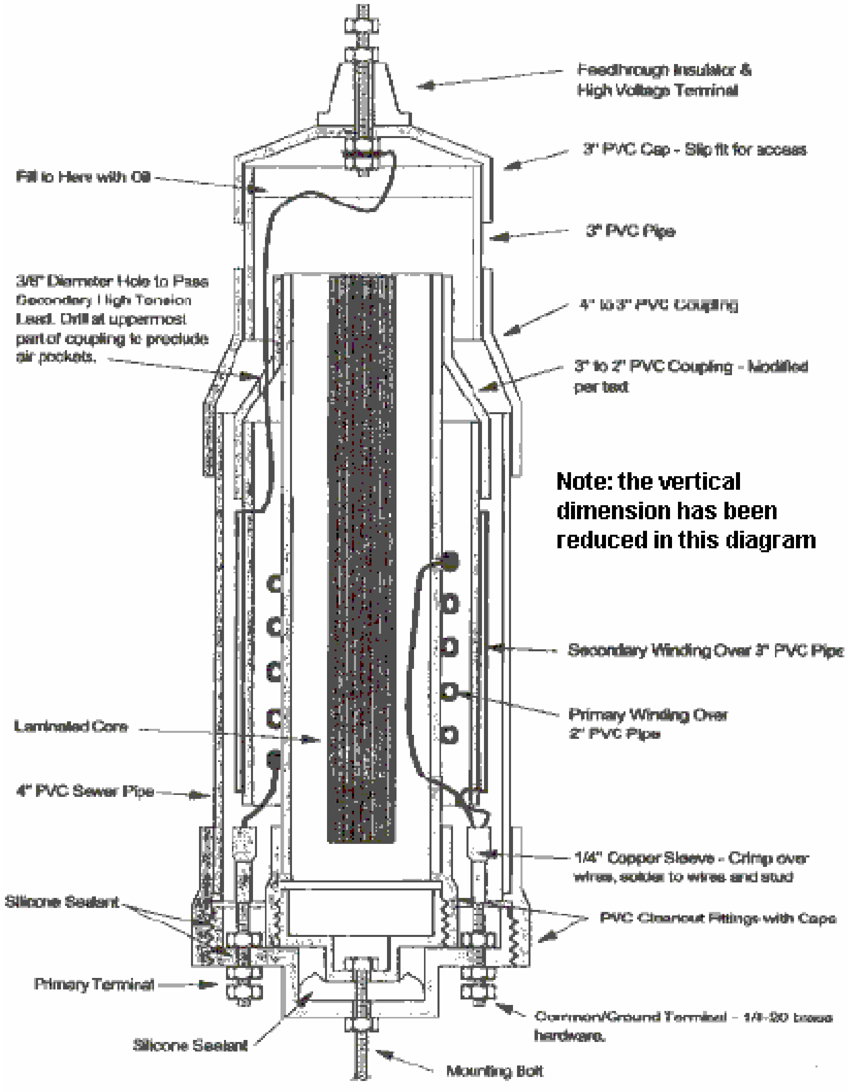
Feedthrough Insulator &
High Voltage Terminal
3" PVC Cap - Slip fit for access
Fill to Here with Oil
3" PVC Pipe
3/8" Diameter Hole to Pass Secondary High Tension Lead. Drill at uppermost part of coupling to preclude air pockets.
4" to 3" PVC Coupling
3" to 2" PVC Coupling - Modified per text
Secondary Winding Over 3" PVC Pipe
Primary Winding Over 2" PVC Pipe
Laminated Core
4" PVC Sewer Pipe
1/4" Copper Sleeve - Crimp over wires, solder to wires and stud
Silicone Sealant
PVC Cleanout Fittings with Caps
Primary Terminal
Common/Ground Terminal - 1/4-20 brass hardware.
Silicone Sealant
Mounting Bolt

Note: the vertical dimension has been reduced in this diagram

Impulse Transformer Assembly

ELECTRICAL PRINCIPLES: TERMINOLOGY & SAFETY

The use of electricity is so commonplace that most people assume that it will always be available on demand. To fully realize our dependence upon electricity, consider the ways in which electricity is being used each day in the home, on the farm and the ranch. Electricity is doing more to increase work efficiency and promote enjoyable living than any other single factor. The use of electricity has grown to the extent that an increasing portion of the home or business budget, is used in paying for this source of energy.

1. Definition of Electricity

Electricity can be defined in several ways. The layman defines electricity as a source of energy that can be converted to light, heat, or power. Electrical Engineers define electricity as a movement of electrons caused by electrical pressure or voltage. The amount of energy produced depends on the number of electrons in motion.

2. The Manufacture and Distribution of Electricity

Electricity is produced from generators that are run by water, steam, or internal combustion engines. If water is used as a source of power to turn generators, it is referred to as hydroelectric generation. There are a number of this type located in areas where huge dams have been built across large streams.

Steam is used as a source of power for generating much of today's electricity. Water is heated to a high temperature, and the steam pressure is used to turn turbines which generate electricity. These are referred to as thermal-powered generators. Fuels used to heat the water are coal, natural gas, and/or fuel oil.

Generators at the power plant generate from 13,800 to 22,000 volts of electricity. From the power plant, electricity is carried to a step-up sub-station which, through the use of transformers, increases the voltage from 69,000 to 750,000 volts. This increase in voltage is necessary for the efficient transmission of electricity over long distances. From the step-up sub-station, the electricity is carried on transmission lines to a step-down sub-station which reduces the voltage to 7,200 to 14,000 volts for distribution to rural and city areas.

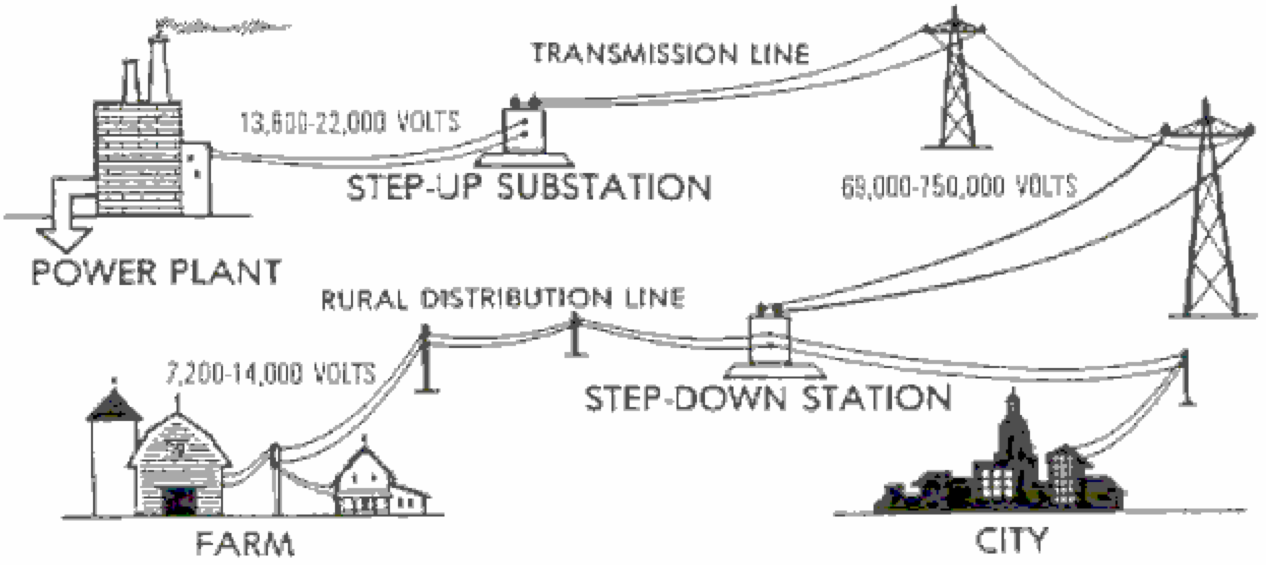
TRANSMISSION LINE
13,800-22,000 VOLTS
STEP-UP SUBSTATION
69,000-750,000 VOLTS
POWER PLANT
RURAL DISTRIBUTION LINE
7,200-14,000 VOLTS
STEP-DOWN STATION
FARM
CITY

Transformers at the business or residence reduce the voltage to 120 or 240 volts to supply the meter of the customer:

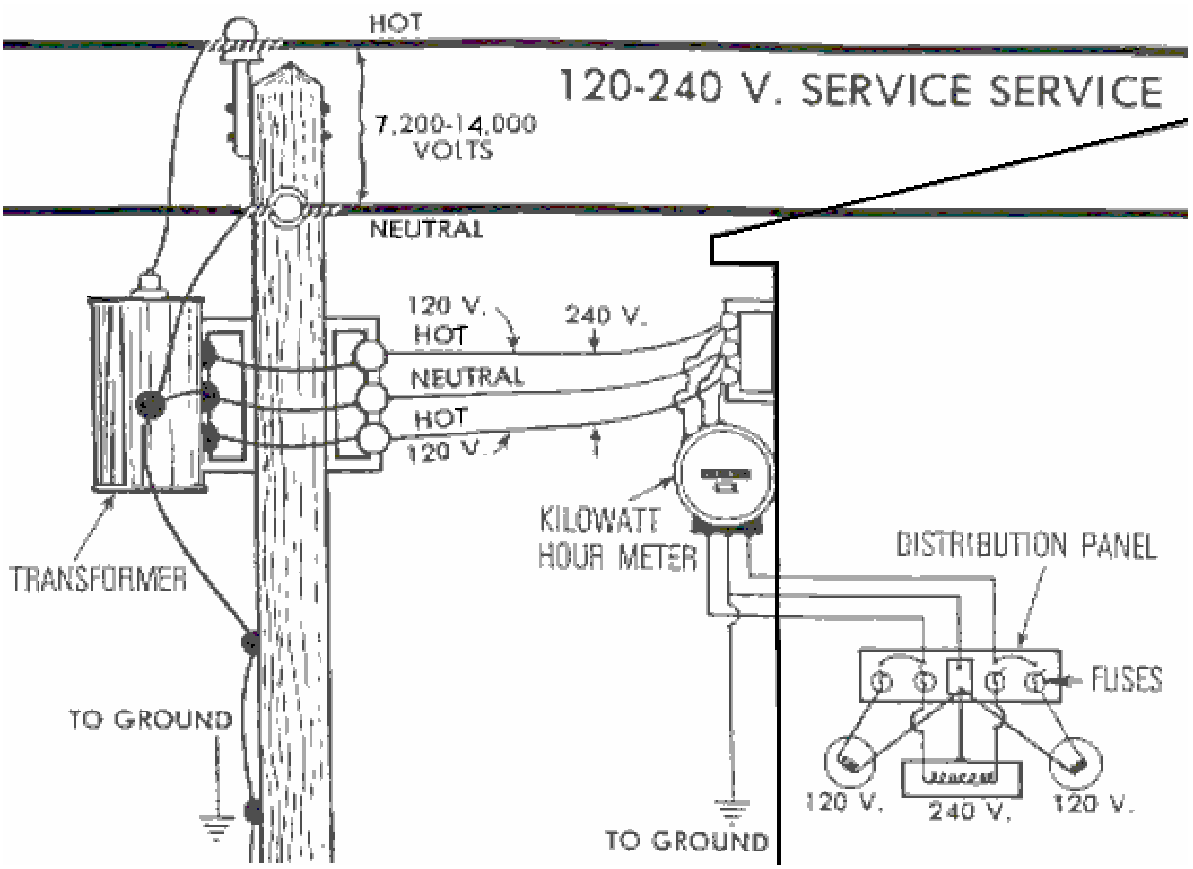
HOT
120-240 V. SERVICE SERVICE
7,200-14,000 VOLTS
NEUTRAL
120 V. HOT
240 V.
NEUTRAL
HOT
120 V.
KILOWATT HOUR METER
DISTRIBUTION PANEL
TRANSFORMER
FUSES
TO GROUND
120 V.
240 V.
120 V.
TO GROUND

3. Common Electrical Terms

In order to work safely and efficiently with electricity and have the ability to converse on the subject, the following terms should be understood:

Ampere (Amp) - A measurement in units of the rate of flow of electrical current. This may be compared with the rate of flow of water in gallons per minute.

Example: A 60-watt incandescent lamp on a 120V circuit would pull 1/2 ampere of electricity (60 divided by 120 = 0.5 or 1/2, Formula: Amperes = Watts / Volts

Volt (V) - A unit of measure of electrical pressure. A given electrical pressure (V) causes a given amount of electrical current (Amps) to flow through a load of given resistance. Voltage may be compared to water pressure in pounds per square inch in a water system. Common service voltages are 120 volts for lighting and small appliance circuits and 240 volts for heating, air conditioning, and large equipment circuits.

Watt (W) - A unit of measure of electrical power. When applied to electrical equipment, it is the rate that electrical energy is transformed into some other form of energy such as light. Watts may be compared to the work done by water in washing a car. (Formula: Volts x Amps = Watts)

Kilowatt (KW) - A unit of measurement used in computing the amount of electrical energy used. Kilowatts are determined by dividing the number of watts by 1000 as 1 kilowatt = 1,000 watts.

Kilowatt-Hour (KWH) - A measure of electricity in terms of power in kilowatts and time in hours. One KWH is 1000 watts used for one hour.

Alternating Current (A.C.) - Electrical current that alternates or changes direction several times per second. The direction current moves depends on the direction in which the voltage forces it.

Cycle - The flow of electricity in one direction, the reverse flow of electricity in the other direction, and the start of the flow back in the other direction. The cycles per second are regulated by the power supplier and are usually 60 in America. Most electric clocks are built to operate on the mains frequency. More or fewer cycles per second would cause mains-operated clocks to gain or lose time. The present practice is to use the term Hertz (Hz) rather than "cycles per second".

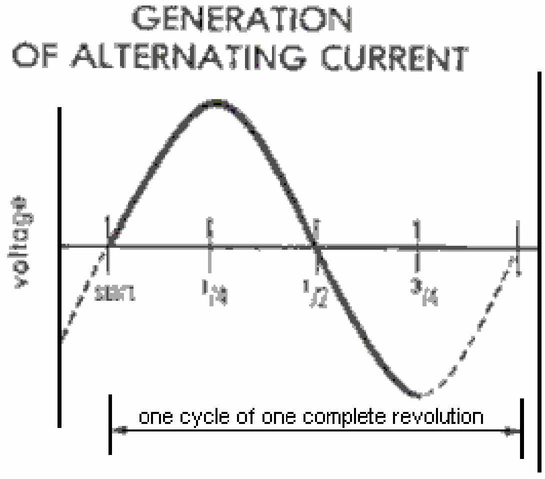
GENERATION OF ALTERNATING CURRENT
voltage
start
1/4
1/2
3/4
one cycle of one complete revolution

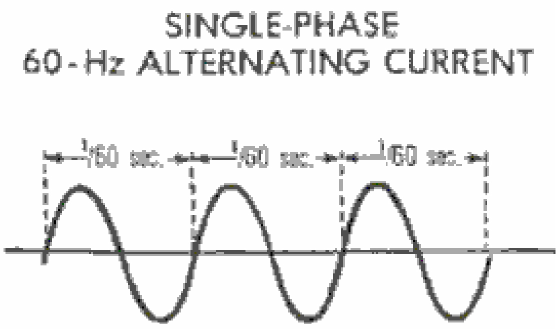
SINGLE-PHASE 60-Hz ALTERNATING CURRENT
1/60 sec.
1/60 sec.
1/60 sec.

Direct Current (D.C.) - Electrical current flowing in one direction. Example: electrical circuits in automobiles and tractors.

Transformer - A device used to increase or decrease voltage.

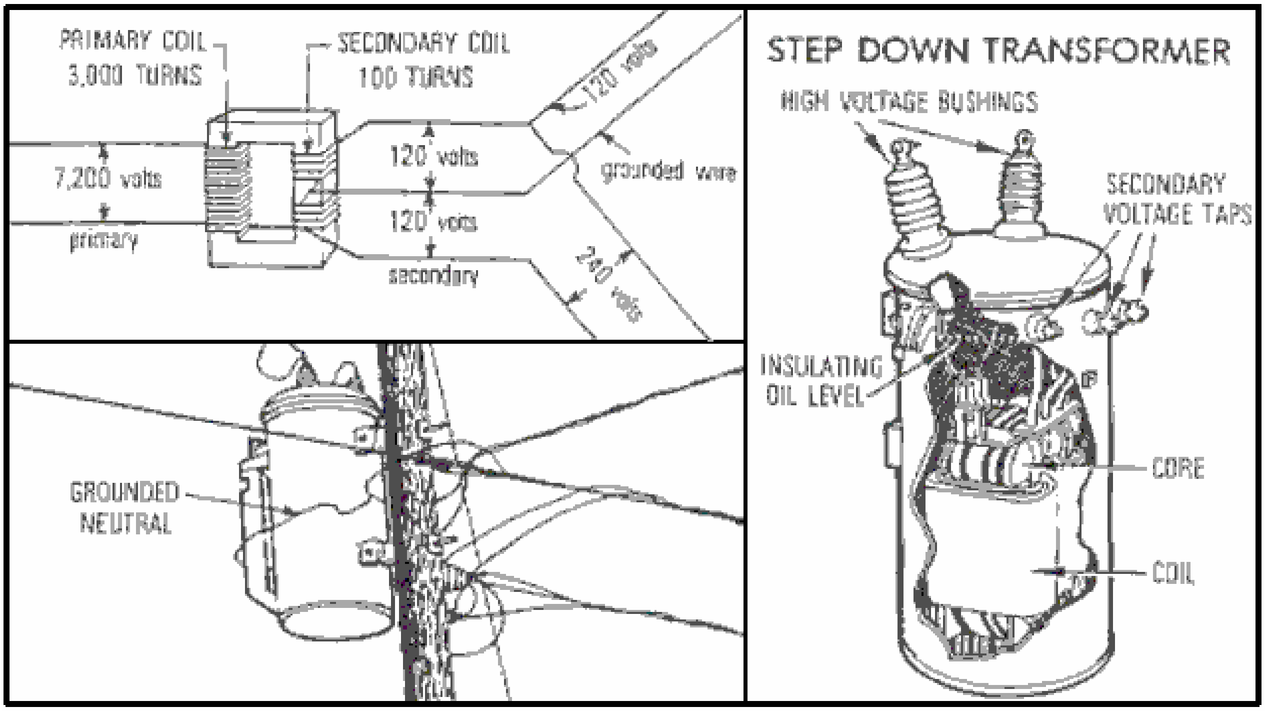
STEP DOWN TRANSFORMER
PRIMARY COIL
3,000 TURNS
SECONDARY COIL
100 TURNS
7,200 volts
primary
120 volts
120 volts
secondary
120 volts
grounded wire
240 volts
GROUNDED NEUTRAL
HIGH VOLTAGE BUSHINGS
SECONDARY VOLTAGE TAPS
INSULATING OIL LEVEL
CORE
COIL

Single Phase - The most common type of electrical service or power available to consumers. One transformer is used between the distribution line and the meter. Usually three wires, two "hot" and one neutral, are installed to provide 120V and 240V single-phase service. Single-phase service may also be supplied with a three-phase service.

Three-Phase - This type of service is designed especially for large electrical loads. It is a more expensive installation due to three wires and three transformers being required. The important advantage of three-phase power is that the total electrical load is divided among the three phases, consequently, the wire and transformers can be smaller. Other advantages exist in the design of three-phase motors.

Short Circuit - A direct connection (before current flows through an appliance) between two "hot" wires, between a "hot" and neutral wire, or between a "hot" wire and ground.

Voltage Drop - A reduction of current between the power supply and the load. Due to resistance, there will be a loss of voltage any time electricity flows through a conductor (wire). Factors that influence voltage drop are size of wire, length of wire, and the number of amps flowing. A drop in voltage may cause a loss of heat, light, or the full power output of a motor. It could cause motor burn-out unless the motor is properly protected (time-delay fuse).

Fuse - A device used to protect circuits from an overload of current.

Circuit Breaker - A device used to protect circuits from an overload of current. May be manually reset.

Time-Delay Fuse - A fuse with the ability to carry an overload of current for a short duration without disengaging the contacts or melting the fuse link.

Horsepower (hp) - A unit of mechanical power equal to 746 watts of electrical power (assuming 74.6% electric motor efficiency). Motors of one horsepower and above are rated at 1000 watts per hp while motors below one horsepower are rated at 1,200 watts per hp.

Conductor - The wire used to carry electricity (typically, copper or aluminum). Copper and aluminum should not be spliced together due to their incompatibility resulting in deterioration and oxidation.

Insulator - A material which will not conduct electricity and is usually made of glass, Bakelite, porcelain, rubber, or thermo-plastic.

"Hot" Wire - A current-carrying conductor under electrical pressure and connected to a fuse or circuit breaker at the distribution panel. (Color Code: usually black or red)

Neutral Wire - A current-carrying conductor not under electrical pressure and connected to the neutral bar at the distribution panel. (Color Code: usually white)

Grounding - The connection of the neutral part of the electrical system to the earth to reduce the possibility of damage from lightning and the connection of electrical equipment housings to the earth to minimize the danger from electrical shock. (Color Code: Can be green or bare wire).

Underwriters' Laboratory (U.L.) - An American national organization which tests all types of wiring materials and electrical devices to insure that they meet minimum standards for safety and quality.

National Electric Code (N.E.C.) - Regulations approved by the National Board of Fire Underwriters primarily for safety in electrical wiring installations. All wiring should meet the requirements of the national as well as the local code.

4. Computing Electrical Energy Use and Cost

If an estimate of cost for electricity used is desired, the name plate data on appliances and equipment and an estimate of operating time may be used. The following formulas should be used for determining watts, amps, volts, watt-hours, kilowatt-hours, and cost.

Watts = Volts x Amperes

Amperes = Watts / Volts

Volts = Watts / Amperes

Watt-Hours = Watts x Hours of operation

Kilowatt-Hours = Watt-Hours / 1000

Cost = Kilowatt-Hours x Local Rate per Kilowatt-Hour (or per "Unit")

Example:

Local electricity rate per Kilowatt-Hour: 8 cents
Equipment plate data: 120 Volts 5 Amps
Monthly hours of operation: 10

1. Watts = Volts x Amperes, so Watts = 120 x 5 = 600 watts
2. Watt-Hours = 600 x 10 = 6,000 watt-hours
3. Kilowatt-Hours = 5,000 / 1,000 = 6 kilowatt-hours (or 6 Units)
4. Cost = 6 x 8 = 48 cents

5. Electrical Circuits

An Electrical Circuit is a completed path through which electricity flows. Insulated conductors (wires) provide the path for the flow of electricity. A water system and an electrical circuit are similar in many respects. Water flows through pipes and is measured in gallons per minute, and electricity flows through conductors and is measured in amperes. A simple circuit is shown here:

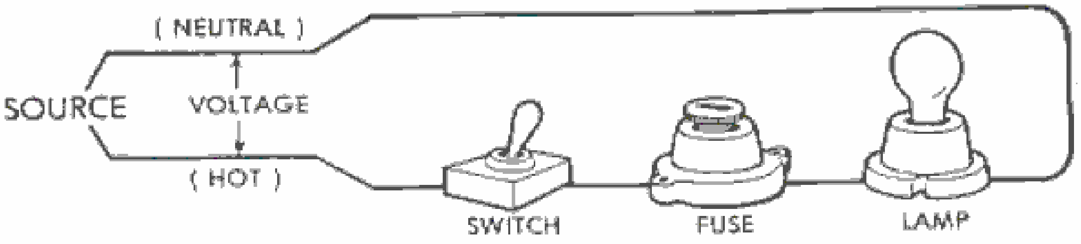

A circuit includes a "hot" wire (red or black) carrying current from the source through a switch, circuit protector (fuse or circuit-breaker), and an appliance. The neutral wire (white) conducts the current from the appliance to the source (ground).

There are two methods for connecting devices in a circuit - "in series" or "in parallel". In a series circuit, all of the current must flow through each device in the circuit. Removing any one of the devices in a series circuit will stop the flow of current. In parallel circuits, the load (lights or appliances) are connected between the two wires of the circuit providing an independent path for the flow of current, and removing a lamp has no effect on the other lamps in the circuit.

Switches, fuses, and circuit breakers are always connected in series. In most cases, except for some Christmas tree lights, appliances and lights are connected in parallel.

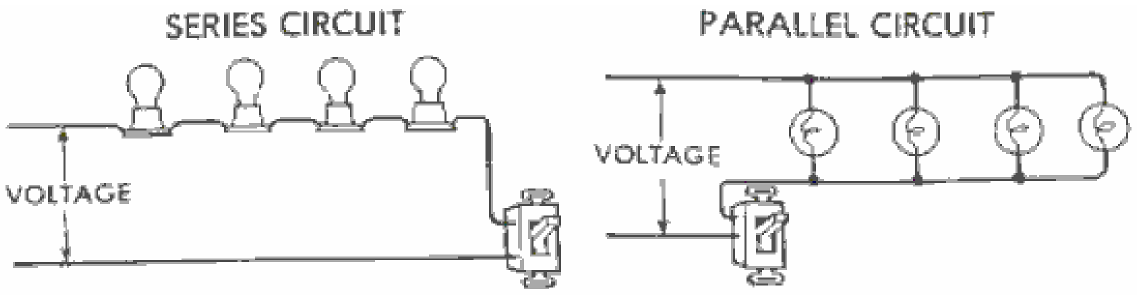

6. 120 Volt and 240 Volt Circuits

The 120V circuit has one "hot" and one neutral wire, with the switch and circuit protector in the hot line. The neutral wire from the appliance is connected to the neutral bar in the fuse or breaker box. For safety, the neutral wire should never be broken or interrupted with a switch or fuse.

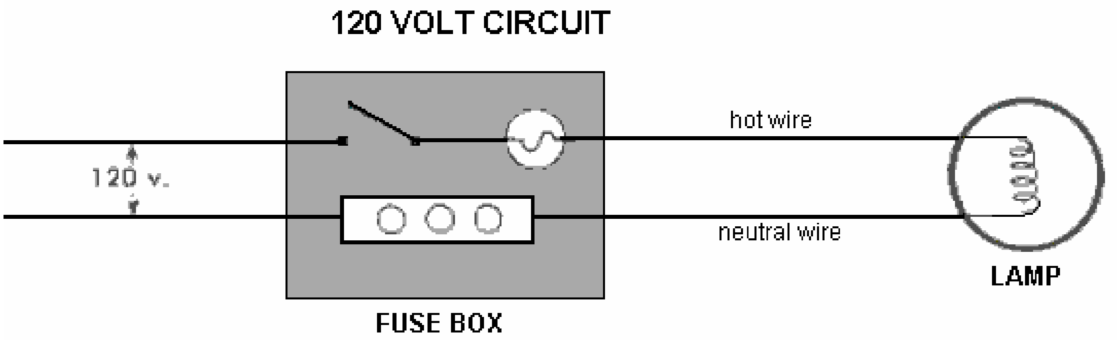

The voltage in a 120V circuit is measured with a voltmeter with one lead on the hot terminal and the other lead on the neutral bar. The number of amperes flowing may be measured with a clamp-on ammeter by encircling the hot or neutral wire with the jaws of the ammeter.

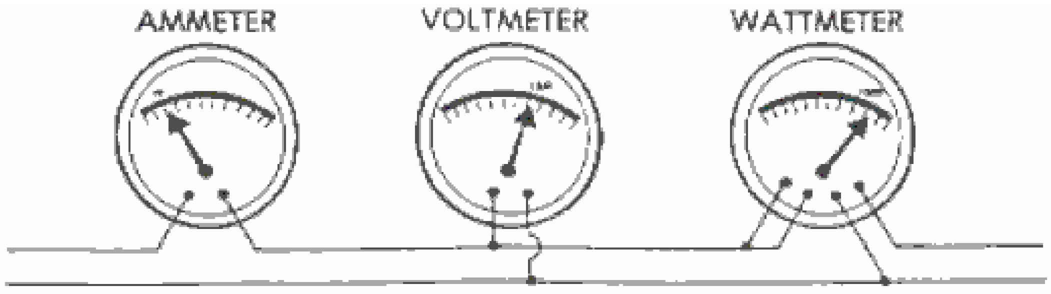

The 240V circuit has two hot wires and one safety-ground wire. Switches and fuses are installed in the hot lines. The two hot wires arc necessary for the operation of 240V welders and motors. The safety-ground wire, connected to the metal frame of the equipment or motor and to the neutral bar, does not carry current unless a "short" develops in the motor or welder. If a short should occur, one of the circuit protectors will burn-out or open, thus opening the circuit.

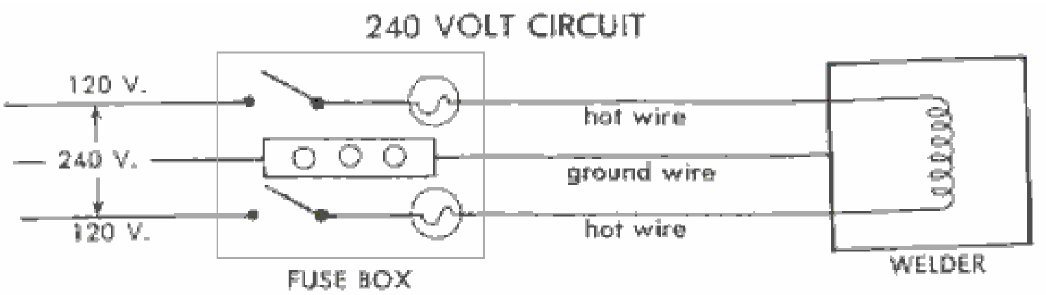

The voltage on a 240V circuit is measured by fastening a lead on the voltmeter to each of the hot wires. Voltage between either hot terminal and the neutral bar will be one-half of the voltage between the two hot wires. The number of amperes flowing can be measured by clamping an ammeter around either of the hot wires.

7. Safety Grounding of Electrical Equipment

Refer back, to the 240V circuit and note the ground wire from the metal frame to the neutral bar. The following illustration shows proper safety grounding when operating a drill in a 120V circuit. The safety-ground wire may be bare, but a three-wire cable is recommended. Safety-ground wire in three-wire cable is usually green in color. A current-carrying neutral wire should never be used for a safety-ground. Likewise, a safety-ground wire should never be used as a current-carrying hot or neutral wire.

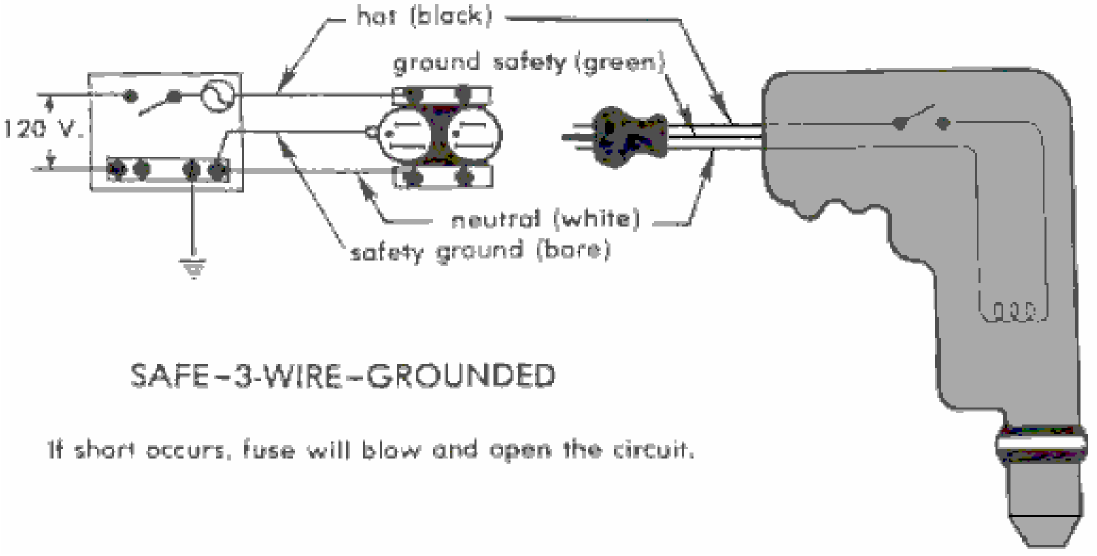

SAFE-3-WIRE-GROUNDED

If short occurs, fuse will blow and open the circuit.

Using grounded receptacles and a safety-ground on all circuits will allow the safety-grounding of appliances when they are plugged into the outlet. An adapter must be used to properly ground appliances connected to receptacles which are not safety-grounded. If an adapter is used, the green pigtail wire must be connected to a known ground to give protection from electrical shock should a short-circuit occur.

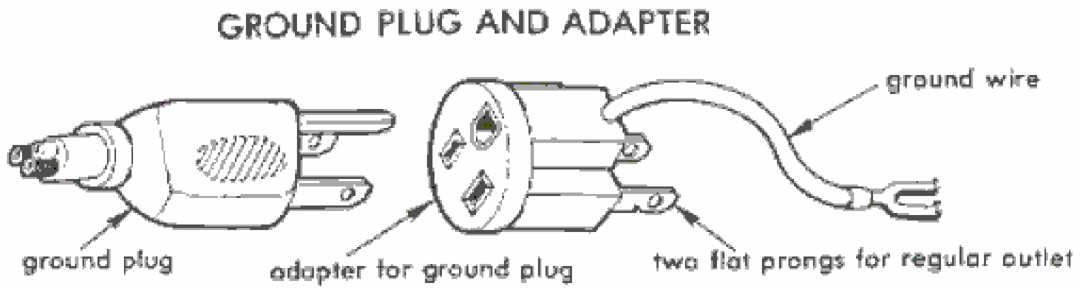

A test lamp can be used to check a circuit completed between a "hot" wire and a neutral wire. Use the test lamp to check appliances for shorts. With the appliance plugged into an outlet, touch the appliance frame with one lead of the test lamp while the other lead of the test lamp is grounded to a water or gas pipe. If the test light does not burn, reverse the appliance plug and check with the test lamp again. If the light burns, a short exists (the hot wire is touching the frame of the appliance). Unplug the appliance and repair or discard it.

8. Electrical Circuit Protection

Electrical circuits should be protected from an overload of amperes. Too many amperes flowing through an unprotected circuit will generate heat, which will deteriorate or melt the insulation and possibly cause a fire. The number of amperes that a given conductor can safely carry, depends upon the kind and size of wire, type of insulation, length of run in feet, and the type of installation. Charts are available in reference texts giving allowable current-carrying capabilities of various conductors.

The four types of circuit protection are: common fuses, fusetrons (time-delay), fustats (two-part time-delay), and circuit-breakers. Fuses are of two basic types: plug, and cartridge.

Common fuses contain a link made from a low-temperature melting alloy which is designed to carry current up to the rating of the fuse. Current higher than the amperage rating causes the link to heat above it's melting point. When the fuse "blows", the link melts and opens the circuit.

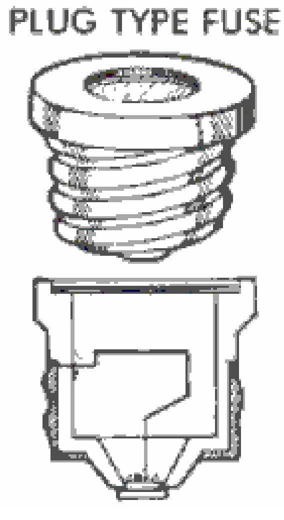

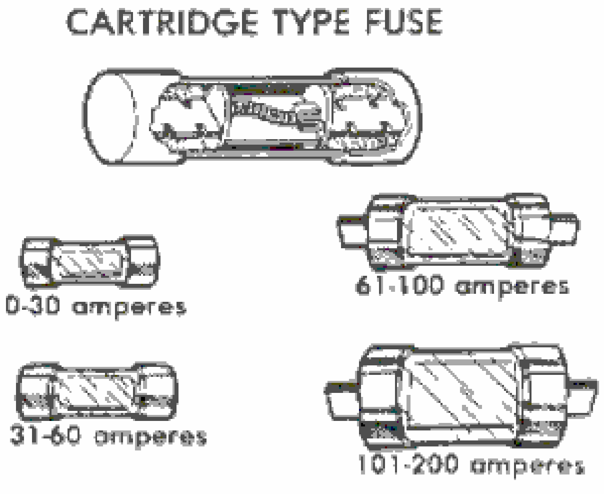

Fusetrons (time-delay fuses) are made to carry a temporary overload, such as the overload caused by the starting of an electric motor. The fuse, however, still provides protection for the circuit, and a short-circuit will melt the fuse link. If a common fuse is used, the fuse link will melt every time an electric motor starts. The use of a larger ampere common fuse will prevent the "blow" resulting from the temporary overload, but will not provide protection for the motor or the circuit.

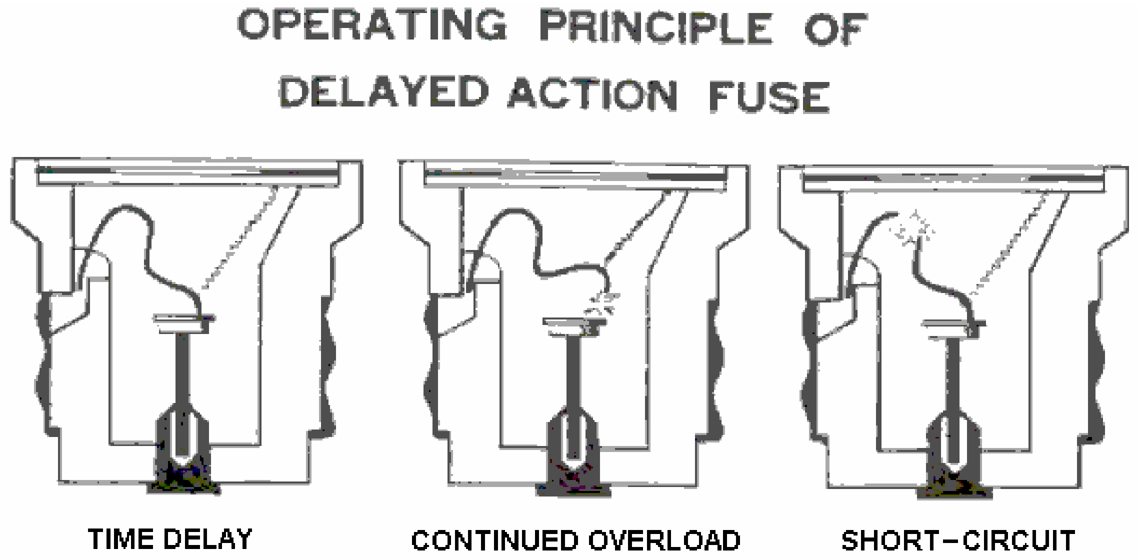
OPERATING PRINCIPLE OF
DELAYED ACTION FUSE
TIME DELAY
CONTINUED OVERLOAD
SHORT-CIRCUIT

Fustats, non-tamperable fuses of the time-delay type, have a different size base and require a special adapter which is screwed into the standard fuse socket. After the adapter is installed, it cannot be removed. For example, the installation of a 15-ampere adapter allows only the use of 15-ampere or smaller fuse.

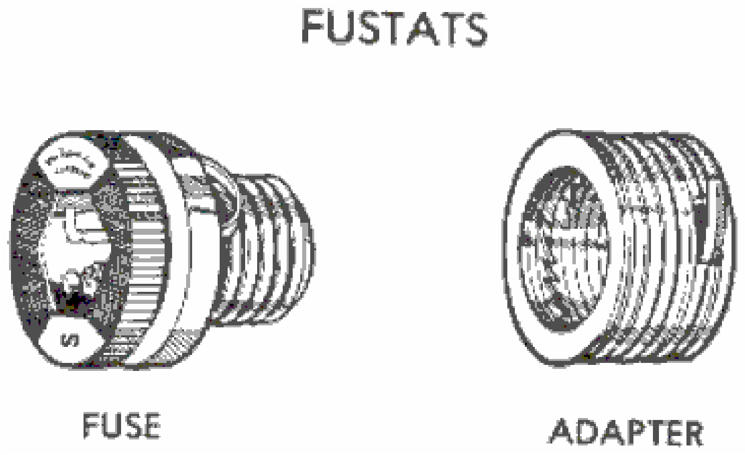
FUSTATS
S
FUSE
ADAPTER

Circuit breakers eliminate the replacement of fuses and are commonly used even though a circuit breaker box costs more than a fuse box. Circuit breakers are of two types, thermal and magnetic. The thermal breaker has two contacts held together by a bi-metal latch. A current overload causes the bi-metallic strip to become heated, the latch releases, and the points spring open. After the bi-metallic strip cools, the switch is reset and service is restored.

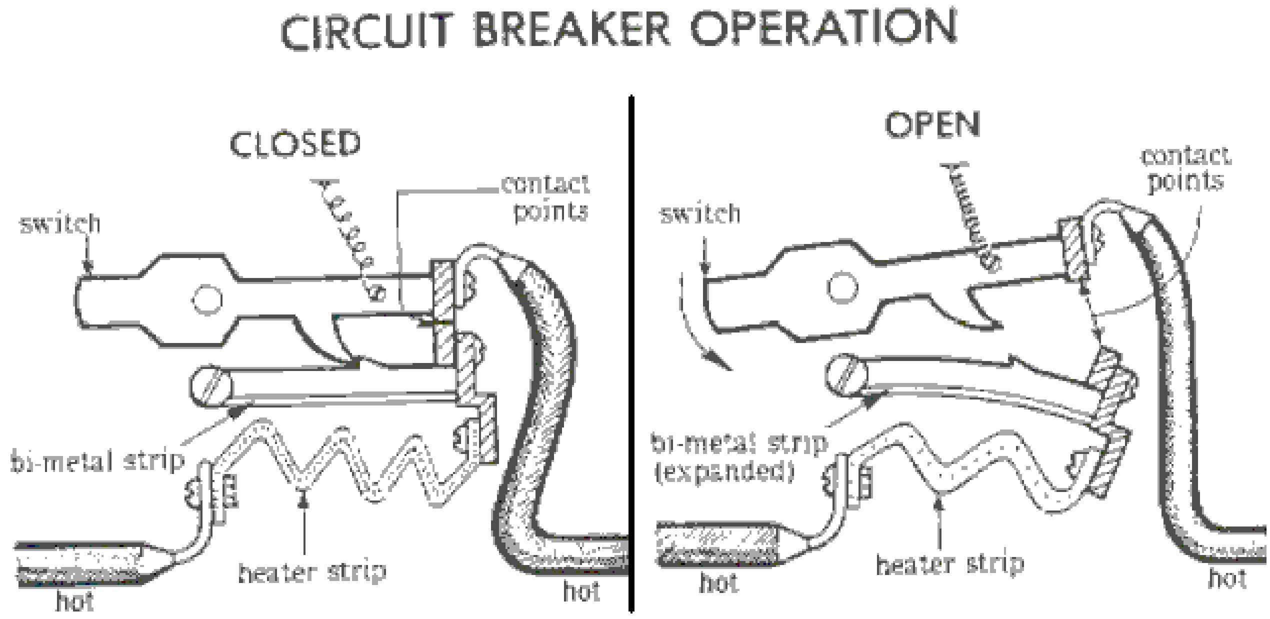
CIRCUIT BREAKER OPERATION
CLOSED
switch
contact points
bi-metal strip
heater strip
hot
hot
OPEN
switch
contact points
bi-metal strip (expanded)
heater strip
hot
hot

The magnetic breaker has contacts that are held together by a latch which is released by the action of an electromagnet. The amount of current flowing through the circuit will determine the size of the electromagnet. This type of circuit-breaker is reset by moving the toggle switch to the "on" position.

The following diagram shows the parts of a circuit breaker.

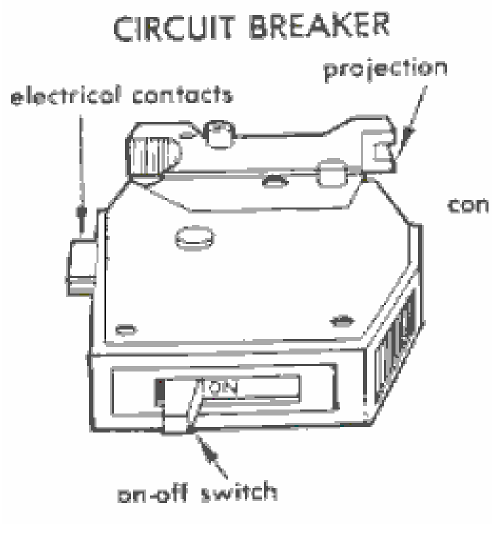
CIRCUIT BREAKER
electrical contacts
projection
ON
on-off switch

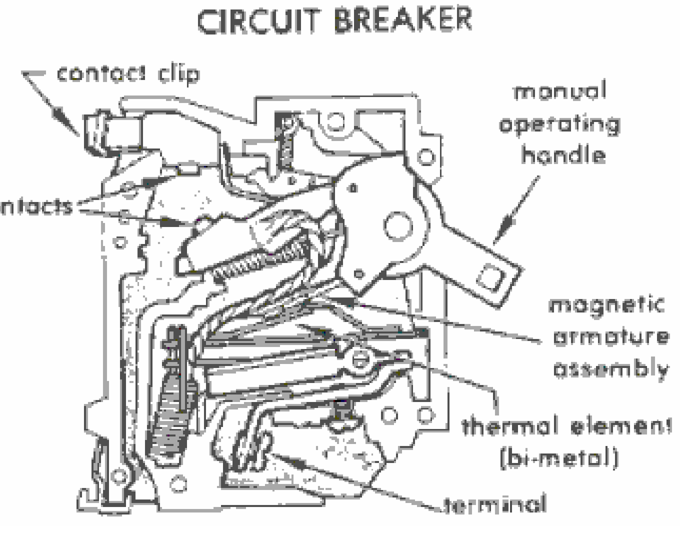
CIRCUIT BREAKER
contact clip
manual operating handle
contacts
magnetic armature assembly
thermal element (bi-metal)
terminal

9. No Fault Grounding

Fuses and circuit-breakers are safety devices which limit current (amperage) in a circuit. Their main function is to protect equipment and wiring from overload. Ground fault circuit interrupters (GFI) are designed to protect humans, equipment, and/or electrical systems from injury or damage if electricity flows in an unintended path (a short-circuit).

A GFI is a very sensitive device that functions by comparing the current moving in the "hot" wire with that in the neutral wire. If these two currents are not equal, a fault exists, and current is "leaking" out of the circuit. If the difference in current between the two wires is 5/1000 of an ampere or greater, the GFI will open the circuit, shutting off the power and eliminating any shock hazard.

The National Electrical Code requires GFI's for all 120V, single phase, 15 and 20 amp receptacles installed outdoors, in bathrooms, and in garages for residential buildings. A GFI is required at construction sites and some other applications. After correcting a circuit fault, the GFI may be reset for further use.

A variety of GFI equipment is made for 120 and 240 volt circuits:

GFI RECEPTACLE FOR CONVENTIONAL OUTLET

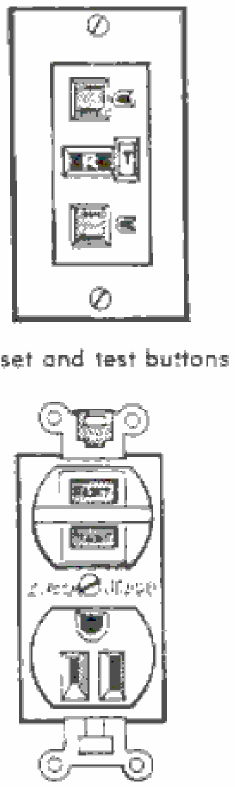

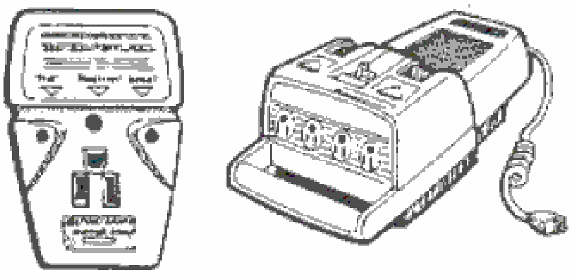

PORTABLE GFI TO PLUG INTO OUTLET

GFI TO INSTALL IN BRANCH CIRCUIT

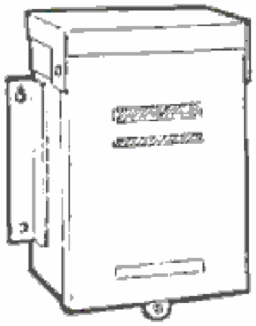

GFI TO REPLACE CIRCUIT BREAKER

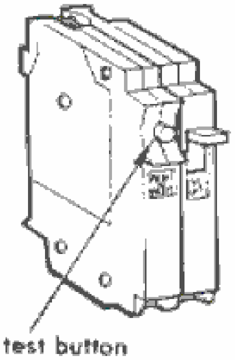

REFERENCES:

COOPER, ELMER L., *AGRICULTURAL MECHANICS: FUNDAMENTALS AND APPLICATIONS*. DELMAR PUBLISHERS INC., ALBANY, NEW YORK

ELECTRICAL WIRING - RESIDENTIAL, UTILITY BUILDINGS, SERVICE AREAS, AAVIM, ATHENS, GEORGIA.

Note: This electrical information does not apply directly to areas outside America and local regulations for electrical supply should be checked.

The following material is the section in Chapter 3 of "The Practical Guide to Free-Energy Devices" which deals with Don Smith's technology:

Don Smith. One of most impressive developers of free-energy devices is Don Smith who has produced many spectacular devices, generally with major power output. These are a result of his in-depth knowledge and understanding of the way that the environment works. Don says that his understanding comes from the work of Nikola Tesla as recorded in Thomas C. Martin's book "The Inventions, Researches, and Writings of Nikola Tesla" ISBN 0-7873-0582-0 available from http://www.healthresearchbooks.com and various other book companies. This book can be downloaded from http://www.free-energy-info.com as a pdf file, but a paper copy is much better quality and easier to work from.

Don states that he repeated each of the experiments found in the book and that gave him his understanding of what he prefers to describe as the 'ambient background energy' which is called the 'zero-point energy field' elsewhere in this eBook. Don remarks that he has now advanced further than Tesla in this field, partly because of the devices now available to him and which were not available when Tesla was alive.

Don stresses two key points. Firstly, a dipole can cause a disturbance in the magnetic component of the 'ambient background' and that imbalance allows you to collect large amounts of electrical power, using capacitors and inductors (coils). Secondly, you can pick up as many powerful electrical outputs as you want from that one magnetic disturbance, without depleting the magnetic disturbance in any way. This allows massively more power output than the small power needed to create the magnetic disturbance in the first place. This is what produces a COP>1 device and Don has created nearly fifty different devices based on that understanding.

Although they get removed quite frequently, there is one video which is definitely worth watching if it is still there. It is located at http://www.metacafe.com/watch/2820531/don_smith_free_energy/ and was recorded in 2006. It covers a good deal of what Don has done. In the video, reference is made to Don's website but you will find that it has been taken over by Big Oil who have filled it with innocuous similar-sounding things of no consequence, apparently intended to confuse newcomers. A website which is run by Conny Öström of Sweden is http://www.johnnyfg.110mb.com/ and it has brief details of his prototypes and theory. You will find the only document of his which I could locate, here http://www.free-energy-info.com/Smith.pdf in pdf form and it contains the following patent on a most interesting device which appears to have no particular limit on the output power. This is a slightly re-worded copy of that patent as patents are generally worded in such a way as to make them difficult to understand.

Patent NL 02000035 A **20th May 2004** **Inventor: Donald Lee Smith**

TRANSFORMER GENERATOR MAGNETIC RESONANCE INTO ELECTRIC ENERGY

ABSTRACT

The present invention refers to an Electromagnetic Dipole Device and Method, where wasted radiated energy is transformed into useful energy. A Dipole as seen in Antenna Systems is adapted for use with capacitor plates in such a way that the Heaviside Current Component becomes a useful source of electrical energy.

DESCRIPTION

Technical Field:

This invention relates to loaded Dipole Antenna Systems and their Electromagnetic radiation. When used as a transformer with an appropriate energy collector system, it becomes a transformer/generator. The invention collects and converts energy which is radiated and wasted by conventional devices.

Background Art:

A search of the International Patent Database for closely related methods did not reveal any prior art with an interest in conserving radiated and wasted magnetic waves as useful energy.

DISCLOSURE OF THE INVENTION

The invention is a new and useful departure from transformer generator construction, such that radiated and wasted magnetic energy changes into useful electrical energy. Gauss meters show that much energy from conventional electromagnetic devices is radiated into the ambient background and wasted. In the case of conventional transformer generators, a radical change in the physical construction allows better access to the energy available. It is found that creating a dipole and inserting capacitor plates at right angles to the current flow, allows magnetic waves to change back into useful electrical (coulombs) energy. Magnetic waves passing through the capacitor plates do not degrade and the full impact of the available energy is accessed. One, or as many sets of capacitor plates as is desired, may be used. Each set makes an exact copy of the full force and effect of the energy present in the magnetic waves. The originating source is not depleted of degraded as is common in conventional transformers.

BRIEF DESCRIPTION OF THE DRAWINGS

The Dipole at right angles, allows the magnetic flux surrounding it to intercept the capacitor plate, or plates, at right angles. The electrons present are spun such that the electrical component of each electron is collected by the capacitor plates. Essential parts are the South and North component of an active Dipole. Examples presented here exist as fully functional prototypes and were engineer constructed and fully tested in use by the Inventor. In each of the three examples shown in the drawings, corresponding parts are used.

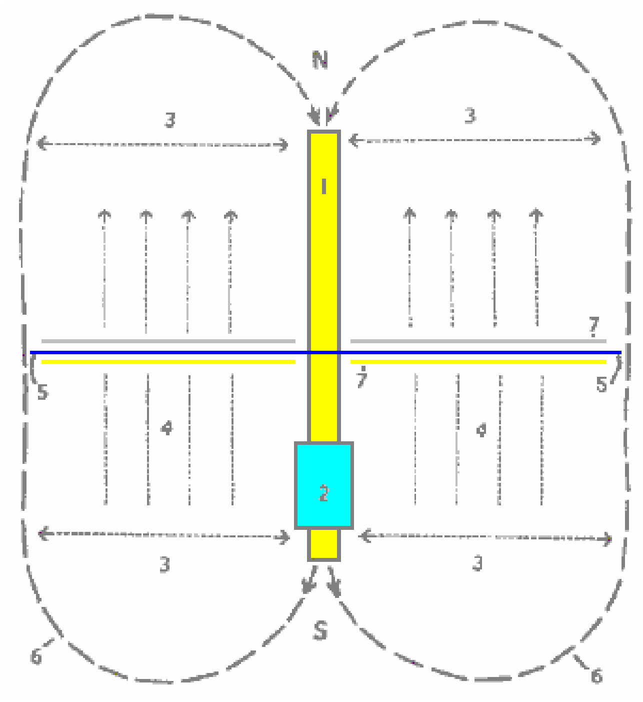

Fig.1 is a View of the Method, where **N** is the North and **S** is the South component of the Dipole.

Here, **1** marks the Dipole with its North and South components. **2** is a resonant high-voltage induction coil. **3** indicates the position of the electromagnetic wave emission from the Dipole. **4** indicates the position and flow direction of the corresponding Heaviside current component of the energy flow caused by the induction coil **2**. **5** is the dielectric separator for the capacitor plates **7**. **6** for the purposes of this drawing, indicates a virtual limit for the scope of the electromagnetic wave energy.

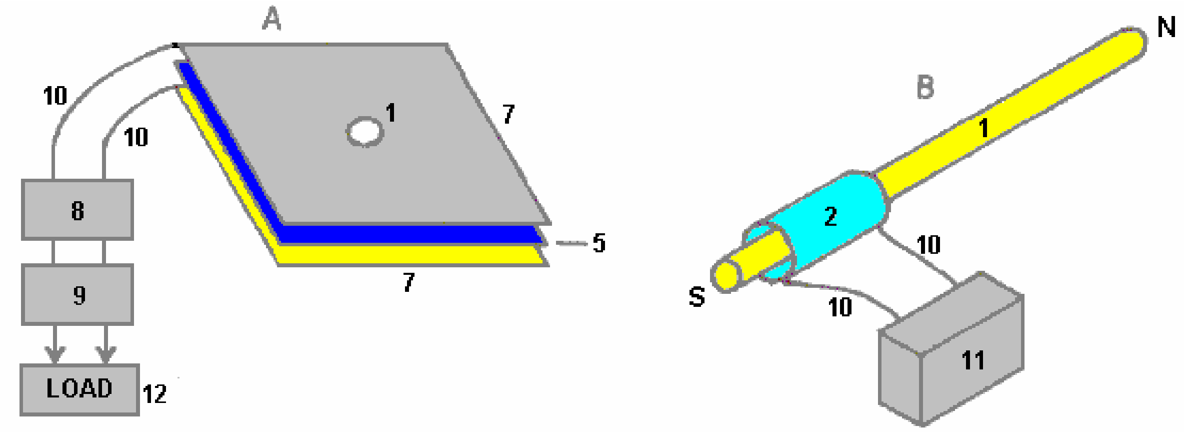

Fig.2 has two parts **A** and **B**.

In **Fig.2A** **1** is the hole in the capacitor plates through which the Dipole is inserted and in **Fig.2B** it is the Dipole with its North and South poles shown. **2** is the resonant high-voltage induction coil surrounding part of the Dipole **1**. The dielectric separator **5**, is a thin sheet of plastic placed between the two capacitor plates **7**, the upper plate being made of aluminium and the lower plate made of copper. Unit **8** is a deep-cycle battery system powering a DC inverter **9** which produces 120 volts at 60 Hz (the US mains supply voltage and frequency, obviously, a 240 volt 50 Hz inverter could be used here just as easily) which is used to power whatever equipment is to be driven by the device. The reference number **10** just indicates connecting wires. Unit **11** is a high-voltage generating device such as a neon transformer with its oscillating power supply.

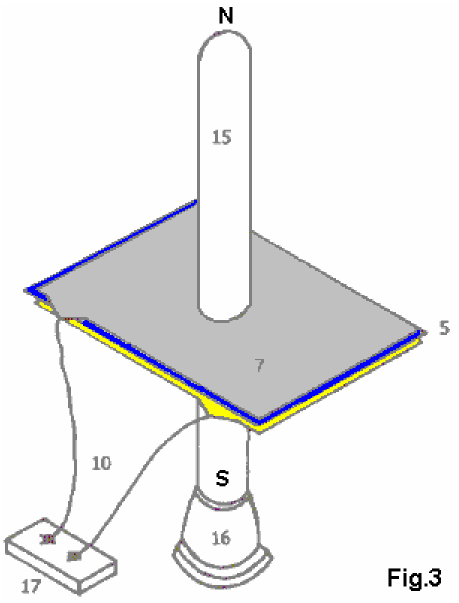

Fig.3

Fig.3 is a Proof Of Principal Device using a Plasma Tube as an active Dipole. In this drawing, **5** is the plastic sheet dielectric separator of the two plates **7** of the capacitor, the upper plate being aluminium and the lower plate copper. The connecting wires are marked **10** and the plasma tube is designated **15**. The plasma tube is four feet long (1.22 m) and six inches (100 mm) in diameter. The high-voltage energy source for the active plasma dipole is marked **16** and there is a connector box **17** shown as that is a convenient method of connecting to the capacitor plates when running tests on the device.

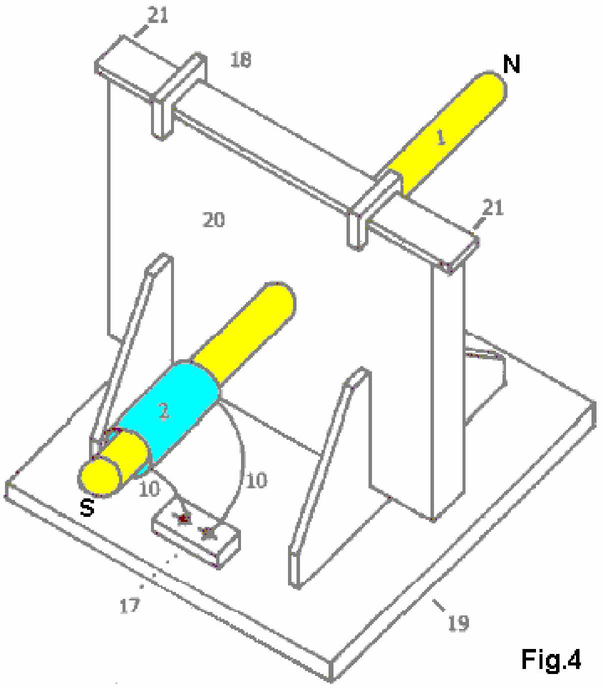

Fig.4

Fig.4 shows a Manufacturer's Prototype, constructed and fully tested. **1** is a metal Dipole rod and **2** the resonant high-voltage induction coil, connected through wires **10** to connector block **17** which facilitates the connection of it's high-voltage power supply. Clamps **18** hold the upper edge of the capacitor packet in place and **19** is the base plate with it's supporting brackets which hold the whole device in place. **20** is a housing which contains the capacitor plates and **21** is the point at which the power output from the capacitor plates is drawn off and fed to the DC inverter.

BEST METHOD OF CARRYING OUT THE INVENTION

The invention is applicable to any and all electrical energy requirements. The small size and it's high efficiency make it an attractive option, especially for remote areas, homes, office buildings, factories, shopping centres, public places, transportation, water systems, electric trains, boats, ships and 'all things great and small'. The construction materials are commonly available and only moderate skill levels are needed to make the device.

CLAIMS

1. Radiated magnetic flux from the Dipole, when intercepted by capacitor plates at right angles, changes into useful electrical energy.

2. A Device and Method for converting for use, normally wasted electromagnetic energy.

3. The Dipole of the Invention is any resonating substance such as Metal Rods, Coils and Plasma Tubes which have interacting Positive and Negative components.

4. The resulting Heaviside current component is changed to useful electrical energy.

****

This patent does not make it clear that the device needs to be tuned and that the tuning is related to its physical location. The tuning will be accomplished by applying a variable-frequency input signal to the neon transformer and adjusting that input frequency to give the maximum output.

Don Smith has produced some forty eight different devices, and because he understands that the real power in the universe is magnetic and not electric, these devices have performances which appear staggering to people trained to think that electrical power is the only source of power. One of his devices is shown here:

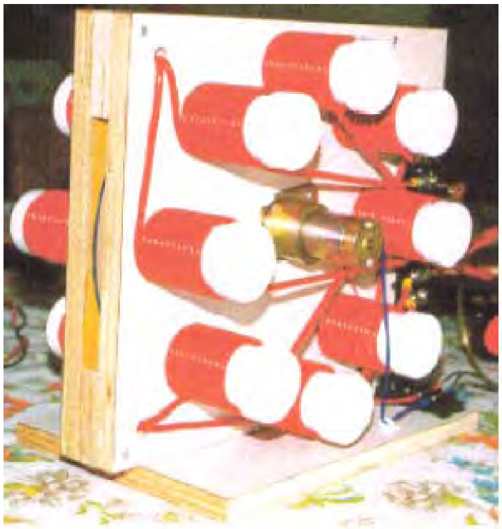

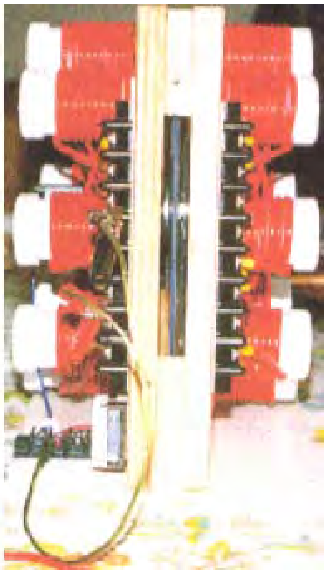

This is a small table-top device which looks like it is an experiment by a beginner, and something which would be wholly ineffective. Nothing could be further from the truth. Each of the eight coils pairs (one each side of the rotating disc) produces 1,000 volts at 50 amps (fifty kilowatts) of output power, giving a total power output of 400 kilowatts. It's overall size is 16" x 14.5" x 10" (400 x 370 x 255 mm). In spite of the extremely high power output, the general construction is very simple:

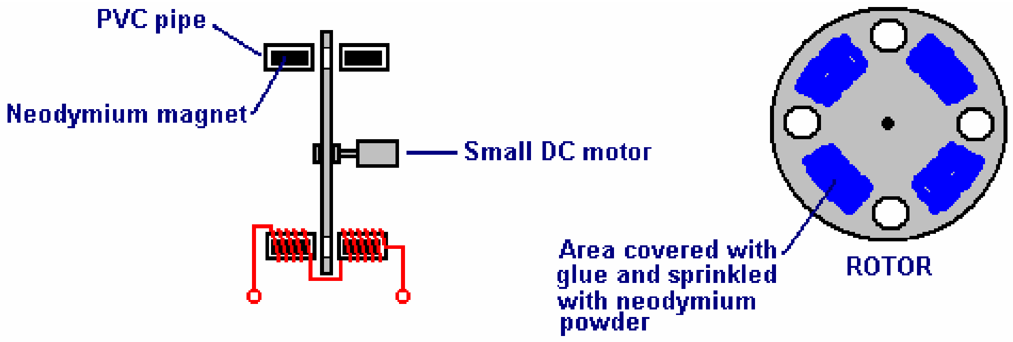

The device operates on a fluctuating magnetic field which is produced by a small low-power DC motor spinning a plastic disc. In the prototype shown above, the disc is an old vinyl record which has had holes cut in it. Between the holes is an area which was covered with glue and then sprinkled with powdered neodymium magnet material. It takes very little power to spin the disc, but it acts in a way which is very much like the Ecklin-Brown generator, repeatedly disrupting the magnetic field. The magnetic field is created by a neodymium magnet in each of the sixteen plastic pipes. It is important that the change in the magnetic flux between the matching magnets on each side of the disc is as large as possible. The ideal rotor material for this is "Terfenol-D" (tungsten zirconate) with slots cut in it but it is so expensive that materials like stainless steel are likely to be used instead. Please understand that all of Don's designs rely on resonant operation and so the coil impedance has to be matched to the pulse frequency used to drive the coil.

For Don Smith, this is not an exceptional device. The one shown below is also physically quite small and yet it has an output of 160 kilowatts (8000 volts at 20 amps) from an input of 12 volts 1 amp (COP = 13,333):

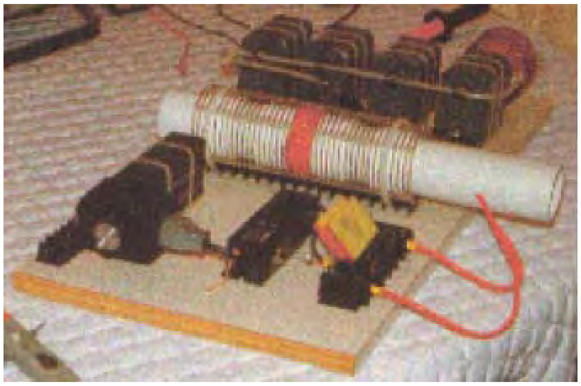

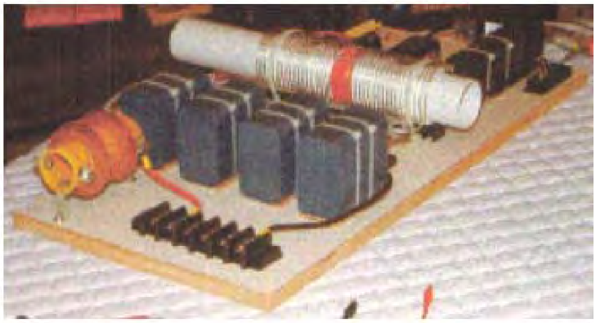

Again, this is a device which can be placed on top of a table and is not a complicated form of construction, having a very open and simplistic layout. However, some components are not mounted on this board. The twelve volt battery and connecting leads are not shown, nor is the ground connection, the step-down isolation transformer and the varistor used to protect the load from over-voltage by absorbing any random voltage spikes which might occur, but more of these things later on when a much more detailed description of this device is given. Again, please understand that Don does not reveal all of the details of any of his designs, and he deliberately omits to mention various important details, leaving us to deduce what is missing from our own understanding of how these devices work.

The device shown above is a typical example of this with various subtle points glossed over in spite of this being one device which Don says that we should be able to reproduce ourselves. Let me state here that reproducing this seemingly simple design of Don's is not an easy thing to do and it is not something which can be thrown together by a beginner using whatever components happen to be at hand at the time. Having said that, with careful study and commonsense application of some obvious facts, it should be possible to make one of these devices.

Another of Don's devices is shown here:

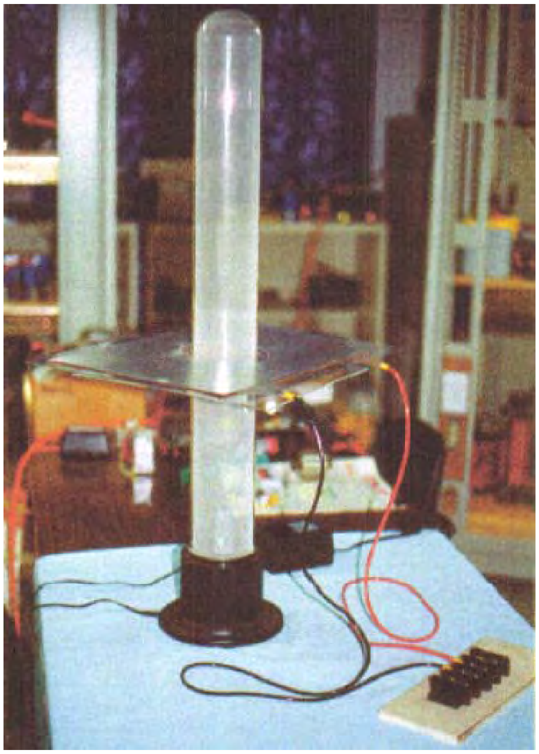

This is a larger device which uses a plasma tube four feet (1.22 m) long and 6 inches (100 mm) in diameter. The output is a massive 100 kilowatts. This is the design shown as one of the options in Don's patent. Being an Electrical Engineer, none of Don's prototypes are in the "toy" category. If nothing else is taken from Don's work, we should realize that high power outputs can be had from very simple devices.

There is one other brief document "Resonate Electrical Power System" from Don Smith which says:

Potential Energy is everywhere at all times, becoming useful when converted into a more practical form. There is no energy shortage, only grey matter. This energy potential is observed indirectly through the manifestation of electromagnetic phenomenon, when intercepted and converted, becomes useful. In nonlinear systems, interaction of magnetic waves amplify (conjugate) energy, providing greater output than input. In simple form, in the piano where three strings are struck by the hammer, the centre one is impacted and resonance activates the side strings. Resonance between the three strings provides a sound level greater than the input energy. Sound is part of the electromagnetic spectrum and is subject to all that is applicable to it.

"Useful Energy" is defined as "that which is other than Ambient". "Electric Potential" relates to mass and it's acceleration. Therefore, the Earth's Mass and Speed through space, gives it an enormous electrical potential. Humans are like the bird sitting unaware on a high voltage line. in nature, turbulence upsets ambient and we see electrical displays. Tampering with ambient, allows humans to convert magnetic waves into useful electricity.

Putting this in focus, requires a look at the Earth in general. During each of the 1,440 minutes of each day, more than 4,000 displays of lightning occur. Each display yields more than 10,000,000 volts at more than 200,000 amperes in equivalent electromagnetic flux. This is more than 57,600,000,000,000 volts and 1,152,000,000,000 amperes of electromagnetic flux during each 24 hour period. This has been going on for more than 4 billion years. The USPTO insist that the Earth's electrical field is insignificant and useless, and that converting this energy violates the laws of nature. At the same time, they issue patents in which, electromagnetic flux coming in from the Sun is converted by solar cells into DC energy. Aeromagnetic flux (in gammas) Maps World-Wide, includes those provided by the US Department of Interior-Geological Survey, and these show clearly that there is present, a spread of 1,900 gamma above Ambient, from reading instruments flown 1,000 feet above the (surface) source. Coulomb's Law requires the squaring of the distance of the remote reading, multiplied by the recorded reading. Therefore, that reading of 1,900 gamma has a corrected value of 1,900 x 1,000 x 1,000 = 1,900,000,000 gamma.

There is a tendency to confuse "gamma ray" with "gamma". "Gamma" is ordinary, everyday magnetic flux, while "gamma ray" is high-impact energy and not flux. One gamma of magnetic flux is equal to that of 100 volts RMS. To see this, take a Plasma Globe emitting 40,000 volts. When properly used, a gamma meter placed nearby, will read 400 gammas. The 1,900,000,000 gamma just mentioned, is the magnetic ambient equivalent of

190,000,000 volts of electricity. This is on a "Solar Quiet" day. On "Solar Active" days it may exceed five times that amount. The Establishment's idea that the Earth's electrical field is insignificant, goes the way of their other great ideas.

There are two kinds of electricity: "potential" and "useful". All electricity is "potential" until it is converted. The resonant-fluxing of electrons, activates the electrical potential which is present everywhere. The Intensity/CPS of the resonant-frequency-flux rate, sets the available energy. This must then be converted into the required physical dimensions of the equipment being used. For example, energy arriving from the Sun is magnetic flux, which solar cells convert to DC electricity, which is then converted further to suit the equipment being powered by it. Only the magnetic flux moves from point "A" (the Sun) to point "B" (the Earth). All electrical power systems work in exactly the same way. Movement of Coils and Magnets at point "A" (the generator) fluxes electrons, which in turn, excite electrons at point "B" (your house). **None of the electrons at point "A" are ever transmitted to point "B"**. In both cases, the electrons remain forever intact and available for further fluxing. This is not allowed by Newtonian Physics (electrodynamics and the laws of conservation). Clearly, these laws are all screwed up and inadequate.

In modern physics, USPTO style, all of the above cannot exist because it opens a door to overunity. The good news is that the PTO has already issued hundreds of Patents related to Light Amplification, all of which are overunity. The Dynode used to adjust the self-powered shutter in your camera, receives magnetic flux from light which dislodges electrons from the cathode, reflecting electrons through the dynode bridge to the anode, resulting in billions of more electrons out than in. There are currently, 297 direct patents issued for this system, and thousands of peripheral patents, all of which support overunity. More than a thousand other Patents which have been issued, can be seen by the discerning eye to be overunity devices. What does this indicate about Intellectual Honesty?

Any coil system, when fluxed, causes electrons to spin and produce useful energy, once it is converted to the style required by its use. Now that we have described the method which is required, let us now see how this concerns us.

The entire System already exists and all that we need to do is to hook it up in a way which is useful to our required manner of use. Let us examine this backwards and start with a conventional output transformer. Consider one which has the required voltage and current handling characteristics and which acts as an isolation transformer. Only the magnetic flux passes from the input winding to the output winding. No electrons pass through from the input side to the output side. Therefore, we only need to flux the output side of the transformer to have an electrical output. Bad design by the establishment, allowing hysteresis of the metal plates, limits the load which can be driven. Up to this point, only potential is a consideration. Heat (which is energy loss) limits the output amperage. Correctly designed composite cores run cool, not hot.

A power correction factor system, being a capacitor bank, maintains an even flow of flux. These same capacitors, when used with a coil system (a transformer) become a frequency-timing system. Therefore, the inductance of the input side of the transformer, when combined with the capacitor bank, provides the required fluxing to produce the required electrical energy (cycles per second).

With the downstream system in place, all that is needed now is a potential system. Any flux system will be suitable. Any amplification over-unity output type is desirable. The input system is point "A" and the output system is point "B". Any input system where a lesser amount of electrons disturbs a greater amount of electrons - producing an output which is greater than the input - is desirable.

At this point, it is necessary to present updated information about electrons and the laws of physics. A large part of this, originates from me (Don Smith) and so is likely to upset people who are rigidly set in the thought patterns of conventional science.

Non - Ionic Electrons

As a source of electrical energy, non-ionic electrons doublets exist in immense quantities throughout the universe. Their origin is from the emanation of Solar Plasma. When ambient electrons are disturbed by being spun or pushed apart, they yield both magnetic and electrical energy. The rate of disturbance (cycling) determines the energy level achieved. Practical methods of disturbing them include, moving coils past magnets or vice versa. A better way is the pulsing (resonant induction) with magnetic fields and waves near coils.

In coil systems, magnetic and amperage are one package. This suggests that electrons in their natural non-ionic state, exist as doublets. When pushed apart by agitation, one spins right (yielding Volts-potential electricity) and the other spins left (yielding Amperage-magnetic energy), one being more negative than the other. This further suggests that when they reunite, we have (Volts x Amps = Watts) useful electrical energy. Until now, this idea has been totally absent from the knowledge base. The previous definition of Amperage is therefore flawed.

Electron Related Energy

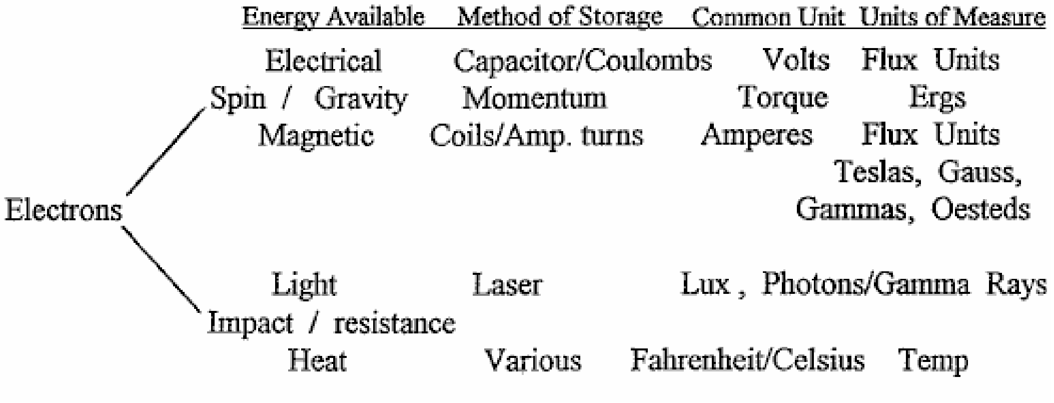

	Energy Available	Method of Storage	Common Unit	Units of Measure
Electrons	Electrical	Capacitor/Coulombs	Volts	Flux Units
	Spin / Gravity	Momentum	Torque	Ergs
	Magnetic	Coils/Amp. turns	Amperes	Flux Units Teslas, Gauss, Gammas, Oesteds
	Light	Laser	Lux ,	Photons/Gamma Rays
	Impact / resistance			
	Heat	Various	Fahrenheit/Celsius	Temp

Left hand spin of electrons results in Electrical Energy and right hand spin results in Magnetic Energy. Impacted electrons emit visible Light and heat.

Useful Circuits, Suggestions for Building an Operational Unit

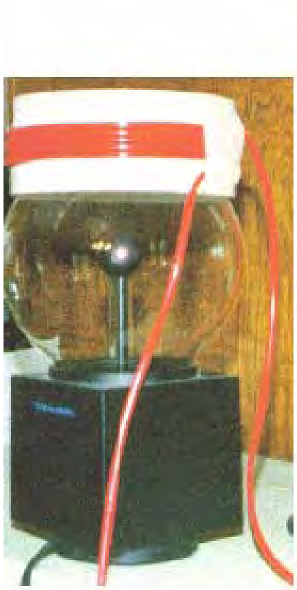

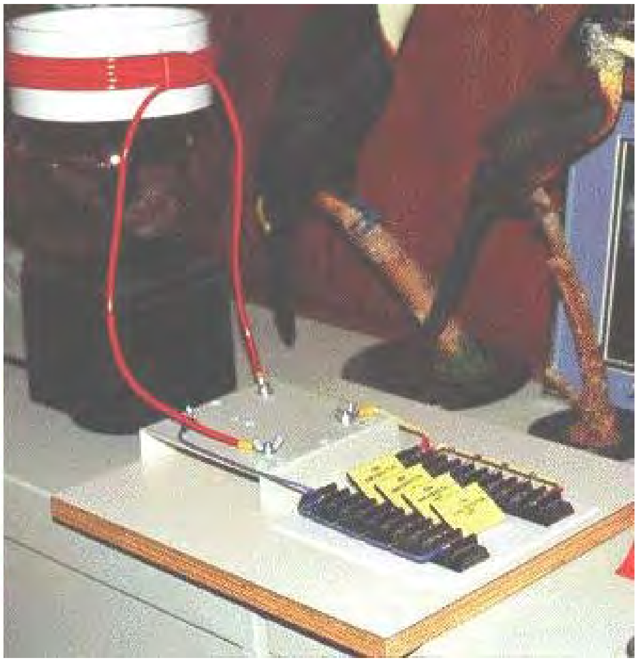

1. Substitute a Plasma Globe such as Radio Shack's "Illumna-Storm" for the source-resonant induction system. It will have about 400 milligauss of magnetic induction. One milligauss is equal to 100 volts worth of magnetic induction.

2. Construct a coil using a 5-inch to 7-inch (125 to 180 mm) diameter piece of PVC for the coil former.

3. Get about 30 feet (10 m) of Jumbo-Speaker Cable and separate the two strands. This can be done by sticking a carpet knife into a piece of cardboard or wood, and then pulling the cable carefully past the blade to separate the two insulated cores from each other. (PJK Note: "Jumbo-Speaker Cable" is a vague term as that cable comes in many varieties, with anything from a few, to over 500 strands in each core. As Don points out that the output power increases with each turn of wire, it is distinctly possible that each of these strands acts the same as individual insulated turns which have been connected in parallel, so a 500-strand cable may well be far more effective than a cable with just a few strands).

4. Wind the coil with 10 to 15 turns of wire and leave about 3 feet (1 m) of cable spare at each end of the coil. Use a glue gun to hold the start and finish of the coil.

5. This will become the "L - 2" coil shown in the Circuits page.

6. When sitting on top of the Plasma Globe (like a crown) you have a first-class resonant air-core coil system.

7. Now, substitute two or more capacitors (rated at 5,000 volts or more) for the capacitor bank shown on the Circuits page. I use more than two 34 microfarad capacitors.

8. Finish out the circuit as shown. You are now in business !

9. Voltage - Amperage limiting resistors are required across the output side of the Load transformer. These are used to adjust the output level and the desired cycles per second.

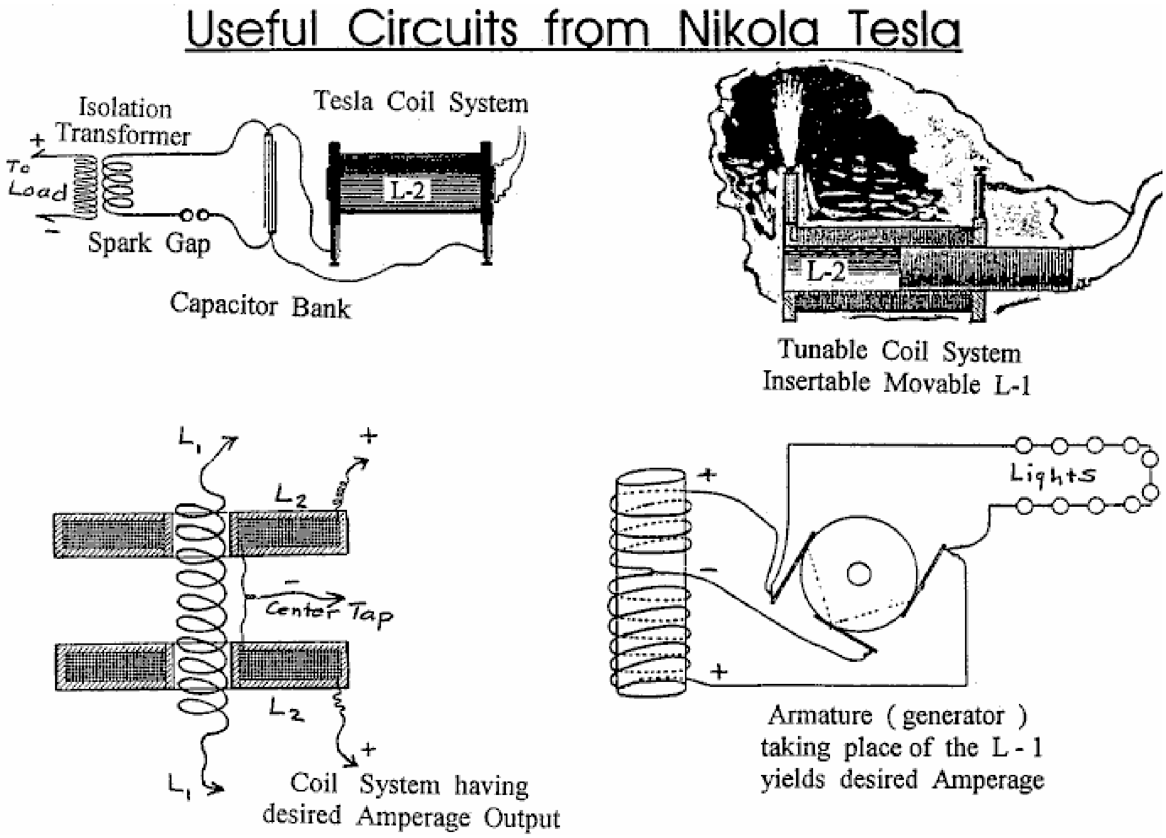
Useful Circuits from Nikola Tesla
Isolation Transformer
+
To Load
-
Tesla Coil System
L-2
Spark Gap
Capacitor Bank
L-2
Tunable Coil System
Insertable Movable L-1
L1
+
L2
-
Center Tap
L2
+
L1
Coil System having desired Amperage Output
+
-
+
Lights
Armature (generator) taking place of the L - 1 yields desired Amperage

Don Smith's Suggestions: Get a copy of the "Handbook of Electronic Tables and Formulas", published by Sams, ISBN 0-672-22469-0, also an Inductance/Capacitance/Resistance meter is required. Chapter 1 of Don's pdf document has important time-constant (frequency) information and a set of reactance charts in nomograph style (*"nomograph": a graph, usually containing three parallel scales graduated for different variables so that when a straight line connects values of any two, the related value may be read directly from the third at the point intersected by the line*) which makes working, and approximating of the three variables (capacitance, inductance and resistance) much easier. If two of the variables are known, then the third one can be read from the nomograph.

For example, if the input side of the isolation transformer needs to operate at 60 Hz, that is 60 positive cycles and 60 negative cycles, being a total of 120 cycles. Read off the inductance in Henries by using the Inductance meter attached to the input side of the isolation transformer. Plot this value on the (nomographic) reactance chart. Plot the needed 120 Hz on the chart and connect these two points with a straight line. Where this line crosses the Farads line and the Ohms line, gives us two values. Choose one (resistor) and insert it between the two leads of the transformer input winding.

The Power Correction Factor Capacitor (or bank of more than one capacitor) now needs adjusting. The following formula is helpful in finding this missing information. The capacitance is known, as is the desired potential to pulse the output transformer. One Farad of capacitance is one volt for one second (one Coulomb). Therefore, if we want to keep the bucket full with a certain amount, how many dippers full are needed? If the bucket needs 120 volts, then how many coulombs are required?

$$\frac{\text{Desired Voltage}}{\text{Capacitance in Microfarads}} = \text{Required frequency in Hz}$$

Now, go to the nomograph mentioned above, and find the required resistor jumper to place between the poles of the Correction Factor Capacitor.

An earth grounding is desirable, acting as both a voltage-limiter and a transient spike control. Two separate earths are necessary, one at the Power Factor Capacitor and one at the input side of the isolation transformer. Off-the-shelf surge arrestors / spark gaps and varistors having the desired voltage/potential and amperage control are commonly available. Siemens, Citel America and others, make a full range of surge arrestors, etc. Varistors look like coin-sized flat capacitors. Any of these voltage limiters are marked as "V - 1" in the following text.

It should be obvious that several separate closed circuits are present in the suggested configuration: The power input source, the high-voltage module, a power factor capacitor bank combined with the input side of the isolation transformer. Lastly, the output side of the isolation transformer and its load. None of the electrons active at the power source (battery) are passed through the system for use downstream. At any point, if the magnetic flux rate should happen to vary, then the number of active electrons also varies. Therefore, controlling the flux rate controls the electron (potential) activity. Electrons active at point "A" are not the same electrons which are active at point "B", or those at point "C", and so on. If the magnetic flux rate (frequency Hz) varies, then a different number of electrons will be disturbed. This does not violate any Natural Law and it does produce more output energy than the input energy, should that be desirable.

A convenient high-voltage module is a 12 volt DC neon tube transformer. The Power Factor Correction Capacitors should be as many microfarads as possible as this allows a lower operating frequency. The 12-volt neon tube transformer oscillates at about 30,000 Hz. At the Power Correction Factor Capacitor bank we lower the frequency to match the input side of the isolation transformer.

Other convenient high-voltage sources are car ignition coils, television flyback transformers, laser printer modules, and various other devices. Always lower the frequency at the Power Factor Correction Capacitor and correct, if needed, at the input side of the isolation transformer. The isolation transformer comes alive when pulsed. Amperage becomes a part of the consideration only at the isolation transformer. Faulty design, resulting in hysteresis, creates heat which self-destructs the transformer if it is overloaded. Transformers which have a composite core instead of the more common cores made from many layers of thin sheets of soft iron, run cool and can tolerate much higher amperage.

RESONATE ELECTROMAGNETIC POWER SYSTEM

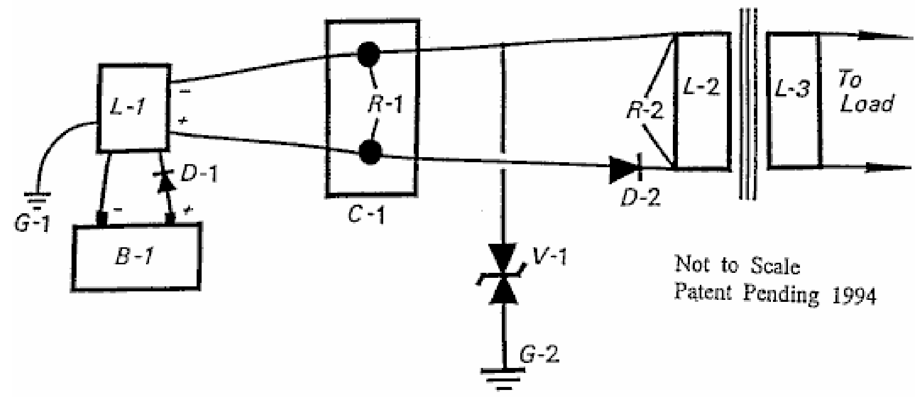

Not to Scale
Patent Pending 1994

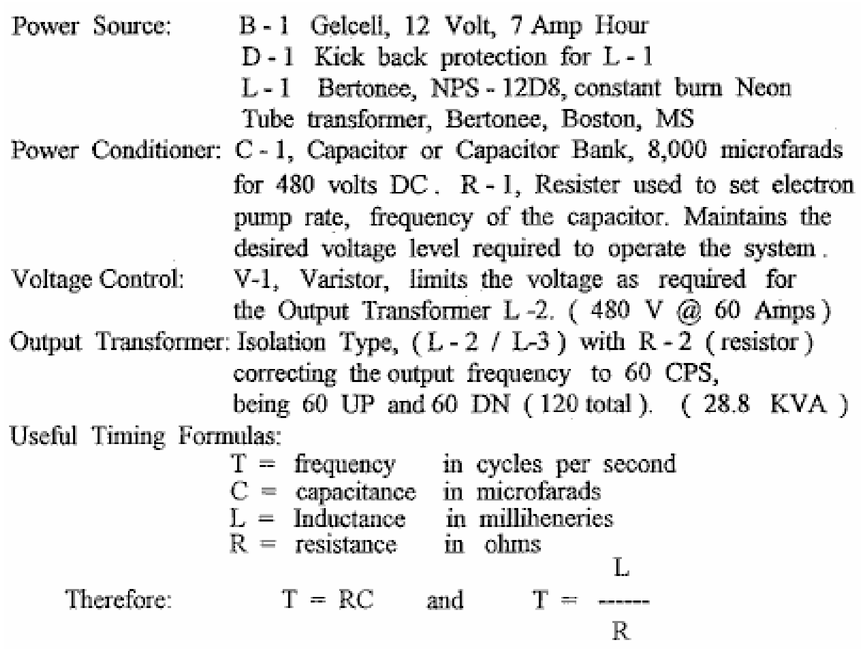

Power Source: B - 1 Gelcell, 12 Volt, 7 Amp Hour
D - 1 Kick back protection for L - 1
L - 1 Bertonee, NPS - 12D8, constant burn Neon Tube transformer, Bertonee, Boston, MS

Power Conditioner: C - 1, Capacitor or Capacitor Bank, 8,000 microfarads for 480 volts DC . R - 1, Resister used to set electron pump rate, frequency of the capacitor. Maintains the desired voltage level required to operate the system .

Voltage Control: V-1, Varistor, limits the voltage as required for the Output Transformer L -2. (480 V @ 60 Amps)

Output Transformer: Isolation Type, (L - 2 / L-3) with R - 2 (resistor) correcting the output frequency to 60 CPS, being 60 UP and 60 DN (120 total). (28.8 KVA)

Useful Timing Formulas:

T = frequency in cycles per second
C = capacitance in microfarads
L = Inductance in milliheneries
R = resistance in ohms

Therefore: $T = RC$ and $T = \frac{L}{R}$

The information shown above, relates to the small Suitcase Model demonstrated at the 1996 Tesla Convention, presented as Don Smiths' Workshop. This unit was a very primitive version and newer versions have atomic batteries and power output ranges of Gigawatts. The battery requirement is low level and is no more harmful than the radium on the dial of a clock. Commercial units of Boulder Dam size are currently being installed at several major locations throughout the world. For reasons of Don's personal security and contract obligations, the information which he has shared here, is incomplete.

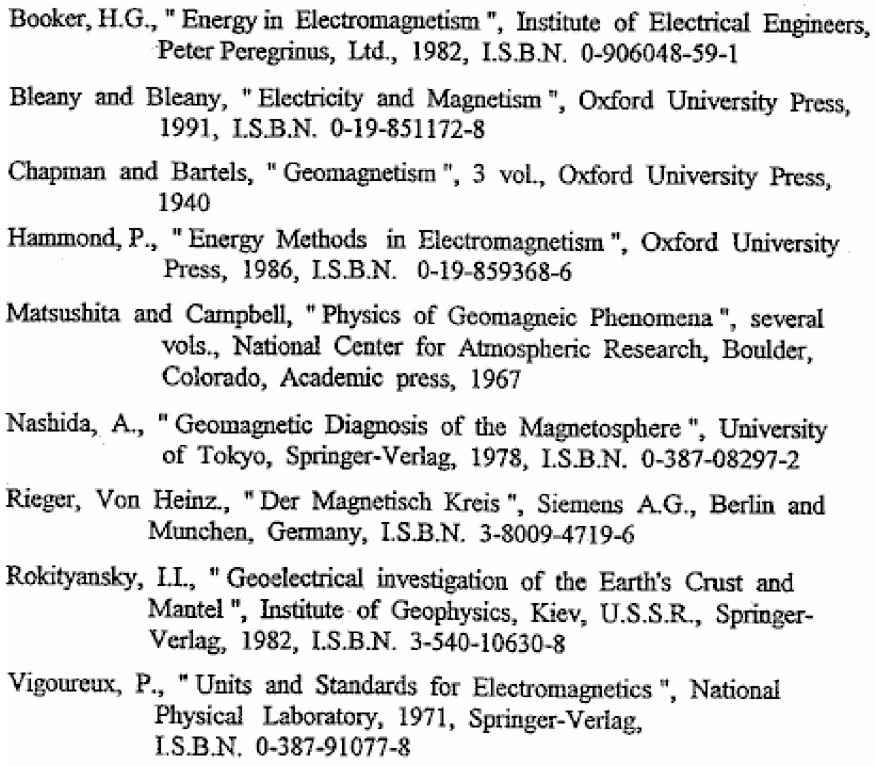

Booker, H.G., " Energy in Electromagnetism ", Institute of Electrical Engineers, Peter Peregrinus, Ltd., 1982, I.S.B.N. 0-906048-59-1

Bleany and Bleany, " Electricity and Magnetism ", Oxford University Press, 1991, I.S.B.N. 0-19-851172-8

Chapman and Bartels, " Geomagnetism ", 3 vol., Oxford University Press, 1940

Hammond, P., " Energy Methods in Electromagnetism ", Oxford University Press, 1986, I.S.B.N. 0-19-859368-6

Matsushita and Campbell, " Physics of Geomagneic Phenomena ", several vols., National Center for Atmospheric Research, Boulder, Colorado, Academic press, 1967

Nashida, A., " Geomagnetic Diagnosis of the Magnetosphere ", University of Tokyo, Springer-Verlag, 1978, I.S.B.N. 0-387-08297-2

Rieger, Von Heinz., " Der Magnetisch Kreis ", Siemens A.G., Berlin and Munchen, Germany, I.S.B.N. 3-8009-4719-6

Rokityansky, I.I., " Geoelectrical investigation of the Earth's Crust and Mantel ", Institute of Geophysics, Kiev, U.S.S.R., Springer-Verlag, 1982, I.S.B.N. 3-540-10630-8

Vigoureux, P., " Units and Standards for Electromagnetics ", National Physical Laboratory, 1971, Springer-Verlag, I.S.B.N. 0-387-91077-8

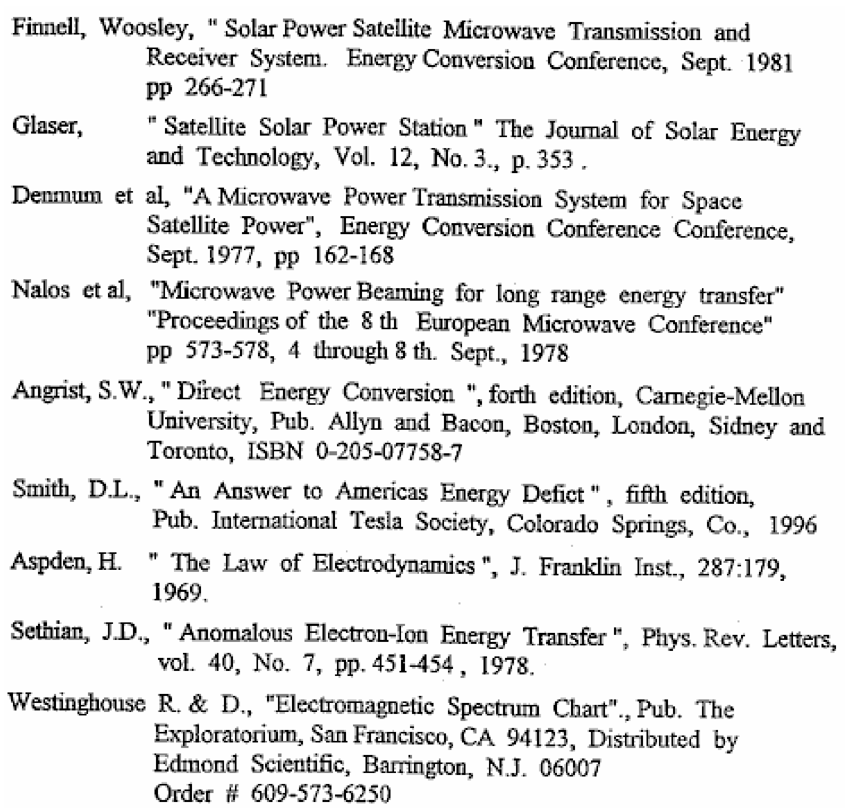

Finnell, Woosley, " Solar Power Satellite Microwave Transmission and Receiver System. Energy Conversion Conference, Sept. 1981 pp 266-271

Glaser, " Satellite Solar Power Station " The Journal of Solar Energy and Technology, Vol. 12, No. 3., p. 353 .

Denmum et al, "A Microwave Power Transmission System for Space Satellite Power", Energy Conversion Conference Conference, Sept. 1977, pp 162-168

Nalos et al, "Microwave Power Beaming for long range energy transfer" "Proceedings of the 8 th European Microwave Conference" pp 573-578, 4 through 8 th. Sept., 1978

Angrist, S.W., " Direct Energy Conversion ", forth edition, Carnegie-Mellon University, Pub. Allyn and Bacon, Boston, London, Sidney and Toronto, ISBN 0-205-07758-7

Smith, D.L., " An Answer to Americas Energy Defict " , fifth edition, Pub. International Tesla Society, Colorado Springs, Co., 1996

Aspden, H. " The Law of Electrodynamics ", J. Franklin Inst., 287:179, 1969.

Sethian, J.D., " Anomalous Electron-Ion Energy Transfer ", Phys. Rev. Letters, vol. 40, No. 7, pp. 451-454 , 1978.

Westinghouse R. & D., "Electromagnetic Spectrum Chart"., Pub. The Exploratorium, San Francisco, CA 94123, Distributed by Edmond Scientific, Barrington, N.J. 06007 Order # 609-573-6250

I am most definitely not an expert in this area. However, it is probably worth mentioning some of the main points which Don Smith appears to be making. There are some very important points being made here, and grasping these may make a considerable difference to our ability to tap into the excess energy available in our local environment. There are four points worth mentioning:

1. Voltage
2. Frequency
3. Magnetic / Electric relationship
4. Resonance

1. Voltage. We tend to view things with an 'intuitive' view, generally based on fairly simple concepts. For example, we automatically think that it is more difficult to pick up a heavy object than to pick up a light one. How much more difficult? Well, if it is twice as heavy, it would probably be about twice as much effort to pick it up. This view has developed from our experience of things which we have done in the past, rather than on any mathematical calculation or formula.

Well, how about pulsing an electronic system with a voltage? How would the output power of a system be affected by increasing the voltage? Our initial 'off-the cuff' reaction might be that the power output might be increased a bit, but then hold on… we've just remembered that Watts = Volts x Amps, so if you double the voltage, then you would double the power in watts. So we might settle for the notion that if we doubled the voltage then we could double the output power. If we thought that, then we would be wrong.

Don Smith points out that as capacitors and coils store energy, if they are involved in the circuit, then the output power is proportional to the **square** of the voltage used. Double the voltage, and the output power is four times greater. Use three times the voltage and the output power is nine times greater. Use ten times the voltage and the output power is one hundred times greater !

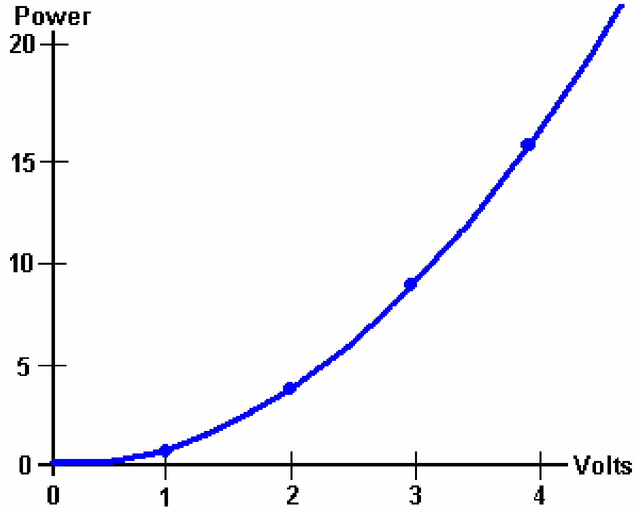

Don says that the energy stored, multiplied by the cycles per second, is the energy being pumped by the system. Capacitors and inductors (coils) temporarily store electrons, and their performance is given by:

Capacitor formula: $\mathbf{W = 0.5 \times C \times V^2 \times Hz}$ where:

W is the energy in Joules (Joules = Volts x Amps x seconds)
C is the capacitance in Farads
V is the voltage
Hz is the cycles per second

Inductor formula: $\mathbf{W = 0.5 \times L \times A^2 \times Hz}$ where:

W is the energy in Joules
L is the inductance in henrys
A is the current in amps
Hz is the frequency in cycles per second

You will notice that where inductors (coils) are involved, then the output power goes up with the square of the current. Double the voltage **and** double the current gives four times the power output due to the increased voltage and that increased output is increased by a further four times due to the increased current, giving sixteen times the output power.

2. Frequency. You will notice from the formulas above, that the output power is directly proportional to the frequency "Hz". The frequency is the number of cycles per second (or pulses per second) applied to the circuit. This is something which is not intuitive for most people. If you double the rate of pulsing, then you double the power output. When this sinks in, you suddenly see why Nikola Tesla tended to use millions of volts and millions of pulses per second.

However, Don Smith states that when a circuit is at it's point of resonance, resistance in the circuit drops to zero and the circuit becomes effectively, a superconductor. The energy for such a system which is in resonance is:

Resonant circuit: $\mathbf{W = 0.5 \times C \times V^2 \times (Hz)^2}$ where:

W is the energy in Joules
C is the capacitance in Farads
V is the voltage
Hz is the cycles per second

If this is correct, then raising the frequency in a resonating circuit has a massive effect on the power output of the device. The question then arises: why is the mains power in Europe just fifty cycles per second and in America just sixty cycles per second? If power goes up with frequency, then why not feed households at a million cycles per second? One major reason is that it is not easy to make electric motors which can be driven with power delivered at that frequency, so a more suitable frequency is chosen in order to suit the motors in vacuum cleaners, washing machines and other household equipment.

However, if we want to extract energy from the environment, then we should go for high voltage and high frequency. Then, when high power has been extracted, if we want a low frequency suited to electric motors, we can pulse the already captured power at that low frequency.

It might be speculated that if a device is being driven with sharp pulses which have a very sharply rising leading edge, that the effective frequency of the pulsing is actually determined by the speed of that rising edge, rather than the rate at which the pulses are actually generated. For example, if pulses are being generated at, say, 50 kHz but the pulses have a leading edge which would be suited to a 200 kHz pulse train, then the device might well see the signal as a 200 kHz signal with a 25% Mark/Space ratio, the very suddenness of the applied voltage having a magnetic shocking effect equivalent to a 200 kHz pulse train.

3. Magnetic / Electric relationship. Don states that the reason why our present power systems are so inefficient is because we concentrate on the electric component of electromagnetism. These systems are always COP<1 as electricity is the 'losses' of electromagnetic power. Instead, if you concentrate on the magnetic component, then there is no limit on the electric power which can be extracted from that magnetic component. Contrary to what you might expect, if you install a pick-up system which extracts electrical energy from the magnetic component, you can install any number of other identical pick-ups, each of which extract the same amount of electrical energy from the magnetic input, **without** loading the magnetic wave in any way. Unlimited electrical output for the 'cost' of creating a single magnetic effect.

The magnetic effect which we want to create is a ripple in the zero-point energy field, and ideally, we want to create that effect while using very little power. Creating a dipole with a battery which has a Plus and a Minus terminal or a magnet which has North and South poles, is an easy way to do create an electromagnetic imbalance in the local environment. Pulsing a coil is probably an even better way as the magnetic field reverses rapidly if it is an air-core coil, such as a Tesla Coil. Using a ferromagnetic core to the coil can create a problem as iron can't reverse it's magnetic alignment very rapidly, and ideally, you want pulsing which is at least a thousand times faster than iron can handle.

Don draws attention to the "Transmitter / Receiver" educational kit "Resonant Circuits #10-416" supplied by The Science Source, Maine. This kit demonstrates the generation of resonant energy and it's collection with a receiver circuit. However, if several receiver circuits are used, then the energy collected is increased several times without any increase in the transmitted energy. This is similar to a radio transmitter where hundreds of thousands of radio receivers can receive the transmitted signal without loading the transmitter in any way.

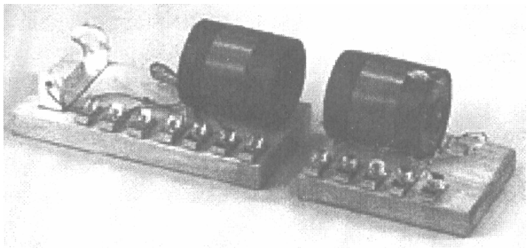

If you get the Science Source educational kit, then there are some details which you need to watch out for. The unit supplied to me had two very nice quality plastic bases and two very neatly wound coils each of 60 turns of 0.47 mm diameter enamelled copper wire on clear acrylic tubes 57 mm (2.25") in diameter. The winding covers a 28 mm section of the tube. The layout of the transmitter and receiver modules does not match the accompanying instruction sheet and so considerable care needs to be taken when wiring up any of their circuits.

The circuit diagrams are not shown, just a wiring diagram, which is not great from an educational point of view. The one relevant circuit is:

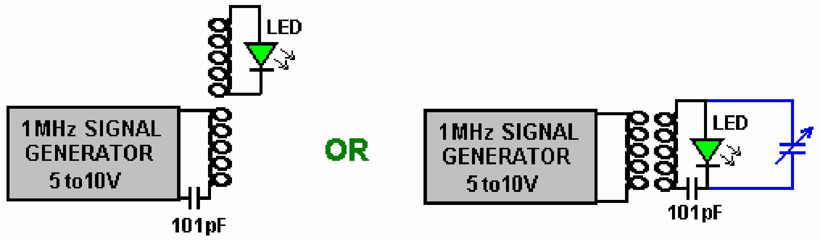

Before you buy the kit, it is not mentioned that in order to use it, you need a signal generator capable of producing a 10-volt signal at 1 MHz. The coil has a DC resistance of just 1.9 ohms but at a 1 MHz resonant frequency, the necessary drive power is quite low.

A variable capacitor is mounted on the receiver coil tube, but the one in my kit made absolutely no difference to the frequency tuning, nor was my capacitance meter able to determine any capacitance value for it at all, even though it had no trouble at all in measuring the 101 pF capacitor which was exactly the capacitance printed on it. For that reason, it is shown in blue in the circuit diagram above. Disconnecting it made no difference whatsoever.

In this particular kit, standard screw connectors have had one screw replaced with an Allen key headed bolt which has a head large enough to allow finger tightening. Unfortunately, those bolts have a square cut head where a domed head is essential if small diameter wires are to be clamped securely. If you get the kit, then I suggest that you replace the connectors with a standard electrical screw connector strip.

In tests, the LED lights when the coils are aligned and within about 100 mm of each other, or if they are close together side by side. This immediately makes the Hubbard device spring to mind. Hubbard has a central "electromagnetic transmitter" surrounded by a ring of "receivers" closely coupled magnetically to the transmitter, each of which will receive a copy of the energy sent by the transmitter:

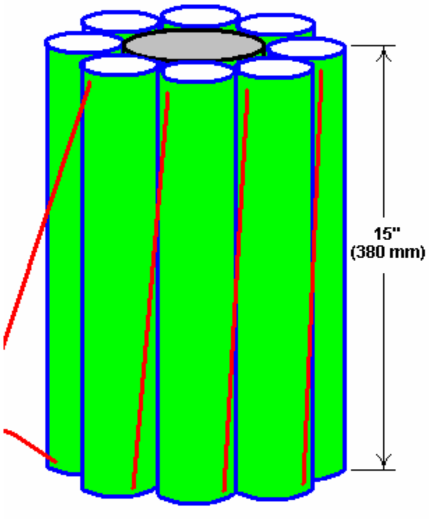

Don points to an even more clearly demonstrated occurrence of this effect in the Tesla Coil. In a typical Tesla Coil, the primary coil is much larger diameter than the inner secondary coil:

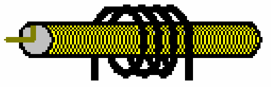

If, for example, 8,000 volts is applied to the primary coil which has four turns, then each turn would have 2,000 volts of potential. Each turn of the primary coil transfers electromagnetic flux to every single turn of the secondary winding, and the secondary coil has a very large number of turns. Massively more power is produced in the secondary coil than was used to energise the primary coil. A common mistake is to believe that a Tesla Coil can't produce serious amperage. If the primary coil is positioned in the middle of the secondary coil as shown, then the amperage generated will be as large as the voltage generated. A low power input to the primary coil can produce kilowatts of usable electrical power as described in chapter 5.

4. Resonance. An important factor in circuits aimed at tapping external energy is resonance. It can be hard to see where this comes in when it is an electronic circuit which is being considered. However, everything has it's own resonant frequency, whether it is a coil or any other electronic component. When components are connected together to form a circuit, the circuit has an overall resonant frequency. As a simple example, consider a swing:

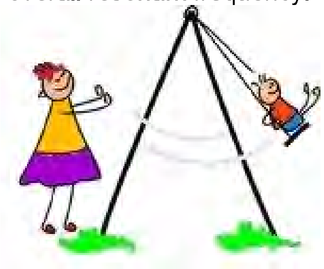

If the swing is pushed before it reaches the highest point on the mother's side, then the push actually detracts from the swinging action. The time of one full swing is the resonant frequency of the swing, and that is determined by the length of the supporting ropes holding the seat and not the weight of the child nor the power with which the child is pushed. Provided that the timing is exactly right, a very small push can get a swing moving in a substantial arc. The key factor is, matching the pulses applied to the swing, to the resonant frequency of the swing. Get it right and a large movement is produced. Get it wrong, and the swing doesn't get going at all (at which point, critics would say "see, see …swings just don't work - this proves it !!").

Establishing the exact pulsing rate needed for a resonant circuit is not particularly easy, because the circuit contains coils (which have inductance, capacitance and resistance), capacitors (which have capacitance and a small amount of resistance) and resistors and wires, both of which have resistance and some capacitance. These kinds of circuit are called "LRC" circuits because "L" is the symbol used for inductance, "R" is the symbol used for resistance and "C" is the symbol used for capacitance.

Don Smith provides instructions for winding and using the type of air-core coils needed for a Tesla Coil. He says:

1. Decide a frequency and bear in mind, the economy of the size of construction selected. The factors are:

(a) Use radio frequency (above 20 kHz).
(b) Use natural frequency, i.e. match the coil wire length to the frequency - coils have both capacitance and inductance.
(c) Make the wire length either one quarter, one half of the full wavelength.
(d) Calculate the wire length in feet as follows:
    If using one quarter wavelength, then divide 247 by the frequency in MHz.
    If using one half wavelength, then divide 494 by the frequency in MHz.
    If using the full wavelength, then divide 998 by the frequency in MHz.
  For wire lengths in metres:
    If using one quarter wavelength, then divide 75.29 by the frequency in MHz.
    If using one half wavelength, then divide 150.57 by the frequency in MHz.
    If using the full wavelength, then divide 304.19 by the frequency in MHz.

2. Choose the number of turns to be used in the coil when winding it using the wire length just calculated. The number of turns will be governed by the diameter of the tube on which the coil is to be wound. Remember that the ratio of the number of turns in the "L - 1" and "L - 2" coils, controls the overall output voltage. For example, if the voltage applied the large outer coil "L - 1" is 2,400 volts and L - 1 has ten turns, then each turn of L - 1 will have 240 volts dropped across it. This 240 volts of magnetic induction transfers 240 volts of electricity to every turn of wire in the inner "L - 2" coil. If the diameter of L - 2 is small enough to have 100 turns, then the voltage produced will be 24,000 volts. If the diameter of the L - 2 former allows 500 turns, then the output voltage will be 120,000 volts.

3. Choose the length and diameter of the coils. The larger the diameter of the coil, the fewer turns can be made with the wire length and so the coil length will be less, and the output voltage will be lower.

4. For example, if 24.7 MHz is the desired output frequency, then the length of wire, in feet, would be 247 divided by 24.7 which is 10 feet of wire (3,048 mm). The coil may be wound on a standard size of PVC pipe or alternatively, it can be purchased from a supplier - typically, an amateur radio supply store.

If the voltage on each turn of L - 1 is arranged to be 24 volts and the desired output voltage 640 volts, then there needs to be 640 / 24 = 26.66 turns on L - 2, wound with the 10 feet of wire already calculated.

Note: At this point, Don's calculations go adrift and he suggests winding 30 turns on a 2-inch former. If you do that, then it will take about 16 feet of wire and the resonant point at 10-feet will be at about 19 turns, giving an output voltage of 458 volts instead of the required 640 volts, unless the number of turns on L - 1 is reduced to give more than 24 volts per turn. However, the actual required diameter of the coil former (plus one diameter of the wire) is 10 x 12 / (26.67 x 3.14159) = 1.43 inches. You can make this size of former up quite easily if you want to stay with ten turns on the L - 1 coil.

5. Connect to the start of the coil. To determine the exact resonant point on the coil, a measurement is made. Off-the-shelf multimeters are not responsive to high-frequency signals so a cheap neon is used instead. Holding one wire of the neon in one hand and running the other neon wire along the outside of the L - 2 winding, the point of brightest light is located. Then the neon is moved along that turn to find the brightest point along that turn, and when it is located, a connection is made to the winding at that exact point. L - 2 is now a resonant winding. It is possible to increase the ("Q") effectiveness of the coil by spreading the turns out a bit instead of positioning them so that each turn touches both of the adjacent turns.

6. The input power has been suggested as 2,400 volts. This can be constructed from a Jacob's ladder arrangement or any step-up voltage system. An off-the-shelf module as used with lasers is another option.

7. Construction of the L - 1 input coil has been suggested as having 10 turns. The length of the wire in this coil is not critical. If a 2-inch diameter PVC pipe was used for the L - 2 coil, then the next larger size of PVC pipe can be used for the L - 1 coil former. Cut a 10-turn length of the pipe (probably a 3-inch diameter pipe). The pipe length will depend on the diameter of the insulated wire used to make the winding. Use a good quality multimeter or a specialised LCR meter to measure the capacitance (in Farads) and the inductance (in henrys) of the L - 2 coil. Now, put a capacitor for matching L - 1 to L - 2 across the voltage input of L - 1, and a spark gap connected **in parallel** is required for the return voltage from L - 1. A trimmer capacitor for L - 1 is desirable.

8. The performance of L - 2 can be further enhanced by attaching an earth connection to the base of the coil. The maximum output voltage will be between the ends of coil L - 2 and lesser voltages can be taken off intermediate points along the coil if that is desirable.

Don provides quite an amount of information on one of his devices shown here:

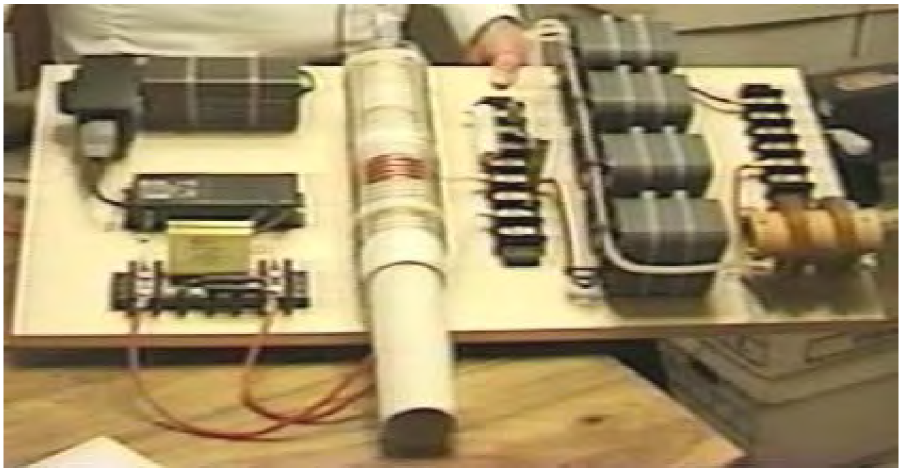

I have recently been passed a copy of Don's circuit diagram for this device, and it is shown here:

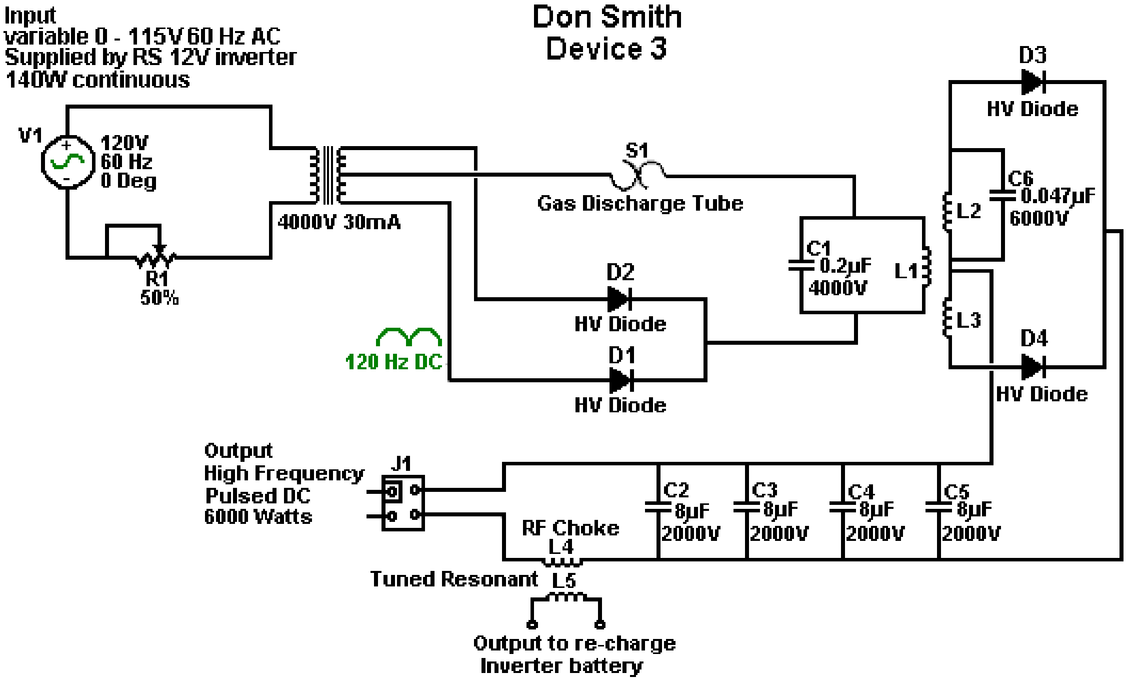
Input
variable 0 - 115V 60 Hz AC
Supplied by RS 12V inverter
140W continuous
Don Smith
Device 3
V1
+
120V
60 Hz
0 Deg
R1
50%
4000V 30mA
S1
Gas Discharge Tube
D2
HV Diode
D1
HV Diode
120 Hz DC
C1
0.2µF
4000V
L1
D3
HV Diode
L2
C6
0.047µF
6000V
L3
D4
HV Diode
Output
High Frequency
Pulsed DC
6000 Watts
J1
RF Choke
L4
C2
8µF
2000V
C3
8µF
2000V
C4
8µF
2000V
C5
8µF
2000V
Tuned Resonant L5
Output to re-charge
Inverter battery

This circuit diagram contradicts what Don says in his video presentations as the output matches the photographs which show the four output capacitors wired in parallel. It appears that while using the prototype shown above as an example of the construction, in the videos Don is actually describing a later version of the circuit. The 4000V

30mA transformer shown in this circuit diagram, is a neon-tube driver module which steps up the voltage and raises the frequency to about 35 kilohertz.

There is always a tendency to think of a spark gap as being a device which is there to protect against excessive voltages rather than seeing it as an active component of the circuit, a component which is in continuous use as it is in this circuit. Tesla, who always speaks very highly of the energy released by the very sharp discharge produced by a spark, shows a high-voltage source feeding a capacitor with the energy passing through a spark gap to the primary winding of a transformer:

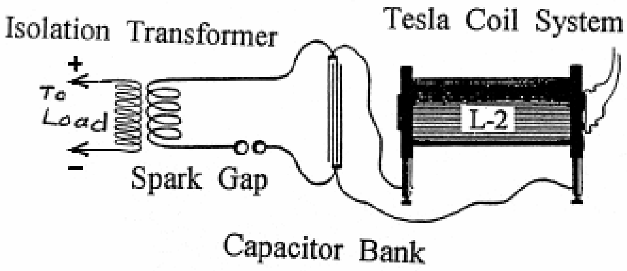

With Don's arrangement, it can be a little difficult to see why the capacitor is not short-circuited by the very low resistance of the few turns of thick wire forming the L1 coil. Well, it would do that if we were operating with DC, but we are most definitely not doing that as the output from the neon-tube driver circuit is pulsing 35,000 times per second. This causes the DC resistance of the L1 coil to be of almost no consequence and instead, the coil's AC "impedance" (effectively, it's AC resistance) is what counts. Actually, the “C1” capacitor and the L1 coil being connected across each other have a combined "impedance" or resistance to pulsing current at this frequency. This is where the nomograph diagram comes into play, and there is a much easier to understand version of it a few pages later on in this document. So, because of the high pulsing frequency, the L1 coil does not short-circuit the capacitor and if the pulsing frequency matches the resonant frequency of the L1 coil (or a harmonic of that frequency), then the L1 coil will actually have a very high resistance to current flow through it. This is how a crystal set radio receiver tunes in a particular radio station, broadcasting on it's own frequency.

Anyway, coming back to Don's device shown in the photograph above, the electrical drive is from a 12-volt battery which is not seen in the photograph. Interestingly, Don remarks that if the length of the wires connecting the battery to the inverter are exactly one quarter of the wave length of the frequency of the oscillating magnetic field generated by the circuit, then the current induced in the battery wires will recharge the battery continuously, even if the battery is supplying power to the circuit at the same time.

The battery supplies a small current through a protecting diode, to a standard off-the-shelf "true sine-wave" inverter. An inverter is a device which produces mains-voltage Alternating Current from a DC battery. As Don wants adjustable voltage, he feeds the output from the inverter into a variable transformer called a "Variac" although this is often made as part of the neon-driver circuit to allow the brightness of the neon tube to be adjusted by the user. This arrangement produces an AC output voltage which is adjustable from zero volts up to the full mains voltage (or a little higher, although Don does not want to use a higher voltage). The use of this kind of adjustment usually makes it essential for the inverter to be a true sine-wave type. As the power requirement of the neon-tube driver circuit is so low, the inverter should not cost very much and Don specifies just 140 watts of continuous current for the inverter specification.

The neon-tube (or "gas-discharge" tube) driver circuit is a standard off-the-shelf device used to drive neon tube displays for commercial establishments. The one used by Don contains an oscillator and a step-up transformer, which together produce an Alternating Current of 4,000 volts at a frequency of 35,100 Hz (sometimes written as 35.1 kHz). The term "Hz" stands for "cycles per second". It should be noted, however, that since Don bought his neon-tube driver module, the design of many of those modules has been altered to include protection circuitry which is liable to prevent the module operating properly in this circuit. It might be worth considering building your own solid-state circuit as shown later on in this chapter. If you do that, then you have full control of the frequency and the output and are not dependent on anyone else’s circuit design.

In this circuit, Don lowers the 4,000 volts as he gets great power output at lower input voltages and the cost of the output capacitors is a significant factor. The particular neon-tube driver circuit which Don is using here, has two separate outputs, so Don connects them together and uses a blocking diode in each line to prevent either of them affecting the other one. Not easily seen in the photograph, the high-voltage output line has a very small, encapsulated, spark gap in it and the line is also earthed. This device is commonly used as a lightning strike protection component but it is an active component in Don's circuit and it lights continuously when the device is running. It looks like this:

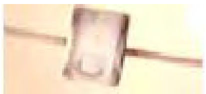

Please note that when an earth connection is mentioned in connection with Don Smith's devices, we are talking about an actual wire connection to a metal object physically buried in the ground, whether it is a long copper rod driven into the ground, or an old car radiator buried in a hold like Tariel Kapanadze uses, or a buried metal plate. When Thomas Henry Moray performed his requested demonstration deep in the countryside at a location chosen by the sceptics, the light bulbs which formed his demonstration electrical load, glowed more brightly with each hammer stroke as a length of gas pipe was hammered into the ground to form his earth connection.

The output of the neon-tube driver circuit is used to drive the primary "L1" winding of a Tesla Coil style transformer. This looks ever so simple and straightforward, but there are some subtle details which need to have attention paid to them.

The operating frequency of 35.1 kHz is set and maintained by the neon-tube driver circuitry, and so, in theory, we do not have to do any direct tuning ourselves. However, we want the resonant frequency of the L1 coil and the capacitor across it to match the neon-driver circuit frequency. The frequency of the "L1" coil winding will induce exactly the same frequency in the "L2" secondary winding with the “C6” tuning capacitor across it. However, we need to pay special attention to the ratio of the wire lengths of the two coil windings as we want these two windings to resonate together.

A rule of thumb followed by most Tesla Coil builders is to have the same weight of copper in the L1 and L2 coils, which means that the wire of the L1 coil is usually much thicker than the wire of the L2 coil. If the L1 coil is to be one quarter of the length of the L2 coil, then we would expect the cross-sectional area of the L1 coil to be four times that of the wire of the L2 coil (as the area is proportional to the square of the radius, and the square of two is four, so doubling the wire diameter gives four times the cross-sectional area of the copper wire).

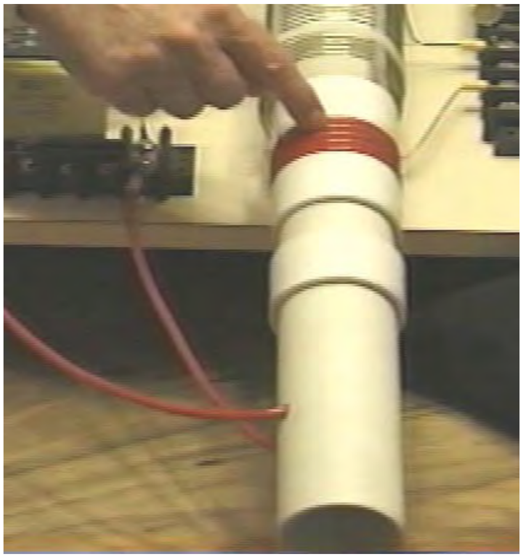

Don uses a plastic tube as the former for his "L1" primary coil winding. As you can see here, the wire is fed into the former, leaving sufficient clearance to allow the former to slide all the way into the outer coil. The wire is fed up inside the pipe and out through another hole to allow the coil turns to be made on the outside of the pipe. There appear to be five turns, but Don does not always go for a complete number of turns, so it might be 4.3 turns or some other value. The key point here is that the length of wire in the "L1" coil turns should be exactly one quarter of the length of wire in the "L2" coil turns.

The "L2" coil used here is a commercial 3-inch diameter unit from Barker & Williamson, constructed from uninsulated, solid, single-strand "tinned" copper wire. Don has taken this coil and unwound four turns at the centre of the coil in order to make a centre-tap. He then measured the exact length of wire in the remaining section and made the length of the "L1" coil turns to be exactly one quarter of that length. The wire used for the "L1" coil looks like Don's favourite "Jumbo Speaker Wire" which is a very flexible wire with a very large number of extremely fine uninsulated copper wires inside it.

You will notice that Don has placed a plastic collar on each side of the winding, matching the thickness of the wire, in order to create a secure sliding operation inside the outer "L2" coil, and the additional plastic collars positioned further along the pipe provide further support for the inner coil. This sliding action allows the primary coil "L1" to be positioned at any point along the length of the "L2" secondary coil, and that has a marked tuning effect on the operation of the system. The outer "L2" coil does not have any kind of tube support but instead, the coil shape is maintained by the stiffness of the solid wire plus four slotted plastic strips. This style of construction produces a very high coil performance at radio frequencies. With a Tesla Coil, it is most unusual to have the L1 coil of smaller diameter than the L2 coil.

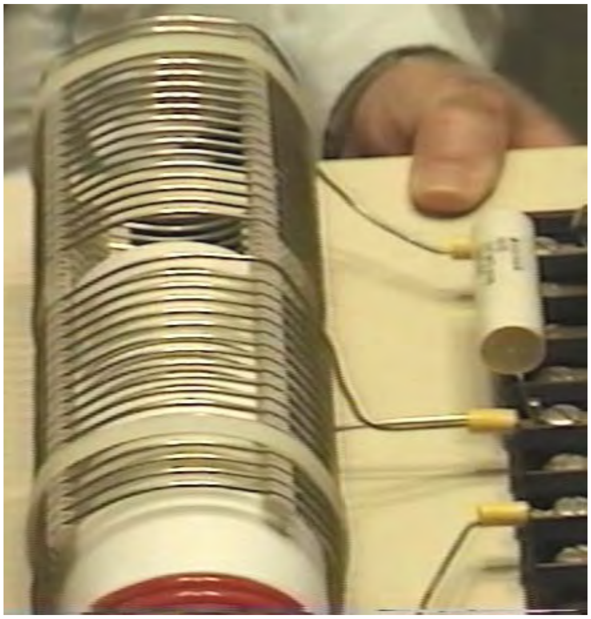

The "L2" coil has two separate sections, each of seventeen turns. One point to note is the turns are spaced apart using slotted strips to support the wires and maintain an accurate spacing between adjacent turns. It must be remembered that spacing coil turns apart like this alters the characteristics of the coil, increasing it's "capacitance" factor substantially. Every coil has resistance, inductance and capacitance, but the form of the coil construction has a major effect on the ratio of these three characteristics. The coil assembly is held in position on the base board by two off-white plastic cable ties.

One point which Don stresses, is that the length of the wire in the "L1" coil and the length of wire in the "L2" coil, must be an exact even division or multiple of each other (in this case, the "L2" wire length in each half of the "L2" coil is exactly four times as long as the "L1" coil wire length). This is likely to cause the "L1" coil to have part of a turn, due to the different coil diameters. For example, if the length of the "L2" coil wire is 160 inches and "L1" is to be one quarter of that length, namely, 40 inches. Then, if the "L1" coil has an effective diameter of 2.25 inches, (allowing for the thickness of the wire when wound on a 2-inch diameter former), then the "L1" coil would have 5.65 (or 5 and 2/3) turns which causes the finishing turn of "L2" to be 240 degrees further around the coil former than the start of the first turn - that is, five full turns plus two thirds of the sixth turn.

The L1 / L2 coil arrangement is a Tesla Coil. The positioning of the "L1" coil along the length of the "L2" coil, adjusts the voltage to current ratio produced by the coil. When the "L1" coil is near the middle of the "L2" coil, then the amplified voltage and amplified current are roughly the same. However, Don stresses that the "height" length of the coil (when standing vertically) controls the voltage produced while the coil "width" (the diameter of the turns) controls the current produced.

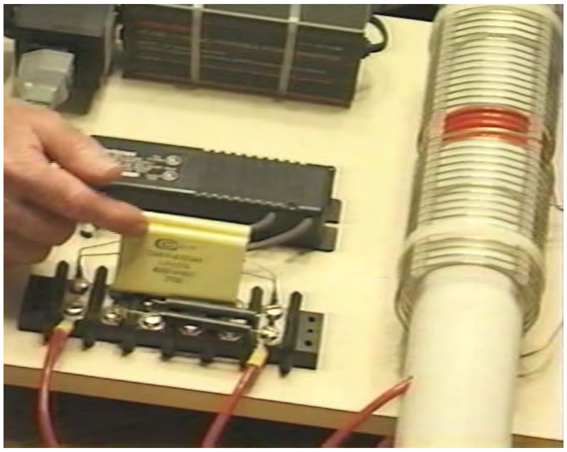

The exact wire length ratio of the turns in the "L1" and "L2" coils gives them an almost automatic synchronous tuning with each other, and the exact resonance between them can be achieved by the positioning of the "L1" coil along the length of the "L2" coil.

Don has also connected a small capacitor across the "L2" coil, and that optional component is marked as "C2" in the circuit diagram and the value used by Don happened to be a single 47nF, high-voltage capacitor. As the two halves of the "L2" coil are effectively connected across each other, it is only necessary to have one capacitor for "L2":

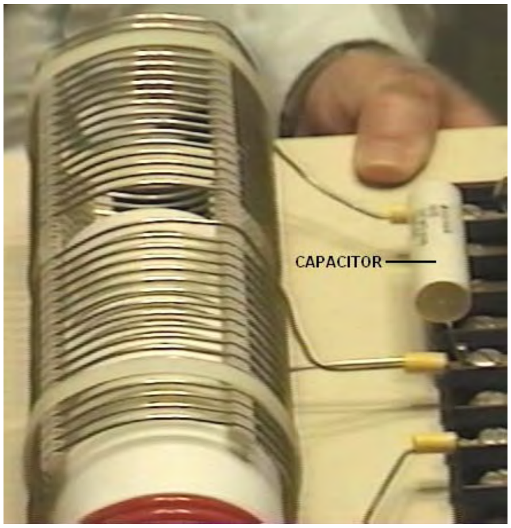

There are various ways of dealing with the output from the "L2" coil in order to get large amounts of conventional electrical power out of the device. The method shown here uses the four very large capacitors seen in the photograph. Don remarks that he has to be very careful to keep the voltage to the neon-tube driver circuit turned down in order to avoid getting more than 2,000 volts on these output storage capacitors.

This capacitor bank is fed through two diodes which are rated for high voltage operation. In the build shown in the photographs and the circuit diagram, the output is rated at 2,000 volts and 6,000 watts, in which case, the output diodes have to be able to handle 3 amps continuously, but at start-up when the capacitor bank is fully discharged that current may well be exceeded while the capacitor bank charges up.

When the circuit is running, the storage capacitor bank behaves like a 2,000 volt battery which never runs down and which can supply 3 amps of current for as long as you want. The circuitry for producing a 220 volt 50 Hz AC output or a 110 volt 60 Hz AC output from the storage capacitors is just standard electronics.

The output current flows through the left hand winding on the brown cylindrical former, and that induces a current in the right-hand winding. This additional current is used to provide charging power for the battery used to drive the inverter. A possible circuit arrangement for doing this might be as shown here:

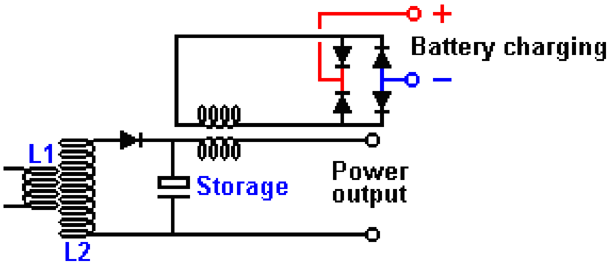

The DC output produced by the four diodes can then be used to charge the driving battery, and the power level produced is substantially greater than the minor current drain from the battery. Consequently, it is a sensible precaution to pass this current to the battery via a circuit which prevents the battery voltage rising higher than it should. A simple voltage level sensor can be used to switch off the charging when the battery has reached its optimum level. Other batteries can also be charged if that is wanted. Simple circuitry of the type shown in chapter 12 can be used for controlling and limiting the charging process. The components on Don's board are laid out like this:

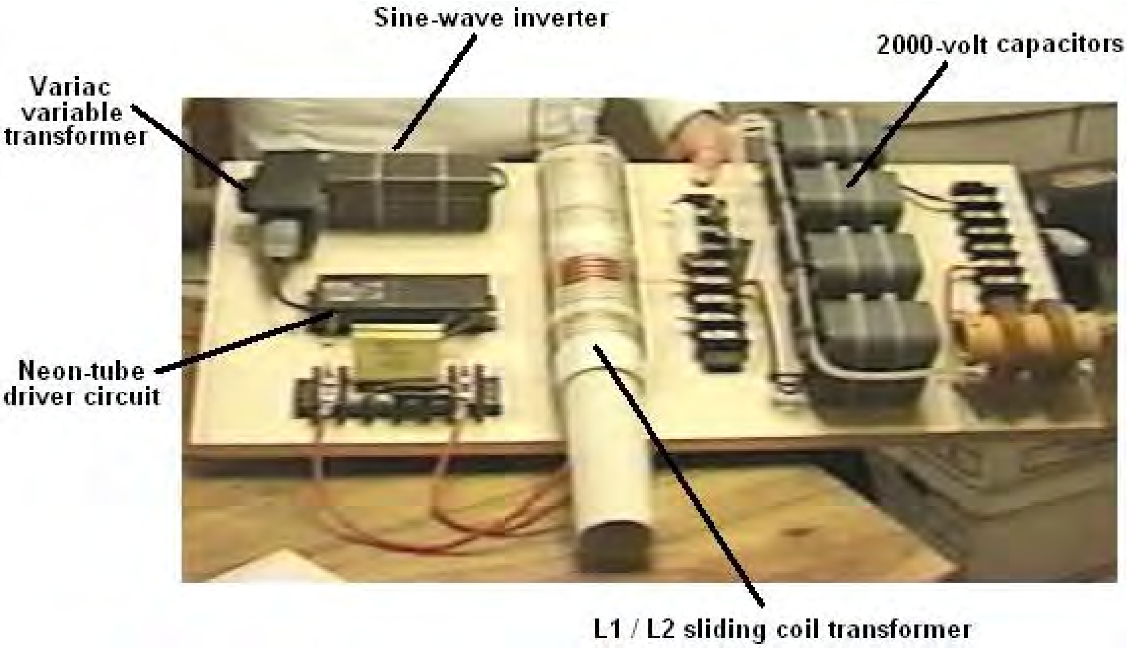

Don draws attention to the fact that the cables used to connect the output of "L2" to the output of the board, connecting the storage capacitors on the way, are very high-voltage rated cables with special multiple coverings to ensure that the cables will remain sound over an indefinite period. It should be remarked at this point, that the outer 3" diameter coil used by Don, is not wound on a former, but in order to get higher performance at high frequencies, the turns are supported with four separate strips physically attached to the turns - the technique described later in this document as being an excellent way for home construction of such coils.

Please bear in mind that the voltages here and their associated power levels are literally lethal and perfectly capable of killing anyone who handles the device carelessly when it is powered up. When a replication of this device is ready for routine use, it must be encased so that none of the high-voltage connections can be touched by anyone. This is not a suggestion, but it is a mandatory requirement, despite the fact that the components shown in the photographs are laid out in what would be a most dangerous fashion were the circuit to be powered up as it stands. Under no circumstances, construct and test this circuit unless you are already experienced in the use of high-voltage circuits or can be supervised by somebody who is experienced in this field. This is a "one hand in the pocket at all times" type of circuit and it needs to be treated with great care and respect at all times, so be sensible.

It may also be necessary to use an earthed screen around the device in order to prevent radio frequency emissions. The remainder of the circuit is not mounted on the board, possibly because there are various ways in which the required end result can be achieved. The one suggested here is perhaps the most simple solution:

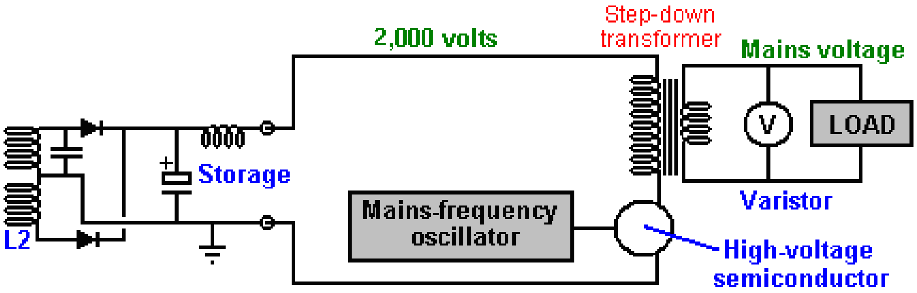

The voltage has to be dropped, so an iron-cored mains-frequency step-down transformer is used to do this. To get the frequency to the standard mains frequency for the country in which the device is to be used, an oscillator is used to generate that particular mains frequency. The oscillator output is used to drive a suitable high-voltage semiconductor device, be it an FET transistor, an IGBT device, or whatever. This device has to switch the working current at 2,000 volts, though admittedly, that will be a current which will be very much lower than the final output current, due to the higher voltage on the primary winding of the transformer. When using this circuit, the available power will be limited by the current handling capabilities of this output transformer, so a custom-wound unit with heavy-duty wire and a substantial frame will be needed to provide the full six kilowatts of power.

As the circuit is capable of picking up additional magnetic pulses, such as those generated by other equipment, nearby lightning strikes, etc. an electronic component called a "varistor" marked "V" in the diagram, is connected across the load. This device acts as a voltage spike suppressor as it short circuits any voltage above its design voltage, protecting the load from power surges.

Don also explains an even more simple version which does not need a Variac, high voltage capacitors or high voltage diodes. Here, a DC output is accepted which means that high-frequency step-down transformer operation can be used. This calls for an air-core transformer which you would wind yourself from heavy duty wire. Mains loads would then be powered by using a standard off-the-shelf inverter. In this version, it is of course, necessary to make the "L1" turns wire length exactly one quarter of the "L2" turns wire length in order to make the two coils resonate together. The operating frequency of each of these coils is imposed on them by the output frequency of the neon-tube driver circuit. That frequency is maintained throughout the entire circuit until it is rectified by the four diodes feeding the low-voltage storage capacitor. The target output voltage will be either just over 12 volts or just over 24 volts, depending on the voltage rating of the inverter which is to be driven by the system. This is effectively two Tesla Coils back-to-back and the circuit diagram might be:

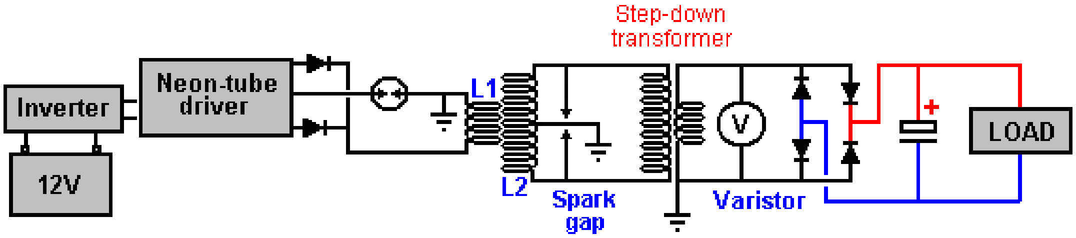

As many people will find the nomograph chart in Don's pdf document very difficult to understand and use, here is an easier version:

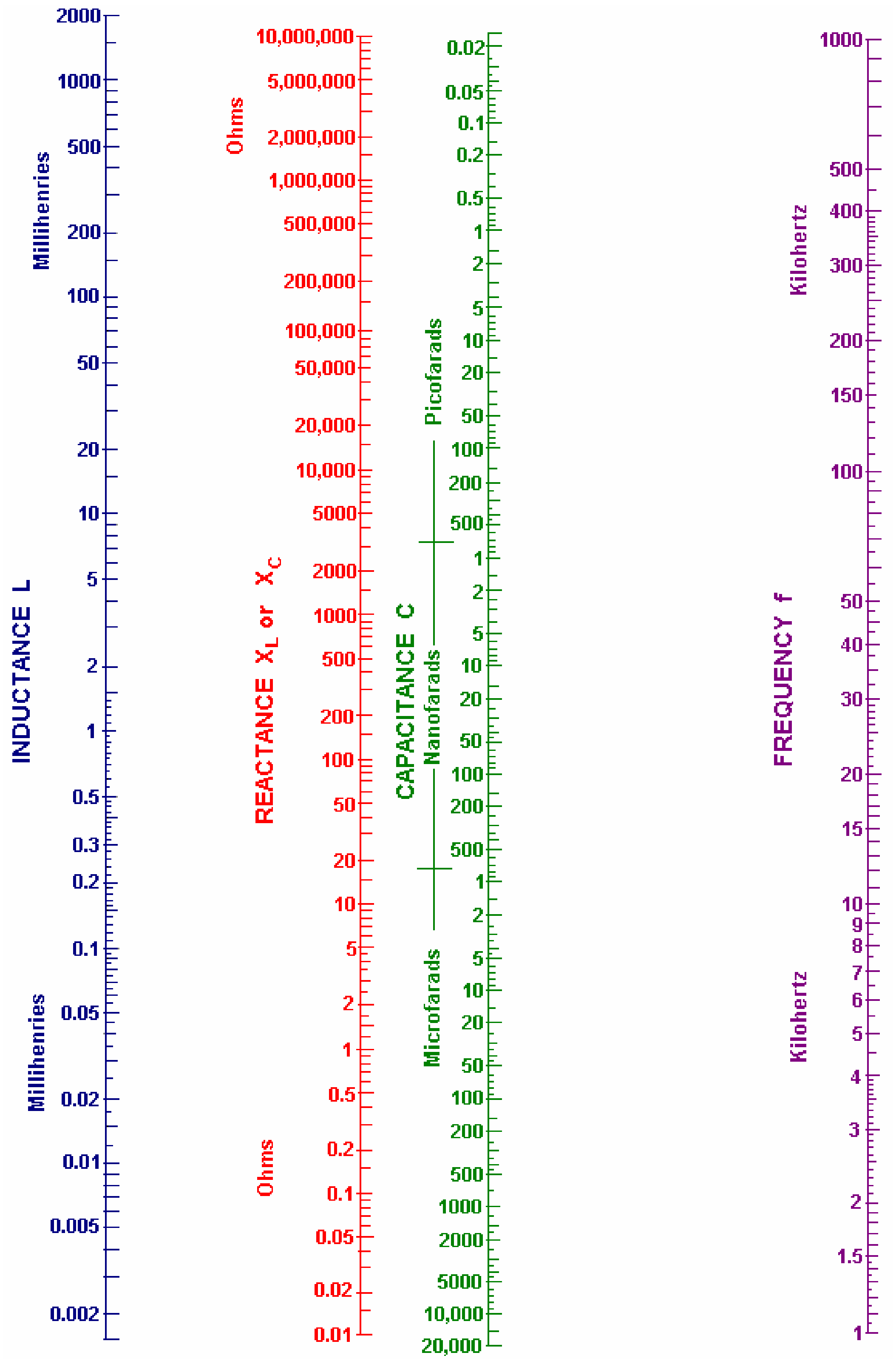

The objective here is to determine the "reactance" or 'AC resistance' in ohms and the way to do that is as follows:

Suppose that your neon-tube driver is running at 30 kHz and you are using a capacitor of 100 nF (which is the same as 0.1 microfarad) and you want to know what is the AC resistance of your capacitor is at that frequency. Also, what coil inductance would have that same AC resistance. Then the procedure for finding that out is as follows:

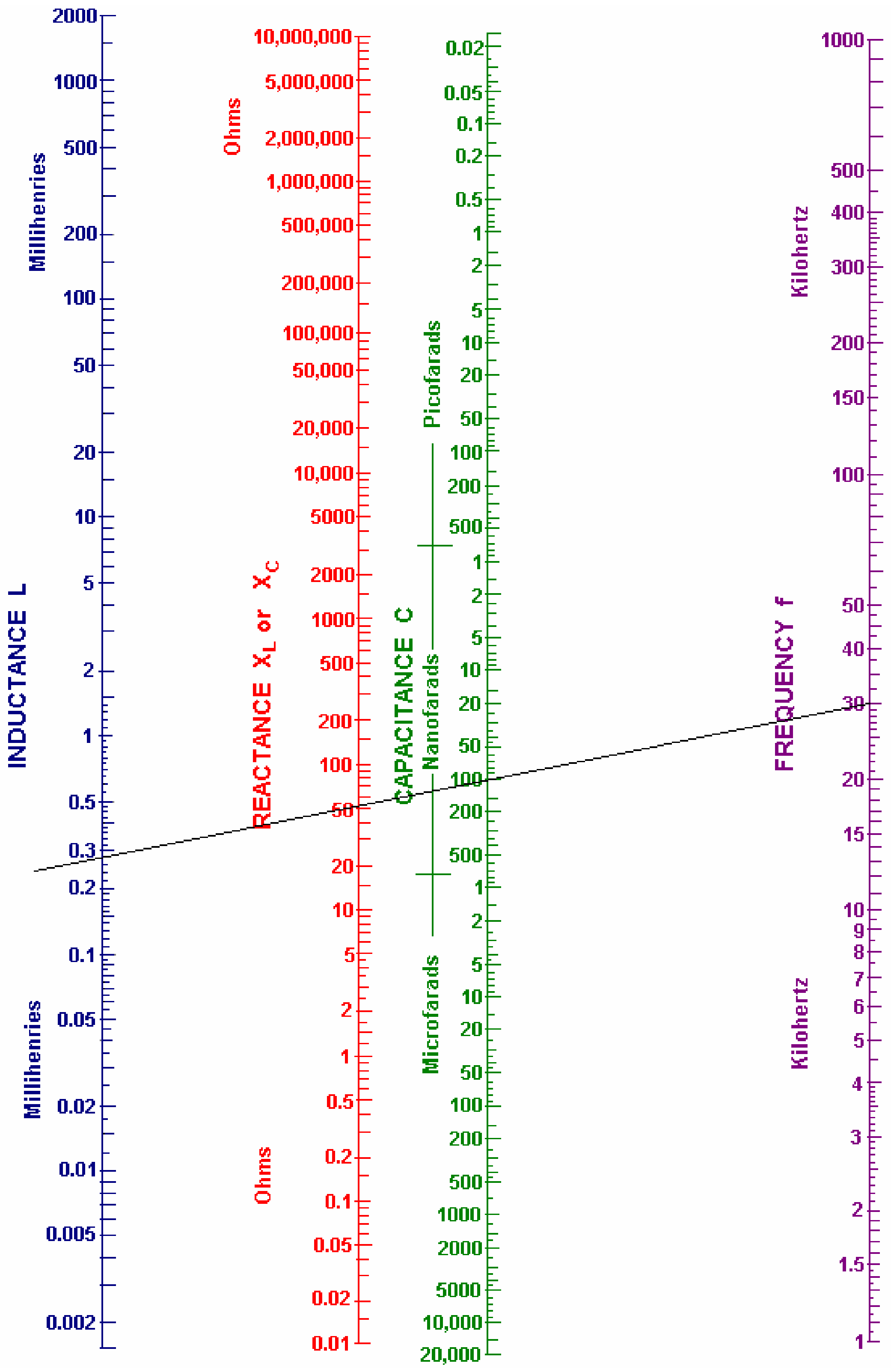

Draw a straight line from your 30 kHz frequency (purple line) through your 100 nanofarad capacitor value and carry the line on as far as the (blue) inductance line as shown above.

You can now read the reactance off the red line, which looks like 51 ohms to me. This means that when the circuit is running at a frequency of 30 kHz, then the current flow through your 100 nF capacitor will be the same as through a 51 ohm resistor. Reading off the blue "Inductance" line that same current flow at that frequency would occur with a coil which has an inductance of 0.28 millihenries.

Another device of Don's is particularly attractive in that almost no home-construction is needed, all of the components being available commercially, and the output power being adaptable to any level which you want. Don particularly likes this circuit because it demonstrates COP>1 so neatly and he remarks that the central transmitter Tesla Coil on its own is sufficient to power a household.

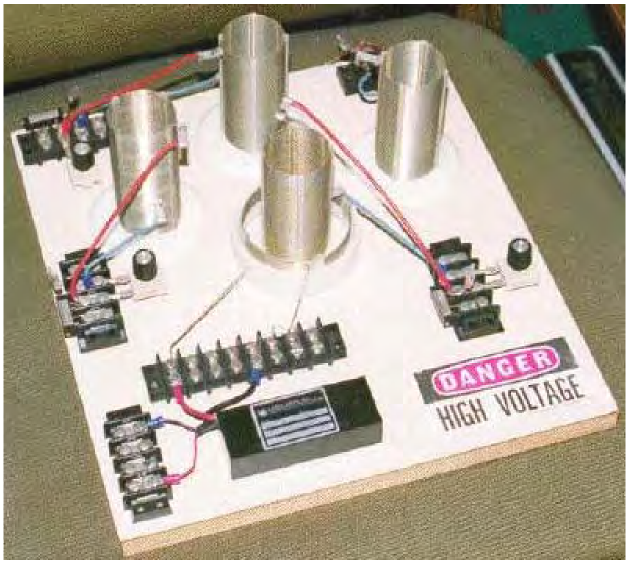

The coil in the centre of the board is a power transmitter made from a Tesla Coil constructed from two Barker & Williamson ready-made coils. Three more of the inner coil are also used as power receivers. The outer, larger diameter coil is a few turns taken from one of their standard coils and organised so that the coil wire length is one quarter of the coil wire length of the inner coil ("L2").

As before, a commercial neon-tube driver module is used to power the "L1" outer coil with high voltage and high frequency. It should be understood that as power is drawn from the local environment each time the power driving the transmitter coil "L1" cycles, that the power available is very much higher at higher frequencies. The power at mains frequency of less than 100 Hz is far, far less than the power available at 35,000 Hz, so if faced with the choice of buying a 25 kHz neon-tube driver module or a 35 kHz module, then the 35 kHz module is likely to give a much better output power at every voltage level.

HIGH VOLTAGE

The "L1" short outer coil is held in a raised position by the section of white plastic pipe in order to position it correctly relative to the smaller diameter "L2" secondary coil. Again, it appears to have five turns:

HIGH VOLTAGE
DANGER

The secondary coils are constructed using Barker & Williamson's normal method of using slotted strips to hold the tinned, solid copper wire turns in place.

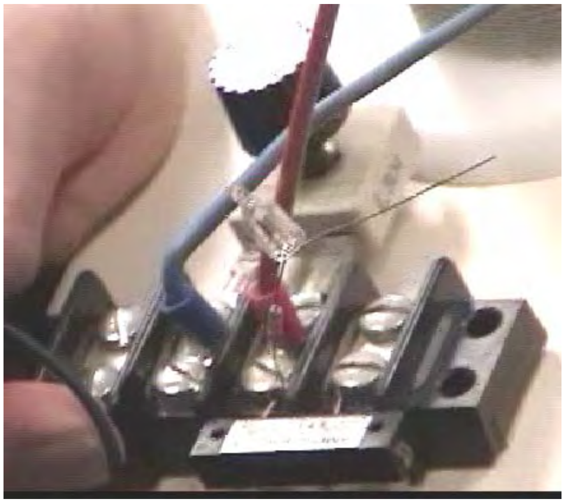

As there are very slight differences in the manufactured coils, each one is tuned to the exact transmitter frequency and a miniature neon is used to show when the tuning has been set correctly.

The key feature of this device is the fact that any number of receiver coils can be placed near the transmitter and each will receive a full electrical pick up from the local environment, without altering the power needed to drive the Tesla Coil transmitter - more and more output without increasing the input power - unlimited COP values, all of which are over 1. The extra power is flowing in from the local environment where there is almost unlimited amounts of excess energy and that inflow is caused by the rapidly vibrating magnetic field generated by the central Tesla Coil. While the additional coils appear to just be scattered around the base board, this is not the case. The YouTube video http://www.youtube.com/watch?v=TiNEHZRm4z4&feature=related demonstrates that the pick-up of these coils is affected to a major degree by the distance from the radiating magnetic field. This is to do with the wavelength of the signal driving the Tesla Coil, so the coils shown above are all positioned at exactly the same distance from the Tesla Coil. You still can have as many pick-up coils as you want, but they will be mounted in rings around the Tesla Coil and the coils in each ring will be at the same distance from the Tesla Coil in the centre.

Each of the pick up coils act exactly the same as the "L2" secondary coil of the Tesla Coil transmitter, each picking up the same level of power. Just as with the actual "L2" coil, each will need an output circuit arrangement as described for the previous device. Presumably, the coil outputs could be connected in parallel to increase the output amperage, as they are all resonating at the same frequency and in phase with each other. Each will have its own separate output circuit with a step-down isolation transformer and frequency adjustment as before. If any output is to be a rectified DC output, then no frequency adjustment is needed, just rectifier diodes and a smoothing capacitor following the step-down transformer which will need to be an air core or ferrite core type due to the high frequency. High voltage capacitors are very expensive. The http://www.richieburnett.co.uk/parts.html web site shows various ways of making your own high-voltage capacitors and the advantages and disadvantages of each type.

There are two practical points which need to be mentioned. Firstly, as the Don Smith devices shown above feed radio frequency waveforms to coils which transmit those signals, it may be necessary to enclose the device in an earthed metal container in order not to transmit illegal radio signals. Secondly, as it can be difficult to obtain high-voltage high-current diodes, they can be constructed from several lower power diodes. To increase the voltage rating, diodes can be wired in a chain. Suitable diodes are available as repair items for microwave ovens. These typically have about 4,000 volt ratings and can carry a good level of current. As there will be minor manufacturing differences in the diodes, it is good practice to connect a high value resistor (in the 1 to 10 megohm range) across each diode as that ensures that there is a roughly equal voltage drop across each of the diodes:

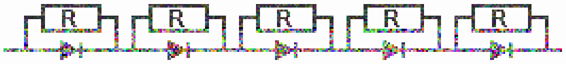

If the diode rating of these diodes were 4 amps at 4,000 volts, then the chain of five could handle 4 amps at 20,000 volts. The current capacity can be increased by connecting two or more chains in parallel.

Various questions from readers indicate that the operation of AC circuits is not really understood, so electronics experts can skip this next section.

AC Circuits. This is a lightweight introduction to Alternating Current circuits and pulsed DC circuits for people who have not read Chapter 12 which is an electronics tutorial. Let me say again, that I am mainly self-taught, and so this is just a general introduction based on my present understanding.

Alternating Current, generally called “AC” is called that because the voltage of this type of power supply is not a constant value. A car battery, for instance, is DC and has a fairly constant voltage usually about 12.8 volts when in it’s fully charged state. If you connect a voltmeter across a car battery and watch it, the voltage reading will not change. Minute after minute it says exactly the same because it is a DC source.

If you connect an AC voltmeter across an AC power supply, it too will give a steady reading, but it is telling a lie. The voltage is changing all the time in spite of that steady meter reading. What the meter is doing is **assuming** that the AC waveform is a sine wave like this:

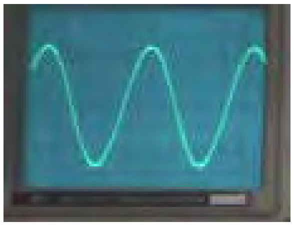

and based on that assumption, it displays a voltage reading which is called the “Root Mean Square” or “RMS” value. The main difficulty with a sine wave is that the voltage is below zero volts for exactly the same length of time as it is above zero volts, so if you average it, the result is zero volts, which is not a satisfactory result because you can get a shock from it and so it can’t be zero volts, no matter what the arithmetical average is.

To get over this problem, the voltage is measured thousands of times per second and the results squared (that is, the value is multiplied by itself) and then those values are averaged. This has the advantage that when the voltage is say, minus 10 volts and you square it, the answer is plus 100 volts. In fact, all of the answers will be positive, which means that you can add them together, average them and get a sensible result. However, you end up with a value which is far too high because you squared every measurement, and so you need to take the square root of that average (or "mean") value, and that is where the fancy sounding "Root Mean Square" name comes from – you are taking the (square) root of the (average or) mean value of the squared measurements.

With a sine wave like this, the voltage peaks are 41.4% higher than the RMS value which everyone talks about. This means that if you feed 100 volts AC through a rectifier bridge of four diodes and feed it into a capacitor the capacitor voltage will **not** be 100 volts DC but instead it will be 141.4 volts DC and you need to remember that when choosing the voltage rating of the capacitor. In that instance I would suggest a capacitor which is made to operate with voltages up to 200 volts.

You probably already knew all of that, but it may not have occurred to you that if you use a standard AC voltmeter on a waveform which is **not** a sine wave, that the reading on the meter is most unlikely to be correct or anywhere near correct. So, please don't merrily connect an AC voltmeter across a circuit which is producing sharp voltage spikes like, for instance, one of John Bedini's battery pulsing circuits, and think that the meter reading means anything (other than meaning that you don't understand what you are doing).

You will, hopefully, have learned that power in watts is determined by multiplying the current in amps by the voltage in volts. For example, 10 amps of current flowing out of a 12 volt power supply, represents 120 watts of power. Unfortunately, that only holds true for circuits which are operating on DC, or AC circuits which have only resistors in them. The situation changes for AC circuits which have non-resistive components in them.

The circuits of this type which you are likely to come across are circuits which have coils in them, and you need to think about what you are doing when you deal with these types of circuit. For example, consider this circuit:

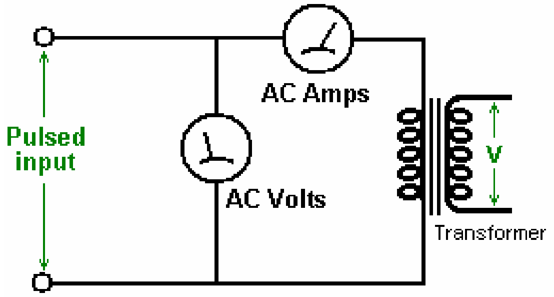

This is the output section of a prototype which you have just built. The input to the prototype is DC and measures at 12 volts, 2 amps (which is 24 watts). Your AC voltmeter on the output reads 15 volts and your AC ammeter reads 2.5 amps and you are delighted because 15 x 2.5 = 37.5 which looks much bigger than the 24 watts of input power. **But**, just before you go rushing off to announce on YouTube that you have made a prototype with COP = 1.56 or 156% efficient, you need to consider the real facts.

This is an AC circuit and unless your prototype is producing a perfect sine wave, then the AC voltmeter reading will be meaningless. It is just possible that your AC ammeter is one of the few types that can accurately measure the current no matter what sort of waveform is fed to it, but it is distinctly possible that it will be a digital meter which assesses current by measuring the AC voltage across a resistor in series with the output, and if that is the case, it will probably be assuming a sine wave. The odds are that both readings are wrong, but let's take the case where we have great meters which are reading the values perfectly correctly. Then the output will be 37.5 watts, won't it? Well, actually, no it won't. The reason for this is that the circuit is feeding the transformer winding which is a coil and coils don't work like that.

The problem is that, unlike a resistor, when you apply a voltage across a coil the coil starts absorbing energy and feeding it into the magnetic field around the coil, so there is a delay before the current reaches it's maximum value. With DC, this generally doesn't matter very much, but with AC where the voltage is changing continuously, it matters a great deal. The situation can be as shown in this graph of both voltage and current:

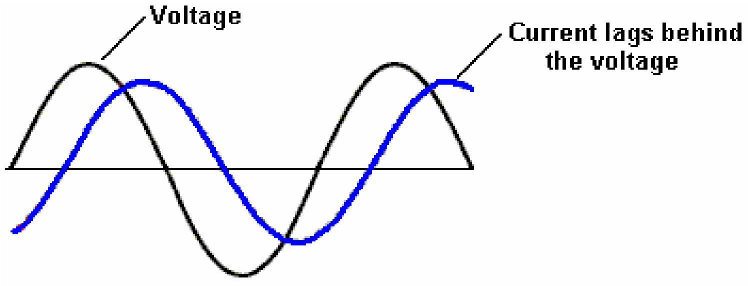
Voltage
Current lags behind the voltage

At first, this does not look like any great problem, but it has a very significant effect on the actual power in watts. To get the 37.5 watts output which we were talking about earlier, we multiplied the average voltage level by the average current level. But these two values do not occur at the same time and that has a major effect.

As this can be a little difficult to see, let's take the peak values rather than the averages as they are easier to see. Let's say that in our example graph that the voltage peak is 10 volts and the current peak is 3 amps. If this were DC we would multiply them together and say that the power was 30 watts. But with AC, this does not work due to the timing difference:

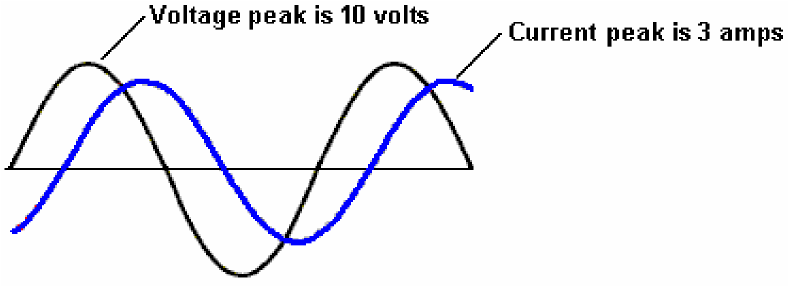
Voltage peak is 10 volts
Current peak is 3 amps

When the voltage is peaking, the current is nowhere near it's peak value of 3 amps:

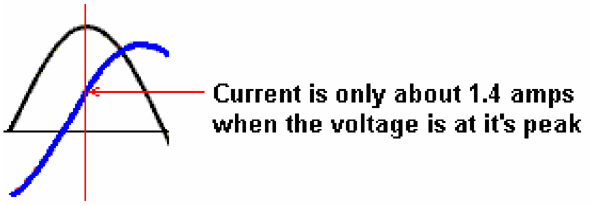
Current is only about 1.4 amps when the voltage is at it's peak

As a result of this, instead of getting our expected peak power at the top of the voltage peak, the actual power in watts is very much lower – less than half of what we were expecting. Not so good, but it gets worse when you look at the situation more closely. Take a look at what the voltage is when the current crosses the zero line, that is, when the current is zero. The output power is zero when the current is zero but this occurs when the voltage is at a very high value:

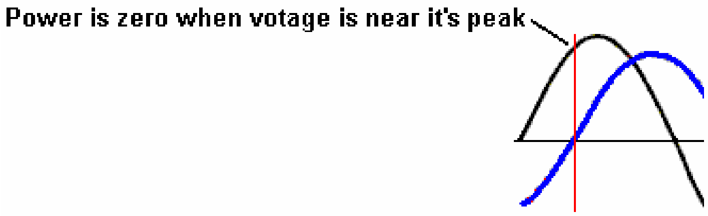
Power is zero when votage is near it's peak

The same goes for when the voltage is zero. When the voltage is zero, then the power is also zero, and you will notice that this occurs when the current is at a high value:

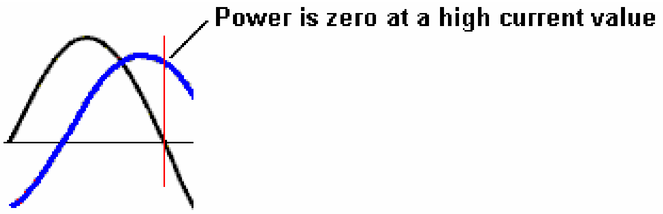

The power is **not** the average current multiplied by the average voltage if there is a coil involved in the circuit – it will be less than that by an amount known as the "power factor" and I'll leave you to work out why it is called that.

So, how do you determine what the power is? It is done by sampling the voltage and current many times per second and averaging those combined results:

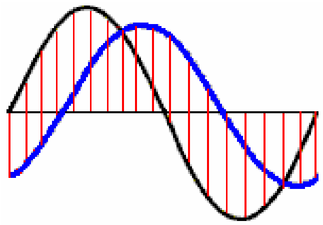

Both the voltage and the current are sampled at the times indicated by the vertical red lines and those figures are used to calculate the actual power level. In this example, only a few samplings are shown, but in practice, a very large number of samples will be taken. The piece of equipment which does this is known as a wattmeter as it measures watts of power. The sampling can be done by windings inside the instrument, resulting in an instrument which can be damaged by overloading without the needle being anywhere near full deflection, or it can be done by digital sampling and mathematical integration. Most digital sampling versions of these meters only operate at high frequencies, typically over 400,000 cycles per second. Both varieties of wattmeter can handle any waveform and not just sine waves.

The power company supplying your home, measures the current and assumes that the full voltage is present during all of the time that the current is being drawn. If you are powering a powerful electric motor from the mains, then this current lag will cost you money as the power company does not take it into account. It is possible to correct the situation by connecting one or more suitable capacitors across the motor to minimise the power loss.

With a coil (fancy name "inductor" symbol "L"), AC operation is very different to DC operation. The coil has a DC resistance which can be measured with the ohms range of a multimeter, but that resistance does not apply when AC is being used as the AC current flow is **not** determined by the DC resistance of the coil. Because of this, a second term has to be used for the current-controlling factor of the coil, and the term chosen is "impedance" or for people who like to make everything sound unduly complicated "reactance". I will stick with the term "impedance" as it is clear that it is the feature of the coil which "impedes" AC current flow through the coil.

The impedance of a coil depends on it's size, shape, method of winding, number of turns and core material. It also depends on the frequency of the AC voltage being applied to it. If the core is made up of iron or steel, usually thin layers of iron which are insulated from each other, then it can only handle low frequencies. You can forget about trying to pass 10,000 cycles per second ("Hz") through the coil as the core just can't change it's magnetic poles fast enough to cope with that frequency. A core of that type is ok for the very low 50 Hz or 60 Hz frequencies used for mains power, which are kept that low so that electric motors can use it.

For higher frequencies, ferrite can be used for a core and that is why some portable radios use ferrite-rod aerials, which are a bar of ferrite with a coil wound on it. For higher frequencies (or higher efficiencies) iron dust encapsulated in epoxy resin is used. An alternative is to not use any core material and that is usually referred to as an "air-core" coil. These are not limited in frequency by the core but they have a very much lower inductance for any given number of turns. The efficiency of the coil is called it's "Q" (for "Quality") and the higher the Q factor, the better. The resistance of the wire lowers the Q factor.

A coil has inductance, and resistance caused by the wire, and capacitance caused by the turns being near each other. However, having said that, the inductance is normally so much bigger than the other two components that we tend to ignore the other two. Something which may not be immediately obvious is that the impedance to AC current flow through the coil depends on how fast the voltage is changing. If the AC voltage applied to a coil completes one cycle every ten seconds, then the impedance will be much lower than if the voltage cycles a million times per second.

If you had to guess, you would think that the impedance would increase steadily as the AC frequency increased. In other words, a straight-line graph type of change. That is not the case. Due to a feature called resonance, there is one particular frequency at which the impedance of the coil increases massively. This is used in the tuning method for AM radio receivers. In the very early days when electronic components were hard to come by, variable coils were sometimes used for tuning. We still have variable coils today, generally for handling large currents rather than radio signals, and we call them "rheostats" and some look like this:

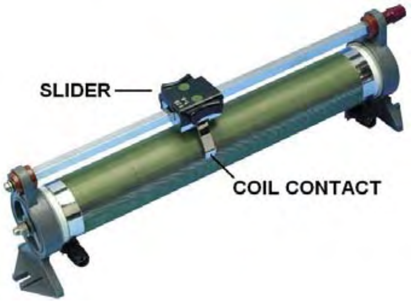

These have a coil of wire wound around a hollow former and a slider can be pushed along a bar, connecting the slider to different winds in the coil depending on it's position along the supporting bar. The terminal connections are then made to the slider and to one end of the coil. The position of the slider effectively changes the number of turns of wire in the part of the coil which is being used in the circuit. Changing the number of turns in the coil, changes the resonant frequency of that coil. AC current finds it very, very hard to get through a coil which has the same resonant frequency as the AC current frequency. Because of this, it can be used as a radio signal tuner:

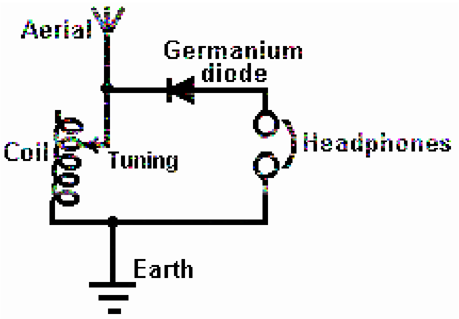

If the coil's resonant frequency is changed to match that of a local radio station by sliding the contact along the coil, then that particular AC signal frequency from the radio transmitter finds it almost impossible to get through the coil and so it (and only it) diverts through the diode and headphones as it flows from the aerial wire to the earth wire and the radio station is heard in the headphones. If there are other radio signals coming down the aerial wire, then, because they are not at the resonant frequency of the coil, they flow freely through the coil and don't go through the headphones.

This system was soon changed when variable capacitors became available as they are cheaper to make and they are more compact. So, instead of using a variable coil for tuning the radio signal, a variable capacitor connected across the tuning coil did the same job:

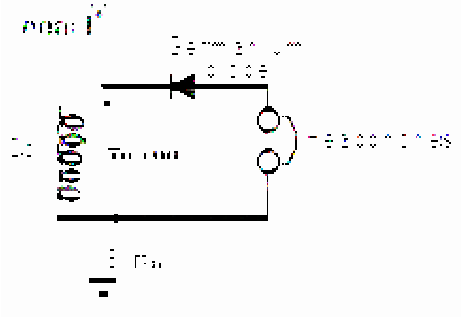

While the circuit diagram above is marked "Tuning capacitor" that is actually quite misleading. Yes, you tune the radio receiver by adjusting the setting of the variable capacitor, **but**, what the capacitor is doing is altering the resonant frequency of the coil/capacitor combination and it is the resonant frequency of that combination which is doing exactly the same job as the variable coil did on it's own.

This draws attention to two very important facts concerning coil/capacitor combinations. When a capacitor is placed across a coil "in parallel" as shown in this radio receiver circuit, then the combination has a very high impedance (resistance to AC current flow) at the resonant frequency. But if the capacitor is placed "in series" with the coil, then there is nearly zero impedance at the resonant frequency of the combination:

This may seem like something which practical people would not bother with, after all, who really cares? However, it is a very practical point indeed. Remember that Don Smith often uses an early version, off-the-shelf neon-tube driver module as an easy way to provide a high-voltage, high-frequency AC current source, typically, 6,000 volts at 30,000 Hz. He then feeds that power into a Tesla Coil which is itself, a power amplifier. The arrangement is like this:

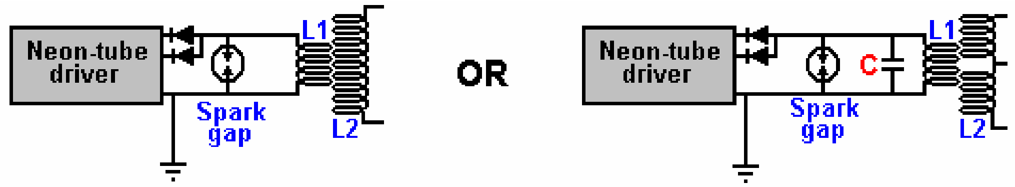

People who try to replicate Don's designs tend to say "I get great sparks at the spark gap until I connect the **L1** coil and then the sparks stop. This circuit can never work because the resistance of the coil is too low".

If the resonant frequency of the **L1** coil does not match the frequency being produced by the neon-tube driver circuit, then the low impedance of the **L1** coil at that frequency, will definitely pull the voltage of the neon-tube driver down to a very low value. But if the **L1** coil has the same resonant frequency as the driver circuit, then the **L1** coil (or the **L1** coil/capacitor combination shown on the right, will have a very high resistance to current flow through it and it will work well with the driver circuit. So, no sparks, means that the coil tuning is off. It is the same as tuning a radio receiver, get the tuning wrong and you don't hear the radio station.

This is very nicely demonstrated using simple torch bulbs and two coils in the YouTube video showing good output for almost no input power: http://www.youtube.com/watch?v=kQdcwDCBoNY and while only one resonant pick-up coil is shown, there is the possibility of using many resonant pick-up coils with just the one transmitter.

It should be stated that people have tried to replicate the devices which Don has described and have not been successful. Consequently, they believe that Don's presentations are a hoax and that he has never had any device which worked as he claimed. That is a rather doubtful conclusion to draw as the only certain thing which can be concluded is that they failed in their attempt. It is my experience that some people who attempt replications do not seem to have the 'knack' or perseverance needed to reach a successful result, while others do have what it takes. For example, many people tried to replicate Stan Meyer's water-splitter which he called his "Water Fuel Cell" or "WFC" for short, and they did not succeed. Dave Lawton in Wales decided to produce a replication and he had zero success for a whole month of effort. After a month of trying, his cell suddenly bust into substantial gas production, showing that the electrode pipes had to be "conditioned" before the cell would work properly. Most people would have given up long before the month was up, and declared that the WFC was a hoax. As a result of Dave's efforts and his generous sharing of what he did (in chapter 10), many people have now been very successful in replicating Stan Meyer's design. So, I will leave it up to you to decide if Don Smith's devices are a hoax. It should be understood, that there is no particular reason why anyone who was successful with Don's designs would necessarily tell me. Most will never even have heard of me.

Bearing that in mind, I have been told of one man who used his common sense and produced an impressive result. He used a Tesla Coil as the driving force, and then used a second Tesla Coil back-to-back with the first one, to step the high voltage back down again. Doing that, he was able to light a series of powerful light bulbs from the "L1" output coils. He also confirmed that doubling the voltage, quadrupled the power output, verifying what Don said. He also found that adding additional coils with bulbs to the output Tesla Coil, did not increase the input power at all, did not cause any of the existing light bulbs to shine any less brightly, and yet lit the additional bulbs. That would appear to be confirmation of Don's statement that any number of magnetic copies of the original oscillating magnetic field of the first Tesla Coil, can provide a full-power electrical output without requiring any additional input power. I'm no expert, but my understanding of the arrangement is:

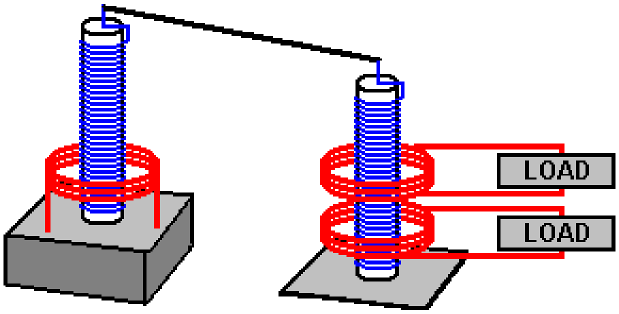

As the large diameter coil is exactly one quarter the length of the smaller diameter coil, there is an automatic resonance of both when the applied frequency is just right. As the first narrow coil is identical to the second narrow coil, they are also automatically resonant together. Again, as the large coils which feed the loads are exactly one quarter the wire length of the narrow coils, they also resonate at the common frequency and at that frequency, the input power is at its minimum while the output power is at its maximum. The spike at the top of each of the narrow coils is connected with a wire to channel the generated power from the first Tesla Coil to the second one.

As some readers may feel that there is some "black magic" about the neon-driver circuit used by Don to drive the Tesla Coil section of his circuitry and that if a suitable unit could not be purchased then the circuit could not be reproduced or tested, it seems reasonable to show how it operates and how it can be constructed from scratch:

The circuit itself is made up of an oscillator to convert the 12-volt DC supply into a pulsating current which is then stepped up to a high voltage by a transformer. Here is a circuit which has been used for this:

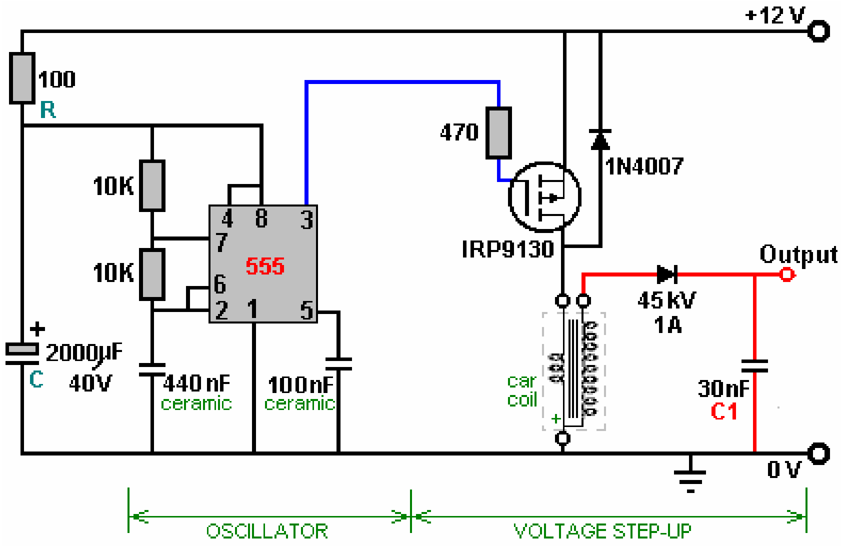

The supply for the 555 timer chip is protected against spikes and dips by the resistor "R" and the capacitor "C". The 555 timer chip acts as an oscillator or "clock" whose speed is governed by the two 10K resistors feeding the 440 nF capacitor. The step-up transformer is an ordinary car coil and the drive power to it is boosted by the IRP9130 FET transistor which is driven by the 555 chip output coming from it's pin 3.

The output from the (Ford Model T) car coil is rectified by the diode, which needs to have a very high voltage rating as the voltage at this point is now very high. The rectified voltage pulses are stored in a very high-voltage capacitor before being used to drive a Tesla Coil. As a powerful output is wanted, two car coils are used and their outputs combined as shown here:

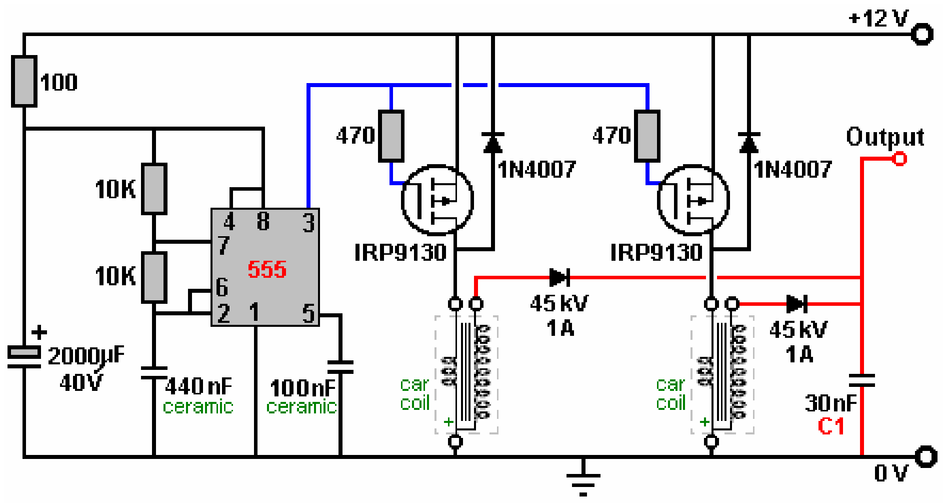
+12 V
100
470
1N4007
470
1N4007
Output
10K
4
8
3
7
555
6
2
1
5
10K
IRP9130
IRP9130
45 kV
1A
45 kV
1A
+
2000µF
40V
440 nF
ceramic
100nF
ceramic
car
coil
+
car
coil
+
30nF
C1
0 V

You will notice that the car coil has only three terminals and the terminal marked "+" is the one with the connection common to both of the coils inside the housing. The coil may look like this:

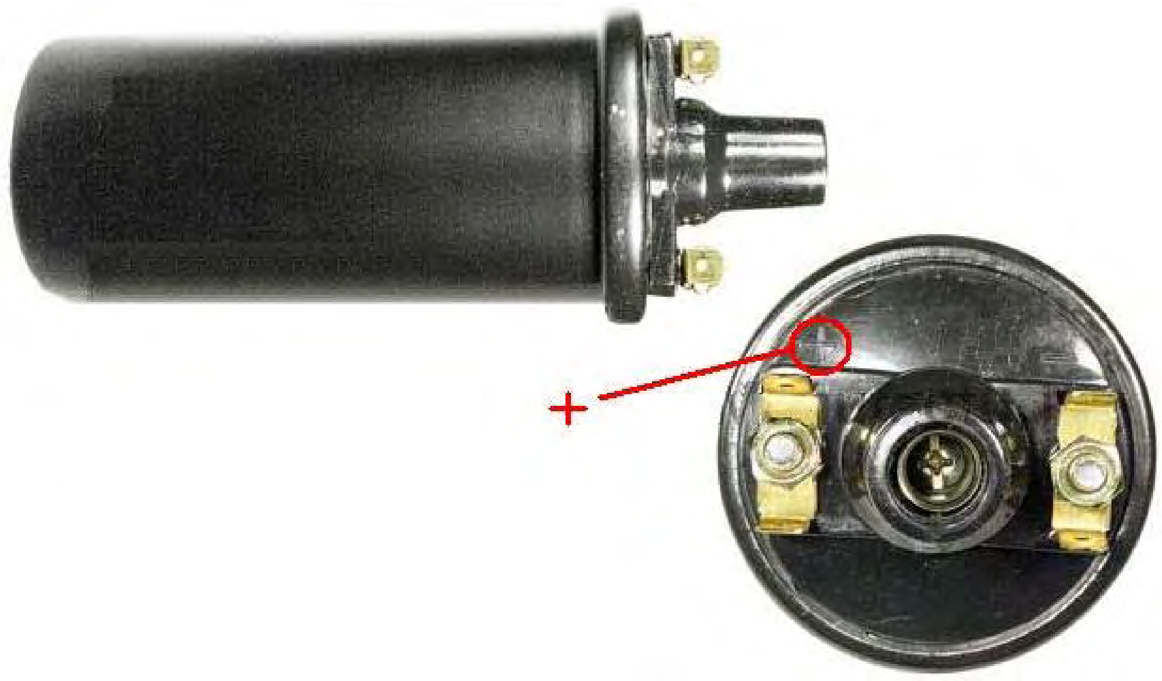

and the "+" is generally marked on the top beside the terminal with the two internal connections running to it. The circuit described so far is very close to that provided by a neon-tube driver circuit and it is certainly capable of driving a Tesla Coil.

There are several different way of constructing a Tesla Coil. It is not unusual to have several spark gaps connected in a chain. This arrangement is called a "series spark gap" because the spark gaps are connected "in series" which is just a technical way of saying "connected in a row". In the chapter on aerial systems, you will see that Hermann Plauston uses that style of spark gap with the very high voltages which he gets from his powerful aerial systems. These multiple spark gaps are much quieter in operation than a single spark gap would be. One of the possible Tesla Coil designs uses a pancake coil as the "L1" coil as that gives even higher gain. The circuit is as shown here:

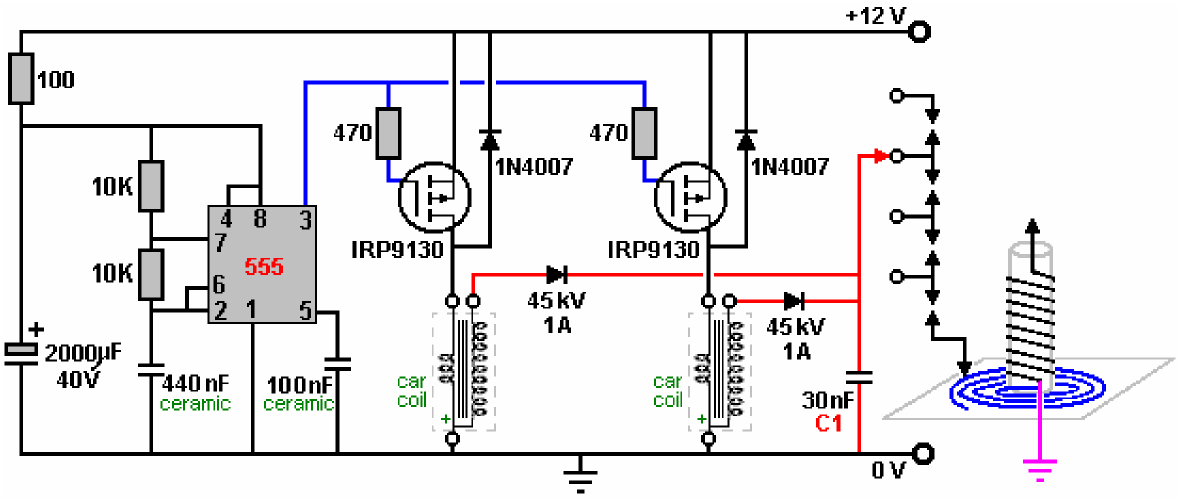

The connection to the pancake coil is by a moveable clamp and the two coils are tuned to resonance by careful and gradual adjustment of that connection. The series spark gap can be constructed in various ways, including using car spark plugs. The one shown here uses nuts and bolts projecting through two strips of a stiff, non-conducting material, as that is much easier to adjust than the gaps of several spark plugs:

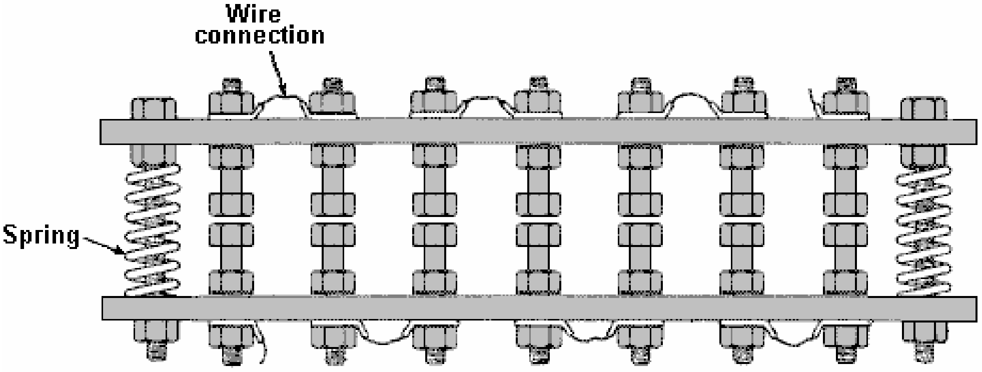

Tightening the bolts which compress the springs moves the bolt heads closer together and reduces all of the spark gaps. The electrical connections can be made to the end tags or to any of the intermediate wire connection straps if fewer spark gaps are required in the chain.

Let me remind you again that this is not a toy and very high voltages will be produced. Also, let me stress again that if you decide to construct anything, then you do so entire on your own responsibility. This document is only provided for information purposes and must not be seen as an encouragement to build any such device nor is any guarantee given that any of the devices described in this eBook will work as described should you decide to attempt to construct a replication prototype of your own. Generally, it takes skill and patience to achieve success with any free-energy device and Don Smith's devices are some of the most difficult, especially since he admits quite freely that he does not disclose all of the details.

The output capacitor marked "C1" in the circuit diagram above has to be able to handle very high voltages. There are various ways of dealing with this. Don dealt with it by getting very expensive capacitors manufactured by a specialist company. Some home-based constructors have had success using glass beer bottles filled with a salt solution. The outside of the bottles are wrapped in aluminium foil to form one of the contacts of the capacitor and bare wires are looped from deep inside each bottle on to the next one, looping from the inside of one bottle to the inside of the next one, and eventually forming the other contact of the capacitor. While that appears to work well, it is not a very convenient thing to carry around. An alternative is just to stand the bare bottles in a container which is lined with foil which forms the second contact of the capacitor.

One method which has been popular in the past is to use two complete rolls of aluminium foil, sometimes called "baking foil", laying them one flat, covering it with one or more layers of plastic cling film and laying the second roll of foil on top of the plastic. The three layers are then rolled up to form the capacitor. Obviously, several of these can be connected together in parallel in order to increase the capacitance of the set. The thicker the plastic, the lower the capacitance but the higher the voltage which can be handled.

The November 1999 issue of Popular Electronics suggests using 33 sheets of the thin aluminium used as a flashing material by house builders. At that time it was supplied in rolls which were ten inches (250 mm) wide, so their design uses 14" (355 mm) lengths of the aluminium. The plastic chosen to separate the plates was polythene sheet 0.062 inch (1.6 mm) thick which is also available from a builders merchants outlet. The plastic is cut to 11 inch (280 mm) by 13 inch (330 mm) and assembly is as follows:

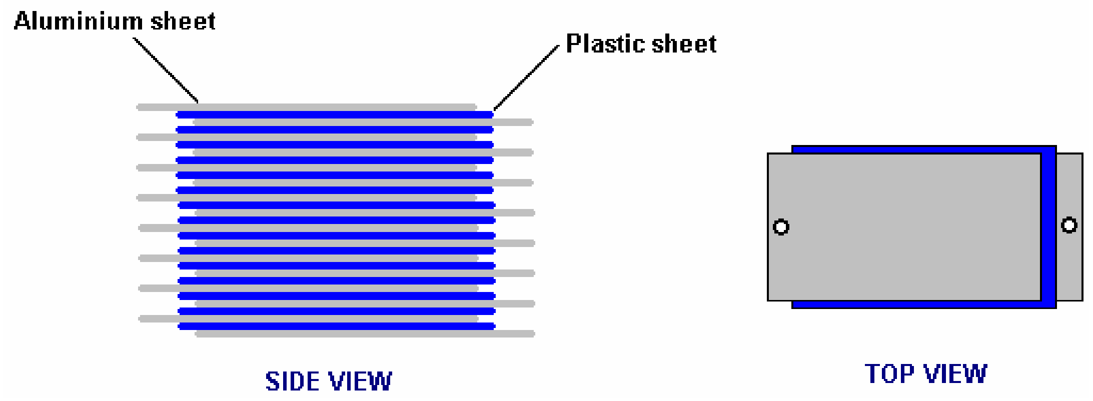

The sandwich stack of sheets is then clamped together between two rigid timber sheets. The tighter that they are clamped, the closer the plates are to each other and the higher the capacitance. The electrical connections are made by running a bolt through the projecting ends of the plates. With two thicknesses of plastic sheet and one of aluminium, there should be room for a washer between each pair of plates at each end and that would improve the clamping and the electrical connection. An alternative is to cut a corner off each plate and position them alternatively so that almost no plate area is ineffective.

As Don Smith has demonstrated in one of his video presentations, Nikola Tesla was perfectly correct when he stated that directing the discharge from a Tesla Coil on to a metal plate (or in Don's case, one of the two metal plates of a two-plate capacitor where a plastic sheet separates the plates just as shown above), produces a very powerful current flow onwards through a good earth connection. Obviously, if an electrical load is positioned between the plates and the earth connection, then the load can be powered to a high level of current, giving a very considerable power gain.

Coil Construction: The Barker & Williamson coils used by Don in his constructions are expensive to purchase. Some years ago, in an article in a 1997 issue of the “QST” amateur radio publication, Robert H. Johns shows how similar coils can be constructed without any great difficulty. The Electrodyne Corporation research staff have stated that off-the-shelf solid tinned copper wire produces three times the magnetic field that un-tinned copper does, so perhaps that should be borne in mind when choosing the wire for constructing these coils.

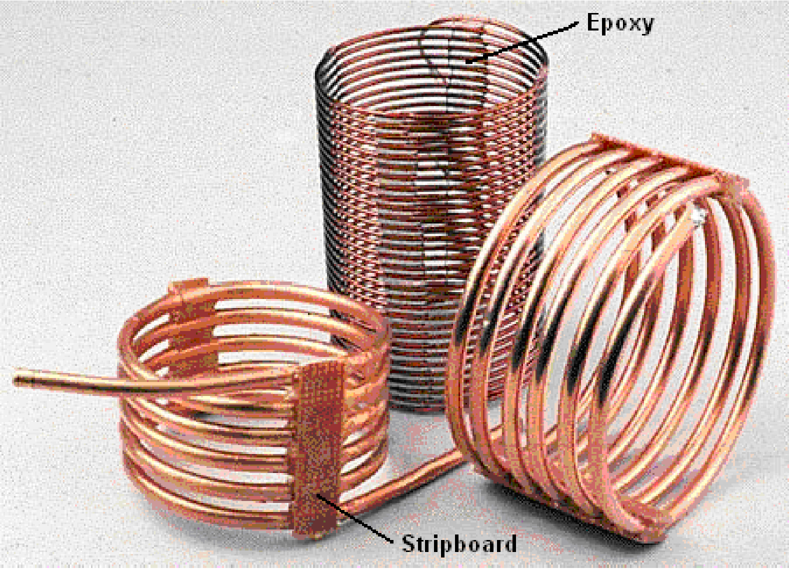

These home-made coils have excellent "Q" Quality factors, some even better than the tinned copper wire coils of Barker & Williamson because the majority of electrical flow is at the surface of the wire and copper is a better conductor of electricity than the silver tinning material.

The inductance of a coil increases if the turns are close together. The capacitance of a coil decreases if the turns are spread out. A good compromise is to space the turns so that there is a gap between the turns of one wire thickness. A common construction method with Tesla Coil builders is to use nylon fishing line or plastic strimmer cord between the turns to create the gap. The method used by Mr Johns allows for even spacing without using any additional material. The key feature is to use a collapsible former and wind the coil on the former, space the turns out evenly and then clamp them in position with strips of epoxy resin, removing the former when the resin has set and cured.

Mr Johns has difficulty with his epoxy being difficult to keep in place, but when mixed with the West System micro fibres, epoxy can be made any consistency and it can be applied as a stiff paste without any loss of it's properties. The epoxy is kept from sticking to the former by placing a strip of electrical tape on each side of the former.

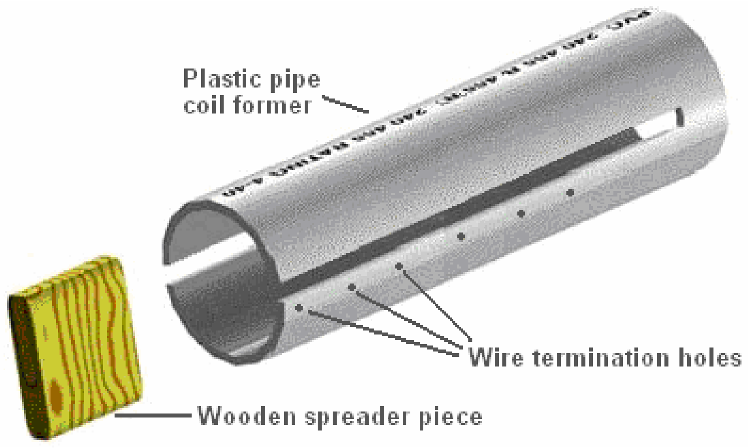

I suggest that the plastic pipe used as the coil former is twice the length of the coil to be wound as that allows a good degree of flexing in the former when the coil is being removed. Before the two slots are cut in the plastic pipe, a wooden spreader piece is cut and it's ends rounded so that it is a push-fit in the pipe. This spreader piece is used to hold the sides of the cut end exactly in position when the wire is being wrapped tightly around the pipe.

Two or more small holes are drilled in the pipe beside where the slots are to be cut. These holes are used to anchor the ends of the wire by passing them through the hole and bending them. Those ends have to be cut off before the finished coil is slid off the former, but they are very useful while the epoxy is being applied and hardening. The pipe slots are cut to a generous width, typically 10 mm or more.

The technique is then to wedge the wooden spreader piece in the slotted end of the pipe. Then anchor the end of the solid copper wire using the first of the drilled holes. The wire, which can be bare or insulated, is then wrapped tightly around the former for the required number of turns, and the other end of the wire secured in one of the other drilled holes. It is common practice to make the turns by rotating the former. When the winding is completed, the turns can be spaced out more evenly if necessary, and then a strip of epoxy paste applied all along one side of the coil. When that has hardened, (or immediately if the epoxy paste is stiff enough), the pipe is turned over and a second epoxy strip applied to the opposite side of the coil. A strip of paxolin board or strip-board can be made part of the epoxy strip. Alternatively, an L-shaped plastic mounting bracket or a plastic mounting bolt can be embedded in the epoxy ready for the coil installation later on.

When the epoxy has hardened, typically 24 hours later, the coil ends are snipped off, the spreader piece is tapped out with a dowel and the sides of the pipe pressed inwards to make it easy to slide the finished coil off the former. Larger diameter coils can be wound with small-diameter copper pipe.

The coil inductance can be calculated from:

Inductance in microhenrys $L = d^2n^2 / (18d + 40l)$

Where:
d is the coil diameter in inches measured from wire centre to wire centre
n is the number of turns in the coil
l is coil length in inches (1 inch = 25.4 mm)

Using this equation for working out the number of turns for a given inductance in microhenrys:

$$n = \frac{\sqrt{L(18d + 40l)}}{d}$$

Patrick Kelly
engpjk@yahoo.co.uk
http://www.free-energy-info.co.uk
http://www.free-energy-devices.com

ELECTRICITY and ELECTRONICS FOR HVAC

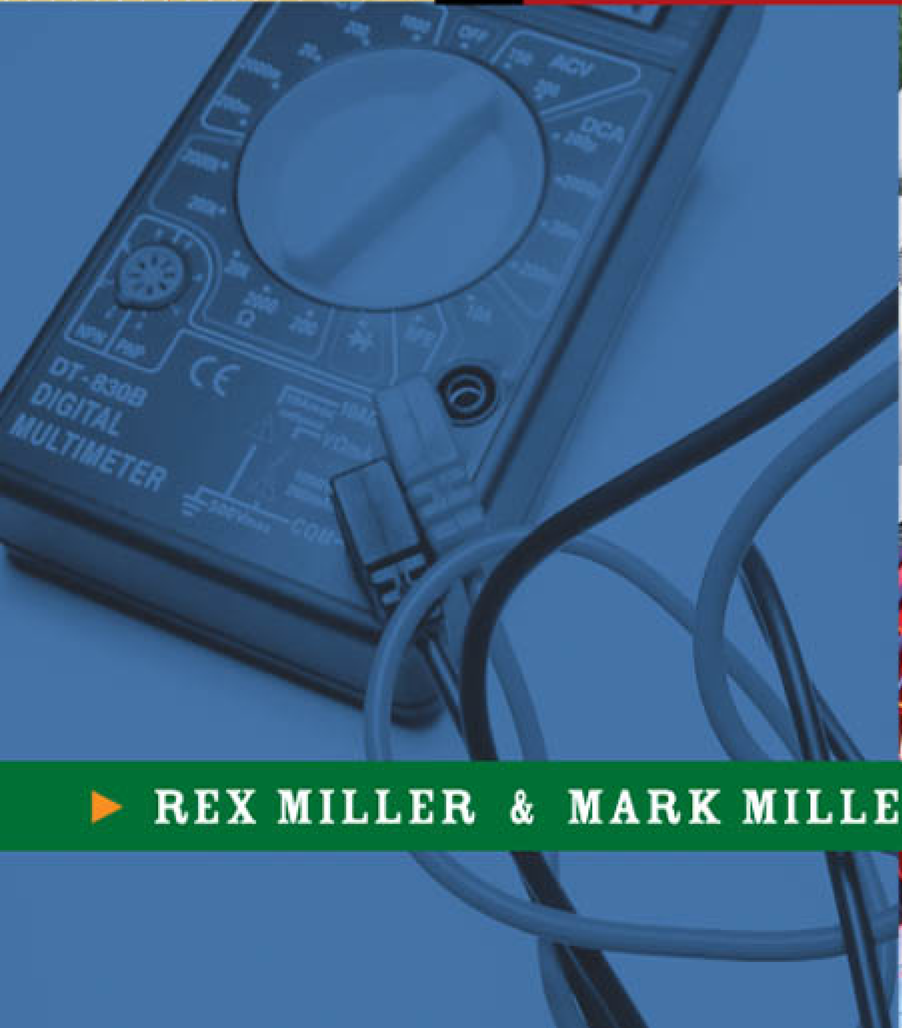
DT-830B
DIGITAL
MULTIMETER
ACV
DCA
OFF
CE
NPN
PNP

REX MILLER & MARK MILLER

Electricity and Electronics for HVAC

ABOUT THE AUTHOR

Rex Miller is Professor Emeritus of Industrial Technology at the SUNY College at Buffalo. He has taught technical curriculums at the high school, technical school, and college levels for more than 40 years. Dr. Miller is the author or coauthor of over 100 textbooks for vocational and industrial arts programs, including McGraw-Hill's best-selling *Carpentry and Construction* and *Electrician's Pocket Manual*.

Mark R. Miller is Professor and Coordinator of Industrial Technology at the University of Texas at Tyler. He has taught technical curriculums at the high school, technical school, and college levels for more than 25 years. Dr. Miller is the author or coauthor of over 30 textbooks for trade and technical programs, including McGraw-Hill's *Welding Licensing Exam Study Guide* and *Air Conditioning and Refrigeration.*

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

Electricity and Electronics for HVAC

Rex Miller
Professor Emeritus
State University College at Buffalo

Mark R. Miller
Professor
University of Texas at Tyler

McGraw Hill

New York Chicago San Francisco Lisbon London Madrid
Mexico City Milan New Delhi San Juan Seoul
Singapore Sydney Toronto

Copyright © 2007 by The McGraw-Hill Companies, Inc. All rights reserved. Manufactured in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

0-07-154270-1

The material in this eBook also appears in the print version of this title: 0-07-149668-8.

All trademarks are trademarks of their respective owners. Rather than put a trademark symbol after every occurrence of a trademarked name, we use names in an editorial fashion only, and to the benefit of the trademark owner, with no intention of infringement of the trademark. Where such designations appear in this book, they have been printed with initial caps.

McGraw-Hill eBooks are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. For more information, please contact George Hoare, Special Sales, at george_hoare@mcgraw-hill.com or (212) 904-4069.

TERMS OF USE

This is a copyrighted work and The McGraw-Hill Companies, Inc. (“McGraw-Hill”) and its licensors reserve all rights in and to the work. Use of this work is subject to these terms. Except as permitted under the Copyright Act of 1976 and the right to store and retrieve one copy of the work, you may not decompile, disassemble, reverse engineer, reproduce, modify, create derivative works based upon, transmit, distribute, disseminate, sell, publish or sublicense the work or any part of it without McGraw-Hill’s prior consent. You may use the work for your own noncommercial and personal use; any other use of the work is strictly prohibited. Your right to use the work may be terminated if you fail to comply with these terms.

THE WORK IS PROVIDED “AS IS.” McGRAW-HILL AND ITS LICENSORS MAKE NO GUARANTEES OR WARRANTIES AS TO THE ACCURACY, ADEQUACY OR COMPLETENESS OF OR RESULTS TO BE OBTAINED FROM USING THE WORK, INCLUDING ANY INFORMATION THAT CAN BE ACCESSED THROUGH THE WORK VIA HYPERLINK OR OTHERWISE, AND EXPRESSLY DISCLAIM ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. McGraw-Hill and its licensors do not warrant or guarantee that the functions contained in the work will meet your requirements or that its operation will be uninterrupted or error free. Neither McGraw-Hill nor its licensors shall be liable to you or anyone else for any inaccuracy, error or omission, regardless of cause, in the work or for any damages resulting therefrom. McGraw-Hill has no responsibility for the content of any information accessed through the work. Under no circumstances shall McGraw-Hill and/or its licensors be liable for any indirect, incidental, special, punitive, consequential or similar damages that result from the use of or inability to use the work, even if any of them has been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause arises in contract, tort or otherwise.

DOI: 10.1036/0071496688

Professional

Want to learn more?

We hope you enjoy this McGraw-Hill eBook! If you'd like more information about this book, its author, or related books and websites, please click here.

This book is dedicated to
Patricia Ann Navara Miller
A great Mother, Wife, and Educator

This page intentionally left blank

For more information about this title, click here

Contents

Preface xvii
Acknowledgments xix

Chapter 1. Introduction to Electricity 1

Performance Objectives 1
Introduction to Electricity 1
Static Electricity and Magnetism 2
Electricity 2
Electricity's Future 3
Matter and Electricity 4
  Solids, gases, and liquids 4
Atom 5
  Electrons, protons, and neutrons 6
Properties of Electrons 8
Orbiting Electrons 8
Electrical Charge 9
Outer Shell 10
  Valence electrons and ions 10
A Practical Unit of Charge 12
The Volt 12
Controlling Electrons 13
Difference of Potential (Voltage) 13
Electron Flow (Current) 14
Conductors 14
Resistance 14
The Electric Circuit 15
Connecting a Circuit 16
Switches Control Electron Flow 16
Schematic 17
Review Questions 17

Chapter 2. Current, Voltage, Resistance, Power, and Ohm's Law 19

Performance Objectives 19
Sources of Electricity 19
  Chemical action 19
  Heat 19
  Light 20
  Pressure 20
  Magnetism 20
Units of Measurement 21
  Unit of charge: coulomb 21
  Unit of current: ampere 22
  Unit of potential difference: volt 22
  Unit of resistance: ohm 22
  Siemens 23
Prefixes 23
Ohm's Law 24
  Ohm's law examples 25
  Ohm's law and the closed circuit 27
  Ohm's law in other forms 29
  Uses for ohm's law 30
Power 31
Joule 31
Watt 32
Putting Electricity to Work 32
  Resistors and heat 32
  Mechanical energy 32
A Practical Unit of Electrical Power: Kilowatt-Hour 34
Review Questions 35

Chapter 3. Resistors, Other Electric Components, and Their Symbols 37

Performance Objectives 37
Color Code 38
  Tolerance 39
  Examples 39
  Gold and silver third bands 40
Variable Resistors 41
Schematic 43
Types of Resistors 43
  Fixed resistors 43
  Tapped resistors 44
  Variable resistors 44
  Fusible resistors 44
  Temperature-compensating resistors 47
Types of Capacitors 47
  Fixed capacitors 47
  Electrolytic capacitors 48
  Variable capacitors 49
Types of Inductors 50
  Fixed inductors 50
  Variable inductors 51
Transformers 51

Semiconductors 52
  Diodes 52
  Transistors 52
Switches 53
Relays 53
Fuses and Circuit Breakers 53
Lamps 55
  Batteries 56
  Meters 57
Reading Schematics 57
Review Questions 57

Chapter 4. Series and Parallel Circuits 59

Performance Objectives 59
Series Circuit 59
  Resistances in series 61
  Voltages in series 61
  Current in series 61
  Series circuit rules 62
  Fuses 62
Parallel Circuit 63
  Cells in parallel 63
  Parallel circuit characteristics 63
  Currents in a parallel circuit 64
  Resistances in a parallel circuit 64
  Parallel circuit rules 65
Series-Parallel Circuits 65
  Series-parallel resistance circuits 66
  Determining resistance 66
  Determining current in series-parallel circuits 68
  Determining voltage in series-parallel circuits 68
Review Questions 72

Chapter 5. Magnetism, Solenoids, and Relays 73

Performance Objectives 73
Permanent Magnets 74
Temporary Magnets 75
Electromagnets 75
Magnetic Theory 76
Magnetic Permeability 76
Shapes of Magnets 76
  Poles of magnets 78
Electromagnetism 79
  Magnetism in a coil of wire 80
  Electromagnets 80
  Using electromagnetism 82
The Solenoids and Relays 82
  Power relays 86
  Solenoid valves 86
  Principles of operation 86
Applications 87
  Liquid line service 87
  Suction line service 89

High-Temperature Applications 89
  Hot-gas defrost service 90
Transformers for Low-Voltage Controls 91
Review Questions 92

Chapter 6. Electrical Measuring Instruments 93

Performance Objectives 93
Electricity 93
Types of Meter Movements 94
  D'Arsonval meter movement 94
  Analog meters 97
Parallax Error 97
Ammeters 98
  Extending the range of an ammeter 99
  Connecting an ammeter in the circuit 99
  Clamp-on ammeters 99
  AC ammeters 101
Voltmeters 101
  Connecting a voltmeter 102
AC Voltmeters 102
Ohmmeters 103
Multimeters 103
Meggers 104
Digital Meters 104
Other Instruments 106
  Automatic halogen leak detector 106
  Electronic sight glass 107
  Electronic charging meters 107
Using an Ohmmeter 108
  Warning 108
  Adjusting the meter 108
Review Questions 110

Chapter 7. Electrical Power: Direct Current and Alternating Current 111

Performance Objectives 111
Power Sources 111
Types of Batteries 112
  Primary cell 112
  Secondary cell 112
Dry Cells 112
Battery Specifications 115
Connecting Cells 115
Battery Maintenance 118
Nickel-Cadmium Cells 119
Alkaline Cells 119
Alternating Current 119
Sine Wave 120
Sine-Wave Characteristics 121
Phase 122
Average, RMS, and Peak Values of AC 122

Polyphase Alternating Current 123
  Three phase 124
  Three-phase connections 124
Electrical Properties of Delta and Wye 126
  Delta connection (Δ) 126
  Wye connection (Y) 126
  Delta and wye summarized 126
Review Questions 127

Chapter 8. Inductors and Transformers 129

Performance Objectives 129
Changing Inductance 130
Self-Inductance 130
Mutual Inductance 133
Inductive Reactance 134
  Measuring X_L 134
Power in an Inductive Circuit 134
Uses for Inductive Reactance 135
Transformers 135
  Mutual inductance in a transformer 135
  Iron-core transformer 136
  Construction 137
  Types of iron-core transformers 137
Voltage Transfer 138
  Step up, step down 139
  Turns ratio 139
Power Transformers 140
Audio Frequency and Radio Frequency Transformers 140
Autotransformers 141
Transformer Losses 142
Inductive Circuits 144
  Special handling 145
Utilizing the Inductive Delay 146
Review Questions 146

Chapter 9. Capacitors and Capacitive Reactance 147

Performance Objectives 147
The Capacitor 149
How the Capacitor Works 149
Capacity of a Capacitor 150
Breakdown Voltage 151
Basic Units of Capacitance 151
Types of Capacitors 151
Electrolytic Capacitors 153
  Making an electrolytic capacitor 154
  Connecting electrolytics 154
Working Voltage, Direct Current (WVDC) 155
Capacitive Reactance 155
Capacitor Causes a Lagging Voltage 156
Checking Capacitors 157
Review Questions 158

Chapter 10. Single-Phase and Three-Phase Alternating Current 161

Performance Objectives 161
Single-Phase and Three-Phase AC 161
Resistance, Capacitance, and Inductance 162
Power Factor 164
Distributing Electric Power 165
Polyphase 166
Circuit Breakers 169
Review Questions 171

Chapter 11. Solid-State Controls 173

Performance Objectives 173
Semiconductors 173
Semiconductor Principles 174
Diode 175
Silicon-Controlled Rectifiers 177
Transistors 178
  Transistor impedances 179
Integrated Circuits 180
  Solid-state demand defrost control 182
Thermistor Sensing 182
Humidity Sensing 184
Bridge Circuit 186
  Unbalanced bridge 187
Sensors 189
Controllers 189
Electronic Controllers 189
Differential Amplifiers 190
Actuators 191
Other Devices 192
Solid-State Compressor Motor Protection 192
  Operation of the module 192
  Troubleshooting the control 194
  Restoring service 196
Review Questions 197

Chapter 12. Alternating Current Motors 199

Performance Objectives 199
Motors 199
Split-Phase Motor 204
  Getting the motor started 204
Repulsion Start, Induction Run Motor 206
Capacitor-Start Motor 208
Permanent Split-Capacitor Motor 210
  Theory of operation 212
Capacitor-Start, Capacitor-Run Motor 215
  Theory of operation 215
Three-Phase Motor 216
  Theory of operation 216

Capacitor Ratings 219
Start Capacitors and Bleeder Resistors 219
  Run capacitors 221
Motor Protectors 221
Compressor Motor Relays 221
  Current-type relay 222
  Potential-type relay 222
Review Questions 224

Chapter 13. Electrical Safety 225

Performance Objectives 225
Safety Devices 225
Safety Precautions 225
Main Switches 226
Portable Electrical Tools 227
Ground-Fault Circuit Interrupters (GFCI) 229
Types of Circuit Protectors 230
Review Questions 231

Chapter 14. Control Devices 233

Performance Objectives 233
Power Relays 233
  Motor start relays 234
  Relays with more than one contact 235
  Thermal overload protectors 236
  Time-delay relays 237
Solenoids 238
Thermostats 239
  Bellows-type thermostats 239
  Bimetallic-type thermostats 241
  Heating and cooling thermostats 242
  Microprocessor thermostats 245
Thermostat Adjustments 245
  Heat anticipators 246
  Cold anticipators 246
Switches of Many Types 246
Pressure Control Switches 247
Water Tower Controls 248
Review Questions 250

Chapter 15. Heating Circuits 251

Performance Objectives 251
Basic Gas Furnace Operation 252
Basic Electric Heating System 253
  Basic operation 253
Ladder Diagrams 254
Manufacturer’s Diagrams 257
Field Wiring 258
Low-Voltage Wiring 258

Heat Pumps 260
  Operation 261
  Special requirements of heat-pump systems 266
  Heat-pump combinations 267
High-Efficiency Furnaces 267
  Operation 268
  Electrical controls 269
  Sequence of operation 270
  Combustion process 271
Troubleshooting the Pulse Furnace 271
Review Questions 278

Chapter 16. Air-Conditioning Circuits 279

Performance Objectives 279
Basic Air-Conditioning Unit 279
  Push-button switches 282
  Fans 283
Schematics 283
Ladder Diagrams 287
  Thermostats 287
  Fan switches 291
Troubleshooting 291
Review Questions 291

Chapter 17. Refrigeration Circuits 293

Performance Objectives 293
Basic Refrigerator 293
  Thermostats 293
  Start relays 295
Refrigerator-Freezer Combination 295
  Manual defrost 295
  Automatic defrost 296
Defrosting 296
  Defrost thermostats 297
  Defrost cycle 299
Other Devices 302
Troubleshooting 304
Rapid Electrical Diagnosis 306
Energy-Saver Switches 309
Review Questions 310

Chapter 18. Troubleshooting 311

Performance Objectives 311
Safety 311
  Handling refrigerants 312
  Testing precaution 312
  Electrical safety 312
Compressor Problems 312
  PSC compressors 312
  Air-conditioner compressors 312

Low-Voltage Operation 315
Using a System to Troubleshoot (Electrical) 316
  Motor testing 316
  Capacitor testing 317
  Motor-protector relay testing 317
Using Meters to Check for Problems 318
Using a Volt-Ammeter for Troubleshooting Electric Motors 320
  Clamp-on volt-ammeters 320
  Looking for grounds 321
  Looking for opens 322
  Looking for shorts 322
Motors with Squirrel-Cage Rotors 323
Testing the Centrifugal Switch in a Single-Phase Motor 323
Testing for a Short Circuit Between Run and Start Windings 324
  Capacitor testing 324
  Measuring the capacity of a capacitor 325
Troubleshooting Procedure 326
Review Questions 327

Chapter 19. Controlling Electrical Power for Air-Conditioning Units 329

Performance Objectives 329
Choosing Wire Size 329
  Limiting voltage loss 330
  Minimum wire size 330
  Wire selection 330
Wire Size and Low Voltage 330
  Voltage drop calculations 332
The Effects of Voltage Variations on AC Motors 333
Selecting Proper Wire Size 335
Unacceptable Motor Voltages 336
Calculating Starting Current Values and Inrush Voltage Drops 336
  Single-phase current 336
  Three-phase circuits 338
  Inrush voltage drop 339
Code Limitations on Amperes Per Conductor 339
Heat Generated Within Conductors 339
Circuit Protection 340
  Standard rule 341
Fuses 341
  One-time single-element fuses 341
  Time-delay two-element fuses 342
  Types of fuses 343
Thermostats 343
  Thermostat as a control switch 344
Service 345
Starter Kits 346
Single-Phase Line Monitors 347
Time Delays 350
Pressure Controls 350
  Head pressure controls 350
  Line voltage head pressure controls 352

Three-Phase Line Voltage Monitor 355
Current Sensing 355
Review Questions 359

Chapter 20. Careers in Air Conditioning and Refrigeration 361

Performance Objectives 361
Industries that Employ Air-Conditioning and Refrigeration Mechanics 361
Job Qualifications 364
The Future 365
Pay and Benefits 365
Teaching as a Career 367
Sources of Additional Information 367
Review Questions 369

Index 371

Preface

This book is intended to provide an introduction to the basic principles of electricity and electronics as they apply to refrigeration, heating, and air-conditioning systems. It combines basic principles, using very little math, with the latest applications found in the HVAC industry.

The fields of refrigeration, air conditioning, and heating are undergoing some very rapid changes. The advent of the computer chip has made it possible to control heating and cooling systems precisely and with a great deal of freedom in programming their applications. The new high-efficiency furnaces utilize the chip both for sequencing and for protection from accidental damage. Most of these programmable controllers provide instructions and technical bulletins. They are numerous, and each has its own approach to solving a given problem. This book will make it possible for you to understand these instructions and improvements.

Future technicians need to deal with the fact that change is inevitable, and that they will have to keep up with the latest developments as long as they work in the field. It is hoped that this book will make that task easier.

REX MILLER
MARK R. MILLER

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

This page intentionally left blank

Acknowledgments

No author works without being influenced and aided by others, and this book is no exception. A number of people cooperated in providing technical data and illustrations, and for this we are grateful. We would like to thank those organizations that so generously contributed information and illustrations. The following have been particularly helpful:

Amprobe Instrument Division of SOS Consolidated, Inc.
Carrier Corporation
Eveready Division of Union Carbide
General Controls
General Electric Company
Honeywell, Inc.
Kelvinator Appliance Company
Lennox Industries (Toronto, Canada)
National Safety Council
Robert Shaw Controls Company
Sporlan Valve Company
Tecumseh Products Company
Tyler Refrigeration Corporation
Wadsworth Electric Company
Wagner Electric Company
Westinghouse Electric Company
Weston Electrical Instruments Company

REX MILLER
MARK R. MILLER

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

This page intentionally left blank

Electricity and Electronics for HVAC

This page intentionally left blank

Chapter

1

Introduction to Electricity

Performance Objectives

- Understand how matter and electricity are related.
- Understand how liquids, solids, and gases are similar, yet different.
- Understand how atoms, electrons, protons, and neutrons are all related to the production of electricity.
- Understand how electricity is put to work in the electric circuit.
- Understand how switches control the flow of electricity.

Introduction to Electricity

Electricity is as old as the universe. But knowledge about it is relatively new. Early humans were aware of electricity in the form of lightning. They learned of its power when they saw it starts fires and kills people and animals. But it was only about 300 years ago that people began to learn the basic laws of electricity. And only about 120 years have passed since electricity was first put to work. It has been only 100 years since the first practical electric lamp was invented, only 80 years since the vacuum tube was invented, and less than 40 years since the transistor was invented. Despite this brief time, electricity has greatly changed people's lives—our lives.

The universe consists of atoms and every atom contains at least one electron. An *electron* is the smallest particle of an atom and has a negative electric charge. When the movement of electrons is controlled, they are capable of doing work.

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

Static Electricity and Magnetism

There are two types of electrical effects: static electricity and magnetism.

The Greek philosopher Thales, who lived about 2500 years ago, is credited with discovering static electricity.

Magnetism is the ability of an object to attract other objects. It was discovered about 2600 B.C., or about 100 years before the discovery of static electricity. It is not certain who discovered magnetism first. Some historians say it was first observed by the Chinese. Others say that it was the Greeks. The discoverer noted that certain heavy stones have the power to attract and lift iron and some other stones. The material in these stones is called *magnetite*. It was named by the Greeks for the province of Magnesia in Asia Minor, where the stones were first found. Today, the power of this stone is called *magnetism*.

These discoveries led to extensive studies of magnetism and static electricity (see Figs. 1-1 and 1-2).

Other people studied electricity and magnetism during the sixteenth century and later in the nineteenth century. Some of these people have electrical terms named for them. You may be familiar with Ampere, Volta, Coulomb, Oersted, Ohm, and Galvani (see Fig. 1-3).

Electricity

One of the most famous experiments on electricity occurred in 1752. In that year, Benjamin Franklin used a kite and a key to successfully draw lightning from the sky. He was trying to prove that electricity is a fluid. From these and other

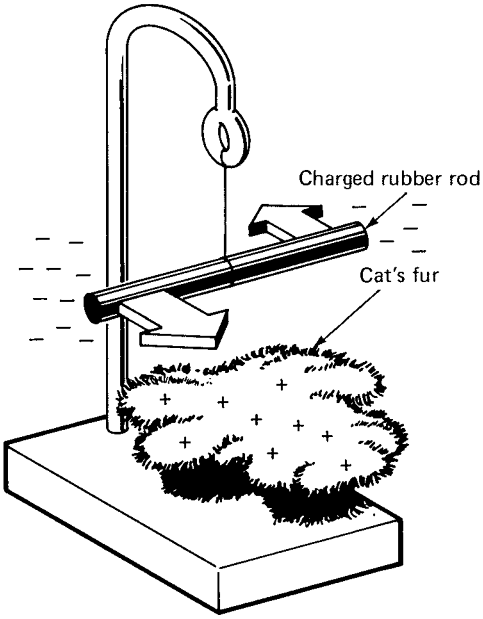

Figure 1-1 Unlike charges attract.

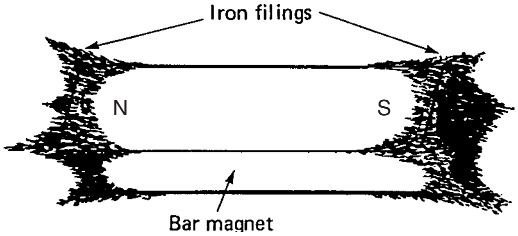

Figure 1-2 Iron filings cling to the ends of a permanent magnet. Note the north and south poles shown as N and S.

experiments, Benjamin Franklin is credited with forming the theory of positive and negative charges. It was another 80 years, though, before someone discovered that there is a relationship between magnetism and electron flow.

Electricity's Future

The effects of the knowledge and use of electricity were profound. Difficult tasks became easy. Old methods were replaced by new. Electrical machines relieved people of back-breaking labor. Machines could do the job better and cheaper than people. Fears that new inventions and methods would displace workers and create widespread unemployment did not prove true. Instead of loss of jobs,

Figure 1-3 George Ohm demonstrates his theories to some of his colleagues.

electricity led to new industries and new jobs. The new industries required more people than were replaced by machines.

Even today we hear about the possibilities of people losing their jobs because of machines and robots. Automation is the use of machines that are controlled by other machines and devices instead of people. It is another step in technical progress. It makes possible more things faster and better. Automation and robots create more jobs and a need for more skilled people. Trained people are needed to design, build, and maintain electrical equipment.

One of the greatest uses of electricity is in the production of ice and cooling for human comfort. Refrigeration and air conditioning rely exclusively on the ability of electricity to pump a fluid or gas through a system. Electricity is also used to control the temperature in heating, air-conditioning, and refrigeration systems.

Matter and Electricity

The name electricity implies the importance of the almost weightless, invisible part of an atom called an *electron*. It is electrons that cause electricity. Electricity is defined as the movement of electrons along a conductor.

An electron is only one part of an atom. An atom is only one part of a molecule. None of these can be seen by the unaided eye. Thus, most actions in electric circuits cannot be seen. An electric circuit can appear motionless although great activity is happening within it at the atomic level.

The electron can be controlled. Control of the electron is the task of an electrician, electrical engineer, or anyone else working with electricity. Electricity can perform work. It can kill. Using it requires knowledge of such things as matter and mass.

Matter surrounds us. It is said to be anything that occupies space. Thus, all physical objects are composed of matter.

Matter has mass. Mass is defined as the resistance an object offers to a change in motion. The tighter the matter is packed together, the greater is its mass. Thus, the greater is its resistance to any change in motion.

Solids, gases, and liquids

The three basic forms of matter, shown in Fig. 1-4, are solid, liquid, and gas. A solid, such as a glass container, is stable and self-supporting. By definition, a solid substance is one that offers a large resistance to forces that might change its shape. A liquid, such as water, maintains a definite volume, but assumes the shape of the container in which it is placed. A gas, such as the air we breathe, has no definite volume. It can be expanded or compressed to the shape or size of any container. The different forms of solid, liquid, and gaseous matter are called *substances*.

Pure water at room temperature is a liquid substance. All samples of pure water are identical. Pure iron is a solid, and pure carbon dioxide is a gaseous substance.

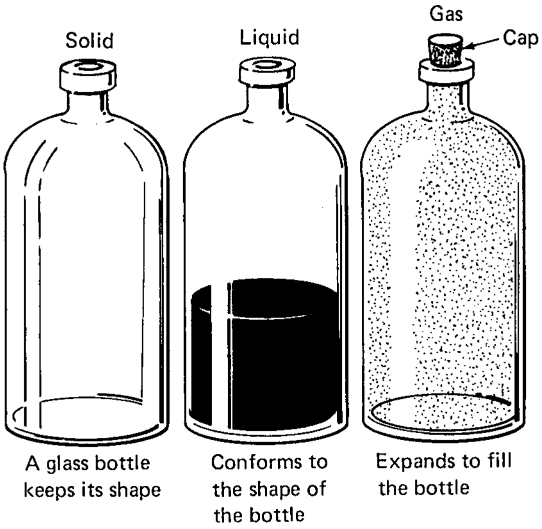
Solid
Liquid
Gas
Cap
A glass bottle keeps its shape
Conforms to the shape of the bottle
Expands to fill the bottle

Figure 1-4 Behavior of three forms of matter: solid, liquid, and gas.

The two classes of substances are elements and compounds. An element is a pure substance that cannot be divided into any more basic substances by chemical change. The elements (there are more than 100) are the simplest forms of matter. Some examples of elements are hydrogen, oxygen, germanium, silicon, gold, and silver. A compound is a substance composed of two or more elements chemically combined. Water, for example, is a compound made up of the elements hydrogen and oxygen. Common table salt (sodium chloride) is a compound made from the elements sodium and chlorine. A molecule is the smallest quantity of any compound that can exist and still retain all the properties of the compound. For example, the elements hydrogen and oxygen, when combined, produce the compound water. A molecule of water, however, is the smallest quantity of water that has all the characteristics of water. Elements and compounds can also form mixtures. A mixture is a mixing of two or more substances in which the properties of each substance are not changed. Water is a compound but salt and water is a mixture because the salt can be separated from the water simply by filtering or by evaporation of the solvent (water in this case).

Atom

An atom is the smallest part of an element that retains all the qualities of the element. Atoms are the building blocks from which all substances are made. Figure 1-5 shows two atoms of hydrogen gas combining to form one molecule of hydrogen gas. Some substances (compounds) are formed by the chemical combination of different elemental atoms. Water is one of these (see Fig. 1-6). A molecule of water consists of two hydrogen atoms and one oxygen atom.

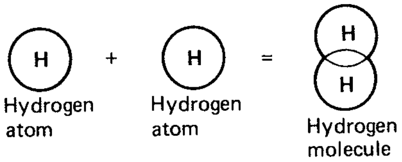

Figure 1-5 Two hydrogen atoms can be combined to form one molecule of gas.

An important part of an atom is the electron. An electric current is the result of the controlled movement of electrons in a substance.

Electrons, protons, and neutrons

The atom is the basic building block of matter, but it can be divided into many particles. The three major particles of the atom are electrons, protons, and neutrons. These three particles are important because they affect the electrical properties of the material. The protons and neutrons form the central mass or nucleus of the atom. One or more electrons circle the nucleus. The nucleus is almost 2000 times heavier than an electron. The electron is the smallest part known. It takes more than 28 billion, billion, billion electrons to weigh 1 ounce.

There is one nucleus in each atom. Elements differ because there are many combinations of orbiting electrons and groupings of protons and neutrons within the nucleus. Each element is made up of atoms having one particular combination of nucleus and orbital electrons. Each compound substance is made up of a particular arrangement of these atoms.

Another major part of the atom is the neutron. The neutron is also a part of the nucleus. However, it has no charge. Neutrons and protons have nearly the same mass. They determine, for the most part, the mass of the atom. The electron has little mass.

The simplest atom is the hydrogen atom (see Fig. 1-7a). It contains a nucleus and one electron. A helium atom (Fig. 1-7b) contains a nucleus and two electrons. In other atoms there is more than one shell of orbiting electrons. Copper has four shells. Some electrons have as many as seven electron shells. The fourth shell around the copper nucleus is made up of only one electron. It is easily

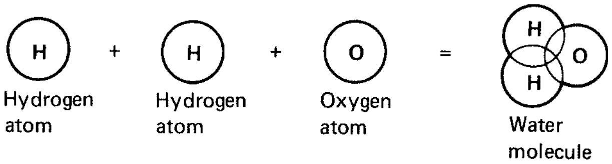

Figure 1-6 Two hydrogen atoms can be combined with one oxygen atom to form a molecule of water.

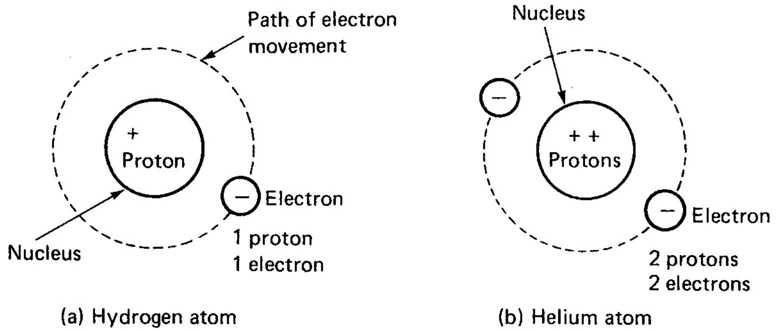
Path of electron movement
+
Proton
Electron
Nucleus
1 proton
1 electron
(a) Hydrogen atom
Nucleus
+ +
Protons
Electron
2 protons
2 electrons
(b) Helium atom

Figure 1-7 Simple atoms that contain one and two orbiting electrons. (a) Hydrogen atom. (b) Helium atom.

moved from one atom to the other by heat or magnetism to produce an electron flow or electric current.

The carbon atom and the copper atom are shown in Fig. 1-8. Notice that the carbon atom has only two shells, but has four electrons in the outer shell. The copper atom has only one electron in the outer shell. These two atoms are very important in electricity.

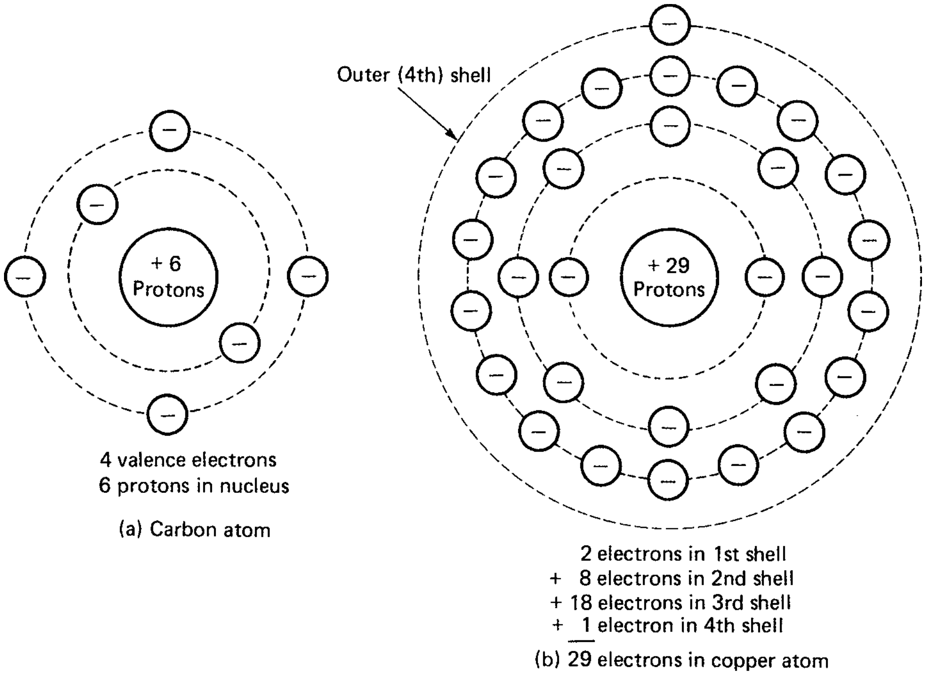
+ 6
Protons
4 valence electrons
6 protons in nucleus
(a) Carbon atom
Outer (4th) shell
+ 29
Protons
2 electrons in 1st shell
+ 8 electrons in 2nd shell
+ 18 electrons in 3rd shell
+ 1 electron in 4th shell
(b) 29 electrons in copper atom

Figure 1-8 Some atoms contain more than one shell of orbiting electrons. (a) Carbon atom. (b) Copper Atom.

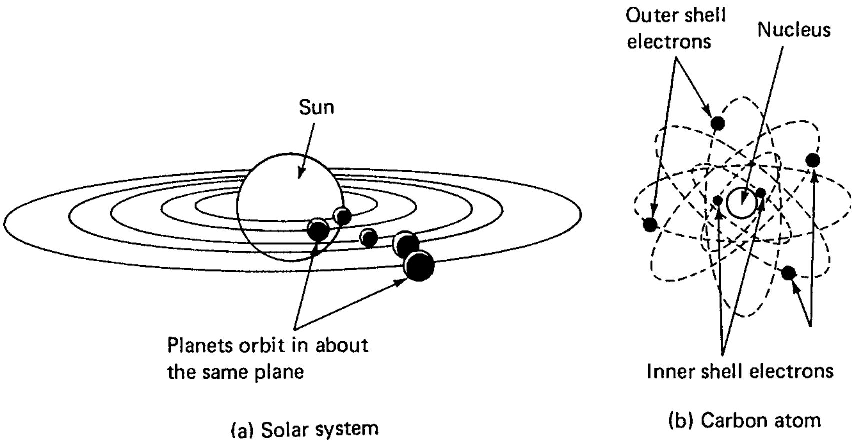

Figure 1-9 Comparison of on atom with the solar system. (a) Solar system. (b) Carbon atom.

A comparison such as the one shown in Fig. 1-9 is often made between our solar system and an atom. The nucleus of the atom is compared to the sun. Electrons revolving around the nucleus are compared to the planets revolving around the sun. A major difference between the two systems is the orbital paths of the planets and electrons. Figure 1-9 shows this difference. The planets have orbits in a fairly common plane during their trips around the sun. In contrast, the orbits of the electrons around the nucleus are constantly changing (this is present-day theory and subject to change later), and their paths eventually produce spherical shells around the nucleus. The arrangement of these spherical paths of the electrons and direction of their rotation around the nucleus determine the magnetic properties of the substance.

Properties of Electrons

The electrical properties of a substance are influenced by the number and arrangement of the electrons in the outermost shell. These electrons, located in the outer shell, are called *valence electrons*. Keep in mind that all electrons are alike. They are the same in all atoms. Electrons can be moved among like and unlike atoms. The application of an electrical force causes electrons to move from atom to atom in a controlled manner. The movement of electrons from atom to atom is called *electric current*. Because all electrons are the same, the basic atomic makeup of a substance (such as copper) is not changed by electron movement.

Orbiting Electrons

Orbiting electrons do not leave the atom. Orbiting planets do not leave the solar system. People can orbit the earth and return without being lost in space.

Two forces prevent the electrons from leaving the atom. One is the force or pull of gravity. It is the same pull that keeps things on earth. Similarly, there is a gravitational attraction between the nucleus and the electrons that causes it to be held in orbit around the atom. But this force is too weak. The main-force is the electrical attraction between the nucleus and the electron.

When a force or energy is applied sufficient to cause it to move from the orbiting path, it moves to the next atom and in moving can produce what we call electricity. This movement can cause work to be done.

Electrical Charge

Figure 1-10a illustrates the electrical charge. A ball is attached to a string and made to swing in a circle. The swinging ball tends to move away from the hand. But it is held by the string. This is like Fig. 1-10b in which the electrons swinging around an atom are pulled to the center nucleus. The speed of rotation causes them to follow an orbital path around the nucleus. The force between the electron and the nucleus is called an *electrical charge.*

The electron possesses a negative charge. The nucleus has the opposite polarity, a positive charge. In the nucleus, however, the positive charge is carried by protons. Thus, for every electron in orbit there is a proton in the nucleus. This is shown in Fig. 1-10b. There the atom has six electrons in orbit and six protons in the nucleus. The simple hydrogen atom has one electron in orbit and one proton in its nucleus.

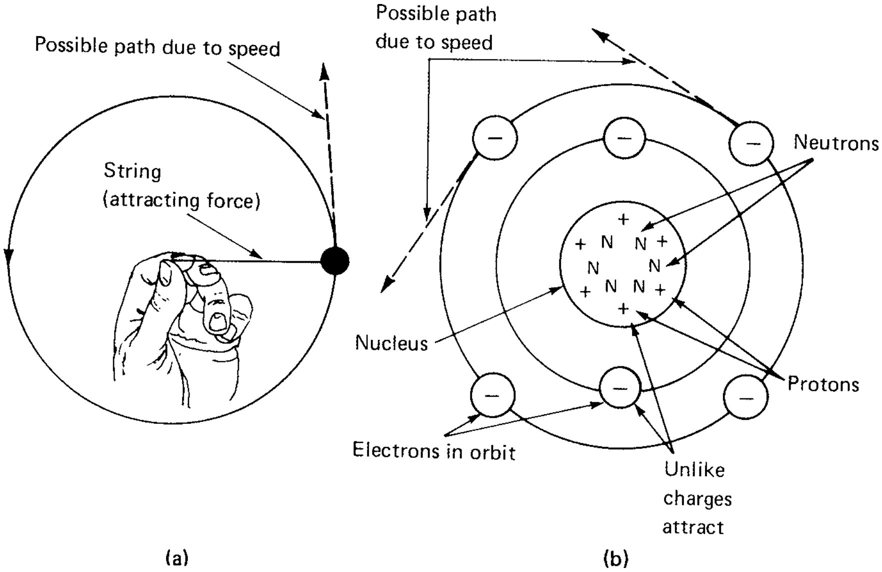

Figure 1-10 Comparison of a swinging ball attached to a string and an electron swinging around the nucleus of an atom. (a) Ball on a string. (b) Location of electrons.

Outer Shell

A basic law of electric charges is that like charges repel and unlike charges attract. The effect of charges on freely moving bodies is shown in Fig. 1-11. In the atom the positive charge of the nucleus (protons) attracts the electrons. However, the speed and energy of the electrons causes them to maintain their orbital paths. Since the forces in the atom are balanced, the electrical charges are balanced. Thus, the atom remains stable and neutral.

Electrons in the outer shell of an atom (called valence electrons) have a higher energy than electrons in the shells closer to the nucleus. External force can add energy to the valence electrons. This added energy permits their escape from the atom. Such free electrons can move from atom to atom. Being free, they are used for electric current.

Valence electrons and ions

Normally, we are concerned only with the valence electrons (outer shell) because these are the easiest to free. When one (or more) electron is removed from or added to the outer shell of an atom, the atom becomes charged. It is no longer neutral. Then the atom is called an *ion*. When an electron is lost, the atom

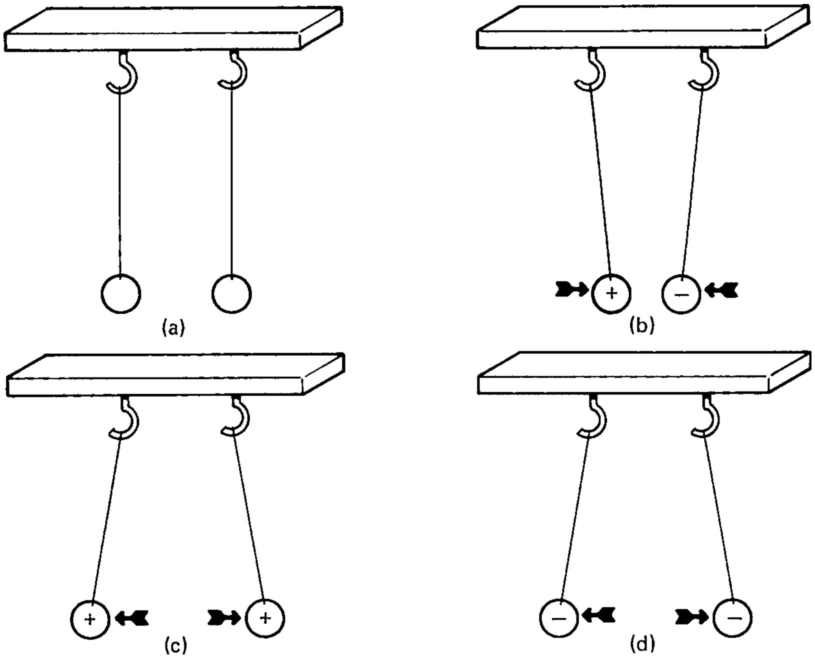

Figure 1-11 Unlike charges attract each other. Like charges repel each other. (a) No charge means there is no attraction or repulsion. (b) Positive and negative charges are attracted to each other (unlikes attract). (c) Positive charges repel each other (likes repel). (d) Negative charges repel each other, again (likes repel).

takes on a positive charge because there are more protons in the nucleus than orbiting electrons. The atom is then called a *positive ion*. Ionization of air is used in some "air-conditioning" systems.

Sometimes it is possible to add an electron to the outer shell. This results in a charged atom that is a negative ion. It is negative because there are more orbiting electrons than protons in its nucleus.

Unlike charges attract. Therefore, a positive ion will attract an electron or any negatively charged body. A negative ion, however, will repel an electron or any negatively charged body.

Unlike charges on two bodies mean there is a difference between them (see Fig. 1-12). A difference in charge exists between the four pairs of charged bodies in Fig. 1-13. This difference is 4. If a conducting path for electrons is made between any pair of bodies in Fig. 1-13, the same number of electrons would have to move from left to right in the illustration to neutralize the charges. When two bodies have the same charge and same polarity, there is no difference between them.

So far we have considered the charge in terms of electrons or numbers. Now we need to give the charge a name. A name is also needed for the difference that exists between these charges.

Any substance, molecule or atom, may have a negative or positive charge. Or it may be neutral. How much more negative or more positive can one body be charged relative to another? If a comparison is made, some unit of measurement must be used. Some standard reference should be used as a basic unit of measurement.

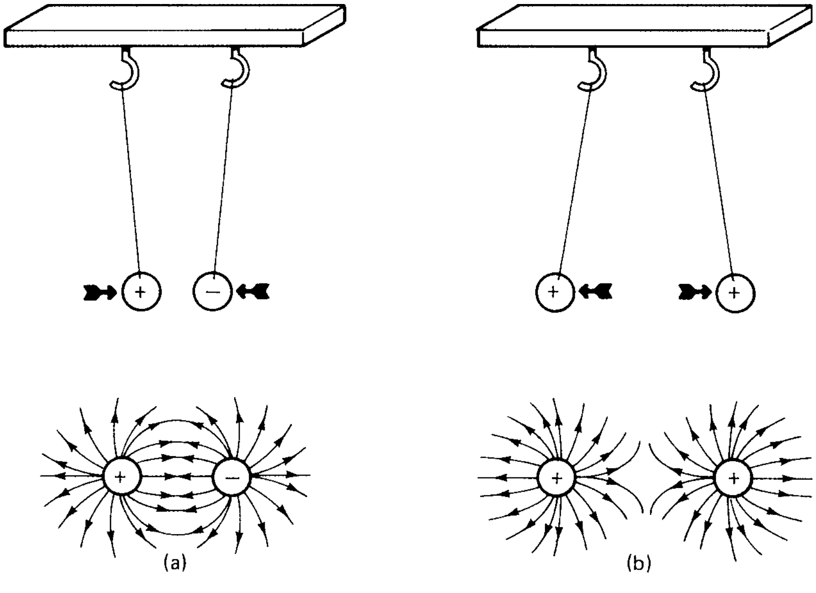

Figure 1-12 Invisible force fields extend outward from charged particles. (a) Lines of force unite and draw the unlike charges together. (b) Lines of force do not unite, so they repel.

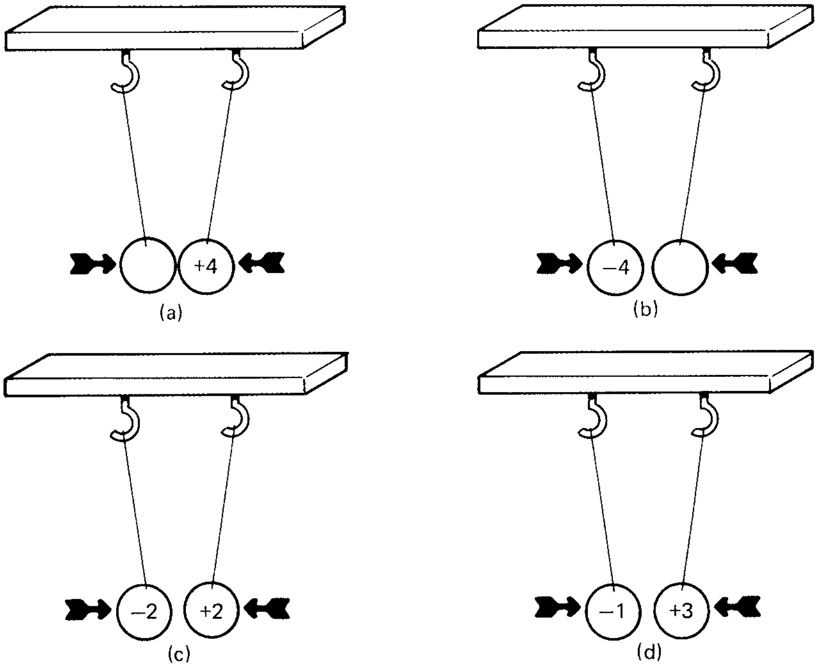

Figure 1-13 Four pairs of charged bodies with some difference of potential. Each pair of charges is attracted by the some amount of force.

The smallest negative charge is already understood to be that of the electron. And the charge of a proton is the smallest positive electric charge. Such charges are too small and not useful in terms of establishing a basic unit of measurement.

A Practical Unit of Charge

The practical unit of charge is the *coulomb*. It is the negative charge made by 6.25×10^{18} electrons. The term 10^{18} means it takes 6,250,000,000,000,000,000 electrons to produce a coulomb. Expressing it any other way than “6.25 times ten to the 18th” is awkward. (Some textbooks use 6.28 instead of 6.25. This is because 6.28 is 2π rounded off.)

The Volt

The volt is the unit for potential difference. It is used to indicate the electrical pressure or force needed to move coulombs of electric charge. The volt is also used to measure a unit of electromotive force (emf). The emf is the moving force behind an electric current. The volt is used and understood everywhere. The term *voltage* is often used to refer to potential difference.

Controlling Electrons

The controlled movement of electrons through a substance is called *current*. Current occurs only when a difference of potential is present. A good example of a difference of potential is observed by connecting battery terminals to a length of copper wire. The pressure from the battery moves the electrons.

Copper is a good path for current because of the relative ease with which electrons can be moved along its length. The one electron in the outer shell of the copper atom is free to move from atom to atom (see Fig. 1-8b). In fact, the electrons of copper drift in random fashion through the copper at room temperature (see Fig. 1-14a). If an imaginary line is set up in a copper wire, it will be found that the same number of electrons cross the line from both directions. This random movement does not produce an electric current. It takes a controlled movement of electrons to produce an electric current.

Difference of Potential (Voltage)

Electric current results when the movement of electrons is in one direction (see Fig. 1-14b). This is done by applying a difference of potential or voltage across the ends of the wire. One end of the wire attracts electrons because it is connected to the battery terminal that has a positive charge or lack of electrons. The electrons in the copper wire drift toward this positive charge. As electrons

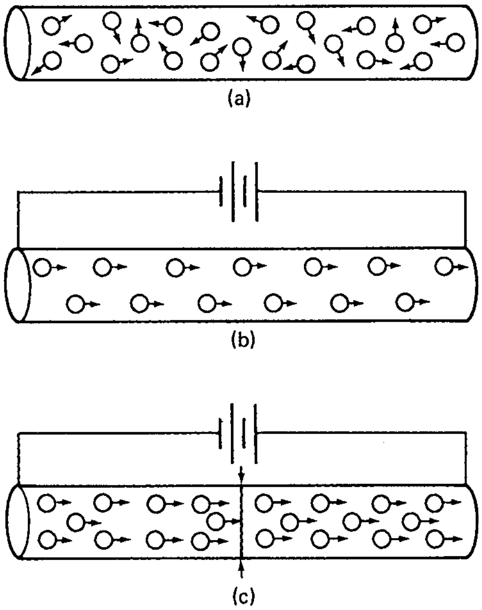

Figure 1-14 Current and the controlled motion of electrons. (a) Drifting electrons with no voltage applied. (b) The applied voltage controls the direction of electron flow. (c) The number of electrons past the line determines current flow.

leave the copper wire and enter the positive terminal, more electrons enter the other end of the copper wire. These electrons are taken from the negative terminal of the battery.

The difference of potential between the terminals of the battery is produced by a chemical reaction. When the chemical activity in the battery stops, the current stops.

Electron Flow (Current)

Current is the rate at which electrons move. If a point is established in the copper wire (see Fig. 1-14c), the current can be measured by the number of electrons that pass this point each second. Recall that a certain number of electrons is a coulomb. When a coulomb of electrons moves past the spot in 1 second, this amount of current is 1 ampere. *One ampere* represents 6.25×10^{18} electrons passing a given point in 1 second. The current is 3 amperes when 18.75×10^{18} electrons pass a given point in 1 second.

Conductors

A *conductor* is a material that allows electrons to move easily. Copper is a good conductor because it has an electron far away from the nucleus that can be easily forced out of orbit. When the electrons in a material cannot be moved as easily as in copper, the material is said to present a higher resistance to the motion of charges. Good conductors are said to have a low resistance; poor conductors (called *insulators*) have a high resistance. When a voltage is applied to a material of high resistance (an insulator), there will be fewer electrons in motion and less current than if the same voltage were applied to a material of low resistance.

Resistance

The ease with which electrons move in a material determines its resistance. A good conductor, such as copper, aluminum, or silver, has electrons that move freely. A low voltage will move a lot of electrons. A good insulator, such as glass, mica, or plastic, has electrons that do not move freely. Even a high voltage will move only a few electrons.

Resistance can have a wide range. It can be as low as that of a good conductor or as high as some good insulators. However, most resistances are somewhere in between good conductors and good insulators. The unit of measurement for resistance is the *ohm* (Ω). The ohm is defined as: One volt of pressure will push 1 coulomb of electrons through 1 ohm of resistance in 1 second.

Another way of saying it is that it takes 1 volt to push 1 ampere of electrons through 1 ohm of resistance.

The Electric Circuit

The workhorse of electricity is the *circuit*. It takes the electrons to where they belong or are needed. A complete circuit has a source of emf, a conducting path between the terminals of the power source, and a resistance, usually called the *load*. Note that all three elements, voltage, current, and resistance, are present in any complete circuit. And each has to be dealt with according to its presence.

The series circuit shown in Fig. 1-15 uses a battery, copper wire, and a light bulb. The battery produces the force needed to move the electrons. Chemical action in the battery makes the electrons available at the negative terminal. The copper wire is the path for the electrons to move along from the battery to the bulb. The copper wire is used because of its low resistance. Its resistance is less than 1 ohm. It is necessary to have a complete path from one terminal of the battery to the other for electrons to flow.

Electrons move only when there is a complete path between the two terminals. The second wire completes the path from the other end of the bulb to the positive terminal of the battery. This permits the electrons to return to the battery. The schematic for this circuit is shown in Fig. 1-16. This can be called a *closed* or *complete circuit*.

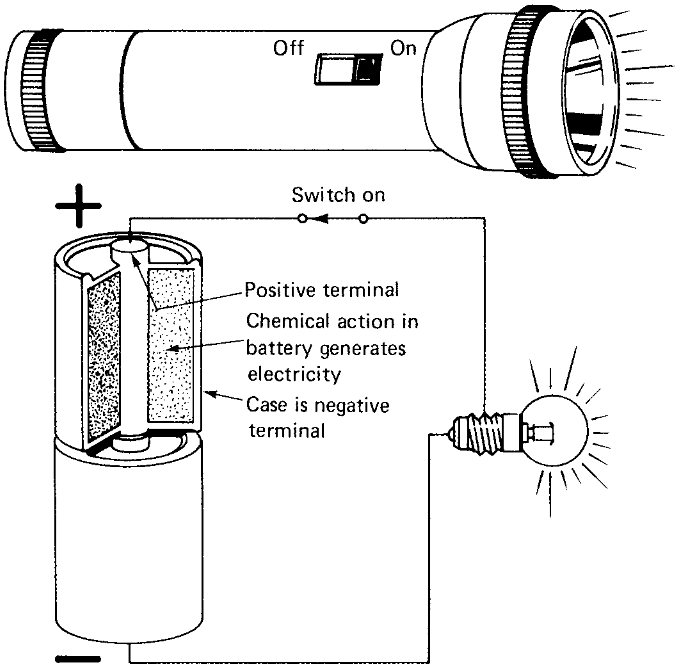
Off
On
Switch on
+
Positive terminal
Chemical action in battery generates electricity
Case is negative terminal
−

Figure 1-15 A simple series circuit has a battery and a bulb connected by two lengths of copper wire.

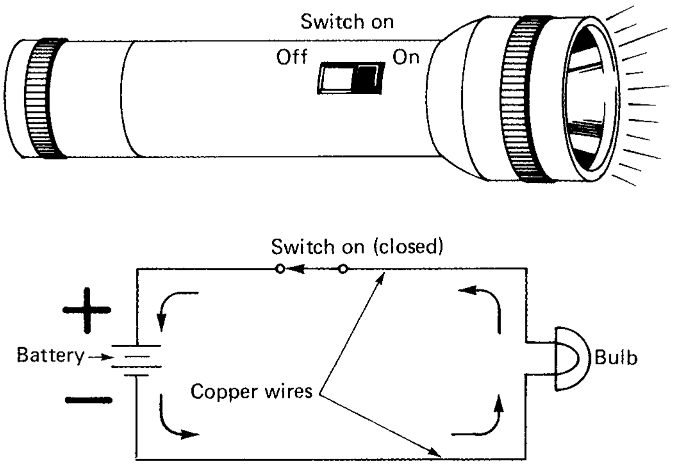

Figure 1-16 Schematic for connecting a bulb and a battery.

Connecting a Circuit

When a circuit is connected to a battery, there is a negative charge at the negative terminal of the battery. Thus, an excess of electrons is located at this point. There is a positive charge at the positive terminal of the battery. Thus, there is a deficiency of electrons at this point. Electrons flow out of the negative terminal into the copper wire. This causes a movement of the electrons in one direction. Electrons also move in one direction in the bulb. The copper wire, connected to the other end of the bulb, conducts the electrons to the positive terminal of the battery. The arrival of the electrons at the positive terminal should end the movement of electrons. However, the chemical action in the battery maintains an emf across the battery terminals, and electrons continue to flow.

The light bulb has a resistance high enough to convert the electrical energy to light and heat energy. The electric current heats the filament of the bulb. It glows brightly. Keep in mind that light bulbs do not change all of the electrical energy into light. As you know, the light bulb also becomes hot. This means that much of the electrical energy is changed to heat.

Chemical action in the battery supplies the electrical energy to the light bulb. The chemical action in the battery has a limited lifetime. Eventually, the chemical action stops because the material in the battery is used up. When this happens, the light goes out because the battery is discharged.

Switches Control Electron Flow

Electron flow can be stopped by opening the circuit at any point. One of the wires can be removed from the battery terminal or from the bulb. In most electric circuits, a switch is used at some point to permit the electron movement to be started and stopped when needed. Figure 1-17 shows a switch placed in

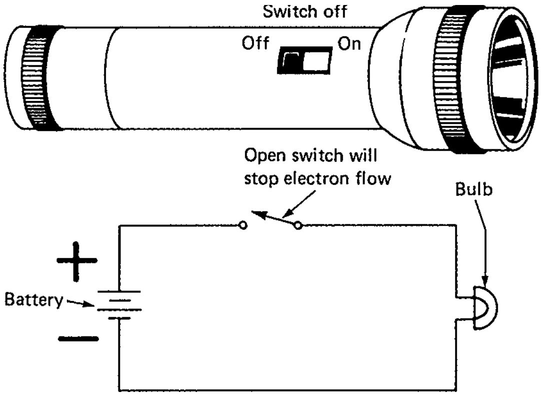

Figure 1-17 A switch is added in series with the bulb and battery to open and close the circuit.

series with the battery and the bulb. The switch can be placed at any point in the circuit.

Schematic

The *schematic* of Fig. 1-17 shows the switch in the open position. A schematic is a shorthand way of drawing a circuit using symbols. This switch can open and close the circuit without moving wires. This simple circuit diagram shows how a flashlight is wired.

Review Questions

1. What is matter?
2. When two or more substances are mixed, what is the result called?
3. What is the result of chemically combining two or more elements?
4. Do all substances contain elements?
5. Which electrons in an atom affect its electrical characteristics?
6. All electrons are alike. (True or False)
7. The atom is made up of a nucleus and one or more ________.
8. What kind of charge does an electron have?
9. What happens to unlike charges?
10. What is the name given to an atom with a negative or a positive charge?
11. What type of electrons make up an electric current?
12. Define the coulomb.
13. Define the unit of measurement of electrical pressure.

14. What is another name for potential difference?
15. What kind of resistance does a conductor have?
16. One coulomb per second indicates what unit of resistance?
17. The unit of measurement for resistance is the _________.
18. Define electric current in terms of electrons.

Chapter 2

Current, Voltage, Resistance, Power, and Ohm's Law

Performance Objectives

- Understand the five ways electricity is produced.
- Understand how the units of measurement for electricity were developed and used. Be able to work Ohm's law problems.
- Understand how volts, ohms, and amperes are related and function in any circuit.
- Understand how to work electrical power problems.
- Understand the terms current, voltage, resistance, power, and Ohm's law.

Sources of Electricity

The five most important sources of electricity for technicians are chemical action, heat, light, pressure, and magnetism.

Chemical action

In the electrical and electronics fields, many sources of electricity are used. In the circuit of Fig. 2-1, the battery is the source of electricity. Batteries produce electrical energy by a chemical action.

Heat

Heat can be used to free electrons from some metals and from specially prepared surfaces. When some materials are heated to a high temperature, electrons are freed from their surfaces. Any nearby metallic surface, if positively charged,

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

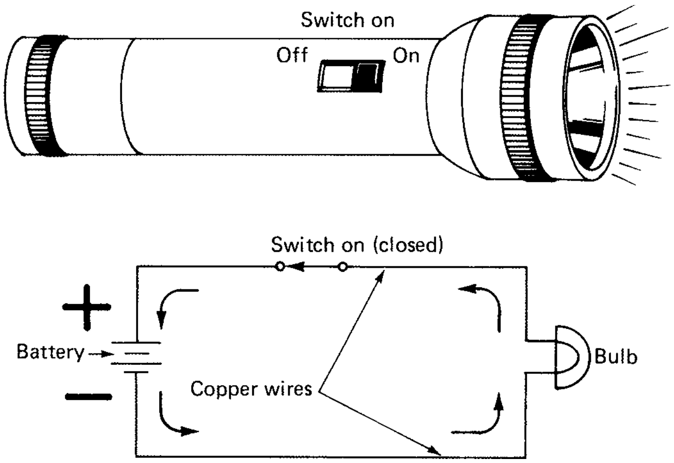

Figure 2-1 A battery produces current that flows in only one direction. Electrons move from negative to positive in the circuit, but positive to negative inside the battery.

attracts these electrons and produces electron flow. The freeing of electrons by heat is called *thermal emission*.

Light

Light striking the surface of certain materials can be used to free electrons. This is called *photoemission*. With a suitable collecting surface, useful electron flow can result. Photoemission is used in photoelectric devices and television camera tubes.

Pressure

Mechanical pressure on certain crystals can be used to produce electricity. The crystal cartridge of an inexpensive record player is a good example. The needle causes a changing pressure. The crystal produces a changing voltage in step with the grooves in the record.

Magnetism

The most common method of generating electrical power is by turning a coil of wire in a magnetic field. This is the method the power station uses to generate the electric power that is used in homes, business, and industry.

Electricity is generated in two forms: direct current (dc) and alternating current (ac). A battery supplies dc electricity because the electrons flow in only one direction, from the negative terminal to the positive terminal, as shown in Fig. 2-1. A battery supplies a constant voltage of one polarity.

The ac generator is shown in Fig. 2-2. It develops a voltage across its output terminals that changes polarity and amplitude as the coil is turned. At one time, the electrons move in one direction (positive polarity). A fraction of a second

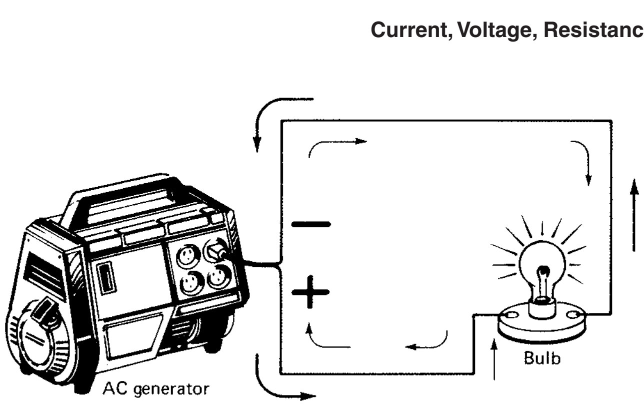

Figure 2-2 The output from an ac generator (alternator) produces a current that reverses direction at regular intervals. The current in the circuit alternates.

later, the electrons move in the other direction (negative polarity). These changes in current direction are called a *hertz*. The number of hertz completed in 1 second is called the *frequency*. In the United States, 60-hertz power is used. Some other countries use 50-hertz power. (Hertz is the newer term for cycles per second.)

When an electric circuit such as the wires and bulb shown in Fig. 2-2 is connected across the output of the generator, the direction of electron movement depends on the polarity of the voltage. When the polarity is positive, the electrons flow from the negative terminal through the load to the positive terminal. A fraction of a second later, the output voltage reverses. The electrons now flow in the direction shown by the heavy black arrows. The electrons in the circuit flow first in one direction and then in the other. Thus, the electron flow alternates. Since the current in the circuit changes direction, the voltage will also change direction. This is shown by the change in polarity (plus and minus) at the generator terminals.

There are advantages to using alternating current. One major advantage is that ac is easier to generate and distribute than dc. But many electronic circuits require both ac and dc. Such devices use power supplies to convert the ac electricity to the dc used in the circuit.

Units of Measurement

The importance of units of electrical measurements cannot be overemphasized. Electrical measurements are useful only if some standards of measurement exist. Let's review the electrical units.

Unit of charge: coulomb

The electrical charge on one electron has such a small value that it is not practical to measure it directly. A practical standard has been set up that says the unit of charge will be the total of 6.25×10^{18} electrons. This unit of charge is the *coulomb* (C). It is a basic measure of a quantity of electrical charge. Electron movement can also be measured in terms of coulombs if we include time as a second value. The electrical work performed by moving electrons depends on charge movement per unit of time. For example, a charge movement of so many coulombs per second is a rate of electron flow. This is a measurement of current or electricity at work. Remember, 1 coulomb per second is 1 ampere when it moves past a given point in 1 second.

Unit of current: ampere

A practical unit for measuring current is the *ampere*. It is used in place of the term coulomb per second. One ampere of electrical current is defined as the movement of 1 coulomb of electrons past a given point in 1 second. The ampere is measured with an ampere meter. This meter is called an *ammeter*.

Unit of potential difference: volt

The force that moves electrons is called a *difference of potential*, *emf*, or *voltage*. The unit of electrical potential is the *volt* (V). The volt is defined as the electrical force needed to produce 1 ampere of current in 1 ohm of resistance. The volt is measured with a voltage meter. This meter is called a *voltmeter*.

Unit of resistance: ohm

The opposition that a material presents to the flow of electrical charges is called *resistance*. The unit for electrical resistance is the *ohm* (Ω). By definition, the ohm represents the resistance a material offers to 1 ampere when 1 volt of electrical potential is present across the material.

The basic relation of units in an electric circuit is shown in Fig. 2-3. If there are 2 volts present across the 2-ohm resistance, there will be a current of 1 ampere. If the resistance remains the same and the voltage across the resistance is increased, the current will rise. Figure 2-4 shows the effect of increasing the voltage across the 2-ohm resistance. If the voltage stays at 2 volts, but resistance to the flow of electrons is increased to 4 ohms, there will be a decrease in current. Figure 2-5 illustrates this. Note that the ammeter indicates to 0.5 ampere.

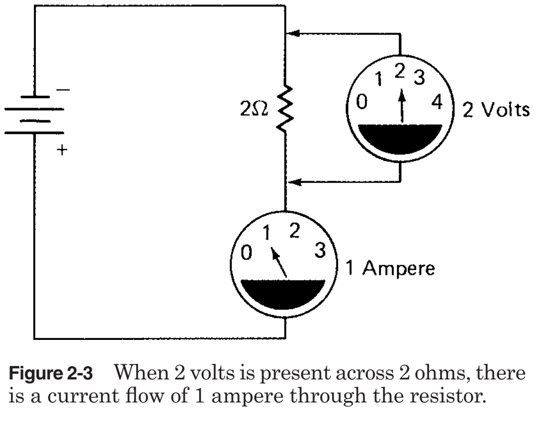

Figure 2-3 When 2 volts is present across 2 ohms, there is a current flow of 1 ampere through the resistor.

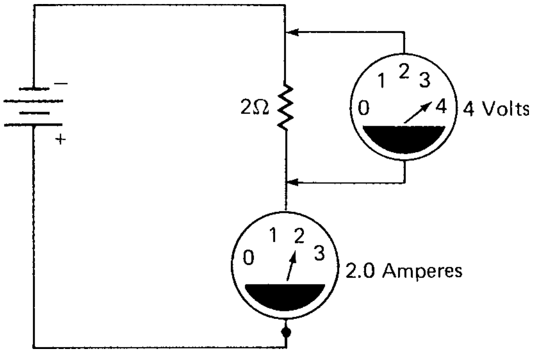

Figure 2-4 Changing the voltage to 4 volts across the 2-ohm resistor results in 2 amperes of current.

Siemens

The term conductance is also used often. It is the opposite of resistance. It is a term used to explain how well a wire or other substance conducts electricity. The unit of conductance is the *siemens* (S). The mho (ohm spelled backward) is the older term. Both are used in the literature, so it is best to be aware of both terms.

Prefixes

The ohm, volt, and ampere are units intended for practical use. For many uses, these basic units are either too large or too small. In such cases a system of prefixes is commonly used to make it easy to work with currents, voltages, and resistances.

Several of the more common prefixes are given in Tables 2-1 and 2-2. For example, a resistance of 10 million ohms can be written as 10,000,000 (10 × 1,000,000), 10 megohms, or 10×10^6 ohms. A current of 0.000003 ampere can

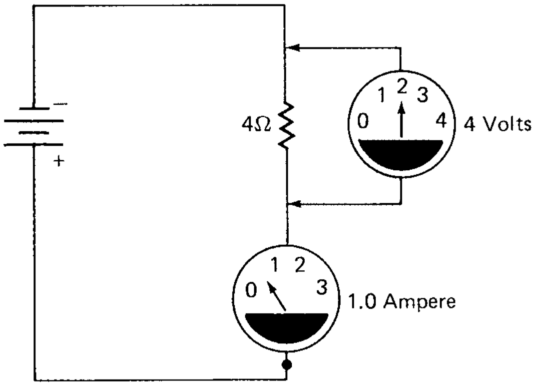

Figure 2-5 If 2 volts is present across 4 ohms, the current is 0.5 ampere.

TABLE 2-1 Common Prefixes and the Values They Represent

Prefix	Abbreviation	Multiply by	Tens power	Value
Mega	M	1,000,000	10^6	Million
Kilo	k	1,000	10^3	Thousand
Milli Thousandth	m	0.001	10^{-3}	
Micro	μ	0.000001	10^{-6}	Millionth

be represented as 3 microamperes or 3×10^{-6} ampere. A difference of potential of 5000 volts can be represented as 5 kilovolts or 5×10^3 volts.

These prefixes are not limited to the ohm, volt, or ampere. They are used frequently with other electrical units or values.

Ohm's Law

Electric current is a flow of electrons. Electrons flow when a voltage is connected to a conducting path. The amount of current depends on the amount of voltage and the value of the resistance. Georg Ohm discovered the relationship between the three factors: voltage, current, and resistance. The exact values can be determined mathematically using Ohm's law.

Current is the result of an applied electrical force. The greater the applied electromotive force or voltage, the greater will be the current or amperes in a given circuit. Thus, the voltage and current are directly related. In other words, *an increase in voltage causes an increase in current. A decrease in voltage will cause a decrease in current, if the resistance remains the same*.

The amount of resistance in the circuit also determines the amount of current that will flow. The lower the resistance (measured in ohms) in the circuit, the higher is the current (measured in amperes). Thus, current and resistance are inversely related. In other words, if one goes up the other goes down, and vice versa. Once you understand this concept, you also have a basic knowledge

TABLE 2-2 Prefixes Used with Volts, Amperes, and Ohms

Prefix	Volts (V)	Amperes (A)	Ohms (Ω)
M	1 megavolt (1 MV)		10 megohms (10 MΩ)
k	5 kilovolts (5 kV)		5000 ohms (5 kΩ)
m	10 millivolts (10 mV)	5 milliamperes (5 mA)	
μ	4 microvolts (4 μV)	3 microamperes (3 μA)	

of Ohm's law. Ohm's law states that the current is equal to the voltage divided by the resistance.

$$\text{Amperes} = \frac{\text{volts}}{\text{ohms}}$$

Usually, symbols are used for the units to form the equation:

$$I = \frac{E}{R}$$

where I = current in amperes (A)
E = voltage in volts (V)
R = resistance in ohms (Ω)

Keep in mind as you work Ohm's law problems that the basic units are the ampere, volt, and ohm. If you run into microamperes or milliamperes, they have to be converted to their decimal equivalent of the ampere. If you run into kilovolts, that unit has to be converted to volts. The same is true with resistance, which may be stated as kilohms or megohms. You have to convert them to ohms before using the formula to obtain the correct answer.

Ohm's law examples

1. What is the current in amperes when the voltage applied is 100 volts and the resistance in the circuit is 50 ohms?

$$I = \frac{E}{R} = \frac{100}{50} = 2 \text{ amperes}$$

Table 2-3 shows the results of changes in voltage and current in a circuit having a fixed resistance of 50 ohms. Our answer, 2 amperes, is circled in the second column of Table 2-3, beside the 100-volt value used in the example.

TABLE 2-3 Relationship between Voltage and Current When the Resistance Is Held at 50 Ohms (Ω) and the Voltage Is Changed

		Resistance (R) = 50 ohms	
Volts E	Amperes I	Voltage change	Current change
400	8	4×	4×
350	7	3.5×	3.5×
300	6	3×	3×
100	②		
75	1.5	0.75×	0.75×
50	1	0.5×	0.5×
25	0.5	0.25×	0.25×

$$I = \frac{E}{R}$$

If the voltage is doubled (increased from 100 to 200 volts), and the resistance remains fixed at 50 ohms, what happens to the circuit current?

$$I = \frac{E}{R} = \frac{200}{50} = 4 \text{ amperes}$$

When the voltage is doubled, or increased from 100 to 200 volts, and the resistance remains fixed at 50 ohms, the current is doubled. This is shown in Table 2-3 in the columns headed Voltage Change and Current Change. On the other hand, when the voltage is halved or decreased from 100 to 50 volts, what happens to the current?

$$I = \frac{E}{R} = \frac{50}{50} = 1 \text{ ampere}$$

When the applied voltage is decreased to half the original value, the current also decreases to half its original 2 amperes.

The direct relationship between voltage and current has been shown in this example. For a fixed amount of resistance, when the voltage goes up, the current goes up. When the voltage goes down, the current goes down. Thus, the current varies directly in step with the voltage.

The next step in understanding Ohm's law is to vary the resistance and keep the voltage constant.

2. What is the current when the resistance is changed to 75 ohms and the applied voltage is held at 100 volts? See Table 2-4.

$$I = \frac{E}{R} = \frac{100}{75} = 1.3333 \text{ amperes}$$

The resistance in the third column of Table 2-4 is 1.5 times the original value. The current in the fourth column is two-thirds of the original value. These values are circled in Table 2-4.

TABLE 2-4 Relationship between Resistance and Current When the Voltage Is Held at 100 Volts and the Resistance Is Changed

		Voltage (E) = 100 volts	
Ohms R	Amperes I	Resistance change	Current change
150	1.500	3×	
100	1.000	2×	
75	1.333	1.5×	0.666×
50	2.000		
25	4.000	0.5×	2×
16.666	6.000	0.333×	3×
12.500	8.000	0.250×	4×
		$I = \frac{E}{R}$	

What is the current value for each of the resistances shown in the first column of Table 2-4? Fill in the fourth column (I) of the chart. Several answers have been filled in as examples.

Complete the third column of the chart to show the amount by which the resistance is increased or decreased. Fill in the fourth column to show the amount by which the current changes. Note that the current decreases by the same ratio that the resistance increases, and the current increases by the same ratio that the resistance decreases. In these simple problems, you have been shown that the current varies inversely with the resistance.

Ohm's law and the closed circuit

It is important to have a grasp of the relation between current, voltage, and resistance. Ohm's law is the key to understanding these relationships. Ohm's law is a tool that can be used to determine what will happen in a circuit before it is turned on. A simple circuit is shown in Fig. 2-6. What happens in the circuit if a battery with a higher voltage is used? Simple, the current increases.

Assume that you have connected the circuit, measured the current, and have found it to be low. What do you know about the circuit? Low current means that

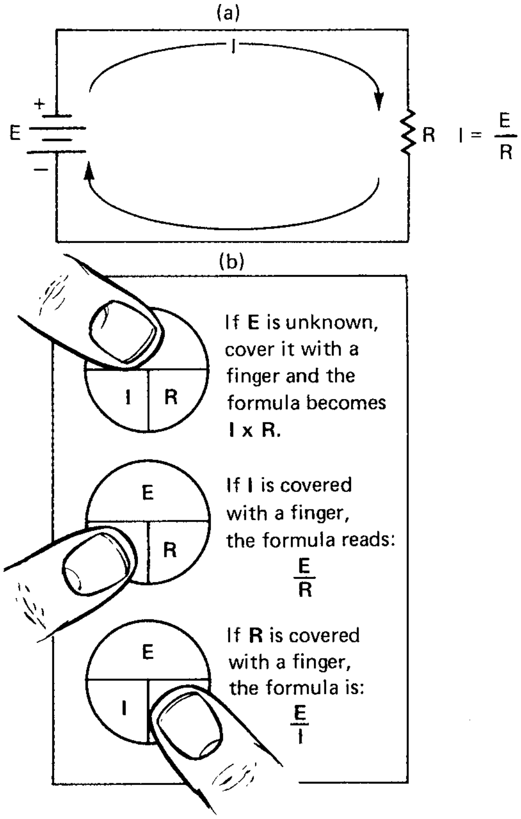
(a)
I
+
E
−
R
I = E/R
(b)
I
R
If E is unknown, cover it with a finger and the formula becomes I x R.
E
R
If I is covered with a finger, the formula reads: E/R
E
I
If R is covered with a finger, the formula is: E/I

Figure 2-6 (a) The use of Ohm's law to plan what happens in a circuit. (b) Using your finger to find the right formula to use.

the circuit resistance is high or the battery voltage is low. You know this because you know current is affected by voltage and resistance.

Suppose you want a certain current in a circuit (such as Fig. 2-6). What can be done to produce this amount of current? You can change to a battery that has a different voltage. Keep in mind that a different voltage will produce a different current. Or you can change the resistance because a change in resistance also produces a change in current.

3. Determine the amount of current in a circuit when the resistance is 25 ohms and the voltage is 50 volts. Use Ohm's law to solve the problem:

$$I = \frac{E}{R} = \frac{50}{25} = 2 \text{ amperes}$$

The answer is 2 amperes. It has been entered in the table of Fig. 2-7. Complete the table of values using 50 volts for each calculation. Place the values of voltage and current from Table 2-5 on the graph of Fig. 2-7.

The point for 2 amperes at 25 ohms is shown on the graph. To place a point, find the value for 25 ohms across the bottom of the graph. The value 25 is marked by an arrow. Find the value for 2 amperes on the left side of the graph. Moving up from the 25 and across from the 2 will locate the point shown on the graph. Locate a point for each of the ohm and ampere values in Table 2-5.

The points you have plotted are now connected. Mark the line 50 volts. The line can now be used to find current and resistance values for a voltage of 50 volts.

If you should need to know the current when 50 volts is applied to a resistance of 12.5 ohms, you can find it by using this line. If you divide 50 by 12.5, you get 4 amperes. That is using Ohm's law for finding the value: $I = E/R$. The

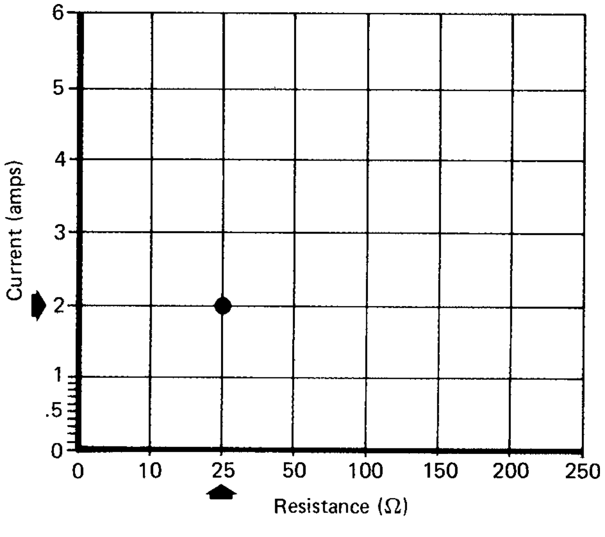

Figure 2-7 Graph to show the relationship of current, voltage, and resistance.

TABLE 2-5 Fill in the Values for Amperes Using Ohm's Law

$E = 50V \quad I = \frac{E}{R}$

R (ohms)	I (amperes)
10	
25	2.000
50	1.000
50	
100	
150	
200	
250	

graph gives you the same answer. Electronics uses graphs for many functions. Get used to graphs for you will be seeing many of them.

Ohm's law in other forms

It is not always necessary to find the current in a circuit. If an ammeter is handy, it will give you this information. To be able to find the resistance or voltage in a circuit when one or the other is missing, it is possible to use Ohm's law to obtain the missing value. Keep in mind that $R = E/I$ and $E = I \times R$. These are nothing more than the other forms of Ohm's law. They come in handy when working with electricity.

If you want to find resistance when voltage and current are given, you use the formula $R = E/I$. For instance, if you have a 120-volt circuit that has 4 amperes flowing in it, the resistance can be found by dividing the voltage (120) by the current (4) to give the result of 30. This 30 is 30 ohms of resistance needed to cause the circuit to draw 4 amperes. Check Table 2-6 and fill in the blanks.

TABLE 2-6 Chart Showing the Relationship between Voltage and Resistance for a Current of 4 Amperes

	I = 4 amperes
E (volts)	R (ohms)
600	
480	
360	90
240	
120	30
60	
40	
30	7.5
24	6.0

$R = \frac{E}{I}$

TABLE 2-7 Chart of Current and Voltage for a Constant Resistance of 30 Ohms

R = 30 ohms	$E = I \times R$
I (amperes)	E (volts)
16	
12	
8	
4	120
2	
1.333	
1	30

The relationship between voltage and current is evident when you use the resistance of 30 ohms and a higher current is found in the circuit. Note when you use $E = I \times R$ that the voltage is directly related to the current. If the current is increased and the resistance stays fixed, the voltage has to go up or increase.

If you have 4 amperes of current and 30 ohms of resistance you use the formula: $E = I \times R$ and get $E = 4 \times 30$ or 120 volts. Place this value in Table 2-7. Now, see if you can fill in the second column of the chart in Table 2-7.

Uses for ohm's law

It is sometimes awkward to express values in basic units because many zeros or decimals are needed. For this reason, several prefixes and symbols are used. These are shown in Table 2-8. These symbols are often used in making Ohm's law calculations. They are also used for electronic terms such as frequency, inductance, and capacitance, which you will encounter in later chapters.

Most circuits have the voltage stated in volts. High-voltage cross-country transmission lines and television picture tube voltages are usually given in kilovolts or thousands of volts. Kilo (k) means 1000. Thus, 138,000 volts can be written as 138 kilovolts and abbreviated to 138 kV.

TABLE 2-8 Prefixes and Multipliers for *E*, *I*, and *R*

Prefix	Symbol	Multiple to get basic unit	Multiplication factor E	I	R
tera	T	1,000,000,000,000	10^{12}		
giga	G	1,000,000,000	10^{9}		
mega	M	1,000,000	10^{6} megavolts		megohms
kilo	k, K	1,000	10^{3} kilovolts	kiloamperes	kilohms
hecto	h	100	10^{2}		
Basic Units		*1*	10^{0} *Volt*	*Ampere*	*Ohm*
centi	c	0.01	10^{-2}		
milli	m	0.001	10^{-3} millivolts	milliamperes	
micro	µ	0.00001	10^{-6} microvolts	microamperes	
nano	n	0.000000001	10^{-9} nanovolts	nanoamperes	
pico	p	0.0000000000001	10^{-12}		
femto	f	0.000000000000001	10^{-15}		
atto	a	0.000000000000000001	10^{-18}		

Some circuits, especially transistor circuits, use the terms milliamperes and microamperes. One milliampere (mA) is 0.001 ampere or one-thousandth of an ampere. The microampere unit is even smaller. One microampere (μA) is 0.000001 ampere or one-millionth of an ampere. These units have to be converted to basic units before they are used in Ohm's law formulas.

Power

Power can be understood when you take the knowledge you have and use it with Ohm's law formulas and examine them in terms of the work done. The formulas are modified slightly to produce the answers needed when figuring the work done by electricity in a circuit.

Work is done when you lift a 100-pound object for a distance of 10 feet. Power is the rate of doing work, or foot-pounds per second. If you take 20 seconds to lift the 100 pounds 10 feet, then you can find the amount of work done in foot-pounds. Power is equal to foot-pounds divided by time. In this case you have lifted 100 pounds for a distance of 10 feet. Thus, you multiply the 100 by 10 to produce 1000. This 1000 is divided by the time it took to do the job (20 seconds) to produce 50 foot-pounds of work.

If you want to convert foot-pounds to horsepower, a term you are more familiar with, you keep in mind that 1 horsepower is equivalent to 550 foot-pounds per second, or lifting 550 pounds for 1 foot in 1 second.

The 50 foot-pounds is about 0.0909 horsepower, or it can be said in fractional form as 1/11th of a horsepower.

Joule

Electrical work (joule) is done when force (E) causes the movement of an electrical charge (Q). This can be stated as joule $= E \times Q$.

The unit of electrical work is the *joule*. One joule represents the work done by a difference of potential of 1 volt (E) while moving 1 coulomb (Q) of charge (6.25×10^{18} electrons).

Power (W for watts) in an electric circuit is the rate of doing electrical work (volt-coulombs per second).

As an equation, it becomes

$$\text{Power (watts)} = \frac{\text{volts} \times \text{coulombs}}{\text{time}}$$

or

$$P = \frac{EQ}{t}$$

where P = power
E = voltage
Q = coulombs
t = time

Watt

The unit of electrical work is the *watt* (W). The letter *P* is used to represent power. One watt is equal to 1 joule per second. Thus, 1 joule and 1 watt are the same and you can convert back and forth between joules and watts easily. It takes 746 watts of electrical energy to produce 1 horsepower (hp).

$$1 \text{ horsepower} = 746 \text{ watts}$$

Putting Electricity to Work

Many jobs can be done by electricity. One useful conversion is from electricity to light. We depend on lights for the home, automobile, street lights, and flashlights. The most common conversion from electrical energy is to heat. An electric toaster or an electric iron such as in Fig. 2-8 provides useful forms of heat energy. Heat is produced in most electrical conversion devices. In some, such as the light bulb, heat is an unwanted result.

Resistors and heat

Resistors, such as those shown in Fig. 2-9, are used in electric circuits. They act as loads, voltage dividers, and current-control elements. However, in doing their job, heat is produced. For this reason, electric circuits must be ventilated to carry away the heat. Also, a resistor must be able to handle the power so that it does not overheat and burn.

Mechanical energy

Electrical energy is often converted to mechanical energy. The electric motor shown in Fig. 2-10 is a good example of a device used to convert electrical to mechanical energy. Another example is a relay, which is a switch that is operated electrically. When electrical energy is applied to the relay, a movable metal

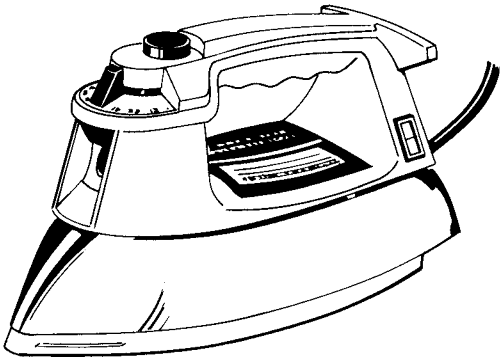

Figure 2-8 The electric iron is an example of a device that converts electrical energy into heat.

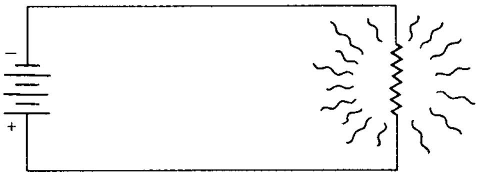

Figure 2-9 A resistor placed across a voltage source gets hot due to the current in the resistor.

arm turns on or off another electric circuit. A relay uses a low-voltage, low-current source to control high voltage or high current from a remote location. Furnaces, air conditioners, and refrigerators all use relays.

Many electrical devices are rated in terms of power and operating voltage. It is often necessary to find the current that results when the device is connected to a certain voltage source. At other times the power and the current are known, and it may be helpful to find how much voltage is needed to operate the device. Both of these problems can be solved by using one of the forms of the basic power law:

$$P = E \times I$$

$$P = I^2 \times R$$

$$P = \frac{E^2}{R} \quad \text{or} \quad P = E^2/R$$

You can also convert these formulas to

$$I = P/E \quad \text{or} \quad I = \frac{P}{E}$$

$$E = P/I \quad \text{or} \quad E = \frac{P}{I}$$

Or you can use $I^2 = P/R$ and convert it to

$$I = \sqrt{P/R} \quad \text{or} \quad I = \sqrt{\frac{P}{R}}$$

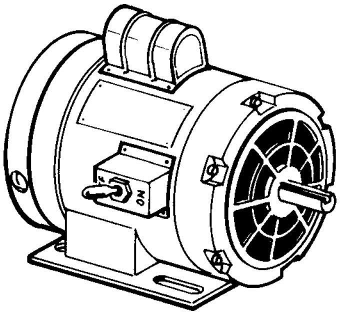

Figure 2-10 A motor is often used to convert electrical energy to mechanical motion.

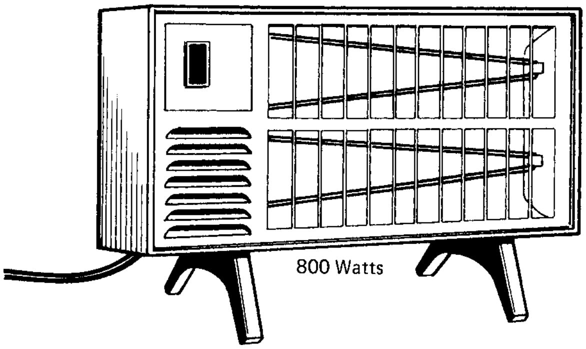

Figure 2-11 An electric heater has low resistance and high current.

A good example of how this formula can be used is found in the following problem: An 800-watt electric heater has a resistance of 2 ohms. What is the current needed to operate the heater?

First, select the formula needed. This is determined by the known values. In this case, the power and the resistance are known. You need to find the current. So, you select $I = \sqrt{P/R}$ for a formula, substitute the values in the formula, and solve.

$$I = \sqrt{P/R} \quad \text{or} \quad \sqrt{800/2} \quad \text{or} \quad \sqrt{400} = 20 \text{ amperes}$$

This is illustrated in Fig. 2-11.

Knowing that the required current is 20 amperes, you can determine the size of the wire needed to cause the heater to operate properly. You look up the current in a chart and find that No. 12 copper is the wire size for safely handling this amount of current.

A Practical Unit of Electrical Power: Kilowatt-Hour

Public utilities that furnish power for your use charge for electrical energy by using a popular unit of electric work, the kilowatt-hour (kWh). One kilowatt-hour means you used 1000 watts of electric power for 1 hour. If the cost of power is 10 cents per kilowatt-hour, and if you operate a 1000-watt iron for 1 hour, you will be billed for 10 cents. If you operate the iron for 6 hours, you will use 60 cents worth of electrical energy.

The kilowatt-hour meter is a motor-type device that accumulates and displays the kilowatt-hour usage, which is then read by a meter reader and subtracted from the last reading to see how much was used in the interval between meter readings (see Fig. 2-12).

One kilowatt is 1000 watts. One kilowatt-hour is 1000 watts for 1 hour. If the kilowatt-hour is not specified, it means the power being consumed in 1 second.

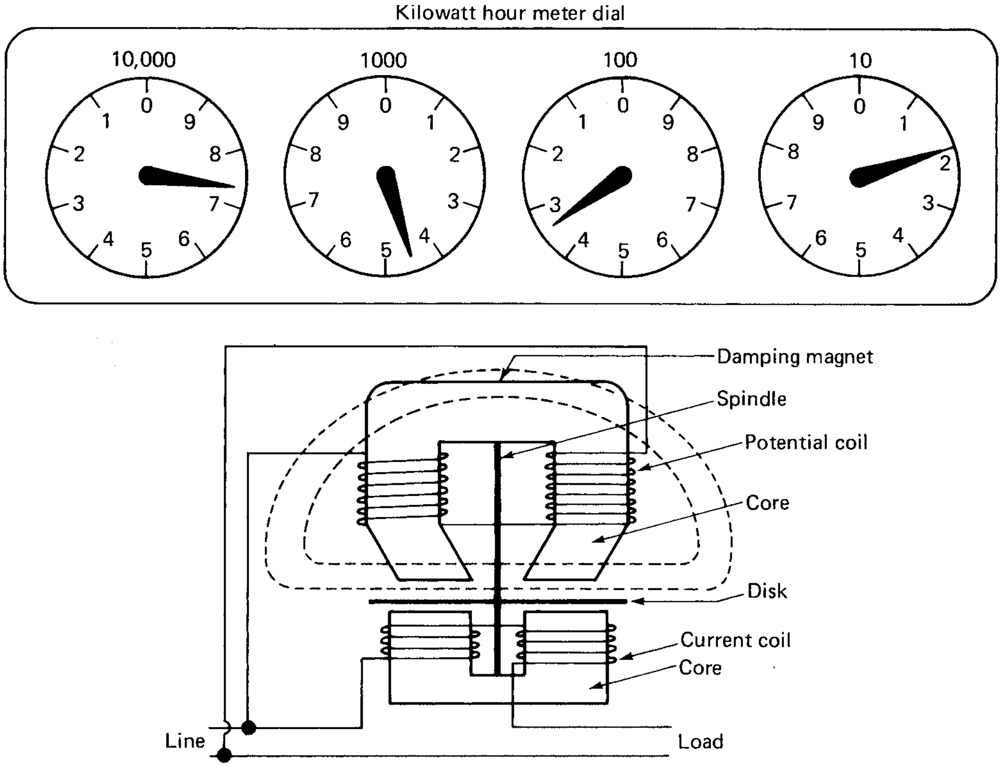
Kilowatt hour meter dial
10,000
1000
100
10
Damping magnet
Spindle
Potential coil
Core
Disk
Current coil
Core
Line
Load

Figure 2-12 Kilowatt-hour meter. Note how it is made and how the dials are read from left to right.

Review Questions

1. The unit of measurement for conductance is _________.
2. Electricity from a battery is produced by a _______ action.
3. How will pressure on crystals produce electricity?
4. What kind of current is produced by using magnetism?
5. Does a battery produce alternating or direct current?
6. The ammeter is used to measure_________.
7. The resistance of a circuit is decreased; what happens to the current if the voltage remains the same?
8. What is a complete circuit?
9. How much current will flow if 1 coulomb per second passes a given point in 1 second?
10. Electrical pressure is measured by a ______meter.
11. What happens to the current in a circuit when the emf applied to a resistance is increased?

12. What happens to the current if the resistance is decreased and the voltage is not changed?
13. Determine the current when the voltage is 64 volts and the resistance is 32 ohms.
14. A voltage applied to a resistance is 64 volts. Resistance is increased from 32 to 64 ohms. What happens to the current?
15. A voltage of 60 volts is applied to a resistance of 20 ohms. What is the current?
16. A voltage of 45 volts is applied to a resistance of 10 ohms. What is the current through the resistor?
17. The current in a resistance is 2 amperes. It is across a 12-volt battery. What is the value of the resistance?
18. A current of 3 amperes is present in a 50-ohm resistance. What is the applied voltage?
19. What is the voltage drop across a resistor that has 8 ohms of resistance and 1.5 amperes of current through it?
20. What is a siemens?
21. How is a siemens related to an ohm?
22. What is the unit of measurement for electrical work?
23. What is the unit of measurement for power?
24. There is a movement of 50 coulombs in 20 seconds; what is the power if the voltage is 6 volts?
25. What is the current if 50 coulombs is moved in 20 seconds by 6 volts of electrical pressure?
26. What is the basic power formula?
27. What is the rate of doing work called?
28. The current in a 50-ohm resistance is 5 amperes. Find the voltage drop and power used.
29. The current in a 30-ohm load is 2 amperes. Find the voltage drop and power used.
30. A 300-watt bulb burns for 4 hours. What is the power consumed in kilowatt-hours?
31. What is the formula used to find power if the voltage and resistance are known?
32. What is the formula for finding power if the current and resistance are known?

Chapter 3

Resistors, Other Electric Components, and Their Symbols

Performance Objectives

- Be able to read the resistor color code.
- Be able to recognize variable and fixed resistors.
- Understand how the capacitor is used in air-conditioning and refrigeration circuits.
- Know the difference between a standard capacitor and an electrolytic capacitor. Know the difference between an inductor and a transformer.
- Know that semiconductors such as diodes, transistors, and silicon chips are used in electronic controls for heating, air-conditioning, and refrigeration devices.

Resistance is found in every electric circuit. Voltage is the pressure that pushes the electrons through the circuit and through the resistance of the consuming device. Everything has resistance to some degree. It is possible to use certain substances in various configurations to produce a device that will limit the amount of current in a circuit. This device is called a resistor. It becomes very useful when you want a number of pieces of equipment to operate from one voltage source. It would be very difficult to use batteries of different voltages for every component in an electric circuit. Resistors of various sizes can be placed in a circuit to cause it to have the correct voltage for the various devices. Resistors are used to drop voltage.

Resistors of various sizes are available. It takes a few pages of any electronics supply house catalog to list all the types handled by the supplier. Our purpose here is to take a look at the method used to mark the value on carbon-composition resistors.

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

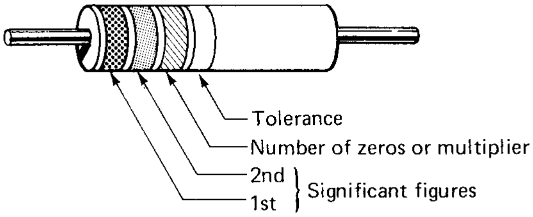

Figure 3-1 Color bands on a carbon composition resistor.

Carbon-composition resistors are small in physical size. They can dissipate $^1/_8$, $^1/_4$, $^1/_2$, 1, and 2 watts. Thus, some method must be devised to mark them. If they are placed in a circuit, it is best to be able to read the value without having to turn the resistor over. Turning it over after it has been soldered in place can cause the connection to break or the leads to break off. Color banding of the resistor seems to be the best answer to the problem of marking.

Carbon-composition resistors have been color coded for a number of years. Three color bands are used to indicate the ohmic value of the resistor. A fourth band is used to indicate the tolerance (see Fig. 3-1).

Resistance is measured in ohms. Wire-wound resistors normally have their values (in ohms) marked on them, since they are physically larger than the carbon-composition type and the lettering or printing can be made on the body of the resistor.

Color Code

The colors used on carbon-composition resistors have a definite value, which is easily read. Each color has a value assigned to it, which is agreed upon internationally, as follows:

Color	Number
Black	0
Brown	1
Red	2
Orange	3
Yellow	4
Green	5
Blue	6
Violet (purple)	7
Gray	8
White	9

When looking at the resistor, you will note that the bands are located at one end. You start reading the resistor value by checking the band nearest the end of the resistor first. Turn it so that you are holding the end with the color band to your left. Then you can read from left to right, as you usually do.

The first two color bands, A and B in Fig. 3-2, indicate the first two digits in the resistance value. The third band is used to show how many zeros are added. C shows

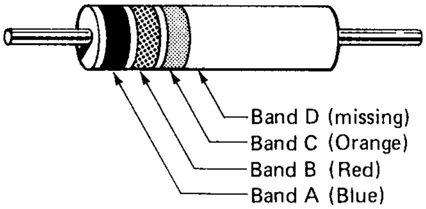

Figure 3-2 Bands A and B indicate the first two digits.

how many zeros follow the first two digits, or it is the multiplier. Sometimes the fourth band (D) is missing. If so, the resistor has a tolerance of ±20%.

Tolerance

When there is a fourth band, it will be either silver or gold. Silver means the tolerance of the resistor is ±10%, while gold indicates ±5%. The lower the tolerance number (5% instead of 10%), the more expensive the resistor is to produce and also buy. The tolerance tells how close the resistor value is to that indicated by the color bands.

Tolerance is indicated in + or – designations to tell you how accurate the resistor is. It may be in the + direction or have more resistance than the value stated by the color bands. Or it may have a – value, meaning the value of the resistor will measure less than the color bands indicate.

Examples

A few examples will make you more familiar with the code and its usefulness. Examine Fig. 3-3. Note the first three bands. The blue-red-orange bands signify 62 followed by three zeros. This produces a value of 62,000 ohms. There is no fourth band, so the tolerance is 20%.

What are the limits of the resistor when you use a meter to check its value? Take the 62,000 and multiply it by 0.20. That produces 12,400. You add the 12,400 to the 62,000 to get 74,400 ohms for the + tolerance value. You take 12,400 away from the 62,000 ohms to get the – or lowest value the resistor can have to still be within tolerance. So, 12,400 from 62,000 produces 49,600 ohms. Thus, the resistor can read between 49,600 ohms and 74,400 ohms and still be called a 62,000-ohm resistor as indicated by its color bands.

Let's take another look at the color code and what it can do for us. Check Fig. 3-4. In this figure you have three color bands to obtain the value of the resistor.

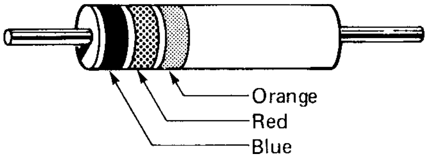

Figure 3-3 A 62,000-ohm resistor.

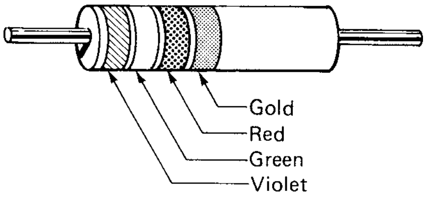

Figure 3-4 A 7500-ohm resistor.

The violet (7), green (5), and red (2) indicate that the resistor has a value of 7500 ohms. The gold band indicates it has a value of 7500 ±5%. Five percent of 7500 ohms is 375 ohms. That gives the 7500-ohm resistor a tolerance range of 7500 + 375 and 7500 – 375. The results are 7875 ohms and 7125 ohms. The resistor can read anywhere between 7125 and 7875 and still be called a 7500-ohm resistor. Sometimes you may find a resistor such as shown in Fig. 3-5. The resistance value is not 750 ohms, but 75 instead. The third band specifies the number of zeros. Since black is zero, these are no zeros after the first two digits. Thus, the black adds nothing and the 75 stands alone. The silver tolerance on the resistor indicates that it has a plus or minus value of 7.5 ohms (75 times 0.10 produces 7.5 ohms). The 7.5 ohms has to be added to and subtracted from 75 to produce the tolerance range of the resistor. So, 75 plus 7.5 equals 82.5 ohms, and 75 minus 7.5 equals 67.5 ohms. The resistor has a tolerance range of 67.5 to 82.5 ohms.

Gold and silver third bands

Carbon-composition resistors have been improved to the point where they can be made in values of less than 10 ohms. Thus, the code has to be altered to fit the situation, and using gold or silver as the third band comes in handy.

Gold as the third band means you divide the first two numbers by 10. Thus, if you have a resistor of red, red, and gold, it has a value of 22 divided by 10, or 2.2 ohms. As you can see, this makes it possible to use the color code now for indicating resistances of less than 10 ohms.

Examine Fig. 3-6. Note that the values of the resistor are indicated by a first band of blue and a second band of yellow, which produces 67. The third band is gold. So, the 67 is divided by 10 to produce a reading of 6.7 ohms for the resistor. If the fourth band is silver, it means the tolerance is ±10%. Ten percent of 6.7 ohms is 0.067 ohm. Add the 0.067 to the 6.7 and you have 6.767 ohms. Subtract 0.67 ohm from the 6.7 and you have a negative tolerance value of 6.633 ohms. The tolerance

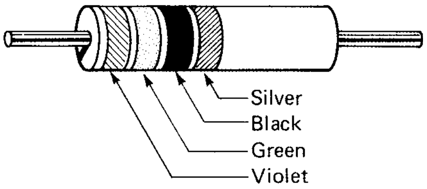

Figure 3-5 A 75-ohm resistor.

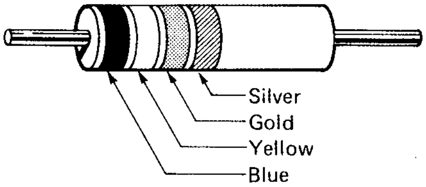

Figure 3-6 Note the third band is gold.

range is 6.633 ohms to 6.767 ohms. It takes a very good ohmmeter to measure this close a tolerance. In fact, it would be hard to see anything near the 6.7 on a conventional ohmmeter. It would probably read 7 or even as much as 10 ohms. The digital ohmmeter is used to measure this type of ohmic value.

Silver as the third band means you divide the first two numbers by 100. Thus, if you have a resistor of red, red, and silver, it has a value of 22 divided by 100, or 0.22 ohm. Take a look at Fig. 3-7. The resistor has bands of yellow, violet, and silver. The values then are 47 divided by 100 to produce the correct answer of 0.47 ohm for the resistor. Most of these resistors will have a fourth band, also. The fourth band still indicates the tolerance. So, if the resistor has a fourth band of silver, it is ±10% as usual. This means that the resistor will vary in resistance by 0.047 ohm in either direction of 0.47. Thus, resistor tolerance may be from 0.423 to 0.517 ohm. It takes a very good ohmmeter to measure this resistance range, but the advent of semiconductors has brought about more sensitive and inexpensive meters and the demand for lower resistance values and closer tolerances.

Resistors come in standard sizes. The Electronics Industries Association (EIA) sets standards for manufacturing of resistors. This helps to standarize the number of sizes of resistors made. It also makes it easy to find spare parts to use for repair jobs.

Variable Resistors

Some resistors are variable. This means that the amount of resistance can be changed. Variable resistors may be either carbon composition or wire wound. These resistors are used for special circuits. On these circuits, the amount of voltage or current that is delivered must be varied. A common example is the volume control on your radio or television set (see Fig. 3-8).

Variable resistors are easily identified because they have three connections for leads. The center lead is usually the variable contact. A variable resistor that is

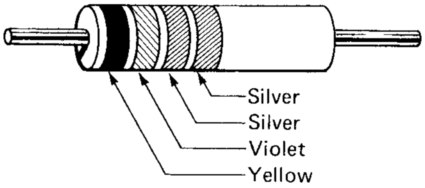

Figure 3-7 Note the third band is silver.

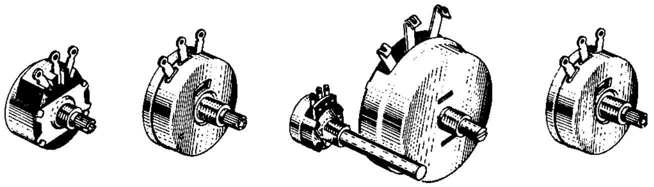

Figure 3-8 Potentiometer.

connected into a circuit at all three points is called a potentiometer (see Fig. 3-8). A potentiometer is often referred to as a pot.

Usually, a potentiometer is used to vary voltage. The device is connected across a voltage source by placing it directly across the battery or power source. The variable arm is then used to change the voltage that is available from the potentiometer. The rheostat is a variable resistor. It is used by connecting it in series not across the voltage source, as was the case with the potentiometer. Rheostats are designed to handle higher currents than potentiometers. Very few rheostats are used today because their jobs are being done by semiconductors. Usually, a rheostat is connected to a circuit at only two points. A symbol for the rheostat is shown in Fig. 3-9.

Variable resistors have a wide range of adjustments. For example, volume controls typically use carbon resistors. Resistance ratings can be adjusted from 0 to 10 million ohms. Another way to state these values is from 0 to 10 megohms (mega means million).

Many potentiometers have what is called a nonlinear resistance element. This simply means that resistance does not change at a fixed, or uniform, rate as adjustments are made. Usually, they are small, or fine, changes at the low end. At the high end, settings lead to large resistance changes. This nonuniform resistance leads to what is called a tapered control. Such devices are usually used to adjust sound volumes and are called audio taper resistors.

There are also linear taper potentiometers. They have a uniform change of resistance as the settings are adjusted. They look exactly the same as the audio taper. When replacing a potentiometer, you must be very careful not to use a linear taper one in a volume control circuit or, worse yet, an audio taper in a control circuit. This is one of the things that you, as a technician, must be aware of in making repairs. Do not try to substitute a volume control of the same resistance for a control circuit potentiometer. You will find it very difficult to make the required adjustments in the control circuit.

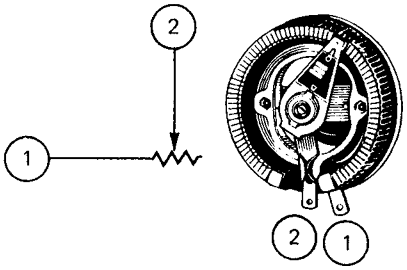

Figure 3-9 A rheostat symbol. Rheostats resemble potentiometers. A potentiometer can be made to serve as a rheostat by connecting the center terminal and one of the end terminals to form a single connection.

Wattage ratings are usually marked on the rheostat or potentiometer. It is difficult to tell the wattage rating by just observing the device. It takes practice to be able to tell the difference between various wattage ratings.

Schematic

Persons who work with electricity use a schematic to tell them where components are located and what the relationships are to one another. A schematic is a diagram something like a road map. It tells you the way from one point to another. Symbols are used to mark certain items along the way. The schematic is the road map of the electrical trade. Many types of diagrams are needed to describe the operation and construction of electrical equipment. The schematic is the most common term for the electrical diagram.

The schematic shows how the electrical or electronic components (parts) are connected. A schematic is a handy item to have when you are analyzing, explaining, or servicing heating, air-conditioning, and refrigeration equipment.

Types of Resistors

The resistor is the most widely used electrical and electronic device. Every radio, television set, and control circuit has a resistor or resistors. This component is used to provide resistance. It is designed to be used at a fixed value or as a variable-value device.

Fixed resistors

The fixed resistor is the simplest of the two types. It is made so that you cannot change the resistance. Some carbon fixed resistors are shown in Fig. 3-10. These are carbon composition and have a cover of black, brown, or green plastic. A color code is used to give the value of the resistor.

Fixed wire-wound resistors are available for use when the wattage rating is higher than 2 watts. Carbon-composition resistors come in $^1/_8$, $^1/_4$, $^1/_2$, 1, and 2 watt sizes. The physical size tells the rating. You get used to the wattage rating when working with resistors. The larger the resistor, the higher the wattage rating is.

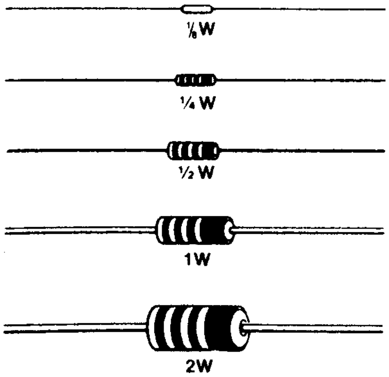

Figure 3-10 Fixed carbon-composition resistors.

The larger the resistor, the easier it is for it to dissipate heat. Since resistors put up resistance to current flow, they also drop voltage. The energy has to be dissipated as heat. Thus, the surface of the resistor must be large enough to allow the heat to be dissipated.

Figure 3-11 shows some fixed wire-wound resistors. These are made of high-resistance wire wound on an insulating core with a ceramic coating. They usually are large enough so that the resistance of the unit can be stamped on it.

The symbol for a fixed resistor is shown in Fig. 3-12. Note how the symbols vary for different users. A is the standard EIA (Electronics Industries Association) symbol. B is usually used by foreign manufacturers and occasionally by American makers of industrial equipment. C is seldom encountered, but it is sometimes used in schematics for industrial equipment and can be seen in some refrigeration and air-conditioning electrical schematics.

Tapped resistors

Tapped resistors are used in some circuits. They have taps for easy connections. They are usually wire wound, although some are carbon. Figure 3-13 shows samples of the tapped resistor. Figure 3-14 shows the schematic representation of tapped resistors.

The ceramic coating is left off the wire where the tap is to be made. This allows a sliding connection so that the tapped resistor can be made into a variable resistor or adjusted as needed.

Variable resistors

A variable resistor may be made of carbon or it may be wire wound. The idea behind the variable is to make it adjustable to meet the needs of the circuit. You are most familiar with the variable resistor as a volume control on a radio or television set. This is a variable carbon-composition type of resistor and controls a circuit to allow for increases or decreases in volume, as you desire.

A variable resistor has a movable contact that is used to adjust or select the resistance value between two terminals. In most uses, the variable resistor is a control device. It is made in many sizes and shapes. Figure 3-15 shows some of these types. The shafts of most variable resistors have knobs placed on them to make them easier to use. However, some are made to be adjusted by the insertion of a screwdriver blade in a slot on the resistor. Many adjustable resistors are used in controls for air-conditioning and refrigeration systems. Figure 3-16 shows the schematic representation for variable resistors.

Fusible resistors

In some cases, the resistor has a purpose other than providing resistance. One type is used to protect the equipment or circuit against excess current surges. This type of resistor, called a fusible resistor, is built to fail before damage is done to more expensive parts. Such units are often made to plug into a socket (see Fig. 3-17). Figure 3-18 shows the schematic symbol for a fusible resistor.

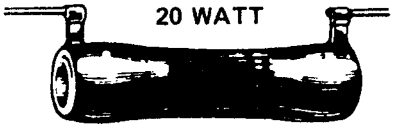

Figure 3-11 Fixed wire-wound resistors.

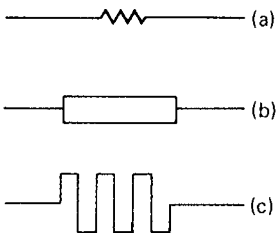

Figure 3-12 Symbols for a fixed resistor.

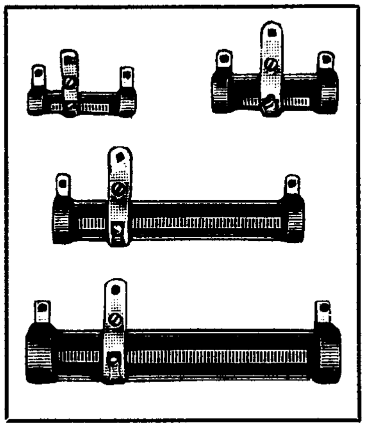

Figure 3-13 Tapped resistors.

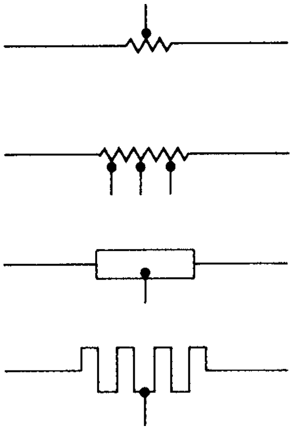

Figure 3-14 Schematic representation of tapped resistors.

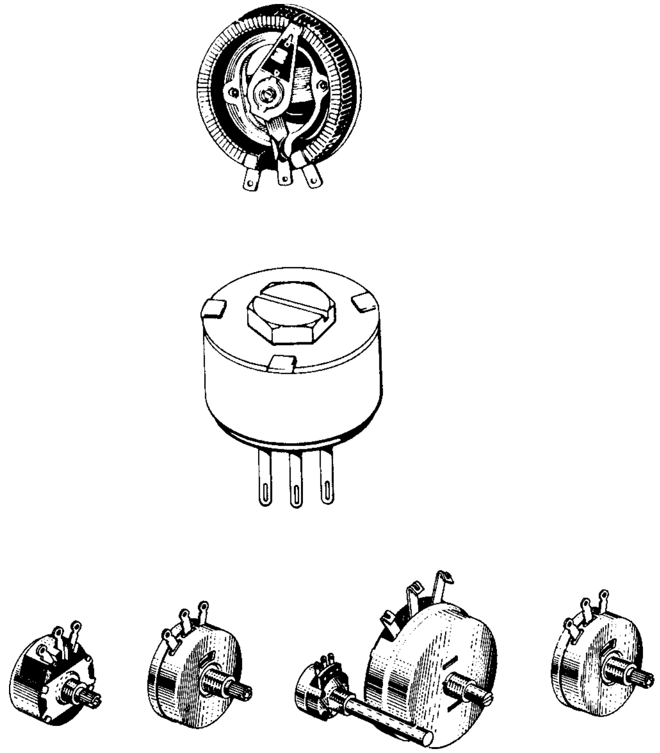

Figure 3-15 Types of variable resistors.

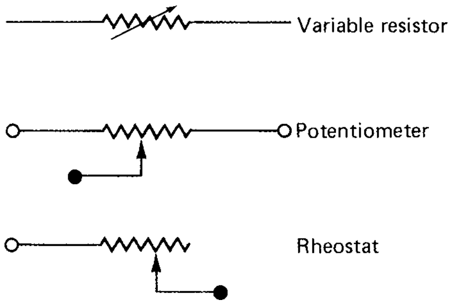
Variable resistor
Potentiometer
Rheostat

Figure 3-16 Schematic representation of variable resistors.

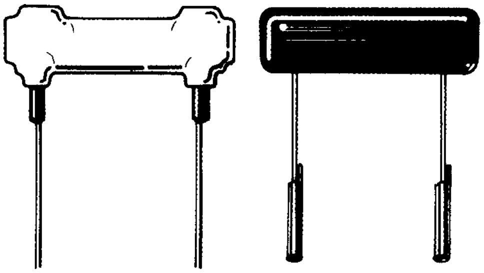

Figure 3-17 Fusible resistors.

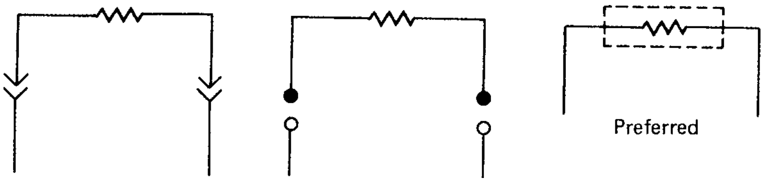

Figure 3-18 Fusible-resistor symbol.

Temperature-compensating resistors

Another type of special resistor is the temperature-compensating resistor. These are designed so that the resistance value changes in a direct or inverse relation with temperature changes. Such resistors are used to provide special control of circuits that must be extremely stable in operation. The symbol is shown in Figure 3-19.

Types of Capacitors

The capacitor is used in many electric circuits in both electronics and in air-conditioning and refrigeration circuits. Two types of capacitors are used in these circuits: fixed and variable.

Fixed capacitors

The fixed capacitor is made for a certain value and is not adjustable. The fixed capacitor is divided into several groupings. It may be made with paper separating two plates of aluminum foil, or it may use plastic, mica, ceramic, or electrolytes.

Most paper capacitors have been replaced by those made of better materials, usually plastic. A typical capacitor is shown in Fig. 3-20. Capacitors are large enough to have their values printed on them. The smaller capacitors use a color code to indicate their value and working voltage. Capacitors come in hundreds of sizes and shapes. It takes a good half-hour to thumb through an electronics catalog that shows all the various types. Each type has a special or particular application. Mica types, for instance, are used for some high-frequency applications with high voltages. The ceramic type is found in circuits that use high voltages, such as television sets and radar equipment.

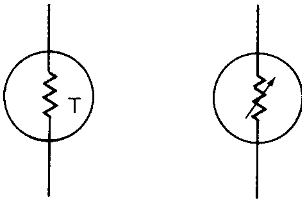

Figure 3-19 Symbol for temperature-compensating resistor.

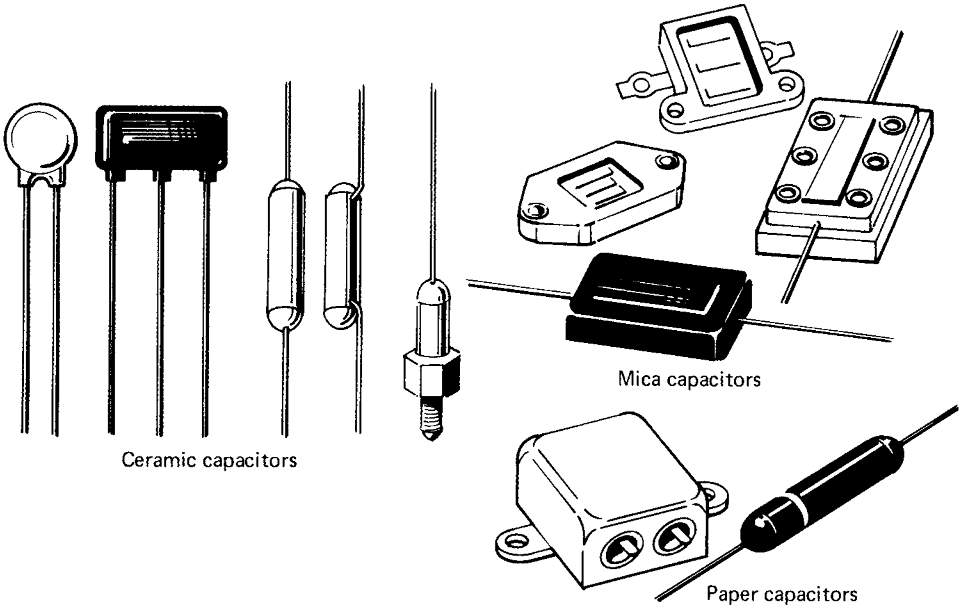
Ceramic capacitors
Mica capacitors
Paper capacitors

Figure 3-20 Fixed capacitors.

Electrolytic capacitors

This type of capacitor deserves special mention, but will be covered in detail in Chapter 10. The electrolytic capacitor has a very high capacitance value when compared with the types mentioned previously. These capacitors may be tubular or square in shape. They have cardboard or metal covers. Values are printed on the cardboard cover and stamped into the metal cover (see Fig. 3-21). They

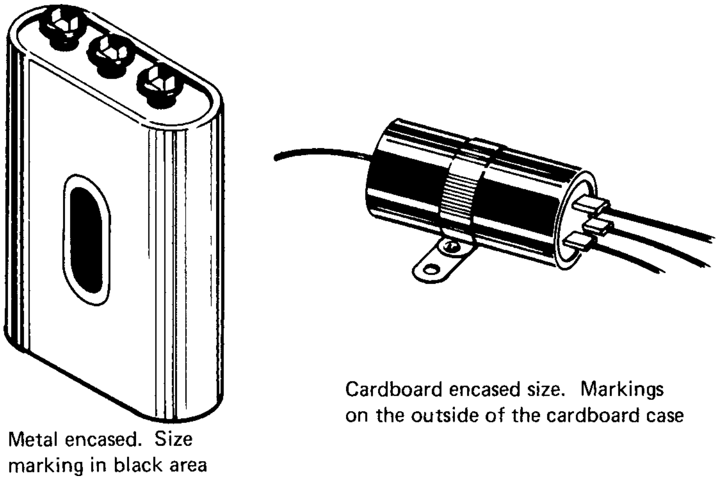
Metal encased. Size marking in black area
Cardboard encased size. Markings on the outside of the cardboard case

Figure 3-21 Electrolytic capacitor with value marked on it.

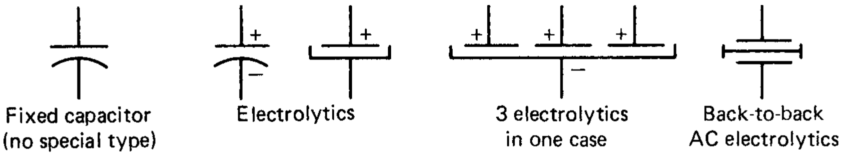

Figure 3-22 Capacitor symbols with electrical polarity marked.

are available in a variety of shapes and sizes. One characteristic of the electrolytic capacitor is its polarity. Its terminals will have − (negative) or + (positive) marked on them. This means that the circuit power must be connected correctly to avoid damage to the electrolytic capacitor. It is not to be used on ac unless it is an ac electrolytic capacitor and so identified.

Caution*: If an electrolytic capacitor marked with a – and a + is connected to ac, it will explode, and can throw its contents over an area as large as 50 square feet. Thus it can be dangerous. Some are manufactured with a small hole in them so that their contents will spew out instead of exploding. However, safety dictates that you treat all electrolytic capacitors as firecrackers and a larger one as a piece of dynamite. This is another reason for wearing eye protection when working with electric circuits.*

Note that the symbols for capacitors in Fig. 3-22 indicate the electrolytic capacitor with polarity markings.

Electrolytic capacitors are 1 microfarad (μF) and larger in size. They can be made to operate on ac by connecting two of them back-to-back as shown in Fig. 3-23. AC electrolytic capacitors are used in electrical motors, crossover networks in speaker systems, and other places needing large capacitances in circuits that contain ac.

Variable capacitors

Variable capacitors are used for tuning purposes in radios and televisions. In most instances, you will not need them for air-conditioning and refrigeration circuits. However, in case you do see one utilized in the electronics control unit, you can identify it by using Fig. 3-24.

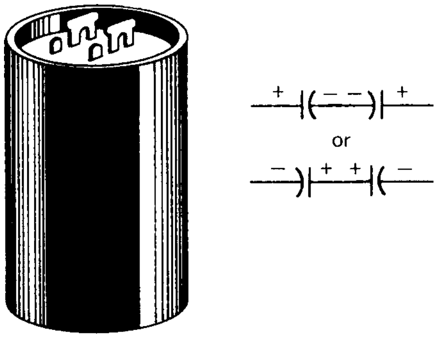

AC electrolytic

Figure 3-23 Back-to-back electrolytics for ac operation.

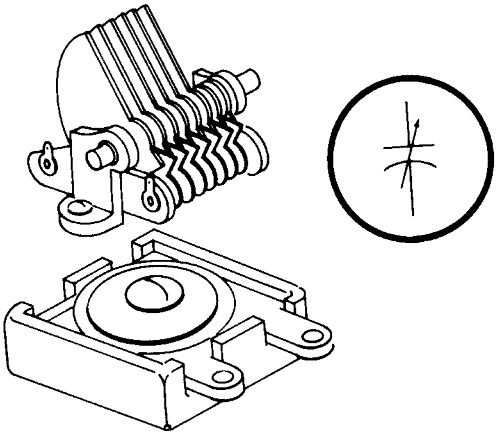

Figure 3-24 Variable capacitors with symbols.

Types of Inductors

Coils, inductors, and chokes are the names used to indicate a coil of wire. "Inductor" is preferred because inductors have inductance, a property that is utilized in many electric circuits.

Fixed inductors

The simplest coil or inductor has an air core and is made by winding a wire in a series of loops, which may or may not have a form to hold them in place. Coils are seldom color coded for value, so we look at the schematic or a parts list for the inductance value of a coil. Inductance is the electrical property of a coil, just as resistance is the electrical property of a resistor. Many coils are wound on plastic forms that support the loops of wire. The form has no effect on the operation of the coil. The symbols for air-core coils are shown in Fig. 3-25. Other types are powdered iron core and iron core. Symbols for these types are shown in Fig. 3-26.

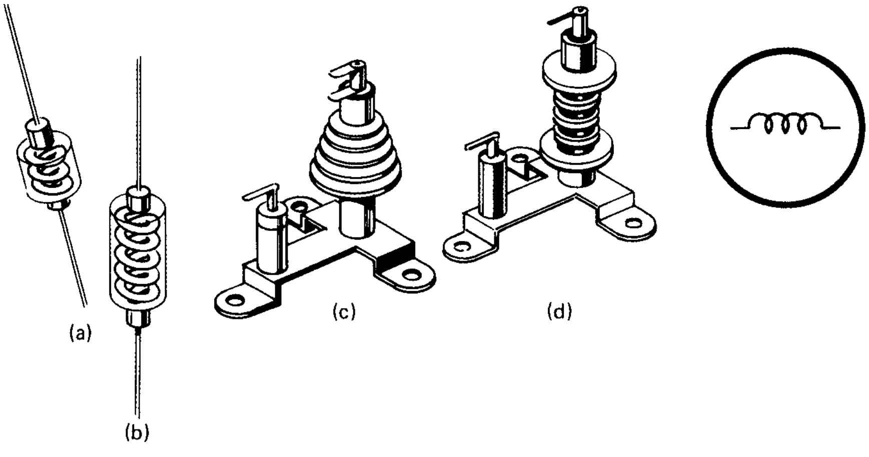

Figure 3-25 Symbols for air-core coils.

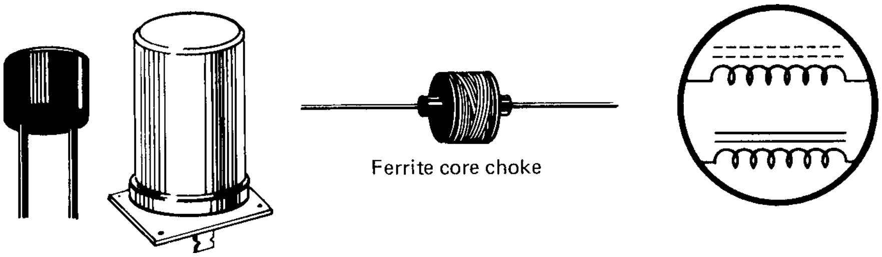

Figure 3-26 Powdered-iron-core and iron-core inductors.

Variable inductors

Some circuits need inductors that can have their values changed, some by screwdriver adjustment and others by changing the core material. Figure 3-27 shows the symbols for variable inductors. Note the differences for iron-core variable inductors. Iron-core (made of iron or steel) chokes are indicated by two straight lines over the loops. Dashed lines indicate powdered iron cores.

Transformers

The inductor has one coil. The transformer has two. The two coils of a transformer are so placed that the energy in one is transferred to the other by magnetic induction. The coil winding connected to the power source is called the primary. The winding connected to the circuit or consuming device is called the secondary. Or keep in mind that the primary is the input and the secondary is the output. Note in Fig. 3-28 that the two coils are not connected. Transformers are used in air-conditioning and refrigeration circuits for stepping down the voltage from 120 or 240 volts to 24 volts. The thermostat circuit then uses the 24 volts. The transformers you will work with have iron cores. Air-core transformers are used in electronics where radio frequencies are present.

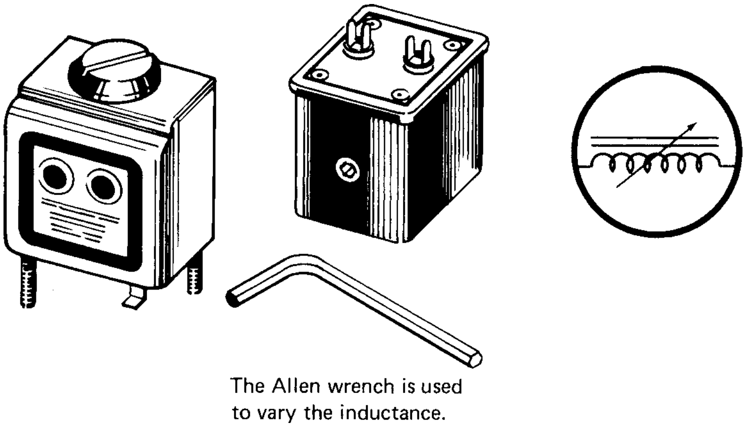

Figure 3-27 Symbol for variable inductors.

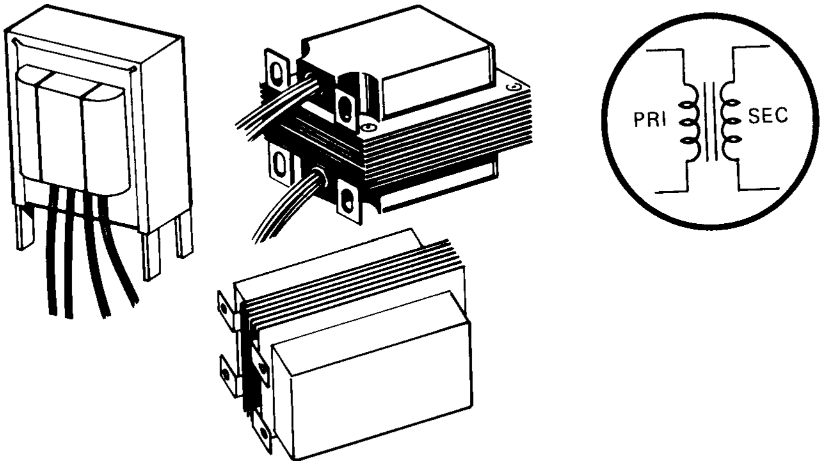

Figure 3-28 Symbol for a transformer.

Semiconductors

You know semiconductors as transistors, diodes, and chips. Silicon material is used to make devices that will conduct a certain amount of current when needed. These devices are made for easy use and with low voltages. You will encounter semiconductor chips in memory devices and computers; transistors are used in many types of circuits, and diodes are used as rectifiers and control devices.

Diodes

Diodes have an anode and a cathode. The cathode end is marked (+) to indicate which polarity it must be connected to in order to work properly. Diodes have many uses. They are found in circuits that are used to change ac to dc for certain control devices. They may also be used as protective devices in circuits. The symbol for a diode is shown in Fig. 3-29.

Transistors

The transistor is used for switching and amplification. It consists of three active elements. There are two basic types of transistors, PNP and NPN (see Fig. 3-30). These terms will be discussed later. E stands for emitter, C stands for collector, and B is used to indicate the base connection.

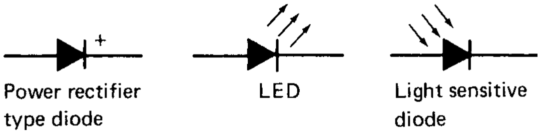

Power rectifier type diode

LED

Light sensitive diode

Figure 3-29 Diode symbol.

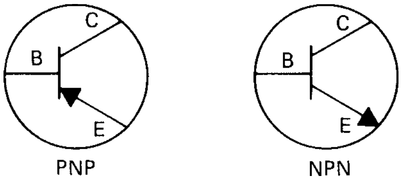

Figure 3-30 Transistor symbols, PNP and NPN.

Switches

Switches are used to turn on or off a circuit. They can be made as a simple on-off device or used to control many functions in a number of circuits.

Switches have names that designate what they can do in terms of turning on or off a circuit or circuits. For instance, there are SPST (single pole, single throw) switches, DPDT (double pole, double throw), DPST (double pole, single throw), single-pole, six-position, and so forth. Symbols for various types of switches are shown in Fig. 3-31.

Relays

Relays are switches that are moved electrically instead of by hand. They can be made in almost any configuration of poles, throws, and construction. The force that operates a relay is magnetic. The magnetic pull is produced by current passing through a coil of wire. The attraction of an armature causes the switch sections to operate.

Some relay types that you may run into in the air-conditioning and refrigeration field are shown in Fig. 3-32. Relays can also be represented by symbols (see Fig. 3-33).

Fuses and Circuit Breakers

Safety devices are needed on a piece of equipment to protect the expensive equipment and the persons who work on it. The fuse is a safety device designed to protect the equipment from itself. The fuse protects the equipment when it develops a condition that causes it to draw too much current.

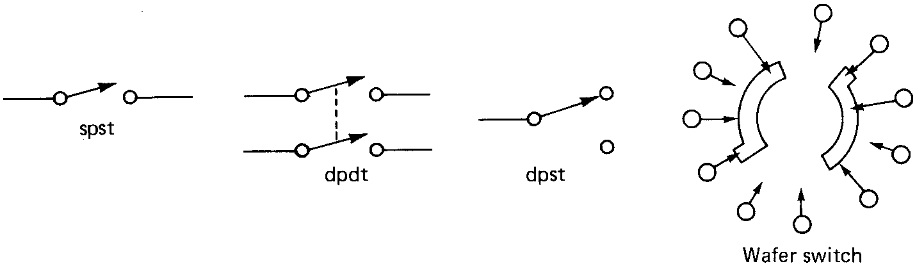

Figure 3-31 Switch symbols.

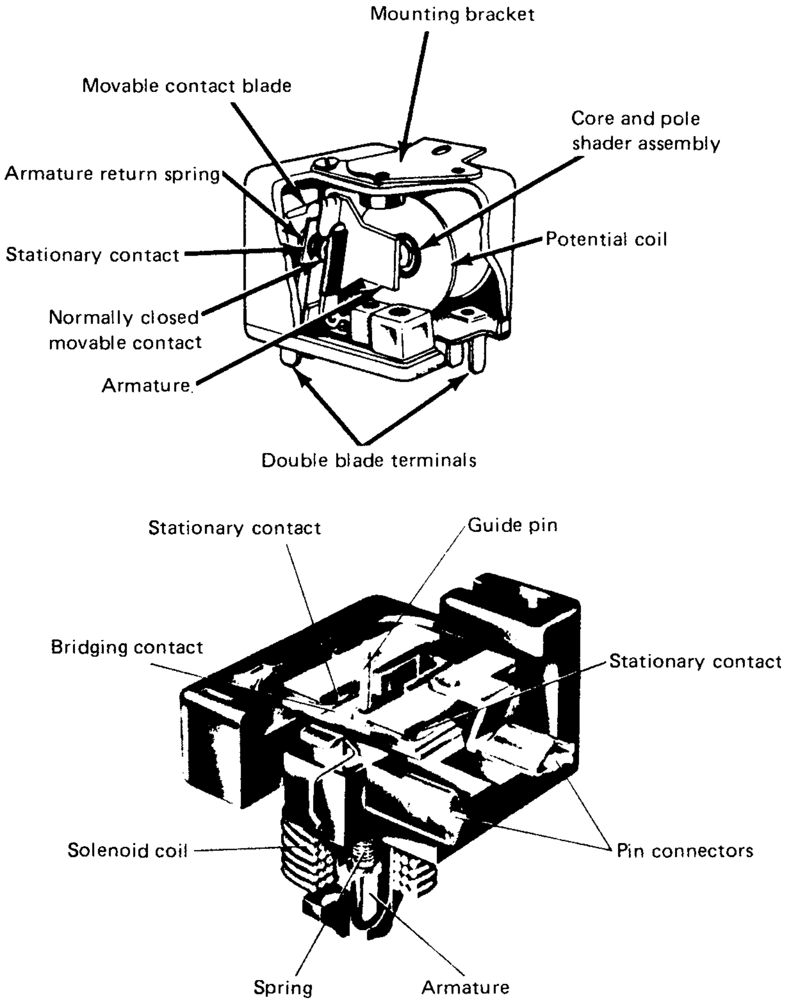
Mounting bracket
Movable contact blade
Core and pole shader assembly
Armature return spring
Stationary contact
Potential coil
Normally closed movable contact
Armature
Double blade terminals
Stationary contact
Guide pin
Bridging contact
Stationary contact
Solenoid coil
Pin connectors
Spring
Armature

Figure 3-32 Relays.

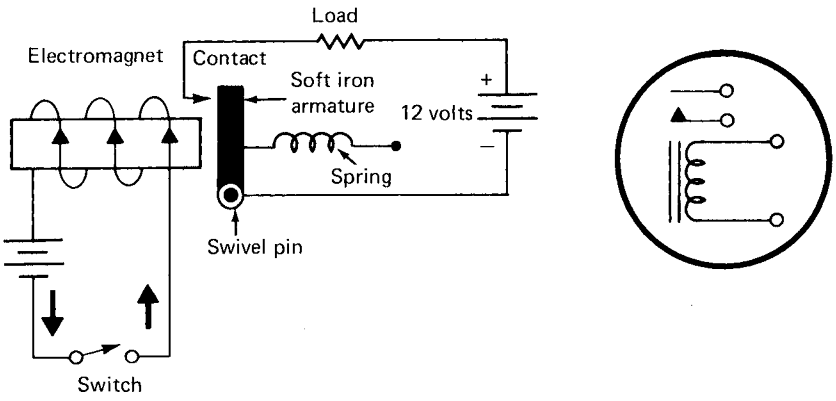
Load
Electromagnet
Contact
Soft iron armature
+
12 volts
−
Spring
Swivel pin
Switch

Figure 3-33 Relay Symbols.

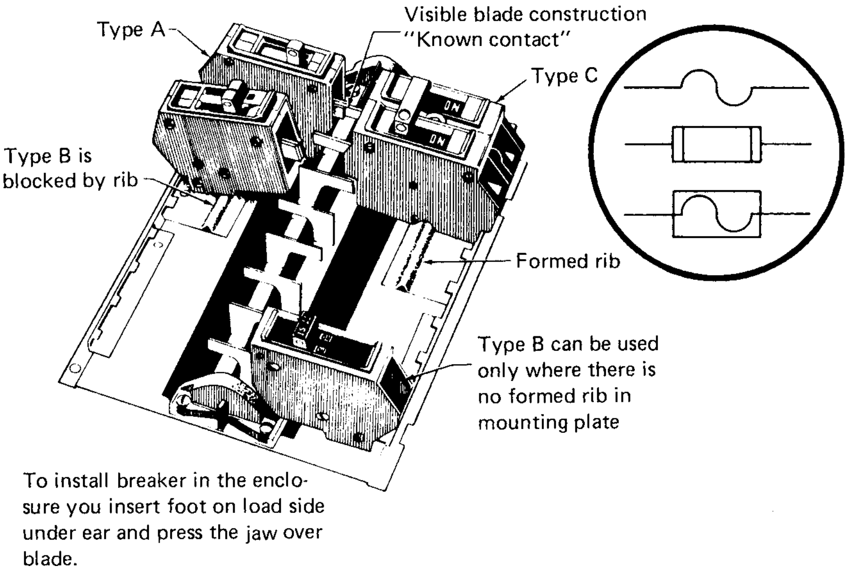

Type B can be used only where there is no formed rib in mounting plate

To install breaker in the enclosure you insert foot on load side under ear and press the jaw over blade.

Figure 3-34 Fuse symbols and circuit breakers.

Many circuits need a device to protect the equipment from shorts. Such devices as a fuse or circuit breaker are called for in most electric circuits. Motors have circuit breakers inserted in their windings that will turn off the motor before it reaches a critically high temperature. Fuses and circuit breakers are represented by a symbol such as that in Fig. 3-34. Typical fuses are shown in Fig. 3-35. They are manufactured in many sizes and values.

Lamps

Indicator lamps come in a variety of shapes and sizes. They may be incandescent, the type used in lamps in the home, or they may be neon types. Light-emitting diodes (LEDs) are also used as indicator lamps. Their symbol is the same as a diode with a couple of arrows pointing away from them to indicate

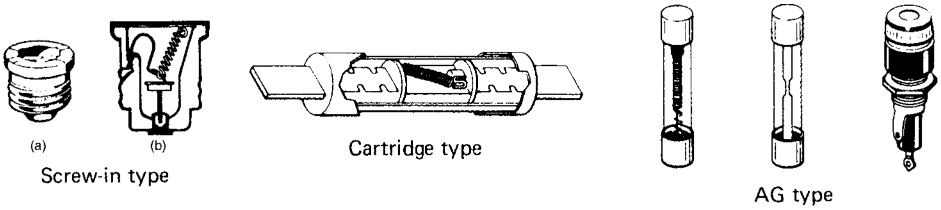

Screw-in type

Cartridge type

AG type

AG stands for automotive glass. They were first used in automobiles.

Figure 3-35 Different types of fuses.

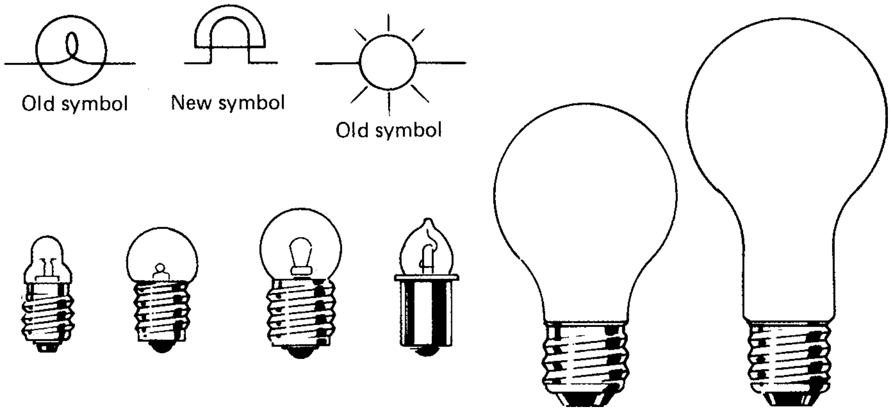

Figure 3-36 Incandescent lamps with incandescent-lamp symbols.

they emit light. However, the incandescent symbol and the neon lamp symbol are quite different (see Figs. 3-36 and 3-37).

Batteries

Many pieces of electronic equipment use batteries. Some types of programmable thermostats use batteries to keep the circuit operating when there is no power. They are required to keep the memory active in the chip so that the settings for times on and off and temperature are not lost with a power failure.

Batteries are made up of cells. A cell symbol is shown in Fig. 3-38 along with that for a battery. The battery is made up of more than one cell. The way to indicate the voltage is by a number alongside the battery symbol, rather than trying to draw the correct number of cells.

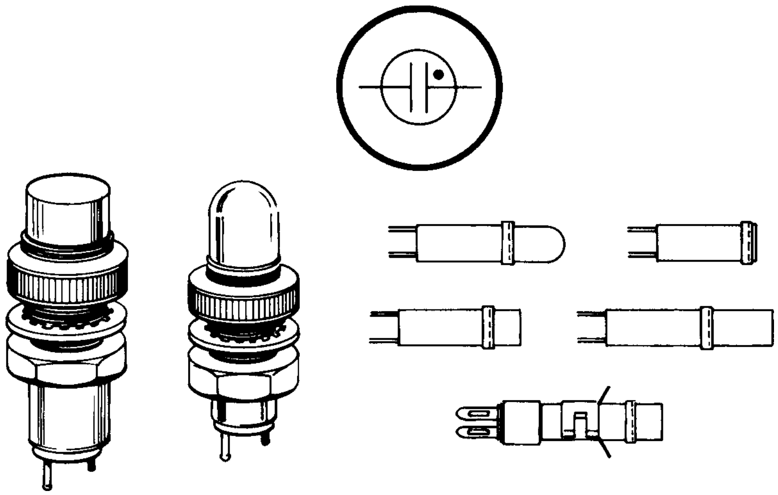

Figure 3-37 Neon lamps with neon-lamp symbols.

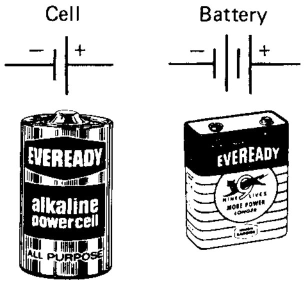

Figure 3-38 Cell and battery symbols.

Meters

Many meters are used in electronic and electric circuits. They are represented by a V or A inside a circle for voltmeter and ammeter, respectively (see Fig. 3-39).

Reading Schematics

As a technician you will be required to read schematics. This is necessary because all air-conditioning and refrigeration equipment use electric circuits for operational control. Study this chapter to be sure you can recognize the various symbols used in schematics. The more you study them, the easier it will be to interpret circuit meanings.

Review Questions

1. What is meant by the term resistance?
2. What is a resistor?
3. Why are there so many types of resistors?
4. What is the unit of measurement for resistance?
5. What is the value of a resistor with green, blue, and red for the three stripes on its body?
6. What is meant by tolerance?
7. How does tolerance affect resistors?
8. What is the difference between a potentiometer and a rheostat?
9. What is meant by linear taper?
10. What is a schematic?

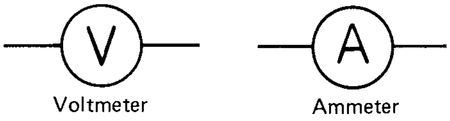

Figure 3-39 Meter symbols.

11. What is a capacitor?
12. Where is the capacitor used?
13. How is an electrolytic capacitor different from others?
14. What is an inductor?
15. How is the inductor put to work in a circuit?
16. Name two types of semiconductors.
17. What is the difference between a relay and a circuit breaker?
18. What is the symbol for a battery?
19. What is the symbol for a lamp?

Chapter

4

Series and Parallel Circuits

Performance Objectives

- Be able to identify a series circuit.
- Be able to identify a fuse and how it is used in a circuit.
- Be able to identify a parallel circuit.
- Be able to determine the total resistance in a parallel circuit.
- Be able to determine the total resistance in a series circuit.
- Be able to determine the total resistance in a series-parallel circuit.
- Be able to solve problems associated with series, parallel, and series-parallel circuits.

To be able to read a schematic, you have to know the types of circuits used in electrical work. Symbols in series and parallel circuits are the first steps in learning to read a schematic.

Series Circuit

A series circuit consists of resistors or other electrical components connected one after another. If you place cells in series, they form a battery. It takes more than one cell to make a battery (see Fig. 4-1). Cells in series produce a higher voltage. Two 1.5-volt cells in series produce 3 volts. The voltages of each cell are added to produce the total voltage. The 9-volt transistor radio battery is actually composed of six 1.5-volt cells connected in series. In a series connection, – is connected to +. The current is limited to whatever is produced by one cell. If one cell is dead, the whole battery is useless.

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

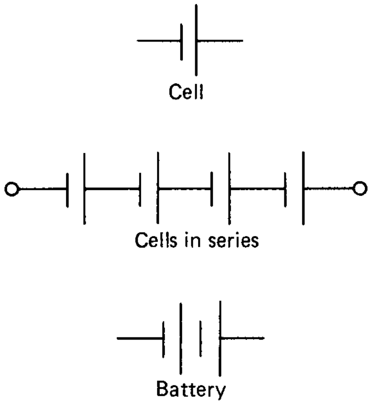
Cell
Cells in series
Battery

Figure 4-1 Cells in series.

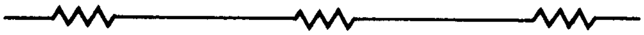

Figure 4-2 Resistors in series.

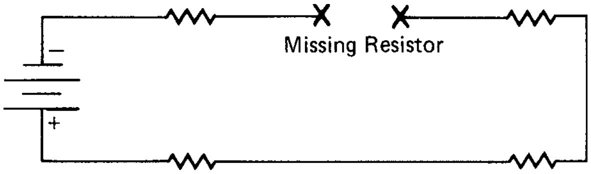
Missing Resistor
–
+

Figure 4-3 Series circuit with one resistor missing.

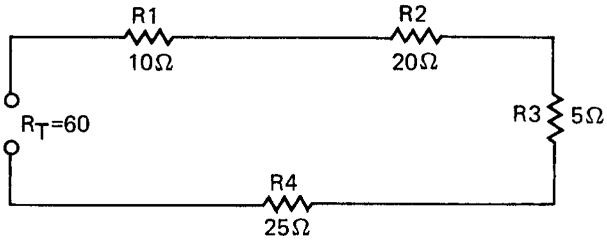
R1
10Ω
R2
20Ω
R3
5Ω
R4
25Ω
R_T=60

Figure 4-4 Total resistance is equal to the sum of the individual resistances.

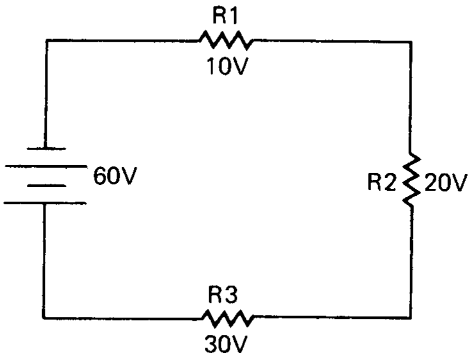
R1
10V
60V
R2
20V
R3
30V

Figure 4-5 Voltage drops around the loop equal the applied voltage.

Series circuits have some peculiar properties. They are made up of resistors connected one after the other (see Fig. 4-2). If one resistor is open or removed from the circuit, the whole circuit is open (see Fig. 4-3). In this case, no current flows because the electrons do not have a complete path from − to + of a power source.

Resistances in series

If the resistances that form a series circuit are added, their total resistance can be found (see Fig. 4-4). When $R_1 + R_2 + R_3 + R_4$ are added, the total resistance in the circuit is 60 ohms. Or R_1 (10 ohms) + R_2 (20 ohms) + R_3 (5 ohms) + R_4 (25 ohms) is equal to 60 ohms. When this is written as a formula it becomes

$$R_T = R_1 + R_2 + R_3 + R_4$$

Voltages in series

The total voltage needed for a series circuit is found by adding the voltages needed by each resistor. Or, as Fig. 4-5 shows.

E_A = voltage applied [total voltage]

E_{R1} = voltage drop across R_1

E_{R2} = voltage drop across R_2

E_{R3} = voltage drop across R_3

Step 1: $E_A = E_{R1} + E_{R2} + E_{R3}$

Step 2: $E_A = 10 + 20 + 30$

Step 3: $E_A = 60$ volts

In a series circuit, the voltage applied is equal to the sum of the voltage drops across individual resistors. Voltage in series divides according to the resistance.

Current in series

Current in a series circuit is the same through all resistors. There is only one path for electrons to move from − to +. If the current in R_1 (Fig. 4-5) is 3 amperes, then the current through R_2 and R_3 is also 3 amperes.

I_T = total current

I_{R1} = current through R_1

I_{R2} = current through R_2

I_{R3} = current through R_3

Therefore, the formula for current in a series circuit is simply stated as

$$I_T = I_{R1} = I_{R2} = I_{R3}$$

Series circuit rules

Three basic elements must be considered in any circuit: current, voltage, and resistance. Three simple rules for series circuits concern these elements.

Voltage: Applied voltage is equal to the sum of the individual voltage drops around the series loop.

$$E_A = E_{R1} + E_{R2} + E_{R3}$$

Current: Total current in a series circuit is the same as the current in any resistor in the circuit.

$$I_T = I_{R1} = I_{R2} = I_{R3}$$

Resistance: Total resistance in a series circuit is equal to the sum of the individual resistances.

There is a disadvantage in using a series circuit. If one resistor opens, the entire circuit will not operate. An open anywhere in the circuit means the circuit will not function. The path of electrons from − to + terminals of the battery is broken. This feature of a series circuit can be useful. A switch can be used to turn the circuit on-off (see Fig. 4-6).

Fuses

A fuse can also be placed in series to protect the power supply (see Fig. 4-7). If the fuse opens, the circuit is turned off. This means you must locate the trouble. Then you replace the fuse. This safety device (fuse) is very useful.

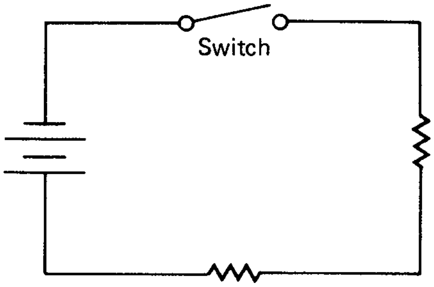

Figure 4-6 A switch controls the flow of electrons.

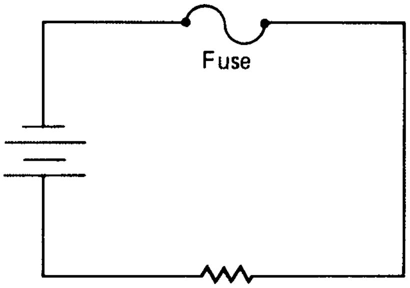

Figure 4-7 The fuse is in series with the load.

A fuse is a safety device. Its symbol is ∿. Fuses are placed in series in a circuit to protect the circuit from excess current. A fuse contains a piece of wire that melts to open the circuit. It melts when too much current flows in the circuit. The size of the fuse determines the amount of current it will safely pass. Fuses are available in a number of sizes and shapes.

Parallel Circuit

Cells in parallel

When cells are connected in parallel, the voltage output from the hookup is the same as from one cell. However, the current available is found by adding the current of each cell. Thus, two cells of 1.5 volts each in parallel produce 1.5 volts output. However, if the current available from each of the two cells is 1 ampere, then the total current available is 2 amperes. The type of material used to make a cell determines its voltage. The physical size of a cell determines its current.

Parallel circuit characteristics

Parallel circuits are used where a number of devices use the same voltage. This is the case in wiring the lighting circuits in your home (see Fig. 4-8). Most appliances

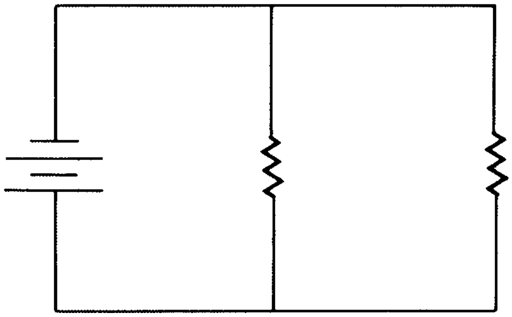

Figure 4-8 Two resistors connected across a battery are in parallel.

in the home are made to operate on the same voltage, and they are connected in parallel to the same voltage, 120 volts.

A parallel circuit has the same voltage applied to each device. However, the current in a parallel circuit will vary with the resistance of the device.

Currents in a parallel circuit

When two resistors are connected in parallel, current can flow through each resistor. Current divides according to the resistance. Two paths for current flow are present. If one path opens, it does not affect the other. The total current in a parallel circuit is equal to the sum of the branch circuits, or

$$I_T = I_{R1} + I_{R2} + I_{R3} + \ldots$$

Resistances in a parallel circuit

Resistances in a parallel circuit cannot be added to yield the total resistance. The following formula must be used to determine the total resistance when more than one resistor is connected in parallel.

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots$$

For instance, if you have three resistors of 5 ohms, 10 ohms, and 20 ohms in parallel, the total resistance is found as follows:

$$\text{Step 1: } \frac{1}{R_T} = \frac{1}{5} + \frac{1}{10} + \frac{1}{20}$$

$$\text{Step 2: } \frac{1}{R_T} = \frac{4}{20} + \frac{2}{20} + \frac{1}{20}$$

$$\text{Step 3: } \frac{1}{R_T} = \frac{7}{20}$$

$$\text{Step 4: } \frac{R_T}{1} = \frac{20}{7}$$

$$\text{Step 5: } R_T = 2\frac{6}{7}$$

A common denominator is necessary to add fractions. Fractions are produced when resistor values are inserted into a formula.

Another formula is used when there are only two resistors in parallel.

$$R_T = \frac{R_1 \times R_2}{R_1 + R_2}$$

If $R_1 = 10$ ohms and $R_2 = 20$ ohms, then

$$R_T = \frac{10 \times 20}{10 + 20} \quad \text{or} \quad R_T = \frac{200}{30}$$

$$\text{thus } R_T = 6\frac{2}{3} \text{ ohms}$$

Parallel circuit rules

Three elements must be dealt with in parallel circuits. They are voltage, current, and resistance. For a quick review, look at these three rules governing parallel circuits:

Voltage: Voltage across each resistor is the same as the applied or total voltage.

$$E_A = E_{R1} = E_{R2} = E_{R3}$$

Current: Total current is equal to the sum of the individual currents.

$$I_T = I_{R1} + I_{R2} + I_{R3} + \ldots$$

Resistance: Total resistance in a parallel circuit is found by using one of two formulas.

For two resistors only,

$$R_T = \frac{R_1 \times R_2}{R_1 + R_2}$$

For two or more resistors,

$$\frac{1}{R_T} = +\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots$$

To connect power sources in parallel, make sure the polarity is observed. Connect + to + and − to − to make a parallel hookup.

Series-Parallel Circuits

Series-parallel circuits use the properties of both series and parallel circuits (see Fig. 4-9). Note that R_1 is in series with R_2 and R_3 and R_2 and R_3 are in parallel. Total current flows through R_1 but divides to go through R_2 and R_3. This type of circuit uses the properties of both series and parallel circuits. It is a good idea to make sure you understand the three rules of series circuits and the three rules of parallel circuits before moving ahead.

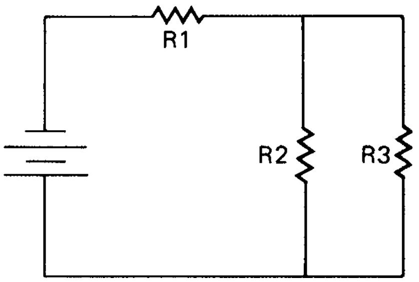

Figure 4-9 Parallel-series circuit.

Series-parallel resistance circuits

It is possible to have circuits that combine series and parallel connections. Within these circuits, some resistors or consuming devices are connected in series. Others are connected in parallel. An example of such a circuit is shown in Fig. 4-10. Another technique for writing a series-parallel circuit is shown in Fig. 4-11. One feature of a series-parallel circuit is obvious from these diagrams. There must be at least three resistors or consuming devices.

Determining resistance

To determine resistance for a series-parallel circuit, follow a simple method: Calculate the resistance for the parallel units first. Then treat the entire circuit as a series circuit. After you have reduced the values of the resistors that are in parallel to a series equivalent, you add all resistance values together. For example, look at the diagram in Fig. 4-12. Both parallel resistors have values of 30 ohms. Thus, the parallel circuit has a total resistance of 15 ohms. Remember the rule: For two resistors of equal value, divide the value of one of the resistors by 2:30 ohms divided by 2 = 15 ohms. In effect, you have a series circuit like the one in Fig. 4-13. You might want to redraw the circuit to aid in figuring the values for series-parallel circuit. Figure 4-13 shows two resistors in series. One has a value of 10 ohms. The other has a value of 15 ohms. So the total resistance of the circuit is 25 ohms. Remember the rule: To determine the total resistance in a series circuit, add the values of all resistors.

To review these steps for a more complex circuit, consider Fig. 4-14. The diagram shows a circuit with five resistors. To figure the resistance for this circuit, follow the steps shown in Fig. 4-15.

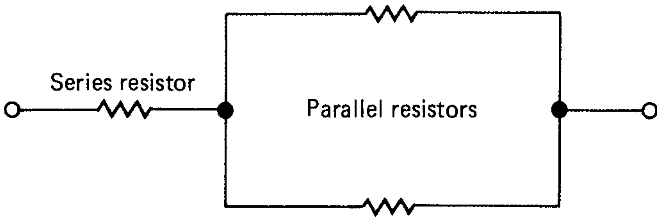

Figure 4-10 Series-parallel circuit.

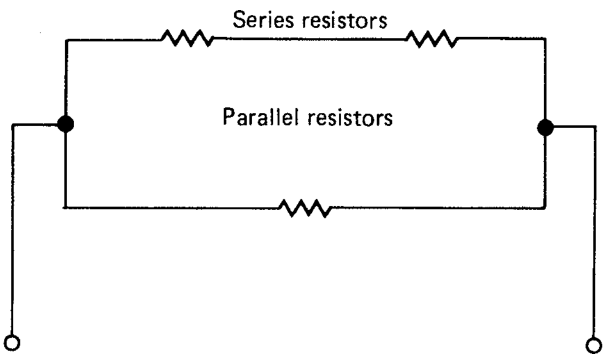
Series resistors
Parallel resistors

Figure 4-11 Alternate series-parallel circuit.

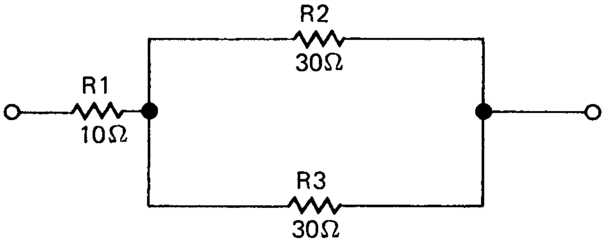
R2
30Ω
R1
10Ω
R3
30Ω

Figure 4-12 Series-parallel circuit showing resistance values.

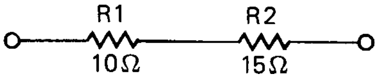
R1
10Ω
R2
15Ω

Figure 4-13 Series equivalent of series-parallel circuit.

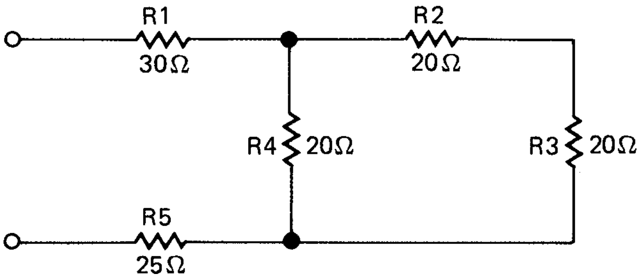
R1
30Ω
R2
20Ω
R4
20Ω
R3
20Ω
R5
25Ω

Figure 4-14 Schematic of complex series-parallel circuit.

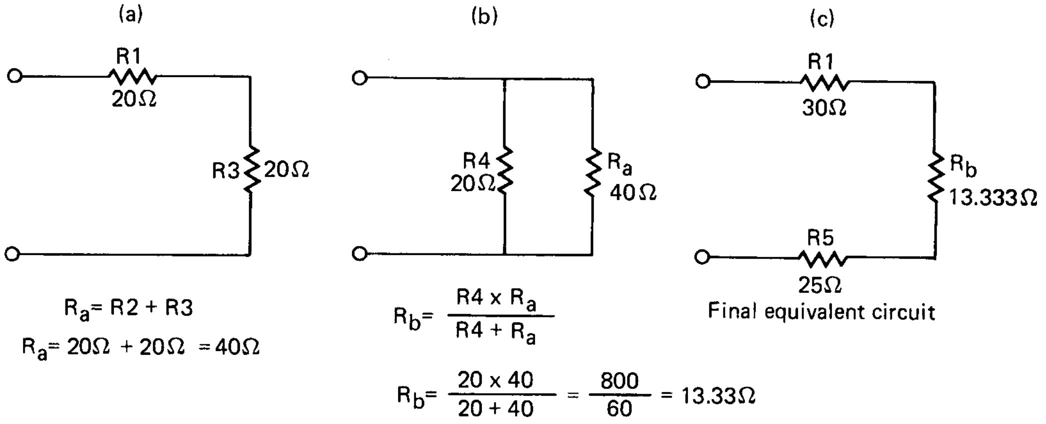

Figure 4-15 Redrawn schematic for determining resistance in series-parallel circuit.

In drawing A of Fig. 4-15, the values of two 20-ohm resistors in series are added together. The total value for these two resistors is 40 ohms. Drawing B then shows this value as part of a parallel circuit. This parallel circuit includes the combined values of the two series resistors. These have been redrawn in parallel with a 20-ohm resistor (R_4). The effect is to combine the series resistors within the parallel portion of the circuit from the original drawing. The 20- and 40-ohm parallel resistors have a total resistance of 13.333 ohms, In diagram C, this value is included as part of a series. The three resistors in the final drawing have values of 30, 13.333, and 25 ohms. Adding these, the total resistance for the circuit is 68.333 ohms. The rules for determining resistance can be summed up as follows:

1. When necessary for clarity, you can redraw the series-parallel circuit.
2. Solve all branches with series resistors first by adding their values.
3. Find the equivalent, or total, resistance of those resistors connected in parallel.
4. After you have series values for all resistors, add all the resistance values.

Determining current in series-parallel circuits

To determine current in a series-parallel circuit, follow the current path through the entire circuit. Remember your rules of current: In a series circuit, current is the same through all resistors. In a parallel circuit, you divide resistance into voltage to determine the current.

Current flow in a series-parallel circuit is shown in Fig. 4-16. The calculations for amperage in this circuit are included in the following section. They are part of the discussion and calculations of voltage values for the same circuit.

Determining voltage in series-parallel circuits

To find the voltage drop across resistors in series-parallel circuits, follow the rules you already know. When resistors are in series, add voltages. When resistors are in parallel, voltage is the same across all resistors.

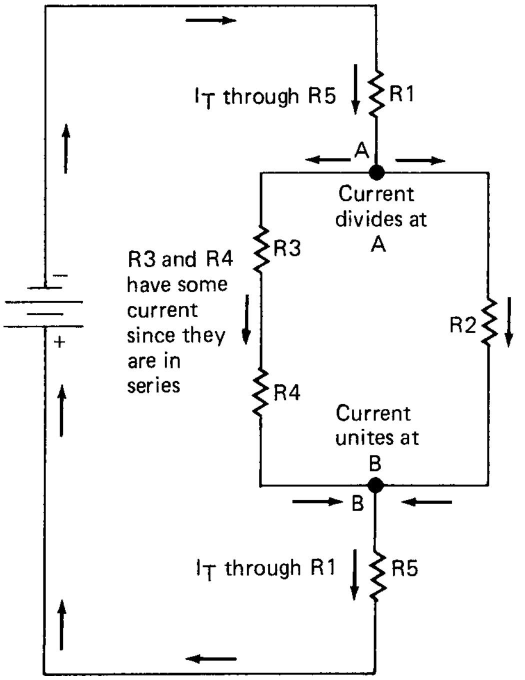

Figure 4-16 Current division in series-parallel circuit.

Examine Fig. 4-16 and see how the current divides; then look at Fig. 4-17. This figure provides total voltage at the source and resistance values for all resistors. To find values for voltage and current, here are the simple steps to follow:

1. Redraw the circuit for clarity, if necessary. Follow the steps illustrated in Figs. 4-18, 4-19, and 4-20.
2. Locate all branches with series resistors. This has been done in drawings 1 and 2 of Fig. 4-18.
3. Locate all branches with parallel resistors. This has been done in drawing 3 of Fig. 4-19.
4. Locate all branches that have series resistors in series with other resistors. Look back at Fig. 4-17. Resistors 1, 2, and 5 are in series. These are shown in loop 1 in Fig. 4-19. Resistors 1, 3, 4, and 5 are also in series. These are shown in loop 2 in Fig. 4-20.
5. Apply the rules for series or parallel circuits to calculate voltage and current.

To figure out current, divide total resistance into total voltage. Total voltage is given as 100 volts. Resistance values are shown in Fig. 4-17. Look at the resistance for the parallel portion of this circuit first. Resistors 3 and 4 are in series. These values of 10 ohms each are added for a total of 20 ohms. This gives the effect of two 20-ohm resistors in parallel. Two 20-ohm resistors in parallel give a total resistance value of 10 ohms. This value (10 ohms) is then added in

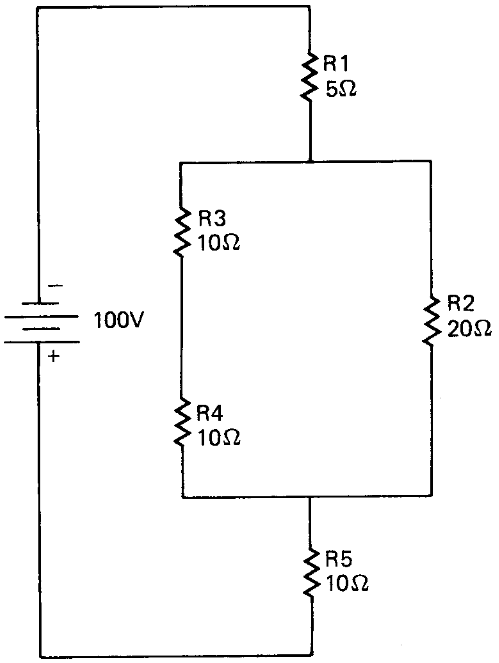
R1
5Ω
R3
10Ω
−
100V
+
R2
20Ω
R4
10Ω
R5
10Ω

Figure 4-17 Complex series-parallel circuit with resistance values given.

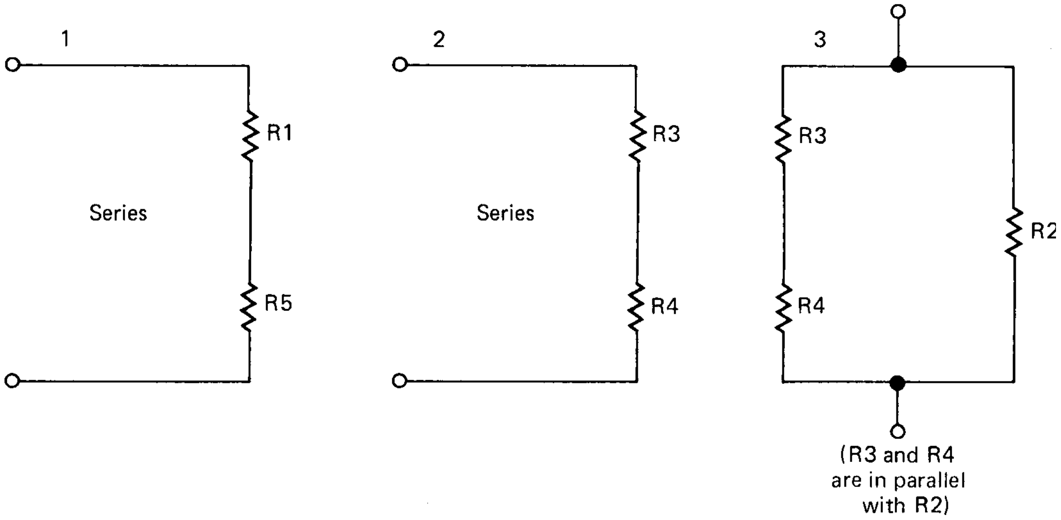
1
R1
Series
R5
2
R3
Series
R4
3
R3
R2
R4
(R3 and R4 are in parallel with R2)

Figure 4-18 Complex series-parallel circuit redrawn in three branches.

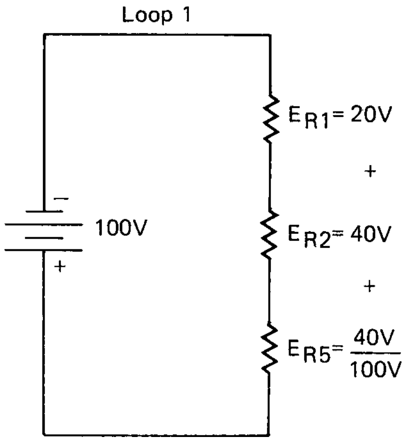
Loop 1
100V
E_{R1}= 20V
+
E_{R2}= 40V
+
E_{R5}= 40V
100V

Figure 4-19 Voltages in loop 1 of series-parallel circuit.

series to the values of resistors 1 and 5. The values added are 5 ohms, 10 ohms, and 10 ohms, for a total of 25 ohms. Dividing 25 ohms into 100 volts (using Ohm's law) produces an answer of 4 amperes.

$$R_3 + R_4 = 10\Omega + 10\Omega = 20\Omega$$

$$20\Omega \div 2 = 10\Omega$$

$$\text{Resistance of parallel circuit} = 10\Omega$$

$$10\Omega + R_1 + R_3 = 10\Omega + 5\Omega + 10\Omega = 25\Omega$$

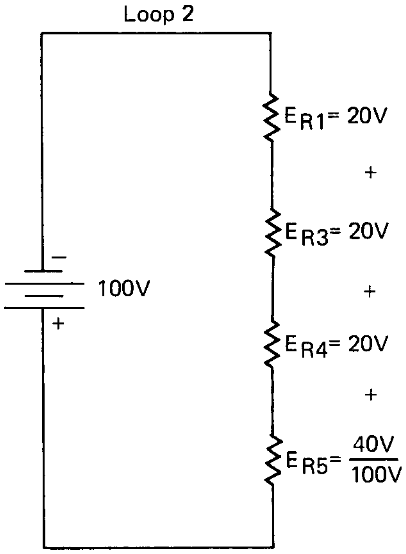
Loop 2
100V
E_{R1}= 20V
+
E_{R3}= 20V
+
E_{R4}= 20V
+
E_{R5}= 40V
100V

Figure 4-20 Voltages in loop 2 of series-parallel circuit.

Total resistance for the circuit is 25 ohms. Now, to determine current, divide the voltage (100 V) by the resistance.

$$100\text{V} \div 25\Omega = 4\text{A}$$

The current through the circuit is 4 amperes. This current is divided through the parallel portion of the circuit. So, current at resistors 2, 3, and 4 is 2 amperes. At resistors 1 and 5, the current is 4 amperes.

Note that resistors 1 and 5 are repeated in the series drawings in Figs. 4-19 and 4-20. This is important to remember: In determining voltage in a series-parallel circuit, reduce the branches to an equivalent series value. In doing so, consider all resistors in each series.

To determine voltage for each resistor in a series branch, multiply the current by the resistance (Ohm's law: $E = I \times R$). Resistor 1 has a rating of 5 ohms and a current of 4 amperes. Use Ohm's law again and get $4 \times 5 = 20$. The voltage is 20. Resistor 2 has 20 ohms of resistance and a current of 2 amperes. Using Ohm's law again, you get 40 volts for R_2. Resistor 5 (R_5) has 10 ohms with 4 amperes of current. This produces 40 volts for R_5. R_3 and R_4 each has 10 ohms resistance. These resistors have 20 volts across each.

You have been working with series, parallel, and series-parallel circuits. They are the basic circuits of anything electrical or electronic. Seeing how voltage, current, and resistance behave in these circuits brings about an understanding of circuits and improves your ability to read schematics. Schematics are symbols placed in series or parallel configurations. They are the shorthand of the electrical field. Every technician should be able to read a schematic and be able to identify how the components are connected in a circuit.

Review Questions

1. What is a series circuit?
2. What is a fuse? How is it used?
3. What is a parallel circuit?
4. What is the formula for finding total resistance in parallel?
5. What is the formula for finding total resistance in series?
6. How do you determine current in a parallel circuit?
7. What is a series-parallel circuit?
8. How is the total resistance found in a series-parallel circuit?

Chapter

5

Magnetism, Solenoids, and Relays

Performance Objectives

- Know the basics of magnetic induction.
- Know how electromagnets are made.
- Know how a solenoid differs from a relay.
- Know the primary purpose of an electrically operated solenoid valve.
- Know the difference between NC and NO.
- Know what happens when a valve leaks hot gas in a hot-gas defrost system.
- Know the usual voltage ratings for solenoid valves.
- Know what VA stands for and why you need to know it.

Magnetism is something everyone knows about and has worked with, even when very young. Magnetism is the property of certain materials that permits them to produce or conduct magnetic lines of force.

Magnetic lines of force surround a current-carrying conductor or any atomic particles in motion. Even in the atom, magnetic fields result from the spinning motions of the electrons. Such magnetic lines of force can also interact with electric fields or other magnetic fields.

Every time you press a doorbell button, an electromagnet causes it to ring or chime. Electromagnets are created by sending electricity through a coil. Permanent magnets are made from materials that will retain magnetism. That is, they keep their ability to attract iron (and other materials). They do not need electricity to function. Electromagnets lose their magnetism when the electric current is removed. Both permanent and temporary (electromagnet) magnets have their use in heating, air-conditioning, and refrigeration control circuits.

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

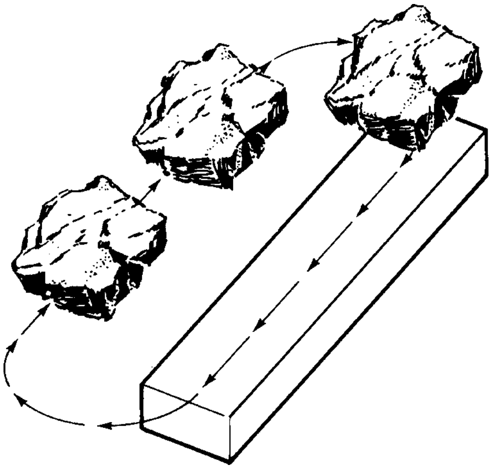

Figure 5-1 Making a permanent magnet by rubbing it with a lodestone.

Permanent Magnets

One type of permanent magnet is the natural magnet, the lodestone. It is also possible to make permanent magnets. For example, if you stroke a piece of high-carbon steel with a lodestone, the steel becomes magnetized (see Fig. 5-1).

Under normal conditions, the steel stroked with the lodestone keeps its magnetism permanently. When this happens, the molecules that form the steel bar line up in the same direction (see Fig. 5-2). The molecules that line themselves up in this way point toward the ends, or poles, of the magnet. Each magnet has two poles, which are identified as north and south.

A permanent magnet can be used to create other magnets. Figure 5-3 shows a permanent magnet used to stroke a bar of high-carbon steel. This creates a second magnet.

Magnets made of high-carbon steel sometimes lose their magnetism. This happens if the magnetized steel bar receives a strong shock, for instance, if the bar is dropped or struck with a hammer. When this happens, the molecules lose their alignment. This leads to a loss of magnetism. Magnetism can also be lost if a magnet is heated to high temperatures.

A magnet may be formed by rubbing a piece of steel with another magnet. This is called *magnetic induction*. This was the only way known of making magnets until electromagnetism was discovered.

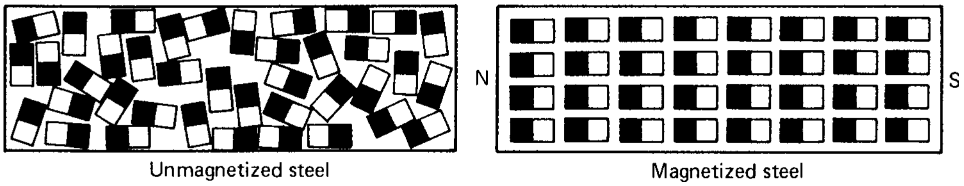

Figure 5-2 A magnetic field aligns iron molecules to produce a permanent magnet.

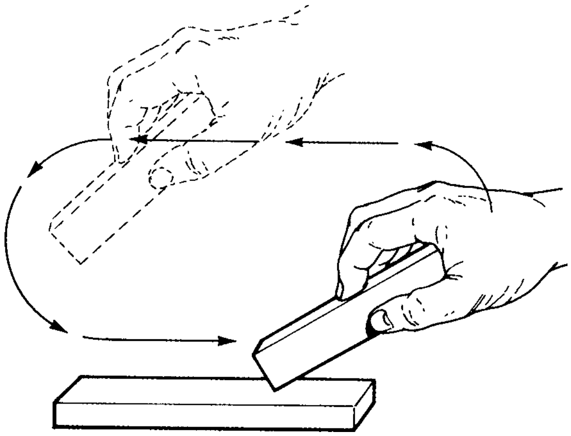

Figure 5-3 Permanent magnet used to make another permanent magnet.

Temporary Magnets

A bar of low-carbon steel attracts iron particles when it is in direct contact with another magnet. For example, you can place a lodestone against one pole of a bar of low-carbon steel as shown in Fig. 5-4. When this is done, the bar attracts iron particles. Removing the lodestone removes the magnetism. Then the iron particles fall away. The same happens if you place a permanent magnet against one end of a bar of low-carbon steel.

Low-carbon steel does not retain, or keep, magnetism. This material can serve as a temporary magnet. But it cannot be used for a permanent magnet.

Electromagnets

Magnetism also results when electric currents are passed through iron materials. Electromagnetism is important to electricity.

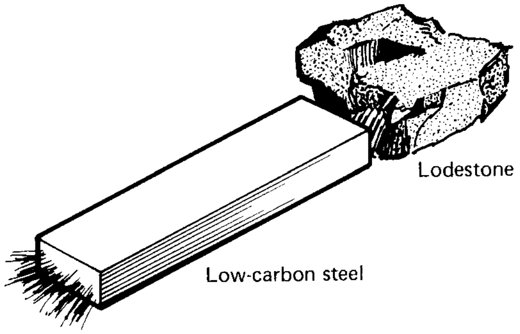

Figure 5-4 A temporary magnet made by placing the lodestone next to a piece of low-carbon steel.

Magnetic Theory

It is important to know the basics about how magnetism behaves for you to understand how certain pieces of equipment operate. Scientists still do not know exactly how magnets behave.

There are no laws or fixed rules for magnetism. Instead, there are a number of theories. These theories explain the behavior patterns that can be observed.

Magnetic Permeability

The ability of a material to become magnetized is called *permeability*. The magnetic force that attracts iron materials is called *magnetic flux*. The greater the permeability of a material, the higher its magnetic flux.

Numbers are used to indicate the permeability of materials. The higher the number, the more easily a material is magnetized. The relative permeabilities of a number of materials are listed in Table 5-1. At lower levels, materials are considered to be nonmagnetic. This applies to aluminum and all materials listed above it in Table 5-1.

Materials with higher ratings, including nickel, cobalt, iron, and permalloy, have high permeability. Permalloy, a combination of materials, has extremely high permeability. When a bar of permalloy is held in a north-south direction, it becomes a magnet. When the bar is turned to face east and west, it loses the magnetism.

Shapes of Magnets

Magnets can be made in a variety of shapes. Some of these are shown in Fig. 5-5. One common shape for magnets is the rod or bar. Another shape is similar to the letter U. This is sometimes called a horseshoe magnet.

Magnets made in a horseshoe shape have an advantage: Their ends, or poles, are close together. Thus, horseshoe magnets have a strong field of magnetic attraction, or flux field (see Fig. 5-6). By comparison, a bar magnet of the same strength has a flux field with less attraction. This is demonstrated in Fig. 5-7.

TABLE 5-1 Selected Permeabilities

Bismuth	0.999833
Quartz	0.999985
Water	0.999991
Copper	0.999995
Liquid oxygen	1.00346
Oxygen (STP)*	1.0000018
Aluminum	1.0000214
Air (STP)	1.0000004
Nickel	40
Cobalt	50
Iron	7,000
Permalloy	74,000

*STP, standard temperature and pressure.

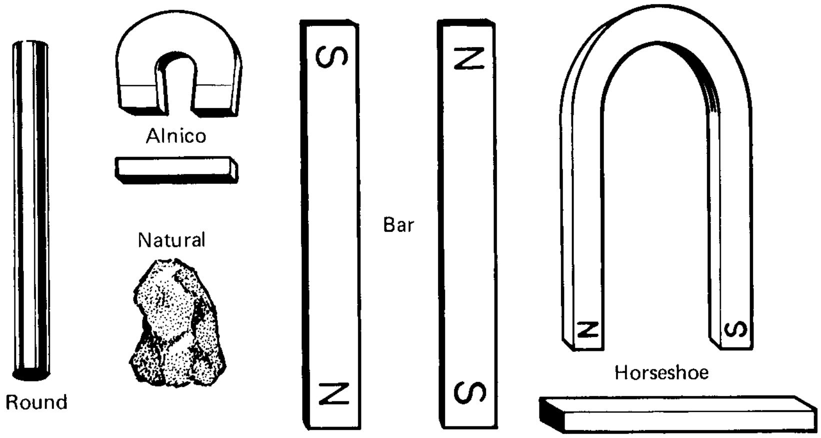
Round
Alnico
Natural
S
N
Bar
N
S
N
S
Horseshoe

Figure 5-5 Various shapes of magnets.

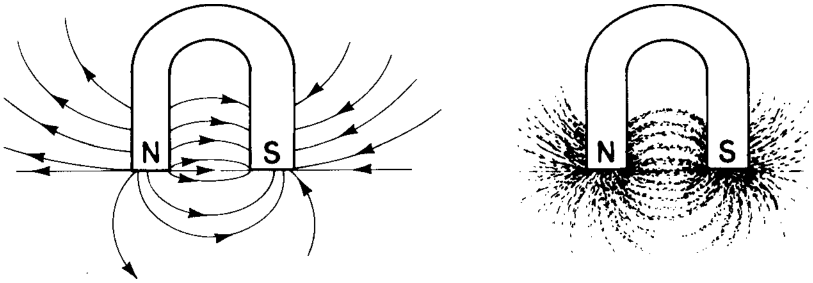

Figure 5-6 Horseshoe-magnet flux pattern.

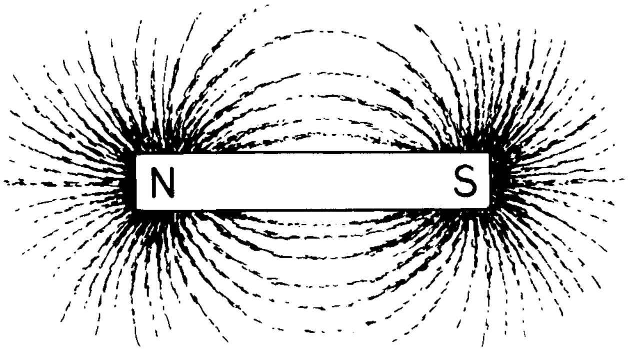

Figure 5-7 Bar magnet flux field.

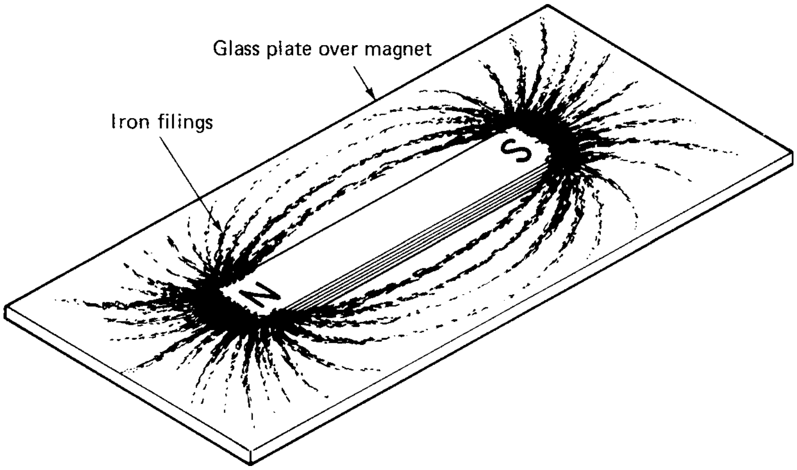

Figure 5-8 Magnetic flux is strongest at the poles.

The horseshoe magnet comes in handy when we start studying electrical measurement instruments.

Poles of magnets

Dip a bar magnet into a pile of iron filings. When it is withdrawn, the filings will gather at the ends of the magnet (see Fig. 5-8). This demonstrates that magnetism is strongest at the ends. These are the poles of a magnet.

Each magnet has two poles. However, there is a difference between the poles. If you suspend a magnet from a string, it turns until one pole points north and the other south (see Fig. 5-9). (This will occur if you do not have any large deposits of steel around, such as steel storage cabinets or a building with a steel framework.) This is the principle behind the operation of a compass: One pole of a magnet is north-seeking. A bar magnet is suspended on a bearing to reduce friction. The north-seeking pole points north and the other pole points south. Therefore, the poles of a magnet are referred to as the north pole and the south pole.

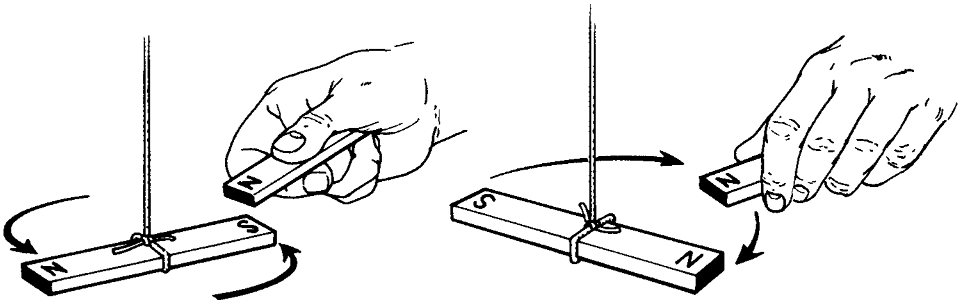

Figure 5-9 Magnetic poles align to north and south.

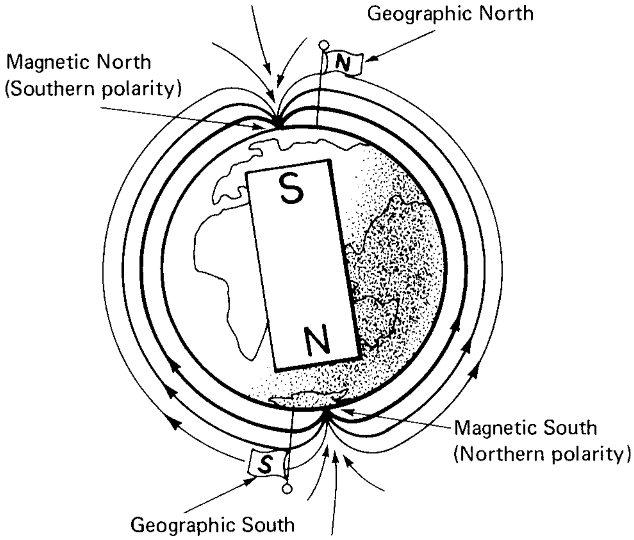

Figure 5-10 Polar attraction of magnets.

If you point the north poles of two bar magnets at each other, they are not attracted. Actually, there is a reverse force pushing the magnets apart. This force repels them from each other. This is a rule of magnetism: Like poles repel.

If you place the north pole of one magnet near the south pole of another, there is a strong attraction. This follows another rule of magnetism: Unlike, or opposite, poles attract. This means that the north-seeking pole of a magnet is actually the south pole (see Fig. 5-10). The pattern of attraction around a bar magnet is shown in Fig. 5-11.

Electromagnetism

The relationship between electric current and magnetism was discovered by Hans Christian Oersted in 1820. This relationship is direct. When current flows through a conductor, there is a magnetic field around the conductor. The direction of the current flow determines the force field of an electromagnet. Currents flowing in the same direction set up connecting fields of force (see

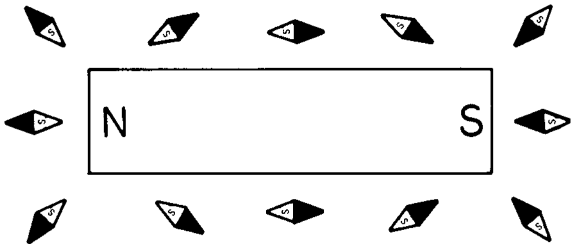

Figure 5-11 Pattern of attraction around a bar magnet.

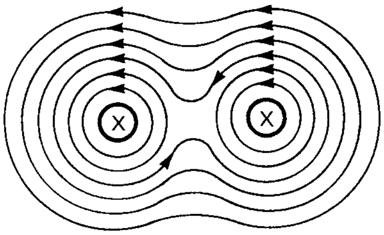

Figure 5-12 Current flow creates an electromagnetic field.

Fig. 5-12). Currents flowing in opposite directions set up repelling force fields (see Fig. 5-13).

The polarity of an electromagnet is determined by the direction of the current (see Fig. 5-14). This is important in an electrical meter.

The strength of an electromagnetic field is proportional to the current flowing through the conductor. More current means more magnetism.

Magnetism in a coil of wire

The magnetic field that surrounds a conductor depends on the form into which the wire is shaped (see Fig. 5-15). The magnetic field surrounding a single loop of wire is shown in Fig. 5-15a. Additional loops are shown in Fig. 5-15b. These loops form a spiral or helix. This shape is also known as a helical coil. Figure 5-15b shows magnetic flux around loops of wire wound next to one another. The more loops in the wire, the stronger the magnetic field becomes. This is shown in Fig. 5-15c.

The strength of an electromagnet or coiled conductor depends on the amount of current flowing in the coil and the number of turns of wire. Keep in mind that the direction of current flow determines the magnetic polarity (see Figs. 5-13 and 5-14).

Electromagnets

The usual form of an electromagnet is a coil of wire wound around a soft iron core. Thus, the core will lose its magnetism when the source of energy is

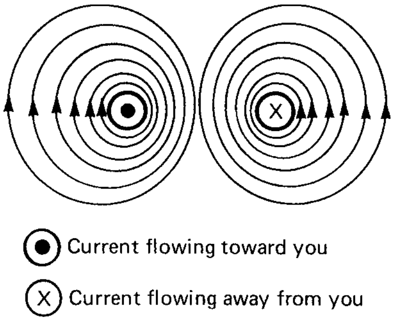

Figure 5-13 Currents in opposite directions create repelling fields.

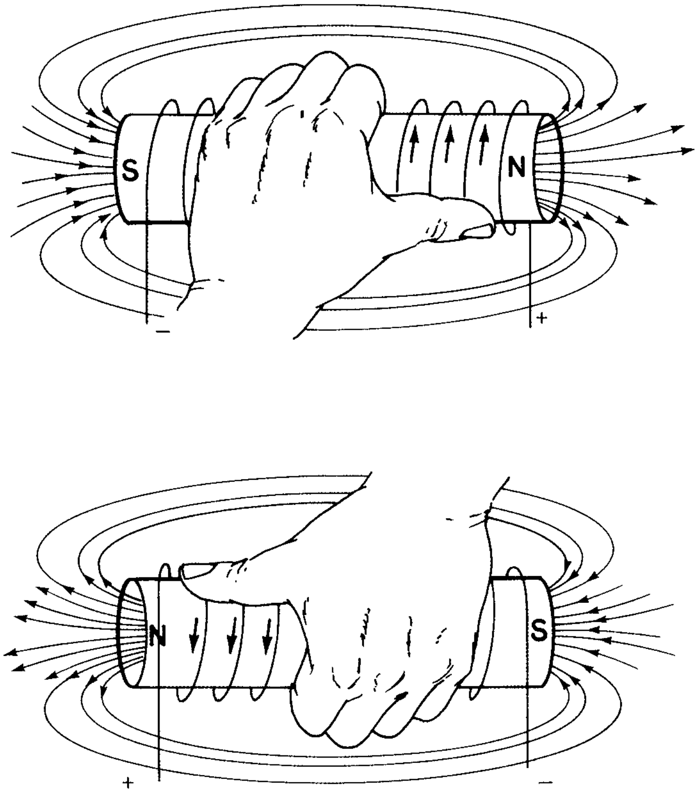

Figure 5-14 The left-hand rule shows how the polarity of an electromagnet is determined by the current flow. Grasp the coil with your fingers pointing in the direction of the current flow: the thumb of the left hand then points toward the north pole.

removed (see Fig. 5-16). The iron core provides an easy path for the magnetic field created by the coil. An electromagnet with an iron core forms a stronger magnet than the same coil of wire without the iron core.

The size of the iron core can help determine the strength of an electromagnet. For full strength, the core should be large enough to absorb all the magnetism

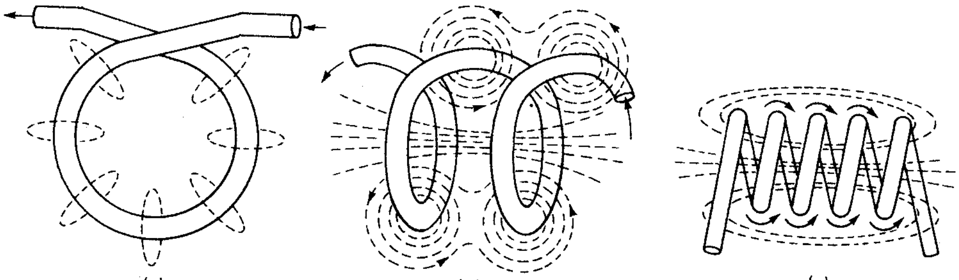
(a) (b) (c)

Figure 5-15 Magnetic fields form around various shapes of magnetic coils.

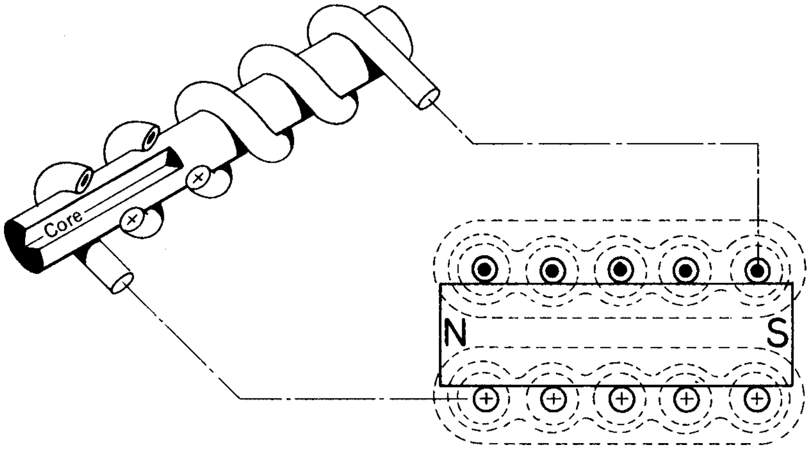

Figure 5-16 Electromagnets can be made by winding wire around a soft iron core.

from the coil. When the coil creates more magnetism than the core can absorb, it is called saturation. If the core is too small, there is too much magnetism to be absorbed. All the magnetism of the coil can be used when the core has a capacity slightly larger than the magnetic flux created.

Using electromagnetism

The generation and use of electricity are directly related to electromagnetism. Some of these uses are in the solenoid and relay. These two items are useful in heating, air-conditioning, and refrigeration equipment control circuits. It is important to understand how they work, since they can be the source of much troubleshooting.

The Solenoids and Relays

A magnetic field seeks the path of minimum reluctance, just as an electric current seeks the path of least resistance. Reluctance and resistance are related. Resistance refers to the opposition to current flow in a circuit. Reluctance refers to the opposition of the flow of a magnetic field. The lower the reluctance, the greater is the attraction of materials to a magnetic field.

A solenoid is a device that uses these principles. A helical coil of wire produces a magnetic field. An iron core fits loosely within the coil of wire. When the current is off, the core rests outside the area of the coil. When the current is applied, the core is sucked into the coil. This is referred to as the sucking effect of a coil (see Fig. 5-17).

The sucking effect is often used in devices that require a small amount of physical movement. One type of solenoid is shown in Fig. 5-18. A common use for solenoids is in devices called solenoid valves (see Fig. 5-19). This valve is a device that opens and closes to permit liquids or gases to flow. In most valves, for

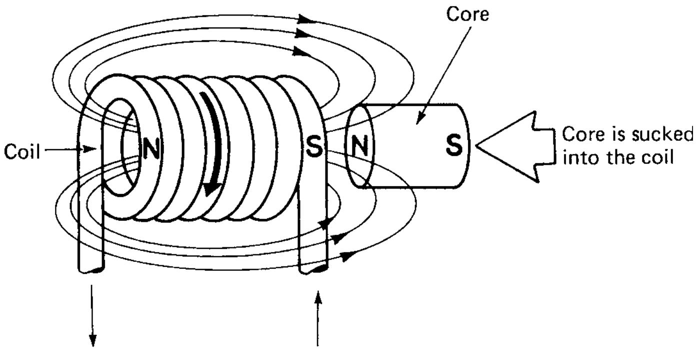

Figure 5-17 The sucking effect of a coil.

example, you turn a handle or knob to start or stop the flow of gas. Solenoid valves are used widely as safety devices. They are found in gas lines, air lines, and water lines. Many types are used in air-conditioning, refrigeration, and heating systems.

The solenoid in the valve in Fig. 5-19 is closed when there is no current through its coil. The valve stem is held in a closed position by a spring that applies light pressure. When the valve is closed, gas cannot flow. Current is applied when a person or thermostat turns on a heater. The current draws the movable core, or plunger, into the coil. This opens the gate to the flow of gas. When the current is turned off, the spring moves the valve system back into a closed position.

This same principle is used widely in solenoid-type relays. A solenoid relay is like an electrically operated switch. The coil in this type of relay pulls a core piece that has a number of electrical contacts. The contacts are designed so that they may either close or open electrical circuits. The relay contacts themselves can be designed to handle large amounts of current. But the coil of the control solenoid may operate on only a fraction of an ampere. The effect is that low-voltage, low-current electricity is used to control the flow of larger amounts of electricity.

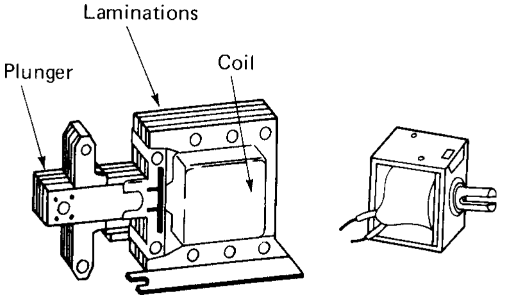

Figure 5-18 Two types of solenoids.

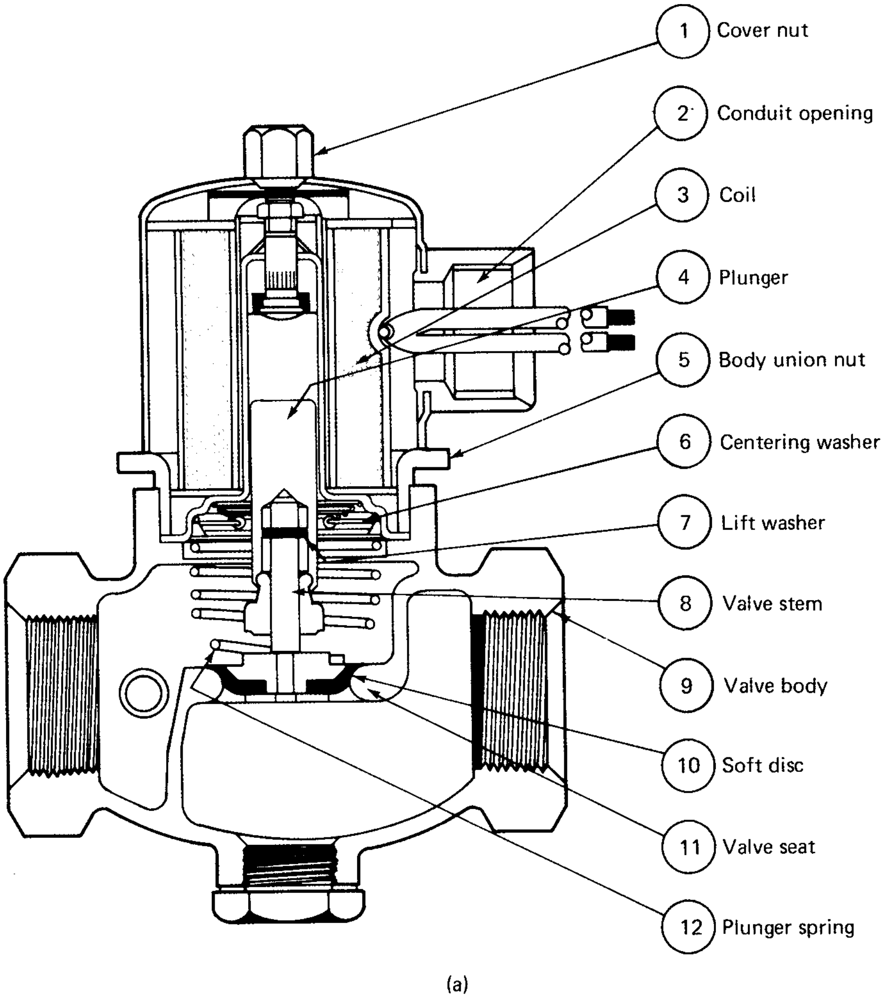
1 Cover nut
2 Conduit opening
3 Coil
4 Plunger
5 Body union nut
6 Centering washer
7 Lift washer
8 Valve stem
9 Valve body
10 Soft disc
11 Valve seat
12 Plunger spring
(a)

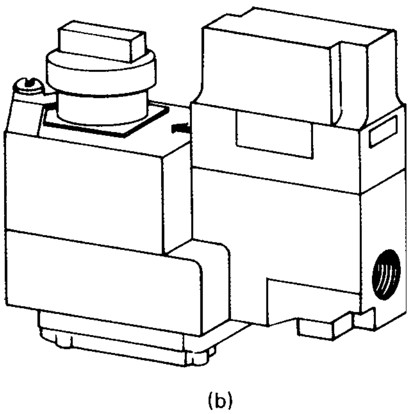

Figure 5-19 (a) Cutaway view of a solenoid valve. (b) Solenoid valve used for hot air furnace gas control. (c) Relay used to start a refrigerator or compressor. (d) Schematic for a refrigerator.

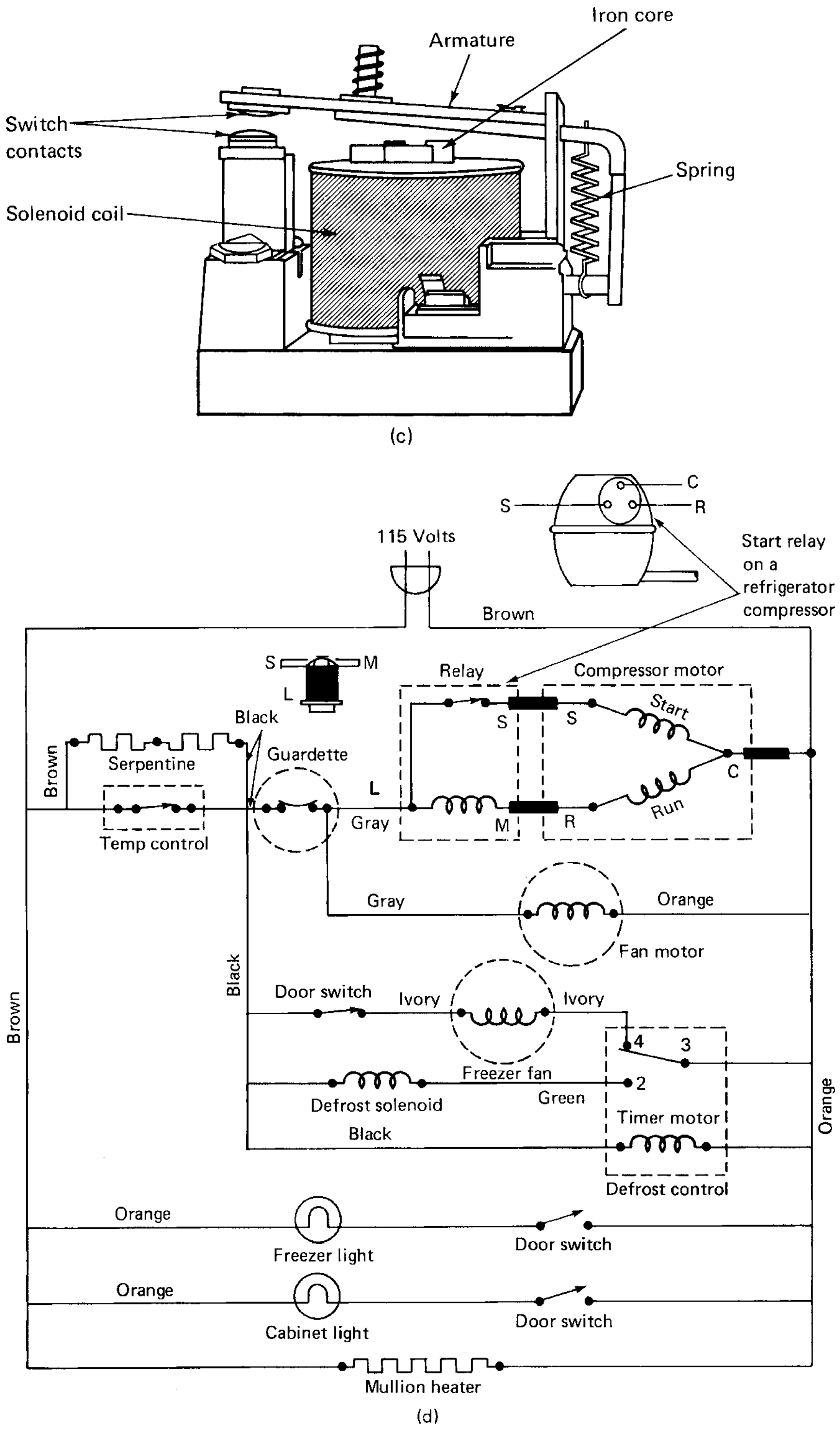
Iron core
Armature
Switch contacts
Solenoid coil
Spring
S
C
R
115 Volts
Start relay on a refrigerator compressor
Brown
S
M
L
Black
Relay
Compressor motor
Start
S
S
C
Serpentine
Guardette
Brown
L
Temp control
Gray
M
R
Run
Gray
Orange
Fan motor
Black
Brown
Door switch
Ivory
Ivory
4
3
2
Freezer fan
Green
Defrost solenoid
Timer motor
Black
Orange
Defrost control
Orange
Freezer light
Door switch
Orange
Cabinet light
Door switch
Mullion heater

(c)

(d)

Figure 5-19 (*Continued*)

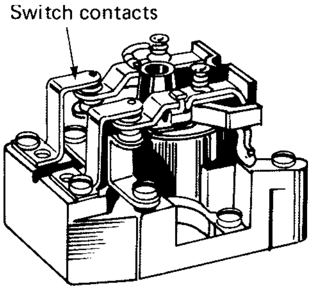

Figure 5-20 Heavy-duty power relay.

Power relays

A heavy-duty power relay that operates from solenoid action is shown in Fig. 5-20. In this device, a spring pulls the solenoid core, or armature, away from the electrical contact when the current is off. When current flows, the armature is pulled toward the coil. An electrical contact connected to the armature is either closed or opened by this action. Note that the electromagnetism can be used either to open or close the relay. The action taken depends on the design of the relay. The relay contacts can be arranged for a variety of functions, such as SPST, SPDT, DPDT, or any other combination.

The advantage of a relay is that a substantial pulling power can be developed with a small coil current. The contacts can be made quite large and can handle or switch high values of electrical power. An extremely small amount of control power thus can be used to switch much higher voltages and currents in a safe manner (see Fig. 5-21).

Solenoid valves

The primary purpose of an electrically operated solenoid valve is to control automatically the flow of fluids, liquid, or gas. There are two basic types of solenoid valves. The most common is the normally closed (NC) type, in which the valve opens when the coil is energized and closes when the coil is de-energized. The other type is the normally open (NO) valve, which opens when the coil is de-energized and closes when the coil is energized (see Fig. 5-22).

Principles of operation

Solenoid valve operation is based on the theory of the electromagnet. The solenoid valve coil sets up a magnetic field when electrical current is flowing through it. If a magnetic metal, such as iron or steel, is introduced into the magnetic field, the pull of the field will raise the metal and center it in the hollow core of the coil. By attaching a stem to the magnetic metal plunger, this principle is used to open the port of the valve. When the electrical circuit to the coil is broken, the magnetic field collapses and the stem and plunger either fall by gravity or are pushed down by the kick-off spring.

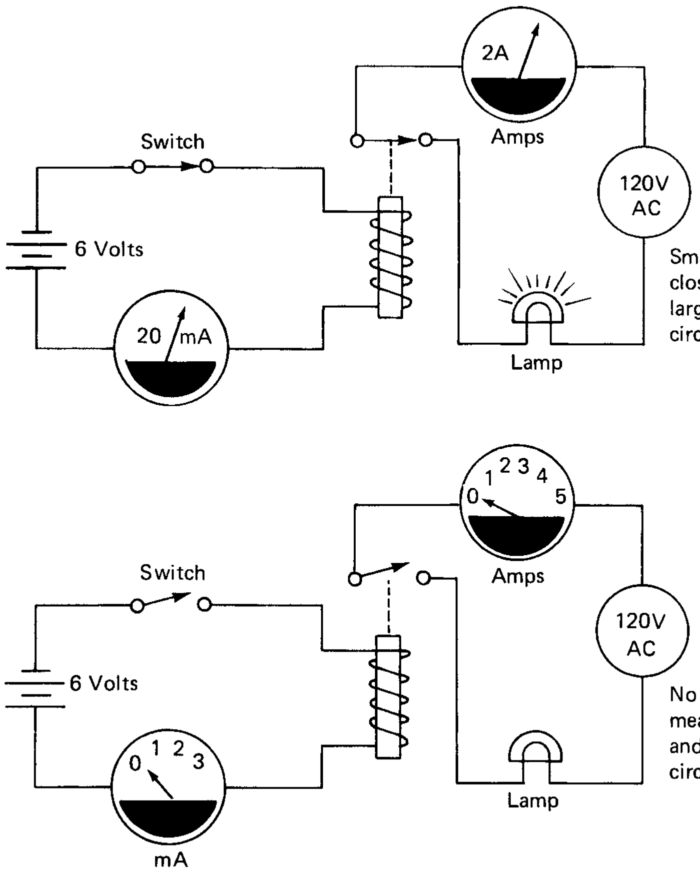

Small current in relay coil closes contacts that control large current in higher voltage circuit.

No current through the coil means relay contacts are open and no current flows in the circuit.

Figure 5-21 Relay in a circuit.

Some solenoid valves are designed with a hammer-blow effect. When the coil is energized, the plunger starts upward before the stem. The plunger then picks up the stem by making contact with a collar at the top. The momentum of the plunger assists in opening the valve against the unbalanced pressure across the port.

Applications

In many cases, valves are used for controlling the flow of refrigerants in liquid or suction lines or in hot-gas defrost circuits. They are equally suitable for many other less common forms of refrigerant control.

Liquid line service

The primary purpose of a solenoid valve in a refrigerant liquid line is to prevent flow into the evaporator during the off cycle. On multiple systems, a solenoid valve may be used in each liquid line leading to the individual evaporators.

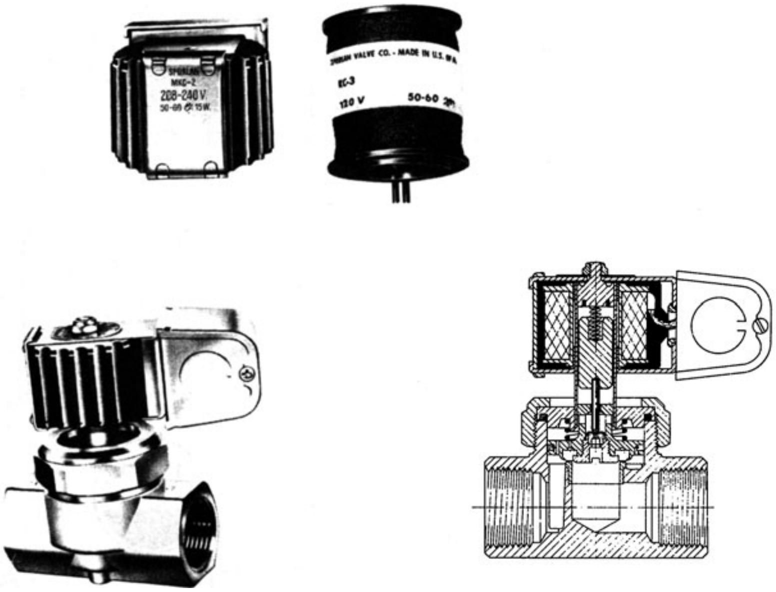
SPORLAN
MKC-2
208-240V
120 V
50-60

Figure 5-22 Solenoid coil and step-down transformer. (Courtesy of Sporlan)

The application of a liquid line solenoid valve depends mainly on the method of wiring the valve with the compressor control circuit. It may be wired so the valve is energized only when the compressor is running. This type of application is shown in Fig. 5-23.

Another application, known as pump-down control, uses a thermostat to control the solenoid valve (see Fig. 5-24 for a wiring and valve location schematic). When the thermostat is satisfied, the valve closes and the compressor continues to run until a substantial portion of the refrigerant has been pumped from the evaporator. A low-pressure cutout control is used to stop the compressor at a predetermined

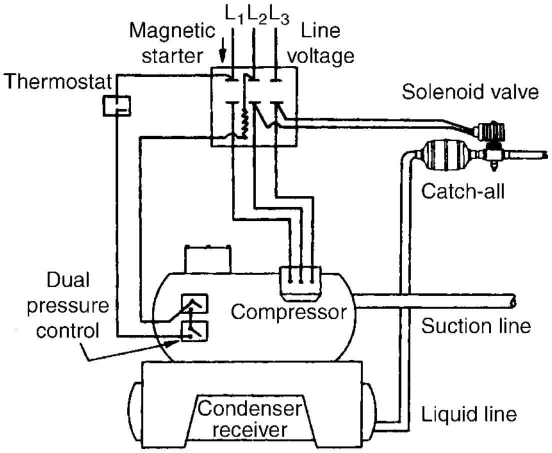
$L_1 L_2 L_3$
Magnetic starter
Line voltage
Thermostat
Solenoid valve
Catch-all
Dual pressure control
Compressor
Suction line
Condenser receiver
Liquid line

Figure 5-23 Liquid-line solenoid valve application.

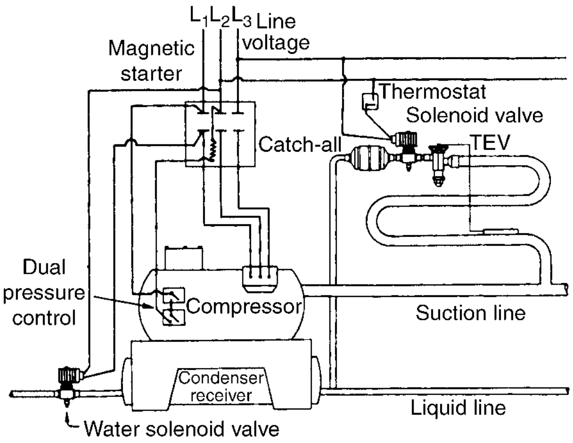

Figure 5-24 A pump-down control uses a thermostat to control the solenoid valve. (Courtesy of Sporlan)

evaporator pressure. When the thermostat again calls for refrigeration, the solenoid valve opens, causing the evaporator pressure to rise and the compressor to start. This arrangement can be used on either single or multiple evaporators.

Suction line service

There are several applications, particularly on suction lines, where pressure drops in the range of 2 to 4 pounds per square inch (psi) cannot be tolerated. Therefore, only valves that are capable of opening at very low pressure drops are suitable for this type of use. Some valves incorporate the floating disc principle and are ideally suited for such special applications. They are capable of opening full at pressure drops of 0.1 psi and below.

Larger-capacity valves are suitable for suction service when supplied with internal parts that are mechanically connected. With this arrangement, the piston is connected to the stem and plunger assembly, and when the coil is energized, the plunger assists in supporting the piston. As a result, the pressure drop through the valve is reduced to a bare minimum. Valves with the direct-connected assembly are designated usually with a prefix D to the type number.

When these valves are required for suction service, they are supplied with a spring support under the piston. The spring counterbalances the major portion of the piston's weight, and therefore the valve will open with far less pressure drop than normal. Valves with the counterbalance spring are identified by the prefix S added to the type number.

High-Temperature Applications

In some high-temperature applications, a high-temperature coil construction is required. The temperature of the fluid or gas flowing through the solenoid valve will generally determine whether a high-temperature coil is necessary.

Hot-gas defrost service

Several piping arrangements are used for hot-gas defrost systems, one of which is shown in Fig. 5-25. A portion of the compressor discharge gas is passed through the solenoid valve into the evaporator. The solenoid valve may be controlled either manually or automatically for this duty.

Hot-gas defrost valve selection requires a slightly different approach from the simple pressure drop versus tonnage. Be sure to consider the evaporator temperature correction factors to make certain that the valve selected has adequate capacity.

Normally open solenoid valves have many uses. Perhaps the most popular is their adaptation to heat-reclaiming systems. The use of one normally closed valve and one normally open valve to shunt the discharge gas to either the outdoor condenser or the indoor heat-reclaiming coil provides positive opening and closing action.

This eliminates the problem found in some three-way valves which have a tendency to leak hot gas into the heat-reclaiming coil when not required. When this leakage occurs during the cooling season, it imposes an extra load on the cooling system that wastes energy, rather than conserving it.

If leakage occurs during the heating season when all the discharge gas should be going to the reheat coil, a good portion of the liquid charge could become logged in the inactive condenser. For a simple schematic of a heat-reclaiming cycle, see Fig. 5-26. Many original equipment manufacturers (OEMs) have developed their own reheat cycle, which may be completely different from the one illustrated. In addition, some may incorporate head pressure control as well, so it is always advisable to consult the manufacturer's bulletin regarding its particular design.

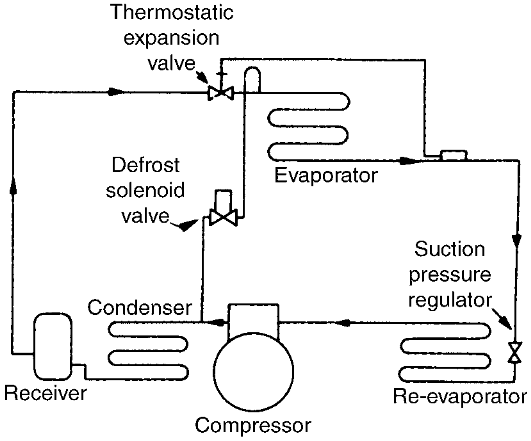

Figure 5-25 Solenoid valves used for hot-gas defrost. (Courtesy of Sporlan)

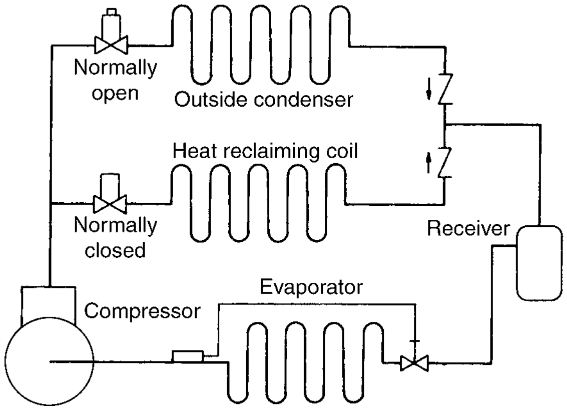
Normally
open
Outside condenser
Heat reclaiming coil
Normally
closed
Receiver
Evaporator
Compressor

Figure 5-26 Heat-reclaiming cycle schematic. (Courtesy of Sporlan)

Transformers for Low-Voltage Controls

The use of low-voltage control systems is becoming more widespread as the demand for residential air conditioning increases. This necessitates the use of a transformer for voltage reduction, normally to 24 volts. The selection of a transformer is not accomplished by merely selecting one that has the proper voltage requirements. The volt-ampere (VA) rating is equally important. To determine the VA requirements for a specific solenoid valve, refer to the manufacturer's data. Insufficient transformer capacity will result in reduced operating power or a lowering of the maximum operating pressure differential (MOPD) of the valve. MOPD and safe working pressure (SWP) are usually noted on the valve nameplate (see Fig. 5-27).

If more than one solenoid valve and/or other accessories are operated from the same transformer, the transformer VA rating must be determined by adding the individual accessories' VA requirements.

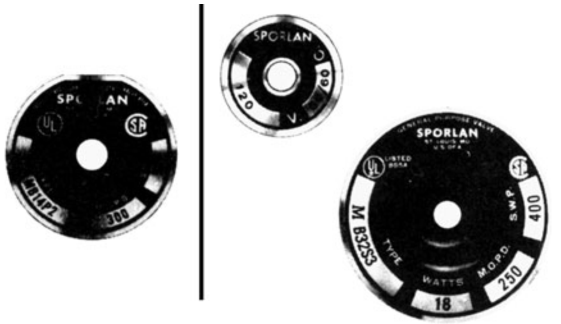
SPORLAN
120
V
60
SPORLAN
M B32S3
TYPE
WATTS
18
M.O.P.D.
250
S.W.P.
400

Figure 5-27 Valve nameplates with voltage, wattage, frequency, MOPD, and SWP. (Courtesy of Sporlan)

Review Questions

1. What is magnetism?
2. What is magnetic induction?
3. How is an electromagnet made?
4. What is a solenoid? Where is it used?
5. What is a relay?
6. How is a relay different from a solenoid?
7. What is meant by like poles repel and unlike poles attract?
8. Where are electromagnets used?
9. What is the primary purpose of an electrically operated solenoid valve?
10. What do NO and NC mean?
11. How are valves with a counterbalance spring identified?
12. What happens when a valve leaks hot gas in a hot-gas defrost system?
13. What is the usual voltage rating for solenoid valves?
14. What does VA mean?
15. Why do you need to know the VA rating of a transformer?

Chapter

6

Electrical Measuring Instruments

Performance Objectives

- Know how electricity is measured.
- Know how the d'Arsonval meter movement works.
- Know what is meant by a meter's full-scale deflection.
- Know how to correct for parallax error.
- Know the difference between milliampere, microampere, and ampere.
- Know how a meter's range is extended.
- Know how an ammeter is connected in a circuit.
- Know the difference between and uses for series and shunt ohmmeters.
- Know how to use a megger.
- Know how to use the voltmeter to test a circuit.

Electricity

Electricity cannot be seen, heard, tasted, or smelled. That is why it is necessary to use some type of device to detect its presence. Once you have detected its presence, you need to know how much of it is present. This is the function of electrical measuring instruments. Another reason for using an instrument is to allow you to detect the presence of electricity without having to feel it. The sense of feel is the only one of our five senses that can be used to detect the presence of electricity. This can be deadly also; therefore, a measuring device is much more reliable and safer.

It takes a meter to measure electricity. It takes a meter to determine the three properties of a circuit: voltage, current, and resistance. The main reason for using meters in heating, refrigeration, and air-conditioning work is to aid

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

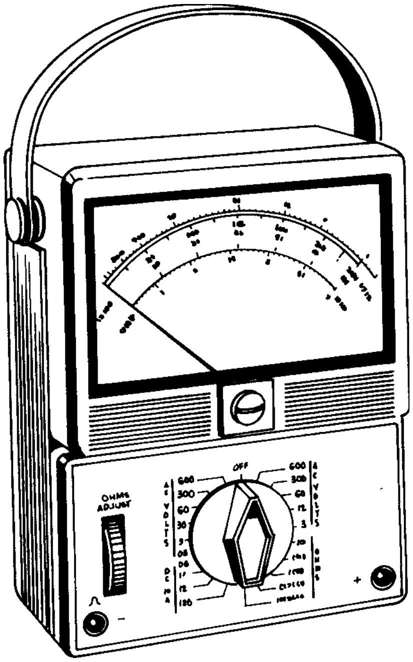

Figure 6-1 Multimeter.

in troubleshooting defective equipment. With this in mind, we will examine a number of types of meters. By looking a little closer at what is under the meter cover, you will better understand how to treat the device to make it last longer and produce the accuracy needed to do your job.

The most frequently used meter is the multimeter (see Fig. 6-1). It can be used to measure voltage, current, and resistance.

Types of Meter Movements

There are many types of meter movements. One of the most often used in the analog configuration is the d'Arsonval meter movement, named after its designer, Arsene d'Arsonval. Actually, it is a revision of the galvanometer. The galvanometer is a measuring device that detects the presence and direction of current flow in a circuit.

Edward Weston made some improvements on the design that d'Arsonval produced and the results are still seen in today's meters.

D'Arsonval meter movement

The d'Arsonval meter movement has a small, rectangular coil of wire. The coil is suspended in a magnetic field created by a permanent magnet. When current is applied to the coil, it becomes an electromagnet. The energized coil then lines up with the poles of the permanent magnet. The amount of current applied to the coil controls its movement. An electromagnet that is free to move will align its axis with the magnetic axis of a fixed magnet. The axis of an electromagnet is a straight line between its poles (see Fig. 6-2).

The coil must be free to rotate in order to align itself with the magnetic axis. The coil in the meter is mounted on pivots that permit easy rotation. Two small springs are mounted on the top and bottom. These springs offer slight resistance to the rotation of the coil. The springs control the position of the coil when there is no current flowing.

When current flows in the coil, it produces a magnetic field around the coil. This magnetic flux overcomes the force of the springs and moves the coil. A pointer on top of the coil rotates to mark the amount of movement. The greater the current through the coil, the more it turns, and the further the pointer moves. The pointer then stops in front of a marked scale on the face of the meter. The position of the pointer indicates the meter measurement (see Fig. 6-3).

A typical meter scale is shown in Fig. 6-4. Note the even divisions between each line on the scale. This is a linear-type meter. Every space indicates the same amount of voltage.

As you look closer at the d'Arsonval meter movement, you will find that it can be extremely sensitive to small currents. For instance, a relatively inexpensive meter can give a full-scale deflection of 10 microamperes (μA). Full-scale deflection is the total range of a meter scale. A microampere is one-millionth (0.000001) of an ampere.

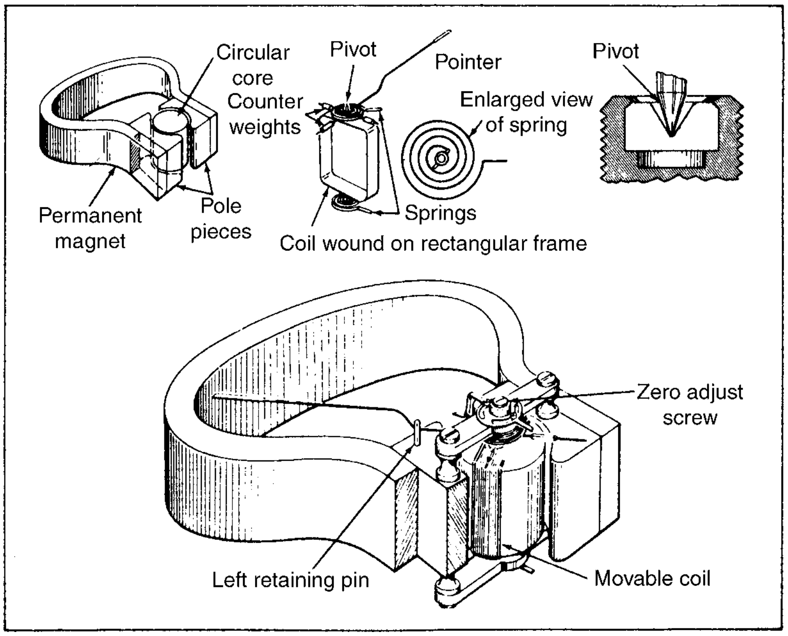
Circular core
Counter weights
Pivot
Pointer
Pivot
Enlarged view of spring
Permanent magnet
Pole pieces
Springs
Coil wound on rectangular frame
Zero adjust screw
Left retaining pin
Movable coil

Figure 6-2 d'Arsonval meter movement.

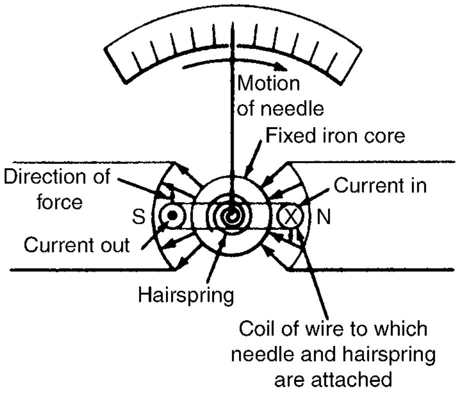
Motion
of needle
Fixed iron core
Direction of
force
Current in
S
N
Current out
Hairspring
Coil of wire to which
needle and hairspring
are attached

Figure 6-3 d'Arsonval meter movement with scale.

Meter deflection is determined by three factors:

1. The number of turns of wire in the coil and the amount of current flowing in the coil.
2. The strength of the magnetic field produced by the permanent magnet affects the positioning of the coil.
3. The tension of the springs and the friction of the bearings determine the sensitivity of the meter movement.

The amount of current flowing through the meter determines its pointer deflection since it is the only variable in the meter movement circuit. The scale of the meter is calibrated (marked) to show the type of reading (volts, ohms, or amperes). The scale in Fig. 6-4 is calibrated to read dc (direct current) volts.

Before you use a meter, check its instruction manual. Each meter is a little different and you should be aware of any important instructions in the instruction manual.

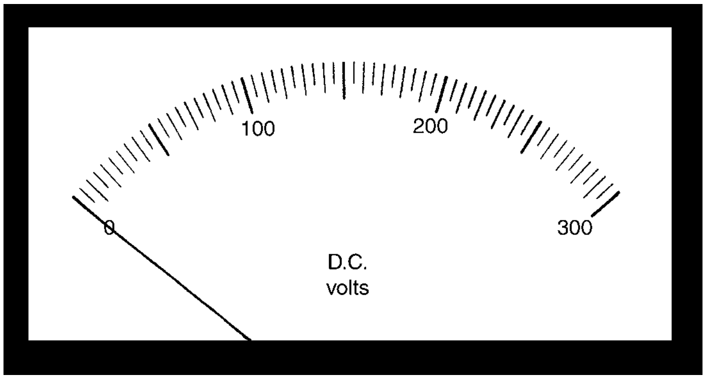
0
100
200
300
D.C.
volts

Figure 6-4 DC volts meter scale.

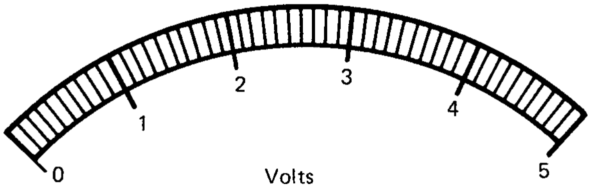

Figure 6-5 Linear scale.

Analog meters

An analog-type meter has a scale that uses a pointer to indicate its reading. The digital meter uses digits similar to a digital watch to indicate the voltage, current, or resistance. There is a place for both types. At this point, it is best to concentrate on the analog type since many of them are still in use and they are still made for all types of applications.

A scale with four separate readings is shown in Fig. 6-5. This shows what is known as a linear scale. A linear scale has marks spaced equally. The full-scale deflection is 5. This could mean 5 amperes, 5 volts, or 5 milliamperes, or 5 of any other value of voltage or current. The information on what is being measured usually appears below the scale. In this case (Fig. 6-5), it is volts.

A nonlinear scale is shown in Fig. 6-6. This particular scale is referred to technically as a square-law scale. The lines in this scale increase in squares. That is, each larger scale marking indicates that the measured value is multiplied by itself or squared. The spacing of nonlinear scales can be easily spotted since they are uneven.

Take a look at a multimeter scale (see Fig. 6-7). A multimeter usually has a number of scales. Each scale on a multimeter is defined individually.

Parallax Error

A parallax error results from the fact that there is a distance between the pointer and the scale behind it. If you look at the pointer from an angle, you will get an incorrect reading. From the side, the reading will be different than from directly in front. It is also possible to have a parallax error if you use both eyes. The best practice is to use one eye. Make sure your view is from directly in front of the pointer. Some meters have a small mirror attached to the scale to reduce parallax error (see Fig. 6-8). You minimize parallax error when you line up the pointer with the reflection in the mirror.

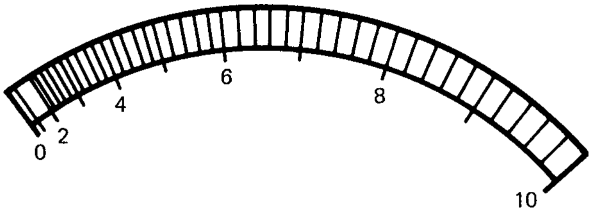

Figure 6-6 Nonlinear scale.

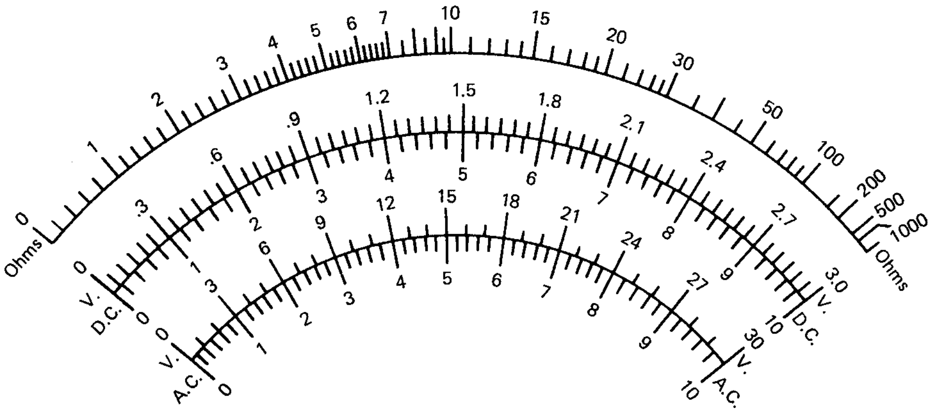
0
1
2
3
4
5
6
7
10
15
20
30
50
100
200
500
1000
Ohms
Ohms
0
.3
.6
.9
1.2
1.5
1.8
2.1
2.4
2.7
3.0
V.
D.C.
0
1
2
3
4
5
6
7
8
9
10
V.
D.C.
0
3
6
9
12
15
18
21
24
27
30
V.
A.C.
0
1
2
3
4
5
6
7
8
9
10
V.
A.C.

Figure 6-7 Multimeter scale.

Ammeters

The ammeter is used to measure current. It measures current in microamperes (μA), milliamperes (mA), or amperes. The microampere is very small (0.000001 ampere). The milliampere is also very small (0.001 ampere). In most instances, you will be working with milliamperes and amperes. Microamperes are usually found in circuits that use transistors and other semiconductor devices.

It is important to fit the meter to the job. You select a meter to fit the work you are doing and the range of currents you will be encountering on the job. Keep in mind that the 1-milliampere meter is used for measuring direct current (dc)

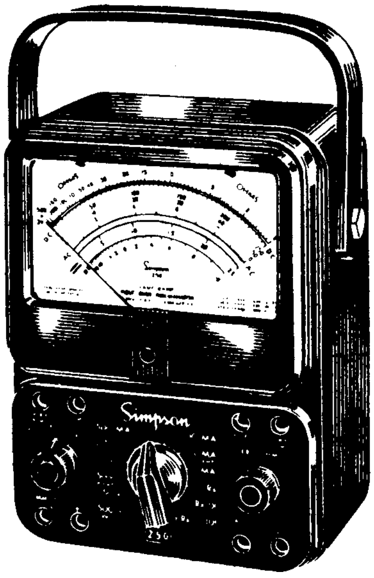
Simpson
260

Figure 6-8 Meter with mirrored scale.

of up to 1 milliampere. Such a meter does not have sufficient sensitivity to measure currents of less than 100 microamperes or 0.1 milliampere.

Keep in mind also that meters are most accurate in the middle 80% of their range. Accuracy falls off in the lowest and highest 10% of the scale.

Extending the range of an ammeter

The 1-milliampere meter movement can be used to measure direct currents that are larger than 1 milliampere. To do this, a resistor is connected in parallel with the meter movement. This resistor is called a shunt. To measure 2 milliamperes, the shunt requires a resistor that can pass 1 milliampere around the meter movement. Thus, 1 milliampere goes through the meter and 1 milliampere through the shunt (see Fig. 6-9).

Various-sized shunts can be connected, by switching, across the meter movement to increase its range from 1 milliampere to 1 ampere, 10 amperes, or even 100 amperes (see Fig. 6-10).

Connecting an ammeter in the circuit

An ammeter is always inserted in a circuit in series with the load.

Clamp-on ammeters

The clamp-on ammeter is shown in Fig. 6-11. It is inserted over a wire carrying alternating current (ac). The magnetic field around the wire induces a small amount of current in the meter. The scale is calibrated to read amperes or volts. Because the wire is run through the large loop extending past the meter movement, it is possible to read ac or current, without removing the insulation from the wire or opening the circuit to install the meter. This meter is very useful

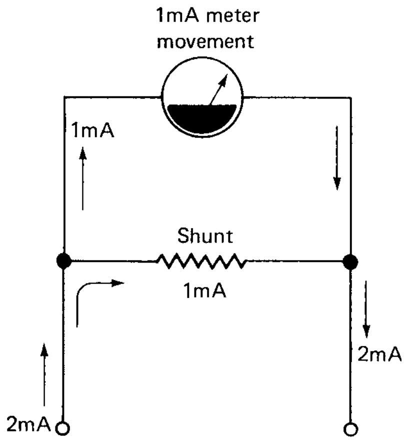

Figure 6-9 Shunting current around the meter movement.

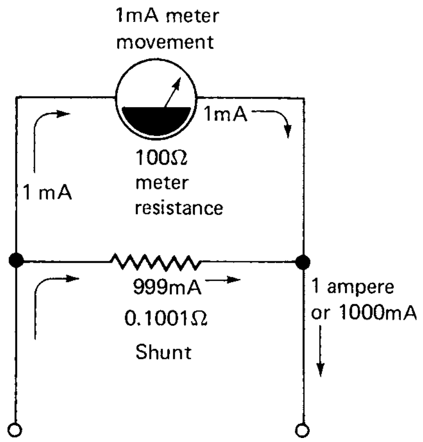
1mA meter movement
1mA
1 mA
100Ω meter resistance
999mA
0.1001Ω
Shunt
1 ampere or 1000mA

Figure 6-10 One-milliampere movement that can read up to 1 ampere.

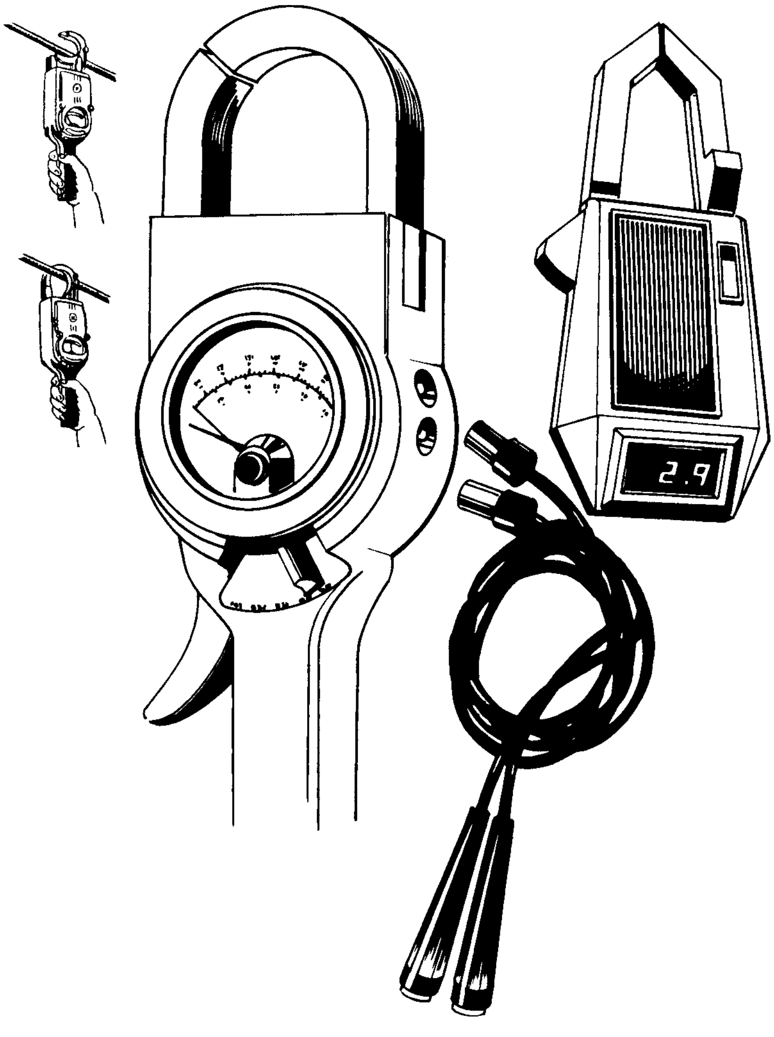

Figure 6-11 Clamp-on ammeter.

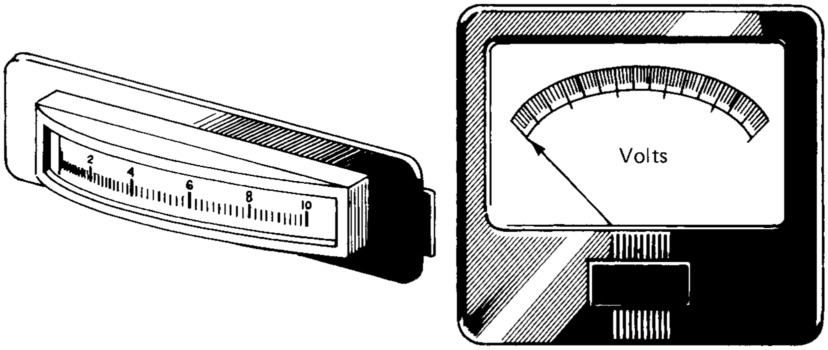

Figure 6-12 AC ammeter circuit.

when working with ac motors and compressors. Leads are provided so that the meter can be used as a voltmeter by connecting the leads across whatever voltage source is being checked.

AC ammeters

DC meters can be modified to measure ac current. This is done by connecting a rectifier in series with the meter movement (see Fig. 6-12). A rectifier is a device that changes ac to dc. An ac scale and rectifier are shown in Fig. 6-12. Today, most rectifiers are semiconductor diodes. In the past, vacuum tubes were often used for this purpose.

Voltmeters

The voltmeter measures electrical pressure, or volts. It is nothing more than an ammeter with a resistor added to the meter circuit. The high resistance of the voltmeter makes it possible to place it across a power source (in parallel) (see Fig. 6-13). A number of resistors, called multipliers or scaling resistors, can be switched into a meter circuit to increase its range to make it capable of measuring higher voltages.

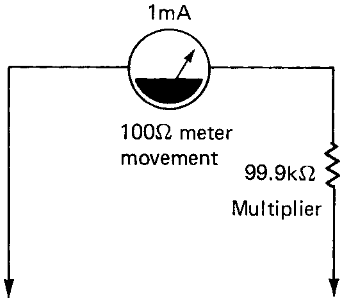

Figure 6-13 Voltmeter.

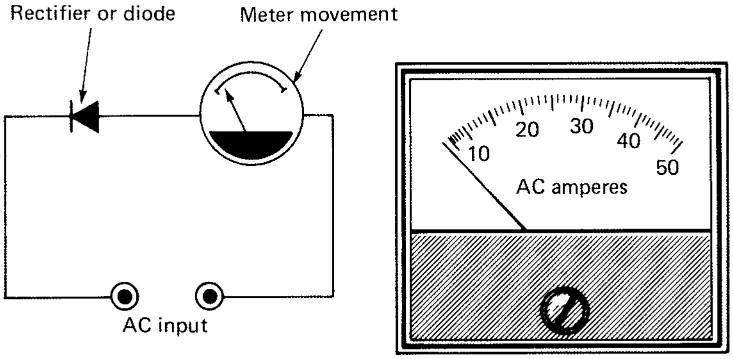
Rectifier or diode
Meter movement
AC input
10
20
30
40
50
AC amperes

Figure 6-14 Voltmeter circuit.

A voltmeter circuit is shown in Fig. 6-14. It has 100-volt full-scale deflection. The resistor indicated in the illustration is the meter multiplier. The multiplier's resistance rating is 99.9 kilohms (kΩ). The notation k represents 1000. So the resistance is 99,900 ohms. Meter multipliers are usually precision-type resistors with a ±1% accuracy.

Connecting a voltmeter

A multirange voltmeter uses a switch to select among multipliers. Scales are arranged so that voltage is measured near the center of the meter scale. The center portion of a meter scale is always the most accurate part of the deflection range.

AC Voltmeters

The ac voltmeter is nothing more than the dc meter movement with a diode inserted in series with the meter movement and the multiplier resistor (see Fig. 6-15).

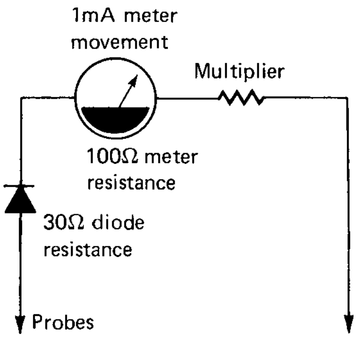
1mA meter movement
Multiplier
100Ω meter resistance
30Ω diode resistance
Probes

Figure 6-15 AC voltmeter circuit.

Ohmmeters

It is sometimes necessary to troubleshoot a piece of equipment when no electricity is available. Also, in some cases, the application of power to the circuit may cause some damage. Therefore, you need to be able to troubleshoot the circuit without having to turn on the power and maybe burn up something. This is when an ohmmeter is used to check a circuit because it has its own power source (batteries).

Two types of ohmmeters are made for use in the field: the series ohmmeter, for measuring high resistances and the shunt ohmmeter for measuring low resistances like motor windings and checking for continuity of wiring.

The ohmmeter is most convenient when it comes to measuring resistance in a circuit. It has two probes that can be placed across an unknown resistance. The value of the resistance is displayed on the meter scale (see Fig. 6-16).

The ohmmeter has its own voltage source or batteries. It is never connected to a live circuit (one with the power on). Full-scale deflection on the series ohmmeter indicates 0. This means that when the probes are shorted, the meter pointer goes to the right and indicates no resistance or zero. When the probes are apart with nothing in between, they indicate infinity (∞), which means too much resistance to measure. Note how the scale is nonlinear. The series ohmmeter measures high resistances and the meter scale reads from right to left.

The shunt ohmmeter is designed to measure low resistance, usually below 200 ohms. See the scale in Fig. 6-17. Note, too, that the scale reads from left to right, with the zero being on the left and infinity on the right. An important thing to remember about the shunt ohmmeter is to be sure you turn it off when you have finished using it. It will run down the batteries if left in the on position. The shunt ohmmeter can be identified because it requires an on-off switch. If it is part of a multimeter, then switching it to amperes or volts will turn it off.

Multimeters

Voltmeters, ammeters, and ohmmeters are often combined into a single instrument. Such an instrument is called a multimeter (see Fig. 6-18). Multimeters have a range switch that selects the scale. Another control is a function switch to control its use as an ohmmeter, ammeter, or voltmeter. Multimeters also have a zero adjust knob for the ohmmeter circuits. There are also jacks for the probes. The functions and ranges of a multimeter are determined by selector switches.

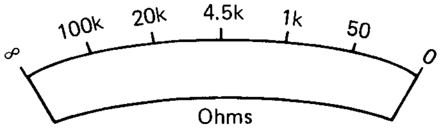

Figure 6-16 Series ohmmeter scale.

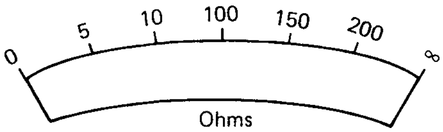

Figure 6-17 Shunt ohmmeter scale.

Meggers

The megger (see Fig. 6-19) is a device that is used for very high resistances. It is used to check for insulation breakdown in motors and compressors. A handcrank is attached to a coil inside the megger. The coil is placed inside a permanent magnet field. A high voltage is produced when the crank is rotated. This high voltage is required to drive small amounts of current through extremely high resistances. The meter movement has to be very sensitive.

The megger can produce a shock if the leads are touched while the handle is turned. Be careful while using a megger. Do not attach the leads to anything that will be easily damaged by high voltage.

Digital Meters

Figure 6-20 shows a digital meter. There are no coils or magnets in digital meters. The meter presents its reading in numbers on a liquid crystal display. This is the type of digital readout found on many instruments, clocks, and watches.

This type of meter can be used to measure ac and dc voltages and currents with great accuracy. In fact, most digital meters can measure voltages to one-hundredth of a volt. They can also measure resistances within one-tenth of an

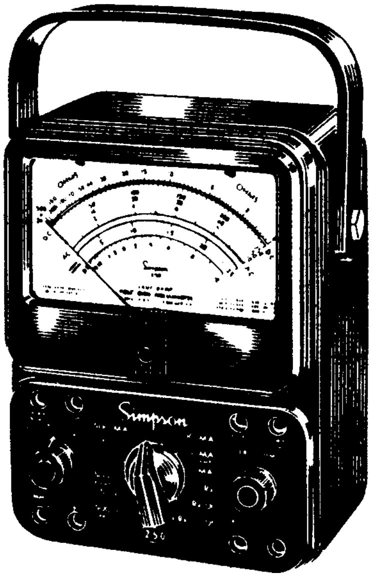

Figure 6-18 Multimeter.

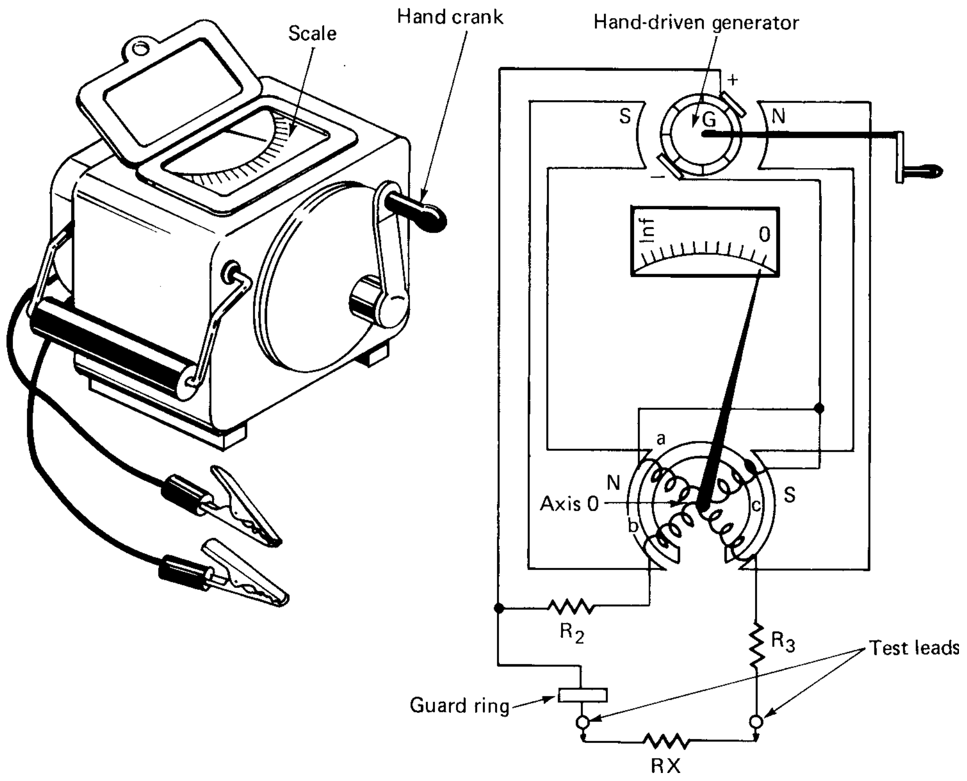
Scale
Hand crank
Hand-driven generator
+
G
S
N
−
Inf
0
a
N
Axis 0
b
c
S
R2
R3
Test leads
Guard ring
RX

Figure 6-19 Megger.

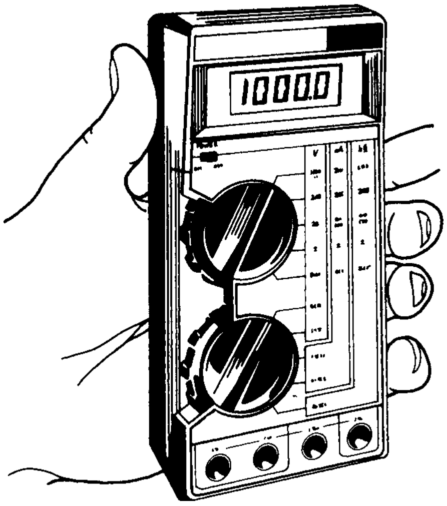

Figure 6-20 Digital meter.

ohm. They are sometimes more accurate than needed for the job. They are also presently more expensive than the analog type. This, of course, is subject to change, as is the case with all electronics.

Other Instruments

If you are going to be working with heating, air-conditioning, and refrigeration circuits, you will need some other types of instruments to do the job properly and, most of all, efficiently. Electronics is being used today to make the job of recharging refrigeration systems easier and more accurate. A few of these newer electronic devices now available to refrigeration and air-conditioning technicians are described next.

Automatic halogen leak detector

A number of halogen leak detectors are available. However, some of the newer electronic detectors are very sensitive. Figure 6-21 shows an automatic type that has a permanently sealed, miniature battery-operated pump that produces a computer-like beeping signal that changes in both speed and frequency as the leak source (halogen or vapor) is approached. It can detect leaks as small as 0.5 ounce per year. An additional feature is that this instrument is capable of calibrating itself automatically while in use. It operates on two C cells. The detector sensor is not ruined by large doses of refrigerant.

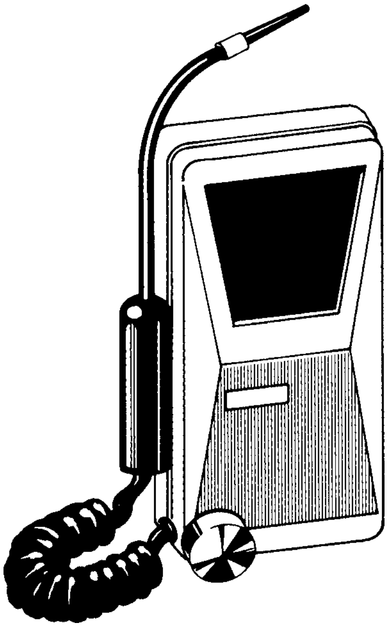

Figure 6-21 Electronic halogen leak detector.

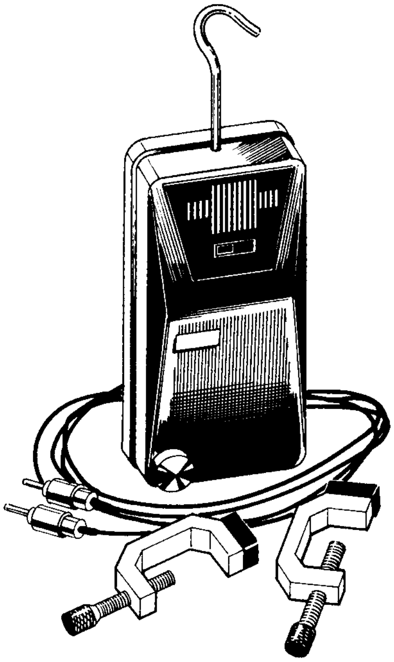

Figure 6-22 Electronic sight glass.

Electronic sight glass

The electronic sight glass emits an audible signal indicating that the refrigeration system being charged is full (see Fig. 6-22). It uses sonar to look inside refrigeration tubing and, in effect, can be considered an "electronic sight glass." It is very handy for capillary systems, which have no sight glass. It is also helpful when charging household and commercial refrigerators and freezers. It is also valuable when checking automobile systems and window air conditioners. This instrument detects flash gas and bubbles in the liquid line. It detects droplets in the suction line and starved evaporators. It can also be used to check for refrigerant flood-back and becomes very useful for multiple evaporator balancing and for adjusting thermostatic expansion valves.

Electronic charging meters

The electronic charging meter is a new tool for dispensing refrigerant into a system in fractions of an ounce without limit from a standard 30-pound cylinder. The charging meter measures the charge dispensed into the system by weight, but it is totally different from a bathroom-type scale. It reads directly like a service station pump. No more wasted time refilling a charged cylinder; no more recorrecting for temperature or type of refrigerant. The refrigerant charge appears on a liquid crystal display, which is easily read even in bright sunlight. There are no controls or adjustments. It is operated at the touch of a

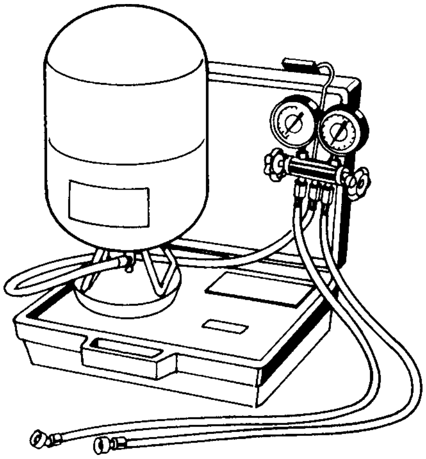

Figure 6-23 Electronic charging meter.

button. It can also be used as a conventional weight scale so that your cylinder can be weighed to ascertain the amount of refrigerant remaining for other jobs. It operates on six C cells (see Fig. 6-23).

Using an Ohmmeter

The basic unit for measuring resistance is the ohm. An ohmmeter is a device used to measure resistance or ohms. It is an ammeter (or milliammeter or microammeter) movement, modified to measure resistance (see Fig. 6-2). The ohmmeter is easily damaged if not properly connected to the circuit under test. Most multimeters are capable of measuring ohms with at least three different ranges: R × 1, R × 10, and R × 10,000. Thus, it is capable of measuring up to 20 million ohms when on the R × 10,000 scale. The meter scale has to be multiplied by the 100 or 10,000 number to obtain the proper value. The R × 1 scale is used to check for shorts and very low resistances, such as in compressor motors, relay contacts, and motor windings and switches.

Warning

An ohmmeter has a battery in it, so it does not use power from the circuit under test. In fact, do not connect it to a live circuit (one with the power on). To do so will result in destruction of the meter movement.

Adjusting the meter

Adjust the ohmmeter so that the meter reads zero before starting to use it to measure resistance. This means that you have to adjust the meter circuit to compensate for the battery voltage changes. Battery voltages decrease with shelf life. It doesn't matter whether or not the battery is used. It will, in time, lose its voltage.

The ohmmeter can be used to test compressor windings for opens or shorts. An open means the meter does not move or reads infinite. A short means the meter reads zero. If you read the same resistance between the three pins of the compressor, it means the motor is three phase and will not run on single-phase power.

One important thing is to obtain an ohmmeter that can measure very low resistances, such as 1 to 10 ohms, with some degree of accuracy. This calls for a digital meter that is somewhat expensive or a multimeter with a shunt ohmmeter capability. The shunt ohmmeter has a scale for low resistance that reads from left to right (see Fig. 6-17). Since, in some instances, a 1-ohm reading can indicate a short, you will probably need a good digital ohmmeter to be sure. The common readings for a compressor are 1, 5, 10, 15, and 20 ohms resistance.

Check Fig. 6-24 for the layout of compressor terminals. Note that the resistance from R to S is 11 ohms. This means that both windings (start and run) are in series and are being added at this point. The reading from R to C is 1 ohm. This means that the run (R) winding has 1 ohm of resistance and is normal. If the reading is less than 1 ohm, it means you have a short in the start winding. The reading from C (common) to S (start winding terminal) is normally 10 ohms. If you obtain infinite for any of these readings, it means the winding is open.

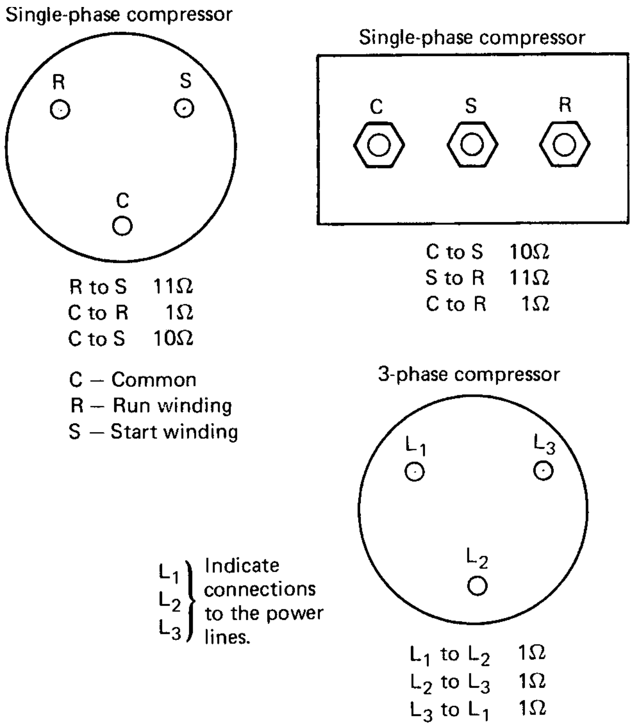
Single-phase compressor
R
S
C
R to S 11Ω
C to R 1Ω
C to S 10Ω
C – Common
R – Run winding
S – Start winding
Single-phase compressor
C
S
R
C to S 10Ω
S to R 11Ω
C to R 1Ω
3-phase compressor
L_1
L_3
L_2
L_1, L_2, L_3 Indicate connections to the power lines.
L_1 to L_2 1Ω
L_2 to L_3 1Ω
L_3 to L_1 1Ω

Figure 6-24 Resistances between compressor terminals.

If you read less than 3 ohms for the run winding, it means the winding is shorted. If you get a reading from any one of the three terminals to the case of the compressor, it means the winding is shorted to the case.

Other steps in troubleshooting will be explained in following chapters.

Review Questions

1. How is electricity measured?
2. What is a d'Arsonval movement?
3. What is meant by full-scale deflection?
4. What is parallax error?
5. What is the difference between milliampere, microampere, and ampere?
6. How is the range of a meter extended?
7. How is the ammeter connected in a circuit?
8. What is the difference between a series and shunt ohmmeter?
9. What is a megger? Where is it used?
10. How is the voltmeter used to test a circuit?

Chapter 7

Electrical Power: Direct Current and Alternating Current

Performance Objectives

- Know the type of current furnished by a battery.
- Know the difference between a cell and a battery.
- Know what is meant by the ampere-hour rating of a battery.
- Know how cells are connected in series and in parallel.
- Know what causes batteries to have a shortened life.
- Know what a sine wave is.
- Know the difference between alternating current and direct current.
- Know what rms means.
- Know the meaning of polyphase.
- Know the difference between delta and wye connections.
- Know how three-phase power is generated.

Power Sources

Today's electronics require sources of direct current to operate the semiconductor devices associated with their control circuits. Direct current can be generated with any number of battery types. However, only the more familiar types are presented here. The heating, air-conditioning, and refrigeration technician must learn to choose the right battery for the job.

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

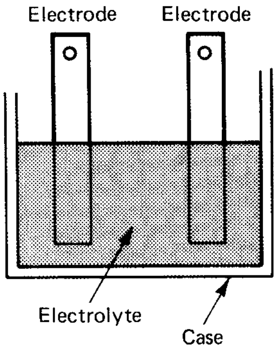

Figure 7-1 Makeup of a cell.

Types of Batteries

A single cell produces a voltage between 1.2 and 2.2 volts (see Fig. 7-1). Two or more cells connected either in series or parallel make a battery. The number of cells in a battery is related to its voltage and current requirements. Cells connected in series produce higher voltages. Cells connected in parallel produce higher current.

Primary cell

A cell is a device that converts chemical energy into electrical energy. The primary cell cannot be recharged. Once the primary cell is discharged, it is thrown away. The chemical action that discharged the primary cell cannot be reversed (see Fig. 7-2).

Secondary cell

The secondary cell can be recharged. The chemical action is reversible. The electrodes and electrolyte that make up the cell can be restored to the same makeup that existed before the discharge (see Fig. 7-3).

The secondary cell, such as the one used in your automobile battery, can be discharged and charged many times. However, the secondary cells that make up such a battery deteriorate. The number of times a secondary cell can be charged and discharged depends on its design, construction, use, and how well it is maintained. During discharge, electrons flow from the − (negative) terminal of the cell or battery to the + (positive) terminal by way of a load (see Fig. 7-4). When the cells are charged, the process is reversed and electrons flow from + to −. This reversal of electron flow causes a chemical action to take place that produces a charged condition when completed.

Dry Cells

The dry cell is a primary cell. It is so named because the electrolyte is not a liquid but a dry, paste-like compound. The many types of dry cells differ mainly in the types of electrodes and electrolyte used. Three types of dry cells in popular use are the carbon-zinc cell, the alkaline cell, and the mercury cell.

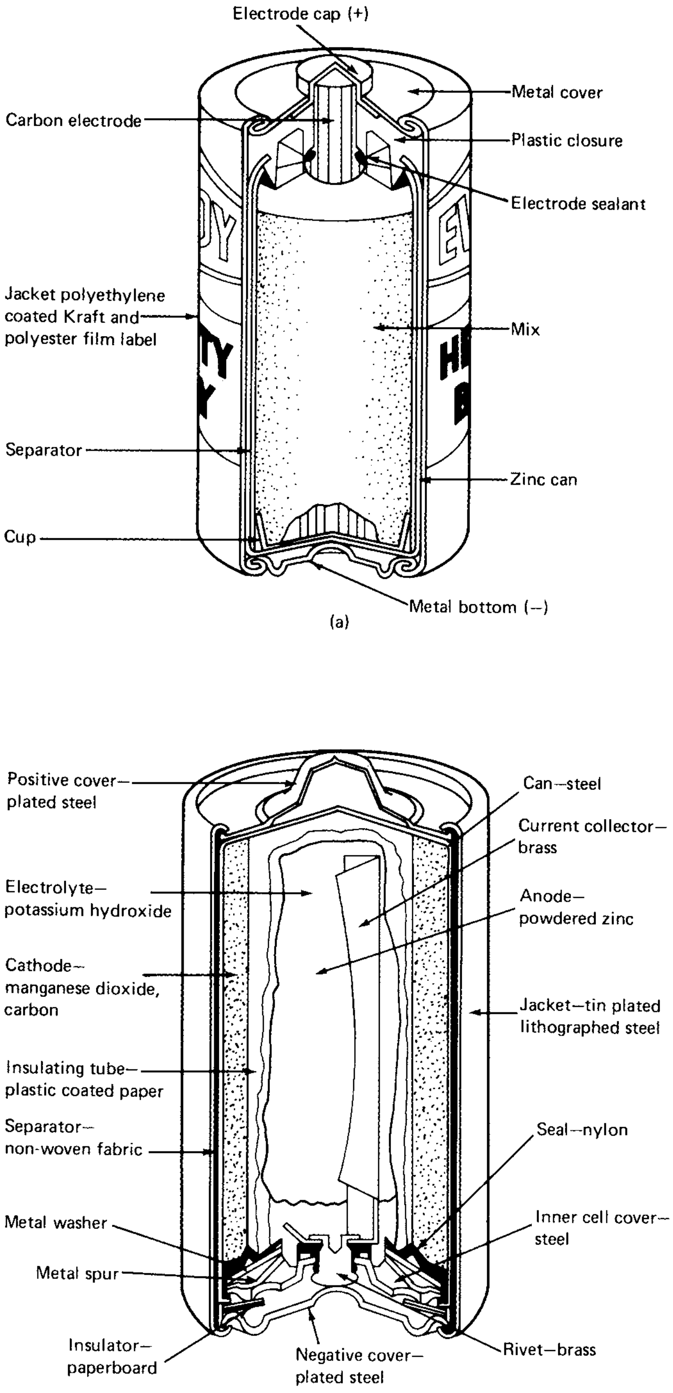
Electrode cap (+)
Metal cover
Carbon electrode
Plastic closure
Electrode sealant
Jacket polyethylene coated Kraft and polyester film label
Mix
Separator
Zinc can
Cup
Metal bottom (−)
Positive cover—plated steel
Can—steel
Current collector—brass
Electrolyte—potassium hydroxide
Anode—powdered zinc
Cathode—manganese dioxide, carbon
Jacket—tin plated lithographed steel
Insulating tube—plastic coated paper
Separator—non-woven fabric
Seal—nylon
Metal washer
Inner cell cover—steel
Metal spur
Insulator—paperboard
Negative cover—plated steel
Rivet—brass

(a)

(b)

Figure 7-2 Dry cells.

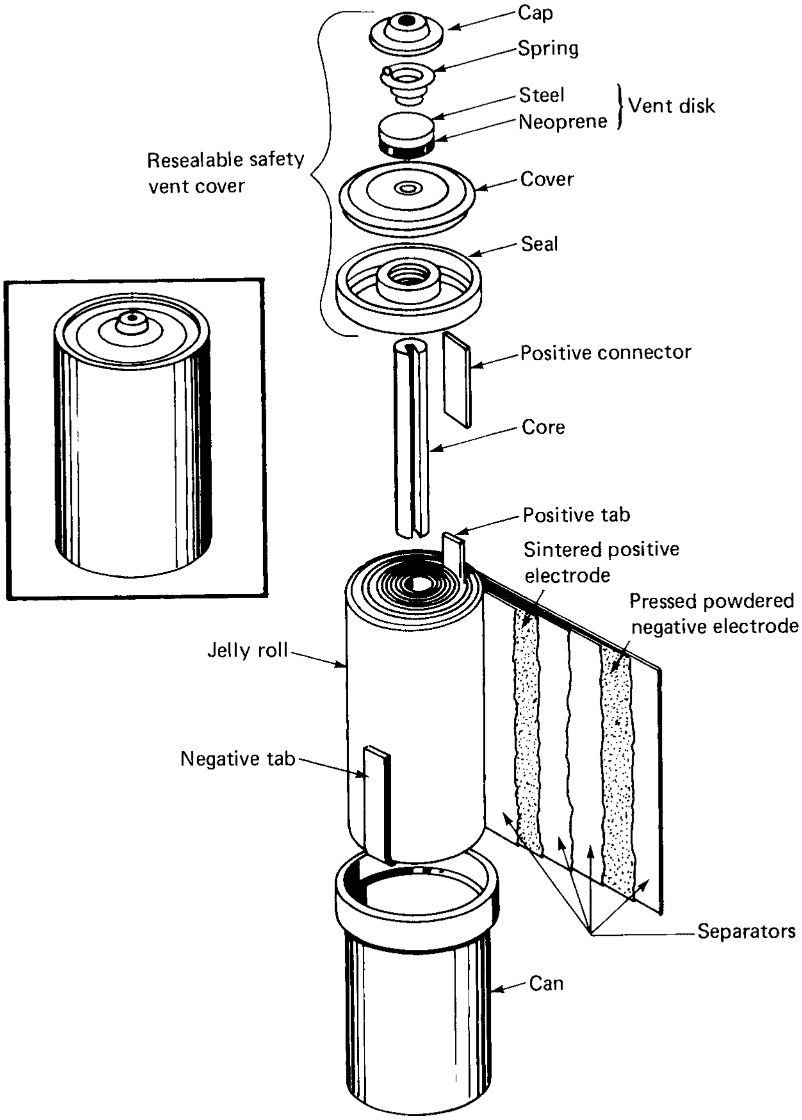
Cap
Spring
Steel
Neoprene
Vent disk
Resealable safety vent cover
Cover
Seal
Positive connector
Core
Positive tab
Sintered positive electrode
Pressed powdered negative electrode
Jelly roll
Negative tab
Separators
Can

Figure 7-3 Rechargeable nickel-cadmium cell. (Courtesy of Eveready)

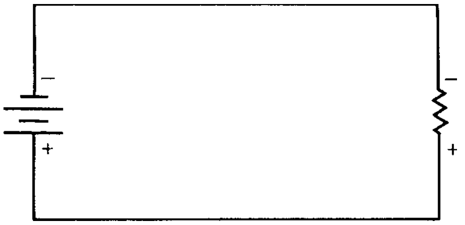

Figure 7-4 Cells connected – to + with the load.

The alkaline cell will last longer than the carbon-zinc because it has a lower internal resistance and higher current output. This is due to the potassium hydroxide used as an electrolyte. Potassium hydroxide has a higher conductivity than the sal ammoniac used in the carbon-zinc cell.

The mercury cell has a positive electrode of mercuric oxide. The negative electrode is made of amalgamated zinc. The entire assembly is mounted in a sealed steel container. It is produced in flat and cylindrical shapes. With the use of pressed powder electrodes, the mercury cell can be made very small. This is an advantage for use in hearing aids and other miniature electronic devices, such as watches and timers used with electronic temperature controls (see Fig. 7-5).

Battery Specifications

Batteries are made in many sizes and shapes. Figure 7-6 shows the standard cell sizes. Batteries come in even larger sizes. They may be as small as the 9-volt transistor battery and as large as a submarine battery that measures over 6 feet high.

Batteries are rated according to voltage. They are also rated according to the recommended current range or in terms of ampere-hour capacity.

Higher current demands and longer operating hours than the specifications call for will shorten the life of the battery. Higher current and longer operating hours will reduce the ampere-hour capacity of the battery. The ampere-hour rating is good for a reasonable range on each side of the point used to establish the rating. Alkaline batteries are made for continuous-drain and high-current operation. Their service life may be 10 times greater than a comparable carbon-zinc battery. For light loads, the service life may be only a limited number of times longer. For light loads, then, alkaline batteries may not be practical because of the higher initial costs. Lithium batteries are becoming commonplace for many electronic devices. They are, however, very expensive.

The mercury battery is expensive in comparison to the alkaline and carbon-zinc. The lithium battery or cell is more expensive than alkaline.

Connecting Cells

Cells can be connected in series or parallel to obtain a desired voltage or current rating. When cells are connected in series, the individual voltages add. Four 1.5-volt cells connected in series provide 6 volts. The 9-volt transistor battery is made up of six 1.5-volt cells connected in series.

Cells are connected in series by putting the − terminal to the + terminal of the next cell. They are connected, − to +, − to +, and − to +, for as many cells as you need to produce the voltage needed (see Fig. 7-7).

Cells connected in parallel, as in Fig. 7-8, provide more current. They are connected by placing the − to − terminals together and the + to + terminals together.

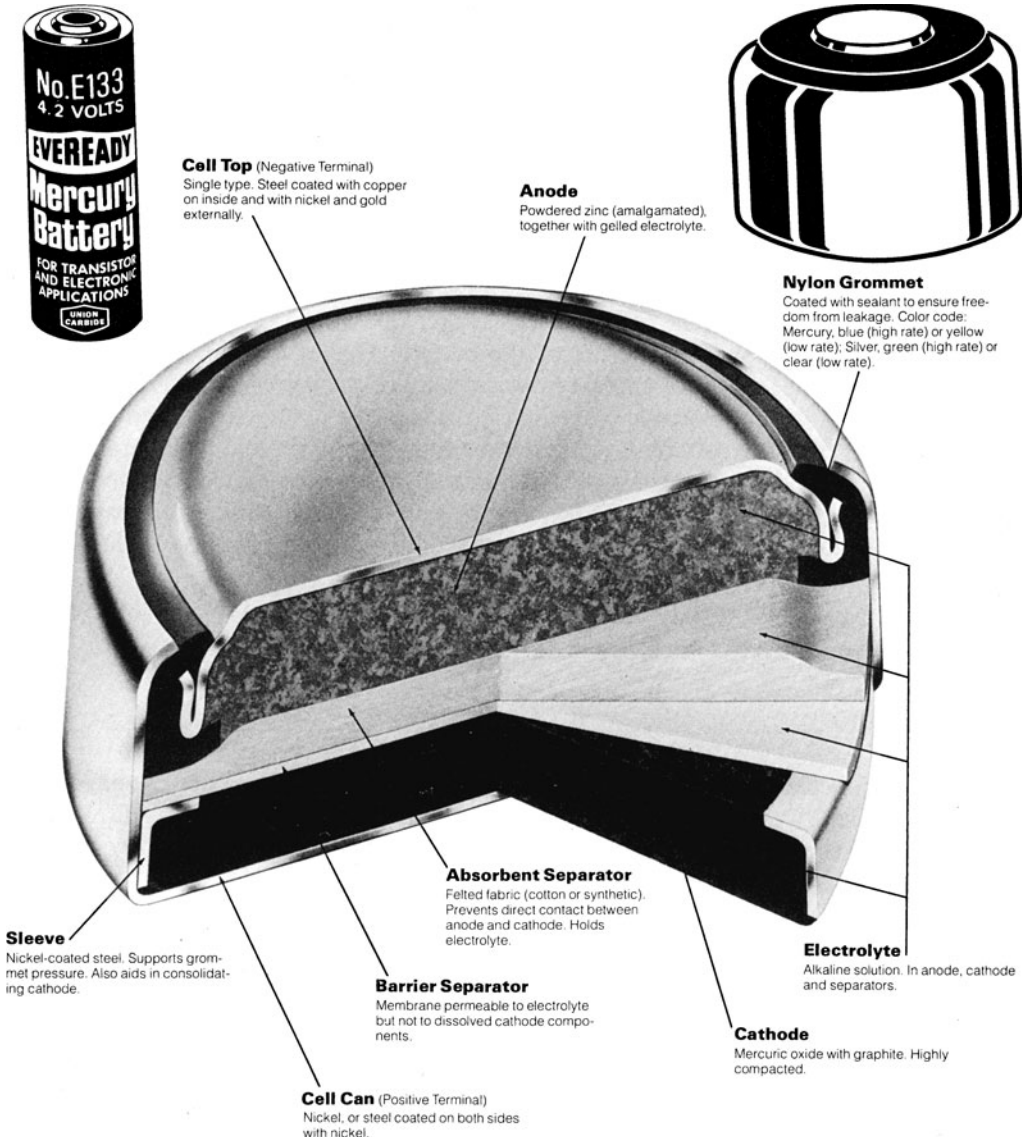
No. E133
4.2 VOLTS
EVEREADY
Mercury Battery
FOR TRANSISTOR AND ELECTRONIC APPLICATIONS
UNION CARBIDE
Cell Top (Negative Terminal)
Single type. Steel coated with copper on inside and with nickel and gold externally.
Anode
Powdered zinc (amalgamated), together with gelled electrolyte.
Nylon Grommet
Coated with sealant to ensure freedom from leakage. Color code: Mercury, blue (high rate) or yellow (low rate); Silver, green (high rate) or clear (low rate).
Absorbent Separator
Felted fabric (cotton or synthetic). Prevents direct contact between anode and cathode. Holds electrolyte.
Sleeve
Nickel-coated steel. Supports grommet pressure. Also aids in consolidating cathode.
Electrolyte
Alkaline solution. In anode, cathode and separators.
Barrier Separator
Membrane permeable to electrolyte but not to dissolved cathode components.
Cathode
Mercuric oxide with graphite. Highly compacted.
Cell Can (Positive Terminal)
Nickel, or steel coated on both sides with nickel.

Figure 7-5 Mercury cell.

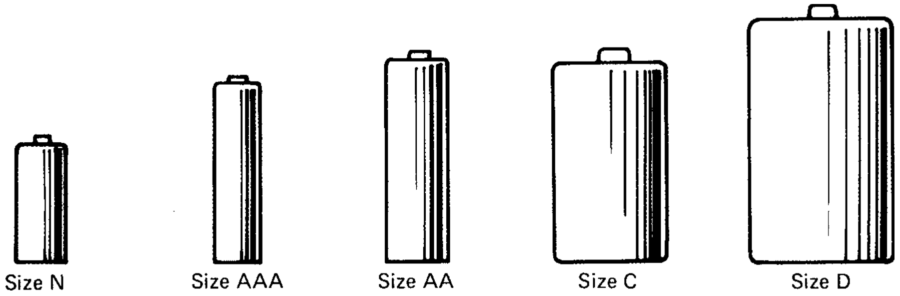
Size N
Size AAA
Size AA
Size C
Size D

Figure 7-6 Five standard cell sizes.

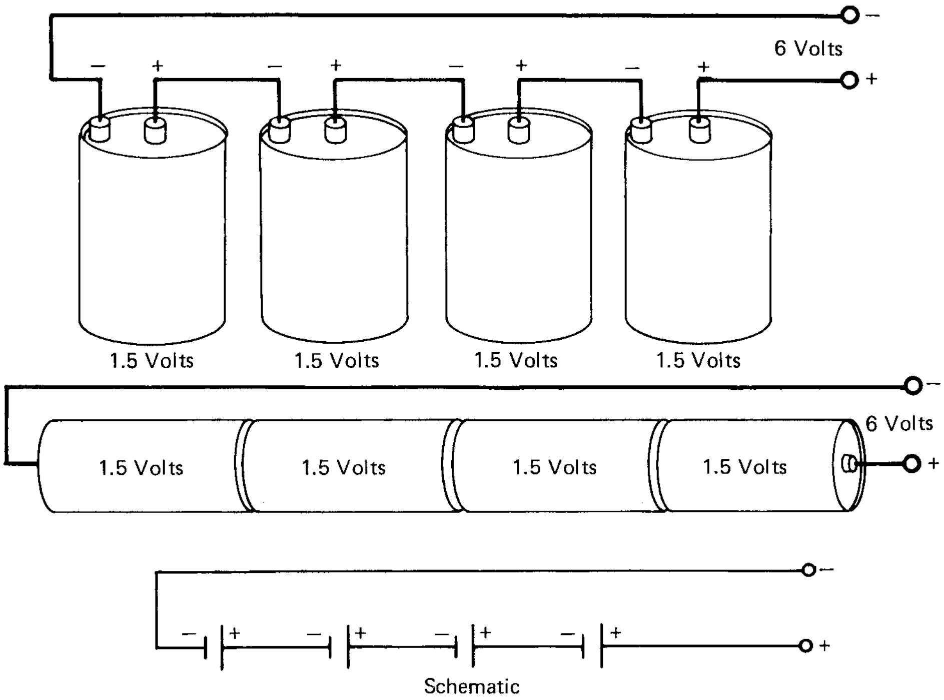
–
6 Volts
+
– + – + – + – +
1.5 Volts
1.5 Volts
1.5 Volts
1.5 Volts
–
6 Volts
+
1.5 Volts
1.5 Volts
1.5 Volts
1.5 Volts
–
– + – + – + – +
+
Schematic

Figure 7-7 Cells in series.

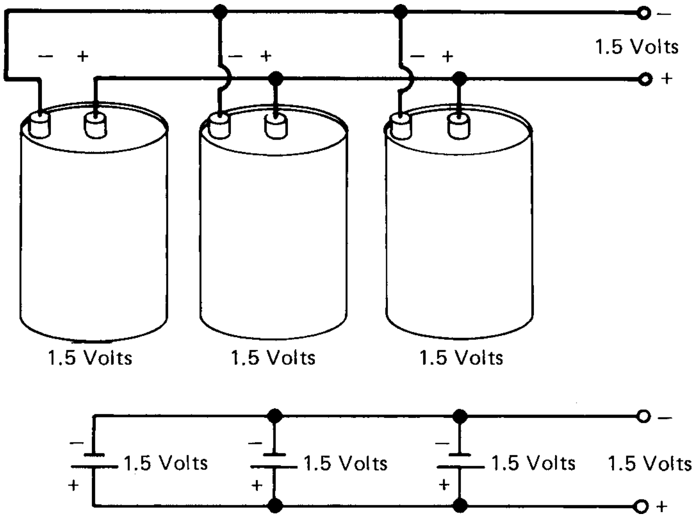
–
1.5 Volts
+
– + – + – +
1.5 Volts
1.5 Volts
1.5 Volts
–
– 1.5 Volts +
– 1.5 Volts +
– 1.5 Volts +
1.5 Volts
+

Figure 7-8 Cells in parallel.

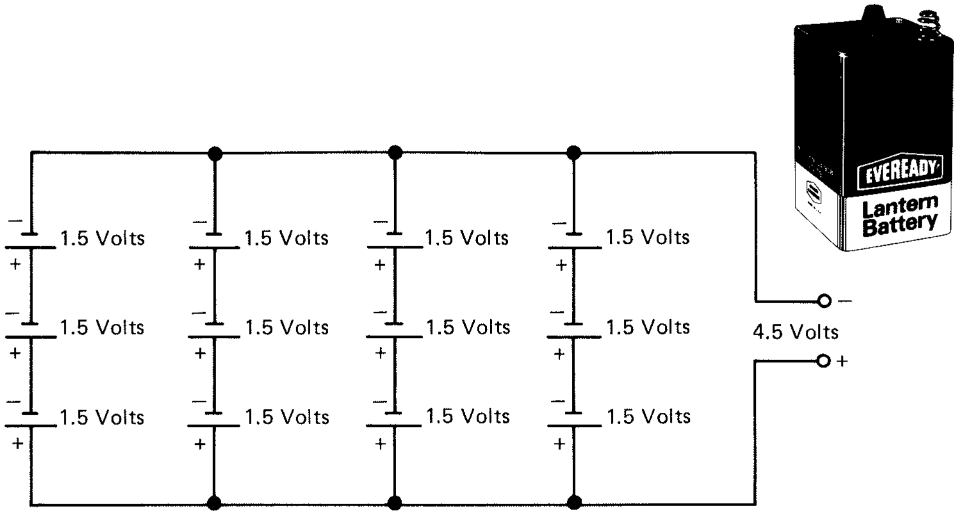

Figure 7-9 Cells in series-parallel.

You obtain the same voltage as from one cell, but the current available from all the cells is added. Do not connect cells with different voltages in parallel. They have a tendency to quickly drain when not in use.

Higher voltage and current batteries such as lantern batteries can be obtained by using cells connected in series-parallel (see Fig. 7-9).

Battery Maintenance

Batteries that have been properly treated will last longer. Two of the most important elements working against batteries are overheating and sulfation. The formation of lead sulfate on the battery plates is normal during the discharge cycle of the lead-acid battery. The lead-acid type is used in the automobile and elsewhere when high currents are needed for an extended period of time. If a battery is allowed to stay in a discharged state for a period of time, the deposit hardens and no longer responds to recharging. The battery is said to be sulfated and cannot be restored to full capacity. Sulfation extended over a long period makes the battery useless.

Other practices that can cause shortened battery life are frequent undercharging, too high a specific gravity, long idle periods in a discharged condition, and defects caused by rapid cycles of high-current discharge and high-current charge.

Charge and discharge currents should not be allowed to overheat the battery. Overheating can cause bending of the battery plates and excessive sulfation. Avoid rapid discharge and keep the proper water level. There is a normal loss of water from the electrolyte, and it is important that the water be replaced. Reasonably pure water should be added when necessary to maintain the proper level and the correct specific gravity range. Overheating tends to encourage faster water evaporation. Whenever the battery is on charge, there is some release of hydrogen and oxygen gas at the electrodes. These gases bubble through the electrolyte to the surface and then through the vent hole in the cell. Indirectly, this also means a loss of water, which must be replaced. However, recent battery technology has improved the battery to such a point that maintenance-free operation is possible for the life of the battery if it is cared for and used for its intended purpose. The sealed battery does not call for the addition of water inasmuch as there is no evaporation.

Nickel-Cadmium Cells

The nickel-cadmium cell comes in a variety of sizes. It can be sealed and made very compact. Figure 7-3 shows the nickel-cadmium construction. It is a high-efficiency cell made so that it can deliver high current at a constant-voltage output. It can be recharged often and still have a long active life. Storage life is not critical and long idle periods, charged or discharged, have no bad effects. It can also be completely discharged without damage.

The voltage output of the nickel-cadmium cell is usually 1.25 volts or a little less than the 1.5 volts of the carbon-zinc cell.

Alkaline Cells

The alkaline cell can be either a primary or secondary cell. It has characteristics comparable to a similar sized nickel-cadmium battery. The number of times that the alkaline battery can be charged and discharged is less. However, the initial cost is much less than a similar sized nickel-cadmium. In some cases, overall economic considerations favor the alkaline type of battery.

The alkaline rechargeable cell should never be discharged below 0.9 volt per cell because the battery life will be shortened. If an alkaline rechargeable cell is completely discharged, it is impossible to recharge it.

When an alkaline secondary battery is discharged for the maximum recommended time and then recharged for the maximum recommended time, it may be cycled as many as 50 times before the battery falls below the rated cutoff voltage at the end of the recommended discharge period. The alkaline cell output voltage is 1.5.

Proper battery care is needed to make sure that electronic and electrical equipment operate properly and when needed. Also, be sure to dispose of the batteries and cells properly according to manufacturer's recommendations.

Alternating Current

Direct current has electricity flowing in one direction. Another type of electricity that you must understand to be able to troubleshoot equipment is alternating current (ac). Most motors and other control devices on refrigerators, air conditioners, and heating units operate on alternating current, or ac as it is most commonly called.

Alternating current is constantly changing in direction and amplitude. In as much as dc was discovered first, the formulas and mathematical equations for working with electricity were developed for dc. However, after the introduction of ac, around the beginning of 1900, a number of changes had to be made in the way electricity was handled mathematically. Alternating current was equated to dc. This was for the sake of making previous formulas work and, at the same time, to produce an understanding of what ac could do when applied to motors and other devices.

Sine Wave

The shape of the ac waveform is that of a sine wave. There are many other types of waveforms, such as the pulse, sawtooth, and square. However, these other waveforms are considered combinations of sine waves of different amplitude, frequency, and phase. Thus, the forms other than sine waves are usually discussed in terms of their sine-wave (sinusoidal) makeup. The basic sine-wave current is a single frequency, and it is this sine wave of single frequency (60 Hz) that we are interested in for our use (see Fig. 7-10).

The sine wave is unique. It can take many shapes. The sine wave has a very special relation to time and motion. Take a close look at Fig. 7-11. Assume that a point at the end of the radial (r) takes one second to travel at a constant speed around the circle, as shown in Fig. 7-11a. How long would it take the point to move 90° or 180°? Since the radial rotates at a constant speed, it would require 0.25 second to rotate 90° (from a horizontal position to a vertical position). At this instant, the point at the end of the radial (on the circumference of the circle) would be at the highest point with relation to the horizontal line. At 180° the point will have been in motion for 0.5 second and will have reached the horizontal line. At 270°, or 0.75 second, the point at the end of the radial line will be at the lowest point with relation to the horizontal reference line. Finally, in 1 second the point at the end of the radial comes back to its starting position, and one cycle (called 1 hertz) has been completed. Keep in mind that time can be stated in degrees of rotation, and the horizontal reference line is time.

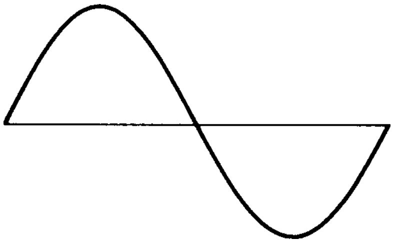

Figure 7-10 Sine wave.

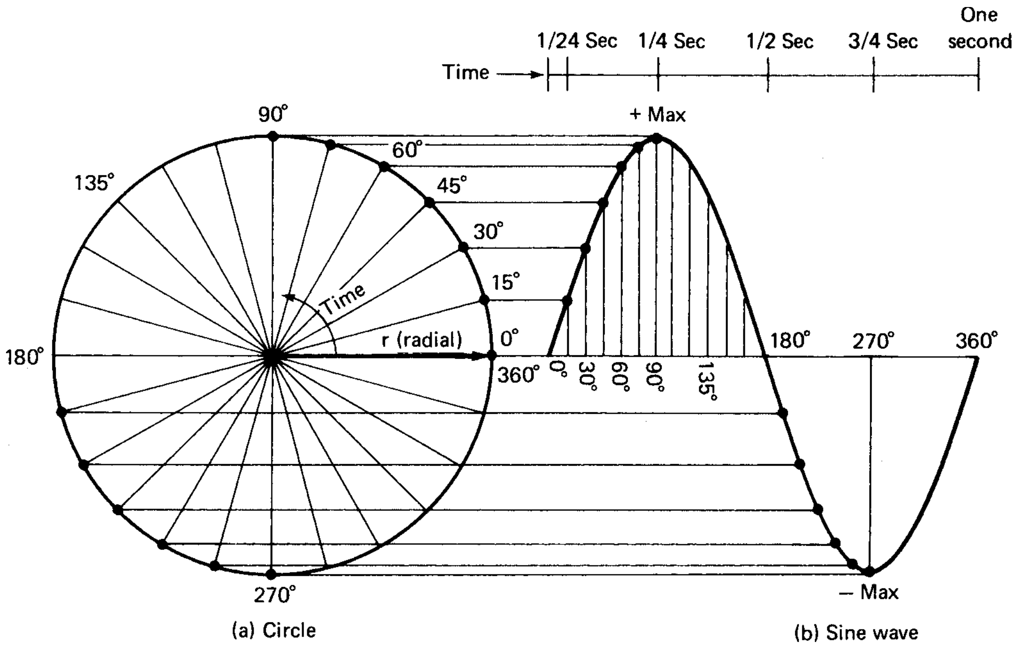

Figure 7-11 Sine wave showing how it is generated by a point rotating in a circle, that point being one coil on a generator rotating in a magnetic field.

Electricity available for home use in the United States is 60 hertz. Thus, the generator must turn 60 times per second to produce 60 hertz. One hertz is one cycle per second. Thus, one sine wave is produced every sixtieth of a second.

Sine-Wave Characteristics

Figure 7-12 shows the output of two cycles of the generator. These two cycles (hertz) of alternating current have a negative alternation and a positive alternation. Each sine wave has a positive alternation that swings above the zero axis and a negative alternation that swings below the zero axis. Unless otherwise marked, it is normal for the alternation that swings above the zero axis to be positive, and the alternation that swings below the zero axis to be negative. Thus, a complete cycle (hertz) is made up of a positive and negative swing.

Two important characteristics of a sine wave are period and frequency. It takes time for the sine wave to go through its positive and negative alternations and return to the origin or zero reference line. The time required for a single cycle (hertz) is called the *period of the sine wave.*

The number of complete cycles (hertz) made each second is called the *frequency of the sine wave.*

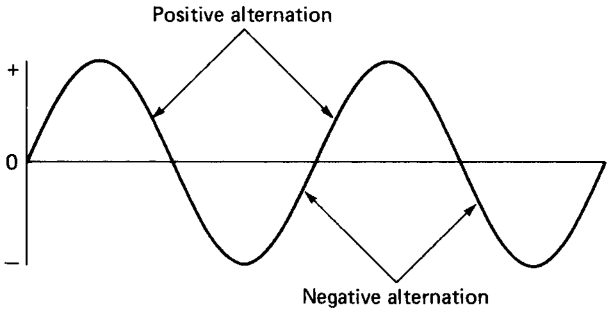

Figure 7-12 Two hertz (cycles) of a sine wave.

Phase

In electrical work it is often necessary to relate one sine wave to another. Two sine waves are shown in Fig. 7-13. This relation between two sine waves, known as phase, can be a comparison between a sine-wave voltage and a sine-wave current. Or it can be a comparison between two sine-wave voltages or between two sine-wave currents. This phase or difference in time is another important characteristic of a sine wave.

Average, RMS, and Peak Values of AC

Three values are used to describe ac:

1. Peak value
2. Average value
3. Root-Mean-Square (rms) value

Peak value. The maximum point on a sine wave is its peak value. Both peaks of a single hertz may be included in a reference. If so, it becomes a peak-to-peak (p-p) value. A peak value of 100 volts means that the peak-to-peak value is 200 volts. This is shown in Fig. 7-14.

Average value. The average of all instantaneous values of a generator is measured at regular intervals (see Fig. 7-11). Instantaneous values are taken

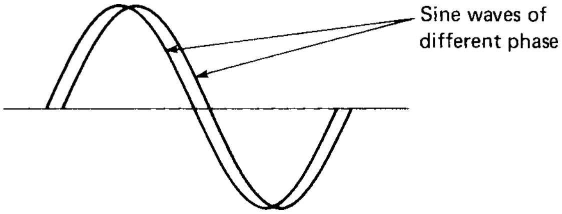

Figure 7-13 Phase difference between two sine waves.

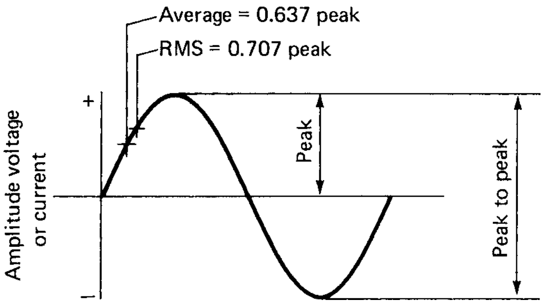

Figure 7-14 Sine-wave values of average, rms, and peak.

at selected points in the generating process. The average of these is the average value of ac current or voltage.

The half-hertz is used to compute this value. This is because the average value for full-hertz (cycles) would be zero. The zero average would come from the fact that half the values are positive and half are negative.

To compute the average value, sine-wave values for angles between 0° and 180° are added. The total is then divided by the number of values taken. As seen in Fig. 7-14, the average is equal to 0.637 of the peak. For example, suppose the peak voltage of a sine wave is 100 volts. The average voltage is 63.7 volts ($100 \times 0.637 = 63.7$).

Root-Mean-Square (RMS) value. A value equivalent to effective voltage or effective current is known as the Root-Mean-Square (rms) value. It is calculated mathematically. Imagine one-quarter of the sine wave broken down into 90 parts. This is one part for each degree of a hertz. The value of each degree is squared, or multiplied by itself. All the squared values are added together. Then the total is divided by 90. This produces the average of the squares. A square root is then figured for this average. This is the rms value.

Polyphase Alternating Current

The armature coils of a single-phase alternator (ac generator) are connected in series. At any given instant, the voltage induced is the sum of the voltages induced separately in each coil. If the coils were not connected in series, and if their connections were brought out separately, the voltages induced would vary in phase with relation to each other. This is because each coil is in a different position with respect to the flux lines of the field (see Fig. 7-15).

When windings, identical in size and of the proper phase relationships, are added to the single-phase alternator, a voltage output is obtained from each coil added. These voltages differ in phase but not in amplitude. The output from such a generator is known as a polyphase-output system. Systems of this type are generally two- or three-phase systems with the three-phase the most widely used. In fact, all the power generated by the large utilities is three-phase. Large air-conditioning and refrigeration installations use three-phase power.

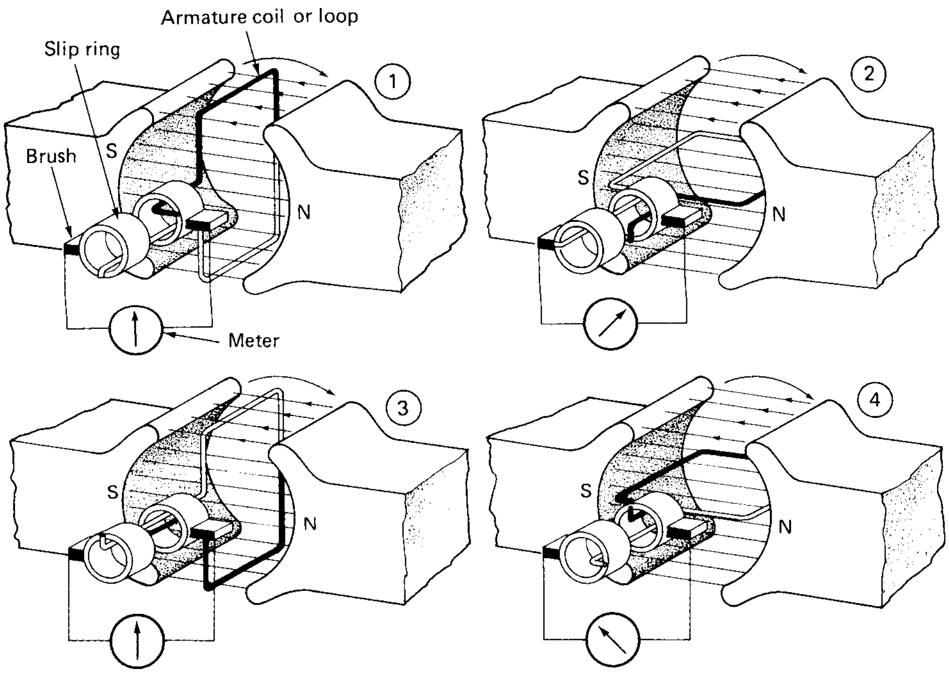

Figure 7-15 Generation of an ac voltage.

Three phase

A three-phase system is one in which the voltages have equal magnitudes and are displaced 120 electrical degrees from each other (see Fig. 7-16). The three windings are placed on the armature 120° apart (Fig. 7-17). As the armature is rotated, the outputs of the three windings are equal but out of phase by 120° (see Fig. 7-18).

Three-phase connections

Three-phase (3ϕ) windings are usually connected in either a delta or wye configuration. Each of these connections has definite electrical characteristics from which the designations delta and wye are derived (see Figs. 7-19 and 7-20).

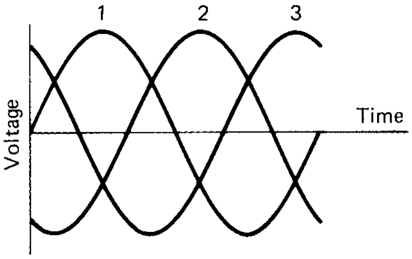

Figure 7-16 Three-phase voltages as induced in a generator stator windings.

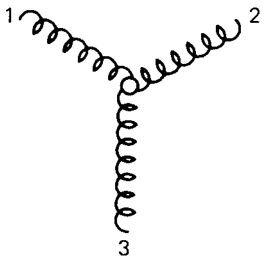

Figure 7-17 Schematic representation of a three-phase stator connected in a wye (Y) configuration.

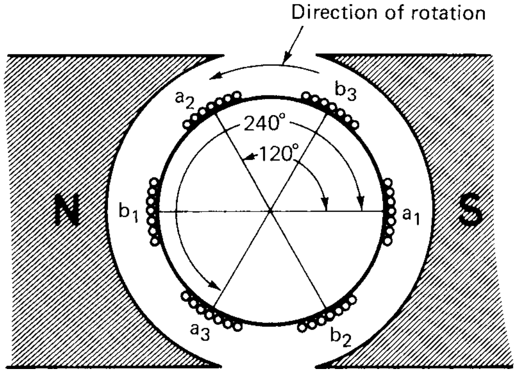
Direction of rotation
N
S
a_2
b_3
240°
120°
b_1
a_1
a_3
b_2

Figure 7-18 Basic three-phase alternator.

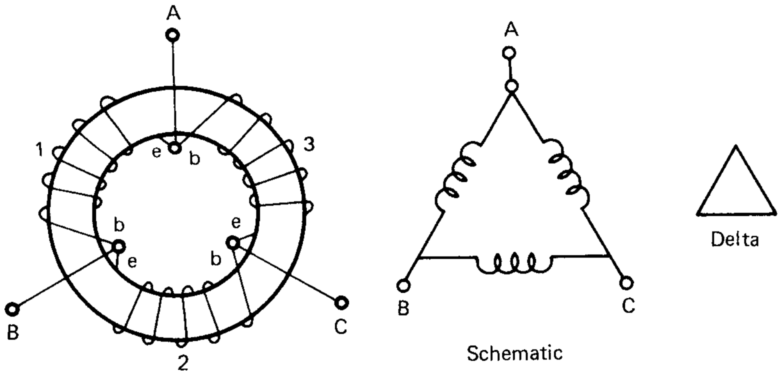
A
1
3
e
b
b
e
e
b
B
C
2
A
B
C
Schematic
Delta

Figure 7-19 Delta-connected ring.

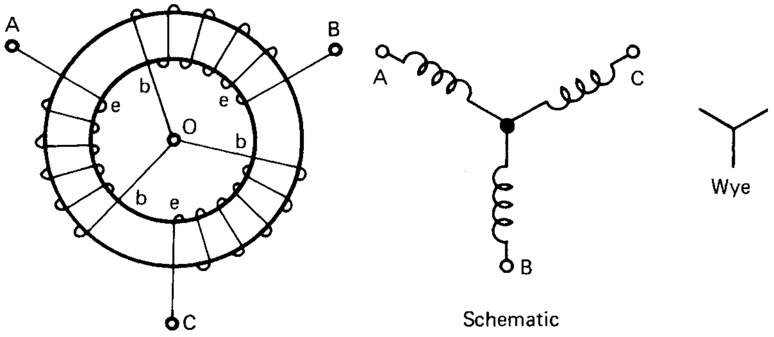
A
B
b
e
e
O
b
b
e
C
A
C
B
Schematic
Wye

Figure 7-20 Wye-connected ring.

Electrical Properties of Delta and Wye

Delta connection (Δ)

In a balanced circuit, when the generators are connected in delta, the voltage between any two lines is equal to that of a single phase. The line voltage and the voltage across any winding are in phase, but the line current is 30° or 150° out of phase with the current in any of the other windings (see Fig. 7-21). In the delta-connected generator, the line current from any one of the windings is found by multiplying the phase current by the square root of 3, which is 1.73.

Wye connection (Y)

In the wye connection, the current in the line is in phase with the current in the winding. The voltage between any two lines is not equal to the voltage of a single phase, but is equal to the vector sum of the two windings between the lines. The current in line A of Fig. 7-22 is the current flowing through the winding L_1; that in line B is the current flowing through the winding L_2; and the current flowing in line C is that of the winding L_3. Therefore, the current in any line is in phase with the current in the winding that it feeds. Since the line voltage is the vector sum of the voltages across any two coils, the line voltage E_L and the voltage across the winding E_ϕ are 30° out of phase. The line voltage may be found by multiplying the voltage of any winding E_ϕ by 1.73.

Delta and wye summarized

The properties of delta connections may be summarized as follows: The three windings of the delta connection form a *closed loop*. The sum of the three equal voltages, which are 120° out of phase, is zero. Thus, the circulating current in the closed loop formed by the windings is zero. The magnitude of any line current is equal to the square root of 3 (1.73) times the magnitude of any phase current.

Properties of the wye connection do not form a closed loop. The magnitude of the voltage between any two lines equals the magnitude of any phase voltage times the square root of 3, that is,

$$E_L = \sqrt{3} \times E_\phi$$

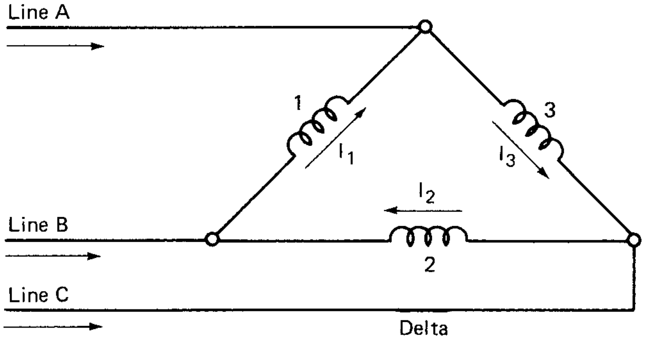

Figure 7-21 Delta-connected currents.

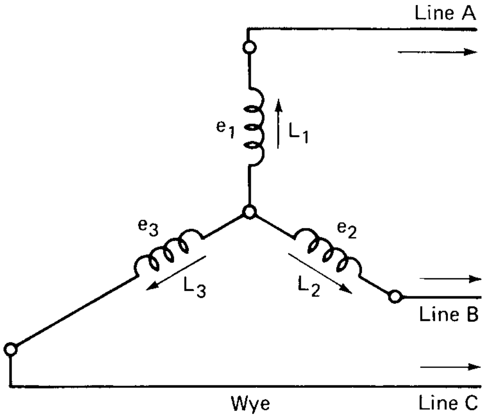

Figure 7-22 Wye-connected currents.

The current in any winding equals the current in the line.

Applications of three-phase power to motors will be discussed in more detail in Chapter 12. Connections of power supply transformers to furnish three-phase power are discussed in Chapter 8. As previously mentioned, three-phase power is used in commercial and industrial air-conditioning and refrigeration equipment and in some heating units.

Review Questions

1. What type of current is obtained from a battery?
2. What is the difference between a cell and a battery?
3. What are three types of dry cells in popular demand today?
4. What is meant by the ampere-hour rating of a battery?
5. How are cells connected in series?
6. Why are cells connected in series?
7. Why are cells connected in parallel?
8. What causes batteries to have a shortened life?
9. What is a sine wave?
10. What is the difference between alternating current and direct current?
11. What is a hertz? Where do you see the term used most frequently?
12. What are the three values used to describe ac?
13. What does rms mean? Where is the term used?
14. What is meant by polyphase ac?
15. What is the difference between a wye connection and a delta connection?
16. Where will you find three-phase power being used?

This page intentionally left blank

Chapter

8

Inductors and Transformers

Performance Objectives

- Know the two names for coil.
- Know what the term inductance means.
- Know the four things that determine inductance.
- Know what self-inductance is.
- Know what mutual inductance is.
- Know the effects of a counter emf.
- Know what out of phase means.
- Know what inductive reactance is.
- Know what a transformer does.
- Know what a variac is.
- Know the three losses associated with a transformer.
- Know what is meant by inductive kickback.

The magnetic field that surrounds a coil produces an opposition to a change in current in that coil. This opposition is not the same as resistance is to direct current. A coil has an opposition to alternating current because the ac is constantly changing its direction and amplitude. The opposition that a coil presents to ac is called *inductive reactance*. Inductive reactance must be taken into consideration when a motor or transformer is in an electrical circuit. It is also very important in the design of start circuits for compressors. Coils are used for filter circuits in electronic power supplies. There are many uses for inductors and for inductive reactance.

Coils are also referred to as inductors and chokes. The inductor is used in a circuit to retard the flow of ac and to allow dc to pass. Inductors are diagrammed in circuits through the use of the symbol shown in Fig. 8-1.

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

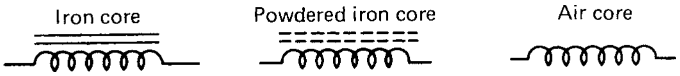

Figure 8-1 Symbols for inductors.

The property of a coil that opposes any change in circuit current is known as *inductance*. The symbol for inductance is L. Inductance is measured in *henrys*, abbreviated H. Since the henry may be too big for most electronic circuits, it is broken down further into millihenrys (mH) and microhenrys (μH). Motors and transformers have inductance and therefore inductive reactance when ac is applied.

Changing Inductance

The inductance of a coil depends on four factors:

1. *Number of turns:* The more turns a coil has, the greater its inductance.
2. *Diameter of the coil:* The wider the cross-section of a coil, the higher the inductance is.
3. *Permeability of the core:* The more permeable the core, the better are the magnetic properties and the higher the inductance.
4. *Length of the coil:* The shorter the coil, the higher the inductance is. Coil length and inductance are said to be inversely proportional.

Figure 8-2 shows how the inductance of a coil can be increased by adding turns, by adding a core, and by the way layers are wound.

Self-Inductance

The ability of a conductor to induce voltage in itself when the current changes is called *self-inductance*. This ability can be valuable. Self-inductance is also measured in henrys.

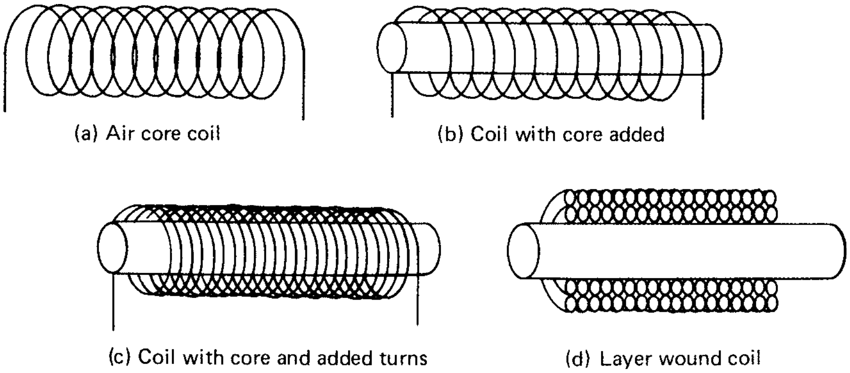

Figure 8-2 Changing the inductance of a coil.

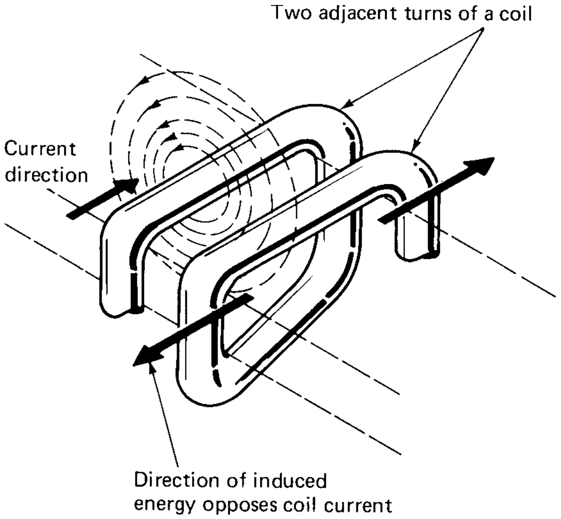
Two adjacent turns of a coil
Current direction
Direction of induced energy opposes coil current

Figure 8-3 Self-inductance created by placing coils adjacent to one another. Note how the magnetic field cuts adjacent turns.

Self-inductance is produced by varying current in a coil. The current produces a magnetic field around each turn of the coil. The field around each turn then cuts across other turns (see Fig. 8-3). A changing current in the turns of a coil produces a changing magnetic field. Self-induced voltage occurs when the changing magnetic field cuts the turns of the coil and induces a voltage across the coil.

The amount of induced voltage depends on the rate of change of current. The faster the current changes, per unit of time, the higher is the self-induced voltage. Take a look at Fig. 8-4 and see how two points on a sine wave demonstrate the difference in ratio of current change.

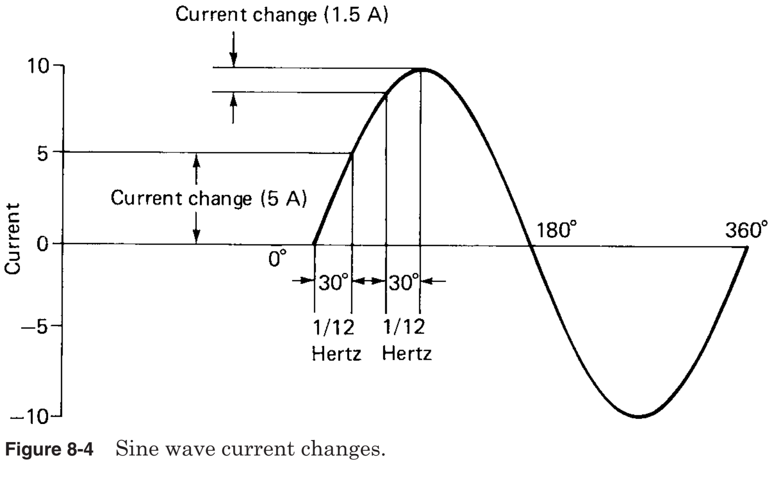

Figure 8-4 Sine wave current changes.

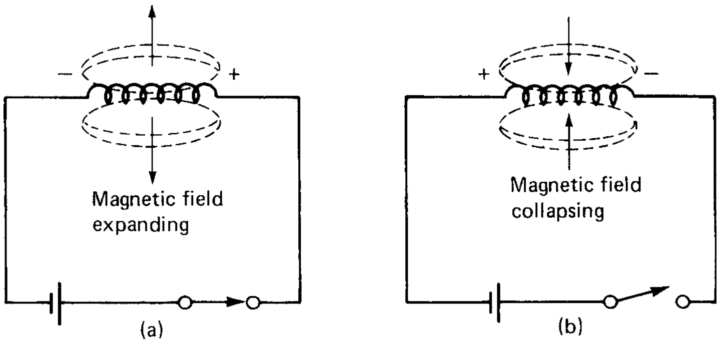

Figure 8-5 Counter emf.

The self-induced voltage is in the opposite direction to the magnetic field that produced it. This voltage is called a *counter emf (cemf).*

The cemf is shown in Fig. 8-5a. A current increases rapidly from zero to maximum. This causes the magnetic field to expand. The expanding magnetic field produces cemf that moves away from the input current. The cemf cuts across windings in the path of the current. The increasing cemf opposes the increase in current in the circuit.

Now take a look at Fig. 8-5b. The circuit is broken by a switch. When this happens, the magnetic field collapses. At this time, the current in the circuit changes from its maximum value to zero. The collapsing field induces voltage across the coil. The induced voltage opposes the decrease in current. The current is thus prevented from dropping quickly to zero. The gradual decline in current level due to self-inductance is shown in Fig. 8-6.

In an inductor, the lag created between current and voltage is 90°. Thus, in a purely inductive circuit the voltage will lead the current as shown in Fig. 8-7. The coil opposes—increases or decreases—in current. This is valuable in filter circuits. In such a situation, inductance protects devices against sudden increases or drops in electrical current. An inductor, then, creates a lag or delay between the voltage and current. To show this, consider a purely inductive ac circuit, one with an inductor only. In such a circuit, current lags voltage by 90°. When this condition exists, the voltage and current are said to be out of phase.

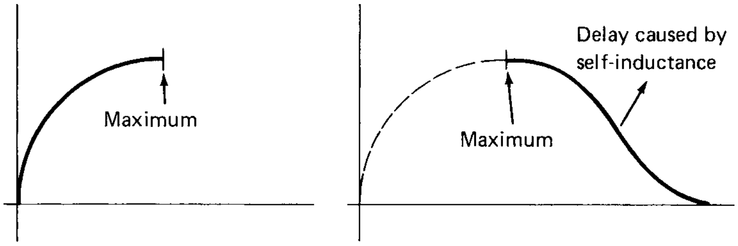

Figure 8-6 Time lag in an inductor.

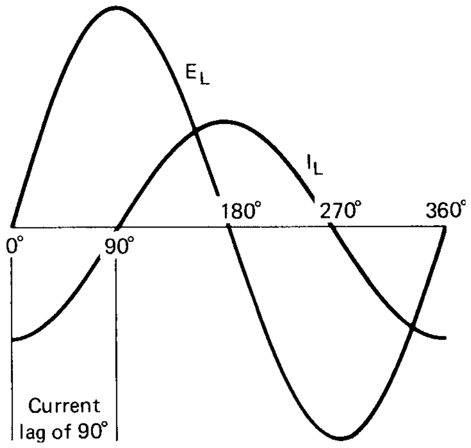

Figure 8-7 Voltage leads current by 90°.

Mutual Inductance

A condition in which two circuits share the energy of one circuit is called *mutual inductance*. The energy in one circuit is transferred to the other circuit. The two coils have a mutual inductance of 1 henry if a current change of 1 ampere per second in one coil induces 1 volt in the other coil. Magnetic lines of flux cause a coupling between the circuits.

A mutual inductance circuit is shown in Fig. 8-8. The symbol at the left of this diagram indicates an ac source. Coil L_2 is placed near coil L_1. The magnetic flux surrounds the wires making up L_2. This induces a voltage across coil L_2.

As soon as the magnetic field is built up by the ac, the current flow is reversed. The magnetic field begins to collapse. As the magnetic field collapses, it also induces a voltage in L_2. This induced voltage moves in the opposite direction from when it built up to a maximum.

L_2 is part of a circuit with an electrical load. Current flows through the load. Current in the circuit of L_2 also produces a magnetic field. This is in opposition to the field that induced the current in L_2. If a heavier load is placed in the circuit with L_2, a greater opposition is presented to the induced current. Thus, L_1 must draw more current for its source. It needs this additional current to make the increased opposition provided by the increased load across L_2. There is no load on L_1 until L_2 has a load connected across its terminals. This is the basis of how the transformer operates.

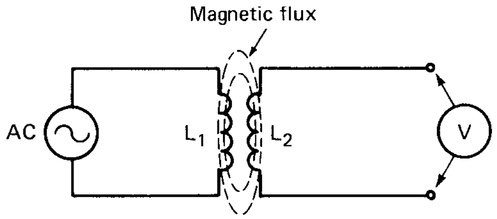

Figure 8-8 Mutual inductance.

Inductive Reactance

As we have already seen, an inductor reacts to a changing magnetic field produced by a changing current. When alternating current passes through an inductor, a phase shift occurs. Voltage and current that were in phase become out of phase. Phase shifting occurs repeatedly. This results from the opposition to ac by an inductor. This opposition is referred to as reactance. X is the symbol for reactance. Since the opposition takes place in an inductor, it is called *inductive reactance* and is represented by X_L. It is an opposition so it is measured in ohms.

This type of opposition, or reactance, does not affect dc. The only opposition to dc comes from the resistance of the conductor (wire) that makes up the coil.

Measuring X_L

A number of factors determine X_L. One is the frequency of the alternating current. Another is the size of the inductor. The formula used to calculate X_L is

$$X_L = 2\pi f L$$

Inductive reactance is not measured by an ohmmeter, but is arrived at by mathematical calculation. In this equation, f is for frequency, measured in hertz. L is inductance, measured in henrys. π is a standard mathematical term with a value of 3.14. So 2π equals 6.28. Thus, if either frequency (f) or inductance (L) is increased, inductive reactance (X_L) also increases. If either of these factors is reduced, inductive reactance also becomes smaller.

Power in an Inductive Circuit

In an inductive circuit, power is absorbed by the coil during the time that current is increasing. This power is stored in the generated magnetic field. However, all the power absorbed during the current rise is returned to the source by the collapse of the magnetic field. This means that the average power consumed by the inductance is zero. Whatever it takes, it puts back.

The current, voltage, and power in a purely inductive circuit are shown in Fig. 8-9. During the first quarter-hertz of the sine wave (0° to 90°), the current is rising and power is being supplied from the source. The instantaneous power, which is represented by the shaded areas in the illustration, is the product of the instantaneous voltage and instantaneous current.

The areas on the positive side of the zero axis represent power being transferred to a magnetic field from the electrical source. The areas on the negative side of the zero axis represent power being returned to the circuit. This occurs when the current decreases from its peak value toward zero on the 90° to 180° part of the hertz or cycle. The magnetic field then decreases in intensity and delivers power back to the source. There is now a transfer of magnetic energy back to electrical energy (the magnetic field is said to be collapsing). The source is now receiving power from the collapsing magnetic field. Since the power supplied and the power taken back are identical, the average power is zero.

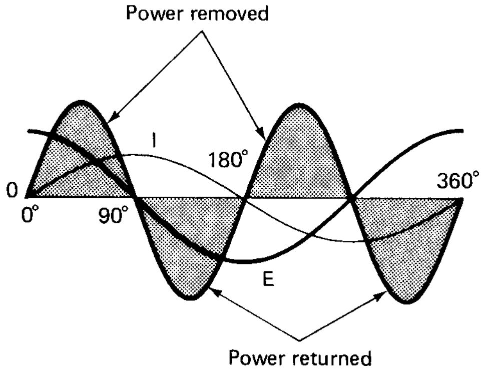

Figure 8-9 Power in an inductive circuit.

Before you think an electric motor or fluorescent lamp can be operated without consuming energy, you have to bear in mind that there is no pure inductor. There is always some resistance associated with any practical coil, because the conductors (wires) that make up the turns of the windings have resistance. It is this resistance that consumes power.

Uses for Inductive Reactance

Inductive reactance can be very useful. It holds back the current in an ac motor and makes sure it does not draw too much from the line. A transformer is able to sit on the line without having a load on it due to inductive reactance. Filters also use the properties of coils to aid in the selection of the proper frequencies for cross-over networks in stereo speaker systems and for making sure the right station is tuned in on the television and the radio. In refrigeration, air-conditioning, and heating work, you will be primarily concerned with what inductance can do when it is present in a solenoid, relay, motor, or transformer.

Transformers

So far you have looked at the terms inductance, mutual inductance, and self-inductance. Inductance is the opposition to current flow presented by a coil or inductor. A changing magnetic field produced by a changing electrical current can also be used in making what is referred to as a transformer. When the magnetic field produced by one coil is allowed to cut the windings of another coil nearby, it produces the transformer effect, or it introduces a voltage in the second nearby winding. This voltage and the amount of current available can be controlled by utilizing knowledge gleaned from years of experience and from some simple mathematical formulas.

Mutual inductance in a transformer

Two coils placed close to one another will allow the magnetic flux from each coil to be linked to the other. This is the basis of mutual inductance. A change in the current in one coil produces a magnetic field that cuts the turns of the second coil and induces a voltage. A mutual inductance thus exists between the two coils. This mutual linkage causes the effective inductance of each coil to be different from that of the self-inductance. The unit of mutual inductance is also the henry and is designated by the symbol L_M. A mutual inductance of 1 henry exists when a current change of 1 ampere per second in the first coil (primary) induces 1 volt into the second coil (secondary).

The mutual inductance between two coils depends on the closeness of the two coils, the way in which the coils are wound, and the reluctance of the mutual magnetic path. Three ways to increase the mutual inductance of two coils are shown in Fig. 8-10. Two coils that are coupled by a magnetic core have a higher mutual inductance. When the magnetic core makes a complete loop, as seen in Fig. 8-8, nearly all the magnetic flux of one coil is coupled to the other coil, and the mutual inductance is maximum. The amount of flux linkage between two coils is called the *coefficient of coupling*.

Iron-core transformer

A device that transfers electrical power from one coil to another is called a *transformer*. Transformers may also change the value of voltage during the transfer. A transformer functions with no physical connection between the source and the receiving conductors. The principle used is mutual inductance.

Current flows into a transformer primary coil. This current creates a magnetic flux. The magnetic flux, in effect, couples the primary coil with the secondary coil. Voltage is induced in the secondary coil. The induced voltage may be varied by increasing or decreasing the magnetic field. The result of a transformer's operation is the induction of emf in the secondary coil. This is transferred to a conductor connected to the secondary coil.

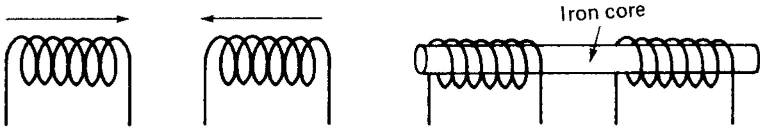

(a) Move the coils closer together.

(b) Use a magnetic core to couple the coils.

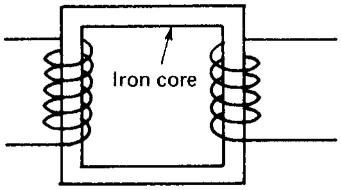

(c) Use a closed magnetic path.

Figure 8-10 Three ways to increase mutual inductance.

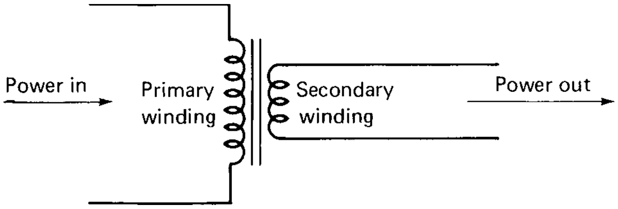

Figure 8-11 Step-down transformer.

Iron-core transformers use the mutual inductance principle to transform power between primary and secondary windings. An iron core provides a low reluctance path for the magnetic flux, and the coefficient of coupling between primary and secondary is so near unity (1) that it is considered as such in the designing of the transformer. The transformer is a highly efficient device since it has no moving parts; only current variations cause the transformer action to occur.

Changing magnetic fields are generated by ac so a transformer can be used to transfer ac power from primary to secondary (see Fig. 8-11). The two windings of a transformer are not connected, and the only means of transferring energy from the primary winding to the secondary winding is by way of the mutual coupling of a magnetic field.

Construction

Every transformer is built with at least two coils. One is a primary coil and the other is a secondary coil. The primary coil is called the *input*. This means that it brings electrical power into the transformer. The secondary coil is the *output*. It carries electrical power out of the transformer.

The transformer's iron core aids in the concentration of the magnetic field. The size of the core is the major determiner of the amount of current that can be drawn from the transformer, along with the size of the wire to handle the current. Wire size and core size determine the weight and physical size of the transformer. The higher the wattage rating or power output of the transformer, the larger the transformer is.

Refrigeration, heating, and air-conditioning circuits utilize a 24-volt transformer in most of the control circuits. This transformer steps down 120 volts to 24 volts. The National Electrical Code classifies this low-voltage control circuit as class 2.

Types of iron-core transformers

There are three basic types of iron-core transformers:

1. *Open-core:* This type is used in power transformers (see Fig. 8-12).
2. *Closed-core:* This type is more efficient than the open-core type. Figure 8-13 shows that the flux path is contained within the core. This increases the strength of the magnetic field. The transfer of energy is improved.

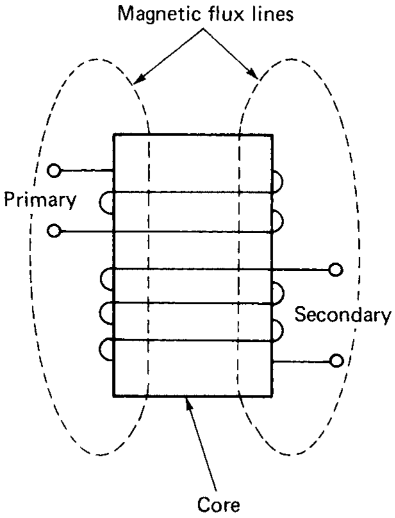
Magnetic flux lines
Primary
Secondary
Core

Figure 8-12 Open-core transformer.

3. *Shell-core:* This is the most efficient type of transformer. Figure 8-14 shows the design characteristics. Note how the magnetic flux pattern containment improves the efficiency.

Voltage Transfer

The transformer is a power transfer device. Power that goes into the primary is transferred to the secondary. Also, power required by the secondary is reflected back as a power requirement to the primary. The rule is simple: Power in is equal to power out, less any losses.

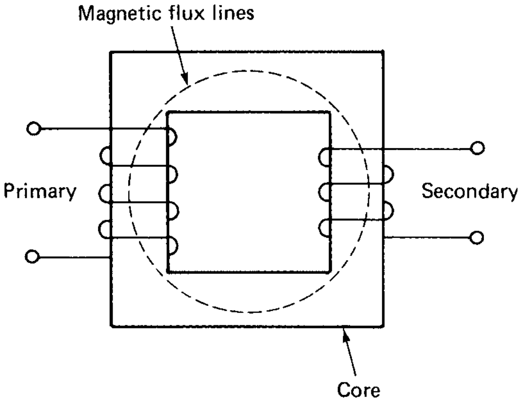
Magnetic flux lines
Primary
Secondary
Core

Figure 8-13 Closed-core transformer.

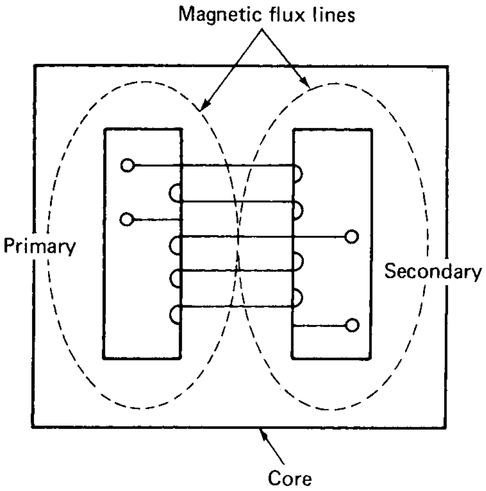

Figure 8-14 Shell-core transformer.

Thus, if no load is connected to the secondary coil, the current flowing into the primary is virtually zero. The only current required is the amount necessary to support part of the losses. This is almost zero. Current flows in the primary only when the secondary is connected and there is output current to a load.

Step up, step down

Transformers are either step up or step down. If input voltage is higher than output voltage, the device is known as a step-down transformer. In the step-up transformer, the output voltage is higher than the input.

Turns ratio

The input and output voltage relationship of a transformer depends on the turns ratio. The turns ratio describes the relationship of the windings of the primary and secondary coils. The first number in a given ratio is for the secondary. The second number is for the primary.

To find the turns ratio of a transformer, divide the number of turns of the secondary by the number of turns of the primary. For example, if there are 1000 turns in the secondary and 100 turns in the primary, divide 1000 by 100. This produces a ratio of 10 to 1. This is written as 10:1. The formula used to find the turns ratio is

$$\text{Turns ratio} = \frac{1000}{100} = \frac{10}{1} \quad \text{or} \quad 10{:}1$$

If the transformer has a turns ratio of 10:1, the voltage ratio is 10:1. If 100 volts is applied to a 10:1 transformer primary, the secondary would have an output of 1000 volts.

There are other types of transformers that you should be familiar with. This will make it possible for you to make some substitutions, if needed, and give you some idea as to which transformers cannot be used for control circuits such as you will experience in refrigeration, air-conditioning, and heating equipment.

Power Transformers

Power transformers use multiple secondary coils. These can deliver a number of secondary voltages. Examples of the power transformer are shown in Figs. 8-15 and 8-16. These transformers are used for a wide variety of jobs. They may provide power for factories. They may power broadcast stations. Or they may be built into special equipment such as the power supply of a transmitter, television set, or electronics equipment. As you can see from the schematic, there are both step-up and step-down transformers. They have one primary coil with various ratios of secondary coils wound on top of the primary. They can be easily identified since they will have more than the usual four leads coming out of the case.

Audio Frequency and Radio Frequency Transformers

Audio transformers change voltages for use in the audio range of frequencies. Audio frequencies are from 16 to 16,000 hertz. An audio transformer is shown in Fig. 8-17.

To deliver voltage at higher frequencies, no core is used. This is because an iron core consumes too much power above 16,000 hertz. Air-core transformers, therefore, are usually used in radio frequency circuits and are called

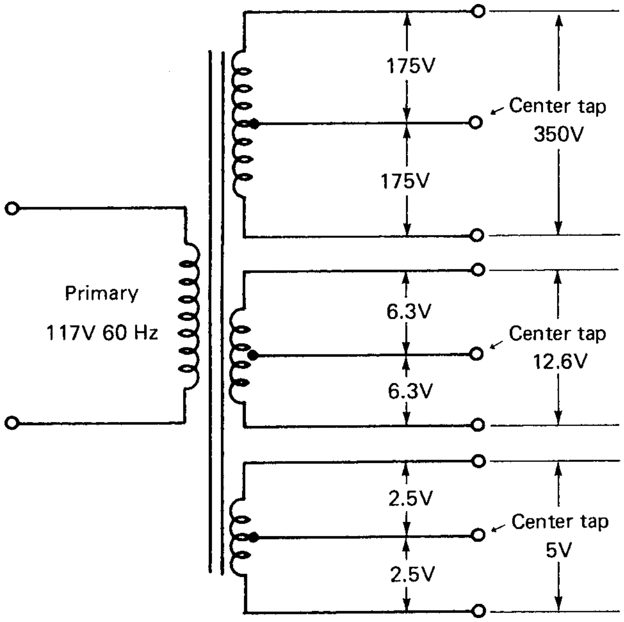

Figure 8-15 Schematic of a power transformer.

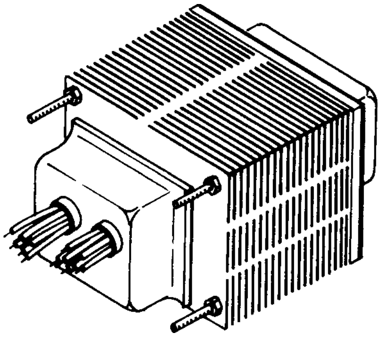

Figure 8-16 Power transformer used in electronic devices.

rf transformers. Electronics equipment have special miniature transformers to do certain jobs.

These are not usable on line frequencies such as 60 hertz for control circuits in heating and air-conditioning equipment.

Autotransformers

Autotransformers have only one coil. Input and output come from a single coil. The amount of voltage output is determined by the position of the secondary tap (connection point). Placement of the secondary tap changes the turns ratio.

When an autotransformer is used to step up the voltage, part of the single winding is used as the primary. The entire winding acts as the secondary (see Fig. 8-18). Power is transferred from the primary to the secondary by the changing magnetic field that is concentrated by the core material. The turns ratio determines the voltage output.

An autotransformer does not provide isolation, as do the regular primary coil and secondary coil types that are not physically connected. The autotransformer will not serve to remove the ground found in most home and industrial wiring systems because the primary and secondary use the same turns. This type of

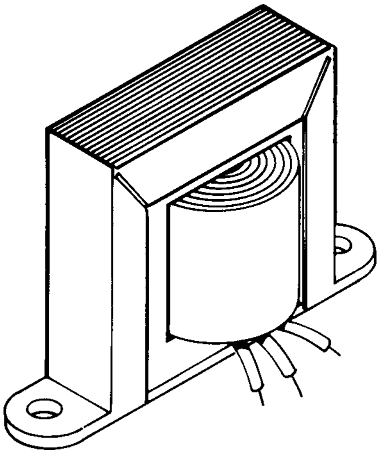

Figure 8-17 Audio-type transformer.

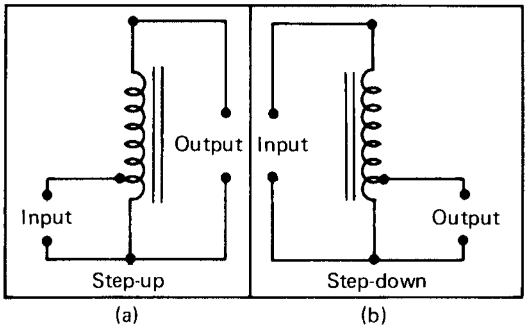

Figure 8-18 Autotransformer.

unit can be made into a variable transformer. A variable transformer can produce differing output voltages. A slider arm is placed over the windings. The slider makes contact with points where the insulation on the wire has been removed. This type of variable transformer is called a *variac*.

Transformer Losses

There are transformer losses, even though most of them can be minimized. Nevertheless, there are some that must be taken into consideration when designing a transformer for maximum efficiency. Three types of losses take power away from the output in a transformer and must be dealt with when the transformer is designed.

1. *Copper losses:* These are due to the resistance of the wire in the primary and secondary coils. Large size wire helps to minimize these losses.
2. *Hysteresis losses:* These are due to the properties of the iron core. Iron is slow to change polarity. The delay is known as hysteresis, or slowness to change properties (see Fig. 8-19). The magnetic domains (or molecules) of the iron core are facing in one direction. The domains (or molecules) are mixed in the middle illustration. This is in response to a change in magnetic fields. In the last illustration, the domains (or molecules) are lined up in the opposite direction. This happens when the magnetizing force of ac is at its peak. Hysteresis losses in

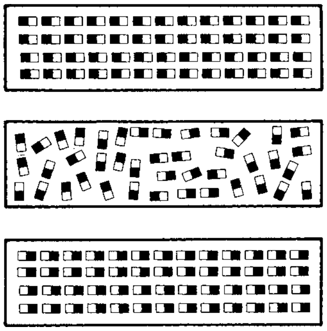

Figure 8-19 Hysteresis losses due to alignment of domains (or molecules) in an iron core.

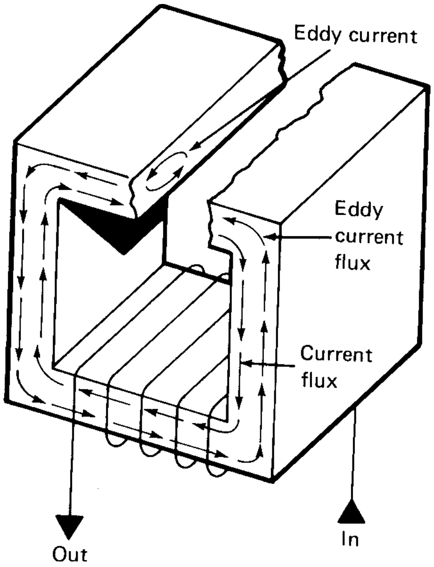

Figure 8-20 Eddy currents in a transformer core.

transformers are minimized by using silicon steel for the iron core. This type of metal offers very little opposition to changing polarity.

3. *Eddy currents:* These are very small but create losses in the transformer core. They are small currents created in the core when the magnetic fields change (see Fig. 8-20). Eddy currents generate heat. They flow in a direction opposite to the current that induced them. The effect is to resist the flow of current in the core. Eddy currents can be minimized by using laminations. Laminating is the building of an object through the use of several layers of material. When this method is used, each lamination is varnished. Varnishing insulates the layers from each other. This increases resistance to eddy currents. Figure 8-21 shows how a solid core has high eddy current losses and a laminated core practically eliminates current losses.

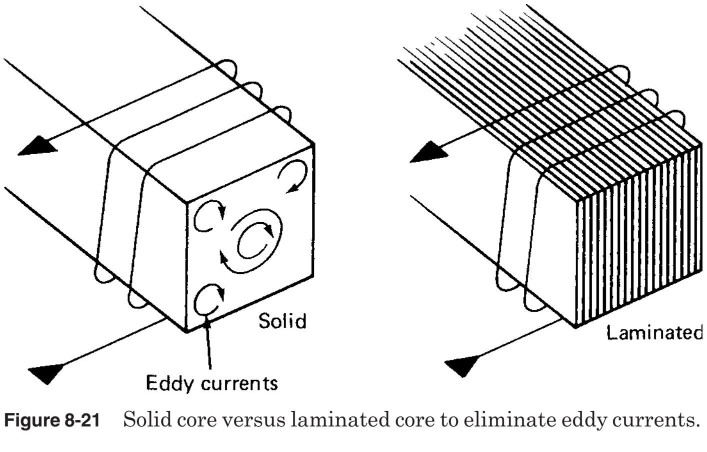

Figure 8-21 Solid core versus laminated core to eliminate eddy currents.

Inductive Circuits

Inductors have a definite effect on the operation of the circuits in which they are located. Inductors, remember, can be a coil of wire, the windings of an electric motor, or the ballast from a fluorescent lamp. Relay coils and solenoid coils have inductance, as dc transformers. These devices have definite functions in circuits used in heating, air conditioning, and refrigeration. Their reaction to one another and to resistors in the same circuit is very important. They can cause some interesting problems that need to be understood for you to be able to properly and safely troubleshoot them.

The response of an inductor to dc is quite different from its response to ac. In ac circuits, the inductor responds to the changing current by producing a self-induced voltage that tends to prevent any change in the amplitude of the current. In a dc circuit, there is no inductive opposition to the current because the current is unchanging. Therefore, in a dc circuit the amount of current depends entirely on the total resistance in the circuit. This includes resistors, wiring, and coil winding resistance.

The dc current in an inductor is determined by the dc resistance. Assume that an ac and a dc voltage of the same value are applied to identical inductors. Will there be a difference in the values of the dc and ac currents? The current is higher for the dc circuit because it is determined by the resistance of the coil windings. This value is usually very small. For the ac circuit, the current is determined by the impedance (impedance is the combination of resistance and reactance). Furthermore, as the frequency of the alternating current is increased, the coil current decreases because of the increasing reactance of the inductor.

Figure 8-22 shows how the coil reacts with dc applied. When the switch in the circuit is closed, a certain time is required for the magnetic field to expand to the maximum value determined by the current in the circuit. During this interval, the magnetic flux is changing, and the counter emf prevents the current from rising immediately to the final value determined by the resistance. The graph in Fig. 8-22b shows what happens. At time zero, the switch is closed and the current begins to increase. When the switch is first closed, the rate of flux change is maximum, and, as shown by the voltage curve E_L, a maximum voltage appears across the coil.

However as the current I rises, the rate of flux change slows down. The voltage across the coil decreases as the current rises.

This process continues until the current builds up to the maximum value determined by the dc resistance of the circuit. At this point, the magnetic field that has built up around the coil becomes stationary and unchanging, and the voltage drop across the coil, assuming the coil has no resistance, falls to zero. This is the steady state condition. As long as the switch remains closed, this amount of current will be present in the circuit. The initial buildup period is brief, and in most dc circuits the time required to reach a steady state is very short. The delay can be witnessed in some circuits where it takes some time for the current to build up to its maximum value.

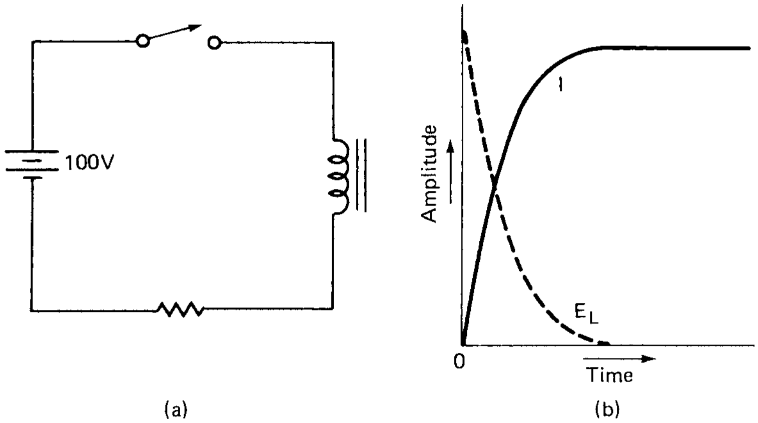

Figure 8-22 Current rises from zero to maximum. Maximum is determined by resistance in the circuit. The voltage across the coil drops from the applied voltage to zero.

Keep in mind that the maximum amount of current is dumped back into the circuit when the switch is opened. This collapsing magnetic field induces an emf of large amounts across the switch when it is opened. This is one reason why switches must be rated for inductive loads. This inductive kickback can cause an arc over and damage the switch if it is not properly rated to handle motors or inductive loads.

In ac resistor-inductor circuits, the current is determined by the impedance of the circuit. The higher the reactance and resistance of the circuit, the lower is the series current. The voltage drop across the inductor is determined by its reactance. Reactance is determined by both frequency and inductance. In a dc resistor-inductor circuit, a short but definite time is needed for the current to rise to its maximum value. This maximum current value is determined by the resistance of the circuit. The higher the resistance in the circuit, the lower the current is. However, assuming that the same two components are used in an ac and a dc circuit, the current present with the application of a dc voltage will be higher than the current present when an ac voltage of the same value is applied to the circuit. The dc voltage drop across the inductor in the dc circuit depends on the winding resistance of the inductor. If the winding resistance is assumed to be zero, the voltage drop across the inductor will be zero in a dc circuit.

Special handling

In handling circuits of this nature, remember the inductor in the circuit can cause a mild shock when you use an ohmmeter to check its continuity. The ohmmeter has a dc power supply that can excite the windings of a motor, solenoid coil, or relay coil. If you get across the terminals being tested, it is possible to obtain a mild shock from a circuit you may assume to be de-energized.

Utilizing the Inductive Delay

When a start capacitor is used in series with a compressor motor to cause it to start, it is taken out of the circuit within a few seconds. This is so that the capacitor does not explode from having ac applied to it. The delay in energizing the relay to remove the capacitor from the circuit is caused by the inductance of the relay coil and the inductance of the motor winding. The delay is caused by inductive reactance and the normal delay caused by a coil to a changing current. This is one practical use for the inductive reactance of a coil.

Review Questions

1. What are two other names for a coil?
2. What does the term inductance mean? What is its symbol?
3. What four things determine the inductance of a coil?
4. What is self-inductance?
5. What is mutual inductance?
6. What is meant by counter emf?
7. What does out of phase mean?
8. What is inductive reactance? What is its symbol?
9. List some uses for inductive reactance.
10. What determines mutual inductance?
11. What does a transformer do? How is it made?
12. How many coils does a transformer have?
13. What is meant by the turns ratio in a transformer?
14. What is the difference between an audio transformer and a radio frequency transformer?
15. How is an autotransformer different from a power transformer?
16. What is a variac?
17. What are the three types of transformers?
18. What are the three losses associated with a transformer?
19. What is meant by inductive kickback?

Chapter 9

Capacitors and Capacitive Reactance

Performance Objectives

- Know what capacitance is.
- Know how a capacitor works.
- Know the three factors that determine a capacitor's capacity.
- Know the basic unit of measurement for capacitance.
- Know six types of widely used capacitors.
- Know how an electrolytic capacitor is made.
- Know what can cause an electrolytic capacitor to explode.
- Know the meaning of WVDC.
- Know why electrolytics are polarized.
- Know what capacitive reactance is.
- Know why a capacitor can cause a lagging voltage.

To cause electric motors to start when single-phase power is applied, it is necessary to make it two phase or more. The split-phase motor was developed to utilize the single-phase power source, but it also needed an additional component in its start circuit to get it moving from a standing condition.

With the development of high-quality and high-capacity electrolytic capacitors, a variation of the split-phase motor known as the capacitor-start motor, has been made. Almost all fractional-horsepower motors in use today on refrigerators, oil burners, washing machines, table saws, drill presses, and similar devices are capacitor start. A capacitor motor has a high starting current and the ability to develop about four times its rated horsepower if it is suddenly overloaded.

In this adaptation of the split-phase motor, the start winding and the run winding have the same size and resistance value. The phase shift between the currents of the two windings is obtained by means of capacitors connected in

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

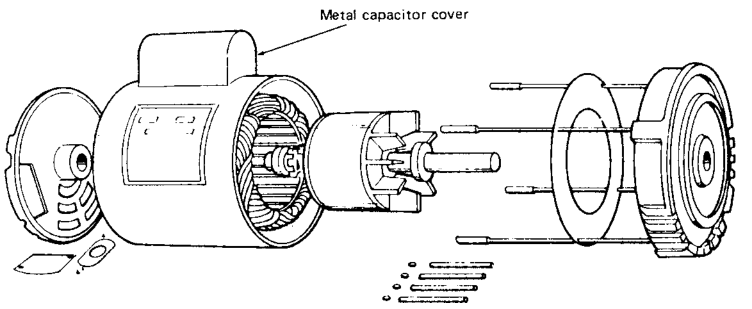

Figure 9-1 Exploded view of a capacitor-start motor.

series with the start winding. Capacitor-start motors have a starting torque comparable to their torque at rated speed and can be used in places where the initial load is heavy. One such application is a compressor in an air conditioner or refrigerator. A centrifugal switch is required for disconnecting the start winding when the rotor speed is up to about 25% of the rated speed. See Fig. 9-1 for a disassembled capacitor-start motor. Note in Fig. 9-2, which also shows a capacitor-start motor, the centrifugal switch arrangement with the governor mechanism. The capacitor on this motor is shown on top in the motor frame. It is enclosed in the metal housing on top of the motor. More about motors will be presented in Chapter 11.

Now that you can see how important the capacitor is in air-conditioning, heating, and refrigeration circuits (especially motor circuits), you can see why you need a better understanding of how they operate in order to be able to make decisions about replacement, repair, and testing of the equipment using them.

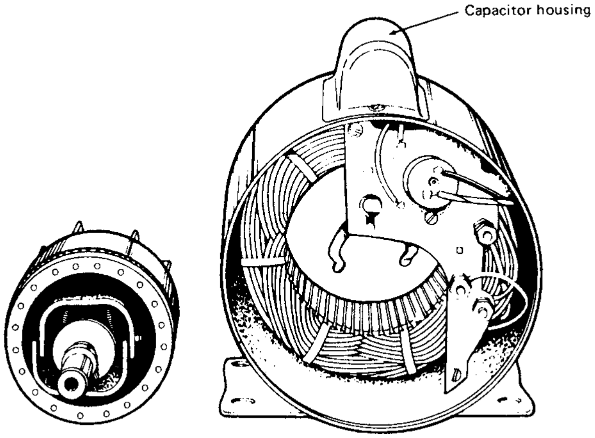

Figure 9-2 Single-phase stator and rotor.

The Capacitor

A *capacitor* is a device that opposes any change in circuit voltage. The property of a capacitor that opposes voltage change is called *capacitance*.

Capacitors make it possible to store electric energy. Electrons are held within a capacitor. This, in effect, is stored electricity. It is also known as an *electrostatic field*. Electrostatic fields hold electrons. When the buildup of electrons becomes great enough, the electrical potential is discharged. This process takes place in nature when clouds build up electrostatic charges; their discharge is seen as lightning.

Figure 9-3 shows the diagram of a simple capacitor. Two plates of a conductor material are isolated from one another. Between the plates is a dielectric. The dielectric material does not conduct electrons very easily. In fact, the dielectric is an insulator. The larger the surface area of the conductive material (the plates), the larger is the capacitor's capacitance.

How the Capacitor Works

If a capacitor has no charge of electrons, it is uncharged. This happens when no voltage is applied to the plates. An uncharged capacitor is shown in Fig. 9-4. Note the symbol for the capacitor in this drawing. This is the preferred way to show a capacitor: a straight and a curved line facing each other. Figure 9-4b shows the switch closed to position 1. This causes current to be applied to the capacitor. A difference in potential is created by the voltage source (the battery). This causes electrons to be transferred from the positive plate to the negative plate of the capacitor. This transfer continues as long as the voltage source is connected to the two plates and until the accumulated charge becomes equal to the voltage of the battery. That is, charging takes place until the capacitor is charged.

In Fig. 9-4c, the voltage has been removed. The switch is open. At this point, the potential difference, or charge, across the capacitor remains. That is, there

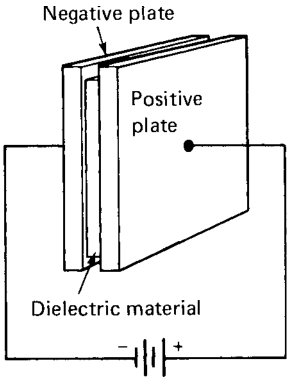

Figure 9-3 Simple capacitor.

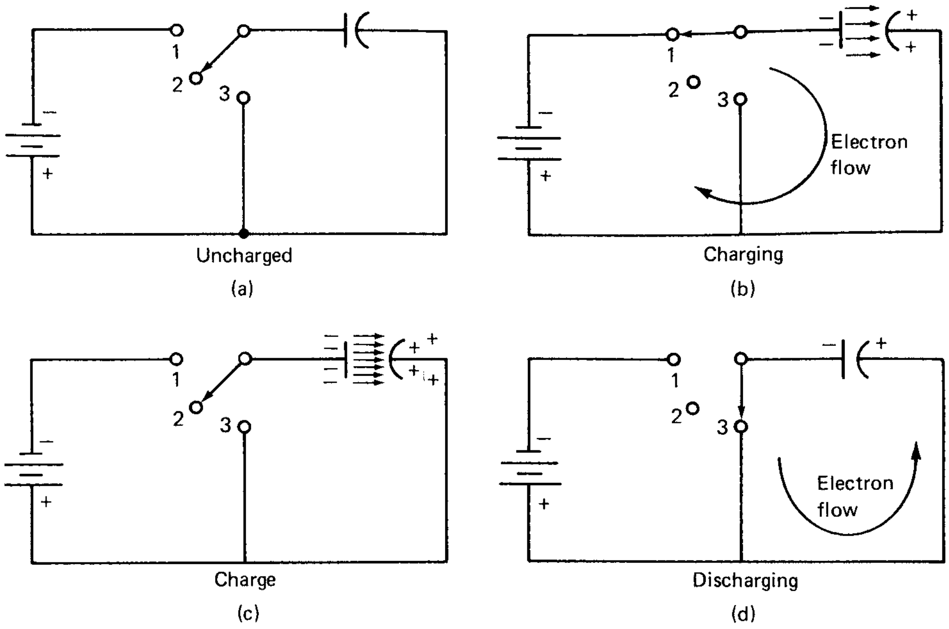

Figure 9-4 Charge-discharge action of a capacitor.

is still a surplus of electrons on the negative plate of the capacitor. This charge remains in place until a path is provided for discharging the excess electrons.

In Fig. 9-4d, the switch is moved to its third position. This opens the path for discharging the surplus electrons. Notice that the discharge path is in the opposite direction from the charge path. This demonstrates how a change in circuit voltages results in a change in the capacitor charge. Some electrons leave the excess or negative plate. They do this to try to keep the voltage in the circuit constant.

Capacity of a Capacitor

The two plates of a capacitor can be made from almost any material. The only requirement is that the material must allow electrons to collect on its surface. The dielectric between the plates of a capacitor is an insulating material. Some dielectrics are air, vacuum, plastic, mica, wood, Bakelite, paper, and oil.

Three factors determine the capacitance of a capacitor:

1. Area of the plates
2. Distance between the plates
3. Material used as a dielectric

The area of the plates determines the ability of a capacitor to hold electrons. The larger the plate area, the greater the capacitance is.

The distance between the plates determines the effect that the electrons on the plates have on one another. There is an electrostatic charge around each electron. Electrons on the plates store energy when voltage is applied. Capacitance increases as the plates are brought closer together. Capacitance decreases as the plates are moved apart.

The material used as a dielectric also determines the charge also. The thickness of the material means the plates are separated more or less by this dimension. But the type of material in the dielectric determines how much the electrostatic charges in the dielectric material react with one another. Some materials have the ability to increase the difference between the negative and positive plates.

Breakdown Voltage

The voltage at which the dielectric of a capacitor no longer functions as an insulator is the *breakdown voltage*. At this level, the capacitor permits the free flow of electrons through the dielectric. The result is known as a *shorted capacitor*. If the breakdown is partial, the result is a *leaky capacitor*. Thus, the dielectric strength of the material used is important to the functioning of a capacitor.

Basic Units of Capacitance

The basic unit for capacitance is the *farad* (F). A 1-farad capacitor stores 1 coulomb of electrons. It has a potential difference of 1 volt between its plates.

In most cases in electronics and electrical applications, the farad is too large a unit to be practical. It is broken down into the microfarad (μF) and the picofarad (pF), formerly called the micromicrofarad. The microfarad has been abbreviated UF, MF, MFD, Mfd. Today's standard abbreviation is μF.

The micromicrofarad was once used as the smaller unit of measurement of capacitance. Some older capacitors may be marked with a MMF, MMFD, UUF, or μμFD. Today's standard abbreviation is pF for picofarad.

The microfarad is one-millionth of a farad (0.000001). The picofarad is just what it says, one-millionth of one-millionth. The pico is today's accepted way of saying one-millionth of one-millionth. The Greek letter mu (μ) is used to represent micro. Micro is used universally as a prefix to indicate one-millionth.

Types of Capacitors

Six general types of capacitors are the most widely used:

1. Air
2. Ceramic
3. Mica
4. Electrolytic

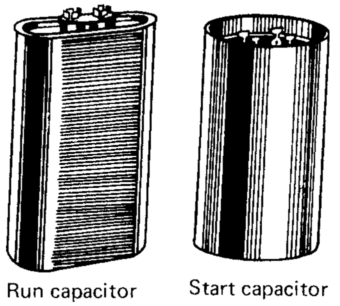
Run capacitor
Start capacitor

Figure 9-5 Electrolytic capacitors.

5. Paper
6. Tantalum

The electrolytic capacitor is marked with + and − and has polarity that must be observed when it is connected in a circuit. The other types do not need a polarity marking. Figure 9-5 shows an electrolytic capacitor that may be found in air-conditioning, refrigeration, and heating applications. It is used as ac motor run and start capacitors (see Fig. 9-6).

Air capacitors. Air capacitors have air for a dielectric. They are usually variable capacitors used in the tuning circuits of radios.

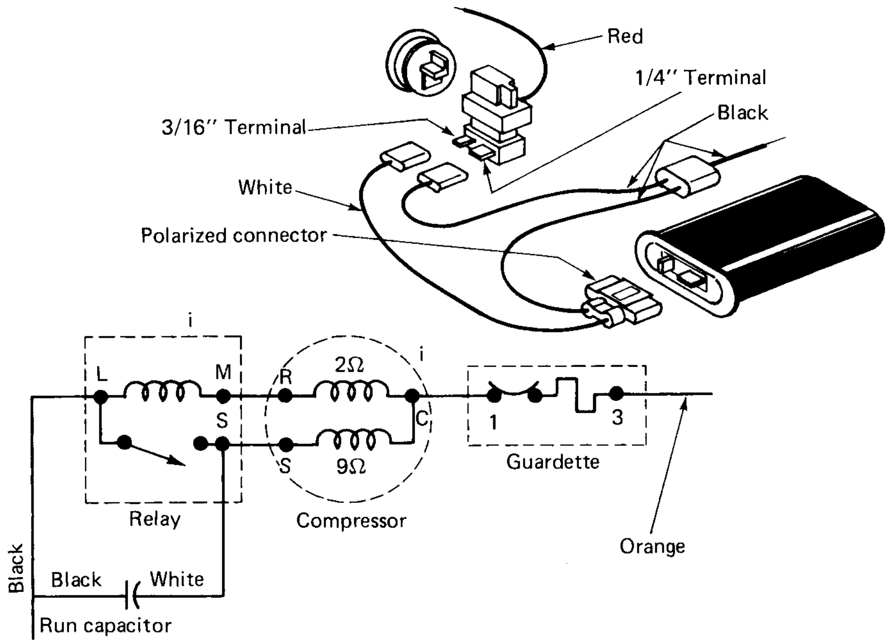
Red
1/4″ Terminal
Black
3/16″ Terminal
White
Polarized connector
i
L
M
S
Relay
i
R
2Ω
C
S
9Ω
Compressor
1
3
Guardette
Orange
Black
Black
White
Run capacitor

Figure 9-6 Electrical diagram of the capacitor-run type of refrigerator compressor. Note that the capacitor has a polarized plug so that it cannot be connected incorrectly.

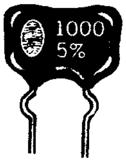

Figure 9-7 Mica capacitors.

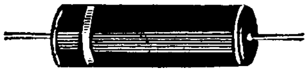

Figure 9-8 Paper capacitors.

Mica capacitors. Aluminum foil is used as the plate material in mica capacitors. Between the aluminum foil plates is a thin sheet of mica. Sometimes the mica is sprayed with a conducting paint. The paint then forms the plate on one side of the mica. Mica capacitors are usually sealed in Bakelite or some type of plastic (see Fig. 9-7).

Paper capacitors. Aluminum is also used as the plate material in paper capacitors (see Fig. 9-8). However, the plates are separated by a paper dielectric. The materials (paper and aluminum) are rolled into a cylindrical shape. A wire is connected to alternate ends of the foil and it is encased in plastic.

Ceramic capacitors. Ceramic dielectric materials make high-voltage capacitors. They have very little change in capacitance due to temperature changes. These small capacitors usually consist of a ceramic disc coated on both sides with silver. They are made in values from 1 picofarad up to 0.05 microfarad. Breakdown voltages of ceramic capacitors run as high as 10,000 volts and more.

Oil-filled capacitors. Oil-filled capacitors are paper capacitors encased in oil. They are sometimes referred to as bathtub capacitors (see Fig. 9-9). The main advantages of these capacitors are sturdy construction and high voltage breakdown ratings. They are used in places where grease and oil are likely to be encountered.

Electrolytic Capacitors

Electrolytic capacitors are very important to people working in the air-conditioning, refrigeration, and heating field.

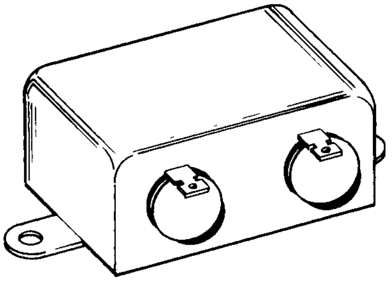

Figure 9-9 Oil-filled capacitor.

There are two types: wet and dry. The wet type uses a liquid electrolyte. It is rather large in physical size. But it has high capacitance. Wet electrolytics are generally used in transmitters and other large, stationary electronic equipment.

The dry electrolytic capacitor is used for capacitances above 1 microfarad. It can be made in sizes up to 1 farad. These units can be kept rather small in physical size for such large capacitances. This is done by using an oxide film as a dielectric (see Fig. 9-10). Very large sizes are used in computer power supplies for filtering the dc power supply.

Making an electrolytic capacitor

The way an electrolytic capacitor is made is very important because it also affects the way in which the capacitor is used in a circuit. To produce a capacitor of this type, dc voltage is applied to the electrolytic capacitor as part of the manufacturing process. This produces an electrolytic action. As a result, a molecule-thin layer of aluminum oxide with a thin film of gas is deposited. This gas is at the junction between the positive plate and the electrolyte.

The oxide film is a dielectric. There is capacitance between the positive plate and the electrolyte through the film. The negative plate provides a connection to the electrolyte. This film allows many layers of film to be placed in a case. Thus, large capacitances can be placed in smaller containers since the insulating film between the plates is so thin that the plates are very close together. And remember, the closer the plates are, the greater the capacitance.

Connecting electrolytics

Electrolytics have polarity (− and +). If these capacitors are not connected in a circuit properly, the oxide film is not formed. If connections are improper, no capacitance is available. Reverse electrolysis forms a gas when the capacitors are connected to opposite polarity. These capacitors become hot and may

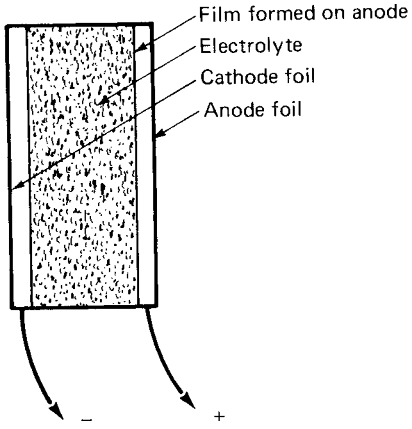

Figure 9-10 Electrolytic capacitor construction.

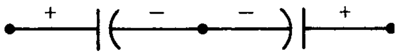

Figure 9-11 Back-to-back connection of electrolytics to provide a nonpolarized capacitor.

explode. Electrolytic capacitors may dry up and become leaky. These cause all kinds of problems since they may re-form to their rated capacitance once the power has been turned on for some time. So, the problems they present may be intermittent. These are the hardest types to pin down when troubleshooting. Electrolytic capacitors should be used in circuits in which approximately 75% of their working voltage is available. This keeps the capacitor formed.

If a refrigerator or air-conditioning unit has been left out of service for a year or more, it is possible for the electrolytic capacitor to present problems. It may have dried up or become leaky. Nonpolarized electrolytic capacitors are available. They do not have − and + polarity. They use two capacitors in series in the same case and connected in what is called a back-to-back configuration. This means that the two negative terminals are connected internally and the positive terminals are brought out of the case to be connected in the circuit. Nonpolarized electrolytics are used for ac circuits since the normal electrolytic capacitor would explode when ac is connected (see Fig. 9-11).

Working Voltage, Direct Current (WVDC)

The maximum safe operating (working) voltage of a capacitor in a dc circuit is identified as working voltage, direct current, or WVDC. When higher voltages are applied, the dielectric will deform and allow electrons to pass through, thus ruining the capacitor.

Capacitive Reactance

The degree to which a capacitor opposes the flow of current is called *capacitive reactance* (X_c). The capacitive reactance of a circuit is determined by capacitance and frequency. These two factors are used in a formula to determine the capacitive reactance of any given circuit. The formula is:

$$X_c = {}^1/_{2}\pi f C$$

In this formula, X_c is the symbol for capacitive reactance in ohms, C is the symbol for capacitance in farads, and f is the frequency in hertz.

You can see, then, that the changing of the frequency from 60 to 50 hertz would make a difference in the operation of the capacitor. To make the capacitors usable for both frequencies, they have been designed for about 55 hertz so they will operate on both frequencies. This means that the electrolytic capacitors you use for motors have to be properly sized for 50/60 hertz to prevent overheating. In most instances today, the capacitors are made for 50/60 hertz operation.

Note also that the size of the capacitor is very important with respect to the size of motor. In most instances, the electrolytic capacitors will be marked with a range of capacitances:

8–106 μF at 110/125 volts, 50/60 hertz

108–130 μF at 110/125 volts, 50/60 hertz

540–648 μF at 110/125 volts, 50/60 hertz

88–106 μF at 220/250 volts, 50/60 hertz

108–130 μF at 220/250 volts, 50/60 hertz

This type of electrolytic is equipped with two quick-disconnect terminals or some variation (see Fig. 9-12).

The main use for electrolytic capacitors in motor or compressor circuits is for providing starting torque for the device. If a motor must start under load, such as a compressor, it is usually a capacitor-start type.

Capacitor Causes a Lagging Voltage

In a circuit with a capacitor only, the voltage lags, or follows the current by 90°. This is because a purely capacitive circuit has no resistance (a condition that is never found in a practical circuit).

In a purely capacitive circuit, current and voltage are said to have a phase relationship. The phase difference is 90° (see Fig. 9-13).

The current and voltage relationships of capacitance and inductance are found in circuits with motors providing the inductance. By the proper use of capacitance, it is possible to bring the circuit back into phase and reduce the cur-

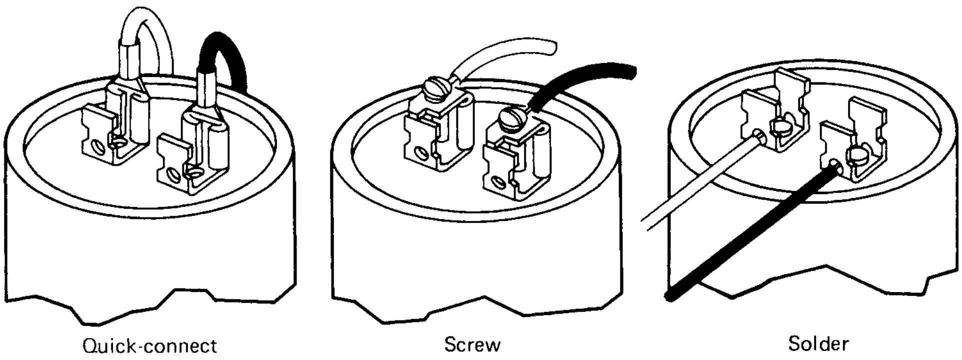

Figure 9-12 Three ways to connect electrolytics in a circuit.

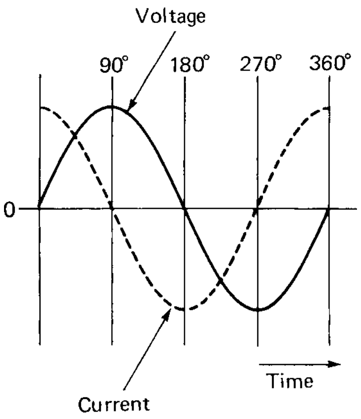

Figure 9-13 Voltage and current waveforms in a purely capacitive circuit.

rent drain from the power source. Run capacitors have this effect since they are left in the circuit at all times. This aids in keeping down the operating costs for electricity.

Checking Capacitors

Start capacitors and run capacitors have a tendency to become open, leaky, and short. These conditions can cause a variety of symptoms that the troubleshooter must become familiar with, to quickly recognize what to do to correct a problem. To check for these conditions, it is necessary to use an ohmmeter.

1. Set the ohmmeter to read its highest resistance range.
2. Sort the leads together and adjust the zero-adjust knob until you get a zero reading on the meter (the battery in the meter will charge the capacitor and, in so doing, a keen observer can detect any malfunctioning of the capacitor).
3. Use a piece of insulated wire with insulation removed at the ends to sort out the capacitor by placing it across the capacitor terminals. You may or may not get a spark. Do this five times to make sure you have discharged the capacitor. If the capacitor has a resistor across its terminals (see Fig. 9-14), remove one end by desoldering or pulling it loose from the connector. This must be done to prevent damage to the meter movement.

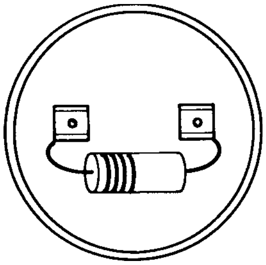

Figure 9-14 Bleeder resistor across capacitor terminals.

TABLE 9-1 Capacitor Troubles

Trouble	Probable cause	Remedy
Compressor will not start. However, it hums intermittently. Cycles with the protector.	Start capacitor open.	Replace start capacitor.
Compressor starts. Motor will not speed up enough to have start winding drop out of circuit.	Start capacitor weak. Run capacitor shorted.	Replace the capacitor. Disconnect run capacitor and check for short.
Compressor starts and runs. However, it cycles on the protector.	Run capacitor defective.	Check capacitance. Replace if found defective.
Start capacitors burn out.	Short cycling.	Reduce number of starts. They should not exceed 20 per hour.
	Prolonged operation with start winding in circuit.	Reduce the starting load. Install a new crankcase pressure limit valve. Increase low voltage if this is found to be the condition. Replace the relay if it is found to be defective.
	Wrong relay or wrong relay setting.	Replace the relay.
	Wrong capacitor.	Check specifications for correct sized capacitor. Be sure the MFD and WVDC are correct for this compressor.
	Water shorts out terminals of the capacitor.	Place capacitor so terminals will not get wet.
Run capacitors burn out. They spew their contents over the surface of anything nearby. This problem can usually be identified with a visual check.	Excessive line voltage.	Reduce line voltage. It should not exceed 10% of the motor rating.
	Light load with a high line voltage.	Reduce voltage if not within 10% overage limit.
	Voltage rating of capacitor too low.	Replace with capacitor of correct WVDC.
	Capacitor terminals shorted by water.	Place capacitor so the terminals will not get wet.

4. Place the meter probes across the capacitor and hold them there for at least 15 seconds. Watch the action of the meter at this time. The needle on the meter should swing all the way to the right or zero. Then it will gradually move back toward infinity and stop. This indicates that the capacitor is taking a charge and is probably all right to put in the circuit.

If the needle goes to zero and stays, the capacitor is shorted. Replace it. There are other ways of finding if a capacitor is doing its job. You may also check from each terminal of the capacitor to its metallic case. If you get a zero reading, replace the capacitor. The meter should read infinity from each terminal to the case to be usable. A troubleshooting guide for capacitors is given in Table 9-1.

Review Questions

1. What is capacitance? What is its symbol?

2. How does a capacitor work?
3. What three factors determine the capacity of a capacitor?
4. What is the basic unit of measurement for capacitance?
5. What is a picofarad? What is a microfarad?
6. What are six types of widely used capacitors?
7. How is an electrolytic capacitor made?
8. What causes electrolytic capacitors to explode?
9. What does working voltage mean?
10. What is polarization and how does it affect electrolytic capacitors?
11. What is capacitive reactance? What is its symbol?
12. How does a capacitor cause a lagging voltage?

This page intentionally left blank

Chapter

10

Single-Phase and Three-Phase Alternating Current

Performance Objectives

- Know the meaning of impedance.
- Know what a vector is and how it works.
- Know what the term kilowatt-hour means.
- Know what is meant by true power.
- Know the difference between apparent power and true power.
- Know what reactive power is.
- Know how most electrical power is generated.

Single-Phase and Three-Phase AC

Alternating current (ac) affects capacitors and inductors differently than direct current. The reactance of inductors and capacitors to ac creates a phase shift between the voltage and current in the circuit. This phase shift causes the kilowatt-hour meter that measures the power consumed in a home, business, or industrial plant to read higher than the power that is actually being used. In as much as compressors for air conditioners and refrigerators are motors, they are also known as inductors and cause a phase shift.

In some cases it is possible by using the right combination of capacitance to lower the power being measured on the kilowatt-hour meter and, of course, lower the cost of operation of the unit. This is why service technicians need to know about power factor and how it affects the equipment they work with.

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

Resistance, Capacitance, and Inductance

The total opposition to the flow of current within a circuit is called *impedance* (Z). It impedes or resists the flow of current. Since it impedes the flow of current in a circuit, impedance is measured in ohms.

In dc circuits, the only opposition is resistance. In ac circuits, opposition to voltage and current comes from a combination of factors. These include resistance (R), capacitance (C), and inductance (L).

The effect of combining these elements (R, L, and C) in a circuit is best examined in steps. Start with resistance. Then add inductance and/or capacitance. When combined in a circuit, a resistor and inductor and/or capacitor produce impedance. Impedance is a combination of effects. It consists of resistance and reactance provided by an inductor and/or a capacitor.

Keep in mind that voltage leads current across an inductor and lags current across a capacitor. This is a phase angle of 90°. The effect on current is of special interest in this situation. The reactance effect is 90° out of phase with the resistive effect. The voltage and current, remember, are in phase in a resistor. To combine these, it is necessary to obtain the vector sum of the two quantities (resistance and capacitive reactance).

Impedance is the vector sum of resistance and reactance. A vector is a line segment used to represent a quantity that has both direction and magnitude. Vectors are used to represent current. A single vector line, or curve, can show two dimensions of current. It can show the direction of flow and it can also show the magnitude or amount of current flowing. A vector sum is a line representing the total of two or more vectors. Impedance is stated in terms of a vector sum (see Fig. 10-1). In this illustration you can see how vectors represent X_L and R. The Z is used to represent the sum of the two acting together to produce the impedance. By measuring the angle made by CAD, you can also find the phase angle produced by the inductor and resistor being in the same circuit. The phase angle is directly related to the power factor. Power factor is important in ac circuits because it tells the ratio of power being consumed to power that appears to be consumed and is actually read by watt-meters. More about power factor will be presented in this chapter as we develop the concepts needed to understand it.

Power is represented by the letter P. It is measured in watts. We also see it represented as kilowatts (kW). The kilo means 1000. Therefore, a kilowatt means 1000 watts. When power is consumed at home or in industry for a period of a month, it is likely to be measured in terms of kilowatts, rather than in watts.

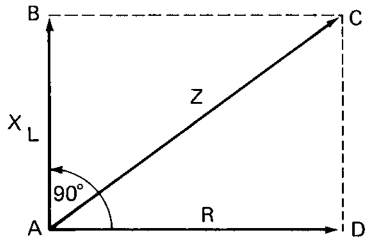

Figure 10-1 Vectors showing impedance.

An ac circuit consumes power differently than a dc circuit or an ac circuit with resistance only. The addition of a capacitor or inductor makes a difference in the actual power consumed. When power is consumed in a dc circuit or an ac circuit with resistance only, the power is found by $P = E \times I$. Or the power is equal to the product of volts and amperes. Once a capacitor or inductor is added to an ac circuit, all this changes because the difference between voltage and current or phase shift has to be taken into consideration to get an accurate reading of what is actually consumed in the way of power. Power that is actually consumed is referred to as true power. True power is that which a dc circuit or ac resistive circuit consumes.

In a purely resistive circuit, energy is dissipated in the resistance as heat, and the resistance does not care in which direction the current is moving at any given instant. As a result, the power cycle looks like the upper waveform in Fig. 10-2. This resistive power cycle is always positive, indicating that power is always being drawn from the circuit. The average power for the cycle is one-half of the maximum value.

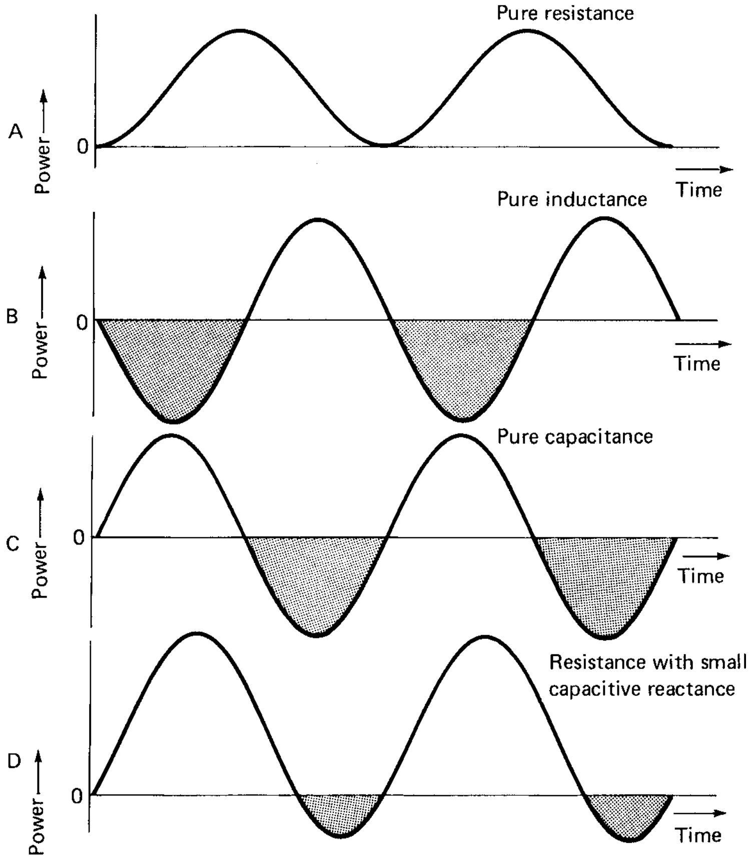

Figure 10-2 Power waveforms.

In a purely inductive or capacitive reactance, the situation is quite different. During parts of the voltage cycle, when the magnetic field is building up around the inductor in either direction or when the capacitor is being charged in either direction, power is being drawn from the circuit. This energy is stored in the magnetic field of the inductor or the electrostatic field of the capacitor. During the remaining parts of the cycle, the magnetic field of the inductor collapses and the capacitor discharges, returning the stored power to the circuit. In these cases, we have a reactive power cycle that is sinusoidal about the zero line, with twice the voltage cycle frequency. The average power consumed for the cycle is zero. The capacitive cycle is 180° out of phase with the inductive cycle, as seen in Fig. 10-2b and c.

The power cycle in any real circuit lies somewhere between the extremes of pure resistive and pure reactive, with more power being drawn from the circuit than is returned to it (see Fig. 10-2d). The circuit is predominantly resistive but slightly capacitive. A slightly inductive power cycle would be similar, except that the negative parts of the cycle would precede the positive parts.

Power Factor

The measure of the relative reactivity of an ac circuit is its power factor. Power factor is defined as the net resistance (*R*) divided by the total impedance (*Z*) of the circuit, or

$$\mathrm{PF} = \frac{\mathrm{R}}{\mathrm{Z}}$$

Power factor is given in both percentage form and in decimal form. That is, 100% is 1.00, or 90% is 0.900.

In a purely resistive circuit, the impedance is made up only of resistance. That means the power factor is unity (100%). In a purely reactive circuit, the impedance comprises only reactance with no resistance, and the power factor is zero (0%). To find the apparent power (AP), you multiply the voltage *E* by the current *I*. This produces what is called *volt-amperes* (VA). Note that this is different from watts. The difference in volt-amperes and watts is significant. The term watt is used to designate a purely resistive circuit or one with only dc resistance. The term volt-ampere designates that the circuit is ac and that there is an inductor or capacitor in the circuit. In other words, it is a reactive circuit when it has volt-amperes. Volt-ampere is also used when referring to the capacity or ability of a transformer to supply a certain amount of volts and amperes. For instance, a door chime transformer may put out 16 volts, but is limited to a maximum of 1 ampere. This means that the volt-ampere rating of the transformer is 16 volt-amperes: $16 \times 1 = 16$. If you want to have two door bells or chimes operate on a transformer, it must be larger, in terms of current available. Thus, if you buy a transformer with the ability to furnish 2 amperes and 16 volts, you have to ask for one that is rated at 32 volt-amperes: $16 \times 2 = 32$.

If you want to know the true power (*W*) instead of the volt-amperes (VA), you have to multiply by the power factor (pf). In other words, if you have 150 volt-amperes and a power factor of 0.5, you will have a true power of 75 watts: $150 \times 0.5 = 75$.

Reactive power averages out to zero. It can be a nuisance in most cases for it puts an instantaneous load on the equipment for part of the cycle, just as if it were a reactive power demand. For this reason, large industrial consumers of power from ac sources will deliberately introduce the opposite kind of reactance into a circuit. That is, they will put a capacitor into an inductive load to reduce the level of reactive power and bring the power factor closer to 100%. This is also the reason why ac generators are always rated in terms of the total apparent power, in kilovolt-amperes (kVA), that they can produce. For loads, on the other hand, it is important to know the two individual components of the apparent power: reactive power in kilovolt-amperes reactive kVAR and the true power in kilowatts (kW).

By adding a capacitor in parallel with an electric motor, the power factor can be improved. This is why some large consumers of electrical power try to reduce the power factor as much as possible. It means they have to pay for the apparent power. Apparent power is always more than true power when there is an inductor or capacitor in the circuit.

Keep in mind, though, that the capacitors used in compressor circuits are not there for power factor correction, but for improving starting torque. The power factor is a secondary consideration. Note that some motors are capacitor start and capacitor run. There are some advantages to having a capacitor in the run circuit. These advantages will be discussed in Chapter 12.

Power factor considerations are not the responsibility of the service technician. Power factor is presented here to give you a better understanding of all aspects of the electrical circuits you will encounter while working on refrigeration and air-conditioning equipment.

Distributing Electric Power

Most electric power is generated as three phase (3ϕ). It is stepped up to 138,000 volts, 345,000 volts, or even to 750,000 volts. The frequency is 60 hertz. Sometimes, however, 25 hertz ac is generated for use by some consumers who may have older equipment that uses this frequency. Motors and other equipment using this power frequency would be very expensive to replace. However, most 25-hertz equipment, when worn out, is replaced with 60-hertz equipment. A separate distribution system is necessary to send 25-hertz power to its destination. This is expensive, but we mention it here so that you will be aware of the fact that it exists in some places where there is a nearby very old generating plant.

Don't make the mistake of connecting 60-hertz equipment to 25-hertz power (Fig. 10-3).

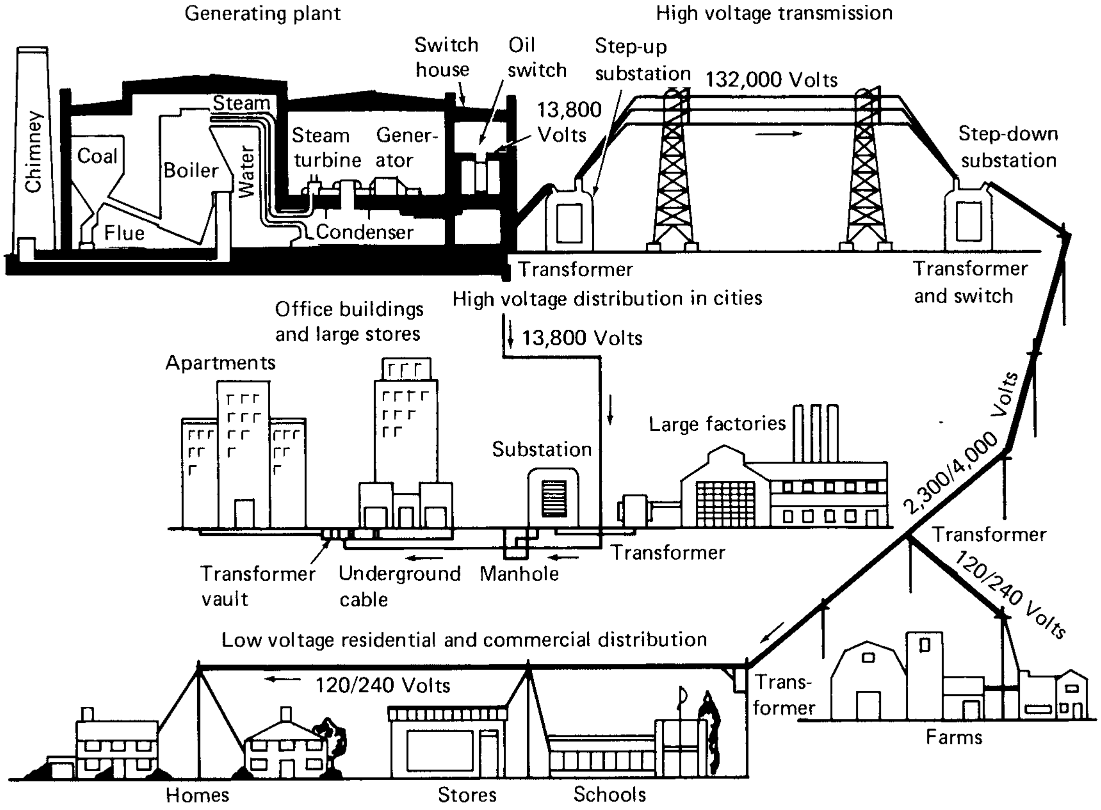
Generating plant
High voltage transmission
Switch house
Oil switch
Step-up substation
132,000 Volts
Chimney
Steam
Coal
Boiler
Water
Steam turbine
Generator
13,800 Volts
Step-down substation
Flue
Condenser
Transformer
Transformer and switch
Office buildings and large stores
High voltage distribution in cities
13,800 Volts
Apartments
Large factories
Substation
2,300/4,000 Volts
Transformer
Transformer
120/240 Volts
Transformer vault
Underground cable
Manhole
Low voltage residential and commercial distribution
120/240 Volts
Transformer
Farms
Homes
Stores
Schools

Figure 10-3 Generation and distribution of electrical power.

Polyphase

The use of more than two wires for a circuit delivering electrical energy offers advantages in economy and in the ways in which the energy can be utilized. As a result, all types of circuits using more than two conductors have been developed and are in common use. These may be either single phase or polyphase (poly means more than one).

In a single-phase system having more than two wires, the voltage between any two pairs of wires will be substantially in phase with the voltage between any other pair of wires. Voltages added in series may be added numerically (taking into account polarity). In the single-phase, three-wire circuit shown in Fig. 10-4, for instance, if the line-to-neutral voltages are each 120 volts, the line-to-line voltage is 240 volts.

In a polyphase system, such voltages are usually not in phase. In a three-phase system (by far the most common polyphase system), three voltages are generated that are equal in magnitude, but 120° apart in phase. Plotted against time,

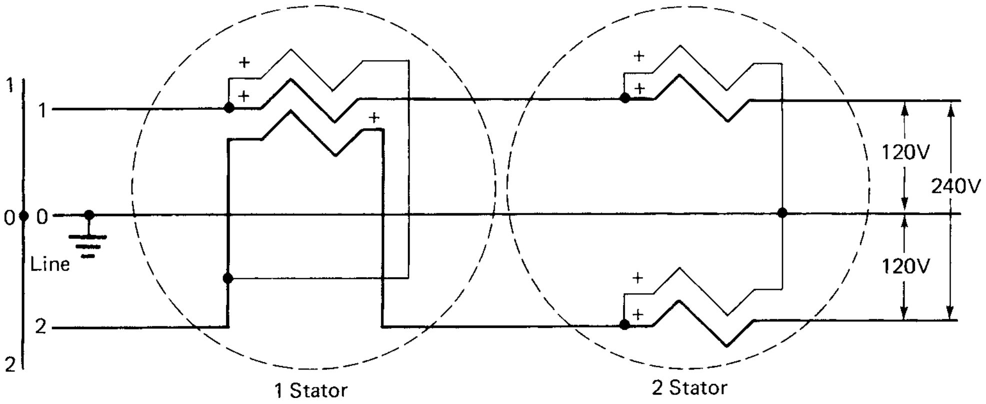

Figure 10-4 Single-phase, three-wire, 240-volt kilowatt-hour meter hookup. The zigzag lines represent the meter coil.

these voltages look like Fig. 10-5. Note that all of the voltages are equal in magnitude, but lag each other by 120°.

Figure 10-6 is a three-wire network being metered with a kilowatt-hour meter. Note how the voltages are distributed. In a conventional three-phase, three-wire system, all line voltages are equal in magnitude, usually at 240 or 480 volts; or for primary metering, at higher distribution voltages, a two-stator meter is required, which is accurate regardless of voltage unbalance. Such a system, usually employed for large commercial or industrial loads, may or may not be grounded. If one line is grounded, the current coils must be in series with the ungrounded lines (see Fig. 10-7).

If a center tap is brought out from one leg of a three-phase delta, the system becomes a three-phase, four-wire delta, such as that frequently used to supply polyphase power at 240 volts for motors and 120 volts, single-phase for lighting. The three-stator kilowatt-hour meter (Fig. 10-8) is rarely used, partly because of the complicated testing procedure. Instead, the two-stator, four-wire delta meter shown in Fig. 10-9 is almost universally used in this service.

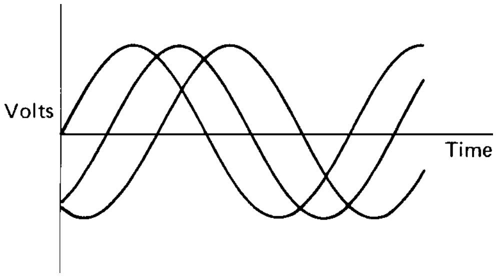

Figure 10-5 Three-phase relationships.

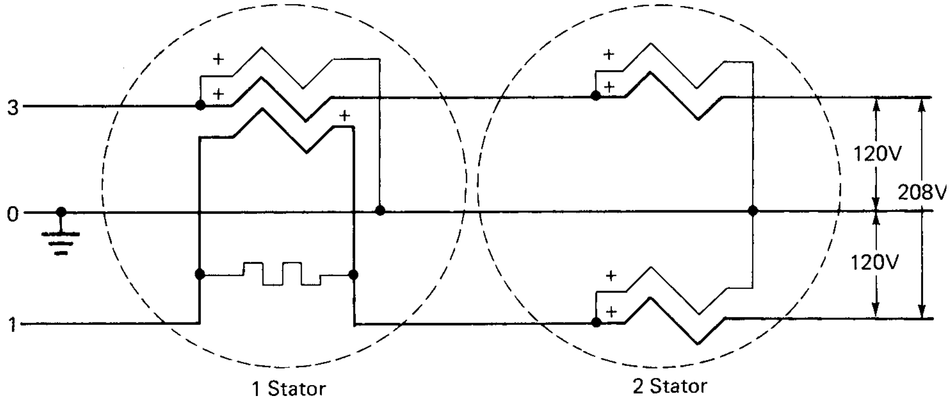
3
0
1
+
+
+
+
+
+
+
120V
208V
120V
1 Stator
2 Stator

Figure 10-6 Three-phase network, 208 volts.

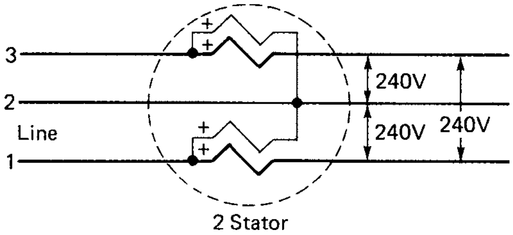
3
2
Line
1
+
+
+
+
240V
240V
240V
2 Stator

Figure 10-7 Three-phase, three-wire, meter hookup (delta).

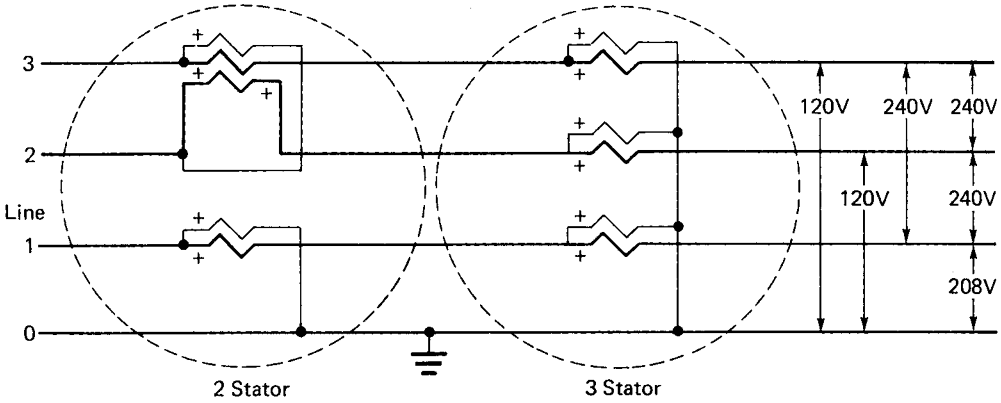
3
2
Line
1
0
+
+
+
+
+
+
+
+
+
+
+
120V
240V
240V
120V
240V
208V
2 Stator
3 Stator

Figure 10-8 Three-stator meter (rarely used).

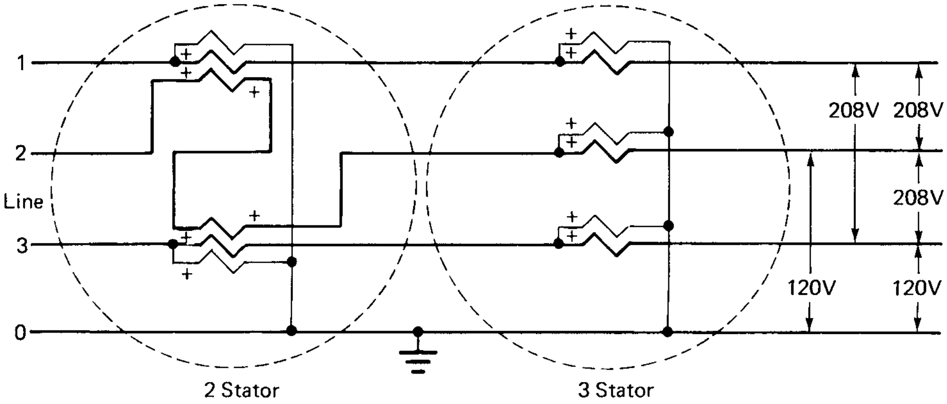

Figure 10-9 Four-wire wye system being metered.

Although a four-wire delta system is normally used only for 240/120-volt service, the three-phase (3ϕ), four-wire system covers a wide range of voltages from 208/120-volt service up to the highest transmission voltages (see Fig. 10-9). The neutral is usually grounded. It is measured with either a three- or two-stator meter, depending on the importance of the load and the degree of accuracy required. Note the distribution voltages between any two of the conductors.

These meter hookups are shown to aid in identifying the type of power supplied to various equipment.

Circuit Breakers

A circuit breaker is a device that, after breaking a circuit, can then be reset by turning the handle to past the off position and then to on (see Fig. 10-10). The circuit breaker is mounted in a distribution box by snapping it into place. The hot wire (black or red) is attached by inserting it under the screw and tightening. A knockout blank in the distribution box must be removed to allow for the handle and top of the circuit breaker to be exposed (see Fig. 10-11).

A number of types of circuit breakers are used in home and industrial applications. However, most of them snap in place much as shown in Fig. 10-12. This illustration shows the snap in action without the sides or top of the box in place. Keep in mind that the circuit breakers are snapped into place in the panel board. The wires are connected before the circuit breakers are snapped into place. Make sure the circuit breaker is in the off position while installing or reinstalling if it becomes necessary. Note there is a difference in the three types of circuit breakers shown here. Types A and C must have a notch so the formed rib fits properly. Type B can be used only where there is no formed rib in the mounting plate.

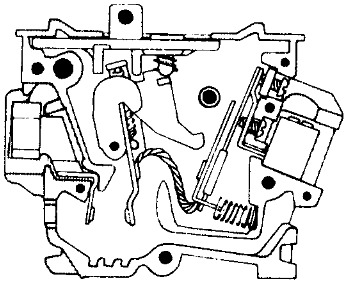

Tripped position
- Contacts open — no current flow
- Handle stationary when tripped

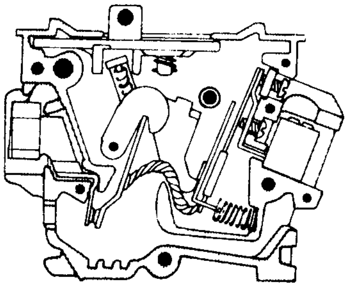

"On" position
- Contacts closed — current on
- Handle in "On" position (shows "On")

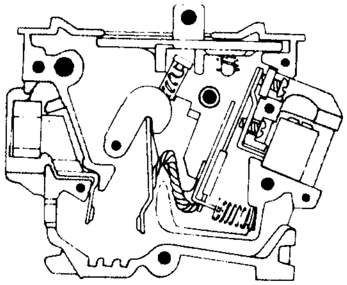

"Off" position
- Contacts open — no current flow
- Handle in "Off" position (shows "Off")

To restore service when fault is cleared you simply move operating handle to "Off" position and then to "On".

Figure 10-10 Cutaway view of circuit breakers.

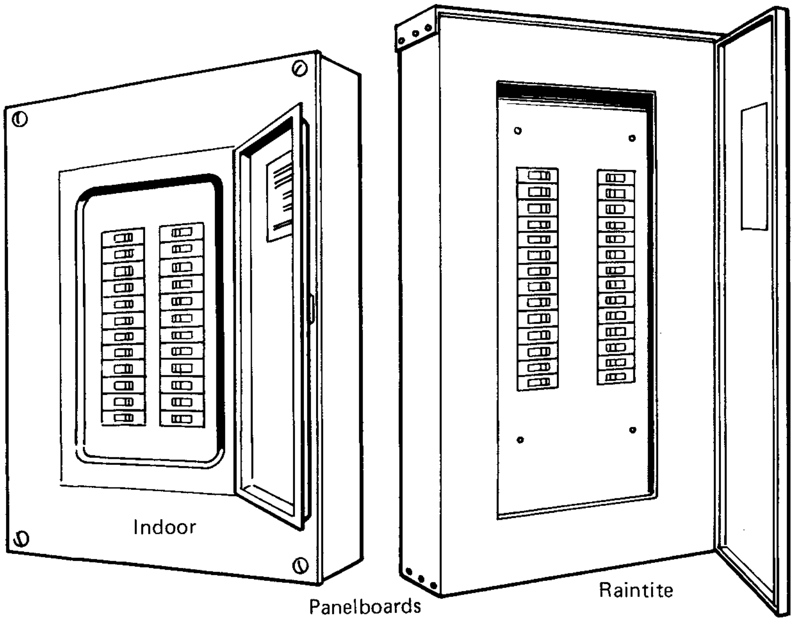

Figure 10-11 Circuit breaker box with breakers in place.

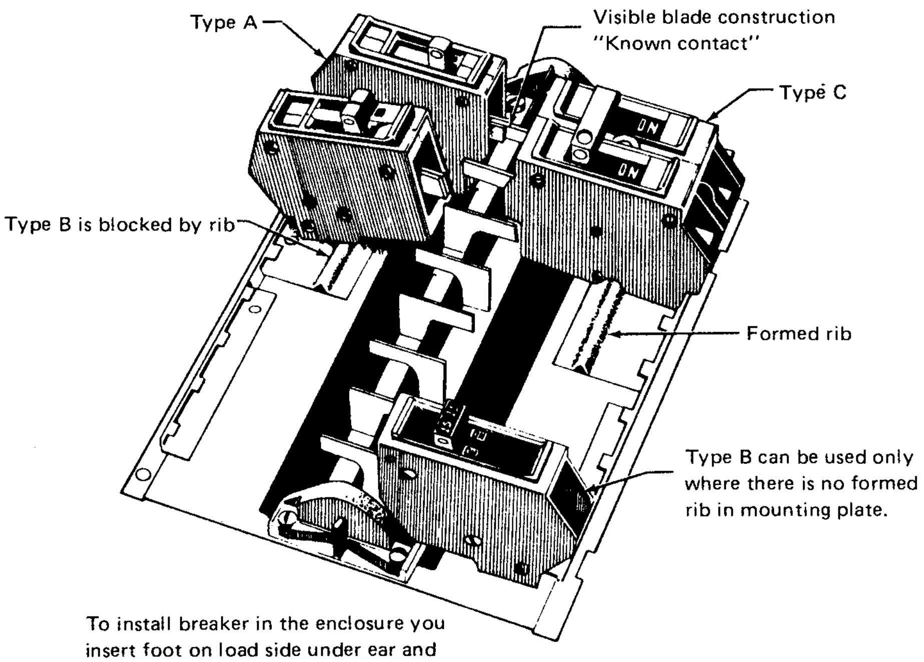
Type A
Visible blade construction
"Known contact"
Type C
Type B is blocked by rib
Formed rib
Type B can be used only where there is no formed rib in mounting plate.

To install breaker in the enclosure you insert foot on load side under ear and press the jaw over blade.

Type A, C and D breakers will clear the rib and can be used at any position on the mounting plate.

Figure 10-12 Circuit breakers are snapped into place in the panel board.

Review Questions

1. What is impedance?
2. What is meant by vector sum?
3. What is meant by the term kilowatt-hour?
4. What is meant by the term true power?
5. How do you measure the relative reactivity of an ac circuit?
6. What is the difference between apparent power and true power?
7. What is resistive power?
8. How is most electrical power generated?
9. What is the purpose of a circuit breaker in a circuit?
10. How are the circuit breakers mounted in the panel board?
11. What is reactive power?
12. What is the term used to measure ac power?
13. What is the term used to express resistive power consumption?
14. What does out-of-phase mean?
15. What's the difference between ac and dc power?

This page intentionally left blank

Chapter
11
Solid-State Controls

Performance Objectives

- Know the meaning of the word semiconductor.
- Know the two materials used for semiconductors.
- Know how diodes are made and used.
- Know how the SCR is used in air-conditioning and refrigeration circuits.
- Know how an integrated circuit is produced and used.
- Know how a thermistor works.
- Know how a humidity element works.
- Know how a bridge circuit is used in control circuits.
- Know how a sensor operates.
- Know what an actuator does.
- Know what a differential amplifier is used for.

Semiconductors

The word semiconductor identifies a type of electronic device. Transistors and diodes are semiconductors. These units are part of the field of electronics. They have been used over the past years in every aspect of control circuits on every type of equipment. Electronic controls for air-conditioning, refrigeration, and heating equipment have been upgraded to use transistors, diodes and computer chips as well, and other types of semiconductor devices (see Fig. 11-1).

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

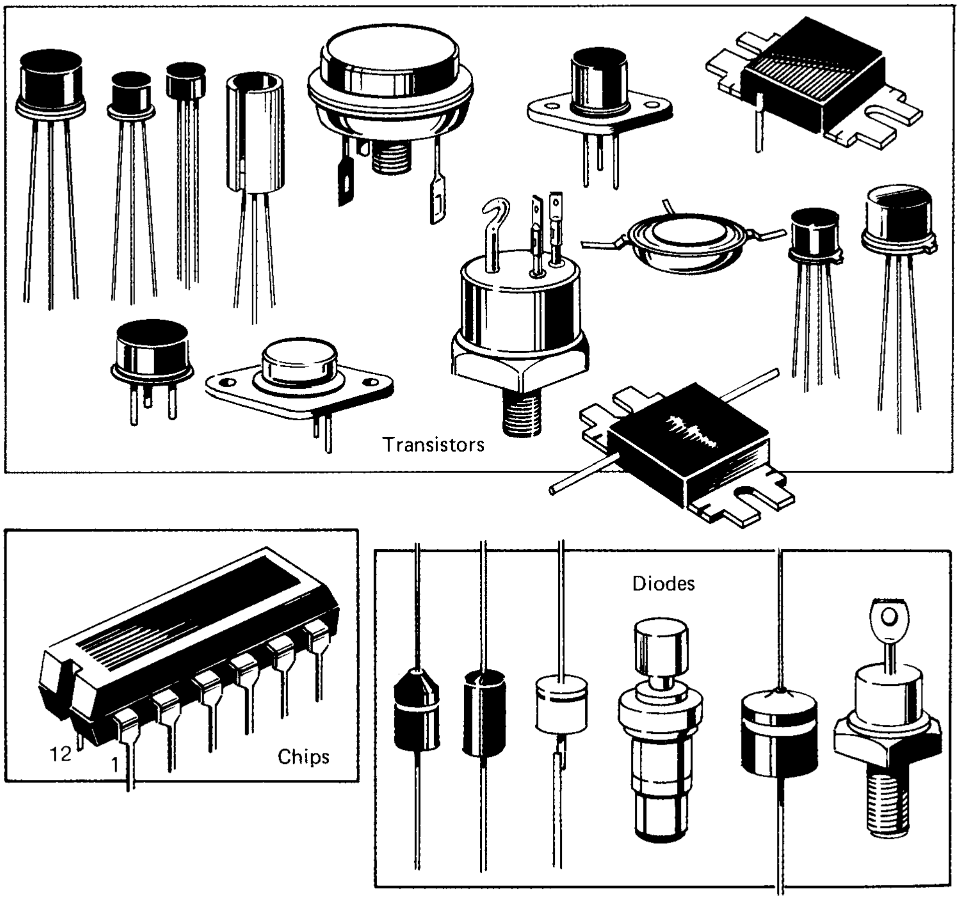

Figure 11-1 Transistors, diodes, and chips.

Semiconductor Principles

Semiconductor technology is usually called *solid state*. This means that the materials used are of one piece. This is in contrast with the vacuum tube, which consisted of a series of assembled parts. Of course, the substances from which semiconductors are made are not really solid. They have atomic structures consisting largely of empty space. The spaces are essential for the movement of electrons.

Many materials can be classified as semiconductors, but two have been used extensively for electronic circuits. They are silicon and germanium. Germanium was used originally for diodes and transistors, but has since been largely replaced by silicon.

Both silicon and germanium are hard, crystal-type materials that are very brittle. They can be made pure to 99.999999999%. Doping agents or impurities

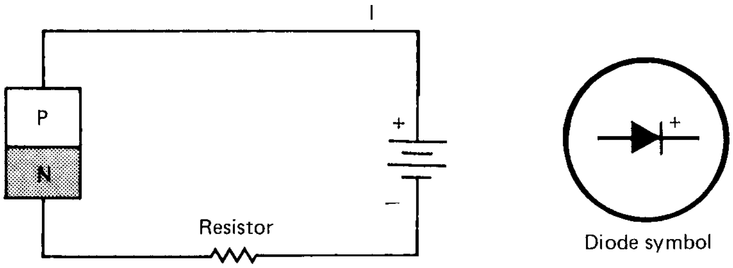

Figure 11-2 PN diode in a circuit. Note the symbol for a diode.

are added in controlled amounts. Impurities used may be boron, aluminum, indium, gallium, arsenic, and antimony.

By doping silicon, it is possible to allow the electrons added by doping (in N-type material) to move easily. In P-type material, the holes or places where an electron should be are moved along by the voltage applied. The addition of arsenic or antimony makes an N-type semiconductor material. The material will have an excess of electrons. Electrons have a negative charge.

When gallium or indium is used as the impurity, a P-type semiconductor material is produced. This means it has a positive charge, or is missing an electron.

Diode

The semiconductor diode is used to allow current to flow in only one direction. It can be used to change alternating current to direct current. It is made by using N-type and P-type semiconductor materials.

When N-type and P-type materials are joined, they form a diode (see Fig. 11-2). The diode is also called a rectifier since it can change (rectify) ac and make it dc. The PN junction (diode) acts as a one-way valve to control the current flow. The forward or low-resistance direction through the junction allows current to flow through it. The high-resistance direction does not allow current to flow. Thus, only one-half of an alternating current hertz is allowed to flow in a circuit with a diode (see Fig. 11-2). The diode used in this circuit is forward biased and allows current to flow. Figure 11-3 indicates the arrangement in the reverse-biased configuration. No current is allowed to flow under these conditions. Also, note the polarity of the battery.

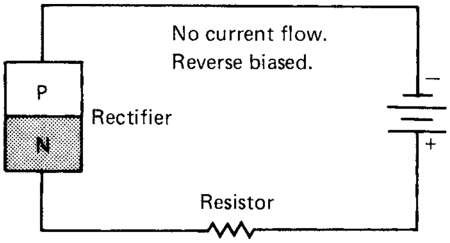

Figure 11-3 Reverse-biased diode in a circuit. Note how this differs from Fig. 11-2.

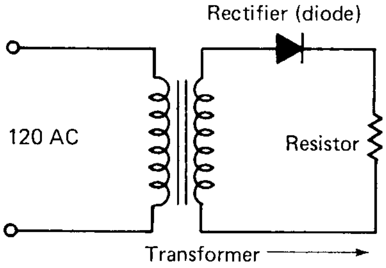

Figure 11-4 Circuit showing how a diode is used to rectify ac to produce dc.

Diodes are used in isolating one circuit from another. A simple rectifier circuit is shown in Fig. 11-4. The output from the transformer is an ac voltage, as shown in Fig. 11-5. However, the rectifier action of the diode blocks current flow in one-half of the sine wave and produces a pulsating dc across the resistor (see Fig. 11-5b).

A number of factors affect the voltage and current ratings of a PN junction diode. As you work with them, you will learn that they have ratings for peak and inverse ac currents. This information is available for both germanium and silicon diodes in transistor handbooks produced by their manufacturers.

Figure 11-6 shows a number of specialized diodes. This illustration demonstrates that diodes are designed for specific jobs. Both internal and external construction is determined by circuit requirements. For example, Fig. 11-6a shows diodes used for radar and computer circuits. The diodes in Fig. 11-6b are designed to carry large currents. So their cases are heat conductors. As the diodes become warm, the generated heat is transferred to the air. Figure 11-6c shows zener diodes. These units protect sensitive meter movements and regulate voltage.

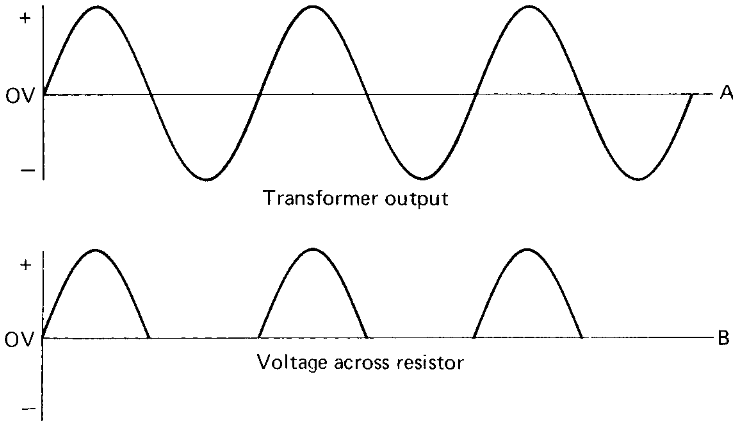

Figure 11-5 Results of the rectifier circuit. Transformer output (ac) is changed to dc.

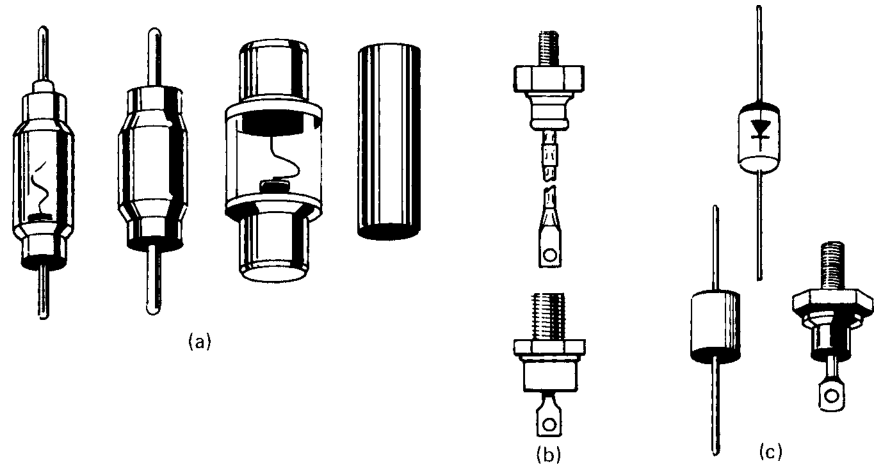

Figure 11-6 Diodes for special circuits. (A) Radar and computer circuits. (B) Large-current circuits. (C) Low-current circuits.

Silicon-Controlled Rectifiers

The silicon-controlled rectifier (SCR) is a specialized type of semiconductor used for control of electrical circuits. This is a four-layer device. The structure can be either NPNP or PNPN.

An SCR conducts current in a forward direction only. The symbol for an SCR is shown in Fig. 11-7. Current always flows through an SCR from the cathode (C) to the anode (A). The illustration indicates that the SCR also has a gate (G).

The function of an SCR is shown in the circuit diagram in Fig. 11-8. The typical use of an SCR is for a controlled circuit. Examples include a light dimmer or a speed control for a motor. This type of circuit is shown in Fig. 11-8. The resistor in the circuit, R_1, is a rheostat, or adjustable resistor: This is used to control the amount of voltage delivered to the gate of the SCR.

The more voltage delivered, the greater the flow. Thus, adjusting the rheostat will control the circuit. If the circuit illuminates a lamp, lowering the voltage to the rheostat dims the bulb. If the load is a motor, its speed is lowered. Figure 11-9 shows typical SCRs.

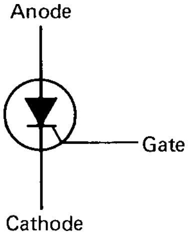

Figure 11-7 Symbol used for representing an SCR in a circuit.

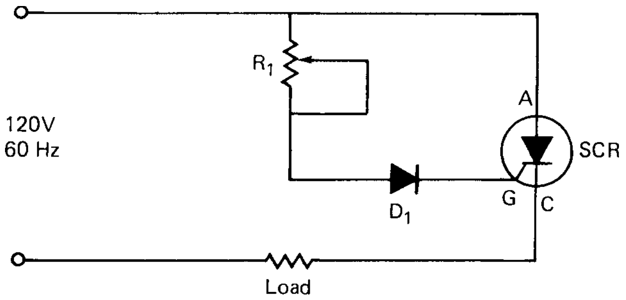

Figure 11-8 Schematic of an SCR-controlled circuit.

Transistors

Transistors are made from N- and P-type crystals. Once joined, the two different types of crystals produce junctions. Transistors are identified according to emitter junction and collector junction. Thus, they are either PNP or NPN types.

A PNP transistor is formed by a thin N region between two P regions (see Fig. 11-10). The center N region is called the base. This base is usually 0.001 inch thick. A collector junction and an emitter junction are also formed. Note the emitter is represented by an arrow either going toward the vertical line or away from it. The collector is the line without an arrow. The third line, the one that has the other two contacting it, is the base. These are usually abbreviated C, B, and E.

Transistors can be used for amplification of signals or they can be used for switching. There are three common configurations for transistor circuits. They are the common base, common collector, and common emitter. The common-emitter circuit is the most often used (see Fig. 11-11). In this type of circuit, the current through the load flows between the emitter and collector. The input signal is applied between the emitter and the base. In normal operation, the collector junction is reverse biased by the supply voltage, B_1. The emitter junction is forward biased by the applied voltage, B_2. Electrons flow across the forward-biased emitter into the base. They diffuse through the base region and flow across the collector junction. Then they flow through the external collector circuit.

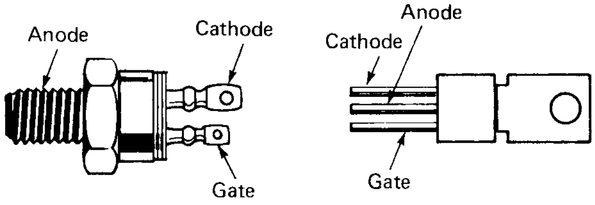

Figure 11-9 Two types of SCRs. Both types are used for control circuits, depending on the amount of current involved.

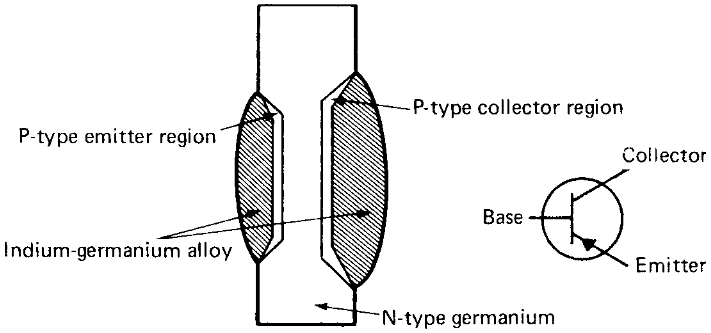

Figure 11-10 Transistor junction using germanium and indium.

Battery B_2 voltage is applied in the forward direction. This means the voltage is positive to the emitter P-type crystal. It also means that voltage is negative to the N-type crystal. Thus, the emitter-base junction has low impedance.

The voltage of battery B_1 is applied in the reverse direction. This means the voltage is positive to the N-type crystal. It also means that voltage is negative to the P-type crystal. This, then, produces a collector-base junction with high impedance.

Transistor impedances

The impedance of the emitter junction is low. Thus, electrons flow from the emitter region to the base region. At the junction, the electrons combine with the holes in the N-type base crystal. If the base is thin enough, almost all the holes are attracted to the negative terminal of the collector. They then flow through the load to B_1.

The collector current is stopped by applying a positive voltage to the base and a negative voltage to the emitter. In actual transistors, however, this cannot be done because of several basic limitations. Some of the electrons in the base region flow across the emitter junction. Some combine with the holes in the base region. For this reason, it is necessary to supply a current to the base. This makes up for these losses.

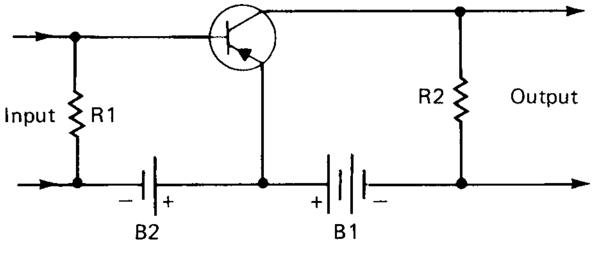

Figure 11-11 Schematic diagram of a common emitter circuit using a PNP transistor.

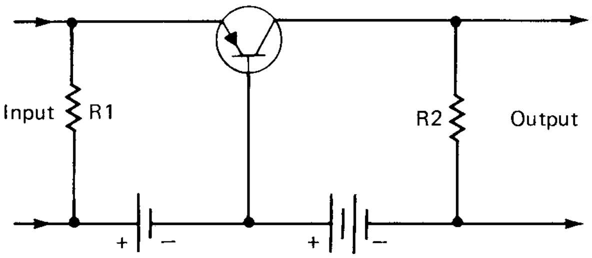

Figure 11-12 Common base circuit using a PNP transistor.

The ratio of the collector current to the base current is known as the current gain of the transistor. Current gain, called beta (β), is found by dividing base current into collector current. At high frequencies, the fundamental limitation is the time for carriers to diffuse across the base region. They move from the emitter to the collector. This is why the base region width or thickness is so important. The thinner the base region, the less time is required for the carriers to diffuse across it. This causes the transistor to operate faster.

Figure 11-12 shows a schematic for a common-base PNP circuit. The signal is injected into the emitter-base circuit. The output signal is taken from the collector-base circuit. An important advantage of the transistor is its ability to transfer impedances. This is where the transistor gets its name. The word transistor comes from *transfer resistor*.

The emitter circuit has low impedance. This low impedance allows current to flow. This current flow then creates a current through the collector circuit. The emitter has low impedance and low current. The collector has high impedance and even slightly less current than the emitter. However, more power is the result in the collector. This is because $P = I^2R$ or, in this case, $P = I^2Z$. Impedance (Z) is high and the current is squared. Thus, the collector circuit has more power than the emitter circuit with its low impedance.

A common-emitter circuit has about 1300 ohms of input impedance. This is compared to an output impedance of about 50,000 ohms. Thus, there is an increase in impedance from emitter to collector of about 39 times. The junction transistor amplifies in this way. It acts as a power amplifier. A small change in emitter voltage causes a large change in the collector circuit. The different impedances cause this reaction.

The power transistor is shown in Fig. 11-13. Along with the power transistor is the much smaller signal amplifying and switching transistor normally seen in electronics equipment.

Integrated Circuits

The semiconductor monolithic chip was developed in 1958 by J.S. Kilby. Active and passive circuit components were successively diffused and deposited on a single chip. Shortly after, Robert Noyce made a complete circuit on a single

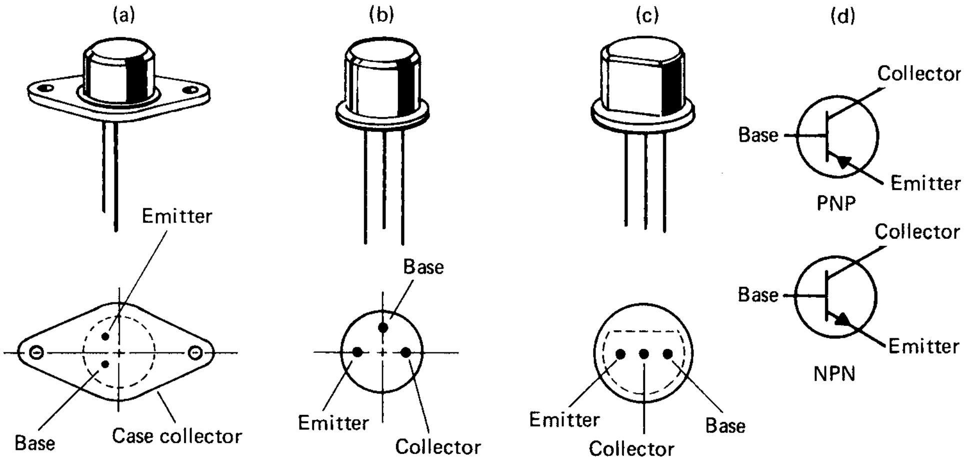

Figure 11-13 Three types of transistor packages. (a) Power transistors. (b) TO-5 case. (c) Small-signal transistors. (d) Transistor symbols.

chip. This was the beginning of the modern, inexpensive integrated circuit (IC or chip). Resistors, capacitors, transistors, and diodes can be placed on a chip. These chips are available in three standard packages, as follows:

Flat pack. The flat pack is hermetically sealed. This means it is vacuum packed. The ceramic flat pack has either 10 or 14 pins. This type of packaging is no longer used. It may be seen in some older-model equipment with chips (see Fig. 11-14).

Multipin circular. This type of packaging is no longer in common use. It was originally used because it fit into the same type of container as a transistor. It had more than three leads coming from a TO-5 case originally made for transistors (see Fig. 11-15).

Dual in-line package (DIP). This is the most commonly used type of integrated circuit today. When you think of a "chip," you think of this type. It may have many leads from each side (see Fig. 11-16). This is an easily used type of IC since it can be placed into a socket and removed if it develops any problem. This type of package is standardized as to size. It can be placed into circuit boards by machines, which increases its usefulness in electronics equipment. Many thousands of ICs can be made at a time. This means the cost of manufacturing

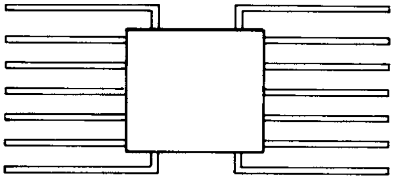

Figure 11-14 Flat pack IC.

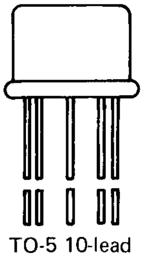

Figure 11-15 TO-5 case used for IC.

is very low. DIPs are used in computers, calculators, and control devices for air-conditioning and heating equipment. Amplifiers are fabricated as complete units. Everything seems to use ICs today. Toys, calculators, computers, and automobiles all use ICs or chips in one form or another. The possibilities are unlimited. ICs will play a more important role in air-conditioning, refrigeration, and heating devices in the future.

Solid-state demand defrost control

An excellent example of a refrigeration control using semiconductor technology is the demand defrost control. The idea of defrosting only when needed saves money and energy and protects the quality of the refrigerated product, since extremes in temperature swings are minimized.

Thermistor Sensing

A temperature difference concept can be used to automatically initiate the defrost cycle on vertical open frozen food display cases that use electric defrost (see Fig. 11-17). If the cover is lifted, you can see the electronic package. There are a few transistors and the usual capacitors, resistors, and a transformer. The sensor that feeds in the information needed to operate properly as a control is shown in Fig. 11-18. These two sensors are simple in construction. They are simple thermistors that change resistance when temperature changes (see Fig. 11-19). Their operation is not affected by dirt, moisture, lint, food particles, or ice. They can sense the least temperature change precisely.

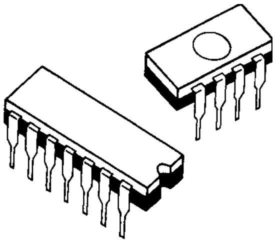

Figure 11-16 Integrated circuits in dual in-line packages (DIPs).

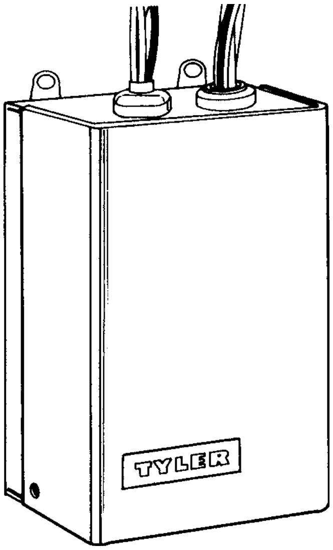

Figure 11-17 DF-1 solid state demand defrost control for vertical open frozen food display cases. (Courtesy of Tyler)

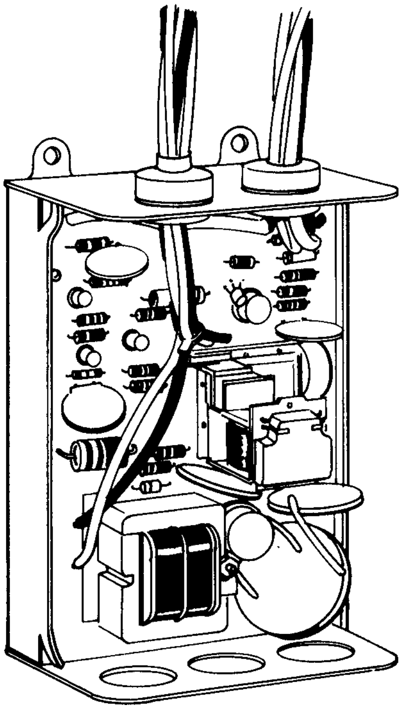

Figure 11-18 Cover removed from the DF-1 defrost control. (Courtesy of Tyler)

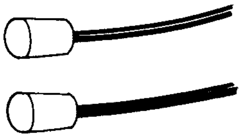

Figure 11-19 Thermistors used to sense temperature differences. (Courtesy of Tyler)

Figure 11-20 shows how frost buildup is sensed to cause the circuit to know what is happening. This tells when defrost should be started and terminated. The ever widening temperature difference is a sure sign of frost buildup. The temperature difference of air flowing over a coil increases in direct proportion to frost buildup. This is what the thermistor monitors to trigger defrost only when needed. Refrigeration fixtures run at peak efficiency all the time. They use less energy and keep the product at a lower, steadier temperature.

Humidity Sensing

Figure 11-21 shows how defrost frequency and humidity are related. Improvements in the design of humidity-sensing elements and the materials used in their construction have minimized many past limitations of humidity sensors. One type of humidity sensor used with electronic controls is a resistance CAB (cellulose acetate butyrate) element (see Fig. 11-22). This resistance element is an improvement over other resistance elements. It has greater contamination resistance, stability, and durability. The humidity CAB element is a multilayered, humidity-sensitive, polymeric film. It consists of an electrically conductive core and insulating outer layers. These layers are partially

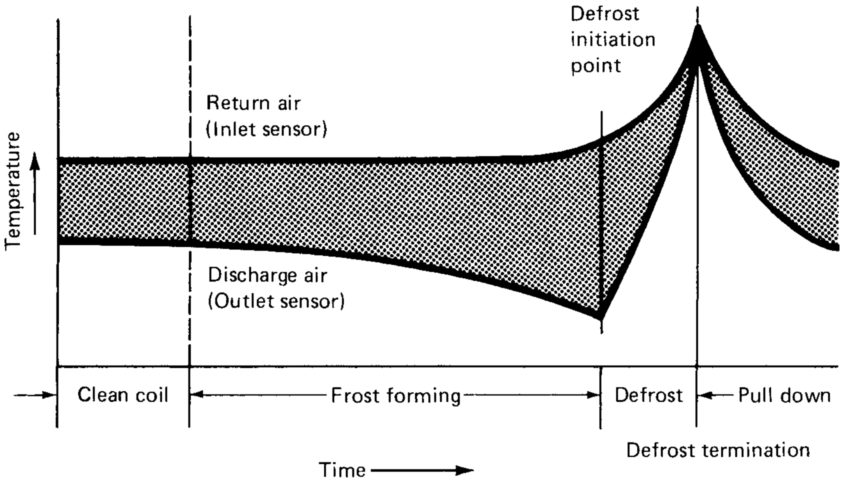

Figure 11-20 Temperature-difference chart. (Courtesy of Tyler)

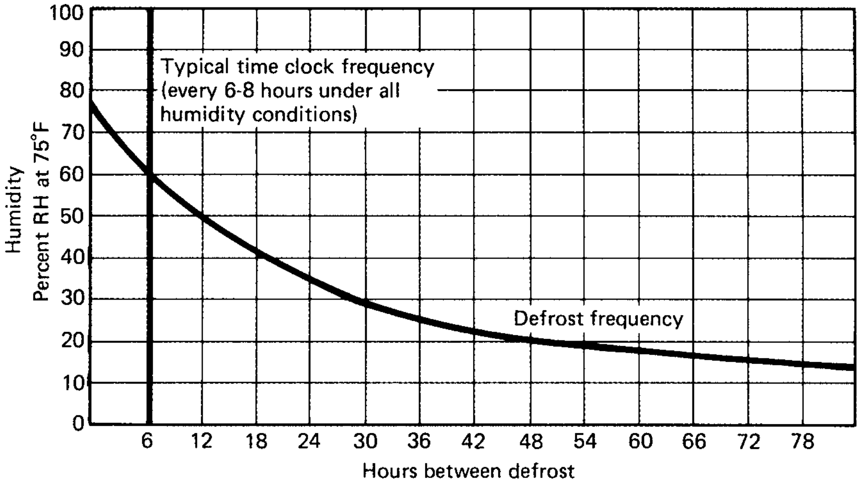

Figure 11-21 Effect of humidity on hours between defrosting. (Courtesy of Tyler)

hydrolyzed. The element has a nominal resistance of 2500 ohms. It has a sensitivity of 2 ohms per 1% relative humidity (rh) at 50% rh. Its humidity sensing range is rated at 0% to 100% rh.

The CAB element consists of conductive humidity-sensitive film, mounting components, and a protective cover (see Fig. 11-23). The principal component of this humidity sensor is the film. The film has five layers of CAB in the form of a ribbon strip. The CAB material is used because of its good chemical and mechanical stability and high sensitivity to humidity. It also has excellent film-forming characteristics.

The CAB resistance element is a carbon element having a resistance/humidity tolerance favorable to inclusion in a control circuit. With an increase in relative humidity, water is absorbed by the CAB, causing it to swell. This swelling of the polymer matrix causes the suspended carbon particles to move farther apart from each other. This results in an increased element resistance.

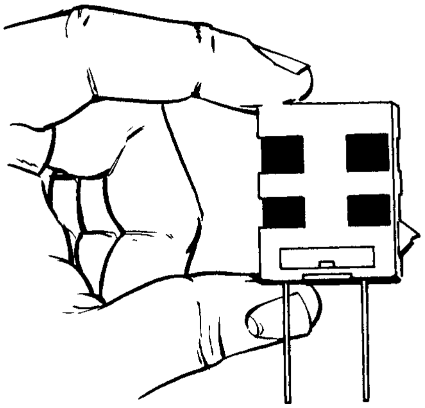

Figure 11-22 CAB resistive element.

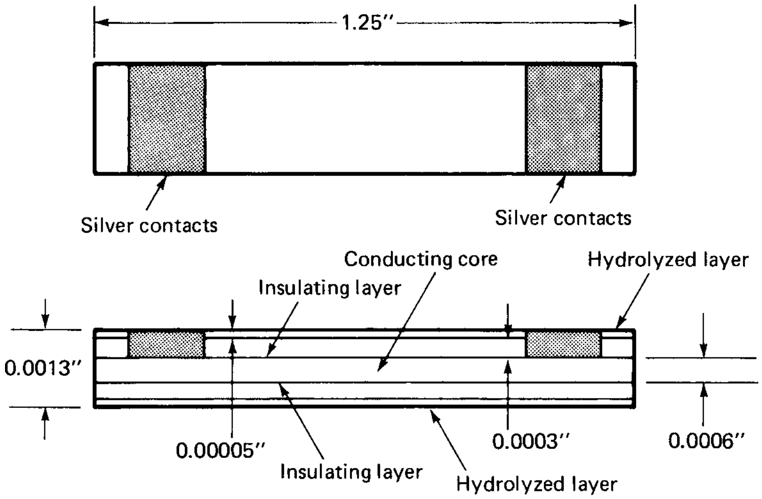
1.25"
Silver contacts
Silver contacts
Conducting core
Hydrolyzed layer
Insulating layer
0.0013"
0.00005"
Insulating layer
0.0003"
Hydrolyzed layer
0.0006"

Figure 11-23 Hydrolyzed humidity element.

When relative humidity decreases, water is given up by the CAB. The contraction of the polymer causes the carbon particles to come closer together. This, in turn, makes the element more conductive or less resistive.

Bridge Circuit

A bridge circuit is a network of resistances and capacitive or inductive impedances. The bridge circuit is usually used to make precise measurements. The most common bridge circuit is the Wheatstone bridge. This consists of variable and fixed resistances. Simply, it is a series-parallel circuit (see Fig. 11-24). The branches of the circuit forming the diamond shape are called legs.

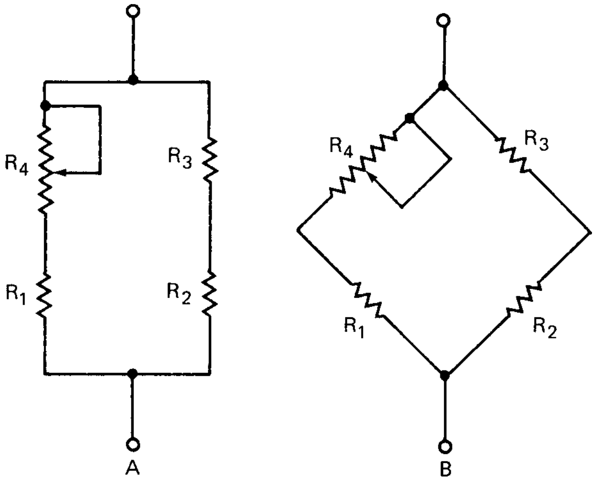
R4
R3
R1
R2
A
R4
R3
R1
R2
B

Figure 11-24 Bridge circuit configurations.

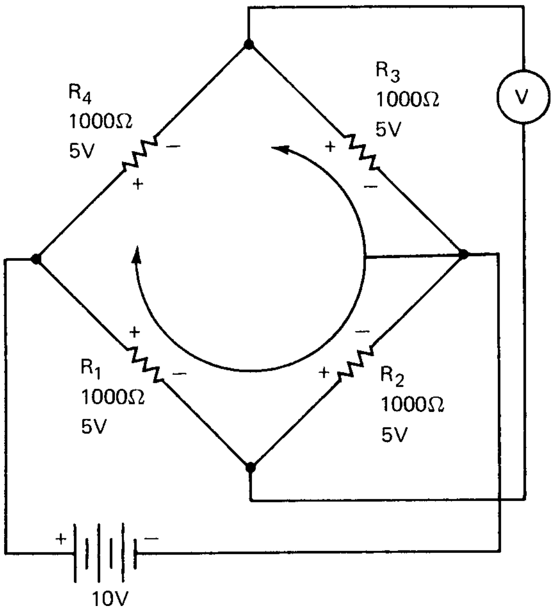

Figure 11-25 Current in a balanced bridge circuit.

If 10 volts dc is applied to the bridge circuit shown in Fig. 11-25, one current will flow through R_1 and R_2, and another through R_3 and R_4. Since R_1 and R_2 are both fixed 1000-ohm resistors, the current through them is constant. Each resistor will drop one-half of the battery voltage, or 5 volts. Thus, 5 volts is dropped across each resistor. The meter senses the sum of the voltage drops across R_2 and R_3. Both are 5 volts. However, the R_2 voltage drop is a positive (+) to negative (−) drop. The R_3 drop is a negative to positive drop. They are opposite in polarity and cancel each other. This is called a *balanced bridge*. The actual resistance values are not important. What is important is that this ratio be maintained and the bridge be balanced.

Unbalanced bridge

In Fig. 11-26, the value of the variable resistor R_4 is 950 ohms. The other resistors have the same value. Using Ohm's law, the voltage drop across R_4 is found to be 4.9 volts. The remaining voltage, 5.1 volts, is dropped across R_3 (see Fig. 11-26). The voltmeter measures the sum of the voltage drops across R_2 and R_3 as 5.0 volts (+ to −) and 5.1 volts (− to +). It registers a total of −0.1 volt.

In Fig. 11-27, the converse is true. The value of R_4 is 1050 ohms. The voltage drop across R_3 is 4.9 volts. The voltmeter senses the sum of 5 volts (+ to −) and 4.9 volts (− to +), or +0.1 volt.

When R_4 changes the same amount above or below the balanced-bridge resistance, the magnitude of the dc output measured by the voltmeter is the same. However, the polarity is reversed.

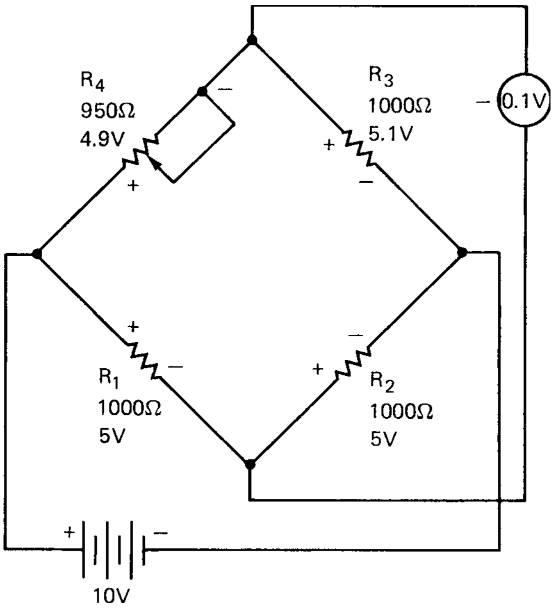
R4
950Ω
4.9V
R3
1000Ω
5.1V
0.1V
R1
1000Ω
5V
R2
1000Ω
5V
10V
+
–

Figure 11-26 Operation of a bridge circuit.

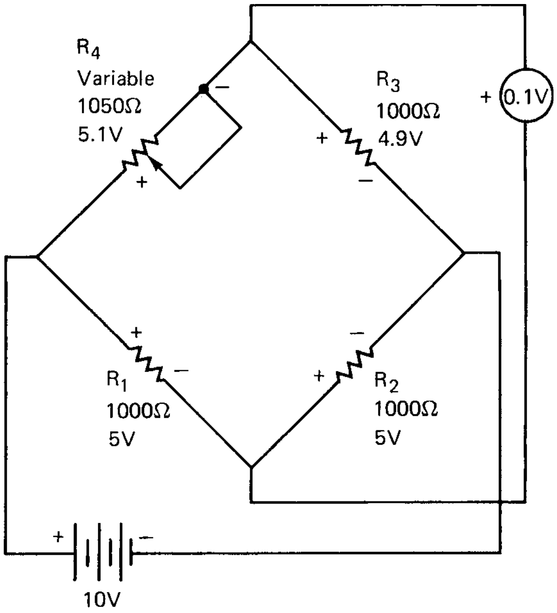
R4
Variable
1050Ω
5.1V
R3
1000Ω
4.9V
0.1V
R1
1000Ω
5V
R2
1000Ω
5V
10V
+
–

Figure 11-27 Operation of a bridge circuit.

Sensors

The *sensor* in a control system is a resistance element that varies in resistance value with charges in the variable it is measuring. This may be humidity or temperature. These resistance changes are converted into proportional amounts of voltage by a bridge circuit. The voltage is amplified and used to position actuators that regulate the controlled variable.

Controllers

The sensing bridge is the section of the controller circuit that contains the temperature-sensitive element or elements. The potentiometer for establishing the set point is also part of the control system. The bridges are energized with a dc voltage. This permits long wire runs in sensing circuits without the need for compensating wires or for other capacitive compensating arrangements.

Both integral (room) and remote-sensing element controllers produce a proportional 0- to 16-volt dc output signal in response to a measured temperature change. Controllers can be wired to provide direct or reverse action. Direct-acting operation provides an increasing output signal in response to an increase in temperature. Reverse-acting operation provides an increasing output signal in response to a decrease in temperature.

Electronic Controllers

Electronic controllers have three basic elements: the bridge, the amplifier, and the output circuit. Two legs of the bridge are variable resistances in Fig. 11-28. The sensor and the set-point potentiometer are shown in the bridge configuration. If temperature changes or if the set point is changed, the bridge is in an unbalanced state. This gives a corresponding output result. The output signal, however, lacks power to position actuators. Therefore, this signal has to be amplified to become useful in the control of devices associated with making sure the right amount of heat of cooled air gets to the room intended.

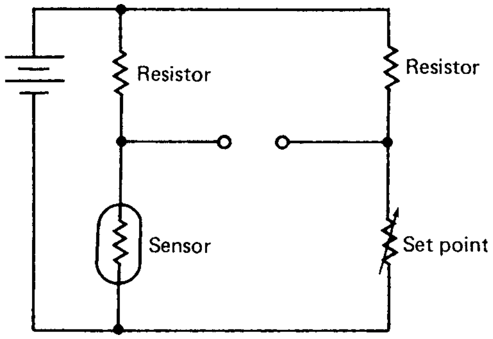

Figure 11-28 Bridge circuit with a sensor and a set point.

Differential Amplifiers

Controllers utilize direct-coupled dc differential amplifiers to increase the millivolt signal from the bridge to the necessary 0- to 16-volt level for the actuators. There are two amplifiers, one for direct reading and the other for reversing signals. Each amplifier has two stages of amplification. This arrangement is shown in block form in Fig. 11-29.

The differential transistor circuit provides gain and good temperature stability. Figure 11-30 compares a single transistor amplifier stage with a differential amplifier. Transistors are temperature sensitive. That is, the currents they allow to pass depend on the voltage at the transistor and its ambient temperature. An increase in the ambient temperature in the circuit shown in Fig. 11-30a causes the current through the transistor to increase. The output voltage decreases. The emitter resistor, R_E, reduces this temperature effect. It also reduces the available voltage gain in the circuit because the signal voltage across the resistor amounts to a negative feedback voltage. That is, it causes a decrease in the voltage difference that was originally produced by the change in temperature at the sensing element.

Since it is desirable for the output voltage of the controller to correspond only to the temperature of the sensing elements and not to the ambient temperature of the amplifier, the circuit shown in Fig. 11-30b is used. Here any ambient temperature changes affect both transistors at the same time. The useful output is taken as the difference in output levels of each transistor, and the effects of temperature changes are canceled. The voltage gain of the circuit shown in Fig. 11-30b is much higher than that shown in Fig. 11-30a. This is

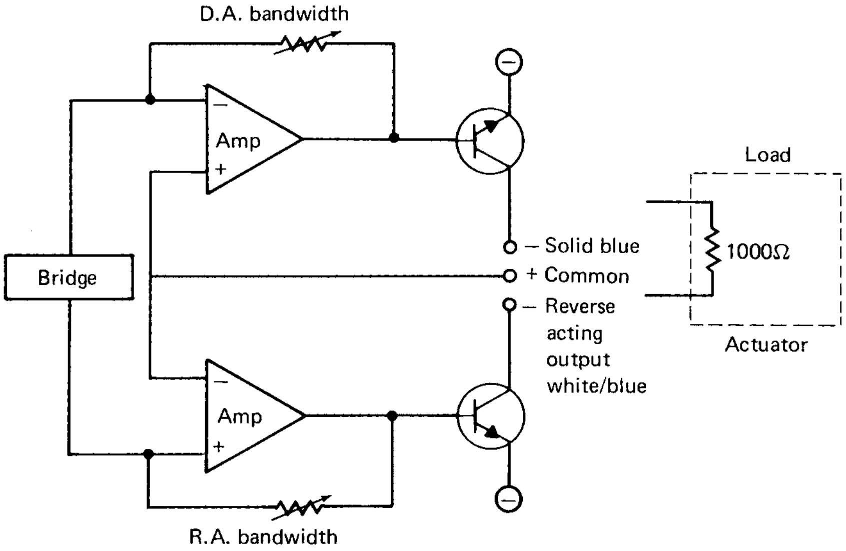

Figure 11-29 DC differential amplifier for use in a controller circuit.

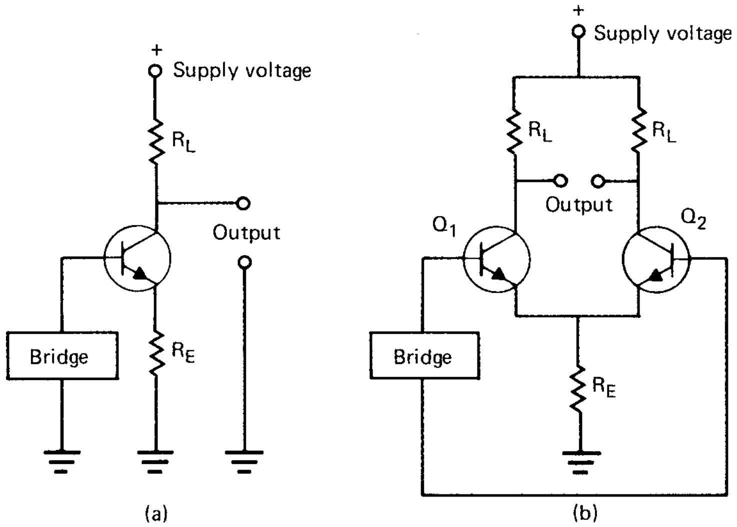

Figure 11-30 (a) Single transistor amplifier stage. (b) Two-transistor amplifier stage.

because the current variations in the two transistors produced by the bridge signal are equal and opposite. An increase in current through Q_1 is accompanied by a decrease in current through Q_2. The sum of these currents through R_E is constant. No signal voltage appears at the emitters to cause negative feedback as in Fig. 11-30a.

The result of sequentially varying dc signals in response to temperature change at the sensing element is shown in Fig. 11-31.

Actuators

Cybertronic actuators perform the work in an electronic system. They accept a control signal and translate that signal into mechanical movement. This is used to position valves and dampers. The electrohydraulic actuators are so called

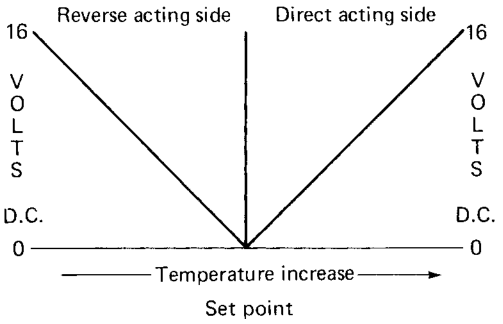

Figure 11-31 Result of sequentially varying dc signals in response to temperature change at the sensing element.

because they convert electrical signals into a fluid movement or force. Damper actuators, equipped with linkage for connection to dampers, and value actuators, having a yoke and linkage to facilitate mounting on a valve body, are also available.

Other Devices

Low- and high-signal selectors accept several control signals. Such selectors then compare the signals and pass the lowest or highest. For example, a high-signal selector can be used on a multi-zone unit to control the cooling coil. The zone requiring the most cooling transmits the highest control signal. This, in turn, will be passed by the high-signal selector to energize the cooling.

The future of electronics in the control of heating, cooling, and refrigeration is unlimited. Only a few examples have been given. More and more mechanically operated thermostats and valves will be replaced and become electronically controlled. A computer has already been utilized in making a programmable controller that can process the information fed to it by many inputs. This will improve process control in industry and increase the comfort of large buildings by more accurately controlling the flow of heat or cooled air. The computer also aids in locating the source of trouble in a system, thereby eliminating a lot of troubleshooting.

Solid-State Compressor Motor Protection

Solid-state circuitry for air-conditioning units has been in use for some time. The following is an illustration of how some of the circuitry has been incorporated into the protection of compressor motors. This module is manufactured by Robertshaw Controls Company.

Solid-state motor protection prevents motor damage caused by excessive temperature in the stator windings. These solid-state devices provide excellent phase-leg protection by means of separate sensors for each phase winding. The principal advantage of this solid-state system is its speed and sensitivity to motor temperature and its automatic reset provision.

There are two major components to the protection system:

1. The protector sensors are embedded in the motor windings at the time the motor is manufactured.
2. The control module is a sealed enclosure containing a transformer and a switch. Figure 11-32 shows two models.

Operation of the module

Leads from the internal motor sensors are connected to the compressor terminals as shown in Fig. 11-33. Leads from the compressor terminals to the control

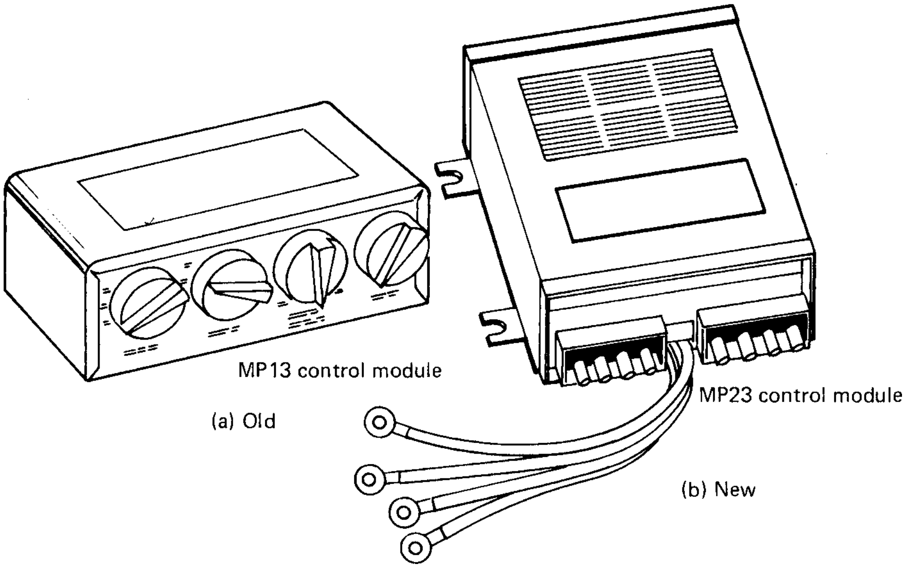

Figure 11-32 Solid-state control modules. (a) Older unit. (b) Newer unit. (Courtesy of Robertshaw)

module are connected as shown in Fig. 11-34. Figure 11-34a shows the older model and Fig. 11-34b the newer model. While the exact internal circuitry is quite complicated, basically the modules sense resistance change through the sensors as the result of motor temperature changes in the motor windings. This resistance change triggers the action of the control circuit relay at predetermined opening and closing settings, which causes the line voltage circuit to the compressor to be broken and completed, respectively.

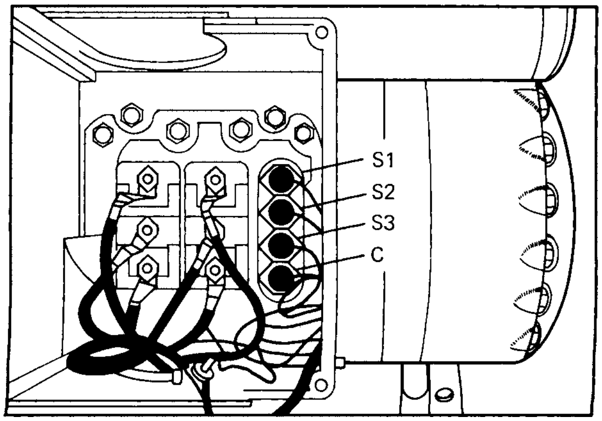

Figure 11-33 Compressor terminal board. (Courtesy of Robertshaw)

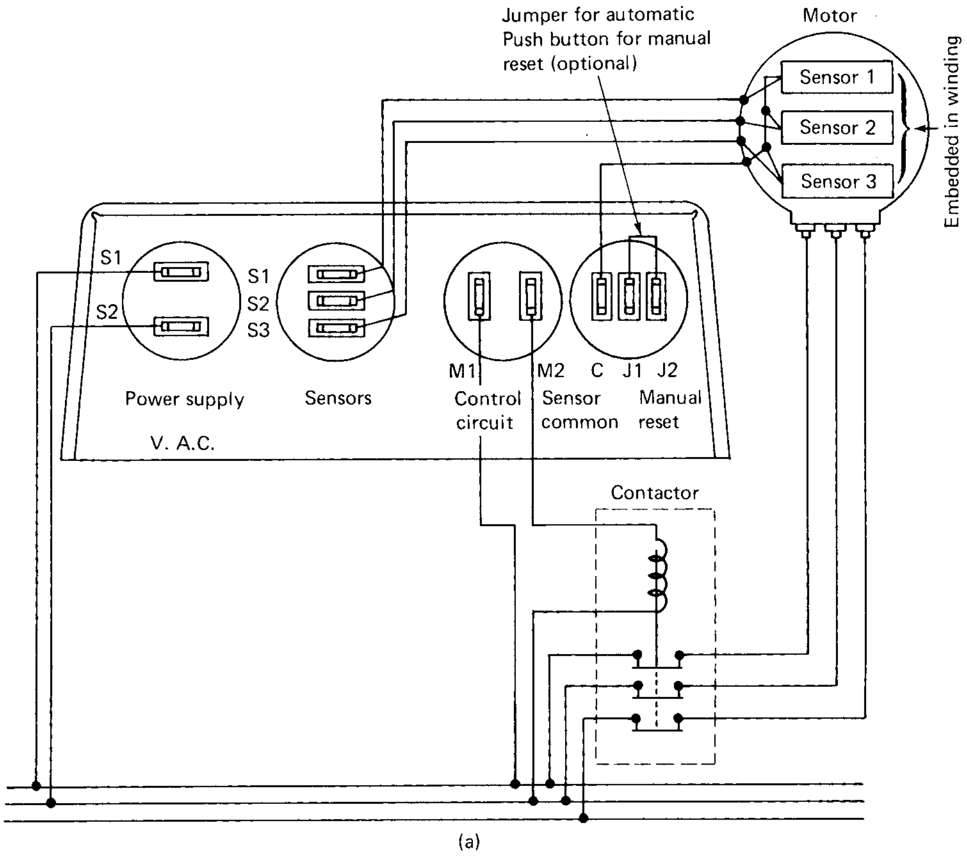
Jumper for automatic
Push button for manual
reset (optional)
Motor
Sensor 1
Sensor 2
Sensor 3
Embedded in winding
S1
S2
S1
S2
S3
Power supply
V. A.C.
Sensors
M1
M2
C
J1
J2
Control
circuit
Sensor
common
Manual
reset
Contactor
(a)

Figure 11-34 Solid-state control modules. (a) Older unit wiring details. (b) Newer unit wiring details.

The modules are available for either 208/240- or 120-volt circuits. The module is plainly marked as to the input voltage. The sensors operate at any of the stated voltages because an internal transformer provides the proper power for the solid-state components.

The two terminals on the module marked Power Supply (T_1 and T_2) are connected to a power source of the proper voltage; normally the line terminals on the compressor motor contactor or the control circuit transformer as required.

Troubleshooting the control

The solid-state module cannot be repaired in the field, and if the cover is opened or the module physically damaged, the warranty on the module is voided. No attempt should be made to adjust or repair this module, and if it becomes defective, it must be returned intact for replacement. This is the usual procedure

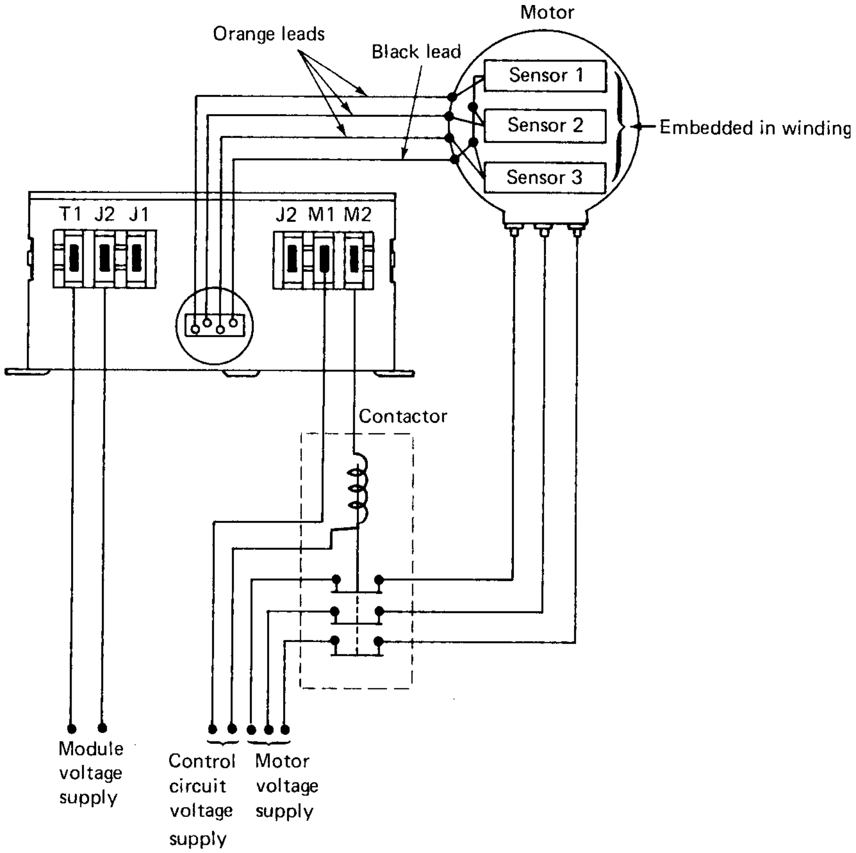
Motor
Orange leads
Black lead
Sensor 1
Sensor 2
Sensor 3
Embedded in winding
T1 J2 J1
J2 M1 M2
Contactor
Module voltage supply
Control circuit voltage supply
Motor voltage supply

Note: Control is automatic reset when terminals J1 and J2 are not included. The control is manual reset when terminals J1 and J2 are included.

(b)

Figure 11-34 (*Continued*)

for most solid-state units. However, if the unit becomes defective, you should be able to recognize that fact and replace it.

If the compressor motor is inoperable or is not operating properly, the solid-state control circuit may be checked as follows:

1. If the compressor has been operating and has tripped on the protector, allow the compressor to cool for at least 1 hour before checking to allow time for the motor to cool and the control circuit to reset.
2. Connect a jumper wire across the control circuit terminals on the terminal board (see Fig. 11-34). This will bypass the relay in the module. If the compressor will not operate with the jumper installed, then the problem is external to the

solid-state protection system. If the compressor operates with the module bypassed, but will not operate when the jumper wire is removed, then the control circuit relay is open.

3. If, after allowing time for motor cooling, the protector still remains open, the motor sensors may be checked as follows:
    a. Remove the wiring connections from the sensor and common terminals on the compressor board (see Figs. 11-33 and 11-34).

WARNING: *Use an ohmmeter with a 3-volt maximum battery power supply. The sensors are sensitive and easily damaged, and no attempt should be made to check continuity through them. Any external voltage or current applied to the sensors may cause damage, necessitating compressor replacement.*

b. Measure the resistance from each sensor terminal to the common terminal. The resistance should be in the following range: 75 ohms (cold) to 125 ohms (hot). Resistance readings in this range indicate the sensors are good. A resistance approaching zero indicates a short. A resistance approaching infinity indicates an open connection. If the sensors are damaged, they cannot be repaired or replaced in the field, and the compressor must be replaced to restore motor protection.

4. If the sensors have proper resistance and the compressor will run with the control circuit bypassed but will not run when connected properly, the solid-state module is defective and must be replaced. The replacement module must have the same voltage specification and made by the same manufacturer as the original module on the compressor.

Restoring service

In the unlikely event that one sensor is damaged and has an open circuit, the control module will prevent compressor operation even though the motor may be in perfect condition. If such a situation should be encountered in the field, as an emergency means of operating the compressor until such time as a replacement can be made, a properly sized resistor can be added between the terminal of the open sensor and the common sensor terminal in the compressor terminal box (see Figs. 11-33 and 11-35). This then indicates to the control module an acceptable resistance in the damaged sensor circuit, and compressor operation can be restored. The emergency resistor should be a 2 watt, 82 ohm, wire wound with a tolerance of ±5%.

In effect, the compressor will continue operation with two-leg protection rather than three-leg protection. While this obviously does not provide the same high degree of protection, it does provide a means of continuing compressor operation with a reasonable degree of safety.

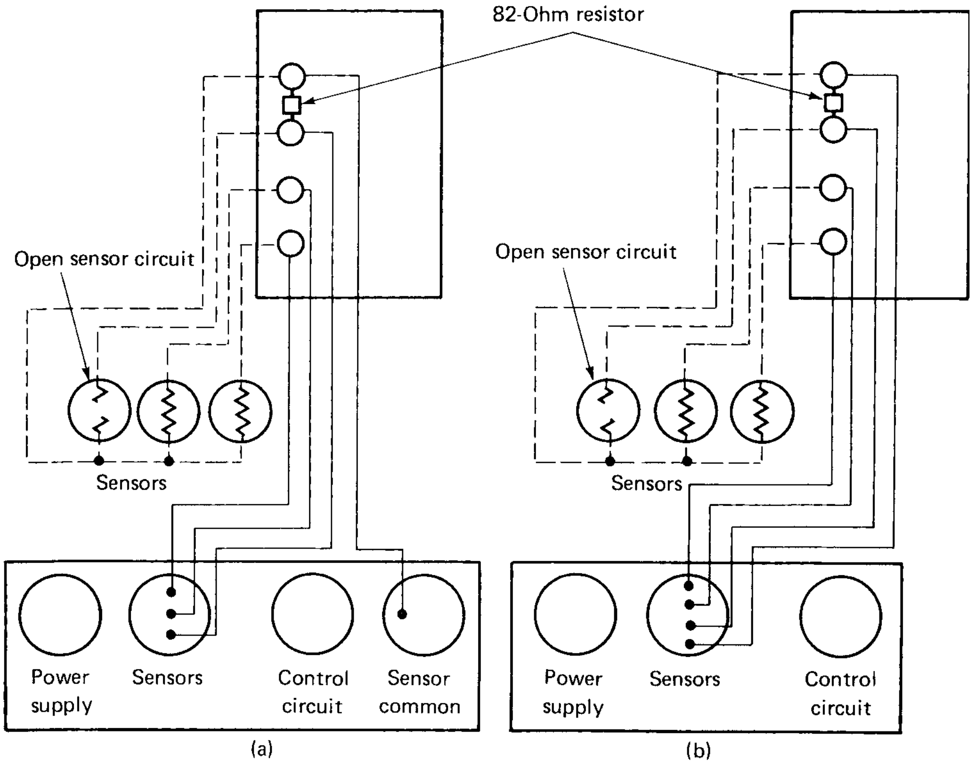
82-Ohm resistor
Open sensor circuit
Sensors
Power supply
Sensors
Control circuit
Sensor common
(a)
Open sensor circuit
Sensors
Power supply
Sensors
Control circuit
(b)

Figure 11-35 Adding a sensor to compensate for an open sensor. (Courtesy of Robertshaw)

Review Questions

1. What does the word semiconductor mean?
2. What are the two materials used for semiconductor devices?
3. What is a diode? What is the PN junction diode used for?
4. What are the uses for diodes?
5. What is an SCR? Where is it used?
6. What are the two main uses for transistors?
7. What is a transistor's current gain?
8. What is an integrated circuit?
9. What is a thermistor?
10. What does CAB stand for in a humidity element?
11. What is a bridge circuit?

12. How do balanced and unbalanced bridges differ?
13. What is a sensor?
14. How is a sensing bridge connected?
15. What is an actuator?
16. What is a differential amplifier used for?

Chapter 12

Alternating Current Motors

Performance Objectives

- Know how an induction motor works.
- Know how to identify a shaded-pole motor and where it is used.
- Know how to identify a split-phase motor and where it is used.
- Know how to identify a three-phase motor and where it is used.
- Know how a capacitor-start motor works.
- Know how a permanent split-capacitor motor works.
- Know where a capacitor-start, capacitor-run motor is used.
- Know how a three-phase motor is electrically connected and properly maintained.

Motors

Most ac motors are of the induction type. They are, in general, simpler and cheaper to build than equivalent dc machines. They have no commutator, slip rings, or brushes, and there is no electrical connection to the rotors. Only the stator winding is connected to the ac source, and, then, as their name implies, induction produces the currents in the rotor. A common and particularly simple form of rotor for this type of motor is the squirrel-cage rotor (see Fig. 12-1). It is so named because of its resemblance to a treadmill-type squirrel cage. The induction motor is based on a rotating magnetic field. This is achieved by using multiple stator field windings (poles), each pair of which is excited by an ac voltage of the same amplitude and frequency as, but phase-displaced from, the voltage supplying the neighboring pair. Figure 12-2 shows how the magnetic field rotates in a four-pole induction motor, where the voltages to the two pairs of poles are 90° out of phase with each other. When the rotor is placed in the stator's rotating field, the induced currents set up their own fields, which react with the stator's field and push the rotor around.

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

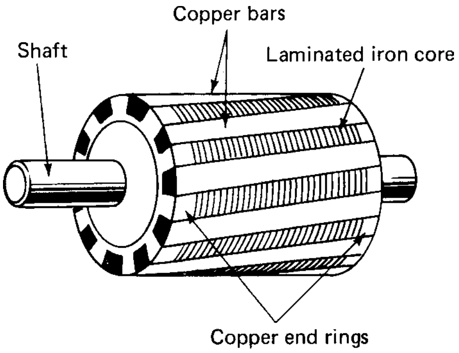

Figure 12-1 Squirrel-cage rotor.

Note that some motors are skewed. The skew of a rotor refers to the amount of angle between the conductor slots and the end face of the rotor laminations. Normally, the conductors are in a nearly straight line, but for high torque applications the rotor is skewed, which increases the angle of the conductors. The term full skew refers to the maximum practical amount (see Fig. 12-3). Figure 12-4 shows how the rotor is located in reference to the stator and the end bells that hold it in place.

A number of types of ac motors are available. The types presented here are those most often encountered when working with heating, air-conditioning, and refrigeration equipment.

The shaded-pole induction motor is a single-phase motor. It uses a unique method to start the rotor turning. The effect of a moving magnetic field is produced by constructing the stator in a special way (see Fig. 12-5).

Portions of the pole piece surfaces are surrounded by a copper strap called a shading coil. The strap causes the field to move back and forth across the face of the pole piece. In Fig. 12-6, a numbered sequence and points on the magnetization curve are shown. As the alternating stator field starts increasing from zero (1), the lines of force expand across the face of the pole piece and cut through the strap. A voltage is induced in the strap. The current that results generates a field that opposes the cutting action (and decreases the strength) of the main field. This action causes certain actions: As the field increases from zero to a maximum of 90°, a large portion of the magnetic lines of force is concentrated in the unshaded portion of the pole (1). At 90° the field reaches its maximum value. Since the lines of force have stopped expanding, no emf is induced in the strap, and no opposite

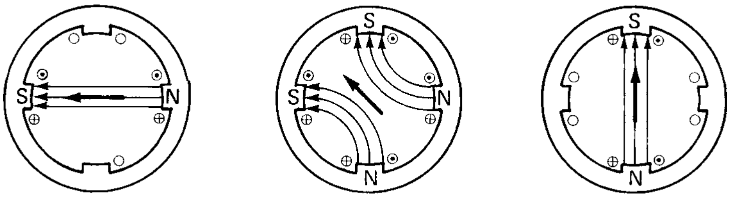

Figure 12-2 Rotating field for a four-pole stator.

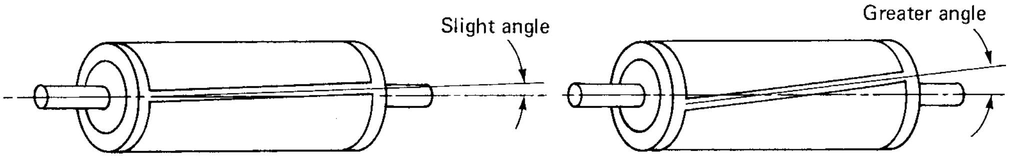
Slight angle
Greater angle

Figure 12-3 Skewed die-cast rotor. Note the angle of the conductor slots.

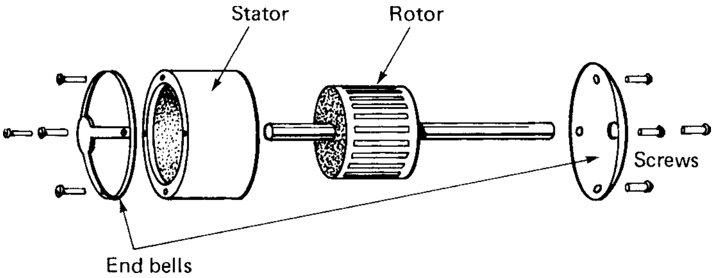
Stator
Rotor
Screws
End bells

Figure 12-4 Exploded view of motor.

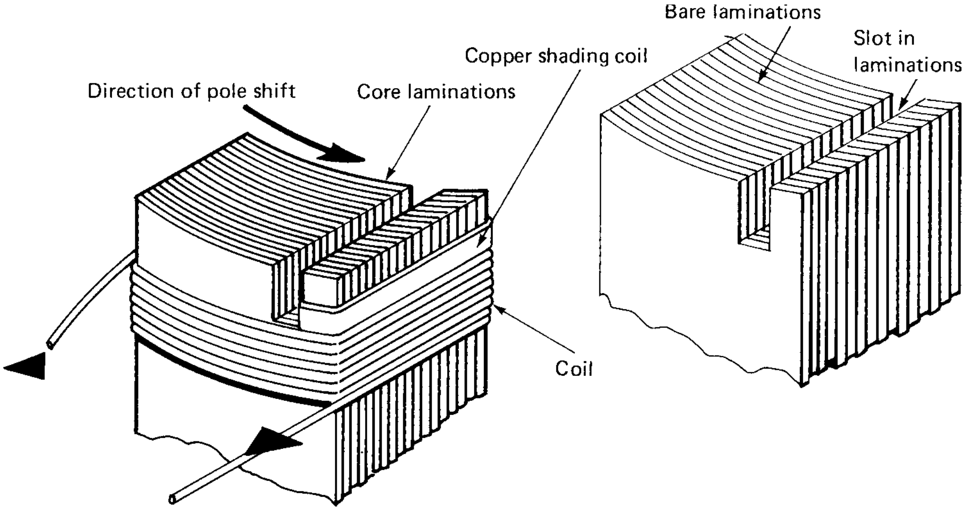
Bare laminations
Slot in laminations
Copper shading coil
Direction of pole shift
Core laminations
Coil

Figure 12-5 Shading the poles of a shaded-pole motor.

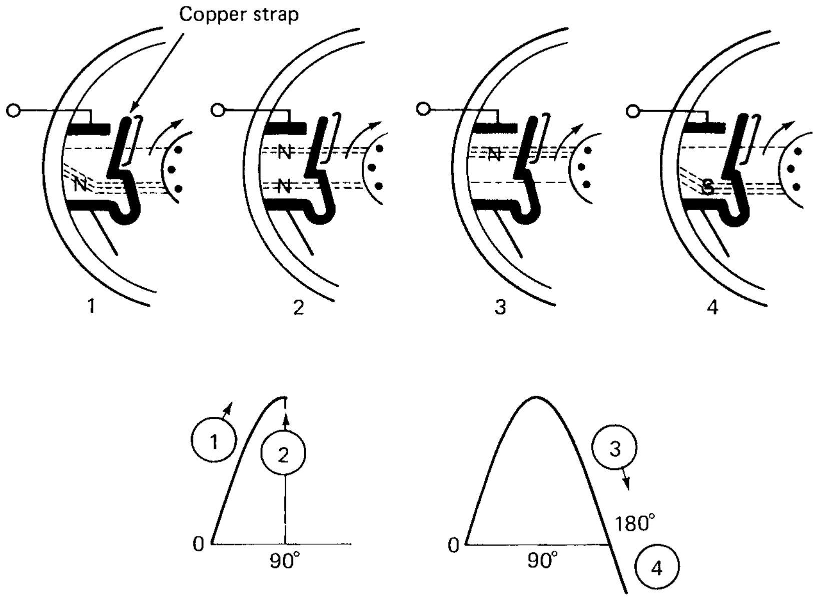

Figure 12-6 Shaded poles used in shaded-pole ac motors.

magnetic field is generated. As a result, the main field is uniformly distributed across the poles as shown in (2).

From 90° to 180°, the main field starts decreasing or collapsing inward. The field generated in the strap opposes the collapsing field. The effect is to concentrate the lines of force in the shaded portion of the poles, as shown in (3).

Note that from 0° to 180°, the main field has shifted across the pole face from the unshaded to the shaded portion. From 180° to 360°, the main field goes through the same change as it did from 0° to 180°. However, it is now in the opposite direction (4). The direction of the field does not affect the way the shaded pole works. The motion of the field is the same during the second half-hertz as it was during the first half-hertz.

The motion of the field back and forth between shaded and unshaded portions produces a weak torque. This torque is used to start the motor. Because of the weak starting torque, shaded-pole motors are built in only small sizes. They drive such devices as fans, timers, and blowers.

Reversibility. Shaded-pole motors can be reversed mechanically. Turn the stator housing and shaded poles end for end. These motors are available from $^{1}/_{25}$ to $^{1}/_{2}$ horsepower.

Uses. As previously mentioned, this type of motor is used as a fan motor in refrigerators and freezers and in some types of air-conditioning equipment where the demand is not too great. It can also be used as part of a timing device for defrost timers and other sequenced operations.

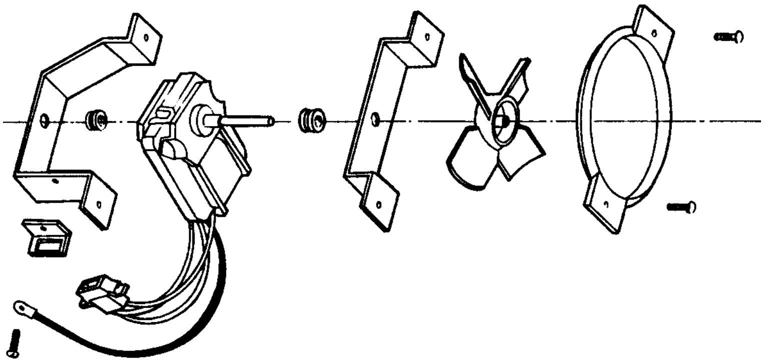

Figure 12-7 Fan, motor, and bracket assembly. (Courtesy of Kelvinator)

The fan and motor assembly are located behind the provisions compartment in the refrigerator, directly above the evaporator in the freezer compartment. The suction-type fan pulls air through the evaporator and blows it through the provisions compartment air dot and freezer compartment fan grille. Figure 12-7 shows a shaded-pole motor with a molded plastic fan blade. For maximum air circulation, the location of the fan on the motor shaft is most important. Mounting the fan blade too far back or too far forward on the motor shaft, in relation to the evaporator cover, will result in improper air circulation. The freezer compartment fan must be positioned with the lead edge of the fan $^1/_4$ inch in front of the evaporator cover. The fan assembly shown in Fig. 12-8 is used on top freezer, no-frost, fiberglass-insulated model refrigerators. The freezer fan and motor assembly is located in the divider partition directly under the freezer air duct.

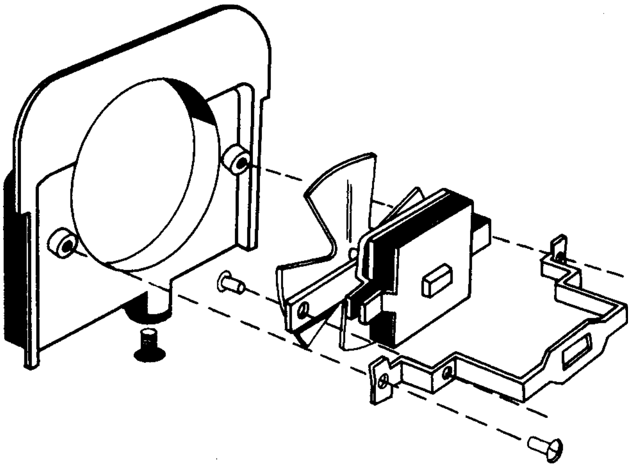

Figure 12-8 Fan and fan motor bracket assembly. (Courtesy of Kelvinator)

Split-Phase Motor

The field of a single-phase motor, instead of rotating, merely pulsates. No rotation of the rotor takes place. A single-phase pulsating field may be visualized as two rotating fields revolving at the same speed, but in opposite directions. It follows, therefore, that the rotor will revolve in either direction at nearly synchronous speed, if it is given an initial impetus in either one direction or the other. The exact value of this initial rotational velocity varies widely with different machines. A velocity higher than 15% of the synchronous speed is usually sufficient to cause the rotor to accelerate to the rated or running speed. A single-phase motor can be made self-starting if means can be provided to give the effect of a rotating field.

To get the split-phase motor running, a run winding and a start winding are incorporated into the stator of the motor. Figure 12-9 shows the split-phase motor with the end cap removed so you can see the starting switch and governor mechanism.

This type of motor is difficult to use with air-conditioning and refrigeration equipment inasmuch as it has very little starting torque and will not be able to start a compressor since it presents a load to the motor immediately upon starting. This type of motor, however, is very useful in heating equipment (see Fig. 12-10).

Getting the motor started

One of the most important parts of the single-phase electric motor is the start mechanism. A special type is needed for use with single-phase motors. A centrifugal switch is used to take a start winding out of the circuit once the motor

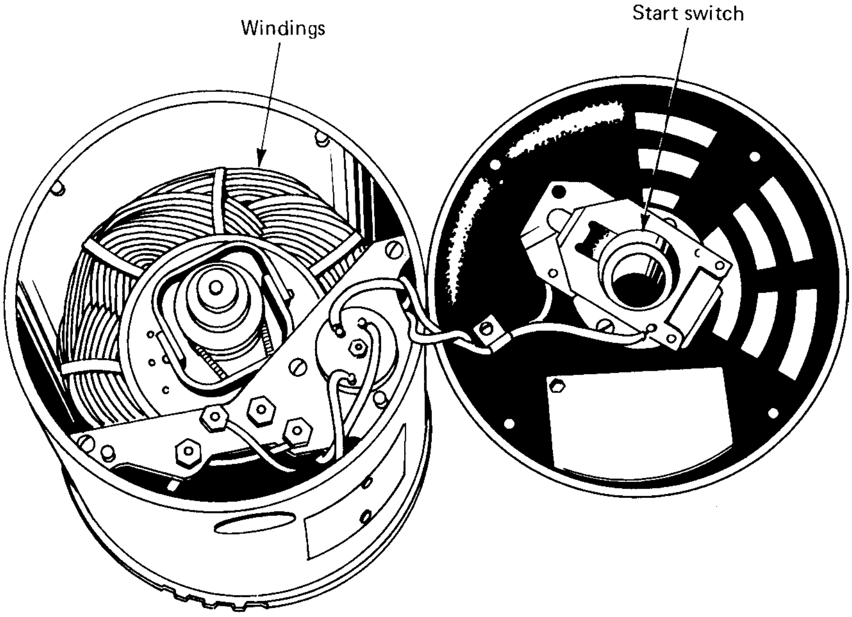

Figure 12-9 Single-phase starting switch and governor mechanism.

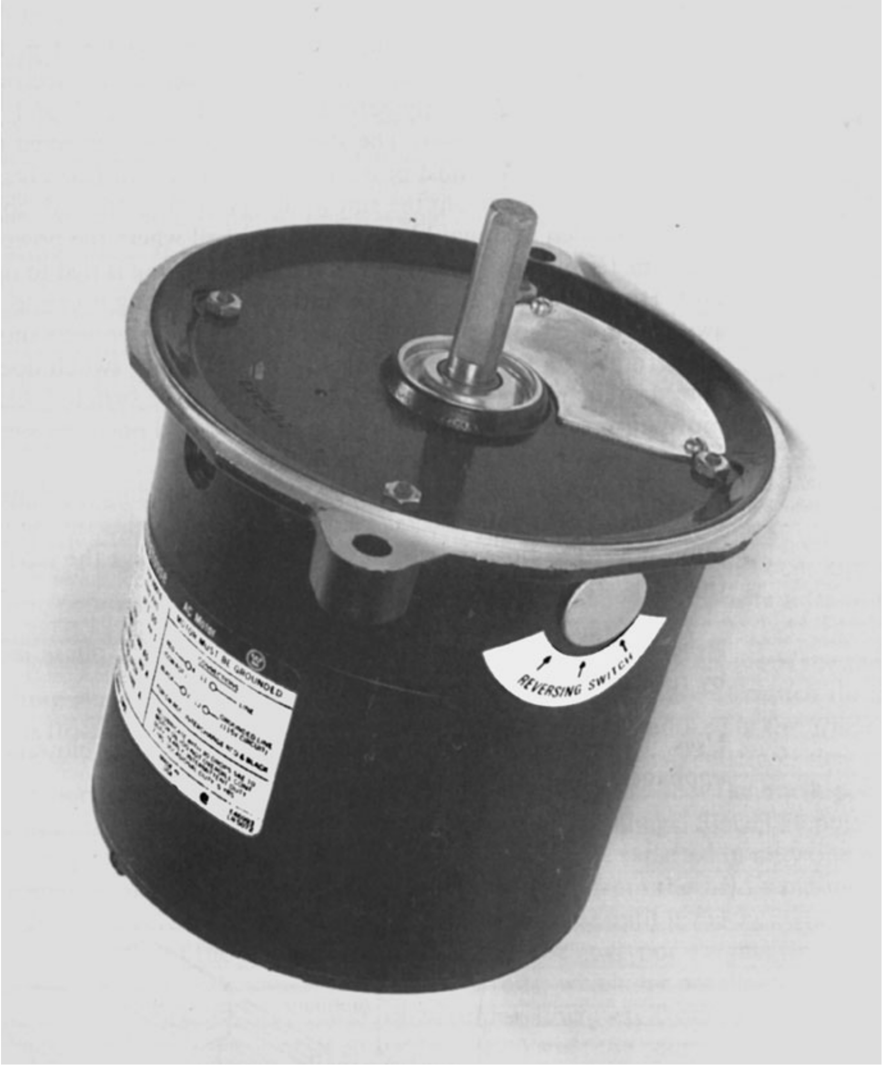

Figure 12-10 Single-phase, split-phase furnace motor. (Courtesy of Westinghouse)

has come up to within 75% of its run speed. The split-phase, capacitor-start, and other variations of these types all need the start mechanism to get them running.

The stator of a split-phase motor has two types of coils; one is called the run winding and the other the start winding. The run winding is made by winding the enamel-coated copper wire through the slots in the stator punchings.

The start winding is made in the same way except that the wire is smaller. Coils that form the start windings are positioned in pairs in the stator directly opposite each other and between the run windings. When you look at the end of the stator, you see alternating run windings and start windings (see Fig. 12-9).

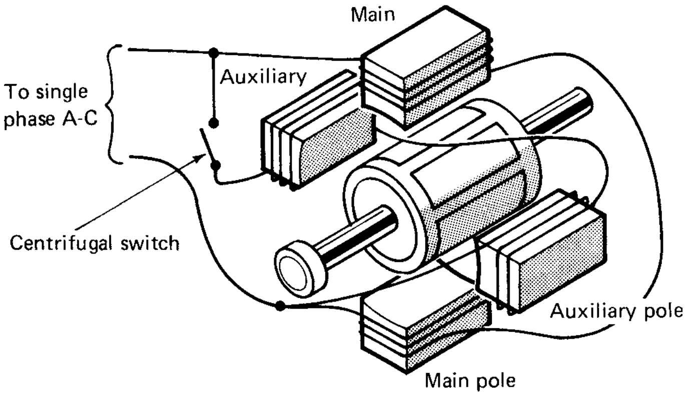

Figure 12-11 Single-phase induction motor.

The run windings are all connected together, so the electrical current must pass through one coil completely before it enters the next coil, and so on through all the run windings in the stator. The start windings are connected together in the same way, and the current must pass through each in turn (see Fig. 12-11).

The two wires from the run windings in the stator are connected to terminals on an insulated terminal block in one end bell where the power cord is attached to the same terminals. One wire from the start winding is tied to one of these terminals also. However, the other wire from the start winding is connected to the stationary switch mounted in the end bell. Another wire then connects this switch to the opposite terminal on the insulated block. The stationary switch does not revolve, but is placed so the weights in the rotating portion of the switch, located on the rotor, will move outward when the motor is up to speed and open the switch to stop electrical current from passing through the start winding.

The motor then runs only on the main winding until such time as it is shut off. Then, as the rotor decreases in speed, the weights on the rotating switch again move inward to close the stationary switch and engage the start winding for the next time it is started.

Reversibility. The direction of rotation of the split-phase motor can be changed by reversing the start winding leads.

Uses. This type of motor is used for fans, furnace blowers, oil burners, office appliances, and unit heaters.

Repulsion Start, Induction Run Motor

The repulsion-induction motor starts on one principle of operation and, when almost up to speed, changes over to another type of operation. Very high twisting forces are produced during starting by the repulsion between the magnetic pole in the armature and the same kind of pole in the adjacent stator field winding. The repulsing force is controlled and changed so that the armature rotational speed increases rapidly, and, if not stopped, would continue to increase beyond a practical operating speed. It is prevented by a speed-actuated mechanical switch that causes the armature to act as a rotor that is electrically the same as the rotor in single-phase induction motors. That is why the motor is called a *repulsion-induction motor*.

The stator of this motor is constructed very much like that of a split-phase or capacitor-start motor, but only run or field windings are mounted inside. End bells keep the armature and shaft in position and hold the shaft bearings.

The armature consists of many separate coils of wire connected to segments of the commutator. Mounted on the other end of the armature are governor weights that move push rods that pass through the armature core. These rods push against a short-circuiting ring mounted on the shaft on the commutator end of the armature. Brush holders and brushes are mounted in the commutator end bell, and the brushes, connected by a heavy wire, press against segments on opposite sides of the commutator (see Fig. 12-12).

When the motor is stopped, the action of the governor weights keeps the short-circuiting ring from touching the commutator. When the power is turned on and current flows through the stator field windings, a current is induced in the armature coils. The two brushes connected together form an electromagnetic coil that

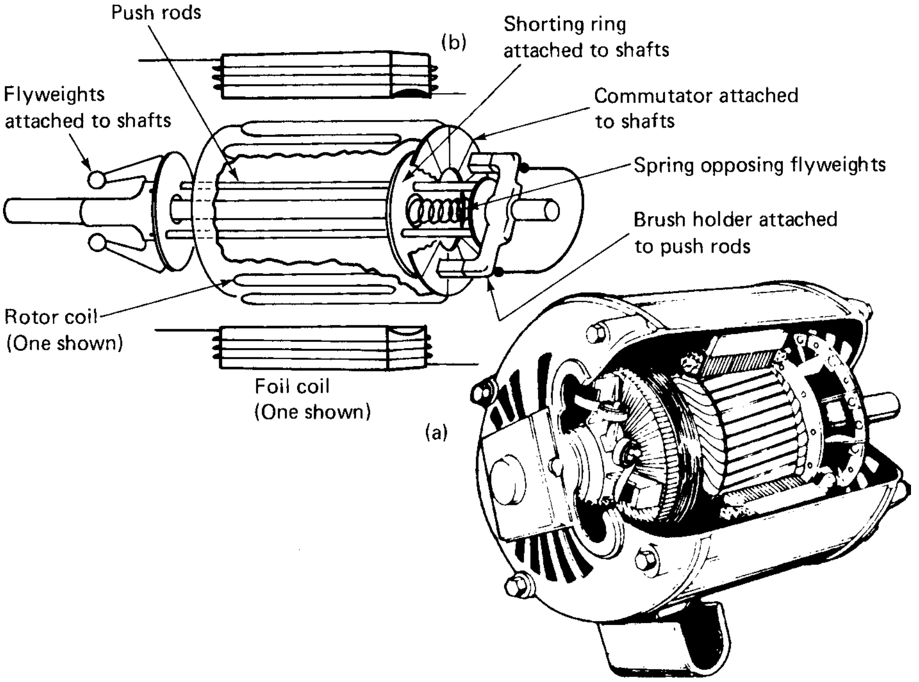
Push rods
(b)
Shorting ring attached to shafts
Flyweights attached to shafts
Commutator attached to shafts
Spring opposing flyweights
Brush holder attached to push rods
Rotor coil (One shown)
Foil coil (One shown)
(a)

Figure 12-12 (a) Brush-lifting, repulsion-start, induction-run motor cutaway. (b) Brush-lifting, repulsion-start, induction motor armature details.

produces a north and south pole in the armature, positioned so that the north pole in the armature is next to a north pole in the stator field windings. Since like poles try to move apart, the repulsion produced in this case can be satisfied in only one way: the armature turns and moves the armature coil away from the field windings.

The armature turns faster and faster, accelerating until it reaches what is approximately 80% of the run speed. At this speed, the governor weights fly outward and allow the push rods to move. These push rods, which are parallel to the armature shaft, have been holding the short-circuiting ring away from the commutator. Now that the governor has reached its designed speed, the rods can move together electrically in the same manner that the cast aluminum discs did in the cage of the induction motor rotor. This means that the motor runs as an induction motor.

The repulsion-induction type of motor can start very heavy, hard-to-turn loads without drawing too much current. They are made from $^1/_2$ to 20 horsepower. This type of motor is used for such applications as large air compressors, refrigeration equipment, and large hoists and is particularly useful in locations where low line voltage is a problem.

This type of motor is no longer used in the refrigeration industry. Some older operating units may be found with this type of motor still in use.

Capacitor-Start Motor

The capacitor motor is slightly different from a split-phase motor. A capacitor is placed in the path of the electrical current in the start winding (see Fig. 12-13). Except for the capacitor, which is an electrical component that slows any rapid change in current, the two motors are the same electrically. A capacitor motor can usually be recognized by the capacitor can or housing that is mounted on the stator (see Fig. 12-14).

Adding the capacitor to the start winding increases the effect of the two-phase field described in connection with the split-phase motor. The capacitor means that the motor can produce a much greater twisting force when it is started. It also reduces the amount of electrical current required during starting to about 1.5 times the current required after the motor is up to speed. Split-phase motors require three or four times the current in starting that they do in running.

Reversibility. An induction motor will not always reverse while running. It may continue to run in the same direction, but at a reduced efficiency. An inertia-type load is difficult to reverse. Most motors that are classified as reversible while running will reverse with a non-inertial-type load. They may not reverse if they are under no-load conditions or have a light lead or an inertial load.

One problem related to the reversing of a motor while it is still running is the damage done to the transmission system connected to the load. In some cases, it is possible to damage a load. One way to avoid this is to make sure the right motor is connected to a load.

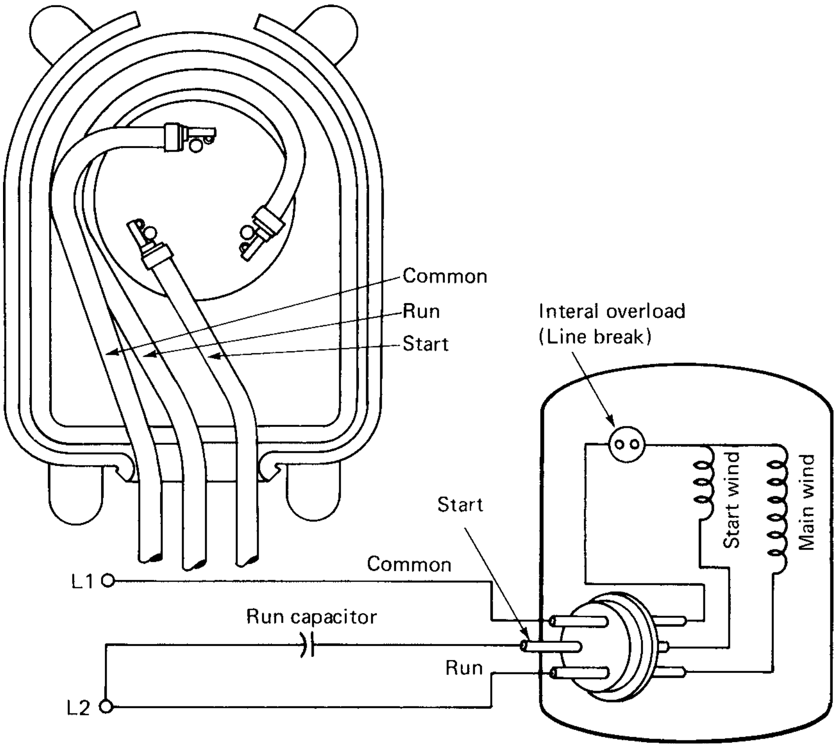
Common
Run
Start
Interal overload
(Line break)
Start wind
Main wind
Start
Common
L1
Run capacitor
Run
L2

Figure 12-13 (A) Single-phase diagram for the AH air conditioner and heat-pump compressor. (Tecumseh) (B) Terminal box showing the position of the terminals on the AH series of compressors. (Courtesy of Tecumseh)

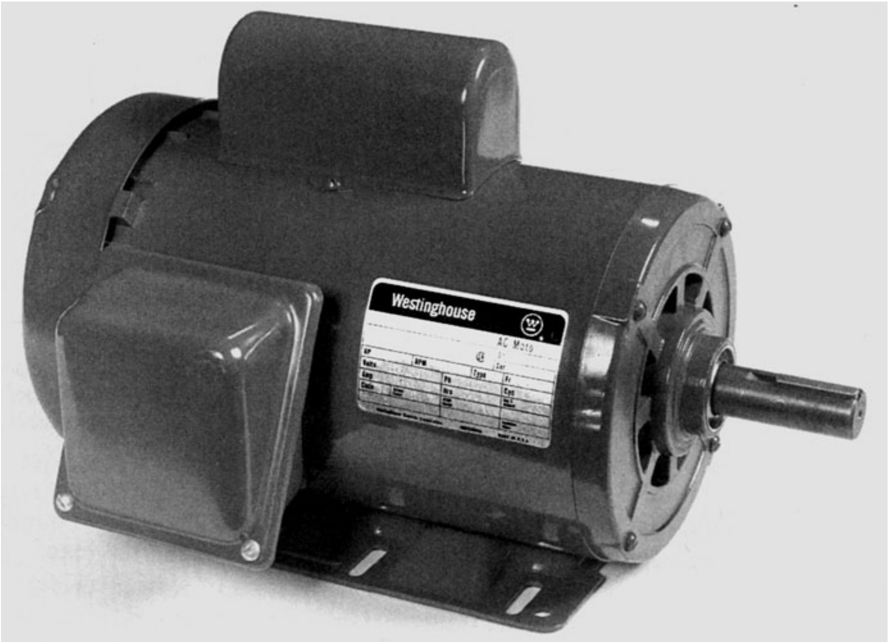
Westinghouse

Figure 12-14 Capacitor-start motor. (Courtesy of Westinghouse)

Reversing (while standing still) the capacitor-start motor can be done by reversing its start winding connections. This is usually the only time that a field technician will work on a motor. The available replacement motor may not be rotating in the direction desired, so the technician will have to locate the start winding terminals and reverse them in order to have the motor start in the desired direction.

Uses. Capacitor motors are available in sizes from $^1/_6$ to 20 horsepower. They are used for fairly hard starting loads that can be brought up to run speed in under 3 seconds. They may be used in industrial machine tools, pumps, air conditioners, air compressors, conveyors, and hoists.

Figure 12-15 shows a capacitor-start, induction-run motor used in a compressor. This type uses a relay to place the capacitor in and out of the circuit. More about this type of relay will be discussed later. Figure 12-16 shows how the capacitor is located outside the compressor.

Permanent Split-Capacitor Motor

The permanent split-capacitor (PSC) motor is used in compressors for air-conditioning and refrigeration units. It has an advantage over the capacitor-start motor inasmuch as it does not need a centrifugal switch with its associated problems.

The PSC motor has a run capacitor in series with the start winding. Both run capacitor and start winding remain in the circuit during start and after the motor is up to speed. Motor torque is sufficient for capillary and other self-equalizing systems. No start capacitor or relay is necessary. The PSC motor is basically an air-conditioner compressor motor. It is also used in refrigerator compressors. It is very common through 3 horsepower. It is also available in the 4- and 5-horsepower sizes (see Fig. 12-17).

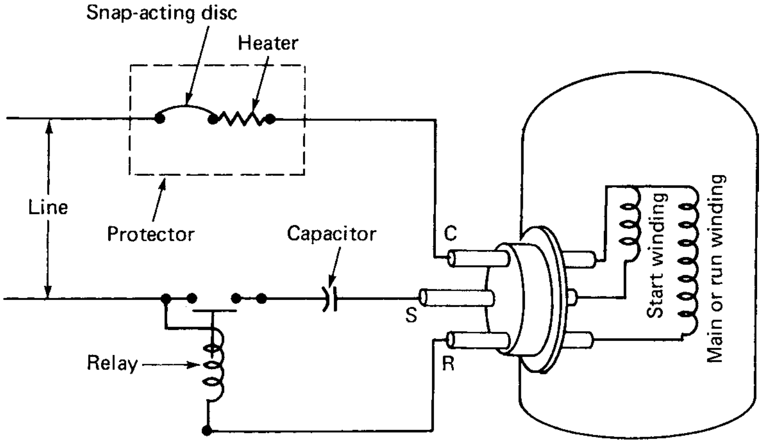

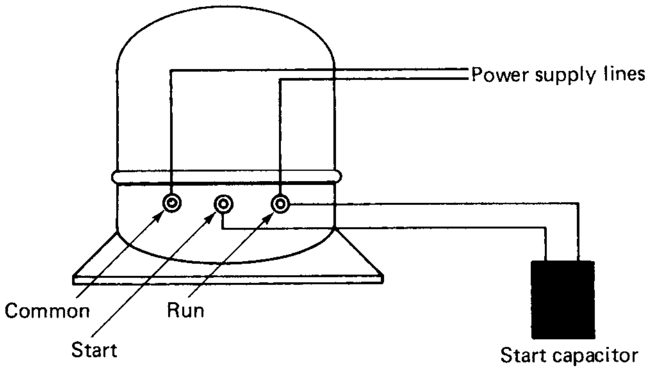
Power supply lines
Common
Start
Run
Start capacitor

Start capacitor* sizes	
Compressor 1/8 hp	Capacitor size is 95 to 200 μF
Compressor 1/6 hp	Capacitor size is 95 to 200 μF
Compressor 1/4 hp	Capacitor size is 200 to 300 μF
Compressor 1/3 hp	Capacitor size is 250 to 350 μF
Compressor 1/2 hp	Capacitor size is 300 to 400 μF
Compressor 3/4 hp	Capacitor size is 300 to 400 μF

*Black case (Bakelite)

Figure 12-16 Location of start capacitor in a compressor circuit. (Courtesy of Tecumseh)

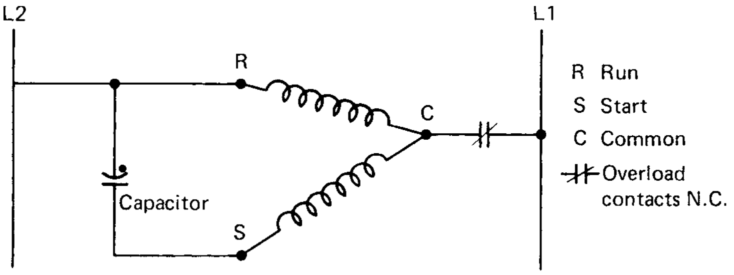
L2
L1
R
C
S
Capacitor
R Run
S Start
C Common
Overload contacts N.C.

(a) Ladder diagram or schematic

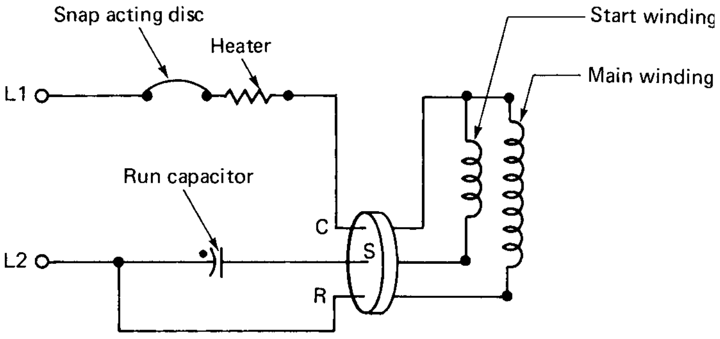
Snap acting disc
Heater
L1
Start winding
Main winding
Run capacitor
C
S
R
L2

(b) Schematic drawing

Figure 12-17 Permanent split-capacitor motor schematic.

Theory of operation

The capacitor is inserted in series with the start winding (see Fig. 12-18). The phase shift produced by the capacitor is similar to that produced by the capacitor-start, induction-run motor. The run capacitor is usually between 5 and 50 μF. Thus, it is smaller than the start capacitor used in the capacitor-start motor. This also means it will have less starting torque than the capacitor-start motor. However, the torque is enough to start the motor running even with a small load. The capacitor is in the circuit all the time. It is not removed by a relay or any other type of device. It is a permanent part of the circuit.

After the compressor or fan is started and begins to run at speed, the motor produces a counter electromotive force. The cemf builds up to within a few volts of the applied voltage when the motor has reached full speed. As long as the difference between the applied voltage and the cemf is small, very little current flows in the start winding. This is because the capacitor will allow more current to pass as the difference in applied voltage and cemf gets larger, and less current to flow when the voltage difference is small. There is a small difference in voltage at full rpm, so the current in the start winding will be small. The 2 to 4 amperes does not constitute enough to cause damage to the compressor start winding (see Fig. 12-19).

The run capacitor left in the circuit aids in speed regulation. One thing to remember on this type of motor is to make sure the capacitor is placed in the circuit properly. The run capacitor has a red dot or some other marking on or

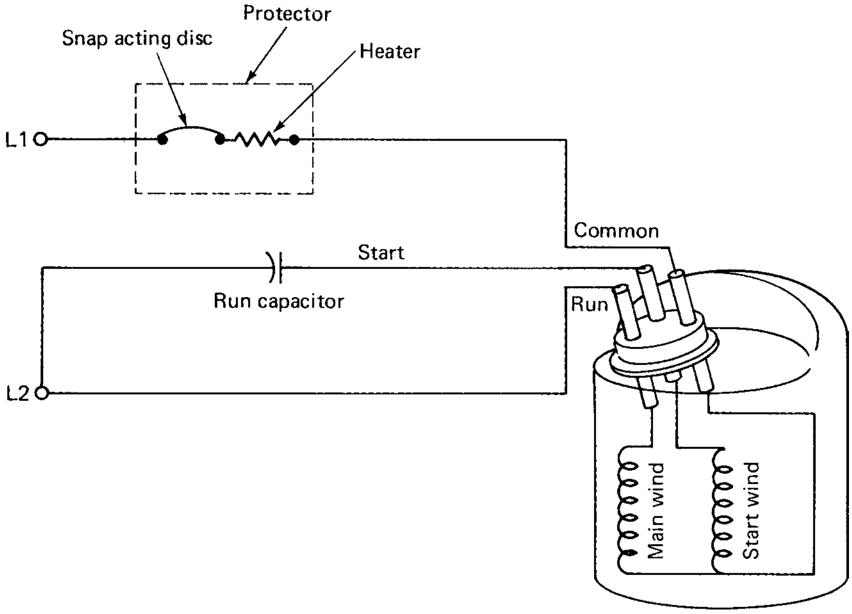

Figure 12-18 Permanent split-capacitor compressor schematic. (Courtesy of Tecumseh)

Run capacitor sizes
1/8 hp compressor or motor uses a capacitor of 4 or 5 μF
1/6 hp compressor or motor uses a capacitor of 4 or 5 μF
1/2 hp compressor or motor uses a capacitor of 10 μF
1/2 to 2 hp compressor or motor uses a capacitor of 10 to 15 μF
3 hp compressor or motor uses 2 capacitors of 10 μF in parallel to equal 20 μF

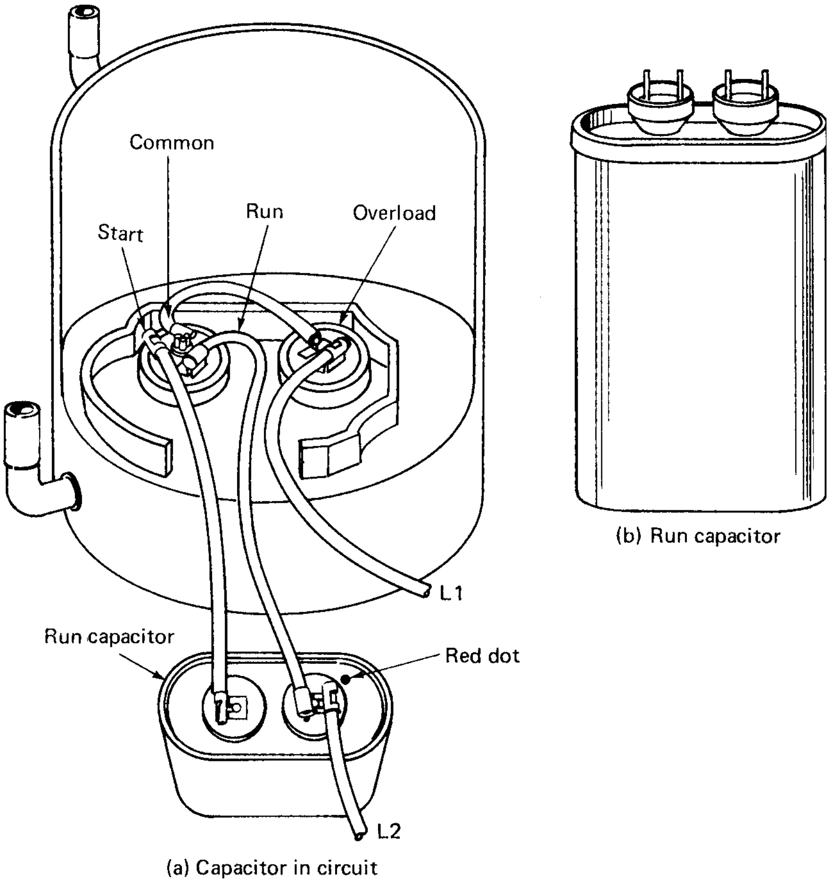

(a) Capacitor in circuit

(b) Run capacitor

Figure 12-19 PSC compressor hookup. (Courtesy of Tecumseh)

near one of the two terminals. This is the outside foil of the capacitor. The run capacitor used here is in a grounded steel case that aids in the dissipation of heat (see Fig. 12-19b). The outside foil is near the capacitor case. If the insulation breaks down or shorts to the outside foil, it will make contact with the case. Excessive current will flow in the circuit when the short occurs. The main objective here is to keep the current as low as possible and have it do the least

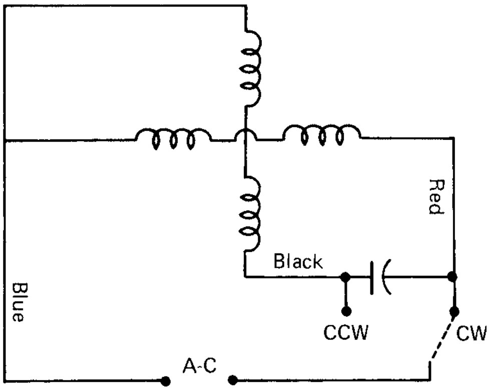

Figure 12-20 Permanent split-capacitor three-lead schematic.

damage possible. That is why the red dot or mark is always placed where it is supposed to be. In Fig. 12-19, the red dot terminal should be connected to L_2. Then, if it shorts, the foil touches the case, shorts to ground, causes excess current to flow, and trips the circuit breaker or blows the fuse.

However, if the red dot is connected incorrectly, the short will burn out the compressor start winding since it puts the winding directly across the power source by grounding one end of the winding. Make sure the capacitor's red dot is connected to the line side and not to the compressor start winding (see Fig. 12-19a).

Reversibility. The PSC motor can be reversed if it has three wires leading from the case (see Fig. 12-20). To reverse, simply connect either side of the capacitor to the line. However, with compressors it is best to leave them as is. They have been designed without the possibility of reversing in most cases. To reverse the four-wire type motor, transpose the black leads as shown in Fig. 12-21.

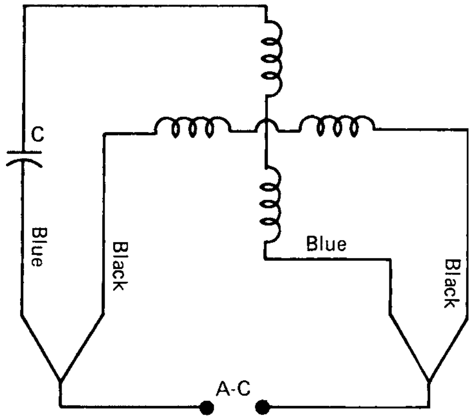

Figure 12-21 Permanent split-capacitor four-lead schematic.

Uses. Permanent split-capacitor motors may also be used for the fans that are mounted behind the condensers on air-conditioning units. They move the air past the condenser and thus remove the heat from the building being air conditioned. This type of motor can also be varied in speed by changing its windings. A number of color-coded windings are brought out so they may be connected to a switching arrangement for low, medium, and high speeds.

Capacitor-Start, Capacitor-Run Motor

The capacitor start and run (CSR or CSCR) motor arrangement uses a start capacitor and a run capacitor in parallel with each other and in series with the motor start windings. This motor has high starting torque and runs efficiently. It is used in many refrigeration and air-conditioning applications up to 5 horsepower. A potential relay removes the start capacitor from the circuit after the motor is up to speed. Potential relays must be as accurately matched to the compressor (see Fig. 12-22). Efficient operation depends on this.

Theory of operation

The CSCR motor is used on air-conditioning and refrigeration units that need large starting torque. You will find the CSCR motor used on equipment that has an expansion valve and on air-conditioning systems when the permanent-split compressor has trouble starting. In some cases, the technician adds a capacitor in the field to make the PSC motor start more easily. This produces the same electrical characteristics as a CSCR type when the additional capacitor is hooked up

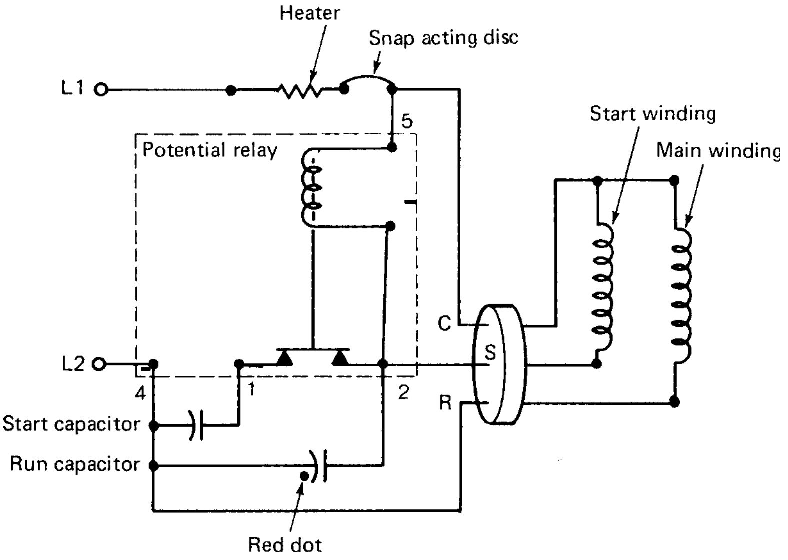

Figure 12-22 Capacitor start and run motor schematic. (Courtesy of Tecumseh)

to the existing arrangement. When the capacitor is added by the technician in the field, it is referred to as a hard-start kit.

Since the CSCR motor has the additional capacitor added during starting time, some way must be provided to move it once its purpose has been served. The potential relay is called upon to do the job. This type of relay will be discussed later in this chapter.

The start capacitor is available in sizes up to 600 μF. This is a large capacitor when compared to the run capacitor of up to 75 μF (see Fig. 12-22). The large capacitance value causes a larger phase shift between the start and run winding voltages. As the compressor motor starts to turn, the cemf begins to build. The cemf is present between the S and C terminals of the compressor. The potential relay coil is connected at terminals 2 and 5 in Fig. 12-22. Note the symbol used for the coil of the potential relay. A PE (for potential relay) and a small resistor-like symbol show the coil part of the relay.

When the compressor reaches approximately 75% of its full rpm, the cemf is strong enough to energize the potential relay coil. This pulls the contacts open. Contactors are shown as two parallel lines with a slant line through them with PR underneath. When the potential relay contacts open, the start capacitor is removed from the circuit. This leaves the run capacitor still in the circuit and in series with the start winding. The result is good starting torque and good running efficiency. The efficiency is increased inasmuch as the power factor is brought closer to unity or 1.00.

Reversibility. This type of motor can be reversed by changing the leads from the start winding. It is, however, difficult to do if the compressor is sealed. In open-type motors, it is possible to reverse the direction of rotation by simply reversing the connections of the start windings.

Uses. This type of motor is used on equipment with a need for good starting torque. This includes some types of refrigeration equipment and some hard to start air-conditioning systems.

Three-Phase Motor

The three-phase motor does not need centrifugal switches or capacitors. The three phases of this type of power generate their own rotating magnetic field when applied to a stator with three sets of windings (see Fig. 12-23). The stator windings are placed 120° apart. The rotor is a form-wound type or a cage type. The squirrel-cage rotor is standard for motors smaller than 1 hp (see Fig. 12-24).

Theory of operation

For the purpose of identification of phases, note that the three phases are labeled A, B, and C in Fig. 12-25. Phase B is displaced in time from A by $^1/_3$ hertz, and phase C is displaced from phase B by $^1/_3$ hertz. In the stator, the different phase windings are plated adjacent to each other so that a B winding is next to an A, a

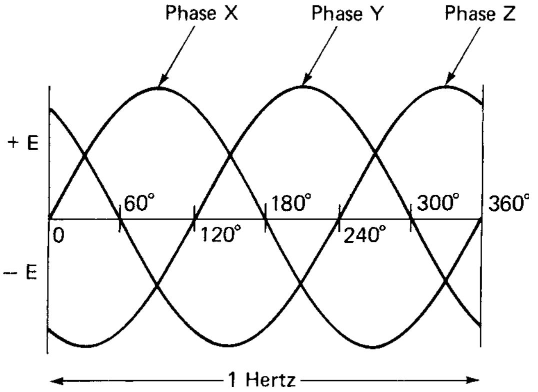

Figure 12-23 Three-phase ac waveform.

C is next to a B, and then an A is next to a C, and so on around the stator (see Fig. 12-25).

The next step is to picture the magnetic fields produced by the three phases in just one group of A, B, and C windings for 1 hertz. Start with the A phase at its maximum positive peak current value. This A winding is a north pole at its maximum strength. As the cycle progresses, the magnetic pole at A will decrease to zero as the current changes direction. It will become a south pole. The strength of the field will increase until the current reaches its greatest negative value. This produces a maximum-strength south pole, and then a decrease. It then passes through zero or neutral and becomes a maximum north pole at the end of the cycle.

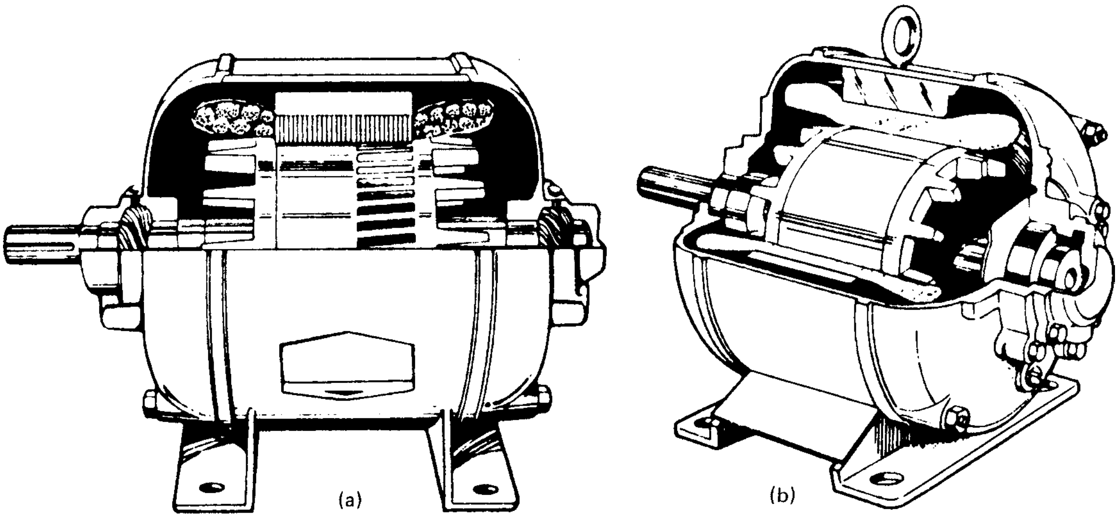

Figure 12-24 Cutaway view of a three-phase motor with (a) a half-etched squirrel-cage rotor and (b) a three-phase motor with a cast rotor.

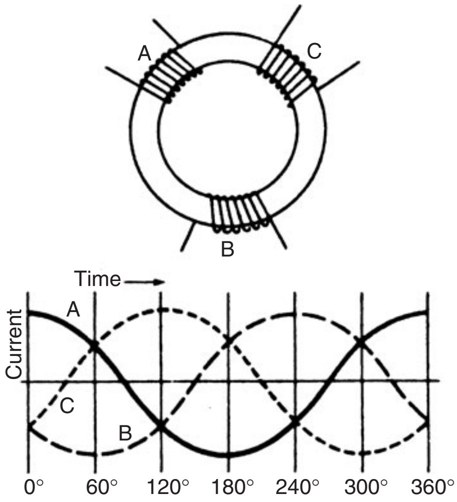

Figure 12-25 Three-phase current and coil placement in a motor.

The B phase winding does exactly the same thing, except that the rise and fall of the magnetic fields follow behind the A phase by $^{1}/_{3}$ hertz, and the C phase winding magnetic field follows behind B by $^{1}/_{3}$ hertz and the A phase by $^{2}/_{3}$ hertz.

Assume that you can see only the maximum north poles produced by each phase. Your view of the motor is from the end of a complete stator connected to a three-phase power source. The north poles will move around the stator and appear to be revolving because of the current relationship of the A, B, and C phases.

The three-phase motor has a rotor that consists of steel discs pressed onto the motor shaft. The slots or grooves are filled with aluminum and connected on the ends to form a cage for the electrical current (see Fig. 12-26). The rotating magnetic field in the stator induces current into this electrical cage, and thereby sets up north and south poles in the rotor. These north and south poles then follow their opposite members in the stator and the shaft rotates. Polyphase induction motors are often called squirrel-cage motors because of this rotor construction. This rotating field makes it possible for the motor to start without capacitors or switches. Thus, it is a simpler motor to maintain and operate.

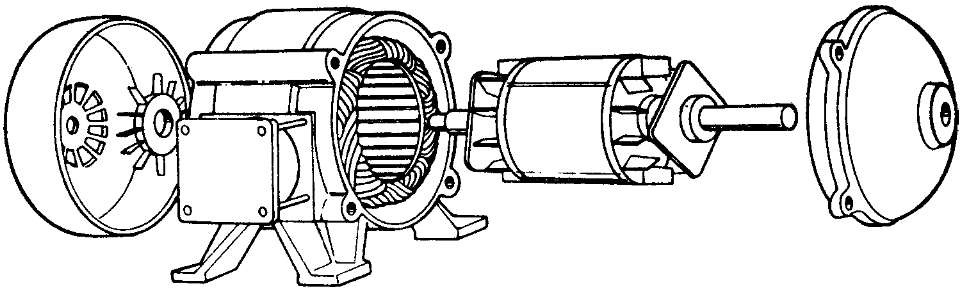

Figure 12-26 Three-phase motor with explosion-proof construction.

Reversibility. Three-phase motors can be reversed while running. It is very hard on the bearings and the driven machine, but it can be done by reversing any two of the three connections. This is usually done by a switch specially designed for the purpose.

Open phase. If a three-phase motor develops an open “leg” or one phase (two instead of three wires are coming into the motor terminal with power), it will slow down and hum noticeably. It will, however, continue to run in the same direction. If you try to start it with only two legs (or phases), it will not start but will rotate if started by hand (in fact, it will start in either direction). Once the other phase is connected, it will quickly come up to speed. The loss of one leg is usually due to a fuse in that leg blowing (that is, if there are three individual fuses in the three-phase circuit, as is normally the case).

Uses. Three-phase motors are used for machine tools, industrial pumps and fans, air compressors, and air-conditioning equipment. They are recommended wherever a polyphase power supply is available. They provide high starting and breakdown torque with smooth pull-up torque. They are efficient to operate and are designed for 208 to 220/460-volt operation with horsepower ratings from one-fourth to the hundreds.

Capacitor Ratings

You have already been introduced to capacitors that are used for starting a motor and for improving the power factor and torque characteristics of the motor. Here we would like to take a look at the ratings and the proper care of capacitors for long-time operation.

Never use a capacitor with a lower rating than specified on the original equipment. The voltage rating and the microfarad rating are important. A higher voltage rating than that specified is always usable. However, a voltage rating lower than that specified can cause damage. Make sure the capacitance marked on the capacitor in microfarads is as specified. Replace with a capacitor of the same size rated in μF, MF, UF, or MFD. All these abbreviations are used to indicate microfarads. See Table 12-1 for a listing of the start capacitors and their ratings for different voltages. Note that the capacitors are not exactly what they are rated for in microfarads. They have a tolerance and the limits are shown.

Start Capacitors and Bleeder Resistors

The development of high-power-factor, low-current, single-phase compressor motors that require start and run capacitors used with potential-type relays has created electrical peculiarities that did not exist in previous designs. In some situations, relay contacts may weld together, causing compressor motor failure. This phenomenon occurs due to the high voltage in the start capacitor discharging

TABLE 12-1 Ratings and Test Limits for AC Electrolytic Capacitors

Capacity rating (microfarads)			110-volt ratings		125-volt ratings		220-volt ratings	
Nominal	Limits	Average	Amps. at rated voltage, 60 Hz	Approx. max. watts	Amps. at rated voltage, 60 Hz	Approx. max. watts	Amps. at rated voltage, 60 Hz	Approx. max. watts
	25–30	27.5	1.04–1.24	10.9	1.18–1.41	14.1	2.07–2.49	43.8
	32–36	34	1.33–1.49	13.1	1.51–1.70	17	2.65–2.99	52.6
	38–42	40	1.56–1.74	15.3	1.79–1.98	19.8	3.15–3.48	61.2
	43–48	45.5	1.78–1.99	17.5	2.06–2.26	22.6	3.57–3.98	70
50	53–60	56.5	2.20–2.49	21.9	2.50–2.83	28.3	4.40–4.98	87.6
60	64–72	68	2.65–2.99	26.3	3.02–3.39	33.9	5.31–5.97	118.2
65	70–78	74	2.90–3.23	28.4	3.30–3.68	36.8	5.81–5.97	128.2
70	75–84	79.5	3.11–3.48	30.6	3.53–3.96	39.6	6.22–6.97	138
80	86–96	91	3.57–3.98	35	4.05–4.52	45.2	7.13–7.96	157.6
90	97–107	102	4.02–4.44	39.1	4.57–5.04	50.4	8.05–8.87	175.6
100	108–120	114	4.48–4.98	43.8	5.09–5.65	56.5	8.96–9.95	197
115	124–138	131	5.14–5.72	50.3	5.84–6.50	65		
135	145–162	154	6.01–6.72	62.8	6.83–7.63	85.8		
150	161–180	170	6.68–7.46	69.8	7.59–8.48	95.4		
175	189–210	200	7.84–8.74	81.4	8.91–9.90	111.4		
180	194–216	205	8.05–8.96	83.8	9.14–10.18	114.5		
200	216–240	228	8.96–9.95	93	10.18–11.31	127.2		
215	233–260	247	9.66–10.78	106.7	10.98–12.25	145.5		
225	243–270	257	10.08–11.20	110.9	11.45–12.72	151		
250	270–300	285	11.20–12.44	123.2	12.72–14.14	167.9		
300	324–360	342	13.44–14.93	147.8	15.27–16.96	201.4		
315	340–380	360	14.10–15.76	156				
350	378–420	399	15.68–17.42	172.5				
400	430–480	455	17.83–19.91	197.1				

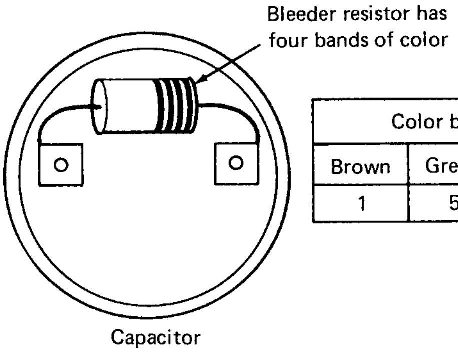

Color bands and values			
Brown	Green	Red and	Silver
1	5	00	± 10%

Figure 12-27 Bleeder resistor across the capacitor terminals.

(arcing) across the potential relay contacts. To eliminate this, start capacitors are equipped with bleeder resistors across the capacitor terminals (see Fig. 12-27).

Bleeder resistor-equipped capacitors may not be available. Then a 2-watt, 1500-ohm resistor can be soldered across the capacitor terminals. This does not interfere with the operation of the circuit, but does allow the capacitor to discharge across the resistor instead of across the relay switch points. This resistor is called a *bleeder* because it bleeds off the capacitor charge.

Run capacitors

The marked terminal of run capacitors should be connected to the R terminal of the compressor and thus to L_2. Check the wiring diagram for the correct terminal. The run capacitor is in the circuit whenever the compressor is running. It is an oil-filled electrolytic capacitor that can take continuous use.

If the start capacitor is left in the circuit too long, in some cases the coil of the potential relay will open from vibration or use and cause the start capacitor to stay in the circuit longer than 10 to 15 seconds. This causes the electrolytic capacitor to explode or spew out its contents.

Motor Protectors

In most compressors, there is a motor protector (see Fig. 12-28). The overload protector is inserted into the motor windings so that if they overheat, the device will open the contacts of the switch. The bimetal element expands to cause the contacts to remove the power to the windings. These protectors are in addition to any circuit breakers that may be mounted outside the compressor.

Compressor Motor Relays

A hermetic compressor motor relay is an automatic switching device designed to disconnect the motor start winding after the motor has come up to running speed (see Fig. 12-29). The two types of motor relays used in refrigeration and air-conditioning compressors are the current and the potential type.

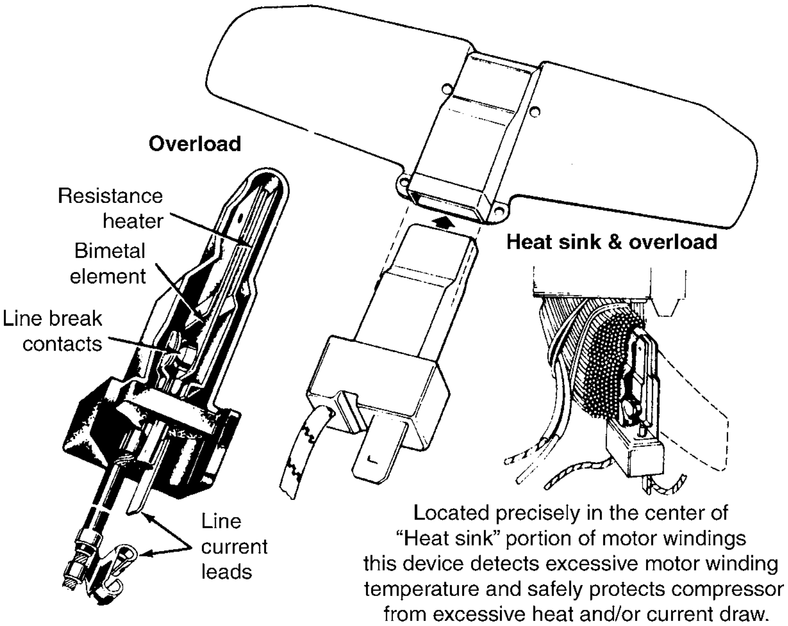
Overload
Resistance heater
Bimetal element
Line break contacts
Line current leads
Heat sink & overload
Located precisely in the center of "Heat sink" portion of motor windings this device detects excessive motor winding temperature and safely protects compressor from excessive heat and/or current draw.

Figure 12-28 Motor internal line-break protector. (Courtesy of Tecumseh)

Current-type relay

The current-type relay is most often used with small refrigeration compressors up to a horsepower. Where power is applied to the compressor motor, the relay solenoid coil attracts the relay armature upward. This causes bridging contact and stationary contact to engage (see Fig. 12-30). This energizes the motor start winding. When the compressor motor comes up to running speed, the motor's main winding current is such that the relay solenoid coil de-energizes. This allows the relay contacts to drop open, which disconnects the motor start winding.

One thing to remember about this type of relay is its mounting. It should be mounted in a true vertical position so that the armature and bridging contact will drop free when the relay solenoid is de-energized.

Potential-type relay

This relay is generally used with large commercial and air-conditioning compressors (see Fig. 12-31). The motors may be capacitor-start, capacitor-run types up to 5 horsepower. Relay contacts are normally closed. The relay coil is wired across the start winding. It senses voltage change. Start winding voltages increase with motor speed. As the voltage increases to the specific pickup value, the armature pulls up, opening the relay contacts and de-energizing the start windings. After switching, there is still sufficient voltage induced in the start winding to keep the relay coil energized and the relay starting contacts open. When the

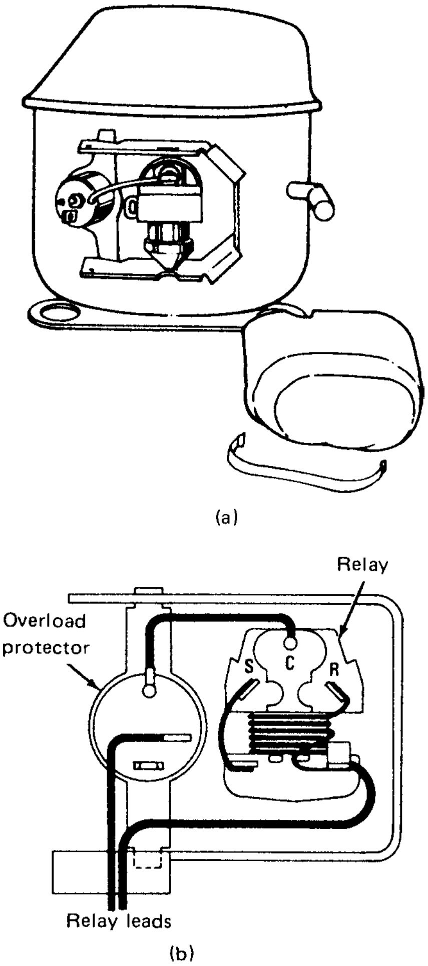
(a)
Relay
Overload protector
S
C
R
Relay leads
(b)

Figure 12-29 (a) Location of the overload and relay on a compressor. (Courtesy of Tecumseh). (b) Starting relay and overload protector. (Courtesy of Kelvinator)

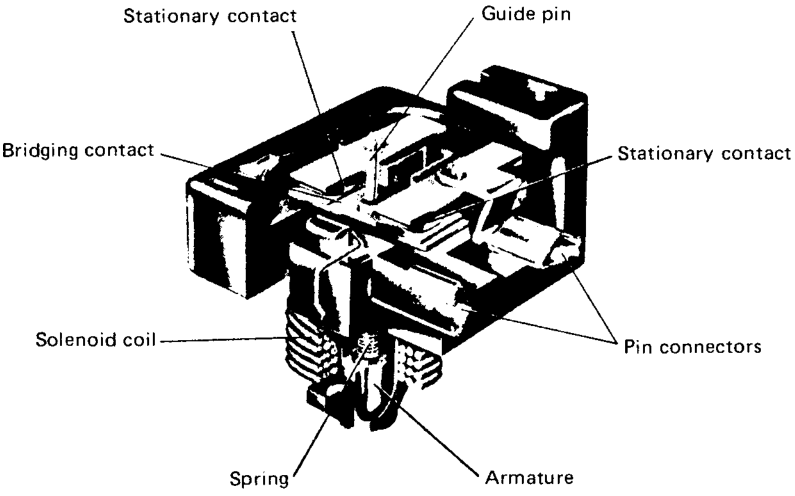
Stationary contact
Guide pin
Bridging contact
Stationary contact
Solenoid coil
Pin connectors
Spring
Armature

Figure 12-30 Current relay. (Courtesy of Tecumseh)

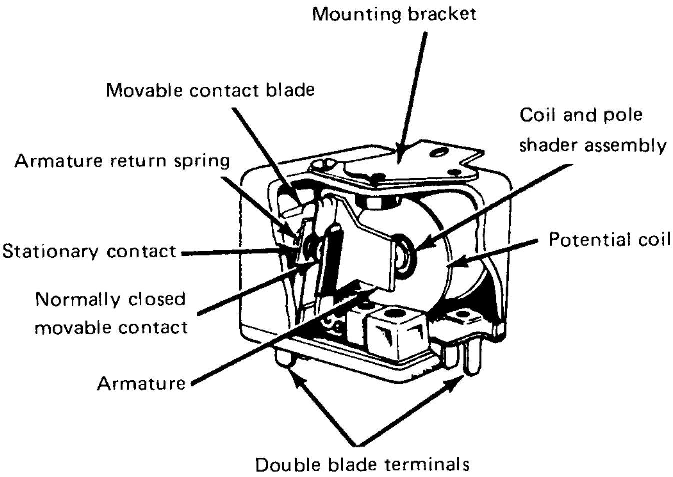
Mounting bracket
Movable contact blade
Coil and pole shader assembly
Armature return spring
Stationary contact
Potential coil
Normally closed movable contact
Armature
Double blade terminals

Figure 12-31 Potential-type relay. (Courtesy of Tecumseh)

power is shut off to the motor, the voltage drops to zero. The coil is de-energized and the start contact is reset for the next start.

Many of these relays are extremely position sensitive. When changing a compressor relay, care should be taken to install the replacement in the same position as the original. Never select a replacement relay solely by horsepower or other generalized rating. Select the correct relay from the parts guide book furnished by the manufacturer.

Visual inspection can distinguish the two relays. The current relay has heavy wire for the coil and the potential relay has fine wire for the coil.

Review Questions

1. On what principle is the induction motor based?
2. What does the skew of a motor mean?
3. What is a shaded-pole motor? Where is it used?
4. What is a split-phase motor? Where is it used?
5. What is a repulsion-start, induction-run motor? Where is it used?
6. What is a capacitor-start motor? Where is it used?
7. What is a permanent split-capacitor (PSC) motor? Where is it used?
8. What is a capacitor-start, capacitor-run motor? Where is it used?
9. What is a three-phase motor? Where is it used?

Chapter

13

Electrical Safety

Performance Objectives

- Know how to work around electricity safely.
- Know what precautions to take when using portable electrical hand tools.
- Know how the ground-fault circuit interrupter works.
- Know where the GFCI is used.
- Know the color-coding of wiring systems.

Safety Devices

Air-conditioning, refrigeration, and heating systems all utilize the convenience of electrical controls. These devices are made with a particular application in mind. Each is designed for a specific purpose. By using low voltage in most instances to control higher voltages, the devices are suited for remote operation since the size of the wire is rather small and inexpensive to install.

Some control devices are designed to protect the technician while repairing or maintaining equipment. Certain safety procedures are necessary to make sure you do not receive a fatal electric shock. The following should be helpful in making it safe to work on electrically controlled and operated equipment.

Safety Precautions

It takes very little current to cause physical damage to the human body. In some cases, death may result from as little as one-tenth (0.1) of an ampere (see Table 13-1). You may react to a slight shock and then move quickly and come in contact with operating machinery. Involuntary actions caused by electric shock are more harmful in most instances than the actual mild shock. It is very easy to become careless. Just keep in mind that the human body has a skin

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

TABLE 13-1 Physiological Effects of Electric Currents

Reading (mA)		Effects
Safe current values	1 or less	Causes no sensation—not felt.
	1–7	Sensation of shock, not painful; individual can let go at will since muscular control is not lost.
Unsafe current values	8–15	Painful shock; individual can let go at will since muscular control is not lost.
	15–20	Painful shock; control of adjacent muscles lost; victim cannot let go.
	20–50	Painful, severe muscular contractions; breathing difficult.
	50–100	Ventricular fibrillation, a heart condition that can result in instant death is *possible.*
	100–200	Ventricular fibrillation occurs.
	200 and over	Severe burns, severe muscular contractions, so severe that chest muscles clamp the heart and stop it for the duration of the shock. (This prevents ventricular fibrillation.)

resistance of between 400,000 and 800,000 ohms. This can be used in an Ohm's law formula to determine the current

$$I = E/R$$

$$I = 120/400{,}000 = 0.3 \text{ milliamperes}$$

This is not enough to cause you to feel it. However, if you have wet hands or make contact with part of the body that is not dry, you may have a body resistance as low as 50,000 ohms and receive a current of 2.4 milliamperes. That is enough to cause you to jump back from a slight tingle. The backward movement on a roof may be enough to cause you to fall a great distance. You may also drop the equipment or tools you are working with and cause further damage. If you are working around 240 volts, the danger is greater. At 240 volts, a contact resistance of 50,000 ohms produces 4.8 milliamperes or enough to cause a heavy shock with tightening of the muscles. A low body resistance of say 200 ohms can be fatal. This can happen whenever the skin is wet or you are standing on a wet surface. With 200 ohms resistance and 120 volts current, you will receive 600 milliamperes, enough to cause death.

Main Switches

If you are working on equipment with the power on, be sure you are standing on a dry surface and your hands do not make contact with anything other than the probes being used to make measurements. It is always a good practice when working around live circuits to keep one hand in your pocket so that you will not complete the path across your chest. It is, of course, always better to have the power off when working on equipment. A good procedure is to turn off the main circuit breakers and put a sign on them so that no one turns them on without your knowledge. If there is a control box or distribution panel with a lock on it, make sure the lock is in place and locked with the lever in the down or off position (see Figs. 13-1 and 13-6). Fusing of the circuits is shown in Fig. 13-1.

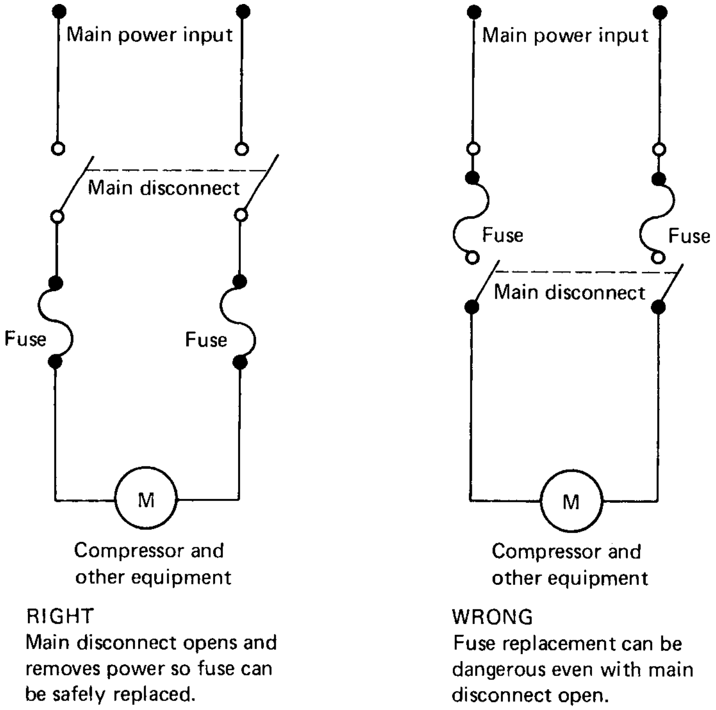
Main power input
Main disconnect
Fuse
Fuse
M
Compressor and
other equipment
RIGHT
Main disconnect opens and
removes power so fuse can
be safely replaced.
Main power input
Fuse
Fuse
Main disconnect
M
Compressor and
other equipment
WRONG
Fuse replacement can be
dangerous even with main
disconnect open.

Figure 13-1 Main disconnect should be located before working on equipment.

Portable Electrical Tools

Grounding portable tools aids in preventing shocks and damage to equipment under certain conditions. A grounding conductor has a white-colored jacket in a two- or three-wire cable (neutral wire). It is terminated to the white- or silver-colored terminal in a plug cap or connector. And it is terminated at the neutral bar in the distribution box. Keep in mind that gray is also used to color code wires in industrial and commercial installations. Gray wires are also used for grounding.

When there is an electrical fault that allows the hot line to contact the metal housing of electrical equipment, in a typical two-wire system, or some other ungrounded conductors, any person who touches that equipment or conductor will receive a shock. The person completes the circuit from the hot line to the ground and current passes through the body. Because a body is not a good conductor, the current is not high enough to blow the fuse. Thus, the current continues to pass through the body as long as the body remains in contact with the equipment (see Fig. 13-2).

A grounding conductor, or equipment ground, is a wire attached to the housing or other conductive parts of electrical equipment that are not normally energized to carry current from them to the ground. Thus, if a person touches a part that is accidentally energized, there will be no shock, because the grounding line furnishes a much lower resistance path to the ground (see Fig. 13-3). The high current passes through the wire conductor and blows the fuses and stops the current. In normal operation, a grounding conductor does not carry current.

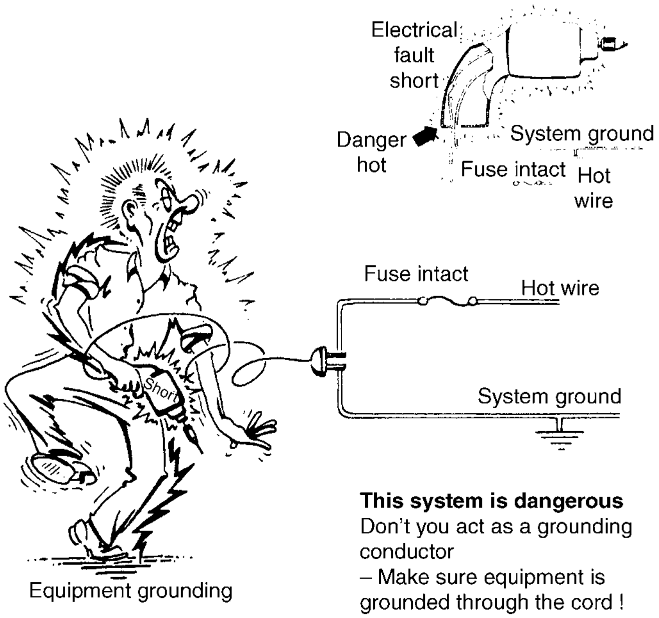
Electrical fault short
Danger hot
System ground
Fuse intact
Hot wire
Fuse intact
Hot wire
Short
System ground
This system is dangerous
Don't you act as a grounding conductor
– Make sure equipment is grounded through the cord !
Equipment grounding

Figure 13-2 A person who contacts the charged housing of a drill or piece of equipment becomes the conductor in a short circuit to ground. (Courtesy of National Safety Council)

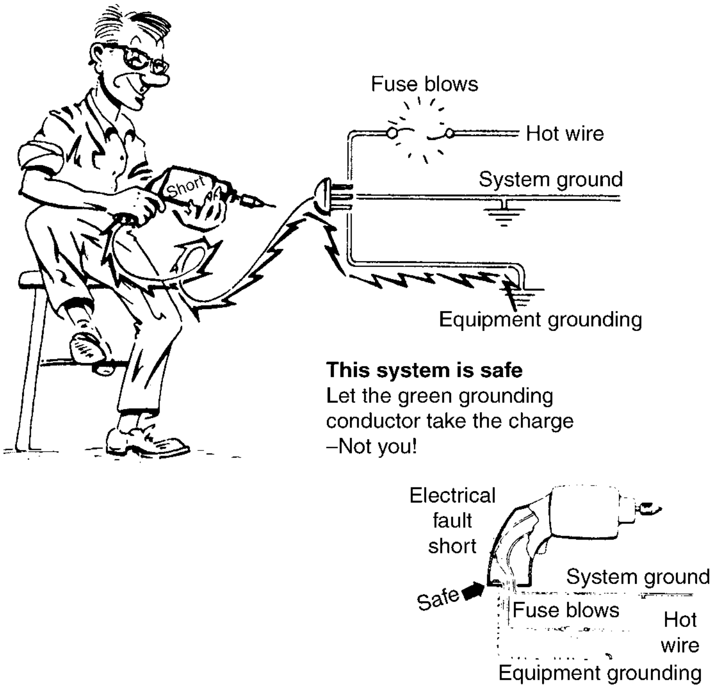
Fuse blows
Hot wire
System ground
Short
Equipment grounding
This system is safe
Let the green grounding conductor take the charge
–Not you!
Electrical fault short
Safe
System ground
Fuse blows
Hot wire
Equipment grounding

Figure 13-3 Properly wired circuit causes the shorted equipment to be sorted to ground instead of through the person to ground. (Courtesy of National Safety Council)

The grounding conductor in a three-wire conductor cable has a green jacket. The grounding conductor is always terminated at the green-colored hexhead screw on the cap or connector. It utilizes either a green-colored conductor or a metallic conductor as its path to ground. In Canada, and England, this conductor is referred to as the earthing conductor, which is somewhat more descriptive and helpful in distinguishing between grounding conductors and neutral wires, or grounded conductors.

Ground-Fault Circuit Interrupters (GFCI)

A safety device that should be used by all technicians in the field who do not have time to test out each circuit before it is used is the ground-fault circuit interrupter (GFCI). It is designed to protect you from shock.

The differential ground-fault interrupter, available in various modifications, has current-carrying conductors passing through the circular iron core of a doughnut-shaped differential transformer. As long as all the electricity passes through the transformer, the differential transformer is not affected and will not trigger the sensing circuit. If a portion of the current flows to ground and through the fault-detector line, however, the flow of electricity through the sensing windings of the differential transformer causes the sensing circuit to open the circuit breaker. These devices can be arranged to interrupt a circuit for currents of as little as 5 milliamperes flowing to ground (see Fig. 13-4).

Another design is the isolation-type ground-fault interrupter. This unit combines the safety of an isolation system with the response of an electronic sensing circuit. In this setup, an isolating transformer provides an inductive coupling between load and line. Both the hot and neutral wires are connected to the isolating transformer. There is no continuous wire connected in between.

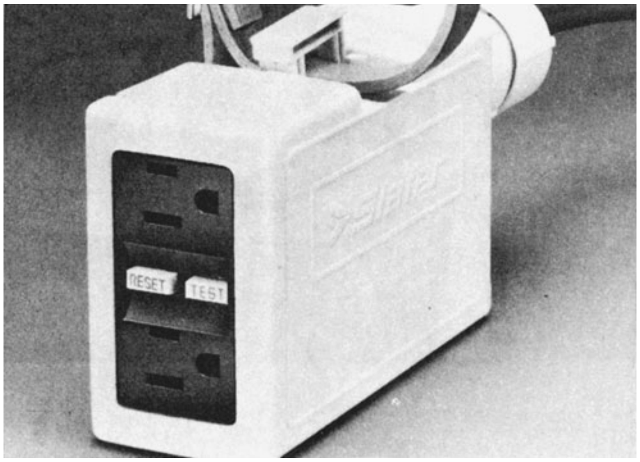

Figure 13-4 Ground-fault circuit interrupter (GFCI) used for portable tools.

In the latter type of interrupter, a ground fault must pass through the electronic sensing circuit, which has sufficient resistance to limit current flow to as low as 2 milliamperes, well below the level of human perception.

Types of Circuit Protectors

There are two types of circuit protectors made in the circuit breaker configuration. They are circuit breakers that work on the heating principle. A bimetallic strip is heated by having circuit current pass through it. When too much current flows, the strip is overheated. The expansion of the strip causes the breaker to trip, thereby causing the circuit to be opened (see Fig. 13-5). This type is slower to cool down than the magnetic type. It takes a little longer for it to cool to a point where it can be reset. If the overload still exists, the breaker will trip again.

The magnetic type uses a coil to operate. The circuit current is drawn through the coil. If too much current flows, the magnetic properties of the coil cause the circuit breaker to trip and open the circuit. This type is quicker in terms of being able to reset it. The overload must, of course, be removed before it will remain in the reset position.

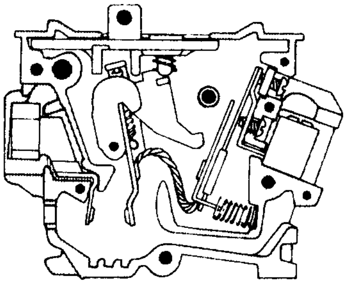

Tripped position
- Contacts open — no current flow
- Handle stationary when tripped

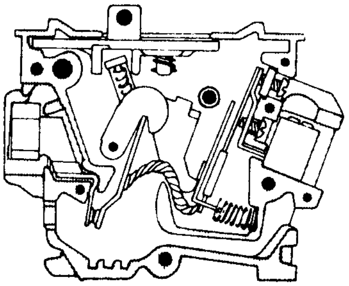

"On" position
- Contacts closed — current on
- Handle in "On" position (shows "On")

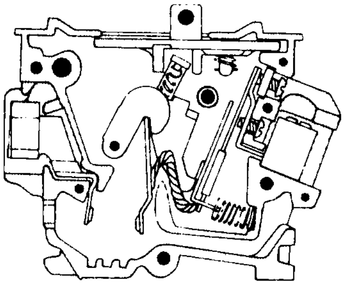

"Off" position
- Contacts open — no current flow
- Handle in "Off" position (shows "Off")

To restore service when fault is cleared you simply move operating handle to "Off" position and then to "On."

Figure 13-5 Cutaway view of a circuit breaker in various positions. (Courtesy of Wadsworth)

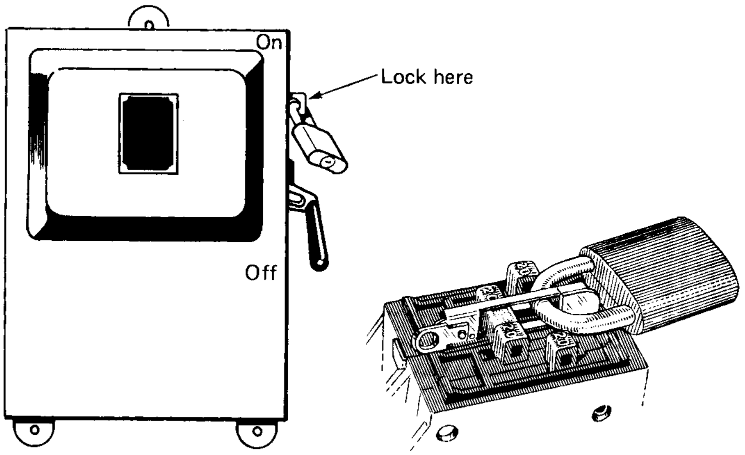

Figure 13-6 Lock the box by using a lock and keep the key until you need to unlock it.

In most cases where the power is 240 volts single-phase, the circuit breakers will be locked together with a pin. If one side of the power circuit is opened, the other circuit is also tripped since they are tied together physically.

In three-phase circuits, the three circuit breakers are tied together so that when one trips, all three circuits are disconnected from the power line.

If air-conditioning or refrigeration equipment is served by a separate box, make sure the lock is in place and the handle is in the down or off position before making an inspection or working on the equipment (see Fig. 13-6).

Review Questions

1. How much current does it take to kill?
2. What precautions should you take when using portable electric hand tools?
3. What is a ground-fault circuit interrupter?
4. Where is the GFCI used?
5. What is the color of the ground wire in a two-wire system?
6. What is the other color used for ground in an industrial or commercial installation?
7. Why should you lock the off switch when working with air-conditioning, refrigeration, or heating equipment?

This page intentionally left blank

Chapter 14

Control Devices

Performance Objectives

- Know how power relays work.
- Know how time delay relays work.
- Know the most common types of thermostats.
- Know the advantages of microprocessor thermostats.
- Know how a cold anticipator works and how to adjust it.
- Know how a heat anticipator works and how to adjust it.
- Know how to adjust a limit switch.
- Know how water tower controls operate.
- Know how pressure control switches operate.

Many types of controls are available for use on air-conditioning, refrigeration, and heating equipment. They come in many sizes and shapes and do the job well for a period of time, but they all require periodic inspections, repairs, and replacement.

Power Relays

One of these controls is the power relay. It is one of the most often used controls for controlling compressors in refrigeration and air conditioning. The power relay is also referred to as the *main conductor*. It is used to apply the main line voltage to the motor circuit. The coil of the relay is usually operated by voltages lower than the line provides. This means that it uses a transformer for the lower control voltages. The symbols used for this type of relay are shown in Fig. 14-1.

Magnetic contactors are normally used for starting polyphase motors, either squirrel cage or single phase. Contactors may be connected at any convenient point in the main circuit between the fuses and the motor. Small control wires (using low voltage) may be run between the contactor and the point of control.

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

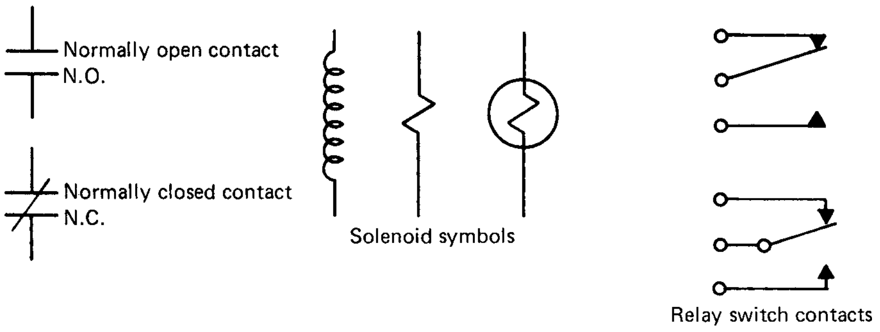
Normally open contact
N.O.
Normally closed contact
N.C.
Solenoid symbols
Relay switch contacts

Figure 14-1 Symbols for the main contactor or power relay.

Motor start relays

Relays are a necessary part of many control and pilot-light circuits. They are similar in design to contactors, but are generally lighter in construction so they carry smaller currents.

Compressors used for household refrigerators, freezers, dehumidifiers, vending machines, and water coolers have the capacitor-start, induction-run type of motor. This type of compressor may have a circuit that resembles Fig. 14-2. When the compressor is turned on by the thermostat demanding action, the relay is closed and the start winding is in the circuit. Once the motor comes up to about 75% of rated speed, there is enough current flow through the relay coil to cause it to energize, and it pulls the contacts of the relay open, thereby taking the start capacitor and start winding out of the circuit. This allows the motor to run with one winding as designed.

Figure 14-3 shows the current type of relay. This is generally used with small refrigeration compressors up to $^3/_4$ horsepower. Figure 14-4 shows the potential

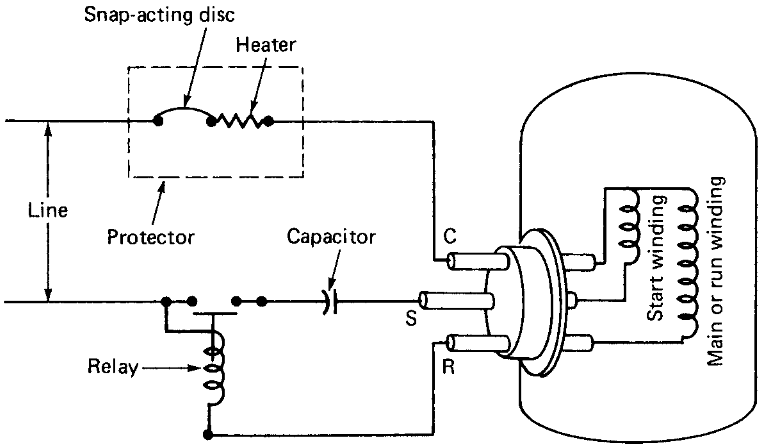
Snap-acting disc
Heater
Line
Protector
Capacitor
C
S
R
Relay
Start winding
Main or run winding

Figure 14-2 Capacitor-start, induction-run motor for a compressor with the potential relay used to take out the start winding once the motor comes up to speed.

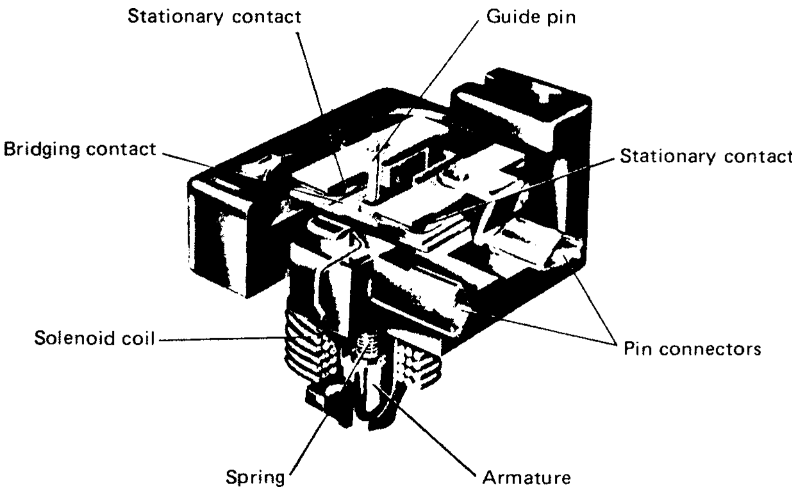
Stationary contact
Guide pin
Bridging contact
Stationary contact
Solenoid coil
Pin connectors
Spring
Armature

Figure 14-3 Current relay. (Courtesy of Tecumseh)

type of relay. This is generally used with large commercial and air-conditioning compressors up to 5 horsepower.

Protection of the motor against prolonged overload is accomplished by time limit overload relays. They are operative during the starting period and running period. Relay action is delayed long enough to take care of the heavy starting currents and momentary overloads without tripping.

Relays with more than one contact

Some power relays are made with more than one set of contacts. They are used to cause a sequence of events to take place. The contacts can be wired into a

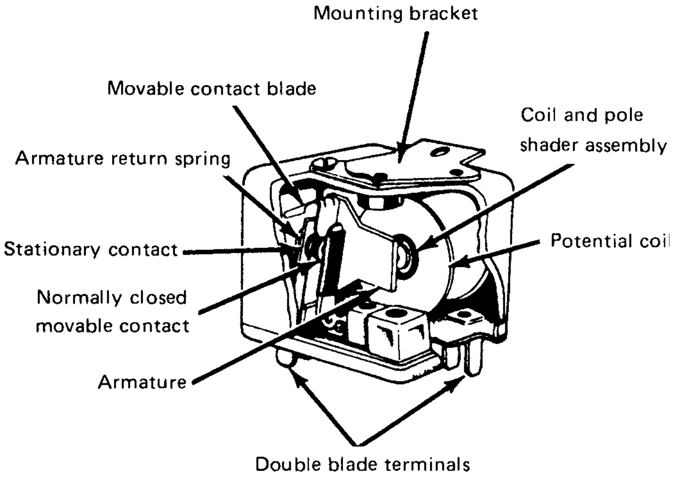
Mounting bracket
Movable contact blade
Coil and pole shader assembly
Armature return spring
Stationary contact
Potential coil
Normally closed movable contact
Armature
Double blade terminals

Figure 14-4 Potential relay. (Courtesy of Tecumseh)

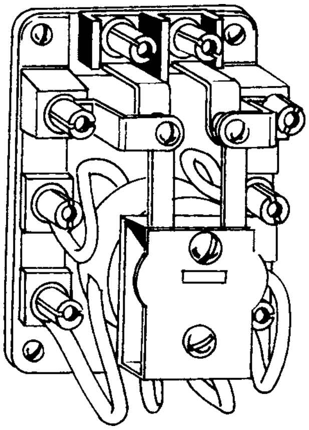

Figure 14-5 Relay with more than one set of contacts.

circuit that controls functions other than the on-off operation of the compressor motor (see Fig. 14-5).

Thermal overload protectors

Motors for commercial units are protected by a bimetallic switch. The switch is operated on the heat principle. This is a built-in motor overload protector (see Fig. 14-6). It limits the motor winding temperature to a safe value. In its simplest form, the switch or motor protector consists essentially of a bimetal switch mechanism that is permanently mounted and connected in series with the motor circuit (see Fig. 14-7). Figure 14-8 shows how the external line-break overload operates.

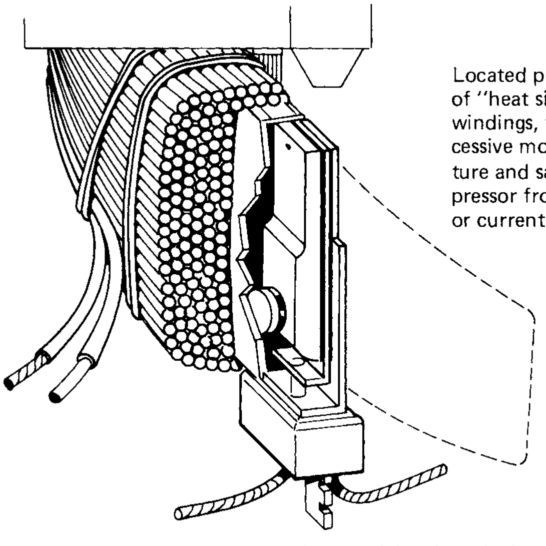

Located precisely in the center of "heat sink" portion of motor windings, this device detects excessive motor winding temperature and safely protects compressor from excessive heat and/or current draw.

Figure 14-6 Motor protector inserted in the windings of the compressor. (Courtesy of Tecumseh)

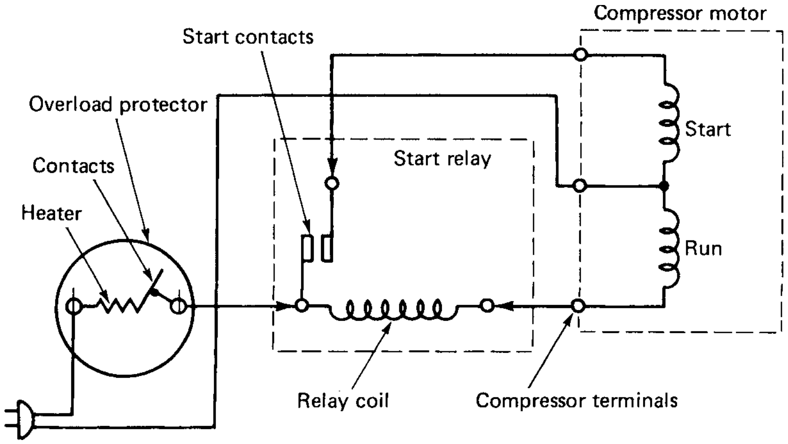
Compressor motor
Start contacts
Overload protector
Contacts
Heater
Start relay
Start
Run
Relay coil
Compressor terminals

Figure 14-7 Domestic refrigerator circuit showing the start contacts and relay coil, as well as the overload protector.

Time-delay relays

In time-delay relays, bimetallic strips are heated with an electrical resistance mounted near or around them. The strips expand when heated. When they expand, they make contact and complete the circuit with their contacts closed (see Fig. 14-9). The time delay can be adjusted by the resistance of the heater unit. This type of unit is different from that shown as a protector in Fig. 14-7.

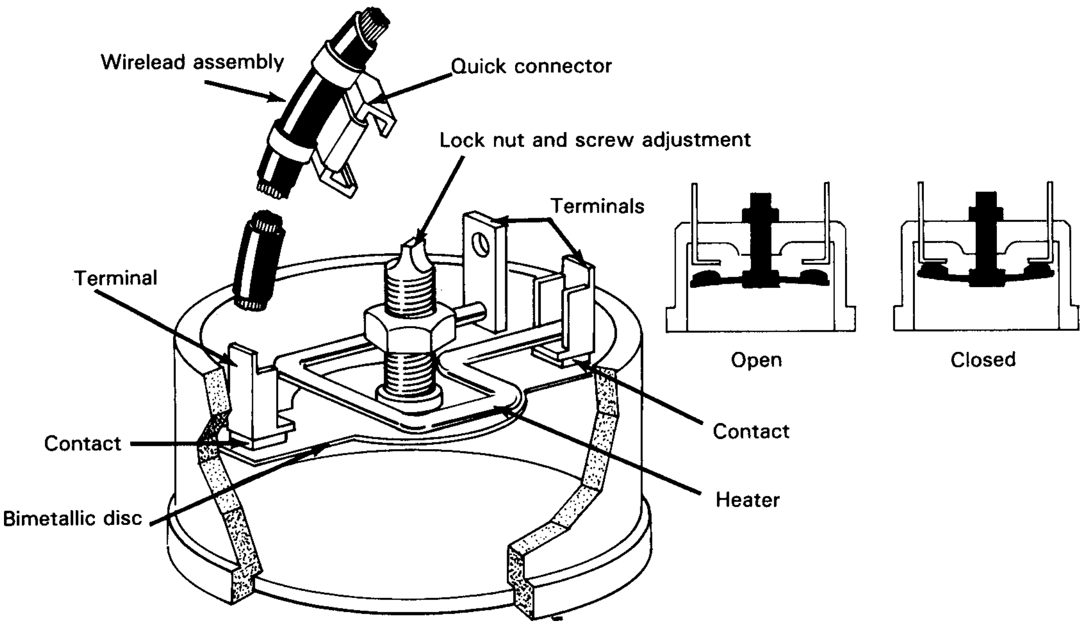
Wirelead assembly
Quick connector
Lock nut and screw adjustment
Terminals
Terminal
Contact
Bimetallic disc
Contact
Heater
Open
Closed

Figure 14-8 Externally located line-break overload. (Courtesy of Tecumseh)

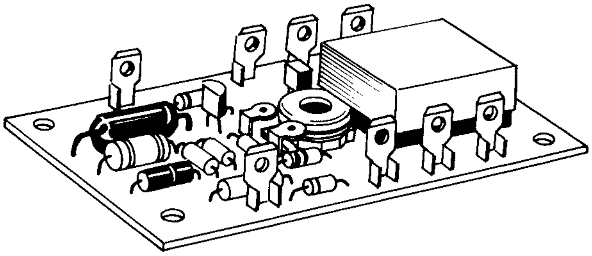

Figure 14-9 Time-delay relay.

The heating element in Fig. 14-7 causes the circuit to open and protect the motor. The time-delay relay is used to make sure that certain things take place within the refrigeration cycle before another is commenced.

Solenoids

Solenoid valves are used in many heating and cooling applications. They are electrically operated. A solenoid valve, when connected as in Fig. 14-10, remains open when current is supplied to it. It closes when the current is turned off. In general, solenoid valves are used to control the liquid refrigerant flow into the expansion valve or the refrigerant gas flow from the evaporator when it or the fixture it is controlling reaches the desired temperature. The most common application of the solenoid valve is in the liquid line, and it operates with a thermostat (see Fig. 14-11).

The solenoid shown in Fig. 14-12 controls the flow of natural gas in a hot-air furnace. Note how the coil is wound around the plunger. The plunger is the core of the solenoid. It has a tendency to be sucked into the coil whenever the coil is

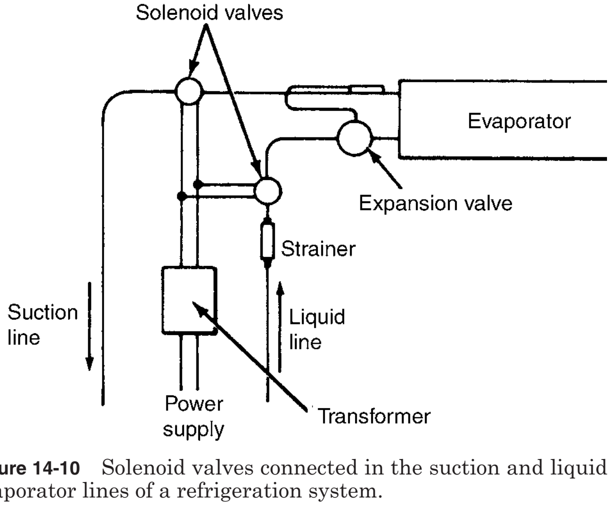

Figure 14-10 Solenoid valves connected in the suction and liquid evaporator lines of a refrigeration system.

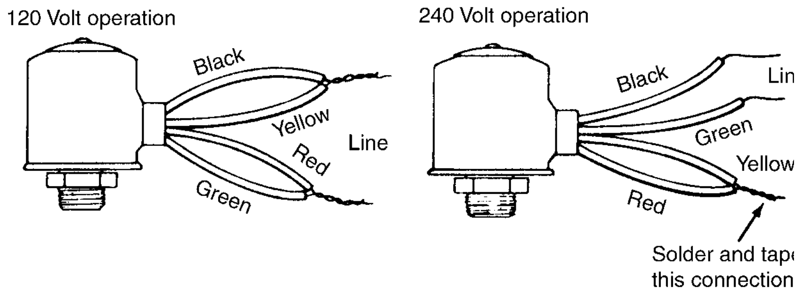

Figure 14-11 Solenoid valves. Note color-coded wires. (Courtesy of General Controls)

energized by current flowing through it. The electromagnetic effect causes the plunger to be attracted upward into the coil area. When the plunger is moved upward by the pull of the electromagnet, the soft disc is pulled upward, allowing gas to flow through the valve. This basic technique is used to control water, gasoline, oil, or any other liquid or gas.

Thermostats

Temperature control by using thermostats is common to both heating and cooling equipment. Thermostats are used to control heating circuits that cause furnaces and boilers to operate and provide heat. Thermostats are also used to control cooling equipment and refrigeration units. Each of these purposes may have its own specially designed thermostat or may use the same one. For instance, in the home you use the same thermostat to control the furnace and the air-conditioning unit.

Bellows-type thermostats

On modern condensing units, low-pressure control switches are largely superseded by thermostatic control switches. A thermostatic control consists of three main parts: a bulb, a capillary tube, and a power element or switch. The bulb is attached to the evaporator in a manner that assures contact with the evaporator. It may contain a volatile liquid, such as a refrigerant. The bulb is connected to the power element by means of a small capillary tube (see Fig. 14-13).

Operation of the bellows is provided by a change in temperature. Or the operation of the thermostatic control switch is such that, as the evaporator temperature increases, the bulb temperature also increases. This raises the pressure of the thermostatic liquid vapor. This, in turn, causes the bellows to expand and actuate an electrical contact. The contact closes the motor circuit, and the motor and compressor start operating. As the evaporator temperature decreases, the bulb becomes colder and the pressure decreases to the point where the bellows contracts sufficiently to open the electrical contacts, thus turning off the motor circuits. In this manner, the condensing unit is entirely automatic. Thus, it is

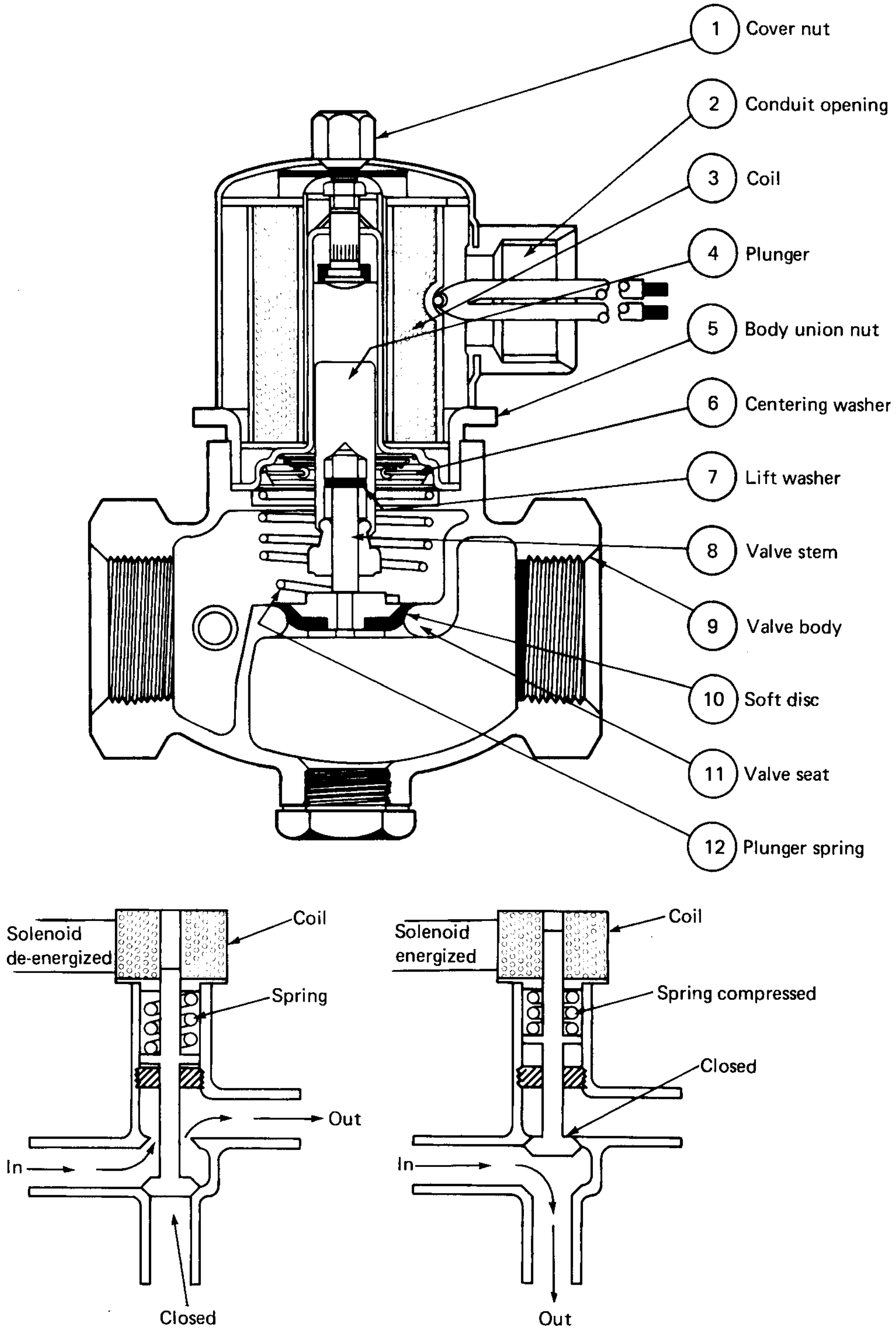
1 Cover nut
2 Conduit opening
3 Coil
4 Plunger
5 Body union nut
6 Centering washer
7 Lift washer
8 Valve stem
9 Valve body
10 Soft disc
11 Valve seat
12 Plunger spring
Solenoid de-energized
Coil
Spring
Out
In
Closed
Solenoid energized
Coil
Spring compressed
Closed
In
Out

Figure 14-12 Solenoid used for controlling natural gas flow to a furnace. (Courtesy of Honeywell)

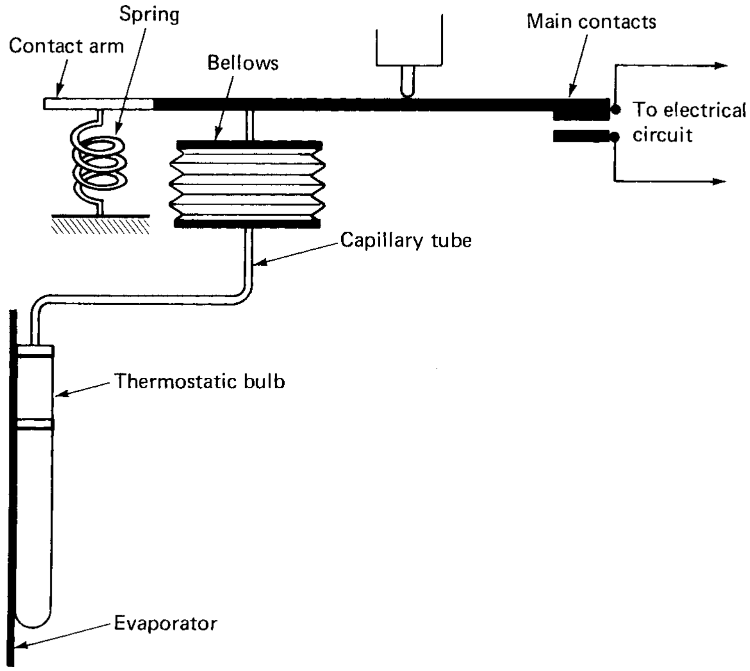

Figure 14-13 Bellows-type switch.

able to produce exactly the amount of refrigeration needed to meet any normal operating condition.

Bimetallic-type thermostats

Temperature changes can cause a bimetallic strip to expand or contract in step with changes in temperature. These thermostats are designed for the control of heating and cooling in air-conditioning units, refrigeration storage rooms, greenhouses, fan coils, blast coils, and similar units. This is the type used in most homes for control of the central air-conditioning and central heating system.

Figure 14-14 shows how the bimetallic strip thermostat works. Two metals, each having a different coefficient of expansion, are welded together to form a bimetallic unit or blade. With the blade securely anchored at one end, a circuit is formed and the contact points are closed. This allows the passage of an electric current through the closed points. Because an electric current provides heat in its passage through the bimetallic blade, the metals in the blade begin to expand. However, they expand at a different rate. The metals in the blade are so arranged that the one with a greater coefficient of expansion is placed at the bottom of the unit. After a certain time, the operating temperature is reached and the contact points become separated. This disconnects the device from its power source.

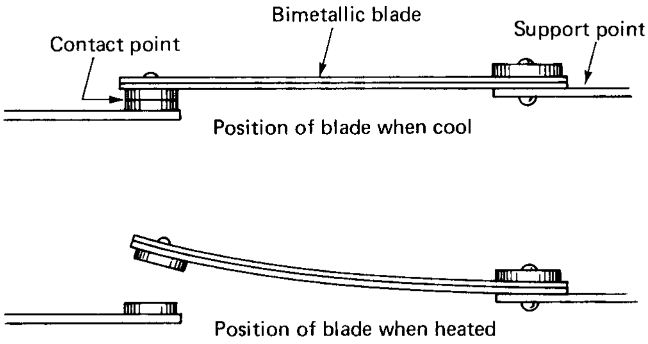
Bimetallic blade
Contact point
Support point
Position of blade when cool
Position of blade when heated

Figure 14-14 Bimetallic strip used in a thermostat.

After a short period, the contact blade will again become sufficiently cooled to cause the contact point to join, thus reestablishing the circuit and permitting the current again to actuate the circuit. The cycle is repeated over and over again. In this way, the bimetallic thermostat prevents the temperature from rising too high or dropping too low.

Heating and cooling thermostats

Some thermostats can be used for both heating and cooling. The thermostat shown in Fig. 14-15 is such a device. The basic thermostat element has a permanently sealed, magnetic SPDT switch. The thermostat element plugs into the subbase and contains the heat anticipation, the magnetic switching, and a room

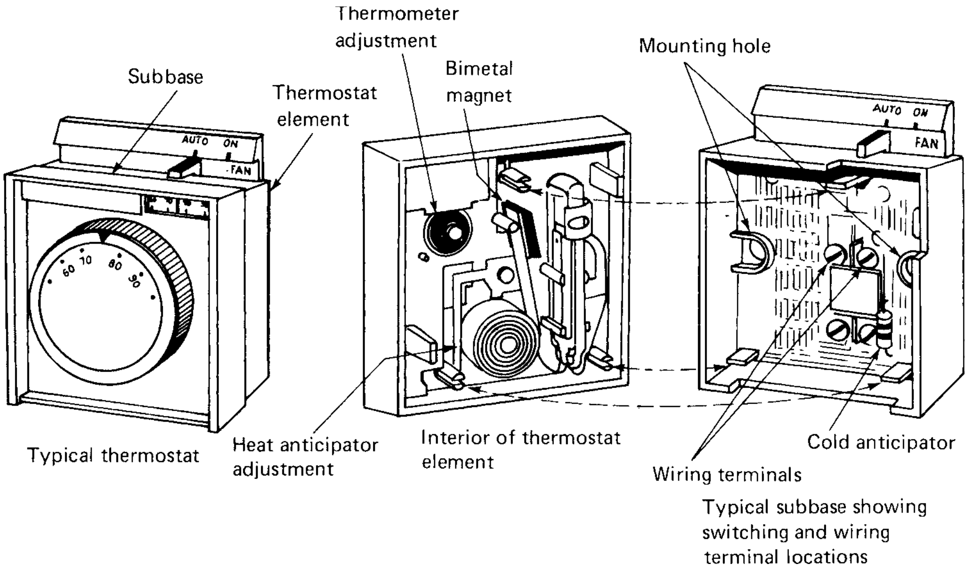
Thermometer adjustment
Mounting hole
Subbase
Thermostat element
Bimetal magnet
AUTO ON FAN
60 70 80 90
Typical thermostat
Heat anticipator adjustment
Interior of thermostat element
Cold anticipator
Wiring terminals
Typical subbase showing switching and wiring terminal locations

Figure 14-15 Modern thermostat for heating and cooling of a house. (Courtesy of General Controls)

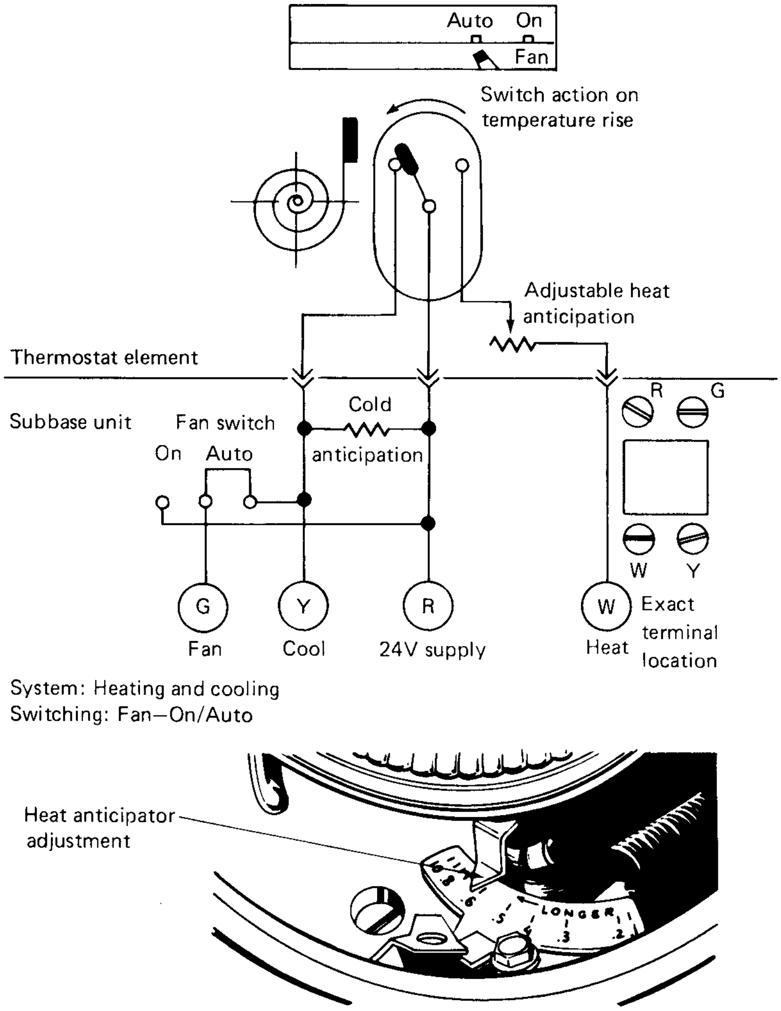
Auto
On
Fan
Switch action on temperature rise
Adjustable heat anticipation
Thermostat element
Subbase unit
Fan switch
On
Auto
Cold
anticipation
R
G
W
Y
G
Fan
Y
Cool
R
24V supply
W
Heat
Exact terminal location
System: Heating and cooling
Switching: Fan—On/Auto
Heat anticipator adjustment
LONGER

Figure 14-16 Wiring diagram for a thermostat.

temperature thermometer. The subbase unit contains fixed cool anticipation and circuitry. This thermostat is used with 24 volts ac. In this case, the thermostatic element (bimetal) does not make direct contact with the electric circuit. Instead, the expansion of the bimetal causes a magnet to move. This, in turn, causes the switch to close or open. Figure 14-16 shows that the bimetal is not in the electric circuit.

Mercury contacts. Some thermostats use the expanding bimetal arrangement to cause a tube of mercury to move. As the mercury moves in the tube, it comes in contact with two wires inserted into the glass tube. When the mercury comes in contact with the two wires, it completes the electric circuit. This type of

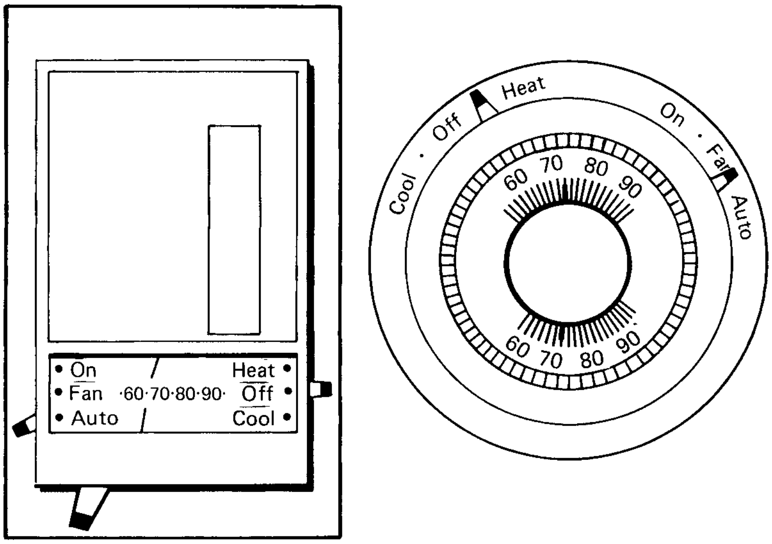
On
Heat
Fan
·60·70·80·90·
Off
Auto
Cool
Heat
On · Fan
Auto
Cool · Off
60 70 80 90
60 70 80 90

Figure 14-17 Thermostat for air conditioning and heating.

thermostat needs to be so arranged that the tube of mercury is pivoted and can be moved by the expanding or contracting bimetal strip, which exerts or releases pressure on the tube of mercury.

Thermostats used in home air-conditioning and heating systems are now equipped with mercury contacts (see Fig. 14-17). They are made so that the mercury contacts two wires that control the air conditioning in one position and two wires that control the heating system in the other position (see Fig. 14-18).

The advantage of the mercury bulb type of switch is the elimination of switch contact points. Contact points are in need of constant attention. In most cases,

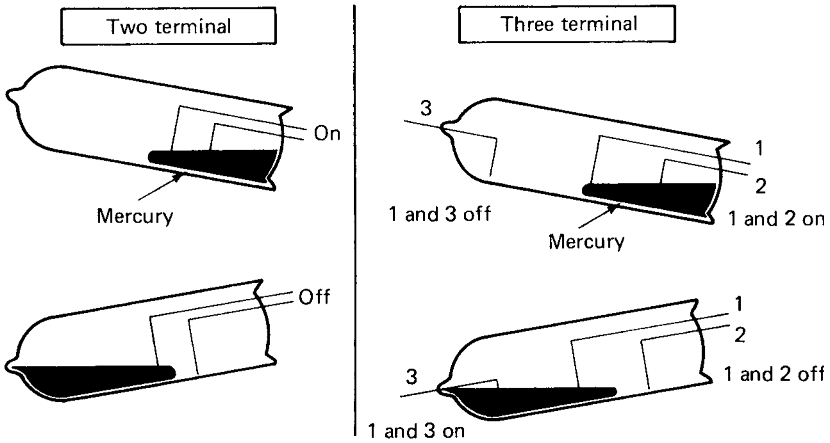
Two terminal
On
Mercury
Off
Three terminal
3
1
2
1 and 3 off
1 and 2 on
Mercury
1
2
3
1 and 2 off
1 and 3 on

Figure 14-18 Mercury-switch operation.

the dust from the air will eventually cause them to function improperly. It is necessary to clean the points by running a piece of clean paper through them to remove the dust particles and arcing residue. Since the mercury type is sealed and the arcing created on make and break of the circuit simply causes the mercury to vaporize slightly and then return to a liquid state, it provides a trouble-free switching operation.

Microprocessor thermostats

Semiconductor technology has produced another means of more accurately controlling air-conditioning and heating systems to provide better regulated temperatures in the home, office, and business. The microprocessor makes use of the semiconductor chip or integrated circuit discussed earlier in Chapter 11. All the external connections are the same as for any other type of thermostat. Only the internal circuitry has changed to provide a better regulated temperature and a variety of operations that allow you to set it for any energy-saving program desired (see Fig. 14-19). Unless a battery is included, it does not retain the program in most instances, and the clock, if there is one on the unit, has to be reset each time the power goes off.

Thermostat Adjustments

In Fig. 14-16, a cold anticipation and a heat anticipation adjustment are placed in the thermostat circuit. The heat anticipation control is placed in series with the switch. The cold anticipation resistor is placed in shunt or parallel with the switch. Thus, when the switch is closed the shunt is shorted out.

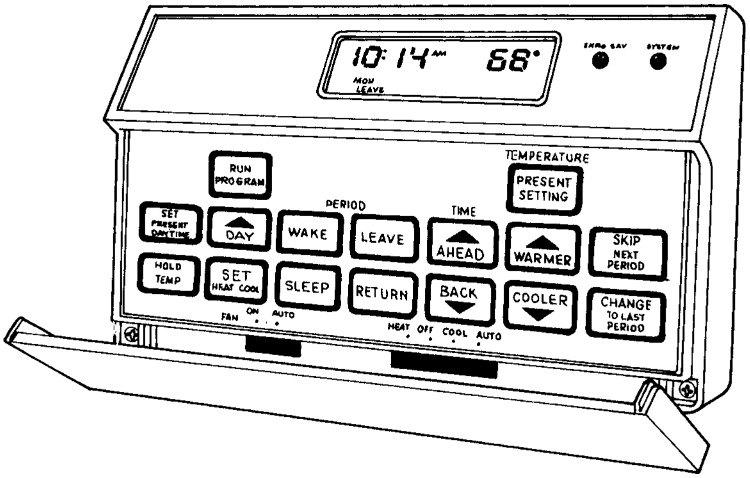
10:14 AM
68°
MON
LEAVE
SYSTEM
RUN PROGRAM
TEMPERATURE
PRESENT SETTING
PERIOD
TIME
SET PRESENT DAY/TIME
DAY
WAKE
LEAVE
AHEAD
WARMER
SKIP NEXT PERIOD
HOLD TEMP
SET HEAT COOL
SLEEP
RETURN
BACK
COOLER
CHANGE TO LAST PERIOD
FAN
ON
AUTO
HEAT
OFF
COOL
AUTO

Figure 14-19 Microprocessor used for air-conditioning and furnace control in a home.

Heat anticipators

The reason for the heat anticipator is to limit the degree of swing between turning on the furnace and the temperature of the room. It is a resistance heater element that is inserted in series with the thermostat line that runs to the heat contactor coil. Once the thermostat contacts are closed, current flows through the resistor. This causes it to heat up. The heat generated by the resistor causes the thermostat to open slightly before the desired room temperature is reached by the heating system. This allows the heat in the plenum of the furnace to continue to heat the room. Thus, the resistor aids the thermostat in anticipating the amount of heat that will be provided to the room by using the heat already produced in the plenum.

Cold anticipators

The cold anticipator is a fixed resistor and is not adjustable. It heats the bimetallic coil that operates the points whenever the air-conditioner compressor is not on. When the compressor is on, the resistor is shorted out by the thermostat points being closed. The heating of the coil while the points are open causes it to close a little earlier than if it waited for the room to heat up sufficiently to cause it to turn on. This way the heat produced by the anticipator resistor causes the compressor to turn on a little before the thermostat would have normally told it to do so. By turning it on before the room has reached the selected temperature, the anticipator causes the temperature swing in the room to be reduced and makes it more comfortable.

Switches of Many Types

Many types of switches are used to limit the amount of heat produced in a furnace. The upper limit has to be controlled so that the furnace does not cause fires by overheating. Limit switches take various forms depending on the manufacturer. However, Fig. 14-20 shows a typical switch and how it works. This is a combination of fan and limit controller that combines the functions of a fan controller and a limit controller in a single unit. One sensing element is used for both controls.

Combination controllers are wired in much the same way as individual controls. These combined controls can be used on line voltage, low voltage, or self-energizing millivolt systems.

Figure 14-21 shows the fluid-filled type of capillary tube used in a limit switch. The one shown in Fig. 14-20 is the bimetal type that twists as it heats up, causing the control unit to move. These limit switches are placed in the plenum of the furnace to control when the fan goes on and off; when the plenum has reached the desired temperature, it turns off the solenoid and shuts off the flow of natural gas to the burner. Limit switches of a slightly different configuration are also used for electrical strip heaters. They may also be of the low voltage (24 volts) or line voltage type.

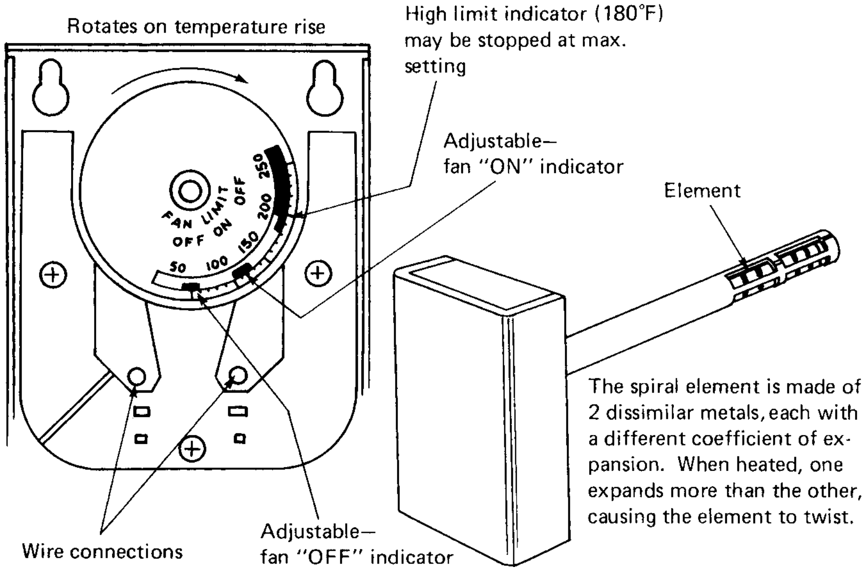
Rotates on temperature rise
High limit indicator (180°F) may be stopped at max. setting
Adjustable— fan "ON" indicator
Element
FAN LIMIT
OFF ON OFF
50 100 150 200 250
The spiral element is made of 2 dissimilar metals, each with a different coefficient of expansion. When heated, one expands more than the other, causing the element to twist.
Wire connections
Adjustable— fan "OFF" indicator

Figure 14-20 Combination fan and limit controller (fluid filled type).

Pressure Control Switches

One safety feature for air-conditioning units with a compressor and condenser is a pressure-controlled switch. This switch is wired into the circuit to protect the system in case the system develops a leak. If a leak develops, it is possible to draw in moisture and air and damage the whole system. If the pressure

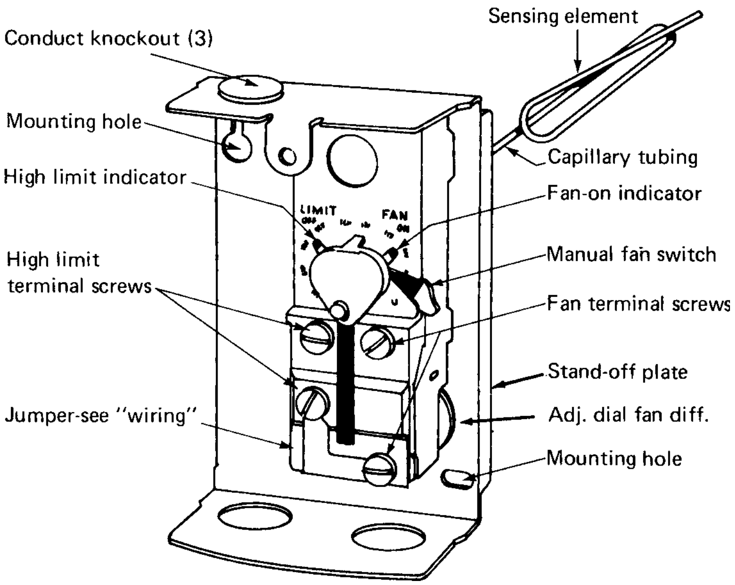
Sensing element
Conduct knockout (3)
Mounting hole
Capillary tubing
High limit indicator
Fan-on indicator
LIMIT
FAN
High limit terminal screws
Manual fan switch
Fan terminal screws
Stand-off plate
Jumper-see "wiring"
Adj. dial fan diff.
Mounting hole

Figure 14-21 Combination fan and limit controller (bimetal type).

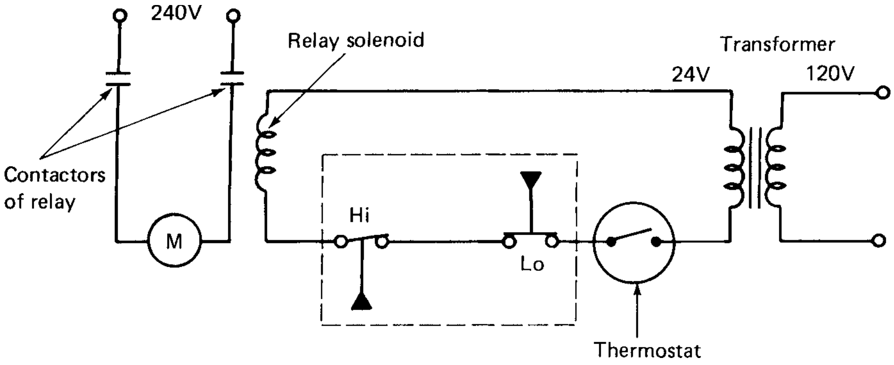

Figure 14-22 Pressure-operated switches control the compressor.

builds too high, it can cause a rupture of any of the joints or weaker points in the system.

A low-voltage (24-volt) relay is wired into the 240-volt line that supplies the compressor motor. The relay contacts are wired into the supply line for the motor (see Fig. 14-22). The solenoid of the relay is wired in series with two pressure-operated switches. If the pressure builds too high, the high switch will open and cause the solenoid to de-energize. If this happens, it causes the contacts of the relay to open. This removes power from the compressor motor. If the low-pressure switch opens, it will do the same thing. This way the compressor is protected from both high and low pressure, causing damage to the system.

Both manual and automatic controls are available. Automatic controls reset when the pressure is stabilized in the system. If it is not stabilized, it will again turn the system off and keep recycling until it reaches the design pressure.

Water Tower Controls

Temperature controls for refrigerating service are designed to maintain adequate head pressure with evaporative condensers and cooling towers. Low refrigerant head pressure, caused by abnormally low cooling water temperature, reduces the capacity of the refrigeration system.

Two systems of control for mechanical and atmospheric draft towers and evaporative condensers are shown in Figs. 14-23 and 14-24. The control opens the contacts when the temperature drops. These contacts are wired in series with the fan motor. Or they can be wired to the pilot of a fan-motor controller. Opening the contacts stops the fan when the cooling water temperature falls to a predetermined minimum value. This value corresponds to the minimum head pressure for proper operation. In the control system shown in Fig. 14-24, the contacts close on a temperature drop and are wired in series with a normally closed motorized valve or a solenoid valve. The contacts open the valve when low cooling temperature occurs. The cooling water then flows through a low header in the atmospheric tower. This reduces its cooling effect and the head pressure increases.

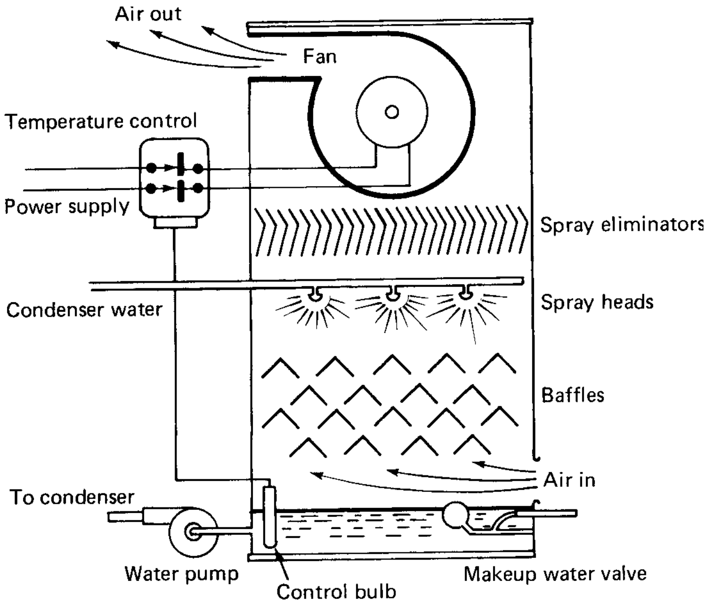
Air out
Fan
Temperature control
Power supply
Spray eliminators
Condenser water
Spray heads
Baffles
Air in
To condenser
Water pump
Control bulb
Makeup water valve

Figure 14-23 Cooling tower with forced-air draft.

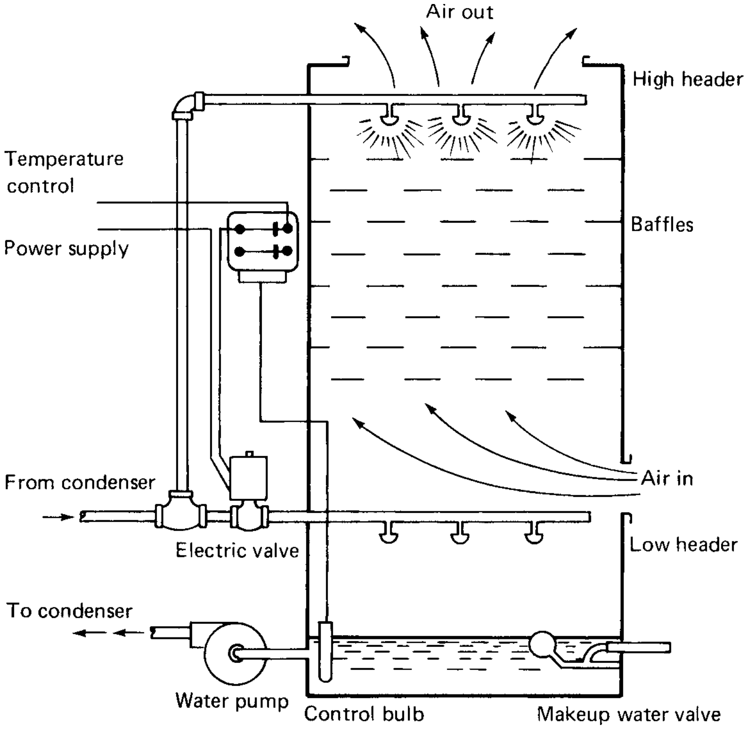
Air out
High header
Temperature control
Power supply
Baffles
Air in
From condenser
Electric valve
Low header
To condenser
Water pump
Control bulb
Makeup water valve

Figure 14-24 Cooling tower with atmospheric draft cooling.

Float switches are used to control the level of water in the cooling tower. Automatic float switches provide automatic control for motors operating tank or sump pumps. They are built in several styles and can be supplied with several types of accessories that provide rod or chain operation and either wall or floor mounting. A sensor system may also be used. There are hundreds of sensor types. They usually sense the level of water by using two probes. When the water contacts the probes, it causes a small electric current (at low voltages) to flow and energize a solenoid or relay that in turn causes the water to be turned off. When the level of water is below the two probes and a complete circuit is not available, the normally closed relay contacts are closed by de-energizing the relay. This causes the water solenoid to be energized. This allows makeup water to flow into the cooling tower until it reaches the point where the probes are immersed in water and the cycle is repeated.

Review Questions

1. What is another name for a power relay?
2. What is the function of the power relay?
3. What is the time-delay relay used for?
4. What is the most common type of thermostat used on air conditioners?
5. Why are mercury contacts better?
6. What is the advantage of a microprocessor thermostat?
7. What is a heat anticipator?
8. Why is a heat anticipator needed?
9. How does the heat anticipator work?
10. What is a limit switch?
11. How does the limit switch work?
12. Where would you find a limit switch?

Chapter 15

Heating Circuits

Performance Objectives

- Know the parts of a basic electrical heating system.
- Know how to read a ladder diagram.
- Know the difference between field wiring and factory wiring.
- Know how to trace the low voltage system in a heating unit.
- Know how to identify parts on a printed circuit board for a furnace.
- Know how a heat pump works.
- Know how a refrigerator can be used to generate heat.
- Know why supplemental heaters are sometimes needed.
- Know how the gas/air mixture is ignited in a Pulse furnace.
- Understand why the Pulse furnace is so efficient.
- Know how to troubleshoot the Pulse furnace using the flowchart provided by the manufacturer.

Hot-air furnaces are self-contained and self-enclosed units. They are usually centrally located within a building or house. Their purpose is to make sure the temperature of the interior of the structure is maintained at a comfortable level throughout. The design of the furnace is determined by the type of fuel used to fire it. Cool air enters the furnace and is heated as it comes in contact with the hot-metal heating surfaces. As the air becomes warmer, it also becomes lighter, which causes it to rise. The warmer, lighter air continues to rise until it is either discharged directly into a room, as in the pipeless gravity system, or is carried through a duct system to warm-air outlets located at some distance from the furnace.

After the hot air loses its heat, it becomes cooler and heavier. Its increased weight causes it to fall back to the furnace, where it is reheated and repeats the

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

cycle. This is a very simplified description of the operating principles involved in hot-air heating and it is especially typical of those involved in gravity heating systems. The forced-air system relies on a blower to make sure the air is delivered to its intended location. The blower also causes the return air to move back to the furnace faster than with the gravity system.

With the addition of a blower to the system, there must be some way of turning the blower on when needed to move the air and to turn it off when the room has reached the desired temperature. Thus, electrical controls are needed to control the blower action.

Basic Gas Furnace Operation

The gas furnace is the simplest to operate and understand. Therefore, we will use it here to look at a typical heating system. This type of natural-gas furnace is used to heat millions of homes in the United States.

Figure 15-1 is a simple circuit needed to control the furnace with a blower. Note the location of the blower switch and the limit switch. The transformer provides low voltage for control of the gas solenoid. If the limit switch opens (it is shown in a closed position), there is no power to the transformer and the gas solenoid cannot energize. This is a safety precaution because the limit switch will open if the furnace gets too hot. When the thermostat closes, it provides 24 volts to the gas solenoid, which energizes and turns on the gas. The gas is ignited by the pilot light and provides heat to the plenum of the furnace. When the air in the plenum reaches 120°F, the fan switch closes and the fan starts. The fan switch provides the necessary 120 volts to the fan motor for it to operate.

After the room has heated up to the desired thermostat setting, the thermostat opens. When it opens, the gas solenoid is de-energized, and the spring

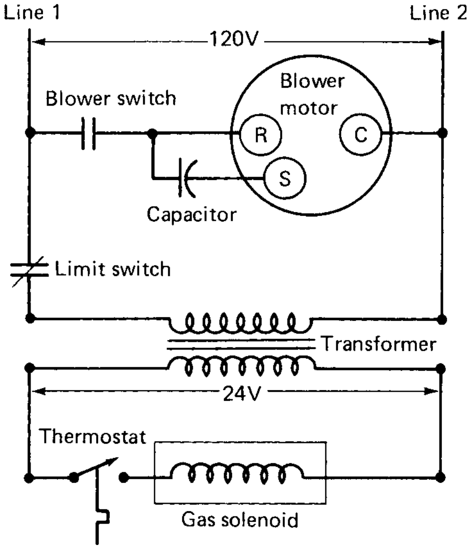

Figure 15-1 Simple one-stage furnace control system.

action of the solenoid causes it to close off the gas supply, thereby turning off the source of heat. When the plenum on top of the furnace reaches 90°F, the blower switch opens and turns off the blower. As the room cools down, causing the thermostat to once again close, the cycle starts over again. The gas solenoid opens to let in the gas and the pilot light ignites it. The heat causes the temperature to rise in the plenum above the limit switch's setting and the switch closes to start the blower. Once the thermostat has been satisfied, it opens, and causes the gas solenoid to turn off the gas supply. The blower continues to run until the temperature in the plenum reaches 90°F and it turns off the blower by opening. This cycle is repeated over and over again to keep the room or house at a desired temperature.

Basic Electric Heating System

Electric-fired heat is the only heat produced almost as fast as the thermostat calls for it. It is almost instantaneous. There are no heat exchangers to warm up. The heating elements start producing heat the moment the thermostat calls for it. A number of types of electric-fired furnaces are available. They can be bought in 5- to 35-kilowatt sizes. The outside looks almost the same as the gas-fired furnace. The heating elements are located where the heat exchangers would normally be located. Since they draw high amperage, they need electrical controls that can take the high currents.

The operating principle is simple. The temperature selector on the thermostat is set for the desired temperature. When the temperature in the room falls below this setting, the thermostat calls for heat and causes the first heating circuit in the furnace to be turned on. There is generally a delay of about 15 seconds before the furnace blower starts. This prevents the blower from circulating cool air in the winter. After about 30 seconds, the second heating circuit is turned on. The other circuits are turned on one by one in a timed sequence.

Once the temperature reaches the desired level, the thermostat opens. After a short time, the first heating circuit is shut off. The others are shut off one by one in a timed sequence. The blower continues to operate until the air temperature in the furnace drops below a specified temperature.

Basic operation

In Fig. 15-2, the electrical heating system has a few more controls than the basic gas-fired furnace. The low-resistance element used for heating draws a lot of current, so the main contacts have to be of sufficient size to handle the current.

The thermostat closes and completes the circuit to the heating sequencer coil. The sequencer coil heats the bimetal strip that causes the main contacts to close. Once the main contacts are closed, the heating element is in the circuit and across the 240-volt line. The auxiliary contacts will also close at the same time as the main contacts. When the auxiliary contacts close, they complete the low-voltage circuit to the fan relay. The furnace fan will be turned on at this time.

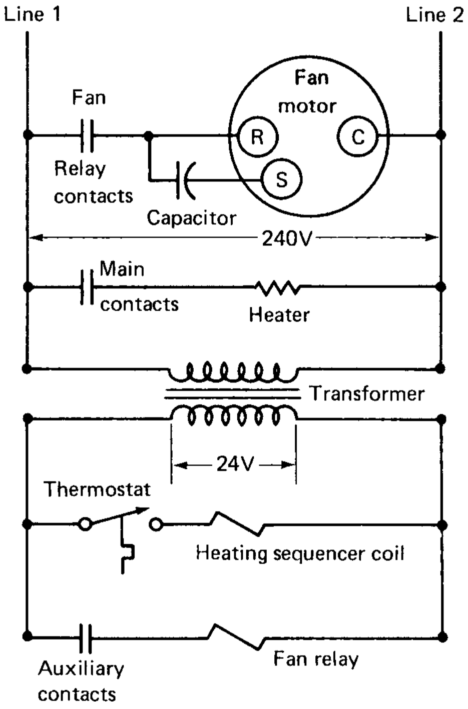
Line 1
Line 2
Fan
Fan motor
R
C
S
Relay contacts
Capacitor
240V
Main contacts
Heater
Transformer
24V
Thermostat
Heating sequencer coil
Auxiliary contacts
Fan relay

Figure 15-2 Ladder diagram for a hot-air furnace.

Once the thermostat has been satisfied, it opens. This allows the heating sequencer coil to cool down slowly. Thus, the main contacts do not open immediately to remove the heating element from the line. So the furnace continues to produce heat after the thermostat has been satisfied. The bimetal cools down in about 2 minutes. Once it cools, it opens the main and auxiliary contacts, which removes the heating element from the line and also stops the fan motor. After the room cools down below the thermostat setting, the thermostat closes and starts the sequence all over again.

Ladder Diagrams

Electrical schematics are used to make it simple to trace the circuits of various devices. Some of these can appear complicated, but they are usually very simple when you start at the beginning and wind up at the end. The beginning is one side of the power line and the end is the other side of the line. What happens in between is that a number of switches are used to make sure the device turns on or off when it is supposed to cool, freeze, or heat.

The ladder diagram makes it easier to see how these devices are wired. It consists of two wires drawn parallel and representing the main power source. Along each side you find connections. By simply looking from left to right, you are able to trace the required power for the device. Symbols are used to represent the devices. There is usually a legend on the side of the diagram to tell you, for example, that CC means compressor contactor, EFR means evaporator fan relay, and HR means heating relay (see Fig. 15-3).

Take a look at the thermostat in Fig. 15-3. The location of the switch determines whether the evaporator fan relay coil, the compressor contactor coil, or the heating relay coil is energized. Once the coil of the EFR is energized by having the thermostat turned to make contact with the desired point (G), it closes the points in the relay and the evaporator fan motor starts to move. This means that the low voltage (24 volts) has energized the relay. The relay energizes and closes the EFR contacts located in the high-voltage (240 volts) circuit. If the thermostat is turned to W or the heating position, it will cause the heating relay coil to be energized when the thermostat switch closes and demands

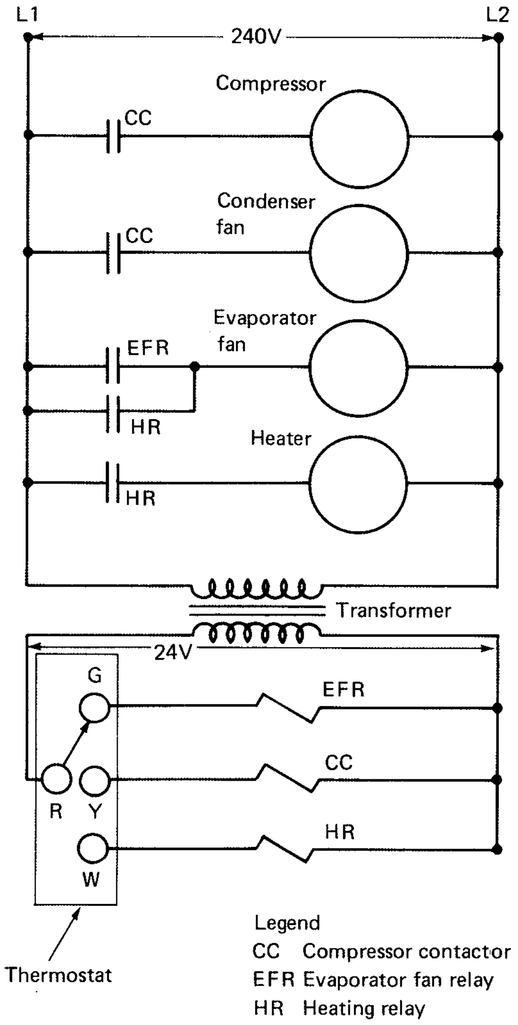
L1
L2
240V
Compressor
CC
Condenser fan
CC
Evaporator fan
EFR
HR
Heater
HR
Transformer
24V
G
EFR
CC
R
Y
HR
W
Thermostat
Legend
CC Compressor contactor
EFR Evaporator fan relay
HR Heating relay

Figure 15-3 Ladder diagram for a heat and cool installation.

heat. The energized heating relay coil causes the HR contacts to close, which in turn places the heating element across the 240-volt line and it begins to heat up. Note that the HR contacts are in parallel with the evaporator fan relay contacts. Thus, the evaporator fan will operate when either the heating relay or the evaporator fan relay is energized.

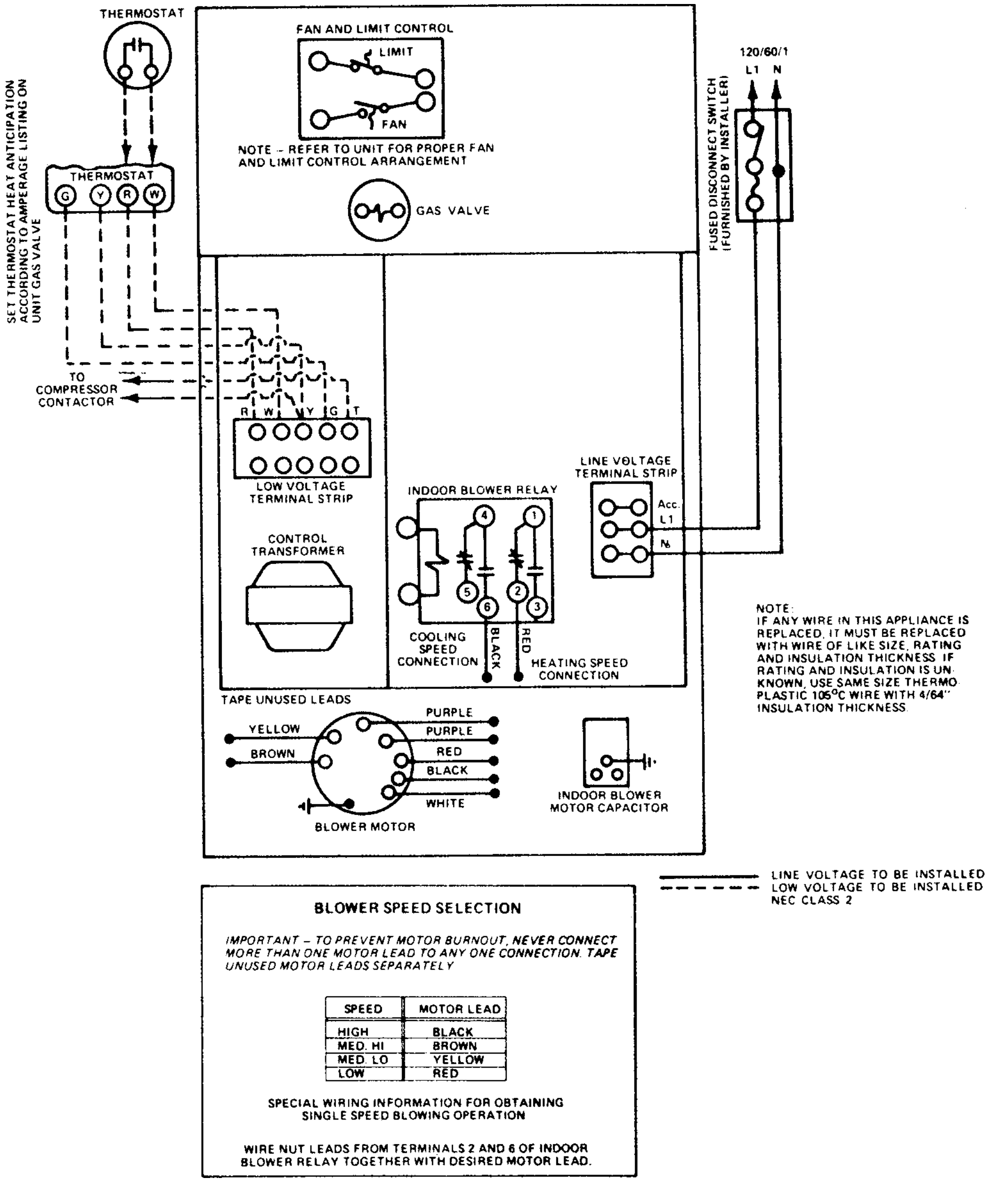
THERMOSTAT
SET THERMOSTAT HEAT ANTICIPATION ACCORDING TO AMPERAGE LISTING ON UNIT GAS VALVE
THERMOSTAT
G
Y
R
W
TO COMPRESSOR CONTACTOR
FAN AND LIMIT CONTROL
LIMIT
FAN
NOTE – REFER TO UNIT FOR PROPER FAN AND LIMIT CONTROL ARRANGEMENT
GAS VALVE
120/60/1
L1 N
FUSED DISCONNECT SWITCH (FURNISHED BY INSTALLER)
R W Y G T
LOW VOLTAGE TERMINAL STRIP
CONTROL TRANSFORMER
INDOOR BLOWER RELAY
4
1
5
6
2
3
COOLING SPEED CONNECTION
BLACK
RED
HEATING SPEED CONNECTION
LINE VOLTAGE TERMINAL STRIP
Acc.
L1
N
TAPE UNUSED LEADS
YELLOW
BROWN
PURPLE
PURPLE
RED
BLACK
WHITE
BLOWER MOTOR
INDOOR BLOWER MOTOR CAPACITOR

NOTE:
IF ANY WIRE IN THIS APPLIANCE IS REPLACED, IT MUST BE REPLACED WITH WIRE OF LIKE SIZE, RATING AND INSULATION THICKNESS. IF RATING AND INSULATION IS UNKNOWN, USE SAME SIZE THERMO PLASTIC 105°C WIRE WITH 4/64" INSULATION THICKNESS.

——— LINE VOLTAGE TO BE INSTALLED
– – – LOW VOLTAGE TO BE INSTALLED NEC CLASS 2

BLOWER SPEED SELECTION

IMPORTANT – TO PREVENT MOTOR BURNOUT, NEVER CONNECT MORE THAN ONE MOTOR LEAD TO ANY ONE CONNECTION. TAPE UNUSED MOTOR LEADS SEPARATELY

SPEED	MOTOR LEAD
HIGH	BLACK
MED. HI	BROWN
MED. LO	YELLOW
LOW	RED

SPECIAL WIRING INFORMATION FOR OBTAINING SINGLE SPEED BLOWING OPERATION

WIRE NUT LEADS FROM TERMINALS 2 AND 6 OF INDOOR BLOWER RELAY TOGETHER WITH DESIRED MOTOR LEAD.

Figure 15-4 Manufacturer's diagram for a hot-air installation.

Manufacturer's Diagrams

Figure 15-4 shows how the manufacturer represents the location of the various furnace devices. The solid lines indicate the line voltage to be installed. The dotted lines are the low voltage to be installed when the furnace is put into service.

The motor is four speed. It has various colored leads to represent the speeds. You may have to change the speed of the motor to move the air to a given location. Most motors come from the factory with a medium-high speed selected. The speed is usually easily changed by removing a lead from one point and placing it on another where the proper color is located. In the schematic of Fig. 15-5, the fan motor has white connected to one side of the 120-volt line (neutral), and the red and black are switched by the indoor blower relay to black for the cooling speed and red for the heating speed. It takes a faster fan motor to push cold air than that required for hot air because cold air is heavier than hot air.

In Fig. 15-5, the contacts on the thermostat are labeled R, W, Y, G, R, and W are used to place the thermostat in the circuit. It can be switched from W to Y manually by moving the Heat-Cool switch on the thermostat to Cool position.

Notice in Fig. 15-5 that the indoor blower relay coil is in the circuit all the time when the "auto on" switch on the thermostat is located at the "on" position. The schematic also shows the cool position has been selected manually, and the thermostat contacts will complete the circuit when it moves from W1 to Y1.

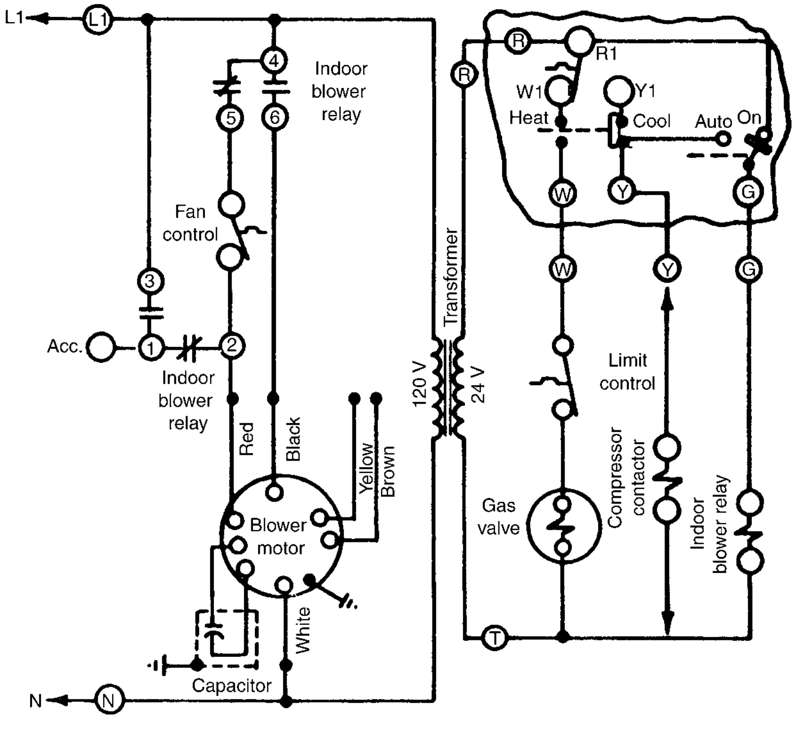
L1
L1
4
Indoor blower relay
5
6
R
R
R1
W1
Y1
Heat
Cool
Auto
On
W
Y
G
Fan control
W
Y
G
3
Transformer
Acc.
1
2
Indoor blower relay
120 V
24 V
Limit control
Red
Black
Yellow
Brown
Compressor contactor
Gas valve
Blower motor
Indoor blower relay
White
T
Capacitor
N
N

Figure 15-5 Schematic for a hot-air installation.

In Fig. 15-4, note that the low-voltage terminal strip has a T on it. This is the common side of the low voltage from the transformer. In Fig. 15-5, the T is the common side of the low-voltage transformer secondary. In Fig. 15-4, the T terminal is connected to the compressor contactor by a wire run from the terminal to the contactor. Note that the other wire to the contactor runs from Y on the terminal strip. Now go back to Fig. 15-5, where the Y and T terminals are shown as connection points for the compressor contactor. Are you able to relate the schematic to the actual device? The gas valve is wired by having wire T of the terminal strip attached to one side of the solenoid and a wire run from the limit switch to the other side of the solenoid.

Figure 15-6 shows how the wiring diagram comes from the factory. It is usually located inside the cover for the cold-air return. In most instances, it is glued to the cover so that it is handy for the person working on the furnace whenever there is a problem after installation.

Field Wiring

The installation of a new furnace requires you to follow a factory diagram furnished in a booklet that accompanies the unit. The wiring to be done in the field is represented by the dotted lines in Fig. 15-7. All electrical connections should be made in accordance with the National Electrical Code and any local codes or ordinances that might apply.

WARNING: *The unit cabinet must have an uninterrupted or unbroken electrical ground to minimize personal injury if an electrical fault should occur. This may consist of electrical wire or approved conduit when installed in accordance with existing electrical codes.*

Low-Voltage Wiring

Make the field low-voltage connections at the low-voltage terminal strip shown in Fig. 15-7. Set the thermostat heat anticipator at 0.60 ampere (or whatever is called for by the manufacturer). If additional controls are connected in the thermostat circuit, their amperage draw must be added to this setting. Failure to make the setting will result in improper operation of the thermostat.

With the addition of an automatic vent damper, the anticipator setting would then be 0.12 ampere. As you can see from this and the schematic (see Fig. 14-16), the anticipator resistor is in series with whatever is in the circuit and is to be controlled by the thermostat. The more devices controlled by the thermostat, the more current will be drawn from the transformer to energize them. As the current demand increases, the current through the anticipator is also increased. As you remember from previous chapters, a series circuit has the same current through each component in the circuit.

Thermostat location. The room thermostat should be located where it will be in the natural circulating path of room air. Avoid locations where the thermostat

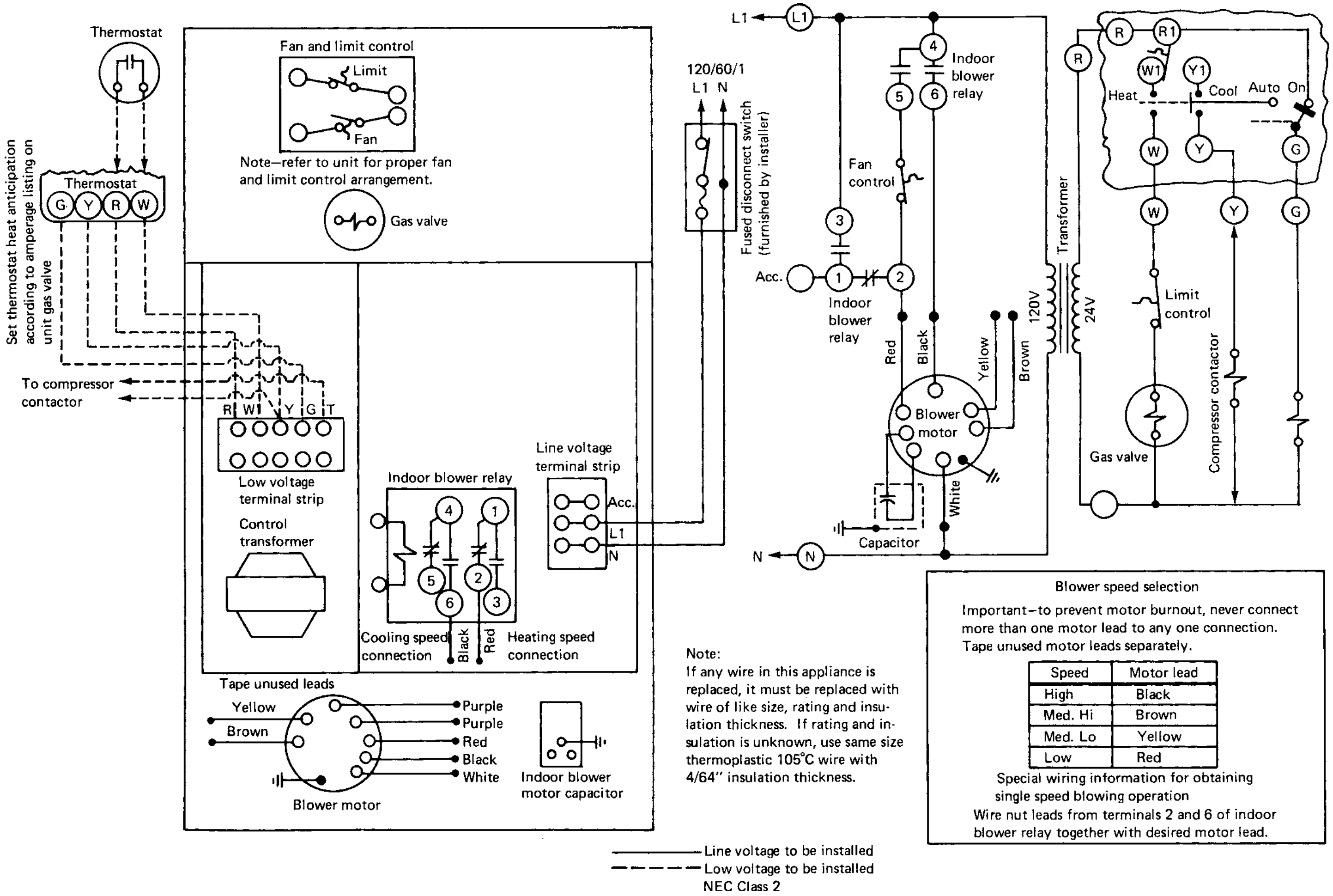
Thermostat
Set thermostat heat anticipation according to amperage listing on unit gas valve
Thermostat
G Y R W
To compressor contactor
R W Y G T
Low voltage terminal strip
Control transformer
Fan and limit control
Limit
Fan
Note—refer to unit for proper fan and limit control arrangement.
Gas valve
Indoor blower relay
4 1 5 2 6 3
Cooling speed connection
Black
Red
Heating speed connection
Line voltage terminal strip
Acc.
L1
N
Tape unused leads
Yellow
Brown
Purple
Purple
Red
Black
White
Blower motor
Indoor blower motor capacitor
120/60/1
L1 N
Fused disconnect switch (furnished by installer)
L1
N
Acc.
Indoor blower relay
Fan control
Red
Black
Yellow
Brown
White
Blower motor
Capacitor
120V
24V
Transformer
R
R1
W1
Y1
Heat
Cool
Auto
On
W
Y
G
Limit control
Gas valve
Compressor contactor

Note:
If any wire in this appliance is replaced, it must be replaced with wire of like size, rating and insulation thickness. If rating and insulation is unknown, use same size thermoplastic 105°C wire with 4/64" insulation thickness.

Blower speed selection

Important—to prevent motor burnout, never connect more than one motor lead to any one connection. Tape unused motor leads separately.

Speed	Motor lead
High	Black
Med. Hi	Brown
Med. Lo	Yellow
Low	Red

Special wiring information for obtaining single speed blowing operation
Wire nut leads from terminals 2 and 6 of indoor blower relay together with desired motor lead.

——— Line voltage to be installed
— — — Low voltage to be installed
NEC Class 2

Figure 15-6 Complete instruction page packaged with a hot-air furnace.

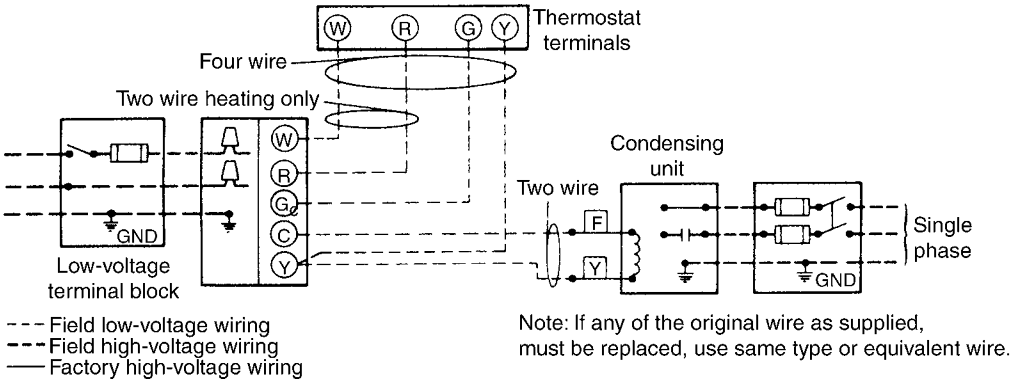
W
R
G
Y
Thermostat terminals
Four wire
Two wire heating only
W
R
G
C
Y
GND
Low-voltage terminal block
Condensing unit
Two wire
F
Y
GND
Single phase
--- Field low-voltage wiring
--- Field high-voltage wiring
—— Factory high-voltage wiring
Note: If any of the original wire as supplied, must be replaced, use same type or equivalent wire.

Figure 15-7 Heating and cooling application wiring diagram. (Courtesy of Carrier)

is exposed to cold-air infiltration, drafts from windows, doors, or other openings leading to the outside, or air currents from warm- or cold-air registers, or to exposure where the natural circulation of the air is cut off, such as behind doors and above or below mantels or shelves. Also keep the thermostat out of direct sunlight.

The thermostat should not be exposed to heat from nearby fireplaces, radios, televisions, lamps, or rays from the sun. Nor should the thermostat be mounted on a wall containing pipes or warm-air ducts or a flue or vent that could affect its operation and prevent it from properly controlling the room temperature. Any hole in the plaster or panel through which the wires pass from the thermostat should be adequately sealed with suitable material to prevent drafts from affecting the thermostat.

Printed circuit board control center. Newer hot-air furnaces feature printed circuit control. The board shown in Fig. 15-8 is such that it is easy for the technician installing the furnace to hook it up properly the first time. The markings are designed for making it easy to connect the furnace for accessories, if needed. Figures 15-9 and 15-10 show the factory-furnished schematic. See if you can trace the schematic and locate the various points on the printed circuit boards.

Heat Pumps

The heat pump is a heat multiplier. It takes warm air and makes it hot air. This is done by compressing the air and increasing its temperature. Heat pumps received more attention during the fuel embargo of 1974 with some emphasis being placed on them now. Energy conservation has become a more important concern for everyone. If a device can be made to take heat from the air and heat a home or commercial building, it is very useful.

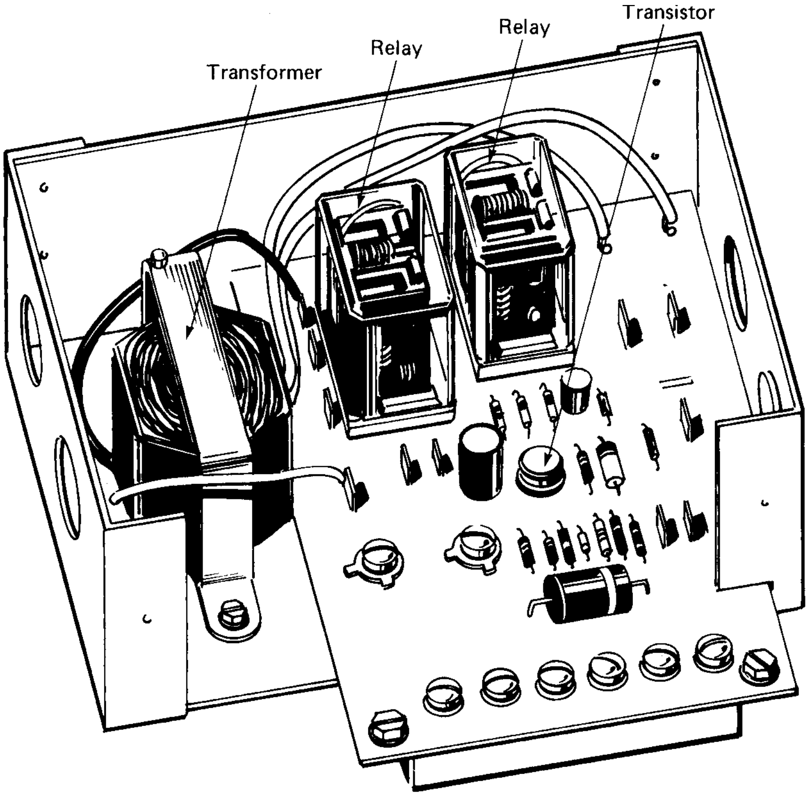

Figure 15-8 Printed circuit control center: heat and cool models. (Courtesy of Carrier)

The heat pump can take the heat generated by a refrigeration unit and use it to heat a house or room. Most of them take the heat from outside the home and move it indoors (see Fig. 15-11). This unit can be used to air condition the house in the summer and heat it in the winter by taking the heat from the outside air and moving it inside.

Operation

On mild temperature heating days, the heat pump handles all heating needs. When the outdoor temperature reaches the balance point of the home, that is, when the heat loss is equal to the heat-pump heating capacity, the two-stage indoor thermostat activates the furnace (a secondary heat source, in most cases electric heating elements). As soon as the furnace is turned on, a heat relay de-energizes the heat pump. When the second-stage (furnace) need is satisfied and the plenum temperature has cooled to below 90 and 100°F, the heat-pump relay turns the heat pump back on and controls the conditioned space until the

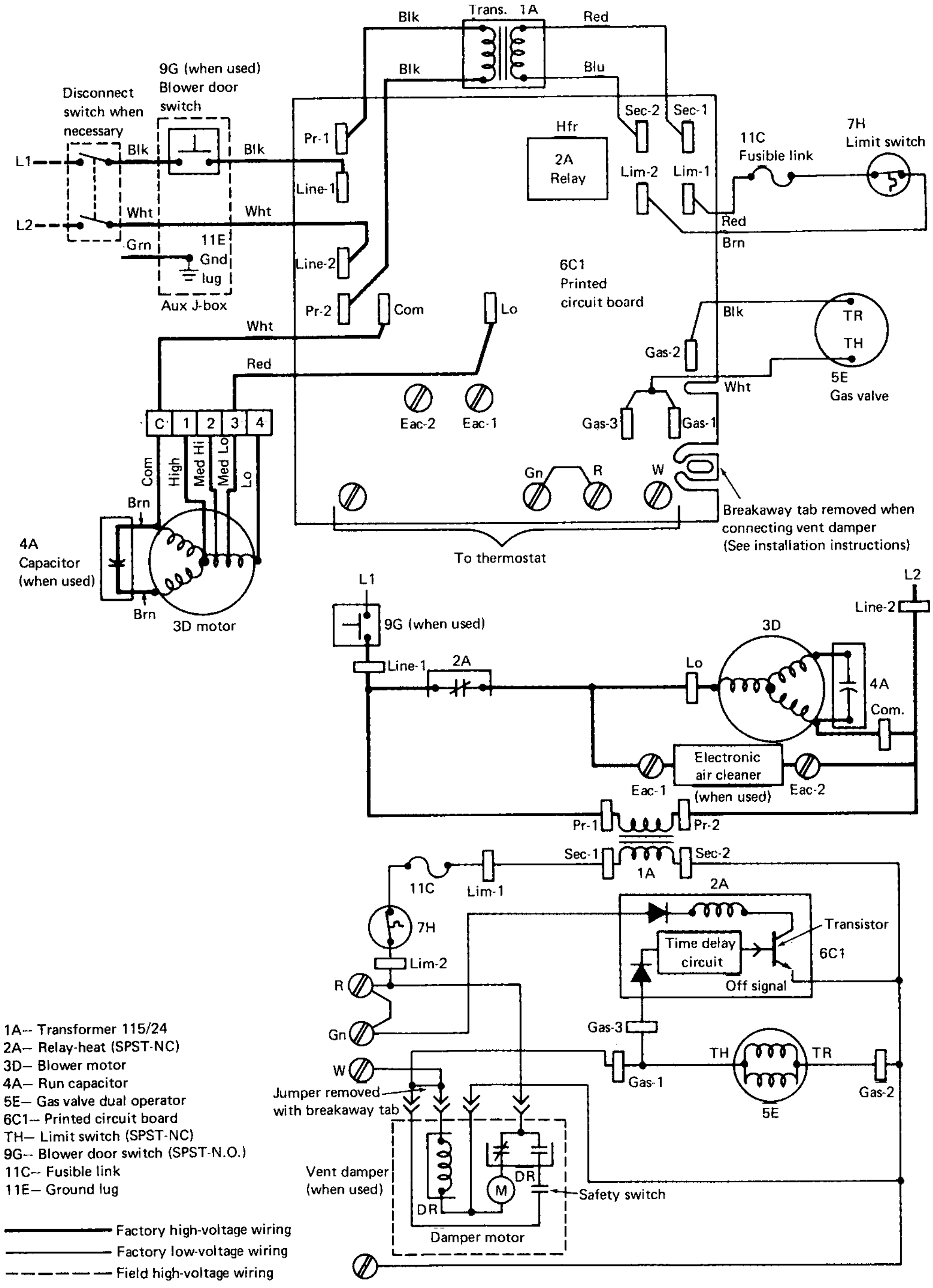
Trans. 1A
Blk
Red
Blk
Blu
9G (when used)
Blower door switch
Disconnect switch when necessary
L1
L2
Blk
Blk
Wht
Wht
Grn
11E
Gnd
lug
Aux J-box
Pr-1
Line-1
Line-2
Pr-2
Com
Lo
Hfr
2A
Relay
Sec-2
Sec-1
Lim-2
Lim-1
11C
Fusible link
7H
Limit switch
Red
Brn
6C1
Printed circuit board
Blk
TR
TH
Gas-2
Wht
5E
Gas valve
Wht
Red
C
1
2
3
4
Com
High
Med Hi
Med Lo
Lo
Brn
Brn
4A
Capacitor
(when used)
3D motor
Eac-2
Eac-1
Gas-3
Gas-1
Gn
R
W
Breakaway tab removed when connecting vent damper
(See installation instructions)
To thermostat
L1
L2
9G (when used)
Line-2
Line-1
2A
Lo
3D
4A
Com.
Electronic air cleaner
(when used)
Eac-1
Eac-2
Pr-1
Pr-2
Sec-1
Sec-2
1A
11C
Lim-1
2A
7H
Transistor
6C1
Time delay circuit
Off signal
Lim-2
R
Gn
W
Gas-3
TH
TR
Gas-1
Gas-2
5E
Jumper removed with breakaway tab
Vent damper (when used)
DR
DR
M
Safety switch
Damper motor
1A— Transformer 115/24
2A— Relay-heat (SPST-NC)
3D— Blower motor
4A— Run capacitor
5E— Gas valve dual operator
6C1— Printed circuit board
TH— Limit switch (SPST-NC)
9G— Blower door switch (SPST-N.O.)
11C— Fusible link
11E— Ground lug
Factory high-voltage wiring
Factory low-voltage wiring
Field high-voltage wiring

Figure 15-9 Wiring diagram. Heating only. (Courtesy of Carrier)

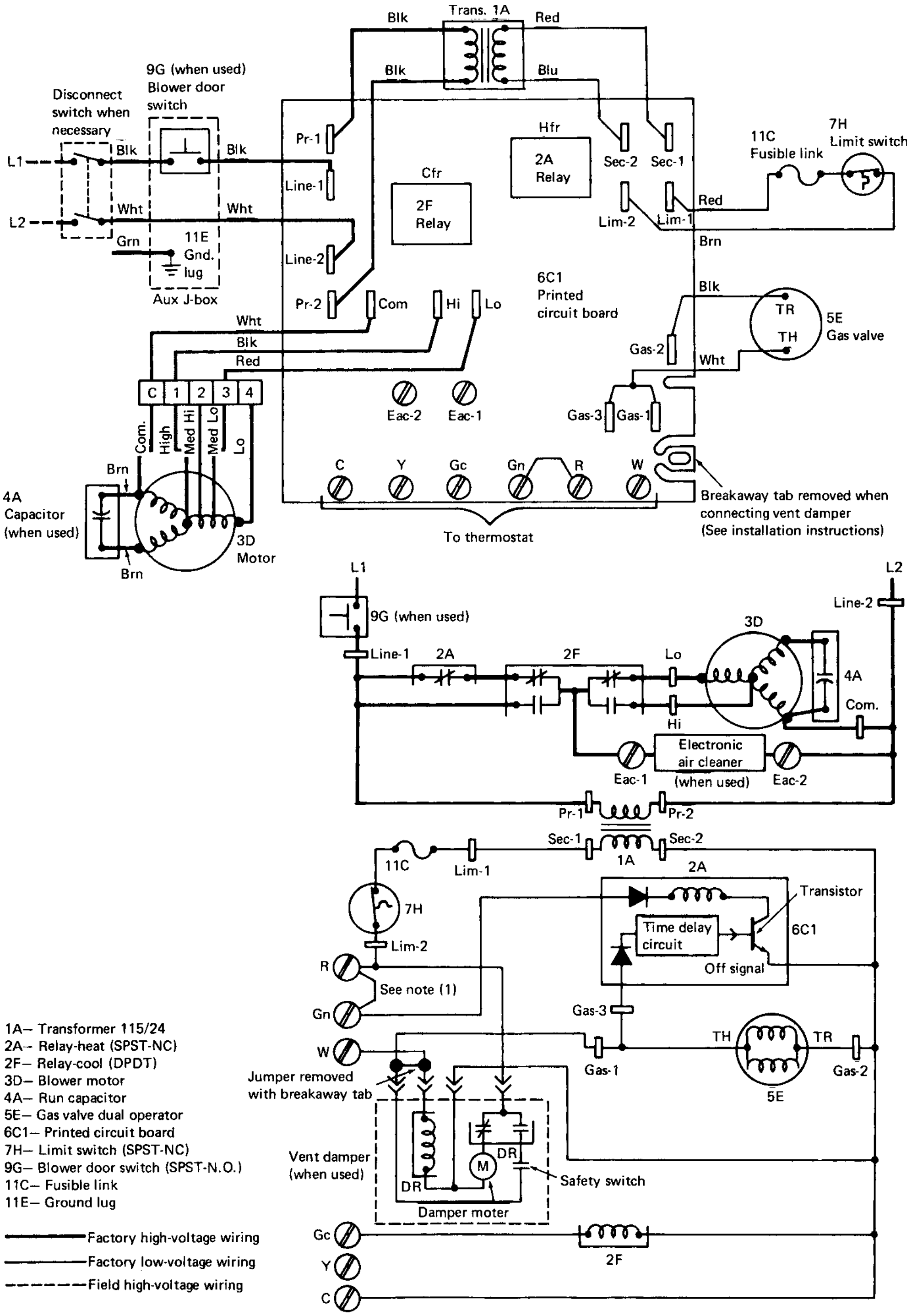
Trans. 1A
Blk
Red
Blu
9G (when used)
Blower door switch
Disconnect switch when necessary
L1
L2
Wht
Grn
11E Gnd. lug
Aux J-box
Pr-1
Line-1
Line-2
Pr-2
Com
Hi
Lo
Cfr
2F Relay
Hfr
2A Relay
Sec-2
Sec-1
Lim-2
Lim-1
11C Fusible link
7H Limit switch
Brn
6C1 Printed circuit board
TR
TH
5E Gas valve
Gas-2
Gas-3
Gas-1
Eac-2
Eac-1
C
Y
Gc
Gn
R
W
To thermostat
Breakaway tab removed when connecting vent damper (See installation instructions)
C 1 2 3 4
Com.
High
Med Hi
Med Lo
Lo
4A Capacitor (when used)
3D Motor
3D
4A
Com.
Electronic air cleaner (when used)
Transistor
Time delay circuit
Off signal
See note (1)
Jumper removed with breakaway tab
Vent damper (when used)
DR
M
Safety switch
Damper moter
1A— Transformer 115/24
2A— Relay-heat (SPST-NC)
2F— Relay-cool (DPDT)
3D— Blower motor
4A— Run capacitor
5E— Gas valve dual operator
6C1— Printed circuit board
7H— Limit switch (SPST-NC)
9G— Blower door switch (SPST-N.O.)
11C— Fusible link
11E— Ground lug
Factory high-voltage wiring
Factory low-voltage wiring
Field high-voltage wiring

Figure 15-10 Wiring diagram for heating and cooling. (Courtesy of Carrier)

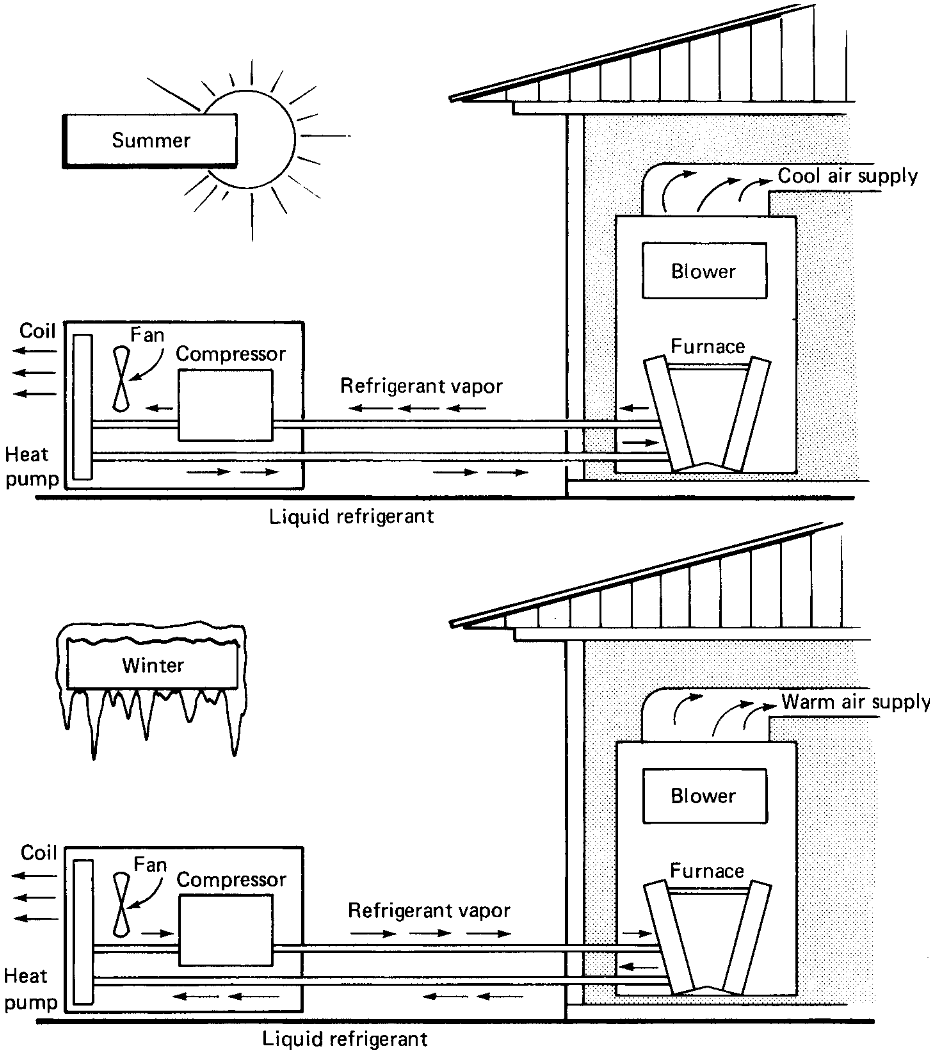
Summer
Cool air supply
Blower
Coil
Fan
Compressor
Furnace
Refrigerant vapor
Heat pump
Liquid refrigerant
Winter
Warm air supply
Blower
Coil
Fan
Compressor
Furnace
Refrigerant vapor
Heat pump
Liquid refrigerant

Figure 15-11 Basic operation of a heat pump.

second-stage operation is required again. Figure 15-12 shows the heat-pump unit. The optional electric heat unit shown in Fig. 15-13 is added in geographic locations where needed. This particular unit can provide 23,000 to 56,000 Btus per hour (Btuh) and up to 112,700 Btuh with the addition of electric heat.

If the outdoor temperature drops below the setting of the low-temperature compressor monitor, the control shuts off the heat pump completely and the furnace handles all the heating needs.

During the defrost cycle, the heat pump switches from heating to cooling. To prevent cool air from being circulated in the house when heating is needed, the

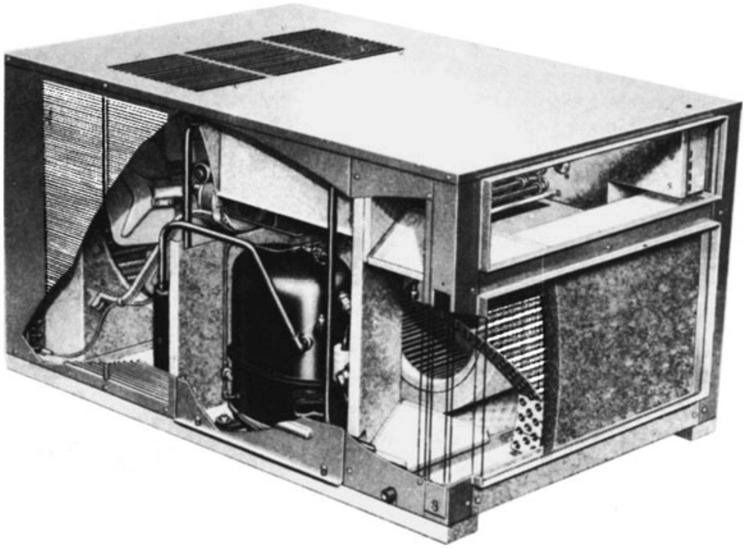

Figure 15-12 Single package heat pump.

control will automatically turn on the furnace to compensate for the heat-pump defrost cycle (see Fig. 15-14). When supply air temperature climbs above 110 to 120°F, the defrost limit control turns off the furnace and keeps indoor air from getting too warm.

If, after a defrost cycle, the air downstream of the coil gets above 115°F, the closing point of the heat-pump relay, the compressor will stop until the heat exchanger has cooled down to 90° to 100° as it does during normal cycling operation between furnace and heat pump.

During summer cooling, the heat pump works as a normal split system, using the furnace blower as the primary air mover (see Fig. 15-15).

In a straight heat pump/supplementary electric heater application, at least one outdoor thermostat is required to cycle the heaters as the outdoor temperature drops. In the system shown here, the indoor thermostat controls the supplemental heat source (furnace). The outdoor thermostat is not required.

Figure 15-13 Optional electric heat for a heat pump.

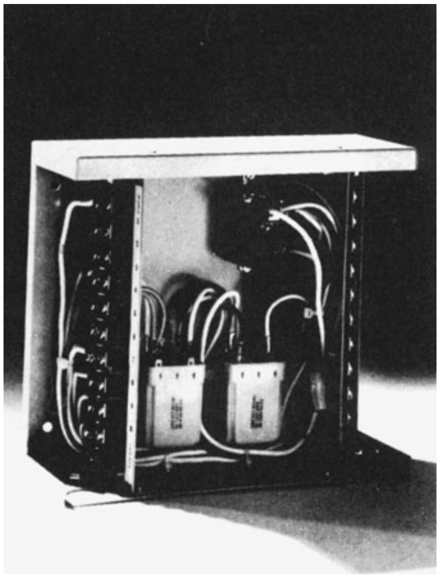

Figure 15-14 Control box for an add-on type heat pump.

Since the furnace is serving as the secondary heat source, the system does not require the home rewiring usually associated with supplemental electric strip heating.

Special requirements of heat-pump systems

The installation, maintenance, and operating efficiency of the heat-pump system are like those of no other comfort system. A heat-pump system requires the same air quantity for heating and cooling. Because of this, the air-moving capability of an existing furnace is extremely important. It should be carefully

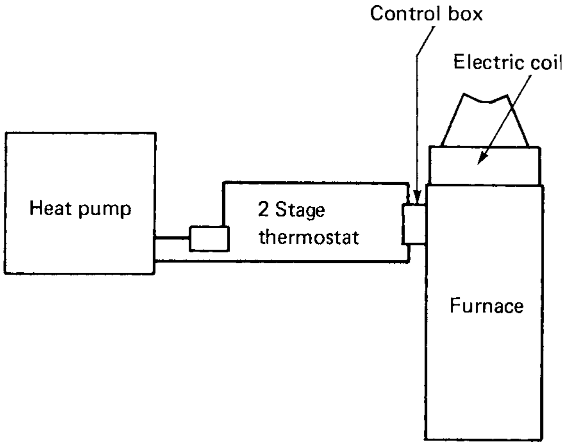

Figure 15-15 Heat pump with a two-stage thermostat and control box mounted on the furnace.

checked before a heat pump is added. Heating and load calculations must be accurate. System design and installation must be precise and according to the manufacturer's suggestions.

The air-distribution system and diffuser location are equally important. Supply ducts must be properly sized and insulated. Adequate return air is also required. Heating supply air is cooler than with other systems. This is quite noticeable to homeowners accustomed to gas or oil heat. This makes diffuser location and system balancing critical.

Heat-pump combinations

There are four ways to describe the heat-pump methods of transporting heat into the house:

1. *Air-to-air:* This is the most common method. It is the type of system previously described.
2. *Air-to-water:* This type uses two different types of heat exchangers. Warmed refrigerant flows through pipes to a heat exchanger in the boiler. Heated water flows into radiators located within the heated space.
3. *Water-to-water:* This type uses two water-to-refrigerant heat exchangers. Heat is taken from the water source (well water, lakes, or the sea) and is passed on by the refrigerant to the water used for heating. The reverse takes place in the cooling system.
4. *Water-to-air:* Well water furnishes the heat. This warms the refrigerant in the heat-exchanger coil. The refrigerant, compressed, flows to the top of the unit, where a fan blows air past the heat exchanger.

Each type of heat pump has its advantages and disadvantages. Each needs to be properly controlled. This is where the electrical connections and controls are used to do the job properly. Before attempting to work on this type of equipment, make sure you have a complete schematic of the electrical wiring and know all the component parts of the system.

High-Efficiency Furnaces

Furnaces have been designed (since 1981) with efficiencies of up to 97%, as compared to older types with efficiencies in the 60% range. The Lennox Pulse is one example of the types available. The G14 series pulse combustion up-flow gas furnace provides efficiency of up to 97%. Eight models for natural gas and LPG are available with input capacities of 40,000, 60,000, 80,000, and 100,000 Btuh. The units operate on the pulse-combustion principle and do not require a pilot burner, main burners, conventional flue, or chimney. Compact, standard-sized cabinet design, with side or bottom return air entry, permits installation in a basement, utility room, or closet. Evaporator coils may be added, as well as electronic air cleaners and power humidifiers (see Fig. 15-16).

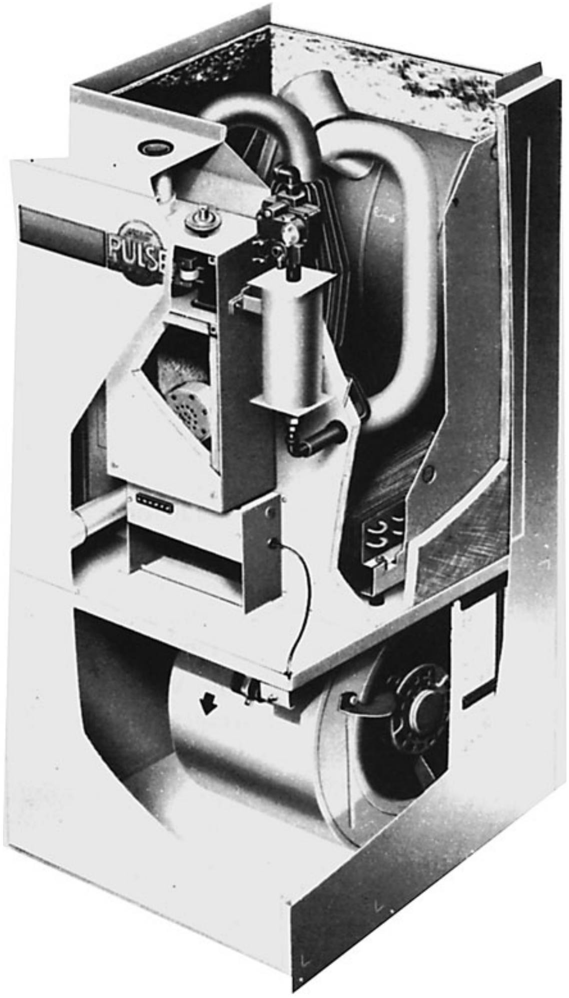

Figure 15-16 Lennox Pulse furnace. (Courtesy of Lennox)

Operation

The high-efficiency furnaces achieve a high level of fuel conversion by using a unique heat-exchanger design. It features a finned cast-iron combustion chamber, temperature-resistant steel tailpipe, aluminized steel exhaust de-coupler section, and a finned stainless-steel tube condenser coil similar to an air-conditioner coil. Moisture, in the products of combustion, is condensed in the coil, thus wringing almost every usable Btu out of the gas. Since most of the combustion heat is utilized in the heat transfer from the coil, flue vent temperatures are as low as 100 to 130°F, allowing for the use of 2-inch diameter polyvinyl chloride (PVC) pipe. The furnace is vented through a sidewall or roof or to the top of an existing chimney with up to 25 feet of PVC pipe and four 90° elbows. Condensate created in the coil

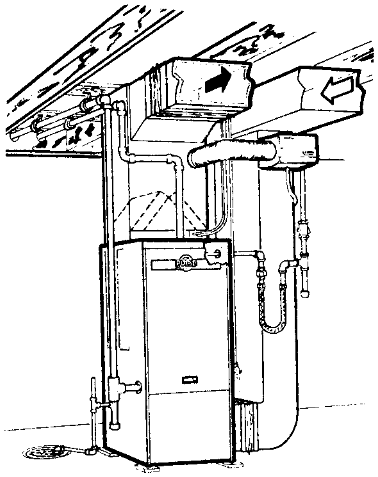

Figure 15-17 Basement installation of the Pulse with cooling coil and automatic humidifier. Note the floor drain for condensate. (Courtesy of Lennox)

may be disposed of in an indoor drain (see Fig. 15-17). The condensate is not harmful to standard household plumbing and can be drained into city sewers and septic tanks without damage.

The furnace has no pilot light or burners. An automotive-type spark plug is used for ignition on the initial cycle only, saving gas and electrical energy. Due to the pulse-combustion principle, the use of atmospheric gas burners is eliminated, with the combustion process confined to the heat-exchanger combustion chamber. The sealed combustion system virtually eliminates the loss of conditioned air due to combustion and stack dilution. Combustion air from the outside is piped to the furnace with the same type of PVC pipe used for exhaust gases.

Electrical controls

The furnace is equipped with a standard-type redundant gas valve in series with a gas expansion tank, gas intake flapper valve, and air intake flapper valve. Also factory installed are a purge blower, spark plug igniter, and flame sensor with solid-state control circuit board. The standard equipment includes a fan and limit control, a 30-VA transformer, blower cooling relay, flexible gas line connector, and four isolation mounting pads, as well as a base insulation pad, condensate drip leg, and cleanable air filter. Flue vent/air intake line, roof or wall termination installation kits, LPG conversion kits, and thermostat are available as accessories and must be ordered extra, or you can use the existing one when replacing a unit.

The printed circuit board is replaceable as a unit when there is a malfunction of one of the components. It uses a multivibrator transistorized circuit to generate the high voltages needed for the spark plug. The spark plug gets very little use except to start the combustion process. It has a long life expectancy. Spark gap is 0.115 inch and the ground electrode is adjusted to 45° (see Fig. 15-20).

Sequence of operation

On a demand for heat, the room thermostat initiates the purge blower operation for a prepurge cycle of 34 seconds, followed by energizing of the ignition and opening of the gas valve. As ignition occurs, the flame sensor senses proof of ignition and de-energizes the spark igniter and purge blower (see Fig. 15-18). The furnace blower operation is initiated 30 to 45 seconds after combustion ignition. When the thermostat is satisfied, the gas valve closes and the purge blower is re-energized for a postpurge cycle of 34 seconds. The furnace blower remains in operation until the preset temperature setting (90°F) of the fan control is reached. Should the loss of flame occur before the thermostat is satisfied, flame sensor controls will initiate three to five attempts at reignition before locking out the unit operation. In addition,

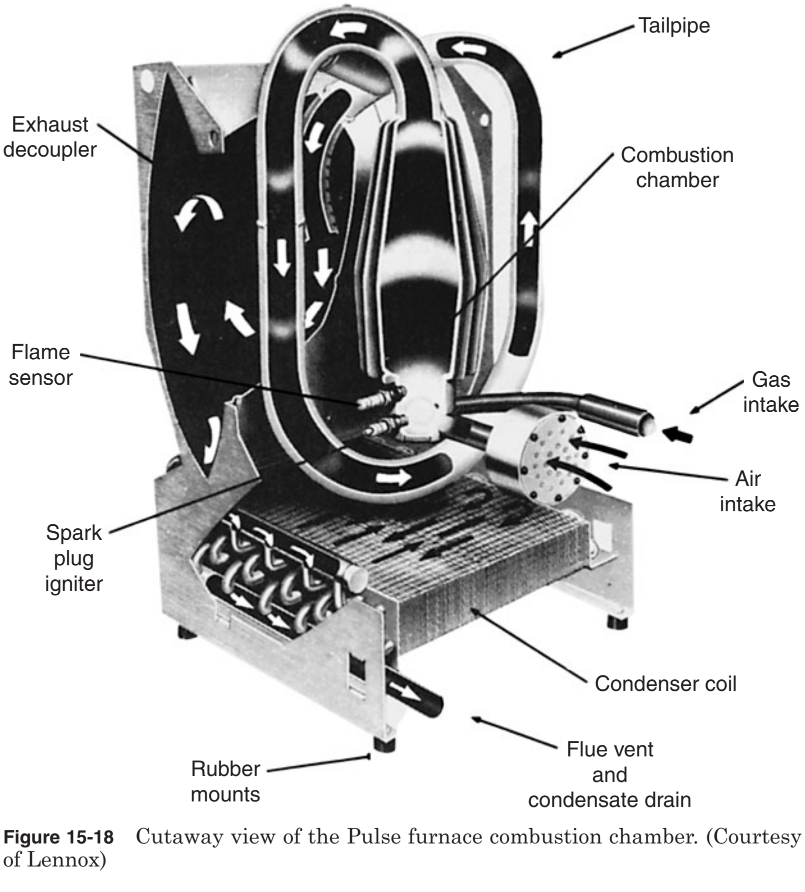
Tailpipe
Exhaust decoupler
Combustion chamber
Flame sensor
Gas intake
Air intake
Spark plug igniter
Condenser coil
Rubber mounts
Flue vent and condensate drain

Figure 15-18 Cutaway view of the Pulse furnace combustion chamber. (Courtesy of Lennox)

loss of either combustion intake air or flue exhaust will automatically shut down the system.

Combustion process

The process of pulse combustion begins as gas and air are introduced into the sealed combustion chamber with the spark plug igniter. Spark from the plug ignites the gas/air mixture, which in turn causes a positive pressure buildup that closes the gas and air inlets. This pressure relieves itself by forcing the products of combustion out of the combustion chamber through the tailpipe into the heat-exchanger exhaust de-coupler and on into the heat-exchanger coil. As the combustion chamber empties, its pressure becomes negative, drawing in air and gas for ignition of the next pulse. At the same instant, part of the pressure pulse is reflected back from the tailpipe at the top of the combustion chamber. The flame remnants of the previous pulse of combustion ignite the new gas/air mixture in the chamber, continuing the cycle.

Once combustion is started, it feeds on itself, allowing the purge blower and spark igniter to be turned off. Each pulse of gas/air mixture is ignited at the rate of 60 to 70 times per second, producing one-fourth to one-half Btu per pulse of combustion. Almost complete combustion occurs with each pulse. The force of these series of ignitions creates great turbulence, which forces the products of combustion through the entire heat-exchanger assembly, resulting in maximum heat transfer (see Fig. 15-18).

Start-up procedures for the GSR-14Q series of Pulse furnaces, as well as maintenance and repair parts, are shown in Fig. 15-19.

Troubleshooting the Pulse Furnace

Troubleshooting procedures for the Pulse furnaces are shown in Fig. 15-20. Figure 15-21 shows the circuitry for the G-14Q series of furnaces. Note the difference in the electrical circuitry for the G-14 and GSR-14. Blower speed color-coded wires are also indicated for the different units. The 40, 60, 80, and 100 after the G-14Q indicates whether it is a 40,000, 60,000, 80,000, or 100,000 Btuh unit. Thermostat heat anticipation is also given for the Robertshaw valve and the Rodgers valve. This type of electrical diagram is usually glued to the cabinet so that it is with the unit whenever there is need for troubleshooting.

The troubleshooting flowchart is typical of those furnished with newer equipment in the technical manuals furnished the dealers who provide the service. After locating the exact symptoms, check with the other part of Fig. 15-20 to find how to use the multimeter to check out all the circuitry to see if the exact cause of the problem can be determined.

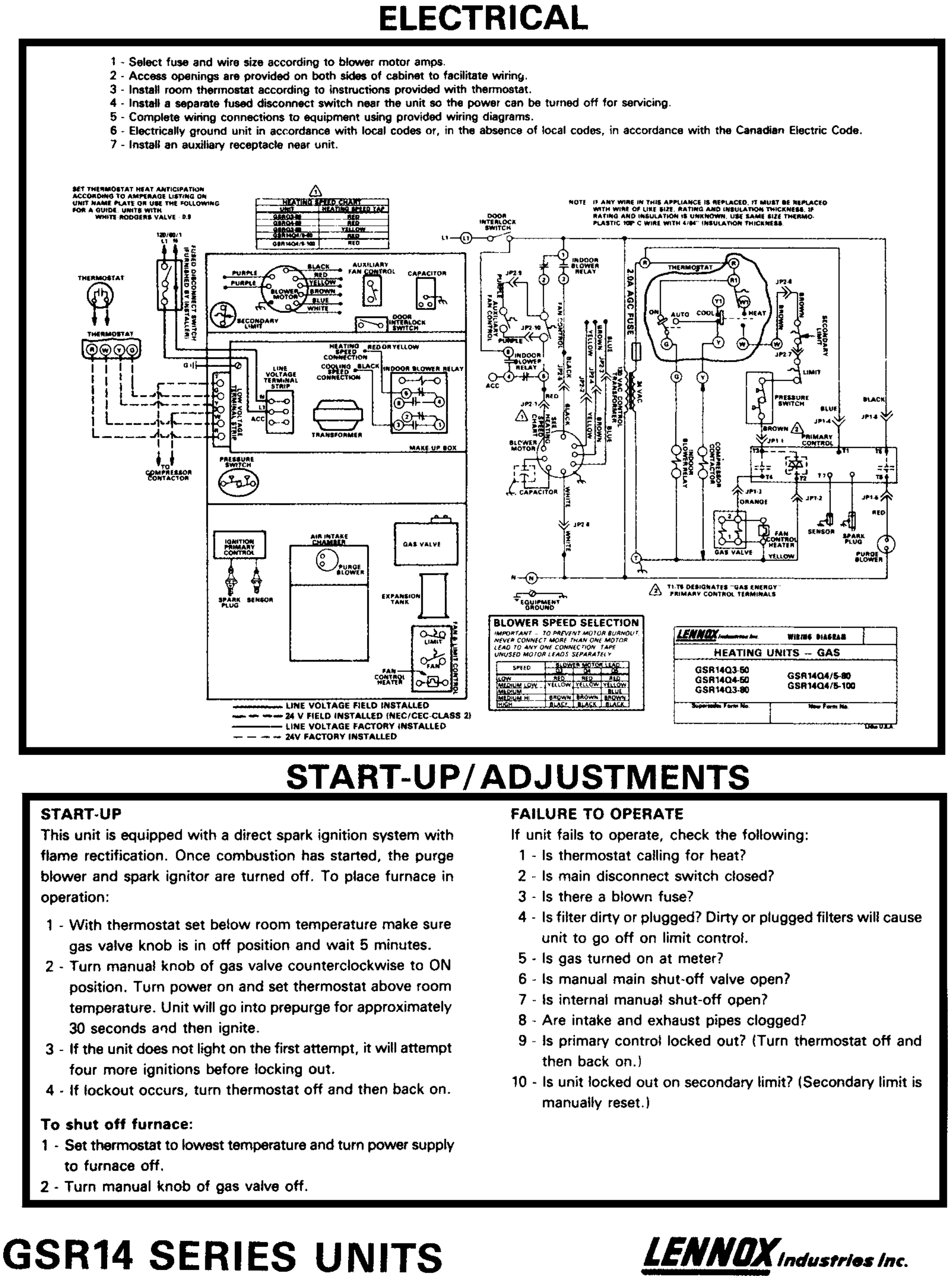

ELECTRICAL

1 - Select fuse and wire size according to blower motor amps.
2 - Access openings are provided on both sides of cabinet to facilitate wiring.
3 - Install room thermostat according to instructions provided with thermostat.
4 - Install a separate fused disconnect switch near the unit so the power can be turned off for servicing.
5 - Complete wiring connections to equipment using provided wiring diagrams.
6 - Electrically ground unit in accordance with local codes or, in the absence of local codes, in accordance with the Canadian Electric Code.
7 - Install an auxiliary receptacle near unit.

SET THERMOSTAT HEAT ANTICIPATION ACCORDING TO AMPERAGE LISTING ON UNIT NAME PLATE OR USE THE FOLLOWING FOR A GUIDE. UNITS WITH WHITE-RODGERS VALVE - 0.8

NOTE: IF ANY WIRE IN THIS APPLIANCE IS REPLACED, IT MUST BE REPLACED WITH WIRE OF LIKE SIZE, RATING AND INSULATION THICKNESS. IF RATING AND INSULATION IS UNKNOWN, USE SAME SIZE THERMOPLASTIC 105° C WIRE WITH 4/64" INSULATION THICKNESS.

HEATING SPEED CHART	
UNIT	HEATING SPEED TAP
GSR14Q3-50	RED
GSR14Q4-50	RED
GSR14Q3-80	YELLOW
GSR14Q4/5-80	RED
GSR14Q4/5-100	RED

THERMOSTAT
THERMOSTAT
R W Y G
120/60/1
L1 N
FUSED DISCONNECT SWITCH (FURNISHED BY INSTALLER)
TO COMPRESSOR CONTACTOR
PURPLE
PURPLE
BLOWER MOTOR
BLACK
RED
YELLOW
BROWN
BLUE
WHITE
AUXILIARY FAN CONTROL
CAPACITOR
SECONDARY LIMIT
DOOR INTERLOCK SWITCH
HEATING SPEED CONNECTION
RED OR YELLOW
COOLING SPEED CONNECTION
BLACK
INDOOR BLOWER RELAY
LINE VOLTAGE TERMINAL STRIP
N
L1
ACC
LOW VOLTAGE TERMINAL STRIP
TRANSFORMER
MAKE-UP BOX
PRESSURE SWITCH
IGNITION PRIMARY CONTROL
SPARK PLUG
SENSOR
AIR INTAKE CHAMBER
PURGE BLOWER
GAS VALVE
EXPANSION TANK
LIMIT
FAN
FAN CONTROL HEATER
FAN & LIMIT CONTROL
DOOR INTERLOCK SWITCH
L1
INDOOR BLOWER RELAY
JP2-9
PURPLE
AUXILIARY FAN CONTROL
JP2-10
PURPLE
FAN CONTROL
INDOOR BLOWER RELAY
ACC
RED
JP2-1
SEE HEATING SPEED CHART
BLACK
JP2-6
JP2-7
JP2-4
JP2-3
YELLOW
BROWN
BLUE
BLOWER MOTOR
CAPACITOR
WHITE
JP2-8
2.0A AGC FUSE
120 VAC CONTROL TRANSFORMER
24 VAC
THERMOSTAT
R
R1
Y1
W1
ON
AUTO
COOL
HEAT
G
Y
W
JP2-6
BROWN
SECONDARY LIMIT
JP2-7
LIMIT
PRESSURE SWITCH
BROWN
JP1-1
PRIMARY CONTROL
BLUE
JP1-4
BLACK
JP1-6
T1 T2 T3 T4 T5 T6 T7
INDOOR BLOWER RELAY
COMPRESSOR CONTACTOR
JP1-3
ORANGE
JP1-2
JP1-5
RED
GAS VALVE
FAN CONTROL HEATER
YELLOW
SENSOR
SPARK PLUG
PURGE BLOWER
N
EQUIPMENT GROUND

T1-T6 DESIGNATES "GAS ENERGY" PRIMARY CONTROL TERMINALS

LINE VOLTAGE FIELD INSTALLED
24 V FIELD INSTALLED (NEC/CEC-CLASS 2)
LINE VOLTAGE FACTORY INSTALLED
24V FACTORY INSTALLED

BLOWER SPEED SELECTION

IMPORTANT — TO PREVENT MOTOR BURNOUT, NEVER CONNECT MORE THAN ONE MOTOR LEAD TO ANY ONE CONNECTION. TAPE UNUSED MOTOR LEADS SEPARATELY

SPEED	BLOWER MOTOR LEAD		
	Q3	Q4	Q5
LOW	RED	RED	RED
MEDIUM LOW	YELLOW	YELLOW	YELLOW
MEDIUM			BLUE
MEDIUM HI	BROWN	BROWN	BROWN
HIGH	BLACK	BLACK	BLACK

LENNOX Industries Inc. WIRING DIAGRAM

HEATING UNITS — GAS

GSR14Q3-50
GSR14Q4-50
GSR14Q3-80
GSR14Q4/5-80
GSR14Q4/5-100

Supersedes Form No.
New Form No.
Litho U.S.A.

START-UP/ADJUSTMENTS

START-UP

This unit is equipped with a direct spark ignition system with flame rectification. Once combustion has started, the purge blower and spark ignitor are turned off. To place furnace in operation:

1 - With thermostat set below room temperature make sure gas valve knob is in off position and wait 5 minutes.
2 - Turn manual knob of gas valve counterclockwise to ON position. Turn power on and set thermostat above room temperature. Unit will go into prepurge for approximately 30 seconds and then ignite.
3 - If the unit does not light on the first attempt, it will attempt four more ignitions before locking out.
4 - If lockout occurs, turn thermostat off and then back on.

To shut off furnace:

1 - Set thermostat to lowest temperature and turn power supply to furnace off.
2 - Turn manual knob of gas valve off.

FAILURE TO OPERATE

If unit fails to operate, check the following:

1 - Is thermostat calling for heat?
2 - Is main disconnect switch closed?
3 - Is there a blown fuse?
4 - Is filter dirty or plugged? Dirty or plugged filters will cause unit to go off on limit control.
5 - Is gas turned on at meter?
6 - Is manual main shut-off valve open?
7 - Is internal manual shut-off open?
8 - Are intake and exhaust pipes clogged?
9 - Is primary control locked out? (Turn thermostat off and then back on.)
10 - Is unit locked out on secondary limit? (Secondary limit is manually reset.)

GSR14 SERIES UNITS

LENNOX Industries Inc.

Figure 15-19 Electrical start-up adjustments, maintenance, and repair list for the GSR-14Q. (Courtesy of Lennox Industries Inc.)

ADJUSTMENTS (CONT.)

GAS FLOW

To check proper gas flow to combustion chamber, determine BTU input from the appliance rating plate. Divide this input rating by the BTU per cubic foot of available gas. Result is the number of cubic feet per hour required. Determine the flow of gas through gas meter for 2 minutes and multiply by 30 to get the hourly flow of gas to burner.

GAS PRESSURE

1 - Check gas line pressure with unit firing at maximum rate. Normal natural gas inlet line pressure should be 7.0 in. (178 mm) w.c. Normal line pressure for LP gas is 11 in. (280 mm) w.c.
IMPORTANT - Minimum gas supply pressure is listed on unit rating plate for normal input. Operation below minimum pressure may cause nuisance lockouts.

2 - After line pressure is checked and adjusted, check regulator pressure. Correct manifold pressure (unit running) is specified on nameplate. To measure, connect gauge to pressure tap in elbow below expansion tank.

HEAT ANTICIPATOR SETTINGS

Units with White Rodgers gas valves — 0.9

FAN/LIMIT CONTROL

Limit Control — Factory set: No adjustment necessary
Fan Control — Factory set: ON — No adjustment necessary
OFF — 90°F (32°C)

TEMPERATURE RISE AND EXTERNAL STATIC PRESSURE

Check temperature rise and external static pressure. If necessary, adjust blower speed to maintain temperature rise and external static pressure within range shown on unit rating plate.

ELECTRICAL

1 - Check all wiring for loose connections.
2 - Check for correct voltage at unit (unit operating).
3 - Check amp-draw on blower motor.
Motor Nameplate__________Actual__________

NOTE - Do not secure electrical conduit directly to duct work or structure.

BLOWER SPEEDS

Multi-tap drive motors are wired for different heating and cooling speeds. Speed may be changed by simply interchanging motor connections at indoor blower relay and fan control. Refer to speed selection chart on unit wiring diagram.

CAUTION - To prevent motor burnout, never connect more than one (1) motor lead to any one connection. Tape unused motor leads separately.

MAINTENANCE

NOTE - Disconnect power before servicing.

ANNUAL MAINTENANCE

At the beginning of each heating season, system should be checked by qualified serviceman as follows:

A - Blower
1 - Check and clean blower wheel.
2 - Check motor lubrication.
Always relubricate motor according to manufacturer's lubrication instructions on each motor. If no instructions are provided, use the following as a guide:
  a - Motors Without Oiling Ports — Prelubricated and sealed. No further lubrication required.
  b - Direct Drive Motors with Oiling Ports — Prelubricated for an extended period of operation. For extended bearing life, relubricate with a few drops of SAE No. 10 non-detergent oil once every two years. It may be necessary to remove blower assembly for access to oiling ports.

B - Electrical
1 - Check all wiring for loose connections.
2 - check for correct voltage at unit (unit operating).
3 - Check amp-draw on blower motor.
Motor nameplate ________ Actual ________
4 - Check to see that heat tape (if applicable) is operating.

C - Filters
1 - Filters must be cleaned or replaced when dirty to assure proper furnace operation.
2 - Reusable foam filters supplied with GSR14 can be washed with water and mild detergent. They should be sprayed with filter handicoater when dry prior to reinstallation. Filter handicoater is RP products coating No. 481 and is available as Lennox part No. P-8-5069.
3 - If replacement is necessary, order Lennox part No. P-8-7831 for 20 X 25 inch (508 X 635 mm) filter.

D - Intake and Exhaust Lines

Check intake and exhaust PVC lines and all connections for tightness and make sure there is no blockage. Also check condensate line for free flow during operation.

E - Insulation

Outdoor piping insulation should be inspected yearly for deterioration. If necessary, replace with same materials.

Figure 15-19 *(Continued)*

REPAIR PARTS LIST

The following repair parts are available through independent Lennox dealers. When ordering parts, include the complete furnace model number listed on unit rating plate. Example: GSR14Q3-50-1.

CABINET PARTS
Blower access panel
Control access panel
Upper vestibule panel
Lower vestibule panel
Control box cover

CONTROL PANEL PARTS
Transformer
Indoor blower relay
Low voltage terminal strip
High voltage terminal strip

BLOWER PARTS
Blower wheel
Motor
Motor mounting frame
Motor capacitor
Blower housing cut-off plate
Blower housing

HEATING PARTS
Heat exchanger assembly
Gas orifice
Gas valve
Gas decoupler
Gas flapper valve
Purge blower
Air intake flapper valve
Primary control board
Ignition lead
Flame sensor lead
Flame sensor
Primary fan and limit control
Secondary limit control
Auxiliary fan control
Differential pressure switch
Door interlock switch
Air filter

Figure 15-19 *(Continued)*

TROUBLESHOOTING

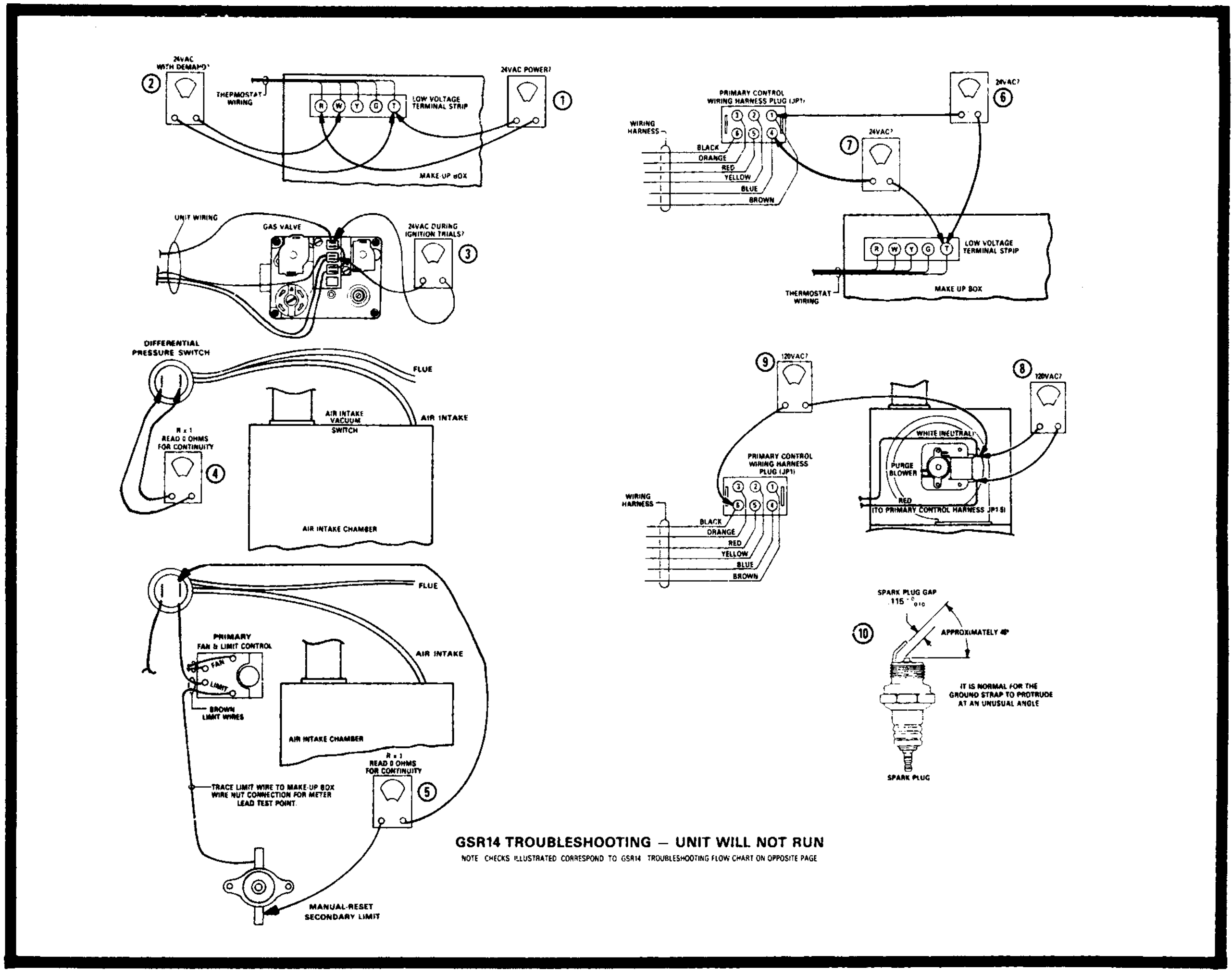
② 24VAC WITH DEMAND?
THERMOSTAT WIRING
R W Y G T
LOW VOLTAGE TERMINAL STRIP
24VAC POWER? ①
MAKE-UP BOX
UNIT WIRING
GAS VALVE
24VAC DURING IGNITION TRIALS? ③
DIFFERENTIAL PRESSURE SWITCH
FLUE
AIR INTAKE VACUUM SWITCH
AIR INTAKE
R x 1 READ 0 OHMS FOR CONTINUITY ④
AIR INTAKE CHAMBER
FLUE
PRIMARY FAN & LIMIT CONTROL
FAN
LIMIT
BROWN LIMIT WIRES
AIR INTAKE
AIR INTAKE CHAMBER
R x 1 READ 0 OHMS FOR CONTINUITY ⑤
TRACE LIMIT WIRE TO MAKE-UP BOX WIRE NUT CONNECTION FOR METER LEAD TEST POINT.
MANUAL-RESET SECONDARY LIMIT
PRIMARY CONTROL WIRING HARNESS PLUG (JP1)
WIRING HARNESS
3 2 1 6 5 4
BLACK
ORANGE
RED
YELLOW
BLUE
BROWN
24VAC? ⑥
⑦ 24VAC?
LOW VOLTAGE TERMINAL STRIP
THERMOSTAT WIRING
MAKE UP BOX
⑨ 120VAC?
⑧ 120VAC?
PRIMARY CONTROL WIRING HARNESS PLUG (JP1)
WIRING HARNESS
WHITE (NEUTRAL)
PURGE BLOWER
RED
(TO PRIMARY CONTROL HARNESS JP1-5)
SPARK PLUG GAP .115 +.0 -.010
⑩
APPROXIMATELY 45°
IT IS NORMAL FOR THE GROUND STRAP TO PROTRUDE AT AN UNUSUAL ANGLE
SPARK PLUG

GSR14 TROUBLESHOOTING — UNIT WILL NOT RUN

NOTE CHECKS ILLUSTRATED CORRESPOND TO GSR14 TROUBLESHOOTING FLOW CHART ON OPPOSITE PAGE

Figure 15-20 Troubleshooting the GSR-14Q furnace with a meter. (Courtesy of Lennox)

TROUBLESHOOTING (CONT.)

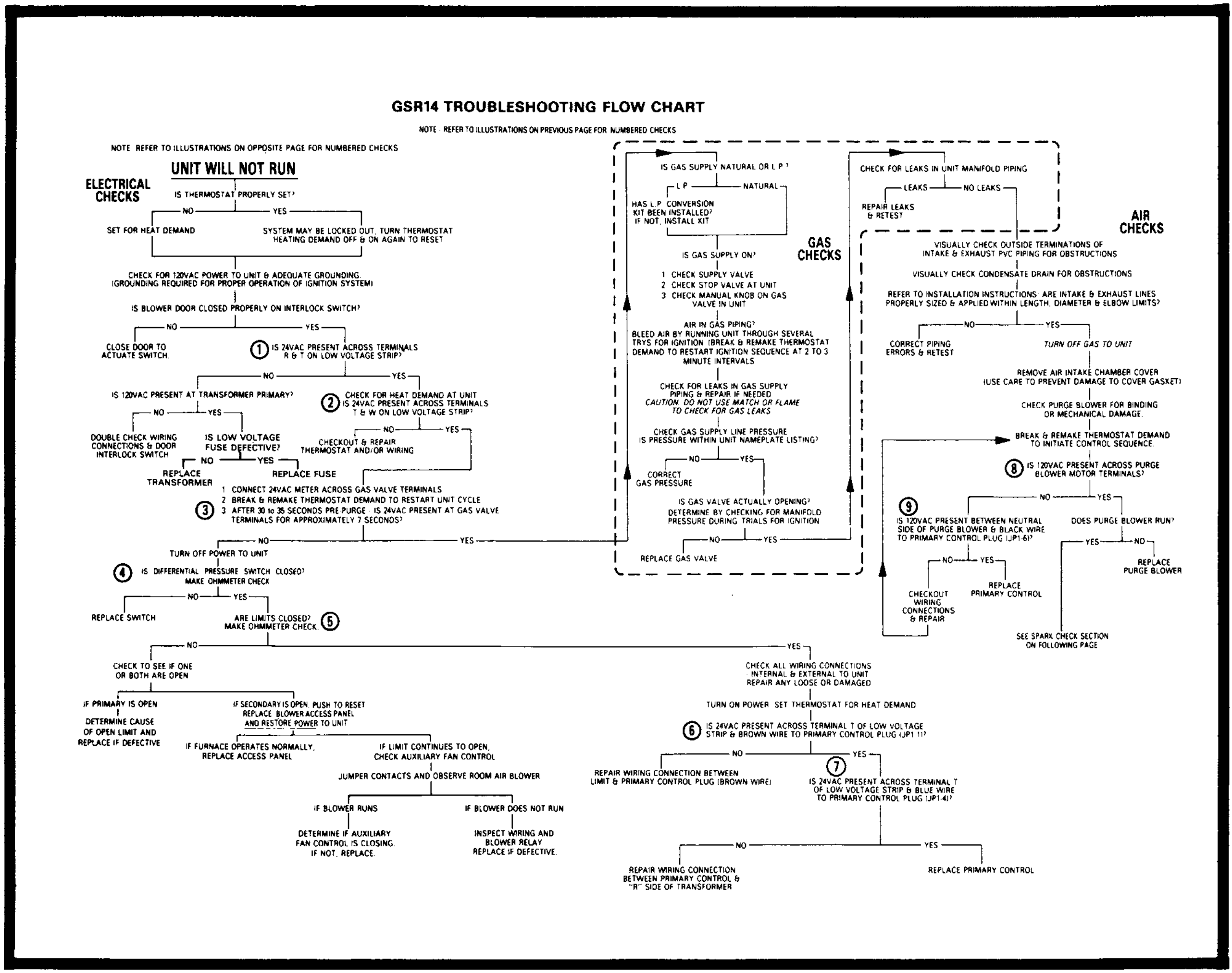
GSR14 TROUBLESHOOTING FLOW CHART
NOTE · REFER TO ILLUSTRATIONS ON PREVIOUS PAGE FOR NUMBERED CHECKS
NOTE REFER TO ILLUSTRATIONS ON OPPOSITE PAGE FOR NUMBERED CHECKS
UNIT WILL NOT RUN
ELECTRICAL CHECKS
IS THERMOSTAT PROPERLY SET?
NO
YES
SET FOR HEAT DEMAND
SYSTEM MAY BE LOCKED OUT, TURN THERMOSTAT HEATING DEMAND OFF & ON AGAIN TO RESET
CHECK FOR 120VAC POWER TO UNIT & ADEQUATE GROUNDING. (GROUNDING REQUIRED FOR PROPER OPERATION OF IGNITION SYSTEM)
IS BLOWER DOOR CLOSED PROPERLY ON INTERLOCK SWITCH?
NO
YES
CLOSE DOOR TO ACTUATE SWITCH.
① IS 24VAC PRESENT ACROSS TERMINALS R & T ON LOW VOLTAGE STRIP?
NO
YES
IS 120VAC PRESENT AT TRANSFORMER PRIMARY?
NO
YES
DOUBLE CHECK WIRING CONNECTIONS & DOOR INTERLOCK SWITCH
IS LOW VOLTAGE FUSE DEFECTIVE?
NO
YES
REPLACE TRANSFORMER
REPLACE FUSE
② CHECK FOR HEAT DEMAND AT UNIT IS 24VAC PRESENT ACROSS TERMINALS T & W ON LOW VOLTAGE STRIP?
NO
YES
CHECKOUT & REPAIR THERMOSTAT AND/OR WIRING
③ 1 CONNECT 24VAC METER ACROSS GAS VALVE TERMINALS
2 BREAK & REMAKE THERMOSTAT DEMAND TO RESTART UNIT CYCLE
3 AFTER 30 to 35 SECONDS PRE-PURGE · IS 24VAC PRESENT AT GAS VALVE TERMINALS FOR APPROXIMATELY 7 SECONDS?
NO
YES
TURN OFF POWER TO UNIT
④ IS DIFFERENTIAL PRESSURE SWITCH CLOSED? MAKE OHMMETER CHECK
NO
YES
REPLACE SWITCH
ARE LIMITS CLOSED? MAKE OHMMETER CHECK. ⑤
NO
YES
CHECK TO SEE IF ONE OR BOTH ARE OPEN
IF PRIMARY IS OPEN
DETERMINE CAUSE OF OPEN LIMIT AND REPLACE IF DEFECTIVE
IF SECONDARY IS OPEN, PUSH TO RESET REPLACE BLOWER ACCESS PANEL AND RESTORE POWER TO UNIT
IF FURNACE OPERATES NORMALLY, REPLACE ACCESS PANEL
IF LIMIT CONTINUES TO OPEN, CHECK AUXILIARY FAN CONTROL
JUMPER CONTACTS AND OBSERVE ROOM AIR BLOWER
IF BLOWER RUNS
IF BLOWER DOES NOT RUN
DETERMINE IF AUXILIARY FAN CONTROL IS CLOSING. IF NOT, REPLACE.
INSPECT WIRING AND BLOWER RELAY REPLACE IF DEFECTIVE.
GAS CHECKS
IS GAS SUPPLY NATURAL OR L P ?
L P
NATURAL
HAS L.P. CONVERSION KIT BEEN INSTALLED? IF NOT, INSTALL KIT
IS GAS SUPPLY ON?
1 CHECK SUPPLY VALVE
2 CHECK STOP VALVE AT UNIT
3 CHECK MANUAL KNOB ON GAS VALVE IN UNIT
AIR IN GAS PIPING? BLEED AIR BY RUNNING UNIT THROUGH SEVERAL TRYS FOR IGNITION (BREAK & REMAKE THERMOSTAT DEMAND TO RESTART IGNITION SEQUENCE AT 2 TO 3 MINUTE INTERVALS
CHECK FOR LEAKS IN GAS SUPPLY PIPING & REPAIR IF NEEDED
CAUTION. DO NOT USE MATCH OR FLAME TO CHECK FOR GAS LEAKS
CHECK GAS SUPPLY LINE PRESSURE IS PRESSURE WITHIN UNIT NAMEPLATE LISTING?
NO
YES
CORRECT GAS PRESSURE
IS GAS VALVE ACTUALLY OPENING? DETERMINE BY CHECKING FOR MANIFOLD PRESSURE DURING TRIALS FOR IGNITION
NO
YES
REPLACE GAS VALVE
CHECK FOR LEAKS IN UNIT MANIFOLD PIPING
LEAKS
NO LEAKS
REPAIR LEAKS & RETEST
AIR CHECKS
VISUALLY CHECK OUTSIDE TERMINATIONS OF INTAKE & EXHAUST PVC PIPING FOR OBSTRUCTIONS
VISUALLY CHECK CONDENSATE DRAIN FOR OBSTRUCTIONS
REFER TO INSTALLATION INSTRUCTIONS: ARE INTAKE & EXHAUST LINES PROPERLY SIZED & APPLIED WITHIN LENGTH, DIAMETER & ELBOW LIMITS?
NO
YES
CORRECT PIPING ERRORS & RETEST
TURN OFF GAS TO UNIT
REMOVE AIR INTAKE CHAMBER COVER (USE CARE TO PREVENT DAMAGE TO COVER GASKET)
CHECK PURGE BLOWER FOR BINDING OR MECHANICAL DAMAGE.
BREAK & REMAKE THERMOSTAT DEMAND TO INITIATE CONTROL SEQUENCE.
⑧ IS 120VAC PRESENT ACROSS PURGE BLOWER MOTOR TERMINALS?
NO
YES
⑨ IS 120VAC PRESENT BETWEEN NEUTRAL SIDE OF PURGE BLOWER & BLACK WIRE TO PRIMARY CONTROL PLUG (JP1-6)?
NO
YES
CHECKOUT WIRING CONNECTIONS & REPAIR
REPLACE PRIMARY CONTROL
DOES PURGE BLOWER RUN?
YES
NO
REPLACE PURGE BLOWER
SEE SPARK CHECK SECTION ON FOLLOWING PAGE
CHECK ALL WIRING CONNECTIONS · INTERNAL & EXTERNAL TO UNIT REPAIR ANY LOOSE OR DAMAGED
TURN ON POWER SET THERMOSTAT FOR HEAT DEMAND
⑥ IS 24VAC PRESENT ACROSS TERMINAL T OF LOW VOLTAGE STRIP & BROWN WIRE TO PRIMARY CONTROL PLUG (JP1-1)?
NO
YES
REPAIR WIRING CONNECTION BETWEEN LIMIT & PRIMARY CONTROL PLUG (BROWN WIRE)
⑦ IS 24VAC PRESENT ACROSS TERMINAL T OF LOW VOLTAGE STRIP & BLUE WIRE TO PRIMARY CONTROL PLUG (JP1-4)?
NO
YES
REPAIR WIRING CONNECTION BETWEEN PRIMARY CONTROL & "R" SIDE OF TRANSFORMER
REPLACE PRIMARY CONTROL

Figure 15-20 (*Continued*)

TROUBLESHOOTING (CONT.)

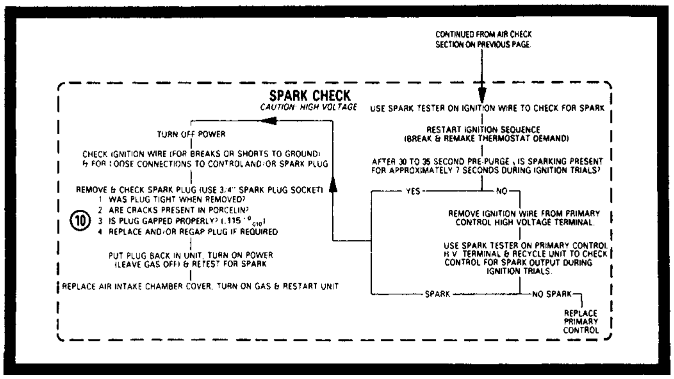
CONTINUED FROM AIR CHECK SECTION ON PREVIOUS PAGE.
SPARK CHECK
CAUTION: HIGH VOLTAGE
USE SPARK TESTER ON IGNITION WIRE TO CHECK FOR SPARK
RESTART IGNITION SEQUENCE (BREAK & REMAKE THERMOSTAT DEMAND)
AFTER 30 TO 35 SECOND PRE-PURGE, IS SPARKING PRESENT FOR APPROXIMATELY 7 SECONDS DURING IGNITION TRIALS?
YES
NO
REMOVE IGNITION WIRE FROM PRIMARY CONTROL HIGH VOLTAGE TERMINAL.
USE SPARK TESTER ON PRIMARY CONTROL H.V. TERMINAL & RECYCLE UNIT TO CHECK CONTROL FOR SPARK OUTPUT DURING IGNITION TRIALS.
SPARK
NO SPARK
REPLACE PRIMARY CONTROL
TURN OFF POWER
CHECK IGNITION WIRE (FOR BREAKS OR SHORTS TO GROUND) & FOR LOOSE CONNECTIONS TO CONTROL AND/OR SPARK PLUG
REMOVE & CHECK SPARK PLUG (USE 3/4" SPARK PLUG SOCKET)
⑩
1 WAS PLUG TIGHT WHEN REMOVED?
2 ARE CRACKS PRESENT IN PORCELIN?
3 IS PLUG GAPPED PROPERLY? (.115 +0 -.010)
4 REPLACE AND/OR REGAP PLUG IF REQUIRED
PUT PLUG BACK IN UNIT, TURN ON POWER (LEAVE GAS OFF) & RETEST FOR SPARK
REPLACE AIR INTAKE CHAMBER COVER, TURN ON GAS & RESTART UNIT

Figure 15-20 (*Continued*)

ELECTRICAL

1 - Select fuse and wire size according to blower motor amps.
2 - Access holes are provided on both sides of cabinet to facilitate wiring.
3 - Install room thermostat according to instruction provided with thermostat.
4 - Install a separate fused disconnect switch near the unit so the power can be turned off for servicing.
5 - Complete wiring connections to equipment using provided wiring diagrams.
6 - Electrically ground unit in accordance with local codes, or in the absence of local codes in accordance with the CSA Standards.
7 - Seal unused electrical openings with snap-plugs provided.

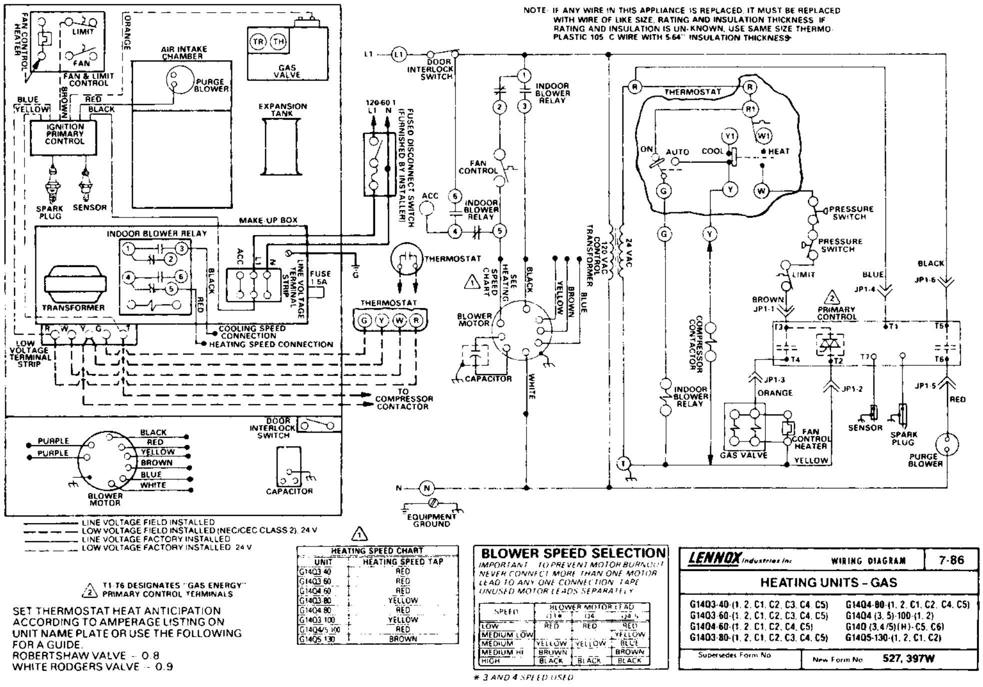
FAN CONTROL HEATER
LIMIT
FAN
FAN & LIMIT CONTROL
ORANGE
AIR INTAKE CHAMBER
PURGE BLOWER
TR
TH
GAS VALVE
EXPANSION TANK
BLUE
YELLOW
BROWN
RED
BLACK
IGNITION PRIMARY CONTROL
SPARK PLUG
SENSOR
MAKE-UP BOX
INDOOR BLOWER RELAY
TRANSFORMER
ACC
L1
N
LINE VOLTAGE TERMINAL STRIP
FUSE 1.5A
G
LOW VOLTAGE TERMINAL STRIP
R W Y G
COOLING SPEED CONNECTION
HEATING SPEED CONNECTION
120-60-1
FUSED DISCONNECT SWITCH (FURNISHED BY INSTALLER)
THERMOSTAT
TO COMPRESSOR CONTACTOR
DOOR INTERLOCK SWITCH
PURPLE
PURPLE
BLACK
RED
YELLOW
BROWN
BLUE
WHITE
BLOWER MOTOR
CAPACITOR
L1
DOOR INTERLOCK SWITCH
INDOOR BLOWER RELAY
FAN CONTROL
ACC
SEE HEATING SPEED CHART
BLOWER MOTOR
CAPACITOR
120 VAC CONTROL TRANSFORMER
24 VAC
THERMOSTAT
R
R1
Y1
W1
ON
AUTO
COOL
HEAT
G
Y
W
PRESSURE SWITCH
PRESSURE SWITCH
LIMIT
BROWN
JP1-1
BLUE
JP1-4
BLACK
JP1-6
PRIMARY CONTROL
T1 T2 T3 T4 T5 T6 T7
COMPRESSOR CONTACTOR
INDOOR BLOWER RELAY
JP1-3
ORANGE
JP1-2
JP1-5
RED
GAS VALVE
FAN CONTROL HEATER
YELLOW
SENSOR
SPARK PLUG
PURGE BLOWER
T
N
EQUIPMENT GROUND

NOTE: IF ANY WIRE IN THIS APPLIANCE IS REPLACED, IT MUST BE REPLACED WITH WIRE OF LIKE SIZE, RATING AND INSULATION THICKNESS. IF RATING AND INSULATION IS UN-KNOWN, USE SAME SIZE THERMOPLASTIC 105° C WIRE WITH 5/64" INSULATION THICKNESS.

LINE VOLTAGE FIELD INSTALLED
LOW VOLTAGE FIELD INSTALLED (NEC/CEC CLASS 2), 24 V
LINE VOLTAGE FACTORY INSTALLED
LOW VOLTAGE FACTORY INSTALLED 24 V

△2 T1-T6 DESIGNATES "GAS ENERGY" PRIMARY CONTROL TERMINALS

SET THERMOSTAT HEAT ANTICIPATION ACCORDING TO AMPERAGE LISTING ON UNIT NAME PLATE OR USE THE FOLLOWING FOR A GUIDE.
ROBERTSHAW VALVE -- 0.8
WHITE RODGERS VALVE -- 0.9

△1 HEATING SPEED CHART

UNIT	HEATING SPEED TAP
G14Q3-40	RED
G14Q3-60	RED
G14Q4-60	RED
G14Q3-80	YELLOW
G14Q4-80	RED
G14Q3-100	YELLOW
G14Q4/5-100	RED
G14Q5-130	BROWN

BLOWER SPEED SELECTION

IMPORTANT — TO PREVENT MOTOR BURNOUT NEVER CONNECT MORE THAN ONE MOTOR LEAD TO ANY ONE CONNECTION. TAPE UNUSED MOTOR LEADS SEPARATELY.

SPEED	BLOWER MOTOR LEAD		
	Q3*	Q4	Q4-5
LOW	RED	RED	RED
MEDIUM LOW			YELLOW
MEDIUM	YELLOW	YELLOW	BLUE
MEDIUM HI	BROWN		BROWN
HIGH	BLACK	BLACK	BLACK

** 3 AND 4 SPEED USED*

LENNOX Industries Inc	WIRING DIAGRAM	7-86
HEATING UNITS - GAS		
G14Q3-40-(1, 2, C1, C2, C3, C4, C5) G14Q3-60-(1, 2, C1, C2, C3, C4, C5) G14Q4-60-(1, 2, C1, C2, C4, C5) G14Q3-80-(1, 2, C1, C2, C3, C4, C5)	G14Q4-80-(1, 2, C1, C2, C4, C5) G14Q4 (3, 5)-100-(1, 2) G14Q (3,4/5)(H)-C5, C6) G14Q5-130-(1, 2, C1, C2)	
Supersedes Form No	New Form No 527, 397W	

Figure 15-21 Electrical wiring for the G-14Q series furnaces. (Courtesy of Lennox)

Review Questions

1. What are the parts of a basic electric heating system?
2. What is a ladder diagram? Why is it useful?
3. What is the source of schematics for specific types of equipment?
4. What is the meaning of the term field wiring?
5. What is the usual low voltage for a heating system?
6. Where should the thermostat be located in a home?
7. What is the advantage of a printed circuit board for a furnace?
8. How is a heat pump different from a furnace?
9. What does a heat pump do that a furnace cannot do?
10. How can a refrigerant be used to generate heat?
11. List four basic types of heat pumps.
12. Why are supplemental heaters sometimes needed on heat pumps?
13. How is the gas/air mixture ignited in a Pulse furnace?
14. Why are these new furnaces so efficient?
15. Where is the electronics package located in the Pulse furnace?

Chapter

16

Air-Conditioning Circuits

Performance Objectives

- Know what is meant by modes of cooling.
- Know where the temperature sensor is located on an air conditioner.
- Know how to adjust the fan speed on an air conditioner.
- Know what an auxiliary winding does.
- Know the purpose of a run capacitor in the compressor circuit.
- Know how to read a ladder diagram.
- Know the function of the AUTO and ON positions of a thermostat.
- Know how the overload protector works in an air conditioner compressor motor.
- Know how to troubleshoot an air conditioner using a troubleshooting chart.

Circuits for air-conditioning units vary slightly with the manufacturer. The best source for schematics of the electrical system is the manufacturer's service manuals and bulletins. In this chapter we will be concerned primarily with the circuits of domestic (home) air-conditioning equipment.

Basic Air-Conditioning Unit

The basic or simple air-conditioning unit consists of a plug, a thermostat, overload protector, compressor, condenser, evaporator, capacitors for the motors, and some kind of switching for control of the fan motor.

The shaft on the fan motor usually extends from both ends of the motor. A fan blade is attached to one end to cool the condenser, and another fan blade is attached to the other end of the motor to cause air to flow over the evaporator

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

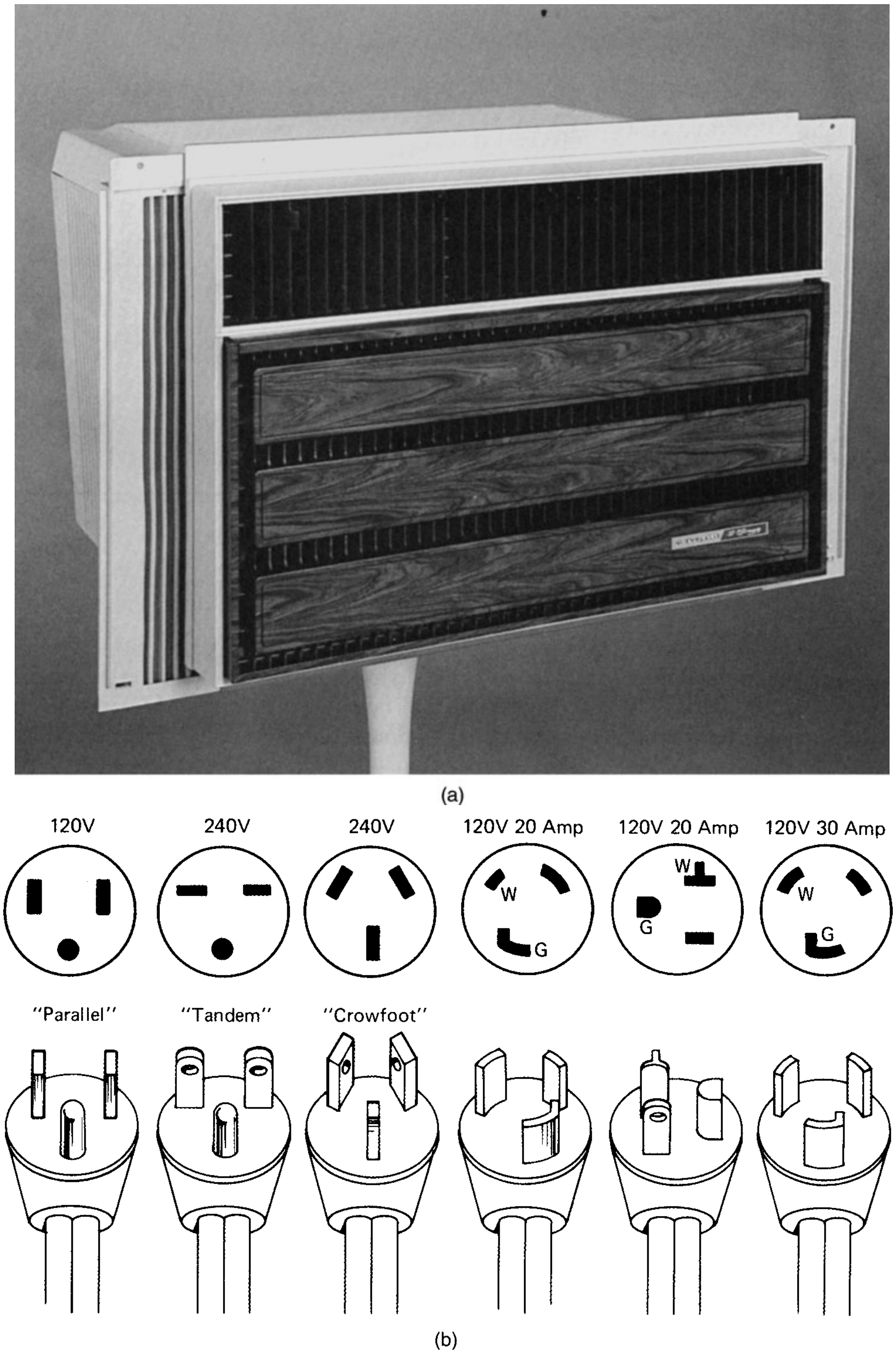
(a)
120V
240V
240V
120V 20 Amp
120V 20 Amp
120V 30 Amp
W
G
W
G
W
G
"Parallel"
"Tandem"
"Crowfoot"
(b)

Figure 16-1 (a) Window air conditioner. (b) Line cords and plugs for various voltages used in air conditioning units.

and cool the room. This way only one motor is used. This is the type of arrangement most often encountered in the window air conditioner.

The air-conditioning unit that fits into the window (see Fig. 16-1) usually has at least two speeds. The fan is used in some cases without the compressor running. Thus, there are three modes of cooling: (1) fan only, (2) fan at high speed and cooling furnished by the compressor, and (3) fan at low speed and cooling furnished by the compressor. This basic window unit consists of the compressor, the fan motor, the thermostat, and push buttons for the selection of the type of cooling desired (see Fig. 16-2).

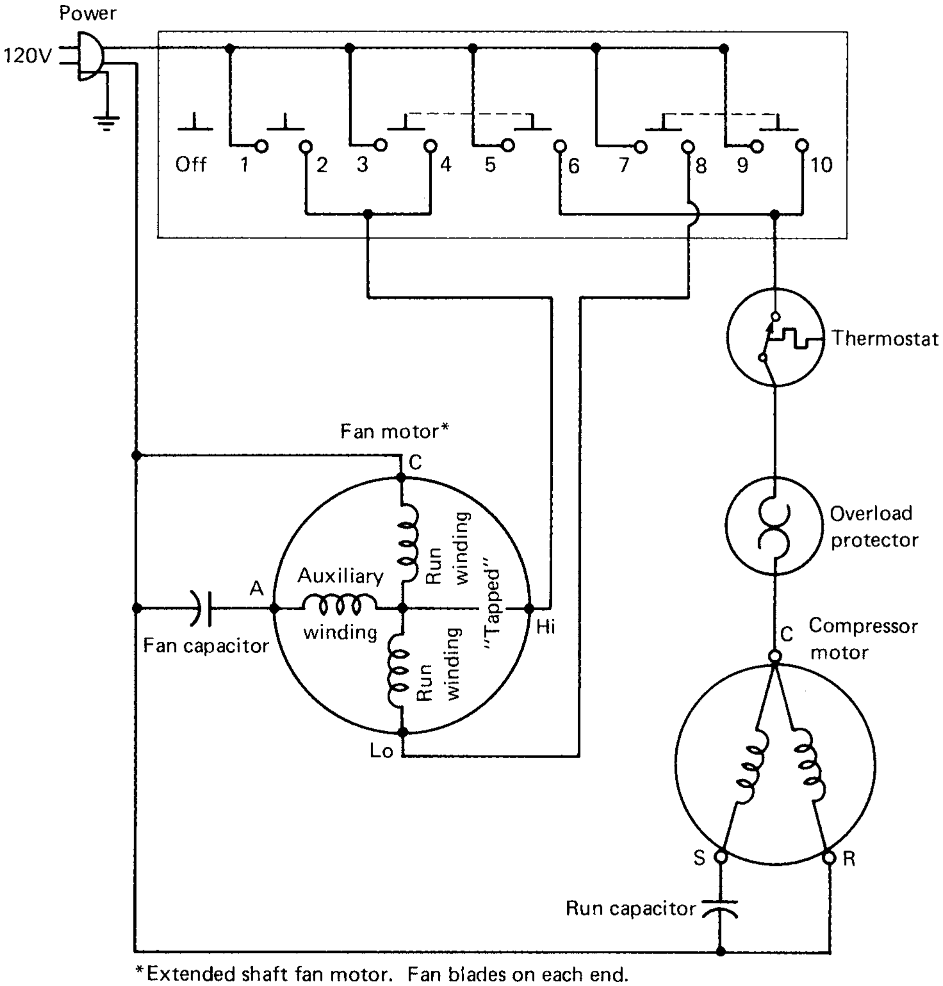
Power
120V
Off
1
2
3
4
5
6
7
8
9
10
Thermostat
Fan motor*
C
Run winding
Auxiliary
winding
A
"Tapped"
Hi
Fan capacitor
Run winding
Lo
Overload protector
C
Compressor motor
S
R
Run capacitor

*Extended shaft fan motor. Fan blades on each end.
One fan for evaporator and one for condenser.

Figure 16-2 (a) Schematic for a two-speed, push-button-controlled window unit. (b) Location of component parts of a window unit.

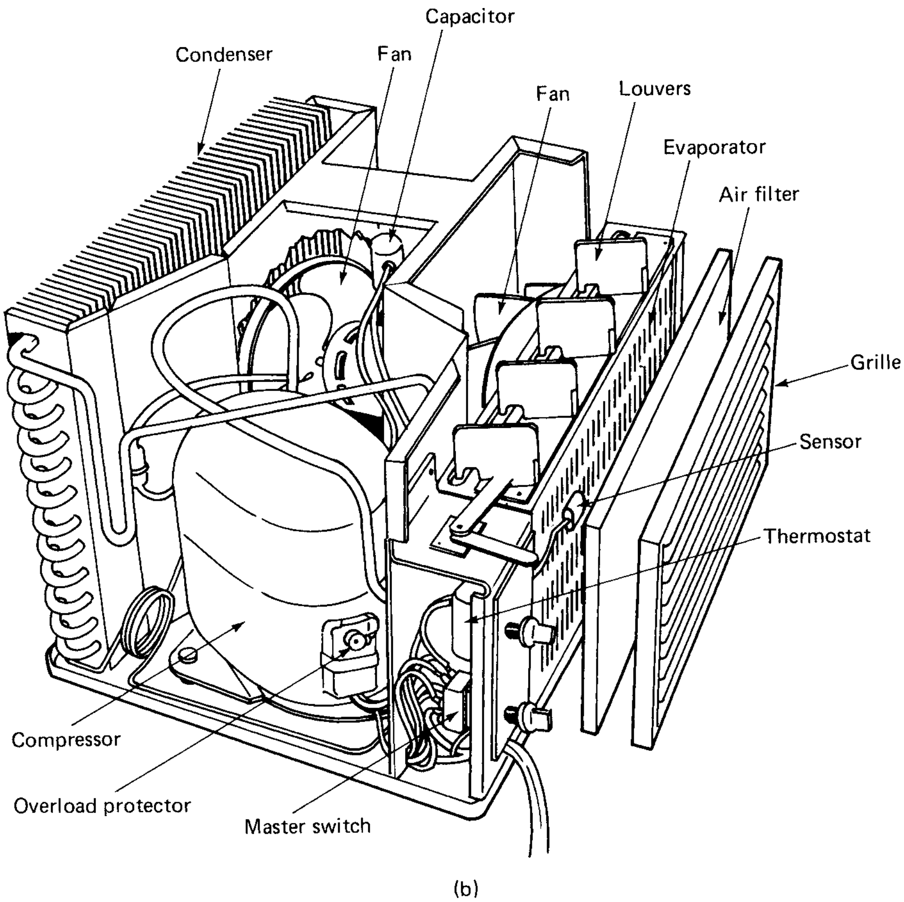
Capacitor
Condenser
Fan
Fan
Louvers
Evaporator
Air filter
Grille
Sensor
Thermostat
Compressor
Overload protector
Master switch
(b)

Figure 16-2 (*Continued*)

Push-button switches

The push-button switches that operate the window unit usually resemble those in Fig. 16-3. If the unit has a two-speed fan and a compressor that operates in conjunction with the high- and low-speed fan positions, it will have an OFF button that disengages the push buttons that are depressed, and thereby opens the circuit and turns off the power to both compressor and fan.

The thermostat is variable by twisting a knob. This moves the bimetallic strips closer together to provide the proper operation of the thermostat in relation to the room temperature. Thermostats can have a number of design configurations, depending on the manufacturer. The knobs differ slightly, but basically the thermostat is the same for all window units. The temperature sensor located on the inlet side of the evaporator senses the air returned to the air conditioner from the room being conditioned.

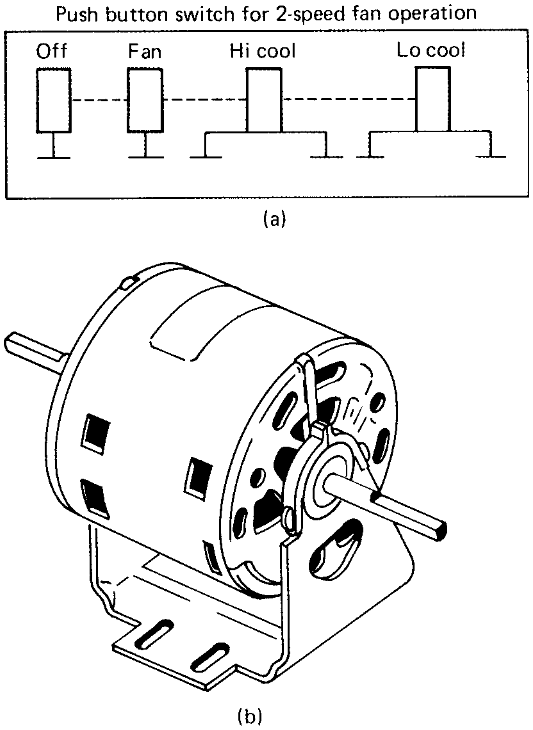

Figure 16-3 (a) Push-button switches used in a window unit for fan and compressor control. (b) Fan with double shaft. Used in window units with a blade on each end.

Fans

The fan is two speed. It has an auxiliary winding with a capacitor in series with it. The run windings are arranged so that when one of them is in the circuit it runs at high speed. When the two run windings are placed in series, it slows down and the fan speed is LO. This permanent split-capacitor (PSC) motor is commonly used in two- or three-speed configurations for fans.

Schematics

To be able to troubleshoot this type of window unit, it is necessary to know how each control functions and what will happen when each malfunctions. The best way to understand the unit is to look at the schematic in Fig. 16-2, and trace it out.

Tracing it out will result in Fig. 16-4. In Fig. 16-4a, you can see how one side of the power source is connected to the fan motor C and run capacitor. Then it goes to the run capacitor on the compressor and the R connection of the compressor. This completes the line for one side of the power. But, for anything to work electrically, two sides of the power line must be taken into consideration.

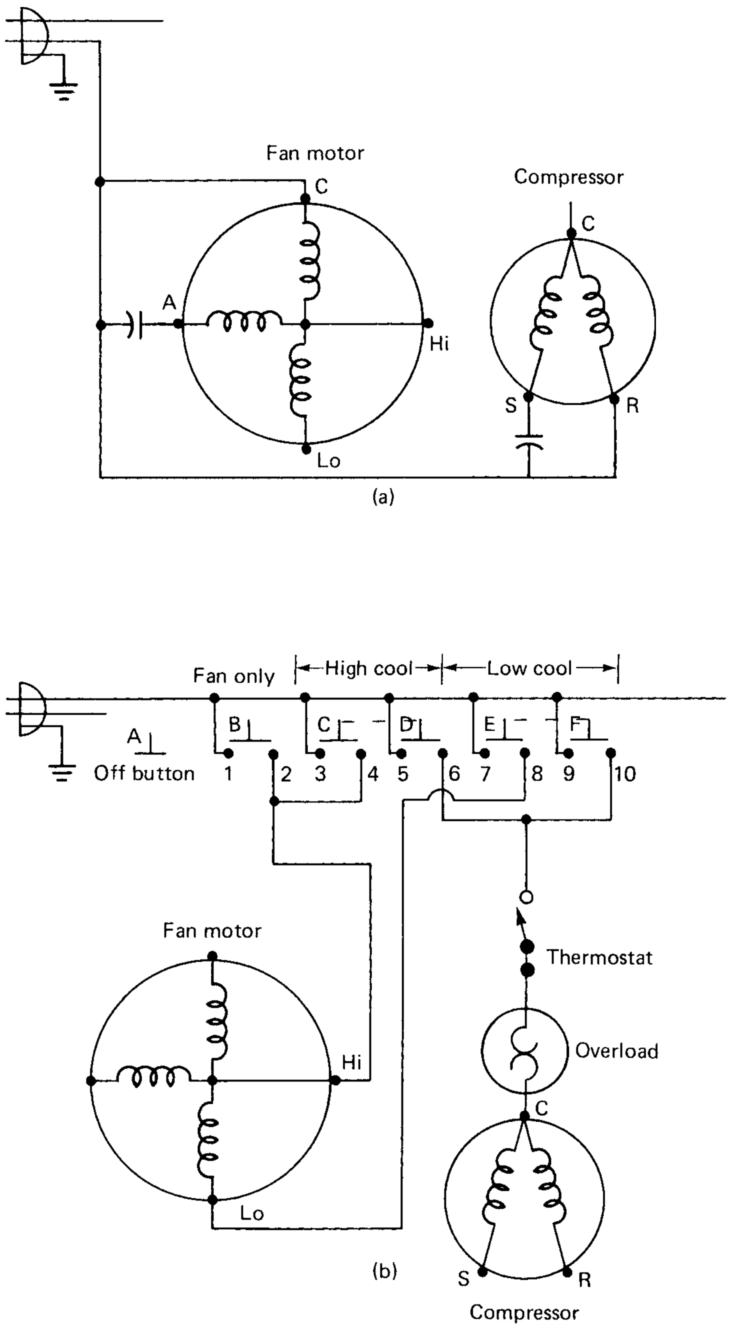
Fan motor
C
Compressor
C
A
Hi
S
R
Lo
(a)
Fan only
High cool
Low cool
A
Off button
B
C
D
E
F
1
2
3
4
5
6
7
8
9
10
Thermostat
Fan motor
Overload
Hi
C
Lo
(b)
S
R
Compressor

Figure 16-4 (a) One side of the power line is connected directly to all components in the unit. (b) The other side of the power line is switched and becomes the controlled power line. (c) Schematic for "fan only" operation of the unit. (d) Schematic for "hi cool" operation of the unit. (e) Schematic for "lo cool" operation of the unit.

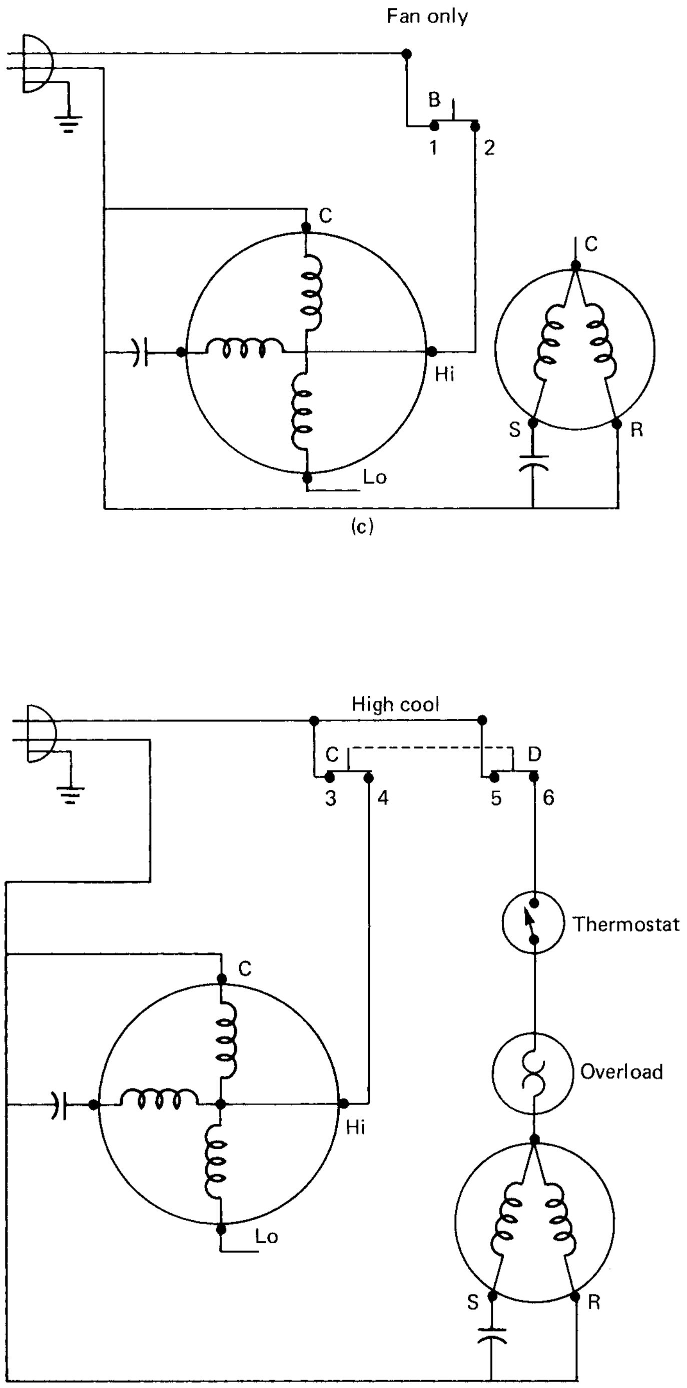
Fan only
B
1
2
C
Hi
Lo
C
S
R
High cool
C
D
3
4
5
6
Thermostat
Overload
C
Hi
Lo
S
R

(c)

(d)

Figure 16-4 (*Continued*)

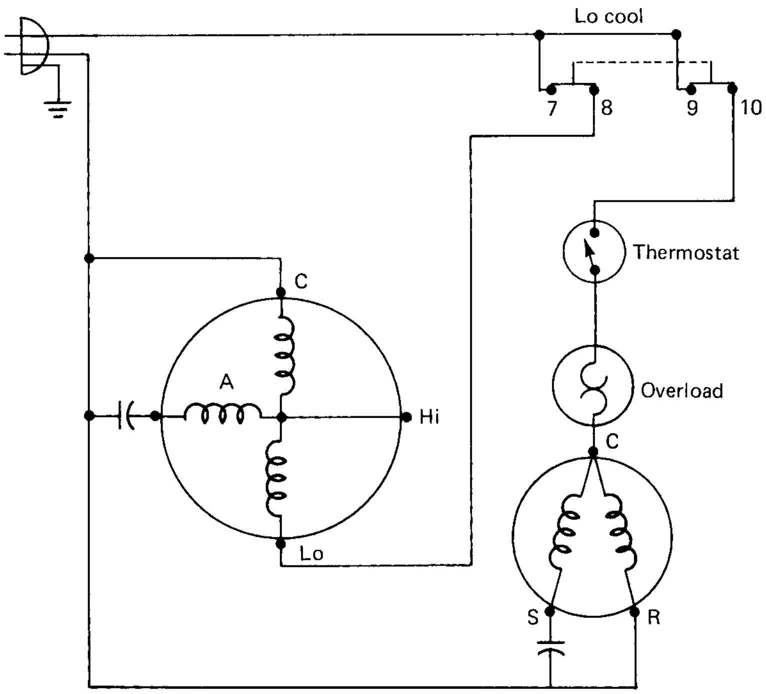

(e)

Figure 16-4 (*Continued*)

It takes both sides to cause the circuit to be complete. Now let's look at the other line as seen in Fig. 16-4b.

All the controls are in this other line. The push-button switch has an "off" button, a "fan only" button, a "hi cool" button, and a "lo cool" button. Each of these has a definite circuit that must be traced out to aid in troubleshooting later when possible switching problems develop.

Figure 16-4c shows the "fan only" circuit. Push-button B is depressed. It makes contact with 1 and 2 on the switch block. This completes the other side of the power line through the switch to the high connection of the fan motor. Thus, power is now available to cause the "hi winding" of the fan motor to be energized. The "aux winding" is also completed to the power line at this time, and the fan motor runs. However, this is the only thing operational in the window unit at this time. The fan is sometimes needed to circulate air in the room when a number of people are assembled. The fan only is often all that is needed to keep the room cool. Keep in mind that the thermostat is not in the circuit, so the fan will run until it is turned off by depressing the "off button."

Figure 16-4d shows how the compressor is operated with a demand for "hi cool." The hi cool push button is depressed and it completes the circuit from 3 and 4 to the hi side of the fan motor. The fan runs. Hi cool also is so arranged that the push button will complete the circuit from 5 to 6. The 5 and 6 combination completes the circuit to the thermostat and, if it is closed, on to the overload and to the compressor motor. The compressor will now run while the fan is also running. Once the thermostat opens (the room has cooled down to where the thermostat has been adjusted), the compressor is turned off. However, as long as the hi cool button is depressed, the fan motor will continue to operate. The compressor is cycled on and off by the demands of the thermostat. The fan will continue to run until the off button is depressed to mechanically push up the hi cool switch.

Figure 16-4e shows the completed circuit for the “lo cool” position of the push buttons. Note how terminals 7 and 8 are connected to complete the circuit to the lo position of the fan motor. This means that the two main run windings of the fan motor are now in series. This causes the fan motor to run at a slower speed than when only one winding is in the circuit.

Terminals 9 and 10 are used to complete the circuit to cause the compressor to operate. Here, again, the fan runs at lo speed all the time, whether the compressor operates or not. The compressor operates at the command of the thermostat.

To troubleshoot the unit, you should understand how it operates when all systems are functional. Then the symptoms can be traced to either the refrigerant system or electrical system.

Here we are primarily interested in the problems associated with the electrical system. Problems are shown in Table 16-1, which covers both systems. Note how many of the problems are electrical.

Window units with three-speed fans are also available. All you do is place windings in series with one another to reduce the fan speed. The push buttons do the same things as just explained for the two-speed unit. In larger compressors for 18,000 Btu units, a start potential relay is used to switch on the start winding. A start capacitor is placed in and out of the circuit as required, and the run capacitor stays in the circuit. This type of arrangement is discussed in detail in Chapter 17 for refrigerators.

Ladder Diagrams

Ladder diagrams make it easier to visualize the circuitry. The two power lines are drawn parallel with one another and the components are connected between the two lines to complete the various system circuits (see Fig. 16-5).

Another air-conditioning unit is shown in Fig. 16-5. It utilizes 220 to 240 volts and is wired into the line permanently. This type of unit usually becomes part of a hot-air type of furnace and provides cooling in the summer and heating in the winter. The evaporator is located in the plenum of the furnace, and the blower motor is used to move air over the evaporator during the air-conditioning season and move the hot air in the plenum from the furnace during the winter.

Thermostats

The thermostat is in the cool position. This means that when the room reaches a high temperature, the thermostat closes. Once the thermostat closes, the circuit to the compressor contactor (CC) relay is completed. The CC contacts are closed

TABLE 16-1 Troubleshooting Hermetic Compressors Used in Air Conditioners

Problem	Probable cause	Remedy
Compressor won't start.	1. Open line circuit.	1. Check the wiring, fuses, and receptacle.
There is no hum.	2. Protector open.	2. Check current drawn from the line after waiting for the reset to cool down.
	3. Contacts open on control relay.	3. Check the control and check the pressure readings.
	4. Open circuit in the motor stator.	4. Replace the stator or the whole compressor.
Compressor hums intermittently but will not start. However, it will cycle with the protector.	1. Incorrectly wired.	1. Check the wiring diagram and the actual wiring.
	2. Low line voltage.	2. Check the line voltage; find the cause of the dropping voltage and correct.
	3. Start capacitor could be open.	3. Replace with one known to be good.
	4. Contacts on relay do not close.	4. Check by manually operating the relay; replace if defective.
	5. Start winding could be open.	5. Check the stator leads; replace the compressor if the leads check out and are all right.
	6. Stator winding could be grounded. (This usually blows the fuse.)	6. Check stator leads; replace compressor if the leads are OK.
	7. Discharge pressure could be too high.	7. Remove the cause of excessive pressure; discharge shutoff and receiver valves should be open.
	8. Discharge pressure too high.	8. Compressor too tight.
	9. Start capacitor weak.	9. Replace the start capacitor.
Compressor starts, but the motor will not speed up enough to have start winding drop out of the circuit.	1. Low line voltage.	1. Increase the voltage.
	2. Compressor incorrectly wired.	2. Rewire according to wiring diagram.
	3. Defective relay.	3. Check operation; if defective, replace.
	4. Run capacitor shorted.	4. Disconnect the run capacitor and check for short.
	5. Start and run windings shorted.	5. Check winding resistance; if incorrect, replace the compressor.
	6. Start capacitor weak.	6. Check the capacitors; replace those found to be defective.
	7. Discharge pressure too high.	7. Check the discharge shutoff valves; check the pressure.
	8. Tight compressor.	8. Check the compressor oil level; check for binding; replace if necessary.
Compressor starts and runs, but it cycles on the protector.	1. Low line voltage.	1. Increase the voltage.
	2. Too much current being drawn through the protector.	2. Check to see if the fans or pumps are wired to the wrong connector.
	3. Suction pressure is too high.	3. Check the compressor; see if it is the right size for the job.

TABLE 16-1 Troubleshooting Hermetic Compressors Used in Air Conditioners (*Continued*)

Problem	Probable cause	Remedy
	4. Discharge pressure is too high.	4. Check the ventilation of the compressor; check for overcharge; also check for obstructions to airflow or refrigerant flow.
	5. Protector is weak.	5. Check the current; replace the protector if it is not clicking out at the right point.
	6. Defective run capacitor.	6. Check the capacitance; replace if found to be defective.
	7. Stator partially shorted or grounded.	7. Check the resistance for a short to the frame; replace if found shorted to the frame.
	8. Insufficient motor cooling.	8. Correct airflow.
	9. Compressor tight.	9. Check oil level; check the cause of the binding.
	10. Three-phase line unbalanced.	10. Check each leg or phase; correct if the voltages are not the same between legs.
	11. Discharge valve leaks or is damaged.	11. Replace the valve plate.
Start capacitors burn out.	1. Short cycling.	1. Reduce the number of starts; they should not exceed 20 per hour.
	2. Start winding left in circuit too long.	2. Reduce the starting load; install a crankcase pressure-limit valve; increase the low voltage if this is found to be the condition; replace the relay if it is found to be defective.
	3. Relay contacts are sticking.	3. Clean the relay contacts; or replace the relay.
	4. Wrong relay or wrong relay setting.	4. Replace the relay.
	5. Wrong capacitor.	5. Check the specs for correct-sized capacitor; be sure the MFD and the WVDC are correct for this compressor.
correct	6. Working voltage of the capacitor is too low.	6. Replace with a capacitor of voltage rating.
	7. Water shorts out the terminals of the capacitor.	7. Place the capacitor so the terminals will not get wet.
Run capacitors burn out. They spew their contents over the surfaces of anything nearby. This problem can usually be identified with a visual inspection.	1. Line voltage too high.	1. Reduce the line voltage; should not be over 10% of motor rating.
	2. Light load with a high line voltage.	2. Reduce voltage if not within the 10% overage limit.
	3. Voltage rating of the capacitor too low.	3. Replace with capacitors of the correct WVDC.
	4. Capacitor terminals shorted by water.	4. Place the capacitor so the terminals will not get wet.
Relays burn out.	1. Low line voltage causes high current drain.	1. Increase the voltage to within the 10% limit.
	2. High line voltage.	2. Reduce voltage to within 10% of the motor rating.
	3. Wrong-sized capacitor.	3. Use the correct-sized capacitor; the proper MFD rating should be

(*Continued*)

TABLE 16-1 Troubleshooting Hermetic Compressors Used in Air Conditioners (*Continued*)

Problem	Probable cause	Remedy
	4. Short cycling.	4. Decrease the number of starts per hour.
	5. Relay vibrates.	5. Make sure you mount the relay rigidly.
	6. Wrong relay.	6. Use the recommended relay for the compressor motor.

There are problems of a general nature. They are identified with hermetic-type compressors.

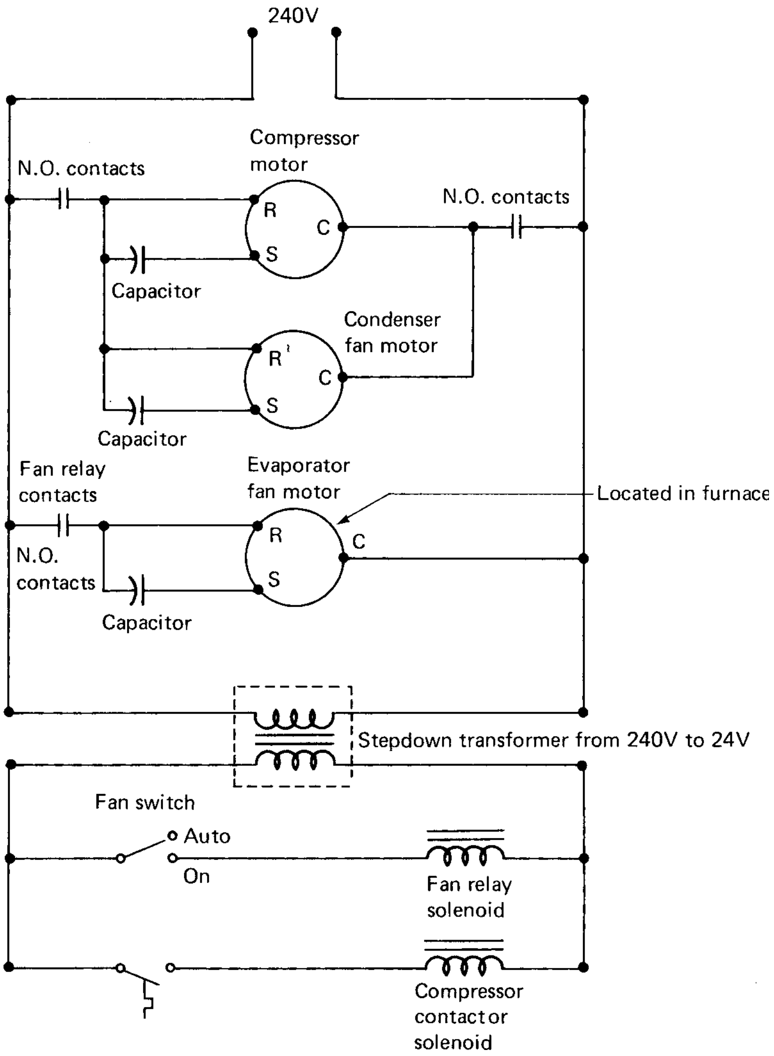
240V
Compressor motor
N.O. contacts
R
C
S
N.O. contacts
Capacitor
Condenser fan motor
R'
C
S
Capacitor
Fan relay contacts
Evaporator fan motor
R
C
S
Located in furnace
N.O. contacts
Capacitor
Stepdown transformer from 240V to 24V
Fan switch
Auto
On
Fan relay solenoid
Compressor contactor solenoid

Figure 16-5 Ladder diagram for an air-conditioning unit.

and complete the circuit to the compressor and condenser. The compressor begins to pump the refrigerant, and the fan starts to cool the condenser. Note how the condenser fan motor is connected in parallel with the compressor. The condenser is on when the compressor is on and off when the compressor is off. Both are controlled by the CC contacts. Opening of the thermostat causes the compressor to stop running, and the fan motor for the condenser also stops until needed again when the thermostat switch closes.

Fan switches

The fan switch has two positions: “auto” and “on”. When it is in the on position, it means the fan relay solenoid is energized and the fan runs continually until the switch is returned to auto. When the fan switch is in the auto position, it works in conjunction with the thermostat. When the thermostat closes, the fan switch closes. Thus, it starts the fan only when the compressor is operating. However, if you wish to move the air in the room without the benefit of the cooling action of the compressor, you may do so by manually moving the switch to the on position.

Note that the fan relay contacts have to be closed in order for the evaporator fan motor to operate. These relay contacts close when the fan relay solenoid energizes. This solenoid and the compressor contactor solenoid are both powered from a 24-volt transformer that provides the low voltage necessary for their operation. This low-voltage control also removes the possibility of shock for persons operating the controls.

Troubleshooting

Problems in air conditioners can be caused by a number of malfunctioning parts. We have already mentioned how the thermostat, overload protector, start relay on larger units, and fan motor work. Knowing how they work means a lot in troubleshooting. If you know how they work, you can more easily identify why they are not working.

Throughout the industry, you will find Problem-Probable Cause-Remedy charts made available by manufacturers of equipment. These are an aid to troubleshooting. They help in locating the possible trouble component or system. Then you need to identify the schematic and read it properly to be able to identify the problems associated with the electrical system.

Table 16-1 is a Problem-Probable Cause-Remedy listing of possible troubles with the hermetic compressors used in air-conditioning units. Take a close look at it and determine which are the troubles caused by electrical failure and those caused by the refrigerant system.

Review Questions

1. What is meant by modes of cooling?
2. Where is the temperature sensor located on an air conditioner?
3. How do you make a high-speed fan run slower?

4. What is an auxiliary winding? What purpose does it serve?
5. What is the purpose of a run capacitor in the compressor circuit?
6. What is the purpose of a ladder diagram?
7. What is the function of the fan switch with “auto” and “on” positions?
8. What is the purpose of the overload protector in series with the compressor motor?

Chapter

17

Refrigeration Circuits

Performance Objectives

- Know three different types of refrigerator thermostats.
- Know how short relays work.
- Know how defrost circuits are timed.
- Know how various defrost timers operate.
- Know how to use the ladder diagram to aid in troubleshooting a refrigerator.
- Know how to locate defrost thermostats on refrigerators.
- Know how hot gas defrost works.
- Know how to troubleshoot using a Probable-Cause-Remedy chart.

Circuits for refrigerators vary slightly according to the manufacturer. The best source for these schematics is the manufacturer's service manuals and bulletins. In this chapter we will be concerned primarily with the circuits of domestic (home) refrigerators with freezers as part of the unit.

Basic Refrigerator

The basic or simple refrigerator has an electric circuit that consists of a plug, terminal board or plug-in connector, a start relay with overload protector, and a thermostat or temperature control. A light switch and lamp are usually included in the basic type of circuit (see Fig. 17-1).

Thermostats

Thermostats can have a number of design configurations, depending on the manufacturer. Figure 17-2 shows three different types of thermostats. Each has a knob on the end of the shaft when installed in a refrigerator.

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

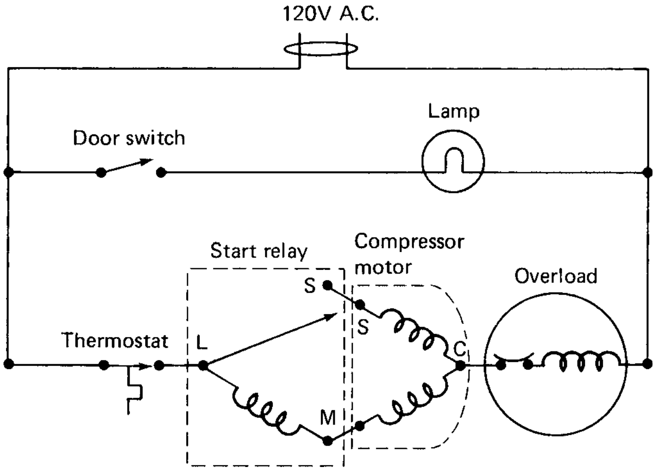
120V A.C.
Lamp
Door switch
Start relay
Compressor motor
Overload
S
S
Thermostat
L
C
M

Figure 17-1 Basic refrigerator circuit.

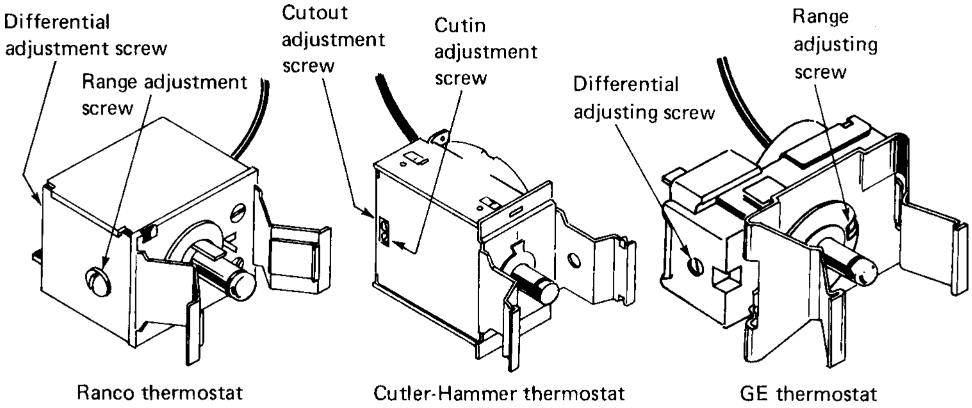
Differential adjustment screw
Range adjustment screw
Cutout adjustment screw
Cutin adjustment screw
Differential adjusting screw
Range adjusting screw
Ranco thermostat
Cutler-Hammer thermostat
GE thermostat

Figure 17-2 Three different types of thermostats. (Courtesy of Kelvinator)

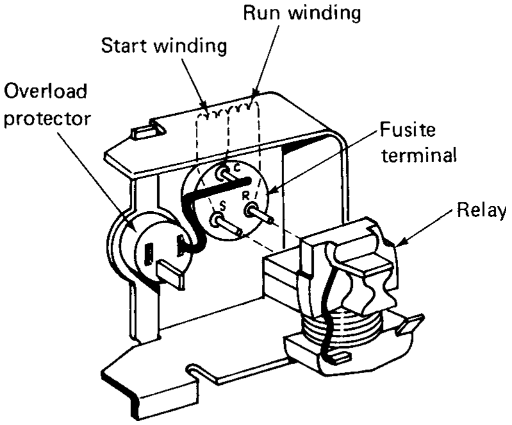
Run winding
Start winding
Overload protector
Fusite terminal
Relay
C
S
R

Figure 17-3 Fusite thermal relay and overload protector. (Courtesy of Kelvinator)

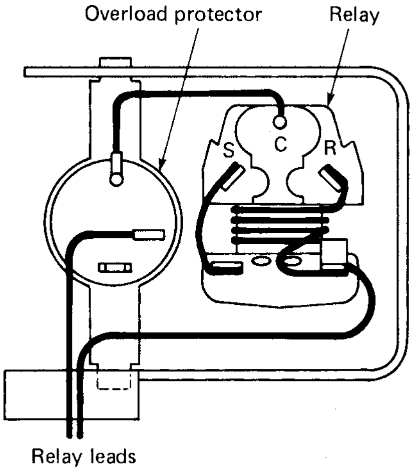
Overload protector
Relay
S
C
R
Relay leads

Figure 17-4 Start relay and overload protector. (Courtesy of Kelvinator)

Start relays

The design of a start relay varies with the manufacturer of the device. Figures 17-3 and 17-4 show two types used for refrigerators. They are usually located on the compressor. S stands for start winding connection, C for the common line, and R for the run winding connection to the compressor motor.

Refrigerator-Freezer Combination

Manual defrost

The refrigerator-freezer combination exists in a manual defrost model for those interested in the most inexpensive device available. It has only the addition of a fan in the freezer compartment and may or may not have a light in the freezer section (see Fig. 17-5).

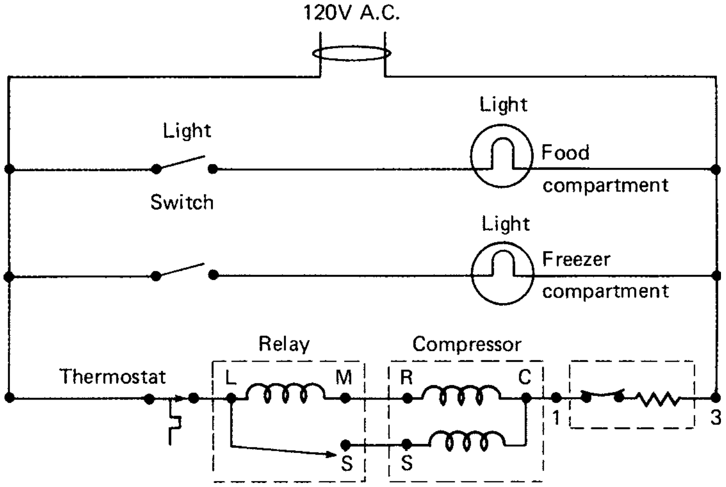
120V A.C.
Light
Light
Food
compartment
Switch
Light
Freezer
compartment
Relay
Compressor
Thermostat
L
M
R
C
S
S
1
3

Figure 17-5 Basic refrigerator-freezer circuit.

Automatic defrost

Frost-free refrigerators and freezers call for some means of timing the defrost cycle. As the frost accumulates in the freezer and refrigerator sections, it must be removed. It is easily removed if done often enough to avoid a buildup.

The defrost timer is used to control how often the defrost cycle takes place (see Fig. 17-6). Contacts 1 and 3 are used for the defrost-motor coils. Contacts 2 and 4 are used to make contact with the proper circuit components to provide energy to the defrost heater or compressor (see Fig. 17-7).

The termination thermostat shown in the circuit is located near or on the evaporator to sense its temperature and indicate that the defrost cycle is complete and has done its job of removing the frost from the evaporator. The defrost timer can be located in any number of places. Figure 17-8 illustrates how it may be placed on the back of the refrigerator.

A wiring diagram of a frost-free refrigerator is shown in Fig. 17-9. Note the addition of other features, such as a drip catcher heater, mullion heater, and door heater. These heaters serve different purposes. The drip catcher heater allows the water that forms during defrost to drain out of the freezer unit to a pan underneath to be evaporated by the heat of the compressor nearby. The mullion heater and door heater are used to prevent a frost buildup on and around the doors of the freezer and refrigerator compartments when the outside air is very humid.

Defrosting

The timing of defrost operations is very important. Too long and you waste energy. Too short a time will result in a job half done, with the resulting decrease in efficiency of the unit since the evaporator is loaded with frost and its ability to cool the compartment is diminished. A number of methods are utilized to defrost. They are all electrically controlled and therefore part of the circuits associated with refrigerator-freezers.

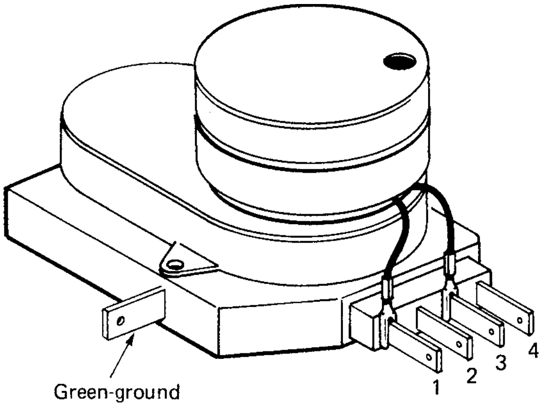

Figure 17-6 Defrost timer. (Courtesy of Kelvinator)

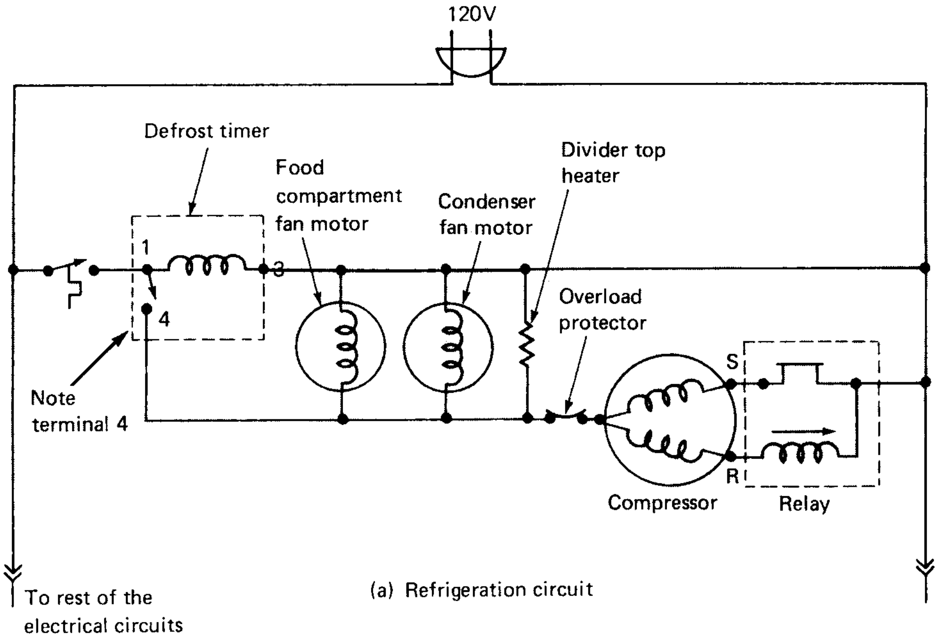
120V
Defrost timer
Food compartment fan motor
Condenser fan motor
Divider top heater
1
3
4
Overload protector
S
R
Note terminal 4
Compressor
Relay
To rest of the electrical circuits

(a) Refrigeration circuit

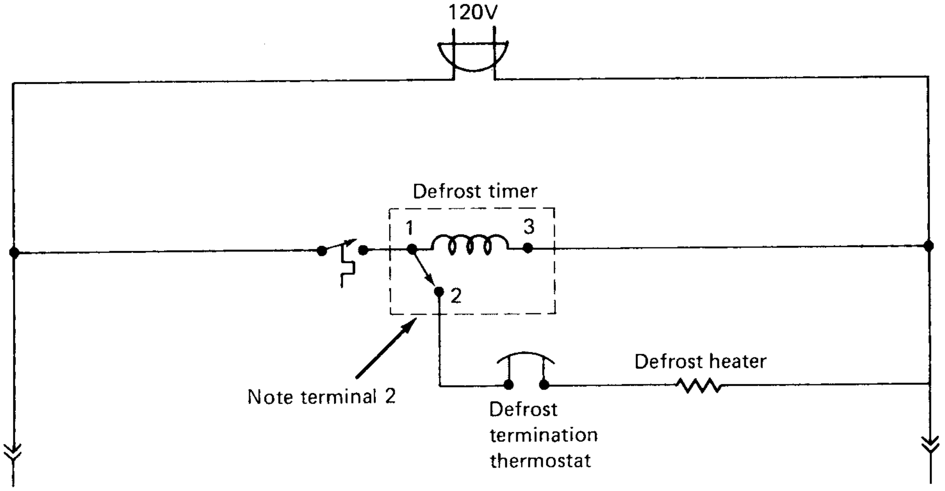
120V
Defrost timer
1
3
2
Note terminal 2
Defrost heater
Defrost termination thermostat

(b) Defrost circuit

Figure 17-7 Note the refrigeration-cycle and the defrost-cycle circuits.

Defrost thermostats

Defrost thermostats are precision built and tested to within ±6°F of the specified limit. A unique characteristic of a bimetal disc is that its calibration is fixed and does not change (see Fig. 17-10). This provides for reliability in excess of the life expectancy of the refrigerator or freezer.

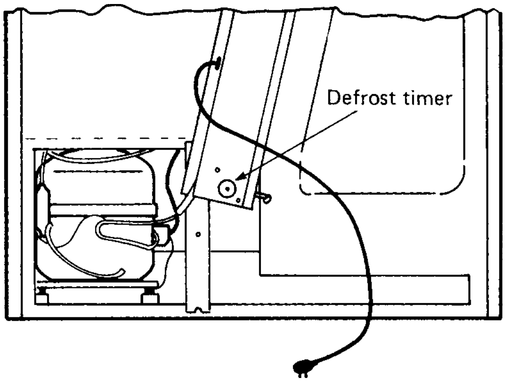
Defrost timer

Figure 17-8 Location of the defrost timer on the refrigerator. (Courtesy of Kelvinator)

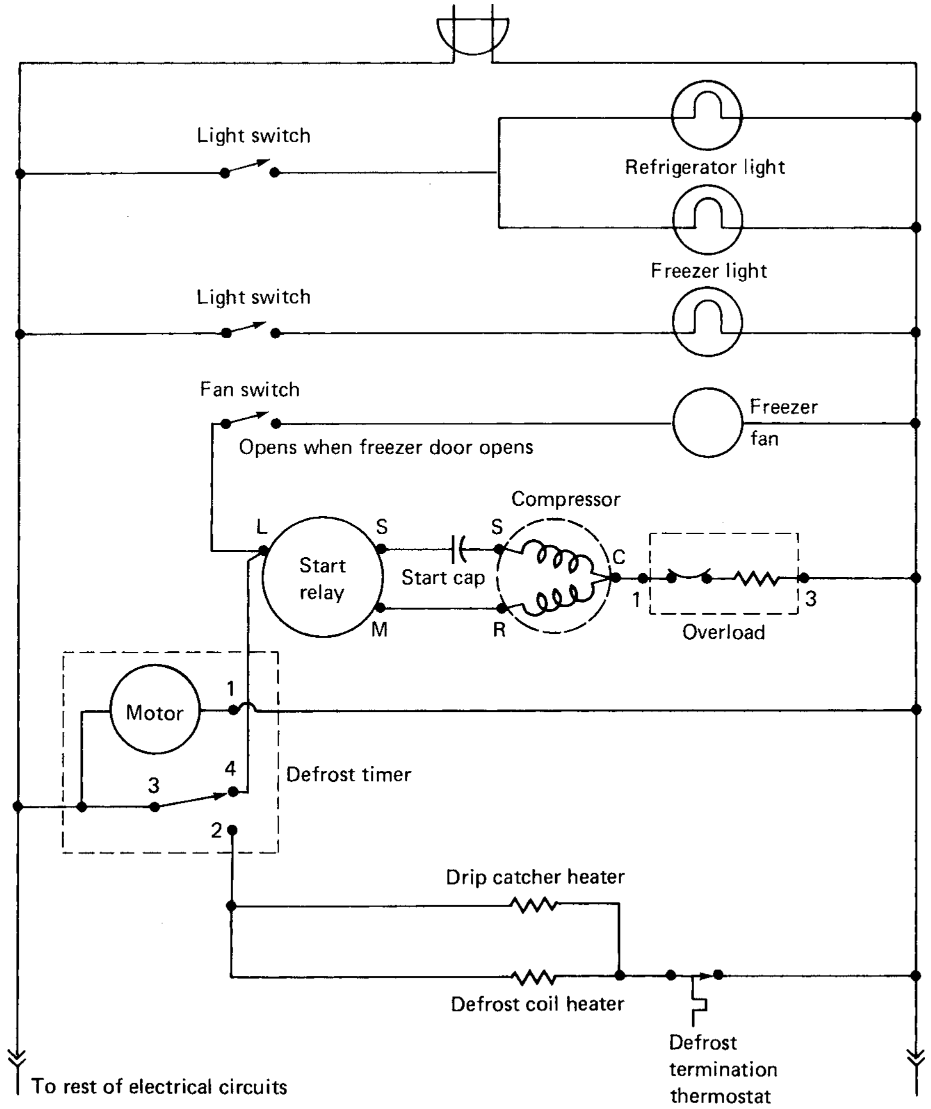
Light switch
Refrigerator light
Freezer light
Light switch
Fan switch
Opens when freezer door opens
Freezer
fan
Compressor
L
S
Start
relay
M
Start cap
S
C
R
1
3
Overload
Motor
1
4
3
2
Defrost timer
Drip catcher heater
Defrost coil heater
Defrost
termination
thermostat
To rest of electrical circuits

Figure 17-9 Refrigerator-freezer circuit.

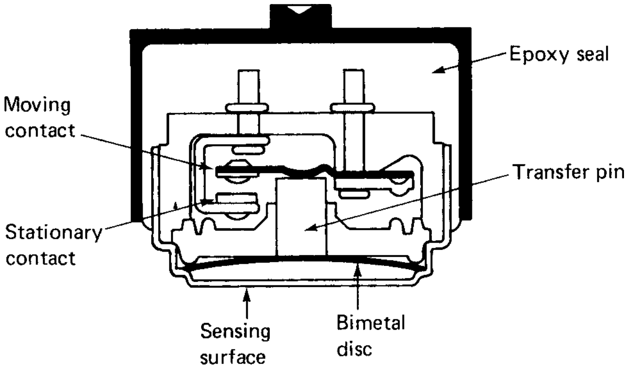

Figure 17-10 Cutaway view of a defrost termination thermostat. (Courtesy of General Electric)

Continual life tests at the factory of 100,000 cycles, corresponding to 100 years, reveal that the calibration of a defrost thermostat does not drift out of tolerance. A slight creep of about 2°F occurs at about 28,000 cycles during the life of all defrost thermostats as the components wear out. A defrost thermostat can, however, be incorrectly calibrated from the beginning. But defrost thermostats never get weak during the life of the refrigerator or freezer. Therefore, if the defrost system has functioned properly for several months before failure, disregard the possibility of an incorrectly calibrated defrost thermostat. Do not suspect that the defrost thermostat has the wrong calibration unless residual ice is found in the evaporator.

The only method for checking the defrost thermostat in the field is to test it for continuity. The contacts should be closed at all times except during the later part of the defrost cycle and for the first 10 minutes thereafter, when the compressor resumes operation.

You can determine that the defrost thermostat contacts are closed on side-by-side models by feeling the mullion for heat. On side-by-side models, the mullion heater is in series with the defrost thermostat. Thus, if the mullion heater is warm, the thermostat contacts must be closed (see Fig. 17-11).

Defrost cycle

Resistance heat is used in automatic defrosting. A radiant heater is attached to the bottom of the evaporator with two aluminum straps. A defrost heater and defrost termination thermostat are wired in series with the defrost heater. The defrost timer initiates and terminates a 21-minute defrost cycle every 6 hours. A termination thermostat cycles the defrost heater off at a predetermined temperature, prior to termination of the 21-minute defrost cycle. The aluminum drain positioned below the evaporator is defrosted by the defrost heater. Some refrigerators have a 25-minute defrost cycle. Another manufacturer makes models with a 17-minute cycle every 8 hours. As you can see, the timer is important in

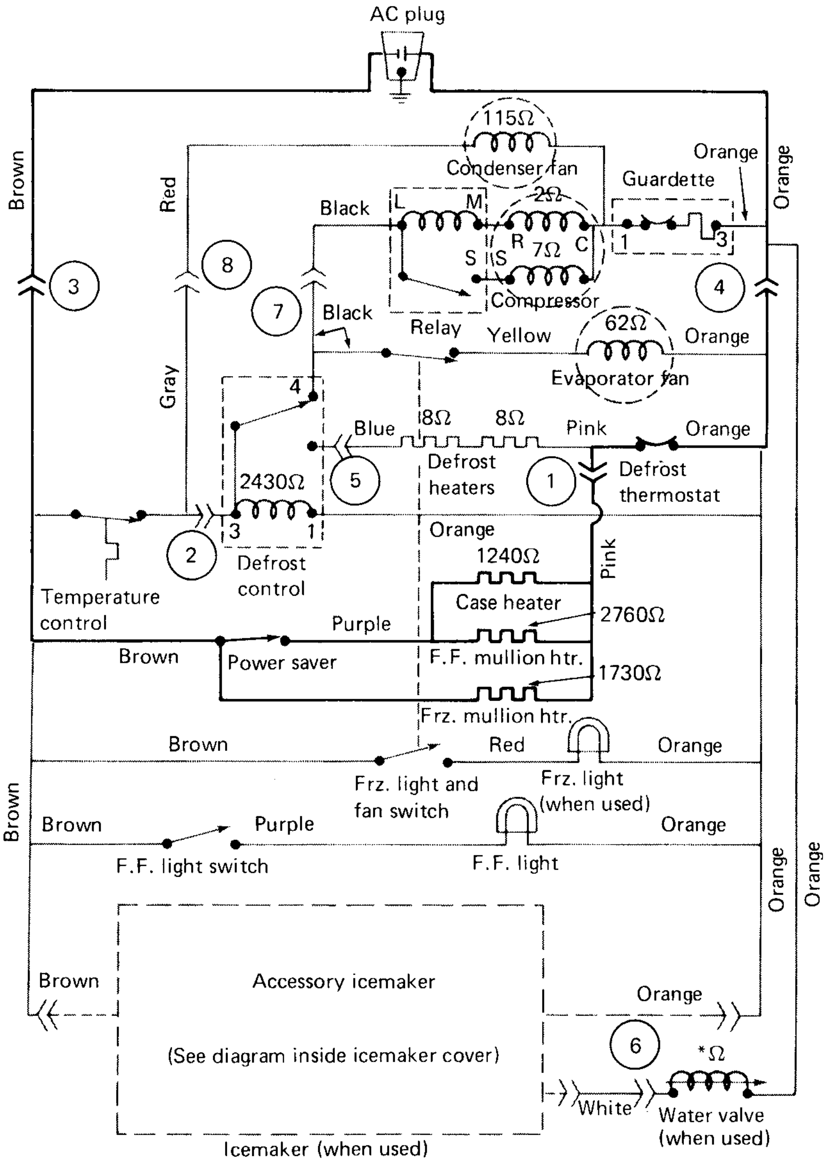
AC plug
115Ω
Condenser fan
Orange
Guardette
Orange
Brown
Red
Black
L
M
2Ω
R
7Ω
C
S
S
1
3
Compressor
Relay
Black
Yellow
62Ω
Orange
Evaporator fan
3
8
7
4
Gray
4
Blue
8Ω
8Ω
Pink
Orange
5
Defrost heaters
1
Defrost thermostat
2430Ω
3
1
2
Defrost control
Orange
1240Ω
Pink
Temperature control
Case heater
2760Ω
Purple
Brown
Power saver
F.F. mullion htr.
1730Ω
Frz. mullion htr.
Brown
Red
Orange
Frz. light and fan switch
Frz. light (when used)
Brown
Brown
Purple
Orange
F.F. light switch
F.F. light
Orange
Orange
Brown
Accessory icemaker
(See diagram inside icemaker cover)
Orange
6
*Ω
White
Water valve (when used)
Icemaker (when used)

Figure 17-11 Schematic for the refrigerator-freezer. (Courtesy of General Electric)

making sure the defrost takes place when needed to keep the evaporator free of frost.

Another method used is the accumulated compressor time. In this case, the defrost operates whether there is any frost on the evaporator or not. The amount of frost buildup is a function of the amount of time the compressor operates. Therefore, it is only reasonable that the accumulated time for which the compressor operates is proportional to the frost buildup. This means defrost timing is typical in some manufacturers' models.

The defrost timer motor in Fig. 17-12 is controlled by the refrigerator thermostat. The defrost timer motor operates only when the refrigerator thermostat calls for compressor operation. After about 6 hours of accumulated compressor operation, the defrost timer switch cam has rotated sufficiently to open the connection between terminals 1 and 4 of the timer. The compressor stops while the refrigerator thermostat continues to call for compressor operation. The circuit through the defrost timer now uses terminals 1 and 2. Power is fed to the defrost heater instead of the compressor. The switch remains in this position (causing the heater to warm up) for approximately 25 minutes. During this time period, the cam on the timer motor has been rotating. This rotation causes the switch to flip from terminal 2 to terminal 4. This action allows power to be restored from terminal 1 to terminal 4. The compressor can now operate and cool until the refrigerator thermostat switch opens.

One other method used to defrost is the accumulated compressor time hot-gas defrost. This type of defrosting uses hot gas from the compressor to melt the frost accumulation on the evaporator. The defrost solenoid is important in this operation (see Fig. 17-13). The solenoid is used to open the line from the compressor

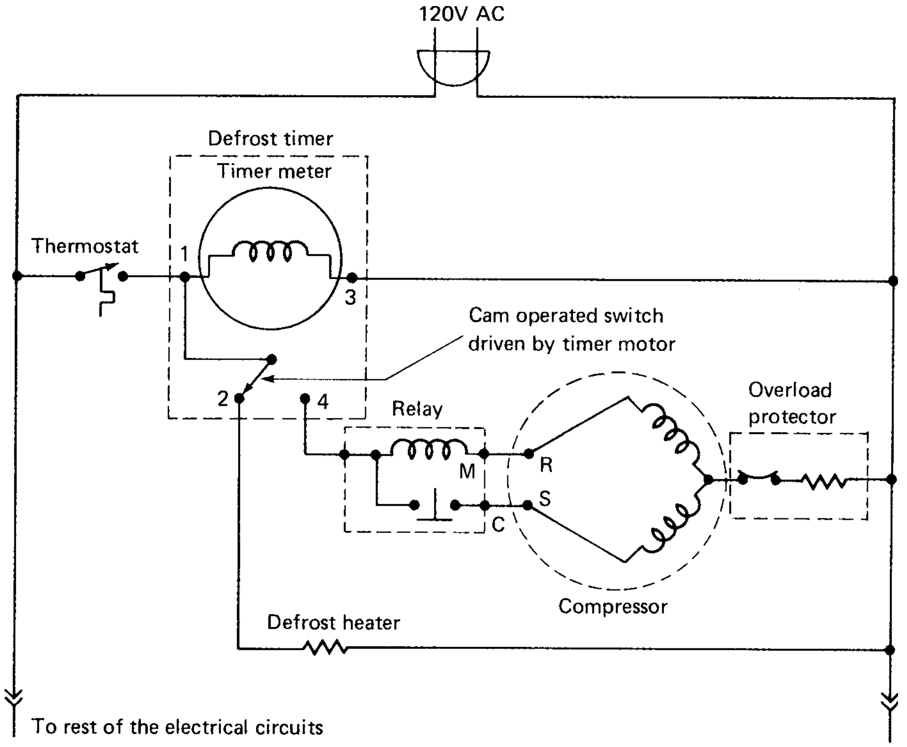
120V AC
Defrost timer
Timer meter
Thermostat
1
3
Cam operated switch driven by timer motor
2
4
Relay
M
C
R
S
Overload protector
Compressor
Defrost heater
To rest of the electrical circuits

Figure 17-12 Accumulated compressor time defrost circuit.

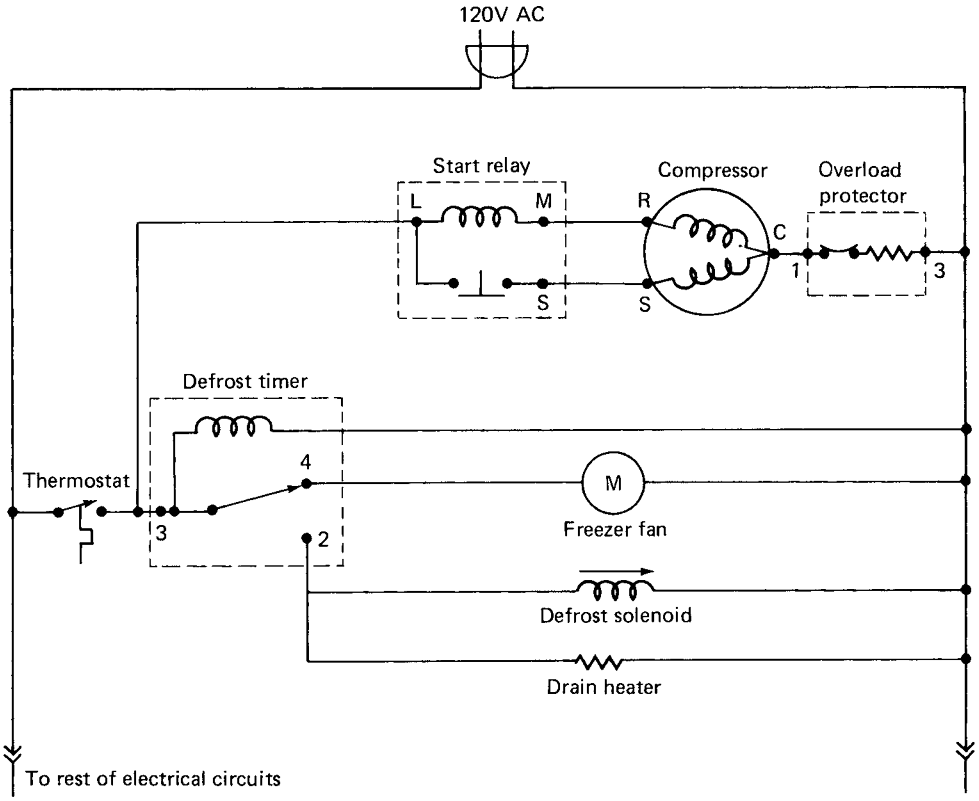
120V AC
Start relay
L
M
S
Compressor
R
C
S
Overload protector
1
3
Defrost timer
4
3
2
Thermostat
M
Freezer fan
Defrost solenoid
Drain heater
To rest of electrical circuits

Figure 17-13 Hot-gas-defrost circuit.

to the piping alongside the evaporator, allowing hot gas to flow in the line. This piping is placed there originally by the manufacturer of the refrigerator. As the hot gas circulates through the piping, it causes the frost on the evaporator to melt. The compressor and evaporator are still in the refrigeration mode, and the cooling process continues while defrosting takes place. Timing is a function of the defrost timer motor and its cams.

Other Devices

As the refrigerator progressed from an "ice-box" replacement to one with a number of conveniences, other circuit changes were necessitated. In most instances they were add-ons. Some of the add-ons were the ice maker, juice dispenser, soft drink dispenser, and cold water faucets. Each refrigerator manufacturer has its own way of designing these devices. You should refer to the manufacturer's manuals and bulletins to be sure of the particular model's electric circuits. See Fig. 17-11 for an illustration of where the ice maker is placed in the circuit.

Operation of the ice maker can be seen in Fig. 17-14. In Fig. 17-14a, note that near the completion of the first revolution the timing cam closes the water valve

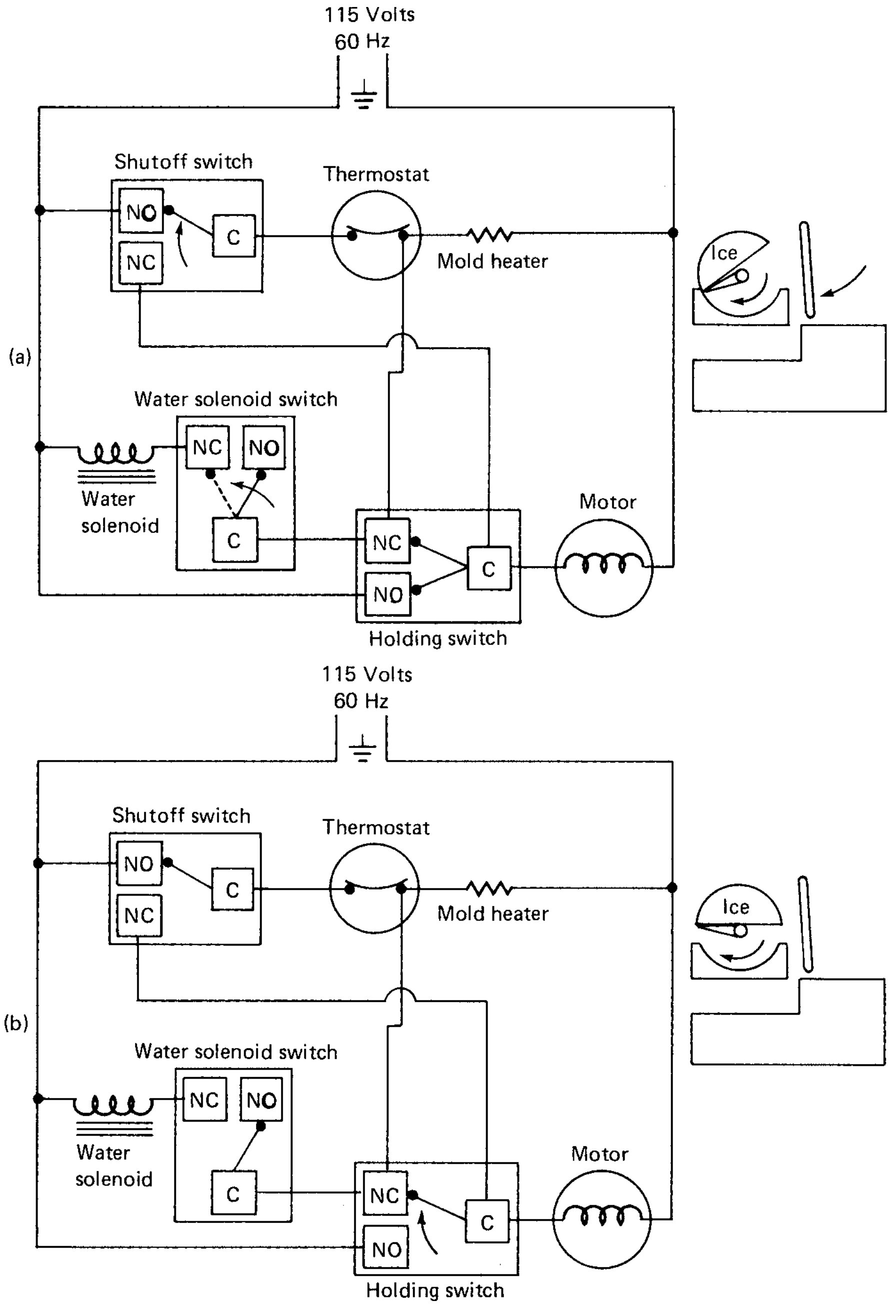
115 Volts
60 Hz
Shutoff switch
NO
NC
C
Thermostat
Mold heater
Ice
(a)
Water solenoid switch
NC
NO
C
Water
solenoid
Motor
NC
NO
C
Holding switch
115 Volts
60 Hz
Shutoff switch
NO
NC
C
Thermostat
Mold heater
Ice
(b)
Water solenoid switch
NC
NO
C
Water
solenoid
Motor
NC
NO
C
Holding switch

Figure 17-14 Circuit for ice maker.

switch. However, since the thermostat is still closed, the mold heater circuit is energized. Current will not pass through the water valve solenoid and its switch since electric current follows the path of least resistance.

At the end of the first revolution (Fig. 17-14b), the timing cam opens the holding switch. However, since the thermostat is still closed, a second revolution begins.

Troubleshooting

To see the importance of the electric circuits used in refrigerators and freezers, study Table 17-1. It shows some of the troubles that occur in a simple domestic freezer. The electrical problems have been emphasized. Compare these with what happens with the mechanics of refrigerant part of the freezer.

TABLE 17-1 Troubleshooting Freezers: Upright Models

Trouble	Probable cause	Remedy or repair
Product too cold.	1. Temperature selector knob set too cold.	1. Set warmer.
	2. Thermostat bulb contact bad.	2. If the bulb contact is bad, the bulb temperature will lag behind the cooling coil temperature. The unit will run longer and make the freezer too cold. See that the bulb makes good contact with the bulb well.
	3. Thermostat is out of adjustment.	3. Readjust or change the thermostat.
Product too warm.	1. Thermostat selector knob set too warm.	1. Set cooler.
	2. Thermostat contact points dirty or burned.	2. Replace thermostat.
	3. Thermostat out of adjustment.	3. Readjust or change the thermostat.
	4. Loose electrical connection.	4. This may break the circuit periodically and cause the freezer to become warm because of irregular or erratic operation; check the circuit and repair or replace parts.
	5. Excessive service load or abnormally high room temperature.	5. Unload part of the contents; move unit to a room with lower temperature or exhaust excess room heat.
	6. Restricted air circulation over wrapped condenser.	6. Allow 6-inch clearance above the top and $3^1/_2$-inch clearance at the sides and between the back of the cabinet and the wall.
	7. Excessive frost accumulation on the refrigerated shelves (manual defrost models).	7. Remove the frost.
	8. Compressor cycling on overload protector.	8. Check the protector and line voltage at the compressor.
Unit will not operate.	1. Service cord out of wall receptacle.	1. Plug in the service cord.
	2. Blown fuse in the feed circuit.	2. Check the wall receptacle with a test lamp for a live circuit; if the receptacle is dead but the building has current, replace the fuse; determine the cause of the overload or short circuit.
	3. Bad service cord plug, loose connection, or broken wire.	3. If the wall receptacle is live, check the circuit and make necessary repairs.
	4. Inoperative thermostat.	4. Power element may have lost charge or points may be dirty; check the points; short out the thermostat; repair or replace the thermostat.

TABLE 17-1 Troubleshooting Freezers: Upright Models (*Continued*)

Trouble	Probable cause	Remedy of repair
	5. Inoperative relay.	5. Replace.
	6. Stuck or burned-out compressor.	6. Replace the compressor.
	7. Low voltage. Cycling on overload.	7. Call utility company, asking them to increase voltage to the house; or move unit to a separate household circuit.
	8. Inoperative overload protector.	8. Replace.
Unit runs all the time.	1. Thermostat out of adjustment.	1. Readjust or change the thermostat.
	2. Short refrigerant charge (up to 4 oz). Cabinet temperatures abnormally low in lower section.	2. Not enough refrigerant to flood the evaporator coil at the outlet to cause the thermostat to cut out; recharge and test for leaks.
	3. Restricted airflow over the wrapper condenser.	3. Provide proper clearances around the cabinet.
	4. Inefficient compressor.	4. Replace.
Unit short cycles.	1. Thermostat erratic or out of adjustment.	1. Readjust or change the thermostat.
	2. Cycling on the relay.	2. This may be caused by low or high line voltage that varies more than 10% from the 115 volts; it may also be caused by high discharge pressures caused by air or noncondensable gases in the system; correct either condition.
Unit runs too much.	1. Abnormal use of the cabinet.	1. Heavy usage requires more operation; check the usage and correct or explain.
	2. Shortage of refrigerant.	2. Unit will run longer to remove the necessary amount of heat and it will operate at a lower than normal suction pressure; put in the normal charge and check for leaks.
	3. Overcharge of refrigerant.	3. Excessively cold or frosted suction line results in lost refrigeration effort; unit must run longer to compensate for the loss; purge off excessive charge.
	4. Restricted airflow over the condenser.	4. This can result if the cabinet is enclosed; this will obstruct the airflow around the cabinet shell; restricted airflow can also be caused by air or noncondensable gases in the system; this results in a higher head pressure; the higher head pressure produces more re-expansion during the suction stroke of the compressor; consequently, less suction vapor is taken; increased running time must compensate for loss of efficiency; correct the condition.
	5. High room or ambient temperature.	5. Any increase in temperature around the cabinet will increase the refrigeration load; this will result in longer running time to maintain cabinet temperature.
Too much frost on refrigerated surfaces—lowside.	1. Abnormally heavy usage in humid weather.	1. Do not leave the freezer door open any longer than necessary to load or remove products.
	2. Poor door gasket seals.	2. This permits the entrance of moisture by migration, which freezes out of the air as frost on the refrigerated surfaces.

Courtesy of Kelvinator.

Rapid Electrical Diagnosis

General Electric and Hotpoint have a quick and accurate device for diagnosing electrical faults in a refrigerator. The handheld dc device is a result of downsizing a computer used in the factory to check production models of refrigerators. The device is called "Big Red" for Rapid Electrical Diagnosis unit (see Fig. 17-15). It allows the technician to check almost every electrical component in the refrigerator within 6 minutes. It is not necessary to unplug the refrigerator, unload any food, move the refrigerator from the wall, or, usually, open the door.

Multi-connectors are designed into the refrigerator wiring harness. They are located behind the front grill. A mini-manual contains information that makes the device useful for a particular refrigerator model. All the latest models with the RED feature have a packet or envelope containing the necessary information. The packet contains the pictorial and schematic wirings, RED component circuits, machine wiring diagram, and components energized through the RED system (see Figs. 17-16 and 17-17).

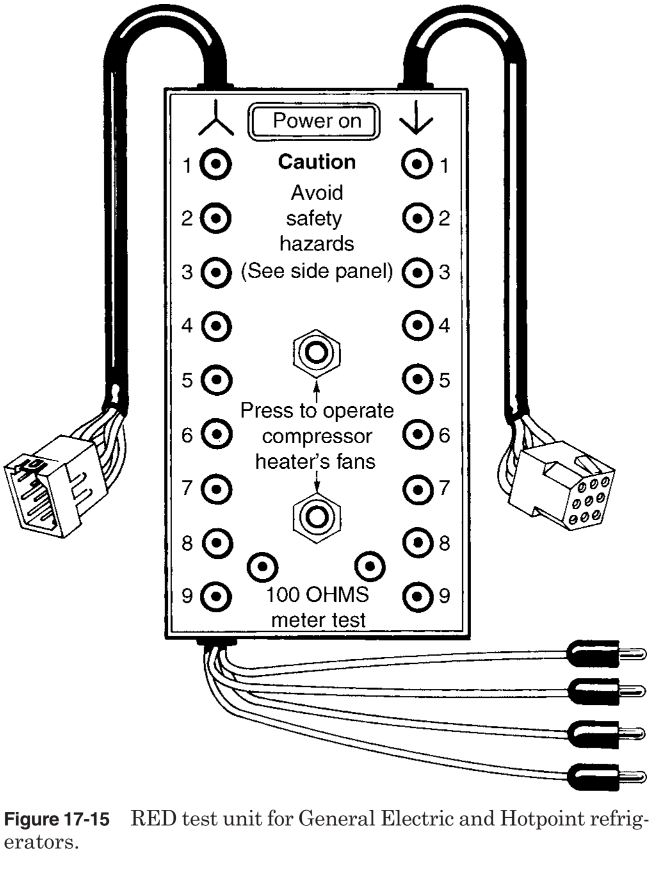
Power on
1
2
3
4
5
6
7
8
9
Caution
Avoid safety hazards
(See side panel)
1
2
3
4
5
6
7
8
9
Press to operate compressor heater's fans
100 OHMS meter test

Figure 17-15 RED test unit for General Electric and Hotpoint refrigerators.

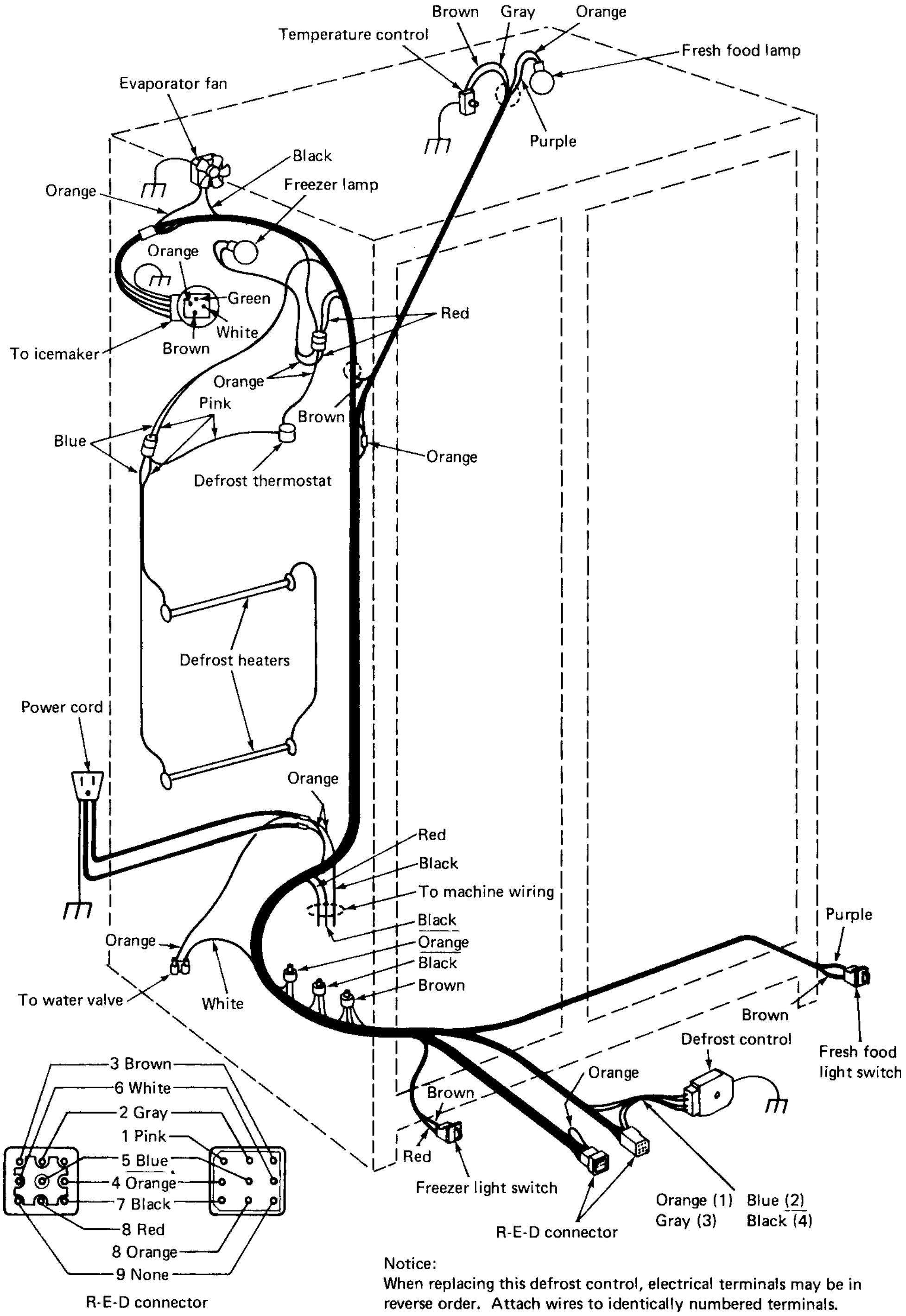
Brown
Gray
Orange
Temperature control
Fresh food lamp
Evaporator fan
Purple
Black
Freezer lamp
Orange
Orange
Green
Red
White
To icemaker
Brown
Orange
Pink
Brown
Blue
Orange
Defrost thermostat
Defrost heaters
Power cord
Orange
Red
Black
To machine wiring
Black
Orange
Black
Brown
Purple
Orange
To water valve
White
Brown
Defrost control
Fresh food light switch
3 Brown
6 White
2 Gray
1 Pink
5 Blue
4 Orange
7 Black
8 Red
8 Orange
9 None
R-E-D connector
Orange
Brown
Red
Freezer light switch
R-E-D connector
Orange (1) Blue (2)
Gray (3) Black (4)

Notice:
When replacing this defrost control, electrical terminals may be in reverse order. Attach wires to identically numbered terminals.

Figure 17-16 Pictorial drawing of the electrical wiring of a side-by-side GE 20-cubic-feet refrigerator that uses the RED system for troubleshooting.

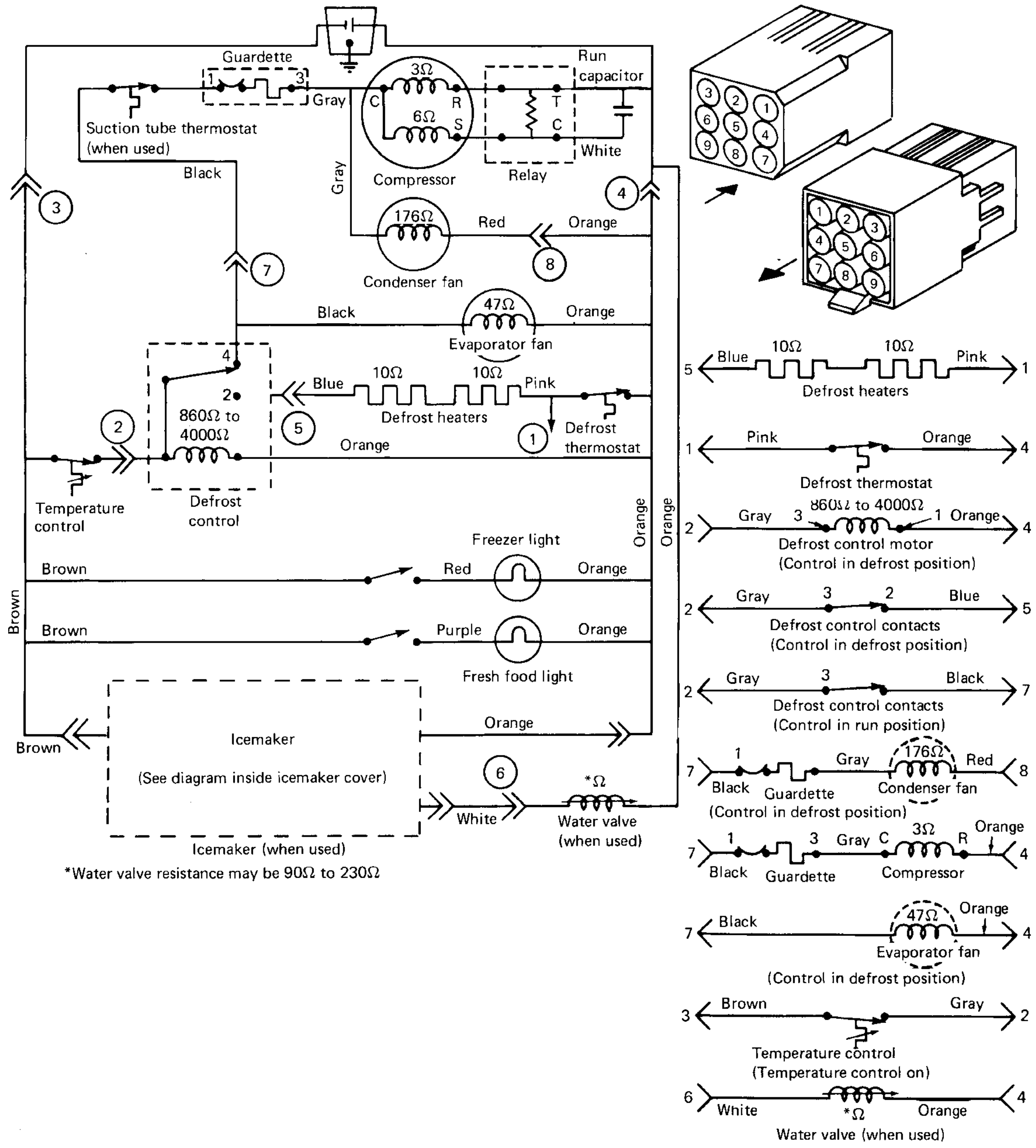
Guardette
1
3
Gray
Suction tube thermostat
(when used)
Black
3Ω
C
R
6Ω
S
Compressor
Gray
Run
capacitor
T
C
White
Relay
176Ω
Red
Orange
Condenser fan
Black
47Ω
Orange
Evaporator fan
4
2
860Ω to
4000Ω
Blue
10Ω
10Ω
Pink
Defrost heaters
Orange
Defrost
thermostat
Temperature
control
Defrost
control
Freezer light
Brown
Red
Orange
Brown
Brown
Purple
Orange
Fresh food light
Orange
Orange
Icemaker
(See diagram inside icemaker cover)
Orange
Brown
White
*Ω
Water valve
(when used)
Icemaker (when used)
*Water valve resistance may be 90Ω to 230Ω
① ② ③ ④ ⑤ ⑥ ⑦ ⑧
3 2 1 6 5 4 9 8 7
1 2 3 4 5 6 7 8 9
5 Blue 10Ω 10Ω Pink 1
Defrost heaters
1 Pink Orange 4
Defrost thermostat
2 Gray 3 860Ω to 4000Ω 1 Orange 4
Defrost control motor
(Control in defrost position)
2 Gray 3 2 Blue 5
Defrost control contacts
(Control in defrost position)
2 Gray 3 Black 7
Defrost control contacts
(Control in run position)
7 1 Gray 176Ω Red 8
Black Guardette Condenser fan
(Control in defrost position)
7 1 3 Gray C 3Ω R Orange 4
Black Guardette Compressor
7 Black 47Ω Orange 4
Evaporator fan
(Control in defrost position)
3 Brown Gray 2
Temperature control
(Temperature control on)
6 White *Ω Orange 4
Water valve (when used)

Figure 17-17 Ladder diagram of a GE 20-cubic-feet refrigerator that uses the RED system.

Figure 17-16 shows the pictorial of the wiring harness and color code for the RED connector. Figure 17-17 gives the electrical schematic and beside it is the same diagram, but with the < and > connections indicating whether it is a female- or male-type connection for that point.

Use of RED may reduce the technician's service time by 20%. It reduces the additional service callbacks for the same problem by 50%. All "no-frost" GE models 16 cubic feet and larger have this feature. Hotpoint uses the same system. Newer refrigerators have more advanced electronics units and testing devices. This explanation provides you with some of the early progress in the refrigerator field.

Energy-Saver Switches

An energy-saver switch is the latest development in refrigerator design (see Fig. 17-18). The energy-saver switch can reduce operating costs by taking some of the heaters out of the circuit.

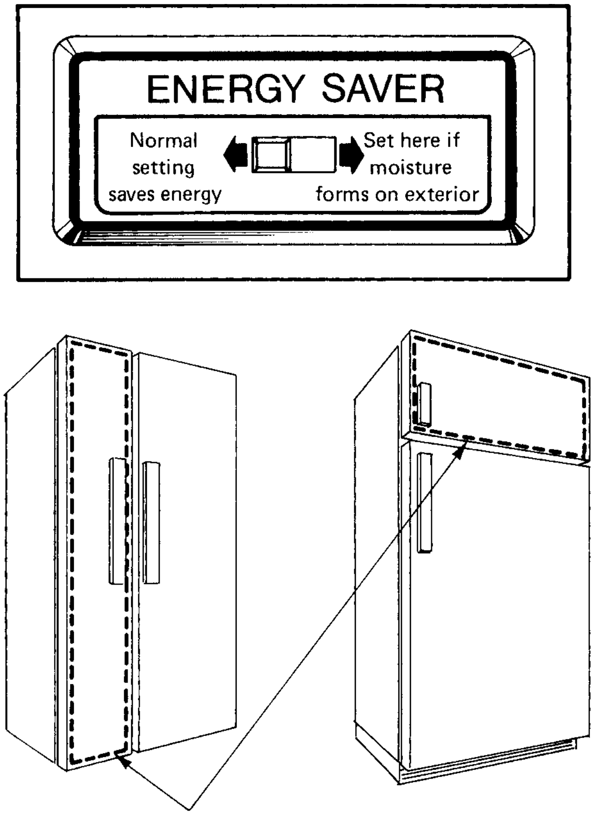
ENERGY SAVER
Normal setting saves energy
Set here if moisture forms on exterior
Antisweat heaters for humid conditions only

Figure 17-18 Energy-saver switches save electricity and one used in conjunction with the antisweat heaters located around the freezer section doors.

Controls for the case and mullion antisweat heaters are incorporated into the energy-saver switch. The heaters prevent moisture formation on the outside of the case in humid conditions. The normal setting may be used about 80% of the time due to low humidity in the winter and air conditioning in the summer. The switch can also be set to an alternate setting only if moisture forms on the outside of the refrigerator.

Review Questions

1. What makes up the basic refrigerator?
2. What is the purpose of the thermostat in a basic refrigerator?
3. What is the function of the start relay in a refrigerator?
4. What is necessary to make a refrigerator frost free?
5. How does the defrost timer work?
6. Why is a defrost thermostat needed in a refrigerator?
7. What is meant by accumulated compressor time hot-gas defrost?
8. How does the ice maker work in a refrigerator?

Chapter 18

Troubleshooting

Performance Objectives

- Know some of the common sense rules for safely working around air-conditioning and refrigeration equipment.
- Know how often filters on air conditioners have to be cleaned.
- Know how the low supply voltage affects the operation of electrical equipment.
- Know the source of most trouble in air-conditioning systems.
- Know the source of most trouble in heating systems.
- Know how to test motor capacitors.
- Know how to use a clamp-on meter to troubleshoot.
- Know how to test the centrifugal switch in a single-phase motor.
- Know how to measure the capacitance of a capacitor in a piece of heating and cooling equipment.
- Know good troubleshooting procedures.

From time to time, air-conditioning and refrigeration units experience problems with the electrical and the mechanical aspects of their operation. They have in some cases been moved and damaged, and in other cases they have been repaired with components that are incorrect or wired into the circuits incorrectly.

The service technician must be aware of the problems that may develop in a period of time so that they may be located quickly, easily, and without undue expense to the owner of the unit. In this chapter, we will attempt to show how troubleshooting procedures can aid in making the job faster and easier.

Safety

Safety first has its direct and implied meaning. You can work with air-conditioning and refrigeration equipment safely if a few common sense rules are followed.

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

Handling refrigerants

The proper use of gloves, eye protection, and clothing to protect the body is necessary. Freon or any refrigerant escaping from a refrigeration unit can cause permanent damage to the skin. It can also cause blindness if you are hit in the face with the gases under pressure. Make sure you wear the proper clothing at all times when troubleshooting or recharging a unit.

Testing precaution

Pressure testing or cleaning refrigeration and air-conditioning systems can be dangerous. Be careful not to exceed 150 psig when pressure testing a complete system.

Electrical safety

The main rule for electrical safety is to see to it that the main circuit breaker is in the off position and locked before starting to remove or check (with an ohmmeter) any refrigeration, air-conditioning, or heating equipment. Keep in mind that some compressors have power applied at all times to the off-cycle crankcase heater. Even if the compressor is not running, the power is applied to the crankcase heater. Some run capacitors are connected to the compressor motor windings even when the compressor is not running. Other devices are energized when the compressor is not running. Thus, electrical power is applied to the unit even when the compressor is not running.

Compressor Problems

A number of compressor problems can be quickly identified from a table of problems, possible causes, and suggested repair (see Table 18-1). The compressor is the heart of the refrigeration system whether it is the air conditioner, refrigerator, or freezer. That makes it of primary concern in any troubleshooting procedure.

PSC compressors

The permanent split-capacitor (PSC) compressor has some problems that should be uppermost in your mind so that you are aware of them (see Table 18-2). The low-voltage problems are listed and possible corrections given. The branch-circuit fuse or circuit breaker trips and causes the unit to become disabled. The possible cause could be that the rating of the protection device is not high enough to handle the current. Some problems are within the scope of the work the technician can handle. Others like low line voltage caused by factors beyond the control of the homeowner call for the utility company to correct them.

A lot of troubleshooting is common sense. If the fuse is blown and you put in another and it blows, it tells you that there is something drawing too much current. Simply check for cause.

Air-conditioner compressors

Home air conditioners, those that fit into a window or those that are part of a central air-conditioning system, are subject to problems with their compressors

TABLE 18-1 **Compressor Troubleshooting and Service**

Complaint	Possible cause	Repair
Compressor will not start. There is no hum.	1. Line disconnect switch open.	1. Close start or disconnect switch.
	2. Fuse removed or blown.	2. Replace fuse.
	3. Overload protector tripped.	3. Refer to electrical section.
	4. Control stuck in open position.	4. Repair or replace control.
	5. Control off due to cold location.	5. Relocate control.
	6. Wiring improper or loose.	6. Check wiring against diagram.
Compressor will not start. It hums, but trips on overload protector.	1. Improperly wired.	1. Check wiring against diagram.
	2. Low voltage to unit.	2. Determine reason and correct.
	3. Starting capacitor defective.	3. Determine reason and replace.
	4. Relay failing to close.	4. Determine reason and correct, replace if necessary.
	5. Compressor motor has a winding open or shorted.	5. Replace compressor.
	6. Internal mechanical trouble in compressor.	6. Replace compressor.
	7. Liquid refrigerant in compressor.	7. Add crankcase heater and/or accumulator.
Compressor starts, but does not switch off of start winding.	1. Improperly wired.	1. Check wiring against diagram.
	2. Low voltage to unit.	2. Determine reason and correct.
	3. Relay failing to open.	3. Determine reason and replace if necessary.
	4. Run capacitor defective.	4. Determine reason and replace.
	5. Excessively high discharge pressure.	5. Check discharge shutoff valve, possible overcharge, or insufficient cooling of condenser.
	6. Compressor motor has a winding open or shorted.	6. Replace compressor.
	7. Internal mechanical trouble in compressor (tight).	7. Replace compressor.
Compressor starts and runs, but short cycles on overload protector.	1. Additional current passing through the overload protector.	1. Check wiring against diagram; check added fan motors, pumps, etc., connected to wrong side of protector.
	2. Low voltage to unit (or unbalanced if three phase).	2. Determine reason and correct.
	3. Overload protector defective.	3. Check current, replace protector.
	4. Run capacitor defective.	4. Determine reason and replace.
	5. Excessive discharge pressure.	5. Check ventilation, restrictions in cooling medium, restrictions in refrigeration system.
	6. Suction pressure too high.	6. Check for possibility of misapplication; use stronger unit.
	7. Compressor too hot—return gas is hot.	7. Check refrigerant charge. (Repair leak.) Add refrigerant if necessary.
	8. Compressor motor has a winding shorted.	8. Replace compressor.
Unit runs, but short cycles on.	1. Overload protector.	1. Check current; replace protector.
	2. Thermostat.	2. Differential set too close; widen.
	3. High-pressure cutout due to insufficient air or water supply, overcharge, or air in system.	3. Check air or water supply to condenser; reduce refrigerant charge, or purge.

(*Continued*)

TABLE 18-1 Compressor Troubleshooting and Service (*Continued*)

Complaint	Possible cause	Repair
	4. Low-pressure cutout due to: a. Liquid line solenoid leaking. b. Compressor valve leak. c. Undercharge. d. Restriction in expansion device.	4. a. Replace. b. Replace. c. Repair leak and add refrigerant. d. Replace expansion device.
Unit operates long or continuously.	1. Shortage of refrigerant.	1. Repair leak; add charge.
	2. Control contacts stuck or frozen closed.	2. Clean contacts or replace control.
	3. Refrigerated or air-conditioned space has excessive load or poor insulation.	3. Determine fault and correct.
	4. System inadequate to handle load.	4. Replace with larger system.
	5. Evaporator coil iced.	5. Defrost.
	6. Restriction in refrigeration system.	6. Determine location and remove.
	7. Dirty condenser.	7. Clean condenser.
	8. Filter dirty.	8. Clean or replace.
Start capacitor open, shorted, or blown.	1. Relay contacts not operating properly.	1. Clean contacts or replace relay if necessary.
	2. Prolonged operation on start cycle due to: a. Low voltage to unit. b. Improper relay. c. Starting load too high.	2. a. Determine reason and correct. b. Replace. c. Correct by using pump-down arrangement if necessary.
	3. Excessive short cycling	3. Determine reason for short cycling as mentioned in previous complaint.
	4. Improper capacitor.	4. Determine correct size and replace.
Run capacitor open, shorted, or blown.	1. Improper capacitor.	1. Determine correct size and replace.
	2. Excessively high line voltage (110% of rated maximum).	2. Determine reason and correct.
Relay defective or burned out.	1. Incorrect relay.	1. Check and replace.
	2. Incorrect mounting angle.	2. Remount relay in correct position.
	3. Line voltage too high or too low.	3. Determine reason and correct.
	4. Excessive short cycling.	4. Determine reason and correct.
	5. Relay being influenced by loose vibrating mounting.	5. Remount rigidly.
	6. Incorrect run capacitor.	6. Replace with proper capacitor.
Space temperature too high.	1. Control setting too high.	1. Reset control.
	2. Expansion valve too small.	2. Use larger valve.
	3. Cooling coils too small.	3. Add surface or replace.
	4. Inadequate air circulation.	4. Improve air movement.
Suction line frosted or sweating.	1. Expansion valve oversized or passing excess refrigerant.	1. Readjust valve or replace with smaller valve.
	2. Expansion valve stuck open.	2. Clean valve of foreign particles. Replace if necessary.
	3. Evaporator fan not running.	3. Determine reason and correct.
	4. Overcharge of refrigerant.	4. Correct charge.
Liquid line frosted or sweating.	1. Restriction in dehydrator or strainer.	1. Replace part.
	2. Liquid shutoff (king valve) partially closed.	2. Open valve fully.
Unit noisy.	1. Loose parts or mountings.	1. Tighten.
	2. Tubing rattle.	2. Reform to be free of contact.
	3. Bent fan blade causing vibration.	3. Replace blade.
	4. Fan motor bearings worn.	4. Replace motor.

Courtesy of Kelvinator.

TABLE 18-2 PSC Compressor Motor Troubles and Corrections

Causes	Corrections
Low Voltage	
1. Inadequate wire size.	1. Increase wire size.
2. Watt-hour meter too small.	2. Call utility company.
3. Power transformer too small or feeding too many homes.	3. Call utility company.
4. Input voltage too low.	4. Call utility company.
Note: Starting torque varies as the square of the input voltage.	
Branch Circuit Fuse or Circuit Breaker Tripping	
1. Rating too low.	1. Increase size to a minimum or 175% of unit FLA (full load amperes) to a maximum of 225% of FLA.
System Pressure High or Not Equalized	
1. Pressures not equalizing within 3 minutes.	1. a. Check metering device (capillary tube or expansion valve). b. Check room thermostat for cycling rate; off cycle should be at least 5 minutes; also check for "chattering." c. Has some refrigerant dryer or some other possible restriction been added?
2. System pressure too high.	2. Make sure refrigerant charge is correct.
3. Excessive liquid in crankcase (split-system applications).	3. Add crankcase heater and suction line accumulator.
Miscellaneous	
1. Run capacitor open or shorted.	1. Replace with new, properly sized capacitor.
2. Internal overload open.	2. Allow 2 hours to reset before changing compressor.

and electrical controls. Most of those that operate in the window are designed for 120- or 240-volt operation. In some cases, during the summer when line voltage is low, they may experience some low-voltage problems.

The units are equipped with some kind of filter to make sure the air is cleaned before it is forced into the room being cooled. These filters usually require a change or cleaning at least once a year. Where dust is a problem, such maintenance should be more frequent. At this time, the condenser coil should be brushed with a soft brush and flushed with water. The filters should be vacuumed and then washed to remove dust. The outside of the case should be wiped clean with a soapy cloth. The cleaner the filter, the more efficient the unit. In some cases a clogged filter can cause compressor problems. Table 16-1 shows some of the problems associated with the compressor in these units. Note how most of the problems associated with air conditioners are electrical.

Low-Voltage Operation

Electrical apparatus designed to produce at full capacity at the voltage indicated on the rating plate can malfunction at lower than designated voltages. Motors operated at lower than rated voltage cannot provide full horsepower without shortening their service life. Relays and solenoids can also fail to operate if low voltage is present.

The Air Conditioning and Refrigeration Institute (ARI) certifies the cooling units after testing them. The units are tested to make sure they will operate with ±10% above or below rated voltage. This does not mean they will operate continuously without damage to the motor. Most air-conditioning compressor burnouts are caused by low-voltage operation. A hermetic compressor is entirely enclosed within the refrigerant cycle; it is very important that it not be abused by overloading or low voltage. Both conditions can occur during peak load periods. A national survey has shown that the most common cause of compressor low voltage is the use of undersized conductors between the utility lines and the condensing unit.

Using a System to Troubleshoot (Electrical)

Refrigeration, air-conditioning, and heating units all have various systems that cause them to operate and do the job they were designed for. First, you have to identify the system and the problems most commonly associated with the particular system. Then you have to be able to determine exactly what caused the symptom you are witnessing. Next, you have to be able to correct the problem and put the unit back into operation.

Electrical tests are the most common because the electrical problems are the most frequently encountered in all three types of equipment. A systematic procedure is necessary to obtain the needed results.

Motor testing

Testing a motor in a sealed condensing unit to determine why it does not operate becomes a very simple process if the correct procedure is followed. Each test made should be one of a series of eliminations to determine what part of the system is defective. By checking other parts of the wiring system before checking the unit itself, a great deal of time can be saved since, in most cases, the trouble will be in the wiring or controls rather than in the unit.

To make a complete electrical test on electrical outlets and on the unit itself, it is advisable to make a test cord (see Fig. 18-1). By connecting the black and white terminals together and placing a light bulb in the socket, the cord may be used to check the wall outlet into which the unit is connected. By connecting the

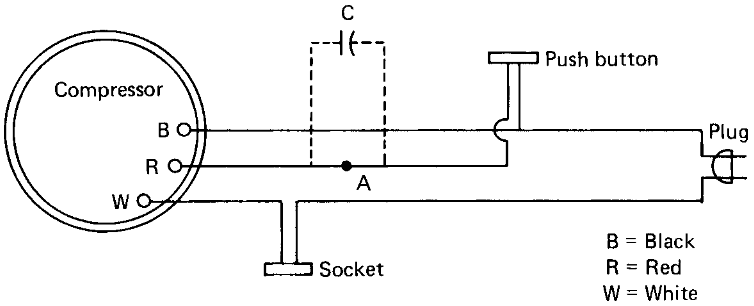

Figure 18-1 Test lamp for troubleshooting.

white and red terminal clips, this same test may be made by depressing the push button. This will serve as a test to make certain the push button is in working order. This test cord should have a capacitor installed in the red lead to the push button if the compressor is a capacitor-start type.

Once these tests have been completed and it is known that current is being supplied to the unit, the next step is to check the three wires on the base of the compressor unit. Pull the plug from the wall receptacle and carefully examine the nuts that hold the wires in position. Then try each wire to be sure it is held firmly in place, as a loose wire may keep the unit from operating. Test the thermostatic switch to determine whether or not contact is being made at that point. Turn the cold-control knob several times. If this fails to start the unit, then short across the thermostatic switch terminals on the switch. To do this, it will be necessary to remove the switch cover from the top of the switch. If the unit starts, it is an indication that the thermostatic switch is not operating properly and must be repaired or replaced. After the thermostatic switch has been checked and if the trouble is not located, it will be necessary to determine whether the trouble is in the motor, motor-protector relay, or capacitor.

Capacitor testing

The capacitor must be checked before testing the unit itself. This is done in the following manner:

1. Disconnect the capacitor wires from the motor-protector relay.
2. Connect these two wires to the black and white terminals of the test cord.
3. Put a 150-watt light bulb in the receptacle on the test cord, and plug it into an outlet.

If the 150-watt bulb does not light up, it is an indication that the capacitor has an open circuit and must be replaced. However, if the bulb does light up, it is not an indication that the capacitor is perfect. This must be checked further by shorting across the two terminals of the capacitor with a screwdriver with an insulated handle. If the brilliance of the light changes, that is, if the light bulb burns brighter when the terminals of the capacitor are shorted, it indicates that the capacitor is in proper operating condition. A decided sparking of the terminals will also be noticed when the terminals are shorted. If the brilliance of the bulb does not change, it is an indication that the capacitor has an internal short and must be replaced.

Motor-protector relay testing

If the overload protector is found defective during the preceding test, replace it (see Fig. 18-2). The motor overload protector is usually accessible for replacement and is located near the compressor (see Fig. 18-3). Remove the clips holding the overload in position, and remove the wire from the terminal by pulling it outward. Notice the position of the electrical lead before replacing it (see Fig. 18-4). Place the electrical leads on the replacement overload and check for correct connections against the wiring diagram. Close up the unit (see Fig. 18-5). Test for proper operation.

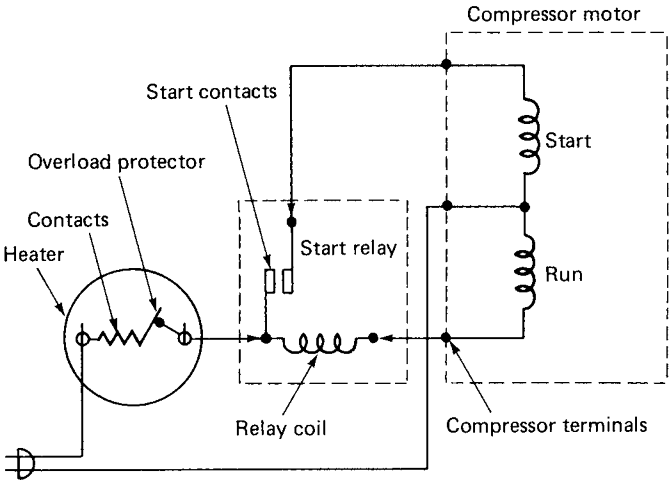
Compressor motor
Start contacts
Overload protector
Contacts
Heater
Start relay
Start
Run
Relay coil
Compressor terminals

Figure 18-2 Location of overload protector in the circuit.

Using Meters to Check for Problems

The voltmeter and the ohmmeter can be used to isolate various problems. You should be able to read the schematic and make the proper voltage or resistance measurements. An incorrect reading will indicate the possibility of a problem.

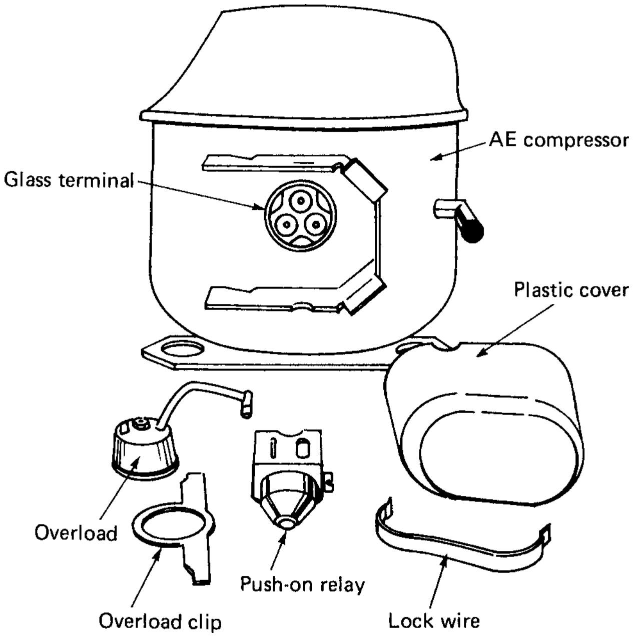
AE compressor
Glass terminal
Plastic cover
Overload
Push-on relay
Overload clip
Lock wire

Figure 18-3 Compressor with its glass terminal and associated components. (Courtesy of Tecumseh)

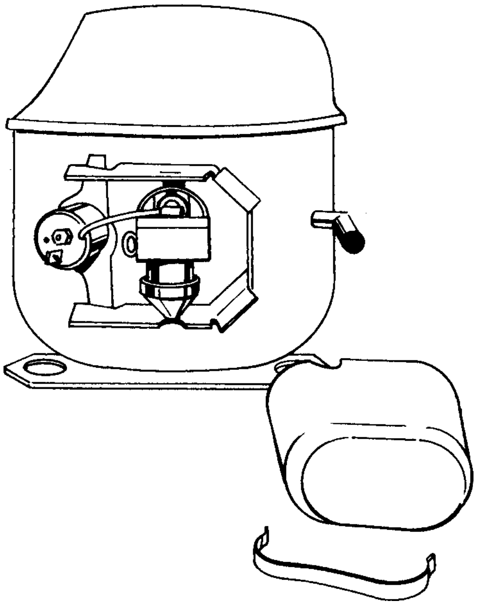

Figure 18-4 Overload protection location on the compressor. (Courtesy of Tecumseh)

Troubleshooting charts will aid in isolating the problem to a given system. Once you have arrived at the proper system that may be causing the symptoms noticed, you will then need to use the ohmmeter with the power off to isolate a section of the system. Once you have zeroed in on the problem, you can locate it by knowing what the proper reading should be. Deviation from a stated reading of over 10% is usually indicative of a malfunction, and in most cases the component part must be replaced to assure proper operation and no callbacks.

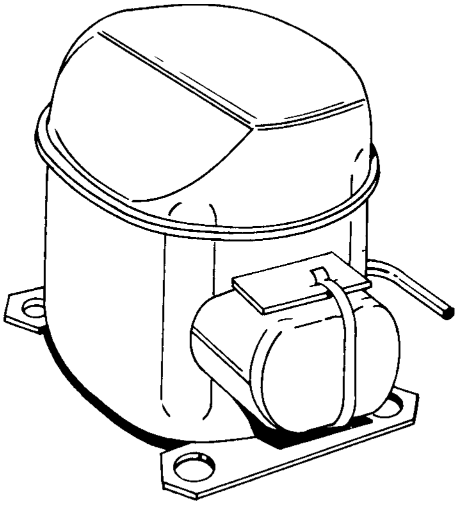

Figure 18-5 Compressor with cover. (Courtesy of Tecumseh)

Using a Volt-Ammeter for Troubleshooting Electric Motors

Most electrical equipment will work satisfactorily if the line voltage differs ±10% from the actual nameplate rating. In a few cases, however, a 10% voltage drop may result in a breakdown. Such may be the case with an induction motor that is being loaded to its fullest capacity both on start and run. A 10% loss in the line voltage will result in a 20% loss in torque.

The full-load current rating on the nameplate is an approximate value based on the average unit coming off the manufacturer's production line. The actual current for any one unit may vary as much as ±10% at rated output. However, a load current that exceeds the rated value by 20% or more will reduce the life of the motor due to higher operating temperatures, and the reason for excessive current should be determined. In many cases it may simply be an overloaded motor. The percentage of increase in load will not correspond with percentage of increase in load current. For example, in the case of a single-phase induction motor, a 35% increase in current may correspond to an 80% increase in torque output.

The operating conditions and behavior of electrical equipment can be analyzed only by actual measurement. A comparison of the measured terminal voltage and current will check whether the equipment is operating within electrical specifications.

A voltmeter and an ammeter are needed for the two basic measurements. To measure voltage, the test leads of the voltmeter are in contact with the terminals of the line under test. To measure current, the conventional ammeter must be connected in series with the line so that the current will flow through the ammeter.

To insert the ammeter, you must shut down the equipment, break open the line, connect the ammeter, and then start up the equipment to read the meter. And you have to do the same to remove the meter once it has been used. Other time-consuming tests may have to be made to locate the problem. However, all this can be eliminated by the use of a clamp-on volt-ammeter.

Clamp-on volt-ammeters

The pocket size volt-ammeter shown in Fig. 18-6 is the answer to most troubleshooting problems on the job. The line does not have to be disconnected to obtain a current reading. The meter works on the transformer principle; it picks up the magnetic lines surrounding a current-carrying conductor and presents this as a function of the entire amount flowing through the line. Remember, in transformers we discussed how the magnetic field strength in the core of the transformer determines the amount of current in the secondary. The same principle is used here to detect the flow of current and how much.

To get transformer action, the line to be tested is encircled with the split-type core by simply pressing the trigger button. Aside from measuring terminal voltages and load currents, the split-core ammeter-voltmeter can be used to track down electrical difficulties in electric motor repair.

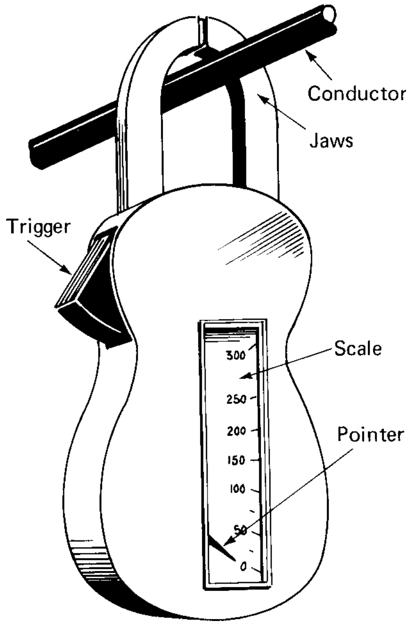
Conductor
Jaws
Trigger
300
250
200
150
100
50
0
Scale
Pointer

Figure 18-6 Clamp-on volt-ammeter. (Courtesy of Amprobe)

Looking for grounds

To determine whether a winding is grounded or has a very low value of insulation resistance, connect the unit and test leads as shown in Fig. 18-7. Assuming the available line voltage is approximately 120 volts, use the unit's lowest voltage range. If the winding is grounded to the frame, the test will indicate full line voltage.

A high-resistance ground is simply a case of low insulation resistance. The indicated reading for a high-resistance ground will be a little less than line voltage. A winding that is not grounded will be evidenced by a small or negligible

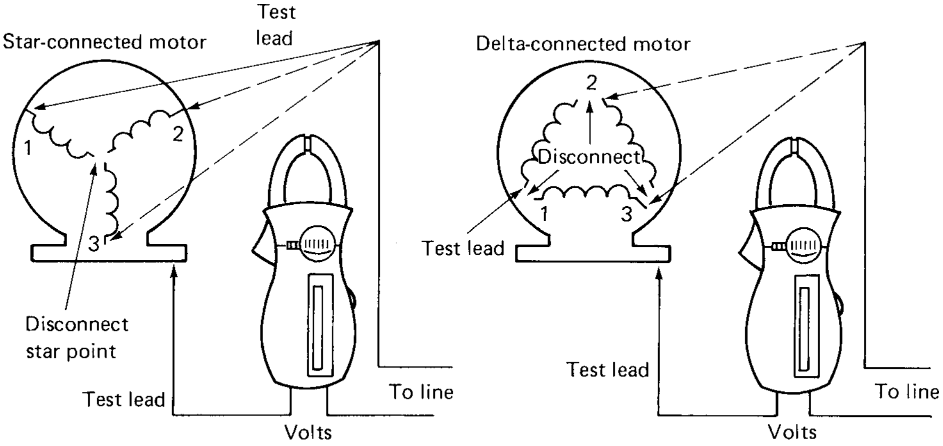
Star-connected motor
Test lead
1
2
3
Disconnect star point
Test lead
To line
Volts
Delta-connected motor
2
Disconnect
1
3
Test lead
Test lead
To line
Volts

Figure 18-7 Grounded phase of a motor. (Courtesy of Amprobe)

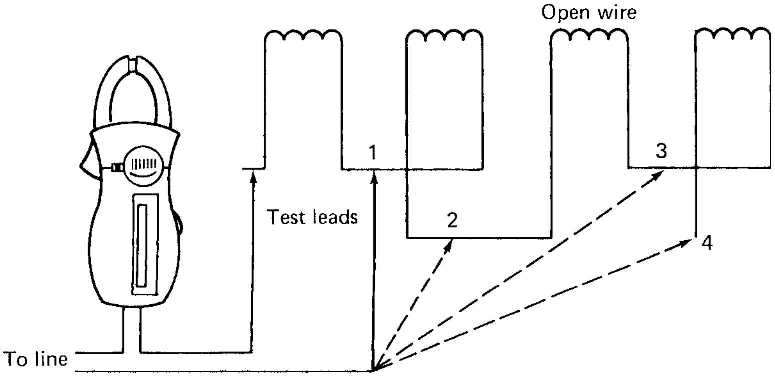
Open wire
1
3
Test leads
2
4
To line

Figure 18-8 Isolating an open phase. (Courtesy of Amprobe)

reading. This is due mainly to the capacitive effect between the windings and the steel lamination.

To locate the grounded portion of the windings, disconnect the necessary connection jumpers and test. Grounded sections will be detected by a full line voltage indication.

Looking for opens

To determine whether a winding is open, connect test leads as shown in Figs. 18-8 and 18-9. If the winding is open, there will be no voltage indication. If the circuit is not open, the voltmeter indication will read full line voltage.

Looking for shorts

Shorted turns in the winding of a motor behave like a shorted secondary of a transformer. A motor with a shorted winding will draw excessive current while running at no load. Measurement of the current can be made without disconnecting lines.

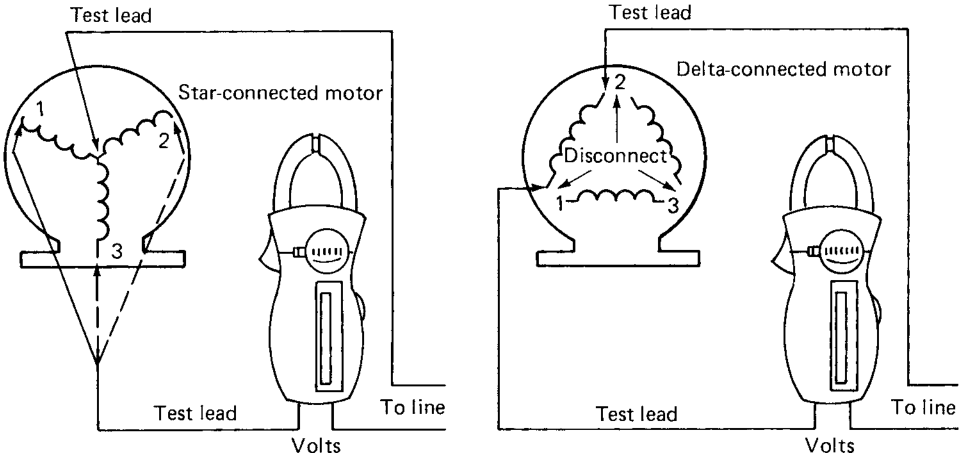
Test lead
Star-connected motor
1
2
3
Test lead
To line
Volts
Test lead
Delta-connected motor
2
Disconnect
1
3
Test lead
To line
Volts

Figure 18-9 Finding an open phase. (Courtesy of Amprobe)

You engage one of the lines with the split-core transformer of the tester. If the ammeter reading is much higher than the full-load ampere rating on the nameplate, the motor is probably shorted.

In a two- or three-phase motor, a partially shorted winding produces a higher current reading in the shorted phase. This becomes evident when the current in each phase is measured.

Motors with Squirrel-Cage Rotors

Loss in output torque at rated speed in an induction motor may be due to opens in the squirrel-cage rotor. To test the rotor and determine which rotor bars are loose or open, place the rotor in a growler. Engage the split-core ammeter around the lines going to the growler, as shown in Fig. 18-10. Set the switch to the highest current range. Switch on the growler and then set the test unit to the approximate current range. Rotate the rotor in the growler and take note of the current indication whenever the growler is energized. The bars and end rings in the rotor behave similarly to a shorted secondary of a transformer. The growler windings act as the primary. A good rotor will produce approximately the same current indications for all positions of the rotor. A defective rotor will exhibit a drop in the current reading when the open bars move into the growler field.

Testing the Centrifugal Switch in a Single-Phase Motor

A defective centrifugal switch may not disconnect the start winding at the proper time. To determine conclusively that the start winding remains in the circuit, place the split-core ammeter around one of the start-winding leads. Set the instrument to the highest current range. Turn on the motor switch. Select the appropriate current range. Observe if there is any current in the start-winding circuit. A current indication signifies that the centrifugal switch did not open when the motor came up to speed (see Fig. 18-11).

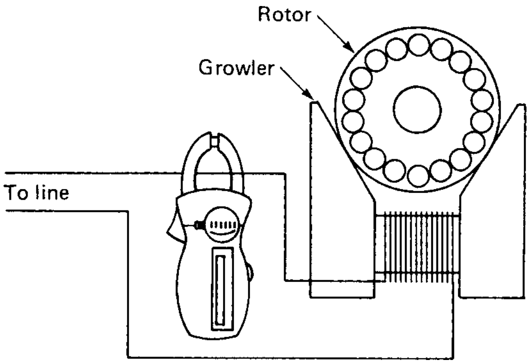

Figure 18-10 Using a growler to test a motor. (Courtesy of Amprobe)

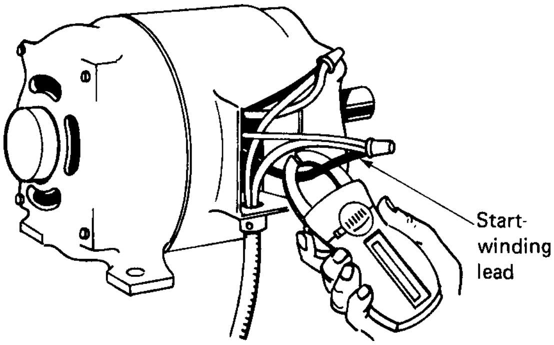

Figure 18-11 Checking the centrifugal switch with a clamp-on meter. (Courtesy of Amprobe)

Testing for a Short Circuit Between Run and Start Windings

A short between run and start windings may be determined by using the ammeter and line voltage to check for continuity between the two separate circuits. Disconnect the run- and start-winding leads and connect the instrument as shown in Fig. 18-12. Set the meter on voltage. A full-line voltage reading will be obtained if the windings are shorted to one another.

Capacitor testing

Defective capacitors are very often the cause of trouble in capacitor-type motors. Shorts, opens, grounds, and insufficient capacity in microfarads are conditions for which capacitors should be tested to determine whether they are good.

You can determine a grounded capacitor by setting the instrument on the proper voltage range and connecting it and the capacitor to the line as shown in Fig. 18-13. A full-line voltage indication on the meter signifies that the capacitor is grounded to the can. A high-resistance ground is evident from a voltage

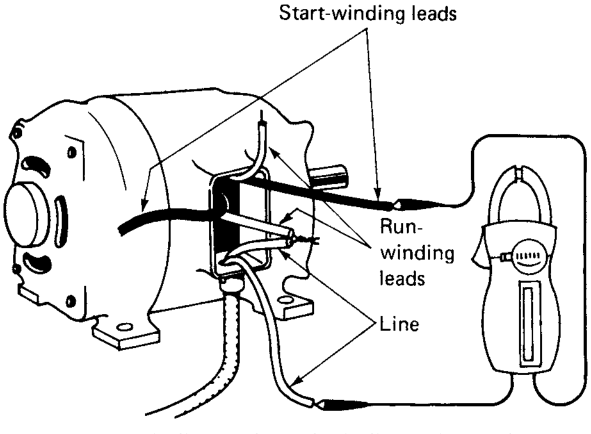

Figure 18-12 Finding a shorted winding using a clamp-on meter. (Courtesy of Amprobe)

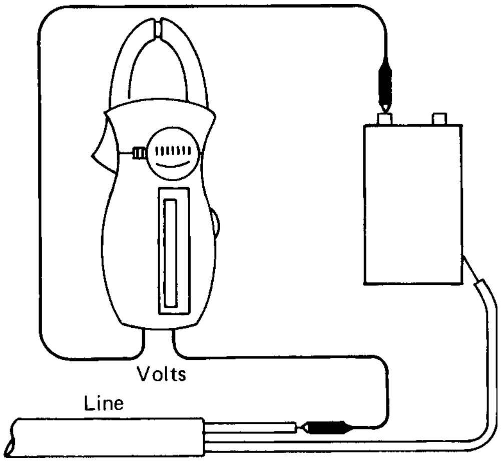

Figure 18-13 Finding a grounded capacitor with a clamp-on meter. (Courtesy of Amprobe)

reading that is somewhat below the line voltage. A negligible reading or a reading of no voltage indicates that the capacitor is not grounded.

Measuring the capacity of a capacitor

To measure the capacity of the capacitor, set the test unit's switch to the proper voltage range and read the line voltage indication. Then set to the appropriate current range and read the capacitor current indication. During the test, keep the capacitor on the line for a very short period of time, because motor-starting electrolytic capacitors are rated for intermittent duty (see Fig. 18-14). The capacity in microfarads is then computed by substituting the voltage and current readings in the following formula, assuming that a full 60-hertz line is used:

$$\text{Microfarads} = \frac{2650 \times \text{amperes}}{\text{volts}}$$

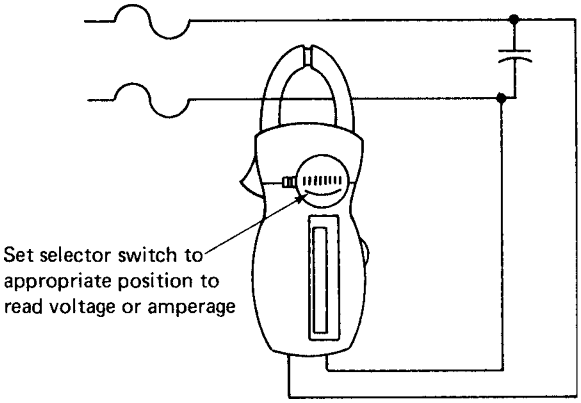
Set selector switch to appropriate position to read voltage or amperage

Figure 18-14 Finding the size of a capacitor with a clamp-on meter. (Courtesy of Amprobe)

An open capacitor will be evident if there is no current indication in the test. A shorted capacitor is easily detected. It will blow the fuse when the line switch is turned on to measure the line voltage. Troubleshooting can be broken down into a simple procedure. Logical thinking is what it takes to be able to accurately diagnose a problem and then correct it. For instance, you can isolate a particular system since you know how the system is supposed to operate. If it is not operating according to expectations, then you know something is causing the problem. By concentrating on a particular system that has been located, it is possible to concentrate your thinking on things that would cause the malfunction.

Once you have determined which components in the system can cause the symptoms being experienced, it is possible to find the faulty component.

After the problem has been located and the component identified, it is then necessary to take the component out and repair it or replace it.

Once the problem has been located, the part removed and a new one inserted or the old one repaired is necessary to check the operation of the system to make sure it is performing according to its design characteristics.

All this can be broken down into key words that will aid you in any troubleshooting situation.

Troubleshooting Procedure

A stepwise set of actions to be taken to troubleshoot has been described next.

Isolate. Isolate the system giving the trouble. The system may be:

- the evaporator
- the condenser
- the compressor
- the controls

Use the Problem-Probable Cause-Remedy charts to aid in isolating the problem.

Concentrate. Concentrate on the isolated system. Concentrate on how it works and why it would malfunction.

Differentiate. Determine which components could cause the identified symptoms; then locate each possible troublemaking part.

Eliminate. Eliminate the possible troublemakers one by one. Use the proper test instrument to do so.

Repair or replace. Either repair or replace the identified troublemaker.

Check. Check to see if the unit operates properly with the repaired or replaced parts.

Review Questions

1. What are some common sense rules for safely working around air-conditioning and refrigeration equipment?
2. How often should filters on air conditioners be cleaned?
3. How does low supply voltage affect the operation of electrical equipment?
4. What does ARI mean?
5. What is the ARI function?
6. What is the source of most trouble in air-conditioning systems?
7. What is the source of most trouble in heating systems?
8. How do you test motor capacitors?
9. How do you use a clamp-on ammeter to test electric motors?
10. How do you test the centrifugal switch in a single-phase motor?
11. How do you measure the capacitance of a capacitor in a piece of cooling equipment?
12. What is a good troubleshooting procedure?

This page intentionally left blank

Chapter

19

Controlling Electrical Power for Air-Conditioning Units

Performance Objectives

- Know how to find the correct wire size for a refrigeration unit.
- Know the permissible maximum voltage drops in a service line.
- Know the effects of voltage variations on AC motors.
- Know how to calculate starting current values and inrush voltage drops.
- Know the code limitations on amperes per conductor.
- Know how to select and correct circuit protection.
- Know the different types of fuses and circuit breakers.
- Know the latest in modules for control of refrigeration units.
- Know some of the electronic pressure controls for refrigeration systems.
- Know how sensing of current is used for refrigeration control.

In order to operate properly, the correct voltage, amperage, and wattage must be furnished to the air-conditioning unit. A number of devices have to be utilized to keep the voltage source clean and available for efficient operation of the unit. It all starts with the choosing of the correct wire from the power pole to the house or store or school where the air-conditioning unit is used.

Choosing Wire Size

There are two criteria for choosing wire size for installation of air-conditioning or refrigeration equipment. The size of the electrical conductor wire recommended for a given appliance circuit depends upon two things: limitation on voltage loss and minimum wire size.

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

Limiting voltage loss

Proper operation of an electrical device must be under the conditions for which it was designed. The wire size selected must be low in resistance per foot of length. This will assure that the full load "line loss" of the total length of the circuit does not cause low voltage of the appliance terminals. Since the length of electrical feeders varies with each installation, wire sizing to avoid excessive voltage loss becomes the responsibility of the installing contractor. The National Electrical Code or local code should be followed.

Minimum wire size

To avoid field wiring being damaged by tensile stress or overheating, national and local codes establish minimum wire sizes. The maximum amperage permitted for a given conductor limits internal heat generation so that temperature will not damage its insulation. This assumes proper fusing that will limit the maximum current flow so that the conductor will always be protected.

Wire size and voltage loss go hand in hand, so to speak. The larger the wire, the more current it can handle without voltage loss along the lines. Each conductor or wire has resistance. This resistance, measured in ohms per unit of wire length, increases as the cross-sectional area of the wire decreases. The size of the wire is indicated by gage number. The higher the gage number, the smaller the wire. American Wire Gage (AWG) is the standard used for wire size. Each gage number has a resistance value in ohms per foot of wire length. The resistance of aluminum wire is 64% greater than that for copper of the same gage number.

Wire selection

The wire size recommended for actual use should be the heavier of the two indicated by the procedures that will follow.

Local approval is usually necessary for any installation that has large current draws. The data presented here are based on the National Electrical Code. Much of the detail has been omitted in the interest of simplification. Thus, there may be areas of incompleteness not covered by a footnote or reference. In all cases it is recognized that final approval must come from the authority having local jurisdiction. The National Electrical Code sets forth minimum standards. It is an effort to establish some standard for safe operation of equipment.

Wire Size and Low Voltage

The voltage at which a motor or device should operate is stamped on the nameplate. This voltage indicates that the full capacity of the device is being utilized when that particular voltage is available. Motors operated at lower than rated voltage are unable to provide full horsepower without jeopardizing their service life. Electric heating units lose capacity even more rapidly at reduced voltages.

Low voltage can result in insufficient spark for oil burner ignition, reluctant starting of motors, and overheating of motors handling normal loads. Thus, it is not uncommon to protect electrical devices by selecting relays that will not close load circuits if the voltage is more than 15% below rating.

ARI (Air-Conditioning and Refrigeration Institute)-certified cooling units are tested to assure they will start and run at 10% above and 10% below their rating plate voltage. However, this does not imply that continuous operation at these voltages will not affect their capacity, performance, and anticipated service life. A large proportion of air-conditioning compressor burnouts can be traced to low voltage. Because the motor of a hermetic compressor is entirely enclosed within the refrigerant cycle, it is important that it not be abused either by overloading or under-voltage. Both of these can occur during peak load conditions. A national survey has shown that the most common cause of compressor low voltage is the use of undersized conductors between the utility lines and the condensing unit.

The size of the wire selected must be one that, under full load conditions, will deliver acceptable voltages to the appliance terminals. The National Electrical Code requires that conductors be sized to limit voltage drop between the outdoor-pole service tap and the appliance terminals to not in excess of 5% of rated voltage under full load conditions. This loss may be subdivided, with 3% permissible in service drops, feeders, meters, and overcurrent protectors at the distribution panel and the appliance (see Fig. 19-1).

In a 240-volt service, the wire size selected for an individual appliance circuit should cause no more than 4.8-volt drop under full load conditions. Even with this 5% limitation on voltage drop, the voltage at the equipment terminals is still very apt to be below the rating plate values. See Table 19-1.

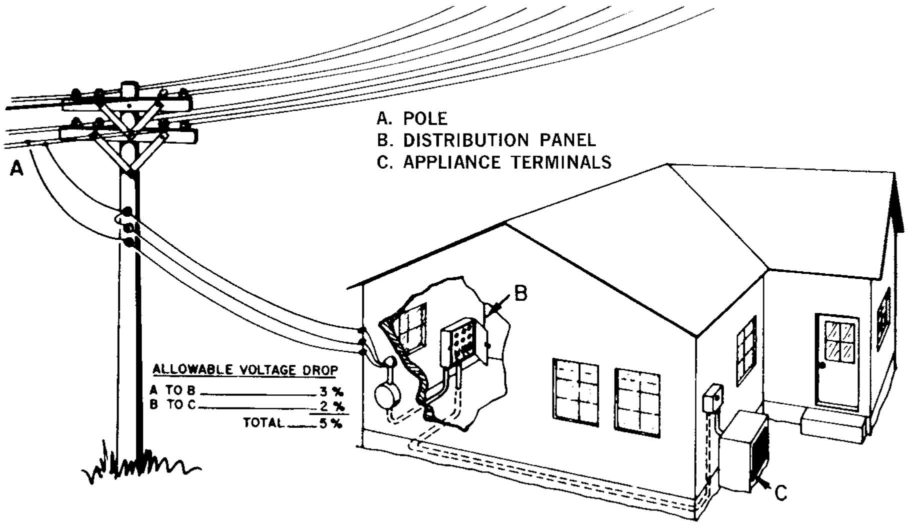

Figure 19-1 Voltage drops from the pole to air conditioner. (Courtesy of Bryant)

TABLE 19-1 Permissible Maximum Voltage Drops

For a line voltage of:	120	208	240	480
Feeders to distribution panel (3%)	3.6	6.24	7.2	14.4
Branch circuit to appliance (2%)	2.4	4.16	4.8	9.6
Total voltage drop fully loaded	6	10.4	12	24
Resultant* voltage at appliance	114	197.6	228	456

*Assumes full-rated voltage where feeders connect to utility lines. If utility voltage runs low, the overall voltage drop should be further reduced so as to make available at the appliance terminals a voltage as close as practical to that voltage range specified on the appliance rating plate.

Voltage drop calculations

Just as friction creates pressure loss in water flow through pipes, so does electrical resistance create voltage drop as current flows through a conductor. The drop increases with the length of the conductor (in feet), the current flow (in amperes), and the ohms of resistance per foot of wire. This relationship may be expressed as follows:

$$\text{Voltage drop} = \text{amperes} \times \text{ohms/foot} \times \text{length of conductor}$$

Figure 19-2 illustrates how voltage drop per 100 feet of copper conductor will increase with the amount of current drawn through the conductor. The wire size

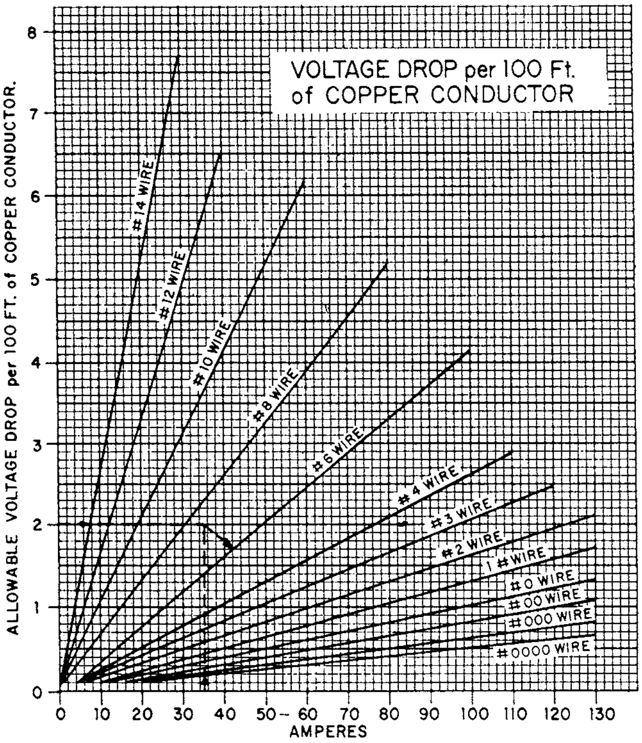

Figure 19-2 Conductor voltage drop per 100 feet.

is indicated on the straight line. Match the amperes with the wire size. Then follow over to the left column to determine the voltage drop. For instance, Fig. 19-2 shows a 2.04-volt drop per 100 feet of copper conductor for 20 ampers of current through a No. 10 wire.

The Effects of Voltage Variations on AC Motors

Motors will run at the voltage variations already mentioned. This does not imply that such operation will comply with industry standards of capacity temperature rise, or normally anticipated service life. Figure 19-3 shows general effects. Such effects are not guaranteed for specific motors.

The temperature rise and performance characteristics of motors sealed within hermetic compressor shells constitute a special case. These motors are cooled by return suction gas of varying quantity and temperature. Thus, Fig. 19-3 is not necessarily applicable to this specialized type of equipment.

The chart shows the approximate effect of voltage variations on motor characteristics. The reference base of voltage and frequency is understood to be that shown on the nameplate of the motor.

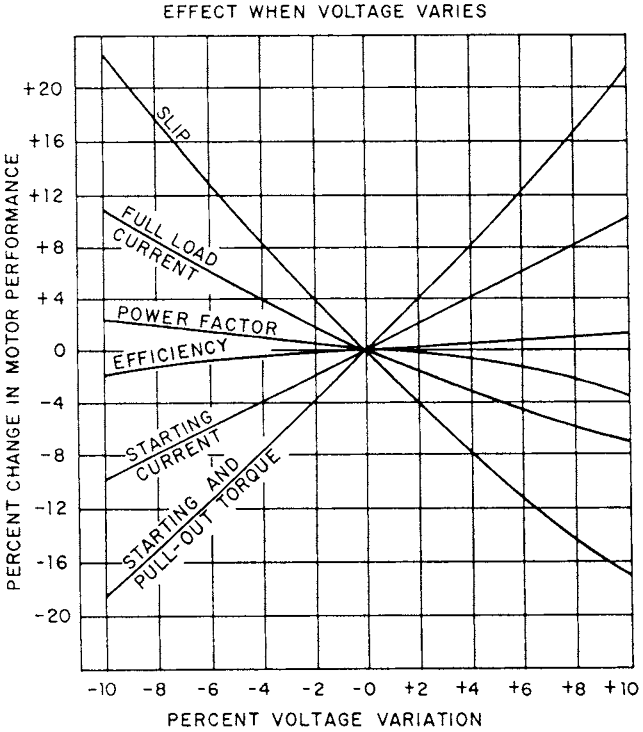

Figure 19-3 General effects of voltage variations on induction motor characteristics.

Some of the terms used in the chart are explained here.

Normal slip = synchronous speed – the rating plate speed.

Slip in the graph indicates the change in normal slip.
Synchronous speeds for 60-hertz motors are:

2-pole—3600 r/min or rpm

4-pole—1800 r/min or rpm

6-pole—1200 r/min or rpm

8-pole—900 r/min or rpm

Table 19-2 indicates the voltage drop that may be anticipated for various ampere flow rates through copper conductors of different gage size. Figure 19-2 provides the same data in graphic form.

These data are applicable to both single-phase and three-phase circuits. In each case, the wire length equals twice the distance from the power distribution panel to the appliance terminals, measured along the path of the conductors. This is twice the distance between B and C in Fig. 19-1, measured along

TABLE 19-2 Voltage Drop per 100 Feet of Copper Conductor of Wire Gage

Amperes*	No. 14	No. 12	No. 10	No. 8.	No. 6	No. 4	No. 3	No. 2	No. 1	No. 0	No. 00	No. 000	No. 0000
5	1.29	0.81	0.51	0.32	0.21	0.13	0.11						
10	2.57	1.62	1.02	0.64	.041	0.26	0.21	0.16	0.13	0.10			
15	3.86	2.3	1.53	0.96	0.62	0.39	0.31	0.24	0.19	0.15	0.12	0.10	
20	5.14	3.24	2.04	1.28	0.82	0.52	0.41	0.32	0.26	0.20	0.16	0.13	0.10
25	6.43	4.05	2.55	1.60	1.03	0.65	0.51	0.41	0.32	0.26	0.20	0.16	0.13
30	7.71	4.86	3.06	1.92	1.23	0.78	0.62	0.49	0.39	0.31	0.24	0.19	0.15
35		5.67	3.57	2.24	1.44	0.91	0.72	0.57	0.45	0.36	0.28	0.22	0.18
40		6.48	4.08	2.56	1.64	1.04	0.82	0.65	0.52	0.41	0.32	0.26	0.20
45			4.59	2.88	1.85	1.17	0.92	0.73	0.58	0.46	0.36	0.29	0.23
50			5.10	3.20	2.05	1.30	1.03	0.81	0.65	0.51	0.41	0.32	0.26
60			6.12	3.84	2.46	1.56	1.23	0.97	0.77	0.61	0.49	0.38	0.31
70				4.48	2.87	1.82	1.44	1.13	0.90	0.71	0.57	0.45	0.36
80				5.12	3.28	2.08	1.64	1.30	1.03	0.82	0.65	0.51	0.41
90					3.69	2.34	1.85	1.46	1.16	0.92	0.73	0.58	0.46
100					4.10	2.59	2.05	1.62	1.29	1.02	0.81	0.64	0.51
110						2.85	2.26	1.78	1.42	1.12	0.89	0.70	0.56
120							2.46	1.94	1.55	1.22	0.97	0.77	0.61
130								2.10	1.68	1.33	1.05	0.83	0.66
140									1.81	1.43	1.13	0.90	0.71
150										1.53	1.22	0.96	0.77
Ohms/100 feet copper	0.257	1.62	0.1018	0.064	0.041	0.0259	0.0205	0.0162	0.0128	0.0102	0.0081	0.0064	0.0051
Aluminum	4.22	0.266	0.167	0.105	0.0674	0.0424	0.0336	0.0266	0.0129	0.0168	0.0133	0.0105	0.0084

*To determine voltage drop for aluminum, enter the chart number then using 1.64 × actual amperes. The conductor's length is twice the length of the branch leads, whether single or three phase. Since resistance varies with temperature, it may be necessary to correct for wire temperature under load conditions if the ambient materially exceeds 80°F. If so, increase ampere values using the multiplier 1.0 + 0.002 × (ambient temperature − 80°F). Example: If current flow and environment result in conductors reaching 140°F under load conditions, the appliance ampere ratings should be increased by the multiplying factor 1.0 + 0.002 (140°F – 80°) = 1.0 + 0.12 = 1.12.

the path of the conductors. For motorized appliances, particularly those that start under loaded conditions, the voltage at the appliance terminals should not drop more than 10% below rating plate values unless approved by the manufacturer. Thus, the voltage drop permissible in the load leads must anticipate any reduction below rated voltage that may be suffered under full load conditions at the point of power source connection (point A in Fig. 19-1).

Troublesome voltage losses may also occur elsewhere if electrical joints or splices are mechanically imperfect and create unanticipated resistance. Such connections may exist in the distribution panel, the meter socket, or even where outdoor power drops are clamped to the feeder lines on poles. Where there is a wide variation between no-load voltages and operating voltage, sources of voltage drop can be determined by taking voltmeter readings at various points in the circuit. These points might be ahead of the meter, after the circuit disconnect switch, at the appliance terminals, and at other locations.

Selecting Proper Wire Size

To provide adequate voltage at the appliance terminals, anticipate the minimum voltage that may exist at the distribution panel. Then determine the allowable voltage drop acceptable in the appliance circuit. This should not exceed 2% of rated voltage. It should, for example, not exceed 4.1 volts for 208-volt service. Table 19-1 shows voltage drops for 120-, 208-, 240-, and 408-volt service.

Determine the length of feed conductor. This is twice the length of the wire path from the source to the appliance. In Fig. 19-1, this is two times the distance from B to C measured along the path of the wire. If it is single-phase or three-phase, consider two conductors in establishing the total length of the circuit.

Determine the allowable drop per 100 feet of conductor.

Example: If for a 230-volt installation, a 4.6-volt drop is permissible, and the wire path is 115 feet from the distribution panel to the appliance (this makes 230 feet of conductor), then the allowable drop per 100 feet will be:

$$4.6 \text{ volts} \times \frac{100 \text{ feet}}{230 \text{ feet}} = 2.0 \text{ volts per 100 feet}$$

Using either Table 19-2 or Fig. 19-2, determine the gage wire required. When using the graph, select the gage number closest below and to the right.

Example: The full load value is 35 amperes. The allowable voltage drop is 2.0 volts per 100 feet. See Fig. 19-2.

Table solution. (Table 19-2): Select No. 6 wire. This results in a drop of 1.44 volts per 100 ft.

Graph solution. (Fig. 19-2): Intersection lies between No. 6 wire and No. 8 wire. Select the larger of the two. In this case it would be No. 6.

Unacceptable Motor Voltages

Occasionally it becomes necessary to determine causes of unacceptable voltage conditions at motor terminals. Often this is necessary where excessive voltage drops are encountered as motors start. During this brief interval, the starting inrush current may approximate a motor's locked rotor amperage rating.

Table 19-3 shows the range of locked rotor amperes (LRA) per motor horsepower. LRA appears on the rating plates of hermetic compressors. Depending on the type of motor, its LRA may be two to six times its rated full load current. Motor-starting torque varies as the square of the voltage. Thus, only 81% of the anticipated torque is available if the voltage drops to 90% of the rating during the starting period.

The full load amperage value must be considered in choosing the proper wire size and making sure the motor has acceptable voltages. These are shown in Table 19-4.

Calculating Starting Current Values and Inrush Voltage Drops

Single-phase current

Wire size and inrush voltage drop can be calculated. The following formula can be used for single-phase current.

TABLE 19-3 Range of Locked Rotor Amperes per Motor Horsepower

NEMA	115	208		230		460	
Code letter	1ϕ	1ϕ	3ϕ	1ϕ	3ϕ	1ϕ	3ϕ
A	0–27.4	0–15.1	0–9.1	0–13.7	0–7.9	0–6.9	0–4.0
B	27.5–30.9	15.2–17.0	9.2–9.8	13.8–15.5	8.0–9.0	7.0–7.7	4.1–4.5
C	31.0–34.8	17.1–34.8	9.9–11.2	15.6–17.4	9.1–10.1	7.8–8.7	4.6–5.0
D	34.9–39.2	19.5–21.6	11.3–12.5	17.5–19.6	10.2–11.3	8.8–9.8	5.1–5.7
E	39.3–43.5	21.7–24.0	12.6–13.9	19.7–21.7	11.4–12.5	9.9–10.9	5.8–6.3
F	43.6–48.7	24.1–26.9	14.0–15.5	21.8–24.4	12.6–14.1	11.0–12.2	6.4–7.0
G	48.8–54.8	27.0–30.3	15.6–17.5	24.5–27.4	14.2–15.8	12.3–13.7	7.1–7.9
H	54.9–61.7	30.4–33.7	17.6–19.5	27.5–30.6	15.9–17.7	13.8–15.3	8.0–8.8
J	61.8–69.6	33.8–38.4	19.6–22.2	30.7–34.8	17.8–20.1	15.4–17.4	8.9–10.1
K	69.7–78.4	38.5–43.3	22.3–25.0	34.9–39.2	20.2–22.6	17.5–19.6	10.2–11.3
L	78.5–87.1	43.4–48.0	25.1–27.7	39.3–43.2	22.7–25.2	19.7–21.8	11.4–12.6
M	87.2–97.4	48.1–53.8	27.8–31.1	43.3–48.7	25.3–28.7	21.9–24.4	12.7–14.1
N	97.5–109	53.9–60.0	31.2–34.6	48.7–54.5	28.3–31.5	24.5–27.3	14.2–15.8
P	110–112	60.1–67.2	34.7–38.8	54.6–61.0	31.6–35.2	27.4–30.5	15.9–17.6
R	123–139	67.3–76.8	38.9–44.4	61.1–69.6	35.3–40.2	30.6–34.8	17.7–20.1
S	140–157	76.9–86.5	44.5–50.0	69.7–78.4	40.3–45.3	34.9–39.2	20.2–22.6
T	158–174	86.6–96.0	50.1–55.5	78.5–87.0	45.4–50.2	39.3–43.5	22.7–25.1
U	175–195	96.1–108	55.6–56.4	87.1–97.5	50.3–56.3	44.5–48.8	25.2–28.2
V	196 and up	109 and up	56.6 and up	97.6 and up	56.4 and up	48.9 and up	28.3 and up

NOTE: Locked rotor amperes appear on rating plates of hermetic compressors. The NEMA code letter appears on the motor rating plate. Multiply above values by motor horsepower.

TABLE 19-4 Approximate Full Load Amperage Values for AC Motors

Motor	Single phase*		Three-phase, squirrel cage induction		
HP	115 Volts	230 CV	230 Volts	460 Volts	575 Volts
1/6	4.4	2.2			
1/4	5.8	2.9			
1/3	7.2	3.6			
1/2	9.8	4.9	2	1.0	0.8
3/4	13.8	6.9	2.8	1.4	1.1
1	16	8	3.6	1.8	1.4
11/2	20	10	5.2	2.6	2.1
2	24	12	6.8	3.4	2.7
3	34	17	9.6	4.8	3.9
5	56	28	15.2	7.6	6.1
71/2			22	11.0	9.0
10			28	14.0	11.0
15			42	21.0	17.0
20			54	27.0	22.0
25			68	34.0	27.0

Example: If a single-phase 230-volts condensing unit, rated at 22 amperes full-load and having a starting current of 91 amperes is located 125 feet from the distribution panel and so utilizes 250 feet of the No. 10 copper wire, the voltage drop expected during full load operation is calculated as follows:

Refer to the lower lines of Table 19-2. Note that the resistance of No. 10 copper wire is 0.1018 ohms per 100 feet.

$$\text{Voltage drop} = 22 \text{ amperes } \times 0.1081 \times \frac{250 \text{ feet}}{100} = 5.6 \text{ volts}$$

(Note that 5.6 volts exceeds the 2% loss factor, which is 4.6 volts.) If the full 3% loss (6.9 volts) allowed ahead of the meter is present, then the voltage at the load terminal of the meter will be 223.1 volts (230 – 6.9 = 223.1). Subtract the voltage drop calculated above and there will be only 197.5 volts at the unit terminals during full load operation.

Thus,

$223.1 - 5.6 = 197.5$. With a total loss of 5.4%, ($230 - 197.5 = 12.5$, or 5.47%),

it is a common practice to move to the next largest wire size. Therefore, for this circuit, AWG No. 8 wire should be used instead of No. 10.

Insofar as motor starting and relay operation are concerned, the critical period is during the initial instant of start-up when the inrush current closely approximates the locked rotor value. For the equipment described in the above example, the voltage drop experienced at 91 amperes flow for No. 10 wire is again excessive, indicating the wisdom of using No. 8 wire.

For No. 10 wire=

$$\begin{aligned}\text{Inrush voltage drop} &= 91 \text{ amperes} \times \frac{0.1018 \text{ ohms}}{100 \text{ feet}} \times 250 \text{ feet}\\ &= 24.57 \text{ volts}\end{aligned}$$

For No. 8 wire

$$\begin{aligned}\text{Inrush voltage drop} &= 91 \text{ amperes} \times \frac{0.064 \text{ ohms}}{100 \text{ feet}} \times 250 \text{ feet}\\ &= 14.56 \text{ volts}\end{aligned}$$

For a 230-volt circuit, the 23.2 volts slightly exceeds a 10% drop between the meter and the appliance. To this must be added the voltage drop incurred in the lead-in wires from the outdoor power line. This total must then be deducted from the power line voltage on the poles, which may be less than 230 volts during utility peak load periods. Although the inrush current may exist for only an instant, this may be long enough to cause a starting relay to open, thus cutting off current to the motor. Without current flow, the voltage at the unit immediately rises enough to re-close the relay, so there is another attempt to start the motor. While the unit may get underway after the second or third attempt, such "chattering relay" operation is not good for the relay, the capacitors, or the motor.

For electrical loads such as lighting, resistance heating, and cooking, inrush current may be considered the equivalent of normal current flow. In the case of rotating machinery, it is only during that initial period or rotation that the start-up current exceeds that of final operation. The same is true of relays during the instant of "pull-in."

Three-phase circuits

Calculating the inrush voltage drop for three-phase circuits is the same as calculating the drop for single-phase circuits. Again, the value for circuit length equals twice the length of an individual conductor. Since more conductors are involved, the normal current and the starting current per conductor are smaller for a motor of a given size. Thus, lighter wire may be used.

Example: Using the same wire length as in the single-phase example and the lower values of 13.7 amperes full load and 61 amperes starting inrush per conductor for the three-phase rating of the same size compressor, the use of No. 10 conductor results in:

$$\begin{aligned}\text{Normal voltage drop} &= 13.7 \text{ amperes} \times \frac{0.1018 \text{ ohms}}{100 \text{ feet}} \times 250 \text{ feet}\\ \circ\ \circ\ \circ\ \circ\ \circ\ \circ\ \circ\ \circ\ \circ &= 3.5 \text{ volts}\\ \text{Inrush voltage drop} &= 61 \text{ amperes} \times \frac{0.1018 \text{ ohms}}{100 \text{ feet}} \times 250 \text{ feet}\\ \circ\ \circ\ \circ\ \circ\ \circ\ \circ\ \circ\ \circ\ \circ &= 15.5 \text{ volts}\end{aligned}$$

Inrush voltage drop

The actual inrush current through an appliance usually is somewhat less than the total of locked-rotor current values. Locked-rotor current is measured with rated voltage at the appliance terminals. Since voltage drop in the feed lines reduces the voltage available at the terminals, less than rated voltage can be anticipated across the electrical components. Consequently, inrush currents and voltage drops are somewhat less. This fact is illustrated in the following, which is based on the same installation as that in the previous single-phase examples. However, here the actual locked-rotor current of 101 amperes is used. The formula can be found on the previous page.

Code Limitations on Amperes Per Conductor

Varied mechanical conditions are encountered in field wiring. Thus, the National Electrical Code places certain limitations on the smallness of conductors installed in the field. Such limitations apply regardless of conductor length. They assure the following:

1. That the wire itself has ample strength to withstand the stress of pulling it through long conduits and chases. With specific exceptions, no wire lighter than No. 14 copper is permitted for field wiring of line voltage power circuits.
2. By stipulating the maximum amperage permissible for each wire gage, self-generated heat can be limited to avoid temperature damage to wire insulation. If wiring is installed in areas of high-ambient temperature, the amperage rating may need to be reduced.
3. By stipulating the maximum amperage of overload protectors for circuits, current flow is limited to safe values for the conductor used. Some equipment has momentary starting currents that trip out overload protectors sized on the basis of full load current. Here heavier fusing is permissible, but only under specific circumstances. Current flow limitations for each gage protect wire insulation from damage due to overheating.

Heat Generated Within Conductors

Heat generation due to current flow through the wire is important for the following two reasons.

1. Temperature rise increases the resistance of the wire and, therefore, the voltage drop in the circuit. Under most conditions of circuit usage, this added resistance generates additional heat in the wires. Finally, a temperature is reached where heat dissipation from the conductors equals the heat that they generate. It is desirable to keep this equilibrium temperature low.

The number of Btu generated can be found by both of the following formulas.

$$\text{Indicated locked} - \text{rotor impedance} = \frac{\text{rated voltage}}{\text{locked} - \text{rotor amperes}}$$

$$= 230 \text{ volts}/101 \text{ amperes} = 2.28 \text{ ohms}$$

$$\text{Resistance of No. 10 leads} = \text{ohms}/100 \text{ feet} \times \text{length of wire (in feet)}$$

$$= 0.1018 \text{ ohms}/100 \text{ feet} \times 250 \text{ feet} = 0.25 \text{ ohms}$$

$$\text{Total indicated load and conductor impedance} = (2.28 + 0.25) = 2.53 \text{ ohms}$$

$$\text{Inrush current} = \frac{\text{distribution panel voltage}}{\text{total indicated impedance}}$$

$$\text{Inrush current} = \frac{230 \text{ volts}}{2.53 \text{ ohms}} = 91 \text{ amperes}$$

$$\text{Calculating inrush voltage drop} = \frac{91.0 \times 0.1018 \text{ ohms} \times 250 \text{ feet}}{100 \text{ feet}}$$

$$= 23.2 \text{ volts}$$

$$\text{Locked rotor voltage drop} = \frac{101 \times 0.1018 \text{ ohms} \times 250 \text{ feet}}{100 \text{ feet}} = 25.7 \text{ volts}$$

$$\text{Btu generated} = \text{amperes}^2 \times \text{resistance in ohms} \times 3.4313^*$$

2. Temperature also damages wire insulation. The degree of damage is dependent upon the insulation's ability to withstand temperature under varying degrees of exposure, age, moisture, corrosive environment, mechanical abuse, and thickness.

Estimating the probable operating temperature of a conductor and its insulation is difficult. The rate of heat dissipation from the wiring surfaces varies with the ambient temperature, the proximity of other heat-generating conductors, the heat conductivity of the insulation and jacket material, the availability of cooling air, and other factors. Freestanding individual conductors dissipate heat more effectively. However, the typical situation of two or three conductors, each carrying equal current and enclosed in a common jacket, cable, or conduit, anticipates limitations as set forth by the National Electrical Code.

Circuit Protection

Circuits supplying power to appliances must incorporate some means for automatically disconnecting the circuit from the power source should there be abnormal current flow due to accidental grounding, equipment overload, or short circuits. Such overload devices should operate promptly enough to limit the buildup of damaging temperatures in conductors or in the electrical components of an appliance. However, devices selected to protect circuits feeding motors must be slow enough to permit the momentary inrush of heavy starting current. They must then disconnect the circuit if the motor does not start promptly, as can happen under low voltage conditions.

---

*Conversion factor

Devices heavy enough to carry continuously the motor starting current do not provide the overload protection desired. Likewise, heavily fused branch circuits do not adequately protect the low amperage components that cumulatively require the heavy fusing. For this reason some literature lists maximum allowable fuse sizes for equipment. While electrical components of factory-built appliances are individually safeguarded, the field combining of two or more units on one circuit may create a problem more complex than that normally encountered. Remember that the final authority is the local electrical inspector.

Standard rule

With a few exceptions, the ampere capacity of an overload protector cannot exceed the ampacity values listed by wire size by the NEC. (Check the National Electrical Code for these exceptions.) If the allowable ampacity of a conductor does not match the rating of a standard size fuse or nonadjustable trip-circuit breaker, the device with the next largest capacity should be used. Some of the standard sizes of fuses and nonadjustable trip-circuit breakers are: 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 125, 150, 175, 200, 225, 250, and 300 amperes.

Fuses

One-time single-element fuses

If a current of more than rated load is continued sufficiently long, the fuse link becomes overheated. This causes the center portion to melt. The melted portion drops away. However, due to the short gap, the circuit is not immediately broken. An arc continues and burns the metal at each end until the arc is stopped because of the very high increase in resistance. The material surrounding the link tends

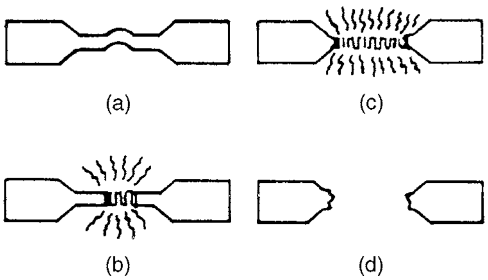

Figure 19-4 How a fuse works.

to break the arc mechanically. The center portion melts first, because it is farthest from the terminals that have the highest heat conductivity. See Fig. 19-4.

Fuses will carry a 10% overload indefinitely under laboratory-controlled conditions. However, they will blow promptly if materially overloaded. They will stand 150% of the rated amperes for the following time periods:

1 minute (fuse is 30 amperes or less).
2 minutes (fuse of 31 to 60 amperes).
4 minutes (fuse of 61 to 100 amperes).

Time-delay two-element fuses

Two-element fuses use the burnout link described previously. They also use a low-temperature soldered connection that will open under overload. This soldered joint has mass, so it does not heat quickly enough to melt if a heavy load is imposed for only a short time. However, a small but continuous overload will soften the solder so that the electrical contact can be broken.

With this type of protection against light overloads, the fusible link can be made heavier, yet blow quickly to protect against heavy overloads. This results

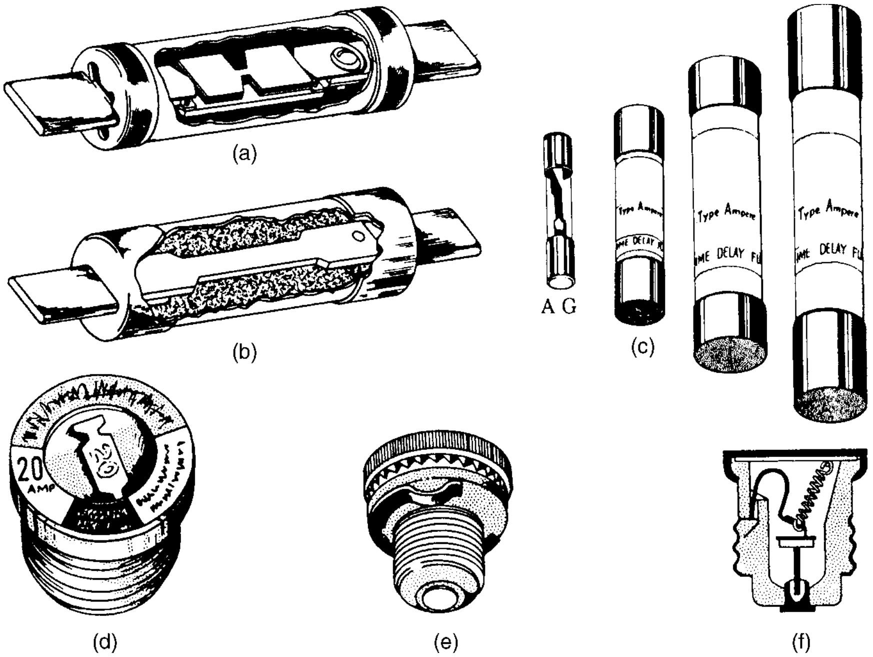

Figure 19-5 Fuse types.

in fewer nuisance burnouts and equipment shutdowns. Two types of dual-element fuses are shown in Fig. 19-5.

Types of fuses

In addition to those fuses just described, there are three general categories based on shape and size.

1. The AG (automotive glass) fuse consists of a glass cylinder with metallic end caps between which is connected a slender metal element that melts on current overload. This fuse has a length of $1^5/_{16}$ and a diameter of $^1/_4$. It is available only for low amperages. While used in specific appliances, it is not used to protect permanently installed wiring.
2. Cartridge fuses are like AG fuses. However, they are larger. The cylindrical tube is fiber, rather than glass. The metallic end pieces may be formed as lugs, blades, or cylinders to meet a variety of fuse box socket requirements. The internal metal fusible link may be enclosed in sand or powder to quench the burnout arc. Cartridge fuses are made in a variety of dimensions, based on amperage and voltage. Blade-type terminals are common above 60 amperes. Fuses used to break 600-volt arcs are longer than those for lower voltages. Fuses are available in many capacities other than the listed standard capacities, particularly in the two-element, time-delay types. Often, they are so dimensioned as to not be interchangeable with fuses of other capacities.
3. Plug fuses are limited in maximum capacity to 30 amperes. They are designed for use in circuits of not more than 150 volts above ground. Two-element time-delay types are available to fit standard screw lamp sockets. They are also available with nonstandard threads made especially for various amperage ratings.

Thermostats

The thermostat (or temperature control) stops and starts the compressor in response to room temperature requirements. Each thermostat has a charged power element containing either a volatile liquid or an active vapor charge. The temperature-sensitive part of this element (thermostat feeler bulb) is located in the return airstream. As the return air temperature rises, the pressure of the liquid or vapor inside the bulb increases. This closes the electrical contacts and starts the compressor. As the return air temperature drops, the reduced temperature of the feeler bulb causes the contacts to open and stops the compressor.

The advent of transistors and the semiconductor chips or integrated circuits produced a more accurate method of monitoring and adjusting temperatures within a system. The microprocessor makes use of the semiconductor's and chip's abilities to compare temperatures. It can also program on and off cycles, as well as monitor the duration of each cycle. This leads to more accurate temperature control.

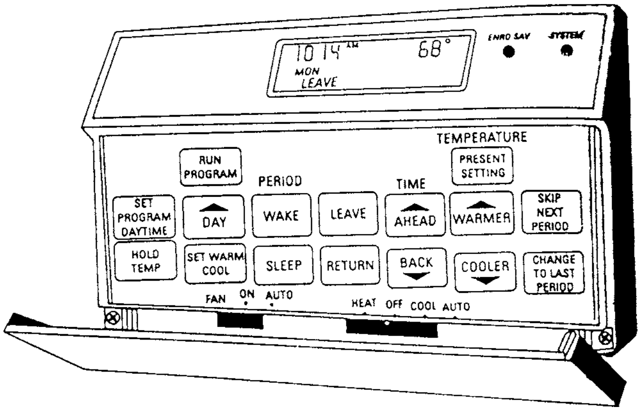
10 14 AM
68°
MON
LEAVE
ENRG SAV
SYSTEM
TEMPERATURE
RUN PROGRAM
PERIOD
TIME
PRESENT SETTING
SET PROGRAM DAYTIME
DAY
WAKE
LEAVE
AHEAD
WARMER
SKIP NEXT PERIOD
HOLD TEMP
SET WARM COOL
SLEEP
RETURN
BACK
COOLER
CHANGE TO LAST PERIOD
FAN
ON
AUTO
HEAT
OFF
COOL
AUTO

Figure 19-6 Microprocessor thermostat used for residential temperature and cooling control.

Figure 19-6 shows a microprocessor-based thermostat. As you can see from the front of the control panel, you can adjust the program to do many things. It can also save energy, whether it is operating the furnace for heat or the air-conditioning unit for cooling. These units usually come with a battery so that the memory can retain whatever is programmed into it. The battery is also a backup for the clock so that the program is retained even if the line power is interrupted.

Thermostat as a control switch

The control switch (thermostat) may be located in the room to be cooled or heated, depending upon the particular switch selection point. The control switch (heat, off, cool, and auto) is of the sliding type and normally has four positions, marked "heat," "off," "cool," and "auto." The thermostat is taken from its socket and programmed according to the manufacturer's directions. Then it is activated by plugging it into the wall socket and replacing a couple of screws to hold it in place.

To operate the unit as a ventilator, the switch on the left is marked "fan" with an "on" and "auto" choice to select the "fan" operation. When a thermostat is installed for automatic cooling, the compressor and fans will cycle according to dial requirements.

Figure 19-7 shows the electrical circuitry for a home heat-cool thermostat. Keep in mind that the thermostat should always be on an inside partition, never on an outside wall. Do not mount the instrument on a part of the wall that has steam or hot water pipes or warm air ducts behind it. The location should be such that direct sunshine or fireplace radiation cannot strike the thermostat. Be careful that the spot selected is not likely to have a floor lamp near it or a table lamp under it. Do not locate the thermostat where heat from kitchen appliances can affect it. Do not locate it on a wall that has cold unused room on the other side.

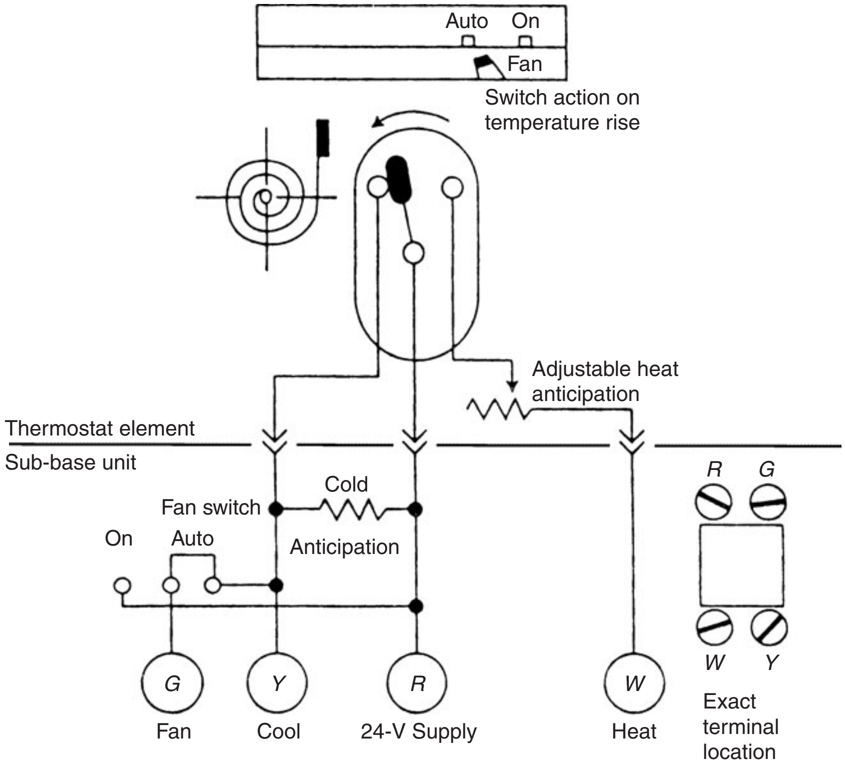
Auto
On
Fan
Switch action on temperature rise
Adjustable heat anticipation
Thermostat element
Sub-base unit
Cold
Fan switch
Anticipation
On
Auto
R
G
W
Y
G
Y
R
W
Fan
Cool
24-V Supply
Heat
Exact terminal location

Figure 19-7 Electrical circuitry of a home heat-cool thermostat.

After a thermostat has been mounted, it is wise to fill the stud space behind the instrument with insulating material. This is to prevent any circulation of cold air. Furthermore, the hole behind the thermostat for the wires should be sealed so that air cannot emerge from the stud space and blow across the thermostat element. It is quite common to find considerable air motion through this hole caused by a chimney effect in the stud space.

Service

Servicemen who have a good knowledge of refrigeration and air conditioning will be able to competently service air conditioners. Most air-conditioning units of present design contain compressors of the hermetic or sealed type. The only parts that can be serviced in the field are the relay, control switch, fan, fan motor, start and run capacitors, air filters, and cabinet parts. The refrigerating system (consisting of the cooling unit, condensers, compressors, and connecting lines) generally cannot be serviced in the field. Most servicemen will find the newer electronic-type thermostats easy to program when following the manufacturer's instructions included with every thermostat and usually given to the homeowner at closing on the house.

Starter Kits

There are kits of components to use with the PSC (Permanent Start Capacitor) that need additional starting torque. These kits are available for most of the pop-

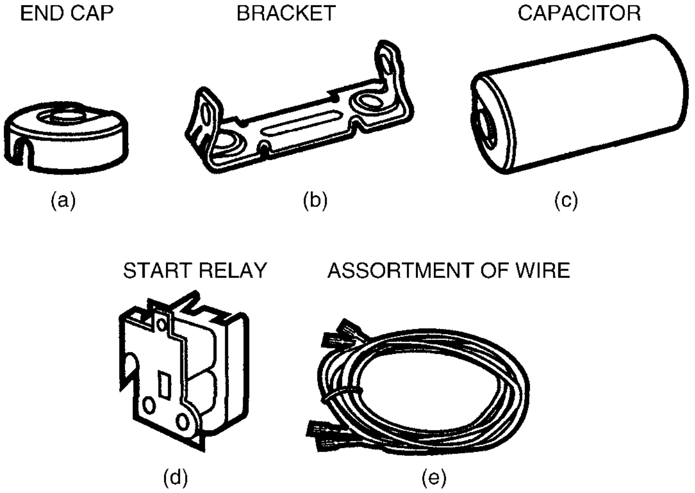

TOTALINE START KIT RATINGS						
START CAPACITOR		START RELAY				
MFD RANGE	VOLTS	CONTACT RATING	MAX OPER. VOLTS	PICK UP VOLTS	DROPOUT VOLTS	PART NO.
21-25	330	50 AMP	336	162-175	55-115	P296-0007
43-53	320	35 AMP	336	162-175	40-90	P296-0006
43-53	320	50 AMP	395	180-195	40-105	P296-0001
72-88	330	35 AMP	336	162-175	40-90	P296-0012
88-108	330	35 AMP	336	162-175	40-90	P296-0005
88-108	330	50 AMP	420	195-224	60-123	P296-0002
88-108	330	50 AMP	395	180-195	40-105	P296-0004
88-108	330	50 AMP	336	162-175	55-115	P296-0009
108-130	330	50 AMP	420	195-224	70-140	P296-C010
135-155	320	50 AMP	420	195-224	60-123	P296-0003
145-175	330	50 AMP	395	180-195	40-105	P296-0008
189-227	330	50 AMP	420	262-290	60-121	P296-0011

Figure 19-8 Starter kits. (Courtesy of Carrier)

TABLE 19-5 Start Capacitors and Relays

Start capacitor			Start relay			
Mfd range	Volts	Contact rating	Max oper. volts	Pick up volts	Dropout volts	Part no.
21–25	330	50 AMP	336	162–175	55–115	P296-0007
43–53	320	35 AMP	336	162–175	40–90	P296-0006
43–53	320	50 AMP	395	180–195	40–105	P296-0001
72–88	330	35 AMP	336	162–175	40–90	P296-0012
88–108	330	35 AMP	336	162–175	40–90	P296-0005
88–108	330	50 AMP	420	195–224	60–123	P296-0002
88–108	330	50 AMP	395	180–195	40–105	P296-0004
88–108	330	50 AMP	336	162–175	55–115	P296-0009
108–130	330	50 AMP	420	195–224	70–140	P296-C010
135–155	320	50 AMP	420	195–224	60–123	P296-0003
145–175	330	50 AMP	395	180–195	40–105	P296-0008
189–227	330	50 AMP	420	262–290	60–121	P296-0011

ular brands of compressors, such as Copeland, Tecumseh, Bristol, and Carlyle. See Fig. 19-8.

The kit consists of the end cap (a), bracket (b), capacitor (c), start relay (d), and an assortment of wire (e) of four different lengths and three colors.

Table 19-5 shows the specifics of each part with the capacitor voltage rating and microfarad range. Note the dropout volts for the start relay.

Single-Phase Line Monitors

The single-phase line monitor is a very low cost, highly accurate, rugged module that was designed to protect single-phase devices from over- or undervoltage, rapid

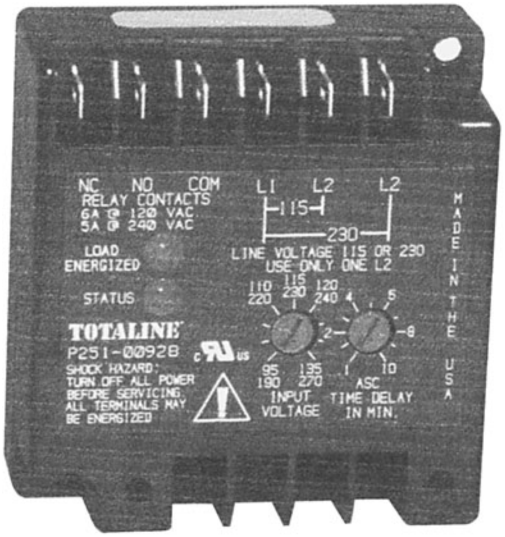
NC NO COM
RELAY CONTACTS
6A @ 120 VAC
5A @ 240 VAC
L1 L2 L2
115
230
LINE VOLTAGE 115 OR 230
USE ONLY ONE L2
LOAD ENERGIZED
STATUS
TOTALINE
P251-0092B
SHOCK HAZARD:
TURN OFF ALL POWER
BEFORE SERVICING.
ALL TERMINALS MAY
BE ENERGIZED
110 220
115 230
120 240
95 190
135 270
INPUT VOLTAGE
4 5 2 8 1 10
ASC
TIME DELAY IN MIN.
MADE IN THE USA

Figure 19-9 Single-phase line monitor. (Courtesy of Carrier)

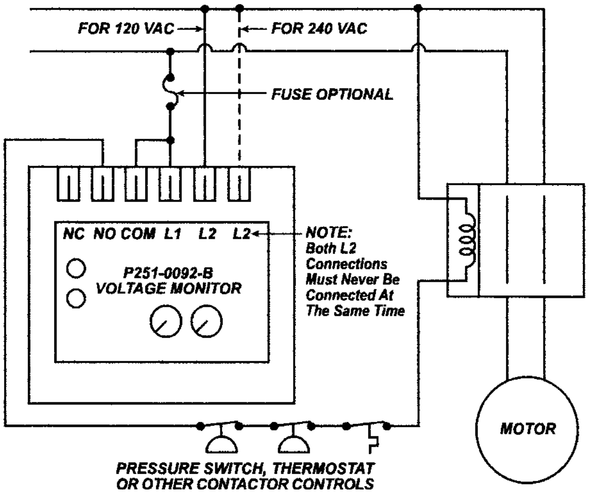
FOR 120 VAC
FOR 240 VAC
FUSE OPTIONAL
NC NO COM L1 L2 L2
P251-0092-B
VOLTAGE MONITOR
NOTE:
Both L2
Connections
Must Never Be
Connected At
The Same Time
MOTOR
PRESSURE SWITCH, THERMOSTAT
OR OTHER CONTACTOR CONTROLS

Figure 19-10 Wiring diagram for single-phase line monitor. (Courtesy of Carrier)

short-cycling, and power interruptions (see Fig. 19-9). Its small design is easy to mount to continuously monitor incoming line voltage for errors. When line power is appropriate, the module closes a set of normally open contacts and lights a green LED. When the incoming power is outside of the user-selected parameters, the normally closed contacts will close and a red LED will illuminate, indicating current

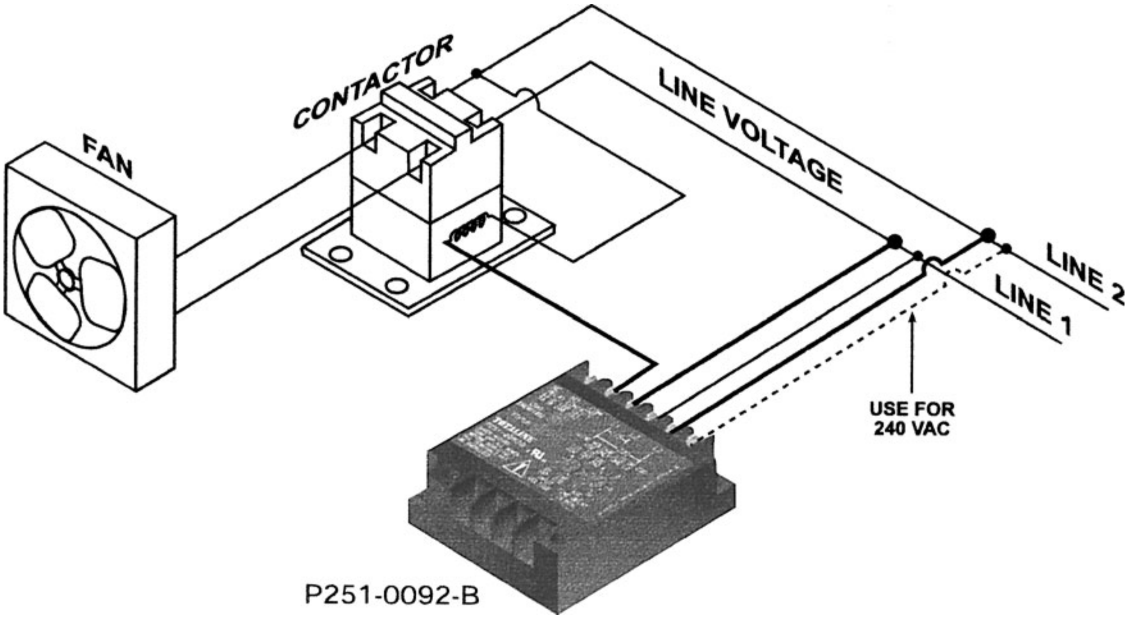
CONTACTOR
FAN
LINE VOLTAGE
LINE 2
LINE 1
USE FOR
240 VAC
P251-0092-B

Figure 19-11 System diagram for single-phase line monitor. (Courtesy of Carrier)

fault conditions. This unit also interrogates the line during fault conditions to reduce nuisance trips from transients or compressor start-ups. The unit trips if power is abnormal for 66% of interrogation time. The time delay on make is 0–10 minutes. The green LED shows that the power is on and valid and the relay is energized. The red LED flash rapidly when the unit currently detects high/low voltage situation. Red LED blinking shows the power is current, valid, and wait-

TABLE 19-6 Competitive Cross-Reference Guide

	P251-0083C model Totaline	800/800A series Hoffman	E31 series Ranco	P66 Johnson controls	Totaline advantage
Hard start	Field adjustable 1–5 seconds	Factory fixed at 5 seconds	No	No	Installer may adjust the "Hard start" period to satisfy different fan size and job requirements.
Low temperature cutoff	Field adjustable 30°F–70°F (min, span)	Fixed at 50°F	No	No	Installer may determine the minimum rpm level at which condenser fan should operate.
High temperature bypass	Yes	Yes	No	No	Applies full voltage to the condenser fan under normal operating conditions.
Set point hysteresis (Dead band)	Fixed at 3°F	Fixed at 3°F	No	No	Prevents system oscillation, which may occur with small temperature fluctuations in the liquid line.
24 VAC supply	Yes	No	No	Yes	A 24-VAC supply provides control from the low-voltage side via a low-voltage sensor. This eliminates costly high-voltage wiring and allows for easier and safer installation.
Surface mount sensor	Yes	Yes	Yes	No	Helps to prevent system penetration and is easier to install.
Independent adjustment	Yes	No	No	No	Variable adjustment is quick, easy and independent... simplifying field calibration to a wide variety of motors.
Multiple voltage operation	One model covers 120–600 VAC!	No Three models required	No 240 VAC only	No Two models required	Reduces inventory. One model is easier, less costly to stock, and helps reduce the chance of mis-wiring in the field.

ing for the end of a delay. The wiring diagram is shown in Fig. 19-10. Figure 19-11 shows the pictorial wiring of the unit.

The unit is epoxy encapsulated for use in extreme environmental conditions. The power loss detection or trip takes place within 45 milliseconds. There is a user selectable anti-short cycle (ASC) delay of 0.0 to 10 minutes. This unit also trips if the power is abnormal for 55% of interrogation period. The relay contacts can handle 6 amperes. The proper unit requires only 21 to 31 milliamperes for power operation. Table 19-6 presents a competitive cross-reference guide to various manufacturers' units.

Time Delays

Time delays also come in epoxy-encapsulated units. They will meet requirements for numerous applications. They are a simple two wire hookup usually and will work with the anticipator-type thermostat. One of those is shown in Fig. 19-12. It is specifically designed to control the circulating fan in heat pumps, air con-

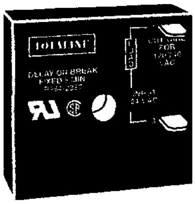

2387

- Simple 2 wire hookup
- Works with anticipator-type thermostat
- 24 VAC

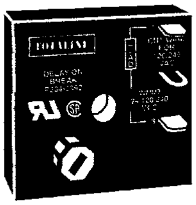

2392

- Simple 2 wire hookup
- Works with anticipator-type thermostat
- Provides jumper wire for 120/240 VAC operation
- 24/120/240 VAC

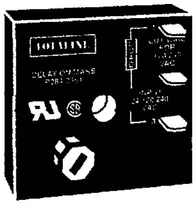

2391

- Simple 2 wire hookup
- Works with anticipator-type thermostat
- Provides jumper wire for 120/240 VAC operation
- 24/120/240 VAC

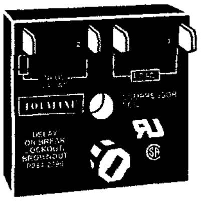

2390

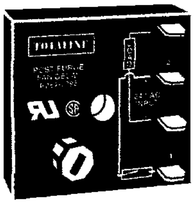

2395

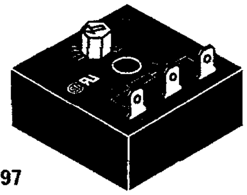

2397

PART NO.	FIXED	ADJUSTABLE	TYPE OF DELAY	USAGE
P284-2387	5 minute	-	On break	Delay period starts at end of cycle
P284-2392	-	0 - 10 minutes	On break	Delay period starts at end of cycle
P284-2390	-	0 - 10 minutes	On break	Delay period starts at end of cycle plus brownout protection
P284-2395	-	0 - 10 minutes	Post purge	Fan remains on at end of cycle
P284-2391	-	0 - 10 minutes	On make	Delay period starts at beginning of cycle
P284-2397	-	0 - 10 minutes	Bypass timer	Delay period starts at beginning of cycle

Figure 19-12 Time delay modules. (Courtesy of Carrier)

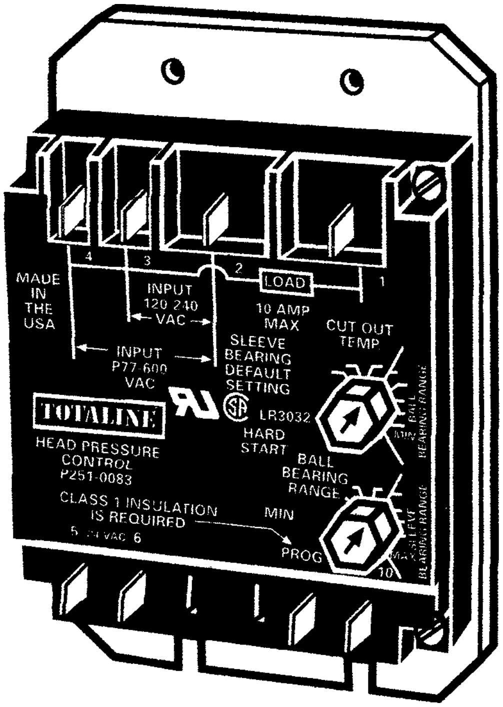
4
3
2
1
LOAD
MADE IN THE USA
INPUT 120 240 VAC
10 AMP MAX
CUT OUT TEMP
INPUT P77 600 VAC
SLEEVE BEARING DEFAULT SETTING
TOTALINE
LR3032
HARD START
HEAD PRESSURE CONTROL P251-0083
BALL BEARING RANGE
CLASS 1 INSULATION IS REQUIRED
MIN
PROG
5
6
10
BALL BEARING RANGE
SLEEVE BEARING RANGE

Figure 19-13 Solid-state head pressure control module. (Courtesy of Carrier)

ditioners, and forced air systems. The table contained in the figure indicates the specifics of the units and their variety.

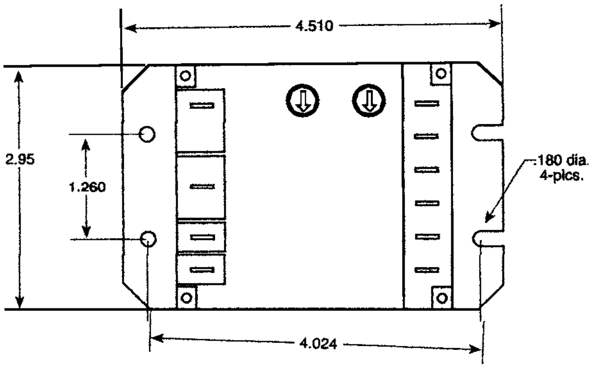
4.510
2.95
1.260
.180 dia. 4-plcs.
4.024

Figure 19-14 Dimensions for head pressure control module. (Courtesy of Carrier)

Pressure Controls

Head pressure controls

The solid-state, epoxy-encapsulated head pressure control was designed specifically for HVAC and refrigeration applications. The unit is a temperature-sensitive fan motor speed control that regulates head pressure at low ambients by varying the air volume through the condenser. Its unique construction permits exceptionally quick response times to change in liquid line temperatures. See Fig. 19-13. The dimensions of the unit are shown in Fig. 19-14. It provides full torque to the motor during start-up to ensure proper fan rotation and lubrication of bearings. It also features high temperature bypass and applies full voltage to the condenser fan under normal operating conditions. It also determines the minimum RPM level at which the condenser fan should operate. The set-point hysteresis (Dead-band) prevents system oscillation which may occur with small temperature fluctuations in the liquid line. This unit can be used with a wide variety of motors.

Table 19-6 shows the head pressure control specifics and compares the Carrier or Totaline with the Hoffmann, Ranco, and Johnson Controls models of the same unit.

Line voltage head pressure controls

Preventing evaporator freeze-ups, liquid slugged compressors, and low pressure cutouts can occur during low ambient conditions. The dual line voltage input head pressure controls are ideal for refrigeration (see Fig. 19-15).

The unit controls temperature-sensitive motor fan speed. They have been designed to regulate head pressure at low ambients by varying the amount of airflow through the condenser and help to ensure sufficient pressure across the expansion valve. This unit has a dual-voltage 240/480 input transformer as part of the package. The user is able to select and adjust set points on the unit:

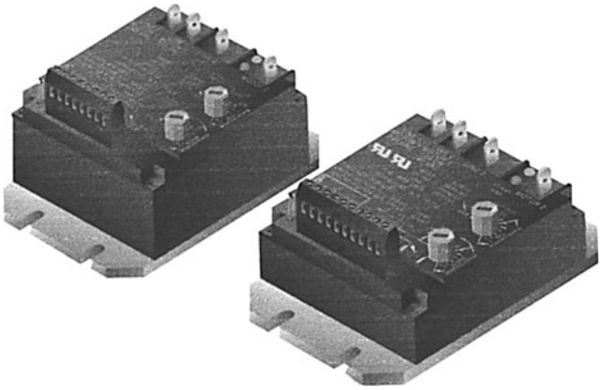

Figure 19-15 Line voltage head pressure controls. (Courtesy of Carrier)

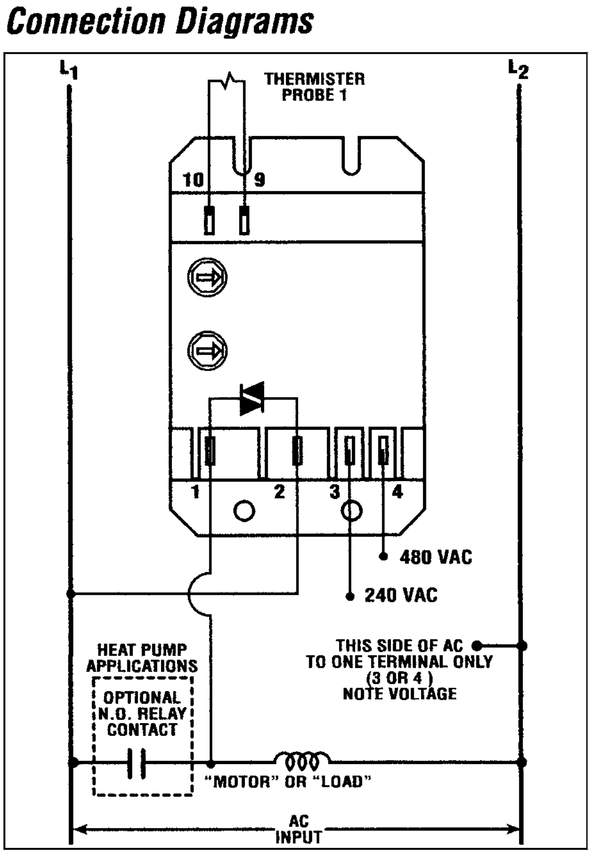
Connection Diagrams
L1
L2
THERMISTER PROBE 1
10
9
1
2
3
4
480 VAC
240 VAC
HEAT PUMP APPLICATIONS
OPTIONAL N.O. RELAY CONTACT
THIS SIDE OF AC TO ONE TERMINAL ONLY (3 OR 4) NOTE VOLTAGE
"MOTOR" OR "LOAD"
AC INPUT

Figure 19-16 Connection diagrams for Fig. 19-14. (Courtesy of Carrier)

TABLE 19-7 Probe Resistance versus Temperature

P251-0084			P251-0085			P251-0086		
		Resistance			Resistance			Resistance
°F	°C	K Ohms	°F	°C	K Ohms	°F	°C	K Ohms
32	0	22.8	59	15	28.3	59	15	42.4
50	10	15.9	68	20	22.5	68	20	33.7
68	20	11.4	77	25	18.0	77	25	27.0
77	25	9.7	86	30	14.5	86	30	21.8
86	30	7.6	95	35	11.8	95	35	17.6
104	40	4.6	11	40	9.6	104	40	14.4
22	50	2.8	113	45	7.9	113	45	11.8
40	60	1.8	22	50	6.5	122	50	9.7
58	70	1.2	131	55	5.4	131	55	81

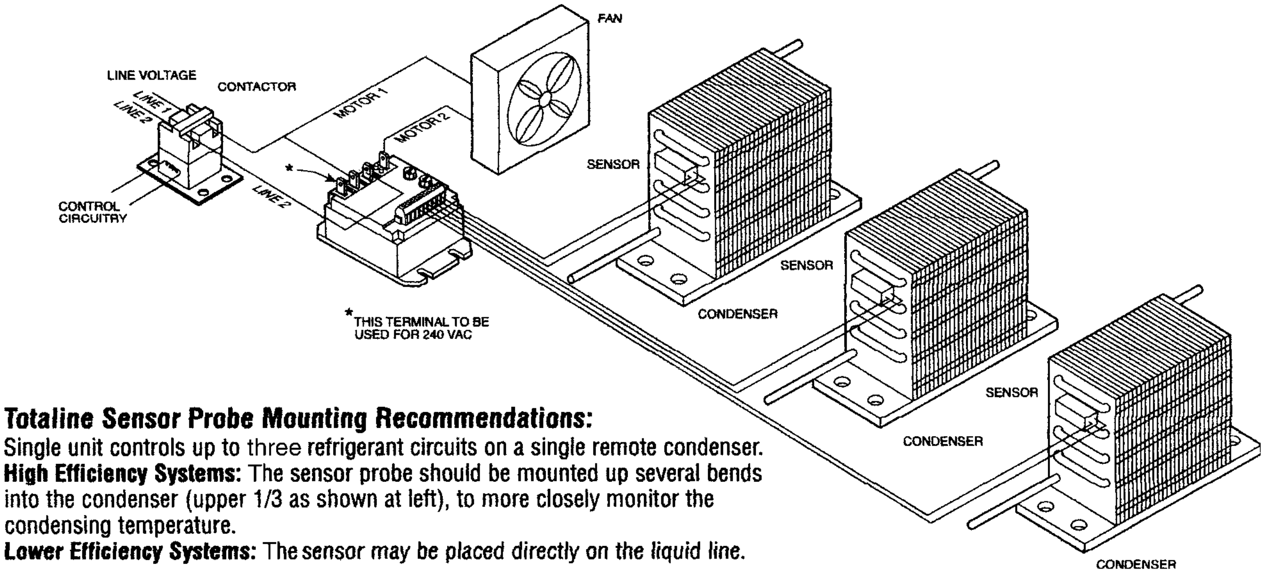
LINE VOLTAGE
LINE 1
LINE 2
CONTACTOR
MOTOR 1
MOTOR 2
FAN
CONTROL CIRCUITRY
*
LINE 2
SENSOR
SENSOR
CONDENSER
*THIS TERMINAL TO BE USED FOR 240 VAC
SENSOR
CONDENSER
CONDENSER

Totaline Sensor Probe Mounting Recommendations:

Single unit controls up to three refrigerant circuits on a single remote condenser.

High Efficiency Systems: The sensor probe should be mounted up several bends into the condenser (upper 1/3 as shown at left), to more closely monitor the condensing temperature.

Lower Efficiency Systems: The *sensor may be placed directly on the liquid line.*

Figure 19-17 Typical system diagram for pressure controls. (Courtesy of Carrier)

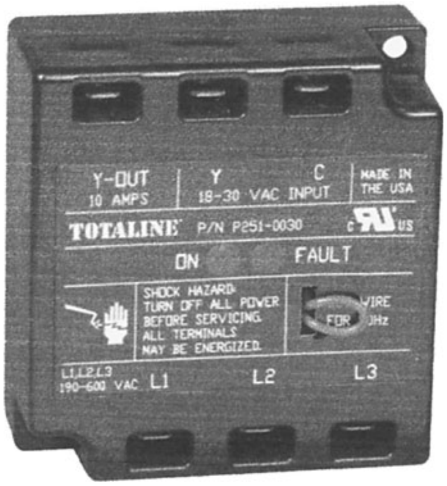

Figure 19-18 Three-phase line voltage monitor. (Courtesy of Carrier)

Hard start is 0.1 to 5 seconds adjustment possible; low temperature cutoff adjustment has adjustment for minimum pressure.

Basically, the unit prevents evaporator freeze-ups, low pressure cutouts, and liquid-slugged compressors. It provides full torque to the motor during start-up to help ensure proper fan rotation and lubrication of bearings.

It is possible to set the low temperature cutoff at the minimum rpm level at which the condenser fan should operate. And, the high temperature bypass applies full voltage to the condenser fan under normal conditions. Note the schematic drawing for connectons in Fig. 19-16. There are two models available, one operates on 120/240 volts and the other on 208/480 volts AC. Table 19-7 shows the probe resistance versus various temperatures. Figure 19-17 gives a graphic illustration of how the sensor probe is mounted. A single unit controls up to three refrigerant circuits on a single remote condenser. The sensor probe should be mounted up several bends into the condenser (upper $^1/_3$ as shown at the left) to more closely monitor condensing temperature. On lower efficiency systems, the sensor may be placed directly on the liquid line.

Three-Phase Line Voltage Monitor

The three-phase line monitor protects against phase loss, phase reversal, and phase unbalance. It is ideally suited to protect scroll and screw compressors from reverse rotation (see Fig. 19-18). A bright LED indicates faults and loads energized. There are no adjustments needed and it will operate on 190 to 600 volts AC. Its response to phase loss is less than 1 second, with an automatic reset from a fault condition.

If all three phases are relatively equal and in proper sequence, the normally open contacts (Y-Y-OUT) will close when 24 volts is applied betweeen C and Y. If the phases are out of sequence or one is missing, the contacts will never close. If a phase is lost while the unit is energized, the contacts will open immediately, and remain open until the error is corrected.

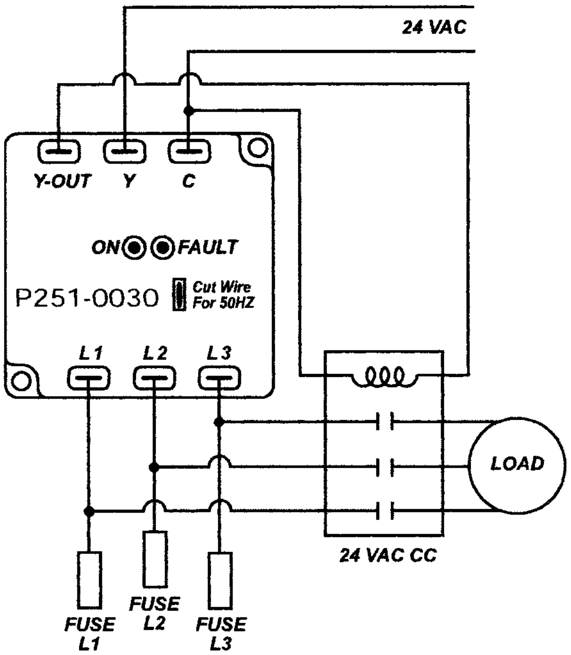
24 VAC
Y-OUT
Y
C
ON
FAULT
P251-0030
Cut Wire For 50HZ
L1
L2
L3
LOAD
24 VAC CC
FUSE L2
FUSE L1
FUSE L3

Figure 19-19 Wiring diagram for three-phase line voltage monitor. (Courtesy of Carrier)

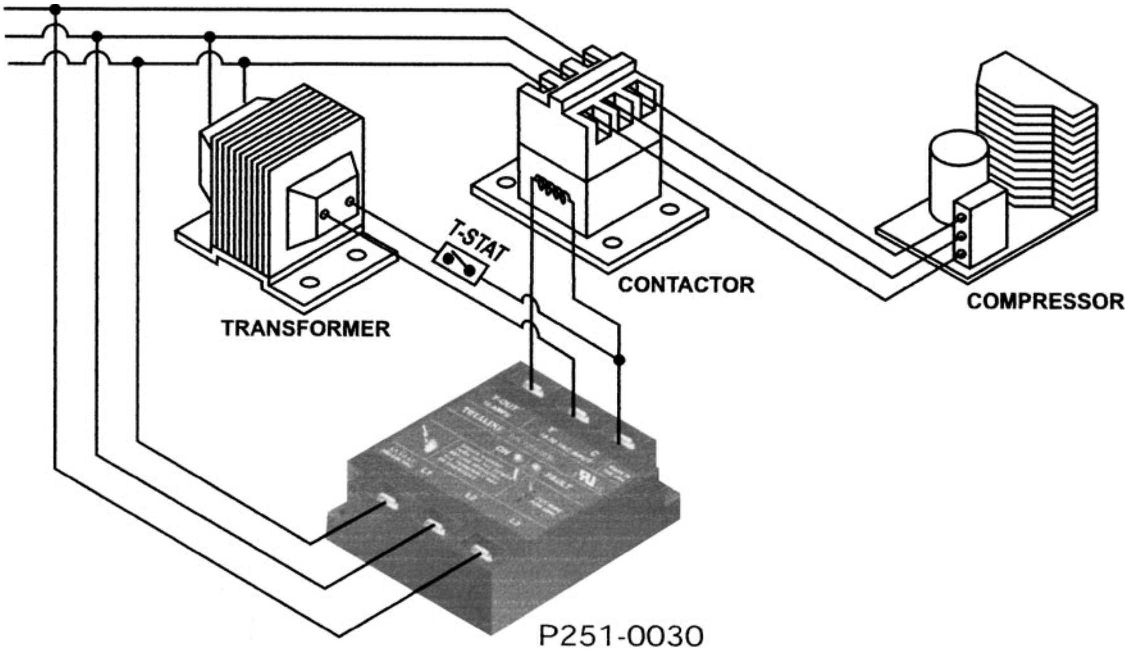
T-STAT
CONTACTOR
COMPRESSOR
TRANSFORMER
P251-0030

Figure 19-20 Typical wiring diagram for three-phase line voltage monitor. (Courtesy of Carrier)

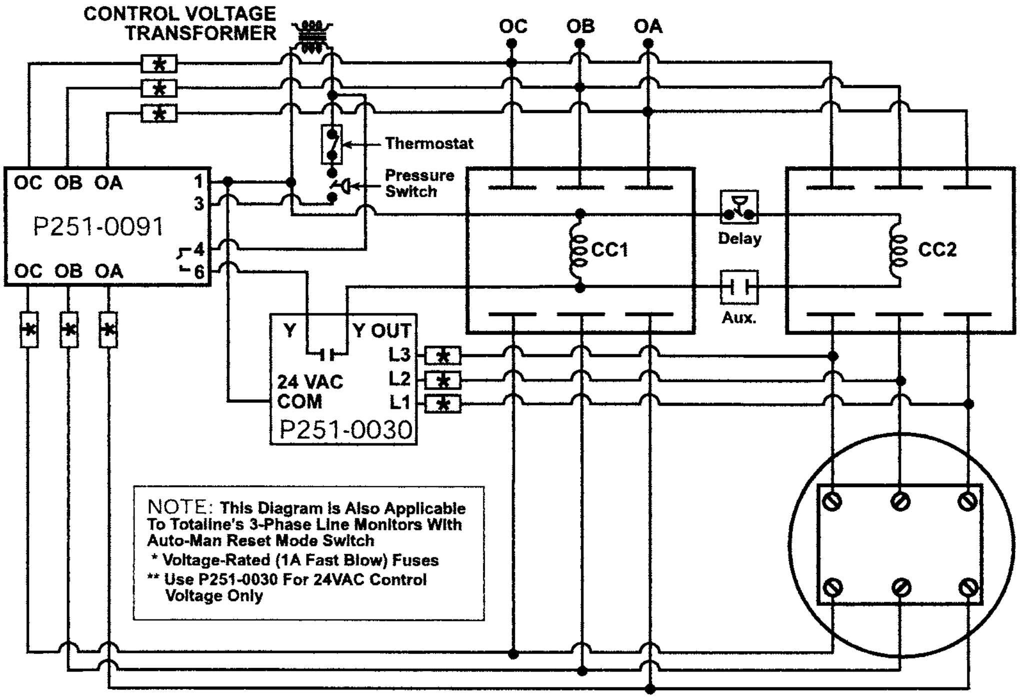
CONTROL VOLTAGE TRANSFORMER
OC
OB
OA
Thermostat
Pressure Switch
OC OB OA
P251-0091
OC OB OA
1
3
4
6
CC1
Delay
Aux.
CC2
Y
Y OUT
L3
L2
L1
24 VAC
COM
P251-0030

NOTE: **This Diagram Is Also Applicable To Totaline's 3-Phase Line Monitors With Auto-Man Reset Mode Switch**
*** Voltage-Rated (1A Fast Blow) Fuses**
**** Use P251-0030 For 24VAC Control Voltage Only**

PWS Motor With 3-Phase And Single Side Monitor

Figure 19-21 Typical part-winding start wiring diagram. (Courtesy of Carrier)

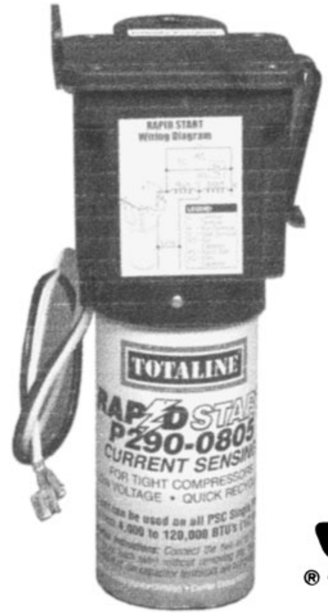
RAPID START Wiring Diagram
TOTALINE
RAPID START
P290-0805
CURRENT SENSING
FOR TIGHT COMPRESSORS
LOW VOLTAGE • QUICK RECYCLE

Figure 19-22 Current sensor. (Courtesy of Carrier)

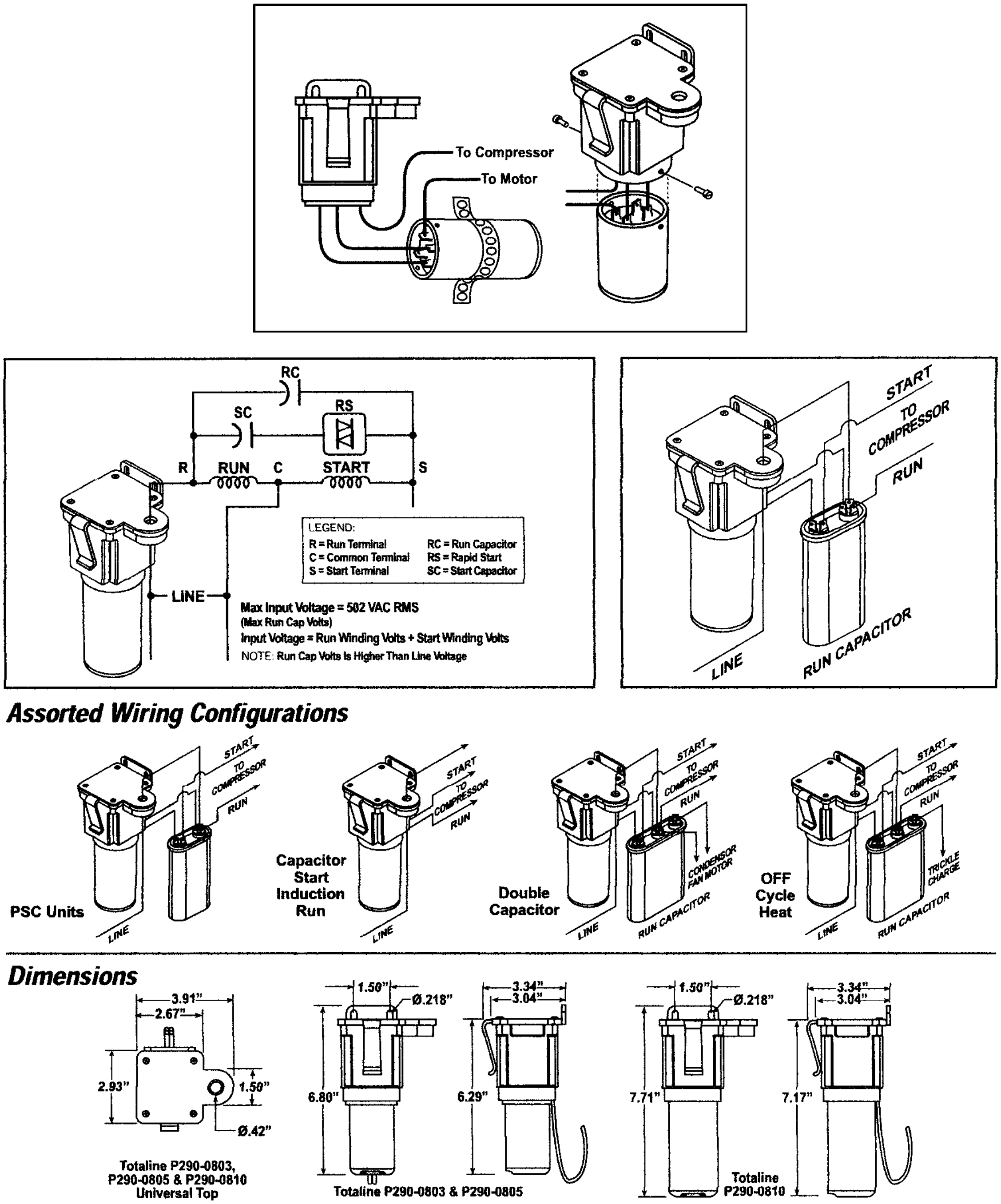
To Compressor
To Motor
RC
SC
RS
R
RUN
C
START
S
LEGEND:
R = Run Terminal
C = Common Terminal
S = Start Terminal
RC = Run Capacitor
RS = Rapid Start
SC = Start Capacitor
LINE
Max Input Voltage = 502 VAC RMS
(Max Run Cap Volts)
Input Voltage = Run Winding Volts + Start Winding Volts
NOTE: Run Cap Volts Is Higher Than Line Voltage
START
TO COMPRESSOR
RUN
LINE
RUN CAPACITOR
Assorted Wiring Configurations
PSC Units
Capacitor Start Induction Run
Double Capacitor
OFF Cycle Heat
CONDENSOR FAN MOTOR
TRICKLE CHARGE
Dimensions
3.91"
2.67"
2.93"
1.50"
Ø.42"
Totaline P290-0803, P290-0805 & P290-0810 Universal Top
1.50"
Ø.218"
6.80"
3.34"
3.04"
6.29"
Totaline P290-0803 & P290-0805
7.71"
7.17"
Totaline P290-0810

Figure 19-23 Wiring diagrams for current sensors along with dimensions. (Courtesy of Carrier)

TABLE 19-8 Rapid Start Comparisons (Carrier)

	Hard start				Soft start
	Differential current relay	Potential		PTCR devices	Timing devices
Totaline Rapid Start comparison	Totaline Rapid Start	Kickstart	Conventional three wire relay capacitor kit	Gemline HS600 & HS650 Mars 32701 & 32702 Robertson 600-052 & 600-057 SUPCO SPP5, SPP6 SPP7 WATSCO WSX-5, WSX-6	SUPCO SPP8 WATSCO WSX-1
Self adjusting	Yes	No	No	No	No
Uses current differential technology	Yes	No	No	No	No
Uses potential motor start relay	No	Yes	Yes	No	No
Two wires, nonpolarized	Yes	Yes	No	No	Yes
Recycles instantly	Yes	Yes	Yes	No	No
Senses whether motor started or not	Yes	Yes	Yes	No	No
Replaces three wire relay and capacitor kit	Yes	Yes	N/A	No	No
UL and CSA recognized	Pending	Yes	No	No	No
Approved by compressor manufacturers	Yes	Yes	Yes	No	No
Approved by equipment manufacturers	Yes	Yes	Yes	No	No
Used by OEM manufacturers	Yes	Yes	Yes	No	No
Safety cutoff	Yes	No	No	No	No
True power factor starting	Yes	No	No	No	No
Factory calibration	Not required	Yes	Yes	Yes	Yes
Voltage sensitive	No	Yes	Yes	Yes	Yes
PTCR device	No	No	No	Yes	No
Timing circuit device	No	No	No	Yes	Yes
Affected by ambient temperature	No	No	No	No	Yes
Stays in circuit too long at start-up	No	No	No	Yes	Yes

An illuminated green LED indicates that the output is energized and a red LED indicates that there is a fault and the output is de-energized (see Fig. 19-19).

Figure 19-20 shows a graphic presentation of a typical wiring diagram, while Fig. 19-21 shows typical part-winding start wiring.

Current Sensing

The Rapid Start current sensing unit is able to engage the hard-start capacitor for precisely the correct amount of time. This is done by monitoring the compressor current upon start-up. It ensures maximum starting torque without the risk of supplying too much current into the start-winding.

The unit is a timed safety circuit that operates in the event the motor fails to start within 600 milliseconds or 0.6 seconds (see Fig. 19-22).

This easy-to-install two-wire unit is solid-state circuitry. It boosts starting torque and disengages upon start. It can fit into tight locations. Various wiring configurations and dimensions are shown in Fig. 19-23. Table 19-8 shows some Rapid Start comparisons to illustrate further the operation and properties of the current-sensing device.

Review Questions

1. What are the two criteria for choosing wire size for installation in air-conditioning and refrigeration equipment?
2. What are some of the results of low voltage to a refrigerating system?
3. What is the most common cause of compressor low voltage?
4. What is synchronous speed?
5. How do you choose the proper size of wire for a job?
6. What does the abbreviation LRA mean in reference to motors?
7. How much torque is available from a motor if the voltage drops to 90% of its rated value?
8. What is the minimum number of wires needed to wire a three-phase compressor?
9. What causes heat generation in wire conductors?
10. What is the purpose of a fuse in a circuit?
11. What is the difference between a cartridge fuse and a plug fuse?

Chapter

20

Careers in Air Conditioning and Refrigeration

Performance Objectives

- Know the industries that employ air-conditioning and refrigeration mechanics.
- Know what job qualifications are needed for work in the air-conditioning/refrigeration field.
- Know the various sources of information for those in the field.
- Know the opportunities for teaching in the field.

The field of air conditioning and refrigeration offers a variety of career opportunities. Air-conditioning and refrigeration mechanics install and service air-conditioning and refrigeration equipment. There are a number of different types of equipment that require service. Some systems are complex, but are easily broken down into smaller units for repair purposes. Some mechanics specialize in a particular part of a system.

Industries that Employ Air-Conditioning and Refrigeration Mechanics

Approximately 259,000 persons work as air-conditioning and refrigeration mechanics. Cooling contractors employ most of these mechanics. Food chains, school systems, manufacturers, and other organizations use the services of air-conditioning and refrigeration mechanics. Large air-conditioning systems use many mechanics to keep the equipment operational. However, not all mechanics work for other people. About one in every seven mechanics is self-employed.

Manufacturers use refrigeration equipment for a variety of processes. Meat packers and chemical makers use refrigeration in some form. Temperature control is very important for many manufacturing processes.

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

Mechanics work in homes, office buildings, and factories. They work anywhere there is air conditioning or refrigeration to be installed or serviced. They bring to the job sites the tools and parts they need. During the repair season, mechanics may do considerable driving. Radio communications and cell phones may be used to dispatch them to the jobs. If major repairs are needed, mechanics will transport parts or inoperative machinery to a repair shop.

Mechanics work in buildings that are often uncomfortable. This is because the air-conditioning or refrigeration system has failed. The mechanic may have to work in a cramped position in an attic, a basement, or a crawl space.

Rooftop units are a common practice in keeping smaller installations cool. Many of the systems have at least one unit on the roof of the building. Cooling towers are usually mounted on top of a building. Thus, the mechanic may be called upon to work in high places. In summer, the rooftop may be very hot. This trade does require some hazardous work. For instance, there are the dangers of electric shock, torch burns, muscle strains, and other injuries from handling heavy equipment.

System installation calls for work with motors, compressors, condensing units, evaporators, and other components. These devices must be installed properly. This calls for the mechanic to be able to follow the designer's specification. In most instances, blueprints are used to indicate where and how the equipment is to be installed. The ability to read blueprints is essential for the air-conditioning and refrigeration mechanic. Such ability will help ensure that the ductwork, refrigerant lines, and electrical service are properly connected (see Fig. 20-1).

After making the connections, it is then necessary to charge the system with the proper refrigerant. Proper operation must be assured before the mechanic is through with the job.

Equipment installation is but one of the jobs the mechanic must perform. If the equipment fails, then the mechanic must diagnose the cause of the trouble and make the proper repairs. The mechanic must:

1. Find defects
2. Inspect parts
3. Be able to know if thermostats and relays are working correctly

During the winter, air-conditioning mechanics inspect parts such as relays and thermostats. They also perform required maintenance. Overhauling may be included if compressors need attention or recharging. They may also adjust the airflow ducts for the change of seasons.

Air-conditioning and refrigeration mechanics use a number of special tools. They also use more common tools such as hammers, wrenches, metal snips, electric drills, pipe cutters and benders, and acetylene torches. Air ducts and refrigeration lines require more specialized tools. Voltammeters are also part of the mechanic's toolbox. Electric circuits and refrigeration lines must be tested and checked. Testing of electrical components is also required. A good background in electricity is necessary for any mechanic (see Fig. 20-2).

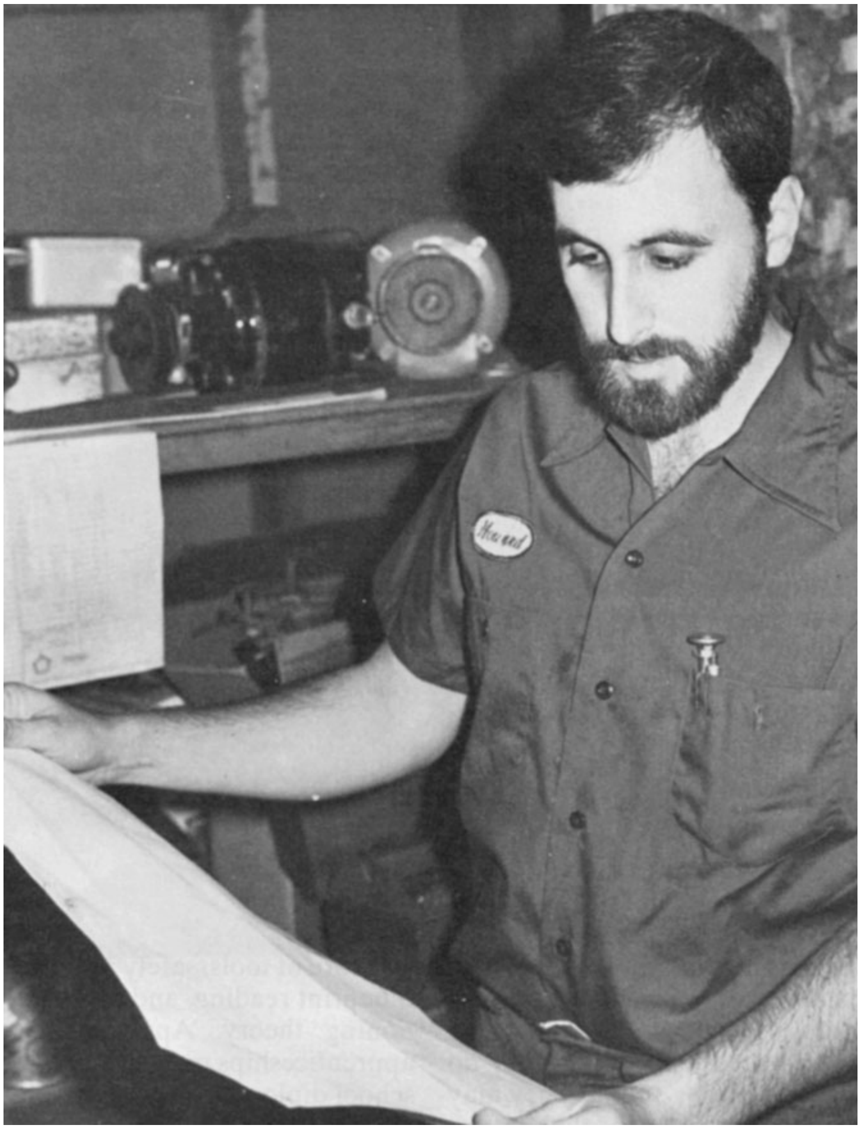

Figure 20-1 The air-conditioning and refrigeration mechanic must be familiar with blueprint symbols. Reading blueprints is essential to the proper installation of systems.

Cooling systems sometimes are installed or repaired by other craft workers. For example, on a large air-conditioning installation job, especially where people are covered by union contracts, sheet-metal specialists might do ductwork. Electricians will do electrical work. The installation of piping will be done by pipe fitters. However, in small towns or small companies, the AC man will probably have to do all these specialized skills.

Figure 20-2 The air-conditioning and refrigeration mechanic must be able to test the effectiveness of a component. This demands skill in using a variety of instruments.

Job Qualifications

Most air-conditioning and refrigeration mechanics start as helpers. They acquire their skills by working for several years with experienced mechanics. New people usually begin by assisting experienced mechanics. They do the simple jobs at first. They may carry materials or insulate refrigerant lines. In time, they do more difficult jobs, such as cutting and soldering pipes and sheet metal and checking electric circuits. In 4 or 5 years, the new mechanics are capable of making all types of repairs and installations.

The armed forces operate their own technical schools. They range from 6 months to 2 years, depending on the specialty and degree of skill needed for the military occupational specialty (MOS).

Civilian apprenticeship programs are run by unions and air-conditioning contractors. In addition to on-the-job training, apprentices receive 144 hours of classroom instruction each year. This is in related subjects. Such subjects include the use and care of tools, safety practices, blueprint reading, and air-conditioning theory. Applicants for apprenticeships must have a high school diploma. They are given a mechanical aptitude test. Apprenticeship programs last 4 to 5 years.

Many high schools, private vocational schools, and junior colleges offer programs in air conditioning and refrigeration. Students study air-conditioning and refrigeration theory and equipment design and construction. They also learn the basics of installation, maintenance, and repair. Employers may prefer to hire graduates of these programs because they require less on-the-job training.

High school graduates are preferred as helpers. If they have mechanical aptitude and have had courses in mathematics, mechanical drawing, electricity, physics, and blueprint reading, they have a better opportunity to be hired. Good physical condition is also necessary. Mechanics sometimes have to lift and move heavy equipment.

To keep up with technological change and to expand their skills, experienced mechanics may take courses, offered by a number of sources. The Refrigeration Service Engineers Society and the Air-Conditioning Contractors of America offer updating courses for mechanics. There are a number of trade magazines that help keep the mechanic up-to-date in the latest equipment and troubleshooting procedures.

Mechanics can advance to become supervisors. Some open their own contracting businesses. However, it is becoming difficult for one-person operations to operate successfully.

The Future

Employment of air-conditioning and refrigeration mechanics is expected to increase. The increase is expected to be about as fast as the average of all occupations during the past 10 years. Many openings will occur as experienced mechanics transfer to other fields of work. As experienced mechanics retire or die, they have to be replaced. The numbers will vary from year to year. However, in the United States, as people move toward the Sunbelt there will be more jobs for air-conditioning mechanics (see Fig. 20-3).

Opportunities for air-conditioning and refrigeration mechanics are expected to follow trends in residential and commercial construction. Even during periods of slow growth, many mechanics will be needed to service existing systems. Installations of new energy-saving air-conditioning systems will also create new jobs. In addition, more refrigeration equipment will be needed in the production, storage, and marketing of food and other perishables. Because these trades have attracted many people, the beginning mechanics may face competition for jobs as helpers or apprentices.

Pay and Benefits

Most mechanics work for hourly wages. These skilled air-conditioning and refrigeration specialists receive pay frequently higher than those who work in similar specialties.

Apprentices receive a percentage of the wage paid to experienced mechanics. Their percentage is about 40% at the beginning of their training. They receive 80% during their fourth year.

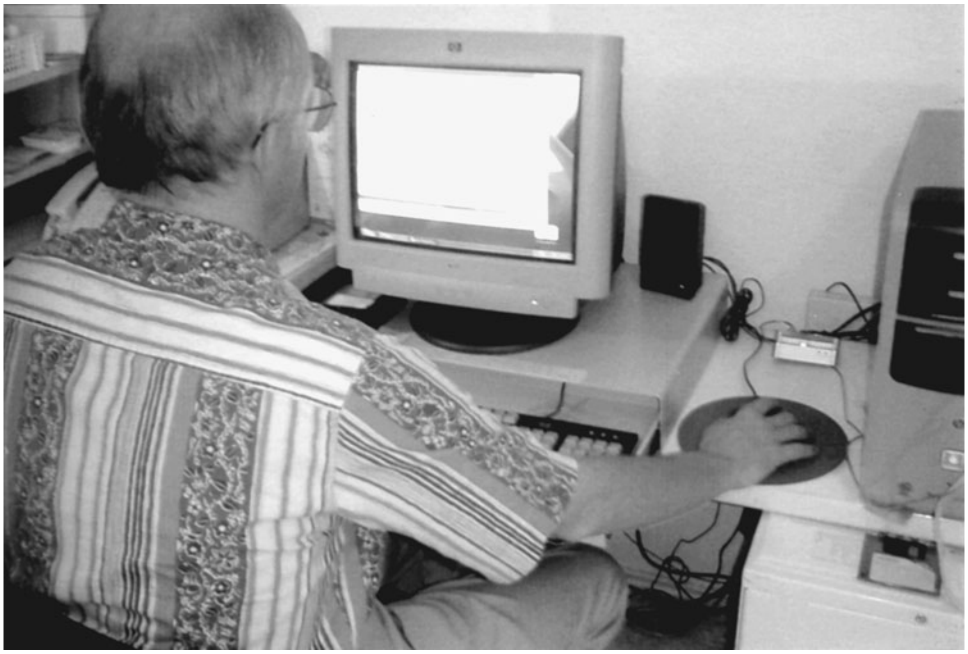

Figure 20-3 The temperature in data processing centers must be carefully controlled if the data processing equipment is to operate properly. Such equipment often generates a great deal of heat. Thus, air conditioning is essential for the equipment to function properly. This operator works in an air-conditioned room.

Mechanics usually work a 40-hour week. However, during seasonal peaks they often work overtime or irregular hours. Most employers try to provide a full workweek for the entire year. By doing this, they have mechanics when they need them most during the summer months. However, they may temporarily reduce hours or lay off some mechanics when seasonal peaks end. In most shops that service air-conditioning and refrigeration equipment, employment is largely stable throughout the year. Median hourly earnings of heating, air-conditioning, and refrigeration mechanics and installers were $16.78 in 2002. The middle 50% earned between $12.95 and $21.37 an hour. The lowest 10% earned less than $10.34, and the top 10% earned more than $26.20. Median hourly earnings in the industries employing the largest numbers of heating, air-conditioning, and refrigeration mechanics and installers in 2002 were as follows:

Merchant wholesalers	$18.78
Commercial and industrial machinery and equipment (except automotive and electronic) repair and maintenance	$17.16
Direct selling establishments	$17.14
Elementary and secondary schools	$16.80
Building equipment contactors	$16.03

Apprentices usually begin at about 40%–50% of the wage rate paid to experienced workers. As they gain experience and improve their skills, they receive periodic increases until they reach the wage rate of experienced workers. Heating, air-conditioning, and refrigeration mechanics and installers enjoy a variety of employer-sponsored benefits. In addition to typical benefits such as health insurance and pension plans, some employers pay for work-related training and provide uniforms, company vans, and tools.

About 20% of heating, air-conditioning, and refrigeration mechanics and installers are members of a union. The unions to which the greatest numbers of mechanics and installers belong are the Sheet Metal Workers International Association and the United Association of Journeymen and Apprentices of the Plumbing and Pipe-Fitting Industry of the United States and Canada.

Teaching as a Career

A person interested in passing on to others his/her knowledge of air conditioning might be to teach in vocational schools. The public schools also offer classes in air conditioning and refrigeration. Teachers of such courses often come from the trade itself. Once they have secured a position in the school, they may have to take certain college-level courses. These courses will deal with teaching methods and other subjects related to education.

Private trade schools are usually in need of good people with experience in the trade. They are needed to organize and teach apprentices. These schools may be sponsored by air-conditioning and refrigeration contractors or by unions (see Fig. 20-4).

Pay and benefits are the same as for any other teacher in the public or the private schools. Working conditionings are similar throughout the country. The demand varies with the temperature. Therefore, the climate has much to do with the demand for air-conditioning and refrigeration specialists.

Sources of Additional Information

For more information about opportunities for training, certification, and employment in this trade, contact local vocational and technical schools; local heating, air-conditioning, and refrigeration contractors; a local of the unions or organizations previously mentioned; a local joint union-management apprenticeship committee; or the nearest office of the State employment service or apprenticeship agency.

For information on career opportunities, training, and technician certification, contact:

Air-Conditioning Contractors of America (ACCA), 2800 Shirlington Rd., Suite 300, Arlington, VA 20206. Internet: http://www.acca.org

Refrigeration Service Engineers Society (RSES), 1666 Rand Rd., Des Plaines, IL 60016-3552. Internet: http://www.rses.org

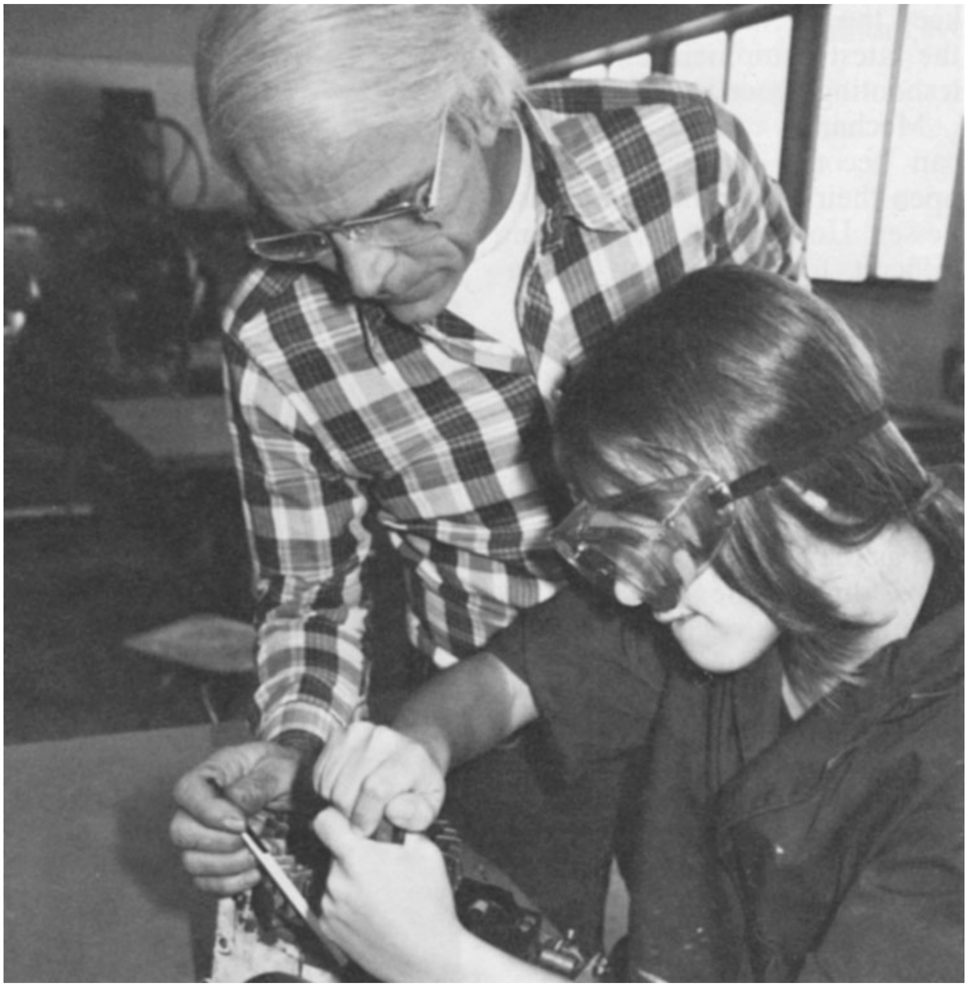

Figure 20-4 The field of air conditioning and refrigeration offers many opportunities to qualified teachers. Such teachers can work in private trade schools, as well as public schools.

Plumbing-Heating-Cooling Contractors (PHCC), 180 S. Washington St., P.O. Box 6808, Falls Church, VA 20046. Internet: http://www.phccweb.org

Sheet Metal and Air-Conditioning Contractors' National Association, 4201 Lafayette Center Dr., Chantilly, VA 20151-1209. Internet: http://www.smacna.org

For information on technician testing and certification, contact:

North American Technician Excellence (NATE), 4100 North Fairfax Dr., Suite 210, Arlington, VA 20203. Internet: http://www.natex.org

For information on career opportunities and training, contact:

Associated Builders and Contractors, Workforce Development Department, 4250 North Fairfax Dr., 9th Floor, Arlington, VA 20203.

Home Builders Institute, 1201 15th St. NW., 6th Floor, Washington, DC 20005-2800. Internet: http://www.hbi.org

Mechanical Contractors Association of America, 1385 Piccard Dr., Rockville, MD 20850-4329. Internet: http://www.mcaa.org

Air-Conditioning and Refrigeration Institute, 4100 North Fairfax Dr., Suite 200, Arlington, VA 20203. Internet: http://www.coolcareers.org or http://www.ari.org

There are more than 500 occupations registered by the U.S. Department of Labor's National Apprenticeship System. For more information on the Labor Department's registered apprenticeship system and links to State apprenticeship programs, check their Web site: http://www.doleta.gov.

Review Questions

1. How many hours of classroom instruction are received in an air-conditioning and refrigeration apprenticeship program?
2. What business employs most air-conditioning and refrigeration mechanics?
3. What percentage of the wage paid to experienced workers are paid to apprentices?
4. What does the future promise for the air-conditioning and refrigeration mechanics?
5. How many people work in this industry today?
6. What are some of the professional organizations that exist for service to the air-conditioning and refrigeration technician?

This page intentionally left blank

Index

Copyright © 2007 by The McGraw-Hill Companies, Inc. Click here for terms of use.

This page intentionally left blank

AC. *See* Alternating current
  ammeter, 22, 98, 101, 320
    symbol, 57
  generator, 21
  window unit, diagram, 290
Actuators, 191
AG fuses. *See* Automotive glass fuses
Air capacitors, 152
Air-Conditioning and Refrigeration Institute (ARI), 331
Air conditioning, 4
  unit, 279
Air core coil, 50, 130
Alternating current (AC), 21, 119
Alternator, three-phase, 125
American Wire Gage (AWG), 330
Ampere, 2, 14, 22
Amplifier stage, 191
Anti-sweat heater, 309
Applications, high temperature, 89
ARI. *See* Air-Conditioning and Refrigeration Institute
Asia-minor, 2
Atom, 5
Automatic defrost, 296
Automation, 4
Automotive glass (AG) fuses, 55
Autotransformers, 141
Average
  value, 122
AWG. *See* American Wire Gage

Bar magnet, 77
Batteries, 56
  types of, 112
Battery maintenance, 118
Bimetallic disc, 297
Bimetallic strip, 237, 241, 242
Bleeder resistor, 157, 219, 221
Breakdown voltage, 151
Bridge circuit, 186
  operation of, 188, 189
Brush-lifting, 207
  induction run motor, 207

CAB resistive element, 185
Capacitance, 162
  units, 151
Capacitive reactance, 155
Capacitor
  capacity of, 150, 325
  charging, 150
  discharging, 150
  how it works, 149
  ratings, 219
  run, 152
  start, 152
  symbols, 49
  testing, 317, 324
Capacitors, 147, 149
  start, capacitor-run motor, 215
  start motor, 208, 209
  types, 151, 153
Carbon atom, 7, 8
Carbon dioxide, 4
Carbon-composition resistors, 38
Cartridge fuse, 55
Cell
  alkaline, 119
  construction of, 112
  sizes, 116
  symbol, 60
Cell and battery symbols, 57
Cells
  connecting, 115
  in parallel, 63, 117
  in series, 60, 117
  nickel-cadmium, 119

cemf. *See* Counter emf
Centrifugal switch, 323, 324
Ceramic capacitors, 153
Characteristics, parallel circuit, 63
Charged bodies, 12
Charges
  attract, 11
  negative, 10
  positive, 10
  repel, 11
Charging meter, 108
  electronic, 107
Checking capacitors, 157
Chemical action, 16, 19
Chemical reaction, 14
Chinese, 2
Chips, 174
  semiconductor, 343
Choke, 51
Circuit, 15
  complete, 15
  connecting a, 16
  open, 15
  protection, 339
Circuit breaker, 55, 169, 170, 171, 341
Circuit protectors, types, 230
Circuits
  air-conditioning, 279
  inductive, 144
  three-phase, 338
Clamp-on ammeter, 99, 100
Clamp-on volt-ammeter, 321
Code limitations, 339
Coefficient of coupling, 136
Coil
  sucking effect, 82, 83
  of wire, magnetism, 80
Coils, 129
Cold anticipator, 246
Color bands, 38, 39
Color code, 38
Common base circuit, 180
Common emitter circuit, 179
Compounds, 5
Compressor, 318
  air conditioning, 312
  motor protection, 192
  motor relays, 221
Compressor (*Cont.*)
  problems, 313
  terminal board, 193
  terminals, 109
  troubleshooting, 313
Conductance, 23
Conductor, 14
  grounding, 227
  heat generated, 339
Connecting
  an ammeter, 99
  a capacitor, 156
  cells, 115
Connection
  delta, 126
  diagrams, 353
  three-phase, 124
  wye, 126
Construction, 137
Control, troubleshooting, 194
Controllers, 189
Controls
  pressure, 350
  solid state, 173–198
Cooling
  thermostats, 242, 244
  towers, 249
Copper, 13
  atom, 7
Coulomb, 12, 21
Counter emf (cemf), 132
Current, 13, 14, 62
  changes, sine wave, 131
  division, 69
  eddy, 143
  in parallel circuits, 64
  in series, 61
  relay, 223, 235
  sensing, 355
  type relays, 222
  voltage, and resistance, relationship, 28, 29
Cybertronic actuators, 191
Cycle, defrost, 299

D'Arsonval meter movement, 93
Dead band, 351
Defrost
  circuit, 297, 300

Defrost (*Cont.*)
  control, 183
    solid-state, 182
  heater, 299
  service, 90
  termination thermostat, 299
  timer, 296, 298, 301
  manual, 295
Defrosting, 296
Defrosts, automatic, 296
Delay, inductive, 146
Determining resistance, 66
Devices, safety, 225
Diagram, manufacturer's, 256, 257
Difference of potential, 22
Differential amplifiers, 190
Digital meter, 104, 105
Diodes, 52, 174–177
DIP. *See* Dual in-line package
Distributing electric power, 165, 166
Doping, 175
Dry cell, 112, 113
Dual in-line package (DIP), 181, 182

Eddy currents, 8, 143
Electric currents, physiological effects, 226
Electric heating system, 253
Electrical
  charge, 9
  safety, 225–232
Electricity, 1, 2, 4, 93
  future of, 3
  putting to work, 32
  sources, 19
Electrolytic capacitors, 48, 49, 153–158
Electromagnetism, uses, 82
Electromagnets, 75, 80, 82
Electron, 1, 4
  flow, 16
  movement, 21
Electronic controllers, 189
Electrons
  controlling of, 13
  location of, 9
  orbiting, 8
  properties of, 8
  protons, neutrons, 6
Elements, 5
EMF, counter. *See* Counter emf
Energy saver, 309
Evaporator pressure, 89
Explosion proof motor, 218

Fan, 283
  motors, 203
  switch, 291
Ferrite core choke, 51
Field wiring, 258
Filings, iron, 3
Fixed
  capacitors, 47, 48
  inductors, 50
  resistors, 43
Flat pack IC, 181
Float switch, 250
Force fields, 11
Four-wire wye system, 169
Franklin, Benjamin, 2, 3
Free electrons, 10
Freezers, upright, 304
Freeze-ups, 352
Frequency, 21
Frost-free, 296
Fuse, 62
  and circuit breakers, 53
  dual element, 342
  how it works, 341
  single element, 341
  time delay, 342
  types of, 343
Fusible resistors, 44, 46
Fusite thermal relay, 294

Gallium, 175
Gas furnace operation, 252
Generation of power, 166
GFCI. *See* Ground-fault circuit interrupter
Gold and silver bands, resistors, 40
Gray wires, 227
Greeks, 2
Ground-fault circuit interrupter (GFCI), 229
Grounded capacitor, 324
Grounded phase, 322
Growler, 323
Guide, cross reference, 349

Halogen leak detector, 106
Handling refrigerants, 312
Heat, 19, 260
  anticipator, 246
  combinations, 267
  and furnace controls, 269
  operation of, 264–267
  systems, 266
Heater,
  defrost, 299
  electric, 34
Heating
  circuits, 251–278
  and cooling wiring diagram, 260
  thermostats, 242, 244
Heat reclaiming, 91
Helium ion, 7
Hermetic compressors, troubleshooting, 288, 290
High-efficiency furnaces, 267
Horsepower, 32
Horseshoe magnet, 77
Hot air furnaces, 251
Hot gas defrost circuit, 302
Hotpoint refrigerator, 307
Humidity sensing, 184
Hydrogen
  atoms, 6
  ion, 7
Hydrolyzed, 185, 186
Hysteresis, 142, 349

Ice maker, 302, 303
Ice-box, 302
Impedance, 162, 179
Incandescent lamps, 56
Indium, 175
Inductance, 162
Inductance, changing, 130
Inductive circuits, 144
Inductive delay, 146
Inductive reactance, 129, 134
  uses of, 135
Inductors, 129–146
  symbols, 130
Inrush current, 338
Inrush voltage
  drops, 336, 339
Instruments, measuring 93–110
Insulator, 14
Integrated circuit, 180
Ionization, 11
Ions, 10
Iron core coil, 130

Joule, 31

Kilby, J.S., 180
Kilohms, 25
Kilowatt-hour, 34
  meter, 35
kVAR, 165
kW, 165

Ladder diagrams, 234, 255, 287, 308
Laminated core, 143
Lamps, 54
Leading and lagging, 132, 156
Leakage, 90
Leaky capacitor, 151
LED, 347
Light, 20
Limit controller, 247
Limit switch, 246
Line cords and plugs for AC units, 280
Line voltage
  monitor, three-phase, 355, 356
  head pressure controls, 352
Linear scale, 97
Lines of force, 11
Liquid line solenoid, 88
Load, 15
Lock box, 231
Locked rotor amperes, 336
Losses, transformer, 142
Low voltage
  operation, 315
  wiring, 258

Magnesia, 2
Magnetic
  contactors, 233
  field, 80, 81
  flux, 78
  permeability, 76
  poles, 78
  theory, 76
Magnetism, 2, 20, 73

Magnetite, 2
Magnets, shapes, 77
Main disconnect, 227
Main switches, 226
Maintenance, battery, 118
Manual defrost, 295
Mass, 4
Matter, 4
  forms of, 5
Measuring X_L, 134
Mechanical energy, 32
Megger, 104, 105
Megohms, 25
Mercury cell, 116
Mercury contacts, 243, 244
Meter
  movements, 94
  scales, 96
  adjusting, 108
Mica capacitors, 48, 152
Microamperes, 25
Milliamperes, 25
Mirrored scale, 98
Mixtures, 5
Module, 192
Molecule, 5
Monitors, 350
  single-phase, 347, 348
Motor, 33, 199
  capacitor-start, 148
  problems, 315
  protection, 235
  protector relay, 317
  protectors, 221, 236
  single-phase, 148
  start relays, 234
  voltages, unacceptable, 336
Multipin circular, 181
Multimeter, 94, 103, 104
  scale, 98
Mutual inductance, 133, 136

National Electrical Code, 331, 339, 341
Neon lamps, 56
Nonlinear
  resistors, 42
  scale, 97
Noyce, Robert, 180
NPNP, 177
N-type crystals, 178

Objectives, performance, 1, 19
Oersted, Hans Christian, 79
Ohm, 2, 14, 22
  George, 3
Ohm's law, 24
  closed circuit, 27
  examples, 25
  other forms, 29
  uses for, 30
Ohmmeter
  scale, 103
Oil-filled capacitors, 153
Open
  capacitor, 324
  phase, 322
Opens, 322
Orbital electrons, 6, 7
Outer shell, 10
Overload, 339
  protector, 319
Oxygen atom, 6

Packages, transistor, 181
Paper capacitors, 48, 153
Parallax error, 97
Parallel circuit, 63
  rules, 65
Peak value, 122
Permanent magnets, 74
Permanent split-capacitor (PSC)
  compressor, 312
  motor, 210
Permeability, 76
Phase, 122
Photoemission, 20
PN junction, 175
PNP transistor, 180
PNPN, 177
Polyphase, 166
  alternating current, 123
  system, 166
Potential
  difference, 13
  relay, 222, 224, 235
Potentiometer, 42
Powdered iron-core inductors, 51

Power, 31, 134, 162
  factor, 164
  waveform, 163
Precautions, safety, 225
Prefixes, common, 23
Pressure, 20
Pressure control
  head, 350
  head, solid state, 351
  switches, 247, 248
Pressure controls, 350
  mounting recommendations, 354
Primary cell, 112
Principles, semiconductor, 174
Printed circuit board control center, 260, 261
Probe resistance, 353
Procedure, troubleshooting, 326
Programmable
  controllers, 192
  thermostats, 56
PSC compressor. *See* Permanent split-capacitor, compressor
  hookup, 213, 214
PTCR device, 359
P-type crystals, 178
Pulse
  combustion process, 273
  furnace, 251, 268
  sequence of operation, 270–273
  troubleshooting, 274–277
Purely resistive, 164
Push-button switches, 282, 283
PWS motor, 357

Range of ammeter, 99
Rapid start comparisons, 359
Reactance, capacitive, 147
Reactive power, 165
Reading schematics, 57
Rechargeable cells, 114
Rectifier circuit, 176
RED, 306–309
Refrigerants, handling, 312
Refrigeration, 4
Refrigerator,
  basic, 293, 294
  circuits, 293
  freezer, schematic, 300
Relay, 53, 54, 73, 82, 85, 86
  applications, 87
  in circuit, 87
  contacts, 236
  symbols, 234
  motor start, 234
  power, 86, 233
  pull in, 338
  switch contacts, 234
  symbols, 234
  time delay, 237, 238
Repair or replace, 326
Repulsion start, induction run motor, 206, 207
Resistance, 14, 37, 62, 162
  in parallel, 64
  in series, 61
Resistor
  and heat, 32
  symbols, 45
  various sizes, 37
Reverse bias, 175
Rheostat, 42
RMS. *See* Root Mean Square
Root Mean Square (RMS), 123
Rotating field, 200
Run and start windings, 324
Run capacitor, 221
  sizes, 213

Safety, 311
  electrical, 312
Schematic, 16, 43, 85, 283
  AC window unit, 281
  permanent split-capacitor compressor, 212
SCR. *See* Silicon-controlled rectifier
  types, 178
Secondary cell, 112
Self-inductance, 130, 131
Semiconductors, 52, 173
  chips, 343
Sensing, current, 355, 357
Sensors, 189, 198
  wiring diagrams, 358
Series circuit, 59, 60
  rules, 62
Series-parallel circuits, 65–72

Service, 345
  technician, 311
Shaded pole, 201
  induction motor, 200
  motor, reversibility, 201
  uses, 201
Shapes of magnets, 76
Shell core transformer, 138, 139
Shorted equipment, 228
Shorts, 322
Shunting current, 99
Siemens, 23
Sight glass, 107
Silicon-controlled rectifier (SCR), 177
Simple rectifier, 176
Sine wave, 120
  characteristics, 121
Single-phase
  starting switch, 204
  system, 166, 167
  AC, 161
  compressor, 109
  induction motor, 206
Slip, normal, 334
Solar system, 8
Solenoid, 73, 82, 238
  coil, 88
  natural gas, 240
  valve, 84, 86, 90, 238, 239
Solid core, 143
Solid state control modules, 193, 194
Solids, gases, and liquids, 4
SPDT switch, 242
Specifications, battery, 115
Speed, synchronous, 334
Split-phase
  motor, 204
  furnace motor, 205
Squirrel cage, 199
Squirrel cage rotor, 200, 323
Start capacitor in compressor circuit, 211
Start capacitors, 219
Start kits, 346
Start relay, 295
Starting current values, 336
Static electricity, 2
Substances, 4
Switch, 16, 246
  symbols, 54
Symbol
  for inductors, 51
  for NPN transistor, 53
  for PNP transistor, 53
  for transformer, 52
Synchronous speed, 334

Tapped resistors, 44
Temperature-compensating resistors, 47
Temporary magnets, 75
Test lamp, 316
Testing, motor, 316
Thales, 2
Thermal emission, 20
Thermistor sensing, 182
Thermistors, 184
Thermostat, 239, 287, 293, 343
  adjustments, 244
  bellows-type, 239, 241
  circuitry, 345
  control switch, 344
  defrost termination, 299
  location, 258
  microprocessor based, 244, 343
  modern, 242
  wiring diagram, 243
Third band, color code, 40
Three phase, 124
  AC, 161
  circuits, 338
  compressor, 109
  meter hookup, 168
  motor, 216, 217
  relationships, 167
Time
  delays, 350
  lag, 132
TO-5 case, 182
Tolerance, 39
Tools, portable, 227
Total resistance, 72
Transfer resistor, 180
Transformer, 51, 129, 135, 136
  audio frequency, 140, 141
  auto, 141
  closed-core, 138
  iron-core, 136
  losses, 142
  mutual inductance, 135
  open core, 138

Transformer (*Cont.*)
  power, 140, 141
  radio frequency, 140
  step-up, 139
  step-down, 137, 139
Transistor, 53, 175
  impedance, 179
  junction, 179
Troubles, capacitor, 158
Troubleshooting, 304
  window units, 291
Turns ratio, 139
Types of
  capacitors, 47
  fuses, 55
  inductors, 50
  resistors, 43
  solenoids, 83
  transformers, 137

Unbalanced bridge, 187
Units of measurement, 21

Valence electrons, 8, 10
Values, full load amperage, 337
Valve
  nameplates, 91
  operation, 86
Variable capacitors, 49
Variable resistors, 41, 44, 46
Vector, 162
Volt, 2, 22
Voltage drop, calculations, 332
Voltage drop, per 100 feet of copper wire, 334
Voltage drops, 331
  permissible, 332
Voltage loss, limiting, 330
Voltage losses, 335
Voltage, 13, 62
  transformer, 138
  variations, general effects of, 333
Voltages in series, 61
Volt-ammeter, 320
Voltmeter, 22, 101
  circuit, 102
  symbol, 57
  connecting, 102

Water tower controls, 248
Water, 5
Watt, 32
Wattage rating, resistors, 38
Window air conditioners, 280
Wire size, 329
  low voltage, 330
  selection, 329, 330, 335
Wire-wound resistors, 43
Wiring diagram, 262, 263
Working voltage, direct current (WVDC), 155
WVDC. *See* Working voltage, direct current

Instant Access to the Engineering Information You Need—Right Now

Download Chapters in PDF!

Visit McGraw-Hill's Digital Engineering Library at www.digitalengineeringlibrary.com to find, purchase, and download in just one-click.

- Pay-per-view access
- Chapters available in PDFs
- Pay-per-view price: $15—$50

This dynamic online repository features:

- 5,000+ chapters derived from 170+ outstanding McGraw-Hill titles, including:
  - —*Marks' Standard Handbook for Mechanical Engineers*
  - —*Perry's Chemical Engineers' Handbook*
  - —*Roark's Formulas for Stress and Strain*
- 500 major topics structured around 12 major areas of engineering
- Ultra-efficient search or topical browse

FULL COVERAGE OF MAJOR ENGINEERING DISCIPLINES:

- Aerospace Engineering
- Bioengineering
- Chemical Engineering
- Civil Engineering
- Communications Engineering
- Construction & Architectural Engineering
- Electrical Engineering
- Engineering Math & Science
- Environmental Engineering
- Industrial Engineering
- Materials Science & Engineering
- Mechanical Engineering

Subscribe Today!

Contact Us: 888-307-5984 (outside the U.S. 1-614-759-3663)
www.digitalengineeringlibrary.com

Learn more. Mc Graw Hill Do more.

ELECTRONICS & BRAIN CONTROL

We are learning more about our gray-matter computers

BY L. GEORGE LAWRENCE

IT IS TIME that we closely examined brain control, now that scientists are actively seeking to unravel the mysteries that shroud that miniature bioelectric giant known as the human brain. Elements of brain control can already be found in anti-collision radar technology involving birds. It has also been substantiated that, by pumping energy in the gigahertz range of frequencies through human heads, subjects can suddenly "hear" without using their ears.

Much remains to be done to correlate and sift out concealed facts. We still know very little about the all-important coding events that take place in the brain's neuron-synaptic complex. How can we measure and reinforce such significant human capacities as love, will, and character? Does uncontrolled electromagnetic pollution affect the brain? These and other questions remain to be answered as scientists delve deeper and deeper into the human brain with the help of electronics.

What Is the Brain? The theory that the brain is a true electronic machine has often been proposed. To a point, the theory is true. As Dr. Wilder Penfield demonstrated some years ago, the electrical stimulation of brain tissue during skull surgery triggers lucid audio-visual recalls of past events in patients. The patient "sees" and "hears" complete increments of true life experiences, all in correct sequence.

But consider Fig. 1. Studies are enormously complicated by the fact that the human brain contains an estimated 10 billion nerve cells called "neurons" and another 100 billion of a second type called "glial" cells. The fluid bath in which these cells are suspended is a vital element in their electrochemical interactions. Is it here where emotional components and memory are stored and where we have susceptance to microwave and other electromagnetic frequencies?

Such questions guide us into the subject of "synaptic" transmission. The term "synapse" is derived from the Greek phrase "to clasp' and was introduced in 1897 by a Dr. Sherrington who used it to describe the junctional region between two nerve cells. In many cases there is a gap or cleft across a synapse. (In other cases, as in fish, a synapse is a real physical joint.)

One big neuron might have on its surface as many as 10,000 points of contact (synaptic knobs) with other neurons. When the latter are stimulated, some of the millions of ribonucleicacid (RNA) molecules inside them give orders to the glial cells to manufacture new proteins. The nature and pattern of these proteins contain an imprint of something that has been perceived and apparently gives rise to a molecular "engram."

We also find electrical dipoles in synaptic mechanisms which, when oriented and arranged in a large array, apparently can produce an electric field strong enough to drive positive ions over the junction barrier of the postsynaptic membrane (in a manner similar to that in a transistor) and thereby initiate excitation or produce depolarization. So, one comes to believe that the nature of synaptic transmission is essentially electrical, be it mediated by electrical or chemical transmitters.

By inference, then, the possibility arises of human brain control by electromagnetic forces directed at it from the outside. If such radiations can be suitably coded to elicit a synchronous response in the neuron-synaptic complex, the brain will trigger motor functions which, in turn, cause man or animal to execute a programmed act. Here we have some fascinating experiments that hold great hopes for the immediate future.

In the case of hearing, for example, we have been taught that our auditory system can respond only to acoustic energy. This "fact" is far from correct. Experiments conducted by Dr. Frey and others clearly indicate that the audio sense in man—and probably in animals, too—can respond to electromagnetic energy in at least a portion of the r-f spectrum. Data shows that, at very low power levels (to preclude biological damage) in tests like that shown in Fig. 2, there are audio sensations at frequencies as low as 200 MHz and at least as high as 3 GHz. When low-level energy was directed at them, the test subjects reported "hearing" a buzzing sound. However, they found it almost impossible to match r-f sounds to a sine wave. The apparent source of the buzzing, clicking, knocking, or hissing sounds is described as being within or immediately behind their heads. This localization persists no matter how a person rotates or twists his head in the r-f field.

It was during these studies that a profoundly important discovery was made: deaf subjects often had the ability to *hear* r-f sound. The clinical criterion was that, if a given person could hear audio above 5 kHz either by bone or air conduction, then r-f sound could be heard as well. This and related work has resulted in the manufacture of r-f type hearing aids for the deaf, one of which is made by Listening, Inc., 6 Garden St., Arlington, Mass., and is known as the Neurophone Model GPF-1. It operates at 100 kHz and employs crystal control.

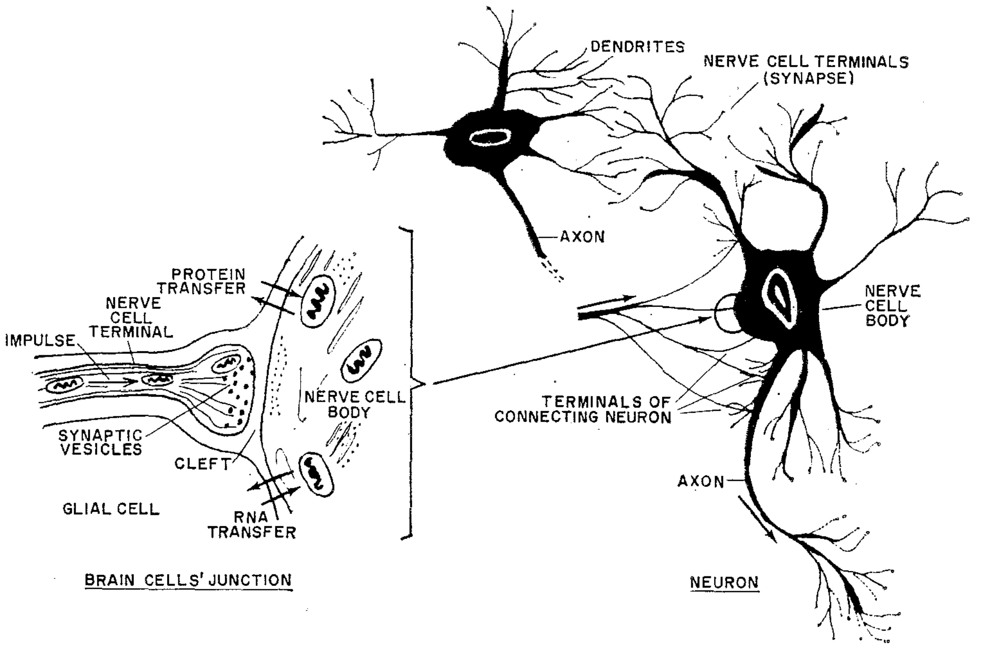

Fig. 1. Neuron/synaptic mechanism in the human nervous system.

These observations tie in with the fact that some individuals can detect radio programs through the fillings in their teeth. This phenomenon was technically verified by interposing shields between respective people who exhibited it and the modulated r-f sources. When the lower half of the head was covered, including the maxillary dental area, the r-f sound was perceived. The sound ceased upon covering the top half of the head. While the mechanism responsible for this phenomenon is only imperfectly understood, it can be assumed to be the result of direct cortical nerve fiber stimulation.

We also have another form of hearing sensations that come about when the human head is placed between two large capacitor plates that are excited by varying electrostatic potentials. "Electrophonic hearing," as it is called, apparently acts on the ear's tympanic membranes in a quasi-mechanical manner. It is useful as a new research tool in specialized psycho-physiological studies on the auditory or vibrotactile system.

Unavoidably, data of this sort begs application. Pilot studies are under way to apply effective electrodynamic brain control to animals and man himself. Consider, for example, the brain-wave proposal based on ideas put forth by the late Dr. Norbert Weiner, the acknowledged father of cybernetics.

According to Weiner, a sheet of tin suspended from the ceiling of a room and connected to a 10-Hz electrostatic generator can cause unpleasant sensations in human subjects. With a field strength of 1 or 2 volts per sq cm, the oscillating field roughly coincides with the human brain's alpha-rhythm frequency but attempts to lock it to a *fixed* frequency—that of the generator. Electronic sleep machines employ similar principles, with currents of fixed amplitude and pulse width (usually square waves) being fed by conductive face masks through the cranium and brain.

Brain Control of Birds. Radar technology is now being used to deal with the problem of birds getting in the way of fast flying aircraft. The idea is to trigger a flying bird's (or a whole flock's) brain into motor functions to initiate collision avoidance by having the bird(s) veer off the flight path of the plane. This area of research was triggered by the staggering incidence of plane/bird collisions that result in equipment damage estimated to be in the millions of dollars each year and the human lives that continue to be lost as a result of such collisions.

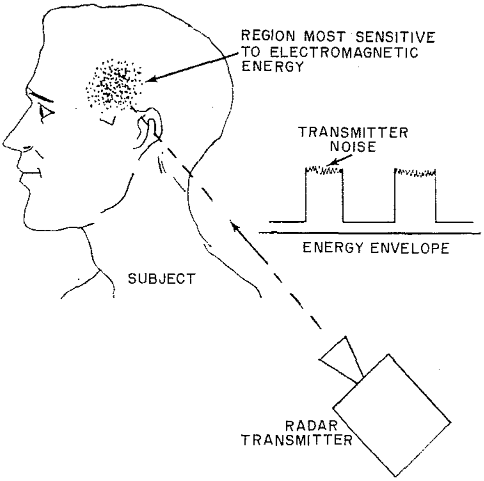

Fig. 2. Microwave susceptance area in brain.

The work being carried out by the National Research Council of Canada bears much promise in bird-brain control. The NRC's test system takes the form of a carousel of bird cages containing live chickens. (Fig. 3) Only one of the many cones suspended above the cages contains a microwave antenna. The program, headed by Dr. Alan Tanner, aims to evolve microwave brain-control systems that will have the greatest possible effect on birds while at the same time deploying the least amount of power.

When exposed to microwave radiation, birds in general exhibit escape reactions. This fact became clear during World War II. Investigators also found that in each case the microwave field through which birds were flying was of very low intensity —too low, in fact, to account for confusion and escape reactions on the basis of heat generated in the animals' bodies.

Different species of birds have different behavioral patterns. In the laboratory, a few seconds after the microwave field has been initiated, the given bird's wing outside the field of radiation became collapsed and the opposite wing became extended. Similar phenomena were observed with the legs. Sometimes the birds heeled over to the outside of the field. In the turning reaction, the outer side of the bird becomes paralyzed. In short, the microwave beam interacts with the nervous system of the test birds. Seagulls and pigeons reacted similarly, though the gulls are more inclined to initiate flight.

It is interesting to note that a bird's feathers appear to play a much greater role in the sensory complex than has been realized before. Defeathered chickens, for example, give little or no reactions to microwave fields until the twelfth day, at which time new feathers start to grow and tips protrude from the surface of the skin. In the case of fully feathered chickens whose tail feathers were exposed to microwave radiation, the birds immediately ceased exploration of their cages and exhibited mounting signs of distress after a period of 10-20 seconds. When the microwave field was switched off, the birds responded by fluffing their body feathers and active preening.

The Canadians believe that the physical properties of quill tissue—particularly the piezoelectric properties that are fundamental to living tissue–suggest mechanisms that have heretofore been overlooked. However, work carried on by Dr. Tanner and his staff as well as research conducted elsewhere should in time yield a microwave beam of the proper wavelength and modulation to cause birds to activate collision avoidance with all possible haste.

Brain-Wave Detection. Some 40-odd years ago, university professor F. Cazzamalli started publishing papers on the subject of brain-wave detection and implied that he had detected radiations from the mind. As shown in Fig. 4, he placed subjects in a shielded room (or Faraday cage), emanated vhf radiowaves through their heads, and claimed to have recorded "beat frequencies" obtained with an untuned receiver consisting of a galena crystal or diode tube, a fixed capacitor, an antenna, and a sensitive light-beam galvanometer.

The trouble is that Cazzamalli never mentioned transmitter power in his somewhat unprofessional papers. His oscillograms meant to show variations of the "beat" when his subjects were emotionally aroused or engaged in creative tasks when they were in the Faraday cage. Later, he told an astounded world that his subjects would hallucinate when under the influence of his "oscillatori telegrafica," its frequency being about 300 MHz at the time.

Tom Jaski, a noted science writer and engineer, duplicated some of Cazzamalli's work with a modern low-power oscillator that was swept from 300 MHz to 600 MHz. His subjects could not see the dial. They were told to sound off as soon as they felt something unusual. At a certain frequency range—varying between 380 MHz and 500 MHz—the subjects repeatedly indicated points with exact accuracy in as many as 14 out of 15 trials. At these "individual"

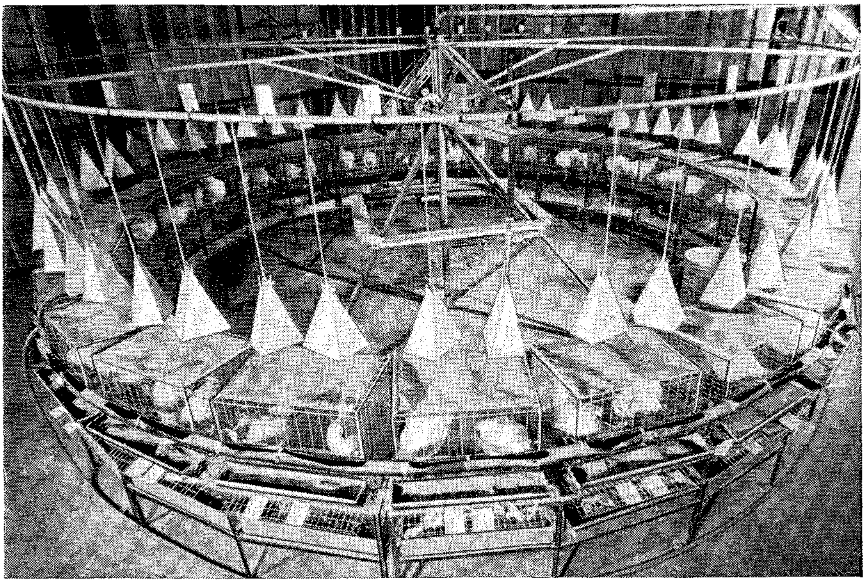

Fig. 3. Microwave test carousel to determine reactions in live birds. Only one cone is active; all others are dummies. (Courtesy National Research Council of Canada)

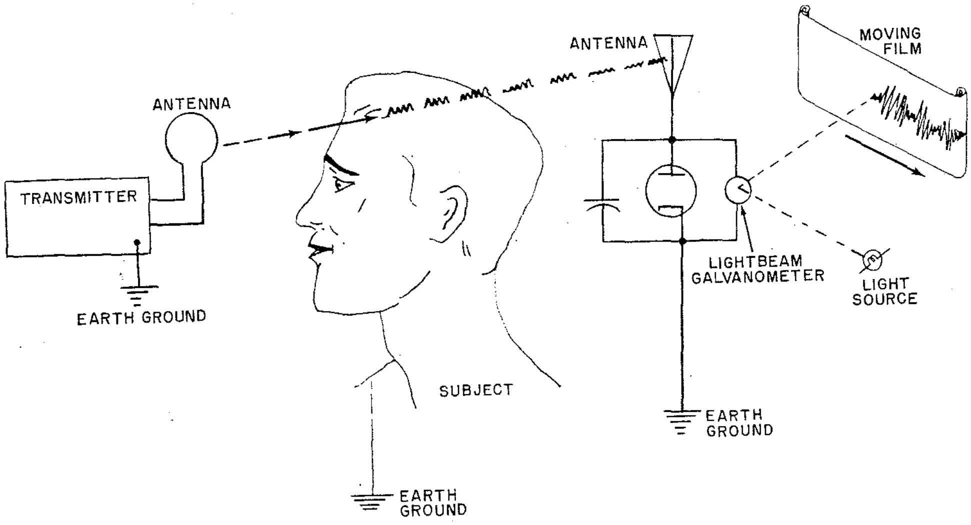
ANTENNA
MOVING FILM
ANTENNA
TRANSMITTER
EARTH GROUND
LIGHTBEAM GALVANOMETER
LIGHT SOURCE
SUBJECT
EARTH GROUND
EARTH GROUND

Fig. 4. Cazzamalli's brain-wave detector of some years ago is illustrated here.

frequencies, the same subjects announced having experienced pulsing sensations in the brain, ringing in the ears, and an odd desire to *bite* the experimenters. The oscillator's output power was only a few milliwatts, while the oscillator itself was located several feet away from the subjects.

The Conclusions. Considering the ingredients of the few sample discussions presented above, it appears that both humans and animals have brains sensitive to r-f energy. The correlating mechanisms are only imperfectly known, but they apparently reside within the neuron-synaptic complex. Nor will we know how this susceptance affects our longevity without proceeding with a great deal more research.

Just how electromagnetic radiation affects our social structure has led to a great deal of speculation in the past. For example, the late Dr. Goldman once insisted that r-f energy allows the id, or primitive brain, to take control over human affairs. Such considerations might open a Pandora's box when applied to an explanation of our sharply increasing crime rates and decline of social fidelity. So, before we rush pell-mell into electromagnetically contaminating our environment, it would serve us well once and for all to discover what adverse effects, if any, it will have on our lives. ◆

T V COMMERCIAL KILLER

FEATURES:
- No interconnecting wires—use a flashlight to key relay
- Low battery drain
- Auto reset

PARTS LIST:
1 HEP 312
1 HEP 156
1 HEP 55
1 Resistor, 1.5K, 1/2 Watt
1 Relay, Potter-Brumfield RS5D, 6VDC
1 Battery, 9VDC

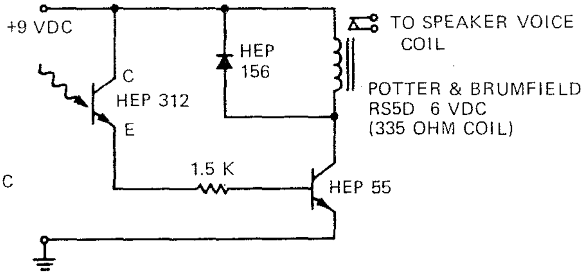
+9 VDC
C
HEP 312
E
HEP
156
TO SPEAKER VOICE
COIL
POTTER & BRUMFIELD
RS5D 6 VDC
(335 OHM COIL)
1.5 K
HEP 55

Electronic Assembly Manual

ELEC 201: INTRODUCTION TO ENGINEERING DESIGN

Department of Electrical and Computer Engineering,
Rice University, P.O. Box 1892, Houston, TX 77251
www.owlnet.rice.edu/legolab

This manual[1] introduces electronic assembly, provides step-by-step instructions for the assembly of the ELEC 201 RoboBoard, and gives wiring and other information for the electronic hardware used in the course. We assume the reader has no prior electronics background. This document is organized into the following sections:

1. Electronic Assembly Technique
2. RoboBoard Assembly Instructions
3. Battery and Charger System
4. Infrared Beacon
5. Cable and Connector Wiring
6. Sensor Wiring
7. Motor Mounting and Wiring

All of this information, except for Section 2, is also in the Course Notes on the ELEC 201 web site.

---

[1] Edited 8/02 JFY; printed August 7, 2002.

1 Electronic Assembly Technique

If there are places in life where neatness counts, electronic assembly is certainly one of them. A neatly-built and carefully soldered board will perform well for years; a sloppily- and hastily-assembled board will cause ongoing problems and failures. While building your board your team should switch tasks frequently to avoid fatigue. Mistakes are very easy to make when you are tired, and can be difficult and costly to fix.

The basics of electronic assembly are covered here: proper soldering technique, component mounting technique, and component polarities. Following these instructions and guidelines should make future debugging less frustrating.

1.1 Soldering Technique

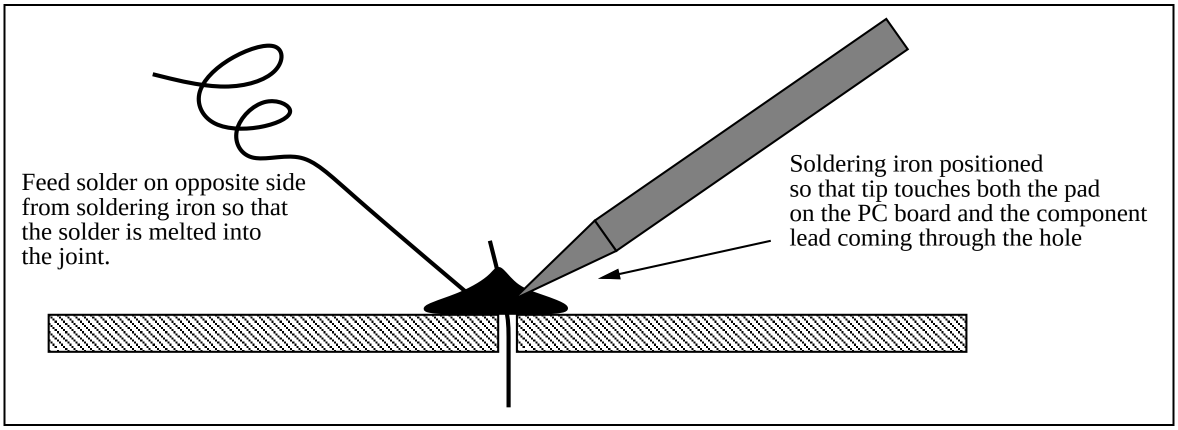
Feed solder on opposite side
from soldering iron so that
the solder is melted into
the joint.
Soldering iron positioned
so that tip touches both the pad
on the PC board and the component
lead coming through the hole

Figure 1: Proper Soldering Technique

Figure 1 shows proper soldering technique: the tip of the soldering iron is inserted into the joint such that it touches both the lead being soldered *and* the surface of the PC board. Use the side of the tip when soldering. The increased surface area will transfer heat more quickly.

Then solder is applied into the joint, *not* to the iron directly. The solder is melted *by the joint*, and both metal surfaces of the joint (the lead and the PC pad) are heated to the necessary temperature to bond chemically with the solder. Once solder flows into the joint, lift the soldering iron tip straight up, so that the finished soldered joint appears cone-shaped. This also tends to wipe excess solder along the part of the component lead that will be cut off.

Figure 2 shows what happens if the solder is “painted” onto the joint after being applied to the iron directly. The solder has “balled up,” refusing to bond with the pad (which did not receive enough heat from the iron).

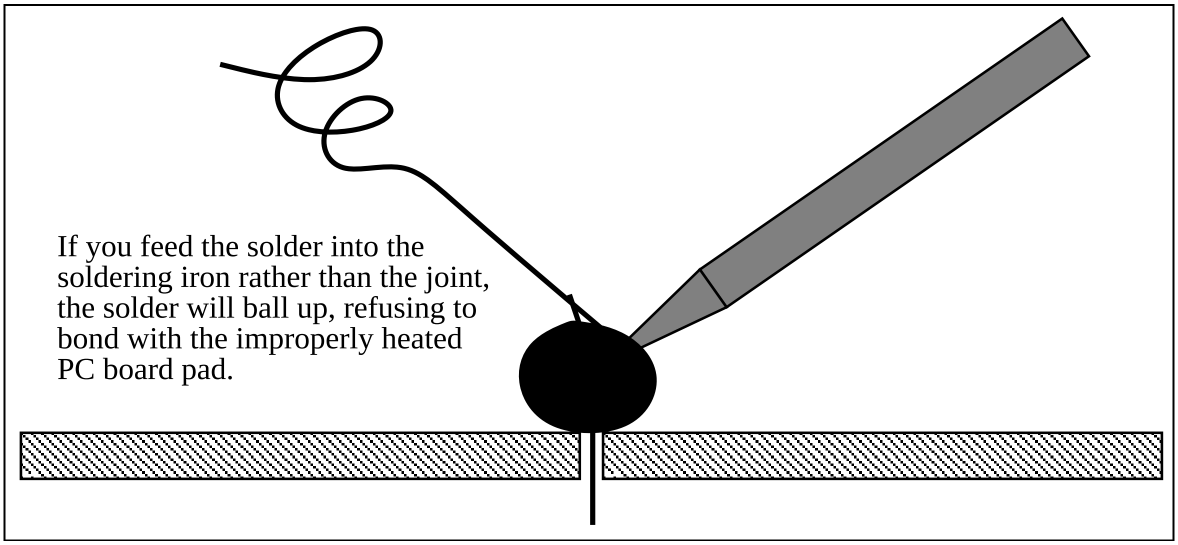
If you feed the solder into the soldering iron rather than the joint, the solder will ball up, refusing to bond with the improperly heated PC board pad.

Figure 2: Improper Soldering Technique

With this technique in mind, please read the following list of pointers about electronic assembly. All of these items are important and will help develop good assembly skills:

1. Keep the soldering iron tip away from everything (including hair) except the point to be soldered. The iron is *extremely* hot and can easily damage parts, cause burns, or even start a fire. The soldering iron should be kept in its holder when not being held.

2. For cleaning off and tinning the tip, a damp sponge should be on hand. Soldering is basically a chemical process, and even a small amount of contaminants can prevent a good joint from being made. “Damp” means wet the sponge and squeeze out all excess moisture.

3. Always make sure that the tip is *tinned* when the iron is on. Tinning protects the tip and improves heat transfer.

   To tin the iron, clean the tip by wiping it on the damp sponge and then immediately melting some fresh solder onto it. Then wipe it on the sponge again. The tip should be shiny and coated with solder.

   If the iron has been idle awhile, always clean and then re-tin the tip before continuing.

4. The tip of the iron is nickel-plated and therefore should not be filed; this would remove the protective plating on the tip.

5. A *cold solder joint* is a joint containing an air bubble or other impurity that entered as the joint was cooling. Cold solder joints can be identified by their

dull and mottled finish. The solder does not flow and wrap around the terminal like it should.

Cold joints are brittle and make poor electrical connections. To fix such a joint, apply the tip (and perhaps a *very* small amount of solder) at the joint until the solder re-melts and flows into the terminal. If a cold solder joint reappears, remove solder with desoldering wick, and re-solder the joint.

6. Do not hold the iron against the joint for longer than 3-4 seconds, since many electronic components can be damaged by prolonged or excessive heat. Too much heat can also cause the traces on the printed circuit board to peel up or even burn off.

   Components that are particularly sensitive to heat damage include diodes, integrated circuits, LEDs, transistors and voltage regulators.

1.2 Desoldering Technique

It takes about ten times as long to desolder a component as it does to solder it in the first place. This is a good reason to be careful when assembling boards. However, errors will inevitably occur, and it is important to know how to fix them. The primary reasons for performing desoldering are removing an incorrectly-placed component, removing a defective component, and removing solder from a cold solder joint to try again with fresh solder.

The most common methods for removing solder are desoldering pumps and desoldering wick. The ELEC 201 lab kit includes a desoldering pump and desoldering wick as standard equipment.

To use the desoldering wick, simply place the wick between the soldering iron and the excess solder, being very careful not to apply heat for too long.

To use a desoldering pump, first load the pump by depressing the plunger until it latches. Grasp the pump in one hand and the soldering iron the other, and apply heat to the bad joint. When the solder melts, quickly remove the soldering iron and bring in the pump in one continuous motion. Trigger the pump to suck up the solder while it is still molten. This takes a little practice to get right, even for experienced builders.

Adding additional solder to a troublesome joint can be helpful in removing the last traces of solder. This works because the additional solder helps the heat to flow fully into the joint. The additional solder should be applied and de-soldered as quickly as possible. Don't wait for the solder to cool off before attempting to suck it away.

The desoldering pump tip is made of Teflon. While Teflon is heat-resistant, it is not invincible, so the Teflon tip should not be jammed directly into the soldering iron. Solder will not stick to Teflon, so the desoldering operation should suck the solder into the body of the pump. You may need to open the desoldering pump occasionally and empty out accumulated solder.

1.3 Component Types and Polarity

A variety of electronic components will be in use when assembling the boards. The following section provides a brief introduction to these components – how to properly identify and install these parts when building the boards.

1.3.1 Component Polarity

Polarity refers to the fact that many electronic components are not symmetric electrically. Reversing a polarized component in a circuit will result in very different, often undesired effects. For example, passing current through an LED in one direction (forward biased direction) will cause the LED to light up. Reversing the LED will stop current flow up to a certain, often much higher, voltage. Incorrectly mounted polarized devices will not work, and in some cases will be damaged or may damage other parts of the circuit. The following components are always polarized:

- all diodes (LEDs, standard diodes, zener diodes)
- transistors and voltage regulators
- integrated circuits

Capacitors are an interesting case, because some are polarized while others are not. As a general rule, large capacitors (values 1 μF and greater) are usually polarized, while smaller ones are not.

Resistors are not polarized, meaning it does not matter in which direction electricity flows through them. However, it is usual to mount them so their markings all face the same way to make it easier to work on the circuit. (It is also prettty.) The RoboBoard also contains *resistor single-in-line packages* (SIPs), and these have non-symmetric internal wiring configurations, making them polarized from a mounting point of view.

The following paragraphs discuss these components individually, explaining standardized component markings for identifying a component's polarity.

Resistors Individual resistors are small cylindrical devices with color-coded bands indicating their value (how to read color-coding is explained in Figure 10). Most of the resistors in the ELEC 201 kit are rated for $\frac{1}{8}$ watts, which is a very low power rating; hence they are very small devices. A few resistors are much larger: A 1 watt resistor has a large cylindrical shape, while a 5 watt resistor has a large, rectangular package.

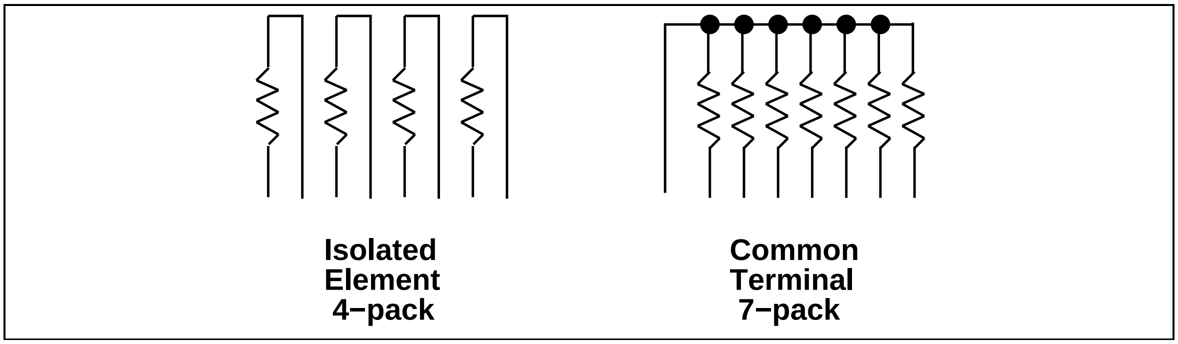

Figure 3: Resistor Pack Internal Wiring

Resistor Packs Resistor packs (or packages) are flat, rectangular packages with four, six, eight, or ten leads. There are two basic types of resistor packs:

Isolated Element. Discrete resistors; usually three, four, or five per package.

Common Terminal. One end of each resistor is tied to a common pin, and the other end is free. The package may contain from three to nine resistors.

Figure 3 illustrates the internal wiring for each style of 8-pin resistor pack.

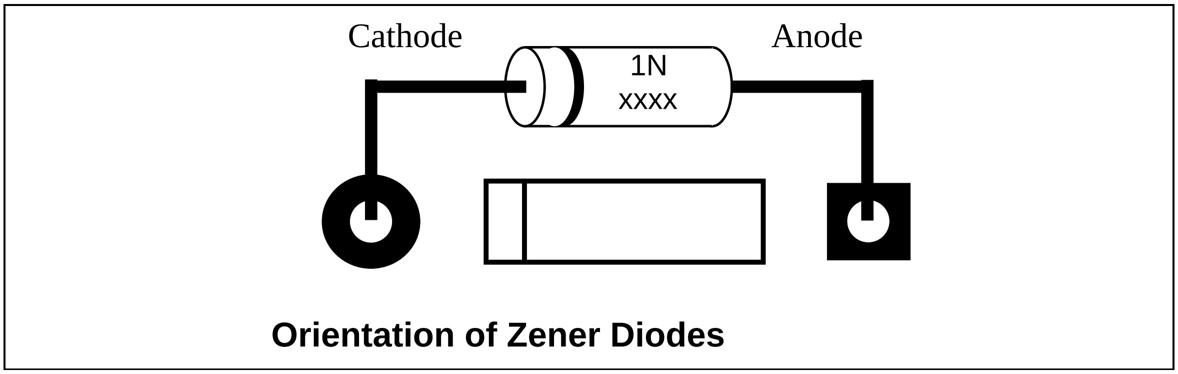

Figure 4: Zener Diode Orientation

All diodes (including the zener diodes and standard diodes used on the RoboBoard) have two leads, called the *anode* and *cathode*. When the anode is connected to positive voltage with respect to the cathode, current can flow through the diode; if polarity is reversed, no current flows. Zener diodes add a twist: at a particular value of reverse polarity voltage (called its "zener," or avalanche voltage), the diode conducts again. This functionality is useful in the design of voltage regulators and (the RoboBoard use) for precise over-voltage protection.

A diode package usually provides a marking that is closer to one lead than the other (a band around a cylindrical package, for example). This marked lead is always the *cathode*, and is always inserted into Pin 2 (the end *opposite* the square pad on the RoboBoard).

Figure 4 shows a the correct orientation of the 1Nxxxx (where "xxxx" is a four digit number denoting the specific diode type) zener diodes used on the RoboBoard and on the battery charger board.

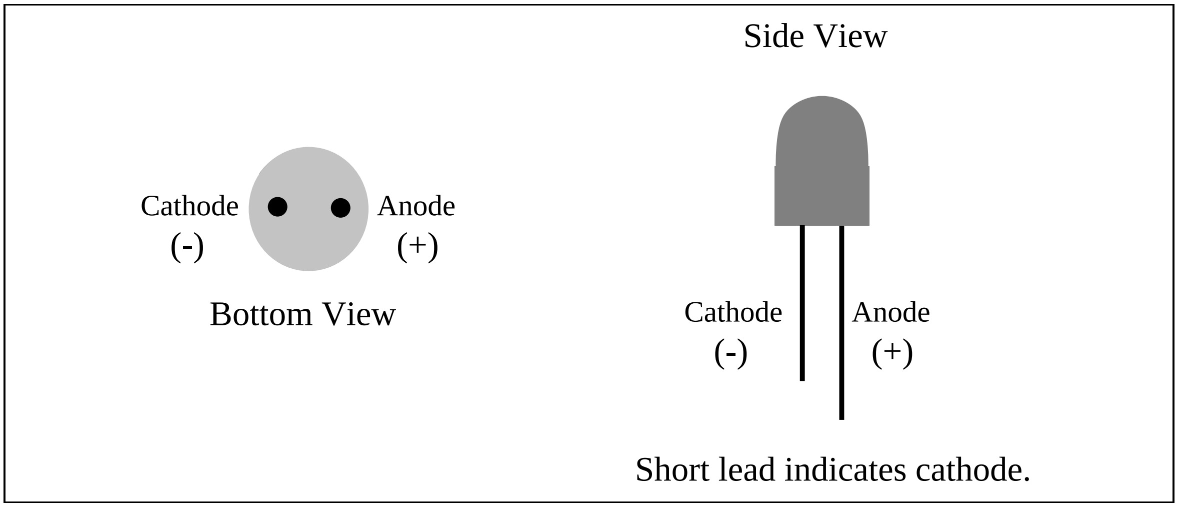
Side View
Cathode
(-)
Anode
(+)
Bottom View
Cathode
(-)
Anode
(+)
Short lead indicates cathode.

Figure 5: Identifying LED Leads

LEDs *LED* is short for "light-emitting diode." An LED's cathode is marked either by a small flat edge along the circumference of the diode casing, or the shorter of two leads. The latter case is the one present in the ELEC 201 kit. Figure 5 shows a typical LED package.

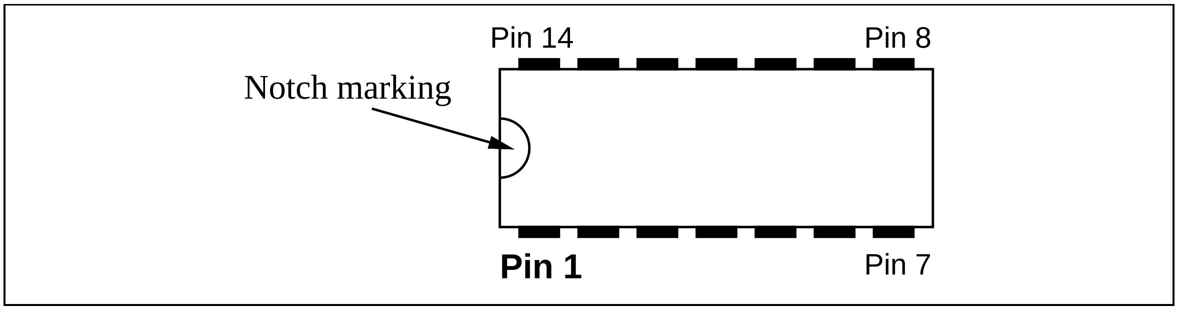
Pin 14
Pin 8
Notch marking
Pin 1
Pin 7

Figure 6: Top View of 14-pin DIP

Integrated Circuits Integrated circuits, or ICs, come in a variety of package styles. Two common types, both of which are used in the RoboBoard design, are called the *DIP* (for dual-inline package), and the *PLCC* (for plastic leadless chip carrier).

In both types, a marking on the component package signifies "pin 1" of the component's circuit. This marking may be a small dot, notch, or ridge in the package. After pin 1 is identified, pin numbering proceeds sequentially in a counter-clockwise direction around the chip package, when looking from the top.

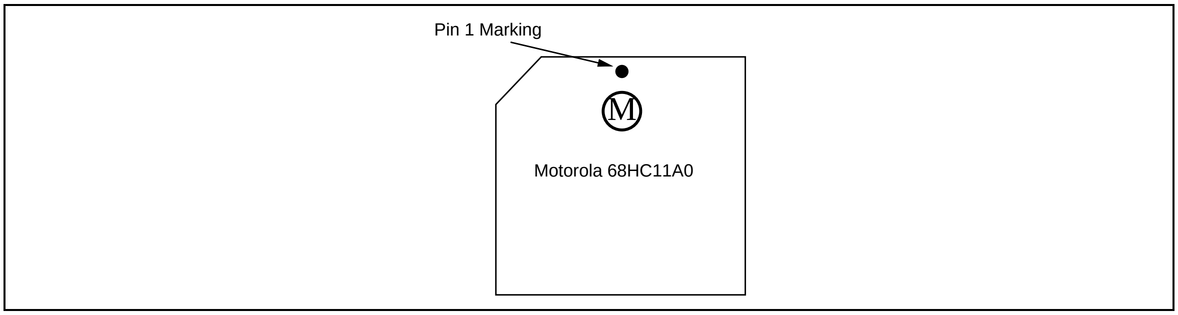

Figure 7: Top View of 52-pin PLCC

Figure 6 shows the typical marking on a DIP package. Figure 7 is a drawing of the PLCC package.

DIP sockets All of the integrated circuits (ICs) on the RoboBoard (except the voltage regulators) are socketed. This means that they are not permanently soldered to the ELEC 201 RoboBoard, but installed in sockets. Components that are socketed can be easily removed if they are damaged or defective. Sockets are also used to avoid the need to solder directly to ICs, reducing the likelihood of heat damage.

Do not place the components into the sockets before you mount the sockets onto the board!

DIP sockets have similar markings to those found on the component that they hold. DIP sockets are not mechanically polarized, but the marking indicates how the chip should be mounted into the socket after the socket has been soldered into the board.

PLCC Sockets sockets *are* polarized, however: a PLCC chip can only be inserted into the socket the "correct" way. Of course, this way is only correct if the socket is mounted right in the first place. When assembling the ELEC 201 RoboBoard, a marking printed onto the board indicates the correct orientation of the PLCC socket.

Capacitors Several different kinds of capacitors are made, each having different properties. There are four different types of capacitors in the ELEC 201 kit:

- **Monolithic.** These are very small-sized capacitors that are about the size and shape of the head of a match from a matchbook. They are excellent for use when small values are needed (0.1 μF and less). They are inexpensive and a fairly new capacitor technology. Monolithic capacitors are never polarized.

- **Electrolytic.** These capacitors look like miniature tin cans with a plastic wrapper. They are good for large values (1.0 μF or greater). They become bulky as the values increase, but they are the most inexpensive for large capacitances.

Electrolytics can have very large values (100 μF and up). They are usually polarized except for special cases; all the electrolytics in the ELEC 201 kit are polarized.

- **Tantalum.** These capacitors are compact, bulb-shaped units. They are excellent for larger values (1.0 μF or greater), as they are smaller and more reliable than electrolytic capacitors. Unfortunately they are decidedly more expensive. Tantalum capacitors are always polarized.
- **Ceramic.** Most often the smallest capacitors in both size and value. Their values are most often measured in picofarads. Ceramic capacitors are never polarized.

As indicated, some capacitors are polarized and will therefore require special attention when mounting. This is very important.

On the ELEC 201 boards, all polarized capacitor placements are marked with a plus symbol (+). The capacitors themselves sometimes are obviously marked and sometimes are not. One or both of the positive or negative leads may be marked, using (+) and (−) symbols. In this case, install the lead marked (+) in the hole marked (+). The (+) hole is also the hole with the square pad. Some capacitors may not be marked with (+) and (−) symbols. In this case, the positive (+) lead will be marked with a dot or a vertical bar.

Polarized capacitors mounted backwards won't work. In fact, they often **overheat and explode**. They should be mounted carefully. When in doubt, ask.

Inductors The inductors used in the ELEC 201 kit are a miniature coil of wire wound around a small ferrite core. They are about the size of a pencil eraser. Some inductors are coated with epoxy and look a lot like resistors. Others are big bulky coils with iron cores. Inductors usually are not polarized.

Transistors and Voltage Regulators The transistors and voltage regulators used in the ELEC 201 kit are small, three-wire devices. They stand out since they are the only three-wire devices used. Transistors and voltage regulators are polarized.

The table shown in Figure 8 summarizes this discussion with regard to polarity issues.

1.4 Component Mounting

When mounting components, the general rule is to try to mount them as close to the board as possible. The main exceptions are components that must be spaced away from the board before being soldered; some resistors fall into this category.

Device	**Polarized?**	**Effect of Mounting Incorrectly**
Resistor	no	
Isolated R-Pack	no	
Common R-Pack	yes	circuit doesn't work
Diode	yes	circuit doesn't work
LED	yes	device doesn't work (no light)
Monolithic capacitor	no	
Tantalum capacitor	yes	explodes
Electrolytic capacitor	yes	explodes
DIP socket	yes	user confusion
PLCC socket	yes	52-pin severe frustration
Integrated circuit	yes	overheating; permanent damage
Inductor	no	
Transistor	yes	circuit doesn't work

Figure 8: Summary of Polarization Effects

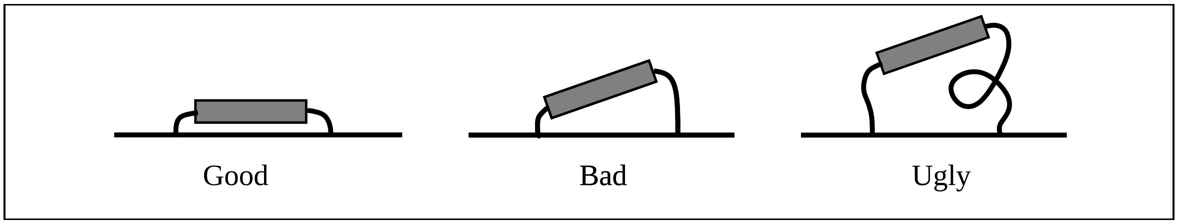

Figure 9: Flat Component Mounting

All LEDs used in the ELEC 201 kit are mounted upright while all resistors are intended to lay flat. See Figure 9 for clarification.

1.5 Component Value Markings

Various electronic components have their values marked on them in different ways. For example, value markings on some resistors are based on a color code (see Figure 10). Other devices, such as transistors and integrated circuits, have their part number printed on the device package. Still others may have several numbers printed on them, some of which are unrelated to their value.

1.5.1 Resistors

The largest resistors – in terms of wattage, not resistive value – simply have their value printed on them.

Other resistors are labeled using a standard *color code*. This color code consists of three value bands (four for a 1% resistor) plus a tolerance band. The first two of the three value bands form the value mantissa. The final value band is the exponent.

It is easiest to locate the tolerance band first. This is a metallic silver-, gold-, or brown-colored band. If it is silver, the resistor has a tolerance of 10%, if gold, the resistor has a tolerance of 5%, and if brown, the resistor has a tolerance of 1%. If the tolerance band is missing, the tolerance is 20%.

The more significant mantissa band begins opposite the tolerance band. If there is no tolerance band, the more significant mantissa band is the one nearer to an end of the resistor. Figure 10 shows the meaning of the colors used in reading resistors.

Color	**Mantissa Value**	**Multiplier Value**
Black	0	1
Brown	1	10
Red	2	100
Orange	3	1000
Yellow	4	10,000
Green	5	100,000
Blue	6	1,000,000
Violet	7	
Grey	8	
White	9	

Figure 10: Resistor Color Code Table

A few examples should make this clear. (See figure 11 for help)

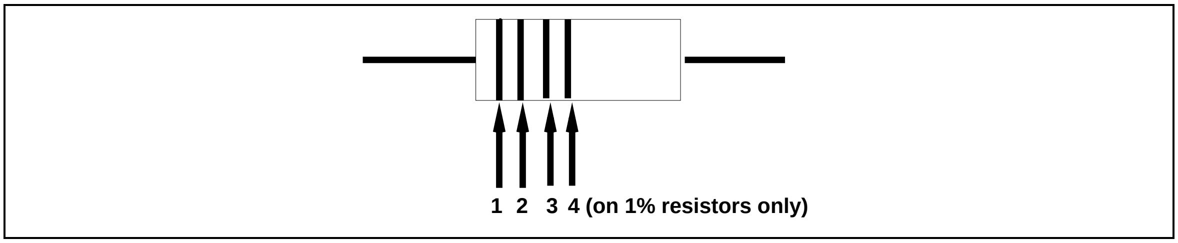

Figure 11: Resistor with Bands

- *brown, black, red*:

  First band: brown-1

  Second band: black-0

  Third band (multiplier value): red-(× 100)

  So have 10 × 100 = 1,000 ohms, or 1kΩ.

- *yellow, violet, orange*:

  First band: yellow-4

  Second band: violet-7

  Third band: orange-(× 1000)

  So have 47 × 1000 = 47,000 ohms, or 47kΩ.

- *brown, black, orange*:

  First band: brown-1

  Second band: black-0

  Third band: orange-(× 1000)

  So have 10,000 ohms, or 10kΩ.

- *brown, black, black, red*:

  First band: brown-1

  Second band: black-0

  Third band: black-0

  Fourth band (multiplier value): black-(× 100)

  So have 100 × 100 = 10,000 ohms, or 10kΩ(for a 1% resistor).

1.5.2 Capacitors

Reading capacitor values can be confusing because there often are numbers printed on the capacitor that have nothing to do with its value. So the first task is to determine which are the relevant numbers and which are the irrelevant ones.

For large capacitors (values of 1μF and greater), the value is often printed plainly on the package; for example, "4.7μF". Sometimes the "μ" symbol acts as a decimal point; e.g., "4μ7" for a 4.7μF value. Capacitors smaller than 1μF have their values printed in picofarads (pF). There are 1,000,000 pF in one μF, and 1,000,000 μF in one Farad.

Capacitor values are similar to resistor values in that there are two digits of mantissa followed by one digit of exponent. Hence the value "472" indicates 47×10^2 picofarads, which is 4700 picofarads or 0.0047 μF.

2 Assembling the ELEC 201 RoboBoard

The ELEC 201 RoboBoard[2] is the heart of the ELEC 201 Robot Controller system. It uses a Motorola 68HC11 family microprocessor coupled with 32K of non-volatile memory. It has outputs to drive six motors, inputs for a variety of sensors, a serial communications port for downloading programs and user interaction, and many other features.

2.1 Assembly Instructions

Before beginning assembly, make sure to have a well-lighted, well-ventilated workspace. Make sure that all of the electronic assembly tools are available. Turn on the light at your workstation.

All of the ELEC 201 boards have component placements silkscreened directly onto the board. In addition, diagrams in these instructions will provide copies of the diagrams printed on the boards, often at better resolution. Refer to the printed diagrams as often as necessary to be sure that components are being placed correctly. **Correct component placement is absolutely critical. A few minutes spent checking placement and orientation prior to soldering will save hours of debugging and rework later.**

Figure 12 illustrates the component placement on the microprocessor board. The component numbering for parts on the microprocessor board increments in a counter-clockwise fashion around the board for resistors, capacitors, and resistor packs.

Each section of the instructions, or the individual listing of the component contains a reference to the bin in the parts kit where the component can be found. The instruction checklist may be marked off as each step is completed.

1–☐ **Get the ELEC 201 V4.1 RoboBoard, and determine which is the "component side."** The RoboBoard is the largest of the ELEC 201 boards.

---

[2] These instructions are for Versions 4.0 and 4.1 of the RoboBoard. Versions 1.0, 1.1, 2.0, and 3.0 of the RoboBoard use different assembly instructions.

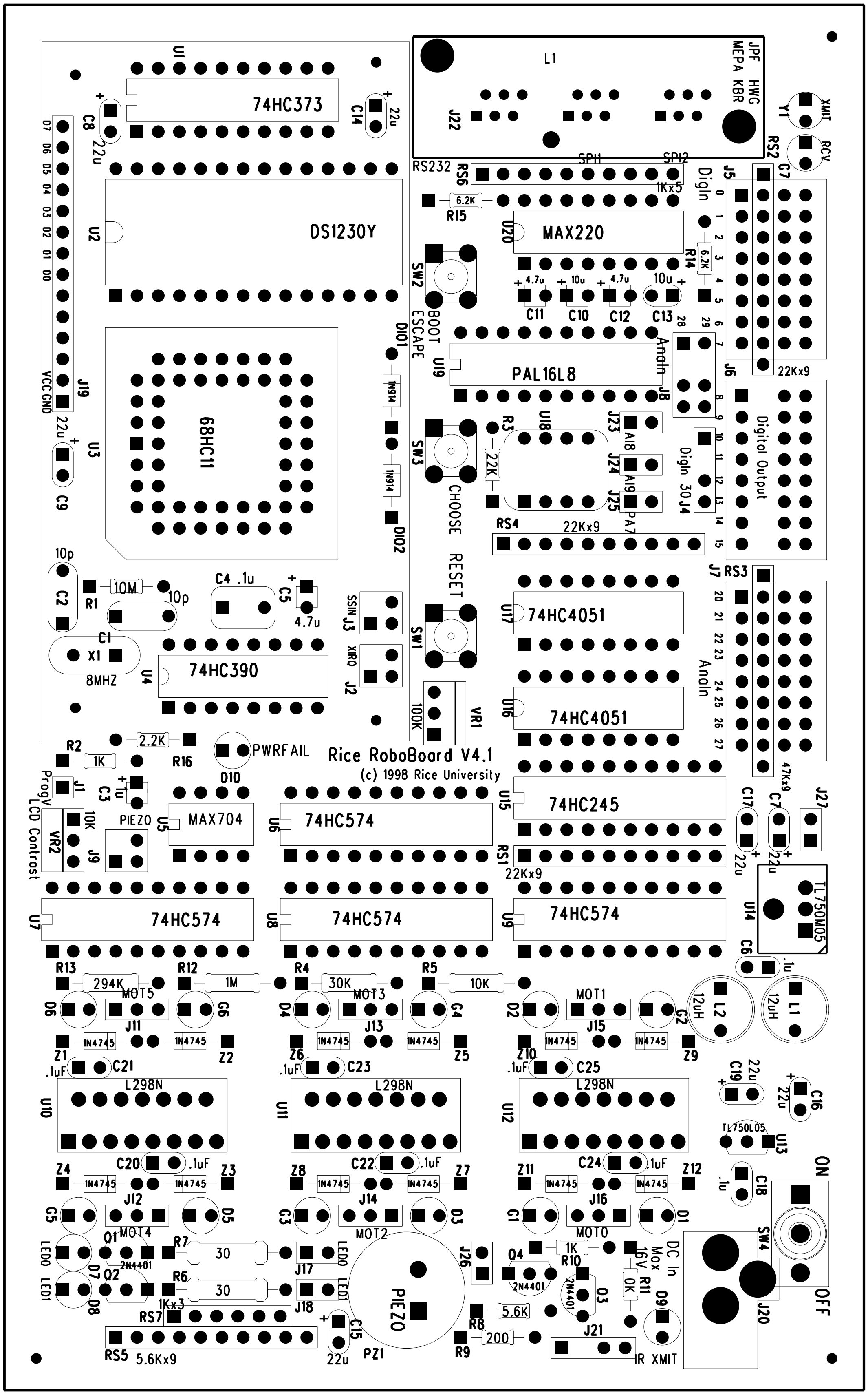
U1
74HC373
C8
22u
C14
22u
L1
J22
JPF HWG
MEPA KBR
Y1
XMIT
RCV
RS2
G7
D7 D6 D5 D4 D3 D2 D1 D0
U2
DS1230Y
RS232
RS6
SPI1
SPI2
1Kx5
J5
DigIn
R15
6.2K
U20
MAX220
R14
6.2K
SW2
BOOT
ESCAPE
4.7u
10u
4.7u
10u
C11
C10
C12
C13
28
29
22Kx9
J6
DIO1
1N914
U19
PAL16L8
AnaIn
J8
VCCGND
J19
22u
U3
68HC11
C9
SW3
R3
22K
U18
J23
AI8
J24
AI9
J25
PA7
DigIn 30
J4
Digital Output
CHOOSE
DIO2
RS4
22Kx9
RESET
10p
C2
R1
10M
10p
C4 .1u
C5
4.7u
SSIN
J3
J7
RS3
U17
74HC4051
C1
X1
8MHZ
U4
74HC390
XIRQ
J2
SW1
AnaIn
VR1
100K
U16
74HC4051
R2
1K
2.2K
R16
PWRFAIL
D10
Rice RoboBoard V4.1
(c) 1998 Rice University
47Kx9
U15
74HC245
ProgV
J1
C3
1u
PIEZO
U5
MAX704
U6
74HC574
C17
C7
J27
22u
22u
LCD Contrast
VR2
10K
J9
RS1
22Kx9
TL750M05
U14
U7
74HC574
U8
74HC574
U9
74HC574
C6
.1u
R13
294K
R12
1M
R4
30K
R5
10K
MOT5
MOT3
MOT1
12uH
L2
12uH
L1
D6
J11
G6
D4
J13
G4
D2
J15
G2
Z1
1N4745
1N4745
Z2
Z6
1N4745
1N4745
Z5
Z10
1N4745
1N4745
Z9
.1uF
C21
.1uF
C23
.1uF
C25
C19
22u
C16
22u
U10
L298N
U11
L298N
U12
L298N
TL750L05
U13
C20
.1uF
C22
.1uF
C24
.1uF
C18
.1u
ON
Z4
Z3
Z8
Z7
Z11
Z12
1N4745
J12
J14
J16
G5
D5
G3
D3
G1
D1
MOT4
MOT2
MOT0
Q1
R7
30
LED0
J17
J26
Q4
1K
R10
DC In
Max
16V
SW4
LED0
LED1
D7
Q2
2N4401
R6
30
LED1
J18
PIEZO
2N4401
Q3
R11
10K
D9
OFF
J20
D8
1Kx3
RS7
C15
R8
5.6K
J21
RS5
5.6Kx9
22u
PZ1
R9
200
IR XMIT

Figure 12: ELEC 201 RoboBoard Component Placement

The side of the board that has been printed with component markings is the "component side." This means that components are mounted by inserting them down from the printed side; then they are soldered on the opposite, unprinted side.

Make sure that the components are mounted on the proper side of the board!

2–☐ Measure the resistance between Power and Ground

Using the multimeter set to resistance ("Ω"), touch the red lead to Pin 20 of U19, and the black lead to Pin 10 (refer to Figure 6 if you do not remember the pin numbering scheme.) The value read should be higher than 10 megaohms. If this is not the case, contact your lab assistant *before* proceeding.

If the reading on the multimeter does not change (e.g. the meter reads "O.L"), all is well. However, to ensure that the meter is working properly, connect the probes together. You should get a low resistance reading.

3–☐ IC Sockets.

Mount the DIP sockets such that the notch in the socket lines up with the notch marking in the rectangular outline printed on the PC board. In the list that follows, "DIP5" means the DIP socket used for integrated circuit U5.

Note that there are two varieties of 16-pin DIP socket. One type has an integral bypass capacitor and costs about four times as much as the standard variety. None of the 8-pin sockets have integral bypass capacitors.

A caddy-cornering soldering technique technique should be used. After inserting a DIP socket into the board, solder its two opposite-corner pins first. This will hold the chip in place. Look at it carefully to check for correct orientation, and that it is level and down as far as it can go. If not, it is easy to heat one or both pins and adjust the socket. Only when everything is correct, solder the other pins.

If you are sure of your technique, a time-saving method is to place all of the sockets where they belong, then hold a book against the top of the board (and therefore the sockets) while the board is turned over. Then solder two pins on each socket, and check them all before soldering the rest of the pins.

NOTE: U13 and U14 are *not* socketed. They will be installed later.

Bin A

○ **DIP1**–20 pins (with bypass capacitor)
○ **DIP2**–28 pins (with bypass capacitor)
○ **DIP4**–16 pins (with bypass capacitor)
○ **DIP5**–8 pins (no bypass capacitor)
○ **DIP6**–20 pins (with bypass capacitor)

- ○ **DIP7**–20 pins (with bypass capacitor)
- ○ **DIP8**–20 pins (with bypass capacitor)
- ○ **DIP9**–20 pins (with bypass capacitor)
- ○ **DIP15**–20 pins (with bypass capacitor)
- ○ **DIP16**–16 pins (with bypass capacitor)
- ○ **DIP17**–16 pins (with bypass capacitor)
- ○ **DIP18**–8 pins (no bypass capacitor)
- ○ **DIP19**–20 pins (with bypass capacitor)
- ○ **DIP20**–16 pins (no bypass capacitor)

4–☐ **Resistor Packs.**

Install the resistor packs. Most of the resistor packs are polarized: the common terminal end is marked with a dot or band. On the ELEC 201 RoboBoard, find a *square metal pad* at one end of the area that each resistor pack will mount. *Insert the resistor pack such that the marked end mounts in the square hole.* (The square hole is sometimes more easily seen on the unprinted solder side of the board.)

The "caddy-cornering" technique of soldering the two end terminals first is helpful here also; check the orientation *before* soldering the other pins. **Bin B**

- ○ **RS1**–22kΩ×9, 10 pins, polarized (marked end goes in the square pad), *marked "101R22K"* or *marked "10R-1-223."*
- ○ **RS2**–22kΩ×9, 10 pins, polarized, *marked "101R22K"* or *marked "10R-1-223."*
- ○ **RS3**–47kΩ×9, 10 pins, polarized, *marked "101R47K."*
- ○ **RS4**–22kΩ×9, 10 pins, polarized, *marked "101R22K"* or *marked "10R-1-223."*
- ○ **RS5**–5.6kΩ×9, 10 pins, polarized, *marked "101R5.6K."*
- ○ **RS6**–1kΩ×5, 10 pins, non-polarized (can be mounted either way), *marked "103R1K."*
- ○ **RS7**–1kΩ×3, 6 pins, non-polarized, *marked "63R1K."*

5–☐ **Non-polarized Capacitors.** These capacitors may be mounted in either orientation. To keep each component in place while soldering, bend the leads slightly out after inserting into the pad. After installing, solder and clip leads close to the board at the edge of the solder joint. **Always hold the end of the lead being clipped, so that the lead cannot become a projectile.**

Bin E

- ○ **C1**–10 pF, *marked "10."* **Caution:** The label "C1" on the board is not printed in quite the right place. It is located on an oval with "X1" in the center; *This is not where C1 goes.* You will solder a crystal here later. Look above and slightly to the right of the printed C1 and find another oval with nothing printed in it. Put C1 there. Ask a laboratory assistant if you have any questions.
- ○ **C2**–10 pF, *marked "10."*
- ○ **C4**–0.1 μF, *marked "104K 100F" or "104K 100E", or "104K 100G"*
- ○ **C6**–0.1 μF, *marked "104K." (or "104")*
- ○ **C18**–0.1 μF, *marked "104K." (or "104")*
- ○ **C20**–0.1 μF, *marked "104K." (or "104")*
- ○ **C21**–0.1 μF, *marked "104K." (or "104")*
- ○ **C22**–0.1 μF, *marked "104K." (or "104")*
- ○ **C23**–0.1 μF, *marked "104K." (or "104")*
- ○ **C24**–0.1 μF, *marked "104K." (or "104")*
- ○ **C25**–0.1 μF, *marked "104K." (or "104")*

6–☐ Microprocessor Socket (U3) (Bin A).

- ○ Install **PLCC1**, a 52-pin square socket for the 68HC11 (U3). The Pin 1 marking is indicated by the numeral "1" and an arrow in the socket; this marking mounts nearest to **C9** (next to the edge). There should be a beveled notch in the lower-right corner of the chip and the outline printed on the board, with respect to the pin 1 marking. *Be absolutely sure to mount this socket correctly; the socket is polarized and will only let you mount the chip into it one way.* Solder.

7–☐ Male and Female Socket Headers.

To cut sections of header, it is usually best to destroy one unit of header. For example, if you are cutting a 3-long single female strip, count three units and cut through the fourth. After cutting you can trim the excess header with diagonal cutters, or by smoothing with sandpaper.

When mounting the headers, pay attention to how well they are lining up vertically. Sometimes reversing the way a strip is mounted will help its connections to line up better with the others. It may be helpful to insert a strip of male header into the socket to hold them at proper horizontal and vertical placement before soldering.

Both single- and double-wide female header strips are used in the following assembly instructions.

Refer to Figure 13 for placement of these parts.

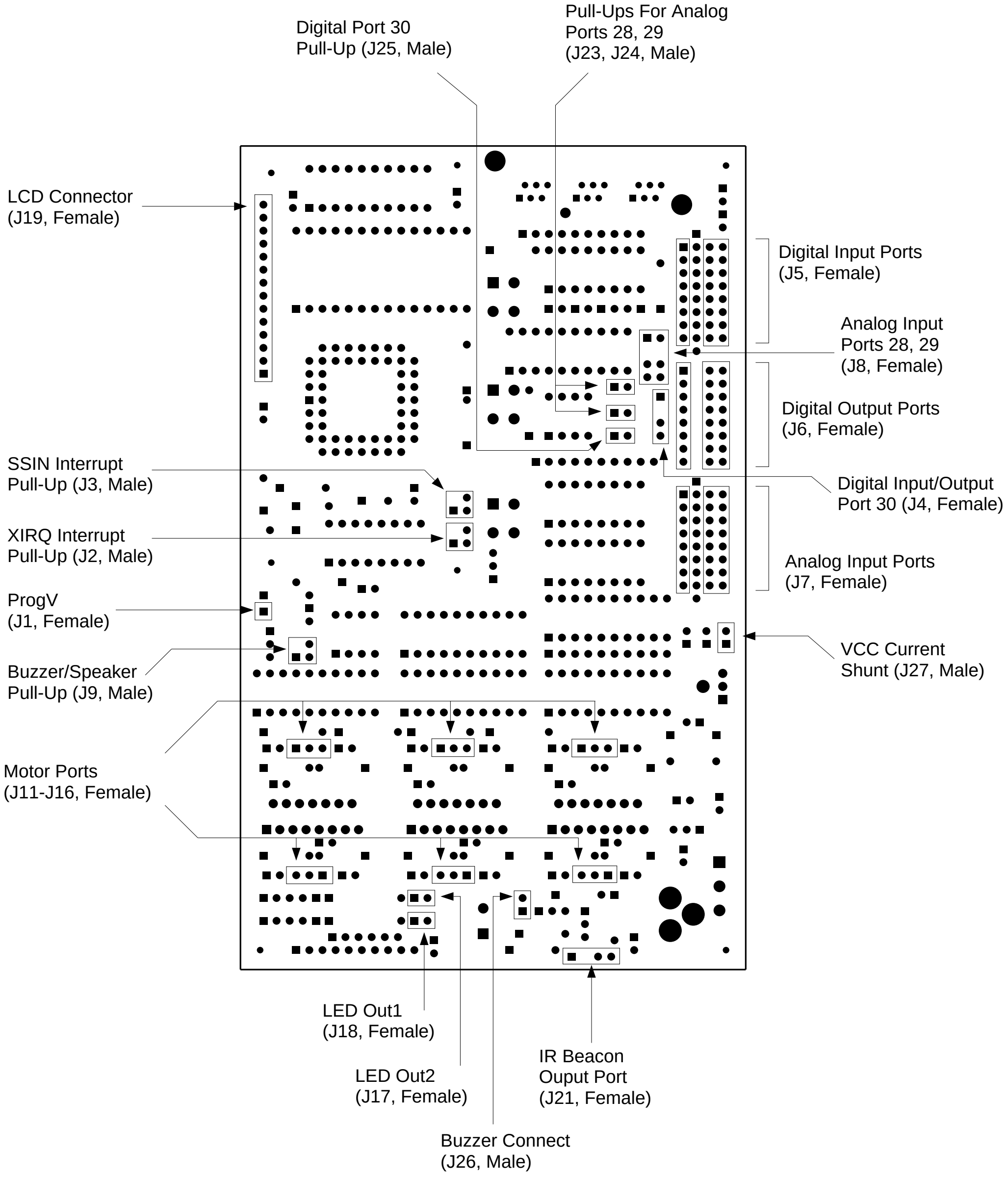
Digital Port 30
Pull-Up (J25, Male)
Pull-Ups For Analog
Ports 28, 29
(J23, J24, Male)
LCD Connector
(J19, Female)
Digital Input Ports
(J5, Female)
Analog Input
Ports 28, 29
(J8, Female)
Digital Output Ports
(J6, Female)
SSIN Interrupt
Pull-Up (J3, Male)
Digital Input/Output
Port 30 (J4, Female)
XIRQ Interrupt
Pull-Up (J2, Male)
Analog Input Ports
(J7, Female)
ProgV
(J1, Female)
VCC Current
Shunt (J27, Male)
Buzzer/Speaker
Pull-Up (J9, Male)
Motor Ports
(J11-J16, Female)
LED Out1
(J18, Female)
LED Out2
(J17, Female)
IR Beacon
Ouput Port
(J21, Female)
Buzzer Connect
(J26, Male)

Figure 13: ELEC 201 RoboBoard Header Placement

- ○ The motor chips have special, pre-cut female header of 7 and 8 pin lengths. Each chip requires one seven pin and one eight pin header. Install the motor mount hardware **(Bin J)** for **U10**, **U11**, and **U12**.
- ○ Cut three 8-long double female strips, and select three additional pre-cut 8-long single female strips (Bin J). Install the **J5**, **J6**, and **J7** connector blocks. Solder.
- ○ Cut one 2-long double female strip, and one 2-long single female strip. Install the **J8** connector block. Solder.
- ○ Cut two 2-long single female strips, and two 1-long single female strips. Install the **J4** and **J21** connector blocks. Solder.
- ○ Cut three 2-long male strips and three 1-long male strips. They will be installed in groups of three (a triangular pattern of three pins). Install the **J2**, **J3**, and **J9** connector blocks. Solder. The ordering of the two pieces of header is not really important, but it is best to be consistent. For example, put the 1-long strips into pin 1 and the 2-long strips into pins 2 and 3.
- ○ Cut six 3-long single female strips. Install the **J11**, **J12**, **J13**, **J14**, **J15**, and **J16** connector blocks. Solder.
- ○ Cut two 2-long single female strips. Install the **J17** and **J18** connector blocks. Solder.
- ○ Cut five 2-long single male strips. Install the **J23**, **J24**, **J25**, **J26**, and **J27** connector blocks. Solder.
- ○ Cut one 14-long single female strip. Install the **J19** connector. Solder.
- ○ Cut one 1-long single female strip. Install the **J1** connector. Solder.

8–☐ **LEDs.**

LEDs must be mounted so that the *long lead* (the anode) is inserted into the *square* (as opposed to round) pad of the LED footprint.

Be sure to mount LEDs properly as it is very difficult to desolder them without damage if they are mounted backward.

To aid in soldering, bend the leads a little bit after inserting into the board to hold in the LED. **Remember, LEDs are heat-sensitive, so do not hold the soldering iron on the pad too long.** Also, after soldering in one pin of an LED, give it a few seconds to cool down. Check that the recently soldered pad is cool to the touch, *carefully!*, before soldering the other lead. **Bin G**

- ○ **D1**–red
- ○ **D2**–red
- ○ **D3**–red
- ○ **D4**–red

- ○ **D5**–red
- ○ **D6**–red
- ○ **D7**–red
- ○ **D8**–red
- ○ **D9**–red
- ○ **D10**–red
- ○ **Y1**–yellow
- ○ **G1**–green
- ○ **G2**–green
- ○ **G3**–green
- ○ **G4**–green
- ○ **G5**–green
- ○ **G6**–green
- ○ **G7**–green

9–☐ Resistors.

All resistors (except **R6** and **R7**) mount flat against the RoboBoard; try to get the body of the resistor very close to the board. These components are not polarized; however, we suggest that resistors be mounted uniformly – with each resistor's tolerance band closest to the top of the board.

If you have trouble discerning colors, you may wish to have your colleagues handle this task. Actually, it is fairly difficult to read the color bands from $\frac{1}{8}$ watt resistors, even for the trained eye. The blue body of some resistors seem to distort the color of the bands. The best technique is to check the value of each resistor with the Digital Multi-Meter.

After soldering each resistor, check the component side of the joint to ensure that the solder flowed all the way through the joint. If it hasn't, turn the board back over to the solder side, and touch the joint again with the iron briefly– this usually helps. Once the resistor is successfully soldered in place, clip the leads. (Again, be sure to hold each lead while clipping to prevent possibly hazardous projectile wires.)

Make sure that you use 1% resistors where they are called for.

Bin B

- ○ **R1**–10MΩ, *brown, black, blue* (beige body)
- ○ **R2**–1kΩ, *brown, black, red* (beige body)
- ○ **R3**–22kΩ, *red, red, orange* (beige body)
- ○ **R4**–30.1kΩ, 1% *orange, black, brown, red* (blue body)

- ○ **R5**–10kΩ, 1% *brown, black, black, red* (blue body)
- ○ **R6**–30Ω, 1 watt *orange, black, black* (large gray body) NOTE: R6 and R7 should be spaced above the RoboBoard, about as high as the header for **J17** and **J18**.
- ○ **R7**–30Ω, 1 watt *orange, black, black* (large gray body)
- ○ **R8**–5.6kΩ, *green, blue, red* (beige body)
- ○ **R9**–200Ω, *red, black, brown* (beige body)
- ○ **R10**–1kΩ, *brown, black, red* (beige body)
- ○ **R11**–0Ω, *use a small piece of wire or a clipped resistor lead.*
- ○ **R12**–1MΩ, 1% *brown, black, black, yellow* (blue body)
- ○ **R13**–294KΩ, 1% *red, white, yellow, orange* (blue body)
- ○ **R14**–6.2kΩ, *blue, red, red* (beige body)
- ○ **R15**–6.2kΩ, *blue, red, red* (beige body)
- ○ **R16**–2.2kΩ, *red, red, red* (beige body)

10–☐ Electrolytic and Tantalum Capacitors.

All of these capacitors are polarized. Make sure that the longer lead (marked (+)) on the capacitor goes into the square pad (marked (+)). If the capacitor leads are not marked (+) or (−), a dot or bar will indicate the (+) lead. Be sure on the polarity of these capacitors. The results are pretty spectacular if these are put in backwards.

Bin E

- ○ **C3**–1 μF, 16 volt tantalum, *marked "1C"*
- ○ **C5**–4.7 μF, 16 volt tantalum, *marked "4.7", and "16"*
- ○ **C7**–22 μF, 16 volt tantalum, *marked "22", and "16"*
- ○ **C8**–22 μF, 16 volt tantalum, *marked "22", and "16"*
- ○ **C9**–22 μF, 16 volt tantalum, *marked "22", and "16"*
- ○ **C10**–10 μF, 16 (or 6.3) volt tantalum, *marked "10", and "16" (or 6.3)*
- ○ **C11**–4.7 μF, 16 volt tantalum , *marked "4.7", and "16"*
- ○ **C12**–4.7 μF, 16 volt tantalum, *marked "4.7", and "16"*
- ○ **C13**–10 μF, 16 volt tantalum, *marked "10", and "16"*
- ○ **C14**–22 μF, 16 volt tantalum, *marked "22", and "16"*
- ○ **C15**–22 μF, 25 volt tantalum, *marked "22", and "25"*
- ○ **C16**–22 μF, 25 volt tantalum, *marked "22", and "25"* Note: Insert SW4 (Bin H) into the board before soldering C16 to ensure that C16 does not interfere with the placement of SW4.

- ○ **C17**–22 μF, 16 volt tantalum, *marked "22", and "16"*
- ○ **C19**–22 μF, 16 volt tantalum, *marked "22", and "16"*

11–☐ Transistors and Voltage Regulators

Transistors and voltage regulators are polarized. Install them so that their flat edge is above the flat edge of the placement marking. *Transistors and voltage regulators are* very *heat sensitive.* Put your finger (have one of your partners feed solder into the pad for you) on the top to gauge heat while soldering, and wait awhile after soldering each lead to let the device cool down.

Bin D

- ○ **Q1**–2N4401
- ○ **Q2**–2N4401
- ○ **Q3**–2N4401
- ○ **Q4**–2N4401
- ○ **U13**–TL750L05 (same package as transistors). This part is extremely heat sensitive, so be careful.
- ○ **U14**–TL750M05 (the large rectangular version of this device). Install the slip-on heatsink for the TL750M05 regulator (**U14**). First, obtain some thermal compound from your lab assistant. Place a *small* amount of this compound on the metal side of the regulator (facing the board). Then gently press the slip-on heat sink onto the regulator. Guide the regulator pins and heatsink pin into the board until both are snug. Ensure a snug fit, solder and trim the support pin of the heat sink. Solder the regulator pins and trim.

12–☐ 8 MHz Crystal (Bin K)

Crystals are not polarized. Install **X1**, the 8 MHz. crystal.

13–☐ Inductors. (Bin K)

Inductors are not polarized. Install **L1** and **L2**, 12 uH.

- ○ **L1**–12 uH (may be marked "120K")
- ○ **L2**–12 uH (may be marked "120K")

14–☐ Switches.

These are slightly rectangular and should only insert easily in the correct orientation.

Bin K

- ○ **SW1**–miniature pushbutton switch

○ **SW2**–miniature pushbutton switch

○ **SW3**–miniature pushbutton switch

Bin H

○ **SW4**–large slide switch; either orientation works.

15–☐ Variable Resistors. (Bin K)

Orient the variable resistors such that the line on the decal and the dial of the resistor are lined up. *Check carefully for the correct value.*

○ **VR1**, 100kΩ.

○ **VR2**, 10kΩ.

16–☐ Power Receptacle. (Bin H)

Install **J20**, DC power receptacle. When soldering, feed ample amounts of solder into the joint, so that solder completely fills the mounting pads. Try not to melt the receptacle.

17–☐ RJ-11 Top-Entry RJ11 Receptacle. (Bin K)

The top-entry RJ11 phone receptacle needs to be modified before it can be mounted on the RoboBoard. The correct procedure is important. Ask your lab assistant if you have any questions.

Hold the receptacle such that the connecting and mounting pins are pointing downward and the notch is facing you and the row of vertical lines is on the opposite side. Locate the larger triangular mounting pins on your left hand side. The inside mounting pin (the one closest to you) needs to be clipped. Bend this pin outwards to form a right angle. Using the diagonal cutters, clip this mounting pin close to the plastic package.

Install **J22**, the one-position RJ11 receptacle in the space nearest to the center of the board, where it will cover the “J22” mark. The notch should be closest to the edge of the board. Use care in aligning the pins. Be extra careful when inserting not to damage the pins. Solder the mounting pins as well.

18–☐ Piezo beeper. (Bin K)

The beeper is polarized, so mount the Piezo beeper with the long lead in the square pad, and make sure that it is centered on the circular outline.

19–☐ Diodes

Zener diodes are polarized. Install them so that the black band on the diode is in the same orientation as white band marked on the RoboBoard (see Figure 4). *Correct orientation is critical!* If in doubt, ask. Zener diodes *are* *very heat sensitive.* Put your finger on the top to gauge heat while soldering, and wait awhile after soldering each lead to let the device cool down.

Bin J

○ **Z1**–1N4745
○ **Z2**–1N4745
○ **Z3**–1N4745
○ **Z4**–1N4745
○ **Z5**–1N4745
○ **Z6**–1N4745
○ **Z7**–1N4745
○ **Z8**–1N4745
○ **Z9**–1N4745
○ **Z10**–1N4745
○ **Z11**–1N4745
○ **Z12**–1N4745

Two standard diodes are also used on the RoboBoard. **These diodes are polarized** like the zener diodes. Install them with the same orientation as the zeners and with the same care.

Bin G

○ **DIO1**–1N914
○ **DIO2**–1N914

CONGRATULATIONS! You have completed basic assembly of the RoboBoard.

2.2 Final Assembly of the RoboBoard

This section explains a few simple tests to be performed *before* installing the ICs in the sockets.

Full board testing and debugging will be handled in the laboratory.

1–☐ Check the solder side of the board for proper solder connections. Specifically: look for solder bridges and cold solder joints.

Solder bridging is when a piece of solder "bridges" across to adjacent terminals that should not be connected.

Cold solder joints are recognized by their dull luster. A cold solder joint typically makes a poor electrical connection. Make sure that all of the solder joints are shiny with a silver color. Everyone in your team should check the board, and then ask a lab assistant or the instructor for a final check.

2–☐ Install the eight shunts onto the male headers as follows (see Figure 14 for how pins are numbered):

Bin H

○ **J2** – pin 2 to pin 3 (pin 1 is the square pin)
○ **J3** – pin 1 to pin 3
○ **J9** – pin 1 to pin 3
○ **J23** – pin 1 to pin 2
○ **J24** – pin 1 to pin 2
○ **J25** – pin 1 to pin 2
○ **J26** – pin 1 to pin 2
○ **J27** – pin 1 to pin 2

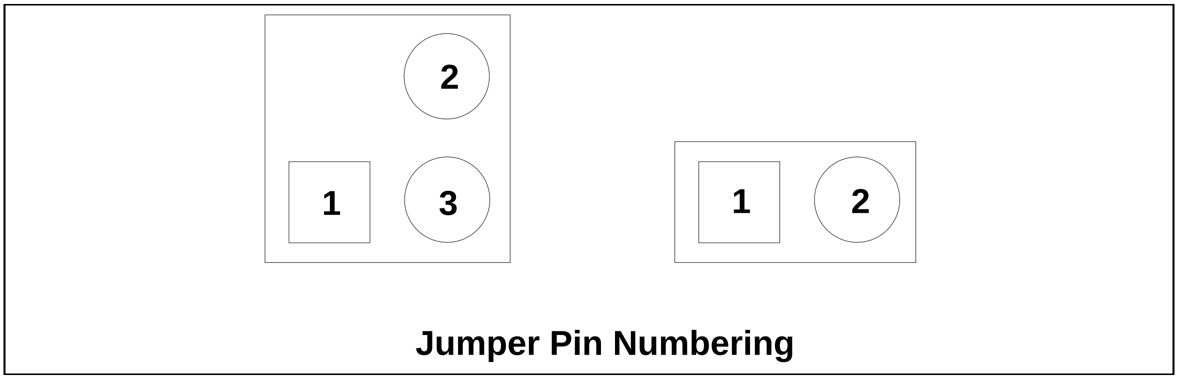

Figure 14: RoboBoard Jumper Pin Numbering

3–☐ Using the multimeter set to resistance ("Ω"), touch the red lead to Pin 20 of U19, and the black lead to Pin 10.

Resistance should increase as the capacitors charge. The board resistance should measure between 1kΩ and 10kΩ. *If a reading of zero ohms is observed, the board probably has a power to ground short.* Do not proceed with testing until this is corrected.

4–☐ Ask a lab assistant for a battery and cable, and note their instructions on how to connect it. Connect battery power to the board.

5–☐ Slide the board power switch to "ON".

6–☐ Measure the board voltages. Using the multimeter set to DC Volts ("DC"), touch the red lead to Pin 20 of **U19**, and the black lead to Pin 10. You should read approximately 5.0 volts.

Then touch the red lead to pin 3 of **U13** (pin 3 is in the outside round pad closest to **U12**). You should read approximately 5.0 volts.

7–☐ Slide the board power switch to "OFF".

8–☐ Disconnect the battery.

9–☐ Install the socketed ICs into their sockets on the board. *Be careful not to damage the component leads when installing the chips into their sockets!* Make sure to get the orientation correct (the notched end of the chip should match the notch on the socket and board)—refer to Figure 12 if necessary. **Always touch the board with a part of your body** *before* **bringing the IC in contact with the socket.** This ensures that any static charge that may be present will discharge through you, rather than through the more delicate integrated circuit.

Pushing ICs into sockets can be difficult. It is usually helpful to bend the leads slightly before insertion; ask a lab assistant to show you how. Make sure that all pins of the IC go into the socket, as they sometimes get bent underneath the chip and are difficult to see. If you are having trouble pushing the ICs into the sockets, take the needle nose pliers and *gently* bend the pins inward so the chip will socket easier. Also, make sure to apply even pressure to both sides of the chip so that both sides go into the socket at the same time. It is difficult to put in one side after the other has been inserted.

Again, double-check to ensure correct orientation, because if they are in the wrong orientation when you turn on the power, they are usually destroyed.

○ **U1**–MC74HC373AN (Bin F)
○ **U2**–DS1230Y 32K static RAM (expensive part – use special care) (Bin I)
○ **U3**–68HC11 microprocessor (expensive part – use special care) (Bin I)
○ **U4**–MC74HC390N (Bin I)
○ **U5**–MAX704 (Bin L)
○ **U6**–SN74HC574N (Bin L)
○ **U7**–SN74HC574N
○ **U8**–SN74HC574N
○ **U9**–SN74HC574N
○ **U15**–SN74HC245N (Bin F)
○ **U16**–MC74HC4051N (Bin C)
○ **U17**–MC74HC4051N

○ **U18**–4-pole Dipswitch (Pin one is the pin below the word “ROCKER” so switch 1 will be nearest the microprocessor) (Bin F)

○ **U19**–PAL16L8 (Bin L)

○ **U20**–MAX220CPE (Bin F)

○ **Motor Drivers**

These devices can only be inserted in one orientation. Insert the three L298N’s into the female headers.

Bin C

1–☐ **U10**–L298N
2–☐ **U11**–L298N
3–☐ **U12**–L298N

○ Have your lab assistant or the instructor examine the board for correct placement, orientation, etc., BEFORE PROCEEDING FURTHER.

2.3 Preliminary Testing

To test the basic functionality of your RoboBoard, borrow an assembled LCD unit from a lab assistant and install it (without screws) into the board. Connect your board to the battery and to the computer as described in the Class Notes on the web. Follow the steps described to load IC, and then test some of the simple IC commands.

2.4 LCD Installation

The LCD display provided in the ELEC 201 kit can display two rows of 16 characters. The system software makes it easy to write code that prints messages to this display, for status, debugging, or entertainment purposes. The LCD installation should be performed only after the board has been demonstrated to work properly. It is difficult to debug a board once the LCD display has been bolted on.

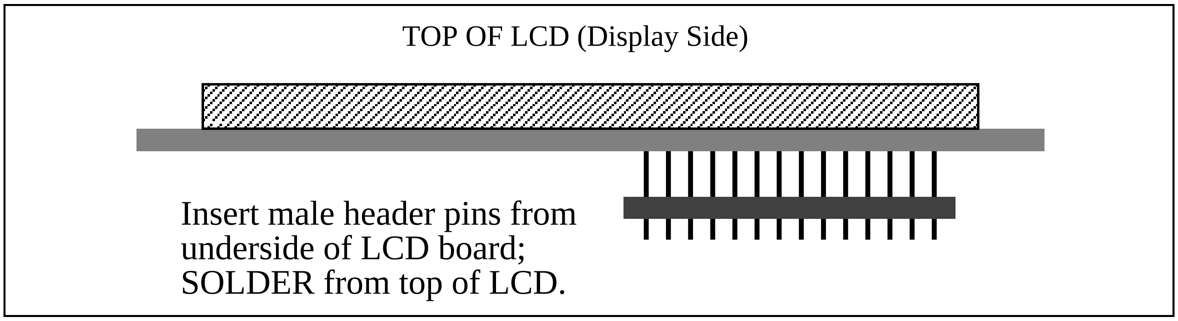
TOP OF LCD (Display Side)
Insert male header pins from underside of LCD board; SOLDER from top of LCD.

Figure 15: LCD Connector Mounting

The display needs to have a 14-pin male header soldered to its interface. Figure 15 shows how these pins should be installed. You may receive an LCD that is already soldered to a header strip. In that case, you can skip the second and third steps below.

1–☐ Mount four threaded nylon spacers to the RoboBoard with 6-32 nylon screws (do not overtighten!)

2–☐ The LCD mounting holes at the corners need to be drilled larger to fit the mounting screws. If you receive a new display, ask a lab assistant to do this.

3–☐ Cut a 14-long single male LCD header strip; the long pins are 0.31 inches long. Insert the *short* side of the header into **J19**, then mount the LCD as indicated in the figure, putting the top header pins through the holes in the LCD board.

4–☐ Attach the LCD display to the spacers with four more 6-32 nylon screws. The LCD display will now be in correct position and alignment. Solder the 14 header pins to the LCD board. To remove the LCD display, simply remove the four 6-32 nylon screws holding the display to the spacers, and gently pull the LCD display and attached male header away from **J19**. Remember, never overtighten nylon screws. They are much softer than steel.

2.5 Testing the Completed RoboBoard

After you have completed and installed the LCD display, you are ready to test all of the features of your RoboBoard. It is better to ensure that all of your ports work now, rather than to discover later that they are faulty. We have written several programs to do this. One is called *testports.lis.* To download, or install, this program perform the following steps.

1. Make sure that the board has been initialized and that p-code has been downloaded to the board. If you do not know how to do this ask a lab assistant or see the Class Notes on the web page.
2. Start IC.
3. Type `load testports.lis` at the IC prompt. This will load the testports program into the RoboBoard.
4. Using *testports*, check every port on your RoboBoard to see if it is working properly. If this is not the case, consult a lab assistant.

Testports provides instructions for use on the RoboBoard's LCD. If you have any questions about how to use the program, please ask a lab assistant. The program QT.c is another test program that is particularly useful for testing all the motor ports quickly.

2.6 Mounting the RoboBoard

The RoboBoard can be LEGOized by using the threaded axles and a long LEGO beam. LEGOizing the RoboBoard could make it easier to mount to your robot. Figure 16 shows a RoboBoard mounted to a 16-Long LEGO beam. The threaded axles are slipped through the mounting holes on the RoboBoard and one of the holes in the LEGO beam. The threaded axles are then secured using two axle nuts. The side of the board with the LCD display can be mounted on LEGO by using one threaded axle and a 1 inch nylon screw. If you wish to do this, ask your lab assistant for a 1 inch nylon screw.

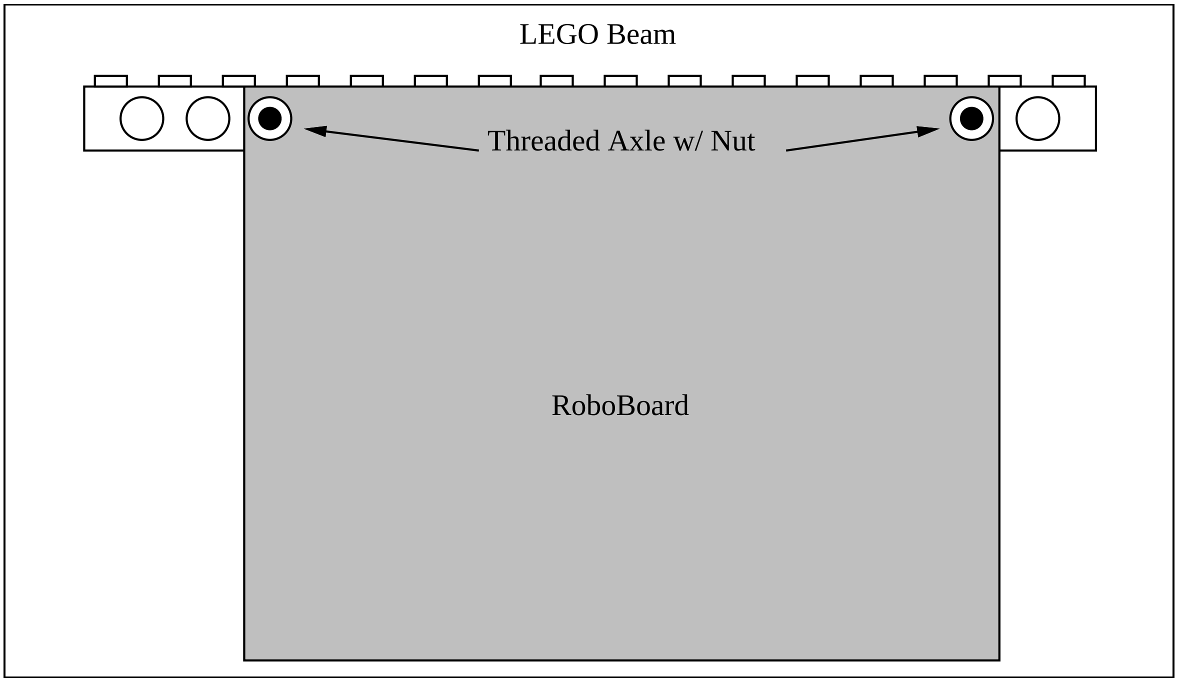

Figure 16: Mounting the RoboBoard on LEGO

2.7 Debugging a Non-working RoboBoard

It is possible that after your best efforts, your RoboBoard is non-functional. This section provides some help to isolating and correcting the cause of the failure. A good working knowledge of the logical design of the RoboBoard and of the function of the various chips is invaluable in finding a fault, so read the chapter of the notes on the ELEC 201 hardware carefully. The steps below will give you additional insight on the operation of the RoboBoard and its components, so read the entire section before starting. Also invaluable is a working RoboBoard, which can provide a comparison for measurements, but most importantly, provides a source of *known-working* chips for substitution. The donor RoboBoard should be loaded with p-code and optionally with a test program like testports. The two different cases of a board that has never worked, and a previously operating board that has failed, are treated separately below. However, there are a few common checks that should be done in any case.

- First, put out the fire. Just kidding, although it has happened. If nothing shows on the display, first check the contrast adjustment pot; also check its value, as it (10KΩ) and the roboknob (100KΩ) are sometimes interchanged.

- More importantly, sometimes a board will act crazy or dead when you turn it on because it is trying to run some (perhaps disfunctional) `main()` program that has been downloaded. It could be waiting for a start light, for example. To avoid this, hold down the CHOOSE button while turning on the power, which bypasses the `main()` function and goes directly to the IC prompt/heartbeat. Also, unfortunately, p-code often gets trashed by an error from a user program and causes problems. Reloading p-code will solve both problems and is a good test. If this doesn't help, or the board is really dead, proceed to the next step.

- Disconnect everything connected to the board.

- Be careful about what you set the board down on. Bench tops are often covered with bits of wire and solder that can short pins. Put the board on a clean piece of paper.

- Measure the battery voltage and polarity.

- Turn on the board and measure the output of the 5 V logic power supply. If the logic supply is low, a bad component or a short is probably drawing too much current. While regulators do fail, they are quite rugged. Note that the regulator will shut off to protect itself if the current draw is too high, in which case the regulator, U14, will be *very* hot, so be careful. A good way to check the regulator is to remove the jumper next to it, J27. This isolates the regulator output from all the components on the board except the filter capacitors C6, C7, and C17. Check them for correct installed polarity and shorts. If the output voltage is still not correct (check right on the pins of the regulator), the regulator is probably bad; replace and recheck. If removing J27 does result in a 5 V output, then the task is to find the component or short that is causing the current draw. Removing all the chips is a good starting point, then proceed, monitoring the current draw at J27, while replacing first the Minimum Functional Chip Set (see below). After that, follow the rest of the steps below as appropriate.

- Measure the output of the 5 V supply, at U13, for the motor chips. If this is low, pull the motor chips and remeasure. If the output is still low or zero, the regulator probably needs to be replaced, as there really is nothing else connected to the output, but check the filter capacitors C18 and C19 as well for correct polarity and shorts.

2.7.1 Failure at Initial Operation

The usual cause of a failure at first startup is an error in construction. So the primary steps involve looking for common errors, and then uncommon errors. First be sure to perform the power supply checks listed above.

- Make sure J27 is in place, or no logic power will get to the board.
- Inspect each chip location carefully to verify that each has the correct chip, that it is oriented properly, and that all pins are actually inserted into the socket (and not bent under the package).
- Check the polarity of all the electrolytic capacitors for correct installation. Installation backwards can cause excess current draw and occasionally sparks, but sometimes no problem will show up for some time.
- Check the polarity of all the zener diodes, both at the motor ports and the data input ports.
- Inspect the soldering for shorts and cold solder joints. In really bad cases, you just have to go over all the joints again and redo them.
- Trim off excess lead length on the solder side of the board with diagonal cutters.

If none of the above helps, go on to the next steps.

2.7.2 Failure of a Previously Working Board

If the source of a problem is not obvious from such information as voltage checks and the nature of the problem, one technique is to pull out chips to get to the minimum working configuration, and then to replace chips sequentially with known-good ones from a working donor board.

The Minimum Functional Chip Set will run the IC logo and heartbeat on the display. The set consists of:

1. The processor, U3, 68HC11.
2. The memory, U2, DS1230Y *loaded with p-code* from a donor board.
3. The 74HC373, U1, memory address latch.
4. The PAL16L8 chip, U19, to provide memory enable on pin 19.
5. The MAX 704, U5, for the reset function on power application.

6. The 74HC245, U15, which senses the CHOOSE button.
7. The LCD, which supplies a status return to the processor.

- Pull all the chips except the Minimum Functional Set above. Use a known-good LCD, and check the contrast pot adjustment.
- Turn on the power; the display should show the IC heartbeat, indicating all the chips are good.
- If "Halted" appears on the display, replace the 74HC245 and try again.
- If nothing appears on the display, start substituting chips from a working donor board. My order is processor, 74HC373, PAL, and MAX 704 (rarely a problem). When you find the problem you can back substitute to verify the bad chip. If you substitute *all* of the Minimum Functional Chip Set from a working donor board, and the voltages are correct, and the board still does not work, you have a big problem: a board fault or a bad on-board component. Fortunately, this is rare, especially if the board has worked previously. Check the solder joints first, and then everything else carefully. Call for help.
- Assuming the board now works minimally, add the next functionality, downloading and digital inputs, by adding the MAX 220, which rarely fails.
- You should now be able to download p-code, robo.lis, and use IC to beep(), etc. Note, if there is no beep, it could be the beeper (jumper in another) or Q4, the driver transistor (check the base drive and output), or a missing jumper J26.
- Digital inputs #0, 1, and 30 should work as they go directly to the processor. Check them from the keyboard; if there is a problem, replace the processor.
- Check digital input ports #2–7. These make use of additional capability of the 74HC245, so if there is a problem, replace it.
- Add *both* the 74HC4051 chips to add the analog input function. U9 supports ports #20–27, and U10 supports ports #28 and 29 plus the roboknob, dip switches, and battery voltage measuring functions. Check all and replace the responsible chip as necessary. Now you can load the program testports.lis for more testing.
- Now add U7, the 74HC574 chip to enable the LED output drivers. Pins 12 and 13 drive the 2N4401 driving transistors, which do burn out if abused. Check the drive at the transistor bases to see if the problem is U7 or the transistor.
- Add IR beacon functionality by installing the 74HC390 chip. This is not necessary for board operation but it checks additional PAL chip functionalities (pin 12). The driver transistor Q3 could also be bad.

- Provide motor port function. First double check the zener diode orientations, then add the three 74HC574 latches; use known-good chips. Add the motor driver chips. Check all the motor ports, both forward and reverse. If nothing works, the problem could be the PAL, which provides enable signals out of pins 16, 17, and 18. Replace the motor driver chips as necessary, and give thanks that they are in sockets in this version of the board. If everything works, you can go back and check the original latches.

- Finally, check the LCD display. Put in the original one, and recheck the contrast adjustment pot. The processor will halt if it does not receive a status signal from the display. Check that the pins *really* insert fully into the socket on the board—a problem if the longer male header was not used. The LCD driver chips are on the LCD board, so replacement of the display is the only option if it is still not working.

And that is all there is, so the RoboBoard should be fully functional at this point. In some cases, and with practice, you can skip some of the steps above if the problem is obviously limited to a particular area, like the motor drivers (a common problem). In that case, just pull the driver chips and the 574 latches and see if everything else works OK. If it does, just substitute latches and motor drivers. For other problems, jump into the sequence above at the appropriate point.

Finally, please destroy and discard all chips found to be bad. Then fill up your donor RoboBoard with new chips as necessary, load it with p-code and testports, and test it, so you will be ready for the next problem.

3 The Battery and Charger System

The ELEC 201 Robot Controller system uses a Panasonic 12 volt lead-acid battery rated for 1.3 ampere-hours of operation. These batteries are rechargeable. The battery is connected to the RoboBoard through the power receptacle J20, just below the power switch.

Lead-acid cells are extremely powerful devices. Car batteries are constructed of similar lead-acid technology. *When handling the batteries, be extremely careful not to short the (+) and (−) terminals of the battery together.* A huge surge of current will flow, melting the wire and causing burns. In extreme cases, batteries can explode and cause serious injury.

3.1 The Battery Charger

The Version 4.1 battery charger is designed to charge one 12 Volt DC battery at a constant charging rate until the battery is fully charged. When the battery is fully charged, the battery charger automatically shuts off. This automatic shut-off feature means that the battery can be left on the charger overnight or longer without risk. A red LED on the board lights up brightly when the battery is charged. You will receive an assembled battery charger board and 12 VDC adapter, but you may need to repair or replace cables. (If you need to build a new charger you will receive additional parts and the assembly instructions.)

3.2 Wiring the Battery Cable

Ask a lab assistant for a battery and cable if you do not already have one. Your battery cable is pre-assembled, but in case you have to repair it, or make a new one, here are instructions.

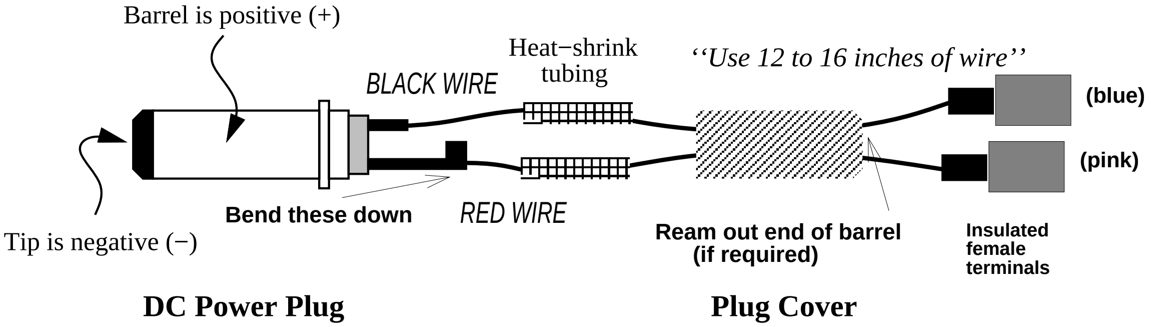

Figure 17: Battery Plug and Cable Wiring Diagram

Figure 17 illustrates how to wire the battery plug and cable assembly.

○ Cut a 12" to 16" length of the black/red twin cable for use in making the battery cable. Strip and tin the wire ends at both ends of the cable. One end should have about $\frac{1}{4}$ inch of exposed tinned wire; the other end (which will be connected to the power plug) should have about $\frac{1}{4}$ inch exposed. At this end the red wire should be about $\frac{1}{16}$ inch shorter than the black wire.

○ If your power jack has a hexagonal barrel and a flexible sleeve there is no need to ream the jack. Proceed to the next step. Otherwise, use the reaming tool borrowed from the lab assistant, enlarge the small end of the black annular connector cover. Return the tool to the lab assistant. Slide the cover, reamed end first, onto the tinned leads. Push back about six inches.

○ Heat-shrink tubing is used on both terminals of the DC power plug. The tubing acts as an insulator to minimize the likelihood of an electrical short at the plug terminals. *It is essential that this wiring be performed carefully because a short in the power plug will short out the battery terminals and create a serious hazard.* Place a $\frac{3}{8}$ inch long piece of heat shrink tubing on each wire before soldering to the power plug. Be sure to keep the heat shrink tubing well away from the point of soldering or it will shrink prematurely.

○ **Proper polarity is important.** The use of red wire to signify the (+) terminal and black wire to signify the (−) terminal is an international standard. Using the "helping hands", position the black wire under the short terminal and solder. Then position the red wire inside the long terminal and bend down the tabs to ensure a good mechanical connection. Solder.

○ After soldering, slide the heat-shrink tubing down over both terminals and shrink.

○ Screw the plug cover onto the plug.

○ Place the pink shielded terminal on the $\frac{1}{4}$ inch of exposed tinned wire of the red lead. Place the blue shielded terminal on the $\frac{1}{4}$ inch of exposed tinned wire of the black lead. Use the crimping tool (borrowed from a lab assistant) to crimp the terminals onto the wire. Use the "red dot" location on the crimping tool, and squeeze the tool until it releases (this takes *a lot* of pressure). *Be sure to return the tool when you are finished!*

○ *Before installing the cable onto a battery, use an ohmmeter to make absolutely sure that the cable is not shorted.* The cable should measure open circuit or infinite resistance. If a conductor is placed across the terminals of lead-acid batteries, a huge surge of current will flow, melting the wire responsible for the short circuit and possibly causing the battery to explode.

○ **Make sure to connect the red wire (pink terminal) to the + terminal of the battery.**

3.3 Wiring the 12 VDC Adaptor

You should have a pre-assembled 12 VDC adaptor. Nevertheless, before using your adaptor, use a voltmeter to make absolutely sure that the connections are correct. Plug in the adaptor to a wall outlet, and and set your multimeter to the DC voltage postion. Place the black multimeter probe inside the barrel of the power connector, and the red probe on the outside of the barrel. You should read **positive** voltage. If not, ask a lab assistant to verify the measurement, and then you will need to put on a new plug, following the directions below.

Figure 18 illustrates how to wire the 12 volt DC adaptor and cable assembly. **NOTE: The adaptor may come out of the box with a plug that looks correct, but it probably has opposite polarity from that required.**

○ Examine the 12 Volt DC power adapter. Clip off and discard the existing power plug near the base of the plug (this may have already been done for you). Strip

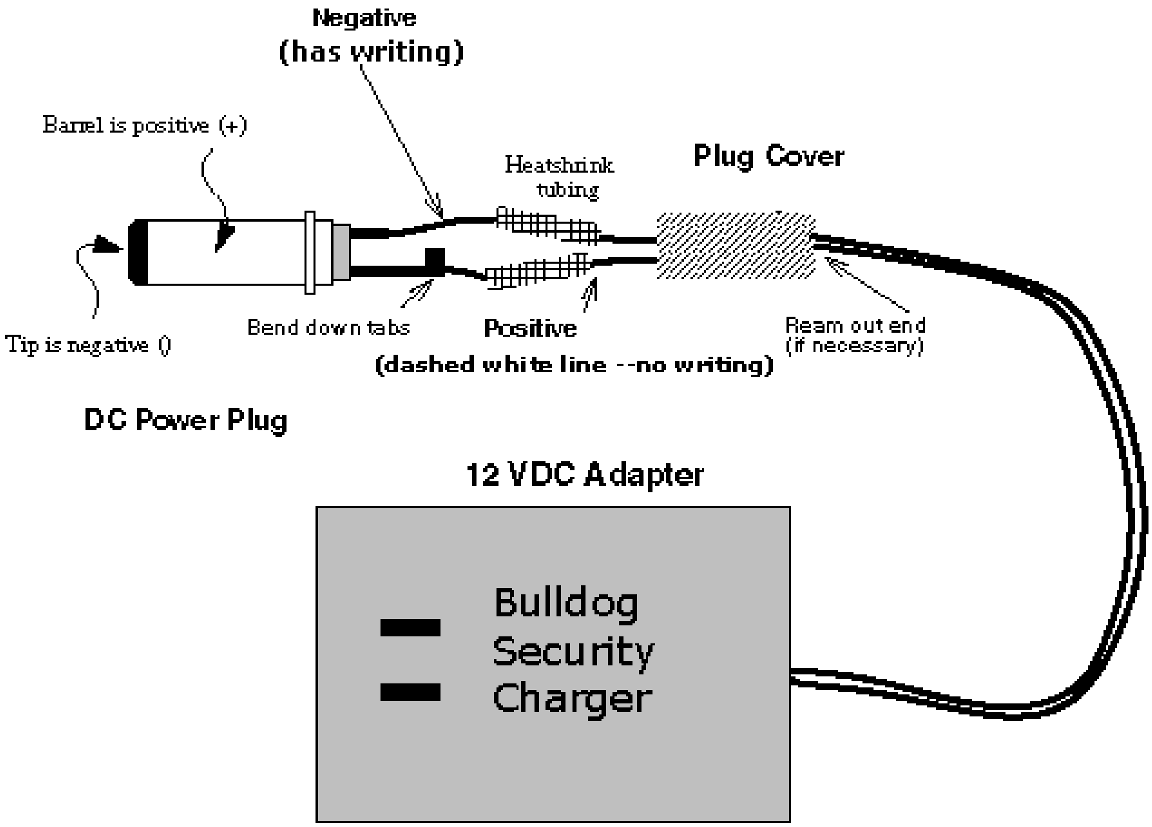
Negative
(has writing)
Barrel is positive (+)
Heatshrink tubing
Plug Cover
Tip is negative ()
Bend down tabs
Positive
(dashed white line --no writing)
Ream out end
(if necessary)
DC Power Plug
12 VDC Adapter
Bulldog
Security
Charger

Figure 18: 12 VDC Adaptor Wiring Diagram

$\frac{1}{8}$ inches of insulation from power wires and tin. At this end the non-ribbed wire should be about $\frac{1}{16}$ inch shorter than the black wire.

○ If your female power jack has a hexagonal barrel and a flexible sleeve there is no need to ream the jack. Proceed to the next step. Otherwise, use the reaming tool borrowed from the lab assistant, enlarge the small end of the black annular connector cover. Return the tool to the lab assistant. Slide the cover, reamed end first, onto the tinned leads. Push back about six inches.

○ Heat-shrink tubing is used on both terminals of the DC power plug. The tubing acts as an insulator to minimize the likelihood of an electrical short at the plug terminals. *It is essential that this wiring be performed carefully because a short in the power plug will short out the battery terminals and create a serious hazard.* Place a $\frac{3}{8}$ inch long piece of heat shrink tubing on each wire before soldering to the power plug. Be sure to keep the heat shrink tubing well away from the point of soldering or it will shrink prematurely.

○ We use different models of adapters and the wire markings vary. **Use your volt meter to determine which lead is Positive and which is Negative,** and make a note on the figure, *which may label the wires incorrectly!*

Using the “helping hands”, position the **Negative** wire under the short terminal and solder. Then position the **Positive** wire inside the long terminal and bend down the tabs to ensure a good mechanical connection. Solder. *It is essential that this wiring be performed carefully because a short in the power plug will short out the DC adaptor.*

○ After soldering, slide the heat-shrink tubing down over both terminals and shrink.

○ Screw the plug cover onto the plug.

○ *Before using the modified adaptor, use a voltmeter to make absolutely sure that the connections are correct.* Plug in the adaptor to a wall outlet and place the black voltmeter probe inside the barrel of the power connector, and the red lead on the outside of the barrel. You should read **positive** voltage.

4 The Infrared Beacon

The infrared (IR) transmitter board emits modulated infrared light that can be detected by the Sharp GP1U52X IR sensors. The IR board features eight infrared emitting diodes positioned in a circle to generate a 360 degree signal. The diodes are driven by an oscillator circuit (the 74HC390 chip) and a transistor amplifier on the RoboBoard.

The infrared LEDs are wired in series to minimize the current draw from the battery. The board also contains a visible LED, that when lighted, indicates all IR LEDs are transmitting. An unlit LED indicates problems with the board. You will receive an assembled IR transmitter, but you may need to add, repair, or replace the cable and connector. (If you need to build a new IR beacon you will receive additional parts and assembly instructions.)

4.1 Cable and Connector

Use two strands from the rainbow colored ribbon wire; the actual color of the wires is not important. Just pick one color for the positive connection, and the other one for the negative connection. In these instructions we will just call them *pos-color* and *neg-color*, respectively.

○ Cut the 2-conductor piece of ribbon cable to the necessary length for your robot; longer is better than too short. Strip $\frac{1}{4}$" of insulation from the wire on both ends and tin.

○ From *underside* of IR board, insert the pos-color wire into the hole marked (+) and the neg-color wire into the hole marked (−). Solder from the *top* of the board.

○ Mount other end of the wire to outside pins of a four-pin male connector with one pin removed (see Figure 19). The pos-color wire goes to the isolated end-pin (no adjacent pin). Use the guidelines shown in Section 5. Be sure to place heat-shrink tubing on the wires, well away from the connection, *before* soldering. After soldering, slide the heat-shrink tubing down over the pins to make a neat connection. Then apply heat.

The infrared transmitter plugs into the connector labeled IR XMIT on the RoboBoard (see Figure 13), with the positive lead inserted into the end marked (+) on the board.

5 Cable and Connector Wiring

This section explains how to build reliable cables and connectors for the motors and sensors that will plug into the robot's controller board. It also gives instructions for building a cable to connect the RoboBoard to a PC.

Sturdy and reliable connectors are critical to the success of a robot. If a robot's connectors are built sloppily, hardware problems will occur. Well-built connectors will help make the robot more reliable overall, and will ease development difficulties.

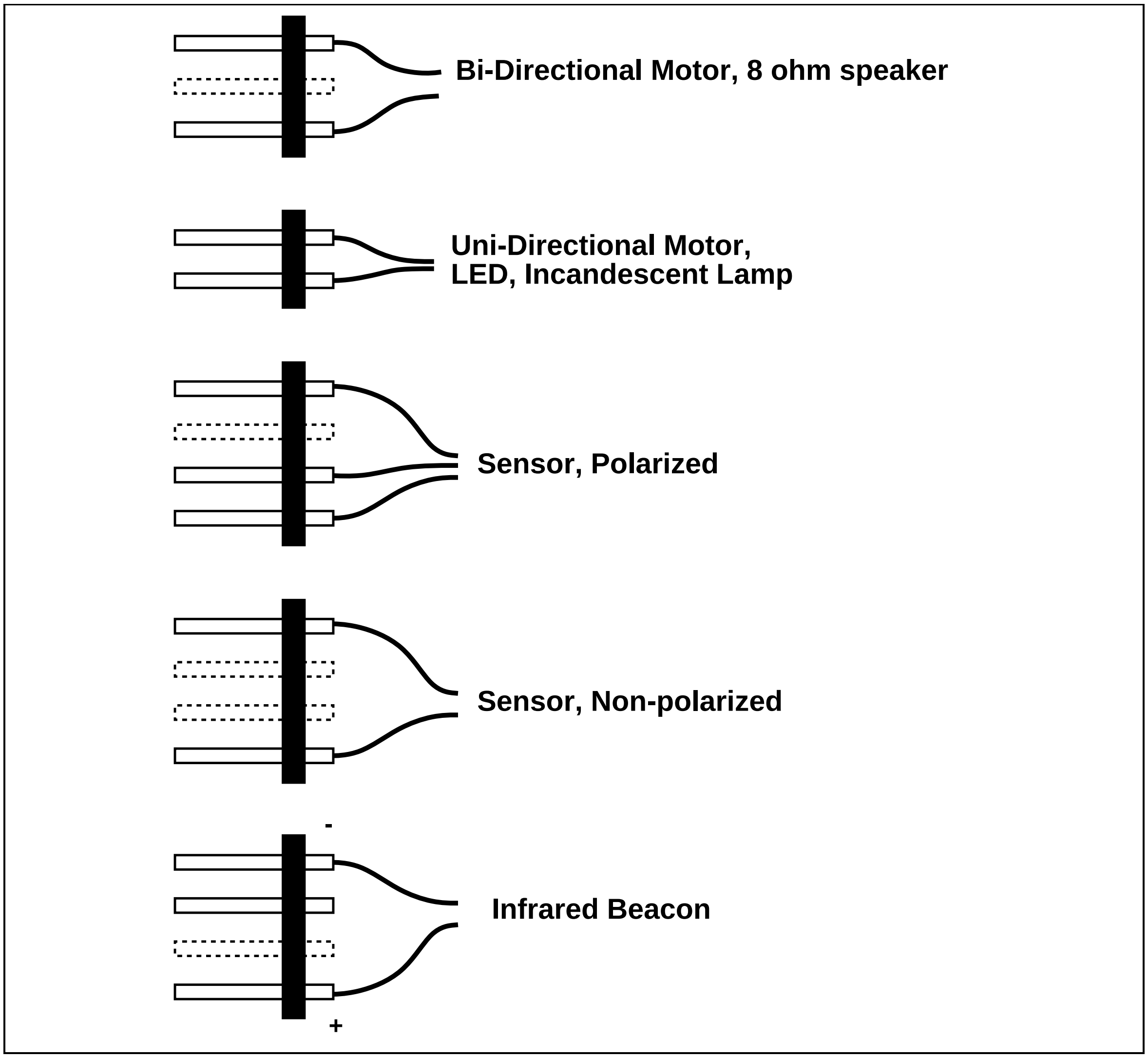
Bi-Directional Motor, 8 ohm speaker
Uni-Directional Motor,
LED, Incandescent Lamp
Sensor, Polarized
Sensor, Non-polarized
-
Infrared Beacon
+

Figure 19: Standard Connector Plug Configurations

Sensors and motors are built with fixed wiring; that is, a sensor or motor will have a fixed length of wire terminating in a connector. It is possible to build extension cables, but it is more time-efficient to build cables that are the proper length initially.

The average robot has its control electronics near the physical center of the robot; hence, motors and sensor cables need to reach from the center of the robot to their mounting position. Given this geometry, most robots will need sensor and motor cables between 6 and 12 inches long.

Several different connector styles are used depending on the device that is being connected to the robot. Figure 19 shows the connector configurations used for bidirectional motors, unidirectional motors, sensors, and the infrared beacon.

The ribbon cable provided in the ELEC 201 kit is best for making sensor and motor cables. Use the "helping hands" to hold the connectors while soldering.

Figures 20 through 23 illustrate the recommended method for wiring to a connector plug. When assembled properly, this method will provide for a sturdy, well-insulated connector that will be reliable over a long period of use.

Be sure to place heat-shrink tubing on the wires, well away from the connection, *before* soldering. After soldering, slide the heat-shrink tubing down over the pins to make a neat connection. Then apply heat.

The example shows wiring to opposite ends of a three-pin plug, as would commonly be used when wiring to a motor. The method, however, is suitable for all kinds of connectors.

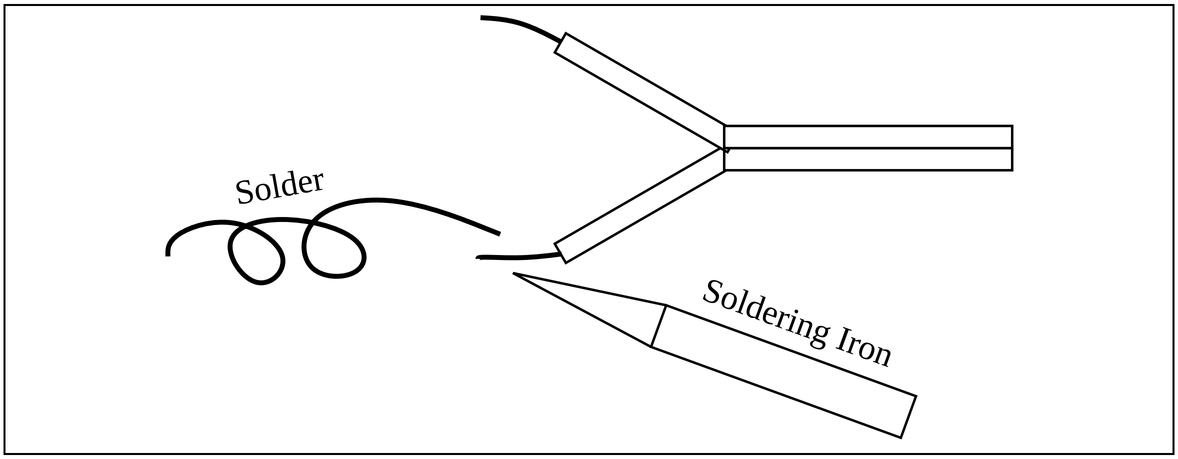

Strip a small amount of insulation off the wire ends. Tin the wire ends by applying a thin coat of solder to them.

Figure 20: Step One of Connector Wiring

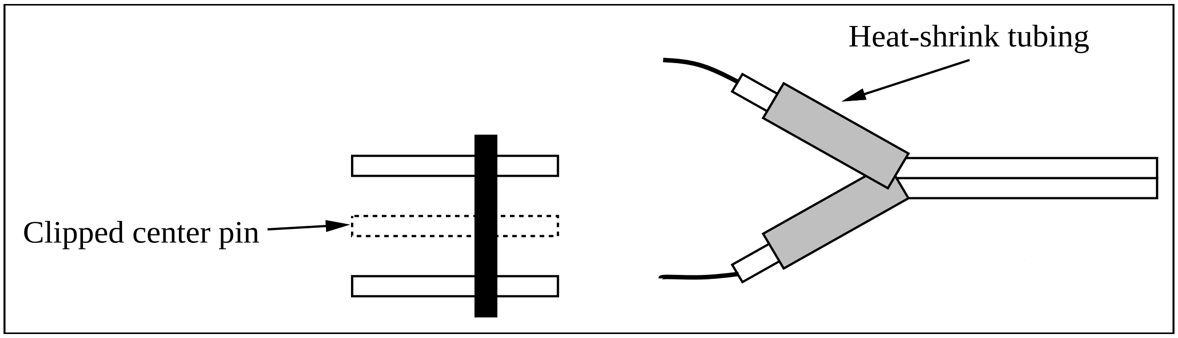
Heat-shrink tubing
Clipped center pin

Cut the male connector to size. This example shows a plug that can be used to wire a motor or the infrared transmitter. Cut $\frac{1}{2}$ inch length pieces of heat-shrink tubing, and slide over tinned wires, well away from the point of soldering.

Figure 21: Step Two of Connector Wiring

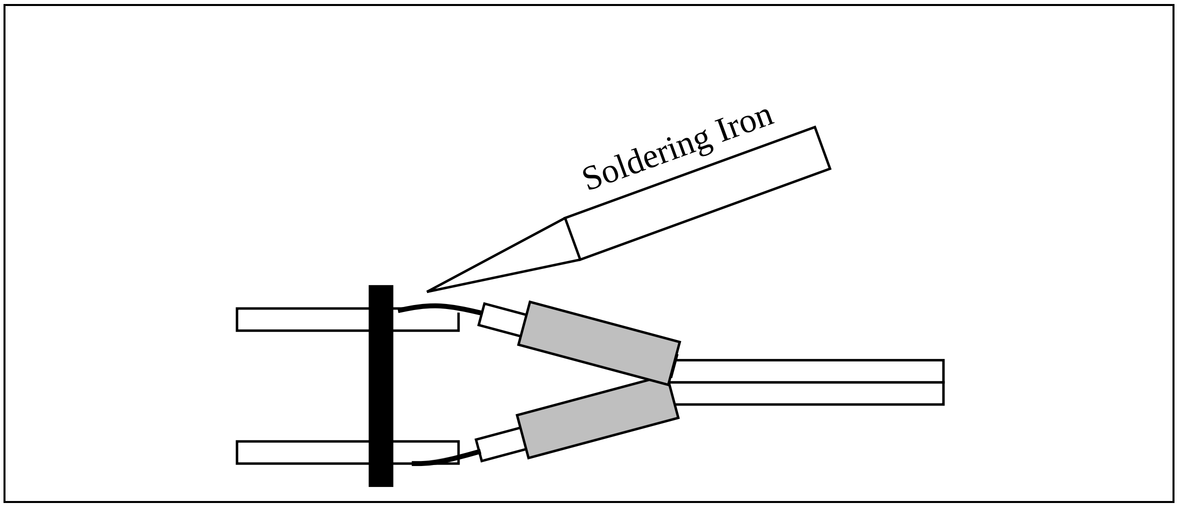
Soldering Iron

Using the “helping hands”, hold the connector and wires in place and solder. Clip wires to a little less than the length of the male pins before soldering.

Figure 22: Step Three of Connector Wiring

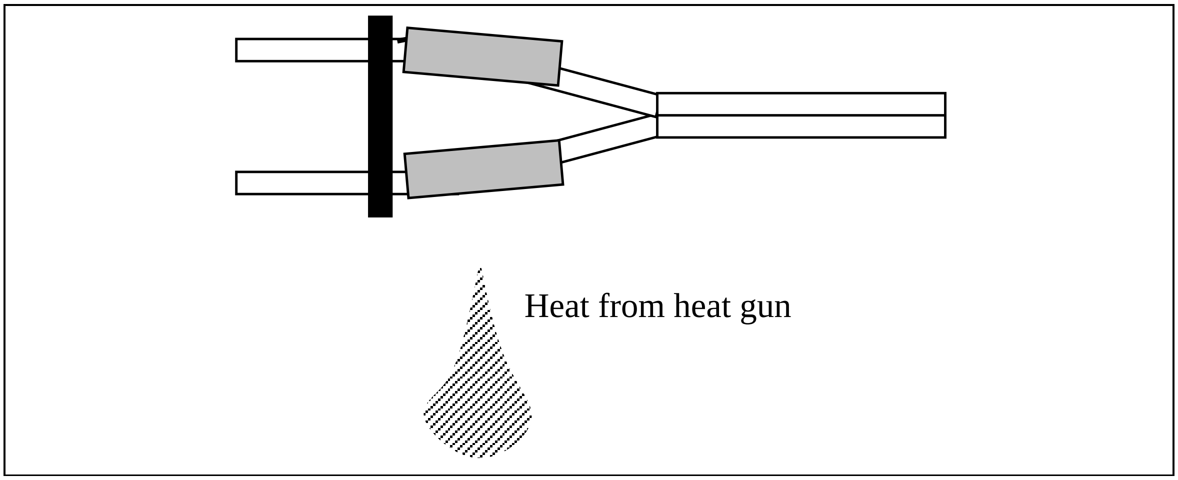

Slide pieces of heat-shrink tubing over connections. Shrink using heat gun.

Figure 23: Step Four of Connector Wiring

5.1 Computer Cable Wiring

Each of the computers in the ELEC 201 lab has a cable from the serial port to a modular phone plug, known as a RJ-11 plug. You may want to make a similar cable for your own computer. You can obtain the parts from a lab assistant. Refer to Figure 24 for wiring details. New Apple computers require an adapter to provide a standard PC DB-9 serial port.

Rice RoboBoard Connection to PC Host Computer

RoboBoard RJ-11 6-Pin Socket

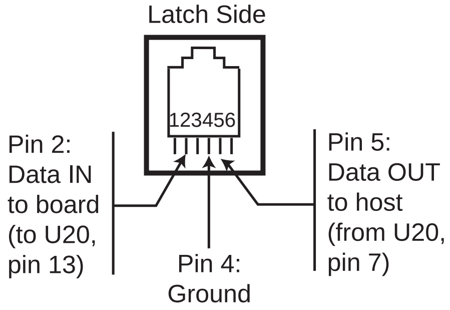
Latch Side
123456
Pin 2: Data IN to board (to U20, pin 13)
Pin 4: Ground
Pin 5: Data OUT to host (from U20, pin 7)

RJ-11 Plug on Phone Cable

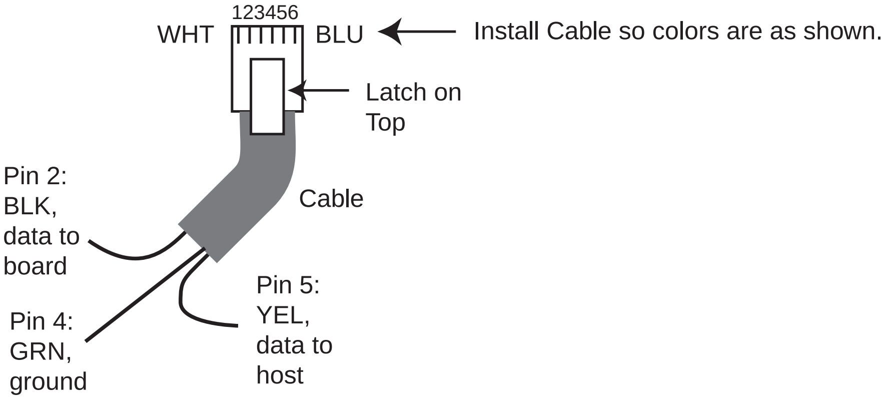
123456
WHT
BLU
Install Cable so colors are as shown.
Latch on Top
Cable
Pin 2: BLK, data to board
Pin 4: GRN, ground
Pin 5: YEL, data to host

IBM PC DB9 Serial Port: (female connector required)

Terminal #2: YEL: Receive data (to host).
Terminal #3: BLK: Transmit data (to board).
Terminal #5: GRN: Signal Ground.

IBM PC DB25 Serial Port: (female connector required)

Terminal #2: BLK: Transmit data (to board).
Terminal #3: YEL: Receive data (to host).
Terminal #7: GRN: Signal Ground.

Macintosh (Old) Serial Port: DIN-8 Round (male connector)

Terminal #3: BLK: Transmit data (to board).
Terminal #4: GRN: Signal Ground.
Terminal #5: YEL: Receive data (to host).

JFY 9/98

Figure 24: Computer Cable Wiring Diagram

6 Sensor Wiring

Sensors are a critical part of your robot. They provide the information that your control program requires to make navigation, action, and strategy decisions during the game. Poorly mounted or wired sensors are prone to breaking, resulting in no, or worse, incorrect, information to the RoboBoard. Bad sensors are usually a major cause of robot failures. The course notes on the class web site shows many mounting ideas, plus you can look at the robots on display in the lab for ideas (good and bad). However, feel free to experiment and come up with new ideas. Pay particular attention to ruggedness and reliability.

This section shows wiring diagrams for the following sensors:

- Touch sensors
- Mercury Tilt Switch
- Photocell light sensor
- Infrared reflectance sensor
- Sharp infrared sensor
- Potentiometer position sensors
- slotted optical switch
- Magnetic Switches

In most cases, some discussion of the sensor's principle of operation accompanies the wiring diagram. Further information on the use of all sensors is found in the Class Notes. Examples of correctly wired sensors are available for your inspection. As a general rule, do not wire a sensor until it is needed. This will conserve material, and will result in cable lengths appropriate for your robot.

6.1 Touch Sensors

Several switches in the ELEC 201 kit may be used to make touch sensors. Diagrams for the white "ALPS" switch, the small red push-button switch, the Burgess lever switch, the lever switch, the mock roller switch, and the roller switch are shown.

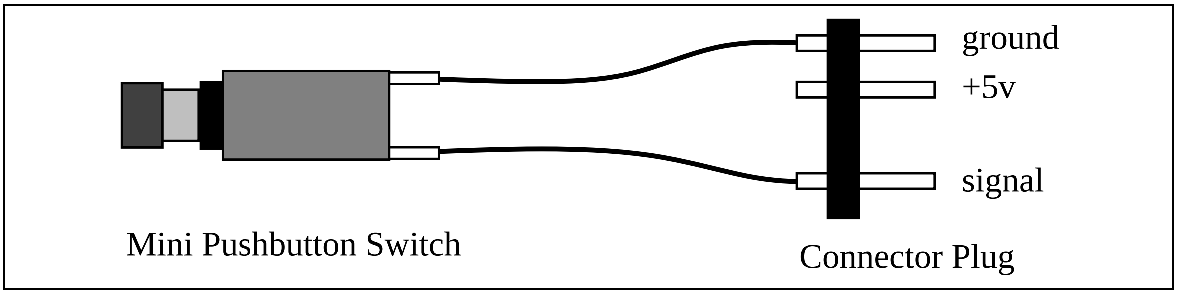
ground
+5v
signal
Mini Pushbutton Switch
Connector Plug

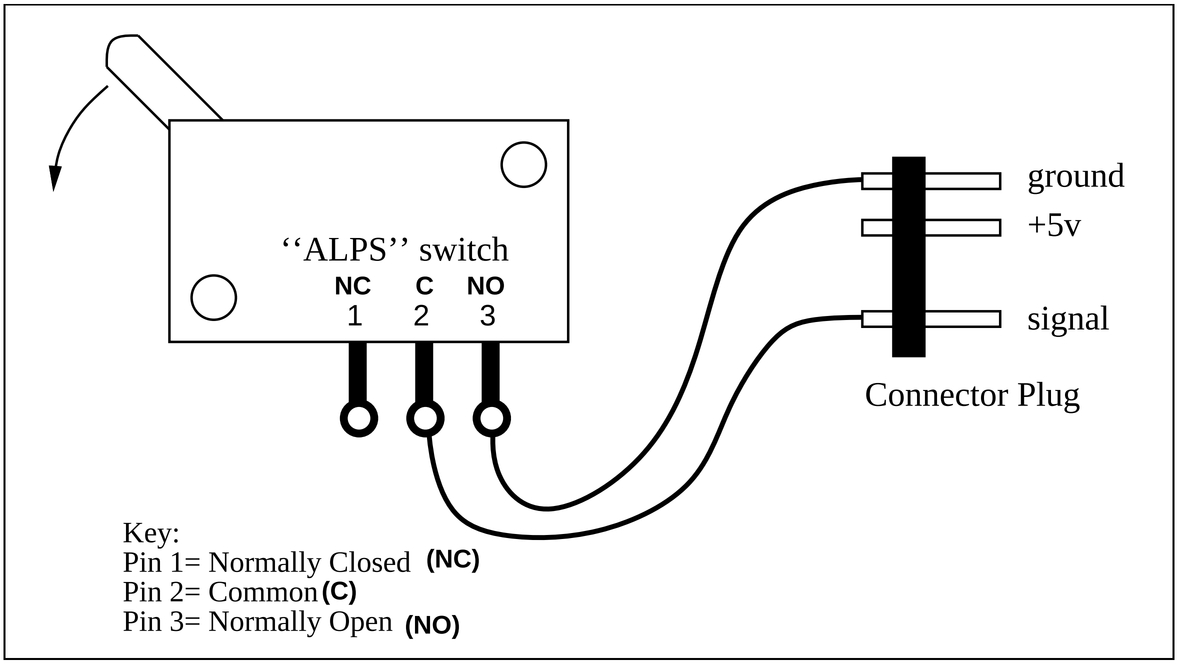
''ALPS'' switch
NC C NO
1 2 3
ground
+5v
signal
Connector Plug
Key:
Pin 1= Normally Closed (NC)
Pin 2= Common (C)
Pin 3= Normally Open (NO)

Touch switches should be wired in a normally open configuration, so that the signal line is brought to ground only when the switch is depressed.

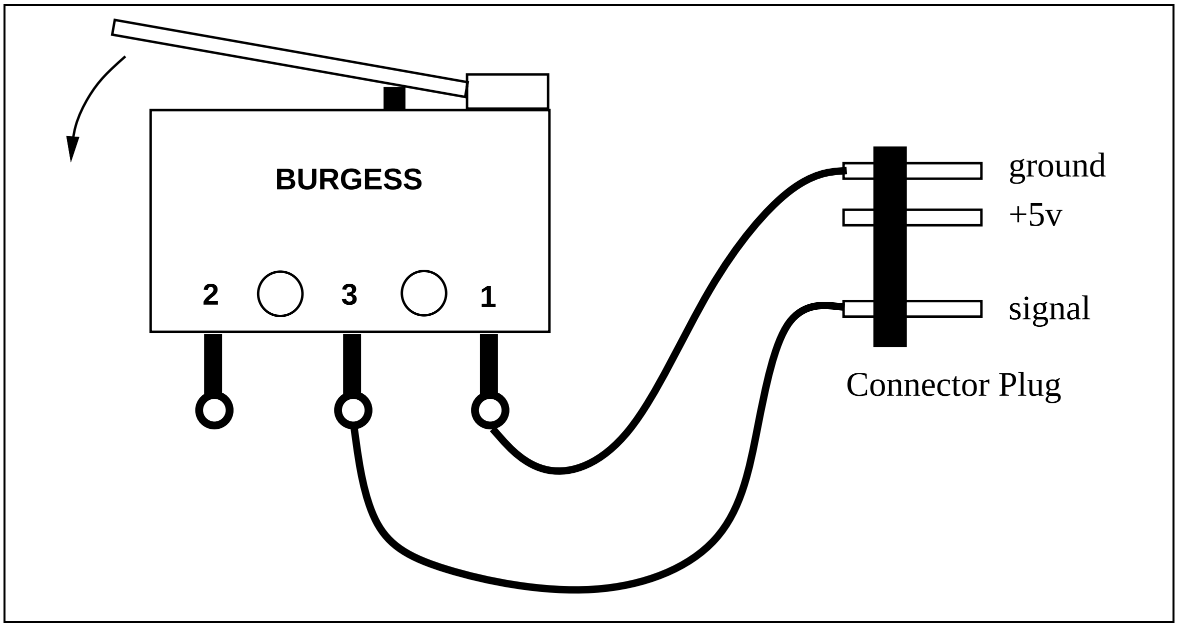
BURGESS
2
3
1
ground
+5v
signal
Connector Plug

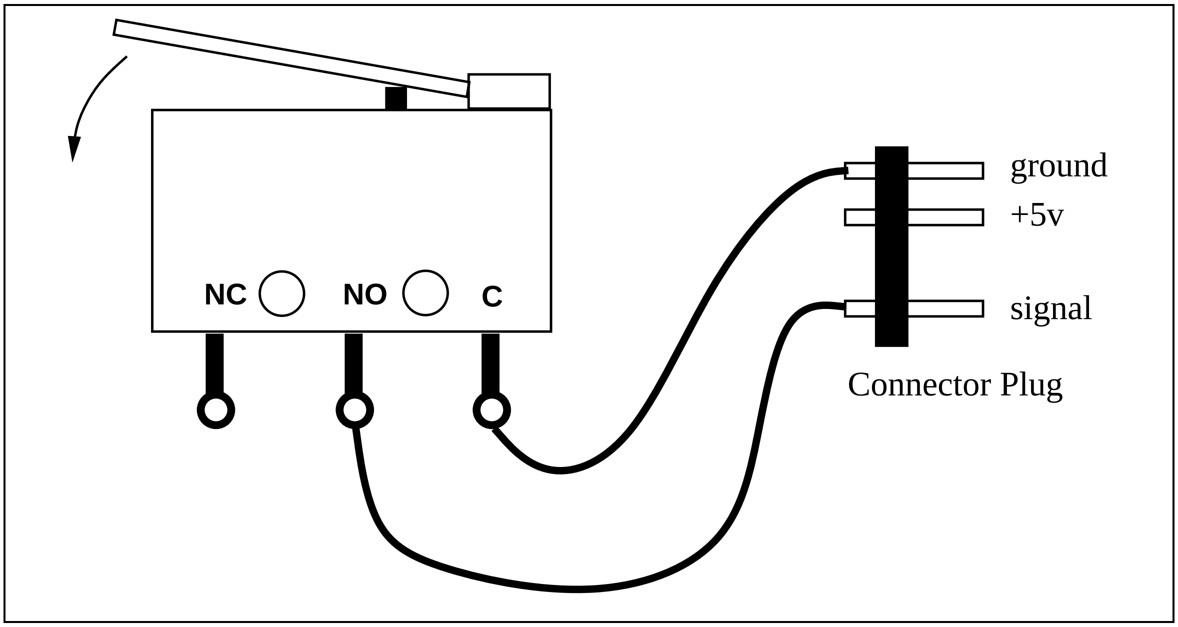
NC
NO
C
ground
+5v
signal
Connector Plug

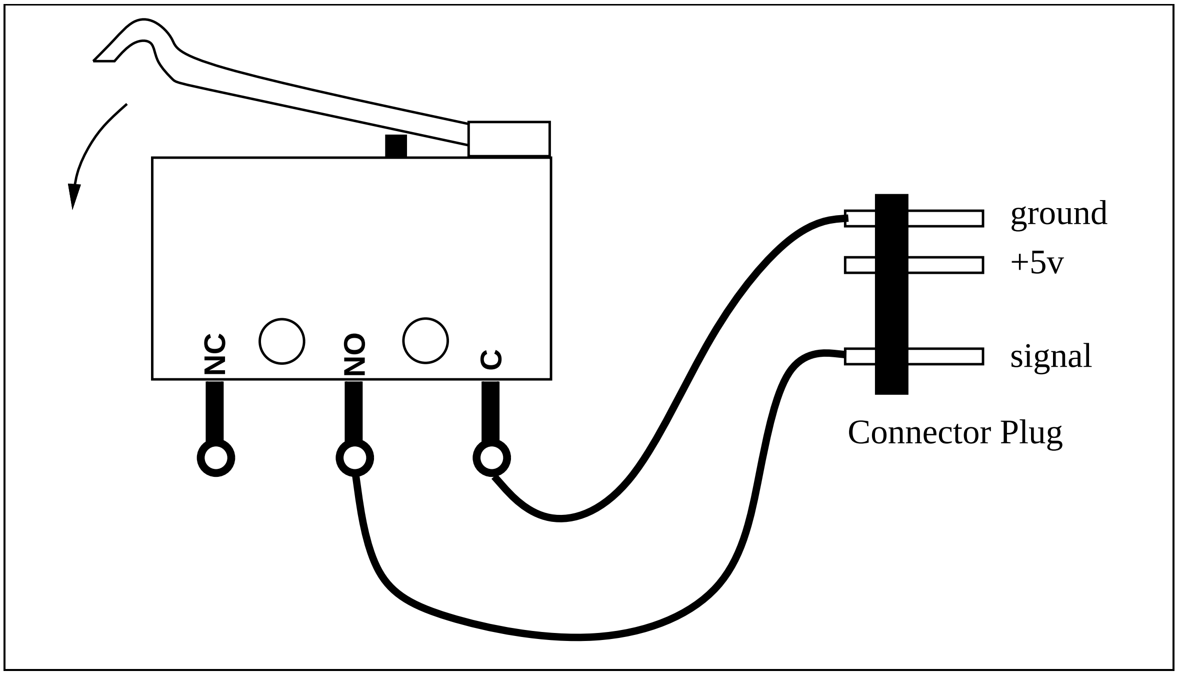
NC
NO
C
ground
+5v
signal
Connector Plug

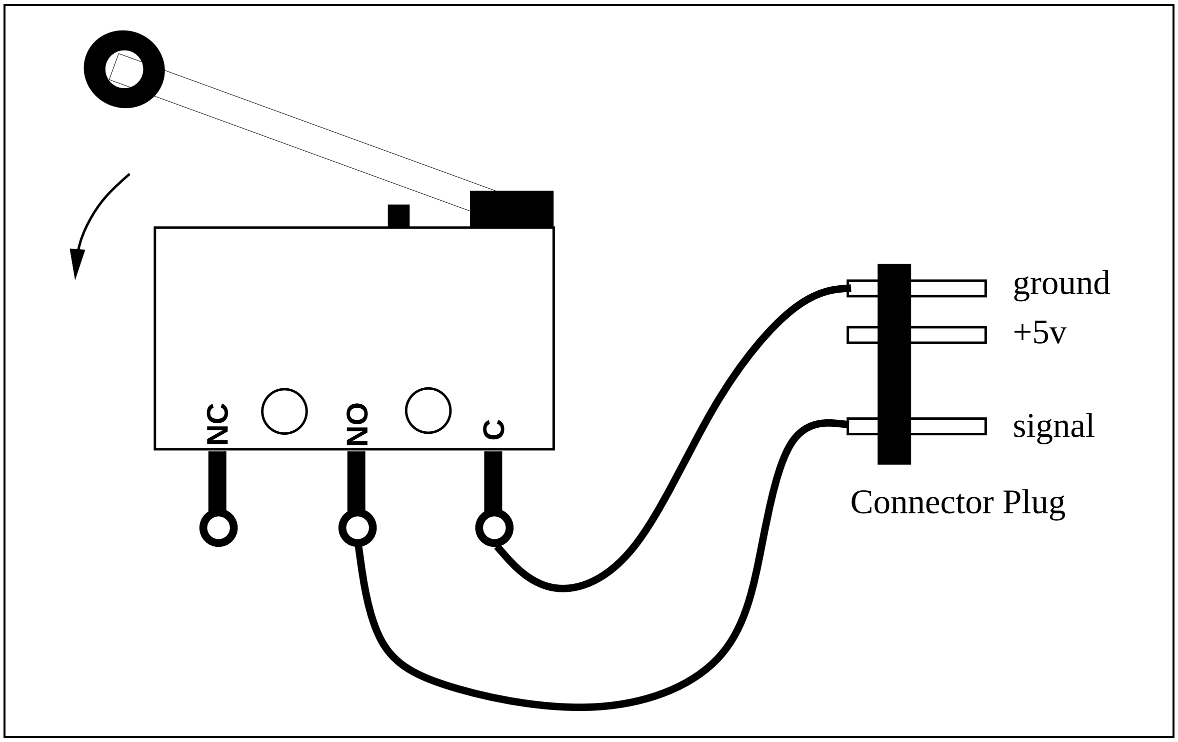
NC
NO
C
ground
+5v
signal
Connector Plug

6.2 Photocell Light Sensor

The photocell is a special type of resistor whose resistance varies in proportion to the impinging light. The more light hitting the photocell, the lower its resistance.

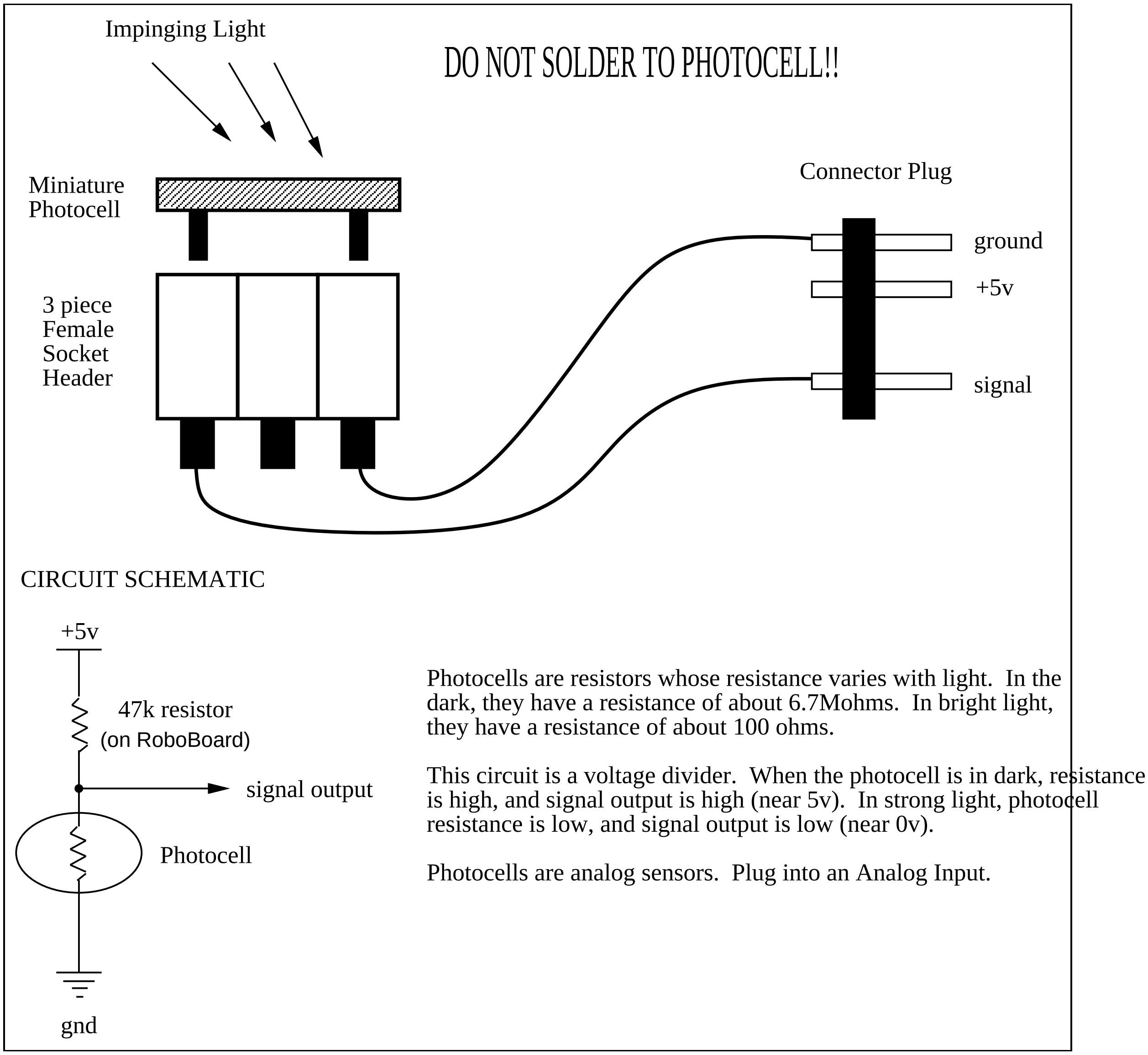
Impinging Light
DO NOT SOLDER TO PHOTOCELL!!
Miniature Photocell
3 piece Female Socket Header
Connector Plug
ground
+5v
signal
CIRCUIT SCHEMATIC
+5v
47k resistor
(on RoboBoard)
signal output
Photocell
gnd

Photocells are resistors whose resistance varies with light. In the dark, they have a resistance of about 6.7Mohms. In bright light, they have a resistance of about 100 ohms.

This circuit is a voltage divider. When the photocell is in dark, resistance is high, and signal output is high (near 5v). In strong light, photocell resistance is low, and signal output is low (near 0v).

Photocells are analog sensors. Plug into an Analog Input.

The output signal of the photocell is n analog voltage corresponding to the amount of light hitting the cell. Higher values correspond to less light.

6.3 Mercury Tilt Switch

Assemble the Mercury Tilt Switch according to the diagram below. Be sure to apply heat-shrink to both ends.

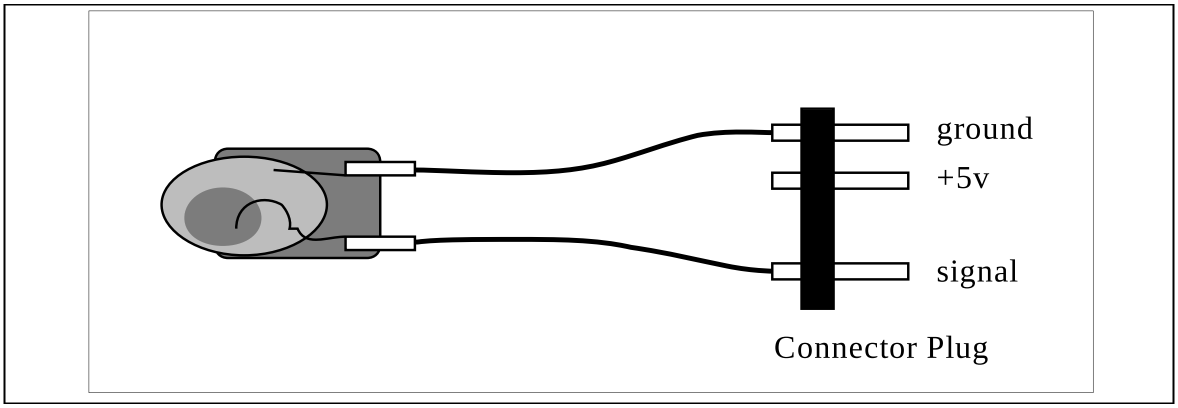
ground
+5v
signal
Connector Plug

6.4 Potentiometer

The potentiometer can be used as a rudimentary rotary position sensor. The first diagram deals with shafted potentiometers. The second deals with slotted potentiometers.

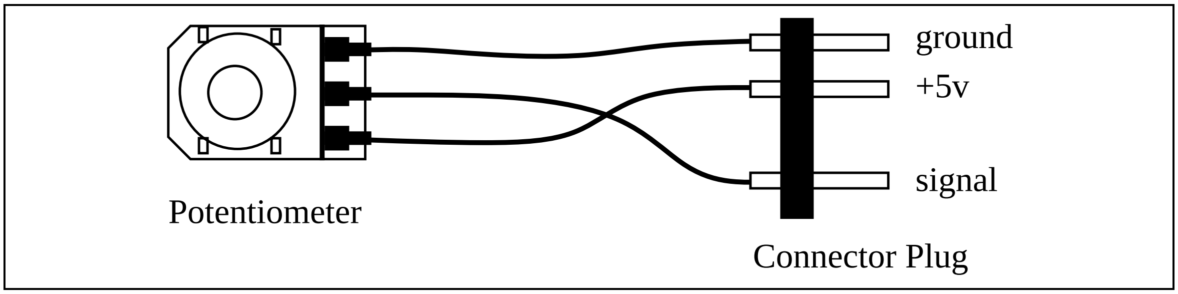
ground
+5v
signal
Potentiometer
Connector Plug

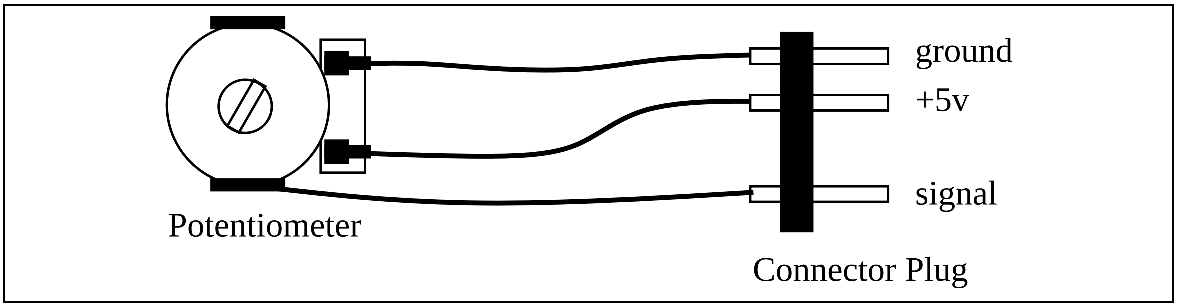
ground
+5v
signal
Potentiometer
Connector Plug

6.5 Infrared Reflectance Sensor

IR reflectance sensors contain a matched infrared transmitter and infrared receiver pair. These devices work by measuring the amount of light that is reflected into the receiver. Because the receiver also responds to ambient light, the device works best when well shielded from ambient light, and when the distance between the sensor and the reflective surface is small (less than 5mm). IR reflectance sensors are used primarily to detect white and black surfaces. As you would imagine, white surfaces generally reflect well, while black surfaces reflect poorly.

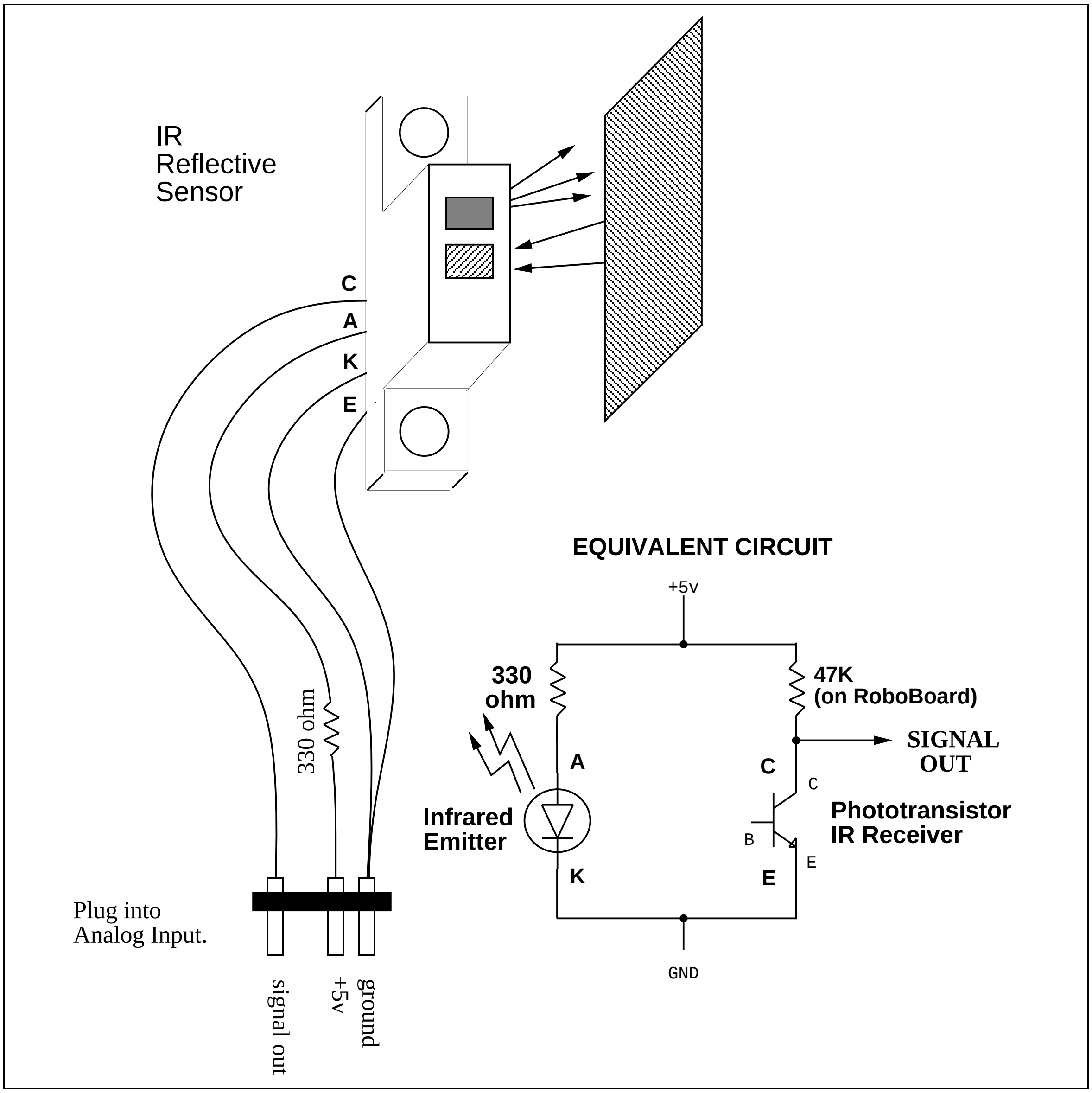
IR Reflective Sensor
C
A
K
E
330 ohm
Plug into Analog Input.
signal out
+5v
ground
EQUIVALENT CIRCUIT
+5v
330 ohm
47K (on RoboBoard)
SIGNAL OUT
A
C
C
Infrared Emitter
Phototransistor IR Receiver
B
K
E
E
GND

The output signal of the phototransistor receiver is an analog voltage corresponding to the amount of light hitting the phototransistor. Higher values correspond to less light, and hence a smaller degree of surface reflectivity.

6.6 Sharp Infrared Sensor

The Sharp GP1U52X sensor detects infrared light modulated (e.g., blinking on and off) at 40,000 Hz. It has an active low digital output, meaning that when it detects the infrared light, its output is zero volts.

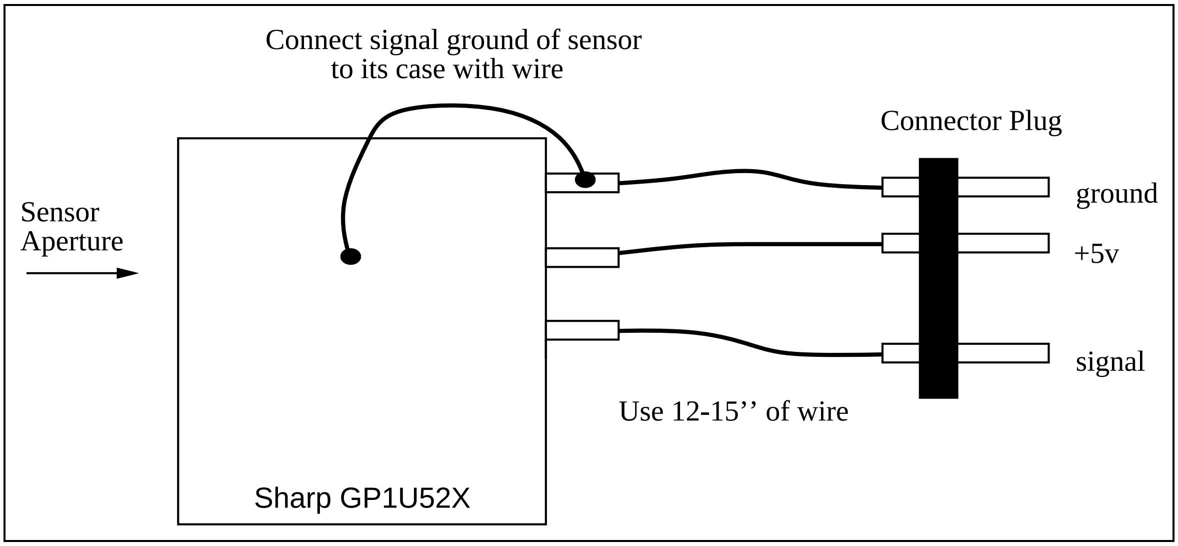
Connect signal ground of sensor
to its case with wire
Connector Plug
Sensor
Aperture
ground
+5v
signal
Use 12-15'' of wire
Sharp GP1U52X

The metal case of the sensor must be wired to circuit ground, as indicated in the diagram. This makes the metal case act as a Faraday cage, protecting the sensor from electromagnetic noise.

An explanation of the infrared system, and use of the Sharp sensor is given in the Class Notes.

6.7 Slotted Optical Switch

Like the IR reflective sensor, the slotted optical switch contains a matched infrared transmitter and infrared receiver. The slotted optical switch also works by measuring the amount of light transmitted from the transmitter to the receiver, but in this case the transmission is direct, rather than reflected. Best results are again achieved when ambient light is shielded from the device. We recommend that you examine the sample before attempting to wire a slotted optical switch.

Slotted Optical Switch Hook-Up

(recommended configuration)

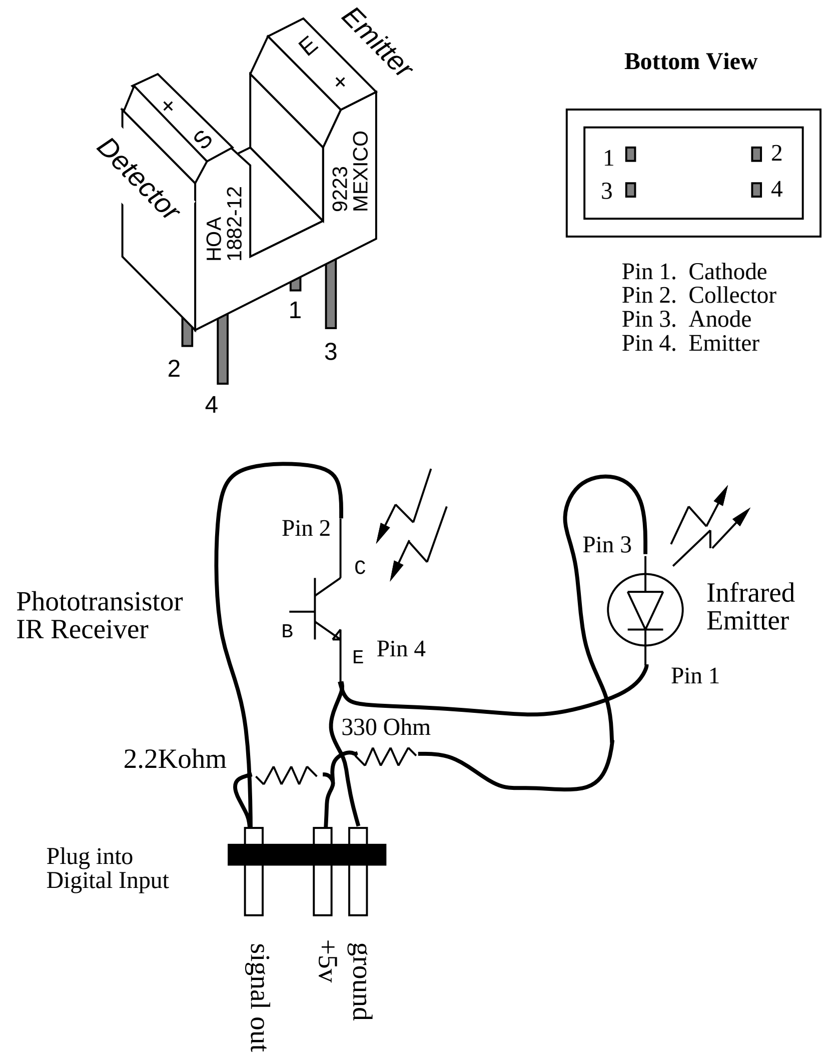
Emitter
E
+
Detector
+
S
HOA
1882-12
9223
MEXICO
1
2
3
4
Bottom View
1
2
3
4
Pin 1. Cathode
Pin 2. Collector
Pin 3. Anode
Pin 4. Emitter
Pin 2
C
B
E
Pin 4
Phototransistor
IR Receiver
Pin 3
Infrared
Emitter
Pin 1
330 Ohm
2.2Kohm
Plug into
Digital Input
signal out
+5v
ground

The output signal of phototransistor element is an analog voltage that corresponds to the amount of light hitting the phototransistor. Higher values indicate less light. Generally, however, the slotted switch is used as a digital sensor, with the two states Blocked and Unblocked.

6.8 Magnetic Switches

These magnetic switches are used as window sensors in home security systems. The switch will close when it comes within 1” of its companion magnet.

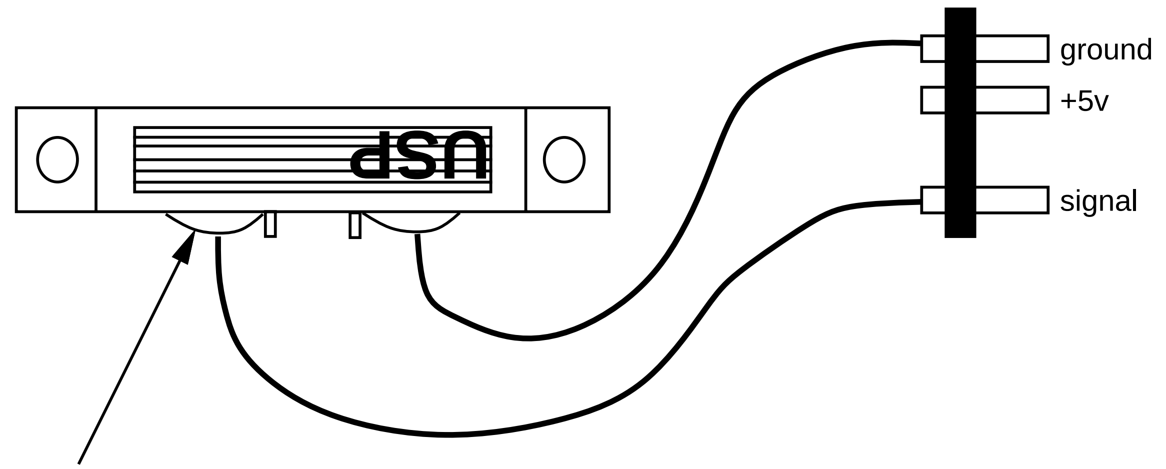

To attach wires to screw terminals, loosen screws, place stripped and tinned wire under screws and tighten the screws. No soldering is required.

7 Motor Mounting and Wiring

Several types of motors are available, but they all are either round, or oblate (round with two flat sides). See the course notes on the class web site for the different motor’s characteristics and for instructions on how to mount them to LEGO pieces. It is important to get the motor aligned correctly so that the gears will mesh properly, with minimum friction.

In addition to mounting the motor on a LEGO piece, you must also attach a gear to the shaft. You will find some motors in the lab that use heat shrink tubing and glue to attach an eight-tooth gear to the motor shaft. We now use a method that is more reliable and more versatile. We attach a small section of LEGO axle to the shaft, and then any gear (and other parts) can be easily attached. Since the adapter will go through a beam hole, you can attach it before mounting the motor; in fact this is the best procedure. The same technique is used for all the motors, but you must use the correct adapter piece because the motor shafts have different diameters.

7.1 LEGOizing Motor Shafts

○ Score the motor shaft slightly with sandpaper, *taking great care not to get sand inside the motor.* Point the motor shaft-down, grip it with a piece of sandpaper near the motor, and pull down away from the motor, so you are putting scratches along the length of the shaft. Repeat several times.

○ Clean the shaft with alcohol. Scoring and cleaning are especially important if you are re-working an old motor, so that the shaft is clean and the glue will adhere.

○ Obtain the correct motor-axle adapter piece. This is a piece of LEGO axle with a hole drilled through the center. The motor shaft is a *snug* fit in the hole. Different motors have different shaft diameters, so there are multiple adapters. *Be sure to use the correct one!*

○ Put a drop of instant glue (Loctite 409) on the end of the motor shaft; spread it around the shaft circumference. Only about 4mm or 3/16 inch of the shaft end needs to be coated.

○ It's best to have two people for the final step. Have your partner hold the motor on the bench with the shaft pointing up; place it on a LEGO beam so you will not bend the contacts or put pressure on the shaft. Your partner should also hold a LEGO beam so the motor shaft goes through a hole and the beam is flat against the brass motor bearing. This will indicate how far to push the adapter onto the shaft.

○ Hold the adapter with needle nose pliers, and put it on the motor shaft. You will have to tap it with a small hammer, or screwdriver handle, etc. Drive it down until it just clears the beam mounting guide.

○ Remove the beam, and give the glue a few minutes to dry.

7.2 Wiring a Cable and Plug to a Motor

○ Motor cables should be constructed with two strands of ribbon cable wire. Cut an 8-inch to 12-inch length of the wire.

○ Strip and tin both ends of the wire.

○ On the side of the motor there should be two metal lead/pads. Solder one wire lead to each pad. After proper soldering, hot glue may be used to hold the wire to the side of the motor for a stress relief.

○ Motor plugs should be wired for bidirectional use, as shown in Figure 19.

For the following steps, cut three-long strips of male socket header as needed.

○ Using the connector plug wiring technique shown in Figure 20 through Figure 23, wire the motor plug. Polarity does not matter since the plug may be inserted into a motor power jack in either orientation. However, placing a tag of tape on one of the wires just above the header, or marking one side of the plug with paint, will keep the motor orientation consistent. Even though the motors will operate when plugged in backwards, labeling the motor leads means your robot really will go forward when you tell it to, and not backwards.

CLIFF

Tools for Electronic Assembly

Quicktest

Preforming Machine

Fume Extractor

Task Lights

PC Workframe

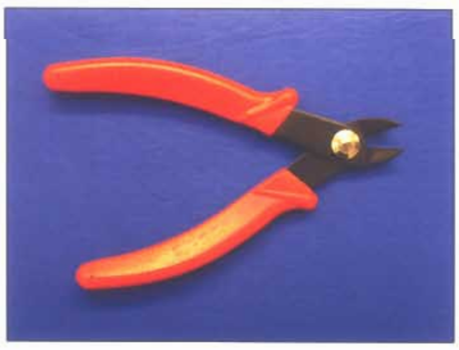
Hand Tools

Cliff Electronic Components, Ltd.
76 Holmethorpe Avenue, Holmethorpe Ind. Est.
Redhill, Surrey RH1 2PF. England
Tel: +44 (0) 1737 771375
Fax: +44 (0) 1737 766012
Email: sales@cliffuk.co.uk

Visit us online at:

www.cliffuk.co.uk

Quicktest

The Quicktest is ideal for electrical shops, factories, labs and service and test departments.

With the lid open the wires of a plug-less lead are inserted into the insulated clips. When the lid is closed, current passes to the equipment demonstrated or tested.

A neon light shows that the connector is live and ready to use. The 13 Amp fuse (BS 1362) is easily accessible for replacement.

The connector is made of tough safe materials. The base is phenolic moulded for insulation and rigidity. The lid is flame retardant glass filled thermo plastic.

The insulated clips are colour coded for easy identification. The flexible cable grip will take any normal mains cable. Only one small screwdriver is needed for both the base entry and the terminal screws.

Part Number CL1850 CE

PM-1 Pre-Forming Machine

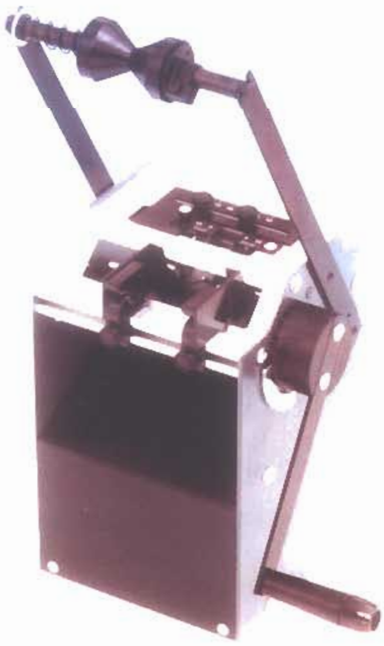

The PM-1 component pre forming machine is designed to cut to length and pre form axial lead components ready for fitting into PC boards. The components may be fed in bandolier form (assuming a 5 mm pitch between components or multiples thereof), or loose by the means of the built in chute slots.

Finished components are collected in a standard size storage box which may be located under the cutter head or in front of the machine.

All adjustments are carried out on the top of the machine, with setting instructions clearly displayed near the setting point.

The general construction features anodised aluminium body with heavy steel sides. The free running blades are 35mm in diameter and produced in M2 high speed steel. The cutting and bending wheels are produced in high carbon steel and are hardened and tempered as appropriate.

The machine will cut and pre form components at a rate of 40,000 per hour.

Guards are fitted all around the cutting and bending wheels to ensure the operator can work safely.

The PM-2 variation has the above features but is designed only to rapidly cut components out of a bandolier producing loose components with straight wires.

Tools for Electronic Assembly

CLIFF

Workframe

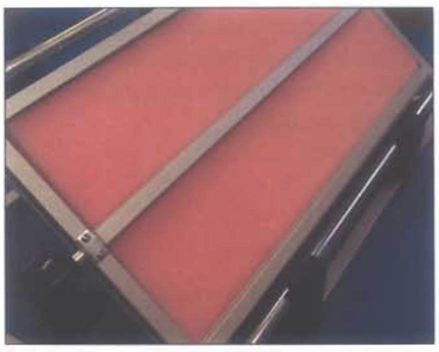

PC workframes are designed to facilitate components loading and soldering of small batch runs of assembled printed circuits. The different size models will hold most common Euro norm sized PCB's in one or two rows depending on the PCB dimensions and size of frame. The working angle is adjustable from vertical to 45 degrees. Components are held in place by an antistatic foam sheet. The PCB's are retained in slots in spring loaded adjustable bars. The PCB's can be removed by just springing the bars—once they have been set correctly for the PCB size. This enables rapid loading / unloading in a production situation

Specification:

Model No/Dimensions of Max sizes of PCB which can be held (mm)

Model	WF / 0	WF / 1	WF / 2
In 2 Rows	2 x 110 x 280	2 x 110 x 510	2 x 175 x 510
In 1 Row	1 x 230 x 230	1 x 130 x 510	1 x 360 x 510

CLIFFUME®

Compact Fumes Extractor

CLIFFUME SFE/1 is a low cost, compact, bench top fumes extractor system. It is primarily designed to be used to be used adjacent to hand soldering operations, where it will remove much of the smoke and fumes away from the operator. It can also be used in a gluing environment. *(note some solvents may attack the case and filter).*

The unit needs to be located as close to the fumes source as is practical, usually 10-20cm. The optimum position can be found by looking for the maximum smoke collection. The stand is designed to hold the fan at an angle, and assist smoke collection.

SFE/1 contains a filter element comprising activated charcoal filled reticulated foam. This will absorb particles down to about 5 microns. Filters are low cost and easily changed. Replace the filter when soiled, or when fume collection is observed as significantly reduced. The case is made from antistatic ABS. The stand can be earthed for optimum static dissipation.

The UK model is supplied individually boxed with fitted 13A plug. Rated AC240V 50Hz. 0.2A. Max CE

Ordering codes: OD 6680 UK 230/240V AC
OD 66805 Europe 220/240V AC (with Euro plug)
OD 66807 USA 110/115V AC (with USA plug)
OD 6681 Pack of 5 filters

t. Contact: sales@cliffuk.co.uk

CLIFF

Tools for Electronic Assembly

Hand Tools

A range of high quality cutters nippers and special uses tools. The range is manufactured for use in the electronic industries. Each tool is precision made from tempered steel for long lasting wear. All models have plastic insulated and moulded grips to provide safety and comfort for the operator. The range of tools comprises many variants for specific tasks, they include:

Clean cutters for ductile wires up to 1.5mm
Semi clean cutters for steel wire up to 1.5mm
Long handled nippers
Sheet metal cutter for thin metal, plastics, paper etc
Special use tools for bending, flattening and for work in narrow spaces.
General purpose crimping tools
Pliers with long jaws, curved jaws, needle nose jaws, smooth or knurled jaws.

Stack Rack

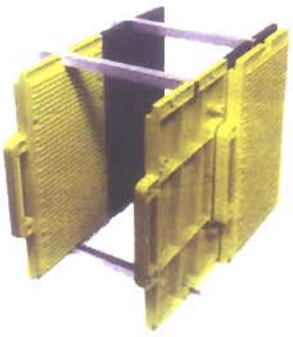

PC-01 is designed for safely holding PC boards during all stages of assembly / cleaning / testing. The width of the 'rack' module is easily adjusted from 0 to 290mm by sliding one side panel along the aluminium extrusion and tightening 2 screws. 1 Mtr lengths of extrusion are avialable, to allow construction of racks up to 970mm wide. The module is stackable in all directions so that racks of unlimited height and depth may be constructed. Plastic sides are antistatic—so sensitive devices may be safely stacked in PCBs—without special precautions other than to earth the rack.

PC-01C Conductive version. Side panels are made from conductive plastic, for use when highly sensitive IC's need protection against anti static charges throughout all handling, up to soak test or 'burn in'.

Ordering Codes PC-01 Rack, CL1887
PC-01C Conductive Rack, CL18875

Artikula Task Lamp

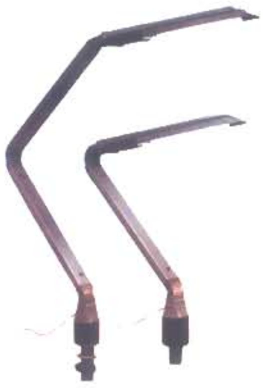

The original Artikula high tech flexible arm system. This design closely simulates the features of the human spine. The clean graceful profile of the arm gives a smart 21st century look to the desk or the workplace.

Available in two versions:

Desk version ART—D. 2 joints with a reach of 75cm
Task version ART-T. 3 joints with reach of 110cm
Each lamp is supplied with an energy saving 11 watt lamp, equivalent to the light of a 75 watt bulb. This provides an even and non glare illumination
The Artikula is made from engineering class plastics and can be fixed by an innovative clamp which opens up to a max 100mm.
EC approved under ENEC-SE-0030, the lamp is supplied with 2.8M of cable a moulded plug and ballast.

© 2004 CLIFF®

Drawings available on request. Contact sales@cliffuk.co.uk

BERNARDS

BUILD YOUR OWN ELECTRONIC EXPERIMENTERS LABORATORY USING IC's

BY

B. B. BABANI

NO.218

85p

BUILD YOUR OWN ELECTRONIC EXPERIMENTERS LABORATORY USING IC's

BY
B. B. BABANI

BERNARDS [publishers] LTD

The Grampians
Shepherd's Bush Road
London W6 7NF
England

CONTENTS

SIMPLE AUDIO FREQUENCY OSCILLATOR USES 2 LOW COST IC'S 5
TTL PULSE DETECTOR USES INEXPENSIVE IC, LED 15
HIGH IMPEDANCE VOLTMETER BASED ON LOW COST OP AMPLIFIER 16
SQUARE WAVE OSCILLATOR/PULSE GENERATOR USES 555 IC 28
DIGI DESIGNER 31
5-STAGE LOGIC PROBE HAS 7-SEGMENT READOUT 35
REPAIRING CRACKED PC BOARDS 40
LOW COST SCALER EXTENDS DIGITAL COUNTER TO 200MHz 41
BUILD OUR DIGI-METER : A DVM WITHOUT TEARS 49
BREAKDOWN VOLTAGE TESTER FOR SEMICONDUCTORS 56
LOW RANGE OHMMETER 65
EXPERIMENTAL BRIDGE TYPE WIDE RANGE OHMMETER 66
SYNCHRONOUS GATING UNIT FOR AUDIO TESTING 74
SIGNAL TRACER USES IC 89
SIMPLIFIED VERSION OF SQUARE WAVE GENERATOR 90
LIST OF EQUIVALENTS FOR CIRCUITS USED IN THIS BOOK 92

Although every care is taken with the preparation of this book the publishers will not be responsible for any errors that might occur.

The compiler of this book wishes to express his most sincere thanks to "ELECTRONICS AUSTRALIA" the leading radio, TV and electronics magazine published in that continent, and to Sungravure Syndication, their agents, for permission to use some of the material in this book which appeared originally as articles in that magazine.

© 1975 by Bernard B. Babani

I. S. B. N. 0 900162 57 0

First Published December 1975

Printed & Manufactured in Great Britain by
C. Nicholls & Co. Ltd, Manchester M11 4AU

SIMPLE AF OSCILLATOR USES TWO LOW COST ICs

Here is the design for an easy to build audio oscillator which uses only two low cost ICs. It generates both sine and square waves at low distortion, covering the frequency range from 3Hz to 30kHz. You can power it from batteries, or from a simple mains supply.

Looking through a recent edition of "Radio Communication", my interest was aroused by a very simple Wien bridge oscillator circuit by D.S. Jones. The circuit uses the ever popular and very low priced type 741 IC op amp. The supply suggested is 12 volts and the current just a few milliamps. Tuning is by a ganged 10k potentiometer and the frequency coverage is from 3Hz to 40kHz in four ranges. This looked like a good starting point for a low cost, simple, yet very worthwhile, economy audio oscillator.

By rearranging some of the time constants in the Wien bridge, the frequency coverage could easily be changed to go from 3Hz to 30kHz in four ranges. Instead of a 12V DC supply, this has been increased to 15V and could even be increased further to 18V. Hopefully, the increase in supply voltage would reduce the harmonic distortion to some degree.

As this audio oscillator set out to be an economy unit, a fairly obvious saving would be to do without a level meter. However, by taking some other quite inexpensive steps, the output voltage may be determined with quite good accuracy without the use of a meter.

The sine wave output is over one volt and by adding a preset potentiometer in series with the output load, the voltage across the load could be adjusted to one volt against a suitable meter. As the output voltage is almost constant over the entire range, it is then only necessary to provide a means of voltage division and the output voltage can be attenuated by a known amount.

By suitable switching, the sine wave output can be put through a type 555 IC used as a Schmitt trigger to give a square wave output. As the output of this stage is also in excess of one volt, the same procedure has been adopted to reduce the output to one volt and this is attenuated in the same way as for the sine wave output.

The oscillator circuit is the Wien bridge and for those unfamiliar with this arrangement it is shown in principle in Figure 1. As may be seen it consists of a high-gain differential amplifier with two separate feedback circuits. One circuit, consisting of resistors R1 and R2 and capacitors C1 and C2, connects from the amplifier output back to the "+" input and so provides positive feedback. The other circuit consists of resistor R3 and the thermistor (negative temperature-coefficient resistor), and is connected between the output and the "–" input to provide negative feedback.

The configuration R1, R2, C1 and C2 is known as the "Wien network". At a particular frequency determined largely by the values of the four elements, the transmission loss of the network falls to a minimum while the phase shift also passes through zero.

If R1 and R2 are equal in value, and C1 and C2 also equal in value, the frequency of oscillation is equal to the reciprocal of 2piRC. Under these conditions, the transmission loss falls to a minimum of 3. In other words, there is a maximum transmission gain of 0.33.

To produce sustained oscillations, the overall loop gain must be at least unity and the phase shift either zero or a multiple of 360 degrees. For the circuit of Figure 1 to produce continuous oscillations, the amplifier gain must be at least 3 to compensate for the loss in the Wien network.

It is necessary for the amplifier to have an overall loop gain of at least unity for sustained oscillations. But if the gain is more than one, the oscillations will continue to grow until the amplifier is driven into clipping. To keep the gain at unity and oscillation at constant amplitude, a thermistor is used in the negative feedback network.

Since the thermistor has a negative temperature coefficient of resistance, its resistance falls as its temperature rises. When power is first applied to the circuit the thermistor will have a relatively high resistance and so there will be little negative feedback. The resulting high gain around the positive feedback loop will rapidly build up oscillation.

As the oscillations grow, the temperature of the thermistor rises, as this and resistor R3 are effectively connected in series across the amplifier output and so draw signal current. Hence the resistance of the thermistor falls, negative feedback increases and the effective amplifier gain drops.

Consequently an equilibrium is reached, as the output amplitude can only rise to the point where the thermistor has increased the negative feedback to correspond to an effective gain of three – giving unity loop gain. If the oscillations tend to rise above this level, the thermistor will reduce the loop gain and the oscillations will begin to die away. Conversely, if the amplitude tends to fall, the thermistor will increase the loop gain slightly above unity to correct it.

In short, the non-linear negative feedback action produced by the thermistor acts to continuously and automatically maintain the loop gain at unity, and the oscillation amplitude constant. By using a suitable thermistor the output amplitude may be maintained at a level well below limiting, giving an output waveform with low distortion.

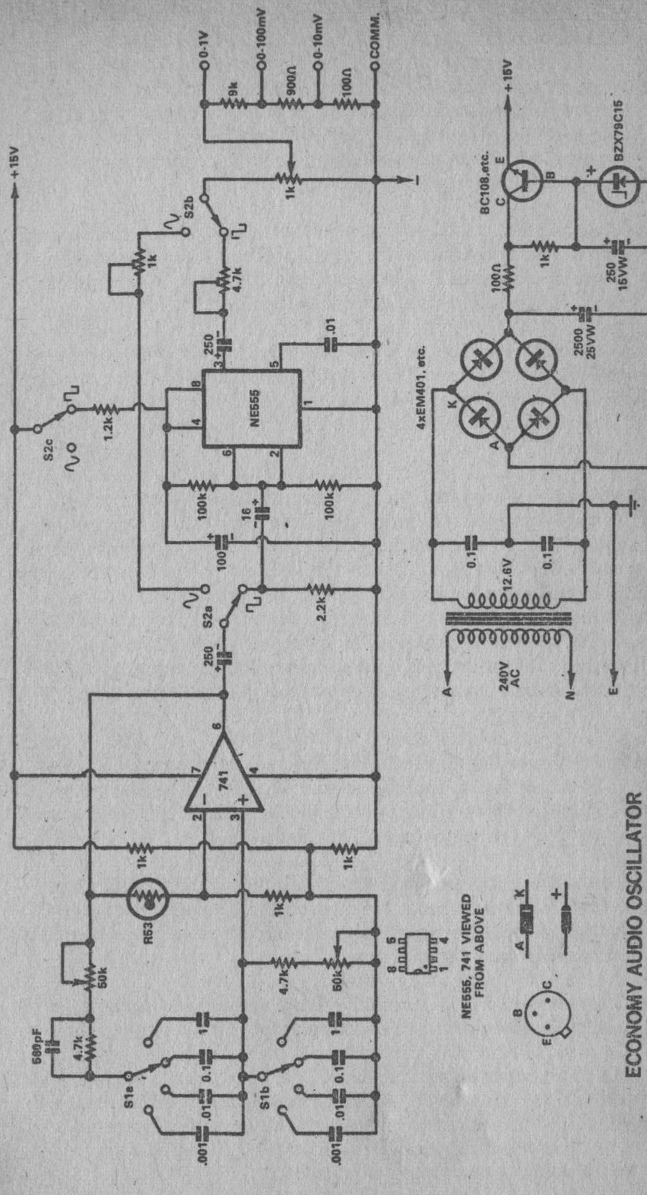

ECONOMY AUDIO OSCILLATOR

The circuit of our economy audio oscillator uses two low cost ICs and offers low distortion.

Looking now at the main circuit, the relative parts may be clearly seen with reference to the foregoing description of the theoretical aspects of this type of circuit. The positive and negative feedback circuits are easy to see, along with the two resistors and capacitors for each frequency range. Before leaving the oscillator section of the circuit, it may be seen that one of the 4.7k resistors is shunted with a 560pF capacitor. This is to compensate in some measure for stray capacitance in the circuit, which becomes significant at the higher frequencies.

The oscillator circuit is followed by a Schmitt trigger which can be used to convert the sine wave output from the oscillator to a corresponding square wave. The Schmitt trigger uses a type 555 timer IC as suggested by Walter G. Jung and the circuit performs this task adequately.

In order to select either sine or square wave output, a three-pole two-position toggle switch is used. This diverts the sine wave either through or around the Schmitt trigger. At the same time, the positive supply rail is switched on or off in accordance with the requirements of the Schmitt trigger.

Whichever output is selected, it is fed into a 1k linear potentiometer. Before being fed into this potentiometer, it is passed through a variable resistor in the form of a preset pot, which is set to give a predetermined output level. More will be said about this later on. To give a wider range of control over the output level, the output of the 1k pot is fed into a voltage divider. The divider provides for divisions of 1, 10 and 100, corresponding to attenuations of 0dB, 20dB and 40dB, respectively. In addition, the 1k pot is fitted with a knob calibrated 0 to 10, which means that attenuations may be achieved fairly accurately down to 60dB.

With the preset pots just referred to adjusted so that we have a level of one volt across the 1k pot, we are able to make settings between 0 to 1V, 0 to 100mV and 0 to 10mV. We may therefore set the output level fairly accurately anywhere between 1mW and 1V.

We elected to incorporate a power supply within the unit, although it would be quite possible to run it from batteries. The power supply is regulated at 15V and although it is a simple circuit, steps have been taken to keep hum and noise to a minimum.

The transformer has a 12.6V secondary winding rated at 150mA. Four diodes in a bridge give about 18V DC across the 2500µF filter capacitor. A BC108 transistor is used in a simple series regulator circuit which is referred to a 15V zener diode in its base. The two 0.1µF capacitors, one on each side of the secondary of the transformer help to reduce noise from the mains. The negative supply line is left floating as is normal practice with this type of instrument. Metal parts of the case are connected to the mains earthing system.

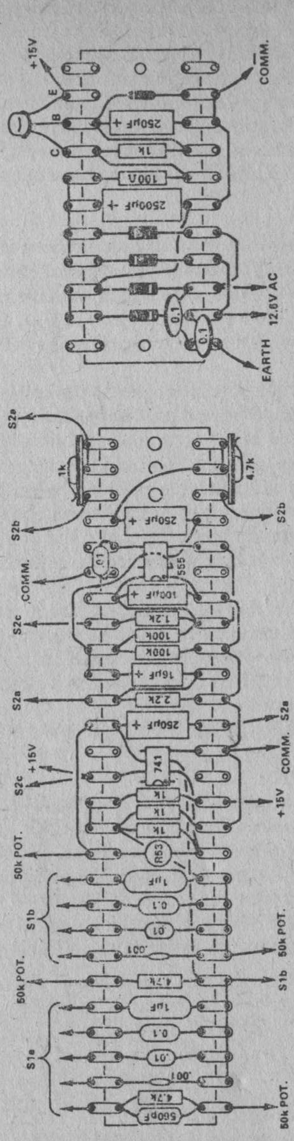

These two wiring diagrams should make construction a straight-forward process. The oscillator tagboard is shown at left, whilst the simple regulated power supply is shown at right. Make sure that the thermistor is firmly secured in position.

Some comments on the more important components may be helpful. Perhaps the item of most interest is the 50k linear ganged potentiometer. In the interest of economy, we used an imported deposited carbon unit under the name of Soanar. This has proved to be quite satisfactory; we have tried a number of them in the circuit and each one follows the calibrations well enough for all practical purposes. For readers who decide to use this potentiometer, we are printing a full size copy of the scale which may be used in any way that is considered best.

Before leaving the subject of carbon potentiometers, we can see no reason why some other brands of potentiometer should not be tried. However, it may be that the scale which we have calibrated against the Soanar unit may not suit another brand. In this, and other circumstances yet to be touched on, it would be necessary to calibrate your own scale.

When considering what potentiometer to use initially, we invested the idea of using a moulded carbon unit but as near as we can ascertain, these are not available in the type and resistance value required. In past projects we have used a wire would potentiometer, made by Naunton. These are still available and may be used if any reader so desires. However, there are two points worth considering. It would be necessary to calibrate your own scale, and the price could be rather daunting. The choice of course is up to the individual.

The output level control is a 1k moulded carbon potentiometer in the prototype. While an ordinary carbon pot may be quite satisfactory, we considered that the moulded element would be more reliable from a linearity point of view. This is important as we have fitted a knob to this control which is calibrated 0 to 10. This facility, in conjunction with the output divider, makes it possible to set the output level with reasonably good accuracy.

The output divider just mentioned is made up of a number of resistors to give the correct ratios. The bottom element is 100 ohms and this presents no problem as it is a preferred value. The next section calls for 900 ohms and this is made up with 220 ohms and 680 ohms in series. The top section has a value of 9k and this is made up with 2.2k and 6.8k in series. We used Philips ½ watt types here and although we did not specify close tolerance, we have found them to be quite satisfactory.

The set of capacitors for the frequency ranges should be of good quality and preferably of 1 percent tolerance. The tolerance applies to all except the 1µF. Although it would be ideal to have this in the same tolerance, it would be quite expensive and we understand that availability is open to question. We found that with 10 percent tolerance units in circuit the amount of error was not significant for normal purposes.

The thermistor is an STC type R53 and this does an excellent job in this circuit. While we have not tried other types, we feel that it would be wise not to substitute for this item.

Both ICs are readily available. The type 741 comes in a number of brands but the 555 IC timer is only available from Signetics. Both ICs are 8-pin DIL are used in the prototype but other packages are also available if required. For our purpose, the smaller package is more convenient.

The prototype was built into a plastic case with an aluminium front panel and one which we have used a number of times before.

Most of the components are mounted on either one of two miniature tag boards and these are shown in diagram form. Care should be taken with the thermistor to stabilise it mechanically and this may be done by simply tying or gluing it to the adjacent 1µF capacitor. Sockets for the two ICs are fixed to the board by drilling either holes on the board carefully with a fine drill. Lugs of the sockets are pushed through the board and leads taken as shown in the diagram. The photographs show the location of all major components. The boards are stood off the case by 5mm or so by means of spacers, or by using a couple of extra nuts on each of the mounting screws.

The power transformer which we used had a centre tap on the secondary winding, but this is not required. We simply cut it off to about 10mm long and located it so that it was not likely to come into contact with any other parts of the unit when it was all assembled. A lead which should not be forgotten is from the mains earth termination to a lug on the metal front panel.

Once the unit is assembled it is ready for adjustment and calibration. Before going ahead with these operations, a simple check that it is working would be in order. Feed the output into a CRO or listen on a pair of headphones. If you choose the latter method, make sure that the oscillator is operating at a frequency which is covered by the phones in use. Assuming that you get sine waves and square waves on the CRO, or sounds resembling a flute with the switch in the sine wave position and sounds like a clarinet when the switch is in the square wave position, you are in a position to carry out adjustments.

Set the dial knob assembly on the potentiometer spindle such that the pointer corresponds with the marks on the scale for extreme limits of travel. Set the output knob on the output pot spindle so that 0 and 10 correspond with minimum and maximum output, respectively.

Set the range switch and frequency dial to give an output of about 100Hz. Set the output level control to 10 and take the output from sockets "E" and "0–1V" and connect to an AC voltmeter of good accuracy and high impedance. Set the toggle switch to sine wave output and adjust the 1k preset pot to give a reading of 1 volt. Set the toggle switch to square wave output and adjust the 4.7k preset pot to give a reading of 1 volt.

These Lissajous figures show (left to right) vertical to horizontal frequency ratios of 1:1, 2:1, and 3:1.

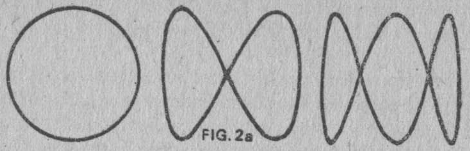

FIG. 2a

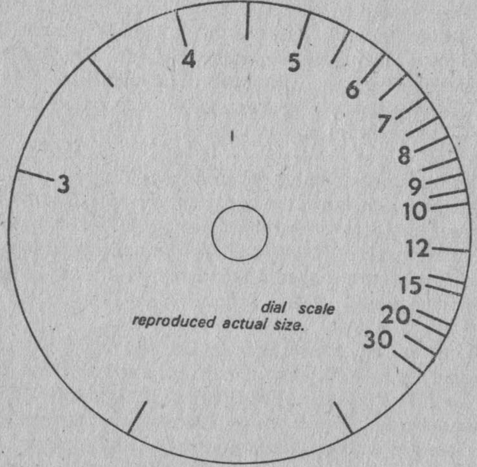

dial scale reproduced actual size.

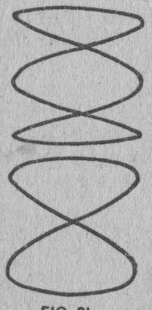

FIG. 2b

These Lissajous figures show vertical to horizontal frequency ratios of 1:3 (top) and 1:2 (bottom).

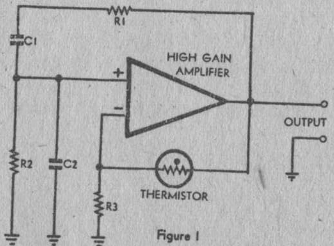

Figure I

Above is the schematic of a Wien bridge oscillator showing the high gain amplifier and its associated positive and negative feedback networks.

If you have access to a calibrated CRO, the two preset pots just mentioned may be set against the CRO graticule. For sine wave output, the level should be set to give a reading of 2.8V peak-to-peak and for square waves, a reading of 2V peak-to-peak should be obtained.

If you have used the same type of 2-gang potentiometer as for the prototype, together with a copy of our scale, then you will not have to carry out any further calibrations. For readers who have used the wire wound, or some type of pot different from that on the prototype, then it will be necessary to calibrate for one range. This could conveniently be from 30 to 300Hz, with the range switch set to "X10".

If you are fortunate enough to have access to a digital frequency meter, then the job may be done quite easily. Another method is to generate Lissajous figures with the aid of an oscilloscope and another known source of audio frequency. The second source could be an accurately calibrated oscillator, a test record, an electronic organ or the mains frequency of 50Hz.

For those not familiar with the Lissajous figure method of frequency comparison, the following briefly outlines the procedure. The output of the known frequency source is fed to the horizontal amplifier of the CRO and the output of the oscillator to be calibrated is fed to the vertical amplifier. When the gains of the two deflection amplifiers are suitably adjusted, a series of patterns known as Lissajous figures are produced. Typical frequency patterns are shown. These enable the unknown frequency to be calibrated easily.

These figures differ from the normal pattern produced by a sine wave oscillator, since the horizontal deflection signal is a sine wave and not the normal sawtooth produced by the oscilloscope timebase generator.

Some examples of Lissajous figures are shown in Figure 2a. Each represents a different frequency ratio – 1:1. 2:1 and 3:1. Each ratio is that of the vertical frequency to the horizontal frequency. If the signal applied to the horizontal amplifier is 50Hz and the resulting ratio is 2:1, then the vertical frequency will be 100Hz.

The method of determining the ratio is to count the loops along the top of the pattern and the number of loops along one side. The ratio of the two numbers if the frequency ratio.

This basic rule makes it possible to identify a variety of ratios. For example, five loops along the top and two at the side would indicate a ratio of 5:2 and, related to 50Hz, this would indicate a frequency of 125Hz. Incidentally, submultiples can be identified in the same way and Figure 2b indicates the pattern shape for ½ and 1/3 of 50Hz, corresponding to 25Hz and 16.7Hz.

If you attempt to simulate these patterns on a CRO you will find that the patterns tend to drift. This is because the phase relationship of the two frequencies is changing because of small frequency drift of the two sources. Unless the two sources are very stable, the circle pattern will slowly change to an ellipse to a straight line as the phase relationship of the two signals changes. A circle represents a phase difference of 90 or 270 degrees while a straight line represents a phase difference of zero or 180 degrees. This drift in the pattern is only a portion of a cycle and so it has negligible effect on calibration accuracy.

The foregoing information on Lissajous patterns may be used to calibrate your oscillator and this may be applied most easily to the range 30 to 300Hz, as mentioned earlier. Once you get the easy points, such as 50Hz, 100Hz, 150Hz, etc., other intermediate and less obvious patterns will become obvious by virtue of interpolation of the already determined points. It is not a difficult operation and indeed, can be quite fascinating.

List of Component Parts

1 Case 184 × 115 × 118mm, with aluminium front panel
1 Transformer, 240V primary, 12.6V secondary at 150mA.
1 50k linear dual-ganged potentiometer (see text)
1 1k linear moulded carbon potentiometer
1 2-pole 4-position rotary switch
1 3-pole 2-position toggle switch
4 Banana sockets, 1-black, 3-red
1 3-way terminal strip
1 Jabel handspan knob for frequency scale
1 Scale (see text)
1 Knob with indicator dot for range switch
1 Knob calibrated 0 to 10 for 1k pot
1 Tagboard with 27 prs tags
1 Tagboard with 11prs tags
4 Diodes, EM401 or RS276-1139 or BY127 or 1N4007
1 Zener diode, BZX79/C15 or RS276-624/564
1 Transistor, BC108 or RS276-2009
1 IC, 8-pin DIL 741 or RS276-010/1741
1 IC, 8-pin DIL 555 or RS276-1723
2 8-pin DIL IC sockets
1 1k linear trimpot
1 4.7k linear trimpot
1 Thermistor, STC type R53

RESISTORS (½ or ¼ watt)
2 100 ohms
1 220 ohms
1 680 ohms

4 1k
1 1.2k
2 2.2k
2 4.7k
1 6.8k
2 100k
CAPACITORS
2 0.001μF 630V polystyrene, low tolerance
2 0.01μF 630V polystyrene, low tolerance
2 0.1μF 100V polycarbonate, low tolerance
2 1μF 200V polycarbonate (may be low tolerance if available)
1 0.01μF 25V redcap
2 0.1μF 100V polycarbonate
1 16μF 40VW electro
1 100μF 12VW electro
2 250μF 12VW electros
1 250μF 15VW electro
1 2500μF 25VW electro
SUNDRIES
3-core flex with 3-pin plug and flex clamp, screws, nuts, spacers, hookup wire, solder.

TTL PULSE DETECTOR USES INEXPENSIVE IC, LED

When one is testing an experimental logic circuit, or troubleshooting in a completed digital instrument or system, it is often necessary to detect single, fast pulses. At times this can be surprisingly difficult even using a wideband scope with extensive triggering facilities. Some logic probes have provision for "stretching" short pulses, but even this can be of limited help.

I have found that this simple circuit is very useful in this sort of situation. It will only detect negative-going pulses, but in practice this has not proved a problem as it is usually possible to look for a pulse where it is negative going.

As you can see, it is a simple R-S slip-flop formed from two gates of a 7400 IC. A third gate is used as an inverting driver for the indicating LED. The circuit is reset before each "search" by pressing the pushbutton.

If the probe is taken down to the low logic level (0. 8V or less) for even as short a time as about 10 nanoseconds, the circuit will quickly flip over to the set state and latch there, indicating that a pulse occurred.

Almost any low cost LED can be used.

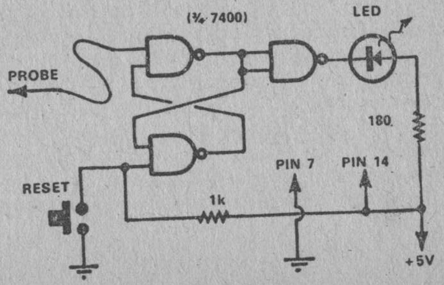

HIGH IMPEDANCE VOLTMETER BASED ON LOW COST OP AMP

Here is an easily built high impedance electronic voltmeter using a single 741 op amp IC. While the basic instrument is designed for DC measurements, probes are also described to allow its use for measuring both RF and low frequency AC signals.

Specifications

Ranges: DC – 0-1, 0-10, 0-100 volts. AC and RF – 0-1, 0-10 volts RMS.
Input impedance: On DC – 22 megohms, all ranges.
On AC and RF – 5 megohms approx., both ranges.
Frequency: 20Hz to 50MHz, using two probes.
Power supply: 2 x 9V batteries. Drain approx. 2mA.

When working on solid-state circuits there are many occasions when a DC voltmeter of really high resistance is very useful, because it can give substantially true readings in high-resistance circuits without upsetting their working conditions to any serious extent.

Needing such an instrument wuickly and having no calibrated vacuum-tube voltmeter handy, the writer considered what might be done by using a 1-milliamp FSD meter which was on hand, in conjunction with an inexpensive linear integrated circuit. Such an arrangement has the advantage, apart from the possibility of high input resistance, that calibration is unnecessary if accurate resistors are available, because a reading of one milliamp full scale can easily be made to mean one volt, with other ranges in proportion merely by switching one resistor. Accuracy and stability can be very good because of the large amount on negative feedback employed with the integrated circuit. Furthermore, a 1-milliamp meter is cheaper and much more robust than those of ten or twenty times greater sensitivity which are usually found in normal high-resistance voltmeters – which, in any case, will have a resistance seldom exceeding half a megohm on the lower ranges.

Having worked out the details of such a DC instrument, the idea of extending its usefulness to radio-frequencies and, consequently, to AC voltage measurements present itself. The need here was for an RF voltmeter of low losses, low capacity, and an input impedance of several megohms, which could be applied to a tuned circuit or oscillator while leaving its "Q" almost unaffected. Therefore it was arranged that an AC or RF probe on a long flexible lead could be plugged into the DC instrument, the probe containing a full-wave peak-to-peak rectifier circuit using low-loss materials. It was then only necessary to change one resistor in series with the meter movement by a DC-AC switch to have the meter show RMS values of AC voltage, with the original linear scale remaining accurate throughout except for the first one-third of the 1-volt scale where the curvature of the rectifier characteristic produces increasing non-linearity. A calibration curve for this part of the scale is shown hereafter.

At first sight it seemed that one AC probe would suffice for both audio and radio frequencies. However, it was desired to extend the audio range with full accuracy down to 20Hz and this required capacitors of $1\mu F$ in the rectifier circuit. Flat lacquered polyester-foil capacitors were used here and were quite satisfactory at the lower frequencies, but it was feared that their losses would be too high at radio frequencies, where much smaller capacity would in any case be sufficient.

Since low RF losses were an essential requirement of the project, a second probe was made for this work, using small polystyrene-foil capacitors and glass-fibre insulation, designed for frequencies from 100kHz upwards; while the other probe covers the range 20Hz to 200kHz.

Figure 1 shows the circuit of the DC instrument and Figure 2 the two probe circuits. Note that the low-frequency probe has a DC-AC switch so that it can be used for both purposes. AC and RF measurements are confined to the 1-volt and 10-volt ranges only, as anything exceeding 40 volts RMS could break down the rectifier diodes in the probes.

The Integrated Circuit

For readers who may not be very familiar with these devices, a very brief explanation may be in order. The type of IC used here is called an operational amplifier. Originally devised for carrying out certain mathematical operations in computer work, it is now used increasingly for all sorts of linear amplification purposes in both DC and AC circuits, and for these purposes it has potent advantages in space, cost, simplicity, stability and economy of working current.

An operational amplifier of this type is, in essentials, a direct-coupled multi-stage DC transistor amplifier of very high overall gain, intended specifically to be used with very heavy negative feedback. The initial gain, called open-loop gain, may be anything from around 2000 times to 200,000 times or more, depending on the type chosen, but the gain after feedback is applied, called the closed-loop gain, is seldom required to be more than 1000 times and is commonly much less.

This massive negative feedback not only extends the frequency response and reduces distortion and noise in the usual manner of NFB; it also makes the whole circuit largely independent of manufacturing tolerances in the IC itself and in its supply voltages. As a result, its performance in a given circuit can be predicted and determined almost completely by the choice of the external feedback components. Feedback can be either DC as in the present design, or AC as in the case of pre-amplifiers for audio frequency work, or both. There are, however, frequency limitations depending upon the type of IC chosen and very few are useful beyond a megahertz or so irrespective of the amount of feedback used. They are, as previously said, primarily DC amplifiers.

Usually an operational amplifier, or "op-amp", has two inputs and one output, plus a few other connections for supply voltages and other purposes. The input connections are called "inverting" and "non-inverting" inputs respectively, meaning simply that an input signal applied to either of them will appear at the output terminal either with the same or with the opposite polarity, according to choice of input terminal. To avoid drawing the whole complex internal circuit of the IC, the symbol shown in Figure 3 is used to represent it, namely, a triangle placed sideways with inputs on the left and output at the apex on the right. The inputs are labelled "+" for the non-inverting one and "–" for the inverting one. A signal applied to the "+" input appears at the output in phase with the input; if applied to the "–" input the output is in phase op-position; i.e. there is phase reversal.

Figure 3 shows the two basic circuits for an "op-amp". In 3a showing the "inverted" circuit, the input goes through R1 to the inverting terminal and negative feedback is applied through R2 to the same terminal. If the source of the signal has low impedance, the closed-loop gain of this system is, practically, R2/R1 and by varying R2 it can be made almost anything from unity up to the full open-loop gain. The input impedance of this arrangement is relatively low – virtually equp to R1 – and its output impedance fairly high; these properties are useful in many circuits but not for the present application where we need very high input and low output impedances, the latter to suit the low resistance of the milliammeter.

We therefore turn to Figure 3b, showing the non-inverting configuration. Here we feed the signal into the "+" terminal, but the feedback remains connected to the inverting terminal. The closed-loop gain will now be equal to R3/R2 and the input impedance will be high – several megohms at least – and output impedance 75 ohms or less. These characteristics evidently suit our present purpose and this circuit was chosen.

Several types of linear op-amps are available at moderate prices having widely-different characteristics. For this instrument the popular type 741 seemed most suitable. This is quite a complex affair containing twenty transistors, many resistors and a capacitor for frequency compensation and having a typical open-loop gain of 200,000 times. It is available either in a TO99 round can or as 14-pin or 8-pin dual-in-line packages; the 14-pin version was chosen. A second version, type 741C, is often available and sometimes at a rather lower price; this appears to be a so-called "Commercial" version with somewhat looser tolerances,

Complete circuit of the Op Amp voltmeter. The required range is selected by varying the feedback network. AC and RF voltages may be measured by means of external probes. (See Fig. 2)

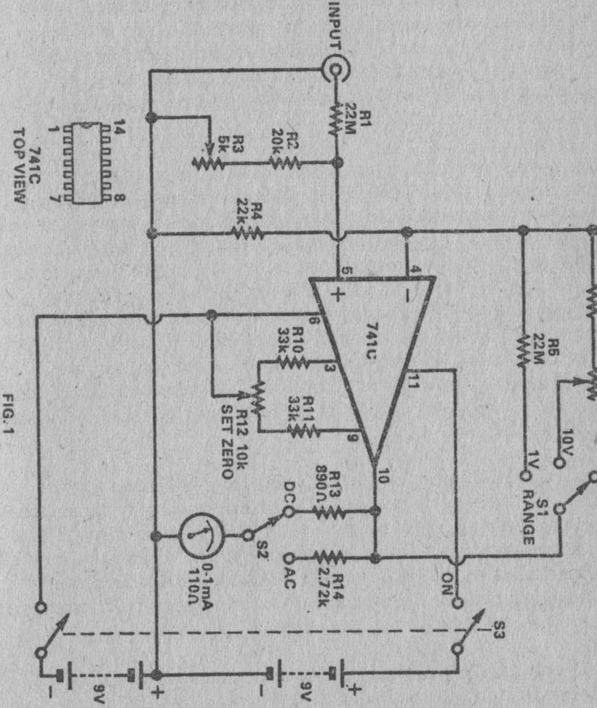

FIG. 1

quite satisfactory and is used in this instrument. A constructor may use whichever type suits him or is procurable, and the package version desired.

The practical circuit evolved into that shown in Figure 1. Starting with the meter, we connect in series with it a resistor of a value which, together with the meter's internal resistance, will total 1000 ohms and thus give a full-scale deflection (1mA) with 1 volt applied across them. This will be the output required from the op-amp on all ranges and the meter scale divisions can be read as volts or fractions of a volt, multiplied by the range switch setting. The meter used here is an imported one of SEW make, type MR-85P with square face measuring $4\frac{5}{8} \times 4\frac{3}{16}$ inches, and its measured resistance was 110 ohms, requiring a series resistor of 890 ohms. This can be made with sufficient accuracy from 820 ohms and 68 ohms in series, both of 1pc tolerance, these being preferred values. (R13 Figure 1.)

While a large meter like this one is nice to work with, easily read and robust, any 1mA meter can be used. If its internal resistance is not known or cannot be measured, it can be calibrated in terms of whatever 1 volt standard meter may be available, using a variable pre-set trimmer resistor of around 2000 ohms as the series resistor; the accuracy will then be no better than that of the standard voltmeter used.

It was desired to have a DC input resistance of at least 20 megohms, this being determined by R1. The nearest preferred value is 22 megohms, and the required tolerance is 1pc; this being difficult to procure, it was made up from two 10 megohm and one 2 megohm units of 1pc in series. Both input circuits of the op-amp have to be returned to the earth line through resistors of only moderate value and of the same resistance, otherwise the slightly differing input currents can set up standing voltages large enough to make zero-balancing difficult. Thus R2 and R3, taken together, form a voltage-divider with R1 which reduces the input signal applied to the op-amp; a value of 22k was considered the highest desirable. Thus an input of 1 volt will be reduced to 1 millivolt at the input of the 741.

Since this must produce an output of 1 volt for application to the meter, the 741 must give a gain of 1000 for this range and the feedback circuit on the inverting input must be proportioned accordingly. This input, as above-mentioned, must be returned to the earth line through a resistance similar to R2 + R3, namely 22k. Thus R4 is determined and the feedback resistor R5 then automatically becomes 22 megohms, the ratio of R4 to R5 producing the required gain of 1000.

For the 10-volt and 100-volt ranges it is only necessary to change the feedback resistor to increase feedback and reduce gain in the required ratio; thus R6 + R7 becomes 2.2 megohms for the 10-volt range and

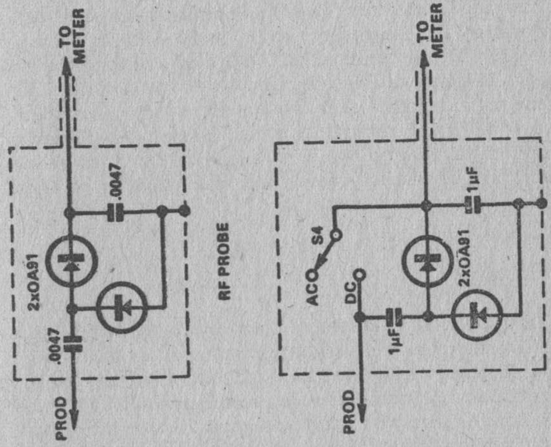

RF PROBE

AC-DC PROBE

FIG. 2

Circuit for the RF and AC-DC probes to suit the voltmeter. Each probe is constructed in a small IF transformer can.

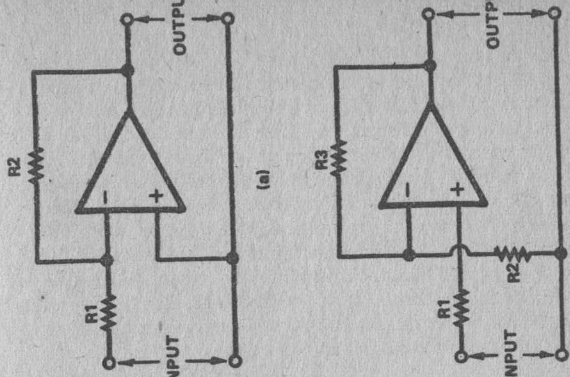

(a)

(b)

FIG. 3

The final circuit is based on (b). This gives a high input and low output impedance, ideal for this requirement.

R8 + R9 becomes 220k to give the 100-volt range. These three conditions are shown in Figure 4 with the respective input and output voltages and gains indicated.

All resistors so far mentioned should ideally be of 1pc tolerance, and if these are available the accuracy of the instrument will be dependent only on that of the meter movement without further calibration. However, close tolerance resistors may sometimes be hard to get, so variable pre-set resistors have been added at the signal input (R3) and the 10-volt and 100-volt feedback resistors (R7 and R9) to permit exact adjustment, either to the required tolerance on a bridge or by direct comparison with another voltmeter. These can be miniature carbon trim-pots if low cost is important but they are awkward for precise adjustment and of doubtful stability of setting. A much better solution is the miniature Cermet trim-pot of the 10-turn type as made by Spectrol, Beckman, Morganite and others; these are probably worth the extra cost in terms of accuracy and stability.

It is necessary, in order to bring the output voltage to zero in the absence of an input signal, to balance the input currents of the two input terminals. Provision is made for this in the internal circuit by connecting a potentiometer (R12) across terminals 3 and 9, with its slider going to the negative supply rail (terminal 6). Additional resistors R10 and R11 were found desirable to make the adjustment less critical. R12 is a panel control with a suitable knob, used to set the meter reading to zero on the 1-volt DC scale only; this setting holds good on all other ranges; It is quite stable and shows very little drift after a few minutes' warmup, but it may be checked occasionally when great accuracy is called for.

AC and RF Ranges

As stated earlier, the rectifier system in the probes actually produces a DC output proportional to the peak-to-peak value of the measured AC voltage, which is roughly 2.83 times the RMS voltage. If applied to the DC voltmeter circuit as it stands, the scale readings would have to be divided by this factor, a very awkward procedure, or else the meter would have to be re-calibrated with an additional peak-to-peak AC scale; both expedients to be avoided. The alternative is to reduce the milliammeter sensitivity until it reads the RMS value of the peak-to-peak voltage, thus preserving the original single scale for all purposes. This can be easily done by changing the multiplying resistor in series with the meter, and switch S2 does this by bringing in R14 to replace R13. Its value can easily be calculated.

The combined resistance of the meter and R14 must now be 1000 × 2.83, or 2830 ohms; subtracting the meter's internal resistance of 110 ohms (in this case) give 2720 ohms for R14. If desired, this could be made a variable resistor to permit fullscale deflection to be adjusted on

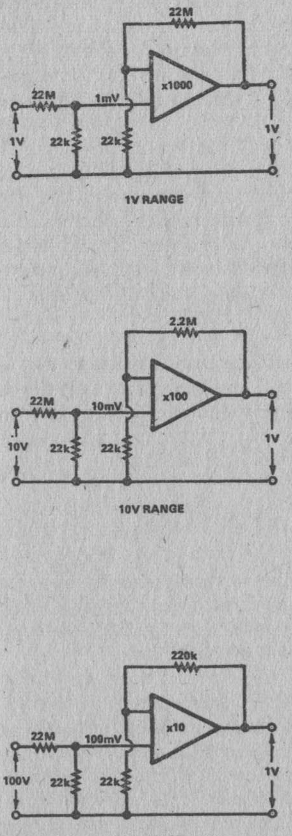

FIG: 4
Showing how the meter ranges are provided by changing the feedback resistor values. Values for other ranges can be calculated.

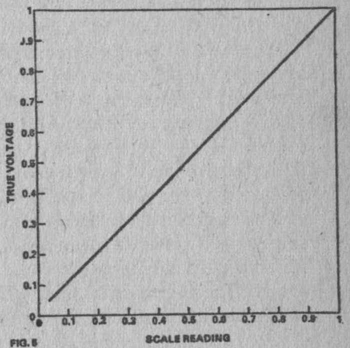

FIG. 5
Correction curve for the 1V AC range. Although the error is small this curve will correct for it where it may be important.

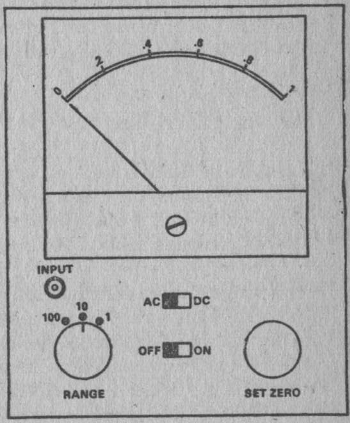

FIG. 6
The physical form of the finished meter is not critical, but the panel arrangement shown should serve as a guide.

the 10-volt scale by comparison with another AC voltmeter, but most such voltmeters (of the rectifier type) are not accurate to better than 4 or 5pc and the use of the calculated resistor value is likely to give greater accuracy.

As mentioned earlier, the lower one-third of the 1V AC scale becomes non-linear, but the curve of Figure 5 shows that this is not serious. Nevertheless, the curve gives quite accurate correction down to 0.1V RMS at least.

It is worth remembering that by leaving S2 in the DC position while measuring AC, one gets, effectively, the peak-to-peak reading instead of RMS. In other words the sensitivity is increased by the factor 2.83 and although the scale is not calibrated for it, this increase is often quite useful for comparative readings of RF voltages too small to be read comfortably on the RMS scale.

The effective input impedance is lower when using the RF and AF probes. For the peak-to-peak rectifier arrangement used the effective load presented to the circuit under measurement is usually taken to be one-quarter of the DC load resistance, in this case 22 megohms (R1). Thus the AC and RF probes load the circuit under measurement with about 5 megohms, with a few picofarads of capacity in parallel; this is negligible for most tuned circuits.

Two small 9-volt transistor radio batteries are adequate for the supply and as the total drain is only around 2 milliamps their life will be long. Lower voltages down to perhaps 3 volts could be used without much effect on performance bu the 9-volt supplies are recommended. Note that both positive and negative supplies are needed.

RF and AC Probes

These are of similar construction, the only difference being in the size of the capacitors, the insulation used and the provision of an AC-DC switch on the AC probe, allowing it to be used for DC measurements without changing cables. Figure 2 shows the circuits.

Each probe is assembled in a square IF transformer can about $1\frac{1}{8}$in square by $2\frac{1}{2}$in high. The open end carries the insulating panel supporting the whole circuit assembly and a sharp $1\frac{1}{4}$in spike to serve as a prod, while the shielded coaxial cable carrying the DC output to the meter is brought out through a hole in the solid end of the can. The shielded coax cable is essential here to avoid pick-up of spurious signals, hum, etc, because of the very high input resistance of the meter circuit. It is about two feet long for convenience and is terminated at the meter end with any small coaxial connector of good insulation resistance, to suit a matching socket on the front panel.

The diodes in both probes are type OA91, of the germanium point-contact type, with a peak voltage rating of about 115 volts. They are chosen in preference to silicon types, firstly because point-contact types are usable to higher radio-frequencies than other constructions; secondly, because their conductivity starts at lower voltage than silicon types, thus contributing to better sensitivity and scale linearity; and thirdly, because their slightly higher leakage gives faster response and quicker recovery after a momentary overload. Silicon diodes of good quality have such low leakage that if the probe capacitors accumulate an overload charge it takes quite a time to leak away, and in the meantime the meter is paralysed.

Capacitors in the RF probe are small tubular polystyrene types (not polyester) of excellent "Q" and insulation. Uncased silvered-mica types would be equally good but probably more bulky, and most available mica capacitors seem to come in "moulded mud" cases whose RF losses are not to be trusted. Ceramic capacitors are quite unsuitable because of high RF losses and should not be used.

The AC-DC probe uses flat lacquer-coated polyester foil capacitors of 1μF which fit the can fairly snugly; the writer used "ELNA" brand. This probe has a miniature slide-switch of the smallest available size mounted on one side wall, to change the probe from AC to DC operation simply by shorting input to output. The prod consists of a $1\frac{1}{2}$in length of 6BA threaded brass rod, sharpened to a point at one end and mounted at the other by nuts in the centre of a square panel of insulating material. The latter is mounted over the open end of the can by whatever mounting means are provided – lugs, spade bolts, or whatever.

The AC-DC probe used $\frac{1}{16}$in bakelite sheet for this panel, but in the RF probe a piece of polyester-glass-fibre sheet salvaged from a high-quality printed circuit was used. Sheet polystyrene would have lower losses, but is hard to find and ceramic materials, also good, are too difficult to work.

Inside the can the components are supported on a vertical panel of the same material, cut to fit the internal dimensions of the can and mounted to the top panel by a generous fillet of Araldite adhesive or else by small lugs and bolts. No specific constructional details are given as so much depends on the components actually available.

The AC-DC probe will need careful arrangement to leave room for the switch without fouling the rather large capacitors. For the screened leads the writer used PVC insulated, screened and PVC sheathed coaxial cable of 3.5mm overall diameter – a pleasantly flexible size which imposes minimum strain on the meter. A short length of tinned-copper braid equipped with an alligator clip is attached to the probe can to serve as earth connection to the circuit being measured.

A useful accessory is a further alligator clip provided with a 6BA nut mounted coaxially on its shank; this can be screwed onto the prod when a temporary connection is needed to a circuit, so leaving both hands free.

Construction

The prototype is assembled on a sloping panel measuring 7¼ x 6in, set at a 30-degree angle on the front of a pine-board box of sufficient depth to house the batteries. Figure 6 gives the panel layout.

The 741 IC is best used in a 14-pin DIL socket to avoid soldering directly to the IC pins, which can be a tricky job. These sockets exactly match the holes in 0.1in matrixboard or Veroboard, and their pins can easily be soldered to wires or strips on the reverse side. The writer favours Veroboard for this, and a suitable connection scheme for the circuit of Figure 1 is easily arrived at. If suitably dimensioned, this board, carrying most of the components, can be mounted directly on the meter terminals by loops of 18-gauge tinned copper wire.

Note carefully the small notch at one end of the 741 to identify it with the pin numbering diagram (see Figure 1). The socket should have a similar small notch at one end and the 741 must be inserted with the notches coinciding. There is nothing else to prevent the 741 being inserted back to front, which may do it no good at all. And when inserting or removing it, be very careful to do so with an absolutely straight pull or push, otherwise the fragile pins are almost certain to be crumpled or otherwise damaged, perhaps beyond repair. Once safely in its socket, the less it is moved the better, but the ability to remove it during servicing or circuit changes is a great advantage over permanent soldering into the circuit.

The Range Switch S1 should preferably be a miniature rotary wafer type to conserve space. The "ON-OFF" switch S3 must be a double-pole type, to switch both batteries. For this and also for S2 the writer used small slide switches, again to save panel space.

Because the instrument is virtually self-calibrating if 1pc resistors have been used throughout, no special instructions are needed in that case.

If the suggested variable resistors have been included, two courses are open. If a Wheatstone Bridge is available or can be borrowed, each combination of fixed and variable resistor can be set up to the exact value needed before wiring into the circuit and further calibration is then unnecessary. The values required are

R2 + R3 = 22k
R6 + R7 = 2.2 megohms
R8 + R9 = 220k.

R4 and R5 cannot well be made adjustable without badly complicating the scale-setting procedure; they must be of 1pc tolerance.

Failing the bridge, each DC range can be set at full-scale deflection by comparison with another voltmeter, starting with the 1-volt range. A supply of finely-adjustable voltage, or a potentiometer across a fixed supply can be used. The 1-volt range is best set up by feeding in an accurate one volt DC and adjusting R3 for full-scale reading, having first zeroed the pointer using the Zero Set knob R12. The 10-volt and 100-volt ranges can then be set up in succession by adjusting the appropriate resistor in the feedback chain (R7 and R9 respectively.)

With the DC ranges correct, the AC ranges should not need further calibration. However, if an AC voltmeter of known accuracy at 10 volts is available for comparison the 10-volt AC range can, if desired, be set correctly at 10 volts full scale by using, say, a 5k variable resistor for R14. The 1-volt scale will then also be correct and both will be accurate when using the RF probe over its designed frequency range.

Parts List

RESISTORS

R1	*22 megohms, 1pc (see text)*
R2	*20k 5pc*
R3	*5k linear trim-pot*
R4	*22k 1pc*
R5	*22 megohms, 1pc (see text)*
R6	*1.5 megohm, 5pc*
R7	*1 megohm linear trim-pot*
R8	*150k 5pc*
R9	*100k linear trim-pot*
R10 *R11*	*33k 10pc*
R12	*10k linear potentiometer, panel mounting*
R13	*890 ohms, 1pc (see text)*
R14	*2720 ohms, 1pc (see text)*

All fixed resistors are ½ watt carbon film or metal oxide film

CAPACITORS

2	*5000pF, tubular polystyrene pigtail type*
2	*1μF flat lacquered polyester foil, 200VW*

SEMICONDUCTORS

1	*Integrated circuit operational amplifier, uA741C or RS276-007/010/1741*
4	*Diodes, OA91 or RS276-1102/1136.AA117-118 1N618*

MISCELLANEOUS

Meter: 0-1 milliamp, SEW Type MR-85P or similar, to choic

S3 *S4*	*Miniature slide switches*
S1	*Any small one-pole three-way switch, panel mounting*
S2	*Miniature slide switch*
1	*Small coaxial socket, panel mounting*
2	*Coaxial plugs, to suit above*
1	*14-pin DIL socket*
2	*1-inch knobs, to choice*
2	*IF transformer cans, approx. 1⅛ inch square x 2¼ inch high*
4	*feet of 3.5mm screened and PVC sheathed coaxial cable*
2	*9-volt transistor radio batteries, P1 size*
2	*battery connectors for above*

Matrix board or Veroboard, 0.1in hole centres

Front panel, cabinet, wire, screws, nuts, screwed brass rod 6BA, three alligator clips.

SQUARE WAVE OSCILLATOR/PULSE GENERATOR USES 555 IC

With a little care, the popular 555 or Tandy Archer RS276-1723 monolithic timer IC can be used as a square wave oscillator or variable duty cycle pulse generator. The circuits and design details are given in this short note.

The popular 555 timer IC can be connected as an oscillator, as shown in Figure 1. The capacitor C is charged by current flowing through R1 and R2 in series, and the time the output is high will be given by:

$$T1 = 0.685(R1 + R2)C \text{ seconds}$$

where the resistance is in megohms and the capacitance in microfarads.

At the end of this time, a ground is applied to pin 7 from within the device, and the capacitor is discharged through R2 in a time:

$$T2 = 0.685R2C \text{ seconds.}$$

The pulse repetition time will therefore be:

$$T3 = 0.685(R1 + 2R2)C \text{ seconds.}$$

With this simple circuit, the duty cycle, or the ratio of the pulse "ON" time to repetition time will be:

$$(R1 + R2)/(R1 + 2R2).$$

The duty cycle can be set from slightly more than 50pc to nearly 100pc by a suitable choice of the values of R1 and R2. The lowest value of R2 will be determined by the maximum current that may flow into pin 7 without exceeding the dissipation rating, when it is held low during the discharge cycle.

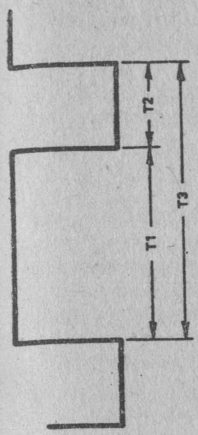

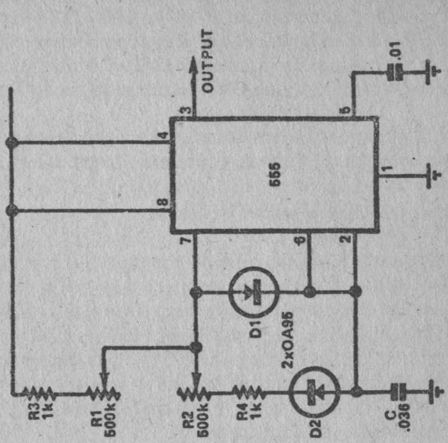

FIG. 3

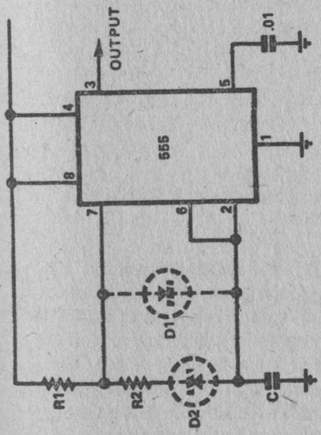

FIG. 1

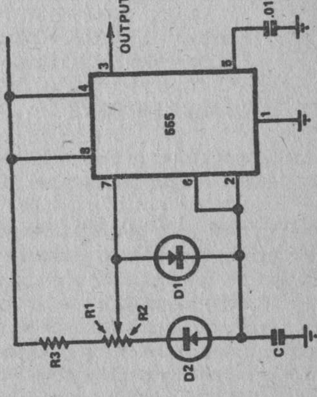

FIG. 2

The addition of the diode D1 (shown dotted) between pins 6 and 7, allows the capacitor to charge through R1 only, and discharge through R2, so the duty cycle is now R1/(R1 + R2). It can therefore be varied over almost the full range from 0 to 100pc. The charge time will not be completely independent of R2, as some current will flow through it due to the voltage drop across D1. This problem can be overcome by inserting a second diode D2 in series with the lower end of R2.

Two applications of this circuit are a fixed frequency, variable pulse width oscillator, Figure 2, and a wide range square wave generator, Figure 3.

Variable Pulse Width Oscillator

As the repetition time is set by the sum of R1 + R2, and the duty cycle by R1/(R1 + R2), the two resistors may be replaced by a potentiometer so that the duty cycle may be varied from 0 to 100pc without changing the repetition time (or frequency). R3 must be included to limit the maximum current into pin 7, and this will determine the minimum pulse width. In practice there is a small variation in the repetition rate, but this is of little importance in applications such as a pulse width control to vary the speed of a DC motor.

Wide Range Wave Generator

If R1 and R2 in Figure 1 are both varied, but remain equal in value, the repetition rate will be varied, while the output remains a square wave. A suitable circuit is shown in Figure 3. Resistor R3 is again added to limit the maximum current into pin 7, and R4 must then be included to balance its effect so that $(R1 + R3) = (R2 + R4)$.

The limits of the frequency range are given by

$$F \text{ min} = 0.73/(R1 + R3) \text{ Hz.}$$
$$F \text{ max} = 0.73/R3C \text{ Hz.}$$

and their ratio is $(R1 + R3)/R3$.

For the values shown in Figure 3, the lowest frequency is 40Hz, and the highest 20kHz, with a range of 500:1.

The output will only be a square wave if R1 and R2 track exactly. The maximum error is likely to occur if a log. potentiometer is used, to avoid crowding the frequencies at one end of the scale. This type of potentiometer is usually made with two linear segments so that it approximates a logarithmic curve and the tracking error is likely to be biggest at the crossover between the two segments. In practice, the ratio of off to on time may vary to as much as 40:60 instead of 50:50 over this part of the range. If a reasonably accurate square wave is needed over the full range, linear potentiometers will provide closer tracking.

THE DIGI DESIGNER

If you're involved with logic circuit design using digital ICs, you'll find this little unit irresistible. It provides all the basic requirements for fast and easy lash-up of new circuits, yet also doubles as a logic trainer.

A breadboarding aid of one sort or another is almost essential when one is attempting to try out even a simple circuit using the usual dual in-line ICs. Without such an aid, progress tends to be painfully slow, and the ICs and other components are likely to be damaged due to repeated lead bending and soldering operations. If you've tried it, you'll know what I mean.

In its most basic form, a digital bread-boarding aid may be little more than a scheme which allows the ICs to be plugged into sockets, with their connections brought out to convenient solder pads or terminals. But before it can be put to much use, this basic setup must be supplemented with a power supply, a source of test signals, provision for monitoring logic levels and a means of setting control signals.

With the Digi Designer, we have provided not just a basic breadboard aid, but all of these supplementary facilities as well. Although it measures only 19 x 21.6 x 10.2 cm. the Digi Designer combines a multi-device breadboard socket, a power supply, four logic level monitors with lamps, a dual-polarity clock generator with six switched frequencies, four logic level switches and two dual polarity pushbutton pulsers with bounce suppression.

It thus forms a complete miniature logic laboratory, with which you can easily and quickly lash up – and test – any number of possible circuits and logic configurations. All you need is the ICs and some lengths of hookup wire, and away you go!

Needless to say, it would also be very suitable for use as a logic trainer in schools and colleges. Provided with a few standard gate and flip-flop ICs, it can readily be used to demonstrate virtually all of the basic laws of digital logic.

Heart of the Digi Designer is a multiway breadboard socket, Based on a precision moulding of high-impact plastic, the socket provides some 328 nickel-silver spring contact clips, grouped to allow multiple connections to IC pins and other devices.

The clips are designed to provide low resistance contact with IC pins, transistor leads, component pigtails and solid-core hookup wire from 22 to 26 gauge. At the same time they require only a relatively low insertion force, which should avoid strain to multi-pin ICs and similar devices.

The fact that ordinary solid-core hookup wire may be used for the interconnections makes the unit very convenient in use. There are

DIGI DESIGNER

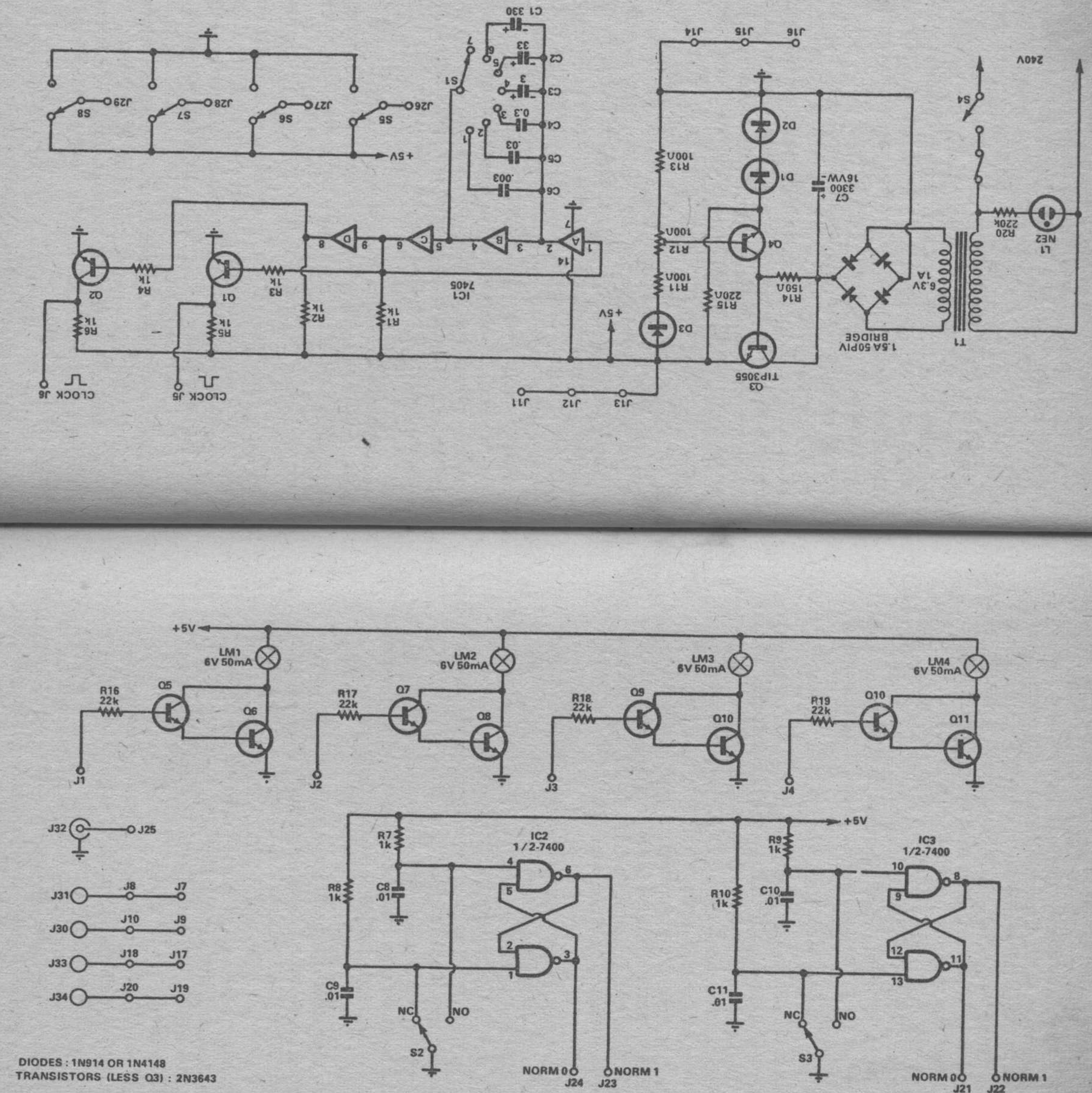
CLOCK J6
CLOCK J5
R6 1k
R5 1k
R4 1k
R3 1k
R2 1k
R1 1k
Q1
Q2
IC1 7405
A
B
C
D
+5V
J11
J12
J13
J26
J27
J28
J29
S5
S6
S7
S8
S1
C1 330
C2 33
C3 3
C4 0.3
C5 .03
C6 .003
D1
D2
D3
R11 100Ω
R12 100Ω
R13 100Ω
R14 150Ω
R15 220Ω
Q3 TIP3055
Q4
C7 3300 16VW
1.5A 50PIV BRIDGE
6.3V 1A
T1
R20 220k
L1 NE2
S4
240V
J14
J15
J16
+5V
LM1 6V 50mA
LM2 6V 50mA
LM3 6V 50mA
LM4 6V 50mA
R16 22k
R17 22k
R18 22k
R19 22k
Q5
Q6
Q7
Q8
Q9
Q10
Q10
Q11
J1
J2
J3
J4
J32
J25
J31
J8
J7
J30
J10
J9
J33
J18
J17
J34
J20
J19
IC2 1/2-7400
IC3 1/2-7400
R7 1k
R8 1k
R9 1k
R10 1k
C8 .01
C9 .01
C10 .01
C11 .01
NC
NO
S2
S3
NORM 0
NORM 1
J24
J23
J21
J22

DIODES : 1N914 OR 1N4148
TRANSISTORS (LESS Q3) : 2N3643

The circuit of the Digi Designer is shown above. While simple and straightforward, it provides all of the required features with no sacrifice in performance. Note the dual polarity clock generator.

no special patchcords to become lost or broken, and leads of any length may be fabricated with a minimum of effort right on the spot.

The socket is mounted centrally on the aluminium front panel of the Digi Designer, with the lamp monitors, clock generator and mains switch, pilot and fuse above it. On the lower part of the panel are the logic switches, the pushbutton pulsers and a coax socket/pin jack combination which may be used either to feed external signals into the unit, or extract signals from it.
Immediately adjacent to the breadboard socket are pin jacks supplying plus 5V and ground, also four pairs of pin jacks connected to screw terminals. The latter again provide a convenient means of interfacing the circuitry with external instruments or other Digi Designers.

The circuit of the unit regeals a very practical engineering approach to its design. Considering the functions performed, there are surprisingly few parts and the wiring is quite straightforward. In fact the Digi Designer is a good example of the all-too-often forgotten art of "getting the most from the least".
The power supply is a fairly conventional series regulator using a plastic version of the well-known 2N3055. Two normal silicon diodes are used as the reference, with a third in the feedback divider presumably for temperature compensation.

The clock generator uses three of the inverters from a 7405 IC in cascade, with a capacitor across the centre element. The capacitor value is switched to provide the six frequencies: 1, 10 and 100Hz, and 1, 10 and 100kHz. Two discrete transistors are used as output buffer stages, with a fourth inverter from the 7405 used to provide the complementary output.
The lamp indicator circuits are very simple, each comprising only a 6V/50mA lamp, an input resistor and two general purpose NPN transistors connected in the Darlington configuration.

In contrast with these, the pushbutton pulser circuits seem to reveal a BBbelt and braces" design philosophy. Each uses two gates from a 7400 IC as a simple R-S flip flop, which is normally quite capable of giving full suppression of button contact bounce. But just be sure, the designer has also fitted the inputs with R-C integrator circuits. This may not be a case of gliding the lily, though, as the SPDT buttons used are of a rather simple and rudimentary type.

Inside the case, all the components are mounted on the rear of the front panel – including the power transformer. The power supply and clock generator wiring are on a fairly large printed wiring board, while the lamp drivers and pushbutton circuits are mounted on two much smaller boards. These are supported solely by the wiring, but this should cause no problems in view of their modest size and weight.

5-STAGE LOGIC PROBE HAS 7-SEGMENT READOUT

For anyone working with logic circuits, it should be invaluable. Detecting no less than five distinct logic conditions, it indicates them simply and elegantly using a 7-segment readout.

The logic probe described is the latest in a line of probes I have constructed. Working with industrial control equipment using TTL devices I have tried most available logic probes and noted the following.

Probes including pulse stretching circuitry can be a problem as it often proves necessary to move the probe supply leads to the IC being tested, to prevent false triggering. Where stretching is included, provision to disable should also be provided.

Secondly the probe must show a large number of logic states or it will be put aside for an oscilloscope.

Finally the probe must obtain power from the circuit under test, be hand held, and have protection on the input and output against the most common forms of abuse.

The probe developed is unusual in that it will detect 5 logic states and display these on a 7-segment LED display. For details of the display and logic states covered, see Figure 1.

Pulse stretching circuitry was omitted from this device, as I use a separate probe for this form of measurement.

Input Detection

Individual detectors for the following 4 logic states are provided.

1. Short to positive (TR2 & 3)
2. True logic '1' (TR4)
3. True logic '0' (Gate B1)
4. Short to negative (TR6 & 7)

The 5th state (open circuit or incorrect logic level) is displayed if all the above states are absent.

Detecting a short to positive or a short to negative relies on a similar arrangement, and only the short to negative circuit will be detailed.

Transistors TR6 & 7 form a Schmitt-Trigger with TR6 normally in conduction. Placing the probe on a voltage less than the 50mV dropped across the 10 ohm emitter resistor, diverts base current from TR6, taking it out of saturation, bring TR7 into conduction.

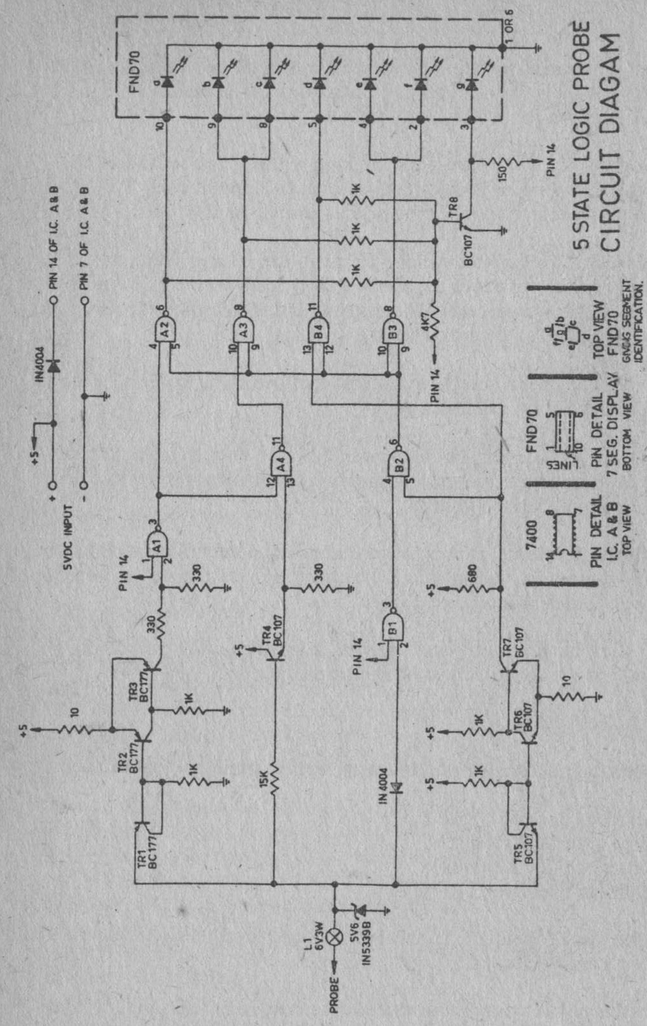
5 STATE LOGIC PROBE
CIRCUIT DIAGAM
FND70
a
b
c
d
e
f
g
1 OR 6
PIN 14 OF I.C. A&B
PIN 7 OF I.C. A&B
IN4004
+5
SVDC INPUT
+
-
A1
A2
A3
A4
B1
B2
B3
B4
PIN 14
1K
4K7
150
TR8
BC107
TR1
BC177
TR2
BC177
TR3
BC177
TR4
BC107
TR5
BC107
TR6
BC107
TR7
BC107
10
330
15K
680
IN 4004
L1
6V3W
5V6
IN5339B
PROBE
7400
PIN DETAIL
I.C. A & B
TOP VIEW
FND70
LINES
PIN DETAIL
7 SEG. DISPLAY
BOTTOM VIEW
TOP VIEW
FND70
GINGS SEGMENT
IDENTIFICATION.

The complete probe circuit. There are four detectors, two for the "short to rail" states. Circuit is fully protected.

Transistor TR5 has a dual role, normally behaving as a blocking diode, but when forward biased cancelling the base emitter drop of TR6.

To activate the Schmitt-Trigger the probe must contact a potential below 50mV with a shunt resistance less than 10 ohms.

To detect logic '1' the emitter follower TR4 drives a standard TTL input. Whereas a gate will respond to logic '1' and open circuit in a similar manner, TR4 will only follow a true logic '1'. The base emitter drop of TR4 also ensures a full logic level must occur to initiate a change of state.

Logic level '0' is taken to a standard 74 series gate input for detection. Including a diode in series with the gate reduce the '0' level accepted to the worst cas level.

Overvoltage Protection

The input probe is by-passed by a zener diode, thereby limiting the maximum voltage to reach the input detectors to 6 volt positive and 0.5 volt negative. Lamp L1 is the load for the zener, having resistance below 2 ohms in normal use, but increasing to 25 ohms with 20 volts on the probe. Fusing of the lamp takes place as the voltage increases. These components are for protection only and voltages above 5 volts should be avoided in normal operation.

Protection from reversal of the supply takes the form of a diode in series with the IC positive pin. Provided the reverse voltage does not exceed 5 volts it is not necessary to similarly protect the input detectors – indeed the short to positive detector would cease to function with diode protection.

Output Decoding

With detection of the input states completed the information enters a gate network for decoding to a form suitable for the 7-segment display. The gates determine which segment (or segments) should be illuminated for a given input. In addition, only one level must be displayed at a time.

The short to negative or positive rail is only a specialized instance of a logic '0' or '1' and hence more than one input detector will be activated by a short condition. However the gating is arranged so the short detectors override the normal logic detectors.

Four gates decode segments 'A' through 'F' while segment 'G' (open circuit indication) is controlled by TR8. The latter transistor by-passes all current from segment 'G' while any other segment is illuminated.

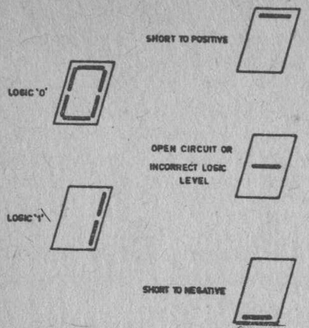
SHORT TO POSITIVE
LOGIC '0'
OPEN CIRCUIT OR INCORRECT LOGIC LEVEL
LOGIC '1'
SHORT TO NEGATIVE

fig. 1: Readout of the 5 states detected.

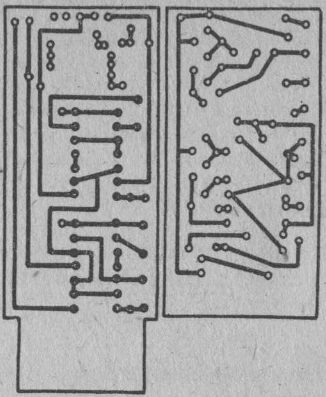

Fig. 2 (above): The board patterns, shown actual size. Below is a close-up of the wired-up board assembly.

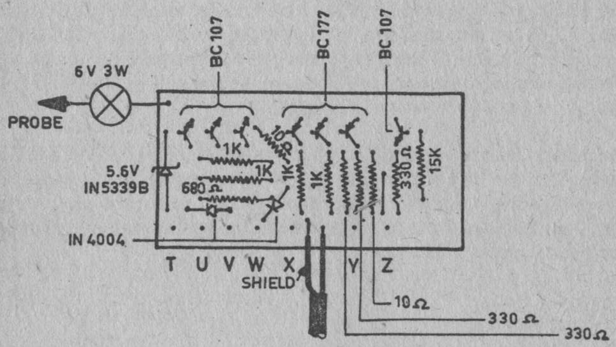
BC 107
BC 177
BC 107
6V 3W
PROBE
5.6V
IN5339B
IN 4004
1K
1K
680 Ω
1K
1K
10
330 Ω
15K
T U V W X
SHIELD
Y Z
10 Ω
330 Ω
330 Ω

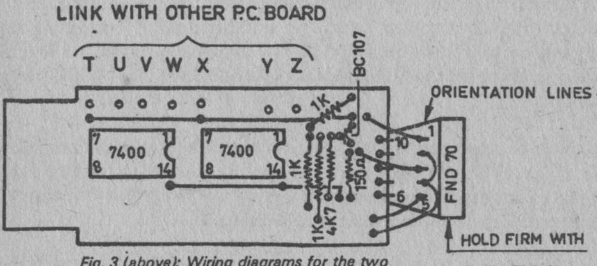
LINK WITH OTHER P.C. BOARD
T U V W X Y Z
BC107
ORIENTATION LINES
7400
7400
1K
1K
150 Ω
10
6
1
FND 70
1K
4K7
HOLD FIRM WITH
5min EPOXY

Fig. 3 (above): Wiring diagrams for the two boards, also showing interconnections.

The display segments are driven directly from the TTL outputs, current limiting being effected by the internal level '1' gate resistance. The circuit has been arranged so that the power dissipation is divided equally between ICs A and B.

The brightness difference that should in theory result from operating certain segments in parallel while others are driven singly, is not discernible in operation.

General Information

A few points may not be clear from the foregoing description.

A short to rail value of 50mV/10 ohms was taken rather than a true short circuit. This overcomes any voltage drop on the supply lines and enables input protection to be incorporated. In practice the '0' level will always exceed 50mV and a '1' level – even in open collector applications – will be many times this value below the rail voltage.

The input loading of this device is 1UL for all conditions excepting a short to positive or negative rail. It is assumed any short to rail will pass the extra 5mA required for these detectors.

Supply current for this probe can reach 100mA, and while in most instances this will be available from the unit under test, it can be provided by a separate battery or supply. However if this is done, ensure that the negative of the extra supply is interconnected with the negative of the unit under test.

The 7-segment display legend, while being easily interpreted is also capable of being read upside down. To prevent ambiguity the top and bottom should be identified by a red dot placed adjacent to segment 'a' and a green dot placed adjacent to segment 'd'.

Construction Details

The accompanying photographs give an idea of construction used for the prototype. The components are mounted on two PC boards measuring 28mm wide by 50mm long. A miniature plastic torch case is used as a housing, requiring only the addition of a probe at the reflector end, and a hole for the 7-segment display to be viewed at the opposite end.

Approximately a metre of shielded cable is provided or the power lead, this terminating with E.Z. mini hooks or similar.

Detail of the PC boards and their assembly is given in Figures 2 and 3. Care should be taken to keep all components as close to the PC boards as possible or trouble may occur when fitting the boards into the housing. Metal can transistors will require an insulating sleeve to prevent adjacent cans shorting.

1	*6 volt 3 watt lamp*
2	*Mini Hooks*
2	*10 ohm ¼ watt resistors*
1	*150 ohm ¼ watt resistor*
3	*330 ohm ¼ watt resistors*
1	*680 ohm ¼ watt resistor*
7	*1K ohm ¼ watt resistor*
1	*4.7k ohm ¼ watt resistor*
1	*15K ohm ¼ watt resistor*
1	*Penlite torch 20 x 35 x 95mm*
1	*Shielded mic. Cable (1 metre long)*

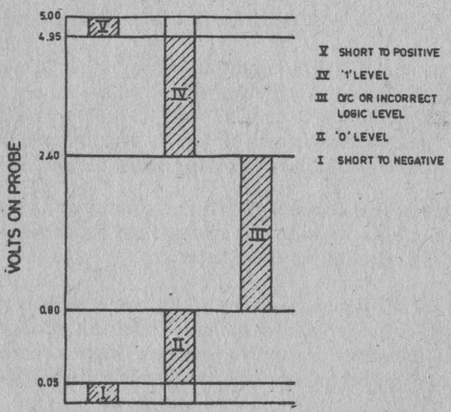

Fig. 4 *The voltage levels which define the five states detected by the probe.*

REPAIRING A CRACKED BOARD

A cracked printed wiring board may usually be repaired with a little care. If both the base and the copper are cracked, drill small holes through each run of copper, on opposite sides of the crack.

From thin tinned copper wire, make up a staple for each pair of holes. The staples should be just wide enough to bridge the two holes when the crack is closed. Make the sides of the staple fairly long.

Push the two halves of the board together, push one staple through the holes from the copper side, then gently twist the ends of the staple together on the component side, with a pair of pliers. Solder the twist, and clip off the surplus. Solder the tinned copper wire to the copper pattern. Where more than one staple is required, insert alternate staples from opposite sides.

LOW COST SCALER EXTENDS DIGITAL COUNTERS TO 200MHz

Here is a new design which should be of special interest to lab workers, mobile two-way radio technicians and VHF radio amateurs. It is easily built at low cost, and uses a state-of-the-art ECL integrated circuit to extend the range of your present DFM or counter to beyond 200MHz.

You may remember, I described our first frequency scaler to extend the range of existing digital counters. It used some of the first emitter-coupled logic (ECL) integrated circuits, which had then just been released. By using some seven of these devices I was able to achieve a maximum frequency of slightly more than 120MHz, with a division ratio of either 10:1 or 20:1 to cope with counters with maximum counting rates as low as 6MHz.

The performance of this design was considered quite good at the time, particularly in view of its modest cost, and many of the units were built up. They have been used in many research, development and teaching laboratories as well as by mobile radio technicians and VHF radio amateurs. However, integrated circuit technology is progressing at a very high rate, and it is not surprising that since the original unit was developed, higher performance devices have been produced.

This fact made itself evident some weeks ago when I was looking through literature describing some of the recently released devices. Despite the relatively short time that has elapsed since it was described, the 1969 scaler is now well and truly obsolete. Hence the reason for developing the new design described in this article.

The new scaler is completely different from the old one in terms of circuit configuration. It uses only three ICs, two transistors and five diodes – three of which are in the power supply. In terms of cost and complexity, it should be both cheaper and easier to build than its predecessor.

Yet at the same time its performance is significantly improved. The upper frequency limit is now slightly above 200MHz, thus extending well past the 144-148MHz amateur band and the upper VHF mobile radio band. The frequency division ratio is now 100:1, so that the scaler will extend to 200MHz any counter capable of counting up to 2MHz at present.

Input sensitivity is high, less than 800mV peak-to-peak input being required for reliable division up to 200MHz. Input impedance is low, approximately 50 ohms, but at VHF this is generally no problem. In most cases a small "sniffer loop" or length of wire for capacitive pickup at the end of the input cable will be quite sufficient. The input is protected from overload damage.

Output from the scaler is approximately 4.5V peak-to-peak, with a low output impedance. This should be more than adequate for all but the most insensitive counters.

Basically the scaler consists of two cascaded decade divider stages, a very high speed input stage followed by a lower speed stage. If the unit is to be used with a counter having a basic frequency capability of 20MHz or better, the second decade may be omitted, reducing the cost and simplifying the unit still further.

In this "shortened" form the scaler may easily be built into existing 70MHz frequency meters to convert them at low cost into full 200MHz instruments. This is a simple procedure, and is described later in the article.

Heart of the scaler in both its "long" and "short" versions is a new ultra-high speed ECL decade divider IC, which performs the input divider function. This device is the Fairchild Semiconductor type 95H90, for which Philips also have an equivalent, type GHJ121.

The 95H90 is a medium-scale integration (MSI) device which contains four very high speed J-K fip-flops. These use ECL circuitry and are arranged for either BCD decade counting, 10:1 frequency division, modulo-11 counting, or 11:1 frequency division. It has a guaranteed minimum input counting rate of 220MHz, and typical devices can exceed 300MHz in a suitable circuit configuration.

Here the device is used for straight 10:1 frequency division, and the simple circuit configuration tends to limit operation to about 210MHz.

The 95H90 is not a cheap device, costing around $15 plus tax. However as it needs only the addition of a few minor components to provide a full 200MHz-plus counting or divider decade, it is still significantly cheaper than any other alternative approach.

The input of the scaler connects to the input of the 95H90 via two coupling capacitors in series, with the junction of the capacitors connected to ground via inverse-parallel diodes. These protect the 95H90 against input overload damage.

A DC bias voltage is also fed to the input of the device to set the conditions for maximum input sensitivity. The bias circuit is decoupled by a ferrite bead and a small RF choke. Without these components, spurious coupling back through the supply line tends to degrade the 95H90 performance and reduce the maximum frequency of operation.

The second decade of the scaler uses TTL (transistor-transistor logic) devices: a 9001 and a 7490 or 9390 or FJJ141. The 7490 / 9390 / FJJ141 is actually a full MSI decade divider itself, but with a guaranteed maximum counting rate of 10MHz. Used alone for the second decade, it could in the worst case limit the upper frequency of the scaler to 100MHz. For this reason the input flip-flop of the device is left unused, and a type 9001 flip-flop IC used instead. With a typical maximum counting rate of 50MHz, the 9001 ensures that the upper frequency limit of the scaler is set solely by the 95H90 input device.

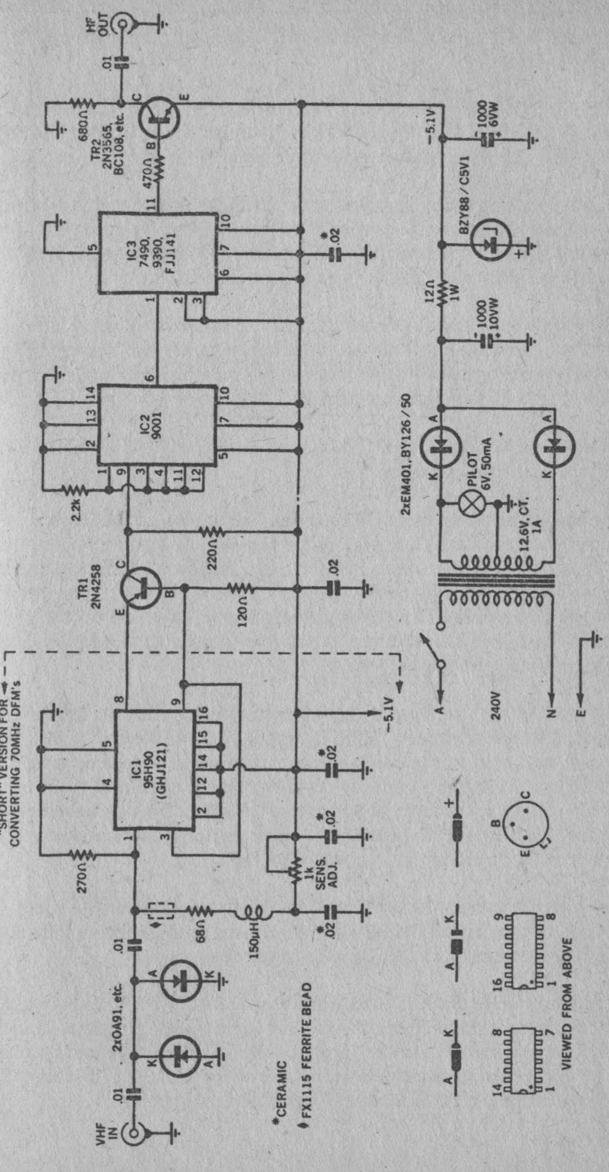
HF OUT
.01
680Ω
TR2 2N3565, BC108, etc.
C
B
E
470Ω
11
IC3 7490, 9390, FJJ141
5
1
2
3
6
7
10
*.02
−5.1V
1000 6VW
BZY88 / C5V1
12Ω 1W
1000 10VW
IC2 9001
14
13
2
1
9
3
4
11
12
6
7
10
5
2.2k
2xEM401, BY126 / 50
K
A
PILOT 6V, 50mA
12.6V, CT. 1A
220Ω
TR1 2N4258
120Ω
.02
"SHORT" VERSION FOR CONVERTING 70MHz DFM's
−5.1V
A
240V
N
E
IC1 95H90 (GHJ121)
4
5
8
9
16
15
14
12
2
3
1
270Ω
1k SENS. ADJ.
*.02
68Ω
150μH
.01
2xOA91, etc.
VHF IN
*CERAMIC
FX1115 FERRITE BEAD
14
8
1
7
16
9
8
VIEWED FROM ABOVE
B
E
C

200MHz DIGITAL SCALER

The circuit of the scaler, which uses only three ICs and two transistors.

Because the second decade uses TTL devices whose logic levels are different from those of ECL, level translation is required between the two stages despite the operation of both from a common positive-ground 5V supply. The translator used is a simple inverter using a 2N4258 very high speed PNP switching transistor.

To prevent disturbance of the second divider decade due to output loading, a buffer stage is used between the 7490 and the output connector. The buffer uses a 2N3565, BC108 or similar general-purpose NPN silicon transistor.

Power for the scaler circuits is provided by a simple supply using a full-wave rectifier connected to a centre-tapped 12.6V transformer winding. A 400mW zener diode is used to stabilise the supply to the ICs to a nominal 5.1V. The current drain of the scaler circuits is approximately 210mA.

With the exception of the power transformer, mains switch, pilot lamp and input / output connectors, all of the scaler components and wiring are mounted on a small printed wiring for ease of assembly. The board measures 4½ x 2¾in (114 x 70mm) and is coded 72 / s10.

The prototype of the "long" version of the unit was built in a small plastic-and-aluminium instrument case,

Inside the case the printed board is attached to the rear of the front aluminium panel, via four 1¼ x ⅛in Whitworth screws with multiple nuts used for spacing. The lower two screws are also each used to fasten one side of the input and output connectors, to reduce the number of screw heads visible on the front panel. The power transformer is mounted on the bottom of the plastic case, as far to the rear as possible to provide maximum clearance between it and the wiring board.

The prototype "short" version of the scaler has been built into the original 70MHz digital frequency meter. Here it simply consists of the "front half" of the board, mounted in the DFM in place of the small board originally used to perform 2:1 input frequency division. The scaler board is simply sawn in half along a line drawn between two "notches" which have been provided for this purpose in the etched copper pattern. The notches are approximately midway along the long edges of the board.

The scaler board has been designed with its width the same as the small divider board in the original DFM, and with its mounting hole centres also the same. It is therefore a simple drop-in replacement.

Because the input gate and first counting decade of the DFM already use ECL devices, there is no need for the scaler's logic translator stage. Thus in this short form the scaler consists simply of the 95H90 device and its power supply and input biasing components. It is powered from the−5V supply already provided in the DFM for the other ECL devices.

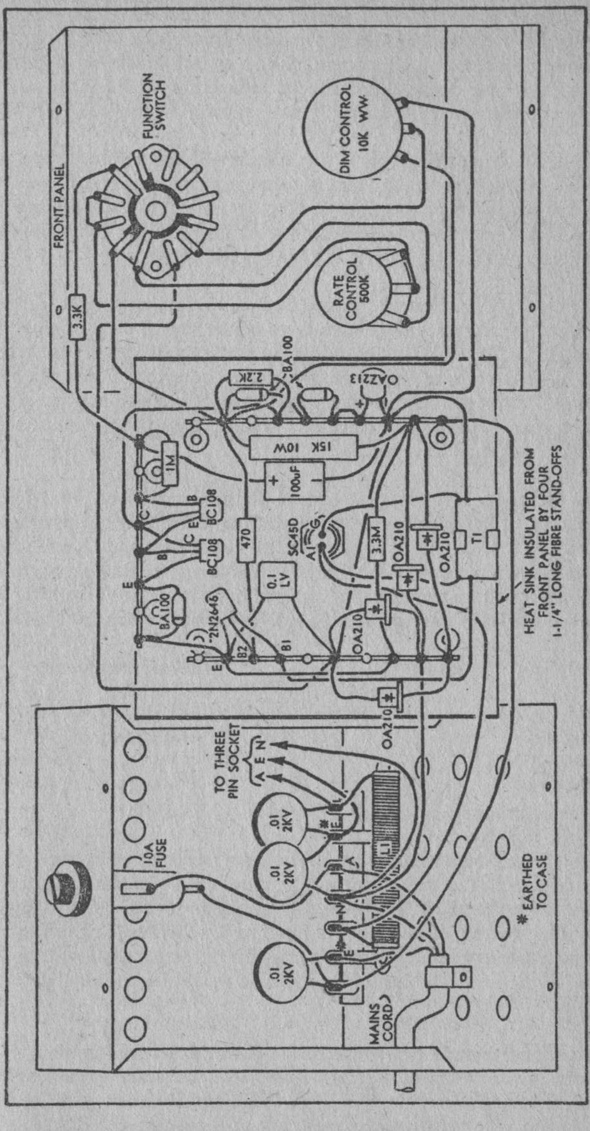
FRONT PANEL
FUNCTION SWITCH
DIM CONTROL 10K WW
RATE CONTROL 500K
3.3K
2.2K
BA100
OAZ213
15K 10W
1M
100uF
BC108
BC108
B
C
E
470
SC45D
A1
G
3.3M
OA210
OA210
OA210
OA210
T1
0.1 LV
BA100
2N2646
B1
B2
E
HEAT SINK INSULATED FROM FRONT PANEL BY FOUR 1-1/4" LONG FIBRE STAND-OFFS
TO THREE PIN SOCKET
A E N
10A FUSE
.01 2KV
.01 2KV
.01 2KV
L1
A
N
E
MAINS CORD
* EARTHED TO CASE

The lower input frequency limit for reliable operation of the 95H90 device with sinewave inputs is approximately 1MHz, so that it is really not suitable for use with all input signals. However, on the other hand, the existing input shaping circuitry of the DFM is not capable of working at VHF, so that it also cannot be used with all input signals. A switch must therefore be used to connect either into circuit alternatively between the DFM input connector and the gating circuitry. The switch may logically be marked "HF-VHF".

Note that the correct position of this switch will depend purely on the frequency of the input signal, and not on the range selected. Even when one switches down to lower ranges to look at the less significant digits of a VHF signal, it is still necessary to have the switch in the "VHF" position, as the input circuitry must still handle the fill signal frequency.

On the prototype instrument, rather than add a further switch to the front panel, I merely changed the connections to the switch used originally to select either AC or DC input coupling, so that this switch has become the "HF-VHF" switch. The DC input facility had never been used since the instrument had been built, so that it is no great loss. The original coupling capacitor is simply wired permanently in series with the HF input circuitry, as shown in the small diagram.

Because the scaler involves a 10:1 input division instead of the 2:1 division used originally, the appropriate timebase frequencies for correct readings will now be 100kHz and its sub-multiples, not 50kHz and its sub-multiples. Happily these frequencies are already available on the timebase board, so that it is simply a matter of changing over the connections from the range switch wafer S1c to the timebase board. The new frequencies used are as follows:

Range (HF)	Timebase Frequency
0-1999Hz	1Hz
0-19.99kHz	10Hz
0-199.9kHz	100Hz
0-1.999MHz	1kHz
0-19.99MHz	10kHz
0-35MHz or more	100kHz

Strictly it would be necessary to divide these frequencies by ten when the input switch is in the VHF position, to compensate for the 10:1 division introduced by the scaler. However,this would involve either a further wafer on the range switch, or another 10:1 frequency divider IC. Even then it would be necessary to add further provision to reduce the lower sampling rate, as for the lowest rating the gating time would be no less than 10 seconds.

This additional complexity seems to me unwarranted, when one can easily perform a mental x10 multiplication whenever the scaler is switched into circuit. All one has to remember is that readings obtained with the input switch in the VHF position are one-tenth of their correct value; thus the lowest range really reads in units of 10Hz instead of 1Hz, for example, whil the highest range reads directly in HMz and not in units of 100kHz.

Note that the original 2:1 input divider board also included a flip-flop used to divide the lower sampling rate to 0.125Hz. This rate was necessary to allow correct operation with the 0.5Hz timebase frequency used for the lowest frequency range. As the lowest timebase frequency will now be 1Hz, this low sampling rate is no longer required. The 0.25Hz output from the input board may now be taken directly to the sampling switch, giving a lower sampling rate of 1 every four seconds.

After the scaler is completed, there is only one adjustment required to ensure that it is operating correctly. This is exactly the same for either the "long" or "short" version: adjustment of bias for maximum input sensitivity.

The best way to make the adjustment is to feed a VHF signal into the scaler input, with its output connected either to a digital counter or to an oscilloscope. Then the bias preset pot on the scaler board is simply adjusted for reliable operation with the smallest possible input voltage. As viewed on a counter, this will be evident as the bias setting at which one can reduce the amplitude of the VHF input furthest before the reading starts to fall from its correct value and being to vary randomly. On an oscilloscope, it will be the setting where the input can be reduced furthest before the output wave form begins to exhibit "drop-out". Its as simple as that!

PARTS LIST

1 Instrument case, as required.
1 Printed wiring board, 4½ x 2¾in, 72 / s10.
1 Stepdown transformer, secondary 12.6V CT at 1A.
2 Coaxial connectors, panel mounting.
1 Miniature SPST toggle switch, 240V rating.
1 Ferrite bead, type FX1115 or similar.
1 6V 50mA pilot lamp bezel.
1 150uH peaking inductor.

SEMICONDUCTORS
1 95H90 or GHJ121 decade divider IC.
1 9001 high speed flip flop IC.
1 7490 or 9390 or FJJ141, RS276-1808 decade divider IC.
1 2N4258 very high speed PNP silicon transistor.
1 2N3565, BC108, RS276-2009 or similar.
2 EM401, RS276-1139, BY126 / 50 silicon diodes.
2 OA91, RS276-1102/1136 or similar germanium diodes.
1 BZY88 / C5V1 or similar 5.1V 400mW zener diode.

RESISTORS
5% half watt: 1 x 68 ohm, 1 x 120 ohm, 1 x 220 ohm, 1 x 270 ohm, 1 x 470 ohm, 1 x 2.2k.
1 12 ohm 1 watt.
1 1k miniature tab pot.

CAPACITORS
3 .01uF disc ceramic
5 .022uF disc ceramic.
2 1000uF 10 VW electrolytic

MISCELLANEOUS
Mains cord and plug, clamp for cord, connecting wire, solder lugs, screws, nuts, etc.

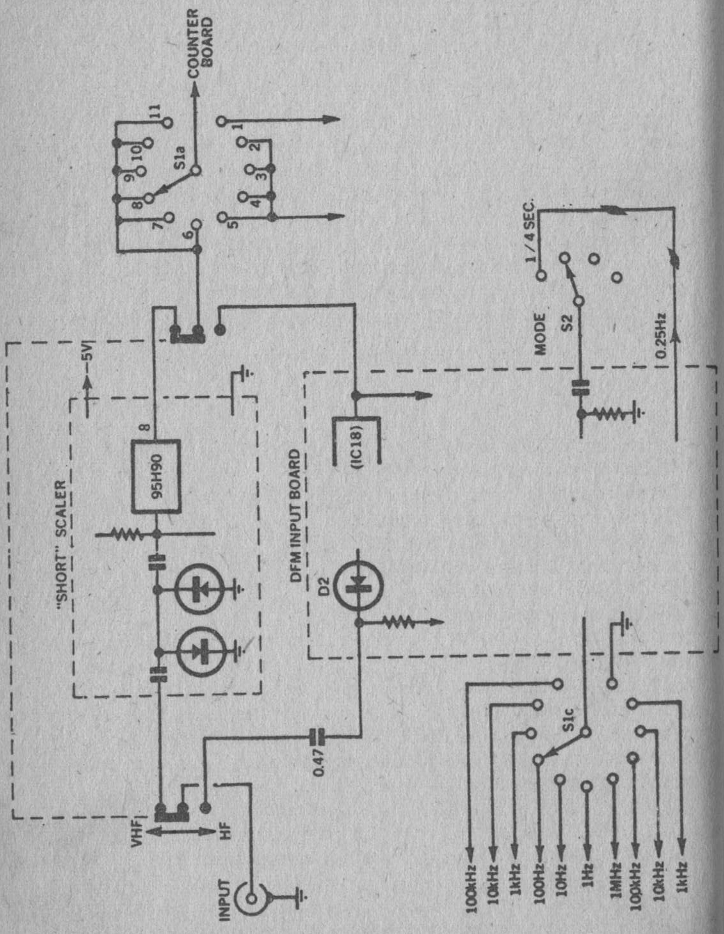

How to convert the DFM to count up to 200MHz. Few parts are needed.

BUILD OUR DIGI-METER: A DVM WITHOUT TEARS

Here's a basic DVM which almost anyone should be able to put together with complete success. Despite this its performance is very impressive.

Not very long ago, before integrated circuits came along, digital voltmeters or "DVMs" were large and costly instruments. Because of their cost and complexity they tended to be found only in the rarified atmosphere of the research laboratory. The idea that a DVM would ever be used for everyday measuring jobs seemed a futuristic pipe-dream; as for building a DVM into a piece of equipment as a direct replacement for a moving coil meter movement, that was virtually unthinkable!

Yet, thanks to modern semiconductor technology, and the increasingly complex ICs which it has produced at ever-lower costs, both these dreams have now become a reality. Low cost multi-range DVMs are now available for little more than the cost of a good analog multimeter, and are already being used even in servicing and hobbyist situations. Not only this but more and more original equipment manufacturers ("OEMs") are fitting single-range DVM "panel meters" into equipment in places where once they would have used moving coil movements.

The latest generation of these digital panel meters are truly impressive little devices, a tribute to modern semiconductor devices and advanced circuit design. A particularly outstanding example is the Model AD2010/E recently released by Analog Devices, Inc., of Norwood, Massachussetts.

The size of the AD2010/E is a compact 45 x 76 x 20mm, less than many moving coil movements. It weighs only 113 grams (4 ounces). Yet crammed into this compact package is a complete 3½-digit DVM with a full-scale sensitivity of 199.9mV and an accuracy of 0.05pc plus/minus 1 digit in the least significant position. It also offers such features as bright seven-segment LED readouts with 7mm-high digits, display storage, automatic polarity selection, automatic zero correction, internal overload protection, blanking of leading zeros, BCD outputs for data logging, a DC input resistance of 100 megohms, a normal-mode 50Hz hum rejection of 40dB, and a common-mode rejection of 60dB.

In short, it is a true state-of-the-art single range DVM, compressed into a little black box half the size of many conventional meter movements.

As soon as we saw the AD2010/E a few weeks ago, we wondered whether it might be possible to use it as the heart of a basic and easy to build multi-range DVM for the home constructor. When we approached Parameters with the idea, they were confident that it could be done. And so, with their help and co-operation, the Digi-Meter was born.

As you can see from the photographs and the circuit, it is essentially little more than the AD2010/E fitted into a case, with a power supply and a simple input attenuator circuit to provide four higher voltage ranges. All the real credit must go to Analog Devices! To make the instrument just that more attractive, however, we have added a simple constant-current source circuit, to provide four additional resistance ranges.

The circuit itself must surely be about as simple and as straightforward as one could get with a practical multi-range meter.

As mentioned earlier, the basic DC input resistance of the AD2010/E is 100 megohms, and its sensitivity 199.9mV. To provide the higher voltage ranges we have used an input voltage divider system similar to that in the more complex DVM. Three series-connected 3.3M resistors are switched into circuit in series with the AD2010/E input, with different shunt resistors switched directly across the input for each range.

It would not be sufficient simply to use standard close tolerance fixed resistors for the range divider shunt resistors, because the closest tolerance normally available is 1pc; this would quite seriously degrade the basic 0.05pc accuracy of the AD2010/E. Hence it is necessary to use preset adjustable pots for each range, to allow more accurate calibration. It is still desirable to use high stability resistors, if they are available, but more in the interests of stability than for close tolerance.

The advantage of the particular input attenuator system we have used is that each of the high voltage ranges may be calibrated independently; there is no adjustment interaction (apart from the internal calibration of the AD2010/E, which may be regarded as a reference). The fixed 9.9M resistance in series with the input on all the higher voltage ranges also gives additional overload protection. Note that the input resistance on the higher voltage ranges varies between about 10 and 11 megohms, which is quite suitable for most measurements.

The constant current source used for the resistance range is a simplified version of that used in the more elaborate DVM, and is also designed to operate from the same 5V DC supply used for the AD2010/E. It uses a single forward-biased 1N914 or similar silicon diode as the basic voltage reference, with a second diode-connected transistor to balance the Vbe drop of the actual current source device. Both transistors are high-gain PNP silicon devices such as the 2N3638A. Preset resistors are used again in the emitter feedback circuit, to allow accurate calibration of the four ranges. The currents for the four ranges are 1μA, 10μA, 1mA and 10mA.

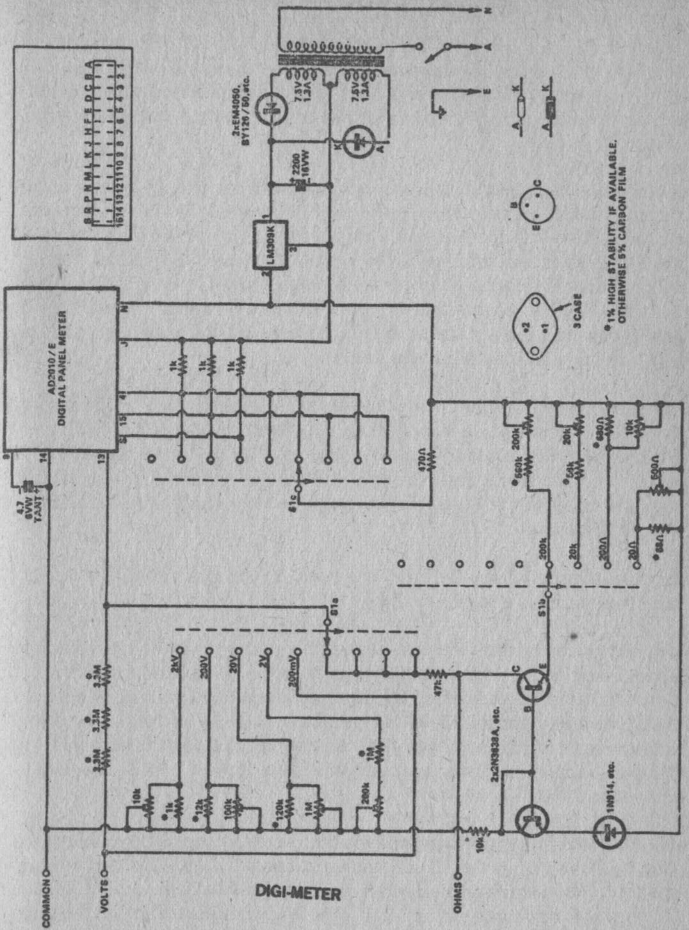
N
A
E
2×EM4050, BY126/50, etc.
7.5V 1.3A
7.5V 1.3A
2200 16VW
LM309K
A
K
A
K
B
E
C
3 CASE
AD2010/E DIGITAL PANEL METER
4.7 6VW TANT
470Ω
1k
1k
1k
200k
*560k
20k
*54k
*680Ω
10k
200Ω
*6.8Ω
200k
20k
20Ω
S1c
S1a
S1b
2kV
200V
20V
2V
200mV
3.3M
3.3M
3.3M
10k
1k
12k
100k
120k
1M
200k
1M
47k
2N3638A, etc.
C
B
E
10k
1N914, etc.
COMMON
VOLTS
OHMS
DIGI-METER
*1% HIGH STABILITY IF AVAILABLE. OTHERWISE 5% CARBON FILM

The circuit is simplicity itself, thanks to the Analog Devices digital panel meter. Only a handful of other components are required for the power supply and regulator, input attenuator and current source circuit.

Provision of the decimal point selection circuit proved to be a little tricky, because the AD2010/E is designed so that all decimal point LEDs are normally "on". For its normal application as a single-range panel meter, this is no problem as the unwanted points are simply blanked by shorting the appropriate pins to the negative rail. But in the Digi-Meter we had basically only one single-pole 9-position switch section available for decimal point selection – hardly the most appropriate for the job.

After a little head-scratching we came up with the system shown on the circuit, and it seems to work quite well. Essentially we have taken each of the AD2010/E decimal point pins permanently to the negative rail via 1k resistors, and then used the switch to re-apply 5V DC to the pin of the decimal point needed for each range, via a 470-ohm protective resistor. The 1k resistors appear to be quite low enough to reliably blank the undesired LEDs, while at the same time being large enough to allow any one to be unblanked when desired.

Our prototype Digi-Meter was built up in a nominal 7 x 5 x 4 inch vinyl-covered metal instrument case. This produces a neat and professional little unit, yet involves a minimum of work. We mounted the AD2010/E directly on the front panel (it clips into a plain rectangular hole 80 x 45 mm), along with the range switch, the input terminals and a small mains on-off switch.

Most of the rest of the wiring is mounted on the rear panel, which like the front panel is also removable.

The mains cord enters the case at the lower far end of the rear panel, its entry being made via the usual grometted hole. It is clamped in the approved manner with a P-clamp, and the active and neutral wires terminated in a section of "B-B" connector strip. The earth wire is soldered to a lug screwed firmly to the metal panel, so that even when the panel and transformer are removed from the rest of the case, both remain earthed.

The only component not actually mounted on either the front or rear panels of the case is the IC voltage regulator, which is mounted on a simple heatsink bracket bent up from a scrap of 16G aluminium sheet. The bracket measures 65 x 60 x 10mm, and is mounted in the bottom of the case by two of the screws used to attach the rubber feet.

If the usual four mounting holes for a TO-3 device are drilled in the heatsink, this will suit either the LM309K device or the TO-3 to measure a voltage in a low resistance circuit, (say less than about 10k), the effect will be very small, and can generally be neglected.

For measurements in higher resistance circuits, the only way of correcting for the error produced is to first make the measurement, then

simulate the resistance of the circuit by connecting a resistor of similar value across the Digi-Meter input terminals. The residual reading produced is then subtracted from the original reading, algebraically as before, to obtain the corrected reading.

When completed and power is to be applied for the first time, it would be a wise precaution to temporarily detach the 30-pin connector from the AD2010/E until you make sure that the power supply is correctly delivering the nominal 5V. This will prevent any possible damage should a wiring error have been made.

When you have satisifed yourself that the voltage socket between pins J and N on the socket is between the acceptable limits of 4.8V and 5.2V, turn off the power temporarily and refit the 30-pin connector to the AD2010/E. Reapplying the power should then result in normal operation, with digits visible on the readout display.

The only remaining operation is that of calibration. This is a little easier than with the more complex DVM, at least in one respect: the 200mV range is already calibrated. If the full accuracy of the AD2010/E is to be retained for the other ranges, however, it will be necessary to perform a full-scale calibration using a reference instrument and a stable source of adjustable DC.

Probably few readers will be in a position to do this, so that in many cases it will be necessary to do the next best thing: calibrate the 2V and 20V ranges using a source of reasonably known voltage, and then use a few suitable voltages present in an amplifier or other piece of equipment to bring the 200V and 2kV ranges into as close agreement as possible with the lower ranges.

The most convenient source of a reasonably known voltage for most readers will probably be a new mercury cell of the type used in cameras having inbuilt cadmium sulphide exposure meters.

When new, these cells have an open-circuit terminal voltage of 1.3566V, which drops to around 1.3524V after 12 months at an average temperature of about 21 degrees C. It is therefore reasonable to assume that virtually any such cell will have an unloaded voltage of very close to 1.35V, providing it is in reasonably good condition.

Even better, a cell bought as new from a large city store, where it is likely to be fairly fresh, could be reasonably assumed to have an unloaded terminal voltage of 1.355V. Such a cell could therefore be used to calibrate the 2V and 20V ranges to better than 1 per cent.

Calibration of the resistance ranges is best done using a precision decade resistance box, if access to such a device is possible. Some readers may be fortunate enough to gain access to such a decade box at their local technical college. For those unable to do this, probably the only practical alternative is to obtain a few representative close-tolerance resistors, and use these as references.

A final note: when you use the Digi-Meter, you may notice that there will be a very small but non-zero residual reading on the 2V range for open-circuit input. Similarly there will be a more significant residual reading on the 200mV range under the same conditions. Both are normal, and are due to the small bias current drawn by the input circuit of the AD2010/E.

On the 2V range the residual reading should be so small that for most measurements it may be ignored. However, for readings where the maximum possible accuracy is desired, it is only necessary to subtract the residual reading algebraically from the reading produced for the voltage being measured, to obtain the best estimate of the true voltage.

With the 200mV range the situation is not quite so simple, because the effect of the bias current will depend upon the resistance of the circuit in which the measurement is being made. If the Digi-Meter is being used to measure a voltage in a low resistance circuit, (say less than about 10k), the effect will be very small, and can generally be neglected.

For measurements in high resistance circuits, the only way of correcting for the error produced is to first make the measurement, then simulate the resistance of the circuit by connecting a resistor of similar value across the Digi-Meter input terminals. The residual reading produced is then subtracted from the original reading, algebraically as before, to obtain the corrected reading.

The Parts Needed

1 Case
1 Digital panel meter, Analog Devices type AD2010/E
1 Power transformer, 15V CT at 1.3A, low profile
1 Miniature rotary switch, 2-sections each 2-poles 9-positions.
1 Miniature toggle switch, SPST 240V rating
1 instrument knob, large
3 Screw terminals, 2 red and 1 black.

SEMICONDUCTORS

2 BY126/50, EM4050, RS276-1101/1105 or similar 50V silicon diodes
1 1N914 or RS276-1139 silicon diode
2 2N3638A, MPS3638, TIS50, 2N5447, SK3025, RS276-2021, or similar silicon PNP transistors
1 LM309K. uA7805 or similar 5V-1A regulator

RESISTORS

Half watt, 5pc: 470 ohms, 3 × 1k, 10k, 47k
High stability for preference: 68ohms, 680ohms, 1k, 12k, 56k, 120k, 560k, 1M, 3 × 3.3M
Miniature preset pots: 500-ohms, 2 × 10k, 20k, 100k, 2 × 200k, 1M (all linear).

CAPACITORS

1 4.7µF 6VW tantalum
1 2,200µF 16VW electrolytic.

MISCELLANEOUS

Mains cord and plug, handle and rubber feet for case, 30-pin connector for AD2010/E, 18-lug pairs section of miniature resistor strip, 3-bay section of "B-B" connector strip. P-clamp for mains cord, scrap of 16G aluminium for regulator heatsink, connecting wire, solder, etc.

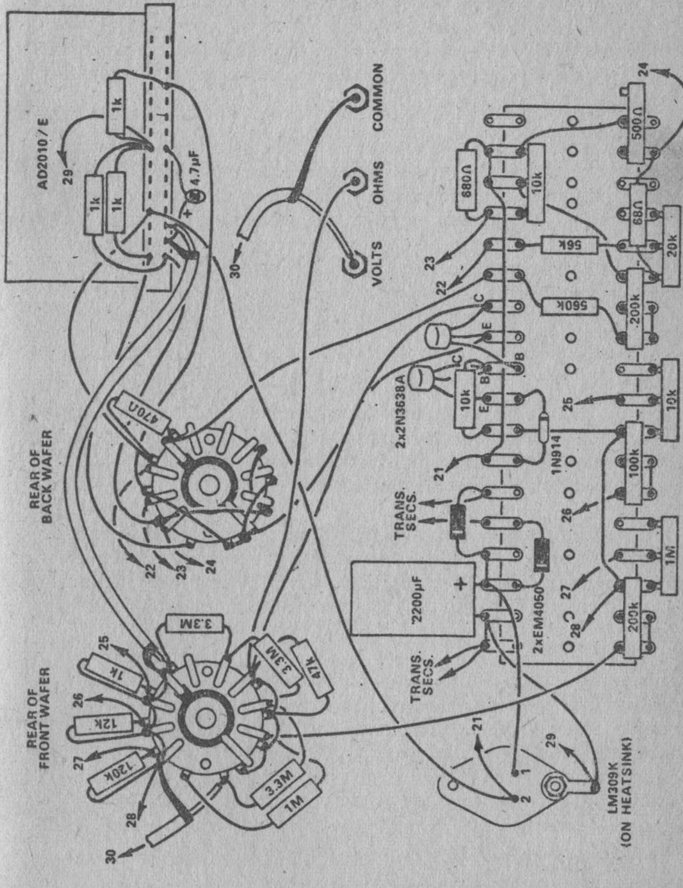

The wiring diagram of the Digi-Meter,

BREAKDOWN VOLTAGE TESTER FOR SEMICONDUCTORS

How often have you wanted to select a transistor with especially high breakdown voltage, for a special circuit? Or picked up a zener diode, only to find its identification has rubbed off? If the answer is often, you need this easily built little tester.

Here is a handy, low cost device which should be useful in the laboratory, to servicemen, amateurs and enthusiasts alike. It is designed to make breakdown voltage tests of a non-destructive nature, on bipolar transistors, FETs, diodes (including zener diodes), and other semiconductor devices.

With this tester, such measurements as BVces, BVcer and BVceo may be made directly. If the type of transistor is unknown, PNP or NPN, this may also be determined directly. Unknown diodes may also be checked as to whether they are silicon or germanium, by virtue of the forward voltage drop, together with the reverse breakdown voltage for each device. The zener voltage may be also readily determined for zener diodes. By a little exercise of the imagination, other tests may also be made on specific components. More will be said about testing procedure later on.

Basically, the tester consists of a high voltage supply which is directed across the item to be tested, in series with a limiting resistor. Facilities are included to introduce the item to be tested into the circuit, via suitable switching. Finally, a high resistance voltmeter is provided, which is connected across the device being tested.

Contrary to what was once thought, it is possible to break down the junctions of solid state devices without causing damage, providing the actual power dissipation within the device is kept to a low value. A typical example in practice is the now common zener diode voltage regulator.

In designing a tester of this sort, a number of arbitrary and other decisions have to be made in order to arrive at a practical unit. The extent or value of the high tension voltage to be used must be considered, in the light of the actual maximum voltage breakdown of the more common devices to be tested, against the economics of generating a useful voltage. In this unit, we settled for just a little over 400V. This is easy and economical to generate, and most devices which one could reasonably consider for testing would come within this figure.

While it is essential to have a voltmeter of very high resistance to measure the breakdown voltage of the device, the accuracy of the voltage reading is not all that important, within reasonable limits. This means that the design of the voltmeter part of the circuit may contain some short cuts in order to keep costs down. More will be said about this and similar items as we go along. Meanwhule, let us have a look at the circuit of the complete instrument.

The power supply, including the 400V high tension and the +12V and −12V supplies for the IC are all obtained from one source. A small transformer with a secondary AC voltage of 150, rated at 30mA is fed into a voltage doubler, giving a total output of about 420V DC. A little over 400V of this is applied via 440k to the device to be tested. Under pre-test conditions, this resistor is shorted to ground, thus dissipating within itself just a little under 400mW.

The maximum power which can be dissipated within a device being tested, is when its breakdown voltage is half the supply voltage, and amounts to a bit less than 100mW.

In order to protect the operator against an electric shock, the 440k resistor is normally shorted to ground via a push-button switch. A socket is provided for small transistors and the three connections are duplicated on three banana sockets for large transistors and other devices. A toggle switch reverses the supply voltage to accommodate NPN and PNP devices, while a second three-position toggle switch permits switching for three different test conditions. These are – collector to emitter with base shorted, with 10k in the base and with the base open circuit.

The voltmeter is designed around the ubiquitous 741 IC. This device is itself quite cheap, and it also allows the use of a relatively cheap 0–1mA meter movement.

In order to keep the input resistance to the device high and to restrict the voltage applied to the non-inverting input, a voltage divider consisting of 22M and 22k is used. By varying the negative feedback from the output to the inverting input, the gain, and so the effective voltage read by the meter, is controlled. The meter has 900 ohms in series with its internal 100 ohms, resulting in a 0–1 volt movement. Combined with the negative feedback just mentioned, this gives the requisite voltage ranges of 10V, 100V, and 400V.

Although the amount of current delivered by the 741 IC is restricted, we considered it desirable to add some extra protection to the meter. An ordinary silicon diode, such as an EM401 or BY127 or RS276-1139 has been connected across the movement and this should prevent any overload from exceeding about six times, giving adequate protection. A 10k trimpot is connected between pins 1 and 5, to adjust the zero meter reading.

To provide the positive and negative supply voltages for the 741 IC, we have connected two 12V zener diodes in series. The junction of the diodes becomes the reference point for all measurements involving the IC. The two zener diodes are fed from the 400V supply via 60k.

So much for the basic circuit details. A number of components need clarification and so some comments may be helpful. The power transformer which we used is a standard stock item. This transformer has a multitapped secondary but we have ignored the taps and used the full available 150V AC. There is also a 6.3V secondary winding, which we have not used. More will be said about these points under constructional details. If you wish, and another suitable type of transformer is available to you, there is no reason why it should not be used.

The 440k series resistor is made up of two 220k 1W resistors. Although one 470k 1W resistor may do the job satisfactorily, it is not desirable to put too high a voltage across any resistor, notwithstanding its power dissipation capabilities. A somewhat similar situation exists with the 60k resistance in series with the two zener diodes. Not only do we have a high voltage to cope with but the power dissipation of nearly 3 watts must also be considered. Although it is all right to apply high voltages across wire wound resistors, providing the power dissipation is adequate, such resistors are expensive and we settled for five 12k 1W units as an economy measure.

The 22M resistor at the input to the IC should be of low tolerance and perhaps the best way to achieve this end is to make up the value with two 8.2M units and a 5.6M unit, all of low tolerance, in series. Both 22k resistors should also be of low tolerance.

The 900 ohm resistance in series with the meter is made up with a 680 ohm and a 220 ohm resistor in series. These should also be low tolerance types. The 56k resistor should, according to calculations, be 55k. The three resistors involved in the voltage range selection should ideally be tailored to size during calibration of the system against a meter of known accuracy. In point of fact, we fitted a 2.7M resistor in place of the 2.2M and shunted it down until we got the required accuracy. Similarly, we used a 270k resistor, shunted down, instead of the 220k. Also, the 56k must be shunted to give a full scale reading of 400V.

Before leaving the subject of resistors, we have used the term "low tolerance". In our unit, we used 5 percent tolerance resistors where this was required and then did the final calibration with the three resistors in the feedback network, as described in the last paragraph. If you have no means of calibration, then it would be wise to use 2, or even 1 percent tolerance resistors where required and hope that this will lead to a satisfactory level of voltmeter accuracy. Although accuracy is not vital, as suggested earlier, we feel that it should be around the 5 percent mark.

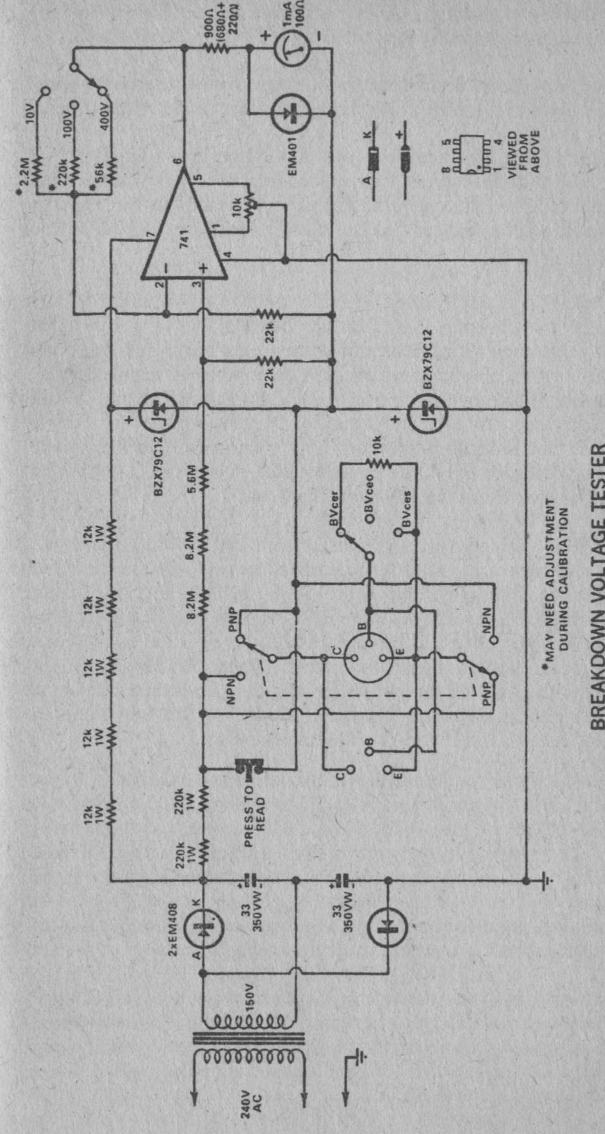

The complete circuit of the breakdown voltage tester. The required range is selected by varying the feedback network.

Although the push-button switch only calls for a single pole, we used a double pole unit as this was available at the time. Also, the single-pole three-position rotary switch, was not available at the time we needed it and we substituted a double pole unit.

The 741 IC is available in a variety of packages but we chose the 8-pin DIL version as this is compact and is easy to fit into the wiring board.

We built the prototype Breakdown Voltage Tester into a plastic moulded case with aluminium front panel, and the result is a neat and compact unit. The actual physical arrangement of the tester is not critical and it may be built into some other type of box if you choose to do so.

Construction is a relatively simple matter but like any other project, it is wise to tackle it in some logical order. The majority of the components are mounted on a length of miniature wiring board and this could be done first. One operation which calls for more than usual care, is the mounting of the socket for the 741 IC. Eight small clearance holes must be accurately drilled in the board so that the socket fits neatly into these holes. The holes should be just large enough to take the pins of the socket. With this done, the pins should then be carefully wired to their terminations as indicated on the diagram.

The rest of the wiring is quite straightforward and should present no difficulties. However, as mentioned earlier, the resistors associated with the calibration of the meter should be given consideration at this stage, according to your approach to this part of the circuit. If you intend to follow our method, then a 2.7M resistors should be fitted instead of the 2.2M and a 270k resistor should replace the 220k. A 56k resistor will be fitted anyway, and all three resistors will then be shunted down after the unit is complete and ready for calibration. For the time being, omit the lead from the 220k resistor to the 8.2M resistor.

Before leaving the wiring board assembly, perhaps we should point out that the two resistors connected in series to make up 900 ohms, share one pair of tags at one end of the board. This means that the resistors are stood off the board at one end, to form a triangle, with a soldered joint at the top. Also, the centre hole in the board at this point is used for screwing the board to the case later on. The equivalent hole at the other end is also used for mounting, which means that the 33μF electrolytic must be moved to one side while mounting is done.

The transformer is then mounted on the back of the box and at an angle. The angle mounting was necessary to fit the transformer into the available space. Before fixing the board which has already been completed, leads for external terminations should be provided. These involve switching, meter leads, etc.

Mounting of the components on the metal panel should present no problems. However, if you have a transistor socket the same as ours,

the question arises how best to fix it. We drilled a hole so that the socket was just a neat fit, set it so that about half of it protruded through the panel and then put a couple of blobs of Araldite between the socket and the back of the panel.

When wiring the components on the panel, some care should be taken with the two toggle switches as these can be a little tricky. The three banana sockets are connected in parallel with the transistor socket. The protective diode is wired directly across the meter terminals. We also ran a lead from the mains earth screw on the terminal strip, to the earth point on the wiring board and continuing onto a solder lug under one of the meter fixing screws.

Having completed your Breakdown Voltage Tester, before proceeding to adjust and calibrate the unit, it is always wise to make a thorough check of the wiring, polarity of components, etc. Satisfied that all is well, set the rotor of the 10k trimpot to its mid-position. Switch on and check for obvious signs of distress.

Measure the voltage from the voltage doubler output, between it and earth. This should be about 420V. Check the voltage between earth and the other end of the two 220k resistors. With the push-button switch operating correctly, this voltage should be about 12V, which is the voltage across one of the zener diodes. A check across both zener diodes in series should give 24V.

At this point, there may be some error in the meter zero reading. This error will be a maximum with the voltage selector set to "10V". Adjust the 10k trimpot for zero reading on the meter, with the selector set to the 10V range.

To calibrate the meter, it will be necessary to have a source of variable voltages around 10V, 100V and 400V. A good multimeter is also needed to calibrate against. It does not matter in which order the voltage ranges are calibrated. Connect the negative lead of the multimeter and the negative lead of the voltage source to the junction of the two zener diodes. With the 8.2M voltmeter input resistor still isolated from the rest of the circuit, this is connected to the positive pole of the voltage source and the positive lead of the multimeter.

Set the appropriate range on the voltage selector switch, adjust the input voltage from the source, for the correct reading on the multimeter. Now proceed to shunt the appropriate resistor in the feedback network until the meter on the tester reads correctly. Solder this new resistor permanently in place. Repeat this procedure for the other two ranges and calibration is complete. Having removed the multimeter and voltage source leads from the unit, connect the 8.2M resistor to the push-button switch side of the 220k resistor. This done and with the voltage switch set to 400V, pressing the push-button should give a meter reading just exceeding full scale.

Although the use of the Breakdown Voltage Tester is more or less self-explanatory, some comments and suggestions may be helpful. The simplest devices to test are diodes in their various forms. Bipolar transistors, whether small or large, NPN or PNP, silicon or germanium, can also be tested simply and easily. Junction FETs may also be tested for BVgso.

Before making a test, always start with the voltmeter set to 400V and then if the subsequent reading is low, the range may be changed to 100V or 10V as required. After taking a reading, always reset the voltmeter to 400V.

To test any diode for its reverse breakdown voltage, or to find the operating voltage of a zener diode, connect the cathode of the diode to the collector point on the transistor socket, or the equivalent banana socket point via an alligator clip and banana plug. The anode of the diode is connected to the emitter point. Set the switch to "NPN" and press the button for the reverse breakdown voltage reading. Setting the switch to "PNP" will give the forward voltage drop, which will be about 0.6V for a silicon device and about 0.2V for a germanium device.

To test small bipolar transistors in TO18, TO5 packages, etc., they are plugged directly into the small socket provided. However, in some cases, the connections are such that they may not be plugged in directly, without bending the leads so that the right connections are made between the socket and the device. The popular BF115 is a case in point. The shield lead will be ignored, but the base and emitter leads must be reversed. Whenever there is a doubt, the appropriate manufacturers reference should be consulted.

With the device connected, the polarity switch should be set to NPN or PNP as required. Pressing the button will give the breakdown voltage readings for BVces, BVceo and BVcer, according to the position selected on the switch. In most cases, there will be little difference in readings between BVcer and BVceo. BVces will generally be somewhat higher.

If the bipolar transistor is of the larger variety and cannot be plugged into the small socket, then it may be literally hung on the three alligator clips, clipped to the appropriate connections on the transistor. Testing is then carried out as previously described.

If you wish to test any bipolar transistor for BVcbo, then this may be done by using the alligator clips, or by plugging the appropriate leads into the socket. In this case, the collector will be connected as before but the base will be connected to the emitter point and the emitter will be left open. The reading will then be taken by pressing the button and with the switch set for polarity. The three-position switch has no effect on this reading. It may be of interest to mention that this test gives virtually the same reading as BVces.

There are many occasions when the type and particularly the polarity of a transistor are unknown. If the transistor is a good one, with the polarity switch the wrong way only a very low reading will be observed. With the polarity switch correctly set, the normal voltage reading will be obtained. If the transistor is short circuited, no reading will be obtained in either polarity. If the transistor is open circuited, the full voltage reading will be obtained in both directions. By thoughtful use of the polarity and the three-position condition switches, the state of a transistor can be determined with reasonable certainty.

To test a junction FET for BVgso, connect the gate to the collector point and the source to the emitter point, with the source left open. Set the polarity switch to PNP for N-channel devices and to NPN for P-channel devices. Pressing the button gives the breakdown voltage reading.

The foregoing is just a brief run down on testing solid state devices for breakdown voltage effects. For a more in-depth treatment, readers may wish to refer to Transistor and Diode Testing in our Bernard's Book No.206 "Practical Transistor Circuits for Modern Test Equipment" on pages 80 to 96.

Parts List

1	*Case 184 × 115 × 118mm, with aluminium front panel.*
1	*Power transformer, 240V primary, 150V, 30mA secondary.*
1	*Meter, 1mA FSD, 100 ohms, 75 × 65mm*
1	*IC, 741 8-pin DIL. or RS276-007/010/1741*
1	*IC socket, 8-pin DIL.*
1	*Miniature tag board, 25prs tags*
1	*3-way terminal strip*
2	*33μF 350VW electrolytics*
2	*Diodes, EM408 or similar or BY127-1N4006*
1	*Diode, EM401 or similar or BY127-1N4002, RS276-1129*
2	*Zener diodes, BZX79C12, RS276-563/623*
1	*Miniature toggle swithh, double-pole, 3-position*
1	*Miniature toggle switch, DPDT*
1	*Miniature push-button switch, single-pole, normally closed*
1	*Rotary switch, single-pole, 3-position*
1	*Transistor socket*
3	*Banana sockets*
1	*Knob*

RESISTORS (½ watt unless stated otherwise)

1	*220 ohms (low tolerance)*
1	*680 ohms (low tolerance)*
1	*10k*

1 10k linear trimpot
5 12k 1W
2 22k (low tolerance)
1 56k (see text)
2 220k 1W
1 220k (low tolerance)
1 2.2M (low tolerance)
1 5.6M (low tolerance)
2 8.2M (low tolerance)

SUNDRIES

3-core flex and plug, flex clamp, screws, nuts, solder lugs, solder, hook-up wire.

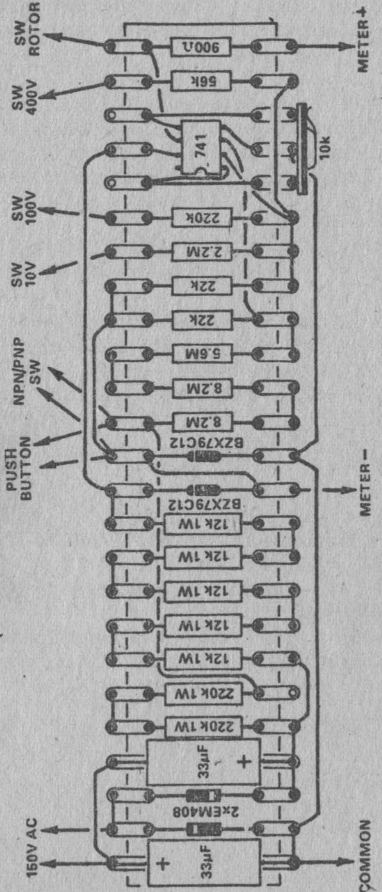
SW ROTOR
SW 400V
SW 100V
SW 10V
NPN/PNP SW
PUSH BUTTON
150V AC
METER+
METER–
COMMON
900Ω
56k
741
10k
220k
2.2M
22k
22k
5.6M
8.2M
8.2M
BZX79C12
BZX79C12
12k 1W
12k 1W
12k 1W
12k 1W
12k 1W
220k 1W
220k 1W
33µF
2xEM408
33µF

This wiring diagram of the miniature tagstrip should make construction a straightforward process. Follow the diagram in conjunction with the circuit

LOW RANGE OHMMETER

This ohmmeter fulfils a need for measurement of resistors from 0.01 ohm to over 100 ohms by comparing a resistor of unknown value with the selected 1% tolerance reference resistor. In operation the meter is zeroed by means of RV1 with the range switch set at position 2. In positions 3, 4 and 5 a voltage of almost 15mV across the reference resistor, coupled with an IC gain of 101 produces an output of about 1.5V and by adjusting RV2. the meter is set to FSD.

When connected, an unknown resistor forms a parallel circuit across the reference resistor lowering the voltage across it, thereby reducing the meter reading. Accuracy is better than 1% + or – the reference resistor tolerance. A meter having a mid-scale reading of 10 is necessary. Any VOM having this type of ohms scale could be used as the read-out by connecting the lower end of the 680 ohm resistor to 0V and connecting the VOM set at the 1V range between points X and Y.

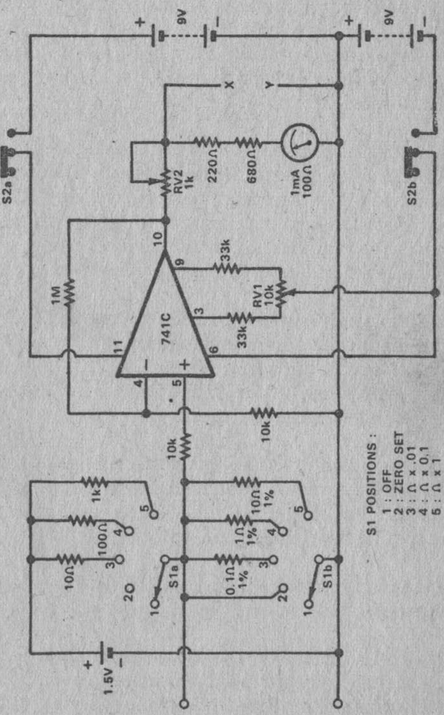
9V
9V
S2a
S2b
X
Y
RV2 1k
220Ω
680Ω
1mA 100Ω
10
9
33k
RV1 10k
1M
741C
3
33k
11
6
4
5
10k
10k
1k
10Ω
100Ω
4
3
5
1O
2O
S1a
10Ω 1%
1Ω 1%
0.1Ω 1%
S1b
1.5V
S1 POSITIONS :
1 : OFF
2 : ZERO SET
3 : Ω x .01
4 : Ω x 0.1
5 : Ω x 1

A low range ohmmeter

EXPERIMENTAL BRIDGE TYPE OHMMETER

Here is an interesting experimental project which should appeal to beginners and advanced workers alike; a simple bridge type ohmmeter using a novel form of null indicator. The circuit would also lend itself to a wide range of other applications.

Following the description of our most recent RC Bridge there was some discussion in our laboratory concerning bridge circuits in general and the manner in which they can be modified to perform specific functions. The discussion considered in particular the various forms of null indicator available and their suitability for unskilled operators who, ideally, need an unambiguous "high" or "low" indication or, in some cases a simple "go" or "no go" indication.

Arising out of this was the circuit we are about to describe; a null indicator based on an operational amplifier IC (uA741) and a pair of light emitting diodes. Not only is it a novel circuit but, as far as we know, also an original one. More importantly, it would seem to have considerable potential in many applications, particularly those involving unskilled operators, as already mentioned.

Fig. 1 is a simplified explanatory circuit, in which the operational amplifier is being used as a comparator. If the voltage at the non-inverting input is higher than that at the inverting input, the op amp will saturate, and the output voltage will rise to a value close to that of the positive supply rail, Vcc. This will forward bias one of the light emitting diodes, which will give a visible indication, and reverse bias the other diode, which will not emit.

When this situation is reversed, and the voltage at the inverting input is higher than that at the non-inverting input, the op amp will saturate in the reverse direction, and the output voltage will fall to a value close to that of the negative supply rail, – Vcc. This will forward bias the other light emitting diode, and reverse bias the first one.

Since the op amp has a very high gain, the change-over between the two conditions will occur very rapidly and over a very small range of differential input voltages. Thus the null of the bridge will be indicated by the point at which the two LED's change from off to on and vice-versa.

To prove the circuit, we made up a simple bridge network similar to that shown on the left hand side of Fig. 1.

It consists of a 1k linear pot, a set of three standard (close tolerance) resistors, a battery, and provision to connect the unknown resistor (Rx). The 1k pot forms a pair of resistors whose ratio is variable over a wide range although, in practice, we limit this range so as to provide a practical scale. From the mid position of the moving arm, which represents a 1 to 1 ratio, we provide for an extremity of 100 to 1 in either direction.

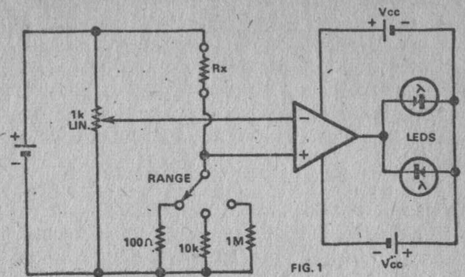

FIG. 1

Fig. 1. Simplified diagram showing the basic bridge configuration. Note that the 1k linear pot provides two of the four resistors making up the bridge. The other two are the unknown resistor and one of the standard resistors selected by the range switch.

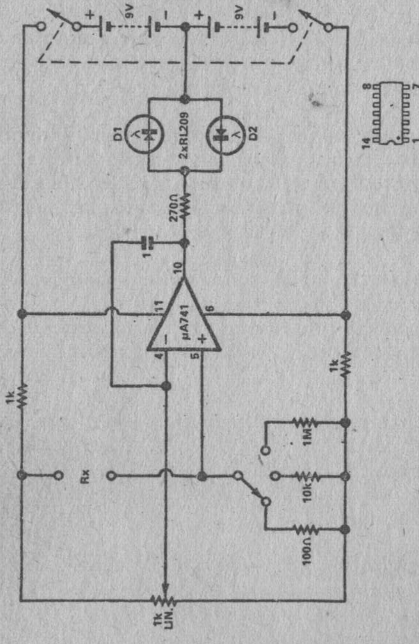

EXPERIMENTAL OP. AMP. BRIDGE

FIG. 2

Fig. 2. The circuit as finally developed. The bridge circuit remains essentially the same, except for the addition of two 1k current limiting resistors. The one battery pack, supplying 9V each side of a centre tap, serves both the bridge and the IC.

The three standard resistors are 100 ohms, 10k and 1M. In conjunction with the pot movement just discussed, this gives a range of measurement from at least 10 ohms to 10M, beyond this with some limitations, and with useful overlap between ranges.

After ironing out a few minor problems we came up with a circuit which we feel is worth passing on to readers. As presented it is a simple battery operated bridge, using relatively few low cost components, but capable of quite useful accuracy. It would seem to be an ideal project for YRCS groups or similar organisations.

Alternatively, by using closer tolerance components, it can be upgraded to provide a much higher order of accuracy, but at some increase in cost. And for the really keen experimental types there is an almost unlimited scope for its exploitation in automatic control circuits.

The practical circuit is shown in Fig. 2, and is worth discussing in some detail. The 741 op amp requires nominal voltages of plus and minus 15, but it will work quite satisfactorily for our purpose from voltages as low as plus and minus six. Since 9V batteries are readily available in convenient sizes, we elected to use two of these.

This gives us a total of 18V, which also becomes the power supply for the bridge. The latter should be as high as possible when high value resistors are being measured, otherwise current due to thermal noise in the resistors may be comparable with the current due to the supply.

On the other hand, a high supply voltage causes problems when low values of resistance are being measured, as the battery drain increases enormously. To avoid this, limiting resistors are connected in series with the bridge. To keep the input voltages to the op amp centered about zero, equal value resistors are fitted at the top and bottom of the bridge.

These resistors also limit the current to the op amp inputs, so that it is not necessary to provide separate input resistors. A protective resistor is required for the LEDs, to keep the forward current within the maximum allowable. Because the two diodes are back to back it is not possible to exceed their reverse voltage rating.

When the prototype was constructed, some trouble was experienced with oscillation at the null point. This took the form of cycling between the LEDs, which made both LEDs appear to be on at the same time. This was prevented by using a 1uF capacitor to provide negative feedback from the output to the inverting input.

In order to prevent oscillation if the wiper arm of the pot should become dirty and go open circuit, it is necessary that the wiper arm be connected to the inverting input, and not the non-inverting input.

The circuit does have one minor disadvantage. Because the null indicator operates from the same supply as the bridge, and because there will be some leakage between the op amp input terminals and the supply rails, there will be some error introduced into the null indication at higher resistance values. In practice this is insignificant up to about 10M, which should be regarded as the upper limit for accurate measurement.

The only close tolerance components are the three standard resistors. We used 1% types but, if these are not available, it may be possible to trim 5 or even 10% types to the correct values. To ensure stability with respect to temperature, only good quality components should be used.

We constructed our unit on a panel measuring 111mm x 161mm, which happened to suit a box we had on hand, Since the size of the panel and the type of box is not critical, we suggest that the reader make his own choice in this regard, possibly using something which is on hand.

A good idea of the panel layout and general assembly can be obtained from the photographs. The pot is in the top centre of the panel and the two switches and the unknown resistor terminals at the bottom. The LEDs we used fit neatly into a 1/8in hole and we mounted them at the top of the panel, about 20mm apart, where they are easily seen.

The only other piece of metal work is a simple battery clamp made from a scrap of aluminium. It is roughly "U" shape, but with wings extending outwards from the uprights of the "U". One 9V cell is held under each of these wings, but they also support a piece of veroboard on which is mounted the IC and other minor components. The bracket and the veroboard are clearly visible in the photograph. The bracket is fastened to the front panel by the pot, avoiding the need for unsightly screws.

We mounted the three standard resistors on the back of the single pole three position range switch. It was necessary to add a dummy terminal to the switch, to provide the common point for the resistors, by attaching a solder lug to a spare hole fortunately provided. Alternatively, there is ample room to mount the resistors on the veroboard with the remainder of the components, although this will entail running more wires to the switch.

The 741IC we used was in a 14 pin dual-in-line package. These devices are also available in an 8 pin dual-in-line package and in an 8 lead metal can. All are suitable for use in the bridge. We used veroboard with the correct hole spacing to suit the dual-in-line package.

In order to reduce the height needed for the op amp, we mounted it directly on the veroboard. As only pins 4, 5, 6, 10 and 11 are used, we did not solder the remaining pins to the veroboard. This makes it easier to remove the op amp if necessary. Take care not to damage the IC while it is being soldered, due to excess heat. It is a good idea to leave it to the last to solder in.

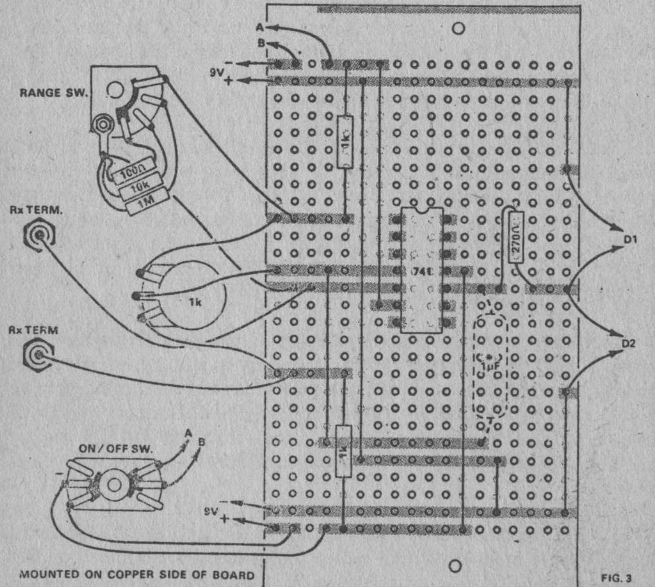

FIG. 3

Fig. 3. Wiring diagram showing the general layout of components and the pattern of the Veroboard. Layout is not critical and almost any other physical arrangement would be acceptable. All components mount on the panel, simplifying the choice of a box.

As shown in the photographs, the IC and the remaining three resistors are mounted on top of the veroboard. The LED's and the 1uF capacitor are mounted underneath. This saves space, in the case of the capacitor, and enables the LED's to fit into their mounting holes in the front panel. The veroboard is attached to the battery mounting clamp using two self-tapping screws and some washers for spacers. Leave sufficient clearance under the board to avoid shorts to the battery case.

We chose to polarise the unknown resistor terminals by selecting one red and one black. While resistors are not normally polarised, the bridge may be used in other applications, such as checking the forward and reverse leakage of a diode. The red terminal should indicate positive.

We used pointer knobs for the two rotary switches and a larger round knob for the potentiometer. To this we added a clear plastic pointer, fastened to the bottom of the knob with a strong epoxy glue. Make sure that the line on the pointer is truly radial and in line with any lines or arrows already on the knob.

Finally, there is the matter of a scale. In the case of previous bridge projects we have been able to publish a calibrated scale and supply details to any firms wishing to supply etched labels. We are deliberately avoiding this approach on this occasion for a number of reasons.

Such a scale is valid only if the pot used by the builder is identical with that used in the prototype. This means, in practice, nothing less than a wire wound type of the same make and model. Unfortunately, wire wound pots are quite expensive and, to keep the cost down in what is, after all, mainly an experimental project, we used a linear carbon pot, which is a good deal cheaper.

Unfortunately, carbon pots, as a class, introduce two unwanted variables. They are not necessarily as linear as the wire wound type, even from unit to unit of the same model, and the angle of rotation varies from make to make. Thus, a scale prepared for brand "A" will not necessarily suit brand "B". But the linearity is another matter.

Because of these variables, the best approach is to make one's own scale; an interesting exercise in its own right. There are two approaches. One is direct calibration against a large number of close tolerance resistors, ideally one for each calibration mark.

The other is direct calibration of only two points, one at each extremity of the scale, and all other points by calculation on the assumption that the pot is linear.

Obviously the first would be the more accurate, but is seldom practical. The second is more practical, but leaves the accuracy at the mercy of the pot linearity. We used the second method.

For this method we require three close tolerance resistors, 1k, 10k and 100k, a protractor, and ordinary drawing instruments. The first job is to determine the angle between two selected values near the extremities of the pot rotation. Place a scrap of paper on the panel where the scale will finally fit and secure it with sticky tape. Fit the pointer so that it swings approximately evenly each side of centre. This setting is not critical at this stage.

Set the range switch to 10k and note the positions at which the 1k and 100k resistors are nulled. The 1k should null near the furthest anti-clockwise position, and the 100k near the furthest clockwise position. If these positions are reversed, reverse the outside connections to the pot.

Remove the pointer and the piece of paper and, with the protractor, measure the angle between the two marked points. This will be easier if the centre of the pot can be marked, using lines at right angles, before the paper is removed from under the pointer. The angle should be about 200 degrees.

Let us call this angle W. Then, if s is the fraction or multiple of 1 by which the standard has to be multiplied, and p is the angle in degrees that actually corresponds to this value of s, we can say $p = 0.611 \times W \times (s-1) / (s+1)$. This formula enables the values of p to be calculated from the values of s. The angle p is measured from the vertical, which corresponds to the position $s = 1$, $p = 0$. The angles are measured positive clockwise and negative anti-clockwise.

As an example, we have calculated the angle at which the scale marking "0.5" would be placed. We have taken the measured value W as 200.

$p = 0.611 \times 200 \times (0.5-1) / (0.5 + 1)$

$p = 122.2 \times (0.5-1) / (0.5 + 1)$

$p = 122.2 \times (-0.5) / (1.5)$

$p = 122.2 \times (-0.333)$

$p = -40.74$ degrees

This means that the scale point marked "0.5" should be positioned 40.74 degrees anti-clockwise from the centre of the scale.

These calculations will be much easier if a slide-rule or a calculator can be used, as well as being much less prone to error.

Once the values for the scale have been obtained, the scale can be drawn on a suitable piece of white card. We made the scale with a diameter of 80mm, the major graduations 6mm long, and the minor graduations 3mm long, using India ink. Black ball point pen would provide an acceptable substitute for ink. We used rub-on lettering to mark the major points on the scale and also to label the functions of the rotary switches.

When the scale is finished, suitable cutouts must be made for the pot and switch shafts, the LED's and the terminals. With the scale fitted, the pointer can be attached to the pot shaft, and correctly aligned with the scale. This is done using the 10k close tolerance resistor, with the range switch on the 10k range. The pointer should be attached to the shaft so that when the bridge is balanced, the pointer reads 1.0. Provided there are no serious linearity errors in the pot, the scale should then be accurately calibrated for all ranges.

Care should be exercised in regard to battery life. The current consumption is highest on the lowest resistance range, and unlike an ohm-meter, current is consumed even when the unknown terminals are open circuit. Fortunately, the LED's give a visible indication that the power has been left on. Otherwise, due to the intermittent nature of the current drain, battery life should be quite long.

PARTS LIST

1 uA741 op amp, RS276-007/010/1741 or equivalent
1 red LED's, CEMA RL209 or similar
2 1k ¼ watt resistors
1 100 ohm 1% ¼ watt resistor
1 270 ohm, ¼ watt resistor
1 10k 1% ¼ watt resistor
1 1M 1% ¼ watt resistor
1 1k linear pot
1 1uF capacitor
2 9V batteries with clips
2 terminals, 1 red, 1 black
1 2 pole 2 pos rotary switch
1 1 pole 3 pos rotary switch
3 knobs to suit (see text)
Vero board, scrap aluminium, hookup wire, solder, self-tapping screws, white card, 5mm spacers, solder lugs, screws and nuts, small piece of clear plastic.

Note: resistor wattage ratings and capacitor voltage ratings are those used in our prototype. Components with higher ratings may generally be used provided they are physically compatible. Components with lower ratings may also be used in some cases, providing ratings are not exceeded.

A SYNCHRONOUS GATING UNIT FOR AF TONE-BURST TESTING

A fully solid-state instrument which produces synchronously grated AF tone bursts suitable for a wide variety of audio measurements. The use of integrated microcircuits gives the instrument a standard of performance comparable with quite costly commercial units, despite the fact that it is easily constructed at low cost.

Although at first sight the instrument described in this article might appear to be suitable only for a few specialised and somewhat esoteric applications, this is not the case. In fact it has many uses, of which probably the most practical and straightforward is the measurement of short-term or "dynamic" power output capability of audio power amplifiers.

Other direct uses include testing of amplifier stability, measurement of loudspeaker system response and damping, measurement of room and auditorium acoustics, measurement of the performance of filters and automatic gain and level control circuits, sonar equipment testing, and psychometric research and testing. The instrument may also be used to produce sound effects for live theatre and motion pictures, and in the production of electronic music.

In addition it may be used directly or indirectly as a signal shaper-squarer, a wide-range binary scaler-frequency divider, or a synchronous digital gate controller suitable for counting, frequency and time measurement, or clock word generation.

In view of the likely interest in the use of tone-bursts to measure the short-term power output capability of audio amplifiers, a brief discussion of this application will now be presented by way of a preamble to the description of the instrument itself. It is hoped that the discussion will be found of value not only to those who may contemplate construction of the instrument but also to those who wish solely to gain a theoretical knowledge of this relatively recently evolved technique.

It is commonly found that the maximum power output which may be obtained from an audio power amplifier, before discernible waveform clipping or other distortion is incurred, depends significantly upon the period of time over which the power is to be delivered. Generally, the shorter the period involved the higher the power which may be delivered —althouth this simple relationship will naturally be limited by transformer losses and other effects influencing the power frequency response.

Thus the maximum power output delivered by a power amplifier when tested under continuous sinewave conditions may often be significantly less than that which it is capable of delivering during the short and relatively infrequent amplitude peaks present in typical musical programme material. Hence for the many audio amplifiers

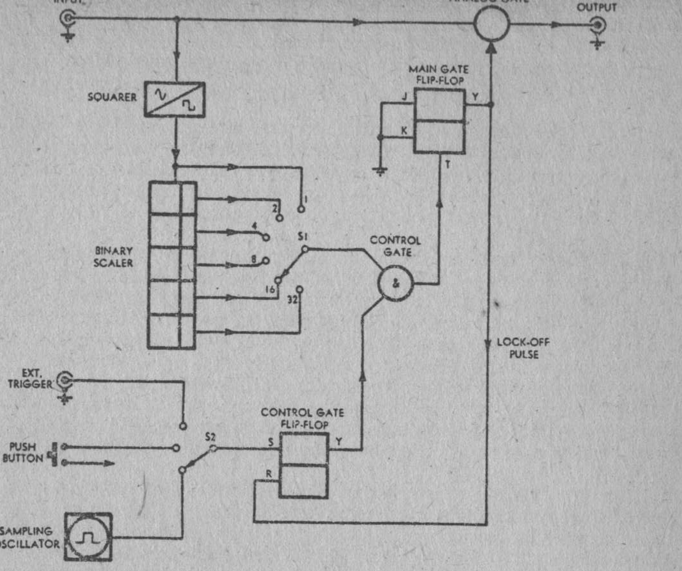

Fig. 1 BASIC OPERATION OF THE TONE-BURST GATE

Specification

A synchronous analog gating unit which may be used to derive pedestal-free bursts of input signals between 10Hz and 60KHz. The integral number of cycles in each burst and the burst repetition rate may be adjusted independently, or locked if desired. The instrument is fully solid-state and employs low-cost RTL digital microcircuits. It may also be used as a signal shaper-squarer, a scaler frequency-divider or a synchronous digital gate controller.

ANALOG INPUT:
Frequency Response: 10Hz—60KHz.
Minimum Input: Less than 150mV RMS.
Maximum Input: Greater than 600mV RMS.
Switching Ratio: Better than 40dB.
Insertion Loss (gate open): Less than 6dB minimum.
Switching Transients: Less than 30mV P-P.
Input Impedance: Approximately 33K.
Output Impedance: Maximum 6K, minimum 500 ohms.
T.H. Distortion (gate open): Less than 0.2% at 250mV RMS input.
Gate offset or pedestal: Negligible.

BURST LENGTH: Adjustable to 1, 2, 4, 8, 16 or 32 cycles; also continuous.

BURST REPETITION RATE: Internal trigger generator provides a range of approx. 3Hz—500Hz. Single bursts triggered manually by panel push-button. External triggering by +3V pulse, duration less than 20uS. Inter-burst interval may also be locked to equal the burst length.

SYNC OUTPUT: Rectangular waveform with level 0V when gate is closed and +3V when gate is open.

intended to handle musical signals rather than sinewaves, a conventional power output rating based upon continuous sinewave measurement may be both unduly conservative and of only incidental relevance, despite the relative ease and high order of objectivity with which the measurement may be made.

Aware of the limitations of a continuous sinewave power rating when used alone as a performance criterion applied to their products, manufacturers of high fidelity amplifiers some time ago adopted other rating methods which have been employed either as supplements to the basic continuous sinewave rating, or as substitutes for it.

It is undoubtedly true that academic purism was by no means the sole motivation behind the adoption of these other–usually more flattering–rating methods. It is also true that some of the rating methods adopted were and still are, of dubious technical value. However, it is neither appropriate not possible in the present article either to discuss in detail the various rating systems or to examine the motivation behind their introduction. Sufficient to say that one of these ratings, the so-called "music power" rating, was introduced in an effort to provide a measure of the actual power output capability of an amplifier when reproducing a musical programme.

Unfortunately while such a rating is easily proposed and validly justified, the measurement ideally required is not at all simple or straightforward in execution. The basically statistical nature of musical signals itself produces both theoretical and practical problems when one attempts to measure the power level involved, quite apart from the perhaps more obvious difficulty in selecting the "typical" musical programme to be used as a basis for inter-amplifier comparison.

The method originally adopted for "music power" measurements bypassed these problems altogether, by making a convenient and usually fairly valid assumption regarding the cause of the diminished power output capability of an amplifier when tested under sinewave conditions. The assumption was that the diminished power output was due solely to imperfect power supply voltage regulation under the steady drain of continuous operation at maximum power output.

The converse corollary of the assumption was that under the transient conditions of a musical programme the power supply was sufficient to cope with the relatively infrequent demands for full power, and was therefore able to maintain supply voltages at their full amplitude. Hence in order to measure the "music power" output capability of an amplifier, it was reasoned, it was not necessary to attempt to take measurements under actual musical conditions but simply necessary to take a "normal" continuous sinewave measurement with the amplifier supply voltages maintained at their full amplitude to duplicate the situation prevailing during brief musical peaks.

Consequently an established technique for the measurement of amplifier "music power" is to substitute for the internal amplifier power supply an external heavy-duty regulated supply–delivering the same voltages–and to measure the maximum continuous sinewave output under these artificial conditions.

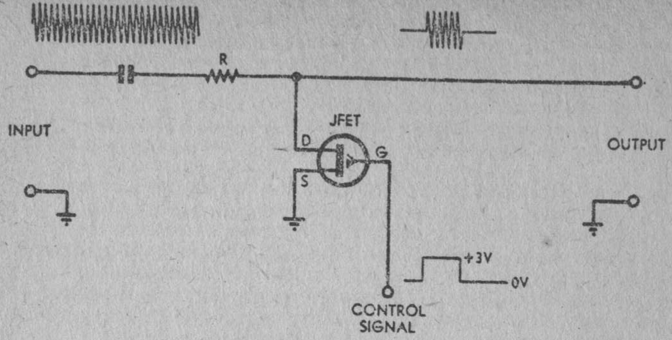

(a) BASIC JFET SHUNT ANALOG GATE

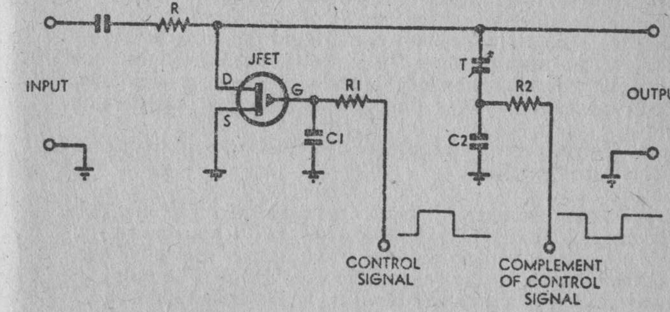

Fig. 2 (b) MODIFIED ANALOG GATE WITH TRANSIENT SUPPRESSION

Although this technique is widely used it is not without its critics. Many are not happy about the assumptions involved, and are dubious about the validity of simply equating sinewave conditions with a substitute power supply to transient conditions with an amplifier's actual power supply. Quite apart from such criticisms there is the practical inconvenience involved in matching a substitute regulated supply to an amplifier and connecting it in circuit for the testing.

Fortunately an alternative approach exists, and one which is not only more valid theoretically but also more convenient in practice. The method involves testing using short bursts of sinewaves delivered to the amplifier from an instrument such as that to be described, with the amplifier itself unaltered and powered from its own internal supply. The "tone bursts" used are quite easily arranged to have a duration and repetition rate which effectively duplicates infrequent musical programme peaks, while their relatively simple and basic nature makes them both easily reproduceable and also more readily measured than a musical signal.

It should be noted that a tone burst power output measurement does not really give the information originally signified by the term "music power"–nor, for that matter, does a measurement made using the "substitute power supply" method. Both methods give essentially the absolute maximum power output capability of an amplifier, under short-term conditions; and this is not strictly equivalent to the somewhat vague notion of "music power output".

Probably the most accurate way of visualising the situation is to regard the "music power" capability of an amplifier as a dynamic or variable parameter which depends at any instant upon the immediately preceding amplifier conditions, and which accordingly varies over a range defined at its lower limit by the normal continuous sinewave output rating and at its upper limit by the normal continuous sinewave output rating and at its upper limit by the rating provided by a tone-burst or "substitute supply" measurement.

Late in 1965 the American Institute of High Fidelity published a revised version of their earlier "IHF Amplifier Standard" rating system, in which they advocated the use of the term "dynamic power rating" to signify the short-term output power capability of an amplifier, replacing the term "music power rating". The method suggested for the measurement of dynamic power rating was the tone-burst method, using bursts of 10mS suration at a low reception rate.

Let us now turn to the instrument to be described, considering first its principle of operation.

The input signal to be gated is fed to both an analog gate and to a shaper-squarer. The latter derives from the signal a squarewave equivalent having extremely fast rise- and fall-times corresponding to the zero crossings of the original signal waveform. The output of the squarer is fed to a group of flip-flops which are arranged in "toggling" mode and connected in cascade to form a binary scaler or frequency divider.

A six-position switch (S1) is arranged to select either the direct output from the squarer or the scaled-down signal from a chosen scaler stage. The selected signal from the switch is then passed to an AND logic gate whose second input is controlled by a simple set-reset state and thus holds the AND gate closed; thus the signal selected by the switch cannot progress further.

The "set" (S) input of the control flip-flop is connected to a second switch (S2) which may be used to select the desired source of burst triggering signals. There are two triggering sources internal to the instrument, one a free-running "sampling oscillator" of adjustable frequency–for repetitive operation–and the other a manual on-off circuit operated by a front-panel push-button. A third position of the switch allows the use of external triggering sources.

On the arrival of a triggering pulse at its S input, the control flip-flop switches to its set state and accordingly opens the control AND gate. This excursion from the signal selected by S1 to pass through the gate and reach the "toggle" (T) input of a second flip-flop, which is a J-K type whose output controls the analog gate.

Under quiescent conditions this "main gate" flip-flop is in the reset state, and accordingly the analog gate is closed. The analog (tone) input signal is therefore prevented from reaching the output terminal. However, the arrival of the first toggling edge from the control gate switches the main gate flip-flop into the set state, and the gate is immediately opened to pass the tone signal.

The period of time for which the gate remains open depends upon the setting of switch S1, because the second toggling edge to reach the main gate flip-flop via the control gate will cause the former to be reset once more–closing the analog gate with it. And depending upon the setting of switch S1, this turn-off edge will correspond to the end of a single tone cycle, or two cycles, or four, eight, sixteen, or thirty-two. Hence the analog gate will have been open for a period which accurately corresponds to an integral number of tone cycles, and a "burst" of the desired number of tone cycles will have been passed to the output terminal.

In order to prevent further opening and closing of the analog gate due to the arrival of further toggling edges at the main flip-flop (from S1), a pulse derived from the main gate output is used to reset the control gate flip-flop immediately following the resetting of the main gate flip-flop. This closes the control gate and prevents further signals from reaching the main gate flip-flop from S1. Both flip-flops thus remain reset, and the analog gate remains closed, until the arrival of another triggering pulse from S2.

From the foregoing it should be apparent that it is the triggering pulses, supplied to the control gate flip-flop from S2, which initiate each tone burst; while the duration of each burst is governed by the tone-derived signal selected by S1. The burst repetition rate and the burst duration may therefore be adjusted independently.

Before turning to consider the full circuit diagram of the new instrument it may be worthwhile to look briefly at the operation of the analog gate circuit employed.

A number of circuits have been developed for this purpose; however, the circuit which has been selected in the present instance employs a p-channel junction field-effect transistor (JFET) as a voltage-controlled shunt switch. Its basic operation is shown in the diagram of figure 2(a).

As may be seen the gate circuit proper consists simply of a series input isolating resistor R and the JFET, whose drain-source channel is connected between the output terminal and ground. The effective drain-source resistance of the JFET is controlled by means of a control signal applied to the gate electrode.

When the control voltage is at zero, there is no applied transverse electric field in the JFET channel. Accordingly the drain-source is low–a few hundred ohms–and the input signal suffers heavy attenuation as a result of the divider action produced by resistor R. Under these conditions the gate is effectively "closed".

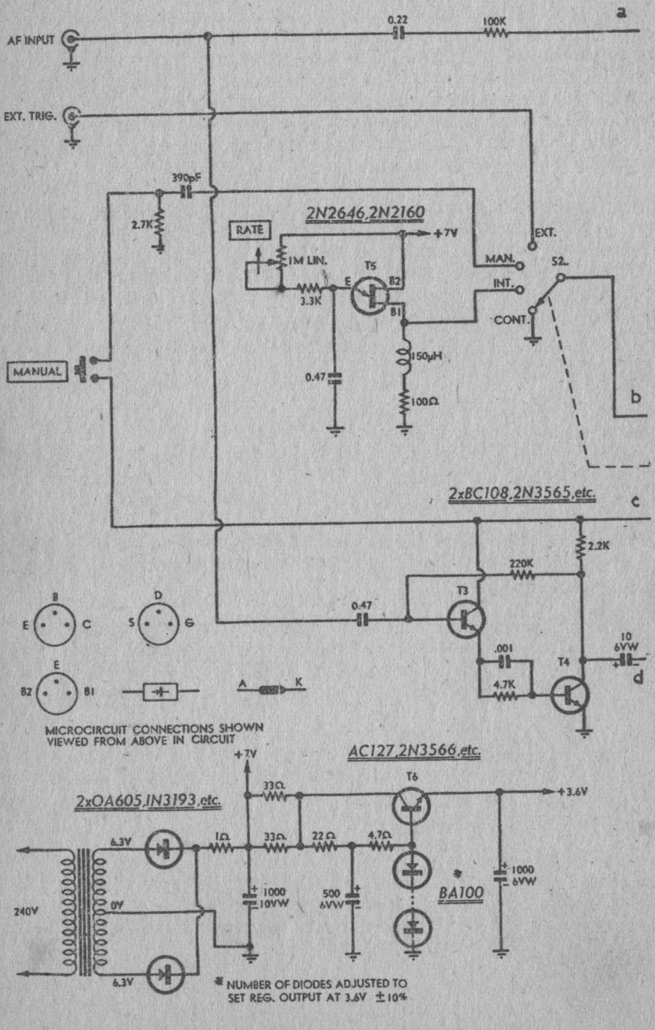
AF INPUT
0.22
100K
a
EXT. TRIG.
390pF
2.7K
2N2646,2N2160
RATE
+7V
1M LIN.
T5
E
B2
B1
3.3K
EXT.
MAN.
INT.
S2.
CONT.
MANUAL
0.47
150μH
100Ω
b
2xBC108,2N3565,etc.
c
2.2K
220K
T3
0.47
.001
4.7K
T4
10 6VW
d
B
E
C
D
S
G
E
B2
B1
A
K
+7V
AC127,2N3566,etc.
T6
33Ω
+3.6V
2xOA605,IN3193,etc.
6.3V
1Ω
33Ω
22Ω
4.7Ω
240V
0V
1000 10VW
500 6VW
BA100
1000 6VW
6.3V

MICROCIRCUIT CONNECTIONS SHOWN VIEWED FROM ABOVE IN CIRCUIT

* NUMBER OF DIODES ADJUSTED TO SET REG. OUTPUT AT 3.6V ±10%

Connects with drawing on Page 81

Connects with drawing on Page 80

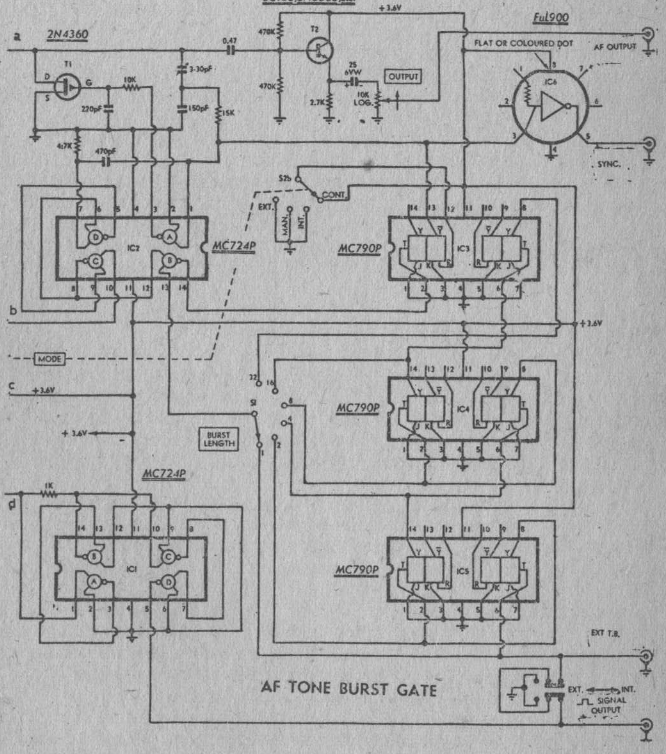
BC108,2N3565,etc.
+3.6V
Ful900
a
2N4360
0.47
470K
T2
FLAT OR COLOURED DOT
AF OUTPUT
T1
D
G
S
10K
3-30pF
220pF
150pF
15K
470K
25 6VW
OUTPUT
2.7K
10K LOG.
IC6
SYNC.
4·7K
470pF
S2b
CONT.
EXT.
MAN.
INT.
IC2
MC724P
IC3
MC790P
b
+3.6V
MODE
c
+3.6V
S1
IC4
MC790P
+3.6V
BURST LENGTH
MC724P
1K
d
IC1
IC5
MC790P
EXT T.B.
EXT.
INT.
SIGNAL OUTPUT

AF TONE BURST GATE

When the gate is taken positive, however, the reverse bias applied to the gate-channel junctions of the JFET produces a transverse field which "pinches off" the channel. The drain-source resistance rises to a very high value, and the input signal therefore appears at the output terminal suffering only slight attenuation. The gate under these conditions is effectively "open".

The switching action of this type of analog gate is free from many of the problems associated with other circuits. Because the drain-source channel requires no DC bias, and furthermore acts over a fairly wide signal amplitude range as an almost perfectly linear resistor (of controllable value), there is virtually no DC offset component or "pedestal" associated with resultant tone bursts.

The available channel resistance range is quite wide, even with moderate control voltages, giving the gate a high "on" to "off" ratio combined with a low insertion loss in the "on" state. Hence the gate circuit used in the final instrument exhibits an on-off switching ratio of more than 40dB, with an insertion loss of only about 5dB.

The shunt JFET gate is also capable of switching at a high speed although the operation at high speeds is complicated by inter-electrode capacitances. Drain-source capacitance tends to increase the insertion loss and reduce the switching ratio at high signal frequencies, while gate-source capacitance tends to slow down the rise and fall of the gate control voltage. However, more troublesome than both of these is gate-drain capacitance, which tends to feed control signal energy into the signal circuit.

The effect of the latter is that sharply differentiated switching "spikes" tend to appear across the analog output terminals. A large positive spike appears as the gate opens, and a smaller and narrower negative spike as the gate closes. The differences in amplitude and duration arise as a result of the differing dynamic R and C situations present in the JFET for turn-off and turn-on. As the spikes represent spurious gate output or "noise", it is important that they be suppressed.

A number of suppression techniques may be used. However, the method which has been adopted in the present instrument is shown in the diagram of figure 2(b). A two-step process, it works partly by slowing down the attack and decay of the JFET gate control signal by means of the low-pass filter formed by resistor R1 and capacitor C1. This reduces the charge stored in the JFET gate-drain capacitance, by reducing the rate of change of voltage. (The reduced switching speed is still high relative to the analogue signals being switched, so that the gate operation is not impaired.)

The second step involves cancellation or nulling the remaining spikes, in effect by injecting into the analog circuit spikes of opposite polarity. This is performed by an adjustable trimmer capacitor T which is connected to a source of signal complementary to the JFET control signal. A second low-pass filter comprising R2 and C2 is used to adjust the attack and decay of the nulling signal to correspond with the JFET gate signal.

At a critical adjustment of the trimmer the net charge transfer to or from the analog circuit during gate opening and closure is reduced to a very low level. The tone bursts produced by the gate are then quite "clean".

Reference to the main circuit diagram of the complete instrument will show that many of the logical functions are performed by integrated digital microcircuits. Six such devices are used, five of which are in the "dual in-line" encapsulation while the remaining device is of the TO-5 eight-lead type. All six devices are of the economy plastic RTL type, the dual in-line devices being from the Motorola Semiconductors' MC700P range and the TO-5 device from the Fairchild Semiconductor FuL900 range.

The cost of these devices is quite modest when one considers the number of discrete components which would be required to perform the same functions—if one includes the functionally identifiable devices within the microcircuits, for example, the total transistor count for the complete tone-burst gate is greater than 90! From this figure it may also be deduced that, quite apart from the cost advantage, the use of microcircuits simplifies quite drastically the construction of the instrument.

The shaping-squaring function of the instrument is performed by transistors T3 and T4 together with IC1, which is an MC724P quadruple 2-input gate. The transistors are employed in a simple input preamplifier circuit having a moderately high input impedance. Output from T4 is coupled to gate A of IC1, and the output of this gate in turn feeds to gate B. Both gates are used simply as squarer/inverters, with a small amount of positive feedback from B's output back to A's input via a 1K resistor to improve input sensitivity. The two remaining gates of IC1 are connected in cascade from the output of B, to give additional squaring; the output from gate D accordingly has quite short rise–and fall–times.

Although in the simplified instrument block diagram of figure 1 the output of the shaper-squarer block is taken directly to the scaler and selector switch, in the instrument itself this connection is made by means of a small slider switch mounted on the rear of the case. In addition both the shaper output and the scaler-selector input are made available at adjacent coaxial connectors. This allows the internal connection between the two to be broken at will, and external arrangements made either for use of the squared signal and/or for use of substitute gating signals.

As a result of this simple and inexpensive refinement not only is the instrument made more flexible in terms of its main function, but it is also rendered capable of performing additional functions. Thus the squarer section becomes available for external use, as does the scaler section. Use of the latter section as an external scaler/frequency divider will be discussed further shortly.

The scaler section itself consists of five binary J-K flip-flops, each of which is half an MC790P microcircuit. Devices IC4 and IC5 provide four of the flip-flops, while the fifth is the element in IC3, remote

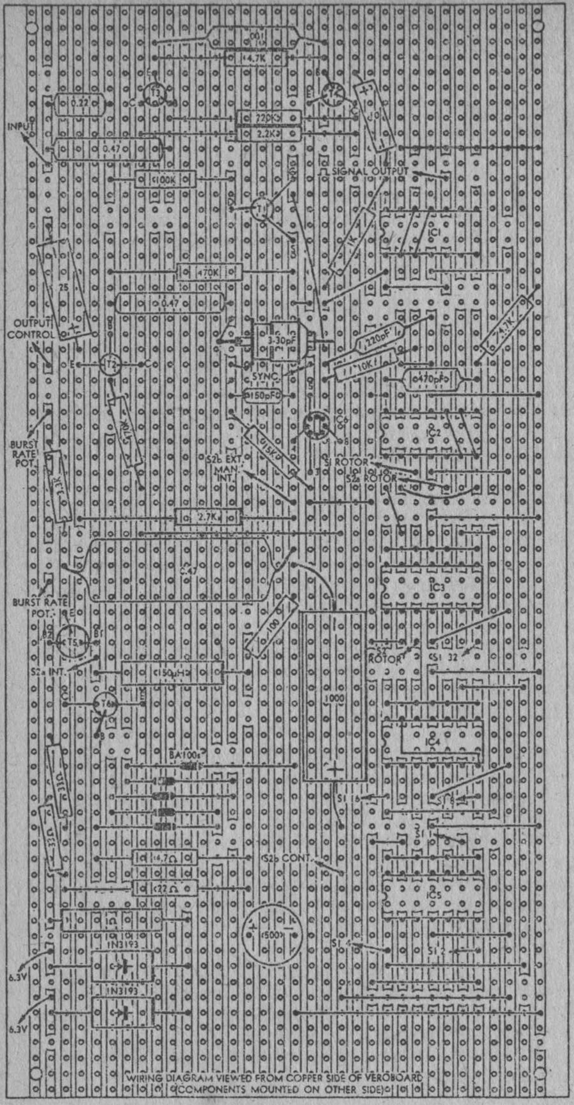

This diagram of the wiring board for the tone-burst gate shows the interconnections and most of the minor components.

from the indexing notch (pins 5-10). Each flip-flop has both J and K gating inputs taken to "1" (ground) to ensure simple toggling mode operation. As before switch S1 is used to select the desired gating signal, the switch being labelled "burst length".

The control gate function is performed by gate B of IC2, which is a second MC724P quadruple gate device. Gates C and D of this device form the control gate flip-flop, being interconnected in the usual manner to form a simple S-R flip-flop. Switch S2a forms the triggering selector as before, with positions for internal, manual and external triggering modes. Note, however, that the switch has an additional fourth position, in which no triggering input is admitted; the position is a "gate continuously open" mode, provided for setting-up purposes, and will be discussed further in a moment.

The internal triggering generator is a simple relaxation oscillator circuit, employing an economy unijunction transistor type 2N2646, or similar and whose frequency may be adjested over the approximate range 3-500Hz. A small inductor is employed in the unijunction B1 circuit to sharpen the output pulse waveform, as a sharp pulse of less than 20uS duration is necessary to preclude the appearance of spurious secondary bursts at high signal frequencies.

The "Manual" triggering circuit consists simply of a panel push-button together with a low-value differentiating capacitor and discharging resistor. The design of the triggering circuitry is such that contact bounce has virtually negligible effect upon instrument operation; however, it is wise to employ a high quality push-button, preferably one with a rapid-acting "micro-switch" movement.

The external triggering input will accept signals varying between OV and +3V relative to groond, with the +3V level corresponding to logical "1". For normal operation, in which the burst length gating is synchronous but the repetition rate is not, the triggering signal should consist of sharp pulses with a duration of less than 20uS.

If it is desired, locked burst operation may be obtained quite simply by merely connecting a source of +3V DC to the external triggering input—with S2 in the appropriate position, of course. When this is done the control gate flip-flop is clamped in the set position and prevented from resetting on the arrival of the "lock-off" pulse from the main gate flip-flop at the end of each burst. The instrument thus produces a continuous series of bursts spaced such that the inter-burst interval is automatically equal to the burst duration.

A +3V DC signal connected to the external triggering input is also used when it is required to use the scaler section of the instrument for external scaling or frequency division.

The main gate flip-flop function is performed by the second element of IC3, again a J-K element connected for simple toggling mode operation. As the output voltage swing from this element is inadequate for direct operation of the analog gate, it is passed through gate A of IC2 for amplification and incidental inversion. The direct flip-flop output is also differentiated to produce the control gate flip-flop "lock-off" signal. Additionally it is used to derive the analog gate transient suppression signal, and also to provide an oscilloscope synchronising signal via IC6. The latter device is employed as an isolating buffer-inverter, and accordingly delivers a +3V rectangular signal corresponding to the "open" state of the analog gate.

It is the IC6 "sync" output connector which may be used as the output of the scale section of the instrument when the latter is required for external use. With the signal to be scaled or divided applied to the "Ext T.B." input connector, the slider switch in the "Ext" position, and +3V DC applied to the external triggering input, the selector switch S1 may be used to adjust the scaling ratio over the binary range from 1:2 to 1:64. Note that when the instrument is used in this way the main gate control flip-flop is used as a sixth scaler stage.

As mentioned earlier there is a fourth position of the "mode" switch S2, in which the analog gate is kept "open" continuously to permit setting-up of auxiliary equipment. To provide this mode of operation a second pole of S2 is used to apply +3.6V to the "reset" input of the main gate flip-flop, holding the latter locked in the reset state. As output is taken from the Y-complement output of the flip-flop rather than from the Y output, this results in the analog gate being locked in the "open" condition.

The analog gate of the instrument is of identical form to that already discussed and illustrated in figure 2. It employs a P-channel economy silicon JFET type 2N4360 or similar, and as a result of the relatively modest demands made by this mode of operation, most such devices should work satisfactorily despite the wide spread in parameters. A small proportion of devices may prove unsuitable by virtue of unduly low gain or channel resistance, but these should be sufficiently rare to make replacement an economic proposition.

In order to achieve maximum switching ratio combined with minimum insertion loss, an emitter follower stage is employed following the analog gate as an impedance transformer and buffer. As the dynamic range of the gate itself is but modest, the emitter follower functions quite satisfactorily from the +3.6V microcircuit supply line.

For optimum performance the analog gate and emitter follower should be operated at a signal input level of between 250 and 500mV RMS, as this provides the best noise/distortion compromise. In view of this, and because in most applications a burst signal of adjustable amplitude will be required, the output of the emitter follower is controlled by a simple potentiometer attenuator.

The power supply of the instrument is a relatively simple circuit supplying approximately 7V unregulated for the unijunction trigger oscillator and a regulated 3.6V + or - 10% for the remaining circuits. The unregulated supply is a simple full-wave rectifier employing two silicon diodes type OA605, 1N3193 or similar, and a miniature centre-tapped 12.6V step-down transformer.

The regulator is a simple series type employing either a germanium type AC127 or a silicon type 2N3566 medium-power NPN transistor. The base reference voltage is well filtered and regulated, the latter being performed by a string of forward-biased silicon diodes in view of the relative scarcity of low-voltage zener diodes. The number of diodes is arranged so that the regulated output voltage at the transistor emitter is within + or - 10% of its nominal 3.6V value.

As may be seen from the photographs the instrument is housed in the small flanged front panel instrument case used for most of our recent instrument designs, measuring 7½in x 5in x 4in. There are only four main front-panel controls, comprising the "burst length" and "mode" switches together with the "burst rate" and "output" potentiometers. The remaining fittings consist of the push-button for manual operation and the four main coaxial connectors.

At the rear of the case are the squared-signal output and external timebase input connectors, the "Int-Ext" slider switch and the mains cord entry.

Inside the case, most of the components are mounted on a "Veroboard" perforated printed-strip conductor panel measuring 7in x 3¾in; the panel is a section of stock Veroboard strip. The position of components and interconnection of copper conductors should be fairly clear from the board diagram provided, and also from the internal photographs.

The Veroboard wiring panel is supported by two small sheetmetal brackets which are attached to the rear of the front panel by the input connector mounting screws and the "burst length" switch mounting bush respectively. The brackets mount the panel at a distance of 2in behind the front panel, and close to the top of the case.

Wiring of the components on the board is a fairly straightforward operation, and should present few problems. However, care should be taken to ensure that components are not damaged either by mechanical strain during the lead bending and insertion, or by overheating during the soldering operation. The soldering itself should be done quickly and carefully with a low to medium power iron having a small bit; a steady and sure hand is required to ensure that the copper conductors are not overheated nor accidentally shorted by tiny solder links.

The microcircuit connection leads should not be cut, but simply passed through the board holes, bent over carefully and soldered. It will be found that the "dual in-line" devices have small moulded projections which automatically space the device from the board. The TO-5 device should be spaced approximately 3/8in from the board, which spacing may also be used for the transistors.

To prevent strain on the Veroboard panel the power transformer is mounted in the lower rear of the case, with the mains cord, primary and secondary leads terminated on a 5-lug tagstrip alongside. The mains cord is clamped inside the case to prevent strain on the tagstrip.

Three minor components are supported by convenient switch and potentiometer lugs behind the front panel rather than on the Veroboard panel. These components are the 2.7K resistor and 390pF capacitor associated with the manual triggering button, and the 1000uF 10VW power supply input reservoir electrolytic capacitor. The first of these connects between the puch-button and the input connector earth bus, while the second connects between the appropriate lug on S2 and an unused lug on S1. The 1000uF electrolytic connects between the +7V lug on the "burst rate" pot and an earth lug on the "mode switch".

When wiring the unit it is wise to connect only four BA100 diodes into the power supply reference circuit. This will ensure that when the unit is first switched on the regulated output will not be excessively high. Upon switch-on the voltage should be checked, and an additional diode or diodes added as necessary to bring the supply within the required range.

The only other adjustment to be made to the completed instrument is that concerning the switching transient suppressor trimmer. This adjustment is most easily made with a sensitive wide-band oscilloscope connected to the AF output, and in the absence of AF signal input. To permit gate operation under these conditions an external square-wave gating signal must be connected to the "Ext T.B." socket, with the slider switch set appropriately.

It will be found that at a critical setting of the trimmer, switching transients are reduced to a minimum. The trimmer should be locked in position at this setting with a small dab of nail lacquer or similar cement, to prevent vibration disturbance. If difficulty is experienced in obtaining a null balance, or if the null is insufficiently accurate to reduce the transients below 30mV P-P, it may be necessary to modify the R-C filter values in either the JFET gate circuit, or the trimmer feed circuit, or both.

LIST OF PARTS:

1 Small instrument case, 7½in x 5in x 4in, with flanged front panel. Handle, rubber feet, scrap aluminium for wiring board brackets.
1 7in length of Veroboard panel
1 Miniature power transformer, 12.6V CT
1 Rotary switch, 1 pole, 6 positions
1 Rotary switch, 2 poles, 4 positions
1 Slider switch, 2 poles, 2 positions
1 Miniature push-button
6 Small coaxial sockets
1 Medium fluted control knob
2 small fluted control knobs
1 Small skirted knob, numbered

SEMI-CONDUCTORS AND I.C.s
3 MC790P dual flip-flops
2 MC724P quadruple gates
1 FuL900 inverting buffer
1 2N4360 junction FET
1 2N2646, 2N2160 or similar UJT
3 BC108, 2N3565 or similar
1 AC127, 2N3566 or similar
2 OA605, 1N3193 or similar
4-5 BA100 or similar (see text)

RESISTORS (5 p.c. tolerance, ½W for fixed values)
1 x 1 ohm, 1 x 4.7 ohm, 1 x 22 ohm, 2 x 33 ohm, 1 x 100 ohm,
1 x 1K, 1 x 2.2K, 2 x 2.7K, 1 x 3.3K, 2 x 4.7K, 1 x 10K, 1 x 15K,
1 x 100K, 1 x 220K, 2 x 470K
1 x 10K log pot
1 x 1M linear pot

CAPACITORS
(Fixed values, low voltage plastic, except electrolytics)
1 x 150pF, 1 x 220pF, 1 x 390pF, 1 x 470pF, 1 x .001uF, 1 x .22uF,
3 x .47uF
1 3-30pF ceramic air trimmer
1 10uF 6VW electro
1 25uF 6VW electro
1 500uF 6VW electro
1 1000uF 6VW electro
1 1000uF 10VW electro

MISCELLANEOUS
1 x 150uH RF inductor, power cord and plug, grommet for cord entry, nuts, bolts, washers, solder lugs, connecting wire, solder, etc.

SIGNAL INJECTOR USES IC

I decided to make up an injector using a Motorola IC type MC799/P and the circuit shows what I came up with. It is basically a battery powered 1000Hz multivibrator that generates square waves. The amplitude of the square wave output is fully variable and sufficient to test a loudspeaker.

The injector also provides a wide bandwidth RF signal suitable for testing RF and IF stages. Current drain is of the order of 80mA which allows a reasonably long battery life. The resistors used were ¼W types and the 1K potentiometer is a carbon type with switch. The two 0.1uF capacitors are 10V disc type, the 100pF capacitor is also a disc type.

Construction may be left to the individual ideas and requirements of the builder. Many ideas will no doubt suggest themselves, possibly dictated by what materials and facilities happen to be available at the time.

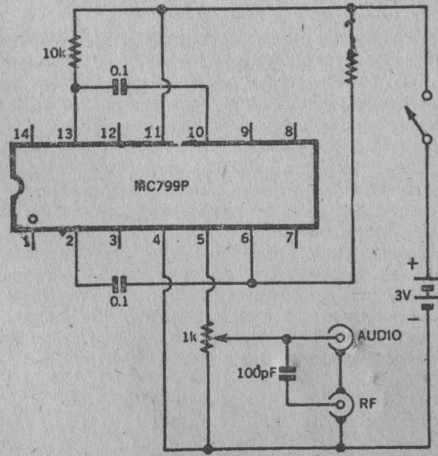

SIMPLIFIED VERSION OF SQUARE WAVE GENERATOR

After experimenting and consulting the Fairchild data sheets, which gave the output of the FuL914 as 16 units and the input of the FuL900 as 6 units, I decided that the FuL914 would not be overloaded by the FuL900. Tests showed that the circuit operated satisfactorily in this mode, so I built up the complete unit as illustrated. In addition, I left the No. 2 lead of the FuL914 floating in order to use it as a sync. input, should this be necessary. AC and DC outputs were also made separate.

The performance seems to be unchanged by the modifications, since my generator has a frequency range of 4Hz to 2MHz. You may be interested in my experience with the modified circuit.

In addition, I would draw attention to the method of construction adopted, which is much less messy and time consuming than etching your own boards. Basically, it involves the removal of thin strips of copper from the board with a suitable scribing tool, to provide the necessary pattern. The method, incidentally, leaves intact a maximum amount of copper and, with it, the ability to make the circuit changes which are often necessary in '1 off' designs.

The approach is first to collect all the components required for the project, since physical placement is most importants. Then draw the circuit on the copper side of the board in pencil, which can be rubbed out, if necessary, with an ordinary eraser. Leave room for the actual size of the components involved and, for preference, arrange them in a neat and orderly fashion. After a certain amount of planning and possible rearrangement, it will become apparent where lines have to be scribed to separate the components.

Fine parallel lines are scribed to cut through the copper and a tool, made from a specially ground section of hacksaw blade, can be used to roll the copper out from between the scribed lines. A little practice is required to get the knack; too steep an angle will cause the tool to dig in; too shallow an angle will cause the blade to skip out of the groove.

Having drilled the holes for the component pigtails, the board can be viewed against the light from both sides to pick out any copper which has been missed or copper feathers which might short across the grooves. The components can then be soldered in, taking care to use a heat shunt when soldering leads to delicate components. If the planning has been done well, the need for jumpers will have been minimised but, if desired, these can often be hidden under components. We use a standard knife devised for scribing Laminex and similar sheeting. During development of the original square-wave generator, we passed through the simpler form suggested but decided on the additional stages: (a) to render the device less dependent for its operation on a degree of unbalance in the basic multivibrator sections and (b) to preserve a near-unity mark-space ratio irrespective of output load conditions. However, if "1-off" units constructed to the simpler design work satisfactorily, a worthwhile economy will have been effected. If particular units fail to operate, it may be necessary to try another FuL914

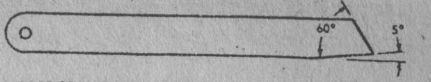

PIECE OF HACKSAW BLADE WITH TEETH GROUND OFF AND GROUND TO ANGLES SHOWN TO MAKE COPPER LIFTING TOOL

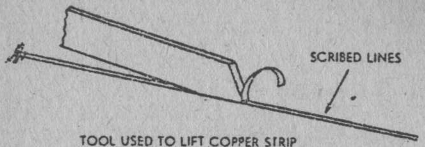

TOOL USED TO LIFT COPPER STRIP

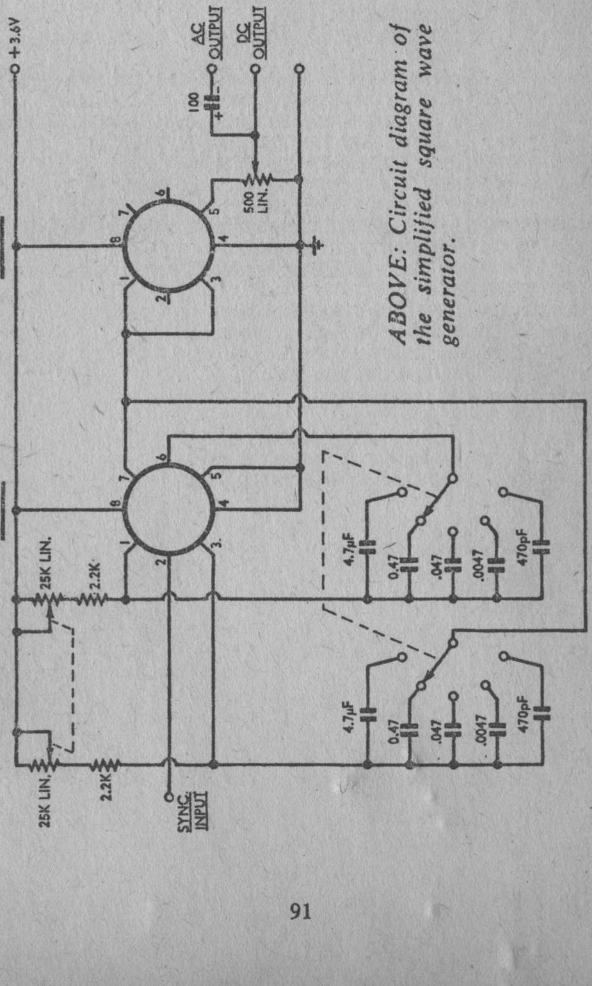

ABOVE: Circuit diagram of the simplified square wave generator.

EQUIVALENTS TABLE FOR SEMICONDUCTORS & IC's

1N914=RS276-1139. AA117-118. 1N618. OA91/5.
1N4004=RS276-1103/1138. SK3031. BY137/400
1N4148=RS276-1139. AA117-118. 1N618. OA91/5.
1N5339B=Zener diode
2N2646=2N2160. RS276-2029. TIS43
2N3565=BC108-209. RS276-2009. 2N4967
2N3638A=MPS3638. TIS50. 2N5447. RS276-2021. SK3025.
2N3643=RS276-2033. BSW54. 2N2219A
2N4258=RS276-2023. BS W25. 2N3546.
95H90. GHJ121. Divide by 10/11 prescaler
741/C=RS276-010/1741. CA3056-3741. MC1539-1741. TBA221/2 OP Amp
7400=RS276-1801. FJH131, FLH101-105. MC408-458-508-558. Quad 2 Input NAND Gate
7490=9390. FJJ141. RS276-1808 Decade counter
9001=7470. 8470. FJJ101. FLJ101-105.
AC127=RS276-2001. 2N3566-647-1304-1308-2430. 2SD96. SK3010. AC141-176-179-181.
BA100=RS276-1136. 1N659-914/A-4863. BA108-127-215. BAV18. BAW63.
BC107=RS276-2031. BC147-167-207-317. 2N2921-3568. 2SC458.
BC108=RS276-2009. BC148-168-208. SK3020. MPS6520. 2N3564-3711-4131.
BC177=RS276-2024. BC204-212-261-307-512-557. 2SA565.
BZX79C12=RS276-563/623. 400mW 12Volt Zener Diode
BZX79C15=RS276-624/564. 400mW 15Volt Zener Diode
BZY88C5V1=400mW 5. 1 Volt Zener Diode
EM401=RS276-1139-1101-1105. BY127. 1N4007. BY126/50. BAY219.
EM408=BY127. 1N4006. RS276-1114. 1S107.
EM4050=RS276-1139-1101-1105. BY127. 1N4007. BY126/50. BAY219.
FuL900=U8B900 Buffer
FuL914=U8A914. RS276-015. Dual 2 input gate
LM309K=UA7805. RS276-1830. Voltage regulator
MC724P=MC824-924. Quad 2 input NOR gate
MC790P=MC890-990. Dual JK Flip-Flop
MC799P=MC899-999=Dual Buffer
NE555=RS276-1723/1728. Timer
OA91=RS276-1102/1136. AA117-118. 1N618. OA95.
OA95=RS276-1102/1136. AA117-118. 1N618. OA91
OA605=1N3193. RS276-1139. 1N4003. 1S101. BY127

BERNARDS & BABANI PRESS RADIO AND ELECTRONICS BOOKS

BP1	First Book of Transistor Equivalents and Substitutes	40p
BP2	Handbook of Radio, TV and Ind. & Transmitting Tube & Valve Equiv.	60p
BP3	Handbook of Tested Transistor Circuits	40p
BP4	World's Short, Medium & Long Wave FM & TV Broadcasting Stations Listing (International Edition)	60p
BP5	Handbook of Simple Transistor Circuits	35p
BP6	Engineers and Machinists Reference Tables	30p
BP7	Radio and Electronic Colour Codes and Data Chart	15p
BP8	Sound and Loudspeaker Manual	50p
BP9	38 Practical Tested Diode Circuits for the Home Constructor	35p
BP10	Modern Crystal and Transistor Set Circuits for Beginners	35p
BP11	Practical Transistor Novelty Circuits	40p
BP12	Hi-Fi, P.A., Guitar & Discotheque Amplifier Design Handbook	75p
BP13	Electronic Novelties for the Motorist	50p
BP14	Second Book of Transistor Equivalents	95p
BP15	Constructors Manual of Electronic Circuits for the Home	50p
BP16	Handbook of Electronic Circuits for the Amateur Photographer	60p
BP17	Radio Receiver Construction Handbook using IC's and Transistors	60p
BP18	Boys & Beginners Book of Practical Radio and Electronics	60p
BP22	79 Electronic Novelty Circuits	75p
BP23	First Book of Practical Electronic Projects	75p
BP24	52 Projects using IC741 (or Equivalents)	75p
BP25	How to Build Your Own Electronic and Quartz Controlled Watches & Clocks	85p
BP26	Radio Antenna Handbook for Long Distance Reception & Transmission	85p
BP27	Giant Chart of Radio, Electronic, Semiconductor & Logic Symbols	60p
100	A Comprehensive Radio Valve Guide – Book 1	40p
121	A Comprehensive Radio Valve Guide – Book 2	40p
126	Boys Book of Crystal Sets	25p
129	Universal Gram-Motor Speed Indicator (Combined 50 & 60 ∿ model)	10p
138	How to Make Aerials for TV (Band 1–2–3)	25p
143	A Comprehensive Radio Valve Guide – Book 3	40p
150	Practical Radio Inside Out	40p
157	A Comprehensive Radio Valve Guide – Book 4	40p
160	Coil Design and Construction Manual	50p
161	Radio, TV and Electronics Data Book	60p
170	Transistor Circuits for Radio Controlled Models	40p
177	Modern Transistor Circuits for Beginners	40p
178	A Comprehensive Radio Valve Guide – Book 5	40p
183	How to Receive Foreign TV Programmes on your Set by Simple Modifications	35p
195	High Fidelity 14 Watt Amplifier Design Chart	15p
196	AF – RF Reactance-Frequency Chart for Constructors	15p
197	Inexpensive Push-Pull Amplifier Construction Chart	15p
200	Handbook of Practical Electronic Musical Novelties	50p
201	Practical Transistorised Novelties for Hi-Fi Enthusiasts	35p
202	Handbook of Integrated Circuits (IC's) Equivalents and Substitutes	75p
203	IC's and Transistor Gadgets Construction Handbook	60p
204	Second Book of Hi-Fi Loudspeaker Enclosures	60p
205	First Book of Hi-Fi Loudspeaker Enclosures	60p
206	Practical Transistor Circuits for Modern Test Equipment	60p
207	Practical Electronic Science Projects	75p
208	Practical Stereo and Quadrophony Handbook	75p
209	Modern Tape Recording Handbook	75p
210	The Complete Car Radio Manual	75p
211	First Book of Diode Characteristics Equivalents and Substitutes	95p
214	Audio Enthusiast Handbook	85p
215	Shortwave Circuits and Gear for Experimenters and Radio Hams	85p
216	Electronic Gadgets and Games	85p
218	Build your own Electronic Experimenters Laboratory Using IC's	85p
RCC	Resistor Colour Code Disc Calculator	10p

BABANI PRESS & BERNARDS (PUBLISHERS) LIMITED
The Grampians, Shepherds Bush Road, London W6 7NF
Telephone Number: 01-603 2581/7296

R6809 ★ $1.45 ★ BANTAM SCIENCE

KNOWLEDGE
THROUGH
COLOR

ELECTRONICS

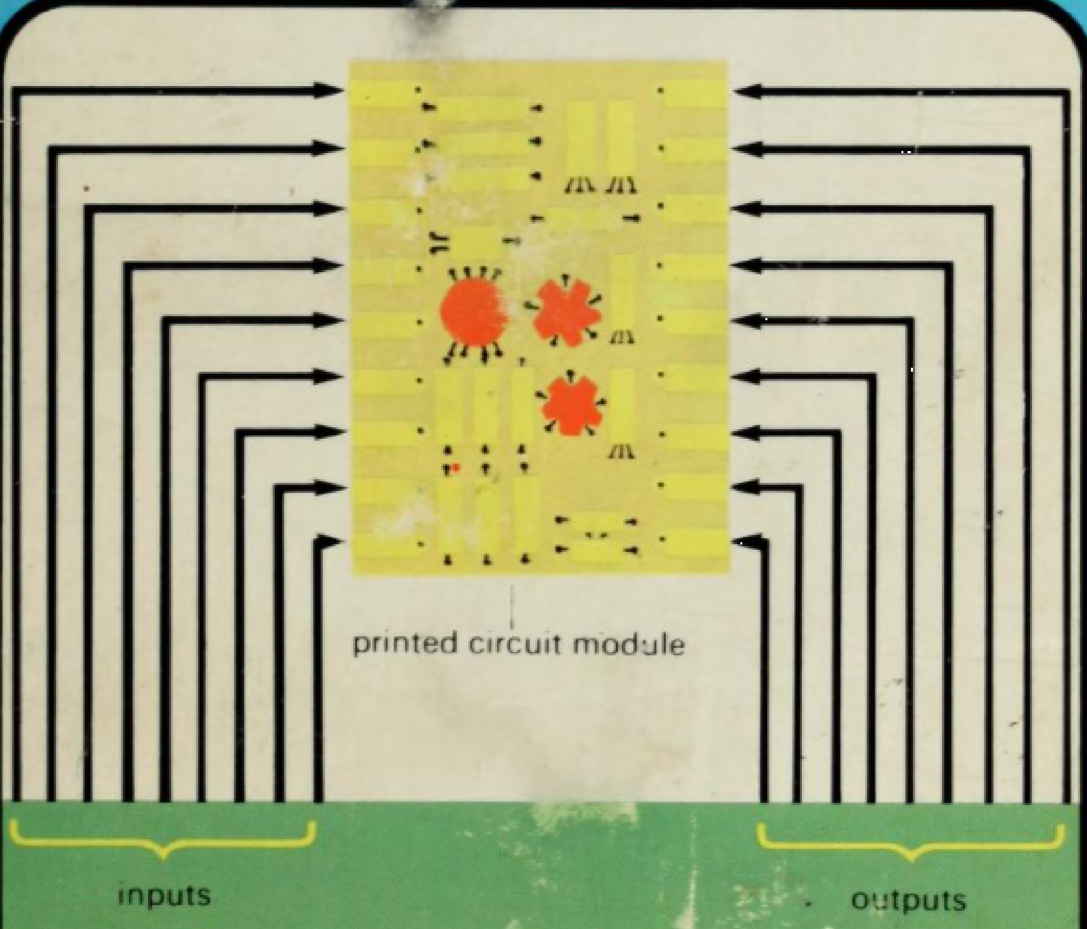
printed circuit module
inputs
outputs

BY ROLAND WORCESTER • FULLY ILLUSTRATED IN COLOR

Electrons in Motion

We can make a picture of an atom as simple or as complicated as we like. We can show a hydrogen atom as a central nucleus or proton, circled by a single electron [1]. The proton has a positive charge, and is 1,840 times the mass of the electron, which has a negative charge.

The electron circles the nucleus millions of times in a millionth of a second, and hundreds of millions of atoms side by side would be needed to occupy an inch of space. Positive and negative charges (proton and electron) neutralize each other.

Helium has two electrons and two protons. The heavier elements have more electrons [2]. Uranium has 92.

There are other particles besides protons and electrons. The neutron is also found in some atoms. It has no charge; however, it is the 'glue' that holds the proton mass together. Other transitory particles arise, change their state, or apparently disappear in energy.

A hydrogen atom electron can absorb a photon or light particle, and take energy from it, moving to an orbit at a higher energy level. Or electrons may collapse to lower energy levels, emitting photons.

Ions

We have seen that, normally, the negative electron charge equals the positive nucleus charge: hence the atom is in equilibrium. However, sometimes an atom may gain an electron and then it becomes a negative ion. If an atom loses an electron, it becomes a positive ion. With semiconductors, we may speak of positive 'holes' or hole carriers. These can be imagined as spaces which could be occupied by electrons.

A body with a considerable excess of electrons has a strong negative charge; one with a relative lack of electrons has a strong positive charge [3]. Thus there is a difference of potential, which exists so long as there is no means of electrons escaping, or nuclei gaining the electrons they lack.

Electrons, being negative, repel one another when they are close together, but are attracted by positive nuclei. A force draws the one to the other. If they are supplied to a *conductor* that allows electrons to move, the force draws them along the conductor to a positive ion region lacking electrons [4]. The force is electromotive force (e.m.f.) and the movement is current.

The Electric Current

If the flow of electrons is to be maintained, a supply must be generated or produced to replace those traveling away from the supply along the conductor. In a circuit, current will flow from the negative battery terminal (supply), along the conductor to a lamp or other apparatus (the load). The electrons that have not been dropped across the load will return along another conductor back to the positive battery terminal.

A supply of electrons can be produced chemically by a

1. Simple picture of a hydrogen atom.

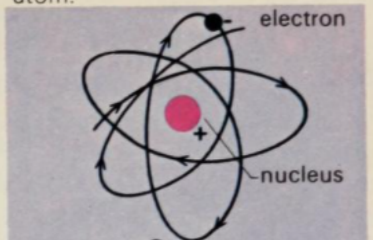

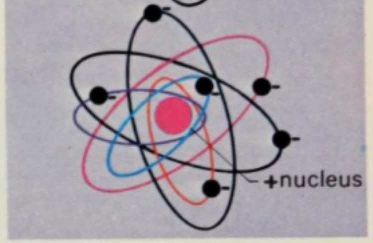

2. Heavier atoms have more electrons.

3. Potential differences between charges of dissimilar polarity.

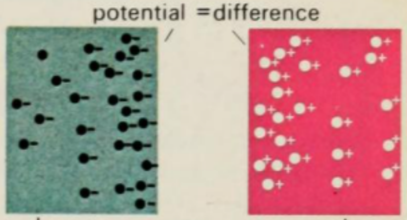

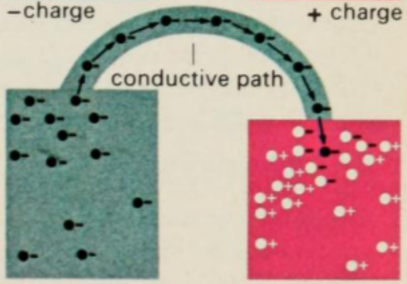

4. Negative electrons flowing to influence the positive charge.

battery, electromagnetically by a generator or by several other methods.

Conductors

We use conductors to take current to the point where it is wanted. A conductor readily allows the exchange of free orbital electrons, so electrons forming the electric current can move from atom to atom in the material [5]. Silver and copper are good conductors, for they allow current to flow freely. Copper is drawn into wires, or wound into coils or used to carry current [6].

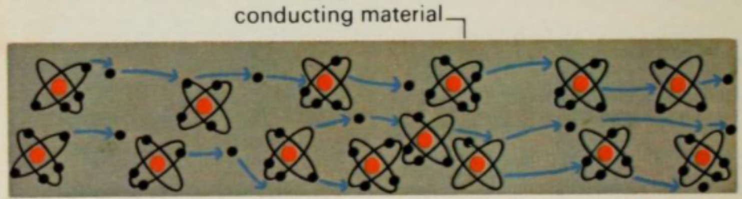

5. Easy exchange of electrons forms current.

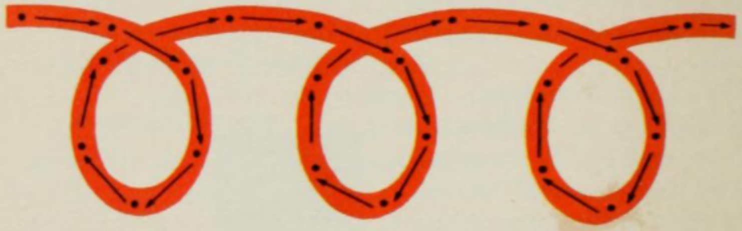

6. We take currents where needed by using conductors.

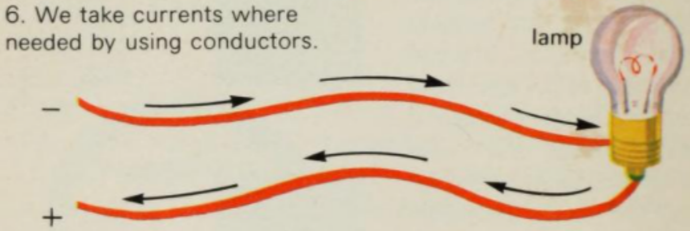

To find resistance of a conductor:

$$R = \frac{kl}{A}$$

where R = Resistance
k = Material constant
l = Length of wire
A = Cross-sectional area (usually in circular mils)

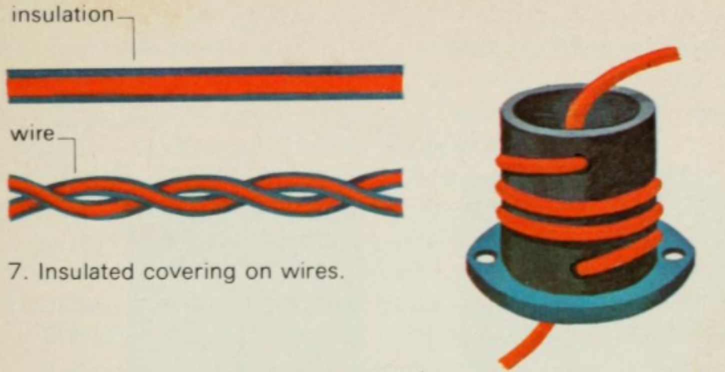

7. Insulated covering on wires.

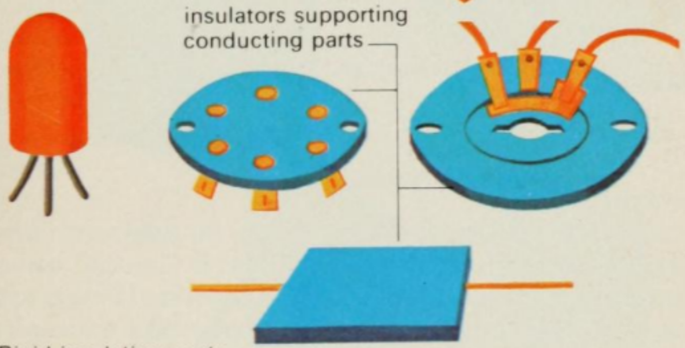

8. Rigid insulating parts.

Insulators

Some materials are poor conductors. This fact makes them good insulators. A perfect insulator would allow no exchange of electrons. Hence there would be no movement from atom to atom, and no current would flow. Pyrex, micalex and steatite are good insulators, but so are many other substances. Rubber, polyvinyl chloride (pvc), cotton, silk, enamel or other insulating material on wires prevents the wires touching each other or nearby metal parts where current is not wanted [7].

Glass, mica, ceramic, phenolic plastic and other rigid insulating parts support conducting items in switches, tube sockets, transistors, capacitors and many other electronic components [8].

What is the best kind of insulator depends on the work it must do. Insulators on wires may need to be thin and flexible. Insulating parts may have to be very strong or may have to withstand high voltages or high temperatures.

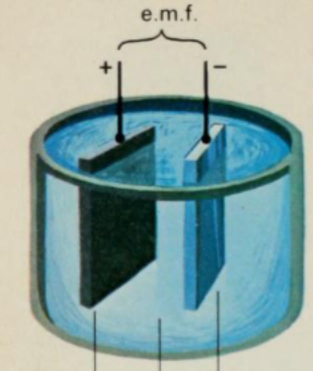

9. Simple cell generates electromotive force chemically.

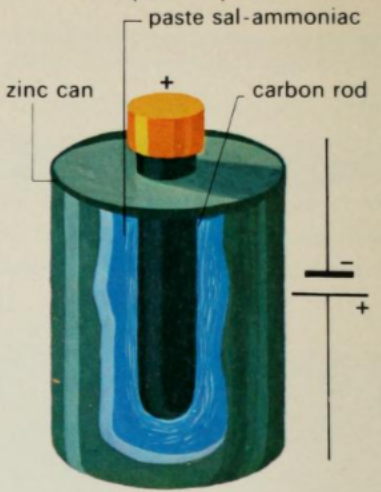

10. Popular dry cell

Primary Cells

We said that a potential difference can be produced chemically. Primary cells or batteries do this. If zinc and carbon plates are placed in a solution of sal-ammoniac [9] this happens. Chlorine atoms in the sal-ammoniac take incomplete atoms from the zinc, but each atom removed leaves two electrons. An available supply of electrons, or negative charge, distributes evenly on the zinc plate.

The removed zinc atoms enter the solution as positive ions, or atoms lacking some electrons, until the solution can accept no more. The solution then has a positive charge and so has the carbon plate immersed in contact with it.

A potential difference or electromotive force (e.m.f.) now exists between the zinc and carbon plates. If a wire is connected from zinc to carbon, electrons flow through it. Chemical action maintains surplus electrons at the zinc until the cell is discharged.

The popular 'dry' Leclanché cell has a zinc canister as negative electrode and central carbon rod as the positive electrode [10]. The rod is surrounded by powdered carbon and manganese dioxide, whose molecules can combine with hydrogen molecules to produce water and prevent hydrogen accumulating. The e.m.f. is about 1.5 volts.

Batteries

If the cell is large, there is more area for chemical action, and a more powerful current is available. If cells have the carbon rods and also the zinc cases joined [11], this is electrically like a larger cell. Voltage is that of a single cell.

If cells are joined positive to negative, the cells are in series [12]. The battery voltage is that of the cells added together—3 volts for two 1.5 volt cells, and so on. A layer battery [12] has flat units placed one upon another. Miniature mercury cells are made for small apparatus.

Secondary cells or lead acid cells are re-charged by passing direct current through them from a generator or other supply. A car battery usually has spongy lead and lead peroxide plates in dilute sulphuric acid, and each cell has an e.m.f. of 2 volts. Six cells form a 12 volt battery.

11. A battery can have cells in parallel.

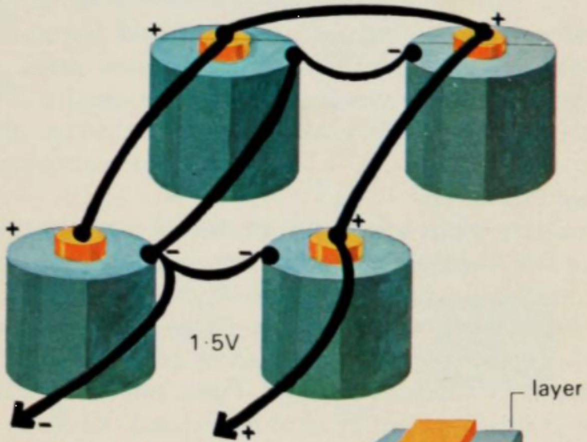

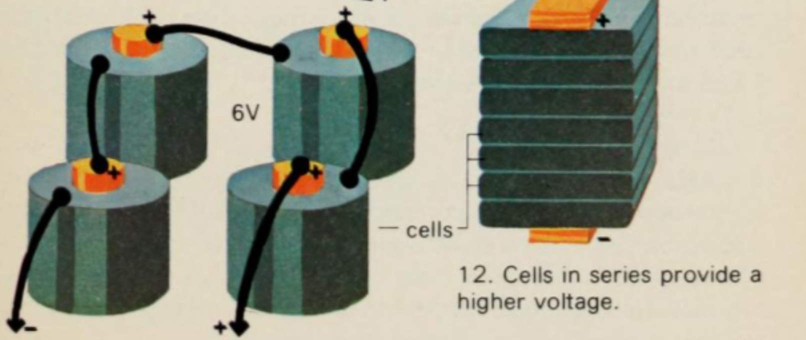

12. Cells in series provide a higher voltage.

13. Attraction of dissimilar charges.

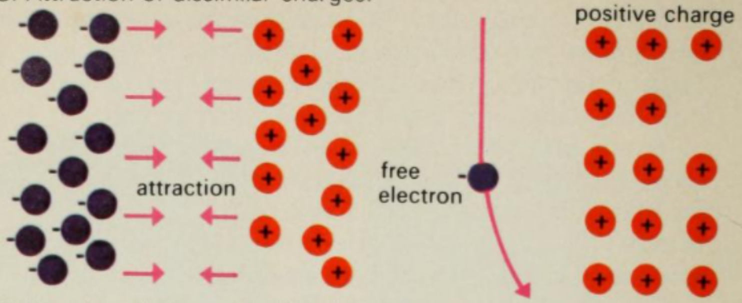
positive charge
attraction
free electron

14. Repulsion between similar charges.

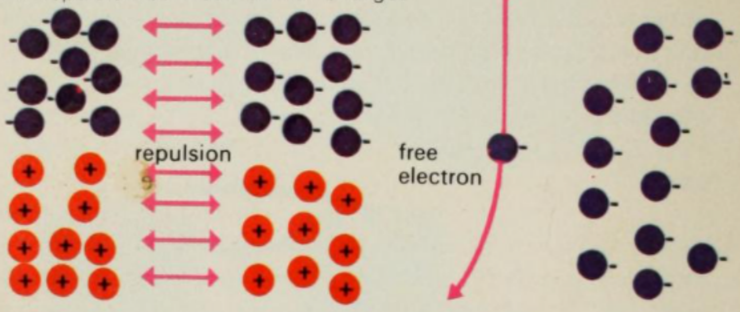
repulsion
free electron

Free Electrons

Negative electrons move toward a positive charge [13]. If a free electron is moving near a positive charge, its path is curved, because of this attraction.

Repulsion exists between charges of the same polarity [4]. The path of a free electron or stream of electrons curves away from a negative charge. These effects are used to direct the electron stream in various electronic devices, such as the cathode ray tube used in oscilloscopes. The path of a free electron is also changed by a magnetic field. Television tubes and magnetrons utilize this effect.

Electron Gun

A stream of electrons is obtained from an electron gun having a cathode heated by an internal filament [15]. The high temperature causes thermionic emission of electrons. The beam passes through a hole in a cylinder, its intensity controlled by potential applied. Making the cylinder negative repels electrons and reduces the beam.

15. Electron gun emits stream of electrons.

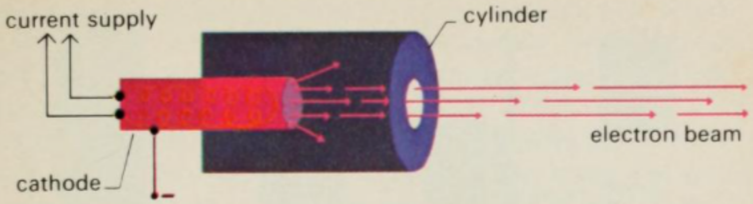
current supply
cylinder
electron beam
cathode
–

Focusing

A convex optical lens will focus light rays to a point. An electron lens has a similar effect on the cathode ray beam, the electric field curvature acting in much the same way as the glass lens on light rays [16]. Cylinders having different potentials have a similar effect.

By controlling the potential on one or more electrodes, the intensity of the field and its curvature can be adjusted to focus the stream of electrons into a narrow beam.

The cathode ray tube has an electron gun, grid or aperture to control beam intensity, and accelerator and focusing electrodes. The beam strikes a fluorescent coating, producing a spot of light visible on the tube. Stray electrons are absorbed by the aquadag coating.

16 Focusing the beam in a cathode ray tube

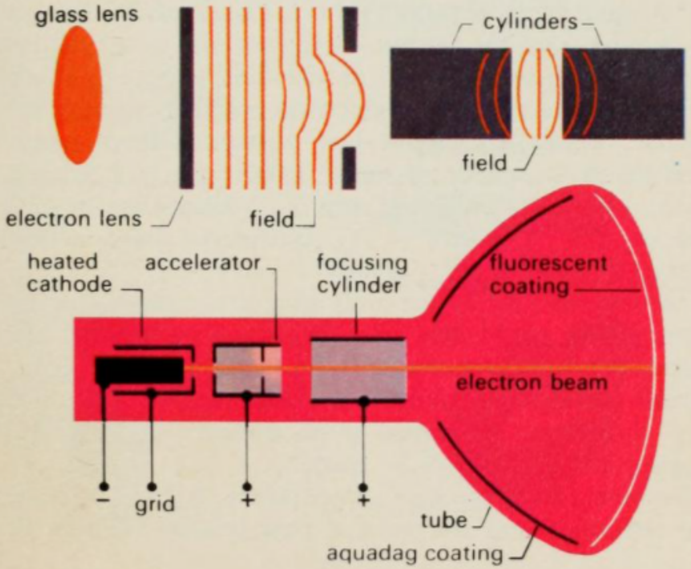
glass lens
cylinders
field
electron lens
field
heated cathode
accelerator
focusing cylinder
fluorescent coating
electron beam
–
grid
+
+
tube
aquadag coating

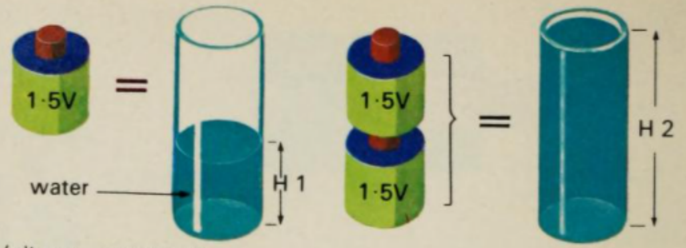
1·5V
water
H 1
1·5V
1·5V
H 2

17. Voltage = pressure

18. Current flowing is in amperes

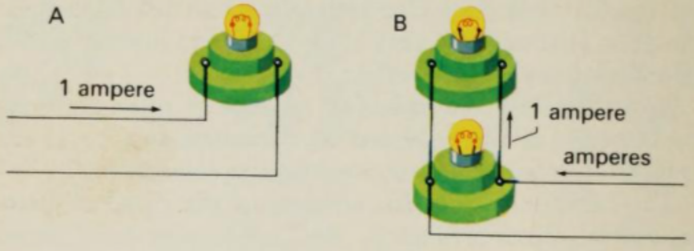
A
1 ampere
B
1 ampere
amperes

Circuit Laws and Units

A volt (V) is the unit of *pressure* from a battery or from some other source. A dry cell has a pressure of 1.5 volts. Two cells in series provide 1.5 + 1.5 = 3 volts. A vessel half full has a depth of water H1 [17]. The depth H2 is twice H1. A pipe from the bottom of the second vessel delivers water at twice the pressure of the first vessel. Similarly, raising the voltage increases electrical pressure.

Some circuits operate at low voltage. Others require hundreds or thousands of volts. 1,000 volts = 1 kilovolt, or 1kV. On the other hand, very low voltages are in millivolts or mV. 1,000mV = 1V. Extremely low voltages are in microvolts, or μV. 1,000μV = 1mV.

The ampere (A) is the unit of current flow. For example, a bulb could take 1A [18a]. Two such bulbs take 1 ampere each, or 2 amperes, or twice as much current [18b].

Small currents are given in milliamperes, or mA. 1,000 mA = 1 ampere. Extremely small currents are in microamperes, or μA. 1,000μA = 1mA.

The ohm (Ω) is the unit of *resistance*. A thick wire has little resistance and lets current pass readily [19]. In the same way a large pipe lets water flow easily. A thin wire resists the ready flow of current, just as the small pipe offers resistance to the flow of water. The greater the electrical resistance of a component or circuit, the more the passage of current is hindered.

Copper wires have little resistance, so they let current pass readily. But we may want resistance in many circuits. Iron and alloy wires can provide this. Making the wire thinner or increasing its length raises the resistance. Carbon compounds resist the ready flow of current, and can be formed into rods having any resistance needed. They are widely used. A component having resistance is a *resistor* [20]. A variable resistor has a sliding contact, bringing into circuit as much of the resistance element as is wanted. A potentiometer has a connection at each end and sliding contact.

For high values we use kΩ, k, or kilo-ohm. 1k = 1,000 ohms. Very high values are in MΩ (megohms). 1 MΩ = 1,000k. Resistors are usually color-coded. Each color shows a figure, or a number of zeros.

19. Ohm = unit of resistance

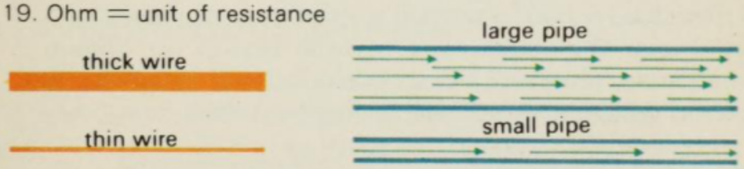
thick wire
thin wire
large pipe
small pipe

20. Resistors and color codes

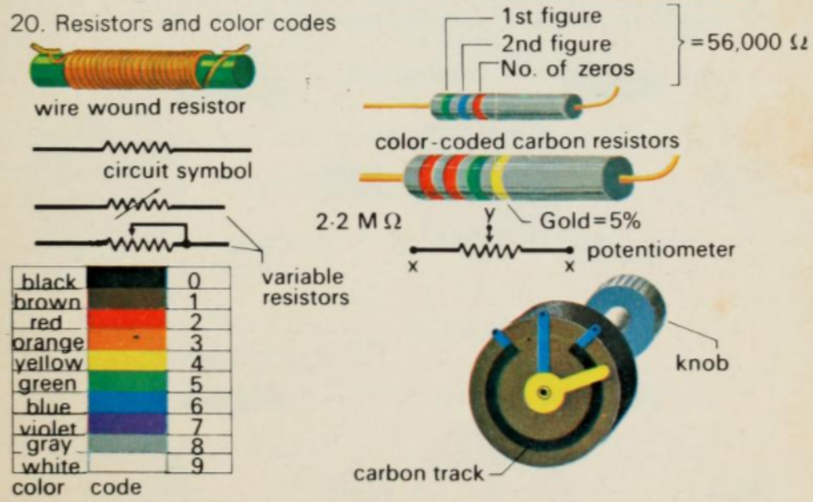
wire wound resistor
circuit symbol
variable resistors
1st figure
2nd figure
No. of zeros
=56,000 Ω
color-coded carbon resistors
2·2 MΩ
Gold=5%
x
x
potentiometer
knob
carbon track
color code

black	⬛	0
brown	🟫	1
red	🟥	2
orange	🟧	3
yellow	🟨	4
green	🟩	5
blue	🟦	6
violet	🟪	7
gray	▨	8
white	⬜	9

Ohm's Law

If we increase the voltage applied to a circuit, the electron flow becomes stronger [21]. Reducing the voltage makes the current weaker.

If we leave the voltage the same, increasing the resistance will make the current weaker [22]. Reducing the resistance naturally allows more current to flow.

These facts are the basis of Ohm's Law. We say:

Current in Amperes × *Resistance in Ohms* = e.m.f. As I is used for current, and R for ohms, this is $I \times R = E$, or $IR = E$. In the same way, $\frac{E}{I} = R$ and $\frac{E}{R} = I$.

Always remember that the basic units are amperes, volts and ohms. As an example, 1mA flows through 100k. What voltage is dropped? 1mA = 0.001 ampere, and 100k = 100,000 ohms. So 0.001 × 100,000 = 100 volts.

The *power* that is usefully employed or perhaps wasted or dissipated as heat in a circuit is given in watts. Useful power could raise the temperature of tube heaters, or operate a loudspeaker or other device. Wasted power might be that lost in a resistor which is used to drop the voltage.

21. Current increases with voltage.

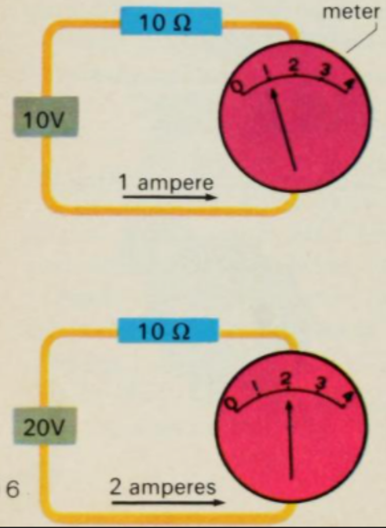
10 Ω
meter
0 1 2 3 4
10V
1 ampere
10 Ω
0 1 2 3 4
20V
2 amperes

22. Increasing resistance reduces current.

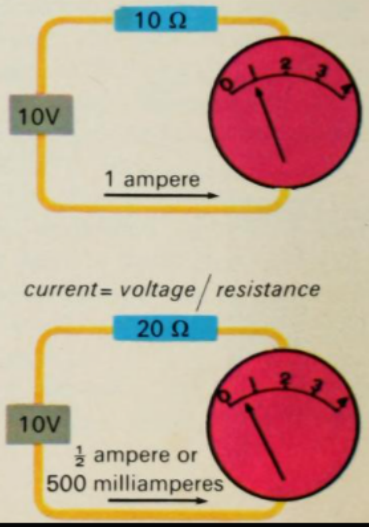
10 Ω
0 1 2 3 4
10V
1 ampere
current= voltage / resistance
20 Ω
0 1 2 3 4
10V
½ ampere or
500 milliamperes

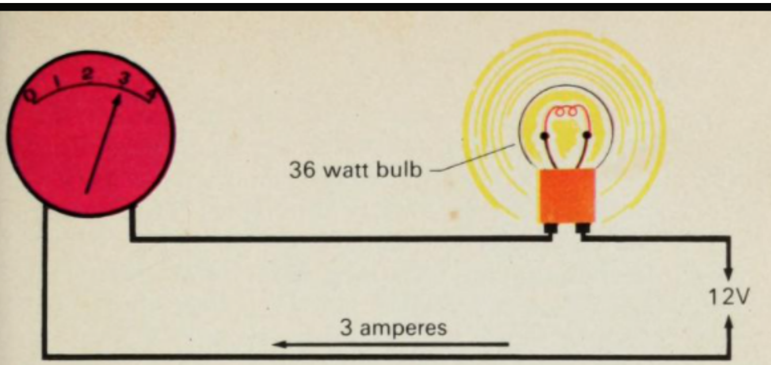
0 1 2 3 4
36 watt bulb
12V
3 amperes

23. Electrical power in a circuit (Power = watts = voltage × current).

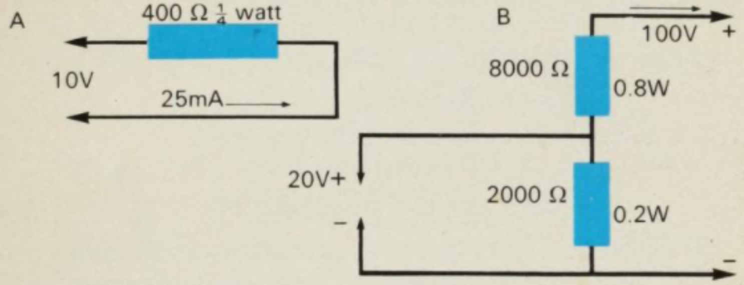
A
400 Ω ¼ watt
10V
25mA
B
100V
+
8000 Ω
0.8W
20V+
−
2000 Ω
0.2W
−

24. Voltage, resistance, current and wattage.

We could give the useful output of an amplifier or other device in watts. Small powers are given in milliwatts. 1,000 mW = 1 watt. Power in watts = voltage × current.

Voltage in Volts × *Current in Amperes* = *Power in Watts*

Or we can say $\frac{\textit{Watts}}{\textit{Volts}} = \textit{Amperes}$ or $\frac{\textit{Watts}}{\textit{Amperes}} = \textit{Volts}$

A 36 watt car bulb is run from a 12 volt battery [23]. So it draws $\frac{36}{12} = 3$ amperes current.

Imagine that a potential of 10 volts is present across a 400 ohm resistor [24a].

From $I = \frac{\textit{Volts}}{\textit{Resistance}} = \frac{10}{400} = 0.025$ ampere.

With resistors in series, 8,000 ohms + 2,000 ohms = 10,000 ohms [24b]. From $I = \frac{E}{R}$ or $\frac{100}{10{,}000}$ we find that 10mA (0.01A) flows.

The Inductor

The inductor is a wire wound as a coil. When current is applied, a magnetic field swells around the coil. When the coil is disconnected, a spark appears between the disconnected lead and the source, and the magnetic field collapses. These events indicate that energy had been stored in the coil field.

Both the building and collapsing fields resist changes in current flow (show *reactance*). With a current alternately changing back and forth, or *alternating current* (AC), the inductor assumes an inductive reactance that acts much like a resistance [25a]. The more turns on the coil, the more its inductance and the more its reactance [25b].

The formula $X_L = 2\pi fL$ gives reactance where:

$X_L =$ Reactance in ohms
$2\pi =$ Constant for one cycle
$f =$ Frequency in Hertz (Hz)
$L =$ Inductance in henrys

In Figure 26A the supply has a frequency of 50 cycles per second, or 50 Hz. It changes direction 50 times a second. At [26b] the frequency is 100 Hz. The reactance of the inductor B is twice that of A (Hertz (Hz) can be written as cps).

Millihenrys (mH) and microhenrys (μH) are smaller values of henry (H) for inductance.

25. Reactance of coil increases with increased inductance.

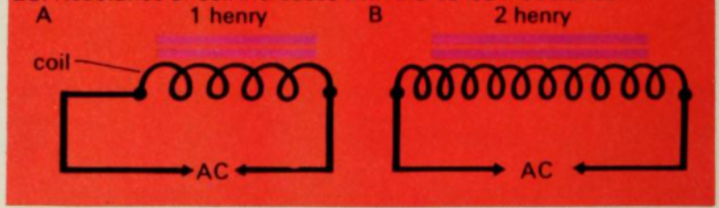

26. Reactance of coil increases with frequency.

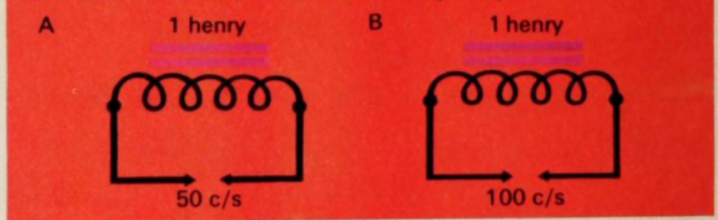

Reactance of Capacitor

A capacitor has two closely spaced plates insulated from each other. If we connect an e.m.f., there will be only a brief surge of current as the capacitor charges to the applied voltage. If we remove the e.m.f. and place a wire across the capacitor, current will surge briefly as it loses its charge. Current flows only with a changing e.m.f. (see page 28).

In Figure 27a, AC is connected to the capacitor and a certain current flows. If we make the capacitor larger (27b), the current is larger too, because the reactance is lower. Reactance falls with an increase in capacity or frequency. In Figure 28a we apply 50 c/s AC to the capacitor. At (28b), we apply 100 c/s. The reactance of the capacitor falls with increased frequency.

$$X_C = \frac{1}{2\pi fC}$$

Where:

$X_c =$ Capacitive reactance in ohms
$2\pi =$ Angular constant
$f =$ Frequency in cycles per second
$C =$ Capacity in farads

Does it Matter?

Figure 29 shows a transformer (see page 40). Primary P takes audio signals from amplifier; secondary S supplies loudspeaker. Capacitor C is across the primary. Higher frequen-

27. Reactance of capacitor falls with increased capacity.

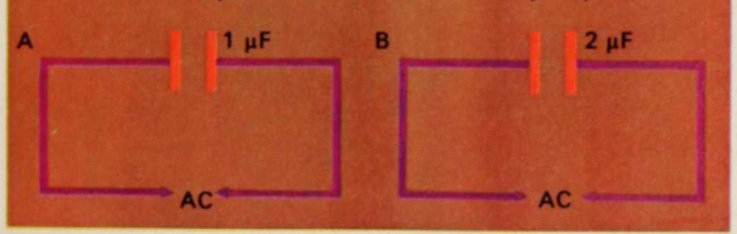

28. Reactance of capacitor falls with increased frequency.

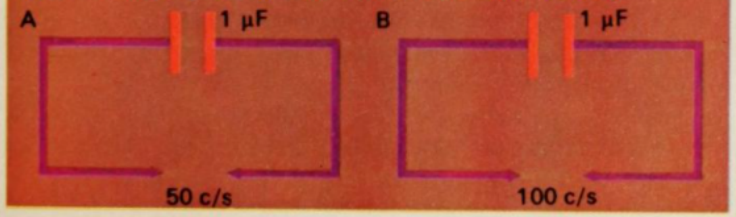

cies more readily pass through C and do not flow through the primary. This acts as a treble tone control.

In Figure 30 capacitor C1 is from an audio amplifier tube plate, and by-passes or loses higher frequencies to ground.

The effect is changed by placing part of the variable resistor VR1 in circuit. VR1 is a variable tone control.

29. Falling reactance of capacitor 'cuts' high frequencies.

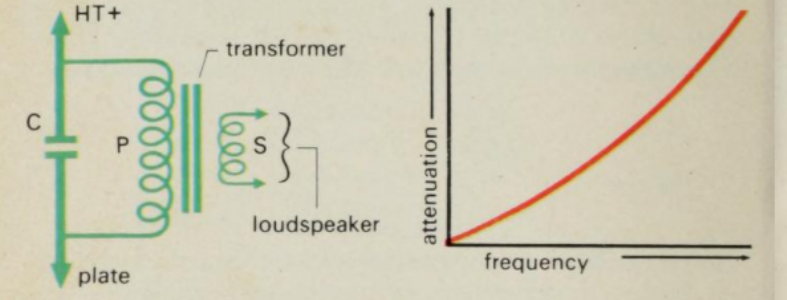
HT+
transformer
C
P
S
loudspeaker
plate
attenuation
frequency

30. Tone control circuits.

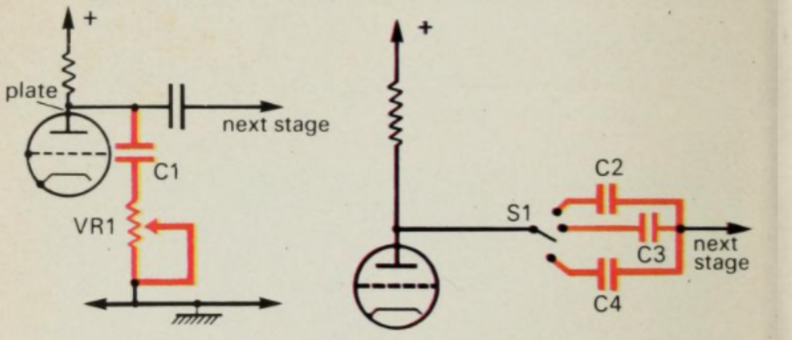
+
plate
next stage
C1
VR1
+
C2
S1
C3
next stage
C4

The switch S1 selects a small capacitor C2, a medium value capacitor C3, or a large capacitor C4. C2 lets high frequencies pass easily. C3 lets some lower frequencies pass. C4 allows low frequencies to pass readily. Tone controls are often found in receivers and amplifiers. Most make use of the fact that capacitive reactance falls with increased frequency.

Inductive reactance also has many uses. In Figure 31 a rectifier supplies a power filter. Output from the rectifier

31. Use for inductive reactance.

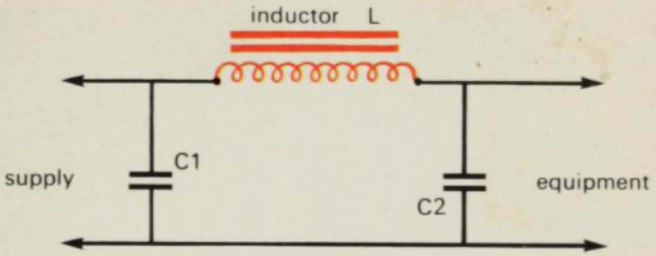
inductor L
supply
C1
C2
equipment

has pronouned ripple. L is a choke with high inductance. It hinders fluctuations in current. Capacitor C1 acts as a reservoir. The inductor L refuses to let ripples pass, and capacitor C2 provides almost pure direct current. This is called 'smoothing' (see page 45).

In Figure 32a radio frequency (RF) signals and audio frequency (AF) signals arise at collector C. The radio frequency choke hinders RF, so this goes through the small capacitor C1. The choke reactance is low at low frequency, so AF passes through it and is taken off by the larger capacitor C2.

At [32b], a high frequency noise filter choke is placed in an audio circuit. It does not let high frequencies pass, and so reduces surface noise when playing records.

32. Blocking the movement of high frequencies.

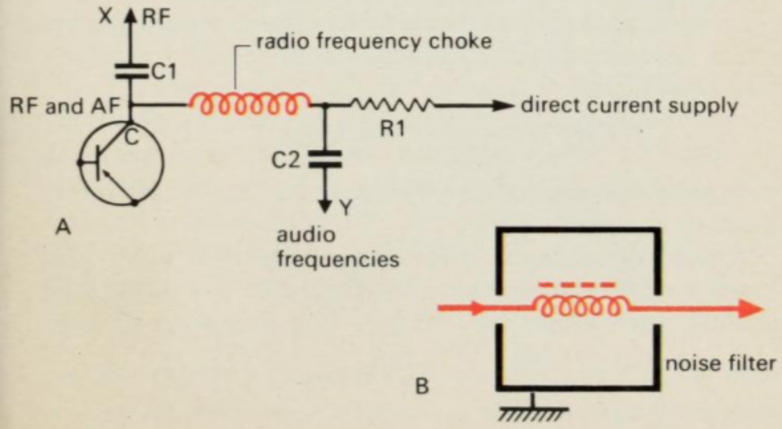
X
RF
C1
radio frequency choke
RF and AF
C
R1
direct current supply
C2
Y
A
audio frequencies
B
noise filter

33. Supplying a screen grid

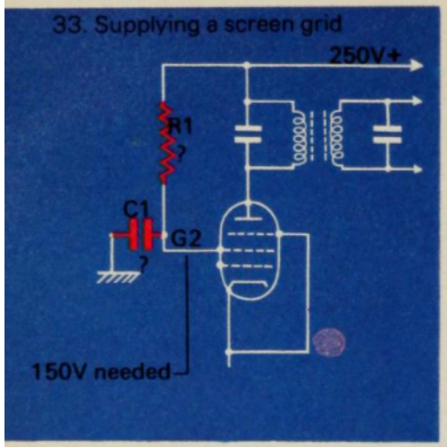

34. A problem in cathode bias

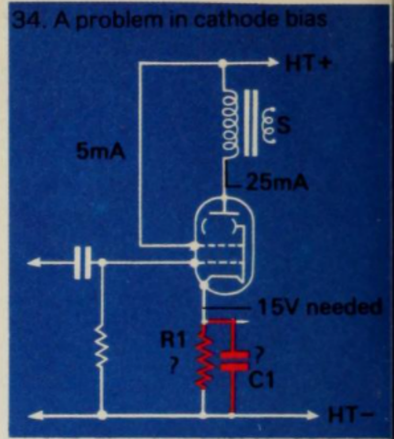

A pentode as amplifier is shown in Figure 33. The supply is 250 volts, but the screen grid G2 should have only 150 volts. Since 250–150 = 100, resistor R1 should drop 100 volts.

The current flowing through R1 is measured or found from a tube manual. It is 5mA. Recalling

that $\frac{E}{I} = R$, we can say that $\frac{100}{0.005} = 20{,}000$ ohms for R1.

What power is dissipated in R1?

$Watts = E \times I = 100 \times 0.005 = 0.5$ watt. So we use a 20K ½-watt resistor. G2 needs to be grounded at operating frequency. So C1 is a 0.05 μF 250 volt capacitor.

At the output stage of the receiver, 15 volts should be present at the cathode [34]. Cathode current is 25 + 5mA = 30mA or 0.03A.

$$\text{So R1 is } \frac{15}{0.03} = 500 \text{ ohms.}$$

For audio frequencies and 15 volts, C1 can be 25 μF 25 working volts.

A test instrument for two voltage ranges—0 to 10 volts and 0 to 100 volts—is shown in Figure 35. For a full-scale reading, 1mA must flow (.001A). This is to show 10 volts with R1. R1 = $\frac{10}{0.001}$ = 10,000 ohms. Similarly, R2 = 100,000 ohms.

35. Resistor values for a voltmeter

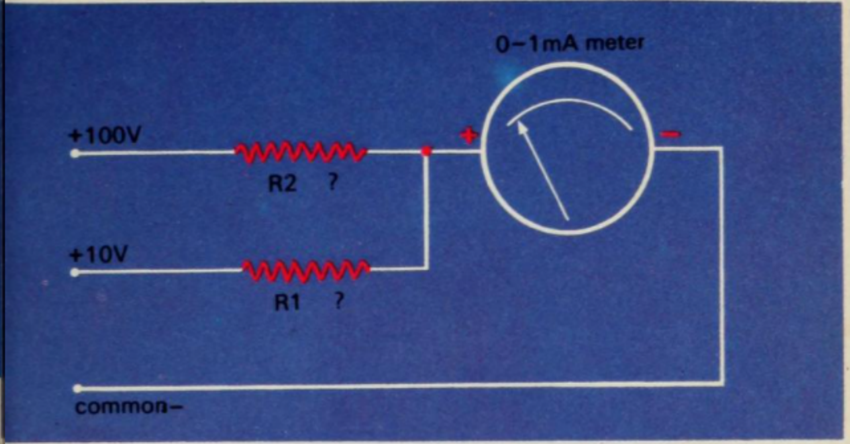

36. DC working conditions of a transistor amplifier.

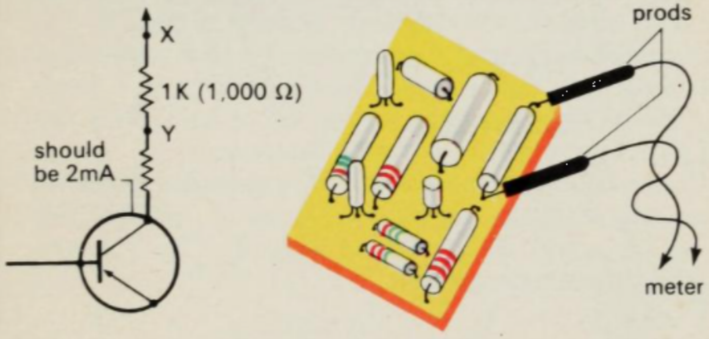

A transistor circuit where we suspect the transistor is not drawing the correct collector current, though all resistors and the supply voltage are correct, is shown in Figure 36. The collector current should be 2mA. The simplest test is to measure the voltage across the resistor, from X to Y. From $E = I \times R$, the voltage should be $0.002 \times 1{,}000 = 2$ volts.

A meter reads only 0.5 volt.

$$\text{From } I = \frac{E}{R} = \frac{0.5}{1{,}000}$$

we see that the transistor collector current is actually 0.5mA.

Electronic Circuit Diagrams

For some electronic equipment, components could be shown in outline, with connections, in a scale drawing just as in the apparatus.

But it is usually better to have a circuit diagram where simple symbols represent various components. Symbols are often simplified representations of the components. By comparing components and their symbols, you will become familiar with them.

Lines represent wires. These may be joined as in Figure 37a. They may cross without contact [37b]. Dots [37c] show connecting points such as sockets, which could be for plugs connected to a loudspeaker, or could be used for some other purpose.

A fixed resistor symbol is shown in Figure 38, together with three resistors in a small part of a circuit. In Figures 39a and 39b is shown a variable resistor. If connections are taken to X and Y, moving the sliding contact (arrow) to the right increases the resistance actually in circuit.

The resistance element may be wire for low values, but carbon is used for high values. The track can be in a semicircle, and rotating a spindle moves the contact along the track [39c]. Figure 39d is a typical component.

B, C and D are potentiometers. If a potential—such as an audio signal voltage—is taken to X and Z, any required level can be tapped off by moving slider Y. This happens with an audio volume control.

Switches

In Figure 40 is shown an on/off switch. It could be a supply voltage toggle switch, low voltage rotary switch, or any other component.

A volume control often has a switch. It can be a single on/off type (single pole), or it can be two on/off switches (double pole) which work together to interrupt two circuits.

The rotary switch in Figure 40 has a contact X, which can transfer a circuit to 1, 2 or 3, the switch having three positions. It is '1-pole 3-way'. It could have more contacts and more ways, as required.

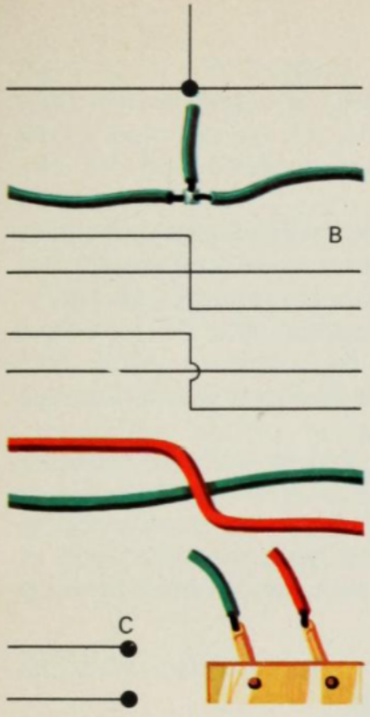

37. How connections are shown.

38. Resistors

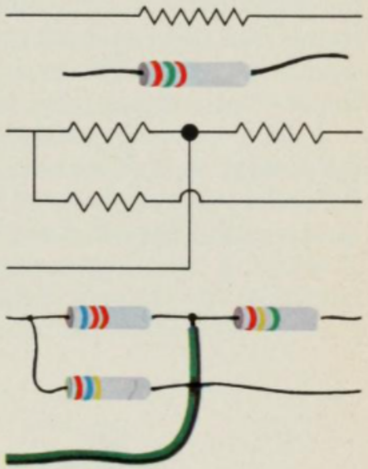

39. Variable resistors and potentiometers

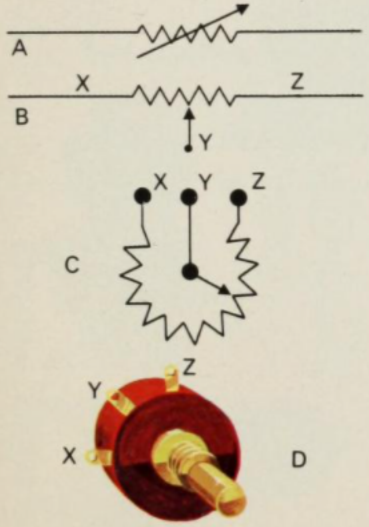

40. How switches are shown

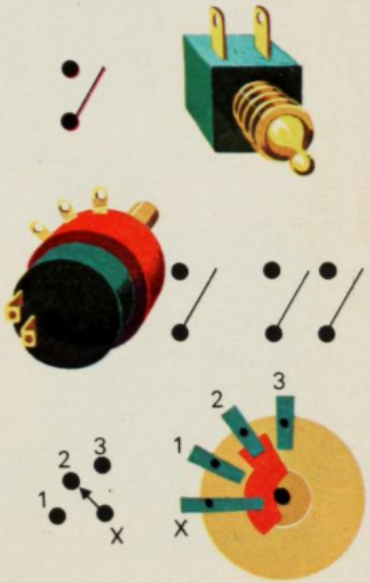

Inductors and Coils

Insulated wire wound upon an insulated tube forms a coil or inductor [41]. We can increase the inductance by using more turns, or having a larger coil. Or we can place a core in the winding to concentrate the magnetic field and raise its inductance.

The inductor might be of very small inductance and have few turns [42a]. Larger inductance coils have more turns [42b]. The coil could have a tapped winding—or two or more windings on the same insulated tube [42c]. If there are very many turns they may be in sections [42d]. This is a 'high frequency choke'—an inductor which prevents the passage of high frequencies.

Special materials have to be used for cores for radio frequency coils (those that work at radio frequencies). A core of ferrite or iron dust material placed in the coil winding intensifies the magnetic field and raises the inductance, so if we want to adjust the inductance, we can use a threaded core, and screw it in or out [43]. In the symbol, the dotted line is the core, and the sloping line shows it is adjustable. To simplify large circuit diagrams, cores are often not shown.

41. Radio frequency (RF) chokes

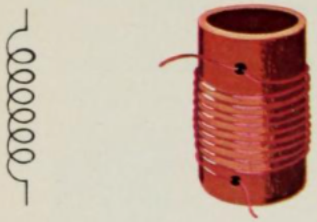

42. Other small inductors

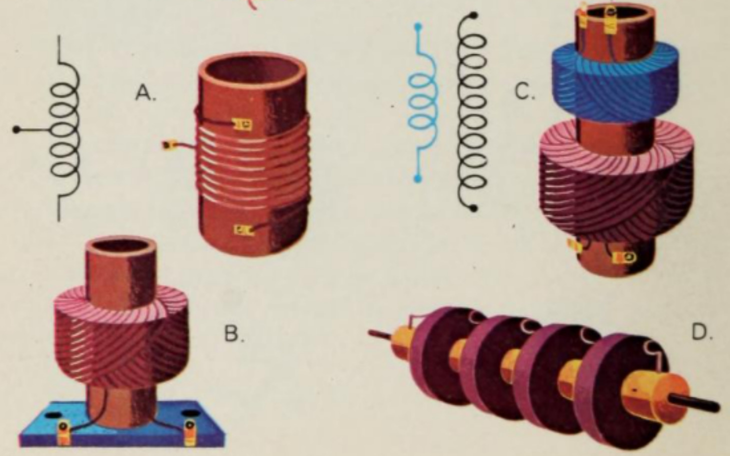

LF Cores

At low frequencies, cores can be made from various alloys [44]. Thin sheet iron was once used.

The winding is often on a bobbin, entirely surrounded by the core. One limb fits inside the winding. This core is often of thin metal 'stampings' or laminations, and enough are used to make up the thickness wanted.

Since this inductor prevents the ready flow of low frequencies it is a 'low frequency choke'. Such a choke is often employed to help remove ripple in a supply from a power circuit, and is then called a 'smoothing choke'. In order to be used for this purpose, it would normally have an inductance of several henrys.

43. Adjustable core allows inductance to be changed.

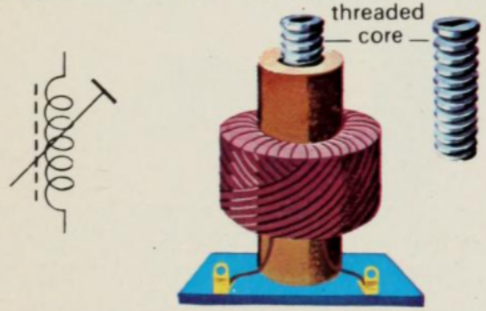

44. Audio chokes with core made from laminations.

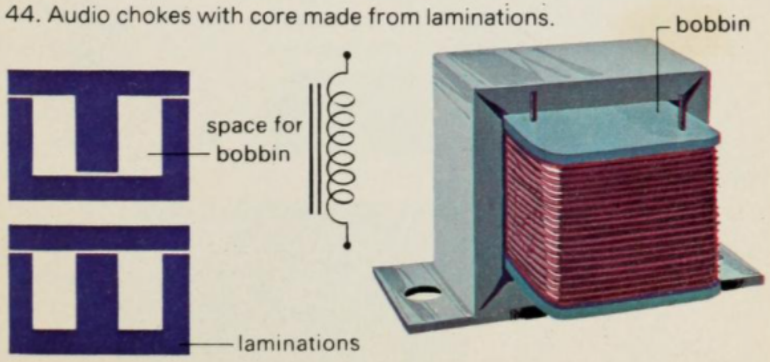

Capacitors

Two metal plates near each other but not in contact form a capacitor and can store a charge [45]. If we connect a voltage to X-X a deficiency of electrons arises at one plate, and a surplus at the other. Electrostatic stress arises in the insulating space between plates. The capacitor is 'charged' with a power depending on 'capacity' or charge-storing ability. Unit of capacitance = farad (F).

If X-X are joined together, the capacitor gives up its charge. We can increase the capacity to hold a charge by placing the plates nearer each other, and in other ways. The insulation between plates is the 'dielectric'. Capacitors are made with mica and ceramic insulation, or foil rolled up with insulating paper and in a tube [46]. Or metallic conductors may be deposited on an insulator.

'Electrolytic' capacitors are so made that an applied voltage must be of the correct polarity. These capacitors are of some set value and are called 'fixed capacitors'.

45. A fixed capacitor

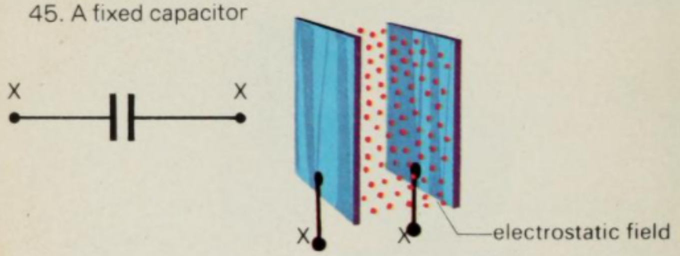

46. Some capacitors for electronic equipment

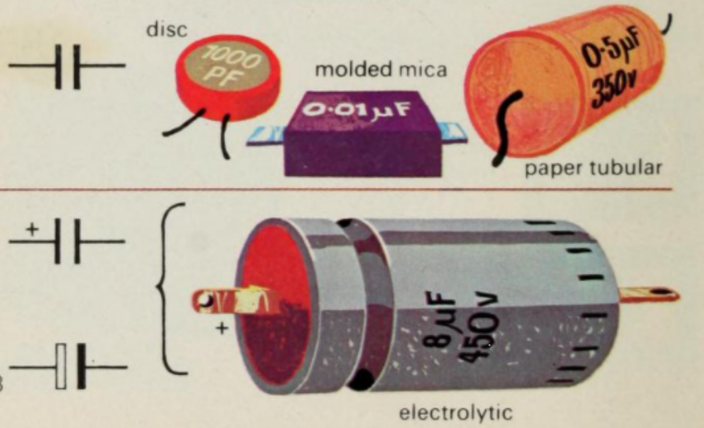

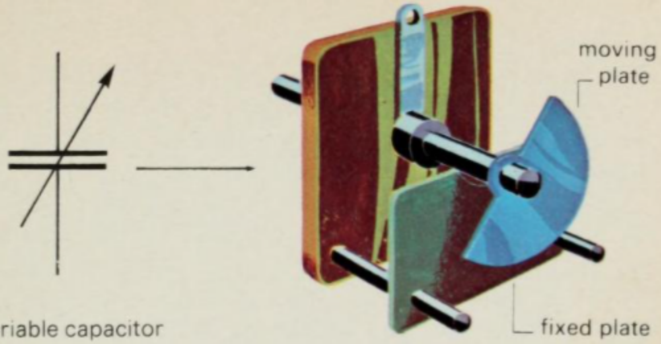

47. Variable capacitor

48. 2-gang and trimmer capacitors

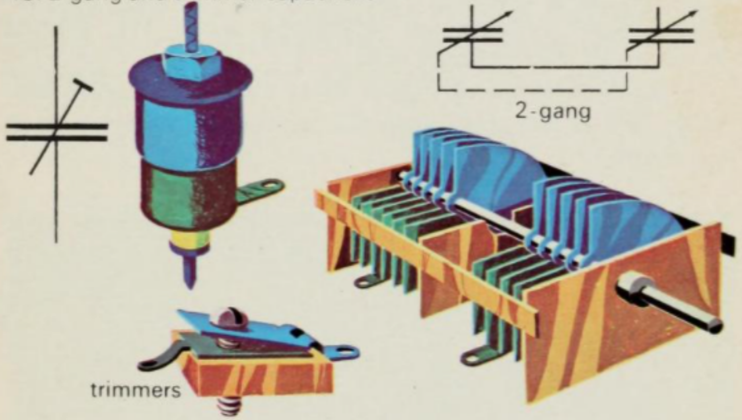

We may need to change the capacitance at any time and for this we want a 'variable capacitor'. A small variable capacitor can have one fixed plate, and one moving plate attached to a spindle [47]. Rotating the spindle changes the extent to which moving and fixed plates overlap each other, and thus alters the capacity. Capacitance is least with the plates not overlapping at all.

We may wish to tune two circuits simultaneously in a radio receiver and can then use a 2-gang capacitor [48]. By having many fixed and many moving plates, the total capacity of each section can be increased. Sometimes the capacity has to be adjusted when first putting a circuit into proper working condition. A 'trimmer' is then used. It is adjusted with a screwdriver or other tool, and brings moving and fixed plates closer to increase the capacitance.

A Voltage Amplifier

The circuit in Figure 49 is the first audio amplifier stage of a receiver or large audio amplifier. It increases the voltage of audio signals obtained from a pick-up employed to play records, or from a receiver detector.

Capacitor C1 isolates earlier direct current circuits from the base B of the transistor, supplied by the resistors R1 and R2. Resistor R3 applies a small negative bias to the emitter E. Bias increases when emitter current rises. This helps to make working conditions stable. C2 grounds the emitter E at audio frequencies.

Current from collector C flows through resistor R4. Amplified signal voltages are developed across R4. Signals are taken off by capacitor C3.

A practical layout of Figure 49 is shown in Figure 50a. Figure 50b is the same amplifier assembled in more compact form.

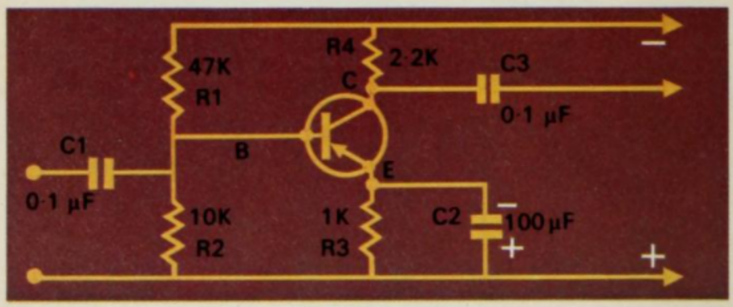

49. One-stage transistor amplifier

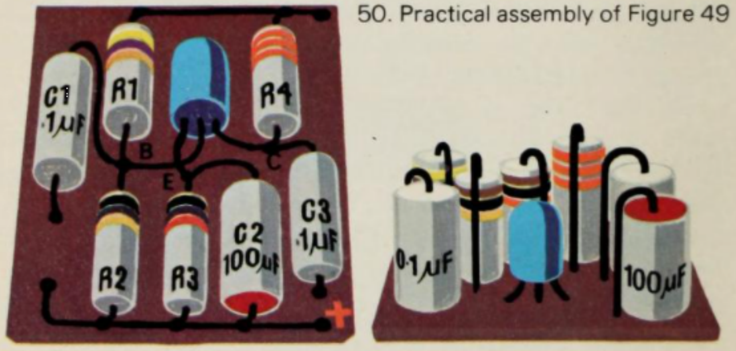

50. Practical assembly of Figure 49

A.

B.

Block Diagrams

We may show the essential parts of a circuit without going into detail. Block diagrams do this. Figure 51 is a block diagram of a large amplifier. Each 'box' represents one section—Section 1 is an audio frequency amplifier like Figure 49. So is Section 2. Section 3 is an output stage, or amplifier which supplies appreciable power. Signals reach it from Section 2, and its output goes to the loudspeaker.

Figure 52a is for playing stereo records. The stereo pick-up produces left and right channel signals. These go to the balancing section, where the relative volume is adjusted. Amplifiers supply outputs for right and left speakers.

Figure 52b is a block diagram showing a photoelectric counter. An item interrupts light falling on the cell, causing electrical output to work a counter.

51. Block diagram of audio amplifier

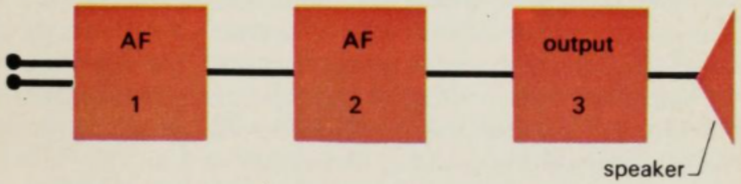

52. Block diagrams of stereo amplifier and photo-electric counter

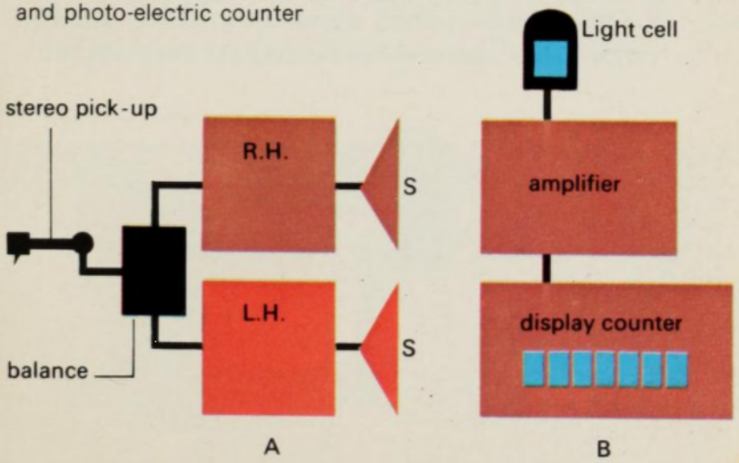

A

B

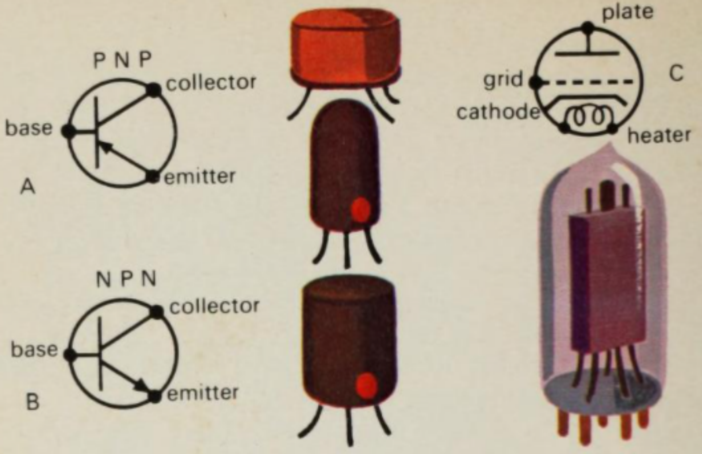
PNP
collector
base
emitter
A
NPN
collector
base
emitter
B
plate
grid
C
cathode
heater

53. Transistor symbols and vacuum triode

Other Symbols

The symbol for a PNP transistor is shown in Figure 53a; that for the NPN transistor at 53b. Figure 53c is the symbol for a triode tube. The tube might be of any type.

Remember, most symbols are simplified representations of the component or its function! The relay [54] illustrates this. Tags X-X take current to an electromagnet winding. The spring-actuated contact Y normally rests against the fixed contact 1. When current flows in the winding, the armature Y moves and switches the circuit to 2.

Sometimes items in a circuit can be omitted, to obtain a diagram which is easier to follow, or to ignore repeated details. In Figure 55 the switch allows inductor L1, L2, L3 or L4 to be selected. This is shown in full at [55a]. At [55b] the actual way in which inductor L1 is connected is shown and inductors L2, L3, and L4 are omitted.

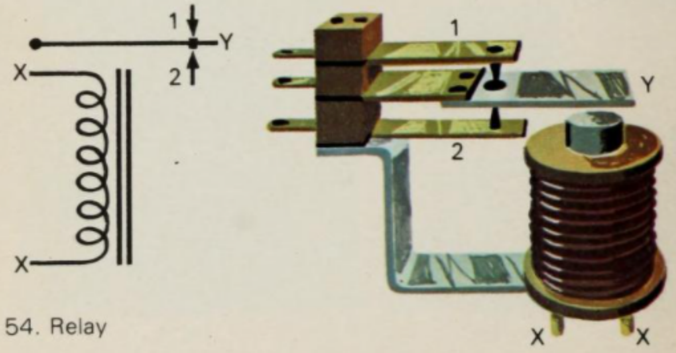
1
Y
X
2
X
1
Y
2
X
X

54. Relay

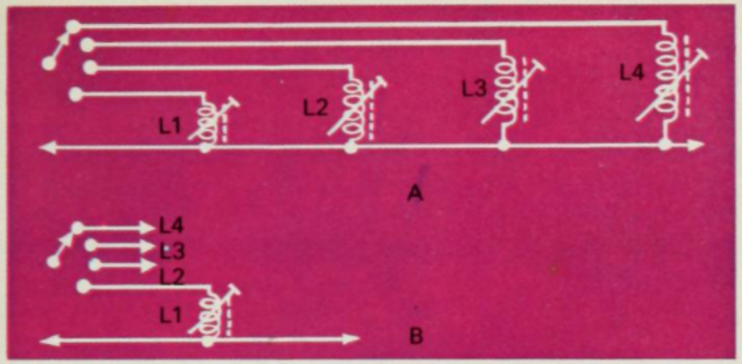
L1
L2
L3
L4
A
L4
L3
L2
L1
B

55. Omissions for clarity

Simplifications of this kind are very useful when there are several switch positions and sections, and possibly quite a large number of inductors or other parts.

Theoretical circuits do not usually take into account the actual way in which the equipment is assembled. For example, an amplifier circuit could be assembled on a panel and occupy quite a large space. Or it could be arranged on a smaller panel, with components close together, or made with miniature components to the smallest possible size. But in all these cases it might be the same circuit.

Various other symbols are shown in Figure 56.

56. Various other symbols

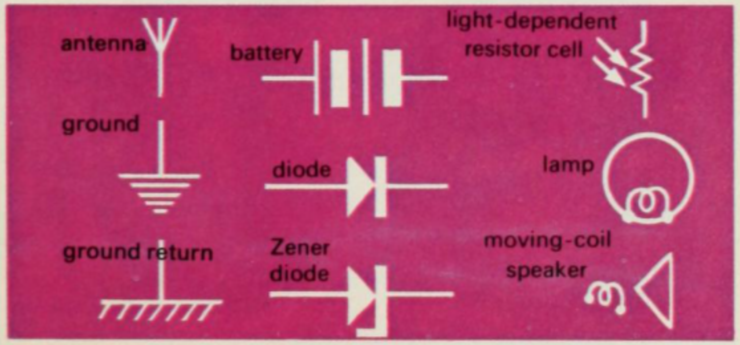
antenna
battery
light-dependent resistor cell
ground
diode
lamp
ground return
Zener diode
moving-coil speaker

Magnetism is Important

An unmagnetized bar has all its particles at random [57a]. Magnetizing it draws them into similar polarity [57b]. The bar magnet [57c] has a north pole at one end, and south pole at the other end. Other shapes are the horses shoe [57d] and ring magnets [57e].

If the north pole of one magnet is near the south pole of another magnet, an attractive force arises. But if a north pole is placed by a north pole, or a south pole by a south pole, a repulsive force pushes the magnets apart. Both poles attract ferrous or iron objects.

Various specially shaped magnet assemblies are used [58]. That at Figure 58a has curved ends to accommodate the rotating coil of a meter; [58b] is for a similar purpose. Figure 58c is for a moving-coil speaker or microphone.

If we connect a coil to a battery or other direct current supply, a magnetic field is created [59a]. It has north and south poles, and exists so long as current flows. The magnetic field arises when the switch is closed and current begins [59b]. When the switch is opened, the field collapses [59c]. This generates an electromotive force or e.m.f. in the coil, of opposite polarity to that from the battery. If the supply polarity is reversed, the magnetic poles are also reversed. If we connect the coil to an alternating current supply [60] current flows back and forth through the coil. At each reversal of direction, the magnetic field collapses and arises with opposite polarity. Most electromagnets have cores allowing concentration of magnetic force. Current in a winding creates a magnetic field.

The reverse arises. If a magnet is moved in or out of a coil, an e.m.f. is produced [61]. Relative motion must exist between the field and conductor. The magnetic field can be from a permanent magnet, or from another electromagnet. These effects are used in generators and other devices where the rapid motion of magnetic fields produces strong currents.

57. Permanent magnets

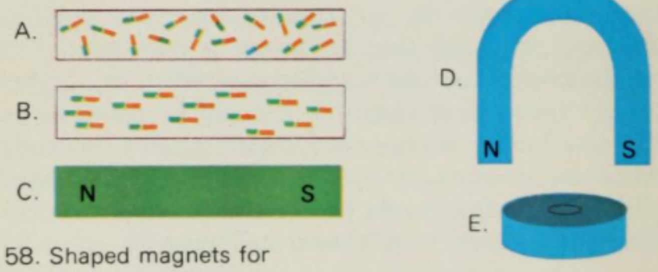

58. Shaped magnets for electronic apparatus

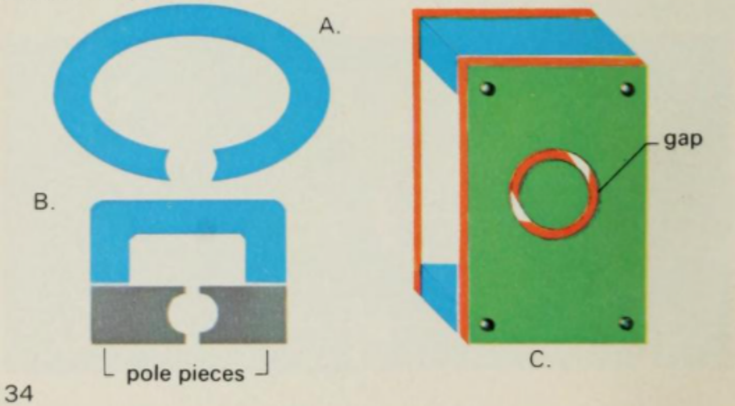

59. Field surrounding an inductor

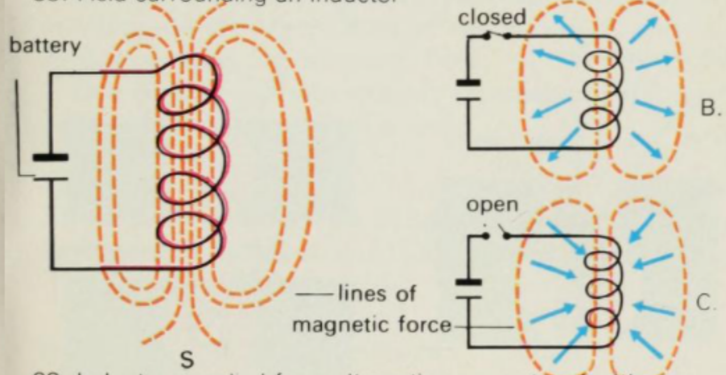

60. Inductor supplied from alternating current and with core

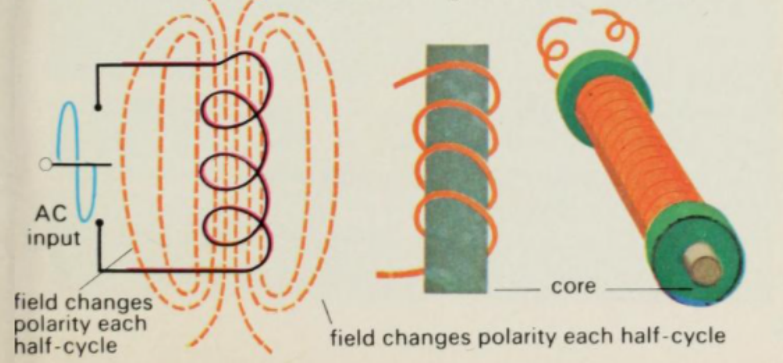

Moving-Coil Meter—The D'Arsonval Movement

In Figure 62, a strong field exists in the circular space between the shaped pole pieces. A coil is pivoted to rotate in this gap. Current is fed to it by hairsprings, and a pointer is attached.

When current flows in the coil, the magnetic field surrounding the conductors interacts with the permanent magnet field, and the coil rotates slightly. This can show current, voltage, or other things we wish to know. When current ceases, the coil returns to its original position.

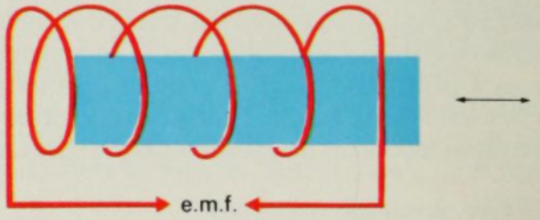

61. Changing magnetic field generates e.m.f.

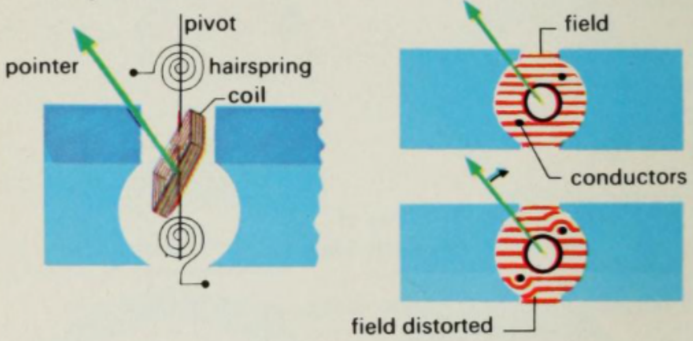

62. Moving-coil meter

Moving-Coil Speaker

This embodies a cone carrying a very light coil, which receives current by flexible leads [63]. The coil occupies a narrow gap in a magnet assembly, and can move in and out in line with the cone axis.

Audio signals taken to the coil generate a fluctuating magnetic field. Interaction between this field and the fixed magnet field causes the coil and cone to move in and out. This is transmitted to the air, producing sound waves.

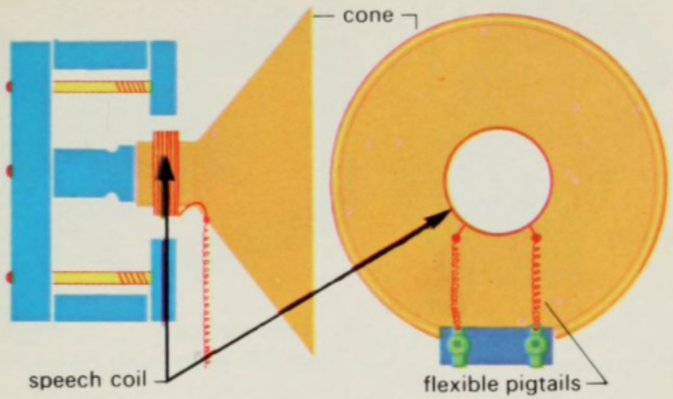

63. Moving-coil loudspeaker

The winding is often called the 'speech coil'. 'Permanent magnet' moving-coil speakers have a permanent magnet assembly. 'Energized' speakers have an electromagnet, or winding, on the center limb of the assembly, and current in this produces the magnetic field.

In a moving-coil microphone, vibrations of the cone and coil produce an e.m.f. in the coil.

'OR' Gate

Small cores are used for various purposes in computers. In Figure 64 an input at A OR at B will give an output. This is called 'OR' gate.

64. Magnetic 'OR' gate for computer

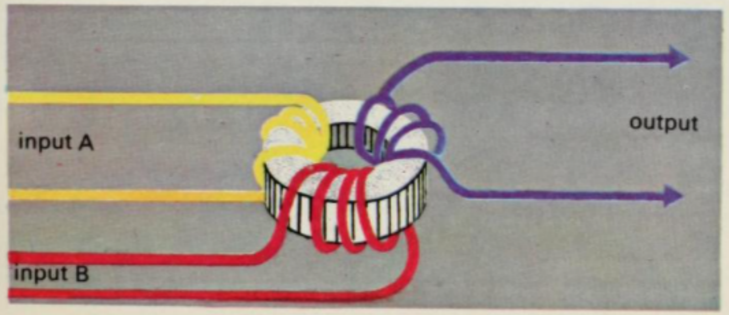

Moving-Coil Meter—The D'Arsonval Movement

In Figure 62, a strong field exists in the circular space between the shaped pole pieces. A coil is pivoted to rotate in this gap. Current is fed to it by hairsprings, and a pointer is attached.

When current flows in the coil, the magnetic field surrounding the conductors interacts with the permanent magnet field, and the coil rotates slightly. This can show current, voltage, or other things we wish to know. When current ceases, the coil returns to its original position.

61. Changing magnetic field generates e.m.f.

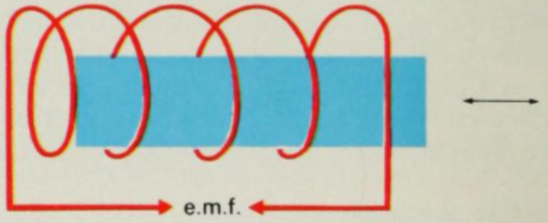

62. Moving-coil meter

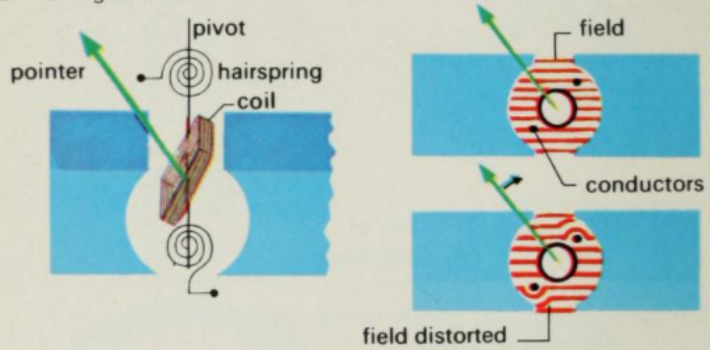

Moving-Coil Speaker

This embodies a cone carrying a very light coil, which receives current by flexible leads [63]. The coil occupies a narrow gap in a magnet assembly, and can move in and out in line with the cone axis.

Audio signals taken to the coil generate a fluctuating magnetic field. Interaction between this field and the fixed

63. Moving-coil loudspeaker

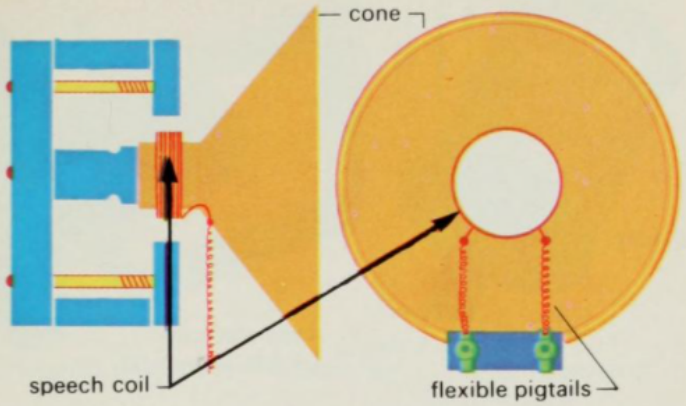

magnet field causes the coil and cone to move in and out. This is transmitted to the air, producing sound waves.

The winding is often called the 'speech coil'. 'Permanent magnet' moving-coil speakers have a permanent magnet assembly. 'Energized' speakers have an electromagnet, or winding, on the center limb of the assembly, and current in this produces the magnetic field.

In a moving-coil microphone, vibrations of the cone and coil produce an e.m.f. in the coil.

'OR' Gate

Small cores are used for various purposes in computers. In Figure 64 an input at A OR at B will give an output. This is called 'OR' gate.

64. Magnetic 'OR' gate for computer

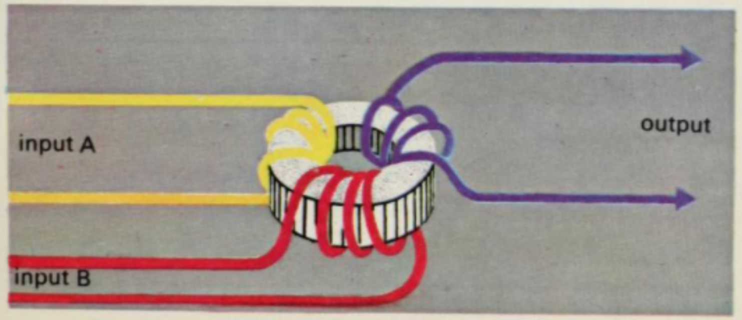

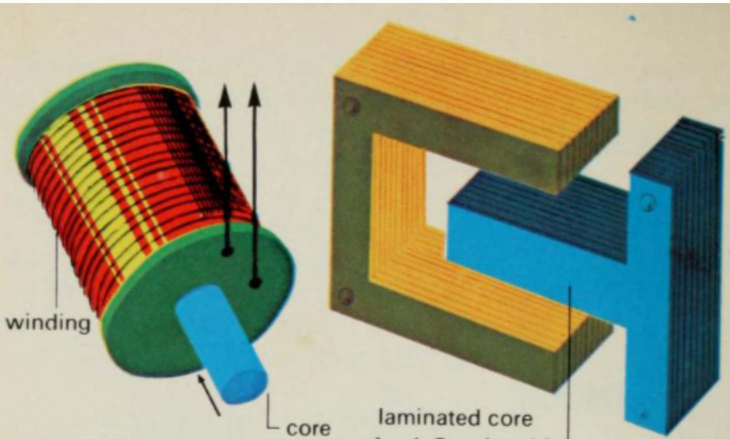

65. Solenoids

Solenoid and Magnetic Field

When current flows in a winding, a core becomes magnetized. If the core is out of the winding, magnetic forces pull it in [65]. The limbs of the plunger and the fixed portion of the assembly can be shaped to assist this action by concentrating the magnetic field. A solenoid may control fuel or liquid flow valves, mechanical step-by-step counters, and many other devices.

Sometimes we wish to avoid magnetic fields, as in 'non-inductive' wire-wound resistors. Then we try to make one part of the field cancel another part [66]. Here, A is a wire-wound resistor, resembling an electromagnet. B has a loop: current reverses and flows back. C uses thin flat mica. One-half of each turn lies flat under the other half. D reverses the winding direction half-way.

Electrons are influenced by the presence of magnetic fields. This is useful with television picture tubes, and with magnetrons, and so on.

66. Non-inductive wire-wound resistors

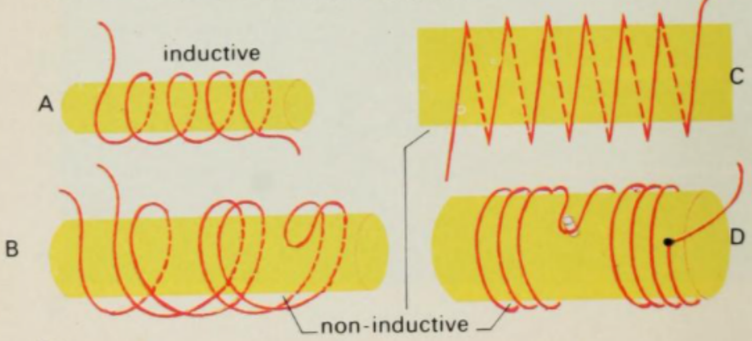

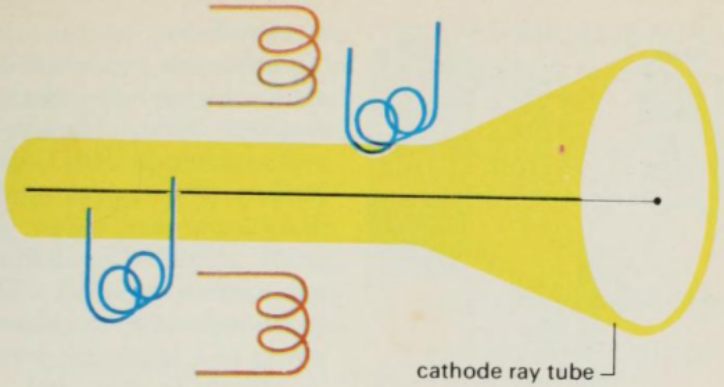

67. Magnetic deflection of cathode ray

A stream of electrons is emitted by the electron gun of a cathode ray tube [67]. The beam passes between coils on the tube. When no current flows in these coils, the beam travels straight to the screen, and produces a central bright spot on the face of the tube.

Current flowing in the coils produces a field which curves the electron path up or down, or from side to side. In this way the beam is made to strike the screen at any point we want.

By altering the strength and polarity of current in the coils, the beam can be made to travel over any part of the picture-tube face. The coils are specially shaped to lie around the tube [68]. Electromagnetic deflection of the electron beam is used in television receivers.

68. Arrangement of coils on cathode-ray tube

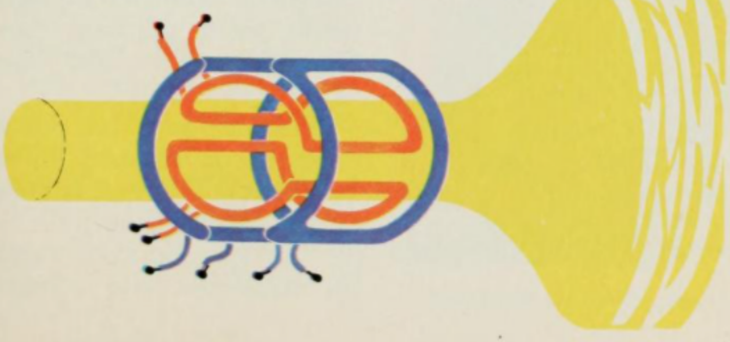

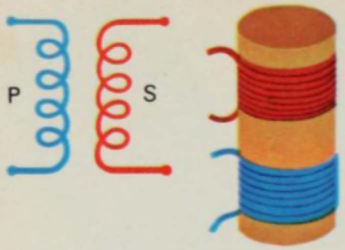

69. An air-core transformer

70. Transformer with core

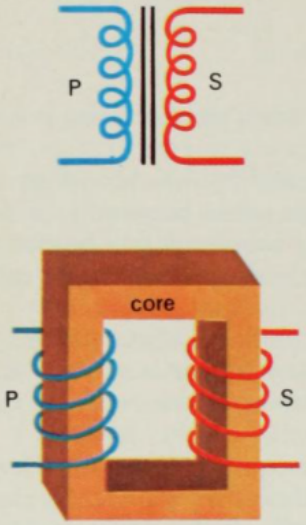

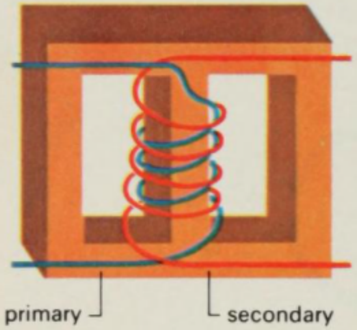

Transformers

Transformers are useful devices relying on electromagnetic effects. They have two windings [69]. The primary (P) receives power from some source (an antenna, amplifier, or alternating current supply). The secondary (S) delivers power to a following device or circuit (a transistor or tube, a loudspeaker, a motor, an electric lamp, etc.).

The transformer may be 'air core' for high frequencies [69]—that is, it may have no solid core material. Currents in the primary develop a fluctuating electromagnetic field, rising and falling rapidly according to the current. This field cuts the turns of the secondary, and induces an e.m.f. in the winding.

For lower frequencies, audio circuits, or alternating current supply, solid cores are placed in the windings. These concentrate the magnetic field, and are generally made up from thin laminations. (See also page 27).

Primary and secondary may be on separate limbs of the core [70] or a bobbin fitted on the center limb.

The transformer changes or transforms one voltage to another, according to the 'turns ratio', or relationship between primary turns and secondary turns [71], given by: Transformer formulas:

$$\frac{N_p}{N_s}=\frac{e_p}{e_s}=\frac{i_s}{i_p}=\sqrt{\frac{Z_p}{Z_s}}$$

The subscripts denote primary and secondary where:

N = Number of turns
e = Voltage
i = Current
Z = Impedance

These relationships presuppose 100% transformer efficiency. Suppose the secondary has one-third the number of primary turns. The ratio is 3:1. If the primary received 240 volts AC, the secondary would deliver 80 volts. But if the secondary had twice as many turns as the primary, the ratio would be 1:2. If the primary received 240 volts, the secondary would receive 480 volts.

Transformers produce the voltages we want, or couple circuits together. A transformer may have several windings [72]. The primary here is tapped for 210/230/250 volt. One secondary supplies 5 volts, another 6.3 volts, and a third 500 volts center tapped, or 250/0/250 volts. That is, 250 volts each side of the tapping.

71. Turns ratio of transformers

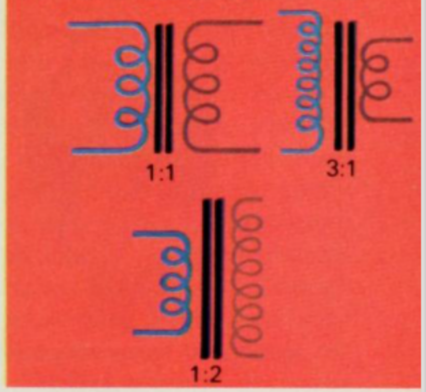

72. Typical supply transformer

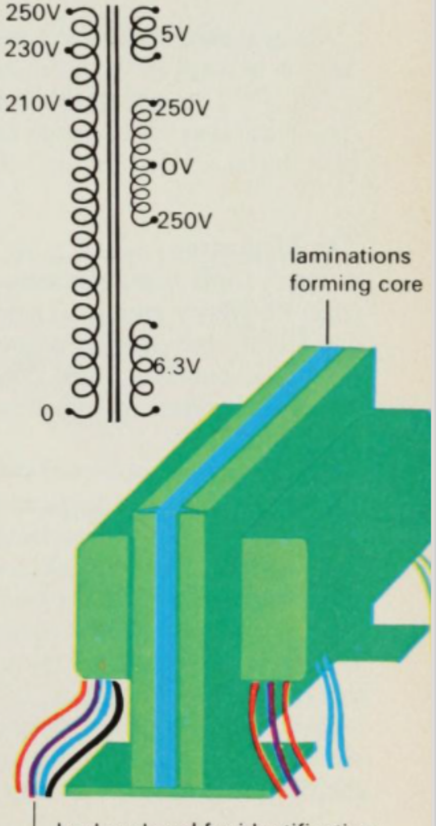

73. A magnetron

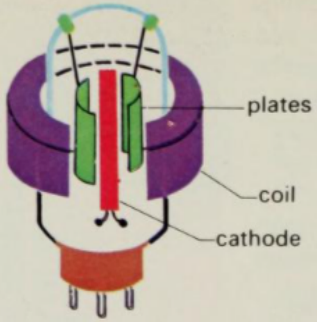
plates
coil
cathode

74. An 8-cavity 10cm. magnetron

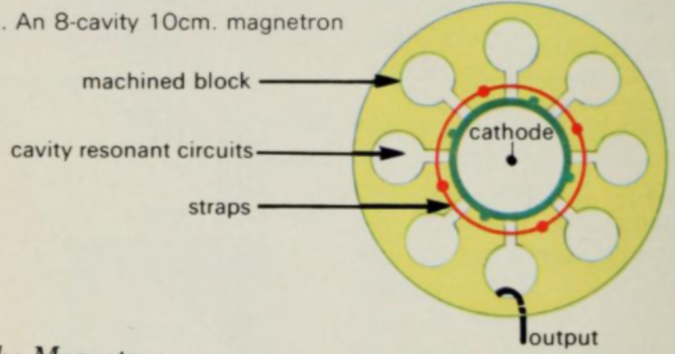
machined block
cavity resonant circuits
straps
cathode
output

The Magnetron

A small 2-plate magnetron—a device which can produce radio frequency energy—is shown in Figure 73. If no magnetic field were present, negative electrons would flow from cathode to positive plates. When the magnetic field intensity is raised the electrons take curved paths and tend to circle the cathode.

An alternating potential is obtained on the plates from a resonant circuit. Electrons approaching the higher potential accelerate. Those approaching the lower potential lose velocity, so that clusters of electrons tend to revolve around the cathode, maintaining the oscillation.

Operation is in brief duty periods (less than one per cent) with high power at high frequency. Large radar magnetrons deliver many thousands of watts, at centimeter wave lengths. For very high frequencies, the plates are formed with cavities which act as the resonant circuit. Plates and cavities are arranged in a ring [74].

Cyclotron

The cyclotron [75] can supply high-energy particles and has two D-shaped members. They are supplied with high frequency alternating current at many kilovolts. The whole is in a vacuum and placed between magnetic poles. Particles travel in a curved path because of the magnetic field. Particles approach each of the D-shaped members as the potential is rising, and this accelerates them. They travel around and around, gaining velocity until ejected through the opening. They bombard other particles, and facts are learned about the nucleus. The synchrotron, with magnets around a circular track, has accelerated particles up to nearly the speed of light. Particles generated include protons, deuterons and alpha particles.

Magnetohydrodynamics, or Gas Flow Electricity

If a conductor is moved across the lines of force of a magnetic field, an electromotive force is generated [76]. When a flow of ionized gas is directed between the poles of a magnet, an e.m.f. is also produced. Current can be taken away by conducting plates.

75. The cyclotron

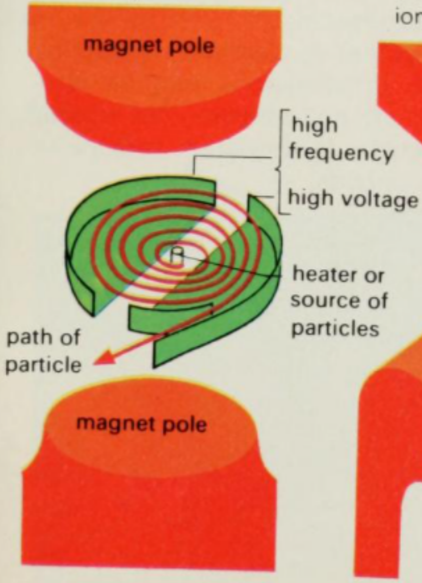
magnet pole
high frequency
high voltage
heater or source of particles
path of particle
magnet pole

76. Generating electricity by ionized gas flow

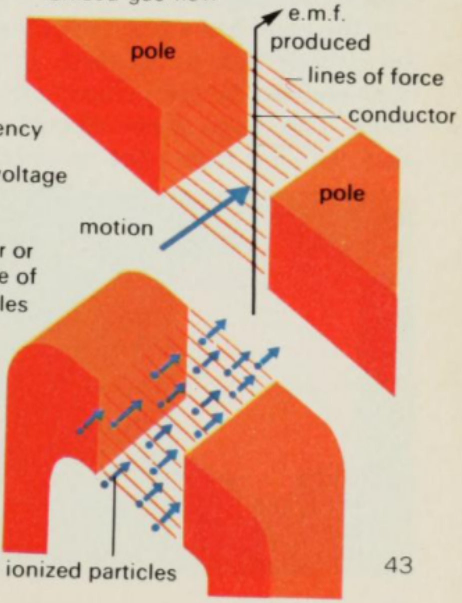
e.m.f. produced
pole
lines of force
conductor
pole
motion
ionized particles

Electron Tubes

A directly heated filament is a thin wire heated by an electric current. It then emits electrons.

Often we have the surface which emits electrons separated electrically from the filament [77]. The filament is electrically insulated from the tube or cathode, which reaches working temperature after an interval. The filament is the heater; the tubes are 'indirectly heated'.

77. Directly heated filament and indirectly heated cathode

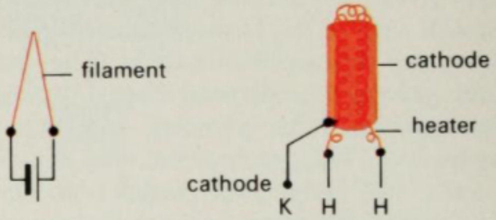

Diode

A thermionic diode has two essential electrodes, and it lets current pass one way only. The diode has a metal plate by the cathode [78]. In working circuits we assume that current will be applied to the heater: so these connections are often omitted in diagrams.

If the plate has a positive potential, electrons emitted by the cathode flow to the plate. If the plate is negative, electrons emitted by the cathode are not attracted. No current flows. The diode allows electrons to flow from cathode to plate, but not from plate to cathode.

78. Thermionic diode

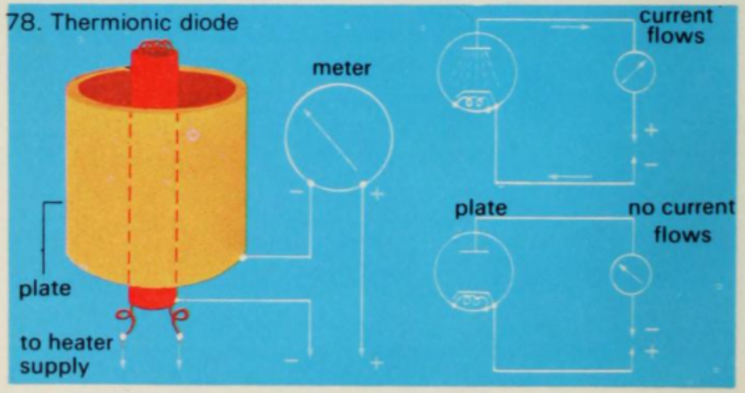

79. Diode as AC power rectifier

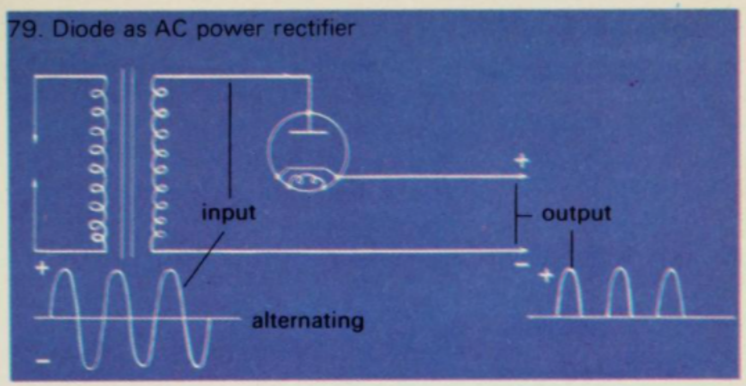

Diode Rectifier

Direct current flows steadily in one direction. Alternating current flows first one way, then the other. With 60 cycle per second supplies there are 60 cycles each second [79]. If we apply this to a diode, current passes only during those intervals when the diode plate is positive. Half-cycles of opposite polarity are suppressed, so we obtain pulsating direct current.

A single diode has cathode and plate. This is also a 'half wave rectifier'. A 'full-wave rectifier' [80] has two plates, and can be connected to make use of both half-waves of the AC supply cycles.

80. A full-wave rectifier

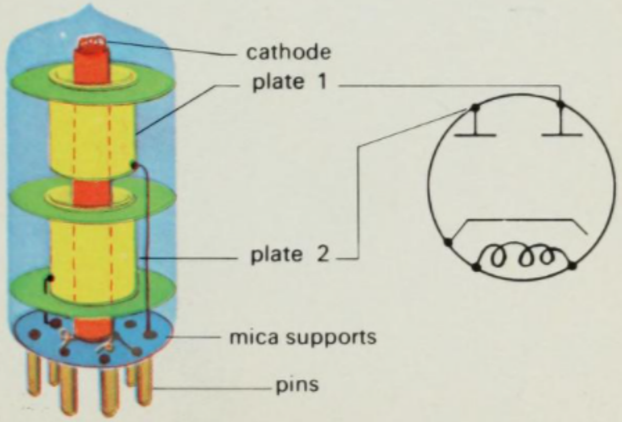

Diode Current

Suppose we provide any chosen voltage by moving the Wiper of a potentiometer VR [81]. A milliammeter indicates the diode current. As the voltage increases, more electrons are attracted to the plate, and current rises. When practically all the electrons are attracted, the current levels off, and scarcely increases at all, even with a much higher plate voltage. This is the maximum, or saturation current.

Full-Wave Rectification

A single diode in a half-wave rectification circuit allowed half-cycles of an alternating supply to flow, so we obtained pulsating direct current. Full-wave rectification is possible with two plates or two separate single diodes [82]. The transformer secondary is center-tapped. For one half-cycle one diode is positive, and for the other half-cycle the other diode is positive. So both half-cycles are utilized.

81. Current flowing through a diode

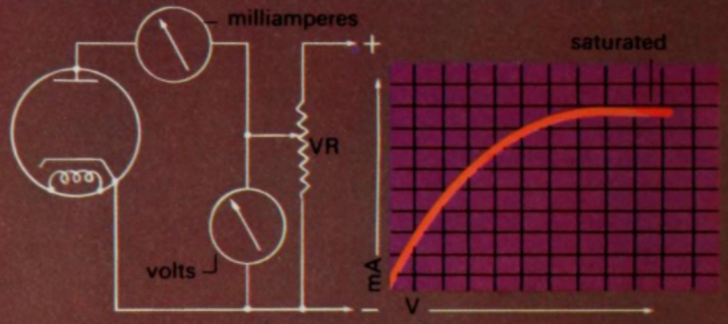

82. Full-wave rectification

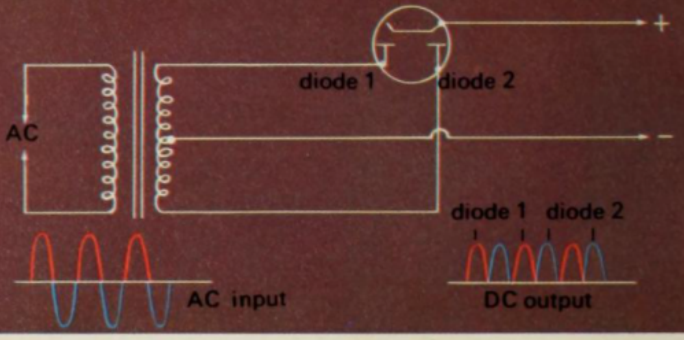

Much AC equipment has a rectifier which supplies direct current for other tubes. The pulsating DC has to be made steady and continuous. 'Smoothing capacitors' across the supply, and a 'smoothing choke' in series do this [83]. The capacitors store energy and give out current. The choke opposes changes in current. Thus a smooth supply, almost like that from a battery, is available. Plate voltages are called the 'B' supply.

Uses for Diodes

Thermionic diode power rectifiers [84] can supply large currents at high voltage. Smaller diodes provide detection or demodulation, or rectify signals for automatic volume control bias. With an indirectly heated cathode, heater and cathode can be connected to different circuits. But if it has a directly heated cathode or filament, there is no separate cathode connection. This can be important with power rectifiers.

83. Power supply smoothing circuit

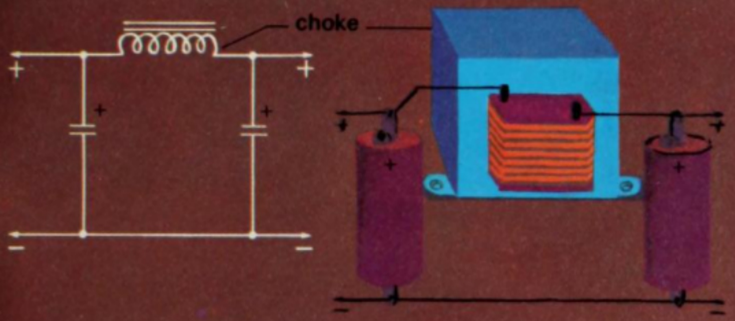

84. Thermionic rectifiers with directly heated and indirectly heated cathodes

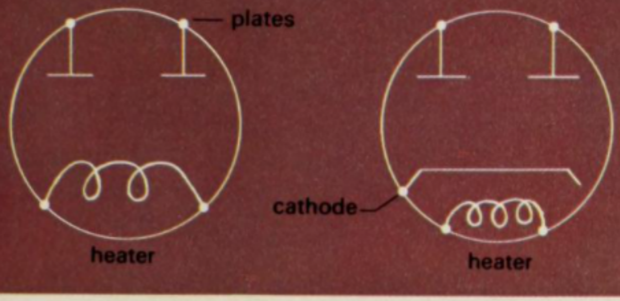

The Triode

A diode has two important electrodes: cathode and plate. The triode has a third electrode, a grid of thin wires, between cathode and plate [85]. Electrons emitted from cathode to plate pass between the meshes of the grid. The triode can 'amplify' signals, or increase their strength. If the grid is at or near zero potential it has little effect on the flow of electrons.

Suppose a meter placed between the plate and supply shows 8mA with−1 volt [86]. If we make the grid more negative, electrons are repelled from its vicinity, so that plate current is reduced. With−2 volts at the grid, plate current has fallen to about 6mA. At−3 volts current is 4mA. If the grid is made very negative, virtually all the electrons emitted by the cathode are repelled by the negative charge on it, and do not reach the plate.

Changes in grid voltage cause changes in plate current. This grid is called the 'control grid'. A steady D-C voltage is called 'grid bias'. Changing the grid potential by 1 volt causes plate current to change by 2mA. The tube has a mutual conductance of 2mA per volt, or 2mA/V.

85. A triode has control grid between cathode and plate

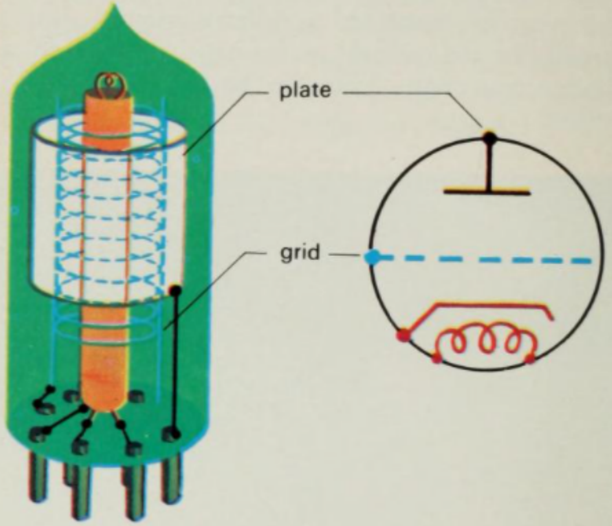

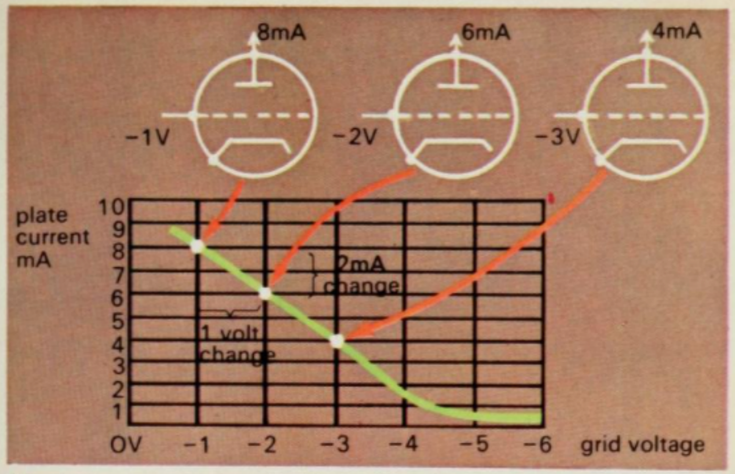

86. Plate current controlled by grid voltage

We can apply various voltages, and read current with a meter [87]. As we raise the voltage, current increases. At 50 volts it is 4mA, but it has risen to 6mA at 100 volts. If the bias voltage is unchanged, altering the plate potential by 50 volts changes the plate current by 2mA [88]. In Figure 86 changing the grid potential by 1 volt changed the plate current by 2mA. So altering grid potential by 1 volt has the same result on plate current as changing the plate potential by 50 volts.

87. Changing plate voltage also changes plate current

88. Plate voltage and current

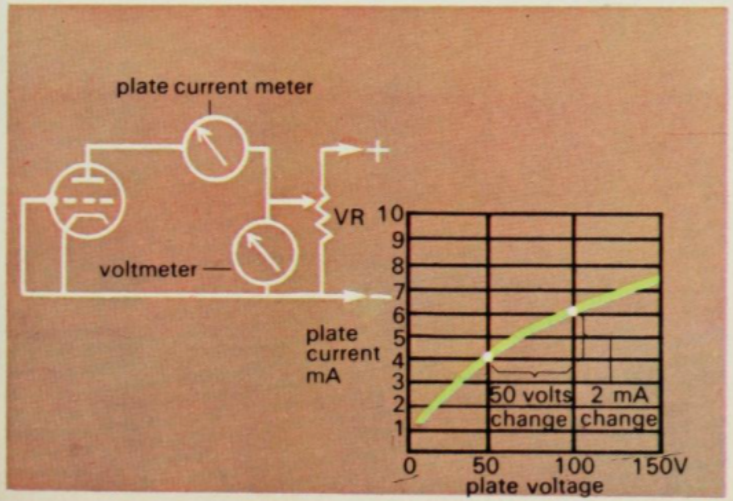

Triode Amplifier

An audio signal is applied to X in Figure 89. Resistor R1 completes the tube grid circuit. Cathode current flows through R2. We need the control grid to be −2 volts relative to the cathode K. So R2 drops 2 volts. C1 by-passes R2 at frequencies to be amplified.

With no signal at X, bias is −2 volts and plate current is 6mA. The audio signal is 2 volts from peak to peak, so swings the grid from 1 volt to 3 volts. Plate current swings from 8mA to 4mA. The voltage dropped in resistor R3 depends on plate current. So the plate voltage also swings up and down. An amplified signal is taken off by capacitor C2. Such capacitors isolate other circuits [90].

More Electrodes

A screen grid G2 betweeen control grid G1 and plate [91] has a positive potential. Plate current depends less on the plate voltage, and amplification is increased. Electrons may be knocked off the plate and return to the positively charged screen grid. The suppressor grid G3 hinders this. This tube is a pentode. Shaped plates between screen grid and plate in a 'beam tetrode' also hinder the flow of electrons from plate to screen grid.

We control the electron stream from two circuits by using a tube with more grids, such as a heptode [92]. Signals applied to grids G1 and G3 control the electron flow. Grids 2 and 4 form a screen each side of G3. G5 is the suppressor grid.

A duo-triode has two triodes in a single bulb. A duo-diode triode has two diodes together with a triode assembly. In the same way, there are other tubes, such as triode-hexodes, triode pentodes, pentodes with diodes, and duo-pentodes.

89. Thermionic triode amplifier

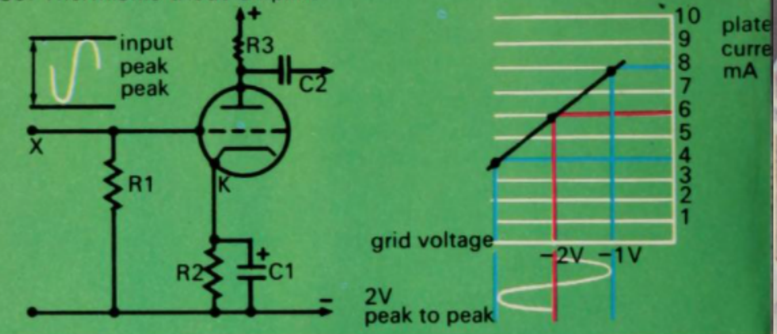
input
peak
peak
+
R3
C2
X
R1
K
R2
C1
+
-
grid voltage
-2V
-1V
2V
peak to peak
10
9
8
7
6
5
4
3
2
1
plate
curre
mA

90. Isolating direct current circuits by coupling capacitors

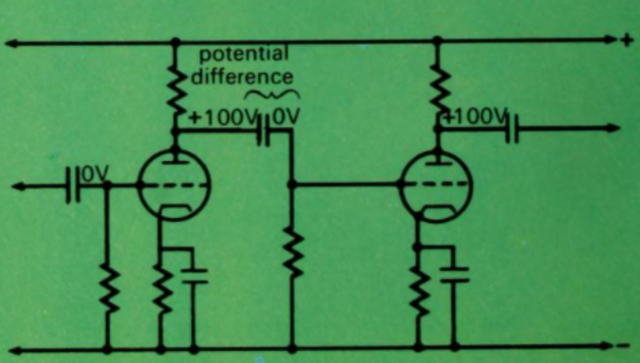
potential
difference
+100V
0V
+100V
0V
+
-

91. Screen grid, tetrode and pentode tubes

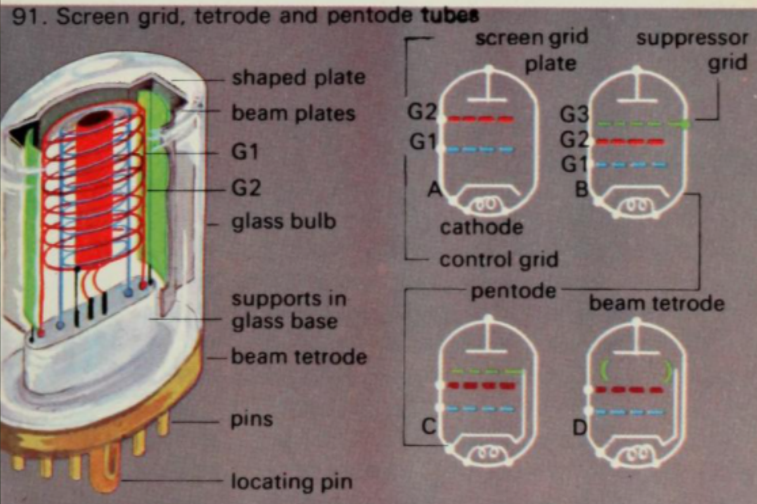
shaped plate
beam plates
G1
G2
glass bulb
supports in
glass base
beam tetrode
pins
locating pin
screen grid
plate
suppressor
grid
G2
G1
A
G3
G2
G1
B
cathode
control grid
pentode
beam tetrode
C
D

92. Two more multi-electrode tubes

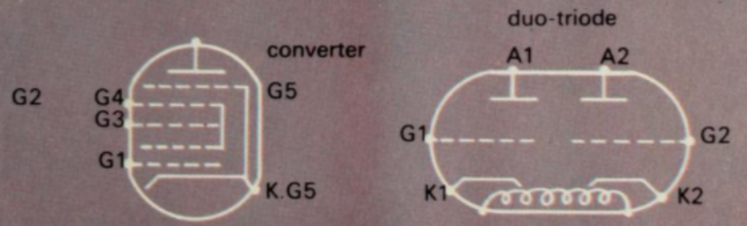
converter
G2
G4
G3
G1
G5
K.G5
duo-triode
A1
A2
G1
G2
K1
K2

Variable-Mu Pentodes or Remote Cut-off Pentodes

Amplification can be controlled by using a grid having irregular spacing [93]. By increasing the negative bias, the amplification can be reduced to a very small fraction. The tube might have a mutual conductance of 5mA/V with zero bias, falling to 0.005mA/V and 20 volt bias.

Negative Resistance Condition

If a screen grid tube has certain screen grid and plate potentials, electrons strike the plate at a velocity causing other electrons to be liberated. These 'secondary emission' electrons return to the screen grid. When potentials are suitable, and the screen grid potential is higher than the plate, secondary emission is considerable. If the plate voltage is increased, electrons strike it with added velocity, and liberate more secondary electrons. So over a certain part of the operating curve plate current *falls* for an increase in plate voltage [94]. This is 'negative resistance'. It can cancel the actual or positive resistance of a circuit, sustaining oscillation.

93. Control grid of variable mu tube

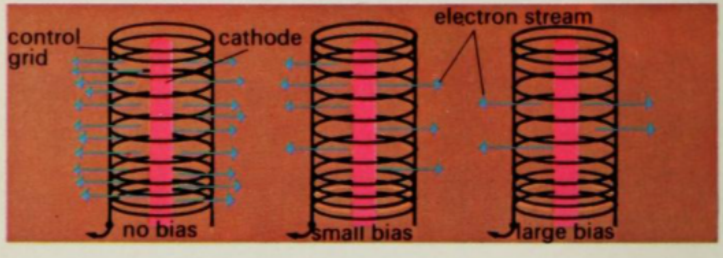
control grid
cathode
electron stream
no bias
small bias
large bias

Tube Component Parts

Figure 95 shows parts of a small transmitting tube. Many receiver tubes are quite similar in design, though generally with all connections to base pins. The heater is in the cathode. Then come control grid, screen grid, beam-forming plates and plate. The whole is supported by mica parts pushed inside a glass bulb. As high voltages are present, the plate has a cap. The insulated base has pins. Smaller tubes may be all glass, with pins in a glass base.

94. Negative resistance screen grid tube

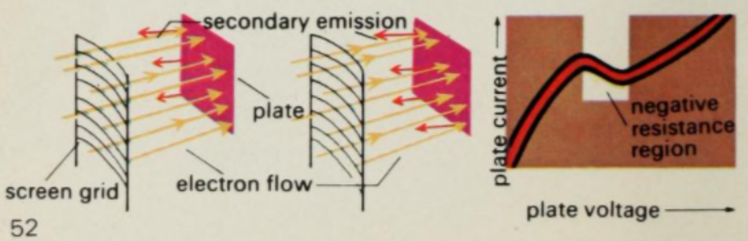
secondary emission
plate
screen grid
electron flow
plate current
negative resistance region
plate voltage

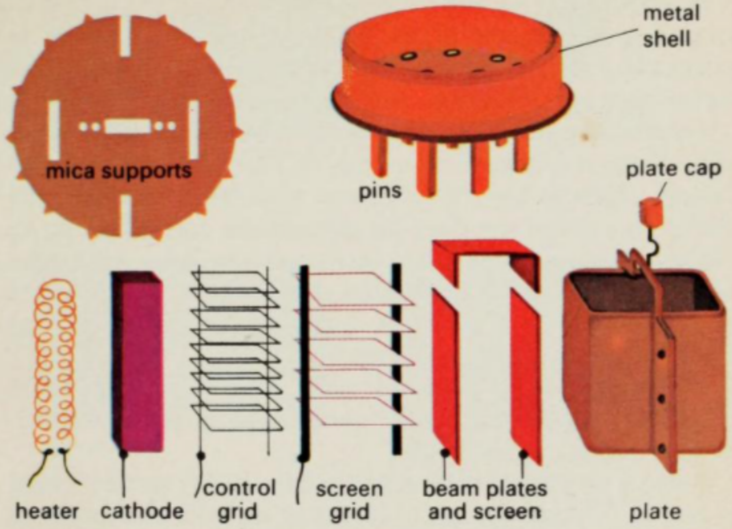
metal shell
mica supports
pins
plate cap
heater
cathode
control grid
screen grid
beam plates and screen
plate

95. A small transmitting tube—component parts

Gas-Filled Thyratron

This contains a gas which ionizes when sufficient current flows, reducing the internal resistance [96]. When the grid is sufficiently negative no electron current flows. Less negative potential causes electrons to flow, ionizing the gas. The tube conducts heavily until the plate potential is reduced.

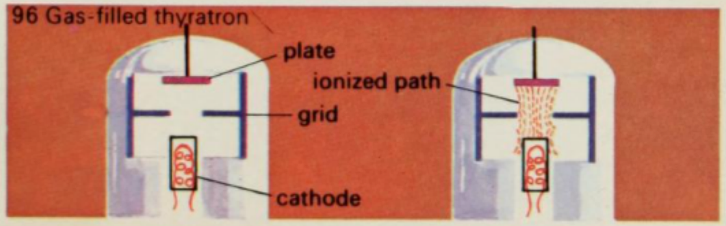
96 Gas-filled thyratron
plate
ionized path
grid
cathode

Semiconductors

The electrons of an atom have various energy levels so that they can be thought of as occupying different bands or positions. Free electrons in the outer position, or shell, can move from atom to atom, forming an electric current.

Germanium has 32 electrons, with four in the outer shell. By adding impurities, an excess or lack of electrons is arranged. Indium has 49 electrons, with three in the outer shell. This is one less than the germanium outer shell, so an electron from a germanium atom can move in to occupy the space, leaving a germanium atom with a deficiency of one electron. This, a positive 'hole', can accept an electron from another germanium atom [97].

Antimony has 51 electrons, with five in the outer shell. This extra electron can displace an adjoining electron, which in turn can move elsewhere.

Lack of an electron leaves a positive charge, so the crystal is termed 'P-type'. Excess of electrons constitutes a negative charge and gives 'N-type' material. Movement of the electrons or hole carriers forms an electric current.

A semiconductor junction diode is a junction of N-type and P-type regions, and is often called a PN diode. Current in N-type germanium is conveyed mostly by electrons, but in the P-type mostly by holes [98].

97 Germanium with positive and negative carriers

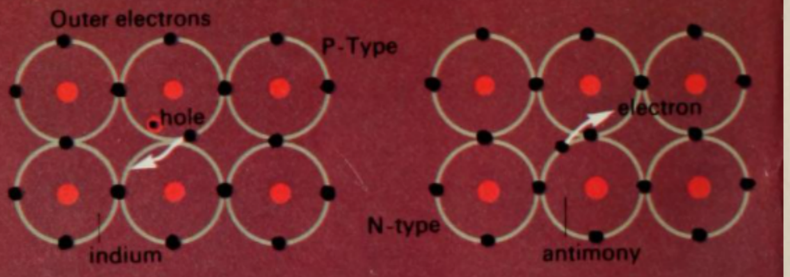

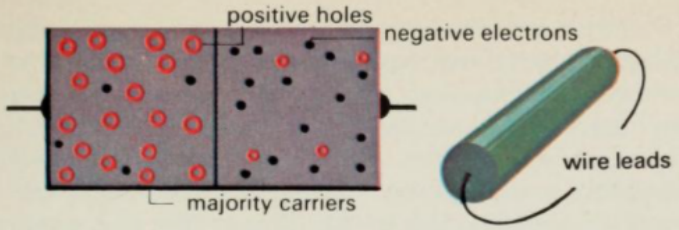

98. PN junction diode

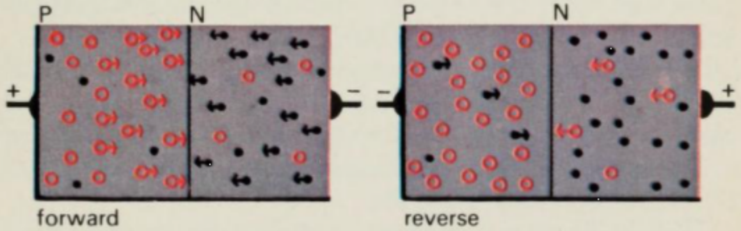

99. Junction with forward and reverse voltage

The diode allows current to pass readily in the 'forward' direction. There are many positive holes to move across to the N-region, and numerous negative electrons to travel to the P-region [99]. This represents a large flow of current across the junction. With the voltage reversed, there are few negative electrons in the P-region, and few positive holes in the N-region, so that the current is small. Semiconductor diodes are made to suit the voltage, current and frequency [100]. Point contact diodes pass small currents, but can operate at very high frequencies.

Germanium rectifiers tend to fail around 75° C, while silicon can withstand up to about 200° C. Thus a silicon rectifier is less affected by rises in temperature, and can operate with currents of many amperes.

100. Some semiconductor diodes

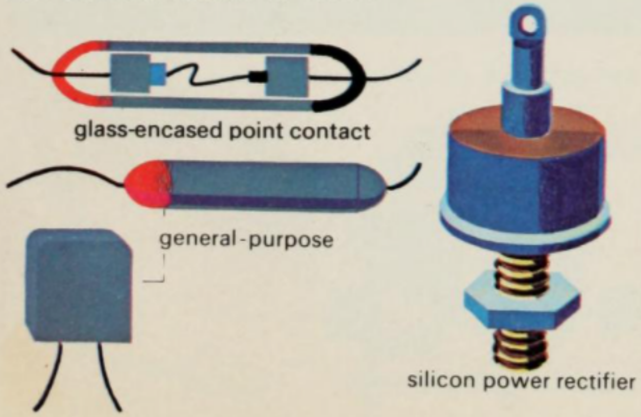

Semiconductor diodes are now much used in power supply circuits for rectification. They are small in size, have low forward resistance, can handle high voltages and current, and need no separate heater supply.

Many electronic circuits, including those of receivers and amplifiers, require direct current. The alternating current supply is usually taken to the primary (P) of a power transformer [101]. This isolates equipment from any actual connection with the supply. Voltage can be reduced or increased by a suitable turns ratio.

The diode allows current to pass readily one way and so can charge a reservoir capacitor. Compare this with the half-wave tube circuit. Semiconductor diodes can also be used in full-wave circuits, to make use of both half-cycles of the AC supply, as with tubes.

When rectifiers are working at maximum current, heat is produced. This is carried away by a metal plate or heat sink. Rectifiers may be connected in series or parallel, to obtain a higher voltage or current rating, or they can be stacked in special assemblies [102].

101. Power diode allows DC to be obtained from AC

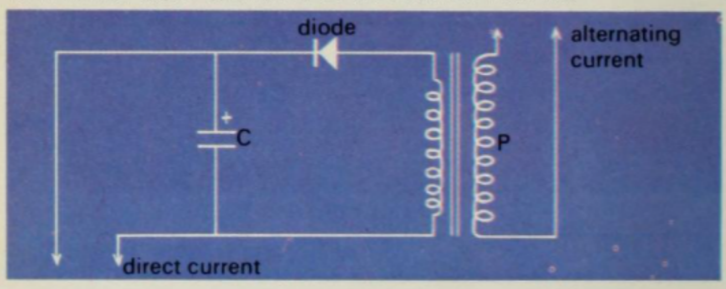

102. Rectifier stack with convection cooling

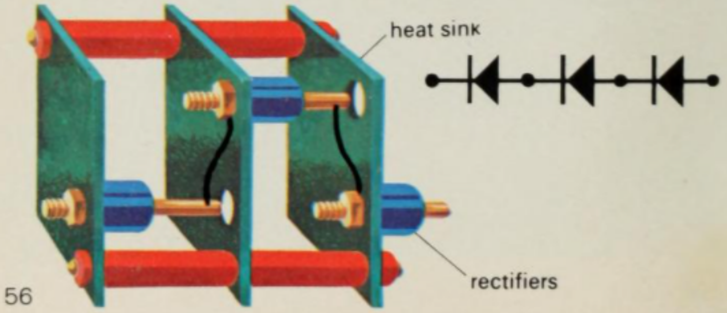

Compensating Diode

In transistor receivers and amplifiers, working conditions of the output transistors which supply the loudspeaker are quite critical. Wrong bias will increase battery drain, or cause distortion.

Base bias which is suitable for a new battery may be unsatisfactory with one that is partly discharged. The temperature at which they are operating affects somewhat the working of transistors.

A compensating diode [103] helps overcome this. The diode's resistance (and transistor operation) varies with the temperature. When battery voltage falls, current through the diode is less. Its resistance rises, helping to shift the base supply more negative and restore satisfactory operation.

RF Probe

The point contact diode has extremely small internal capacity, and can operate at very high frequencies. It is suitable for a radio frequency probe [104]. The probe can be taken to any circuit point where RF is present.

103. Diode to stabilize performance

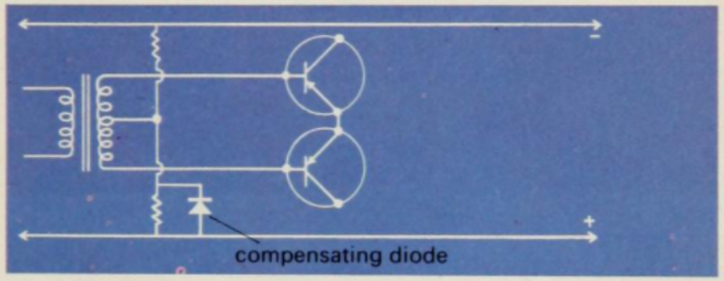

104. Isolated VHF probe

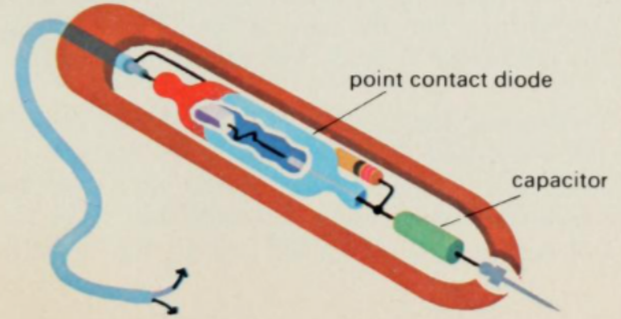

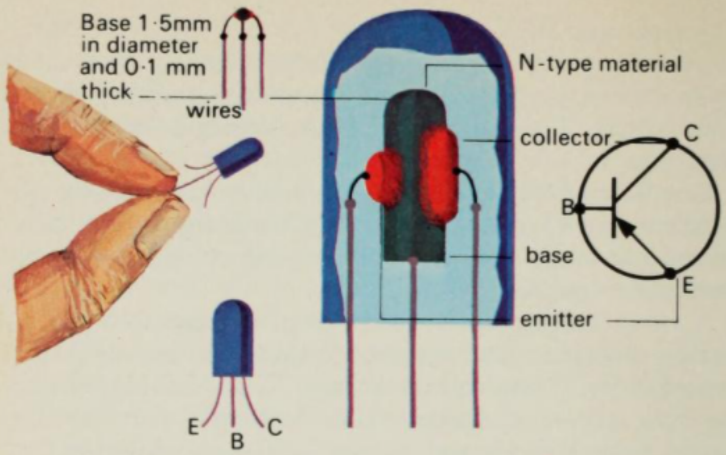
Base 1·5mm in diameter and 0·1 mm thick
wires
N-type material
collector
C
B
base
E
emitter
E
B
C

105. A PNP transistor

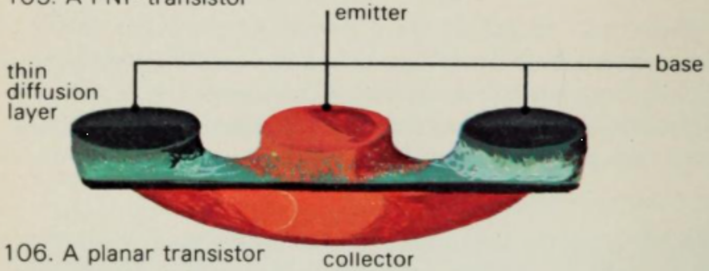
emitter
base
thin diffusion layer
collector

106. A planar transistor

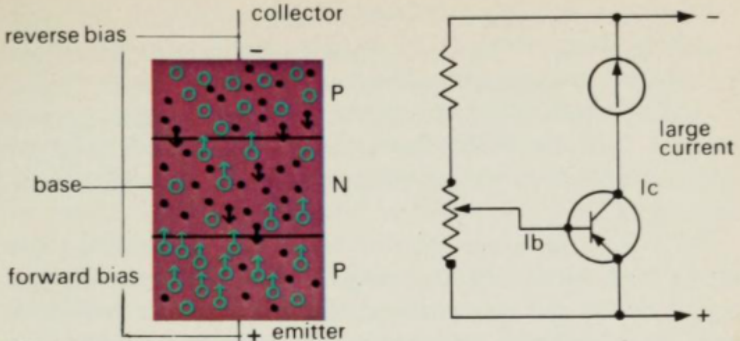
collector
reverse bias
–
P
base
N
forward bias
P
+ emitter
–
large current
Ic
Ib
+

107. Large flow of current

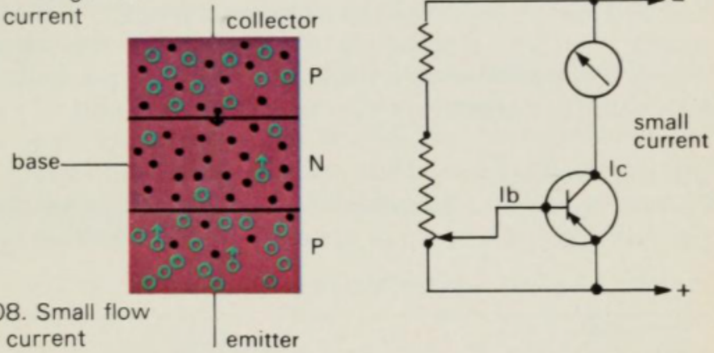
collector
P
base
N
P
emitter
–
small current
Ic
Ib
+

108. Small flow of current

Transistors

A transistor has a central region, or base, of one kind of semiconductor material, with junctions of an opposite kind of material on each side. A PNP transistor has positive-negative-positive material, with a negative central base [105]. The NPN transistor has negative-positive-negative material—a positive central region or base.

For the PNP transistor, pellets of indium are fused on by heating, and diffuse into the base to form P-type regions. The whole is fitted with leads and sealed in a protective case.

Such a transistor becomes less efficient at high frequencies. An alloy junction transistor has a base region only a small fraction of a millimeter wide. Planar transistors [106] made by alloy diffusion have an extremely thin active base region, and construction which reduces the time required for the movement of the current carriers.

The PNP emitter is connected to a positive supply, and the collector to a negative [107]. If the base is appreciably negative, relative to the emitter, the emitter-base junction acts as a diode junction with forward bias. Holes pass into the base, and are drawn by the negative charge of the collector, so that they usually flow to the collector. Current passing through the transistor is large.

If the forward bias is reduced, fewer holes travel to the collector. The current flow is quite small [108]. In a PNP transistor, there is a small movement of 'minority' or negative electron carriers, the 'majority' or most numerous carriers being positive holes—a positive charge equal to an electron. Base and collector currents added together form approximately the emitter current. Almost all the emitter current passes to the collector.

Note the important difference between PNP and NPN transistors [109]. The PNP transistor base is negative-type material, and majority current carriers are positive holes, caused by a deficiency of negative electrons. Emitter goes to *positive*. With the NPN transistor, the base is positive-type material, and electrons are majority carriers. Emitter goes to *negative*.

Note that the arrow symbol is always the emitter, and agrees with the motion of positive hole carriers. We can regard this as the conventional electrical current flow from positive to negative, encountered in some text books. The flow of electron current is from negative to positive. A heavy current may flow if a supply is connected in the wrong polarity to a transistor. If resistors and so on do not impede this current, transistor junction may be damaged.

Internal Feedback

The electrodes of a transistor are very close together and this and other effects cause unwanted feedback from the collector. In Figure 110, an imaginary capacitor connection from collector to base stands for the unwanted feedback. This is usually unimportant at low frequencies, but troublesome at high frequencies, where special precautions have to be taken to avoid oscillation or other effects.

Transistors for high frequencies may have a fourth lead, or shield, between base and collector [110]. It helps to reduce unwanted feedback.

Receiver transistors [11] are usually quite small, with insulated or metal cases.

They are made for radio frequency amplification, audio amplification, etc. They are also produced in small sizes for miniature equipment.

Power transistors [112] for RF and audio amplification are larger; usually they can be clamped to a metal plate or heat sink, to help avoid overheating. Some transistors are made in similar or matched pairs, in a single case. Other transistors are intended for computer ‘logic’ and various special applications.

109. PNP transistor and NPN transistor compared

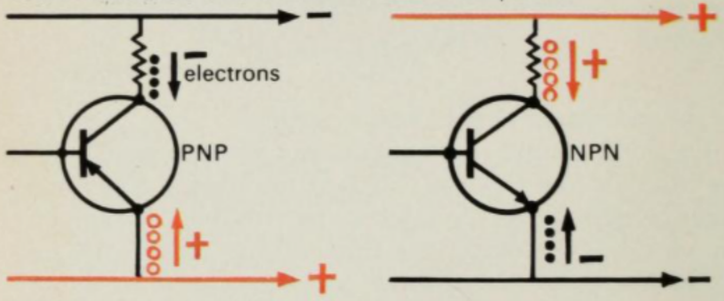

110. Screened transistor

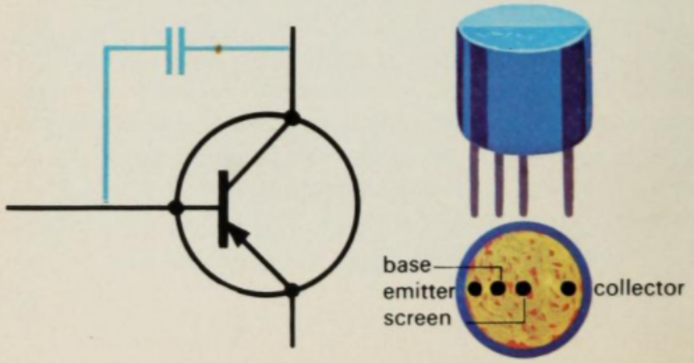

111. Small receiver and amplifier transistors

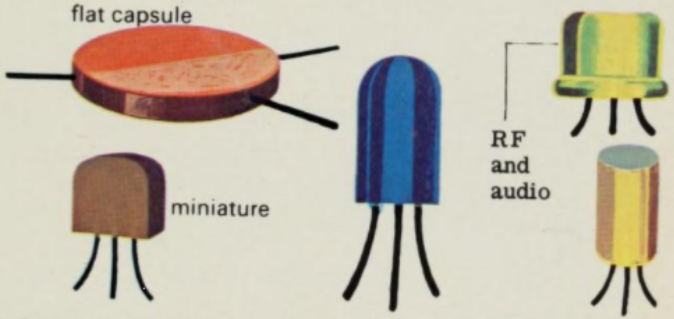

112. Power and matched pair transistors

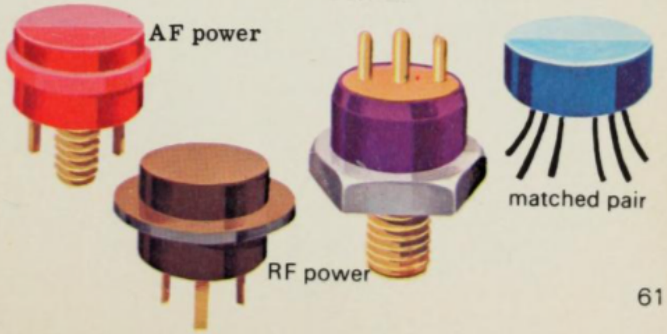

Some Circuit Points

We use transistors in various ways. Some are shown in Figure 113. Here A has input to emitter and base, with output from collector and base. R is a load resistor across which power is developed, or a tuning coil or other device. Both input and output circuits use the base, so A is a *common and grounded base* arrangement.

B has input between base and emitter, and output developed across collector and emitter circuits. It is a *common and grounded emitter*. In the same way C is a *common emitter grounded collector* arrangement.

A, B and C are NPN transistors. D is a PNP transistor with input to base and emitter, and output developed across collector and emitter circuits. It is a *common emitter grounded emitter* circuit.

Some methods provide best power gain. Others operate better at high frequency. Some have high impedance input and low impedance output, or the reverse. The 'impedance' is the 'resistance' at operating frequency. Figure 114 shows the transistor as a current generator, with input impedance R1 and output impedance R2. Maximum useful power results when input and output circuits match R1 and R2.

Dynamic Operation

Figure 115 is a PNP common emitter amplifier, with base bias set by R1 and R2 and output developed across R3. C1 and C2 isolate the stage from other direct current points. R4 provides emitter bias to stabilize operation, and has a by-pass capacitor, C3.

Base current Ib is at some small steady value. Collector current Ic is also at a steady value, and emitter current Ie approximately equals Ib plus Ic. Currents are indicated without regard to polarity. When a signal to be amplified arrives through C1, base current swings up and down with each cycle, as shown in [116-Ib]. Slight increases in base current cause large increases in emitter-collector current. So collector current Ic swings up and down in value, developing an output voltage across R3.

A moving-coil meter would show the average currents [115], an oscilloscope dynamic conditions.

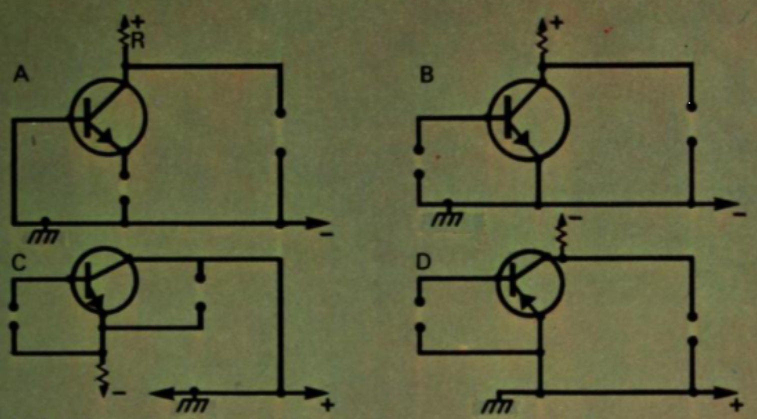

113. Ways of using a transistor

114. Impedance matching

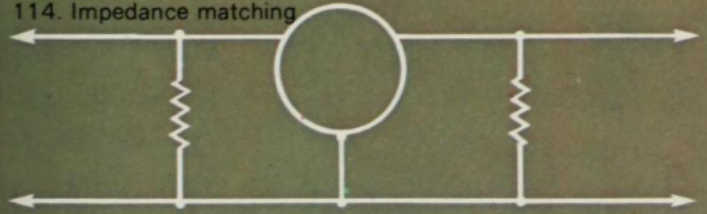

115. Static conditions of transistor

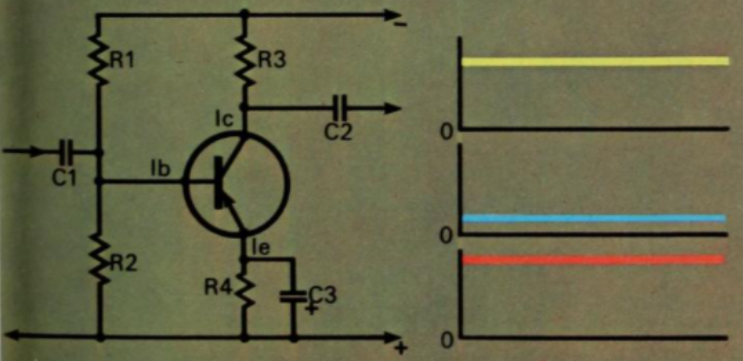

116. Dynamic working during amplification

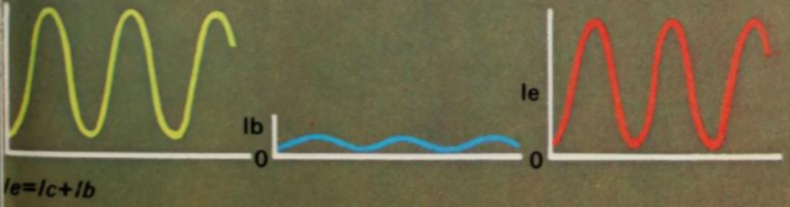

AUDIO, RADIO, TELEVISION, TAPES

Audio Amplifiers

A signal is 'amplified' to make it stronger. An audio amplifier deals with frequencies from about 15 c/s to 15,000 c/s. Musical items mostly lie in the 50 c/s to 2,000 c/s range, but 'harmonics' must be reproduced for good results. The electrical output of a record-player pick-up may be a few millivolts. A loudspeaker needs much more power, so signals are amplified [117].

Figure 118 is a triode tube audio amplifier. There is 2 volts negative bias through R1 for the control grid. The tube passes 1mA from its cathode to its plate. Plate current flows through R2. From Ohm's Law ($E = I \times R$) the voltage dropped at R2 is 100 volts. Suppose a positive signal peak of 1 volt reaches the grid [119a]. The grid voltage is now —1 volt. Plate current rises to 1.5mA. So 1.5mA flows through R2. The voltage drop is now 150 volts, leaving 50 volts at the plate. Imagine that a 1 volt negative peak reaches the grid, B. This adds to the—2 volts present, giving—3 volts. Suppose plate current falls to 0.5mA. The voltage lost in R2 is now 50 volts, so the plate potential is 200–50 = 150 volts.

Hence by applying a signal which swings 1 volt above and below zero we have obtained an output or plate voltage which swings from 150 volts to 50 volts.

C1 is the usual coupling capacitor [120]. It interrupts any direct current path, but it has little reactance at audio frequency, and therefore lets signals pass. In the same way, C2 feeds another stage.

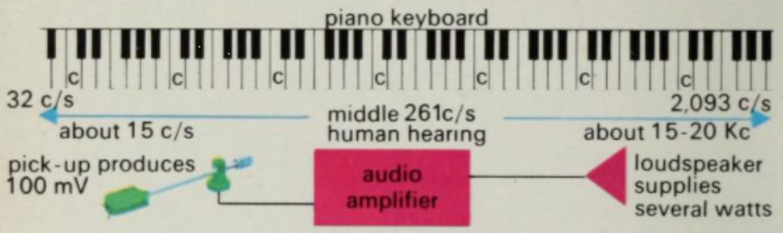
piano keyboard
c
c
c
c
c
c
c
32 c/s
about 15 c/s
middle 261c/s
human hearing
2,093 c/s
about 15-20 Kc
pick-up produces
100 mV
audio
amplifier
loudspeaker
supplies
several watts

117. Audio spectrum and need for amplification

118. A triode tube audio amplifier

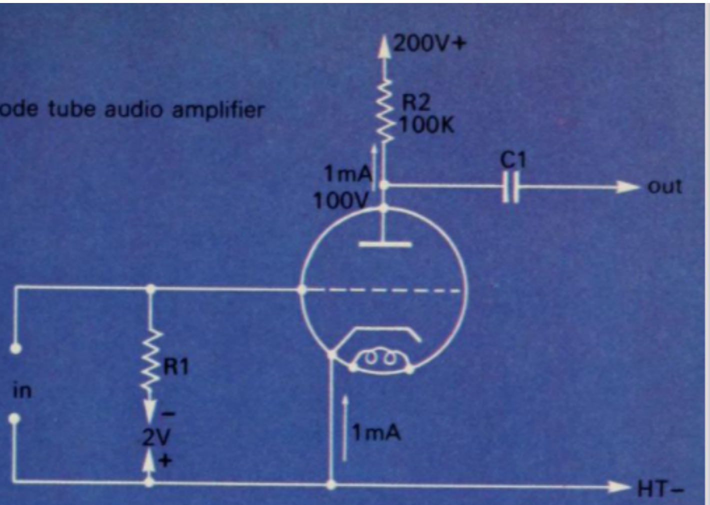
200V+
R2
100K
C1
1mA
100V
out
R1
in
2V
1mA
HT–

119. The triode amplifying a signal

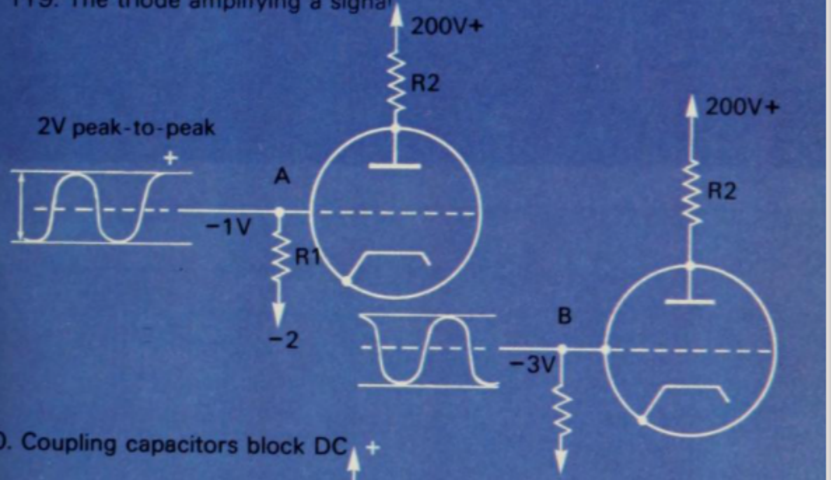
200V+
R2
2V peak-to-peak
+
A
–1V
R1
–2
200V+
R2
B
–3V

120. Coupling capacitors block DC

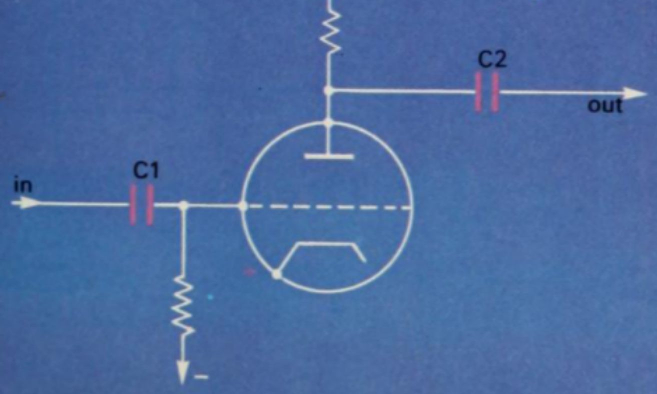

We need the control grid to receive a small negative bias voltage. Figure 121 has cathode bias resistor R3, which must drop 2 volts with 1mA passing. $\frac{E}{I} = R$. So R3 is 2,000 ohms.

C3 across R3 'by-passes' the cathode to ground at operating frequency. We often use a pentode and get higher 'gain' or amplification than from a triode [122]. VR is a potentiometer which allows any portion of the signal voltage to be taken, and so acts as volume control. RS supplies the screen grid G2, and CS is the screen grid by-pass capacitor. The suppressor grid is at the same potential as the cathode.

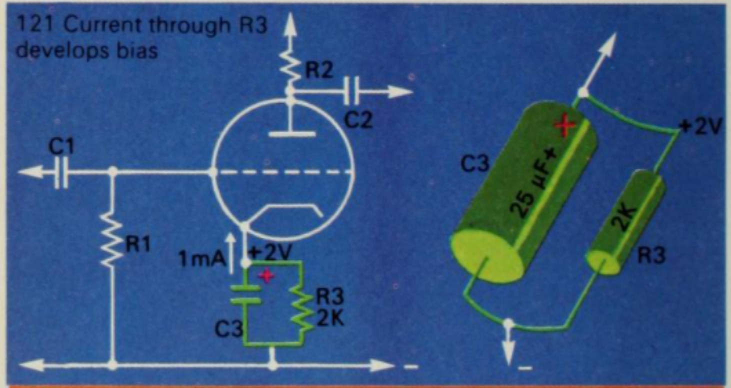
121 Current through R3 develops bias

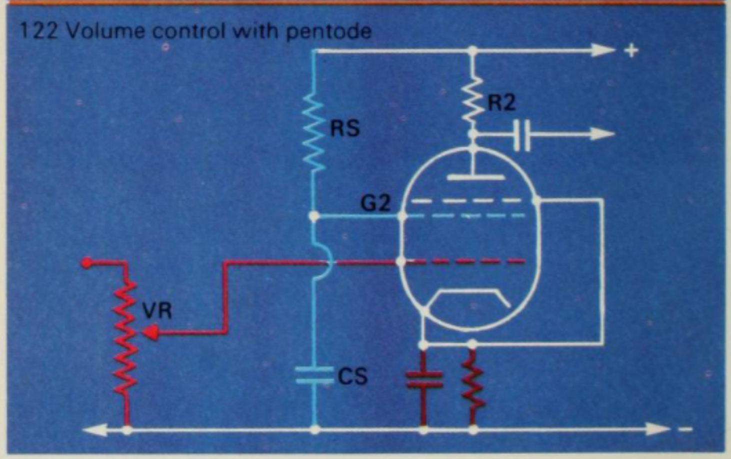
122 Volume control with pentode

Useful work is done by the plate load resistor R2—a signal voltage is developed across it. The tube, with resistors and other components, is termed a 'stage'. The 'stage gain' or amplification achieved is less than the amplification factor of the tube.

Imaginary resistor Ra [123a] represents the tube internal impedance and the plate load resistor R2 is chosen to suit this. With output tubes this is termed the 'optimum' (best) load. In Figure 123b it is furnished by the transformer and speaker. By choosing a suitable transformer ratio, we can provide any optimum load we choose.

Loudspeaker impedance will often be about 2 to 15 ohms. Transformer ratio is usually from about 15:1 to 60:1.

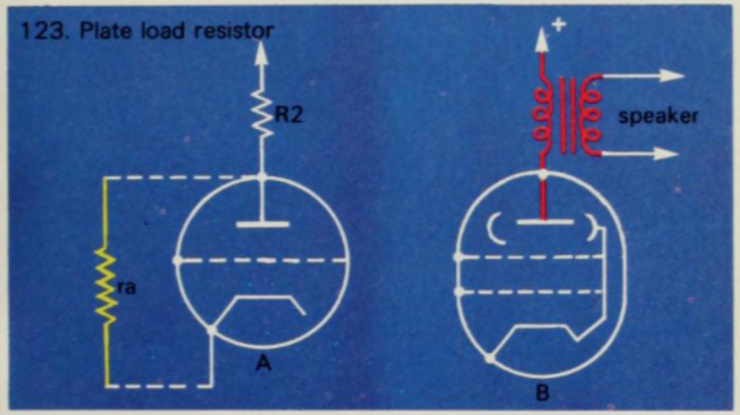
123. Plate load resistor

124 Three-stage amplifier

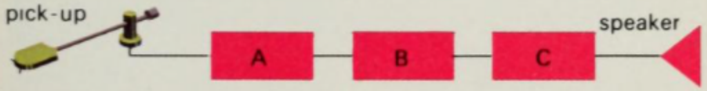

Cascaded Stages

We usually need more amplification than is provided by one stage, so use two or more stages [124]. Here A could be an initial voltage amplifier [122], B a further amplifier stage [121], and C an output stage working a loudspeaker.

Transformer Coupling

In Figure 125 the transformer primary winding receives audio signals from an earlier stage. The secondary delivers signals to the next grid. This system is little used today, but the diagram will help to explain circuits given later. When one end of the secondary swings negative, the other end is positive. For the next half-cycle, this is reversed.

Figure 126 has a center-tapped secondary. For one half-cycle V1 receives a negative signal, but V2 a positive signal. So V1 passes less plate current, and V2 more plate current. During the next half-cycle the opposite occurs so that V1 current increases and V2 current falls.

The 'half outputs' from V1 and V2 go to the output transformer primary. They are combined and work the speaker, connected to secondary S. This is called 'push-pull'. We can also provide so much bias that plate current is nearly cut off except when a tube grid receives a positive signal voltage. This is called 'Class B' operation, and is less drain on the supply.

125. Audio transformer for coupling

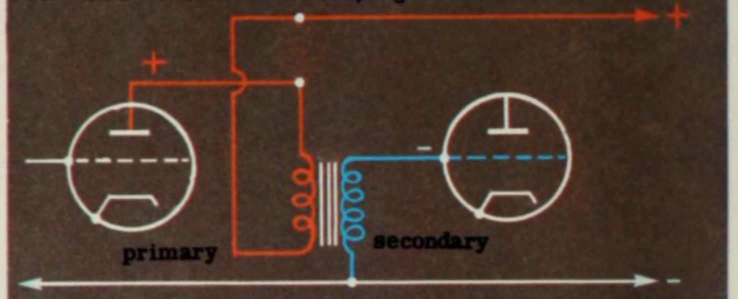

126. Push-pull circuit

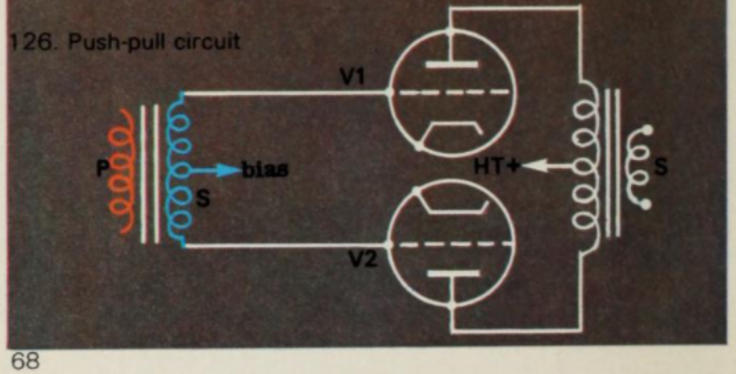

Resistance Capacity Coupling can be employed instead. In Figure 127, V1 receives audio signals at G1. When its plate current rises, the voltage drop in R1 and R2 increases. So point X swings negative, and Y swings positive. The result is like using a center-tapped transformer to supply V2 and V3. Signals go via C1 and C2. R3 and R4 are grid resistors. R5 supplies cathode bias.

Negative Feedback

We can take part of the output signal and return it to *oppose* the original signal. This reduces distortion. Components such as capacitors have varying effects as frequency changes. Suppose the amplifier gives most amplification to high tones, or has too much treble. If we feed back signals to oppose the input, these top frequencies will be most strongly present. This will reduce high frequencies more than lower frequencies. The overall result is better 'fidelity', or faithfulness of amplification. In Figure 128, feedback is through R1 to the tube cathode.

127. Push-pull circuit from a typical amplifier

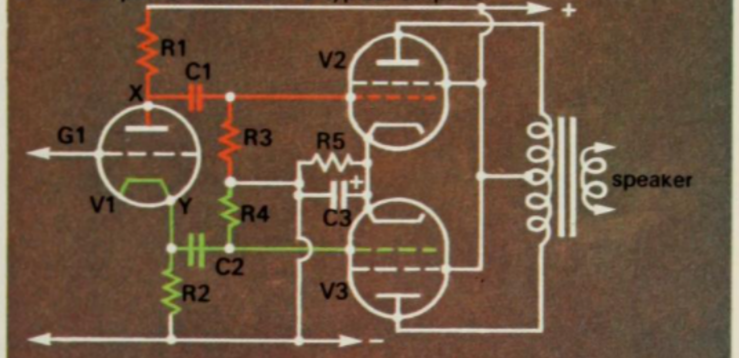

128. Negative feedback

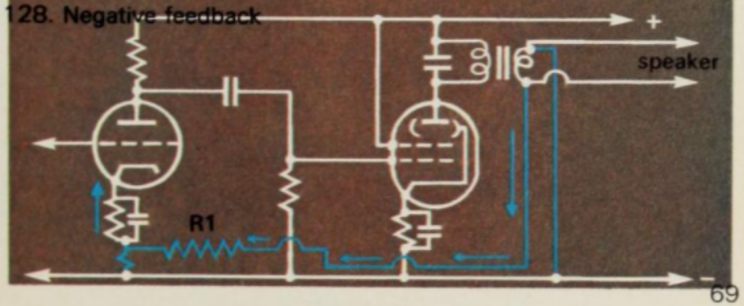

129 Two stages of transistor amplifier

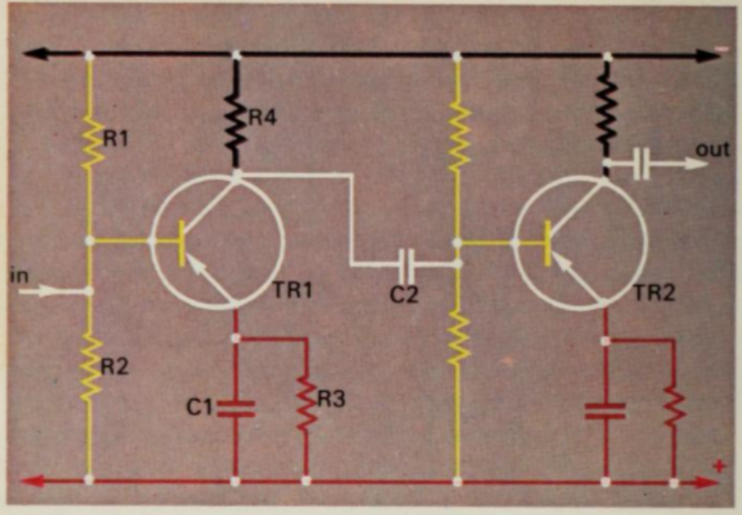

Transistor Audio Amplifier

Figure 129 has TR1 as audio amplifier. R1 and R2 are chosen so that a small base current flows and the PNP transistor base is slightly more negative than the emitter. R3 stabilizes the working conditions by emitter bias. If current rises, more voltage is lost in R3. So the emitter is more negative. This lowers the base-to-emitter voltage and reduces current. C1 by-passes R3 at working frequency.

Changes in base current from the signal source cause changes in collector current. These develop a signal voltage across the load resistor R4. C2 has low reactance at audio frequencies, so signals arrive at the base of the second transistor TR2. Because of C2 the circuit is called 'capacitor coupled'. TR2 gives further amplification.

In Figure 130, potentials on TR1 collector and TR2 base are arranged to be the same, so that they can be directly connected. This is 'direct coupling'.

Figure 131 has three stages of amplification, all directly coupled. The collector voltages on TR1 and TR2 will be low, but can be sufficient for good results.

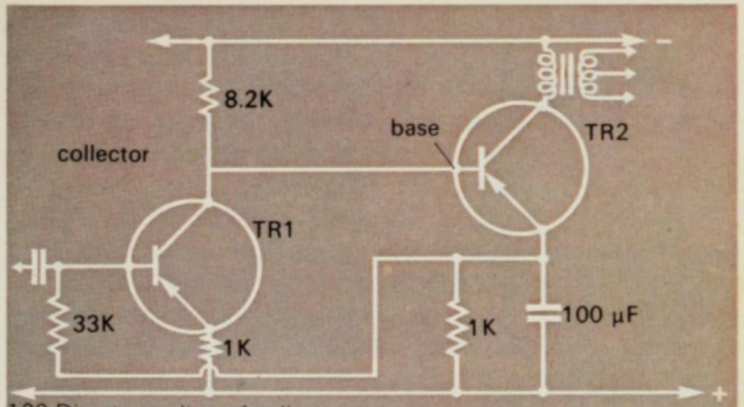

130 Direct coupling of collector to base

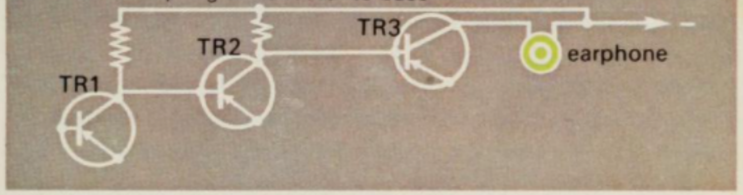

131. Direct coupling for miniature hearing aid

Car Radio

Figure 132 is a typical car radio amplifier. The output transistor needs a large current, but can work directly from a 12 volt battery. Only one transistor is present. It must amplify the whole audio cycle. This is termed Class A operation.

Speaker matching is by a tapped winding, instead of a transformer with separate primary and secondary. Such a tapped component is an 'auto-transformer'.

132. Car radio Class A output stage

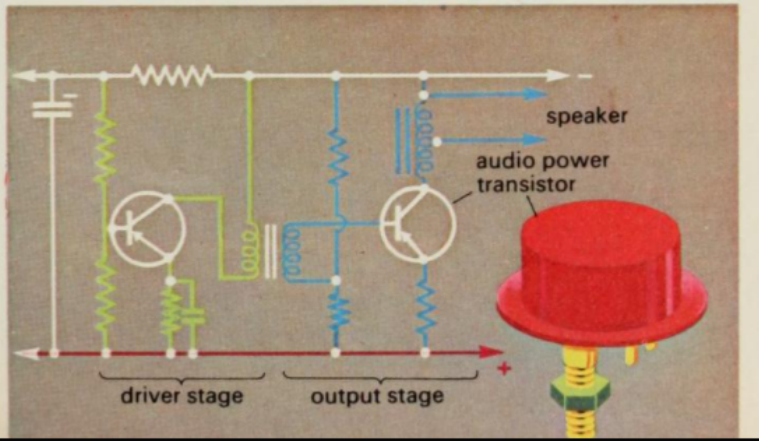

Push-Pull Too

'Class B' working with transistors uses base bias which causes collector current to be lower than with 'Class A'. Figure 133 has resistors R1 and R2 to supply the bases of TR1 and TR2 through the transformer secondary. When audio signals drive TR1 into the conducting region for one half-cycle, collector current rises. For the other half-cycle TR2 is made to conduct [134]. This is the same as using electron tubes in push-pull. Battery drain is small at low volume, or when no signal is present, but rises to allow adequate loudspeaker volume when required.

Figure 135 shows an arrangement often found in miniature transistor receivers. The transformer has two secondaries, each driving one transistor. The transistors conduct on alternate halves of the signal cycle. No output transformer is needed, and a miniature 8 ohm or other speaker can be fed directly.

Stereophonic Amplifier

A stereo amplifier [136] has two channels. Volume control VR1 takes off the required signal level from a stereo pick-up, which supplies right and left outputs. Signals pass through R1 and R2 to V1A. V1A and V1B are assemblies of a triode-tetrode. R4 is the plate load, and C3 couples to the control grid of V1B. R5 and R8 with by-pass capacitors C2 and C5 provide cathode bias. R7 and C4 drop and smooth the supply for V1A. Transformer T couples to the speaker. Higher frequencies pass back through C6 and R9 to the junction of R2 and R3, to provide selective feedback controlled by VR2, for adjusting treble.

The left channel output of the pick-up goes to a second identical amplifier, and to the left speaker. A common power pack supplies both amplifiers, and the equipment is generally assembled on a single chassis.

During the recording, right and left channels depend on strength according to the position of the musical instruments or performers, and so a person listening to the speakers in Figure 136 obtains a sensation of depth. (See also page 104). A similar effect is obtained when listening to stereo broadcasts. Special means are used to provide right and left channels.

133 Transistor TR1 base driven so TR1 conducts

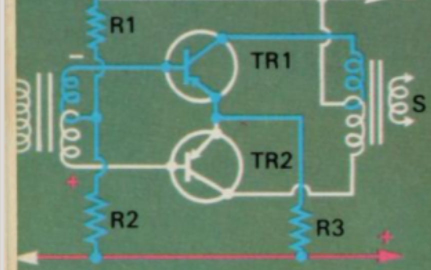

134 Transistor TR2 base driven so TR2 conducts

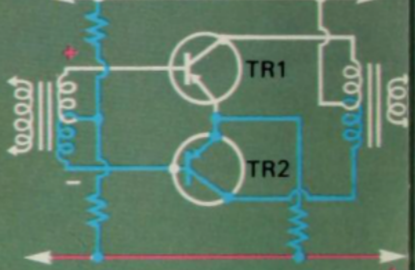

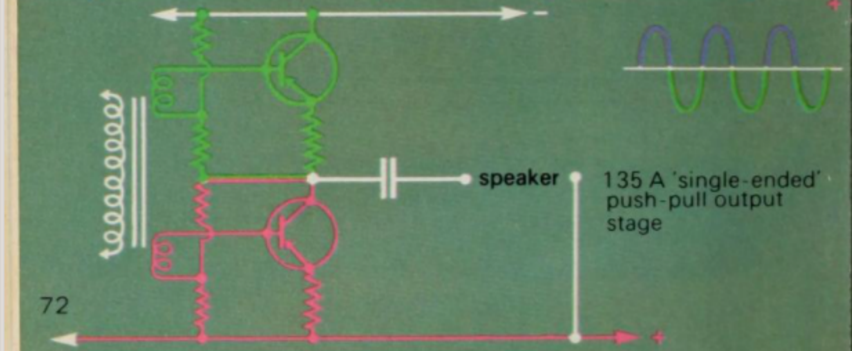

135 A 'single-ended' push-pull output stage

136 Amplifier and system for stereo reproduction

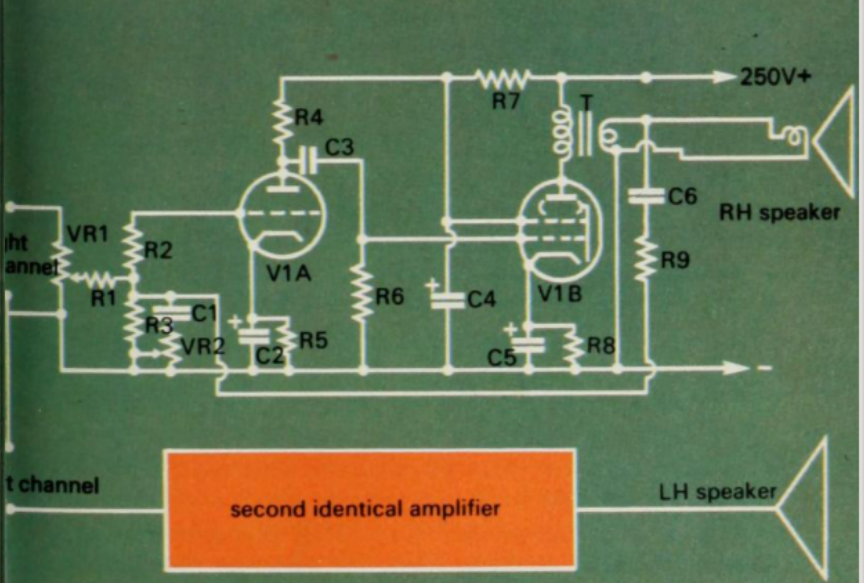

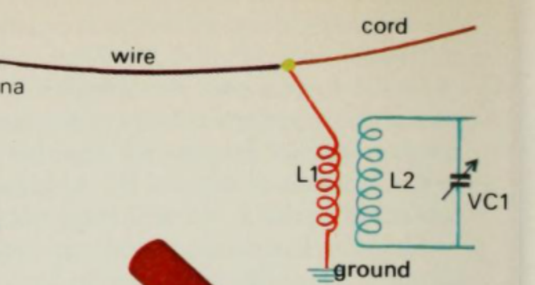

137. Long wire antenna

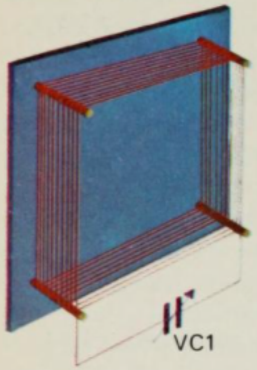

138. Frame antenna

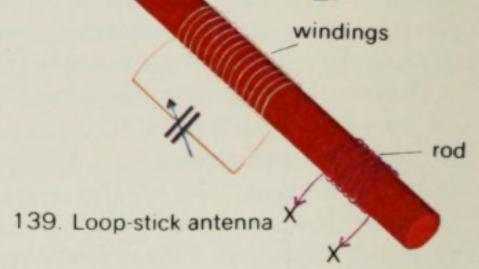

139. Loop-stick antenna

How a Radio Receiver Works

Radio waves travel 300 million meters a second, with the electromagnetic field and the electrostatic field at right angles. We speak of 'wave length' or 'frequency'. If the wave changes from positive to negative and back 300,000 times a second, there will be 1,000 meters from the crest of one wave to the next. So a wave length of 1,000 meters equals a frequency of 300,000 cycles per second (c/s).

1,000c/s = 1 kilocycle, 1kc/s, or 1 kilo Hertz

1,000m = 300kc/s

1,000kc/s = 1 megacycle, 1Mc/s, or 1MHz.

A long wire can intercept part of the radio wave [37]. A signal voltage across the coil L1 induces a voltage in the coil L2, which is tuned by the variable capacitor VC1 to select the desired frequency.

Portable receivers may use a frame antenna [138]. The wave induces a signal voltage in the winding, which is tuned to the desired station by VC1. Small portables generally have a loop-stick antenna [139]. A winding is tuned by VC1. A smaller coupling winding X-X supplies a transistor.

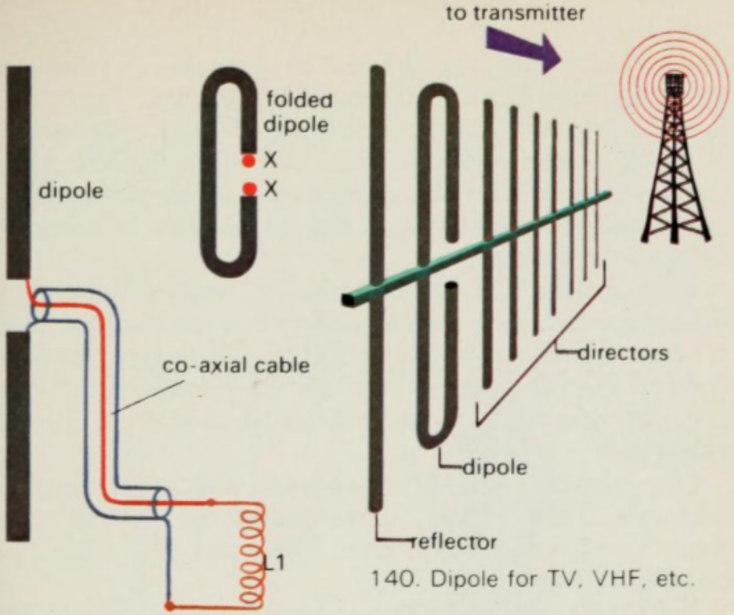

140. Dipole for TV, VHF, etc.

Frame and loop-stick antennas are directive. They pick up signals best when an imaginary axis through the windings is at right angles to the direction of the transmitter. Pick-up is least when the axis points to the transmitter. This is used in direction-finding equipment. (See also page 122).

Television and Very High Frequency

Television and very high frequency (VHF) transmissions are of very short wave length, and so a dipole can be used [140]. It is about the same length as a half-wave. It is cut at the center and a 75 ohm co-axial or twin feeder connected. This carries signal currents to the coil L1.

A slightly longer element or 'reflector' can be put behind the dipole. This increases signal strength. Shorter elements or 'directors' may be positioned in front of the dipole. The elements may be arranged like an H, or an X.

Adding a reflector and directors changes the center impedance of the dipole, making it low. Therefore it is folded and the feeder is attached to X-X. The antenna has its elements placed vertically and horizontally to match the 'polarization' of the television or VHF signal.

Resonance

Radio transmitters use different frequencies. To choose a signal we tune a circuit to 'resonance' with it. Figure 141 is a parallel-tuned circuit. The 'resonant frequency' is that at which a signal voltage is developed across the coil. Altering the capacity VC1 changes the resonant frequency. VC1 is operated by a drive having a dial marked with frequencies, stations, or wavelengths.

With a fixed coil L1, moving VC1 from minimum to maximum capacity covers a band of frequencies. One band is about 1,500kc to 600kc, or 200 to 500 meters (medium waves). For other wavebands, we select coils of different size and inductance. Figure 142 has S1 to select L1 or L2. Other coils can be present.

Often the coils have other windings, such as the coupling primaries. Then section S2 can choose these at the same time. A receiver may have several coils and a switch with a number of sections.

We may tune two or more circuits at once. Figure 143 has the antenna coil with primary P1 and secondary S1, tuned by VC1. Amplified signals go to Y and to the primary P2. The secondary S2 is tuned by VC2. The ganged capacitor VC1—VC2 has two sets of plates with a single shank. In some receivers, VC1 and VC2 both tune the same frequency—that of the transmission wanted. Two tuned circuits make the receiver more 'selective': we can better choose or select one station from others on near frequencies.

141. Parallel tuned current

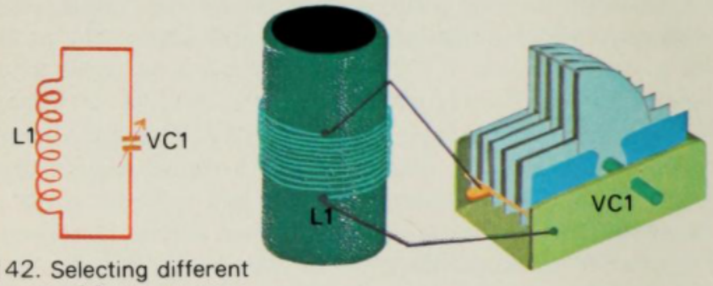

142. Selecting different bands of frequency

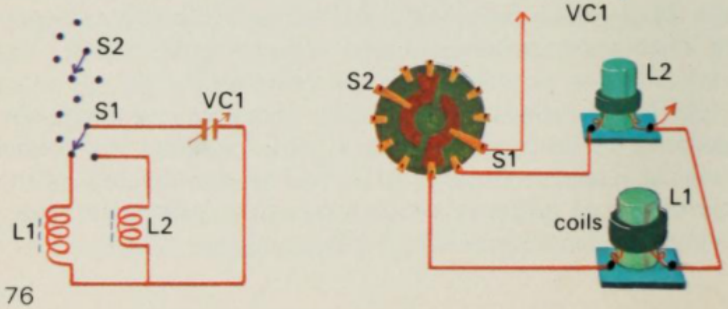

143 Ganged tuning

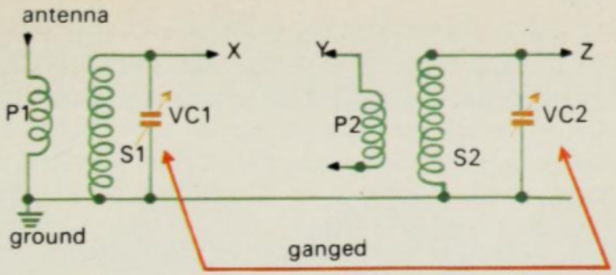

144. Adjustable trimmers and cores

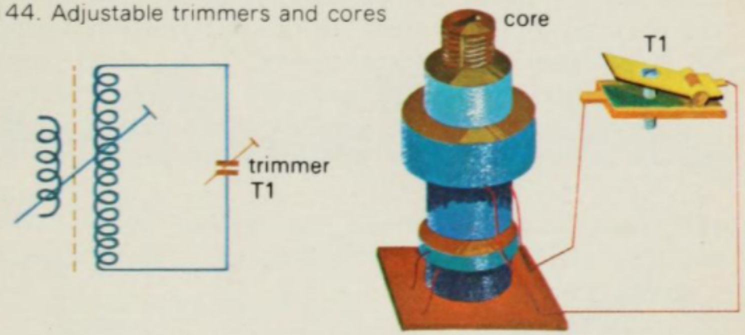

In superhet receivers, VC2 tunes to a different frequency. By a process called heterodyning or beating all stations produce a signal of the same frequency. Later circuits can be permanently tuned to this new frequency.

Often each coil has an adjustable core [144]. Its inductance can be matched with other coils. Coils generally have a trimmer T1. This compensates for stray circuit capacity in wiring, tubes, transistors, etc. The technician adjusts cores and trimmers so that the circuits 'gang' as exactly as possible together.

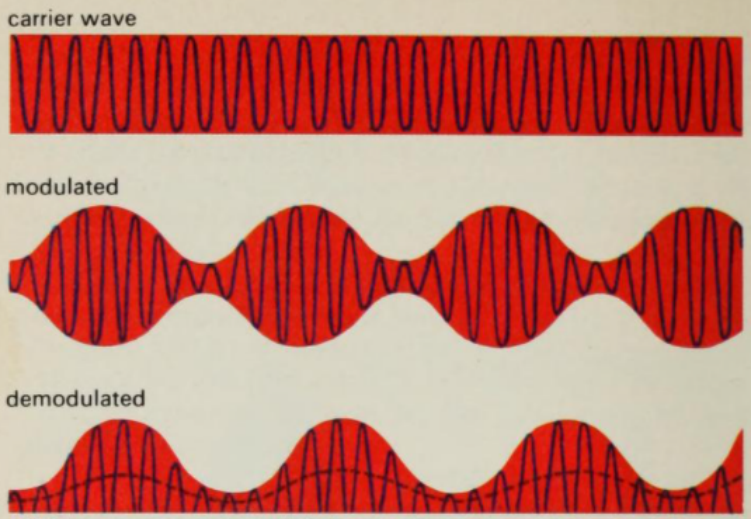

145 The process of demodulation

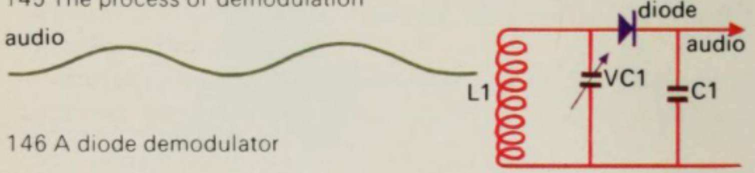

146 A diode demodulator

Demodulation

Putting the audible program into the radio wave is called 'modulation'. Recovering the sound from the signal is 'demodulation' or 'detection'. A radio wave with no modulation is of uniform strength [145]. Many broadcasts use 'amplitude' modulation. A radio wave modulated with a single tone is represented, but for speech and music the shape is very complicated. The modulated wave reaches a diode in the receiver. This lets current pass one way only, and so removes one-half of the modulated waves, as shown in the diagram.

The peaks of the radio signals are smoothed by later circuits, leaving a copy of the audio signal used to modulate the transmitted signal. In Figure 146, L1 is tuned to the desired station by VC1. Signal voltages resemble the 'modulated' wave. The diode produces the 'demodulated' wave. Capacitor C1 helps smooth the radio frequency peaks to produce the audio output. This could be heard with headphones, but as a rule will pass to an amplifier. When the signals are strong, enough output is obtained from the diode to work headphones [147]. This is the simple crystal radio receiver, similar in performance to the crystal sets so popular in the early days of radio. Semiconductor diodes were not then made, but an efficient detector consisted of a crystal, and cat's-whisker of thin wire. This resembled the point contact diode.

In tube receivers, a thermionic diode often provides demodulation. The heated cathode emits electrons, and the diode conducts from cathode to plate only [148]. The audio signal is developed across R1, and taken from C2 to an amplifier. L1 is usually the secondary of an 'intermediate frequency transformer'. The diode is generally part of a diode-triode or duo-diode-triode tube. The triode section is used for audio amplification. C1 is a radio frequency by-pass capacitor.

Nearly all receivers are 'superhets'—abbreviated from 'super-heterodyne'. In a superhet, an oscillator generates

147 A crystal diode receiver

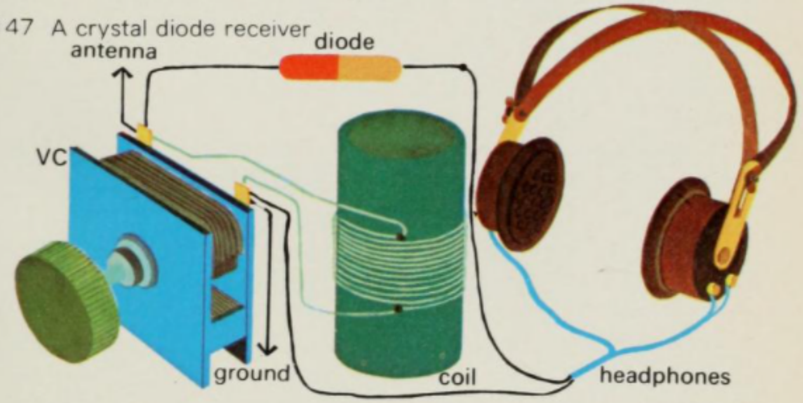

148. Thermionic tube diode as demodulator

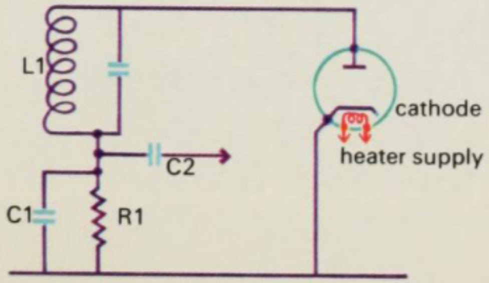

a signal which is made to beat with, or heterodyne received signals. This beating produces a fixed intermediate frequency, which passes to an amplifier with fixed tuned circuits. The frequencies are above audibility.

In Figure 149 the station desired is tuned in by L1 and VC1. L2 is an oscillator coil, tuned by VC2. If two frequencies are mixed, another frequency produced is equal to the difference. Suppose L1 is tuned to 200kc/s and L2 to 670kc/s. The difference is 470kc/s. Should L1 be adjusted to 1,400kc/s, and L2 to 1,870kc/s, the difference is still 470kc/s. By arranging that the *difference* between L1 and L2 frequencies remains unchanged, the same frequency output is obtained.

Figure 150 is a transistor mixer. L1 is tuned to the desired transmission. L2 transfers the signals to the base B. L3 is the oscillator coil, with windings connected to collector C and emitter E. R3 is for emitter bias.

If L1 tunes from 500kc/s to 1,500kc/s for medium waves, and the later stages are to operate at 470kc/s, then L3 tunes from 970kc/s to 1,970kc/s. The *ratio* of frequencies covered by L3 (970:1,970) is smaller than frequencies covered by L1 (500: 1,500) so VC2 is smaller, or has a capacitor C in series.

The desired frequency from the mixer is the 'intermediate frequency'. It is often around 455kc/s to 470kc/s, but much higher in receivers for high frequencies.

Switching can select other coils, for more wavebands [151]. Each coil can have its own trimmer and adjustable core.

Some receivers have a radio frequency amplifier before the mixer [152]. This is tuned to the desired station frequency ('signal frequency'). Tube receivers operate in a similar way, with the circuit arranged for tubes.

When signals have been converted to a chosen fixed frequency, they pass to the 'intermediate frequency' amplifier. Signals have been converted from radio frequency to this intermediate frequency. Later they will become a lower, audio frequency.

149. What a superhet mixer does

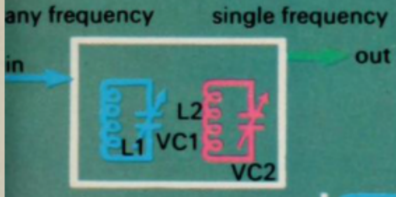

150. Transistor mixer

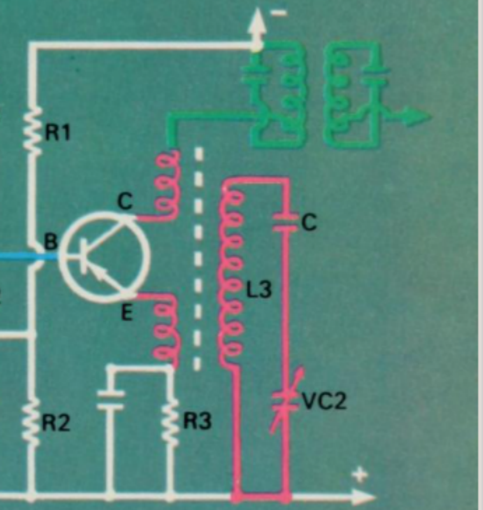

151. Switching for other wavebands

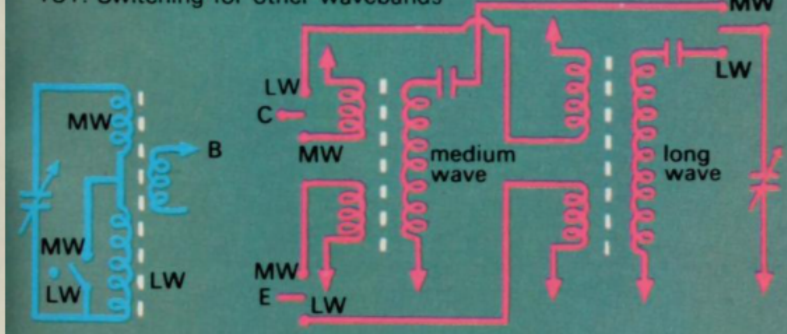

152. Radio frequency amplifier

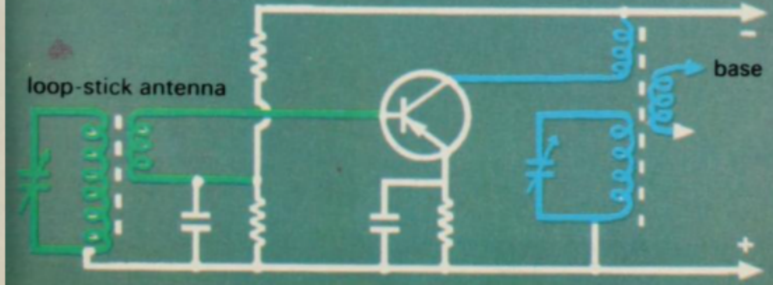

IF Amplifier

The intermediate frequency transformer IFT1 [153] is permanently tuned to the intermediate frequency. After amplification by transistor TR1, signals pass to IFT2, then to transistor TR2. Each transistor has base supplies at X and Y. IFT3 feeds the demodulator diode D1. Audio signals are obtained across the volume control VR1. The sliding contact takes off audio signals as required.

Transformers, used to couple tubes, transistors, or other circuits, generally have two windings [154]. Each is permanently tuned by a fixed capacitor C. The cores are screwed in or out, allowing changes in inductance and precise adjustment to the correct frequency. Often the diode winding is not tuned [IFT3, Figure 153]. In some receivers, none of the secondaries is tuned.

A transistor can provide demodulation [155]. Audio signals are taken off through capacitor C. Thermionic diodes, triodes, and other tubes can be used as detectors. A semiconductor diode is very popular.

Figure 156 shows the stages of a receiver. The mixer converts all incoming signals to one frequency. The intermediate frequency amplifier tuned circuits help eliminate possible interference from other transmissions, and also raise the signal strength. The detector recovers the audio signal. But this signal is not yet strong enough to operate a loudspeaker and must be raised in level by the audio amplifier. The output stage or last amplifier increases the power even more, and signal strength becomes great enough to drive the loudspeaker. The same sequence of stages is often used in a tube receiver.

153 Intermediate frequency amplifier

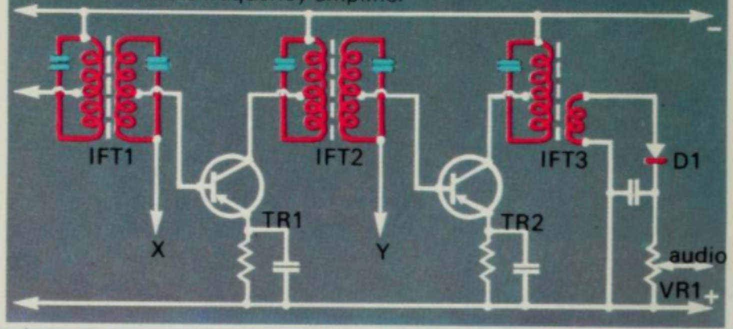

154 Intermediate frequency transformer

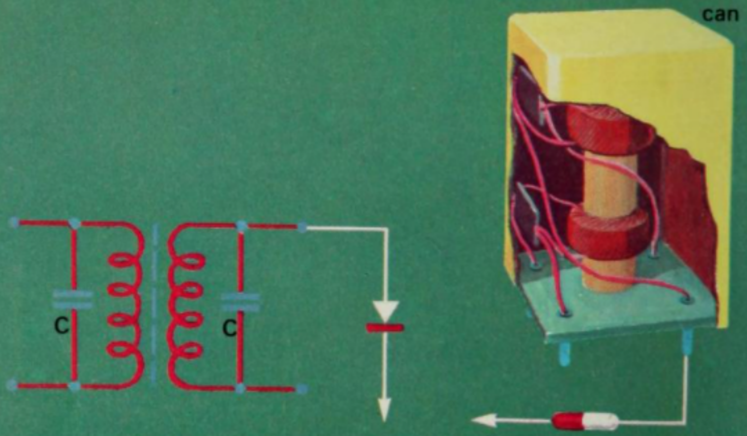

155. Transistor detector

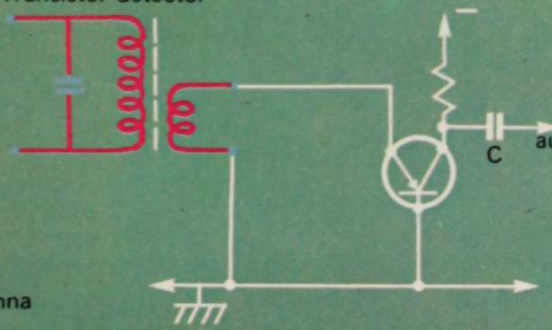

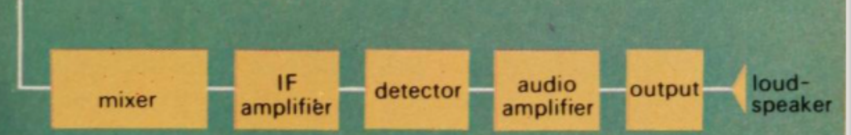

156. Block diagram of a superhet receiver

How a Television Receiver Works

The television station transmits two signals—one for audio and one for video. It would be possible to use separate receivers and antennas, one for audio and the other for picture reception [157].

But sound and picture frequencies are near each other, so we may employ the same antenna. We can also build the equipment as a single receiver, and employ some sections for both audio and video [158]. A radio frequency amplifier increases the strength of both audio and video signals, which are then separated. A common power pack provides current for both sound and picture sections.

TV Sound

The sound program may be provided by amplitude modulation of the carrier wave. The apparent strength of the wave changes with the audio signals—speech or music, etc. A detector recovers the sound, as with radio.

Sound can also be provided by frequency modulation. The amplitude of the wave remains the same, but its frequency is changed to carry audio signals. The sound is recovered by a 'ratio detector' or other frequency modulation detector. FM is used for television sound.

157. Television—audio and video

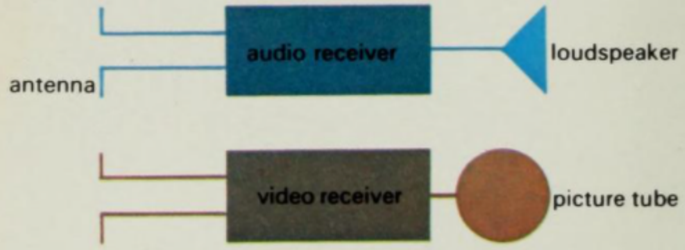

158. Circuits combined in a TV receiver

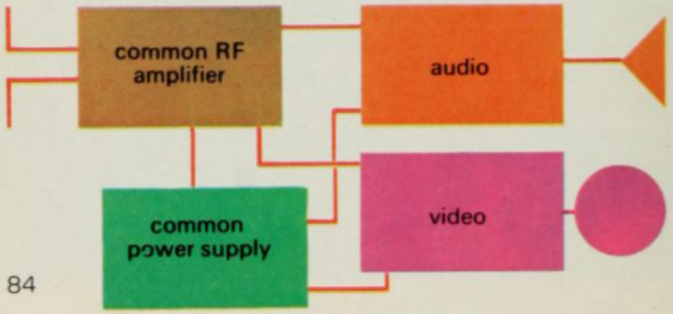

159. Some TV channels showing adjacent sound and picture frequencies

Channel	Megacycles
2	54-60
3	60-66
4	66-72
5	76-82
6	82-88
7	174-180
8	180-186
9	186-192
10	192-198
11	198-204
12	204-210
13	210-216
UHF	470-890

160. Antennas are made for the channels required

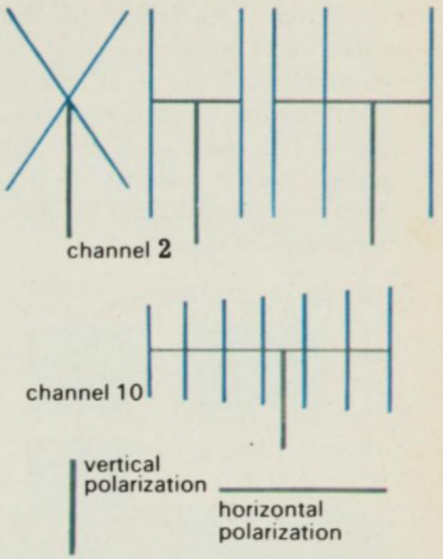

Frequencies

The frequency or wavelength determines the channel, just as in sound broadcasting [159]. As already mentioned, each station has two adjacent frequencies—one for sound, and one for video. We usually select the station by some form of rotary or other preset tuning arrangement, to which a variable fine tuner may be added.

The antenna is arranged to suit the channel we want, or we may have a combined antenna for two or more frequencies. The receiver antenna elements must be vertical or horizontal, in order to suit the 'polarization' of the signals [160], depending on the transmitter.

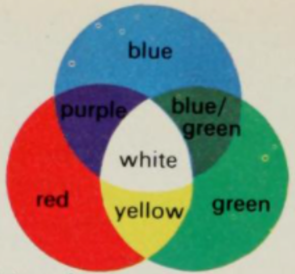
blue
purple
blue/
green
white
red
yellow
green

177. Additive mixing of colors

178. A method of obtaining color signals for transmission

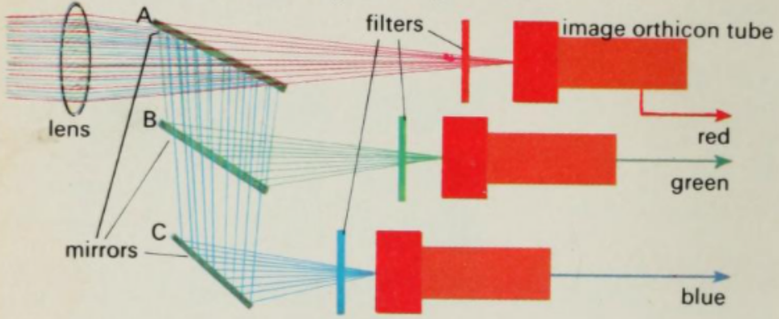
A
filters
image orthicon tube
lens
B
red
green
C
mirrors
blue

Color Television

Colors other than black can be made up by mixing three chosen colors. The additive combination of light differs from results obtained with pigments. Absence of all colors forms black. Red and green together provide yellow. Blue and red give purple, and blue and green provide blue-green [177]. When red, blue and green are present in an intensity of 40 percent, 26 percent and 34 percent, combining them provides white.

To transmit a color picture, the scene is scanned by cameras sensitive to three colors, which are reproduced in the receiver. Other colors are reproduced by changing the intensity of the red, blue and green channels.

Figure 178 is a color camera system. Mirror A lets red pass, but reflects blue and green. B reflects green but transmits blue, which is reflected by C. In this way the scene scanned is separated into the three colors, each providing its own signal. Other arrangements of selective mirrors may be used in order to obtain the desired result. The picture could be reproduced by three projection tubes, one each for red, green and blue, and each receiving the red, green or blue signal [179].

A shadow mask color tube has three separate cathode assemblies [180]. Each emits a ray controlled by the red, green or blue channels. A mask with many small holes is fitted inside the tube. The cathode rays come from slightly different directions, and after passing through the same hole strike different spots on the screen. The screen is coated so that rays from the blue channel gun strike areas producing blue. The cathode ray responsible for green strikes areas which produce green, and in the same way the red-producing ray strikes areas which cause red. The different color spots are so small that the eye blends them together into the various desired shades. Deflection coils move the three cathode rays over the tube face together.

179. Color receiving system with three tubes

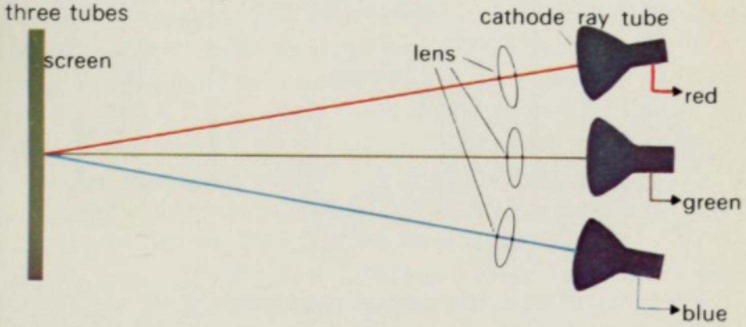
cathode ray tube
screen
lens
red
green
blue

180. Color tube

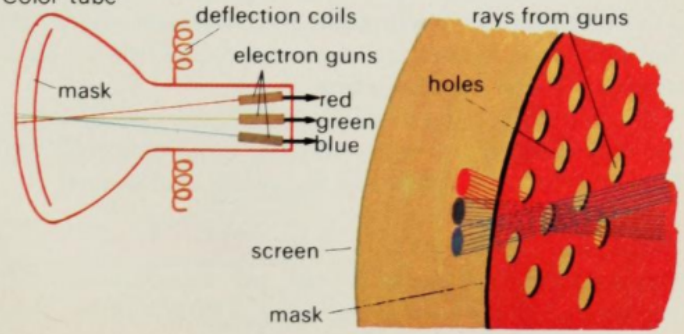
deflection coils
rays from guns
electron guns
mask
holes
red
green
blue
screen
mask

Recording on Tape, and Disc Stereo

Tape recordings are much used for video or television picture recording, input and storage of computer information, and other purposes.

For sound recording, the signal level is raised with an audio amplifier [181], and taken to a recording head that has a winding and magnetic core with a narrow gap. Signals in the winding cause a fluctuating magnetic field across this gap.

The tape is polyvinyl chloride or other insulating material, coated on one side with tiny particles of iron oxide. It is often ¼ in. wide for sound, and 2 ins. for video.

181. Amplification and recording head

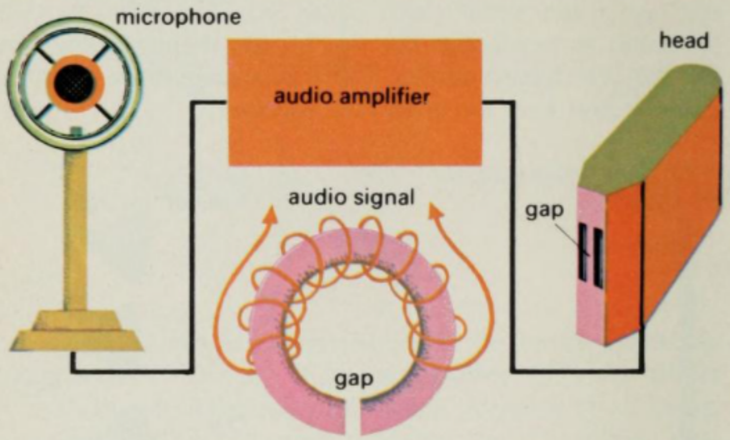
microphone
audio amplifier
head
audio signal
gap
gap

The unused tape has magnetic particles lying at random [182]. A mechanism draws the tape smoothly by the head gap. The particles are brought into magnetic alignment, and the tape has magnetized areas corresponding to the audio currents in the head winding.

For longer playing time, about half the track is used with the tape passing one way, and the remaining half with the tape reversed. Quarter-track recording is also used.

Frequency Response

Suppose the tape moves 15 ins. per second and a 15 cycle per second audio tone is recorded. There will be 1 in. tape length from one cycle peak to the next. If the tape moves at 7½ ips the distance would fall to ½ in. For slower tape speeds, the distance between peaks becomes smaller [183]. Raising the audio frequency also brings the magnetized sections closer to each other.

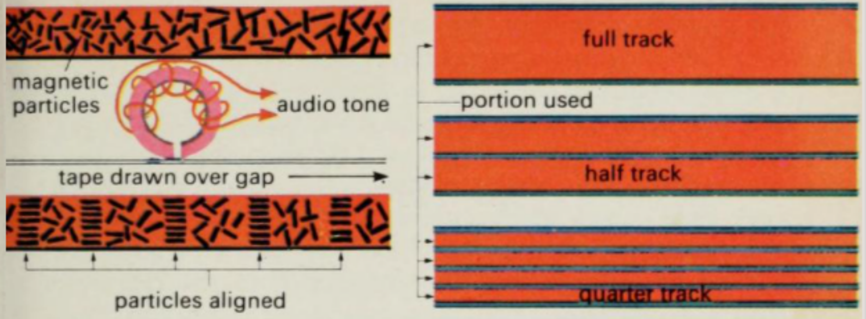
magnetic particles
audio tone
tape drawn over gap
particles aligned
full track
portion used
half track
quarter track

182. Magnetic recording on tape

A limit arises when the magnetic 'domain' length from one cycle to the next is smaller than the gap of the head. Then results soon fall off. To help avoid this limit, head gaps are very small—as small as 0.00025 in.

For music, or when a response up to 10,000cps or so will be wanted, the tape may move at 15 ips, while 7½ ips and 3¾ in. per second speeds suit less exacting needs, and 1⅞ ins. per second or slower movement allows satisfactory recording of speech, and gives long playing time with small reels. Two or more tracks also give longer time.

The tape is drawn by a 'capstan', a revolving roller. It may have a heavy flywheel and belt drive [184].

183. Tape speed and frequency response

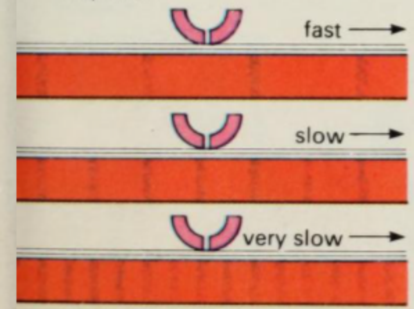
fast
slow
very slow

184. Tape driving mechanism

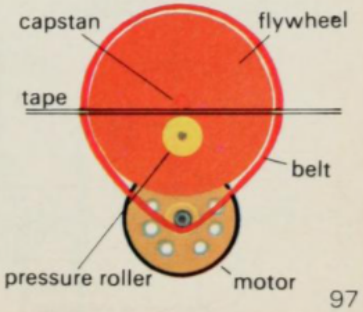
capstan
flywheel
tape
belt
pressure roller
motor

Playback

To hear the recording, the tape is wound on its original reel, then drawn past a playback head at the same speed as when recorded [185]. The magnetized tape causes a small magnetic field in the head core, thus generating an electrical output. When recording, audio currents produce a varying magnetic field; during playback, the varying of the magnetic field causes audio currents.

In many spheres, such as broadcasting, news gathering and so on, signals may be recorded from sources other than a microphone (e.g., from other tape recordings, from discs or from radio programs). Signals can also be recorded from television so that playback supplies a picture. Equipment for this is complicated.

For house current use, recorders can employ about four stages, resistance capacity coupled [186]. For better volume and fidelity, push-pull output may be provided.

Record/Playback

The same head may be used to record on the tape, and obtain signals when playing the recording. Switching of input and output circuits allows the same amplifier to be employed for both purposes [187]. With the switch at R (Record) the microphone provides the amplifier input.

For playback (P) the head is switched to the amplifier input, and the output to a loudspeaker. There are two 'outputs', because that for the head is taken from an earlier stage—the full power of a large amplifier is not used while recording.

Recording Level Indicator

An indicator shows us that recording is at a suitable level. In Figure 188, X goes to an audio amplifier. The diode rectifies audio signals, building up a negative voltage. This reduces plate current, so the deflector plates become more positive. Electrons from the cathode strike a target, and the beam width is controlled by the deflector plate potential. Hence the display depends on the average audio power, and we adjust volume for correct recording.

If we apply a high frequency 'bias' oscillation to the head while recording, results are improved. The power output stage can be utilized as a bias oscillator. In Figure 189, C1

185. Playback of the recording

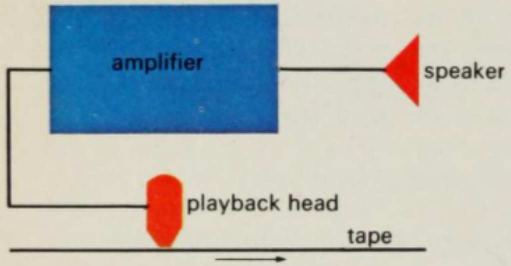

186. Vacuum tube playback

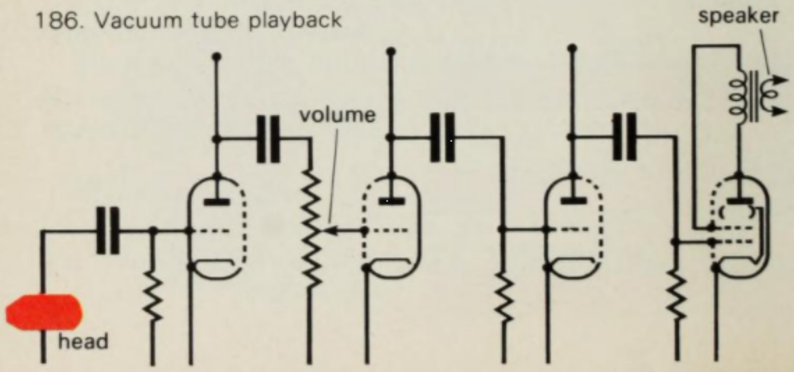

187. Recorder/playback switching

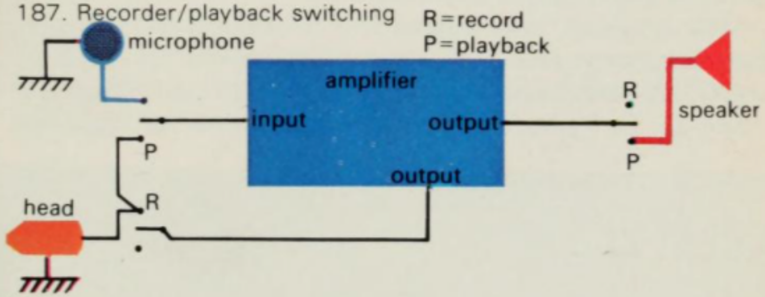

188. Recording level indicator

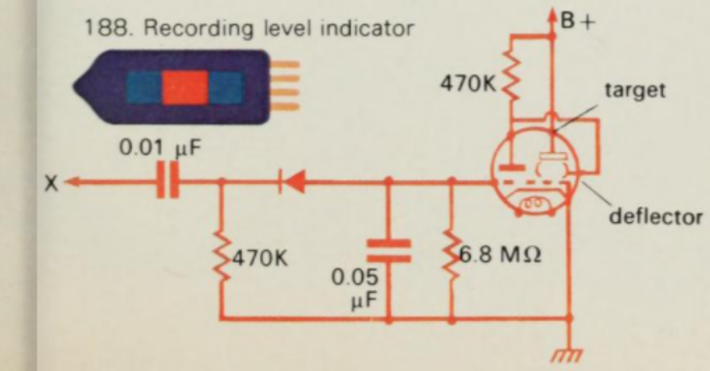

OTHER APPLICATIONS OF ELECTRONICS

Counting Electronically

Counting circuits record the number of objects produced by a machine, the number of oscillations each second, or speed, or distance. They also show the results of addition, and other operations.

Mechanical Counter

An odometer shows miles covered. An electricity meter counter shows units consumed. Seeing what actually happens mechanically is useful.

Revolving mechanical counters are often used [201]. Shaft A rotates to display numbers up to 9. When it has rotated from zero to the 10th unit, tooth X rotates the wheel on shaft B one place, moving the tens counter from 0 to 1. This happens with each revolution of A. One complete turn of B rotates the hundreds counter one place, and 99 revolutions of shaft A, counting up to 999, can be accommodated.

A solenoid, pawl and toothed wheel could rotate shaft A. An electrical circuit closes, current flows, and the pawl moves the counter one place. Such devices are occasionally connected after high-speed electronic counters, to record very large numbers.

What actually happens? As shown in Figure 202, each time the full 'units' capacity is reached, one impulse is transferred to the 'tens' counter and the units return to zero.

201. Mechanical counter

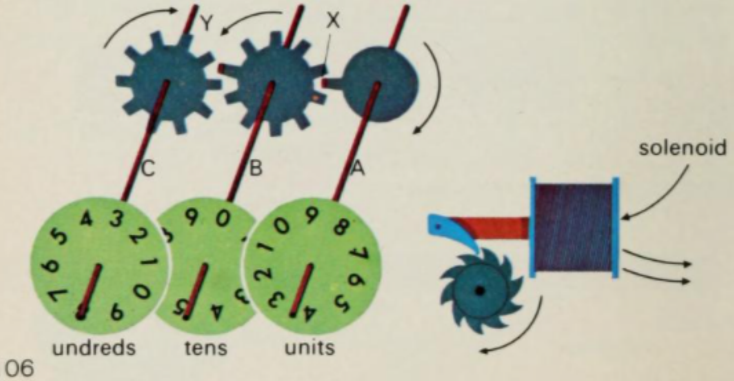
Y
X
C
B
A
solenoid
undreds
tens
units

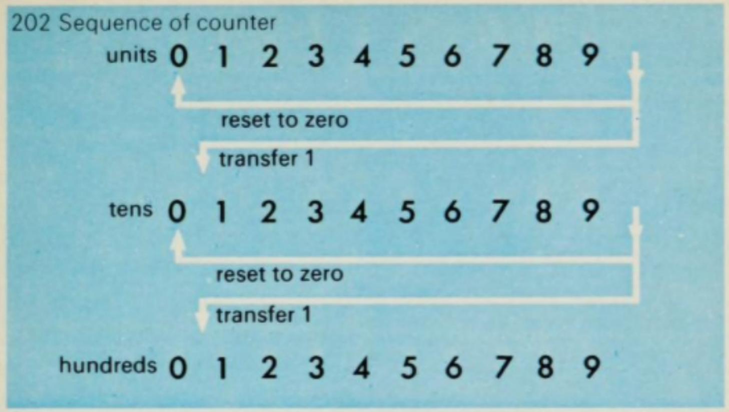
202 Sequence of counter
units 0 1 2 3 4 5 6 7 8 9
reset to zero
transfer 1
tens 0 1 2 3 4 5 6 7 8 9
reset to zero
transfer 1
hundreds 0 1 2 3 4 5 6 7 8 9

Similarly, the full capacity of tens results in one impulse being transferred to the 'hundreds' position.

Electronic circuits giving one output pulse for a number of input pulses similarly form a section of a counter.

Mechanical processes that are repeated can be counted by a closing circuit [203]. The counter is electrically operated. Current from the supply reaches the solenoid each time the circuit closes, moving the units counter. The switch may be operated by weight, or by any means.

Electronic methods perform the same function more reliably, easily and rapidly. Figure 204 has an infra-red light beam falling on a photocell. This causes an electrical output which is amplified. When the beam is interrupted an impulse is passed to the units counter. Every tenth impulse is transferred to the tens counter, and so on.

203. Electro-mechanical counter

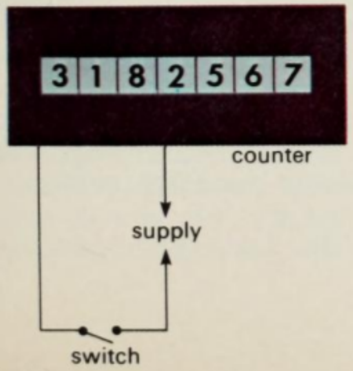
3 1 8 2 5 6 7
counter
supply
switch

204. Electronic counter

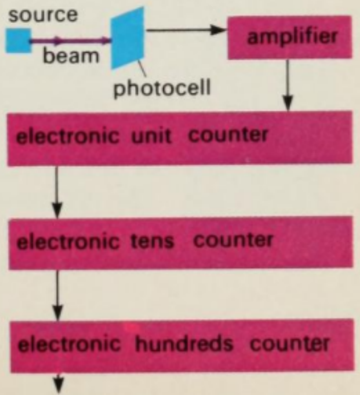
source
beam
photocell
amplifier
electronic unit counter
electronic tens counter
electronic hundreds counter

A ring counter gives one output pulse for a selected number of input pulses—for the decimal system, one output pulse for ten input pulses [205]. We may use controlled rectifiers, thyratrons or other devices which remain stable until 'triggered' by an input pulse, retaining the new state until returned to the original.

Pulses reach the rectifiers through diodes D. Diodes except the first have a large bias voltage from the associated controlled rectifier. So the first pulse triggers only the first rectifier. This one conducts, causing a low voltage across the second diode. Hence the next pulse can trigger the second controlled rectifier, and ten pulses progress along the circuit. The last rectifier gives an output pulse to a 'tens' panel, and back to the input to restore the units circuit to zero. Six panels would count up to 999,999.

205 A ring counter

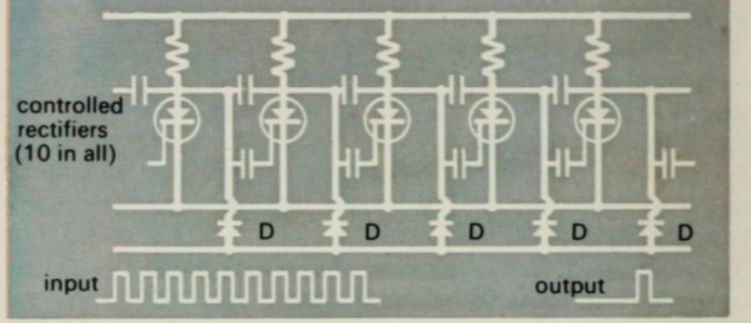

Trochotron and Dekatron

The trochotron has a central cathode, ten grids, ten targets or plates, and ten electrodes [206]. Grids 1, 3, 5, 7 and 9 are connected to one circuit, and 2, 4, 6, 8 and 0 to another circuit. Suppose the electron stream is at 0. An input pulse cuts off the stream, while grid 1 causes the beam to form on target 1. Each subsequent pulse moves the beam around one step. The final position can transfer one pulse to a 'tens' counting tube. The trochotron can count one million a *second*.

The dekatron has ten cathodes. Ionization causes the discharge path to glow [207]. The X pins are joined and fed from pulses to be counted. Pulses reach the Y pins through a delay circuit. When a pulse pin X near 7 deflects the discharge clockwise, a further deflection arises as the delayed pulse reaches adjacent pin Y. When the pulse ceases, the discharge has been stepped to 8. This can be repeated as many times as required.

206 The trochotron has 10 electron stream positions

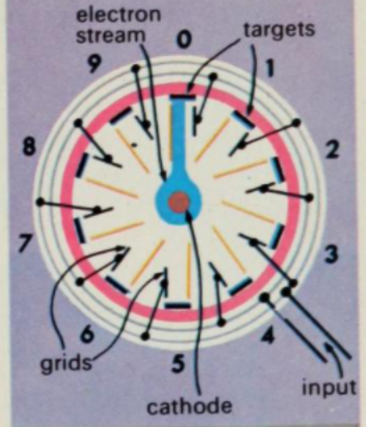

207 Dekatron counter

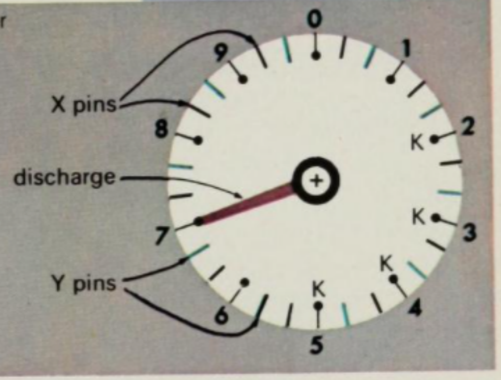

Read-Out Tube

Figure 208 is a tube with electrodes in the shape of the numerals 0 to 9 arranged close together with a common plate. The other ends of the numerals are connected to separate pins. The discharge causes a visible glow around the numeral receiving voltage.

208 A tube to display numerals

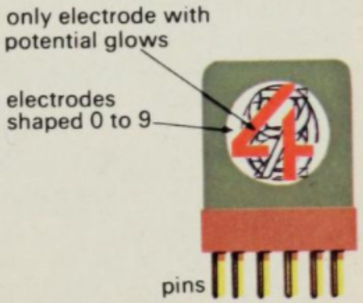

209 Decimal and binary counters

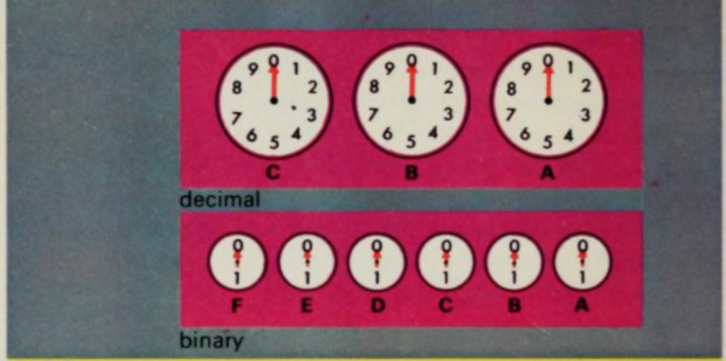
C
B
A
decimal
F
E
D
C
B
A
binary

210 Decimal and binary compared

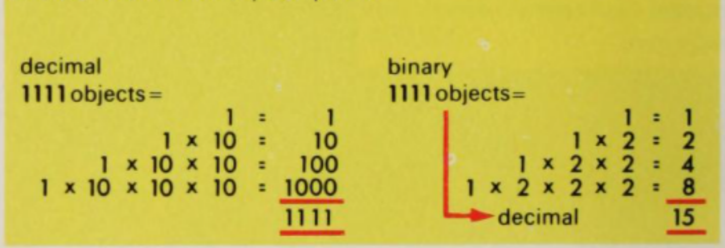
decimal
1111 objects =
1 = 1
1 × 10 = 10
1 × 10 × 10 = 100
1 × 10 × 10 × 10 = 1000
1111
binary
1111 objects =
1 = 1
1 × 2 = 2
1 × 2 × 2 = 4
1 × 2 × 2 × 2 = 8
decimal
15

The Binary System

We need not count in the form '1, 2, 3, 4, 5, 6, 7, 8, 9, 0 and carry 1'. Electronically there are advantages in counting in other ways. The binary system counts by using 0 and 1. Figure 209 shows two counters, one with 0–9 dials and one for binary. They have rotating pointers. We record objects up to 9 on A of the decimal counter. For 10 we see there is no higher number than 9 here. So we set A to zero and transfer 1 to B. We proceed in this way to 999.

With binary the highest number on any dial is 1. When one object arrives, this is put on A, giving 000,001. For object 2 there is no number higher than 1 on A. So we restore A to 0, carrying one to B, to get 000,010. For object 3 we put A to 1 again: 000,011. The next object results in 0 for A and carry 1, 0 and carry one for B, turning C to 1 = 000,100. Similarly, 5 = 000,101; 6 = 000,110; 7 = 000,111; 8 = 001,000; 9 = 001,001, 10 = 001,010, and so on.

Each time we move to the left in decimal counting, we multiply by 10 [210]. Each move to the left in binary multiplies by 2 because there are two possibilities for each column (0 and 1). To register numbers from 0 to 9 in decimal systems, we need *ten* reliable states for an electronic device. For binary, we need *two*—for 0 to 1.

A lack of voltage or current can represent 0, and full voltage or current can show 1. The result is a definite 'on'—'off' system for each column [211].

Computers

Relays close when energized [212]. Early computers used thousands of relays. Vacuum tube plate current is low with strong negative grid bias, but high with little bias or a positive pulse. The PNP transistor does not conduct with a positive signal at the base, but conducts with a negative signal,

211 Switches acting as binary numbers

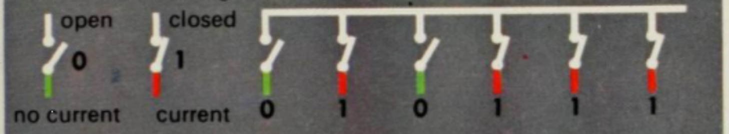
open
closed
0
1
no current
current
0
1
0
1
1
1

212 Relays, valves and transistors as binary numbers

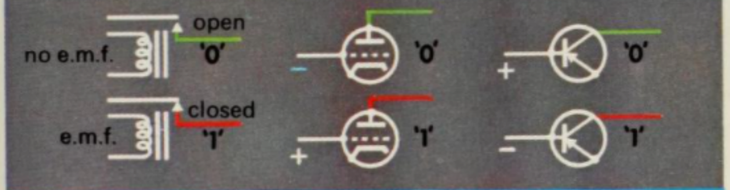
no e.m.f.
open
'0'
'0'
'0'
e.m.f.
closed
'1'
'1'
'1'
+
+
–
–

213 Computers with tape input and output

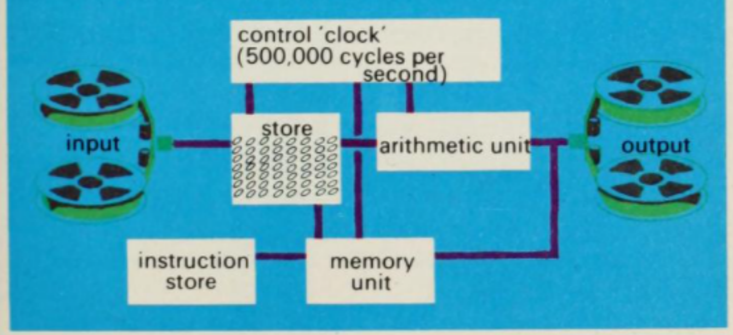
control 'clock'
(500,000 cycles per second)
input
store
arithmetic unit
output
instruction store
memory unit

and operation is only 'on' or 'off'. The computer may take input pulses from a tape, store and process them, and deliver the solution or output to a second tape [213].

Bi-Stable Multivibrator or Flip-Flop Circuit

Figure 214 shows a circuit that will remain in either of two conditions. If transistor A base is negative, transistor A collector current is large, hence collector C1 is positive, and so is base B2 of transistor B. As a result, collector C2 current is small, and there is little voltage drop across R2, and so base B1 is held negative. An input pulse reverses these conditions. Collector C1 no longer passes a large current and the voltage drop in R1 has fallen, shifting base B2 of transistor B in a negative direction. This transistor conducts, so the increased voltage drop across R2 makes collector C2 positive, and therefore base B1. The circuit thus remains in either state, until a pulse is received. Output can be taken from collector C2.

Any device which can be placed in a steady 'on' or 'off' condition can be used for counting. Binary counters divide by 2 [215]. Decade circuits divide by 10. In counting frequencies, an electronic gate (switch) allows signals to pass for a known interval, while they are counted.

Geiger counters help measure radiation intensity. A central plate of high positive potential is used in an envelope of gas mixture [216]. High velocity particles cause ionization. Liberated electrons pass to the positively charged plate, causing small output pulses.

Analog to Digital Conversion

'Analog' information represents a fact in an analogous way. A voltmeter pointer or speedometer pointer position is an analog of voltage or of miles per hour. One means of converting analog to binary digital is shown in Figure 217. A photo-electric eye registers divisions as they pass.

The second method shown has sensing devices opposite the white 0000 sectors. Black sectors are 1. Rotating the disc one step at a time produces the binary count—0001, 0010, 0011, 0100, etc.

214. Bi-stable or flip-flop circuit

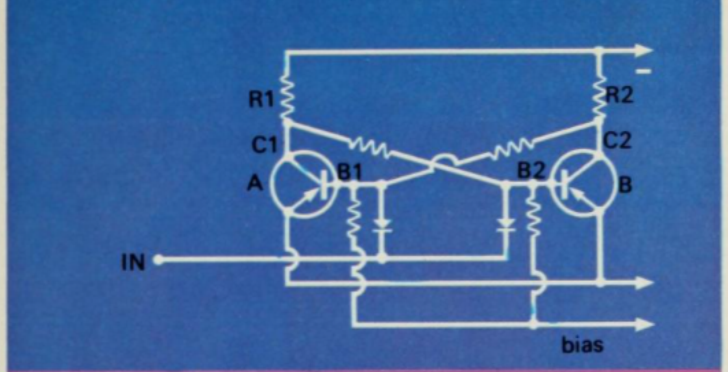

215. Counters in sequence

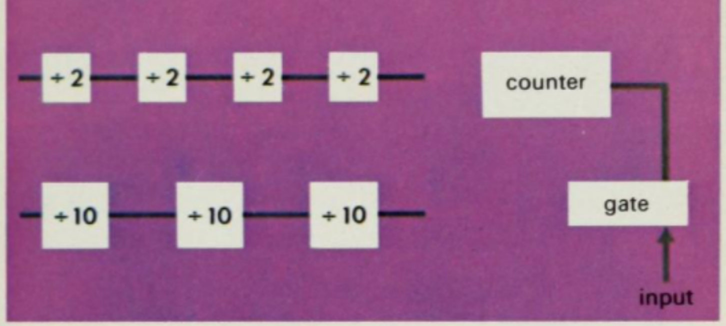

216. A Geiger counter

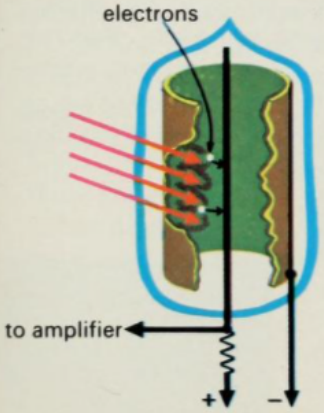

217. Methods of converting analog to digital

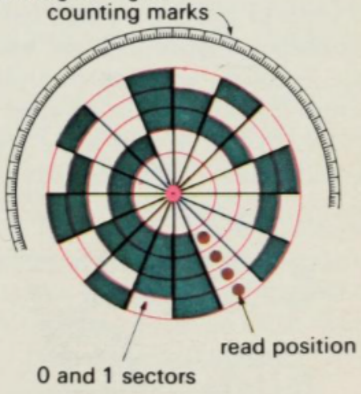

Output-Producing Circuits

Outputs are provided by oscillators, amplifiers, and other circuits, as needed.

If an inductor L and capacitor C are in parallel [218], and a potential momentarily arises in the inductance, the inductor electromagnetic field stores energy. When the field collapses, an e.m.f. charges the capacitor. The capacitor then discharges, returning power to the inductor. Resistance and other losses cause the oscillations to get smaller and finally cease.

The oscillation frequency or 'resonant' frequency in cps

$$\text{is } \frac{1}{2\pi\sqrt{LC}}$$

where L is henrys and C farads.

Resonance occurs when inductive reactance is equal to capacitive reactance.

218. Oscillation in circuit having inductance and capacitance

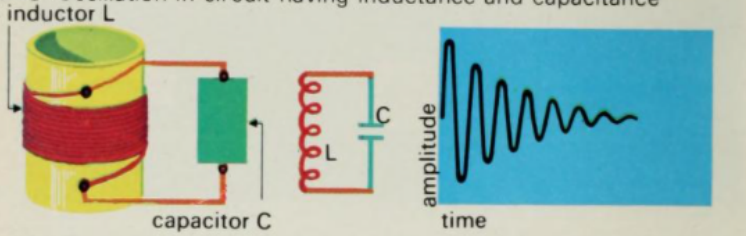

Sustained Oscillation

If we add a device which cancels circuit losses, oscillation continues. Figure 219 has resonant circuit L1 and C. L2 is a coupling winding returning energy to L1. The e.m.f. across L1 reaches the control grid G. An amplified signal arises at plate A, and if L2 feeds back energy to L1 in the correct phase this sustains oscillation.

219. Positive feedback to sustain oscillation

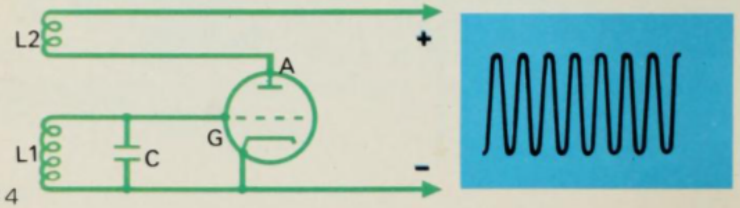

220. Variable frequency oscillator

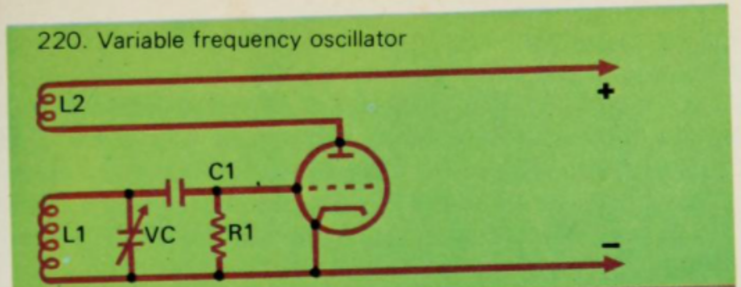

221. Superhet receiver oscillator and mixer

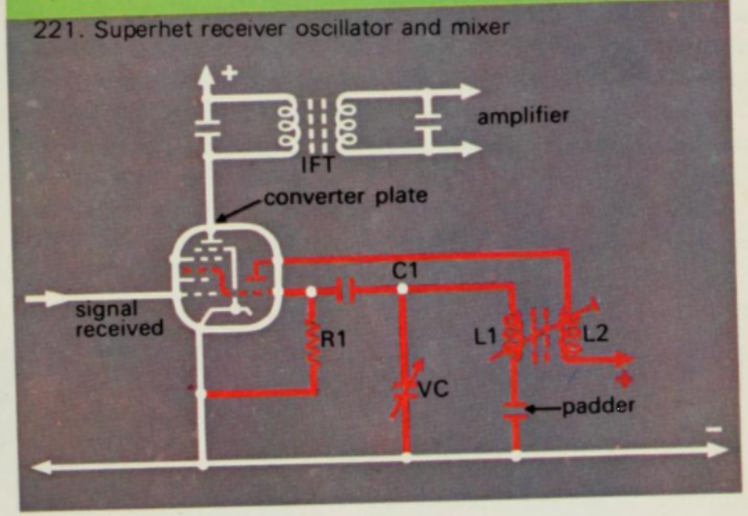

A variable capacitor VC [220] allows tuning over a band of frequencies. The inductor L1 is changed for other wavebands. Rectification between cathode and grid causes a negative charge to appear across C1 and R1, producing stable operation.

A circuit like this is found in most tube radio receivers. The oscillator is the triode part of a converter, or other mixer. Figure 221 has R1, C1, VC, L1 and L2 with the triode section. The triode grid also controls the flow of current to the converter plate. L1 is tuned so that the difference between the received signal and frequency of L1 is always the same. Mixing in the converter results in a signal of fixed frequency appearing at the plate. This output passes to the fixed tuned circuits of the intermediate frequency transformer IFT. The 'padder' helps make the coverage of L1 suitable.

Signal Generator

Oscillating circuits can produce a radio frequency output or radio signal. A signal generator is a source of low-power energy that we can tune to any frequency.

Figure 222 has R1 and R2 to provide suitable base working conditions, with emitter resistor R3. L1 is tuned with VC and has a feedback winding L2 to the base B. The output from the arrangement in Figure 222 is a radio frequency signal without modulation. If it is tuned in on an ordinary receiver, no sound is heard. It is termed an 'unmodulated' or 'carrier wave' output. When we need a wave modulated by an audio tone we can add an audio oscillator [223]. This modulates the RF signal. When the signal is tuned in, the audio tone is heard.

A use for the signal generator is in aligning the intermediate frequency transformers of a receiver [224]. Most IFT's have two tuned circuits, generally with adjustable cores for the coils. Trimmers could be used instead. We set the generator tuning to get the required output—often about 455kc/s. This is fed into the receiver circuit at a point ahead of the IFT circuits to be adjusted. Then we rotate the cores or trimmers with a trimming tool. Best output from the receiver shows that the circuits are suitably tuned.

The antenna and other tuned circuits in a receiver can be adjusted in this way. We set the generator to suitable frequencies, and adjust trimmers and cores so that wavelengths or frequencies are tuned in at the correct positions on the receiver dial.

Figure 225 is the radio frequency section of a signal generator for the 10–300 megacycle range, and is thus suitable for very high frequency alignment or testing.

222. Tunable R-F signal generator

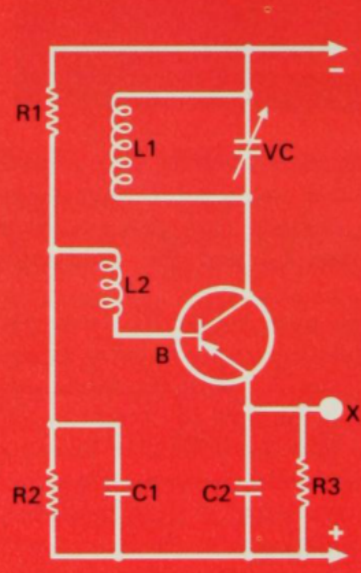

223. Audio modulator

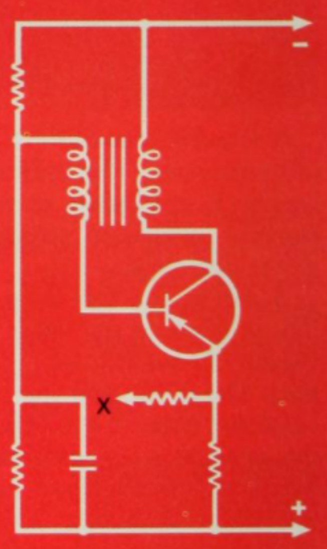

224. Use of signal generator to align receiver

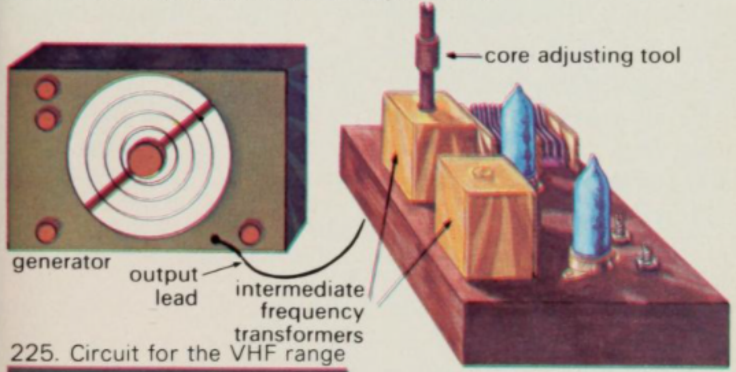

225. Circuit for the VHF range

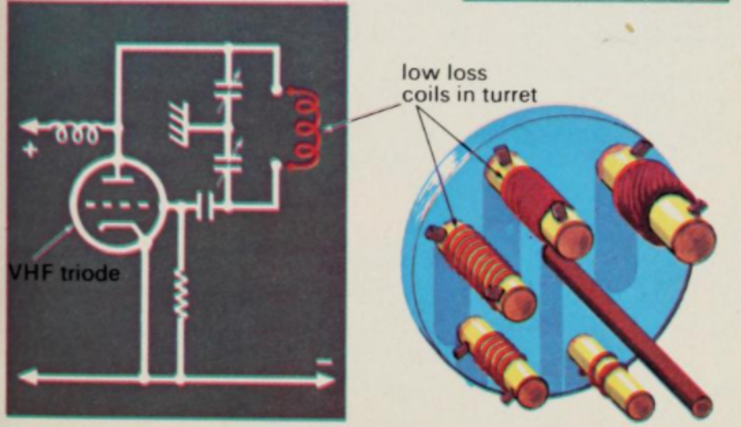

Multivibrator

The 'multivibrator' circuit produces a wide range of outputs, according to working conditions. It uses two vacuum tubes, or two transistors [226]. Transistor A has collector resistor R1 and base resistor R2. B has R3 and R4. C1 couples A collector to the base of B, while C2 couples B back to the base of A. With each transistor driving the other, frequency depends largely on capacitor and resistor values. This circuit can produce outputs for receiver or amplifier testing. The multivibrator will produce both sawtooth and square waves.

Phase-Shift Oscillator

If the voltage of a PNP transistor base is moved positive, collector current falls. Less current flows in the collector resistor, less voltage is dropped here, and the collector potential moves negative. Feedback from collector to base would be in the wrong phase to sustain oscillation—negative when it should be positive, and vice versa. Two transistors, each reversing phase [226], overcome this.

So for a single transistor or vacuum tube we may need some way of reversing phase. Feedback then sustains oscillation. Capacitors and resistors can form a phase-shift circuit for this purpose [227].

Bias Oscillator

Figure 228 obtains oscillation by suitably connecting the primary P and secondary S of the oscillator coil. For example, reversing the connections to S would reverse the phase of feedback to the base. The frequency of oscillation depends on the windings and value of C1. The primary P is tapped for a suitable collector load. Output is taken from another tap through C2 to provide high frequency bias when tape recording.

Electronic Clock

Figure 229 is a circuit producing low-speed output pulses of great accuracy of occurrence. A piezo-electric crystal oscillator runs at set frequency, and this is stepped down by dividers, to obtain the low-speed output pulses used for timing. The apparatus is controlled by the crystal.

226. Multivibrator using two transistors

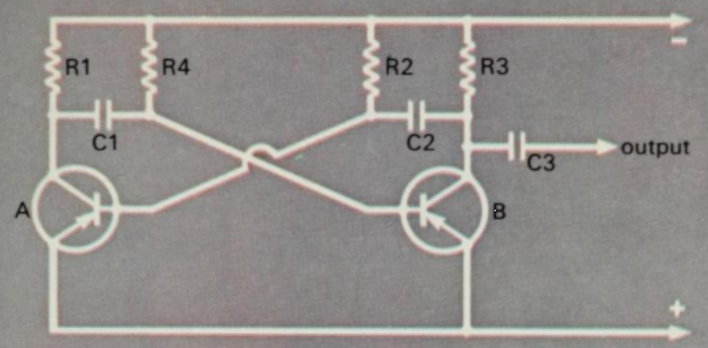

227. Oscillator with phase-shift network

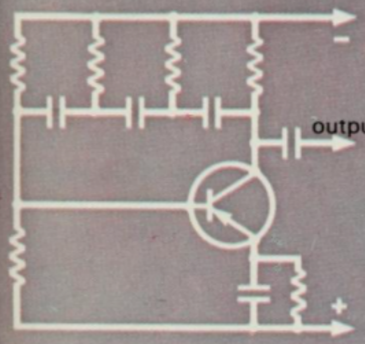

228. Tape recorder high-frequency bias oscillator

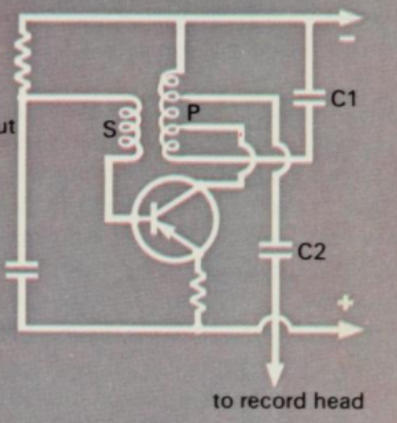

229. An electronic clock

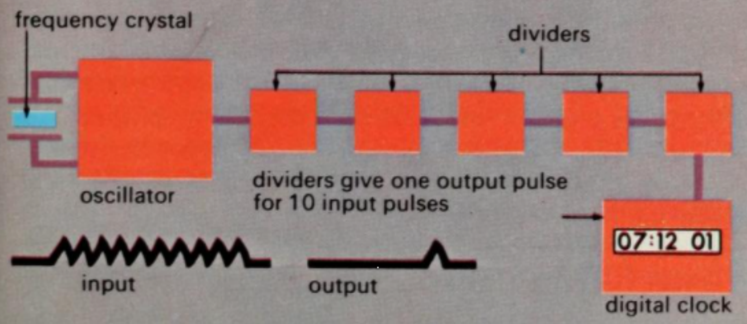

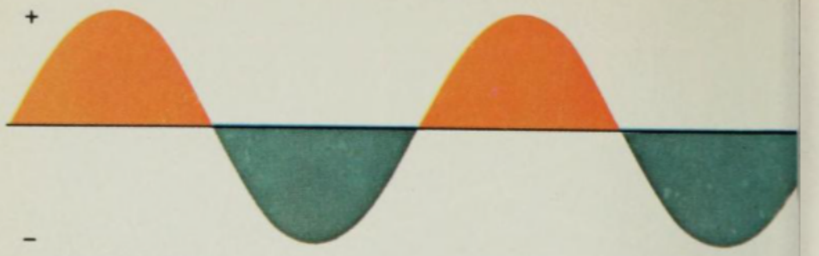

230. Two cycles of a sine wave

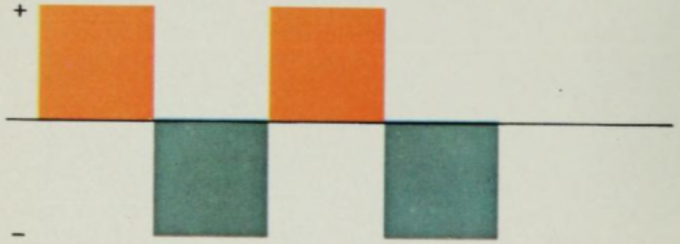

231. Two cycles of a square wave

Waveforms

An important wave is the sine wave [230]. This is obtained from alternating current supplies. The voltage rises smoothly to a maximum, and falls through zero to a negative value of equal amplitude. If we used an oscilloscope, we would see this happening 60 times each second with 60 cycle per second house supplies. The sine wave could be of other frequency, perhaps a pure audio tone, from a signal generator. 'Frequency' means complete cycles per second.

The smooth rise and fall of voltage and current enables mathematical relations to exist between the peak voltage or current, and the effective voltage or current (which is less). An instrument such as an AC voltmeter cannot respond to the rapid changes, and indicates the effective value. The square wave is useful for test and other purposes. It rises and falls almost instantaneously [231]. It can be produced by generating a sine wave and clipping positive and negative peaks.

The sawtooth wave is quite different [232]. If a capacitor is charged through a series resistor from a high potential

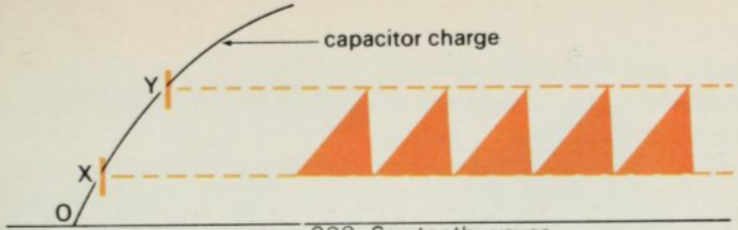

232. Sawtooth waves

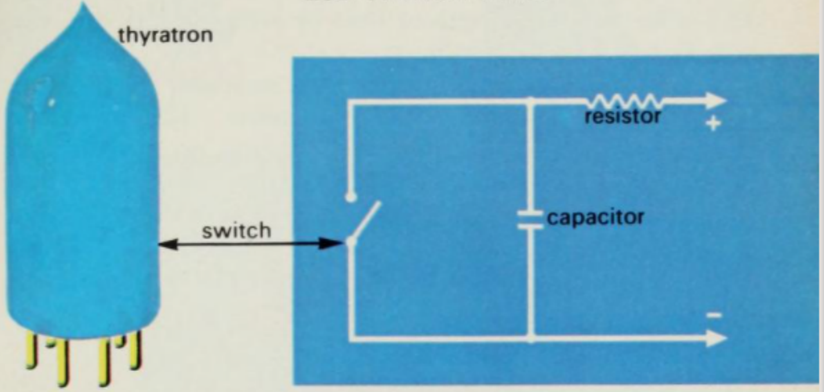

233. Circuit for producing sawtooth waves

the voltage across the capacitor rises. One section, from X to Y, is fairly linear ('straight'). To produce sawtooth waves, the capacitor is discharged rapidly when it reaches Y, so the voltage falls to that at X. The capacitor charges again. This is repeated. Rise from X to Y should be without curvature. Descent to X should be almost instant.

Figure 233 shows a circuit for producing sawtooth waves. Current flows through the resistor into the capacitor, so that the charge rises (X to Y). The speed and extent of this rise can be adjusted by selecting capacitor and resistor values and the supply potential. When the charge reaches Y the switch is closed, discharging the capacitor. The switch is then opened, so the charge begins to rise again. In practical circuits, a thyratron can replace the switch. This tube conducts abruptly each time the potential reaches a pre-arranged level (see page 53).

Sawtooth waves are used in an oscilloscope. The slow rise of voltage makes the cathode ray traverse the screen, and the ray returns almost instantaneously when the capacitor is discharged, to begin another sweep.

Radar and Other Devices

Radio waves travel outward at 300 million meters or 186,000 miles per second.

So if a signal has a frequency of 300 million oscillations a second, the wave length is 1 meter [234]. The radiated signal has magnetic and electric fields at right angles. Unless reflected by the earth or bent by ionized belts, etc. the signals travel in straight lines.

A directional antenna or dish gives strongest reception of signals when pointed at the transmitter. The direction-finding receiver Rx [235a] takes bearings on transmitters Tx. These bearings on a chart cross to show the receiver position.

If the position of a transmitter Tx is not known [235b] bearings are taken from two positions. The lines cross at the transmitter.

Radio Astronomy

Observations made directly by ordinary telescope are confined to visual frequencies. Photographic time exposures record dim stars. Visible light and similar rays are absorbed by haze, but radio frequency radiations are of longer wave length, and are not hindered.

234. Speed of radio waves

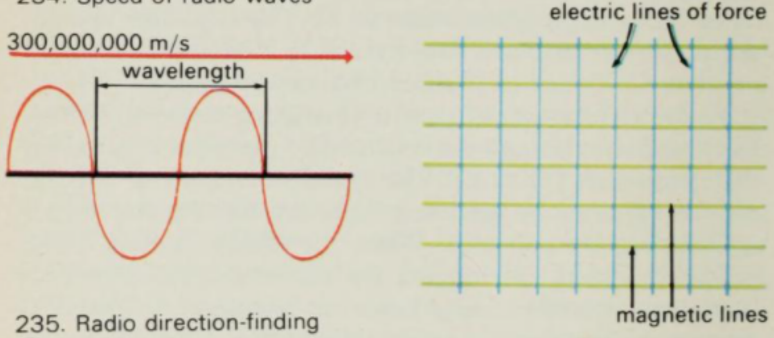
300,000,000 m/s
wavelength
electric lines of force
magnetic lines

235. Radio direction-finding

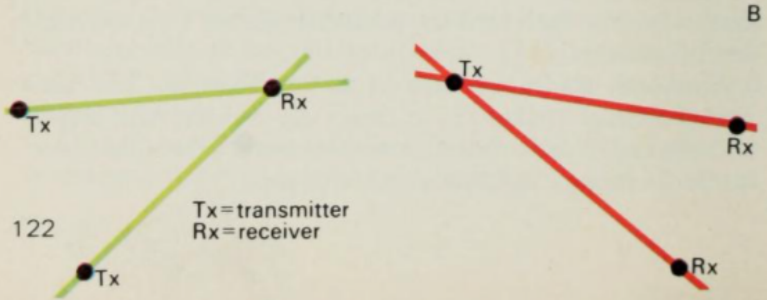
B
Tx
Rx
Tx
Rx
Rx
Tx
Tx=transmitter
Rx=receiver

236. Radio telescope

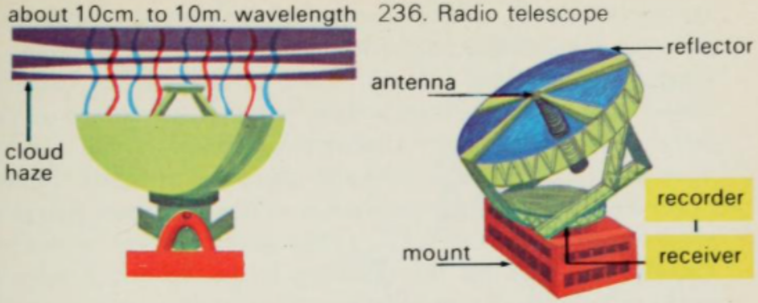
about 10cm. to 10m. wavelength
cloud
haze
reflector
antenna
recorder
receiver
mount

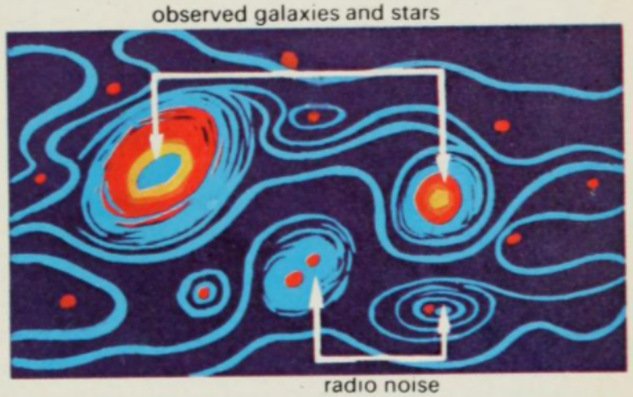
observed galaxies and stars
radio noise

237. Visual observation and results of radio astronomy

Remote galaxies and systems generate radio 'noise'—that is, random bursts of radio-frequency radiation. A directive antenna can locate the bearing of these. The dish reflector [236] may be hundreds of feet in diameter, and concentrates signals on an antenna. The dish can be pivoted. Signals pass to a sensitive receiver, and are recorded. Radio noise sources are often located near systems observed optically [237]. Sources of radio noise have been mapped at enormous distances, beyond the reach of an optical telescope and where no systems can be seen or photographed. In this way, information has been gathered outside the range of optical observation. Radio waves travel 186,000 miles per second. Radio telescopes have located galaxies over 4,500 million light-years away.

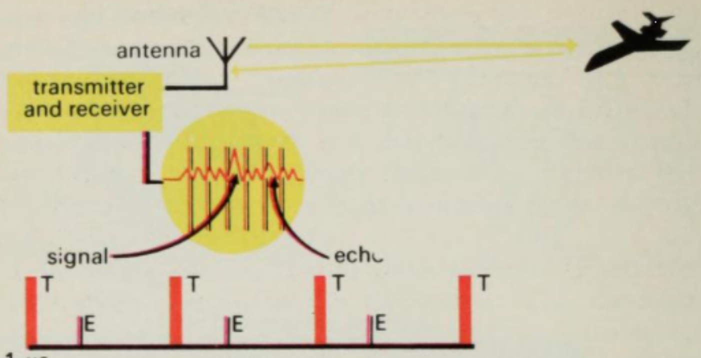
antenna
transmitter and receiver
signal
echo
T
E
1 µs
2,500 µs

238. Cathode ray tube showing echo

Radio Detection And Ranging (Radar)
If aircraft are in the path of a radio signal, some energy is reflected as an 'echo' back to the transmitter. In Figure 238 a transmitter radiates 1 micro-second pulses T of very high frequency at 2,500μs intervals and followed by periods in which a receiver picks up echoes E. As signals travel at known speed, the distance of an object is measured by noting the time taken for the echo to return. The interval can be shown by a cathode ray traveling rapidly from left to right across the screen [238]. A scale can be marked in distances.

Plan position radar equipment shows the distance of an object. By rotating a directional antenna [239], and causing the trace to scan from the center of the tube outward and rotate with the dish, the compass bearing of an object is shown, as well as its distance. As the antenna and trace revolve, echoes from surrounding features build up a plan or picture on the screen. The cathode ray may be brightened at regular intervals to produce range rings [240]. These can be expanded or contracted at the same time as the trace speed is adjusted, to change the 'scale' so that large or small areas can be examined. Bearing marks can also be provided [240].

239. Rotating dish and trace for bearing

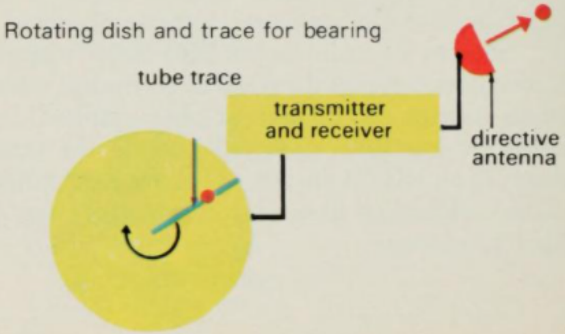
tube trace
transmitter and receiver
directive antenna

Some precision-approach radar for ground aid to landing aircraft has two antenna systems. One directive system sweeps a narrow beam in a horizontal plane, and the other sweeps a vertical plane [241]. The two sweeps combine to give a 3-dimensional position for an aircraft—one set of equipment produces the 2-dimensional 'map', the other the height reading, or other dimension.

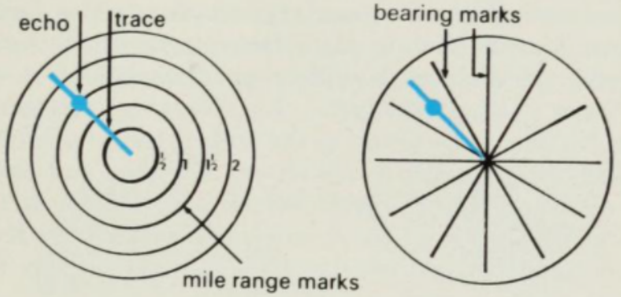
echo
trace
½
1
1½
2
mile range marks
bearing marks

240. Range and bearing marks on a plan position indicator

241. Precision-approach radar

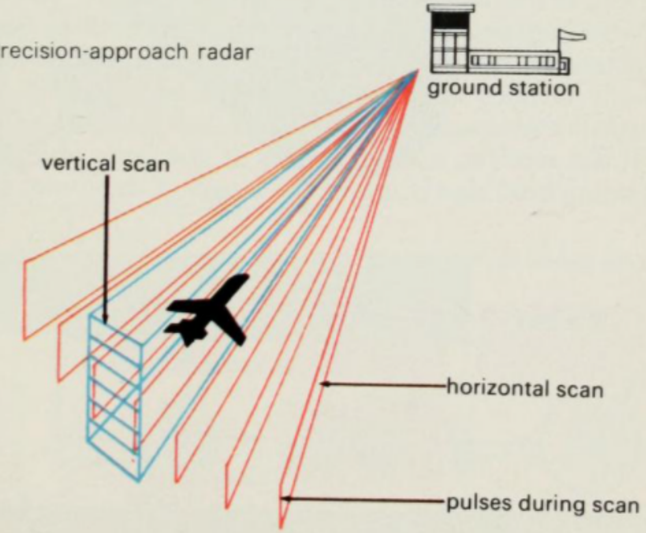
ground station
vertical scan
horizontal scan
pulses during scan

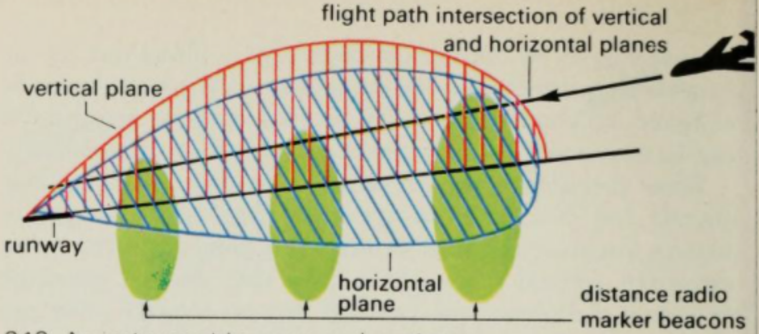
flight path intersection of vertical and horizontal planes
vertical plane
runway
horizontal plane
distance radio marker beacons

242. An instrument low-approach system

One instrument low-approach system for landing aircraft has a vertical signal to mark the runway, and a horizontal signal for the landing path. Beacons radiate vertically at 'outer', 'middle' and 'boundary' positions [242].

Sonar (Sound Navigation And Ranging) systems allow the distance to be found to the seabed and to a submerged vessel. High-frequency signals are amplified and passed to an electronic switch circuit, which directs them to a transducer [243]. Echoes are diverted to a receiver by the electronic switch, and the delay allows the cathode ray tube to give a distance reading.

Pictures by Radio (Facsimile Transmission)

News pictures can be sent across the world by radio. A picture secured to a rotating drum can be scanned electronically in lines which lie closely side by side [244]. Radio signals corresponding to light and dark areas are produced. At the receiver, a drum rotates at the same speed, and a writing head slowly moves the length of the drum, producing light and dark marks. So a complete picture is built up during many rotations of the drum.

Radio signals for pictures may be tape-recorded for storage, or for re-transmission when long-distance short-wave radio conditions are suitable. To synchronize a tape, the tape can have two tracks, one for picture information, and the other recording from a crystal oscillator. These signals are amplified and applied to a synchronous motor, the speed of which depends on the frequency [245]. A second motor receives amplified signals from a second crystal oscillator, and rotation operates a differential gear. If the motor running from the taped signals changes speed, the differential gear cage turns. This operates a correction circuit which increases or reduces the tape motor speed. So movement of the tape is synchronized.

243. Sonar sound-navigation and ranging

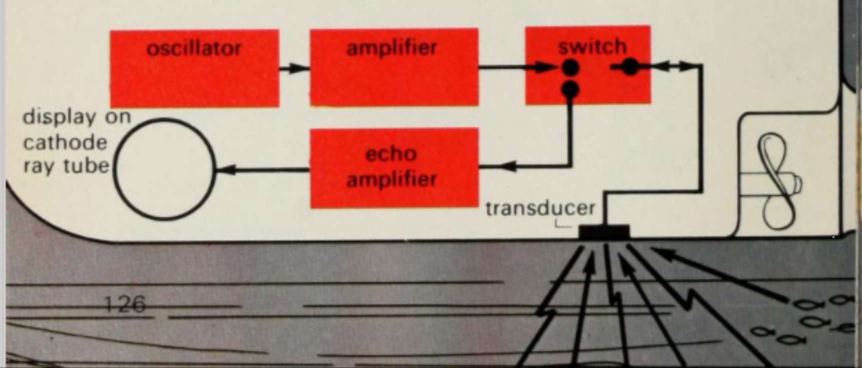
oscillator
amplifier
switch
display on cathode ray tube
echo amplifier
transducer

244. Pictures by radio

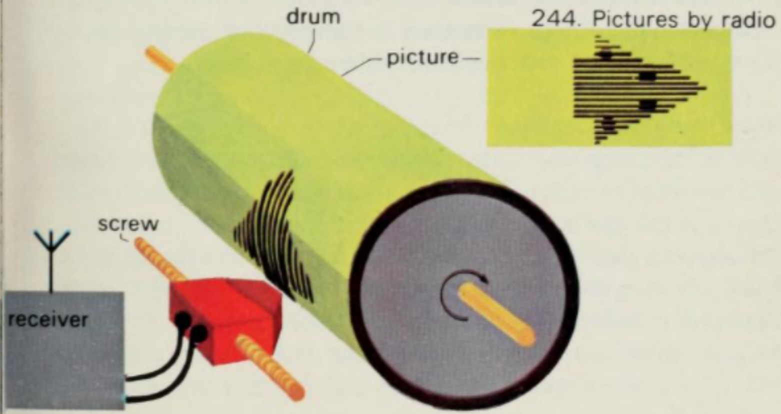
drum
picture
screw
receiver

245. Synchronizing radio picture tape

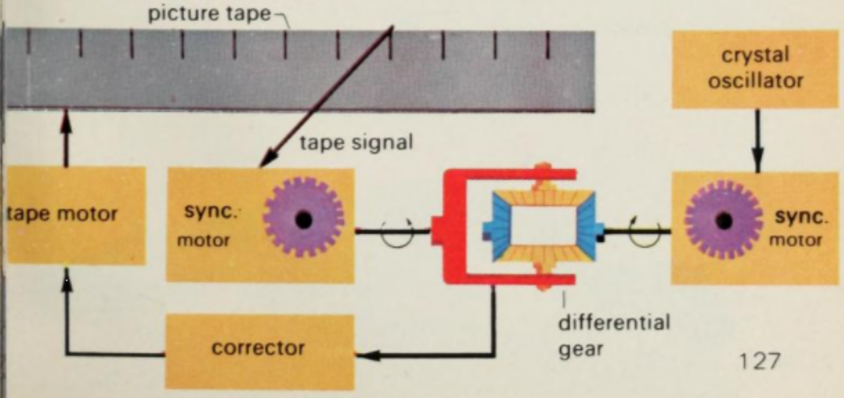
picture tape
tape signal
crystal oscillator
tape motor
sync. motor
sync. motor
differential gear
corrector

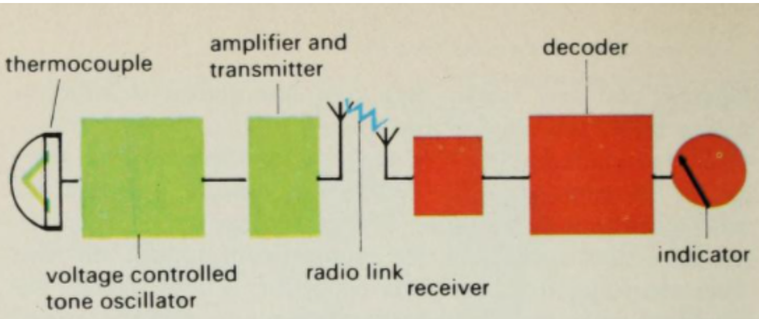
thermocouple
amplifier and transmitter
decoder
voltage controlled tone oscillator
radio link
receiver
indicator

246. Radio telemetry for temperature information

Telemetry is conveying information from instruments to a recording device at a distance—for example, from a satellite to earth. One way is to convert the 'information' into an audio tone, which changes to convey data. As an example, a thermocouple produces a voltage which changes with temperature. This can be taken to an oscillator whose frequency changes with voltage. The tone is transmitted, picked up by a ground station, and decoded [246].

Satellites

Satellites can investigate conditions far above the earth, and send radio reports to ground stations. One employs an orthicon camera to take cloud pictures [247]. Pictures may be taken in daylight and are stored on tape in the satellite. When it is within range of a receiving station, the taped signals are radiated for decoding and printing.

A satellite can measure electron temperature, galactic radio noise, earth-generated noise, oxygen density and other factors. In Figure 248 winged booms with solar cells produce current to operate internal equipment, and to charge storage batteries. Loop, extended and other antennas transmit and receive signals, to convey information to earth, or receive commands. Channels for radio signals are in the high frequency bands. Sun attitude cells receive illumination according to their bearing with relation to the sun's rays, and this can convey information on the tilt or rotation of the satellite.

247. Weather satellite

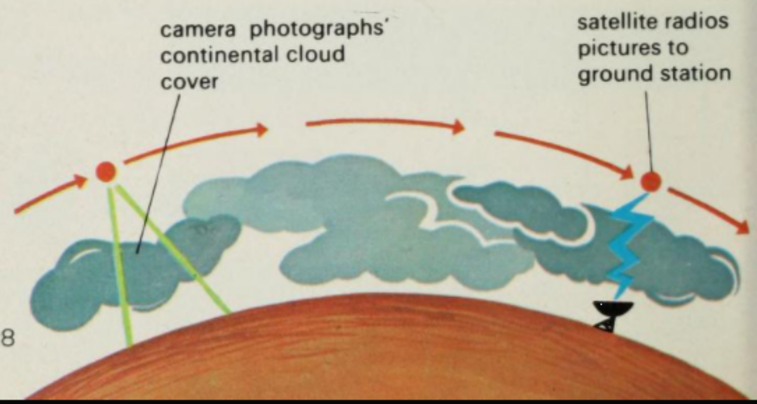
camera photographs' continental cloud cover
satellite radios pictures to ground station

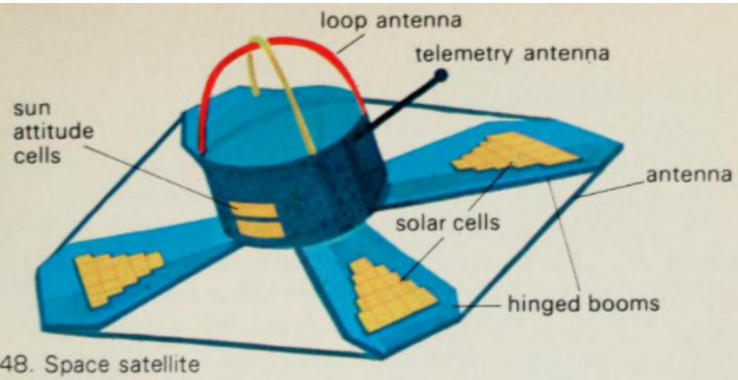
loop antenna
telemetry antenna
sun attitude cells
antenna
solar cells
hinged booms

248. Space satellite

A relay satellite can be situated thousands of miles above the earth, in a stationary orbit. Solar cells generate power. The satellite has antennas and equipment to receive transmissions, and re-transmit them on a frequency [249]. Transmissions are received on the ground by a large paraboloid dish, which reflects signals to a focus at a pick-up point. The efficiency of such dish parabolics is many thousands of times greater than that of an ordinary antenna and the receiver is specially designed for high sensitivity.

249. Radio relay by space satellite

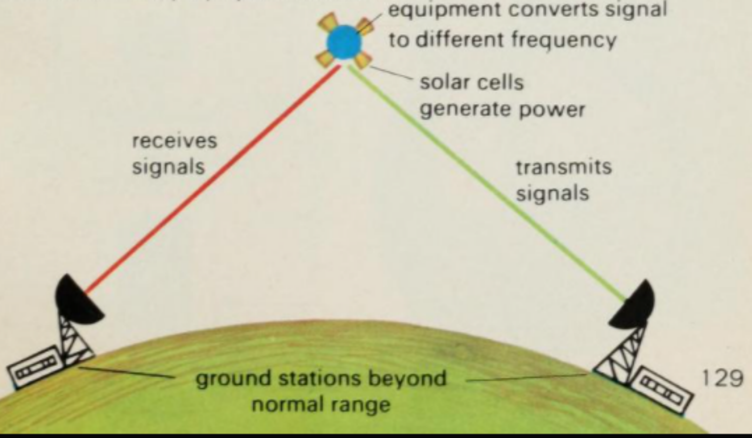
equipment converts signal to different frequency
solar cells generate power
receives signals
transmits signals
ground stations beyond normal range

Measurements in Electronic Circuits

Various measurements allow the working conditions of a circuit to be found. The simplest measurements are those which show a steady or uniform voltage or current.

A moving-coil meter winding is in a strong magnetic field, and receives current through hairsprings [250]. It turns slightly when current flows. This meter has a 'full scale deflection' or FSD of 6mA. A current of 6 milliamperes would move the pointer fully across the scale.

Current Through a Resistance = Voltage Divided by Resistance. So resistors in series with the meter allow various voltage ranges to be read.

Figure 251 has R1, R2 and R3 for ranges of 2.5 volts, 10 volts and 100 volts, and a 0.25mA meter. So R1 is 10,000 ohms; R2, 40,000 ohms; R3, 400,000 ohms.

To measure heavier currents we connect a resistor across the meter. Suppose $\frac{9}{10}$ of the total current flows through this resistor, and only $\frac{1}{10}$ through the meter [252]. The original range is multiplied 10 times.

The multi-range instrument in Figure 253 has a switch with nine positions.

250. The moving coil or d'Arsonval meter—basis of most test instruments

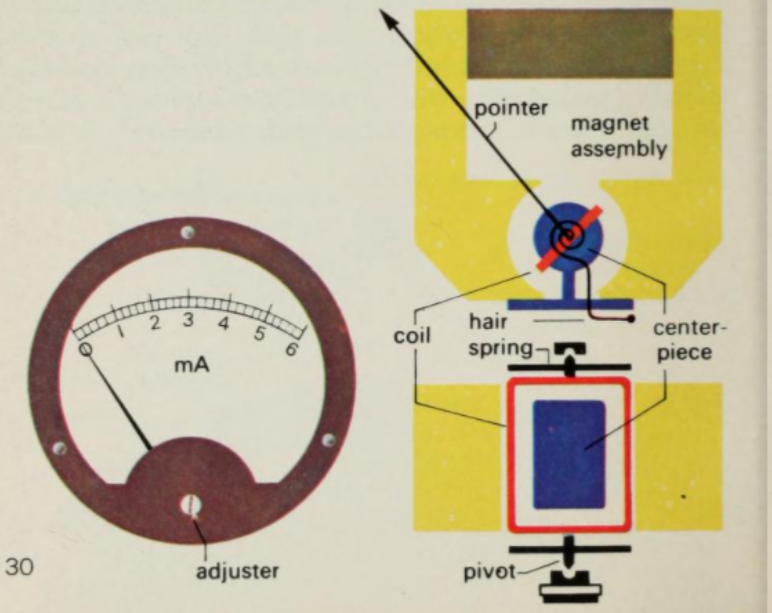
pointer
magnet assembly
coil
hair spring
center-piece
pivot
adjuster
mA
0
1
2
3
4
5
6

At 1, R1 is in circuit for a voltage range. Positions 2, 3 and 4 give other voltage ranges. R2, R3 and R4 are in series, so that the higher voltages are not across a single resistor. Positions 6, 7, 8 and 9 move along a tapped shunt for higher current ranges. The shunt is permanently across the meter. Suppose the meter has a FSD of 0.5mA, or 1mA FSD with the shunt added. Then R1 is 2.5k for 2.5 volts, R2 is 10k for 10 volts, R3 is 90k for 100 volts (10k already present) and R4 is 150k for 250 volts (10k plus 90k already in the circuit).

251. Multipliers for voltage ranges

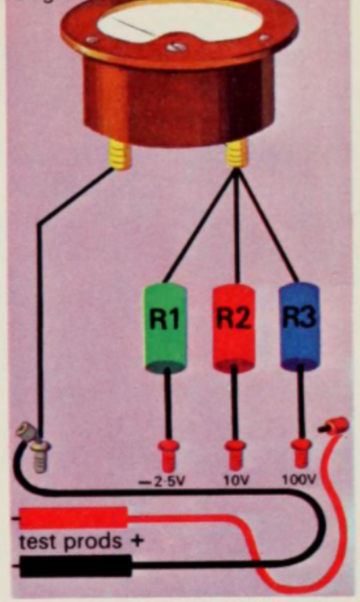
R1
R2
R3
−2·5V
10V
100V
test prods +

252. Parallel resistors or shunts

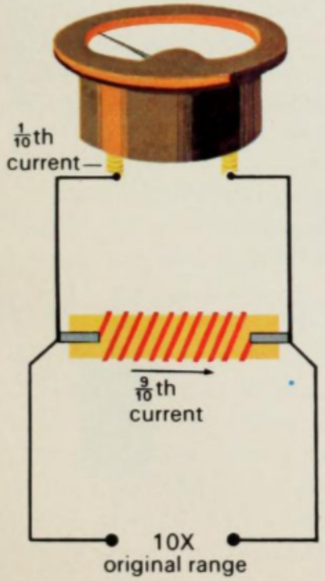
$\frac{1}{10}$th current—
$\frac{9}{10}$th current
10X original range

253. A multi-range meter circuit

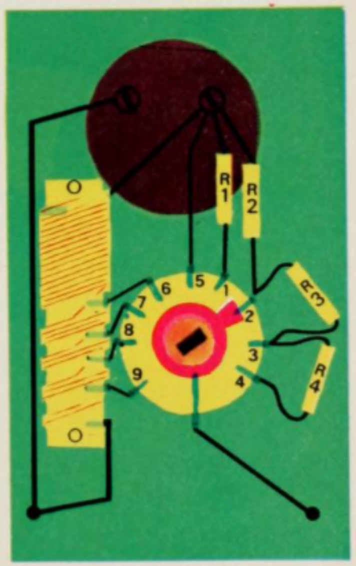
R1
R2
R3
R4
O
O
1
2
3
4
5
6
7
8
9

254. Use of rectifier instrument

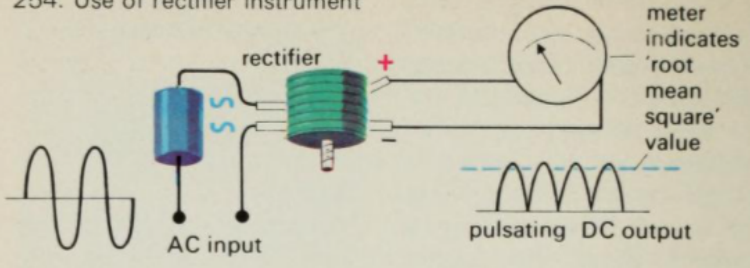

Alternating Current

The moving coil cannot respond to AC. For AC ranges we add a rectifier [254] to a meter like that in Figure 250. This changes AC to pulsating DC. The meter shows the 'Root mean square' or effective value—what we generally need to know. 'Effective' value is that which has the same heating effect as equivalent direct current.

Resistances Too

We may need to measure resistance values. We use a battery, variable resistor VR, fixed resistor R, and the meter all in series [255]. We join test leads from the sockets together and adjust VR until the meter reads full scale. This is *zero ohms*. If we separate the test leads, no current flows, and the meter reads zero current, which is infinitely high resistance, or *infinity*. If we take the leads to a resistor, or circuit being tested, the meter will indicate some intermediate current, depending on the resistance in the circuit. From Ohm's Law we can make a scale for the meter, showing resistance values.

255. This circuit measures resistance

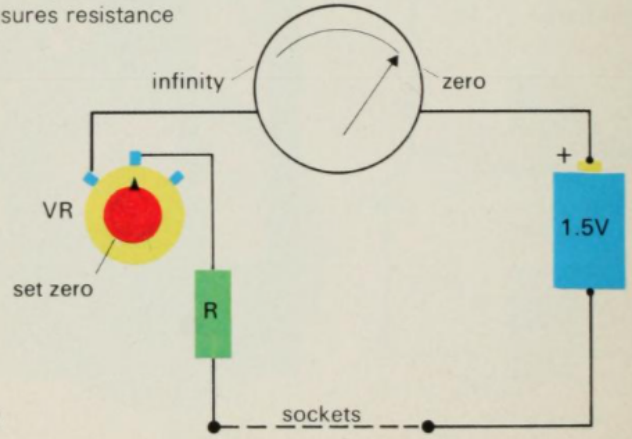

A 'bridge' [256] permits accurate determination of resistance.

If $\frac{Ra}{Rb} = \frac{Rx}{Rc}$ there is no e.m.f. across the detector.

Rx is is the resistance to be measured and Rc an accurately calibrated variable resistor, or a decade box with values selected by switches. Ra and Rb are of known value or ratio. We find the value of Rx by noting what value of Rc produces no current at the detector.

256. Bridge measuring circuit

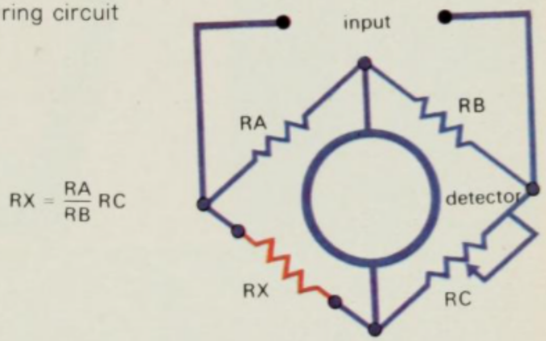

Power Also

Power in Watts = I^2R that is, *Current Squared* × *Resistance*. A known resistance in series with a meter allows watts to be found [257]. This is handy for some uses. Suppose R is 50 ohms and I is 2 amperes: 2 × 2 × 50 = 200 watts.

Sometimes voltage drop across a resistance may be more readily found. Suppose R is a 2 ohm resistor replacing a loudspeaker and E proves to be 3 volts. From $watts = \frac{E^2}{R} = \frac{3 \times 3}{2} = \frac{9}{2}$ we find that our amplifier is delivering 4.5 watts of audio power.

For direct current circuits, a DC meter is used. But for radio frequency, audio frequency, or alternating current circuits we must use a meter of appropriate type. Popular multi-range meters have scales, with range, AC/DC and ohms controls [258]. The switches bring into use the various parts of the circuit wanted.

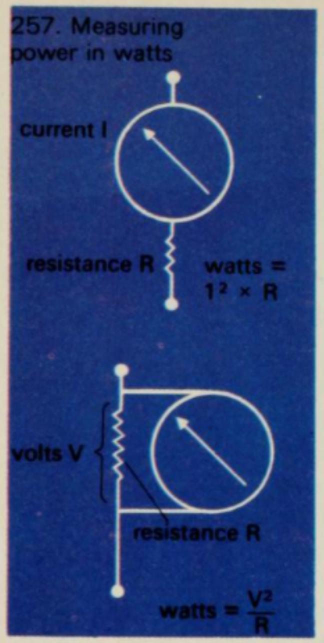
257. Measuring power in watts

current I
resistance R
watts = $I^2 \times R$
volts V
resistance R
watts = $\frac{V^2}{R}$

258. Multi-range meter

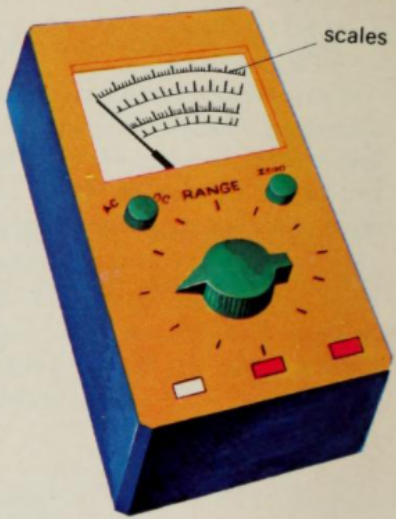
scales
AC
DC
RANGE

Millivoltmeters

A vacuum tube plate current changes with grid voltage. A plate meter can be calibrated in terms of voltage at the grid. Voltages to be tested are placed across a high value resistor R1 [259]. A range switch with resistors R1, R2, R3, etc. could be added.

A vacuum tube voltmeter allows extremely small voltages to be measured. The tube is very sensitive to grid voltage,

259. Vacuum tube voltmeter

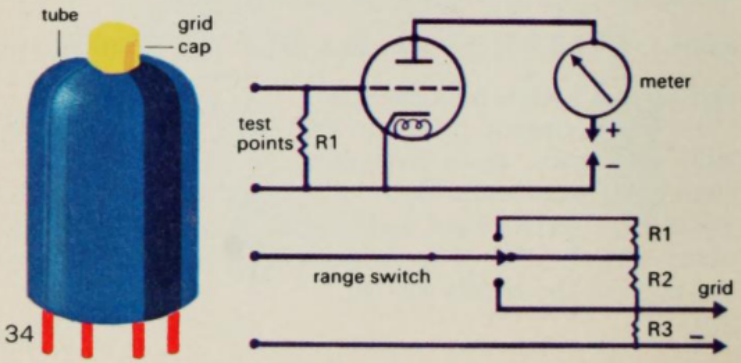
tube
grid cap
test points
R1
meter
+
–
range switch
R1
R2
R3
grid
–

and so it has a cap for this connection, to avoid leakage.

Sensitive meter circuits are used to investigate exceedingly low voltages and currents, and to test circuits where current is too small for a moving-coil meter.

Sensitivity can be raised by using an amplifier [260]. V1 is the vacuum tube. Changes in plate current produce a change in voltage across R2. This voltage is taken across R3 to V2, so changing plate current through R4. With no input, VR1 is adjusted until the current through R5 is equal to that through R4, when the meter gives no reading. Low

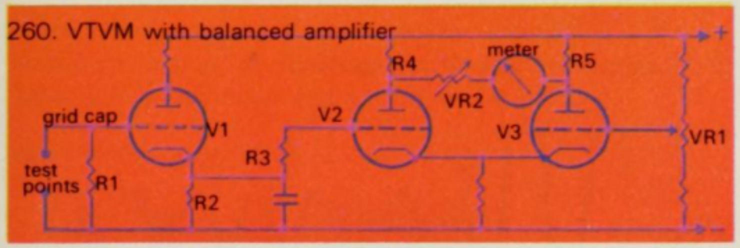
260. VTVM with balanced amplifier

+
meter
R4
R5
VR2
grid cap
V1
V2
V3
VR1
R3
test points
R1
R2
–

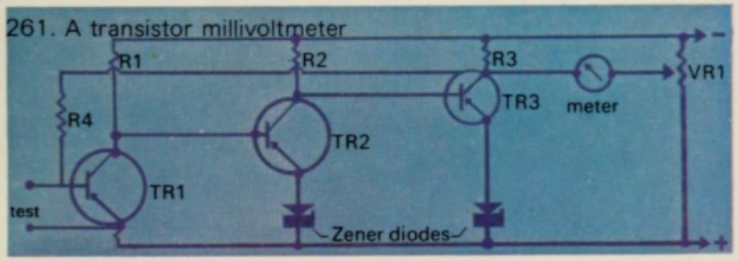
261. A transistor millivoltmeter

–
R1
R2
R3
VR1
R4
TR3
meter
TR2
TR1
test
Zener diodes
+

current or voltage at the 'Test Points' produces a reading on the meter. VR2 adjusts sensitivity.

The amplification from transistors allows sensitive measuring instruments to be made [261]. Here, a low voltage or current at 'Test' changes TR1 current, which flows through R1. This changes TR2 current through R2, which in turn changes TR3 current through R3. VR1 is first set for no reading, with no input. Zener diodes stabilize the emitter voltages of TR2 and TR3 because they maintain a certain voltage. The three transistors each amplify in series. Balanced circuits are often used to avoid errors due to transistors drifting.

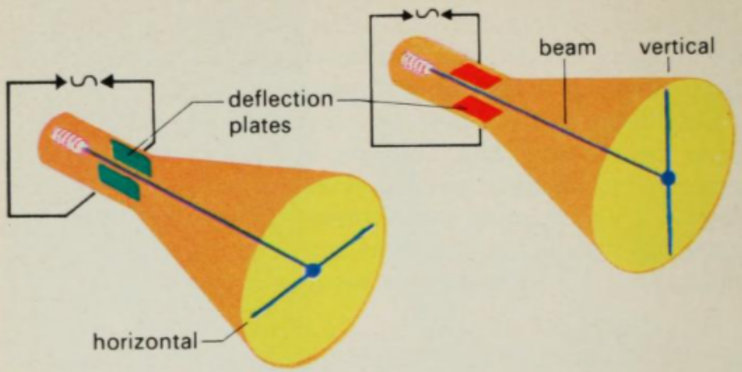
beam
vertical
deflection plates
horizontal

262. Use of oscilloscope for testing

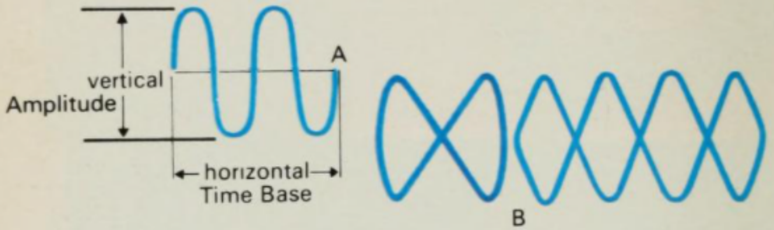
A
vertical
Amplitude
horizontal
Time Base
B

263. Traces on the scope screen

Oscilloscope

A cathode ray tube will display visually changes in current or voltage not shown by a meter. The cathode ray strikes the screen center, producing a bright spot [262]. Applying an alternating potential to the vertical plates deflects the beam up and down, making a vertical line. This deflection shows voltage amplitude. An alternating potential on the horizontal plates results in a horizontal line. By combining horizontal and vertical deflections we can display instantaneous events in a continuous pattern.

If we let an AC potential cause vertical scan, with horizontal scan from a sawtooth oscillator, we get a picture of the wave [263a]. Or we can take inputs to vertical and horizontal plates, causing wavy patterns in which the number of loops shows the relationship between one frequency and the other [263b].

The signals may need amplification [264]. Signals pass to the 'vertical amplifier' which raises the amplitude of weak inputs. The 'time base' is the generator of sawtooth waves, which sweep the spot across the screen. The 'sync' or synchronizing circuit takes a small part of the input signal and uses it to trigger the time-base, so that complicated traces remain stationary on the screen. A 'horizontal' amplifier is used. Another signal moves the spot horizontally. A 'power supply' provides current for each section, and for the tube, which may require 1,000 volts or more, at low current.

All sorts of phenomena can be displayed by the oscilloscope. For example, if an amplifier introduces third harmonic distortion, it combines with the fundamental frequency to produce an unpleasant result [265].

264. Cathode-ray oscilloscope

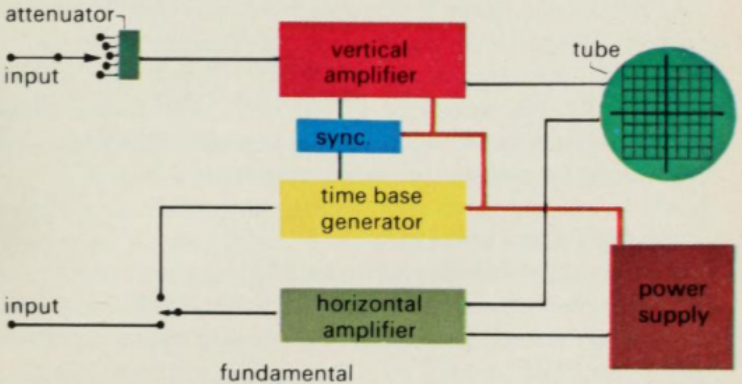
attenuator
input
vertical amplifier
tube
sync.
time base generator
input
horizontal amplifier
power supply

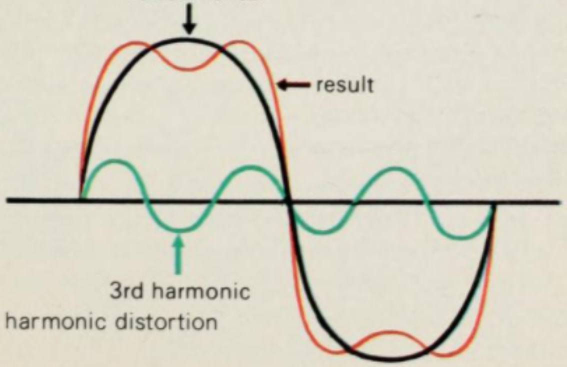

265. Third harmonic distortion

More Uses for Electronics

Electronics does remarkable things—if a process is wanted, an electronic circuit is usually possible.

Gasoline Pump

An electronic gas pump can replace the familiar electric pump. Coils on the pump are the windings for a low frequency oscillator [266], so the collector and coil current rise and fall. Valves in the piston and cylinder transform the rapid piston movement into a steady gasline pressure, exerted by a spring. This action provides a constant supply to the carburetor.

Interval Timer

An interval timer can automatically time a short period, and this can be repeated as often as is needed.

An electronic timer in Figure 267 makes use of the time constant of the large capacitor C. Transistor collector current holds the relay armature down. When the collector current falls, the armature is released. The potentiometer can be calibrated in seconds. Closing and opening of the relay contacts can control other apparatus.

Controlling Currents

The silicon-controlled rectifier or SCR can be turned on by applying current to a control electrode [268]. The SCR can deal with high power for motors and other appliances.

266. Transistorized gas pump for vehicles

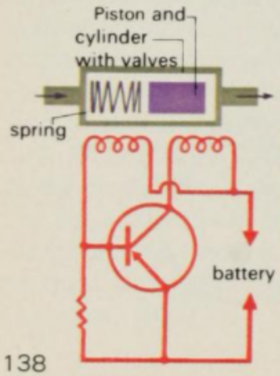

267. Transistorized interval timer

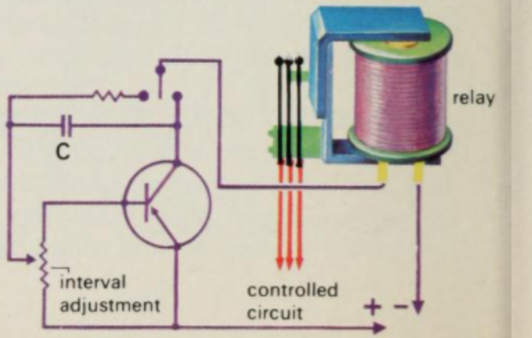

268. Silicon-controlled rectifier and circuit

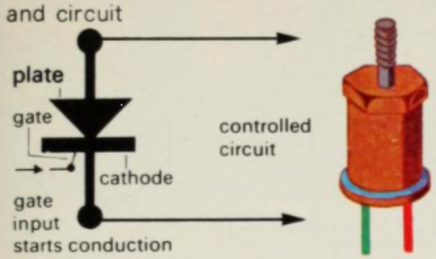

269. Pulse width power control

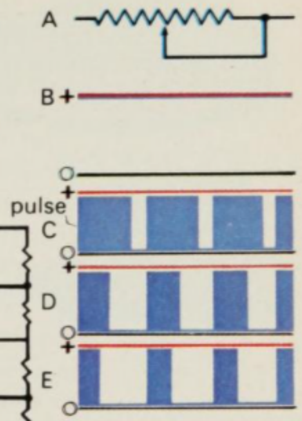

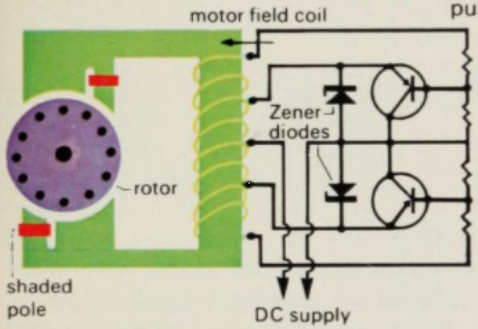

270. Brushless direct current motor

A gate-controlled switch or GCS is somewhat similar in effect. The speed of a direct current electric motor can be reduced by a variable resistance in one conductor: voltage to the motor is less [269a]. This dissipates unused power as heat in the resistance. Another way is to employ a 'pulse width' oscillator to control a solid state rectifier which allows current to reach the motor. At [269b] current is present the whole time for full speed. In Figure 269c short 'off' intervals reduce the average motor current. In Figures 269d and 269e the 'off' intervals are longer, reducing speed further. No power is dissipated in heat.

Brushless DC Motor

The brushes of a DC motor can be noisy; they wear out, and the sparking may cause R-F interference. The brushless direct current motor employs power transistors to cause an oscillating current [270]. The motor has shaded pole windings, as used for an AC machine, and needs no brushes or commutator. R-F interference and sparking is eliminated.

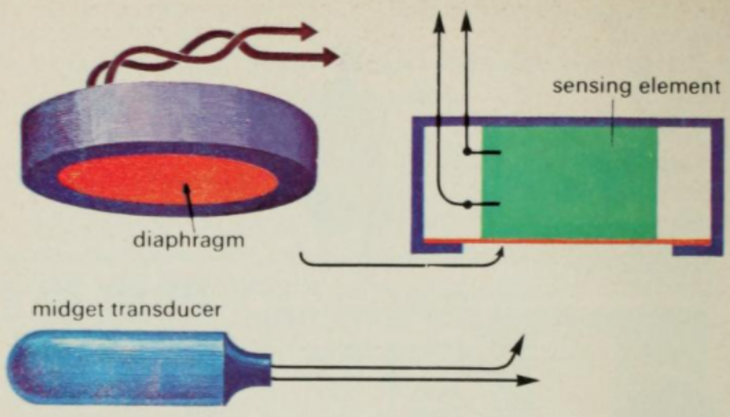
sensing element
diaphragm
midget transducer

271. Transducers

Transducers and Ultrasonics

A ‘transducer’ supplies an electrical signal under the influence of pressure or other effects. A pressure transducer [271], alters its electrical resistance with changes in pressure, and so can operate a meter or other indicator. Microminiature pressure transducers have elements formed in silicon crystal, and may be fitted to engine parts, or even in the human body.

In ultrasonic flaw detection [272], steel bars or other materials tested receive the ultrasonic waves upon one surface. The bars may move through a tank, or the equipment may rest on the objects tested. The ultrasonic wave is reflected by the more distant surface, and by internal flaws. A cathode ray tube shows the interval of time taken for the reflection, and thus whether a flaw is present.

272. Ultrasonic flaw detection

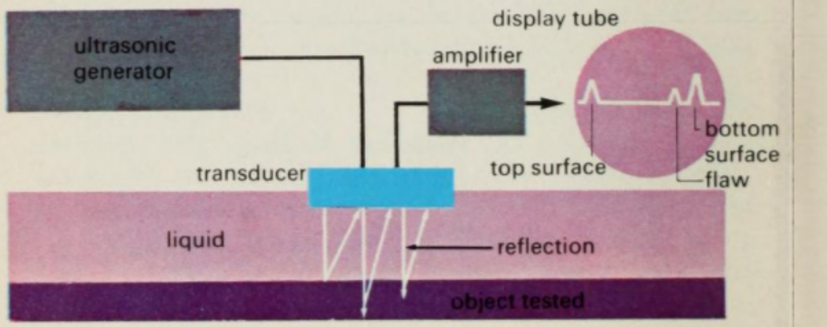
ultrasonic generator
display tube
amplifier
bottom surface
flaw
transducer
top surface
liquid
reflection
object tested

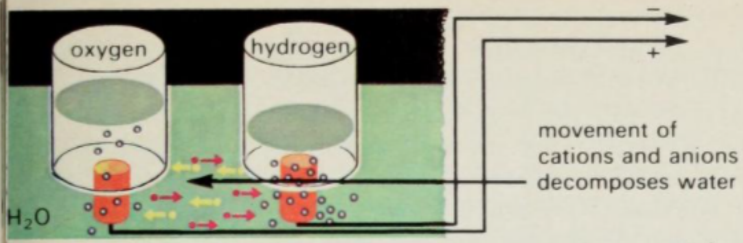
oxygen
hydrogen
−
+
movement of cations and anions decomposes water
H_2O

273. Electrolysis by passing a current in a chemical bath

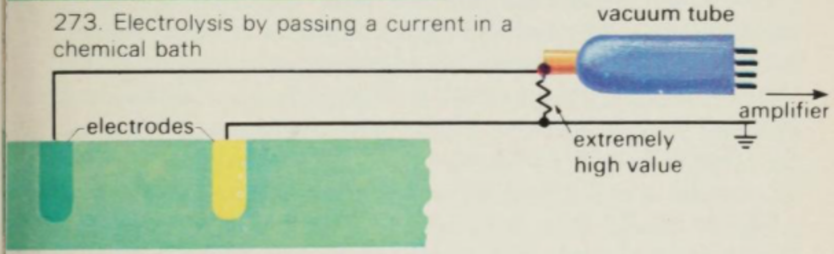
vacuum tube
amplifier
electrodes
extremely high value

274. Finding pH value of a liquid

Another use for ultrasonic equipment is in transmitting signals from a vessel into the water and receiving echoes from the seabed, fish shoals, etc. The time for the echo to return depends on the distance. (See also page 126.)

Electro-Analysis

Various methods allow chemical analysis by electronic means. In electrolysis, compounds are chemically separated by the passage of an electric current. In the electrolysis of water, hydrogen escapes at the negative electrode, and oxygen at the positive electrode [273]. Ions migrating in the direction of the positive current are called cations, while those moving in the direction of the negative current are anions. Substances in solution may be deposited upon neutral electrodes, and afterward weighed.

The degree of acidity or pH value can be found by a sensitive instrument with immersed electrodes. One electrode is neutral and the other responds to the pH value, so the output from the electrodes depends on the latter. An extremely sensitive measuring circuit is needed [274].

More Semiconductor Effects

Gallium phosphide alloyed junction diodes and gallium arsenide diffused junction diodes produce light when current is passed in the forward or conducting direction. The gallium arsenide diode junction region has a high concentration of electrons and positive carriers. Photons are emitted when electrons change their energy level. The diode is a form of semiconductor laser or light amplifier—photons traveling along the junction region stimulate re-combination and extra photon emission, so producing more light [275].

Stimulation of a junction by light photons produces a current. A small-current diode may produce one electron for each photon [276]. In the avalanche silicon photo-diode, a potential is applied across the junction. The photons have a higher energy level, and more electrons are liberated. In the photo-transistor, the photo-diode junction forms the base-emitter elements, and transistor action amplifies the base current, so a larger collector current flows from a battery [276].

275. Light-producing diode

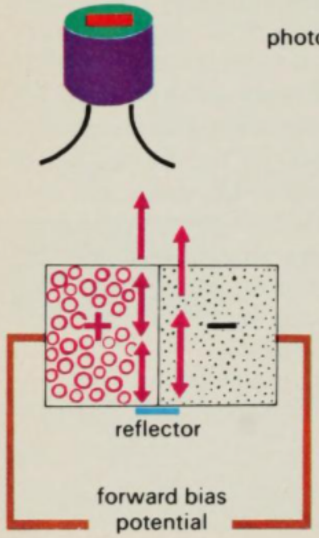

276. Photo-diodes and photo-transistor

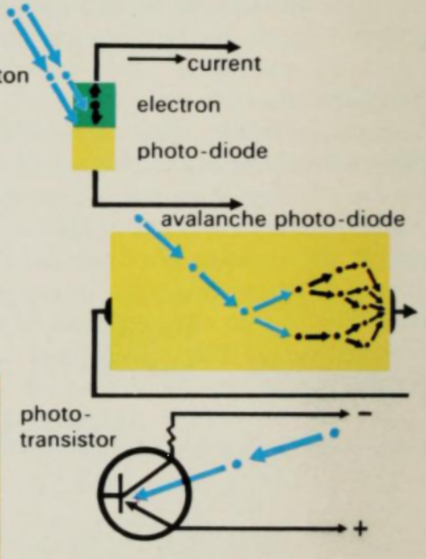

The Field-Effect Transistor

This has source, gate and drain connections. It is P-channel or N-channel. Figure 277 is a field-effect transistor with a P-type gate in the N-type channel. Electron carriers flow from the source to the drain. Changing the gate voltage alters the current which can flow through channel and external circuit. With no gate potential, the channel is at maximum width, A. Current passes readily. With a small negative voltage at the gate, the channel is depleted of carriers, and therefore is smaller, B. Current is reduced. Increasing the voltage causes the channel to become narrower until it is 'pinched off' or scarcely able to conduct at all, C. The transistor has a high input resistance. It imposes a smaller load on a circuit that supplies the gate. Ordinary transistors are operated by the current flow at the base, and thus draw a current which loads the base circuit. The field-effect transistor is controlled by the voltage. Figure 278 is a field-effect transistor audio amplifier. This type of transistor is rapidly becoming more popular in radio and other equipment.

277. A field-effect transistor

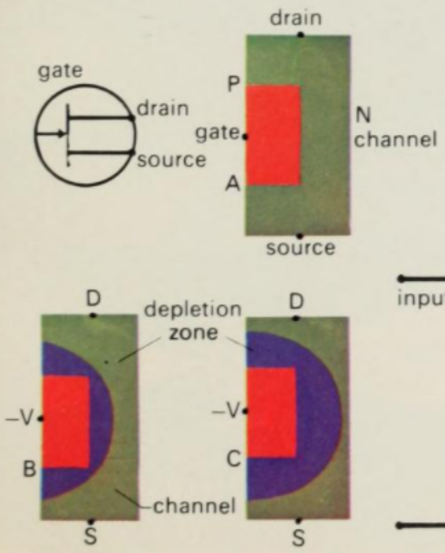

278. F.E.T. amplifier circuit

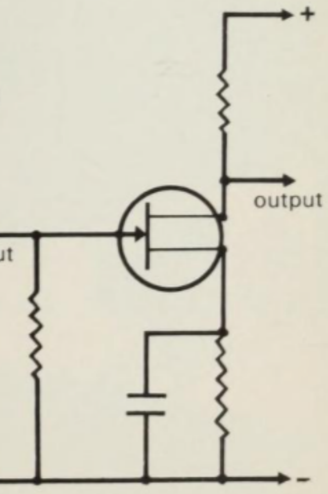

Television Cameras

A television camera supplies a video signal, afterward amplified and radiated from a transmitter. As the whole scene cannot be changed simultaneously into electrical signals, it is 'scanned' or covered in regular lines. Covering the picture area constitutes a 'frame'. Frames follow each other with sufficient rapidity to avoid flickering at 30 frames per second.

A television picture is built up by a tiny spot of light from the cathode ray, which scans the screen rapidly. A television camera performs the reverse action, generating signals which correspond to the light and shade of the scene transmitted. With space vehicles, the impulses may be stored on tape, later used to supply the transmitter.

Figure 279 shows how a picture area is scanned or sampled by a moving area X of small size. During one frame, the small area X has covered the whole picture. (See the section on television receivers, page 88.)

A camera lens throws an image of the scene on a film [280]. A variable zoom lens, or interchangeable lenses, can give a choice of ordinary, wide-angle, close-up and other views.

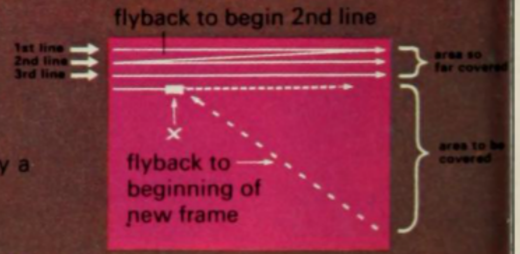
flyback to begin 2nd line
1st line
2nd line
3rd line
area so far covered
X
area to be covered
flyback to beginning of new frame

279. Scanning of a picture image area

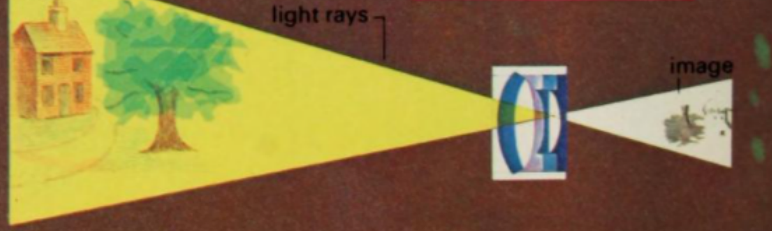
light rays
image

280. Formation of image by a convex lens

The image orthicon camera optical system projects an image on light-sensitive photocathode [281]. When light strikes the photocathode, photoelectrons are emitted from the other side. They are most numerous at bright areas of the image. An electric field attracts them to the target. This is extremely thin glass covered by a wire mesh having very many holes per square inch. The photoelectrons knock electrons from the glass, leaving a positive charge. So the positions and intensity of the positive charges correspond to light areas of the image.

The beam from an electron gun scans the back of the glass. When passing over an area where there is a positive charge, the beam leaves electrons to neutralize the charge. When the beam passes over an uncharged area, few if any of the electrons are left. Thus the beam current corresponds to light or dark picture features. An electron multiplier amplifies the changes in beam current. A series of electrodes receive progressively higher voltages [282]. Electrons striking these electrodes cause secondary electrons to be liberated, and the intensified signal can be passed on to other amplifiers.

281. Image orthicon used in TV camera

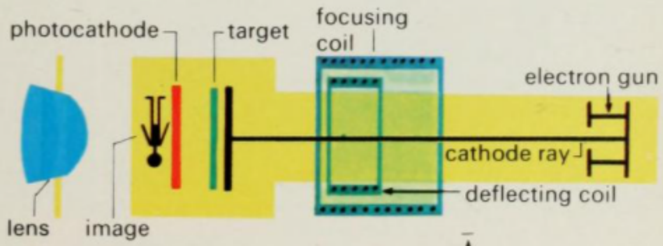
photocathode
target
focusing coil
electron gun
cathode ray
deflecting coil
lens
image

282. Electron multiplier

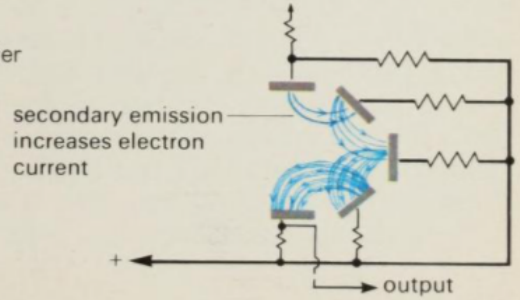
secondary emission increases electron current
−
+
output

Logic Gates

Processes of logic can be carried out by 'gates' whose output depends on the logical process or 'truth' required. Computers use very many of these gates. Figure 283 is a 2-input AND gate. Input pulses arrive at A and B. Transistors 1 and 2 are normally non-conducting. There is no current through R1, no voltage drop here. Transistor 3 base is negative. Current flow through R2 causes a large voltage drop. Output level remains near zero (positive).

A pulse at A or B makes one transistor conduct, but no current flows in R1 unless *both* conduct simultaneously—that is, when pulses arrive at A *and* B together. Then voltage drop in R1 moves collector C positive, transistor 3 ceases to conduct, and a negative pulse appears in the output. Figure 284 is a 2-input OR gate. A pulse at either A OR B will cause an output pulse. Such units may operate at 500,000 pulses or more a second. They work with low voltage (typically 5 volts) and require little power.

283. A 2-input 'AND' gate and inverter

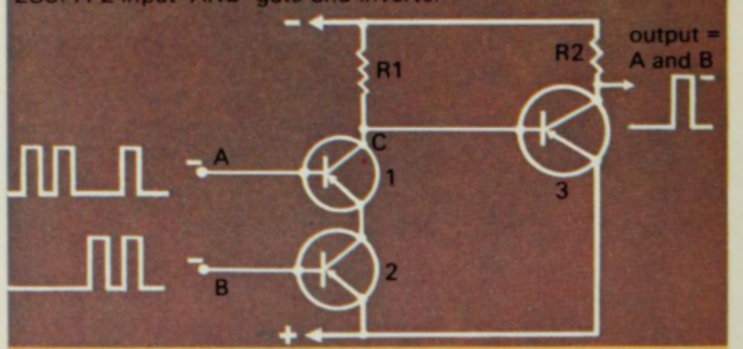

284. A 2-input 'OR' gate and inverter

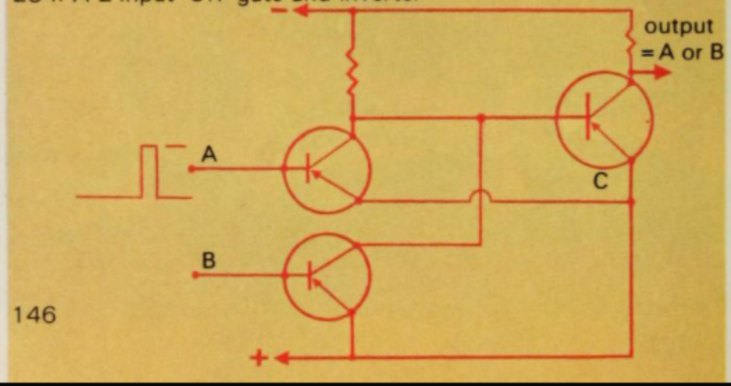

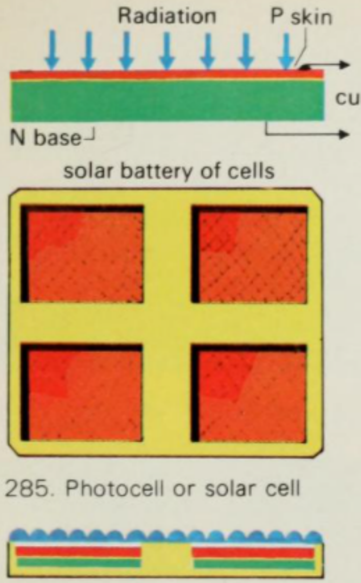

285. Photocell or solar cell

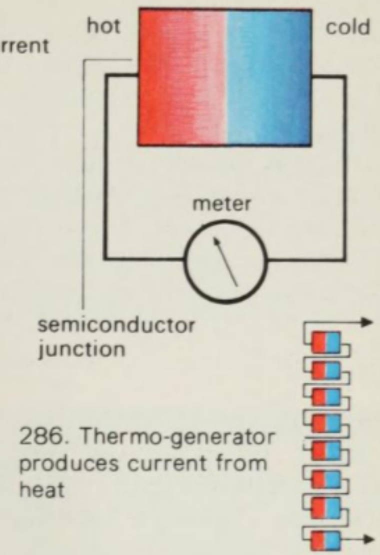

286. Thermo-generator produces current from heat

Power in Space

Various unconventional ways of generating electrical power have been used in Polar weather stations, for automatic warning lights, satellites and space probes.

A convenient source of power for a space probe is the solar cell. Silicon solar cells have an N-type base containing antimony or other impurity, and a P-type surface layer. Photons or light from the sun striking the cell cause extra electron and hole pairs, and a current which flows through an external circuit [285]. This produces about 0.5 volt. The solar battery can charge accumulator-type cells, which provide power for equipment requiring more current and when the satellite is in darkness.

The thermo-generator is an improvement on the thermocouple. In N-type material, electrons move away from a heat source. So a device can be made in which current is produced if one surface is hot, and the other cold [286]. The heat is a by-product of environment (solar) or of relevant activity. In a reverse or 'refrigeration' system the passage of an electric current makes a semi-conductor junction drop in temperature.

Looking Ahead

Analysis of gases, once needing considerable time, can be completed almost instantly. After ionization, the particles are accelerated down a drift tube [287]. Low mass ions accelerate most rapidly, so the ions are separated according to mass.

'Printed circuits' have revolutionized construction of miniature equipment. Conducting paths are arranged to run without overlapping. A large drawing is prepared, and reduced by photocopying methods. Further processing leaves printed copper conductors [288]. Large numbers of similar panels are readily made. When components are connected, the printed conductors act as wiring.

287. High-speed spectrometer

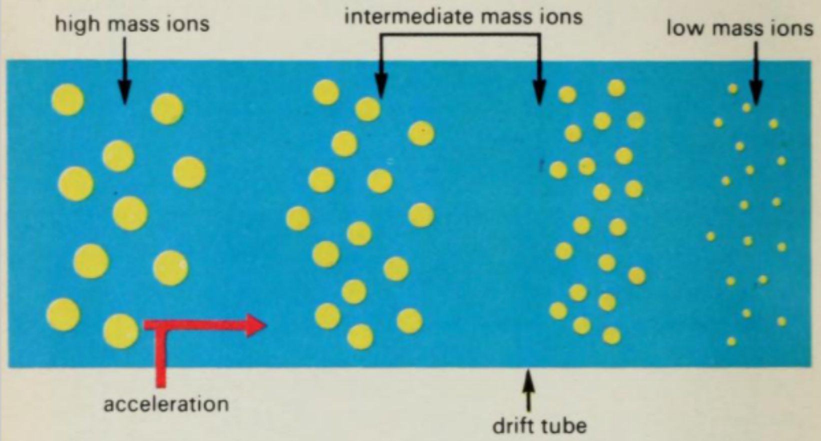
high mass ions
intermediate mass ions
low mass ions
acceleration
drift tube

Printed circuits and similar methods lead to the production of modules, or small units of similar type. A defective unit is removed and replaced by one in working order.

Computers contain thousands of 'logic' elements, such as those for AND and OR functions. These and other circuit units can be tested automatically at high speed by a machine which applies suitable signals, and records the outputs [289]. Errors in outputs from the unit result in it being set aside for repair.

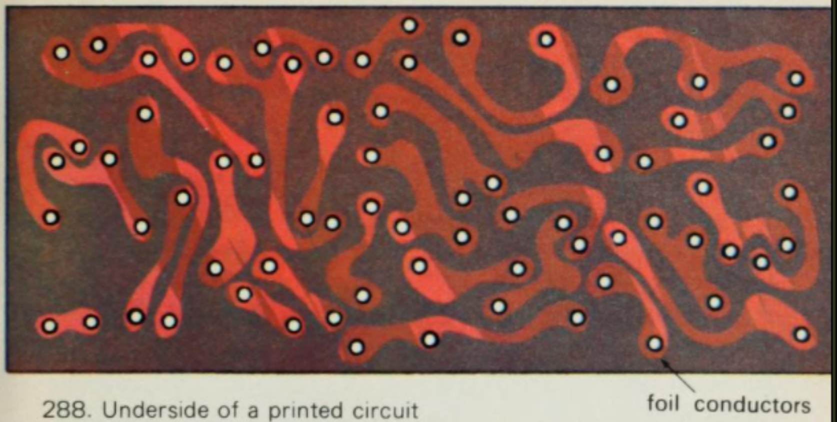
foil conductors

288. Underside of a printed circuit

289. Automated module testing

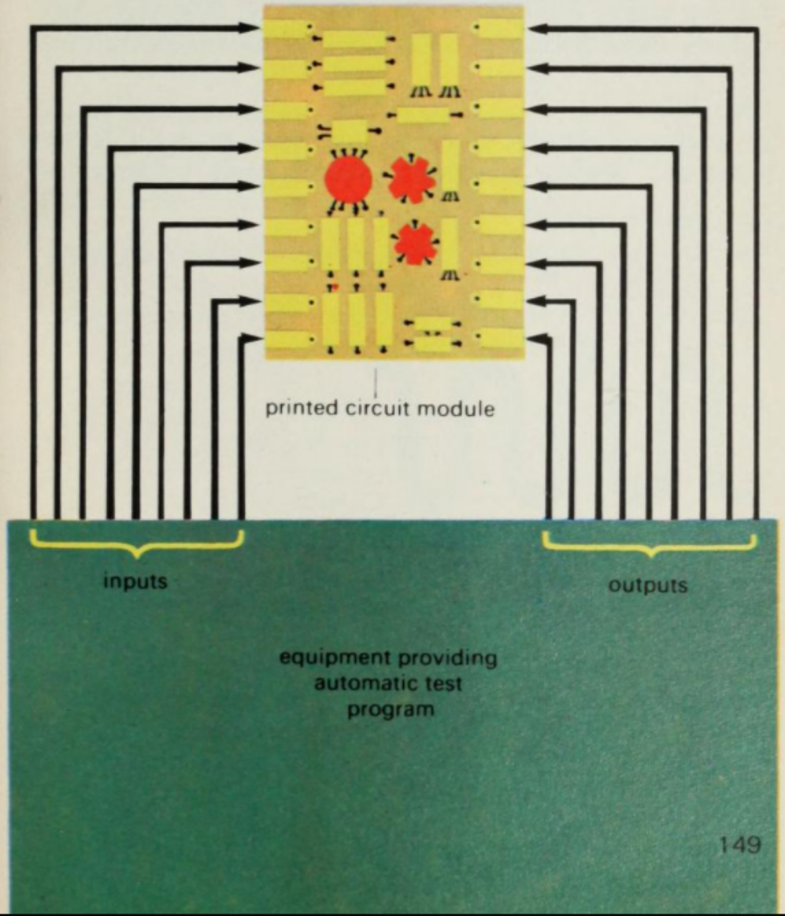
printed circuit module
inputs
outputs
equipment providing automatic test program

A Step Smaller

Even miniature individual resistors have wire ends and occupy unnecessary space. Instead, resistor material can be deposited directly on an insulated panel, between conductors [290]. The resistor has almost no thickness. If the process is carefully controlled, the resistors will be reasonably accurate in value. For high accuracy, the resistor is made a little low in value, and the value is increased by cutting or etching material away.

Conventional capacitors are of foil, separated by insulating paper; or have sheets of mica to separate the plates, or consist of conductive deposits on an insulator. Capacitors of almost no thickness are made [291]. Conducting layer A is deposited, then the insulator or dielectric B. A further conducting layer C is deposited on the insulator. Surfaces A and C form the capacitor.

Thin Films

In integrated circuits, the insulating base or panel is the 'substrate'. Surface roughness would result in irregular deposits, so ceramic, glass, and other smooth substrates are used.

290. Thin film resistors

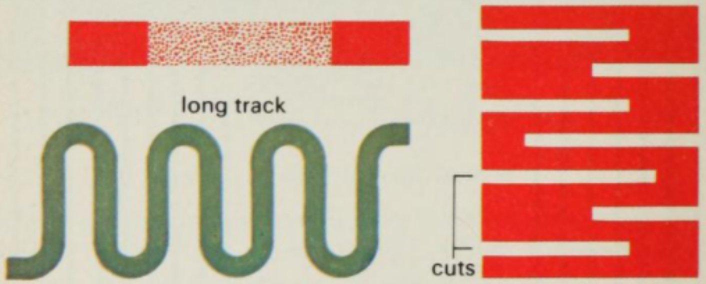

291. Thin film capacitors

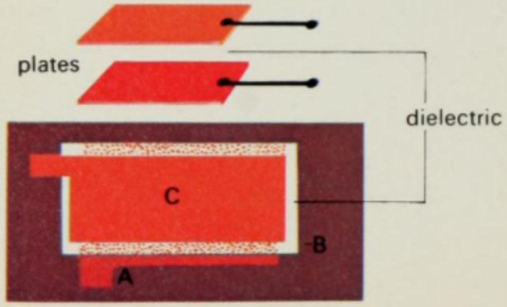

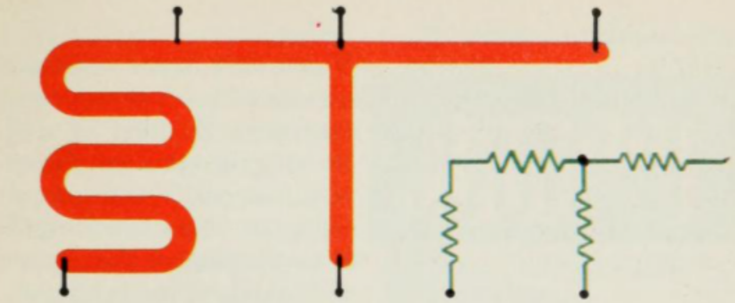

292. Thin film resistor network

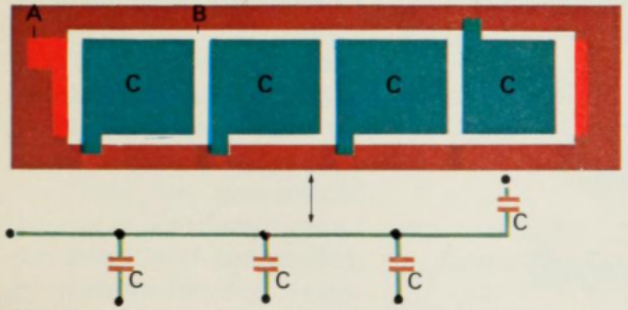

293. Thin film capacitors and equivalent circuit

Resistors can be deposited nichrome, of extreme thinness, and resistance paths can be shaped as wanted, and tapped if required [292].

One method is to cover a prepared substrate chip with lacquer and photoresist, so that when the pattern is printed and etched the nichrome can afterward be deposited, giving the required thickness and shape.

Capacitors are made by vacuum depositing aluminum, then the dielectric (insulator), then aluminum again. The dielectric may be magnesium fluoride, zinc sulphide, silicon monoxide, or other material. For high capacities in small size, the film is very thin.

Deposits can be shaped to give the desired result. In Figure 293 the single conductor surface A is covered with dielectric B, then four top surfaces C, thus obtaining the equivalent of four capacitors.

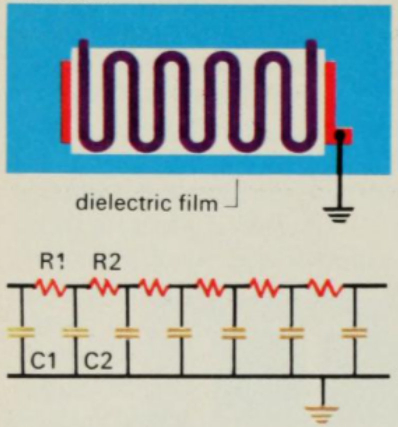

294. Resistance capacitor network

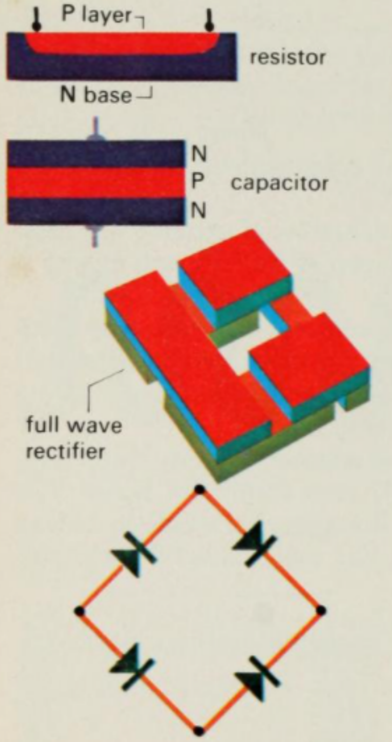

295. Semiconductor components

Other Units

Figure 294 has a resistance element, then a dielectric or insulator. A film of conducting material is deposited on the insulator, so there is capacitance from it to the resistor throughout its length. This is approximately equivalent to a number of individual resistors R1, R2, etc. with small capacitors C1, C2, etc. Phase-shift networks for oscillators and other purposes have been made in this way.

A completed integrated chip having resistors, capacitors and interconnecting conductors can be of extremely small size. The chip could be 0.1mm. thick, or less, and have many components in an area of a few square millimeters. A number of chips can be placed one on top of another, to build up a thicker unit, or cube, in which are many circuits.

Semiconductors Too

Semiconductor material can be fabricated into microminiature circuit elements. Silicon is preferred, for some increase in temperature is less important than it is with germanium.

A resistor is obtained by having a thin track of P-type material in the N-type base [295]. Capacitors can be derived by having N-type material separated by P-type material. Arrangements of suitably chosen material can provide more complicated circuit equivalents, such as the full-wave bridge rectifier.

Effects present in conventional semiconductor devices can be employed. In Figure 296 the variable capacitor has two metal plates, which are moved to change the capacitance. In the semiconductor variable capacitance diode, capacitance exists between the carrier areas. Applying a bias voltage effectively moves these areas together or apart, changing capacitance.

Planar transistors occupy little space and thickness. Another method adopted is shown in Figure 297. A base carries a resistance element acting as three resistors in series, with conductor strips A, B and C. Transistors are formed on the material. Input to X or Y will cause a voltage drop in R1, so that a pulse is obtained at B. This is a computer 'OR' gate.

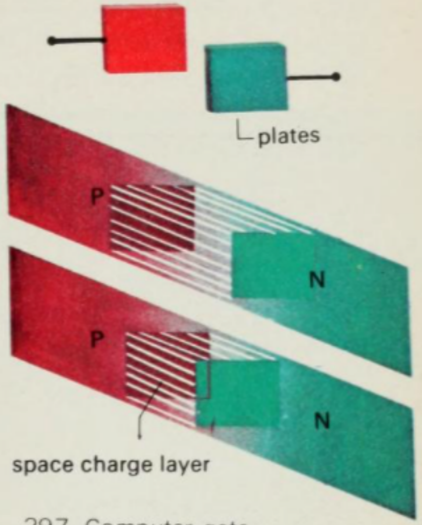

296. Semiconductor variable capacitor

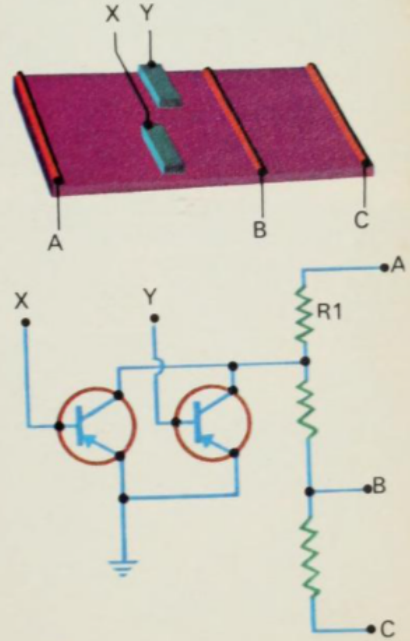

297. Computer gate

Figure 298 is an integrated circuit formed by diffusion processes in a piece of semiconductor. With suitable external connections, it may form part of an amplifier, counter, oscillator, etc. With direct couplings and complementary circuits having PNP and NPN transistors, full use can be made of I.C. techniques.

298. Integrated microminiature circuit

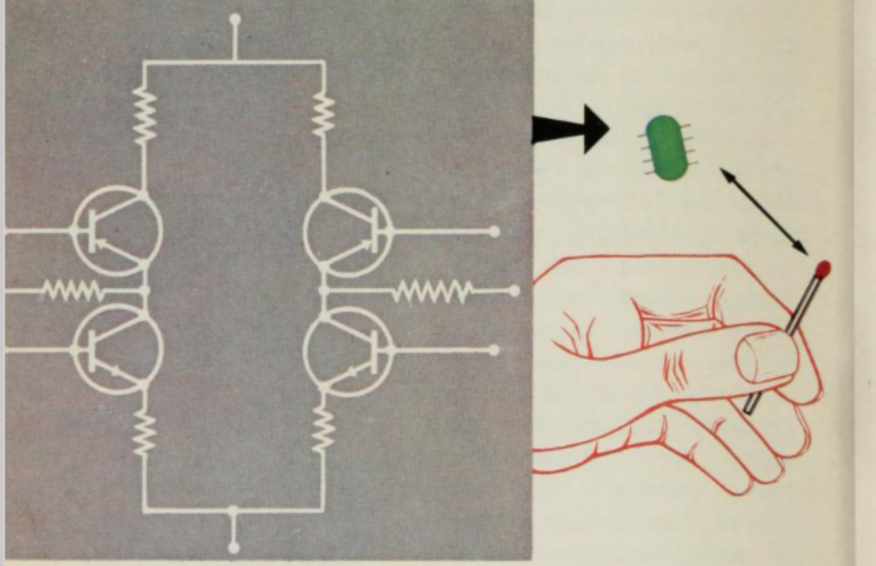

299. Thin film multivibrator

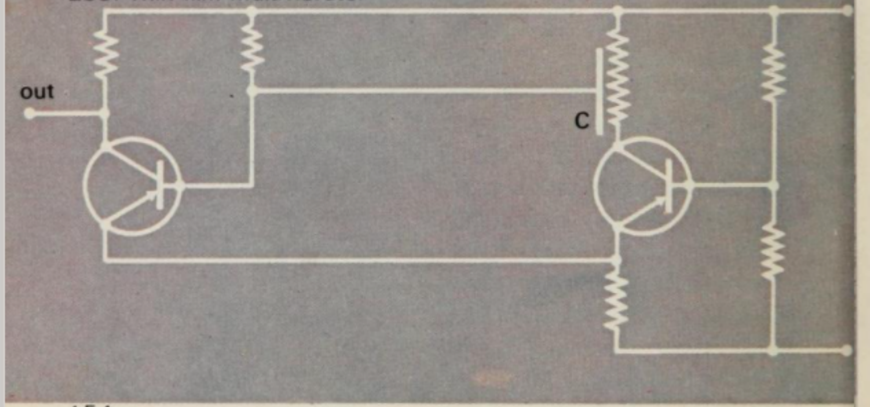

Figure 299 is a thin film multivibrator or oscillator, which requires only the distributed capacitance resistor C for feedback, in conjunction with directly connected emitters.

Thin metal film resistors can be cut by electron beam. A powerful discharge tube produces a beam concentrated electromagnetically to a point 0.0005 in. in diameter. The beam is moved by magnetic deflection [300]. The process is carried out by automatic mechanisms [301].

It is interesting to imagine what electronics may accomplish in the future. A tidy picture of a central nucleus surrounded by negative electrons racing around at tremendous speed does well for many purposes. But much is still unknown. Many discoveries will be made.

300. Thin film resistors cut by electron beam

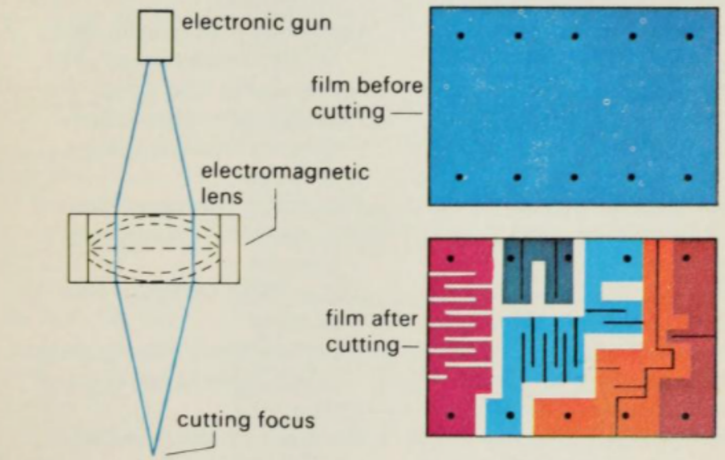
electronic gun
film before cutting
electromagnetic lens
film after cutting
cutting focus

301. Automated production of electronic units

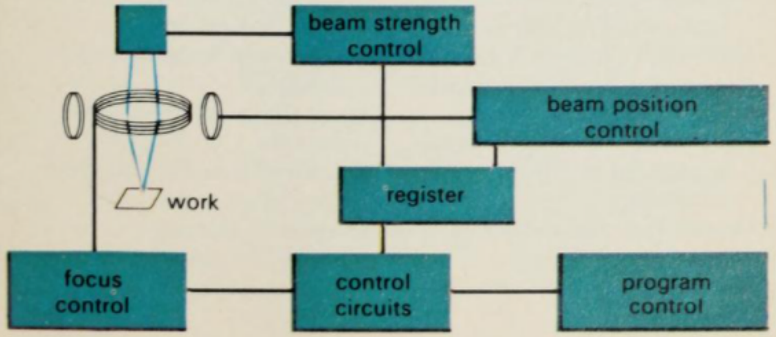
beam strength control
beam position control
work
register
focus control
control circuits
program control

NO. 9
IN A SERIES
KNOWLEDGE THROUGH COLOR

Bantam Books presents a unique home reference library that will give your family clear, concise information on a vast range of subjects in the fascinating world of knowledge. This invaluable series encompasses many fields, including the natural sciences, the arts, cultural history, popular and space science, transportation, history, horticulture, animals and hobbies. Each handy volume has been thoroughly researched and written in plain language and packed with hundreds of specially created *full-color* illustrations. ************************

KNOWLEDGE THROUGH COLOR — a home library in one handsome paperback series. Perfect as source material for student papers. Useful as guidebooks. Good for general reading or just plain browsing. KNOWLEDGE THROUGH COLOR — Learning in the home! ************************

ELECTRONICS FOR KIDS

PLAY WITH SIMPLE CIRCUITS AND EXPERIMENT WITH ELECTRICITY!

ØYVIND NYDAL DAHL

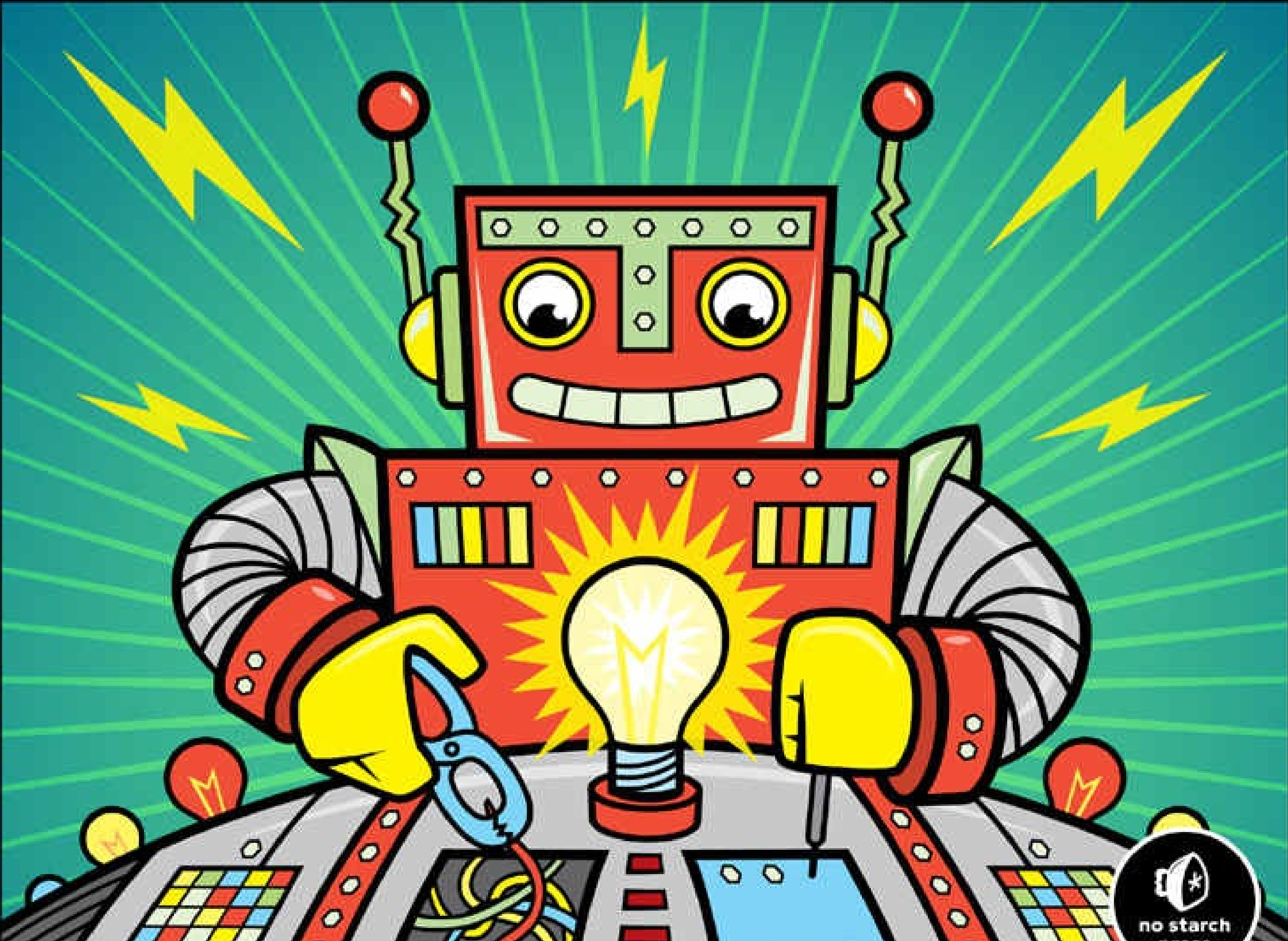

no starch press

ELECTRONICS FOR KIDS

PLAY WITH SIMPLE CIRCUITS AND EXPERIMENT WITH ELECTRICITY!

ØYVIND NYDAL DAHL

no starch press
San Francisco

ELECTRONICS FOR KIDS. Copyright © 2016 by Øyvind Nydal Dahl.

All rights reserved. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage or retrieval system, without the prior written permission of the copyright owner and the publisher.

Printed in Canada

First printing

20 19 18 17 16   1 2 3 4 5 6 7 8 9

ISBN-10: 1-59327-725-3
ISBN-13: 978-1-59327-725-3

Publisher: William Pollock
Production Editor: Riley Hoffman
Cover Illustration: Garry Booth
Interior Design: Beth Middleworth
Developmental Editor: Jennifer Griffith-Delgado
Technical Reviewer: John Hewes
Copyeditor: Julianne Jigour
Compositor: Riley Hoffman
Proofreader: Paula L. Fleming

For information on distribution, translations, or bulk sales, please contact No Starch Press, Inc. directly:
No Starch Press, Inc.
245 8th Street, San Francisco, CA 94103
phone: 415.863.9900; info@nostarch.com
www.nostarch.com

Library of Congress Cataloging-in-Publication Data

Names: Nydal Dahl, Øyvind, author.
Title: Electronics for kids : play with simple circuits and experiment with electricity! / by Øyvind Nydal Dahl.
Description: San Francisco : No Starch Press, [2016] | Audience: Ages 10+ | Includes index.
Identifiers: LCCN 2015048986 (print) | LCCN 2016005706 (ebook) | ISBN 9781593277253 (pbk.) | ISBN 1593277253 (pbk.) | ISBN 9781593277475 (epub) | ISBN 1593277474 (epub) | ISBN 9781593277482 (mobi) | ISBN 1593277482 (mobi)
Subjects: LCSH: Electronics--Juvenile literature. | Electronic circuits--Juvenile literature.
Classification: LCC TK7820 .N93 2016 (print) | LCC TK7820 (ebook) | DDC 621.381--dc23
LC record available at http://lccn.loc.gov/2015048986

No Starch Press and the No Starch Press logo are registered trademarks of No Starch Press, Inc. Other product and company names mentioned herein may be the trademarks of their respective owners. Rather than use a trademark symbol with every occurrence of a trademarked name, we are using the names only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

The information in this book is distributed on an “As Is” basis, without warranty. While every precaution has been taken in the preparation of this work, neither the author nor No Starch Press, Inc. shall have any liability to any person or entity with respect to any loss or damage caused or alleged to be caused directly or indirectly by the information contained in it.

ABOUT THE AUTHOR

Øyvind Nydal Dahl has been an electronics enthusiast since he was a kid—he's always loved figuring out how things worked so he could try to build them for himself. He studied electronics and computer science at the University of Oslo, where he received a master's degree after building his own microchip. He then co-founded the company Intelligent Agent to develop sensors that allow robots to see through walls.

After a few years at Intelligent Agent, Øyvind set out on a mission to teach the world electronics. He gives workshops, develops courses, and writes about electronics and technology for a variety of outlets. He's posted hundreds of articles, tutorials, and videos on his blog (*http://www.build-electronic-circuits.com/*) and maintains Ohmify (*http://ohmify.com/*), a membership site that makes learning electronics fun and easy.

ABOUT THE TECHNICAL REVIEWER

John Hewes began connecting electrical circuits at an early age, moving on to electronics projects as a teenager. He later earned a physics degree and continued to develop his interest in electronics, helping school students with their projects while working as a science technician.

John has taught electronics and physics up to an advanced level in the United Kingdom and ran a school electronics club for children aged 11 to 18 years, setting up the website *http://www.electronicsclub.info/* to support the club. He believes that everyone can enjoy building electronics projects, regardless of their age or ability.

BRIEF CONTENTS

Foreword by Joe Grand
Acknowledgments
Introduction

PART 1: PLAYING WITH ELECTRICITY

Chapter 1: What Is Electricity?
Chapter 2: Making Things Move with Electricity and Magnets
Chapter 3: How to Generate Electricity

PART 2: BUILDING CIRCUITS

Chapter 4: Creating Light with LEDs
Chapter 5: Blinking a Light for the First Time
Chapter 6: Let’s Solder!
Chapter 7: Controlling Things with Electricity
Chapter 8: Building a Musical Instrument

PART 3: THE DIGITAL WORLD

Chapter 9: How Circuits Understand Ones and Zeros
Chapter 10: Circuits That Make Choices
Chapter 11: Circuits That Remember Information
Chapter 12: Let’s Make a Game!

Handy Resources
Index

CONTENTS IN DETAIL

Foreword by Joe Grand
Acknowledgments

INTRODUCTION

- About This Book
  - Who Should Read This Book
  - How to Read This Book
  - What's in This Book?
- Your Electronics Lab
  - Useful Supplies
  - Safety First!

PART 1: PLAYING WITH ELECTRICITY

1
WHAT IS ELECTRICITY?

- Project #1: Turn on a Light!
  - Shopping List
  - Step 1: Inspect the Light Bulb
  - Step 2: Connect the Light Bulb to the Battery
- How Does Electricity Light a Bulb?
  - What Is an Electron?
  - Voltage Pushes Electrons
  - Current Flows
  - Resistance Reduces Current
  - Lighting the Bulb
- How Is a Circuit Like a Pipe System?
- Meet the Switch
- Project #2: Intruder Alarm
  - Shopping List
  - Tools
  - Step 1: Does the Buzzer Beep?
  - Step 2: Prepare the Aluminum
  - Step 3: Foil Your Door
  - Step 4: Prepare a Trigger Wire
  - Step 5: Connect the Buzzer and Trigger Wire

- Step 6: Mount the Buzzer and Trigger Wire
- Step 7: Add a Power Source
- Step 8: Stage an Intruder Alert!
- Step 9: What If the Intruder Alarm Doesn’t Work?

What’s Next?

2
MAKING THINGS MOVE WITH ELECTRICITY AND MAGNETS

How Magnets Work
- *Try It Out: Find Some Magnetic Objects!*

Meet the Electromagnet

Project #3: Create Your Own Electromagnet
- Shopping List
- Tools
- Step 1: Check Your Bolt
- Step 2: Remove Insulation from One End of the Coil Wire
- Step 3: Wind the Wire
- Step 4: Connect the Negative Battery Terminal to the Coil
- Step 5: Connect the Switch
- Step 6: Test Your Super Electromagnet
- Step 7: What If the Electromagnet Isn’t Working?

Meet the Motor

Project #4: Create a Motor
- Shopping List
- Tools
- Step 1: Create the Rotor
- Step 2: Build the Motor’s Structure
- Step 3: Place the Magnets
- Step 4: Reinsulate Part of the Coil
- Step 5: Rev Up Your Motor
- Step 6: What If the Motor Doesn’t Work?

What’s Next?

3
HOW TO GENERATE ELECTRICITY

Generating Electricity with Magnets
- A Changing Magnetic Field Creates Electricity
- How Does a Generator Work?
- *Creating Electricity from Water or Wind*

Meet the Multimeter
- How to Measure Voltage

- What Are AC and DC?

Project #5: Make a Shake Generator
- Shopping List
- Tools
- Step 1: Prepare Your Tube
- Step 2: Wind Your Coil
- Step 3: Connect the Multimeter
- Step 4: Shake That Thing!
- Step 5: What If There’s No Voltage?
- *Try It Out: Using a Motor as a Generator*

How Do Batteries Work?
- What’s Inside a Battery?
- The Chemistry Behind Batteries
- What Determines a Battery’s Voltage?

Project #6: Turn On a Light with Lemon Power
- Meet the LED
- Shopping List
- Tools
- Step 1: Prepare Your Wires
- Step 2: Insert Electrodes into a Lemon
- Step 3: Create Four Lemon Batteries
- Step 4: Connect the Lemons in Series
- Step 5: Test Your Lemon Battery
- *Try It Out: More Food Batteries!*
- Step 6: What If Your Lemon Light Doesn’t Work?

What’s Next?

PART 2: BUILDING CIRCUITS

4
CREATING LIGHT WITH LEDS

Meet the Resistor
- Resistor Color Codes
- *How to Write Large Values*
- What Are Resistors Made Of?
- Resistors Control Current and Voltage

Introducing Ohm’s Law

Project #7: Let’s Destroy an LED!
- Shopping List
- Step 1: Identify Which LED Leg Is Which

Step 2: Break That LED!
Step 3: What If Nothing Happens to the LED?
How to Use an LED Correctly
How to Write Small Values
Protecting Your LED with a Resistor
Calculating the Resistance You Need
Project #8: Powering an LED
Shopping List
Step 1: Twist the Resistor and LED
Step 2: Wire the Battery Clip
Step 3: Let There Be Light!
Step 4: What If the LED Doesn’t Work?
Building Circuits on a Breadboard
How to Connect Components and Wires
Wires to Use on a Breadboard
Project #9: Your First Breadboard Circuit
Shopping List
Step 1: Place the Resistor
Step 2: Place the LED
Step 3: Place the Battery Clip
Step 4: What If the LED Doesn’t Work?
What’s Next?

5
BLINKING A LIGHT FOR THE FIRST TIME

Meet the Capacitor
How Capacitors Work
Polarized and Nonpolarized Capacitors
Capacitor Values
Project #10: Test a Capacitor
Shopping List
Step 1: Start with the LED Circuit
Step 2: Add the Capacitor
Step 3: Charge the Capacitor
Step 4: Use the Capacitor to Light the LED
Step 5: What If the Circuit Doesn’t Work?
Describing Circuits with Symbols
Meet the Relay
Using the Relay to Blink a Light
Slowing Down the Blinking
Project #11: Blink a Light!

- Shopping List
- Step 1: Identify the Relay Pins
- Step 2: Make the Relay Switch Fast
- Step 3: Make the Relay Stay On Longer
- Step 4: Make the Relay Stay Off Longer
- Step 5: Add the LED and Resistor
- Step 6: What If the LED Won’t Blink?
- *Try It Out: Make Your Intruder Alarm More Effective*

What’s Next?

6
LET’S SOLDER!

How to Solder
- Soldering Safety Tips
- Heat the Soldering Iron
- Clean the Soldering Iron Tip
- Tin the Soldering Iron Tip
- Heat Both the Pin and the Pad
- Add Solder
- Remove the Soldering Iron
- Watch Out for Bad Solder Joints!

Project #12: Solder Your First LED Circuit
- Shopping List
- Tools
- Step 1: Place the Components
- Step 2: Bend the Component Legs
- *How to Read the Circuit Board Illustrations*
- Step 3: Heat and Clean the Soldering Iron
- Step 4: Solder the Resistor and LED
- Step 5: Trim the Legs
- Step 6: Solder the Battery Clip
- Step 7: Let There Be Light!
- Step 8: What If the Soldered LED Circuit Doesn’t Work?

Oops! How Do I Remove a Soldered Component?

Project #13: Desolder the Battery Clip
- Shopping List
- Tools
- Step 1: Heat the Soldering Iron
- Step 2: Place the Solder Wick on the Solder Joint
- Step 3: Heat the Solder Joint and the Solder Wick
- Step 4: Trim the Used Solder Wick

  - Step 5: Remove the Other Battery Clip Wire
  - *Try It Out: Solder More Stuff!*
- What's Next?

7
CONTROLLING THINGS WITH ELECTRICITY

- Meet the Transistor
  - Why Use a Transistor?
  - How the Transistor Works
  - Controlling an LED with a Transistor
- Project #14: Build a Circuit that Senses Touch
  - Shopping List
  - Tools
  - Step 1: Place Components on the Prototyping Board
  - Step 2: Check Your Component Placement
  - Step 3: Solder the Components and Trim Excess Legs
  - Step 4: Solder the Touch Pad
  - Step 5: Power It Up!
  - Step 6: Test the Sensor
  - Step 7: What If the Touch Sensor Doesn't Work?
  - *Try It Out: Can the Touch Sensor Detect Different Touches?*
- Resistors That Can Change Value
  - Meet the Potentiometer
  - Meet the Photoresistor
- Dividing a Voltage with Resistors
  - What Does a Voltage Divider Look Like?
  - Calculating the Voltage from a Voltage Divider
  - How a Voltage Divider Can Help Measure Light
- Project #15: Build a Sunrise Wake-Up Alarm
  - Shopping List
  - Tools
  - Step 1: Place Components on the Prototyping Board
  - Step 2: Solder the Components and Trim the Legs
  - Step 3: Add the Buzzer to the Board
  - Step 4: Make the Remaining Connections with Wires
  - Step 5: Add the Battery Clip to the Board
  - Step 6: Set a Wake-Up Call
  - Step 7: What If There's No Sound?
  - *Try It Out: Temperature-Controlled Fan*
- What's Next?

8
BUILDING A MUSICAL INSTRUMENT
- Meet the Integrated Circuit
- ICs and Datasheets
- How to Make Sound with Electricity
- Sounds That Humans Can Hear
- Meet the 555 Timer
  - Powering an IC
  - How to Set the Output Speed of the 555 Timer
- Project #16: Make Your Own Sound with the 555 Timer
  - Shopping List
  - Step 1: Place the 555 Timer on the Breadboard
  - Step 2: Set the Frequency
  - Step 3: Connect the Speaker and Coupling Capacitor
  - Step 4: Connect the Power and Reset Pins
  - Step 5: Make Some Sound!
  - Step 6: What If There’s No Sound?
  - *Checking Connections as a Team*
- Turning an Annoying Beep into Music
- Project #17: An Instrument That Beeps and Boops
  - Shopping List
  - Step 1: Connect the 555 Timer and the Capacitors
  - Step 2: Connect Jumper Wires
  - Step 3: Connect the Note Controller and Resistor
  - Step 4: Add the “On” Button
  - Step 5: Add the Loudspeaker
  - Step 6: Play Some Music!
  - Step 7: What If the Instrument Doesn’t Work?
  - *Try It Out: Make a Motion-Controlled Instrument*
- What’s Next?

PART 3: THE DIGITAL WORLD

9
HOW CIRCUITS UNDERSTAND ONES AND ZEROS
- Ones and Zeros as Voltages
- Meet the Binary Number System
- Project #18: Convert from Binary to Decimal
  - Tools
  - Step 1: Write It Down on Paper

  - Step 2: Write the Position Values
  - Step 3: Find the Value of Each Digit
  - Step 4: Add the Numbers
  - *Try It Out: Transform More Binary Numbers*
- Bits and Bytes
- Numbers Can Be Anything
- Project #19: Color Guessing Game
  - Meet the RGB LED
  - Shopping List
  - Step 1: Place the Color Chooser’s Push Buttons
  - Step 2: Connect the RGB LED
  - Step 3: Connect the Color Revealer Button
  - Step 4: Test the Colors
  - Step 5: What If the Game Doesn’t Work?
  - *Try It Out: Solder the Color Guessing Game*
- How Binary Numbers Can Create Words
- Project #20: The Secret Message Machine
  - Meet the DIP Switch
  - Shopping List
  - Step 1: Connect the Push Button
  - Step 2: Connect the DIP Switch
  - Step 3: Connect the LEDs
  - Step 4: Send a Secret Message!
  - Step 5: What If the Secret Message Machine Doesn’t Work?
  - *Why Computers Use Ones and Zeros*
- What’s Next?

10
CIRCUITS THAT MAKE CHOICES

- It’s Only Logical
- Meet the Logic Gates
  - AND Gates Check for Two True Inputs
  - OR Gates Check for One True Input
  - NOT Gates Flip Inputs
  - A Bigger AND Gate
- How to Draw Logic Circuit Diagrams
  - A Logic Equation for a Secret Code
  - Converting a Logic Equation into a Circuit Diagram
  - *Try It Out: Draw More Logical Statements as Circuits*
- Using Logic Gates in Real Life
  - *More About Current, Components, and Transistors*

Project #21: A Secret Code Checker
    - Shopping List
    - Tools
    - How to Use Other Voltages with a Breadboard
    - *Making Circuits More Reliable*
    - Step 1: Place the Switches and Resistors
    - Step 2: Place the ICs
    - Step 3: Place the Transistor and LED
    - Step 4: Build the Logic Circuit
    - Step 5: Finish Wiring the Transistor
    - *Protecting Your Components*
    - Step 6: Power and Test the Secret Code Checker
    - Step 7: What If the LED Doesn’t Light Up?
    - *Try It Out: Disarm Your Intruder Alarm*

Negative Logic Gates
    - NAND Looks for One False Input
    - NOR Looks for Two False Inputs

What’s Next?

11
CIRCUITS THAT REMEMBER INFORMATION

Saving One Bit at a Time

A Better Memory Circuit

Memory That Changes Only at a Certain Time

An Output That Toggles

Project #22: An Electronic Coin Tosser
    - Shopping List
    - Step 1: Build the Oscillator Circuit
    - Step 2: Add the Start Button
    - Step 3: Build the Toggle Circuit
    - Step 4: Add the Heads and Tails LEDs
    - Step 5: Toss that “Coin”!
    - Step 6: What If the Coin Tosser Doesn’t Work?

What’s Next?

12
LET’S MAKE A GAME!

Meet the Reaction Game Circuits
    - Meet the V_{CC} and GND Symbols
    - *Why Is It Called V_{CC}?*
    - A 555 Timer to Set the Light Speed

    - A Counter to Turn the LEDs On
    - A Latch to Start and Stop the Light
- Project #23: An LED Reaction Game
    - Shopping List
    - Tools
    - Step 1: Build the 555 Timer Circuit
    - Step 2: Build the LED-Controlling Circuit
    - Step 3: Build the Start and Stop Circuit
    - Step 4: Practice Your Reaction Time!
    - Step 5: What If the Game Isn’t Working?
    - *Try It Out: Change the Light’s Speed*
- Add a Buzzer to Your Game
- What’s Next? Go Make Cool Stuff!

HANDY RESOURCES

- Component and Unit Value Cheat Sheets
    - Resistor Color Codes
    - Capacitor Codes
    - Standard Prefixes
- A Quick Review of Ohm’s Law
- A Basic Voltage Divider Circuit
- Online Electronics Shops
- Online Resources

INDEX

FOREWORD

There’s something special about bringing a project to life that you read about in a book or that started as an idea in your head. And sometimes the simplest things are the most satisfying.

One of my favorite childhood projects was a mischievous little device made of a single resistor connected between the tip and ring of a telephone line. I used a piece of one-sided copper circuit board with rub-off symbols to lay out the design, and then I etched the unprotected copper away using ferric chloride in my basement. You could still use the phone normally to make outgoing calls, but anyone calling the house would receive a busy signal. This was the perfect way to make sure my parents didn’t receive any phone calls from my teachers during dinner!

A few years later, I modified a garage door opener to open any door of the same brand. In normal operation, the passwords on the transmitter and receiver were manually set with a series of 10 DIP switches. If the transmission signal matched what the receiver was expecting, then the garage door would open. I replaced the switches on my transmitter with a common 555 timer IC, to generate a clock signal, and a 10-stage binary counter, a type of digital logic device, to automatically try every single possible combination (that’s 2^{10} or 1,024 attempts). Within a few minutes of holding down the button, the correct password would be transmitted and the garage door would open! I never used my universal “brute-force” garage door opener for malicious purposes, but it reinforced my hacker mindset—solving problems with unconventional solutions, pushing the limits of technology, harming no one, and learning through constant questioning and experimentation. I also thought it was pretty cool to be able to modify an off-the-shelf device and make it do something the original designers probably never anticipated.

When I was much younger, I somehow ended up with a 6 V lantern battery and a spring from an adjustable lamp. I wondered, “What would happen if I connected the spring between the battery terminals?” So of course, I tried it. The spring got hotter and hotter until I freaked out, plucked it off the terminals, and threw it into the bathroom sink. I had created a short circuit by connecting the positive and negative terminals of the battery together, causing current to flow between them. I never looked at batteries and springs the same way again.

I remember trying to build my own alarm system for my bedroom door, sort of a low-tech version of the one you’ll build in Chapter 1. I hung an old AM/FM radio from a hook on the back of my door, tuned it to static, turned the volume up to maximum, and “armed” it by connecting the sliding power switch to a wire I had attached to my wall. In theory, when the door opened, the wire would pull the switch and turn on the radio, blasting white noise at the intruder. That didn’t happen. Instead, when my dad opened the door, the radio slid off the hook and crashed onto the floor. Back to the drawing board on that one!

These stories are meant to do one thing: inspire you to explore the wonderful, wild world of electrons—and this book is the perfect launch pad! Øyvind breaks down complex electronics fundamentals in an enjoyable, fun way. His passion for electronics and his love for teaching shine on every page. Starting with the basics and building up from there, you’ll end up with the power to create bigger, better, faster, and more intelligent projects on your own. There’s no better way to learn than by *doing*. So go ahead, turn the page and begin your adventure into all that electronics has to offer!

Joe Grand
Product Designer, Hardware Hacker, and Daddy
Portland, Oregon

ACKNOWLEDGMENTS

First of all, thanks to my father for explaining how things work based on practice instead of theory when I was a kid. His great explanations got me started in the world of electronics. Also, many thanks to my mother, who had to endure all those technical discussions around the dinner table.

Thanks to Jennifer Griffith-Delgado, Riley Hoffman, Tyler Ortman, and the rest of the team at No Starch Press—first of all for believing in me, but also for guiding me through the editorial process in such a good way. You have been a dream to work with!

Thanks to my technical reviewer, John Hewes, for finding my errors, challenging me in some areas, and making me think through some parts of the book a few extra times.

Finally, a special thanks to Garry Booth for the cover illustration, to Beth Middleworth for designing the layout and background illustrations, and to Riley once more, for drawing the technical diagrams. Those three really made this book come alive.

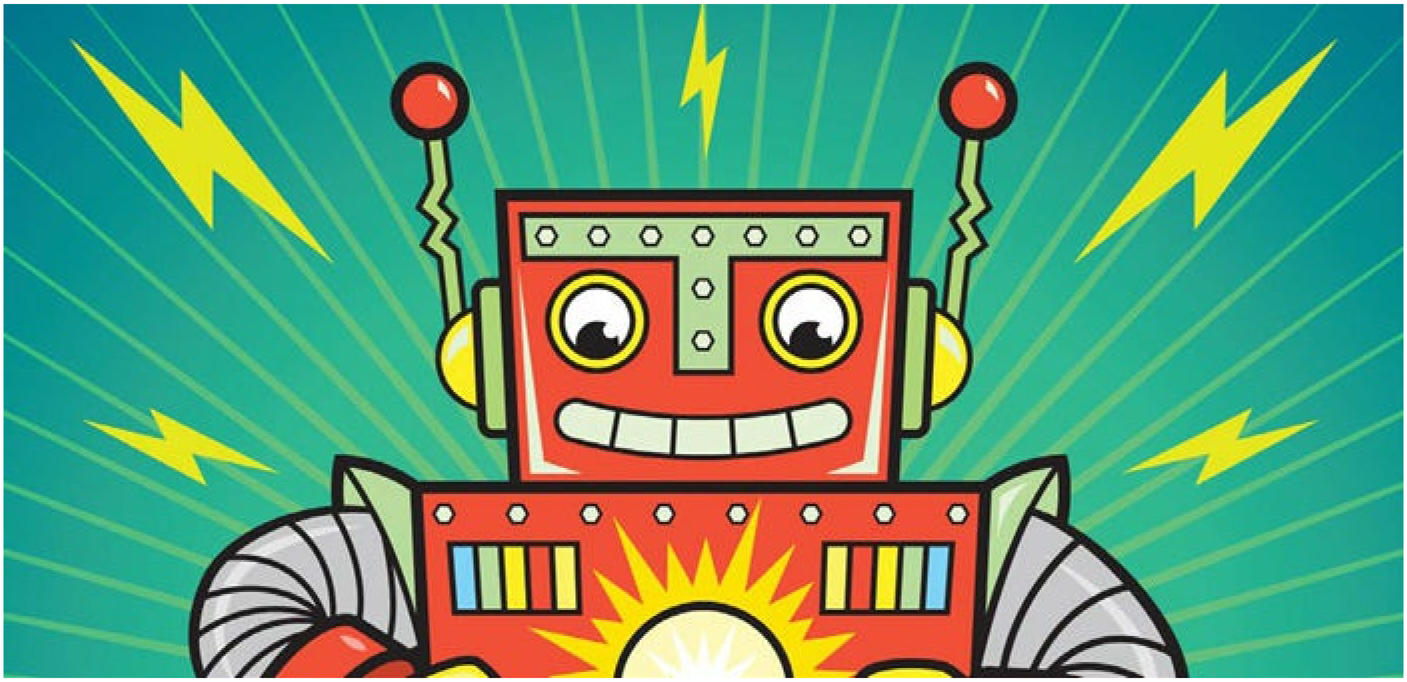

INTRODUCTION

Welcome to *Electronics for Kids*! This book will teach you how to make cool things by putting together the same parts that are inside televisions, electronic toys, radios, and all the other gadgets in the world. You'll build fun experiments, like a light powered by lemons, as well as useful (but still fun) projects, like an intruder alarm and a musical instrument.

You'll do more than just follow directions, however: you'll also learn how every component in each project works.

My hope is that when you know how those components work, you'll see how to create your own inventions by combining the components in different ways. Blinking a light is one of the first things I learned how to do with electronics. When I saw how that worked, a whole new world suddenly opened up to me. Since then, I've built robots, music players, miniature computers, and even a device that lets you see through a wall! With practice, you can build those things, too—and this book will teach you the basic skills you need to start the journey.

ABOUT THIS BOOK

When I was about 14 years old, I thought computers were cool, but I had no idea how they worked. They seemed magical, and I thought I'd never understand them or be able to build one. Luckily, my dad was an engineer, and he had a very good way of explaining things. When I asked questions, he showed me not only how things worked but also how I could build something similar myself.

I wrote this as the book I would have loved to have had as a kid, and I hope you enjoy it!

Who Should Read This Book

If you've ever looked at an electronic gadget and thought, "How does that work?" or "How can I make that?" as I did, then you're in the right place. Whether you're 8 or 100 years old, as long as you're curious and have a playful spirit, this book is for you.

How to Read This Book

I recommend you read this book in order, because every chapter builds upon concepts and skills covered in previous chapters.

Each chapter has at least one hands-on project. Build these projects! Electronics is a very practical skill, and reading about how a component works or what a project should do is different from experiencing it yourself. Just be sure to read a project in full before you dive into it so you understand the steps involved.

If you encounter problems as you build a project, don't worry: that happens to everyone at some point when working with electronics—even me. Just keep at it, study your circuit, and rewire the whole project if needed to get it working. When you've been battling to get a circuit working for a couple of hours, then suddenly find the error and your circuit works, you'll feel amazing! If you get stuck, grab a friend or family member and ask them to help out.

If there are parts of the book you don't understand right away, I recommend you keep on reading. Don't let details stop you. Come back to that particular topic later when you have some more projects under your belt.

What's in This Book?

As you work through this book, you'll build your knowledge of electronics gradually, starting with basic—but essential—information and simple circuits. After the basics, you'll build more complex circuits and meet components like resistors, capacitors, transistors, and integrated circuits. To see how the components work and to understand electronics in a practical way, you'll build fun projects in every chapter.

At the end of the book, you'll build one final, epic project: a game to play with your friends. By then, you'll have enough experience and knowledge to modify the game or even build a totally new game you invent yourself!

This book is divided into three parts. **Part 1: Playing with Electricity** is the foundation for the rest of the book. It's all about fundamental knowledge and how electricity actually works.

- **Chapter 1: What Is Electricity?** introduces the science behind electricity and describes the basic requirements for a circuit to turn something on.
- **Chapter 2: Making Things Move with Electricity and Magnets** shows you how you can move objects with electricity. In this chapter, you'll build a motor from scratch.
- **Chapter 3: How to Generate Electricity** describes how batteries and power plugs in the wall provide electricity. Of course, you'll build your own electricity sources, too!

Part 2: Building Circuits is where you'll really get your hands dirty. You'll meet some of the most important components in electronics, and you'll learn how to build both permanent and temporary circuits.

- In **Chapter 4: Creating Light with LEDs**, you'll build circuits on a breadboard for the first time to create a *prototype*, which is just a temporary circuit. You'll learn about resistors, light-emitting diodes (LEDs), and how to use those parts together.
- **Chapter 5: Blinking a Light for the First Time** shows how two new components, capacitors and relays, work. You'll even combine these with an LED to create a circuit that blinks a light.
- **Chapter 6: Let's Solder!** teaches you how to solder. With soldering, you can transform a circuit from a prototype to a proper device that will last for years to come.
- **Chapter 7: Controlling Things with Electricity** introduces the transistor, a component that lets a circuit control other circuits. You'll learn how transistors work and how to use them to build a touch sensor and a simple alarm clock.
- In **Chapter 8: Building a Musical Instrument**, you'll learn what an integrated circuit is and how circuits can make sound. You'll combine this knowledge to build a musical instrument.

Part 3: The Digital World introduces digital electronics, which almost all modern technology is based upon.

- In **Chapter 9: How Circuits Understand Ones and Zeros**, you'll learn about 1s and 0s, bits and bytes, and how to use them to communicate.
- **Chapter 10: Circuits That Make Choices** teaches you how to build smart circuits that use logic to make decisions. You'll build a secret code checker and learn how you can combine it with your intruder alarm.
- **Chapter 11: Circuits That Remember Information** shows how you can use logic gates to create circuits that remember information in a way similar to a computer. Then, you'll use this to create an electronic coin tosser.
- **Chapter 12: Let's Make a Game!** is dedicated to one large project. You'll get to show off your new skills by combining all the knowledge from the book to make a reaction speed game.

Finally, you'll find a **Handy Resources** appendix at the back of the book, which includes cheat sheets for figuring out component values, doing some essential electronics calculations, and so on. You'll learn about those concepts in detail throughout the book, but even electronics experts need a quick reference every now and then!

YOUR ELECTRONICS LAB

The wonderful thing about electronics projects is that your "lab" can be anywhere you want—it doesn't have to be a garage or workshop. All you need is a flat surface to work on, with enough room for your tools and components. Just gather the supplies to build your latest invention, and you're set.

Each project in this book includes a convenient list of the electronic components and tools needed to build it. Before you dig into a project, check its Shopping List to make sure you have all the materials. I've also created a complete list of all the components and tools you'll need for all the projects in this book, which you can find linked from the book's web page at *https://www.nostarch.com/electronicsforkids/*. This list should always have the most up-to-date part numbers and links to kits you can buy that contain all the necessary components.

Useful Supplies

Whether you're building the projects in this book or other projects on your own, there are a few supplies that will always come in handy:

- **A digital multimeter** (Jameco #2206061, Bitsbox #TL057, Rapid Electronics #55-6662) for testing connections and making sure a project is working correctly.
- **A pair of wire cutters** (Jameco #35482, Bitsbox #TL008)
- **A big spool of insulated wire** (Jameco #36792, Bitsbox #W106BK)
- **Electrical tape** to protect bare wires or fasten stuff.
- **9 V batteries**—nearly every project in the book uses one!
- **A bunch of LEDs** (Jameco #18041, Bitsbox #K033)

- **A bunch of resistors** (Jameco #2217511, Bitsbox #K017)
- **Safety glasses** to wear when snipping component leads, stripping wires, or soldering.

You can buy most of these from your local hardware store or from any online electronics retailer, like Jameco (*http://www.jameco.com/*), SparkFun (*http://www.sparkfun.com/*), or Bitsbox (*http://www.bitsbox.co.uk/*). Check out "Online Electronics Shops" on page 286 for more options.

You might also want to have a pair of scissors, some scrap paper, and pencils to take notes.

Safety First!

All the circuits in this book use a low voltage, and they're not dangerous to build and play with. That said, there are a few safety tips to keep in mind when using electronic components and tools:

- Wear safety glasses when trimming components or soldering.
- Use tools only for their intended purpose. Soldering irons are hot, and wire cutters are sharp—if used improperly, they can hurt you. If you're confused about how to use a tool, ask an adult for help.
- An adult should supervise younger children when they're working with small components, solder, tools, and so on to teach them how to use everything safely.
- Keep electronic parts out of reach of babies and very young children.
- Most projects in this book use batteries, but some do use power from a wall outlet. Follow the instructions for those circuits carefully. Never plug components directly into a power outlet, or you will get hurt.

Some projects do have steps you should take special care with, and I will clearly state that in the instructions with a warning, like this:

WARNING

When you see this type of note, be careful with the step it talks about.

Electronics is a safe hobby, though, so you won't see very many of these warnings. When you do see one, don't let it stop you from having fun. If you use common sense and follow the directions, you'll have nothing to worry about.

Now let's get started!

PART 1

PLAYING WITH ELECTRICITY

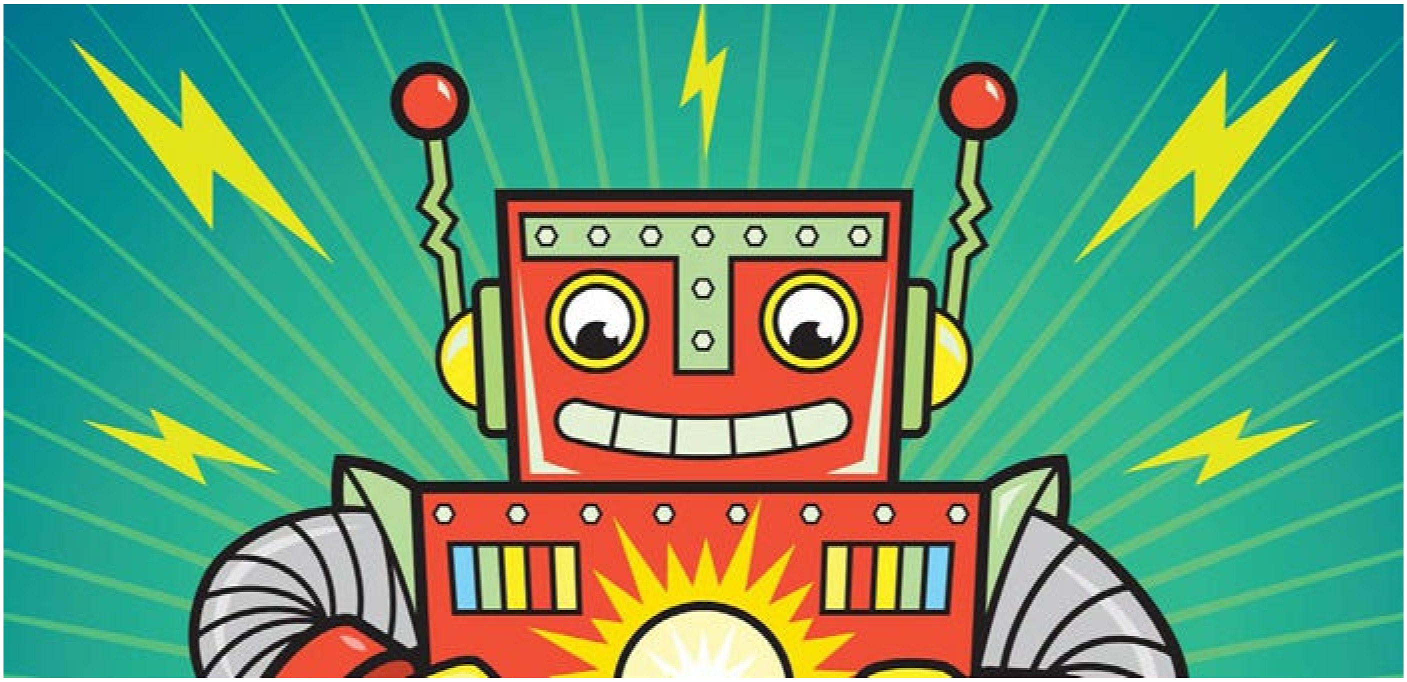

1

WHAT IS ELECTRICITY?

Push a button on a music player, and a song suddenly comes out of the speakers. Push a button on a TV's remote control, and your favorite shows come to life instantly. These wonders happen thanks to the magic of *electricity*, a type of energy that powers all the technology in your home. By the time you finish this book, you'll be an electronics wizard, and then you can try using your powers to build any invention you can imagine!

This book is all about understanding electricity and using it to make amazing things. In this chapter, we'll explore how electricity works, and then you'll build a complete electronics project: a burglar alarm that warns you if intruders have entered a room. Once you get the hang of using electricity, you can build all sorts of fun contraptions, like a musical instrument or a light-up game to play with your friends. In fact, you'll build these in this book.

PROJECT #1: TURN ON A LIGHT!

When you flip the light switch in a room, the bulb brightens right away. Let's look at how electricity makes that bulb shine, starting with a little experiment.

Shopping List

For this project, you'll need the following parts:

- **A standard 9 V battery** to power the circuit.
- **A small, incandescent light bulb** rated for 9 to 12 V (DigiKey #CM394-ND, Bitsbox #OP037, or a similar light bulb from a hardware store).

Step 1: Inspect the Light Bulb

Look closely at your light bulb; you should see a thin metal wire filament inside the glass. One end of this filament is connected to the metal side of the base, and the other end is connected to the metal contact on the bottom.

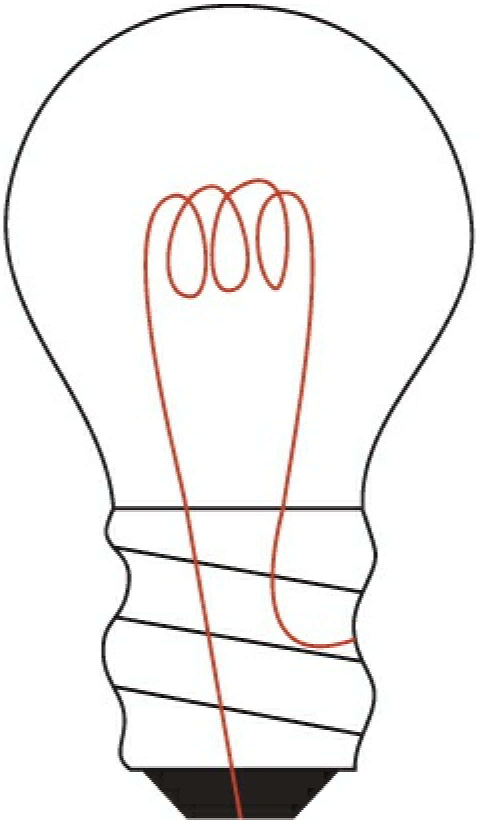

Step 2: Connect the Light Bulb to the Battery

Place your 9 V battery upright on a table. Take the light bulb and gently place it so that the bottom point touches one battery terminal and the metal side touches the other battery terminal. When both the bottom and the side are touching the battery, the bulb should light up.

Congratulations: you just generated light with electricity! The bulb lights because when you touch it to the battery contacts, electricity runs through the wire filament inside. The filament then heats and starts to glow, creating light.

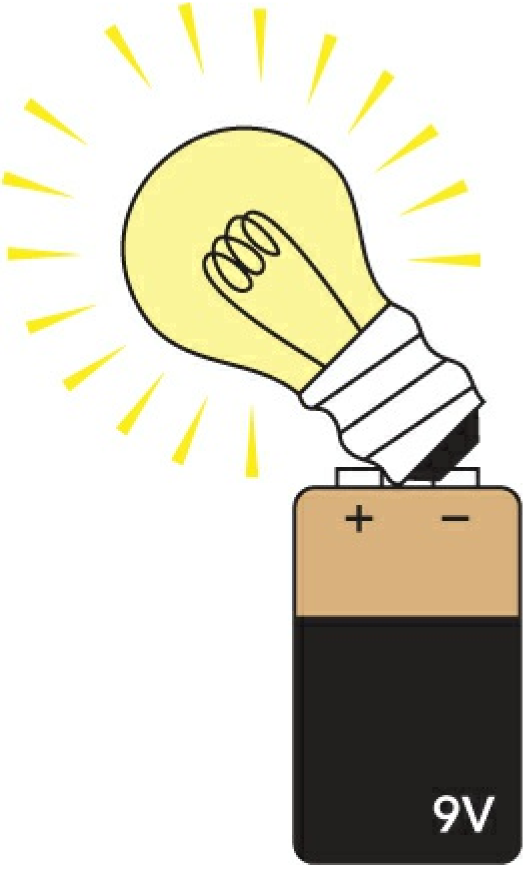

HOW DOES ELECTRICITY LIGHT A BULB?

But *how* does electricity cause the wire to heat up, and why does the light turn on instantly? There are four concepts that combine to make that happen:

- Electrons
- Current
- Voltage
- Resistance

These fundamental concepts of electricity all depend on each other, and we’ll explore them in this section.

What Is an Electron?

Everything you see around you is made of *atoms*, which are particles so small you can’t see them without a special type of microscope. But atoms are made of even smaller particles, called *protons*, *neutrons*, and *electrons*.

Protons and neutrons form an atom’s *nucleus* (its center), and electrons orbit the nucleus like planets orbiting the sun. Protons and electrons are both *electrically charged*: protons have a positive charge, and electrons have a negative charge. That’s why the electrons stick with an atom in the first place. The positive and negative charges act like opposite sides of a magnet and attract each other.

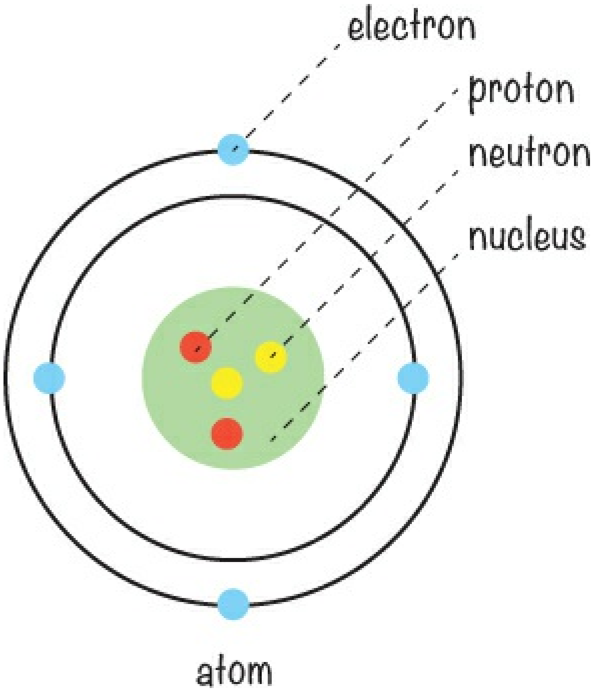

Certain materials are *conductive*, which means that if you apply energy to them (like the energy stored in a battery), the electrons will start moving from one atom to the next!

The filament inside your light bulb is made of a conductive metal, meaning it's full of electrons just waiting for a push to move them.

Voltage Pushes Electrons

When you attach a battery to the light bulb, you're applying a *voltage* across the filament inside. Voltage pushes electrons through the wire and is measured in *volts (V)*. The higher the voltage, the more electrons will flow through the wire.

Think of a wire like a tube filled with marbles: when you put a marble in on one side, a marble pops out on the other side at the exact same time, with no delay.

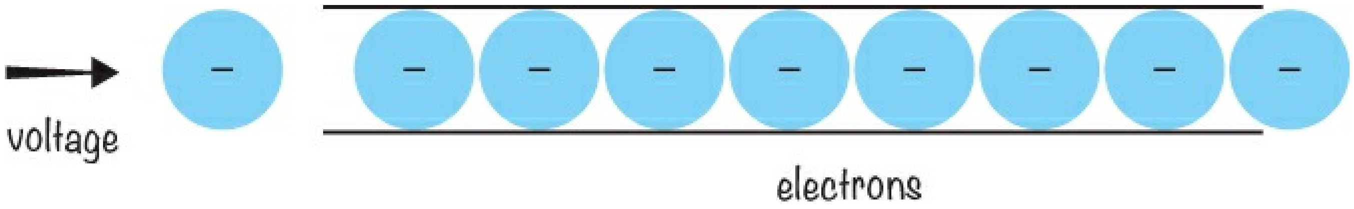

The more marbles you push in one side, the more pop out of the other. That's how electrons behave inside a wire, when a voltage is applied to them.

Current Flows

Current is the amount of electrons flowing through a wire, and it's measured in *amperes (A)*, which we usually shorten to *amps*. You might have also heard the word *current* used to describe a river, as in "This river has a strong current." That means there's a lot of water moving down the river.

Electrical current is similar: a strong current means there are a lot of electrons flowing through a wire. When you increase the voltage in a circuit, the current also increases.

Just as water flows downhill due to gravity, electric current flows from the positive battery terminal toward the negative battery terminal. Actually, the electrons themselves flow in the opposite direction, from the negative side of your battery to the positive side.* But when we talk about electrical current, we say that it flows from positive (+) to negative (–).

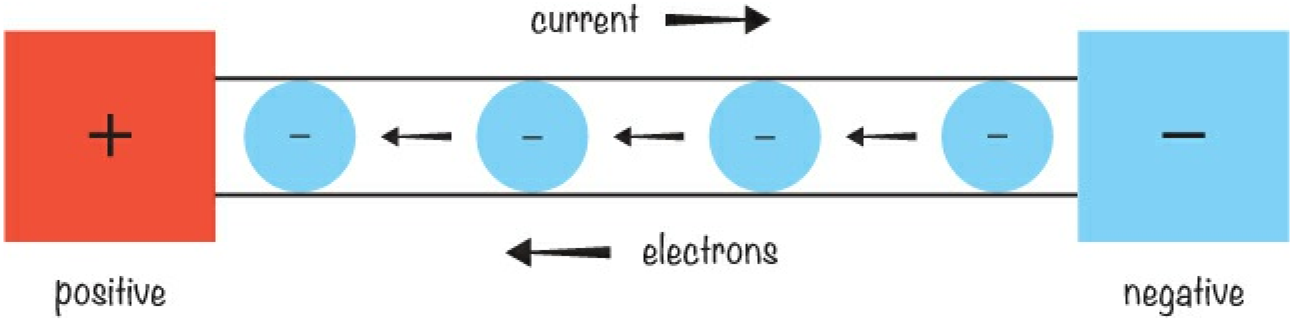

Resistance Reduces Current

Voltage pushes electrons to form a current, and *resistance* restricts the current. It's like playing with a garden hose: if you squeeze the hose, you add resistance to the flow of water so that less water comes out. But if you turn the tap more (like increasing the voltage), the pressure increases, and more water flows even though you're still squeezing the hose in the same way. Resistance in electricity works just like this, and it's measured in *ohms* (Ω).

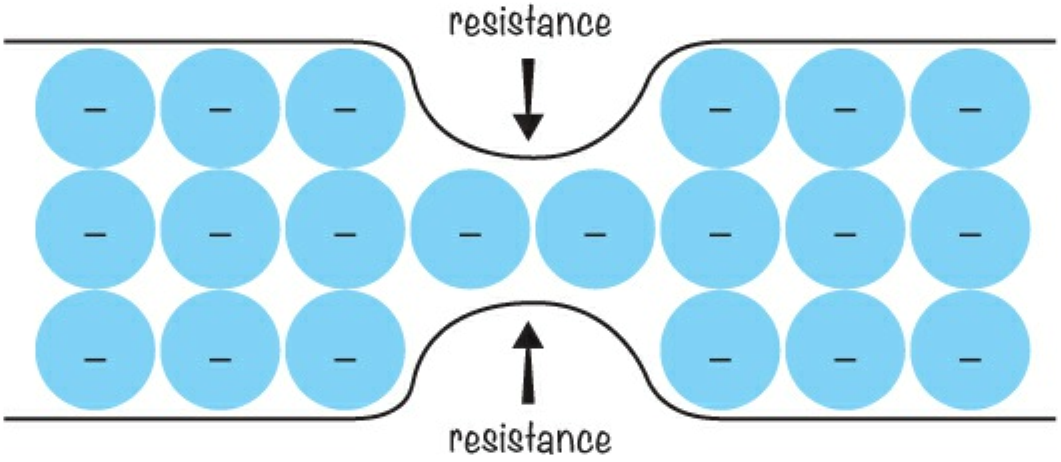

Now that you know about electrons, current, voltage, and resistance, I'll explain how they work together to turn on your light bulb.

Lighting the Bulb

The two ends of the wire filament inside your light bulb are connected to the outside of the bulb: one end is connected to the metal side of the base, and the other end is connected to the metal contact on the bottom. When you connect a battery to the light bulb, you create what's called a *circuit*. A circuit is just a closed path that allows current to flow from the positive terminal of your voltage source to the negative terminal.

The voltage on the battery pushes electrons through the circuit, including the filament inside your light bulb. The filament has resistance and restricts the current in your circuit. As the electrons struggle to make their way through the resistance of the filament, the filament becomes so hot that it starts to glow and generate light. For the battery to be able to push the electrons, there must be a *closed loop* going from the positive terminal on the battery to the negative terminal.

Electricity always needs a *closed circuit* to work. If you disconnect even one of the sides, the light bulb turns off right away! Let's look at circuits in a little more detail.

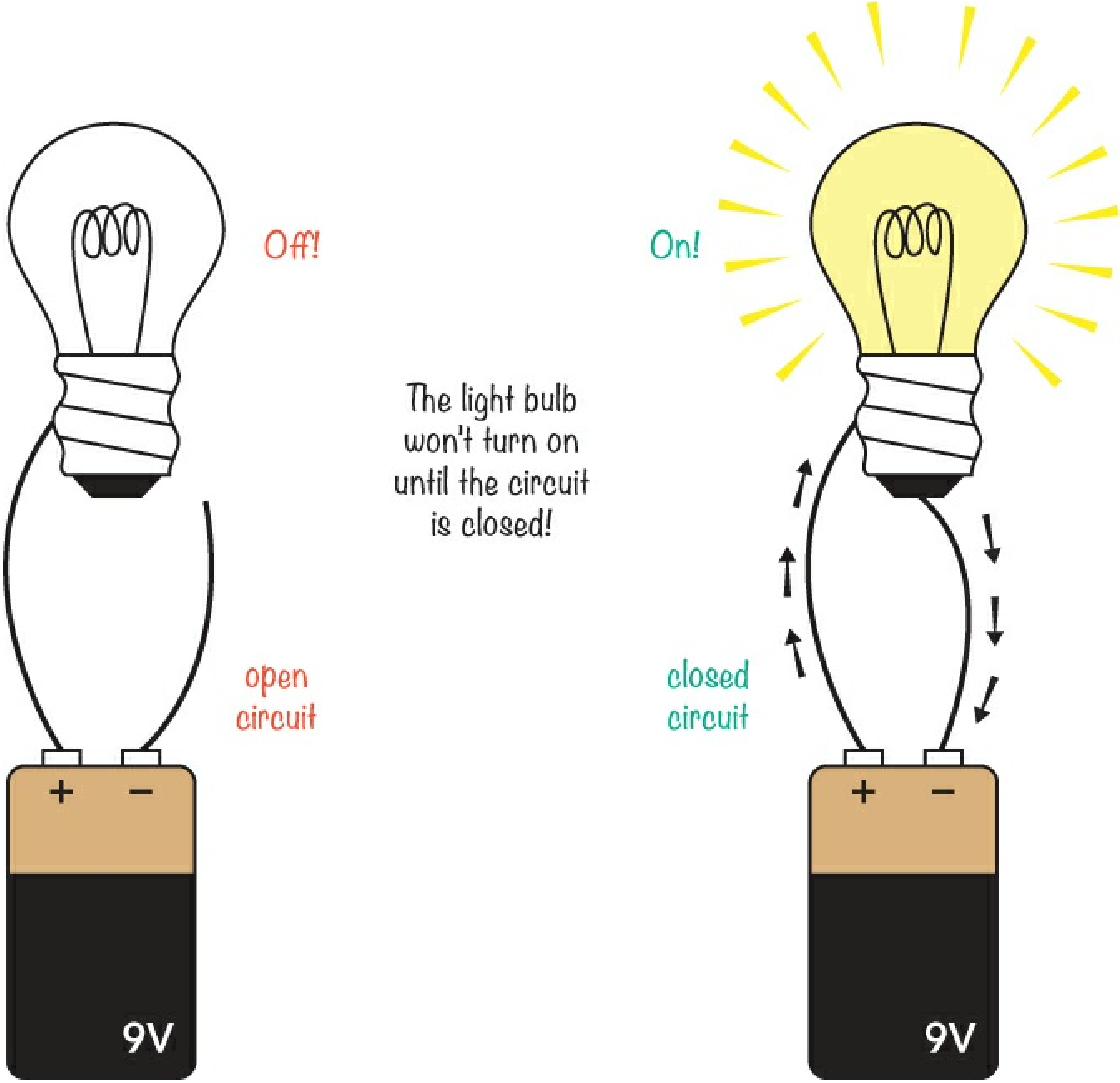
Off!
On!
The light bulb won't turn on until the circuit is closed!
open circuit
closed circuit
+
−
9V
+
−
9V

HOW IS A CIRCUIT LIKE A PIPE SYSTEM?

Let's continue to think about electricity by comparing it to water. Imagine a pipe system in a loop with a pump, and imagine the pipe is always totally filled with water. At one point, the pipe is narrower.

The pump is like a battery that gives power to a circuit. The narrow part of the pipe reduces the flow of water. This narrow pipe is the resistance. The amount of water that flows through the pipes is the current.

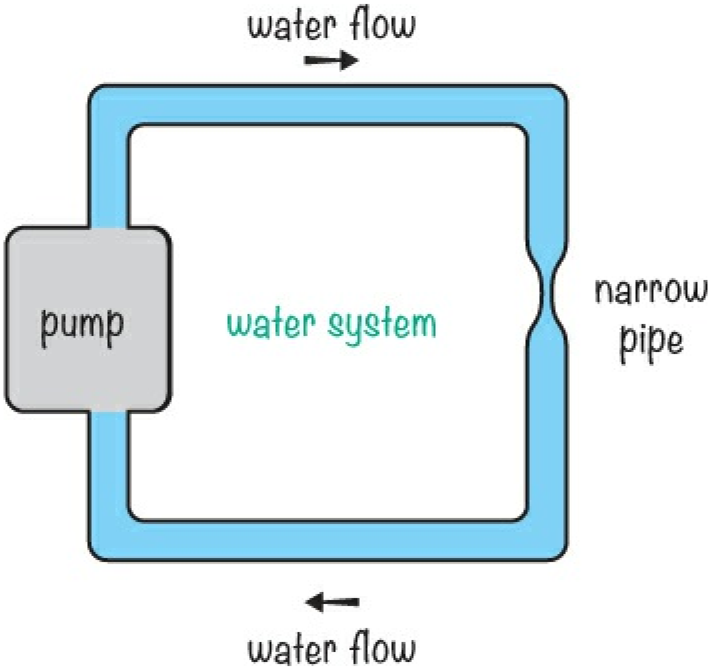

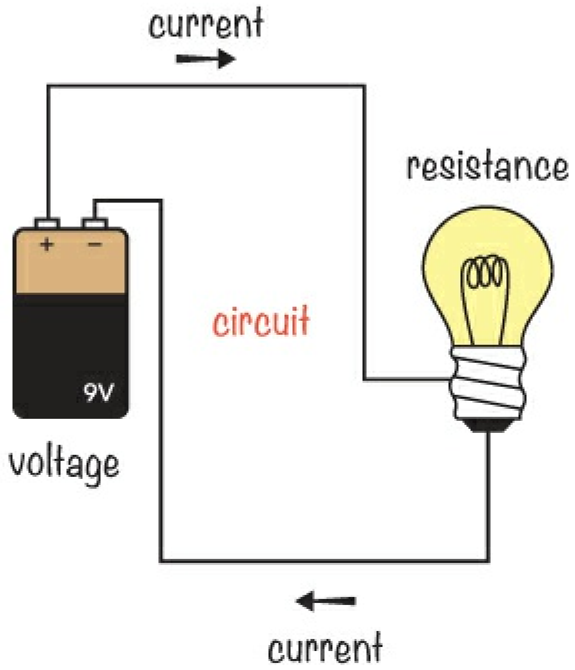

Now imagine that you could insert a measuring device somewhere in the pipe system that tells you how much water is flowing through it per second. Note that I’m talking only about how much water flows through one randomly selected point in the pipes, not the total amount of water in the pipes. This is how we’ll talk about current in an electric circuit, too: it’s the amount of electrons flowing through a certain point per second.

MEET THE SWITCH

When you look around your home, you probably see switches everywhere. You use them all the time to turn lights on and off! When the light in a room is turned on, it must be part of a closed loop because the bulb has current running through it. But what happens when you flip the switch off? Flipping the switch off is the same as disconnecting a wire in the loop: it stops current from flowing and turns the light off, just like the disconnected circuit we saw earlier.

What other switches can you find around you? You might find a switch that turns a computer on and off, a switch to ring a doorbell, a switch to determine whether a refrigerator door is open, and more.

Switches control electricity, and they’re very simple devices. They connect two wires to close a loop or disconnect two wires to open a loop. Inside, a switch is just some pieces of metal that connect or disconnect.

When the switch is open, the light is off. When the switch is closed, the light turns on! It’s very simple, but very useful. With only this knowledge, you can create some nifty circuits, and that’s what we’re going to do next.

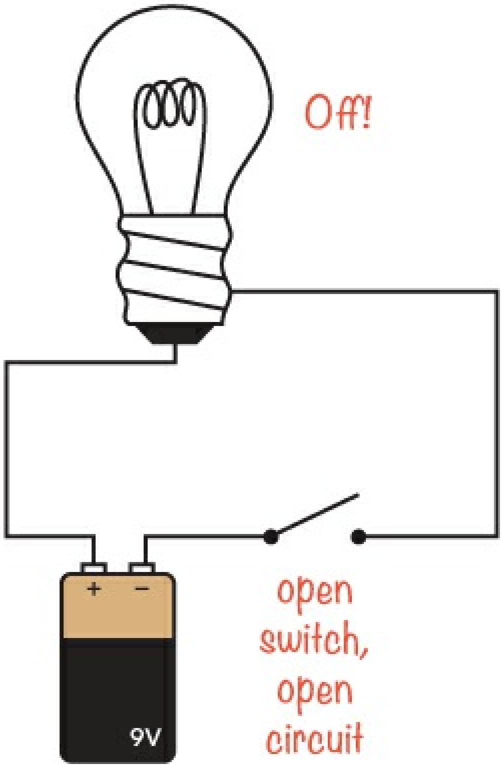

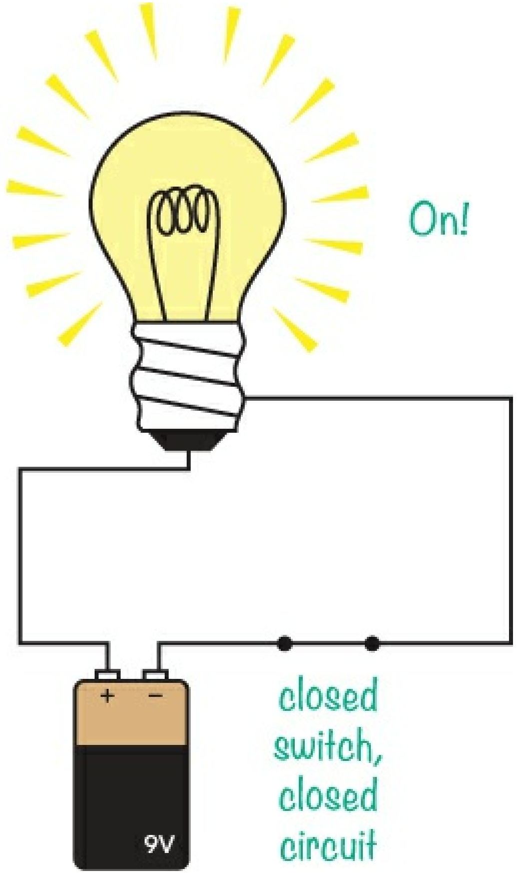

PROJECT #2: INTRUDER ALARM

In this chapter, we learned that electricity needs a closed circuit to make a circuit do anything interesting, and we’ve looked at how a switch works. Let’s build a circuit with a switch!

You can build a switch out of a lot of different things—even a door. In this project, you’ll turn a door into a gigantic switch and use it to build an intruder alarm that tells you when someone tries to enter the room.

To create the alarm, we’ll attach some wires and aluminum foil to your door so that when the door is closed, your circuit is open and nothing happens. But when the door opens, the loop closes and a buzzer sounds to set off a red alert.

We’ll hang an exposed wire down from above the door, place a strip of aluminum foil on top of the door, and connect each of these parts to a different side of the circuit. Then, when the door opens, the exposed hanging wire will touch the aluminum foil and close the loop so that the buzzer will sound.

Shopping List

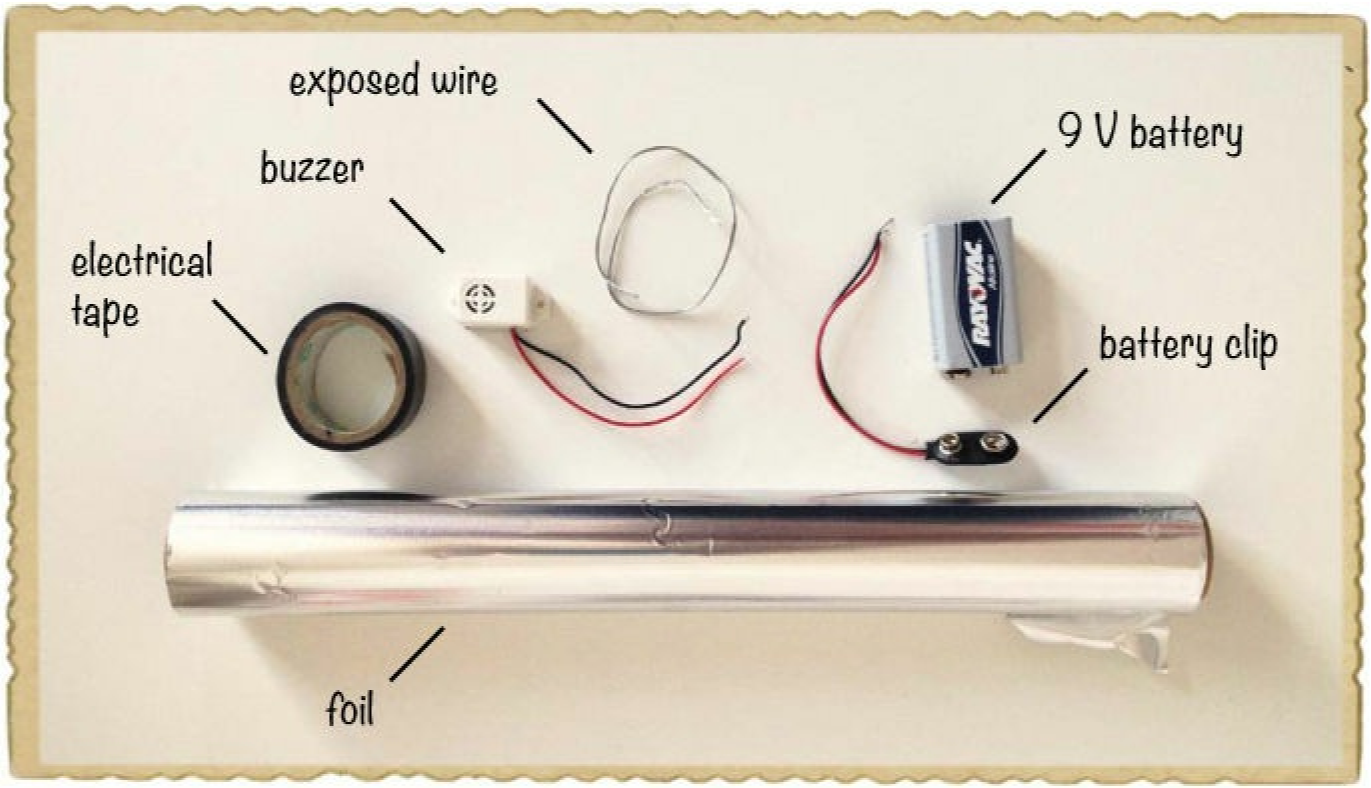

- **A buzzer** (Jameco #2173870, Bitsbox #ST016) that beeps. Buzzers come in both passive and active versions. Passive buzzers need an audio frequency input, while active buzzers need only a voltage. For this project, you need an active buzzer that works with 9 V.
- **A standard 9 V battery** to power the circuit.
- **A 9 V battery clip** (Jameco #11280, Bitsbox #BAT033) to connect the battery to the circuit.
- **Aluminum foil**
- **Exposed wire** (Jameco #2098478, Rapidonline.com #05-0320), like some soft electrical wires, an old steel guitar string, or something similar.
- **Tape** to fasten everything. You can use masking tape, electrical tape, or whatever you have.

Tools

wire cutter
scissors

- **A wire cutter** (Jameco #35482, Bitsbox #TL008) to cut or remove the insulation from wire.
- **A pair of scissors** (optional) is useful for cutting the aluminum foil nicely.

TIP

If you want to make stripping wires even easier, you could buy a pair of actual wire strippers, like Jameco #78992, which are slotted so that you won't cut through the wire by accident.

Step 1: Does the Buzzer Beep?

First, test the buzzer to see that it beeps. Hold the red wire from the buzzer to the positive terminal on your battery (marked +) and touch the black wire to the negative terminal on your battery (marked –).

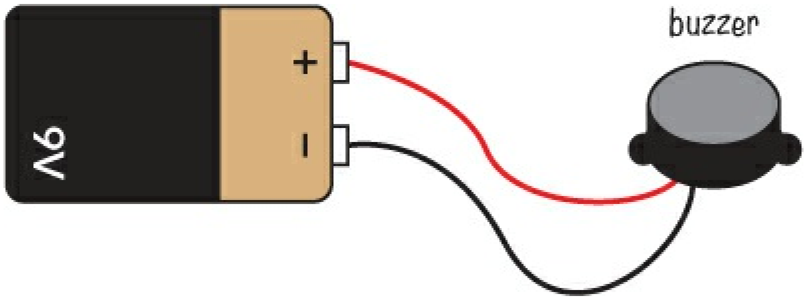

The buzzer should now make a loud, clear buzz or beep. If you disconnect one of the wires from the battery, the buzzer should stop making a sound because the circuit won't have a closed loop anymore.

NOTE

If your buzzer just made a click or didn't make any sound at all, you might have a passive buzzer. A passive buzzer can't create sound by itself, so you need an active buzzer for this *project. The buzzer recommended in this project's Shopping List (page 12) should do the trick.*

Step 2: Prepare the Aluminum

Use a pair of scissors to cut a big, shiny strip of aluminum foil to go on the top of the door. Cut a straight piece of foil, about 1 inch wide and as long as the roll of aluminum foil is wide.

Step 3: Foil Your Door

Fasten the strip of aluminum foil on the top of the door by using a piece of tape on each side of the strip. The foil will act as a contact for the battery and buzzer wire.

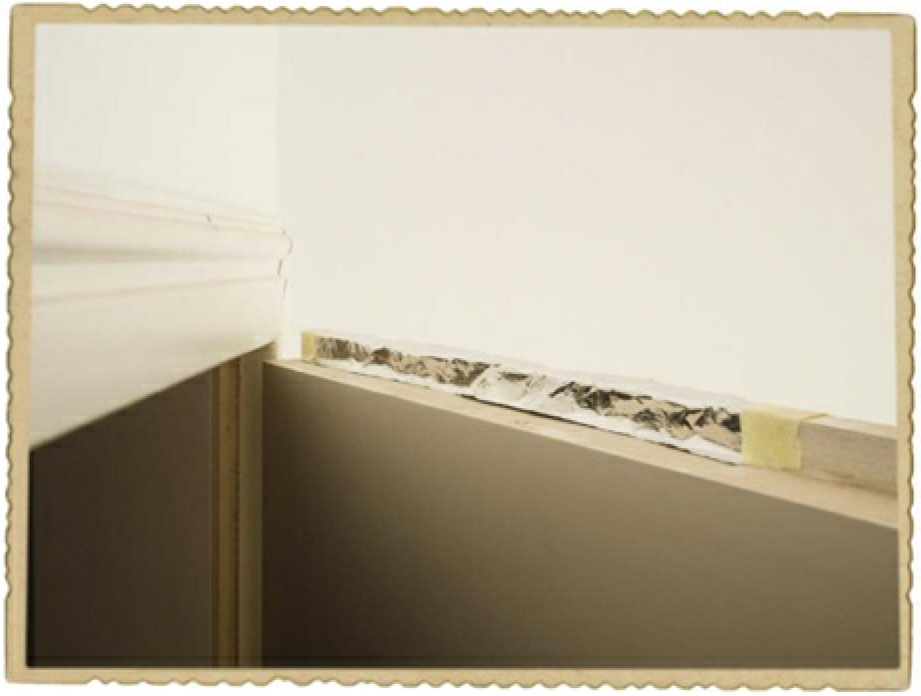

Step 4: Prepare a Trigger Wire

Get a piece of exposed soft wire about 10 inches long. An *exposed wire* is a wire that doesn't have any plastic around it, as opposed to *insulated wire*, which is metal enclosed in plastic. Just find some wire that is already exposed, such as a steel string from a guitar, or use your wire cutters to snip a piece from the spool in the Shopping List (page 12). This is going to be your trigger wire.

NOTE

You could also use a wire cutter to remove, or strip, the plastic from insulated wires. If you want to do that, ask an adult for help!

Step 5: Connect the Buzzer and Trigger Wire

Connect one side of the exposed trigger wire to the exposed metal end of the battery clip's black wire with some tape. Connecting two wires is easy. Here's how you do it: Pick up the two wires you want to connect and twist their ends together. Make sure the two pieces of metal are touching! Then wrap them inside the tape together.

Following the same process, connect the red wire from the battery clip to the red wire on your buzzer.

Step 6: Mount the Buzzer and Trigger Wire

Now, let's place the trigger wire and the buzzer above the door. First, tape the trigger wire onto the door frame above the door so that it's hanging in front of the door when the door is closed and lying on top of the door and the aluminum strip when the door is open.

Next, tape the buzzer to the door frame in such a way that the black wire can touch the aluminum foil on top of the door. Tape the black wire onto the foil so that the exposed part of the wire touches the foil.

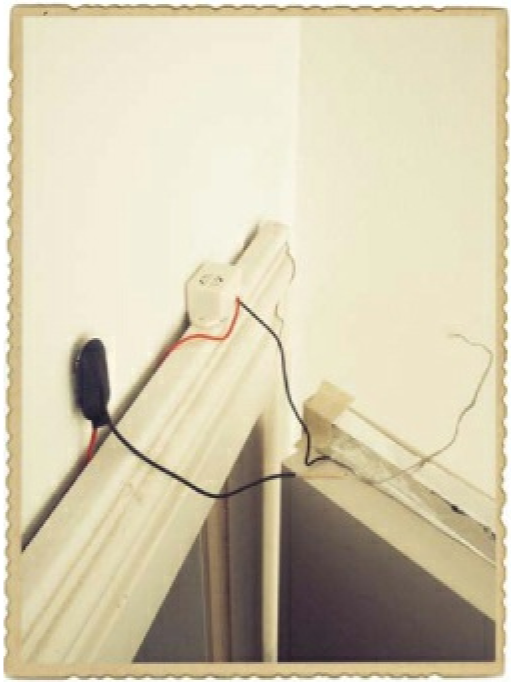

Step 7: Add a Power Source

Place the battery on top of the door frame, close to the battery clip. Use some tape to hold it in place if necessary. Then connect the battery clip to the battery.

Once the battery is connected, your finished intruder alarm should look something like this:

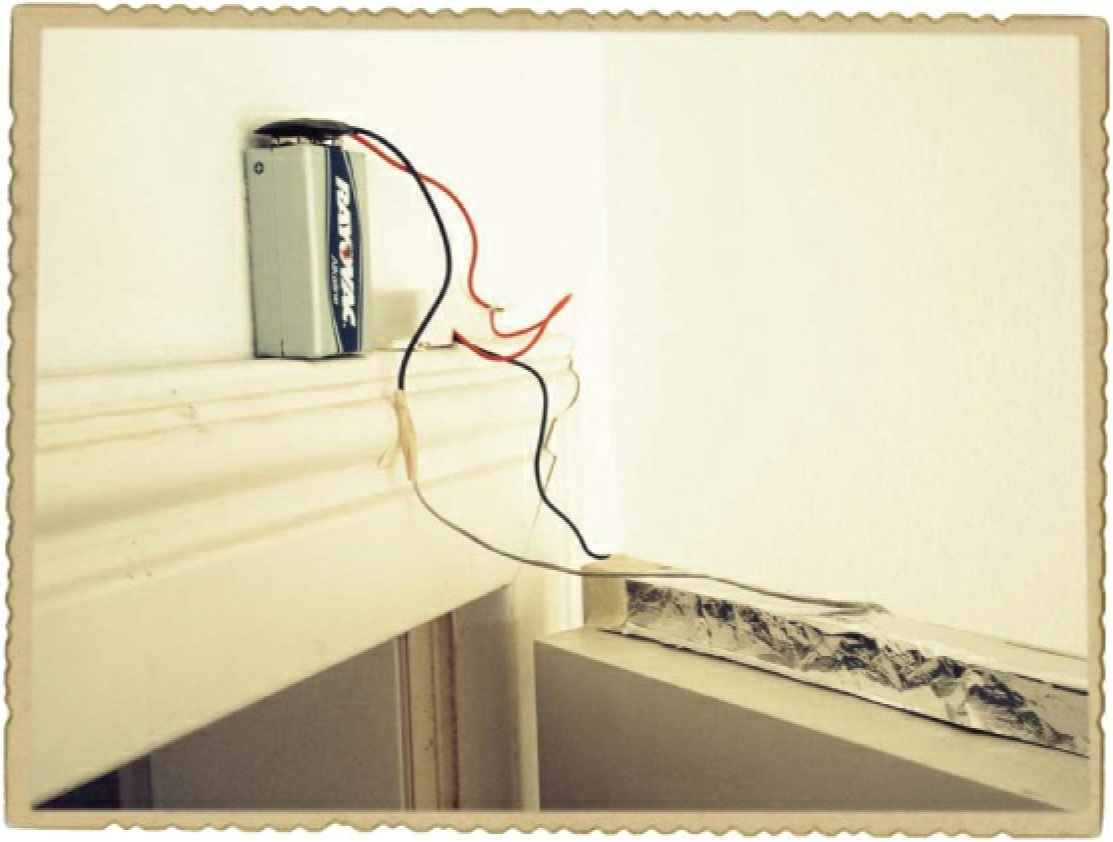

Step 8: Stage an Intruder Alert!

Test the alarm by opening and closing the door. As the door opens, the exposed wire should hit the foil, causing the buzzer to sound a loud alarm. For a more realistic test, invite someone else to open the door instead!

Step 9: What If the Intruder Alarm Doesn’t Work?

If the buzzer doesn’t go off, you might need to adjust the position of the trigger wire a bit, just to make sure the wire touches the strip of aluminum foil when the door opens. If the trigger wire touches the foil just fine, try a different battery. If that doesn’t work, you might need to retape your battery leads to their wires.

WHAT’S NEXT?

Now you know the basics of electricity—a current of electrons flows through wires and makes something happen, like lighting a light bulb or sounding an alarm. And you also know that to make the electrons flow through the circuit, you need a voltage source, such as a battery, and a closed loop. That’s all you need to start tinkering with electronics!

What else can you think of to make with what you’ve learned? There are many other things that can be made into switches. For example, try making an alarm for your closet to keep nosy siblings or friends away from your personal stuff. Or how about making a silent alarm? Just replace the buzzer with a light bulb!

In the next two chapters, we’ll look at how electricity is generated and how we can use electricity to make things move.

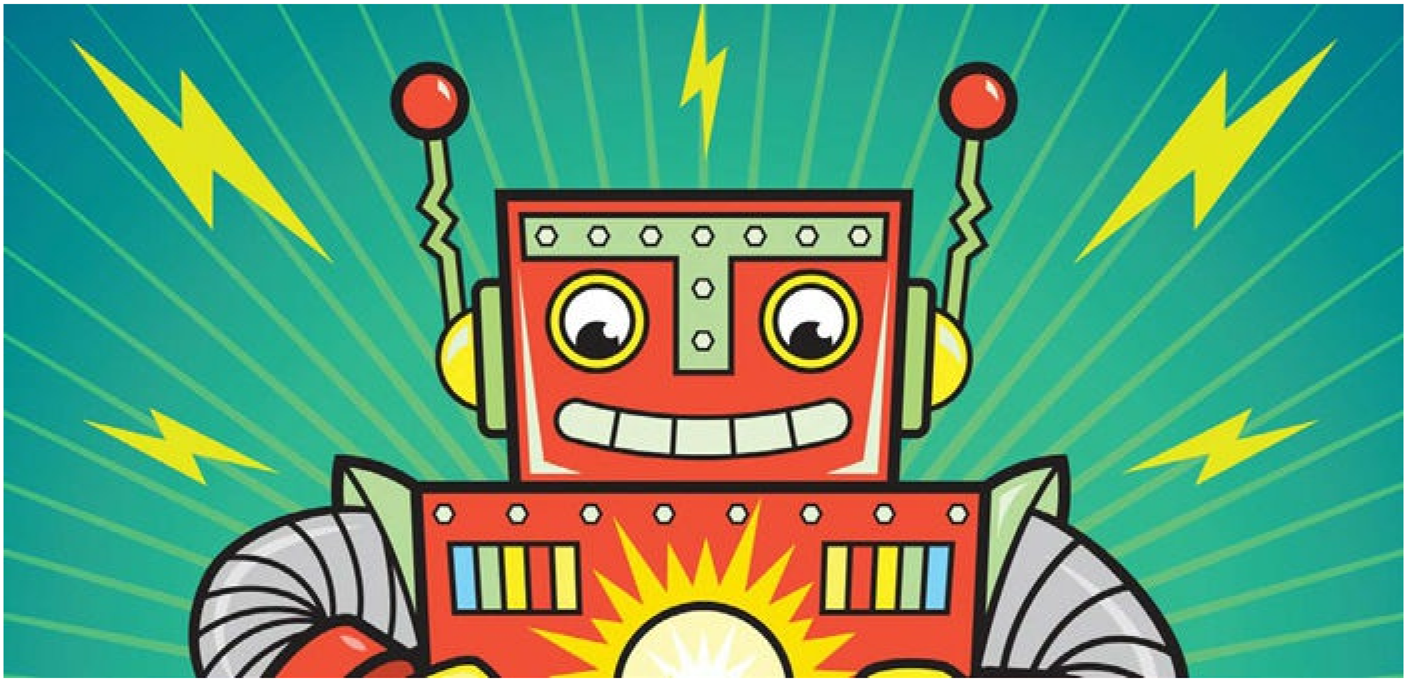

2

MAKING THINGS MOVE WITH ELECTRICITY AND MAGNETS

Big magnets attract small metal objects; small magnets stick to large metal objects. For example, refrigerator doors are usually big pieces of metal, so it's easy cover them with tiny, decorative magnets. You've probably seen magnets in cartoons, too: characters like to use giant horseshoe-shaped magnets to cause mischief. You can find magnets in nature or create them with electricity. A magnet created with electricity is called an *electromagnet*.

You can use an electromagnet to make things move, and you don't even have to be a superhero to do it! In fact, many things you see every day—like motors, loudspeakers, and the automatic doors in shops—work because electromagnets make something in them move.

An electromagnet is very easy to make, and in this chapter, you'll build an electromagnet that you can turn on and off with a switch. Then, you'll use an electromagnet to build your very own motor!

HOW MAGNETS WORK

Magnets have two poles, the *north pole (N)* and *south pole (S)*, and they're surrounded by a magnetic field.

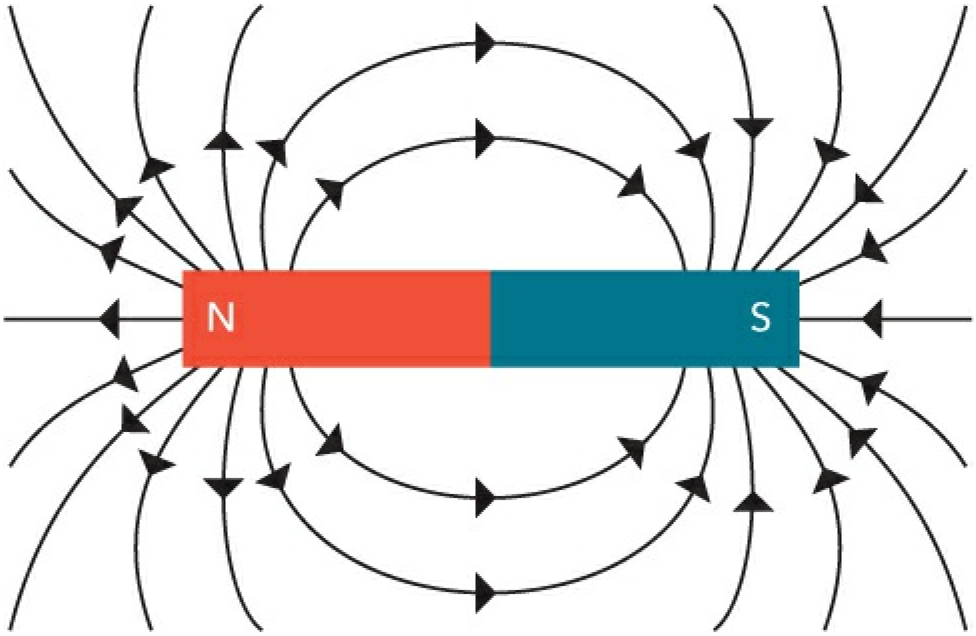

If you place two magnets side by side, the north pole of one magnet attracts the other magnet's south pole and repels that magnet's north pole. Try pushing two magnets together. If you don't force them, they should naturally attach to each other at their opposite poles. Now, try to force two of the same poles toward each other. That's harder, isn't it? Opposite poles are attracted to each other, and identical poles repel each other.

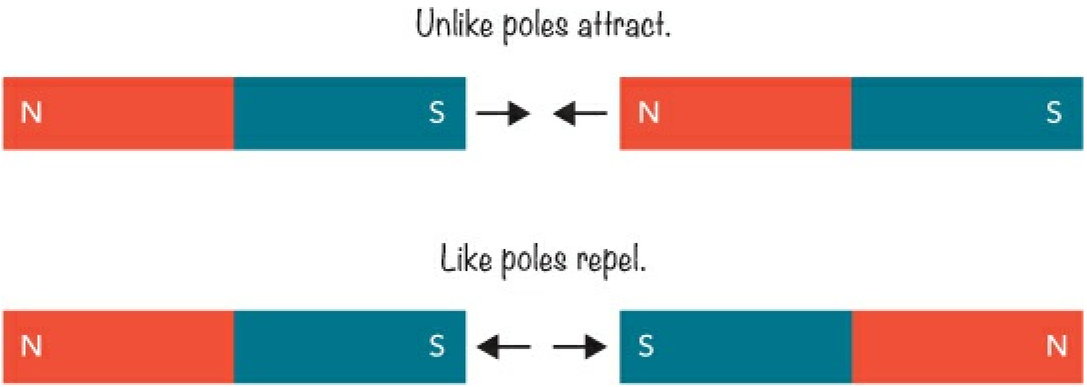

NOTE

Thin, flexible refrigerator magnets don’t have two distinct poles. Instead they have many poles of opposite polarity next to each other, so it’s harder to feel the magnets attract and repel.

But magnets don’t attract all materials. For example, plastic is unaffected by magnets. Try testing some metal objects around you!

TRY IT OUT: FIND SOME MAGNETIC OBJECTS!

Take any magnet and place it over objects made out of different materials, such as:

- Aluminum foil
- A stainless steel spoon
- A soda can
- An iron nail
- A piece of metal jewelry
- A few different coins

Which objects does the magnet attract or stick to? You should find that the magnet attracts some metals, but not all metals. What happens with aluminum foil?

It turns out that some metals can turn into magnets if you apply a little electricity. That’s where electromagnets come in.

MEET THE ELECTROMAGNET

When current flows through a wire, something strange happens: the current creates a magnetic field around the wire.

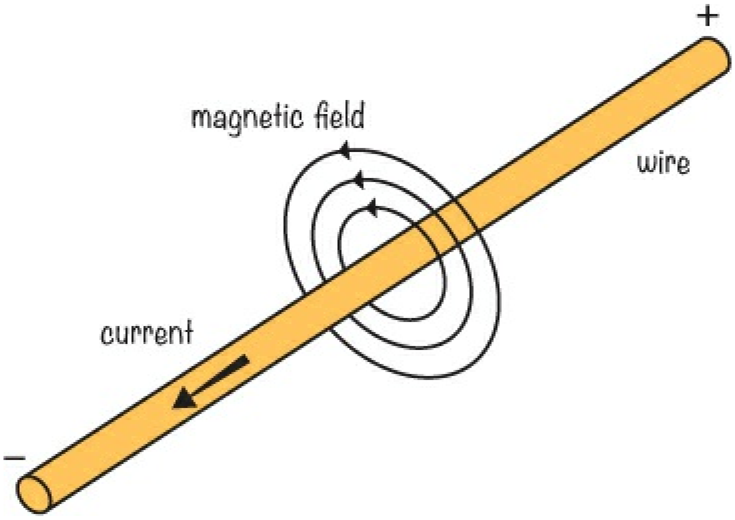

The magnetic field of one wire, however, is very weak. To make a stronger magnetic field, you need to run current through lots of wires placed next to each other. But you still need only a single wire: you can just wind that wire into many loops to make a coil, and then send a current through it. The magnetic fields from each loop in the coil overlap and combine to create a stronger magnetic field. If you wind your wire around a piece of iron—like a nail, a bolt, or a screw—you'll get an even stronger magnetic field.

All you have to do to create an electromagnet is connect a battery to the ends of the coiled wire, making a closed circuit. When current flows through the wire, the piece of iron it's wrapped around starts to behave like a magnet, with the south pole at one end and the north pole at the other end. Which pole is which depends on the direction of the current, as well as the direction of the coil windings. When you disconnect the battery, the current stops and the magnetic field disappears.

Building an electromagnet will help you start to understand how you can use electricity to make things like a loudspeaker in the real world, so let's make one! With enough current, enough wire, and the right circuit, you could build a supermagnet straight out of your favorite cartoon, but for now, we'll start with a small one.

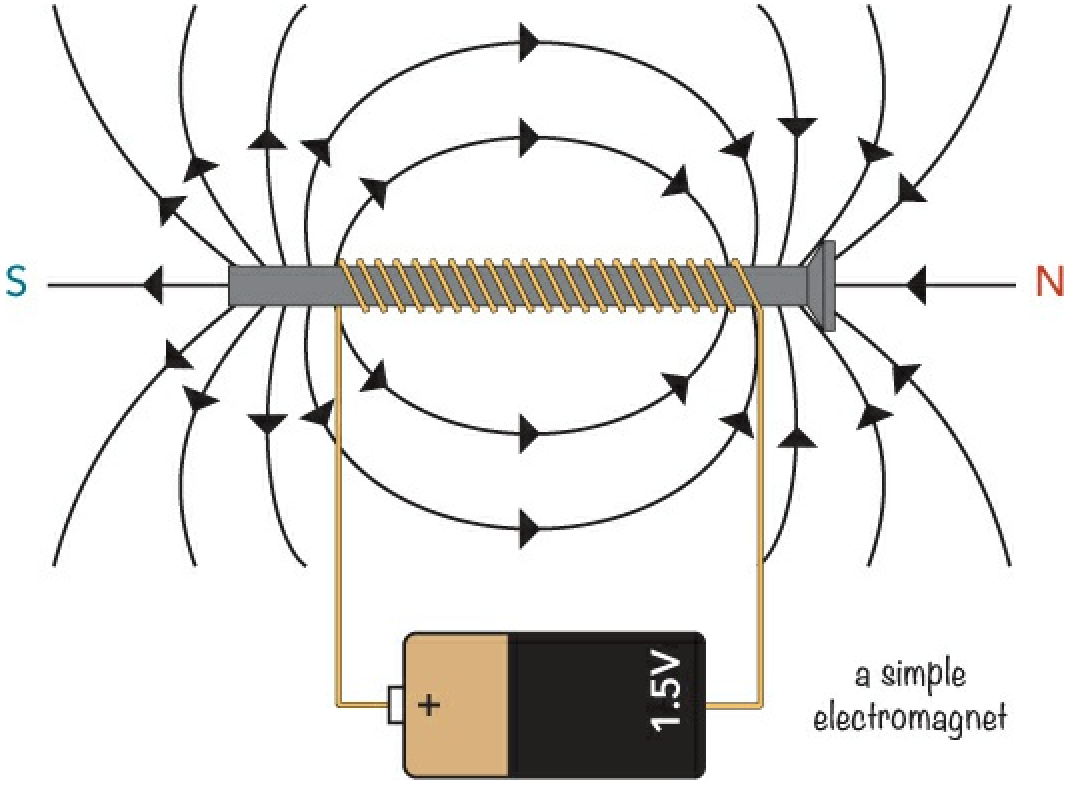

PROJECT #3: CREATE YOUR OWN ELECTROMAGNET

You know the theory behind how to build your own electromagnet. But reading the theory isn't the same as making something in real life, so it's time to have some fun!

You're going to build your own electromagnet with wire and a bolt. All you need to do is to wrap the wire around the bolt several times and connect the battery to the wire. To make it easy to turn the electromagnet on and off, you'll also add a switch to the circuit so that you can control whether or not current flows through the wires.

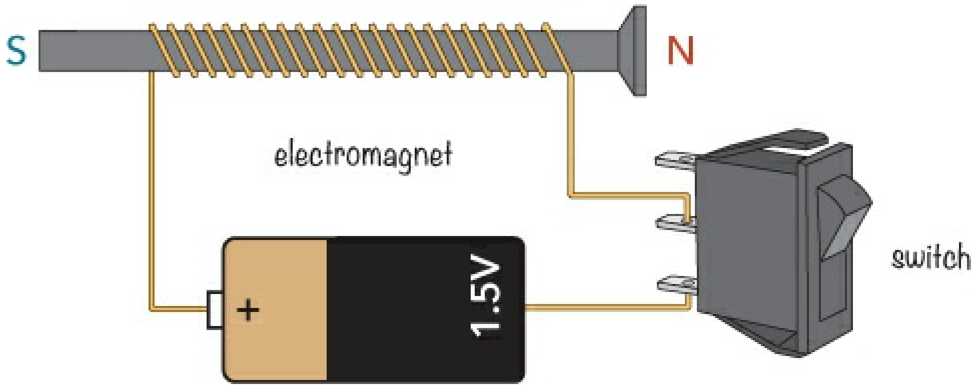
S
N
electromagnet
+
1.5V
switch

Shopping List

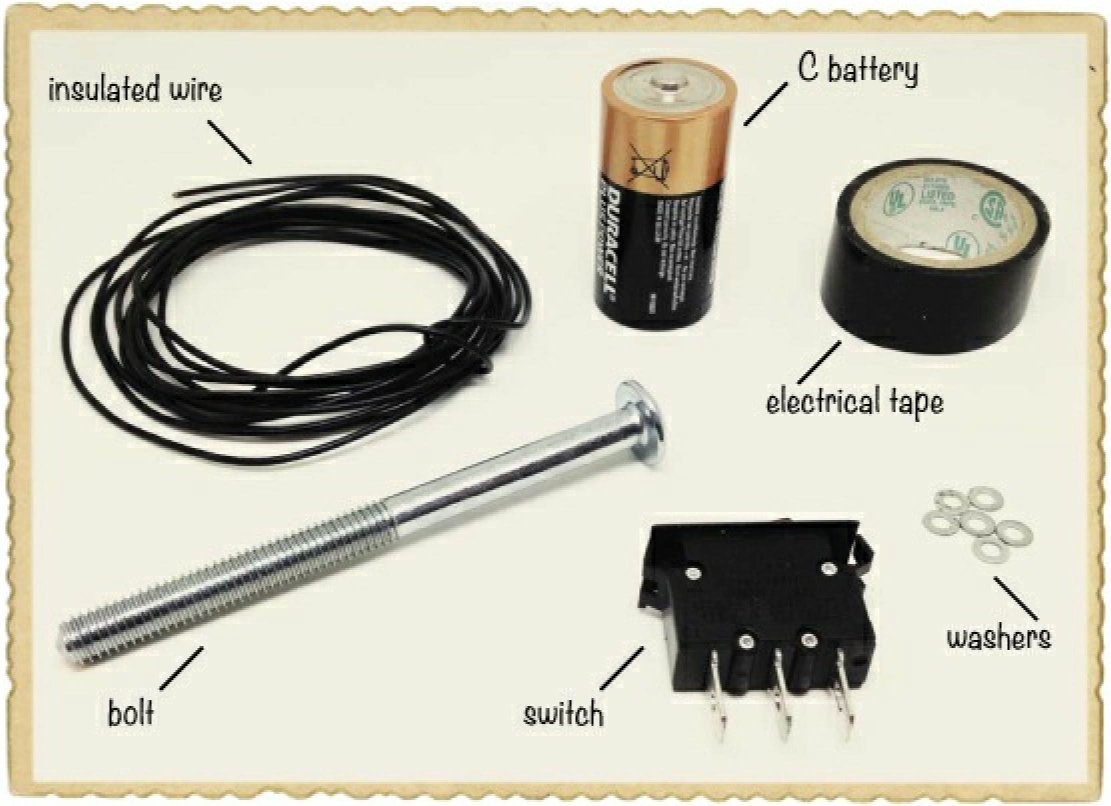
insulated wire
C battery
electrical tape
washers
bolt
switch

- **A 1.5 V alkaline (C) battery** (Jameco #2112428, Bitsbox #BAT040), like the big round ones used in older flashlights. Don't use a rechargeable battery or plug-in power supply.
- **Insulated solid-core wire** (Jameco #36792, Bitsbox #W106BK), about 7 feet. Standard hookup wire works fine.

- **Tape** to fasten everything. You can use masking tape, electrical tape, or whatever you have.
- **Washers or paper clips**, or other small metal objects that your electromagnet can lift.
- **A bolt** to wind the wire around. Choose a big one to make room for many turns with the wire. The bolt I used was 0.3 inches thick and 4 inches long.
- **A switch** (Jameco #581685, Bitsbox #SW018) to turn the electromagnet on and off.

Tools

wire cutter
magnet

- **A wire cutter** (Jameco #35482, Bitsbox #TL008) to cut or remove the insulation from wire.
- **A standard magnet**

Step 1: Check Your Bolt

Your bolt is going to be the core of your electromagnet, making it stronger. But not all materials will work as an electromagnet's core! Most metal bolts should work, but if you're unlucky and find one that is made of nonmagnetic material, your electromagnet won't be very effective.

To check whether a bolt is okay to use in this project, just hold it close to any standard magnet. If the magnet attracts the bolt, then the bolt is a good one.

Step 2: Remove Insulation from One End of the Coil Wire

To connect the coil wire to the battery and the switch, you need to expose the metal of the wire at both ends. You'll use a wire cutter to strip away about 0.5 inches of insulation from the beginning of your wire. After you've wound the coil, you'll do the same with the end of your wire. Stripping wires can be a bit difficult if you've never done it before, so ask a parent or teacher for help to get started.

First, gently grasp the end of the wire with the cutters.

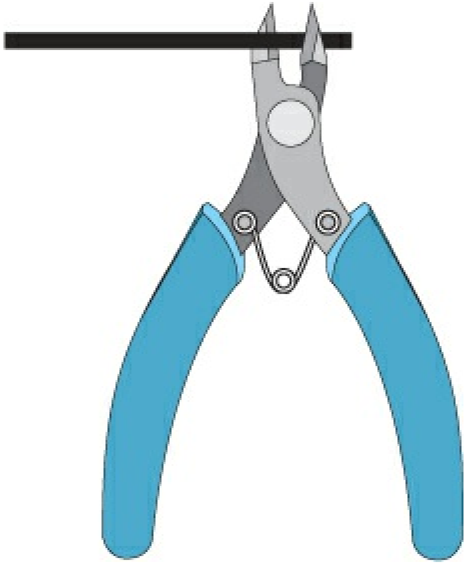

Apply just enough pressure with the wire cutter to cut the plastic around the wire, but not the wire itself. When you've cut through the insulation, your wire should look something like this:

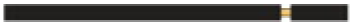

Then, place the wire cutter in the cut you made. Squeeze the wire cutter enough to grip the loose plastic with the blades. Use the wire cutter to gently pull off the plastic without cutting into the metal of the wire.

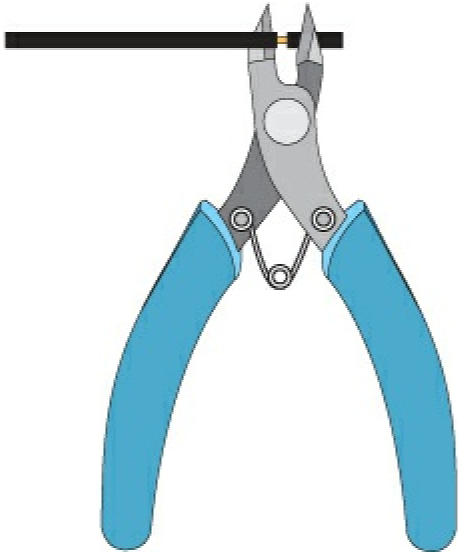

Now, you should have a wire with some exposed metal at the end, like this:

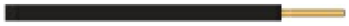

If stripping wires seems tricky in the beginning, don't worry: it becomes much easier with practice.

Step 3: Wind the Wire

Take the wire and wrap it around your bolt 50 to 100 times. Leave about 3 inches of each end of the wire hanging loose. Make sure you don't use all the wire; you'll need a piece of wire about 4 inches long in a later step.

Wrap the wire as tight as possible and tape the end to make sure the turns stay in place. We call this wound wire the *coil* of the electromagnet.

Repeat Step 2 to strip the insulation off the other end of your coil.

Step 4: Connect the Negative Battery Terminal to the Coil

Connect one end of the coil—it doesn't matter which—to the negative terminal of the battery. Fasten it to the battery with tape.

WARNING

Be sure you're using the recommended 1.5 V battery! Anything more powerful could send too much current through your coil, which could make both the battery and the coil hot enough to burn you.

Step 5: Connect the Switch

In Chapter 1, I showed you how to build your own switch and described how you can use one to turn something on and off. Now, you’re going to connect a prebuilt switch to your electromagnet to turn it on and off. A switch often has three *pins* that you can connect to.

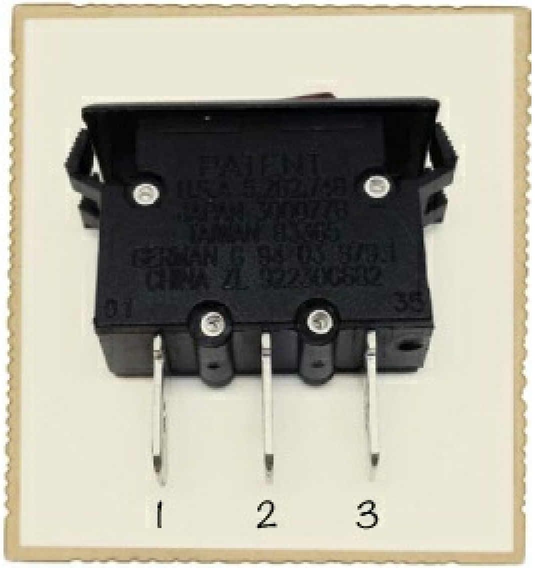

On the switch in this project’s Shopping List (page 24), pin 2 is the *common pin*, which is connected to either pin 1 or pin 3, depending on the position of the button. If the button is pushed toward pin 1, then pins 2 and 1 are connected. If the button is pushed toward pin 3, then pins 2 and 3 are connected.

Some switches have only two pins. In that case, the two pins are connected when the button is in one position, and not connected in the other—just like the switch you built in “Project #2: Intruder Alarm” on page 11.

Fasten the other end of the coil wire to pin 1 of the switch and make sure the button of the switch is pushed toward pin 3. Then, cut a brand-new piece of wire from your spool, about 4 inches long, and strip some insulation from both ends to expose the metal. Connect one end of the new wire to the positive battery terminal and one end to the middle pin of the switch. Use tape to make sure the wires are properly connected and stay in place.

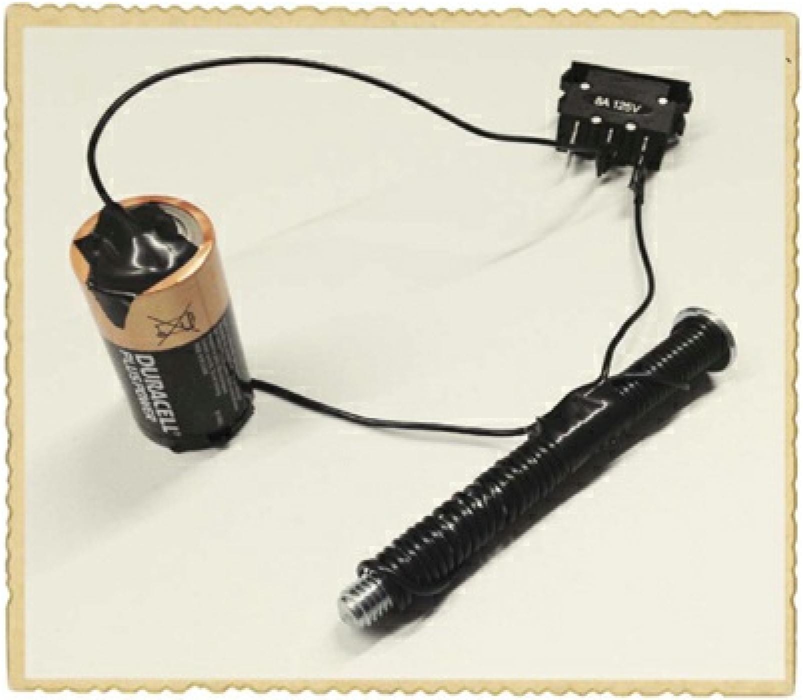

Step 6: Test Your Super Electromagnet

That's it for building the circuit! Now, let's test it. If you've connected everything correctly, your electromagnet should be off now.

First, find a good piece of metal to attract with your electromagnet. A small metal paper clip should do the trick, though I used a little pile of steel washers. Magnets won't attract all metals—for example, aluminum foil is not magnetic—so hold a regular magnet next to the metal you want to attract first to make sure it's magnetic.

Then, flip your switch and place your electromagnet close to your paper clip or whatever other metal object you're using. If you've found the *on* position, the bolt should pull the metal object toward it.

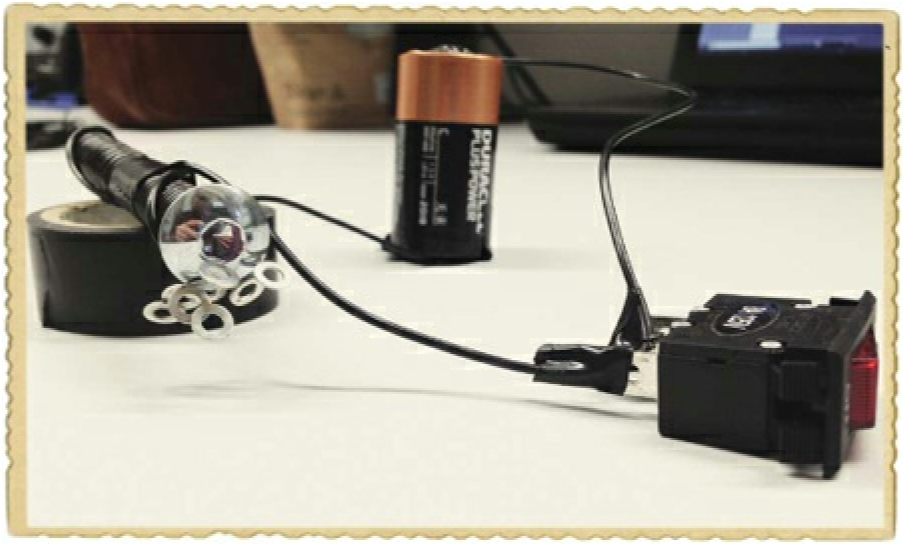

If nothing happens, press your switch into the other position; the bolt should start to pull the metal object now.

The electromagnet consumes a lot of power, so if you keep the switch flipped on for too long, your battery will drain quickly. You might also notice that the battery and the coil become hot. Try to limit the time your electromagnet is on to only a few seconds, and always disconnect the battery before you leave your circuit.

Step 7: What If the Electromagnet Isn't Working?

Make sure you used insulated wire to make the loops around the bolt. The wire must have some kind of insulating layer on the outside of the metal; otherwise, it won't work. The reason for this is that without the insulating layer, the current won't follow the wire loops around the bolt. Instead, they'll go through the bolt if the bolt is conductive or through to the neighboring wire if the loops of wire are touching. In either case, the current will function as if you had one thick wire.

Another possible problem is that your battery is dead. Try switching to a different battery that you're sure is working.

If you're sure you're using insulated wire and that the battery has power, check that the connections on the switch and battery are connected, as I described in Steps 4 and 5. If you're unsure, it might be a good idea to redo the connections.

MEET THE MOTOR

A wire with flowing current creates a magnetic field, as I described in "Meet the Electromagnet" on page 22. When powered, the coil from Project #3 will have a magnetic field with south and north poles, just like any other magnet. Like poles repel each other and opposite poles attract each other. So, if you put a magnetized coil of wire over a regular magnet with the same poles close to each other, the coil will try to twist itself around.

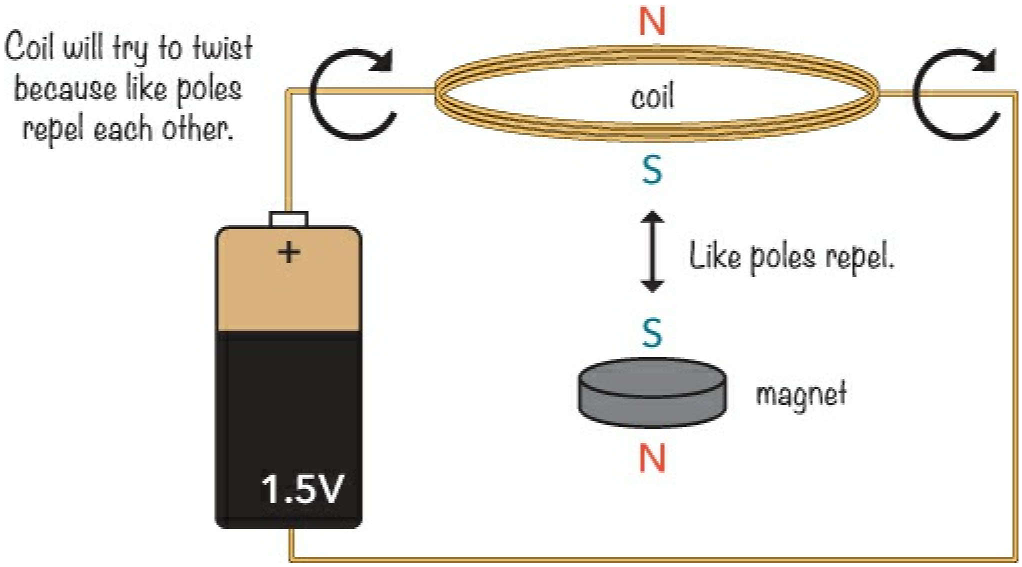

If you placed the wire coil on some kind of stand so that it could rotate freely over the magnet, it would flip back and forth without making a full spin. This is because when the coil has made a half spin, the opposite poles face and attract each other, which will force the coil in the opposite direction.

How can you make the coil continue to spin in one direction? You just need to find a way to disconnect the battery halfway around and turn the battery back on when the coil is back in its starting position. Then, here's what happens. The coil starts moving when it's powered and pushes the wire coil halfway through one round. Because you disconnect the battery halfway through, the existing motion keeps the coil moving forward. When it comes back to its original position, the battery gets reconnected and gives the coil another push forward, and it continues the same way.

Electric motors are based on this basic principle of magnetic poles attracting and repelling each other.

PROJECT #4: CREATE A MOTOR

In this chapter, you've built your own electromagnet, and you've learned how motors work. Now, it's time to combine these two concepts. In this project, you'll build your very own motor from scratch!

You'll use a magnet together with a coil of wire. The coil will spin, and this spinning coil is called the *rotor* of the motor. You're going to build the motor so that the rotor coil has current through it for only half of the spin. The magnet should push the electromagnet for half of the spin, and the rotor coil should continue around the second half of its spin with the energy it gets from the first push.

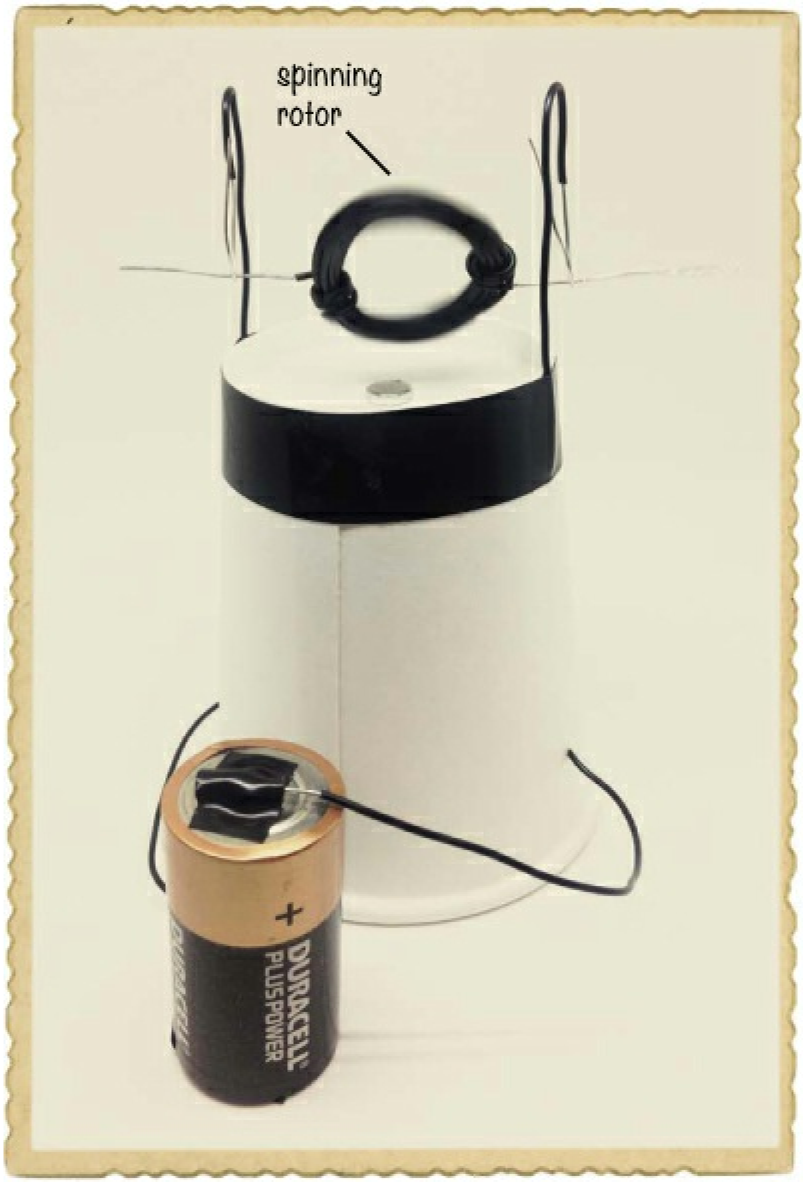
spinning rotor
+
DURACELL
PLUS POWER

Shopping List

cup
insulated wire
C battery
magnets
electrical tape

- **A 1.5 V alkaline (C) battery** (Jameco #2112428, Bitsbox #BAT040), like the big round ones used in older flashlights.
- **Insulated solid-core wire** (Jameco #36792, Bitsbox #W106BK), about 13 feet. The stiff insulated wire will be used both for the coil and to support the coil.
- **Tape** to fasten everything. You can use masking tape, electrical tape, or whatever you have.
- **A paper or plastic cup** to hold everything in place.
- **Two disc magnets** (Jameco #2181319, Bitsbox #HW145), the stronger the better.

WARNING

Always keep small supermagnets like these away from babies and young children. These magnets are very dangerous if swallowed.

Tools

wire cutter

▸ **A wire cutter** (Jameco #35482, Bitsbox #TL008) to cut or remove the insulation from wire.

Step 1: Create the Rotor

First, we'll create a new coil of wire; this coil will be the rotor, or spinning part, of your motor. To create the rotor, first take your spool of wire and strip the insulation from about 1.5 inches of the free end. Then, wind the wire around the battery.

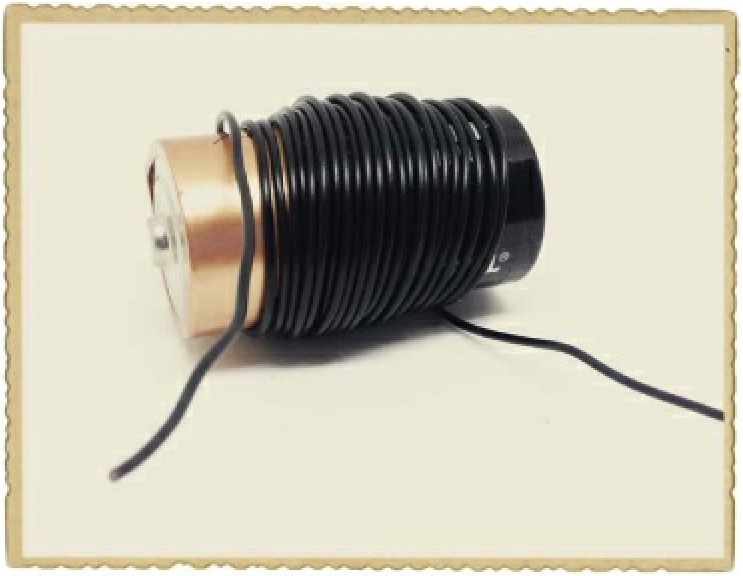

If you buy the wire I recommend in this project's Shopping List (page 34), try making around 30 windings; if you use thinner wire, wind it more. The point is to make the coil as magnetic as possible, without making it too heavy. More windings make the rotor more magnetic, but also heavier.

Carefully slide your coiled wire off the battery. Gather the windings into a loop and wrap the ends of the wire around your loop a few times on each side so that the coils stay together. Cut your loop from the spool of wire, leaving the other end about 1.5 inches long. Then, remove the insulation from this end, too, so that the metal inside is exposed. If you're using wires with plastic insulation, you can use a wire cutter, as described in Step 2 of Project #3 (page 26).

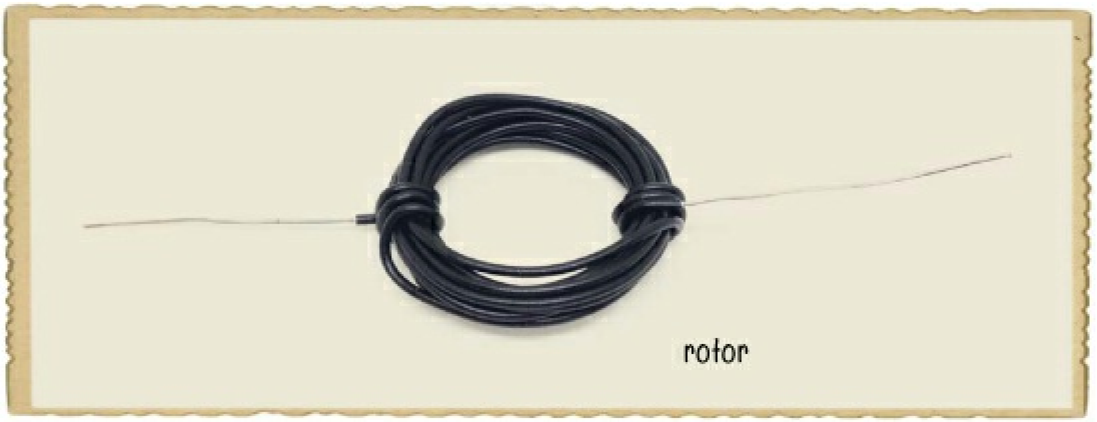

Step 2: Build the Motor's Structure

Set your coil aside for now and take out your paper cup. Punch a hole in one side of the cup about 0.4 inches from the top and another one about 0.4 inches from the bottom. Pull a piece of the stiff wire around 8 inches long through these two holes. Then, do the same on the other side of the cup. Turn the cup upside down, remove the insulation from the ends of both wires, and tape the wires to the cup to ensure they stay in place.

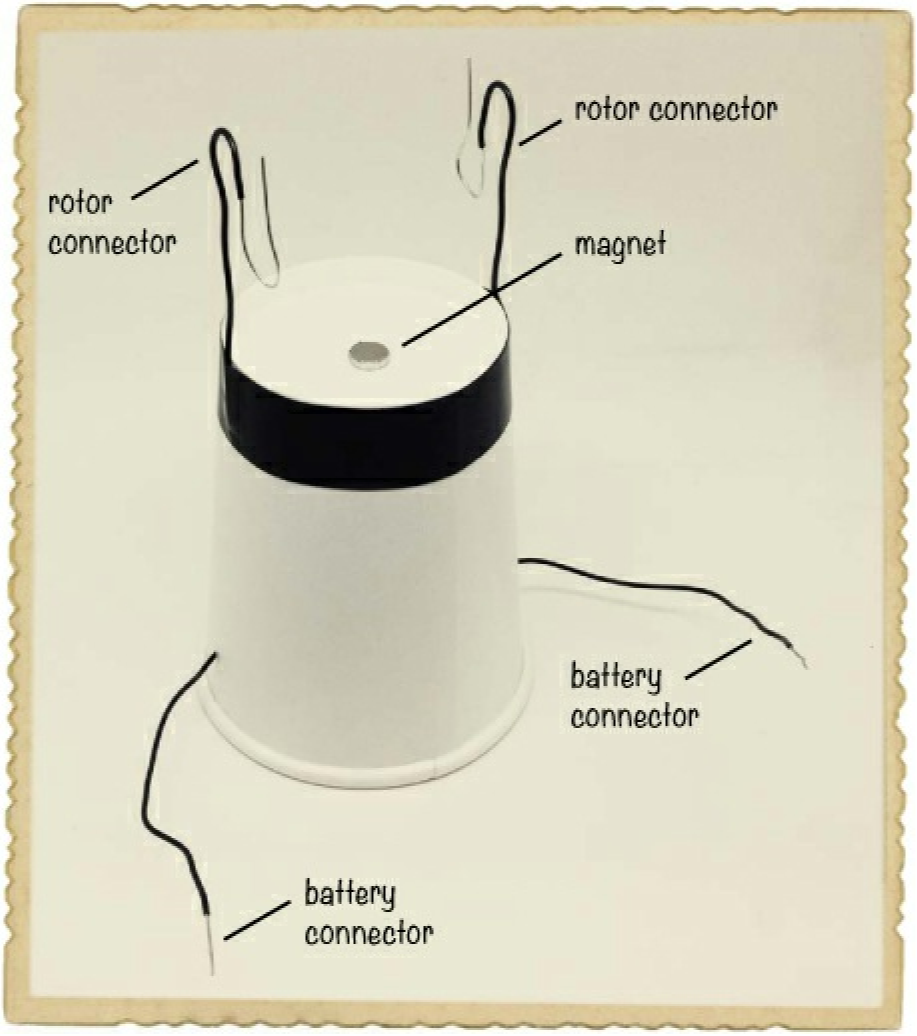

The ends that are now on the bottom will connect to the battery, and the top ends are going to make up the connection to the rotor and support it. Bend the top ends of the two wires into two U-shapes that can hold the rotor. Make sure the bottom part of each U has exposed metal so that it will touch the exposed wires of the rotor. This U-structure will be the battery's connection to the rotor.

Step 3: Place the Magnets

Place one magnet on top of the cup. Then place one magnet inside the cup so that the two magnets stick to each other through the cup. Place your rotor into the U-structure and adjust the position of the magnets to make sure they are at the center, just under the coil.

Step 4: Reinsulate Part of the Coil

If you connected the battery now, the motor wouldn't work. With your coil rotor attached, you'd see movement, but the rotor would just be pushed back and forth in opposite directions because it's always connected to the battery. You need a way to disconnect the coil from the battery halfway through so that it's first pushed away from the magnet and then released until it has spun the rest of the way around. Then, it can reconnect with the magnet and get pushed again, and so on. You can make this happen by insulating the wire on one side with a permanent marker. Do this on only one arm of the rotor.

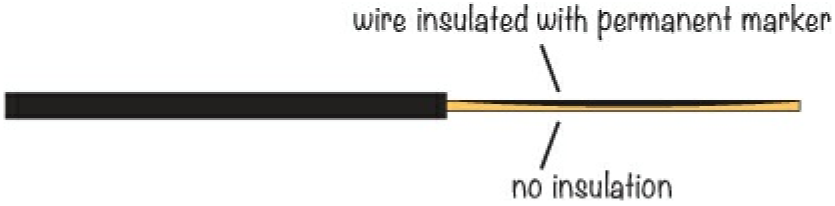

Lay your coil flat on the table and use a permanent marker to draw along the wire on one side to make it non-conductive. Draw your line so that the rotor disconnects from the battery when the loop lies horizontally above the magnet.

Step 5: Rev Up Your Motor

Let's get that motor running! Connect the battery by taping the two wires to the positive and negative terminals.

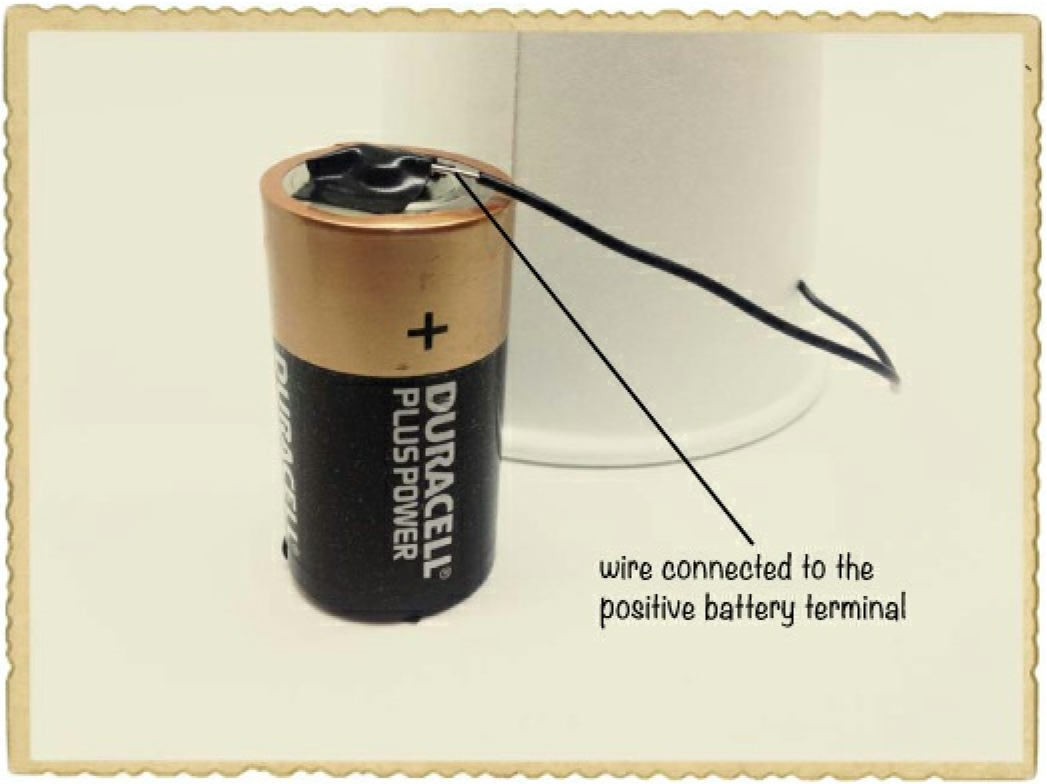
+
DURACELL
PLUS POWER
DURACELL
wire connected to the
positive battery terminal

Now, place the rotor into the U-structure. The motor should start spinning. You might need to give it a little push. It won't run any cars, but if it works, then you definitely just made something move with electricity. Congratulations!

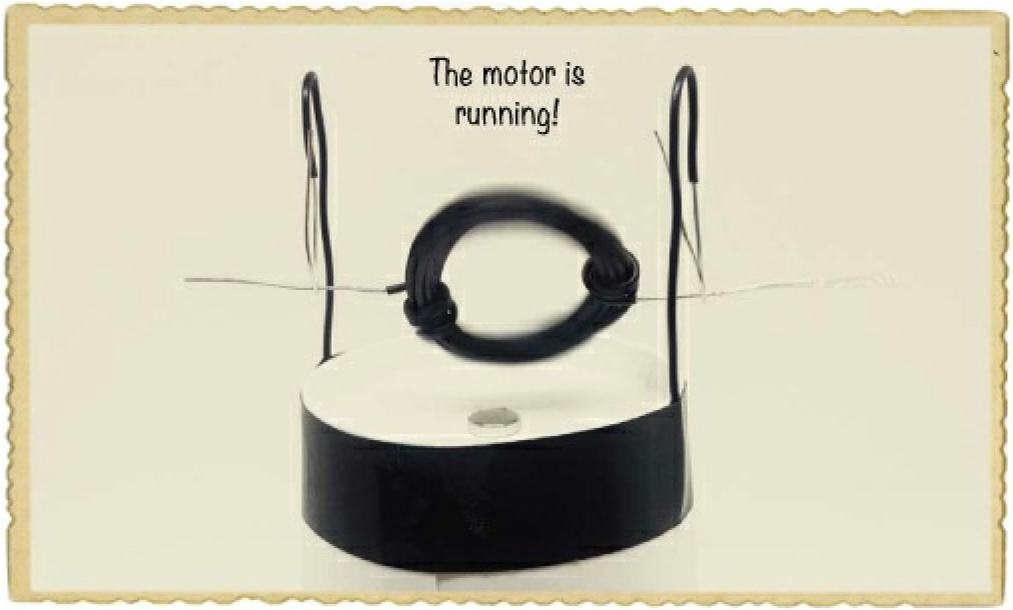
The motor is
running!

Step 6: What If the Motor Doesn't Work?

Can you see any movement? If you're very lucky, it'll work right away, but you'll most likely need to make some adjustments. Here are some places to start:

1. Make sure your coil is placed so that it starts with the exposed wire—that is, not the part you covered with the marker—touching the exposed wire of the U-shaped structure. That way, when you connect the battery, the coil becomes magnetic.
2. Figure out which way the battery should be connected. You might find that the rotor spins better

in one direction than the other, so try to connect the battery the other way around to see what's best for your motor.

3. If your coil is a bit too heavy, the magnetism won't be enough to push the coil all the way around the loop. Try unwinding a few loops to make the coil lighter.
4. You might need to adjust the position of the magnets under your rotor. They should be as centered as possible.

If your motor still doesn't run, your rotor may just need a little push to get started. Try tapping it lightly with your finger to see whether that unleashes a speed demon.

WHAT'S NEXT?

In this chapter, you've learned that magnets can be created by winding a wire around a bolt and connecting it to a battery, and you've tested this by building your own electromagnet. At the end, you learned how electric motors work, and you even built one for yourself. You really got things moving!

Now, take that knowledge and explore electricity a little further. Try adding even more magnets under the rotor of your motor. Then, wind a rotor coil that is twice as big or even bigger. You can create a much larger structure for the motor. How fast can you make your motor go?

So far, you've only used electricity, but you can actually generate it, too. In the next chapter, you'll learn a couple of different ways to generate electricity, and you'll be playing around a bit more with magnets.

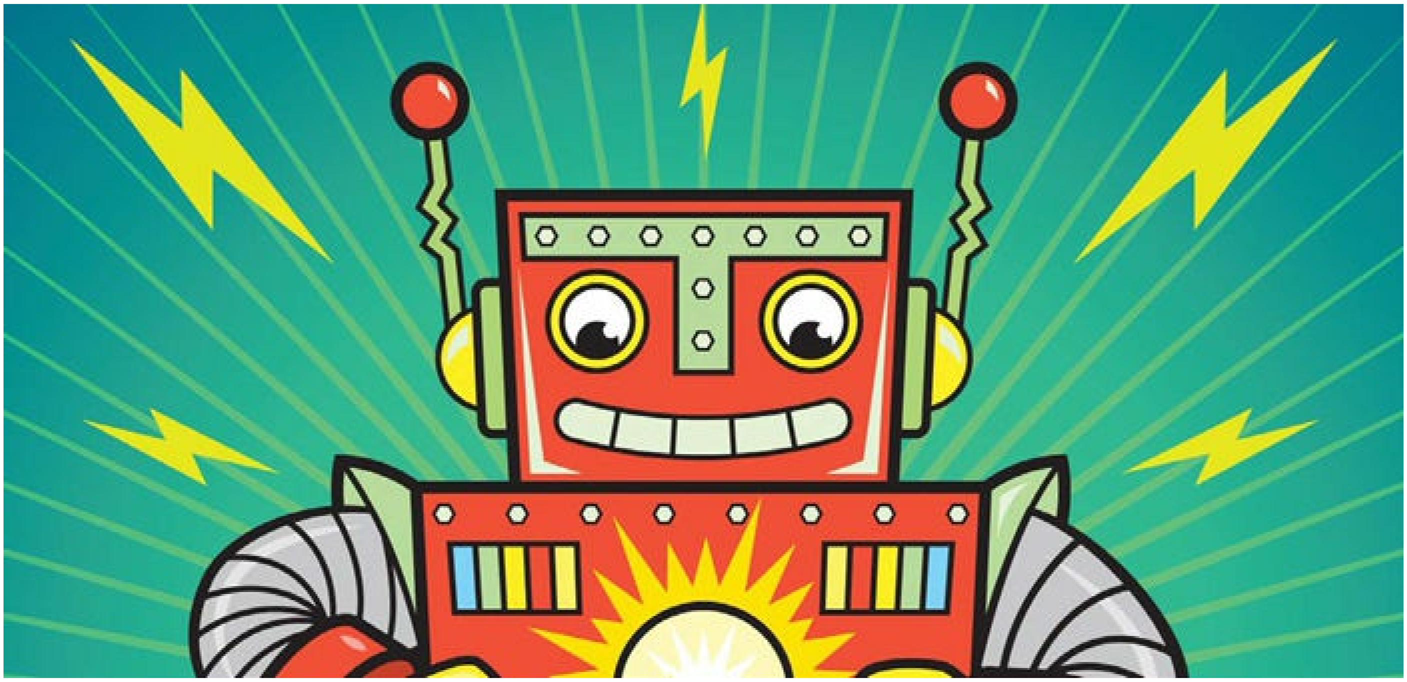

3

HOW TO GENERATE ELECTRICITY

Chapter 1 described why you need a closed loop to get current flowing through a circuit, and Chapter 2 showed you how to build your own electromagnet and motor. The projects in those chapters used electricity from a battery, but in this chapter, you'll make your own electricity sources!

Specifically, you'll learn how to build your own generator, which creates electricity from movement, and your own battery, which creates electricity through chemical reactions. These are two of the most common ways to obtain electricity.

GENERATING ELECTRICITY WITH MAGNETS

When you run current through a wire, it creates a magnetic field around the wire, but there's another connection between electricity and magnetism. You can also create electricity using a wire and a magnet!

A Changing Magnetic Field Creates Electricity

If you move a magnet back and forth over a wire connected in a closed loop, you'll create a current in the wire. Moving the magnet changes the magnetic field around the wire, and the changing magnetic field pushes the electrons through the wire.

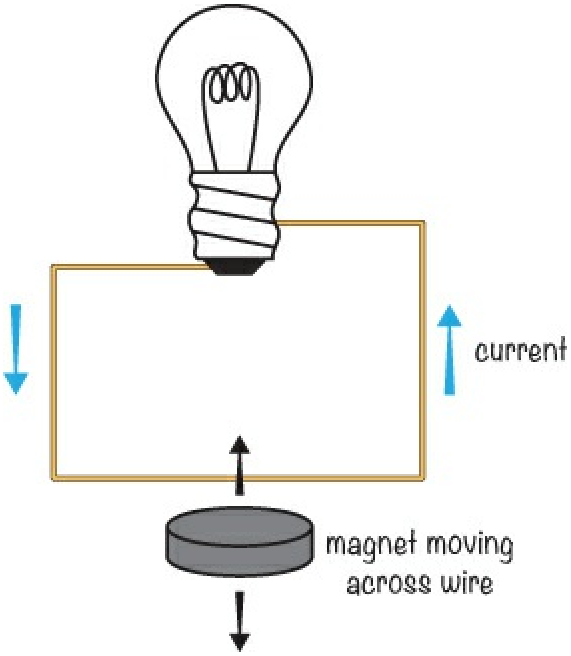

If you stop moving the magnet, the current also stops—even if the wire is still within the magnetic field—because the magnetic field is no longer changing.

If you connect the two ends of the wire to a light bulb and create a closed loop, then the current can flow. Unfortunately, however, the current created by moving a magnet over a single wire doesn't provide enough energy quickly enough to actually light the bulb. To light a bulb, or to power anything else, you need to find a way to generate more *power*, which is the amount of energy produced in a certain time.

How Does a Generator Work?

A *generator* is a device that turns movement—such as the movement of a magnet over a wire—into electricity. To create more power with a wire and a magnet, you can wind that wire into a coil. The coiled wire acts like a group of wires, and when the magnetic field passes through it, a current flows through each coil, creating more power than you could with a straight wire.

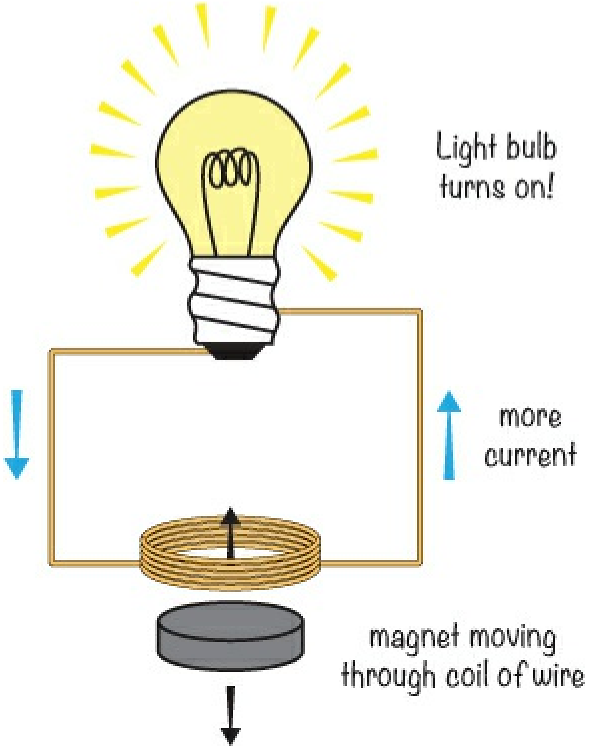

CREATING ELECTRICITY FROM WATER OR WIND

If you place a coil in a magnetic field and rotate the coil with a handle, you're converting your own movement into electricity. If you replaced the handle with a water wheel and placed it into a stream of water, the water would push the wheel so that the coil would rotate in the magnetic field and create a current. This is how some power plants generate electricity! The power plant just lets water run through a wheel that's connected to a generator. Then this electricity is transferred, through power lines, to the power outlets in people's houses.

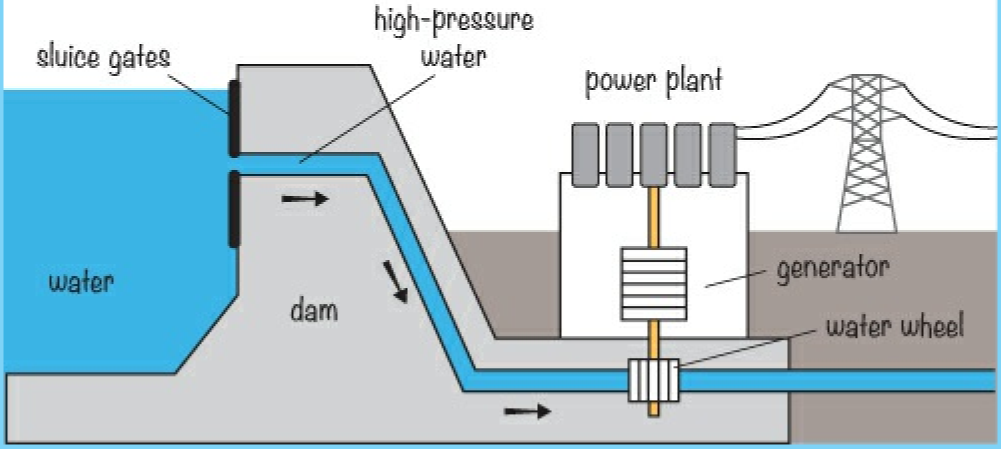
sluice gates
high-pressure water
power plant
water
dam
generator
water wheel

You can make electricity out of other natural forces in the same way. For example, to create electricity out of wind, you can connect the coil to a windmill so that when the wind blows, it rotates the coil.

MEET THE MULTIMETER

You can measure exactly how much energy a simple generator creates with a basic *multimeter*. Multimeters are handy when building any circuit because they can measure a lot of different values, including resistance, current, and voltage.

VOLTCRAFT
VC155

The red lead is the positive lead, the black lead is the negative lead, and the big dial in the middle lets you tell the multimeter what to measure. If you're having problems with a circuit, measuring the voltage at key points in your circuit is one practical way to figure out what's wrong.

How to Measure Voltage

To measure voltage with a multimeter, first turn the dial to one of the V options. (In this book, I'll tell you which setting to choose, but in your own projects, pick one that has a number higher than the highest voltage you expect to see in your circuit.) Then, at the bottom of the multimeter, connect the black lead to the COM socket and the red lead to the V socket. Finally, place one lead on each side of the part you want to measure the voltage across.

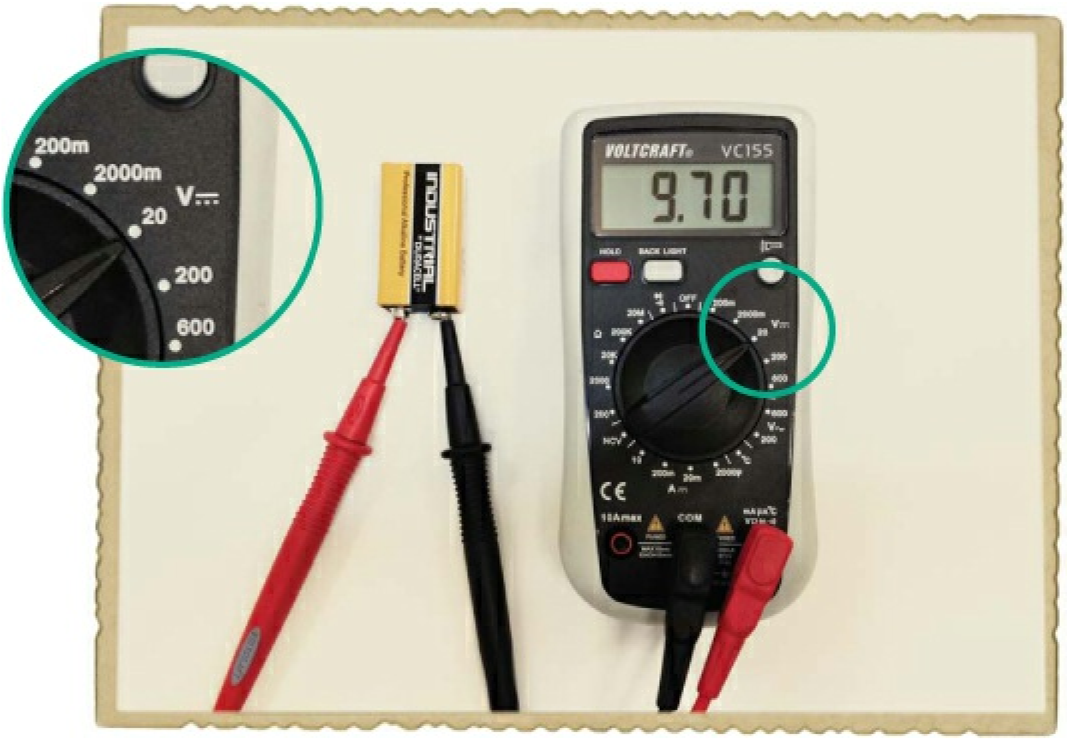

In this example, the meter is measuring the voltage between the positive and negative terminals of a 9 V battery. Notice that my dial is turned to 20 V, in the range showing a V with a straight-line symbol. But there's another V on the multimeter with a wavy line next to it. Let's look at what these symbols mean.

What Are AC and DC?

How you set your multimeter depends on whether you want to measure the voltage from a battery or a generator. A battery has a positive and a negative side, but a generator doesn't! A generator has two wires that alternate between being positive and negative. This is because when one side of the magnet moves past the coil, current in the coil flows in one direction, and when the other side of the magnet moves past the coil, current flows in the other direction.

When the current direction switches like that, we call it *alternating current (AC)*; when the direction of the current stays the same all the time, we call it *direct current (DC)*. Usually, you'll find these symbols on your multimeter to indicate the AC and DC ranges of measurement:

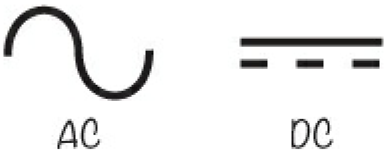

You need to set the multimeter to measure either AC or DC to get the correct reading. For example, batteries have a DC voltage.

PROJECT #5: MAKE A SHAKE GENERATOR

Grab your multimeter—this project will show you how to make a generator and measure its voltage. One quick way to create a simple generator is to manually move a magnet back and forth inside a coil. In this project, you'll put a magnet inside a tube and wind a coil around the tube. When you shake the tube, the magnet should move back and forth inside the coil and create a voltage.

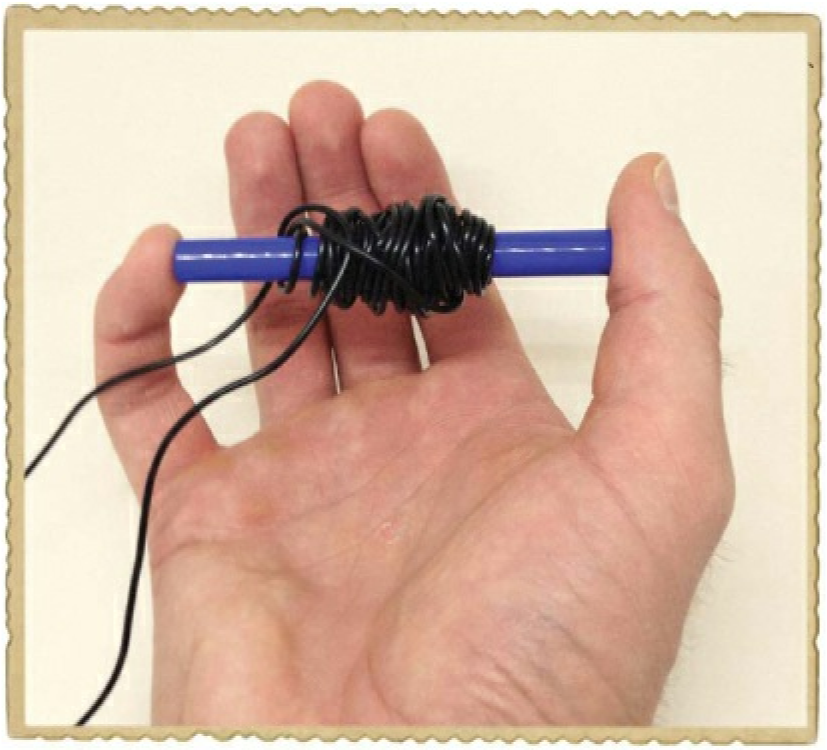

Shopping List

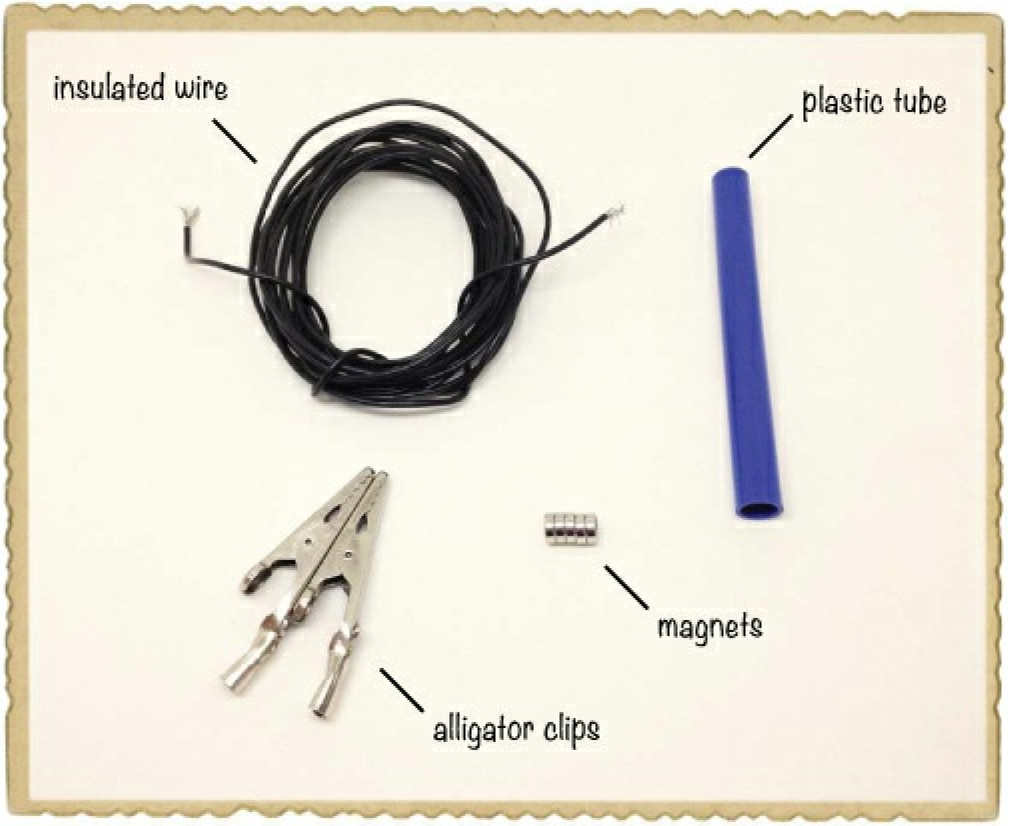
insulated wire
plastic tube
magnets
alligator clips

- **Insulated solid-core wire** (Jameco #36792, Bitsbox #W106BK), about 9 feet. Standard hookup wire works fine.
- **A small plastic tube**, such as an old pen.
- **Five disc magnets** (Jameco #2181319, Bitsbox #HW145) stacked to form a magnet rod.
- **Two alligator clips** (Jameco #256525, Bitsbox #CN262) to connect the multimeter to the coil.

Tools

- **A multimeter** to measure the voltage of your generator. The multimeter should be able to measure very low AC voltages, down to 0.01 V or less. Suitable multimeters are Jameco #2206061, Bitsbox #TL057, or Rapid Electronics #55-6662. These multimeters are a bit more expensive than the cheapest ones, but they will serve you for many years to come.

multimeter

Step 1: Prepare Your Tube

Find a tube that's big enough to let the magnets slide easily back and forth. If you're using a pen, disassemble the pen and make sure your magnets fit inside the tube.

Step 2: Wind Your Coil

Wind about 50 turns of wire around the middle of your tube. After winding, make a simple knot with the two ends to keep your coil together. Then, strip the insulation from the two wire ends, as shown.

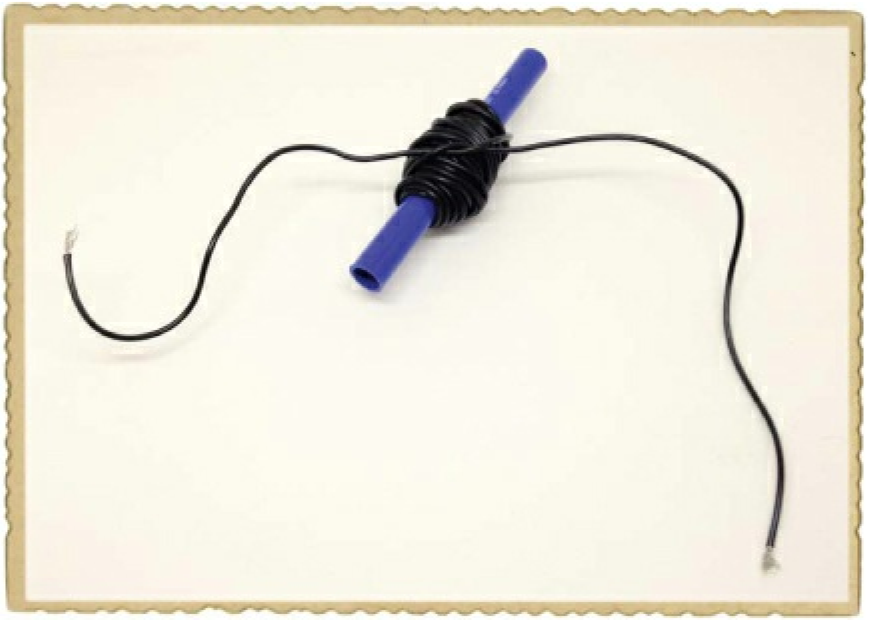

Step 3: Connect the Multimeter

Connect the multimeter to both ends of the coil using alligator clips and set the multimeter to measure AC. Choose the lowest AC voltage setting available.

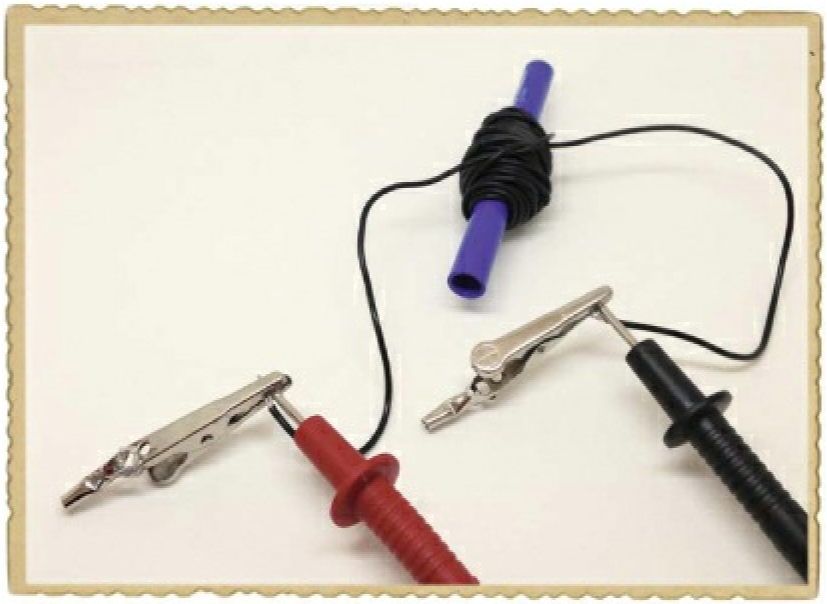

Step 4: Shake That Thing!

Next, put the magnets inside the tube. They should fit inside without coming apart.

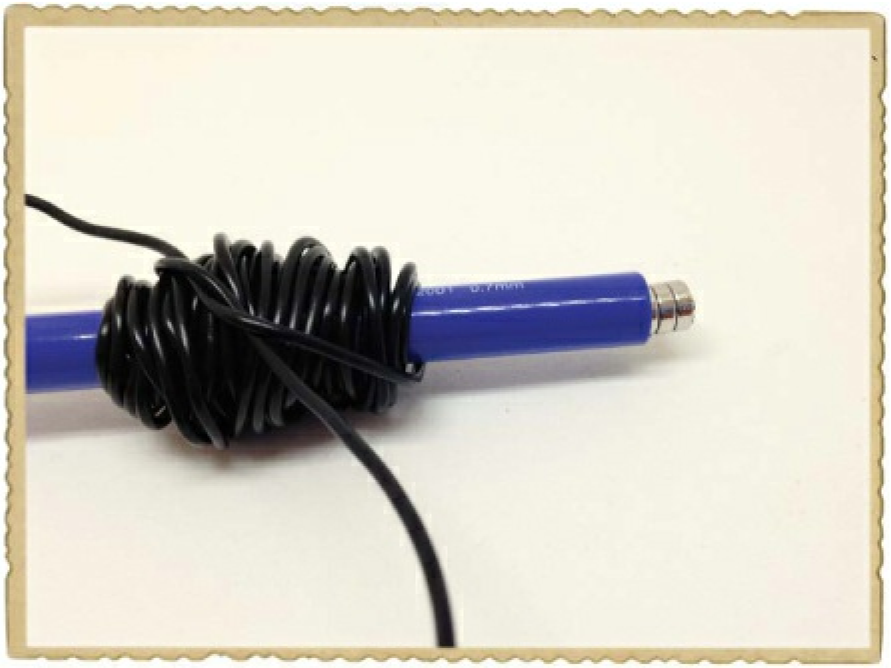

Holding the tube and multimeter leads in your hand, place one finger on each side of the tube so that the magnets don’t fall out. Then, shake it like you mean it!

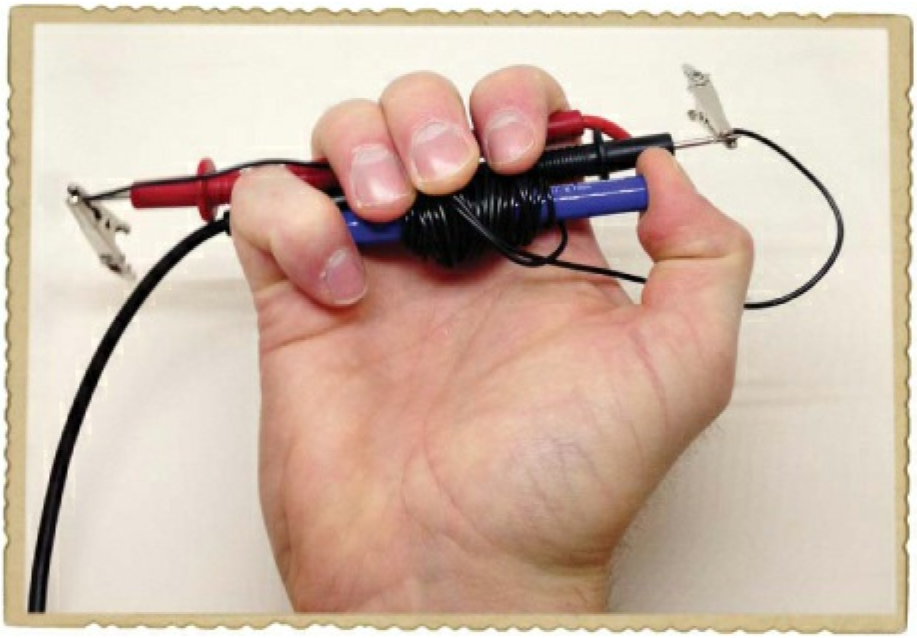

Observe the voltage value on the multimeter. How much voltage do you get? I was able to get only 0.02 V from my generator, so it's not very powerful.

BS3604W
.020
V
AUTO POWER OFF
Ω

Step 5: What If There's No Voltage?

If you can't measure any voltage from your generator, first check that your multimeter leads are connected well to the exposed coil wires. If you still don't see a voltage higher than 0 V, make sure your multimeter is set to measure really low voltages; my dial was turned to 2 V AC. You won't get a high voltage from this simple generator, so if the multimeter isn't on the lowest setting possible, it will keep reading 0 V. Note that not all multimeters are able to measure such low voltages.

This generator isn't very powerful right now. How can you make it more powerful? Try to increase the voltage from the generator by shaking it faster, adding more loops of wire to the coil, or using a more powerful magnet.

NOTE

Standard hookup wire is a bit bulky; even 50 turns take up a lot of space! If you want to get a lot more turns, try using magnet wire *instead. It's really thin wire with a thin layer of insulating coating.*

TRY IT OUT: USING A MOTOR AS A GENERATOR

A motor already has a magnet and a coil of wire that can rotate in the magnet's magnetic field. If you rotate the rotor with your hand, you can generate a voltage on the motor's wires.

You could create a generator by reversing the motor you built in Chapter 2, but the power you'd get from it would be too small to measure. Instead, try to find an old motor from a computer fan or a radio-controlled toy car that you don't want to play with anymore. Then, set your multimeter to a low-voltage DC range, such as 2 V DC. Attach the multimeter leads to the motor wires, just as you did with the shake generator, and turn the rotor with your fingers. Some motors have internal circuits that control the motor, and those circuits can prevent the electricity generated inside the motor from going out to the wires. But if you're lucky and find a motor that doesn't have such circuits, you should see a reading on the multimeter. Try a low-voltage AC range on your multimeter if you see nothing with DC.

HOW DO BATTERIES WORK?

I've shown you how to generate electricity manually, but that doesn't explain how you've powered circuits up to this point in the book. You've been using batteries, and in this section, we'll look at what lets those batteries create electricity.

What's Inside a Battery?

To create a battery, you need three things:

- A positive electrode
- A negative electrode
- An electrolyte

An *electrode* is a wire that is used to make contact with something nonmetallic, like the inside of a battery. An *electrolyte* is a substance that can release or gain electrons.

Here's how these three pieces fit inside a typical battery:

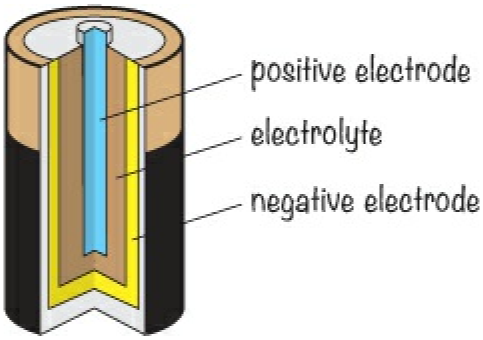

You can actually make your own battery by using a simple nail for one electrode and a copper wire for the other. Stick both into a lemon, and the lemon juice is your electrolyte.

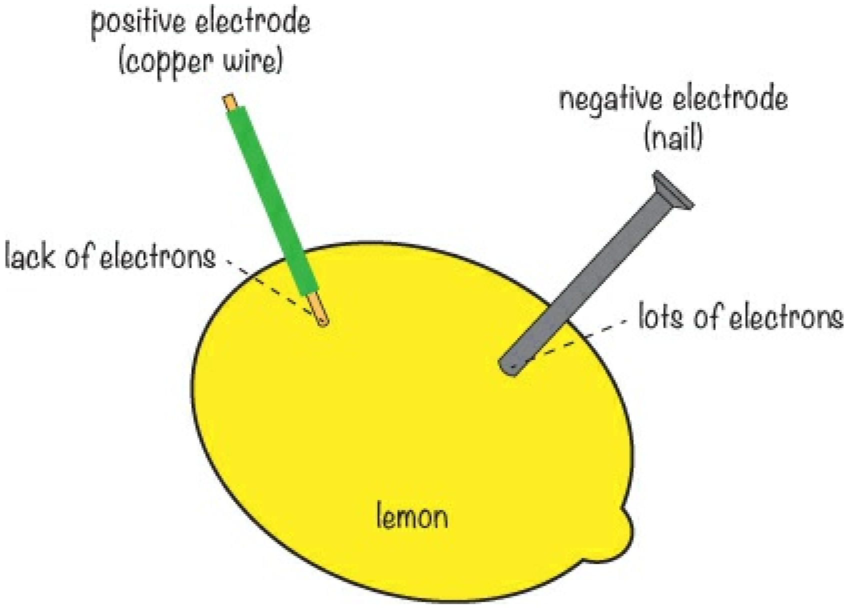

The copper wire becomes the positive terminal of the battery, and the nail becomes the negative terminal.

The Chemistry Behind Batteries

When you combine the lemon, the copper wire, and the nail, two chemical reactions happen: one between the lemon juice and the nail, and another between the lemon juice and the copper wire. In the first reaction, electrons build up on the nail; in the second, electrons leave the copper wire. The nail gets too crowded with electrons, and the copper wire ends up with too few. Electrons don't like to be in crowded places, so the electrons on the nail want to go over to the copper wire to even things out. But the chemical reactions with the lemon juice are pushing the electrons the other way.

Now, what do you think will happen if you connect a light bulb between the nail and the copper wire? The electrons on the nail really want to get to the copper wire, so they'll take the easiest path they can find, and when you create this closed-loop circuit, they flow from the nail to the copper wire through the light bulb. Recall that current is just electrons flowing in a wire; if you have enough current flowing through the light bulb, it lights up!

After a while, the chemical reactions in the battery stop. When this happens, the battery is dead. Some batteries can be recharged when they die, while others must be thrown away. The materials chosen for the electrodes and electrolyte determine whether the battery can be recharged or not.

The batteries you buy in the store are not made of lemons, of course! Modern batteries are made from different materials, and scientists are always looking for new ways to create batteries that have more energy, while being small and lightweight.

What Determines a Battery's Voltage?

The materials used for the electrodes and electrolyte determine the voltage you get from a battery, but the size of the electrodes and the amount of electrolyte don't matter when it comes to voltage.

To create higher battery voltages, several battery cells are connected in *series*. Connecting two battery cells in series means that you connect the positive side of one battery to the negative side of the other. The two unconnected terminals become the bigger battery's new positive and negative terminals, and the resulting voltage is the sum of the voltages from the two batteries. For example, in a standard 9 V battery, you have six 1.5 V battery cells, as shown. Notice that the connectors on the outside are attached to just two terminals.

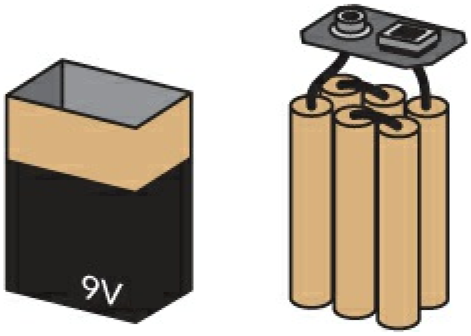

PROJECT #6: TURN ON A LIGHT WITH LEMON POWER

You can make a battery out of many different things; for example, in "What's Inside a Battery?" on page 55, I showed you how a lemon battery might work. In this project, you'll learn how to build a lemon battery of your own and power a light with it.

WARNING

When you're finished with this project, throw the lemons away. The chemical reactions that happen with the nail and copper wire will leave the lemons unsuitable for eating.

Meet the LED

A lemon battery can't create a lot of electricity, so you need to connect the battery to something that needs very little power to see the effect. Most light bulbs need more power than you'll generate in this project, so let me introduce a component called a *light-emitting diode*, or *LED*.

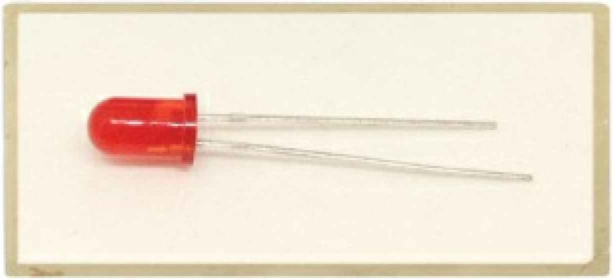

This little electronic component gives off, or *emits*, light when you apply a little bit of power to it. LEDs come in many colors: red, green, yellow, blue, and more. You'll learn more about this component in Chapter 4, and you'll use LEDs a lot in this book. For now, you're just going to use an LED to see the power generated by your lemon battery.

Shopping List

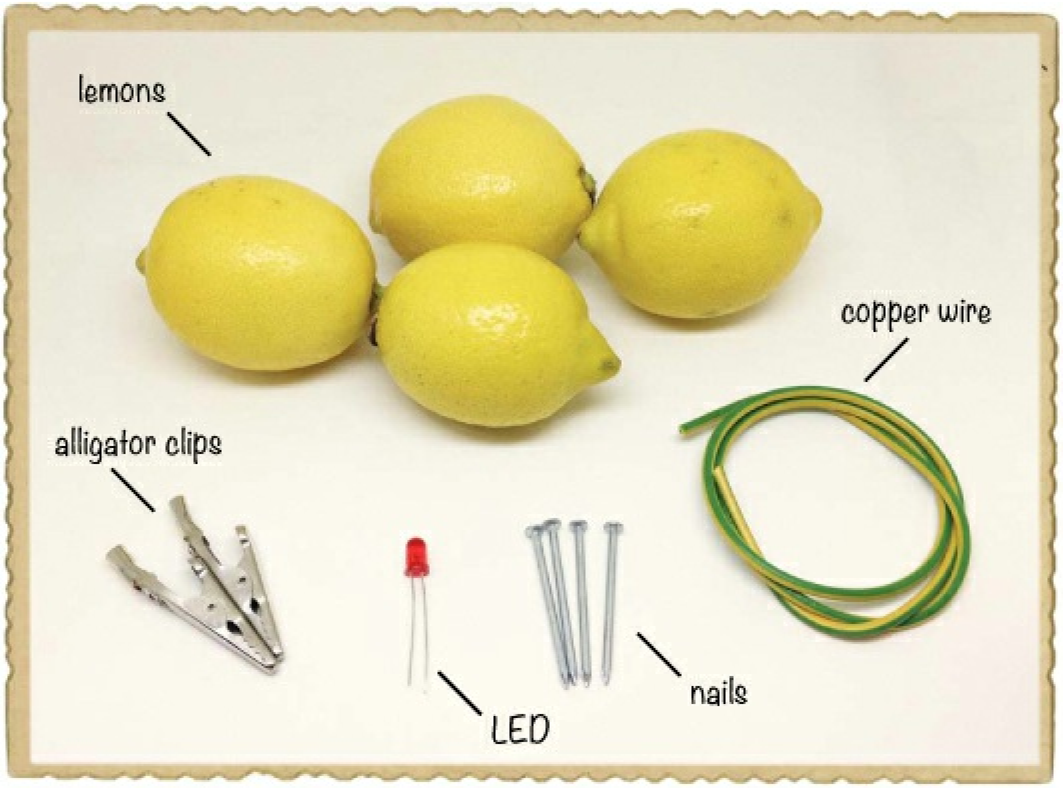

- **Four lemons** or one lemon cut into four pieces.
- **24 inches of copper wire** (any copper wire will do, but it's important that the wire be copper).
- **Four galvanized nails** (most common nails for outdoor projects are galvanized).
- **Two alligator clips** (Jameco #256525, Bitsbox #CN262) for connecting the LED.
- **A standard LED** (Jameco #333973, Bitsbox #OP002 for just this one, or Jameco #18041, Bitsbox #K033 for a variety pack). You'll need several LEDs for the projects in this book, so order at least 10 or a variety pack.

Tools

multimeter and leads
wire cutter

- **A wire cutter** (Jameco #35482, Bitsbox #TL008) to prepare the copper wire.
- **A multimeter** (Jameco #2206061, Bitsbox #TL057, Rapid Electronics #55-6662) to see whether your battery is working correctly.

Step 1: Prepare Your Wires

First, cut your copper wire into four 6-inch lengths. Strip about 1 inch of insulation from both ends of each wire. These will become the electrodes.

Step 2: Insert Electrodes into a Lemon

Roll and squeeze a lemon so that you break up the small juice packets inside it, but not enough to break the skin. Then, use a nail to make one hole in one end, push a copper wire into that hole, and push the nail into the other end, as shown. This is the first lemon battery!

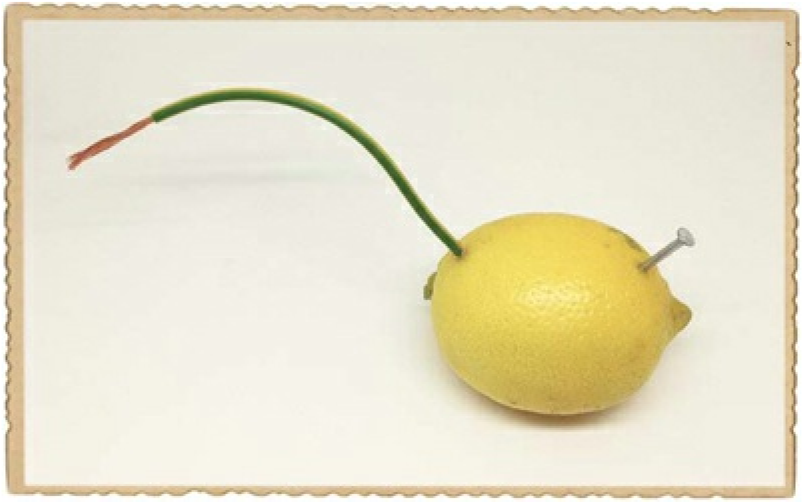

Get your multimeter, set it for DC voltage measurement, and test your lemon battery now. Place the positive test lead on the copper wire and the negative test lead on the nail. If everything works correctly, you should see a voltage of around 1 V on your multimeter.

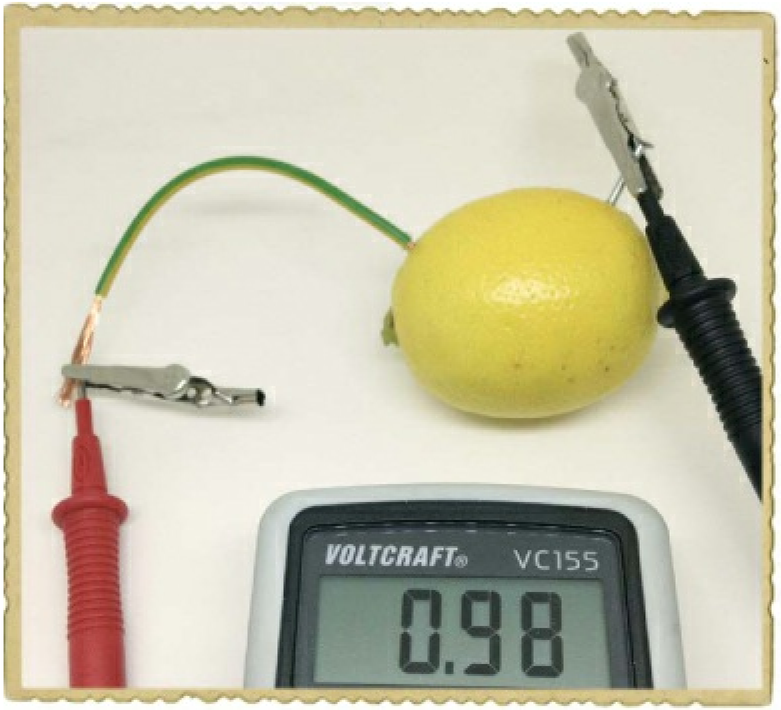

Step 3: Create Four Lemon Batteries

Even if you get 1 V out of your lemon, that's not enough to light an LED. Let's create several lemon batteries so we can get more electricity!

Just repeat the process described in Step 2 for the other lemons; each will become a battery. (If you don't have four lemons to spare, you can cut one lemon into four pieces.) Now you should have four lemon batteries.

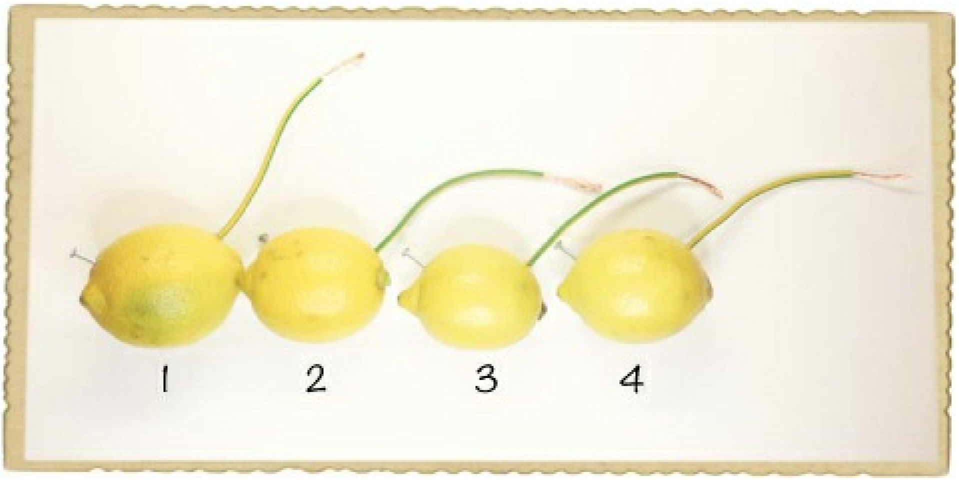

Step 4: Connect the Lemons in Series

To get a higher voltage with your lemon batteries, you'll need to connect them in series. To connect two lemons in series, you just connect the positive side of one lemon to the negative side of another. Remember, the copper wire is positive, and the nail is negative.

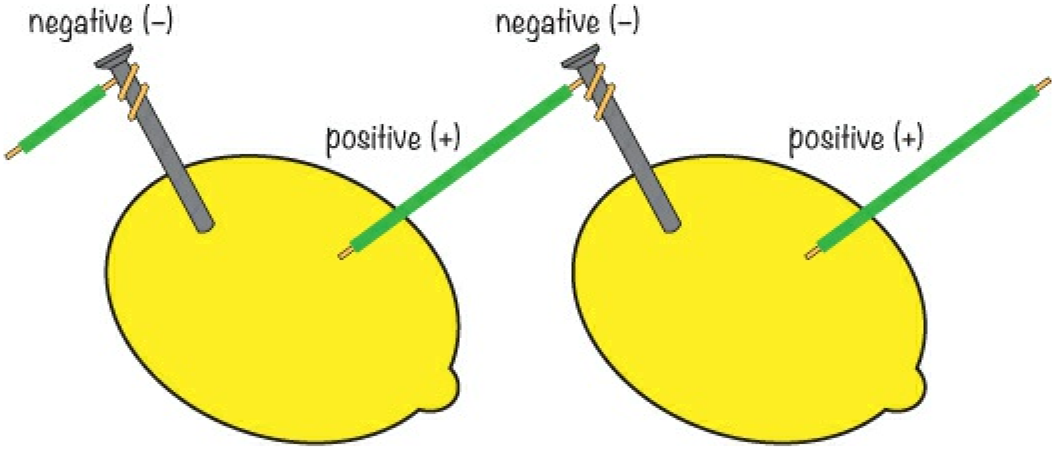

To wire four lemons in series, just repeat that process a couple more times. Line your lemons up in a row with the copper wires pointing to the right and number the lemons from 1 to 4, beginning from the left. Connect the copper wire from lemon 1 to the nail in lemon 2. Twist the wire onto the nail so that the metals connect without coming apart.

Connect the copper wire of lemon 2 to the nail in lemon 3, and connect the copper wire from lemon 3 to the nail in lemon 4. This should give you a row of four lemons, with an unconnected nail on lemon 1 and an unconnected copper wire on lemon 4. These are the positive and negative terminals for your big lemon battery, respectively.

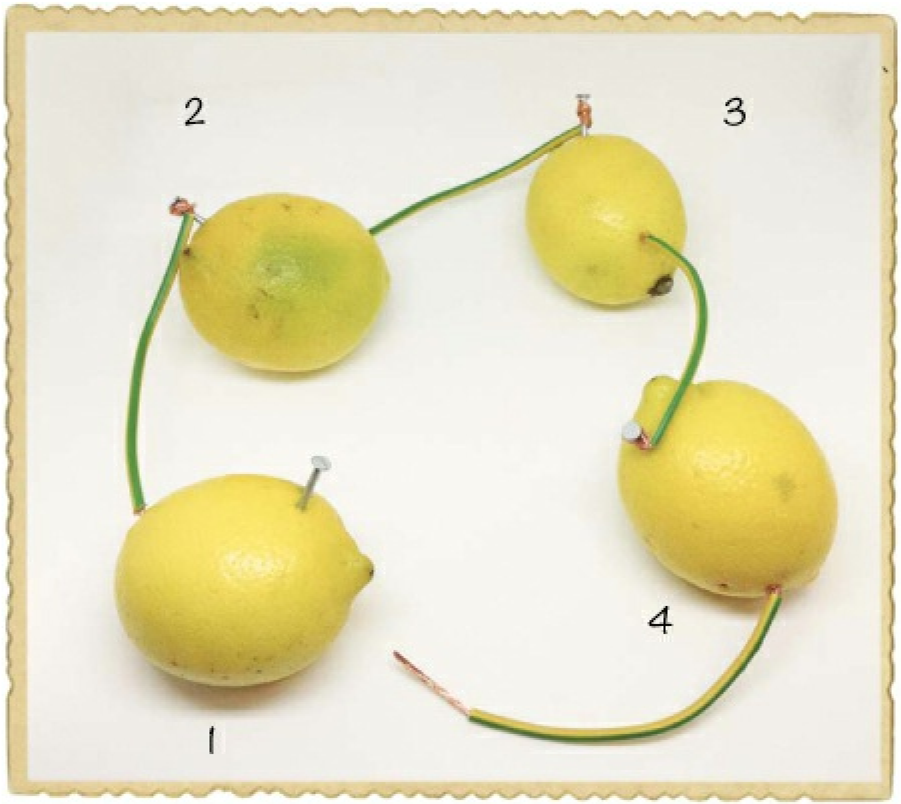

When you connect batteries in series, you can add their voltages to find your total. Four 1 V lemon batteries should give you 4 V. If you have a multimeter, measure the voltage between the two ends to see whether everything is connected. You should get a voltage of around 3.5 to 4 V.

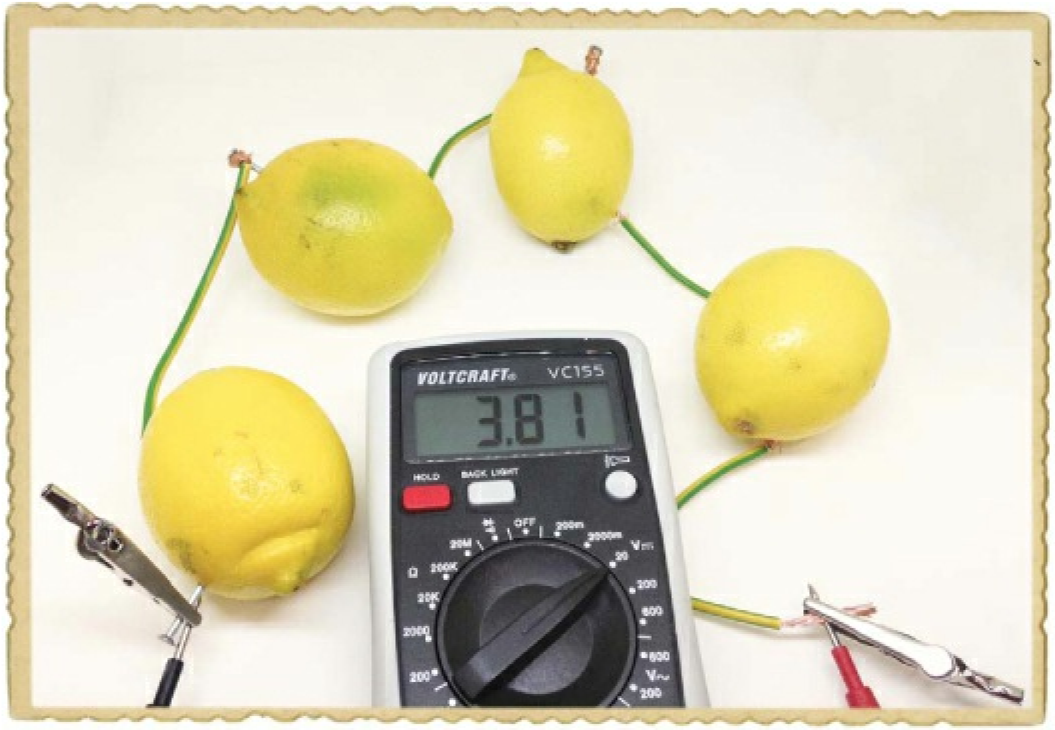

Step 5: Test Your Lemon Battery

Let’s connect the LED to the lemons! Connect the long leg from the LED to the copper wire, and connect the short leg to the nail, as shown. The LED should now light up.

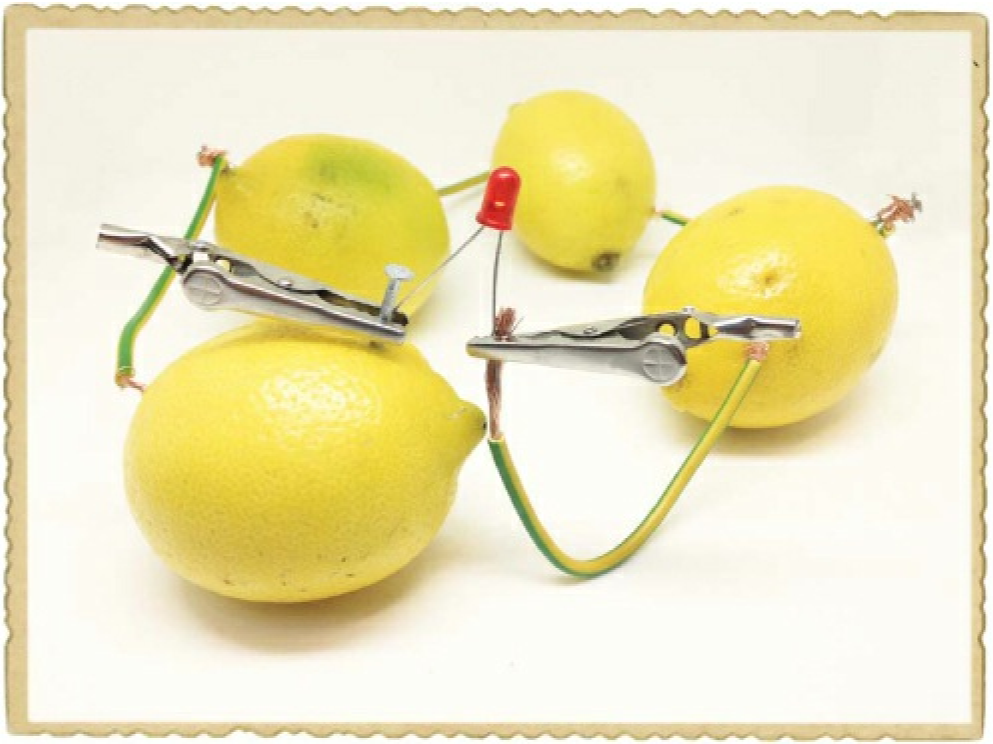

Lemons aren't super powerful batteries (you'd never see anyone with a lemon connected to their computer, for example), so your LED will probably be very dim. After you finish building your lemon-powered circuit, turn off the light in your room, and you should see the LED glow.

Remember, when you're finished with your lemon battery, throw the lemons away—don't eat them!

TRY IT OUT: MORE FOOD BATTERIES!

When you're done making lemon batteries, test to see whether you can make batteries out of other fruits or vegetables. For example, what about a potato battery? Are you able to get more voltage, or is it the same as the voltage from the lemon?

Step 6: What If Your Lemon Light Doesn't Work?

If you can't see light from your LED, even in a dark room, check to see whether your LED is connected the right way. The long leg should be connected to the positive side of the battery, which is the copper wire.

Make sure the lemons are connected to each other only through the wires and nails. For example, if your lemons are sitting in a puddle of lemon juice, they could be connected through that. Just dry them off and move them somewhere else. Next, check that the copper wires are properly connected to the nails and that the nails and copper wires are actually touching the juice inside the lemons. Also, check that the nails and copper wires are not touching each other inside any of the lemons.

If the circuit still doesn't work, disconnect all the lemon batteries from each other. Then, use a multimeter to check that each lemon battery has some voltage. Connect two lemons in series, and check that you see a higher voltage. Connect the third lemon, and check that the voltage has increased again. Then, connect the fourth lemon and check that you have even more voltage.

If you see a voltage but the LED doesn’t light, then you probably just need some more power. Get another lemon or two, create some more batteries, and connect them in series with the rest.

WHAT’S NEXT?

In this chapter, you learned how to create your own electricity from magnetism and chemical reactions. You made your own shake generator, and you built a lemon battery to power an LED.

If you want to explore generators even more, I suggest trying to find a *dynamo* from an old bike. Unlike the generator you built in this chapter, a dynamo is a generator that gives you a DC voltage, like a battery, and dynamos are commonly used to power headlights on bikes. Cut some windmill blades out of some stiff cardboard or plastic, connect them to the dynamo, and see whether you can harvest energy from the wind.

You’ve now met a few electronic components, including switches, LEDs, and motors. In the following chapters, you’ll learn about even more components and graduate to building some real electronic circuits, like lights that blink, a touch-sensitive switch, and even your own electronic musical instrument!

PART 2

BUILDING CIRCUITS

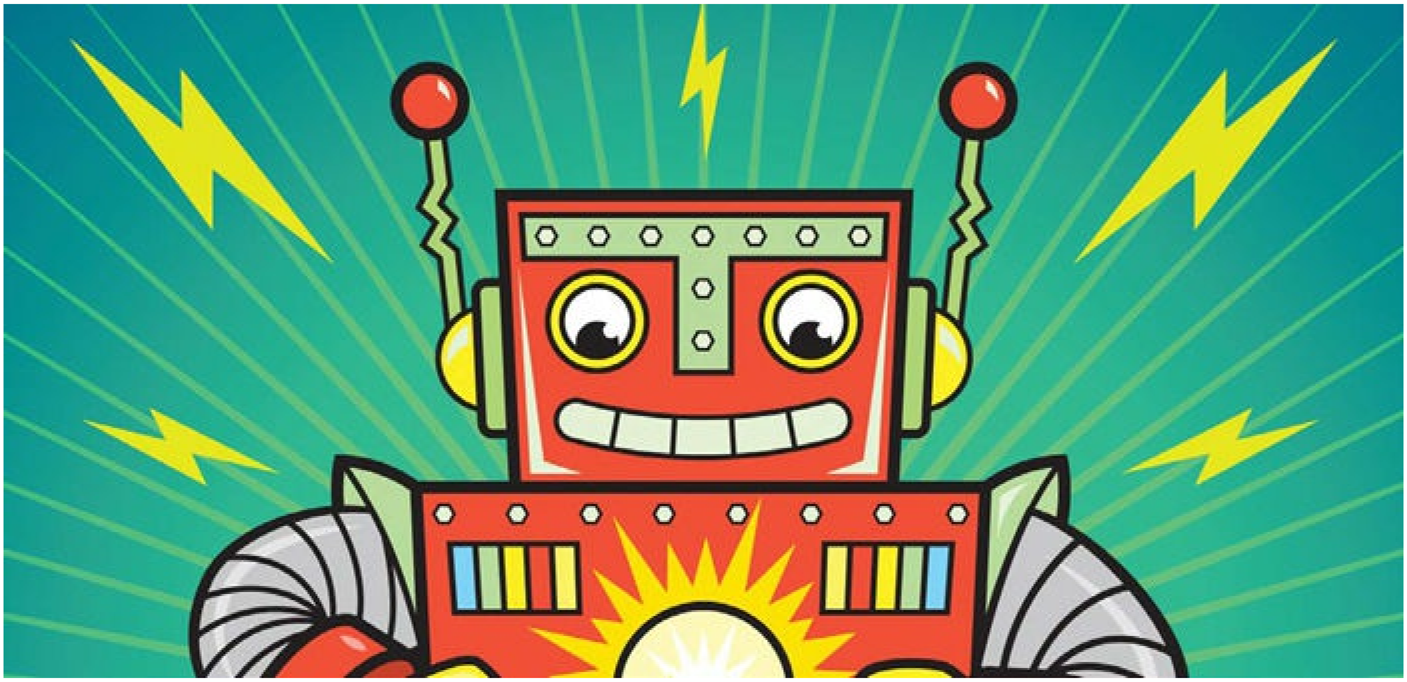

4

CREATING LIGHT WITH LEDS

Lights, especially LEDs, are used all the time in electronics. Sometimes, they're just simple indicators that show whether a device is on or not, but they can also be part of more complicated devices, like computer displays. In fact, some displays are actually made up of thousands of tiny LEDs.

In this chapter, you'll learn how two of the most common basic components in electronics work: the resistor and the LED. I'll show you how to kill an LED, but don't worry: you'll learn how to use resistors to keep LEDs alive, too. In this chapter's projects, you're also going to start using a new tool, called a *breadboard*, to connect circuits. Many projects in this book use breadboards, and you can also use them to build a lot of cool projects on your own.

MEET THE RESISTOR

Recall that resistance restricts current from flowing freely in a circuit. A *resistor* is a component that adds resistance to a circuit. The more resistance your circuit has, the less current will flow through it.

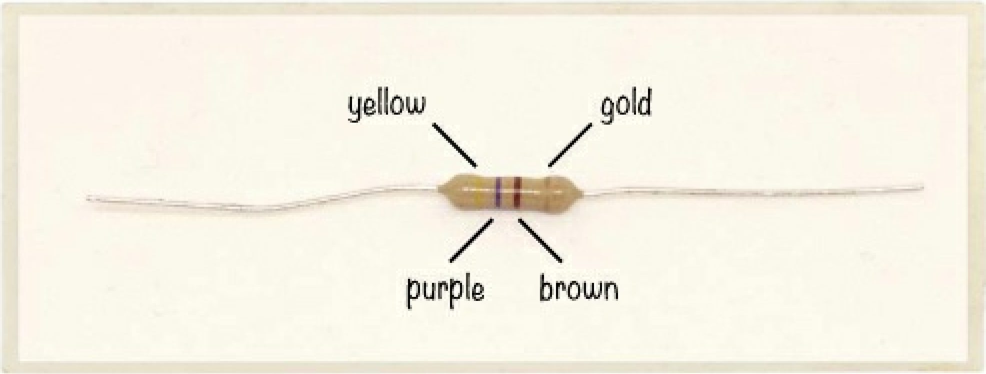

Resistor Color Codes

When you look at a resistor, you'll notice that it has several colored bands. These colors tell you the value of the resistor. Resistance is measured in *ohms*, but when we write about it, we'll use an *omega* symbol, *Ω*, for short. More ohms means more resistance.

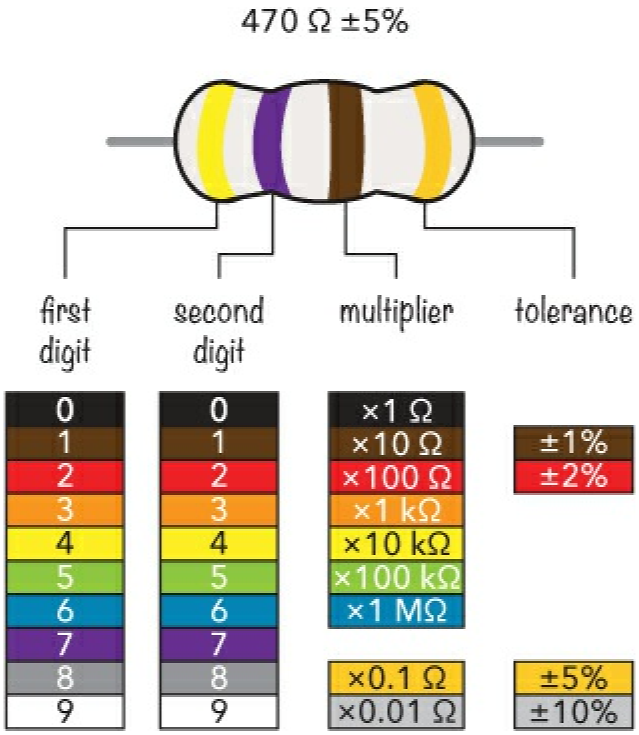

Most resistors have four color bands. From the left, the first band gives the first digit of the resistance value. In this example, the first band is yellow, so the first digit is 4. The second digit is given by the second band, which is purple for 7. Together, this gives us the base value of 47. Next we multiply 47 by the value of the third band—the *multiplier*. In this example, the brown band stands for 10 Ω, so we multiply 47 by 10:

$$47 \times 10\ \Omega = 470\ \Omega$$

NOTE

If a resistor has five bands instead of four, then the first three bands are digits, and the fourth is the multiplier.

But the actual resistance of a resistor usually won't match the value written on it! That sounds a bit crazy, right? It's hard for manufacturers to create resistors with a very exact resistance value, so instead, they make sure the resistors are somewhere around that value and tell you how far off the real value could be.

This is where *tolerance* comes in. Our example resistor is labeled 470 Ω with a tolerance of 5 percent. This means that the resistor's real resistance could be any value 5 percent higher or 5 percent lower than 470 Ω. Because 5 percent of 470 is around 24, the real resistance could be anywhere between 446 Ω and 494 Ω.

Usually, the three bands that tell you the resistance value are grouped together, and the band that tells you the tolerance is spaced a bit farther away. But sometimes the bands are so close that it's hard to see which three bands give the resistance. Fortunately, the fourth band is typically gold or silver, so if you see a gold or silver band, it's safe to assume this is the tolerance band.

HOW TO WRITE LARGE VALUES

Our resistance color chart shows some resistance values written with *k* and *M* in front of the Ω symbol. These are part of a shorthand that makes it easier to write really large values. If you have a resistor that's 300,000 Ω, it's common to shorten that to 300 kΩ instead, where *k* is short for *kilo*, which means one thousand. The *M* stands for *mega*, which means one million. So instead of writing 3,000,000 Ω, you could write 3 MΩ.

What Are Resistors Made Of?

To create a resistor, you could just use a really long piece of standard wire. Wires have a bit of resistance, and the longer your wire is, the more resistance you'll get. But using miles of wire to reduce current isn't very efficient. It's better to use a material that has more resistance, such as carbon. Often the resistors that you buy in stores are made of carbon wrapped inside an insulating material.

Resistors Control Current and Voltage

At first, you might find the resistor a bit boring. If you connect one to a battery, you probably won't see anything happen; the resistor might just get warm, and you might wonder what the big deal is. On the other hand, if you use a resistor with a very low resistance value, such as 10 Ω, then it could get *really* hot—hot enough to give you a burn—and the battery might die pretty quickly.

WARNING

Connecting a low-value resistor directly between positive and negative points can be dangerous on some types of batteries. Some batteries are strong enough to make your resistor burst into flames. Be careful!

But the cool thing about resistors is that you can use them to change the voltages and currents in your circuit! That means that you get to be the master of your circuit and decide how it should behave.

INTRODUCING OHM'S LAW

The key to controlling the current and voltage in your circuit is a formula called *Ohm's law*. Ohm's law relates resistance, voltage, and current as follows:

$$V = I \times R$$

Here's what those letters mean:

V Voltage, measured in volts (V)

I Current, measured in amps (A)

R Resistance, measured in ohms (Ω)

Given these definitions, in English, Ohm's law reads, "Voltage equals current multiplied by resistance." You can also write the Ohm's law formula in the two following forms:

$$R = \frac{V}{I} \qquad I = \frac{V}{R}$$

Let's put Ohm's law to work. Imagine you have a resistor and a 9 V battery, and you want 0.05 A of current to flow in the resistor. How much resistance do you need in the resistor to get the right amount of current flowing? Use Ohm's law to find out:

$$R = \frac{V}{I}$$

$$R = \frac{9\text{ V}}{0.05\text{ A}}$$

$$R = 180\ \Omega$$

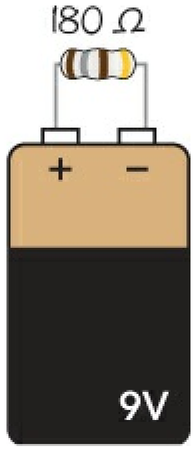

After dividing the voltage by the current, you'll find that to get 0.05 A of current flowing in the resistor, you need a 180 Ω resistor.

PROJECT #7: LET'S DESTROY AN LED!

Almost all electronics have some LEDs, which I introduced in Chapter 3. Where there are LEDs, there are also resistors. Look around a house, and there's a big chance you'll see a few. For example, check a computer, a washing machine, a television, or a Wi-Fi router. Do you see some blinking lights when you push buttons? Those are very likely LEDs in series with resistors.

In "Project #6: Turn On a Light with Lemon Power" on page 58, you just connected an LED to your homemade lemon battery, and that was it. In most circuits, however, you need to take a bit more care to make sure you don't break your LED. If too much current flows through an LED, it becomes really hot and burns out. The lemon battery was too weak to provide enough current to break the LED.

Of course, I could tell you all this forever, but trying things in real life is the best way to learn! I had to break a few LEDs myself before I accepted that I couldn't connect them directly to a battery without a resistor, and I want you to see what that's like, too. That's why in this project, you're going to destroy an LED!

Shopping List

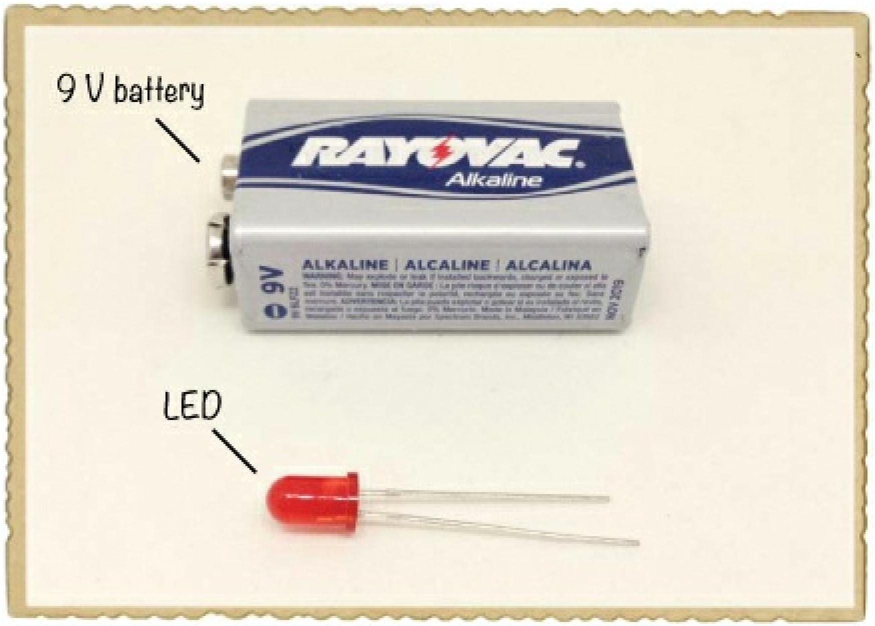

- **A standard LED** (Jameco #333973, Bitsbox #OP002).
- **A standard 9 V battery** to power the circuit.

Step 1: Identify Which LED Leg Is Which

Look at your LED closely, and you should see that one leg is longer than the other. LEDs are *polarized*, which means that current flows through them only if you connect them a certain way in your circuit. The longer leg is called the *anode*; it's the leg that you connect to the positive side of the battery. The shorter leg is called the *cathode*, and you connect it to the negative side of the battery.

On some LEDs, the legs are the same length. In that case, find the flat side on the bottom of the LED itself. The leg on the flat side is the cathode.

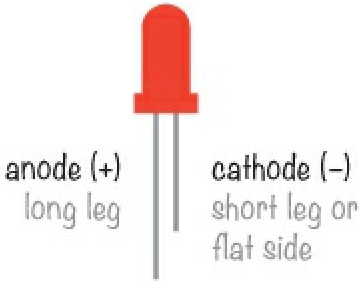

Step 2: Break That LED!

To avoid burning your fingers, hold your LED by one of the legs. Then, place the 9 V battery on the table and touch the legs of the LED directly to the battery terminals.

The LED should glow brightly for a short moment, become hot, and then go dark. Parts of it may actually turn black. Congratulations: You just broke your first LED!

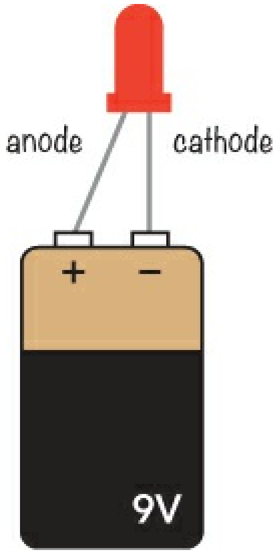

NOTE

Some LEDs stop working after a second when connected directly to a battery. Others may give a bit of light for a few seconds.

Step 3: What If Nothing Happens to the LED?

If nothing happens, there are three likely causes:

- You connected the LED backward.
- Your LED is already broken.
- Your battery is dead.

First, try connecting your LED to the battery the other way around. If you're sure it's connected the right way, then either your LED is already broken or your battery is dead. Try replacing the battery first; if that doesn't work, replace the LED. Now, you should be able to break your LED.

HOW TO USE AN LED CORRECTLY

Even though it's pretty fun to destroy LEDs, it's better to know how to *avoid* destroying an LED. Your LED burned because it had too much current running through it, but you can prevent that with your trusted friend the resistor. Resistors resist the flow of current, and if you choose the right resistance value, they'll resist the current enough to get just the right amount of current for your LED.

HOW TO WRITE SMALL VALUES

In electronics projects, you often need to deal with very small values, especially when measuring or calculating current. For example, most current values in the circuits in this book are less than 0.1 A, and many are closer to 0.02 A. To simplify writing these values, I typically use the prefix *milli*, which is written as a lowercase *m*. It means one thousandth, so 1 mA is 0.001 A. Because

1,000 mA is the same as 1 A, 0.02 A becomes 20 mA, and 0.1 A becomes 100 mA.

Protecting Your LED with a Resistor

An LED in a circuit should always have a resistor in series with it. Of course, resistors come in many different values, and to figure out the right one for your circuit, you need to do a little math.

Most standard LEDs need a voltage of about 2 V and a current of about 20 mA, or 0.02 A, to light up. These two values, together with the voltage of your battery, are all you need to figure out the correct resistance. Just put these two values into the following formula:

$$R = \frac{V_{\text{BAT}} - V_{\text{LED}}}{I_{\text{LED}}}$$

If this formula looks familiar, that's because it's actually just another version of Ohm's law. The two Vs and the I are still voltage and current, but V_{BAT} is the battery voltage, V_{LED} is the voltage your LED needs to light up (often 2 V), and I_{LED} is the current your LED needs (often 20 mA). You'd read this formula as "To find the resistance, subtract the LED voltage from the battery voltage and divide the result by the LED current."

Calculating the Resistance You Need

Imagine you have a 9 V battery, a resistor, and a standard LED. What resistance value should the resistor be? Using the formula from the previous section, you should get:

❶ $$R = \frac{V_{\text{BAT}} - V_{\text{LED}}}{I_{\text{LED}}}$$

❷ $$R = \frac{9\text{ V} - 2\text{ V}}{20\text{ mA}}$$

❸ $$R = \frac{7\text{ V}}{0.02\text{ A}}$$

❹ $$R = 350\ \Omega$$

That means you need a resistor of 350 Ω to get the right amount of current flowing through the circuit.

PROJECT #8: POWERING AN LED

Now let's power a standard LED with a protective resistor so the LED doesn't burn out. We just calculated that to power an LED with a 9 V battery, you need a resistor of 350 Ω.

But as I explained in "Resistor Color Codes" on page 70, standard resistor values aren't always exactly the resistance you need. If you buy a 350 Ω resistor, it isn't necessarily 350 Ω, but maybe 370 Ω. And not all resistance values are even available. For a resistor in an LED circuit, having the exact value isn't important. That's fortunate because you won't find any 350 Ω resistors in standard resistor packs. Instead, you can use a 330 Ω resistor, which is a standard value that's easier to find.

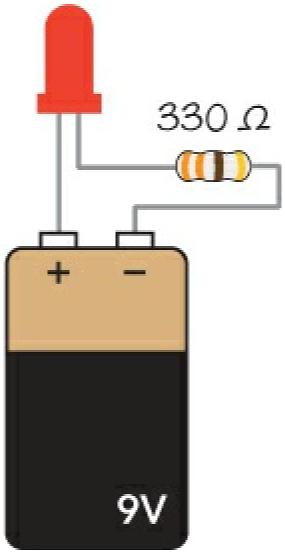

Shopping List

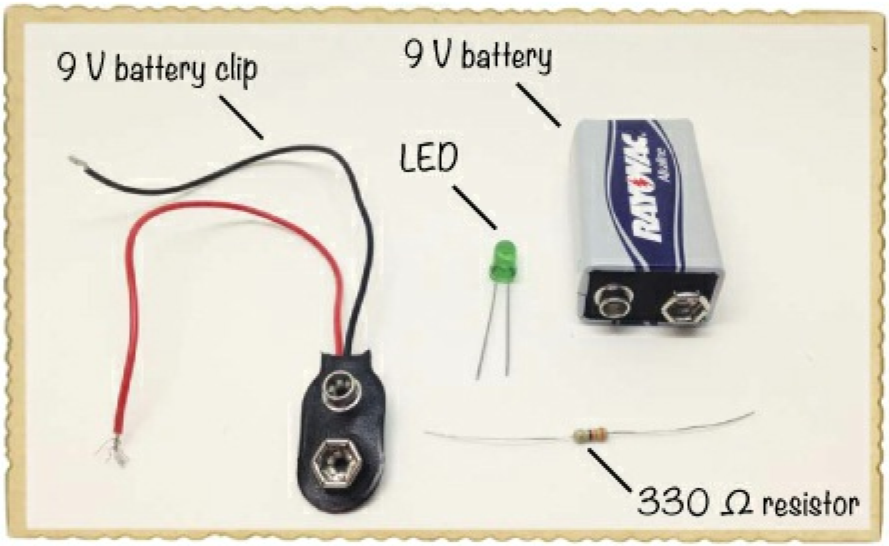

- **A standard 9 V battery** to power the circuit.
- **A 9 V battery clip** (Jameco #11280, Bitsbox #BAT033) to connect the battery to the circuit.
- **A standard LED** (Jameco #333973, Bitsbox #OP002)
- **A 330 Ω resistor** (Jameco #661386, Bitsbox #CR25330R for just this value or Jameco #2217511, Bitsbox #K017 for a variety pack) for limiting the current to the LED.

Step 1: Twist the Resistor and LED

First, connect the short leg, or the cathode, of the LED to one side of the resistor. It doesn’t matter which side of the resistor you connect; just twist the resistor leg around the LED leg.

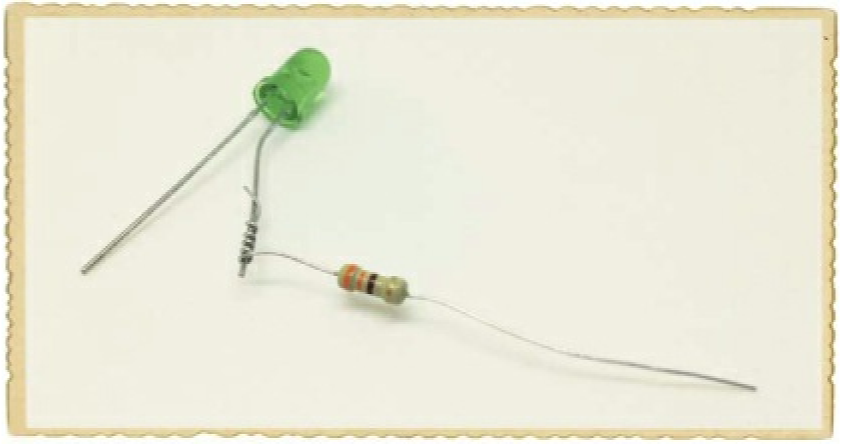

Step 2: Wire the Battery Clip

Twist the battery clip's red wire onto the long leg of the LED. Then twist the black wire to the unconnected side of the resistor.

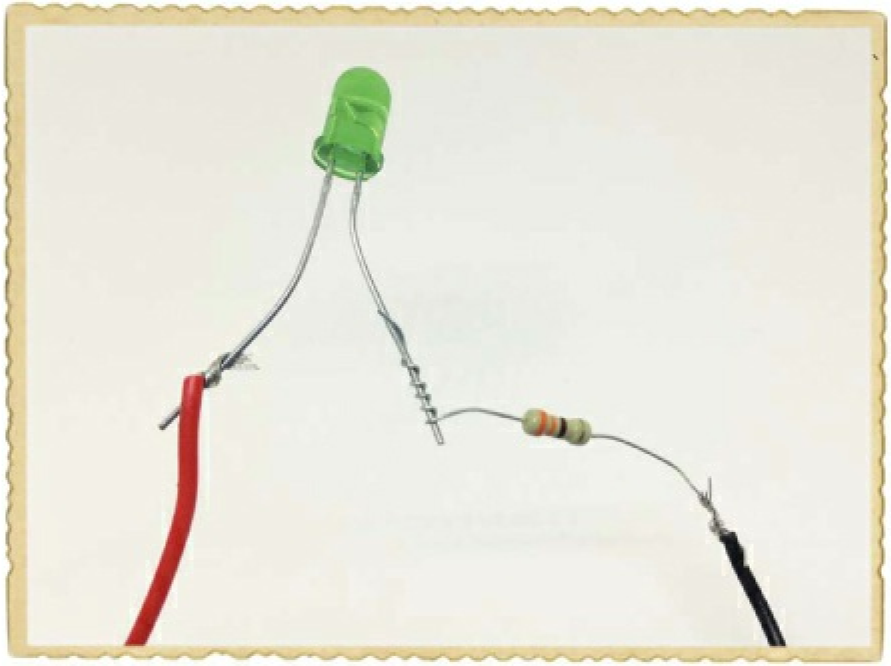

Step 3: Let There Be Light!

Now, plug your battery into the clip, and your LED should glow!

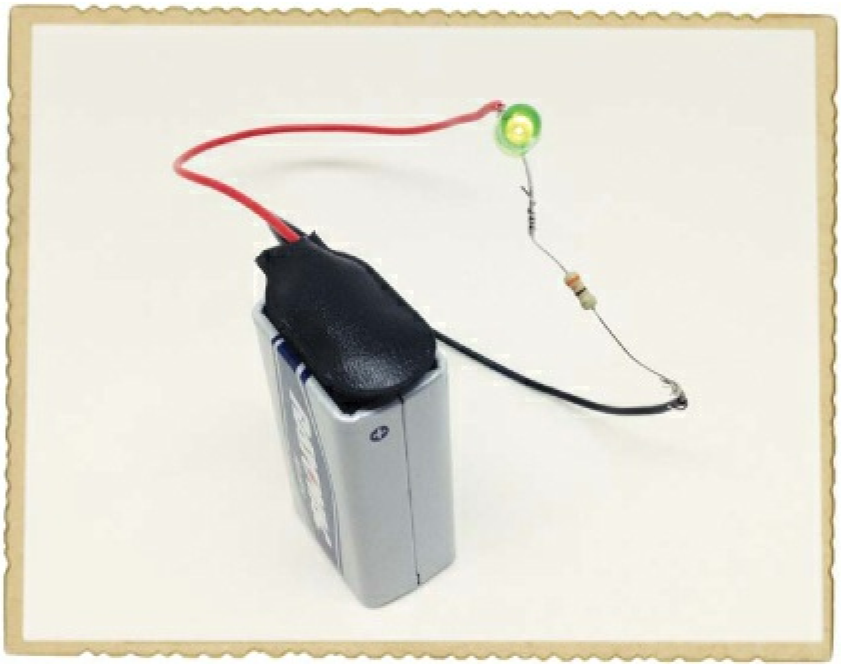

Step 4: What If the LED Doesn't Work?

If your LED doesn't turn on, first disconnect the battery and make sure you've connected the components exactly as I described in Steps 1 through 3. Having someone else review your wiring can be helpful, too; ask a parent, sibling, or friend to look it over.

If your connections look right and the LED is still dark, then double-check the LED's orientation; just about anyone who's ever built an electronics project has connected an LED backward at least once. The long leg is the anode, and in this project, it should connect to the positive side of the battery.

BUILDING CIRCUITS ON A BREADBOARD

Up to now, you've connected circuits with tape or by twisting component legs together, but this isn't very practical when a circuit has more than a few components. Fortunately, a *breadboard* can make connecting components easier. Breadboards have holes that you can stick component leads into to create circuits. When you're done, you can just unplug all the components and reuse them in different projects!

How to Connect Components and Wires

Inside a breadboard, metal plates connect the holes you see on the outside in a certain pattern. Let's look at a breadboard with four connection areas—two supply areas and two component areas.

In the supply areas on both sides, all the holes in each *column* are connected. You'll typically plug the positive side of your circuit's *power supply*—like the batteries you've used so far—into the red columns, and you'll typically plug the negative side of the power supply into the blue columns. Throughout this book, I'll refer to the supply column marked with a red line as the *positive supply column*, and I'll refer to the supply column marked with a blue line as the *negative supply column*.

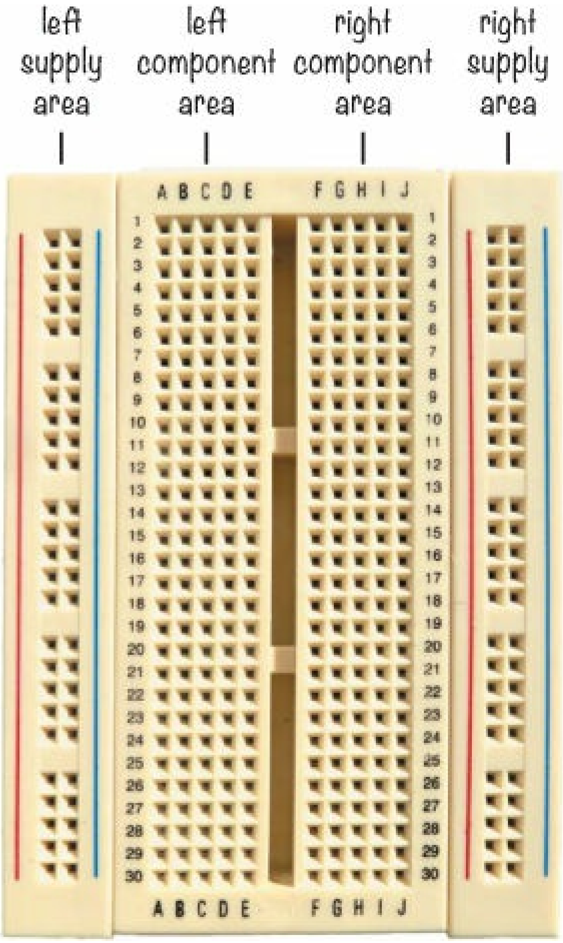

In the component areas, all the holes in each *row* are connected, and the columns are not connected. The left and right component areas are separated so that there's no connection between them. For example, holes A, B, C, D, and E in row 1 are connected, and holes F, G, H, I, and J in row 1 are connected, but holes E and F in row 1 are not connected.

To plug a component into a breadboard, simply push it into the hole where you want the connection. For example, if you wanted to connect one side of a resistor to the positive side of an LED, you'd just insert both the leg from the resistor and the leg from the LED into two holes on the same row in the left or right component area. If you have two component legs or wires that shouldn't connect, just make sure they are either on different rows in the component area or on opposite sides of the component area.

Wires to Use on a Breadboard

Eventually, you want to connect one row on your breadboard with a different row. You can use a wire to make that connection, but not all wires work well on a breadboard. The wire has to be stiff enough that you can push it into the hole without it bending, and it has to be thick enough to fit all the way inside the breadboard hole without falling out. *Single-strand* wires are the best wires for building circuits on a breadboard because they have one solid core inside, instead of many tiny wires wrapped together. The thickness of wire you need depends on your breadboard, but wires with 0.016- to 0.028-inch diameters should work. Wire thickness is often given in *American wire gauge (AWG)*, and I recommend using wire that is 21 to 26 AWG. You can buy wires that are cut and stripped for simple use with breadboards, or you can cut and strip your own wires using a wire cutter.

Another option is to use breadboard *jumper wires*. These wires have stiff ends that are very easy to connect to a breadboard. If you plan to connect a lot of circuits on a breadboard (you should!), keep a bunch of breadboard jumper wires on hand to make your life easier.

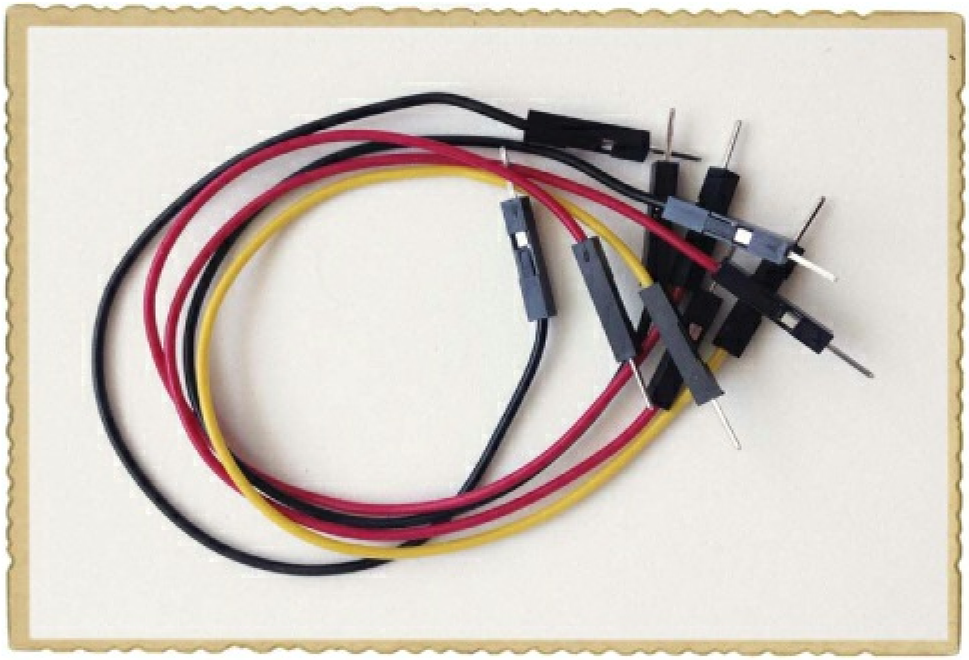

PROJECT #9: YOUR FIRST BREADBOARD CIRCUIT

Let's connect a simple circuit on a breadboard! Just as in "Project #8: Powering an LED" on page 78, this circuit lights up an LED, but this time we'll build the circuit on a breadboard. In this project, we're not going to use the supply rails on the side because the circuit is so simple that it makes more sense just to connect it all on the component area.

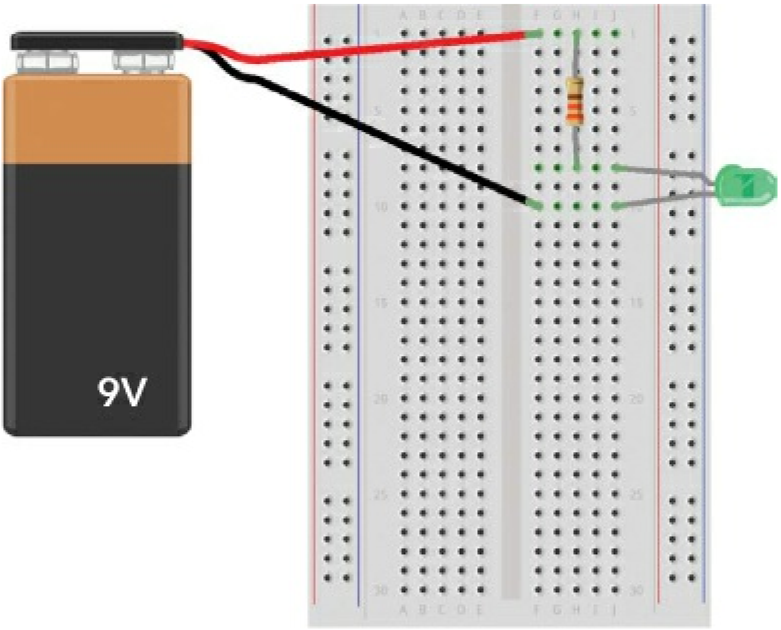

Shopping List

- **A breadboard** (Jameco #20601, Bitsbox #CN329) with at least 30 rows.

- **A standard 9 V battery** to power the circuit.
- **A 9 V battery clip** (Jameco #11280, Bitsbox #BAT033) to connect the battery to the circuit.
- **A standard LED** (Jameco #34761, Bitsbox #OP003).
- **A 330 Ω resistor** (Jameco #661386, Bitsbox #CR25330R for just this value or Jameco #2217511, Bitsbox #K017 for a variety pack) for limiting the current to the LED.

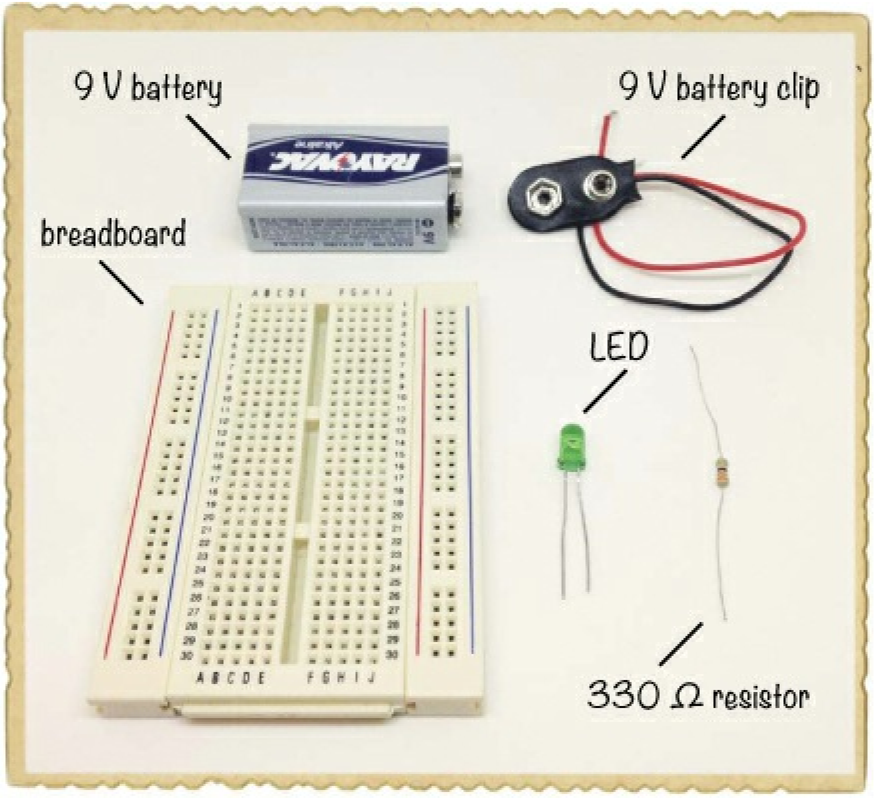

Step 1: Place the Resistor

First, place one leg of the resistor in row 1 and the other in row 8.

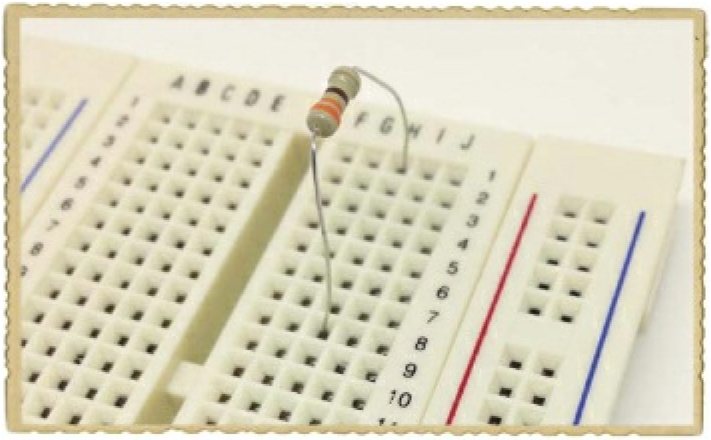

Step 2: Place the LED

Remember, LEDs are polarized, and they must be connected the right way to work. Connect the long leg of the LED to row 8, where the resistor leg is connected. Because the resistor and LED legs are on the same row, they're now connected. Connect the other leg of the LED to row 10.

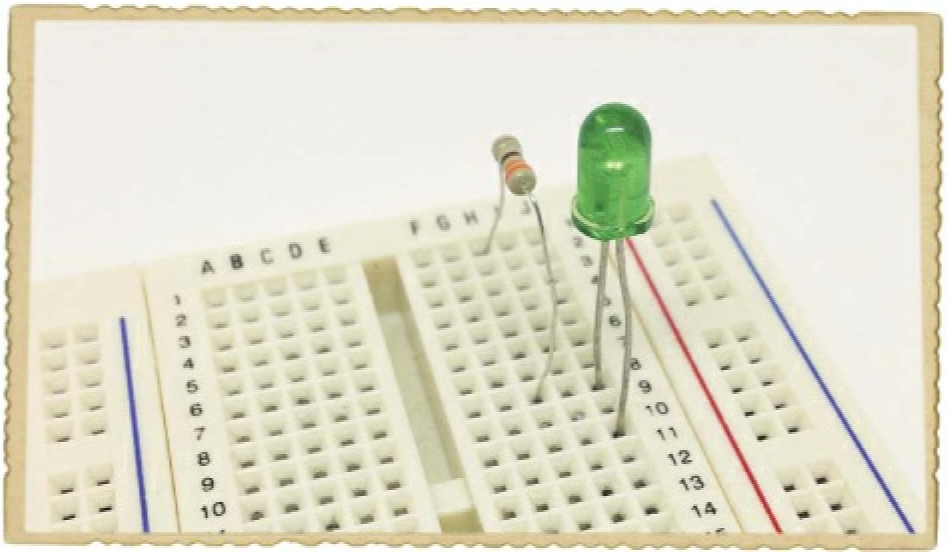

Step 3: Place the Battery Clip

Now, connect the battery to the LED and resistor. Connect the battery clip with the red wire at row 1 and the black wire at row 10. Plug your battery into the clip, and your LED should light up!

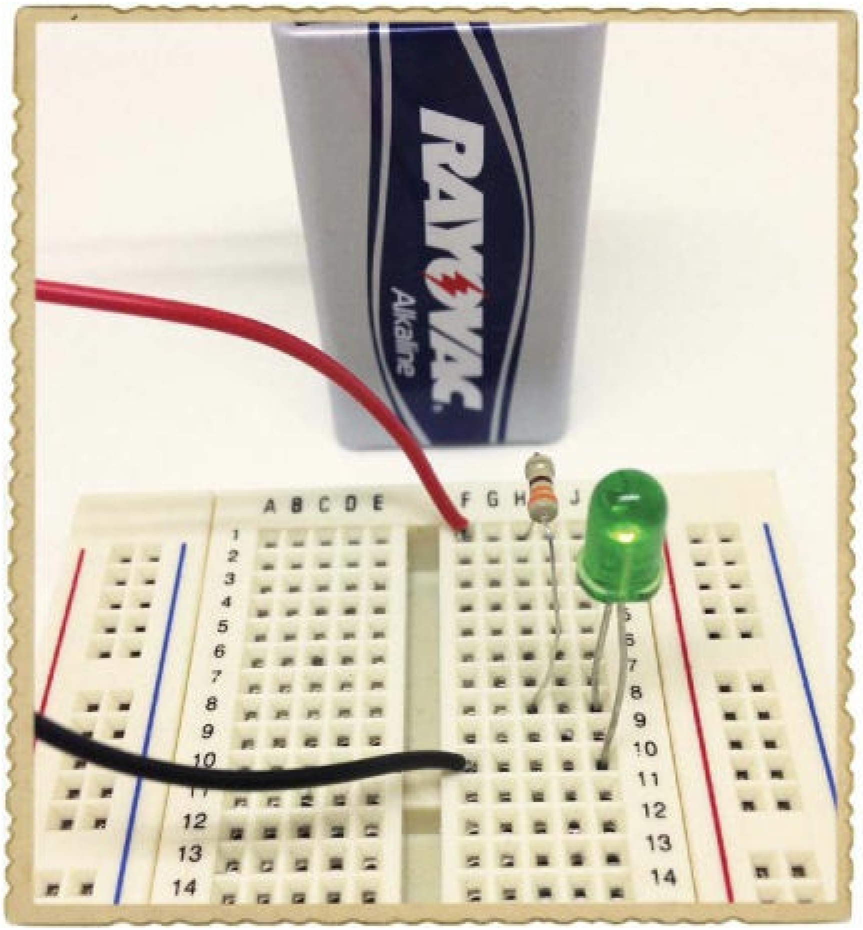

Step 4: What If the LED Doesn't Work?

If your LED doesn't glow, first disconnect the battery. You should always disconnect the battery when making changes to your circuit. Then, check whether the short leg of the LED is connected to the negative side of the battery.

If your LED is still not working once it's oriented correctly, check that your components are connected exactly as described in Steps 1 through 3. Are the long leg of the LED and one of your resistor's legs in row 8? Is the positive battery lead in the same row as the other resistor leg? Is the negative battery lead in row 10 with the short LED leg? Ask someone else to have a look at your circuit, too; maybe they can help you find the problem.

WHAT'S NEXT?

In this chapter, you've learned about two very common components: the resistor and the LED. You also learned how to use Ohm's law to calculate resistance, current, and voltage values. This knowledge will form the foundation for many aspects of electronics that you'll explore throughout this book.

You also learned to use a breadboard, which is a useful skill. To practice placing circuits on a breadboard, try building one of the projects you did earlier in this book without tape! How would you re-create "Project #2: Intruder Alarm" on page 11 on a breadboard?

In the next chapter, you'll learn about two more components: the capacitor and the relay. Then, I'll show you how to build one of my favorite circuits—a circuit that blinks a light!

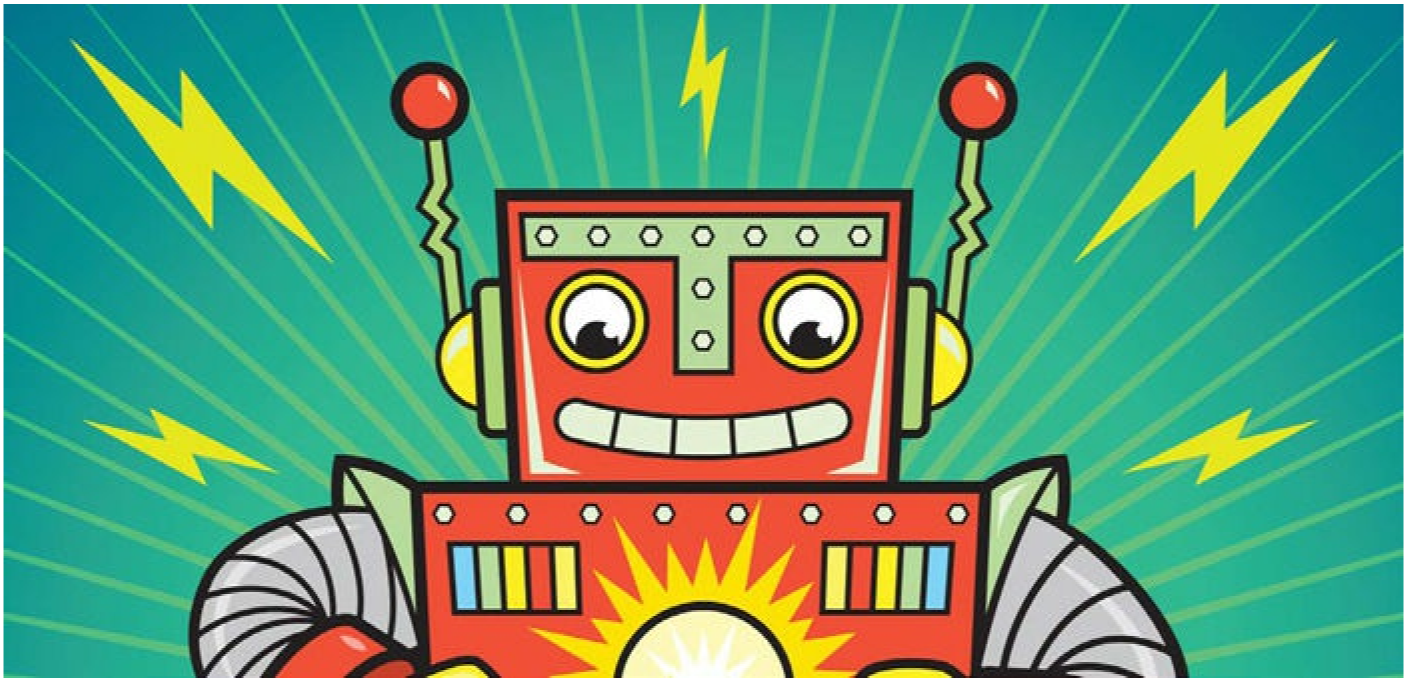

5

BLINKING A LIGHT FOR THE FIRST TIME

When I was a kid, one of my very first electronics projects was making a light blink. It was amazing to see the circuit work for the first time, and now I want to share that experience with you. In this chapter, you'll learn how the capacitor and the relay work. These are two common and very interesting electronics components, and I'll show you how to have some fun with them. At the end, you'll build your very own blinking light!

MEET THE CAPACITOR

The *capacitor* is like a rechargeable battery; you can charge a capacitor and use its energy to power something. But a battery can hold much more energy than a capacitor. A battery can power an LED for days without recharging, while most capacitors can power one for only a few seconds at most.

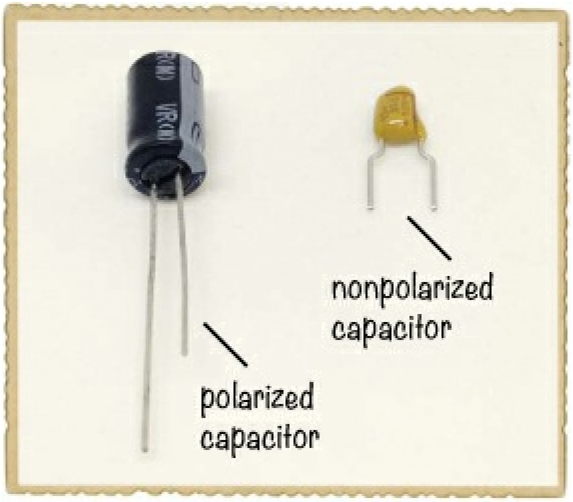

Capacitors are often used to introduce time delays in a circuit. For example, because a capacitor stores energy, it could be used to keep an LED on for a bit more time, even after the power is shut off. This little trick can be used with other components, too, to create interesting results.

How Capacitors Work

On the inside, capacitors are very simple devices. They're made of two metal plates placed very close together, with a material such as paper in between. To save space, the metal plates and the material between them are folded or rolled into a compact package.

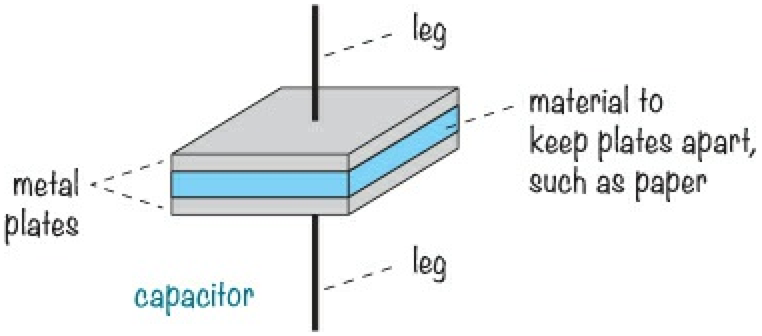

When you connect a battery to the two sides of a capacitor, a current flows as the battery tries to push electrons through the capacitor. But electrons can't flow across the gap between the plates, so instead, electrons build up on one plate and leave the other plate. Eventually, one plate can't hold anymore electrons, and the current stops flowing; at that point, we say that the capacitor is *fully charged*.

But just like the electrons in a battery, the electrons in the capacitor don't like being crowded together on the one plate. They want to go over to the side with fewer electrons. This means you have stored potential energy in your capacitor. If you disconnect the battery and connect, for example, a resistor between both sides of the capacitor, the stored electrons on one plate will start flowing the other way around, through the resistor, in order to reach the plate with few electrons.

Polarized and Nonpolarized Capacitors

Capacitors can be either *polarized* or *nonpolarized*. Like an LED, a polarized capacitor has a positive and a negative leg, and its positive leg must always face the positive terminal of the battery. The black capacitor shown in the photo is polarized, and its negative leg is marked with a stripe and minus signs down the side. The yellow capacitor isn't polarized, so it doesn't matter which leg goes where.

WARNING

Take care when you use polarized capacitors for the projects in this book and in your own projects. You must connect them the correct way to prevent them from being damaged.

In all circuits that call for a capacitor, you can use a nonpolarized capacitor as long as you can find one with the right *capacitance*. Capacitance is measured in farads (F), and the more capacitance a capacitor has, the more energy it can store. Nonpolarized capacitors with large capacitances aren't made because they'd have to be physically very large. Polarized capacitors are able to hold more energy in a smaller space, but they have the disadvantage of needing to be connected the correct way around.

When building circuits with high capacitance values, you'll be using polarized capacitors. Always make sure the positive leg of any polarized capacitor is connected closest to the positive side of the battery.

Capacitor Values

The capacitors you’ll use in this book will have capacitances in the µF (microfarad), nF (nanofarad), or pF (picofarad) range. Capacitances tend to be very small, and they’re often written with the *micro*, *nano*, and *pico* prefixes, which are defined as follows:

- µ (micro) means *millionth*, so 1,000,000 µF = 1F
- n (nano) means *billionth*, so 1,000,000,000 nF = 1F
- p (pico) means *trillionth*, so 1,000,000,000,000 pF = 1F

Polarized capacitors are big enough to have their values written on them. Nonpolarized capacitors, on the other hand, are a bit trickier. They’re usually really small, so they tend to have cryptic codes like *104* or *202*. I always forget what they mean, so when I need to figure out what a certain code means, I just look it up in a table. You can find a table of common codes in “Capacitor Codes” on page 283.

But like most things, capacitors are much more interesting to play with than to read about. Build the next project, and you’ll see for yourself how a capacitor works.

PROJECT #10: TEST A CAPACITOR

This project demonstrates that a capacitor stores energy. It’s nearly the same circuit you built in “Project #9: Your First Breadboard Circuit” on page 84, but this time, you’re going to add a capacitor. When you remove the battery from the circuit, as shown here, you’ll see that the LED remains lit for a second or two. That’s because the capacitor is powering the LED using its stored energy.

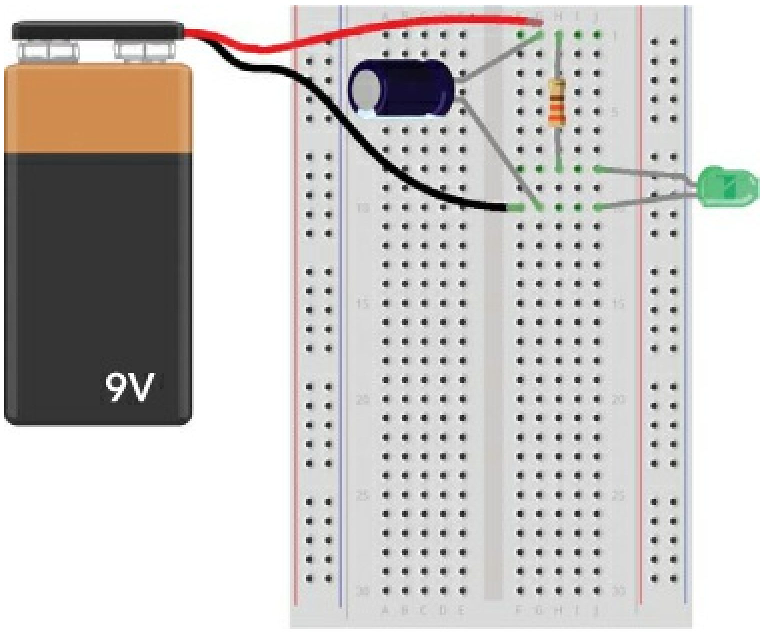

Shopping List

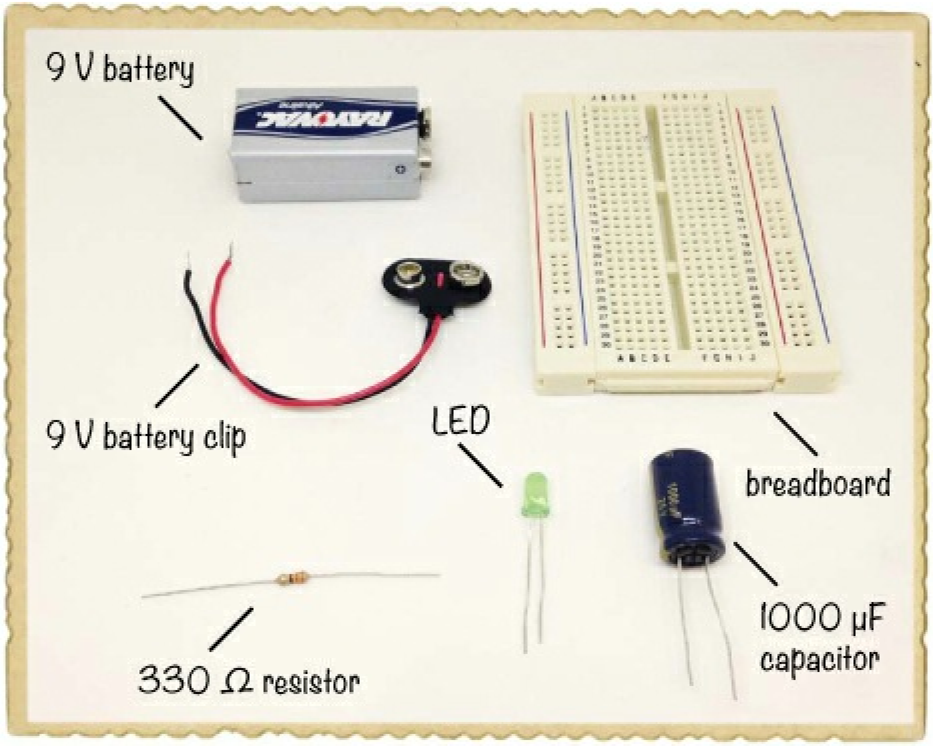

- **A breadboard** (Jameco #20601, Bitsbox #CN329) with at least 30 rows.
- **A standard 9 V battery** to power the circuit.
- **A 9 V battery clip** (Jameco #11280, Bitsbox #BAT033) to connect the battery to the circuit.
- **A standard LED** (Jameco #34761, Bitsbox #OP003)
- **A 330 Ω resistor** (Jameco #661386, Bitsbox #CR25330R) for limiting the current to the LED.
- **A polarized 1000 µF capacitor** (Jameco #158298, Bitsbox #EC1KU25)

Step 1: Start with the LED Circuit

Follow the instructions from "Project #9: Your First Breadboard Circuit" on page 84, making sure you end up with a working circuit that lights an LED. Then, disconnect the battery and move on to the next step.

Step 2: Add the Capacitor

Connect your capacitor to the battery. Because the capacitor is polarized, place the pin marked with a minus sign or a zero in the same breadboard row as the battery's negative leg. Connect the other leg to the same row as the battery's positive leg, as shown.

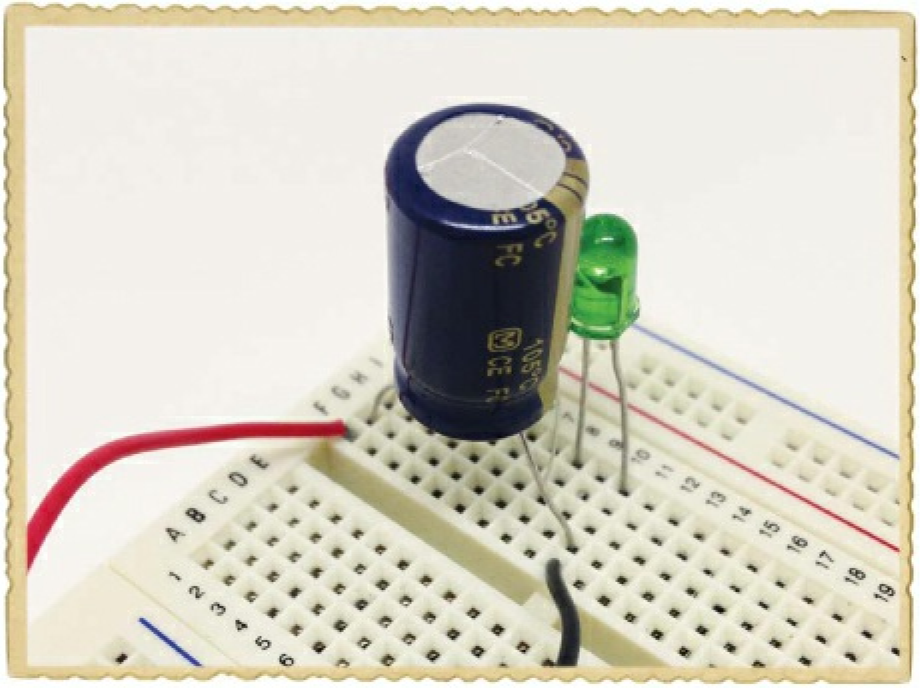

Step 3: Charge the Capacitor

Connect the battery to the clip, and the LED should light up. At the same time, the battery should have very quickly charged the capacitor.

Step 4: Use the Capacitor to Light the LED

Watch the LED while removing the battery. The LED shouldn't turn off right away when you disconnect the battery. Instead, it should stay lit for a second or so and then fade out slowly until there's no more energy left in the capacitor.

Step 5: What If the Circuit Doesn't Work?

First, check whether the circuit works without the capacitor. If not, then go back to Step 1 and get the LED circuit working before you move forward.

If the LED lights up with the battery connected but turns off the instant you remove the battery, then something is wrong with the capacitor part of your circuit. Check that the capacitor's positive leg is connected to the positive battery leg (row 1 in the photo) and that the other leg is connected to the negative battery leg (row 10 in the photo).

If the circuit looks correct, confirm that the capacitor value is at least 1000 µF; the value should be written on the capacitor. If it's less than 1000 µF, try a bigger capacitor.

DESCRIBING CIRCUITS WITH SYMBOLS

So far, you've built circuits with only a few components. To build more interesting electronics projects, you usually need more components. But drawing out every component in a big circuit exactly as it looks in real life can be messy and time-consuming. This is where *circuit diagrams*, also called *schematics*, come to the rescue.

In a circuit diagram, each component has its own simple symbol, which lets you draw the whole circuit quickly. As with words and books, until you know what the different symbols mean, a circuit diagram might look a bit complex. Let's jump right in and learn a few symbols! Here's a circuit with an LED, a resistor, and a battery, alongside its circuit diagram:

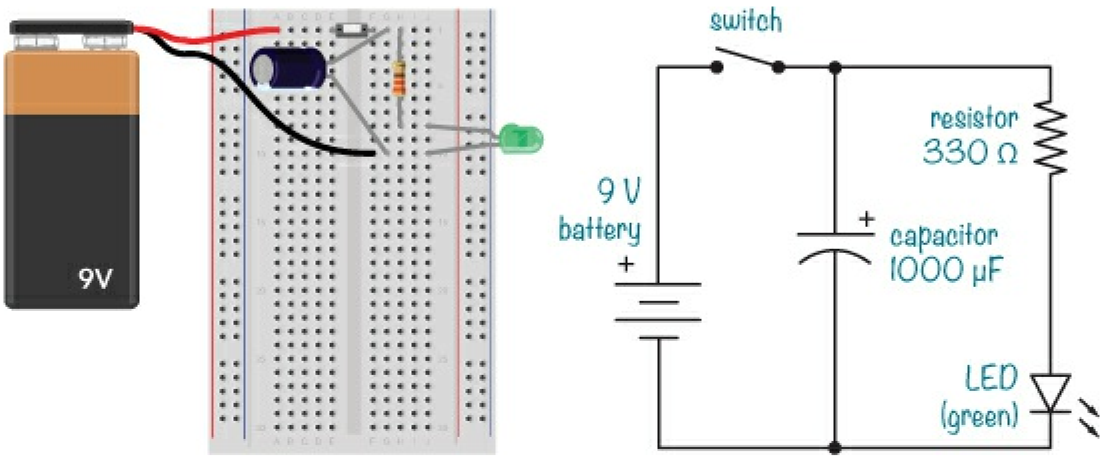

The battery symbol has a + sign to tell you where the positive leg goes. Sometimes you'll see the battery symbol without the plus sign; in that case, the positive side is the one with the longest line.

NOTE

The battery symbol may be drawn with two lines, with four lines, and sometimes with even more lines. Regardless of the number of lines, just look for the specified voltage and use a battery with the same voltage.

The switch symbol is very simple, and it doesn't matter which way you draw it. In the LED symbol, the line on the point of the triangle indicates the negative side of the LED, or the cathode. Check that against the "real-life" version, and you can see that the LED's positive leg is indeed connected to the resistor in both diagrams. The resistor, on the other hand, isn't polarized, so its symbol doesn't have any directional markers. There are both polarized and nonpolarized capacitor symbols, but our example shows the polarized version, which has the positive side marked with a + sign. Look at both the circuit and the circuit diagram again—do you agree that they are the same?

Once you learn how to build something just by looking at a circuit diagram, a whole new world will open up for you. You can find circuit diagrams for almost anything on the Internet these days, like radios, MP3 players, walkie-talkies, or whatever you want to build! I'll teach you more circuit symbols throughout this book as we use more components.

MEET THE RELAY

I was a very curious child, and I always wondered how things worked. To me, electronics devices like the radio or the television were just magical. I had no idea how they worked, and I didn't believe I'd ever understand how to make one.

But one day, I asked my dad how it was possible to blink a light automatically. I thought that if I could only understand that, I'd be able to understand more. Fortunately, my dad was an engineer, and he was also good at explaining things in a practical way. When I asked him how to blink a light, he introduced me to a *relay*, like the one shown here.

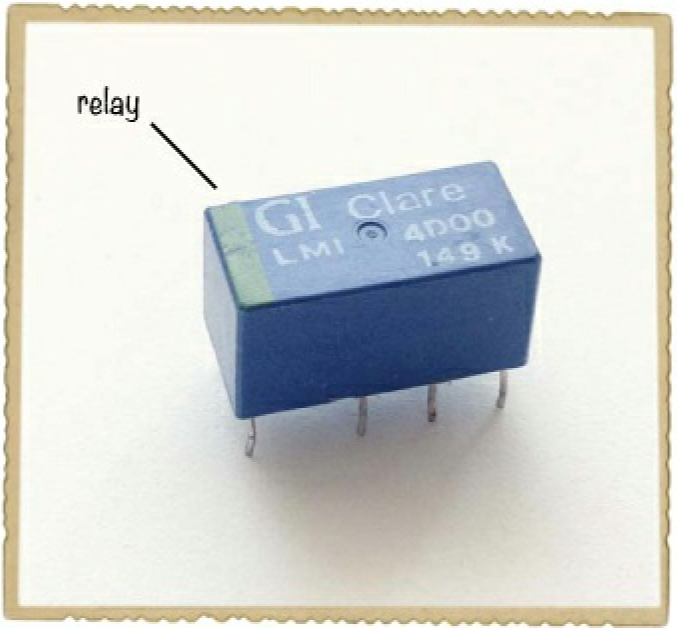
relay
GI Clare
LMI 4D00
149 K

Chapter 1 described how you can use switches to turn things on and off. Chapter 2 showed you how to use an electromagnet to move things. Imagine combining the electromagnet with a switch: instead of pushing a button to change the switch's position, you add an electromagnet that can change the switch position for you. That's the idea behind the relay, and this illustration shows how it works:

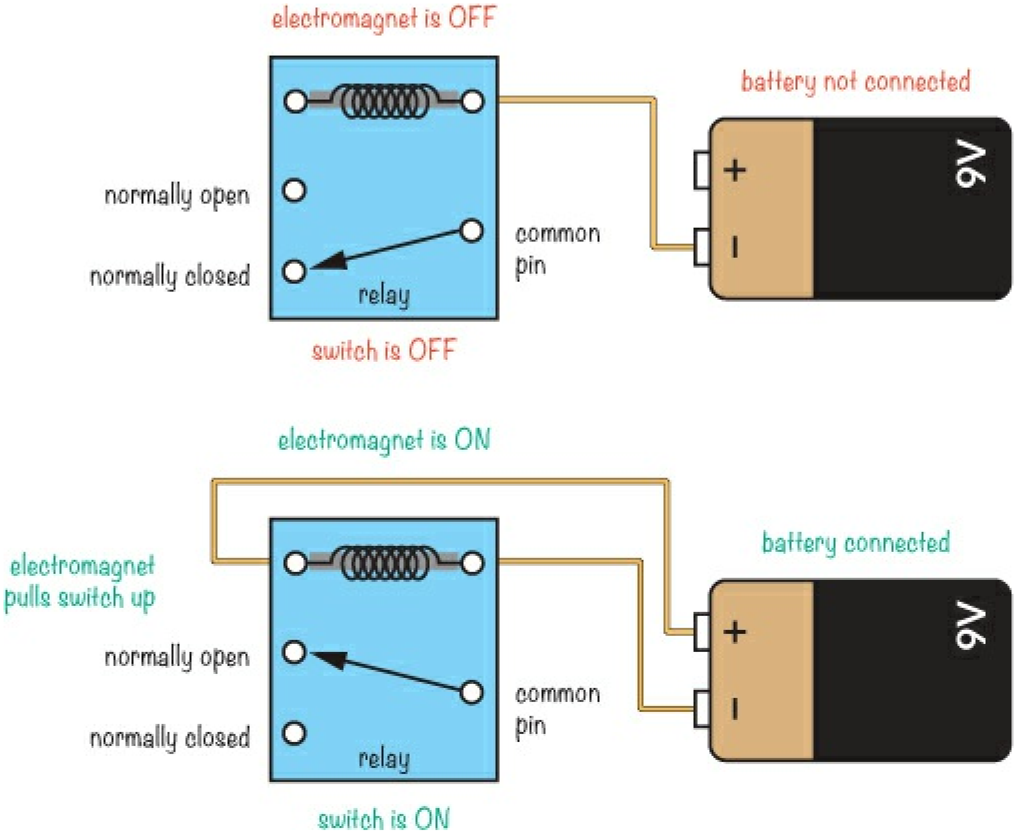
electromagnet is OFF
battery not connected
normally open
normally closed
common pin
relay
switch is OFF
+
-
9V
electromagnet is ON
battery connected
electromagnet pulls switch up
normally open
normally closed
common pin
relay
switch is ON
+
-
9V

The white dots represent the relay's pins. Notice that the connections are labeled *common pin*, *normally closed*, and *normally open*. These labels are defined as follows:

Common pin (COM) Connected to either NC or NO

Normally closed (NC) Connected to COM when coil is off

Normally open (NO) Connected to COM when coil is on

When the battery isn't connected to the relay coil, the electromagnet doesn't pull, and COM from the switch is connected to NC. But when you connect the battery to the relay coil, the electromagnet turns on, and the switch is pulled so that COM gets connected to NO instead. You can connect or disconnect a battery from the electromagnet to change the position of the switch!

Using the Relay to Blink a Light

If you were to connect a relay to a battery so that the electromagnet connects to the battery through the relay's COM and NC switch contacts, the electromagnet would turn on and off continuously. Here's an example relay circuit that's also connected to a light bulb:

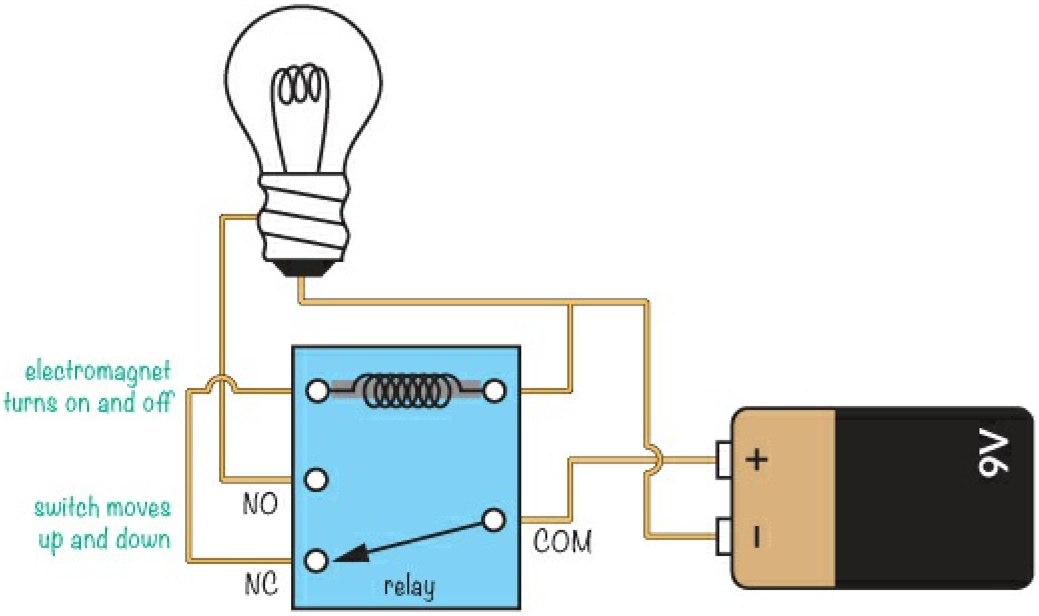

Before you connect the battery in this circuit, the electromagnet would be off, leaving the COM and NC switch contacts connected. With the battery connected to the circuit, the electromagnet would receive power from the battery through the switch. This means it'd pull the switch and connect COM and NO instead, giving the light bulb power from the battery.

But with the switch in this position, the battery would no longer be connected to the coil, and the electromagnet would lose its power. When the electromagnet has no power, the switch falls back to its original position, disconnecting the battery from the light bulb. The battery would power the electromagnet again, and the process I just described would repeat.

In this example, it seems like you'd get a blinking light, right? In theory, yes. But the relay would switch on and off so fast you wouldn't see the light turn on and off properly! Instead, you'd hear a really fast ticking sound as the relay switched back and forth, but the light would appear to stay dark.

Slowing Down the Blinking

To build a circuit that lets you actually see the light blink, you need to slow down the relay. The capacitor can help with this. In "Project #10: Test a Capacitor" on page 92, adding a capacitor to the LED circuit made the LED stay on for a short time after the battery was disconnected. If you connected a capacitor across the electromagnet in our too-fast blinking-light circuit, the electromagnet would stay on for a bit, too.

But the electromagnet wouldn't stay off for long, so in this case, the light would appear to be on all the time. To make it stay off longer, you'd need to slow down the charging of the capacitor so it wouldn't jump back to fully charged as soon as it discharged. To do this, you could reduce the amount of current flowing into the capacitor. And how do you reduce the amount of current? With a resistor! To blink an LED light with a relay, you'd use a circuit like this one, which we're going to build next:

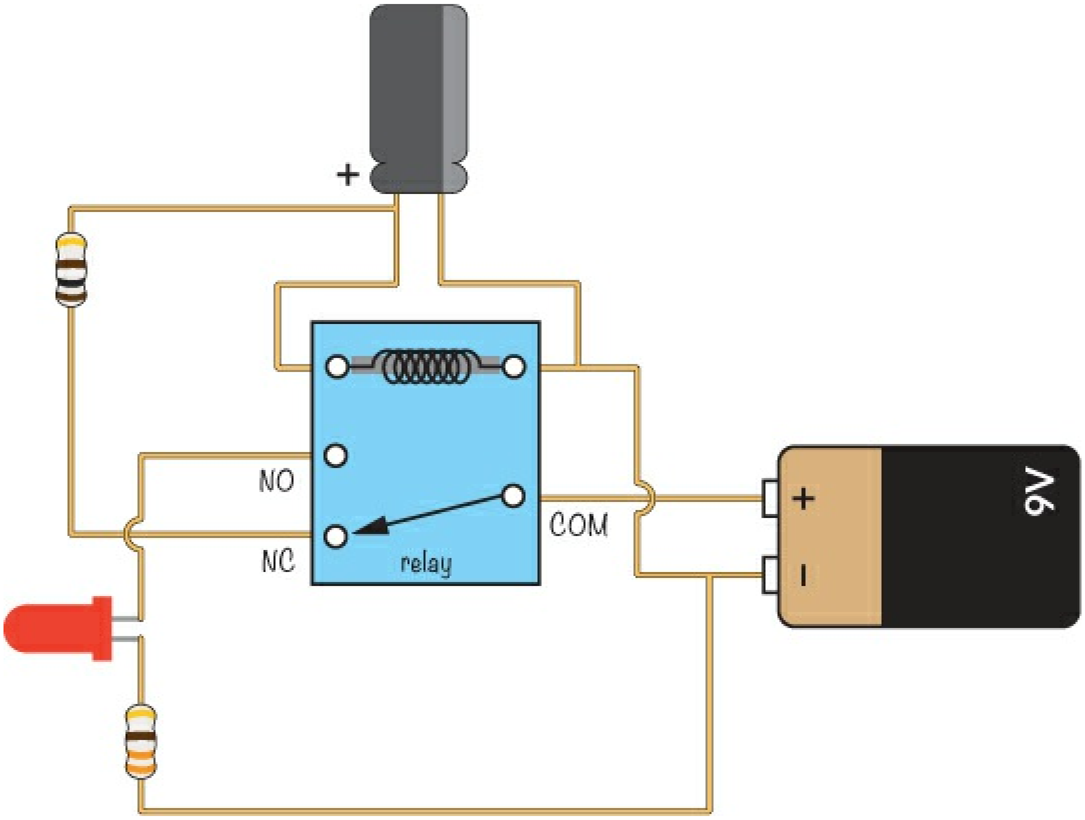

PROJECT #11: BLINK A LIGHT!

It's time for you to build your first blinking light by flashing an LED. Here's the complete circuit diagram—do you recognize the components?

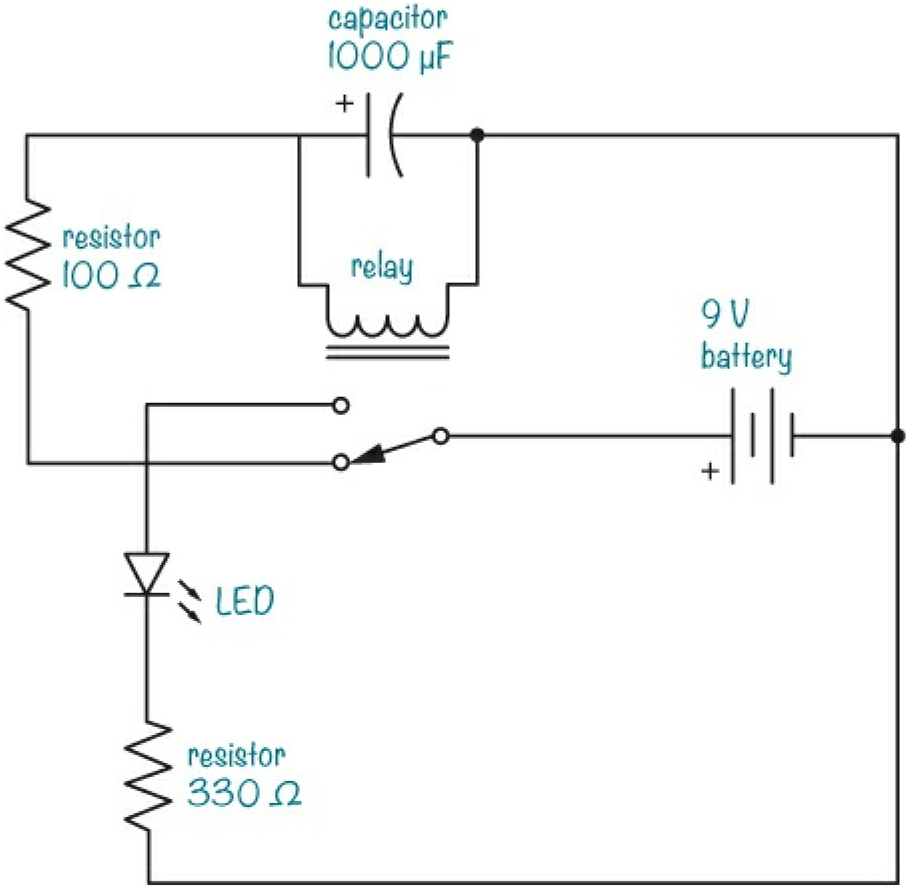
capacitor
1000 µF
+
resistor
100 Ω
relay
9 V
battery
+
LED
resistor
330 Ω

When connecting a circuit, it's helpful to connect it in a way that looks like the circuit diagram. That makes it easier to find out what's wrong if your circuit isn't working later—and you'll often find that your circuit doesn't work on the first try. Tracking down errors and figuring out how to fix them is part of the game!

Shopping List

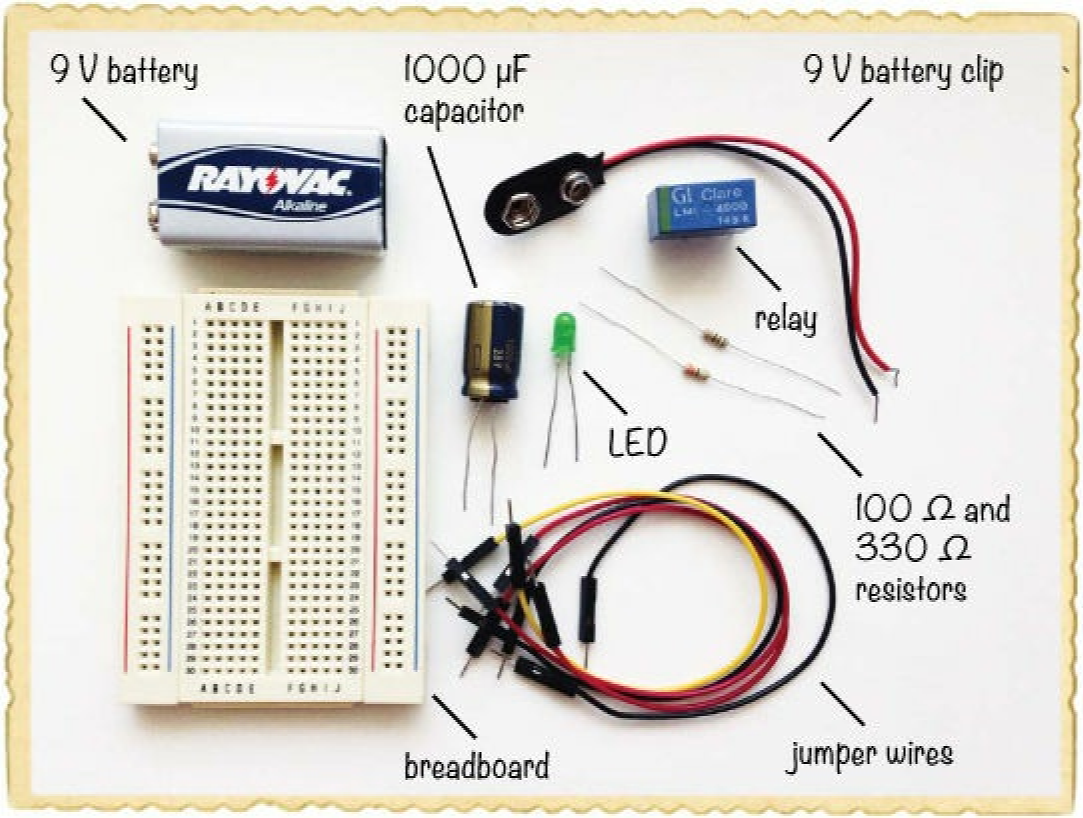
9 V battery
1000 µF
capacitor
9 V battery clip
relay
LED
100 Ω and
330 Ω
resistors
breadboard
jumper wires

- **A breadboard** (Jameco #20601, Bitsbox #CN329) with at least 30 rows.
- **Breadboard jumper wires** (Jameco #2237044, Bitsbox #CN236) for making easy connections. (Standard hookup wire works, too.)
- **A standard 9 V battery** to power the circuit.
- **A 9 V battery clip** (Jameco #11280, Bitsbox #BAT033) to connect the battery to the circuit.
- **A DPDT or SPDT relay** (Jameco #842996, Bitsbox #SW073) with a 5 V, 6 V, or 9 V coil.
- **A standard LED** (Jameco #34761, Bitsbox #OP003)
- **A polarized 1000 µF capacitor** (Jameco #158298, Bitsbox #EC1KU25)
- **A 100 Ω resistor** (Jameco #690620, Bitsbox #CR25100R) for limiting the current to the capacitor.
- **A 330 Ω resistor** (Jameco #661386, Bitsbox #CR25330R) for limiting the current to the LED.

Step 1: Identify the Relay Pins

Not knowing which pin on the relay is which is the biggest source of error when building this circuit, so let's look at these pins now. To find the function of each pin, look up the *datasheet* for your relay. A datasheet is a document that tells you how an electrical component works. For a relay, it should say what voltage you need for the electromagnet coil, how much current you can run through the contacts, and so on. You should be able to find a link to the datasheet on the product page where you bought your relay.

For the relay I recommend, the pins are placed as shown here:

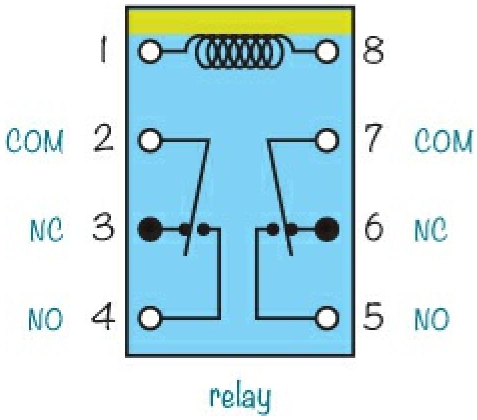

This type of diagram is called a *pinout*, and it shows the function of each pin. The view is from above with the pins out of sight below the plastic case of the relay. I suggest drawing the pinout on a piece of paper and keeping it on your desk while you connect the circuit so you can check it repeatedly and make sure you're connecting it the right way.

In this relay, pins 1 and 8 are for the electromagnet coil—you can identify these pins by the line at the top. When you look at the real relay with the line at the top, the pins should be in the same places as in the pinout.

The pinout also shows that there are two switches inside. Pins 2 to 4 make up one switch, and pins 5 to 7 make up the other. Don't worry about memorizing the relay pins, though. You can always refer to the pinout if you get stuck, and I encourage you to do so whenever you need to. Also, different relay types can have different pinouts.

Step 2: Make the Relay Switch Fast

First, let's connect the relay so that it switches on and off automatically. Place the relay in the middle of your breadboard, centered over the notch, with one side in each component area. This way, no pin should be connected to any other pins.

Plug your battery clip's positive leg into the positive supply column on the left, and plug the negative leg into the negative supply column on the same side. Connect a jumper wire from the negative supply column to the same row as the right coil pin (pin 8), shown as row 9 this diagram:

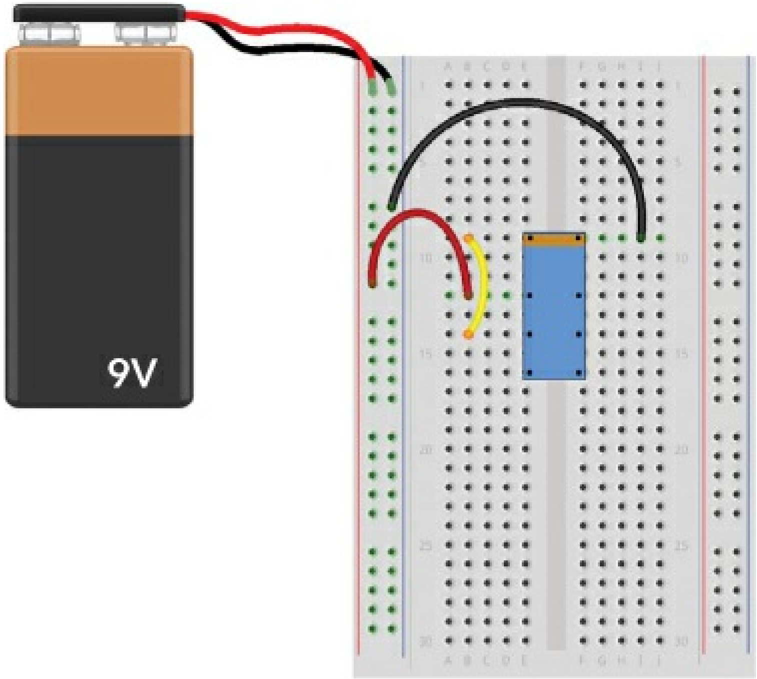

Next, connect a wire from the plus column on the left to the relay's common switch pin (pin 2). In this diagram, the common pin is in row 12, column B. Then, connect a wire from pin 3 on the relay (the NC pin) to the coil pin (pin 1); this is the yellow wire going from row 14 to row 9 in column B in the diagram. Connect the 9 V battery, and you should hear a really fast ticking. This is the relay turning on and off. Disconnect the battery for now.

Step 3: Make the Relay Stay On Longer

The next step is to slow down the relay by placing a capacitor across the coil of the electromagnet. Connect the capacitor as shown here:

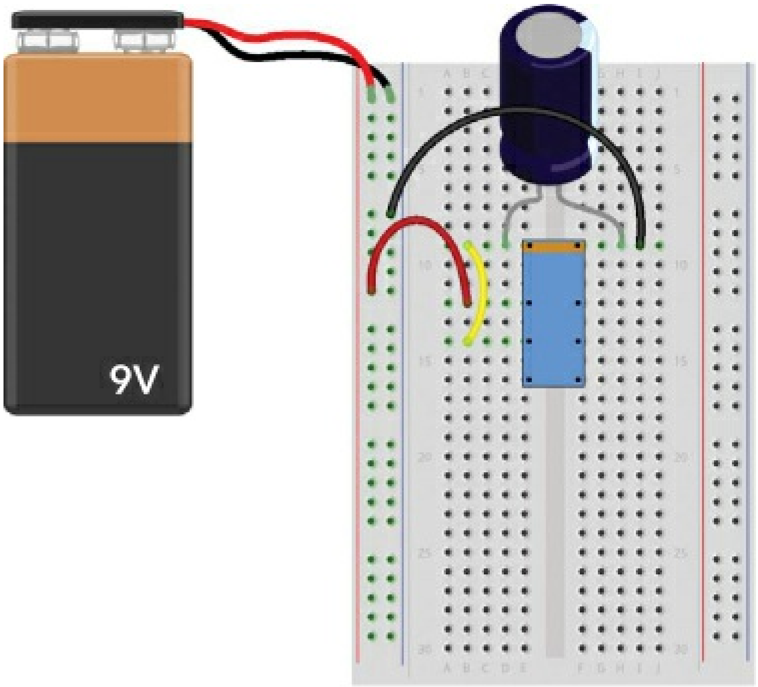

In the diagram, I’ve connected the negative capacitor leg to pin 8 on the relay by plugging it into row 9, column H and connecting the positive leg to pin 1 on the relay at row 9, column D. Whichever holes you use, just make sure you connect the capacitor’s negative leg to the same relay pin as the one you connected the negative supply to. The negative capacitor leg is usually marked with a stripe, a zero, or a minus sign.

Connect the battery to test your circuit. You’ll know it’s working if you hear the telltale ticking sound. The ticking should be much slower now, which means that the relay is staying on for a longer time. But in the instant the relay is turned off, the capacitor should charge again, leaving the relay off for only a fraction of a second. You’ll fix that in the next step, so disconnect the battery again now.

Step 4: Make the Relay Stay Off Longer

Let’s add a resistor before the capacitor to reduce the amount of current that flows and force the capacitor to take more time to charge. To do this, simply replace the wire between pins 1 and 3 on the relay (the yellow wire connected to rows 9 and 14 in my diagrams) with a resistor of about 100 Ω as shown here:

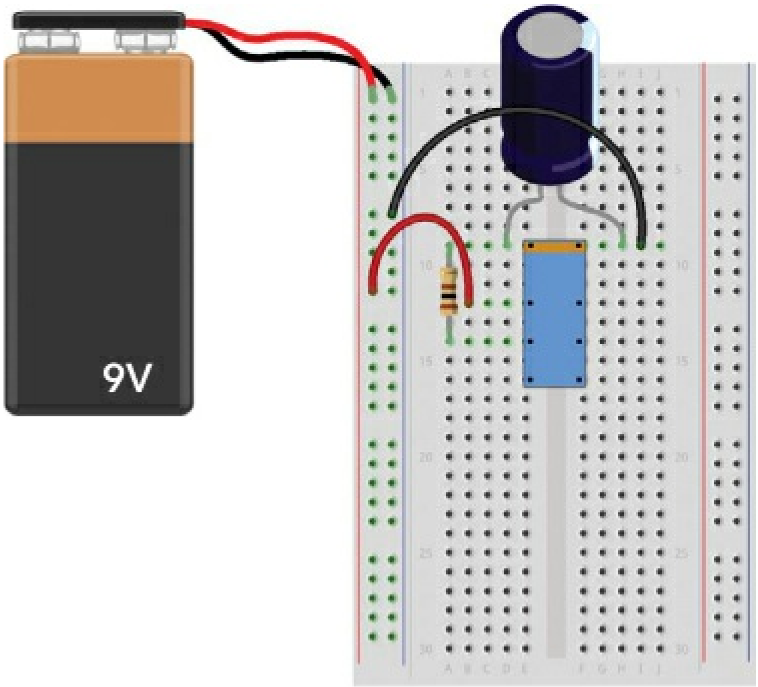

Connect the battery to test it. The relay should stay off a bit longer now, and you’ll recognize this by the “tick-tock” sound of the relay.

Step 5: Add the LED and Resistor

Disconnect the battery and add the LED and the 330 Ω resistor to the circuit. Connect the long leg of the LED to the NO pin of the relay (pin 4). My LED’s long leg is connected to row 16, because the NO pin is plugged into that row. Connect the other leg to an unconnected row below, such as row 19. Connect the resistor from the same row to the negative column on the supply side.

Your breadboard should look something like this:

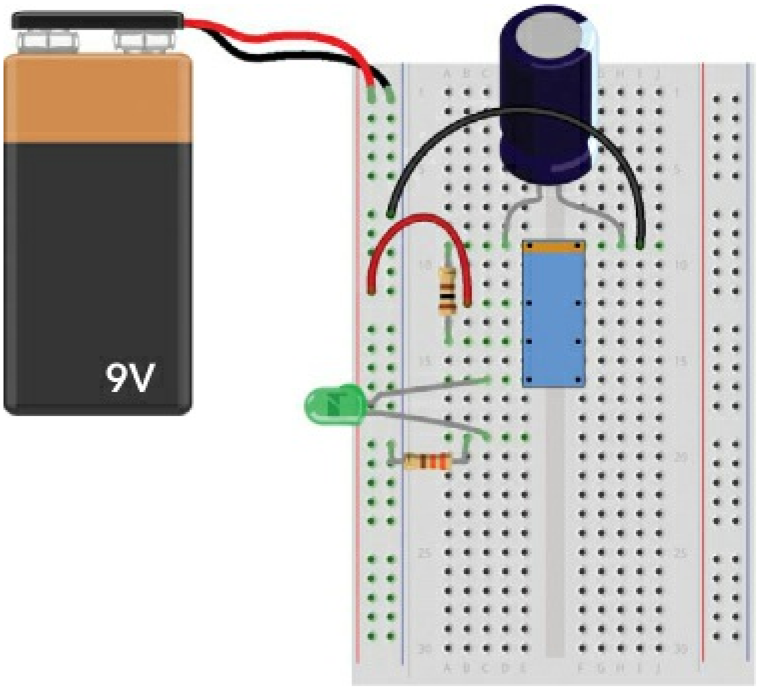

Compare this to the circuit diagram at the beginning of the project to see how the symbols match. Now, connect the battery and watch the light blink!

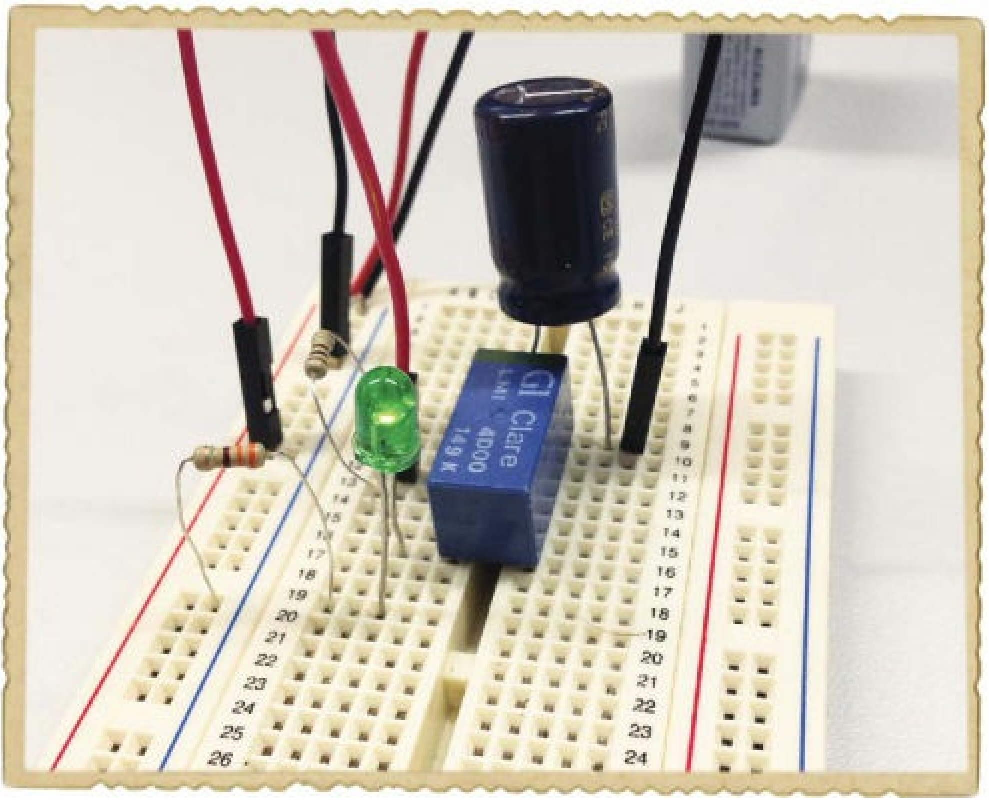

Step 6: What If the LED Won't Blink?

If you can't get the LED to blink, go back to Step 1 and check your work through to Step 4. This should give you a relay that turns on and off automatically. If that works, then you should be able to connect the LED and resistor to have a working circuit.

If the LED still doesn't blink, confirm you've connected it according to the diagrams. Still no luck? Remove the LED and the resistor from the circuit, and then connect only these two to the battery until you can make the LED light up. (Follow the instructions in "Project #9: Your First Breadboard Circuit" on page 84.) If you can't make the LED light, then your LED may be broken or your resistor may be the wrong value.

TRY IT OUT: MAKE YOUR INTRUDER ALARM MORE EFFECTIVE

Try connecting a relay to the Intruder Alarm you built in Project #2 on page 11 so that when someone tries to enter a room, they can't turn off the alarm by closing the door. When your relay is triggered by a switch, it should stay on. This diagram shows how you can connect the relay (shown in blue) to the rest of the circuit (shown in black). When the alarm is triggered, the beep will continue to sound until you disconnect the battery.

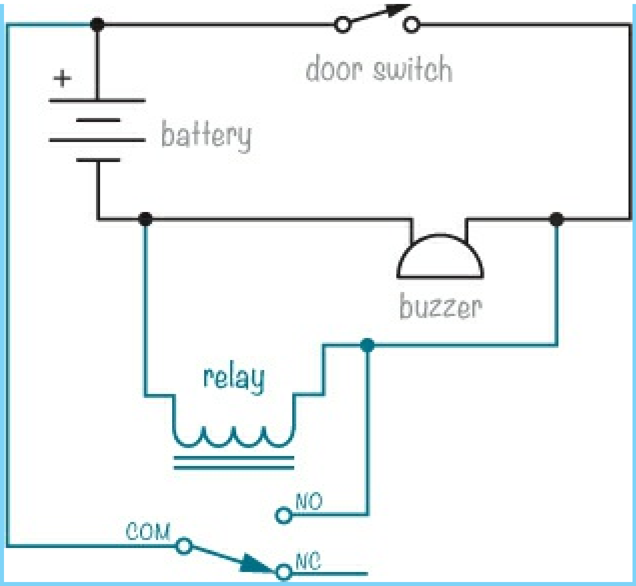

WHAT'S NEXT?

You've come a long way in this chapter! So far, you've learned how four common components in electronics work (resistors, LEDs, capacitors, and relays), you've connected a few circuits on a breadboard, and you've even made your own project that blinks a light. Hopefully, you understand why the light blinks, too.

You've already connected a few circuits on a breadboard, which is a useful skill when you want to test a circuit or build simple prototypes. But when you want a circuit's components to stick together forever, you need to solder them, and I'll show you how in the next chapter.

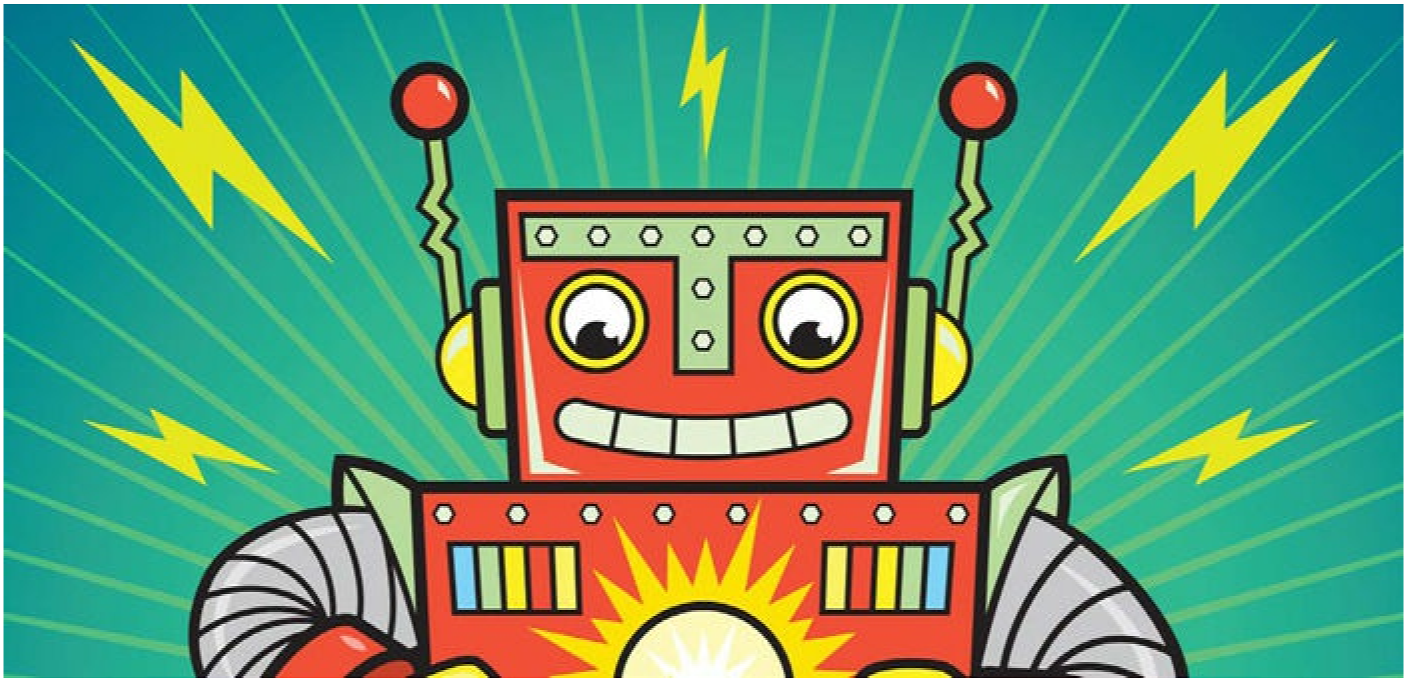

6

LET'S SOLDER!

In Chapter 5, you used a breadboard to build new circuits. A breadboard lets you quickly test new ideas and experiment with different components to see the result, but it isn't a permanent solution. Components can easily fall out, and there can be loose wires all over the breadboard. When you want to use a circuit for a long time without changing it, it's better to solder the circuit to a *circuit board*. Soldering is kind of like gluing: you melt a material called *solder* onto the component legs so the components stick to the circuit board. A circuit board has holes like a breadboard, and when you solder components to it, they become connected by plates of copper.

This chapter will teach you how to solder, starting with a simple LED circuit. This will give you the foundation you need to solder your own circuits later. After you've practiced the basics in this chapter, you'll continue soldering in Chapter 7 to build a touch-enabled switch and a circuit that wakes you up when the sun rises in the morning.

HOW TO SOLDER

To solder a circuit, you need a few tools:

- Solder
- A soldering iron
- A stand to hold the soldering iron
- A damp sponge
- Safety goggles

Solder is a mix of metals that can easily melt at high temperatures. The most commonly used solder melts at around 360 to 370 degrees Fahrenheit.

A *soldering iron* is a pen-shaped tool that heats up to a temperature high enough to melt solder. If you put solder on the tip of the soldering iron, the solder melts.

soldering iron

Soldering is quite easy once you learn the basics, and one important basic step is to take a few safety precautions.

Soldering Safety Tips

Soldering irons get extremely hot, and you can burn yourself if you touch the metal part of a soldering iron while it's plugged in. Always return the soldering iron to its stand when you're not using it; never lay it directly on a surface. Avoid touching *joints* (soldered connections) and components for a few seconds after you solder them, too. They can get hot enough to burn.

Here are some other important safety tips to keep in mind while soldering:

- Keep the hot parts of the soldering iron away from the power cord.
- If you're soldering on a table, protect the surface with a piece of wood or some thick cardboard.
- Always wash your hands after handling solder.

- If you do get burned, don’t panic. For a minor burn, immediately cool the burn with running cold water, keeping the burned area under water for at least five minutes. Putting ice on a burn is good, too, but make sure to do the initial cooling with water right away.

Ask an adult to supervise your first few attempts at soldering, and remember: soldering is fun, but soldering irons must still be used with great care. With those tips in mind, read the rest of this section for a step-by-step guide to basic soldering.

Heat the Soldering Iron

The first step of soldering is to plug in your soldering iron and place it on your stand. Don’t forget to put on your safety goggles!

After a minute or two, check whether the iron is hot enough by touching some solder to the tip of the iron. If the solder melts, your iron is ready to go.

Clean the Soldering Iron Tip

Touch the tip of your soldering iron to a damp sponge to clean it. A clean tip transfers heat much better than a dirty one, so clean the tip often.

Tin the Soldering Iron Tip

Here’s a little trick: to heat the joint faster, add a bit of solder to the tip of the soldering iron just before you start soldering. This is called *tinning*. This must be done just a second or two before soldering to be effective.

Heat Both the Pin and the Pad

Place the tip of the soldering iron onto both the component leg and the copper strip. Heat the strip and leg for a couple of seconds before you move on to the next step.

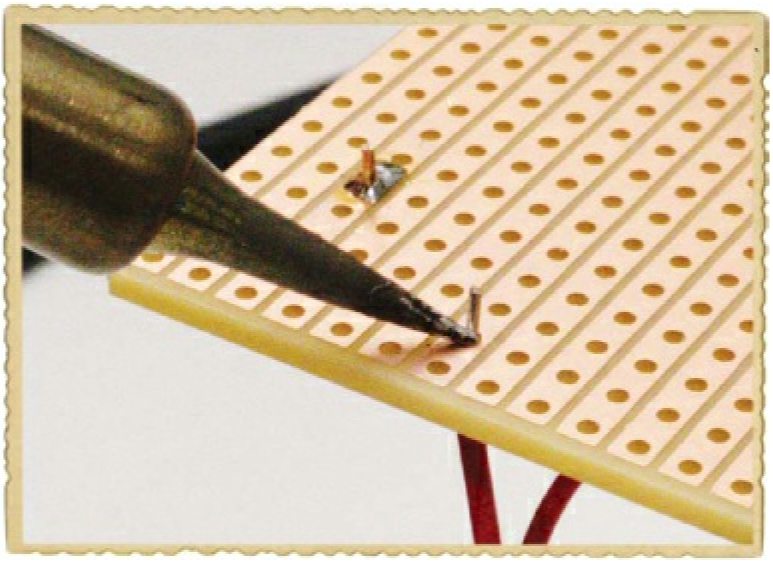

Add Solder

While keeping the soldering iron tip on the joint, touch some solder to the leg and the copper strip. As the solder melts, slowly add more until there’s just enough to coat both the leg you’re soldering and the copper strip you want to connect it to. When you have enough melted solder, remove the solder wire from the joint while still keeping the soldering iron tip on the joint.

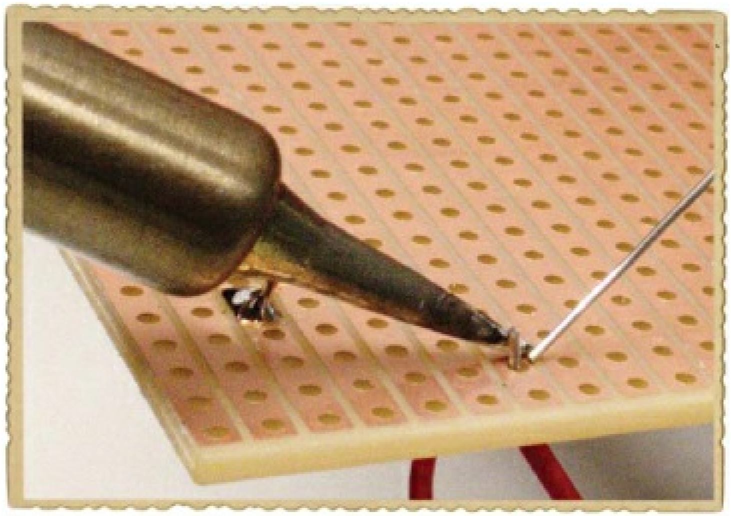

Remove the Soldering Iron

Finally, remove the soldering iron tip from the joint and place the iron in its stand. Always do this last. If you remove the iron while the solder wire is touching the solder joint, the solder wire may get stuck to the circuit board when the solder joint hardens.

Your solder joint should have a cone shape.

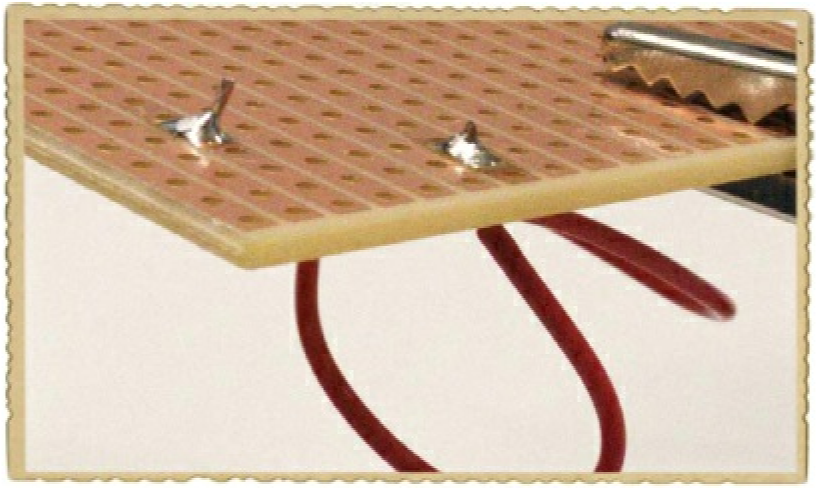

If you're done soldering, then unplug the iron so it can cool down.

Watch Out for Bad Solder Joints!

It's important to heat both the component leg and the copper strip with the iron before touching solder to either. When you heat only the component leg and not the copper strip, the solder sticks to the leg, but there's no connection between the solder and the copper strip. If only the strip is heated, the solder will stick to the strip, but not the leg. It could look like a good soldering joint from a distance, but the solder probably won't be connected to the leg.

You must also make sure the solder doesn't float over to another copper strip next to it. This will create an unintended connection between the two strips.

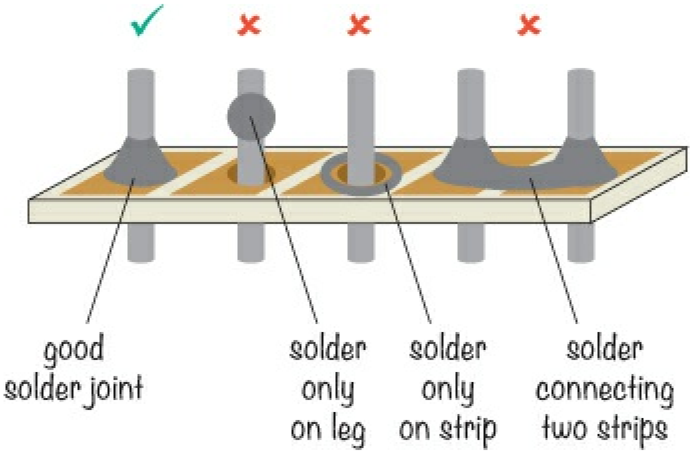
✓
×
×
×
good solder joint
solder only on leg
solder only on strip
solder connecting two strips

If your solder joint doesn't look right, don't worry. Just reheat the joint, the component leg, and the copper strip and add more solder to achieve the cone shape of a good solder joint. Then, you should be good to go.

PROJECT #12: SOLDER YOUR FIRST LED CIRCUIT

Now, let's get soldering! In this project, you'll solder the resistor and LED circuit from "Project #8: Powering an LED" on page 78 to a circuit board. The battery should make current flow in the circuit, the resistor should make sure there isn't too much current, and the LED should light up.

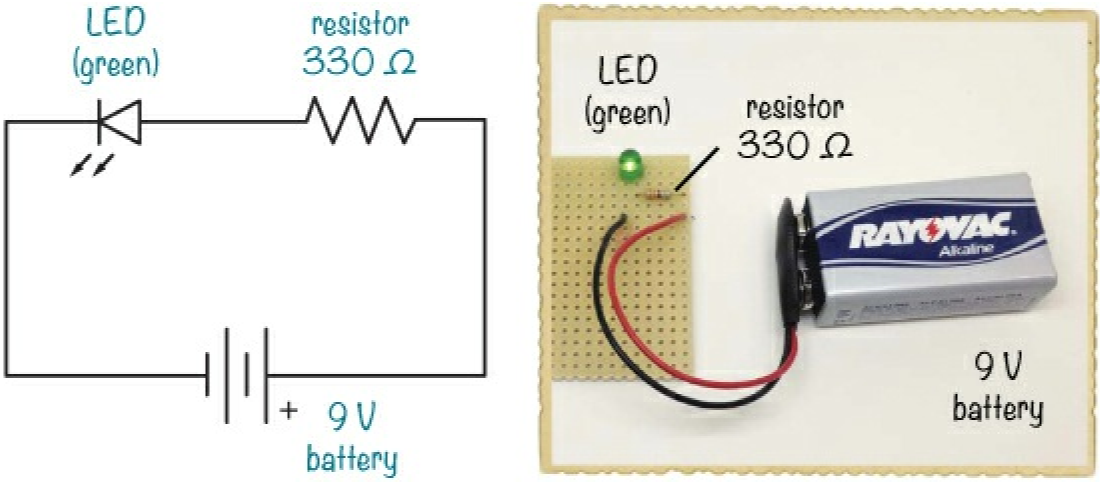
LED (green)
resistor 330 Ω
+ 9 V battery
LED (green)
resistor 330 Ω
RAYOVAC
Alkaline
9 V battery

Shopping List

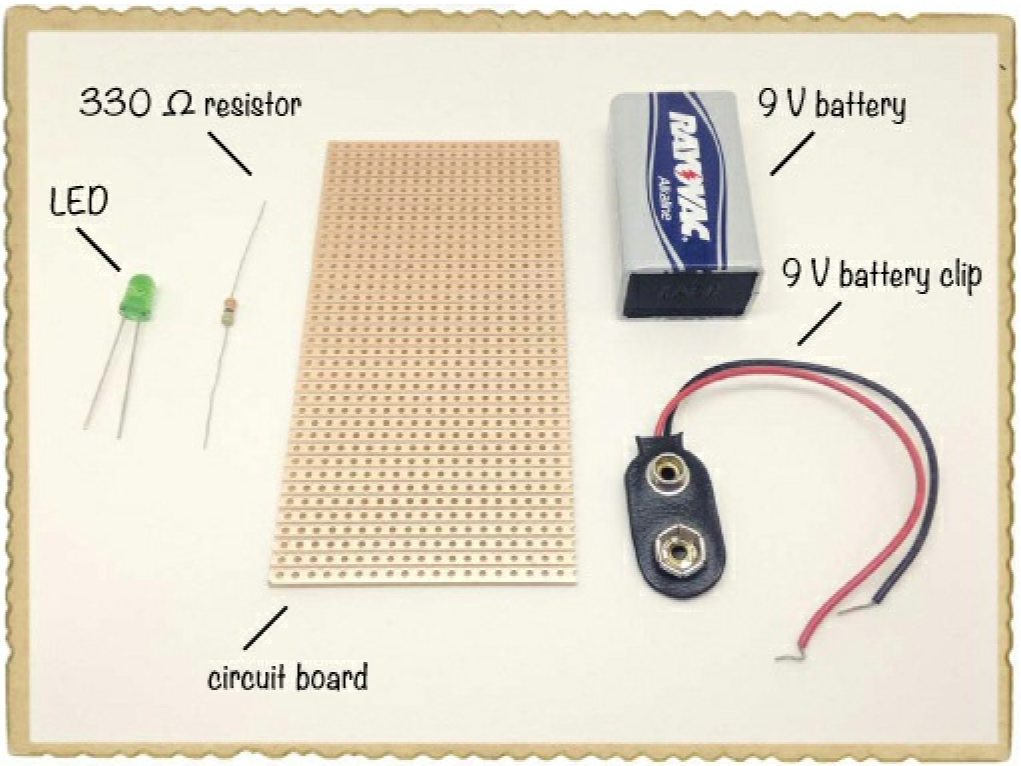
330 Ω resistor
LED
9 V battery
9 V battery clip
circuit board

- **A standard 9 V battery** to power the circuit.
- **A 9 V battery clip** (Jameco #11280, Bitsbox #BAT033) to connect the battery to the circuit.
- **A circuit board** (Jameco #2191488, Bitsbox #HW005) with copper strips.
- **A standard LED** (Jameco #34761, Bitsbox #OP003)
- **A 330 Ω resistor** (Jameco #661386, Bitsbox #CR25330R) for limiting the current to the LED.

Tools

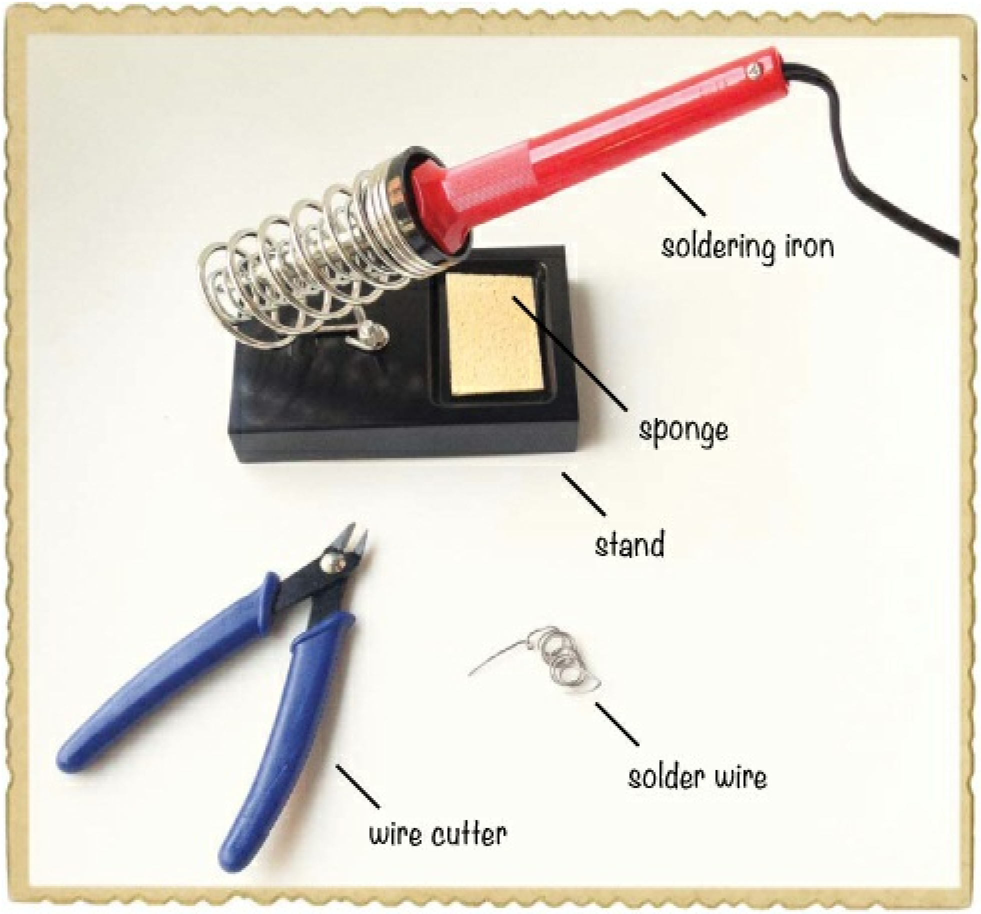

- **A soldering iron** (for example, Jameco #116572, Bitsbox #TL031)
- **A stand** (for example, Jameco #36329, Bitsbox #TL032) to hold the soldering iron.
- **A roll of standard solder wire** (for example, Jameco #94570, Bitsbox #HW022)
- **A wet sponge** to clean the tip of the soldering iron.
- **A wire cutter** (Jameco #35482, Bitsbox #TL008) to cut off excess legs from components.

Step 1: Place the Components

Look at your circuit board and get familiar with the connection pattern on it. The circuit board I recommend is called a *prototyping board*, and it's covered with little straight strips of copper, each with lots of holes. Any component leads that you solder into holes that share a strip of copper will become electrically connected.

Place your resistor and LED legs through the circuit board. The resistor should share a copper strip with the LED's positive lead. Note that the strips shown in the figure are on the bottom side of the board, which is why they're not copper colored in the picture. Anytime a project involves soldering to a circuit board, I'll show shadows of the copper strips so you can see where the connections are underneath.

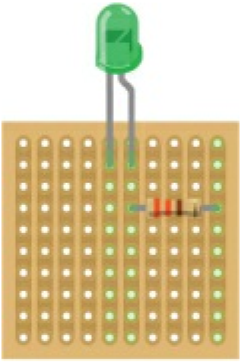

Step 2: Bend the Component Legs

Carefully flip your circuit board over so you can see the copper; hold the LED and resistor in place with a finger if needed so they don't fall out. Then, slightly bend each component's legs outward so that the LED and resistor stay in place even with the copper side up. Keep the board copper-side-up for now.

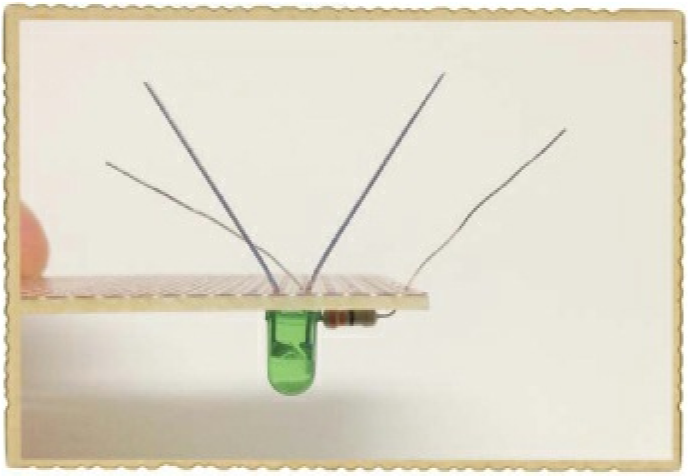

HOW TO READ THE CIRCUIT BOARD ILLUSTRATIONS

Throughout this book, you'll see circuit board illustrations I created in a program called Fritzing. These illustrations show the top side of the circuit board so you can see how the components look, and they show the bottom side—a slightly darker color than the top—so you can see how the copper strips look.

When following my instructions, always orient your circuit board with the copper strips on the bottom going in the same direction shown in the illustration. If you ever get confused about how you should read the connections in a circuit board illustration, just refer back to this example.

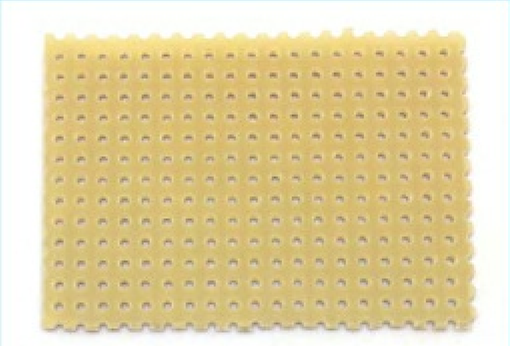

top view photo of real circuit board

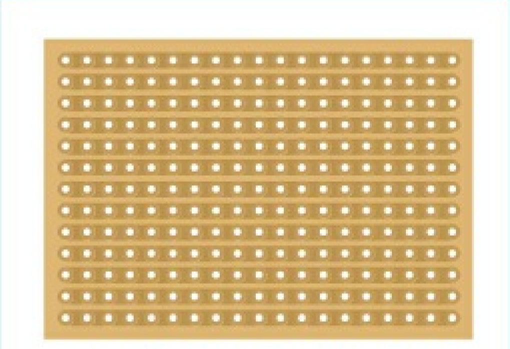

top view illustration showing copper strip locations underneath

If you want to create your own circuit illustrations with Fritzing, you can download the free, open source software from *http://www.fritzing.org/*.

Step 3: Heat and Clean the Soldering Iron

Your components should be in place and ready to solder, so plug in and heat up the iron now. This may take a few minutes. As always, check to see whether the iron is hot enough by touching some solder to the tip to see whether it melts.

Before you start soldering, clean the tip of the soldering iron, too. Wet your sponge just a little bit and wring out any excess water. Then, wipe the tip of the iron on the sponge to remove any old solder.

Step 4: Solder the Resistor and LED

With the copper side of the circuit board facing up, solder each component leg to the board, as described in “How to Solder” on page 112. Your board should look like this:

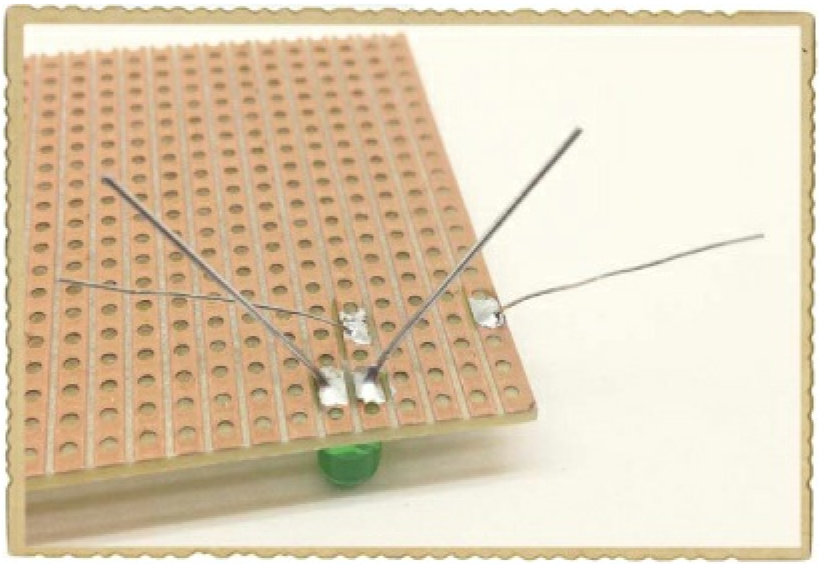

Step 5: Trim the Legs

Right now, your LED and resistor have really long legs. You don’t want to leave them sticking out because stray metal can create unintentional paths between components. At best, that could prevent your circuit from working until you move the leg away from the component it shouldn’t be touching; at worst, an unintended connection could break a component. To prevent accidental connections, cut off each leg just above the solder joint. The legs could go flying when you cut them, so turn the board away from your eyes or wear safety goggles for protection.

TIP

Hold the circuit board over an empty box while you cut the legs off, to prevent the scraps from getting lost and to make cleanup easier. You could also rest the circuit board on a table and hold the leg with one hand while you clip.

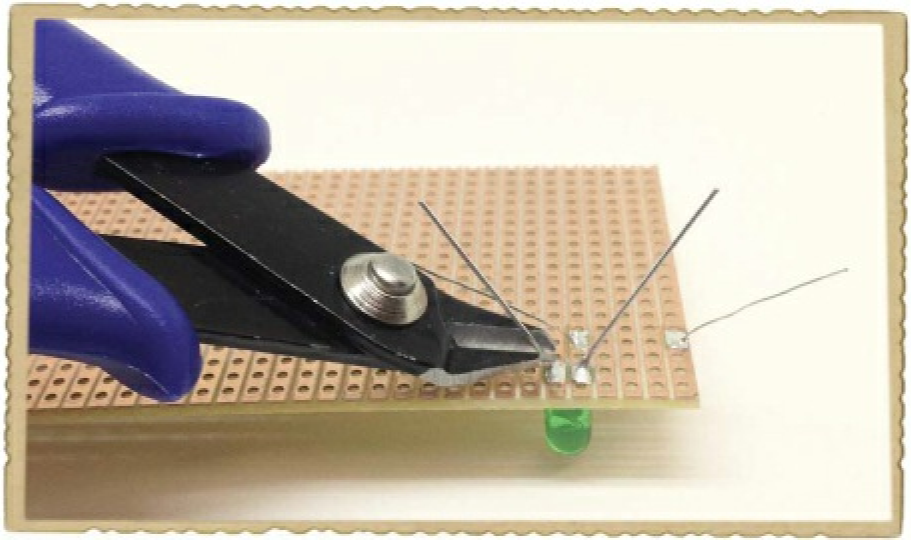

Step 6: Solder the Battery Clip

This circuit also needs a way to connect the battery to the components, so solder the battery clip to the circuit board now.

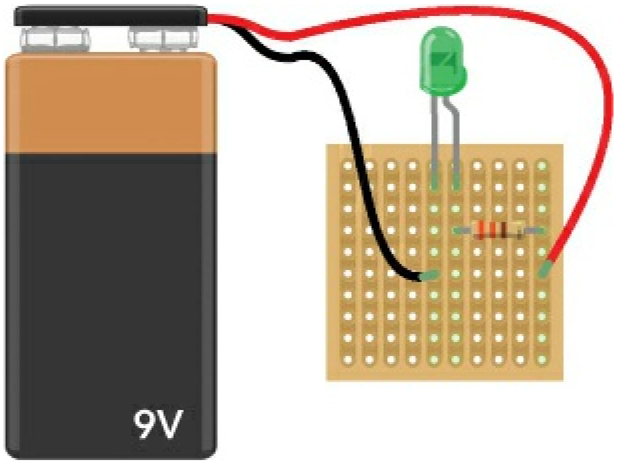

The red wire from the battery clip is the positive wire; connect it to the copper strip that connects only to the resistor, not the LED. The black wire is the negative lead, so connect it to the negative side of the LED. Your circuit board should now look like this:

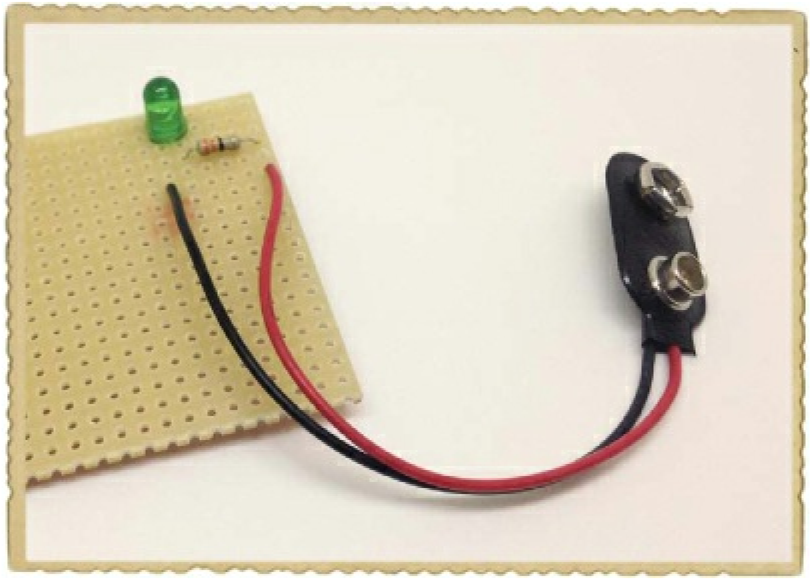

Step 7: Let There Be Light!

Now, let's test your circuit. Plug in the battery, and your LED should light up!

Step 8: What If the Soldered LED Circuit Doesn't Work?

If your LED doesn't light, check that you don't have any unintentional connections. Is solder connecting two joints that shouldn't connect? Are extra long component legs making contact?

Next, inspect your solder joints closely. Do any of them look like the examples of bad connections described in "Watch Out for Bad Solder Joints!" on page 117? If so, you might need to flow a little more solder: warm up your iron again and try to make sure your solder joints look like little pyramids.

Check the placement of the LED, too. Did you connect it correctly? Because you've cut off the excess legs, you can't see which leg is longer anymore, but if you look closely at the LED, you should see that one side of the base of the plastic housing is flattened. This is the negative side, which should be connected to the battery clip's black wire. If you've placed it the wrong way, here's one quick fix: use your wire cutters to cut the battery clip leads, removing the clip from the circuit board. Now, solder the positive battery lead where the negative lead used to be, and solder the negative battery lead where the positive lead used to be. This works because it doesn't matter which side of the LED the resistor is on.

OOPS! HOW DO I REMOVE A SOLDERED COMPONENT?

Sometimes, you'll make mistakes when soldering. For example, what if you soldered the battery clip the wrong way or used a 33,000 Ω resistor instead of a 330 Ω one? Well, don't worry: even the most experienced engineers mess up at soldering sometimes, and there's a way to fix your circuit when it happens. In these cases, you need to *desolder*, which just means to remove the solder from a solder joint.

Solder wick is a very useful tool for desoldering, and I suggest keeping some at your side any time you solder. The solder wick is made of braids of copper thread, so it's also called *desoldering braid*.

SOLDER-EX
2.0 mm
1.6 m
4 006062 071015
STANNOL-Ablötlitze
Art. Nr. 907101

When you place a piece of solder wick on top of a solder joint and heat both together, the solder should melt as usual, and the solder wick should absorb the liquid solder just as a dry cloth absorbs water. When you remove the solder wick, the solder should be on the wick instead of on the circuit board!

PROJECT #13: DESOLDER THE BATTERY CLIP

Knowing how to desolder is very useful, and not just for fixing soldering mistakes. For example, desoldering will let you exchange a broken component in a circuit for a fresh one—or you can desolder a component to reuse it in another circuit. In this project, I'll show you how to desolder your battery clip from the previous project.

Shopping List

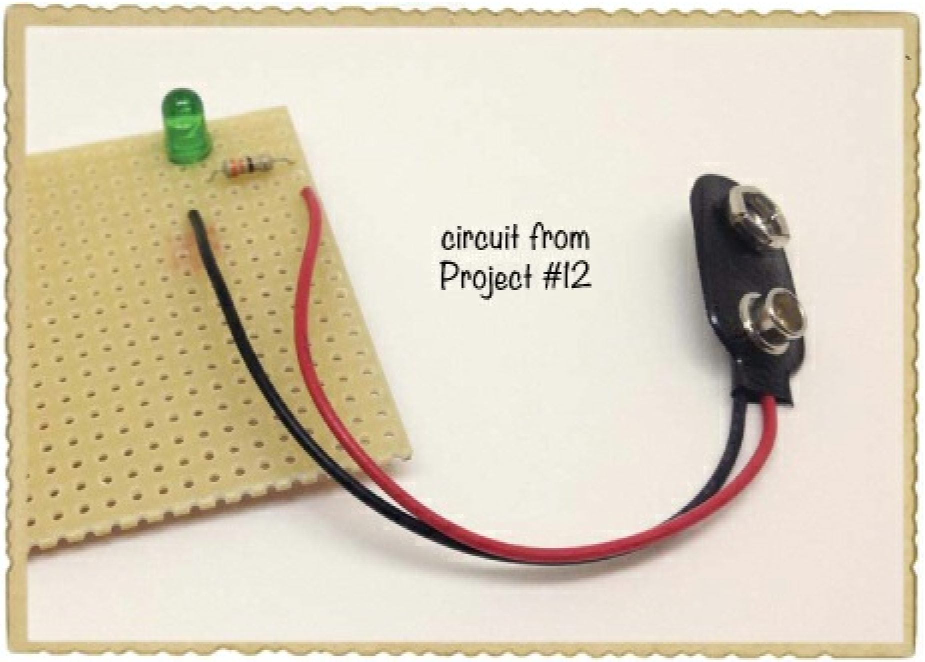
circuit from
Project #12

- **The circuit from "Project #12: Solder Your First LED Circuit" on page 118**

Tools

- **A soldering iron** (for example, Jameco #116572, Bitsbox #TL031)
- **A stand** (for example, Jameco #36329, Bitsbox #TL032) to hold the soldering iron.
- **A wire cutter** (Jameco #35482, Bitsbox #TL008) to cut the solder wick.
- **Solder wick** (Jameco #153462, Bitsbox #HW082) to remove the solder.

soldering iron
stand
solder wick
wire cutter

Step 1: Heat the Soldering Iron

Plug in your soldering iron and wait for it to heat. To test the temperature, touch some solder wire to the tip; the solder will melt when the iron is warm enough.

Step 2: Place the Solder Wick on the Solder Joint

Lay the end of your solder wick on top of the solder joint for one battery clip wire.

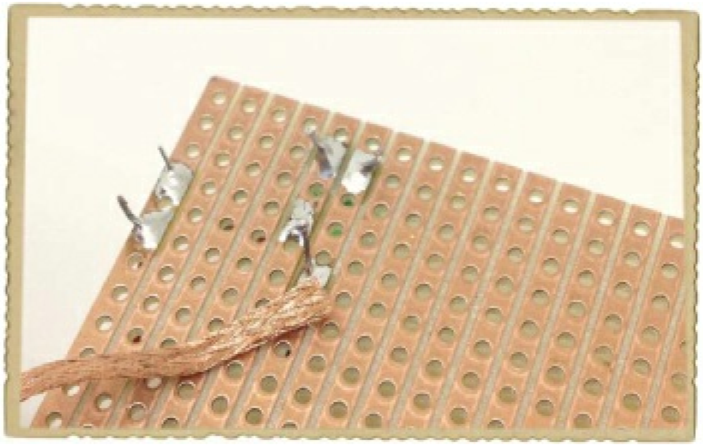

WARNING

Solder wick can get very hot when you heat it with the soldering iron, so don't hold the metal part by your fingers.

Step 3: Heat the Solder Joint and the Solder Wick

Place the tip of the heated soldering iron on top of the solder wick, directly over the solder joint you want to desolder.

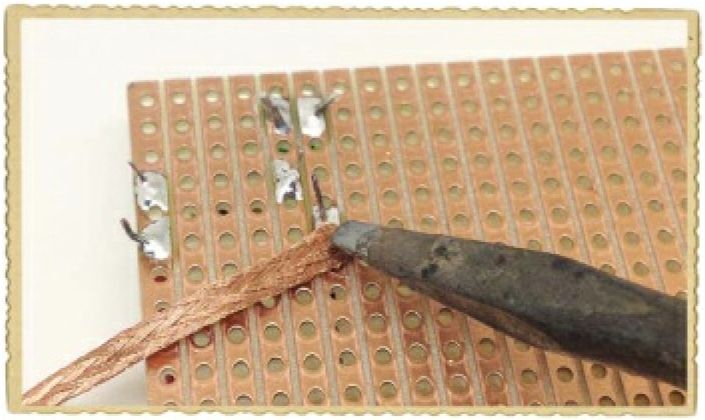

After a few seconds, the solder should melt and flow onto the braid. Lift the braid from the board, together with the soldering iron.

Step 4: Trim the Used Solder Wick

Remove the braid from the joint and look at it. The braid should have solder in it.

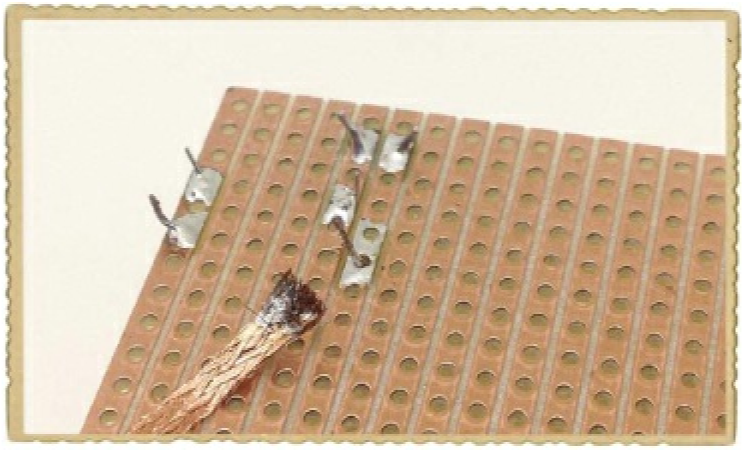

Much like a dirty cloth, this bit of wick is too dirty to use anymore. Use your wire cutter to cut off the piece that has solder inside.

Now, inspect your solder joint. If there's no solder left connecting the component leg to the circuit board, then you should be able to remove the wire from the hole. If not, repeat Steps 2 and 3 with a fresh bit of solder wick, and then repeat Step 3 until you can remove the wire.

Step 5: Remove the Other Battery Clip Wire

Repeat Steps 2 to 4 for the other battery clip wire, and you should be able to completely remove the battery clip from the circuit board. At this point, the copper side of your circuit board should look something like this:

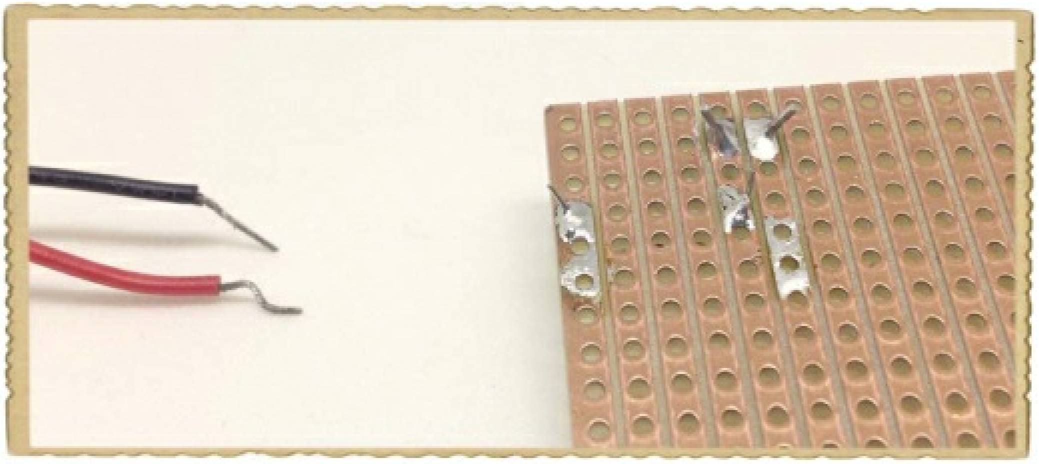

TRY IT OUT: SOLDER MORE STUFF!

Are you itching to get some extra soldering practice? One fun way to practice is to buy and build electronics *kits*, which include components and a circuit board that you can solder. You can also find circuit diagrams online, buy the necessary components, and solder them on a prototyping board, just as you did in this chapter.

Here are some online stores that sell both kits and components:

- **Jameco:** *www.jameco.com*
- **Adafruit:** *www.adafruit.com*

- **SparkFun:** *www.sparkfun.com*

WHAT’S NEXT?

In this chapter, you learned a very useful skill: soldering. This means you can make permanent circuits for your projects without worrying that they’ll fall apart. And with a little practice at desoldering, you can take projects apart on purpose, too.

In Chapter 7, I’ll introduce you to the transistor, the photoresistor, and the potentiometer. These are some really exciting components that you can use to make your circuit come alive and respond to its surroundings. For example, I’ll show you how to create a circuit that tells you when the sun rises in the morning!

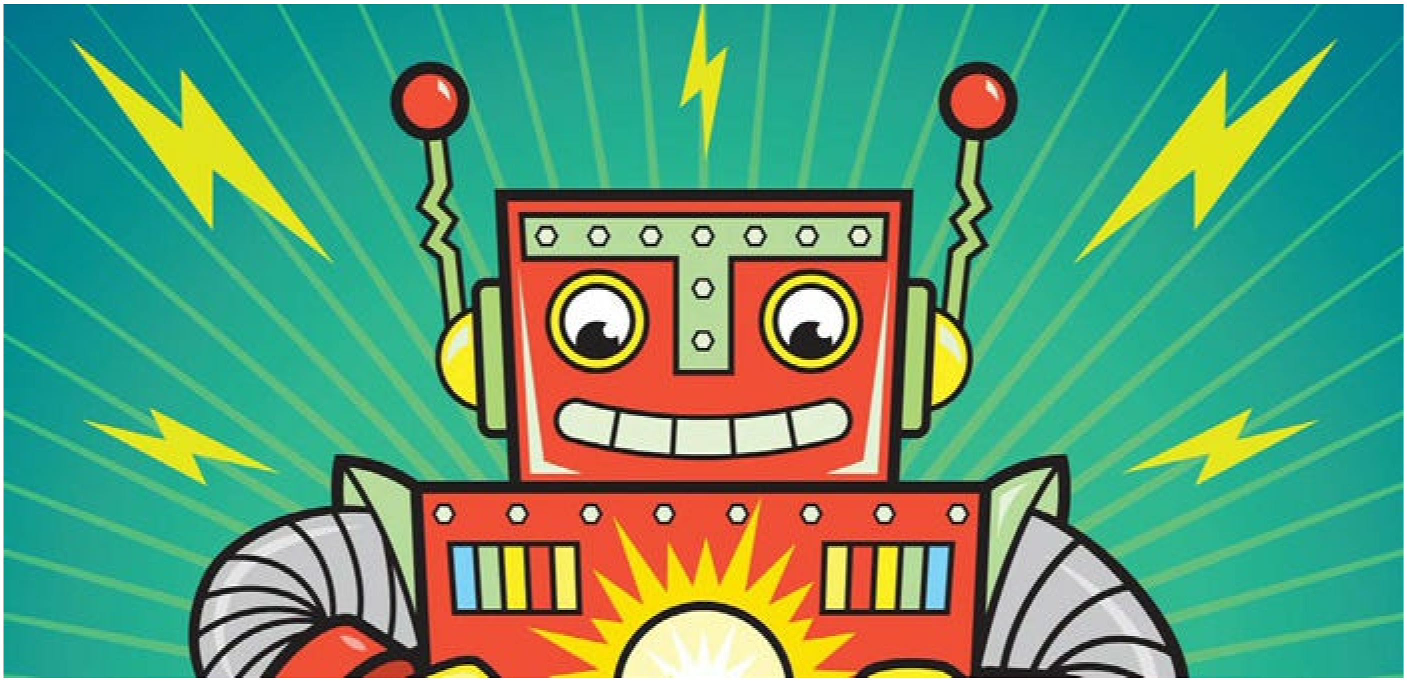

7

CONTROLLING THINGS WITH ELECTRICITY

Electronic components let you build smart things, like a lamp that turns on when it gets dark or a door that opens automatically when you approach it. In this chapter, you'll learn about a new component that will help you build such smart objects: the transistor.

The previous chapter showed you how to solder, and this chapter's projects—a touch sensor and an alarm that wakes you up when the sun rises—will give you more chances to practice your soldering skills. All you need for each project is the transistor and a few additional components.

MEET THE TRANSISTOR

The *transistor* is the most important component in electronics, and if you've ever heard experienced hardware enthusiasts talk about it, you've probably noticed that they tend to use a lot of difficult words. But the transistor really isn't hard to understand; in fact, you've already used something that acts a lot like one! Do you remember the relay you learned about in Chapter 5? The transistor is similar to the relay in many ways: it's like a switch that you can open and close with electricity.

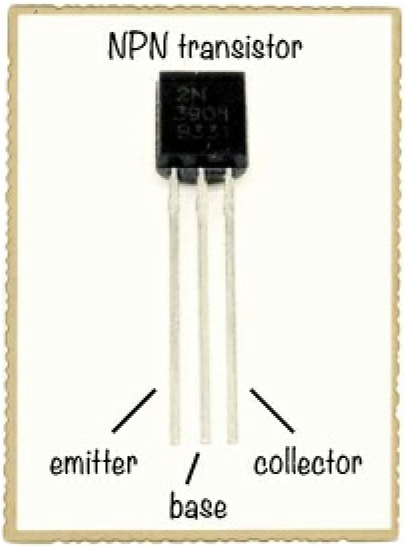

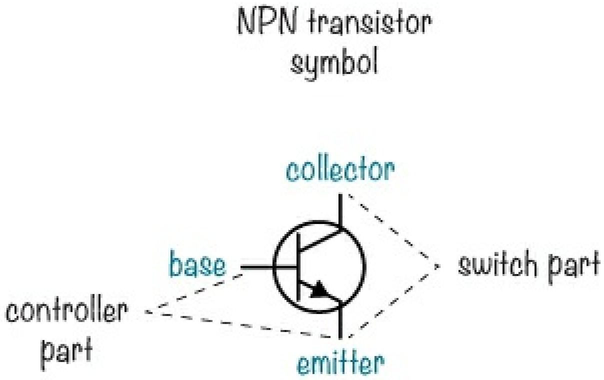

A transistor has three legs. In a standard *NPN transistor*, the three legs are called *emitter*, *base*, and *collector*. You'll often see these labeled *e*, *b*, and *c* in circuit diagrams. You turn the transistor on and off—that is, you open and close the switch—with the base and emitter legs, and you connect a circuit that you want to control between the collector and the emitter legs.

When you look at our example transistor with the flat side facing you, the leftmost leg is the emitter, the middle leg is the base, and the rightmost leg is the collector. But this is not the case for all transistors, so always check the datasheet for your transistor to find out which leg is which.

Why Use a Transistor?

If a transistor acts like a switch, then you might be wondering when you'd want to use a transistor instead of a switch in the first place. Well, think about a fan: if it's hot in your room and you want to turn a fan on, you have to manually flip a switch. But if that fan were part of the right kind of circuit with a transistor and a few other components, you could make the fan turn on automatically when the temperature in your room rises above 75 degrees Fahrenheit. To make this happen, you'd need one circuit that could sense temperature and another circuit that could turn on a fan.

Now, imagine a temperature-sensing circuit that gives a voltage when the temperature is above 75 degrees and no voltage if the temperature is below. If you were to connect one wire from the fan to an NPN transistor's collector, connect the fan's other wire to the positive terminal of the battery, and connect the transistor's emitter to the battery's negative terminal, the transistor would control when the fan is switched on.

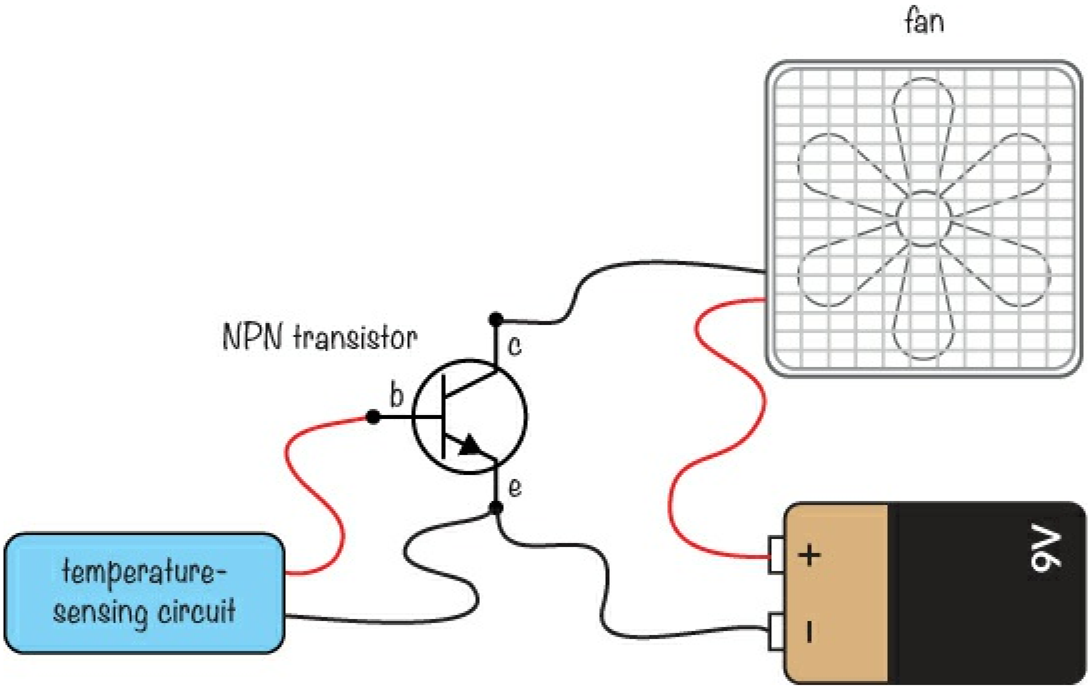

Then, you could connect the output of the temperature-sensing circuit to the controlling part of the transistor—that is, the base and emitter. In such a circuit, the fan would turn on when the temperature rises above 75 degrees and turn off when the temperature is less. Let's look at how the transistor's "switch" closes in the first place.

How the Transistor Works

When a little bit of current flows from the base of a transistor to the emitter, the transistor "closes the switch" so that current can also flow from the collector to the emitter.

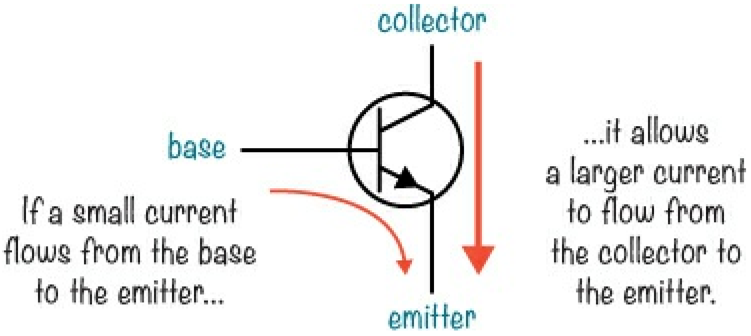
collector
base
If a small current flows from the base to the emitter...
...it allows a larger current to flow from the collector to the emitter.
emitter

In Chapter 1, I explained that voltage pushes current through a circuit. For current to flow from the base to the emitter, there needs to be a voltage to push that current. When the current flows from the base to the emitter, this opens a path for the current to flow between the collector and the emitter. You can also control the amount of current that flows between the collector and the emitter by changing the current flowing from the base to the emitter.

The specific type of transistor we're talking about is an *NPN-type bipolar junction transistor*. This long name describes the materials inside the transistor. There's some advanced chemistry and physics behind how it works, but you don't need to know all of that to build cool circuits with a transistor—you just need to know what the transistor does.

The important thing to remember for now is that different transistor types exist. For each transistor project in this book, just use the type I describe in the Shopping List sections, and your circuits will work just fine. And of course, when you build circuits outside this book, be sure to use the type of transistor specified in the circuit diagram.

Controlling an LED with a Transistor

Other projects in this book have used this simple LED circuit, complete with a resistor and a battery:

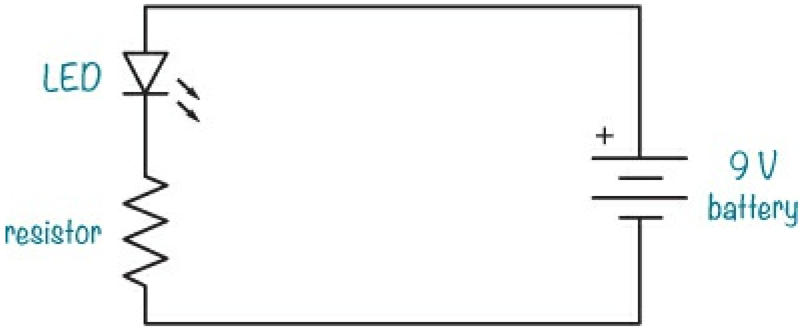
LED
resistor
+
9 V battery

Based on what you've learned so far, what do you think would happen if you put a transistor between the resistor and the battery's negative terminal?

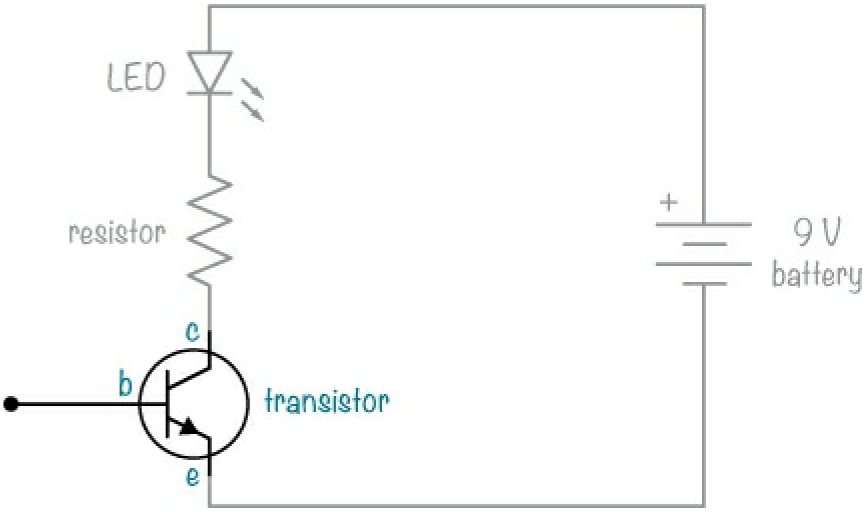

With no voltage on the transistor's base, or the controlling pin, no current flows from the base to the emitter. That means no current can flow between the collector and the emitter either, and the LED would be off.

But if you were to apply a small voltage to the base—for example, by connecting a small battery to it—the transistor would let current flow from the collector to the emitter, and the LED would glow. A transistor that allows current to pass is considered *on*; a transistor that doesn't allow current to pass is considered *off*. The amount of voltage needed to turn an NPN transistor on is about 0.7 V, so a circuit like this one would allow current to pass through properly to light the LED:

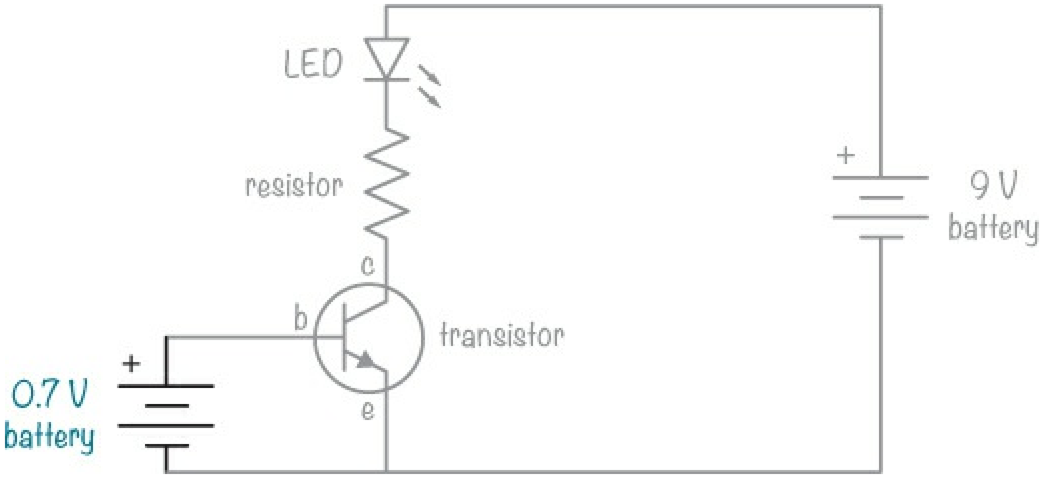

Normally, you wouldn't connect a battery to the base. Instead, you'd connect another circuit that you want to control the transistor, like a light-sensing circuit that gives a small voltage when it detects light. A circuit like that would turn the transistor on and off according to the light.

You can build a circuit like this to control LEDs and other components. In the next project, I'll show you how to use a transistor to turn an LED on with the touch of your finger.

PROJECT #14: BUILD A CIRCUIT THAT SENSES TOUCH

Did you know that your finger can act as a resistor? Your finger has a resistance of a few megohms (MΩ), and that's a lot! This resistance varies, though. If your finger is sweaty, for example, the resistance decreases.

In this project, you'll use your finger as a resistor to complete a circuit that turns on an LED, creating a touch sensor. A *sensor* is a component that can measure things in the world around you, like light or temperature. In many cases, a sensor is a resistor that changes its value based on light level, temperature, or some other physical quantity you might want to know about.

If you connect a resistor of a few megohms between the positive side of your battery and the base of the transistor in this circuit, a small current should flow through the base of the transistor to the emitter. This current should be large enough to turn the transistor on and let a little bit of current flow from collector to emitter.

This circuit has a transistor, a resistor, an LED, and a battery, just like the one in the previous section. This time, instead of a separate battery, you'll connect the 9 V battery to the transistor's base with your finger through a kind of touch pad. The touch pad will just be two exposed wires placed close enough together that you can touch both with your finger at the same time.

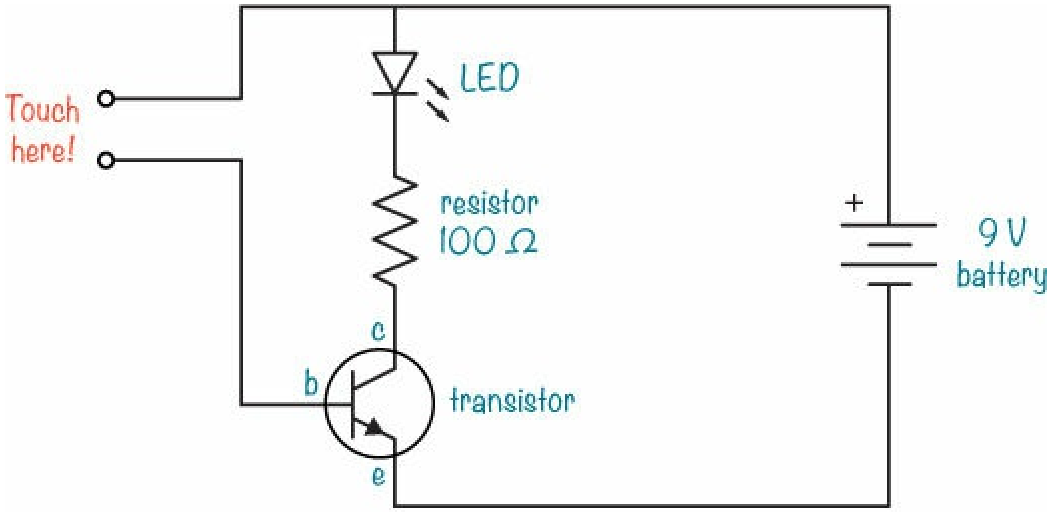

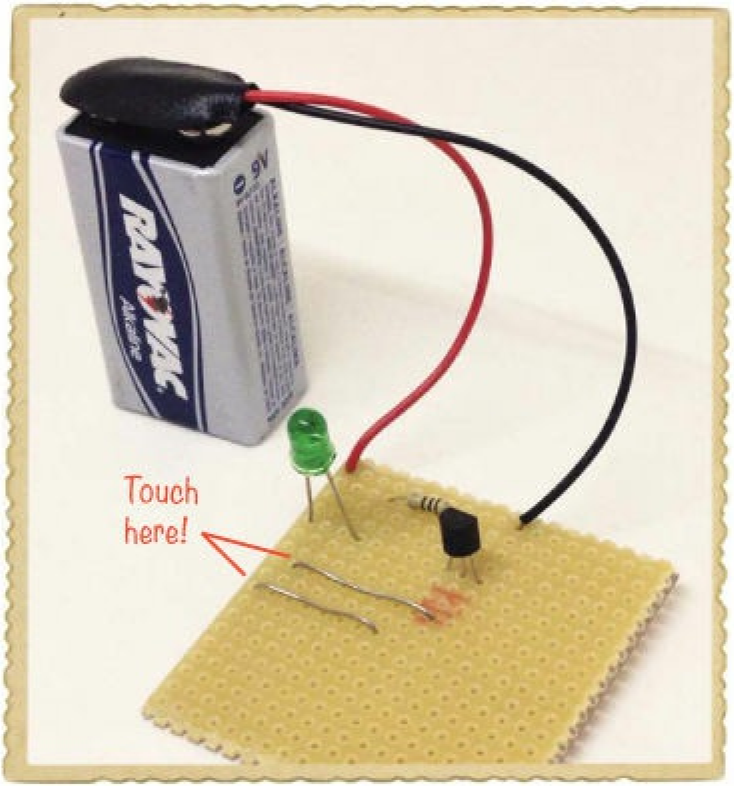

Notice that instead of a 330 Ω resistor with the LED that you used before, this circuit uses a 100 Ω resistor. Often the resistance in your finger is so high that the transistor won't turn on fully. With a smaller resistor, you should still get a bright LED, even if your finger's resistance is a bit high.

Shopping List

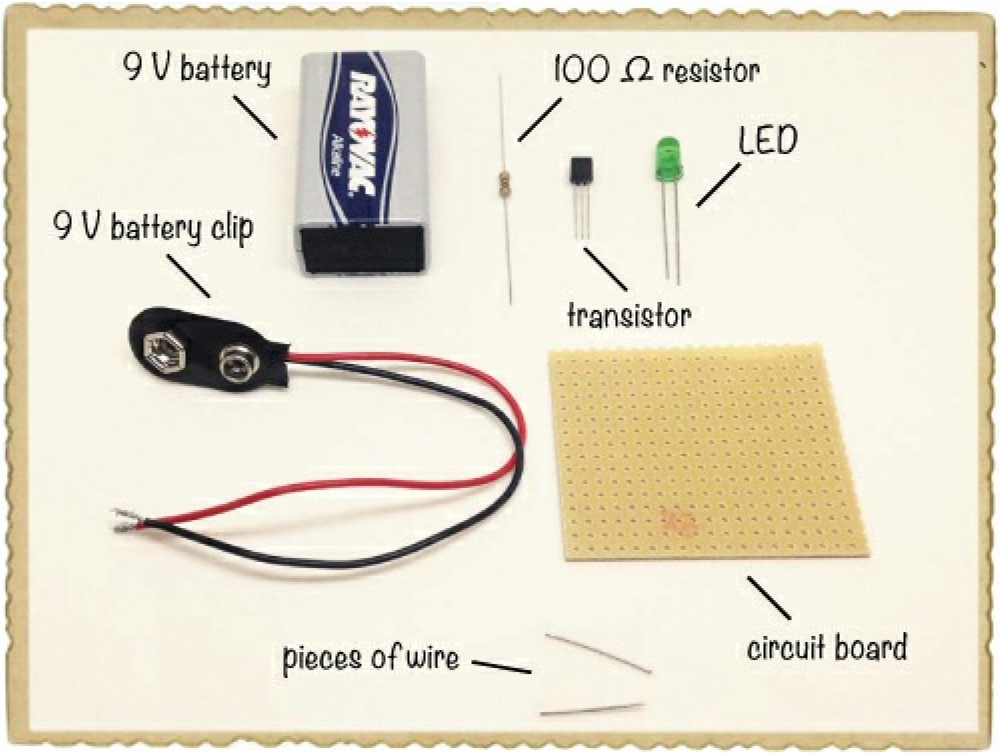
9 V battery
100 Ω resistor
LED
9 V battery clip
transistor
pieces of wire
circuit board

- **A standard 9 V battery** to power the circuit.
- **A 9 V battery clip** (Jameco #11280, Bitsbox #BAT033) to connect the battery to the circuit.
- **A circuit board** (Jameco #2191488, Bitsbox #HW005) with copper strips.
- **A standard LED** (Jameco #34761, Bitsbox #OP003)
- **Two pieces of exposed wire**, each about 1 inch long. You could also use two legs you cut from components in other projects.
- **A transistor 2N3904** (Jameco #38359, Bitsbox #QD018)
- **A 100 Ω resistor** (Jameco #690620, Bitsbox #CR25100R) for limiting the current to the LED.

Tools

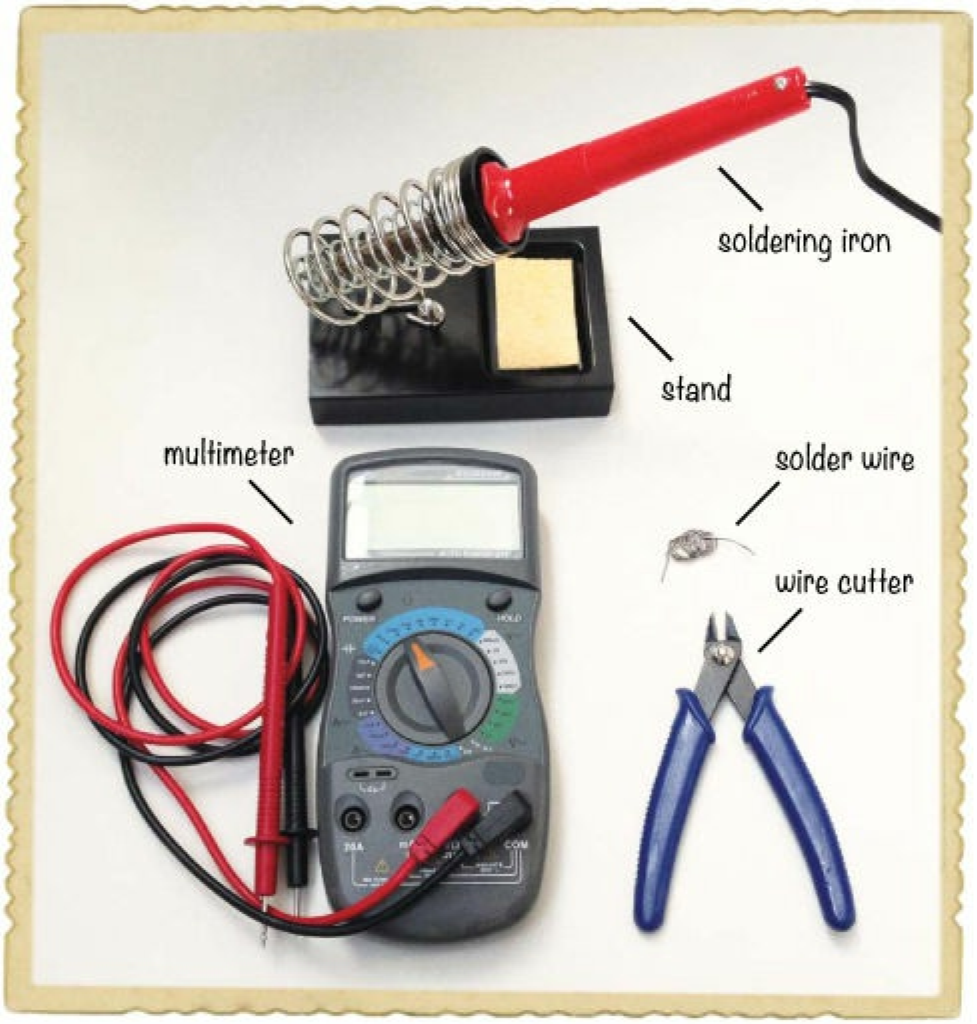

- **A soldering iron** (for example, Jameco #116572, Bitsbox #TL031)
- **A stand** (for example, Jameco #36329, Bitsbox #TL032) to hold the soldering iron.
- **Solder wire** (for example, Jameco #94570, Bitsbox #HW022)
- **A multimeter** (Jameco #2206061, Bitsbox #TL057, Rapid Electronics #55-6662) to measure voltages if the circuit doesn't work.
- **A wire cutter** (Jameco #35482, Bitsbox #TL008) to cut off excess legs.

Step 1: Place Components on the Prototyping Board

Place the LED, the resistor, and the transistor into the prototyping board as shown. Make sure a copper strip connects the resistor to (1) the LED's cathode leg and (2) the transistor's collector leg. Bend the legs on the copper side of the board so that the components stay in place.

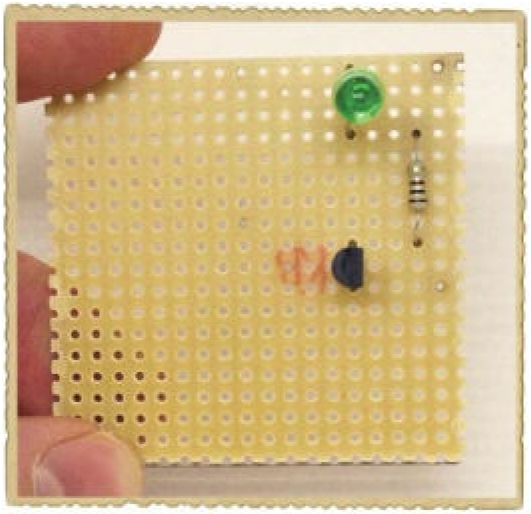

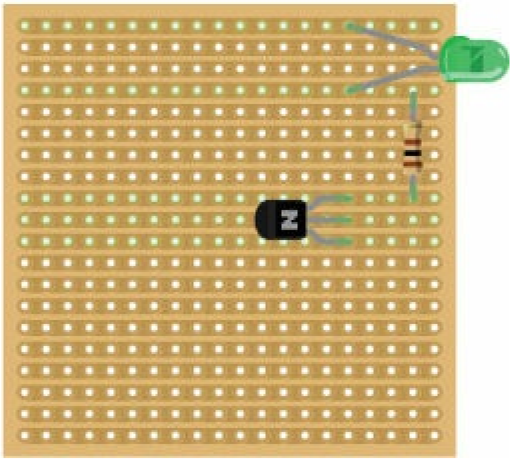

Step 2: Check Your Component Placement

Before soldering, look on the copper side of the board and double-check that your components are placed according to the directions in Step 1. Inspect the orientation of the LED and the transistor specifically, as these must be soldered the correct way for your circuit to work.

Step 3: Solder the Components and Trim Excess Legs

Solder the components to the board, just as you learned in “How to Solder” on page 112, and then use your wire cutters to cut off the excess legs. Wear your safety goggles and take care to turn the board away from yourself in case bits of wire fly off as you cut.

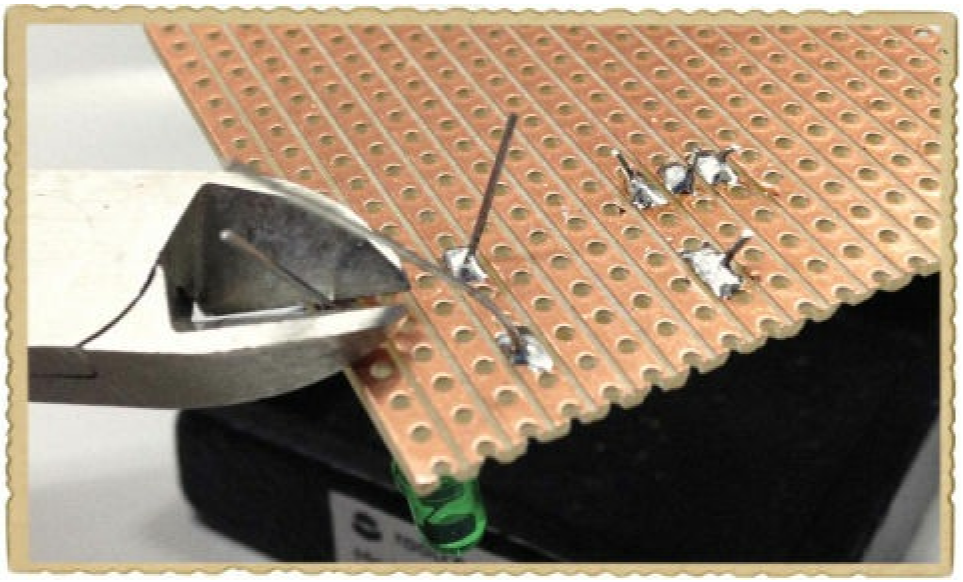

Step 4: Solder the Touch Pad

Next, solder the two pieces of exposed wire. Connect one wire to the LED’s anode leg, with the other end on an empty copper strip. Connect the other wire to the base of the transistor, also with the other end on an empty copper strip. Solder them in such a way that you can touch both with your finger.

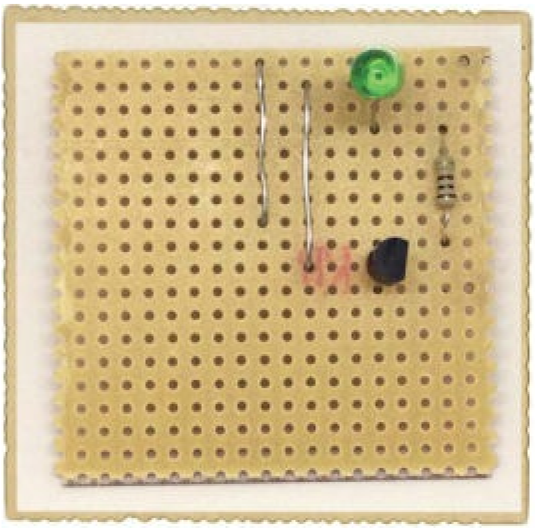

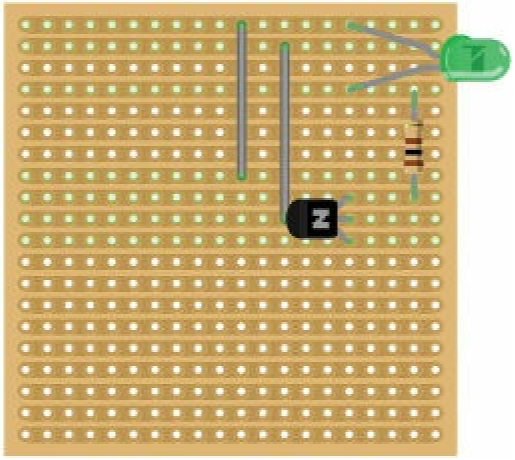

Step 5: Power It Up!

Now you need power! To finish the job, first solder the battery clip to the board, placing the red and black wires as shown at ❶. After you've soldered the battery clip, plug the battery into it at ❷.

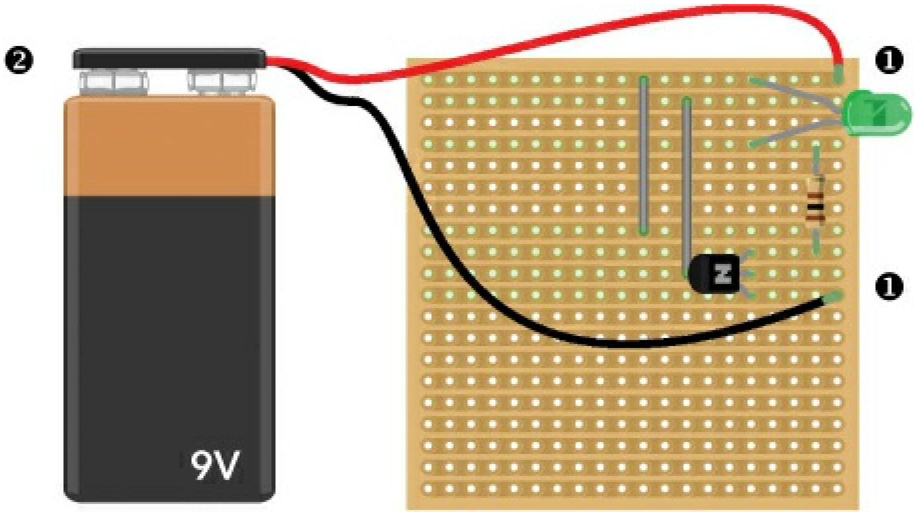

This touch-sensor circuit is ready to test!

Step 6: Test the Sensor

Touch both exposed wires at the same time with your finger. Your LED should light up. If you can't see the LED glow, try turning out the light in your room; the LED might just be dim. If you still can't see the LED glowing, dip your finger in water and try again, as wetting your finger reduces its resistance.

WARNING

Use only a finger to touch the exposed wires. If you use something with a very low resistance, like a piece of wire, you can destroy the transistor.

Step 7: What If the Touch Sensor Doesn't Work?

If nothing happens when you touch the wires, start by checking the direction of your LED and your transistor. It's very common to mix up the pins of these components, so go back to Step 1 and make sure they're connected on the prototyping board's copper strips, according to the images.

If the LED and the transistor are connected correctly, you can use your multimeter to measure the voltage between the base and the emitter of the transistor on the controller side, without touching the touch pad. Set your multimeter to a DC voltage range—20 V DC, for example—and then connect one multimeter cable to the base leg and one to the emitter leg, as shown. The multimeter should show a value around 0 V. Now, place your finger on the touch pad and measure the voltage again. The multimeter should show around 0.7 V.

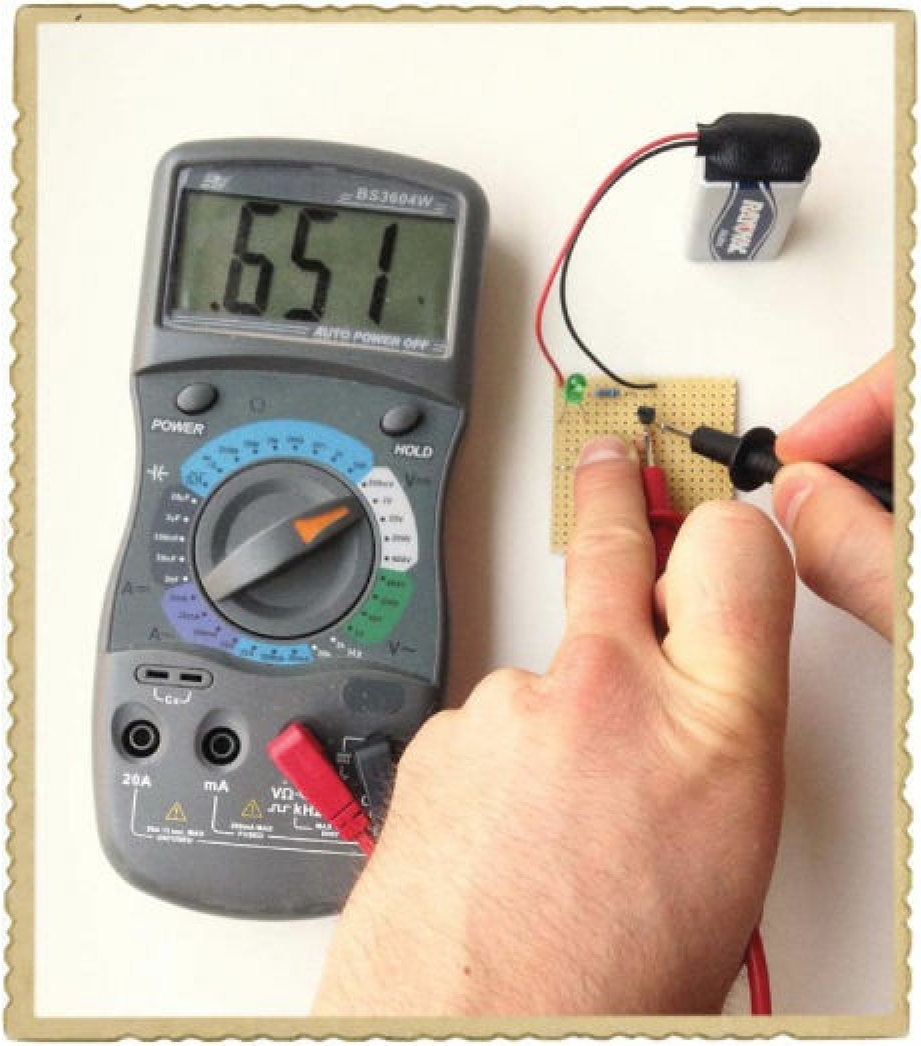

If the controller side of the transistor is fine, check the switch side. Measure the voltage from one leg of the LED to the other, without touching the touch pad. You should see 0 V on the multimeter. Place your finger on the touch pad and measure again; now it should be around 1.7 to 2 V.

If either multimeter measurement gives you incorrect values, then check your connections against the circuit diagram one more time, paying extra attention to the two exposed wires. Finally, check your solder joints, look for stray pieces of wire or solder bridging the gap between copper strips, and fix any bad connections you find.

TRY IT OUT: CAN THE TOUCH SENSOR DETECT

DIFFERENT TOUCHES?

Try touching the two exposed wires lightly and note the brightness of the LED. Now, try pressing really hard. Can you see any difference? Have a friend repeat the experiment. Was the LED’s brightness when your friend touched the wires different from when you touched them? If so, that means your friend’s finger has a different resistance!

When you press harder, the connection between the wires and your finger improves, reducing the resistance so it’s easier for the current to flow.

Here’s another fun thing to try with a friend: Touch one of the wires yourself and have your friend touch the other. Then, hold your friend’s free hand with your free hand. Can you see the LED light up? Now, the current goes from the battery through both of your bodies and into the transistor. But don’t worry—the current is so low that it’s not dangerous and you won’t feel anything.

RESISTORS THAT CAN CHANGE VALUE

Up until now, you’ve used only resistors that have a fixed resistance, but you can also find resistors that have a *variable* resistance value, which means the resistance can change. For example, some resistors change value when you turn a knob, while others change their value based on the temperature or the amount of light. This section introduces two variable resistors: the potentiometer and the photoresistor.

Meet the Potentiometer

In Chapter 4, you learned about the standard resistor, a component that has a certain, unchangeable resistance. The *potentiometer* is also a resistor, but it has a variable resistance, and it’s often used to control things such as the volume of a speaker. (You know the volume control on the radio? That’s often a potentiometer.) A potentiometer usually has three pins and a shaft you can rotate to change the resistance.

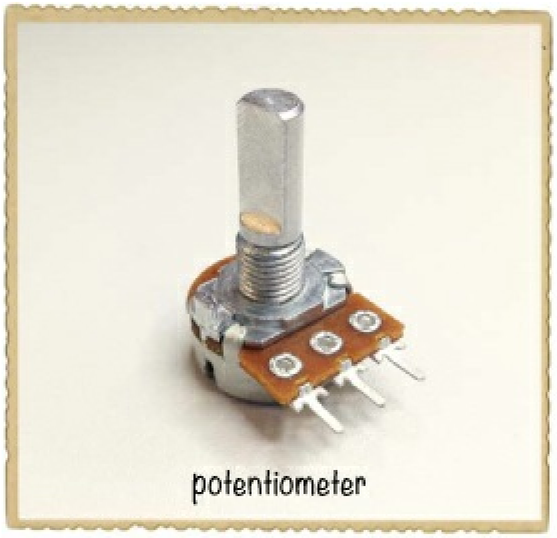
potentiometer

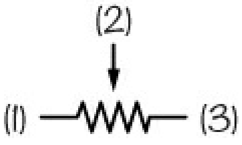

potentiometer symbol

The potentiometer symbol represents how the potentiometer works and the functions of the three pins. The resistance between pins 1 and 3 is a fixed resistance of a certain value. This value is equal to the value listed when you buy the potentiometer. If you have a 10 kΩ potentiometer, for example, then the resistance between pins 1 and 3 will be 10 kΩ.

Pin 2 is called the *wiper*. It connects to the resistor somewhere between pins 1 and 3. You can change the position of the wiper by turning the shaft of the potentiometer. If you turn the shaft so that the wiper comes closer to pin 1, the resistance between pin 1 and the wiper gets smaller, but the resistance between pin 3 and the wiper becomes larger.

Meet the Photoresistor

A *photoresistor* is another variable resistor. *Photo* means *light*, and this component's resistance changes with the amount of light shining on the top of it. Sometimes this component is also called a *light-dependent resistor (LDR)* because its value depends on light.

Photoresistors are made of a material with some special properties. In the dark, this material has a high resistance, but when light shines on it, the light energizes electrons that would otherwise be bound in the material. Those energized electrons can flow freely through the material, reducing the resistance. The more light that shines on the photoresistor, the less resistance it has.

photoresistor

Credit: Wikimedia Commons

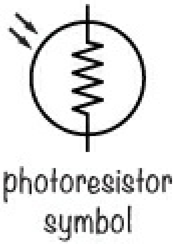

DIVIDING A VOLTAGE WITH RESISTORS

When the resistance of a photoresistor or potentiometer in a circuit changes, the voltage and/or the current must change, too, according to Ohm's law. (See "Introducing Ohm's Law" on page 73 if you want to walk through the math again.) If you make your variable resistor part of a circuit called a *voltage divider*, you'll get a voltage output that varies with the resistance. You can use that changing voltage to control another component in your circuit. Knowing how to identify voltage dividers can also help you understand how other circuits work.

What Does a Voltage Divider Look Like?

If you connect two resistors with the same resistance to each other and to the positive and negative sides of a battery, the voltage where your resistors meet will be half the battery voltage—for example, 4.5 V if you use a 9 V battery. This circuit is called a *voltage divider*.

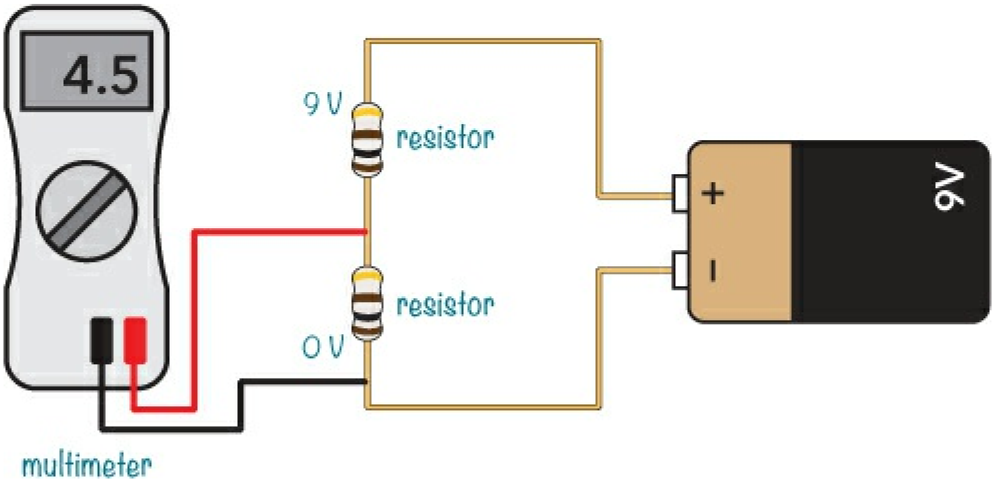
4.5
9 V
resistor
resistor
0 V
multimeter
+
-
9V

If you use unequal resistors instead of identical ones, you can use a voltage divider to get voltages anywhere from 0 V up to your battery voltage. You just need to do a little bit of math.

Calculating the Voltage from a Voltage Divider

Let's say you have the following circuit. What's the output voltage (V_{out}) from this circuit?

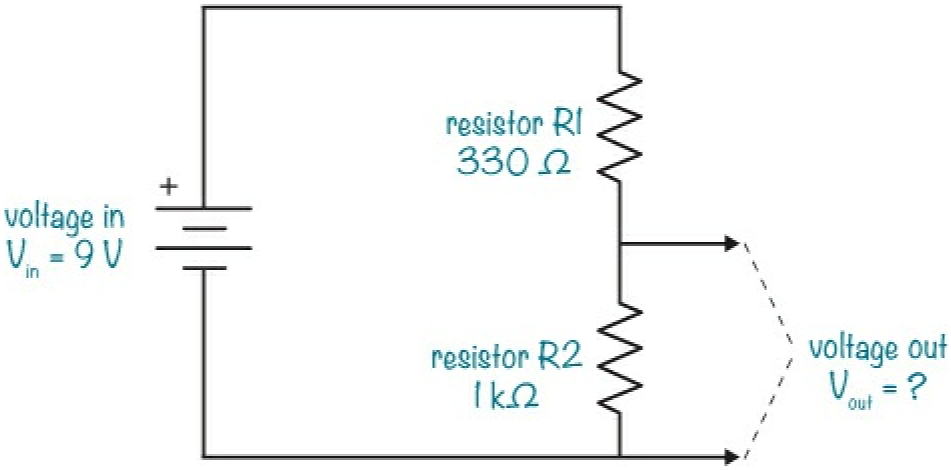
resistor R1
330 Ω
+
voltage in
V_{in} = 9 V
resistor R2
1 kΩ
voltage out
V_{out} = ?

To find V_{out}, enter the values from the circuit into the following formula:

$$V_{out} = V_{in} \times \frac{\mathrm{R2}}{\mathrm{R1} + \mathrm{R2}}$$

$$V_{out} = 9\ \mathrm{V} \times \frac{1000\ \Omega}{330\ \Omega + 1000\ \Omega}$$

$$V_{out} = 9\ \mathrm{V} \times 0.752$$

$$V_{out} = 6.77\ \mathrm{V}$$

In this example, the output voltage is 6.77 V—roughly two-thirds of the 9 V input from the battery.

How a Voltage Divider Can Help Measure Light

In the beginning of this section, I mentioned that the voltage divider can help you measure light, but how do you do that? Just replace one resistor in the voltage divider with a photoresistor, and you'll get a circuit that outputs a voltage based on the amount of light shining on it. And by adjusting the other resistor in the voltage divider, you can set the circuit to give a certain voltage at a certain light level. Connect the output of this circuit to a transistor that controls a buzzer, and you've got yourself a light-controlled alarm!

PROJECT #15: BUILD A SUNRISE WAKE-UP ALARM

It's time to combine all the concepts you've learned in this chapter into one fun project: a sunrise wake-up alarm!

This circuit starts an alarm when it detects light. After you build it, you can place it in your window (between any curtains and the glass) when you go to bed. When the sun rises, your circuit should detect the light and start the alarm, leaving you with no choice but to get up and turn it off.

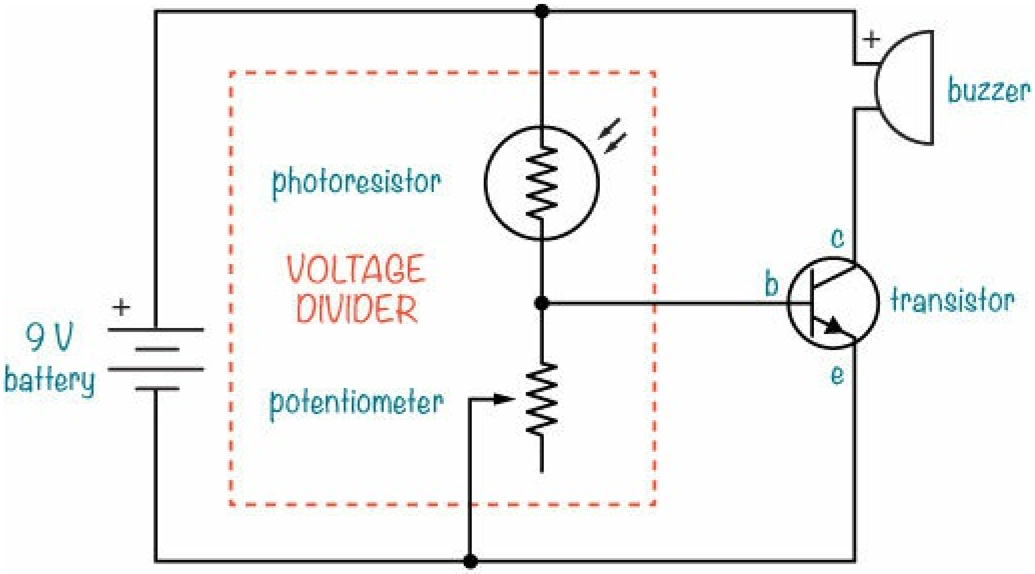
photoresistor
VOLTAGE DIVIDER
potentiometer
9 V battery
+
buzzer
+
c
b
e
transistor

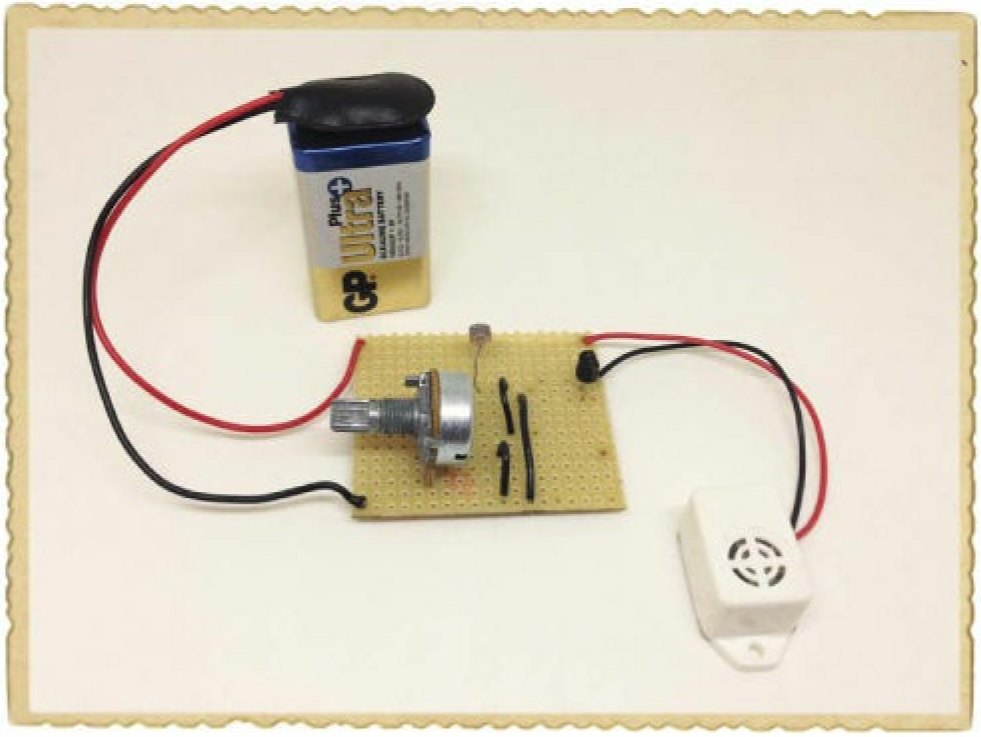

This project uses a transistor, a photoresistor, a potentiometer, and a buzzer. The photoresistor and the potentiometer create the voltage divider, and the output of the voltage divider is connected to the base of the transistor. The amount of light the photoresistor detects will determine whether the transistor is on or off. If the transistor is on, current should flow through the buzzer, and the buzzer should make sound. The potentiometer is there to set how much light the circuit needs before the alarm triggers.

Shopping List

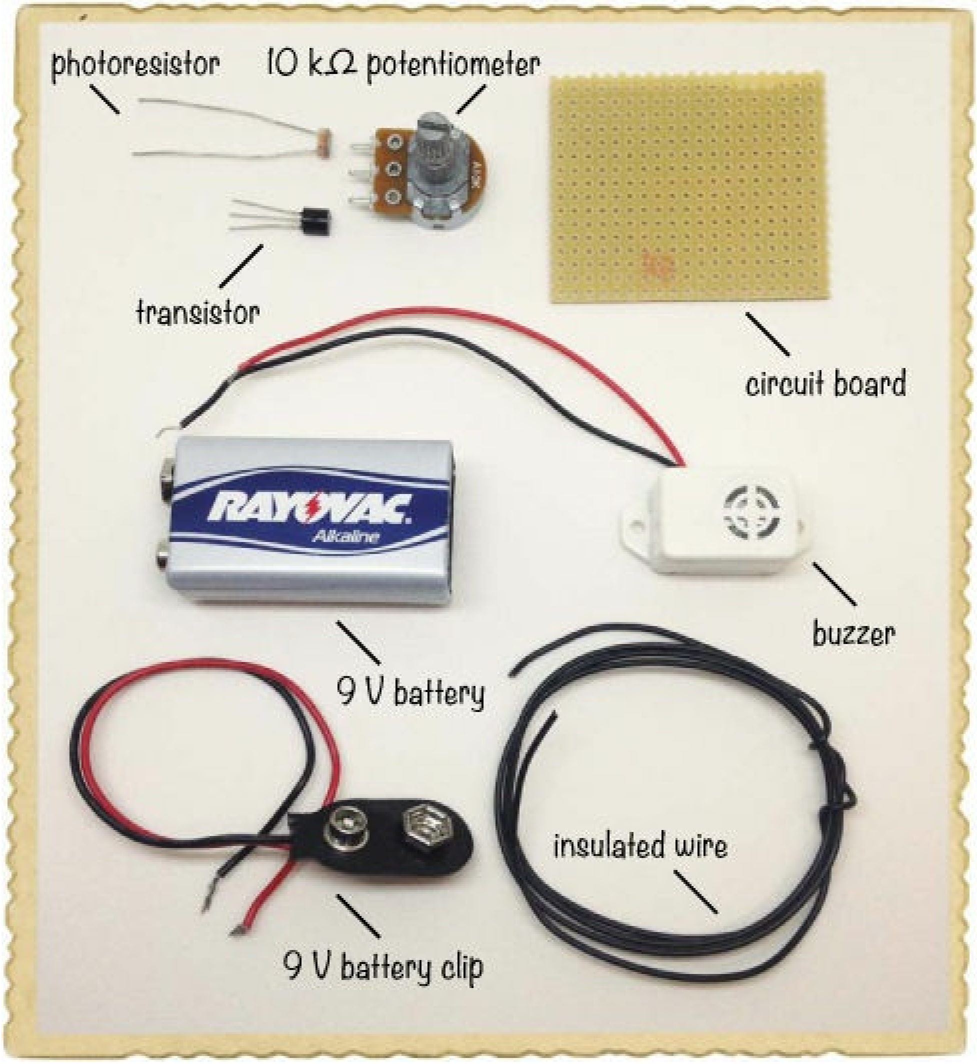
photoresistor
10 kΩ potentiometer
transistor
circuit board
9 V battery
buzzer
9 V battery clip
insulated wire

- **A standard 9 V battery** to power the circuit.
- **A 9 V battery clip** (Jameco #11280, Bitsbox #BAT033) to connect the battery to the circuit.
- **A circuit board** (Jameco #2191488, Bitsbox #HW005) with copper strips.
- **Insulated wire** (Jameco #36792, Bitsbox #W106BK), about 10 inches in length. Standard hookup wire works fine.
- **A transistor 2N3904** (Jameco #38359, Bitsbox #QD018)
- **A 10 kΩ potentiometer** (Jameco #2118791, Bitsbox #VR004)
- **A photoresistor** (Jameco #202454, Bitsbox #ST004) for detecting light.
- **A buzzer** (Jameco #2173870, Bitsbox #ST016) that beeps. Buzzers come in both active and passive versions. You'll need an active buzzer that works with 9 V in this project, just like the one you used in "Project #2: Intruder Alarm" on page 11.

Tools

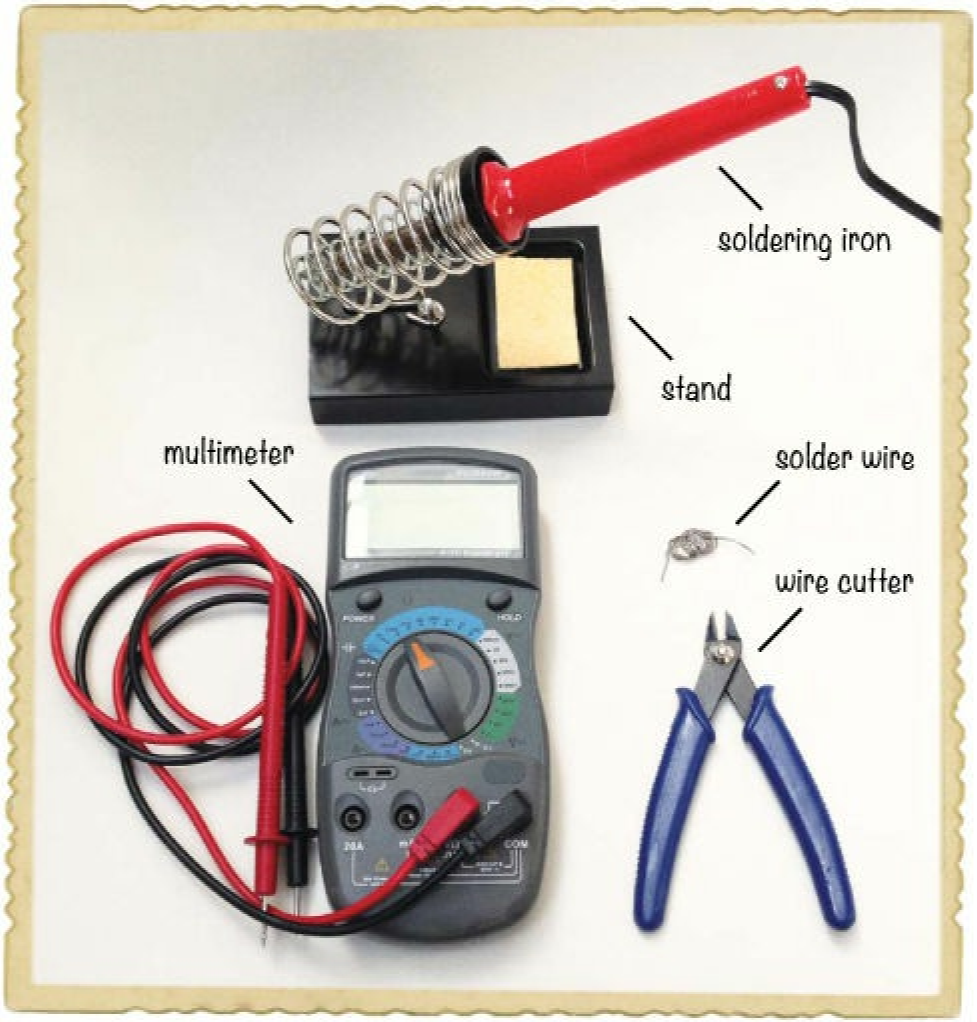

- **A soldering iron** (for example, Jameco #116572, Bitsbox #TL031)
- **A stand** (for example, Jameco #36329, Bitsbox #TL032) to hold the soldering iron.
- **Solder wire** (for example, Jameco #94570, Bitsbox #HW022)
- **A multimeter** (Jameco #2206061, Bitsbox #TL057, Rapid Electronics #55-6662) to measure voltages if the circuit doesn't work.
- **A wire cutter** (Jameco #35482, Bitsbox #TL008) to cut off excess legs.

Step 1: Place Components on the Prototyping Board

Start by placing the transistor, the photoresistor, and the potentiometer on the board, as shown. Bend the legs of the photoresistor and transistor on the other side so that they stay in place. The top copper strip will be the positive battery connection, and the bottom copper strip will be the negative one.

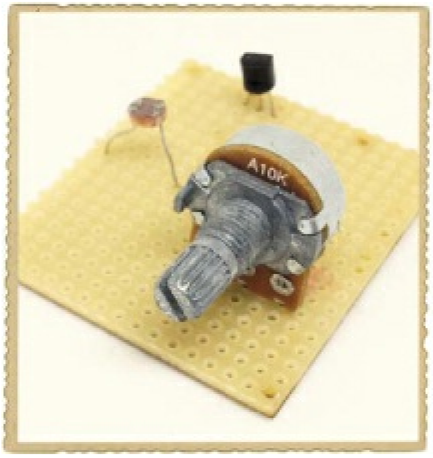

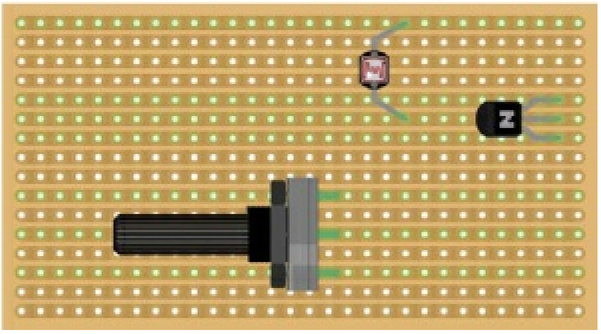

The diagram shows the components and the copper tracks underneath, while the photo shows what your board should look like in real life.

Step 2: Solder the Components and Trim the Legs

Check the placement of the components closely against the circuit diagram at the beginning of the project and the illustration in Step 1. Make sure the transistor is connected the right way and that the photoresistor has one leg on the same copper row as the center pin of the transistor. Each of the other pins, including each pin on the potentiometer, should be on a row of its own.

When you're sure your components are placed correctly, solder the pins to the board.

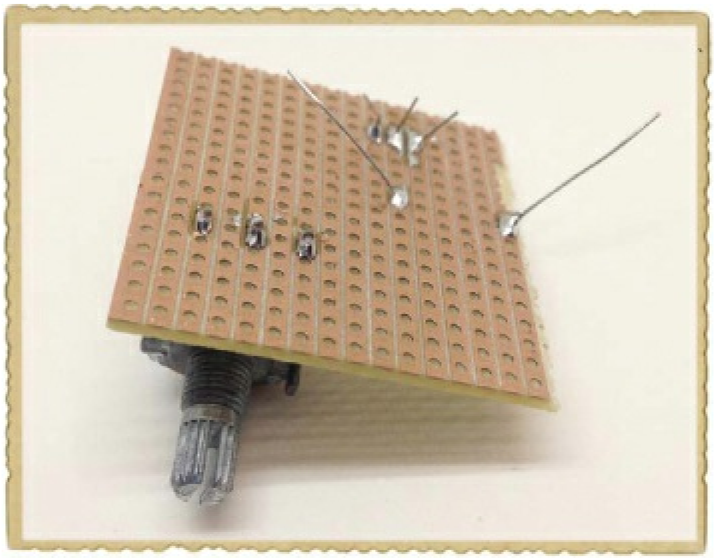

When you're finished soldering, cut off the excess legs.

Step 3: Add the Buzzer to the Board

Next, solder the buzzer to the board. Place the buzzer's positive (red) lead through a hole in the top copper row together with the photoresistor. Then place its negative (black) lead in the same row as the collector leg of the transistor. Solder both now.

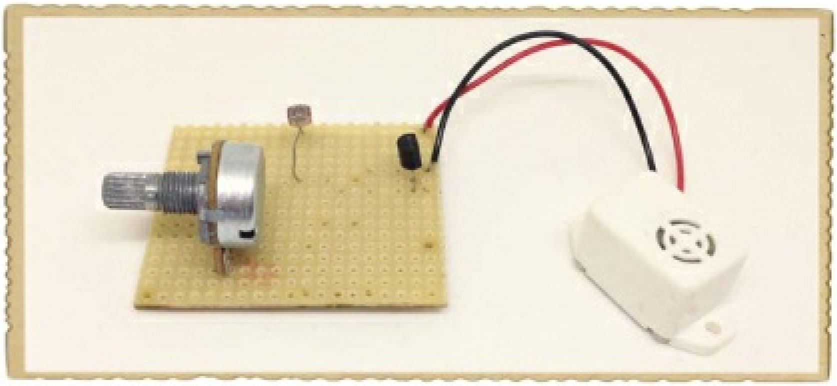

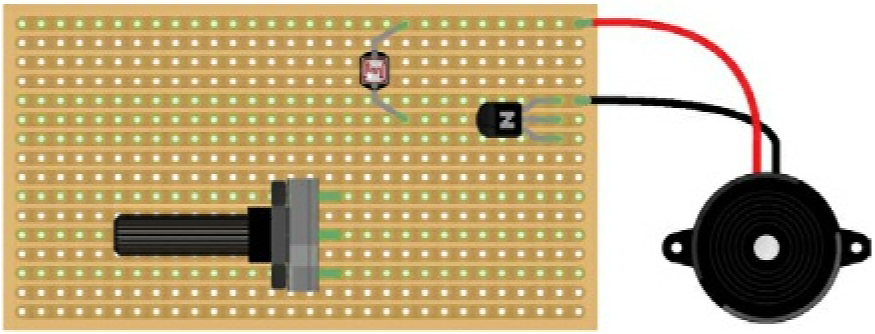

Step 4: Make the Remaining Connections with Wires

If you compare your circuit to the circuit diagram, you'll see that you're still lacking a few connections to make the circuit complete. The circuit board is missing the battery clip, but to work when powered, the circuit also needs the following connections:

- The base of the transistor needs to connect to the upper potentiometer pin.
- The middle potentiometer pin needs to connect to the negative battery terminal.
- The emitter of the transistor needs to connect to the negative battery terminal.

To create these remaining connections, you can solder three small wires like this one to the circuit board as jumper wires:

Cut a piece of wire about 2 inches long and strip about 0.3 inches of insulation from both ends. Removing insulation from shorter wires can be tricky, so if you're struggling with this step, use a longer wire instead. When you know what length of wire you can strip most easily, prepare two more wires. Then, solder the three wires to make the remaining connections.

Run one wire from the base of the transistor to the upper pin of the potentiometer. Next, connect the one wire from the middle pin of the potentiometer to an empty row at the bottom of the board.

Finally, run one wire from the same row of the prototyping board, where you just connected the previous wire, to the row connected to the emitter of the transistor. Check that your wires match the illustration and then solder them to the board.

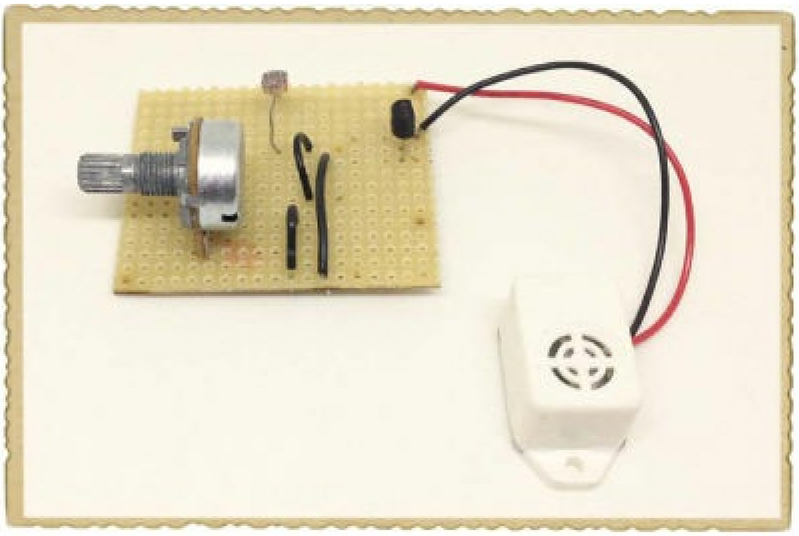

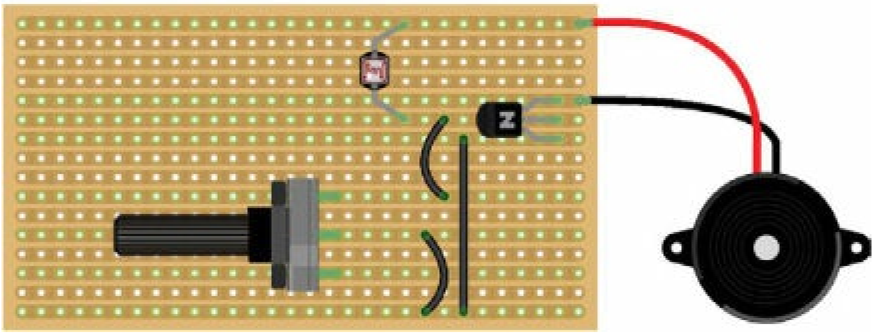

Step 5: Add the Battery Clip to the Board

All that's missing is the battery clip. Solder the battery clip's red wire to the top copper row of the board. The top row should also contain the photoresistor leg and the buzzer's red wire. Solder the battery clip's black wire to the bottom row.

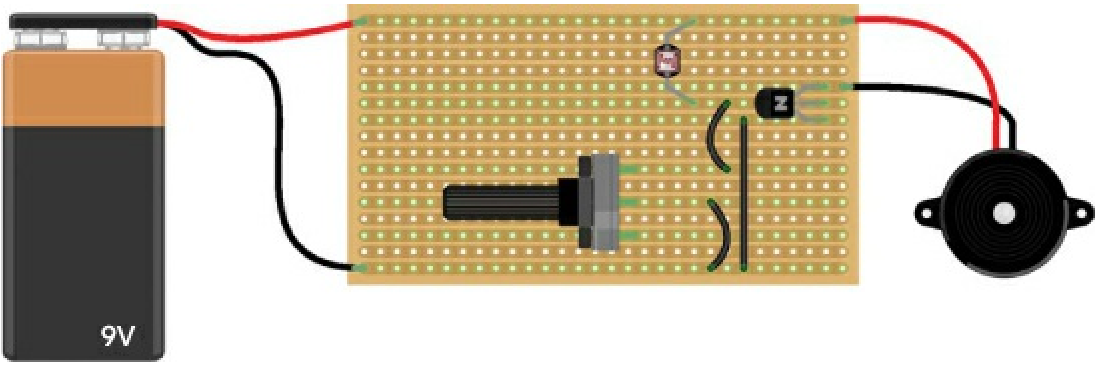

Step 6: Set a Wake-Up Call

Plug in the battery and put the circuit in an area that has the amount of light that you want to activate your alarm. Turn the shaft of the potentiometer until you find a position where the sound turns on and off with just a little nudge back and forth on the shaft. Now, turn the shaft just enough to make the sound turn on all the way. Place your hand over the photoresistor to block the light, and the sound should stop. Remove your hand, and the sound should turn on again.

Remove the battery from the circuit and wait until you're about to go to bed. With the lights off, place the circuit in your window, connect the battery, and go to sleep. When the sun rises and the light outside is as bright as the light you used to set your potentiometer, you'll be awakened by your very own sunrise alarm!

Step 7: What If There's No Sound?

Go through the circuit component by component, preferably with a friend. Check that your prototyping board has every connection shown in the circuit diagram. Check that there are no *short circuits*, too. A short circuit is an unintentional connection between something in a circuit. For example, make sure solder joints that are close to each other aren't actually touching.

If all looks good, use your multimeter to measure the voltage between the base and the emitter of the transistor. First, turn the shaft of your potentiometer all the way to one side, measure the voltage, and write it down. Then, turn the potentiometer all the way to the other side and measure the voltage again. Your multimeter should show 0 V on one side and around 0.7 V on the other side. If it doesn't, check your connections again.

If you're still not sure what's wrong, turn the potentiometer all the way to one side and measure the voltage between the red and black buzzer wires. Then, turn the potentiometer all the way to the other side and measure again. You should get 0 V on one side and about 8 to 9 V on the other.

If all else fails, check your solder joints and redo any that look like the connections might not be complete. And if you've built your circuit flawlessly, you may just have some broken components. Try breadboarding the project with new components, and if that works, solder those components instead.

TRY IT OUT: TEMPERATURE-CONTROLLED FAN

You can use the circuit you built in Project #15 for other things as well. For example, if you exchange the photoresistor for a *thermistor*, a resistor that changes resistance value based on temperature, the circuit will respond to temperature instead of light.

You can also change what's being controlled. Instead of an LED and resistor, try putting in a fan. Now you have a temperature-controlled fan! The circuit diagram looks like this:

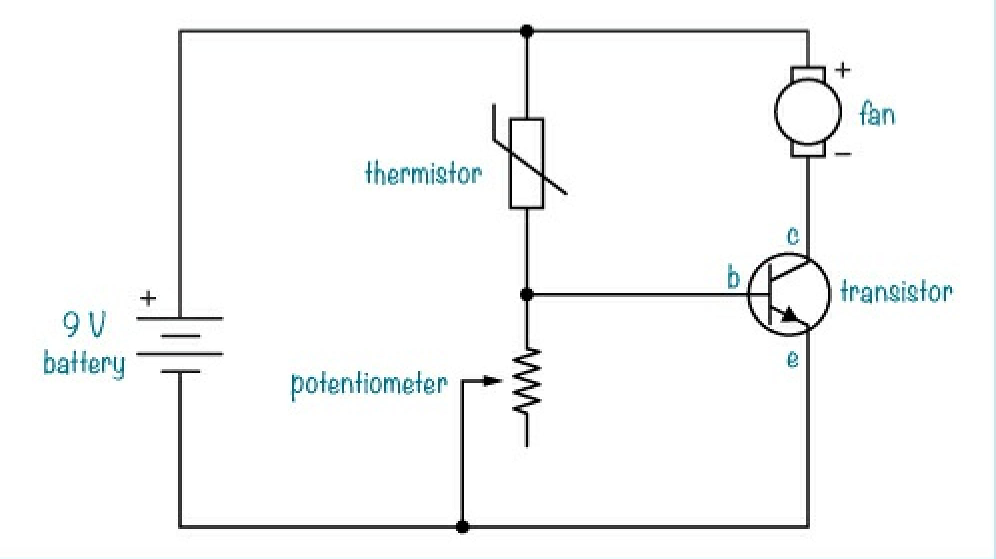

You'll need the following parts:

- **A standard 9 V battery** to power the circuit.
- **A 9 V battery clip** (Jameco #11280, Bitsbox #BAT033) to connect the battery to the circuit.
- **A circuit board** (Jameco #2191488, Bitsbox #HW005) with copper strips.
- **Insulated wire** (Jameco #36792, Bitsbox #W106BK), about 10 inches in length. Standard hookup wire works fine.
- **A 10 kΩ thermistor** (Jameco #207037, Bitsbox #ST021)
- **A 10 kΩ potentiometer** (Jameco #2118791, Bitsbox #VR004)
- **A transistor PN2222A** (Jameco #178511, Bitsbox #QD101)
- **A 12 V DC fan** (Jameco #1708465, Bitsbox #AF002)

For this circuit, I've specified a different transistor. Different transistors can handle different amounts of current. A fan often draws much more current than the LED, so I've listed a transistor that can handle more current than the one you used previously.

To test the circuit, first use an ice cube to cool down the thermistor and then warm it up with your fingers to see the fan come on.

WHAT'S NEXT?

In this chapter, you learned how the most important component in electronics—the transistor—works. You also learned about the potentiometer and the photoresistor, and you combined these to build a sunrise wake-up alarm.

Now you've met almost all of the most common components used in electronics! And you've gotten some practice with soldering. In the following chapters, you're going to build circuits on a breadboard again because it's much easier to fix mistakes and reuse your components. If you want to make a permanent version of any of the circuits, just refer back to Chapter 6 on how to solder.

In Chapter 8, you'll learn about *integrated circuits*. These are small components that contain circuits that have been shrunk down to fit on a tiny chip. These chips can be used to create the most amazing types of circuits. You'll see what I mean when you build your very own electronic instrument in Chapter 8!

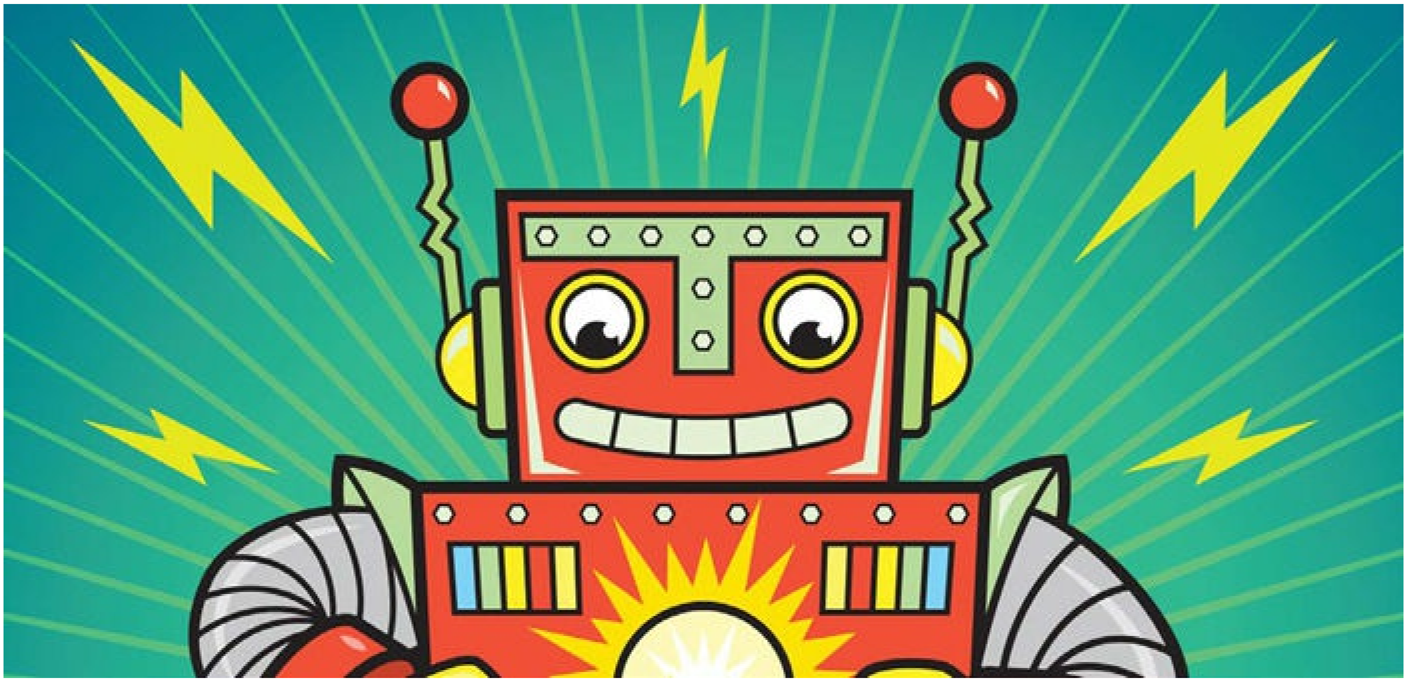

8

BUILDING A MUSICAL INSTRUMENT

This chapter will show you how to create sound with electronic components! Unlike other projects you've built so far, this chapter's projects use an *integrated circuit (IC)*, which is a whole circuit that's been shrunk down and packaged inside a tiny box. All sorts of circuits can be built as ICs, and most household devices are full of them. Peek inside a computer or an electronic toy, and you're sure to find a few.

To start, I'll explain what an IC is in more detail and describe how to figure out what a particular IC does. Then, you'll get some practice with ICs by building a simple circuit that makes some strange noises. At the end of the chapter, I'll show you how to build your very own electronic instrument that you can use to play music!

MEET THE INTEGRATED CIRCUIT

ICs are really small circuits that are conveniently packaged into little plastic boxes, which are often colored black. Any given IC has metal *pins* sticking out of it so you can connect other components to the circuit inside.

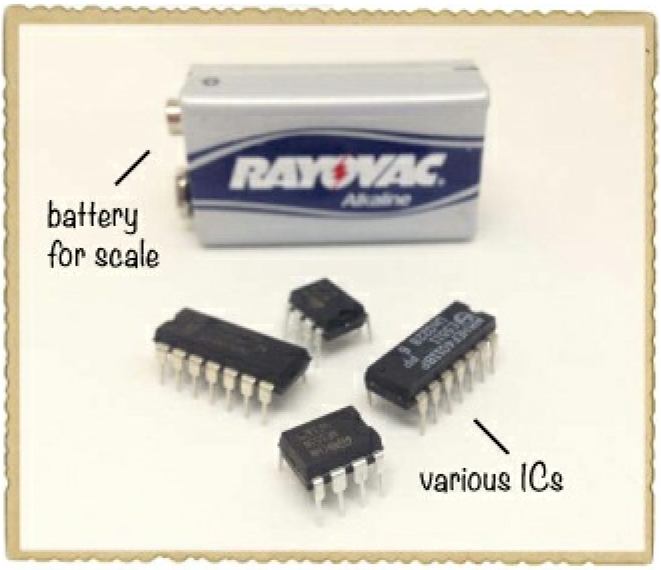

On the inside, an IC might contain a music player circuit, a radio circuit, a remote control circuit, or something else entirely. Because one little IC can contain a very complicated circuit, you can use ICs to make circuits with many different functions without building everything from scratch.

ICS AND DATASHEETS

Some ICs have only 8 pins, but other ICs have more than 100! To figure out what each pin on a specific IC does, you need to check that IC's datasheet, just as you did with the relay in "Project #11: Blink a Light!" on page 101. The datasheet tells you the function of each pin, and datasheets often show examples of how to use ICs in a circuit, too. To find a component's datasheet, try searching for the component name plus the term "datasheet" online, or check the website where you bought the component.

Datasheets often have tables with lots of numbers and technical terms, so the first time you see a datasheet for an IC, it might look very complex. But you usually don't need to read the whole datasheet. Instead, you can just look up the information you need and then go back to building your circuit.

HOW TO MAKE SOUND WITH ELECTRICITY

Sound is what you hear when air moves back and forth, or *vibrates*, really fast. Many devices that make noises, like the sound system in a car, do so with a *loudspeaker*, a component that vibrates air fast enough to make sound.

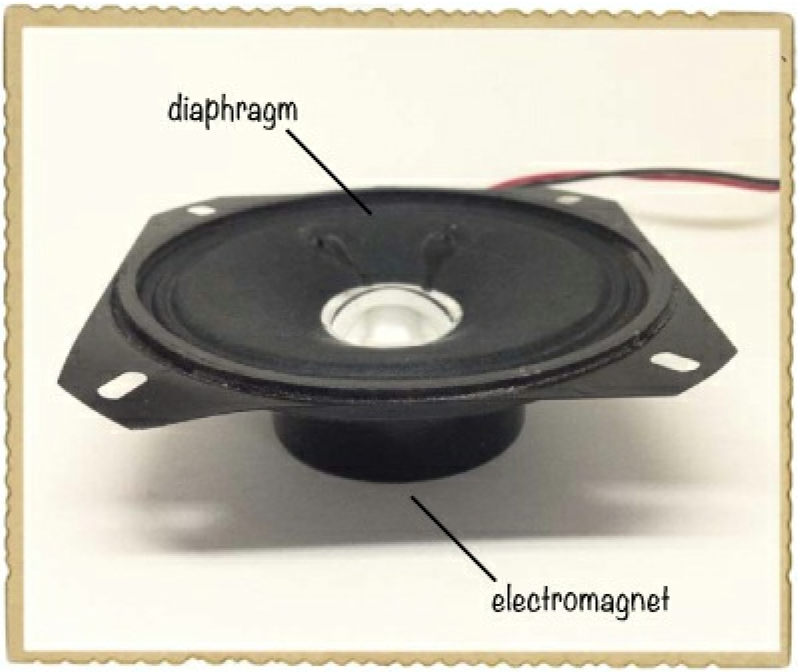

A loudspeaker has an electromagnet inside that moves a *diaphragm*, which is a surface that pushes the air in front of it. If a circuit turns the power to the loudspeaker's electromagnet on and off at, say, 1,000 times per second, then the diaphragm pushes the air back and forth 1,000 times per second. This is called the *frequency* of the sound, and frequency is measured in hertz (Hz). A frequency of 1,000 times per second, or 1,000 Hz, creates a steady beep.

SOUNDS THAT HUMANS CAN HEAR

Humans can hear sound only from around 20 Hz up to around 20,000 Hz. In "Project #11: Blink a Light!" on page 101, you built a circuit that blinked a light about once per second. If you had connected the output of the circuit to a speaker instead, the sound from the speaker would have a frequency of 1 Hz, which would be too low to hear. Instead, you would hear clicks from the speaker as it was switched on and off.

This means that to create sound you can hear, you need to create a circuit that can turn the voltage to a speaker on and off hundreds or thousands of times per second! Fortunately, there's an IC that can help you do just that.

MEET THE 555 TIMER

One classic IC is the *555 timer*, which you can use to switch things rapidly on and off. For example, you could use a 555 timer to blink a light every second, or you could connect it to a loudspeaker to make sound. Along with the 555 timer, you'd need to add a few extra resistors and capacitors, and by carefully selecting the values of those components, you could control how fast the light blinks or the sound's frequency. The 555 timer is very popular with hobbyists because it's cheap and pretty easy to use once you learn the basics. You'll use a 555 timer IC like the one shown here for the projects in this chapter.

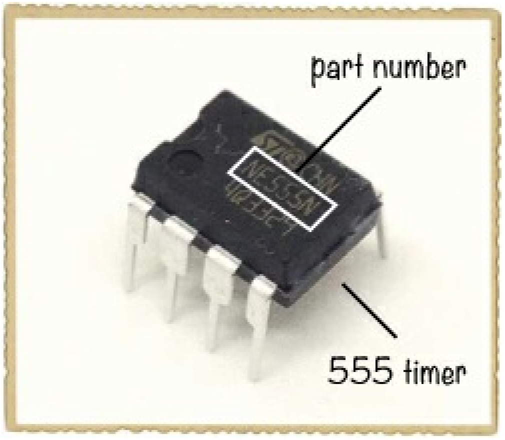

Powering an IC

The datasheet for the 555 timer should tell you what each pin does. Try searching online for "555 timer datasheet" now. When you find the datasheet and open it, you should see a *pinout* like this:

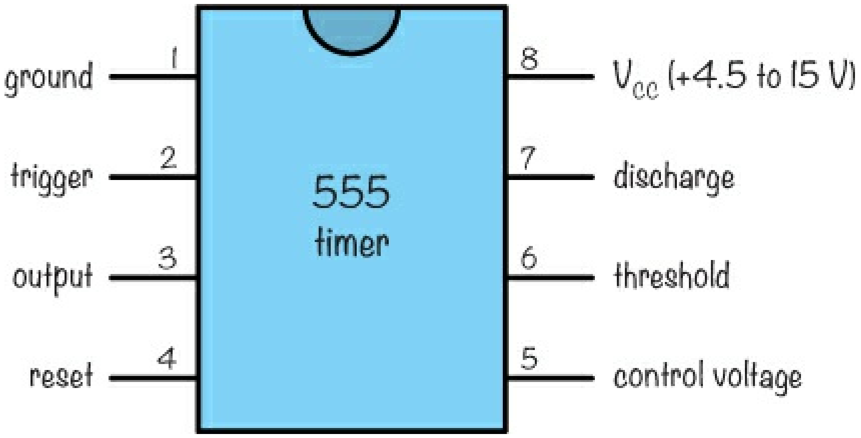

The 555 timer has eight pins. The pins of an IC are numbered counterclockwise, starting from the top left. (The top is indicated by a little notch or circle; some ICs have both.) The datasheet tells you the functions of each of the eight pins, but the following two pins are the most important to notice right now:

Pin 1, ground This pin must be connected to the negative terminal of your battery. Circuit diagrams that use ICs often label the negative battery terminal in a circuit schematic "ground," or GND for short.

Pin 8, V_{CC} This pin must be connected to the positive side of your battery, which must be at a voltage between 4.5 and 15 V. This means that a 9 V battery will work fine. On some ICs, this pin is called V_{DD} instead.

You'll find V_{CC} and ground pins on all ICs. They're the first pins to learn from any datasheet because you'll use them to power the circuit inside the IC.

The datasheet also shows how to connect the 555 timer in a circuit, and the circuit I find the most interesting connects the 555 timer in *astable mode*. *Astable* means that something is continually changing (not stable), and in astable mode, a 555 timer switches its output on and off constantly. That's perfect to blink a light or to create sound! Here's the circuit that tells the 555 timer to turn its output on and off:

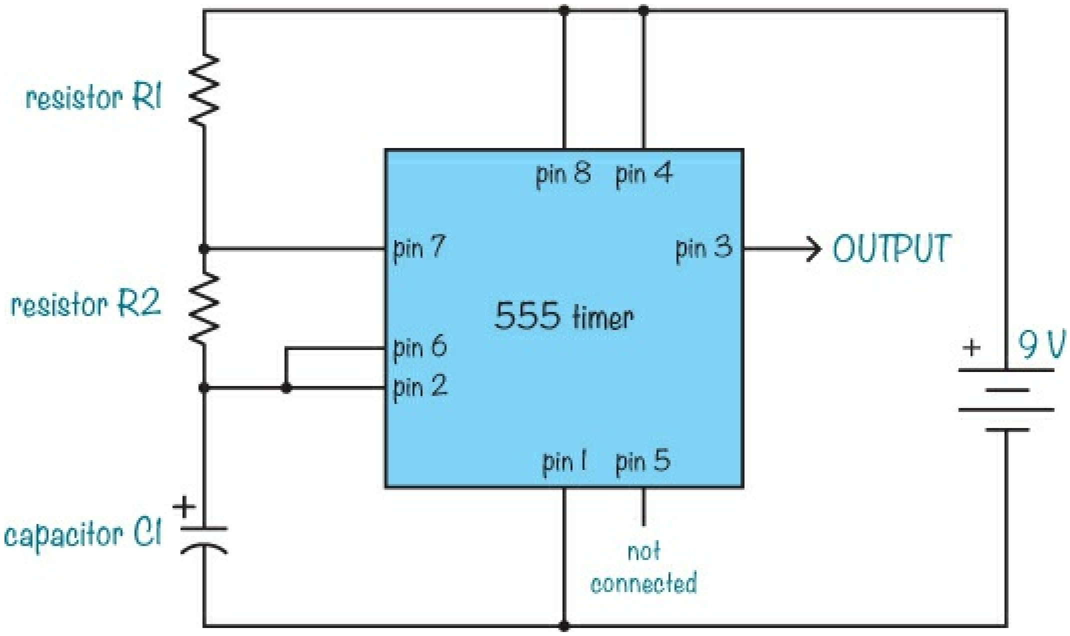

NOTE

The positions of the 555 pins have been arranged to suit the diagram; they're not in the same order as on the IC.

How to Set the Output Speed of the 555 Timer

The values of R1, R2, and C1 in the circuit diagram determine how long the output pin stays *high*—meaning it outputs a voltage close to the battery voltage—and how long it stays *low*—meaning it outputs a voltage close to 0 V. The speed of the change in output, or the frequency, is the number of times the output goes from high to low in one second. When the 555 timer is connected in astable mode, as it is in the circuit diagram, the frequency of the output is controlled by resistors R1 and R2 and capacitor C1 according to the following formula:

$$\text{Frequency} = \frac{1.44}{(\text{R1} + \text{R2} + \text{R2}) \times \text{C1}}$$

Here, the resistor values are in ohms, and the capacitance is in farads. To find the frequency, first replace R1, R2, and C1 in the formula with the values of these parts and then punch the stuff on the right side of the equal sign into a calculator.

Let's try an example. Imagine a circuit with the following component values:

- R1 = 100 kΩ
- R2 = 10 kΩ
- C1 = 10 nF

What is the frequency of the output? Enter these values into the formula:

$$\text{Frequency} = \frac{1.44}{(100\ \text{k}\Omega + 10\ \text{k}\Omega + 10\ \text{k}\Omega) \times 10\ \text{nF}}$$

$$\text{Frequency} = \frac{1.44}{(120\ \text{k}\Omega) \times 10\ \text{nF}}$$

Now convert the units so the values are easier to multiply (120 kΩ = 120,000 Ω and 10 nF = 0.00000001 F):

$$\text{Frequency} = \frac{1.44}{120{,}000\ \Omega \times 0.00000001\ \text{F}}$$

$$\text{Frequency} = 1{,}200\ \text{Hz}$$

According to this calculation, the output should turn on and off 1,200 times per second with those values.

PROJECT #16: MAKE YOUR OWN SOUND WITH THE 555 TIMER

This project will show you how to play a sound with a frequency of about 1,200 Hz through a speaker. That's pretty cool! But a 1,200 Hz sound isn't very pleasant to listen to, and if you have any pets, they might appreciate the sound even less than you. In fact, when I turned on this circuit at my parents' home, their dog came running to me looking very confused and a bit scared. I turned it off quickly and moved my experiments to a place with no pets around, and I suggest not building this project around your pets, either.

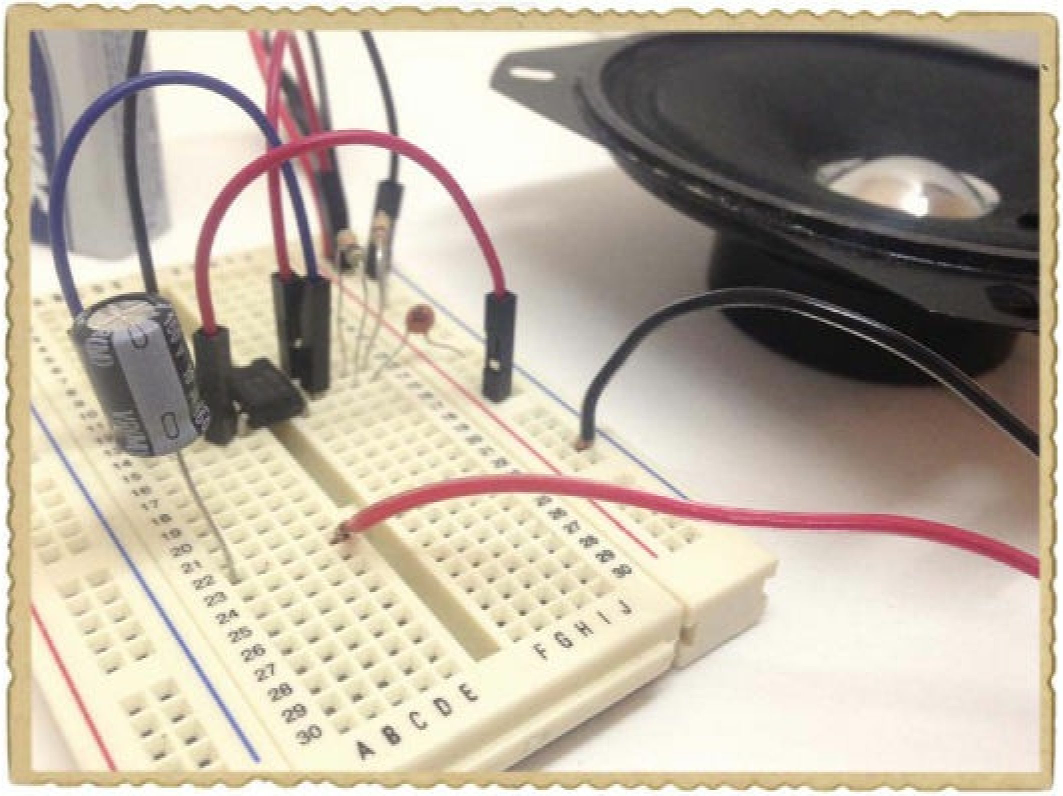

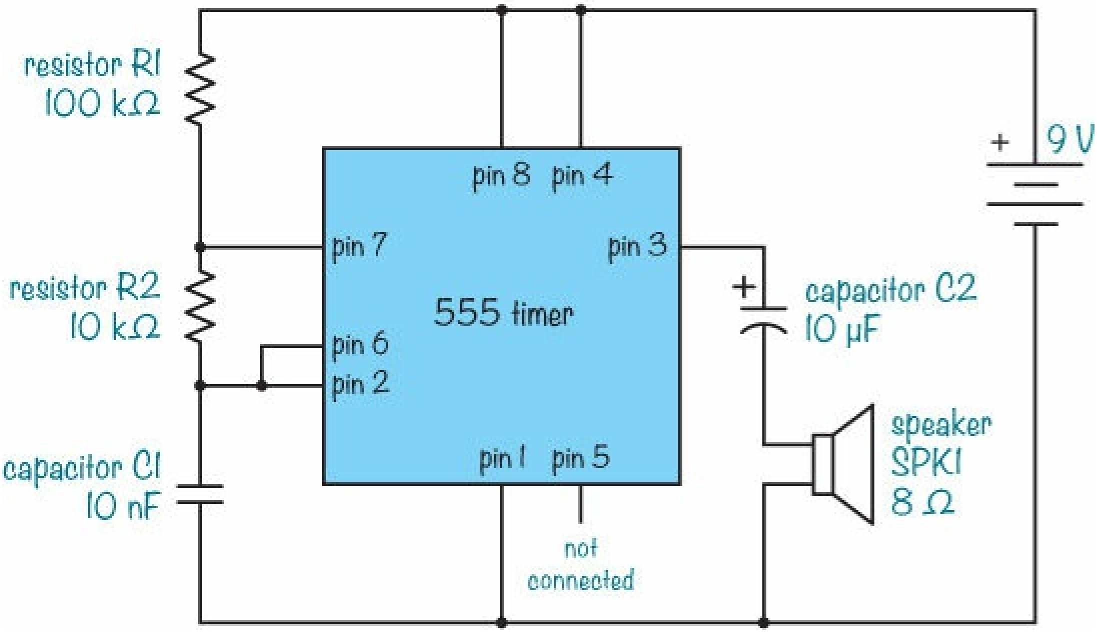
resistor R1
100 kΩ
resistor R2
10 kΩ
capacitor C1
10 nF
pin 8
pin 4
pin 7
pin 3
555 timer
pin 6
pin 2
pin 1
pin 5
not connected
+
capacitor C2
10 µF
speaker
SPK1
8 Ω
+
9 V

Shopping List

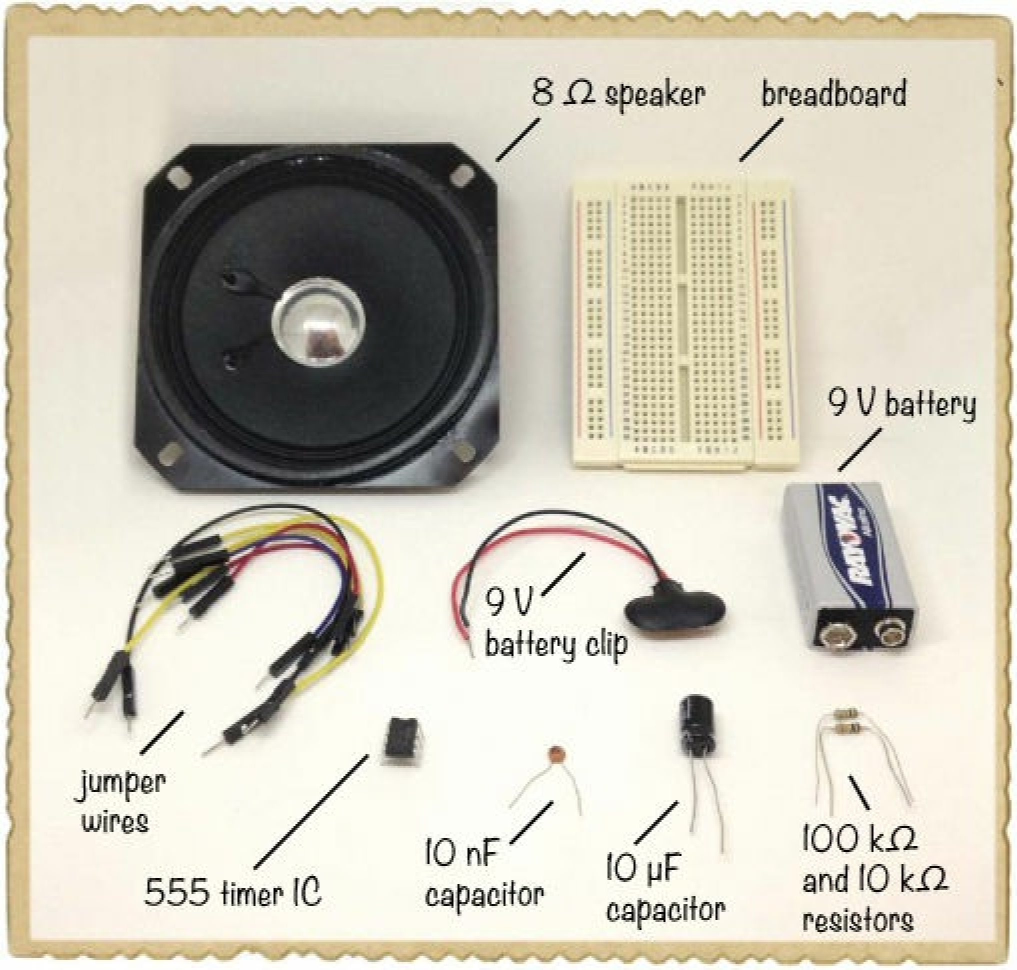
8 Ω speaker
breadboard
9 V battery
9 V battery clip
jumper wires
555 timer IC
10 nF capacitor
10 µF capacitor
100 kΩ and 10 kΩ resistors

- **A standard 9 V battery** to power the circuit.
- **A 9 V battery clip** (Jameco #11280, Bitsbox #BAT033) to connect the battery to the circuit.
- **A breadboard** (Jameco #20601, Bitsbox #CN329) with at least 30 rows.
- **Breadboard jumper wires** (Jameco #2237044, Bitsbox #CN236) for making easy connections. (Standard hookup wire works, too.)
- **A 555 timer IC** (Jameco #904085, Bitsbox #QU001) to create the timing.
- **A 8 Ω speaker** (Jameco #1954818, Bitsbox #ST063) to play the sound.
- **A 10 µF capacitor** (Jameco #29891, Bitsbox #EC10U25) to connect to the speaker.
- **A 10 nF capacitor** (Jameco #15229, Bitsbox #CC10N) to help set the frequency of the sound.
- **A 100 kΩ resistor** (Jameco #691340, Bitsbox #CR25100K) to help set the frequency of the sound.
- **A 10 kΩ resistor** (Jameco #691104, Bitsbox #CR2510K) to help set the frequency of the sound.

Step 1: Place the 555 Timer on the Breadboard

This circuit is built around the 555 timer IC, so first, place that IC in the middle of the breadboard, making it easy to connect all the components around it.

NOTE

When connecting an IC to a breadboard, always place it over the middle notch, with one set of pins on the left side and the other on the right side. Otherwise, any pins that share a row on the breadboard will be connected to each other.

Place the IC's notch marker toward the top of the board so that pin 1 is at the upper-left corner and pin 8 is at the upper-right corner, as shown. Triple-check that you've oriented the chip as I describe; otherwise, the rest of the instructions in this project won't work.

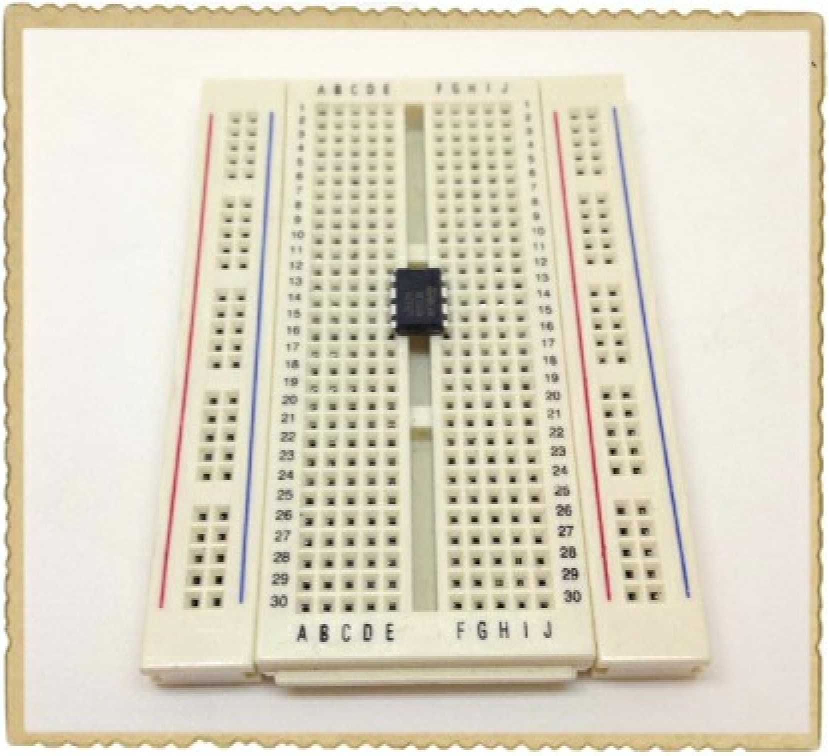

Step 2: Set the Frequency

Next, connect the resistors and capacitor that set the frequency: R1, R2, and C1. The C1 capacitor isn't polarized, so it doesn't matter which way you connect it. Use the two vertical columns on the right for positive and negative: the red column for positive and the other for negative.

Connect R1, the 100 kΩ resistor, from pin 7 of the 555 timer to the positive column. Connect R2, the 10 kΩ resistor, from pin 6 to pin 7. And connect C1, the 10 nF capacitor, from pin 6 to the negative column. Then, connect a jumper wire from pin 2 to pin 6 to complete the rest of the connections that are needed to set the frequency.

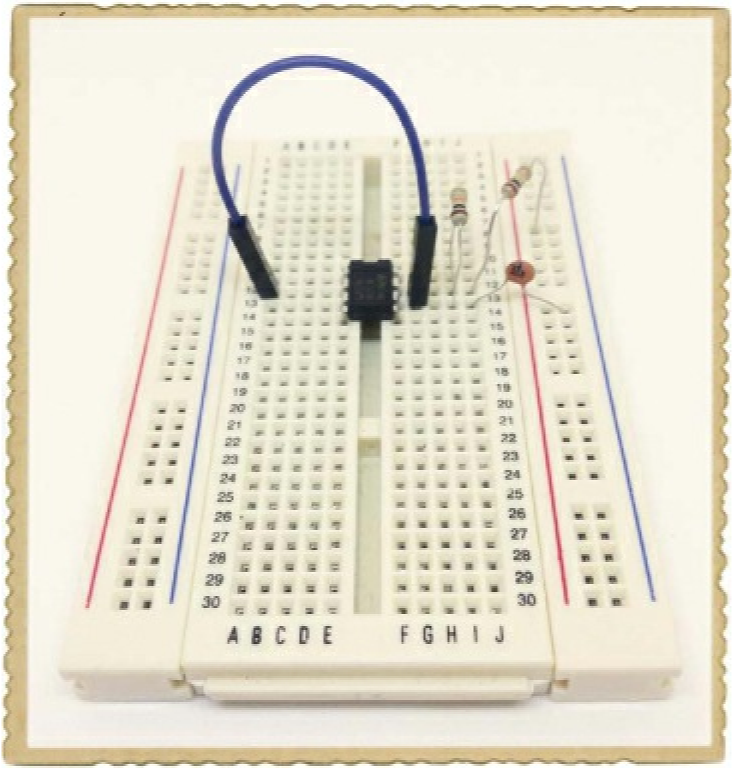

Step 3: Connect the Speaker and Coupling Capacitor

All the components to make the 555 timer output pin 3 turn on and off around 1,000 times per second should be in place on your breadboard now. If you were to connect the speaker directly to the output, a large current would flow through the speaker, possibly damaging both the speaker and the 555 IC. You could connect the speaker through a resistor to reduce the current, but using a capacitor is even better. When a capacitor is connected to an AC voltage, the capacitor acts a bit like a resistor, but if you connect it to a constant DC voltage, the current will be blocked. This means that there won't be any current running through the speaker unless there's actually a frequency to be played. When a capacitor is used like this, it's called a *coupling capacitor*.

The coupling capacitor for this project is a 10 µF polarized capacitor, so first identify which leg is negative. Connect the positive capacitor leg to the 555 timer's output pin 3. Then connect the capacitor's negative leg to an empty row on your breadboard.

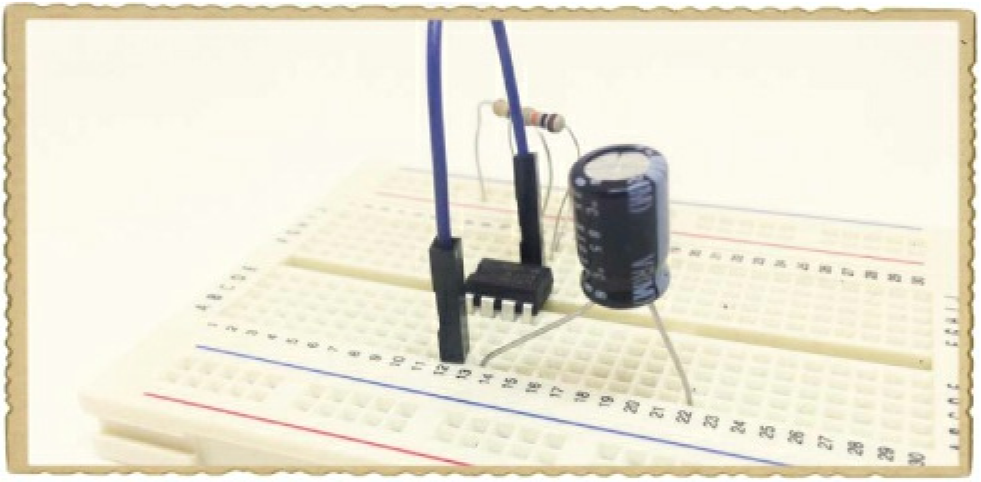

Next, if your speaker comes without wires, solder a wire about 6 inches long to each of the two contacts on the back of the speaker (ignore any + or – labels by the speaker contacts). Then, connect one of the speaker wires to the same row as the negative capacitor leg and the other wire to the negative supply column.

MOD.
BOC-400
STEREN
7 W
Bocina
8 Ohms

Step 4: Connect the Power and Reset Pins

Compare your breadboard to the circuit diagram at the beginning of the project, and you'll see that there are a few connections missing. Once you have all of your components in place on the breadboard, you can make those final connections with jumper wires. Add a jumper wire for each of the following connections:

- From pin 1 on the 555 timer to the negative supply column
- From pin 8 on the 555 timer to the positive supply column
- From pin 4 on the 555 timer to the positive supply column

Note that there's no connection to pin 5 of the 555 timer in this circuit.

Step 5: Make Some Sound!

Connect the battery clip to your supply columns on the right side of the breadboard. The red wire goes to the positive supply column, and the black wire goes to the negative supply column.

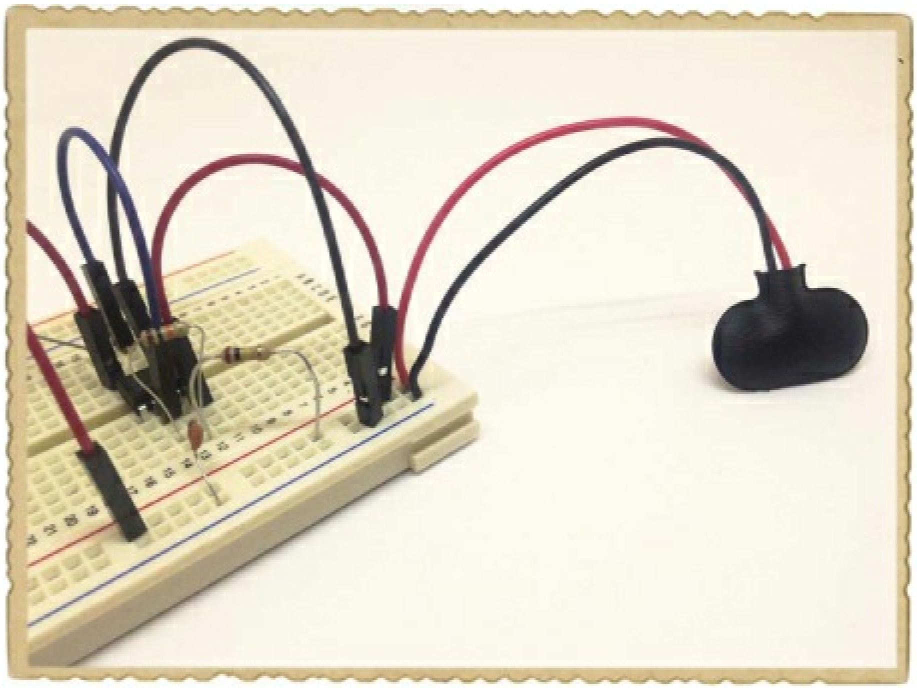

When you're ready, connect the battery, and you should hear a loud beep. Congratulations: You just made your first electronic sound!

Step 6: What If There's No Sound?

There are a lot of connections in this circuit, so if your circuit doesn't work on your first attempt, don't worry: that happens to everyone who plays with electronics at some point. First, disconnect the battery, and then check the resistor and capacitor legs. These legs are long, and they can easily end up in contact with each other by accident, creating a short circuit. (For example, if you accidentally connect the positive terminal of the battery directly to the negative terminal, that short-circuits the battery.)

Next, check the component connections to the 555 timer. The 555 timer's pins must connect to the rest of the circuit according to the circuit diagram, or the IC won't work. When there are a lot of connections, it's easy to plug a wire into the wrong row of the breadboard.

CHECKING CONNECTIONS AS A TEAM

Finding circuit problems is called *debugging*, and it's easier to do with some help. When you get stuck, ask someone else to look at the circuit diagram and say the connections out loud one by one while you check the real connections. For example, if your friend is reading the schematic and you're looking at the breadboard, you might have a conversation like this:

> Friend: "The positive side of the battery is connected to one side of R1."
>
> You: "Got it!"
>
> Friend: "The positive side of the battery is also connected to pins 4 and 8 of the IC."
>
> You: "Got it!"
>
> Friend: "The other side of R1 is connected to pin 7 of the chip and to one side of the resistor R2."

You: “Oh, wait! I don’t have the connection to pin 7!”

And just like that, you’ll discover the problem.

If all your components appear to be oriented correctly and you see no short circuits, then check all your breadboard connections to make sure components that should be connected share a row. Start with the connection from the positive terminal of the battery on your circuit diagram. Is it connected on the board just as the circuit diagram shows? If yes, then move on to the next connection; keep going like this until you’ve checked all the connections.

TURNING AN ANNOYING BEEP INTO MUSIC

The sound you created in the previous project isn’t very pleasant. So how can you turn it into music? Musical notes are just sound waves that vibrate at specific frequencies, and that means it’s possible to make an electronic instrument with the 555 timer. If you change the frequency of the signal that goes into the speaker, then the tone of the sound will change. You just have to find a way to change the 555 timer’s output frequency at will, without rebuilding the circuit every time.

Chapter 7 introduced two components that can change their resistance value: the potentiometer and the photoresistor. If you use one of those components to control the frequency of the 555 timer’s output signal, then when that component’s resistance changes, the sound will change, too. That’s how you’ll make the instrument in the next project.

PROJECT #17: AN INSTRUMENT THAT BEEPS AND BOOPS

This project shows you how to combine what you’ve learned so far to build your very own electronic instrument. Specifically, you’re going to build an instrument with a button to play sound and a potentiometer shaft to change the tone.

This instrument is like a very simple *synthesizer* that uses electricity to make sounds. Synthesizers have been used to add all kinds of bleeps, bloops, and glitchy sounds to songs since the beginning of the electronic music genre. This one is pretty basic, but it still makes plenty of fun noises. In fact, the circuit for this project looks very similar to the circuit from “Project #16: Make Your Own Sound with the 555 Timer” on page 167, but it has a few twists.

This circuit replaces the separate R1 and R2, which the circuits in Project #16 used, with a potentiometer (and a protective 1 kΩ resistor). The potentiometer acts like two resistors, meaning that you can effectively change the values of R1 and R2 whenever you want to change the tone you hear.

Notice that this circuit also has a switch connected to the positive battery terminal. This switch is a push button, and with a switch between power and your circuit, the instrument should make a sound only when you push the button.

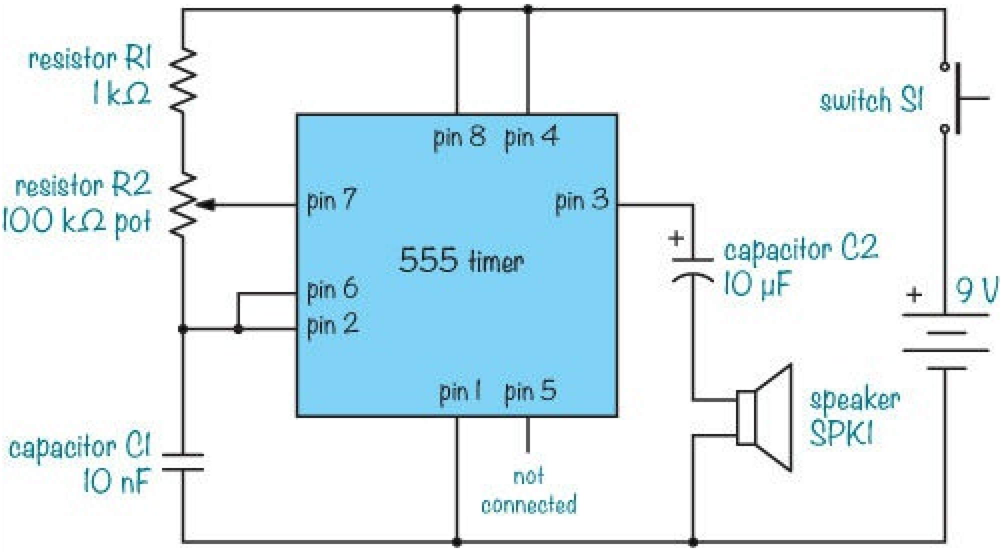
resistor R1
1 kΩ
resistor R2
100 kΩ pot
capacitor C1
10 nF
pin 8
pin 4
pin 7
pin 3
555 timer
pin 6
pin 2
pin 1
pin 5
not connected
switch S1
capacitor C2
10 μF
+
9 V
+
speaker
SPK1

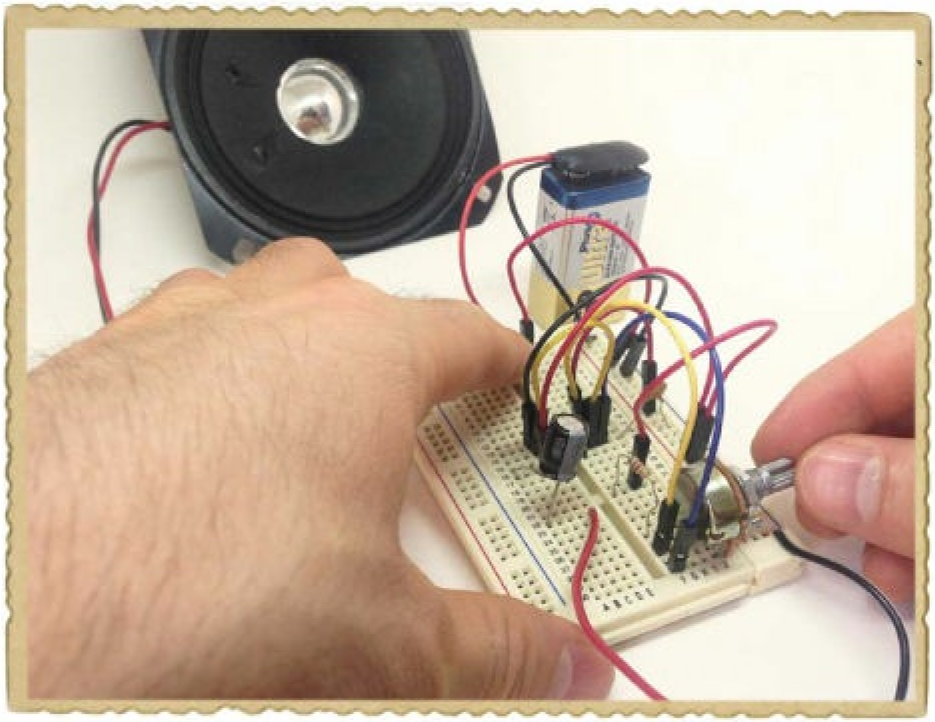

Shopping List

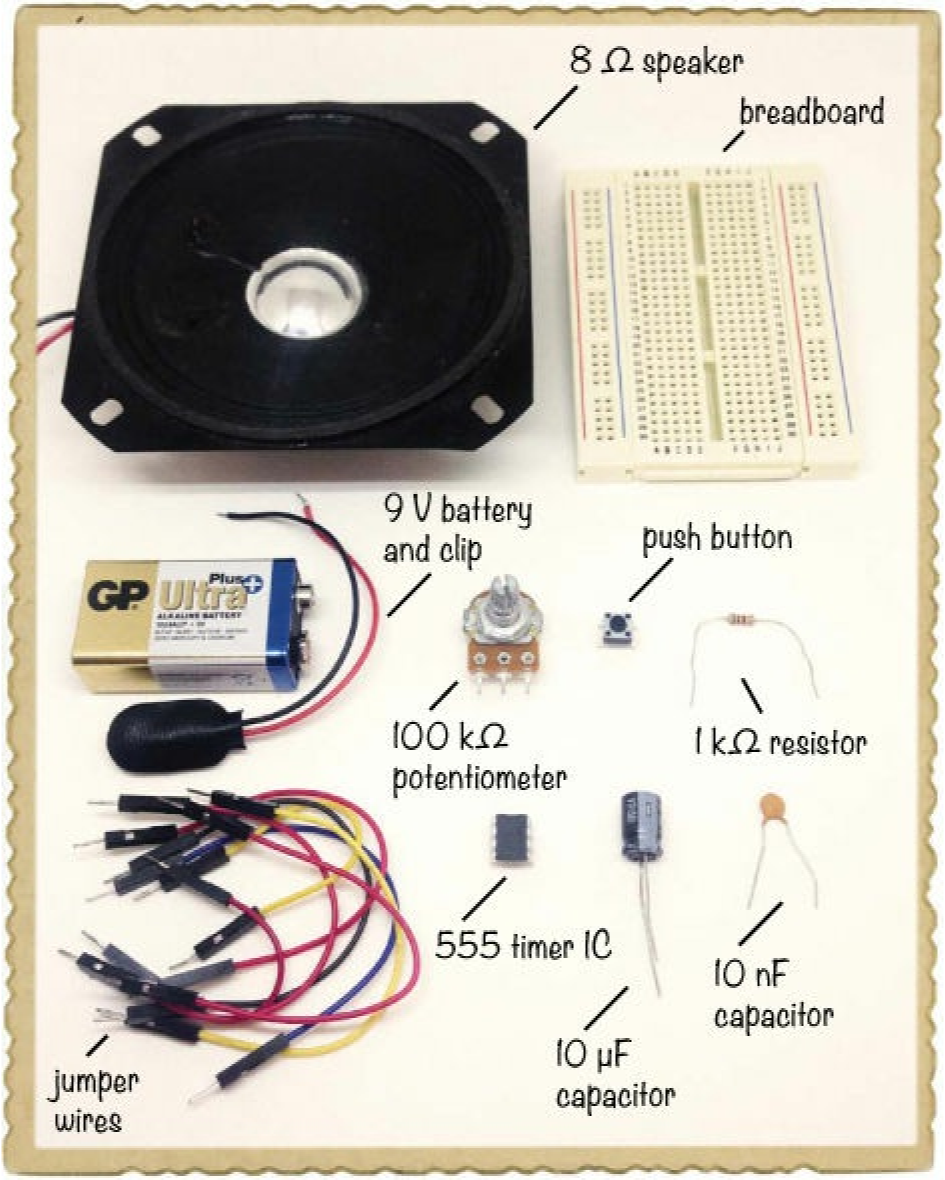
8 Ω speaker
breadboard
9 V battery and clip
push button
100 kΩ potentiometer
1 kΩ resistor
555 timer IC
10 nF capacitor
10 µF capacitor
jumper wires

- **A standard 9 V battery** to power the circuit.
- **A 9 V battery clip** (Jameco #11280, Bitsbox #BAT033) to connect the battery to the circuit.
- **A breadboard** (Jameco #20601, Bitsbox #CN329) with at least 30 rows.
- **Breadboard jumper wires** (Jameco #2237044, Bitsbox #CN236) for making easy connections. (Standard hookup wire works, too.)
- **A 555 timer IC** (Jameco #904085, Bitsbox #QU001) to create the timing.
- **A 8 Ω speaker** (Jameco #1954818, Bitsbox #ST063) to play the sound.
- **A 10 µF capacitor** (Jameco #29891, Bitsbox #EC10U25) to connect to the speaker.
- **A 10 nF capacitor** (Jameco #15229, Bitsbox #CC10N) to help set the frequency of the sound.
- **A 100 kΩ potentiometer** (Jameco #2161406, Bitsbox #VR006) to control the tone.
- **A 1 kΩ resistor** (Jameco #690865, Bitsbox #CR251K) to protect pin 7 from being connected directly to V_{CC}.
- **A push button** (Jameco #119011, Bitsbox #SW087) to play tones with.

Step 1: Connect the 555 Timer and the Capacitors

Start by connecting the 555 timer in the middle of the board. Then, connect both capacitors. Connect C1, the 10 nF capacitor, from pin 6 to the negative power column. And connect C2, the 10µF capacitor, with its positive leg at pin 3 of the chip and the negative leg on an empty row farther down. When you're done, your board should look something like this:

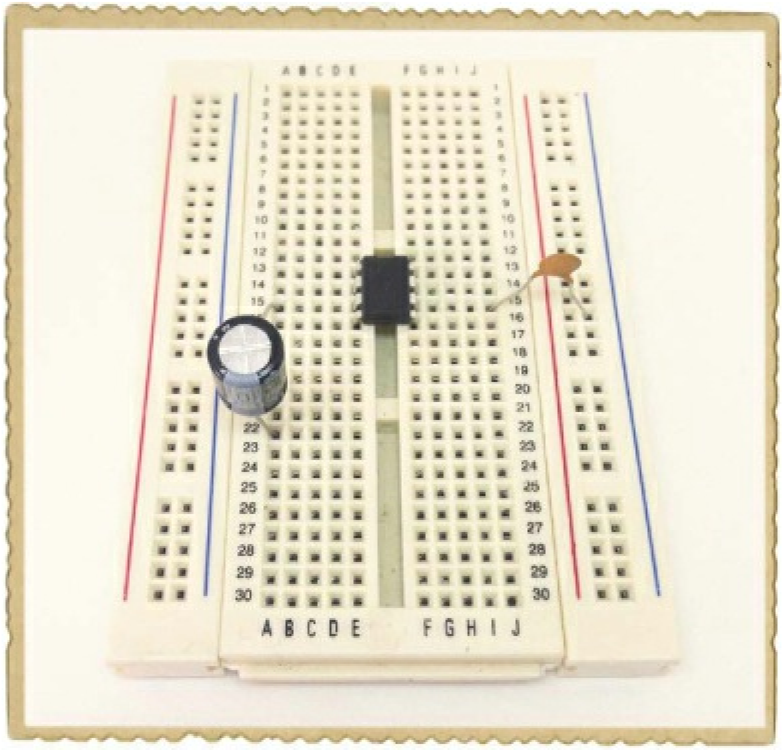

Step 2: Connect Jumper Wires

Use jumper wires to create the following connections:

- Pin 8 to the breadboard's positive supply column
- Pin 4 to the breadboard's positive supply column
- Pin 1 to the breadboard's negative supply column
- Pin 2 to pin 6

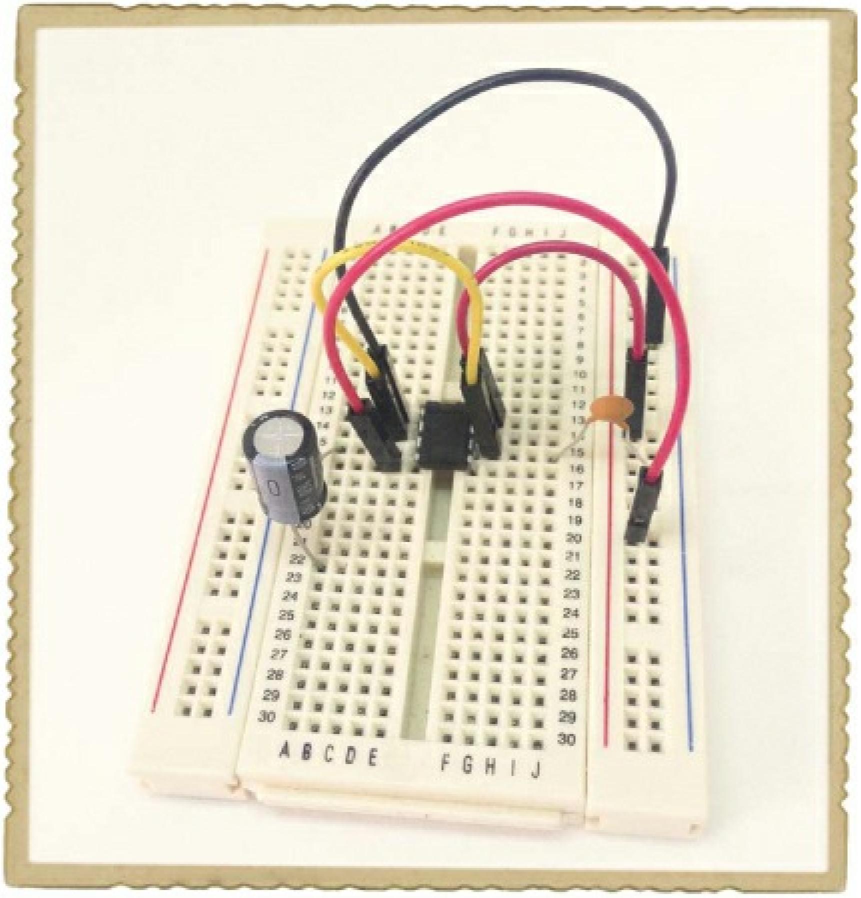

Step 3: Connect the Note Controller and Resistor

Next up is the potentiometer, which controls the note that plays. Because it takes up more space, plug this in at the bottom of the breadboard, as shown, and be sure each leg is in its own row.

Connect the 1 kΩ resistor from the uppermost leg of the potentiometer to an unused row away from the 555 IC and connect a jumper wire from this row to the positive supply. Then, connect one jumper wire to each of the two remaining potentiometer legs. Connect the wire on the middle potentiometer leg to pin 7 on the 555 timer, and connect the wire on the lower leg to pin 6 on the 555 timer.

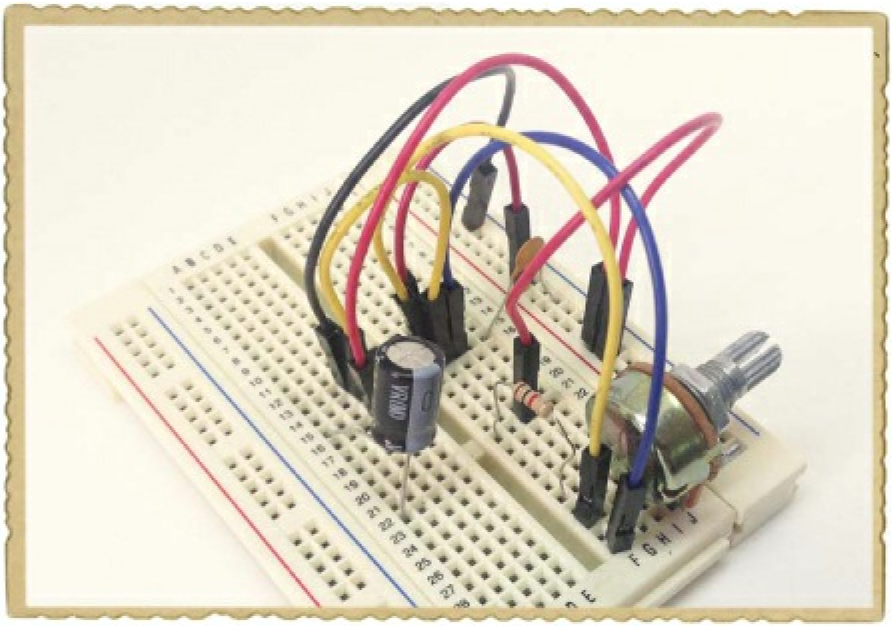

Step 4: Add the “On” Button

You’re almost done! Instead of wiring the battery clip directly to the positive supply column on the breadboard, you’re going to connect it through a push button, which will act as an “on” button for your instrument. This way, the circuit will get power from the battery only when the button is pushed and, therefore, make sound only when you push the button.

The push button in the Shopping List (page 177) has four pins, but the two legs in the front are connected and the two pins in the back are connected. That means this button works exactly like a two-pin switch. When you push the button, you connect the front pair of pins with the back pair. Your circuit will get power and you’ll hear the sound. When you release the button, the circuit loses power again and the sound will stop.

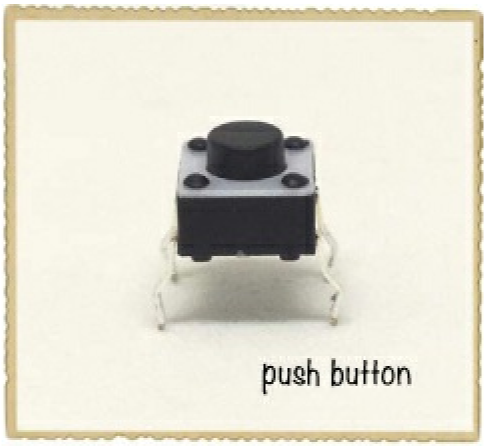

Plug the push button into your breadboard over the notch in the middle. Then, use a jumper wire to connect one side of the push button to the breadboard’s positive supply column. Next, connect the positive wire from your battery clip to the other side of the push button. Finally, connect the negative battery clip wire to the negative supply column.

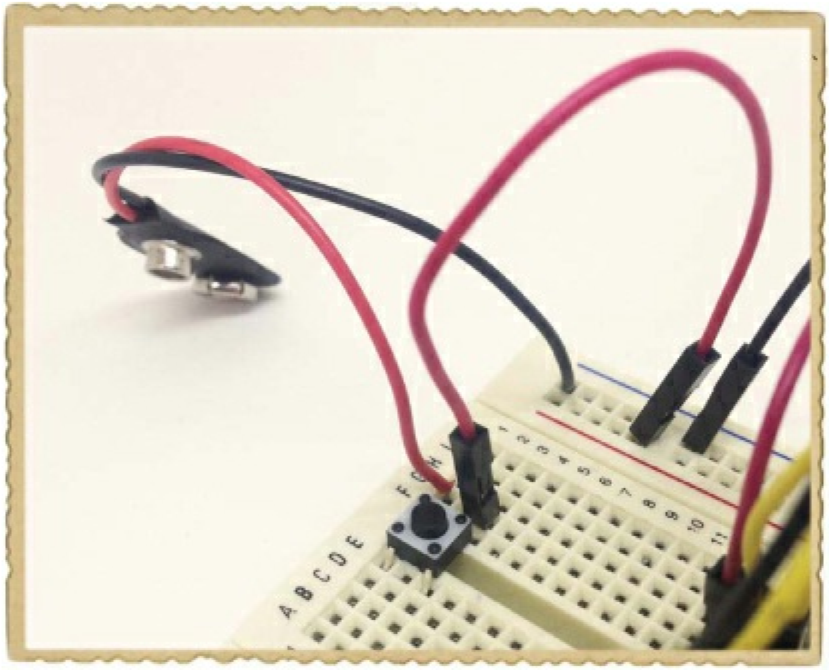

Step 5: Add the Loudspeaker

Connect one of the speaker wires to the same row as the negative capacitor leg and the other wire to the negative supply column.

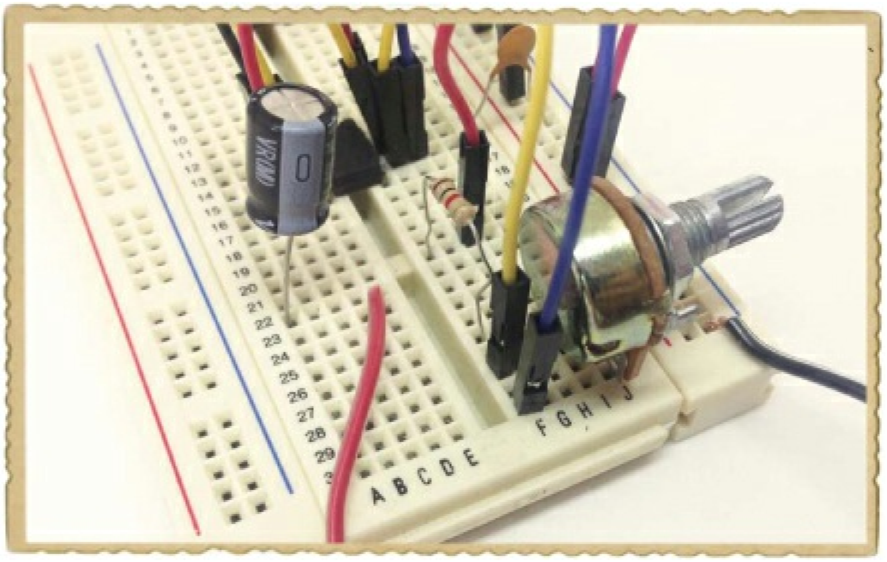

Step 6: Play Some Music!

Connect your battery to the battery clip and push the button. Did you hear a sound? Try turning the potentiometer spindle back and forth until you hear something.

You can make a beat by pushing the button in any rhythm you like, and you can control the tone by turning the shaft of the potentiometer in between. Beep, baap, booop!

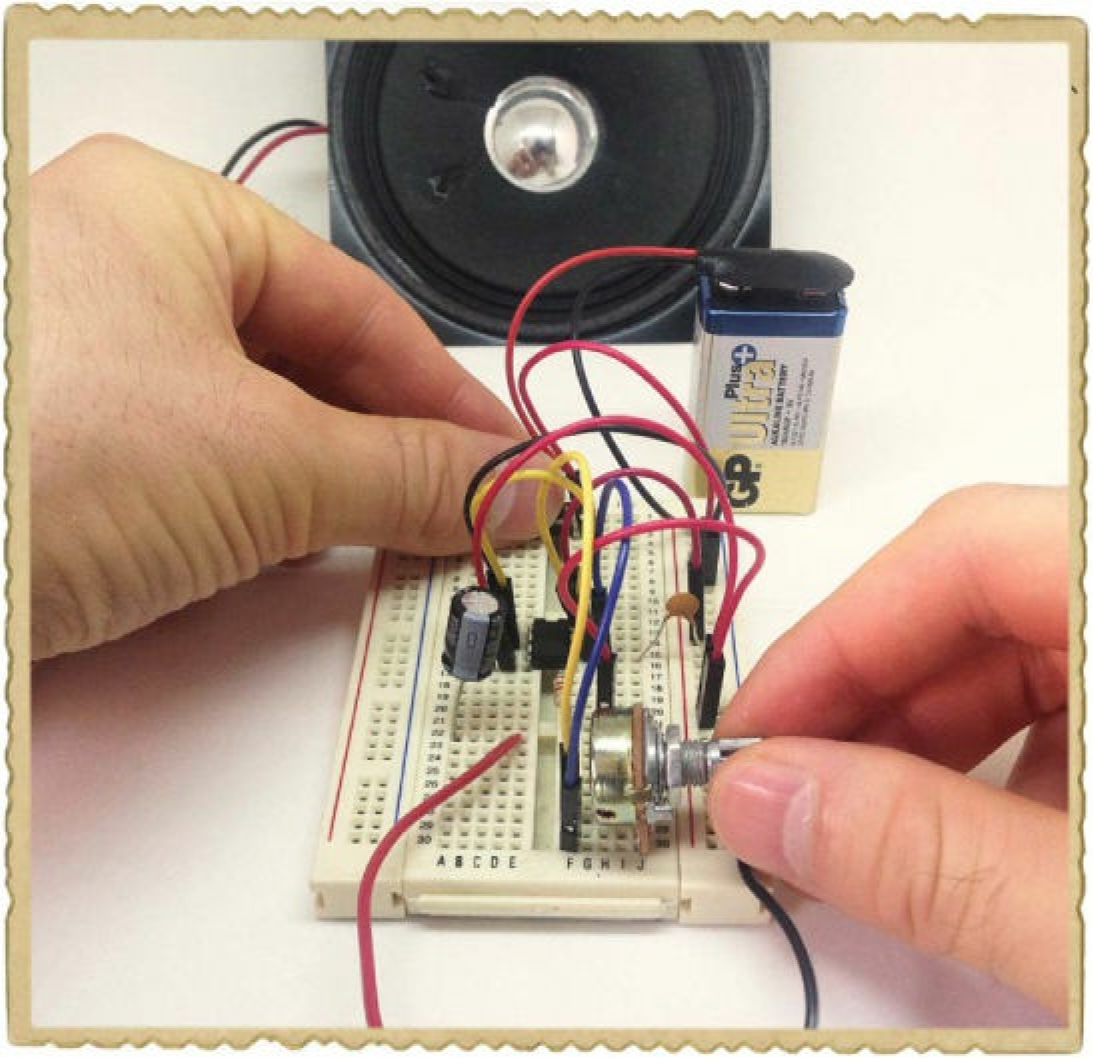

Step 7: What If the Instrument Doesn’t Work?

Some potentiometers aren't perfectly fitted for breadboards, so start by checking whether your potentiometer is properly connected to the breadboard. Next, check that the two capacitors and the push button are connected according to the circuit diagram at the beginning of the project.

There are a lot of wires on this board, so it's easy to mix up connections. If you still aren't able to hear any sound, then go through all the connections on your board and check that they're correct. I suggest using the process I described in "Checking Connections as a Team" on page 174.

TRY IT OUT: MAKE A MOTION-CONTROLLED INSTRUMENT

Instead of replacing both R1 and R2 with a potentiometer, try using a 1 kΩ resistor for R1 and a photoresistor for R2, as shown in this circuit diagram. As you move your hand over the photoresistor, the amount of light hitting it should change, and the instrument should play different notes!

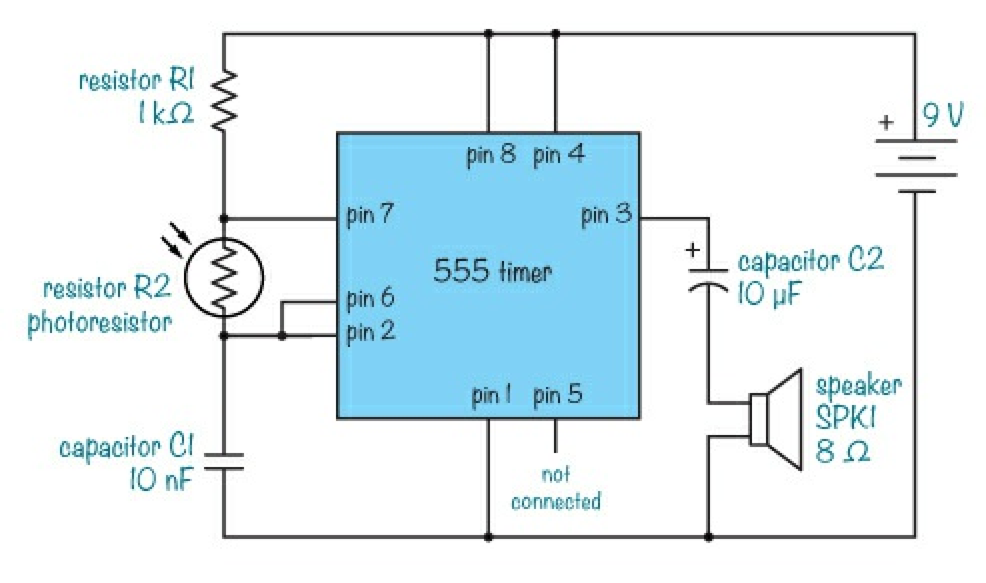

WHAT'S NEXT?

There are millions of integrated circuits in the world, and they do all kinds of awesome things. Remember, whenever you want to use an IC in your projects, you can use its datasheet to find out both what the IC does and what kind of circuit you need to make it work.

In this chapter, you learned how to use a 555 timer IC to make music. These circuits had lots of connections, so if you didn't get them right the first time, then you're in good company. Most people mix up wires the first time they connect a circuit, and fixing the connections is great practice! And speaking of practice, now would be a good time to do some more soldering. Why not solder your instrument on to a prototyping board—and maybe even put it in a box?

Integrated circuits are very commonly used in *digital circuits*. Digital circuits can do many things simply by switching the right parts on and off, often very rapidly, just as in the musical instrument you made in this chapter. Modern gadgets like your computer use digital circuits, and you’ll learn more about digital electronics in the last few chapters of this book.

PART 3

THE DIGITAL WORLD

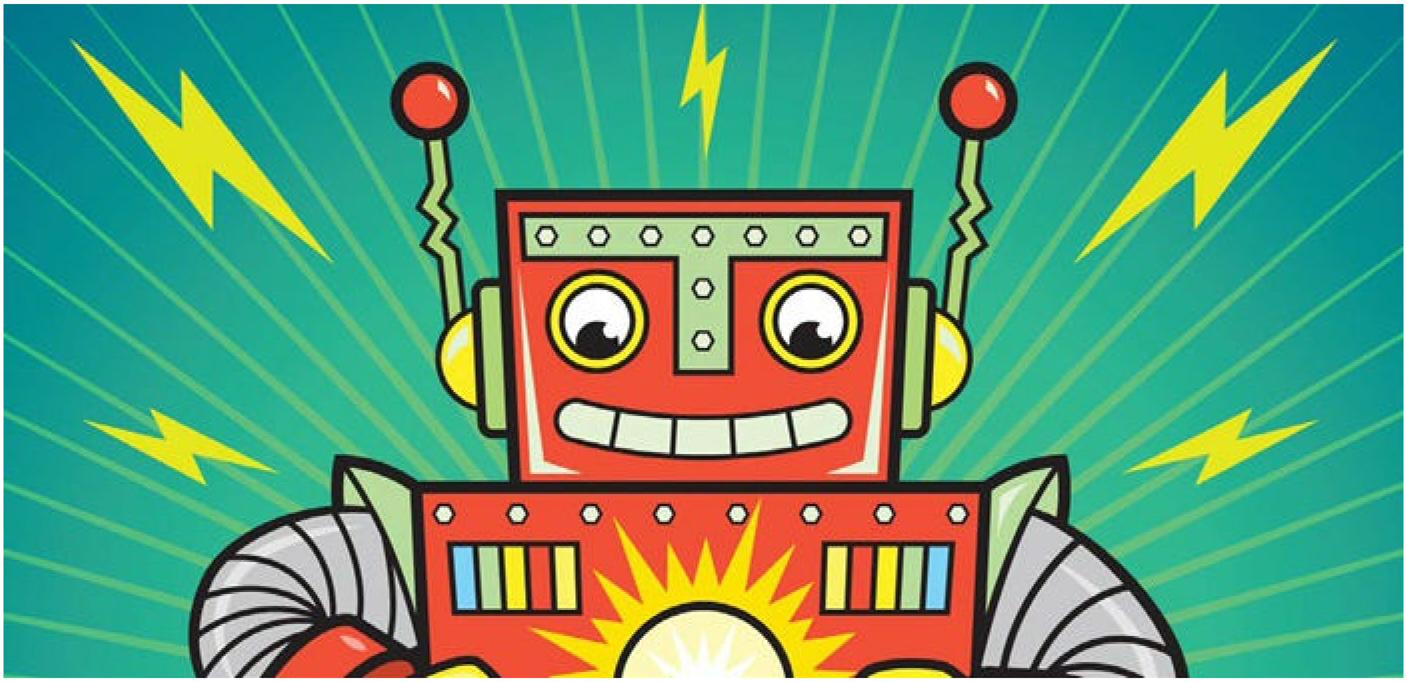

9

HOW CIRCUITS UNDERSTAND ONES AND ZEROS

Mobile phones, computers, TVs, video game consoles, and almost all the other technology you see around you is built with digital electronics. If you've ever looked inside a computer, the circuits might seem complex, but when you break that hardware down into small pieces, it's actually quite simple: every action a computer takes is based on whether some circuit sees a high or low voltage.

You've used transistors to make circuits that control things, and in this chapter, you'll learn the basics of using digital electronics to make even smarter circuits. You're going to learn a new number system, what bits and bytes are, and how you can use bits and bytes to create messages with electricity.

ONES AND ZEROS AS VOLTAGES

When I was in school, my teacher told me that computers used ones and zeros to communicate, but he didn't explain it any further. I wanted to know more, so I decided to ask my dad. Because computers are complex machines, I assumed the way they use ones and zeros was going to be complex, too, but my dad told me, "In a computer, a one is just a wire with voltage, and a zero is a wire without voltage." (By "without voltage," my dad meant the voltage was zero.)

In "How to Set the Output Speed of the 555 Timer" on page 166, I mentioned briefly that the pins of the 555 timer could be either high or low. The pins on the ICs in a computer work the same way: high is one, and low is zero. Digital circuits work by flipping the voltages on different wires between high and low.

MEET THE BINARY NUMBER SYSTEM

Normally, when you talk about numbers, you use *decimal numbers*, which are in the *base-10* number system. In base-10 numbers, a digit can have one of 10 possible values, from 0 to 9. But digital circuits can work with only two voltages (high and low), so they can understand only *binary numbers*. Binary is also called the *base-2* number system, and a binary number's digits have only two possible values: 0 and 1.

Here's how you'd count to 10 in both decimal and binary:

Decimal number	Binary number
0	0
1	1

2	10
3	11
4	100
5	101
6	110
7	111
8	1000
9	1001
10	1010

What's going on here? The number 2 in decimal is 10 in binary, and 4 in decimal is 100 in binary!

You can figure out the value of any number in any system using the digits and their position. The rightmost digit always gets multiplied by 1. Going left from there, each position's value equals the base times the value of the previous position. In decimal, the base is 10, so the second position's value is 10, or 10 × 1; the third position's value is 100, or 10 × 10; the fourth position's value is 1000, or 10 × 100; and so on. Here's an example:

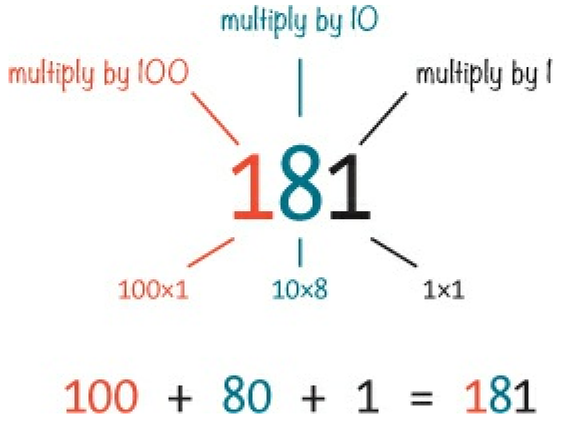

In the number 181, the first digit is 1, the second digit is 8, and the third digit is 1. Once you multiply those digits by their position values, you can add the resulting values to get 181. You don't need to use this method to figure out decimal numbers, though. It's pretty clear that 181 equals 181! But this method is very useful for finding the value of binary numbers. Binary follows the same rules, but it uses a base of 2 instead.

PROJECT #18: CONVERT FROM BINARY TO DECIMAL

In this project, you're going to *convert* a binary number into its decimal form. Converting a number to a different base just means calculating that number's value and writing it using digits in the new base.

Tools

- **A pen or pencil** to write with.
- **Paper** to write on.
- **A calculator** to add numbers. If you're good at doing math in your head, try this project without one.

Step 1: Write It Down on Paper

First, write an eight-digit binary number on paper, leaving space between the digits, both above and below the number. I'm going to convert the binary number 1011 0101, and I wrote it out like this:

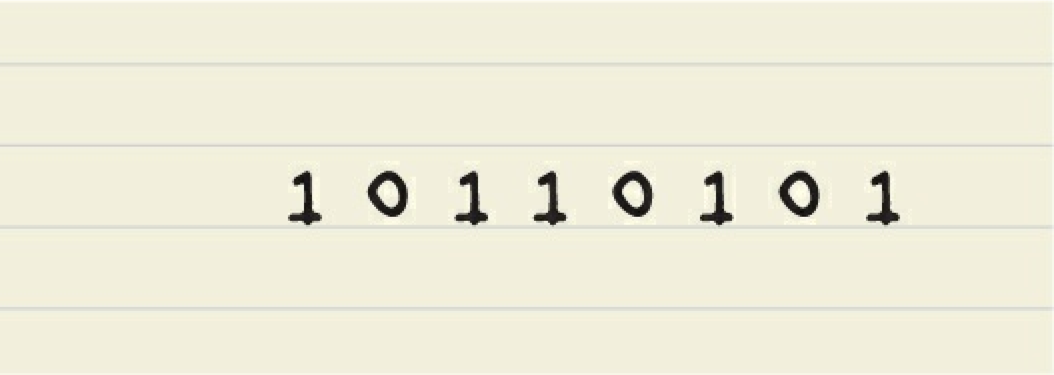

Step 2: Write the Position Values

Next, write the value of each position above each digit. Binary works the same way as the decimal example, but the base is 2. That means the value for the rightmost position is 1, and to find the next position to the left, you'd multiply the previous position's value by 2. For example, the second position's value is 2 × 1 = 2; the third position's value is 2 × 2 = 4; the fourth value is 4 × 2 = 8; and so on. Use a calculator if you don't want to calculate it all in your head. When you write those values down, you should have something like this:

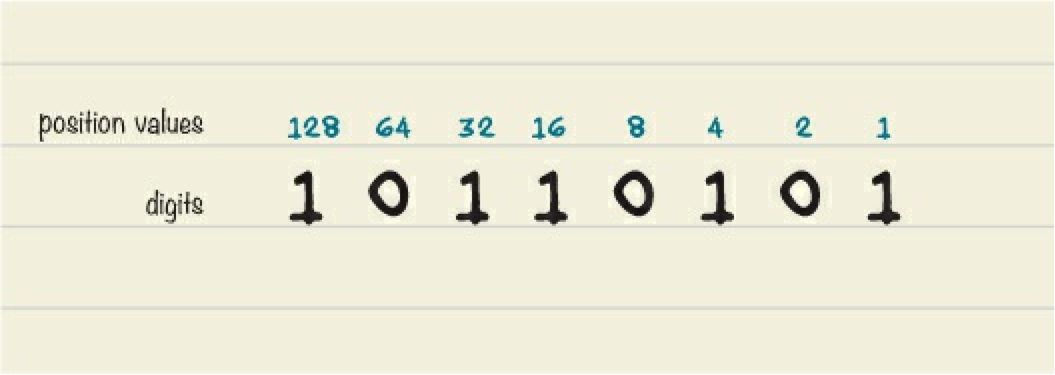

Step 3: Find the Value of Each Digit

Look at each digit of the binary number you wrote down. If a digit is 0, write 0 below it. If the digit is 1, write the value of that position below it. Those are the values you'll add together to get your final decimal number.

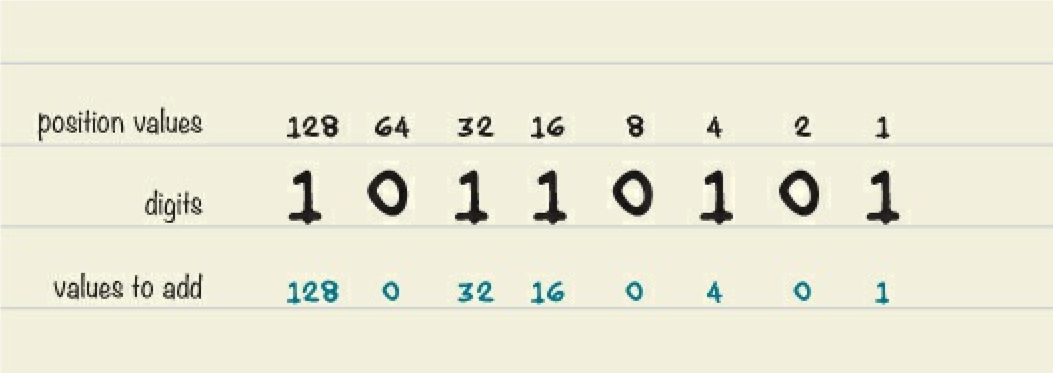
position values
128 64 32 16 8 4 2 1
digits
1 0 1 1 0 1 0 1
values to add
128 0 32 16 0 4 0 1

Step 4: Add the Numbers

You should now have three rows of numbers. Add the numbers in the bottom row (you might find it helpful to write a + sign between each pair of numbers) to get the sum, which is the decimal value of your binary number.

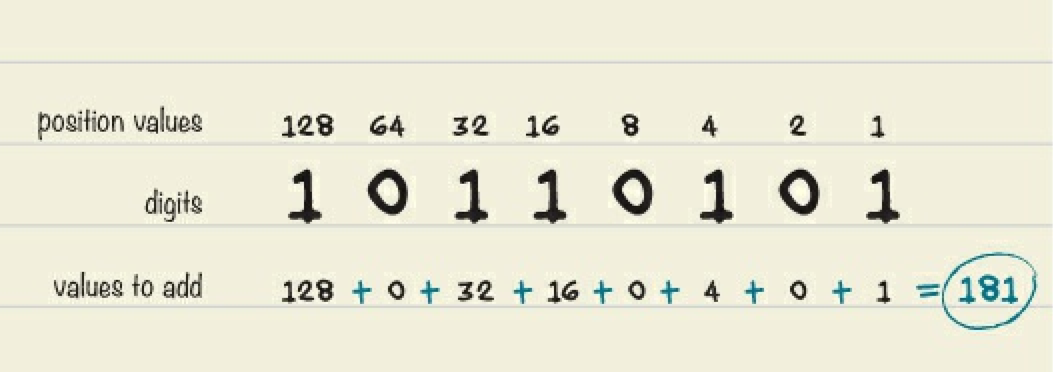
position values
128 64 32 16 8 4 2 1
digits
1 0 1 1 0 1 0 1
values to add
128 + 0 + 32 + 16 + 0 + 4 + 0 + 1 = 181

The binary number 1011 0101 is 181 in decimal. If you've been following along with that number and your answer is different, go through the steps again and check your results against mine to find out where the error is.

TRY IT OUT: TRANSFORM MORE BINARY NUMBERS

To get more practice with ones and zeros, let's bring some more binary numbers into the decimal world. Here are a few to try:*

1010
0011 1111
1000 0000
0011 1011 0101

* In decimal, these binary numbers are 10, 63, 128, and 949.

BITS AND BYTES

Each digit in a binary number is called a *bit* (that's short for *binary digit*), and computers handle numbers in blocks of eight bits called a *byte*. The binary number 1011 0101 has eight bits, so it's a byte. You can interact with a computer in many ways, but all your mouse clicks, key presses, webcam videos, and so on need to be translated into bits and bytes before the computer can understand them.

In fact, when working with computers and other digital gadgets, you usually see much, much larger numbers than just 1 byte. All the files on a computer are collections of bytes, but if you tried to describe their size in bytes alone, those numbers would be huge! That's why files are usually described in larger units, like kilobytes* (kB), megabytes (MB), gigabytes (GB), terabytes (TB), and so on. Here's what those units mean:

1 kB = 1,000 bytes

1 MB = 1,000 kB = 1,000,000 bytes

1 GB = 1,000 MB = 1,000,000 kB = 1,000,000,000 bytes

1 TB = 1,000 GB = 1,000,000 MB = 1,000,000,000 kB = 1,000,000,000,000 bytes

If a computer's hard disk can hold 1 TB of data, then it can hold one trillion bytes, which is eight trillion ones and zeros!

NUMBERS CAN BE ANYTHING

Now you might be thinking, "Why on Earth would I ever need eight trillion ones and zeros in my computer?" The fact is, whether you use a computer to write stories, draw pictures, talk to your friends, play video games, or anything else, you're actually using those ones and zeros.

For example, how does a computer show an image on a screen? A computer screen is made of a lot of small points called *pixels*, and each pixel can be set to a color that is a mix of red, green, and blue light. If you want a pixel to be the brightest yellow possible, you'd use numbers to tell the computer to turn that pixel's red and green settings to full intensity and its blue setting to zero (because mixed red and green light looks like yellow). This way you can translate numbers into an image on the screen.

PROJECT #19: COLOR GUESSING GAME

In this project, you'll build a pixel-color guessing game using binary values. It's a two-player game where you and a friend take turns being "the computer" and "the user."

The person who is the computer sets the color of the pixel by pushing a combination of three buttons. When the computer is ready, the user must try to guess which color the pixel is. When the user is ready to make the guess, they should say the chosen color out loud and then push the color revealer button to show the actual color. If the user guesses correctly, they get one point and get to guess again. If the guess is wrong, you switch roles. The first one to reach three points wins the round.

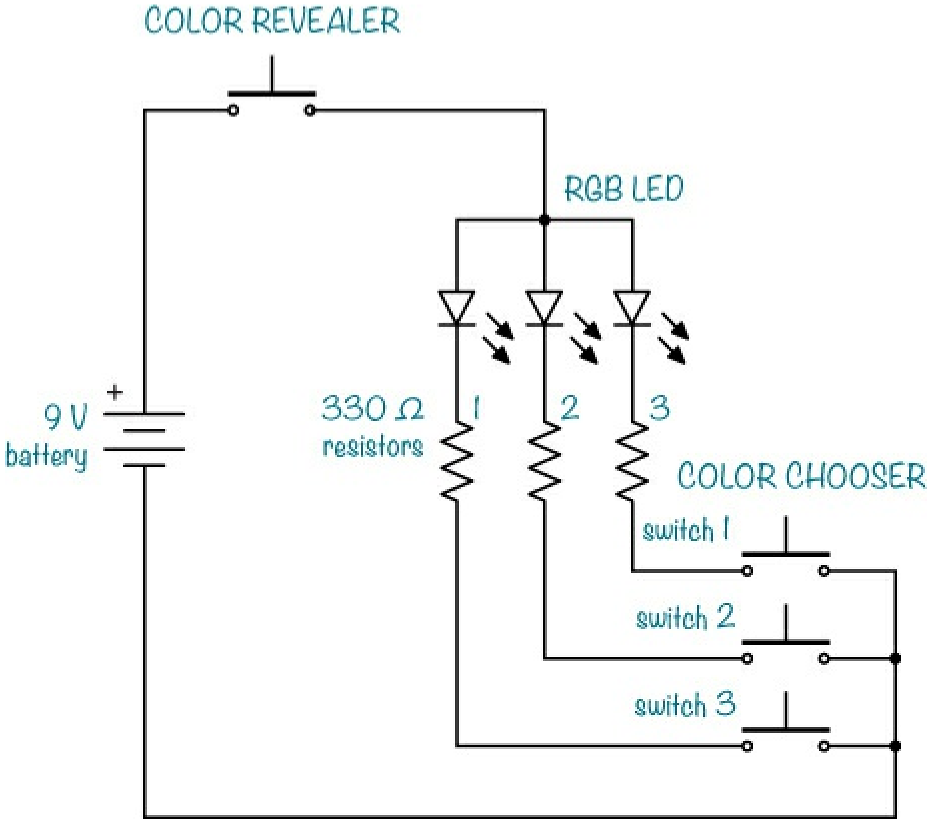
COLOR REVEALER
RGB LED
9 V
battery
+
330 Ω
resistors
1
2
3
COLOR CHOOSER
switch 1
switch 2
switch 3

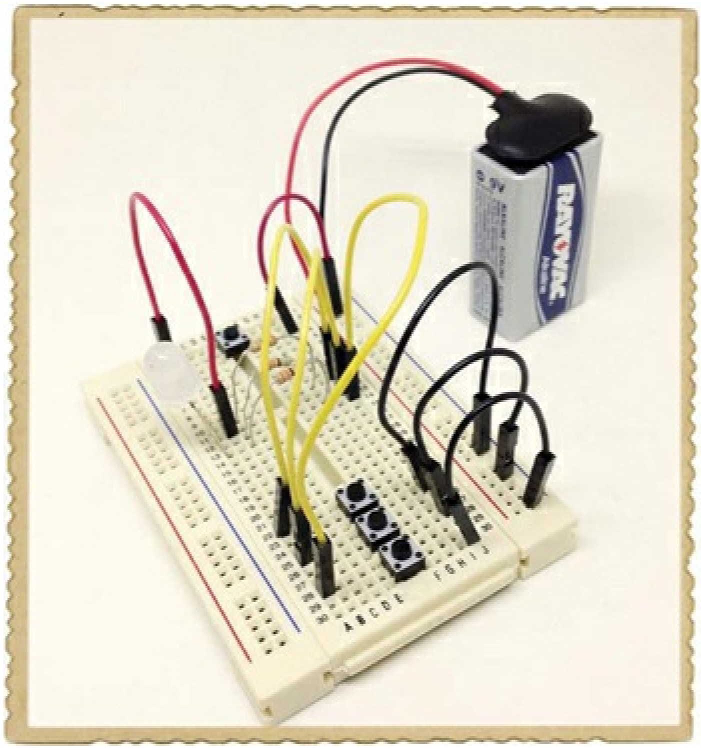

Meet the RGB LED

To create the pixel for this project, you're going to use a *red-green-blue (RGB) LED*, which combines a red LED, a green LED, and a blue LED into one component. By turning on and off each of these LEDs, you can create different colors. For example, turning on only the red and green LEDs would give you yellow.

There are two types of RGB LEDs: *common anode* and *common cathode*. This is what a common anode RGB LED looks like, along with its symbol:

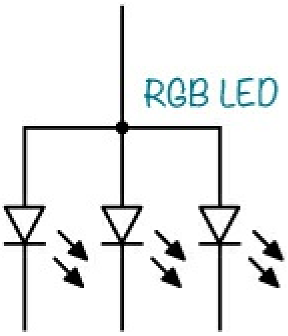

RGB LEDs have four legs. You’ll use three legs to set the color, and the fourth leg is the *common* leg. In common anode RGB LEDs, the positive (anode) sides of the three LEDs are combined into one pin; common cathode RGBs combine the negative (cathode) sides into one pin.

In this project, you’re going to build the following circuit, which uses a common anode RGB LED and some buttons.

Shopping List

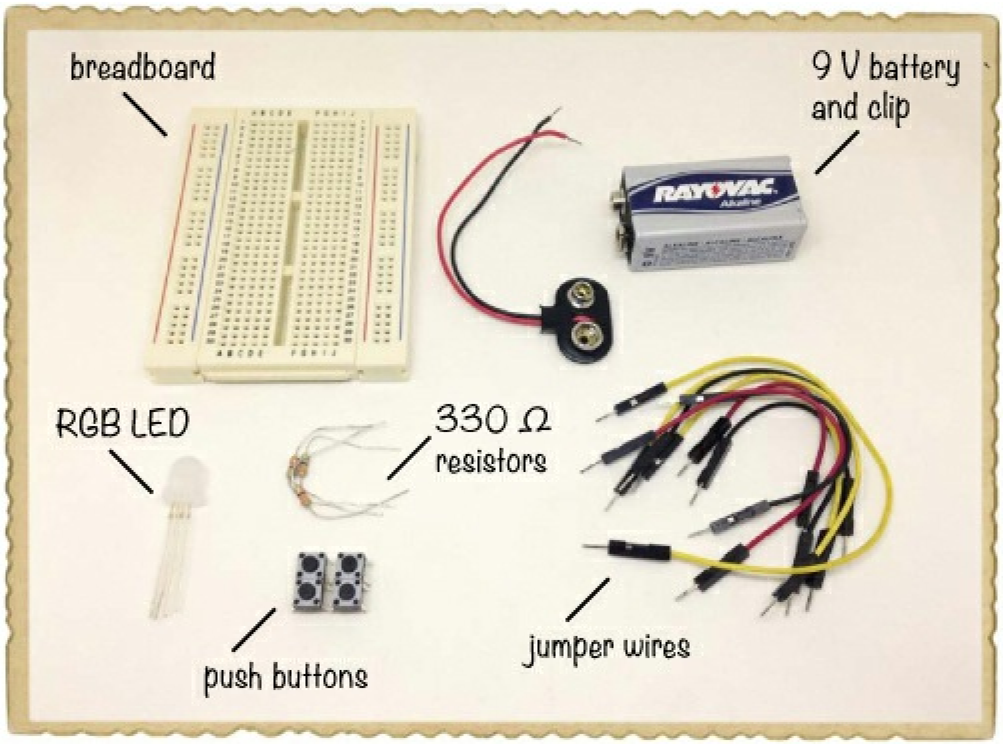

- **A standard 9 V battery** to power the circuit.
- **A 9 V battery clip** (Jameco #11280, Bitsbox #BAT033) to connect the battery to the circuit.
- **A breadboard** (Jameco #20601, Bitsbox #CN329) with at least 30 rows.
- **Breadboard jumper wires** (Jameco #2237044, Bitsbox #CN236) to easily connect parts. (Standard hookup wire works, too.)
- **Three 330 Ω resistors** (Jameco #661386, Bitsbox #CR25330R) for limiting the current to the LED.
- **RGB LED with common anode** (Jameco #2219567, Bitsbox #OP100) to create colors.

- **Four push buttons** (Jameco #119011, Bitsbox #SW087) to set and reveal colors.

Step 1: Place the Color Chooser’s Push Buttons

Orient your breadboard so that the notch down the middle runs from top to bottom. Then, place three push buttons at the bottom of your breadboard; each should have one pair of legs on the left side of the notch and one pair on the right. This should leave you plenty of room to connect components on both sides of the buttons, and you’ll be able to separate the computer’s buttons from the user’s button more easily. With your buttons in position, connect three jumper wires from the bottom-right pin of each push button to the negative supply column.

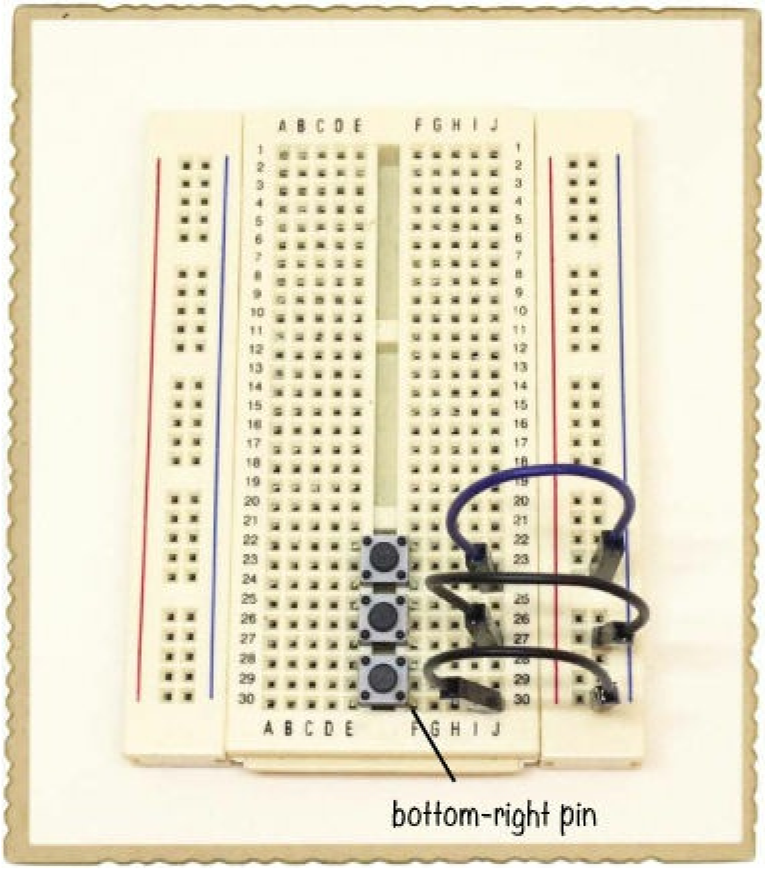

Step 2: Connect the RGB LED

In the middle of the board on the left, connect your RGB LED. The longest leg of the RGB LED is the common anode leg. Leave this unconnected for now. Connect a 330 Ω resistor from each of the remaining LED legs to an empty row on the right side of the breadboard. From each of these three rows, connect a jumper wire to the top-left pin of each push button.

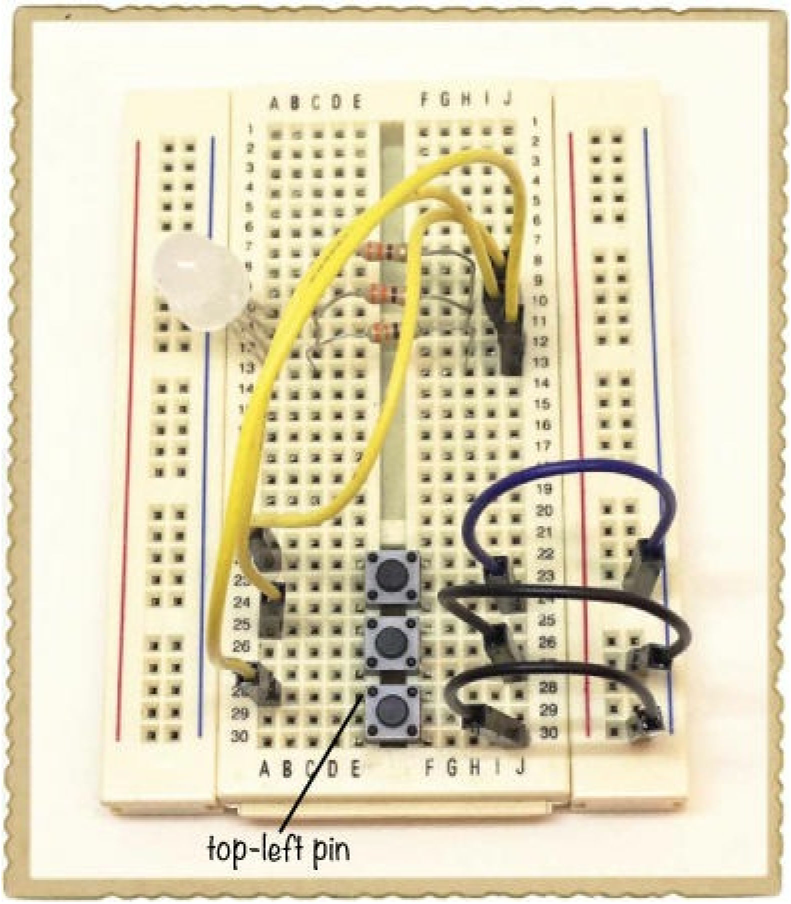

Step 3: Connect the Color Revealer Button

Next, add the button that reveals the computer's chosen color to the user. Place a fourth push button all the way at the top of the breadboard, across the middle notch. Connect a jumper wire from the common anode leg of the RGB LED to the top-left pin of the fourth push button. Then, connect one jumper wire from the bottom-right pin of this push button to the positive supply column on the right.

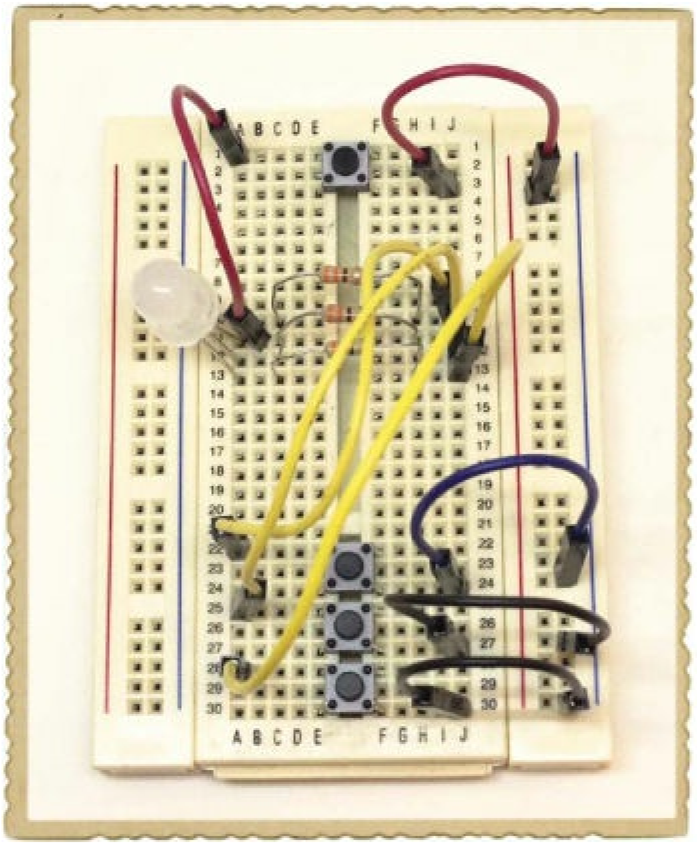

Step 4: Test the Colors

Connect the battery clip to the supply columns on the right, connect a battery, and try it out! Push some of the buttons on the bottom and then push the color revealer button to see the color. In this photo, I set my RGB LED to green.

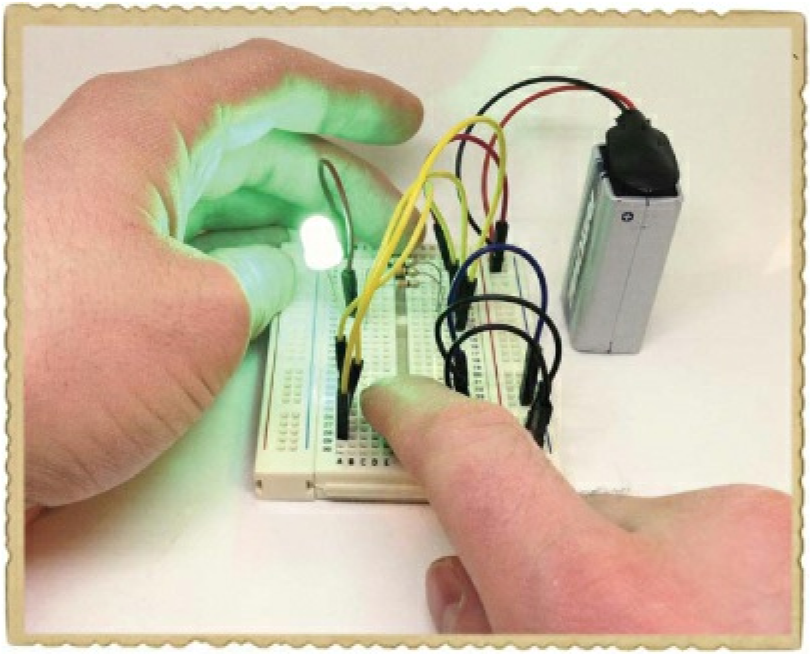

You should be able to see the following seven colors, depending on which buttons you push:

Color	Top button	Middle button	Lower button
Red	Pushed (1)	Not pushed (0)	Not pushed (0)

Green	Not pushed (0)	Pushed (1)	Not pushed (0)
Blue	Not pushed (0)	Not pushed (0)	Pushed (1)
Yellow	Pushed (1)	Pushed (1)	Not pushed (0)
Cyan	Not pushed (0)	Pushed (1)	Pushed (1)
Magenta	Pushed (1)	Not pushed (0)	Pushed (1)
White	Pushed (1)	Pushed (1)	Pushed (1)

If your buttons do not match these colors, switch the wires between the three resistors to the buttons so that they match.

Step 5: What If the Game Doesn't Work?

If you see no colors at all, check that your connections match the circuit diagram. If the connections are correct and you still can't see any colors when pushing the color revealer button, you may have a common cathode LED instead of a common anode. To check for this, simply switch the positive and negative connections from the battery.

When you've verified that all the colors work, invite a friend to play!

TRY IT OUT: SOLDER THE COLOR GUESSING GAME

This is a perfect circuit to practice your soldering skills on. Get a prototyping board and solder the circuit onto it so that you have a permanent Color Guessing Game to bring with you on a long car ride.

HOW BINARY NUMBERS CAN CREATE WORDS

Images aren't the only things that can be stored as binary numbers; you can represent letters as numbers, too. One way to do this is to use the *ASCII code*, which is a standard set of bytes that computers understand as upper- and lowercase letters, numbers, punctuation marks, and so on. The numbers in this table represent the lowercase letters in the English alphabet.

Decimal ASCII code	Binary value	Lowercase letter
97	0110 0001	a
98	0110 0010	b
99	0110 0011	c
100	0110 0100	d
101	0110 0101	e

102	0110 0110	f
103	0110 0111	g
104	0110 1000	h
105	0110 1001	i
106	0110 1010	j
107	0110 1011	k
108	0110 1100	l
109	0110 1101	m
110	0110 1110	n
111	0110 1111	o
112	0111 0000	p
113	0111 0001	q
114	0111 0010	r
115	0111 0011	s
116	0111 0100	t
117	0111 0101	u
118	0111 0110	v
119	0111 0111	w
120	0111 1000	x
121	0111 1001	y
122	0111 1010	z

For example, the letter *a* can be represented by the decimal number 97, which is the binary number 0110 0001. You can use this table to encode and decode secret messages written with only ones and zeros.

PROJECT #20: THE SECRET MESSAGE MACHINE

This project is a circuit that shows eight-bit binary numbers with LEDs. An LED that is lit is a 1, and an LED that is off is a 0. You'll use switches to set the binary number and a push button to show the binary number on the LEDs.

Eight bits are enough to represent an ASCII character, so you can use this project to generate secret messages that can only be decoded by someone who knows binary! This finished Secret Message Machine shows the letter *w*, or 0111 0111.

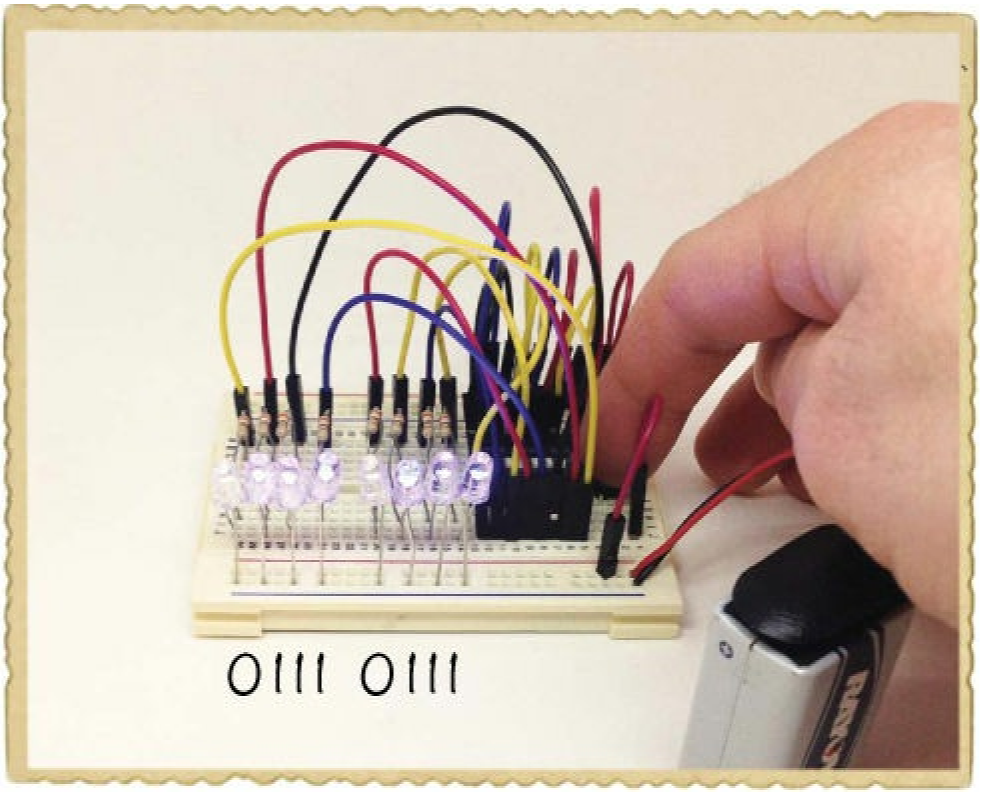

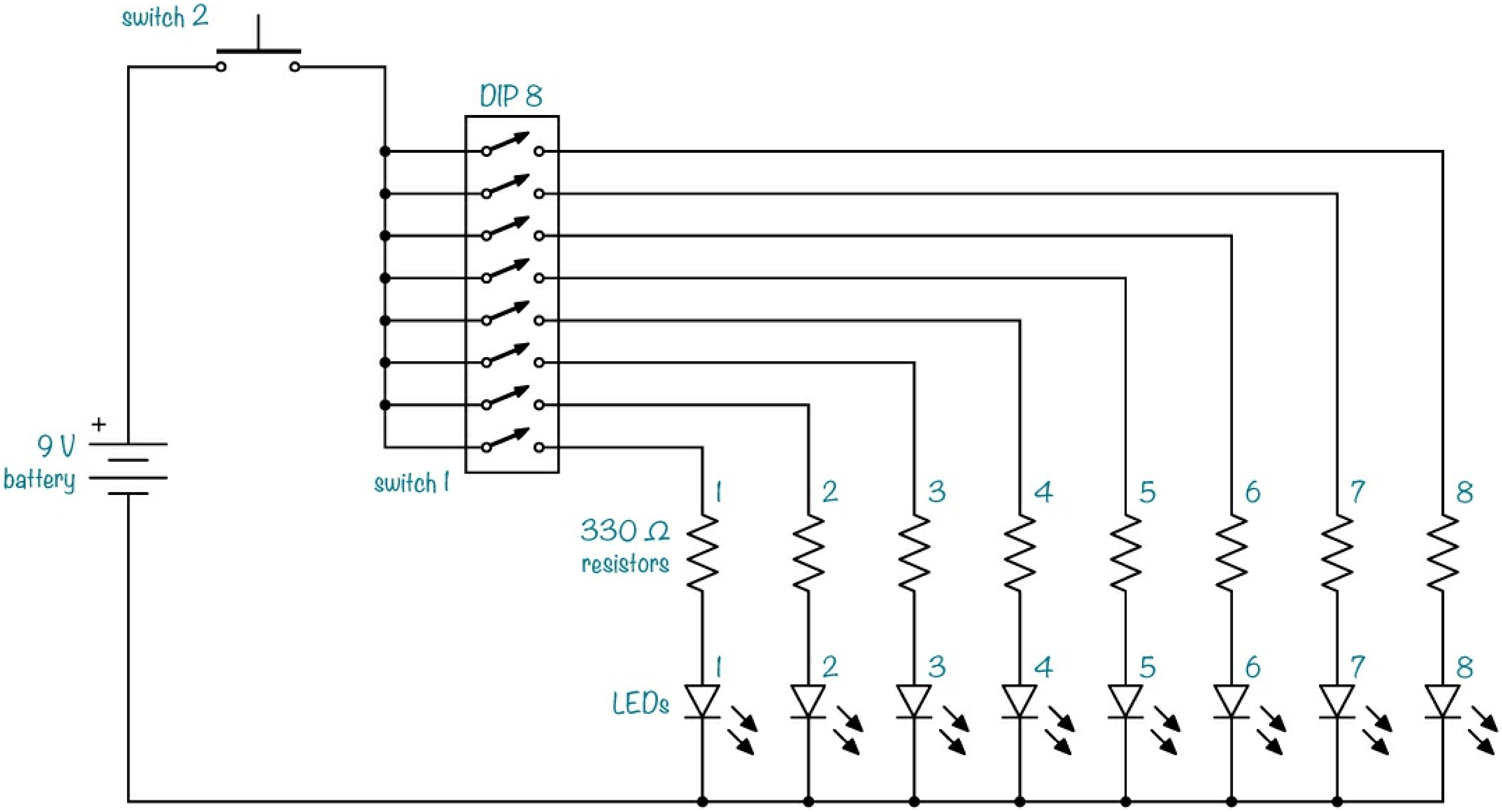
switch 2
DIP 8
+
9 V battery
switch 1
1
2
3
4
5
6
7
8
330 Ω resistors
1
2
3
4
5
6
7
8
LEDs

Meet the DIP Switch

This circuit uses eight switches to set the binary value, but it would be a pain to plug that many individual buttons into your breadboard. Fortunately, you can just use a *DIP switch*, which is a component with a row of one or more mini switches.

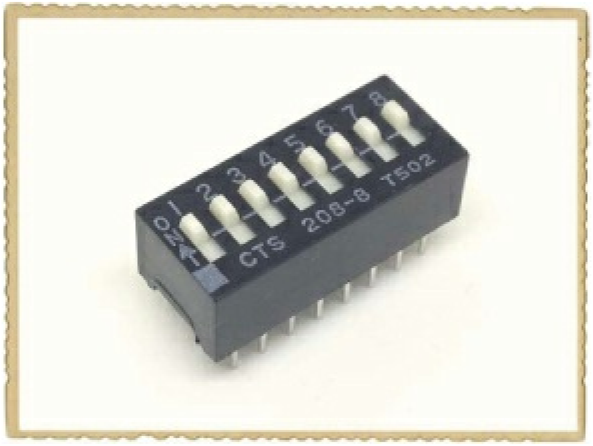

NOTE

DIP stands for dual in-line package, *which is how the pins are arranged. A component with "DIP" in its name has two rows of pins that you can plug into a breadboard. The DIP switch in this project has eight individual switches. This is perfect for creating a binary value input to a row of LEDs.*

Shopping List

- **A standard 9 V battery** to power the circuit.
- **A 9 V battery clip** (Jameco #11280, Bitsbox #BAT033) to connect the battery to the circuit.
- **A breadboard** (Jameco #20601, Bitsbox #CN329) with at least 30 rows.
- **Breadboard jumper wires** (Jameco #2237044, Bitsbox #CN236) to easily connect parts. (Standard hookup wire works, too.)
- **Eight 330 Ω resistors** (Jameco #661386, Bitsbox #CR25330R) for limiting the current to the LEDs. Any value between 270 Ω and 470 Ω works well.
- **Eight blue LEDs** (Jameco #2193889, Bitsbox #OP033) to show the binary number.
- **One 8-position DIP switch** (Jameco #696984, Bitsbox #SW098) to set the binary number.
- **A push button** (Jameco #119011, Bitsbox #SW087) to turn on the LEDs.

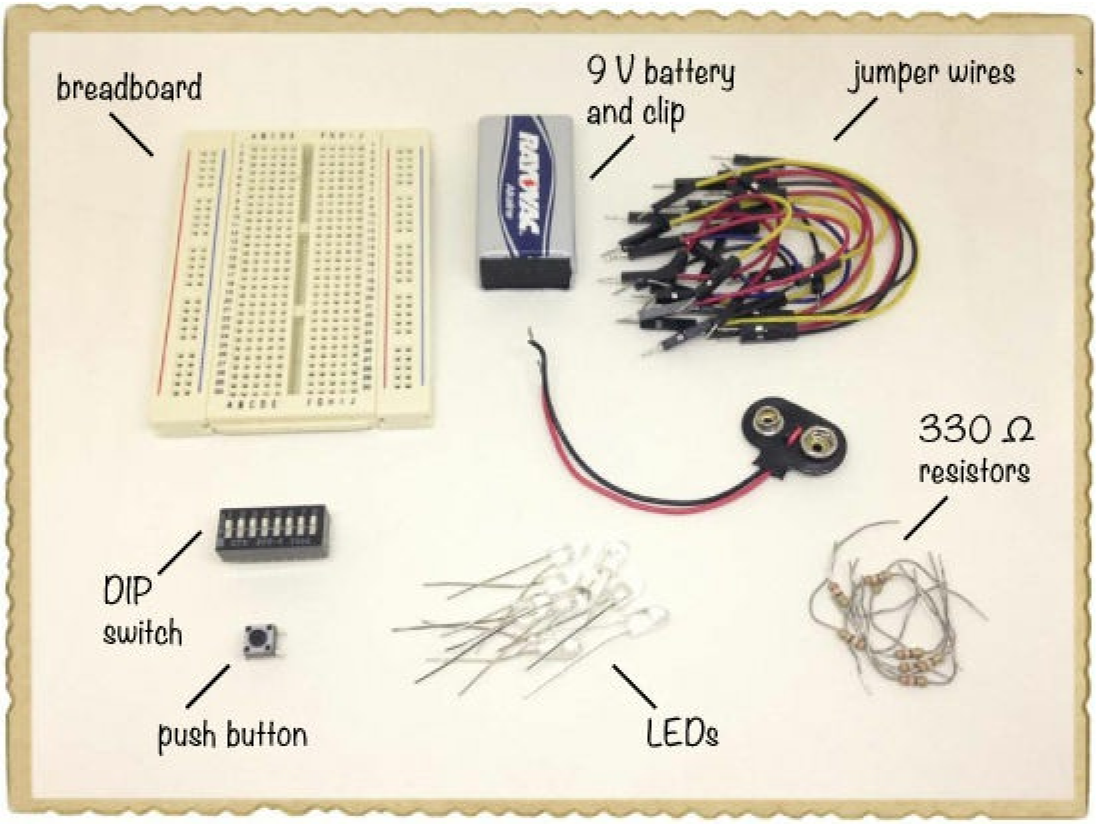
breadboard
9 V battery and clip
jumper wires
330 Ω resistors
DIP switch
push button
LEDs

Step 1: Connect the Push Button

Plug the push button into the top of your breadboard. This project will take up a lot of space on your breadboard, so from here on, try to keep your components as close together as possible.

Connect a jumper wire from the top-right breadboard row, which should be connected to the upper pin of the push button, to the positive supply rail on the right side. Then, connect a jumper wire from the lower pin to the positive supply column on the left.

Your breadboard should look something like this:

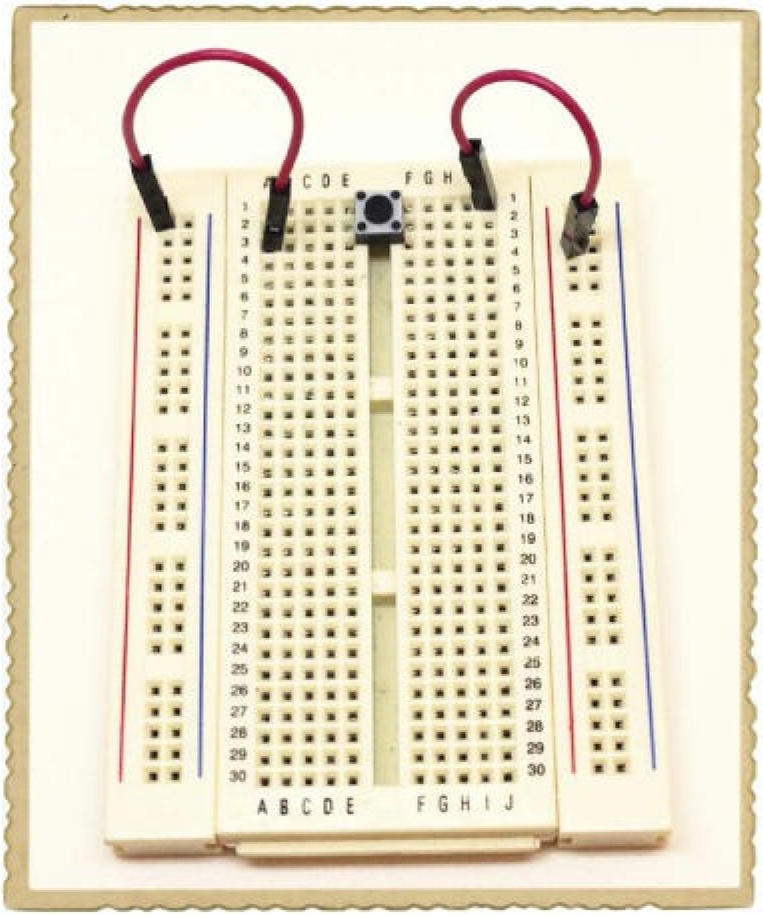

This push button controls the connection from the battery's positive terminal to the rest of the circuit. When the button is not pushed, there's no closed loop for the current to flow through, and the circuit doesn't have power.

Step 2: Connect the DIP Switch

Next, connect the DIP switch to your breadboard just below the push button, with the numbers on the right side of the middle notch. Run one jumper wire from each of the eight rows to the positive supply column on the left, for a total of eight wires. You'll need to be able to set the switches on and off, so connect the wires as far away from the switch as possible to leave room for your fingers.

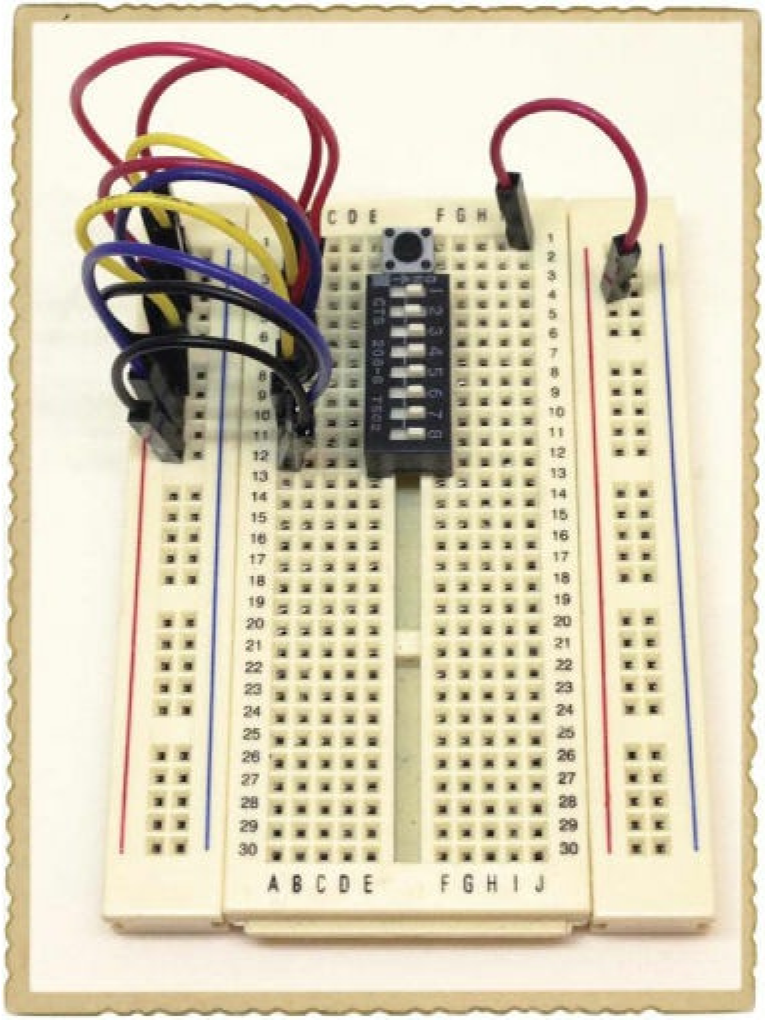

Step 3: Connect the LEDs

Now, connect one LED’s short leg to the negative supply column on the right, and connect its long leg to the last row at the bottom of your breadboard. Follow the same pattern upward to plug the rest of the LEDs into the breadboard on rows of their own, leaving one empty row between LEDs. (If you want to make it easier to read the LEDs, put your LEDs in two groups of four by leaving a couple of extra rows after the fourth LED.) Then, connect a resistor from each row with an LED to the same row on the left side of the notch in the middle of the breadboard.

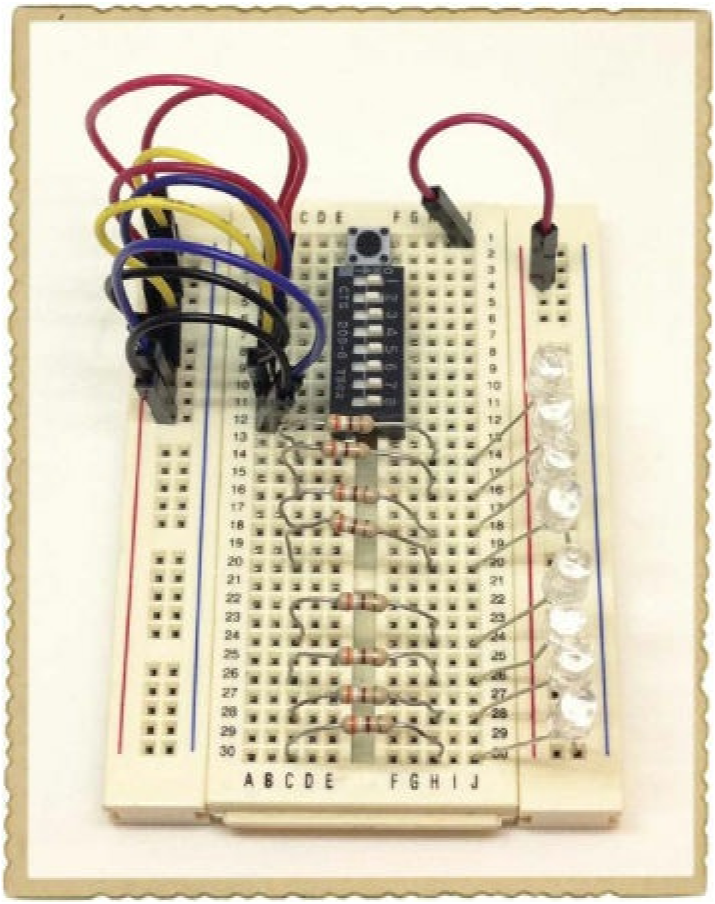

It will be best to connect the LEDs and switches so that you and your friend can sit on opposite sides of the board and see the bits in the same order, so pay careful attention to these connections. With your resistors in place, connect a jumper wire from each resistor row on the left side of the breadboard notch to a DIP-switch pin on the right side of the notch. Connect the resistor closest to the bottom of the breadboard to Switch 1, the next LED to Switch 2, and so on; you should end by connecting the top resistor to Switch 8.

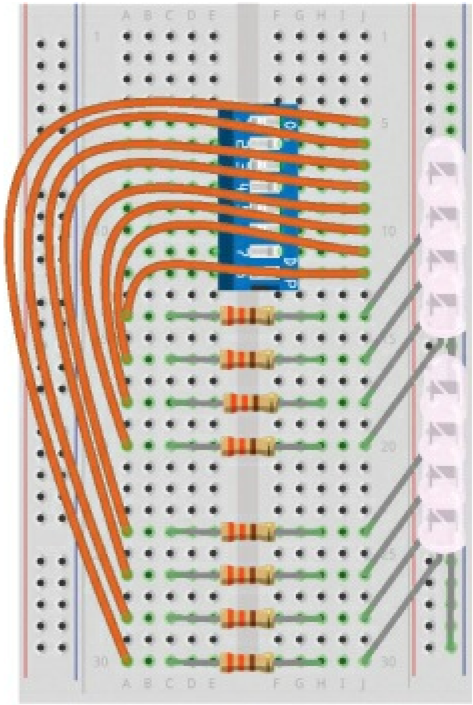

Step 4: Send a Secret Message!

Connect the negative side of the battery clip to the negative supply column, and then connect the positive side to the positive supply column, as shown.

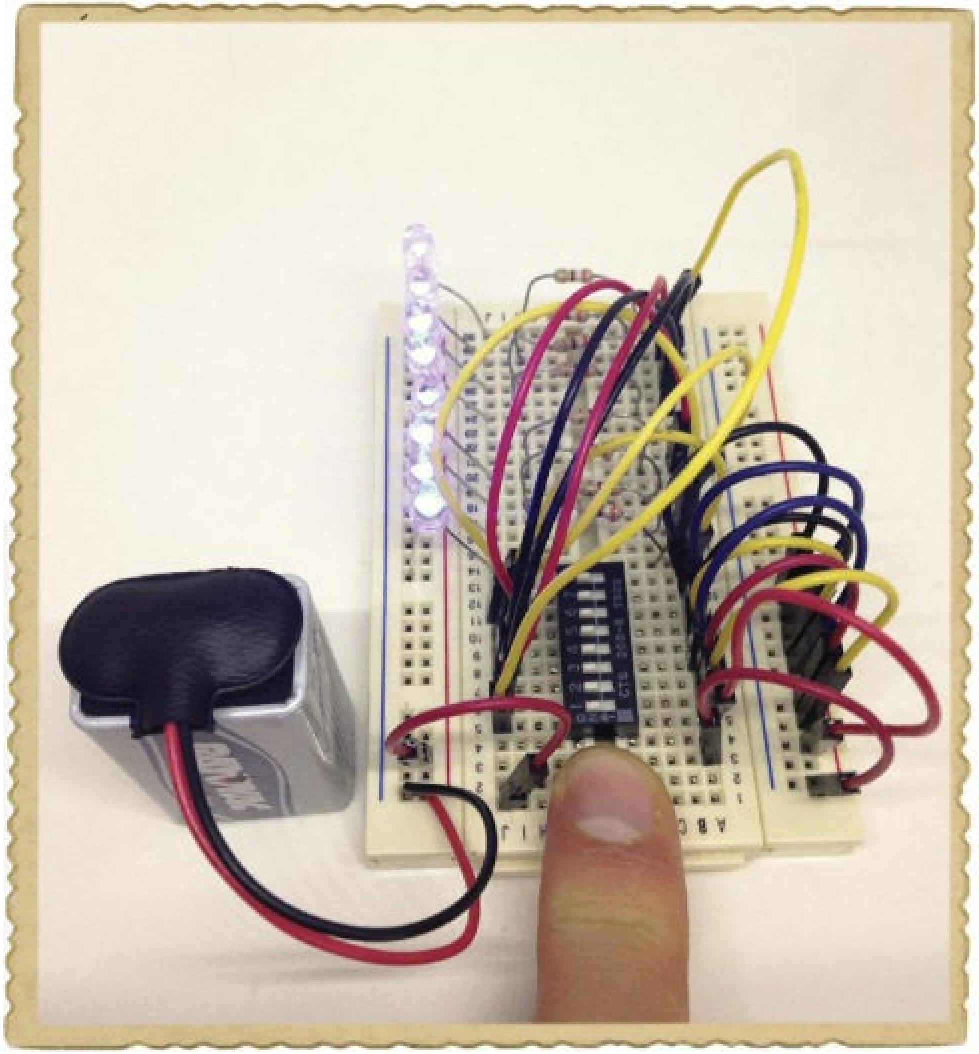

Then, push all the switches in the “on” direction, press and hold the button to see whether all the lights light up, and turn off each LED one by one.

If all the LEDs work, it’s time to play! Invite a friend over and use this circuit to “talk” to them without speaking. Sit on opposite sides of a table, write down a word on a piece of paper for your own reference, and keep it hidden from your friend. Look at the ASCII code table and set the switches to the binary values for the first letter in your word. For example, the switch in the next photo is set to display a lowercase letter *a*.

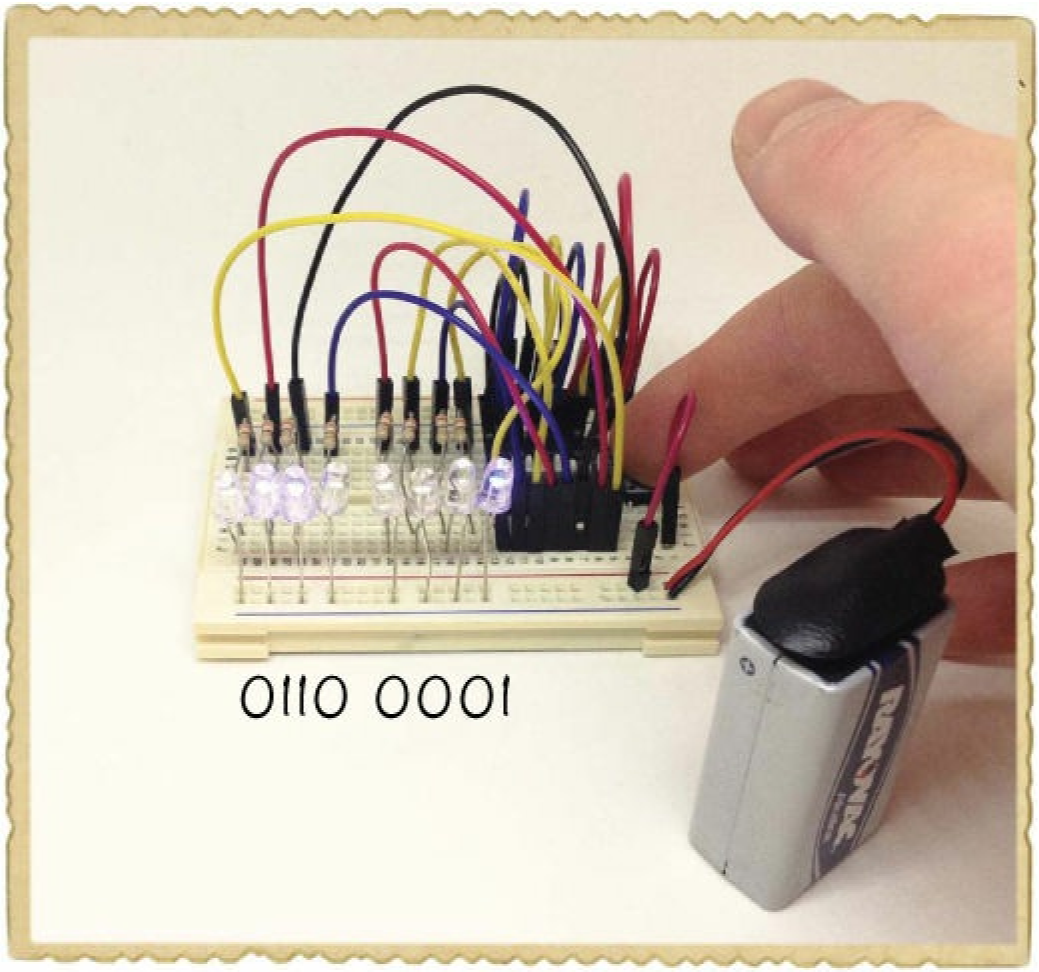

When your switch bank is ready to show a letter, push the button so that the LEDs light up. Hold the push button while your friend writes down the binary value. Then, release the push button and set the binary value for the next character. Keep doing this until you've shown all the characters in your word. When you're done, show the ASCII code table to your friend so they can try to figure out your word.

Step 5: What If the Secret Message Machine Doesn't Work?

If you've been following along for the whole book, you've connected a lot of LED circuits, but this one does have a lot of components. It's easy to make a wrong connection somewhere, so if your circuit doesn't work right away, check each connection carefully against this project's circuit diagram.

If none of the lights work, first check that the battery's positive and negative terminals are connected correctly. Next, check that the LEDs are oriented the right way and that the resistors are the right values. If some lights work while others don't, then some of the LED or resistor legs may be touching when they shouldn't. Inspect them closely to find the error.

WHY COMPUTERS USE ONES AND ZEROS

Computers use ones and zeros instead of the decimal numbers from 0 to 9 because when you have only two values, it makes constructing electronic circuits to do calculations and store the values much easier.

For example, because each digit has only two possible values, one or zero, it's easy to create a simple memory block to store binary numbers using switches, as you did with the DIP switch in Project #20.

It's possible to save ones and zeros in many different ways, thereby creating *memory*. In the early days of computing, it was common to save sets of ones and zeros as physical cards with holes in them. Today, many hard disks save ones and zeros on magnetic disks, or even via electrons stored inside an integrated circuit.

WHAT'S NEXT?

In this chapter, you learned how binary numbers work. You've seen how ones and zeros can be used to show images on your computer screen and how to decode a bunch of seemingly random ones and zeros into readable text. To explore binary numbers a bit more, come up with your own binary codes for the letters in the alphabet. There are 26 letters from *a* to *z*. How many digits do you need to represent all 26 letters? To figure this out, you can start by saying that the binary number 1 means *a*. The next binary number, 10, is *b*. Next, 11, is *c*. Write this down and keep increasing until you reach *z*. Then, count the number of digits you needed for the *z*. This is the minimum number of digits you need.

Here's another thing to try. Normally, if you count with your fingers, you can count to 10, right? If you don't use your thumbs, you can only count to 8. But what if you count in binary instead? A straight finger is 1, and a bent finger is 0. How high can you count with eight fingers now?

Digital values have another very important use: they allow you to create *logic circuits*, which make decisions based on whether certain wires see a high voltage or a low voltage. High is typically considered "true," and low is typically "false." If those wires are associated with simple true-or-false questions and the voltages are the answers, then you can think of the output from a logic circuit as a conclusion reached after asking a series of questions.

With logic, you can make circuits that do math, like adding two numbers, or circuits that do something only under a certain condition, like a door lock that opens only if you input the right combination of numbers. In Chapter 10, I'll show you some of the building blocks of digital logic and how you can make smart circuits with them.

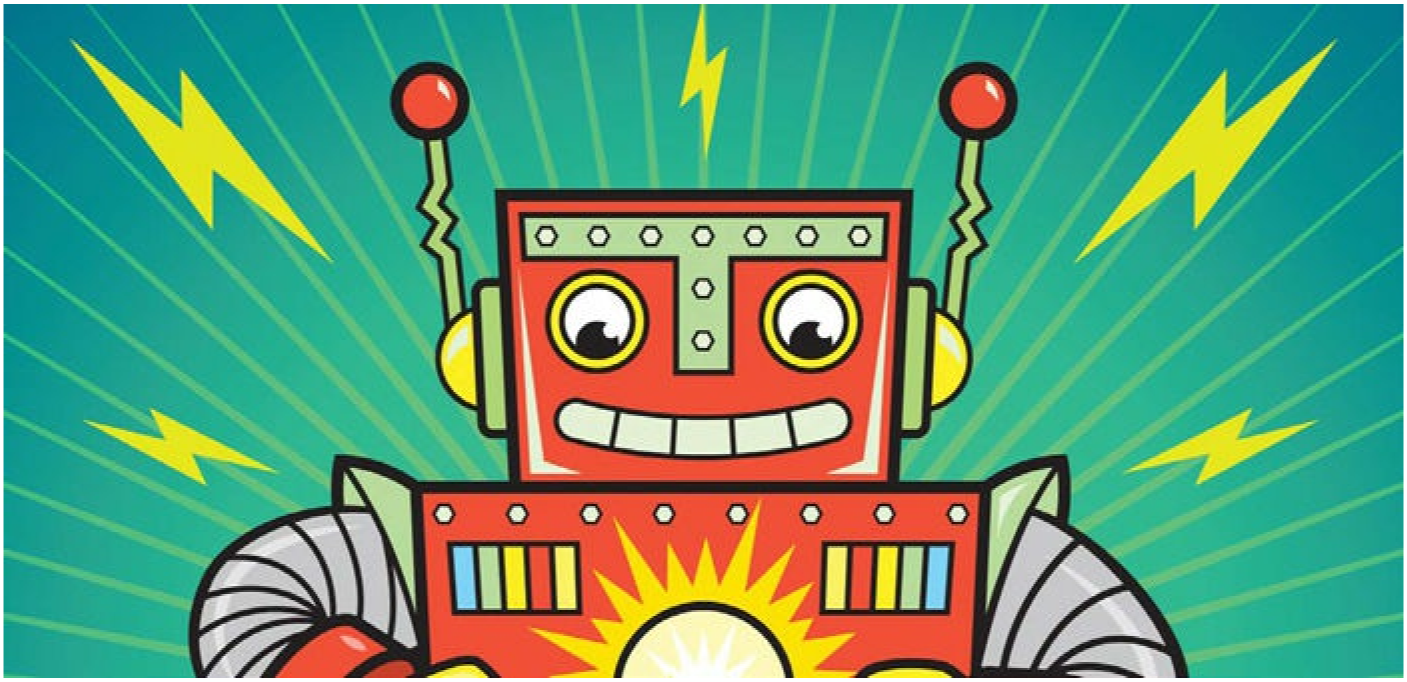

10

CIRCUITS THAT MAKE CHOICES

Chapter 9 was all about ones and zeros, and you played with bits and bytes in a couple of projects. Now in this chapter, you'll make circuits that actually use ones and zeros to make decisions. *Logic gates* are components that check for the voltages that represent those ones and zeros and output a voltage accordingly. I'll show you a few types of logic gates and how you can use them to create a secret code detector.

IT'S ONLY LOGICAL

Logic is a way of reaching a conclusion based on pieces of information that you know to be true or false. For example, imagine you know the following statement is true, beyond a doubt:

> **Statement 1: If there are oranges in your fridge AND you have an orange squeezer, THEN you're able to make orange juice.**

If you trust the preceding statement, then there are two conditions to check before you can make orange juice:

> **Condition 1: There are oranges in your fridge.**
>
> **Condition 2: You have an orange squeezer.**

If you check your kitchen and find that these conditions are true, then you can logically conclude that you can make orange juice.

Computers use *Boolean logic*, which is a system of logic that works only with the values *true* and *false* to convert ones and zeros into actions. For a computer to know whether you can make orange juice or not, it would have to reach that conclusion through Boolean logic. Let's try thinking like a computer!

First, look for the conditions in Statement 1 that affect whether you can make orange juice or not. In this case, the conditions are the two phrases between "if" and "then," joined by "and." Assign them letters as follows:

> **There are oranges in your fridge. = A**
>
> **You have an orange squeezer. = B**

The conclusion is the statement after "then." Give it a letter, too:

> **You're able to make orange juice. = Q**

With these letters, you could rewrite Statement 1 as "If A and B, then Q." In Boolean logic shorthand, that looks like this:

A AND B = Q

This is a *logic equation*, where *AND* is an operator like addition or subtraction. When both statements on either side of AND are true, the conclusion Q is true.

Given Condition 1 and Condition 2, A and B are both true. Substitute both into the equation to get:

True AND True = Q

Q = True

Because both A and B are true, then Q must be true. Time to make orange juice!

How a Computer Decides When It Can Make Orange Juice

Condition A (There are oranges in your fridge.)	Condition B (You have an orange squeezer.)	Result Q (You're able to make orange juice.)
False	False	False
False	True	False
True	False	False
True	True	**TRUE!**

MEET THE LOGIC GATES

Many of the circuits inside your computer are physical versions of logic equations, complete with smaller circuits called *logic gates*, which are physical logical operators. A logic gate takes ones and zeros—representing true and false, respectively—as inputs and then outputs a 1 or 0 based on the results of the equation inside.

You can make really awesome projects with logic gates yourself, too!

I remember the first time my dad told me about logic gates: I went straight to my room and spent hours trying to combine them on paper in different ways to add binary numbers. I hope you have as much fun with them as I did! Now, let's look at how a few different logic gates work.

AND Gates Check for Two True Inputs

The *AND gate* is the physical form of the AND operator you used to decide whether you were able to make orange juice. An AND gate has two or more inputs—A and B, for example—and one output—Q, for example. It checks whether A and B are both 1, and if they are, then Q is 1; otherwise, the output is 0. Q is 1 only if both A *and* B are 1; if one or both inputs are 0, the output is 0.

I find it helpful to write out the values of Q that result from different input combinations in a *truth table*. This truth table shows all possible input combinations for the AND gate and what the output will be for each. In a truth table, 0 stands for *false*, and 1 stands for *true*.

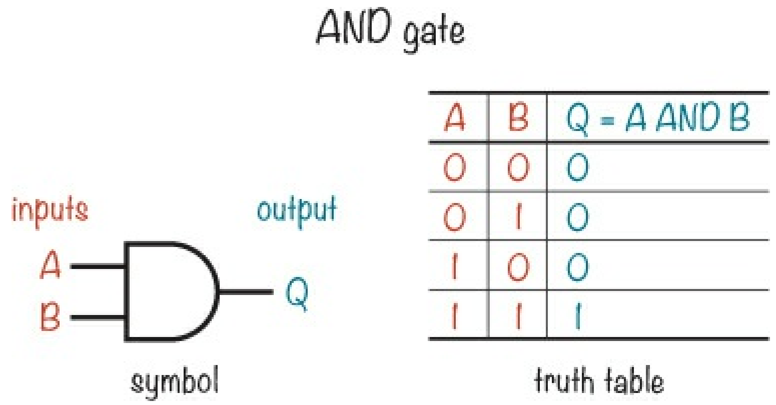
AND gate

A	B	Q = A AND B
0	0	0
0	1	0
1	0	0
1	1	1

inputs
output
A
B
Q
symbol
truth table

OR Gates Check for One True Input

The *OR gate* checks whether input A *or* input B is 1. If either is 1 or both are 1, then the output Q is also 1. But if both inputs are 0, then the output is 0.

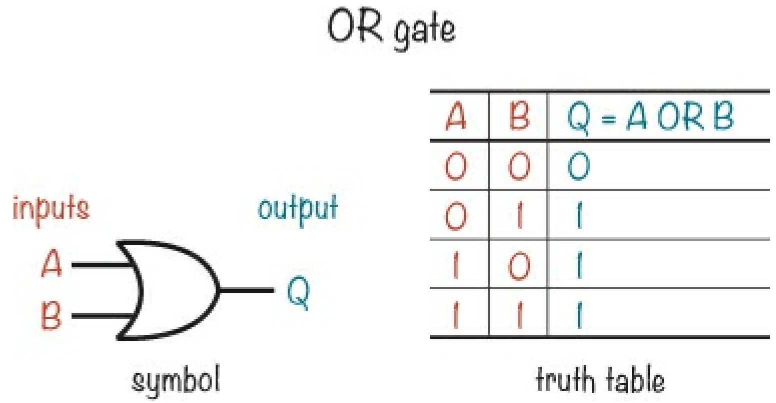
OR gate

A	B	Q = A OR B
0	0	0
0	1	1
1	0	1
1	1	1

inputs
output
A
B
Q
symbol
truth table

NOT Gates Flip Inputs

The *NOT gate*, also called an *inverter*, has only one input and one output, and its function is very simple: the output is the opposite of the input. If the input is 1, then the output is 0. If the input is 0, the output is 1.

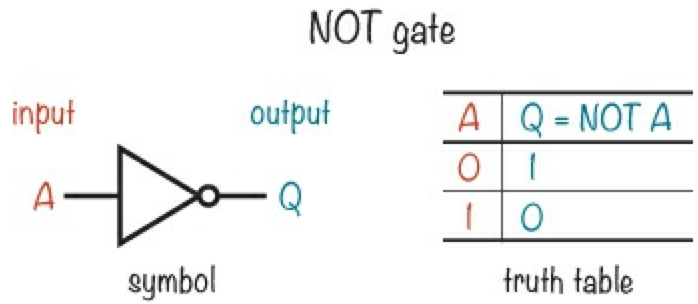
NOT gate

A	Q = NOT A
0	1
1	0

input
output
A
Q
symbol
truth table

A Bigger AND Gate

AND gates and OR gates can have more than two inputs. For example, here’s a 4-input AND gate symbol:

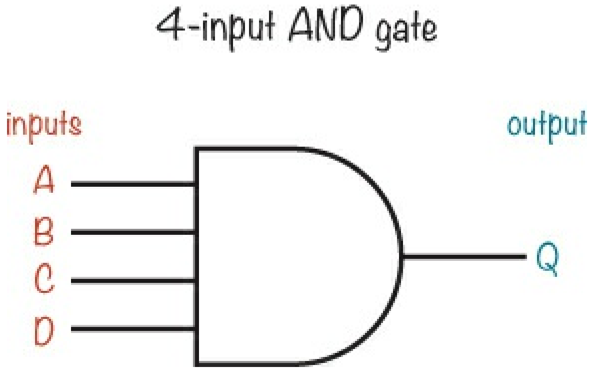

Because it’s an AND gate, the result will be 1 only if all four inputs are 1; otherwise, it will be 0. That is, the output Q is true (1) if all four inputs—A, B, C, and D—are true (1):

Q = A AND B AND C AND D

We can also make a 4-input AND gate from three 2-input AND gates, like this:

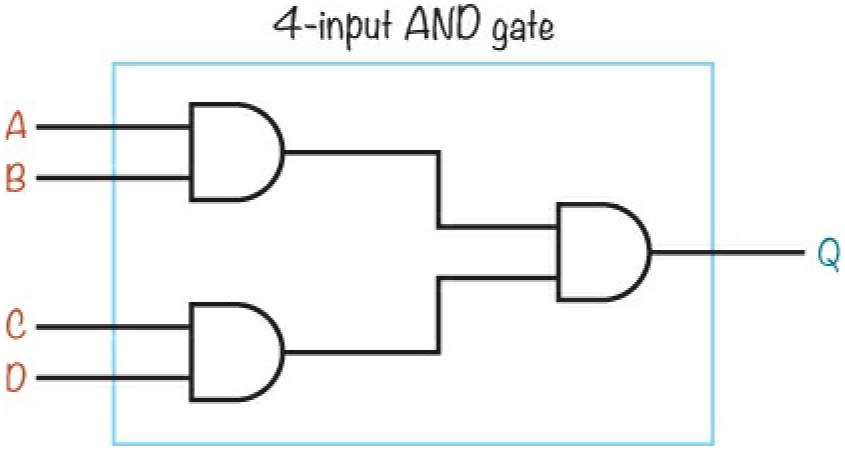

HOW TO DRAW LOGIC CIRCUIT DIAGRAMS

You can use logic gates to build a circuit that checks for conditions and decides what to do based on them. For example, imagine you could deactivate your alarm system from Chapter 1 by entering a secret code. Then, you could leave the alarm system on while you’re gone, and if someone opened the door, they’d have to know the right code to turn the alarm off. With logic gates, a circuit can check easily whether the right code was entered.

A Logic Equation for a Secret Code

Let’s say the secret code is 1001, and when the secret code is detected, an LED should turn on to indicate success. When building logic circuits, it’s helpful to write the logic equation for your circuit before building, so let’s practice.

First, think about what each 1 and each 0 in the secret code represents in terms of logic gates. In this case, you want the LED to turn on only when four conditions are true, and you can connect those conditions with AND operators as follows.

Let’s represent the four bits in the secret code with the letters *W*, *X*, *Y*, and *Z*. Then, you can check each bit to see whether it’s the correct value, testing for W = 1, X = 0, Y = 0, and Z = 1.

You’ll need to AND the four secret code bits together using a 4-input AND gate. But simply connecting W, X, Y, and Z straight to the AND gate would give this logic equation:

Q = W AND X AND Y AND Z

This would test whether all the bits are 1 because Q = 1 only if W, X, Y, and Z are all 1. Instead, you need to test for the secret code where W and Z are 1, but X and Y are 0.

Fortunately, in Boolean logic, you have only two options: 1 or 0 (true or false). If something is 0, then it is NOT 1; in words, if something is false, then it is NOT true. This means that if X = 0 (false), then NOT X = 1 (true). Knowing that, you can rewrite the equation as follows:

Q = W AND (NOT X) AND (NOT Y) AND Z

This equation uses NOT on bits X and Y, which should be 0 for the secret code. The NOT will invert their values, changing 0 to 1 and 1 to 0.

Converting a Logic Equation into a Circuit Diagram

Now we'll draw the secret code equation as a circuit. The final output will be a single 1 or 0. You need a 4-input AND gate to test all four code bits at once, and you're going to make this using three 2-input AND gates as explained earlier. Because you need to test whether the X and Y bits are 0, you'll need to use a NOT gate for each to invert the 0 to 1.

Here's the final circuit:

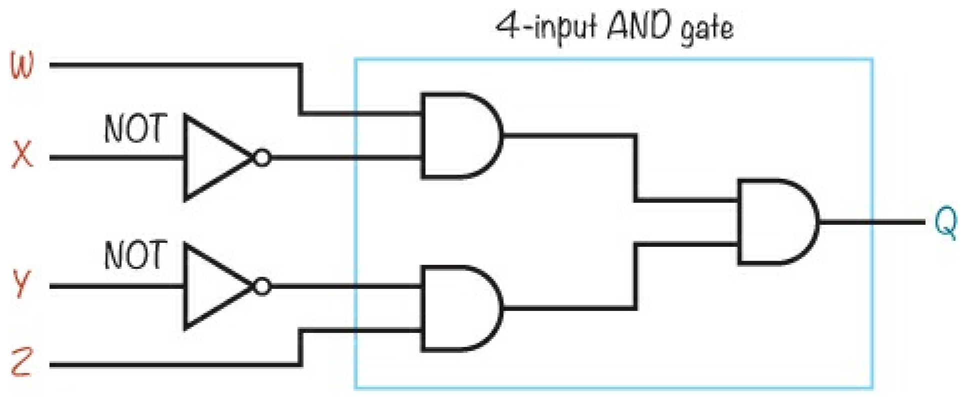

A logic circuit that checks for a secret code (1001)

The first bit (W) should be a 1 and the second (X) should be a 0, so the second bit gets a NOT gate. The third bit (Y) has a NOT gate as well, and it goes into an AND gate with the fourth bit (Z).

The first AND gate should output 1 if it sees W = 1 and X = 0, and the second AND gate should output 1 if it sees Y = 0 and Z = 1. If both of those AND gates output 1, then so will the third AND gate, which ultimately confirms that 1001 was entered.

TRY IT OUT: DRAW MORE LOGICAL STATEMENTS AS CIRCUITS

Now that you've seen the basics, transform a few logical statements of your own into logic circuits on paper. For example, how might you draw up plans for a "Can I Make Orange Juice?"

decision maker? Create one based on the statements under “It’s Only Logical” on page 214.

USING LOGIC GATES IN REAL LIFE

When I learned about logic gates, I thought they’d be small two- and three-legged components. But logic gates come boxed inside integrated circuits (ICs). Each IC contains several gates, so even if you want to use just one, you’ll have to use an IC anyway.

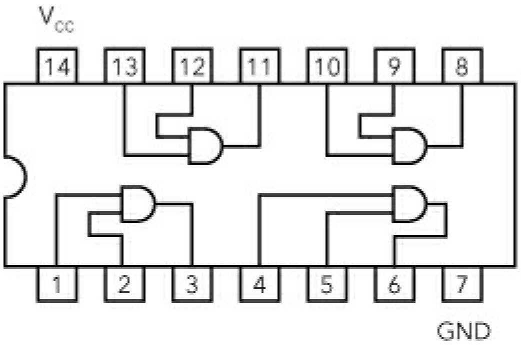

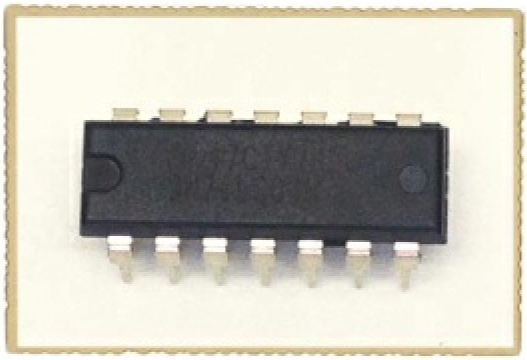

It’s also important to know that logic gate outputs don’t supply much current. Even if a logic gate outputs 5 V, that doesn’t mean that you can connect your 5 V motor to it. The logic gate simply can’t give enough current for the motor to run.

Recall from Chapter 7 that a transistor needs only a little current flowing into its base to turn on and let a lot more current flow from its collector to emitter. When you want to use a logic gate to turn on a circuit or component that requires more current, you can connect the logic gate to a transistor. Do you remember the circuit from “Project #14: Build a Circuit that Senses Touch” on page 136? You can modify this circuit to turn on an LED from a logic gate, like this:

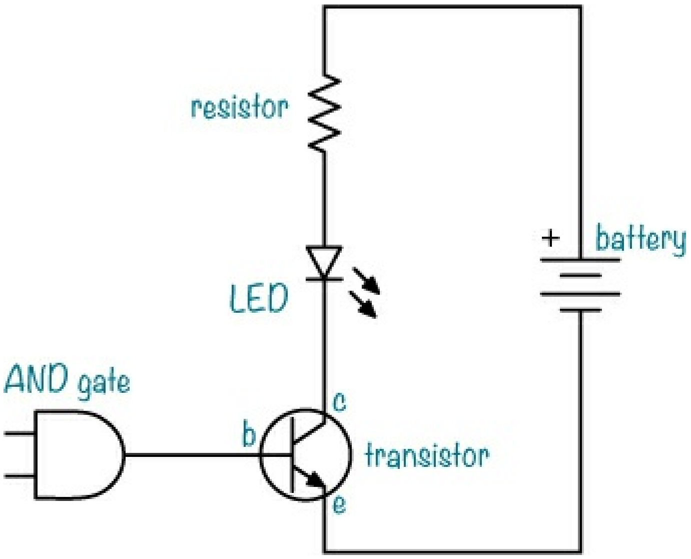

I’ll show you how to incorporate this into a project in the next section.

MORE ABOUT CURRENT, COMPONENTS, AND TRANSISTORS

You can easily replace the LED and resistor with something else that you want to control, like a motor, a fan, or a relay. But when doing this, be mindful of the current. How much current does your motor need, and how much current can the transistor handle?

Both values can be found in the components' datasheets. For transistors, the value you are looking for is called I_C, or *collector current*. According to the datasheet of a BC547 transistor, its maximum collector current is 100 mA. That's more than enough to power an LED, which usually uses about 15 to 20 mA at the most.

But what if you want to connect a motor? First, you'd need to find out how much current the motor needs, and you'll find that in the motor's datasheet. If a motor needs 500 mA, you'll need to connect it to a transistor that can handle more than 500 mA of current. For example, a PN2222 transistor can handle up to 600 mA, so it should be able to switch the motor on and off.

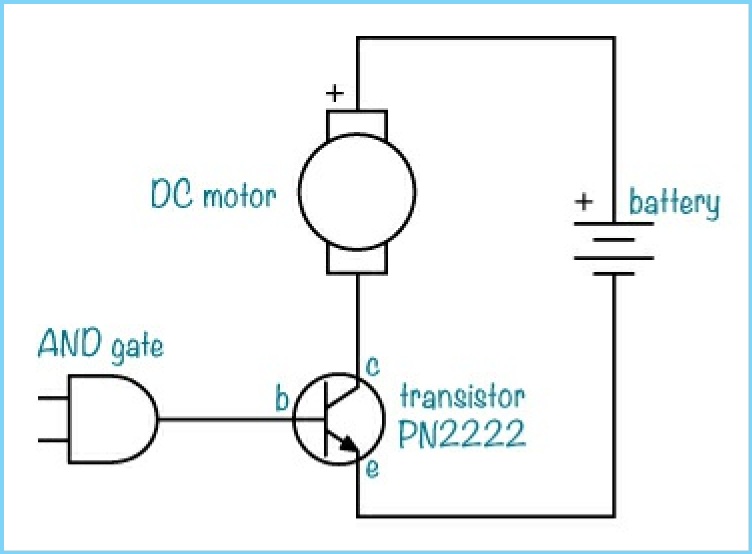

PROJECT #21: A SECRET CODE CHECKER

In this project, you'll build a logic circuit that checks whether a set of four input bits matches a secret code. You'll use four switches, inside one DIP switch, to set the code. If the input bits match the code, then the logic circuit should output a voltage, representing a 1; otherwise, it should output zero voltage, to indicate 0. This final output will go to a transistor so you can use it to control something—like an alarm!

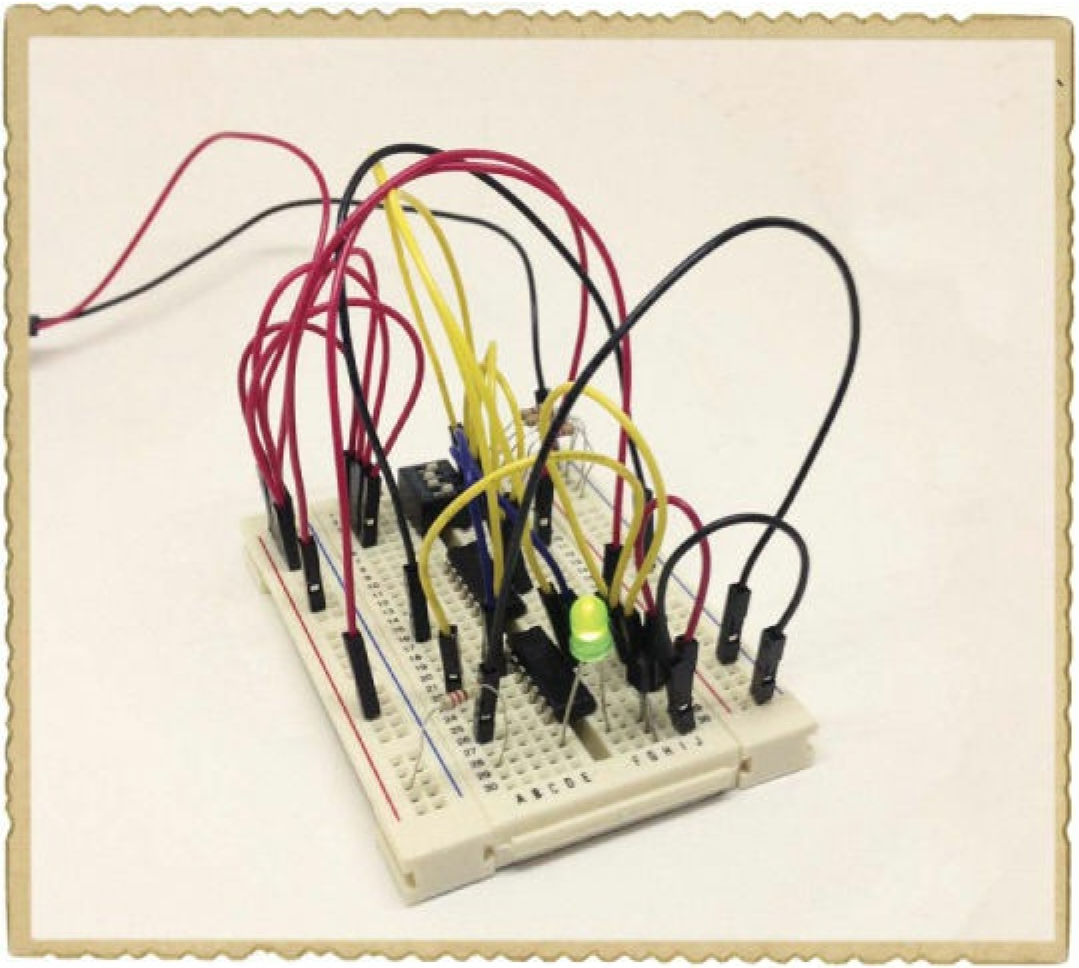

The basic Secret Code Checker circuit turns on an LED when you input the right code. At the end of the project, I'll show you how to use the Secret Code Checker to disarm your intruder alarm from Chapter 1.

Here's the complete circuit diagram for this project:

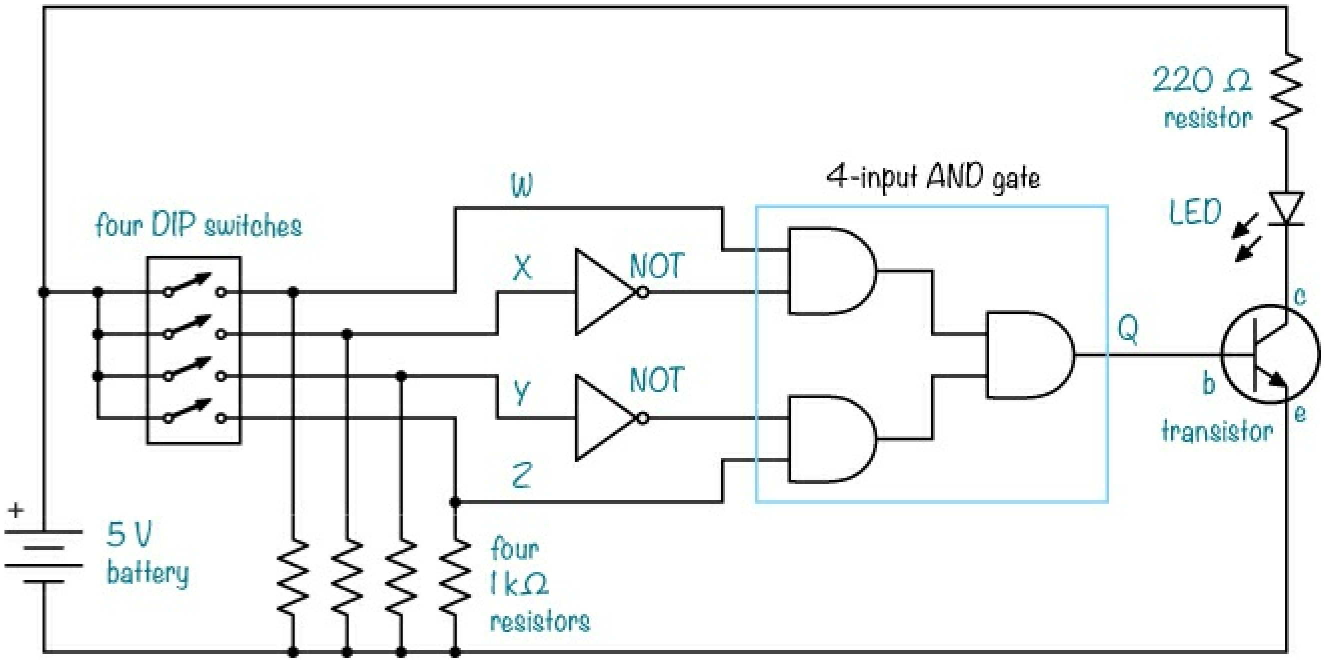
220 Ω
resistor
4-input AND gate
LED
four DIP switches
W
X
NOT
Y
NOT
Z
Q
c
b
e
transistor
+
5 V
battery
four
1 kΩ
resistors

Shopping List

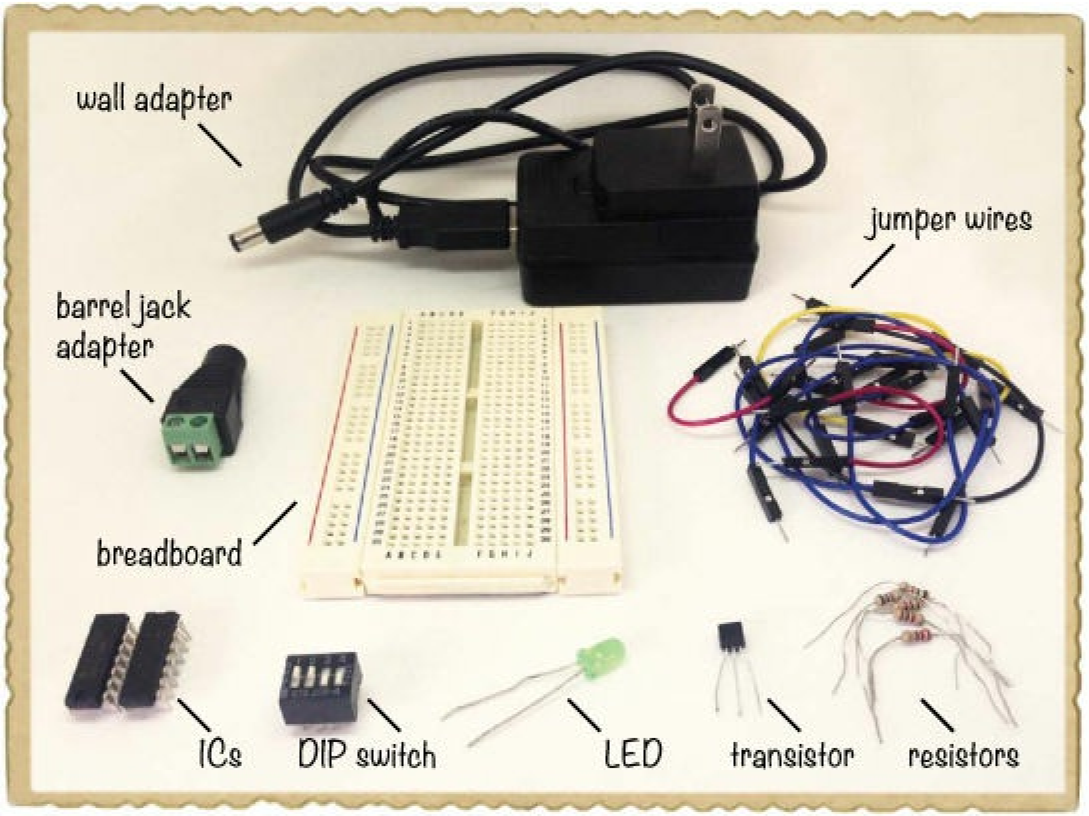
wall adapter
jumper wires
barrel jack adapter
breadboard
ICs
DIP switch
LED
transistor
resistors

- **A breadboard** (Jameco #20601, Bitsbox #CN329) with at least 30 rows.
- **Breadboard jumper wires** (Jameco #2237044, Bitsbox #CN236)—you'll need around 20 for this project.
- **A DIP switch** (Jameco #38820, Bitsbox #SW042) with four individual switches.
- **A 74LS04 inverter IC with six NOT gates** (Jameco #46316, Bitsbox #QU108)
- **A 74LS08 IC with four AND gates** (Jameco #46375, Bitsbox #QU109)
- **A general-purpose NPN transistor** (Jameco #254801, Bitsbox #QD011), such as BC547.
- **A standard LED** (Jameco #34761, Bitsbox #OP003)
- **A 220 Ω resistor** (Jameco #690700, Bitsbox #CR25220R) to limit the current to the LED.
- **Four 1 kΩ resistors** (Jameco #690865, Bitsbox #CR251K) to use as pull-down resistors.
- **A 5 V DC wall adapter** (Jameco #2126125, Bitsbox #TF010) to power the circuit.
- **A DC barrel jack adapter** (Jameco #2227209, Bitsbox #CN424) to connect the wall adapter to the breadboard.

Tools

- **A screwdriver** that fits the screw terminal of the barrel jack adapter.

How to Use Other Voltages with a Breadboard

You've used 9 V batteries for each circuit in this book so far, but most digital circuits need to use lower voltages. For example, a lot of ICs with logic gates inside use 5 V instead. But 5 V is not a standard battery value; there are 4.5 V and 6 V batteries, but not 5 V.

What can you do when your circuit requires 5 V? Say hello to the *wall adapter* and the *barrel jack adapter*.

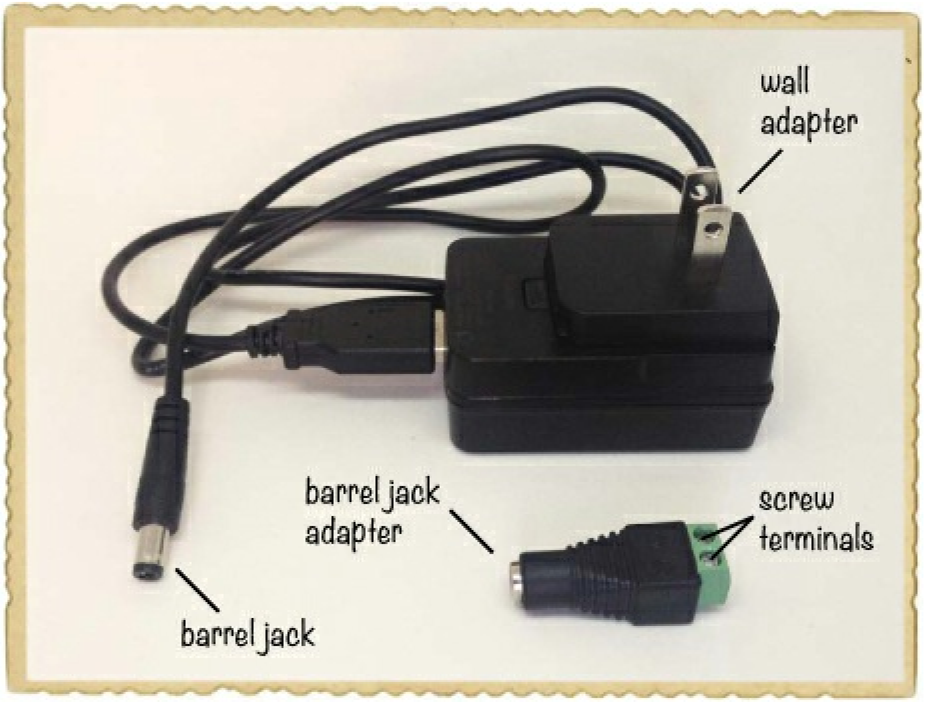

Many electronic devices use wall adapters to recharge batteries or just to stay powered. The pronged side of a wall adapter plugs into a wall socket, and the other side plugs into something you want to power. Wall adapters come in many values, and this project uses a 5 V DC regulated adapter.

To supply power to a circuit, the wall adapter in this project needs to plug into a barrel jack adapter. The barrel jack adapter listed in this project's Shopping List (page 224) has two *screw terminals*, where you can plug in jumper wires that connect to the breadboard. You can connect any wall adapter with the standard round plug into this barrel jack adapter.

MAKING CIRCUITS MORE RELIABLE

If a circuit needs an input voltage and you don't connect the input to anything, then that input is *floating*. A floating input is unreliable because the circuit may see it as a 1 or a 0, and you can't control which.

The individual switches on a DIP switch are either open or closed. When the switch is open, the input it controls will be floating if it's not connected to anything else. To fix this, you can attach a *pull-down resistor* to each input on the logic gates, like this:

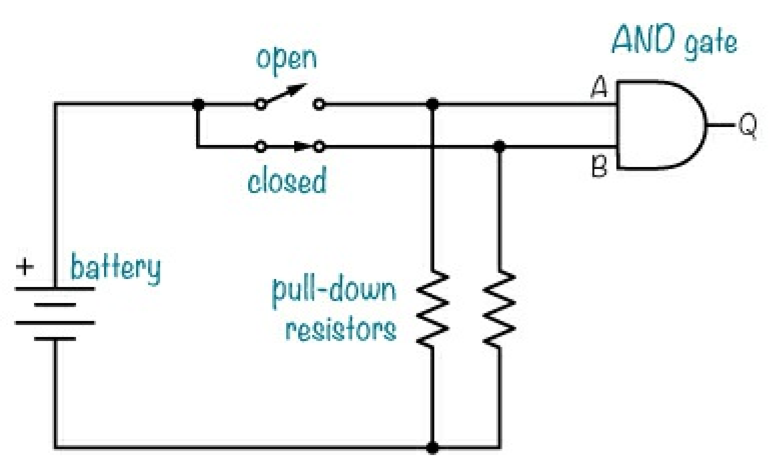

Each pull-down resistor in this circuit diagram connects to a switch and a gate input on one side and to the negative battery terminal on the other. When a switch is open, the resistor "pulls" the gate input down to 0 V, which is a 0. When a switch is closed, the gate input connects to the positive terminal and gets the positive supply voltage, which is a 1.

In this project's circuit diagram on page 224, there are four switches with 1 kΩ pull-down resistors. All of them are shown as open, and all of the AND gate inputs would be 0 in that state.

Step 1: Place the Switches and Resistors

Plug your DIP switch in at the top of the breadboard, with one side of the switch on each side of the notch in the middle. Use jumper wires to connect the left side of each DIP switch to the positive supply rail on the left, and connect a 1 kΩ resistor from the right side of each DIP switch to the negative supply rail on the right.

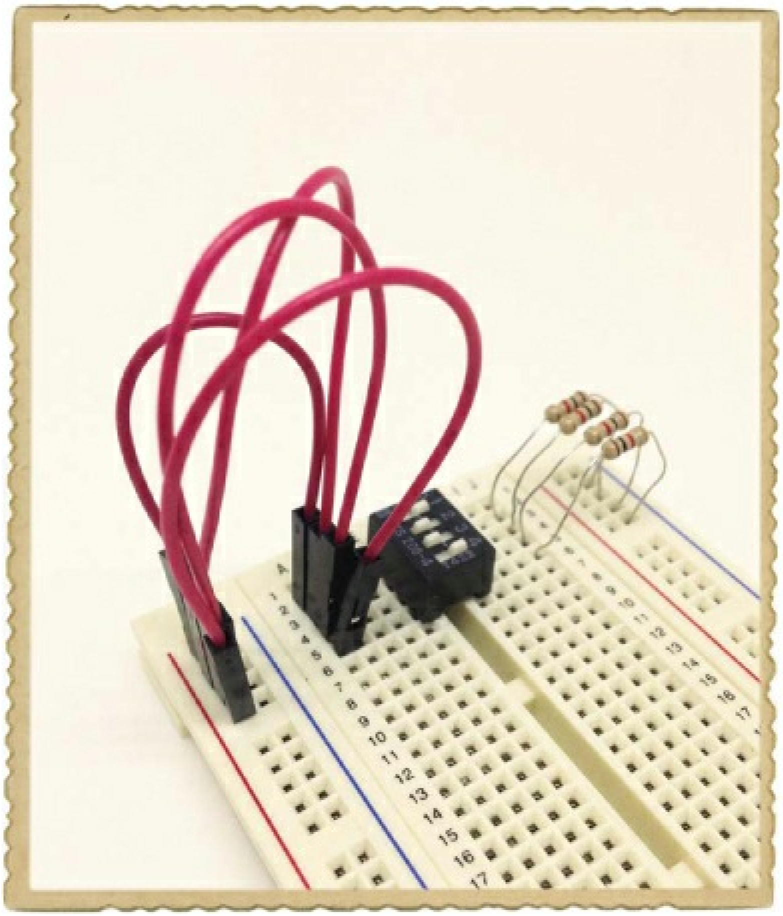

Step 2: Place the ICs

Place the IC with NOT gates, marked *74LS04*, in the middle of the breadboard and place the IC with AND gates, marked *74LS08*, farther down. For both ICs, point the rounded notch toward the DIP switch. Leave at least three rows at the bottom of the breadboard for the transistor.

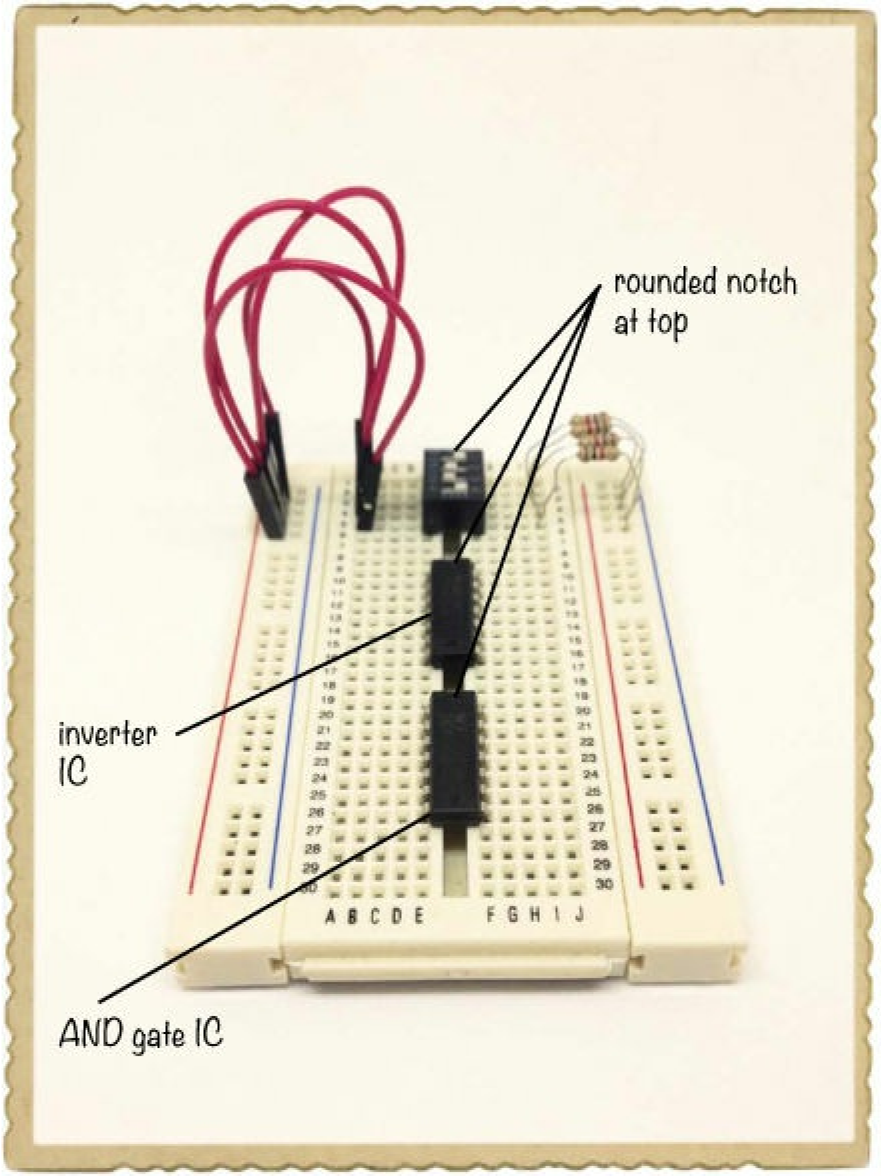

Step 3: Place the Transistor and LED

Plug your transistor into three rows at the bottom of the breadboard. If you used the BC547 transistor from this project’s Shopping List (page 224), face the flat side left so that the collector is the upper pin, the base is the middle pin, and the emitter is the bottom pin. If you used a different NPN transistor, check its datasheet to see which pin is which.

Connect the LED’s short leg, the cathode, to the same row as the collector. Connect the LED’s long leg, the anode, to an empty row on the left side of the breadboard. Finally, connect the 220 Ω resistor from the LED’s anode to the positive supply rail.

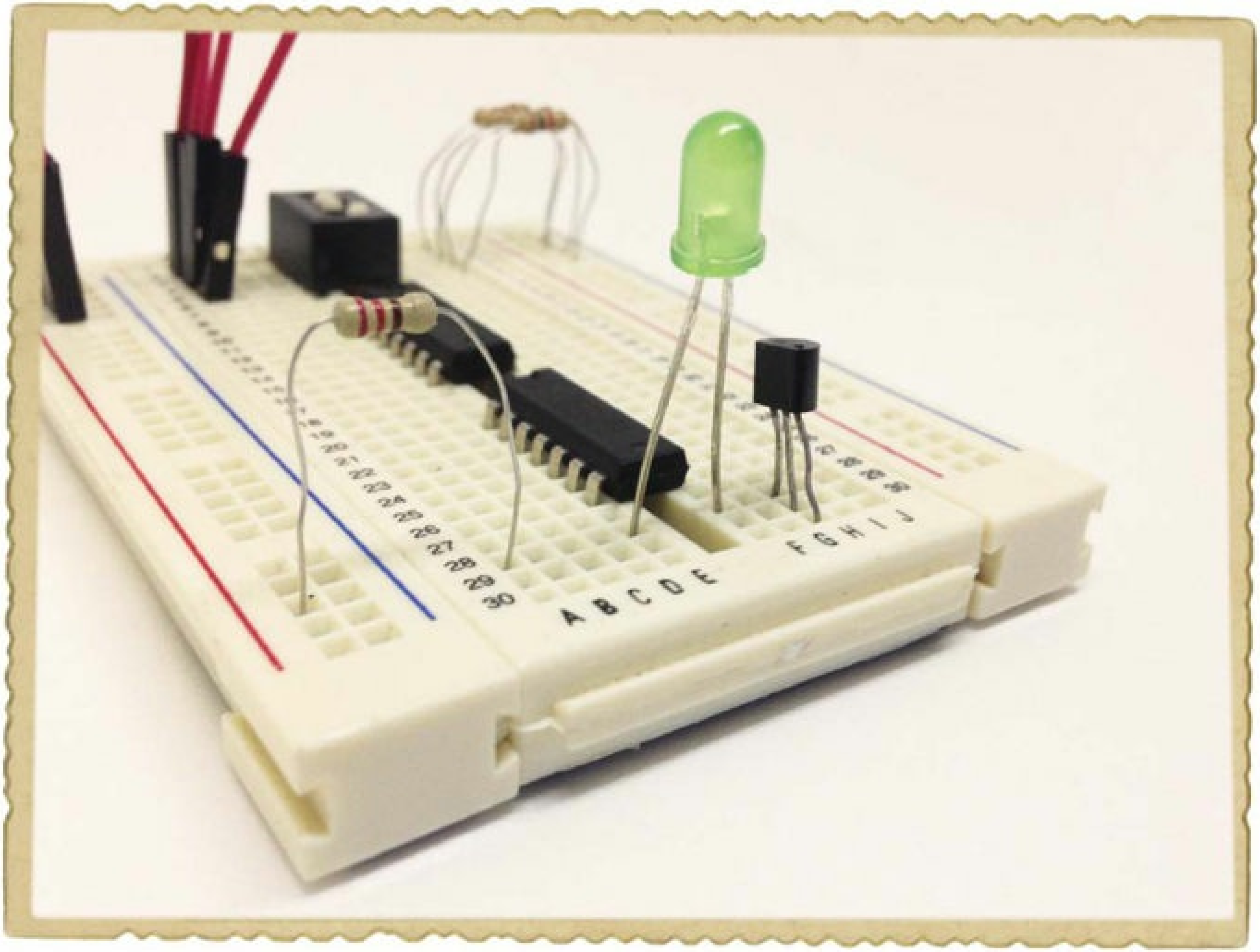

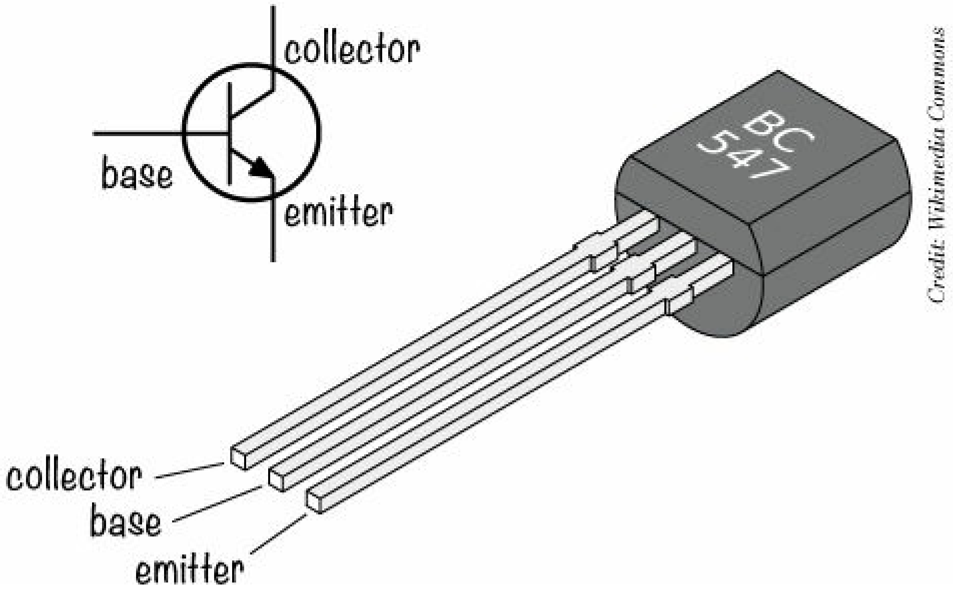
collector
base
emitter
BC
547
collector
base
emitter

Credit: Wikimedia Commons

Step 4: Build the Logic Circuit

First, look at the following diagram to see where the AND and NOT gates are inside your ICs and to see the connections you need to make.

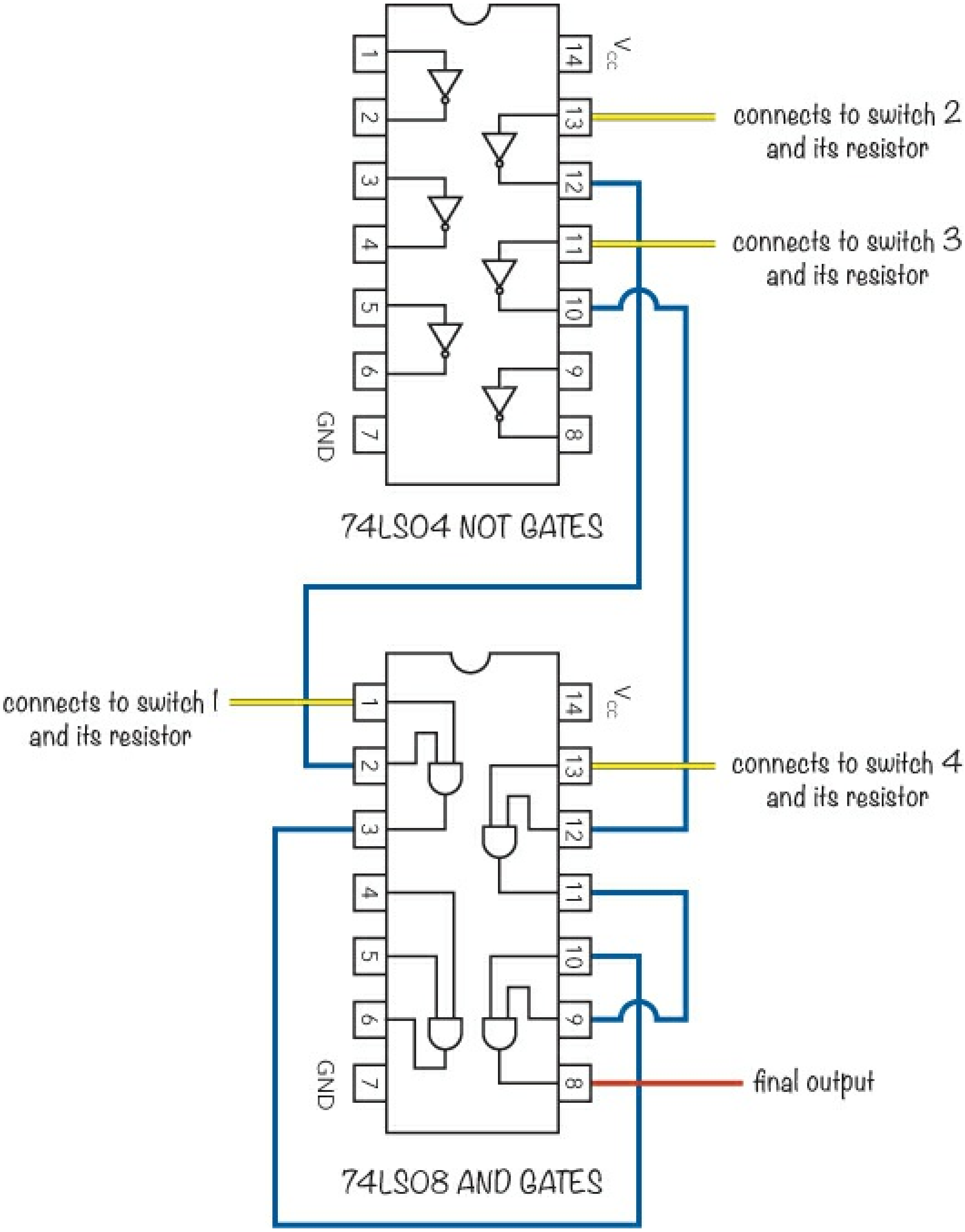
V_{cc}
connects to switch 2 and its resistor
connects to switch 3 and its resistor
GND
74LS04 NOT GATES
connects to switch 1 and its resistor
connects to switch 4 and its resistor
final output
74LS08 AND GATES
1 2 3 4 5 6 7 8 9 10 11 12 13 14

Take four jumper wires and connect them from the switch outputs to the gate inputs as follows:

- The output from switch 1, the uppermost switch, goes to the AND gate input on pin 1 of the lower IC (74LS08).
- The switch 2 output goes to the NOT gate input on pin 13 on the upper IC (74LS04).
- The switch 3 output goes to the NOT gate input on pin 11 on the upper IC.
- The switch 4 output goes to the AND gate input on pin 13 of the lower IC.

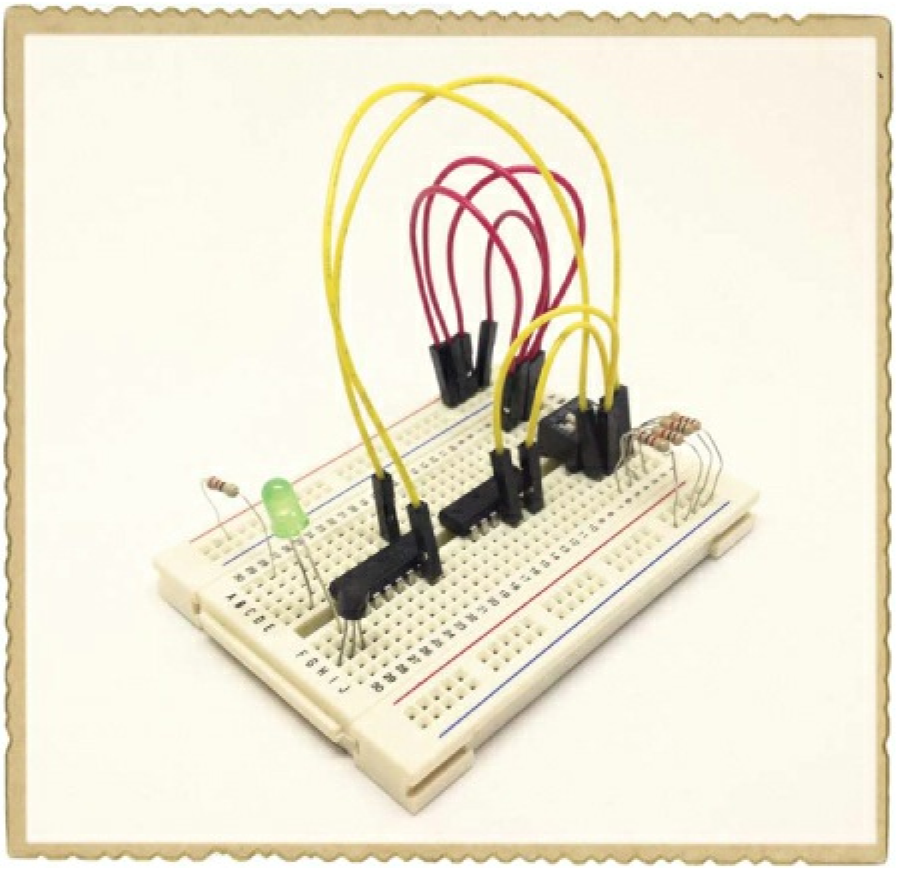

Next, get two jumper wires to connect the outputs of the NOT gates to the AND gate inputs like this:

- One jumper wire goes from pin 12 on the upper IC to pin 2 on the lower IC.
- The other jumper wire goes from pin 10 on the upper IC to pin 12 on the lower IC.

Now, each output from the AND gates needs to go into the third AND gate, as follows:

- Connect one jumper wire from pin 3 to pin 10 of the lower IC.
- Connect another jumper wire from pin 11 to pin 9 on the lower IC.
- Finally, connect a jumper wire from the final output of the AND gate with the other side hanging loose for now.

Your IC connections should look like this:

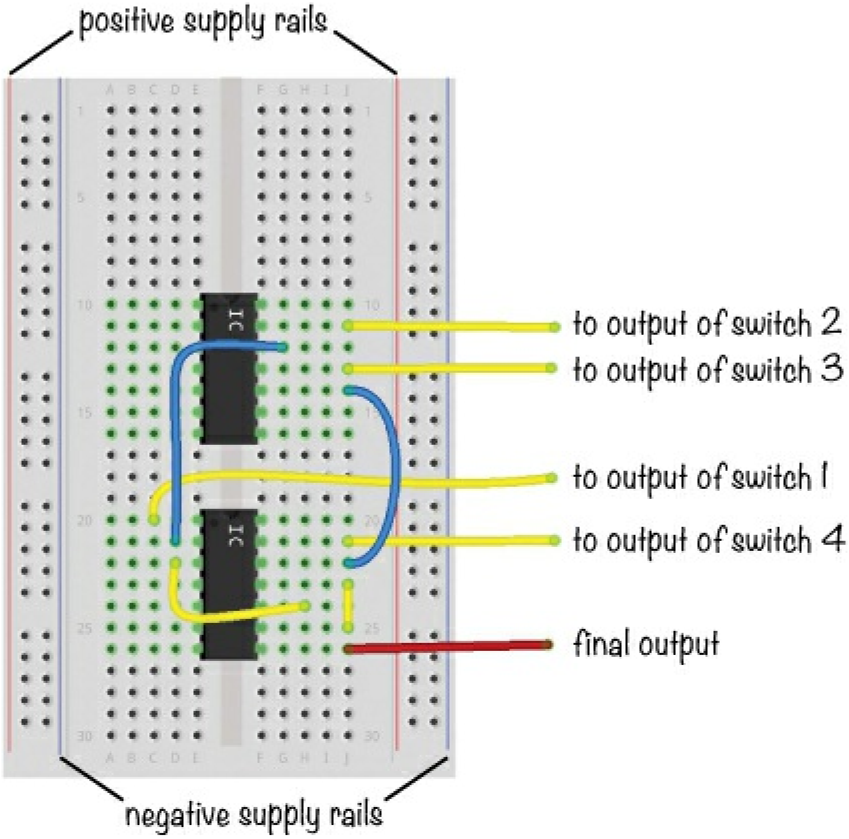
positive supply rails
IC
IC
to output of switch 2
to output of switch 3
to output of switch 1
to output of switch 4
final output
negative supply rails

Step 5: Finish Wiring the Transistor

Now, connect the output from the final AND gate—pin 8 of the lower IC—to the base of the transistor. This output will control whether the transistor allows current to pass through to the LED or not. Connect a jumper wire from the emitter of the transistor to the negative supply rail.

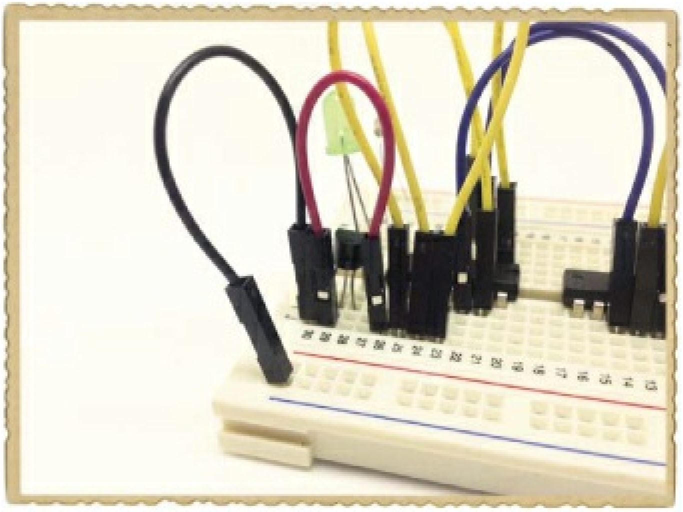

PROTECTING YOUR COMPONENTS

The circuit for this project relies on the fact that gates supply only a tiny amount of current to the transistor base. Your transistor will be perfectly fine while you use your secret code checker, but to protect the transistor base from damage by larger currents in other circuits, you should place a resistor around 1 to 10 kΩ between the transistor base and its current source—in this case, the AND gate output.

Transistors also should usually be protected in motor circuits, like the one in “More About Current, Components, and Transistors” on page 222. That circuit should work as is, but to be really careful, it’s good to connect a diode across the motor, with the cathode on the positive side. That should protect the transistor from any high-voltage spikes that might happen when the motor is switched off.

Step 6: Power and Test the Secret Code Checker

Connect jumper wires from pin 14 on both ICs to the positive supply rail and from pin 7 of both chips to the negative supply rail. Then, turn off all the switches on the DIP switch and connect your 5 V source, with plus to the positive supply on the left and minus to the negative supply on the right.

Use your barrel jack adapter with a couple of jumper wires to make this connection. The barrel jack adapter should have + and – markings to tell you which supply is which. Just loosen the screws on the adapter, insert a wire into each, and tighten the screws again. Follow the conventional color code by using a red wire for positive and a black wire for negative so you will be sure to connect them the correct way on your breadboard.

The LED should stay dark when the switches are off, but when you set the code to 1001 by switching on the top and bottom switches, it should light up.

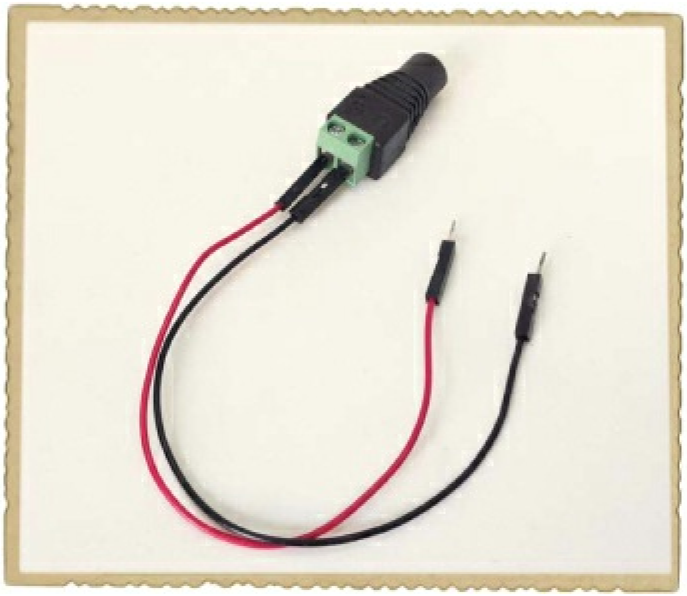

Step 7: What If the LED Doesn’t Light Up?

First, check that the two ICs have power. Do both ICs have pin 14 connected to the positive supply column and pin 7 connected to the negative supply column? If you find the ICs become too hot to touch, disconnect the 5 V power supply from the wall immediately and wait for the ICs to cool down. Then, make certain you have the supply connections the correct way before trying again. The wire plugged into the barrel jack adapter’s positive (+) terminal should be plugged into the positive supply column on the breadboard, and the wire in the barrel jack adapter’s negative (–) terminal should be plugged into the negative supply column.

If the ICs are powered and the circuit still doesn’t work, then check the input values on the switches. Use a multimeter to measure the voltage from the negative supply rail to the pins on the

AND and NOT gates that take inputs from the switches. You should get 5 V on pins 1 and 13 of the AND IC and 0 V on pins 11 and 13 of the NOT IC. Check that the output from each AND gate you're using is 5 V, too; you should see 5 V on pins 3, 8, and 11. If any AND gate in the IC doesn't output 5 V, then one of its inputs is 0 V. Figure out why it's 0 V, and you should find the problem.

TRY IT OUT: DISARM YOUR INTRUDER ALARM

Instead of an LED and a resistor, you can connect a relay to the Secret Code Checker and combine this project with the intruder alarm you built in Chapter 1. Connect the 9 V battery to the intruder alarm through the relay so that when you input the right code, the power to the alarm is cut and the noise stops. Refer back to "Meet the Relay" on page 97 to see how to connect a relay.

Notice that the secret code checker, with its 5 V supply, is being used to control the completely separate intruder alarm circuit with a 9 V supply. Connecting two circuits with separate power supplies this way is okay because there's no electrical connection between the two circuits. Relays are useful when you need to control a circuit with a different type of power supply!

Here's the circuit:

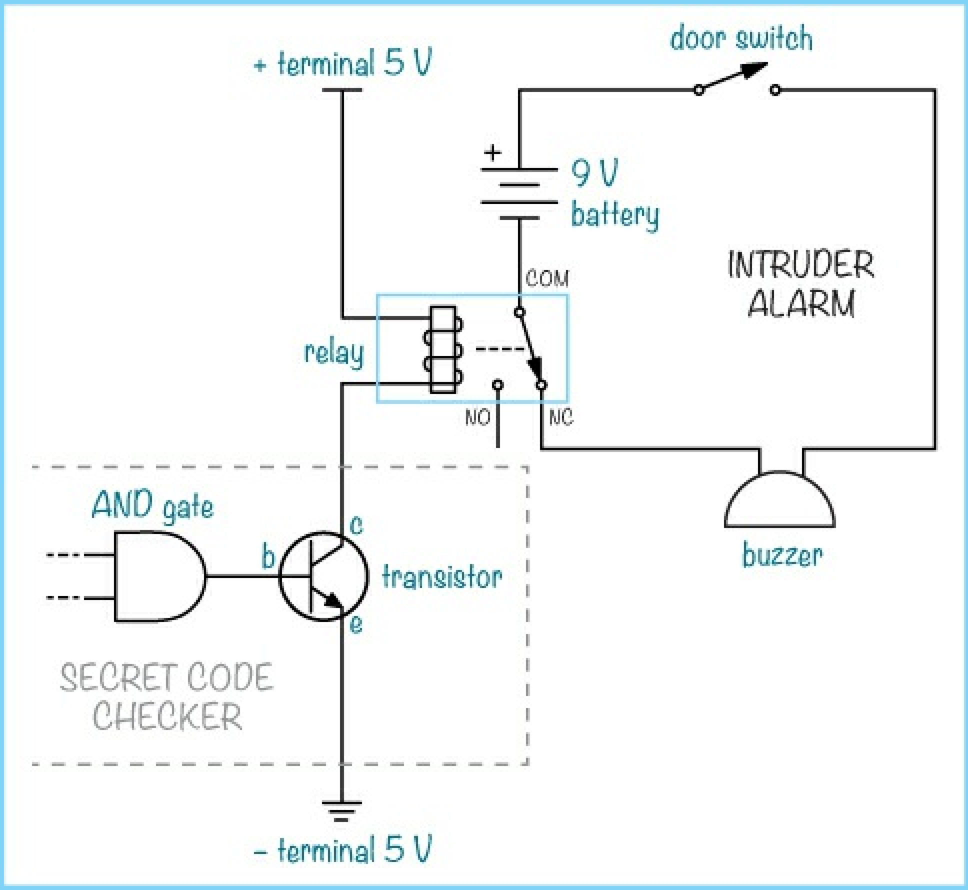

And these are the components you'll need:

- **The circuit from "Project #2: Intruder Alarm" on page 11**
- **The circuit from "Project #21: A Secret Code Checker" on page 223**
- **A 5 V relay** (Jameco #842996, Bitsbox #SW073)

NEGATIVE LOGIC GATES

AND, OR, and NOT are basic logic gates, and you can combine them to create new ones. Let's look at two more gates that are created this way.

NAND Looks for One False Input

The *NAND gate* works like an AND gate with a NOT gate inverter on the output. The little circle on the output means NOT. That means the output from the NAND gate is 0 when both A and B are 1.

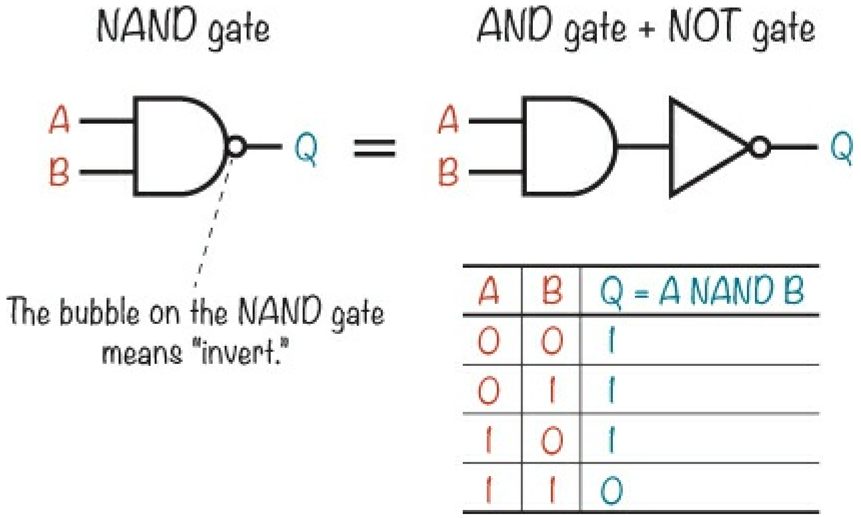

The bubble on the NAND gate means "invert."

A	B	Q = A NAND B
0	0	1
0	1	1
1	0	1
1	1	0

NOR Looks for Two False Inputs

The *NOR gate* works like an OR gate with an inverter on the output. The output is 1 when both A and B are 0.

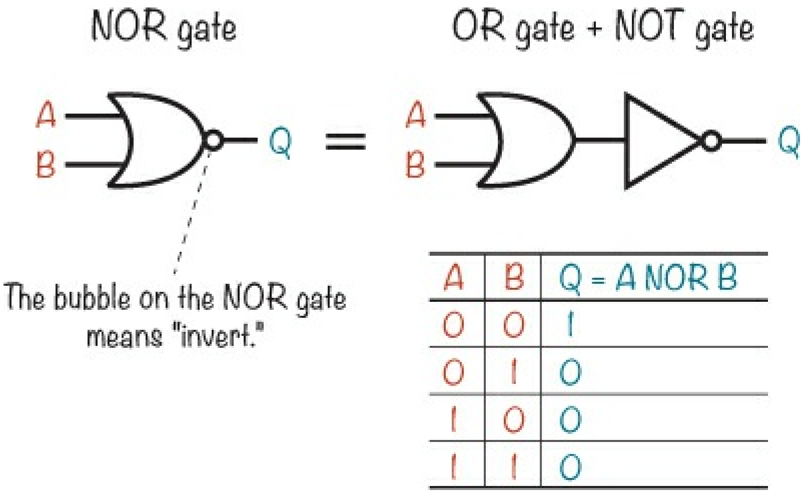

The bubble on the NOR gate means "invert."

A	B	Q = A NOR B
0	0	1
0	1	0
1	0	0
1	1	0

WHAT'S NEXT?

In this chapter, you learned how to use logic gates to build circuits that "decide" things, like whether a code is correct or not. And at the end, you got to see some negative logic gates as well. Understanding how negative logic gates work is helpful, because they are often used in real-life circuits. In fact, you'll use them in Chapter 11.

If you want to explore gates a bit further, I suggest you try combining some logic gates you've learned about on paper to create an *XOR gate*. An XOR gate gives out 1 only if the inputs are different from each other.

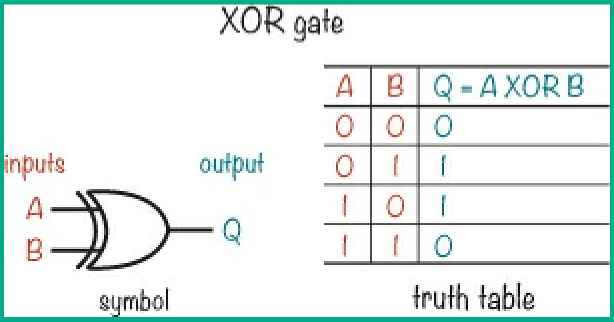

A	B	Q = A XOR B
0	0	0
0	1	1
1	0	1
1	1	0

By combining logic gates in different ways, you can create almost anything you can imagine. But that might be a bit hard to see right now, so in the next chapter, I'm going to show you some more building blocks you can create with logic gates. You'll learn how to build your own memory circuit, and then you'll build your own electronic coin tosser!

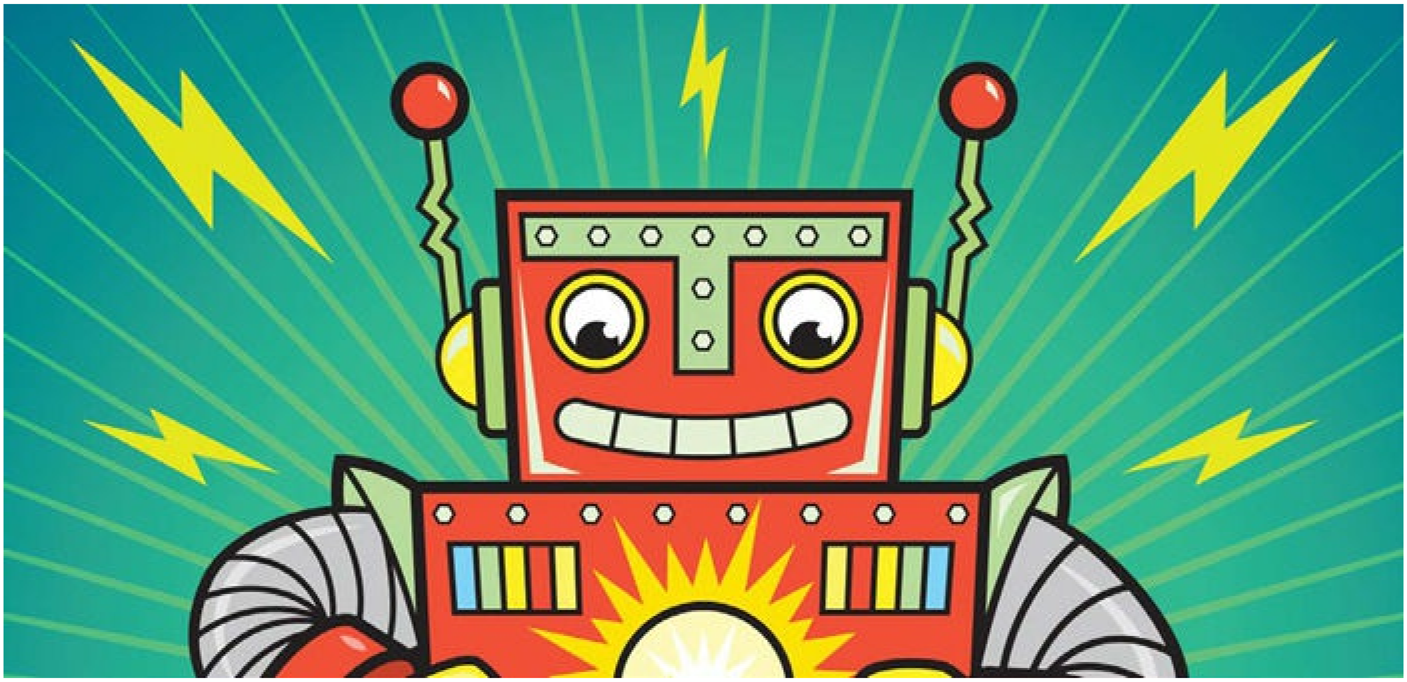

11

CIRCUITS THAT REMEMBER INFORMATION

In Chapter 9, you learned how to store bits using switches. As long as the switches don't change, the bits stay the same. But you have to manually set the switches, and that's not very efficient. In Chapter 10, you learned about logic gates and how you can use them to play with ones and zeroes. Now, I'll show you how to use logic gates to make electronic memory that saves bits, even after you change the input. At the end of this chapter, you'll build your own electronic coin tosser!

SAVING ONE BIT AT A TIME

One simple memory circuit is an *SR latch*. You can create an SR latch with two NOR gates, and one latch can store one bit of data.

SR latch

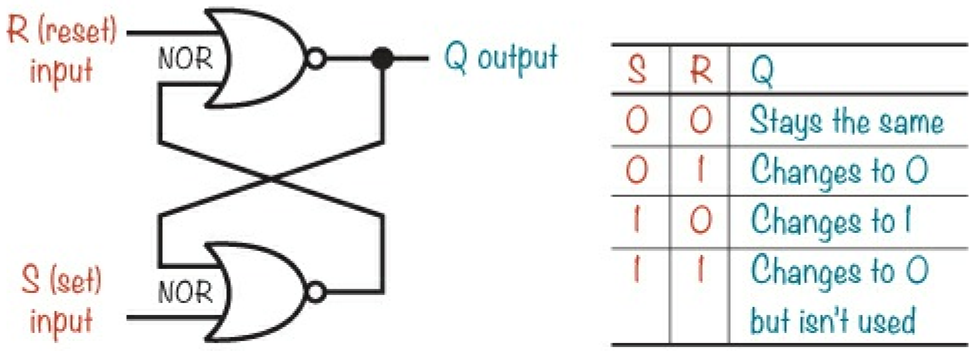

S	R	Q
0	0	Stays the same
0	1	Changes to 0
1	0	Changes to 1
1	1	Changes to 0 but isn't used

symbol

truth table

Q starts at 0, and this circuit won't update its output until you use R or S to tell it to, which is called *latching* the bit. S and R stand for *set* and *reset*, respectively: when you set the latch, Q changes to 1. When you reset it, Q changes to 0. As the SR latch truth table describes, you can set the latch by putting a 1 on the S input and a 0 on the R input. To reset it, you'd put a 1 on the R input and a 0 on the S input. Let's look at how setting Q to 1 works.

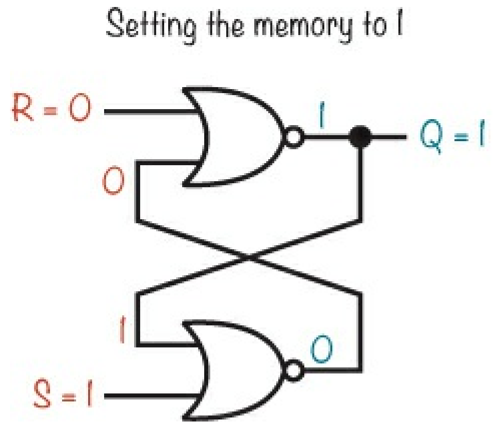
Setting the memory to 1

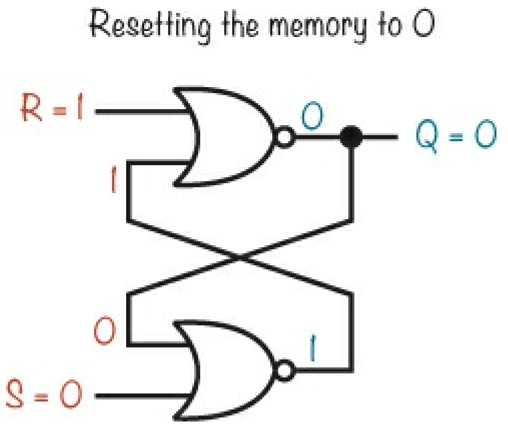
Resetting the memory to 0

The output from a NOR gate is 1 when all inputs are 0. When S (set) is 1, the output of the lower NOR gate is 0, no matter what that gate's other input is. The output is connected to an input on the other NOR gate, together with R (reset). Because R is 0, you have two 0s into the upper NOR gate, which makes Q output 1.

A BETTER MEMORY CIRCUIT

If you add a few more gates to the SR latch, you can create a *D latch*, which sets the output Q to whatever the D input is when the C input is 1.

D latch

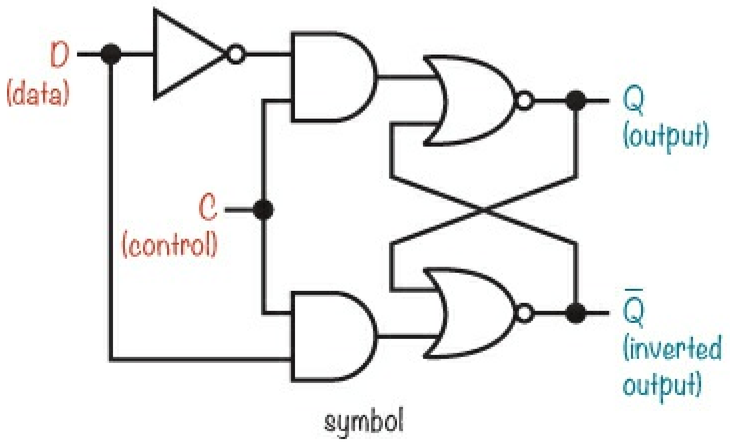

symbol

D	C	Q
0	0	Doesn't change
0	1	Changes to 0
1	0	Doesn't change
1	1	Changes to 1

truth table

The D latch is an improvement over the SR latch because you can change D, the *data input*, as much as you want, and Q won't change unless you set C, the *control input*, to 1. The lower NOR gate output will always be the opposite of the output Q, and that output is labeled $\bar{Q}$ to show this.

MEMORY THAT CHANGES ONLY AT A CERTAIN TIME

The D latch has one weakness: when C is 1, changing D also changes the output Q. What if you don't want the output to change immediately?

Computers use a *clock signal* to tell the circuits inside when something should happen, like when to store new data from a wire. A clock signal is just a voltage that turns on and off continuously—that is, it keeps switching between 1 and 0. This signal is similar to the one you sent to the speaker in "Project #16: Make Your Own Sound with the 555 Timer" on page 167.

To reduce the chance of errors, actions like calculations or storing data happen only when the clock signal switches from off to on or from on to off. This is called *edge-triggering*. If a circuit does something when the clock signal changes from off to on, then that action happens on the rising edge, and the circuit is *positive edge-triggered*. A circuit that triggers an action on the falling edge, when the clock changes from on to off, is *negative edge-triggered*.

A *flip-flop* is a latch that updates its output when triggered by the edge of a clock signal, and you can create one by combining two D latches and a NOT gate. This is called a *D flip-flop*.

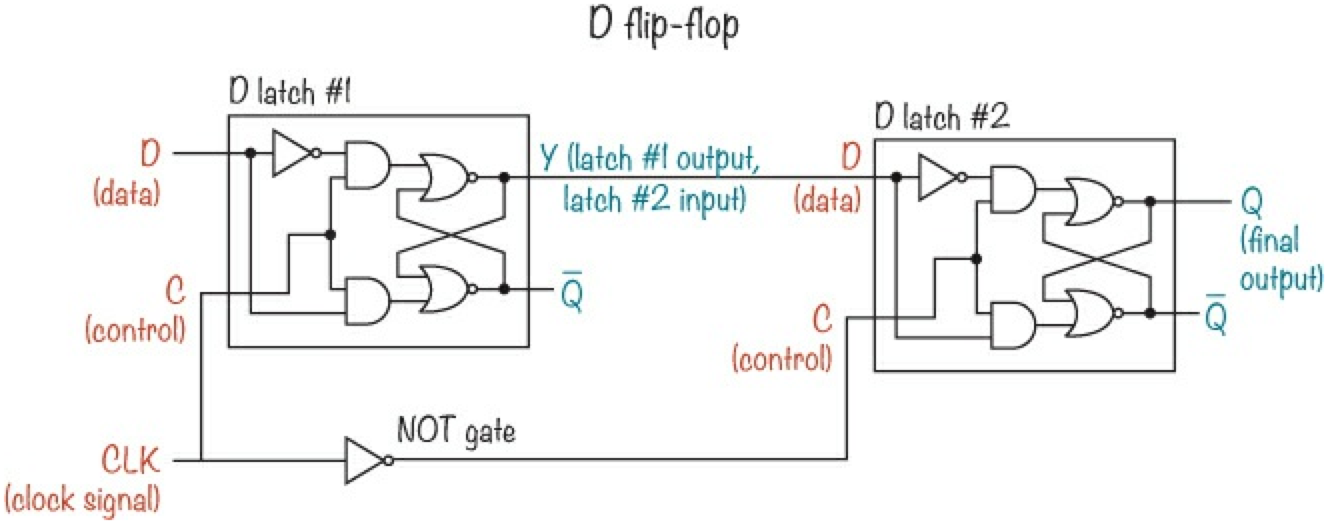
D flip-flop
D latch #1
D latch #2
D
(data)
C
(control)
CLK
(clock signal)
Y (latch #1 output,
latch #2 input)
$\bar{Q}$
D
(data)
C
(control)
NOT gate
Q
(final
output)
$\bar{Q}$

The output voltage Q can change only when the CLK voltage changes from high to low, from 1 to 0. Here's how that works.

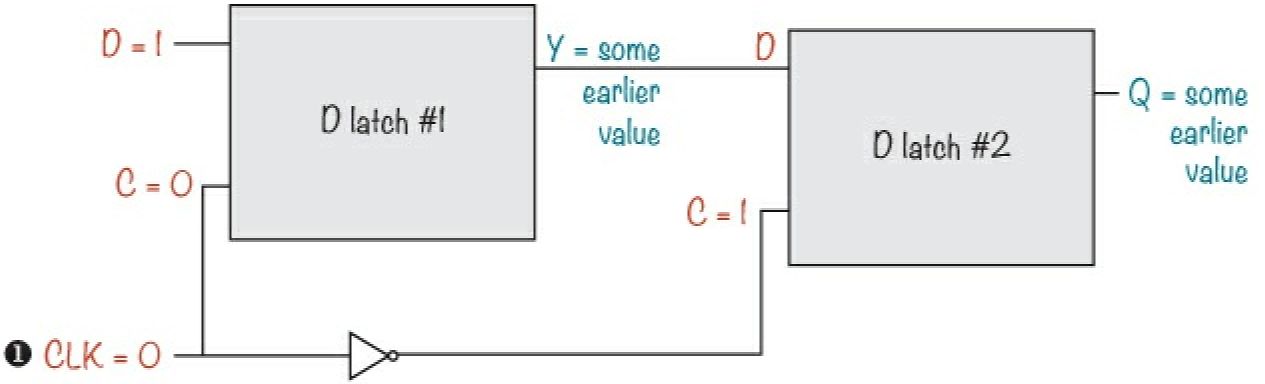
D = 1
D latch #1
C = 0
❶ CLK = 0
Y = some earlier value
D
D latch #2
C = 1
Q = some earlier value

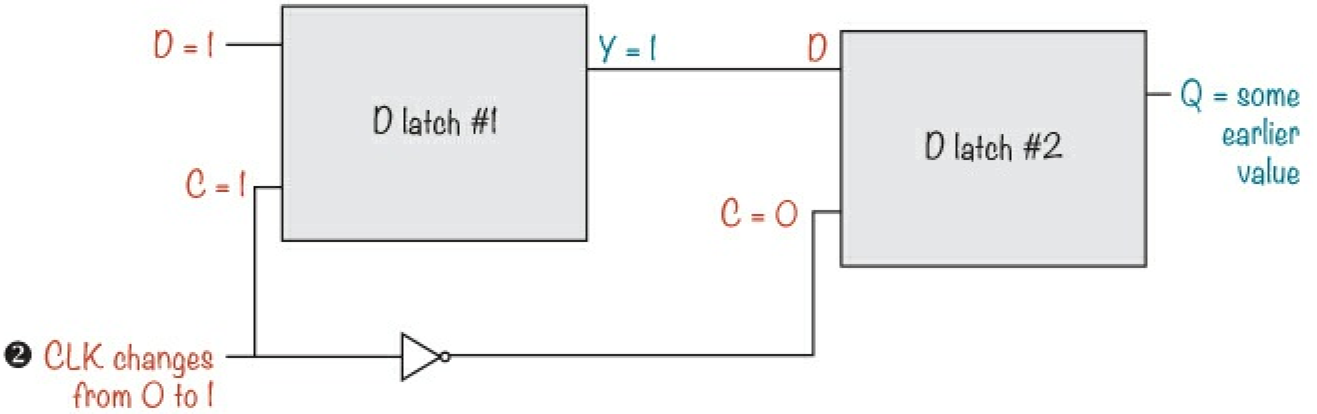
D = 1
D latch #1
C = 1
❷ CLK changes from 0 to 1
Y = 1
D
D latch #2
C = 0
Q = some earlier value

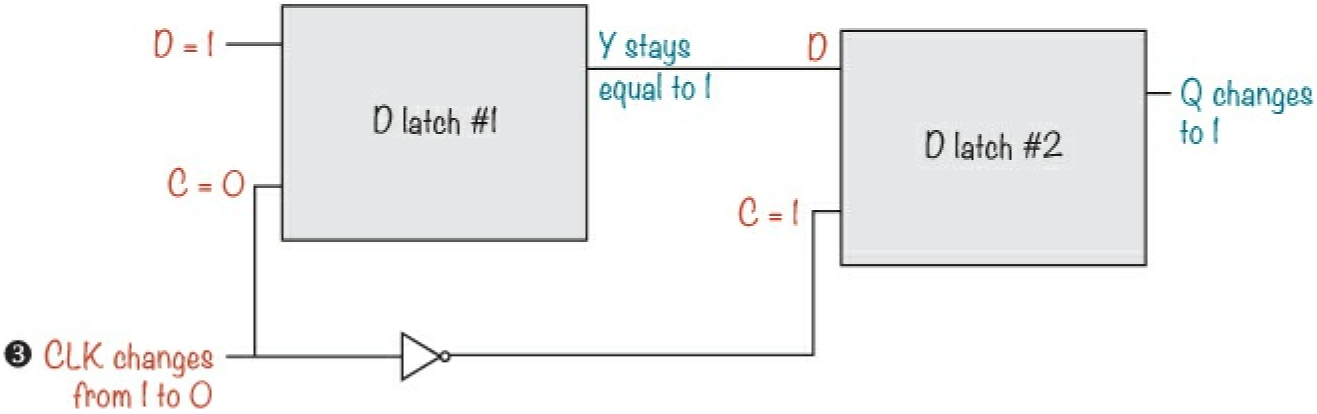
D = 1
D latch #1
C = 0
❸ CLK changes from 1 to 0
Y stays equal to 1
D
D latch #2
C = 1
Q changes to 1

When CLK is 0, Y and Q don't change ❶. When CLK changes to 1 ❷, Y changes to match the D input to D latch #1. But the NOT gate inverts the 1, setting C on D latch #2 to 0 so Q doesn't change. As CLK goes back to 0 ❸, C on D latch #2 changes to 1, the value on Y is saved in latch #2, and Q changes to match Y.

NOTE

Q updates when the clock signal changes from high to low voltage, so this flip-flop is negative edge-triggered.

This is the circuit diagram symbol for a positive edge-triggered D flip-flop:

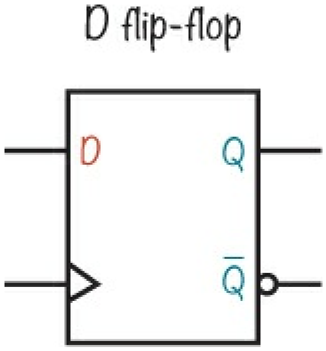
D flip-flop
D
Q
$\overline{Q}$

Instead of writing CLK for the clock input, the D flip-flop symbol has a > marking. Notice the bubble on the $\overline{Q}$ output. Just like on the NOT gate symbol, the bubble means $\overline{Q}$ is the inverted version of Q.

AN OUTPUT THAT TOGGLES

With a simple wire, you can turn a D flip-flop into a circuit that toggles another circuit on and off. For example, imagine you want to turn a light on and off. The D flip-flop stores a value on its input, D, when its clock goes from low to high voltage (if positive edge-triggered). If you connect the D flip-flop's inverted output $\overline{Q}$ to D, the flip-flop's input will always be the opposite of Q. Every time the clock input triggers, the output will change to the opposite value, and the light will toggle.

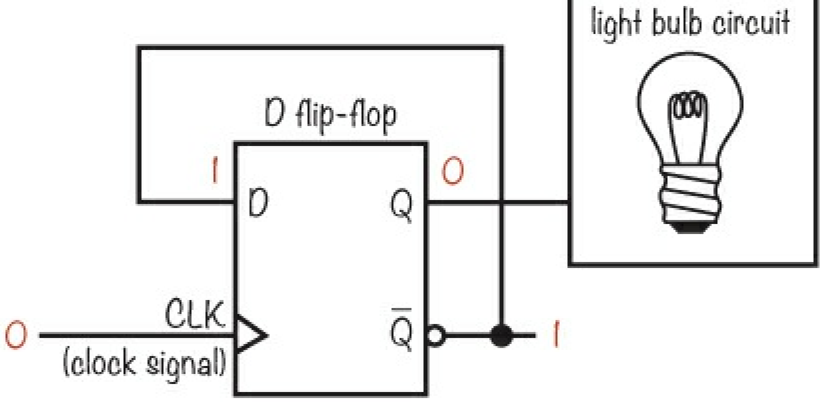
We start with these values:
light bulb circuit
D flip-flop
1
D
0
Q
0
CLK
(clock signal)
$\overline{Q}$
1

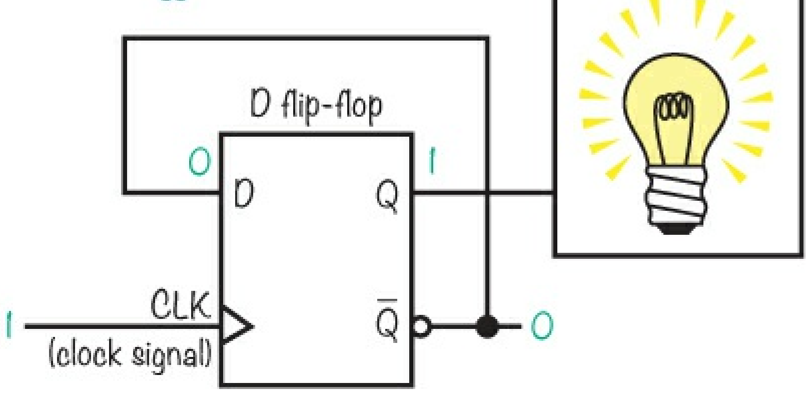
When the clock input changes from 0 to 1, the values are toggled:
D flip-flop
0
D
1
Q
1
CLK
(clock signal)
$\overline{Q}$
0

Let's see this concept in action!

PROJECT #22: AN ELECTRONIC COIN TOSSER

For this project, you'll build an electronic coin tosser with a 555 timer, a D flip-flop, a push button, and two LEDs.

In Chapter 8, you built several 555 timer circuits that switched voltages on and off. A circuit that turns a voltage on and off continuously is called an *oscillator*, and in this project, you're going to use an oscillator circuit as an input to the toggling D flip-flop. Do you recognize the oscillator circuit in this diagram?

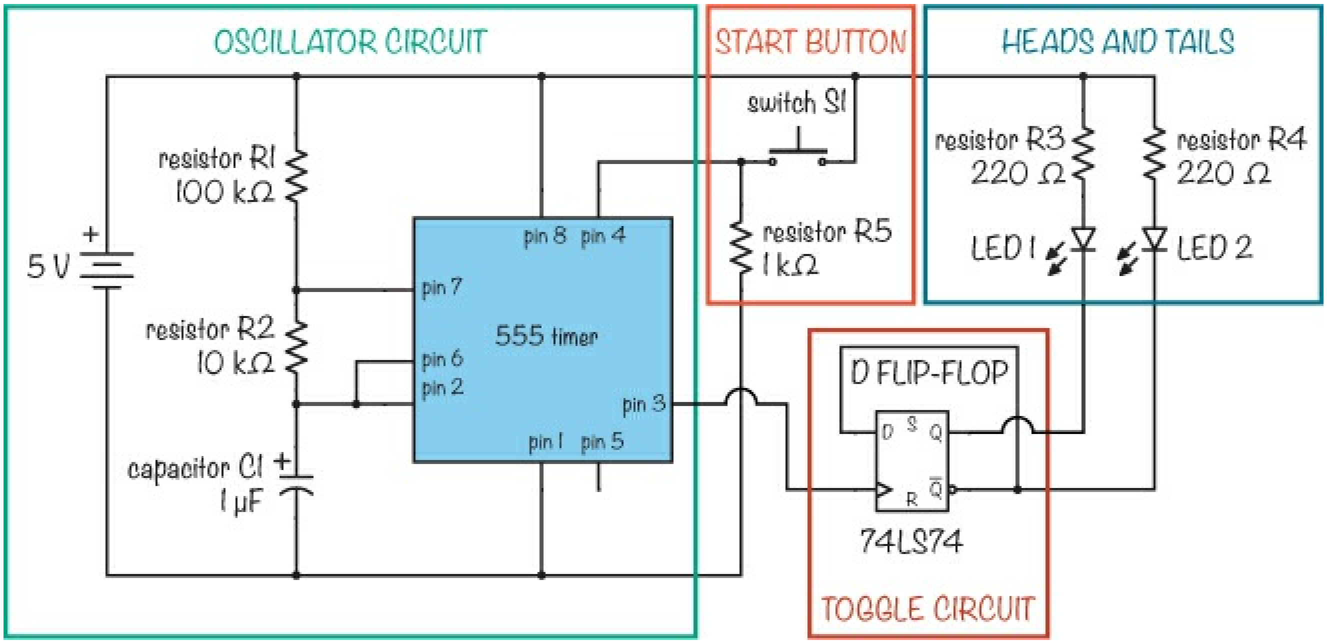
OSCILLATOR CIRCUIT
START BUTTON
HEADS AND TAILS
switch S1
resistor R1
100 kΩ
+
5 V
resistor R2
10 kΩ
capacitor C1
1 µF
pin 8
pin 4
pin 7
555 timer
pin 6
pin 2
pin 3
pin 1
pin 5
resistor R5
1 kΩ
resistor R3
220 Ω
resistor R4
220 Ω
LED 1
LED 2
D FLIP-FLOP
D
S
Q
R
Q̄
74LS74
TOGGLE CIRCUIT

The 555 timer oscillator circuit creates a clock signal that goes to the D flip-flop, causing the output to switch on and off, or *toggle*, continuously as long as you press a push button. The changing output from the D flip-flop turns the LEDs on and off.

When you let go of the push button, the clock signal from the 555 timer stops. The flip-flop output will stop alternating, and only one of the two LEDs will be on: one for heads or the other for tails.

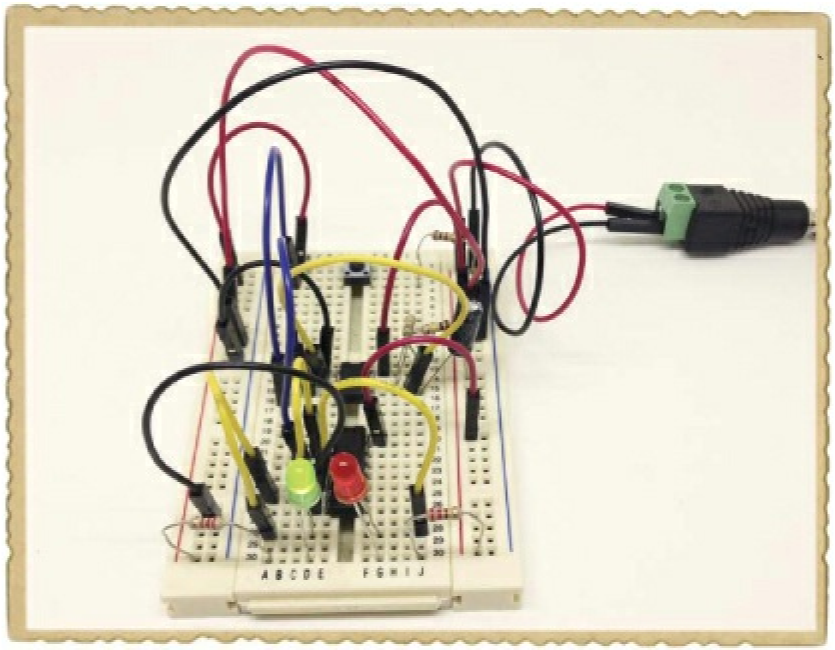
A B C D E
F G H I J

Shopping List

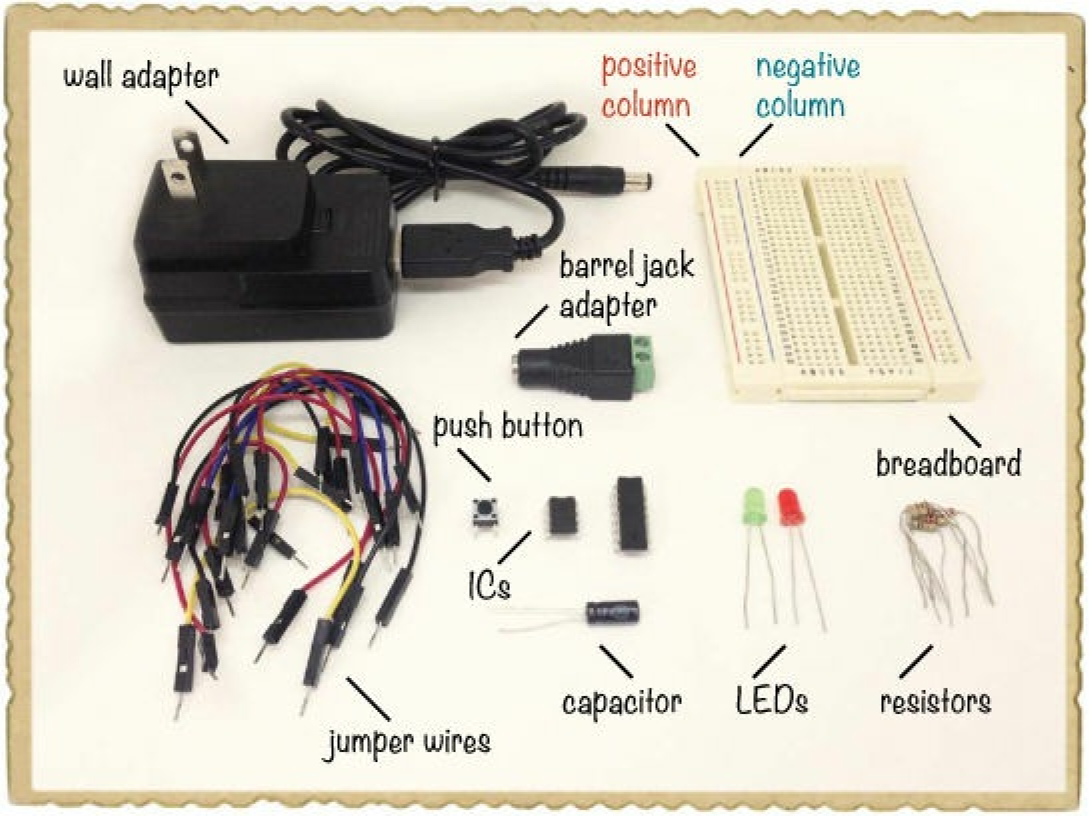
wall adapter
positive column
negative column
barrel jack adapter
push button
breadboard
ICs
capacitor
LEDs
resistors
jumper wires

- **A breadboard** (Jameco #20601, Bitsbox #CN329) with at least 30 rows.
- **Breadboard jumper wires** (Jameco #2237044, Bitsbox #CN236)—you'll need around 20 for this project.
- **A 555 timer IC** (Jameco #904085, Bitsbox #QU001) to create the input signal to be counted.
- **An IC with two D flip-flops** (Jameco #48004, Bitsbox #QU193)
- **A standard green LED** (Jameco #34761, Bitsbox #OP003)
- **A standard red LED** (Jameco #333973, Bitsbox #OP002)
- **Two 220 Ω resistors** (Jameco #690700, Bitsbox #CR25220R) for limiting the current to the LEDs.
- **A 100 kΩ resistor** (Jameco #691340, Bitsbox #CR25100K) to help set the frequency of the sound.
- **A 10 kΩ resistor** (Jameco #691104, Bitsbox #CR2510K) to help set the frequency of the sound.
- **A 1 kΩ resistor** (Jameco #690865, Bitsbox #CR251K) to use as a pull-down resistor for the start button.
- **A 1 µF capacitor** (Jameco #29831, Bitsbox #EC1U063) to help set the frequency of the sound.
- **A push button** (Jameco #119011, Bitsbox #SW087) to "toss the coin."
- **A 5 V DC wall adapter** (Jameco #2126125, Bitsbox #TF010) to power the circuit.
- **A DC barrel jack adapter** (Jameco #2227209, Bitsbox #CN424) to connect the wall adapter to the breadboard.

This circuit uses the positive and negative supply columns on both sides of the breadboard. When I say to connect a component to the negative or positive supply column “on the left,” that means you should use one of the supply columns on the left side of the breadboard. On both sides, the positive column is marked with a red line to the left, and the negative is marked with a blue line to the right.

Step 1: Build the Oscillator Circuit

First, let’s wire up the 555 timer:

1. Plug the 555 timer into the breadboard near the middle.
2. Connect R1, the 100 kΩ resistor, from pin 7 of the 555 timer to the positive supply column on the right.
3. Connect R2, the 10 kΩ resistor, from pin 6 to pin 7.
4. Connect C1, the 1 µF capacitor, from pin 6 to the negative supply column on the right. If you’re using a polarized capacitor like the one I suggest in the Shopping List, make sure you connect the negative leg to the negative supply column. The negative leg should be marked with a minus or a zero on the capacitor itself.
5. Connect a jumper wire from pin 2 to pin 6 of the 555 timer.

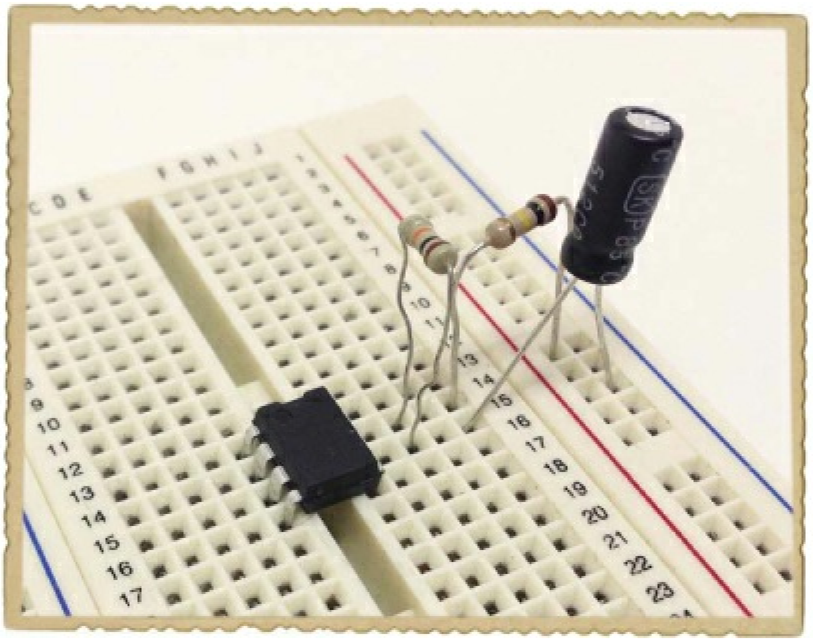

Step 2: Add the Start Button

Now, connect the push button between pin 4 on the 555 timer and the positive supply column as follows:

1. Place the push button at the very top of the breadboard, across the notch in the middle. By connecting it like this, you should have one side of the switch on the top row and the other side on row 3.
2. Connect a wire from pin 4 of the 555 timer to the lower pins of the push button (row 3). Connect a wire from the upper pins of the push button (row 1) to the positive supply column on the left.
3. Connect R5, the 1 kΩ pull-down resistor, from the lower pins of the push button to the negative

supply column on the right.

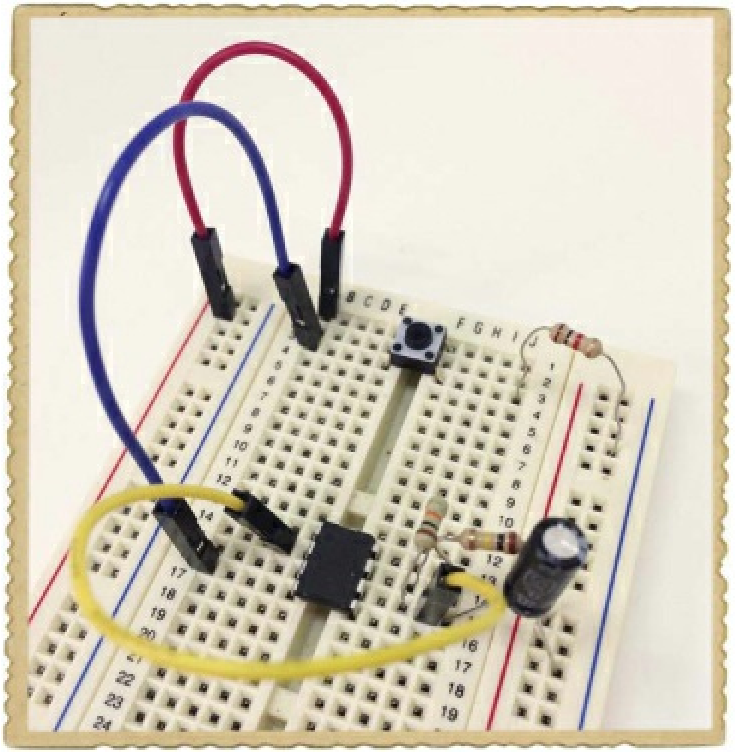

The 555 timer also needs to be powered. Connect a jumper wire from pin 1 to the negative supply column on the left. Use another jumper wire to connect pin 8 to the positive supply column on the right.

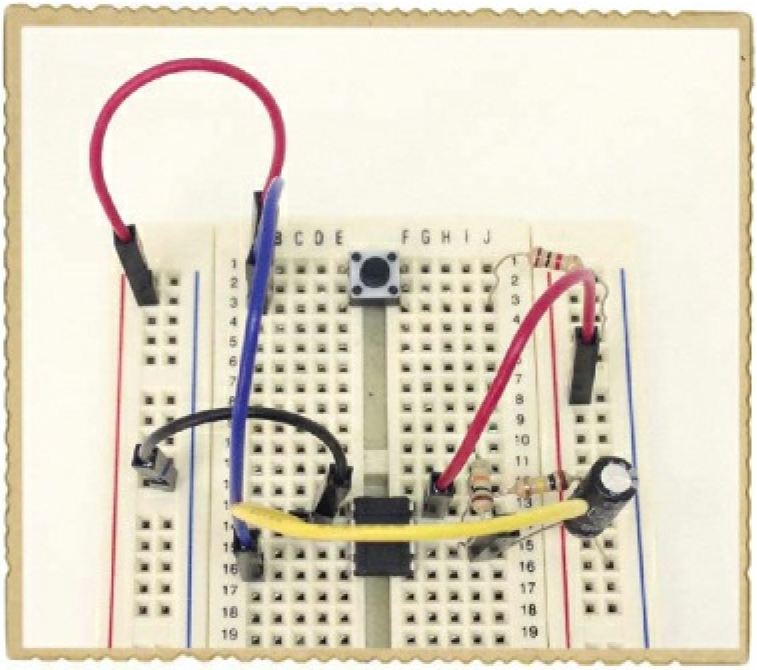

Step 3: Build the Toggle Circuit

Place the IC with the D flip-flops—marked *74LS74*—below the 555 timer so that it straddles the notch in the breadboard, with the notch pointing to the top of the breadboard. This IC contains two D flip-flops, but you'll use only the D flip-flop on pins 1 to 6.

Run a jumper wire from the inverted output $\overline{Q}$ on pin 6 of the 74LS74 D flip-flop to the D-input at pin 2. Connect the output from pin 3 on the 555 timer to pin 3 on the D flip-flop, which is the clock input.

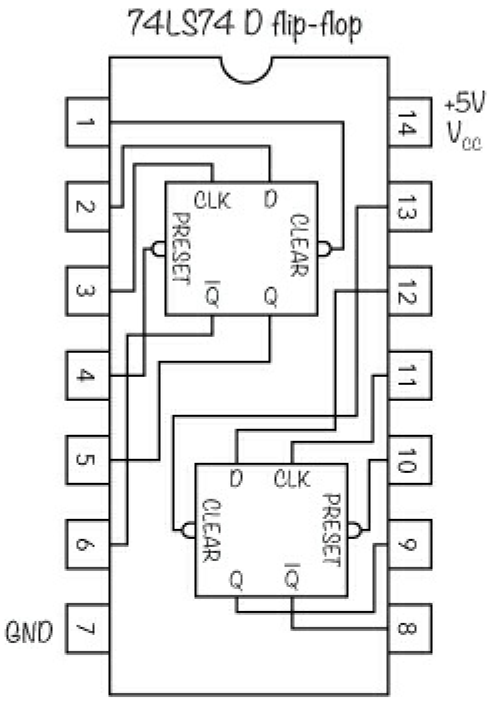
74LS74 D flip-flop
+5V
V_{cc}
GND
1
2
3
4
5
6
7
8
9
10
11
12
13
14
CLK
D
PRESET
CLEAR
Q̄
Q
D
CLK
CLEAR
PRESET
Q
Q̄

The D flip-flop needs power, too. Connect pin 14 to the positive supply column on the right and connect pin 7 to the negative supply column on the left.

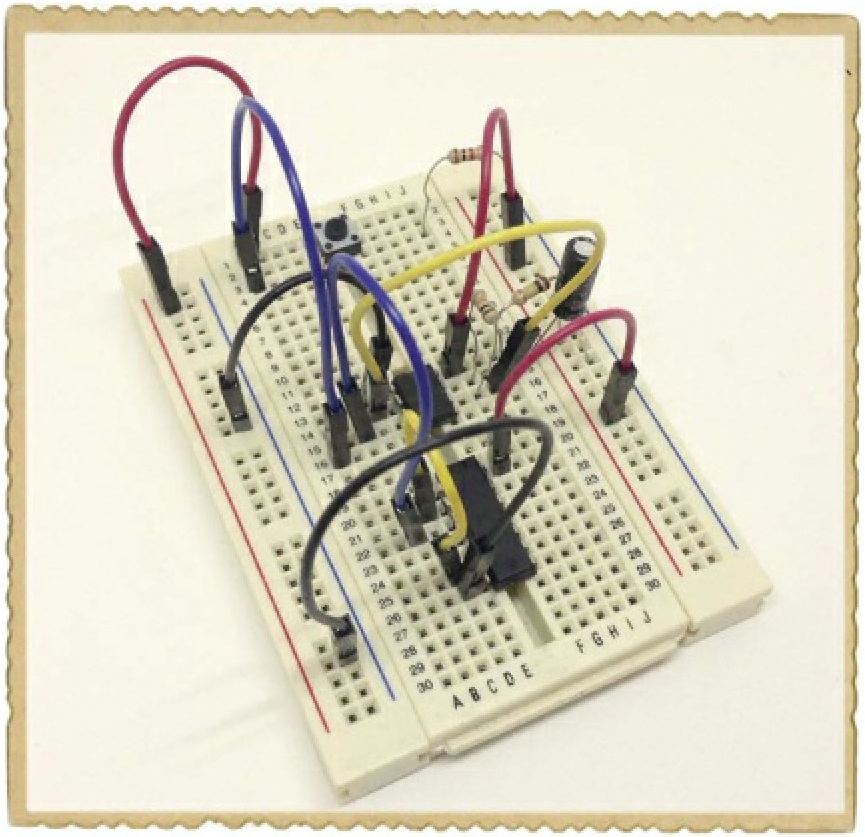

Step 4: Add the Heads and Tails LEDs

In the previous project, you used a transistor to power an LED from a logic gate output because the gate couldn't provide enough current. This circuit has the same challenge because D flip-flops are just a few logic gates in an IC, but there's a little trick you can use to get around that limitation.

The 74LS74 D flip-flop's datasheet says this IC will allow only about 0.5 mA to flow when the output voltage is high, but it allows 8 mA when the output voltage is low. (Search online for *74LS74 datasheet* if you're curious to read the datasheet for yourself.) If you connect the LEDs and resistors to the positive supply column on one side and to the flip-flop output on the other, the LEDs should get 8 mA of current when the output is low, turning them on. It may seem strange to connect the LEDs like this, instead of connecting them to the negative supply column, but doing so makes the LEDs light when the output from the gate is 0 instead of 1.

Whatever value Q has, $\bar{Q}$ will always be the opposite. If you connect an LED to each output, one will light up and the other won't. Add the heads and tails LEDs as follows:

1. Plug the two LEDs into the bottom of the breadboard, with the red LED on the right side of the middle notch and the green LED on the left side. Place the longer leads (the anodes) in the bottom row and the shorter leads (the cathodes) a couple of rows above.
2. Connect one jumper wire from pin 5 of the D flip-flop to the short leg of the red LED. Then connect another jumper wire from pin 6 of the D flip-flop to the short leg of the green LED.
3. Connect a resistor from each bottom row to the positive supply column on each side (R3 and R4 from the circuit diagram).

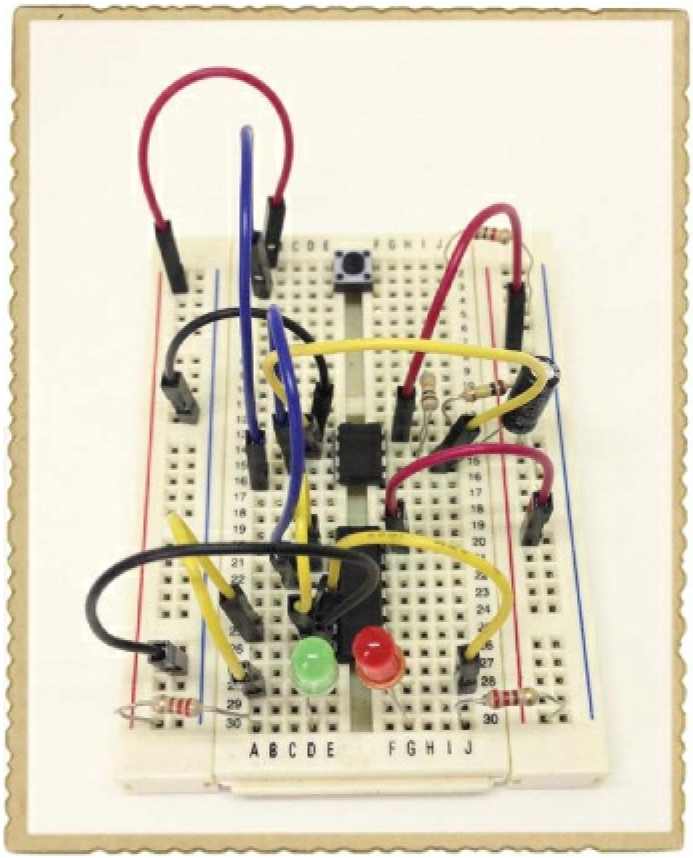

Step 5: Toss that "Coin"!

Use a jumper wire to connect the left negative supply column to the one on the right, and do the same for the two positive supply columns. Then, connect two jumper wires to your barrel jack adapter—connect the plus side to one of the positive supply columns and connect the minus side to one of the negative supply columns.

Finally, plug the wall adapter into the barrel jack adapter first and then into the wall socket. One LED should light up right away. When you push the button, the LEDs should alternate quickly between on and off. Release the button, and only one should be lit.

Now, you can use this circuit to make decisions. For example, say you ask yourself, "Should I go out and play football this weekend, or should I play baseball?" Green means go out and play football; red means baseball. Or, if you're arguing with one of your friends about who gets the last cookie, then let the coin tosser decide!

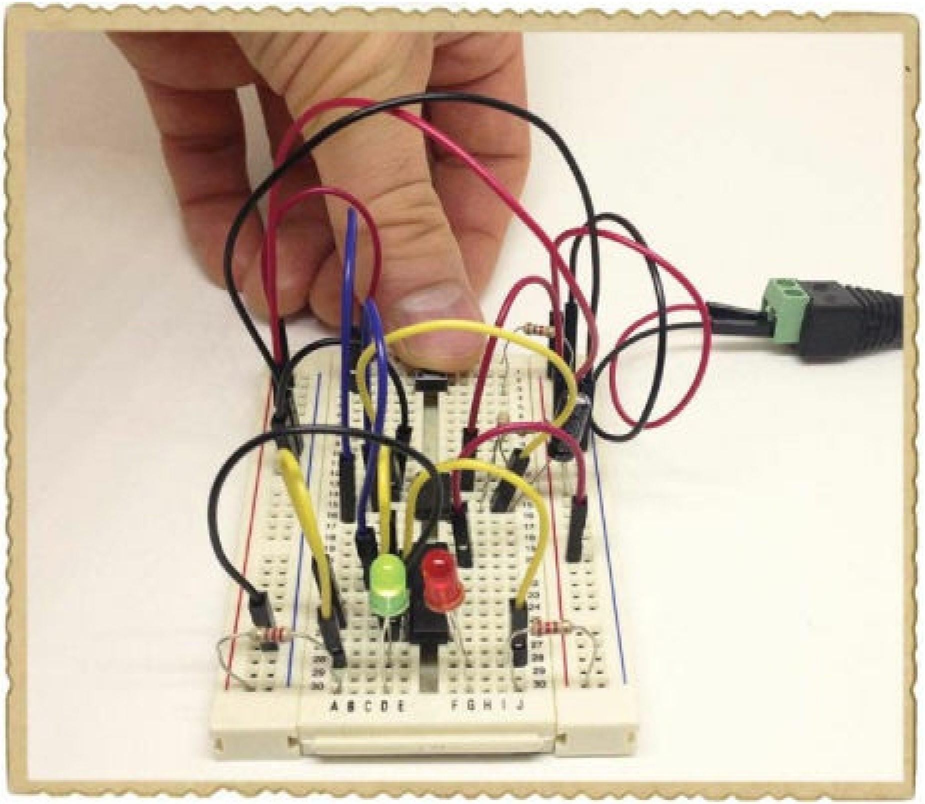

Step 6: What If the Coin Tosser Doesn't Work?

First, check that you're using a 5 V wall adapter. Any other voltages might not work.

Then, if one LED is lit but nothing happens when you push the button, check that the D flip-flop is connected correctly. If no LEDs are lit, there's definitely something wrong with the D flip-flop and LED parts of the circuit. Carefully compare your connections with the circuit diagram at the beginning of this project. If your circuit isn't working after you know the LED and flip-flop circuits are connected correctly, then check the 555 timer connections.

To help you avoid the same mistakes, I'll share the problems I had when building this for the first time:

- I connected the LEDs to pins 4 and 5 of the D flip-flop instead of pins 5 and 6.
- I connected the capacitor to pin 5 of the 555 timer instead of pin 6.
- I forgot to connect the positive supply column on the left to the one on the right.

WHAT'S NEXT?

You've built a lot of different circuits in this book! At this point you have a solid foundation in electronics, both in theory and practical experience. Now, the next step is to focus on what you think is fun. Find a project you really want to build—and go make it!

The best way to continue learning is to build lots of circuits and read about anything you're curious about. Follow tutorials online and find more books on different topics in electronics.

In Chapter 12, the final chapter, I'll show you how to build one last project: a really cool game where you have to test your reaction speed by "catching" a light. After that, I hope you'll continue exploring, playing, and having fun with electronics. There are so many great things you can build!

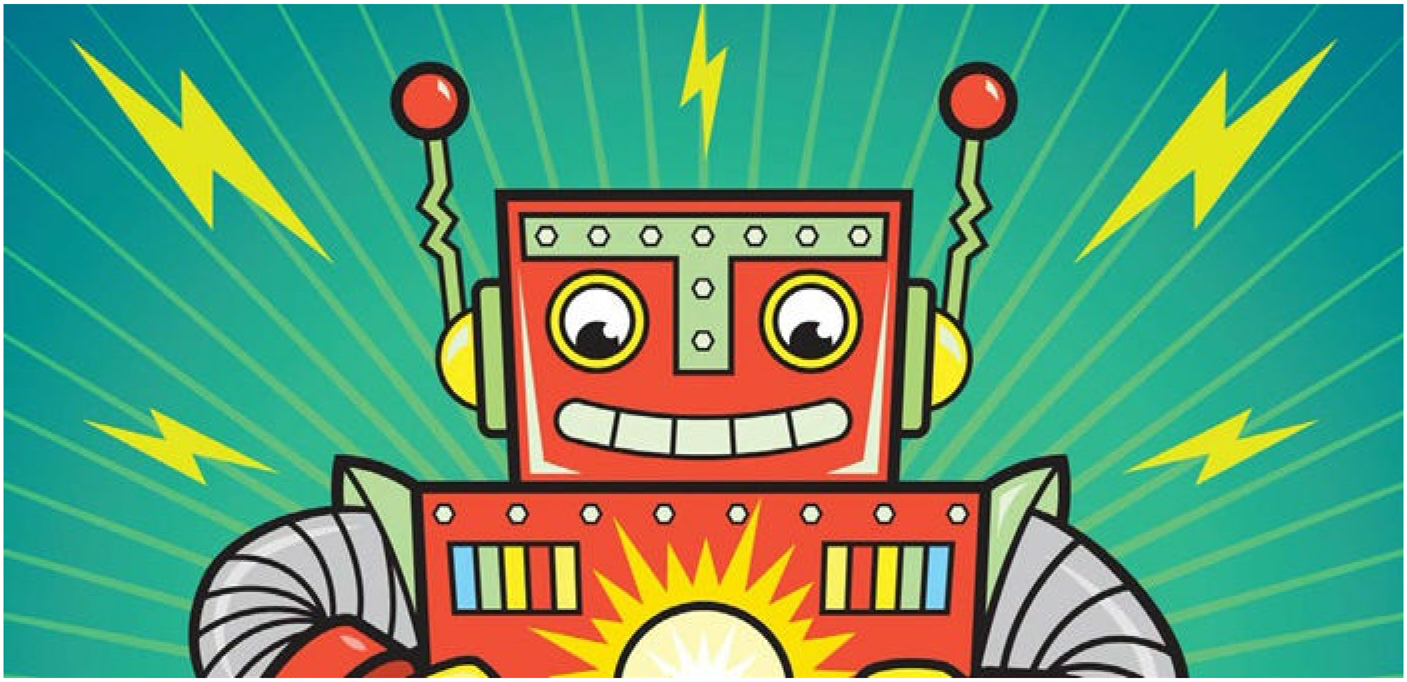

12

LET’S MAKE A GAME!

You’ve built all sorts of small circuits in this book, and each circuit was designed to teach you a particular concept. In this chapter, you’ll combine all your new skills to build a reaction game. The game has a row of five LEDs that light up one at a time so that a light appears to run back and forth.

The goal of the game is to stop the light when it’s in the middle of the five LEDs. That gives you 10 points. If you stop it on an LED next to the middle one, you get 5 points. But if you stop it on one of the end LEDs, you lose all your points and have to start over from 0. Try to reach 50 points!

You can play this game by yourself to practice your reaction time, or with as many friends as you want. If you’re competing with friends, I suggest giving each player only one attempt at stopping the light before the next player gets a turn.

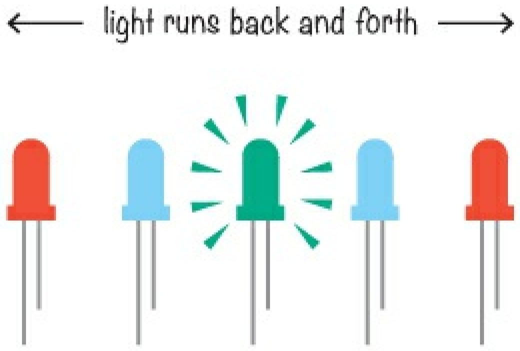
← light runs back and forth →

If you stop on the green light, you get 10 points!

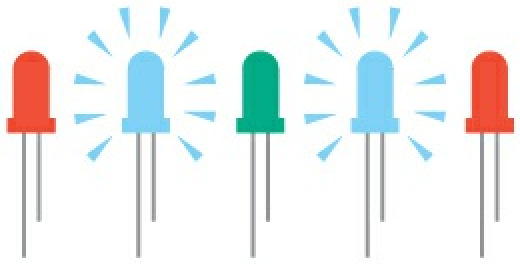

If you stop on a blue light, you get 5 points!

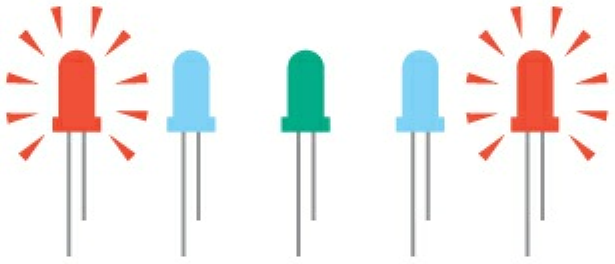

If you stop on a red light, you lose all your points!

MEET THE REACTION GAME CIRCUITS

The reaction game will consist of three circuits:

- A 555 timer circuit that determines the speed of the game
- A counter that controls which LED light to turn on
- An SR latch that will add a reset button and an action button

This section explains each circuit, but to help you understand their diagrams, let's meet two new circuit symbols.

Meet the V_{CC} and GND Symbols

Circuit diagrams don't always use a battery symbol like the one used throughout this book. Sometimes they use the V_{CC} (or V_{DD}) and *GND* symbols instead.

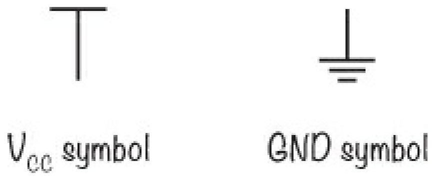

If nothing in the circuit diagram or its description says otherwise, you can assume that V_{CC} represents the positive side of the battery and that GND represents the negative side, or ground. The symbols sometimes look a little different, but the V_{CC} symbol usually shows a wire connecting down from its symbol to the circuit, while the GND symbol shows a wire connecting up from the symbol to the circuit.

In bigger circuit diagrams, like the one you're going to build from in this chapter, these symbols make the diagram much easier to draw and understand.

WHY IS IT CALLED V_{CC}?

The positive voltage symbol is called V_{CC} because of old naming conventions. V_{CC} was the voltage supplied to the collector side of a transistor in common transistor circuits, usually through a resistor or some other components. The collector is where the "CC" comes from.

You've used a bipolar junction transistor throughout this book, but there's another type of transistor called a *field-effect transistor (FET)*. The pin that equals the collector on this type of transistor is called the *drain*, so the voltage that was supplied to the drain side of the FET was called V_{DD}.

A 555 Timer to Set the Light Speed

The circuit that sets the reaction game's speed will be built around a 555 timer, and it's similar to the circuits you built in Chapter 8. The components in this circuit diagram will set the game to a "medium" speed: it's not super fast, and it's not super slow.

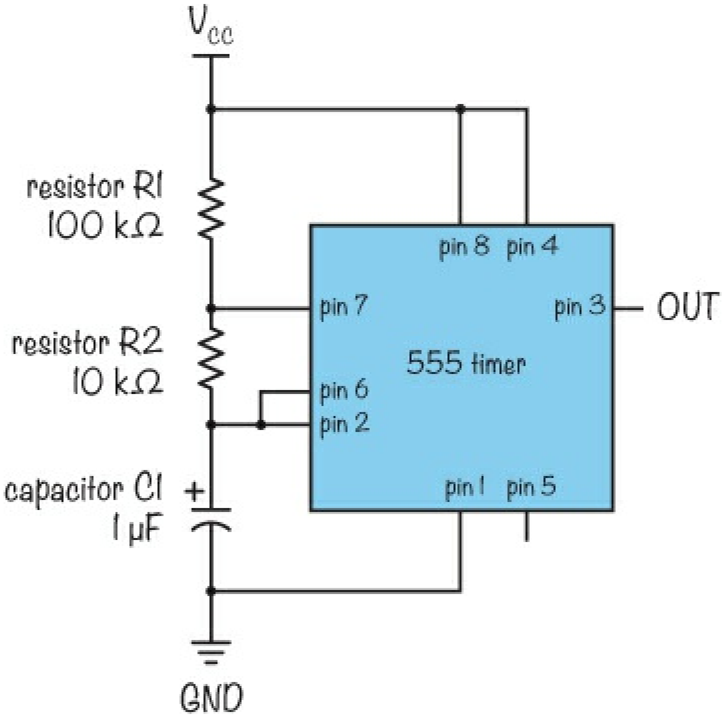
V_cc
resistor R1
100 kΩ
resistor R2
10 kΩ
capacitor C1
1 μF
+
GND
pin 8
pin 4
pin 7
pin 3
OUT
555 timer
pin 6
pin 2
pin 1
pin 5

Every time the output from the 555 timer goes from low to high, the light moves one step to the side. The number of times the output from the 555 timer goes high per second is the frequency of the output. As I showed in Chapter 8, the formula for calculating the frequency of the output of the 555 timer is

$$\text{Frequency} = \frac{1.44}{(\text{R1} + \text{R2} + \text{R2}) \times \text{C1}}$$

The following values from the 555 timer circuit diagram correspond to that formula:

R1 = 100 kΩ
R2 = 10 kΩ
C1 = 1 μF

Plug these into the formula, keeping in mind that 1 μF = 0.000001 F and 120 kΩ = 120,000 Ω, and you get this:

$$\text{Frequency} = \frac{1.44}{(100\ \text{k}\Omega + 10\ \text{k}\Omega + 10\ \text{k}\Omega) \times 1\ \mu\text{F}}$$

$$\text{Frequency} = \frac{1.44}{120\ \text{k}\Omega \times 1\ \mu\text{F}}$$

$$\text{Frequency} = \frac{1.44}{120{,}000\ \Omega \times 0.000001\ \text{F}}$$

$$\text{Frequency} = 12\ \text{Hz}$$

This means the output will go high 12 times per second and the light will change places 12 times per second. You can experiment with the component values for R1, R2, and C1 later to speed up or slow down the game.

A Counter to Turn the LEDs On

To control the LEDs, you'll use a *decade counter*, which is an IC that counts input pulses. Every time the clock input on pin 14 goes from low to high, the counter increments by one. It counts from 0 to 9, and it has 10 outputs marked 0 to 9.

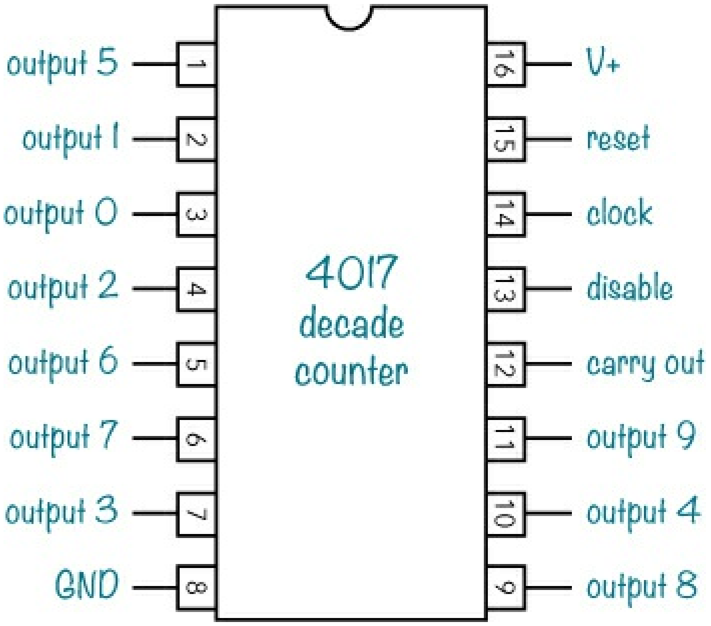
output 5
output 1
output 0
output 2
output 6
output 7
output 3
GND
1
2
3
4
5
6
7
8
4017
decade
counter
16
15
14
13
12
11
10
9
V+
reset
clock
disable
carry out
output 9
output 4
output 8

For example, when the counter has counted three input pulses, output 3 (that is, pin 7) is high, and the other pins are low. If you connect an LED to output 3, then when the counter is at three, the LED will turn on.

If you connect LEDs to several output pins, then as the counter increases, the LEDs turn on in order, according to their output pins. When the counter is at 9 and receives a 10th input pulse, it goes back to 0 and turns the output pins on in order again.

But the counter counts pulses only if pin 13 is low. This means you can use pin 13 to tell the game when to start moving the light across the LEDs and when to stop the light.

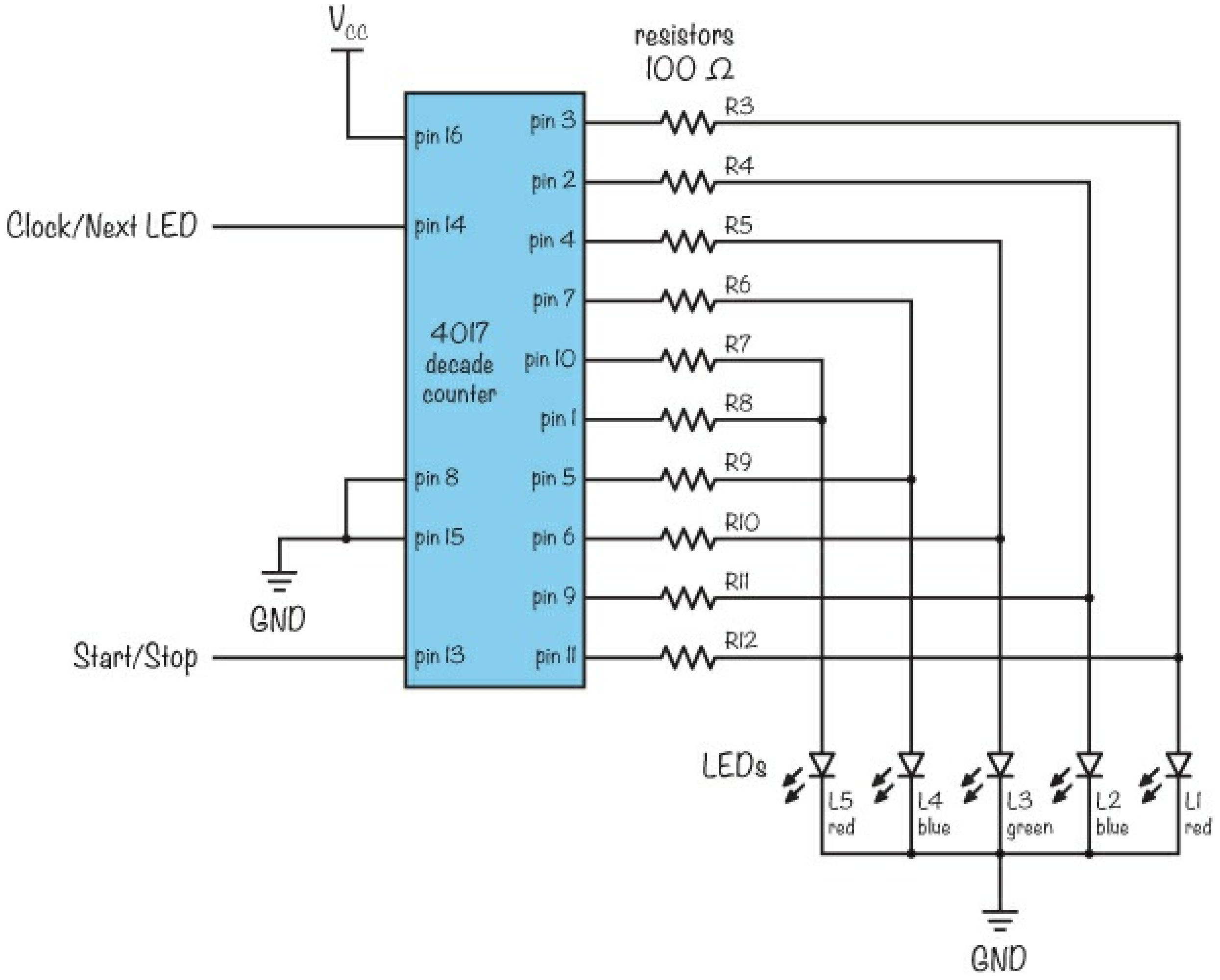
V_{cc}
resistors
100 Ω
R3
R4
R5
R6
R7
R8
R9
R10
R11
R12
pin 16
pin 3
pin 2
Clock/Next LED
pin 14
pin 4
pin 7
4017
decade
counter
pin 10
pin 1
pin 8
pin 5
pin 15
pin 6
GND
pin 9
Start/Stop
pin 13
pin 11
LEDs
L5
red
L4
blue
L3
green
L2
blue
L1
red
GND

Each output has a resistor to reduce the current through the LED and make sure the LED doesn't get destroyed. Because two output pins connect to each LED, the resistors keep the voltage to each LED high, even though one output will be low and one will be high. The resistors also ensure that two outputs aren't connected directly together, which could damage the IC when one output is high and the other low.

A Latch to Start and Stop the Light

Do you remember the SR latch from "Saving One Bit at a Time" on page 240? The start/stop circuit for this game is a similar SR latch but built with two NAND gates. (The SR latch in Chapter 11 used NOR gates.)

The SR latch is a circuit that can remember a single bit. Its output is either 0 or 1, and it keeps that number until it gets set or reset with a new input.

You can create a circuit that tells the latch what to output with two buttons: one for setting the output to 1 and one for setting the output to 0. Using NAND gates instead of NOR gates means the buttons must make the inputs low to output a 1.

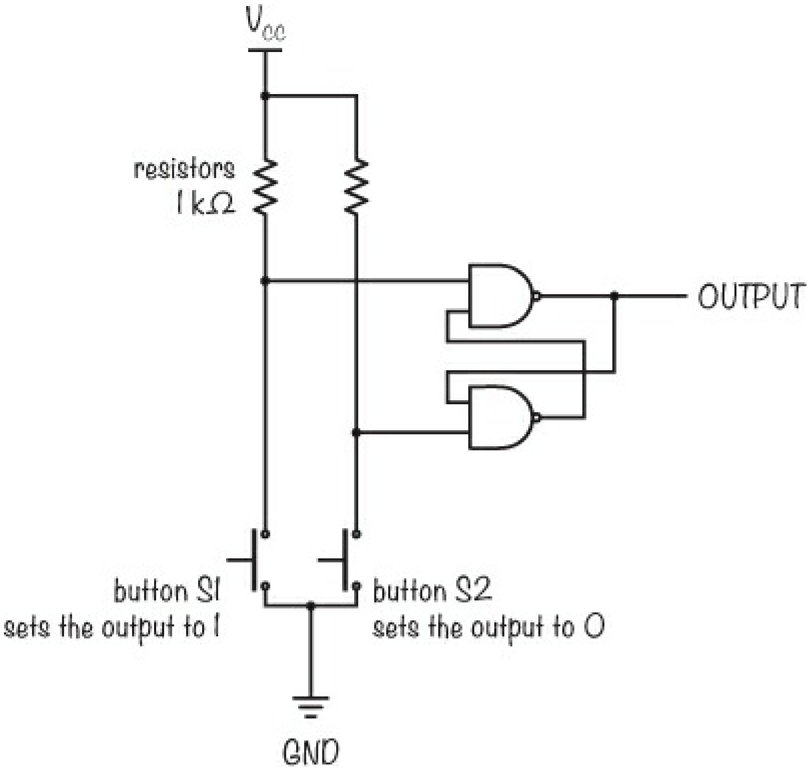
V_{cc}
resistors
1 kΩ
OUTPUT
button S1
sets the output to 1
button S2
sets the output to 0
GND

In this circuit, it doesn’t matter whether you click the buttons quickly or slowly. The 1-button always sets the output to 1, and the 0-button always sets the output to 0.

That’s perfect for the reaction game! Connecting the output to the start/stop pin, or pin 13, on the decade counter gives you a button for starting and stopping the LEDs.

PROJECT #23: AN LED REACTION GAME

It’s time to put all the pieces I showed you together to build the reaction game. This circuit has a lot of connections, but I know you can make it. Just don’t rush. Take your time and test each part of the circuit after the step where I explain how to build it.

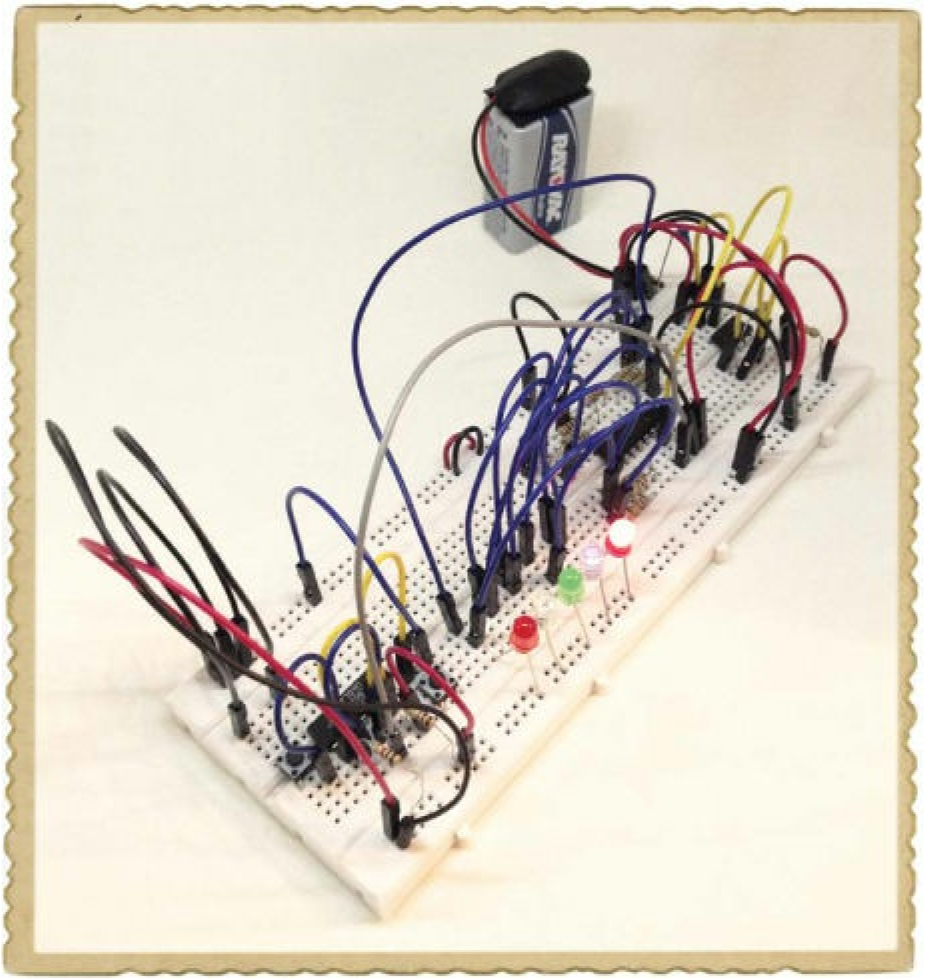

I also recommend using a bigger breadboard than you've used in the previous projects, because this circuit is huge!

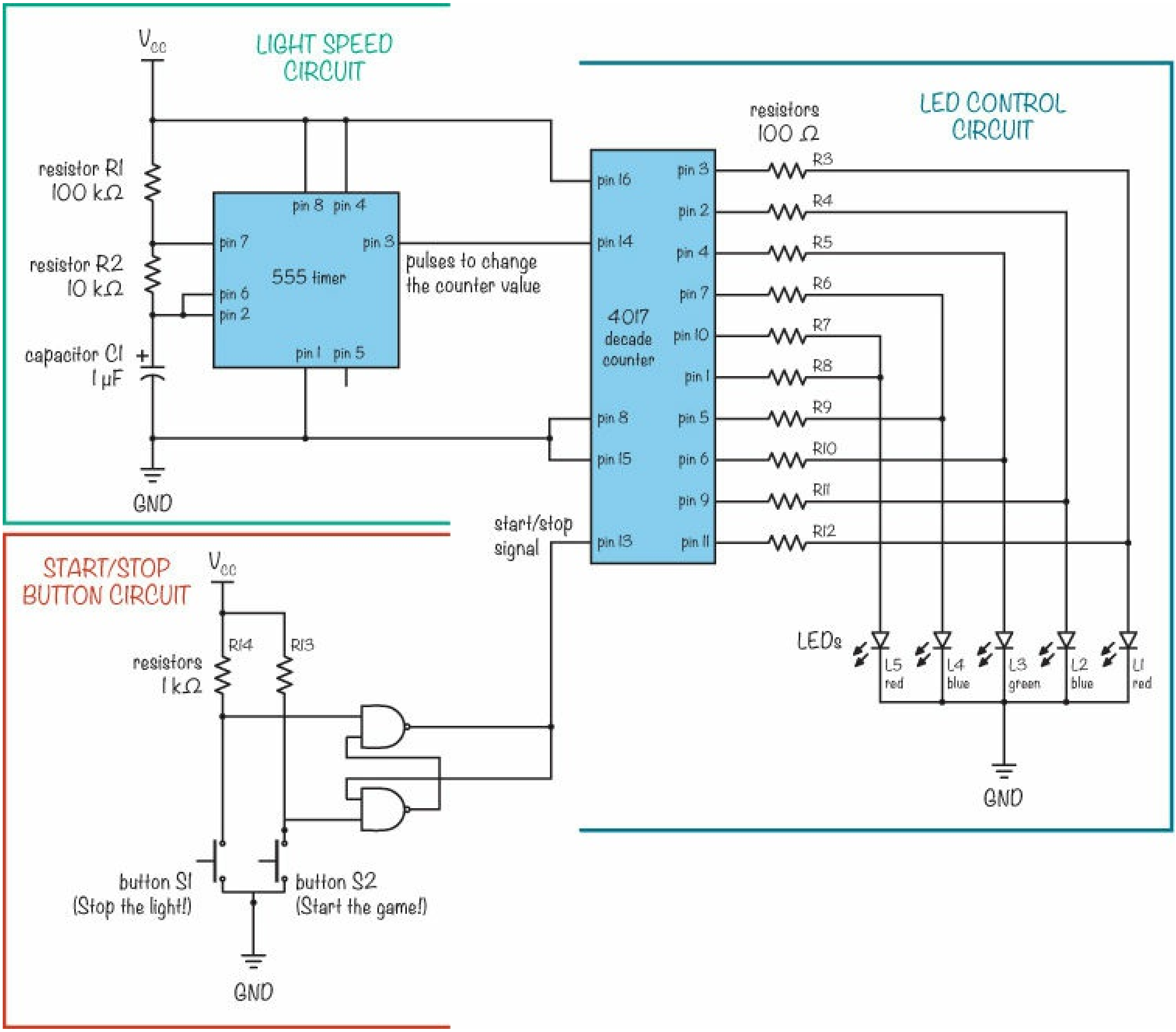
LIGHT SPEED CIRCUIT
Vcc
resistor R1 100 kΩ
resistor R2 10 kΩ
capacitor C1 1 µF
+
GND
pin 8
pin 4
pin 7
pin 3
555 timer
pin 6
pin 2
pin 1
pin 5
pulses to change the counter value
LED CONTROL CIRCUIT
resistors 100 Ω
4017 decade counter
pin 16
pin 14
pin 8
pin 15
pin 13
pin 3
pin 2
pin 4
pin 7
pin 10
pin 1
pin 5
pin 6
pin 9
pin 11
R3
R4
R5
R6
R7
R8
R9
R10
R11
R12
LEDs
L5 red
L4 blue
L3 green
L2 blue
L1 red
GND
start/stop signal
START/STOP BUTTON CIRCUIT
Vcc
resistors 1 kΩ
R14
R13
button S1 (Stop the light!)
button S2 (Start the game!)
GND

Shopping List

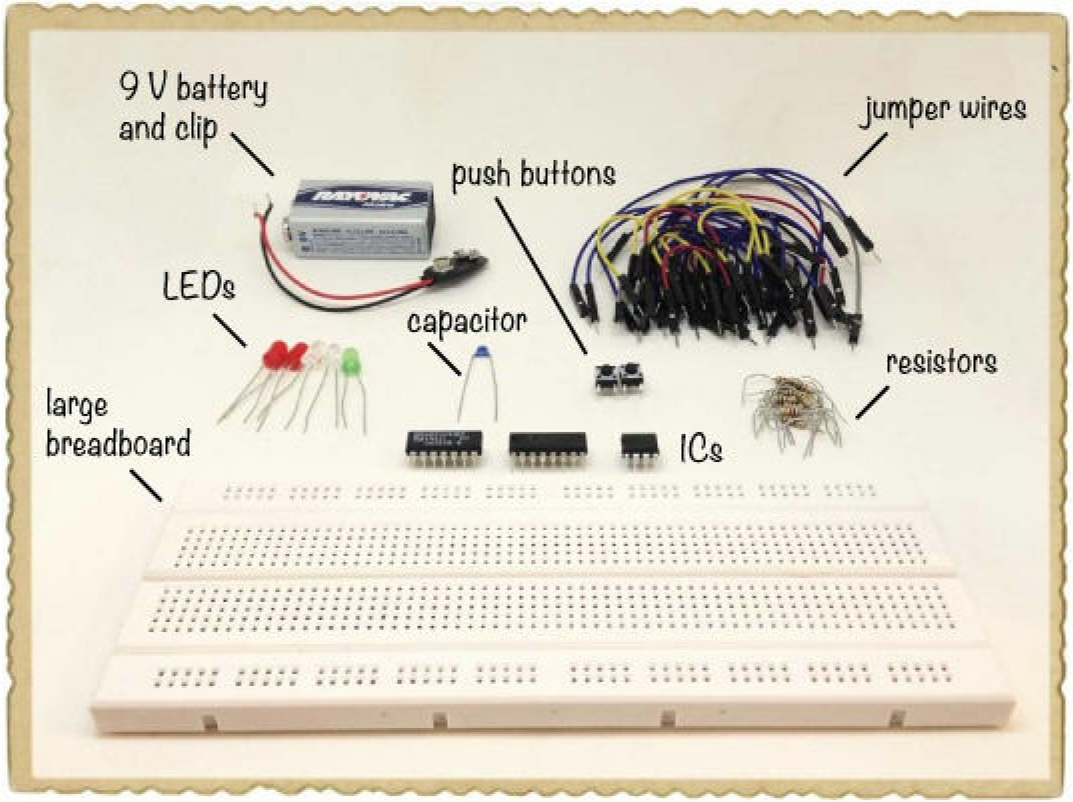

- **A breadboard** (Jameco #2212218, Bitsbox #CN204) with at least 60 rows.
- **Breadboard jumper wires** (Jameco #2237044, Bitsbox #CN236)—you'll need around 35 for this project. Standard hookup wire works, too.
- **A standard 9 V battery** to power the circuit.
- **A 9 V battery clip** (Jameco #11280, Bitsbox #BAT033) to connect the battery.
- **A 555 timer IC** (Jameco #904085, Bitsbox #QU001) to create the timing.
- **A 10 kΩ resistor** (Jameco #691104, Bitsbox #CR2510K) to set the game speed.
- **A 100 kΩ resistor** (Jameco #691340, Bitsbox #CR25100K) to set the game speed.
- **A 1 µF capacitor** (Jameco #768183, Bitsbox #CC006) to set the game speed.
- **A 4017 decade counter IC** (Jameco #12749, Bitsbox #QU020) to control the LEDs.
- **Two standard blue LEDs** (Jameco #2193889, Bitsbox #OP033)
- **Two standard red LEDs** (Jameco #333973, Bitsbox #OP002)
- **A standard green LED** (Jameco #34761, Bitsbox #OP003)
- **Ten 100 Ω resistors** (Jameco #690620, Bitsbox #CR25100R) for limiting the current to the LEDs.
- **A 4011 NAND-gate IC** (Jameco #12634, Bitsbox #QU018) to create the SR latch for starting and stopping the game.
- **Two 1 kΩ resistors** (Jameco #690865, Bitsbox #CR251K) to act as pull-up resistors for the start/stop circuit.
- **Two push buttons** (Jameco #119011, Bitsbox #SW087), one for resetting the game and one for

playing.

Tools

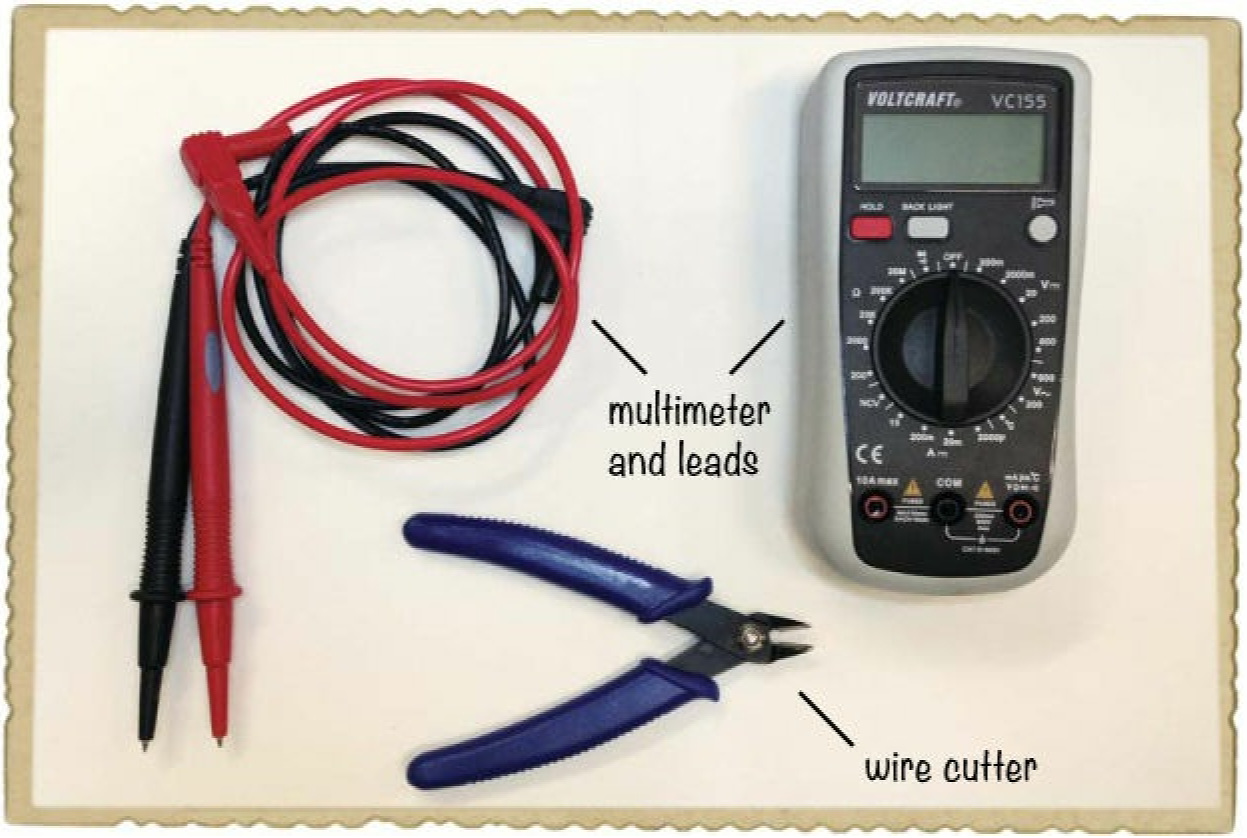

- **A wire cutter** (Jameco #35482, Bitsbox #TL008) to cut small pieces of wire.
- **A multimeter** (Jameco #2206061, Bitsbox #TL057, Rapid Electronics #55-6662) to debug your circuit if it's not working correctly.

Step 1: Build the 555 Timer Circuit

Plug the 555 timer into the breadboard all the way at the top so that you'll have room for the other parts of the circuits farther down. Then, connect the capacitors and resistors to the IC according to this project's circuit diagram. The capacitor I suggest in this project's Shopping List is a nonpolarized capacitor, so it doesn't matter which way you connect it. If you use a polarized capacitor instead, connect it according to the plus marking in the circuit diagram.

Use wires to make connections as needed, as I show in this breadboard diagram.

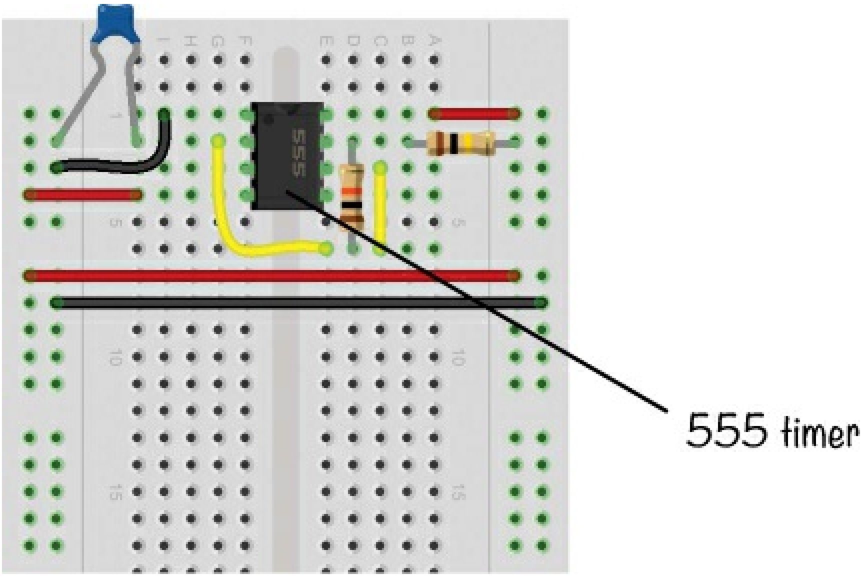

In this project, it’s best to use the supply column pairs on both sides to make connections easier and keep everything as tidy as possible. The breadboard that I recommend in this project’s Shopping List doesn’t have blue and red markings, but the positive and negative columns are the same as in breadboards with the stripes. The left and right sides of the breadboard each have a pair of supply columns. The positive supply column is the left column in each pair, and the negative supply column is the right column in each pair. Use a red wire to connect the positive column on one side to the positive column on the other side, and do the same using a black wire with the negative columns.

As you follow my instructions, connect everything in the 555 timer circuit that should connect to V_{CC} to one of the positive supply columns, and connect everything that should connect to GND to one of the negative supply columns.

NOTE

This circuit connects the 555 timer in astable mode, just like the 555 timer circuits in Chapter 8. Read “Meet the 555 Timer” on page 164 for a description of exactly how this IC works. You can also build the projects in Chapter 8 to practice using the 555 timer.

Before you move on to the next step, check that this circuit is working by connecting an LED with a resistor to the output of the 555 timer as follows:

1. Connect the negative side (short leg) of an LED to the output on pin 3 of the 555 timer.
2. Connect the positive side (long leg) of the LED to a 100 Ω resistor, and connect the other side of this resistor to the positive supply column.
3. Connect your battery clip to one of the supply column pairs as usual. Then plug in the battery to check that the circuit works.

If your LED blinks really fast, then you’re ready to move on. If not, recheck your connections to find out where the error is.

When you know the 555 timer circuit works, unplug the LED, 100 Ω resistor, and battery clip.

Step 2: Build the LED-Controlling Circuit

Now, you're going to connect the 4017 decade counter with resistors and LEDs. There are a lot of connections, so take as much time as you need to get them all correct.

Plug the 4017 decade counter into the breadboard so that the middle of the decade counter is around row 20, with the chip marker pointing up toward row 1. Then, take out five LEDs and ten 100 Ω resistors.

Connect each LED's negative (short) leg to the negative supply column on the right, and connect each positive (long) leg to its own empty row in the component area on the right. Place the green LED in the middle, the two blue ones on each side of the green LED, and the red ones on each end.

Then, connect the ten 100 Ω resistors. In the circuit diagram, notice that pins 1 to 7 and pins 9 to 11 of the 4017 decade counter each connect to one side of a resistor. The other side of each resistor needs to be on a row by itself. Take care to ensure the resistor legs don't accidentally touch one another. Look at the following breadboard circuit to see how I connected them:

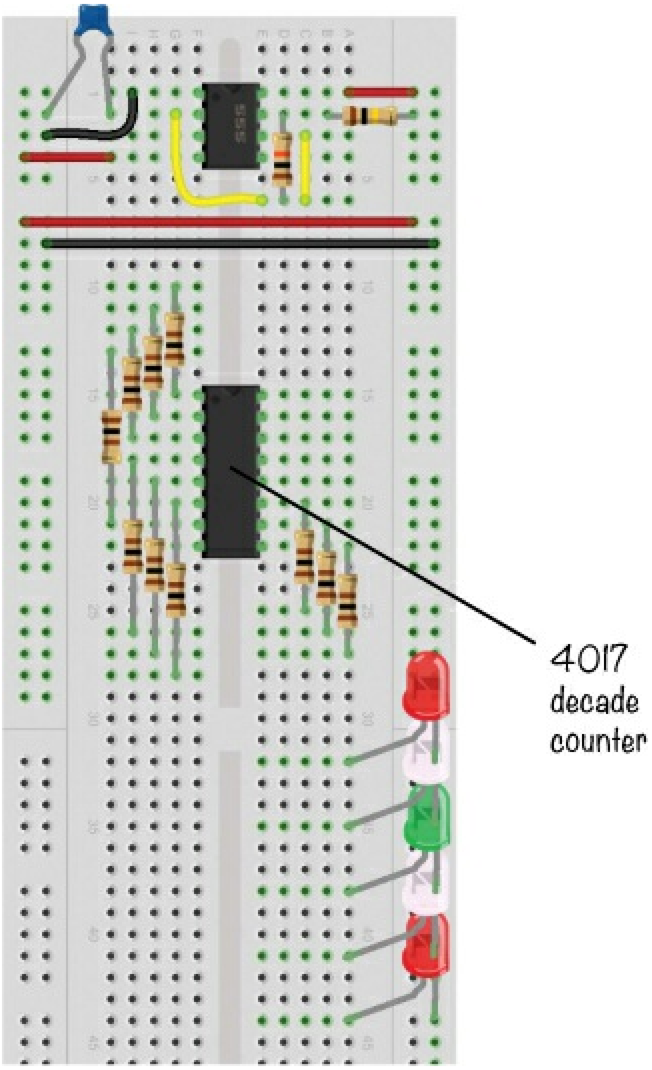

Now, connect the LEDs to the resistors on the 4017 decade counter, and connect the decade counter circuit to the 555 timer circuit according to the circuit diagram. Jumper wires are the best way to make those connections.

From each resistor, connect a jumper wire to the corresponding LED. Look at the circuit diagram and notice, for example, that the other side of the resistor connected to pin 4 of the 4017 decade counter should connect to the positive pin of the green LED in the middle. Go through the pins in the circuit diagram to figure out which LED to connect each resistor to.

Connect pins 8 and 15 of the 4017 decade counter to the negative supply column, and connect pin 16 to the positive supply column. Use a wire to connect the output from the 555 timer (pin 3) to the clock input of the 4017 decade counter (pin 14).

Make sure that you have positive and negative connections in all of your power supply columns. The breadboard I recommend in this project's Shopping List (page 267) divides its power supply columns into two sections, one upper and one lower. Just connect each of the upper and lower halves on the left side with a wire to bridge the gap, as shown. Do the same on the right side. Alternatively, use two jumper wires from the left columns to the right columns.

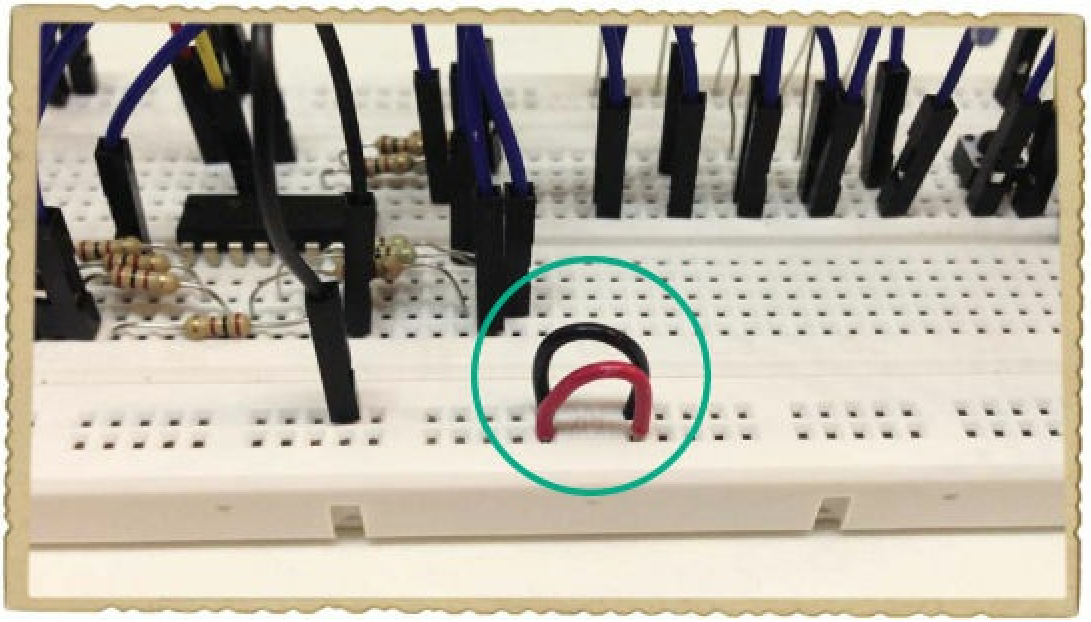

You can use a jumper wire, or you can cut off a small piece of wire, as I've done in this photo. Then, use two long jumper wires to connect the lower-left power supply columns with the two lower-right columns. When you're done connecting the two circuits and all the power supply columns, your breadboard should look like this:

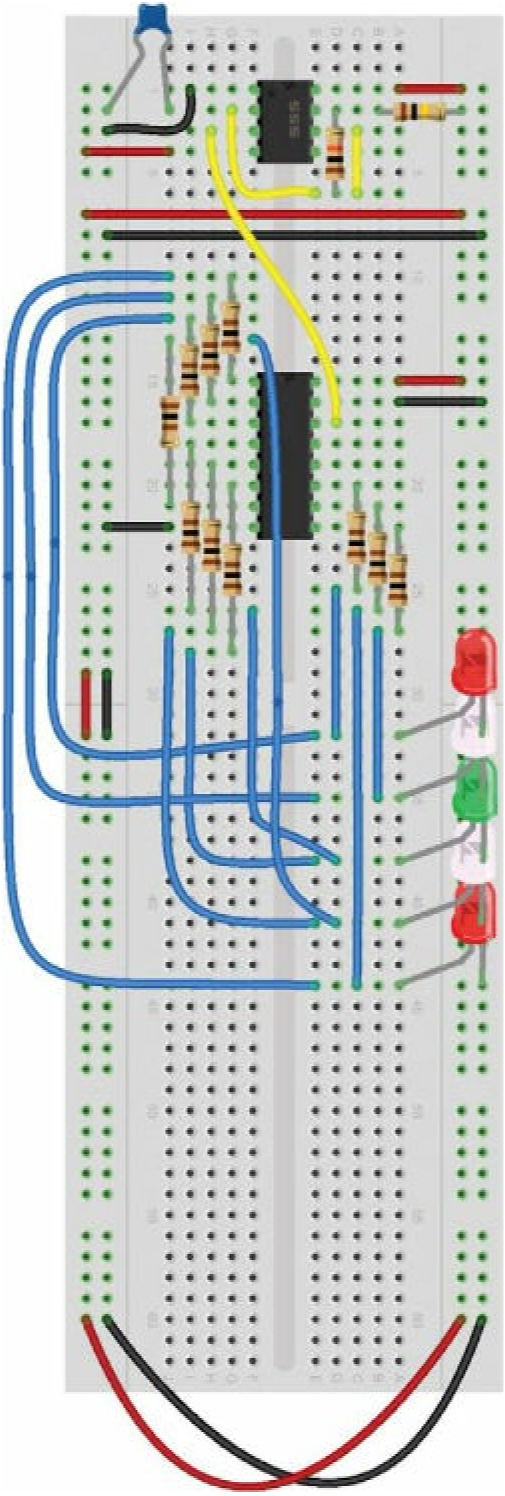

Before building the next part of the circuit, check that your LED-controlling circuit is working correctly, too. To test it, just connect pin 13 on the 4017 decade counter—that is, the "disable" pin—to the negative supply column with a jumper wire, and plug your battery clip and battery into the breadboard as usual. You should see a light "running" back and forth across the row of LEDs.

If no LEDs light up, first check that you've connected the 4017 decade counter with the notch pointing upward. Connecting the chip the wrong way is an easy mistake to make. I've done it many times!

Next, check that pin 16 of the 4017 decade counter is connected to the positive supply column and that pins 8 and 15 connect to the negative supply column. Also, confirm that you've connected the LEDs with their short legs in the negative supply column.

If some LEDs work and some don't, or if the light doesn't run smoothly back and forth, look over all the connections of resistors and jumper wires to find the fault.

After verifying that your circuit works, remove the wire connecting pin 13 of the 4017 decade counter to the negative supply column, and disconnect the battery from the breadboard.

Step 3: Build the Start and Stop Circuit

The last piece of this project is the button circuit that starts and stops the LEDs. Make these connections now:

1. Connect one push button at the bottom of the breadboard, across the notch in the middle. Plug the 4011 NAND-gate IC into the breadboard, a couple of rows above the button. Make sure its chip marking points toward row 1 on the breadboard.
2. Place the second button above the IC on the right component side so that it’s easy to reach it with your finger.
3. Connect the two resistors, R13 and R14, as shown in the circuit diagram. Then, use jumper wires to make the remaining connections in the SR latch circuit, as shown in the following breadboard diagram. Connect the positive and negative supply columns to the NAND-gate IC (pins 14 and 7, respectively), and connect the wire from pin 11 of the NAND-gate IC to pin 13 of the 4017 decade counter.

Compare your connections to the following image.

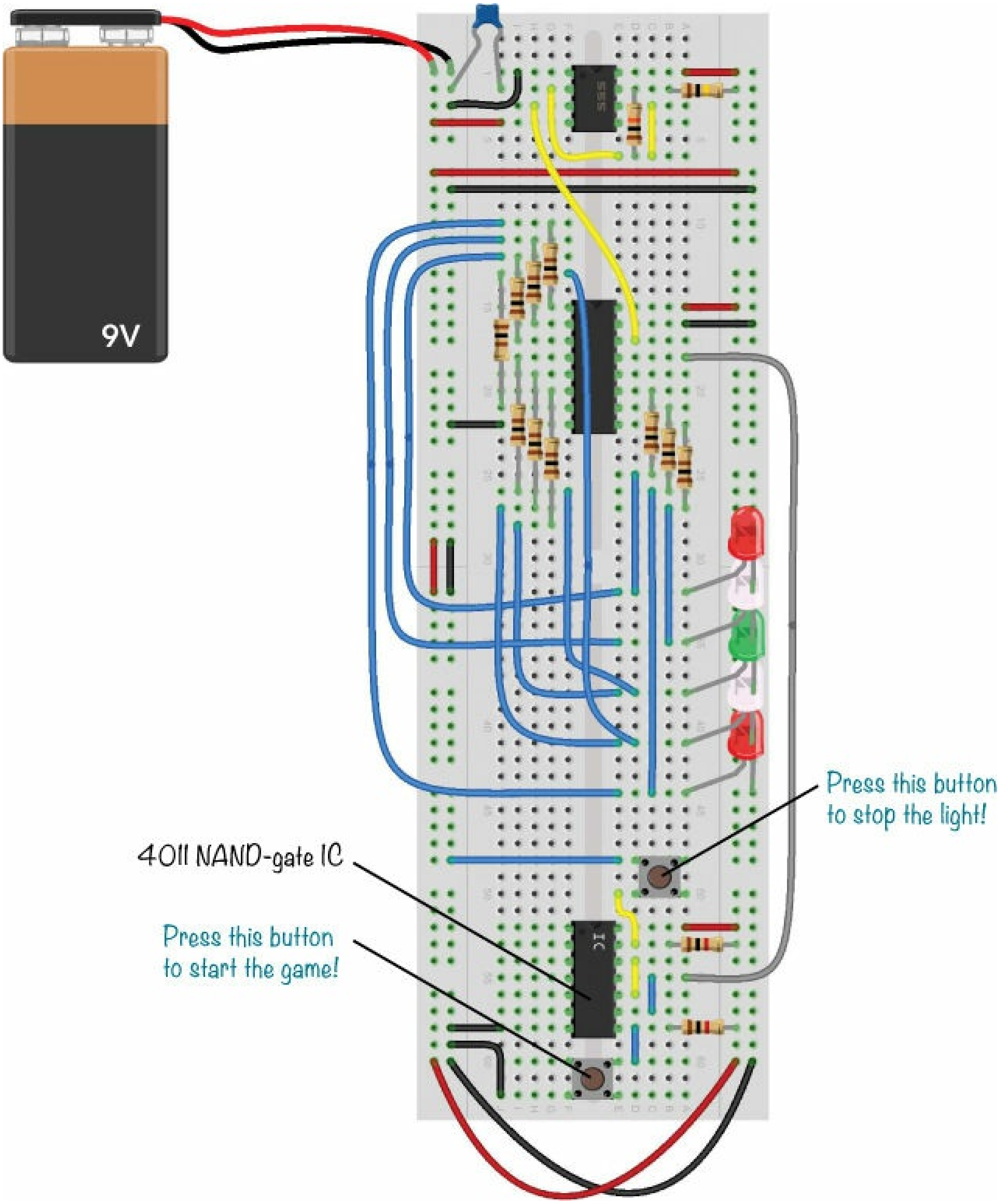
9V
555
4011 NAND-gate IC
Press this button to start the game!
Press this button to stop the light!

Step 4: Practice Your Reaction Time!

All that's left is to connect the battery to the supply columns. The button at the bottom of the board is the Reset button. Use this to start the game and to restart the game after each player attempts to stop the light.

The button next to the LEDs should stop the light when the game is running. See how many turns it takes you to get to 50 points!

Step 5: What If the Game Isn't Working?

If you've followed my instructions so far, the circuits from Steps 1 and 2 should be working. If your circuit isn’t working, the only sources of error left are the start/stop circuit you just built and the connection from this circuit to the 4017 decade counter.

A. Check the Continuity

First, check that you don’t have a short circuit between the positive and negative columns. To do this, use the *continuity* function on your multimeter. A continuity test checks for a direct connection between two points in a circuit. The symbol for the continuity tester usually looks like the one shown here.

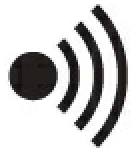

continuity
symbol

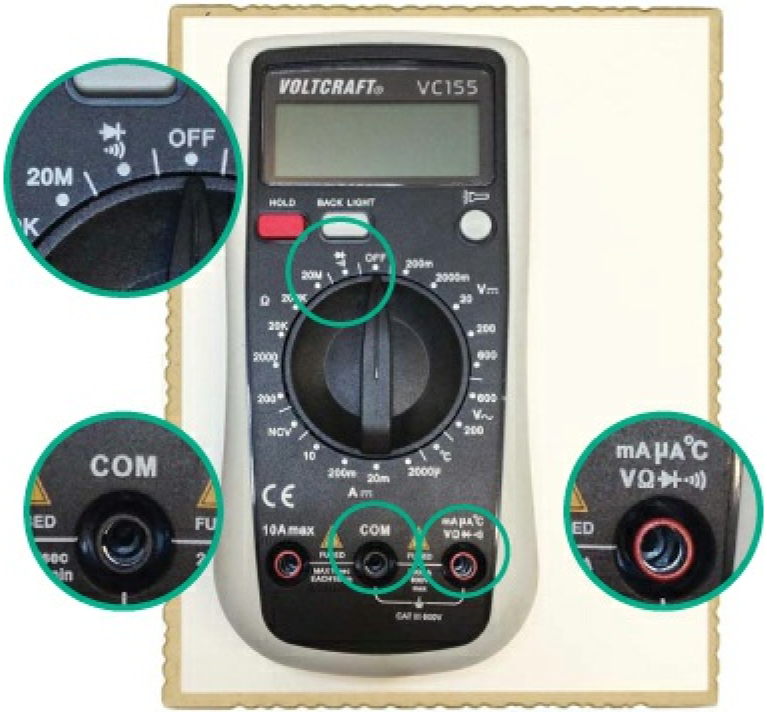

You don’t want a direct connection between the positive and negative columns because that would short-circuit the battery and stop the game from working. Use the continuity tester to check for short circuits.

Turn the dial on your multimeter so that it points toward the continuity symbol. Plug the black measurement lead into the multimeter’s COM socket, and plug the red measurement lead into the multimeter’s V socket. Touch the tip of the black and red measurement leads to each other, and you should hear a beep to indicate that there’s a direct connection.

NOTE

Many electronics enthusiasts also call continuity mode beep mode.

B. Check for Bad and Good Beeps

Now, plug your battery clip into the breadboard as you would normally, but without the battery. Touch one measurement lead tip to the positive connector and the other to the negative connector. If you hear a beep, there's a short circuit, and you need to fix it! Check all your connections to the positive and negative supply columns.

BS3604W
AUTO POWER OFF
POWER
HOLD
Ω
200M
20M
2M
200k
20k
2k
200
200mV
2V
20V
200V
600V
OUT
20uF
2uF

Next, check the connection between pin 11 on the 4011 NAND-gate IC and pin 13 on the 4017 decade counter to make sure they're connected correctly. Use the continuity tester to check that you have a connection by carefully touching the lead tips on the IC pins. There isn't much space between IC pins, so take care to be sure each tip only touches the correct pin. This time, a beep is a good sign.

C. Check for Power

If the connection to the NAND-gate IC is correct, use a multimeter to measure the output voltage from the start/stop circuit to see whether it's working correctly. Set your multimeter to measure voltage. Make sure the black measurement lead is connected to the multimeter's COM socket, and the red measurement lead is connected to the V socket.

Touch the tip of the black lead of the multimeter to the negative side of the battery, and touch the red lead to pin 11 on the 4011 NAND-gate IC. You should see a high signal—about 9 V—after clicking the stop button and a low signal—about 0 V—after pushing the start button. If not, check the connections of the SR latch circuit to find the error.

TRY IT OUT: CHANGE THE LIGHT'S SPEED

To change the speed and difficulty of the game, play around with different values for R1, R2, and C1 around the 555 timer. Smaller values will make the game go faster. Larger values will make the game go slower. Flip to "How to Set the Output Speed of the 555 Timer" on page 166 for the calculations to figure out specific resistor and capacitor values based on the frequency you want.

Note that R1 should not be less than 1 kΩ, as lower values might damage the 555 timer.

But what if you want to change the difficulty on the go? Just replace resistor R2 with a potentiometer. Then you can change that resistance value by rotating the potentiometer shaft, which changes the speed!

ADD A BUZZER TO YOUR GAME

Congratulations: You've finished the last project in the book! Now, it's up to you to decide what to make next. If you're not sure where to start, why not add more circuits to your reaction game?

The LED in the middle is where you want the light to stop, and I suggest adding a sound circuit to bring some excitement to hitting your target. To do this, you could use an active buzzer like the one in "Project #2: Intruder Alarm" on page 11, as shown in this partial circuit diagram.

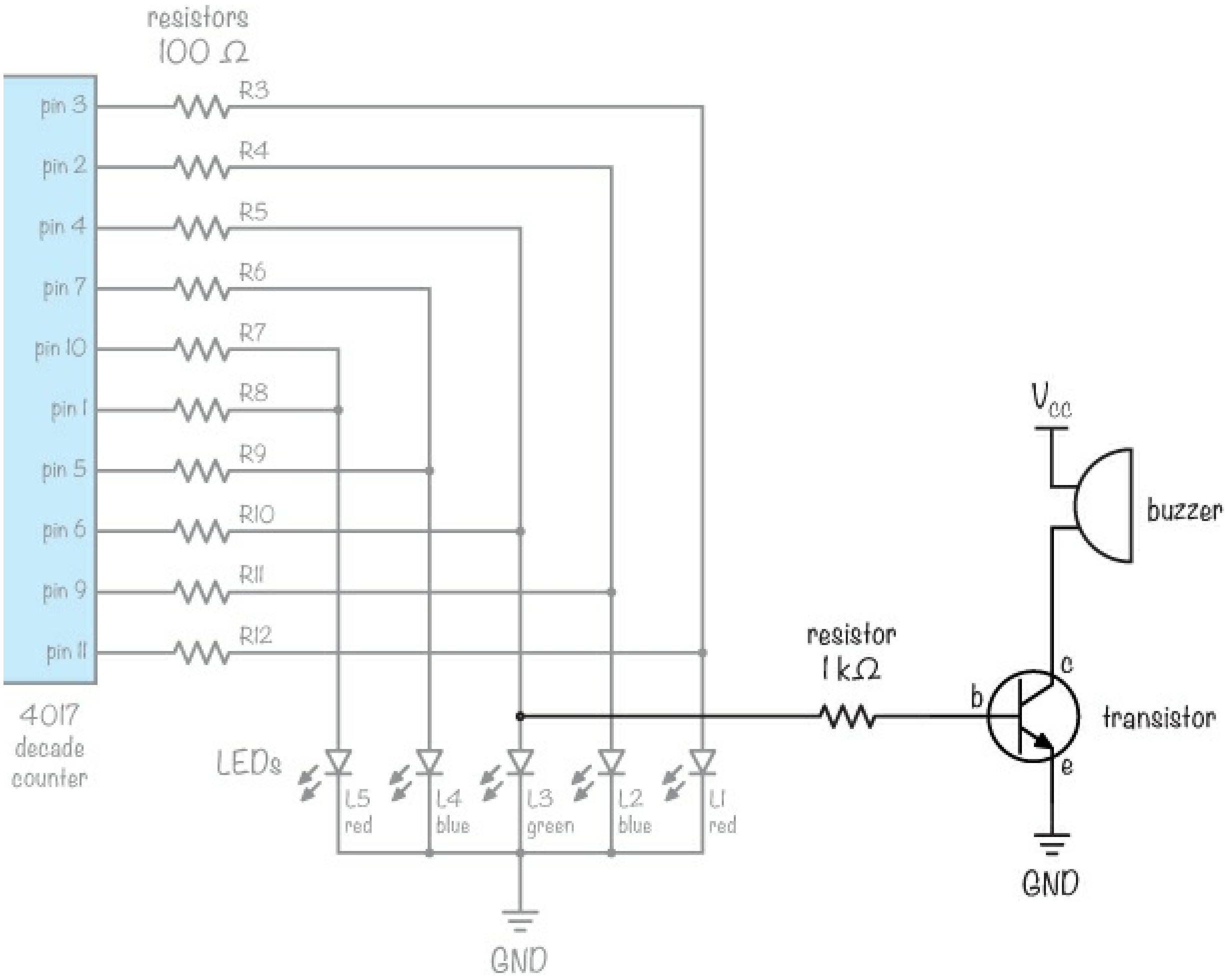
resistors
100 Ω
pin 3
pin 2
pin 4
pin 7
pin 10
pin 1
pin 5
pin 6
pin 9
pin 11
R3
R4
R5
R6
R7
R8
R9
R10
R11
R12
4017
decade
counter
LEDs
L5
red
L4
blue
L3
green
L2
blue
L1
red
GND
resistor
1 kΩ
V_{cc}
buzzer
b
c
e
transistor
GND

The darker part of this circuit shows new components you'd need in order to add a buzzer to the reaction game project. The lighter components are just a section of the original circuit diagram.

Connect the positive leg of the middle LED through a 1 kΩ resistor to the base of an NPN transistor. Then connect the buzzer to the transistor's collector. Connect the positive side of your battery to the other side of the buzzer, and connect the negative side of the battery to the transistor's emitter.

You should end up with a circuit that makes a little beep every time the light passes the middle LED. If you can stop the light on the middle LED, the buzzer should beep continuously to indicate that you've hit the main target.

When you've customized the game to your liking, solder it onto a prototyping board. Maybe you'll even want to place it in a nice box to hide the electronics and show only the buttons and LEDs.

WHAT'S NEXT? GO MAKE COOL STUFF!

I'm so happy you've read the book all the way to the end! I hope you've enjoyed the projects, and I hope you'll continue building cool things with electronics. One way to practice is to find a circuit diagram for something exciting, buy the components, and build the circuit. You can find circuit diagrams for almost anything online.

I'd also like to invite you to continue learning on my website, *http://www.ohmify.com/*. You can watch video courses, read lots of project tutorials, and visit a discussion forum where you can ask questions and make friends with people from all over the world who also like to build electronics.

Just make sure to ask your parents' permission to join the site, as it's subscription based. If they say yes, then use the following link for a special offer available only to the owners of this book: *http://ohmify.com/e4k/*.

Also be sure to check out the "Online Resources" on page 286—you'll find loads of tutorials and more circuits to build at those websites. Have fun!

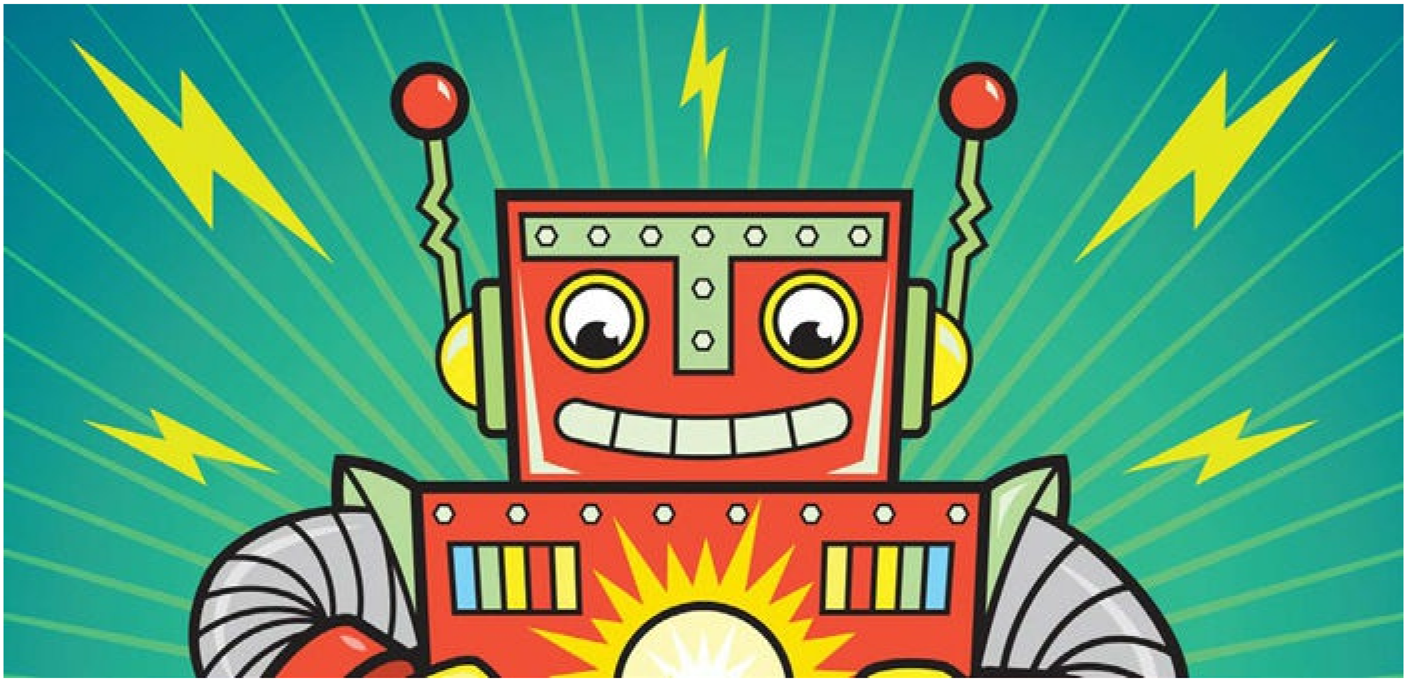

HANDY RESOURCES

Here are some resources that you'll find useful when building projects with electronics. Use this as a reference when, for example, you need to figure out what the color bands of a specific resistor mean. I've also listed some websites where you can continue learning and find more projects to build.

COMPONENT AND UNIT VALUE CHEAT SHEETS

Throughout this book, you'll use lots of components, and there are as many ways to read those components as there are component types. Here are some handy cheat sheets to help you read resistors and capacitors and to help you remember what the different prefixes on units like volts and amperes mean.

Resistor Color Codes

Most of the resistors in this book have four color bands. To determine a resistor's value, just look up its colors in the following diagram and multiply accordingly. For example, to get 470 Ω, you'd multiply the number 47 (given by the yellow and purple bands) by 10 (given by the brown band). For more details on resistors, see "Meet the Resistor" on page 70.

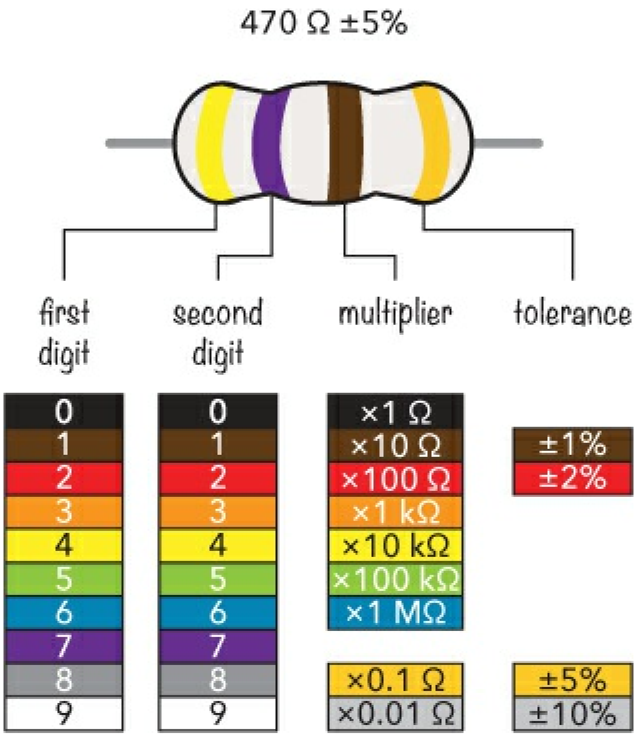

first digit	second digit	multiplier	tolerance
0	0	×1 Ω	
1	1	×10 Ω	±1%
2	2	×100 Ω	±2%
3	3	×1 kΩ	
4	4	×10 kΩ	
5	5	×100 kΩ	
6	6	×1 MΩ	
7	7		
8	8	×0.1 Ω	±5%
9	9	×0.01 Ω	±10%

Capacitor Codes

In the following table, I've listed the most common capacitor codes. Refer to this table when you're using ceramic or tantalum capacitors because, unlike the electrolytic capacitors used in much of this book, those won't have their capacitance written straight out for you.

Code	Picofarad (pF)	Nanofarad (nF)	Microfarad (µF)

101	100	0.1	0.0001
102	1,000	1	0.001
103	10,000	10	0.01
104	100,000	100	0.1
105	1,000,000	1,000	1

If you have a capacitor with a different code from those listed here, you can find the value in picofarads by taking the first two digits and adding the number of zeros of the third digit.

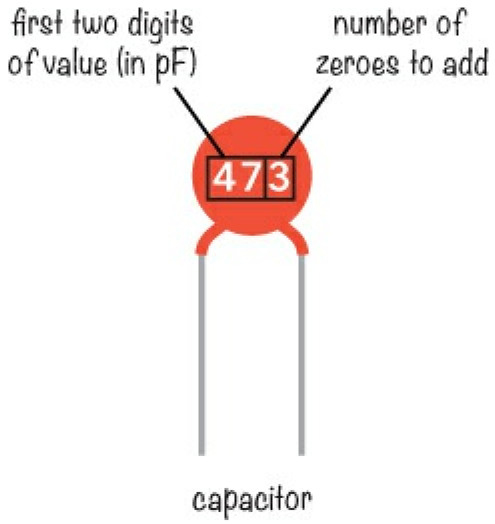

In this example, the capacitor has the code 473. Take the first two digits, 47, and add the number of zeros specified by the third digit, 3. That gives you 47,000 pF, which is 47 nF, or 0.047 µF.

Standard Prefixes

When building electronics projects, as in many areas of science, we sometimes have to deal with really small or really big numbers. Fortunately, there’s a set of standard prefixes in the International System of Units to make those numbers easier to write. The prefixes are multipliers, as shown in the table.

Prefix	Name	Multiply value by	Example usage
p	pico	× 0.000 000 000 001	Capacitor values (example: 47 pF capacitor)
n	nano	× 0.000 000 001	Capacitor values (example: 100 nF capacitor)
µ	micro	× 0.000 001	Capacitor values (example: 10 µF capacitor)
m	milli	× 0.001	Currents in a circuit (example: 20 mA current)

-	-	× 1	Voltages often don't have a prefix (example: 9 V battery)
k	kilo	× 1,000	Resistor values above 1,000 (example: 10 kΩ resistor)
M	mega	× 1,000,000	File sizes (example: 2MB photo)
G	giga	× 100,0000,000	File sizes (example: 1GB video)
T	tera	× 1,000,000,000,000	Hard disk sizes (example: 2TB hard drive)

A QUICK REVIEW OF OHM'S LAW

Ohm's law is such an essential part of calculating values in circuits that you'll keep coming back to it again and again as you build more projects. Whenever you need a refresher on figuring out a voltage, current, or resistance in a circuit, just flip to this section.

$V = I \times R$ Voltage (in volts) equals the current (in amps) multiplied by the resistance (in ohms)

$I = \frac{V}{R}$ Current (in amps) equals the voltage (in volts) divided by the resistance (in ohms)

$R = \frac{V}{I}$ Resistance (in ohms) equals the voltage (in volts) divided by the current (in amps)

In the Ohm's law equation, you must use volts (V), amps (A), and ohms (Ω), so remember to convert units if necessary: 1 mA = 0.001 A and 1 kΩ = 1,000 Ω.

A BASIC VOLTAGE DIVIDER CIRCUIT

The voltage divider is a circuit that is very useful, for example, when you have a sensor based on resistance, such as a thermistor, which senses temperature, or a photoresistor, which senses light. See "Project #15: Build a Sunrise Wake-Up Alarm" on page 148 for a project that uses a voltage divider like this. You can also use your knowledge of the voltage divider to calculate voltages within a circuit in order to understand what's going on.

When you have two resistors in series, they form a voltage divider. The input voltage gets divided between the two resistors, and the output voltage (across R2) is given by the formula:

$$V_{out} = V_{in} \times \frac{R2}{R1 + R2}$$

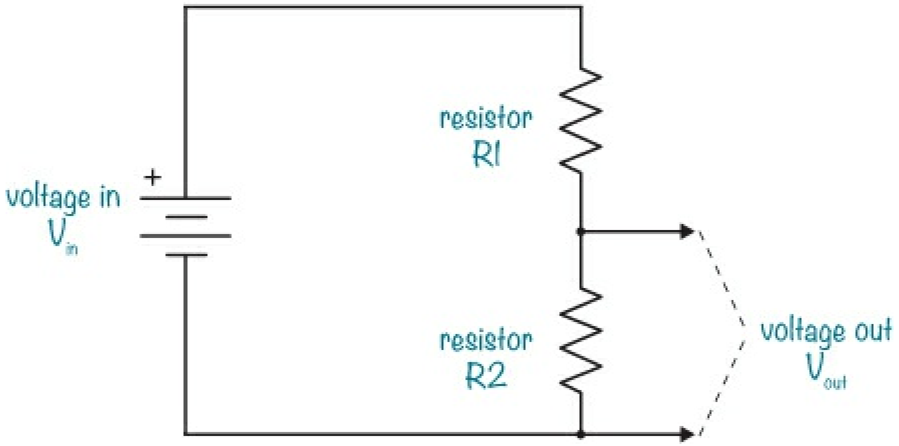

ONLINE ELECTRONICS SHOPS

Throughout the book, I recommend shops where you can buy your parts, but those aren't the only great electronics stores in the world! Try these, too:

- Adafruit (US) *www.adafruit.com*
- DigiKey (US) *www.digikey.com*
- Jameco (US) *www.jameco.com*
- SparkFun (US) *www.sparkfun.com*
- Bitsbox (UK) *www.bitsbox.co.uk*
- Quasar Electronics (UK) *www.quasarelectronics.co.uk*
- Rapid Electronics (UK) *www.rapidonline.com*
- Spiratronics (UK) *www.spiratronics.com*
- Farnell (worldwide) *www.farnell.com*
- Protostack (Australia) *www.protostack.com*
- Seeed Studio (China) *www.seeedstudio.com*
- Tayda Electronics (Thailand, US) *www.taydaelectronics.com*

ONLINE RESOURCES

When you're done with this book, you can keep learning about electronics online. (Ask your parents first!) You'll find tons of fun tutorials and other projects at these sites:

Adafruit (*https://learn.adafruit.com/*) Lots of guides based on the components they sell.

Build Electronic Circuits (*http://www.build-electronic-circuits.com/*) My personal blog where I post tutorials, videos, articles, and so on—all about electronics. I also have a free newsletter with useful tips and tricks for your projects.

Electronics Club (*http://www.electronicsclub.info/*) A website for anyone wishing to learn about electronics or build simple projects, created and maintained by John Hewes, the technical reviewer of this book.

Ohmify (*http://www.ohmify.com/*) My online learning platform with courses, project tutorials, discussion forums, and more. Get cool step-by-step project tutorials, ask questions, make friends, and learn. Owners of this book get a special offer by going to *http://www.ohmify.com/e4k/*.

SparkFun (*https://learn.sparkfun.com/*) Lots of guides based on the components they sell.

You can also visit this book's website at ***https://www.nostarch.com/electronicsforkids/*** for additional resources, updates, and more.

INDEX

NUMBERS AND SYMBOLS

- 555 timer, 164–166
  - making sound using, 167–175
  - setting output speed of, 166–167
  - using for reaction game speed, 260–261
- µF (microfarad), 92
- Ω (ohms), 7, 70

A

- A (amperes), 7
- AC (alternating current), 48–49
- adapters, 226
- alarm projects
  - intruder, 11–17, 108, 236–237
  - sunrise, 148–157
- alternating current (AC), 48–49
- American wire gauge (AWG), 83
- amperes (A), 7
- amps, 7
- AND gates, 216, 217
- AND operator, 215
- anode, 75
- ASCII code, 201–202
- astable mode, 165
- atoms, 6
- AWG (American wire gauge), 83

B

- barrel jack adapters, 226
- base, 132
- base-2 number system, 188
- base-10 number system, 188
- batteries
  - chemistry of, 56–58
  - food, 65
  - lemon, 58–66

  - what's inside, 55–56
- beep mode, 277
- binary numbers, 188–190
  - converting to decimal, 190–192
  - counting on fingers, 212
  - why computers use, 211
- bits, 193
- blinking lights
  - for reaction game, 265–278
  - using a relay, 99–100, 101–108
- Boolean logic, 214
- braid, desoldering, 125
- breadboards, 81–83, 84–87
- buzzer
  - adding to reaction game, 279–280
  - in intruder alarm, 13–14
  - in sunrise alarm, 153
- bytes, 193

C

- capacitance, 91
- capacitors, 90–91
  - codes for, 283
  - coupling, 171–172
  - polarized and nonpolarized, 91–92
  - testing, 92–95
- cathode, 75
- charge, electric, 6
- cheat sheets, 282
- circuit boards, 111
  - illustrations of, 121
- circuits, 8–10
  - diagrams of, 95–97
  - logic, 212
  - making reliable, 227
- clock signal, 242
- closed circuit, 8
- closed loop, 8
- codes

- for capacitors, 283
- for resistors, 70–72, 282
- secret, 218–219, 223–237

coin tosser project, 245–254

collector, 132

collector current (I_c), 222

color codes, for resistors, 70–72, 282

color guessing game project, 194–200

COM (common pin), 98

common anode, 195

common cathode, 195

common pin
- in relay, 98
- in switch, 29

components
- gathering, xxvii–xxviii
- where to buy, 286

conditions, logical, 214–215

conductive materials, 6

continuity, 276

counter, decade, 262

coupling capacitors, 171–172

current, 5, 7–8, 73, 222

D

datasheets, 103, 162

DC (direct current), 48–49

debugging, 174

decade counter, 262

decimal numbers, 188

desoldering, 125
- braid, 125
- how to, 126–129
- safety tip, 128

D flip-flop, 242

diaphragm, 163

DIP (dual in-line package) switches, 204

direct current (DC), 48–49

D latch, 241

drain, on FET, 260
dual in-line package (DIP) switches, 204
dynamo, 66

E

edge-triggering, 242
electrical current, 5, 7–8, 222
electrically charged particles, 6
electricity, 3–10
  creating from water or wind, 46
  generating with magnets, 44–46
electrodes, 56
  inserting into lemon, 61
electrolytes, 56
electromagnets, 19, 22–23
  creating your own, 23–31
  in loudspeakers, 163–164
  in relays, 97–100
electronics stores, 286
electrons, 6–8
emitter, 132

F

F (farads), 91
false (Boolean value), 214
fan, temperature-controlled, 157–158
farads (F), 91
FET (field-effect transistor), 260
flip-flop, 242
floating input, 227
frequency, 164, 166
Fritzing (software), 121

G

games
  color guessing, 194–200
  LED reaction, 265–278
generators, 45–46

  - creating your own, 49–54
  - using motors as, 55
- GND symbol, 259

H

- Hz (hertz), 164

I

- I_c (collector current), 222
- IC (integrated circuit), 161–163
- instrument, musical (project), 175–183
- insulation
  - adding to wire with marker, 38
  - removing from wire, 26–27
- integrated circuit (IC), 161–163
- International System of Units, 284
- intruder alarm project, 11–17, 108, 236–237
- inverter, 217

J

- joints, soldered, 113, 116, 117
- jumper wires, 83

K

- kilo prefix (k), 72
- kits, electronics, 129

L

- large values, prefixes for, 72, 284
- latching, 240–241
- LDRs (light-dependent resistors), 146
- LEDs (light-emitting diodes), 58–59
  - blinking, 101–108, 265–278
  - brightness of, varying with resistance, 144
  - controlling with transistor, 135–136
  - destroying, 74–76
  - guessing color game, 194–200
  - identifying legs, 75

  placing on breadboard, 85
  powering, 78–81
  reaction game, 265–278
  RGB (red-green-blue), 195
  using correctly, 76–78
lemon batteries, creating, 56, 58–66
light bulb, 4. *See also* LEDs (light-emitting diodes)
  connecting to battery, 5
  project, 4–5
light-dependent resistors (LDRs), 146
light-emitting diodes (LEDs). *See* LEDs (light-emitting diodes)
logic, 214–215
  circuit diagrams, how to draw, 218–220
  circuits, 212
  equation, 215
  gates, 215–218
    negative, 237–238
    in real life, 220–221
loudspeaker, 163–164

M

M (mega) prefix, 72
magnetic fields, 20, 44–45
magnets, 19–21. *See also* electromagnets
magnet wire, 54
measurement units
  amperes (A), 7
  farads (F), 91
  hertz (Hz), 164
  ohms (Ω), 7, 70
  prefixes for, 72, 77, 92, 284
  volts (V), 6
mega prefix (M), 72
memory
  circuits, 240, 241
  in computers, 211
messages, secret, 202–211
micro prefix (μ), 92
microfarad (μF), 92
motion-controlled instrument, 183

motors, 31–32
  creating your own, 32–40
  current needed for, 222
  using as generators, 55
  protecting transistors used with, 234
multimeter, 47–49
musical instrument project, 175–183

N

NAND gate, 237
nano prefix (n), 92
nanofarad (nF), 92
NC (normally closed), 98
negative edge-triggered circuit, 242
negative logic gates, 237–238
negative supply column, 81
neutrons, 6
nF (nanofarad), 92
NO (normally open), 98
nonpolarized capacitors, 91
NOR gate, 237
normally closed (NC), 98
normally open (NO), 98
north pole, on a magnet, 20–21
NOT gate, 217
NPN transistor, 132, 134
nucleus, 6
numbers, binary and decimal, 188

O

Ohmify (website), 280
ohms (Ω), 7, 70
Ohm's law, 73–74, 284
ones and zeros. *See* binary numbers
online resources, 286
OR gate, 216
oscillator, 245

P

p (pico) prefix, 92

parts

  gathering, xxvii–xxviii

  where to buy, 286

pF (picofarad), 92

photoresistors, 146, 148, 149, 183

pico prefix (p), 92

picofarad (pF), 92

pinout, 103, 165

pins

  in IC, 162

  in potentiometer, 145

  in relay, 98

  in switch, 28–29

pixels, 193–194

polarized components

  capacitors, 91–92

  LEDs, 75

poles, on a magnet, 20–21

positive edge-triggered circuit, 242

positive supply column, 81

potato batteries, 65

potentiometers, 145, 146, 149

power, 45

power plants, 46

prefixes, units, 72, 77, 92, 284

projects

  alarms

    intruder, 11–17

    sunrise, 148–157

  binary numbers, converting to decimal, 190–192

  breadboard circuit, 84–87

  capacitor, testing, 92–95

  coin tosser, 245–254

  color guessing game, 194–200

  desoldering, 126–129

  electromagnet, 23–31

  generator, 49–54

  - LEDs
    - destroying, 74–76
    - powering, 78–81
  - lemon battery, 58–66
  - light
    - blinking, 101–108
    - turning on, 4–5
  - motor, 32–40
  - musical instrument, 175–183
  - reaction game, 265–280
  - secret code checker, 223–235
  - secret message machine, 202–211
  - soldering, 118–125
  - sound, with 555 timer, 167–175
  - touch sensor, 136–144
- protons, 6
- prototyping boards, 120, 129, 140
- pull-down resistor, 227
- push buttons, 180–181

R

- reaction game project, 265–278
- red-green-blue (RGB) LEDs, 195
- relays, 97–100
  - adding to intruder alarm project, 108, 236–237
  - blinking a light using, 99–100
- resistance, 5, 7–8, 73
  - calculating for LEDs, 77–78
  - illustration of, 8
  - variable, 144–145
- resistors, 70
  - color codes for, 70–72, 282
  - materials in, 72
  - placing on breadboard, 85
  - pull-down, 227
  - variable, 146
- resources, 281–288
- RGB (red-green-blue) LEDs, 195
- rotor, 33, 35–36

S

safety, xxviii–xxix
  with battery lemons, 58
  soldering, 113–114, 128
  trimming wires, 122–123
  using resistors, 73
  using supermagnets, 34
schematics, 95–97
screw terminals, 226
secret code checker
  adding to intruder alarm project, 236–237
  project, 223–235
secret message machine project, 202–211
sensors, touch, 136, 144
series, batteries connected in, 57–58
shake generator project, 49–54
short circuits, 156, 174
single-strand wires, 83
small values, prefixes for, 77, 284
solder, 111–112
  melting temperature, 112
  wick, 125
soldering
  avoiding bad joints, 117
  desoldering, 125–129
  how to, 112–117
  iron, 113
    cleaning, 115
    stand, 114
    tinning, 115
  safety tips, 113–114, 128
  supplies, 112
sound, 163–164
  555 timer project, 167–175
  creating with electricity, 163–164
  musical instrument project, 175–183
south pole, on a magnet, 20–21
sponge (for cleaning soldering iron), 115
SR latch, 240–241, 263–264

stripping wires, 26–27
sunrise wake-up alarm project, 148–157
supplies
    gathering, xxvii–xxviii
    where to buy, 286
switches, 10–11
    connecting typical, 28–29
    controlling a light bulb with, 11
    DIP, 204
    push buttons, 180–181
    relays, 97–98
    symbol for, 96
    vs. transistors, 133–134
synthesizer, 175

T

teamwork, for debugging, 174
temperature-controlled fan, 157–158
thermistors, 157
timer (555), 164–166
    making sound using, 167–175
    setting output speed of, 166
    using for reaction game speed, 260–261
tinning, 115
toggling output, with D flip-flop, 246
tolerance, of resistors, 72
touch sensor project, 136–144
transistors, 132–136
    bipolar junction, 134
    controlling LEDs with, 135–136
    current, finding maximum, 222
    FET (field-effect), 260
    how they work, 134
    NPN, 132, 134
    why to use, 133–134
true (Boolean value), 214
truth tables, 216–217, 238, 240, 241
tutorials, online, 286

U

units of measurement
  amperes (A), 7
  farads (F), 91
  hertz (Hz), 164
  ohms (Ω), 7, 70
  prefixes for, 72, 77, 92, 284
  volts (V), 6

V

V (volts), 6
variable resistance, 144, 145
variable resistors, 146
V_{CC} symbol, 259, 260
V_{DD} symbol, 259, 260
voltage, 5, 6–7, 73
  of batteries, 57–58
  dividers, 146, 147, 285
    calculating the voltage from, 147
    measuring light with, 148
  how to measure, 47–48, 54
  from wall adapters, 226
volts (V), 6

W

wall adapter, 226
water, generating electricity with, 46
water analogy, for electrical current, 9–10
wick, solder, 125
wind, generating electricity with, 46
wiper (pin), 145
wire cutter, 26–27
wires
  adding insulation to, 38
  connecting, 16
  hookup, 54
  jumper, 83
  preparing for lemon battery, 60
  single-strand, 83
  stripping insulation from, 26–27

X

XOR gate, 238

Electronics for Kids is set in Century Schoolbook, Filmotype Candy, Housearama Kingpin, TheSansMono Condensed, and Billy the Flying Robot. The book was printed and bound by TC Transcontinental Printing in Beauceville, Québec, Canada. The paper is 70# Husky Offset, which is certified by the Forest Stewardship Council (FSC).

The book uses an Otabind binding, in which the pages are bound together with a cold-set, flexible glue, and the first and last pages of the resulting book block are attached to the cover. The cover is not actually glued to the book's spine, and when open, the book lies flat and the spine doesn't crack.

RESOURCES

Visit *https://www.nostarch.com/electronicsforkids/* for updates, errata, and more information.

MORE SMART BOOKS FOR CURIOUS KIDS!

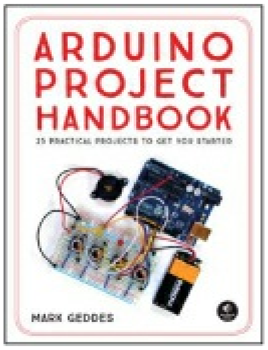

ARDUINO PROJECT HANDBOOK

25 Practical Projects to Get You Started

by MARK GEDDES
JUNE 2016, 272 PP., $24.95
ISBN 978-1-59327-690-4
full color

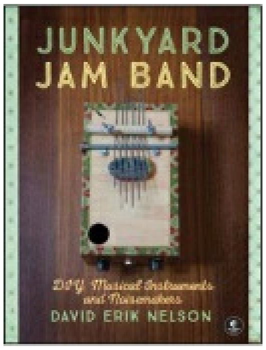

JUNKYARD JAM BAND

DIY Musical Instruments and Noisemakers

by DAVID ERIK NELSON
OCTOBER 2015, 408 PP., $24.95
ISBN 978-1-59327-611-9

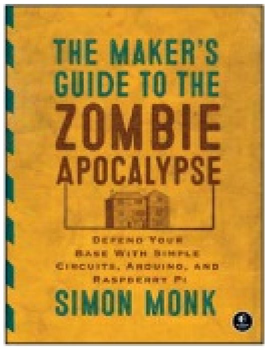
THE MAKER'S GUIDE TO THE ZOMBIE APOCALYPSE
DEFEND YOUR BASE WITH SIMPLE CIRCUITS, ARDUINO, AND RASPBERRY PI
SIMON MONK

THE MAKER’S GUIDE TO THE ZOMBIE APOCALYPSE

Defend Your Base with Simple Circuits, Arduino, and Raspberry Pi

by SIMON MONK
OCTOBER 2015, 296 PP., $24.95
ISBN 978-1-59327-667-6

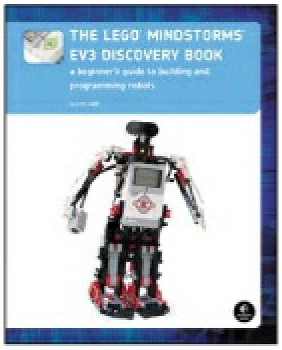
THE LEGO MINDSTORMS EV3 DISCOVERY BOOK

THE LEGO® MINDSTORMS® EV3 DISCOVERY BOOK

A Beginner’s Guide to Building and Programming Robots

by LAURENS VALK
JUNE 2014, 396 PP., $34.95
ISBN 978-1-59327-532-7
full color

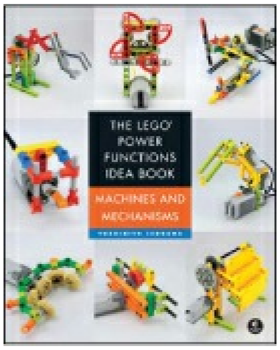
THE LEGO POWER FUNCTIONS IDEA BOOK
MACHINES AND MECHANISMS

THE LEGO POWER FUNCTIONS IDEA BOOK, VOL. 1

Machines and Mechanisms

by YOSHIHITO ISOGAWA
OCTOBER 2015, 324 PP., $24.95
ISBN 978-1-59327-688-1
full color

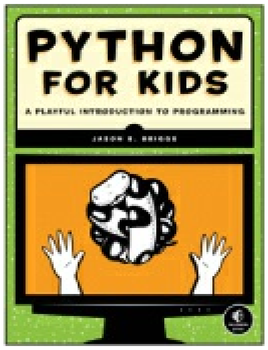

PYTHON FOR KIDS

A Playful Introduction to Programming

by JASON R. BRIGGS
DECEMBER 2012, 344 PP., $34.95
ISBN 978-1-59327-407-8
full color

800.420.7240 or 415.863.9900 | sales@nostarch.com | www.nostarch.com

REAL ELECTRONICS.
REAL EASY.

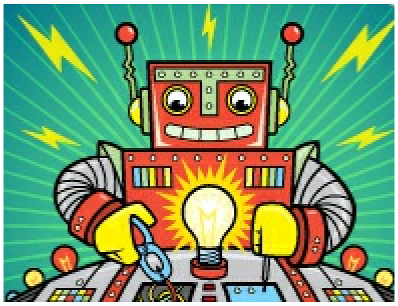

Why do the lights in a house turn on when you flip a switch? How does a remote-controlled car move? And what makes lights on TVs and microwaves blink? The technology around you may seem like magic, but most of it wouldn't run without electricity.

Electronics for Kids demystifies electricity with a collection of awesome hands-on projects. In Part 1, you'll learn how current, voltage, and circuits work by making a battery out of a lemon, turning a metal bolt into an electromagnet, and transforming a paper cup and some magnets into a spinning motor. In Part 2, you'll make even more cool stuff as you:

- Solder a blinking LED circuit with resistors, capacitors, and relays
- Turn a circuit into a touch sensor using your finger as a resistor
- Build an alarm clock triggered by the sunrise
- Create a musical instrument that makes sci-fi sounds

Then, in Part 3, you'll learn about digital electronics—things like logic gates and memory circuits—as you make a secret code checker and an electronic coin flipper. Finally, you'll use everything you've learned to make the LED Reaction Game—test your reaction time as you try to catch a blinking light!

With its clear explanations and assortment of hands-on projects, *Electronics for Kids* will have you building your own circuits in no time.

ABOUT THE AUTHOR

Øyvind Nydal Dahl built his first circuit at 14 and has been passionate about electronics ever since. He has a master's degree in electronics from the University of Oslo, helps companies develop new products, and travels the world while teaching electronics workshops. He also writes beginner-friendly tutorials at *http://www.build-electronic-circuits.com/*.

no starc press

THE FINEST IN GEEK ENTERTAINMENT™

Footnotes

Chapter 1: What Is Electricity?

* The electron is a negative particle, but in some materials the current is made up of positive particles instead, and they flow in the opposite direction. So, at the atomic level, the particles can flow in either direction.

Chapter 9: How Circuits Understand Ones and Zeros

* Sometimes kilobyte refers to 1024 bytes (2^10).

TECHNOLOGY IN ACTION™

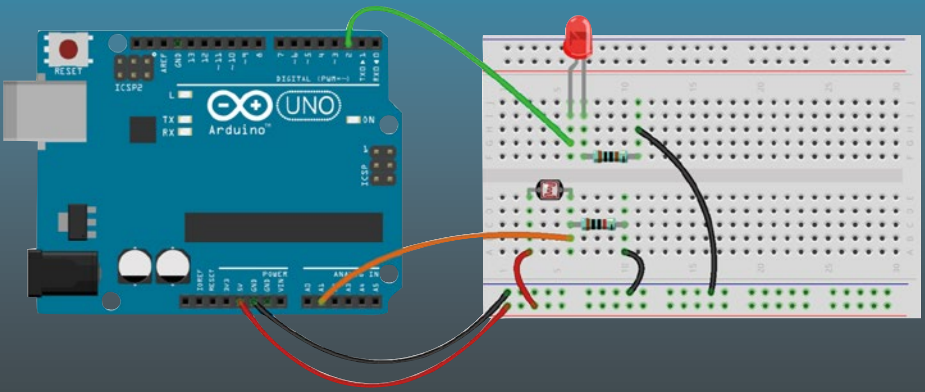
RESET
ICSP2
UNO
Arduino
TX
RX
ON
DIGITAL (PWM~)
POWER
ANALOG IN

Electronics for Beginners

A Practical Introduction to Schematics, Circuits, and Microcontrollers

—

Jonathan Bartlett

Apress®

Electronics for Beginners

A Practical Introduction to Schematics, Circuits, and Microcontrollers

Jonathan Bartlett

Apress®

Electronics for Beginners: A Practical Introduction to Schematics, Circuits, and Microcontrollers

Jonathan Bartlett
Tulsa, OK, USA

ISBN-13 (pbk): 978-1-4842-5978-8 ISBN-13 (electronic): 978-1-4842-5979-5
https://doi.org/10.1007/978-1-4842-5979-5

Copyright © 2020 by Jonathan Bartlett

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

Trademarked names, logos, and images may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, logo, or image we use the names, logos, and images only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Managing Director, Apress Media LLC: Welmoed Spahr
Acquisitions Editor: Natalie Pao
Development Editor: James Markham
Coordinating Editor: Jessica Vakili

Distributed to the book trade worldwide by Springer Science+Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a **Delaware** corporation.

For information on translations, please e-mail booktranslations@springernature.com; for reprint, paperback, or audio rights, please e-mail bookpermissions@springernature.com.

Apress titles may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Print and eBook Bulk Sales web page at http://www.apress.com/bulk-sales.

Any source code or other supplementary material referenced by the author in this book is available to readers on GitHub via the book's product page, located at www.apress.com/978-1-4842-5978-8. For more detailed information, please visit http://www.apress.com/source-code.

Printed on acid-free paper

This book is dedicated to Forrest M. Mims III, whose Engineer's Mini-Notebook series of books I read endlessly as a youth and whose work as a citizen scientist has been an inspiration to me and to many others.

Table of Contents

About the Author xix

About the Technical Reviewer xxi

Acknowledgments xxiii

Chapter 1: Introduction 1

- 1.1 Working the Examples 2
- 1.2 Initial Tools and Supplies 3
- 1.3 Safety Guidelines 5
- 1.4 Electrostatic Discharge 7
- 1.5 Using Your Multimeter Correctly 7

Chapter 2: Dealing with Units 9

- 2.1 SI Units 9
- 2.2 Scaling Units 12
- 2.3 Using Abbreviations 14
- 2.4 Significant Figures 15
- Apply What You Have Learned 18

Part I: Basic Concepts 21

Chapter 3: What Is Electricity? 23

- 3.1 Charge 23
- 3.2 Measuring Charge and Current 26
- 3.3 AC vs. DC 28

3.4 Which Way Does Current Flow? 30
Review 31
Apply What You Have Learned 32

Chapter 4: Voltage and Resistance 35
4.1 Picturing Voltage 35
4.2 Volts Are Relative 36
4.3 Relative Voltages and Ground Potential 37
4.4 Resistance 38
Review 41
Apply What You Have Learned 42

Chapter 5: Your First Circuit 45
5.1 Circuit Requirements 45
5.2 Basic Components 47
5.3 Creating Your First Circuit 49
5.4 Adding Wires 52
5.5 Drawing Circuits 53
5.6 Drawing the Ground 56
Review 57
Apply What You Have Learned 58

Chapter 6: Constructing and Testing Circuits 61
6.1 The Solderless Breadboard 61
6.2 Putting a Circuit onto a Breadboard 64
6.3 Using Fewer Wires 69
6.4 Testing Circuits with a Multimeter 72
6.5 Using a Multimeter with a Breadboard 75
6.6 Measuring Current with a Multimeter 77

6.7 Using a Power Regulator....78
Review....80
Apply What You Have Learned....83
Chapter 7: Analyzing Series and Parallel Circuits....85
7.1 Series Circuits....85
7.2 Parallel Circuits....87
  7.2.1 Kirchhoff's Current Law....88
  7.2.2 Kirchhoff's Voltage Law....90
7.3 Equivalent Parallel Resistance....93
7.4 Wires in a Circuit....97
7.5 Wiring Parallel Circuits onto a Breadboard....99
Review....100
Apply What You Have Learned....102
Chapter 8: Diodes and How to Use Them....105
8.1 Basic Diode Behavior....105
8.2 Circuit Calculations with Diodes in Series....107
8.3 Circuit Calculations with Diodes in Parallel....109
8.4 Diode Short Circuits....113
8.5 Nonconducting Diodes....115
8.6 Usage of Diodes....115
8.7 Other Types of Diode Protection....120
8.8 Zener Diodes....122
8.9 Schottky Diode....123
8.10 Diode-Like Behavior in Other Components....123
Review....123
Apply What You Have Learned....126

Chapter 9: Basic Resistor Circuit Patterns 129

9.1 Switches and Buttons 129

9.2 Current-Limiting Resistor Pattern 131

9.3 Voltage Divider Pattern 132

9.3.1 Calculating the Voltages 134

9.3.2 Finding Resistor Ratios..135

9.3.3 Finding Resistor Values 136

9.3.4 General Considerations 138

9.4 The Pull-Up Resistor...138

9.5 Pull-Down Resistors..141

Review 141

Apply What You Have Learned..143

Chapter 10: Understanding Power 145

10.1 Important Terms Related to Power..145

10.2 Power in Electronics 147

10.3 Component Power Limitations 149

10.4 Handling Power Dissipation with Heatsinks 150

10.5 Transforming Power...151

10.6 Amplifying Low-Power Signals 153

Review 154

Apply What You Have Learned..156

Chapter 11: Integrated Circuits and Resistive Sensors 159

11.1 The Parts of an Integrated Circuit 159

11.2 The LM393 Voltage Comparator 163

11.3 The Importance and Problems of Datasheets 164

11.4 A Simple Circuit with the LM393...167

11.5 Resistive Sensors and Voltages 169
11.6 Sensing and Reacting to Darkness 170
Sources and Sinks 172
Review 173
Apply What You Have Learned 175

Part II: Digital Electronics and Microcontrollers 177

Chapter 12: Using Logic ICs 179
12.1 Logic ICs 179
12.2 Getting a 5 V Source 185
12.3 Pull-Down Resistors 187
12.4 Combining Logic Circuits 190
12.5 Understanding Chip Names 193
Review 196
Apply What You Have Learned 197

Chapter 13: Introduction to Microcontrollers 201
13.1 The ATmega328/P Chip 203
13.2 The Arduino Environment 204
13.3 The Arduino Uno 206
13.4 Programming the Arduino 207
Review 210
Apply What You Have Learned 211

Chapter 14: Building Projects with Arduino 213
14.1 Powering Your Breadboard from an Arduino Uno 213
14.2 Wiring Inputs and Outputs to an Arduino Uno 214
14.3 A Simple Arduino Project with LEDs 216

14.4 Changing Functionality Without Rewiring 219
Review 220
Apply What You Have Learned...221
Chapter 15: Analog Input and Output on an Arduino........................223
15.1 Reading Analog Inputs 223
15.2 Analog Output with PWM 226
Review 229
Apply What You Have Learned...231
Part III: Capacitors and Inductors..233
Chapter 16: Capacitors...235
16.1 What Is a Capacitor? 235
16.2 How Capacitors Work...237
16.3 Types of Capacitors..241
16.4 Charging and Discharging a Capacitor...243
16.5 Series and Parallel Capacitances...245
16.6 Capacitors and AC and DC 248
16.7 Using Capacitors in a Circuit..249
Review 251
Exercises...253
Chapter 17: Capacitors as Timers 255
17.1 Time Constants 255
17.2 Constructing a Simple Timer Circuit 258
17.3 Resetting Our Timer 263
Review 265
Apply What You Have Learned...267

Chapter 18: Introduction to Oscillator Circuits....................................269
18.1 Oscillation Basics....................................269
18.2 The Importance of Oscillating Circuits....................................271
18.3 Building an Oscillator....................................273
18.4 Calculating On and Off Times with the 555....................................279
18.5 Choosing the Capacitor....................................283
Review....................................284
Apply What You Have Learned....................................287

Chapter 19: Producing Sound with Oscillations....................................289
19.1 How Sound Is Produced by Speakers....................................289
19.2 Graphing Electricity....................................290
19.3 Outputting a Tone to Headphones....................................292
19.4 AC vs. DC....................................294
19.5 Using Capacitors to Separate AC and DC Components....................................295
19.6 Speaker Wattage....................................297
19.7 Sound Control....................................298
Review....................................300
Apply What You Have Learned....................................302

Chapter 20: Inductors....................................303
20.1 Inductors, Coils, and Magnetic Flux....................................303
  20.1.1 What Is an Inductor?....................................303
  20.1.2 What Is Magnetic Flux?....................................304
  20.1.3 What Is the Difference Between Electric and Magnetic Fields....................................304
20.2 Induced Voltages....................................306
20.3 Resisting Changes in Current....................................307
20.4 Analogy from Mechanics....................................308

20.5 Uses of Inductors 308
20.6 Inductive Kick........ 309
Review 311
Apply What You Have Learned........ 312

Chapter 21: Inductors and Capacitors in Circuits........ 315
21.1 RL Circuits and Time Constants 315
21.2 Inductors and Capacitors as Filters 317
21.3 Parallel and Series Capacitors and Inductors 318
Review 319
Apply What You Have Learned........ 320

Chapter 22: Reactance and Impedance 323
22.1 Reactance 323
22.2 Impedance 326
22.3 RLC Circuits........ 329
22.4 Ohm’s Law for AC Circuits........ 331
22.5 Resonant Frequencies of RLC Circuits 334
22.6 Low-Pass Filters 335
22.7 Converting a PWM Signal into a Voltage 336
Review 337
Exercises........ 339

Chapter 23: DC Motors........ 343
23.1 Theory of Operation........ 343
23.2 Important Facts About Motors........ 344
23.3 Using a Motor in a Circuit........ 345
23.4 Attaching Things to Motors 347

23.5 Bidirectional Motors....348
23.6 Servo Motors....349
23.7 Stepper Motors....350
Review....350
Apply What You Have Learned....352

Part IV: Amplification Circuits....353

Chapter 24: Amplifying Power with Transistors....355

24.1 An Amplification Parable....356
24.2 Amplifying with Transistors....357
24.3 Parts of the BJT....358
24.4 NPN Transistor Operation Basics....361
  Rule 1: The Transistor Is Off by Default....361
  Rule 2: V_{BE} Needs to Be 0.6 V to Turn the Transistor On....361
  Rule 3: V_{BE} Will Always Be Exactly 0.6 V When the Transistor Is On....361
  Rule 4: The Collector Should Always Be More Positive Than the Emitter....362
  Rule 5: When the Transistor Is On, I_{CE} Is a Linear Amplification of I_{BE}....362
  Rule 6: The Transistor Cannot Amplify More Than the Collector Can Supply....362
  Rule 7: If the Base Voltage Is Greater Than the Collector Voltage, the Transistor Is Saturated....363
24.5 The Transistor as a Switch....364
24.6 Connecting a Transistor to an Arduino Output....367
24.7 Stabilizing Transistor Beta With a Feedback Resistor....368
24.8 A Word of Caution....370
Review....370
Apply What You Have Learned....372

Chapter 25: Transistor Voltage Amplifiers..375
25.1 Converting Current into Voltage with Ohm's Law....................................375
25.2 Reading the Amplified Signal....................................378
25.3 Amplifying an Audio Signal....................................380
25.4 Adding a Second Stage....................................385
25.5 Using an Oscilloscope....................................387
Review....................................388
Apply What You Have Learned....................................390
Chapter 26: Examining Partial Circuits..391
26.1 The Need for a Model....................................391
26.2 Calculating Thévenin Equivalent Values....................................393
26.3 Another Way of Calculating Thévenin Resistance....................................396
26.4 Finding the Thévenin Equivalent of an AC Circuit with Reactive Elements....................................397
26.5 Using Thévenin Equivalent Descriptions....................................398
26.6 Finding Thévenin Equivalent Circuits Experimentally....................................399
Review....................................403
Apply What You Have Learned....................................404
Chapter 27: Using Field Effect Transistors for Switching and Logic Applications..407
27.1 Operation of a FET....................................408
27.2 The N-Channel Enhancement Mode MOSFET....................................410
27.3 Using a MOSFET....................................410
27.4 MOSFETs in Logic Circuits....................................414
Review....................................415
Apply What You Have Learned....................................416

Chapter 28: Going Further 417

Appendix A: Glossary...419

Appendix B: Electronics Symbols 435

Appendix C: Integrated Circuit Naming Conventions........................439

C.1 Logic Chip Basic Conventions...439

Appendix D: More Math Than You Wanted to Know..........................443

D.1 Basic Formulas 443

D.1.1 Charge and Current Quantities 443

D.1.2 Volt Quantities 444

D.1.3 Resistance and Conductance Quantities.......................................444

D.1.4 Ohm's Law...445

D.1.5 Power 445

D.1.6 Capacitance...445

D.1.7 Inductance...446

D.2 Semiconductors...447

D.2.1 Diodes 447

D.2.2 NPN BJT 448

D.3 DC Motor Calculations 448

D.4 555 Timer Oscillator Frequency Equation 450

D.5 Output Gain Calculations in BJT Common Emitter Applications 452

D.6 The Thévenin Formula...461

D.7 Electronics and Calculus..463

D.7.1 Current and Voltage..463

D.7.2 Capacitors and Inductors 464

D.7.3 Time Constants...465

Appendix E: Simplified Datasheets for Common Devices 467

E.1 Batteries..468

E.1.1 Overview..468

E.1.2 Variations...468

E.1.3 Notes 469

E.2 Resistors 469

E.2.1 Overview..469

E.2.2 Finding a Resistor Value 470

E.3 Diodes 471

E.3.1 Overview..471

E.3.1 Variations...472

E.3.3 Forward Voltage Drop 472

E.3.4 Usages...473

E.4 Capacitors 473

E.4.1 Overview..473

E.4.2 Variations...474

E.4.3 Finding Capacitance Values...475

E.5 Inductors 475

E.5.1 Overview..475

E.5.2 Uses..476

E.5.3 Inductive Kick 476

E.5.4 Inductor Color Codes 477

E.6 NPN BJTs 477

E.6.1 Overview..477

E.6.2 Variations...478

E.6.3 Pin Configuration 478

E.6.4 Design Considerations...479

E.7 YwRobot Power Module 479
    E.7.1 Overview 479
    E.7.2 Variations 480
    E.7.3 Pin Configuration 480
    E.7.4 Limitations 480
E.8 555 Timer 481
    E.8.1 Overview 481
    E.8.2 Variations 482
    E.8.3 Pin Configuration 482
    E.8.4 Specifications 483
    E.8.5 Implementation Example 483
E.9 LM393 and LM339 Voltage Comparator 483
    E.9.1 Overview 483
    E.9.2 Variations 484
    E.9.3 Specifications 484
E.10 CD4081 and 7408 Quad-AND Gate 485
    E.10.1 Overview 485
    E.10.2 Variations 485
    E.10.3 Specifications (CD4081) 486
    E.10.4 Specifications (7408) 486
E.11 CD4071 and 7432 Quad-OR Gate 487
    E.11.1 Overview 487
    E.11.2 Variations 487
    E.11.3 Specifications (CD4071) 488
    E.11.4 Specifications (7432) 488

E.12 CD4001 and 7402 Quad-NOR Gate....489
  E.12.1 Overview....489
  E.12.2 Variations....489
  E.12.3 Specifications (CD4001)....490
  E.12.4 Specifications (7402)....490
E.13 CD4011 and 7400 Quad-NAND Gate....491
  E.13.1 Overview....491
  E.13.2 Variations....491
  E.13.3 Specifications (CD4011)....492
  E.13.4 Specifications (7400)....492
E.14 CD4070 and 7486 Quad-XOR Gate....493
  E.14.1 Overview....493
  E.14.2 Variations....493
  E.14.3 Specifications (CD4070)....494
  E.14.4 Specifications (7486)....494
E.15 LM78xx Voltage Regulator....495
  E.15.1 Overview....495
  E.15.2 Variations....495
  E.15.3 Specifications....496
  E.15.4 Usage Notes....497

Index....499

About the Author

Jonathan Bartlett is a senior software R&D specialist at Specialized Bicycle Components, focusing on creating initial prototypes for a variety of IoT (Internet of Things) projects. Jonathan has been educating the tech community for well over a decade. His first book, *Programming from the Ground Up*, is an Internet classic and was endorsed by Joel Spolsky, co-founder of Stack Exchange. It was one of the first open source books and has been used by a generation of programmers to learn how computers work from the inside out, using assembly language as a starting point. He recently released *Building Scalable PHP Web Applications Using the Cloud* as well as the calculus textbook *Calculus from the Ground Up*. Jonathan also writes a mix of technical and popular articles for a number of websites, including the new MindMatters.ai technology blog. His other articles can be found on IBM's DeveloperWorks website, Linux.com, and Medium.com. He is also the head of Tulsa Open Source Hardware, a local group focusing on do-it-yourself electronics projects.

Jonathan also participates in a variety of academic work. He is an associate fellow of the Walter Bradley Center for Natural and Artificial Intelligence. There, he does research into fundamental mathematics and the mathematics of artificial intelligence. He also serves on the editorial board for the journal *BIO-Complexity*, focusing on reviewing information-theoretic papers for the journal and assisting with LaTeX typesetting.

Additionally, Jonathan has written several books on the interplay of philosophy, math, and science, including *Engineering and the Ultimate* and *Naturalism and Its Alternatives in Scientific Methodologies*. Jonathan served as editor for the book *Controllability of Dynamic Systems: The Green's Function Approach*, which received the RA Presidential Award of the Republic of Armenia in the area of “Technical Sciences and Information Technologies.”

Jonathan serves on the board of Homeschool Oklahoma along with his wife, Christa, of 20 years. They inspire their community in several ways including writing educational material, creating educational videos, tutoring students through Classical Conversations, and sharing their own stories of tragedy and success with others.

About the Technical Reviewer

Mike McRoberts is the author of *Beginning Arduino* by Apress. He is the winner of Pi Wars 2018 and a member of Medway Makers. He is an Arduino and Raspberry Pi enthusiast.

Mike McRoberts has expertise in a variety of languages and environments, including C/C++, Arduino, Python, Processing, JS, Node-RED, NodeJS, Lua.

Acknowledgments

I would like to thank, first and foremost, my homeschool co-op community. This book originally started from a series of classes that I taught in our local co-op, and my students were the guinea pigs for this content. I received a lot of encouragement from that class, with both the students and the parents enjoying the material. I want to thank my wife who put up with me always typing on my computer to put this together. I also want to thank the Tulsa Open Source Hardware community (as well as the larger Tulsa WebDevs community), who gave me a lot of encouragement while putting together this book and who also sat through many presentations based on this material.

CHAPTER 1

Introduction

Welcome to the world of electronics! In the modern world, electronic devices are everywhere, but fewer and fewer people seem to understand how they work or how to put them together. At the same time, it has never been easier to do so as an individual. The availability of training, tools, parts, instructions, videos, and tutorials for the home experimenter has grown enormously, and the costs for equipment have dropped to almost nothing.

However, what has been lacking is a good guide to bring students from *wanting* to know how electronic circuits work to actually understanding them and being able to develop their own. For the hobbyist, there are many guides that show you how to do individual projects, but they often fail to provide enough information for their readers to be able to build projects of their own. There is plenty of information on the physics of electricity in physics books, but they fail to make the information practical. One exception to this is Horowitz and Hill's *The Art of Electronics*. This book is a wonderful reference guide for practical circuit design. However, its target audience is largely electrical engineers or other very advanced circuit designers. Not only that, the book itself is prohibitively expensive.

What has been needed for a long time is a book that takes you from knowing nothing about electronics to being able to build real circuits that you design yourself. This book combines theory, practice, projects, and design patterns in order to enable you to build your own circuits from scratch. Additionally, this book is designed entirely around safe,

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_1

low-current DC (direct current) power. We stay far away from the wall outlet in this book to be sure that you have a fun and largely worry-free experience with electronics.

This book is written with two groups of people in mind. First, this book can be used as a guide for hobbyists (or wannabe hobbyists) to learn on their own. It has lots of projects to work on and experiment with. Second, this book can also be used in electronics classes for high school and college students. It has problems to be worked, activities to do, and reviews at the end of each chapter.

The needs of these groups are not so different from each other. In fact, even if you are a hobbyist and plan on using this book to learn on your own, I suggest that not only do you read the main parts of the chapter but that you also do the activities and homework as well. The goal of the homework is to train your mind to think like a circuit designer. If you work through the example problems, it will make analyzing and designing circuits simply a matter of habit.

1.1 Working the Examples

In this book, all examples should be worked out using decimals, not fractions. This is an engineering course, not a math course, so feel free to use a calculator. However, you will often wind up with very long strings of decimals on some of the answers. Feel free to round your answers, but always include at least a single decimal point. So, for instance, if I divide 5 by 3 on my calculator, it tells me 1.66666667. However, I can just give the final answer as 1.7. This only applies to the final answer. You need to maintain your decimals while you do your computations.

Also, if your answer is a decimal number that *begins* with a zero, then you should round your answer to include the first two to four nonzero digits. So, if I have an answer of 0.0000033333333, I can round that to 0.00000333. If you want to be precise about the proper way to round results, see the section on significant figures in the next chapter.

For beginners and hobbyists, this is less of a concern, and we will generally be in a hobbyist mindset for the book.

In short, as engineers, we wind up being, at minimum, as precise as we *need* to be or, at maximum, as precise as we *can* be. The amount of precision we need will vary from project to project, and the amount of precision that we can be will depend on our tools, our components, and other things we interact with. Therefore, there is not a lot of focus on this book on how many decimals exactly to use. You can get more detailed descriptions in other science books for dealing with significant figures. In the problems in the chapters, if you are off by a single digit due to rounding errors, don't worry about it.

1.2 Initial Tools and Supplies

You can get started in electronics with a minimum set of tools, but you can also be as fancy as you have money to afford. This book will focus on the more modest tools that are within the reach of pretty much every budget.

While the book will walk you through a wide variety of parts for different types of circuits, every electronics hobbyist should start out with the following components:

1. Multimeter: Multimeters will measure voltage, current, resistance, and other important values. For these projects, the cheapest digital multimeter you can find will work just fine. You only need one of these.

2. Solderless Breadboards: Solderless breadboards will hold your projects in place and connect your components together. Breadboards are sold based on the number of holes, known as "tie points," the breadboard contains. If you want to keep your projects around, you should have a separate breadboard for each project. However, the beauty of solderless breadboards is that they are in fact reusable if you want.

3. Jumper Wires: Jumper wires are just like normal insulated wires, except that their ends are solid and strong enough to be pushed into your breadboard. The wires themselves may be flexible or rigid. Jumper wires with female ends (a hole instead of a wire) also exist for plugging into circuits which have metal pins sticking out of them (known as headers) to connect to. Every hobbyist I know has a huge mass of jumper wires. They usually come in bundles of 65 wires, which is plenty to get started.

4. Resistors: Resistors do a lot of the grunt work of the circuit. They resist current flow, which, among other things, prevents damaging other parts of the circuit. Resistors are measured in ohms (Ω). Most hobbyists have a wide variety of resistors. You should have a range of resistors from 200 Ω to 1, 000, 000 Ω. However, if you had to pick one value for your resistors, 1, 000 Ω resistors work in a wide variety of situations. Resistors for this book should be rated for 1/4 watt of power.

5. LEDs: LEDs (light-emitting diodes) are low-power lights often used in electronic devices. I recommend getting a variety of colors of LEDs just because it makes life more fun. Most standard single-color LEDs have about the same specifications, so the main difference is the color.

6. Buttons and Switches: Buttons and switches will be the primary method of input and output in these circuits. You should buy buttons and switches *which are specifically made to go on breadboards*.

7. Power Regulator: While most of these projects can be operated directly from a battery, a power regulator board will make sure that, no matter how well charged or drained the battery is, you get a predictable voltage from your battery. The YwRobot breadboard power supply is extremely cheap (cheaper than most batteries) and also provides your project with an on/off switch. You should buy one of these for each breadboard you have. Other breadboard power supplies are available as well (make sure they output 5 volts), but our drawings will assume the YwRobot one.

8. 9 V Battery and Connector: The easiest way to supply power to the power regulator is with a 9 V battery with a standard barrel plug (2.1 mm × 5.5 mm), which will fit into the YwRobot power supply.

Later projects will require specialized components, but these are the components that are needed for nearly every project you will encounter or design yourself. If you would like to order a kit with all of the components you need for this book, you can find them at `www.bplearning.net`.

1.3 Safety Guidelines

This book deals almost entirely with direct current from small battery sources. This current is inherently fairly safe, as small batteries are not capable of delivering the amount of current needed to injure or harm. For these projects, you can freely touch wires and work with active circuits without any protection, because the current is incapable of harming you. The main issue that sometimes arises is that, in poorly made circuits, components can overheat and occasionally (but rarely) catch fire.

Additionally, the battery itself may become overheated/compromised, and batteries are often made from potentially toxic chemicals.

Please follow the following safety guidelines when working on projects (both projects from this book and projects you build yourself). They will help keep you safe and help prevent you from accidentally damaging your own equipment:

1. If you have any cuts or other open areas on your skin, please cover them. Your skin is where most of your electric protection exists in your body.
2. Before applying power to your circuit, check to be sure you have not accidentally wired in a short circuit between your positive and negative poles of your battery.
3. If your circuit does not behave as you expect it to when you plug in the battery, unplug it immediately and check for problems.
4. If your battery or any component becomes warm, disconnect power immediately.
5. If you smell any burning or smoky smells, disconnect power immediately.
6. Dispose of all batteries in accordance with local regulations.
7. For rechargeable batteries, follow the instructions on the battery for proper charging procedures.

Please note that if you ever deal with alternating current (AC) or large batteries (such as a car battery), you must exercise many more precautions than described in this book, because those devices generate sufficient power in themselves and within the circuits to harm or kill you if mishandled (sometimes even after the power has been disconnected).

1.4 Electrostatic Discharge

If you have ever touched a doorknob and received a small shock, you have experienced electrostatic discharge (ESD). ESD is not dangerous to you, but it can be dangerous to your equipment. Even shocks that you can't feel may damage your equipment. With modern components, ESD is rarely a problem, but nonetheless it is important to know how to avoid it. You can skip these precautions if you wish, just know that occasionally you might wind up shorting out a chip or transistor because you weren't careful. ESD is also more problematic if you have carpet floors, as those tend to build up static electricity.

Here are some simple rules you can follow to prevent ESD problems:

1. When storing IC components (i.e., electronics chips), store them with the leads enmeshed in conductive foam. This will prevent any voltage differentials from building up in storage.
2. Wear natural 100% cotton fabrics.
3. Use a specialized ESD floor mat and/or wrist strap to keep you and your workspace at ground potential.
4. If you don't use an ESD strap or mat, touch a large metal object before starting work. Do so again any time after moving around.

1.5 Using Your Multimeter Correctly

Even though we haven't covered the details we need to use our multimeter yet, since we are covering proper handling of devices, I am including this section here with the others. Feel free to skip over this until we start using multimeters in the book.

In order to keep your multimeter functioning, it is important to take some basic precautions. Multimeters, especially cheap ones, can be easily broken through mishandling. Use the following steps to keep you from damaging your multimeter or damaging your circuit with your multimeter:

1. Do not try to measure resistance on an active circuit. Take the resistor all the way out of the circuit before trying to measure it.
2. Choose the appropriate setting on your multimeter *before* you hook it up.
3. Always err on the side of choosing high values first, especially for current and voltage. Use the high value settings for current and voltage to give your multimeter the maximum protection. If you set the value too large, it is easy enough to set it lower. If you had it set too low, you might have to buy a new multimeter!

CHAPTER 2

Dealing with Units

Before we begin our exploration of electronics, we need to talk about **units of measurement**. A unit of measurement is basically a standard against which we are measuring something. For instance, when measuring the length of something, the units of measurement we usually use are feet and meters. You can also measure length in inches, yards, centimeters, kilometers, miles, and so on. Additionally, there are some obscure units of length like furlongs, cubits, leagues, and paces.

Every type of quantity has its own types of units. For instance, we measure time in seconds, minutes, hours, days, weeks, and years. We measure speed in miles per hour, kilometers per hour, meters per second, and so on. We measure mass in pounds, ounces, grams, kilograms, grains, and so on. We measure temperature in Fahrenheit, Celsius, Kelvin, and Rankine.

Units for the same type of quantity can all be converted into each other using the proper formula.

2.1 SI Units

The scientific community has largely agreed upon a single standard of units known as the **International System of Units**, abbreviated as **SI Units**. This is the modern form of the metric system. Because of the large number of unit systems available, the goal of creating the SI standard was

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_2

to create a single set of units that had a basis in physics and had a standard way of expressing larger and smaller quantities.

The imperial system of volumes illustrates the problem they were trying to solve. In the imperial system, there were gallons. If you divided a gallon into four parts, you would get quarts. If you divided quarts in half, you'd get pints. If you divided a pint into twentieths, you'd get ounces.

The imperial system was very confusing. Not only were there an enormous number of units but they all were divisible by differing amounts. The case was similar for length—12 inches in a foot, but 3 feet in a yard and 1,760 yards in a mile. This was a lot to memorize, and doing the calculations was not easy.

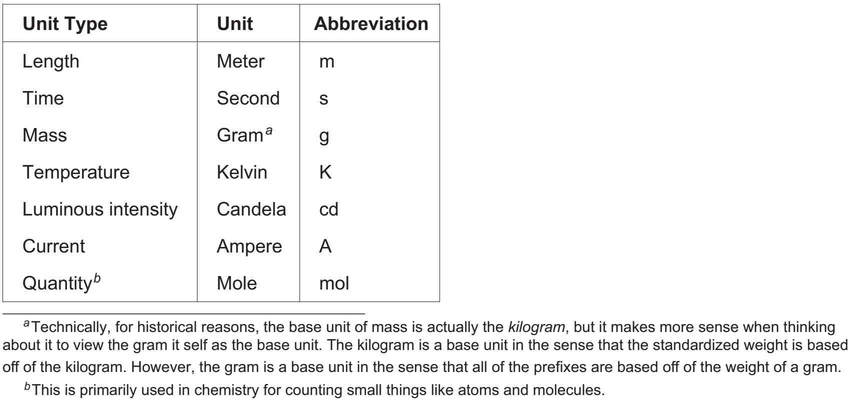

Unit Type	**Unit**	**Abbreviation**
Length	Meter	m
Time	Second	s
Mass	Gram[a]	g
Temperature	Kelvin	K
Luminous intensity	Candela	cd
Current	Ampere	A
Quantity[b]	Mole	mol

[a] Technically, for historical reasons, the base unit of mass is actually the *kilogram*, but it makes more sense when thinking about it to view the gram it self as the base unit. The kilogram is a base unit in the sense that the standardized weight is based off of the kilogram. However, the gram is a base unit in the sense that all of the prefixes are based off of the weight of a gram.

[b] This is primarily used in chemistry for counting small things like atoms and molecules.

Figure 2-1. *SI Base Units*

The imperial system does have some benefits (the quantities used in the imperial system match the sizes normally used in human activities—few people order drinks in milliliters), but for doing work which requires a lot of calculations and units, the SI system has largely won out. Scientific quantities are almost always expressed in SI units. In engineering it is more of a mix, just as engineering itself is a mix between scientific inquiry and human usefulness. However, the more technical fields have generally moved to SI units and stayed with them.

There are only seven base units in the SI system. Other units are available as well, but they all can be measured in terms of these base units. The base units for the SI system are shown in Figure 2-1.

Many other units are derived from these and are known as SI derived units. For instance, for measuring volume, liters are often used.[1] A liter, however, is not defined on its own, but in terms of meters. A liter is a thousandth of a cubic meter. Thus, we can take the unit of length and use it to describe a unit of volume.

A more complicated example is the newton, which is a unit of force. In the SI system, the newton is defined as being a “kilogram-meter per second squared.” This is another way of saying that a newton is the amount of force which accelerates 1 kilogram 1 meter per second, per second.

All of the things of interest to us in this book are ultimately defined in terms of SI base units. For the purposes of this book, it is not important to know which units are base units or derived units, and it is especially unnecessary to know how they are derived. The important thing to keep in mind is that you will be using a well-thought-out, standardized system of units. If the units seem to fit together well, it’s because they were designed to do so.

---

[1] Technically, the liter is not defined by the SI, but it is listed here because it is the simplest illustration of the point.

Conversion Factor	Prefix	Abbreviation	Examples
1,000,000,000,000	Tera	T	Terameter, terasecond, teragram
1,000,000,000	Giga	G	Gigameter, gigasecond, gigagram
1,000,000	Mega	M	Megameter, megasecond, megagram
1,000	Kilo	K	Kilometer, kilosecond, kilogram
1			Meter, second, gram
0.001	Milli	m	Millimeter, millisecond, milligram
0.000001	Micro	μ or u	Micrometer, microsecond, microgram
0.000000001	Nano	n	Nanometer, nanosecond, nanogram
0.000000000001	Pico	p	Picometer, picosecond, picogram

Figure 2-2. *Common SI Prefixes*

2.2 Scaling Units

Now, sometimes you are measuring really big quantities, and sometimes you are measuring very small quantities. In the imperial system, there are different units altogether to reach a different scale of a quantity. For instance, there are inches for small distances, yards for medium-sized distances, and miles for large distances. There are ounces for small volumes and gallons for larger volumes.

In the SI system, however, there is a uniform standard way of expressing larger and smaller quantities. There are a set of modifiers, known as **unit prefixes**, which can be added to *any unit* to work at a different scale. For example, the prefix *kilo-* means thousand. So, while a meter is a unit of length, a kilometer is a unit of length that is 1,000 times as large as a meter. While a gram is a unit of mass, a kilogram is a unit of mass that is 1,000 times the mass of a gram.

It works the other way as well. The prefix *milli-* means thousandth, as in $\frac{1}{1000}$. So, while a meter is a unit of length, a millimeter is a unit of length that is $\frac{1}{1000}$ of a meter. While a gram is a unit of mass, a milligram is a unit of mass that is $\frac{1}{1000}$ the mass of a gram.

Therefore, by memorizing one single set of prefixes, you can know how to modify all of the units in the SI system. The common prefixes occur at every power of 1,000, as you can see in Figure 2-2.

To convert between a prefixed unit (i.e., kilometer) and a base unit (i.e., meter), we just apply the conversion factor. So, if something weighs 24.32 kilograms, then I could convert that into grams by multiplying by 1,000. $24.32 * 1000 = 24{,}320$. In other words, 24.32 kilograms is the same as 24,320 grams.

To move from the base unit to a prefixed unit, you *divide* by the conversion factor. So, if something weighs 35.2 grams, then I could convert that into kilograms by dividing it by 1,000. $35.2/1000 = 0.0352$. In other words, 35.2 grams is the same as 0.0352 kilogram.

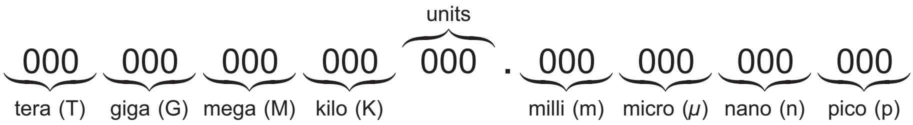

Figure 2-3. *Visualizing Common Unit Prefixes*

You can also convert between two prefixed units. You simply multiply by the starting prefix and divide by the target prefix. So, if something weighs 220 kilograms and I want to know how many micrograms that is, then I will multiply using the kilo- prefix (1,000) and divide by the micro- prefix (0.000001):

$$\frac{220 \cdot 1000}{0.000001} = 220000000000$$

In other words, 220 kilograms is the same as 220,000,000,000 micrograms.

You can do all of the unit scaling that you need just by knowing the multipliers. However, what usually helps me deal with these multipliers intuitively is to simply visualize where each one lands in a single number. Figure 2-3 shows all of the prefixes laid out in a single number.

So let's say that I was dealing with fractions of a meter and I had something that was 0.000000030 meter. If you line this number up with the chart in Figure 2-3, there are only zeros in the unit, the milli-, and the micro- areas. The first nonzero digits appear in the "nano-" group. When lined up with the chart, the number in the nano- area is 030. Therefore, the number under consideration is 30 nanometers.

2.3 Using Abbreviations

Typing or writing words like kilogram, microsecond, and micrometer isn't terribly difficult, but, when it occurs a lot (as what can happen in equations), it can get overwhelming. Therefore, every prefix and every unit has an abbreviation. Since the abbreviation for gram is g, and the abbreviation for kilo- is k, we can abbreviate kilogram as kg. Occasionally the abbreviation for the unit and the scaling prefix are identical, as in the case of meter (m) and milli- (also m). That's fine, as, when you put them together, you get millimeter, which is abbreviated as mm.

The hardest one to write is the one for micro-, μ. This is the Greek letter mu (pronounced "mew"). It's essentially the Greek way of writing the letter m, and it is used because we already have a lowercase m (milli-) and an uppercase M (mega-) in use. Since micro- begins with an m, lacking any additional English/Latin way of writing an m, it is written with the Greek μ. However, this is sometimes hard to type. Therefore, since the squashed way that it is written makes it look kind of like a u, sometimes people will write u instead of μ if they don't know how to type out μ with their computer/keyboard. In this book, we never use u for this purpose, but, if you are reading elsewhere something like 100 us, that means 100 microseconds.

2.4 Significant Figures

Significant figures are the bane of many science books. Nearly everything in science has to be rounded, and significant figures are basically the rounding rules for science. We need to talk about them simply so that you are aware of how I achieve the rounding that I do in my exercises.

These rules aren't hard, but they can cause some newer students to stumble. If you just are wanting to play with electronics, you can skip this section; just be aware that I may have rounded the answers to problems differently than you do.

The goal of significant figures is to prevent us from thinking that we are being more precise than we really are. Let's say that I measured a piece of wood to be 1 meter long, but I wanted to cut that wood into thirds. How should I report the distance, in decimal, of the length of each desired piece? Well, 1 divided by 3 is 0.33333333333… I can keep writing three's until the cows come home. But do I really need the length to be that precise? Is my measurement of the initial length of the wood precise enough to warrant that sort of precise request? Significant figures allow us to answer that question and report numbers with a justifiable precision.

So, for any measured quantity, we need to count the number of significant figures. For the most part, the number of significant figures in a number is the same as the number of digits, with a few exceptions. First, significant figures ignore all leading zeros. So, if I measured something as being 102 feet, it has three signficant figures, even if I add leading zeros. So 102 feet and 0000000000102 feet both have three significant figures. Additionally, these leading zeros are still ignored even if they are after the decimal point. So the number 0.00042 has two significant figures.

The second rule in counting is that trailing zeros aren't counted as significant if the measuring device isn't capable of measuring that accurately (or the quantity isn't reported that accurately). For instance, the value 1 meter and the value 1.000 meter refer numerically to the same number. However, the second one is usually used to indicate that we can actually measure that precisely. We wouldn't report 1.000 meter unless our scale can actually report accurately to a thousandth of a meter.

The situation is a little more complicated with zeros on the left side of the decimal. If I say, "1,000 people attended the event," how precise is that number? Did you count individuals and get exactly 1,000? Is it possibly estimated to the tens or hundreds place? This gets murky. To simplify the issue for this book, you can assume that all digits on the right count as significant figures. So, if we say "1,000 people" attended the event, that is a number with four significant digits. However, if we say "1 kilopeople" attended the event, that is a number with one significant digit. If we say " 1.03 kilopeople" attended the event, that is a number with three significant digits. We will use this convention when writing down problems, but results may just be a rounded number with trailing zeros (i.e., instead of writing "1.03 kilopeople," we might write the result as 1,030 people).

One other small rule: If a number is exact, then it is essentially considered to have infinite significant digits. So, for instance, it takes two people to have a baby. This is an exact number. It isn't 2.01 or 2.00003 people, it is exactly two. So, for the purposes of significant figures, this value has an infinite number of digits. Conversion factors are generally considered to be exact values.

So that is how to count significant digits. This is important because the significant digits affect how caclulations are rounded. There are two rules—one for multiplying and dividing and another for adding and subtracting.

For multiplying and dividing, you should find the input value with the fewest significant figures. The result should then be rounded to that many significant figures. For instance, if we had $103 * 55$, then the result should be rounded to two significant figures. So, even though the result is 5,665, we should report it as 5,700. Let's say we have 55.0 ÷ 3.00. Since both of the input values have three significant figures, then the result should have three significant figures. Therefore, the result is 18.3.

For addition and subtraction, instead of using significant figures, the decimal points for the numbers are lined up, the operation is performed, and the result is rounded to the number of decimal places available in the input value that has the least precision (fewest numbers to the right of the decimal). So, for instance, if I have 1.054 + 0.06, the result is 1.104. However, this would be rounded to 1.10 because that is how many digits to the right the decimal 0.06 had.

If there are a series of operations, significant figures are usually applied at the end of the whole calculation, or when necessary to limit the complication of intermediate results. Calculators will round for you at some point anyway, so there is no getting around some amount of intermediate rounding. Therefore, you should recognize that if your answer differs from the book's answer by the least significant digit, it is likely that you are correct, but that you rounded in different stages in your calculation.

In professional science and engineering data reporting, significant figures are important. In playing around with electronics, they are much less so. Additionally, even the rules for significant figures aren't perfect—there are places where their usage leads to problematic results. Entire books have been written on the subject.[2] Significant figures are there not because they are perfect, but so that we all have a common, straightforward way of communicating the precision of our results. The most important thing to keep in mind is that the degree to which you are precise in your measurements affects the degree to which you can be precise in your calculations.

---

[2] For instance, see Nicholas Higham's *Accuracy and Stability of Numerical Algorithms*.

Apply What You Have Learned

1. How many nanometers is 23 meters?
2. How many seconds is 23.7 microseconds?
3. How many grams is 89.43 megagrams?
4. How many meters is 15 nanometers?
5. How many kilograms is 0.3 microgram?
6. How many milliseconds is 45 kiloseconds?
7. What is the abbreviation for picosecond? What is the abbreviation for microgram?
8. What is the abbreviation for a terameter?
9. How many significant figures does the number 476 have?
10. How many significant figures does the number 5 have?
11. How many significant figures does the number 000352 have?
12. How many significant figures does the number 0.00043 have?
13. How many significant figures does the number 1.0004 have?
14. How many significant figures does the number 2.34000 have?

Calculate the following problems taking into account significant figures.

15. What is 23 ∗ 5?
16. What is 23 + 0.6?
17. What is 0.005 ∗ 209?
18. What is 0.0023 ∗ 45?
19. What is 0.5 + 0.5?
20. Why are significant figures important?

PART I

Basic Concepts

CHAPTER 3

What Is Electricity?

The first thing to tackle in the road to understanding electronics is to wrap our minds around what electricity is and how it works. The way that electricity works is very peculiar and unintuitive. We are used to dealing with the world in terms of physical objects—desks, chairs, baseballs, and so on. Even if we never took a class in physics, we know the basic properties of such objects from everyday experience. If I drop a rock on my foot, it will hurt. If I drop a heavier rock, it will hurt more. If I remove an important wall from a house, it will fall down.

However, for electricity, the only real experience we have is that we have been told to stay away from it. Sure, we have experience with computers and phones and all sorts of devices, but they give us the result of processing electricity a million times over. But how does electricity itself work?

3.1 Charge

To answer this question, we need to answer another question first: What *is* electricity? Electricity is the flow of **charge**. So what is charge?

Charge is a fundamental quantity in physics—it is not a combination (that we know of) of any other quantity. A particle can be charged in one of three ways—it can be positively charged (represented by a + sign), negatively charged (represented by a – sign), or neutrally charged (i.e., has no charge). Figure 3-1 shows what an atom looks like. In the center of the atom are larger, heavier particles called **protons** and **neutrons**. Protons are positively charged particles, and neutrons are neutrally charged particles.

© Jonathan Bartlett 2020

J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_3

Together, these form the **nucleus** of the atom and determine *which* atom we are talking about. If you look on a periodic table, the large printed number associated with an element is known as its **atomic number**. This number refers to how many protons it has in its nucleus. Sometimes, there is a number in smaller print as well. This is the total number of protons and neutrons combined.[1] Note that the number of neutrons in an element can vary, so this number is often a decimal representing the average number of combined protons and neutrons in any particular element.

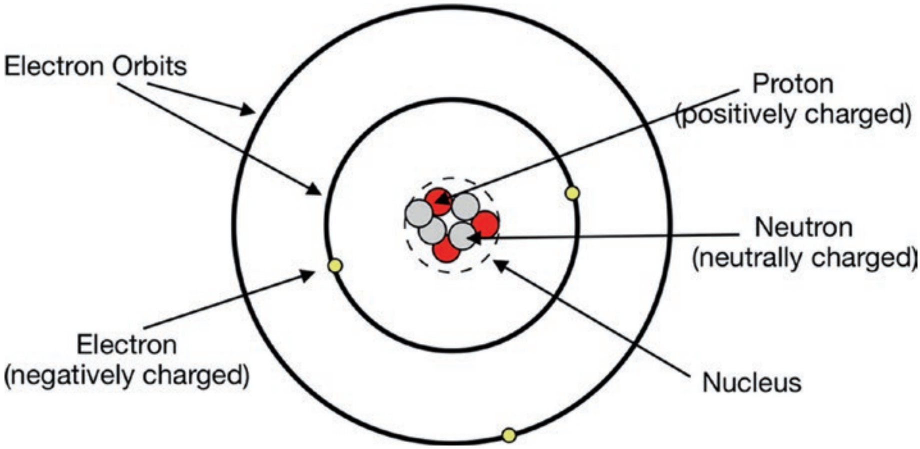

Figure 3-1. *Charged Particles in an Atom*

Circling around the nucleus are **electrons**. Electrons are negatively charged particles. Even though electrons are much smaller and lighter than protons, the amount of negative charge of one electron is equal to the amount of positive charge of one proton. Positive and negative charges attract each other, which is what keeps electrons contained within the atom.

---

[1] This is technically the *atomic mass* of the element, given in special units called atomic mass units (AMUs). However, the weights of protons and neutrons are each extremely close to 1 AMU, and the weight of electrons is extremely close to zero, so the atomic mass and the combined number of protons and neutrons are nearly identical.

Electrons are arranged in shells surrounding the nucleus. The outermost shell, however, is the most important one when thinking about how atoms work.

When we think about individual atoms, we think about them when they are isolated and alone. In these situations, the number of electrons and the number of protons are equal, making the atom as a whole electrically neutral. However, especially when atoms interact with other atoms, the configuration of their electrons can change. If the atoms gain electrons, then they are negatively charged. If the atoms lose electrons, then they are positively charged. Free electrons are all negatively charged.

If there are both positively and negatively charged particles moving around, their opposite charges attract one another. If there is a great imbalance of positive and negative charges, usually you will have a *movement* of some of the charged particles toward the particles of the opposite charge. This is a *flow* of charge, and this flow is what is referred to when we speak of electricity.

The movement of charge can be either positively charged particles moving toward negatively charged ones, negatively charged particles moving toward positively charged ones, or both. Usually, in electronics, it is the electrons which are moving through a wire, but just know that this is not the only way in which charge can move.

Electricity can be generated by a variety of means. The way that electricity is generated in a battery is that a chemical reaction takes place, but the reactants (the substances that react together) are separated from each other by some sort of medium. The positive charges for the reaction move easiest through the medium, but the negative charges for the reaction move easiest through the wire. Therefore, when the wire is connected, electricity moves through the wire to help the chemical reaction complete on the other side of the battery.

This flow of electric charge through the wire is what we normally think of as electricity.

MAKING YOUR OWN BATTERY

You can make a simple battery of your own out of three materials: thick copper wire or tubing, a galvanized nail (it *must* be galvanized), and a potato or a lemon. This battery operates from a reaction between the copper on the wire and the zinc on the outside of the galvanized nail. The electrons will flow from the zinc to the copper through the wire, while the positive charge will flow within the lemon or potato.

To build the battery, you must insert the thick copper and the nail into the potato. They should be near each other, but *not touching*. This battery will not produce very much current—less than a milliamp of current at less than 1 volt (we will discuss milliamps later in this chapter and volts in Chapter 4). This is not quite enough to light up an LED, but it should register on a multimeter. See Chapter 6 for how to measure voltage with a multimeter.

Note that the lemon/potato is not actually supplying the current. What the lemon/potato is doing is creating a barrier so that only the positive charges can flow freely in the potato, and the negative charges have to use the wire.

3.2 Measuring Charge and Current

Atoms are very, very tiny. Only in the last few years have scientists even developed microscopes that can see atoms directly. Electrons are even tinier. Additionally, it takes a *lot* of electrons moving to have a worthwhile flow of charge. Individual electrons do not do much on their own—it is only when there are a very large number of them moving that they can power our electronics projects.

Therefore, scientists and engineers usually measure charge on a much larger scale. The **coulomb** is the standard measure of electric charge. One coulomb is equivalent to the electric charge of about

6,242,000,000,000,000,000 protons. If you have that many electrons, you would have a charge of −1 coulomb. That's a lot of particles, and it takes particles on that scale to do very much electrical work. Thankfully, protons and electrons are very, very small. A typical 9-volt battery can provide about 2,000 coulombs of charge, which is over 10,000,000,000,000,000,000,000 charged particles (ten thousand billion billion particles).

However, electricity and electronics are not about electric charge sitting around doing nothing. Electricity deals with the *flow* of charge. Therefore, when dealing with electricity, we rarely deal with coulombs. Instead, we talk about how fast the electric charge is flowing. For that, we use **amperes**, often called amps and abbreviated as A. One ampere is equal to the movement of 1 coulomb of charge out of the battery each second.

For the type of electronics we will be doing, an ampere is actually a lot of current. In fact, a full ampere of current can do a lot of physical harm to you, but we don't usually deal with full amperes when creating electronic devices. Power-hungry devices like lamps, washers, dryers, printers, stereos, and battery chargers need a lot of current—that's why we plug them into the wall. Small electronic devices don't usually need so much current. Therefore, for electronic devices, we usually measure current in **milliamperes**, usually called just milliamps and abbreviated as mA. Remember that the prefix *milli-* means one-thousandth of (i.e., $\frac{1}{1000}$ or 0.001). Therefore, a milliamp is one-thousandth of an amp. If someone says that there is 20 milliamps of current, that means that there is 0.020 amp of current. This is important, because the equations that we use for electricity are based on amps, but since we are dealing with low-current devices, most of our measurements will be in milliamps.

So, to go from amps to milliamps, multiply the value by 1,000. To go from milliamps to amps, divide the value by 1,000 (or multiply by 0.001) and give the answer in decimal (electronics always uses decimals instead of fractions).

Example 3.1 If I were to have 2.3 amps of electricity, how many milliamps is that? To go from amps to milliamps, we multiply by 1,000. 2.3 ∗ 1,000 = 2,300. Therefore, 2.3 amps is the same as 2,300 milliamps.

Example 3.2 If I were to have 5.7 milliamps of electricity, how many amps is that? To go from milliamps to amps, we divide by 1,000. 5.7/1,000 = 0.0057. Therefore, 5.7 milliamps is the same as 0.0057 amp.

Example 3.3 Now, let's try something harder—if I say that I am using 37 milliamps of current, how many coulombs of charge has moved after 1 minute? Well, first, let's convert from milliamps to amps. To convert from milliamps to amps, we divide by 1,000. 37/1000 = 0.037. Therefore, we have 0.037 amp. What is an amp? An amp is 1 coulomb of charge moving per second. Therefore, we can restate our answer as being 0.037 coulomb of charge moving each second.

However, our question asked about how much has moved after 1 *minute*. Since there are 60 seconds in each minute, we can multiply 0.037 by 60 for our answer. 0.037 ∗ 60 = 2.22. So, after 1 minute, 37 milliamps of current moves 2.22 coulombs of charge.

3.3 AC vs. DC

You may have heard the terms AC and DC when people talk about electricity. What do those terms mean? In short, DC stands for **direct current**, and AC stands for **alternating current**. So far, our descriptions of electricity have dealt mostly with DC. With DC, electricity makes a route from the positive terminal to the negative. It is the way most people envision electricity. It is “direct.”

However, DC, while great for electronics projects, very quickly loses power over long distances. If we were to transmit current that simply flows from the positive to the negative throughout the city, we would have to have power stations every mile or so.

So, instead of sending current in through one terminal and out through another, your home is powered with **alternating current**. In alternating current, the positive and negative sides continually reverse, switching back and forth 50–60 times per second. So the current direction (and thus the direction the electrons are moving) continually switches back and forth, over and over again. Instead of moving in a continuous flow, it is more like someone is pushing and pulling the current back and forth. In fact, at the generator station, that is almost exactly what is going on! This may seem strange, but this push and pull action allows for much easier power generation and also allows much more power to be delivered over much longer distances.

AC, such as the current that comes out of a wall socket, is much more powerful than we require for our projects here. In fact, converting high-power AC to low-power DC voltage used in electronic devices is an art in itself. This is why companies charge so much money for battery chargers—it takes a lot of work to get one right!

Now, not all AC is like this. We call this current AC “mains,” because it comes from the power mains from the power stations. It is supposed to operate at about 120 volts, and the circuits are usually rated for about 15–30 amps (that’s 15,000–30,000 milliamps). That’s a lot of electricity!

In addition to AC mains, there are also ACs which we will call AC “signal.” These currents come from devices like microphones. They are AC because the direction of current does in fact alternate. When you speak, your voice vibrates the air back and forth. A microphone converts these air vibrations into small vibrations of electricity—pushing and pulling a small electric current back and forth. However, these ACs are so low powered as to be almost undetectable. They are so small we have to actually amplify these currents (see Chapter 25) just to work with them using our DC power!

So, in short, while we will do some work with AC signal voltages later in the book, all of our projects will be safe, low-power projects. We will often touch wires with our projects active or use multimeters to measure currents and voltages in active circuits. This is perfectly safe for battery-operated projects. But *do not* attempt these same maneuvers for anything connected to your wall outlet unless you are properly trained.

3.4 Which Way Does Current Flow?

One issue that really bungles people up when they start working with electronics is figuring out which way electric current flows. You hear first that electric current is the movement of electrons, and then you hear that electrons move from negative to positive. So one would naturally assume that current flows from negative to positive, right?

Good guess, but no—or, at least, not quite.

Current is not the flow of physical stuff like electrons, but the flow of *charge*. So, when the chemical reaction happens in the battery, the positive side gets positively charged, and the influence of that charge moves down the wires. The electrons are a negative charge that moves toward that positive charge. The positive charge is just as real as the negative charge, even though physical stuff isn't moving with the positive charge.

Think about it this way. Have you ever used a vacuum cleaner? Let's say we are tracing the action of a vacuum cleaner. Where do you start? Usually, you start at the inside where the suction happens and then trace the flow of suction through the tube. Then, at the end of the tube, the dust comes into the tube.

Engineers don't trace their systems from the dust to the inside, they trace their systems from the suction on the inside out to the dust particles on the outside. Even though it is the dust that moves, it is the suction that is interesting.

Likewise, for electricity, we usually trace current from positive to negative even though the electrons are moving the other way. The positive charge is like the suction of a vacuum, pulling the electrons in. Therefore, we want to trace the flow of the vacuum from positive to negative, even though the dust is moving the other way.

The idea that we trace current from positive to negative is often called **conventional current flow**. It is called that way because we conventionally think about circuits as going from the positive to the negative, and it is the common convention to draw them that way (any arrow in an electronics diagram is pointing toward the movement of *positive* charge).

If you are tracing charge the other way, that is called **electron current flow**, but it is rarely used.

Review

In this chapter, we learned the following:

1. Electric current is the flow of charge.
2. Charge is measured in coulombs.
3. Electric current flow is measured in coulombs per second, called amperes or amps.
4. A milliampere is one-thousandth of an ampere.
5. In an atom, protons are positively charged, electrons are negatively charged, and neutrons are neutrally charged.

6. Batteries work by having a chemical reaction which causes electricity to flow through wires.
7. In DC, electricity flows continuously from positive to negative.
8. In AC, electricity flows back and forth, changing flow direction many times every second.
9. Even though electrons flow from negative to positive, in electronics, we usually think about circuits and draw circuit charges as flowing from positive to negative.
10. AC mains (the kind in your wall outlet) is dangerous, but battery current is relatively safe.
11. Small signal AC (like that generated by a microphone) is not dangerous.

Apply What You Have Learned

1. If I have 56 milliamps of current flowing, how many amps of current do I have flowing?
2. If I have 1,450 milliamps of current flowing, how many amps of current do I have flowing?
3. If I have 12 amps of current flowing, how many milliamps of current do I have flowing?
4. If I have 0.013 amp of current flowing, how many milliamps of current do I have flowing?
5. If I have 125 milliamps of current flowing for 1 hour, how many coulombs of charge have I used up?

6. What is the difference between AC and DC?
7. In AC mains, how often does the direction of current go back and forth?
8. Why is AC used instead of DC to deliver electricity within a city?
9. In working with electronic devices, do we normally work in amps or milliamps?

CHAPTER 4

Voltage and Resistance

In Chapter 3, we learned about current, which is the rate of flow of charge. In this chapter, we are going to learn about two other fundamental electrical quantities—**voltage** and **resistance**. These two quantities are the ones that are usually the most critical to building effective circuits.

Current is important because limiting current allows us to preserve battery life and protect precision components. Voltage, however, is usually the quantity that has to be present to do any work within a circuit.

4.1 Picturing Voltage

What is voltage? Voltage is the amount of power each coulomb of electricity can deliver. If person A has 1 coulomb of charge at 5 volts and person B has 1 coulomb of charge at 10 volts, that means that person B's coulomb can deliver twice as much power as person A's.

A good analogy to electronics is the flow of water. When comparing water to electricity, *coulombs* are a similar unit to *liters*—coulombs measure the amount of electric charge present just like a liter is the amount of water volume present. Both charge and water move as a flow. In water, we can measure the flow of a current of a stream in liters per second. Likewise, in electronics, we measure the flow of charge through a wire in coulombs per second, which are also called amperes.

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_4

Now, I want you to image the end of a hose through which water is flowing. Normally, the water just falls out of the hose, especially if the hose is just sitting on the ground. That hose just sitting on the ground is like a current with 0 volt—each unit of water or charge is just not doing that much.

Let's pretend we added a spray nozzle to the hose. What happens now? Water shoots out of the nozzle forcefully. We haven't added any more water—it is actually the same amount of water (i.e., current) flowing. Instead, we increased the pressure of the water, which is just like increasing the voltage on an electric charge. By increasing the pressure, we changed the amount of work that each liter of water is available to perform.

Likewise, when we increase voltage, we change the amount of work that each coulomb of electricity can do.

One way we might measure the pressure of water coming out of a hose is to measure how far up it can shoot out of the hose. By doubling the pressure of the water, we can double how far out of the hose it can shoot. Similarly, with voltages, large enough voltages can actually jump air gaps across circuits. However, to do this, it takes a lot of voltage—about 30,000 volts per inch of gap. If you have been shocked by static electricity, though, this is what is happening! The power of the charge is extreme (thousands of volts), but the amount of charge in those shocks is so small that it doesn't harm you (about 0.00000001 coulomb).

4.2 Volts Are Relative

While charge and current are fairly concrete ideas, voltage is a much more relative idea. You can actually never measure voltage absolutely. All voltage measurements are actually relative to other voltages. That is, I can't actually say that my electric charge has exactly 1, 2, 3, or whatever volts. Instead, what I have to do is say that one charge is however many volts more or less than another charge. So let's take a 9-volt battery. What that means is not that the battery is 9 volts in any absolute sense, but rather that there is a 9 -volt *difference* between the charge at the positive terminal and the charge at the negative terminal. That is, the pressure with which charge is trying to move from the positive terminal to the negative terminal is 9 volts.

4.3 Relative Voltages and Ground Potential

When we get to actually measuring voltages on a circuit, we will only be measuring voltage *differences* on the circuit. So, to measure voltage, I can't just put a probe on one place on the circuit. Instead, I have to put my probe on two *different* places on the circuit and measure the voltage difference (also called the **voltage drop**) between those two points.

However, to simplify calculations and discussions, we usually choose some point on the circuit to represent "0 volt." This gives us a way to standardize voltage measurements on a circuit, since they are all given relative to the same point. In theory, this could be any point on the circuit, but, usually, we choose the negative terminal on the battery to represent 0 volt.

This "zero point" goes by several names, the most popular of which is **ground** (often abbreviated as **GND**). It is called the ground because, historically, the physical ground has often been used as a reference voltage for circuits. Using the physical ground as the zero point allows you to also compare voltages between circuits with different power supplies. However, in our circuits, when we refer to the ground, we are referring to the negative terminal on the battery, which we are designating as 0 volt.

Another, lesser-used term for this designated 0-volt reference is the **common** point. Many multimeters label one of their electrodes as **COM**, for the common electrode. When analyzing a circuit's voltage, this electrode would be connected to whatever your 0-volt point is.

This “ground” analogy also makes sense with our water hose analogy. Remember that a voltage is the potential for a charge to do work. What happens to water after it lands on the ground? By the time the water from my hose lands on the ground, it has lost all its energy. It is just sitting there. Sure, it may seep or flow around a bit, but nothing of consequence. All of its ability to do work—to move quickly or to knock something over—has been drained. It is just on the ground. Likewise, when our electric charge is all puttered out, we say that it has reached “ground potential.”

So, even though we could designate any point as being zero, we usually designate the negative terminal of the battery as the zero point, indicating that by the time electricity reaches that point, it has used up all of its potential energy—it now has 0 volt compared to the end destination (i.e., the other battery terminal).

4.4 Resistance

Resistance is how much a circuit or device resists the flow of current. Resistance is measured in **ohms** and is usually represented by the symbol Ω. Going back to our water hose analogy, **resistance** is how small the hose is, because a smaller hose will resist the flow of water more than a larger hose will. Think about a 2-liter bottle of pop. The bottle has a wide base, but the opening is small. If I turn the bottle upside down, the small opening limits the amount of liquid that flows out at one time. That small opening is giving *resistance* to the flow of liquid, making it flow more slowly. If you cut off the small opening, leaving a large opening, the liquid will come out much faster because there is less resistance.

Ohm’s law, which we will use throughout this book, tells us about the relationship between resistance, voltage, and current flow. The equation is very simple. It says

$$V = I * R \tag{4.1}$$

In this equation, V stands for voltage, I stands for current (in *amperes*, not milliamperes), and R stands for resistance (in ohms). To understand what this equation means, let's think again about water hoses. The water that comes out of the faucet of your house has essentially a constant current. Therefore, according to the equation, if we add resistance, it will increase our voltage.

We know this to be true from experience. If we have a hose and just point it forward, water usually comes out about a foot or two. Remember, voltage is how much push the water has, which determines how far the water will go when it leaves the hose. However, if my children are on the other side of the yard and I want to hit them with a water spray, what do I do? I put my thumb over the opening. This increases the resistance, and, since the current is relatively constant, the voltage (the force the water will have when it leaves the hose) will increase, and this force will cause the water to spray further.

However, in circuits, we usually don't have a constant current source. Instead, batteries provide a constant voltage source. A 9-volt battery will provide 9 volts in nearly every condition. Therefore, for electronics work, we usually rearrange the equation a little bit. Using a little bit of algebra, we can solve our equation for either current or resistance, like this:

$$I = \frac{V}{R} \tag{4.2}$$

$$R = \frac{V}{I} \tag{4.3}$$

Equation 4.2 is the one that is usually most useful. To understand this equation, think back to the example of the bottle turned upside down. There, the liquid had a constant amount of push/voltage (from gravity), but we had different resistances. With the small opening, we had a large resistance, so the liquid came out slower. With the large opening, we had almost no resistance, so the liquid came out all at once.

> **Example 4.4** Let's put Ohm's law to use. If I have a 5-volt voltage source with 10 ohms of resistance in the circuit, how much current will flow? Since we are solving for current, we should use Equation 4.2. This says $I = \frac{V}{R}$. Therefore, plugging in our voltage and resistance, we have $I = \frac{5}{10}$, which is $I = 0.5$ ampere (remember, Ohm's law always uses amperes for current).

> **Example 4.5** Now let's say that we have a 10-volt source and we want to have 2 amps worth of current flowing. How much resistance do we need in order to make this happen? Since we are now solving for resistance, we will use Equation 4.3, which says $R = \frac{V}{I}$. Plugging in our values, we see that $R = \frac{10}{2} = 5\ \Omega$. Therefore, we would need 5 Ω of resistance.

> **Example 4.6** Now let's say that I have a 9-volt source and I want to limit my current to 10 *milliamps*. This uses the same equation, but the problem I have is that my units are in milliamps, but my equation uses amps. Therefore, before using the equation, I have to convert my current from milliamps to amps. Remember, to convert milliamps to amps, we just divide by 1,000. Therefore, we take 10 milliamps and divide by 1,000, and we get 0.010 amp. Now we

can use Equation 4.3 to find the resistance we need. $R = \frac{V}{I} = \frac{9}{0.010} = 900\Omega$. Therefore, with 900 Ω of resistance, we will limit our current to 10 milliamps.

Review

In this chapter, we learned the following:

1. Voltage is the amount of power that each coulomb of charge delivers.
2. The volt is the electrical unit that we use to measure voltage.
3. Voltage is always given relative to other voltages—it is not an absolute value.
4. The ground of a circuit is a location on the circuit where we have chosen to use as a universal reference point—we define that point as having zero voltage for our circuit to make measuring other points on our circuit easier.
5. In DC electronics, the chosen ground is usually the negative terminal of the battery.
6. Other terms and abbreviations for the ground include common, GND, and COM.
7. Resistance is how much a circuit resists the flow of current and is measured in ohms (Ω).
8. Ohm's law tells us the relationship between voltage, current, and resistance: $V = I * R$.

9. Using basic algebra, we can rearrange Ohm's law in two other ways, depending on what we want to know. It can be solved for current, $I = \frac{V}{R}$, or it can be solved for resistance, $R = \frac{V}{I}$.

Apply What You Have Learned

1. If I have a 4-volt battery, how many volts are between the positive and negative terminals of this battery?
2. If I choose the *negative* terminal of this battery as my ground, how many volts are at the *negative* terminal?
3. If I choose the *negative* terminal of this battery as my ground, how many volts are at the *positive* terminal?
4. If I choose the *positive* terminal of this battery as my ground, how many volts are at the *negative* terminal?
5. If I have a Point A on my circuit that is 7 volts above ground and I have a Point B on my circuit that is 2 volts above ground, what is the voltage difference between Point A and Point B?
6. Given a constant voltage, what effect does increasing the resistance have on current?
7. Given a constant current, what effect does increasing the resistance have on voltage?

8. If I have a 10 V battery, how much resistance would I need to have a current flow of 10 amps?
9. If I have a 3-volt battery, how much resistance would I need to have a current flow of 15 amps?
10. Given 4 amps of current flow across 200 ohms of resistance, how much voltage is there in my circuit?
11. If I am wanting to limit current flow to 2 amps, how much resistance would I need to add to a 40-volt source?
12. If I am wanting to limit current flow to 2 milliamps, how much resistance would I need to add to a 9-volt source?
13. If I am wanting to limit current flow to 20 milliamps, how much resistance would I need to add to a 5-volt source?

CHAPTER 5

Your First Circuit

In the last two chapters, we have learned about the fundamental units of electricity—charge, current, voltage, and resistance. In this chapter, we are going to put this information to use in a real circuit.

5.1 Circuit Requirements

For a circuit to function properly, you usually need several things:

1. A source of power which provides electricity for your circuit (this is usually a combination of both a source *and destination* for the electricity)
2. A network of wires and components that ultimately lead from the source to the destination
3. Some amount of resistance in your circuit

We need the source because, without a source of power, the charge won't move! If we have a circuit, but no electrical power, it will just sit there. In our circuits, batteries will usually provide the power we need. They will do this by providing a (relatively) constant voltage to our circuits.

We need the wires because unless we provide a *complete pathway* from a higher voltage (the source) to a lower voltage (the destination), the electricity won't be able to move. If we want the charge to move, we have to

© Jonathan Bartlett 2020

J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_5

make a *pathway* from the higher voltage to the lower voltage. Without this pathway, we have what is known as an **open circuit**. No electricity flows in an open circuit.

The components in the pathway from the source to the destination do the "electrical stuff" that we want to accomplish, whether that is turning on a light, running a motor, running a computer, or whatever else it is we want to do. In order for them to take advantage of the electrical power, they have to be in the pathway in which the current is moving. Think of this like a watermill—those structures that sit near a river and use the power of the water moving through the river to turn a wheel and therefore operate whatever is inside the house. In order for that to work, the water must flow through the watermill-connected structure. If the watermill is not built next to the river or if the river runs dry, the watermill doesn't work. Similarly, if the components are not in the flow of moving electric charge, they will not operate.

However, in addition to the wires and components, we must also have resistance (though sometimes this is added by the components themselves). Without resistance, the current would be too high. If you had zero resistance, the current would be so high that it would immediately drain your battery and likely destroy all of your components that you have connected. You can actually see this using Ohm's law. If we have a 10-volt source with no resistance, the current is given by the equation $I = V/R = 10/0 \approx \infty$. Dividing by zero gives you, essentially, infinite current. Now, wires and batteries *themselves* have some amount of resistance, so the current wouldn't be infinite, but it would be very, very large and would quickly drain your battery and destroy any sensitive components you had connected. Therefore, every pathway from the positive side of the battery to the negative *must* have some measurable amount of resistance. When a pathway from positive to negative occurs without resistance, this is known as a **short circuit**.

In other words, to accomplish real tasks with electricity, we must control its flow. If it doesn't flow (as in an open circuit), it can't do anything. If it flows without resistance (as in a short circuit), it does damage rather than work. Therefore, the goal of electronics is to provide a controlled route for electric charge to follow so that the power of electricity does the things we want it to do on its way from the source to the destination.

SHORTING OUT

If you've ever heard the phrase "shorting out," that refers to the fact that if you have a short circuit, it will often run too much current through a component and break the component. Sometimes this causes sufficient damage to actually damage the connections within the circuit. So, although the *original problem* was that there was a short circuit (i.e., something that created a pathway which bypassed resistance in the circuit), the *resulting problem* is that the damage leaves you with a permanent open circuit.

Interestingly, shorting out is actually usually a better result than having a permanently short circuit, which would leave you open to your project heating up and potentially catching on fire. This is why many projects which have larger power sources often use fuses. The goal of the fuse is to break either before other parts of the circuit break or before something catches fire.

5.2 Basic Components

The first circuits that we will build will only use four basic types of components:

- Batteries (9-volt)
- Battery regulator
- Resistors
- LEDs

As we have discussed, batteries provide a relatively constant amount of voltage between the positive and negative terminals. A 9-volt battery, therefore, is supposed to have a 9-volt difference between the positive and negative terminals. However, there is an actual range that 9-volt batteries have. A 9-volt battery, depending on the battery chemistry, will deliver between 7 and 10 volts. Additionally, when the battery is worn out, its voltage will decrease as well.

Therefore, in order to make the projects operate more uniformly across different power sources, our projects will often include a regulator, which will take the battery's voltage (which can vary from 7 to 10 volts) and return a constant 5-volt output. So keep in mind that if the project does not have a regulator attached, the actual voltages in the circuit may vary depending on the battery. If the project does have a regulator attached, it should be reliably yielding a 5-volt difference from the positive to the negative. We will treat the regulator just like a battery, but one that is more stable and reliable than a battery.

A resistor is a device that, as its name implies, adds resistance to a circuit. Resistors have colors that indicate how much resistance they add to the circuit. You don't need to know the color codes yet, but if you are curious, you can see the resistor datasheet in Appendix E. So, if we want to add 100 Ω to our circuit, we just find a resistor with a value of 100 Ω. Resistors are not the only devices that add resistance to a circuit, but they are usually what are used when you want to add a fixed amount of resistance. Resistors have two terminals (connecting points), but they both function identically—unlike other components, there is no backward or forward for a resistor. You can put them in your circuit either way, and they will function just fine.

Of the components in this section, the LED is probably the strangest. LED stands for "light-emitting diode." A diode is a component that only allows current to flow in one direction. It blocks the flow of electricity in the other direction. However, more importantly, LEDs emit light when current passes through them. However, LEDs do not resist current, so they must be used with a resistor to limit the amount of current flowing through them (most of them will break at 20–30 milliamps). Also, since LEDs only allow current to flow one way, they have to be wired in the correct direction. To find out which way to wire an LED, look at the legs of the LED—one is longer than the other. To allow current flow, the longer leg of the LED should be on the more positive side of the circuit.

Most of your components (especially your resistors) come with very long legs. You can feel free to bend or cut these legs however you please to better fit in your circuit. However, on LEDs (and any other component where leg length matters), be sure to keep the longer legs longer so you don't get confused about which way it should be inserted in your circuit.

5.3 Creating Your First Circuit

Now we will put together a simple first circuit. What you will need is

- One 9-volt battery
- One red LED (other colors will work too)
- One 1,000 Ω resistor (anything from 400 ohms to 2,000 ohms should work)

If you already have resistors, you can use a multimeter to determine the resistance, or you can use the color chart in Appendix E. Most electronics enthusiasts I know don't actually use the color codes, because they are too small. If you have better vision than we do, then the color chart may work great. Otherwise, use a multimeter to check the resistance, or, better yet, keep your resistors organized so it is easy to find the value you want. If you don't have resistors, it is even easier—just buy a resistor with the resistance you are looking for!

READING A RESISTOR'S VALUE FROM A MULTIMETER

Multimeters are incredibly flexible tools, and they are pretty cheap too. For the purposes of this book, you don't need anything fancy—the cheapest one will do just fine! Be sure that the resistor is not connected to anything else before testing.

To find the value of the resistor, look for values on the multimeter marked with the ohm sign (Ω). Set the dial to the lowest resistance value it has (remember to think about the metric suffixes from Chapter 2—for instance, 10 M means 10,000,000). Then, test the resistor by putting the two multimeter probes on the two legs of the resistor. It is okay to hold the probes on the resistor legs with your fingers. Many multimeters require you to only use the pointed tips of the probes in order to measure values accurately.

If the screen shows a 0 or 1, that means that the resistor's value is actually too large for it to register at that setting. In that case, move the dial on the multimeter to the next setting higher and try again. If the screen shows a more precise value, then use that value plus whatever metric suffix the multimeter is set to. For instance, if the multimeter is set to 10 M, then you would add the M suffix to whatever value is shown on the screen. That is the value of the resistor.

If you did buy a fancy, expensive multimeter, it may only have one spot for measuring resistance, marked as ohm without listing any number by it. In this case, it will automatically do that same process internally and report the results to you like magic.

More information about multimeter use is found in Chapter 6.

To make this circuit, take one leg of the resistor and twist it together with the *short* leg of the LED. It should look like Figure 5-1.

Now, take the long leg of the LED and touch it to the positive terminal of the battery. Nothing happens—why not? Nothing happens because, even though we have connected the wires to the positive side of the battery, the electricity has nowhere to go to, so the current won't move. We have an open circuit because there is not a complete path from positive to negative.

Now, touch the long leg of the LED to the battery and, at the same time, touch the unattached end of the resistor to the battery. The LED should give a nice glow of its color. Congratulations! You have built your first circuit!

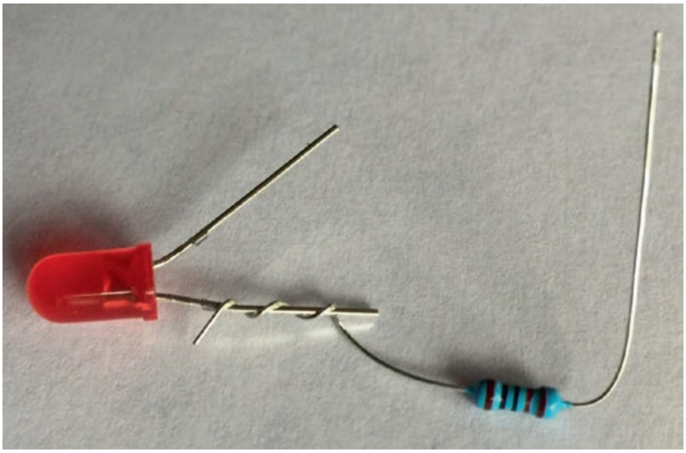

Figure 5-1. *Wrapping the Resistor Around the LED's Short Leg*

Even though we can't see the electricity moving, I hope you understand how it will flow through the circuit. We can trace the current flow from the positive terminal of the battery through the LED. The resistor limits the amount of current flowing through the circuit and therefore through our LED (the resistor can actually go on either side of the LED; it will limit the flow no matter which side it is on). Without the resistor, the battery would easily go over the 30-milliamp rating of our LED, and it would short out and no longer work. If you connected it without a resistor, you might see it turn on for a moment and then very quickly turn off, and then it would never work again. If you have an extra LED, you can try this out if you want. It is not dangerous—it will just cost you the price of an LED.

If your LED is backward, no current will flow at all. It won’t hurt the LED, but it won’t turn on unless it is oriented in the right direction.

5.4 Adding Wires

We are not going to physically add wires to our circuit at this time, but I did want to make a note on wires. Changing the lengths of wires will not affect our circuits in any way. On advanced projects (usually projects with extremely high precision or extremely long wires), the length of a wire will sometimes have an effect on such circuits. We are not doing any high-precision circuits, and our wire lengths are all less than a meter. Therefore, for the electronics we are doing, we can totally ignore wire length.

Therefore, if we connected our components using wires rather than directly wrapping their legs around each other directly, it would have no effect on the circuit at all. What is important is not the wires but the connections—what components are connected together and how are they connected. The length of the wire used to connect them is not important. You can connect them with a really long wire or directly touch the legs of each component together with no wires. It doesn’t matter because the result is the same—the components are connected.

Symbol	Component	Description
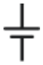	Battery	A battery is represented by a long line and a short line stacked on top of each other. Sometimes, there are two sets of long and short lines. The long line is the positive terminal and the short line is the negative terminal (which is usually used as the ground).
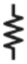	Resistor	A resistor is represented by a sharp, wavy line with wires coming out of each side. This represents that the resistor is not an easy way for current to flow.
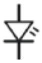	LED	An LED is represented by an arrow with a line across it, indicating that current can flow from positive to negative in the direction of the arrow, but it is blocked going the other way. The LED symbol also has two short lines coming out of it, representing the fact that it emits light.

Figure 5-2. *Basic Component Diagram Symbols*

5.5 Drawing Circuits

So far, we have only described circuits in words or by showing you pictures. This, however, is a lousy way of describing circuits. In complicated circuits, trying to trace the wires in a photograph is difficult. If you wanted to draw a circuit that you wanted built, you would have to be an artist to render it correctly. Likewise, reading through text describing a circuit takes a long time, and it is easy to get lost when discussing large circuits.

Therefore, in order to communicate information about how a circuit is put together in a way that is easy to read and write, engineers have developed a way of drawing circuits called **circuit diagrams** or **electronic schematics** (often just called *diagrams* or *schematics*). In a circuit diagram, each component is represented by an easy-to-draw symbol that helps you remember what the component does. Figure 5-2 shows the symbols for the components we have used so far. Note that everybody draws the symbols slightly differently and some components have more than one symbol. However, these are the symbols we will use in this book. For more symbols, see Appendix B.

Then, the components are connected together using lines to represent the wires and connections between the components.

Therefore, we can redraw our original circuit using these symbols as you see in Figure 5-3.

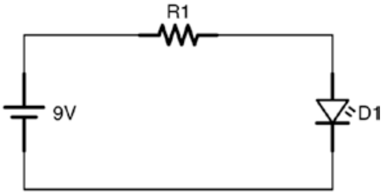

Figure 5-3. *Basic LED Circuit Drawn as a Diagram*

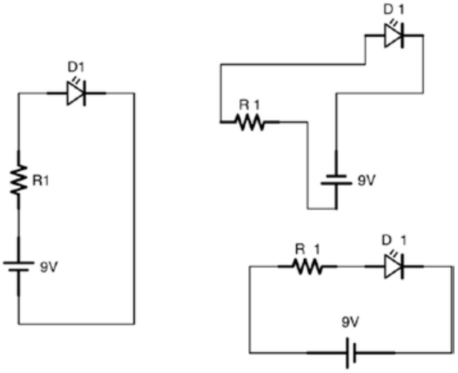

Figure 5-4. *Alternative Ways of Drawing the Basic LED Circuit*

Notice that our components are laid out on the diagram with wires connecting them. Remember that it doesn't matter if we have very long wires or very short wires or if the components are directly placed end to end—the resulting circuits will operate identically. Also notice that each component is labeled (R1 and D1) because, as we make more complicated circuits, it is important to be able to refer back to them.

It does not matter in a diagram which way you have your components turned, how long or short your wires are, or what the general spacing looks like. When you actually wire it, all of those things will change. The important part of a circuit diagram is to convey to the reader what the parts are, how they are connected, and what the circuit does in the way that is easiest to read.

For instance, all of the circuits in Figure 5-4 are equivalent to the circuit in Figure 5-3, they are just drawn differently. All of them have the positive terminal of a battery connected to a resistor, with the other leg of the resistor connected to the positive end of an LED, with the other leg of the LED connected to the negative side of the battery. Since the connections are identical, the circuits are identical. Schematics don't care how long the lines are that you draw or how the components are oriented or even where the components are drawn on the page. The only thing important in the schematic is how the components are connected.

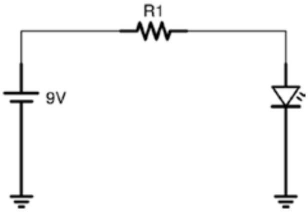

Figure 5-5. *Basic LED Circuit Drawing Using the Ground Symbol*

For consistency, I like to draw all of my batteries to the left of the drawing with the positive side on top. By keeping the battery positive-side-up, components with higher voltage are usually closer to the top, and components with lower voltages are usually closer to the bottom, with the ground (i.e., 0 volt) coming back into the negative terminal. I also try to make my wire lines as simple as possible in order to make following them easier.

By keeping some amount of consistency, it is easier to look at a drawing and see what is happening.

5.6 Drawing the Ground

Remember that for electricity to move, every circuit must be fully connected from the positive side to the negative side. That means that in larger circuits, there are numerous connections that come from the positive or go back to the ground/negative. Because of this, a special symbol has been adopted to refer to the ground point in a circuit. This symbol, the ground symbol, has three lines, each shorter than the next. Points on a circuit that have this symbol connected to them are connected to each other (usually they are all connected to the negative side of the battery).

Therefore, the circuit in Figure 5-5 is the same circuit as before, just drawn using the ground symbol. Since points with the ground symbol are all connected together, using this symbol on both the negative terminal and the negative side of the LED means that they are wired together.

This doesn't help us a lot for this circuit (and, in fact, it makes it a little less easy to read). However, in complex circuits, it is much easier to write the ground symbol than trying to have 20 lines drawn back to the negative terminal.

Additionally, the same is true with the positive side of the battery. Many components require a direct connection to a specific voltage to work correctly. These are usually marked with just a disconnected wire with the end of the wire marking what voltage it requires. We make less use of that symbol in this book than the ground symbol, but it does come in handy sometimes.

So, using both the voltage source and ground symbols, we could rewrite the same circuit again in the manner shown in Figure 5-6. This circuit, again, is not *wired* any differently than before. We are just *drawing* it differently. For this circuit, it doesn't matter, but in more complex circuits, if we need a specific voltage at a specific location, this symbol tells us to put it there.

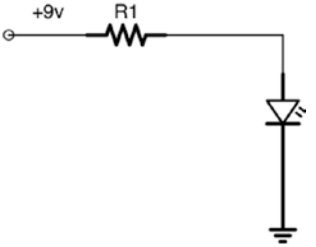

Figure 5-6. *Simple LED Circuit Using Positive and Ground Symbols*

Review

In this chapter, we learned the following:

1. Every circuit requires a source of power (usually a battery), wires and components, some amount of resistance, and a complete path back to the negative side of the power source.
2. An open circuit is one that does not connect back to the negative side (and thus does not provide any electricity), and a short circuit is one that connects back to the negative side without any resistance (and thus overwhelms the circuit with current).
3. A battery supplies a fixed voltage between its two terminals.
4. A resistor provides a fixed resistance (measured in ohms) within your circuit.
5. An LED allows current to flow in only one direction, gives off light when current is flowing, but is destroyed when the current goes above 20–30 milliamps.

6. The longer leg of the LED should be on the positive side of the circuit.
7. Wires on simple circuits can be almost any length (from zero to a few meters) without changing the functionality of the circuit.
8. A circuit diagram is a way of drawing a circuit so that it is easy to read and understand what the circuit is doing.
9. Each component has its own symbol in a circuit diagram.
10. Components labeled with the ground symbol are connected together, usually at the negative side of the battery.
11. Voltage sources can be similarly labeled by a wire connected on one side labeled with the voltage that it is supposed to be carrying.

Apply What You Have Learned

Special Note: In the following problems, since we have not yet studied LED operation in depth, we are ignoring the electrical characteristics of the LED and just focusing on the resistor. If you know how to calculate the circuit characteristics using the LED, please ignore it anyway for the purpose of these exercises.

1. Calculate the amount of current running in the circuit you built in this chapter using Ohm's law. Since Ohm's law gives the results in amps, convert the value to milliamps.

2. Let's say that the minimum amount of current needed for the LED to be visibly on is 1 milliamp. What value of the resistor would produce this current?
3. Let's say that the maximum amount of current the LED can handle is 30 milliamps. What value of the resistor would produce this current?
4. Draw a circuit diagram of a short circuit.
5. Take the circuit drawing in this chapter, and modify it so that it is an open circuit.
6. Draw a circuit with just a battery and a resistor. Make up values for both the battery and the resistor and calculate the amount of current flowing through.

CHAPTER 6

Constructing and Testing Circuits

In Chapter 5, we learned how a circuit works. However, the method of putting together a circuit in the chapter doesn't translate to the real world well. In this chapter, we will learn how to use solderless breadboards to construct circuits in a more robust manner. Additionally, we will put our understanding of Ohm's law to the test as we learn how to measure voltages in circuits using our multimeter.

6.1 The Solderless Breadboard

The most important piece of equipment to use for making circuits is the **solderless breadboard**. Before solderless breadboards, if you wanted to put together a circuit, you had to attach them to a physical piece of wood to hold them down and then **solder** the pieces together. Soldering is a process where two wires are physically joined using heat and a type of metal called solder, which melts at much lower temperatures than other types of metal. So what you would have to do is attach the electrical components to the board, wrap the components' legs around each other, and then heat them up with a soldering iron and add solder to join them permanantly.

This was an involved process, and, though it was sometimes possible to get your components back by reheating the soldered joints, you were generally stuck with your results. The solderless breadboard is an amazing

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_6

invention that allows us to quickly and easily create and modify circuits without any trouble at all. Figure 6-1 shows what a solderless breadboard looks like, and Figure 6-2 has the different parts of the breadboard labeled.

The solderless breadboard has a number of spring clips (usually about 400 or 800 of them) called **connection points** which will allow you to insert wires or component leads and will hold them in place. Not only that, the breadboard itself will connect the components for you!

The way that this works is that the breadboard is broken up into little half-rows called **terminal strips**. Each terminal strip has multiple connection points—usually five. Each connection point on a given terminal strip is connected by wire *inside* the breadboard. Therefore, to connect two wires or leads together, all you need to do is connect them to the same terminal strip. Any two wires or leads connected to the same terminal strip are themselves connected.

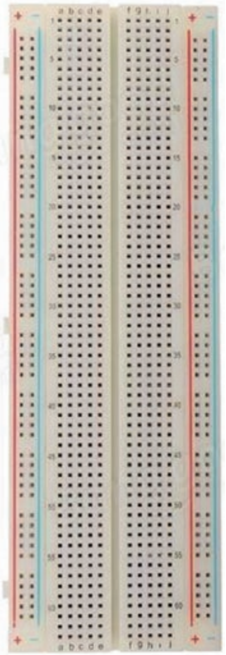

Figure 6-1. *A Solderless Breadboard*

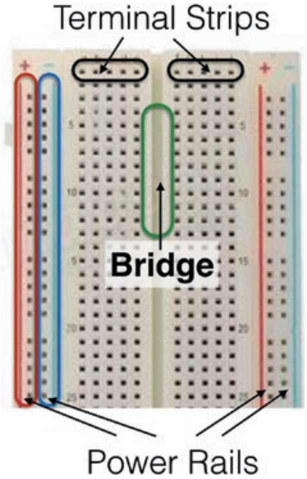

Figure 6-2. *Parts of a Solderless Breadboard*

In most breadboards, the two sides of the breadboard are separated by a gulf known as the **bridge**. The bridge is a visual indication that the two sets of terminal strips are not connected, but it also serves a practical purpose. If you have an integrated circuit (a small chip), the bridge is the right width so that you can place your integrated circuit right over the bridge, and each leg of the chip will receive its own terminal strip for you to easily connect them to what you need. We will cover this in more depth in later chapters.

In addition to the terminal strips, most breadboards have two strips running down each side, one with a red line and one with a blue line. These are known as **power rails** (some people call them **power buses**).

Power rails are very similar to terminal strips, with a few exceptions. The main difference is that, in terminal strips, only the five connection points grouped together are connected. On power rails, many more of the connection points are connected together, even when there are short gaps. Some boards will split the power rails at the halfway point, but others go all the way down the board. A split in the power rails is usually visually indicated by a break in the red and blue lines that indicate the power rails.

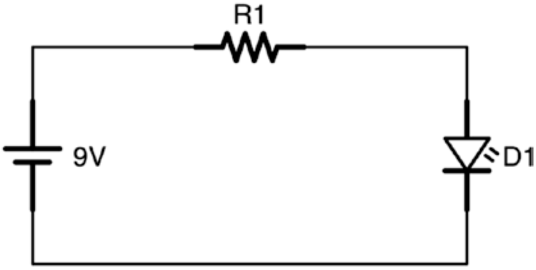

Figure 6-3. *Basic LED Circuit*

Note that the positive and negative are *not* connected to each other (that would create a short circuit) and they are *not* connected to the power rails on the other side of the breadboard (unless you connect them manually). On some breadboards, even a single side isn't connected all the way down, but may be broken into sections at the halfway point.

In many projects, many components need direct access to the positive or negative power supply. Power rails make this easy by providing a connection point with positive and negative power a very short distance away from wherever you need it on the breadboard. If you plug your power source's positive and negative terminals into the positive and negative rails on the breadboard, then any time you need a connection to the positive or negative terminal, you can just bring a wire to the closest connection point on the appropriate power rail.

6.2 Putting a Circuit onto a Breadboard

To see how a simple circuit works on a breadboard, let's go back to the circuit we first looked at in Chapter 5. Figure 6-3 has the drawing again for ease of reference.

So how do we translate what we see in the drawing to what we need to put in the breadboard? Well, let's take a look at what is in the circuit—a 9-volt battery, an LED (we will go with a red LED), and a 1,000 Ω resistor.

Let us not concern ourselves with the battery at the moment. So, without the battery, we have a resistor connected to an LED.

Let us start out by simply placing our components onto the breadboard. What you will want is to place them on the breadboard so that each of their legs is on a *different* terminal strip. It doesn't matter *which* terminal strips you use—just make sure the legs all get plugged into different ones. Figure 6-4 shows how your breadboard should look so far. Note that the longer leg of the LED is closer to the resistor. The longer leg is depicted in the diagram as having an extra bend in the leg.

Figure 6-5 shows the *wrong* way to do it. In that figure, the legs of both the components are on the same row, which is the same thing as placing a wire between the legs, creating a short circuit. Don't do that! Make sure each leg goes into its own row.

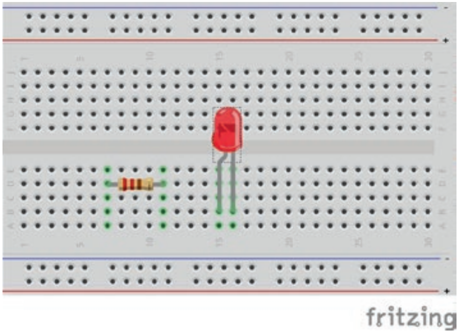

Figure 6-4. *Putting the Components onto the Breadboard*

Now, to connect the resistor to the LED, we need to add a wire. So all we need to do is connect a wire to any empty connection point that is on the same terminal strip of the right leg of the resistor and connect the other side of that wire to the left leg of the LED as shown in Figure 6-6.

Note that wires used on breadboards are often called "jumper wires." The difference between these and other kinds of wires is that jumper wires usually have solid, hard ends, making them easier to push into the breadboard. If you use regular wire or speaker wire, the ends are flexible, and it is almost impossible to get them into the breadboard properly to connect.

A common mistake that people will make is to connect the wire to the row right before or after the component. Take some time and be extra certain that the wire is connected to the *very same* row as the leg of your components.

Now, we need to connect our project to the power rails. So take a red wire from the left leg of the resistor to the positive power rail (remember, as long as it is in the same terminal strip as the resistor, they will be connected). Likewise, take a black wire from the right leg of the LED to the negative power rail. I always use red wires for connecting to the positive power rail and black wires for connecting to the negative/ground rail, as it makes it more clear when I am looking at my project what wire carries what. Your project should look like Figure 6-7.

Now your project is almost done. All you need to do now is to connect your power rails to a power supply. Connect a T-connector to a 9-volt battery, and then connect the red (positive) wire to the positive power rail on the breadboard.[1] You can plug it in anywhere on the rail, but I usually connect the power to the edge of the rail to leave more room for components. Then, connect the black (negative) wire to the negative power rail on the breadboard. As soon as you do this, the LED should light up! Figure 6-8 shows the final circuit.

---

[1] A T-connector is just a cap on the battery that has wires coming out of it. If you don't have a T-connector, you can manually plug a jumper wire into the breadboard and hold the other end on the battery terminal. T-connectors have their own drawbacks, as most of them have flexible ends, making them difficult to plug into breadboards. Shortly, we will move to an easier way of handling power with power regulators.

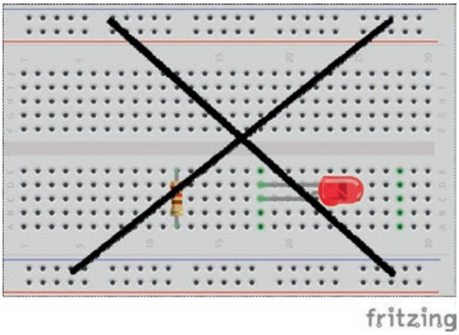

Figure 6-5. *The Wrong Way to Put Components onto the Breadboard*

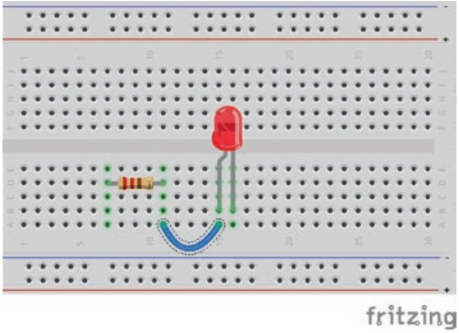

Figure 6-6. *Adding a Wire to Connect the Components*

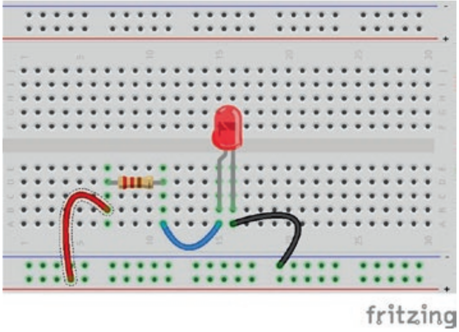

Figure 6-7. *Adding Wires to the Power Rails*

Note that many T-connectors for 9-volt batteries have very flimsy wires that are difficult to insert into a breadboard. Usually, as long as you can get both terminals in far enough to touch the metal within the connection point, it will work. We will move to using power regulators shortly, which makes the process a bit easier.

If your circuit doesn't work, here is a list of things to check:

1. Make sure your battery is properly connected to the breadboard—the red should go to positive and the black to negative.
2. Make sure there are *no* wires directly connecting positive to negative on the board. Any direct pathway from positive to negative without going through a component will cause a short circuit and can destroy your components and battery.
3. Make sure that your wires are connected to the same terminal strip as the component lead that they are supposed to be connected to. If they are on a different row, *they are not connected*!
4. Make sure the LED is inserted in the right way. The longer leg should be connected to the resistor, and the shorter leg should be connected to the negative power supply.
5. Make sure your components are good. Try replacing your LED with another LED to see if it works.
6. If all of those things fail, take a picture of your project and post it to an online forum. Someone will likely be able to spot your problem and/or lead you in the right direction. Many forums are also available on the Web for this.

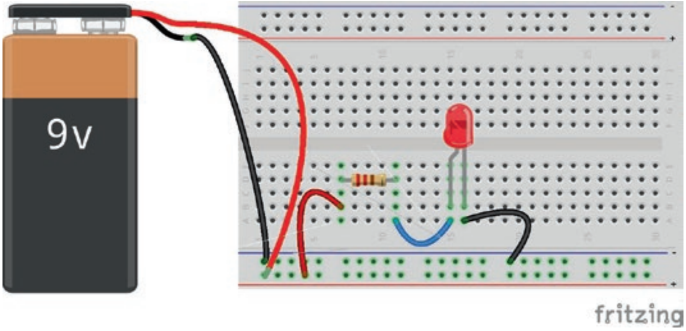

Figure 6-8. *Final LED Circuit with Power Connected*

6.3 Using Fewer Wires

In the previous section, we used three wires to connect our components, plus two more wires from the battery. We can improve our project by reworking it so that most of the wires are not necessary.

Remember that any two leads or wires plugged in next to each other on the same terminal strip are connected. Therefore, we can remove the wire that goes from the LED to the resistor simply by moving the LED and resistor so that the right leg of the resistor is on the same terminal strip. Figure 6-9 shows what this looks like.

However, that middle wire is not the only redundant wire. If you think about it, we could also save a wire by actually using the LED's own lead to go back to the negative rail. Figure 6-10 shows how this is set up. Now, in order to make the LED fit better, it is now on the *other* side of the resistor in the terminal strip. Remember that this does not matter at all! No matter where a component is connected on the terminal strip, it is joined with a wire to every other component on the same terminal strip.

Now, there is one last wire that we can get rid of. Can you think of which one it is? If you said the wire going from the positive rail to the resistor, you were right.

What we can do is to directly connect the resistor to the positive rail. Doing this gives us what is shown in Figure 6-11.

Therefore, as you can see, there are any number of ways that you can arrange parts on a breadboard to match a given schematic. All of these arrangements we have seen match the schematic given in Figure 6-3. As long as your circuit matches the configuration in the schematic, the specifics of where you put the wires and components is up to you. Some people like to place the components on their breadboard first, spaced out, and then add wires to connect them as needed. This works, though it does make for a messier board, with lots of wires going every which way. Other people like to use as few wires as possible and have their layouts as clean as possible (i.e., they don't like a tangled mess).

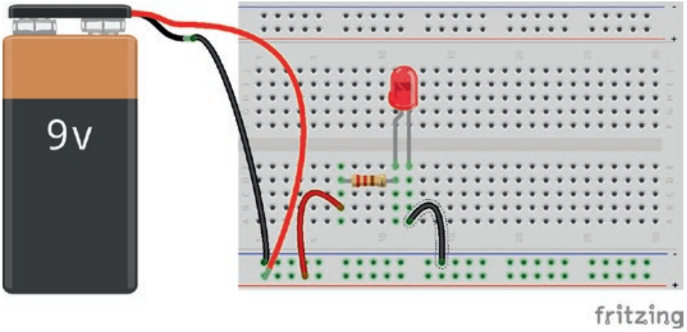

Figure 6-9. *Joining Components by Putting Their Leads on the Same Terminal Strip*

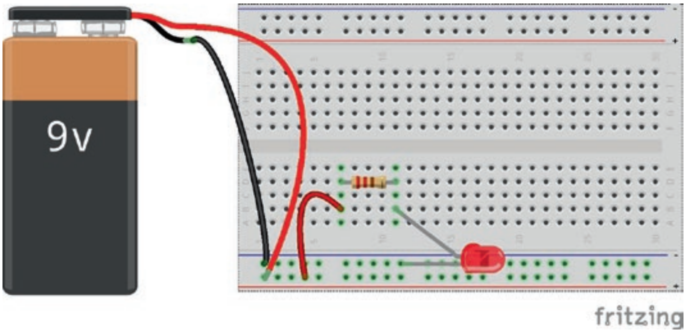

Figure 6-10. *Directly Connecting the LED to the Negative Rail*

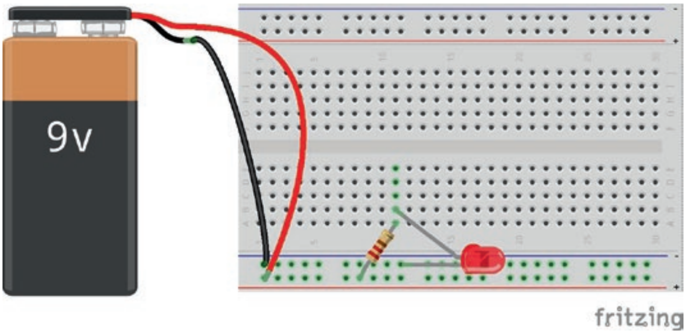

Figure 6-11. *Connecting the Resistor Directly to the Positive Rail*

Some people like to use flexible jumper wires that go up and over the board. Other people like to use rigid jumper wires that lie down close to the board and are the exact length needed. The flexible wire allows more flexibility in building your circuits (they are easier to move around and reconfigure), while the solid, rigid wire makes the final result a lot cleaner and easier to follow.

You can also trim the legs of your components to make them fit better, if you want. Some people like to leave their components as intact as possible, while others like to trim the legs of their leads to be the exact right size for their project. However, if you do trim the leads on your LEDs, be sure to keep the positive leg longer!

However you like to work with electronics is up to you. There are lots of options, and they all end up with the same circuit.

6.4 Testing Circuits with a Multimeter

Now that we know how to put circuits together, we need to know how to *test* our circuits. The main tool used to test simple circuits is the **multimeter**. It is called a multimeter because it measures *multi*ple different things about a circuit. Figure 6-12 shows a typical low-cost multimeter.

There are a lot of different multimeters around which have a lot of different functions. However, almost all of them will measure voltage, current, and resistance. Each of these values is measured by testing two different points on the circuit. Most multimeters have a red lead and a black lead. The red lead should connect to the more positive side of the circuit, and the black lead should connect to the more negative side of the circuit. However, if you get it reversed, it is usually fine—the multimeter may just report negative values if you are measuring for voltage or current.

To illustrate how to use a multimeter, we will start out measuring the voltage in a 9-volt battery. Remember from Chapter 4 that there is no absolute zero voltage—voltages are merely measured with reference to each other. Therefore, a multimeter doesn’t tell you the exact voltage of something—there is no exact voltage. Instead, a multimeter allows you to choose two points on your circuit and measure the voltage difference (also known as the **voltage drop**) between them.

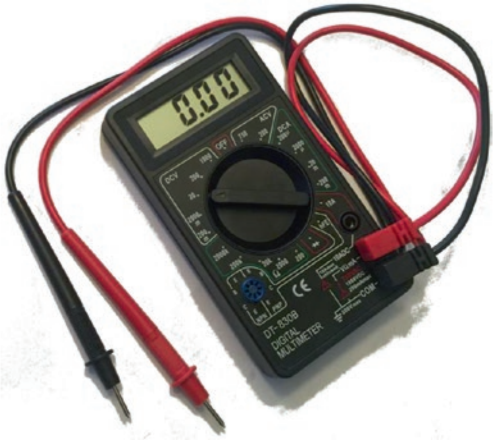

Figure 6-12. *A Low-Cost Multimeter*

Now, remember that a 9-volt battery means that the battery should have 9 volts between its positive and negative terminals. Don't try it yet, but when we measure voltage, we will expect that the multimeter will tell us that the voltage difference is near 9 volts.

When using your multimeter, you must set *what* you are going to test *before* you test it. Otherwise, you can easily damage your multimeter or your circuit. Therefore, since we are going to measure voltage, select the DC Voltage setting on your multimeter (*do not* select the DC Current or DC Resistance setting!). If you are using a high-quality **auto-ranging** multimeter, that is all you need to do. However, when starting out, most people buy the bottom-of-the-line multimeter. That's not a problem, but just know that you will probably accidentally break it at some point.

If you are using a lower-quality multimeter, you will need to select not only *what* you want to measure but the *estimated range* of values that you want to measure. On my multimeter, the DC Voltage has five different settings—1000, 200, 20, 2000 m, and 200 m. These are the upper boundaries (in volts) that these settings can read.

So, for a 9-volt battery, using the 1,000-volt setting is probably unwise. It may give a reading, but it probably won't be accurate. However, if I try it on too low of a setting (say, 2000 m), it either won't read, or it will blow out my multimeter. So the safe thing to do is to start with the highest reasonable setting (or just the highest setting if you don't know what's reasonable) and test it and then reduce the setting until it gives you a good reading.

So, for instance, for my 9-volt battery, let's say I didn't know the voltage. Therefore, I'm going to measure the battery using the 1,000-volt setting. After setting the multimeter to 1,000 volts, I will put the red lead on the positive terminal of the battery and the black lead on the negative terminal. Be sure that you *firmly* press the *tip* of your leads against the positive and negative terminals. If it is not firm or if you use the sides of your terminals, you will not get a good reading.

When I do this, my multimeter reads 9.

Now, notice that this reading is significantly less than our 1,000-volt setting. Therefore, it may not be entirely accurate. So I will reduce the setting to the 200-volt setting and measure again. This time, my multimeter reads 9.6. This is definitely a more accurate reading—it is giving me an extra digit of accuracy! However, this reading is still significantly below the setting.

Therefore, I will reduce the setting again to the 20-volt setting and remeasure. This time, the measurement is 9.66. Again, it is more accurate. Now, can I reduce the setting even more? Well, the next setting is 2000 m, which is basically 2 volts. Our reading is 9.66 volts, so it is above the cutoff point for the next setting. Therefore, I should not try it on a lower setting, both for the sake of accuracy and for the sake of my multimeter's lifespan.

However, I should note that if I did use a lower setting, since the setting is listed as being in millivolts (i.e., 2000 m), then the reading would also be in millivolts. That is, if we were to read the value of the battery on that setting, it would say 9660, because that is how many millivolts the battery has.

Now, you could be wondering, why is a 9-volt battery anything other than exactly 9 volts? Well, it turns out that in electronics, no value is exact, and no formula works perfectly. When we talk about a 9-volt battery, we are actually talking about a battery that runs anywhere from 7 volts to

9.7 volts. In fact, my battery that started out at 9.66 volts will slowly lose voltage as it discharges. This is one of the reasons why measurement is so important.

Also, this means that in our circuits, we will have to find ways to compensate for varying values. Our circuits should work across a wide range of possible values for our components. We will discuss strategies for this as we go forward.

The next thing we will measure is resistance. Pull out a resistor—any resistor. The resistor datasheet in Appendix E shows you how to find the resistor values based on the color bands on the resistor. I don't know about you, but my eyes are not that good at looking at those tiny lines on the resistor and figuring out which color is which. Many times, it is easier just to test it with the multimeter.

The process is the same as with measuring the voltage, except that you start at the lowest setting and work up. First, find the resistance settings on your multimeter (perhaps just marked with the symbol for ohms—Ω). Start with the smallest value (200 in my case). On this setting, the multimeter read 1, which means that it didn't pick up a signal. So I turned it up to the next setting, 2000. This time, it read 1002. Remember that if the value that the meter is set to includes a metric suffix (i.e., m, k, or M), that suffix gets applied to the value that is displayed on the screen.

Note that you should *never test for resistance in a live circuit*. The multimeter uses power to measure resistance, and if there is already power in the circuit, it can damage the multimeter and/or the circuit. However, to test for current and voltage the circuit must be powered.

6.5 Using a Multimeter with a Breadboard

We can use our multimeter with our breadboard too. Let's say that we wanted to measure the voltage between the positive and negative rails of the breadboard.

There are two ways to do this. The first, if the size of your multimeter probes and the size of your breadboard connection points allow it, is to simply shove the probes of your multimeter into connection points on the positive and negative rails. Since these will be connected to the power by a wire, these will be at the same voltage levels as the battery itself.

Second, if your breadboard/multimeter combination does not support this, you can do the same thing by simply connecting two jumper wires into the positive and negative rails and then testing the voltage on the other end of the wires.

Also, if you are testing components for voltage, you can also use your multimeter on the exposed legs of the component. This is often easier than either trying to push your probes into the breadboard or running extra wires to your multimeter.

To try out using your multimeter with your breadboard, configure your breadboard similar to Figure 6-8. Use this layout, and *not* one of the ones with fewer wires (you will see why in a minute). With the battery connected to the breadboard, set your multimeter to the highest voltage setting, and put the red lead in any empty hole in the positive rail. While that lead is there, put the black lead in any empty hole in the negative rail.

This should give you the same reading that you received for the battery terminals. Remember that the power rails are connected all the way across—that is why putting your probes in any hole on the line works! If you work your way down the ranges on your multimeter, you should find that you get the same value that you did when you measured directly on the battery's terminals. Again, if your probes do not fit inside the connection points, you can also use wires to connect out from your breadboard to your multimeter probes.

You can now do the same to any component on your board. Let's find the voltage difference between one side of the resistor and the other. To do this, find an empty hole on the same terminal strip as the left-hand side of the resistor, and put the red lead from your multimeter in that hole. Then, find an empty hole on the same terminal strip as the right-hand side of the resistor, and put the black lead from your multimeter in that hole. Now you can measure the voltage difference. Note that to measure voltage differences, the circuit *must* be active. If the power is gone, the voltage difference will likely drop to zero. Use the same ranging procedure to find the voltage drop between the left-hand and right-hand sides of the resistor.

Even though we have not discussed diodes, this doesn't prevent you from measuring the voltage difference between the legs of the diode in your circuit. Use the same procedure as before to measure the voltage drop.

6.6 Measuring Current with a Multimeter

Now we will learn to measure current using the same circuit layout from Figure 6-8. Like voltage, measuring current requires that the power to your circuit be on. To measure current, use the DC Amperage (sometimes called DC Current) settings on your multimeter.

Measuring current is a little different than measuring voltage in a circuit. Instead of just placing your probes in the breadboard as it is, you are going to use your probes to *replace a wire*. You will remove a wire and then place your probes in the holes (connection points) where the wire used to be. Alternatively, if your multimeter probes do not fit into the connection points, you can again run two wires, one from each hole, from the breadboard to your multimeter probes.

Using either of these approaches, the circuit will then use your multimeter *as the wire* that was removed, and the multimeter will then measure how much current is running through it and report it to you on the screen. You will need to use the same ranging technique as you used before with voltages to get an accurate report.

Let's say that you wanted to measure the current going through the wire that connects the resistor to the LED. To do this, we will start by *removing* that wire, and connecting the red lead to where the wire used to be on the left (since it is more positive) and the black lead to where the wire used to be on the right (since it is more negative). The multimeter should now report back how much current the circuit is using. This will vary for a number of reasons, but should be about 17 mA.

Now, put the wire back, and remove another wire and measure current there. No matter which wire you choose, they should all measure the same current. The reason is that, since all of these components are in series (one right after the other), they must all have the same amount of electricity flowing through them (otherwise, where would the electricity be going?).

6.7 Using a Power Regulator

So far, we have discussed several problems of using batteries in electronics projects. First, with batteries, we don't know exactly how much voltage they will deliver. A 9 V battery will deliver between 7 V and 10 V, and the actual amount will vary across its life. Additionally, the connectors available from batteries usually don't have the solid jumper wire ends, so it is hard to connect them to breadboards.

Both of these problems can be solved with a power regulator module. A power regulator module will take power from other sources (such as a battery) and reduce the voltage to a value which, while lower, will be constant. The power regulator we will use is the YwRobot breadboard power supply.

This low-cost device will take an input between 6.5 and 12 volts and deliver a constant voltage at the output of either 3.3 or 5 volts (we will use the 5 V setting). This device can be connected to a battery with an appropriate battery clip.[2] Figure 6-13 shows what this looks like.

---

[2] This can also be connected to a wall outlet using an appropriate adapter or connected to your laptop for a USB power supply. For the purposes of this book, we recommend against both of these options as it adds additional risk if something is wired incorrectly.

As you can see, the power module clips onto the power rails of the breadboard. The most important thing is to make sure that the power module is oriented correctly so that the positive and negative markings on the power module line up with the positive and negative power rails. Additionally, this power module has a shunt which, depending on where it is placed, sets whether it outputs 3.3 V or 5 V or is switched off. You need to be sure that the shunts on both sides are set to the 5 V position.

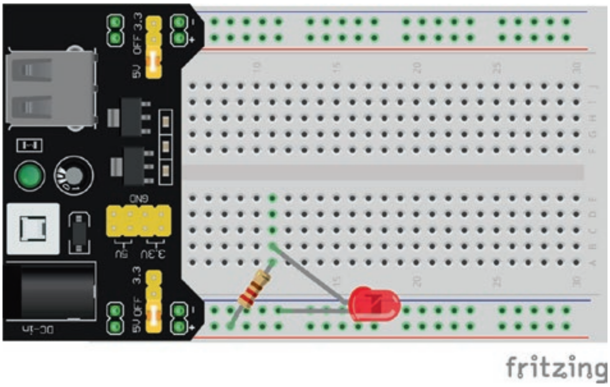

Figure 6-13. *Breadboard with a Power Module*

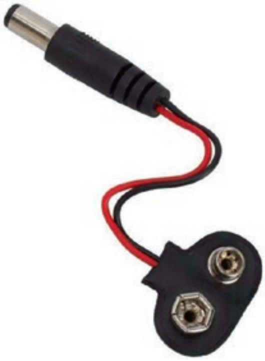

Figure 6-14. *9 V Battery Clip*

Finally, you need to plug the battery in. A 9 V battery can be plugged into the power module using a battery clip designed for CCTV cameras. They have a 9 V clip on one side and a barrel plug on the other side, which plugs nicely into the power module. You can see one in Figure 6-14.

After that is set up, don't forget to turn it on! The power module has an LED that lights up when the power module is turned on. This is really convenient because you don't have to include an on/off switch or an on light on your project, because the power module has one for you.

So how does the power module show up in the schematic? Well, actually, it doesn't directly. The combination of battery and power module basically just yields a 5 V battery in the schematic. So we will just represent it with the normal battery symbol, but set to 5 volts.

There are other power modules available as well, and they all basically work the same way. Just be sure that the voltage is set to 5 volts.

Review

In this chapter, we learned the following:

1. Solderless breadboards can be used to quickly create circuits.
2. Solderless breadboards allow circuits to be easily constructed and destructed in such a way that the components are reusable from one project to the next.
3. Both wires and the legs of a component are attached to connection points on the breadboard.
4. Connection points in the same terminal strip are connected by a wire behind the breadboard.

5. To connect two components together, all you have to do is put their legs on the same terminal strip of the breadboard.
6. The power rails on a breadboard extend either all the way down the board or sometimes split at the halfway point.
7. The bridge of a breadboard divides and separates different groups of terminal strips. This allows a chip to be placed over the bridge, allowing each of its pins a separate terminal strip.
8. The schematic drawing of a circuit can be assembled onto a breadboard, giving a definite implementation of the drawing.
9. There are multiple different ways to place a given circuit drawing onto a breadboard.
10. Components on a breadboard can be connected by wires, or they can be connected by placing their legs in the same terminal strip.
11. There are many different styles of placing components onto breadboards, which have tradeoffs between how easy it is to reconfigure and how clean the result is.
12. A multimeter allows you to measure several important values on a circuit, including resistance, voltage, and current.

13. If your multimeter is not auto-ranging, you must test your value several times, starting with the highest range setting for the value you are looking for and decreasing it through the settings until you find a precise value. For resistance values, this is reversed—we start with the lowest range setting and move upward.
14. Always be sure your multimeter is set to the right setting *before* measuring.
15. Always turn your circuit off before measuring resistance.
16. Your circuit must be on to measure voltage or current.
17. Voltage is measured by connecting your multimeter to empty connection points in the terminal strips that you want to measure. This can be done either by putting your multimeter leads directly into the relevant connection points or by running wires from those connection points to your multimeter leads.
18. Current is measured by using your multimeter to replace a wire that you want to measure current running through.
19. Many circuit values vary much more than what you might think, so it is good to design circuits in a way that will handle these variances.
20. Power modules can deliver consistent voltages even when the components feeding into them (such as batteries) may have a significant amount of variance.

Apply What You Have Learned

All measured values should be measured using the ranging technique discussed in this chapter:

1. Start with the circuit you built in Figure 6-8. Measure the voltage drop across the resistor, and then measure the voltage drop across the LED. Now, measure the voltage drop across both of them (put the red multimeter lead on the left side of the resistor and the black multimeter lead on the right side of the LED). Write down your values.
2. Using the same circuit, change the LED from red to blue. Measure the values again and write them down. Measure the current going through the circuit using any wire. Is it the same or different than before?
3. Add another LED in series with the one you have already. Measure the voltage drops on each side of each component in the circuit. Measure the current going through any given wire. Write down each value.
4. Take the new circuit you built in the previous problem and draw the schematic for the circuit.

CHAPTER 7

Analyzing Series and Parallel Circuits

In Chapter 5, we looked at our very first circuit and how to draw it using a circuit diagram. In this chapter, we are going to look at different ways components can be hooked together and what they mean for your circuit.

7.1 Series Circuits

The circuit built in Chapter 5 is considered a **series circuit** because all of the components are connected end to end, one after another. In a series circuit, there is only one pathway for the current to flow, making analyzing the circuit fairly simple.

It does not matter how *many* components are connected together—as long as all of the components are connected one after another, the circuit is considered a series circuit. Figure 7-1 shows a series circuit with several components included.

© Jonathan Bartlett 2020

J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_7

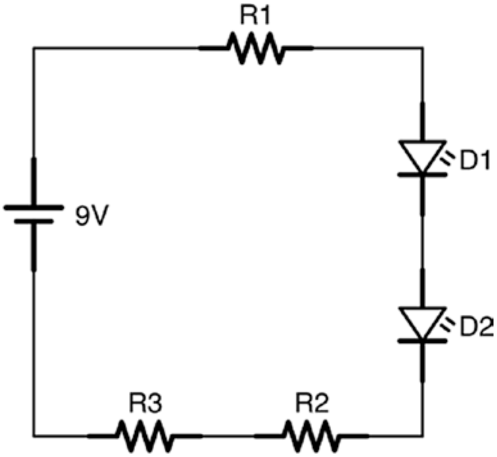

Figure 7-1. *A Series Circuit with Several Components*

If all of the components are in a series, then, even if there are multiple resistors scattered throughout the circuit, you can figure out the total resistance of the circuit just by adding together all of the resistances. This is known as the **equivalent resistance** of the series.

In this example (Figure 7-1), if R1 is 100 Ω, R2 is 350 Ω, and R3 is 225 Ω, then the total series resistance of the circuit will be 100 + 350 + 225 = 675 Ω.

That means that the current is easy to figure out as well. If we ignore the LEDs (since we have not yet learned to calculate using them), then we can use the total series resistance to calculate current the same way we did with the single resistor.

Since the voltage is 9 volts, then we can use Ohm's law to find out the current going through the system.

$$I = V / R = 9 / 675 = 0.013\text{A}$$

Note that A stands for ampere, and we will be using this in our calculations from here on out. However, we usually measure in milliamps (abbreviated as mA), so let us convert:

$$0.013 * 1000 = 13\text{mA}$$

So our circuit will draw about 13 milliamps of current. This amount of current is the same amount running through all of the components in the series.

7.2 Parallel Circuits

Circuits are wired into a **parallel circuit** if one or more of their components are arranged into multiple branches.

Figure 7-2 shows a simple circuit with two resistors in parallel. In this figure, the circuit has *two* branches. R1 is in the first branch, and R2 is in the second branch. The place where the branch occurs is called a **junction**, and is usually marked with a dot to show that all the wires there are connected.

In a parallel circuit, electricity will flow through both branches simultaneously. Some of the current will go through R1 and some of it will go through R2. This makes determining the total amount of current more difficult, as we have to take into account more than one branch.

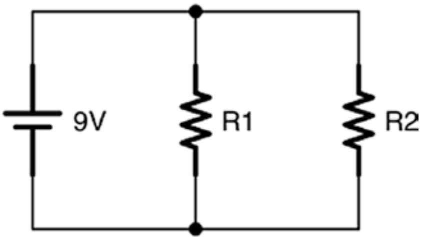

Figure 7-2. *Two Resistors Wired in Parallel*

However, there are two additional laws we can use to help us out, known as **Kirchhoff's circuit laws**. The guy's name is hard to spell, but his rules are actually fairly easy to understand.

7.2.1 Kirchhoff's Current Law

The first law is known as **Kirchhoff's current law**. Kirchhoff's current law states that, at any junction, the total amount of current going *into* a junction is exactly the same as the total amount of current going *out* of a junction. This should make sense to us. Think about traffic at a four-way intersection. The same number of cars that enter that intersection must be the same number of cars that leave the intersection. We can't create cars out of thin air; therefore, each car leaving must have come in. Cars don't magically disappear; therefore, each car entering must leave at some point. Therefore, Kirchhoff's circuit law says that if you add up all of the traffic going in, it will equal the amount going out.

ADVANCED: ANOTHER WAY OF LOOKING AT IT

Another way to say this is that the total amount of all of the currents at a junction is zero. That is, if we consider currents coming in to the junction to be positive and currents going out of the junction to be negative, then their total will be zero since the size of the currents coming in must equal the size of the currents going out.

So let's look at a junction. Figure 7-3 shows a junction where one wire is bringing current in and it branches with two wires bringing current out. The first wire going out has 0.75 A of current, and the second wire going out has 0.34 A of current. How much current is going into the junction from the left?

Since the total coming in must equal the total coming out, then that means the total coming in must be

$$0.75\text{ A} + 0.34\text{ A} = 1.09\text{ A}$$

Therefore, the total amount of current coming into the circuit is 1.09 A.

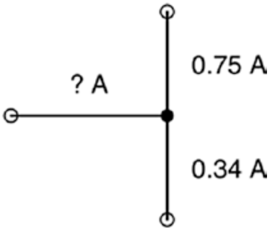

Figure 7-3. *A Simple Junction*

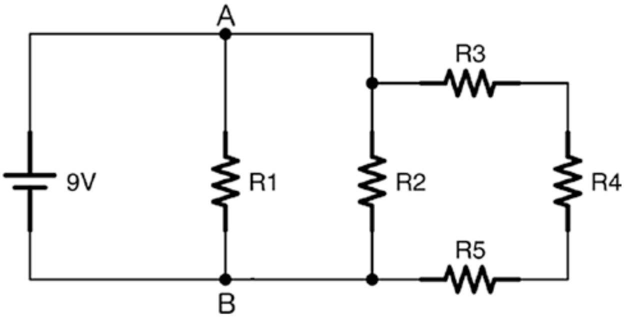

Figure 7-4. *A Circuit with Many Parallel Paths*

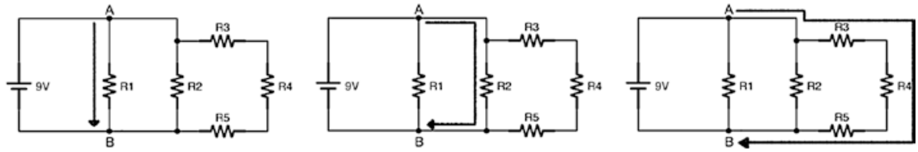

Figure 7-5. *All Paths Between Two Points Have the Same Voltage Drop*

Now, let's say we had a junction of four wires. On the first wire, we have 0.23 A of current coming in. On the second wire, we have 0.15 A of current going out. On the third wire, we have 0.20 A of current going out. What must be happening on the fourth wire? Is current coming in or going out on that wire?

To figure that out, we have to look at the totals so far. Coming in, we have the one wire at 0.23 A. Going out, we have the two wires for a total of 0.15 A + 0.20 A = 0.35 A. Since we only have 0.23 A coming in, but there is 0.35 A going out, that means that the fourth wire must be bringing current in. Therefore, the amount that this fourth wire must be bringing in is 0.35 A – 0.23 A = 0.12 A.

7.2.2 Kirchhoff's Voltage Law

Kirchhoff's current law makes a lot of sense, because the amount of "stuff" coming in is the same as the amount of "stuff" going out. This is similar to our everyday experience. Kirchhoff's voltage law, however, is a bit more tricky. **Kirchhoff's voltage law** states that, given any two specific points on a circuit at a particular time, the difference in voltage between the two points (known as the **voltage drop**) is the same *no matter what pathway you take to get there*.

Figures 7-4 and 7-5 illustrate this point. If we wanted to measure the voltage drop between the two points indicated (A and B), then that voltage drop, at least at a particular point in time, would be the same no matter what pathway electricity travels. The direct route between the two points has the same voltage drop as the more winding pathways, no matter what the values of the resistors are.

So how does that square with Ohm's law?

The way it works is that Ohm's law will cause all of the *currents* through each part of the circuit to adjust in order to make sure that the *voltage* stays the same.

As you can see, the voltage drop between A and B *must* be 9 volts because the battery is a 9-volt battery, and there are no components (only wires) between the battery terminals and A and B. Since batteries always have a constant voltage between their terminals, that means that A and B will have the same voltage—9 volts.

Therefore, that means that the voltage drop across R1 is 9 volts, because it is one of the pathways between A and B, and all pathways get the same voltage. Let's put in some real values for these resistors and see if we can figure out how much voltage and current is happening in each part of the circuit. Let's set R1 = 1,000 Ω, R2 = 500 Ω, R3 = 300 Ω, R4 = 400 Ω, and R5 = 800 Ω. Now, let's find out what our circuit looks like.

As we have noted, *every* path must have the same voltage drop—9 volts. So let's start with the easiest one, the current going across R1. Since we have a 9-volt drop and 1,000 Ω, we can just use Ohm's law for current:

$$I = V / R = 9\text{V} / 1{,}000\Omega = 0.009\text{A}$$

So we have 0.009 A running across R1.

Now, what about R2? R2 is connected to point A simply by a wire. As we mentioned in Chapter 5, wires can be considered to be zero length. Therefore, R2 is just as much directly connected to point A as R1 is. Therefore, the voltage drop across R2 is also going to be 9 volts. Again, using Ohm's law, we can see that

$$I = V / R = 9\text{V} / 500\Omega = 0.018\text{A}$$

So the current going across R2 is 0.018 A.

What about the current going across R3, R4, and R5? Well, if you notice, those resistors are all in series, so we can add them all up and just use the total resistance.

So the total resistance for this section of the circuit will be

$$R3 + R4 + R5 = 300\Omega + 400\Omega + 800\Omega = 1{,}500\Omega$$

So, using Ohm's law, the current running through this part of the circuit will be

$$I = V / R = 9\text{V} / 1{,}500\Omega = 0.006\text{A}$$

Now, remember that the total current flowing into any junction has to be equal to the current flowing out of it. So let's look at the junction between R2 and R3. We calculated that the current flowing to R2 is 0.018 A and the current flowing to the series starting with R3 is 0.006 A. Therefore, there has to be 0.018 + 0.006 = 0.024 A flowing into that junction.

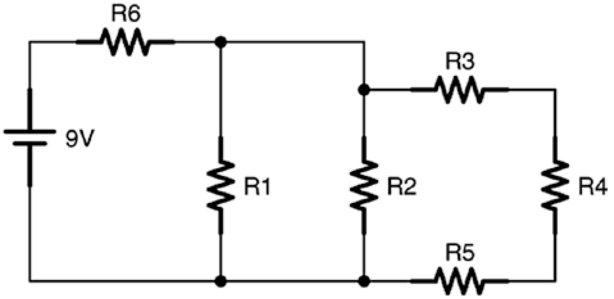

***Figure 7-6.** Kirchhoff's Voltage Law with Series and Parallel Components*

Now, how much current is flowing out of junction A? Well, earlier, we noted that the amount of current flowing across R1 was 0.009 A, and we just calculated that there is 0.024 A flowing out the junction between R2 and R3. That means that there must be 0.033 A total flowing into junction A.

While there were a lot of steps to determine this, each individual step was fairly straightforward. We simply combined Ohm's law, Kirchhoff's voltage law, and Kirchhoff's current law to figure out each step.

Now, one important thing to notice is that there is *less* current running through the pieces of the circuit with more resistance than there is with the pieces of the circuit with less resistance. The electric current is more likely to go down the path of least resistance. This is a very important point and should not be overlooked, as it will come in handy in later chapters.

7.3 Equivalent Parallel Resistance

The sort of calculation that we have done in the previous section gets trickier if there is a series resistance before or after the parallel resistance. Figure 7-6 gives an example of this. The setup is just like the previous circuit, except there is a single resistor (R6) in series with the battery *before* the parallel branches. This will prevent our simple calculations from working because the current flowing in each of the branches of the circuit will all add together to tell us the amount of current flowing through R6. However, the voltage drop across R6 will depend on the current flowing through it. If this voltage changes, then it will change our starting voltage for our calculations to figure out the parallel branches.

Thus, we have ourselves in a loop—in order to find out the current flowing through the parallel branches, we have to know their starting voltage. In order to find out their starting voltage, we have to know how much the voltage dropped across R6. In order to know how much the voltage dropped across R6, we have to know how much current was flowing through it!

This may seem like an impossible problem, but basic algebra allows us to work it out, though the details are kind of ugly. Instead, there is an equation that we can use for parallel resistors which will give us an **equivalent resistance** for a group of parallel resistors. That is, we can take a group of parallel resistors, and we can calculate the total resistance of those resistors, as if they were acting together as one resistor. In other words, we can find out what value we would need for a single resistor to replace all of the other resistances.

If you have resistors in parallel to each other (let's call them R_1, R_2, and R_3) and you want to know the resistance of their *combined* action (which we will call total RT), then you would use the following equation:

$$R_T = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}} \qquad (7.1)$$

This equation works for any number of resistances that we have in parallel. We can just keep on adding them to the end of the list:

$$R_T = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \ldots + \frac{1}{R_N}} \qquad (7.2)$$

So let's look at our circuit and see how we can find out the current flowing through each resistor. For this example, we will again say that $R1 = 1{,}000\ \Omega$, $R2 = 500\ \Omega$, $R3 = 300\ \Omega$, $R4 = 400\ \Omega$, and $R5 = 800\ \Omega$. Additionally, $R6 = 250\ \Omega$.

In order to compute this, we first have to figure out *what* is in series and what is in parallel. Notice the loop made by R3, R4, and R5. Those are all connected end to end, so they are in series. Because they are in series, we can get their equivalent resistance just by adding them together—300 + 400 + 800 = 1,500 Ω. Therefore, we can actually *replace* these resistors with a single, 1, 500 Ω resistor. We will call this "combined" resistor R7. Now, if you look at the new picture, with R7 standing in for the loop, you will see that R1, R2, and R7 are in parallel with each other.

Therefore, we can find out their combined resistance by using Equation 7.2:

$$R_T = \frac{1}{\frac{1}{R1} + \frac{1}{R2} + \frac{1}{R7}}$$

$$R_T = \frac{1}{\frac{1}{1{,}000} + \frac{1}{500} + \frac{1}{1500}}$$

$$R_T = \frac{1}{0.001 + 0.002 + 0.00067}$$

$$R_T = \frac{1}{0.00367}$$

$$R_T = 272.5\Omega$$

Therefore, the equivalent resistance of all of the parallel resistances is about 272.5 Ω, which means that we could in theory replace *all* of these resistors (R1, R2, R3, R4, and R5) with a single resistor that is 272.5 Ω. Also notice that this resistance is actually *less* than each of the individual resistances.

Now, to get the total resistance of the circuit, we notice that this parallel resistance (272.5 Ω) is in series with R6, which is 250 Ω. Since they are in series with each other, we can simply add them together. The total resistance of this circuit is 250 + 272.5 = 522.5 Ω. We can now use Ohm's law to find the total amount of current running through this circuit:

$$I = \frac{V}{R}$$

$$I = \frac{9}{522.5}$$

$$I = 0.0172\text{A}$$

Thus, the whole circuit has 0.0172 ampere of current running through it. Using this, we can now go back through and identify how much current and voltage is flowing through each individual piece.

Because the entirety of the 0.0172 ampere is going through the first resistor (R6), that means that the voltage drop of the R6 resistor will be, using Ohm's law

$$V = I \cdot R$$

$$V = 0.0172 \cdot 250$$

$$V = 4.3\text{V}$$

That means that this resistor will chew up 4.3 V. This leaves us with 9 – 4.3 = 4.7 V after the series resistor.

We now know the starting and ending voltages of each branch of the parallel resistors—4.7 V at the beginning (what we just calculated the voltage to be after the series resistor) and 0 V at the end (because it connects to the negative terminal of the battery, which we have designated as the 0-volt reference).

Therefore, we can use Ohm's law to find the amount of current flowing through each of them. For R1:

$$I = \frac{V}{R}$$
$$I = \frac{4.7}{1{,}000}$$
$$I = 0.0047\,\text{A}$$

For R2:

$$I = \frac{V}{R}$$
$$I = \frac{4.7}{500}$$
$$I = 0.0094\,\text{A}$$

And finally, for the series that is in a loop at the right (R3, R4, and R5):

$$I = \frac{V}{R}$$
$$I = \frac{4.7}{1500}$$
$$I = 0.0031\,\text{A}$$

Since the loop is all in series, that means all of the resistors in that series will have 0.0031 A going through them.

If we add all of these currents, we will see that 0.0031 + 0.0094 + 0.0047 = 0.0172 A, which is the amount of current we originally figured out.

What we have learned is that we can replace the entire circuit with a single value for its resistance to figure out how the circuit will behave as a whole. For a simple circuit like this, having all of these parallel branches doesn't do much, so it may seem pointless. However, in a real circuit, each of these branches may be, instead of a resistor, a more complicated component that has some amount of resistance. If you know the resistance, you can calculate how much current is flowing through it the same way.

However, we start with only resistors in order to make the problems simpler.

7.4 Wires in a Circuit

In complicated circuits, sometimes we run out of room and must draw wires on top of each other even though the wires aren't connected. In this book, we try to make clear which wires are connected by placing a dot on the junction point. Another way to show this is to show the wires crossing if the wires are joined but showing one of the wires as being broken across the intersection point if the wires are not joined. Figure 7-7 demonstrates the difference. The wires on the left are joined together as indicated by the dot. The wires on the right are not joined in any way; they just had to be drawn across each other because of space reasons in the diagram. In this book, the convention we follow is to use dots to indicate joined wires, but we do not break lines for unjoined wires.

Also, the lengths of wires that we draw are irrelevant. Usually, in simple circuits, we should consider that wires are all zero length. If, after a resistor, the voltage in the circuit has dropped to 5 V, then we can consider that the *whole wire* until the next circuit is at 5 V. If a wire branches into multiple branches, even though each branch will have a different amount of *current* running on it, each branch of the wire will all have the *exact same voltage* until they reach another component.

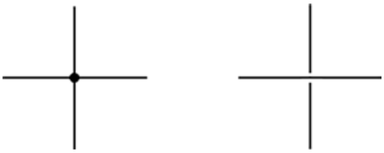

Figure 7-7. *Showin Joined Wires (Left) vs. Showing Unjoined Wires (Right)*

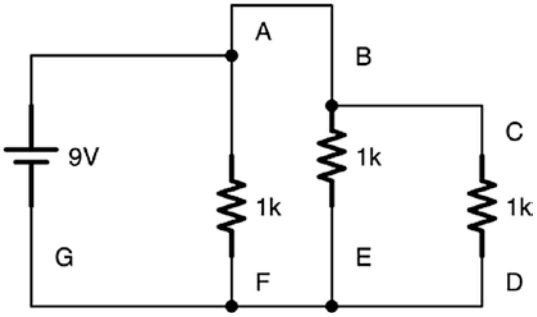

Figure 7-8. *Several Points on a Circuit*

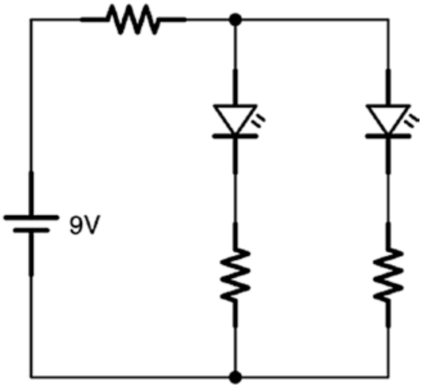

Figure 7-9. *A Circuit with Series and Parallel Components*

Therefore, in the circuit in Figure 7-8, you can see several points labeled A, B, C, D, E, F, and G. In this circuit, A, B, and C all have equivalent voltages (though not equivalent currents) since there are only wires (and not components) between them. Likewise, D, E, F, and G all have equivalent voltages since there are only wires between them. Also, since D, E, F, and G are all connected to the battery negative (i.e., ground) with no components between them, that means that they are all at 0 volt. Likewise, since A, B, and C are all directly connected to the battery positive with no intervening components, they are all at 9 volts.

7.5 Wiring Parallel Circuits onto a Breadboard

One other issue we need to look at is how to wire parallel circuits onto the breadboard. It is actually very simple to do. In this section, we are going to put the circuit in Figure 7-9 onto a breadboard.

Notice that, in this circuit, there is a series resistor at the beginning and then two parallel circuits that branch off from it. It doesn't matter much what the value of the resistors are, but we will put them at 1,000 Ω if you need a specific value (anything between 200 and 2,000 Ω should work fine).

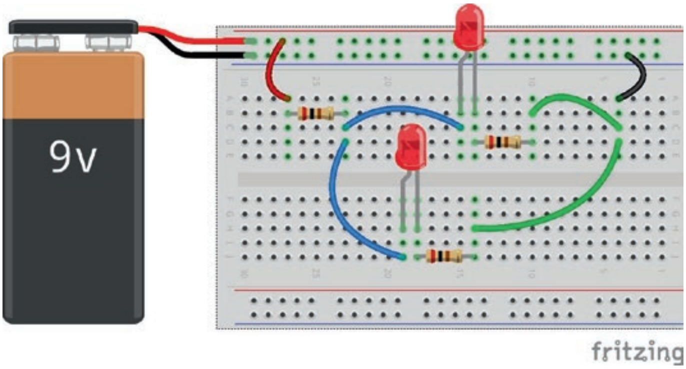

Figure 7-10. *A Circuit with Series and Parallel Components on the Breadboard*

To get the circuit onto the breadboard, remember that anything that is connected in the same terminal strip on a breadboard is connected together. This means that, when we have a parallel subcircuit, we connect all of the branches of the subcircuit into the same terminal strip. Figure 7-10 shows what this looks like.

Let's follow the path of electricity through the breadboard. First, the current flows from the positive terminal to the positive rail on the breadboard. A wire then pulls the positive +9 V power onto the board. This wire is connected to a resistor by putting one leg of the resistor in the same terminal strip as the wire. On the other leg of the resistor, there are *two* wires that are in the same terminal strip. Each of these goes to a different part of the circuit. We have one LED with a resistor on the top and one LED with a resistor on the bottom. These are just normally connected components.

However, after the resistor, the two subcircuits come back together to a terminal strip on the right. Then, a wire on the same terminal strip takes that back to the negative rail on the breadboard (which connects to the negative terminal on the battery).

Note that I could have used a lot fewer wires to accomplish the same circuit. However, I thought that using more wires would make it more clear what is happening at each junction, especially when the circuit gets divided into branches or comes back together.

Take a moment to look at both the schematic drawing and the breadboard picture, and be sure that you can trace the flow of the schematic on the actual breadboard.

Review

In this chapter, we learned the following:

1. In a series circuit, electricity flows in a single line through all of the components.
2. In a parallel circuit, electricity branches and flows in multiple branches.

3. Most real circuits are combinations of series and parallel circuits.
4. When you have resistors together in series, the total resistance of all of the resistors combined is simply the sum of their individual resistances. $R_T = R_1 + R_2 + \ldots + R_N$.
5. In a parallel circuit, Kirchhoff's current law says that the total amount of current entering a branch/junction is the same as the total amount of current leaving the branch.
6. In a parallel circuit, Kirchhoff's voltage law says that, between any two points on a circuit at a given point in time, the voltage difference between those two points will be identical no matter what pathway the electricity follows to get there.
7. When resistances are in parallel, the total resistance for the parallel circuit is given by the equation

$$R_T = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \ldots + \frac{1}{R_N}}$$

8. By using these laws in combination, we can predict how current will flow in each part of our circuit.
9. Series circuits are placed onto a breadboard by putting the connected legs of two connected components onto the same terminal strip.
10. Parallel circuits are placed onto a breadboard by connecting each subcircuit branch to the same terminal strip.

Apply What You Have Learned

1. There is a junction in a circuit that has one wire with current flowing in and two wires with current flowing out. There is 1.25 A of current coming in, and the first wire going out has 0.15 A of current. How much current is leaving through the second wire?
2. There is a junction in a circuit that has two wires with current flowing in and two wires with current flowing out. The first wire with current flowing in has 0.35 A of current, the first wire with current flowing out has 0.25 A of current, and the second wire with current flowing out has 0.42 A of current. How much current is flowing in on the second incoming wire?
3. At a junction of four wires, wire 1 has 0.1 A of current flowing in, wire 2 has 0.2 A of current flowing in, and wire 3 has 0.4 A of current flowing out. Is the current in wire 4 going in or out? How much current is flowing on it?
4. If I have three 100 Ω resistors in series, what is the total resistance of the series?
5. If I have a 10 Ω resistor, a 30 Ω resistor, and a 65 Ω resistor in series, what is the total resistance of the series?
6. If I have a 5 Ω resistor and a 7 Ω resistor in series, what is the total resistance of the series?

7. If I have two resistors in parallel, a 30 Ω resistor and a 40 Ω resistor, what is the total resistance of this circuit?

8. If I have three resistors in parallel—25 Ω, 40 Ω, and 75 Ω—what is the total resistance of this circuit?

9. If I have four resistors in parallel—1,000 Ω, 800 Ω, 2,000 Ω, and 5,000 Ω—what is the total resistance of this circuit?

10. If I have three resistors in parallel—100 Ω, 5,000 Ω, and 10,000 Ω—what is the total resistance of this circuit? Which of the resistors is the total resistance most similar to?

11. Take a look at the following circuit diagram. If the voltage drop between B and C is 2 volts and the voltage drop between C and D is 3 volts, what is the voltage drop between A and E? What is the voltage at E? What is the voltage at A?

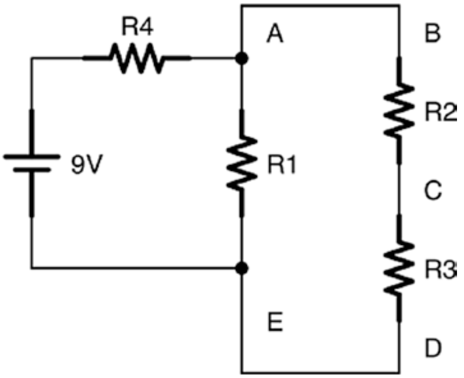

12. If the preceding circuit runs with 2 A total current, what is the value of the resistor R4?

13. The following circuit is a combination of series and parallel resistances. Each resistor is labeled with its resistance value, given in ohms. Find out how much current is flowing through each resistor and how much each resistor drops the voltage.

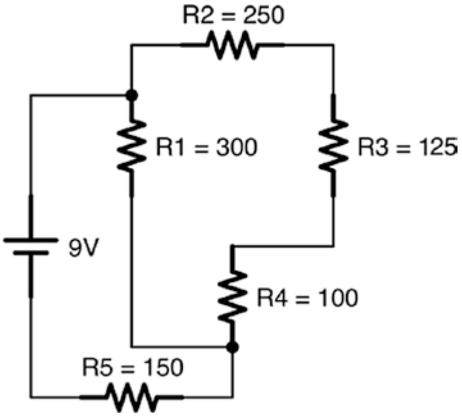

14. Build the circuit in Figure 7-10 on your own breadboard. Measure the voltage drops across every component, and measure the amount of current flowing into the first series resistor.

CHAPTER 8

Diodes and How to Use Them

This chapter introduces the **diode**. We have used light-emitting diodes (LEDs) in previous chapters, but have not really discussed their function except for emitting light. In this chapter, we are going to look at regular diodes, light-emitting diodes, and Zener diodes, to get a feel for what these devices are and how they might be used in circuits for more than just light.

8.1 Basic Diode Behavior

Unlike resistors, diodes have both a positive and negative side. For any component with both positive and negative legs, the positive leg of the component is termed the **anode** and the negative leg is termed the **cathode**. On LEDs, the anode is longer than the cathode. However, for other types of diodes, the cathode is marked with a line. You can remember this because in the schematic of a diode, the cathode has the blocking line.

The diode performs two fundamental "actions" with electric current. The first action of a diode is to drop voltage by an essentially fixed amount *without* affecting or limiting current. This amount of voltage is called the **forward voltage drop** and it is usually around 0.6 V for most non-LED diodes. Forward voltage is often abbreviated as V_F. For most LEDs, the

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_8

forward voltage drop depends on the color of the LED, with a red LED dropping about 1.8 V and a blue LED dropping about 3.3 V. The forward voltage drop can vary among the other colors as well.

To remind you, when we say “voltage drop,” we are referring to the difference in voltage between the positive and negative legs of the diode. Therefore, no matter what the voltage is coming into the diode, the voltage coming out of the diode will be that voltage minus the voltage drop.

The second action of a diode is to limit the direction of current to a single direction. With some exception, the diode only allows current to flow one way in the circuit. The “normal flow” of a current through the diode is to flow from the anode to the cathode. Looking at the schematic symbol, current flows in the direction that the arrow is pointing, and current is blocked from flowing the other way (you can think of the line as a “block” preventing reverse flow).

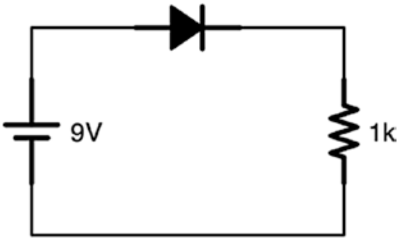

Figure 8-1. *A Diode and Single Resistor Circuit*

However, diodes are limited in the amount of reverse flow they can block. At a certain point, diodes reach their **breakdown voltage**. The breakdown voltage is the voltage at which they will stop blocking voltage. In regular diodes, this is a failure mode, and the precise value shouldn’t be relied upon (we will see an exception to this with Zener diodes). Usually, though, this value is high enough not to worry about (i.e., around 100 volts).

8.2 Circuit Calculations with Diodes in Series

Now let's talk about how to properly calculate the behavior of circuits with diodes. Remember, the key to a diode is that if current is flowing through the diode, the voltage drop across the diode will be essentially constant. For a normal non-LED diode, this voltage drop is almost always 0.6 V. This is so common it is usually assumed, and never listed in the circuit itself.

Therefore, take a look at the circuit in Figure 8-1. Since the voltage source is 9 volts, that means that the total voltage drop between the positive and negative is 9 volts. The diode will eat 0.6 V of the voltage, but will not limit the current in any way. The resistor, then, since it is the only component left, will use up the rest of the voltage—8.4 V. Therefore, we can calculate the current in the circuit using Ohm's law:

$$I = V / R = 8.4 / 1000 = 0.0084\text{A} = 8.4\text{mA}$$

Therefore, our circuit will use 8.4 mA of current. So, as you can see, when doing calculations, the diodes simply provide a drop in voltage, they do not limit (or even affect) current.

This is true no matter how many diodes or resistors I have in my circuit. Figure 8-2 shows a circuit like that. To figure out the behavior of this circuit, remember that *every* diode has a 0.6 V voltage drop. Since the circuit has three diodes, that means that the diodes will drop a total of $0.6 * 3 = 1.8$ V.

Therefore, we will have 1.8 V taken up by diodes, which drop voltage without limiting current. Since we have a 9 V source, there will be $9 - 1.8 = 7.2$ V that is not eaten by diodes. This voltage will go to our two resistors. These resistors, even though they are separated by diodes, are essentially in series with each other. Therefore, we can treat them as a single resistor in series. So the total resistance on the circuit will be $1k + 2k = 3k\ \Omega$.

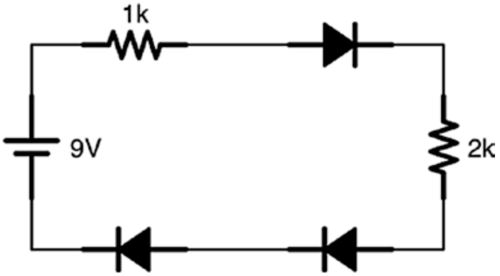

***Figure 8-2.** A Circuit with Multiple Diodes and Resistors*

We can then find the total current running through the circuit using Ohm's law:

$$I = V / R = 7.2\text{V} / 3{,}000\Omega = 0.0024\text{A} = 2.4\text{mA}$$

However, it is possible to put too many diodes in your circuit. Since they each eat 0.6 V in their forward voltage drop, that puts a limit on how many you can string together in series from a given battery. For a 9 V battery, if I try to put 20 diodes together in series, I will have used *more* than the total 9 volts that I have available. Therefore, current will not flow. With 20 diodes, the voltage drop will be 20 $*$ 0.6 = 12 V. Since this is more voltage than the battery can put out, no current will flow.

So we have seen two conditions for which current will not flow through a diode—the first is that the diode will block current from flowing in the wrong direction, and the second is that the diode will not conduct if the voltage source cannot provide enough voltage to bridge the forward voltage drop of the diode.

8.3 Circuit Calculations with Diodes in Parallel

The real magic of diodes comes when using them in parallel circuits. Remember the rules that we learned in Chapter 7. Kirchhoff's voltage law says that between any two points, the voltage drop between those two points will be the same *no matter what path the current follows*. Therefore, since the voltage drop across a diode is *fixed*, that means that we can guarantee a maximum voltage drop between two points on a circuit by putting diodes between them.

Figure 8-3 shows what this looks like. The voltage drop from one side of the diode to the other is 0.6 V. Period. End of story (actually, it could be less, which would keep the diode from conducting altogether, but we won't consider that at the moment).

Kirchhoff's voltage law tells us that *no matter what path is traveled*, the voltage difference between those two points will be the same. Since the 2k resistor is attached to the same two points that the diodes are attached to, Kirchhoff's voltage law tells us that the voltage across the resistor *must* be the same as the voltage across the diode. Thus, the voltage across the resistor *must* be 0.6 V.

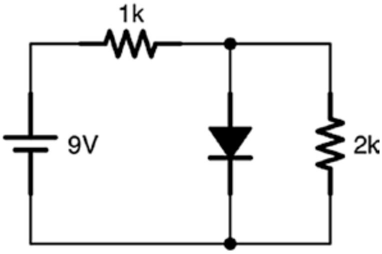

Figure 8-3. *A Single Diode in Parallel with a Resistor*

Using Ohm’s law, we can deduce the amount of current flowing through that resistor:

$$I = V / R = 0.6 / 2{,}000 = 0.0003\text{A} = 0.3\text{mA}$$

So how much current is flowing through the diode? To find this out, we need to use Kirchhoff’s current law. The amount entering the junction where the diode and the 2k resistor split off is the same as the amount leaving it. We know that 0.3 mA leaves the junction to go to the resistor. Therefore, if we could figure out how much current was coming *into* the junction, we could figure out how much is going through the diode.

To discover that value, we need to know how much current is going through the 1k resistor. To figure that out, we need to know the voltage drop across the resistor. However, we can figure this out easily. Since the tail end of the diode is connected to ground (the negative terminal), that means that after the diode, we are at 0 V. Therefore, since the diode voltage drop is 0.6 V, then the voltage before the diode must be 0.6 V. That means that the rest of the voltage must have been consumed by the 1k resistor.

Since the voltage source is 9 V, that means that the voltage for the entire circuit from positive to negative is 9 V, and therefore, the voltage drop across the resistor must have been 9 – 0.6 = 8.4 V. Using this value, we can determine the current going through the resistor using Ohm’s law:

$$I = V / R = 8.4 / 1{,}000 = 0.0084\text{A} = 8.4\text{mA}$$

In this circuit, there is 8.4 mA going through the first resistor. That means that there is 8.4 mA coming into the junction. We know that 0.3 mA is going out of the junction through the 2k resistor. That means that the rest of the current is going through the diode. Therefore, we can calculate that the current going through the diode is 8.4 – 0.3 = 8.1 mA.

Diodes don’t make the math harder, but they do force you to *think* a little harder about how you apply the rules.

Let's take a look at a slightly harder example, using Figure 8-4. In this circuit, we have two diodes in parallel with two resistors in parallel. This parallel circuit is in series with a resistor on the front and another diode at the end.

It is almost always easiest to analyze circuits starting with the diodes because their voltage drops are fixed. If we look at this circuit, the diode at the end gives the circuit a voltage drop of 0.6 V.

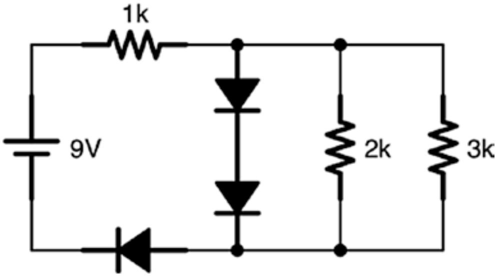

Figure 8-4. *A Circuit with Multiple Diodes*

Now let's look at the parallel part of the circuit. Here, we have three pathways—one through two diodes and two other pathways through resistors. However, one of the pathways contains all diodes. We know that diodes give a constant voltage drop, and we know that Kirchhoff's voltage law says that all pathways between two points have the same voltage drop. Since we have two diodes, the voltage drop of this parallel pathway is 0.6 ∗ 2 = 1.2 V.

Now we have the parallel resistors to worry about. However, we can use the formula for parallel resistance (Equation 7.2) to figure out the total resistance of the parallel resistors.

$$R_T = \frac{1}{\frac{1}{R1} + \frac{1}{R2}} = \frac{1}{\frac{1}{2,000} + \frac{1}{3,000}} \approx \frac{1}{0.0005 + 0.0003333} = \frac{1}{0.0008333} \approx 1200\Omega$$

However, we technically don't even need that number, because, since we already know the voltage drop across each resistor (it *must* be 1.2 V because of Kirchhoff's voltage law), we can just apply Ohm's law to each resistor.

Now, using Ohm's law, we can calculate the current flowing through the resistors. So, for the 2k resistor, we have

$$I = V / R = 1.2\text{V} / 2{,}000\Omega = 0.0006\text{A} = 0.6\text{mA}$$

For the 3k resistor, we have

$$I = V / R = 1.2\text{V} / 3{,}000\Omega = 0.0004\text{A} = 0.4\text{mA}$$

The total current going into both resistors is simply the sum of each of the currents, 0.4 + 0.6 = 1.0 mA. So there is 1 mA total flowing through the two resistors. To find out how much current is flowing through the diode, we will need to know how much current is coming into the circuit itself out of the first resistor.

So how much current is flowing through the first resistor? Well, the voltage drops we have calculated so far include a 0.6 V drop at the end and a 1.2 V drop in the middle. That is a total of 1.2 + 0.6 = 1.8 V. Since the battery is 9 V, that means that there is 9 − 1.8 = 7.2 V left to be consumed by the circuit. Therefore, that must be the voltage drop of the first resistor. Using Ohm's law, we find that

$$I = V / R = 7.2\text{V} / 1{,}000\Omega = 0.0072\text{A} = 7.2\text{mA}$$

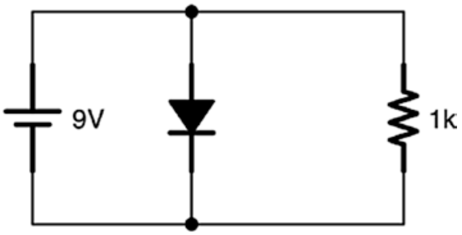

Figure 8-5. *A Bad Diode Circuit*

Therefore, we have 7.2 mA flowing through that first resistor. So, if we have 7.2 mA coming into the parallel circuit in the middle and 1 mA flowing to the resistors, the amount of current flowing through the diodes down the middle is 7.2 – 1 = 6.2 mA. The amount of current going through the final diode is the full 7.2 mA of current in the circuit.

Again, there are a lot of steps, but none of the steps are individually very hard. You simply start with the easiest-to-find values (in this case, the voltage drops from the diodes) and work out from there.

8.4 Diode Short Circuits

Next, let's look at a very bad design with a diode. Let's say that someone wanted to use a diode to regulate the amount of voltage across a resistor to 0.6 V. Therefore, they built the circuit shown in Figure 8-5. Can you figure out what the problem is here?

Well, the voltage drop across the diode will be 0.6 V. However, the battery operates at 9 V. This means that there is 8.4 V left in the circuit with *zero* resistance. Using Ohm's law, that gives us

$$I = V / R = 8.4 / 0 = \infty$$

Thus, having a diode going direct from the positive to the negative of your voltage source is essentially the same as a short circuit. That is why the circuit is bad! To use a diode, there must *always* be resistance somewhere in series with the diode in order to dissipate the current from the excess voltage. The resistor can be before the diode or after it, but it must be there to absorb the extra voltage.

Additionally, it is important to make sure the amount of current flowing through your diode is within the specifications of the diode you are using. Most LEDs, for instance, are only rated for small currents up to about 30 mA.

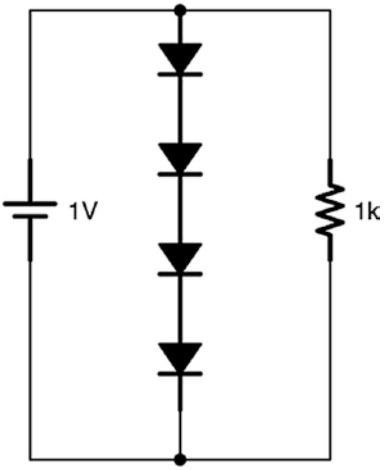

Figure 8-6. *Nonconducting Diodes*

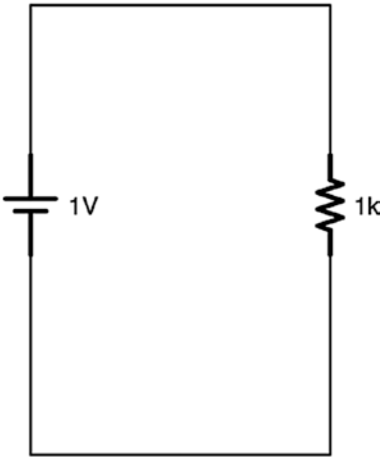

Figure 8-7. *A Circuit Equivalent to Figure 8-6 Because of Nonconducting Diodes*

8.5 Nonconducting Diodes

There is one case where diodes do not maintain a constant voltage drop, and that is where there is not enough voltage to go across them. Figure 8-6 shows an example of this.

In this figure, the voltage drop across the center diode bridge is 0.6 + 0.6 + 0.6 + 0.6 = 2.4 V. However, the battery source is only 1 V. Since the voltage drop of the diodes is larger than the available voltage, this part of the circuit is basically turned off. In reality, there is a small but ignorable leakage current (about 0.00000022 mA in this case if you were really curious), but we can think of this circuit as effectively switched off.

So keep that in mind—if you ever wind up with more forward voltage drop from a diode than you have available voltage, you may treat the diode as if it were an open (i.e., unconnected) circuit, as if it were not even there.

Therefore, we would analyze Figure 8-6 as if it were just like Figure 8-7.

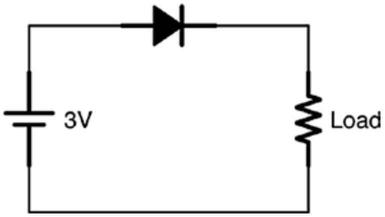

Figure 8-8. *A Simple Diode Protection Circuit*

8.6 Usage of Diodes

Diodes can be thought of as circuit traffic cops—they regulate the flow of electricity. They make sure that everything within a circuit is happening in a controlled manner. They regulate the circuit in two different ways—by limiting the direction of current and, in parallel circuits, by establishing fixed voltages between two points.

The simplest usage of a diode is to use it as a device that makes sure that your battery is plugged in the right orientation. If you have a device that will be damaged if someone puts the battery in backward, a simple diode will make sure that the current can only flow in one direction.

The circuit in Figure 8-8 shows what this looks like. Note that we have a resistor labeled "Load." Many times in circuits, a load resistor is shown to represent whatever else is happening in another part of a circuit. Thus, this circuit shows that the diode is protecting the rest of the circuit (whatever it is) from the user putting the battery in backward. However, this has a cost—the diode will eat 0.6 V in order to provide this protection.

Another problem often solved by diodes is voltage regulation. Because diodes provide a fixed voltage drop between two points, you can use diodes to ensure a fixed voltage for devices that require it.

For instance, when we looked at batteries in Chapter 6, we noted that their voltage actually varied quite a bit. A 9 V battery might give you anywhere from 7 V to 9.7 V. This is true of any battery, not just the 9 V variety. In that chapter, we showed how we could use a voltage regulator to accomplish this, but a similar effect can be accomplished with diodes. If you need a fixed amount of voltage, you can use diodes to provide that at the cost of some extra current.

This isn't recommended for a whole project (the excess current can generate a lot of heat), but sometimes makes sense in limited circumstances. However, for demonstrating the concept, let's look at what it would take to provide voltage regulation using diodes.

Figure 8-9 shows a simple voltage regulator using diodes. Given a 5 V battery (actually, a battery of any size significantly over 3 V will work), this circuit will provide a regulated 3 V of electricity to whatever is connected to it as a load (remember a "load" resistor is just a stand-in for whatever we want to attach to this).

Because the diodes provide a fixed voltage drop of 0.6 V each, then that pathway has a total voltage drop for five diodes of $5 * 0.6 = 3.0$ V. Kirchhoff's voltage law says that between two points, *every* pathway between them will have the same voltage drop. This means that the load, since it is connected to those two points, will have the same voltage drop, thus regulating the amount of voltage used by the load.

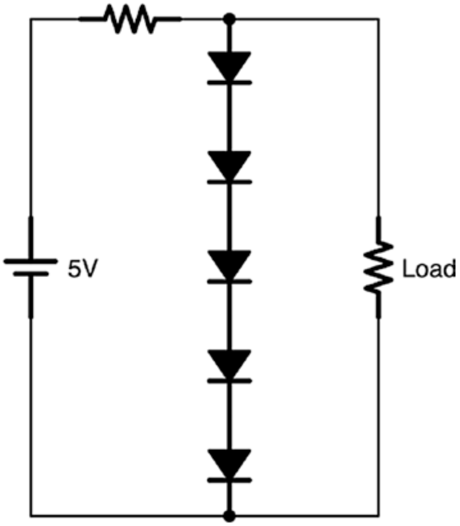

Figure 8-9. *A Simple 3 V Voltage Regulator*

But what about that first resistor at the beginning of the circuit? Remember, if we provide a pathway through diodes only, we create a short circuit in the system. There has to be some element (usually a resistor) that eats up whatever voltage is left over. Therefore, putting a small resistor before our diode pathway will give the circuit a place to use the excess voltage and limit the amount of current that flows.

The size of the resistor will depend on how much current the load requires and how much current you are willing to waste. A small resistor wastes more current, but allows for a higher maximum current used by the load. A larger resistor wastes less current, but if the load needs a lot of current, it could interfere with the voltage regulation.

To see how this could happen, let's say that our load is equivalent to 1,000 Ω. This means that the amount of current that the load will draw can be calculated using Ohm's law:

$$I = V / R = 3\mathrm{V} / 1{,}000\Omega = 0.003\mathrm{A} = 3\mathrm{mA}$$

So the load will use up 3 mA of current. That means that, however much current goes through our first resistor, however much that is over 3 mA will be drained off through the diodes. So let's calculate what that looks like on a tiny, 20 Ω resistor. The voltage will be 5 – 3 = 2 V.

$$I = V / R = 2\mathrm{V} / 20\Omega = 0.1\mathrm{A} = 100\mathrm{mA}$$

So, if we just used a 20 Ω resistor, that means that even though the load only used 3 mA, the whole circuit will be using 100 mA to operate! Therefore, we need a bigger resistor to limit the amount of current we use. Let's try a 500 Ω resistor:

$$I = V / R = 2\mathrm{V} / 500\Omega = 0.004\mathrm{A} = 4\mathrm{mA}$$

This is much better—we are only using 4 mA in this circuit, so we are only wasting 1 mA. There has to be *some* amount of waste to do the voltage regulation.

Let's say that, after a while, our battery sags down to only providing 4 V of power. What happens now? Well, using the 500 Ω initial resistor, that means that there is only 1 V extra to drain off, so the current will be

$$I = V / R = 1\mathrm{V} / 500\Omega = 0.002\mathrm{A} = 2\mathrm{mA}$$

Here, the current is only 2 mA, but we need 3 mA to power the circuit! Thus, if we use a 500 Ω resistor, we can't handle our power supply dropping down to 4 V.

What about using the 20 Ω resistor? In this case, Ohm's law would give us

$$I = V / R = 1\text{V} / 20\Omega = 0.05\text{A} = 50\text{mA}$$

So, here, the 20 Ω resistor will still provide plenty of excess current to allow our regulator to keep working. However, it is still eating up an extraordinary amount of current compared to our load.

So how do you choose the right resistor? The way to work situations like this is to think about what are the maximum cases you are designing for and then calculate appropriately. So, if I want this circuit to work when the battery sags down to 4 V, I need to decide how much excess current I'm willing to have at that level. Let's say I decide that I always want at least half a milliamp to run through the diode (somewhat of an arbitrary number, but if the diode has no power, it is not providing any regulation—this is a low number that is still "noticeable" on the circuit). That means that, since my load will be using 3 mA, the total amount of current running through the initial resistor will be 3.5 mA, or 0.0035 A. Therefore, I calculate what the resistor will need to be at 4 V using Ohm's law:

$$R = V / I = 1\text{V} / 0.0035\text{A} \approx 286\Omega$$

Therefore, for this situation, the initial resistor should be 286 Ω. Now, let's figure out how much current this wastes when the battery is at full charge—5 V (which means that there will be a 2 V drop across this resistor).

$$I = V / R = 2\text{V} / 286\Omega \approx 0.007\text{A} = 7\text{mA}$$

So, at full charge, there is 7 mA going through this resistor, which means we will waste 4 mA. Whether or not that is acceptable to your circuit depends on what you are going to do with it!

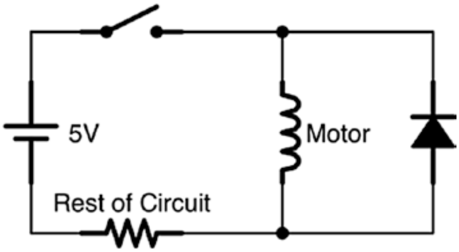

Figure 8-10. *A Diode Protection Circuit*

Note that there are better means of regulating battery voltage for a whole circuit than using diodes. However, many times you need a regulated voltage somewhere *within* a more complex circuit. Diodes are great for that, and the same calculations and considerations apply.

One final note: As mentioned earlier, although we think of diodes as providing a fixed voltage, they actually do vary a little bit with the amount of current flowing through them. In any circuit, you should allow for ±10% variance in the voltage drop of a diode.

8.7 Other Types of Diode Protection

Diodes can provide other types of protection for circuits as well. As single-direction control valves, they can be used to prevent a variety of overvoltage conditions. Oftentimes they are wired in such a way that they will not normally conduct, but will conduct under certain conditions to redirect extra voltage in a safe manner.

Figure 8-10 shows an example circuit. In this circuit, a DC motor is connected to a switch. Notice the diode that is wired backward. Normally, this diode does absolutely nothing because the current is flowing in the other direction, so it just goes through the motor. DC motors, however, tend to generate very large voltages for a short time when switched off (see Chapters 20 and 23 for more information about DC motors and inductive loads). Therefore, when the motor is switched off, the motor can produce a very large voltage—up to 50 V in this circuit!

To protect the rest of the circuit from this sudden influx of voltage, the diode provides an alternative path back through the motor. Thus, when the voltage starts to build up after the switch closes, the diode provides a safe pathway back through the motor for it to flow, allowing the voltage buildup to slowly dissipate through the motor, rather than overloading a circuit expecting 5 V with 50 V.

Oftentimes when looking at circuit diagrams, you may find diodes in funny places and oriented in funny ways. These are oftentimes providing some sort of protection to the circuit from potential failure conditions or exceptional circumstances. Many microchips, for instance, use diodes to shunt off excess voltage from static electricity shocks.

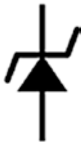

Figure 8-11. *The Zener Diode Schematic Symbol*

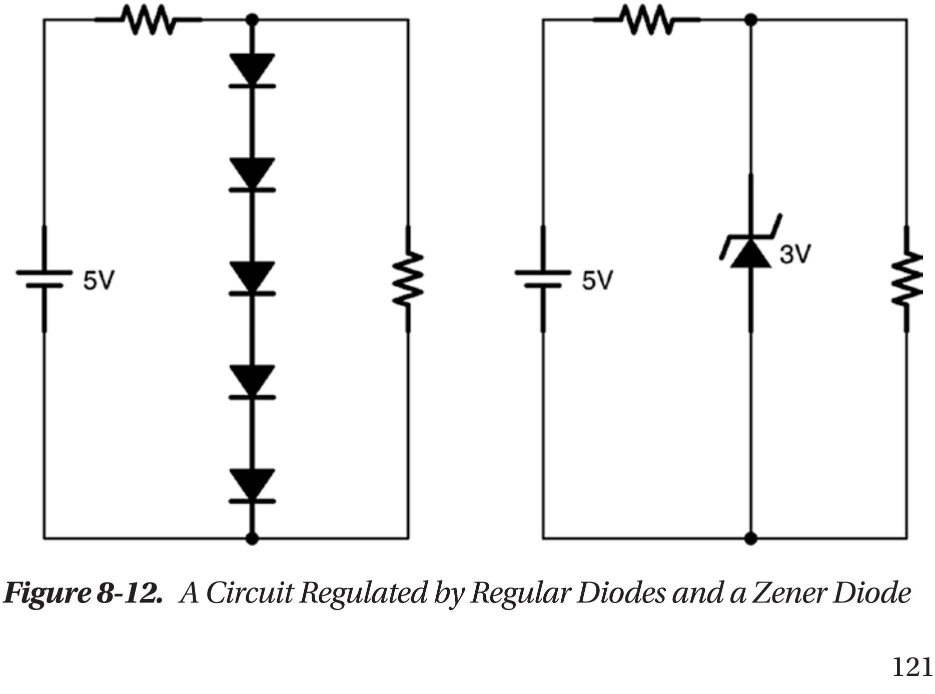

Figure 8-12. *A Circuit Regulated by Regular Diodes and a Zener Diode*

8.8 Zener Diodes

One problem with using diodes for voltage regulation is that their forward voltage drop is pretty small, so you have to have quite a few of them to regulate larger voltages. Zener diodes can help out in these situations. Figure 8-11 shows the symbol for a Zener diode.

Remember that most diodes have a breakdown voltage if you try to pass voltage the wrong way. However, in most diodes, this is a failure mode—doing this can, at best, be unpredictable, and, at worst, damage your diode. A Zener diode, however, is built so that it has a very predictable operation at its breakdown voltage. In fact, at its breakdown voltage, it acts like a normal diode with a larger voltage drop.

However, since you are using the breakdown voltage rather than the forward voltage, Zener diodes are wired into your circuit *backward*. Figure 8-12 shows what this looks like. In this figure, on the left you have the same 3 V regulated circuit as Figure 8-9. On the right, you have an equivalent circuit regulated by a Zener diode instead. Since we are using the Zener diode's breakdown voltage rather than its forward voltage, it has to be wired backward for it to work.

Not only that, the breakdown voltage drop for a Zener diode is much more constant over a larger range of current than the forward voltage drop of most regular diodes. Because of this, Zener diodes are used much more often for voltage regulation than a series of regular diodes.

Zener diodes come in a variety of breakdown voltages. In any exercise in this book which involves drawing a circuit with a Zener diode, you may presume that the Zener diode with the voltage drop you are looking for exists. When drawing a circuit, be sure to label the Zener diode with the breakdown voltage you are needing.

8.9 Schottky Diode

A Schottky diode is very similar to a normal diode, but it is made using a different process, meaning that its voltage drop is significantly reduced. While a “normal” diode’s forward voltage drop is about 0.6 V, a Schottky diode’s voltage drop can be as low as 0.15 V. This can be very useful in cases where a diode is needed to regulate the direction of electrical flow, but the voltage requirements are sensitive enough that you can’t afford the 0.6 V of a regular diode.

8.10 Diode-Like Behavior in Other Components

Another important reason for being familiar with diode operation is that other types of components have parts which contain diodes or behave in diode-like ways. When we discuss bipolar junction transistors (BJTs) (see Chapter 24), we will see that these components partially consist of diode-like behavior.

Review

In this chapter, we learned the following:

1. Diodes only allow current to flow in one direction.
2. Diodes have a fixed forward voltage drop across the diode—0.6 V for normal diodes and a range of about 1.8 to –3.3 V for LEDs.
3. Different color LEDs have different voltage drops.

4. Diodes also have a breakdown voltage—the amount of voltage for which, if applied in the reverse direction, the diode will no longer block current.
5. When analyzing circuits with diodes, it is often easier to analyze the diodes first, since the voltage drop is fixed.
6. Because of Kirchhoff's voltage law, anything wired in parallel with diodes will have the same voltage drop as the diodes (or possibly less).
7. If the forward voltage drop on a diode is greater than the available voltage in the circuit, the diode will not conduct, and it can be treated as an open circuit.
8. If diodes are connected to the positive and negative terminals of a voltage source (such as a battery) with no resistance in series, this will create a short circuit, causing extremely large amounts of current to flow through the diodes.
9. Diodes are often used as control valves to regulate the direction of current flow on a circuit.
10. Diodes are often used to regulate the amount of voltage between two points on a circuit.
11. The series resistor used with voltage-regulating diodes determines both how much current is wasted and how much current the load can draw—lower-value resistors waste more current but allow the load to draw more current, and higher-value resistors waste less current but don't allow as much potential current to your load.

12. When designing circuits, it is often useful to account for the most extreme variations possible. This will allow your circuit to be more flexible.
13. Diodes can also provide protection to circuits against strange failure conditions, such as voltage spikes and static electricity. Diodes in strange places in circuit diagrams are often there to protect the circuit against certain types of failures or events.
14. Zener diodes are built so that they have very reliable operation on their breakdown voltage.
15. Wiring a Zener diode backward gives you the equivalent of several forward diodes in series and can be used for simple voltage regulation.
16. The breakdown voltage of a Zener diode is much more constant than the forward voltage of a regular diode and therefore works even better for voltage regulation.
17. Zener diodes come in a wide variety of breakdown voltages and are usually labeled on the circuit with the necessary breakdown voltage.
18. Schottky diodes are like ordinary diodes except they have a much smaller forward voltage.
19. Understanding diode behavior is important because a number of components contain diodes or exhibit diode-like behavior.

Apply What You Have Learned

1. If you have a 9 V voltage source, a blue LED, and a 500 Ω resistor all in series, how much current is running through the LED?
2. If you have a 3 V voltage source and a red LED, what size resistor do you need to put in series with the LED to have it use 3 mA of current?
3. If you have a 10 V voltage source, a blue LED, a red LED, and a 200 Ω resistor all in series, how much current is running through the LEDs?
4. If I have a 12 V voltage source, a blue LED, and a red LED and the LEDs have a maximum current of 30 mA before they break and a minimum current of 1 mA before they turn on, what range of resistors can I put in series with the LEDs to get them to light up without breaking?
5. In the following circuit, calculate how much current is flowing through each component and each component's voltage drop if R1 is 500 Ω (note that the diode is just a regular diode, not an LED).

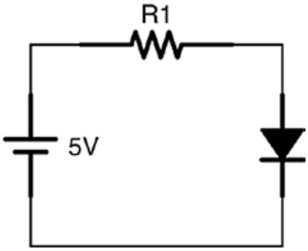

6. Let's say instead of an ordinary diode, the diode is a blue LED. Recalculate the current going through each component and the voltage drops for each component.

7. In the following circuit, calculate how much current is flowing through each component and each component's voltage drop if R1 is 300 Ω, R2 is 400 Ω, and R3 is 500 Ω.

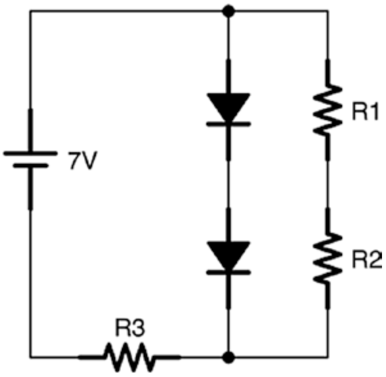

8. Draw a circuit that provides a 6-volt regulated power supply to the circuit load from a 9-volt battery using regular diodes. Choose a resistor that works efficiently for a circuit load of 500 Ω and operates with a battery voltage from 7 V to 9.6 V. What is the current at the lowest and highest ranges of the battery? How much is used by the circuit load, and how much is wasted through the diodes in each configuration?

9. Draw an equivalent circuit to the previous question using a Zener diode instead of normal diodes.

CHAPTER 9

Basic Resistor Circuit Patterns

When most people look at a schematic drawing, all they see is a sea of interconnected components with no rhyme or reason combining them. However, most circuits are actually a collection of **circuit patterns**. A circuit pattern is a common way of arranging components to accomplish an electronic task. Experienced circuit designers can look at a circuit and see the patterns that are being used. Instead of a mass of unrelated components, a circuit designer will look at a schematic and perceive a few basic patterns being implemented in a coherent way.

In this chapter, we are going to learn three basic resistor patterns and learn to work with switches as well.

9.1 Switches and Buttons

Switches and buttons are very simple devices, but nonetheless we probably need to take a moment to explain them. A switch works by connecting or disconnecting a circuit. A switch in the "off" position basically disconnects the wires so that the circuit can't complete. A switch in the "on" position connects the wires.

There are different types of switches depending on their operation. The ones we are concerned with are called "single pole single throw" (SPST) switches, which means that they control only one circuit (single pole) and the only thing they do is turn it on or off (single throw).

© Jonathan Bartlett 2020

J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_9

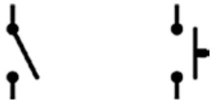

Figure 9-1. *Schematic Symbols for an SPST Switch (Left) and an SPST Button (Right)*

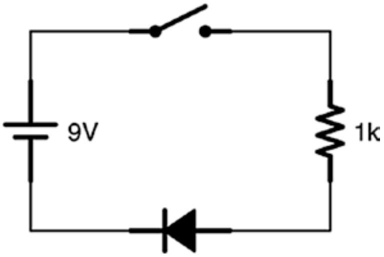

Figure 9-2. *A Simple Switch Circuit*

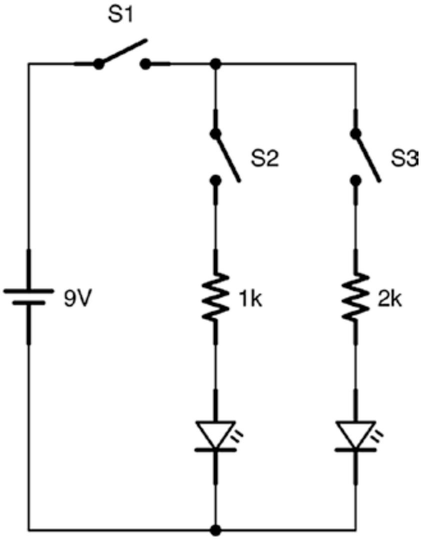

Figure 9-3. *A Circuit with Multiple Switches*

Figure 9-1 shows what the schematic symbols for an SPST switch and an SPST momentary switch (i.e., a button) look like. As the drawing indicates, when the switch is open, the circuit disconnects. When the switch closes, it connects the circuit. While the switch holds its position stable (someone has to manually switch it back and forth), the button only connects the circuit *while it is being pushed*. While the button is being pushed, the circuit is connected, but as soon as someone stops pushing the button, the circuit opens back up.

Figure 9-2 shows what a simple circuit with a switch looks like. It is just like a normal LED circuit, but with a switch controlling whether or not electricity can flow. Note that the switch is just as effective on the other side of the circuit. If the switch was the last part of the circuit, it would be equally as effective. Remember, in order for current to flow, there must be a full circuit from positive back to negative.

Switches can also be used to turn on or off individual parts of a circuit—basically reconfiguring the circuit while it is running. In the circuit given in Figure 9-3, a master switch (S1) turns the whole circuit on or off, and two individual switches (S2 and S3) turn parallel branches of the circuit on and off.

To analyze a circuit with switches, you need to analyze the way the circuit behaves with each configuration of switches. In this case, obviously when S1 is open, no current at all flows. However, this circuit will use different amounts of current when S2 is closed, S3 is closed, and both S2 and S3 are both closed. Therefore, to truly know the behavior of the circuit, you need to calculate the current usage in each of these situations.

9.2 Current-Limiting Resistor Pattern

The first resistor pattern we are going to learn is one that we already know—the current-limiting resistor pattern. The idea behind this pattern is that a resistor is added to limit the amount of current that can flow through a device. The size of the resistor needed depends on the size of the voltage source, the action of the device itself, and the maximum amount of current to allow. Then, the resistor size needed can be calculated using Ohm's law.

Many resistors are added to circuits to limit current flow. At the beginning, we used resistors to make sure we didn't destroy our LEDs. In Chapter 8, we used a resistor to limit the amount of current flowing through our voltage regulation circuit.

In many different circuits, we will need resistors to limit current for two different reasons—to avoid breaking equipment and to save battery life. Oftentimes, we are actually choosing resistor values to accomplish both of these tasks.

If an LED breaks with 20 mA, then we need a resistor big enough to keep the current that low. However, if the LED light is sufficiently visible with 1 mA, then, to save battery life, we might want a bigger resistor. Battery capacity is often measured in milliamp-hours (mAh), with a typical 9 V battery holding 400 mAh. So, with such a battery, an LED circuit at 10 mA will drain the battery in 40 hours (400 mAh/10 mA = 40 h), but the same LED circuit with a bigger resistor, limiting the current to 1 mA, will take a full 400 hours (400 mAh/1 mA = 400 h) to drain the same battery! That will save you a lot of money in the long run.

9.3 Voltage Divider Pattern

A **voltage divider** occurs anytime there are two resistors together with a subcircuit coming out from in between them. They usually are connected to a fixed positive voltage on one side of the first resistor and the ground on the other side of the second resistor, but this isn't strictly necessary. A simple schematic of a voltage divider is shown in Figure 9-4. Notice that there are two resistors between the voltage source and the ground (a 1k on top and a 2k on the bottom) and a subcircuit (indicated by the load resistance) branching off from between them. Under certain circumstances (which will be covered in a moment), we can basically ignore the parallel resistance of the subcircuit and just look at the voltages at each point in the main voltage divider circuit.

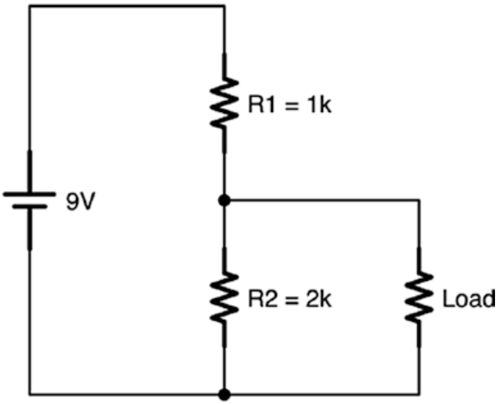

Figure 9-4. *A Simple Voltage Divider Circuit*

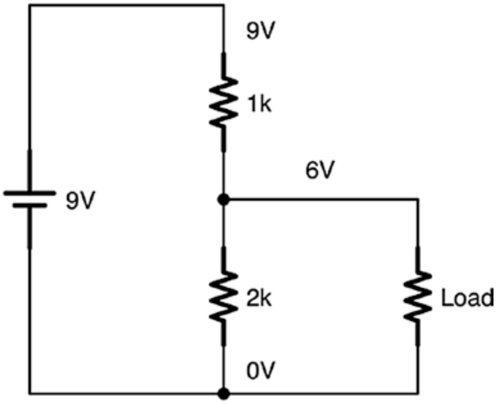

Figure 9-5. *Voltage Divider with Voltages Labeled*

9.3.1 Calculating the Voltages

We can see that the voltage at the top of the voltage divider is 9 V (because it connects to the positive terminal) and at the bottom of the voltage divider it is 0 V (because it connects to the negative terminal). Therefore, the total voltage drop across both resistors must be 9 V. Since the resistors are in series (remember, we are ignoring the load for now), we can find the total resistance in the circuit by just adding their resistances. So 1,000 Ω + 2,000 Ω = 3,000 Ω. Since the current in a series is the same for the whole series, we can now use Ohm's law to calculate the current flow:

$$I = \frac{V}{R} = \frac{9\text{V}}{3{,}000\Omega} = 0.003\text{A} = 3\text{mA}$$

So there is 0.003 A (3 mA) in this circuit. That means that *each* resistor in the series will have this amount of current flowing through them. Therefore, we can calculate the voltage drop across each resistor. Let's look at the 1k resistor:

$$V = I * R = 0.003\text{A} * 1{,}000\Omega = 3\text{V}$$

So the voltage drop across the first resistor is 3 V. That means that, since the battery started at 9 V, at the end of the resistor the voltage compared to ground is 6 V. We can calculate the voltage drop across the second resistor either by Ohm's law again or just by noting the fact that since the other end of the resistor is connected to ground, the voltage *must* go from 6 V to 0 V.

Figure 9-5 shows the voltages at each point (relative to ground). As you can see, the wire from the middle of the voltage divider has a new voltage that can be used by the load. This is what voltage dividers are normally for—they have a simple way of providing a scaled-down voltage to a different part of the circuit.

9.3.2 Finding Resistor Ratios

But how do we choose the values of the resistors?

One thing to note is that the second resistor consumed exactly twice as much voltage as the first resistor. Additionally, the second resistor was exactly twice as large as the first resistor. Thus, as a general principle, the relative sizes of the resistors will determine the relative amounts of voltage they eat up. So, if we needed a 4.5 V output—that is half of our input voltage—we would need both resistors to be the same.

A more explicit way of stating this is with an equation. Given a starting voltage V_{IN} connected to the first resistor, R_1, and the second resistor (R_2) connected to ground, the output voltage (V_{OUT}) coming out between the resistors will be given by the equation

$$V_{OUT} = V_{IN} * \frac{R_2}{R_1 + R_2} \tag{9.1}$$

Note that the specific resistance values don't matter yet—it is the *ratio* we are concerned about so far. To get 4.5 V, we can use two 1 kΩ resistors, two 200 Ω resistors, or two 100 kΩ resistors. As long as the values are the same, we will divide the voltage in half.

If we wanted an 8 V output, we would do a similar calculation. Since we start at 9 V, we need to use up $\frac{1}{9}$ of the voltage in the first resistor, and $\frac{8}{9}$ of the voltage in the second resistor. Therefore, our resistors need to be in similar ratio. We could use a 100 Ω resistor for the first resistor and an 800 Ω resistor for the second resistor. Alternatively, we could use a 10 Ω resistor for the first resistor and an 80 Ω resistor for the second resistor. It is the ratio that matters most.

9.3.3 Finding Resistor Values

So how do you determine exactly what value to use? Here is where we start thinking about the load again. While we have been treating the voltage divider as a series circuit, in truth we have one resistor in series and then a parallel circuit with the other voltage divider resistor in parallel with the load. Our simplified model (where we ignore the parallel resistance) will work, *as long as the load resistance does not impact the total parallel resistance by a significant amount*. Therefore, let's look at how the load resistance affects the parallel resistance.

So, using Equation 7.2, we can write a formula for the total resistance of these two, with R_2 being our second voltage divider resistor and R_L being our load resistance:

$$R_T = \frac{1}{\frac{1}{R_2} + \frac{1}{R_L}}$$

Now, let's look back at the circuit in Figure 9-5. Let's say that the resistance of the load (R_L) is 400 Ω, which is much less than the resistance of the voltage divider resistor (R_2, which is 2,000 Ω). So what is the total resistance?

$$R_T = \frac{1}{\frac{1}{R_2} + \frac{1}{R_L}} = \frac{1}{\frac{1}{2{,}000} + \frac{1}{400}} = \frac{1}{0.0005 + 0.0025} = \frac{1}{0.003} \approx 333\Omega$$

This is significantly different from our simplified model which ignored the load resistance, which gave 2,000 Ω. That means that our simplified model won't work with this value.

Now, let's imagine a bigger load resistance so that it is equal to the R_2 resistance (2,000 Ω) and recalculate:

$$R_T = \frac{1}{\frac{1}{R_2}+\frac{1}{R_L}} = \frac{1}{\frac{1}{2{,}000}+\frac{1}{2{,}000}} = \frac{1}{0.0005+0.0005} = \frac{1}{0.001} \approx 1{,}000\Omega$$

This is still significantly off, but it is much closer. So, now, let's look at what happens if the load resistance is double of R_2, or 4,000 Ω:

$$R_T = \frac{1}{\frac{1}{R_2}+\frac{1}{R_L}} = \frac{1}{\frac{1}{2{,}000}+\frac{1}{4{,}000}} = \frac{1}{0.0005+0.00025} = \frac{1}{0.00075} \approx 1{,}333\Omega$$

Here, we are getting much closer to our original value. Now, let's say that the load is ten times the resistance of our R_2 resistor, or 20,000 Ω. That gives us this:

$$R_T = \frac{1}{\frac{1}{R_2}+\frac{1}{R_L}} = \frac{1}{\frac{1}{2{,}000}+\frac{1}{20{,}000}} = \frac{1}{0.0005+0.00005} = \frac{1}{0.00055} \approx 1{,}818\Omega$$

This is very close to the resistance of $R2$ by itself. So what we can say is that our voltage divider circuit can ignore the resistance of the load *if the resistance of the load is significantly more than the resistance of the voltage divider resistor*. A way of writing this down is that $R\text{L} \gg R2$. What "significantly" means depends on how sensitive your circuit is to voltage changes, but, generally (and for the purposes of the exercises), I will say that "significantly more" should mean at least ten times as much.

9.3.4 General Considerations

So, for low-resistance loads, a voltage divider does not work well, because it puts too little resistance between the voltage source and ground. However, in Chapter 11, we will see that many circuits have loads of approximately infinite resistance, so voltage dividers work really well in those cases.

In general terms, a voltage divider with smaller resistors is "stiffer" because it varies less in response to variations in a load, but it also eats up more current. A voltage divider with larger resistors doesn't work with low-resistance loads, but it also uses up much less current.

If you need to know full equations for figuring out voltage divider resistors, I have them listed in Equations 9.2 and 9.3. In these equations, V_{IN} is the voltage coming into the voltage divider, V_{OUT} is the voltage going out of the divider into your load, and R_{L} is the load resistance. R_1 is the resistor connected to the positive power supply, and R_2 is the resistor connected to the ground. If you are good at algebra, you might see if you can deduce these formulas yourself from the information given in this section:

$$R_2 = \frac{R_L}{10} \tag{9.2}$$

$$R_1 = \frac{R_2 * \left(V_{IN} - V_{OUT}\right)}{V_{OUT}} \tag{9.3}$$

9.4 The Pull-Up Resistor

The pull-up resistor is a strange circuit, but we will find very good applications for it once we start dealing with ICs in Chapter 11. It is probably easiest to describe by simply showing you a circuit and then describing how it works.

Figure 9-6 shows the circuit diagram for a basic pull-up resistor circuit. Normally, we think of lighting up an LED by pushing the button. However, in this case, pushing the button causes the current to bypass the LED, which turns it off.

If you look at the path from where the circuit branches, when the button is not pressed, the current can only go one way—through the LED. However, when the button is pressed, the electricity has two options—either through the LED or directly to ground through the button. The electricity would always rather go directly to ground rather than through an intermediary, so *all* of the current goes through the closed button, and none of it goes through the LED.

Since the branch point is directly connected to ground when the button is pushed, that means that the voltage *at the branch point* is also zero. Kirchhoff's voltage law says that no matter what path is taken, the voltage drop will always be the same. However, an LED induces a voltage drop, but the voltages on both sides of the LED are zero. Therefore, electricity cannot flow through the LED.

So what is the function of the resistor? The resistor connects the switch and the LED to the positive voltage source and provides a limitation on the current that runs through it. The resistor must be *before* the branch point for it to work.

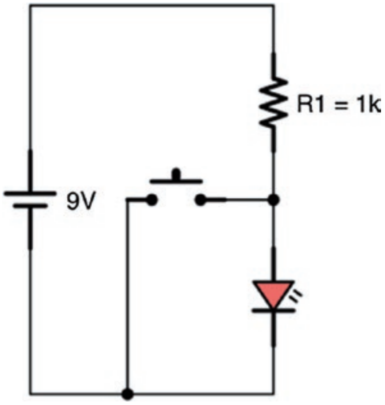

Figure 9-6. *Basic Pull-Up Resistor*

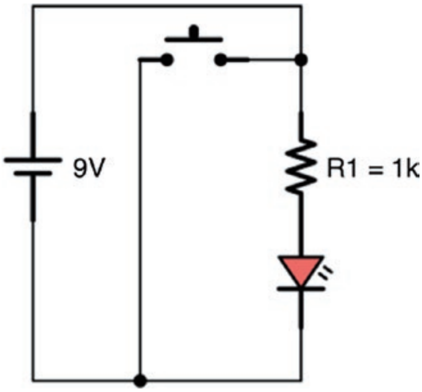

Figure 9-7. *Incorrect Way to Wire the Circuit*

Think about what happens without the resistor or if the resistor is after the branch point. The electricity will have a path directly from the positive voltage source to ground with no resistance—in other words, a short circuit. This will draw an enormous amount of electricity. Figure 9-7 shows what this would look like. Notice that when the button is pushed, you can trace a path from the positive voltage source to ground with no intervening resistance.

The resistor is called a pull-up resistor because it is connected to the positive voltage source and is used to "pull up" the voltage on the circuit to a positive value when the switch is open while still providing safety (by limiting the current) when the switch is closed.

In short, a pull-up resistor is usually used to supply positive voltage to a circuit which might be turned off by redirecting the voltage to ground. The resistor provides both the electrical connection to the positive source and a limit to the amount of current that will flow if the current flow is then routed to ground (usually through some kind of switching mechanism).

9.5 Pull-Down Resistors

One more basic resistor circuit that is often used is the pull-down resistor. While the pull-up resistor is connected to the positive power rail, the pull-down resistor is instead connected to ground. So, in a pull-up resistor, if part of the circuit is disconnected, the voltage goes high. In a pull-down resistor, if part of the circuit is disconnected, the voltage goes low. However, we don't yet have enough background really to understand how they are used here. They are covered more fully in Chapter 12. They are merely mentioned here because they are one of the basic resistor circuit patterns that are seen throughout electronics.

Review

In this chapter, we learned the following:

1. Buttons and switches allow circuits to be altered while they are running by connecting circuits (allowing pathways for electric current) and disconnecting circuits (blocking pathways for electric current).
2. Most circuits are a combination of common, well-understood circuit patterns.
3. The more experienced you are with the basic circuit patterns, the easier it is to see these circuit patterns when you look at a schematic drawing.
4. A current-limiting resistor is a resistor that is used to limit the maximum current flow within a circuit, either to protect other components or to limit overall current usage.

5. A voltage divider is a pair of two resistors connected in series with one another (usually connected to a positive voltage on one side and the ground on the other), but with another wire coming out in between them to provide voltage to another circuit (called the *load*).
6. In a voltage divider, it is assumed that the resistance of the load is significantly more than (i.e., greater than ten times) the resistance of the second half of the voltage divider because then the load can be basically ignored for calculating voltage drops.
7. For a voltage divider, the ratio of the voltages consumed by each resistor is the same as the ratio of their resistances. The output voltage coming out of the first resistor is the level of the voltage that will be supplied to the load.
8. Another way of stating the output voltage is $V_{OUT} = V_{IN} * \frac{R_2}{R_1 + R_2}$, where R_1 is the resistor connected to the positive voltage and R_2 is the resistor connected to ground.
9. Voltage dividers with smaller resistances are “stiffer”—they are impacted less by the resistance of the load. Voltage dividers with larger resistances are not as stiff but waste much less current.
10. A pull-up resistor circuit is a circuit in which a positive voltage which may be switched to ground at some point is provided through a resistor.

11. The pull-up resistor both (a) connects the circuit to the positive voltage to supply a positive current when the circuit is not switched to ground and (b) limits the current going to ground (i.e., prevents a short circuit) when the output is switched to ground.
12. It is called a pull-up resistor because it pulls the voltage up when the circuit is not switched to ground.

Apply What You Have Learned

1. In Figure 9-3, calculate the amount of current used by the whole circuit for each configuration of the switches S2 and S3 when S1 is closed. You can assume that the LEDs are red LEDs.
2. Build the circuit given in Figure 9-3 (you may swap out resistors with different but similar values—anything from 300 Ω to about 5 kΩ should work).
3. Given a 15 V voltage supply, what size of a resistor would be needed to make sure that a circuit never went over 18 mA?
4. Given a 9 V battery source, design a voltage divider that will output 7 V to a load that has a resistance of 10 kΩ.
5. Given a 3 V battery source, design a voltage divider that will output 1.5 V to a load that has a resistance of 1 kΩ.

6. In Figure 9-6, how much current is going through the circuit when the switch is open? How much when it is closed? You can assume that the LED is a red LED.
7. How would you modify the circuit in Figure 9-6 to keep the maximum current in the circuit under 2 mA? Draw the full circuit out yourself.
8. Build the circuit you designed in the previous question. If you do not have the right resistor values, use the closest ones you have.

CHAPTER 10

Understanding Power

So far we have covered the basic ideas of voltage, current, and resistance. This is good for lighting up LEDs, but for doing work in the real world, what is really needed is **power**. This chapter on its own adds very little to your capabilities as a circuit designer, but it is absolutely critical background information for the chapters that follow. Additionally, this chapter contains information critical to the safe usage of electronics. Knowing about power, power conversions, and power dissipation will be critical to taking your electronics abilities into the real world.

10.1 Important Terms Related to Power

To understand what power is, we need to go through a few terms from physics (don't worry—they are all easy terms):

1. **Work** happens when you move stuff.
2. Work is measured in **joules**. A joule is the amount of work performed when a 1-kilogram object is moved 1 meter.
3. The capacity to perform work is called **energy**. Energy is also measured in joules.
4. **Power** is the sustained delivery of energy to a process.

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_10

5. Power is measured in **watts** (abbreviated W). Watts are the number of joules consumed or produced per second.
6. Another measurement of power is horsepower (abbreviated hp). One horsepower is equivalent to 746 watts. Horsepower is not important for electronics, but I wanted to mention this because horsepower is a term you have probably heard, and I wanted you to be able to connect its importance to the ideas in this chapter.

One of the interesting things about work, energy, and power is that they can take on a number of forms that are all basically equivalent. For instance, we can have mechanical energy, chemical energy, and electrical energy (as well as others). We can also perform mechanical work, chemical work, and electrical work. All these types of energy and work can be converted to each other. They are all also measured in joules. Therefore, we have a common unit of energy for any sort of task we want to accomplish.

Now, when we actually apply energy to perform work, we do not get a 100% conversion rate. In other words, if we want to do 100 joules of work, we will probably need more than 100 joules of energy to perform the task. That's because the process of converting energy into work (as well as converting between different types of energy and work) is **inefficient**—not all of the energy gets directed to the task we want to perform. There is no perfectly efficient process of converting energy to work. Additionally, there is no way to create energy from nothing—any time you need additional energy, you will need a source for it.

When energy is converted to work, *all* of the energy does something, even if it isn't work on the task you want. Usually, the inefficiencies get converted to **heat**. So, if I have a process that is only 10% efficient and I give that process 80 joules of energy, then that process will do only 8 joules of work, leaving 72 joules of energy that is converted to heat.

Work and energy are usually the quantities we care about for systems that do a fixed task (i.e., require a fixed amount of energy). In electronics, however, we are usually building systems that stay on for sustained periods of time. Therefore, instead of measuring energy, we measure power, which is the continuous delivery of energy (or the continuous usage of energy in doing work).

As we mentioned, power is measured in watts, where a watt is 1 joule per second. So, if you have a 100 W light bulb, that bulb uses 100 joules of energy each second. True, 100 W incandescent light bulbs are very inefficient, which is why they get so hot—the energy that is not converted to light gets converted to heat instead. Today, incandescent bulbs are rarely used for that very reason. For instance, the LED bulbs that are labeled as "100 W" equivalent generally use about 11 W, but they are as bright as the old 100 W incandescents. What this means is that in an incandescent light, less than 11 W of energy was being put toward actually giving off light, and the rest was wasted as heat, which is why they got quite so hot.[1]

10.2 Power in Electronics

So we have a basic idea about what power is in general. In electronics, there are a few equivalent ways of calculating power.

The first is to multiply the number of volts being consumed by the number of amps of current going through a device:

$$P = V * I \tag{10.1}$$

---

[1] It is actually much less than 11 W, since LEDs are not perfectly efficient either. The point is that since they are both emitting the same amount of light (i.e., producing the same light energy), we can see that the difference in watts used is probably lost in heat.

Here, P indicates power measured in watts, V indicates volts, and I indicates current measured in amps. So, if my circuit is on a 9-volt battery and I measure that the battery is delivering 20 mA to the circuit, then that means I can calculate the amount of power that my circuit is using (don't forget to convert milliamps to amps first!):

$$\begin{aligned} P &= V * I \\ &= 9\text{V} * 20\text{mA} \\ &= 9\text{V} * 0.02\text{A} \\ &= 0.18\text{W} \end{aligned}$$

So our circuit uses 0.18 watt of power.

You can also measure the amount of power that individual components use. For instance, let's say that a resistor has a 3 V voltage drop and has 12 mA of current running through it. Therefore, the resistor uses up $3 * 0.012 = 0.036$ watt of power.

The second way of calculating power comes from applying Ohm's law. Ohm's law say

$$V = I * R \tag{10.2}$$

So, if we have the equation $P = V * I$, Ohm's law allows us to *replace* V with $I * R$. Therefore, our new equation becomes

$$P = (I * R) * I \tag{10.3}$$

Or we can simplify it further and say that

$$P = I^2 * R \tag{10.4}$$

We can also substitute $I = V/R$ and wind up with a third equation for power:

$$P = \frac{V^2}{R} \tag{10.5}$$

So, if we have 15 mA running through a 200 Ω resistor, then we can calculate the amount of power being used using Equation 10.4:

$$\begin{aligned} P &= I^2 * R \\ &= (15\text{mA})^2 * 200\Omega \\ &= (0.015\text{A})^2 * 200\Omega \\ &= 0.000225 * 200 \\ &= 0.045\text{W} \end{aligned}$$

10.3 Component Power Limitations

Now, if you think about it, the resistor in the previous example isn't actually *doing* anything. It is just sitting there. Therefore, since we are not accomplishing any *work* by going through the resistor, the energy gets converted to heat. Electronic components are usually rated for how much power they can **dissipate**, or easily get rid of. Most common resistors, for instance, are rated between 1/16 W and 1/2W (most that I've seen for sale are 1/4 W). This means that they will continue to work as long as their power consumption stays under their limit. If the power consumption goes too high, they will not be able to handle the increased heat and will break (and possibly catch fire!).

So far, our projects have dealt with low enough power that this isn't a concern. In fact, using 9 V batteries, it is hard to generate more than 1/4W of power—you would have to have less than 350 Ω of resistance on the *whole* circuit and have the entire voltage drop occur on the resistor.

In any case, whenever you are building circuits, you should keep in mind how much power the component is rated to handle and how much power it is actually consuming. You can use any of the equations given here to calculate power consumption. The component itself should have information on its maximum ability to handle power consumption and dissipation.

10.4 Handling Power Dissipation with Heatsinks

As we mentioned earlier, when power dissipates without doing any work, it is converted to heat. Some devices need to dissipate large quantities of heat under regular workloads. One common device that often needs to dissipate heat is the voltage regulator, like the 7805 regulator we will encounter in Chapter 12.

The way that this regulator performs its job is essentially by dissipating power until the voltage is at the right level. When used with any serious amount of current, this can actually get very, very hot. As such, the back side of these contains a metal plate which is used to dissipate heat. Additionally, it has a metal tab with a hole that can be used to attach a **heatsink**.

A heatsink is a metal structure with a large surface area that helps an electronic component dissipate heat. By being made of metal, it quickly moves the heat into itself. By having a large surface area, it can transfer the heat to the air, where it will then disperse into the environment.

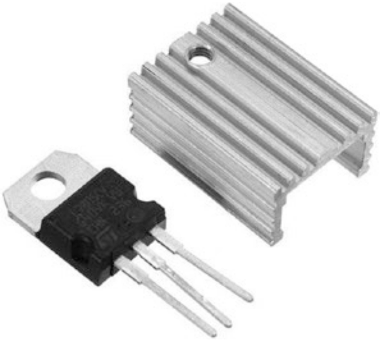

Figure 10-1. *A 7805 voltage regulator and its heatsink*

Figure 10-1 shows a 7805 chip next to its heatsink. To attach the heatsink, just screw it into the 7805. On the 7800 series of regulators, the tab is electrically connected to ground, so it should not produce a voltage. However, other types of chips in the same TO-220 package may actually have a voltage on the tab. In such a case, it would be wise to buy an isolation kit to electrically isolate the chip from the heatsink; otherwise, incidental contact with the heatsink could cause a short circuit. The isolation kit will only allow heat, not electricity, flow into the heatsink.

10.5 Transforming Power

As we have discussed, energy (and therefore power) can be transformed among a variety of forms—mechanical, electrical, chemical, and so on. In any case, always remember that energy is only reduced or lost, never gained.

The essence of energy transformation is at the heart of what makes batteries work. A battery contains energy stored in a chemical form. Chemical reactions in the battery allow electrons to move. By drawing the electrons through a specific path (drawing them to the positive from the negative), this reaction generates electrical energy. So we have a conversion from chemical energy (the reaction of the chemicals in the battery) into electrical energy (the pull of the electrons through the circuit).

This can also go the other way. Electrical energy can be used to stimulate chemical reactions. A common one is the separation of water into hydrogen and oxygen.

The same conversion can happen with mechanical energy. In an electric motor, electrical energy is converted into mechanical energy in the motor. But the reverse also can occur. A power generator is made by converting mechanical energy into electrical energy.

We won't go into details on how each of these transformations works (you would need to take courses in chemistry, mechanics, etc. to know more), but the essential ideas are that

1. Energy and power can be transformed between a variety of forms.
2. These forms of energy can all be measured with the same measuring stick (joules).
3. Every energy transformation will lose (never gain) some amount of energy through inefficiencies.

Power and energy are known as **conserved quantities**, because, although they are transformed, they are never created or lost. Note that when we talk about power lost through inefficiencies, the power actually isn't lost in total; it is merely converted to heat. You can think of heat as power that is applied in a nonspecific direction. Energy is lost when it is transformed not because it disappears but rather because all processes which channel one form of energy into another are imperfect.

In Section 10.2, "Power in Electronics," we noted that, in electronics, the power (measured in watts) is determined by *both* the voltage and the current—by multiplying them together. Therefore, what will be conserved in electronics will not be the voltage or the current individually, but their product. What this means is that we can, at least in theory, increase the voltage without needing a power gain, but at the cost of current. Likewise, we can, at least in theory, increase the current without needing a power gain, at the cost of voltage. In both of these cases, rather than transforming electrical power to another form of power altogether, we are transforming it into a different configuration of electrical power.

Devices that convert electrical power between different voltage/current configurations are known as **transformers**. A step-up transformer is one that converts a low voltage to a higher voltage (at the cost of current), and a step-down transformer is one that converts a high voltage to a lower voltage (but can supply additional current).

Technically, for DC circuits, these are usually known as **DC-DC power converters** or **boost converters** instead of being called transformers, but the same rules apply—the total number of watts delivered can never increase, but the voltage can be converted up and down at the expense or gain of the current.[2]

So, if I had a source of 12 V and 2 A, then I would have $12 * 2 = 24$ W. Therefore, it may be possible to convert that to 24 V, but I would only be able to get 1 A of current ($24 * 1 = 24$). However, I could drop the voltage to get more current. If I needed 4 A, I could reduce the voltage to 6 V.

Also remember that in doing these transformations, there is always some amount of power loss as well, but these calculations will give you what the maximum possibilities are. The actual mechanisms that these devices employ for doing power conversions are outside the scope of this book.

10.6 Amplifying Low-Power Signals

Many devices, especially integrated circuits, are only capable of processing and generating low-power signals. Microcontrollers (like the ATmega328/P) have limits to how much power they can send or receive. The ATmega328/P can only source up to 40 mA per pin and only about 200 mA total across all pins simultaneously. At 5 V, 40 mA would yield a maximum of 0.2 W. Therefore, if you want to turn on a device that requires more power than that, you will need to **amplify** your signal.

Now, as we discussed previously, you can't actually create more power out of nothing. What you can do is, instead of trying to *create* power, you can instead *control* power. We will discuss several specific techniques on how to do this starting in Chapter 24, but the essential idea is that you can amplify a signal by using a small signal to control a larger one.

---

[2] As a sidenote, if you tried to wire a regular AC transformer to a DC power supply, it would not deliver any power at all—DC-DC converters work on very different principles than AC transformers.

Think about your car. The way that you control your car is by taking a low-power signal, such as the gas pedal, and using it to control a high-power signal, such as the engine. My foot is not directly powering the car. My foot is merely using the pedals to tell another power source—the engine—how much of its energy it should move. My foot doesn't actually interact directly with the engine at all, except as a valve to unleash or not unleash the power available in the gas tank and engine.

In the same way, since the output signals from the microcontrollers are low power, instead of using these signals directly, we will use the signals to control larger sources of power. Devices which can do this include relays, optocouplers, transistors, op-amps, and darlington arrays. We will cover more about amplification starting in Chapter 24.

Review

In this chapter, we learned the following:

1. Work is what happens when you move stuff and is measured in joules.
2. Energy is the capacity to do work and is also measured in joules.
3. Power is the sustained delivery of energy and is measured in joules per second, also called watts.
4. Power can be converted to a number of different forms.
5. Converting power to another form or using it to do work always has inefficiencies, and these inefficiencies result in energy lost to heat.

6. In electronics, power (in watts) is calculated by multiplying the voltage by the current ($P = V * I$). It can also be calculated as $P = I^2 * R$ or $P = V^2/R$.
7. To calculate the power consumption of an individual component, use the voltage drop *of that component* and multiply it by the current flowing through it (measured in amps, not milliamps).
8. Most components have a maximum rating for the amount of power they can safely consume or dissipate. Be sure you design your circuits so that your components stay under that limit.
9. Some components can handle additional power dissipation by adding a heatsink, which will more effectively dissipate the excess heat into the air.
10. Power can be transformed into other types of power (mechanical, chemical, etc.), but can never go beyond the original amount of power.
11. Electrical power can also be transformed into different combinations of voltage and current, as long as the total power remains the same.
12. Components that do this transformation are called transformers for AC power and DC-DC converters for DC power.
13. Because power cannot be created, the way that signals are amplified is by using a small power signal to control a larger power source.

Apply What You Have Learned

1. If I have 50 joules of energy, what is the maximum amount of work I could possibly do with that amount of energy?
2. If I am using up 10 joules of energy each second, how many watts am I using up?
3. If I convert 30 watts of mechanical power into electrical power with 50% efficiency, how many watts of electrical power are delivered?
4. If I have a circuit powered by a 9 V battery that uses 0.125 A, how many watts does that circuit use?
5. If a resistor has a 2 V drop with a 0.03 A current, how much power is the resistor dissipating?
6. If a resistor has a 3 V drop with a 12 mA current, how much power is the resistor dissipating?
7. If a 700 Ω resistor has a 5 V drop, how much power is the resistor dissipating?
8. If a 500 Ω resistor has 20 mA flowing through it, how much power is the resistor dissipating? If the resistor was rated for 1/8 of a watt, are we within the rated usage for the resistor?

9. In the following circuit, calculate the voltage drop, current, and power dissipation of every component (except the battery). If the resistors are all rated for 1/8 of a watt, are any of the resistors out of spec?

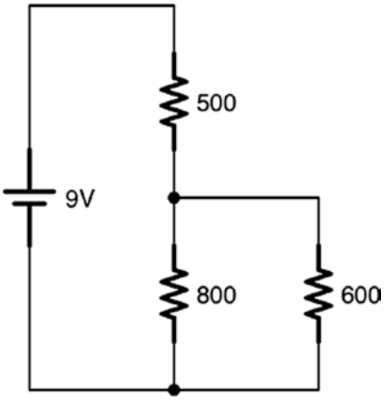

CHAPTER 11

Integrated Circuits and Resistive Sensors

So far, the components we have studied are simple, basic components—batteries, resistors, diodes, and so on. In this chapter, we are going to start to look at **integrated circuits**, also called **chips**, **microchips**, or **IC**s. An IC is a miniaturized circuit placed on silicon. It is a whole collection of parts geared around a specific function. These functions may be small, such as comparing voltages or amplifying voltages, or they may be complex, such as processing video or even complete computers. A single chip may hold just a few components, or it may hold billions.

Miniaturized circuits have several advantages—they are cheaper to produce in mass, they use less power, and they take up less space in your overall circuit—all because they have a reduced area and use fewer materials. These miniaturized circuits are what allowed for the computer revolution over the last century.

11.1 The Parts of an Integrated Circuit

Integrated circuits, as we have noted, are basically miniaturized circuits placed on a silicon plate, called the die. This die is where all of the action of the integrated circuit takes place.

© Jonathan Bartlett 2020

J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_11

The die is then placed into a **package**, which then provides connection points for circuit designers to interface with the IC. These connection points are often called **pins** or **pads**. Each pin on an IC is numbered, starting with pin 1 (we will show you how to find pin 1 shortly). Knowing which pin is which is important, because most of the pins on a chip each have their own purpose, so if you attach a wire to the wrong pin, your circuit won't work or you will destroy the chip. Most packages are marked with the chip's manufacturer and part number. If they weren't, it would be nearly impossible to tell one chip from another.

There are many different types of packaging available, but there are two general types that are often encountered:

Through-Hole: In this packaging type, the connection points are long pins which can be used on a breadboard. This type of packaging is easiest for amateur usage.

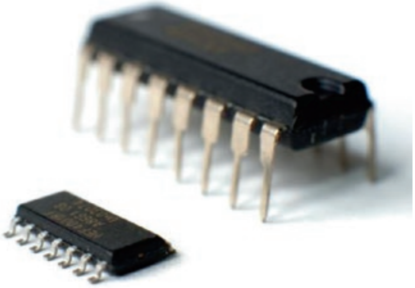

Figure 11-1. *Comparison of the Same IC in SMD (Left) and DIP (Right) Packages (Image Credit: Shutterstock/Youra Pechkin)*

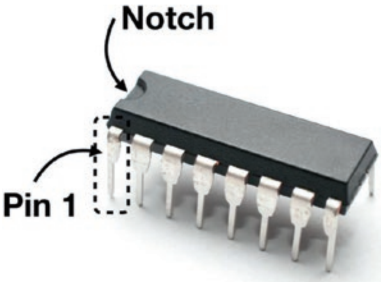

Figure 11-2. *Pin 1 is Immediately Counterclockwise of the Notch (Image Credit: Shutterstock/Cristian Storto)*

Surface Mount: In this packaging type, the connection points are small pads which are meant to be soldered to a circuit board. These packages are much smaller (and therefore less expensive) and can be more easily managed by automated systems to build completed circuits. These are also referred to as SMDs (surface mount devices) or SMT (surface mount technology).

Since we are only using breadboards in this book and not doing any soldering, we will only concern ourselves with through-hole packaging. However, through-hole packaging itself comes in a variety of styles. The main one we will concern ourselves with is called a **dual in-line package**, or **DIP**. Figure 11-2 shows the same chip in SMD and DIP configurations.

An integrated circuit in a DIP package has two rows of pins coming out of the package. Most chips either mark the top of the chip with a notch or indentation (where pin 1 is immediately counterclockwise of the notch) or mark pin 1 with an indentation or both. See Figure 11-2 to see how to use the notch to find pin 1. The rest of the pins are numbered counterclockwise around the chip.

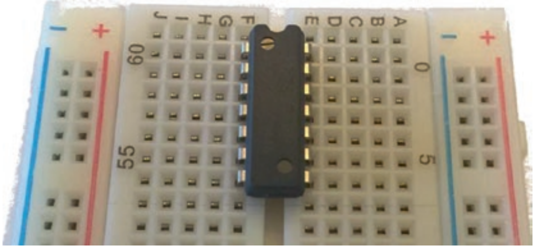

Figure 11-3. *A DIP IC Inserted Into a Breadboard*

The beauty of a DIP-packaged IC is that it fits perfectly onto most breadboards. Figure 11-3 shows how you can place your IC across the breadboard's bridge and each pin on the chip will have its own terminal strip to connect to.

Be careful, though, when inserting ICs into breadboards. The spacing of the pins on an IC is often slightly wider or shorter than the breadboard's bridge, and pins have to be inserted carefully. If you just jam the IC into the breadboard, you will likely accidentally crush one or more of the pins that aren't exactly aligned on the hole. Instead, compare the spacing of the pins to the bridge spacing on your breadboard. If it doesn't match up, *very gently* bend the pins with your fingers or with pliers to get them to match up. It's usually just a very small amount, but if you don't take care, you can easily damage your IC.

Usually, the ICs that I purchase are just a little wide, and I will squeeze the pins on each side slightly between my thumb and finger until they move close enough together. However you adjust the pins, make sure they line up before pushing them into their connection points on the breadboard. Also, with larger ICs, you may also need to slightly wiggle the IC back and forth as you gently insert it into place on the breadboard.

11.2 The LM393 Voltage Comparator

There are thousands and thousands of available chips which do a dizzying array of functions. In this chapter, we are going to focus on a very simple chip—the LM393 voltage comparator. This chip does one simple task. The LM393 compares two input voltages and then outputs either a high-voltage signal or a low-voltage signal depending on which input voltage is greater. The LM393 is actually a *dual* voltage comparator, which means that it will do two separate comparisons on the same chip. Like most chips, the LM393 is an *active* device, which means that it additionally requires a voltage source and a ground connection to provide power to the device.

Figure 11-4 shows the pin configuration (also called the **pinout**) of the LM393. The first thing to note on any pinout is where the voltage and ground connections are. In this case, the voltage is marked as `VCC`, and the ground is marked as `GND`. Even though the LM393 has *two* voltage comparators on the chip, they both share the power (`VCC`) and ground (`GND`) pins. The left side of the chip diagram shows the inputs and output for the first voltage comparator (`1IN+`, `1IN-`, and `1OUT`), and the inputs and output for the second voltage comparator are on the right (`2IN+`, `2IN-`, and `2OUT`). In your projects, you can use whichever one is more convenient for you, or even both at the same time if you have more than one voltage comparison task.

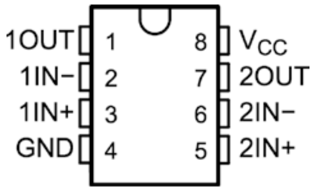

Figure 11-4. *The Pin Configuration of an LM393*

So the `1IN+` pin (pin 3) and the `1IN-` pin are where the two voltages are being fed that are being compared by the first comparator. The `1OUT` is the pin which will contain the output. If the voltage at `1IN+` is less than the voltage on `1IN-`, the output pin will be at a low (i.e., near-zero/ground) voltage. If the voltage at `1IN+` is greater than the voltage on `1IN-`, the output pin *will not conduct at all*, but this will be considered a "high" (positive voltage) state. This sounds counterintuitive, but, as we will see, this lets us set our own output voltage to whatever we want without causing too much complexity. This configuration where high-voltage outputs don't conduct is called an **open collector** configuration. Don't worry if this is a little confusing, we will discuss it more in depth later in the chapter.

VOLTAGE SOURCES ON INTEGRATED CIRCUITS

Note that the voltage pins on integrated circuits can be marked in a number of different ways. The positive voltage source is often labeled as V_{CC} , V_{DD}, or *V*+. The ground connect is often labeled as *GND*, V_{EE}, V_{SS}, or *V*–. There are additional ways that these are labeled as well. Finding the positive and ground connections for an IC should always be the first thing you do with it.

11.3 The Importance and Problems of Datasheets

Every IC (and, usually, any other part as well) has a **datasheet** supplied by the manufacturer which tells you important details about how you should use their chip in your circuit. Reading datasheets is one of the worst parts of electronics, in my opinion. For me, datasheets rarely have the information I am actually looking for in a way that is easy to find.

In fact, most datasheets assume that you already know how to use the device, and the datasheets are just there to supply additional details about the limitations of the device. For instance, looking through the LM393 datasheet from Texas Instruments, the actual operation of the device isn't even listed until page 11, and there it is buried within a sub-subsection, almost as a sidenote.

These datasheets are written *by* people who have spent a lot of time being electrical engineers, and they are written *for* people who have spent a lot of time being electrical engineers, so when mere mortals read the datasheets, the important pieces are often shrouded in unintelligible gibberish. For instance, the fact that the "high" output state of the device doesn't conduct isn't mentioned explicitly anywhere at all in the datasheet. Instead, it is implied by the configuration.

The reason for this is that the datasheets are usually read by professionals familiar with the type of device and who just need to know the electrical details so they don't accidentally bend the device beyond the breaking point. Thus, the datasheets oftentimes spend more time just showing and describing the layout of the circuit on the chip and graphs of different chip properties, and then you are left to interpret what that means for your circuit. For advanced circuit designers, this is great. For students and hobbyists, however, this is oftentimes more frustrating than helpful.

However, datasheets do often provide a few basic details that are helpful to everyone. They will often tell you

- What each pin does
- What the power requirements are
- What the outer limits of the chip's operation are
- An example circuit that you can build with the device

For all of these reasons, Appendix E contains simplified datasheets for a number of common devices that are easier to read than the standard ones.

For the LM393, the important points are as follows:

1. The input voltage on V_{CC} can be anywhere between 2 V and 36 V.
2. When sensing voltage, the LM393 doesn't really draw any (or at least much) current, so there are no parallel resistances we need to worry about.
3. The output is high when `IN+` is greater than `IN-` and low (i.e., near-ground) when `IN+` is less than `IN-`, with an error range of about 2 millivolts.
4. When the output is low, this means the output pin will conduct current into itself (since it is at ground, positive charge will naturally flow into it), but if it sinks more than 6 mA into it, you will destroy it, so you have to be sure that you know the maximum amount of current that could flow through the pin when it is at ground.
5. When the output is high, this means the output will not conduct any current, but instead act as if it were disconnected. As we will see, this will allow us to supply our own "high" voltage level.

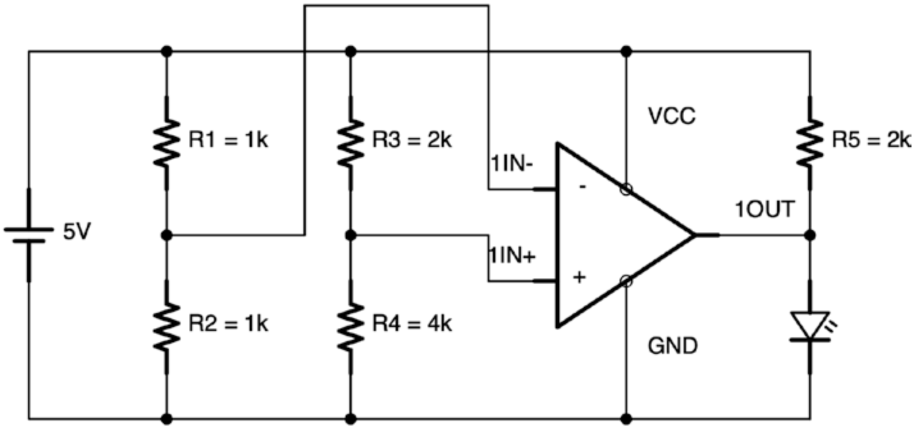

Figure 11-5. *A Simple Comparator Circuit*

That isn't to say that the datasheets aren't important, but for a beginner, the datasheets usually aren't what you need to get started.

11.4 A Simple Circuit with the LM393

In this section, I am going to show a simple circuit using the LM393 chip. In doing so, we are going to be using several of the resistor circuit patterns that we learned in Chapter 9.

The circuit we are discussing is shown in Figure 11-5. Can you identify the resistor circuit patterns? Take a minute and see if you can find some. Note that the wire coming out of `1IN-` crosses two wires that it is *not* joined with.

The first thing to notice is that we have *two* voltage dividers. The first voltage divider is between R1 and R2. Since R1 and R2 are the same resistance and are connected to both 5 V and 0 V, that means that they divide the voltage in half, giving a 2.5 V output. The second voltage divider is between R3 and R4. Since R3 is half of the resistance of R4, that means that it only uses up half as much voltage as R4. Thus, since R3 eats up 1.7 V and R4 eats up 3.3 V, the wire coming out from the middle is at 3.3 V.

Then, to the right of the circuit, you can see that we have a current-limiting resistor in front of the LED. That is not its only function, though. It also functions, as we will see shortly, as a pull-up resistor.

So what is the big triangle? Comparators (and several other circuits commonly placed on ICs) are represented as triangles in the schematic (we could have also placed the chip itself there). Each of the connections is labeled the same as they are labeled in the pinout diagram in Figure 11-4 so they would be easy to locate.

The way that the circuit works is very simple. The voltage coming in to `1IN+` is 3.3 V, and the voltage coming in to `1IN-` is 2.5 V. Since `1IN+` is greater than `1IN-`, then that will turn `1OUT` to high (positive voltage). However, remember that we said that `1OUT` *does not conduct* when it is high. It acts like an open switch. Therefore, R5 acts like a pull-up resistor and supplies the positive voltage for us to our LED to turn it on.

Now let's say that the input voltages were reversed. What would happen? If `1IN-` is greater than `1IN+`, then `1OUT` will go low (0 volt) and also conduct. It will act like a closed switch going to ground.

Therefore, current will go the easy route—it will go through `1OUT` (directly to ground) instead of through the LED (1.8 V or more), effectively denying power to the LED and switching it off. This works just like the switch in the circuit in Chapter 9, Section 9.4, "The Pull-Up Resistor." When `1OUT` is low, it acts like a closed switch to ground and acts as a sink for all current at that point in the circuit. When `1OUT` is high, it acts like an open switch, and whatever voltage/current you provided at that point is allowed to continue on.

The resistor R5 does several jobs. The first job is to act as a pull-up resistor, as we just described. Remember that a pull-up resistor prevents the load to ground from going too high when the switch is closed. Without the pull-up resistor, when the switch is closed (`1OUT` goes low), we would have a short circuit from the voltage source to ground. This would not only waste a large amount of electricity; it would break the LM393, because it can only sink a maximum of 6 mA of current. Having a 2 kΩ resistor, we limit the current for the closed switch to $I = V/R = 5/2{,}000 = 0.0025\ A = 2.5$ mA.

When the switch is open, the current flows through the resistor to the LED, and then the resistor acts as a current-limiting resistor for the LED. The amount of current to the LED will be calculated as $I = V/R = (5 - 1.8)/2{,}000 = 3.2/2{,}000 = 0.0016$ A = 1.6 mA.

11.5 Resistive Sensors and Voltages

One of the more practical uses of the voltage comparator circuit is to measure the values of sensors which act as variable resistors. Many different materials in the world act as resistors. What's really interesting is that many of these materials *change their resistance* depending on external factors. Some of them change their resistance based on temperature, pressure, light, humidity, and any number of other environmental factors.

Now, changing resistance doesn't tell us much by itself. If we put a resistor between a voltage source and ground, it will always eat up that voltage source. However, if you use it in concert with a fixed resistor to make a voltage divider, you can then get the output voltage to vary based on the changes in resistance.

Figure 11-6 illustrates this principle. It is a simple voltage divider, where the top resistor is actually a photoresistor (a resistor that varies based on light) and the bottom resistor has a fixed resistance. Thus, as the light varies, the top resistance will vary. This will change the ratio between the top and bottom resistors, which will affect the output voltage.

To use this circuit, you will need to know the resistances of your photoresistor on the different conditions you are interested in. I usually use the GL5528, which ranges from 10 kΩ in bright light to 1 MΩ in complete darkness. However, depending on your specific photoresistor as well as the light conditions that you think of as "light" and "dark," the resistance values that are relevant for light and dark will be different for you. So, whatever photoresistor you use, it is worthwhile to measure the resistance using your multimeter in the different conditions you think of as light and dark.

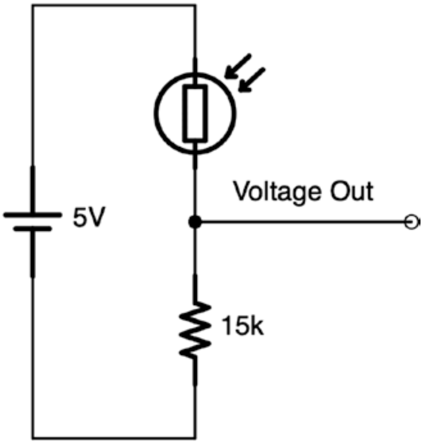

Figure 11-6. *A Simple Resistor Sensor Circuit*

11.6 Sensing and Reacting to Darkness

So far in the book, we have focused entirely on example circuits that didn't really do anything. They lit up, they had voltage and current, but there wasn't much interesting that they were doing. However, now, we finally have enough knowledge to start building circuits that *do* something.

We have

1. A way to generate a fixed voltage (using a voltage divider)
2. A way to generate resistances from real-world events (photoresistors and other resistance sensors)

3. A way to convert changes in resistance to changes in voltage (using a voltage divider with one fixed resistor)
4. A way to compare our varying voltage to our fixed voltage (using the LM393 comparator)
5. A way to utilize the output signal from the LM393 to do work (using the pull-up resistor and the LED)

There are a lot of pieces to put together this simple circuit, which is why it has taken so long to do anything worthwhile. However, if you have followed along carefully, now that you are here, you should be able to see how all of this fits together.

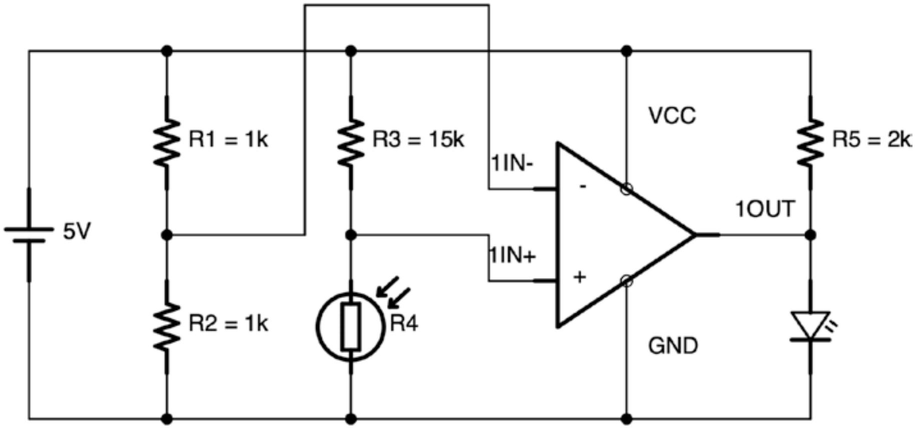

Figure 11-7. *Darkness Sensor Schematic*

What we will do is to take the circuit given in Figure 11-5 and modify R4 to be our photoresistor and R3 to be a fixed resistor. In my own testing, I discovered that the light/dark switchover point for my photoresistor was about 15 kΩ. Therefore, I am going to use a 15 kΩ resistor as the fixed resistor for R3. Yours may need to vary based on your experimentation with your photoresistor.

When there is light in the room, the photoresistor will have a lower resistance than 15 kΩ, which will make the fixed resistor R3 use up more of the voltage. Thus, the voltage at the divider will be less than 2.5 V, which will turn `1OUT` to low (which closes the switch and makes a path to ground on the output before it gets to the LED, which turns the LED off).

In low-light conditions, the resistance will jump way up above the resistance of the fixed resistor. If the upper, fixed resistor has less resistance than the bottom resistor, then the voltage at the divider will be larger than 2.5 V, activating the comparator and turning `1OUT` to high (i.e., opening the switch and allowing power to flow through the LED).

The final circuit is given in Figure 11-7. You can see a way to lay it out on the breadboard in Figure 11-8.

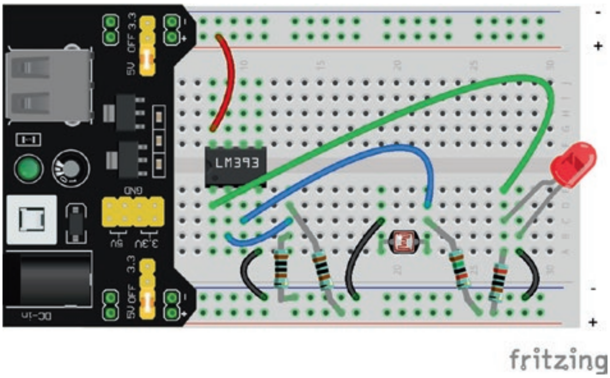

Figure 11-8. *Darkness Sensor Breadboard Layout*

Sources and Sinks

Two terms that often come up when dealing with circuits are the concepts of a current **source** and a current **sink**. A source is a component whose pins might provide current to other parts of the circuit. A sink is a component whose pins might pull current from other parts of the circuit.

For the LM393, its input pins neither source nor sink current (at least not any significant amount). The input pins more or less just sense the voltage without pulling any measurable current. Therefore, they are neither sources nor sinks of current. Technically, they probably sink a few nanoamps (billionths of an amp), but not nearly enough to affect our circuit analysis.

The output pin, even though it is called an *output*, doesn't source any current. Instead, it acts either as a sink (when low) or as a disconnected circuit (when high). This is known as an **open collector** output.

Anytime an IC sources or sinks current, be sure to read the datasheets on the maximum amount of current it can source or sink. These are usually quantities that *you* have to limit—they are merely telling you at what point their circuit will physically break. Therefore, you must use resistors to limit the currents to make sure that they are within limits.

However, be aware that many (but certainly not all) ICs do not source current, using open collectors for their output operations. This has the disadvantage that you have to supply your own voltage and pull-up resistor to the output pin, but it also has the advantage that the output is set to *whatever voltage level you choose*. In other words, you don't need to pick a new comparator IC to get a different output voltage.

Review

In this chapter, we learned the following:

1. Integrated circuits (called ICs or chips) are miniaturized circuits packaged up into a single chip that can be added to other circuits.
2. ICs can have a few or several billion components on them, depending on the function.

3. ICs have different types of packages, including through-hole (optimized for breadboards) and surface mount (optimized for soldering and machine placement).
4. Dual in-line packages (DIPs) are the most common through-hole packaging type used for students, hobbyists, and prototype builders.
5. DIP chips should be placed in the breadboard saddling the bridge, so that each IC pin is attached to its own terminal strip.
6. On most chips, pin 1 is located immediately counterclockwise of the notch in the chip, and remaining pins are numbered counterclockwise.
7. Most ICs are active devices, meaning that they have a direct connection to a power supply and ground in addition to their normal input and output pins.
8. An IC datasheet is a document that tells about the electrical characteristics of an IC. However, most of them are difficult to read and assume you are already familiar with the part. However, they are very useful for getting a pinout for the chip as well as telling the maximum ratings for voltages and currents.
9. The LM393 is a dual voltage comparator IC—it compares two voltages and alters its output based on which is larger.
10. The LM393's inputs do not consume any significant current when sensing the input voltages.

11. The LM393's outputs are open collectors—which means that they act as a switch to ground. When the output is "low," the pin acts as a closed switch to ground. When the output is "high," the pin acts as a disconnected circuit.
12. Because the LM393 acts as a disconnected circuit when high, a pull-up resistor circuit is required to get an output voltage.
13. Many sensors are based on the fact that the resistance of many materials will change with environmental factors. Therefore, the sensor acts as a variable resistor, with the resistance telling you about the environment.
14. A resistive sensor can be used with a fixed resistor to make a variable voltage divider, essentially converting the resistance to a voltage, which then can be detected.
15. By putting the sensor-based resistive voltage divider in comparison with a fixed reference, we can use the LM393 comparator to trigger an output when the sensor crosses some threshold of resistance.

Apply What You Have Learned

1. Calculate the amount of current flowing through each element of the circuit in Figure 11-5. You can presume that the LM393 uses about 1 mA for its own (internal) operation and that the LED is a red, 1.8 V LED. What is the total amount of current used by the circuit?

2. Take the circuit in Figure 11-5 and swap which voltage divider is attached to `1IN+` and `1IN-`. Now calculate the total amount of current used by this circuit.

3. The Spectra flex sensor is a resistive sensor that changes its resistance when bent. When it is straight, it has a resistance of 10 kΩ. When it is bent, it has resistances of 60 kΩ and above. Draw a circuit that turns on an LED when the resistor is bent. Use a resistor symbol for the flex sensor, but label it as `FLEX`.

4. Build the circuit in Figures 11-7 and 11-8.

5. If you wanted to wait until the room was even darker before the LED went on, how would you change the circuit?

PART II

Digital Electronics and Microcontrollers

CHAPTER 12

Using Logic ICs

In Chapter 11, we worked with our first integrated circuit, the LM393 voltage comparator. In this chapter, we are going to look at other ICs and talk more about how they are named and used in electronics.

12.1 Logic ICs

One of the easiest class of ICs to use are the *logic* ICs. A logic IC is a chip that implements a basic function of **digital logic**. In digital logic, electric voltages are given meanings of either “true” or “false,” usually with “false” being a voltage near zero and “true” being a positive voltage (often between 3 and 5 volts). These values are also referred to by a number of other designations—a 1 (for true) and 0 (for false) or HIGH (for true) and LOW (for false).[1] Then, the digital logic ICs implement logic functions that combine different signals (usually designated as A and B) and give an output signal (usually designated as Y or Q).

---

[1] Very rarely, some ICs will reverse the meaning of the positive and zero voltages.

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_12

For instance, the **AND** function will output a “true” (positive voltage) value if both of its inputs are true and will output a “false” (near-zero voltage) value otherwise. In other words, if A *and* B are true, Y is true. As another example, the **OR** function will output a “true” value if either of its inputs is true. In other words, if A *or* B is true, Y is true. Figure 12-1 shows the most common types of logic operations and how they work.

As we have seen, AND yields a true result when both A and B are true, and OR yields a true result when either A or B is true. So what are the others? **XOR** is *exclusive OR*, which means that it is just like OR, but is also false when both inputs are true. **NOR** is *not OR*, which means that it is the exact opposite of OR. Likewise, **NAND** is *not AND*, which means that it is the exact opposite of AND. Finally, **NOT** only has one input and simply reverses its value.

Each digital logic function, when implemented in electronics, is called a **gate**. The nice thing about building circuits with logic gates is that, rather than using math, you can build circuits based on ordinary language.

Operation	**A**	**B**	**Y** (output)
AND	false	false	false
AND	false	true	false
AND	true	false	false
AND	true	true	true
OR	false	false	false
OR	false	true	true
OR	true	false	true
OR	true	true	true
XOR	false	false	false
XOR	false	true	true
XOR	true	false	true
XOR	true	true	false
NOR	false	false	true
NOR	false	true	false
NOR	true	false	false
NOR	true	true	false
NAND	false	false	true
NAND	false	true	true
NAND	true	false	true
NAND	true	true	false
NOT	false	N/A	true
NOT	true	N/A	false

Figure 12-1. *Common Logic Operations*

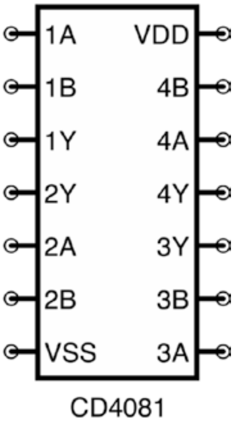

Figure 12-2. *The Pinout of a CD4081 Chip*

If you were to say, "I want my circuit to output a signal if both button 1 *and* button 2 are pressed," then it is obvious that you would use an AND gate to accomplish this.

INVERTED INPUTS AND OUTPUTS

Sometimes an input or an output on a chip will be labeled with a line on top of the input or output name. This means that the result (or requirement) will be the *opposite* of what you would otherwise expect. For instance, if the output is Y, then $\overline{\text{Y}}$ refers to the opposite value of Y. So, if Y is true, then $\overline{\text{Y}}$ will be false.

The same can occur on inputs as well. For instance, some chips have a reset on the chip which will reset the chip. However, we would normally consider "signaling" a pin to be sending it a voltage. Sometimes the chips would prefer that you signal the reset by bringing the voltage low instead of high. In these cases, they usually designate this by putting a bar over the name of the input pin, so the reset pin would be labeled $\overline{\text{RESET}}$ in order to indicate this behavior.

These are known as **inverted** inputs and outputs.

Most logic gates are implemented in chips that contain multiple (often four) implementations of the same gate. For instance, the CD4081 chip is a quad–NAND gate chip. The pinout for this chip is shown in Figure 12-2. Note that it has a supply voltage pin (pin 14) as well as a ground pin (pin 7) to supply power to all the gates on the chip. Each logic gate is numbered 1–4, and the inputs are labeled A and B with the output of Y.

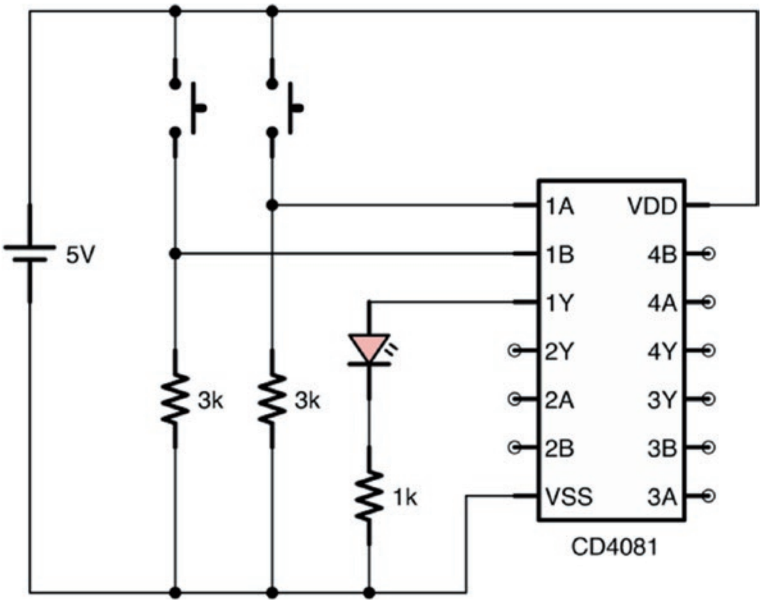

Figure 12-3. *Example Circuit Using an AND Gate*

To use the chip, you pick which one of the four gates you are going to use (it doesn't really matter which). If we want to use Gate 1, then we put our inputs on 1A and 1B, and then our output signal goes to 1Y. Note that, unlike the IC from the last chapter, this logic gate has a powered output—it actually supplies voltage and current to drive a (small) output signal.

Logic gates are wired to expect relatively fixed, predefined voltages on their inputs and output the same voltage levels. They do not need current-limiting resistors for their inputs because the inputs themselves are usually high resistance (i.e., 1,000,000–10,000,000 Ω). Because of the high resistance on the inputs, it also means that even if there is a resistor on the input, it will not affect the input voltage significantly. It also means that the current going into the gate is essentially ignorable: $I = \frac{V}{R} = \frac{5}{10{,}000{,}000} = 0.0005\text{mA}$.

For some logic chips, the input voltage is expected to be around 3.3 V or 5 V, while for others, it is based on the supply voltage. However, nearly all ICs are limited in how much current they can put out before they fry. This is usually somewhere in the range of 8–20 mA, depending on the chip. Because of this, if you use a logic gate to directly power a device (such as an LED), you probably will need a current-limiting resistor to keep the output current down below these limits.

There are logic chips that have open collector outputs (like the LM393 from Chapter 11), but they are more rare because they are harder to use.

Let’s say that we want to build a circuit which will turn on an LED if *both* of two buttons are pushed at the same time. Figure 12-3 shows a circuit to accomplish this. It has two buttons, one wired to 1A (pin 1) and one wired to 1B (pin 2). The output 1Y (pin 3) then goes to an LED with a current-limiting resistor. You may wonder what the resistors attached to the buttons are doing. Those will be explained in Section 12.3, “Pull-Down Resistors.”

For most logic chips, the manufacturers recommend that unused inputs (but not outputs!) be connected to ground. This makes the chip more efficient in power consumption, but for simple projects like these, it isn’t really necessary. If you wish to connect the unused inputs to ground, then it is a better circuit design.

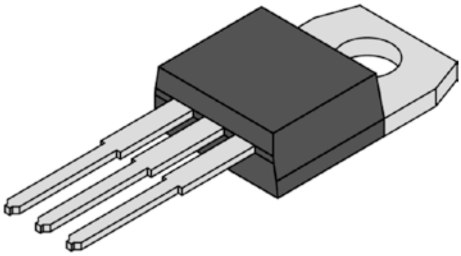

Figure 12-4. *A 7805 Voltage Regulator in a To-220 Package*

Note that the circuit shows a 5 V source. While the CD4081 is tolerant of a wide range of input voltages and would operate just fine at 9 V, many digital logic chips are not. Many digital logic chips operate at pre-specified voltages, usually either 5 V or 3.3 V. Therefore, we will take a moment and look at how we can get an input source for a specific voltage.

12.2 Getting a 5 V Source

So far in this book, we have used a custom power regulator to regulate the voltage on the board. Thankfully, this voltage is actually selectable—jumpers on the board allow you to select 5 V or 3.3 V, which are precisely the voltages usually required.

However, if we didn't have such fancy regulators, how would we achieve a 5 V source? There are several options for doing this, all depending on your requirements and/or the supplies you have available to you.

One option is to build a simple 5 V power supply using the knowledge you already have. In Chapter 9, we showed how to build a voltage divider to step down the voltage from a higher voltage source to a lower one. Although not ideal, this could work fine for simple test circuits. A better option would be to build the Zener diode voltage regulator that was shown in Chapter 8 if you have a 5 V Zener diode handy (however, you would need to be careful to calculate the output wattage, as these can dissipate a lot of power).

A better option is to use a voltage regulator IC. The LM7805 is a simple voltage regulator circuit you can use to convert a 7–24 V voltage source into a 5 V voltage source with somewhat minimal current loss. It is itself an IC, though with a different kind of packaging than we've seen, known as a **TO-220 package**. You can see what this looks like in Figure 12-4. On these packages, if you are reading the writing on the package, pin 1 (input voltage) is on the left, pin 2 (ground) is in the middle, and pin 3 (output voltage) is on the right. Figure 12-5 shows what this looks like in a circuit diagram.

Figure 12-6 shows how to attach the LM7805 regulator to your breadboard. First, plug the regulator into your breadboard so that each pin is on its own terminal strip. Next, plug the positive wire from the battery into the terminal strip with the voltage regulator's pin 1 and the negative wire from the battery to the negative/ground power rail on the breadboard. Then, connect the voltage regulator's pin 2 (ground) to the negative/ground power rail. Finally, connect the voltage regulator's pin 3 (output voltage) to the positive power rail.

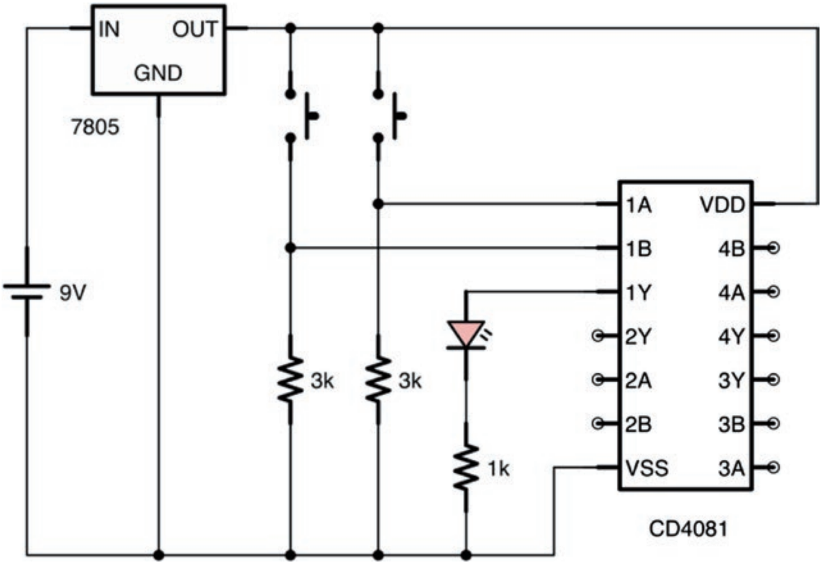

***Figure 12-5.** Logic Gate Circuit with a Voltage Regulator*

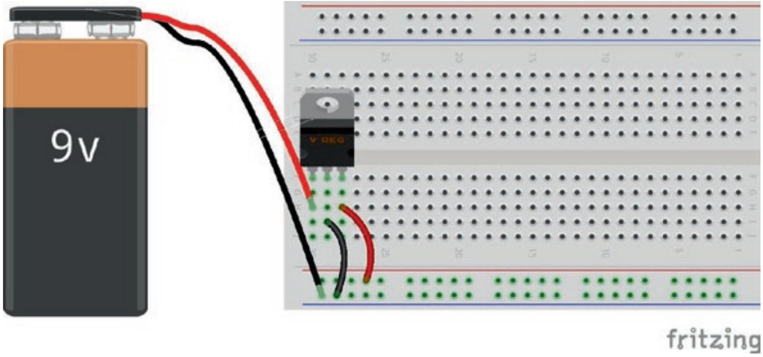

Figure 12-6. *Simple Way to Attach the LM7805 to Your Breadboard*

This will give you a 5 V supply without all of the circuitry of our previous power regulator. This is especially helpful if you have strange power needs or need to use a minimum of space for your power regulation.

Note that some LM7805s have pins that are too big for breadboards. That's unfortunate, but they are pretty rare. As long as you buy from companies that target hobbyists, you are likely to get a component that will work well with breadboards.

In any case, we will continue to use our power regulators, but I wanted to note that there are other ways to supply the correct amount of power to chips that require fewer components.

12.3 Pull-Down Resistors

In Figure 12-3, we looked at the circuit diagram for a simple AND gate. We noted that each button had a resistor connecting it to ground, but we did not mention why. In digital logic circuits, buttons and single-pole switches, when they are open, essentially disconnect the circuit. Because the inputs are high-resistance inputs (i.e., they use very little current), simply disconnecting the input circuits is not always enough to turn them off! Think of it this way—when you connect the circuit by pushing the button, the whole wire becomes positive. When you let go of the button, the state of the wire has not changed. Eventually the positive charge will drain out through the gate, but, since the input uses so little current, it may take a while for that to happen. Therefore, we have to provide another path for the electricity to go when the button is not pressed. These resistors are called pull-down resistors because, when the button circuit is not connected, they pull the voltage level down close to zero.

The resistor is very important because, when the button is connected, it keeps the voltage high and limits the amount of current that leaks out across the resistor. If you directly connected the button to ground without the resistor, then pushing the button would not raise the voltage because it is still directly connected to ground and would therefore remain at 0 volt. Having the resistor there makes sure that the voltage on the inputs remains high while the button is pressed and bleeds off when the button is released.

In short, without a path to ground, when you let go of the button, the input could remain high. Moreover, without the resistor, pushing the button would cause a short circuit. Therefore, a pull-down resistor allows voltage to drain off quickly when the button is not pressed, but also prevents disasters and wasted current when the button is pressed.

Additionally, static electricity in the air can cause the actual voltage on the gate to fluctuate when it is not connected. Physically connecting the gate to the ground through a resistor will make sure there is always *some* complete circuit which provides a deterministic value for the voltage at the gate.

The value of a pull-down resistor is usually somewhere between 1 kΩ and 10 kΩ. Beyond 10 kΩ, the actual function of pulling the voltage down to zero can be slowed down. Additionally, even above 4 kΩ, it is possible to interfere with the actual logic operation of the chip. Having a resistor below 1 kΩ, however, means that you are just wasting current.

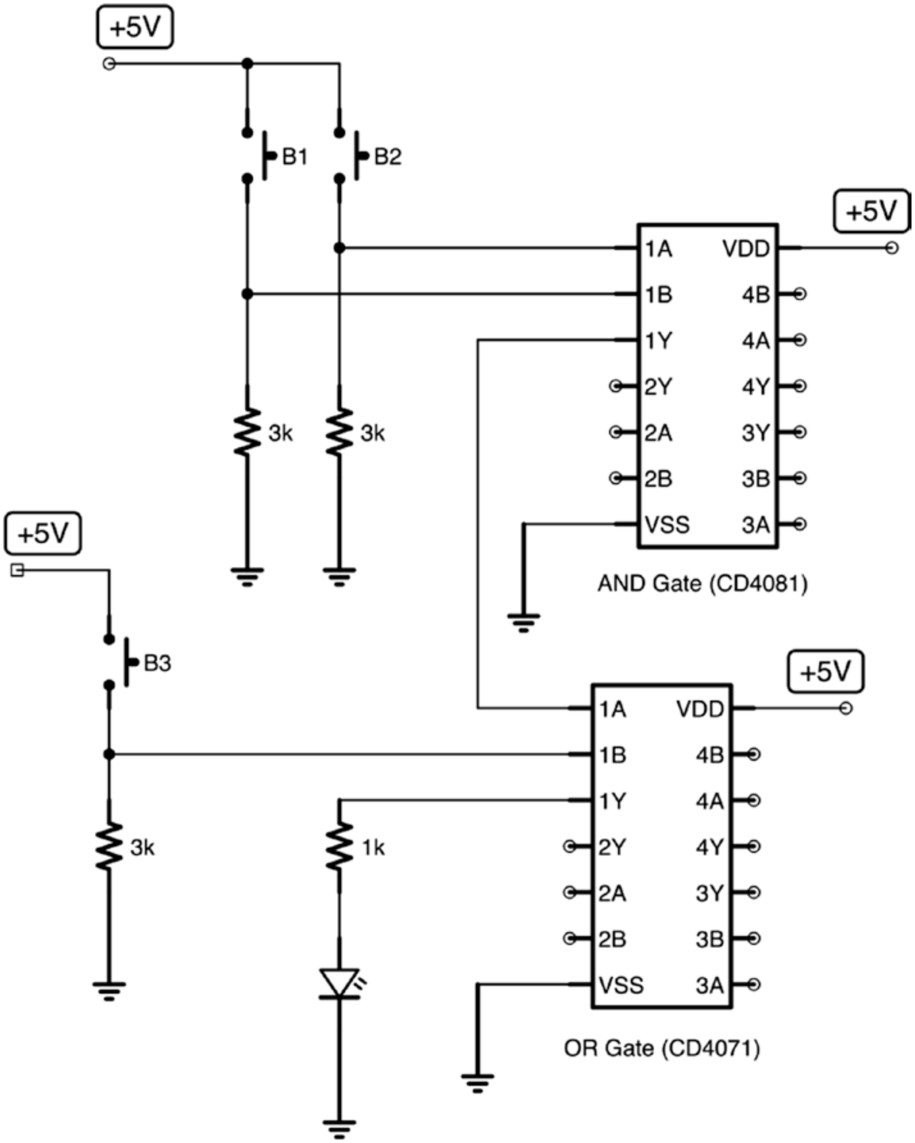
+5V
B1
B2
+5V
1A
VDD
1B
4B
1Y
4A
2Y
4Y
2A
3Y
2B
3B
VSS
3A
3k
3k
AND Gate (CD4081)
+5V
B3
+5V
1A
VDD
1B
4B
1Y
4A
2Y
4Y
2A
3Y
2B
3B
VSS
3A
3k
1k
OR Gate (CD4071)

Figure 12-7. *Multiple Logic Gates Combined in a Circuit*

So, for any button-type input to a digital logic circuit (where the circuit is *phyiscally disconnected* when the input is off), a pull-down resistor is needed to make sure that the input *actually* goes low when the button stops being pushed or the switch turns off.

12.4 Combining Logic Circuits

Logic chips that operate at the same voltage are very easy to combine together. Let's say that you had three buttons that you wanted to monitor and you wanted the light to come on if someone pushed either buttons 1 *and* 2 together *or* button 3 (or all of them). To do that, you would need an AND gate and an OR gate. Buttons 1 and 2 would be wired with the AND gate, and button 3 would be combined with the output of the AND gate through an OR gate.

Figure 12-7 shows what this looks like. Since there are so many voltage/ground connections, the figure does not have an explicit battery drawn; instead, it simply shows +5 V wherever it should connect to the voltage source and a ground symbol wherever it should connect to the battery negative. As you can see here, there are two logic ICs—the CD4081 having the AND gate and the CD4071 having the OR gate. The output of the first AND gate is wired into one of the inputs of the OR gate.

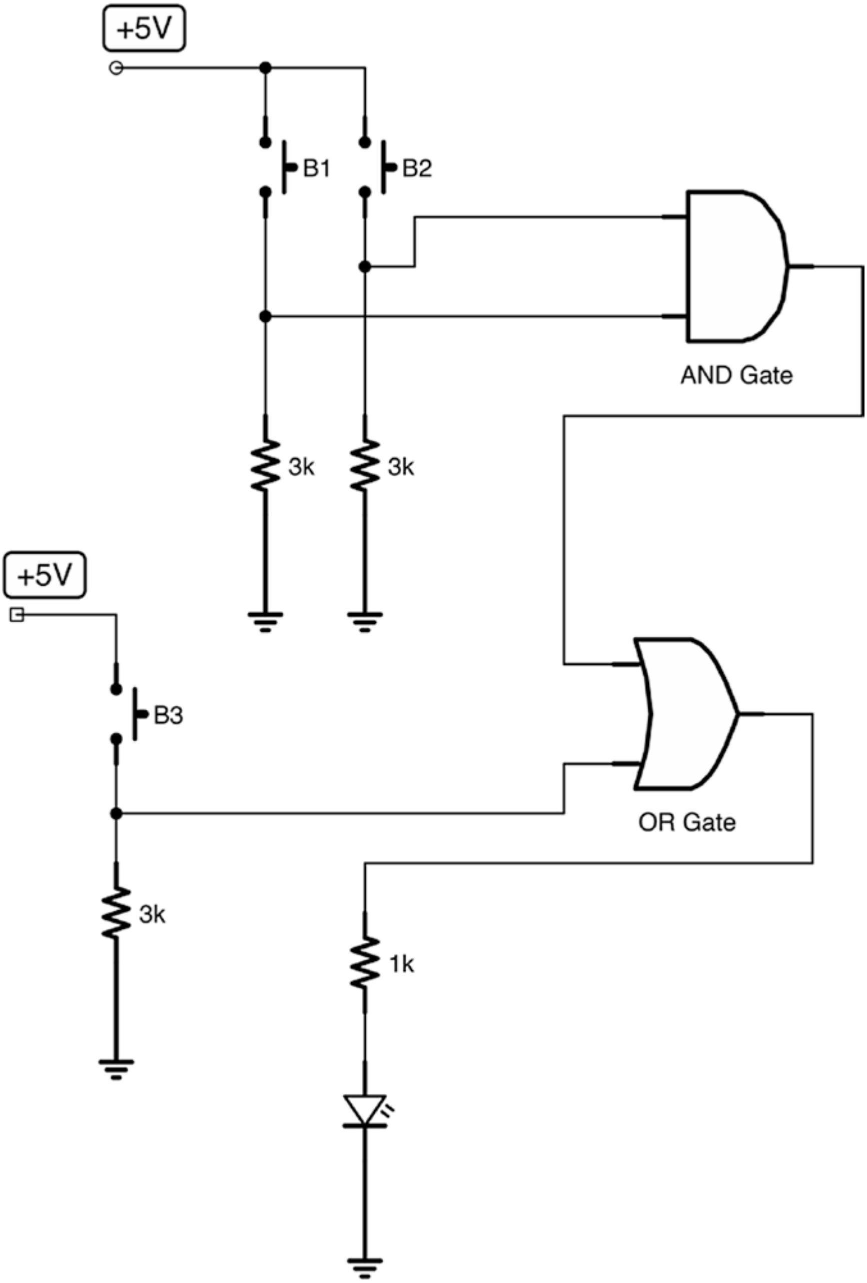
+5V
B1
B2
AND Gate
3k
3k
+5V
B3
OR Gate
3k
1k

Figure 12-8. *Logic Gates Represented as Shapes Instead of IC Pins*

This works because, unlike the LM393 (discussed in Chapter 11), these logic gates actually supply output voltage and current as well. Because the inputs to the logic gates are high resistance (they act like there is an extremely large resistor connected to the input), there is no need for a current-limiting resistor when combining gates in this way.

Now, it is fine to draw logic circuits the way we have in Figure 12-7. However, as the logic becomes more complex, actually drawing all of the connections to voltage and ground becomes tiring, and trying to get all of the wires to the right spot on the chip can get messy as well. Because of this, engineers have devised a simpler way of describing logic gates and logic circuits in schematics.

Instead of representing the entire chip on a schematic, engineers will represent only the logic gates themselves.

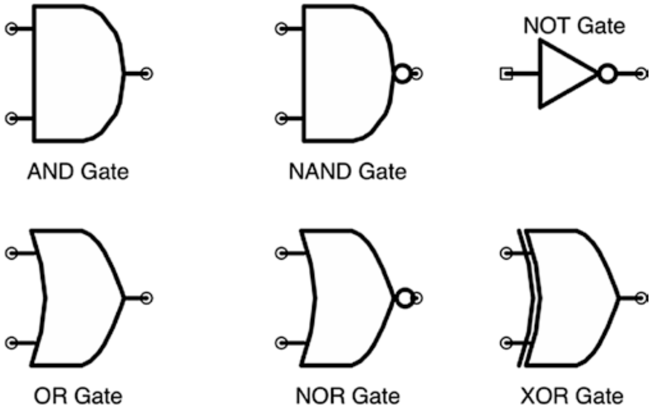

Figure 12-9. *Common Gates Used in Schematic Drawings*

Additionally, since the power goes to the whole chip (and not the individual gates), in such a drawing, the power connections for the gates are not shown. The standard that was developed represents each type of gate with a shape. Figure 12-8 shows what this circuit drawing looks like if it is drawn using shaped gates instead of IC pins. The actual physical circuit is the same; this is only to simplify the schematics to make them easier to understand and follow.

Figure 12-9 shows what these gate drawings look like. The AND gate has a flat back panel and a simple, rounded front. The OR gate looks a bit like a space shuttle, with both the back and the front angled. The NOT gate is a triangle with a circle at the tip. This circle can also be added to other gates to show that the gate is the opposite one. For instance, a NAND gate is drawn by first drawing an AND gate and then adding a circle to the front, indicating that the gate behaves like an AND gate with a NOT gate in front of it. Similarly, the NOR gate is an OR gate with a circle in front of it. The XOR gate is similar to the OR gate, but with an extra line going across its inputs.

Many times, the internal schematics of a chip are shown using gate symbols, in order to help you understand the operation of the chip and how the pins work. For instance, Figure 12-10 shows how the CD4081 chip is wired up internally. You can see the inputs going through the logic gate and out toward the output. While this isn't any new information you didn't already know, if may help you understand why the pins are laid out the way that they are.

As an interesting sidenote, every logic function can actually be built from NAND gates, though you have to wire them up in strange ways. You can actually build a computer almost entirely from NAND gates if you wanted to. It is not incredibly important, but Figure 12-11 shows how to build each type of logic gate from NAND gates. As an activity, go through the truth tables in Figure 12-1 and see if you can follow how each set of values becomes the result.

12.5 Understanding Chip Names

One of the biggest problems in learning to build electronic devices is the bewildering array of chips, each with some weird name. "Oh, for that you want an NE555P," or "You could use a SN74HC00P or a CD4011BE for that task." What language are such people speaking?

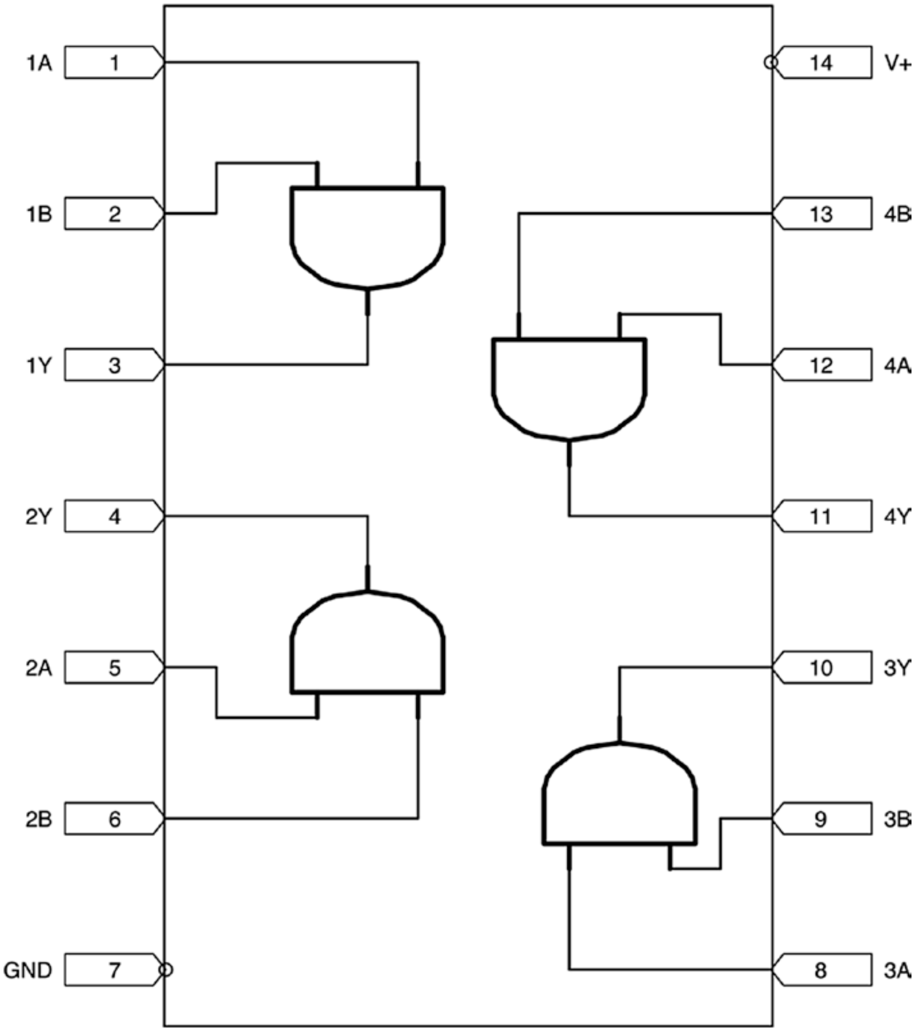
1A
1
1B
2
1Y
3
2Y
4
2A
5
2B
6
GND
7
14
V+
13
4B
12
4A
11
4Y
10
3Y
9
3B
8
3A

Figure 12-10. *The Internal Layout of the CD4081*

There is a huge selection of chips available, and learning their names is a daunting task. Appendix C attempts to offer some method to the madness, but, at the end of the day, chip names are like people's names—you get to know them by using them. Nobody knows everybody's name, but, for the types of projects you like to work on (whatever that happens to be), there will be standard chips whose names you will eventually come to know.

Once you make it through this book, you should have a solid enough background to search for the chips you need, have some understanding of the part names, and be able to find the chips you need for your projects. If you buy the chips from a seller geared toward amateurs and hobbyists, they will likely also include tutorials and additional information available in an easier-to-understand format than just datasheets.

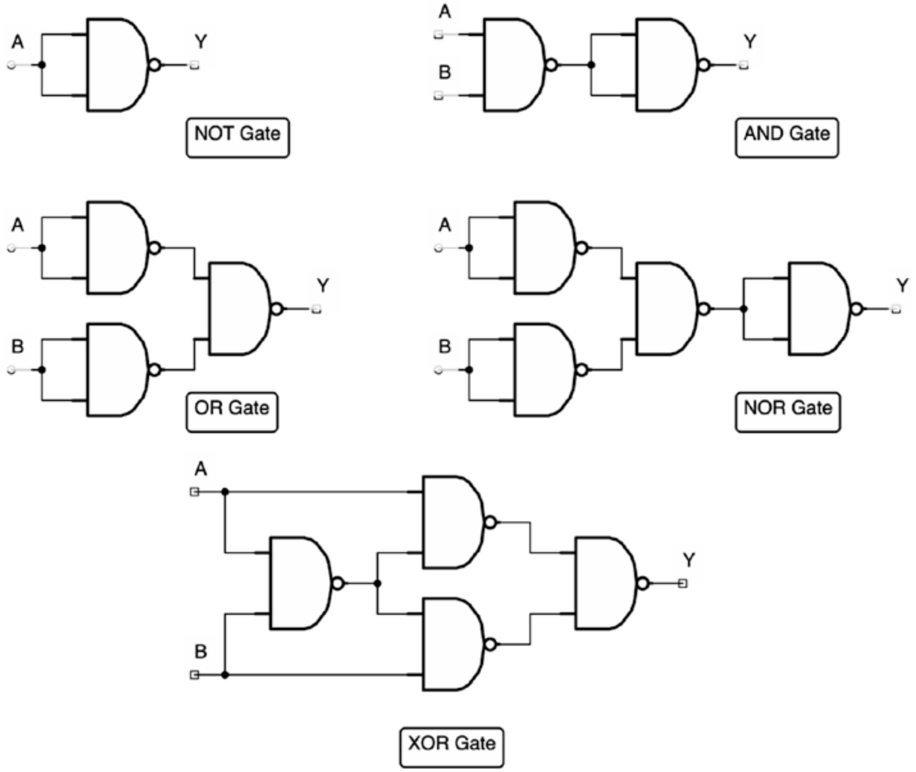

Figure 12-11. *How Each Gate Can Be Built from NAND Gates*

Review

In this chapter, we learned the following:

1. A logic IC implements basic digital logic functions such as AND, OR, NOT, and so on.
2. These logic functions refer to essentially the same ideas that they mean in ordinary language and exactly what they mean in formal logic.
3. Logic ICs use different voltage levels for true and false—usually with true being near the supply voltage and false being near zero voltage.
4. True is sometimes referred to as HIGH or 1. False is sometimes referred to as LOW or 0.
5. The inputs of a logic function are usually designated as A and B, and the output is usually designated as Y or Q.
6. If an input or output name has a line/bar printed on top of it, that means that the value is the logical *opposite* of what would be expected. So, if Y is the output value, then $\overline{Y}$ would be the opposite of the output value.
7. A single digital logic function is called a gate. Most logic ICs have more than one gate on a single chip.
8. Logic ICs require a supply voltage and a ground connection to power the logic.
9. Most logic ICs provide powered logic outputs, so that a “true” value supplies both voltage and a small amount of current on its output. However, a current-limiting resistor is usually required.

10. If an input to a logic IC may be disconnected for its “false” state (as is common with button inputs), then it needs a pull-down resistor to connect it to ground when the button is not being pushed.
11. Logic ICs can usually be combined by wiring the output of one to the input of another to create more complex logical conditions.
12. Logic gates are often drawn in schematics using basic shapes to indicate their operation, rather than as connections to chips. In these cases, the power connections are not shown by schematics.
13. Every logic gate can actually be built from NAND gates wired together.
14. IC names are very confusing and take time and experience to get to know them well.
15. Many ICs require specific voltage levels to operate, often at 5 V or 3.3 V.
16. Many solutions are available for generating specific voltages, including voltage dividers, Zener diodes, voltage regulators, and add-on breadboard power units.
17. The LM7805 is a very common 5 V voltage regulator.

Apply What You Have Learned

1. Draw the circuit in Figure 12-3 yourself. Identify the function of each resistor.

2. Build the circuit in Figure 12-3 (don't forget that the power source should be 5 V).
3. If you assume that a negligible amount of current flows through the inputs of the AND gate and that the output functions as a 5 V voltage source (and the LED is red), how much current flows through each resistor when all of the buttons are pressed? What, then, is the total current used by the circuit if you ignore the logic gate?
4. Measure the actual current that flows through each resistor. If you are having trouble pushing the buttons while you measure the current, just replace the buttons with wires for this test.
5. Measure the current that is used by the AND gate itself. You can do this by measuring the supply current of the AND gate. Measure it both when its output is true and false.
6. Draw a schematic of a circuit that has two buttons (B1 and B2) which light up an LED if either button is pressed. Use the logic gate shapes for the schematic.
7. Draw a schematic of a circuit that has two buttons (B1 and B2) which light up an LED if neither button is pressed. Use the logic gate shapes for the schematic.
8. Draw a schematic of a circuit that has four buttons (B1–B4) which light up an LED either if B1 and B2 are pressed or if B3 and B4 are pressed. Use the logic gate shapes for the schematic.

9. Look at the construction of the different gates from NAND gates in Figure 12-11. Copy down the OR gate construction four times, and trace how the output is generated for each possible set of inputs (true/true, true/false, false/true, false/false). Show the inputs and outputs on each NAND gate. Compare the outputs to the truth table for the OR function in Figure 12-1.
10. Take the circuit in Figure 12-3 and draw a schematic to use pull-up resistors on the inputs rather than pull-down resistors. How will this change the behavior of the circuit?
11. Let's say that we want to create a door buzzer so that someone outside a door can push a button to be let in. However, the person inside also wants a switch to be able to disable the buzzer. The buzzer can be thought of as a simple device that buzzes when any positive voltage is applied. Draw a circuit diagram of this setup using logic gates. The buzzer can be drawn as a resistor labeled "buzzer" (don't forget to connect the other side to ground!).

CHAPTER 13

Introduction to Microcontrollers

In Chapter 12, we learned the basics of digital logic. However, I think we can all agree that those chips wound up taking up a lot of space on our breadboards. If we wanted to do a lot of complicated tasks, we would wind up needing a lot of chips, we would need more and more breadboards to put them on, and our project would get unwieldy very quickly. Additionally, as the number of chips increased, it would get very expensive to build such projects.

Additionally, when all of the logic of a circuit is hardwired into the circuit through logic chips, it is very difficult to change. If you need to change the logic (convert an `AND` gate to an `OR` gate or do anything else), you wind up needing to wade through masses of circuitry to make the change you want. Then, if you are mass-producing the circuit, you have to set up for mass production all over again.

To solve all of these problems (and more), the microcontroller was introduced. A microcontroller is essentially a low-power, single-chip computer. A "real" computer chip usually relies on a whole slew of other chips (memory chips, input/output chips, etc.) to operate. A microcontroller contains all of these (though usually on a smaller scale) in a single chip that can be added to an electronics project.

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_13

Unlike your typical computer, most microcontrollers can't be connected to a mouse, keyboard, or other typical inputs and can't be connected to a monitor, disk drive, or other typical outputs. Instead, microcontrollers usually communicate entirely through digital (true/false) electrical signals on their pins.

So, instead of wiring complex logic onto their boards, many people opt to have microcontrollers provide the bulk of their digital logic. One complication that this adds, however, is that since the microcontroller is essentially a computer, then just, like a computer, it has to be programmed. This means that not only must circuit designers be familiar with electronics, they must also be familiar with computer programming.

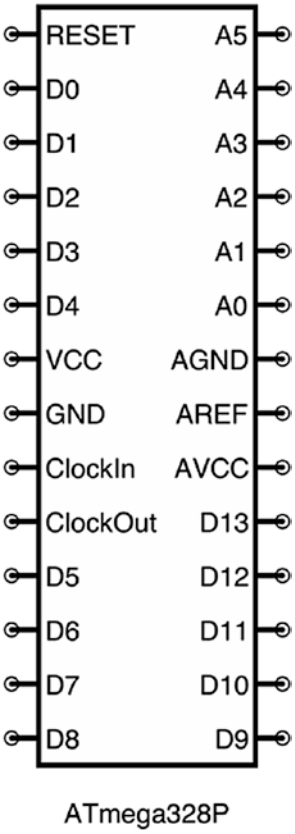
RESET
D0
D1
D2
D3
D4
VCC
GND
ClockIn
ClockOut
D5
D6
D7
D8
A5
A4
A3
A2
A1
A0
AGND
AREF
AVCC
D13
D12
D11
D10
D9
ATmega328P

Figure 13-1. *A Simplified Pinout of the ATmega328/P*

13.1 The ATmega328/P Chip

The microcontroller we will be focusing on is the ATmega328/P. Actually, we will focus less on this specific chip than the overall environment surrounding it, known as Arduino (we will cover more of what this means in Section 13.2, "The Arduino Environment").

Even though the Arduino environment tries to minimize how much you have to know about the hardware itself, it is good to have a quick introduction to this chip and how it works.

The ATmega328/P is part of a family of microcontrollers developed by Atmel known as the AVR family. The AVR became popular because it was one of the first chips to use flash memory to store its programs, which allowed the onboard programs to be more easily changed.

Figure 13-1 shows a simplified pin configuration of the chip, focused on how it is used in the Arduino environment. The VCC pin and the GND pins are the primary power pins. The chip can run on a range of voltages, but 5 V is a very common and safe setting. AVCC and AGND power the chip's analog-to-digital converter unit.

All of the pins labeled "D" are digital input/output pins. They can be configured as inputs for buttons or other signals or as outputs for driving LEDs or other output devices. The pins labeled "A" are analog input pins (we will cover how to use these in Chapter 15). While the digital input pins can only read that a value is true/false, the analog pins can read voltages and convert them to numbers. AREF is a "reference voltage" used for setting the maximum voltage for analog inputs, but is usually unconnected (it should also not be a higher voltage than AVCC).

Microcontrollers, like most processors, control their operation by using a "clock." This is not a clock as you normally think of—it doesn't tell time. Instead of thinking of it as a clock, a better way to think of the "clock" is as a heartbeat. Basically, there is a continuous signal of pulses that are provided through the clock, and the pulses allow the chip to synchronize all of its activities. The ATmega328/P has an internal clock, but it can also be more efficiently operated by connecting an external clock (quartz crystals, for instance, provide a *very* steady pulse for this purpose).

As we have noted, the chip has a number of input and output pins. You might wonder, what does the chip *do* with its input and output pins? That is entirely up to you. It does *whatever you program it to do*. The ATmega328/P has **flash memory** on the chip which can store a computer program (flash memory means that it will remember the program even after the power turns off). You have to upload your program to the chip, and then after that, it will do whatever you like with its inputs and outputs. The D0 and D1 pins, in addition to providing input and output, can also be used to reprogram the chip. We will learn how to program the chip in Section 13.4, "Programming the Arduino."

13.2 The Arduino Environment

The chip itself is just one piece of the puzzle. In order to use the chip, you have to be able to program it. Programming requires the use of programming tools on your main computer. In addition, you also need some way to take the program that you built on your computer and load it onto the chip. That takes both software and hardware.

Then, once the program is on the chip, you have to build a circuit to properly power the chip. This requires voltage regulation for the VCC pin and several other recommendations from the manufacturer about how to set up the other pins. To make it run optimally, it would also be good to supply an external clock. All of this can be quite a lot of work, and a lot of pieces that need to be brought together.

Thankfully, most chips have what is called a **development board** that can be purchased. A development board is a prebuilt circuit that has a microcontroller chip pre-connected in its recommended manner. It is made to simplify the work of developing circuits. Likewise, most chips have a recommended **programming environment** as well. A programming environment is a set of tools for your computer that allow you to create programs for your microcontroller. Additionally, a device called an **in-system programmer** (ISP) connects your computer to your chip or development board and will transmit the program from your computer to the chip.

In 2005, a complete, simplified system for doing all of these tasks called **Arduino** was created, based off of an earlier system called Wiring. Arduino consists of (a) a simplified development environment for your computer to write software for microcontrollers, (b) a simplified development board to make it very easy to build electronics projects, and (c) integrating the in-system programmer into the development board so that all that is required to transfer the program to the microcontroller is a USB cable.

The Arduino environment supports a number of different microcontroller chips. Because it is a simplified environment, many of the special features of individual chips are not directly supported. However, for getting started and doing basic projects, the Arduino environment is excellent.

Even though there is a company behind Arduino, there are many Arduino-compatible boards made by other manufacturers. These boards use the same ATmega328/P microcontroller and often have very similar development boards and functions. Most importantly, they are compatible with the programming tools on the Arduino environment.[1]

---

[1] One important difference between "official" Arduino boards and most Arduino-compatible boards is the USB controller chip. Official Arduino boards use a chip from FTDI to do this, because the drivers for it are preinstalled on every operating system. To save on costs, many Arduino-compatible boards use a lower-cost chip to perform this function (often the CH340). Usually, the only disadvantage of this is that, on some operating systems, an additional driver is needed to communicate with these chips.

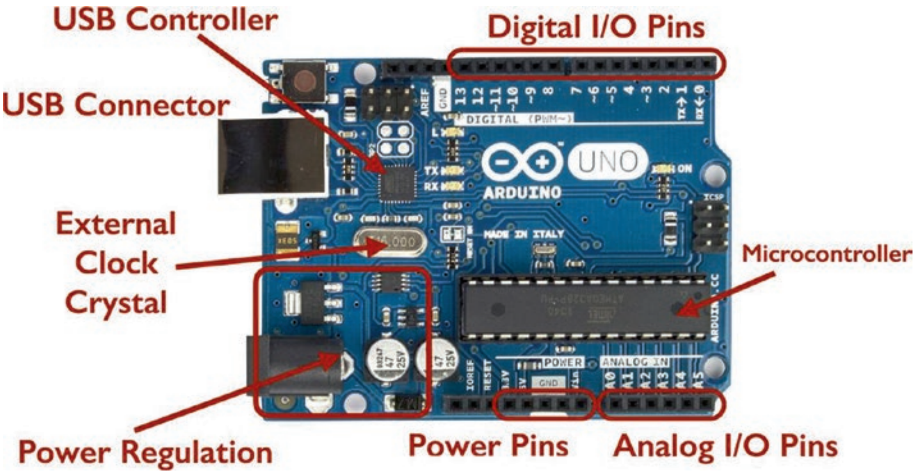
USB Controller
Digital I/O Pins
USB Connector
External Clock Crystal
Microcontroller
Power Regulation
Power Pins
Analog I/O Pins

Figure 13-2. *The Major Components of the Arduino Uno*

13.3 The Arduino Uno

This book focuses on the Arduino Uno development board. Figure 13-2 shows what the board looks like, as well as a general idea of what the different areas of the board accomplish. The Uno is very nice because the USB port allows a whole slew of functions—not only can it be used to receive programs for the chip but you can also power the board through the USB, as well as send data back and forth to the computer through it. If you are not connected to a USB, there is a separate power plug that can be plugged into the wall or to a 9 V battery.

The Arduino Uno provides chips for power regulation, USB communication, as well as the ATmega328/P microcontroller itself. It also supplies an external clock for the microcontroller. Finally, it provides **headers** (places on the board to plug in wires) for the major pins on the microcontroller. Thus, everything you need to make use of the chip is provided for you in this development board. You simply connect the input and output pins to your own breadboard, and you can have a working project that you can program.

13.4 Programming the Arduino

Now that we have seen the pieces of the Arduino environment, let's tackle programming the Arduino. This book is not a book about programming, so we will only cover the absolute basics.

The programming tool for the Arduino is called the **Arduino IDE**. IDE stands for "Integrated Development Environment"—in other words, the thing that you develop with. The Arduino IDE is available for pretty much any computer—Mac, Windows, or Linux. You can download the IDE from `http://arduino.cc/`.

Depending on whether you purchased an "official" Arduino or a clone, you may also have to download an additional driver for the USB interface. If you have an Arduino-compatible clone, you can download the driver from `http://bplearning.net/drivers`. Once you have installed the Arduino IDE and the USB driver, you are ready to start!

We will begin by using an example program (called a "sketch" in Arduino terminology) that ships with the Arduino. First, connect your Arduino to your computer with a USB cable. Open up the Arduino IDE; then click "File," then "Examples," and then "01.Basics"; and finally click "Blink." This loads up a ready-made program for your Arduino. This program simply turns the D13 pin on and off. On an Arduino, the D13 pin already has an LED attached to it, so you don't even need to add any components!

Now that you have the program loaded up, click the button with the checkmark icon. This verifies that the program is written in a way that the computer can understand. If it has any errors, it will show them in the black panel on the bottom.

Now you need to make sure that the IDE is targetted at your board. Go to "Tools" and then "Board" and make sure that "Arduino/Genuino Uno" is selected. Then, click "Tools" and then "Port" and make sure your Arduino was detected and that it is selected. If you don't see your Arduino listed here, you may need to check the USB driver installation.[2]

Once your configuration is verified, click the button with the arrow icon to upload it to the Arduino. It should take about 2–5 seconds, and then the LED on your Arduino should start blinking. If there are any errors, they will display in the black status area below. Note that some Arduinos come with this program preinstalled. If this is the case, your Arduino may not have changed what it is doing much. You can verify that it is all working by changing the program slightly. If you change all of the numbers that say `1000` to `500`, it should blink twice as fast. Remember, though, that you must verify it (click the checkmark button) and upload it (click the arrow button) to get your new code onto the board.

Now, let's take a look at the code and how it functions. The first part of the code should be grayed out. That's because it is a **comment**, or a note telling you about the program. This isn't read by the computer at all. Comments start with the characters /*, and they continue until they reach the characters */ (this can even span multiple lines). Shorter comments are sometimes made with the characters //. Those comments only continue until the end of the line.

After the comments, there are two **functions** defined—`setup()` and `loop()`. A function is simply a piece of code that is named. In the Arduino environment, the `setup` and `loop` functions are special. The `setup` function runs exactly once when the chip first turns on or resets or when a new program is loaded. It is used for things such as telling the chip which pins will be used for input and which ones will be used for output and doing other setup-related tasks. After the `setup` function completes, the `loop` function runs continually over and over again for as long as the chip is on.

---

[2] You might try the following instructions with a device listed there even if you don't recognize it as your Arduino in case your computer had difficulty detecting the device type.

If you look at the code, the `setup` function contains one command:

```
pinMode(13, OUTPUT);
```

This tells the microcontroller that digital output 13 (D13) will be used for output. Note that this refers to the D13 pin in Figure 13-1 (usually just labeled `13` on the Arduino Uno), not to pin 13 of the chip itself (which would be D7). With the Arduino Uno, we don't actually need to worry about the pinout of the ATmega328/P, we just need to read the names of the pins next to the pin headers.

The `loop` function is the main part of the code. It looks like this:

```
digitalWrite(13, HIGH);
delay(1000);
digitalWrite(13, LOW);
delay(1000);
```

The first line, `digitalWrite(13, HIGH);`, says to turn D13 to HIGH, which is about 5 V. This provides the power for the LED attached to D13. This pin will remain high until we tell it to do something else.

The next line, `delay(1000);`, tells the chip to wait for 1,000 milliseconds (which is 1 second). During this time, nothing happens—the chip just waits. Changing this number changes the amount of time that the chip will wait for. Changing it to `500` will cause it to wait for half a second, and increasing it to `2000` (note that there are no commas in the number!) will cause it to wait for 2 seconds.

The next line turns D13 to LOW/false/off/0 V. This turns off the LED, because there is no longer any voltage supplied to it. D13 will stay in this state until told otherwise. The next line then waits for 1 second. Once this function finishes, the chip will simply run the `loop` function again from the start.

In Chapter 14, we will look at how to build projects with the Arduino that use additional components.

Review

In this chapter, we learned the following:

1. A microcontroller is a small computer in a single microchip that provides customizable logic for handling digital signals.
2. A development board is a circuit board that simplifies the process of building circuits with a microchip by providing most of the standard connections for you, allowing you to focus your efforts on the things that make your project distinctive.
3. In order to use a microcontroller, it has to be programmed from a computer.
4. The Arduino environment is a combination of software and hardware meant to make building microcontroller projects easier.
5. The Arduino Uno is a development board for the Arduino environment that includes a microcontroller, USB connection, power regulation, and headers for connecting the microcontroller's input/output pins to other circuits.

6. The ATmega328/P is the microcontroller used in the Arduino Uno.
7. An Arduino program (called a sketch) has two standard functions—`setup` (which is run once when the chip powers on) and `loop` (which runs over and over again as long as the chip is on).
8. Once a program is uploaded to the Arduino Uno, it will be saved on the device until another program is loaded.

Apply What You Have Learned

1. Practice modifying and uploading the Blink program to the Arduino Uno. Change the numbers given to `delay` to different values, and see how that affects the operation of the program.
2. The ATmega328/P is only one of many different microcontrollers available in the AVR family. Research online to find one or two other AVR chips and what different features they have.
3. The AVR family is only one of many microcontroller families. Research one or two other microcontroller families and look at what features are claimed for each. Examples of other microcontroller families include PIC, STM32, MSP432, and the Intel Quark.
4. Go to the `arduino.cc` website and look at the different types of Arduinos that are available. What makes them different? Why might you choose one for a project over another?

CHAPTER 14

Building Projects with Arduino

Chapter 13 covered the basics of what microcontrollers are, what the Arduino environment is, and how to load a program onto an Arduino board. In this chapter, we will go into more depth on how to include an Arduino Uno into a project.

14.1 Powering Your Breadboard from an Arduino Uno

The first thing to understand about the Arduino Uno is that one of its main jobs is power regulation. As we saw in Chapter 13, the Arduino can use a variety of power sources—USB, battery, or a wall plug. Additionally, there is a connection on the Uno's headers that allows for you to supply power from some other source.

If you have a power source that just has wires coming out of it—like a 9 V battery with a simple wire connector—the Uno has a place that you can plug it in. The pin labeled V_{IN} is used for supplying an unregulated voltage supply (7–12 volts) to the Uno (*do not* use the one labeled 5 V—it is for *outputting* power). Therefore, if you plug the positive wire into V_{IN} and the negative wire into any of the *GND* pins (it doesn't matter which one), then the Uno will spring to life.

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_14

On the flip side, you can actually power the rest of your project from the Uno and take advantage of its voltage regulation, as well as its numerous methods for getting power. To do this, simply take a wire from the 5 V connection on the Uno and connect it to the positive rail on your breadboard. Then, take another wire from one of the GND connections on the Uno and connect it to the ground rail on the breadboard. Voilà! A very flexible 5 V power supply for your breadboard. Figure 14-1 shows how to use this to light a simple LED circuit.

Also note that there is also a 3.3 V connection if you need it, as many small devices are powered at that level.

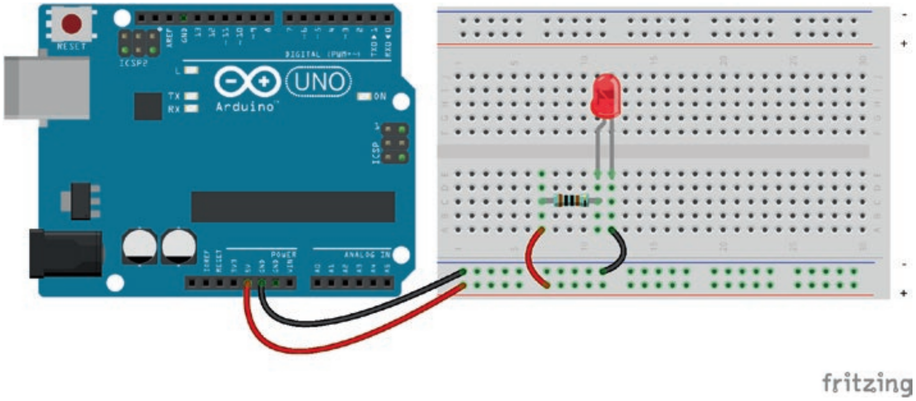

Figure 14-1. *Powering a Simple Project from an Arduino Uno*

14.2 Wiring Inputs and Outputs to an Arduino Uno

Now that we know how to power a breadboard from an Arduino Uno, we can now see how to connect inputs and outputs to the Uno.

Wiring inputs and outputs to an Uno is actually very easy. Outputs of the Uno can be viewed as simple voltage sources, like a battery, which operate at either 5 V (if set to HIGH) or 0 V (if set to LOW). Remember that any of the digital I/O pins can be set to be input or output pins in your Arduino program with the `pinMode` command.

However, even though an output pin can act as a voltage source, the current needs to be limited to prevent damage to the Arduino. Each output pin should only be sourcing up to 20 mA of current, and the total amount of output of all pins combined should never exceed 100 mA. So, for instance, if you have an LED output, be sure to add a resistor to limit the amount of current. Microcontrollers generally cannot drive high-power devices such as motors directly and must rely on some method of amplifying the signal after it leaves them (we will cover amplification in Part 4 of the book).

Inputs to an Arduino are essentially voltage sensors. They will detect a HIGH (around 5 V) or LOW (near 0 V) signal on the pin. You can think of them as having a very large resistor attached to them (about 100 MΩ—100 million ohms), so they don't actually use up any serious amount of current (in other words, you don't have to supply a current-limiting resistor because it is already built in). However, because they use so little current, that means that, just like our inputs in Chapter 12, they cannot be left disconnected, or the results may be randomized from static electricity in the air. Thus, for inputs, you should always attach a pull-up or a pull-down resistor (usually a pull-down) to the input to make sure that the input is *always* wired into the circuit in a known-valid state.

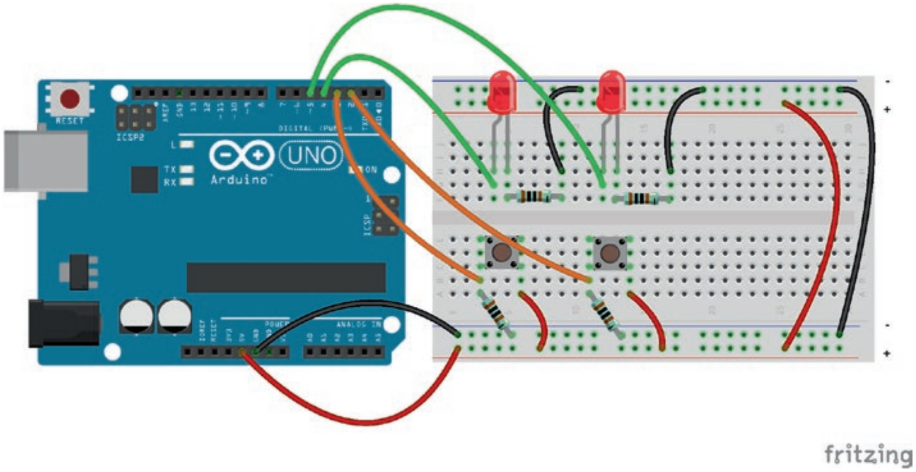

Figure 14-2. *Wiring a Simple Button-Based Arduino Project*

14.3 A Simple Arduino Project with LEDs

In this section, we are going to look at making a simple Arduino project with two buttons, each controlling one of two LEDs. This would actually be simpler to wire without the Arduino, but the goal is to make a baby step to understanding how Arduino projects work. Later we can do more complicated things, but, for right now, we will just see how we get an input signal to the Arduino and send an output signal back out.

This project is going to have buttons wired into Digital Pin 2 and Digital Pin 3 of the Arduino and LEDs wired into Digital Pin 4 and Digital Pin 5. Let's think about what these need to look like. The LEDs will each need a current-limiting resistor, and the buttons will each need a pull-down resistor.

Figure 14-2 shows how this should be wired up. The breadboard is being powered from the 5 V and GND terminals on the Arduino. The wires on the right side make sure power is connected to both sides of the breadboard. On the bottom, buttons are wired up with pull-down resistors and connected to Digital Pins 2 and 3 on the Arduino. On the top, the LEDs are connected to Digital Pins 4 and 5, with current-limiting resistors making sure they don't draw too much current.

Now, of course, for an Arduino, this is not enough. The Arduino also needs a program to control it! Figure 14-3 shows the program you will need to type in to control the LEDs.

Note that, as usual, the project is divided into two pieces—`setup()`, which only occurs once when the chip starts up, and `loop()`, which continuously runs over and over again until the chip is turned off or reset:

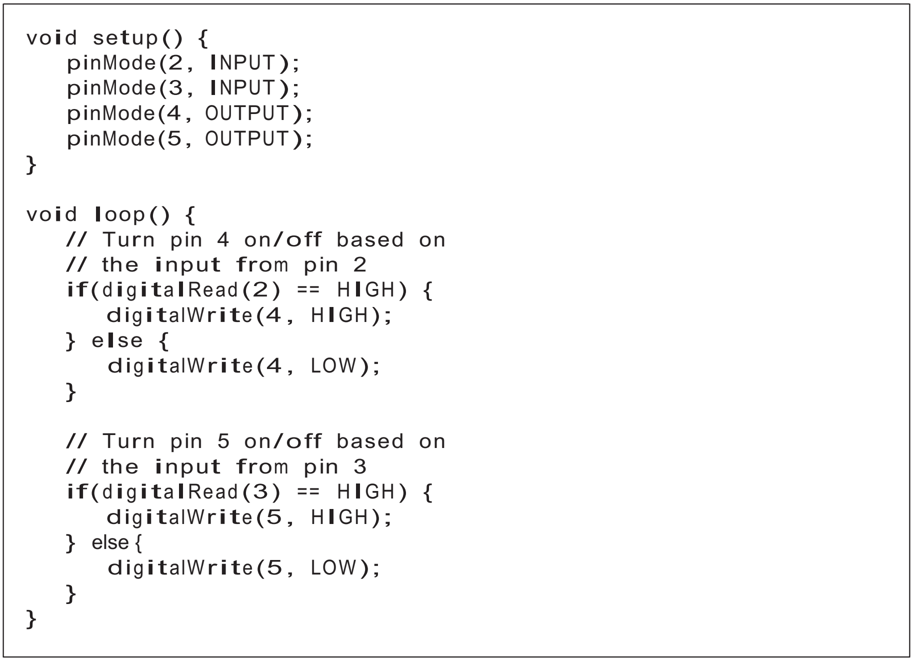

```
void setup() {
   pinMode(2, INPUT);
   pinMode(3, INPUT);
   pinMode(4, OUTPUT);
   pinMode(5, OUTPUT);
}

void loop() {
   // Turn pin 4 on/off based on
   // the input from pin 2
   if(digitalRead(2) == HIGH) {
      digitalWrite(4, HIGH);
   } else {
      digitalWrite(4, LOW);
   }

   // Turn pin 5 on/off based on
   // the input from pin 3
   if(digitalRead(3) == HIGH) {
      digitalWrite(5, HIGH);
   } else {
      digitalWrite(5, LOW);
   }
}
```

Figure 14-3. *An Arduino Program to Control Two Buttons and Two LEDs*

`setup()` simply tells which pins should be in which mode. Note that, unless it includes the calls to the `delay()` function, the `loop()` function will literally run thousands of times per second (or more). The Arduino Uno can execute approximately 16 million instructions per second. Each line of code translates to many instructions, but nonetheless, it goes *really fast*. Just keep this in mind when you are writing programs.

Inside the `loop()` function, we have a new Arduino function—`digitalRead()`. The `digitalRead` function takes a pin number and returns whether that pin is HIGH or LOW. We have put this into a conditional—*if* the read from the button pin is HIGH, then the corresponding LED pin is turned HIGH. Alternatively (i.e., *else*), if the read from the button pin is not HIGH (i.e., it is LOW), then the corresponding LED pin is turned LOW. Note that there are *two* equal signs used in the comparison. In many programming languages, you use two equal signs to tell the computer to compare values. A single equal sign often means that you are *setting* a value (not comparing them).

This book is not a book on computer programming, so we are not going to cover all of the details. Because of this, most of the programs will be given to you, and you will only need to make minor modifications. However, if you are interested in learning more, the programming language being used in the Arduino environment is C++, and the Arduino focuses on the easier-to-understand portions of it. The book *Beginning Arduino Programming* by Brian Evans is a good place to start. If you want to learn more about programming in general (not related to Arduino), you can check out my own book, *New Programmers Start Here*. If you want to understand better how programming works from the perspective of the computer, you can check out my book *Programming from the Ground Up*.

14.4 Changing Functionality Without Rewiring

Now, you might reasonably be thinking, *Wouldn't this be a lot easier if we just directly attached the buttons to the LEDs to turn them off and on?* Indeed, it would. However, by having the inputs and outputs all wired to the Arduino, we can actually *change* the functionality of the project *without* having to do a single bit of rewiring! For instance, if we wanted to have the left button control the right LED and the right button control the left LED, then all we would have to do is swap all of the 2s for 3s in the program and vice versa.

Now imagine if we had spent time developing such a device and even had it sent off to manufacturing, but later decided that we wanted to change the functionality. If all of the logic of the device is implemented by hardware, then that means that you have to throw away all of your old inventory to modify your functionality. If, instead, you use software to connect your components, then oftentimes you can update your device merely by updating your software.

Other modifications we can think of to this simple device might include the following:

1. Making the LEDs turn off rather than on when the buttons are pressed
2. Making the LEDs blink when the buttons are pressed
3. Changing the buttons to be simple toggles, so that you don't have to keep on holding the buttons down to keep the LED on

4. Requiring that the buttons be pushed in a particular order in order to turn on the LED
5. Requiring that both buttons be pushed to turn on the LEDs

This list could go on and on. By routing all control processing through your microcontroller, you make your devices much more flexible. Additionally, at some point, they also become cheaper. When mass-producing, microcontroller chips like the ATmega328/P can be had for just over a dollar. Some chips, when purchased in bulk, cost less than 50 cents! So, if a microcontroller is replacing a complex sequence of logic gates and other control functionality, moving all of your control logic to a microcontroller can actually be much less expensive than hardwiring it, and you get added flexibility as a side bonus.

Review

In this chapter, we learned the following:

1. You can power your breadboards using the 5 V and `GND` pins on the Arduino to simplify power regulation in your projects.
2. Input pins on the Arduino are high-impedance inputs (they act like they have a very large resistor attached, so they don't eat a lot of current).
3. Because input pins use so little current, buttons need to be sure they have a pull-up or pull-down resistor attached to make sure the input is always detecting real voltage values and not static electricity in the air.

4. Care has to be taken when using output pins to make sure the current is never too much for the chip.
5. Using a microcontroller allows you to rewrite the logic of the project without changing the wiring.

Apply What You Have Learned

1. Which Arduino pin do you use when supplying an unregulated voltage (i.e., a voltage above the 5 V that the Arduino runs at)?
2. What Arduino pins would you use to extract power out from an Arduino connected to a power supply?
3. What is the voltage of an output pin set to HIGH?
4. What is the maximum current that should be sourced by any particular Arduino pin?
5. If you have a red LED attached to an Arduino output pin, what is the minimum size of a resistor that you need?
6. If an Arduino input pin is completely disconnected from a circuit, what state does the Arduino read it as?
7. How much current does an Arduino input use?
8. What is the best way to wire a button to an Arduino?
9. What is an advantage of storing a program in a microcontroller even if the logic could be built directly in hardware?

CHAPTER 15

Analog Input and Output on an Arduino

In Chapter 14, we learned how to do basic digital input and output with an Arduino using its I/O pins. In this chapter, we will cover how to do analog input and output as well.

15.1 Reading Analog Inputs

So far we have been focusing on digital input and output—HIGH and LOW states. The Arduino Uno also supports some amount of analog input (through its analog pins) and a sort of "faked" analog output (through its pulse-width modulation (PWM) pins, which will be covered in the next section).

On the Arduino Uno, the analog input pins are grouped together in a section labeled "Analog In." These pins are voltage sensors similar to the digital I/O pins, but they can detect a range of values between 0 V and 5 V (you should never exceed 5 V on any Arduino pin). The `analogRead()` function is similar to the `digitalRead()` function in that it takes a pin number and returns an output value. The difference is that the pin number given to `analogRead()` corresponds to an analog input pin number, not a digital pin number, and the output, rather than being LOW or HIGH, is a number from 0 to 1023 (10 bits of resolution). 0 V will give you a 0, and 5 V will give you 1023. In-between values will give you an in-between result.

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_15

Because of this, we can rework the darkness sensor we developed in Chapter 11 to use the Arduino. Since the photoresistor is just a resistor, we *still need* to make it part of a voltage divider to convert the resistance value to a voltage. However, we no longer need the voltage comparator to get it to work—we can just directly connect it to an analog port on the Arduino!

Figure 15-1 shows the darkness sensor rebuilt for the Arduino. Notice that we have a lot fewer custom components, because the control has been moved from hardware to the Arduino's software. We don't need a reference voltage, and we don't need a voltage comparator. We just have one voltage divider to convert the photoresistor's resistance to a voltage (with a wire going to the Arduino's Analog Input Pin 1) and an LED with a current-limiting resistor for the output (fed from Digital Pin 2). Everything else comes from software.

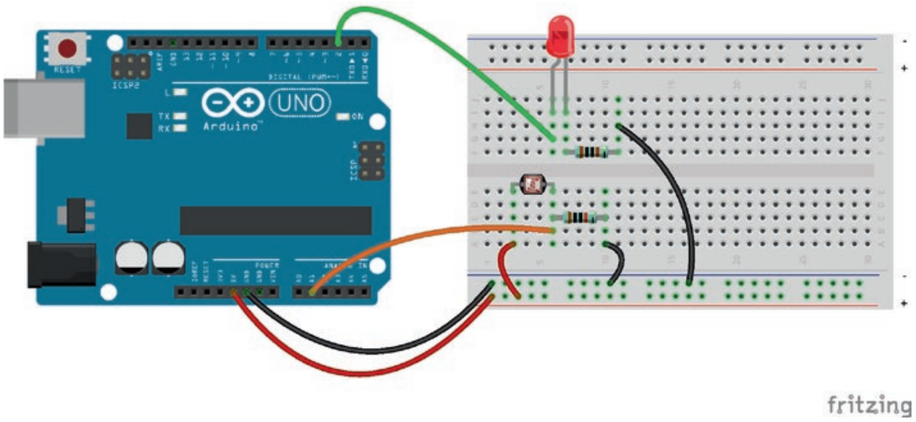

Figure 15-1. *The Darkness Sensor Rebuilt for Arduino*

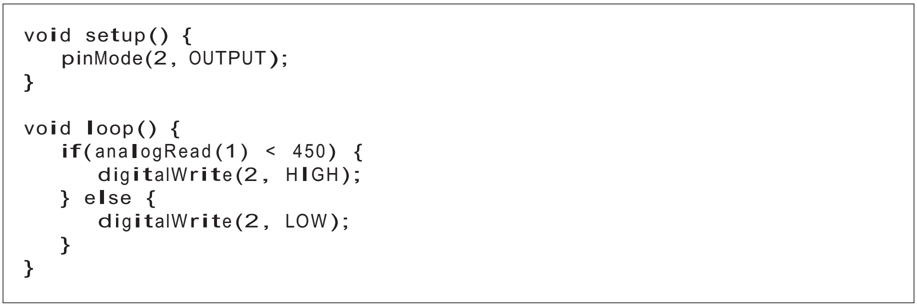

```
void setup() {
   pinMode(2, OUTPUT);
}

void loop() {
   if(analogRead(1) < 450) {
      digitalWrite(2, HIGH);
   } else {
      digitalWrite(2, LOW);
   }
}
```

Figure 15-2. *Code for the Arduino Darkness Sensor*

Figure 15-2 shows the code that will run the sensor. Note that in the `loop()` function, we are now using `analogRead()` rather than `digitalRead()`. Now, instead of it returning HIGH or LOW, it is returning a number. We can then compare that number to a baseline number to tell us whether we should turn the LED on or off.

Now, you may wonder where I got the value to compare against (i.e., 450). What I did was to test the sensor in a variety of conditions and see which value turned off the light when I wanted it to!

However, you may be wondering what exactly the values are that it is reading. Thankfully, the Arduino environment allows a way for us to get feedback from the device while it is running, if it is connected to the computer. To do this, we use what is known as the **serial** interface to the Arduino. This interface communicates over USB so that we can let our computer know how things are going in the program.

To use the USB serial interface, in your setup function, you add the following line:

```
Serial.begin(9600);
```

This tells the chip to initialize its serial interface at 9600 baud (**baud** is an old term meaning "bits per second"), which allows us to talk back to the computer. However, it is important to note that if you use the `Serial` functions, you should not have anything connected to Digital Pin 0 or Digital Pin 1 of the Arduino.

Now, in your program, you can do `Serial.println()` to output any value you want. We will do

```
Serial.println(analogRead(1));
```

to let us know what the current value of the analog input is reading at. The new program, with the added feedback, is shown in Figure 15-3. After uploading this code to the Arduino, to see your output, click the magnifying glass on the top right of the screen. You can also go to the "Tools" menu and select "Serial Monitor." Either way gets you to the same screen. When the code is running, it should be spewing out pages and pages of numbers. Each of these numbers is the current value of `analogRead()` when it is encountered in the code, which happens hundreds or thousands of times each second (it is slowed down a bit by the USB communication, and you could slow it down even further by adding `delay` commands at the end of the `loop` function). This outputting of data in order to better see what is happening within your program is known as **debugging** and is useful for tracking down problems within your code.

15.2 Analog Output with PWM

So we have discussed analog *input*, but what about analog *output*? Truthfully, the Arduino does not support analog output as such. However, analog output is *faked* on an Arduino using a technique known as **pulse-width modulation**, abbreviated as **PWM**. The Arduino only outputs 5 V on its output pins. But let's say we wanted to fake a 2.5 V signal. What might we do? Well, if we turned the pin on and off rapidly so that it was only on for half the time, that would give us about the same amount of electrical power as a continuous output of 2.5 V. That's what PWM does—it fakes lower voltages by just flipping the power to the pin on and off very rapidly so that it "looks" like a lower voltage. In Chapter 22, we will see how to actually convert this to a "true" voltage.

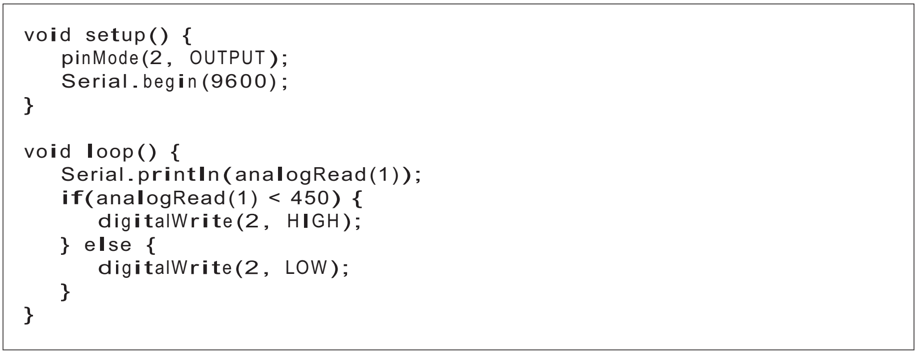

```
void setup() {
   pinMode(2, OUTPUT);
   Serial.begin(9600);
}

void loop() {
   Serial.println(analogRead(1));
   if(analogRead(1) < 450) {
      digitalWrite(2, HIGH);
   } else {
      digitalWrite(2, LOW);
   }
}
```

Figure 15-3. *Darkness Sensor with Serial Feedback*

Arduino programs use the function `analogWrite()` to use a pin for PWM. This function is a little confusingly named because (a) it uses digital pins, not analog pins, and (b) the value is between 0 and 255, not 0 and 1023 like `analogRead()`. Other than that, it basically does what you might expect. `analogWrite(3, 0);` will turn off Digital Pin 3, `analogWrite(3, 255);` will turn it all the way on, `analogWrite(3, 127);` will flick it on and off pretty evenly, and `analogWrite(3, 25);` will keep Digital Pin 3 on only a short time relative to how long the pin stays off.

To get a flavor for PWM, we will do a very simple PWM project—a dimmed LED. Figure 15-4 shows what the connection will look like—just an LED with a current-limiting resistor attached to Digital Pin 3 (which is marked with a ˜ to indicate that it is capable of PWM). Figure 15-5 shows the code to dim the output.

This code is a little more complicated. It's alright if you don't understand it fully, as this isn't a programming book. But I think if you think about it long enough, you can get it.

In short, it creates a **variable**, which is a named temporary storage location for a value (we are calling it `i` for a short name). It is declared an `int`, which means it will hold an integer, and we set it with a starting value of zero.

The `while` command means that this will repeat (or **loop**), executing everything within the block of code between `{` and `}` over and over again, as long as `i` is less than `255`. Within this block, we write the value of `i` to pin `3` using `analogWrite()`. Then, we delay for 10 milliseconds to make sure the change is visible (you can adjust this value to the speed you want the brightness to change at). Then, we increase `i` by one to go to the next value.

The next `while` loop does the same thing but goes the other way. It starts at `255` and progresses down to `0`. Then, when it is all the way to zero, the `loop()` function completes. Then, the Arduino runs the `loop` function over again.

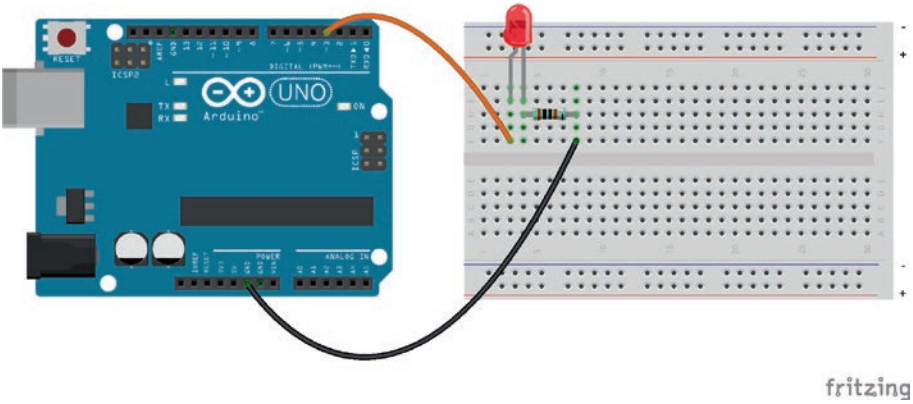

Figure 15-4. *A Simple Analog Dimmer*

Even though the Arduino is outputting “analog” by switching the pin on and off at different rates, it *looks like* the LED is dimming on and off. It is flickering so fast that we merely perceive it as a lower-energy light than as a pulsing light. In fact, when it gets to about 180 (about 70% on and 30% off), there is not a lot of difference between that and full brightness.

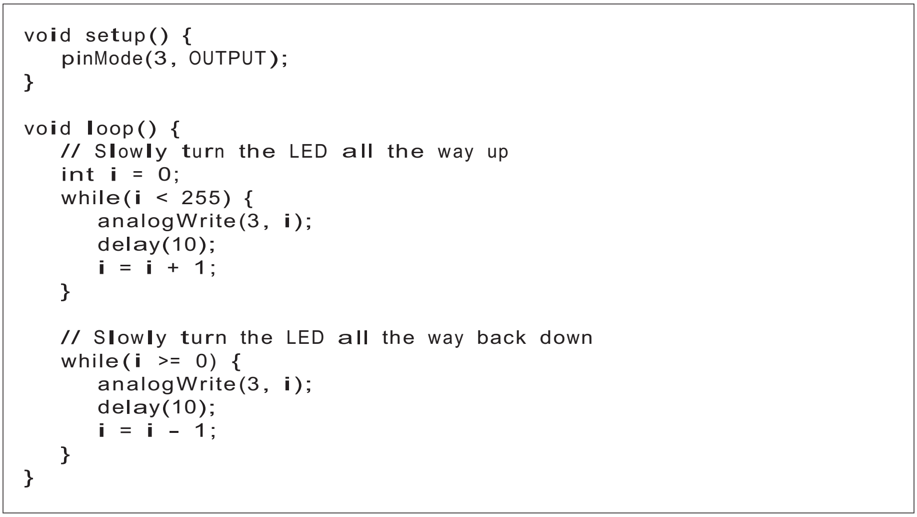

```
void setup() {
   pinMode(3, OUTPUT);
}

void loop() {
   // Slowly turn the LED all the way up
   int i = 0;
   while(i < 255) {
      analogWrite(3, i);
      delay(10);
      i = i + 1;
   }

   // Slowly turn the LED all the way back down
   while(i >= 0) {
      analogWrite(3, i);
      delay(10);
      i = i - 1;
   }
}
```

Figure 15-5. *Code for the Analog Dimmer*

Review

In this chapter, we learned the following:

1. The Arduino can read voltage values between 0 V and 5 V through the analog input pins, using the `analogRead()` function.
2. The values given from the analog input pins will be integers between 0 and 1023, where 0 represents 0 V, 1023 represents 5 V, and the numbers in between represent in-between voltages.

3. When using resistive sensors, you still need to incorporate the resistive sensors into a voltage divider circuit in order to convert the varying resistance into a varying voltage.
4. We can send debugging information back to the Arduino IDE with the USB serial interface, using the `Serial` object and its related commands.
5. Analog output on Arduino is done with the Arduino *digital* pins (not the analog pins), but only on the ones labeled with a tilde (˜), using the `analogWrite()` function.
6. Analog output on Arduino is done using pulse-width modulation (PWM), which switches the output pin on and off very rapidly to simulate a partial voltage.
7. The analog level of the analog output on the Arduino is specified by a number between 0 and 255, where 0 means no voltage, 255 means to keep the pin at 5 V the whole time, and the numbers in between cause the pin to switch back and forth rapidly to emulate a voltage between the two values.
8. In computer programming, a variable is a temporary storage location for a value.
9. In computer programming, a `while` loop repeats the inside of the while loop over and over again until its condition is no longer valid.

Apply What You Have Learned

1. What do you have to do to a resistive sensor in order to read the value of the sensor on the analog input?
2. In the code that dimmed the lights, how would we slow down the dimming process?
3. Modify the darkness sensor circuit and code so that it will output a dimmed (analog) value to the LED.
4. Further modify the darkness sensor circuit so that, at different darkness levels, the LED has different levels of brightness.
5. Think of your own modifications to the circuits in this chapter. Come up with a new way of putting together the pieces to which you have learned to create some amount of modified functioning. Implement this new design.

PART III

Capacitors and Inductors

CHAPTER 16

Capacitors

In this chapter, we will start looking at the **capacitor.**

16.1 What Is a Capacitor?

Before we begin discussing the capacitor, we need to quickly review the concepts from Chapter 3 on the relationship between charge, current, and voltage. In fact, it might be helpful to reread that chapter if you find that you have forgotten how those terms related to each other.

To review

- Charge is essentially the amount of electrical "stuff" (positive or negative) that something contains, measured in coulombs.
- Current is the *movement* of charge, measured in coulombs per second, also known as amperes.
- Voltage is the amount of force that each coulomb will produce. You can think of it as the amount of electric energy that each coulomb is capable of producing or the amount of power that each ampere of current yields.

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_16

A capacitor is a storage device which stores electric energy by holding two opposing charges (i.e., positive and negative). The amounts of charge that the capacitors that we will be working with can store is very small, but some capacitors can store very large amounts of charge. One way to think of a capacitor is as a very, very, very tiny rechargeable battery. However, unlike batteries, instead of storing a fixed voltage, a capacitor stores opposing charges. Unlike a battery, the actual voltage a capacitor yields when it discharges will depend on both the size of the capacitor and the amount of charge it is holding, and the voltage will decrease as the charge is reduced.

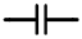

Figure 16-1. *The Symbol for a Capacitor*

The size of a capacitor is known as its **capacitance**, and it is measured in **farads** (abbreviated with the letter F), named after the influential scientist Michael Faraday. A capacitance of 1 farad means that if a capacitor stores 1 coulomb of charge, it will discharge with a force of 1 V. Most capacitors, however, have a capacitance much lower than a farad. Capacitors are usually measured in microfarads (1 millionth of a farad, abbreviated as μF or uF), nanofarads (1 billionth of a farad, abbreviated as nF), or picofarads (1 trillionth of a farad, abbreviated as pF). Capacitors are rarely rated in millifarads (1 thousandth of a farad).

So, to convert from microfarads to farads, multiply the capacitance by 0.000001. To go back from farads to microfarads, multiply the capacitance by 1000000. To convert from picofarads to farads, multiply the capacitance by 0.000000000001. To go back from farads to picofarads, multiply the capacitance by 1000000000000.

16.2 How Capacitors Work

The symbol for a capacitor in a circuit is shown in Figure 16-1. This symbol provides a visual reference for how a capacitor works. Capacitors usually work by having two conductive plates or surfaces that are separated by some sort of nonconductive material. Since the two plates are *near* each other, having a charge in one of the plates will pull charge into the other plate due to the **electric field**. An electric field is generated any time a charge accumulates. Essentially, an electric field pulls an opposing charge closer in.

Electric fields influence nearby charges even though they don't directly touch. The field will pull opposite charges closer to itself. Therefore, having a charge in one plate will attract the opposite charge in the other plate. However, since the plates are not actually touching, the electrons cannot actually jump the gap. Therefore, the capacitor will accumulate a certain amount of charge and hold it in its plates.

To understand this better, imagine that you are a positive charge. You are moving through the circuit, but why? What are you moving toward? As a positive charge, you are trying to move to the negative charge. So then, moving along, you see this big swimming pool (i.e., a capacitor). At the bottom of the swimming pool, the barrier is so thin that you can see to the other side. And what do you see there? It's the negative charges—right there at the bottom of the swimming pool! The negative charges have their own swimming pool the size of your own, separated by a barrier so thin that you can see each other.

Because you can see them, you go down into the swimming pool to see if you can interact. A lot of other bits of positive charge see this too, and they go down to see what is going on. However, when you get there, you realize that no matter how hard you try, you can't get to the negative charge that you can see. As more and more charge fills up the swimming pool, it starts to get crowded in the swimming pool. This creates *pressure* in the swimming pool—also known as *voltage*. As the swimming pool gets more and more crowded, it is harder and harder to fit new charge into it, and so the rate that it gets filled goes down, and the voltage (the pressure pushing the charge out of the pool) goes up. The same thing is happening to the negatively charged swimming pool on the other side.

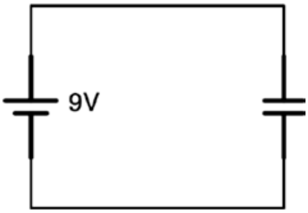

Figure 16-2. *A Simple Circuit to Charge a Capacitor*

When the pressure (voltage) to push charge out of the swimming pool equals the pressure (voltage) of the pipe (wire) leading to the swimming pool (capacitor), then the capacitor is full. When the voltage on the wire goes down (i.e., the battery disconnects), then the pressure of the charges in the capacitor pushes them back out into the wire, discharging the capacitor.

So, if one terminal of the capacitor is connected to the positive side of a battery and the other terminal is connected to the negative side of a battery, charge will quickly flow into the capacitor, as happened in our example with the swimming pool. Figure 16-2 shows this circuit setup. The exact amount of charge that flows in depends on the voltage of the battery and the capacitance of the capacitor (i.e., the size of the swimming pool). Note that this is not a short circuit, because the two plates of the capacitor are *not actually connected*. They are just close enough to attract the opposite charge onto the opposite plate.

In a setup like Figure 16-2, there is an equation that tells you the amount of charge that can be stored in a capacitor for a given voltage:

$$Q = C \cdot V \tag{16.1}$$

In this equation, V is the voltage applied, C is the capacitance (in farads), and Q is the amount of charge (in coulombs) stored in the capacitor's plates.

> **Example 16.7** If I attached a 66 μF capacitor to a 9 V source, how much charge gets stored on the capacitor?
>
> Well, first, we need to convert the capacitance from microfarads to farads. 66 μF is 66 millionths of a farad, so it would be 0.000066 F. Now, we just need to plug the numbers into the equation:
>
> $$\begin{aligned} Q &= C \cdot V \\ &= 0.000066.9 \\ &= 0.000594 \text{ coulomb} \end{aligned}$$
>
> Therefore, if I attached a 66 μF capacitor to a 9 V source, the capacitor would store 0.000594 coulomb of charge.

After my capacitor is charged up, I can use it as a very tiny battery. That is, I could disconnect it from the circuit in Figure 16-2 and reconnect it as the battery for another circuit. However, as I mentioned, this will be a very, very tiny battery. Additionally, as the charge leaves the capacitor, the voltage of the battery will decrease as well.

If we use the capacitor as a battery, if we know how much charge is stored in the capacitor, we can rearrange Equation 16.1 slightly to figure out how much voltage it will deliver when it begins to discharge:

$$V = \frac{Q}{C} \tag{16.2}$$

Now, as the capacitor discharges, since the charge it is holding will decrease, so will the voltage it delivers. If we know how much charge is left at any given moment, we can calculate the present voltage.

> **Example 16.8** If I have a charge of 0.0023 coulomb stored in a 33 μF capacitor, if I discharge the capacitor, what voltage will it begin to discharge at?
>
> The first thing to do here is to convert the capacitance to farads. A microfarad is 1 millionth of a farad, so 33 μF = 0.000033 F. Now we can just plug numbers into the equation:
>
> $$\begin{aligned} V &= \frac{Q}{C} \\ &= \frac{0.0023}{0.000033} \\ &\approx 69.7\text{V} \end{aligned}$$
>
> Therefore, when the capacitor discharges, it will discharge at 69.7 V.

Within a circuit, capacitors are typically used in places where voltages vary. Until the voltage in the capacitor equals the voltage in the circuit, the capacitor will pull charge out of the circuit. When the voltage in the capacitor exceeds the voltage in the circuit, the capacitor will put charge back into the circuit. This dynamic will serve as the core principle used when designing circuits with capacitors.

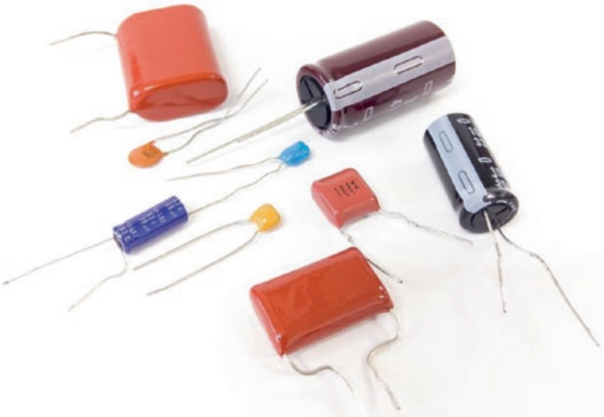

Figure 16-3. *Several Different Types of Capacitors (Shutterstock/ Muhammad Anuar bin Jamal)*

16.3 Types of Capacitors

There are numerous types of capacitors available, each varying in the types of materials they are made of, the internal geometry of the capacitor, and the packaging. Figure 16-3 shows many different types of capacitors that are available.

However, the most important feature of capacitors besides their capacitance is whether they are **polarized** or **non-polarized**. In a *non-polarized capacitor*, it doesn't matter which way you attach the leads. Either side can be the more positive or more negative side. The most common type of non-polarized capacitor is the circle-shaped ceramic disk capacitor. While ceramic disk capacitors are easy to use, they suffer from having limited capacitance.

In a *polarized capacitor*, however, one lead *must* stay more positive than the other lead, or you risk damaging the capacitor. The most common type of polarized capacitor is the electrolytic capacitor. Electrolytic capacitors look like little barrels with leads coming out of them. They usually have much higher capacitances than ceramic disk capacitors, but you have to be sure that the polarity is correct and never switches direction.

On polarized capacitors, it is important to know which lead is positive and which is negative. There are several different ways that a manufacturer might indicate this:

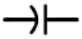

Figure 16-4. *A Polarized Capacitor Symbol*

1. One or both of the leads can be marked with their respective polarities (+ or -).
2. The negative lead can be marked with a large stripe.
3. The positive lead can be longer than the negative.

Many manufacturers do all three.

In a circuit schematic, if a polarized capacitor is called for, it will use a special capacitor symbol as shown in Figure 16-4. The only difference is that one side is curved. In a polarized capacitor, the straight side is the positive side, and the curved side is the negative side. Sometimes polarized capacitors are marked with plus (+) and minus symbols (-) instead of (or in addition to) having a straight side and a curved side.

On any capacitor, it is important to know the capacitance of the capacitor you are looking at. On larger capacitors (especially on electrolytics), manufacturers can print the full capacitance including the units directly on the capacitors. However, many capacitors are extremely small and can't fit that much information on them.

For smaller-sized capacitors, the capacitance is described by three digits and an optional letter. The third digit is how many zeros to add to the end of the other two digits, and then the whole number is the capacitance in picofarads. So, if the number is 234, then the capacitance is 230,000 pF (23 followed by four zeros). If the number is 230, then the capacitance is 23 pF.

The letter at the end tells the *tolerance* of the capacitor—how much the capacitance is likely to vary from its markings. Common letters are J (±5%), K (±10%), and M (±20%).

16.4 Charging and Discharging a Capacitor

To charge up a capacitor almost instantly to a specific voltage, you can (as we also did earlier in the chapter) directly connect the positive and negative leads of the battery to the leads of the capacitor. Once it it charged, you can use the capacitor as a very tiny battery for a project.

In Figure 16-5, we see two simple circuits and an electrolytic capacitor. For this circuit to really work, it helps to have an electrolytic capacitor at least 100 µF and a resitor at least 10 kΩ. So, first, build the LED circuit on the right side of Figure 16-5 on a breadboard. However, don't connect any power to the power rails. Next, take the capacitor, and touch the positive lead of the capacitor to the positive terminal of the 9 V battery and the negative lead of the capacitor to the negative terminal. *Do not let the leads of the capacitor touch each other*. Hold it there for a second or two to allow the capacitor to fully charge. Now, *without touching the leads of the capacitor*, place the capacitor so that the positive lead goes into the positive rail of your breadboard and the negative lead goes into the negative rail of your breadboard. When you do this, the capacitor will power your LED project for a few seconds.

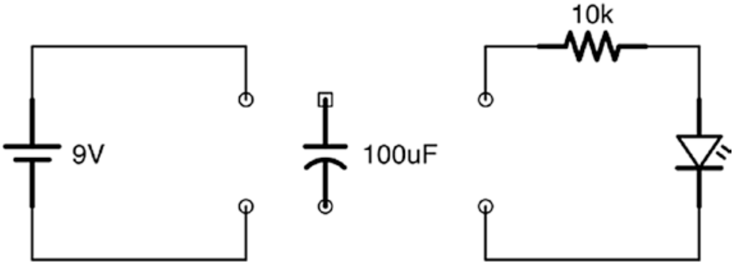

***Figure 16-5.** Using a Capacitor as a Battery*

Now, you will notice that the LED gets dimmer before it goes out. Why does this happen?

Remember that the voltage that the capacitor yields is based on the charge that is present in it. So, going back to Equation 16.2, we can see that the voltage is based on the charge that is in the capacitor. When the capacitor is connected, it will start with a 9 V discharge, since that is what the battery was able to put into the capacitor. However, the capacitor is using up its charge to power the project. This means that as soon as its charge starts leaving, the voltage starts going down, since the voltage is related to how much charge is inside the capacitor.

> **Example 16.9** If we use the components listed in Figure 16-5, how much charge does the battery initially store in the capacitor?
>
> We can use Equation 16.1 to determine this:
>
> $$\begin{aligned} Q &= C \cdot V \\ &= 100\mu F.9\text{V} \\ &= 0.000100\text{F}.9\text{V} \\ &= 0.0009\text{coulombs} \end{aligned}$$

> **Example 16.10** After 0.0003 coulomb of charge has been discharged, what voltage is the capacitor discharging at?
>
> To find this out, we first have to find out how much charge is *remaining* in the capacitor. So, to find this out, we just subtract the amount of charge that has been discharged from our starting charge. This gives 0.0009 – 0.0003 = 0.0006 coulomb.
>
> Now, we can use Equation 16.2 to find out what voltage the capacitor is discharging at:

$$\begin{aligned} V &= \frac{Q}{C} \\ &= \frac{0.0006}{0.000100} \\ &= 6\text{V} \end{aligned}$$

Therefore, the capacitor is discharging at 6 V.

16.5 Series and Parallel Capacitances

Just like resistors, capacitors can be used either in series or in parallel. In fact, they use the same equations for series and parallel capacitance as do resistors. However, there is *one big difference*. The parallel and series versions of the equations are *reversed* for capacitors.

If we want to double resistance in a circuit, we simply add another resistor of the same size in series. If I want to double my capacitance, I can add in another capacitor of the same size. However, for the capacitor, we would add the capacitor *in parallel*. The following is the formula for adding capacitance in parallel:

$$C_T = C_1 + C_2 + \dots \qquad (16.3)$$

If we want to put capacitors in series, we would use a formula that is exactly like the formula for resistors in parallel:

$$C_T = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2} + \dots} \qquad (16.4)$$

Example 16.11 If I have a 100 µF capacitor in series with a 200 µF capacitor, how much total capacitance do I have?

To do this, we use Equation 16.4. It is best to convert our capacitances to farads first. It does not matter as long as the units are the same, but it is good practice to convert to farads so you don't forget when you start doing capacitances in different units. 100 µF is the same as 0.0001 F, and 200 µF is the same as 0.0002 µF.

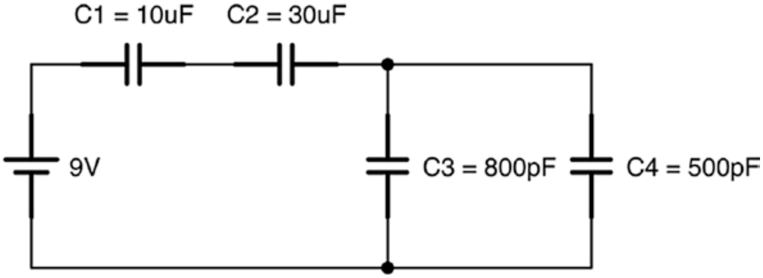

Figure 16-6. *Capacitors in Series and Parallel*

Now, we will plug these values into the equation:

$$
\begin{aligned}
C_T &= \frac{1}{\frac{1}{C_1}+\frac{1}{C_2}} \\
&= \frac{1}{\frac{1}{0.0001\mathrm{F}}+\frac{1}{0.0002\mathrm{F}}} \\
&= \frac{1}{10000+5000} \\
&= \frac{1}{15000} \\
&= 0.0000666\mathrm{F} \\
&= 66.6\mu F
\end{aligned}
$$

Example 16.12 Just as we did for resistors, we can combine a variety of parallel and series capacitances for a single capacitance value. For instance, take the circuit in Figure 16-6. What is the total capacitance of this circuit?

First, we can start by converting all of the capacitances to farads. This will make combining everything easier down the road. In that case, $C_1 = 0.00001$ F, $C_2 = 0.00003$ F, $C_3 = 0.0000000008$ F, and $C_4 = 0.0000000005$ F. C3 and C4 are in parallel, so we can combine them using the parallel formula (Equation 16.3):

$$\begin{aligned} C_T &= C_3 + C_4 \\ &= 0.000000000\text{F} + 0.0000000005\text{F} \\ &= 0.0000000013\text{F} \end{aligned}$$

Now, if we substitute that capacitance in for C3 and C4, we can use the series formula (Equation 16.4) to find the total capacitance of the circuit:

$$\begin{aligned} C_T &= \frac{1}{\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_{3\&4}}} \\ &= \frac{1}{\frac{1}{0.00001} + \frac{1}{0.00003} + \frac{1}{0.0000000013}} \\ &= \frac{1}{100000 + 33333.33 + 769230769.23} \\ &= \frac{1}{769364102.56} \\ &= 0.0000000013\text{F} \\ &= 1.3\text{nF} \end{aligned}$$

So the total capacitance of the circuit is 1.3 nF.

16.6 Capacitors and AC and DC

One important characteristic of a capacitor is that it *allows* the flow of AC (alternating current) but it blocks the flow of DC (direct current). To understand why this is, let's think about how capacitors operate.

Capacitors, when they are charging, essentially act as short circuits. As positive charge flows into the capacitor at one terminal, negative charge flows into the other side. An *inflow* of negative charge to that side means that there is a net *outflow* of positive charge on that side. So, *even though the physical electrons never cross the boundary between the plates*, the total charge *actually moves* from one side to the other.

However, this situation is *temporary* because, as the capacitor gets more full of charge, new charge is less likely to enter. Once the capacitor is fully charged (based on the voltage), no new charge can flow into one side to cause the charge to go out to the other.

Once the capacitor is charged, *current stops flowing through it*. If the voltage level on one of the leads *changes*, then charge will flow until the change is compensated for by the capacitor charging or discharging.

Thus, because there is a barrier between the two sides of a capacitor, it is only when the voltage *changes* that the current flows *through* a capacitor. If the voltage stays the same, then charge will stop flowing through the capacitor as soon as it reaches capacity for that voltage.

Because of this, we say that capacitors *allow* AC but *block* DC. This rule of thumb will help us use capacitors in a variety of circuits later on.

16.7 Using Capacitors in a Circuit

In this section, we will discuss the uses of capacitors in a circuit. We haven't discussed enough to actually use capacitors in these ways, but I thought it would be helpful to see why capacitors are used so you can see why you should care about these things.

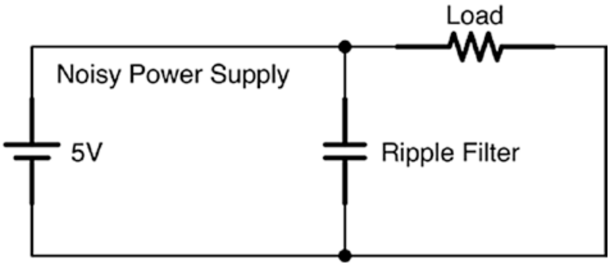

Figure 16-7. *A Capacitor Filtering a Noisy Power Supply*

The first thing that capacitors are used for is for filtering noisy signals, especially in power sources. Imagine that you had a power source which, instead of delivering a constant voltage, the voltage would wobble a bit. If you placed a capacitor in parallel with this circuit, the power source would charge the capacitor up. Then, if the power source dropped a bit, the capacitor would start discharging to compensate. Likewise, if the power source increased, the capacitor would absorb some of that increase by storing the charge.

Thus, by acting as a temporary location to store extra charge, the capacitor can smooth out ripples in a signal, as seen in Figure 16-7. This sort of usage is known as a *filtering* capacitor because it filters out noise.

This can be seen as another implementation of our rule that capacitors allow AC to flow but block DC. When the source voltage has a ripple, the *ripple itself* gets shunted to ground by the capacitor. However, under normal operation, the DC part of the current can't flow through the capacitor and just continues on to the load.

On many ICs that have fluctuating power requirements, you will find many of their spec sheets recommending one or more capacitors on their power supply or other pins in order to filter out the noise.

Another way of using capacitors is as a **coupling capacitor**. A coupling capacitor is used when you have *both* an AC signal and a DC signal combined. This will happen a lot when we talk about amplification. What happens is that we will have an amplified signal, but we *only* want the AC portion of the signal. We can do that by adding in a capacitor to join the segment of the circuit that has both AC and DC components to the segment of the circuit that only wants the AC component. Only the variations in the voltage, not the base voltage, will be transmitted through the capacitor.

Another way that capacitors are used is for filtering specific frequencies. Higher-frequency signals are easily transmitted through a capacitor, but low-frequency signals are essentially blocked as if they were DC. The frequencies that are allowed are based on the capacitance. If you put a capacitor in series with the signal path, then it will only allow higher frequencies. If you put a capacitor in parallel with the signal path, then it will ony allow lower frequencies (the higher frequencies will be shunted to ground).

CAPACITOR SAFETY

If you see a capacitor in any sort of equipment, do not touch it, even with the power off. Because capacitors store charge, they can still hold onto the charge even after the power is off. If you need to touch a capacitor for some reason, use a resistor to discharge it. Placing a large (megaohm) resistor between the leads of a capacitor (without touching the leads yourself!) for a few seconds should drain most capacitors. You can also use a multimeter to determine that the capacitor is drained (0 volt).

Review

In this chapter, we learned the following:

1. An electric field is generated wherever there is an accumulated charge.
2. A capacitor is a device that stores electric energy by holding two opposing charges.
3. Capacitors are sized by their capacitance which is measured in farads.
4. Because a farad is so large, capacitors are usually measured in microfarads, nanofarads, or picofarads.
5. Polarized capacitors (usually electrolytic capacitors) have distinct positive and negative terminals, while non-polarized capacitors (usually ceramic disk capacitors) can go either way.
6. When a battery is connected to a capacitor, it will store a charge on the capacitor.
7. The equation for the charge (Q) stored on the capacitor when a battery is connected is $Q = C \cdot V$ where C is the capacitance in farads, V is the voltage, and Q is the charge in coulombs.
8. By rearranging the equation, we can determine, for a given amount of charge, how much voltage a capacitor will discharge when it is allowed to: $V = \frac{Q}{C}$.
9. As the capacitor discharges its charge through the wire to create current, the amount of charge remaining will decrease, which will also lower the voltage it is putting out.

10. Capacitors with small packages are often marked with a three-digit code, where the third digit is the number of zeros to add to the other two digits and the final number is in picofarads.
11. When multiple capacitors are used together, their capacitances can be combined similar to resistors, but with the series and parallel equations switched.
12. The series capacitance equation is as follows:

    $$C_T = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2} + \ldots}$$

13. The parallel capacitance equation is as follows: $C_T = C_1 + C_2 + \ldots$
14. Capacitors allow the flow of AC through them but block the flow of DC.
15. Capacitors can be used to split AC (high-frequency) and DC (low-frequency) portions of a signal. The AC portion will travel through the capacitor, while the DC portion will remain on one terminal side.
16. Capacitors can be used to filter audio frequencies, filter power supply ripples, operate as voltage boosters, and couple together a combined AC/DC circuit with an AC-only circuit.

Exercises

1. Convert 23 F to microfarads.
2. Convert 15 μF to farads.
3. Convert 35 pF to farads.
4. Convert 0.0002 F to microfarads.
5. Convert 0.0030 μF to farads.
6. If a voltage of 6 V is applied to a 55 μF capacitor, how much charge would it store?
7. If a voltage of 2 V is applied to a 13 pF capacitor, how much charge would it store?
8. If a 132 μF capacitor is holding 0.02 coulomb of charge, how many volts will it produce when it begins to discharge?
9. If a 600 pF capacitor is holding 0.03 coulomb of charge, how many volts will it produce when it begins to discharge?
10. A 121 μF capacitor is connected to a battery. After some fluctuation, the capacitor has 0.00089 coulomb of charge stored in it, and the battery is reading 8.9 V. Is the capacitor going to be charging or discharging at this point?

CHAPTER 17

Capacitors as Timers

In this chapter, we are going to learn how to measure the time it takes for a capacitor to charge. Once we learn this, we can use capacitors for timers—both for delaying a signal and for creating an oscillating circuit.

17.1 Time Constants

As we learned in Chapter 16, when a voltage is applied to a capacitor, it will store energy by storing a charge on its plates, the amount of charge being based on the voltage supplied and the capacitance of the capacitor (see Equation 16.1).

The cases we examined in Chapter 16 generally had the battery connected directly to the capacitor. In such cases, the capacitor charges to the source voltage almost immediately. However, if a capacitor charges through a resistor (instead of being directly connected to the battery), then it takes much longer to fill the capacitor to capacity than if it were connected to the battery directly. In fact, it never *fully* reaches capacity, though it gets close enough that we say that it does.

Having a resistor in series with a capacitor is a configuration known as an **RC (resistor-capacitor) circuit**. The amount of time it takes a capacitor to charge is based on both the resistance of the resistor and the capacitance of the capacitor. The actual equation for this is kind of complicated, but there is a simple trick that suffices for nearly every situation, known as the RC time constant.

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_17

The RC time constant is merely the product of the resistance (in ohms) multiplied by the capacitance (in farads) which will yield the RC time constant in seconds. The RC time constant can be used to determine how long it will take to charge a capacitor to a given level (we will talk about what this level is in a moment). So, for instance, if I have a 100 μF capacitor and a 500 Ω resistor, the RC time constant is 0.0001 ∗ 500 = 0.05 second.

This constant can then be used with the table in Figure 17-1 to determine how long it will take to charge a capacitor to a given level.[1]

# of Time Constants	**% of Voltage**	**% of Current**
0.5	39.3%	60.7%
0.7	50.3%	49.7%
1	63.2%	36.8%
2	86.5%	13.5%
3	95.0%	5.0%
4	98.2%	1.8%
5	99.3%	0.7%

Figure 17-1. *RC Time Constants*

For instance, if I wait for 2 time constants (in this case, 0.05 * 2 = 0.1 second), my capacitor will be charged to 86.5% of the supply voltage. The current flowing through it will be at 13.5% of what current would be flowing if there was just a straight wire instead of a capacitor.

---

[1] For those curious how these numbers are calculated, see Appendix D, Section D.7.3, "Time Constants."

While a capacitor is never really “fully charged” (because it never fully reaches 100%), in this book, we will use 5 time constants to consider a capacitor fully charged.

> **Example 17.13** I have a power supply that is 7 V and a capacitor that is a 100 μF capacitor. I want it to take 9 seconds to charge my capacitor. What size of a resistor do I need to use to do this?
>
> To solve this problem, we need to work backward. Remember, we are considering 5 time constants to be fully charged. Therefore, the time constant we are hoping to achieve is 9/5 = 1.8 seconds. The capacitance is 100 μF, which is 0.0001 F. Since the time constant is merely the product of the capacitance and resistance, we can solve for this as follows:

$$\begin{aligned} RC\ \text{TimeConstant} &= \text{capacitance} \cdot \text{resistance} \\ 1.8 &= 0.0001 * R \\ 0.0001 * R &= 1.8 \\ R &= \frac{1.8}{0.0001} \\ R &= 18000 \end{aligned}$$

Therefore, to make it take 9 seconds to charge the capacitor, we need to use an 18 kΩ resistor.

In the same way, if we connect the capacitor to ground through the resistor instead of to the voltage supply, the capacitor will discharge in the same way that it charged. It will begin discharging a lot, but then, as it gets closer to zero, it will level off the amount that it is discharging. You would then read Figure 17-1 to know what percentage of the voltage/current the capacitor has *discharged*.

Example 17.14 Suppose a 100 μF capacitor has been charged up to 7 V and then is disconnected from the power supply, but the ground connection remains. We then connect the positive terminal of the capacitor to a 5 kΩ resistor that connects to ground. After 2 seconds, how much voltage is remaining in the capacitor?

To find this out, we first need to find the RC time constant of the circuit.

$$\begin{aligned} T &= 5k\Omega * 100\mu F \\ &= 5{,}000\,\Omega * 0.0001\,\mathrm{F} \\ &= 0.5\ \text{seconds} \end{aligned}$$

So the RC time constant is 0.5 second.

So, after 2 seconds, we have performed 4 RC time constants. Looking at Figure 17-1, we have discharged 98.2% of the capacitor's voltage. This means that the amount of voltage *lost* should be 0.982 * 7 V = 6.874 V. If we have lost that much voltage, that means we should subtract it, so the remaining voltage is 7 V – 6.874 V = 0.126 V.

So the voltage after 2 seconds will be 0.125 V.

17.2 Constructing a Simple Timer Circuit

Let's say that we want a circuit that, when turned on, *waits* for a certain amount of time and then does something. How might we do it?

Think of it this way. We can use capacitor charging to give us a time delay. However, we need something that "notices" when the time delay is finished. In other words, we need a way to trigger something when a certain threshold is crossed.

What happens to the capacitor as it charges? The voltage across its terminals changes. When it is first connected to the circuit, there is zero voltage across its terminals. As it charges, the voltage across the capacitor's terminals keeps going up until it matches the supply voltage. Therefore, we know when we have hit our target time based on when the voltage is at a certain level. But how will we know when we are at the right voltage? Have we done a circuit so far that detects voltages? What component did we use?

If you remember back in Chapter 11, we used the LM393 to compare voltages. We supplied the LM393 with a *reference* voltage, and then it triggered when our other voltage went above that voltage. We can do the same thing here.

What we will construct is a circuit that waits for 5 seconds and then turns on an LED. In order to do this, we will need to choose (a) a reference voltage to use, (b) a resistor/capacitor combination that will surpass the reference voltage after a certain amount of time, and (c) an output circuit that lights up the LED.

There are virtually infinite combinations we could choose from for our reference voltage, resistance, and capacitance. In fact, the supply voltage doesn't matter so much since what we are looking at are *percentages* of the supply voltage, which will be the same no matter what the actual voltage is.

For this example, we will use a basic 5 V supply and make the reference voltage to be half of the supply voltage. This allows us to use any two equivalent resistors for a voltage divider to get our reference voltage.

Now, since the reference voltage is half of our supply voltage, we will use the table in Figure 17-1 to determine how many time constants that needs to be. The table says that for 50% of voltage, it will take 0.7 time constant (it is actually 50.3%, but we are not being that exact).

Therefore, the equation for our resistor and capacitor is

$$\begin{aligned} T &= 5 \text{ seconds} \\ 5 \text{ seconds} &= 0.7 \cdot R \cdot C \\ \frac{5}{0.7} &= R \cdot C \\ 7.14 &= R \cdot C \end{aligned}$$

So we can use any resistor and capacitor such that the ohms multiplied by the farads equals 7.14. I usually use 100 μF capacitors for larger time periods such as this because they are larger and because being exactly 0.0001 F, it makes it easier to calculate with. Therefore, we can very simply calculate the needed resistor for this capacitor:

$$\begin{aligned} 7.14 &= R \cdot C \\ 7.14 &= R \cdot 0.0001 \\ \frac{7.14}{0.0001} &= R \\ 71400 &= R \end{aligned}$$

Therefore, we need a resistor that is about 71, 400 Ω. We could choose a combination of resistors that hits this exactly, but for our purposes, we only need to get close. In my case, the closest resistor I have is 68,000 Ω (i.e., 68 kΩ). That is close enough (though I could get even closer by adding in a 3 kΩ resistor in series).

Figure 17-2 shows the full circuit. When building this circuit, don't forget that the comparator also has to be connected to the supply voltage and ground as well! As you can see, R2 and R3 form the voltage divider that provides the reference voltage for the comparator at half of the supply voltage. The values for R2 and R3 are arbitrarily chosen, but they must be equal to get the reference voltage. I chose medium-high values for these resistors so as to not waste current with the voltage divider. The circuit made by R1 and C1 is the timing circuit. When the circuit is first plugged in, C1 is at voltage level 0, essentially acting as a short circuit while it first begins to store charge. At 0 volt, this is obviously less than our reference voltage. However, as the capacitor charges, less and less current can flow into C1. Its voltage level increases under the rules of the RC time constant. After 0.7 time constant, the voltage will be above the voltage level set by the reference voltage, and our voltage comparator will switch on.

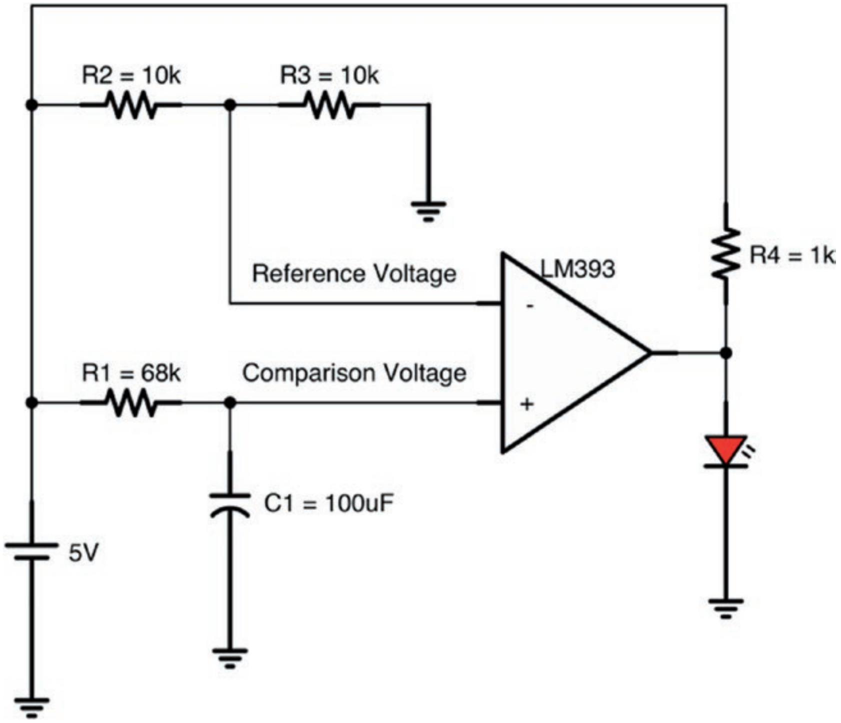

Figure 17-2. *A Simple Timer for an LED*

Remember, though, from Chapter 11, that the LM393 operates using a pull-up resistor. That is, the comparator doesn't ever source current. It will sink current (voltage level 0 when the + input is less than the – input) or disconnect (when the + input is greater than the – input). Therefore, R4 is providing a pull-up resistor to supply power to the LED when the LM393 disconnects.

Figure 17-3 shows this circuit laid out on the breadboard. We are using the second voltage comparator of the LM393 just because it was a little easier to show the wiring for it. If you need to see the pinout for the LM393 again, it was back in Figure 11-4.

Notice the prominence of our basic circuits from Chapter 9 on the breadboard. On the top, we have a combination pull-up resistor which is also acting as a current-limiting resistor (as they often do). On the bottom left, we have a voltage divider.

On the bottom right, we have our new timing circuit, which looks a lot like a voltage divider. In fact, it acts like one too, where the voltage varies by time! If you think about what the capacitor is doing, when we first apply voltage, it acts like a short circuit—in other words, no resistance. This means that there is 0 volt across the capacitor and the resistor is eating up all of the voltage. But, as the capacitor fills up, it increases voltage, which it is *dividing* with the resistor! Just like before where we had a sensor which created a sensor-based variable voltage divider, here we have what is essentially a time-based variable voltage divider.

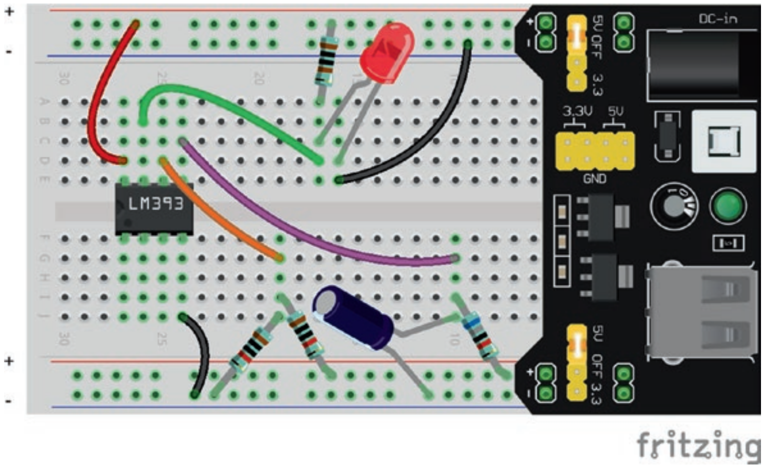

Figure 17-3. *The Capacitor Timer Circuit on a Breadboard*

17.3 Resetting Our Timer

The timer is great, except for one thing—how do we turn it off? You might have noticed that, even if you disconnect power to the circuit, when you turn it back on, it doesn't do any timing anymore! Remember, the capacitor is *storing* charge. When you turn off the circuit, it is *still* storing the charge.

Now, you can very simply get rid of the charge by putting a wire between the legs of the capacitor. However, for larger capacitors, you would need to do this with a resistor instead of a wire in order to keep there from being a dangerous spark (the resistor limits the current, which makes the discharge slower). But how do we do this with a circuit?

What we can do is add a button or switch that will do the same thing as putting a wire or resistor across the legs of the capacitor. Figure 17-4 shows how to do this.

Now let's look at how this circuit works. First of all, we added two components, the switch and the R5 resistor connected to it. To understand what this does, pretend for a minute that the resistor isn't there. What happens when we push the button? That would create a direct link from the positive side of the capacitor to ground. Since the negative side of the capacitor is also connected to ground, that means that these two points would be directly connected. Thus, they would be at the *same* voltage. This would be achieved by the charge suddenly rushing from one side to another to balance out.

If the capacitor were larger or the voltages greater, this might be somewhat dangerous. You would have a large current and a large voltage for a short period of time (which would yield a large wattage), which could blow something out. Therefore, it is good practice to use a small resistor between the button and ground. The resistor should be *much* smaller than the resistor used to charge the capacitor. The specific size doesn't matter too much—it needs to be small enough to discharge it quickly and large enough to prevent a spark when you push the button. You can use the same time constants for discharging the capacitor that you used for charging it. However, keep in mind that the charging circuit is still running! That's why the resistor has to be small—it has to discharge the capacitor *while* the other resistor is trying to charge it.

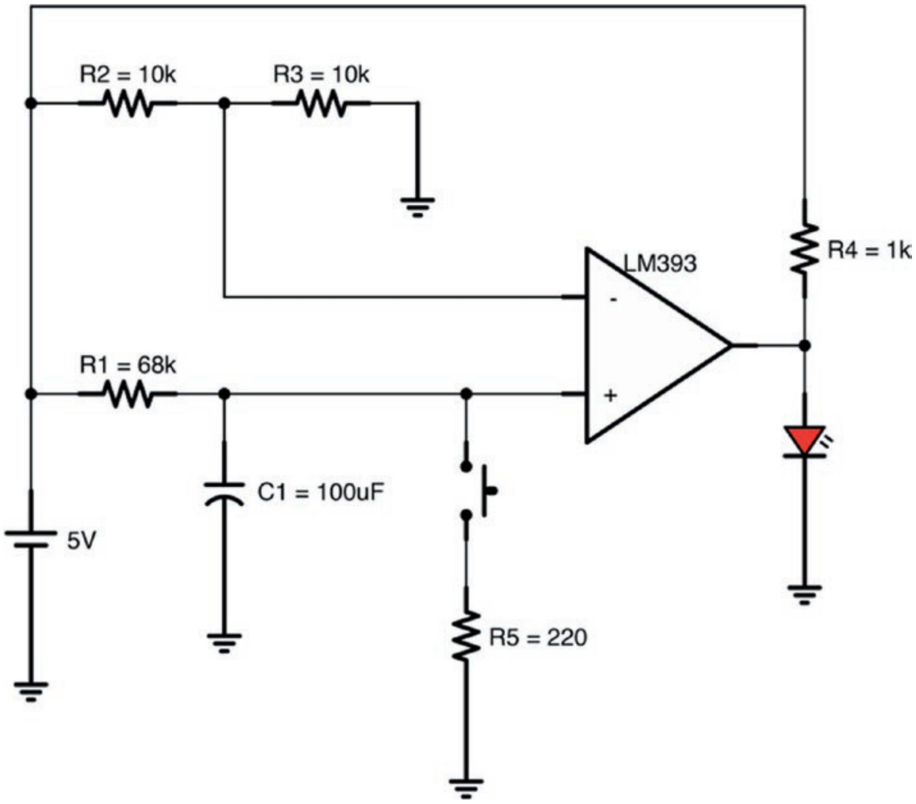

Figure 17-4. *Adding a Reset to the Capacitor*

Also notice R1 and R5. When the button is being pushed, what is happening to them? Think back to our basic resistor circuits. If we have two resistors, with a wire coming out of the middle, what is that? It's a voltage divider! So not only is R5 being a current-limiting resistor for the button, it is also acting as a voltage divider in concert with R1.

What this means is that the capacitor will only discharge down to the level of voltage division between these resistors. That's fine, as long as it is low enough. You will find that in electronics, "close enough" often counts. The trick is knowing how close is close enough and how close is still too far.

However, usually we can perform some basic calculations to figure this out. Since R1 is 68 kΩ and R5 is 200 Ω, what is the voltage at the point of division? If we use a 9 V supply, using the formula from Equation 9.1, then we can see that

$$\begin{aligned} V_{OUT} &= V_{IN} \cdot \frac{R_5}{R_1 + R_5} \\ &= 9 \cdot \frac{220}{68000 + 220} \\ &= 9 \cdot \frac{220}{68220} \\ &\approx 9 \cdot 0.00322 \\ &\approx 0.029\text{V} \end{aligned}$$

So, as you can see, when the button is pushed, the final voltage of the capacitor, while not *absolutely* 0 volt, is pretty darn close.

Also note that, if we wanted, we can reverse the action of this circuit. By swapping our two inputs to the voltage comparator, we can make the circuit be on for 5 seconds and then switch off.

Review

In this chapter, we learned the following:

1. Having a resistor and capacitor in series with each other is known as an RC circuit.
2. In an RC circuit, the amount of time it takes for a capacitor to charge to the supply voltage is based on the capacitance of the capacitor and the resistance of the resistor.

3. The RC *time constant* is a convenient way to think about how long it takes for a capacitor to charge in RC circuits.

4. The RC time constant is calculated by multiplying the resistance (in ohms) by the capacitance (in farads) with the result being the number of seconds in 1 RC time constant.

5. The table in Figure 17-1 shows how long it takes to charge a capacitor to different percentages of supply voltage as a multiple of RC time constants.

6. The RC time constant chart can also be used to calculate the amount of time it takes for a capacitor to discharge to ground if it is disconnected from its source and connected through a resistor to ground. In this case, Figure 17-1 is used to tell what percentage of the voltage has been *discharged*.

7. A timer can be constructed by using a comparator and an RC circuit along with a reference voltage provided by a voltage divider. By tweaking the RC circuit, the timing can be changed.

8. After charging a capacitor, a means needs to be provided to discharge it as well, such as a button leading to ground.

9. Such discharge methods need to have resistance to prevent sparks and other failures, but not too much resistance as they can accidentally form voltage dividers with other resistors in the circuit and prevent full discharge.

10. Even as our circuits get more advanced, the basic circuits we found in Chapter 9 are still dominating our circuit designs.

Apply What You Have Learned

1. If I have an RC circuit with a resistor of 10 Ω and a capacitor of 2 F, what is the RC time constant of this circuit?
2. In the previous question, how many seconds does it take to charge my capacitor to approximately 50% of supply voltage?
3. If I have an RC circuit with a resistor of 30,000 Ω and a capacitor of 0.001 F, what is the RC time constant of this circuit?
4. In the previous question, what percentage of the capacitor's voltage is charged after 60 seconds?
5. If I have an RC circuit with a resistor of 25 kΩ and a capacitor of 20 μF, what is the RC time constant of this circuit?
6. Give a resistor and capacitor combination that will yield an RC time constant of 0.25 second.
7. Reconfigure the circuit in Figure 17-2 to wait for 3 seconds. Draw the whole circuit.
8. Redraw the previous circuit and circle each basic resistor circuit pattern and label it.

CHAPTER 18

Introduction to Oscillator Circuits

In Chapter 17, we learned how to use RC (resistor-capacitor) circuits to create timers. In this chapter, we are going to use our concept of timing circuits to move from one-time timer circuits to *oscillating* circuits.

18.1 Oscillation Basics

So far, most of the circuits we have made have been fairly directional. You do an action (i.e., press a button) and then something happens, but then the circuit just maintains a steady state after that. In Chapter 17, we at least added a delay—allowing the circuit to do something *later*.

However, if you want actions to continue on into the future, you need to not only have delays, you need to have *oscillations*. **Oscillation** means to go back and forth. An oscillating circuit is one that goes back and forth continually between two states—usually zero voltage and some positive voltage. Imagine blinking lights at Christmas. These lights go from a state of zero voltage (off) to a state of positive voltage (on). And they go back and forth between these states over and over again as long as there is power in the circuit. These are oscillating circuits.

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_18

An oscillator is usually described by either its **period** or its **frequency**. The period of an oscillator is how many seconds it takes to go through one complete cycle. So, if I had lights that blinked on for 1 second and then off for 2 seconds (and continued repeating in that fashion), the period would be 3 seconds.

The frequency of an oscillation is the number of times that the system cycles every second, which is merely the reciprocal of the period (i.e., one divided by the period). So, in the example given, since the period was 3 seconds, the frequency is $\frac{1}{3}$ cycle per second. The "cycles per second" unit of frequency also has a special name—**hertz** (often abbreviated as Hz). Therefore, we would say that our blinking lights blinked at a frequency of $\frac{1}{3}$ Hz.

> **Example 18.15** Let's say that we have an oscillator that turns a light on for 4 seconds and then turns off for 4 seconds and repeats this continually. What is the period and frequency of this oscillator?
>
> The period is merely the total time it takes to go through one complete cycle. Therefore
>
> $$\begin{aligned} \text{period} &= 4 \text{ seconds on} + 4 \text{ seconds off} \\ &= 8 \text{ seconds} \end{aligned}$$
>
> So what is the frequency of this oscillator? Simple—just take the reciprocal of the period. That makes the frequency $\frac{1}{8}$ Hz.
>
> **Example 18.16** On a piano, the middle-C key plays a sound with a frequency of 261.6 Hz. What is the period of this sound?

> Since the frequency is the reciprocal of the period, it works the other way around as well—the period is the reciprocal of the frequency. Therefore, the period is simply $\frac{1}{261.6}$ second. Or, as a decimal, this is 0.00382 second.

There are many other factors that are important to various kinds of oscillations, such as the speed of transition between states, what percentage of time each state is achieved, and so on. However, the period/frequency is a good way to summarize the behavior of an oscillator into a single number.

18.2 The Importance of Oscillating Circuits

Oscillating circuits are important in electronics for a number of reasons. First, obviously, is blinky lights. Who goes into electronics without being enamored by blinking lights? But, more importantly, the following are all applications of oscillators in circuits:

> Sound Production: Sounds and tones are made by moving a speaker back and forth, which is moved back and forth by electricity oscillating between different voltages.
>
> Time Clocks: Every clock on earth operates by an oscillator. The clock simply counts the number of oscillations that have occurred to know whether or not to advance another time step (second, minute, etc.).
>
> Hardware Coordination Clocks: In computers and other advanced hardware, system events are coordinated based on signals from an oscillating circuit. When the clock changes state from off to on (or the reverse), then the circuit does the next step in the process. The clock keeps the various circuits synchronized and keeps them from stepping on each other.

> Radio Transmissions: Oscillators are used in radios in order to encode signals onto “carrier waves,” which are just oscillating signals run at a specific frequency.
>
> Servo Motors: A servo is a motor which moves an arm to a specific angle (i.e., think of a car’s steering wheel). Servos are usually operated by frequencies, where each frequency specifies a different angle to move.

Let’s look in more depth about how sound is produced by oscillation. You hear sound through your eardrum, which communicates vibrations it detects to your brain which you interpret as sound. Therefore, any sound that you hear is merely the vibrations of your eardrum. In other words, your eardrum *oscillates* back and forth, which you interpret as sound.

What makes your eardrum move back and forth? The answer is oscillations in the air. What makes those oscillations happen? Oscillations in the sound speaker. When the speaker moves back and forth, it moves the air which produces sound that you hear.

But what moves the speaker back and forth? Oscillations in the voltage and current supplied to the speaker. The speaker has a magnet attached to it which responds to changes in voltage in the wire. As the voltage in the wire increases, it moves the speaker in one direction. As the voltage decreases, the speaker moves in the other direction. Thus, the oscillations in voltage eventually wind up as sound for your ears.

18.3 Building an Oscillator

Let's think for a moment of what it would take to build an oscillator. An oscillator has (at least) two different states—on and off.[1] Then, you have a time period of changing from one state to the other.

What sort of circuit have we looked at that provides a time period?

As we saw in Chapter 17, RC (resistor-capacitor) circuits can provide us with time delays. We can then use these time delays as the time periods between the states of the oscillation. What we need is a circuit that will give out a state (we'll called it S1) for a time period (we'll call it T1) and then move to another state (we'll call it S2) and stay there for another time period (we'll call it T2). After T2 is finished, the circuit moves back to state S1 and begins again.

Now, it is possible that we might be able to build such a circuit using multiple RC timers with multiple comparators. It is doable, but it is harder than it sounds. Thankfully, there is an integrated circuit that is built for making such timers—the NE555, often referred to as just a "555 timer" or just a "555." The NE555 is a very flexible component, and engineers are constantly finding new ways to make use of it. However, we will just focus on using it as a basic oscillator.

It is easiest to describe how a 555 works if we start out by showing a simple circuit with it. The circuit in Figure 18-2 blinks an LED on and off.

The order of the pins on the actual 555 chip is a little confusing (see Figure 18-1). In order to make understanding the circuit in Figure 18-2 easier, I have simply put tags for each pin in the diagram, so you can see where they each belong *functionally*.

---

[1] You can build an oscillator with more than two states, but we won't be concerned with those circuits in this book, though the setup is similar.

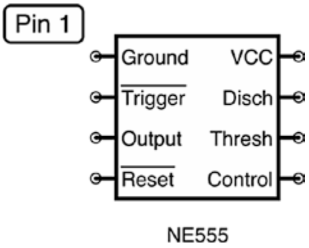
Pin 1
Ground
VCC
Trigger
Disch
Output
Thresh
Reset
Control
NE555

Figure 18-1. *The 555 Pinout Diagram*

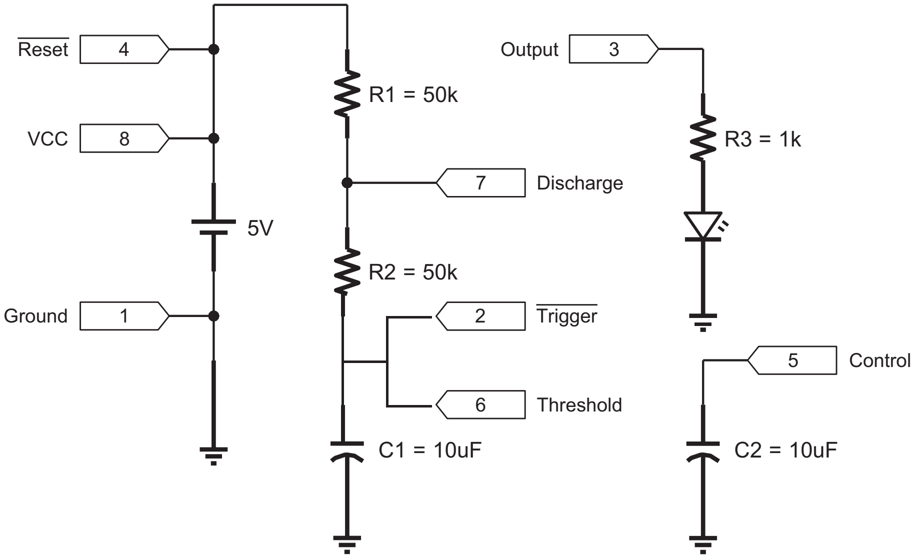
Reset
4
VCC
8
5V
Ground
1
R1 = 50k
7
Discharge
R2 = 50k
2
Trigger
6
Threshold
C1 = 10uF
Output
3
R3 = 1k
5
Control
C2 = 10uF

Figure 18-2. *A Schematic of a Simple 555 Oscillation Circuit*

First, let's look at the pins on the left-hand side of Figure 18-1 that are directly connected to the power rails. Pin 1 (`Ground`) and pin 8 (`VCC`) are easy enough—they are connected to ground and positive voltage, respectively. Pin 4 ($\overline{\text{Reset}}$) gets connected to the voltage source too. That is a reset pin which is activated if the pin receives a *low* voltage signal. Pins that are activated by low-voltage signals are often shown with a line over them.[2] In our case, we never want to reset the chip, so we will just tie the reset pin to positive voltage which will prevent any accidental resets from changes in voltages in the environment.

Now, on the right-hand side at the top, you can see the LED connected to pin 3 (`Output`). Pin 3 is simply the output pin. It is an active output, meaning that it will supply current to the output on its own. We don't need a pull-up resistor as we did for the LM393. Instead, we just need a current-limiting resistor for the LED. The regular NE555 output yields a voltage that is about 1.7 V less than the supply voltage and sources up to 200 mA before it breaks.

Note that there are other, low-power versions of the 555 timer that have other output characteristics. For instance, the LMC555 has an output voltage that is equal to the input voltage, but can only source about 100 mA. For our purposes, either one would work, as we are not doing anything precise with our output, nor are we sourcing very much current. However, any calculations we do will assume a typical NE555 component.

On the bottom right, pin 5 (`Control`) is connected through a 10 μF capacitor to ground. This is just a standard part of using the chip. You can't effectively include capacitors in integrated circuits larger than a few picofarads, so many chips specify certain capacitors be attached to certain pins. The NE555 uses a 10 μF capacitor on pin 5 to provide voltage stability. In Chapter 16, we learned that capacitors act essentially like little batteries. This capacitor is doing just that. It is providing a short-term supply of charge in case there is a sudden change in current needs within the chip. For instance, when the output goes active, there may be a sudden need for charge. This capacitor supplies a quickly available reservoir of charge to the chip so that sudden changes in current needs will not affect other chip properties.[3]

---

[2] Since most pins are activated by a higher voltage, pins that are activated with a lower voltage get the special designator.

[3] For most of our uses of the NE555, you can actually leave pin 5 unconnected—our usage of the chip is not so precise or power-hungry as to be affected in this way. Nonetheless, I show it connected so that you can see how it is *supposed* to be connected.

Down the middle of the schematic is where all of the action happens. This is what controls the oscillation. It is essentially an RC circuit with (a) an extra resistor and (b) a few tie-ins to the chip. As we will see shortly, the capacitor will alternately charge and discharge. The capacitor itself is attached to two voltage sensors on the chip. The first sensor, $\overline{\text{Trigger}}$, watches the voltage on the capacitor and activates if the voltage falls below $\frac{1}{3}$ of the supply voltage (the line over the name of the pin indicates that it activates on a low value). The second sensor, `Threshold`, watches the voltage and will activate if the voltage goes above $\frac{2}{3}$ of the supply voltage.

The oscillator works by moving the capacitor voltage back and forth between $\frac{1}{3}$ and $\frac{2}{3}$ of the supply voltage. It is fairly obvious to see how the capacitor charges—it is a basic RC circuit using both of the resistors R1 and R2. So how does the capacitor discharge? The capacitor discharges through the action of the `Discharge` pin, pin 7. When the 555 starts up, pin 7 essentially acts as if it were not connected to anything, so you can basically ignore it. When the capacitor goes over the $\frac{2}{3}$ supply voltage and triggers the `Threshold` pin, the 555 will then connect pin 7 to ground.

Think about what will happen if pin 7 is connected to ground. Will the capacitor charge? No, the current coming in through R1 will immediately go to ground because it is the easiest path. If the capacitor is charged up, is it above or below ground? It is obviously above ground. Therefore, current will flow *out of* the capacitor, through R2, and to ground. Thus, the capacitor discharges, but at a *different* rate than it charged. The RC circuit for the charging of the capacitor used both R1 and R2 for the resistance. However, the discharge of the capacitor only uses R2. Therefore, the discharge from $\frac{2}{3}$ voltage down to $\frac{1}{3}$ voltage will be faster than the charge up.

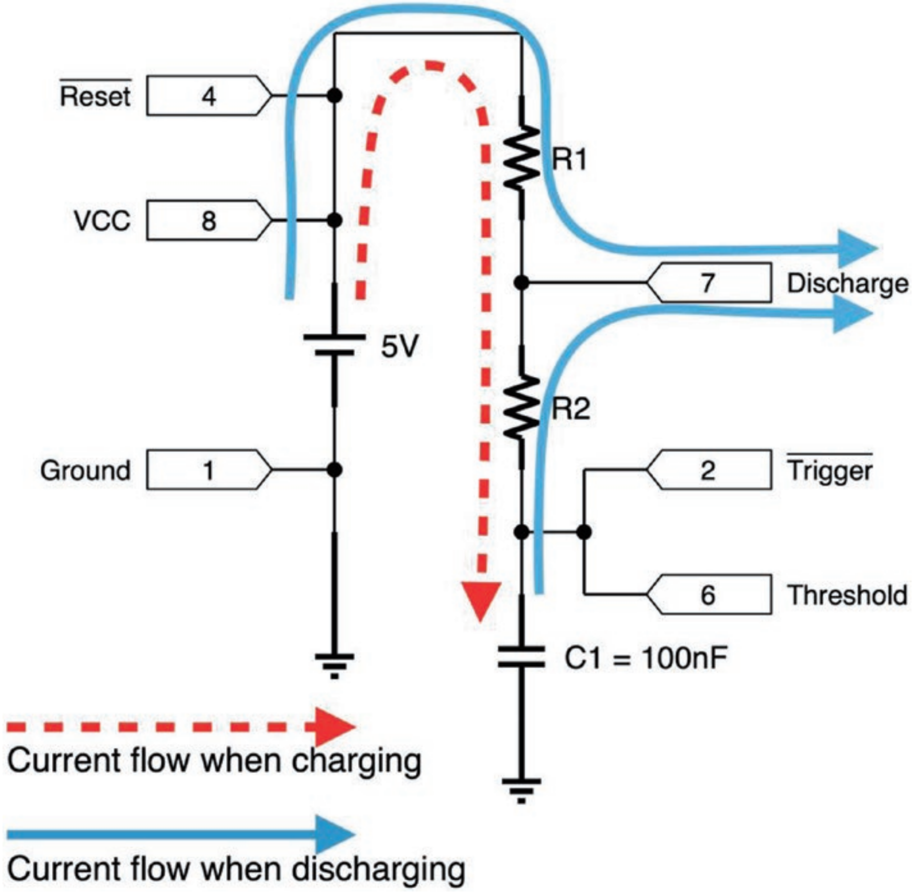
$\overline{\text{Reset}}$
4
VCC
8
5V
Ground
1
R1
7
Discharge
R2
2
$\overline{\text{Trigger}}$
6
Threshold
C1 = 100nF
Current flow when charging
Current flow when discharging

Figure 18-3. *Current Flow During the Charge and Discharge Phases*

Once the capacitor discharges down to $\frac{1}{3}$ of supply voltage, pin 2 ($\overline{\text{Trigger}}$) will detect this event and turn pin 7 off so that the capacitor can recharge again. Therefore, the capacitor will be continually charging and discharging from $\frac{1}{3}$ to $\frac{2}{3}$ of supply voltage, as the 555 turns pin 7 (connected to ground) on and off. This is also why having two resistors is so important. If we didn't have R1, when pin 7 connected to ground, it would make a short circuit between the supply voltage and ground.

Figure 18-3 shows the different current flows of this part of the circuit during the charging and discharging phases.

You may be wondering why there are two detectors connected to the capacitor and not just one. The reason for this is that having two different detectors allows for a *lot* of flexibility in how the 555 is wired. As mentioned earlier, there are a huge number of ways the 555 has been used, and some uses tie $\overline{\text{Trigger}}$ and `Threshold` to different parts of the circuit. For our purposes, both of these will always be used together.

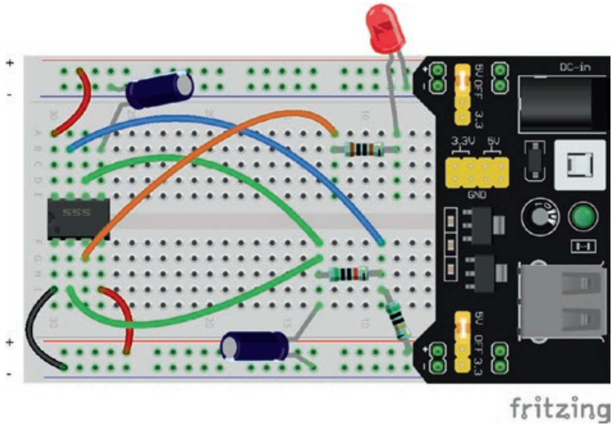

Figure 18-4. *The 555 Oscillator Circuit on a Breadboard*

So how does this charging and discharging of the capacitor affect the output? Quite simply, when the capacitor is *charging*, the output is *on*. When the capacitor is *discharging*, the output is *off*. Note that when we first turn on the 555, the output will be on for a little bit longer than for the rest of the time. This is simply because when the circuit first starts up, it is charging all the way from 0 V instead of from $\frac{1}{3}$ of supply voltage.

You can see the completed 555 circuit on a breadboard in Figure 18-4. This should blink the light on for about two-thirds of a second and off for about a third of a second. This gives the total period of about 1 second and a frequency of about 1 Hz. In the next section, we'll see how to use RC time constants to calculate our own on and off times.

18.4 Calculating On and Off Times with the 555

Remember that the capacitor is charging and discharging through an RC circuit. Therefore, we can use what we know about RC circuits to determine how long the output circuit will be high and low.

When the capacitor is *charging*, the capacitor is charging through *both* R1 and R2. In this situation, what is the RC time constant? First, we need to know the resistance. Since the resistance here is a series resistance, we can simply add them together:

$$\begin{aligned} R_T &= R_1 + R_2 \\ &= 50,000 + 50,000 \\ &= 100,000 \end{aligned}$$

So the resistance is 100,000 Ω, and we are using a 10 μF capacitor (which is 0.00001 F). Therefore (based on Chapter 17), the RC time constant is 100,000 $*$ 0.00001 = 1. So our RC time constant is 1 second. However, we are charging/discharging the capacitor in a strange way. We are not starting from zero (except for the first time)—instead we are usually starting from $\frac{1}{3}$ of supply voltage.

Go back to Chapter 17 and look at Figure 17-1. While we don't have a time constant for *exactly* $\frac{1}{3}$ of supply voltage, we do have one for 39.3%, which is pretty close. So, in the table, that occurs as 0.5 time constants. In this case, that is our *starting* value. Our ending value is when the capacitor is $\frac{2}{3}$ charged. That is very close to 63.2%, which is at 1 time constant.

Therefore, the *difference* between the points where we are $\frac{1}{3}$ charged and $\frac{2}{3}$ charged is $1 - 0.5 = 0.5$ time constants.

Since we were using approximate values from the table, this is only an approximation to see how this idea works. The distance from $\frac{1}{3}$ to $\frac{2}{3}$ of supply voltage actually takes 0.693 time constant. This is an important number to remember when using the NE555, as it will be used in all of your time calculations. Since our time constant is 1 second, that makes the calculation really easy: $1 * 0.693 = 0.693$ seconds.

Now, for discharge, remember that the point that the capacitor is discharging to is pin 7. This is *between* R1 and R2. That means that it is *only* using R2 to discharge, so the RC time constant will be based only on R2 and the capacitor. So the RC time constant is $50{,}000 * 0.00001 = 0.5$ seconds. Since the charge/discharge time between $\frac{1}{3}$ and $\frac{2}{3}$ is 0.693 time constants, the resulting time to discharge from $\frac{2}{3}$ down to $\frac{1}{3}$ is $0.693 * 0.5 = 0.347$ seconds.

The total period will be the total time for one cycle. This will be our charge time (0.693 second) plus the discharge time (0.347 second) which will give us a total of 1.04 seconds.

The value 0.693 looks like a strange number, but that will *always* be the value used for the number of time constants for charging/discharging between $\frac{1}{3}$ and $\frac{2}{3}$ of supply voltage. If you are going to use the 555 timer, it is best to just memorize it.

Also note that since the charging goes through *both* resistors while the discharge only goes through *one* resistor, the charge time (with the output on) will *always* be longer than the discharge time (with the output off).

> **Example 18.17** Let's say we have the basic oscillator circuit shown in Figure 18-2, but with a 2,000 Ω resistor for R1, a 6,000 Ω resistor for R2, and a 10 μF capacitor for C1. What will be the charge time, the discharge time, the period, and the frequency for our oscillator?

To find this out, it is easiest to first calculate the charge and discharge times separately and then use those to find period and frequency. The charge time will be calculated based on *both* resistors in our RC circuit. So the resistance will be 2,000 Ω+6,000 Ω = 8,000 Ω. The capacitance is 10 μF = 0.00001 F. This means that the RC time constant will be 8,000 Ω ∗ 0.00001 F = 0.08 seconds.

The number of time constants it takes to charge from $\frac{1}{3}$ to $\frac{2}{3}$ is 0.693. This means that the charge time will be 0.08 ∗ 0.693 = 0.0554 seconds.

The discharge happens only through R2. This means that the RC time constant will be 6,000 Ω ∗ 0.00001 F = 0.06 seconds. Since we will use 0.693 time constants, the total time it takes to discharge from $\frac{2}{3}$ down to $\frac{1}{3}$ is 0.693 ∗ 0.06 = 0.0416 seconds.

Now that we have the charging and discharging times, we find the period by merely adding them together. This means the period is 0.0554 + 0.0416 = 0.097 seconds.

The frequency is merely the reciprocal of this number, or $\frac{1}{0.097}$, which gives us 10.3 Hz.

Example 18.18 Now let's say that we want to

build an oscillator for which the LED stays on for 3 seconds and then goes off for 2 seconds. Assuming we keep our 10 μF capacitor, what values should we use for each resistor?

To do this, we need to solve for R2 first, since it is much easier. The discharge will be 2 seconds, which will be the same as 0.693 time constants. Therefore, we can write this as an equation:

$$\begin{aligned}
&2\text{seconds} = R \cdot C \cdot 0.693\\
&2\text{seconds} = R \cdot 0.00001 \cdot 0.69\\
&2\text{seconds} = R \cdot 0.0000069\\
&\frac{2\text{seconds}}{0.0000069} = R\\
&289{,}855\Omega \approx R
\end{aligned}$$

So the resistor, R2, needs to be 289,555 Ω.

The time to charge the capacitor, however, needs to be 3 seconds. This means the resistance (which will be *both* R1 and R2) will need to be calculated for this as well. Therefore, we will use a similar equation:

$$\begin{aligned}
&3\text{seconds} = R \cdot C \cdot 0.693\\
&3\text{seconds} = R \cdot 0.00001\text{F} \cdot 0.693\\
&3\text{seconds} = R \cdot 0.00000693\\
&\frac{3\text{seconds}}{0.00000693} = R\\
&434{,}783\Omega = R
\end{aligned}$$

So the resistance for charging the capacitor needs to be 434,783 Ω. However, this is the *combined* resistance for R1 and R2. Thankfully, though, we already know what we want for R2—289,855 Ω. Therefore, we can put these into an equation and solve for R1:

$$
\begin{aligned}
R_1 + R_2 &= 434,783\Omega \\
R_1 + 289,855\Omega &= 434,783\Omega \\
R_1 &= 434,783\Omega - 289,855\Omega \\
R_1 &= 144,928\Omega
\end{aligned}
$$

So now we know our values for R1 and R2—144,928 Ω and 289,855 Ω. Depending on our application, we would probably simply choose resistors that were close to that amount (like 150 kΩ and 300 kΩ) rather than trying to find a combination of resistors that hit that precise resistance. But, for solving equations, it is best to use exact values.

18.5 Choosing the Capacitor

Ultimately, there are no hard-and-fast rules for choosing capacitors. As long as the RC time constant yields the right value, then you can compensate for pretty much any size capacitor with the right size of resistor. However, sometimes just having a little guidance helps people get started, and there are a few situations that you need to watch out for.

First of all, the size of the capacitor will affect the size of the resistor you need for a given time constant. If all you have are smaller-sized resistors, then you should probably use a larger capacitor to compensate.

However, using a smaller capacitor with larger resistors gives a very large advantage in the amount of wasted current. If you are using smaller resistors, then Ohm's law indicates that we will have larger currents for a given voltage. Since $V = I \cdot \mathrm{R}$, if you lower the R, you will increase the I. So having a higher resistance means that your RC circuit will use much less current.

Remember that we aren't actually *using* the current to *do* anything except keep time. The current in our RC circuit is not used to power the LED (the 555 does that through the power source), and it isn't used for anything else; it is just used to keep time. Therefore, pretty much all current used by the RC circuits is wasted. We have to use it, but any smaller currents we can get away with, we should!

This gets even more pronounced when the capacitor is discharging. When pin 7 (`Discharge`) switches to ground, not only does it discharge the capacitor but it also creates a useless waste of current going from the voltage source through R1. In our oscillator configuration, you can't get rid of this, but having larger resistors will reduce the amount of waste.

So, in short, if you have higher-valued resistors to compensate, your circuit will waste much less current by using smaller capacitors.

Review

In this chapter, we learned the following:

1. Because this material is based so heavily on RC circuits, you may want to review the material in Chapter 17.
2. When something oscillates, it moves back and forth between a set of values.
3. In electronics, oscillators usually refer to circuits whose output voltages go back and forth between two different values.
4. Oscillations are usually described by their period or frequency.
5. The period of an oscillation is the time that it takes to go through one entire cycle (usually in seconds).
6. The frequency of an oscillation is the number of times that an oscillation occurs per second, which is simply the reciprocal of the period (i.e., $\frac{1}{\text{period}}$), with a unit of hertz (cycles per second).

7. Oscillating circuits are used in a number of circuit applications, including sound production, time keeping, coordinating activities, radio transmissions, and motor control.

8. Oscillating circuits are often built with RC time circuits which control the time periods of the oscillations.

9. The NE555 (often referred to as a 555 timer or just the 555) is an integrated circuit which allows for several timing applications including making oscillating circuits.

10. The NE555, when set up as an oscillator, is controlled by two resistors and a capacitor.

11. The NE555 can be set up to monitor the charge of the capacitor. It has two pins for monitoring the voltage, $\overline{\text{Trigger}}$ (which checks for the capacitor to drop below $\frac{1}{3}$ of supply voltage) and `Threshold` (which checks to see when the capacitor is charged above $\frac{2}{3}$ of supply voltage).

12. The NE555 cycles between charging and discharging the capacitor from $\frac{1}{3}$ of supply voltage to $\frac{2}{3}$ of supply voltage and back (using $\overline{\text{Trigger}}$ and `Threshold` to check the voltage level).

13. The `Discharge` pin of the NE555 acts like it is disconnected when the capacitor is charging and is connected to ground to allow for a path for the capacitor to discharge when in the discharge part of the cycle.

14. The `Output` pin of the NE555 is turned on (higher voltage level) while the capacitor is charging and off (low voltage level) when the capacitor is discharging.
15. The NE555 output operates at about 1.7 V less than the source voltage when on and can source a maximum of 200 mA to the output circuit.
16. The $\overline{\text{Reset}}$ pin can be given a low voltage to reset the whole device. If it is unused, it should just be tied to a positive voltage line.
17. The `Control` pin is attached to a capacitor in order to regulate and stabilize the circuit operation.
18. The RC circuit takes 0.693 time constants to charge from $\frac{1}{3}$ voltage to $\frac{2}{3}$ voltage or to discharge the other way. Therefore, once you have the RC time constant, multiply by 0.693 to find the actual time it will take.
19. The RC circuit utilizes *both* resistors when charging, but only *one* resistor when discharging. This means that charging and discharging each have their own RC time constants.
20. Because the charge circuit uses two resistors while the discharge circuit only uses one, the charging portion of the oscillation period will always be at least a little longer than the discharging portion.
21. The period of the oscillation is the combination of both the charge time and the discharge time.

22. When calculating for resistors for an RC time circuit, it is best to calculate the discharge resistance first, since it only uses one resistor (R2). Then you can calculate the combined resistance for the charge circuit, which allows you to deduce what the R1 resistor should be.

23. When choosing components for a given RC time constant, you can choose a variety of combinations of capacitor/resistor values for the same result. Choosing a smaller capacitor and a larger resistor will save power in the circuit.

24. An equation for determining the frequency of a 555 oscillator circuit can be found in Appendix D, Section D.4, "555 Timer Oscillator Frequency Equation."

Apply What You Have Learned

Figure 18-2 will be used as the basis for the problems in this section:

1. Copy Figure 18-2 to a piece of paper including the current flows on charge and discharge.

2. Why is R1 important? What would happen if we just replaced it with a wire?

3. Why are there two different pins on the NE555 connected to the capacitor? What type of circuit (that we have discussed in this book) do you think they are connected to inside the chip?

4. Why is the charging time of the NE555 always at least a little longer than the discharging time?

5. Why does the NE555 stay in the on state a little longer when the circuit is first turned on?
6. Let's say that we wanted our circuit to be on for 2 seconds and off for 1 second. Keeping the same capacitor, what values should we use for R1 and R2 to accomplish that?
7. Let's say that we wanted our circuit to be on for 10 seconds and off for 3 seconds. Keeping the same capacitor, what values should we use for R1 and R2 to accomplish that?
8. The factory called and said that they were out of the capacitor we wanted for the circuit and instead only had a 23 µF capacitor that we could use. Recalculate the previous problem using this new capacitor value.
9. How much current is our output sourcing from the chip?
10. When the chip first turns on (and thus the capacitor is empty and at 0 V), how much current is the RC circuit in Figure 18-2 using?

CHAPTER 19

Producing Sound with Oscillations

In Chapter 18, we learned to make an oscillator circuit. Oscillations affect a lot of areas of electronics, but the one that is most directly usable (well, besides making blinky lights) is in making sounds.

19.1 How Sound Is Produced by Speakers

As we mentioned in Chapter 18, sound is the result of vibrations (i.e., oscillating motions) in the air. These vibrations are produced in electronics with a speaker when the speaker oscillates back and forth. The speaker oscillates back and forth because it is connected to a magnet and a coil. When the electricity in the coil oscillates between more positive and more negative charges, the magnet is attracted to and repelled from the coil, moving the speaker back and forth.

The frequency of this oscillation will determine the pitch of the sound (i.e., how high or low the sound is). If the frequency is higher (shorter periods), then the pitch is higher. If the frequency is lower (longer periods), then the pitch is lower.

© Jonathan Bartlett 2020

J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_19

Now, compared to the blinking lights we worked with in Chapter 18, sounds that you can hear are all of a much higher frequency. The bottom end of the hearable audio is about 20 Hz—that is, 20 cycles per second. The top end is about 20 kHz (i.e., 20,000 Hz). These values vary by individual, but this is a pretty good guide for most people. If we tried to blink an LED that fast, we just would not be able to see it—it would just look like a dimmer-than-normal light.

19.2 Graphing Electricity

Now, when we deal with circuits that are oscillating quickly, we need to have a way to visualize what is happening. With circuits that have one or two states, we don't need to visualize much—we just need to calculate what the values are. However, with circuits that are continually oscillating (such as sound circuits), it is helpful to be able to graph what the voltages are at different points in time.

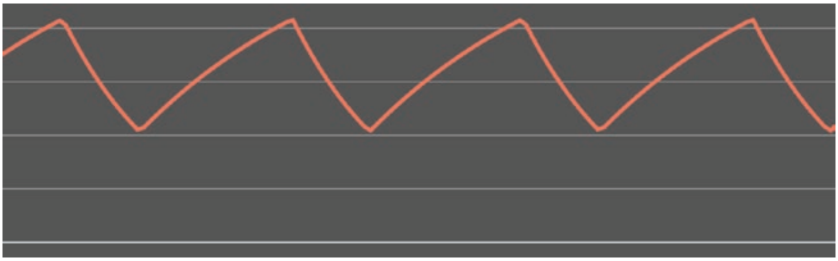

Figure 19-1. *A Graph of the Capacitor Voltage in a 555 Oscillator*

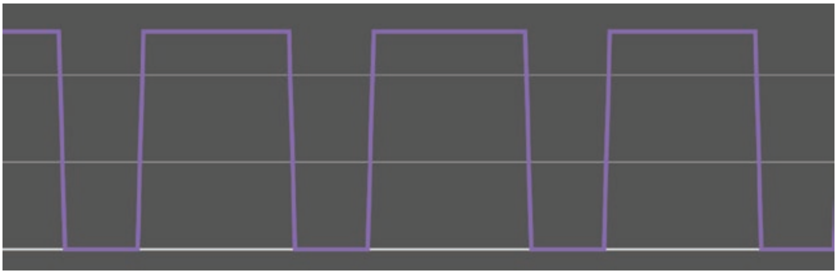

Figure 19-2. *A Graph of the Output Voltage in a 555 Oscillator*

There are several ways to graph electricity. One of the most common ways is to plot *time* on the *x*-axis and *voltage* on the *y*-axis. This gives a visible representation of how the voltage is varying within the circuit. Now, because each point on a circuit has a different voltage, you have to specify *which point* on the circuit you are graphing. For instance, if you take the 555 oscillator circuit from Chapter 18 (i.e., Figure 18-2), the two places whose voltages you might want to graph are (a) right before the capacitor and (b) right after the output pin.

These will give *very* different graphs, but both of them are instructive. The graph of (a) will show you what the input side of the circuit looks like, and the graph of (b) will show you what the output side of the circuit looks like.

Notice that Figure 19-1 changes fairly smoothly (except when it changes directions). This is because capacitor charging causes the voltage to change smoothly over time. It charges up until it reaches $\frac{2}{3}$ of supply voltage and then discharges down to $\frac{1}{3}$ of supply voltage. It does this over and over again as you can see on the graph.

The output, however, is either *on* or *off*. Therefore, Figure 19-2 shows a very jagged graph. This is known as a **square wave**, because it is basically rectangular. The voltage just goes from zero to the target voltage almost instantly, and then after it stays there a bit, the voltage goes back down again just as instantaneously as it rose.

In addition to graphing voltage, we can also graph current in the same way, though it is harder to measure. Because voltage can be measured with simple probes and measuring current requires replacing a wire, most of the time, people just focus on voltage measurements.

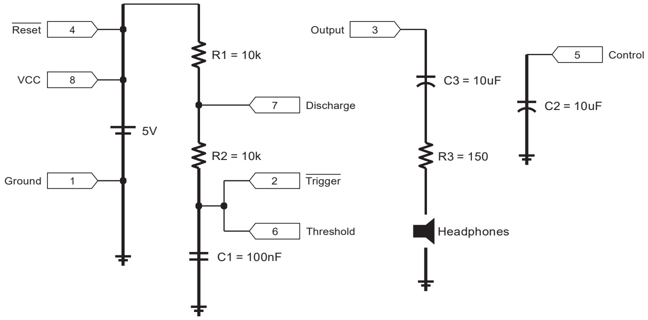

Figure 19-3. *Schematic for a Simple 555-Based Tone Generator*

Special devices known as **oscilloscopes** can be used to measure voltage changes over time. Graphs such as those in Figures 19-1 and 19-2 can be made by hooking up an oscilloscope to the circuit while it is in operation.[1] Oscilloscopes used to cost thousands of dollars, but now you can pick up a small handheld oscilloscope for under $50.

Usage of an oscilloscope is outside the scope of this book, but the basics are simple enough. An oscilloscope usually has an active probe and a ground probe. You hook up the ground probe to the ground on your circuit, and you hook up the active probe to the point on the circuit that you want to measure. Then, you look at the screen, and it will graph for you what the voltage is at that point over a period of time. Most oscilloscopes have all sorts of adjustments, such as how wide of a voltage range you want to measure and how long of a time window you want to show on the screen. In any case, oscilloscopes provide graphical views of oscillating voltages.

19.3 Outputting a Tone to Headphones

In this section, we are going to modify the circuit in Figure 18-2 to add in a headphone output. Figure 19-3 shows the final circuit, and Figure 19-4 shows how to build it. This circuit has several changes to our original circuit that we will discuss in this section.

---

[1] The graphs depicted in the figures are actually idealizations, however, not actual output from an oscilloscope.

The first changes to this circuit are to the RC time circuit. Because we want the circuit to be in audio range, we are making a much higher frequency (i.e., a shorter period). To do this, we swapped out the electrolytic capacitor for a ceramic disk capacitor of 100 nF. Then, we dropped the resistors down to 10 kΩ each. This will make a tone with a period of 0.003 seconds. Taking the reciprocal gives us a frequency of about 333 Hz. This is well within hearing range for most people. You can adjust the resistors and capacitors how you wish for different tone frequencies. We are merely using the equations from Chapter 18 to determine the period and frequency.

The strange change is in the output circuit. When we were blinking the LED, we just used a medium-sized resistor to limit output current. Here, however, we are using a combination of a capacitor and a resistor. This *looks like* an RC time circuit—and it would be, if it had a constant voltage going to it. However, in this circuit, the voltage going to the RC circuit is fluctuating, and that introduces a whole host of new considerations.

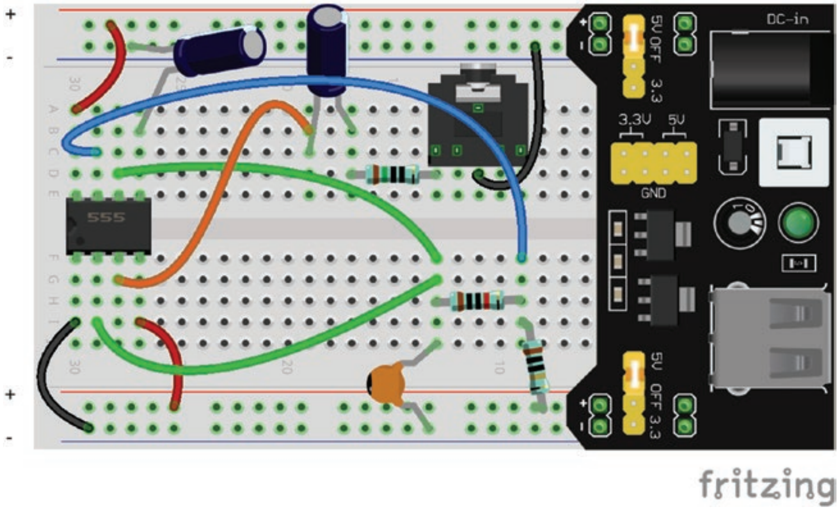

Figure 19-4. *Breadboard Layout for a 555-Based Tone Generator*

19.4 AC vs. DC

In Chapter 3, we introduced the ideas of *direct current* (DC) and *alternating current* (AC). In DC circuits, the circuit has a more-or-less constant voltage. Our power supply is at a certain voltage, our output is at a certain voltage, and so on. In AC circuits, however, the voltage oscillates back and forth. These oscillations introduce a whole new set of considerations.

In pure AC circuits, voltage and current actually swing both to positive and negative values. What is a negative current? A negative current happens when the electricity flows in the *opposite* direction. Thus, in a *pure* AC circuit, electricity flows back and forth, switching between directions while it is running, spending about even time in each direction.

Some components, such as speakers, assume that they are going to receive or process pure ACs. That is, they *expect* the current to switch direction at some point. As mentioned earlier, the speaker is driven by changing voltages in a coil, which alters the attraction of the coil to a magnet. These variations in attraction and repulsion move the coil back and forth, which moves the speaker cone, which vibrates the air, which vibrates your eardrum, which you hear. Speakers expect the oscillating current to go back and forth each direction about evenly.

However, in our circuit, the voltage only swings between a positive voltage and 0 V. If you look at the graph in Figure 19-2, notice that while the voltage goes up to a positive voltage, the *lowest* that it ever goes is zero. Therefore, it is not a pure AC circuit, because it is not centered around 0 V. Instead, the AC has a **DC bias voltage**. That is, the output voltage can be thought of as a combination of an AC voltage signal added to a DC voltage.

19.5 Using Capacitors to Separate AC and DC Components

Sometimes having a DC bias is good, and sometimes we need to separate out our AC and DC signals. Capacitors are excellent at this. Capacitors are said to *allow* AC to flow and to *block* DC. This is a simplistic view, and we will get a little more nuanced in future chapters, but it works for now.

To understand why this happens, let's remember what a capacitor does physically. It stores charge—like a small battery. Remember, even though the pathway between the plates of a capacitor is blocked, when a positive charge accumulates on one side, it pulls electrons onto the other side, which means that the other side beyond the capacitor gets more positive as well. Do you remember what happens when the capacitor *first* starts charging? The capacitor basically acts as a short circuit. As the capacitor fills up, it starts acting more and more like a resistor, preventing current from flowing.

Thus, when given a constant voltage, once the capacitor is nearly full, it essentially stops getting more charge (or comes very close to it). Therefore, once it is full, it is basically *blocking* any more current from flowing. Since what goes onto one side of the capacitor affects what is on the other side, this same blockage will block the *transmission* of the current to the other side as well. Thus, in the long run, if you have a DC voltage, the capacitor will eventually block current flow. Thus, the current on the *other* side of the capacitor will be *zero*.

However, now consider what happens if the voltage *changes*. If the voltage changes, then the capacitor will allow current to flow through. If the voltage goes up, then the capacitor has more storage available (since the amount of storage depends on voltage). While it is charging to the new level, current will be flowing. If the voltage goes down, then the capacitor will actually have to discharge back into the circuit.

Now, if the capacitor discharges back into the circuit, *which way is the current flowing*? Well, that means the current from the capacitor is now flowing in the *opposite* direction!

Thus, what ends up happening is that the capacitor winds up only sending *changes* in voltage to the other side of the circuit! While the capacitor is charging and discharging, charge is moving through both sides of the capacitor. Thus, when the voltage (and therefore the charge) on the capacitor is changing, current is indeed flowing. Now, when the voltage stays the same, no electrons move, no matter what the voltage is. Therefore, even though you have a positive voltage on one side, if the voltage is not changing, no current is flowing. Therefore, on an unchanging current, the voltage on the other side of the capacitor will be zero.

Since the changes go up and down, but the DC bias voltage does not go up and down, this has the effect of sending AC through the capacitor, but blocking the DC bias voltage.

Therefore, capacitors are often used in series in a circuit to couple together parts of the circuit that are pure AC with other parts of the circuit that have a DC bias. Since the output of the 555 is DC biased and speakers generally are optimized for pure AC output, the capacitor in our output circuit filters out the DC bias and gives a pure AC signal.

A capacitor used in this way is known as a **coupling capacitor**.

Alternatively, if for some reason you wanted *only* the DC component, you could wire the capacitor to go to ground. This sends the AC part of the signal to ground, leaving only the DC component in the circuit. In such a scenario, the capacitor is being used as a **filter capacitor**.

Now, using a coupling capacitor does add in a small amount of distortion, both because DC bias voltage does have some effect on the signal and because different capacitor values are optimal for coupling different frequencies. For audio applications, a capacitor somewhere between 10 nF and 10 μF should be used.

19.6 Speaker Wattage

Next in the circuit is the R3 resistor. The R3 resistor is there to limit the output signal to the headphones. Headphones and speakers are all a little strange—there is very little standardization. Different headphones have different resistances in the speakers and different tolerances for how much wattage can be put through them. Speakers are usually rated by (a) how much resistance they normally add to the circuit, (b) how loud they are at a given wattage of power, and (c) what their maximum wattage is. Even though there is quite a bit of variance, for headphones, you can be fairly safe by estimating that the headphones will act as 16 Ω resistors and that they can consume a maximum of 20 mW.

The output of the 555 timer is 1.7 V less than my supply voltage. So, for a 5 V source, the output will be 3.2 V. However, since we have now centered it on 0 V, the output will actually only be half of that—1.6 V. We can use the formula $P = \frac{V}{(R^2)}$ from Chapter 10 to determine how many watts this will supply. If the speaker is 16 Ω, then the power used by the headphone without the extra resistor will be $\frac{(1.6^2)}{16}$ = 0.16 W = 160 mW. Therefore, the R3 resistor pulls out some of that power.

With the resistor, the combined power usage of the resistor and the headphone speaker is 166 Ω. Therefore, the power here is $\frac{(1.6^2)}{166}$ = 0.015 W = 15 mW. Remember, though, that this is the *combined* power of both the resistor and our speaker. The speaker is only supplying about $\frac{1}{10}$ of the resistance, so it is only using up to $\frac{1}{10}$ of the power, or 1.5 mW.

Now, the calculation for this is relatively easy because we are working with square waves. The calculation becomes harder for different types of output, because you have to take into account the *shape* of the wave.

However, we won't go into that in this book. This calculation using the peak voltage provides a simple method which will give you a good starting point, and you can judge for yourself if you need more or less resistance for your own circuits.

19.7 Sound Control

Now, as I mentioned, all headphones are different, and, with the exception of *having* a coupling capacitor, they all have different ranges and need different amounts of power. Usually this is accommodated by having a volume control.

This is usually done with a **variable resistor**. On a variable resistor, you usually turn a knob and the resistance changes. Variable resistors go by different names—variable resistor, adjustable resistor, rheostat, and trimming resistors, though they each have slight differences. Trimming resistors are the easiest to use for this, but probably the hardest to find.

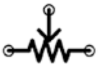

Figure 19-5. *The Potentiometer Symbol*

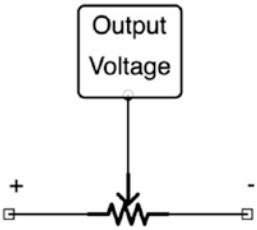

Figure 19-6. *The Potentiometer as a Voltage Divider*

To use a variable resistor, you would need to find one with the right range of values and then simply replace the output resistor R3 with your variable resistor.

A more common type of component which can be used as a variable resistor is a potentiometer. A potentiometer is basically a variable voltage divider. The symbol for the potentiometer is shown in Figure 19-5. It has three leads—input, ground, and output. When you turn the knob, it adjusts the relative resistance of each side of the voltage divider to vary what is coming out of the output.

Potentiometers were built for working as a variable voltage divider. Figure 19-6 shows what this looks like. Essentially, the user controls a knob which alters the voltage coming out by changing the voltage division between the resistors.

However, you can also use a potentiometer as a simple variable resistor by just hooking up one side of the potentiometer. Figure 19-7 shows this configuration.

The simplest way of using the potentiometer for a volume control is simply setting it up like Figure 19-7, with the input coming from the coupling capacitor and the output going to the speaker. Potentiometers vary widely in how much resistance they offer, so if we were to add one as a volume control, we would need to make sure its resistance range approximately matches what is needed (around 150 Ω).

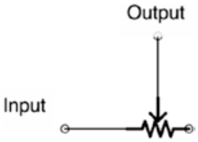

Figure 19-7. *The Potentiometer as a Variable Resistor*

Review

In this chapter, we learned the following:

1. Sound is the result of oscillations in the air that is picked up by your eardrums.
2. Sound can be generated by an oscillating speaker cone vibrating against the air.
3. The oscillations in a speaker cone are made by a magnet being drawn and repelled to another magnet by oscillations in electrical power.
4. Sound generally has higher frequencies (i.e., shorter periods) than what we use for flashing lights.
5. Audible sound frequencies range from 20 Hz to 20 kHz.
6. On circuits where voltage and/or current changes with time, it is often helpful to plot their changes at specific points on a graph to help visualize what is happening.
7. The output of the 555 timer is known as a square wave because it has steep changes in voltage at specific points, making a rectangular-shaped graph.
8. An oscilloscope can be used to measure and record these types of graphs.
9. Voltages that oscillate quickly with time are known as alternating current or AC voltages.
10. Oscillating voltages that do not spend roughly equal amounts of time on either side of 0 volt are said to have a DC bias voltage.

11. The DC bias voltage is how much voltage you would have to subtract from the voltage in order for the voltage to be centered on 0 volt.
12. Most circuits have DC bias voltages, but most output devices (such as speakers) need the DC bias voltage removed and only use the AC component of the signal.
13. DC bias voltage can be removed by placing a coupling capacitor in series with the output signal. The capacitor will only transmit voltage *changes*, thus removing the DC bias voltage.
14. Speakers and headphones vary considerably in their properties.
15. The most important considerations are how many ohms of resistance they have and how many watts are needed to drive the sound.
16. For headphones, a good starting assumption is that the speaker is 16 Ω and needs under 20 mW of output.
17. If your output signal is larger than this, be sure to add a resistor in series to consume some of the output signal's power.
18. The output resistor can be replaced with a variable resistor or a potentiometer to have a user-controlled output setting, also known as a volume control.
19. A potentiometer is a variable voltage divider. However, by using only the positive and output terminals, it can be used as a variable resistor instead.

Apply What You Have Learned

1. Why is the input voltage signal to the 555 timer smooth, but the output is square?
2. Will increasing the resistance of R1 and R2 make the pitch higher or lower?
3. Can you think of a way to modify the circuit so that the pitch of the sound can be adjusted while the circuit is on?
4. Given the circuit in Figure 19-3, what power would be delivered to your headphones if you were using an 8 Ω speaker instead of the 16 Ω speaker? What about a 100 Ω speaker?
5. The frequency of the A note above middle-C on the piano is 440 Hz. Design a modification to this circuit that will yield this frequency.
6. The output of the 555 is 1.7 V less than the power supply used. What is the effect on the wattage to the headphones if we change the power supply to be a 9 V power?

CHAPTER 20

Inductors

In this chapter, we will begin our study of **inductors** and **coils**.

20.1 Inductors, Coils, and Magnetic Flux

In Chapter 16, we learned that capacitors stored charge by using an electric field. We also learned that capacitors continued to hold their charge even after they were disconnected from the rest of the circuit. We learned in Chapter 19 that capacitors blocked DC signals but allowed AC signals to pass through.

20.1.1 What Is an Inductor?

An inductor is kind of like the capacitor's evil twin. It behaves in some ways that remind us a lot of capacitors, but it is operating on a different aspect of electrical behavior. Capacitors store charge and release it in response to voltage changes. Inductors, instead, store **magnetic flux** and release it in response to current changes. Thus, capacitors tend to even out voltages, and inductors tend to even out currents.

Figure 20-1 shows what an inductor looks like in a schematic. Inductors are non-polarized, so they can be wired in either direction. The symbol loosely represents a coil of wire, because that is what inductors are essentially made of.

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_20

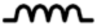

Figure 20-1. *The Symbol for an Inductor*

20.1.2 What Is Magnetic Flux?

When current moves in a wire, it creates a very tiny magnetic field. It's not large enough to pick up anything, but it is present. Flux is the total amount of magnetism present in an inductor. Since this is an engineering book, not a physics book, we won't spend a lot of time trying to come to grips with what flux is and how it works. We will simply understand that it is the total size of the magnetic field of a component, and it is measured in Webers (Wb). To get a feel for what a Weber is like, a small bar magnet has a flux of around $\frac{1}{1000}$ Wb.

As a sidenote, when we say that something is magnetic, we are usually not talking about the total flux of something, but of the *density* of that flux—the amount of flux in a given area. Flux density is measured in Teslas (1 Tesla is $\frac{1\text{Wb}}{1\text{meter}^2}$), but we will not concern ourselves with Teslas or flux densities for this course.

20.1.3 What Is the Difference Between Electric and Magnetic Fields

Now, when we talked about capacitors, we talked a little bit about electric fields. Electric fields and magnetic fields are related, but different. An electric field is generated by a difference in charge between two surfaces. Electric fields can exist even when no current is moving at all—there just has to be two differently charged surfaces. Basically, any charged object will induce a force on another charged object for a certain distance, and this is the electric field.

The magnetic field, however, only exists when a charge is moving. When a charge is in motion, in addition to the electric field, it also creates a *magnetic* field surrounding itself, going perpendicular to the wire. This field also exhibits a force on charged objects within a small distance, but a much larger distance than the electric field. Additionally, when the current in the wire slows down, the magnetic field will convert its own energy into current in the wire.

The size of a magnetic field within a single wire is very, very small and can be ignored. However, the strength of the magnetic field of a wire can be increased by wrapping (coiling) it around a cylindrical core. This aligns the magnetic fields of the wires so that the fields can each be added together to produce a combined magnetic field. Each turn around the core adds an additional bit to the magnetic field. Devices based on this principle are called coils or **inductors**.

When an inductor is first turned on, most of the current goes into creating the inductor's magnetic field. Then, as the magnetic field is setting up, it allows more and more current to flow, until it hits a steady state. Then, if the current flow reduces, the energy from the magnetic field gets transformed into current and goes out to try to make up the difference.

While a capacitor stores energy as a charge and releases it when there is a voltage change, an inductor stores energy as magnetic flux and releases it when there is a current change. An inductor's inductance is measured using the Henry (H). The core equations of capacitors and inductors are related as well. The defining equation of a capacitor is $Q = C \cdot V$, where Q is the charge in the capacitor. The defining equation of an inductor is just as simple:

$$\phi = L \cdot I \tag{20.1}$$

In this equation, φ is the magnetic flux in the inductor (in Webers), L is the inductance (in Henries), and I is the current (in amperes). The **Henry** is the primary unit of inductance and has an abbreviation of H.

Example 20.19 If I have a 50-microhenry inductor and have a constant current of 2 milliamps, how much magnetic flux is being stored?

$$\begin{aligned}
&\phi = L \cdot I \\
&L = 50\mu H = 0.00005 \text{ H} \\
&I = 2mA = 0.002 \text{ A} \\
&\phi = 0.00005 \cdot 0.002 \\
&\quad = 0.0000001\, Wb = 0.1\ \mu Wb
\end{aligned}$$

Therefore, we have 0.0000001 Weber of flux, or 0.1 microweber of flux.

Therefore, when a relatively constant amount of current flows, it is easy to calculate the amount of flux in the magnetic field.

20.2 Induced Voltages

Changes in the flux of a magnetic field induce voltages. The amount of voltage induced depends on how fast the flux changes. The induced voltage is essentially the change in the flux divided by the number of seconds that it took for the change to occur. If the flux decreases, the energy is converted into voltage, which goes up. If the flux increases, the energy is taken from the voltage, which goes down. So, if there was a decrease of 10 Wb over the course of 2 seconds, then the amount of voltage induced would be $\frac{10\text{Wb}}{2\text{seconds}} = 5\text{V}$.

The formal statement of this is known as Lenz's law. It is given as follows:

$$V_{\text{average}} = -\frac{\text{change in}\,\phi}{\text{change in time}} \tag{20.2}$$

The average voltage produced by the change in a magnetic field over a given time period is the same as the amount of flux that was reduced by the field divided by the same time period.

Now, the amount of flux in the inductor will change with the current, as indicated by Equation 20.1. When the current decreases, it will cause the flux to be converted into voltage. When the current increases, it will add more flux into the magnetic field, using up voltage.

When the amount of flux changes, it creates a voltage. The amount of flux change (in Webers) divided by the amount of time it took to change (in seconds) will yield the voltage produced by the change. Therefore, when the current drops, the flux will be converted to voltage. If we add in voltage, it will also increase the current.

20.3 Resisting Changes in Current

Ultimately, what an inductor is doing is resisting changes in current. When the current goes up, the current is first used to construct the magnetic field, and therefore it takes some time to allow the full amount of current through. When the current goes down, the flux of the magnetic field decreases, which gets converted back to voltage, which induces more current.

The inductor, with regard to current, is the opposite of the capacitor. With direct current, the capacitor, when first connected to the circuit, acts as a short circuit—the current basically flows right through until the capacitor starts to fill up. When the capacitor is full, then it acts like an open circuit—blocking any more current from flowing.

The inductor is the opposite. When first connected to the circuit, the inductor acts like an open circuit—the current is basically being used to construct the magnetic field. As the magnetic field strength increases, more and more current is allowed to flow. Once the magnetic field flux matches what it should be for the given current, the current flows freely, as if the inductor were not even there, like a short circuit.

In summary, we can say that a capacitor acts as an open circuit for sustained DC and as a short circuit for alternating current. The inductor acts as a short circuit for sustained DC and as an open circuit for alternating current.

20.4 Analogy from Mechanics

If you are into mechanics, you can think of an inductor as a flywheel. Flywheels store mechanical energy into the motion of a large or heavy wheel. It takes energy to get the wheel spinning, but once it is spinning, the wheel itself can continue to drive the mechanism. The wheel is used to even out the drive of the other components of the mechanism.

For instance, an engine driven by pistons will have somewhat irregular motion, because the force of the pistons is not constant through their motion. However, attaching a flywheel will even out the energy output. When the pistons are delivering more power, most of the excess will be going to the flywheel, and when the pistons are delivering less power, the mechanism can be driven by the force of the flywheel.

20.5 Uses of Inductors

Inductors are used in a very wide variety of applications—much more so than capacitors. Capacitors are used primarily for their effects *within* electronic circuits—filtering noise, blocking DC, coupling DC-biased circuits, and so on. However, the electric field created on the plates of the capacitor is fairly useless on its own because the field fades so quickly with distance. Inductors are used for similar things as capacitors within circuits (their operation being largely the counterpart of what a capacitor would do), but additionally, the magnetic field of the inductor has innumerable uses because its effects are stronger further away. These uses include the following:

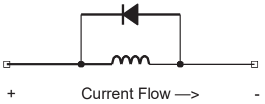

Figure 20-2. *A Snubber Diode*

- Inductors are used to make **electromagnets**, for picking up and setting down large objects.
- Inductors are used to make **transformers**, which use the magnetic field in one inductor to generate a current in another inductor.
- Inductors are used to make motors, where the electromagnetic coils are used to turn a shaft.
- Inductors are used to make speakers, which use the magnetic field to move the speaker diaphragm back and forth.
- Inductors are used to make relays, which use the magnetic field to open and close switches.

These are just some of the uses of inductors. In short, the magnetic field created by the inductor is what allows many of the interfaces between electronic devices and the real world.

20.6 Inductive Kick

One important thing to keep in mind when using inductors is the concept of inductive kick. Lenz's law (Equation 20.2) shows that decreases in the magnetic flux will induce a voltage. Since the magnetic flux is divided by time to get the voltage, that means that a shorter timespan will generate a larger voltage. Thus, a sudden decrease in magnetic flux will cause a voltage spike.

But what controls the flux? If you go back to Equation 20.1, we can see that a decrease in current will decrease the magnetic flux. Therefore, a sudden decrease in current will cause a sudden decrease in the magnetic flux which will then cause a large voltage spike within the inductor itself. This spike is known as an **inductive kick** and can damage nearby electronic components. To avoid an inductive kick, a **flyback diode**, also called a **protection diode** or a **snubber diode**, can be used.

A flyback diode is a regular diode (*not* a Zener diode) wired backward across the terminals of an inductor. Under normal operation, the diode won't conduct. However, when the diode is shut off, it provides a safe path for the generated voltage to pass through (back through the other side of the inductor) and allows the inductor to bleed off voltage in a much more controlled manner.

Figure 20-2 shows how this is wired up. As you can see, under normal operation, the current flows in the direction of the arrow. Since the left-hand side is more positive than the right-hand side, no current can flow through the diode. However, when the current is shut off, the right-hand side becomes more positive very quickly, and the diode feeds the excess current back through the inductor to allow the extra energy to dissipate more slowly.

You might be wondering why the voltage doesn't just shift back to the left side on its own once the right-hand side is more positive. The answer is that the magnetic field that is generating the voltage is itself directionalized within the inductor. The field's movement is such that it is essentially pushing current to the right (which is why a voltage is being produced). Therefore, the flyback diode carries this excess voltage back to the left-hand side of the diode.

Flyback diodes are used with all sorts of inductors (motors, electromagnets, etc.) anywhere where there are components that could be damaged by an inductive kick. For basic applications, nearly any type of diode will work for a flyback diode. However, diodes differ in their speed in reacting to voltage changes. In some circuits, the speed at which a diode reacts to changes in voltage can be important.

Review

In this chapter, we learned the following:

1. Inductors are electronic components that are made of winding wire around a core which allows them to store charge in their magnetic fields.
2. The strength of a magnetic field is the magnetic flux and is measured in Webers (Wb).
3. While an *electric* field is made from the *presence* of charges, the *magnetic* field is made from the *movement* of charges (i.e., the current).
4. When the current going into an inductor increases, the extra current is first used to build up the flux of the magnetic field before achieving a new steady state.
5. When the current going into an inductor decreases, the flux of the magnetic field is converted into a voltage, which increases the output current, resisting the change in current.
6. Inductors oppose changes in current. They act as resistors to AC and as short circuits to DC.
7. An inductor can be thought of similar to a flywheel in mechanics, which stores energy in a rotating wheel.
8. Inductors are useful because not only do they have interesting electric properties, their magnetic field also allows them to interface into the real world because magnetic field effects have a much longer reach than electric field effects.

9. When current to an inductor is suddenly shut off, it creates a large voltage, known as an inductive kick.
10. A snubber diode is used to mitigate against the negative effects of an inductive kick.

Apply What You Have Learned

1. If I want a circuit to block DC but allow AC, should I use an inductor or a capacitor?
2. If I want a circuit to allow DC but block AC, should I use an inductor or a capacitor?
3. Why are inductors used so much in systems that interact with the outside world?
4. Why does inductive kick happen?
5. Draw a schematic of an inductor where the positive side of the inductor has a switch and the negative side is connected to an LED and a resistor. Add in a snubber diode to protect the LED from the effect of turning off the switch.
6. What is the purpose of the snubber diode?
7. If I have an inductor that is 5 H and has a steady current going through it of 2 A, what is the size of the magnetic flux in the inductor's magnetic field?
8. If I have an inductor that is 7 μH and has a steady current going through it of 3 mA, what is the size of the magnetic flux in the inductor's magnetic field?

9. If I have a 4 H inductor with 0.3 Wb of flux in its magnetic field, how much current is flowing through it?
10. If an inductor's magnetic field has 5 Wb of flux and it decreases to 3 Wb of flux over 2 seconds, what is the average voltage produced over that timeframe?
11. If an inductor's magnetic field has 1 μWb of flux and it increases to 2 μWb of flux over 0.4 second, what is the average voltage produced over that timeframe?
12. If the current flowing through a 3 H inductor drops from 7 mA to 1 mA over a period of 0.01 second, what is the average voltage produced over that time period?

CHAPTER 21

Inductors and Capacitors in Circuits

In this chapter, we will look at a few of the basic uses of simple inductors.

21.1 RL Circuits and Time Constants

Just as we had an RC circuit for capacitors, there is a similar circuit for inductors—the RL circuit. Just as inductors are the alter ego of capacitors, RL circuits are the alter ego of RC circuits.

An RL circuit is a circuit consisting of a resistor (R) and an inductor (L) in series with each other. They are similar in construction and usage to the RC circuits we looked at in Chapter 17, but have some important differences.

In RC circuits, the RC time constant was the amount of time it took to charge a capacitor to 63.2% of full voltage when coupled through a resistor. That is, after 1 time constant has passed, the voltage between the legs of the capacitor will be 63.2% of supply voltage.

In RL circuits, the RL time constant is the amount of time it takes to charge the inductor's magnetic field 63.2% of its final value. Since the size of the magnetic field and the current are directly related to each other, this is also the amount of time it takes to get the current up to 63.2% of its final rate.

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_21

The way that the RL time constant is calculated *is different from* the way that RC time constants are calculated. The RL time constant is found by *dividing* the inductance in Henries (L) by the resistance in ohms (R).

$$\tau = \frac{L}{R} \tag{21.1}$$

Figure 21-1 shows the relationship between the number of time constants, the current through the inductor, and the voltage across the inductor. This is the same table as the one for RC time constants (Figure 17-1) except that current and voltage are swapped with each other.

# of Time Constants	% of Current	% of Voltage
0.5	39.3%	60.7%
0.7	50.3%	49.7%
1	63.2%	36.8%
2	86.5%	13.5%
3	95.0%	5.0%
4	98.2%	1.8%
5	99.3%	0.7%

Figure 21-1. *RL Time Constants*

Example 21.20 Let's say that we have a 2 H inductor in series with a 500 Ω resistor connected to a 5 V source. How long will it take before the inductor has 2.5 V across its terminals?

To solve this, recognize that 2.5 V is basically half of the voltage source. Therefore, we need to look at Figure 21-1 and find the one which is closest to having 50% of the voltage. That is 0.7 time constant.

Now, we need to figure out what the time constant for this circuit is. The inductance is 2 H, and the resistance is 500 Ω. Therefore, according to Equation 21.1, we divide the inductance by the resistance:

$$\begin{aligned}\tau &= \frac{2}{500} \\ &= 0.004\text{second}\end{aligned}$$

Therefore, the time constant is 0.004 second. We are wanting 0.7 time constants, so the final answer is 0.004 · 0.7 = 0.0028 seconds.

21.2 Inductors and Capacitors as Filters

As we have mentioned, a general way of thinking about capacitors and inductors is that capacitors allow AC but block DC and inductors are the other way around. Inductors allow DC but block AC.

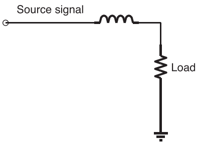

Figure 21-2. *Removing High Frequencies Using an Inductor*

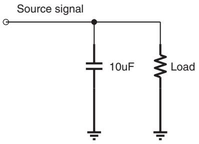

Figure 21-3. *Removing High Frequencies Using a Capacitor*

This can also be thought of in terms of frequency response. When dealing with signals (audio circuits, radio circuits, etc.), you oftentimes want to deal with certain frequencies and not others. With an inductor, the lower the frequency is, the easier it is to pass through from one side to the other. With a capacitor, the higher the frequency is, the easier it is to pass through from one side to the other.

Many audio systems have different speakers optimized for different frequencies. A common setup is to have two speakers—a woofer which handles the low pitches and a tweeter which handles the higher pitches. By using capacitors and inductors, circuit designers can customize which frequencies go to which speakers.

Radio systems also utilize capacitors and inductors. Each radio station operates on a specific carrier frequency. A **carrier frequency** is the dominant frequency used to carry a signal. When building a radio receiver, capacitors and inductors are used to isolate the specific frequency from all of the frequencies being transmitted over the air.

Chapter 22 will discuss the mathematics behind this further.

21.3 Parallel and Series Capacitors and Inductors

Capacitors and inductors can each often take on the role of the other in frequency filtering. As we have mentioned, you can use a capacitor to allow AC signals and block DC and low-frequency signals. Inductors do the opposite—they allow DC and low-frequency signals and block AC signals. However, in a pinch, you can actually get each one to do the job of the other.

Imagine that you want to get rid of noise in a circuit. Noise is essentially high-frequency AC. Using the previously defined rules, we might want to put an inductor in series with the circuit to remove noise. Figure 21-2 shows what this looks like. There is, however, another option.

Instead of putting an inductor in series with the circuit, we can instead wire a capacitor that goes to ground in parallel with the circuit. Figure 21-3 shows this configuration. The way to think about this is that, since AC signals travel through a capacitor, the capacitor is shunting off AC signals to ground before they get to the load.

Likewise, the same can be done for low-frequency filtering. Normally a capacitor is used to block low-frequency or DC signals when wired in series. However, an inductor can be used in parallel to pass off low frequencies to ground while letting high frequencies pass through.

In both of these situations, if the currents going to ground will be significant at all, you might also consider putting a small resistor in the parallel circuit as well. This will lessen the ability of the circuit to pass signals off to ground, but will also prevent short circuit behavior if you get significant signals being shunted off to ground.

Additionally, when you switch from the inline to the parallel version of these circuits, you wind up wasting the power that gets shunted off to ground. In the series versions of the circuit, any unused power is either blocked or stored. Therefore, it doesn't really waste much power. However, in the parallel versions, the filtered power is simply sent to ground, essentially wasting it.

Review

In this chapter, we learned the following:

1. A resistor-inductor series circuit is known as an RL circuit.

2. RL circuits have time constants very similar to RC circuits.

3. The time constant for an RL circuit is given by *dividing* the inductance by the resistance.

4. For RL circuits, the voltage and current values for each time constant are swapped.
5. Inductors and capacitors can be used to filter specific frequencies.
6. Capacitors allow high frequencies to go through, while inductors allow lower frequencies to go through.
7. Using capacitors and inductors together allows a person to define a specific range of frequencies that they wish to either block or allow.
8. Radios use this feature which allows people to filter only the specific radio station frequency they want.
9. Capacitors and inductors can be used for the other's job in filters by wiring them in parallel so that they carry their type of current (AC or DC) to ground instead of letting it pass to the load.

Apply What You Have Learned

1. What is the time constant of a series circuit consisting of a 50 Ω resistor and a 2 H inductor?
2. What is the time constant of a series circuit consisting of a 10 Ω resistor and a 5 μH inductor?
3. If I have a 9 V battery and I connect it to a series circuit consisting of a 1 kΩ resistor and a 23 μH inductor, how much time will it take before the current through the inductor reaches approximately 87% of its maximum value?

4. If I have a 5 V source and I connect it to a series circuit consisting of a 2 kΩ resistor and a 6 μH inductor, how much time will it take before the voltage across the inductor falls below 0.25 V?
5. If I have a circuit that has unwanted high-pitched noise, what component can I wire in series with the circuit to remove the noise?
6. If I have a circuit that has unwanted high-pitched noise, what component can I wire in parallel with the circuit to remove the noise?
7. What types of currents does an inductor (a) block and (b) allow?
8. What types of currents does a capacitor (a) block and (b) allow?
9. If I am building a radio, I need to allow through only very specific frequencies. What component or combination of components would I use to do this?

CHAPTER 22

Reactance and Impedance

22.1 Reactance

We have discussed resistance quite a bit in this book. Resistance is specifically about the ability of a component to be a good conductor of electricity. When a circuit encounters resistance, power is lost through the resistor.

However, another way of preventing current from flowing is known as *reactance*. With reactance, the power isn't dissipated, but rather the current is *prevented from flowing altogether*.

Let's think again about what happens when a voltage is connected to a capacitor in series. The capacitor starts to fill up. As the capacitor gets more and more full, there is less charge that can get onto the plate of the capacitor. This prevents the other side from filling up as well, and current cannot get through the capacitor. This acts *in a similar way to resistance*—it is preventing (impeding) the flow of current. However, it is not dissipating power because it is actually preventing the current from flowing. This is known as **reactance**. For capacitors, it is called **capacitative reactance,** and for inductors, it is called **inductive reactance**.

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_22

Reactance is usually frequency dependent. Again, going to the capacitor example, with high frequencies, the capacitor is continually charging and discharging, so it never really gets full, so it never impedes the current flow very much. Therefore, capacitors add very little reactance with high-frequency AC. The lower frequencies, however, give the capacitors time to get full, and when they are full, they impede the current flow. Therefore, for capacitors, lower frequencies create more reactance.

Reactance is measured in ohms, just like resistance. However, they cannot be simply added to resistances, so they are usually prefixed with the letter *j*. So 50 ohms of reactance is usually labeled as *j*50 Ω so that it is understood as a reactance. Reactances can be added to each other, and resistances can be added to each other. We will see how to combine them in Section 22.2, "Impedance." Reactances are usually denoted using the letter *X*.

The reactance of a capacitor (X_C) is given by the following formula:

$$X_C = -\frac{1}{2\pi \cdot f \cdot C} \qquad (22.1)$$

In this formula, *f* is the AC frequency of the signal (in hertz), and *C* is the capacitance. It might seem strange that this produces a negative value. The reason for this will make sense as we go forward. However, it is *not* producing negative impedance—we will see this when we combine reactance with resistance.

> **Example 22.21** What is the reactance of a 50 nF capacitor to a signal of 200 Hz? To find this, we merely use the formula:

$$
\begin{aligned}
X_C &= -\frac{1}{2\pi \cdot f \cdot C} \\
&= \frac{1}{2\pi \cdot 200 \cdot 0.00000005} \\
&\approx \frac{1}{2 \cdot 3.14 \cdot 200 \cdot 0.00000005} \\
&= \frac{1}{0.00000624} \\
&\approx -j160256\Omega
\end{aligned}
$$

You can see from this formula why it is said that a capacitor blocks DC. DC is, essentially, current that does not oscillate. In other words, the frequency is zero. Therefore, the formula will reduce to $\frac{1}{0}$, which is infinite. Therefore, it has infinite reactance against DC.

Also note what happens as the frequency increases. As the frequency increases, the denominator gets larger and larger. That means that the reactance is getting smaller and smaller—closer and closer to zero. As the frequency goes up, the reactance is essentially heading toward zero, but will never get there because the frequency can't be infinite.

The formula for **inductive reactance** (X_L) is very similar:

$$X_L = 2\pi \cdot f \cdot L \qquad (22.2)$$

In this equation, f is the signal frequency, and L is the inductance of the inductor in Henries.

> **Example 22.22** Let's say that I have an 3 H inductor with a 50 Hz AC signal. How much reactance does the inductor have in this circuit?

$$
\begin{aligned}
X_L &= 2\pi \cdot f \cdot L \\
&= 2 \cdot 3.14 \cdot 50 \cdot 3 \\
&= j9420\Omega
\end{aligned}
$$

The reactance in this circuit is $j9{,}420\ \Omega$.

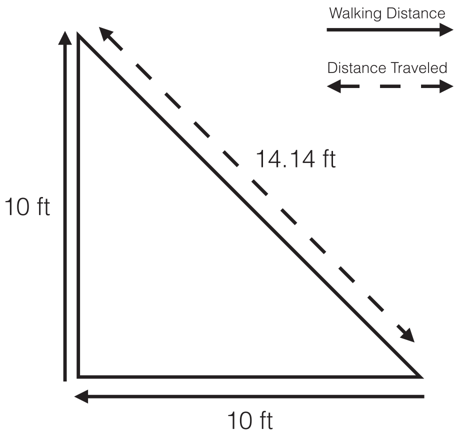

Figure 22-1. *Total Distance Traveled vs. Total Displacement*

22.2 Impedance

In fact, resistance and reactance are usually combined together in a circuit to get a quantity known as **impedance**, which is simply the combination of resistive and reactive quantities. Impedance is often designated using the letter Z.

Resistance and reactance aren't added together directly; instead, you can think of them acting at angles to each other. Let's say that I start at my house and walk 10 feet out my front door. Then, I turn 90 degrees right and walk another 10 feet. While I have walked 20 feet, I am *not* 20 feet from my door. I am, instead, a little over 14 feet from my door. I can call this my displacement. Figure 22-1 shows what this looks like visually.

Since this is a right triangle, the distance from the start to the end is found on the hypotenuse of this triangle. We can calculate the total displacement using the Pythagorean theorem ($A^2 + B^2 = C^2$). If we solve for C (total displacement), we get

$$C = \sqrt{A^2 + B^2} \tag{22.3}$$

So, in our distance example, if I went forward 10 feet, turned left, and went another 10 feet, the total distance traveled would be

$$\begin{aligned} C &= \sqrt{A^2 + B^2} \\ &= \sqrt{10^2 + 10^2} \\ &= \sqrt{100 + 100} \\ &= \sqrt{200} \\ &\approx 14.14 \end{aligned}$$

The total impedance is like the total distance from your door. Resistance and reactance are like different walking directions (at right angles to each other), and impedance is the total displacement. That's why we use the letter *j* to signify reactance—it is just like resistance but in a different direction.

So how do we calculate impedance? In fact, it is calculated *precisely like* the displacement calculation in Equation 22.3:

$$\text{impedance} = \sqrt{\text{resistance}^2 + \text{reactance}^2} \tag{22.4}$$

Or, using their common abbreviations, we can say:

$$Z = \sqrt{R^2 + X^2} \tag{22.5}$$

Let's see how we can use Equation 22.4 to calculate total impedance. If I have a circuit that has 30 Ω of resistance and *j*20 Ω of reactance, then the formula for total impedance is

$$\begin{aligned}\text{impedance} &= \sqrt{\text{resistance}^2 + \text{reactance}^2}\\ &= \sqrt{30^2 + 20^2}\\ &= \sqrt{900+400}\\ &= \sqrt{1300}\\ &\approx 36.1\Omega\end{aligned}$$

As you can see, when using this formula, we are calculating a total impedance, so the *j* drops away.

> **Example 22.23** If I have a 1 kΩ resistor in series with a 100 nF capacitor with a 800 Hz signal, what is the total impedance to the signal that my circuit is giving?
>
> To find out impedance, we need both resistance and reactance. We already have resistance—1 kΩ. The reactance is found by using Equation 22.1:

$$\begin{aligned}X_C &= -\frac{1}{2\pi\cdot f\cdot C}\\ &= \frac{1}{2\pi\cdot 800\cdot 0.0000001}\\ &\approx \frac{1}{0.0005024}\\ &\approx -j1990\Omega\end{aligned}$$

So the reactance is about $-j1990\ \Omega$.

So, if the resistance is 1000 Ω and the reactance is $-j1990\ \Omega$, what is the impedance? The impedance is found by using Equation 22.4:

$$\begin{aligned}\text{impedance} &= \sqrt{\text{resistance}^2 + \text{reactance}^2}\\ &= \sqrt{1000^2 + (-1990)^2}\\ &= \sqrt{1000000 + 3960100}\\ &= \sqrt{4960100}\\ &\approx 2227\Omega\end{aligned}$$

22.3 RLC Circuits

So far we have discussed RC (resistor-capacitor) and RL (resistor-inductor) circuits. When you combine all of these components together, you get an RLC (resistor-inductor-capacitor) circuit.

When you calculate the impedance of such circuits, you have to be sure you include the reactance of *both* the capacitative and the inductive components. While inductors and capacitors both offer reactance to certain frequencies, their reactances actually oppose each other. That is, the reactance of one cancels out the reactance of the other. This is why the capacitative reactance is negative and the inductive reactance is positive.

Therefore, when calculating reactances that include *both* inductance and capacitance, you can add the reactances just like you would add resistances. However, since the capacitive reactances are negative and the inductive reactances are positive, they wind up canceling each other out to some degree.

Example 22.24 If I have an inductor of 5 mH and a capacitor of 5 μF in series with a 200 Ω resistor, what is the impedance of the circuit for a frequency of 320 Hz?

To solve this problem, we need to first find the capacitative reactance (X_C) and the inductive reactance(X_L). To get the capacitative reactance, we use Equation 22.1:

$$\begin{aligned} X_C &= -\frac{1}{2\pi \cdot f \cdot C} \\ &= \frac{1}{2\pi \cdot 320 \cdot 0.000005} \\ &\approx \frac{1}{0.01} \\ &\approx -j1990\Omega \end{aligned}$$

The inductive reactance is found using Equation 22.2:

$$\begin{aligned} X_L &= 2\pi \cdot f \cdot L \\ &= 2\pi \cdot 320 \cdot 0.005 \\ &\approx j10\Omega \end{aligned}$$

Now, we can just add these reactances together:

$$\begin{aligned} X_{total} &= X_L - X_C \\ &= j10\Omega + -j100\Omega \\ &= -j90\Omega \end{aligned}$$

The fact that this is negative is not a problem because it will be squared (which will get rid of the negative) in the next step. Now that we know the resistance (200 Ω) and the reactance (– *j*90 Ω), we just need to use Equation 22.4 to calculate the total impedance:

$$\begin{aligned} \text{impedance} &= \sqrt{\text{resistance}^2 + \text{reactance}^2} \\ &= \sqrt{200^2 + (-90)^2} \\ &= \sqrt{40000 + 8100} \\ &= \sqrt{48100} \\ &\approx 219\Omega \end{aligned}$$

So the total impedance (opposition to current) in this circuit is 219 Ω.

22.4 Ohm's Law for AC Circuits

In an AC circuit, the current and voltage are continually varying. Therefore, using the traditional Ohm's law, you would have to calculate Ohm's law over and over again in order to find out the relationships between voltage, current, and resistance.

However, there is a form of Ohm's law that works directly on AC circuits of a given frequency. That is, it is essentially a *summary* of the voltages and currents that happen on each cycle. Ohm's law for AC circuits is basically identical to the previous Ohm's law, but the terms are slightly different:

$$V_{\mathrm{RMS}} = I_{RMS} \cdot Z \qquad (22.6)$$

Here, the voltage we are referring to (V_{RMS}) is an *average* voltage through one cycle of AC. This average is known as the **RMS** average. It is a little different than the typical average you might think of. If an AC voltage is swinging back and forth from positive to negative, the actual average voltage is probably around zero. However, RMS voltage is about calculating the average amount of push in any direction—positive or negative. Therefore, the RMS voltage will always yield a positive answer.[1]

Likewise, the current refers to the RMS current (I_{RMS}). Just like the RMS voltage, the RMS current will always be positive, because it is the measure of the average amount of flow in *any* direction.

Finally, the impedance Z is calculated as we have noted in this chapter—by combining resistances and reactances together into an impedance.

---

[1] RMS stands for "root mean square." It is obtained by (a) squaring every data point, (b) averaging the squares, and then (c) taking the square root of the average. This is why it will be positive—it deals in squares.

Thus, Ohm's law for AC circuits can be used to express summary relationships about average voltage, average current, and impedance in an AC signal.

> **Example 22.25** I have an AC circuit whose RMS voltage is 10 V. I have calculated the impedance of this circuit to be 20 Ω. What is the RMS current of this circuit?
>
> To find this out, we just rearrange Ohm's law a little bit:

$$V_{\text{RMS}} = I_{RMS} \cdot Z$$
$$I_{\text{RMS}} = \frac{V_{\text{RMS}}}{Z}$$

Now I just use the values given to fill in the blanks:

$$\begin{aligned} I_{RMS} &= \frac{V_{RMS}}{Z} \\ &= \frac{10}{20} = 0.5 \text{ A} \end{aligned}$$

In this circuit, we would have an average of half an amp (500 milliamps) flowing through the circuit.

> **Example 22.26** I have an AC voltage source with an RMS voltage of 5 V running at 200 Hz. It is connected in series with a 50 kΩ resistor and a 50 nF capacitor. What is the current in this circuit?
>
> To find this out, we have to first find the total impedance of the circuit. That means we have to use Equation 22.1 to find the reactance of the capacitor:

$$
\begin{aligned}
X_C &= -\frac{1}{2\pi \cdot f \cdot C} \\
&= \frac{1}{2 \cdot 3.14 \cdot 200 \cdot 0.000005} \\
&= \frac{1}{0.00000628} \\
&\approx -j159236\Omega
\end{aligned}
$$

Now that we have the resistance (50 kΩ) and the reactance ($-j159236\,\Omega$), we can use them in Equation 22.5:

$$
\begin{aligned}
Z &= \sqrt{R^2 + X^2} \\
&= \sqrt{50000^2 + 159236^2} \\
&= \sqrt{2500000000 + 25356103696} \\
&= \sqrt{27856103696} \\
&\approx 166901\Omega
\end{aligned}
$$

Now we can use Ohm's law for AC circuits (Equation 22.6 to find the current):

$$
\begin{aligned}
I_{\text{RMS}} &= \frac{V_{\text{RMS}}}{Z} \\
&= \frac{5}{166901} \\
&\approx 0.00003\text{A}
\end{aligned}
$$

This means that the average current (I_{RMS}) will be approximately 0.00003 A, or 30 μA.

22.5 Resonant Frequencies of RLC Circuits

As we have seen, the capacitative reactance goes closer to zero when the frequency goes up. Likewise, the inductive reactance increases when the frequency goes up. Additionally, the capacitative reactance and the inductive reactance have opposite signs—negative for capacitative reactance and positive for inductive reactance.

What is interesting is that if you have a combination of inductors and capacitors, there is always some frequency at which their reactances exactly cancel each other out. This point is known as the **resonant frequency** of the circuit.

When a capacitor and an inductor are in series with each other, it is termed an LC series circuit. Because the inductor inhibits high frequencies and the capacitor inhibits low frequencies, LC circuits can be used to let through a very specific frequency range. The center of this range is known as the **resonant frequency** of the circuit.

Now, if you are good with algebra, you can combine Equation 22.1 and Equation 22.2 to figure out the resonant frequency of a circuit (i.e., set them to add up to zero and then solve for the frequency f). However, to spare you the trouble, there is a formula that you can use to find the resonant frequency of a circuit:

$$f = \frac{1}{2\pi\sqrt{L \cdot C}} \tag{22.7}$$

At this frequency, there is no total reactance to the circuit—the only impedance comes from the resistance.

> **Example 22.27** Let's say that you have a 20 μF capacitor in series with a 10 mH inductor. What is the resonant frequency of this circuit?
>
> To find the resonant frequency, we only need to employ Equation 22.7:

$$
\begin{aligned}
f &= \frac{1}{2\pi\sqrt{L\cdot C}}\\
&= \frac{1}{2\pi\sqrt{0.01\cdot 0.00002}}\\
&= \frac{1}{2\pi\sqrt{0.0000002}}\\
&\approx \frac{1}{2\pi\cdot 0.000447}\\
&\approx \frac{1}{0.00279}\\
&\approx 358\text{Hz}
\end{aligned}
$$

> Therefore, this circuit has a resonant frequency of 358 Hz. This means that, at this frequency, this circuit has no reactive impedance.

Resonance frequencies are important in signal processing. They can be used in audio equipment to boost the sound of a specific frequency (since all other frequencies will have resistance). They can be used to select radio stations in radio equipment (since it will be the only frequency allowed through without resistance). You can also remove a specific frequency by taking a resonant frequency circuit to ground, thereby having a specific frequency short-circuited to ground with no resistance.

22.6 Low-Pass Filters

As mentioned, capacitors allow AC and block DC. If you wanted to remove the AC component of a signal, a capacitor can be connected to take the AC portion of your signal to ground. This is known as a low-pass filter, because, since the high-frequency (AC) component of the signal is being sent to ground, the low-frequency (DC) component of the signal is being allowed to pass.

How do you know what size of capacitor to use? Well, a rule of thumb is to have a capacitor whose reactance to the signal that you want to *keep* is greater than the resistance coming in to the capacitor (we want a larger resistance because we don't want that signal coming through the capacitor).

Therefore, we can modify Equation 22.1 to say

$$f = \frac{1}{2\pi \cdot R \cdot C} \tag{22.8}$$

or

$$C = \frac{1}{2\pi \cdot R \cdot f} \tag{22.9}$$

If you know the incoming resistance, you can use that last equation to just solve for the size of capacitor you will need (if you don't have any incoming resistance, then you can simply add a resistor of your choosing and calculate for C).

22.7 Converting a PWM Signal into a Voltage

In Chapter 15, we talked about how the Arduino microcontroller can't emit an analog voltage, but can mimic it with a PWM signal, where it turns its 5 V signal on and off to mimic in-between voltages. We can use a low-pass filter to transform a PWM signal into a voltage.

The circuit to do so is fairly straightforward and is given in Figure 22-2. The resistor and capacitor serve to tune the lower end of the frequency of the signal that gets removed.

According to the specs, the PWM frequency is about 490 Hz. In theory, we could find the reactance of the capacitor using Equation 22.1. However, in practice, because the PWM is a square wave (it has very sharp edges), it requires a larger capacitor. In any case, this filter removes the AC component of the PWM and just leaves the residual voltage that it represents.

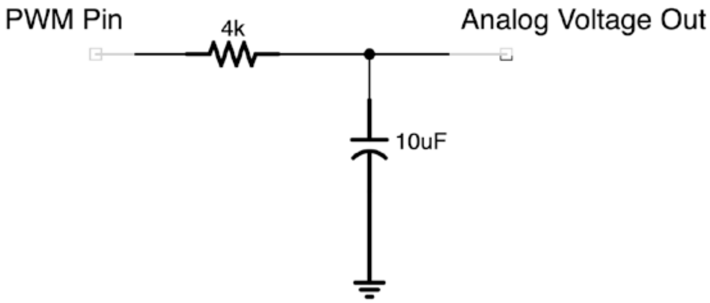

Figure 22-2. *Converting PWM to Analog with a Capacitor*

Review

In this chapter, we learned the following:

1. Reactance (X) is a property of some electronic components that is similar to resistance, but it *prevents* the flow of current instead of *dissipating* the flow (i.e., converting it to heat).
2. Reactance, like resistance, is measured in ohms. A j is placed in front of the reactance to specify that it is a reactance value.
3. Reactance is frequency dependent—the amount of reactance depends on the frequency of the signal.
4. Capacitors and inductors each have formulas that can be used to calculate the reactance of the components.

5. Capacitors yield negative reactance and inductors yield positive reactance. This means that their reactances will oppose and cancel each other out to some degree.
6. Impedance (Z) is the total inhibition of the flow of current, combining both resistive and reactive elements.
7. Reactance and resistance are combined into impedance in the same way that walking two different directions can be combined into a total distance from your originating point—using the Pythagorean theorem.
8. RMS voltage (V_{RMS}) is the average voltage of an AC circuit, regardless of the direction (positive or negative) of the voltage. This can be used to summarize the effects of an AC voltage.
9. RMS current (I_{RMS}) is the average current of an AC circuit, regardless of the direction (positive or negative) of the voltage. This can be used to summarize the effects of an AC current.
10. Ohm's law for AC circuits yields the summary relationship between RMS voltage, RMS current, and impedance in an AC circuit. It is identical to the previous Ohm's law, but uses the summary values for the circuit at a particular frequency, rather than the values at a particular point in time.
11. The resonant frequency of a circuit is the frequency at which inductive reactance and capacitative reactance cancel each other out.

12. Resonant frequencies can be used in any application where isolating a frequency is important, because the resonant frequency will be the only frequency not encountering resistance.
13. Low-pass filters remove AC signals and allow lower-frequency signals to pass.
14. A low-pass filter can be used to convert an Arduino PWM signal into a voltage.

Exercises

1. As the frequency of a signal goes up, how does that affect the reactance from a capacitor? What about with an inductor?
2. As the frequency of a signal goes down, how does that affect the reactance from a capacitor? What about with an inductor?
3. What is true about the relationship between the capacitative reactance and the inductive reactance at the resonant frequency?
4. Why is power not used up with reactance?
5. How are reactance and resistance combined to yield impedance?
6. Calculate the capacitative reactance of a 3 F capacitor at 5 Hz.
7. Calculate the capacitative reactance of a 20 μF capacitor at 200 Hz.

8. Calculate the inductive reactance of a 7 H inductor at 10 Hz.
9. Calculate the inductive reactance of a 8 mH inductor at 152 Hz.
10. Calculate the impedance of a circuit with a 200 Ω resistor in series with a 75 μF capacitor with a signal of 345 Hz.
11. Calculate the impedance of a circuit with a 310 Ω resistor in series with a 90 nF capacitor with a signal of 800 Hz.
12. Calculate the impedance of a circuit with no resistor and a 60 mH inductor with a signal of 89 Hz.
13. Calculate the impedance of a circuit with a 50 Ω resistor in series with a 75 μH inductor with a signal of 255 Hz.
14. If I have an AC circuit with an RMS voltage of 6 V and an impedance of 1 kΩ, what is the average (RMS) current of this circuit?
15. If I have an AC circuit and I measure the AC voltage as 10 V RMS and I measure the AC at 2 mA RMS, what is the impedance of this circuit?
16. If I have an 80 Hz AC circuit that has an 8 V RMS voltage source in series with a 500 Ω resistor, a 5 H inductor, and a 200 nF capacitor, what is the RMS current flowing in this circuit?
17. Calculate the impedance of a circuit with a 250 Ω resistor in series with a 87 μH inductor and a 104 μF capacitor with a signal of 745 Hz.

18. What is the resonant frequency of the circuit in the previous question?

19. What is the reactance of a circuit at its resonant frequency?

20. If I have a 10 μF capacitor, what size inductor do I need to have a resonant frequency of 250 Hz?

CHAPTER 23

DC Motors

A **DC motor** is a device that converts DC electrical power into mechanical power. It operates by rotating a shaft using electromagnetism. DC motors are fairly simple to use, though they require slightly different reasoning from the way we have been examining circuits so far.

23.1 Theory of Operation

While there are more than one kind of motor, the "typical" DC motor is known as a **brushed electric motor**. This motor has a fixed outside container (called a **stator** because it is stationary) that contains magnets with opposite poles. Within the stator is an armature with electromagnets that is connected to the **shaft** (the shaft is the thing that turns). The electromagnets on the shaft's armature are then energized to be the same pole of the current magnetic orientation.

If you have ever tried to push two magnets together that have the same pole, you know that the magnetic fields exhibit a lot of physical force to try to prevent this from happening. Therefore, the fixed magnets' magnetic fields push on the temporarily created electromagnets' fields on the shaft's armature. When the shaft turns to align itself with the fixed magnets, it also turns the connector that connects the shaft to the electricity. When it is turned, the connectors on the shaft are now connected to the *opposite* polarity than they were, so their fields are no longer aligned again! Thus, the shaft continues rotating on.

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_23

A certain type of motor, called a **brushless motor**, has a similar operation, except that instead of having the rotating shaft change the polarity of the current, it uses fixed magnets on the shaft and uses electromagnets for fixed magnets. Then, by pushing AC through the electromagnets, it changes the polarity of the fixed magnets back and forth, moving the shaft through the magnetic field. Brushless motors tend to be more expensive, but can rotate faster and have fewer parts which can wear out.

Even though brushless motors use AC internally, they are operated just like DC motors from the outside. Brushless motors have internal circuitry that converts the DC into an AC waveform for energizing the electromagnets appropriately.

23.2 Important Facts About Motors

Motors require a lot of current, especially to start up. Starting up, they require even more current (often ten times their normal current). This is also the amount of current they require if they ever stop rotating (known as stalling).

Motors are rated according to their most efficient operating voltage. If a motor is rated at 5 V, then it's not that you can't operate it at a different voltage, but that operating it at different voltages will change some of its characteristics and probably won't be as efficient.

Motors generate a negative voltage called *back EMF* that is in opposition to the voltage provided to them. Motors generate more back EMF the faster that they spin. At the point at which the back EMF is equal to the voltage provided, the motor will no longer spin faster. There is also some voltage drop by the internal resistance of the motor's coils, but it is relatively minor compared to the back EMF drop.

If there is nothing attached to the shaft, the amount of current the motor operates with is known as the **no load current**, and the speed at which the motor spins is known as the **no load RPM** (**RPM** means "revolutions per minute").

However, the most difficult thing that a motor does is to *start* rotating. When the motor isn't moving and is exhibiting its maximum rotational power (known as **torque**) to try to move, this is known as a **stall** condition. This is the most strenuous thing a motor can do, and many motors can't survive more than a few seconds in stall conditions. The amount of torque that a motor supplies in stall conditions is known as the **stall torque**, and the amount of current that the motor uses in these conditions is known as the **stall current**. Oftentimes, if you think you have a circuit that is properly sized for a motor, but it won't start spinning, this is because it isn't getting enough current for stall conditions. You can sometimes even check this by spinning the shaft by hand and seeing if the motor then springs to life.

Note that when a motor first turns on, it is always in stall conditions.

23.3 Using a Motor in a Circuit

DC motors are nice because they basically self-regulate. They spin fast enough to match whatever voltage is given to them and draw enough current to manage the proper torque for the load. If you want to know the details and calculations behind these things, see Appendix D, Section D.3, "Motor Calculations." However, if you stick with the voltage listed in the specs, the RPM of the motor generally gets fairly close to the specifications. In general, lower voltages lead to lower RPMs, and higher voltages lead to higher RPMs.

It is pretty rare to add series resistance to a motor circuit. There are several reasons for this. The first is that, as mentioned, motors basically self-regulate. This means that

1. They will draw the amount of power they need to draw for the given torque and no more. Thus, a current-limiting resistor is not needed. Additionally, if you calculate the current-limiting resistor for steady-state conditions, it will provide too little current for stall conditions.

2. They will utilize the given voltage to achieve the RPM that balances it out. Therefore, there will not be any excess voltage to consume.

Additionally, series resistance can actually cause problems. The amount of current a motor can draw can vary very widely. As mentioned, the stall current can sometimes be ten times the steady-state current. With series resistance, the voltage drop across the resistor will then vary up to ten times as well, which will affect the voltage available to the motor and thus alter the performance specifications of the motor. Under stall conditions, it will decrease the performance of the motor at precisely the wrong moment.

Therefore, while you might want to down-step voltage for a motor to get the proper RPM, you want to do so in an essentially non-resistive way, using diodes, transistors, or other components that are not likely to add significant series resistance to the motor.

However, you may want to prevent problematic conditions as well. You don't want a short circuit in your motor to fry your whole circuit. Because of this, you may want to add a fuse (instead of a current-limiting resistor) that will trip if it experiences excessively high current to protect the rest of your circuit.

Additionally, since motors contain electromagnets, they utilize induction. As mentioned in Chapter 20, when current to an inductor stops, it creates a voltage spike due to the collapsing of the electromagnetic fields. Therefore, we have to provide a snubber diode to prevent that current from blowing out our sensitive equipment. You can refer back to Figure 20-2 to see how this is wired.

Because motors are operated through induction, they are often represented with the same symbol as an inductor.

23.4 Attaching Things to Motors

Motors just spin a shaft. What happens after the shaft starts spinning is, technically, outside the scope of this book and entirely up to you. Nonetheless, I thought I would take just a moment to talk about attaching things to motors.

The first thing that comes up is, *how* do you attach something to a motor's shaft? There are several common options:

1. Some items are built to have a **mechanical fit** onto a shaft. That is, you can push them onto the shaft with some force, but they are unlikely to just fall off on their own.
2. Some items are attached with a **set screw**. This is similar to the previous method, except that what keeps the attachment in place is not the pure fit, but an extra screw that applies the force after your attachment is in position. This allows for a little wider range of variance in size, so if the sizes are off slightly, the set screw will make up for the lack of mechanical force.
3. Finally, some items need to be glued, welded, or in some other way permanently affixed to the shaft. The problem with this method is that, once attached, you can't really use your motor for anything else.

So what are we attaching to our motor?

Well, you *can* directly attach things such as wheels to the shaft of the motor. For free-spinning things such as propellers, this is exactly what you do. It works for wheels too, kind of. The problem is that small DC motors often can't generate the kind of torque you need for typical usages. Therefore, you need gears to translate the low torque/high RPM of the motor into a high-torque/low-RPM output to the wheels.

Many places sell gearboxes which handle all of this for you. In fact, you can even buy motors with the gearboxes inside the motor packages as well. Then, you connect your gearbox to the axle of the wheels using a gear or a belt, and the wheel is what performs the final actions.

I find it interesting that all of these are just a series of conversions. We first convert electrical power into mechanical power. We then convert low-torque/high-RPM power into high-torque/low-RPM power. This is actually very similar to the operation of a transformer in electronics, which converts low current/high voltage into high current/low voltage and vice versa. Then, the wheels convert rotational power into linear power, by driving your project along the ground or driving a belt around and so on. We have a source of power (the battery) and the thing we want to do (moving things), and what we have to figure out is the best way to convert the kind of power we have into the kind of power that we need.

23.5 Bidirectional Motors

Oftentimes we want motors to spin both forward and backward. Most motors can spin backward just by reversing the positive and negative leads to the motor. However, we need to be able to do that electrically instead of forcing someone to go and reverse wires on our circuit board whenever they want to back up instead of go forward.

Creating a circuit that can have current go either direction is hard enough, but added to that is the problem of positioning the snubber diodes—they will be needed in *both* directions! However, if we just attached snubber diodes in each direction, we would wind up with a short circuit no matter what (the current would prefer to just go through the diode pointing our way rather than through the motor).

A type of circuit to manage these considerations has been developed, known as an **H-Bridge**. We will not delve into the details of how H-Bridges work or how to use one, but I thought I should mention them.

If you are interested in H-Bridges, a popular chip is the L293.

23.6 Servo Motors

A **servo motor** (usually just called a servo) is a motor that, instead of spinning, creates an angle. Think about a remote-controlled car. To turn the car, the wheels have to be at an angle. A regular DC motor is what spins the wheels, but a servo motor is what causes the wheels to turn to a specific angle.

Servos are operated by sending an oscillating signal, much like the output of a 555 timer (see Chapter 18). Servos expect a signal every 20 milliseconds, and the length of the signal will determine the angle of the servo:

- If the signal is 1.5 milliseconds, the servo will be in the middle (neutral) position.
- If the signal is 1.0 millisecond, the servo will rotate 90 degrees counterclockwise.
- If the signal is 2.0 milliseconds, the servo will rotate 90 degrees clockwise.
- Between these positions, the servo will rotate in a proportional way.

As you can see, the servo can swing through 180 degrees of rotation. The servo expects to continually receive signals while it is on in order to maintain its position, but that position can be changed at any time with a change in the length of the signal. The behavior of a servo when no signal is applied depends on the particular servo you are using.

Servo motors have separate inputs for their voltage supply and their signal. The signal does not have to power the motor—it is powered through its supply voltage. Instead, the servo motor just needs to receive a small signal to know its rotation, and the actual power is supplied through its voltage input.

23.7 Stepper Motors

Stepper motors are motors that rotate in very precise increments. If you need something that can freely rotate, but need an *exact* number of revolutions or even half-revolutions, quarter-revolutions, and so on, a stepper motor supplies this capability. Stepper motors are often used in 3D printers, where the print head needs to move a precise distance. The integrity of the generated structure depends on the print head being in the exact correct location. Therefore, the rotation of the shaft must be precisely controlled.

The control of stepper motors is far outside the scope of this book and usually requires additional hardware (known as a stepper motor driver circuit), but I wanted to be sure you are aware of what they are and why they are important.

Review

In this chapter, we learned the following:

1. A DC motor converts electrical power into mechanical rotational power by either changing the magnetization of the shaft (in brushed motors) or of the fixed magnets (in brushless motors).
2. Rotational mechanical power is known as torque.
3. Motors tend to require a lot of current, especially at startup.
4. The rated voltage of a motor is related to the motor's RPM—decreasing the voltage decreases the RPM and increasing the voltage increases the RPM. This is due to the negative voltage (back EMF) generated by the movement of the motor.

5. The current flowing through a motor is related to the torque needed to spin the motor—more required torque draws more current.
6. When a motor is not spinning (either because it is just starting or because its load is too high), it is in a stall condition.
7. Stall current is the amount of current the motor draws in stall conditions (usually about ten times more than in a steady state).
8. Stall torque is the amount of torque generated in stall conditions.
9. Loading a motor so that it stays in stall conditions for any significant amount of time can damage the motor.
10. Series resistance is almost never added to a motor circuit.
11. Snubber diodes are used to mitigate against inductive spikes that happen when the electromagnetic field of the motor shuts off.
12. Gears can convert the low torque/high RPM of a motor into a high-torque/low-RPM output.
13. Motors can be used bidirectionally using an H-Bridge.
14. A servo motor is a motor that, instead of spinning, creates a fixed angle based on the length of a signal that is sent to it every 20 milliseconds.
15. A stepper motor is a motor that provides precise control over the amount of rotation of the shaft.

Apply What You Have Learned

1. If you wanted to prevent too much current from going through a motor, what should you use for this?
2. What happens to the RPM of a motor when I reduce the voltage to my motor?
3. What should be attached to a motor to protect the other parts of the circuit when a motor is shut off?
4. List one reason why a series resistor should not be used with a motor.
5. If I was building a robot and wanted to control the angle of a robot's arm, what type of motor could I use?
6. If I was building a machine to place components into a precise position onto a board, what type of motor should I pick to achieve that?

PART IV

Amplification Circuits

CHAPTER 24

Amplifying Power with Transistors

Amplification is the conversion of a low-power signal to a higher-power signal. Normally when we think of amplification, we think of sound amplifiers for musical instruments. Indeed those are amplifiers, and we will build a sound amplifier later in this book. However, *anytime* you convert a low-power input to a higher-power output, you have amplified the signal, whether that was a DC signal or an AC signal. In this chapter, we will focus on amplifying DC signals.

Many devices have limits to the amount of power they can output or even the amount of power we want them to output. Microcontrollers, for instance, are generally very sensitive with the amount of power they can source. The ATmega328/P (used in the Arduino Uno), for instance, has a maximum current rating of 40 mA, which is actually quite generous for a microcontroller. Other types of microcontrollers, such as many PIC microcontrollers, are rated for current outputs of 25 mA or less. Additionally, the voltage is more or less fixed at the operating voltage of the chip, which is usually 3–5 V.

However, numerous applications require more output (whether current, voltage, or power) than these can provide. A typical toy DC motor, for instance, needs about 250 mA for operation. So, if you want to control the motor with a microcontroller, then you have to have some way to convert your small output current into a larger output current to drive your motor.

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_24

Another scenario to consider is a situation where we want to control a high-power motor or another device from a button or switch. Sometimes it can be problematic to have a switch for a device to have high power run through it. It could be dangerous, or it could cost money. Let's say that we had a switch that we wanted to control a high-power device that was 1,000 feet away. If we had the full power running through the switch, that means that we have to run larger cables and will have to account for the large resistance in the wires. Running power across that distance will result in power loss because of the resistance in the wires, which will increase the cost for running the unit! However, if we instead use a low-power circuit to run the switch, we don't get nearly as much power loss. We just need to *amplify* the power output of the switch to control our device once the signal reaches it. The amount of amplification that occurs is known as the **gain** of an amplifier.

24.1 An Amplification Parable

As we noted in Chapter 10, physics tells us that we can't actually increase the power of something. What we can do, though, is use a smaller power signal to control a larger power source, and that is what we refer to as amplification.

Let's say that we have a dam on a river. The river wants to go downstream, but it is blocked by the dam. Let's say that we have a giant named Andre, which is able to lift the dam using his strength. If Andre lifts the dam a little bit, a little bit of water flows. If Andre lifts the dam a lot, all of the water can flow. The power of the water itself is actually more than Andre's power. Therefore, Andre can "amplify" his power by raising and lowering the dam.

Let's say that at the end of the riverbed was a structure that Andre wanted to destroy, but Andre himself isn't strong enough to do it. However, the river is strong enough to do it. Therefore, rather than try to destroy the structure on his own, Andre decides to raise the dam and let the river's power do it.

Note that Andre didn't actually increase the amount of power in the river. Instead, he (a being with lesser power) *controlled* the operation of the river (a higher-powered entity) by adjusting the dam (the physical resistance against the water). Thus, he *amplified* the effects of his actions by controlling the resistance on the higher-powered current.

24.2 Amplifying with Transistors

A very common method of amplifying power output in projects is with **transistors**. The term "transistor" is short for "transconductance varistor," which means that it is an electrically controlled variable resistor. In other words, it helps you amplify a signal in your project the same way that the dam allowed Andre to amplify his power. The transistor operates as a controllable electric dam, allowing a smaller-powered electric signal to control a higher-powered electric signal by adjusting resistance.

Another way you can think of it is like an outdoor faucet—the water going through the faucet is controlled by the wheel knob on the top, which provides a variable amount of resistance to the flow. The faucet can be fully on, fully off, or somewhere in between, all based on where the knob is set. The knob setting is like the input current—a specific level of input current will control the amount of flow that comes out.

There are many different types of transistors, with a wide variety of ways in which they work. What they all have in common is that they have three (sometimes four) terminals, and one of the terminals acts as a control valve for the flow of electricity between the other two.

The main way that the types of transistors differ is in whether the knob on the faucet is controlled by voltage or by current. The transistors operated by current are known as **bipolar junction transistors** (BJTs), and the ones operated by voltage are known as **field effect transistors** (FETs).

As a gross overgeneralization, FETs are most useful for lower-power switching applications, while BJTs are most useful for power amplification. This chapter will cover BJTs.

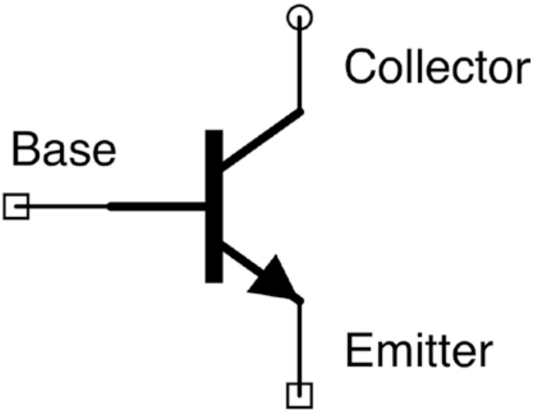

Figure 24-1. *The Schematic Symbol for a Transistor*

BJTs come in two basic forms based on whether the knob is normally closed but turned on by positive current (an NPN transistor) or whether the knob is normally open but closed off by positive current (a PNP transistor). Here we will focus on NPN transistors.

24.3 Parts of the BJT

A BJT has three terminals—the collector, the base, and the emitter. In BJT NPN transistors, a small positive current comes in at the base (which you can think of as the knob of a faucet), and it controls the current moving from the collector to the emitter. The more current that comes through the base, the more current is allowed to flow between the collector and the emitter.

Figure 24-1 shows what an NPN BJT looks like in a schematic. The collector is at the top right of this diagram. The emitter is the line with the arrow pointing out. The current being controlled is the current between the collector and the emitter. The base is the horizontal line coming into the middle of the transistor. The base acts as a knob which can limit the flow of current between the collector and emitter.

Figure 24-2 shows a conceptual picture of how the transistor operates. The connection from the base to the emitter operates as a diode, and the connection from the collector to the emitter operates as a variable resistor, offering resistances from zero (completely open dam) to infinity (completely closed dam). The resistance of the variable resistor is based on the current flowing through the base. The variable resistor is adjusted so that, under certain conditions, the *current* flowing from the collector to emitter is a certain multiple of the current flowing from the base to the emitter.

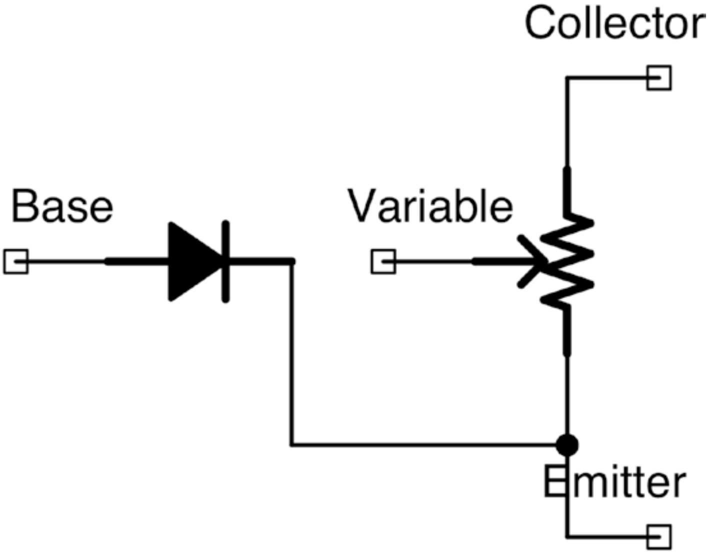

Figure 24-2. *A Conceptual View of Transistor Operation*

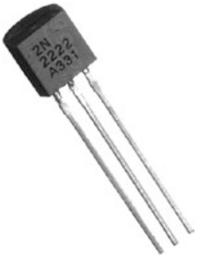

Figure 24-3. *A 2N2222A Transistor in a TO-92 Package*

When we talk about the voltages and currents flowing through the resistor, there are special names you need to remember and keep in mind. Each of the legs of the transistor is named by the first letter of its role. The collector is C, the emitter is E, and the base is B. Then, each of the voltages/currents is labeled based on which leg it is going from and to. Therefore, V_{BE} is the voltage difference between the base and the emitter, V_{CE} is the voltage difference between the collector and the emitter, and V_{CB} is the voltage difference between the collector and the base. I_{BE} is the current flowing between the base and the emitter, and I_{CE} is the current flowing between the collector and the emitter. The total current coming out of the emitter is $I_{BE} + I_{CE}$. Take time to think about these designations as we are going to be using them extensively when we talk about transistors and their usage in a circuit.

A photo of a transistor can be seen in Figure 24-3. This is the transistor we will focus on in this chapter—the 2N2222A (sometimes called the PN2222A). It is important to read the datasheet to find out information about your transistor—especially to know which transistor leg is which! For the picture in Figure 24-3, the collector is on the right, the base is in the middle, and the emitter is on the left. Note that some other transistors have *different* pin configurations, which is why it is so important to check the datasheets. For instance, the P2N2222A (very similar name!) has the collector and emitter pins reversed!

Transistors can also come in a variety of shapes and sizes, known as **packages**. The package depicted in Figure 24-3 is known as a TO-92 package. Other packages you might see are a TO-18 package (which looks like a tiny metal cylinder) or a package similar to an integrated circuit.

24.4 NPN Transistor Operation Basics

To fully understand NPN transistor operation requires a lot of complicated mathematics. However, you can get a "good enough" understanding of it just by remembering a few simple rules.

Rule 1: The Transistor Is Off by Default

By default, if there is no current flowing in the base, there will be no current flowing from the collector to the emitter. NPN transistors default to an "off" state.

Rule 2: V_{BE} Needs to Be 0.6 V to Turn the Transistor On

Remember that there is essentially a diode connecting the base to the emitter. Diodes have a voltage drop of about 0.6 V. Therefore, once the base voltage rises to 0.6 V *above* the emitter voltage, the transistor will turn on, and current will start to flow.

Rule 3: V_{BE} Will Always Be Exactly 0.6 V When the Transistor Is On

This is a corollary of the previous rule. Remember from Chapter 8 that we used diodes to give us fixed voltage differences between points on a circuit. This is the same in a transistor. Because the BE junction acts as a diode, the base will always be 0.6 V above the emitter while the transistor is turned on.

Rule 4: The Collector Should Always Be More Positive Than the Emitter

While technically you can have the collector go below the voltage of the base or the emitter, it is generally a bad idea with NPN transistors. It makes the circuit much harder to analyze. This book will assume that the circuit is set up in this manner.

Rule 5: When the Transistor Is On, I_{CE} Is a Linear Amplification of I_{BE}

So, when the transistor is on, the transistor amplifies the *current* flowing from the base to the emitter by adjusting the floodgates between the collector and the emitter. The multiplier that the transistor amplifies by is known as the transistor's **beta**—this is the current gain that an NPN transistor provides. The symbol for this value can be either β or h_{FE}. The problem with a transistor's beta is that it isn't very exact or very constant. A batch of "identical" transistors can have betas that vary quite a bit. And, while they are operating, their temperature and other environmental factors will affect the beta as well. There are things you can do to compensate for this, but for now just realize that it happens. While there are transistors with a wide variety of ranges of their betas, the most common NPN transistors have a beta of around 100.

The exceptions to this are in Rules 6 and 7.

Rule 6: The Transistor Cannot Amplify More Than the Collector Can Supply

This is mostly a reminder that the amplification comes *from* the collector current. If the collector can't supply the amplification, it won't happen. Basically, we need to think of the transistor as controlling a resistor from the collector to the emitter which will adjust itself to maintain the ratio (beta) between the base-emitter current and the collector-emitter current. Thus, it can't provide less resistance than no resistance.

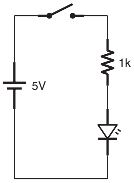

***Figure 24-4.** Using a Switch at the Start of a Circuit*

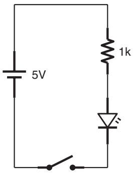

***Figure 24-5.** Using a Switch at the End of a Circuit*

Rule 7: If the Base Voltage Is Greater Than the Collector Voltage, the Transistor Is Saturated

If the base voltage rises above the collector voltage, this causes the transistor to behave as if there was no resistance going from the collector to the emitter. This is known as **saturation mode**.

Using these rules, thinking about transistor action is fairly straightforward. In the next section, we are going to put these rules into practice.

24.5 The Transistor as a Switch

One of the issues with transistors is that it takes a while before using a transistor becomes intuitive. Transistors, as we will see in the forthcoming chapters, wind up needing a lot of special considerations. Because of this, many people don't use transistors directly and instead choose to only use integrated circuits (see Chapter 11). Integrated circuits, since they are based on the needs of a circuit instead of simple physical properties, tend to be much easier to work with. Most of the guess work has been taken out, and the chips are built so that they can be inserted into a circuit in a straightforward way. Even though some people opt to use integrated circuits instead of using transistors directly, it is worthwhile to understand their operation. It is always better to choose an option because you understand your available alternatives instead of choosing an option because that choice is the only one you understand.

Now, when we think of buttons and switches, most people naturally put the switch at the *beginning* of the circuit that is being turned on or off. Figure 24-4 shows what this looks like. However, it is just as valid to place the switch at the end, as shown in Figure 24-5.

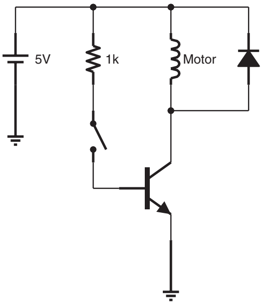

Figure 24-6. *Using a Transistor as a Switch for a Motor*

When using transistors as a switch, we almost always place them at the *end* of a circuit. The reason for this should become clear as we examine circuits.

The first circuit we will look at is Figure 24-6. In this circuit, there is no real *need* for a transistor—we could just as easily have put the switch where the transistor is. However, understanding how this setup works will help us understand other transistor circuits. In this circuit, the base is controlled by a small signal (the size of the signal is set by the size of the resistor). When the base turns on, it switches on the connection from the collector to the emitter, which controls current to the motor (the diode is simply a snubber diode as described in Chapter 20).

VISUALIZING INDUCTIVE KICK

If you wanted to see inductive kick in action, you can replace the snubber in the circuit in Figure 24-6 with an LED. Every time you turn of the switch, the LED will light up for a moment.

If the emitter of a transistor is simply connected directly to ground as it is in this example, the easiest way to analyze the circuit is by looking at the flow of current through the base. When the switch is closed, current will flow from the voltage source, through the resistor, across the transistor to ground. How much current? Well, this is actually a simple question. Remember that the junction from the base to the emitter can be treated as a simple diode (Rules 2 and 3). Therefore, we have a very simple circuit to analyze—voltage source to resistor to diode to ground. The voltage source is 5 V, and we have a diode (0.6 V) in the circuit path, so the voltage going through the resistor will be 4.4 V. Since the resistor is a 1 kΩ resistor, then, using Ohm's law, the current will be $I = \frac{V}{R} = \frac{4.4}{1000} = 0.0044\text{A}$, or 4.4 mA.

Since the current going through the base (I_{BE}) is 4.4 mA, how much current is flowing from the collector to the emitter? We will assume for our exercises that the beta of transistors is exactly 100. According to Rule 5, that means that the current going from the collector to the emitter is 100 times the base current. Therefore, the current going from the collector to the emitter can be determined by $I_{CE} = \beta \cdot I_{BE} = 100 \cdot 4.4 = 440\text{mA}$. Therefore, the current going from the collector to the emitter is 440 mA. Because the current at the collector is 440 mA, that means that the current going through the motor is also 440 mA.

Depending on the current we *wanted* flowing through the motor, we could choose other resistor values to set the current to the appropriate level. However, because of Rule 6, the collector current will be limited by the motor's characteristics as well.

So why did we put the transistor at the end of the circuit? Let's imagine that we added one component to the circuit—a resistor after the emitter before the ground. All of a sudden, the circuit is a lot harder to analyze. Why? Well, we can no longer determine the base current by simply looking at the current path through the base. The voltage across that final resistor will *not* be based on the base current, but based on the *combined* current from both the base and the collector. Thus, when there are components after the emitter, in order to analyze the base current, you have to analyze how the components after the base respond to the *combined* current of the base and collector, but we won't know that until *after* you calculate the base current. It is *possible* to do this, but the math isn't fun. Instead, by connecting the emitter *directly* to ground, we simplify our calculations because we *know* the voltage after the emitter—since it is connected to ground, it will be zero.

Thus, by connecting the emitter directly to ground, we can analyze the current flow from the base *independently* of the current flow from the collector. Additionally, connecting the emitter to ground will automatically make sure our DC circuits follow Rule 4 and make it easy to analyze when the transistor turns on and off by Rule 2.

Just as we have looked at several common resistor circuit patterns, there are also several common transistor circuit patterns. The transistor pattern we are looking at in this chapter is known as a **common emitter** circuit. This is because the "interesting" parts of the circuit are at the base (which provides the current to be amplified) and the collector (where the preceding circuit enjoyed the amplified current) and the emitter is connected to a common reference point (the ground in this case).

24.6 Connecting a Transistor to an Arduino Output

To understand *why* we would use a transistor for a switch in the first place, let's imagine that we have a motor just like in Figure 24-6, but in this case we want it to be controlled by an output pin from an Arduino running on an ATmega328/P.

Now, on this chip, output pins put out 5 V, but their current rating maxes out at 40 mA. However, motors usually require much more current than that. If we wanted our motor to use as much current as in Figure 24-6, it would blow out the chip if we tried to connect it directly to the output.

However, we can instead wire the output of the Arduino to the base of a transistor. Then, the current coming out of the Arduino will be *much smaller* than the current used by the motor.

Figure 24-7 shows how this is configured. The schematic for this setup is almost identical to that of Figure 24-6. The only difference is that the electrical output of the Arduino is being used to control the base current instead of a mechanical switch. We can even calculate exactly how much current will be used. Since there is a diode (in the transistor), the resistor will be using 4.4 V. Therefore, since $I = \frac{V}{R}$, then $I = \frac{4.4}{1000} = 4.4\text{mA}$. With a beta of 100, that means that our motor can source 440 mA.

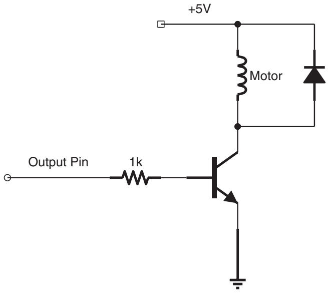

Figure 24-7. *Using a Transistor to Control a Motor with a Microcontroller*

24.7 Stabilizing Transistor Beta With a Feedback Resistor

As we mentioned earlier, the transistor's beta is not a very stable parameter. If we mass-produced something with a transistor, each device would wind up having a transistor with a different beta. Additionally, as the device was used, the beta would *drift*, meaning that things such as the temperature of the transistor would affect the beta, so it wouldn't even have the same value the entire time it was turned on!

In order to get around this, engineers have developed ways of stabilizing the *actual* gain of a circuit even when the transistor's beta shifts. The way that this is achieved depends greatly on the type of circuit used. In a common emitter circuit like the one we have been looking at in this chapter, we can add a resistor to the emitter to stabilize the transistor's gain. I know that I *just said* not to do this because it is hard to analyze the effects. However, if you add a single resistor to the emitter, other people have worked out the math for you in order to make this work easily.

Now, to imagine *why* adding a resistor to the output stabilizes the transistor beta, we have to think about what happens when you do this. Without the resistor, when the transistor beta increases, it simply increases the current flow at the collector. Since the emitter is connected to ground, there aren't any more real effects. However, if I add a resistor to the emitter, then increasing the current flow at the collector will *also* increase the voltage at the emitter. Now, since the voltage between the base and the emitter (V_{BE}) *must* be at 0.6 V, this will actually *reduce* the current in the base.

This is known as **feedback**—where the *output* of the circuit comes back in some way to affect the input. Sometimes, feedback occurs because the output is wired back to the input, but in this case, the resistor simply increases the voltage at the emitter, causing the base to pull back.

What the final effect of adding a resistor to the base will do is to *limit* the gain of the transistor to a *fixed* value that won't change either due to manufacturing issues or due to changes during operation as the transistor beta changes due to temperature. Additionally, the computation for this is very simple. If your transistor beta varies between 50 and 200, you can add a resistor to the emitter which will limit the actual transistor gain below 50, and it will keep this value very stable even as the transistor's beta drifts. In such a configuration, the most stable gain you can get for a given transistor beta is about $\frac{1}{4}$ of the transistor's lowest beta.

The equation for the effect that a feedback resistor has on a transistor circuit depends on the circuit design itself—there isn't a one-size-fits-all rule. However, feedback resistors are usually fairly small compared to the rest of the circuit.

For most circuits, simply start with a small feedback resistor (10–100 Ω). The feedback resistor will usually improve the stability of your design, even if you can't calculate exactly how that affects the beta of the transistor. It means that your design will likely continue to work even with different transistors and operating at different temperatures. Just know that while you add stability, this will also limit the gain (which is sometimes problematic with motors).

24.8 A Word of Caution

One thing to keep in mind with transistors is that they can get very hot! Transistors have a maximum current rating, and that often is based on the amount of heat that they are able to dissipate. Transistors that are able to handle large currents are called **power transistors**, and usually have an attachment for a **heatsink**, which helps them dissipate heat to the air more efficiently.

If you build a circuit similar to the one in this chapter, be sure to keep in mind both the specs of the motor (how much voltage and current is required for it to operate) and the specs of the transistor (how much current the transistor can handle). If the motor doesn't have enough voltage or current, it might not turn on, and if the transistor can't handle the current, it could easily burn out.

Review

In this chapter, we learned the following:

1. It is often beneficial to be able to control a high-power signal using a low-power signal.
2. The two major types of transistors are bipolar junction transistors (BJTs) and field effect transistors (FETs).
3. BJTs are current-controlled devices and FETs are voltage-controlled devices.
4. The terminals of a BJT are the base (B), the collector (C), and the emitter (E).

5. Voltages and currents going through a transistor are labeled using the terminals that the current is passing through. For instance, I_{BE} refers to the current flowing between the base (B) and the emitter (E), and V_{CE} refers to the voltage difference between the collector (C) and the emitter (E).

6. BJTs come in two main configurations—NPN and PNP.

7. With NPN transistors, a small positive current at the base causes a larger current to flow from the collector to the emitter.

8. With PNP transistors, current at the base reduces the current flow from collector to the emitter.

9. NPN transistors can be analyzed using these seven rules:

    - The transistor is off by default.
    - V_{BE} needs to be 0.6 V to turn the transistor on.
    - V_{BE} will always be *exactly* 0.6 V while the transistor is turned on.
    - The collector should always be more positive than the emitter.
    - When the transistor is on, I_{CE} is a linear amplification of I_{BE}.
    - The transistor cannot amplify more than the collector can supply.
    - If the base voltage is greater than the collector voltage, the transistor is saturated (it will offer no resistance from the collector to the emitter).

10. The amount of current amplification a transistor provides is known as the transistor's beta (also known as β or h_{FE}).
11. The actual beta of a transistor is not very stable and fluctuates quite a bit both because of manufacturing differences between transistors and even within a single transistor due to environmental changes such as temperature.
12. When used as a switch, transistors are usually placed at the *end* of a circuit in order to make the circuit analysis easier.
13. Transistors are often used to couple the outputs of devices with low current limits (such as a microcontroller) with devices that require higher output currents (such as motors).
14. To stabilize a transistor circuit's gain so that it doesn't vary with the transistor's beta, you can add in a small emitter resistor.

Apply What You Have Learned

Unless otherwise specified, assume that the transistor is a BJT NPN transistor and that the beta is stable:

1. If the base of a transistor is at 3 V and the transistor is on, what will be the emitter voltage?
2. If the base of a transistor is at 45 V and the emitter is on, what will be the emitter voltage?

3. If the base of a transistor is at 5 V and the emitter, if conducting, would have to be at 4.5 V, is the transistor on or off?

4. If the base of a transistor is at 0.6 V and the emitter is at ground, is the transistor on or off?

5. If the base of a transistor is at 0.4 V and the emitter is at ground, is the transistor on or off?

6. If a transistor has a base current (I_{BE}) of 2 mA and the transistor has a beta of 55, how much current is going through the collector (I_{CE})?

7. In the previous problem, how much total current is coming out of the emitter?

8. If a transistor has a base current of 3 mA and the transistor has a beta of 200, how much current is going through the collector?

9. If the base voltage is greater than the collector voltage, what does this mean for our transistor operation?

10. The output of your microcontroller is 3.3 V and supports a maximum output current of 10 mA. Using Figure 24-7 as a guide, design a circuit to control a motor that requires 80 mA to operate. Assume that the transistor beta is 100.

11. Redesign the previous circuit so that it utilizes a stabilizing resistor on the emitter to prevent variations in the transistor beta.

CHAPTER 25

Transistor Voltage Amplifiers

In Chapter 24, we started our study of the BJT NPN transistor. We noted that what the transistor actually amplified was *current*, so that the current coming into the collector was a multiple (known as β) of the current coming into the base.

Even though what a transistor does is provide *current* amplification, in this chapter, we will learn how to transform that into voltage amplification.

25.1 Converting Current into Voltage with Ohm's Law

If the transistor provides us with current amplification, how might we translate an amplification in the amount of current into an amplification in the amount of voltage? The answer is simple—Ohm's law describes the relationship between current and voltage: $V = I \cdot R$. Therefore, a current amplification can be transformed into a voltage amplification if we use a resistor! The larger the resistor, the larger the change in voltage drop that a given change in current will induce for that resistor.

To see that happening, take a look at the circuit in Figure 25-1. Note that this circuit on its own is rather useless, but it is helpful for illustrating how the calculations work. In this circuit, the current at the

© Jonathan Bartlett 2020

J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_25

base is controlled by the resistor R_B. This current will be amplified into an increased current from the collector. However, the current at the collector is driven through a resistor, R_C. Because this is through a resistor, that means that Ohm's law will take effect, and the size of the voltage drop across R_C will depend on the current running through it.

Remember Ohm's law states that $V = I \cdot R$, so any increase in current will increase the voltage drop across R_C, at least until the voltage at the collector is equal to the base voltage (which, in this circuit, is 0.6 V). If that happens, there is nothing more the transistor can do—it will just treat the collector-emitter junction as a short circuit.

Let's calculate what our circuit is actually doing. The voltage across the base is 5 V – 0.6 V = 4.4 V (remember we have to account for the diode-like voltage drop in the transistor from the base to the emitter). Therefore, using Ohm's law, we can calculate the base current at I = V/R = 4.4/10000 = 0.0004 A.

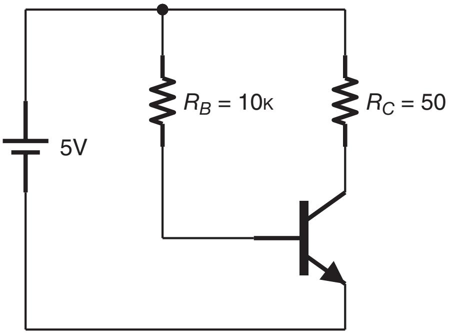

Figure 25-1. *A Simple Current-to-Voltage Amplifier*

Let's assume the transistor beta is 100. Therefore, the current flowing at the collector will be 0.0004·100 = 0.040 A. So the voltage drop across the resistor can be calculated using Ohm's law. $V = I{\cdot}R$ = 0.040·50 = 2 V.

Now, let's say that we change R_B so that we have more current running in the transistor. Let's decrease R_B from 10 kΩ to 6 kΩ. Now the base current will be $I = V/R = 4.4/6000 = 0.000733$ A. Now the current flowing at the collector will be $100 \cdot 0.000733 = 0.0733$ A. So the voltage drop across the resistor is now $V = I \cdot R = 0.0733 \cdot 50 \approx 3.67$ V.

When we increase the current, we increase the voltage drop across the resistor. You may be wondering what happens to the extra voltage. That is, since the emitter of the resistor is at ground and the voltage across R_C keeps changing, where is the remainder of the voltage? The transistor essentially swallows it up.

Remember, in our model of the transistor in Figure 24-2, the transistor acts as a variable resistor for the collector current. Therefore, the rest of the voltage drop happens *within* the transistor.

So, in effect, what we are doing is to translate changes in current at the base into changes in the voltage drop across R_C (and likewise the V_{CE} of the transistor).

As you might have noticed, when dealing with transistors, the place where the "action" occurs is not always right where you might expect it. In this circuit, the location where the voltage amplification *actually occurs* is at a resistor (R_C) connected to the *collector*. So the transistor, acting as an amplifier, is not sending the amplified signal to the emitter. Instead, you can think of it as sending the amplified signal to the collector, where the resistor at the collector converts the amplified current into a corresponding voltage drop.

> **Example 25.28** In the circuit given in Figure 25-1, what is the voltage across the resistor R_C if the base resistor R_B goes up to 20 kΩ?

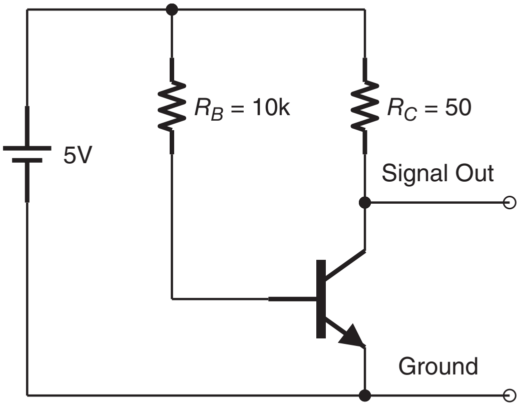

Figure 25-2. *Reading the Amplified Signal from a Voltage Amplifier*

$$I_B = 4.4 / 20000 = 0.00022\text{A}$$
$$I_C = 100 \cdot I_B = 100 \cdot 0.00022 = 0.022\text{A}$$
$$V_{RB} = I_C \cdot R_B = 0.022 \cdot 50 \approx 1.1\text{V}$$

Just to see where we are going, eventually we will use small voltage changes in the base to trigger current changes in the base which will then be amplified into a larger change in the voltage across R_C.

25.2 Reading the Amplified Signal

So we have managed to create a voltage drop which changes in response to changes in current at the base. But how do we read this voltage drop? It is rather difficult to read it directly, but we can read its *inverse* directly.

Take a look at Figure 25-2. In this figure, we added some output signal line to show where we would read the output of the amplifier (i.e., where we would connect the rest of the circuit that receives the amplification). We put the output line *between* the collector resistor R_C and the transistor. What this will do is give us the voltage of the source voltage (5 V) *minus* the voltage across R_C. So, when we have a large voltage across R_C, that will be reflected in a low voltage in our output. Likewise, when there is a low voltage across R_C, that will be reflected in a high voltage in our output.

This sort of an output is known as an **inverted output**, because the output voltage is essentially reverse-amplified. That actually works just fine for audio signals, as it does not matter to the listener if the signal is inverted or not. However, if we needed to get it back to the non-inverted form, we could just add another amplification stage onto the end (we will see how to do this in Section 25.4, "Adding a Second Stage").

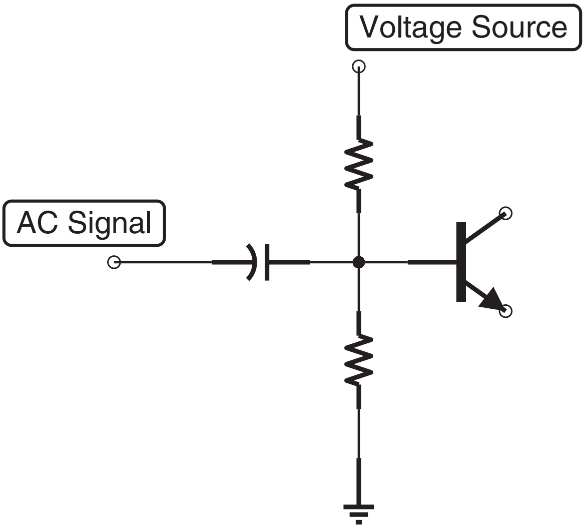

Figure 25-3. *Components of a Transistor Biasing Circuit*

Having said all that, I should point out that we still don't know how to amplify an audio signal—yet. That is coming in the next section.

25.3 Amplifying an Audio Signal

What we really want to do is to amplify an audio signal. Imagine that someone is singing into a microphone, and we want to amplify the signal we get so that we can send it to a speaker. How would we do that?

There are a number of problems that you have to solve in order to get this done. You might imagine that you could just connect a microphone to the base of the transistor and just amplify directly. That's a good idea, but sadly life is not always that easy. To understand why, remember that audio signals are basically alternating current. That means that the signal will swing both positive *and* negative. Also remember that the base voltage has to remain *above* the emitter voltage and the emitter is tied to ground. Therefore, if we tried to do this, we would lose the bottom (negative) half of the signal. In fact, if it was a small signal, we might lose the *entire* signal if it never reached the required 0.6 V above ground.

Therefore, the signal coming into the transistor needs to be **biased**. We basically need to bring the voltage of the incoming signal up so that it keeps the transistor on, but doesn't overwhelm the signal coming in.

If you remember in Chapter 19 when we built the tone generator, we had to unbias the signal. We did that by taking the signal and coupling it through a capacitor. That allowed the two sides to exist at different base voltages, but transmit the *changes* in the voltage through the capacitor. This will be the same idea, but where the input is unbiased and the output is biased.

Figure 25-3 shows what this looks like in general terms. A simplistic way to think about this is to think of it as a voltage divider. The resistors are basically setting a bias voltage, and that voltage is varied based on the current that comes in through the capacitor.

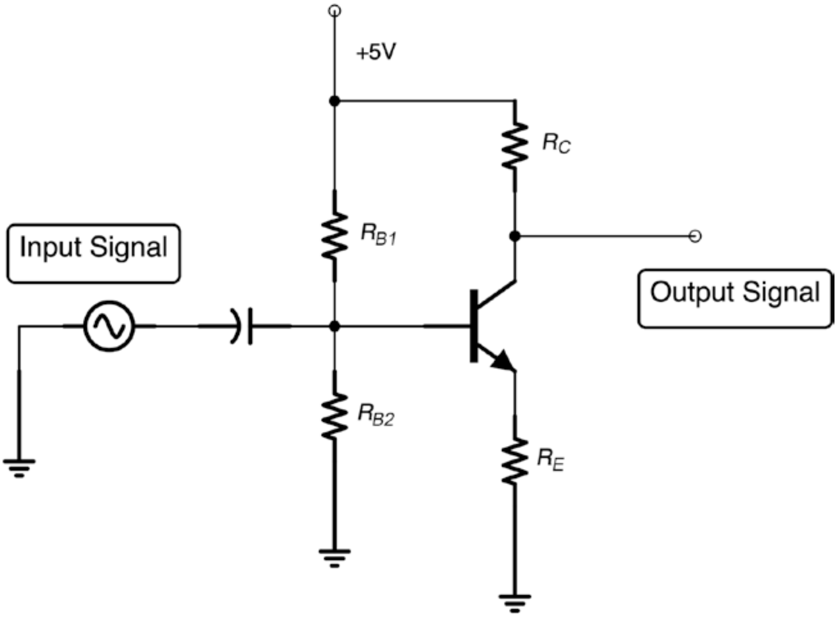

Figure 25-4. *A Single-Stage Transistor Voltage Amplifier*

While that is somewhat true, the details get more complicated as we add more components, especially the transistor. The emitter of the transistor will be connected to ground though through a resistor, R_E. This means that, for the transistor to be on, the actual voltage coming in to the base of the transistor must always be 0.6 V (or above, based on R_E). That means that the voltage divider itself must always be 0.6 V (or slightly above, due to R_E) as well.

Figure 25-4 shows all of the parts of our simple transistor amplifier. The main question is, what do we set the values of each of these resistors to? There are many considerations, and they affect both the amount of gain you will get in your amplifier and whether or not the signal is clipped.

Clipping will happen if the base voltage goes below 0.6 V because the transistor will simply stop conducting (and actually a little above that because of R_E). Clipping will also happen if the R_C resistor is too large, because if the current going through R_C causes the voltage at the collector to get too low, the transistor will have reached its saturation point and therefore will stop amplifying. Since R_C is converting the current into voltage, it also needs to be large enough to yield a wide voltage swing with the amount of current expected.

For the base resistors (R_{B1} and R_{B2}), you need to keep in mind how much current you expect the input source to have. Remember that this current is going to be going to be alternating between the positive and negative directions. Therefore, when the input source swings negative, there has to be enough current coming through R_{B1} to keep the voltage positive and keep the transistor on. When the input source swings fully positive, R_{B2} can prevent clipping as well, as it provides an alternate path for incoming current that isn't through the transistor.

You may be wondering what the resistor at R_E is doing. R_E is a feedback resistor. The gain (β) of a transistor varies both within manufacturing and based on temperature. A feedback resistor, by raising the voltage at the emitter of the transistor, will limit the amount of gain that the transistor can deliver based on its source input.

Feedback means that the output of a circuit is affecting the input in some way. Sometimes, feedback occurs because the output is wired back to the input. However, in this case, the resistor simply increases the voltage at the emitter, causing less current to flow in through the base and more current to flow through R_{B2}.

Limiting the gain of the resistor stabilizes the gain over a variety of operating parameters, so that your circuit will continue to work as the transistor heats up. Additionally, were you to manufacture the circuit, limiting the gain would make it less susceptible to manufacturing differences between individual parts.

In any case, if the output of the circuit has a resistance R_O, then the current gain from the input through the output can be calculated as

$$\text{gain} = \frac{R_{B2} R_C}{R_E R_O} \qquad (25.1)$$

To understand why this formula works, see Appendix D, Section D.5, "Output Gain Calculations in BJT Common Emitter Applications."[1] The formula just gives you a starting point, though you normally have to play around with resistor values a little to get it to work.[2] Also note that the formula only works when $R_O > R_C$.

When you think about your circuit, you will want to analyze it at two points—when the source current coming in at the input is at its highest and at its lowest (which in this case will be negative). If you have developed your circuit correctly, this should give a good swing in the voltage across R_C, but without ever clipping (the transistor turning off or going into saturation).

The best place to start is to recognize that your voltage divider must always provide enough current to keep the transistor on even when the source current swings negative. Therefore, the current coming in to the base through the voltage divider should be a little larger than the signal's current. This is known as the **quiescent current**.

In our case, we are going to develop our circuit to handle passive microphones, so the source signal current will only be on the order of about 10 microamps (0.01 mA). Therefore, we will set our quiescent current to be 0.1 mA. The means that R_{B1} should be a little less than 50 kΩ. We will use a standard 47 kΩ resistor here. R_{B2} can tolerate a wide range of values—basically anything that is at least a quarter the size of R_{B1} (we can't let the voltage drop below the emitter voltage, or the transistor will stop conducting). We'll make it 20 kΩ here.

---

[1] I will warn you that the calculations are pretty ugly (which is why they are in the appendix). Also, to fully understand what is happening, you need to read the rest of this chapter through Chapter 26.

[2] The reason for this is that the formula itself was made by making a *lot* of assumptions and simplifications. Those don't always hold in the real world, so you have to make adjustments.

We want a value for R_C that will swing the voltage about half as far as we can go (we have to save headroom for both the 0.6 V drop on the transistor and the actual output to the rest of the system). If we want a 25x gain, then, if our max base current will be 0.11 mA (quiescent current plus source current), that will put our max collector current at 2.75 mA (25 ∗ 0.11). We want R_C to be dropping about 1.5 V at this level (we need to save some headroom for a number of purposes, including staying above the voltage drop at R_E, the additional diode drop to stay above the base voltage, and some additional which will be used for the current going out of this stage of the amp). Therefore, the resistance for a 1.5 V voltage drop at 2.75 mA is $R = \frac{V}{I} = \frac{1}{0.00275} \approx 545\Omega$. Anything near that value should work fine, so we will go with a 510 Ω resistor.

Then, we mentioned earlier that the total gain will be given by $\frac{R_{B2}RC}{R_E R_O}$. So far, we have $\frac{20{,}000 \cdot 510}{R_E R_O}$. We can estimate the next stage of the amplifier as having a resistance of 2,000 Ω (this is the base resistor of the next stage of the amplifier—see Appendix D, Section D.5, "Output Gain Calculations in BJT Common Emitter Applications," for why this works as an estimate). Therefore, to get a 25x gain, we can do

$$\begin{aligned}
\text{gain} &= \frac{R_{B2}R_C}{R_E R_O} \\
25 &= \frac{20{,}000 \cdot 510}{R_E 2{,}000} \\
R_E &= \frac{20{,}000 \cdot 510}{2{,}000 \cdot 25} \\
R_E &= 204
\end{aligned}$$

In testing, I found that a smaller resistor worked better for R_E, so I used a 47 Ω resistor. Again, the equations are greatly simplified and only give a starting point for experimentation.

25.4 Adding a Second Stage

A single amplification stage is not always enough. If your signal source is weak enough, sometimes you need more power just to hear it. This can be accomplished in a number of ways. The simplest, conceptually, is to add another output stage to your amplifier.

In order to do this, we need to design the next stage to take into account the output of our first stage. We will need to use a coupling capacitor to handle the change in voltage from the output of the first stage into our bias circuit for the second stage. Without the coupling capacitor, the voltage characteristics of the transistor would severely alter the way that our previous stage works. Therefore, the coupling capacitor helps us to isolate the different parts of the circuit.

Now, the current coming in to this stage will be significantly larger than the previous stage. Our previous stage had about 0.01 mA coming in. While not exact, with an actual 25x gain, the current coming in will probably be about 0.25 mA. Therefore, we have to plan this stage differently (i.e., using a different quiescent current).

Using the same procedure as we used for the previous stage, we can use 0.75 mA as the quiescent current and utilize a 5.1 kΩ resistor (to choose a standard value) for R_{B1} and a 2 kΩ resistor for R_{B2}. Our largest swing forward will now be 1 mA, and therefore, the amplified current will be $25 * 1 = 25$ mA. To get the 1.5 V drop, we will need a resistor for R_C of $\frac{1.5}{0.0025} = 60\Omega$. We will approximate to a 100 Ω resistor.

The output resistance will be our headphones, which we can estimate at 16 Ω. Doing the calculation gives

$$\text{gain} = \frac{R_{B2}R_C}{R_E R_O}$$

$$25 = \frac{2{,}000 \cdot 100}{16 \cdot R_E}$$

$$R_E = \frac{2{,}000 \cdot 100}{25 \cdot 16}$$

$$R_E = 100\Omega$$

Again, I had to modify the R_E to get better performance. This is unsurprising because the equation is only even moderately reliable when $R_O > R_C$. Lowering R_E of this stage to 10 Ω seemed to work well.

The output of the second stage will again be coupled through a coupling capacitor to your headphone output and can be attenuated by another resistor (or a variable resistor for volume control) if needed.

Figure 25-5 shows the complete two-stage amplifier. With this circuit, you can use most microphones and most headphones to get enough output power to hear yourself in your headphones (in fact, if you don't have a microphone, headphones can actually work as a microphone in a pinch). Figure 25-6 shows the two-stage amplifier built into a breadboard.

All of the capacitors can be pretty much any standard capacity above around 100 nF. The ones that couple the DC bias signal, if they are polarized, should have their negative side toward the unbiased signal. The coupling capacitor between the two stages should ideally be non-polarized, though with these power levels it doesn't matter too much. If you do need to use a polarized capacitor here, it is best to have the negative side facing the second stage, since that should be right around 0.6 V, while the positive side should be swinging in the positive range.

One thing in this circuit we haven't discussed is the bleed-off resistors connected right next to the coupling capacitors for the input and output circuit. These are very large (1,000,000 Ω) resistors. They are large so that they do not have any effect on the calculations in the circuit itself, but they do provide a place for the negative side of the capacitor to drain out when nothing is connected. Otherwise, any residual charge on the capacitor has no place to go and will remain even when the circuit is turned off and the devices are unplugged.

25.5 Using an Oscilloscope

Designing transistor amplifiers can be tricky because there are a lot of things that can go wrong. Although there is some leeway, using the wrong resistors can result in distortion or a loss of gain. Misconnecting a single component can render the entirety of the circuit inaudible.

Because of the variety of things that can go wrong when building audio circuits, it is best to use an oscilloscope. An oscilloscope can help you visually see what the voltages look like at each point in your circuit. Oscilloscopes can cost as little or as much as you want them to. There are pocket oscilloscopes that you can purchase for less than a nice dinner out, and there are bench oscilloscopes that you would need several months' salary to purchase. Any of those are helpful in analyzing your circuit, so you can see what is happening to your voltages at every point in the circuit.

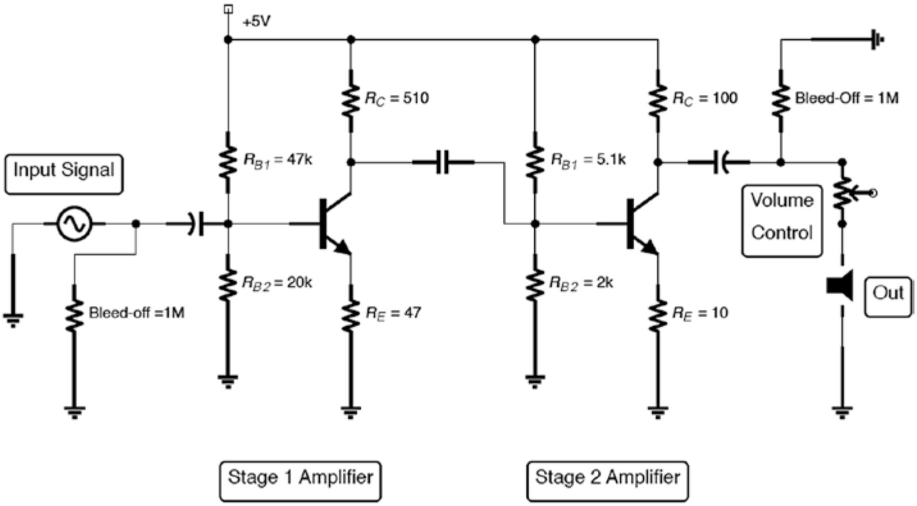
+5V
RC = 510
RB1 = 47k
Input Signal
RB2 = 20k
Bleed-off =1M
RE = 47
RC = 100
RB1 = 5.1k
RB2 = 2k
RE = 10
Bleed-Off = 1M
Volume Control
Out
Stage 1 Amplifier
Stage 2 Amplifier

Figure 25-5. *The Complete Two-Stage Amplifier Circuit*

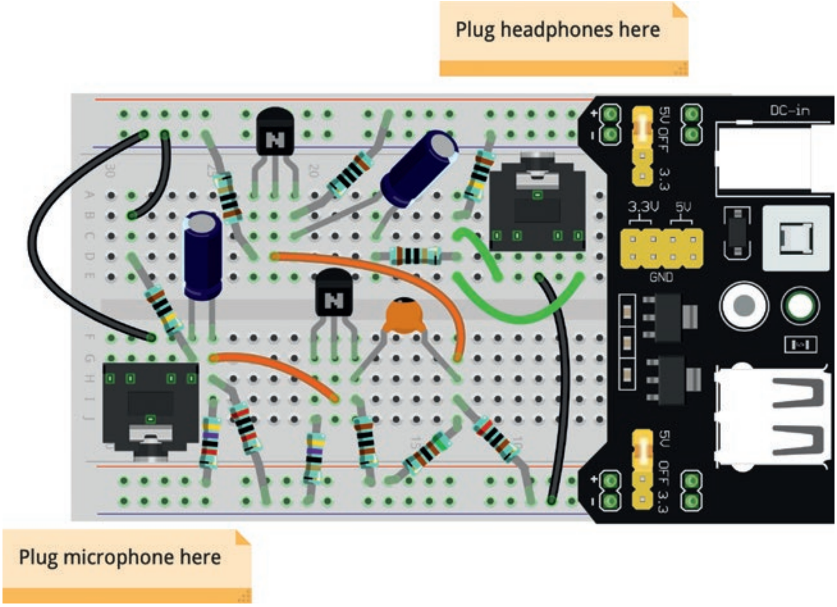
Plug headphones here
Plug microphone here
DC-in
5V OFF 3.3
3.3V 5V
GND
5V OFF 3.3
N
N

Figure 25-6. *The Two-Stage Amplifier Built into a Breadboard*

Review

In this chapter, we learned the following:

1. Although transistors provide current amplification, current changes can be transformed into voltage changes using Ohm's law.
2. Using a resistor at the collector allows us to "read" a voltage change based on the changes in the currents going through the transistor.
3. Using a resistor in this way *inverts* the waveform—it will show low voltage when there is a lot of current and high voltage when there isn't much current.

4. Adding in an emitter resistor limits the amount of voltage gain in the circuit in order to compensate for variable/drifting transistor betas.
5. In order to amplify an audio signal, we have to add a DC bias to the signal so that the transistor stays in its operating range (positive voltage).
6. The simplest way to use a transistor to amplify a signal is to add a voltage divider to the base of the transistor, with a coupling capacitor feeding the signal into the voltage divider.
7. The neutral, “no signal” design point is known as the quiescent point of a circuit. The quiescent point is the state of the circuit when the AC signal coming in is neutral (0 V).
8. The AC signal is coupled into the voltage divider through a coupling capacitor in order to manage the difference between the pure AC signal and the DC biased signal.
9. The circuit should be analyzed both at the highest and lowest swings around the quiescent point from the input signal.
10. The collector resistor should be chosen in order to maximize voltage swing while preventing overdrive.
11. Weak AC signals often need multiple amplification stages to provide sufficient output power for driving outputs.
12. The output of one amplifier can be coupled through a capacitor into the input of a second amplifier.

13. If the signal sources and outputs are not permanently connected, adding a bleed-off resistor will enable the coupling capacitors to drain out after the jacks are disconnected.
14. There are numerous things that can go wrong in a transistor amplifier, including clipping the audio signal, having a bad quiescent current, or accidentally losing your gain through bad resistor choices, not to mention just simply building the circuit wrong.
15. Because of the number of things that can go wrong, it is easiest to diagnose problems in an amplifier using an oscilloscope, which allows you to visualize what is happening at each point in the circuit.

Apply What You Have Learned

1. What is the purpose of the resistor in the collector of a transistor amplifier?
2. What is the purpose of the resistor in the emitter of a transistor amplifier?
3. Why is there a bias voltage on the base of the transistor? Why can't the signal just be connected in directly to the base?
4. Why is the signal coupled in through a capacitor?
5. Why does the single stage voltage amplifier discussed in this chapter invert its output?
6. If the output of the two-stage amplifier is coupled into a third stage, the signal current would swing 1.85 mA in either direction. Design a third amplification stage which can handle this amount of current.

CHAPTER 26

Examining Partial Circuits

We will end our discussion of amplification by discussing partial circuits. Oftentimes you will need to design a circuit which connects to another circuit, either powering it or receiving power from it. For instance, in the amplification circuits from Chapter 25, the outputs were connected to a speaker. They could also be connected to another amplifier or to a stomp box (a device to modulate the incoming signal in some way) or to a recording circuit.

26.1 The Need for a Model

In order to connect circuits together, we need to be able to describe, in general terms, the ways that circuits fit together. When dealing with transistors and other power amplification devices, we often need to come up with a simplified model for how the input to a circuit or the output from a circuit behaves. Early on (in Chapter 7), we learned how to take multiple resistors in series and parallel and combine them into an equivalent single resistor.

When dealing with a power amplification circuit, it is often necessary to look at various parts of the circuit by themselves and figure out how they *look* to other parts of the circuit. The way that a partial circuit looks to other parts of the circuit is called the circuit's **Thévenin equivalent** circuit.

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_26

A Thévenin equivalent circuit takes a partial circuit and reduces it to

- A single voltage source (AC, DC, or DC-biased AC, expressed in RMS voltage)
- A single impedance (i.e., resistance) in series with the voltage source

Note that the single voltage source may be *different* than the voltage source that is actually connected. What you are doing is seeing what the circuit looks like to another circuit. For instance, a voltage divider circuit makes the output of the voltage divider *look like* it is coming from a lower voltage source.

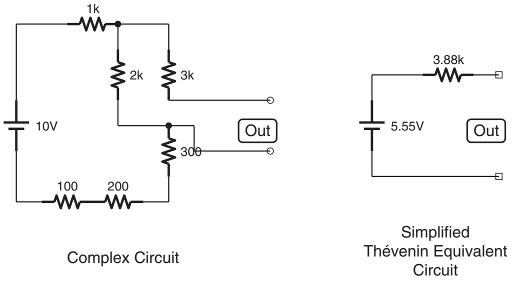

Figure 26-1. *A Complicated Circuit and Its Thévenin Equivalent Circuit*

Figure 26-1 shows a circuit and its Thévenin equivalent circuit. For purposes of thinking about and understanding the relationship between the circuit and things attached to the circuit, we can view the circuit as being the same as its Thévenin equivalent. Thus, having a Thévenin equivalent circuit greatly simplifies our modeling, calculating, and understanding of how circuits work together.

Any network of power sources and resistances can be converted into a Thévenin equivalent circuit. You can also get a Thévenin equivalent circuit for a circuit that includes capacitors and inductors, but the calculations become more difficult and the results are only valid for a specific frequency (each frequency will have a different Thévenin equivalent circuit). For simplicity, we will just focus on resistive circuits.

26.2 Calculating Thévenin Equivalent Values

To see how to calculate the voltage and resistance for a Thévenin equivalent circuit, this section will take a classic voltage divider circuit and analyze how it "looks" to other attached circuits. Figure 26-2 shows an example of a partial circuit. Like most partial circuits, this circuit has two output points—A and B. What we are wanting to know is this—if we attach another circuit up to A and B, is there a model that we can use to understand how the other circuit "sees" our circuit? The goal of making a Thévenin equivalent circuit is to understand what our circuit will look like to other attached circuits.

So, since our Thévenin equivalent circuit will have a voltage source and a single resistor, we need to calculate what the voltage and resistance of this circuit will be. To calculate the voltage, find out what the voltage of the circuit at the output is when there is *nothing connected*. That is, if we were to leave A and B disconnected and I were to connect my multimeter to A and B, what would the voltage be? This is your Thévenin voltage. Since this is a voltage divider, you can just use normal voltage divider calculations to find this out. In this case, we have a 12 V source, and the voltage divider divides it exactly in half (1 kΩ for each half). Therefore, the output voltage is 6 V. Therefore, our Thévenin equivalent circuit will have a 6 V source.

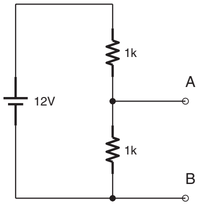

Figure 26-2. *A Voltage Divider Partial Circuit*

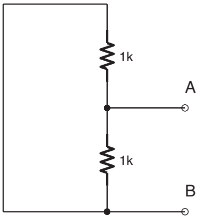

Figure 26-3. *Calculating the Thévenin Resistance of the Circuit*

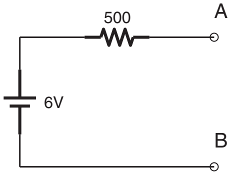

Figure 26-4. *The Thévenin Equivalent of the Voltage Divider*

Now we need to find our Thévenin resistance. There are multiple tricks to do this, but the simplest one is to replace all voltage sources in your circuit with a wire (i.e., a short circuit) and simply compute the total resistance between A and B.[1]

Figure 26-3 shows what this looks like. Therefore, to calculate the Thévenin resistance of this circuit, simply calculate the total resistance from A to B. In this case, there are two parallel paths from A to B—one through the first resistor and one through the second. Therefore, we add up the resistors as parallel resistances. As a result, our Thévenin resistance will be

---

[1] We haven't talked about *current* sources much in this book. However, for completeness, I should note that if you have a current source, you should replace it with an open circuit (i.e., a gap in the wire) when calculating Thévenin resistance.

$$\begin{aligned} R_T &= \frac{1}{\frac{1}{R_1}+\frac{1}{R_2}} \\ &= \frac{1}{\frac{1}{1000}+\frac{1}{1000}} \\ &= \frac{1}{0.001+0.001} \\ &= \frac{1}{0.002} \\ &= 500\Omega \end{aligned}$$

Therefore, we would say that this partial circuit has a Thévenin voltage of 6 V and a Thévenin resistance of 500 Ω. Whenever we attach a circuit to this circuit, what that other circuit will “see” is a circuit like the one in Figure 26-4.

If you wanted to prove this to yourself, you can imagine a variety of different circuits attached to both our original circuit and to the Thévenin equivalent circuit. You will find that, in all cases, the amount of voltage and current the Thévenin equivalent circuit provides to the other circuit is the exact same as what the original circuit will provide.

That isn’t to say that the circuits themselves are exactly equivalent. Our original voltage divider uses up a lot of current stepping down the voltage of the voltage source. Not only does that waste energy from our battery but it probably also causes a lot of heat. However, *any subcircuit that gets attached to A and B* will see both our original circuit and the Thévenin equivalent circuit as providing the same output.

26.3 Another Way of Calculating Thévenin Resistance

There is another way of calculating Thévenin resistance. In this method, we first calculate what the current would be if you shorted A to B directly with a wire. This is known as the short-circuit current, or I_{SHORT}. Then, after calculating this, you can divide the Thévenin voltage by I_{SHORT} to obtain the Thévenin resistance.

When doing this, you have to remember that anything in parallel with our short will be essentially ignored—the current will always want to go through our short circuit.

Figure 26-5 shows what this looks like. What we want to do is to calculate the current going from A to B. Since A to B is a short circuit in parallel with our second resistor, we know that *all* of the current will prefer the short circuit. This means that the current going through A and B will simply be the current that is limited by the first resistor.

So, since we have a 12 V source and a 1 kΩ resistor, the short-circuit current will be

$$\begin{aligned} I_{\text{SHORT}} &= \frac{V}{R} \\ &= \frac{12}{1000} \\ &= 0.012\,\text{A} \end{aligned}$$

Now, to determine the Thévenin resistance, we divide the Thévenin voltage by this number:

$$\begin{aligned} R_{\text{Thevenin}} &= \frac{V_{\text{Thevenin}}}{I_{\text{SHORT}}} \\ &= \frac{6}{0.012} \\ &= 500\Omega \end{aligned}$$

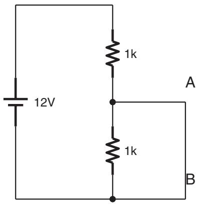

Figure 26-5. *Finding the Short Circuit Voltage*

As you can see, this is the same value that we got from the previous method.

26.4 Finding the Thévenin Equivalent of an AC Circuit with Reactive Elements

If a circuit has reactive elements (inductors and capacitors), we have to do a little more work to find the Thévenin equivalent circuit.

For DC circuits, this is relatively simple. Since capacitors block DCs and inductors are a short circuit for DCs, we can simply treat the capacitors as open circuits (i.e., unconnected) and treat the inductors as short circuits (simple wires). For AC circuits, you can get a feel for what this will be by assuming the opposite—that capacitors will be short circuits and inductors will be open circuits.

However, if you were to try to solve it explicitly, the problem is a little more difficult. The problem is that a full analysis of such circuits requires math involving complex numbers (i.e., numbers involving the imaginary unit *i*). While the technique is roughly equivalent to adding resistances in series and parallel as we have done before, it is much more difficult to do the math with complex numbers.

For the purposes of this book, the previous statements about DC and AC should suffice for a general understanding of how your circuit works.

26.5 Using Thévenin Equivalent Descriptions

Many circuits are described to their users using Thévenin equivalent descriptions. For instance, many circuits are described by their input or output impedance. This gives you a rough guide to imagine what will happen if you connect your own circuit to such circuits.

Imagine that you have a circuit that has a Thévenin equivalent output impedance of 500 Ω. If you connect an output circuit that only has 250 Ω of resistance, what do you think that will do to the signal? Well, since the output of the circuit is equivalent to going through a 500 Ω resistor (that's what Thévenin equivalence means), then if I connect a 250 Ω resistor, then I will have created a voltage divider in which two-thirds of the voltage will be dropped by the circuit I am connecting to and I will only get one-third of the output voltage.

On the other hand, if the impedance of my circuit is 50, 000 Ω, then the voltage drop coming out of the circuit in question is negligible compared to the voltage drop within my circuit. This means that my circuit will essentially receive the full Thévenin equivalent voltage.

We can also use this to calculate the amount of current that our circuit will draw. Let's say that a circuit yields a Thévenin equivalent output of 4 V with an 800 Ω impedance. If I connect a 3,000 Ω output circuit, how much current will flow? The total resistance will be 3,800 Ω, so the current will be $V/R = 4/3800 \approx 1.05$ mA.

The same is true for connecting an input circuit that you make to an output circuit someone else made. For instance, speakers and headphones are normally rated as an impedance—8 Ω, 16 Ω, and so on. They aren't, strictly speaking, resistors, but at normal audio frequencies, they behave essentially like one—they have a Thévenin equivalent impedance (their Thévenin equivalent voltage is zero).

Example 26.29 If I have an output circuit which is a Thévenin equivalent to 3 V RMS and 200 Ω and I connect it to a set of 16 Ω headphones, what will the power of the headphones be in watts?

We can understand this circuit as simply being a voltage source followed by two resistors in series. The voltage source will be 3 V, and the resistances will be 200 Ω and 16 Ω, totaling 216 Ω. The current will therefore be $V/R = 3/216 \approx 0.0139$ A. The voltage drop in the headphones will be $I \cdot R = 0.0139 \cdot 16 \approx 0.222$ V. Therefore the power delivered to the headphones will be $V \cdot I = 0.222 \cdot 0.0139 \approx 0.00309$ W, or 3.09 mW.

26.6 Finding Thévenin Equivalent Circuits Experimentally

In addition to using circuit schematics to determine Thévenin equivalent circuits, it is also possible to determine them experimentally. This way, if you are unsure of the input or output characteristics of your device, you can measure it yourself. The problem with measuring it yourself is that it requires attaching a load to the circuit. Some circuits will fry if a wrongly sized load is attached. You have been warned.

The easiest way to determine Thévenin equivalency experimentally is rather unsafe, but it will help us understand better why the method works. Imagine a voltage divider where the bottom resistor has an extremely large resistance—say, 100 MΩ. In such a voltage divider, the bottom resistor will have almost the entirety of the voltage drop, right? In fact, if the bottom resistor was infinite, it would in fact have all of the voltage drop.

Because of this, we can determine the Thévenin equivalent voltage by measuring the output voltage when there is nothing connected, because no connection means that there is infinite resistance between the output and ground. Measuring this value will give us the Thévenin equivalent voltage.

To determine the Thévenin equivalent current, we can short-circuit the output. This is physical equivalent to the conceptual method we used in Section 26.3, "Another Way of Calculating Thévenin Resistance." When doing this, the *only* impedances to the current will be within the device itself. Therefore, using Ohm's law, the amount of current this draws will tell us how large of a resistance the output is yielding.

> **Example 26.30** If I measure the open-circuit (i.e., disconnected) voltage of the output of an unknown circuit as 8 V and the short-circuit current of the output as 10 mA, what is the Thévenin equivalent circuit?
>
> To find this out, we simply use Ohm's law. What resistance would cause an 8 V source have 10 mA of current?
>
> $$\begin{aligned} R &= V / I \\ &= 8 / 0.010 \\ &= 800 \end{aligned}$$

Therefore, our Thévenin equivalent circuit is 8 V with an impedance of 800 Ω.

The problem with this method is that you don't normally want to short-circuit your output. Additionally, some circuits require some sort of a load to work properly. In order to adjust to such scenarios, there is a set of equations that allow us to measure the voltage drop across a large and small resistance (instead of infinite and no resistance) and come up with a Thévenin equivalent circuit.

The equations are a little complex, but you can actually derive them directly from Ohm's law if you work at it (see Appendix D, Section D.6, "The Thévenin Formula," to see it derived). The first one calculates the Thévenin equivalent voltage (V_T) from the voltage with a high resistance (V_H), the high resistance value (R_H), the voltage with a low resistance (V_L), and the low resistance value (R_L):

$$V_{\mathrm{T}} = \frac{\frac{V_{\mathrm{H}}}{R_{\mathrm{H}}}\left(R_{\mathrm{H}} - R_{\mathrm{L}}\right)}{1 - \frac{V_{\mathrm{H}}R_{\mathrm{L}}}{R_{\mathrm{H}}V_{\mathrm{L}}}} \tag{26.1}$$

Then, we can calculate the Thévenin equivalent resistance:

$$R_{\mathrm{T}} = \frac{V_{\mathrm{T}}R_{\mathrm{L}}}{V_{\mathrm{L}}} - R_{\mathrm{L}} \tag{26.2}$$

For a fairly safe and basic starting point, you can use 1 MΩ for the high resistance value and 1 kΩ for the low resistance value.

> **Example 26.31** I have a circuit that generates an output for which I need to know its Thévenin equivalent properties. I tested the circuit with a 200 Ω resistance for my low resistance and a 1000 Ω resistance for my high resistance. With the 200 Ω resistance, there was a 2 V drop across the resistance. With the 1000 Ω resistance, there was a 5 V drop across the resistance. What is the Thévenin equivalent circuit for this circuit?

First, we find the Thévenin equivalent voltage using Equation 26.1:

$$
\begin{aligned}
V_{\mathrm{T}} &= \frac{\frac{\mathrm{V_H}}{\mathrm{R_H}}\left(R_{\mathrm{H}} - R_{\mathrm{L}}\right)}{1 - \frac{\mathrm{V_H R_L}}{\mathrm{R_H V_L}}} \\
&= \frac{\frac{5}{1000}\left(1000 - 200\right)}{1 - \frac{5 \cdot 200}{1000 \cdot 2}} \\
&= \frac{0.005 \cdot 800}{1 - \frac{1000}{2000}} \\
&= \frac{4}{0.5} \\
&= 8\mathrm{V}
\end{aligned}
$$

Next we can find the Thévenin equivalent resistance using Equation 26.2:

$$
\begin{aligned}
R_T &= \frac{V_T R_L}{V_L} - R_L \\
&= \frac{8 \cdot 200}{2} - 200 \\
&= 800 - 200 \\
&= 600\Omega
\end{aligned}
$$

Therefore, our unknown circuit has a Thévenin equivalent voltage of 8 V and a Thévenin equivalent impedance of 600 Ω.

What makes this method valuable is that it allows a way to *experimentally* determine the Thévenin equivalent of a partial circuit that you don't have a schematic for or for which determining the Thévenin equivalent circuit might be difficult due to nonlinear components such as transistors.

Review

In this chapter, we learned the following:

1. In order to be able to connect circuits together without knowing all of the details of how they are implemented, we need a simplified model of how those circuits work with other circuits they are connected to.
2. A Thévenin equivalent circuit is a combination of a single voltage source and a single series impedance which models the way that the given circuit will respond to other attached circuits.
3. To calculate Thévenin equivalent voltage, calculate the voltage drop for an open circuit between the two terminals. This is the Thévenin equivalent voltage.
4. To calculate Thévenin equivalent impedance, calculate the impedance from one terminal to another (or to ground if there is only one terminal), replacing any voltage sources with short circuits.
5. Alternatively, to calculate Thévenin equivalent impedance, calculate the current flowing from one terminal to another if there was a short circuit between them. Then use Ohm's law to calculate the resistance.
6. Thévenin equivalent circuits can be used to understand how the resistances of attached circuits will affect the signal coming out of or into a circuit.
7. Thévenin equivalent circuits can also be found experimentally.

8. Although not recommended, the Thévenin equivalent voltage and resistance can be found easily by simply measuring the voltage drop of an open circuit across the terminals and the current flowing through a short circuit between the terminals.
9. A better option for experimentally measuring Thévenin equivalencies is by measuring the voltage with two different load resistances across the terminals. Then the Thévenin equivalencies can be found using Equations 26.1 and 26.2.

Apply What You Have Learned

1. Why would we want to know what a circuit's Thévenin equivalent circuit is?
2. What are the two components of a Thévenin equivalent circuit?
3. Think about the two-stage amplifier that you built in Chapter 25. How would you go about finding the Thévenin equivalent circuit as it is seen by the headphones?
4. Suppose I have a circuit where the output terminals have a 2 V drop when it is an open circuit and have 2 mA of current flowing through it when it is a short circuit. Draw the Thévenin equivalent circuit.
5. If I have a Thévenin equivalent circuit of 4 V with an impedance of 400 Ω, what will be the voltage drop of the load if I attach a 2000 Ω resistor across the output?

6. If I have a Thévenin equivalent circuit of 3 V with an impedance of 100 Ω, what will be the voltage drop, the current, and the power of the load if I attach headphones rated at 32 Ω?

7. Calculate and draw the Thévenin equivalent circuit of the following circuit:

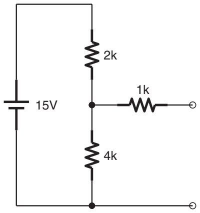

8. Suppose I have a circuit where when I add a load of 350 Ω, I get a 7 V drop and when I add a load of 2000 Ω, I get an 8 V drop. Calculate and draw the Thévenin equivalent circuit.

CHAPTER 27

Using Field Effect Transistors for Switching and Logic Applications

In Chapter 24, we discussed two kinds of transistors—bipolar junction transistors (BJTs) and field effect transistors (FETs). The previous chapters have focused primarily on BJTs because they are, in fact, great at various tasks that are commonly considered "amplification."

However, there are two amplification tasks for which BJTs are terribly suited—switching and logic circuits. Many people don't think of amplification when they think of logic circuits, but, indeed, they do require amplification. Think about the memory of a computer (which is a type of logic circuit). The memory of the computer has to continue to last after it has been written to. If there was no amplification, the signal would fade over time.[1] Therefore, in order to just "hold on" to the signal, amplification is required.

---

[1] Note that there are indeed some types of memory that can last significant amounts of time without amplification. I'm simplifying to help make the point.

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_27

While this can be done with BJTs, since BJTs are *current* amplifiers, they tend to utilize a lot of current to get this done. Field effect transistors (FETs) operate instead on voltage and require very little current to operate. Therefore, they are very efficient for logic-type applications.

27.1 Operation of a FET

A FET has a very similar set of parts to a BJT. There are three parts—a **gate**, a **source**, and a **drain**.[2]

The basic operation of the FET is that a *voltage* at the gate controls the *current* conduction between the drain and the source. The gate controls the conduction between the drain and the source. Like BJTs, FETs have fully off, fully on, and partially on states. However, unlike BJTs, even when "fully on," FETs still offer a resistance between the drain and the source. This is why, for current-focused applications, BJTs are often used—in their fully on state, they don't limit the current from the collector to the emitter.

Unlike the BJT, the FET is operated by voltages rather than current. Because of this, the gate essentially consumes almost no current whatsoever (it can have a Thévenin equivalence of hundreds of megaohms of resistance). So, while BJTs consume current to perform their operation, FETs consume almost none.

FETs can be built to perform in one of two operating modes:

- A **depletion mode** FET is normally "on" (allowing current to flow between the drain and the source), but turns off when a sufficient voltage (compared to the source) is applied at the gate. The voltage works to "deplete" the transistor of its conducting ability.

---

[2] Some FETs also have a **substrate** (also called a **body**), but that is beyond the scope of this book.

- An **enhancement mode** FET is normally "off" (blocking current flow between the drain and the source), but turns on when a sufficient voltage (compared to the source) is applied at the gate. The voltage works to "enhance" the transistor's ability to conduct.

FETs are distinguished as **N-channel** or **P-channel** based on their orientation. In an N-channel FET, the current flows from the drain to the source (i.e., the drain acts like the collector, and the source acts like the emitter). In a P-channel FET, the current flows from the source to the drain (i.e., the source acts like the collector, and the drain acts like the emitter). For an enhancement mode FET, the gate activates the channel when the gate voltage (compared to the source) is sufficiently positive if it is an N-channel and sufficiently negative if it is a P-channel. For a depletion mode FET, the gate closes the channel when the gate voltage (compared to the source) is sufficiently negative if it is an N-channel and sufficiently positive if it is a P-channel.[3]

So, for instance, in a P-channel depletion mode FET, the source will be positive, the drain will be negative, and the channel will close as the gate voltage becomes more and more negative compared to the source.

Additionally, most FETs have a "threshold voltage" that the gate must achieve before it switches on.

In addition to the channel and the mode, FETs are made in a variety of different ways, which leads to numerous different types of FETs—MOSFETs, JFETs, QFETs, MNOSs, and others. Each of these has its own distinct operating characteristics (threshold voltage, channel resistance, etc.). We will focus on one particular FET to make life easier—the N-channel enhancement mode MOSFET. This is the most commonly used FET by hobbyists.

---

[3] I know this is a lot of information, but there are so many terms that you are likely to encounter that I want to be sure you are familiar with what they mean on a practical basis. As you will see in the following, we are only going to actually use one kind of FET here.

27.2 The N-Channel Enhancement Mode MOSFET

The most popular general FET is probably the N-channel enhancement mode MOSFET. Because it is N-channel, the drain operates as the collector, and the source operates as the emitter (if we think of it in terms of our BJT terminology). That is backward from the way we normally think of things, but that is the reality of the naming scheme. Because it is an enhancement mode device, the bridge between the drain and the source is closed, unless a positive voltage (compared to the source) is applied at the gate.

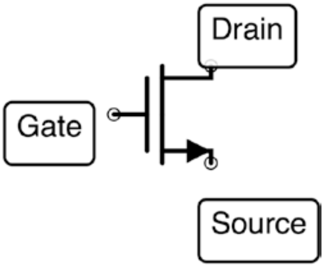

Figure 27-1. *The Schematic Symbol for an N-Channel MOSFET Transistor*

Figure 27-1 shows the schematic symbol for the N-channel MOSFET. Notice that the arrow is pointing out from the *source*. Again, that is because in this transistor, the source performs the same function as the emitter on a BJT.

27.3 Using a MOSFET

Figure 27-2 shows a schematic of a simplistic usage of a MOSFET. In this figure, the gate of the MOSFET is controlled by a switch. When the switch is on, the MOSFET conducts. However, the important thing to note is that the pathway into the gate of the MOSFET consumes basically no current whatsoever!

This is especially important when dealing with small microcontrollers. Some of them have extremely low tolerances for what currents they can deliver. By putting a MOSFET on the output pin, you can make the actual current output from the microcontroller effectively zero and simply have the current delivered through the MOSFET. Figure 27-3 shows this configuration. This is basically the same circuit as Figure 27-2 except that (a) a microcontroller is controlling the gate and (b) we added a pull-down resistor at the gate (note—Figure 27-2 also needs a pull-down resistor, but we left it off for simplicity). The pull-down resistor will make sure that even when the pin is not actively sending a signal, the gate is tied to a known voltage. This is especially important with MOSFETs since they do not consume very much current at all, so charge can stay on the circuit after disconnection.

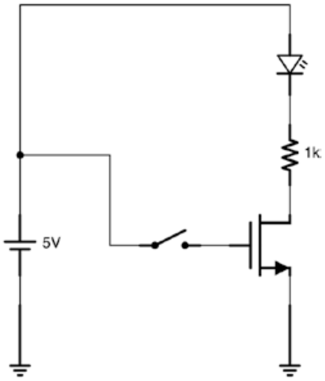

Figure 27-2. *A Simple MOSFET Circuit*

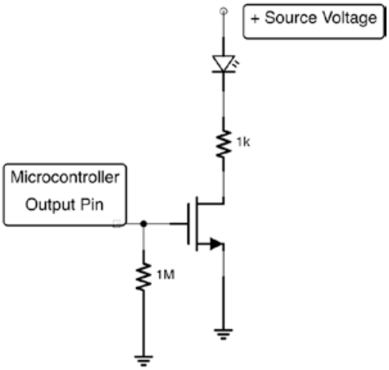
+ Source Voltage
1k
Microcontroller
Output Pin
1M

***Figure 27-3.** A Microcontroller Pin Controlling a MOSFET*

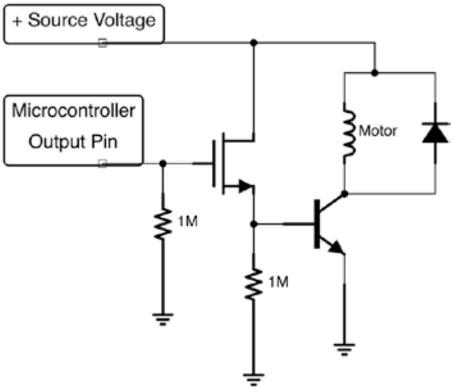
+ Source Voltage
Microcontroller
Output Pin
Motor
1M
1M

***Figure 27-4.** Combining a MOSFET with a BJT to Power a Motor*

One issue, however, is that, as we mentioned, MOSFETs always have resistance between the drain and source, limiting the amount of current that a MOSFET can source. If we wanted to power a motor, for instance, we would need to increase the current. We can do this by adding a BJT to the output of our MOSFET. Figure 27-4 shows this configuration. Notice that both transistors have a pull-down resistor attached. This prevents any accidental conduction of our devices.

In this configuration, when the gate to the MOSFET is pulled high, it will open up the connection from the drain to the source. Current will flow into the base of the BJT, which will allow current to flow from the collector to the emitter. The amount that it allows is based on the β of the transistor. This current is used by the motor.

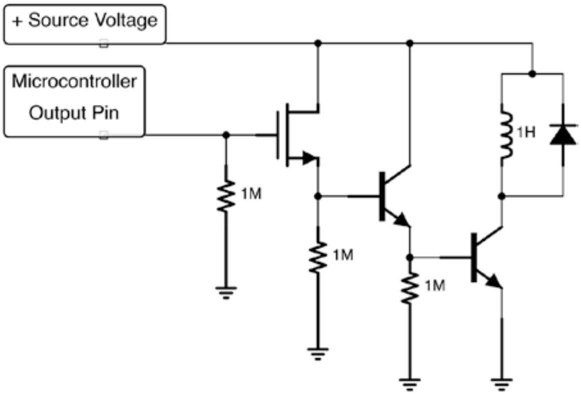

Figure 27-5. *A MOSFET with Two BJTs for Higher Current*

However, even this configuration may not provide enough current. The transistor can only amplify what it gets from the MOSFET, and the current output from the MOSFET is sometimes just not high enough for a single-stage amplification. Therefore, we can use the output from the first BJT as an input to the base of a second BJT to open up the power even more. Figure 27-5 shows this configuration. Note that each transistor has a pull-down resistor to make sure that it gets shut off when the pin goes low.

This configuration provides practically unlimited current through the collector of the transistor and doesn't utilize any current output from the microcontroller.

27.4 MOSFETs in Logic Circuits

MOSFETs are also great to use in logic circuits, again, because they use practically no current at all. While the pathway from the drain to the source can draw current, if the output coming from the source goes to *another* MOSFET, then it won't utilize any current either. As we mentioned earlier, we can think of the gate of a MOSFET as having a resistance on the order of a hundred megaohms. Therefore, Ohm's law will prevent any significant amount of current running into them.

In Chapter 12, we learned about integrated circuits that could perform logic functions such as AND, OR, and NOT. We can mimic these same logic functions with MOSFET transistors (we can also mimic them with BJTs, but we would waste a lot of current doing it).

Figures 27-6, 27-7, and 27-8 show how these gates could be wired using MOSFETs. Note that each transistor requires a pull-down resistor as well, which is not pictured for simplicity of understanding.

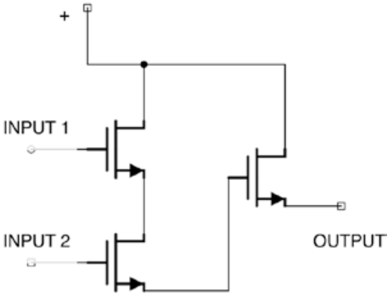

Figure 27-6. *An AND Gate Using MOSFETs*

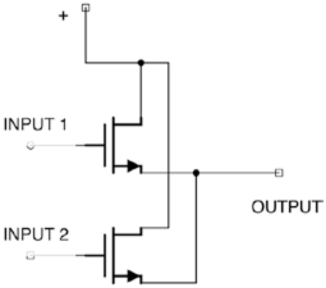

Figure 27-7. *An OR Gate Using MOSFETs*

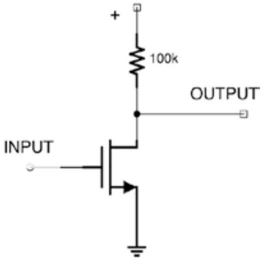

Figure 27-8. *A NOT Gate Using MOSFETs*

Review

In this chapter, we learned the following:

1. Field effect transistors (FETs) are similar to BJTs, but they are controlled by voltage instead of current.
2. The parts of the FET transistor are the gate, the source, and the drain, though whether the source or the drain is on the positive side depends on the type of transistor.

3. The N-channel depletion mode MOSFET is wired (somewhat confusingly) so that the drain is on the positive side and the source is on the negative side. Applying a positive voltage at the gate allows current to flow from the drain to the source.
4. Because FETs operate using voltage, the gate consumes almost no current, making it ideal for controlling devices from microcontrollers.
5. When a device is truly power-hungry, a FET can be used in combination with one or two BJTs to provide additional current output to the device.
6. FETs can also be used in combination with each other to produce logic circuits that utilize very little current.

Apply What You Have Learned

1. Why is it especially important to include pull-down resistors with FET transistors?
2. Which types of FETs are normally conducting between the drain and the source? Which ones are normally not conducting?
3. Create a circuit that combines an AND gate and an OR gate using N-channel MOSFETs. It should turn on an LED if buttons 1 and 2 are pressed or if button 3 is pressed.
4. Why is a BJT needed to power an electric motor in addition to the MOSFET?

CHAPTER 28

Going Further

Congratulations! You have reached the end of this book! Hopefully, though, this is just the beginning of your electronics journey. From here, there are many directions you can go.

If you are mostly into digital electronics (programming and using microcontrollers), then this book probably has all of the electronics knowledge you need. From here, you should learn more about programming and more about the devices you want to use for your projects. Additionally, there are a ton of add-on devices that can be controlled with these sorts of microcontrollers. Learning about how to program and what devices are available would be your best next step. A good start to this would be Michael McRobert's book *Beginning Arduino*, Brian Evans's book *Beginning Arduino Programming*, or Bob Dukish's *Coding the Arduino*.

If you were more interested in the sound and amplification sections, the next steps would be to either learn how to do this with microcontrollers or learn to do it directly with capacitors, inductors, and other electronic components. For the microcontroller path, try out Mike Cook's *Arduino Music and Audio Projects* or Alexandros Drymonitis's *Digital Electronics for Musicians*. For direct work in electronics, check out Craig Anderton's classic, *Electronic Projects for Musicians*, or, if you can find a copy, Nicholas Boscorelli's *The Stomp Box Cookbook*. Ray Wilson's *Make: Analog Synthesizers* is also good.

© Jonathan Bartlett 2020

J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5_28

If you want to learn more about electric motors and devices that move, there are a lot of options as well. David Cook's *Robot Building for Beginners* is a great place to start, as well as the follow-up book *Intermediate Robot Building*. Jeff Cicolani also has a book, *Beginning Robotics with Raspberry Pi and Arduino*, that is worth looking into. The book *Arduino Robotics* has a lot of more advanced ideas as well, including a DIY Segway clone.

If you want to know more about the operation of the components themselves, the best in-depth guidebook (though very expensive!) is Paul Horowitz's *The Art of Electronics*. If electronics is going to be your profession, it's worth every penny. Otherwise, it is probably a little too advanced (and expensive) for the casual hobbyist.

If you enjoyed the math and wanted to know more, I suggest that you start by learning calculus, as much of the math of electronics is based in some way on calculus. The book I would suggest for that is my own *Calculus from the Ground Up*.

Anyway, I wish you well on your electronics journey. You are well prepared to begin!

APPENDIX A

Glossary

AC See *alternating current.*

AC mains This is the type of current that is supplied to your house by the public utility companies. This is usually 120 volts AC and cycles back and forth 50–60 times per second.

AC signal This is the type of current usually picked up by a microphone or antenna. It has very low current and usually must be amplified before processing.

alternating current Current where the flow of current continually changes direction. This can also refer to a continual stream of fluctuations of voltage.

amp A shorthand way of saying ampere. See *ampere.*

ampere An ampere is a measurement of the movement of charge. It is equivalent to 1 coulomb of charge per second moving past a given point in a circuit.

amplification Taking a low-power signal and converting it to a high-power signal. This is usually done by using the low-power signal to control a high-power source and manipulate it to mimic the low-power signal.

AND function A logic function which yields a true value when all of its inputs are true.

AND gate A logic gate which implements the AND function.

anode An anode is the "positive side" of a polarized device. Note that, in batteries, the anode is actually the negative terminal (because the positive charge *enters* here).

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5

Arduino A popular environment for programming microcontrollers, including a standard set of development boards for a variety of microcontrollers.

Arduino IDE A software program used to write and upload programs for Arduino-based microcontrollers.

atomic number The number of protons that a given element has.

auto-ranging multimeter A multimeter which does not require that you know the likely value of the measured value ahead of time.

baud A way of expressing rates of communication. Means “bits per second.”

beta The amount of gain that a BJT can generate.

bit Binary digit. A true or false (or 1 or 0) value.

BJT Bipolar junction transistor. In this type of transistor, current at the base of the transistor controls the conduction between the collector and the emitter.

boost converter See *DC-DC power converter*.

breadboard A device which makes it easier to attach components together. See also *solderless breadboard*.

breakdown voltage The voltage drop that occurs when a diode is blocking current, but the voltage goes beyond the capacity of the diode itself and “breaks down” the blockade. This is problematic in normal diodes, but an expected part of operation of Zener diodes.

bridge A space in the middle of a breadboard separating terminal strips.

brushed electric motor A motor that operates by changing its interior magnetic configuration based on wires that are rotating in the shaft.

brushless electric motor A motor that operates by changing its exterior magnetic configuration based on generating an alternating current inside the motor. Brushless motors are less prone to fail because of fewer parts that wear down.

capacitance Capacitance is the ability to hold an electric charge.

capacitative reactance The reactance that comes from capacitor-like components.

carrier frequency A frequency on which another signal is laid.

cathode A cathode is the “negative side” of a polarized device. Note that, in batteries, the cathode is actually the positive terminal (because the positive charge *leaves* here).

charge Charge is a fundamental quantity in physics. A particle can be positively charged (like a proton), negatively charged (like an electron), or neutrally charged (like a neutron). Charge is measured in coulombs.

chip See *integrated circuit.*

circuit diagram The drawing of a circuit using special symbols to show how the components are connected without worrying about how the circuit will be physically laid out on the breadboard.

circuit pattern A circuit pattern is a common way of building (and understanding) certain types of circuits or subcircuits.

closed circuit A circuit is closed if there is a complete pathway from the positive to the negative.

CMOS CMOS is a technology for building integrated circuits (especially MOSFETs) and stands for “Complementary Metal Oxide Semiconductor.” It is also a standard for true/false voltage levels used by logic chips made using CMOS technology.

coil See *inductor*.

comment In computer programming, a comment is a line of code that the computer ignores and is meant to communicate information to the person reading the code.

common point See *ground*.

common emitter A transistor circuit pattern where the emitter is tied to a known, fixed voltage.

connection point A place on a breadboard where a wire or component can be attached.

conserved quantity A quantity is conserved if, in the normal course of action, it cannot be created or destroyed. Energy is a conserved quantity—it cannot be created or destroyed but only transferred.

conventional current flow Conventional current flow traces the flow of positive charge in a circuit from a positive source (often a battery) to either ground or a negative terminal.

coulomb A coulomb is a quantity of electric charge. One coulomb is roughly equivalent to the charge of 6.242Œ10^{1}8 protons. The same number of electrons produces a charge of −1 coulomb. Coulombs are represented by the symbol C.

coupling capacitor A coupling capacitor is a capacitor that links to circuits, usually at different DC offsets. The coupling capacitor transmits changes in voltage (i.e., AC) while blocking the DC offset voltages from influencing each other.

datasheet The documentation provided by a component or an integrated circuit which provides technical details about the component's operation.

debugging The process of finding errors in electronic circuits or computer code.

DC-DC power converter A device which converts voltage levels (like a transformer) but for DC circuits.

DC bias voltage See *DC offset*.

DC See *direct current*.

DC motor A device that converts DC electrical power into mechanical power.

DC offset If an alternating current is centered around a voltage other than zero, the center voltage is the DC offset.

depletion mode A FET which uses depletion mode means that the FET normally conducts between the drain and source, but will be inhibited by sufficient gate-to-source voltage (sufficiently positive for a P-channel or sufficiently negative for an N-channel).

development board A development board is a prebuilt circuit board that makes trying out certain chips and components (and building example projects with them) much easier.

digital logic A convention where voltages (usually a positive voltage and zero voltage) are used to represent "true" and "false" values (or "1" and "0").

diode An electric component that only allows current to flow in one direction.

DIP See *dual in-line package*.

direct current Current where the flow of current maintains the same direction. It can also be viewed as steady voltage from a positive terminal to a negative terminal (i.e., voltage which does not continually vary/change directions).

dissipation Dissipation refers to the ability to get rid of something. Heat dissipation deals with the removal of extra heat generated in circuits.

drain (transistor) One side of the main current pathway of a FET. This is the positive side in an N-channel FET and the negative side in a P-channel FET.

dual in-line package A type of package for integrated circuits which has two parallel sets of pins which fits nicely across the bridge of a breadboard.

electric field This is the force of attraction of opposite charges which are separated but near to each other.

electromagnet A magnet that is created through an inductor. Electromagnets allow electronics to manipulate the larger physical world.

electron current flow Electron current flow traces the flow of negative charges from a negative source or ground to its destination (often a positive terminal on a battery).

electron A negatively charged particle that is usually on the outside of an atom.

energy The ability to do work, usually measured in joules.

enhancement mode A FET which uses enhancement mode means that the FET normally does not conduct between the drain and source, but sufficient gate-to-source voltage will allow conduction (sufficiently positive if it is an N-channel and sufficiently negative if it is a P-channel).

equivalent resistance If multiple resistors are connected in series or in parallel, their operation can be mimicked by a single resistor. This resistor is said to have equivalent resistance of the resistor circuit that it replaces. See also *Thévenin equivalent circuit*.

farad The basic unit of capacitance named after Michael Faraday.

feedback When the output of a process feeds back into the input of a process.

FET Field effect transistor. This is a type of transistor where the gate controls the conduction between the drain and the source.

filtering capacitor A capacitor used to remove signals of a certain range of frequencies from the signal.

flash memory A type of memory in integrated circuits that is retained even after the power to the device is removed.

flux See *magnetic flux*.

flyback diode See *snubber diode*.

forward voltage drop The amount of voltage that a diode utilizes when the current flows forward.

frequency How often a cycle of an oscillator completes a full cycle, measured in hertz (cycles per second).

function In computer programming, a function is a block of code that accomplishes a task and may return a value to the code that invoked the function.

gain The amount of amplification that a component provides to a signal.

gate (logic circuit) See *logic gate*.

gate (transistor) In a FET, the gate is the input that controls the pathway between the drain and the source using voltage levels. Gates generally utilize a negligible amount of current.

ground The chosen point on a circuit by which to measure other voltages. On DC battery circuits, this is usually simply the negative terminal of the battery.

H-Bridge A device which makes it easier to control a two-way motor drive.

headers Places on a circuit board where additional circuits or wires can be attached.

heat Undirected energy. In electronics, heat is usually due to inefficiency in circuits. All circuits are inefficient to some degree, and that inefficiency is released as heat into the environment.

heatsink A heatsink is an object, usually large and metallic, which attaches to a part of a circuit to dissipate heat into the environment and away from the circuit itself.

henry The unit of measure for inductance.

hertz A unit of measurement of frequency/oscillation. Abbreviated Hz. Hertz refers to the number of cycles per second.

IC See *integrated circuit*.

in-system programmer A device that connects your computer to a microcontroller which will transmit a new program to the microcontroller.

inductance The amount of resistance an inductor offers to changes in current. Measured in Henries.

inductive kick When current is suddenly blocked to an inductor, the magnetic field suddenly collapses and generates a sudden large voltage at the cathode.

inductive reactance The reactance that comes from inductor-like components.

inductor A component, usually made of coiled wire, that stores energy in a magnetic field. Inductors resist changes in current.

impedance The amount of resistance a circuit offers to alternating current which combines resistance and reactance. Impedance is usually based on a specific frequency.

inefficiency In energy terms, inefficiency is the amount of energy that is wasted in performing a process. All processes are inefficient to some degree.

integrated circuit A miniaturized circuit that is built on a silicon chip.

International System of Units See *SI Units*.

inverted input or output An input or output is inverted if you send it (or receive from it) the opposite of what you might normally expect.

ISP See *in-system programmer*.

joule A standard unit of work or energy. It is defined as the amount of work performed when a 1-kilogram object is moved 1 meter. It can also be defined as the work performed when moving a coulomb of charge through 1 volt of electric potential difference.

junction The point where two or more wires or components are connected together.

Kirchhoff's current law At any junction, the total amount of current going into a junction is exactly the same as the total amount of current going out of that junction.

Kirchhoff's voltage law Given any two specific points on a circuit (at a particular point in time), no matter what path is traveled to get from one point to the other, the difference in voltage between the two points is the same. This can also be stated equivalently as saying that any complete, closed path around a circuit has a total voltage drop of zero.

latch A simple type of memory that stores a single bit using logic gates.

lead A metallic connection point of a component.

leg See *lead*.

logic function A function that takes in one or more logic (true/false) values as input and produces one or more (usually one) logic values as output.

logic gate A circuit which implements a logic function. This is different from a transistor gate. Confusingly logic gates are built from transistors which can have a gate.

loop In computer programming, a loop is a group of commands that is repeated multiple times.

magnetic flux The energy stored in an electric field.

mechanical fit The ability of two objects to be joined just by pushing them together.

memory The capacity of a circuit to hold on to previous values. Often used in computer programming to store values.

microchip See *integrated circuit*.

microcontroller An integrated circuit which operates as a standalone computer. Usually used for controlling small electronic devices.

milliamp A short way of saying milliampere. See *milliampere*.

milliampere One thousandth of an ampere. See *ampere*.

MOSFET A specific (and common) type of FET transistor.

multimeter A device used to measure a variety of electrical quantities in circuits and devices, usually including voltage, current, and resistance.

N-channel A type of FET. Refers to the physical construction of the FET.

NAND function A logic function which yields a false value only when all of its inputs are true. A combination of the AND and NOT functions.

NAND gate A logic gate which implements the NAND function.

NPN A type of BJT.

neutron An uncharged particle in the nucleus of an atom.

no load RPM The speed at which the shaft of a motor rotates with nothing attached to the shaft at its rated voltage.

no load current The amount of current that a motor utilizes with nothing attached to the shaft at its rated voltage.

non-polarized component A circuit component (such as a resistor) where it does not matter which end is used as the anode or cathode.

NOR function A logic function which yields false if any of its inputs (or combination of inputs) are true. A combination of the OR and NOT functions.

NOR gate A logic gate which implements the NOR function.

NOT function A logic function which takes one input and yields its opposite.

NOT gate A logic gate which implements the NOT function.

nucleus The nucleus is the part of the atom where protons and neutrons reside.

ohm The unit of measurement of resistance.

open circuit An open circuit is a condition where there is no electrical pathway for the current to flow. See also *closed circuit*, *short circuit*.

open collector A type of output from a circuit where the “positive” state is actually disconnected, while the “zero” state is an actual ground voltage. This is often used to allow the user of the circuit to supply their own voltage through a pull-up resistor.

OR function A logic function which yields a true value when any of its inputs are true.

OR gate A logic gate which implements the OR function.

oscillation Oscillate means going back and forth. In electronics, oscillation usually means periodic back-and-forth fluctuations in current or voltage.

oscillator An oscillator is a circuit that switches back and forth between two or more states (e.g., positive voltage to zero voltage) in a regular fashion.

oscilloscope A device used to probe circuits and visualize the behavior of circuits over time.

P-channel A type of FET. Refers to the physical construction of the FET.

package The way that an integrated circuit is encased, including the physical casing, the pinout, and the type of pins.

parallel circuit A circuit is a parallel circuit if one or more components are arranged into multiple branches.

period The length of a full cycle of an oscillating circuit. Usually measured in seconds.

pad See *pin*.

pin On an integrated circuit, this is one of the legs protruding from the chip that you can connect to your circuit. Can also refer to the leg/lead of other types of components.

pinout The pinout refers to the meaning/purpose assigned to each pin of a particular integrated circuit.

PNP A type of BJT.

polarized component A circuit component (such as a diode) which has an explicit anode and cathode (i.e., the sides of the component are not interchangeable).

power The continual supply of energy, usually measured in watts. In DC circuits, power is calculated as the voltage (in volts) multiplied by the current (in amperes).

power buses See *power rails*.

power rails The part of a breadboard (usually indicated by red and blue lines) which is usually used for power connections.

power transistor A transistor made to handle large amounts of voltage and current.

programming environment A set of tools which makes programming for certain devices, or in certain ways, easier.

protection diode A diode that protects a circuit, usually either from putting in a battery backward or, in the case of a snubber diode, from inductive kick.

proton A positively charged particle in the nucleus of an atom.

pull-up resistor A resistor connected to a positive voltage which gives a default positive voltage if no other voltage is connected on the line.

pull-down resistor A resistor connected to a positive voltage which gives a default zero voltage if no other voltage is connected on the line.

pulse-width modulation Sending a signal using a fixed pulse frequency, but where the percentage of time the signal is high vs. low varies. This is a way of outputting voltage levels on devices that don’t support outputting at different voltage levels.

PWM See *pulse-width modulation*.

quiescent current In a transistor amplifier, the quiescent current is the amount of current that is flowing into the base of the transistor when no signal is applied.

RC circuit A resistor in series with a capacitor, usually used for timing.

reactance This is a part of impedance which is similar to resistance but operates by preventing flow rather than dissipating energy. Reactance is usually based on the frequency of the signal.

resistance Resistance measures how much a component resists the flow of electricity. Resistance is measured in ohms (Ω).

resonant frequency The resonant frequency of a circuit is the frequency of the circuit where the inductive and capacitative reactances cancel each other out.

RMS A way of computing averages that focuses on magnitudes.

RPM Revolutions per minute. How often the shaft of a motor rotates in a minute.

RS-232 A protocol for serial communication describing voltage levels, timing, and other important signaling features for sending and receiving bits using two wires.

saturation mode This occurs in a transistor when it is fully on and cannot amplify the signal further. In saturation mode, the transistor acts more like a switch that is on rather than as an amplifier.

schematic See *circuit diagram*.

serial interface A type of input/output mechanism that sends data a single bit at a time, oftentimes using the RS-232 protocol.

series circuit A series circuit is a circuit or part of a circuit where all of the components are connected one after another.

servo motor A motor that, instead of spinning, creates an angle.

set screw A screw whose purpose is to apply pressure to another part to join it together.

shaft The rotating part of a motor.

short circuit A short circuit is what happens when the current pathway has no resistance from the positive to the negative.

SI Units The international standard for physical units of measure.

sink When a component or circuit receives incoming current.

snubber diode A diode that is wired backward across the terminals of an inductor to protect the circuit against inductive kick.

solder (noun) A substance which can be heated up to attach wires, components, and other metallic objects together.

solder (verb) The activity of attaching things together using solder.

solderless breadboard A device consisting of a number of connected terminals where components can be inserted and removed at will, which allows for the construction of circuits in a nonpermanent way.

source (general) When a component or circuit provides current output.

source (transistor) One side of the main current pathway of a FET. This is the negative side in an N-channel FET and the positive side in a P-channel FET.

square wave An oscillating signal where the signal spends almost all of its time either in the high or low state and almost no time in between.

SR latch A simple type of memory which can store a single bit.

stall condition In an electric motor, this is the condition where the motor is not turning while it is exhibiting its maximum torque.

stall current The amount of current that the motor uses in a stall condition.

stator The fixed outside container of a motor.

terminal See *lead*. Can also refer to a female receiver for a lead.

terminal strip On a breadboard, a terminal strip is a row of connection points which are connected together within the breadboard, so that putting in multiple components or wires on a terminal strip connects them together.

Thévenin equivalent circuit A generalization of how one circuit “sees” another circuit, as if it solely consisted of a power source and an impedance/resistance.

TO-220 package A way of packaging an integrated circuit, usually for transistors, voltage regulators, or other power-oriented devices. It consists of three terminals and a heatsink.

torque Rotational power. In a motor, this is the amount of rotational power that the motor applies to the shaft.

transformer A pair of inductors which utilize the magnetic field in one coil to generate a current in the other inductor. Usually used to convert voltage levels in different circuits and to isolate parts of a circuit. Transformers only operate on AC circuits. For DC circuits, see *DC-DC power converter*.

transistor A circuit component used for switching, logic, and amplification. Generally, a transistor operates by having one terminal control the flow of current between two others. See also *FET* and *BJT*.

TTL An acronym for transistor-to-transistor Logic. This is an old standard for logic chips for signaling true and false values, usually operating at 5 V.

unit of measurement A standard marker used for quantification, such as grams (weight), seconds (time), or meters (length).

unit prefix A unit prefix is a modifier that can be added to a word to indicate that it should be multiplied by some amount. For instance, a kilometer means a thousand meters. A millimeter means a thousandth of a meter.

variable In computer programming, a variable is a storage location where a value is held that can change.

variable resistor A resistor whose value changes.

voltage A voltage is the “power” of an electric charge. Formally, it is the ratio of potential energy of a charge to the magnitude of a charge. It can also be described as the change in magnetic flux over time.

voltage divider A voltage divider occurs when there are two components (usually resistors) connected together with a terminal coming out from between them meant to have a voltage between that of the two components.

voltage drop The voltage difference between two points on a circuit. This usually refers to the amount of voltage consumed within a specific component on a circuit.

watt A unit of power. A watt is the delivery of one joule per second.

work Work occurs when force causes a displacement (change of position). Work is often measured in joules.

XOR function A logic function which yields a true value when any one (and only one) of its inputs is true.

XOR gate A logic gate which implements the XOR function.

Zener diode A diode which is meant to be wired in such a way as to take advantage of its breakdown voltage.

APPENDIX B

Electronics Symbols

Symbol	Component	Description
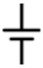	Battery	A battery is represented by a long line and a short line stacked on top of each other. Sometimes, there are two sets of long and short lines. The long line is the positive terminal, and the short line is the negative terminal (which is usually used as the ground).
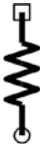	Resistor	A resistor is represented by a sharp, wavy line with wires coming out of each side.
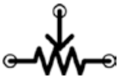	Potentiometer	A potentiometer is represented by a resistor with an arrow in the middle. This indicates that the positioning of the dial will affect how the voltage is divided.

(*continued*)

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5

Symbol	Component	Description
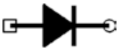	Diode	A diode is represented by an arrow with a line across it, indicating that current can flow from positive to negative in the direction of the arrow, but it is blocked going the other way.
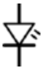	LED	An LED symbol looks similar to a diode (since it is a diode), but it also has two short lines coming out of it, representing the fact that it emits light.
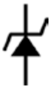	Zener diode	A Zener diode is represented like a normal diode, but with angled lines coming out of the blocking bar, representing the fact that the "blocking" action is not its only use.
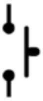	Button	A button is represented by a line that could connect two wires. This represents the fact that if the button is pushed down, it will connect the two parts of the circuit.
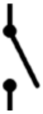	Switch	A switch is like a button, but it is anchored on one side. This represents the fact that the position of the switch is semipermanent. Unlike a button, which pushes and then releases, the switch will stay where you put it.
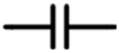	Capacitor (unpolarized)	A capacitor is represented by two parallel lines separated by a space, representing the fact that a capacitor is built from two plates that are not physically touching.

(*continued*)

Symbol	Component	Description
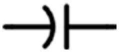	Capacitor (polarized)	A polarized capacitor is similar to an unpolarized capacitor, except that the negative side of the capacitor has a curve applied to it.
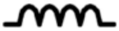	Inductor	An inductor is represented by a curling line to represent the fact that the wires in an inductor are coiled. This symbol is often used to represent not just basic inductors, but also applications of inductors such as motors and electromagnets.
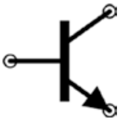	BJT NPN	The BJT NPN transistor is a three-terminal device with a line going in from the collector and a line going out to the emitter, with an arrow pointing toward the emitter. The line coming to the solid flat line is the base, which controls the flow. You can remember that this is an NPN transistor because the arrow "Never Points iN."
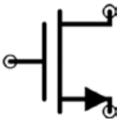	N-channel enhancement mode MOSFET transistor	The N-channel enhancement mode MOSFET transistor has a similar symbol to that of the BJT, but the gate doesn't actually touch the line connecting the drain to the source. This is because the gate of a FET does not actually participate in current flow, but only controls the gate itself with voltage.

APPENDIX C

Integrated Circuit Naming Conventions

The naming conventions for ICs can be bewildering at first. In truth, there is no official standard for chip names, but there are some conventions that are often followed. When a chip is invented, the company that invented it assigns it a part number. However, the courts have ruled that part numbers cannot be copyrighted. Therefore, if another manufacturer makes a similar or identical chip with the same pinout, they will often use the same part number.

C.1 Logic Chip Basic Conventions

Logic chips are often broken up into two families based on the voltage levels that they expect and produce. **TTL** (which stands for transistor-to-transistor logic) is an old standard for logic chips which usually operates at 5 V. TTL chips consider a signal to be "false" when it is below 0.8 V and "true" when it is above 2.2 V and can break if they receive voltages significantly higher than 5 V. Between 0.8 V and 2.2 V is a region where the resulting value is unpredictable. TTL originally referred to *how* the logic chips were constructed, but now it usually refers to the expected input/output levels of the chip.

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5

CMOS is a newer technology, and with it came a newer standard for how logic levels are interpreted. CMOS chips support a much wider supply voltage range than TTL, but their logic levels are a little different. For CMOS, false is from 0 V to one-third of the supply voltage (whatever it is—CMOS can usually operate from 3 V to 15 V), and true is from two-thirds of the supply voltage to the full supply voltage.

Chips are often made in a series—a whole set of chips which all perform complementary functions. The most common series of logic chips is the 7400 series originally designed by Texas Instruments. The 7400 series started as a set of TTL chips. Some common chips in this series are the 7400 itself (a quad–NAND gate), the 7408 (a quad–AND gate), and the 7432 (a quad–NOR gate).

Chips names will often have a prefix that relates to either their manufacturer or the company that originally designed them. As some examples, National Semiconductor chips are usually prefixed with LM; Texas Instruments chips have a whole slew of prefixes, including SN and TI; Motorola chips usually have an MC prefix; and Signetics chips usually have an NE prefix. There are many others, but this is just to give you an example. The 7400 series usually has part numbers starting with SN74 because TI built them. So a SN7408 is an AND gate based on designs by TI.

Now, the series is 7400. The last two digits refer to the function and pinout of the chips. That is, in the 7400 series, "08" will refer to quad–AND gates which all have the same pin configuration. However, sometimes they will insert letters in between "74" and "08." This usually refers to some modification to the electrical characteristics of the chip. For instance, a low-power version of the 7400 series has a "74LS" prefix. So the "SN74LS08" chip is a version of the 7408 (i.e., has the same pinout and function) that was originally designed by TI (the SN prefix) but is built for lower-power consumption (LS).

Then, part numbers often have suffixes as well. Suffixes often refer to some external characteristics of the chip. For instance, in Chapter 11, we mentioned different chip packages, such as DIP and SMD. These different packages will have different suffixes. For the 7400 series, the DIP package is usually suffixed with "P," so an SN74LS08P is the DIP version of the SN74LS08, and the NS74LS08NSR is an SMD version. You may also have suffixes which are based on temperature, hardiness, and even occasionally electrical output characteristics.

Sometimes, if a different manufacturer builds the same chip, they may change the manufacturer code and use different suffixes. For instance, Texas Instruments sells a SN74HC08P, which is a DIP 7408 which uses CMOS levels up to 6 V (that's what the HC is for). Essentially the same chip is available from Fairchild, which calls it the MM74HC08N. The MM prefix is one that Fairchild uses, and it is the same 74HC08 chip, and, for Fairchild, they use an "N" suffix to designate a DIP chip.

As I mentioned, these are only conventions, not standards, so they don't always apply. However, they can be helpful, so that you know that if someone specifies a 7432 chip and you see part numbers that say SN74LS32P, you might be able to determine that this at least has some relationship to the chip you are looking for.

APPENDIX D

More Math Than You Wanted to Know

This appendix is a catalog of equations in electronics and where they came from for those who are curious. This book is meant more for an introductory approach, but nonetheless many people are curious. This chapter isn't for the faint of heart, and it may involve lots of math you haven't taken. That's why it is stuck in an appendix.

However, if you are curious, these are the mathematical answers to your questions.

D.1 Basic Formulas

D.1.1 Charge and Current Quantities

- 1 coulomb = $6.241509 * 10^{18}$ electrons (how many electrons in a coulomb)
- $I = \frac{dc}{dt}$ (current is the derivative of charge with respect to time)
- $1A = 1\frac{C}{s}$ (A = ampere; C = coulomb; s = second)
- $3.6C = 1mAh$ (C = coulomb; mAh is milliamp-hour, a common unit for batteries)

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5

D.1.2 Volt Quantities

Volts are basically measures of energy per unit of charge. Volts are also known as electromotive force (EMF), or $\in$. Volts can be expressed in a number of ways:

- $V = \frac{J}{C}$ (J = joules; C = coulombs)
- $V = \frac{\text{potential energy}}{\text{charge}}$
- $V = \frac{N \cdot m}{C}$ (N = newtons; m = meters; C = coulombs)
- $V = \frac{kg \cdot m^2}{A \cdot s^3}$ (kg = kilograms; m = meters; A = amperes; s = seconds)
- $V = \frac{d\phi}{dt}$ (Faraday's law of induction—voltage is the derivative of the flux of the magnetic field with respect to time)

D.1.3 Resistance and Conductance Quantities

Resistance is in ohms. The inverse of resistance is conductance (the ability of current to flow through a wire) and is measured in siemens (S). The siemens unit is also called a mho (ohm spelled backward) and is sometimes marked by an upside-down omega (Ω):

- $G = \frac{1}{R}$ (G = conductance in siemens, R = resistance)
- $G = \frac{I}{V}$ (G = conductance; I is current; V is voltage)
- $R = \frac{V}{I}$ (Ohm's law)

Individual materials have a resistivity *(ρ)*.

$$R = \rho \cdot \frac{\text{length}}{\text{cross-sectional area}} \tag{D.1}$$

In other words, from beginning to end, resistance decreases with cross-sectional area and increases with length.

D.1.4 Ohm's Law

V is voltage (in volts), I is current (in amperes), and R is resistance (in ohms):

$$V = I \cdot R \tag{D.2}$$

D.1.5 Power

P is in watts. The following hold true for DC circuits. For AC circuits, they hold true if the resistance is actually an impedance:

- $P = V \cdot A$
- $P = I^2 R$
- $P = \frac{V^2}{R}$

D.1.6 Capacitance

Capacitance is the ability to store charge.

The fundamental equation for a capacitor is

$$Q = V \cdot C \tag{D.3}$$

Q is the amount of charge stored, V is the voltage across the terminals, and C is the capacitance in farads.

The derivative of this equation with respect to time is

$$\frac{dQ}{dt} = \frac{dV}{dt} \cdot C \tag{D.4}$$

Because current is the derivative of charge, we can then say

$$I = C\frac{dV}{dT} \tag{D.5}$$

The capacitance of capacitors is given by the equation

$$C = \in_r \in_0 \frac{A}{d} \tag{D.6}$$

Here C is capacitance, $\in_r$ is the dielectric constant of whatever separates the capacitor's plates, $\in_0$ is the dielectric constant of free space, A is the area of the plates in square meters, and d is the distance between the plates in meters.

D.1.7 Inductance

The fundamental equation for an inductor is

$$\phi = L \cdot I \tag{D.7}$$

Here, ϕ is the flux of the magnetic field in Webers, L is inductance in Henries, and I is current in amperes. The derivative gives you voltage:

$$\frac{d\phi}{dt} = L\frac{dI}{dt} \tag{D.8}$$

$$V = L\frac{dI}{dt} \tag{D.9}$$

In other words, the voltage produced is proportional to the change in current.

The inductance of a coil of wire can be calculated by

$$L = \frac{\mu \cdot N^2 \cdot A}{l} \tag{D.10}$$

Where N is the number of turns of wire, A is the area of the coil, l is the length of the coil, and μ depends on the core being used.

D.2 Semiconductors

Components made from silicon are known as semiconductors and have very useful nonlinear properties.

D.2.1 Diodes

Diodes do not have a fixed voltage drop as we assume in this book. It is an exponential function, but is steep enough to act like a fixed 0.6 V voltage drop for most purposes. The actual equation is

$$I = I_S\left(e^{\frac{V}{\eta V_T}} - 1\right) \tag{D.11}$$

I_S is the saturation current (depends on the construction of the diode), V is the voltage, η is either 1 for germanium or 2 for silicon, and V_T is known as the thermal voltage (the amount of voltage created just by particles moving around at a given temperature, usually about 0.026 V at room temperature).

D.2.2 NPN BJT

While we discussed general rules about BJTs, the technical model used to model them is known as the Ebers-Moll model. This model is much more complex to use, which is why we don't discuss it much in the chapter.

There are also several different Ebers-Moll models, depending on the level of detail required. The basic Ebers-Moll model for a conducting but unsaturated transistor is as follows:

$$I_E = I_S\left(e^{\frac{V_B E}{V_T}} - 1\right) \tag{D.12}$$

In this I_E is the emitter current (you can also use it for the collector current, since they are approximately equal). I_S is the saturation current of the base-emitter diode, and V_T is the thermal voltage, just like for diodes.

D.3 DC Motor Calculations

The voltage drop across a motor (V_m) is defined by the following equation:

$$V_m = V_b + R_m I_m$$

where V_b is the back EMF of the motor, R_m is the internal resistance of the motor's wiring, and I_m is the current flowing through the motor—so, basically, just Ohm's law plus the back EMF generated by the spinning of the motor.

So how much back EMF is created? We can determine that like this:

$$V_b = K_e \omega$$

In this equation, K_e varies by the motor and is usually given in volts per RPM or volts per radians per second. ω is merely the rotational speed in the units given.

Likewise, the torque generated can be determined from this equation:

$$T = K_T I_m$$

In this equation, K_T is the torque constant for the motor (be careful of the units), and I_m is the current going through the motor. Knowing the peak (stall) current, you can find the maximum torque available for the motor.

Interestingly, you can see that increasing the torque will actually affect, to some degree, the RPM of the motor. The full equation for the voltage across the motor is

$$V_m = K_e \omega + R_m I_m$$

The torque will affect I_m. That, in effect, will increase the voltage drop given by $R_m I_m$. Given a fixed voltage source, this will leave less voltage available for the $K_e \omega$ part of the equation. Since K_e is a constant, that means that ω will be reduced to some extent.

D.4 555 Timer Oscillator Frequency Equation

In Chapter 18, we learned to make oscillators using the 555 timer chip. In the actual chapter, I wanted you to focus on actually learning what was happening with the 555 timer rather than using a formula. However, there is a nice, simple formula that allows you to relate the resistor/capacitor network of the 555 timer to the final output frequency.

The formula is as follows:

$$f = \frac{1.44}{C(R_1 + 2R_2)} \tag{D.13}$$

In this equation, f is the frequency, R_1 is the resistor coming from the supply voltage, R_2 is the resistor next to the capacitor, and C is the timer capacitor.

To understand where this equation comes from, remember that frequency is just $\frac{1}{\text{period}}$. We can use time constant formulas to find the period and then just flip it to find the frequency.

If you recall, the period is just the total time it takes to complete a charge/discharge cycle. The 555 charges through *both* R_1 and R_2, but only discharges through R_2. Additionally, since it is just bouncing back and forth between $\frac{1}{3}$ and $\frac{2}{3}$ full, it only uses 0.693 time constant.

Therefore, we can have two formulas, one for the time charging and one for the time discharging:

$$\begin{aligned} T_{\text{CHARGE}} &= 0.693C(R_1 + R_2) \\ T_{\text{DISCHARGE}} &= 0.693R_2 \end{aligned}$$

The total period is just these two time periods added together. Therefore, you get

$$
\begin{aligned}
T_{\text{PERIOD}} &= 0.693C(R_1 + R_2) + 0.693CR_2 \\
&= 0.693C((R_1 + R_2) + R_2) && \text{Factoring out } 0.693C \\
&= 0.693C(R_1 + 2R_2) && \text{Regrouping}
\end{aligned}
$$

Since $f = \frac{1}{T_{\text{PERIOD}}}$, we can flip the preceding equation and get

$$
\begin{aligned}
f &= \frac{1}{0.693C(R_1 + 2R_2)} \\
&= \frac{1}{0.693}\frac{1}{C(R_1 + 2R_2)} && \text{Regrouping} \\
&\approx 1.44\frac{1}{C(R_1 + 2R_2)} \\
&= \frac{1.44}{C(R_1 + 2R_2)}
\end{aligned}
$$

At the end of the day, this is exactly what you did when you solved those problems; you just did it by hand instead of using a nice little formula. All a formula does is encapsulate the things that you normally do anyway, but simplifies it down to a set of predefined steps.

I have a love/hate relationship with formulas. Formulas are nice because they are easy to use. However, when you use them, it makes it easy to forget the basic facts behind them. The basic facts are more important than the formula, because you can rearrange the basic facts and develop all sorts of formulas depending on your needs. In fact, if you know the basic facts and you know how to make formulas, if you ever forget a formula, it is easy to determine one from the basic facts. Therefore, while memorizing formulas is important, knowing *why* formulas work is just as important, as it allows you to think more deeply and broadly and adapt your knowledge to new situations.

D.5 Output Gain Calculations in BJT Common Emitter Applications

In Chapter 25, we showed how to put together a multistage amplifier for sound coming in from a microphone.

However, we handwaved away a lot of the calculations needed for it. That's because they are, well, difficult.

Anyway, if you don't like handwaving, then this section is for you.

In Chapter 25, we said that, on a common emitter amplifier, the gain on the output current compared to the incoming current was determined by $\frac{R_{B2}R_C}{R_E R_O}$, where R_{B2} is the bottom resistor from the voltage divider that the signal is fed into, R_C is the collector resistor, R_E is the emitter resistor, and R_O is the output resistance (see Figure 25-4 for the diagram).

Recognize that any incoming current into the circuit from the input (as opposed to the quiescent current coming from the voltage source through R_{B1}) will have two alternative paths to take—through R_E or through R_{B2}. The relative amounts that do this are going to be based on the amount of voltage that R_E utilizes.

What we want to do is to analyze the gain between our *incoming* current (let's call that I_{IN}) and the current at the collector (I_C). That is our real gain from the incoming current.

The problem is knowing how much the actual amount of the incoming current (I_{IN}) will be going through R_{B2} or R_E. It will depend on the voltage in R_E. More voltage in R_E will mean that the current is likely to go down R_{B2} instead. This already tells us that the size of R_E will limit the amount of gain, as increased beta in the transistor will also increase the voltage of R_E which will also limit the amount of incoming current, which will have an equivalent effect of reducing gain.

Now, our circuit will primarily be working at a quiescent currentc, Q. We have already noted that this current will be the dominant signal. That means that the voltage between the base and the ground will be a combination of the diode voltage drop, the current in R_E from Q, and the current in R_E from the transistor's beta (βQ). Therefore, the total voltage drop from the base to the ground will be $V_{BG} = QR_E + \beta QR_E + 0.6$.

We can treat this as a "virtual resistance." The current coming in at I_{in} will be small comparatively, so we can reverse Ohm's law to see what this looks like as a resistance to the incoming current. Because $R = \frac{V}{I}$, we can say that the resistance here to new current is

$$R_{BG} = \frac{QR_E + \beta QR_E + 0.6}{Q} = R_E + \beta R_E + \frac{0.6}{Q} \approx R_E + \beta R_E \tag{D.14}$$

We dropped the $\frac{0.6}{Q}$ term because it would give us a very small comparative resistance, and we are trying to simplify our lives.

So now, when we have new current coming in at I_{IN}, it has two branches it can follow, each with its own resistance. Some of it will be wasted through R_{B2}, and some of it will be amplified through R_{BG}. The amount going through R_BG we will call I_U (for *usable* incoming current). The amount going through R_{B2} will just be $I_{IN} - I_U$.

In parallel resistances, the relative amount of current going through each resistor can be calculated as

$$\frac{I_1}{I_2} = \frac{R_2}{R_1} \tag{D.15}$$

You can see this from Ohm's law. Since $I_1 = \frac{V_1}{R_1}$ and $I_2 = \frac{V_2}{R_2}$, the ratio of the currents becomes $\frac{I_1}{I_2} = \frac{\frac{V_1}{R_1}}{\frac{V_2}{R_2}}$. Since $V_1 = V_2$ (the resistors are parallel and so have the same voltage drop), then this reduces to Equation D.15.

So, for this, I_{IN} will be divided between R_{B2} and R_{BG}. Therefore, we can say

$$\frac{I_U}{I_{\mathrm{IN}} - I_U} = \frac{R_{B2}}{R_{BG}} \tag{D.16}$$

$$I_U = I_{\mathrm{IN}} \frac{R_{B2}}{R_{BG}} - I_U \frac{R_{B2}}{R_{BG}} \tag{D.17}$$

$$I_U + I_U \frac{R_{B2}}{R_{BG}} = I_{\mathrm{IN}} \frac{R_{B2}}{R_{BG}} \tag{D.18}$$

$$I_U = \frac{I_{\mathrm{IN}} R_{B2}}{R_{BG}\left(1 + \frac{R_{B2}}{R_{BG}}\right)} \tag{D.19}$$

$$I_U = \frac{I_{\mathrm{IN}} R_{B2}}{R_{BG} + R_{B2}} \tag{D.20}$$

The amount of current through the collector that comes through the usable current from the signal (and not from the quiescent current, Q) we can denote as I_{CQ}. Therefore

$$I_{CQ} = \beta I_U \tag{D.21}$$

So the total amplification of our signal will be the ratio of the amplified output I_{CQ} to the incoming current I_{IN}:

$$\frac{I_{CQ}}{I_{\mathrm{IN}}} = \frac{\beta I_U}{I_{\mathrm{IN}}} \tag{D.22}$$

$$= \frac{\beta\left(\dfrac{I_{\mathrm{IN}} R_{B2}}{R_{BG} + R_{B2}}\right)}{I_{\mathrm{IN}}} \tag{D.23}$$

$$= \frac{\beta}{I_{\mathrm{IN}}} \frac{I_{\mathrm{IN}} R_{B2}}{R_{BG} + R_{B2}} \tag{D.24}$$

$$= \frac{\beta R_{B2}}{R_{BG} + R_{B2}} \tag{D.25}$$

$$\approx \frac{\beta R_{B2}}{R_E + \beta R_E + R_{B2}} \tag{D.26}$$

As you can see, this attenuates the action of β since it is in both the numerator and the denominator.

To see the extremes, if the β is low (i.e., single digits) and R_E is small compared to R_{B2}, this reduces to

$$\frac{I_{CQ}}{I_{\mathrm{IN}}} \approx \beta \frac{R_{B2}}{R_{B2}} = \beta \tag{D.27}$$

As β increases, the dominating terms become

$$\frac{I_{CQ}}{I_{\mathrm{IN}}} \approx \frac{\beta R_{B2}}{\beta R_E} = \frac{R_{B2}}{R_E} \tag{D.28}$$

We will call this particular gain γ (a lowercase Greek “g”). So therefore

$$\gamma = \frac{R_{B2}}{R_E} \tag{D.29}$$

Now, the actual output current (that goes through the capacitor to the next stage or to the speaker) is actually going to be determined largely by the resistance of the next stage!

What we have determined so far is the amount of current that goes through I_C from I_{IN}. The total amount of current in I_C is going to be $\beta(Q + I_U)$. However, Q is going to be essentially stable and therefore is not going to change the voltage on R_C, and capacitors only see voltage *changes*. Therefore, the only part of that which will be transmitted through the capacitor is the voltage change on R_C due to I_U, which will be βI_U or, in terms of our actual input signal, γI_{IN}.

Next, we have to figure out how much current we will actually be transmitting out of our amplification stage. This is the current that goes out through the coupling capacitor. Now, since the coupling capacitor is *only* transmitting changes, the only current that goes through that branch of the circuit is changing current. Therefore, the quiescent current does not travel through the capacitor.

So what we have to do is figure out how to estimate the amount of current coming into the capacitor. We have already designated I_{IN} as our input current, I_U as the usable current that is amplified at the base, and I_C as the current coming in to the collector.

Current coming out of R_C has two choices of where to go. It can either go out to the next part of the circuit through the capacitor, or it can go through the transistor. This will depend on the resistance of each route.

We need to start by calculating the quiescent resistance of the route that goes through the transistor and R_E. If V is our positive voltage, then, when the current goes through R_C, it will have dropped. Then it will drop again in the rest of the circuit.

The amount that the voltage drops going through R_C is $I_T R_C$ (because the total voltage goes through R_C). We can replace I_T with $I_O + I_C$, which means the voltage drop will be $(I_O + I_C)R_C$. The rest of the voltage drop we will call the "transistor path" voltage. It is simply

$$V_{\text{TPATH}} = V - \left(\left(I_O + I_C\right)R_C\right) \tag{D.30}$$

Ohm's law says that $R = \frac{V}{I}$, so the resistance on this path will be

$$R_{\text{TPATH}} = \frac{V - \left(\left(I_O + I_C\right)R_C\right)}{I_C} \tag{D.31}$$

Since $I_C = \beta Q$, then

$$R_{\text{TPATH}} = \frac{V - \left(\left(I_O + \beta Q\right)R_C\right)}{\beta Q} \tag{D.32}$$

The resistance of the other path will essentially be the Thévenin resistance of the output circuit (whatever lies beyond the capacitor). We will talk about what it should be shortly, but for now let's just call it R_O.

So that gives us two resistance values: R_O and R_{TPATH}. When we have a change in current (i.e., from I_{IN}), it will draw extra current, which will go down both paths. The amount that goes down each path will depend on the resistance of each path.

Because the current changes are small compared to the quiescent current, we can ignore the impact that these changes have on the resistances themselves (i.e., how the change in current affects R_{TPATH}). We will use I_T to represent the total *additional* current that comes across R_C due to the currents coming in at I_{IN}. I_O will be the amount that actually goes on out of the circuit.

As mentioned earlier, given two parallel resistances, R_1 and R_2, $\frac{I_1}{I_2} = \frac{R_2}{R_1}$. Therefore, we can say that the relative amount of current between the amount of current that goes out I_O and the amount of current that goes through the transistor (γI_{IN}) is

$$\frac{I_O}{\gamma I_{\mathrm{IN}}} = \frac{\dfrac{V - \left(\left(I_O + \beta Q\right)R_C\right)}{\beta Q}}{R_O} = \frac{V - \left(\left(I_O + \beta Q\right)R_C\right)}{\beta Q R_O} \tag{D.33}$$

We can then solve for I_O:

$$I_O \beta Q R_O = \left(V - \left(\left(I_O + \beta Q\right)R_C\right)\right)\gamma I_{\mathrm{IN}} \tag{D.34}$$

$$I_O \beta Q R_O = V\gamma I_{\mathrm{IN}} - \gamma I_{\mathrm{IN}} R_C I_O - \gamma I_{\mathrm{IN}} \beta Q R_C \tag{D.35}$$

$$I_O \beta Q R_O + \gamma I_{\mathrm{IN}} R_C I_O = V\gamma I_{\mathrm{IN}} - \gamma I_{\mathrm{IN}} \beta Q R_C \tag{D.36}$$

$$I_O\left(\beta Q R_O + \gamma I_{\mathrm{IN}} R_C\right) = I_{\mathrm{IN}}\gamma\left(V - \beta Q R_C\right) \tag{D.37}$$

$$I_O = \frac{I_{\mathrm{IN}}\gamma\left(V - \beta Q R_C\right)}{\beta Q R_O + \gamma I_{\mathrm{IN}} R_C} \tag{D.38}$$

Therefore, the complete gain (we will denote with a capital gamma, Γ) will be the ratio of I_{IN} (the actual input signal) to I_O (the actual output to the next stage). By using the preceding equivalency for I_O, we can determine

$$\Gamma = \frac{I_O}{I_{\mathrm{IN}}} = \frac{\dfrac{I_{\mathrm{IN}}\gamma\left(V - \beta Q R_C\right)}{\beta Q R_O + \gamma I_{\mathrm{IN}} R_C}}{I_{\mathrm{IN}}} \tag{D.39}$$

$$= \frac{I_{\mathrm{IN}}\gamma\left(V - \beta Q R_C\right)}{\left(\beta Q R_O + \gamma I_{\mathrm{IN}} R_C\right) I_{IN}} \tag{D.40}$$

$$= \frac{\gamma\left(V - \beta Q R_C\right)}{\beta Q R_O + \gamma I_{\mathrm{IN}} R_C} \tag{D.41}$$

Because we are using relatively low voltages, V is going to be essentially irrelevant and can be removed. Because I_{IN} is small and R_C is likely small compared to R_O, we can remove the term $\gamma I_{\mathrm{IN}} R_C$. Therefore, that leaves

$$\Gamma = \frac{I_O}{I_{\mathrm{IN}}} \approx \frac{-\gamma\beta Q R_C}{\beta Q R_O} \tag{D.42}$$

$$\approx -\gamma\frac{R_C}{R_O} \tag{D.43}$$

Since $\gamma = \frac{R_{B2}}{R_E}$, we can also write this as

$$\Gamma = \frac{I_O}{I_{\text{IN}}} \approx -\gamma \frac{R_C}{R_O} = -\frac{R_{B2} R_C}{R_E R_O} \tag{D.44}$$

Also note that this assumes that $R_O > R_C$ and $R_C > R_E$.

This is the total current gain from the input to the output. Notice that this value is negative. As we mentioned earlier, this kind of amplifier *inverts* the signal. Therefore, the gain is negative.

As you probably noticed, there were a lot of steps and a lot of simplifications. There were a lot of things where we handwaved away "insignificant" values and the sort. That's the way that electronics often works—you have to figure out which pieces really matter.

Now, thankfully, not everyone has to write these equations out because someone came in before them and deduced them. Personally, I think it's fun, but I recognize that isn't everyone's experience. Most electronics projects utilize more basic and straightforward applications of Ohm's law. But, occasionally, you have to pull out your whiteboard and go to town.

Of course, one thing I left out was how to determine R_{O}, the resistance of the circuit that follows.

If the output goes to another amplification stage, you may have to do more math to find the resistance of the next stage. The hard way is to look at the calculations earlier in this section for calculating R_{BG} and R_{B2} and calculate them for the next stage and treat them as parallel resistors. However, an easy way is to simply use R_{B2} to stand in for R_O. It's not exactly correct, but then that isn't what we are aiming for. We're just looking for the ballpark.

If the output goes to your headphones, then all you need to do is plug in the resistance of your headphones. Most headphones are less than 50Ω, with 16Ω being a fairly typical value.

D.6 The Thévenin Formula

In Chapter 26, we used two formulas which allowed us to calculate the Thévenin equivalent circuit for circuits experimentally. Equations 26.1 and 26.2 seem strange and complicated, but they are actually directly deducible from Ohm's law and the concept of an equivalent circuit.

The Thévenin Theorem states that any combinations of voltage sources and resistances can be replaced by a single voltage source and a single resistance. We will call this our Thévenin voltage source (V_T) and our Thévenin impedance (R_T). If we hook up a load (i.e., a fixed resistance) across the output terminals of this circuit, we will know the resistance that was added (because *we* added it), and we can measure the voltage drop across the resistor easily enough with a multimeter or oscilloscope.

We will need to measure this using two different loads because we have two unknowns—V_T and R_T. Using two different loads will give us two different equations using Ohm's law that will allow us to solve for two variables. We will call our lower-resistance load R_L, and the voltage drop across the R_L resistor will be V_L. Likewise, our higher-resistance load we will call R_H, and the voltage drop across it will be V_H. The current running through each of these loads (I_L and I_H) can be given by

$$V_L = I_L \cdot R_L$$
$$V_H = I_H \cdot R_H$$

That is just simply Ohm's law. We can also use Ohm's law to develop equations for the whole circuit, including the Thévenin equivalent voltage and impedance. Remember, because of the current rules, whatever current is flowing through our resistor must also be flowing in our Thévenin equivalent impedance. Therefore, the Thévenin equivalent voltage will be the current multiplied by the two impedances together. Therefore, this yields the following equations:

$$V_T = I_L\left(R_L + R_T\right)$$
$$V_T = I_H\left(R_H + R_T\right)$$

Both of these equations solve for V_T, given an unknown of R_T. We can also rearrange either of these to solve for R_T. Let's rearrange the first one to do that:

$$\begin{aligned} V_T &= I_L\left(R_L + R_T\right) && \text{Original equation} \\ \frac{V_T}{I_L} &= R_L + R_T && \text{Divide both sides} \\ \frac{V_T}{I_L} - R_L &= R_T && \text{Subtract } R_L \\ R_T &= \frac{V_T}{I_L} - R_L && \text{Solved for } R_T \end{aligned}$$

This is the same as Equation 26.2. However, it requires V_T to work. Now that we have an equation for R_T, we can substitute that back in and get an equation for V_T without using R_T. Using basic algebra manipulations, we can do the following:

$$\begin{aligned} V_T &= I_H\left(R_T + R_H\right) && \text{Original equation} \\ V_T &= I_H R_T + I_H R_H && \text{Distributive rule} \\ V_T &= I_H\left(\frac{V_T}{I_L} - R_L\right) + I_H R_H && \text{Substituting for } R_T \end{aligned}$$

$$V_T = I_H \frac{V_T}{I_L} - I_H R_L + I_H R_H \quad \text{Distributing}$$

$$V_T - I_H \frac{V_T}{I_L} = -I_H R_L + I_H R_H \quad \text{Get the} V_T \text{s together}$$

$$V_T \left(1 - \frac{I_H}{I_L}\right) = I_H \left(R_H - R_L\right) \quad \text{Factor both sides}$$

$$\frac{V_T}{I_L} = R_L + R_T \qquad \text{Divide both sides}$$

$$V_T = \frac{\frac{V_H}{R_H}\left(R_H - R_L\right)}{1 - \frac{V_H R_L}{R_H V_L}} \quad \text{Replace currents with Ohm's law equivalents} \left(\frac{V}{R}\right)$$

As you can see, this is Equation 26.1.

D.7 Electronics and Calculus

Calculus is a favorite subject of mine, and many parts of electronics make a lot of sense in the light of calculus.

D.7.1 Current and Voltage

First of all, recognize that electronics includes both static and dynamic quantities. Charge, for instance, is a static quantity. Current, though, is the *movement* of charge and thus is a dynamic quantity. Current entering or leaving a point can be written as a differential:

$$I = \frac{\mathrm{d}Q}{\mathrm{d}t} \tag{D.45}$$

Like current, voltage is a dynamic quantity, which is the change in the magnetic flux (ϕ):

$$V = \frac{\mathrm{d}\phi}{\mathrm{d}t} \tag{D.46}$$

D.7.2 Capacitors and Inductors

The static equation governing a capacitor is

$$Q = V \cdot C \tag{D.47}$$

where Q is the charge, V is the voltage, and C is the capacitance. Taking the derivative of both sides, this can be converted into a dynamic equation:

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = C\frac{\mathrm{d}V}{\mathrm{d}t} \tag{D.48}$$

Since $\frac{\mathrm{d}Q}{\mathrm{d}t} = I$, we can rewrite this as

$$I = C\frac{\mathrm{d}V}{\mathrm{d}t}. \tag{D.49}$$

What this means is that the current is proportional to the *change* in voltage.

The static equation governing an inductor is similar:

$$\phi = L \cdot I \tag{D.50}$$

Taking the derivative of both sides yields

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = L \cdot \frac{\mathrm{d}I}{\mathrm{d}t} \tag{D.51}$$

Since $V = \frac{\mathrm{d}\phi}{\mathrm{d}t}$, this can be rewritten as

$$V = L \cdot \frac{\mathrm{d}I}{\mathrm{d}t}. \tag{D.52}$$

In other words, on inductors, the voltage is proportional to the change in current.

D.7.3 Time Constants

If an ideal capacitor were connected to an ideal voltage source, then the capacitor would charge instantaneously. However, as described in Chapter 17, when charged through a resistor, it takes a certain amount of time to do the charging. The RC time constant combined with Figure 17-1 gives a simple way of figuring out these timings.

A more complete way of understanding this is with the following equation:

$$V_C = V_S\left(1 - e^{\frac{-t}{RC}}\right) \tag{D.53}$$

Here, V_C is the voltage across the capacitor, V_S is the voltage source, t is the time from the start of charging (in seconds), and R and C are the resistance and capacitance. e is Euler's constant. To find the amount of time, we can solve for t:

$$t = -RC\ln\left(1 - \frac{V_C}{V_S}\right) \tag{D.54}$$

Here, $\frac{V_C}{V_S}$ is the voltage percentage (expressed as a decimal). Therefore, the number of time constants is given by $-\ln\left(1 - \frac{V_C}{V_S}\right)$.

So how was Equation D.53 even determined? Since the voltage across the capacitor depends on the charge in the capacitor, then $V_C = \frac{Q}{C}$. Since the amount of current flowing through the resistor can be determined by the voltage and the voltage will be based on the difference between the source and the capacitor, then $I = \frac{V_S - V_C}{R}$. From the capacitor side, $I = C\frac{dV}{dt}$. Since the current had to come from somewhere, it had to come from the resistor. So that means these currents are equal. Therefore, $\frac{V_S - V_C}{R} = C\frac{dV_C}{dt}$.

We can solve this as a differential equation, by separating the variables. This gives $\frac{1}{R}dt = \frac{dV_C}{V_S - V_C}$. Since battery voltages are constant, V_S is a constant, so this integrates to $\frac{t}{R} + D = -C\ln\left(V_S - V_C\right)$, where D is the constant of integration.[1] Rearranging, this becomes $\frac{-t}{RC} + D = \ln\left(V_S - V_C\right)$. Exponentiating both sides gives $De^{\frac{-t}{RC}} = V_S - V_C$. At time $t = 0$, $D = VS - \text{VC}$. Since we are assuming that the starting capacitor voltage is zero, this means at that time, $D = V_S$. Now this becomes $V_S e^{\frac{-t}{RC}} = V_S - V_C$. Rearranging to solve for V_C yields $V_C = V_S - V_S e^{\frac{-t}{RC}}$. This can be simplified to yield Equation D.53.

---

[1] *D* undergoes lots of changes, and I'm not going to justify them all, but they are based on the fact that *D* is an unknown constant, and any manipulation of other constants will still yield an unknown constant.

APPENDIX E

Simplified Datasheets for Common Devices

Datasheets are what electronics manufacturers use to communicate the specifications of their components to the engineers who will use them in projects and products. However, datasheets tend to be horrendously complex. Typical datasheets run about 10–30 pages, and most of that is completely useless for someone just trying to get a circuit to work. Sometimes the first pages of the datasheet are all of the different packages the chip is available in, and the pages detailing what the component *actually does* are almost in the very back. I've even seen many sheets that never speak of what a device is for or why you would want to use it.

So, to simplify your life, I have created some simplified datasheets. All of these focus on the typical component styles used in solderless breadboards. These datasheets are also oriented toward *generic* parts. Each manufacturer has their own parts with their own specs and their own benefits and drawbacks. Some manufacturers may have higher or lower current specifications, faster or slower switching times, or other variations. These datasheets should be used as a starting point, but not as a final authority.

Use these datasheets to find the component you want and find its general characteristics, what each of the pins are for, and what they should be hooked up to. However, the manufacturer's datasheet should be consulted for final specifications.

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5

E.1 Batteries

E.1.1 Overview

A battery is a voltage source, meaning that it provides a fixed voltage over a wide range of currents.

Batteries have two terminals—positive and negative (or ground). The rated voltage for the battery is the typical voltage between these two terminals. However, a battery's actual voltage will vary quite a bit over the battery's lifetime.

Batteries also exhibit a small amount of *internal resistance* (also known as *equivalent series resistance*) which limits the amount of current they can provide.

E.1.2 Variations

Battery Type	**Typical Voltage**	**Voltage Range**	**Typical Capacity**
AAA	1.5	1.1–1.5	540 mAh
AA	1.5	1.1–1.5	1100 mAh
C	1.5	1.1–1.5	3800 mAh
D	1.5	1.1–1.5	8000 mAh
E(9 V)	9	7.2–9.6	1200 mAh
Coin cell (all sizes)	3	2–3.6	30–620 mAh

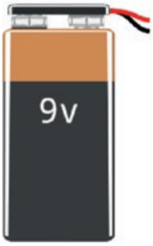

Battery

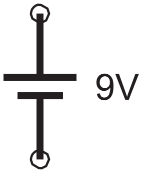

Schematic Symbol

E.1.3 Notes

Putting batteries together in parallel increases their available current but does not alter their voltage. Putting batteries together in series adds their voltages but does not change the available current.

Coin cell batteries are designated CRxxxx, where the first two digits of xxxx are the diameter in millimeters and the remaining digits tell the height in tenths of millimeters.

E.2 Resistors

E.2.1 Overview

Resistors are devices that resist the flow of current from one terminal to the other. The resistance of a resistor is measured in ohms.

The power (in watts) dissipated by a resistor is measured by the current flowing through it multiplied by the voltage drop across it. Resistors are rated for their maximum safe amount of power dissipation.

Resistors are often used to provide a safe limit to the amount of current flowing in a circuit or to provide a voltage drop for an input to another circuit.

E.2.2 Finding a Resistor Value

The colored bands on a resistor tell you how much resistance it has. However, these are very tiny, and sometimes it is easier just to measure it with a multimeter.

To interpret the colored bands, you first need to orient your resistor so that the last band is on the right (it is usually slightly thicker and set off from the others). This band is the *tolerance* band (use the tolerance column on the right).

The next band to the left is the *multiplier*. Use the multiplier column to read this value. The other bands are used as digits of a number which are then multiplied by the multiplier.

Resistor

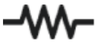

Schematic Symbol

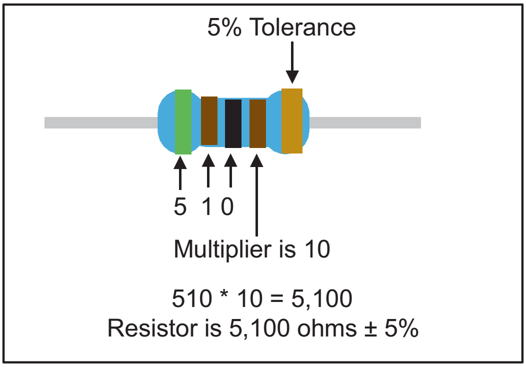

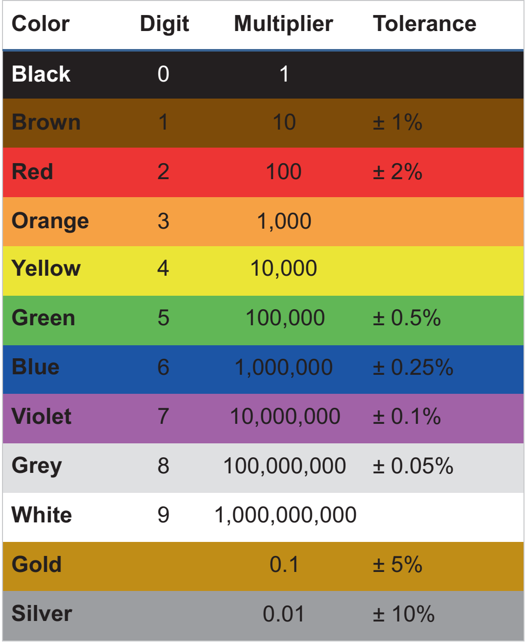

Color	Digit	Multiplier	Tolerance
Black	0	1	
Brown	1	10	± 1%
Red	2	100	± 2%
Orange	3	1,000	
Yellow	4	10,000	
Green	5	100,000	± 0.5%
Blue	6	1,000,000	± 0.25%
Violet	7	10,000,000	± 0.1%
Grey	8	100,000,000	± 0.05%
White	9	1,000,000,000	
Gold		0.1	± 5%
Silver		0.01	± 10%

E.3 Diodes

E.3.1 Overview

A diode is a device that only permits current to flow in one direction and blocks current in the other direction.

All diodes exhibit a *forward voltage drop*, which is the amount of voltage they consume when current is flowing. This is typically 0.6 V, though it varies a little based on current flow.

Diodes also have a *reverse breakdown voltage* which is the amount of voltage they can handle in the blocking direction before they start conducting anyway.

E.3.2 Variations

- An LED is a light-emitting diode. LEDs vary in their forward voltage drop based on their color (red=1.8; blue=3.3).
- A Zener diode is a diode that is built for a specific (usually lower than usual) reverse breakdown voltage. The reverse breakdown voltage of a Zener is more fixed/reliable than its forward voltage drop.
- A Schottky diode has a lower forward voltage drop (0.15–0.45V) and is able to turn on and off faster than ordinary diodes and accordingly also dissipates less heat.

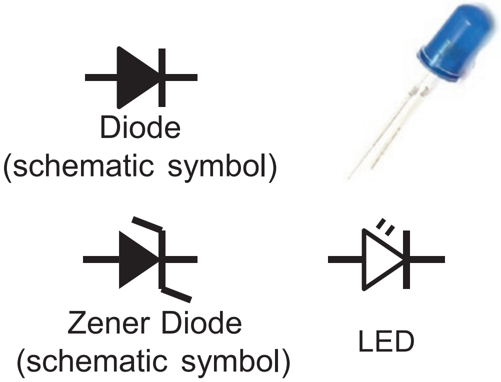

E.3.3 Forward Voltage Drop

Diodes are often considered to not have a resistance, but instead to have a simple voltage drop. That is, you apply the voltage drop and ignore resistance. That works for the math, but the reality is a bit different. Essentially, a diode is a variable resistor, where the resistance varies in order to keep the voltage drop essentially fixed. Therefore, diodes dissipate power just like other components, with the voltage drop multiplied by the current flowing through them.

E.3.4 Usages

- Prevent batteries from being inserted backward in a circuit
- Provide a fixed voltage drop between two points on a circuit, either through the forward voltage drop (regular diode) or the reverse breakdown voltage (Zener diode)
- Convert AC to DC by only allowing the positive flows through the circuit

E.4 Capacitors

E.4.1 Overview

A capacitor is a device that stores energy using a charged electric field. The size of the capacitor is proportional to the amount of charge that it can store. This size is measured in farads.

Changes in the voltage of one side of the capacitor will be transmitted through to the other side, but a steady-state DC voltage will not after the capacitor initially charges up.

This means that capacitors are ideal for sending AC signals through different DC bias levels, as merely the changes in voltage (the AC component) will be transmitted through the capacitor, while the bias voltages themselves will be blocked. A simplified description says that capacitors allow AC and block DC.

E.4.2 Variations

- Capacitors can be made of a variety of materials, which affect several of their properties.
- Capacitors can be non-polar (it doesn't matter which side is which) or polar (one side should always be more positive than the other).
- Ceramic disk capacitors are the classic example of non-polar capacitors, and electrolytic capacitors are the classic example of polar capacitors.
- Capacitors also vary in the amount of voltage that can be used with them.

Common Capacitor Types
(tantalum, electrolytic, ceramic disk)

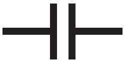

Schematic Symbol (nonpolar)

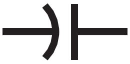

Schematic Symbol (polar)
(straight line is the positive side)

E.4.3 Finding Capacitance Values

Larger capacitors simply have their capacitance values printed on them. Smaller capacitors just have a set of digits printed on them, such as “104,” sometimes with a letter after them. These are a little harder to interpret.

On these capacitors, the first digits should be read as is, and the last digit is the number of additional zeroes that should be tacked onto the end. This value is then interpreted as picofarads.

So the value “104” means we take “10” and then add “4” zeroes, giving us “100000.” This is the capacitance in picofarads. This is the same as 100 nanofarads. If there is any code starting with a letter, this is a tolerance marking.

E.5 Inductors

E.5.1 Overview

An inductor is a device that stores energy in its magnetic field. The size of the inductor (in Henries) is proportional to the size of the magnetic flux it can create for a given level of current.

Inductors prevent changes in current by varying their electric field. Reducing the current through an inductor causes a partial collapse of the field, creating a voltage on the other side, which feeds current.

A simplified way of understanding inductors is to think of them as allowing DC but blocking AC.

Inductors are generally constructed by winding wire around an iron core material.

E.5.2 Uses

- Inductors serve as the main gateway between electrical and mechanical actions through the magnetic field.
- Inductors are used to open and close valves and switches magnetically.
- Inductors are used to drive electric motors.
- Inductors are used to limit certain frequencies in circuits.
- Inductors are used in speakers, where changes in magnetism cause the speakers to move.
- Inductors can be used as transformers to convert power from one voltage/current combination to another of equal power.

Various styles of inductors

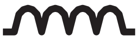

Schematic Symbol

E.5.3 Inductive Kick

Inductors use current to maintain their magnetic fields. When the current decreases suddenly, the magnetic field is converted into a large voltage, creating what is known as an inductive kick.

This kick can be damaging to electronic components. To mitigate this, a diode can be placed backward across the inductor so that, under normal circumstances, it does not conduct, but when an inductor produces a voltage spike, it re-routes the current back through the inductor where it can dissipate slowly in a loop rather than building up and damaging other components with a large voltage spike.

E.5.4 Inductor Color Codes

Simple inductors (i.e., those in packaged casings) are marked identically to resistors, except that the resulting value is in *micro*henries rather than just Henries.

E.6 NPN BJTs

E.6.1 Overview

Transistors come in a variety of configurations. BJTs (bipolar junction transistors) are *current-amplifying* devices.

In an NPN (negative-positive-negative) transistor, small changes to the current at the base result in large changes in current coming into the collector.

In a transistor, the voltage at the emitter is one diode drop (0.6 V) below the base, and if it is not, the transistor does not conduct. The transistor's beta is the multiplier between the base current and the collector current. The collector's voltage must be above the base voltage, or the transistor is saturated (acts as a short circuit from collector to emitter).

If these conditions are met, the transistor acts as a variable resistor which keeps the collector current as an amplification of the base current.

E.6.2 Variations

- A PNP transistor is the opposite of an NPN—the current applied at the base *reduces* the flow of current into the collector.
- A FET (field effect transistor) operates using voltages rather than current. This allows for negligible current usage on the inputs, but also reduced gain and more complicated support considerations.

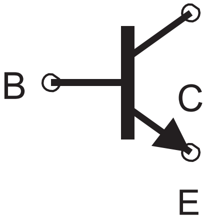

Schematic Symbol

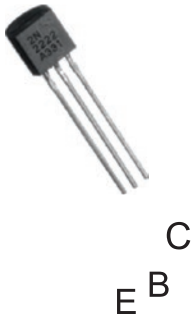

Transistor

E.6.3 Pin Configuration

- Base (B): This is the terminal that contains the current to be amplified.
- Collector (C): This is the terminal where the amplified current will flow into.

- Emitter (E): This is the terminal where the currents from both the base and the collector will flow out of, at a voltage level of one diode drop (0.6 V) less than the base.

E.6.4 Design Considerations

Transistor designs are based on which terminals the signal itself passes through (the third terminal is referred to as the “common” terminal).

In common collector configurations (also called emitter followers), the signal passes from base to emitter with an increased current but no increased voltage.

In common emitter configurations, the signal passes from the base to the collector, with the increase in current converted to a voltage through a collector on the resistor.

E.7 YwRobot Power Module

E.7.1 Overview

This device allows you to supply power to your breadboard projects from a variety of sources through its barrel jack. The module down-steps the voltage to 5 V or 3.3 V (selectable with jumpers).

The power module comes with an on/off switch and a set of header pins that can be used to wire power to other places.

The module is made to fit on a standard breadboard, where the output pins align directly onto the power rails of the breadboard.

Be sure to align the positive and negative markings on the module with their matching power rails!

Note that this part is not technically part of any schematic and simply acts as a voltage source within a schematic.

E.7.2 Variations

- On some boards, the USB jack can be used as a power input, and on some it is a power output only.
- The popularity of this module has prompted a number of manufacturers to build similar devices with a variety of shapes and input/output methods.

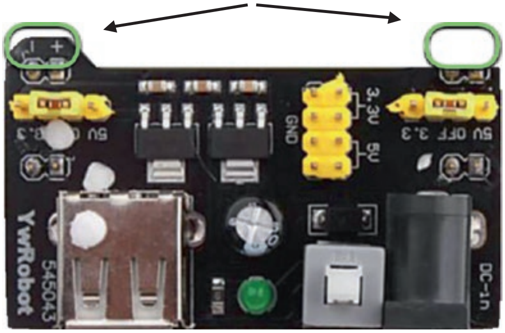

Align these with the +/- power rails on your breadboard

E.7.3 Pin Configuration

The power module connects directly to your breadboard. Each side can be independently selected for 3.3 V or 5 V power via jumpers.

In the middle is a set of male headers for 3.3 V, 5 V, and ground.

E.7.4 Limitations

- Minimum input voltage: 6.5 V (DC)
- Maximum input voltage: 12 V (DC)

- Output voltage: 3.3/5 V (selectable)
- Maximum output current: 700 mA
- Barrel jack plug size: 5.5 mm × 2.1 mm

E.8 555 Timer

E.8.1 Overview

The 555 timer is a collection of components that can be configured to provide timings and oscillations.

It uses two voltage levels—one-third supply voltage and two-thirds supply voltage. Internally, it consists of

- Two comparators (one for each voltage level)
- A flip-flop (single-bit storage) to know what state it is in and to switch states at the appropriate time
- An output driver
- A reset input

The timer relies on external circuitry (such as an RC time circuit) to supply timings.

The timer effectively has two states. In the “charging” state, the **Discharge** pin is disconnected, and the **Threshold** pin is waiting for a high (2/3) voltage. In the “discharging” state, the **Discharge** pin is connected to ground, and the **Trigger** pin is waiting for a low (1/3) voltage. The typical usage is to provide an oscillating circuit.

E.8.2 Variations

- Can be implemented using CMOS/FETs or BJTs. FET implementation consumes less power, but can source less output.
- Many variations in maximum oscillation frequency.

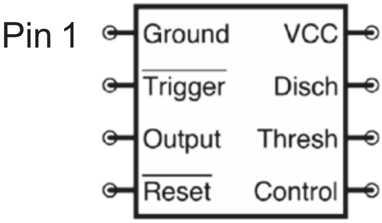

E.8.3 Pin Configuration

- **Trigger** and **Threshold** detect going below 1/3 and above 2/3 voltage, respectively.
- **Discharge** provides a ground that is only attached when the chip is in the discharging state.
- **Output** supplies a high voltage when the chip is in the charging state and a low voltage when it is in the discharging state.
- **Reset** should be normally tied to a positive supply—it resets the circuit when it goes low.
- **Control** is normally connected to ground with a capacitor (10 μF recommended).

E.8.4 Specifications

- Supply Voltage: Usually 2–15 V
- Output Current: 100–200 mA

E.8.5 Implementation Example

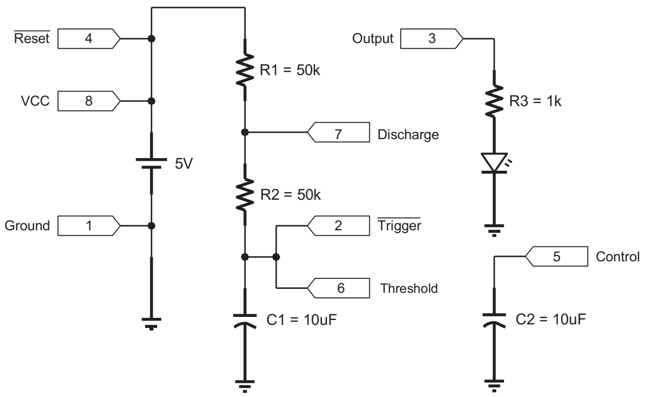
Reset
4
VCC
8
5V
Ground
1
R1 = 50k
7
Discharge
R2 = 50k
2
Trigger
6
Threshold
C1 = 10uF
Output
3
R3 = 1k
5
Control
C2 = 10uF

E.9 LM393 and LM339 Voltage Comparator

E.9.1 Overview

The LM393 is a dual voltage comparator. Each channel has two inputs—designated IN+ and IN-. When the voltage at IN+ is greater than the voltage at IN-, the output (OUT) is positive (actually, it is disconnected; see more later). Otherwise, the output is connected to ground.

The output is called an "open collector" output, which means that the "positive" state has no output current and is essentially disconnected. However, in the negative state, the output is connected to ground. This means that to use the output, you need to provide your own positive voltage through a pull-up resistor (this allows you to set your own output voltage).

The LM339 is identical except that it has four channels instead of two. Note that the schematic symbol for this is the same as that of an op-amp, largely because they perform similar (though not identical) functions.

E.9.2 Variations

Various chips differ in the amount of current they can sink, how fast they respond to changes in input voltage, how smooth the transition is from one state to the other, and the minimum amount of difference required to trigger.

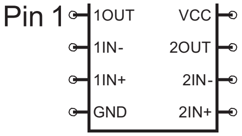
Pin 1
1OUT
1IN-
1IN+
GND
VCC
2OUT
2IN-
2IN+

LM393

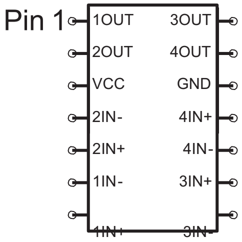
Pin 1
1OUT
2OUT
VCC
2IN-
2IN+
1IN-
1IN+
3OUT
4OUT
GND
4IN+
4IN-
3IN+
3IN-

LM339

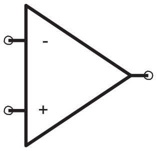

Schematic Symbol
(per channel)

E.9.3 Specifications

- Supply Voltage: 2–36 V
- Input Voltage: -0–1.5 V less than supply voltage
- Maximum Sink Current: 20 mA (when output is grounded)

- Input Impedance: High (inputs use very negligible amounts of current)
- Quiescent Chip Current: 1 mA

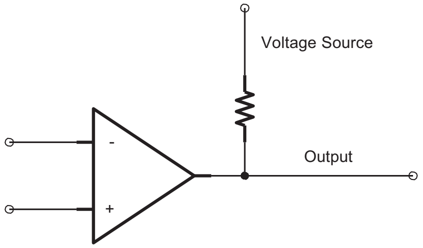

Usage of the Comparator with a pull-up resistor

E.10 CD4081 and 7408 Quad-AND Gate

E.10.1 Overview

The quad-AND gate is a set of four AND logic gates on a single chip. The output (Y) will be high only if *both* inputs (A and B) are high.

The required voltages for low and high on input and the guaranteed voltages for low and high on output are listed in the specifications section.

For the output to be high, both A and B must be high. Otherwise, the output (Y) will be low.

The 7408 and the CD4081 are the TTL and CMOS versions of the chip, respectively.

E.10.2 Variations

- 74HC08: Pin compatible with the 7408; voltage and current characteristics of the CD4081

- 74HCT08: Pin compatible and voltage-level compatible with 7408; similar current usage as the CD4081
- 74LS08: Fast-switching version of the 7408

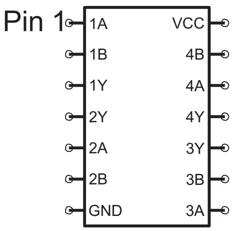
Pin 1
1A
1B
1Y
2Y
2A
2B
GND
VCC
4B
4A
4Y
3Y
3B
3A

CD4081

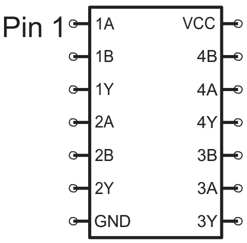
Pin 1
1A
1B
1Y
2A
2B
2Y
GND
VCC
4B
4A
4Y
3B
3A
3Y

7408

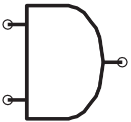

Schematic Symbol
(per channel)

E.10.3 Specifications (CD4081)

- Supply Voltage: 3–15 V
- Input (High) Voltage: > 2/3 supply
- Output (High) Voltage: Supply – 0.05
- Input (Low) Voltage: < 1/3 supply
- Output (Low) Voltage: 0–0.5 V
- Maximum Output Current: ~5 mA

E.10.4 Specifications (7408)

- Supply Voltage: 5 V
- Input (High) Voltage: > 2 V
- Output (High) Voltage: > 2.7 V
- Input (Low) Voltage: < 0.8 V
- Output (Low) Voltage: < 0.4 V
- Maximum Output Current: 100 mA

E.11 CD4071 and 7432 Quad-OR Gate

E.11.1 Overview

The quad-OR gate is a set of four OR logic gates on a single chip. The output (Y) will be high if *either or both* inputs (A and B) are high.

The required voltages for low and high on input and the guaranteed voltages for low and high on output are listed in the specifications section.

For the output to be high, either A or B (or both) must be high. Otherwise, the output (Y) will be low.

The 7432 and the CD4071 are the TTL and CMOS versions of the chip, respectively.

E.11.2 Variations

- 74HC32: Pin compatible with the 7432; voltage and current characteristics of the CD4071
- 74HCT32: Pin compatible and voltage-level compatible with 7432; similar current usage as the CD4071

- 74LS32: Fast-switching version of the 7432

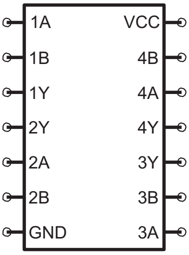
1A
VCC
1B
4B
1Y
4A
2Y
4Y
2A
3Y
2B
3B
GND
3A

CD4071

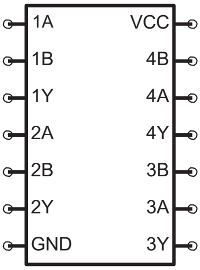
1A
VCC
1B
4B
1Y
4A
2A
4Y
2B
3B
2Y
3A
GND
3Y

7432

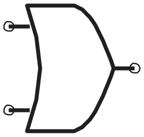

Schematic Symbol
(per channel)

E.11.3 Specifications (CD4071)

- Supply Voltage: 3–15 V
- Input (High) Voltage: > 2/3 supply
- Output (High) Voltage: Supply – 0.05
- Input (Low) Voltage: < 1/3 supply
- Output (Low) Voltage: 0–0.5 V
- Maximum Output Current: ~5 mA

E.11.4 Specifications (7432)

- Supply Voltage: 5 V
- Input (High) Voltage: > 2 V

- Output (High) Voltage: > 2.7 V
- Input (Low) Voltage: < 0.8 V
- Output (Low) Voltage: < 0.4 V
- Maximum Output Current: 100 mA

E.12 CD4001 and 7402 Quad-NOR Gate

E.12.1 Overview

The quad-NOR gate is a set of four NOR logic gates on a single chip. The output (Y) will be high if *both* inputs (A and B) are high or if *both* inputs are low.

The required voltages for low and high on input and the guaranteed voltages for low and high on output are listed in the specifications section.

For the output to be high, either both A and B must be high or neither must be high. Otherwise, the output (Y) will be low.

The 7402 and the CD4001 are the TTL and CMOS versions of the chip, respectively.

E.12.2 Variations

- 74HC02: Pin compatible with the 7402; voltage and current characteristics of the CD4001
- 74HCT02: Pin compatible and voltage-level compatible with 7402; similar current usage as the CD4001
- 74LS02: Fast-switching version of the 7402

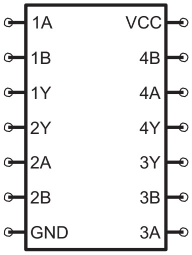
1A
1B
1Y
2Y
2A
2B
GND
VCC
4B
4A
4Y
3Y
3B
3A

CD4001

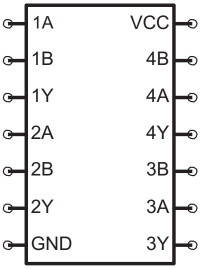
1A
1B
1Y
2A
2B
2Y
GND
VCC
4B
4A
4Y
3B
3A
3Y

7402

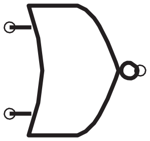

Schematic Symbol
(per channel)

E.12.3 Specifications (CD4001)

- Supply Voltage: 3–15 V
- Input (High) Voltage: > 2/3 supply
- Output (High) Voltage: Supply – 0.05
- Input (Low) Voltage: < 1/3 supply
- Output (Low) Voltage: 0–0.5 V
- Maximum Output Current: ~5 mA

E.12.4 Specifications (7402)

- Supply Voltage: 5 V
- Input (High) Voltage: > 2 V

- Output (High) Voltage: > 2.7 V
- Input (Low) Voltage: < 0.8 V
- Output (Low) Voltage: < 0.4 V
- Maximum Output Current: 100 mA

E.13 CD4011 and 7400 Quad-NAND Gate

E.13.1 Overview

The quad-NAND gate is a set of four NAND logic gates on a single chip. The output (Y) will be low unless *both* inputs (A and B) are high.

The required voltages for low and high on input and the guaranteed voltages for low and high on output are listed in the specifications section.

For the output to be high, A and B can be anything as long as they are not both high. Otherwise, the output (Y) will be low.

The 7400 and the CD4011 are the TTL and CMOS versions of the chip, respectively.

E.13.2 Variations

- 74HC00: Pin compatible with the 7400; voltage and current characteristics of the CD4011
- 74HCT00: Pin compatible and voltage-level compatible with 7400; similar current usage as the CD4011
- 74LS00: Fast-switching version of the 7400

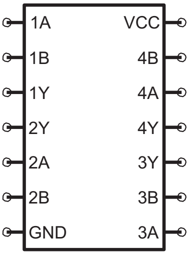
1A
1B
1Y
2Y
2A
2B
GND
VCC
4B
4A
4Y
3Y
3B
3A

CD4011

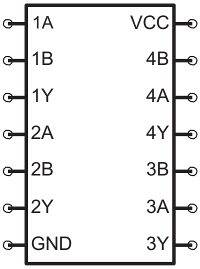
1A
1B
1Y
2A
2B
2Y
GND
VCC
4B
4A
4Y
3B
3A
3Y

7400

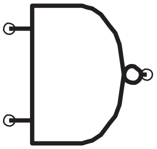

Schematic Symbol
(per channel)

E.13.3 Specifications (CD4011)

- Supply Voltage: 3–15 V
- Input (High) Voltage: > 2/3 supply
- Output (High) Voltage: Supply – 0.05
- Input (Low) Voltage: < 1/3 supply
- Output (Low) Voltage: 0–0.5 V
- Maximum Output Current: ~5 mA

E.13.4 Specifications (7400)

- Supply Voltage: 5 V
- Input (High) Voltage: > 2 V

- Output (High) Voltage: > 2.7 V
- Input (Low) Voltage: < 0.8 V
- Output (Low) Voltage: < 0.4 V
- Maximum Output Current: 100 mA

E.14 CD4070 and 7486 Quad-XOR Gate

E.14.1 Overview

The quad-XOR (exclusive OR) gate is a set of four XOR logic gates on a single chip. The output (Y) will be high if either input (A or B) *but not both* is high.

The required voltages for low and high on input and the guaranteed voltages for low and high on output are listed in the specifications section.

For the output to be high, either A or B must be high but not both of them. Otherwise, the output (Y) will be low.

The 7486 and the CD4070 are the TTL and CMOS versions of the chip, respectively.

E.14.2 Variations

- 74HC86: Pin compatible with the 7486; voltage and current characteristics of the CD4070
- 74HCT86: Pin compatible and voltage-level compatible with 7486; similar current usage as the CD4070
- 74LS86: Fast-switching version of the 7486

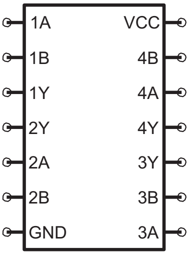
1A
1B
1Y
2Y
2A
2B
GND
VCC
4B
4A
4Y
3Y
3B
3A

CD4011

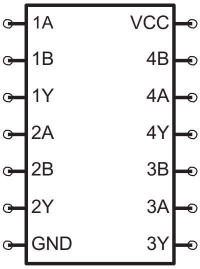
1A
1B
1Y
2A
2B
2Y
GND
VCC
4B
4A
4Y
3B
3A
3Y

7400

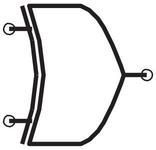

Schematic Symbol
(per channel)

E.14.3 Specifications (CD4070)

- Supply Voltage: 3–15 V
- Input (High) Voltage: > 2/3 supply
- Output (High) Voltage: Supply – 0.05
- Input (Low) Voltage: < 1/3 supply
- Output (Low) Voltage: 0–0.5 V
- Maximum Output Current: ~5 mA

E.14.4 Specifications (7486)

- Supply Voltage: 5 V
- Input (High) Voltage: > 2 V

- Output (High) Voltage: > 2.7 V
- Input (Low) Voltage: < 0.8 V
- Output (Low) Voltage: < 0.4 V
- Maximum Output Current: 100 mA

E.15 LM78xx Voltage Regulator

E.15.1 Overview

The LM78xx voltage regulator is actually a series of chips to provide a consistent voltage output from a variety of voltage inputs. Each chip is named with the number of volts it supplies in its output. For instance, the LM7805 outputs a constant 5 V, and the LM7812 outputs a constant 12 V.

These chips are linear voltage regulators, which means that they regulate voltage by dissipating excess power as heat. If significant heat develops, the LM78xx can have a heatsink attached to the back plate, which also serves as a second ground.

The LM78xx requires an input voltage at least 2.5 V above the regulated voltage. This is known as the “drop-out” voltage of the chip.

E.15.2 Variations

- The 78xxSR is a line of switching regulators, meaning that they do not dissipate significant power when regulating (they operate by turning the power on and off quickly rather than dissipating excess power). They waste significantly less current but do have a significant cost.

- The TL750Mxx chips are similar to the LM78xx chips, but have a very low “drop-out” voltage (~0.6 V).
- The LM79xx chips are similar to the LM78xx chips, but act as *negative* voltage supplies (-5 V, etc.).

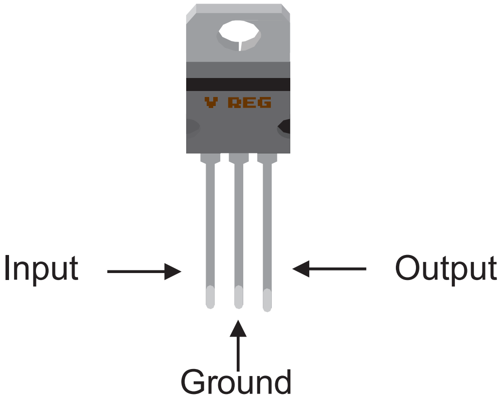

E.15.3 Specifications

- Maximum Input Voltage: 35 V
- Maximum Output Current: 1 A
- Built-in overcurrent protection
- Protection against short circuits
- Overheating protection (shuts off when overheating)

E.15.4 Usage Notes

- The specifications require two capacitors for operation—a 330 nF capacitor on the input and a 100 nF capacitor on the output (see below).
- These capacitors are generally not required for very simple projects—you can simply hook the input directly to your positive power source, the ground to your ground, and the output to your project.

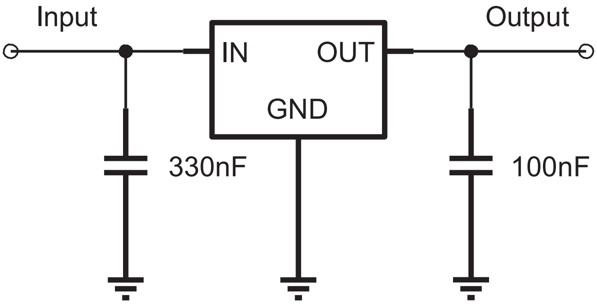

Index

A

Abbreviation, 14, 15
Adjustable resistors, 298
Amplification, 455
Amplification
  definition, 355
  gain of an amplifier, 356
  microcontroller, 355
  NPN transistors, 371, 372
  parable, 356
  parts, 358, 360
  transistors, 357
Analog Inputs
  current value, 226
  darkness sensor, 224, 225
  debugging, 226
  photoresistor, 224
  pin number, 223
  serial interface, 225
  values, 223
Analog output
  Arduino, 229
  darkness sensor, 227
  dimmer, 227–229
  pin numbers, 227
  PWM, 226
  variable, 228
  voltages, 227
  while command, 228
  while loop, 228
analogRead() function, 223, 229
analogWrite() function, 227, 230
AND Gate, 183, 193
Arduino environment, 205
  programming, 207
Arduino IDE, 207, 208
Arduino Programming/Bob Dukish's Coding, 417
Arduino Uno, 206
  breadboard, 213, 214
  LEDs, 216
  modifications, 219
  rewiring, 219
  wiring inputs/outputs, 214, 215
ATmega328/P, 203, 355
Auto-ranging multimeter, 73

B

Bipolar junction transistors (BJTs), 357, 370, 407
Boost converters, 153
Breakdown voltage, 106, 122
Brushed electric motor, 343
Buttons/switches, 4

© Jonathan Bartlett 2020
J. Bartlett, *Electronics for Beginners*, https://doi.org/10.1007/978-1-4842-5979-5

C

Capacitance, 236, 445, 446
Capacitative reactance, 323
Capacitors, 317, 318, 464
  AC *vs.* DC, 295, 296
  adding reset, 263, 264
  alternating current, 248
  capacitance, 236
  charge, 236
  charging/discharging, 243, 244
  circuit, 238
  current-limiting resistor, 264
  definition, 236
  direct current, 248
  electric field, 237
  farads, 236
  filtering, 249, 250
  hard-and-fast rules, 283
  non-polarized capacitor, 241
  polarized capacitor, 241, 242
  R5 resistor, 263
  resistor circuit, 264
  resistors, 283
  ripple, 249
  series/parallel capacitances, 245–248
  size, 263, 283
  swimming pool, 237
  symbol, 236
  tolerance, 242
  types, 241
  voltage, 237
  voltage divider, 264
Carrier frequency, 318
CD4081 Chip, 182
Ceramic disk capacitor, 241, 293
Charge, 235
  amperes, 27
  atomic number, 24
  coulomb, 26
  electrons, 24
  fundamental quantity, 23
  milliamperes, 27, 28
  movement, 25
  neutrons, 23
  nucleus, 24
  protons, 23
  reactants, 25
Charge and current quantities, 443
Circuit
  components, 47–49
  electricity, 46
  original problem, 47
  pathway, 46
  requirements, 45
  resistance, 46
  resulting problem, 47
  watermill—connected structures, 46
Circuit diagrams/electronic schematics, 53
Circuit pattern, 129
Coils, 305
Common emitter circuit, 367
Common emitter amplifier
  coupling capacitor, 456
  denominator, 455

  - headphones, 460
  - input current, 452
  - numerator, 455
  - Ohm’s law, 454
  - output current, 452
  - parallel resistances, 458
  - resistance, 453
  - simplifications, 460
  - Thévenin resistance, 457
  - transistor path, 457
  - volatge change, 456

Conserved quantities, 152
Conventional current flow, 31
Coupling capacitor, 250, 296, 456
Current, 235
Current-limiting resistor pattern, 131, 132
Current-limiting resistors, 217
current sink, 172
Current source, 172
Current/voltage, 463

D

Darkness Sensor Breadboard Layout, 172
DC motor
  - back EMF, 344
  - bidirectional, 348
  - calculations, 449
  - circuit, 345, 346
  - definition, 343
  - RPM, 347
  - servo, 349
  - shaft, 347
  - stall current, 351
  - stepper, 350, 351
  - theory of operation, 343
  - torque, 345

Digital logic, 179
Digital output 13 (D13), 209
digitalRead() function, 223
Diodes
  - actions, 105, 106
  - anode, 105
  - bad circuit, 112
  - cathode, 105
  - circuit calculations
    - parallel, 109, 110
    - series, 107, 108
  - components, 123
  - multiple, 111, 112
  - nonconducting, 115
  - protection circuit, 115, 120
  - short circuit, 113
  - single resistor, 106
  - types of protection, 120, 121
  - usage, 115, 116
  - voltage regulator, 117–119

Drawing circuits, 53–55
Dual in-line package (DIP), 161

E

Ebers-Moll model, 448
Electric fields *vs.* magnetic fields, 304–306

Electricity
  AC *vs.* DC, 28–30
  charge, 23, 25
  computers and phones, 23
  current flow, 30, 31
Electrolytic capacitor, 293
Electromotive force (EMF), 444
Electron current flow, 31
Electronics
  battery and connector, 5
  buttons and switches, 4
  components, 3
  home experimenter, 1
  jump wires, 4
  LEDs, 4
  low-current DC, 2
  math course, 2
  multimeters, 3
  power regulator board, 5
  practical circuit design, 1
  resistors, 4
  safety guidelines, 5, 6
  solderless breadboards, 3
Electronics symbols, 435, 436
Electrostatic discharge (ESD), 7
Equivalent resistance, 93

F

Field effect transistors (FETs), 357, 370, 407
  BJTs, 415
  modes, 408
  N-channel/P-channel, 409
  parts, 408
Filtering capacitor, 249, 296
555 oscillating circuits
  breadboard, 278
  current flow, 277, 278
  on/off times, 279, 280
  pinout diagram, 273–276
  schematic diagram, 274
555 timer oscillator frequency equation, 450, 451
5 V source, 185–187
Flash memory, 204
Flyback diodes, 310
Flywheels, 308
Forward voltage drop, 105

G

Galvanized nail, 26
Graphing electricity
  capacitor voltage, 290
  555 oscillator circuit, 291
  oscilloscopes, 292
  output voltage, 290
  square wave, 291
  tone generator, 292

H

Heatsink, 150, 151

I

Impedance
  calculation, 328
  definition, 326
  formula, 327
  symbol, 326
Induced voltage, 306, 307
Inductance, 447
Inductive kick, 309, 310
Inductive reactance, 323, 325
Inductors, 317, 318, 464
  definition, 303
  resisting changes, 307
  symbol, 303, 304
  uses, 308, 309
In-system programmer (ISP), 205
Integrated circuit naming conventions, 439, 440
Integrated circuits (IC), 159
  datasheets, 164–166
  DIP, 162
  DIP/SMD, 160
  LM393 chip, 167, 168
  LM393 voltage comparator, 163, 164
  parts, 160
  pins/pads, 160
  resistive sensors/voltages, 169, 170
  sensing/reacting, 170, 172
  voltage pins, 164
International System of Units (SI units)
  base units, 10
  calculations, 10
  derived units, 11
  imperial system, 10
  newton, 11
  prefixes, 12
  scientific quantities, 10
  single set, 10
Inverted inputs, 182
Inverted outputs, 182

J

Jumper wires, 4

K

Kirchhoff ’s current law
  four-way intersection, 88
  junction, 88, 89
  parallel circuit, 89
Kirchhoff ’s voltage law, 90–92, 101

L

Lenz’s law, 306
Light-emitting diodes (LEDs), 4, 105
LM393 voltage comparator, 163
Logic Gates, 191–193
Logic IC, 179
Logic Operations, 181
loop() function, 218, 225, 228
Low-pass filers, 335

M

Magnetic flux, 303, 304
Microcontroller, 201
  ATmega328/P, 202, 203
  development board, 204
Milliampere, 31
Milliamps, 40
Multimeters, 3, 7, 8
  auto-ranging, 73
  different functions, 72
  low-cost, 73
  measurement, 74
  metric suffix, 75
  millivolts, 74
  9-volt battery, 72, 74
  resistor datasheet, 75
  voltage drop, 72

N

N-channel enhancement mode MOSFET
  BJT, 412
  circuit, 411
  logic circuits, 414, 415
  microcontrollers, 411
  pull-down resistor, 412, 413
  schematic symbol, 410
NPN transistor operation
  Arduino output, 367
  beta, feedback resistor, 368, 369
  collector, 362, 363
  I_{CE}, 362
  off by default, 361
  saturation mode, 363
  transistor, as switch, 364–366
  V_{BE}, 361

O

Ohm's law, AC circuit, 331–333
Open circuit, 46
Open collector configuration, 164
Oscillations, 269
  AC *vs.* DC, 294
  DC bias voltage, 294
  frequency, 270, 289
  headphone output, 292
  RC time circuit, 293
  speaker, 289
  tone generator, 293
Oscillators, 270
  555 timer, 273
  hardware coordination clocks, 271
  radio transmissions, 272
  servo motors, 272
  sound production, 271
  time clocks, 271
  time period, 273
Oscilloscopes, 292

P

Packages, 360
Parallel circuits, 101
  components, 87

  junction, 87
  resistors, 87
  wiring, 99, 100
Parallel resistance, 95
  loop-in order, 93, 94
  real circuit, 97
  single resistor (R6), 93
  starting and ending voltages, 95
  voltage drop, 93
Parallel/series capacitors, 318, 319
Parallel/series inductors, 318, 319
Partial circuits
  model, 391, 392
  Thévenin equivalent circuit
    calculating values, 393–395
    circuits, 399–402
    equivalent descriptions, 398, 399
    reactive elements, 397
    resistance, calculating, 396, 397
    voltage, 403, 404
  Thévenin resistance, 396
pinMode command, 215
Potentiometer
  definition, 299
  symbol, 298
  variable resistor, 299
  voltage divider, 298
Power, 445
  component limitations, 149, 150
  dissipate, 150
  electronics, 147, 148
  energy/work, 146
  heat, 146
  heatsink, 150, 151
  signals, 153
  terms, 145
  transforming, 151, 152
Power rails, 63
Power regular module
  breadboard, 79
  9 V battery clip, 79, 80
  YwRobot breadboard power supply, 78
Power transistors, 370
Protection diode, 310
Pull-down resistors, 141, 188, 216
Pull-up resistor, 138–140
Pulse-width modulation (PWM), 223, 226
PWM signal, 336, 337

Q

Quiescent current, 383

R

Radio systems, 318
Radio transmissions, 272
Raspberry Pi, 418
RC time constants, 256, 316
Relative voltages/ground potential, 37, 38
Resistance
  circuit/device, 38
  constant amount, 39

Resistance (*cont.*)
  constant current, 39
  definition, 323
  formula, 324
  ohms, 38
  symbol, 324
  total distance traveled *vs.* total displacement, 326
Resistance and conductance quantities, 444
Resistor-capacitor (RC) circuit, 255
Resistor-inductor-capacitor (RLC) circuit, 329, 330, 334, 335
Resonance frequency, 334, 335
Rheostat resistor"s, 298
RL circuit, 315, 316
RL time constant, 316

S

Saturation mode, 363
Scaling units, 12–14
Schottky diode, 123
Semiconductors
  diodes, 447
  NPN BJT, 448
Series circuit
  components, 85, 86
  equivalent resistance, 86
  LEDs, 86
  milliamps, 86
  pathway, 85
Servo motor, 272, 349
Setup function, 208
Short circuit, 46
Short-circuit current/SHORT, 396
Simple first circuit
  adding wires, 52
  diagram symbols, 53
  multimeters, 50
  resistors, 49
  short leg LED, 51
Simple timer circuit
  breadboard, 262
  LED, 261
  pull-up resistor, 261
  reference voltage, 259
  resistors, 260
  supply voltage, 259
  voltage, 259
  voltage divider, 260, 262
Single pole single throw (SPST), 129
Snubber diode, 309, 310
Solderless breadboard, 62
  adding wire, components, 67
  bridge, 63
  components, 61, 65, 80, 81
  connection points, 62
  LED, 64, 68, 69
  LED circuit, 64
  measuring current, multimeter, 77, 78
  multimeter, 76, 77, 82
  parts, 63
  physical piece of wood, 61
  power rails, 63, 64, 66
  power rails, adding wires, 67

  - T-connectors, 68
  - terminal strips, 62, 65, 80
  - wires, 69–72
- Sound control, 298, 299
- Speaker wattage, 297, 298
- Stepper motors, 350, 351
- Surface mount devices (SMD), 161
- Surface mount technology (SMT), 161
- Switches and buttons, 129, 131

T

- Thévenin equivalent circuit, 391, 392
- Thévenin formula, 461–463
- Through-Hole, 160
- Time constant, 315, 316, 465, 466
- TO-220 package, 186
- Torque, 345
- Transformers, 152
- Transistor voltage amplifiers
  - amplification stage, 385–387
  - audio signal, 380–384, 390
  - collector resistor, 389
  - emitter resistor, 389
  - Ohm's law, convert current into, 375–378
  - oscilloscope, 387, 388
  - reading signal, 378, 379
- Trimming resistors, 298

U

- Unit of measurement, 9

V

- Variable resistors, 298
- Voltage, 235
  - coulomb of charge, 41
  - coulomb of electricity, 35
  - DC electronics, 41
  - electronics, 35
  - measurements, 36
  - universal reference point, 41
- Voltage divider, 132, 133
  - calculation, 134
  - resistor ratio, 135
  - resistor values, 136, 137
  - stiffer, 138

W, X

- Wires in circuit, 97–99

Y

- YwRobot breadboard power, 5

Z

- Zener diodes, 122

61218

Electronics for the Electrician

Newton C. Braga

PROMPT®
PUBLICATIONS

San José Public
LIBRARY
Benefit Assessment
Funds at Work

Electronics for the Electrician

By Newton C. Braga

Other Prompt® books by Newton C. Braga

Sourcebook for Electronics Calculations, Formulas, and Tables

Fun Projects for the Experimenter

Electronics for the Electrician

By Newton C. Braga

PROMPT®
PUBLICATIONS

©2000 by Sams Technical Publishing

PROMPT© Publications is an imprint of Sams Technical Publishing, 5436 W. 78th St., Indianapolis, IN 46268.

All rights reserved. No part of this book shall be reproduced, stored in a retrieval system, or transmitted by any means, electronic, mechanical, photocopying, recording, or otherwise, without written permission from the publisher. No patent liability is assumed with respect to the use of the information contained herein. While every precaution has been taken in the preparation of this book, the author, the publisher or seller assumes no responsibility for errors or omissions. Neither is any liability assumed for damages resulting from the use of information contained herein.

International Standard Book Number: 0-7906-1218-6
Library of Congress Catalog Card Number: 00-108430

Acquisitions Editor: Alice J. Tripp
Editor: Cricket Franklin
Assistant Editor: Kim Heusel
Typesetting: Kim Heusel
Indexing: Kim Heusel
Cover Design: Christy Pierce
Graphics Conversion: Bill Skinner, Phil Velikan, Christy Pierce
Illustrations: Courtesy the author

Trademark Acknowledgments:
All product illustrations, product names and logos are trademarks of their respective manufacturers. All terms in this book that are known or suspected to be trademarks or services have been appropriately capitalized. PROMPT® Publications and Sams Technical Publishing cannot attest to the accuracy of this information. Use of an illustration, term or logo in this book should not be regarded as affecting the validity of any trademark or service mark.

PRINTED IN THE UNITED STATES OF AMERICA

9 8 7 6 5 4 3 2 1

Contents

Preface vii
About the Author ix

Section 1

The Differences Between Electricity and Electronics 1
Fundamentals 2

Section 2

Electronic Components 41
Wires 41
Fuses 43
Switches 45
Cells and Batteries 47
Resistors 50
Variable Resistors 55
Incandescent Lamps 58
Neon Lamps 60
Light-Dependent Resistor or CdS Cells 63

NTC/PTC 65
VDR 67
Capacitors 69
Variable Capacitors 74
Coils or Inductors 76
Transformers 78
Relays 81
Solenoids 84
Motors 86
Loudspeakers and Headphones 88
Magnetic Transducers 91
Piezoeletric Transducers 92
Semiconductors 95
Diodes 97
Zener Diodes 101
LEDs 104
Special Diodes 109
Bipolar Transistors 111
Phototransistors 122
Optocouplers 126
Darlington Transistors 127
Unijunction Transistor (UJT) 132
Programmable Unijunction Transistor—PUT 137
Field-Effect Transistor—FET 140
MOSFETS 145
Power FETs 148
Isolated-Gate Bipolar Transistor—IGBT 152
Silicon Controlled Rectifier—SCR 154
Triac 162
Electromagnetic Interference (EMI) 167
Diac 171
Quadrac 172
Silicon Unilateral Switch—SUS 172
Silicon Bilateral Switches—SBS 174
Liquid Crystal Displays—LCD 175
Tubes 179
Integrated Circuits 184

Section 3

Troubleshooting and Repair 217
The First Step 218
Safety Rules of Troubleshooting 218
The Multimeter 221
How to Use a Multimeter 223
Reading and Interpreting Multimeter Values 227
Finding Parts 227
Practical Circuits: How They Work 230
Flashers 237
Automatic Lighting/Emergency Lighting 243
Alarms 245
Doorbells And Chimes 249
Battery Chargers 250
Fluorescent Lamps 253
AC/AC Converters 254
Power Conditioners 255
Intercoms 256
Wireless Systems 258
High-Frequency Appliances 262
Surge Protectors 264
Conclusion 265
Index 267

Digitized by the Internet Archive
in 2018 with funding from
Kahle/Austin Foundation

https://archive.org/details/electronicsforel00newt_0

Preface

The time of the electrician dealing only with simple components such as wires, outlets, switches, and incandescent lamps in his daily work is gone. Each day, electronics extends its branches to other activities. The home and industrial electrician, the automobile electrician, the construction engineer, and the architect are now finding strange components they may not be able to identify, which can cause embarrassing situations.

Electronic components in electric circuits and installations are becoming more common. Timers, dimmers, heater controls, intercom phones, alarms, doorbells, battery chargers, emergency lights, illumination controls, air conditioners, washing machines, and electronic injection systems are all electric devices and appliances where electronic circuits are found. The electrician and other professionals who come into contact with electric installations and who want to be updated with new technologies in these fields must know how these circuits work. This means learning a new science: electronics.

Until some years ago electronics and electricity were independent fields. The electronics technician fixed and repaired only electronic devices, concentrating his efforts in radio and TV, while the electrician worked with cars, home and industrial installations, and energy distribution. They were entirely different fields. Now the electrician must know how a timer or a dimmer works because he needs to know how to repair one if necessary, and a timer is a device of the other field—electronics. Other professionals such as architects and construction engineers must also know about many electronics devices because they must include them in their projects.

Of course, an electronics person, electrician, and other professionals in these areas must know these devices are not as advanced as the ones needed by a telecommunications, hardware, or other advanced specialist in the electronics field. Electricians must know the basic electronics devices found in their work. They must know how to recognize components and devices, how to work with simple circuits, and even detect problems in them. The electrician must know the basic electronics of common circuits used in electric installations of homes, industries, and cars to be secure when installing devices using this technology. The electrician also must know how to examine electronic circuits found in electric installations using basic instruments such as the multimeter.

Where can the electrician and other professionals find basic information about electronics used in electric installations? It was in thinking about this matter that I decided to prepare this book. It gives the reader the fundamentals of electronics, but with an approach familiar to the electrician and professionals who have had some contact with electric installations. The book explains how the basic electronic components work, how they are used in practical applications (homes, cars and industries), and how the electrician can work with them.

This isn't a simple book of electronics directed to all people. It is a book of electronics for electricians and other professionals from construction engineers to architects. It starts with knowledge about electricity and installations that electricians use. Eliminating some steps and going directly to the level where the reader is makes it easier to add important information about electronic circuits and devices. This makes the electrician much more prepared to make money when finding devices with new technology in his work.

This book is intended to give professionals basic knowledge about electronics in order to understand how electronic circuits and devices work in electric installations. The only thing I recommend to the reader is to proceed, and keep in mind that electronics will be increasingly present each day and that electric installations are not the exception.

About the Author

Newton C. Braga has written more than 50 books about electronics and electricity that have been published in his native Brazil as well as Europe and the United States. The main training network of the Brazilian Federation of Industries (SENAI) recommended *Electric Installations For Beginners* and *The Basic Electronics Course,* published in Brazil, as basic texts in its courses.

Electronics for the Electrician

Section I

The Differences Between Electricity and Electronics

The aim of this book is to explain how electronic circuits and devices work without assuming the reader has prior knowledge. Many electricians or professionals of nonelectronic areas probably have some elementary knowledge of electricity. We must assume there are many differences between the work of an electrician and an electronics technician. Not only are the components different, but in many cases the way they work is different as well.

So, it is normal that the first thing the electrician should consider when starting with these studies in electronics is if the knowledge he has about electricity is the same knowledge necessary to learn electronics.

The answer is yes.

The small differences that exist are in the way electricity is used in a domestic, industrial, or automobile installation and the way it is used in an electronic circuit. In essence the electricity is the same.

This means that the basic principles of electricity one needs to know when working on electronic devices and circuits are also the same. The electricians who want to upgrade their knowledge starting with a new science don't have to begin at zero and forget everything studied prior or learned during years of work experience. But, to be successful when learning electronics, it is a good idea to change some points of view about the way the principles are studied, especially when applied to electronics.

Other than electricians who are more practical than theoretical, two other kinds of electricians had to be considered. First, I considered those with a solid base of the theory and apply it in their profession, and secondly, those with a theoretical knowledge that perhaps has been dulled by time. For the first and third groups, a section on the fundamentals of electricity is included.

A section on approaches of the electricity principles applied to electronics is also included for that second group with the solid concept base. Perhaps you may not need this review of basic concepts, but a review is not a bad idea. After all, nobody is free from small slips in memory.

Through a different approach that involves redirecting the ideas, laws, principles, and concepts of electronics, the reader who is familiar with applications in electric installations can form a new image of electricity. This is very important when trying to understand how devices and circuits of a newer science work.

Finally, please realize a complete course of all basic principles needed to know about electronics needs much more space than is available here. Of course, my efforts were directed to include all the essential information a reader needs to have to understand how the main electronic circuits and devices work. If the reader wants to go a step further, much more detailed explanations can be found in many of the books offered by PROMPT®.

Fundamentals

Both electricians and electronics technicians work with the same kind of energy or "fluid" electricity.

As previously stated, the basic differences between an electronic device and an electric device are in the way they are mounted, the parts they use, and the way the electricity is used. Although the ends and the components used to find them are different, the principles of operation of each part of an electronic device and an electric device are the same.

Electricity is not found in different types. This means that the laws of physics that determine how electricity works and can be used are, in essence, universal. This is very important for the electrician who wants to learn about electronics. There are no differences between the laws and principles applied to electronic circuits and devices and the electric circuits and devices.

The Electric Current

Every material substance is made up of small particles called atoms. Atoms are so small that they are not visible under normal conditions. Then scientists discovered that atoms are made of even tinier particles called elementary or atomic particles as shown in Figure 1.

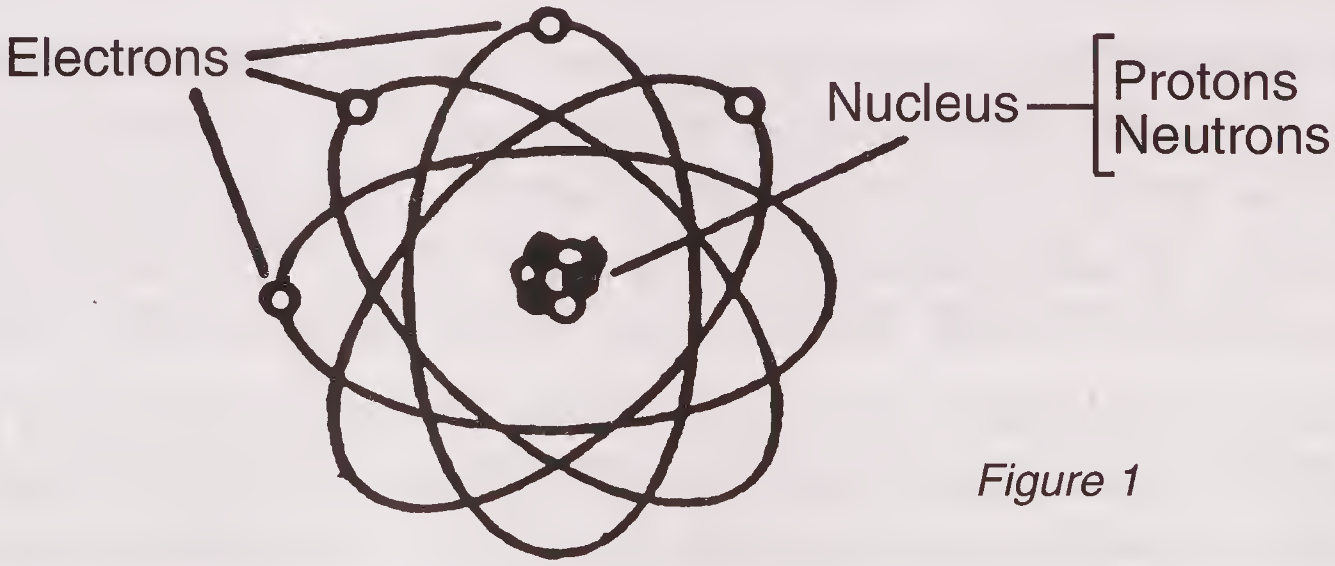

Figure 1

The nucleus of an atom is formed by protons and neutrons surrounded by electrons in an orbital shell. Electrons and protons manifest a special property; they are electrically charged. Electric properties of electrons are negative (–) whereas protons are positively charged (+). Bodies with different charges attract each other, and bodies having charges of the same type, either positive or negative, repel each other. This is why the electrons stay in an orbital movement attracted by the nucleus' charge.

In normal conditions, the number of electrons of an atom is equal to the number of protons. This means the electric charge of the protons and electrons is the same, so the atom is balanced (Figure 2).

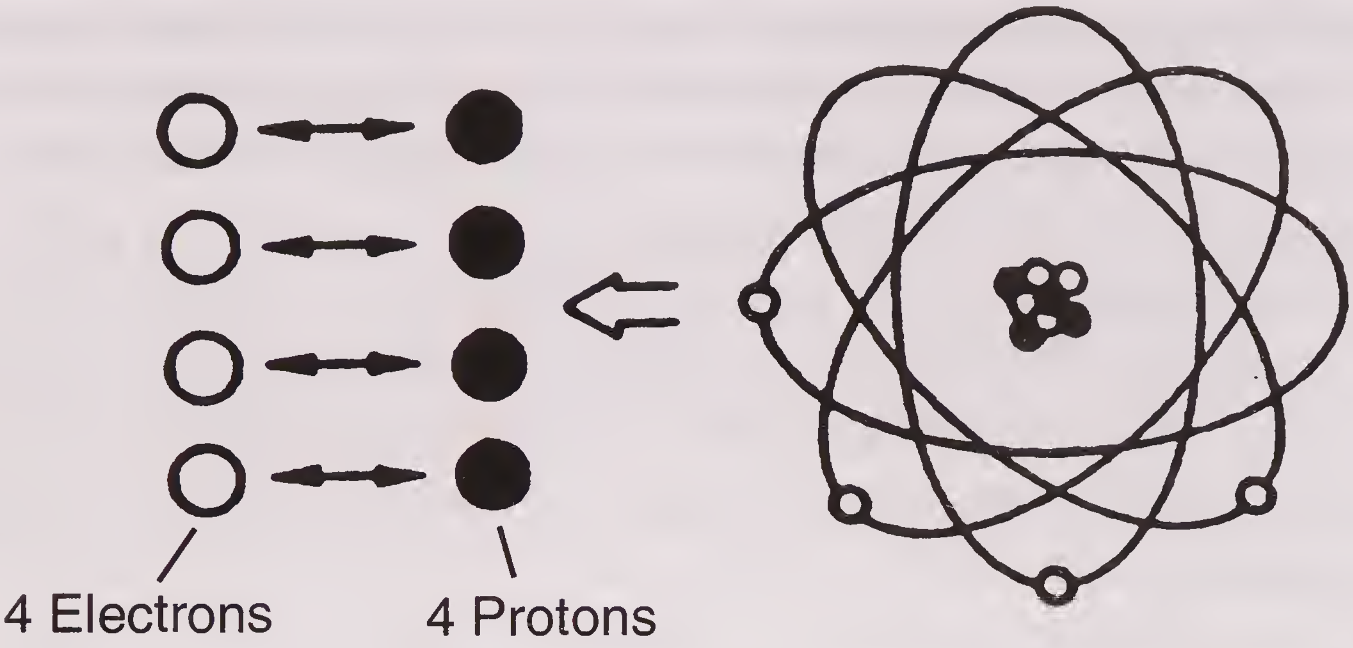

Figure 2

Although the forces act to maintain an atom in neutral or balanced condition, this stability can be upset under certain conditions. Of course, it is impossible to remove protons from the nucleus of an atom without destroying or changing its nature. This involves special conditions such as those created in a nuclear power plant or by an atomic bomb, where a fantastic amount of energy is released.

It is easier to play with the electrons to break the electrical balance of an atom because electrons can be removed from the orbital shell. They can also be added to the orbital shell without large amounts of energy. In some cases, simple natural or artificial processes can remove electrons from the atoms of a body and put them in the atoms of other bodies.

In such a situation, the body losing electrons becomes positively charged with an excess of protons. At the same time, the body that received the electrons now has a negative charge. Each body has a charge so, we say that the bodies have electric charges or are "electrified." An example of this is shown in Figure 3.

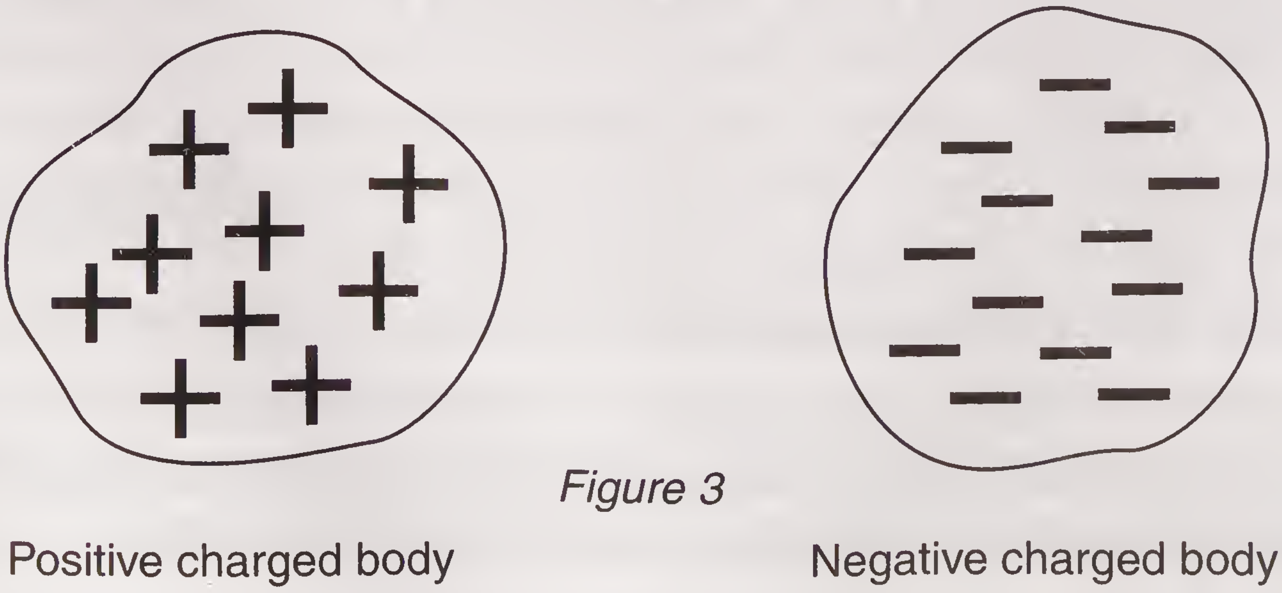

Figure 3

When you rub a plastic pen or glass rod in a piece of tissue (silk or wool works best), electrons are removed from the atoms of the plastic or glass and transferred to the tissue. Both objects become charged, but with opposite charges as shown in Figure 4.

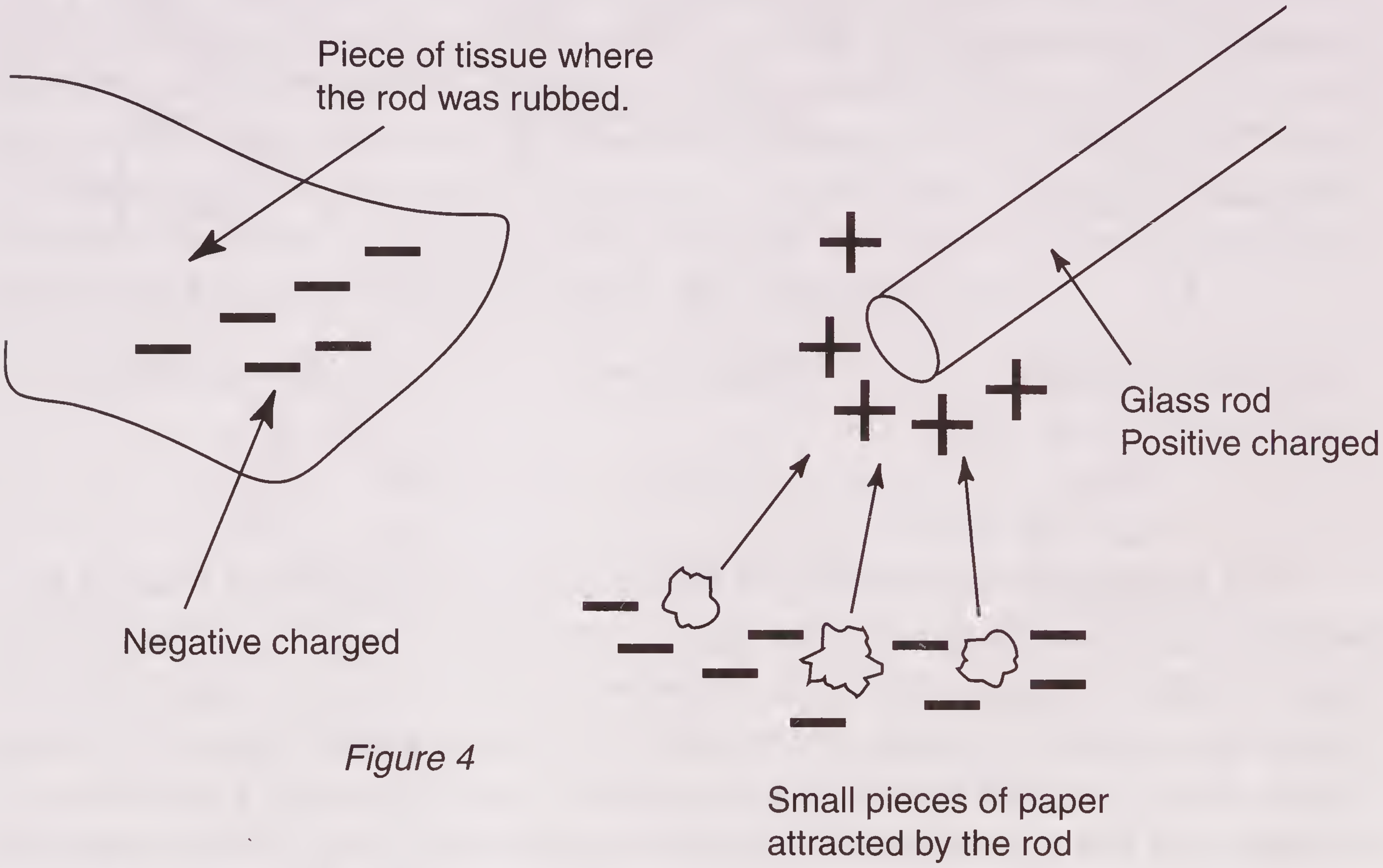
Piece of tissue where the rod was rubbed.
Negative charged
Glass rod Positive charged
Small pieces of paper attracted by the rod

Figure 4

Putting the pen near your hair or small pieces of paper will attract the hair or paper to the pen that is now positively charged. The pen needs the electrons in your hair or the paper to balance its charge. The pen and the piece of tissue have been electrified by attrition, but there are other ways to remove or add electrons to a body, thereby charging them with electricity. The electricity or electric charge in these bodies is fixed or static. The branch of physics that studies the properties of electricity accumulated in objects is static electricity.

Now, going a step farther, if we bridge two charged bodies using a metal wire (a material through which the electrons can move easily—a conductor), a flux of electrons can be noted; the excess electrons in one body move to the body that needs them. This flux of electrons or electron flow is called electric current and is represented in Figure 5.

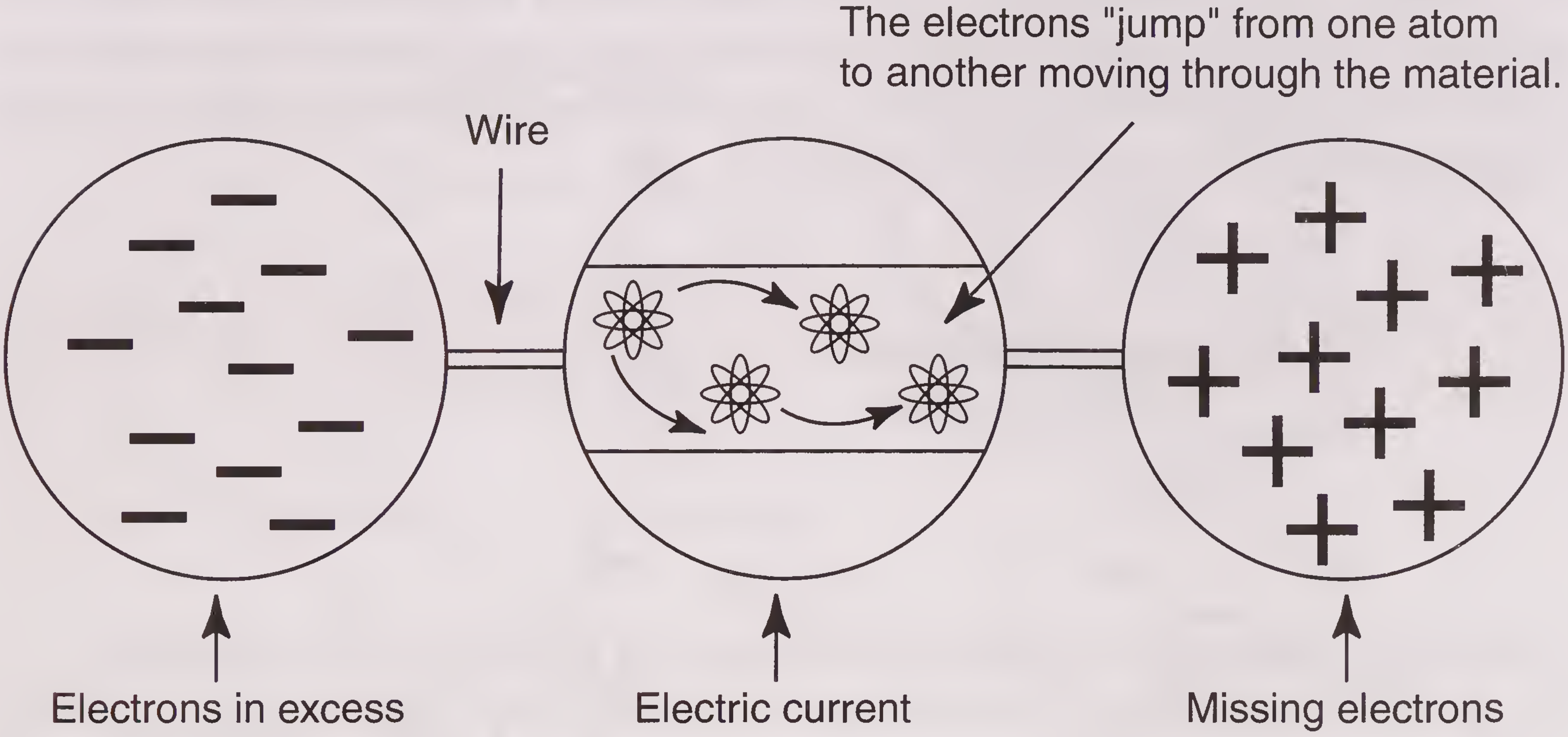
The electrons "jump" from one atom to another moving through the material.
Wire
Electrons in excess
Electric current
Missing electrons

Figure 5

It is important to note that in this process the electrons flowing between the bodies release energy. In this case, the energy is converted into heat. Another important point to observe in this process is that a flux of current can release energy to be used by devices placed along the way. Notice that in all the described processes the only moving particles or charges are the electrons. This explains why the names of the sciences are “electricity” and “electronics” and not “protoncity” or “protonics”!

Scientists found that the existence of atoms with positively charged electrons and negatively charged protons is possible somewhere in the universe. But, these atoms result in antimatter bodies that explode when in contact with our normal atoms. These “antielectrons,” or positively charged electrons, are called “positrons.”

The current flow between two charged bodies, as shown in Figure 5, is very short. As soon the electrons from one body reach the other body neutralizing any charge, the electron flow stops and the current ceases. The practical use of the energy that is transported by an electric current comes from the fact that we can create ways to maintain the difference in charges between two bodies and maintain a flux of electrons or current between them. Current flow is a dynamic process countering the static electricity.

a) Basic Circuits

The electric current can transport energy from one point to another, maintaining the difference of a charge's concentration between the ends of the wire with a current flowing through it.

In the example, two bodies are shown, one with negative charges and the other with positive charges. This is not the only possible configuration that can produce a current flowing between two bodies. They can both be negative or positive but in different degrees or with a different concentration of charges as shown in Figure 6. The charges in these bodies are in different degrees of compression meaning that different repulsion forces move them, either attracting electrons or repelling them.

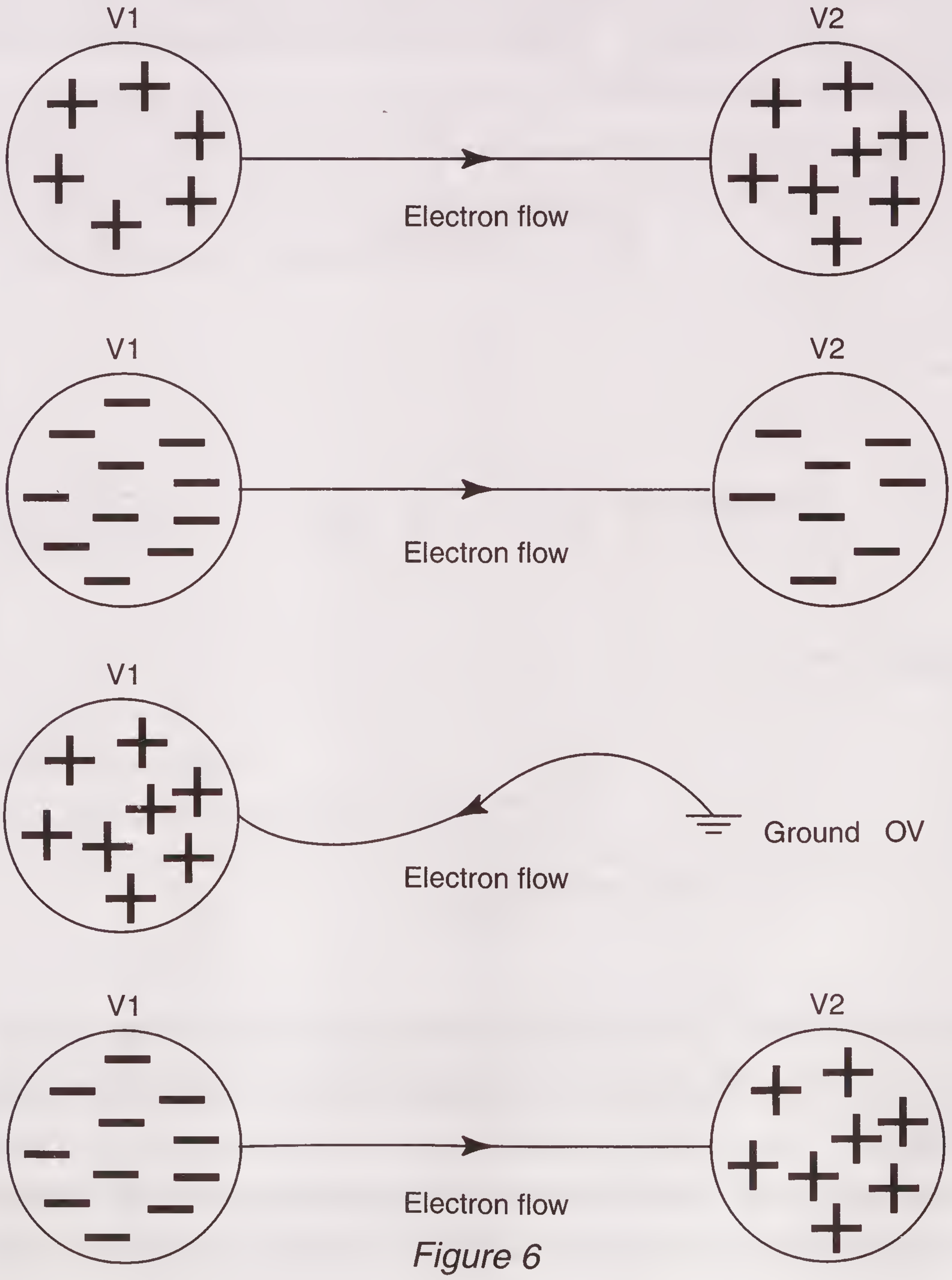

Figure 6

If we place a metal wire between these bodies, a current flow will be formed. The current tends to flow from the more compressed negative charges to the less compressed negative charges. This compression force of charges is very important to establish an electric current. In the case of positively charged bodies, electrons will flow from the body with the least amount of positive charge to where the most electronics are missing. As a practical rule, it is easy to see the natural tendency for a balance to be reached.

Returning to our problem, how do we maintain the current if we want to create a constant supply of energy? First, we need an efficient way to replace the electrons that leave the negatively charged body. Since electrons can't be created from nothing, we need a body that can furnish them as long as they are needed. This problem can be solved by the device shown in Figure 7, a generator.

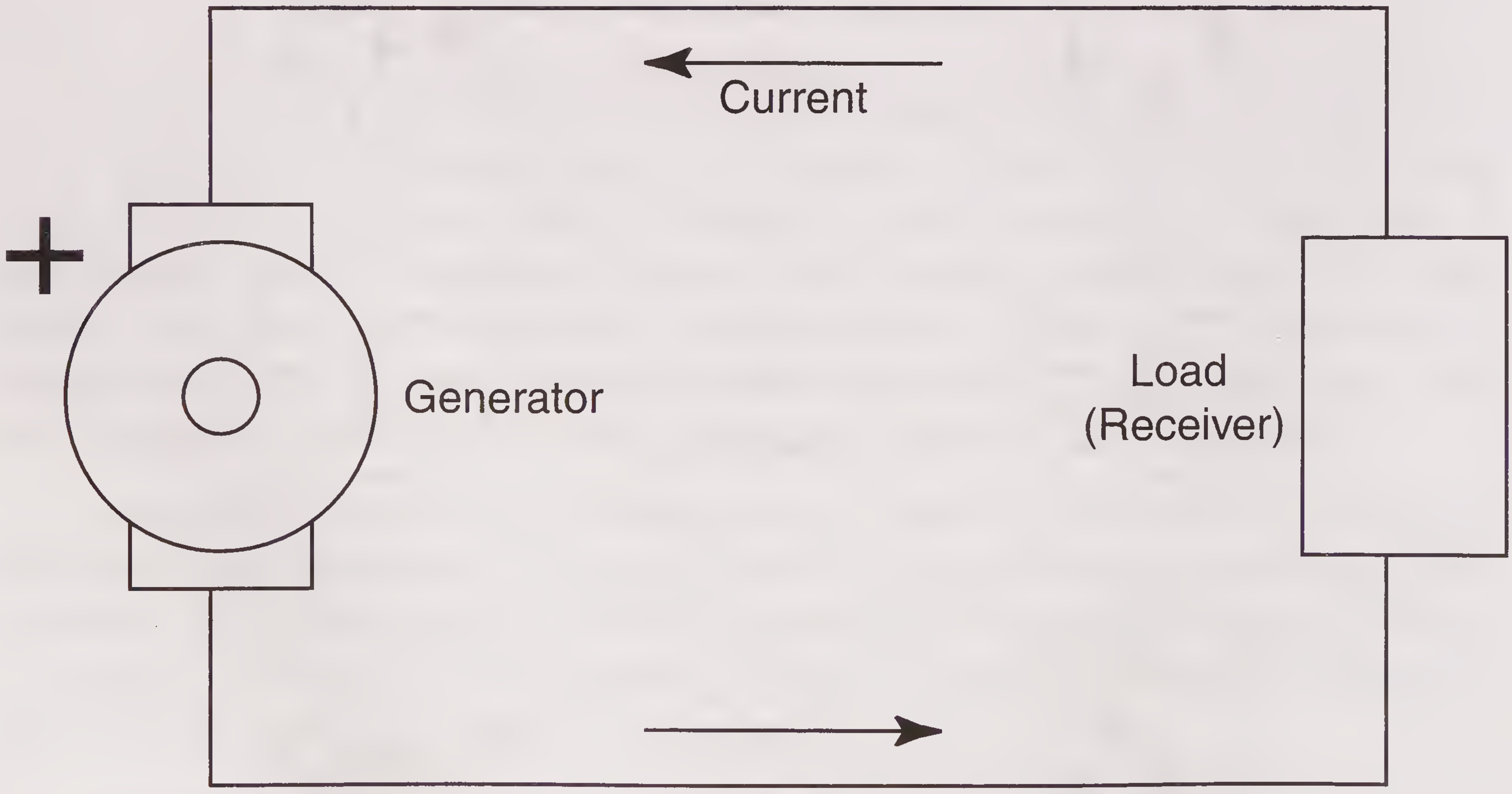

Figure 7

This arrangement is formed by the generator, metal wires, and a device that can absorb the energy transported by the current flow. This is an electric circuit.

So, how do things work in this arrangement and how does the generator supply the electrons to keep the current flowing? The generator can be seen as a two-pole device where one pole is a body with excess electrons (negative pole) and the other has missing electrons (positive pole).

The excess electrons are "pumped" in one direction to the pole where they are missing, passing through a device where the released energy is transformed into heat. Arriving at the pole where the electrons are missing, a replacing process occurs: inside the generator the electrons are taken from the positive pole and replaced to the negative pole again. In this process the generator uses energy. Chemical generators like cells and batteries use energy released in chemical reactions. Mechanical generators known as dynamos use the force applied to a mechanism.

It is important to observe that the generator only needs to replace energy when the current is flowing because in a cell there is no chemical reaction until the moment when it is used. Another important fact to observe is that energy can't be created from nothing and when the current is carrying energy to any device, it must be produced from another kind of energy. The generator doesn't create energy, but transforms one kind of energy into another as suggested in Figure 8.

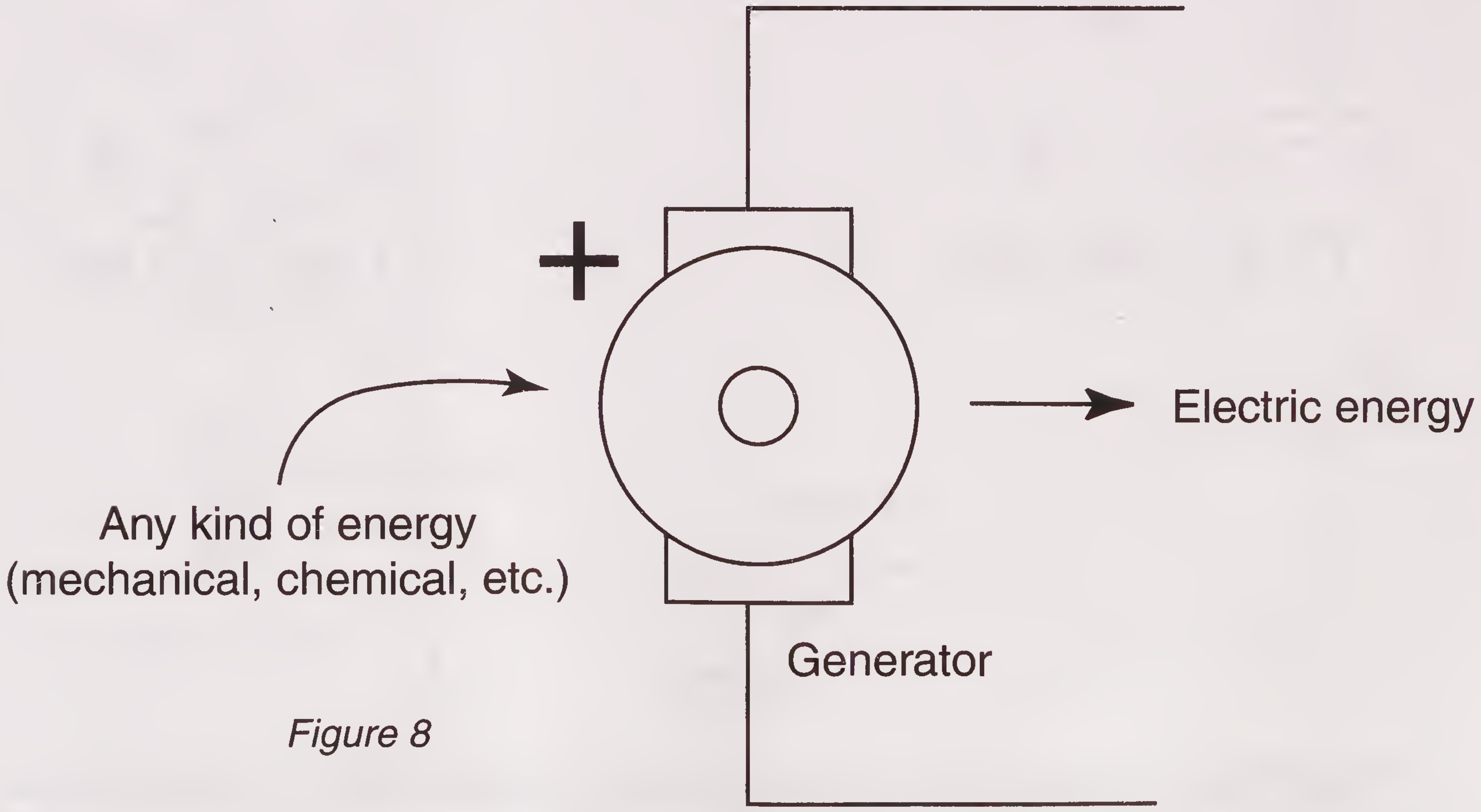

Figure 8

Another important thing to observe in the example is the movement of the electrons and the electric current flowing through a closed loop. This means that the current flow needs a closed circuit to be established. This configuration is called an electric circuit. Going a little bit further with our explanations, a switch can be included in this simple electric circuit to control the current. This is shown in Figure 9.

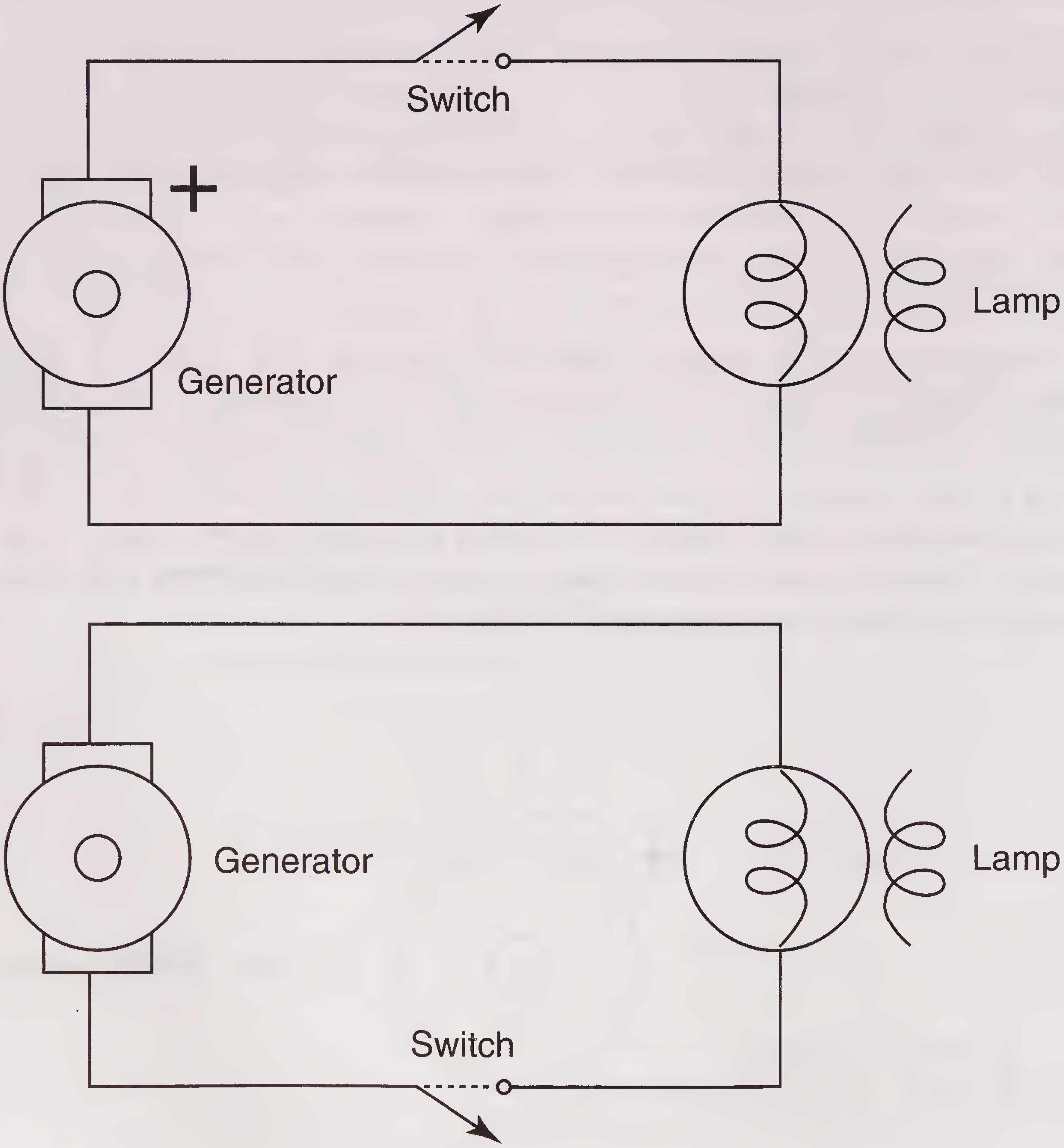

Figure 9

Opening the switch interrupts the current flow and the process of transferring energy to a device. If the device is a lamp we can turn it on and off. Closing the circuit, the current can flow and energy is transferred from the generator to the device; the lamp glows. Opening the switch, no current flows and no energy is transferred to the lamp; it remains off. The important thing to observe in this circuit is the position of the switch. Whether we place the switch after or before the device (having the current flow as reference), the effects are the same. The current can be interrupted in the same manner and the device doesn't receive any energy.

Another point that causes some doubts in electricity students is that the electrons aren't the energy and so they are not consumed when passing through a device. The electrons after and before the device are the same, and the amount of current in the entire circuit is also the same. The electrons are used only to transport energy. They are not the energy!

b) Voltage and Current

The excess electrons in a body and the electrons missing in another body represent a state of compression or tension. The situation can be compared to the situation found in compressed springs. The force that tends to recover the balanced state of a body depends on how the charges are compressed as shown in Figure 10.

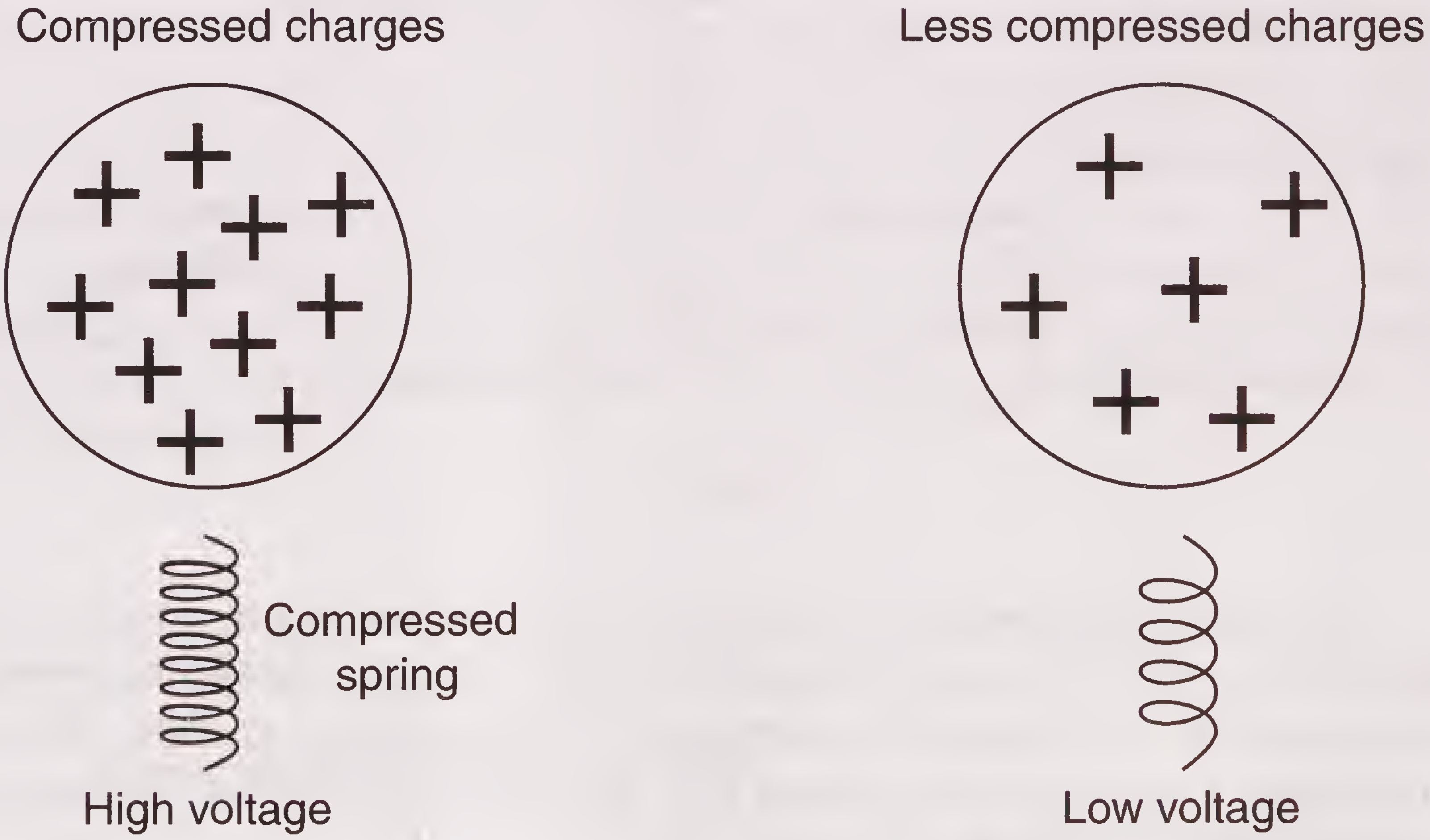

Figure 10

It is this tension that pumps the charges through a metal wire when that body is connected to another. Of course, if the charges in the other bodies are in the same state of tension, the forces pumping the charges are equal and no current flows. It is easy to conclude from this explanation that it is necessary to have a state of tension to produce current. The electric tension is the cause and the current is the effect.

This state of tension or compression is called electric tension or electric voltage and is measured in volts (V).

In a generator, like a cell for instance, there exists a state of tension between its poles—an electric tension. This means that when any circuit is wired between its poles a current can flow. This state of tension in a generator is called electromotive force (emf) and it is also measured in volts. For example, the emf of a single AA cell is 1.5 V. Current is measured in amperes (A).

Figure 11 depicts an analogy between electric current, electric voltage, and the water flowing from a reservoir.

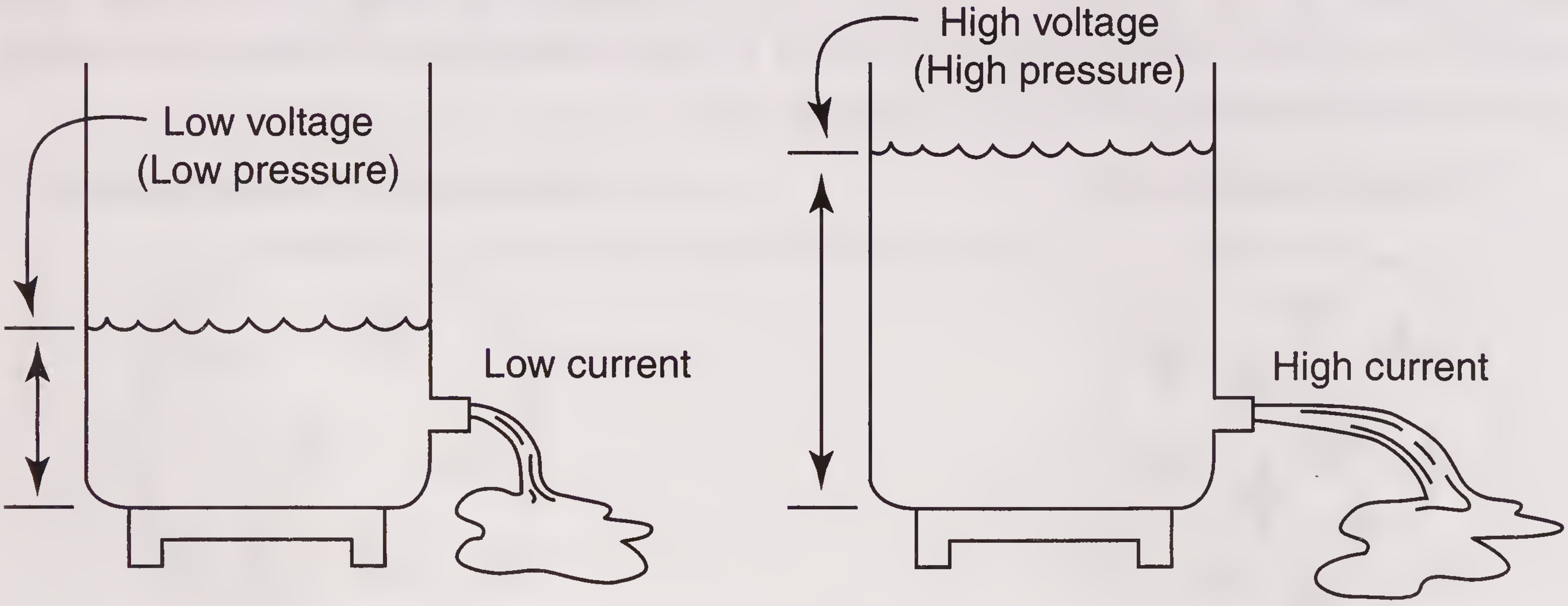

Figure 11

The water compression is compared with the electric tension or voltage and the water flow with the electric current. The higher the water level in the reservoir, the higher the compression and amount of water flow. In comparison, the higher the voltage of a generator, the higher the amount of current flowing through a circuit.

c) Resistance

When describing the basic circuit and the analogy between electric current and water flow, we added an important element: the device that uses the energy released by the generator converting it into heat or light. When reaching this device, the electron flow finds a certain degree of opposition to its movement as suggested in Figure 12.

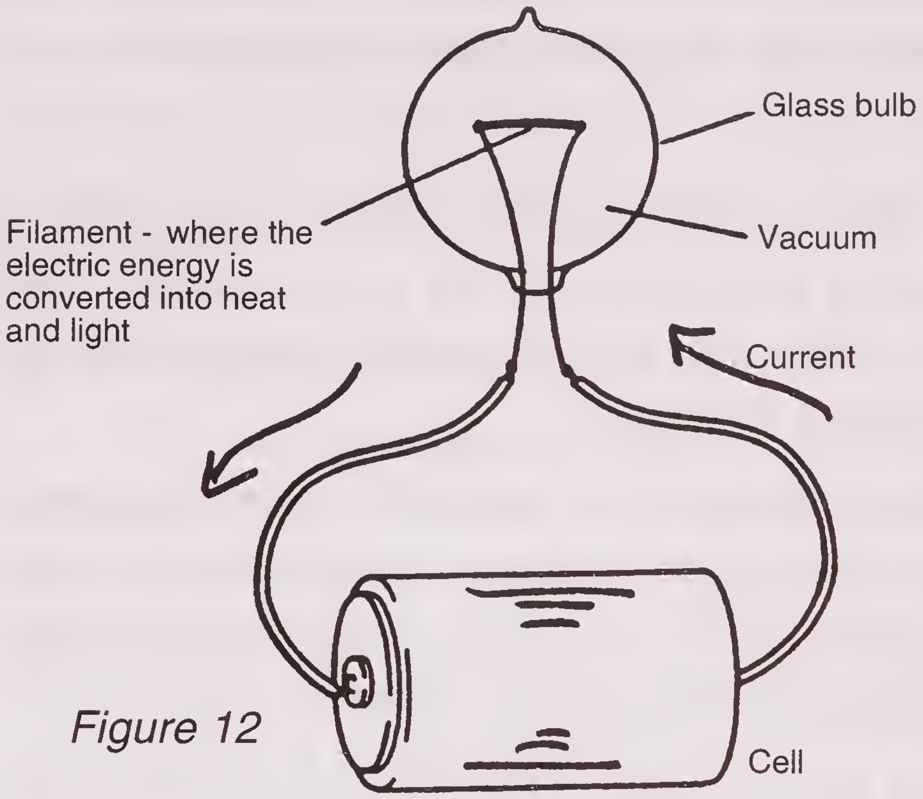
Glass bulb
Filament - where the electric energy is converted into heat and light
Vacuum
Current
Cell

Figure 12

Energy must be released to break this resistance. This energy is converted into heat. The opposition to the current flow is called "electric resistance" and is measured in ohms (Ω). The amount of resistance presented by a device determines how much current can flow through it when a determined voltage is applied to the circuit. The relationship among current, resistance, and voltage in a circuit can be calculated by Ohm's Law (Figure 13).

In a simple manner, considering that the voltage is the cause of a current, we can say that the amount of current flowing through a circuit with a certain resistance is proportional to the applied voltage. The electrician will find that all electronic circuits and devices presenting a resistance will follow this law the same as any other electric device does. In electronic circuits the electrician will also find a device that is used to add a certain resistance to a circuit. It is used in applications where the current flow can be reduced to a determined value. It is a resistor and uses the symbol shown in Figure 13.

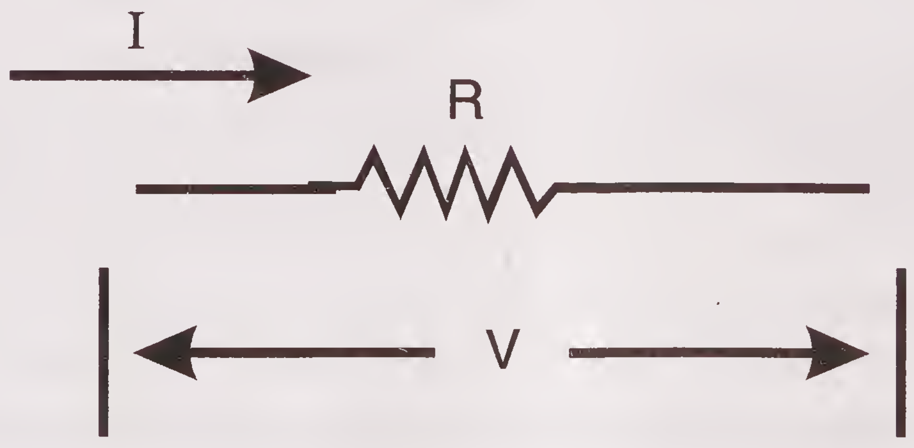

$$R = \frac{V}{I}$$

R = Resistance in ohms

I = Current in amperes (A)

V = Voltage in volts (V)

Figure 13

Because resistors are very important devices in electronic circuits, more information about them is presented in the discussion of components.

d) Units

Electricity can be measured. Due to this fact, the use of electricity in many applications is possible. Making calculations, we can preview what happens with electricity in a circuit under any condition and create circuits and devices that produce all the effects we want.

We measure voltages or electric tension in volts (V), the current in amperes (A), and the resistance in ohms (Ω). There are many other electric quantities, each one using its own unit. Of course, many electricians are familiar with these units as they also are used in electric circuits.

Because electricians can also find circuits where voltages and currents are very high or very low, it is important to know how to express high values of current and voltages. Commonly used are the Greek prefixes kilo (k) and mega (M). So, instead of saying 1000 volts, the electrician can say 1 kilovolt (kV) where the kilo (k) represents x 1000. Thus, 1000 volts is also 1 kV.

In electronics high voltages and high resistances are common, but much more common are the small values of currents and other quantities such as capacitances, inductances, etc. To represent the small quantities we can also use Greek prefixes.

Table 1-1 shows the most common Greek prefixes for representing electric quantities:

Prefix	Abbreviation	Times	exponential
Tera	T	1,000,000,000,000	10^{12}
Giga	G	1,000,000,000	10^{9}
Mega	M	1,000,000	10^{6}
Kilo	K	1,000	10^{3}
Deca	D	10	10
Deci	D	0.1	10^{-1}
Mili	M	0.001	10^{-3}
Micro	μ	0.000,001	10^{-6}
Nano	N	0.000,000,001	10^{-9}
Pico	P	0.000,000,000,001	10^{-12}

Table 1-1

In electronic circuits, it is common to see specifications of current like 10μA, meaning 0.000001 A, or specifications of resistances like 2.2 M_{Ω}, meaning 2,200,000 ohms. In some cases, the decimal point can be replaced by the prefix. So, instead of 2.2 k_{Ω} to express 2,200 ohms, we can write 2k2.

e) Effects of Electric Current

The device used as load in our example presents a pure electric resistance or an "Ohmic" resistance, converting all the electric energy into heat. Many devices and electric appliances use this effect of the electric current for heaters, incandescent lamps, etc. But, this is not the only effect of an electric current. When the electrons flow through certain materials or devices other effects can be noted.

Since the electric and electronic devices as well as appliances use all the effects of an electric current it is very important to any electrician to remember what these effects are and how they are used.

Thermal Effect or Joule Effect

When passing through any device or circuit presenting an electric resistance, the result of a current flow is that the energy spent in the process is converted into heat. This effect is also known as Joule's Effect. The amount of energy released can by calculated by Joule's Law.

Joule's Law says that the amount of heat dissipated each second or converted into heat by a device with a constant resistance is given by the voltage applied, multiplied by the current flowing through it.

The energy is measured in joules (J), and the amount of energy in each second is another quantity measured in watts (W). One watt is one joule per second.

$P = V \times I$

Where:

P is the power converted in heat in watts (W)

V is the applied voltage in volts (V)

I is the current through the circuit in amperes (A)

Since current and voltage depend on the resistance, the next formulas are useful to calculate the converted power:

$P = R \times I^2$

and

$P = V^2/R$

Where:

R is the resistance in ohms (Ω)

Be aware there are many other formulas involved in electronic calculations that depend on the application and the complexity of the involved circuit. These are very simple and important to the electrician who wants to start with electronics. To those who want to go on to higher levels and have a useful handbook of electronic formulas and calculations for reference, the *Sourcebook for Electronics Calculations: Formulas and Tables* is an excellent source. (Also written by Newton C. Braga and published by Prompt® Publications.)

Many electronic devices use the Joule Effect as a basis of operation; others have the production of heat only as a result of their operation. Some devices where heat is the basis of operation are incandescent lamps, heaters, and boilers. Other devices produce heat using this effect, but as a consequence of their operation. In this case, the heat must be eliminated in order to avoid causing damage to the device.

Some electronic components producing large amounts of heat must be mounted in heat sinks (Figure 14).

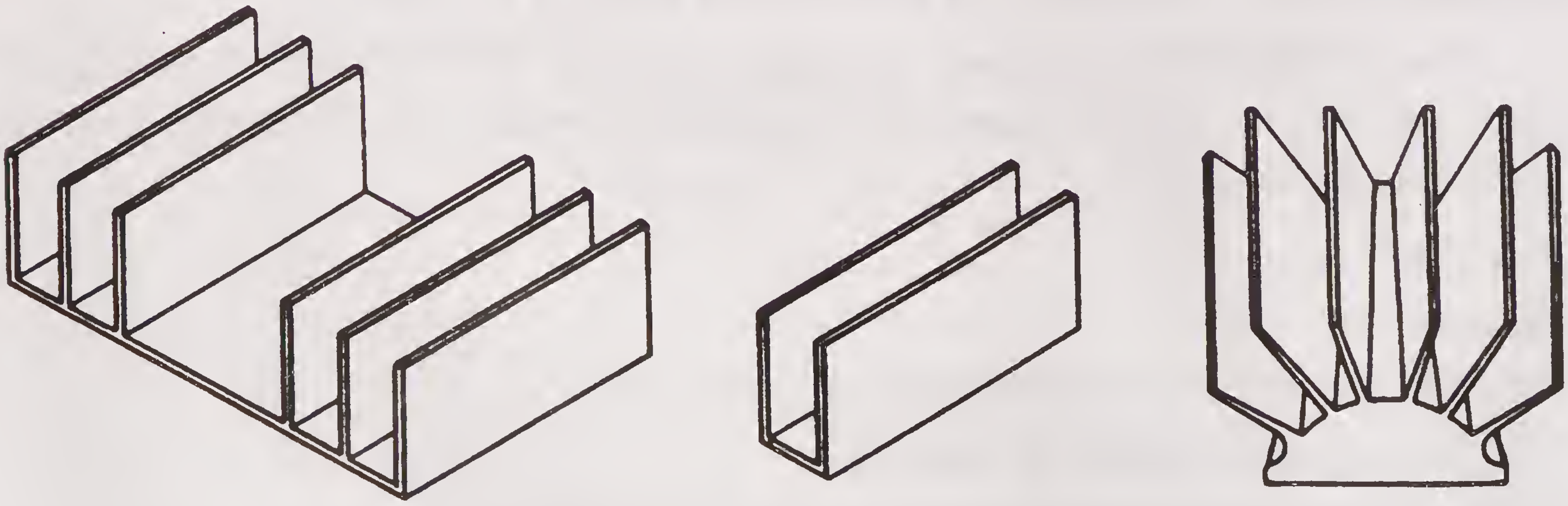

Figure 14

The function of a heat sink is to transfer heat generated by the devices to the air, and in some cases, with the aid of some forced process. Small fans can be used. The electrician will also observe that many circuits are very sensitive to the excess heat produced by components. Heat is the origin of many device failures.

Luminous Effect

There are many ways to produce light from electric energy. The most common way can be considered a consequence of Joule's Effect. In a common incandescent lamp a metal filament (tungsten) is heated by the electric current to a temperature high enough to produce light. Light can also be produced by a current flowing through a gas tube as in neon, xenon and fluorescent lamps. The gas becomes a conductor by ionization and in this process electromagnetic waves in the range of visible, infrared, and ultraviolet are produced. Another way to produce light from electric current is found in electronic components called LEDs (Light Emitting Diodes) whose operation principle is addressed later in the book.

Magnetic Effect

This is the only effect that ever occurs independent of the material in which the electric current flows. A Danish researcher named Oesterd discovered, with a simple experiment, that the movement of electric charges through a wire (current) produces a magnetic field as shown in Figure 15.

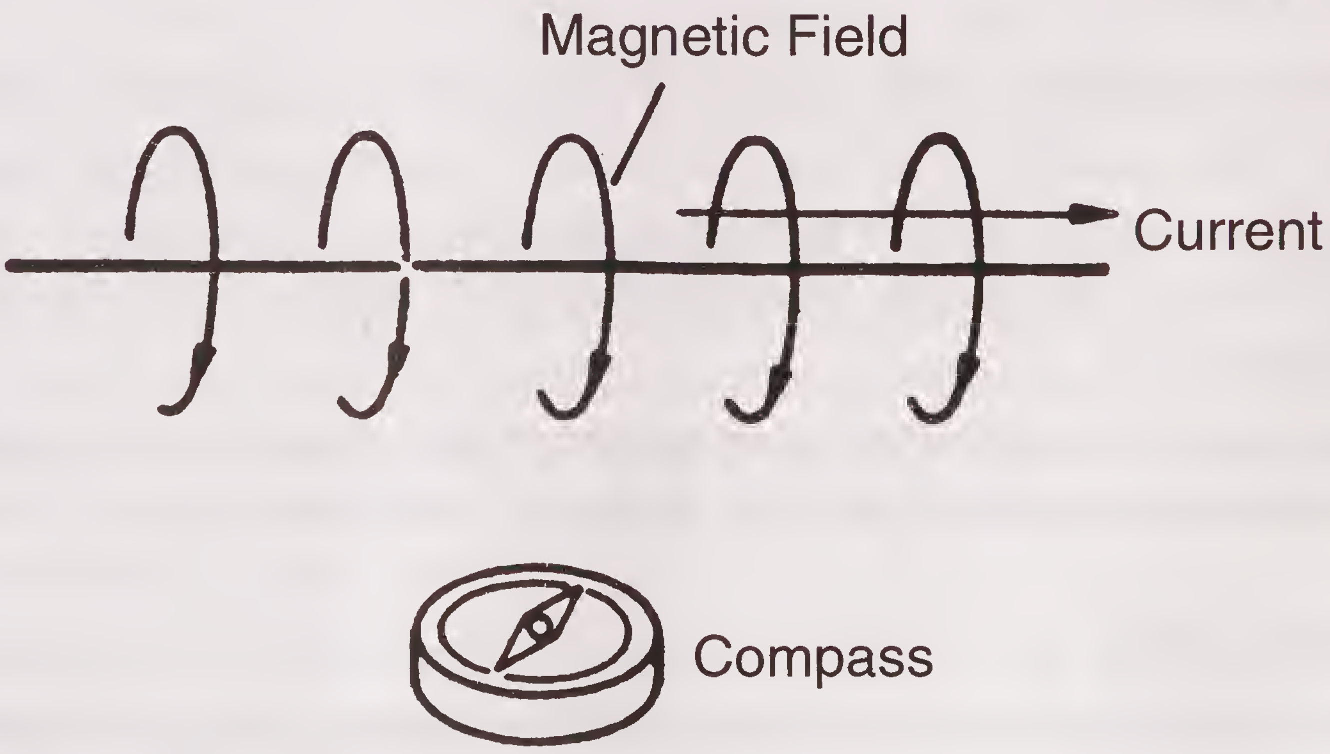

Figure 15

Magnetic field lines involve having the wire oriented in a manner that depends on the direction of the current flow. This effect can be used in many electronic devices. If a current flows through a wire formed into a coil, for example, the magnetic field can be concentrated inside it as shown in Figure 16.

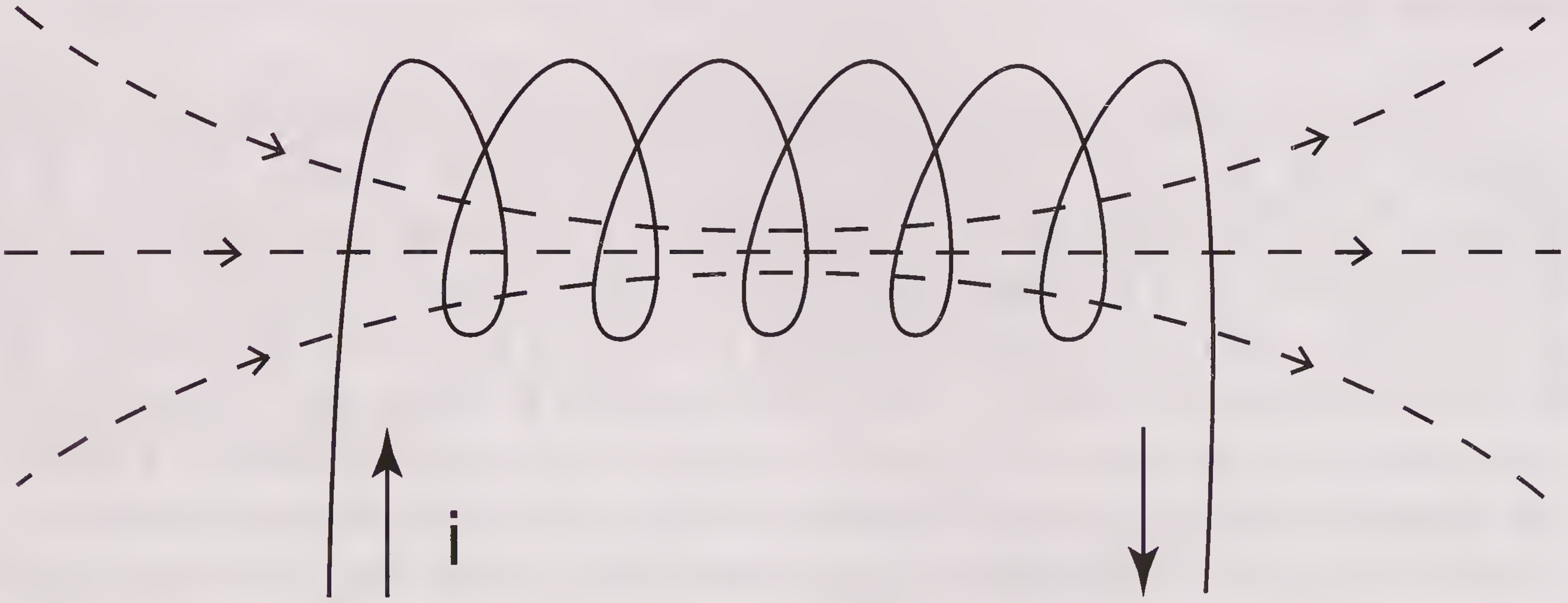

Figure 16

Other researchers found that the inverse effect also exists; if a wire is cut by a magnetic field, an electric current is induced. Devices such as solenoids, electromagnets, relays, motors, loudspeakers, and many others operate based on this effect. These devices include transducers and sensors.

Chemical Effect

When passing through chemical solutions, the electric current can be the agent that causes chemical reactions. The best known of them is decomposition of water and its elements (oxygen and hydrogen) called electrolysis. It can be done with an electric current made with a simple experiment (Figure 17).

Some electronic devices such as electrolytic capacitors and cells when charging are examples of devices that use this effect.

Physiologic Effect

Electric current, when flowing through a living body, can act on that body. For example, in humans electric current can excite the nervous sys-

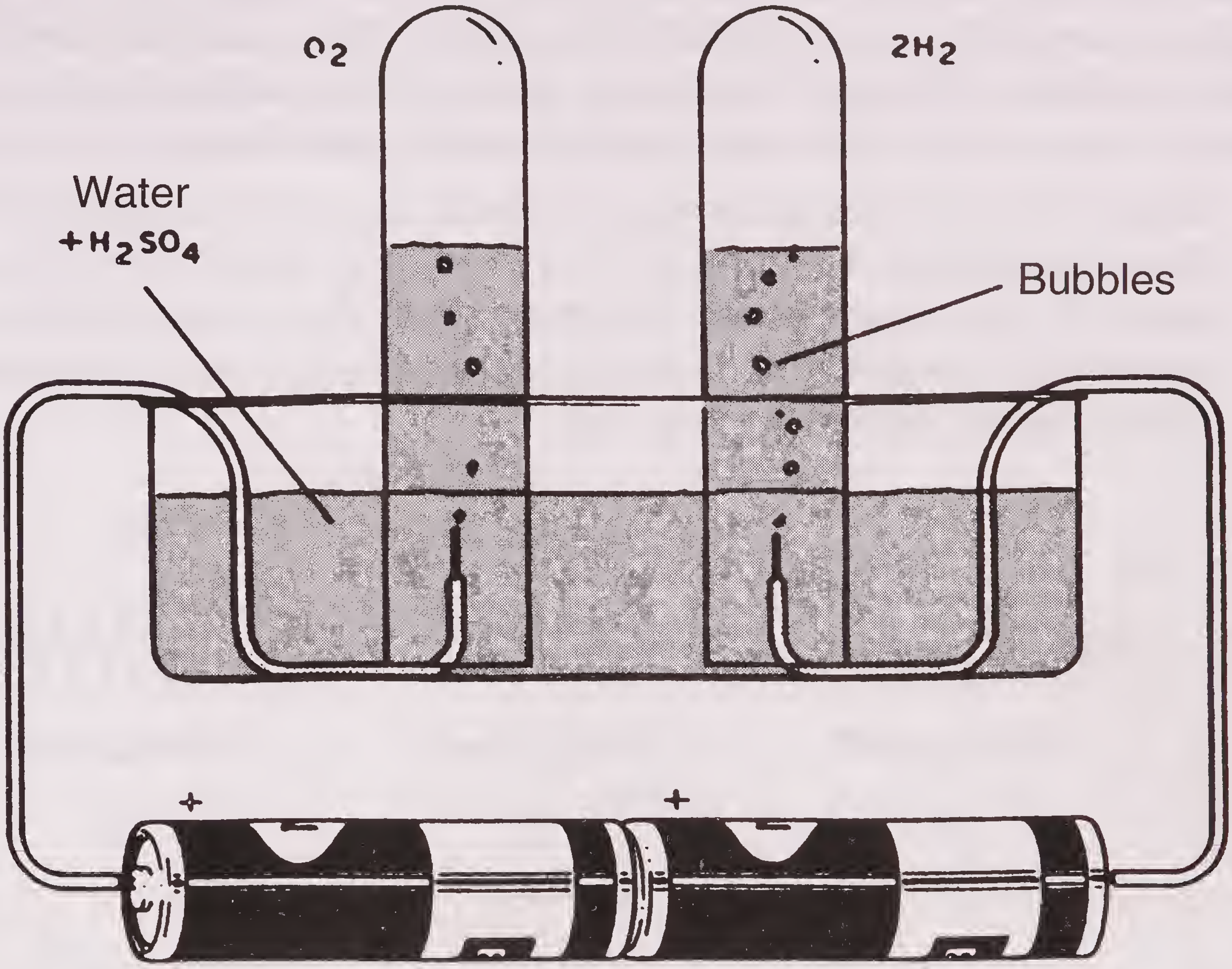

Cells

Figure 17

tem. It can cause a little itching sensation when the current flow is very low or severe shocks and burns if the current flow is high. This effect is used in some applications for nerve stimulators and other medical applications.

f) Series and Parallel

In an electric circuit one power supply can be used to power many elements. Even in a simple circuit several elements are used. The manner in which these elements are wired determines how the current flows, how the voltages are divided, and how the heat is produced by each element. In electrical applications the analysis of the circuits is generally simple when few elements are used.

Otherwise, in electronics the circuits used to be very complex, with hundreds or even thousands components. Knowing the way in which the components are wired is fundamental to find failures or to calculate component values. Any electrician must know the differences in the way the various components are wired to be able to make a correct analysis of their operation. Although the circuits can be formed of hundreds or thousands components, there are only two ways to wire them:

1. Series—Two or more elements of a circuit are wired in series if they are placed as shown in Figure 18. The current in all the elements of this circuit (in our case, resistors) is the same. The voltage across each element is proportional to its resistance. The components with the larger resistance will dissipate more heat.

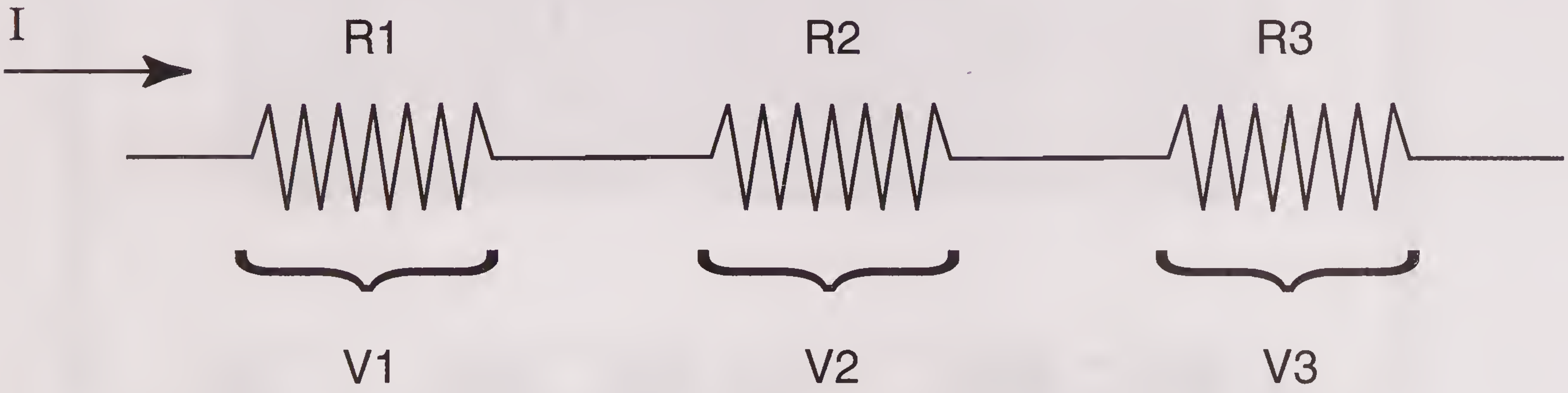

Figure 18

2. Parallel—The components of a circuit (in our example lamps) are wired in parallel when placed as shown in Figure 19. The voltages applied to all the components are the same, and the element presenting the lower resistance dissipates more heat as the current in this element is also the higher.

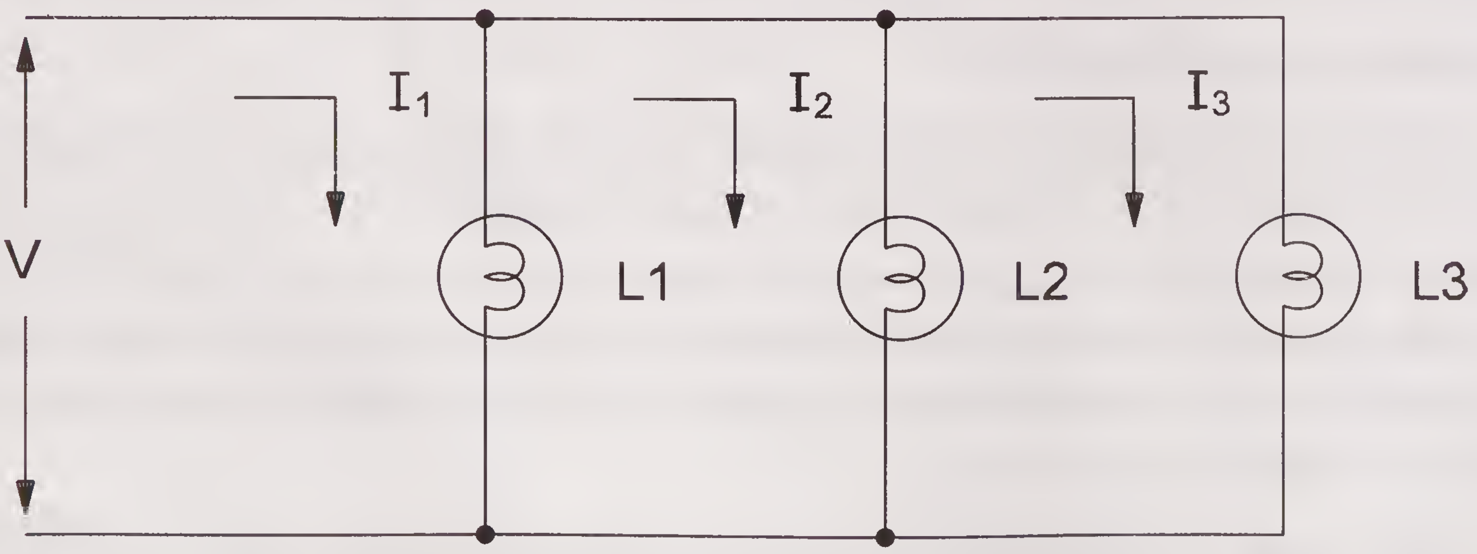

Figure 19

When working with electronic circuits, the electrician can find many types of components and, they can be combined in complex associations where some devices are in parallel and others are in series. So, series-parallel associations are common (Figure 20).

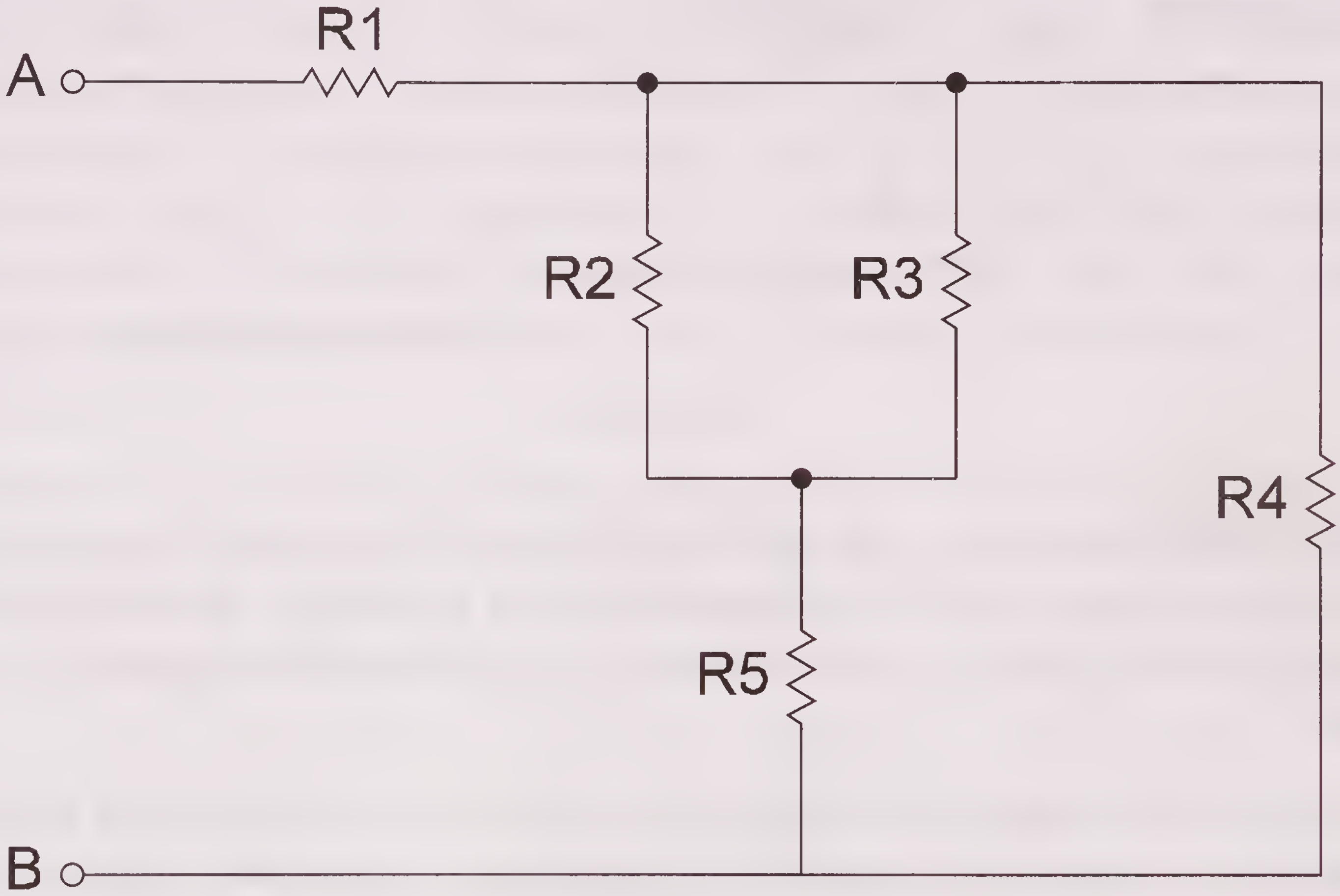

Figure 20

The electrician can analyze one of these configurations taking into consideration which components are in series and which are in parallel. For readers who want to learn more about this topic, a specific electronics course in calculus of equivalence resistance associations is recommended.

g) AC and DC

In basic examples used to show how an electric circuit works, we supposed energy sources or generators with special characteristics. In the poles of these generators we found a constant voltage meaning that when a constant resistance circuit was wired to them, the current flow was established in one direction with a constant value.

This circuit operates with a DC or direct current generator and the current through all its elements is referred to as DC (Figure 21).

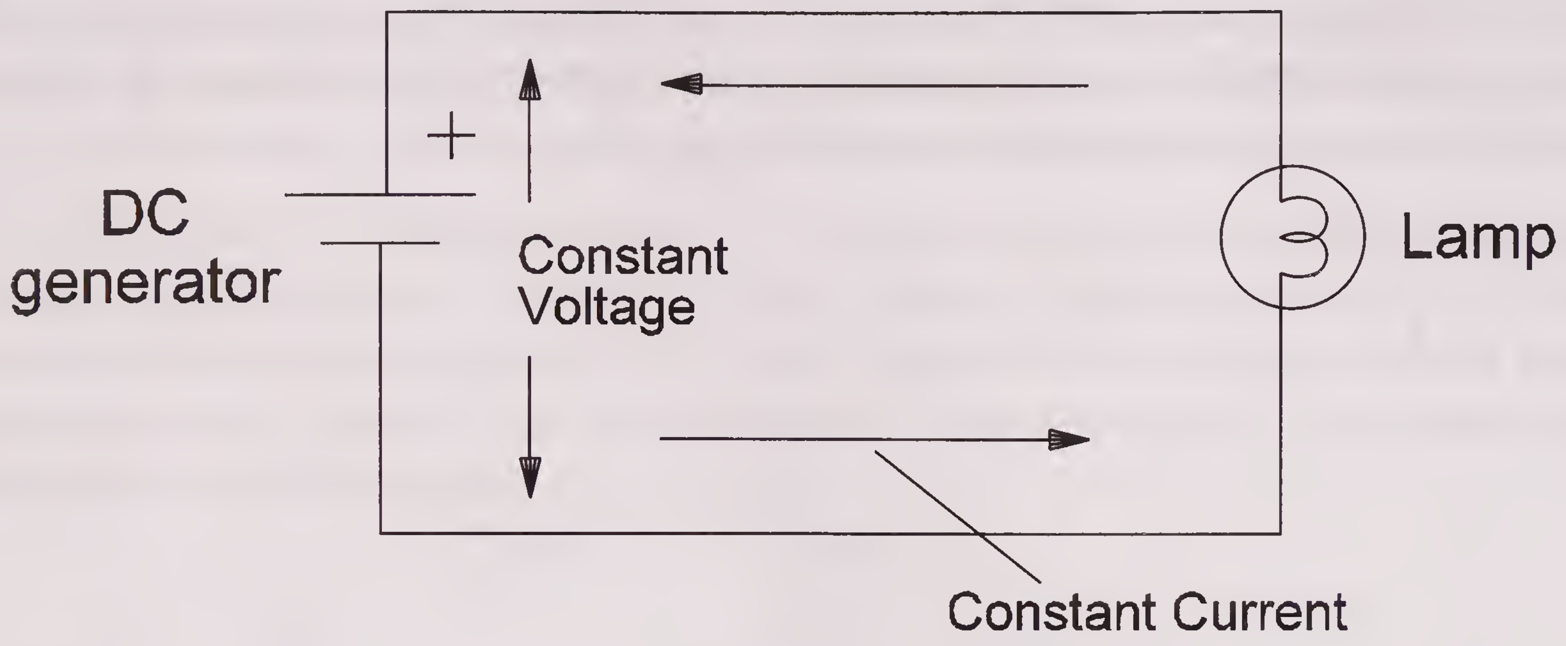
DC
generator
+
Constant
Voltage
Lamp
Constant Current

Figure 21

Through this circuit, the electrons can move only in one direction, transporting energy from the generator to a device or load. Cells, batteries, and dynamos are DC generators as the voltage found in their poles is constant.

On the other hand, the electrician who works with electric installations knows that the electricity supplied to buildings is not DC, but AC or alternating current.

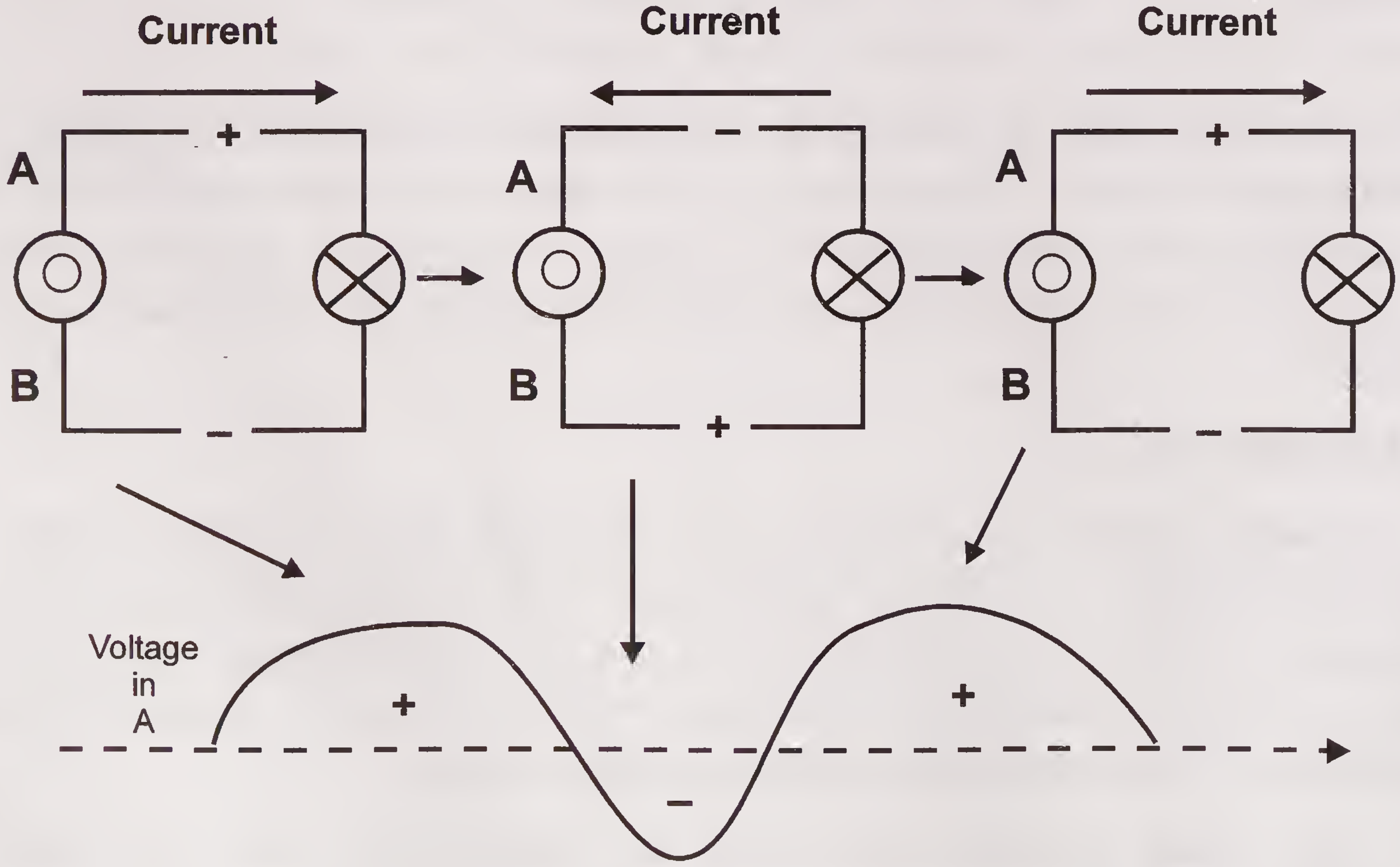
Current
Current
Current
A
B
+
–
A
B
–
+
A
B
+
–
Voltage
in
A
+
–
+

Figure 22

What is the difference and what does it mean? It is not necessary to have a constant flux of charges or constant current flow through a device to transport energy to it. Figure 22 shows a generator that changes the voltage in its terminals at high speed.

One moment pole A is positive and B is negative. The electrons are pumped through the device being used as load (a lamp, for example) releasing their energy and heating the filament. The lamp glows. An instant later, the poles of A and B are inverted. The electrons are now pumped in the opposite direction but must pass through the filament of the lamp again releasing energy. The lamp continues to glow. If the generator continues to change its poles, the electrons are pumped forward and backward through the circuit.

Observe that in this process, the charges move only a microscopic distance in one direction, then reverse motion so the charge retraces its path. The action is repeated continually so that the actual average movement of the charges is zero. But, a current flows because the charges are in motion. This is enough to cause an oscillating movement of the electrons in the circuit and to transfer energy from the generator to the load (any device that can absorb energy). This kind of current moving forward and backward is alternating current (AC).

The curious fact to be observed is, although the electrons move only a few centimeters each second, the force that pumps the electrons acts as a wave that moves near the speed of the light through the wires. In domestic installations the AC energy found in outlet jacks changes its poles 120 times by second, or 60 times in a second a pole is positive and 60 times is negative. We say that the frequency of the AC power line is 60 hertz (60 Hz). This is standard in most countries, but in some it is 50 Hz.

The alternating current found in the AC power line has a special waveform of variation represented by a curve shown in Figure 23.

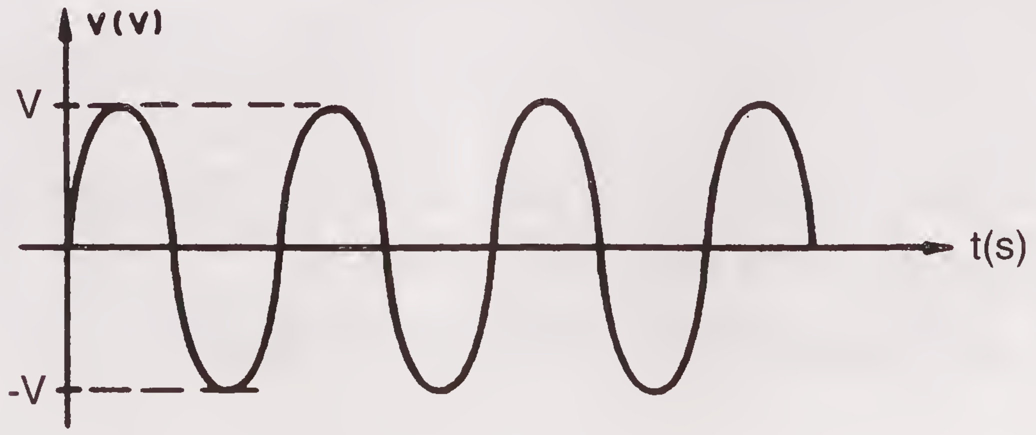

Figure 23

This is a the curve of a trigonometric function called sine and it shows that the voltage across of the poles of an AC generator changes slowly from the positive to the negative passing across the zero point, pumping the electrons forward and backward through a circuit.

When working with electronics, the electrician can find both DC circuits and AC circuits. The circuits powered from batteries or cells are DC circuits. The stages of these circuits are powered from DC sources but they can have some parts where AC currents are found. In general, AC currents ranging from low to high frequencies found in these circuits are called signals. They are used to transport some kind of information like sounds or images.

For the circuits powered from AC power, there are three situations to take into consideration:

1. Some use only AC in their circuits such as the ones used to control AC appliances (motors, lamps, etc).
2. Some use AC in the input but convert it to DC and then DC is used in almost all the internal circuits.
3. Some can use both AC and DC depending on the stage. And, in many stages we can find 60 Hz AC currents and signals ranging from a few hertz to hundreds or thousands of megahertz.

This means that the electrician working with electronic equipment must be prepared to find both AC and DC voltages independent of the used power supply.

h) Sound and Electromagnetic Waves

Many electronic circuits are designed to operate with sounds or electromagnetic waves (radio waves). Telephones, wireless alarms, electronic doorbells, and remote control systems are examples of electronic devices that work with sounds and/or electromagnetic waves.

When working with these devices it is important for the electrician to know something more about the "vibrations." Of course, many electric devices such as doorbells, lamps, and buzzers also work with light and sound, but in a more simple way than many electronic devices.

First of all, we must separate the two kinds of vibrations:

1. Sounds

When a metal bar is struck the vibrations produce disturbances in the surrounding air. These disturbances are caused by compression and decompression waves that propagate through the air in all directions, as shown in Figure 24.

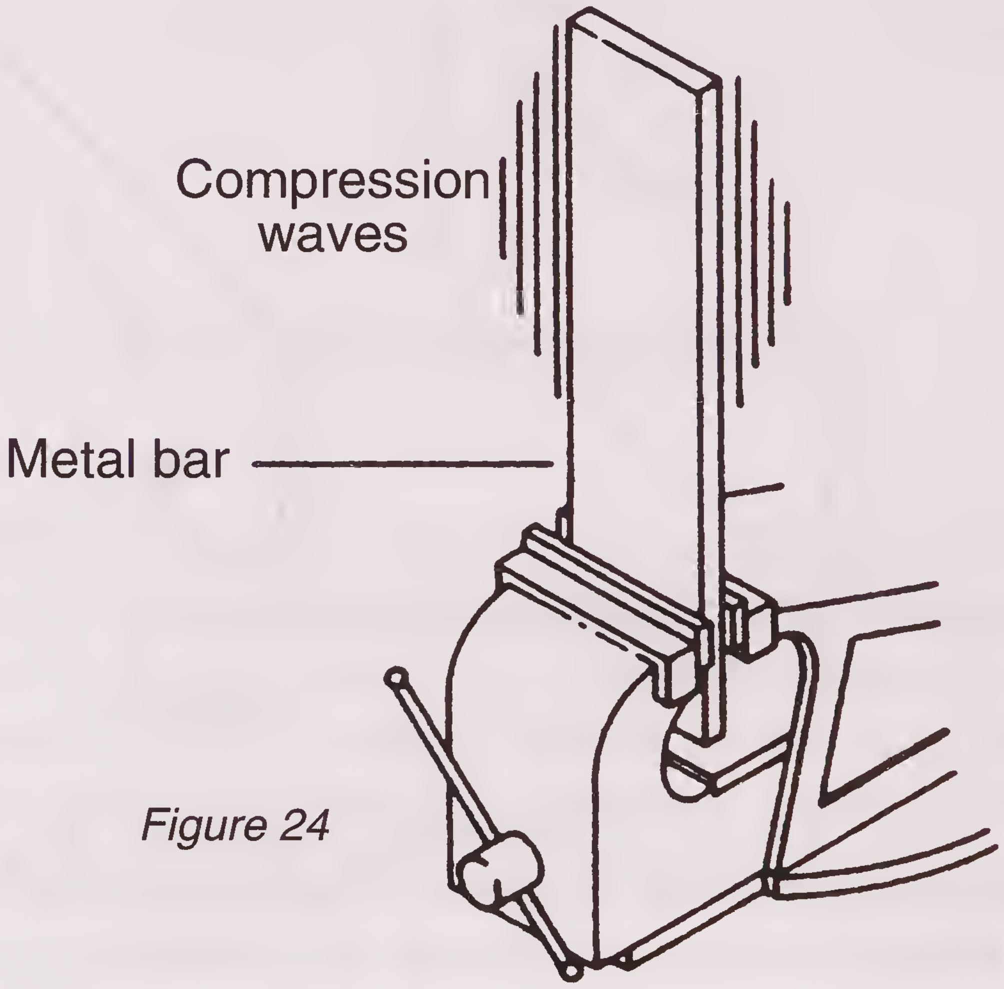

Figure 24

When reaching our ears these waves give us the sensation of sound. But, to be heard as sounds the disturbances or waves must have some special characteristics. First, the human ear can only perceive vibrations inside a limited range of frequencies or "speed" of vibrations. We can hear only the vibrations in the range of about 16 to 18,000 Hertz (vibrations per second). For example, if the bar vibrates at 30,000 Hz it produces sounds that are beyond the audible limit—ultrasonic vibrations. Many animals (dogs, bats, and dolphins) can hear ultrasonic sounds.

Another important characteristic of the sounds is that the human ear can differentiate them by their frequency. Low-frequency sounds, between 16 and 500 Hz, are perceived as bass. High-frequency sounds are perceived as treble. Our ears are sensitive enough to differentiate two sounds if their frequencies are separated of 1/16 of their value. It is because of this that we can differentiate two adjacent musical notes by the number of vibrations or frequency.

Also important is the way the metal bar is put into vibration. The natural way of vibration for a "perfect body" like a tuning fork is the one that produces a sine sound wave as shown in Figure 25.

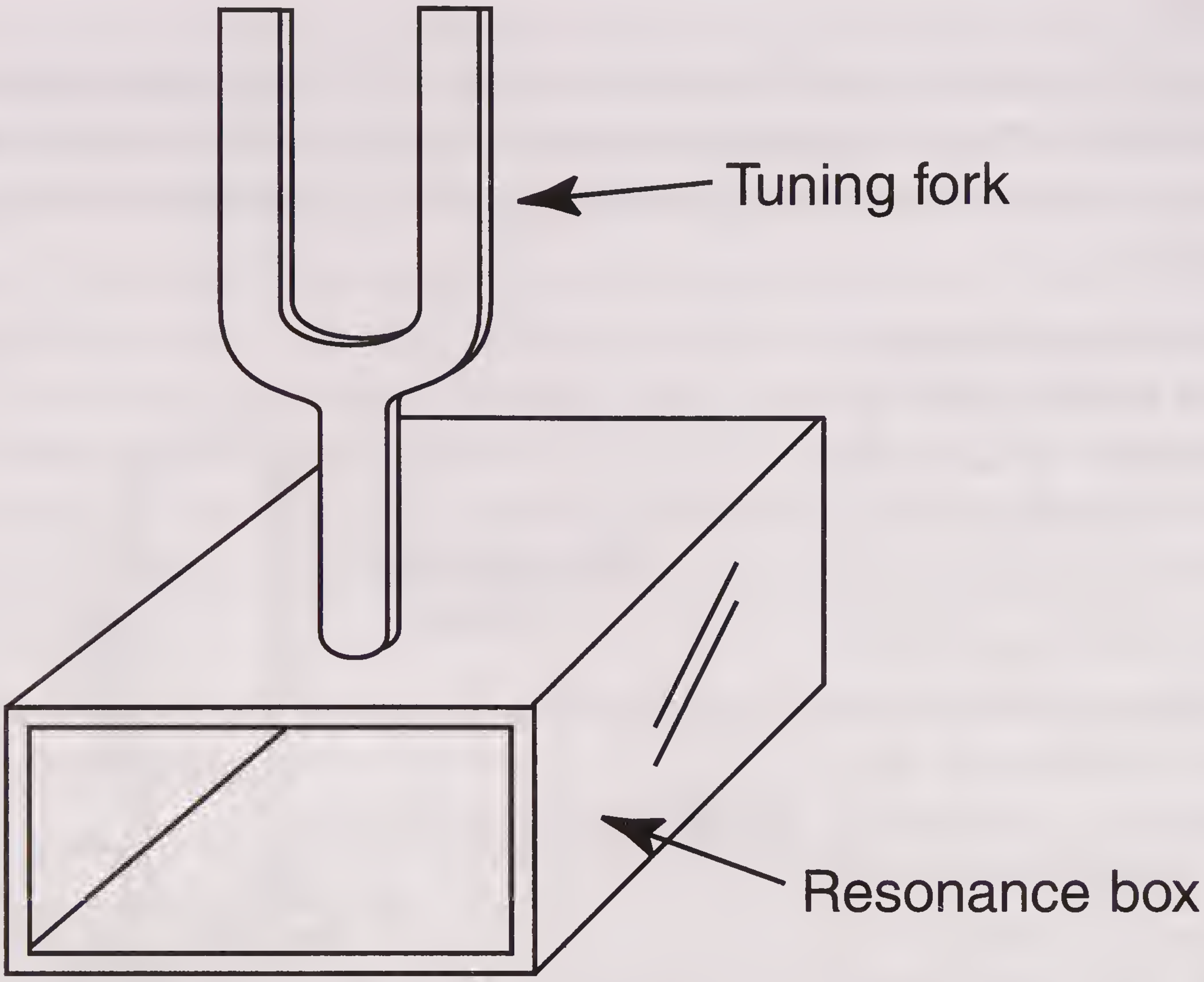

Figure 25

But, if the vibrations are produced by "imperfect bodies" or bodies with special formats such as musical instruments that have resonance boxes or special sizes and formats, the waveform can be different. The different waveforms give the sound a characteristic referred to as timbre or color. Timbre allows us to differentiate between the same tone or musical note produced by a violin and a piano; they have the same frequency but different timbre.

Sounds need a material to support their propagation. What this means is that they can't propagate in a vacuum. In the air, the speed of a sound wave is about 340 m/s.

2. Radio or Electromagnetic Waves

It was in 1865 that the physicist and astronomer James Clerk Maxwell published a theory explaining the existence of electromagnetic fields or electromagnetic waves.

To explain this, begin with the idea that a standard electric charge is surrounded by an electric field and a moving electric charge is surrounded by a magnetic field as shown in Figure 26.

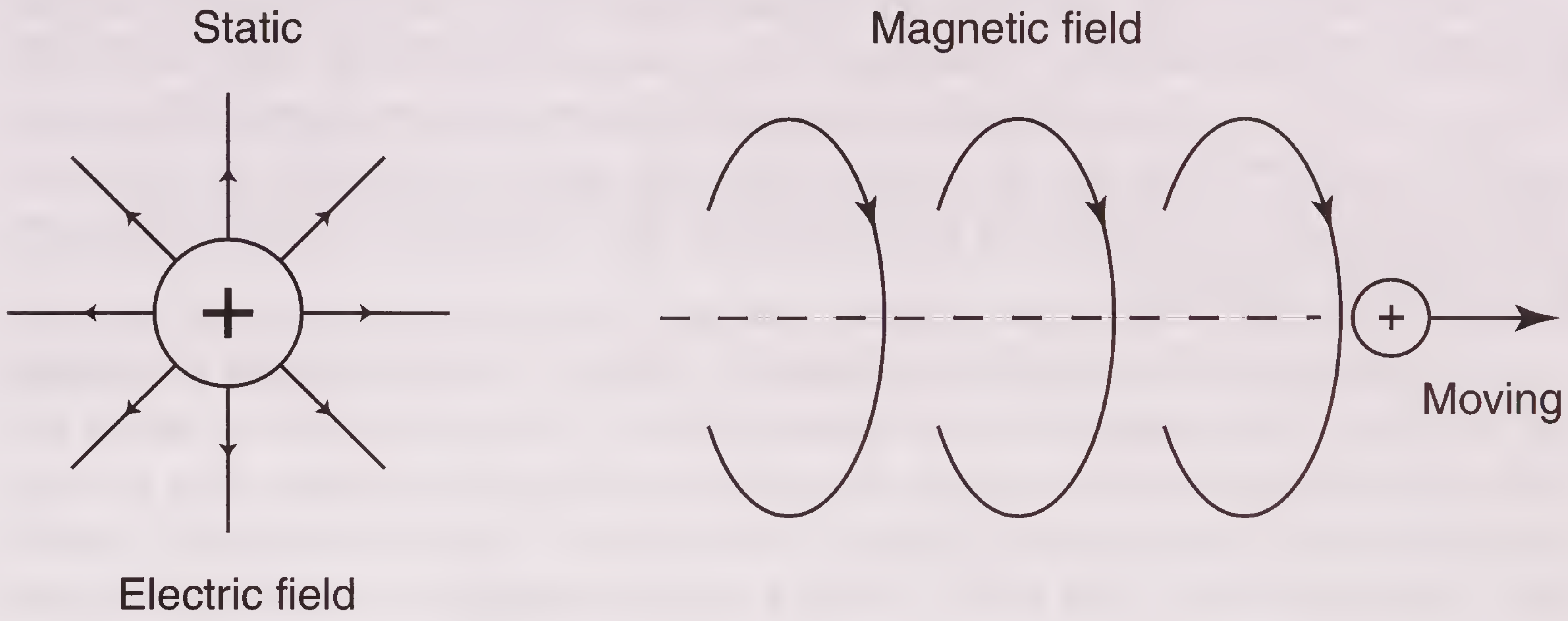
Static
Electric field
Magnetic field
Moving

Figure 26

From this, a simple imaginary experiment can be done to show Maxwell's theory using a row of swinging metal balls (often sold as desk-top toys) as shown in Figure 27. If metal balls are placed end to end in a row and then one end is struck by another moving ball, as shown in (a), the motion of the striking ball is transmitted through entire row. There is little loss as the motion is passed to the ball at the other end of the row as shown in (b).

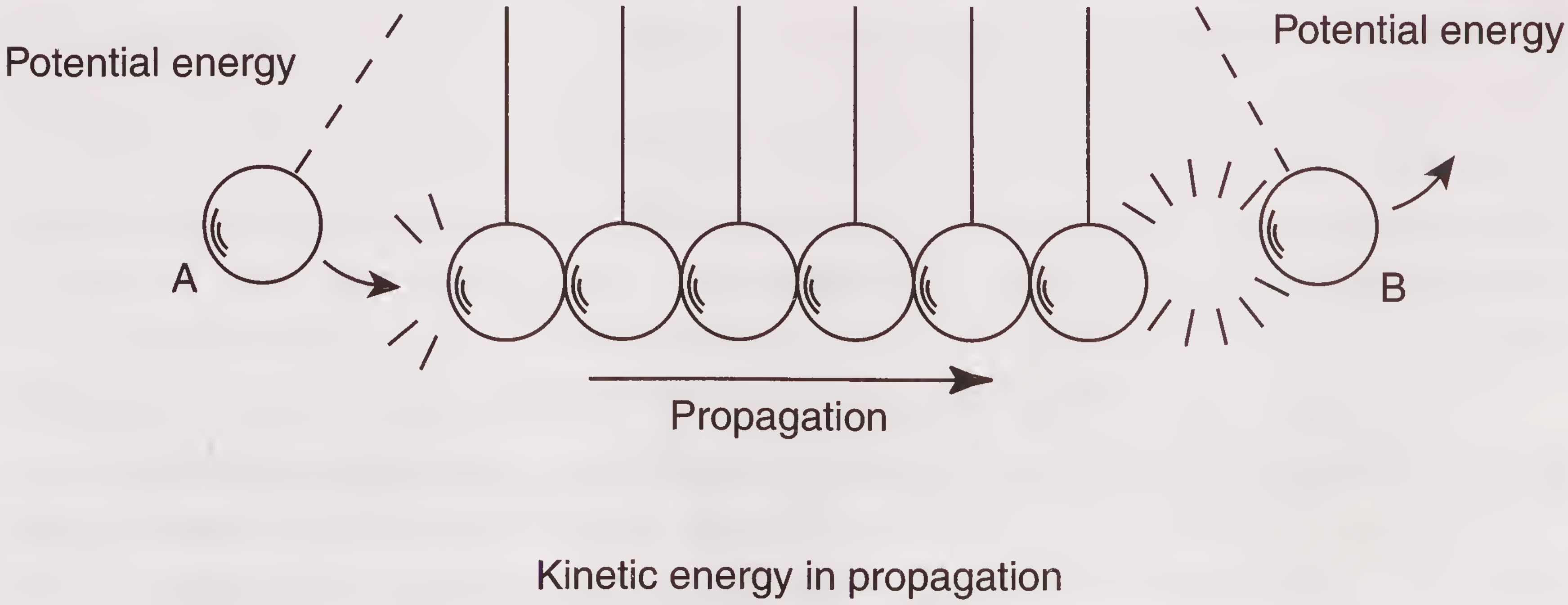
Potential energy
Potential energy
A
B
Propagation
Kinetic energy in propagation

Figure 27

Maxwell showed with this process that a detectable time delay exists between the departure of energy from a source (striking ball) and that energy's arrival to the last ball, during the interval the energy exists in the moving medium (the row of metal balls). He also concluded that the energy exists half in the form of the motion of the medium and half in the form of an elastic reliance.

In the case of an oscillating electric charge, the two forms of energy are stored in the magnetic and electric fields, correspondent to what we call potential and kinetic energy. So, when a charge oscillates, the energy is transferred to surrounding space in the form of electromagnetic waves that combine in the two fields. These electromagnetic waves can travel through space with the speed of light or 300,000 km/s (as we will see, light is an electromagnetic wave).

Electromagnetic waves are present in our world in many forms. Because electromagnetic waves can be produced in any frequency from zero to infinity, they are represented by a continuum or spectrum as shown in Figure 28.

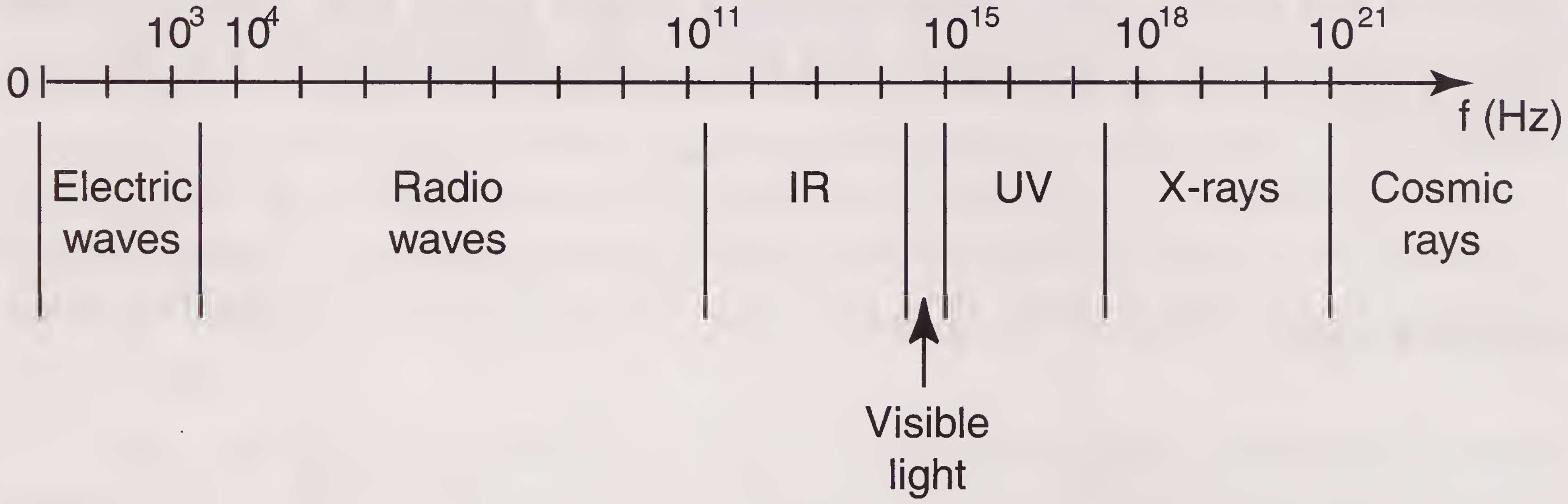

Figure 28

The lower part of the diagram is where radio waves used in broadcasting are found, as well as telecommunications, radar, cellular telephones, TV, remote controls, etc. A little further up is the IR or Infrared part of the spectrum. This corresponds to heat waves. Radiation in this part of the spectrum is due to thermal agitation of an object's atoms. The next segment consists of visible light with the colors, followed by UV waves (ultraviolet radiation). At the upper end of the spectrum are X-rays, gamma rays, and cosmic rays with very high frequency and energy. For the electrician who wants to work with electronics, the most important parts are those corresponding to radio waves, infrared, and visible light.

Two important quantities are associated with electromagnetic waves—wavelength and frequency. Any electromagnetic wave can be represented by sine curves corresponding to the electric and the magnetic fields as shown in Figure 29.

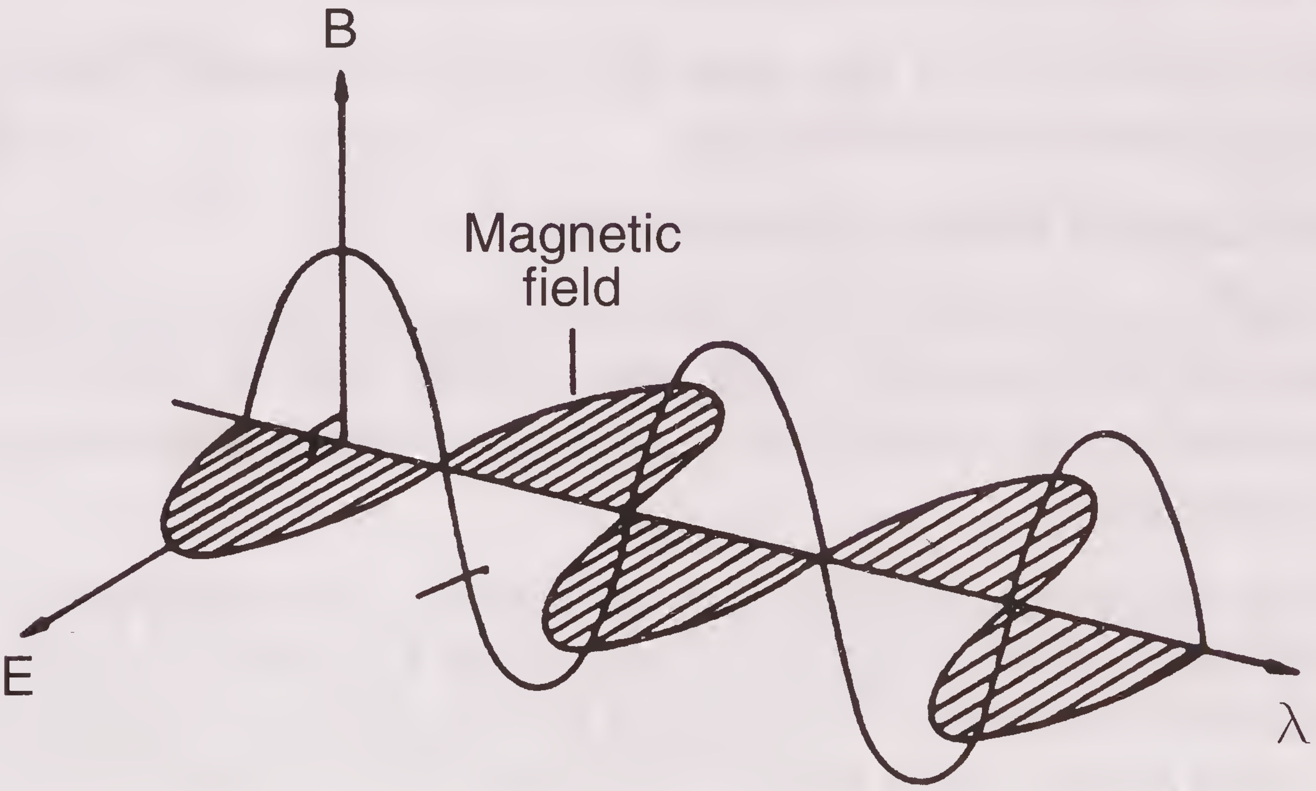

Figure 29

The frequency is the number of waves produced in each second and is measured in hertz (Hz). As high frequencies are common in electronic circuits, the use of the prefixes kilo (k), mega (M), giga (G), and tera (T) are common. When starting from the source, the distance traveled during one cycle can be defined as the wavelength (λ). This distance depends on the speed of the electromagnetic wave that, in a vacuum, is 300,000,000 m/s (or 300,000 km/s). So, to calculate the wavelength of an electromagnetic wave, divide 300,000,000 by the frequency in hertz.

If, for example, the wavelength of a 100 MHz (100,000,000 Hz) signal of an FM radio broadcast station is:

λ= 300,000,000/100,000,000 = 3 meters

Knowing the wavelength of a radio wave is very important since it determines the characteristic of the antenna used to transmit or receive it.

The sizes of the elements of any radio antenna depend on the wavelength of the signal it is designed to use.

The fundamentals of electricity are the same that command electronic circuit operations. In electronic circuits, AC and DC currents pass through components and use all the common effects as you most likely studied before. Light, heat, magnetic fields, motion, images, and sounds are generated by devices that operate based on principles found in any electric devices or are explained by the physics principles.

For the electrician, the next step is to know what tools must be used when working with electronic circuits.

i) Printed Circuit Board and Solder

The small components used in electronic equipment can't stand alone without any physical support. They need some kind of support to keep them fixed while at the same time providing an electrical connection with the rest of the circuit.

Opening any piece of electronic equipment, the electrician will find that the electronic components are mounted on a special board of fiber or another insulating material. This support or chassis for the components is called a printed circuit board (PCB). The board, as shown in Figure 30, is made of an insulating material where copper strips are printed on one or both sides.

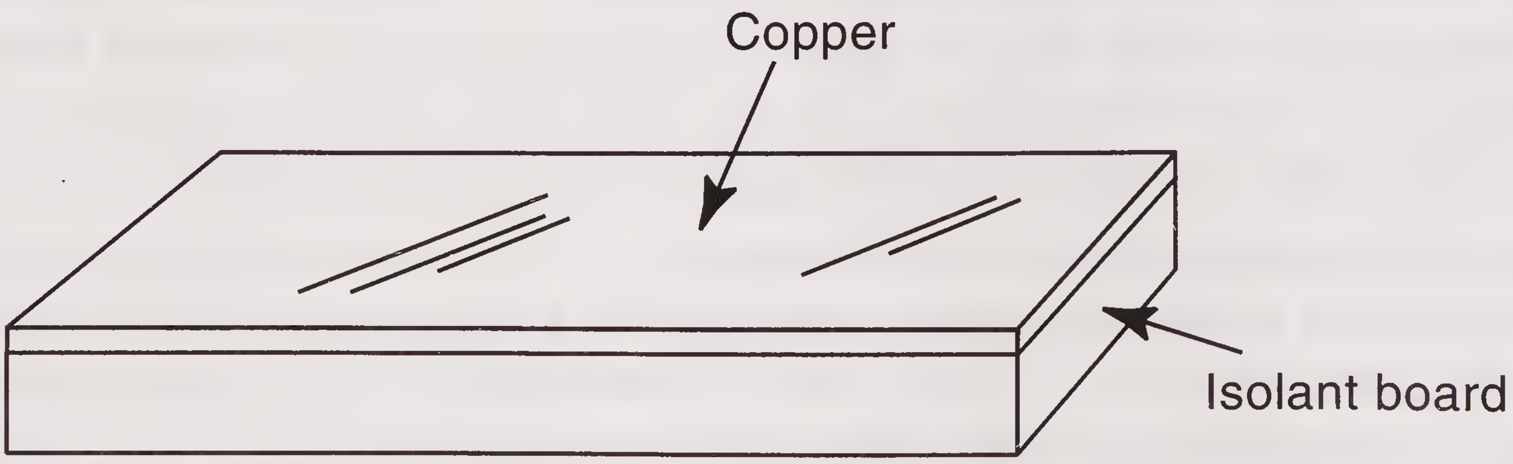

Figure 30

The strips act as wires, conducting the currents from one component to another. The pattern of the strips is determined by the function of the circuit. All the strips are planned before the manufacturing process to provide the necessary connections between components in the desired function. This means that a printed circuit board produced to receive components that form a radio can't be used to mount TV circuits or any other equipment.

As shown in Figure 31, the small components are soldered onto the board so their terminals make contact with the copper strips.

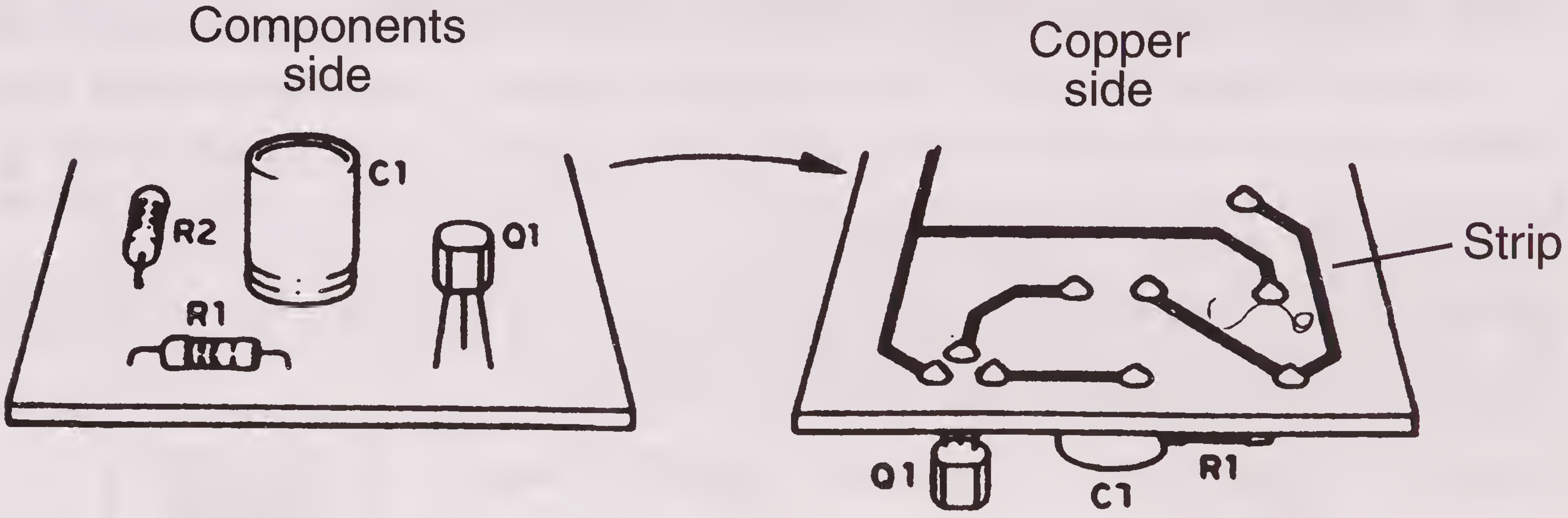
Components side
C1
R2
Q1
R1
Copper side
Strip
Q1
C1
R1

Figure 31

In some cases, when the components are very small as in surface-mounted devices (SMD), they can be placed on the board at the same side of the strips as shown in Figure 32.

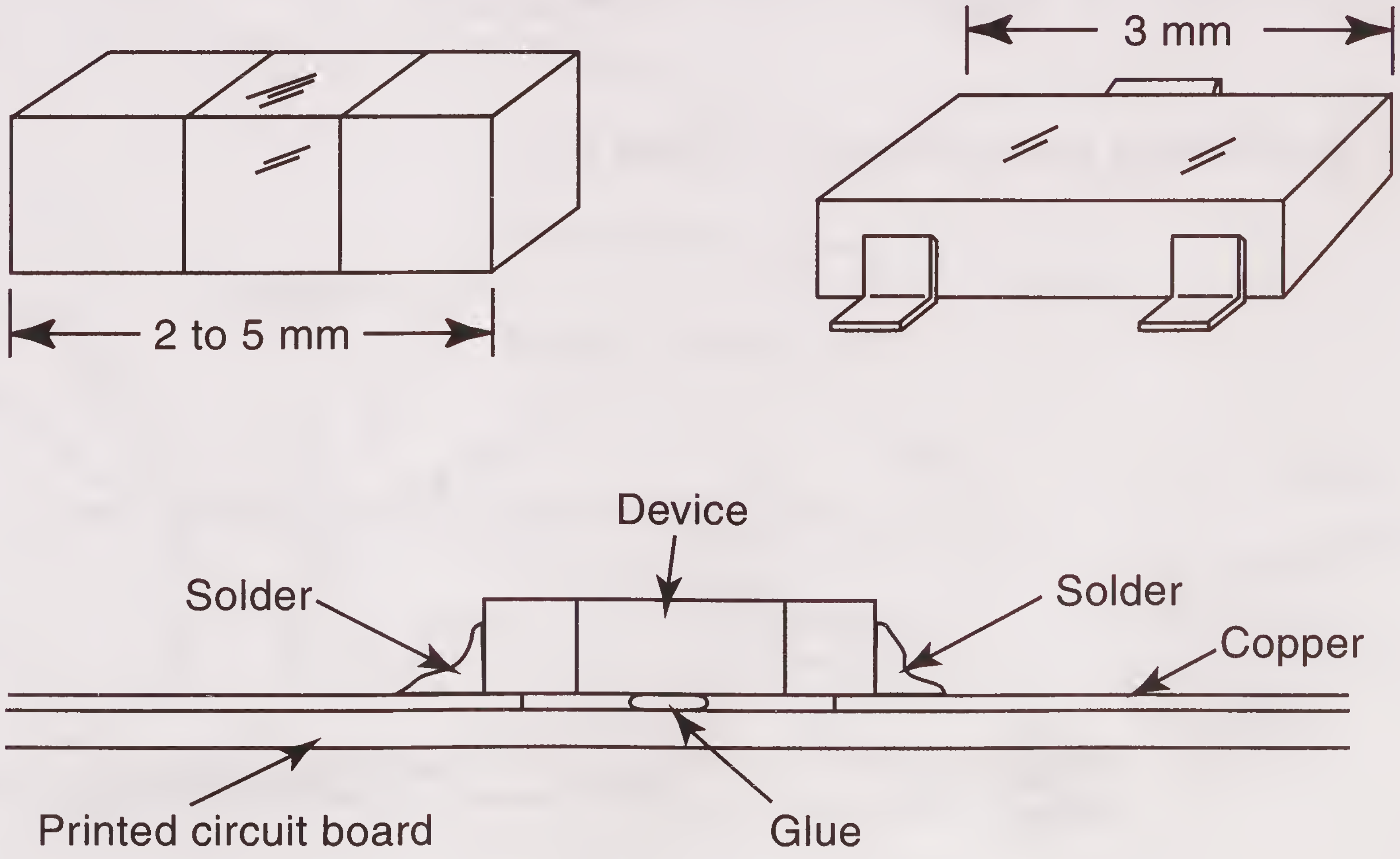
3 mm
2 to 5 mm
Device
Solder
Solder
Copper
Printed circuit board
Glue

Figure 32

Removing and reinstalling components on a printed circuit board is a delicate operation that an electrician who wants to work with electronics, to repair and mount circuits, must know. Special techniques are required for this work because the components are so small and extremely sensitive to heat. The solder used in electronics is an alloy formed of 60 percent tin and 40 percent lead with some rosin. It is common to call this kind of solder transistor solder, radio-TV solder, or 60-40 solder.

When heated to about 273 degrees Celsius, the solder melts and involves the terminal of the component which fixes it to the board, while at the same time providing electric contact with the copper strips or other components. When working on component mounting or replacement, an electrician will need some solder and a soldering iron. The solder can be bought in small quantities as shown in Figure 33.

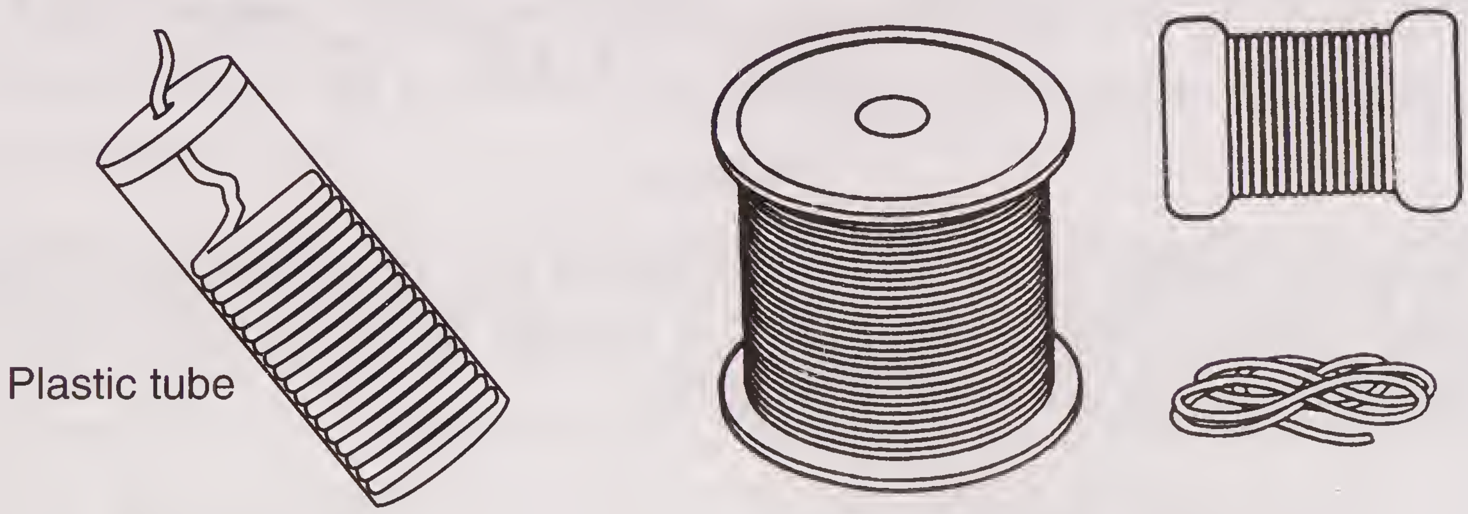

Figure 33

A soldering iron is shown in Figure 34.

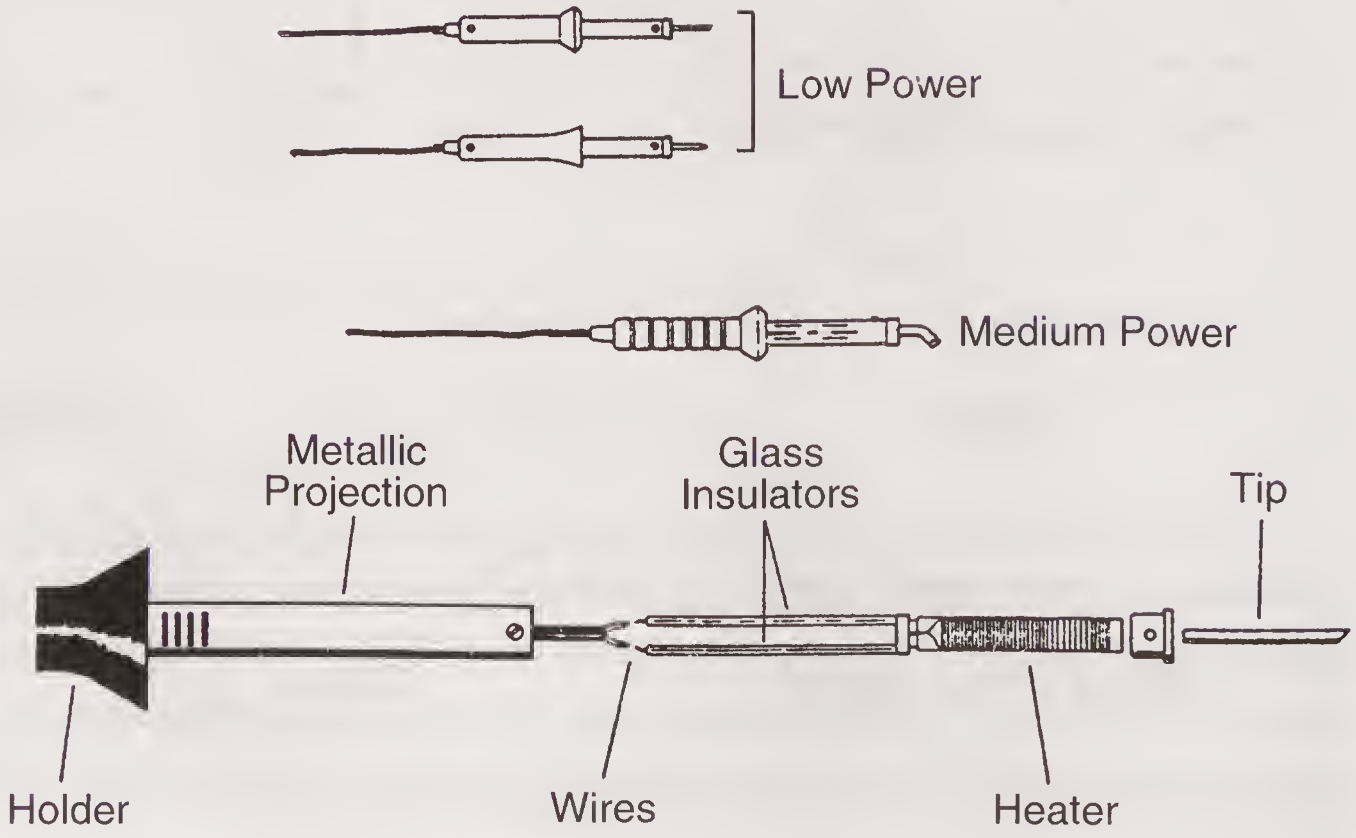

Figure 34

A 25- to 40-watt soldering iron with a shiny tip is recommended when working with small electronic components found in circuits. Of course, the electrician can also have a heavy-duty soldering iron to remove or place larger components like those found in some electric and electronic applications. Radio Shack has soldering irons suitable for this task as well as the next task:

- Dual Wattage Pencil Iron, 15 to 30 W (Radio Shack 64-2060)
- 15 W Pencil (Radio Shack 64-2052)
- 25 W Pencil (Radio Shack 64-2072)

Soldering is a simple operation and many electricians are familiar with this kind of work. But, electronic devices are very special and care should be taken when soldering to avoid damaging them. Many electronic devices are easily damaged by excess heat or the incorrect soldering procedure. The basic procedures to solder electronic components (remove or install on a PCB) are as follows:

- Plug in the soldering iron and heat for at least five minutes. This is long enough to bring the tip to the correct temperature for good soldering operation.
- Touch the soldering iron to the work area, allow a very short time for the connection to heat up, and then touch the solder to the connection, not the iron, as shown in Figure 35.

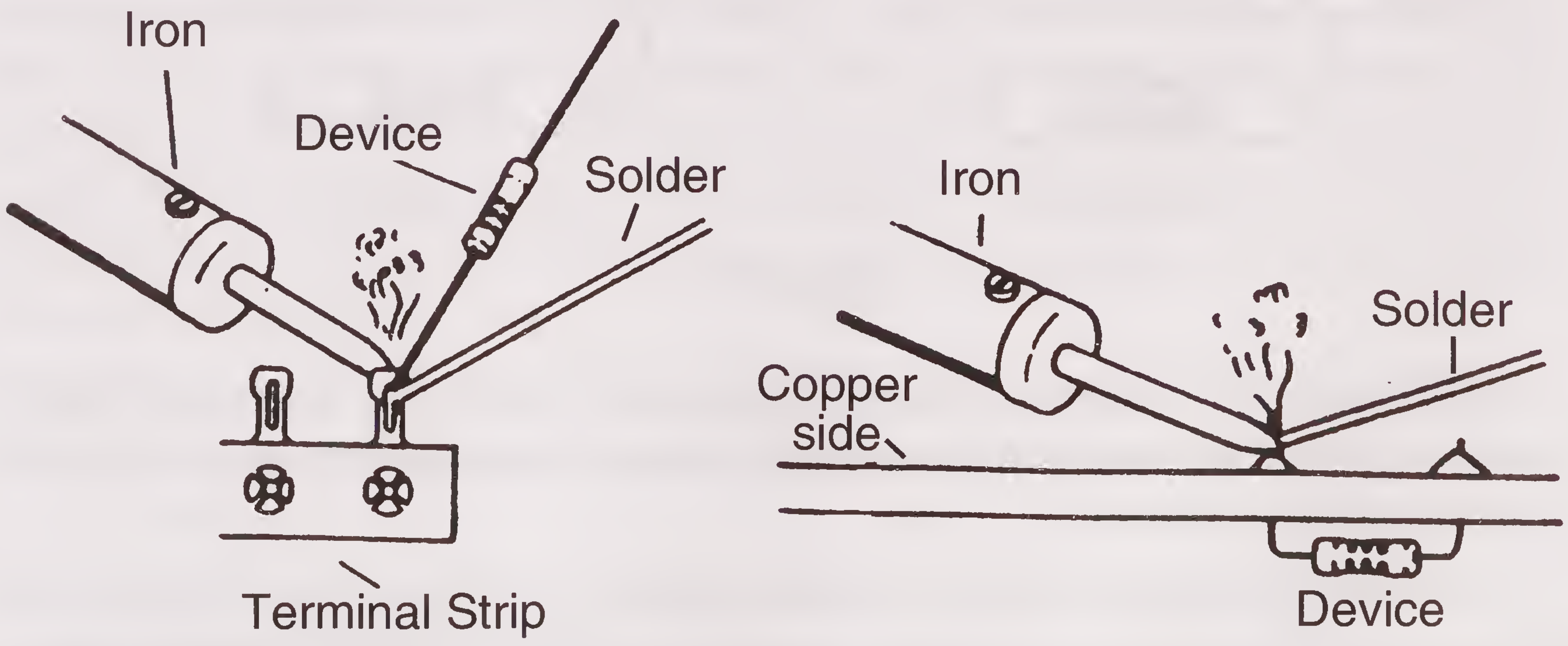

Figure 35

- You will notice that when the solder melts it penetrates every part of the solder joint.
- Remove the iron and do not move the joint until it has had time to cool. It is easy to see when the joint is cool. A peculiar haze will pass over the hot metal after which, the joint is cool enough and strong enough to withstand movement.

Figure 36 shows a perfect solder joint and some solder joints with problems. One of the principal causes of problems in electronic equipment is "cool solder." The solder seems to involve the component, but no electric contact is established because the joint was not heated enough to penetrate the metal creating an isolated layer of moisture or oxide to form between them.

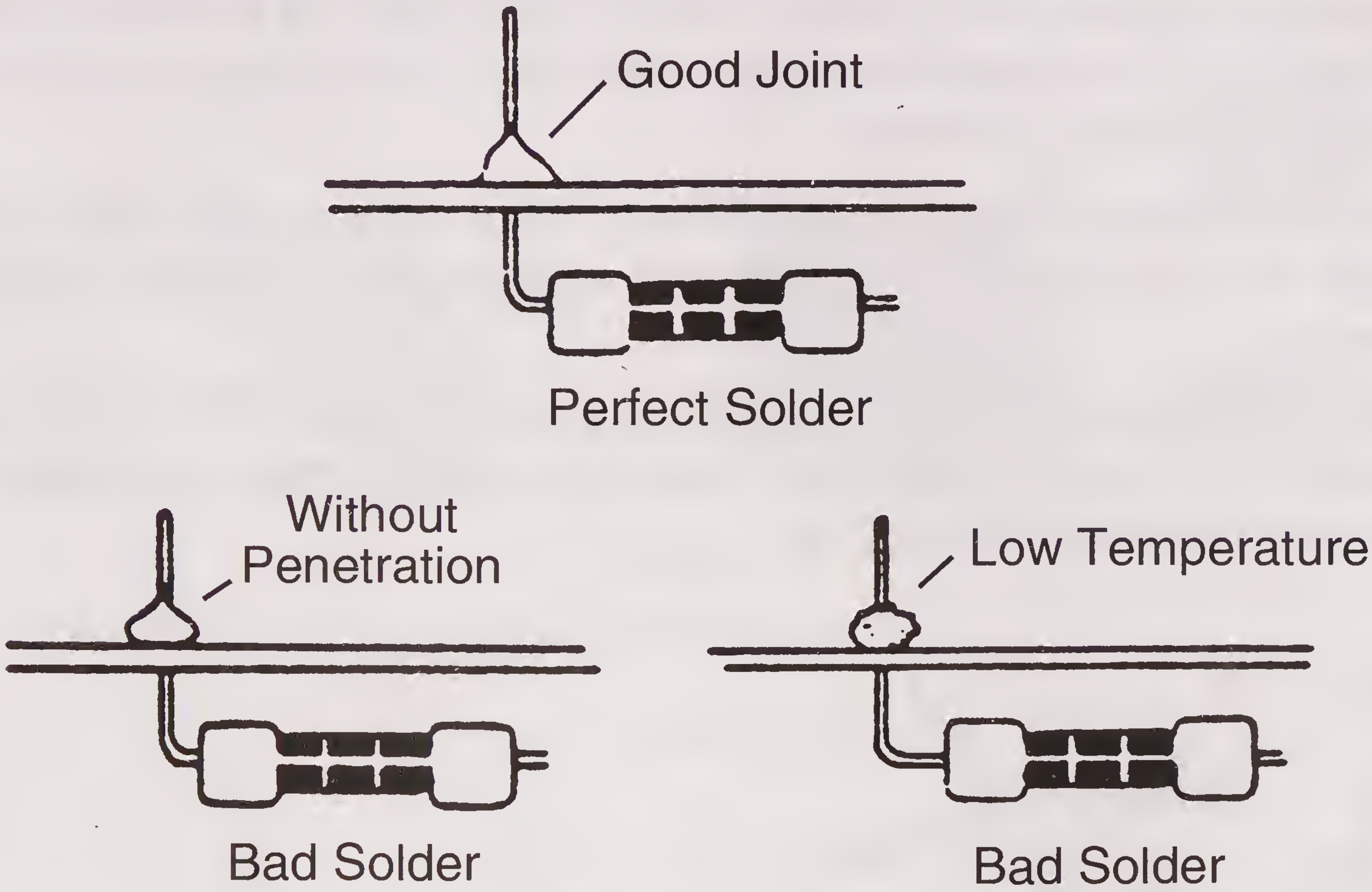

Figure 36

Although the use of the soldering gun, such as the one shown in Figure 37, is not forbidden, there are many cases in electronic works in which it is not recommended.

This type of soldering tool has its resistive tip heated by a strong current coming from the low-voltage winding of a transformer. When working with sensitive components such as integrated circuits, transistors, and others during the soldering operation, this current can be enough to burn them.

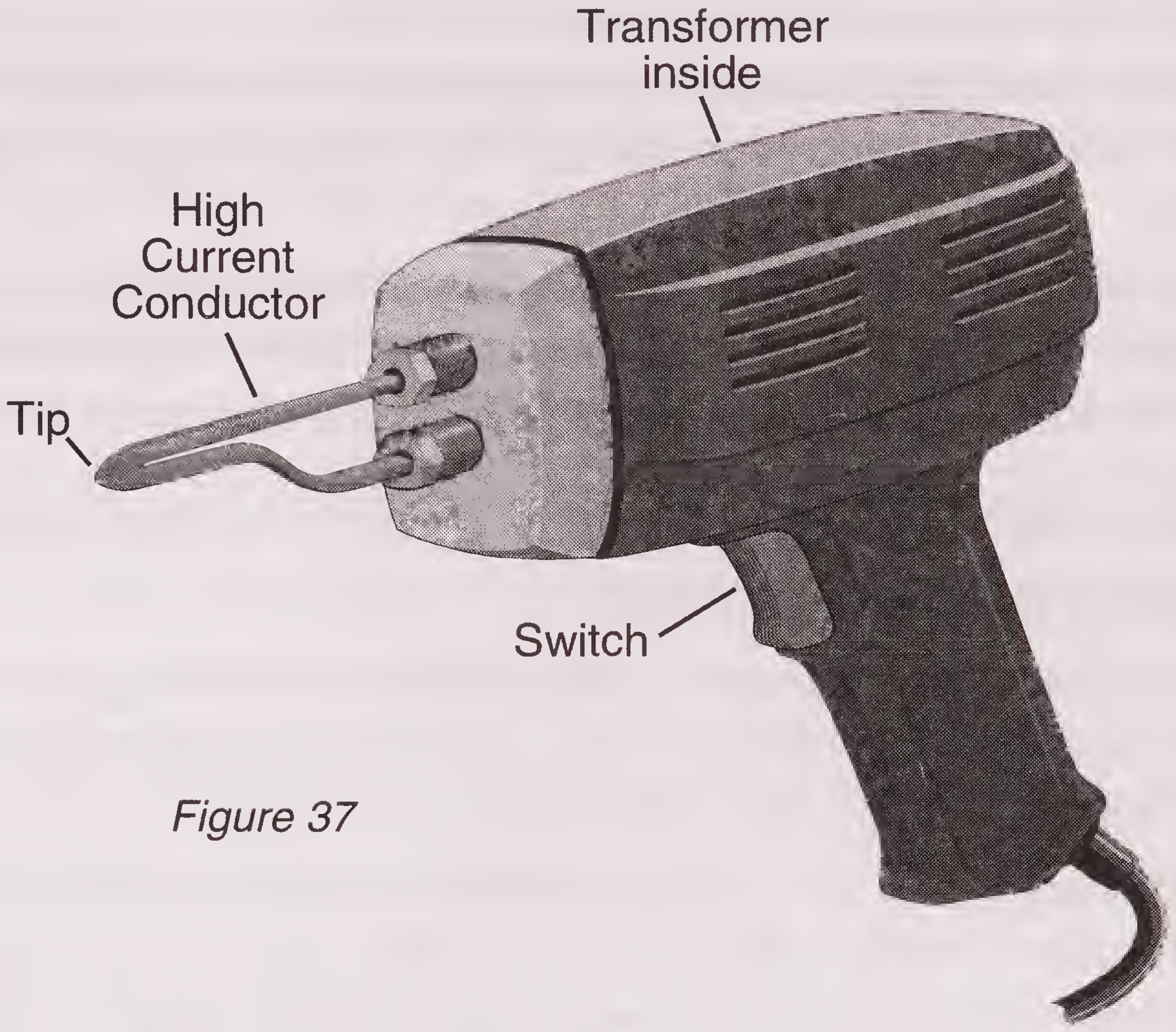

Figure 37

j) Other Tools

The soldering iron isn't the only special tool needed by electricians who intend to work with electronic circuits.

When working with electronic parts, many of the tools used in electric installations or automotive electricity are suitable. On the other hand, many electronic components are very small and delicate and need special tools and care. The use of improper tools when working with these components is often the cause of damage. If an electrician intends to work with electronic circuits, some special tools are needed. I suggest having at least some of these tools:

- Cutting pliers or diagonals (often called dykes) in sizes from 4 to 6 inches.
- Chain-nose or needle-nosed pliers with very narrow tips in sizes from 4 to 5 inches.
- Two or more screwdrivers sized between 2 and 8 inches. (Seven-piece set: Radio Shack 64-1823).
- Crimping tools—stripper and cutter for 10 to 22 wire gauges (Radio Shack 64-2129).

- Precision tool set (10 to 16 tools) with small screwdrivers of hex, common, and Phillips types (Radio Shack 64-1948).
- Soldering and desoldering accessories such as a desoldering bulb (Radio Shack 64-2086) and a soldering iron holder/cleaner (Radio Shack 64-2078).
- Extra hands to hold the work in the form of a mini-vise with vacuum base (Radio Shack 64-2094) or a Project Holder (Radio Shack 64-2093).
- Mini hand drill (five-piece hand drill, Radio Shack 64-1779).

Many other tools can be found in electronic and tool catalogs.

k) Diagrams and Symbols

As in the case of electric installations in buildings or automotive installations, schematic diagrams are used to represent the way the many parts of equipment are interconnected.

There are many differences between an electric diagram and an electronic diagram. These differences start with the symbols used to represent components and end with the way they are connected. These differences are one of the problems to be solved by an electrician wanting to know something about electronics. He needs to be acquainted with new symbols since there are many electronics components represented by symbols that the electrician doesn't know. Learning the symbols used to represent each component and what they do in a circuit are the next steps in this book.

Let's begin with the schematic diagrams.

Figure 38 is a diagram of a simple piece of electronic equipment: a DC power supply similar to those used to power small calculators, CD players, and other appliances from the AC power line.

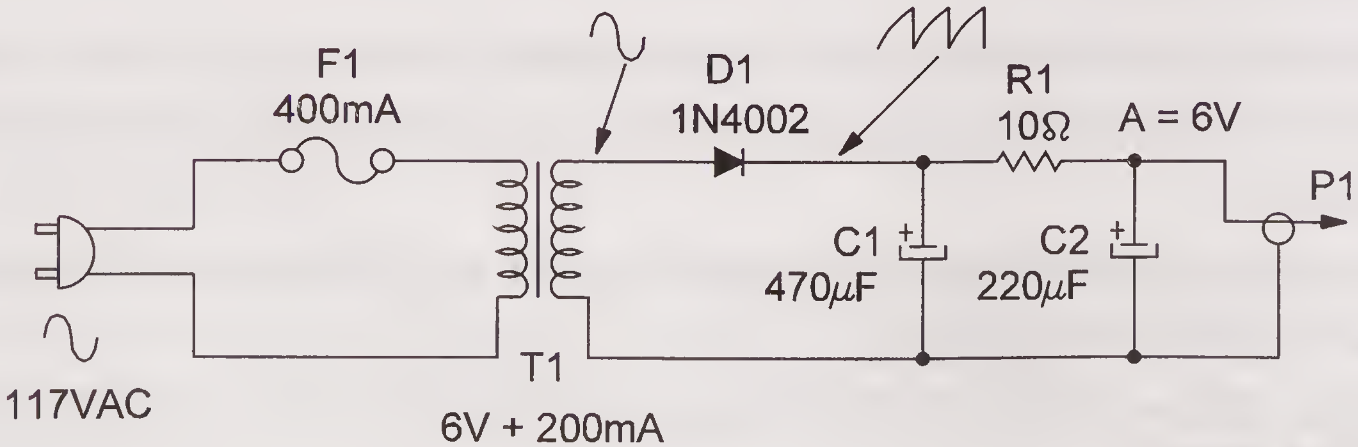

Figure 38

This schematic diagram represents all components by their symbols, and in many cases, the identification, values, and other important information are given. At the side of the symbol of each component is the identification number. This is important since it can help the technician find the component on the PCB or inside the equipment.

For example, all the resistors are identified by the letter R followed by the number of the device in the equipment. This means that many resistors can be found in a device identified by R1, R2, R3, etc. Capacitors are usually noted by the use of the letter C. The capacitors of a circuit are numbered starting from C1, C2, C3, etc. Transistors can be identified by the letters Q, T, or TR. They can be represented as Q1, T1, or TR1.

In many cases, a second number can identify the “block” or a “stage” in which the component is placed. So, the resistors of the first stage can begin at R101 and the resistors in the second stage from R201. Near the identification and the symbol we can also find the value or type of the component.

Resistors have the value of resistance to the side, such as R1, 1000 ohms or 1k. If it is a transistor, you might find 2N3906, meaning that when used, the transistor must be replaced with a 2N3906. The usual identification number for transistors begins with a 2N, but many manufacturers have started using letters signifying their names such as TIP (Texas Instrument) and MPS or MM (Motorola). There is also a European code that uses the letters BC or BD in the identification of devices, and a Japanese configuration that uses the letters 2SB, 2SC, or 2SD to indicate transistors.

Depending on the circuit, other important information can be found in the schematic diagram. For example, the voltage at different points in the circuit can be found.

In the example presented, it is indicated that, between A and the ground (normally taken by reference or 0 V), the measurement is 6 V when a multimeter is used to take the voltage (Figure 39).

Another important piece of information found in a diagram is the waveform of the signal present in a point.

As we said, electronic circuits don’t operate only with DC voltages but also with AC voltages in a large range of frequencies and waveforms. These AC voltages or signals are not only sine signals, but can have several

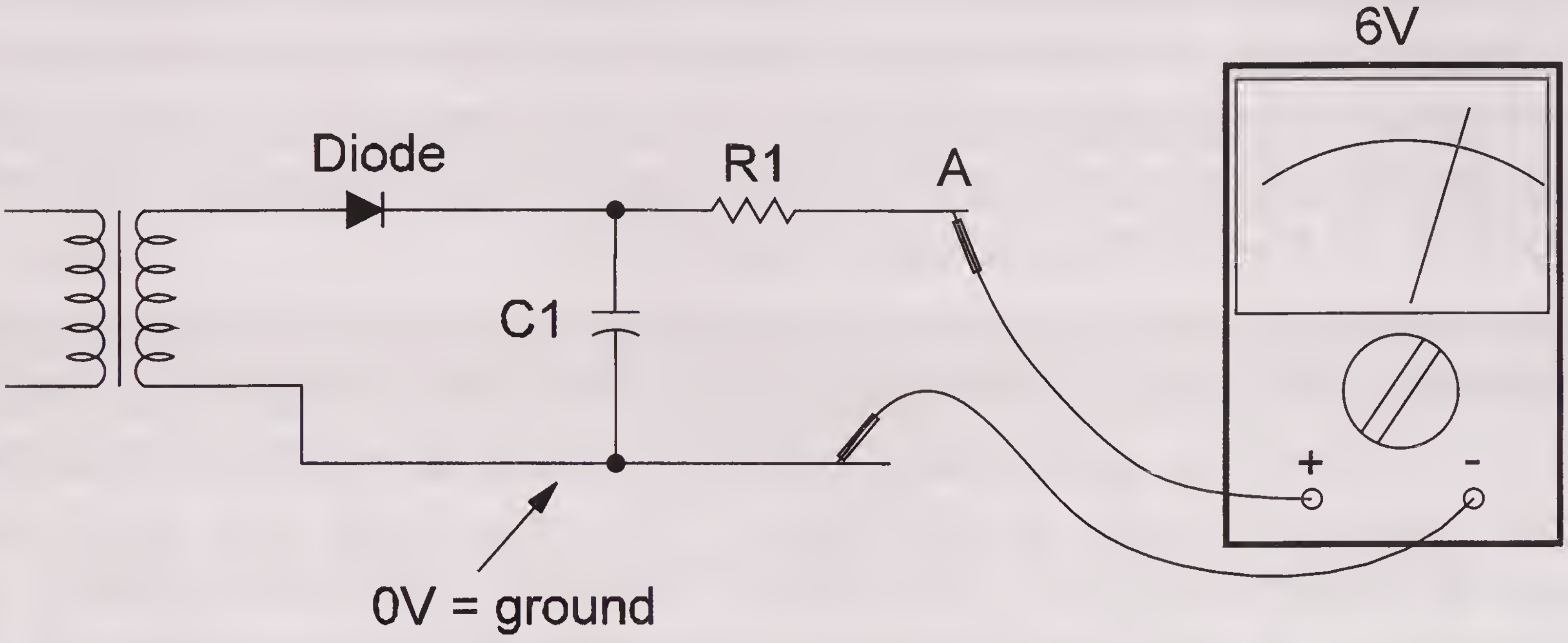
6V
Diode
R1
A
C1
+
-
0V = ground

Figure 39

waveforms and amplitudes. Oscilloscopes are used to observe waveforms of a signal in a circuit. Figure 40 shows a common oscilloscope used on electronic workbenches for troubleshooting and diagnostics.

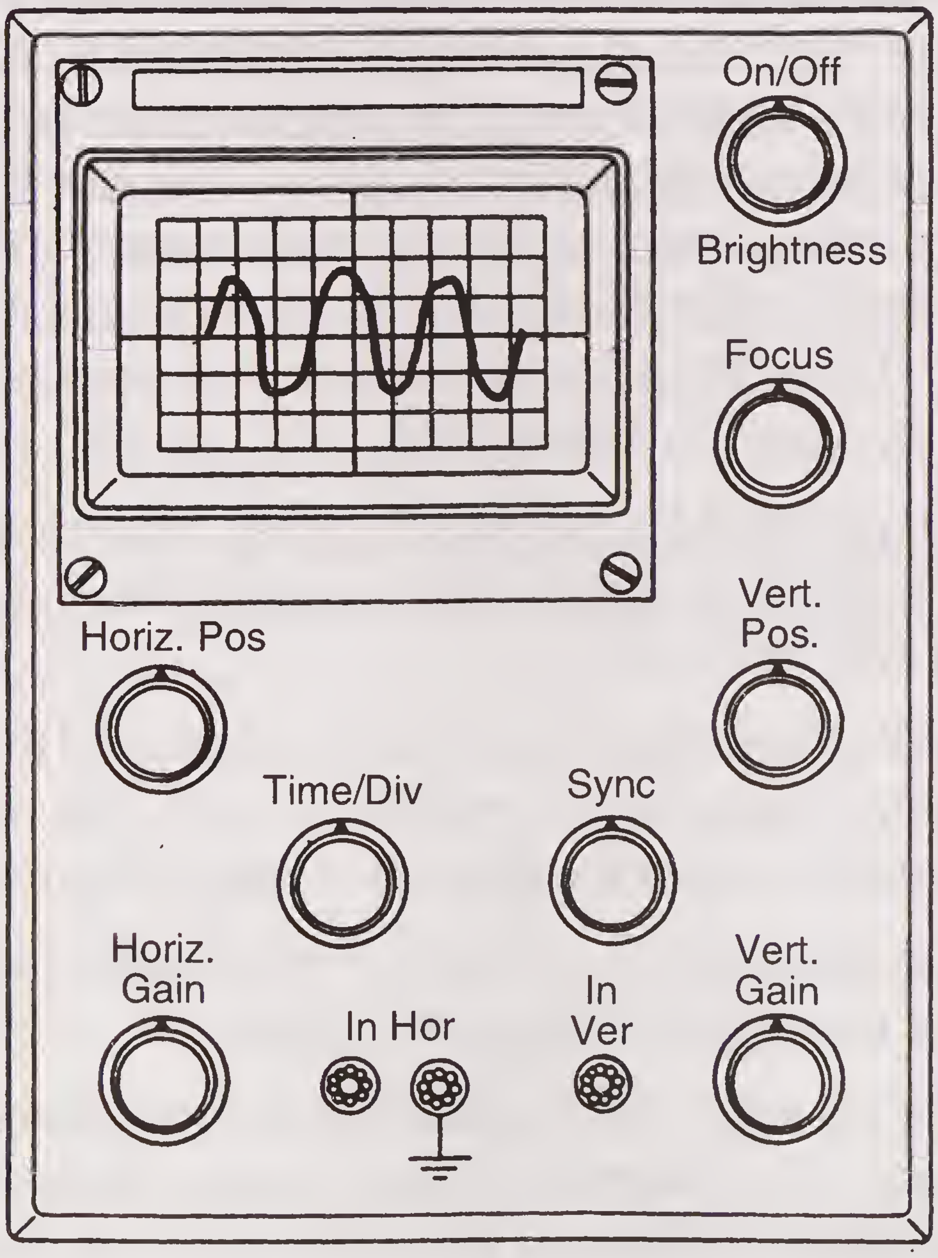
On/Off
Brightness
Focus
Vert.
Pos.
Horiz. Pos
Time/Div
Sync
Horiz.
Gain
In Hor
In
Ver
Vert.
Gain

Figure 40

By observing the waveform of a signal at a point in a circuit a technician can determine what is wrong. The oscilloscope is one of the most powerful tools in the electronic lab. In the schematic diagram example there is the indication of a waveform at point B.

Also found on a schematic are procedures for installation, diagnostics, equivalence, etc. Diagrams for many types of equipment can be found with their manuals, but this is not common. A solution for the electrician who has problems with a piece of equipment and needs the schematic diagram is to purchase one. Sams Photofact is the most important supplier of schematic diagrams for commercial electronic equipment. You can access the company's Web site at www.samswebsite.com.

Section 2

Electronic Components

The parts found in electronic equipment are very different from the ones the electrician finds in electric installations of buildings and cars. Knowing how these electronic components work, how they are used, and how to read and interpret their specifications is an important item to everyone who intends to be an electronic technician. It is also very important to know how these parts can be tested.

So, as a brief introduction of electronics to the electrician, the principal electronic components are described. These are found in some equipment for use in domestic installations and also the automobile. From this description, the electrician will be able to recognize these components when working with electronic equipment. This knowledge is also very important as a starting point for a basic course in electronics.

Wires

Wiring material is used to connect the circuit parts, devices, equipment, etc., electrically. There are various types of wiring material for use in electronic applications. The wires used in electronic equipment have some differences from the ones found in electric installations. The current flow, in general, is lower in electronic appliances, which need thinner wires, but some devices use special wires. Basically the electrician will find three types of wires in electronic work:

A. Cable—formed by a core of several copper conductors and covered by plastic insulation.

B. Rigid—formed by a core with only one conductor covered by plastic insulation.

C. Bare—formed by a copper conductor without cover (insulation).

D. Screened and Coaxial—formed by a core with one or more plastic covered conductors. These conductors are enveloped by a copper screen. Over the entire cable there is a plastic cover.

The plastic-covered wires can be found in several forms including single wires, twisted wires, flat cable, etc. Figure 41 shows some types of wire used in common wiring of electronic circuits and parts.

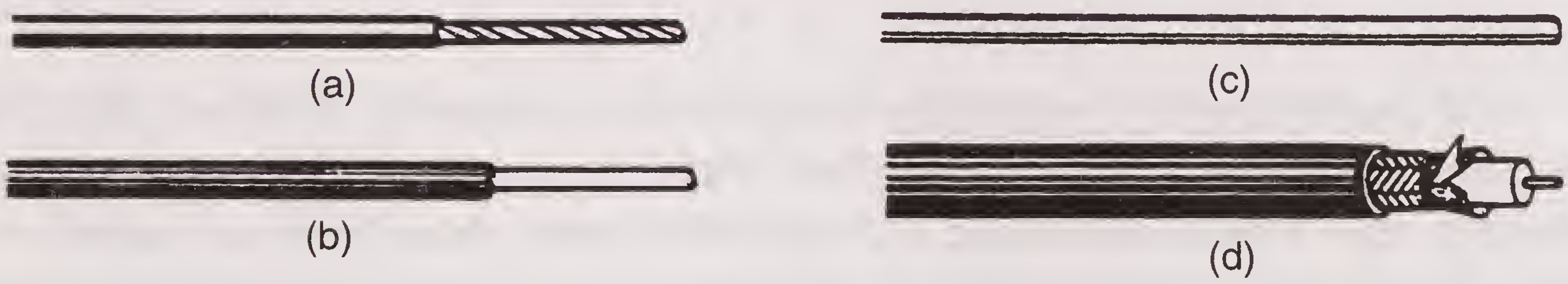

(a) (b) (c) (d)

Figure 41

AWG (American Wire Gauge) is a special type of enameled wire that is used in some electronic (and also electric) components such as coils, relays, solenoids, motors, magnetic sensors, loudspeakers, earphones, and many others. AWG is a standard U.S. set of nonferrous wire conductor sizes. The gauge is referring to the diameter. This standard is mostly applied to copper household wiring and telephone wiring. AWG is sometimes known as Brown and Sharpe (B&S) Wire Gauge, also.

AWG enameled wire is formed by a rigid copper wire insulated by a cover of enamel. The diameter of the wire determines the amount of current it can conduct. The diameter of a wire can be expressed in millimeters, mils, or by an AWG number. The higher the AWG number of a wire, the smaller the diameter. Table 1-2 shows the characteristics of the AWG wires used in electronics.

AWG No.	Dia.	Cross Sec. (mm²)	Resistance (ohms/km)
0000	11.86	107.2	0.158
000	10.40	85.3	0.197
00	9.226	67.43	0.252
0	8.252	53.48	0.317
1	7.348	42.41	0.40
2	6.544	33.63	0.50
3	5.827	26.67	0.63
4	5.189	21.15	0.80
5	4.621	16.77	1.01
6	4.115	13.30	1.27
7	3.665	10.55	1.70
8	3.264	8.36	2.03
9	2.906	6.63	2.56
10	2.588	5.26	3.23
11	2.305	4.17	4.07
12	2.053	3.31	5.13
13	1.828	2.63	6.49
14	1.628	2.08	8.17
15	1.450	1.65	10.3
16	1.291	1.31	12.9
17	1.150	1.04	16.34
18	1.024	0.82	20.73
19	0.9116	0.65	26.15
20	0.8118	0.52	32.69
21	0.7230	0.41	41.46
22	0.6438	0.33	51.5
23	0.5733	0.26	56.4
24	0.5106	0.20	85.0
25	0.4547	0.16	106.2
26	0.4049	0.13	130.7
27	0.3606	0.10	170.0
28	0.3211	0.08	212.5
29	0.2859	0.064	265.6
30	0.2546	0.051	333.3
31	0.2268	0.040	425.0
32	0.2019	0.032	531.2
33	0.1798	0.0254	669.3
34	0.1601	0.0201	845.8
35	0.1426	0.0159	1,069
36	0.1270	0.0127	1,339
37	0.1131	0.0100	1,700
38	0.1007	0.0079	2,152
39	0.0897	0.0063	2,669
40	0.0799	0.0050	3,400
41	0.0711	0.0040	4,250
42	0.0633	0.0032	5,312
43	0.0564	0.0025	6,800
44	0.0503	0.0020	8,500

Table 1-2
Standard annealed copper wire (AWG & B&S)

Fuses

Fuses are the protection elements of a circuit.

The whole purpose of a fuse is to switch off the current in the circuit before anything other than the fuse can be harmed when a problem occurs. They operate based on the old principle of a chain being only as strong as the weakest link. It is easy to see that the best way to protect everything else was to make one link deliberately weakest so it would fail first.

The name fuse comes from fusible link, as suggested by the previous analogy. A fuse is designed to conduct its rated current and act as a conductor. When the current increases beyond the rated value the fuse melts, acting as an SPST switch and opening the circuit.

Symbols and types

Figure 42 shows the symbols used in electronics diagrams to represent a fuse and also the types of common fuses found in electronic equipment. These fuses are basically formed by a piece of wire in which thickness is the determining factor of the melting current.

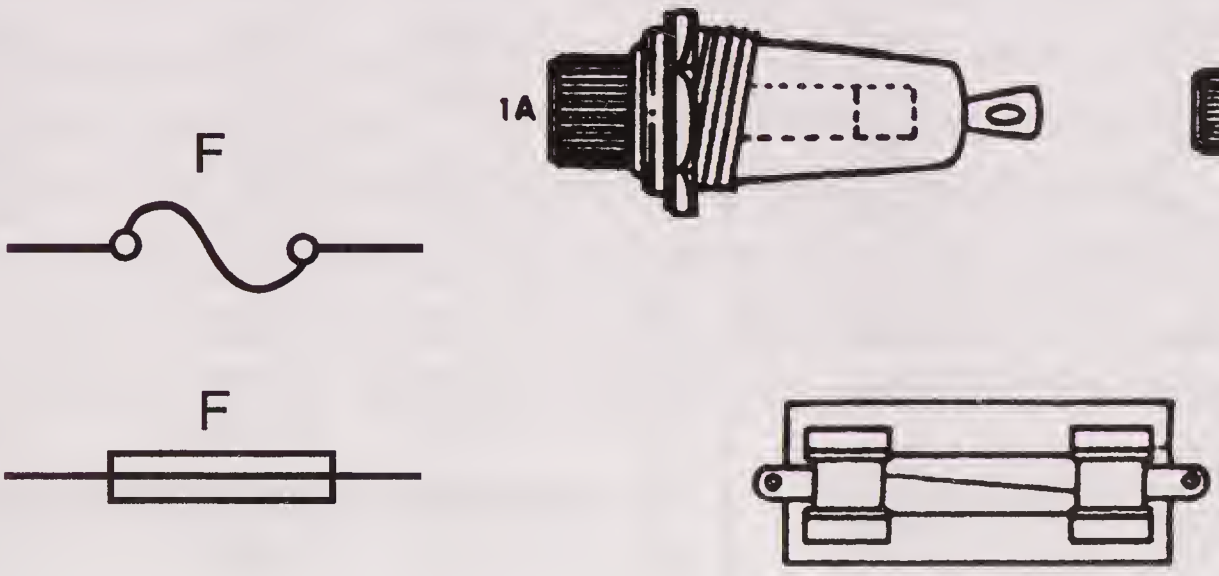

Examples

Figure 42

Where they are found and how they are used

The fuse is wired in series with the device or circuit to be protected as shown in Figure 43. It is usual in some electronic equipment to have several fuses, each one protecting one stage or part in such a manner that if one stage fails and the corresponding fuse melts, the other stages can continue their normal operation unless they are interdependent.

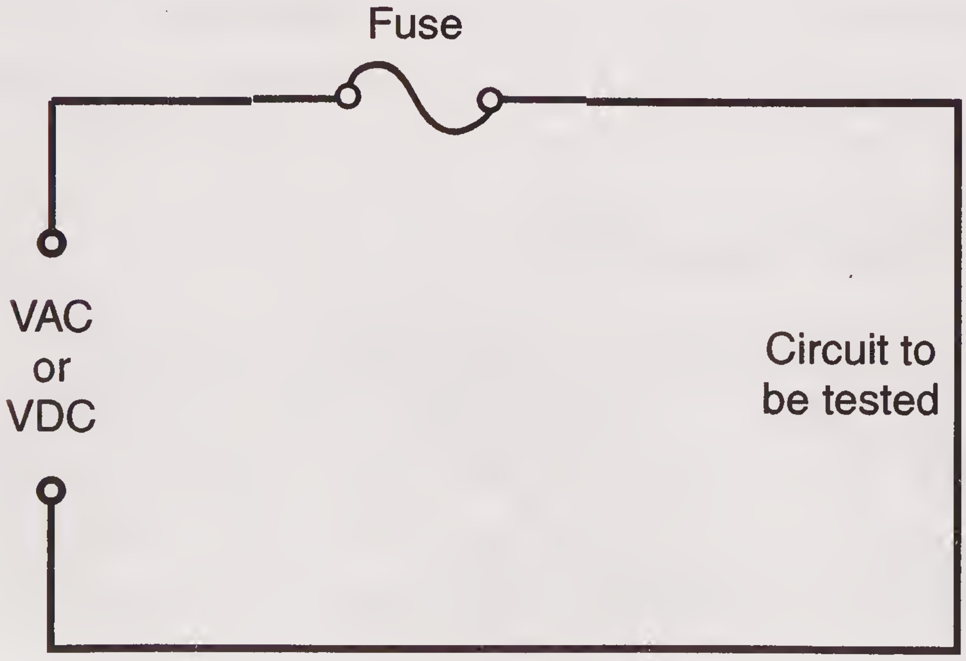

Figure 43

Specifications

The most important specification of a fuse is the melting current or the current that causes the fuse to burn. This nominal current can be specified in miliamperes or amperes according to the value and application.

When replacing a fuse, never use a unit with current specification larger than the original. The specification of current must be the same. If a larger fuse is used and a failure occurs in the circuit, increasing the current to a dangerous value, the fuse may not be the first component to burn out. It may be some component of the circuit or possibly many of them that will burn.

Other specifications of a fuse to consider are how fast it burns when problems occur and the highest operation voltage. In some electronic circuits the fuse is a small piece of bare wire placed between two terminals. The wire gauge is chosen to melt with the desired current.

How to Test a Fuse

A good fuse must present a low electric resistance. To know if a fuse is in good condition, test its continuity—whether it can conduct the electric current. A high resistance means a burned fuse (open circuit). A good fuse measures a resistance near zero when tested. A multimeter can be used for this task.

Switches

In electronic appliances the electrician will find many types of switches. As in any other electrical application, the switches are used to control the current flow throughout the circuits or parts of them.

The difference between the switches used in common electric installations and cars is not important. In electronic equipment the switches, in many cases, are used to control many currents at the same time. The design also depends the applications.

Symbols and Types

Figure 44 shows the symbols and types of the main types of switches found in electronic circuits. The symbol indicates what the switch makes. For instance, (a) shows an SPST (Single-Pole Single-Throw), which is a switch that controls the current flow in only one circuit (single-pole). As the name suggests, this switch controls one circuit turning the current on and off.

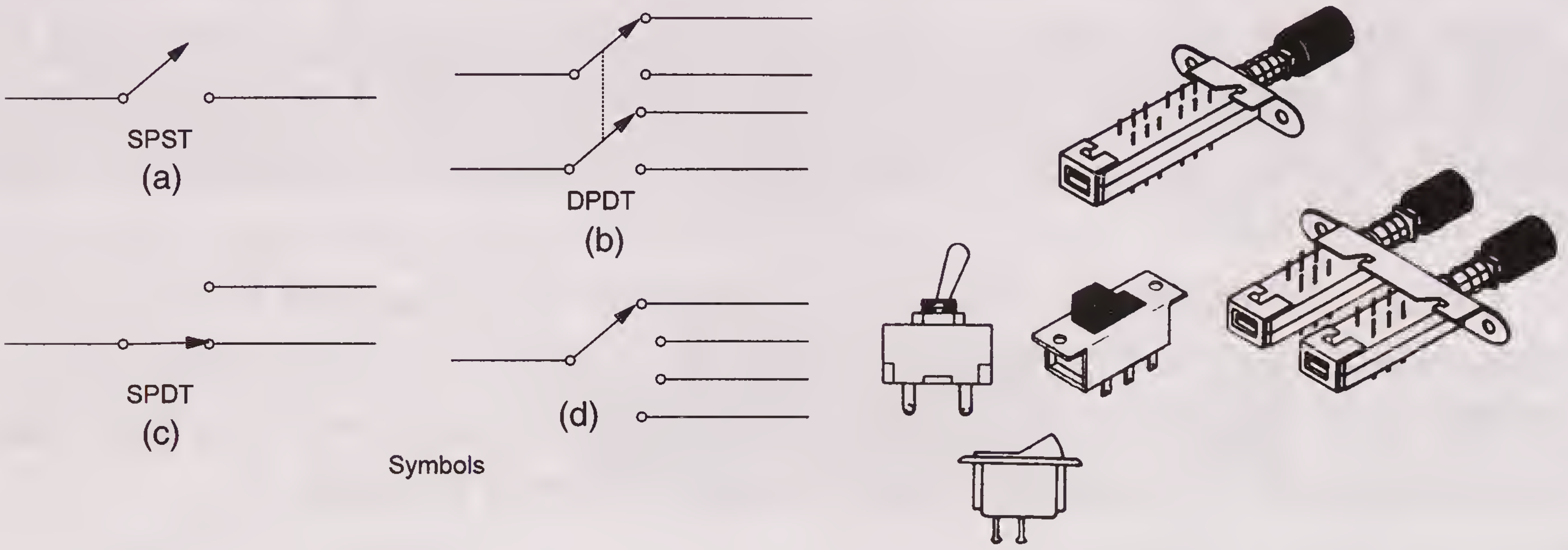

Figure 44

A DPDT (Double-Pole Double-Throw) is shown in (b). This switch can control the current in two independent circuits (double-pole) transferring it to one or another circuit (double throw). The switches also are different in the way they are operated. A circuit can have slide switches (b), toggle switches (c), and rotary switches (d).

Some special switches are also shown in Figure 44, such as thumbwheel types, pushbuttons, and subminiatures that are suitable for installation in PCBs (Printed Circuit Boards), etc.

Where they are found

Switches are found in any application where current must be controlled. Every piece of electronic equipment needs at least one switch for its operation—the on/off switch.

Specifications

The switches are identified by the number of poles, the number of positions, the maximum operating voltage, and current.

The poles and positions are indicated by symbols or letters. These are the main types:

SPST—Single-Pole Single-Throw
SPDT—Single-Pole Double-Throw
DPDT—Double-Pole Double-Throw

Maximum Operating Current and Voltage

The maximum operating current is indicated in amperes (A). Never use a switch in a circuit with a current larger than the maximum recommended. The contacts can heat causing problems. The maximum voltage is indicated in volts (V).

How to Test

An open switch must present a very high resistance (near infinite) and when closed a very low resistance (near zero). To test a switch, take a simple continuity measurement using a multimeter in the lowest resistance scale. It is important, in some cases, to make a visual inspection: a switch with contact problems due to excess current can present some visible signs such as deformations or dark areas in the contact.

Cells and Batteries

Cells are primary chemical energy sources for much electronic equipment. The cell is basically formed by two different metals immersed in some kind of chemical substance. The cells used to power electronic equipment can be found in a large collection of types, shapes, and sizes. The type, size, and shape are determined by the energy requirements of the application, the length of time the battery must source energy, and the overall size of the equipment to be powered. Many types of technology are used by the manufacturers to provide all the batteries found in electronic equipment. The technology involved is with the materials used inside the cell and its form as they are installed.

The most common cell is the carbon-zinc dry cell or Leclanché cell that is found in many battery sizes (button, AA, AAA, C, and D). This cell consists of one electrode of carbon and one of zinc with an electrolyte of ammonium chloride. Actually, the carbon electrode is composed of a mixture of manganese dioxide, carbon, and solid ammonium chloride. Most of the energy produced by this cell comes from decomposition of the manganese dioxide.

A variation of this type of cell is the alkaline cell. In the alkaline-manganese cell the chemical reaction is the same as in the carbon-zinc cell, but it takes place in an alkaline medium rather than the acid mixture used in the carbon-zinc cell. This kind of cell can provide up to twice the current output with a much longer shelf life.

These common cells are not rechargeable as they operate based in a nonreversible chemical reaction, thus the name primary cells. On the other hand, rechargeable or storage cells are called secondary cells—such as the ni-cad (nickel-cadmium) types.

These cells can be recharged by a current flowing in the reverse direction than the current source flows when they are in operation. Special devices, such as the one shown in Figure 45, are used to source this current.

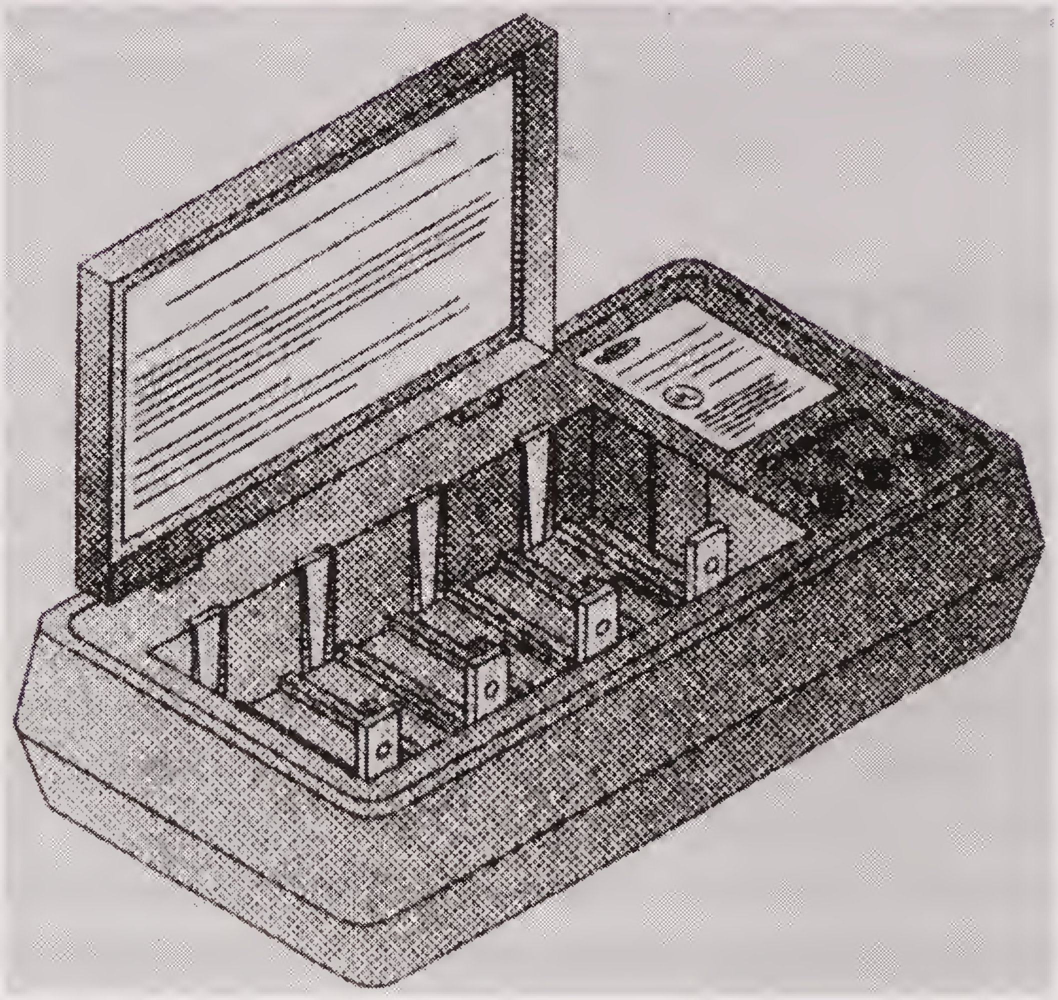

Figure 45

Common cells source voltages between 1.2 V and 1.5 V.

When many cells are associated in series we call the association a battery. The most common type of battery found in electronic applications is the 9 V battery. It is formed by the association of 6 single 1.5 V flat cells.

Symbol and Types

Figure 46 shows the most common types of cells and batteries found in electronic applications. Each type can be found in dry, alkaline, ni-cad, and some other versions based on the application.

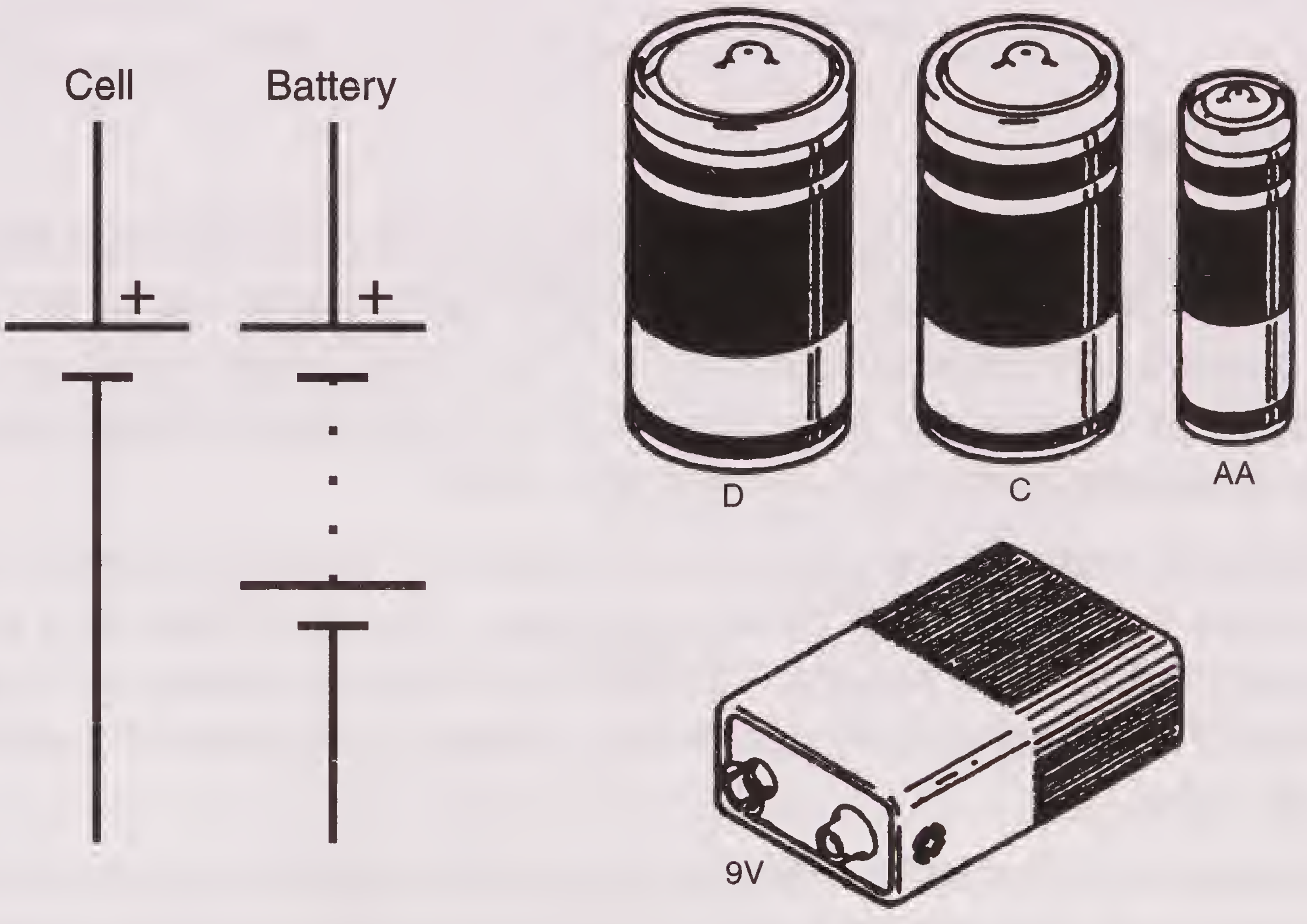

Figure 46

Specifications

A. The first and principal specification of a cell or battery is the type and size. Standard designations are groups of letters to indicate the format and size (AA, A, C, D, etc.).

B. The second specification, voltage, can be omitted because it is included in the type.

C. Depending on the application it is important to know how many hours the battery can source a determined amount of energy to a device. This is indicated by the battery or cell Ah or mAh (amperes x hour or milliamperes x hour). For example, a battery with a 100 mAh specification can source 10 mA of current to a circuit during 10 hours or 1 mA during 100 hours.

Testing

Many battery testers are available in specialized stores. But, a simple multimeter can also be used to make this test. Many of them have a battery test function. The measure of the voltage of a cell is not a conclusive way to test a battery or cell.

Resistors

The presence of resistance in an electric circuit is not desirable. Only when electric power must be transformed into heat such as in air heaters, water heaters, and electric showers, are elements present with an electric resistance to the current flow used. Forcing passage through those elements allows the current flow to produce heat.

Devices that add resistance to a circuit can be used to limit or reduce the current flow or to step down a voltage—two functions that are very important in electronic circuits. To add an amount of resistance to a circuit or reduce the flow of electric current by a circuit, components called resistors are used.

The common resistor or fixed resistor is a component made from some conducting material the size, shape, and conductivity of which are arranged to determine the amount of electrical resistance needed. The amount of resistance is measured in ohms (Ω).

Inside electronic equipment are resistors of many types and sizes with resistance values ranging from less than 1 ohm to many million ohms. The resistors are classified into fixed or variable resistors. Let's look at fixed resistors, also called resistors.

Resistors are found in many sizes and shapes according to the amount of resistance they present, some special performance feature they must have, and the amount of heat generated when working.

Symbols and Types

Figure 47 shows the symbols used to represent a resistor and the types of resistors commonly found in electronic circuits. The symbol shown in (a) was adopted by the American system of symbols. In European equipment the schematics or circuit drawings represent the resistor as shown in (b).

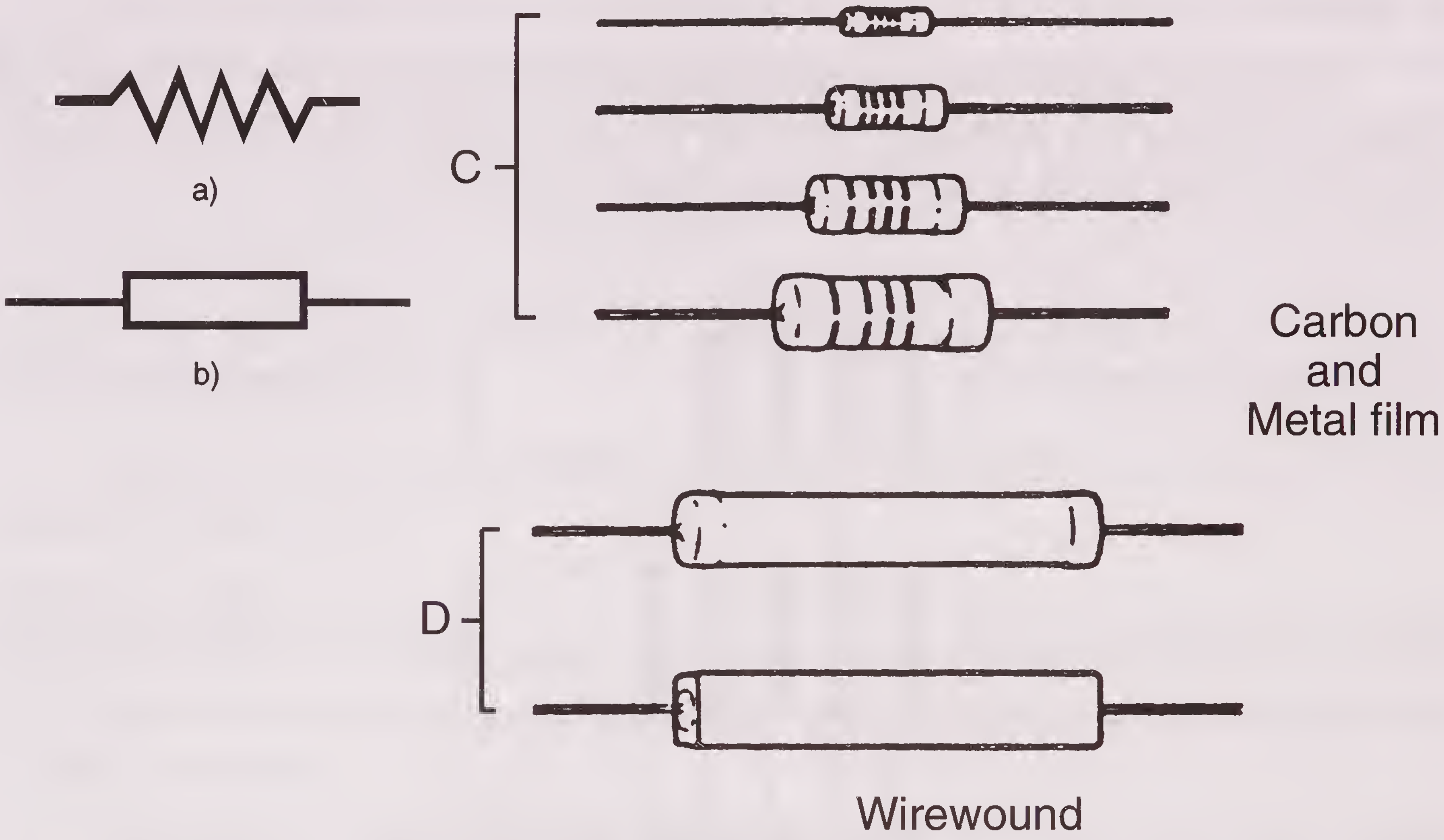

Figure 47

The most commonly used resistors are carbon composition and carbon film types, as shown in Figure 47 (c) where the resistive element is a thin layer of carbon strips deposited onto a ceramic rod. Another type is the metal film resistor or film resistor, with the same external aspect as the carbon type. In this type, metal is evaporated onto ceramic rods forming a thin film. These resistors are cheap and can be easily fabricated in a large range of values. They are intended for applications in low-power or small-current circuits. The disadvantage of carbon resistors is that they are noisy.

When desiring resistors that can operate with large currents, the wirewound type shown in Figure 47 (d) is preferred. Those resistors are formed by turns of metal wires (alloys such as nickel, chromium, iron, silver, etc.) onto a ceramic rod. The material, gauge, and number of turns determine the amount of resistance, the size, and how much heat they can handle.

Specifications

When replacing or testing any electronic device it is necessary to know its specifications. The main specifications for resistors are:

A. Value

The value is the amount of resistance in ohms (Ω) of a resistor. In large-size types the resistance can be found on the body of the device, the same as in the wirewound resistors (See Figure 47).

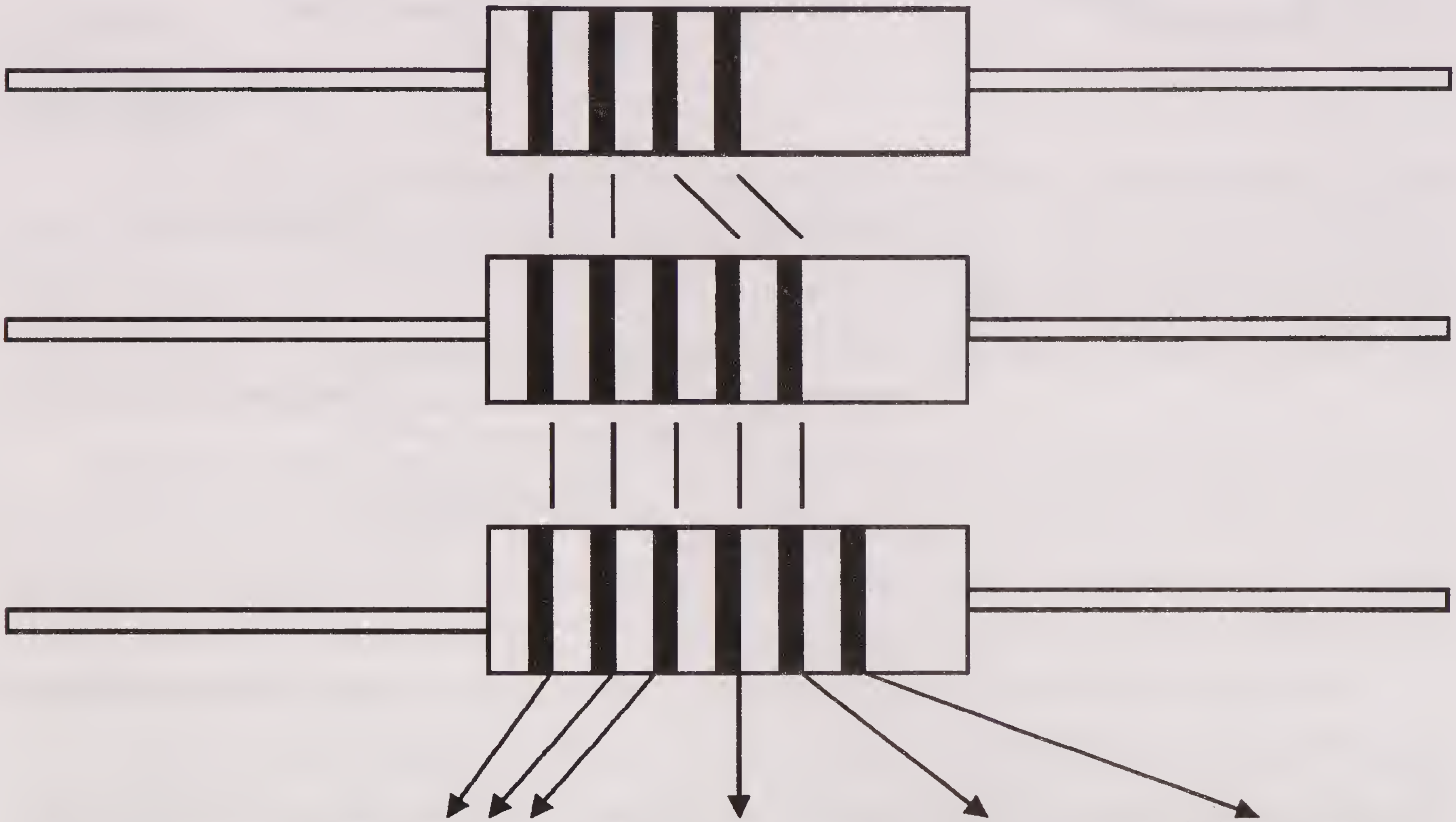

COLOR	Significant Figures	Multiplier	Tolerance	Temperature Coefficient ppm/°C
Black	0	1	-	-
Brown	1	10	1%	100
Red	2	100	2%	50
Orange	3	1,000	-	15
Yellow	4	10,000	-	25
Green	5	100,000	0.5%	-
Blue	6	1,000,000	0.25%	10
Violet	7	10,000,000	0.1%	5
Gray	8	100,000,000	0.05%	-
White	9	1.000.000.000	-	1
Gold	-	0.1	5%	-
Silver	-	0.01	10%	-

Table 1-3 Color Code for Resistors

But, in smaller types such as carbon and metal film types as well as SMDs, there isn't enough space to display the resistance value. In this case, and also to make the manufacturing process a bit easier, a color code is used. Color bands or stripes are printed on the resistor body. Accordingly, the position in the equipment is part of the meaning of a band. Combining the two, the resistance value of any resistor can be found. In addition to the value in ohms, the color bands also indicate the tolerance of the component.

Table 1-3 indicates how the value and tolerance of a resistor is read using the color code.

Notes on Table 1-3

1. If the resistor is a three-band type, the tolerance can be assumed as 20%
2. If the resistor is a three- or four-band type the temperature coefficient is not indicated.

Example—A resistor has colored bands in this sequence: red, violet, orange, and silver. (Read the colored bands from the end to the center.) Bands 1 and 2 form the number 27. Band 3 indicates that we must multiply this value by 1,000. The resistance is 27,000 ohms. Band 4 indicates that the tolerance is 10%.

B. Dissipation

The size and the material the resistor is made of determines how much heat it can produce without burning. Based on the application the resistors are to be used for, resistors must be chosen in different sizes or dissipation capacities.

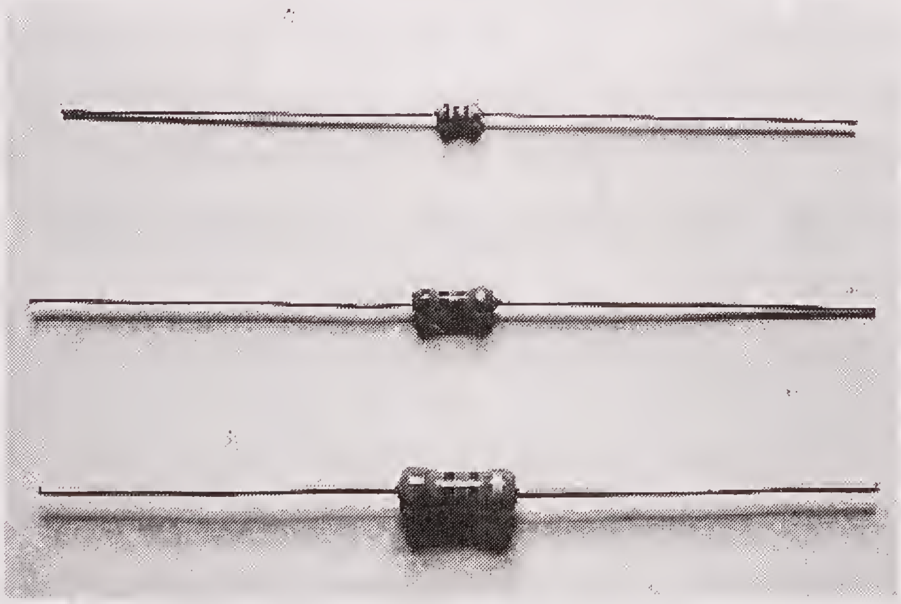

Figure 48

The dissipation or power rate of a resistor is specified in watts (W). Small resistors (metal film or carbon) have dissipation in the range between 1/8 and 2 W. Wirewound types can be found in dissipation starting from 1 or 2 W to more than 200 W. Figure 48 shows the differences in size between carbon resistors according to their dissipation.

(Figure 48) Carbon resistors with different dissipations are shown. From the smaller to the greater the dissipations are 1/8, ¼, and 1/2 W.

C. Tolerance

It is impossible to manufacture a resistor with an exact value of resistance. This is not necessary as the electronic circuits are designed to operate within a certain band of value of components and voltages. The difference between the value specified for a component and the real value or measured value is called tolerance.

Resistors can be found in tolerance ranges from 1 percent to 20 percent. This means that a resistor indicated as 1000 ohms/10 percent unit can really present resistances between 900 and 1100 ohms. As a consequence, it is not necessary make resistors with all values between 900 and 1100 ohms, as the 1000 ohms resistor covers that range. So, the electrician will be surprised when observing that the resistors are found in few common values, according to their tolerances. The basic values are determined by a commercial series of values.

20% SERIES OR E-6	10% SERIES OR E-12
10	10
	12
15	15
	18
22	22
	27
33	33
	39
47	47
	56
68	68
	82
100	100

Table 1-4

Table 1-4 shows that resistors of 47 ohms, 470 ohms, 4700 ohms, 47,000, 470,000 x 20 percent and many other values in multiples of 10 can be found, but not 39 ohms, 390, or 3,900 ohms x 20 percent. These values are found only in 10 percent resistors. In the same manner, you'll never find a 3.4 ohms, 34 ohms, 340 ohms, etc., resistor with 10 or 20 percent tolerance because it doesn't exist.

Where Resistors are Used

The resistors are the most common of the electronic components. They are used in all applications where a current must be limited or a voltage steps down.

In electronic circuits the main application for resistors is in bias or stepping down a voltage. Many elements of a circuit must operate with voltages lower than supplied by the power supply. This means that resistors must be used to step down the voltage to the desired value. These resistors are called bias resistors.

Testing

When heated by excess current, the resistors can open or burn. Although a dark area in the body of the component is an evident sign that it is damaged, in some cases the component doesn't present any other visible signs of problems.

Resistors are tested using the multimeter (VOM). Desolder one terminal of the component and take it off of the printed circuit board or out of the circuit and measure the resistance. Don't measure it while still in the circuit because the multimeter can indicate the resistance of the circuit, not that of the component.

Figure 49 shows how to do this test.

Remember to consider the tolerance when reading the value displayed by the multimeter and comparing it with the value displayed by the component. When replacing a resistor use a unit of the same type and same value. The dissipation, in some cases, can be higher if enough space is available for its placement.

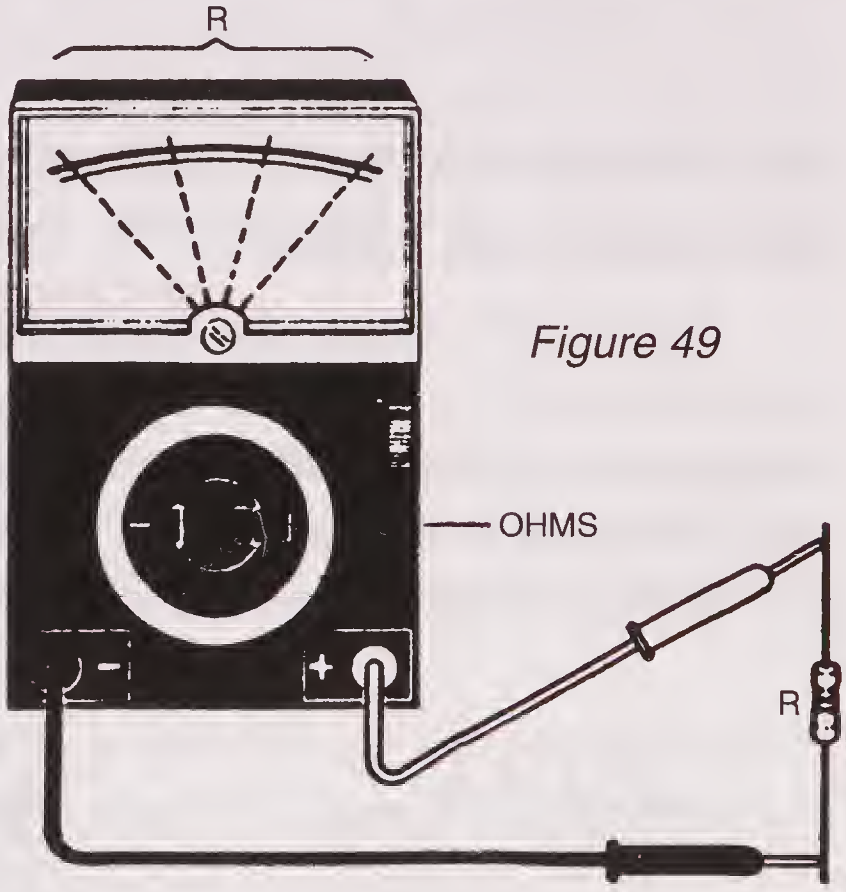

Figure 49

Variable Resistors

In many applications something in a circuit can be controlled or adjusted by changing a resistance. You can increase the volume of an amplifier or the brightness of a lamp by altering the current in a circuit. The common way to alter currents and voltages in circuit is altering a resistance placed in the circuit for this task. The principal component that can be used to make this happen is the variable resistor. The electrician will find variable resistors in a large number of functions in electric and electronic appliances.

As the name indicates, a variable resistor is a device presenting a resistance that can be changed. Variable resistors are found in different sizes and aspects. There are two types. One allows the resistance value to be changed easily and often, like the volume or tone adjustment of an audio amplifier. The other is the semifixed resistor that ordinarily doesn't change the resistance value. It is used to adjust the operational condition of the electronic circuit.

Basically, variable resistors are formed by a resistive material (metal or carbon) with a cursor sliding onto it. When the resistance between the cursor and any corner of the resistive material changes normally, between 0 and a determined value of resistance, it is called nominal resistance. So in a 1000 ohms variable resistor, the resistance between the cursor and any of the corners can assume values between 0 and 1000 ohms.

Symbols and Types

Figure 50 shows the symbols and types of the two basic types of variable resistors. In (a), a trimmer potentiometer or adjustable potentiometer is shown. This small variable resistor is used to adjust the operational condition of a circuit one time. They are found inside the equipment. Size and form of mounting can change according to the application.

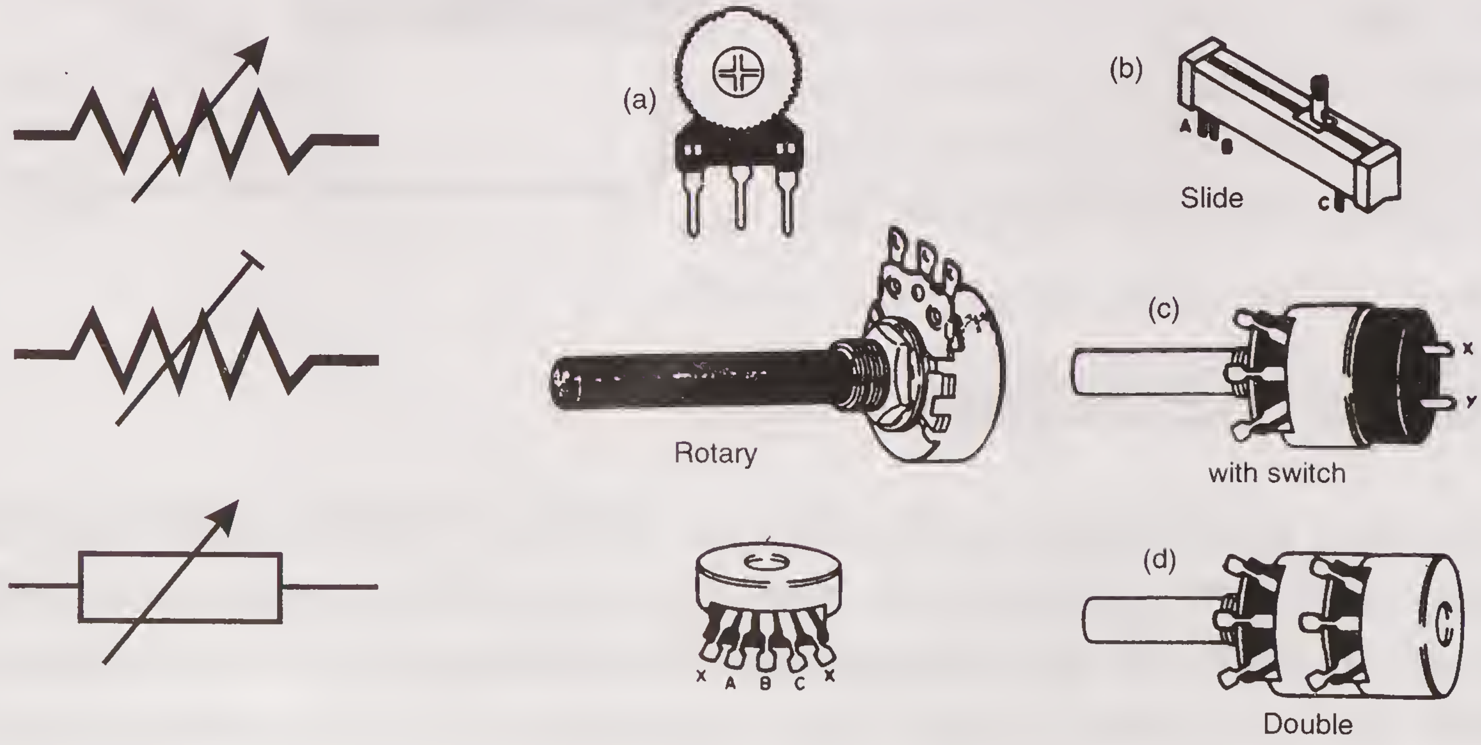

Figure 50

In (b) are potentiometers (slide and rotary) that are used to control many functions of equipment. The slide potentiometers are common in audio equipment and the rotary types are found in TV, radio, dimmers, etc. Many potentiometers are double (d) as those used to control the volume in both channels of a stereo amplifier. Others (c) can include switches to turn on and off an appliance.

Specifications

A. Value or nominal resistance

The most important specification in a variable resistor (potentiometer or trimmer potentiometer) is the nominal resistance. This resistance is indicated in ohms (Ω) and represents the maximum value of resistance assumed by the component. For example, a trimmer potentiometer preset with 10,000 ohms or 10kΩ can be adjusted to any resistance between 0 and 10,000 ohms.

B. Variation of resistance

In some applications the way the resistance changes when the potentiometer is adjusted is important. There are basically two types of variation of resistance curves: linear and logarithmic.

Where they are found

The electrician can find trimmer potentiometers and potentiometers in many types of electronic equipment. In every application where a function of a circuit must be adjusted or controlled by changing current in a circuit, a variable resistor can be used.

An example of a common use is in a dimmer where the amount of current in a circuit determining the brightness of a lamp is controlled by a potentiometer. Timers also use a potentiometer to adjust the time.

Testing and replacing

Trimmer potentiometers and potentiometers can burn or present many functional problems. One of the problems is the cursor contact can fail. Another common problem is caused by noise generated when the cursor slides on the resistive element. This occurs mainly in audio equipment. This problem is the cause of the scratching noise produced when the volume control of a radio or amplifier is adjusted.

A basic test can be done with a multimeter to measure the resistance of the resistive element (between the extreme terminals in potentiometers). Another test is done by sliding the cursor and listening to hear if noise is produced (if a multimeter doesn't indicate abrupt jumps in the resistance change).

Incandescent Lamps

Small incandescent lamps or bulbs are used in electronic equipment as panel indicators, flashlights, or even as active elements of a circuit. The nonlinear characteristic of the filament (voltage vs. current) makes incandescent lamps ideal for use as a current regulator in some applications.

Incandescent lamps are formed by a tungsten filament placed inside a glass enclosure where the oxygen is removed. When the electric current heats the filament to a very high temperature, it produces light. The metal of the filament doesn't burn because there is no oxygen to be a source of a combustion chemical reaction.

Electronic equipment usually uses small incandescent lamps rated for low voltage in the range between 1.5 and 12 V. This is an important difference from the lamps found in electric applications that are primarily used to illumination purposes and are powered directly from the AC power-line voltage.

Symbols and Types

Figure 51a shows the symbols used to represent small incandescent lamps. In (a), the American symbol is shown and in (b), the symbol found in some diagrams from European manufacturers is shown. Figure 51b illustrates the common types used as panel or instrument indicators and for other functions.

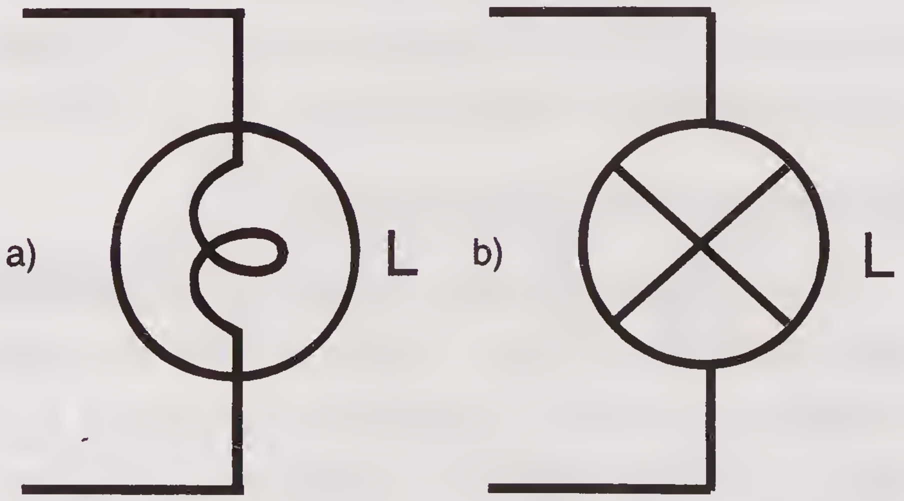

Figure 51a

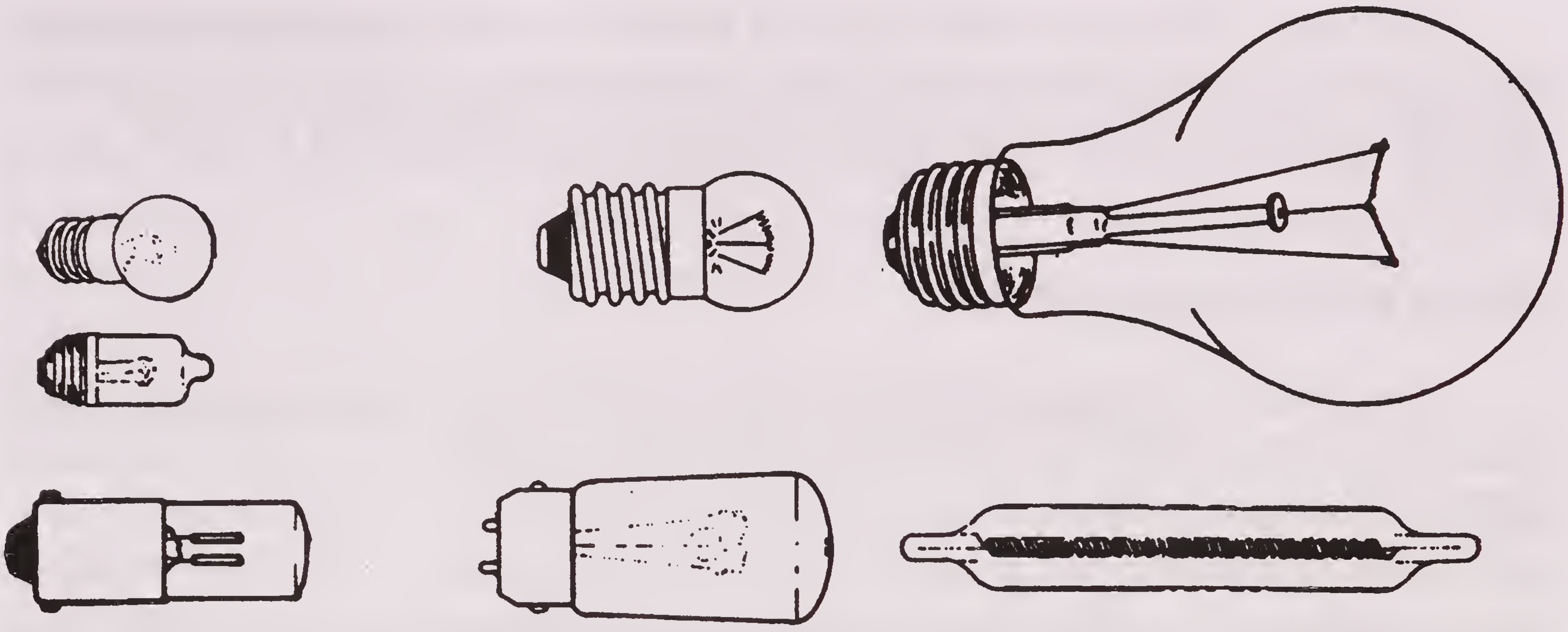

Figure 51b

Observe that the base of the lamp changes in format or size according to the application. The main types are the bayonet base, telephone slide, flange, and threaded.

Specifications

A. Nominal voltage

This is the operational voltage of the lamp. Common types are indicated to operate in the range between 1.5 and 12 V.

B. Nominal current

This is the current flowing through the filament when the nominal voltage is applied. Common types can be specified to operate with currents in the range between 0.01 A (10 mA) and 1 A.

C. Type number

Many catalogs indicate the type of a lamp by a part number giving tables where the electric and mechanical specifications are placed. Types such as "47" are popular. This is a lamp rated to 6.3 V x 150 mA using a bayonet base.

Where they are found

The incandescent lamps can be found basically in all applications where illumination of a panel, instrument, or control is needed. In many pieces of equipment, small incandescent lamps are placed inside the panel instruments.

Although LEDs are replacing the lamps in many applications, those with white light production capabilities are becoming popular. These lamps are used as replacements in antique equipment and some modern applications.

Testing and Replacement

The small incandescent lamps have low filament resistances (varying according to their voltage and current). If the filament burns up, the resistance becomes infinite. Using a multimeter, it is easy to see if an incandescent lamp is good or not good. When replacing a lamp, if the part number is not available, the electrician must know the voltage and current and take care to choose one with the same base format.

Neon Lamps

Neon lamps are used in many electronic circuits as indicators or important active elements because their electric characteristics (negative resistance) make them ideal as trigger devices in timing circuits or waveform generators. The common neon lamp is formed by a small glass bulb filled with neon gas. Two electrodes are placed inside without one touching the other. When the voltage across the gas increases reaching the triggering point (between 60 and 80 V), the gas ionizes and becomes a conductor. The electric resistance falls allowing the circulation of a current. With the ionization, the gas emits an orange light.

Symbols and Types

Figure 52 shows the symbol used to represent the neon lamp and types of this device.

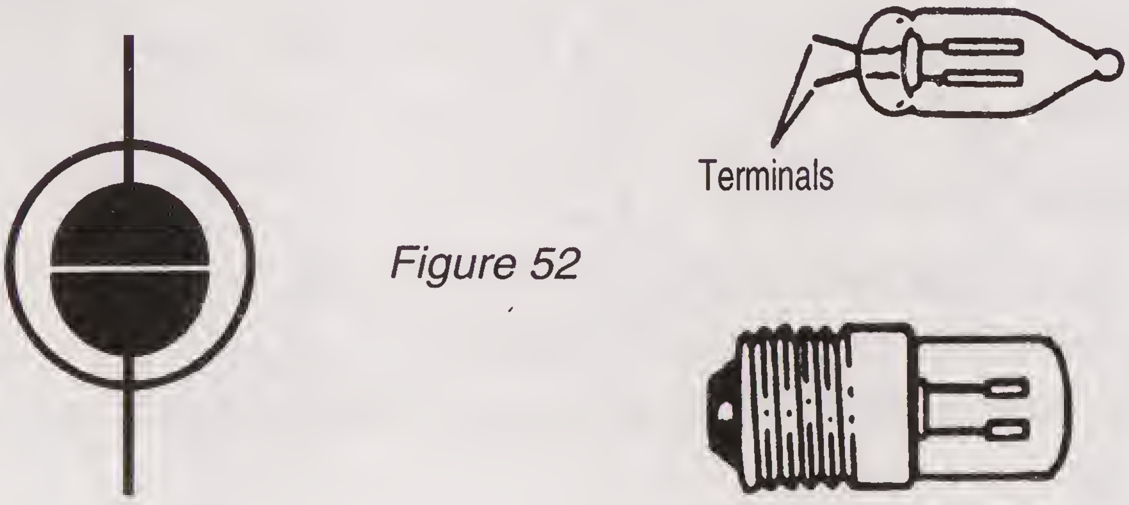

Figure 52

Specifications

The main specification of a neon lamp is the trigger voltage. Normally it is in the range between 60 and 80 volts. The manufacturers indicate neon lamps by a part number such as NE-2H or HE-51. Sometimes it is important to know the type of lamp and if it does or does not have a current-limiting resistor inside.

Where they are found and how they are used

Neon lamps can be used both as indicators and elements of an active circuit. As indicators, they can be found in the input of equipment monitoring the AC power line voltage as shown in Figure 53.

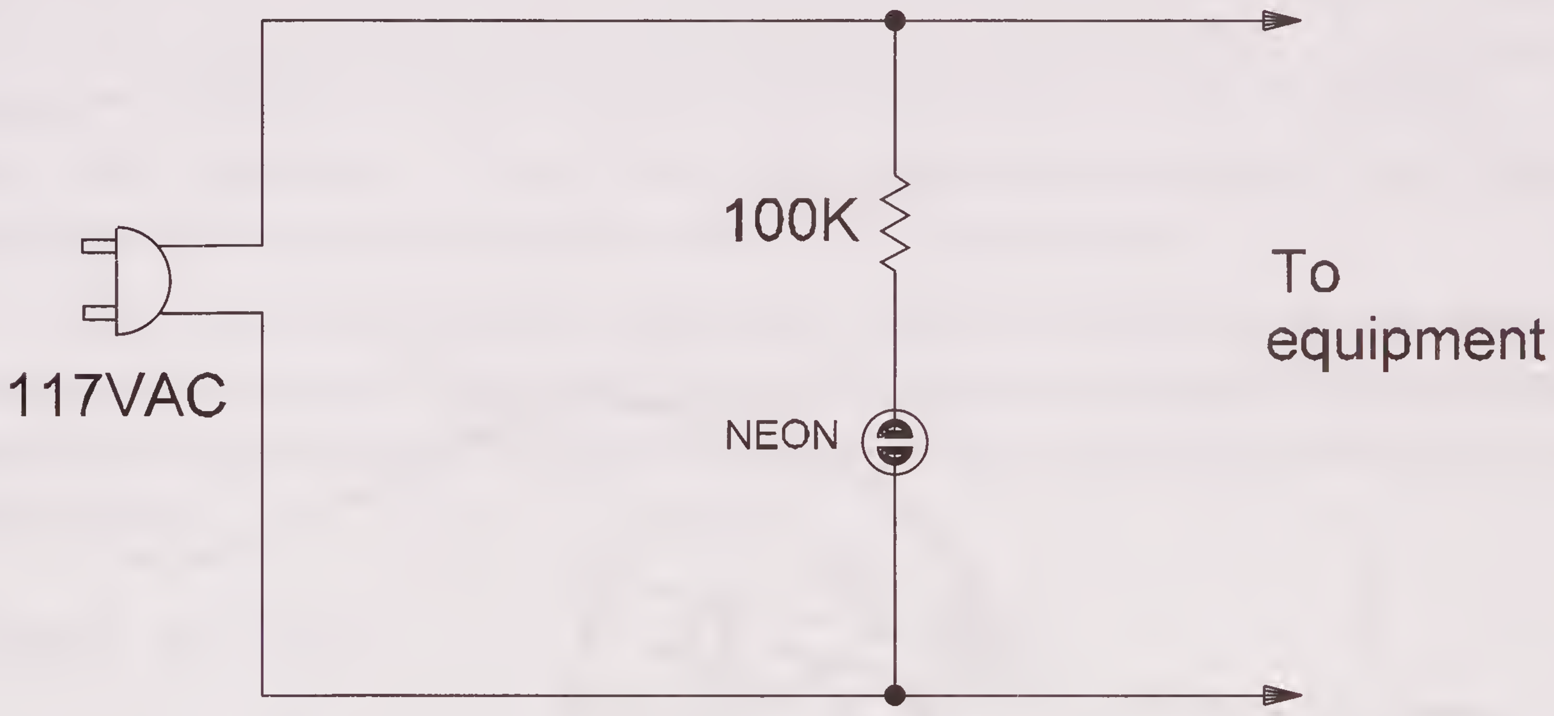

Figure 53

The electrician will find that in circuits like this, the lamp has a series current-limiting resistor (typically between 47k and 1M). Some types of lamps such as the NE-2 have this resistor inside for 120 V use. Others need an external resistor.

Another application is an oscillator for generating signals as in the configuration shown in Figure 54. Depending on C and R values, the circuit can generate sawtooth audio signals or make the lamp flash.

An interesting application of the neon lamp is as a phase indicator as shown in Figure 55. If placed in the phase terminal of an AC power line, the lamp will glow. The electrician will not feel any shock because the series resistor limits the current to a very low value, not enough to excite the nervous cells, giving a shock.

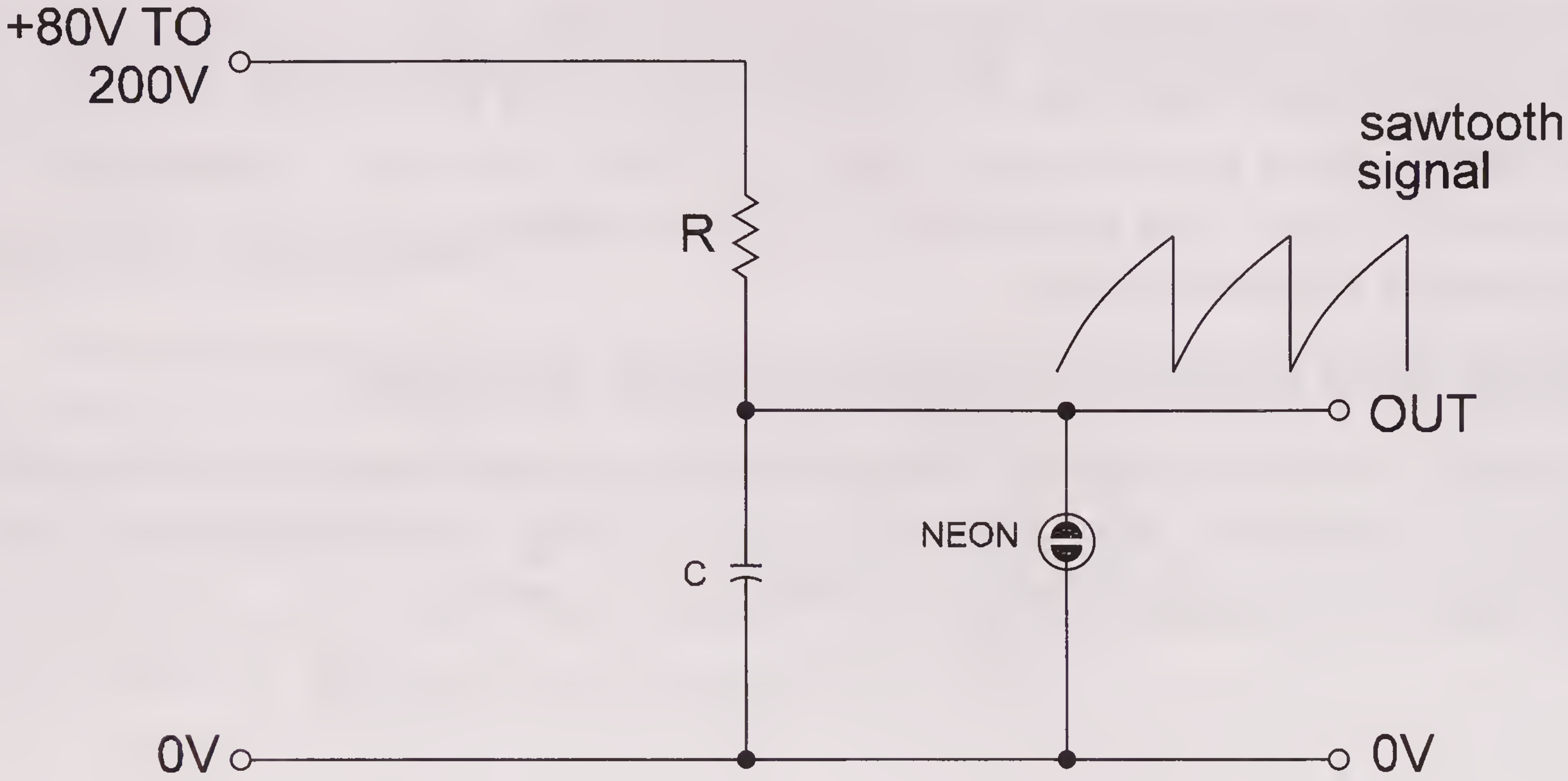
+80V TO
200V
R
sawtooth
signal
OUT
NEON
C
0V
0V

Figure 54

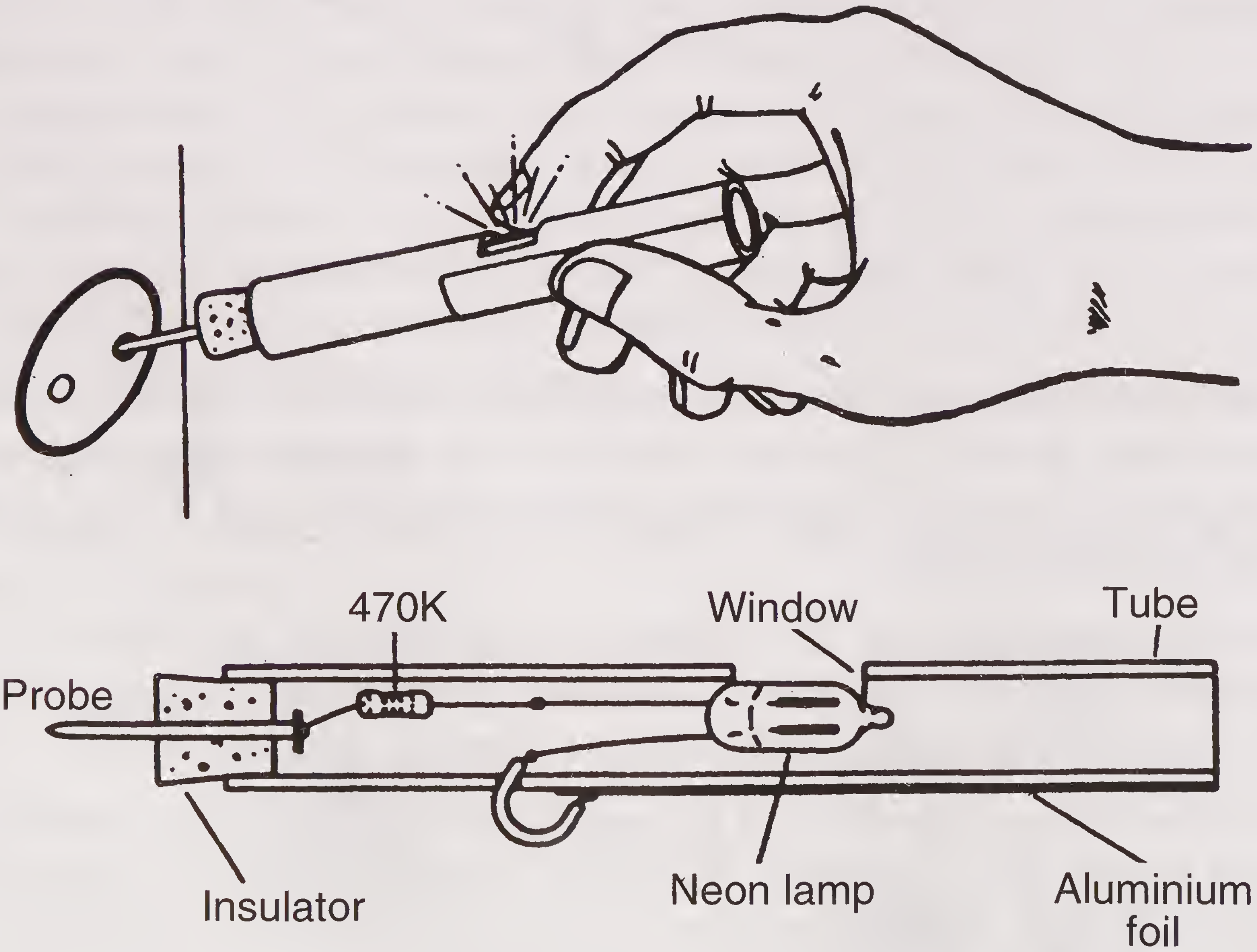
470K
Window
Tube
Probe
Insulator
Neon lamp
Aluminium
foil

Figure 55

Testing

The simplest way to test a neon lamp is to plug it into the AC power line using a 220k ohms series resistor.

> ***WARNING:*** *Do NOT plug the lamp directly into the AC power line. Without the current limiting resistor it can explode!*

Light-Dependent Resistor or CdS Cells

Light-dependent resistors (LDR) or CdS cells or CdS elements are light-sensitive resistors presenting a resistance value that changes with the light. The resistance between the terminals of these components falls when the amount of light falling onto a sensitive surface increases. In the dark, the resistance is high and falls under the sunlight. The LDRs are used as light sensors in many electronic applications.

They are formed by a cadmium sulfide surface—the sensitive element of the device. This surface and its base are placed inside a transparent plastic enclosure. The size and the format of the device depend on the application.

Symbols and Type

Figure 56 shows the common symbols used to represent this component and Figure 57 shows what they look like.

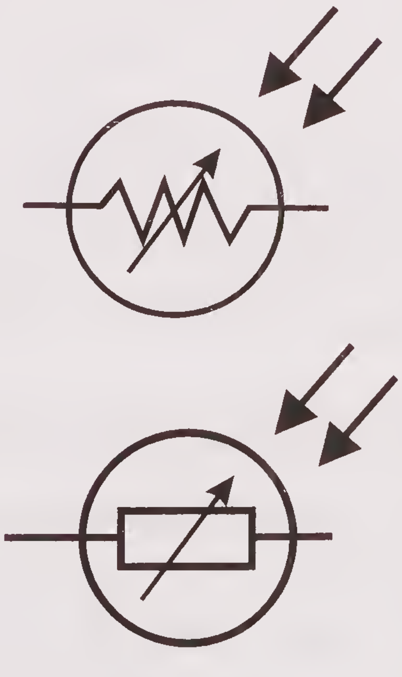

Figure 56

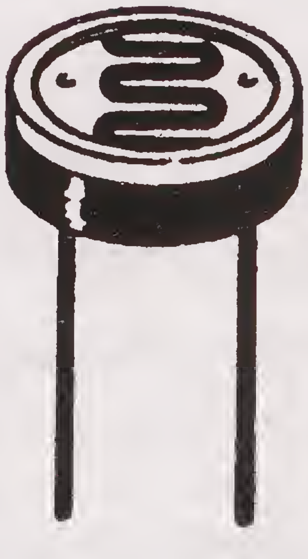

Figure 57

Specifications

The LDRs can by specified by the diameter (size) or by a manufacturer number. When replacing LDRs, types of the same size are normally interchangeable.

Where they are found

The electrician will find LDRs in many electric devices related to domestic applications. They are used as sensors in alarms, automatic lights, emergency lights, and many other circuits. In a typical application the LDR is placed in a way that it receives only the ambient light as shown in Figure 58.

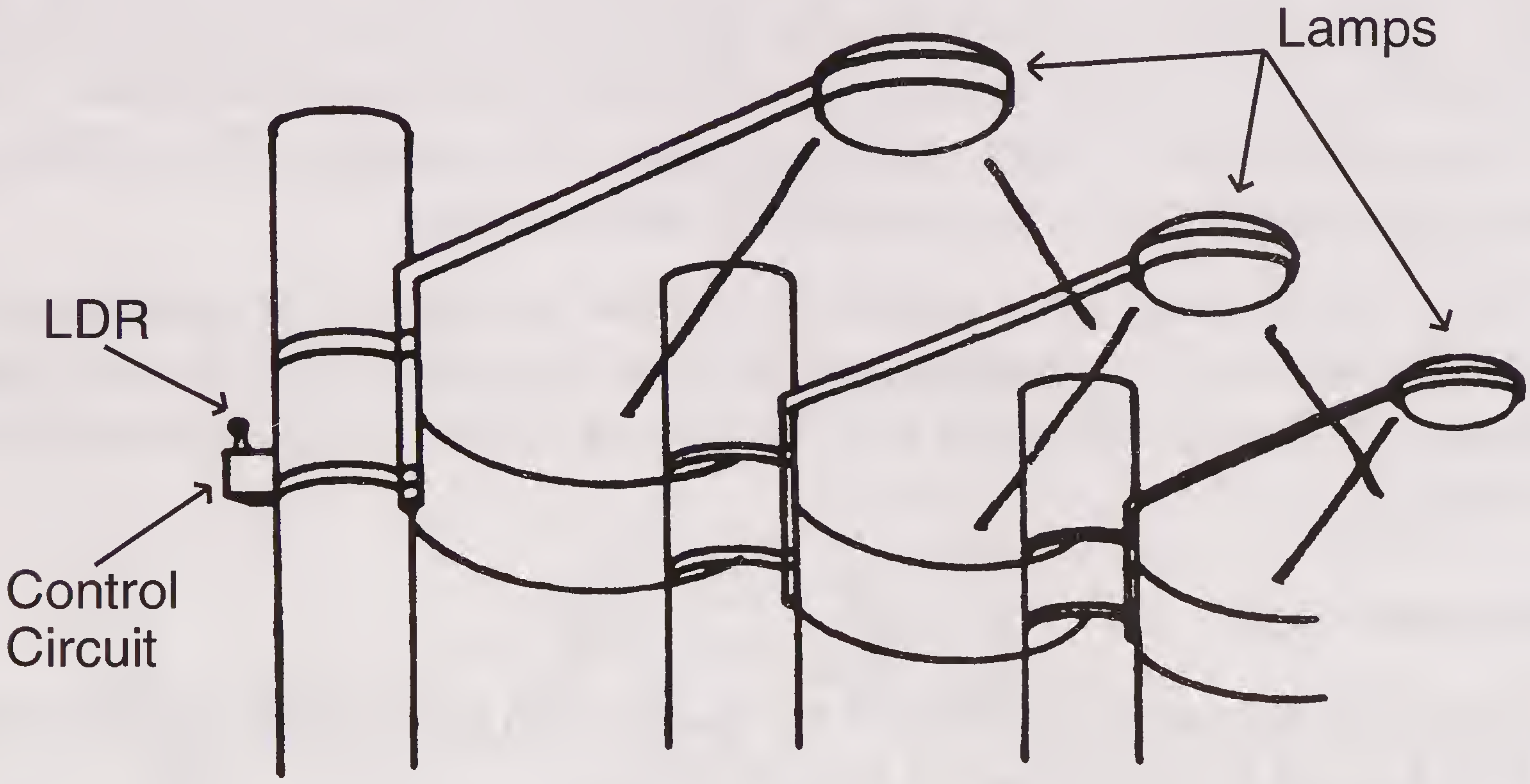

Figure 58

When the light falls at dusk, the LDR's resistance increases acting on the electronic circuit triggering a lamp. If the LDR receives the lamp's light a feedback process can interfere in the circuit. Oscillations and instabilities can be induced this way.

Although the LDRs are very sensitive devices, they are not fast enough for certain applications and can only be used to control simple circuits from slow light changes. When a fast action is needed, such as when reading bar codes or detecting fast changes of light, other devices are preferred like photodiodes and phototransistors.

In many applications the amount of light on the LDR or CdS cell can be increased with the use of a convergent lens. The use of a convergent lens, as shown in Figure 59, is important to help the device to detect light from smaller angles.

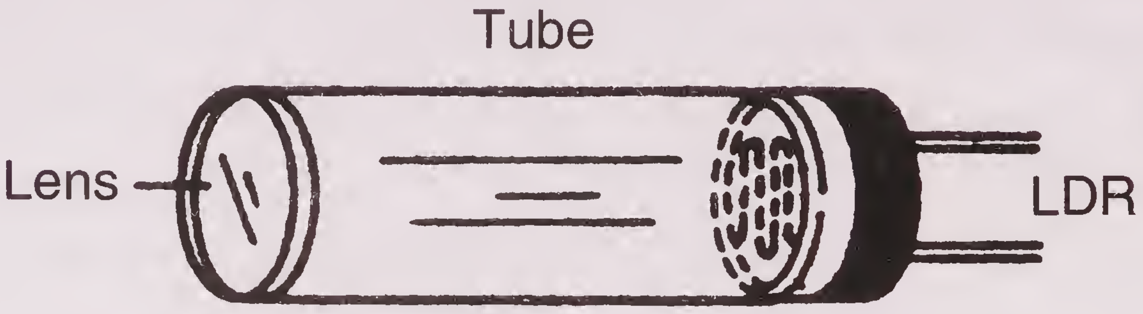

Figure 59

Testing

A simple multimeter (digital or analog) can be used to test any LDR. Put the multimeter on a resistance scale and plug the probes into the LDR. Passing your hand in front of it (to cut the ambient light falling in the sensitive surface) you'll observe the resistance changes in the multimeter's scale.

NTC/PTC

Negative Temperature Coefficient resistors (NTC) or Positive Temperature Coefficient (PTC) resistors (also called thermistors or thermally sensitive resistor) are components having a resistance that changes with their temperature. In the NTC the resistance falls when the temperature increases and in the PTC the resistance increases with the temperature.

They are formed by materials (alloys and mixtures) that have special thermal properties suitable to the task.

Symbols, Types, and Characteristic Curve

Figure 60 shows the symbols used to represent these components, types, and the characteristic curve.

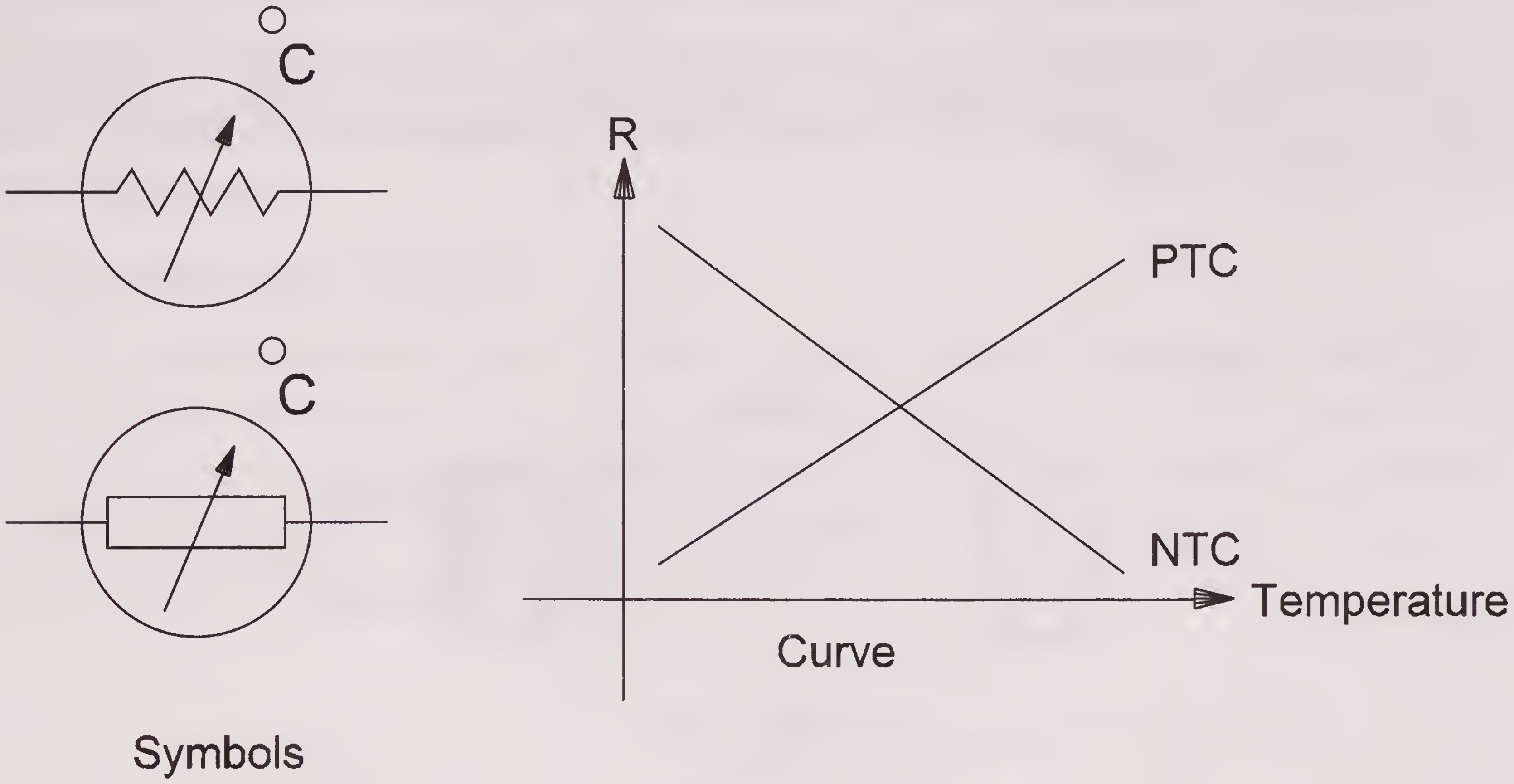

Symbols

Curve

Figure 60

The size of the component determines how fast it is when altering its resistance during a temperature change. This is an important factor in some applications and is referred to as readiness. The format and size depend on the application.

Specifications

The NTCs and PTCs can both be specified by a manufacturer or part number and a nominal resistance, or a resistance in the ambient temperature (normally 20 degrees Celsius). Also refer to Figure 60 for the characteristic curve, a graphic representation of the changes of resistance with the temperature.

Where they are found

The electrician will find NTCs and PTCs in many electronic circuits used in electric installations or even in electronic appliances. Some fire alarms and ambient temperature controllers used with air conditioners, as well as many other devices, use one of these components as a sensor. They are installed in a place where the temperature can be detected. The signal is then sent to a remote device that acts on the element to be controlled (air conditioner, alarm, fan, relay, etc).

Testing

The NTCs and PTCs can burn or break. In this case, the resistance will increase and it can open the circuit. The simplest test consists of taking the resistance measurement using a common multimeter.

Moisture can alter the nominal resistance of high-resistance NTC or PTC inducing a circuit into a wrong operation. This is important to be considered if the NTC or PTC is used as a sensor and not protected from moisture or ambient action.

VDRs

Voltage Dependent Resistors (VDR) or Zinc-Oxide Varistors and Metal Oxide Varistors or Transient Surge Absorbers are devices having a resistance that changes with the applied voltage. They are nonlinear resistors and can be used in circuit protection and many other applications such as power supplies, microprocessor protection, surge protection for consumer and industrial electronics, and many other applications.

The VDR is a device that becomes a conductor when the voltage applied to it crosses a determined value.

Symbols and types

The symbols and examples of common VDRs are shown in Figure 61. The size of the VDR determines the amount of energy it can absorb when becoming conductor.

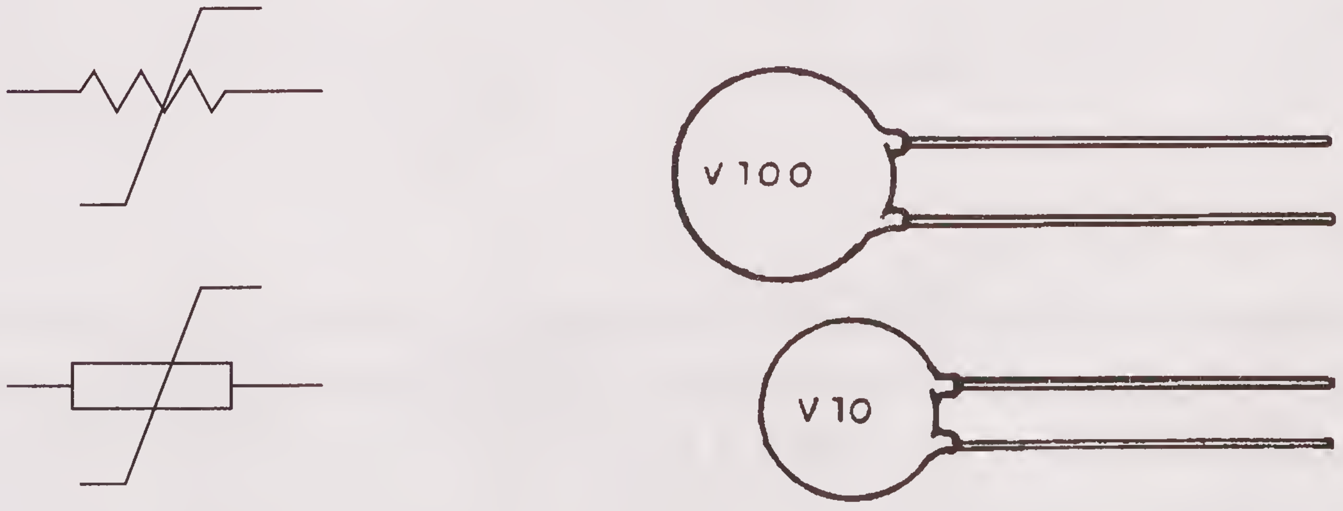

Symbols

Examples

Figure 61

Specifications

The main specification of a VDR is the breakdown voltage. It is the voltage at which the device becomes conductive, passing from a high-resistance state to a very low-resistance state. This component can be found with voltage specifications between 18 and 1800 volts.

Where they are found and how they are used

The electrician will find VDRs at the input of many electronic devices operating as a protective element. In fact, when voltage spikes appear in the AC power line, if their value is high enough to make the VDR conductive, the snap action of this component puts a short in the spike. So, the high voltage that could damage the equipment is shunted to the ground, or the energy of the spike absorbed by the VDR.

Computer outlet sockets with surge protection use this device as a main element in the protection circuit. Outlets for computers with surge protection such as the one shown in Figure 62 use VDRs as protection elements.

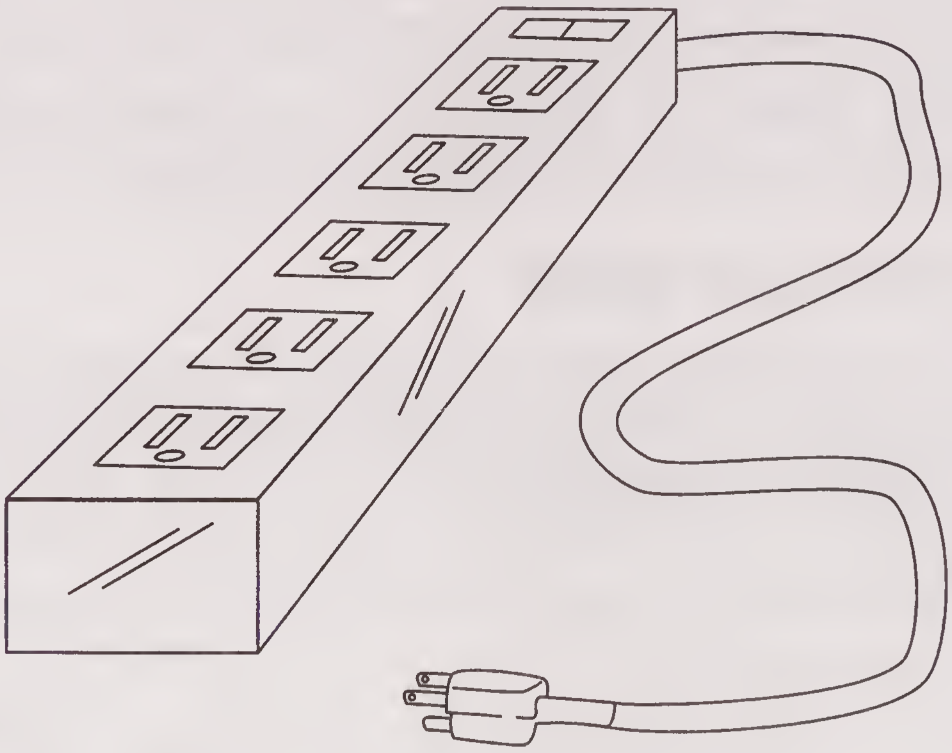

Figure 62

Testing

The effect of the absorption of the energy spikes coming from an AC power line or telephone line makes the VDR dissipate their energy as heat. This effect is cumulative, and after a lag in time of use, the VDR becomes conductive independently from the input voltage acting as a short. In this case, it must be replaced. The replacement type must have the same voltage as the original and be the same size.

The multimeter can be used for testing. A good VDR presents a high resistance above a predetermined hundred thousand ohms.

Capacitors

Capacitors have the function of storing up electrical energy. The basic capacitors are devices formed by a piece of insulating material between two pieces of conductive material. When DC voltage is applied to the capacitor, the electricity carried by the electric charge is stored up to each electrode. The capacitor energy lies in the electrical field formed between the pieces of metal or electrodes called plates. The insulator material is also referred to as dielectric. In electronic equipment, capacitors are used in many important functions. They can be found in many formats, sizes, and with different dielectric materials. The name of a capacitor is given by the material used as the dielectric. For example, a capacitor that uses ceramic as its dielectric is a ceramic capacitor, and a capacitor using mica as its dielectric is a mica capacitor.

Symbols and Types

The types of commonly used capacitors and the symbols used to represent them are shown in Figure 63.

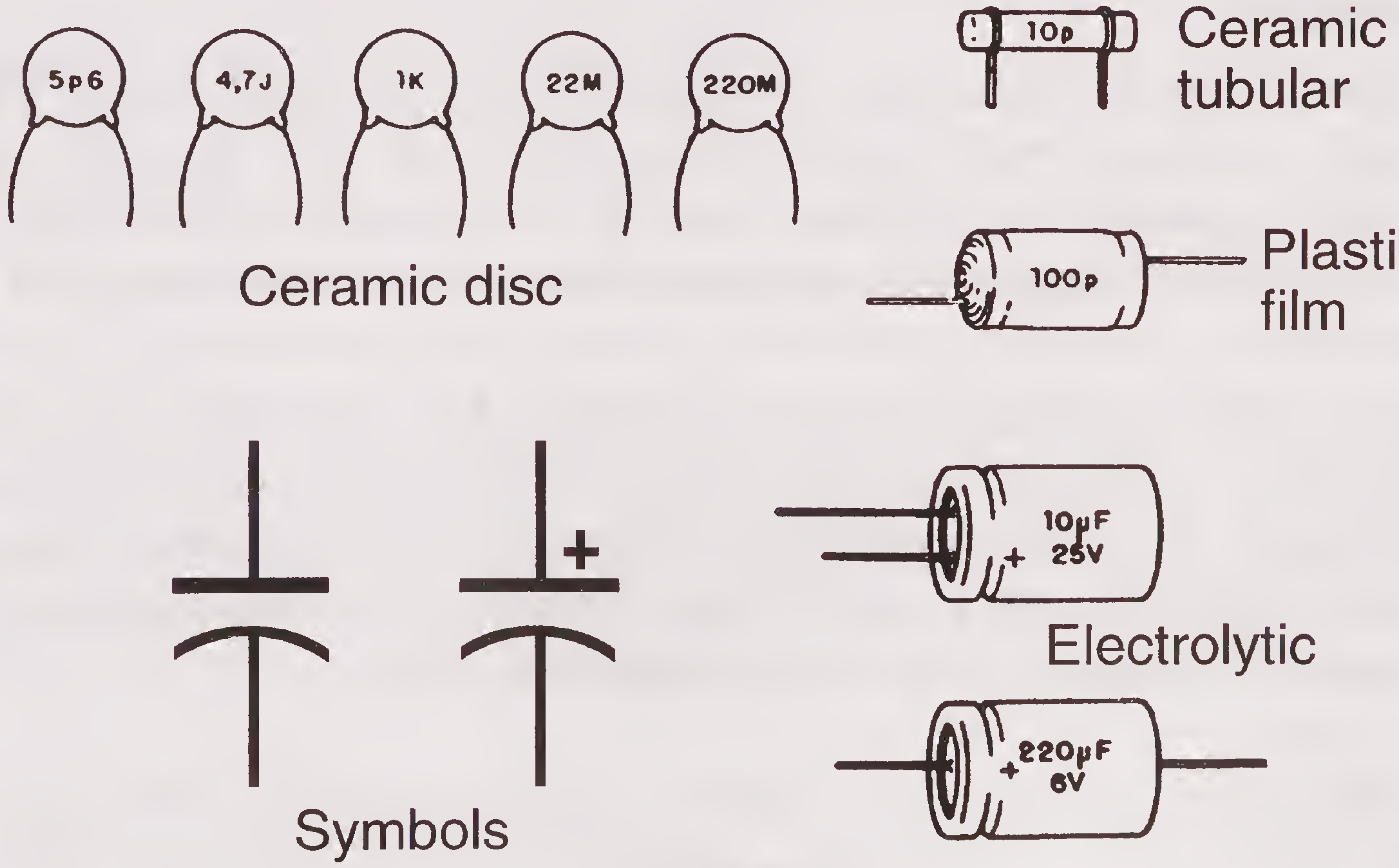

Figure 63

Some differences in representations are also found between diagrams as the European symbol used to represent capacitors is different than the American.

About the types, notice the differences in the manufacturing processes that result in tubular types (where the metal plates and dielectric form a coil) and the plate (where the dielectric is a planar piece).

The differences in types are very important as each capacitor has specific applications according to some electric properties of the material used and the format. For example, ceramic capacitors are suitable for high-frequency applications although polyester film is not. The plastic film capacitors such as polyester and polypropylene are used in low and medium frequency applications.

It is also important to know that electrolytic capacitors are polarized components. This means that they have definite positive and negative electrodes or plates. They can't be inverted when used in a circuit because this can cause the breakdown of the material used as dielectric conductor. In some cases, if a large current flows through the component during a fail, an explosion can occur.

Specifications

The electrician, when replacing a capacitor, must be aware of these specifications:

A. Capacitance

Capacitance is measured in farads. In normal applications the capacitances are very small when compared to this unit. So, the capacitances are expressed in multiples such as microfarads, nanofarads, and picofarads. Table 1-5 gives the conversion factor, the symbol, and the value of these units.

As in a resistor, some capacitors are too small to have the value printed on them. Some codes are used to express the capacitance of a capacitor. One of these is the "three-digit code." This code is formed by three numbers, or two numbers and a letter. If two numbers and a letter are used, it is easy to read the value as shown by these examples:

10n = 10 nF
47p = 47 pF
3.3n = 3.3 nF

Unit	Symbol	Value in Farads (F)
Microfarad	μF	0.000 001 = 10^{-6} F
Nanofarad	nF	0.000 000 001 = 10^{-9} F
Picofarad	pF	0.000 000 000 001= 10^{-12} F

Table 1-5

To convert	In	Multiply by:
Microfarads	Nanofarads	1,000
Microfarads	Picofarads	1,000,000
Nanofarads	Picofarads	1,000
Nanofarads	Microfarads	0.001
Picofarads	Microfarads	0.000 001
Picofarads	Nanofarads	0.001

If three numbers are used, the first two digits show the first and second figures of the capacitance and the third becomes a multiplier which determines how many zeros are to be added to the capacity as in these examples.

102 = 10 00 = 1000 pF

473 = 47 000 = 47 kpF or 47 nF (k = 1000)

Under 100 pF only two digits are displayed: 27 is 27 pF.

Figure 63, where many types of capacitors are shown, gives some samples with this kind of value indication.

Working Voltage

The capacitance of a capacitor depends on the distance between plates and the nature of the material used as a dielectric. The thicker the insulator in a capacitor, the greater its volume, but the smaller its capacitance. Otherwise, if the insulator is very thin, it can't isolate high voltages. Also, the physical construction of a capacitor is important to indicate the highest voltage that can be applied between plates. This is the maximum operating voltage, sometimes expressed as a WVDC (Working Voltage DC).

Types with working voltages in the range from very low voltages to more than 1kV are found in electronic equipment.

Type

The principal types of capacitors found in electronic circuits are:

- Electrolytic capacitors (Electrochemical type)

This type of capacitor uses aluminum as the electrode and a thin oxidization membrane as the dielectric. This layer is so thin that high capacitances can be achieved. But, since the layer is formed in an electrolytic process (thus the name), they are polarized. The pole is identified by a mark in the component (see Figure 64). Electrolytic capacitors have a pole indication that must be observed when installing them. Electrolytic capacitors are found in a capacitance range from 1 uF to more than 220,000 uF and are used in DC and low-frequency circuits.

- Tantalum capacitors

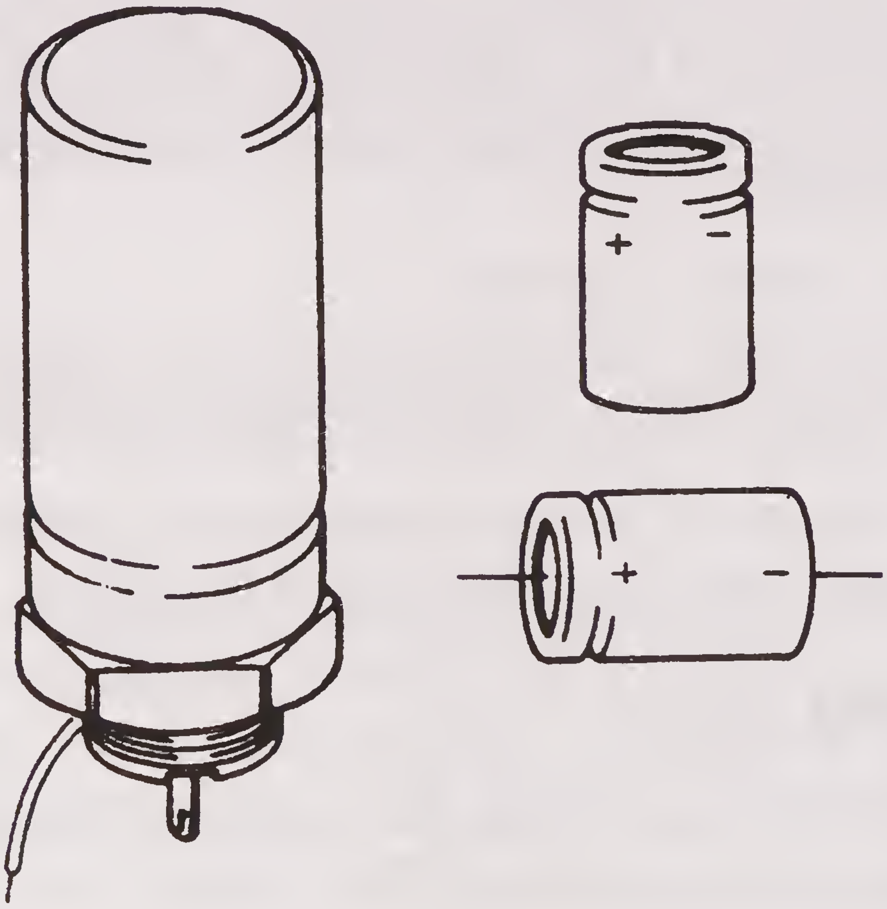

Figure 64

This is a special type of electrolytic capacitor using the material called tantalum for the electrode. Since tantalum oxide has greater dielectric constant when compared with aluminum, tantalum capacitors can be made with very large capacitances and small sizes. Even this capacitor there is polarity. It usually uses the symbol (+) to show the positive lead. Tantalum capacitors are found in a range of values starting from 0.1 uF.

- Ceramic capacitors

Special ceramics, such as titanium acid barium, are used as a dielectric in this kind of capacitor. The electric properties of the dielectric make this capacitor useful in high-frequency applications. The most common type is the disk style. The capacitance of ceramic capacitors is comparatively small. They are typically found in values between 1 pF and 470 nF.

- Polystyrene film capacitor

This capacitor is included in the family of the plastic film capacitors. A thin film of polystyrene is used as a dielectric in this capacitor. Since the plates and the dielectric form thin foils that are wound in a coil, the component presents some inductance. Because of this factor, this capacitor is not recommended for use in high-frequency circuits.

- Polyester film capacitor

Another type of plastic film capacitor is the polyester film capacitor. This kind of capacitor is recommended for high-frequency applications and can be found in a range from very low picofarads to 10 uF or more.

- Other capacitors

Other capacitors in this component family are polypropylene capacitors used in applications where low tolerances are needed, mica capacitors used in applications where high stability is needed, and metalized polyester film capacitors used in miniaturization applications.

In antique equipment other types such as oil/paper capacitors can be found. These capacitors use a paper infused in oil or dry paper as a dielectric. When restoring a piece of antique equipment where a damaged oil or paper capacitor is found, it can be replaced by an equivalent plastic film capacitor.

Where they are found

The electrician is familiar with capacitors since they are used with motors and in other electrical applications. However, the capacitors found in electronic equipment are found in a much larger assortment of types, values, and sizes as well as in every part.

Capacitors have special electric properties that make them very important to the performance of all electronic circuits. They can be used to measure time intervals, to perform calculus in computers, and to separate

AC signals from DC signals in audio and high-frequency circuits. They are also used as filters and for energy storage in many applications.

Testing

When connecting the probes of an analog meter to the terminal of an electrolytic capacitor (10 uF or greater) a current flow is registered by the needle's movement. This current flows while the electric charge is storing up. The current stops flowing when the charge finishes storing. The oscillating movement of a needle in a multimeter can be used to test an electrolytic capacitor. Use an intermediate resistance scale for this task.

Once charged, capacitors are like open circuits. When perfect, they can't let the current pass through them. This means that good capacitors have a very high resistance that can be measured with the aid of any common analog or digital multimeter. If a capacitor presents low resistances (under a hundred thousand ohms) we say that it presents losses or is shorted and must be replaced.

Variable Capacitors

In some applications a capacitance presented by a capacitor must be changed. This is the case in a radio receiver where the station tuned is determined by a capacitance of a resonant circuit. Every time we need to change the station the capacitance in that circuit must be altered. This task needs the aid of a special capacitor. The capacitor that can change the capacity is called a variable capacitor.

Like variable resistors, variable capacitors are found in several types. All of them are manufactured starting from the same principles: a moving plate can slide in front of a fixed plate with an insulator between them, or a moving plate can change the distance from a fixed plate also with a dielectric between them.

Two basic variable capacitor types are found in electronic applications: the trimmer, one that can be adjusted by a screwdriver, and the variable capacitor, or varicon, that is adjusted by a button placed in its axis.

Symbols and Views

Figure 65 gives the symbols used to represent variable capacitors and views of the most common types.

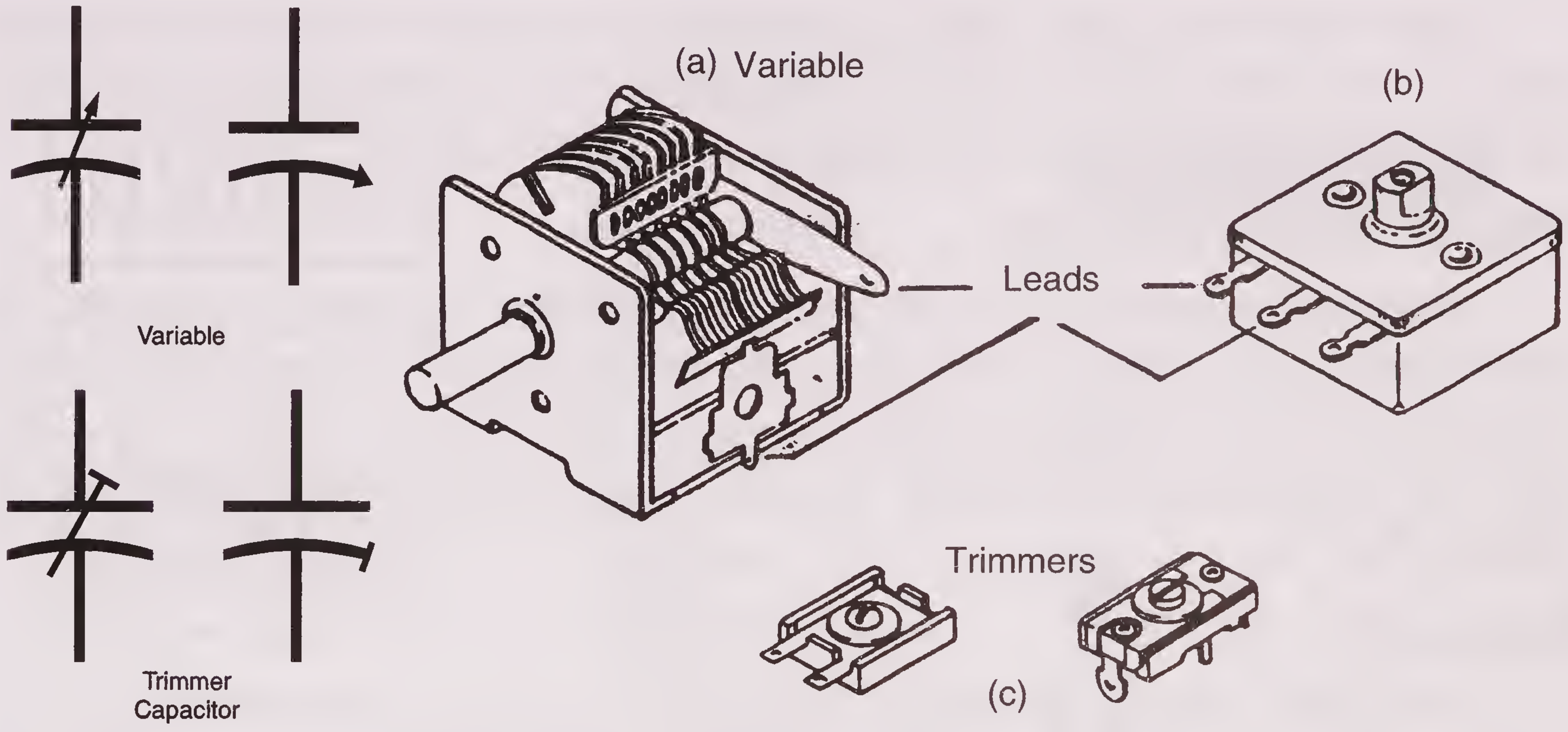

Figure 65

The variable capacitor shown in (a) is found in antique devices such as radios and tuners. The dielectric is the air gap between the moving plates and the fixed plates. In (b) are variable capacitors like those found in modern small transistor radios. In this type the dielectric is a polyester foil between the moving plates and the fixed plates.

The trimmer capacitors can be found with plastic, ceramic, and mica as the dielectric. Adjustment is made by a screw that moves one of the plates as shown in Figure 65 (c).

Specifications

Two main specifications are important when choosing variable or trimmer capacitors.

1. Capacitance

For variable capacitors, normally maximum capacitance is shown as nominal capacitance. Thus, a 120 pF capacitor is a capacitor that can change capacitance between a smaller value and 120 pF. However, even when totally open the capacitance isn't zero, which is normal, to indicate the minimum capacitance. For trimmer capacitors the capacitance range is given. For example, a 2-20 pF trimmer capacitor can be adjusted to present any value between 2 and 20 pF.

2. Type

Type indicates how many sections the capacitor has and of what material the dielectric is made. A two-section, air-dielectric capacitor with 365 pF sections is a type commonly found in old radios.

Where they are found

Variable capacitors are found in equipment operating with RF or radio signals like radios, remote control receivers and transmitters, wireless telephones, wireless communicators, and many others. They are used for adjustment or tuning circuits, placing them in a determined frequency.

Testing

The main problem of variable capacitors is shorts between plates. To test this, measure the resistance between the fixed and moving plates (the component must be disconnected from the circuit for this test). The resistance must be infinite in any position of the adjustment to indicate a good device.

Coils or Inductors

Coils, chokes, or inductors are components made of a certain number of turns of wire on a form. They are used to provide an inductance wherever it may be required. Inside the form is air (no core or air core) or some magnetic material such as ferrite (iron dust), iron, etc. The coils present opposition to current changes characterized by an electric quantity called inductance. The inductance is measured in henry (H) but, the fractions—millihenry (mH) and microhenry (μH)—are widely used in practical components. The number of turns and the gauge of wire determine the inductance of an inductor or coil.

Three major classes of coils or fixed inductors are available. They are the filter choke, audio frequency inductor, and the RF inductor or choke. (RF means radio frequency).

Symbols and types

Figure 66 shows some types of coils or inductors the electrician can find in electronic equipment.

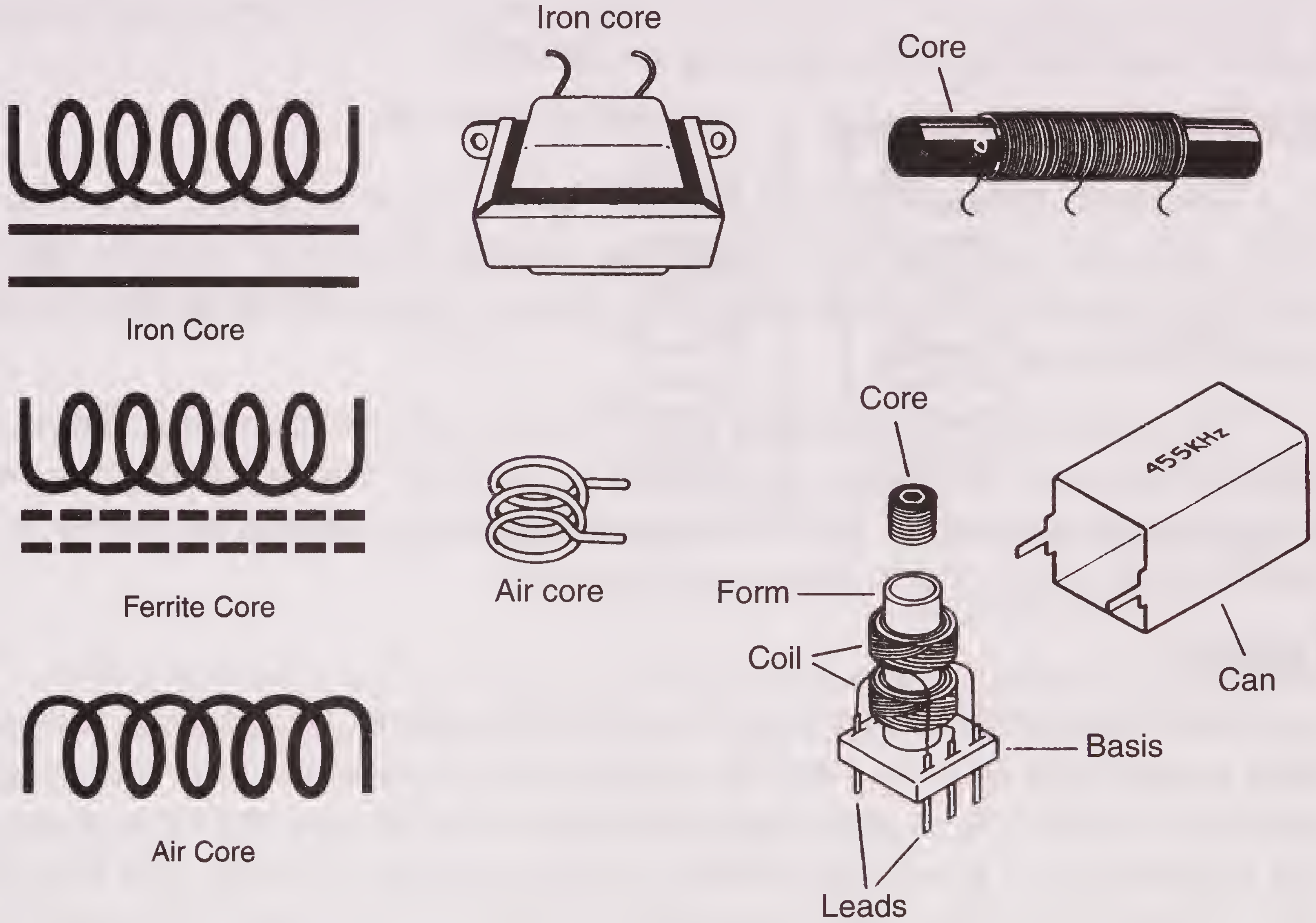
Iron core
Core
Iron Core
Core
455KHz
Ferrite Core
Air core
Form
Coil
Can
Basis
Air Core
Leads

Figure 66

The presence of a continuous line in the symbol indicates an iron core. The presence of an intermittent line in the symbol indicates a ferrite core. If the line is absent the coil is an air core type.

Adjustable coils can also be found in some applications. These coils have an adjustable core that can be slid or screwed in their form.

Specifications

A. Inductance

The inductance of a coil is expressed in henry (H) or its fractions. Low values in the range of microhenry are found in high-frequency applications. Values in the range of millihenry and henry are found in low-frequency circuits and as filter chokes.

B. Core

The nature of the core is an important specification for a coil. Low-frequency and DC coils (filters and audio) use iron cores. High-frequency coils use ferrite and air cores. It is also important to indicate the size and shape of the core. Ferrite toroids or rods are used for high-current chokes and RF applications. Pot cores are an example.

Where they are found

The lower the number of turns, the lower the inductance. Low-inductance coils are used in high-frequency circuits like tuned circuits, filters, etc. High-inductance coils are found in power supply filters, audio circuits, musical instruments, etc.

The use of large inductance coils in a circuit is not common. Because large inductance coils are heavy and expensive components, they are avoided in most projects. Circuits that can simulate inductance using other components are found in some applications.

Testing

The main problem of a coil is the interruption of the wire and the short caused by excess heat or mechanical problems. The test of coil continuity can be done using a multimeter on the lowest resistance scale. The resistance of a coil, depending on the number of turns and the wire used, should be in the range between 0 and 10,000 ohms. The test for a short is more difficult. Generally, a coil in short, when caused by excess current, presents burning signals such as dark spots and the smell of smoke.

Transformers

Transformers are components formed by two or more coils or inductors coupled so closely together that the magnetic field of one causes induction of current in the others. For the total coupling the coils are wound in a common form or core. When an alternating voltage is applied to one coil of a transformer (called the primary winding), this voltage induces alternating voltages in the other coils (secondary windings) depending on the ratio of turns of the two windings. For example, if a transformer has a primary winding with 200 turns and a secondary with 400 turns, and 110 VAC is applied to the primary, 220 VAC can be found in the secondary, as shown in Figure 67.

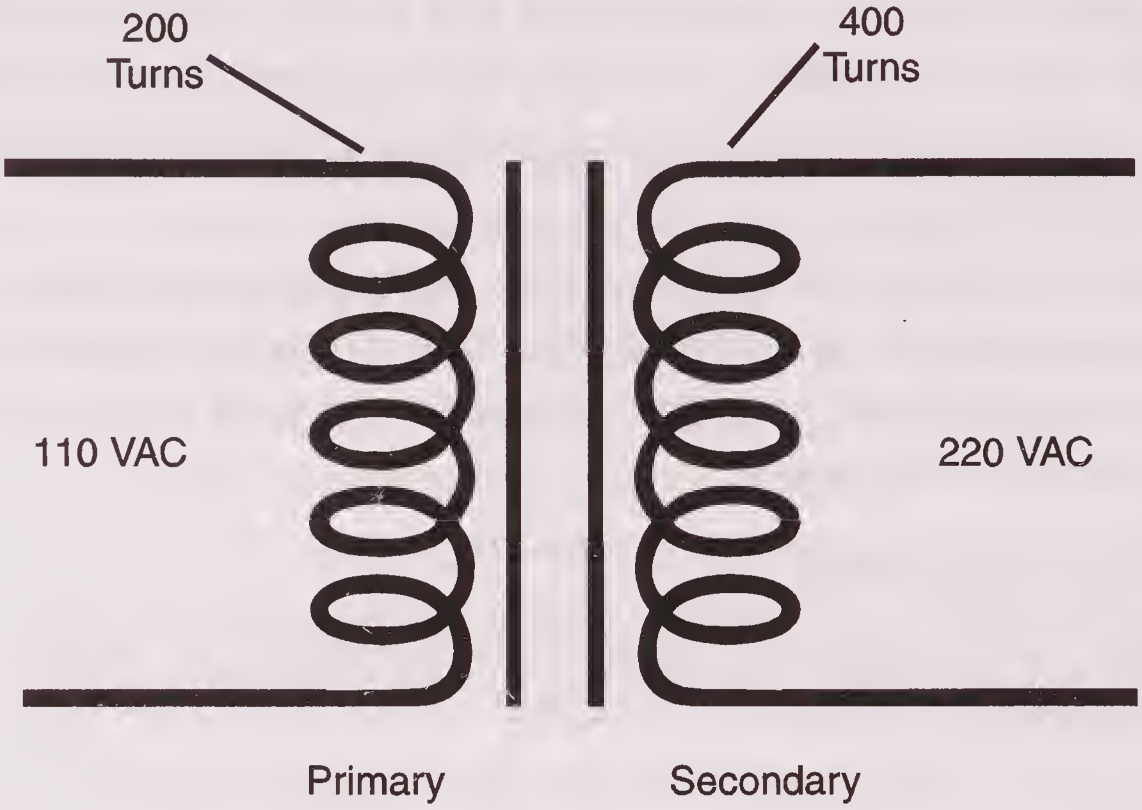

Figure 67

Transformers can be used to change voltages, currents, and impedances in an alternating-current circuit or high-frequency circuit. The transformers can be wound in forms with air cores, ferrite cores, or iron cores according to the application. Air and ferrite cores are used in high- and medium-frequency circuits and iron cores are used in medium- and low-frequency circuits.

Symbols and Types

Figure 68 shows the different symbols used to represent transformers and their types.

The presence of a material core (ferrite or iron) is represented by a continuous or intermittent line. If the line is absent it is an air core transformer.

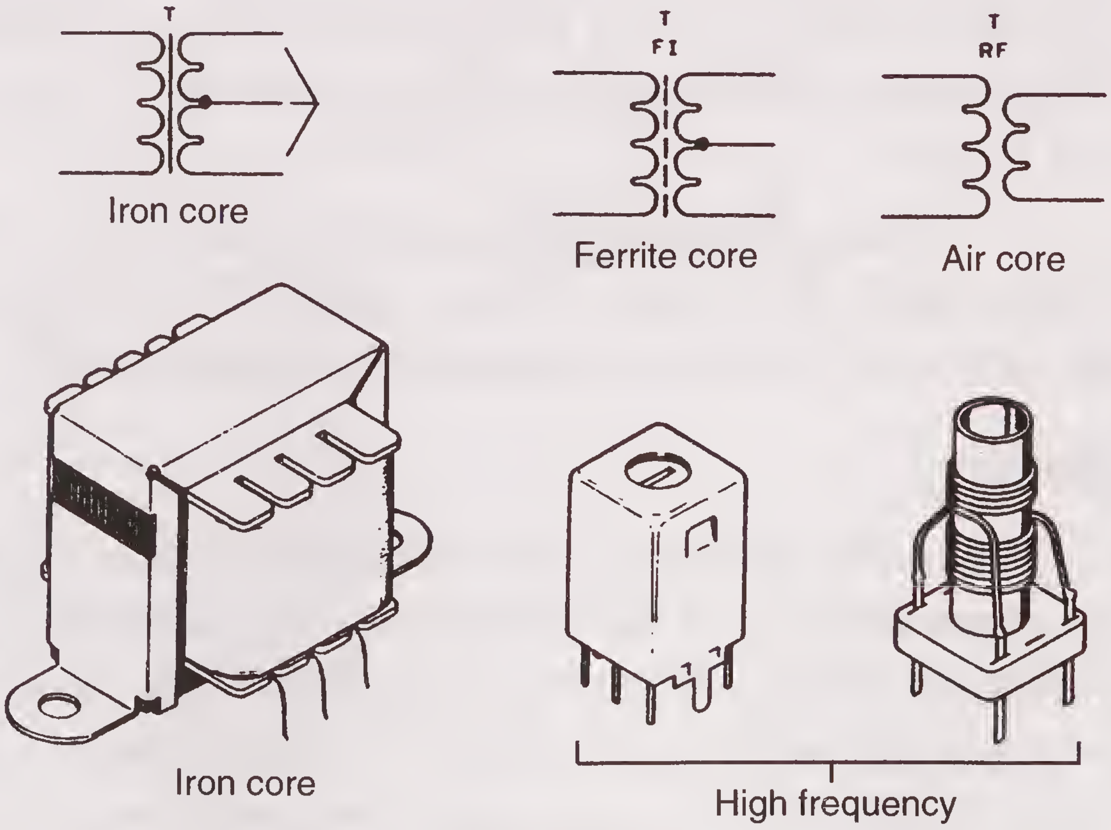

Iron core

Ferrite core

Air core

Iron core

High frequency

Figure 68

Specifications:

Audio and power supply transformers

- Input voltage or impedance—The primary winding or the input of a transformer is specified by the voltage such as the transformers found in power supplies. In transformers used as an impedance match in audio or other applications, the primary winding is specified by an impedance in ohms.
- The output voltage—The secondary winding of a power supply transformer is specified by the output voltage and current. For audio transformers the output impedance, and in some cases, also the output power, are given.

RF transformers

- Number of turns—In RF transformers the number of turns and the wire gauge are specified, mostly in circuits to be mounted.
- Core—The type of core and format is an important specification for high-frequency transformers.

Where they are found

The electrician will find transformers of many sizes and types in all electronic equipment. The most common types are the transformers used in the power supply of any equipment. The windings are isolated one from the other (the electric energy is transferred from one to the other by the magnetic field), so they are important as an insulation element in circuits. The presence of a transformer avoids shock hazards when touching parts of a circuit.

In audio and high-frequency circuits it is also possible to find transformers with air cores or ferrite cores. Computers and some audio amplifiers use toroid transformers in the power supply.

Testing

The simplest way to test a transformer is to check that the windings are not interrupted. This can be done using a common multimeter on the lowest resistance scale. RF transformers with few turns of wire must present very low resistances if they are good. Audio or power-supply transformers present higher resistances, in the range between low ohms and 10,000 ohms.

Important: the measured resistance of the winding of any transformer isn't its impedance but the ohmic resistance of the wire used to wind it.

Relays

Relays are electromagnetic switches. They are formed by an electromagnet coupled to switch contacts so that the switch is operated automatically when the electromagnet is energized. Applying a voltage to the electromagnet allows the devices wired to the switch to be controlled. Relays are important as they can be used to control high power loads from weak signals or currents. Another advantage in the use of relays is that the contacts are isolated from the control circuit, meaning a relay can be used to control a high-voltage circuit from a low-voltage circuit. The disadvantage is that the coupling between the coil and the contacts is mechanical, limiting the speed of response.

Figure 69 shows the basic construction of a relay as found in many electronic applications.

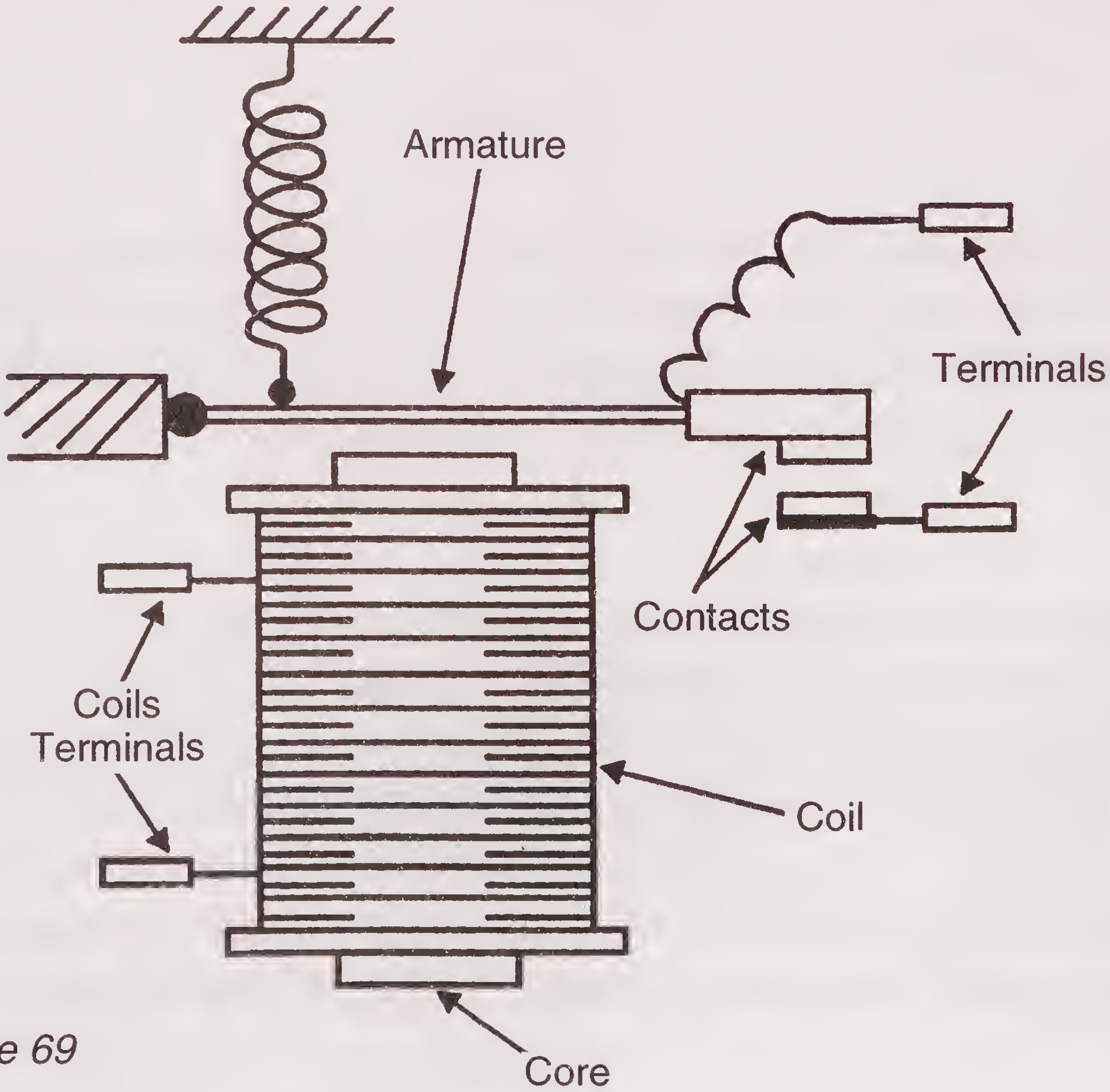

Figure 69

The coil is formed by hundreds or thousands of turns of thin, enameled wire. The contacts are designed to control high currents based on the final use of the device.

Symbols and Views

Figure 70 shows the symbols and views of the relays used in common electronic applications.

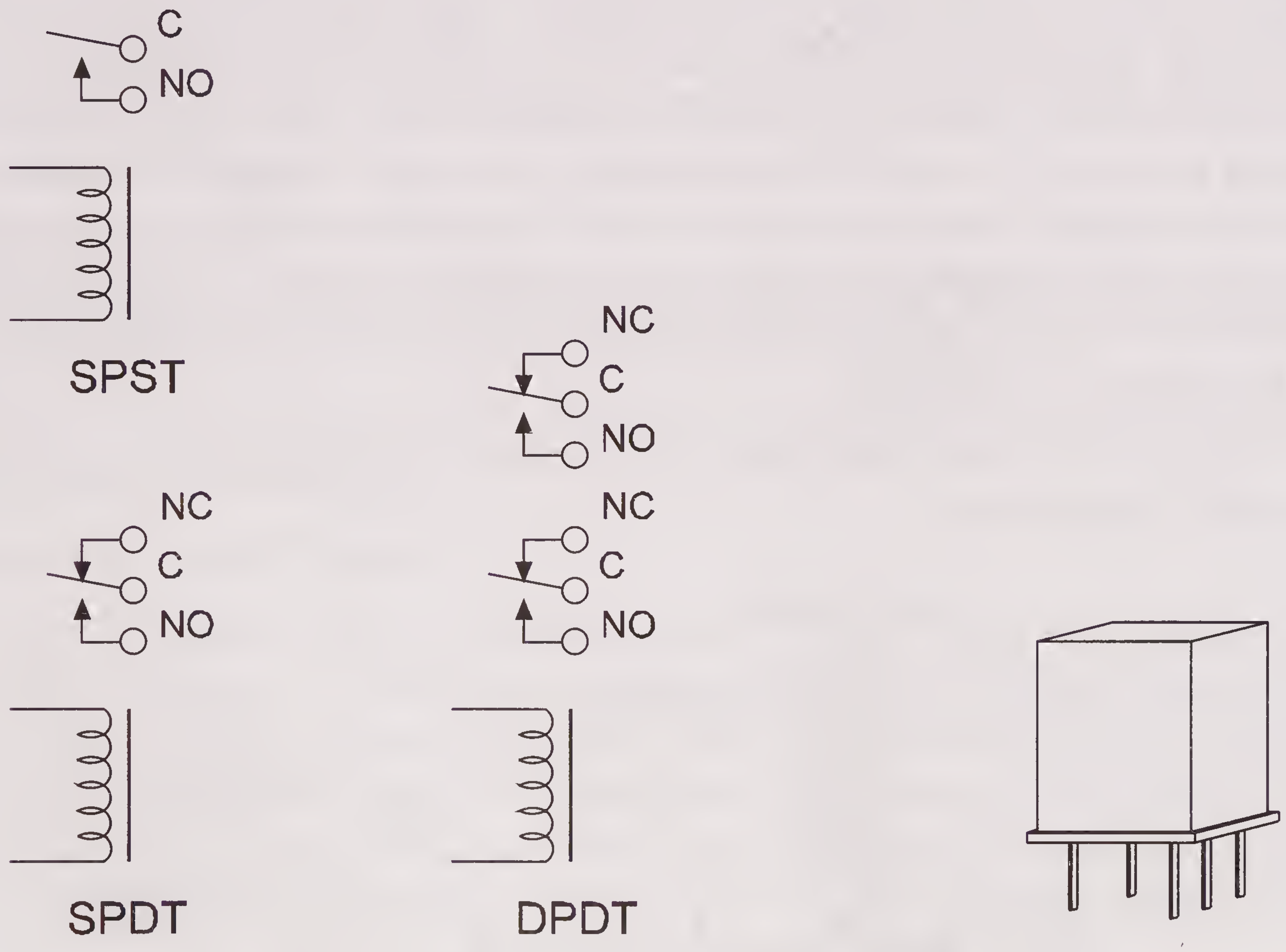

Figure 70

A relay can provide SPDT or DPDT depending on the contact placement. The numbers of contacts that can be controlled by the relay are shown in the symbols. Notice when the relay can have “normally open” contacts (NO) and “normally closed” contacts (NC). When controlling a load from the NO contacts, the load will be turned on when the relay’s coil is energized. Otherwise, if the NC contacts are used, the load will be turned off when the coil is energized. Figure 71 shows how these contacts should be used.

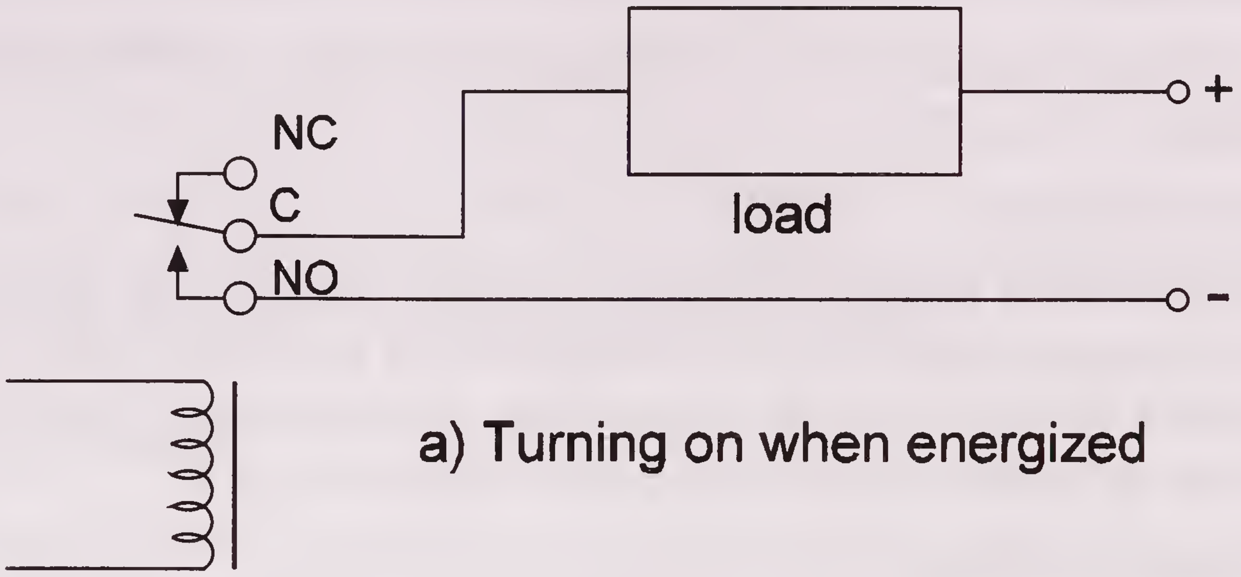

a) Turning on when energized

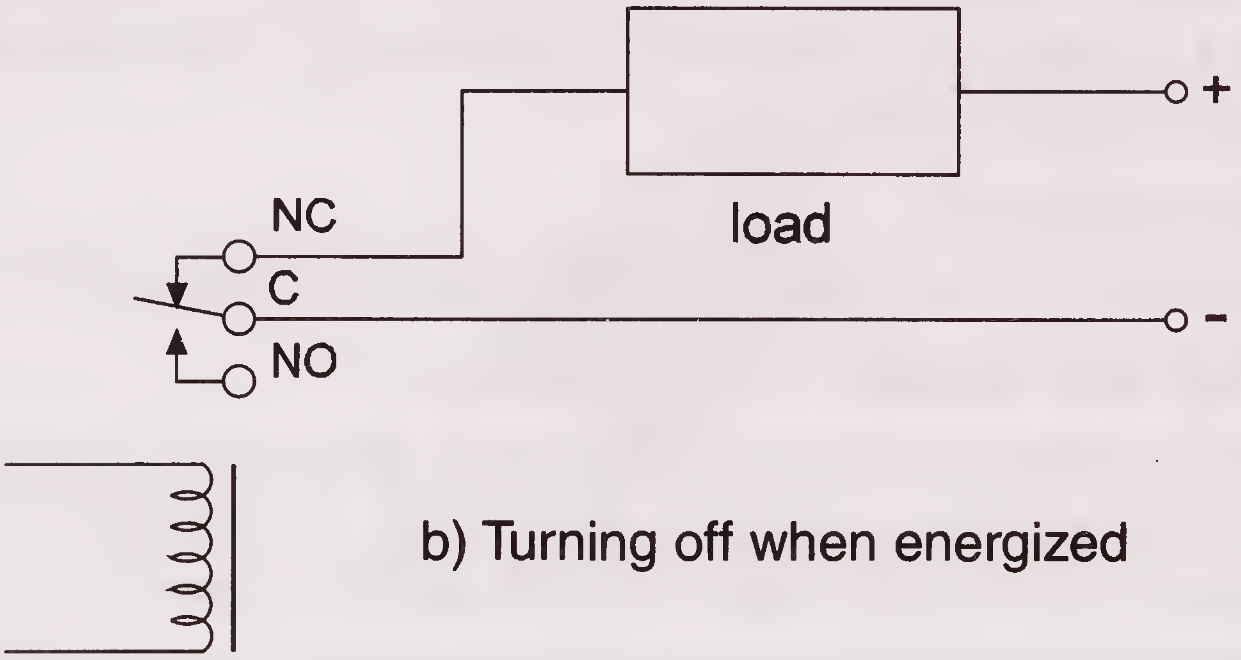

b) Turning off when energized

Figure 71

Figure 72 shows another important type of relay—the reed relay. This relay is formed by a coil involving a reed switch. The reed switch is formed by metal plates inside a glass enclosure that is filled with an inert gas. When a magnetic field acts on these plates, they touch one another, closing the circuit.

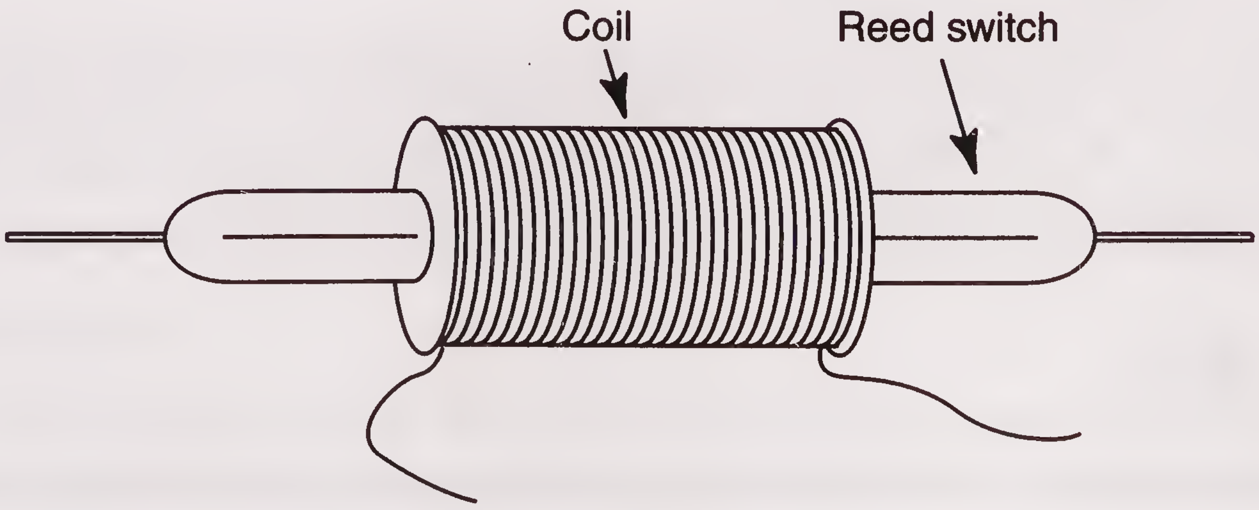

Figure 72

Specifications

When replacing a relay, the electrician must take care to observe these specifications:

A. Coil specifications

The current that turns a relay on is given by the coil specifications. The relay can be specified by the voltage to be applied to the coil in volts and the current, or by its ohmic resistance. For example a 12 V x 50 mA relay can also be indicated as a 240-ohm relay for 12 V.

B. Contact specification

This is the amount of current (and the voltage of the external circuit) that can be controlled by the relay.

C. Contact configuration

This is given by the number of contacts and their type—SPST or DPDT.

Where they are found

In electric circuits, the electrician usually finds relays controlling high electric currents such as in motors, lamps, etc. In electronic circuits it is most common to find small relays controlling small currents.

Small relays can be found in alarms controlling the siren, lamps, or other loads. In automatic doors, remote controls, and other applications, relays interface the control circuit with the controlled load.

Testing

Two kinds of tests can be done in a relay:

A. Coil: The state of the coil can be verified by a multimeter. A good coil presents resistances between some ohms and 10,000 ohms. This probe doesn't indicate a shorted coil.

B. Contacts: The contacts are tested by applying a voltage to the coil and measuring the resistance when they close. This resistance must be almost zero when tested with a common multimeter.

Solenoids

Solenoids are components formed by a coil with a moving core inside. When a current flows through the coil, the magnetic field acts on the core attracting it. Sliding inside the coil, the core can be used to move some mechanism.

Solenoids with different sizes and formats are found in electronic equipment. They are used to move mechanical parts of the appliances such as the tape-ejecting mechanisms in VCRs. Washing machines and dishwashers also use solenoids to control the flow of water.

Symbol and Views

The symbol of a solenoid and views of the common types found in electronic equipment are shown in Figure 73.

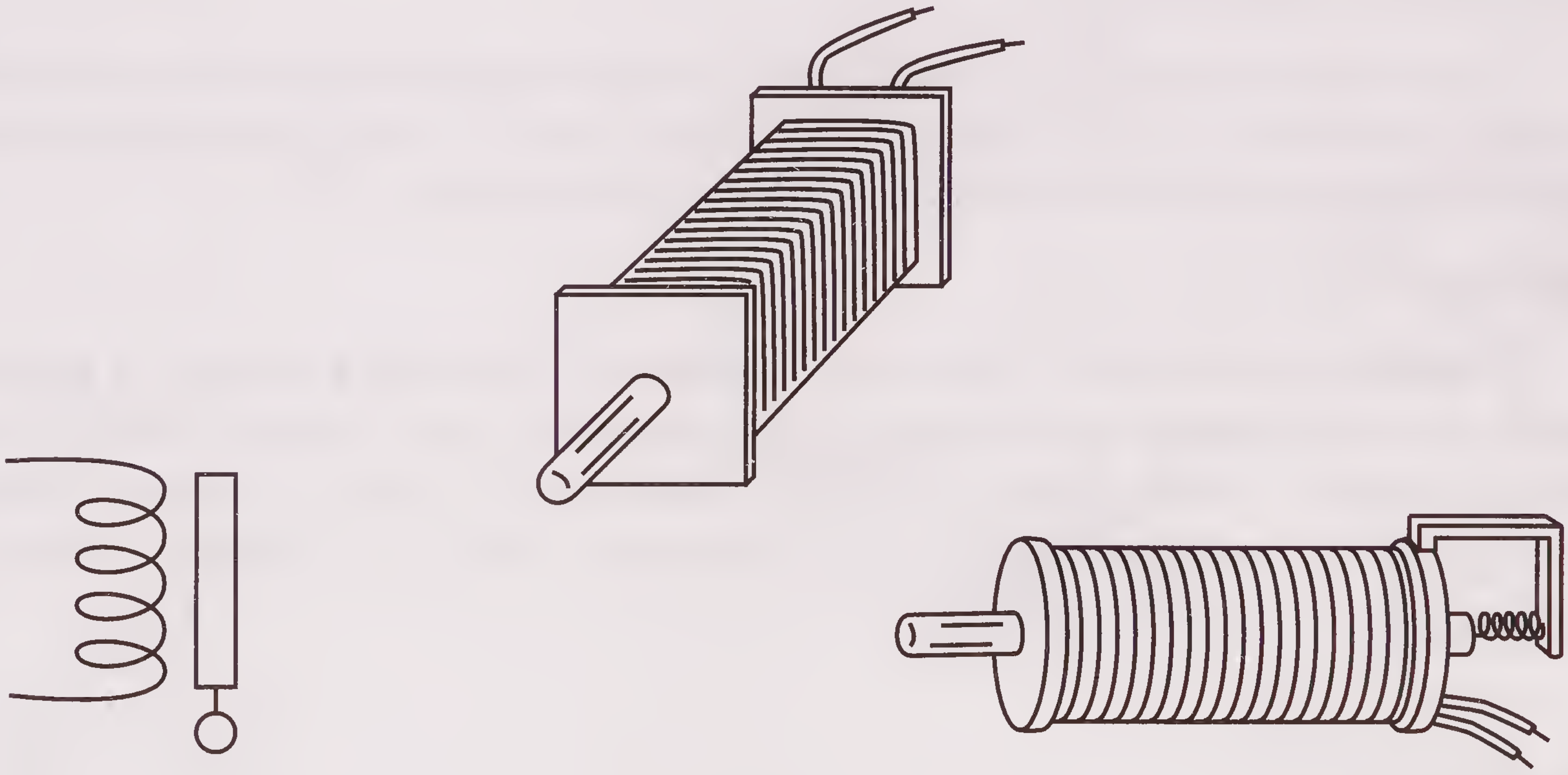

Symbol

Aspects

Figure 73

Small solenoids are formed by a coil with thousands of turns of a very thin, enameled wire. The core is a rod of iron that slides inside the coil. A spring and some mechanism can be found to return the core to the initial position when the current that causes the magnetic field disappears.

Specification

The main specification of a solenoid is the voltage that must be applied to energize it. As an important complement to this information, know the current drain when the nominal voltage is applied. In electronic circuits, solenoids with voltages ranging from some volts DC to AC power line voltage are found.

Where they are found

The electrician can find many types of solenoids in common electric appliances such as washers and electric doors. In electronic appliances, solenoids are found in items such as tape recorders, VCRs, CD players, and a large number of items with circuits where mechanisms must be controlled from electric currents. The solenoids found in these applications are generally very small and delicate. They are different from the ones found in electric appliances that are designed to operate with the AC voltage from the power line.

The important fact for the electrician to recognize is that solenoids used in electronic circuits are usually driven by DC currents supplied by active circuits using transistors and other components.

Testing

The electric test of a solenoid consists in the coil's probe. A multimeter, on the lowest resistance scale, can be used to see if the coil is not interrupted. A resistance varying from some ohms to a thousand ohms must be measured in a good solenoid. This test doesn't show a shorted coil.

Motors

In electronic circuits many types of small DC motors can be found. Basically a motor is a device that converts electric energy into motion. They can be used to move parts of a device from an in signal or control voltage. The electric motor is formed by one or more coils that produce a magnetic field when a current flows through them. The magnetic field acts on metal pieces or other coils causing a force that puts it in motion. In electronic applications, there are two main types of electric motors—DC motors and step motors.

These motors are formed by coils that, when energized, can place a shaft in a determined position. By energizing the coils in sequence, the shaft can run at a determined speed. DC motors produce continuous movement when a voltage is applied. Step motors can be used to place a piece in a predetermined position with great precision.

AC motors are more rare in electronic appliances. They are mainly high-power motors driving large mechanisms.

Symbol and Views

Figure 74 shows a parts view and symbols of some DC motors and step motors.

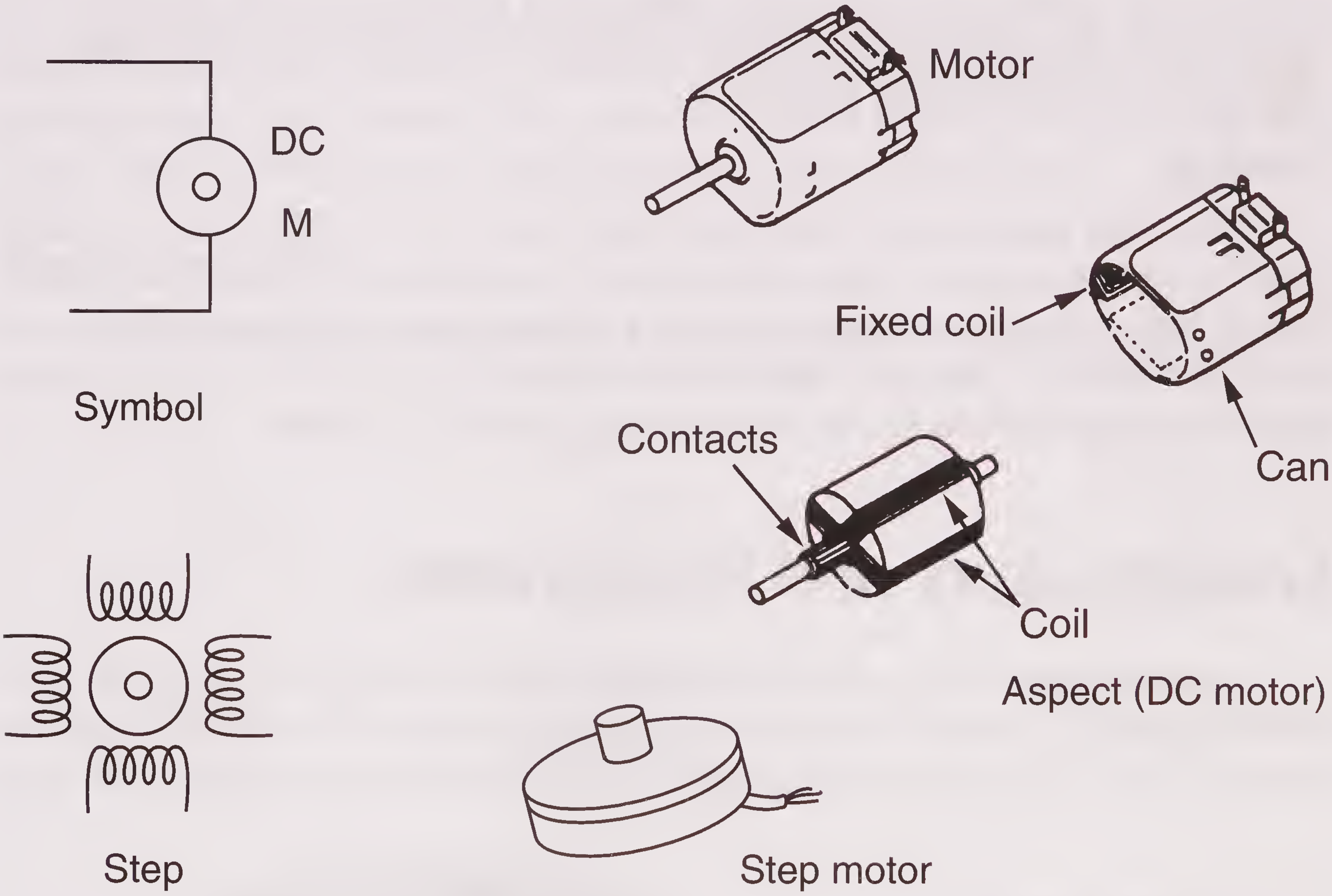

Figure 74

Specifications

DC motors are specified by the operational voltage and the nominal current. In these motors the current and the absorbed power are functions of the load. When running without a load the current is very low and the speed rises to the maximum value. When the load increases, the speed falls down and the drained current increases to the maximum. Another indication sometimes found is the speed given in rpm (rotations per minute) under nominal operation conditions.

The step motors are specified by the type (number of poles) and the voltage. The current drained is also an important specification for this kind of motor.

Where they are found

As stated, there are many electronic applications where motors can be found. Since they are delicate mechanisms, electricians will find that in more of the cases than not, the origin of a problem is not electric, but mechanical. Rotors and the coupling pieces such as the shaft can break or present other mechanical problems.

Testing

The only electric test the electrician can do on a DC motor or step motor is in the coils. An interrupted coil or a shorted coil is the main problem. A test using a multimeter adjusted to the lowest resistance scale can reveal if the coil is perfect. Resistances ranging from some ohms to some hundred ohms can be found when testing small DC motors.

Loudspeakers and Headphones

Loudspeakers and ear/headphones are transducers or devices that convert electric energy into sound (acoustic energy). The electric energy comes in the form of an electric signal whose frequency corresponds to a

Figure 75

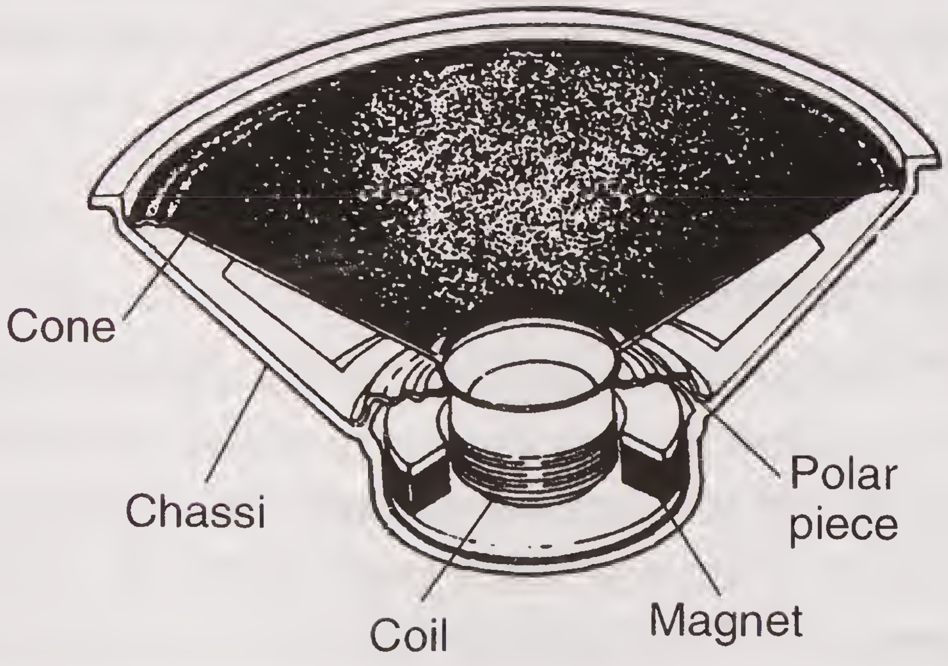

sound. The moving coil loudspeaker (the most common in audio applications) is formed by a coil involving a polar metal piece where a strong magnetic field is produced by a magnet as shown in Figure 75.

The coil is fixed in a cardboard or plastic cone. When the signal is applied to the coil, its magnetic field interacts with the magnetic field produced in the polar piece, resulting in a force that moves the coil and the cone forward and backward. Compression and decompression waves are produced in the air surrounding the cone. These waves have a waveform corresponding to the original sound being reproduced.

In magnetic headphones small devices that operate according to the same principle can be found. Remember that both loudspeaker and headphones exist that operate from other principles such as the piezoelectric units. (See in the item on transducers for more information.)

Symbols and Types

Figure 76 shows the symbols and the types of some loudspeakers.

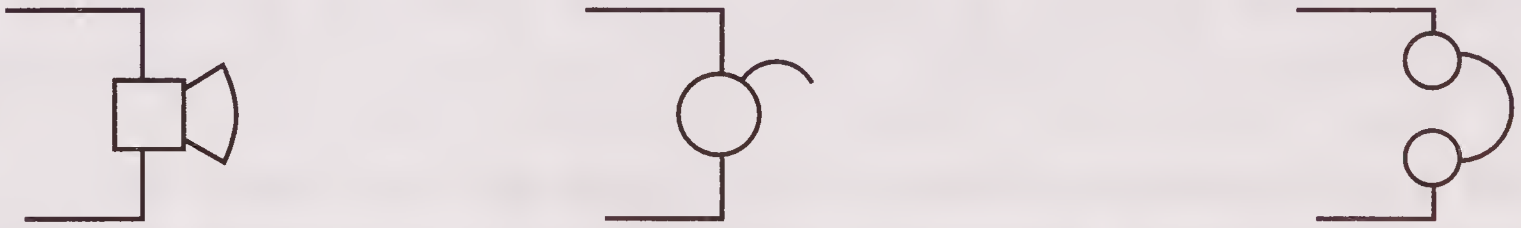

Loudspeaker (Symbol)

Earphone (Symbol)

Headphones (Symbol)

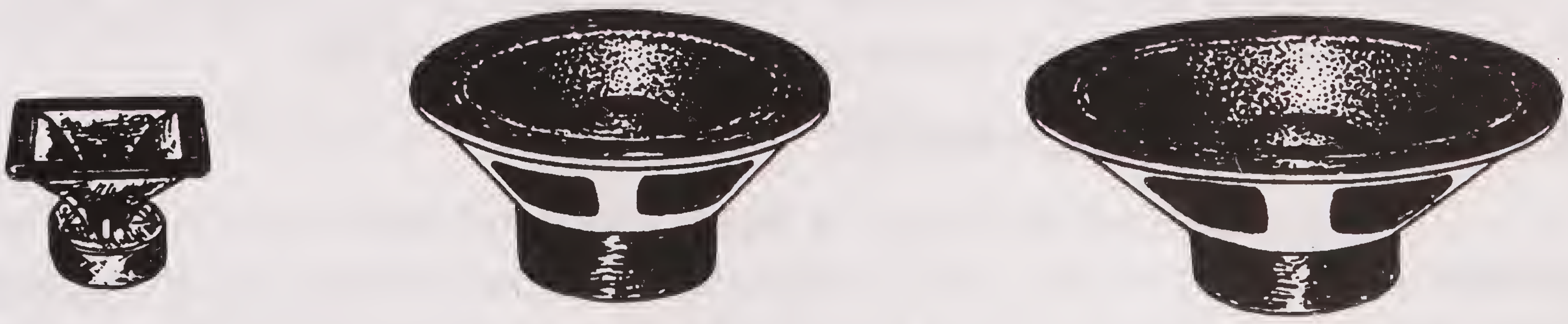

Tweeter

Mid-range

Woofer

Figure 76

The larger the size of a loudspeaker, the easier the production of low-frequency sounds. This means that heavy magnets and large diameters are characteristics of low-frequency loudspeakers, also named bass speakers or woofers. Small loudspeakers are used for the reproduction of high-frequency sounds. These small speakers are called treble speakers or tweeters.

It is important to remember that many loudspeakers are designed to operate only inside enclosures.

Specifications

A. Impedance: When replacing a loudspeaker, next to the size, the impedance is the first electric specification to be observed. The impedance of a loudspeaker is measured in ohms and should not be confused with the ohmic resistance of the coil. Common types have impedances in the range between 3.2 and 16 ohms. Never use a loudspeaker with less impedance than the recommended; the output stage of the driver circuit can be forced and the components damaged.

B. Power: The amount of electric power that can be applied to a loudspeaker and converted in sound is another important specification for these components. Power is measured in watts (W). You'll never use a loudspeaker with less power capabilities than the output of your audio equipment. The loudspeaker can be overloaded and the coil burned.

C. Frequency range: In some cases is important to make sure the loudspeaker isn't a special type designed for the reproduction of a determined frequency range. Loudspeakers for high-frequency sounds (treble) such as the tweeters have different characteristics when compared with the low-frequency (bass) loudspeakers.

Headphone characteristics

The main characteristic of a headphone or earphone is the impedance. Two types are used in electronic applications—low-impedance types ranging from 3.2 to 600 ohms and high-impedance types ranging from 1000 to 10,000 ohms. Never use a low-impedance headphone or earphone in an application where a high-impedance type is recommended.

Testing

The electric test of a magnetic loudspeaker or headphone consists in the measure of the continuity of the coil. A common multimeter on the lowest resistance scale can be used for this task. The measured resistance of a good loudspeaker or headphone is very low, below 100 ohms for the low-impedance types. See that the measured resistance isn't the impedance but the ohmic resistance of the coil. Higher values, up to several thousand ohms, are common in high-impedance types.

Magnetic Transducers

Transducers are devices that convert one form of energy into another. A magnetic transducer transforms variable magnetic fields into electric currents. In the inverse mode, we can also consider a magnetic transducer to be any device that converts electric signals into magnetic fields. Many magnetic transducers can be found in electronic applications functioning as sensors. Basically they are formed by a coil that can be used to pick up any change in the magnetic field produced by a magnet under certain conditions, or a coil that produces a magnetic field from a current.

Loudspeakers and headphones are magnetic transducers, as explained previously, but are treated as a separate item because of their importance in electronics. Microphones and diskettes that read heads and position sensors are other examples of magnetic transducers used as sensors. The writing heads of a Winchester disk or disk drive are magnetic transducers operating in the inverse mode.

Symbols and Types

The symbols and types of some magnetic transducers are shown in Figure 77.

Specifications

Depending on the application, the transducer can show specifications as current, voltage, type of signal, or a part number. The most common is a part number specification.

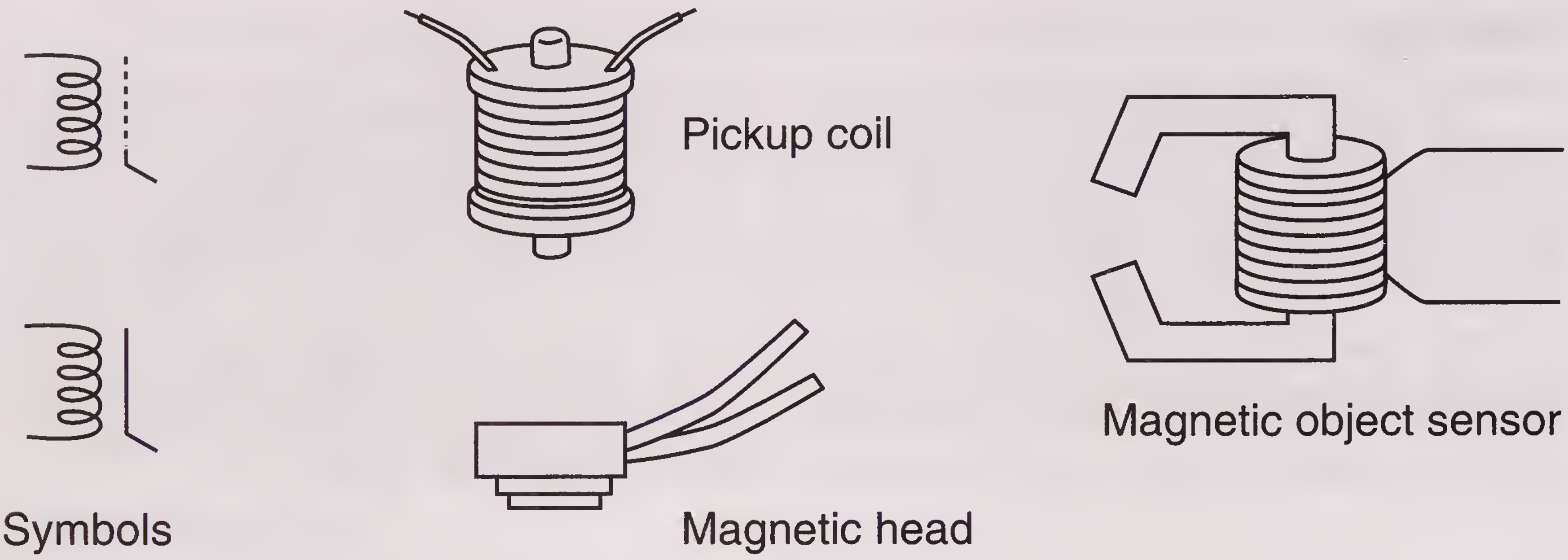

Figure 77

Where they are found

Pickup coils, sensors, microphones, Winchester disk, and disk drives are all examples of magnetic sensors.

Testing

A multimeter can be used to test the continuity of the coil. Magnetic transducers in general must present resistances ranging from less than 1 ohm to more than 1000 ohms based on the application. This test doesn't reveal other problems such as shorts or mechanical failures.

Piezoeletric Transducers

Some materials present piezoelectric properties due to their unusual crystal structure. Among these materials are quartz, Rochelle salt, and several kinds of barium and titanium ceramics. When a piezoelectric material is pressed, bent, or submitted to mechanical efforts or any change of form, electric charges appear in its surfaces. The same effect also works in the inverse mode; if we apply electric voltages to its surface, a mechanical deformation can occur. Some ceramic materials such as barium titanate and quartz crystals present piezoelectric properties.

These materials can be used to make piezoelectric transducers, piezoelectric sounders, piezoelectric loudspeakers and earphones, piezoelectric buzzers, and many other devices that convert electric energy into mechanical energy (sound or super sounds) or the opposite, such as microphones, vibration pickups, etc. Some components are very important in this family:

A. Quartz crystal: When excited by an electric signal quartz crystal tends to vibrate, producing signals of only one frequency. They can be used to determine the operational frequency of many circuits such as those in clocks, watches, computers, radio transmitters and receivers, and many others. The vibration frequency is established by the mechanical resonance of the crystalline plate.

B. Piezoelectric sound generators: A small piece of piezoelectric ceramic can be used to produce sounds in alarms, earphones, and high-frequency loudspeakers (tweeters). When an audio signal is applied, the ceramic vibrates producing sound waves. Ceramic transducers can also be used to produce supersonic sounds.

C. Piezoelectric high-voltage generators: Beating a piece of piezoelectric ceramic with a small, spring-triggered hammer, a ceramic piezoelectric transducer can generate very high-voltage pulses. The high voltage, up to 7000 volts, will produce a spark that is used in gas igniters and in other applications.

D. Piezoelectric microphones: Placing a diaphragm onto a piece of piezoelectric ceramic can be used to convert sounds or beats in electric signals. The sound waves beating the diaphragm make it press the ceramic piece producing the deformations needed to cause the electric signals to appear.

Symbol and Types

Figure 78 shows some piezoelectric transducers and their symbols.

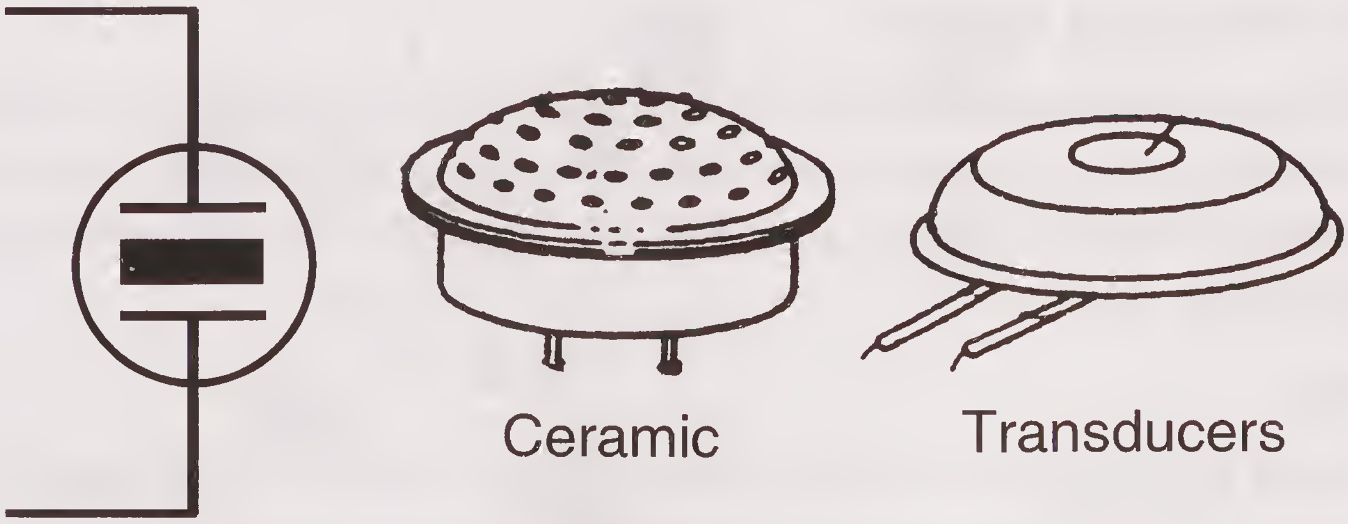

Figure 78

Specifications

Normally the piezoelectric components are indicated by a part number determined by the manufacturer. The quartz crystals are indicated by the frequency. They are high-impedance devices because the ceramic and the quartz are good isolators. In some cases, a drive circuit can be installed inside the component, such as in piezoelectric buzzers that include an audio oscillator. Piezoelectric tweeters, for example, include a small impedance match transformer to drive the transducer as shown in Figure 79.

Figure 79

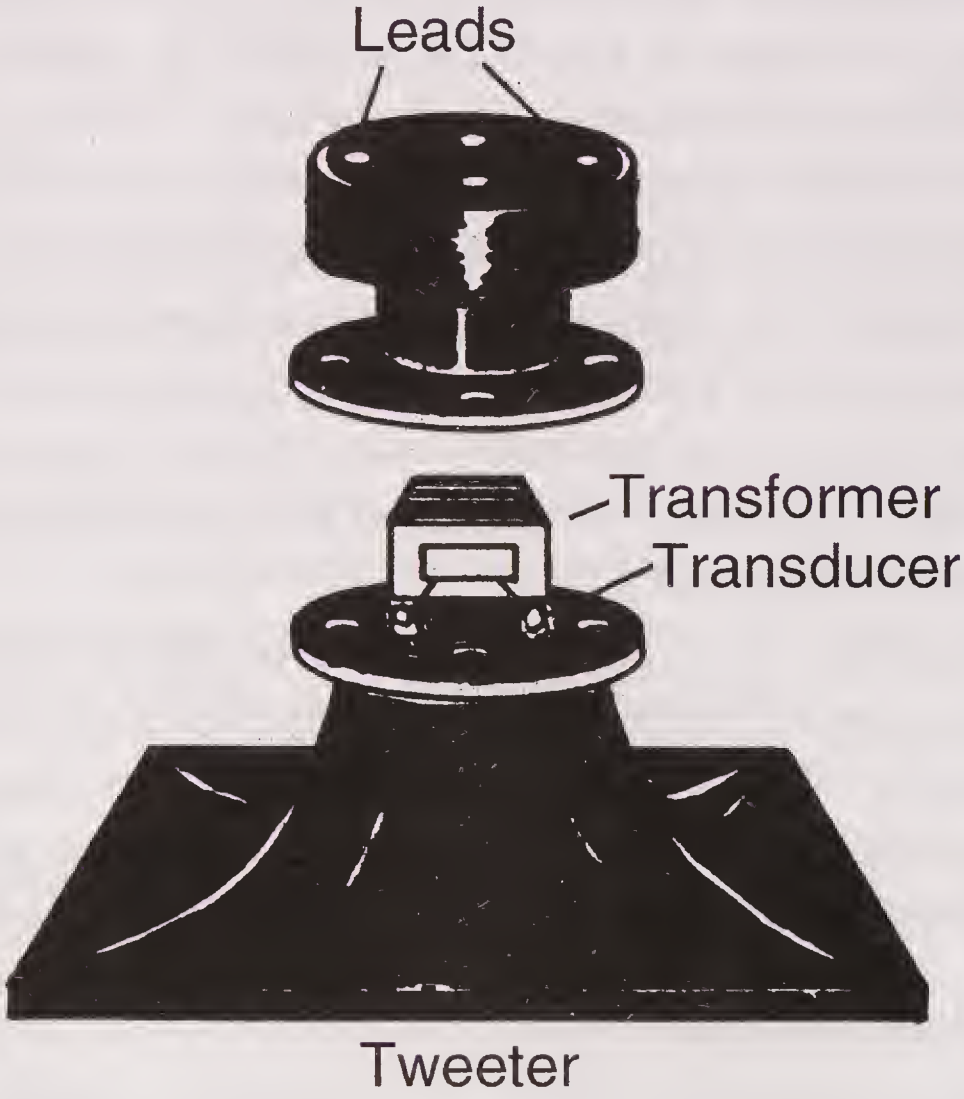

Where they are found

Piezoelectric transducers are found in many applications. Quartz crystals are found in computers, watches, clocks, transmitters, and many other applications. The sound transducers are found in earphones, alarms, audio equipment, etc.

Testing

Quartz crystals are tested using proper circuits. The other transducers (ceramic) are tested using oscillators to drive them.

Semiconductors

Today, any practical application of electronics involves the use of semiconductors. Transistors, diodes, and their relatives form the great semiconductor family found in most any type of modern electronic equipment. With help, the electrician can understand how these components work and why it is important to start with some fundamentals of semiconductors.

Semiconductor Principles

According to modern physics all substances can be placed into one of three classes: insulator, conductor, and semiconductor. The main factor that determines in which of these classes the substance can be placed is the spacing of the energy band in its molecular makeup. Each energy band can hold two electrons at any time.

If the energy bands of a substance are filled, that substance cannot accept or donate electrons and so is an insulator. If there is one electron per band, or if the bands are so closely spaced that no gap exists between a filled band and an empty band, electrons can move through the material and it is a conductor. Now, the intermediate case: If small gaps exist between filled energy bands and vacant ones, the material acts like an insulator at low temperatures and becomes a conductor when the temperature rises. This is a semiconductor material.

Semiconductor types such as the silicon, germanium, gallium, and others have additional properties that make them ideal for use in electronics. They are elements that have four valence electrons each. Because of the valence bonds, they form a basic crystal structure as shown in Figure 80.

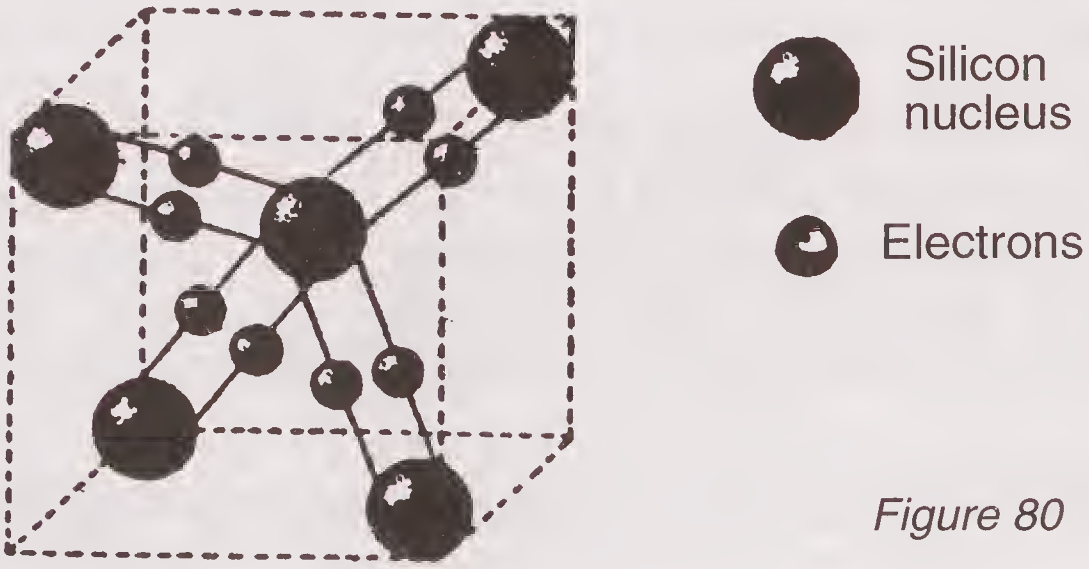

Figure 80

Any crystal is held together by the sharing of electrons between the atoms. In a conducting material the free electrons can drift through the structure under the influence of electric forces. On the other hand, in a semiconductor, availability of a path for free electrons to move along depends upon temperature. As the temperature rises, the path becomes available and high-energy electrons can move.

A crystal of semiconductor like silicon or germanium is composed of billions of atoms bonded together in an arrangement similar to the one shown in Figure 80. Crystals of pure semiconductors can be grown in laboratory conditions. They are called intrinsic materials and have no practical use. But, if we add small amounts of specific impurities to the intrinsic material, these impurities have the capability of joining right into the crystal structure at the atomic level.

There are two kinds of impurities resulting in different effects on the electric properties of the material. If a five-electron element, such as the antimony of boron and phosphorus, is added to the crystal, each of its atoms has one valence electron left over that finds no partner in the crystal structure. The net result is a surplus of electrons. Elements such as this are called donor substances and their presence provides an excess of negative charges. The semiconductor crystal is now called an n-type material.

Now, if we add a three-electron element, like aluminum, gallium, and indium, to the structure of the crystal, the net result is the presence of holes where the surrounding semiconductor atoms have electrons, but the impurity has no matching electrons to fill the valence bond. Elements with three electrons in the valence shell are called acceptors and form p-type materials when added to semiconductors.

Junctions

When crystals are being grown it is possible to add donor materials at one point and acceptors at others, producing single crystals composed of two different types of materials. The surface at which the type changes from one to the other is called a junction. The junction between semiconductor materials is the key of all events that made possible modern solid-state electronics. (As an aside, in early electronics circuits, vacuum tubes were the main elements. Inside these circuits, all events occurred in the vacuums. Today, the main elements of any circuit are semiconductors—solid materials—thus, the name "solid-state" electronics.)

Combining junctions and different semiconductor materials, engineers have created a large number of components that are found in every kind of electronic appliance. Let's take a look at the main types of semiconductor components, how they work, and where they are found.

Diodes

The first important electronic component of the semiconductor family is the diode. To understand how a semiconductor diode functions, look at what happens in a semiconductor junction such as the one described previously.

At a junction the surplus electrons of the n-type material diffuse across the junction filling the holes of the p-type material. This process is called combining and forms valence bonds that can no longer diffuse throughout the crystal. This means that in a small area, a depletion region is formed that is free of holes and electrons.

If, in a semiconductor junction, we apply a positive voltage to the n-type part of the junction, it tends to draw electrons out from the n-type material. At the same time, the electrons in the negative side of the power supply fill the holes in the p-type material. The net result is that the depletion region expands to the entire structure of the material. This situation prevents current flow. The diode is now reverse-biased.

If we reversed the applied voltage, the electrons would be forced away from the power supply toward the junction, and, at the same time, holes from the power supply forced to the junction. The electrical force involved in this process compresses the depletion region until it disappears. At this moment, the barrier is broken and the current can flow through the device. With germanium devices this occurs at about 0.2 V and with silicon devices at about 0.6 V. This way the junction is forward-biased.

The device formed by a single p-n junction is called a diode. Figure 81 shows what happens in the two described bias conditions.

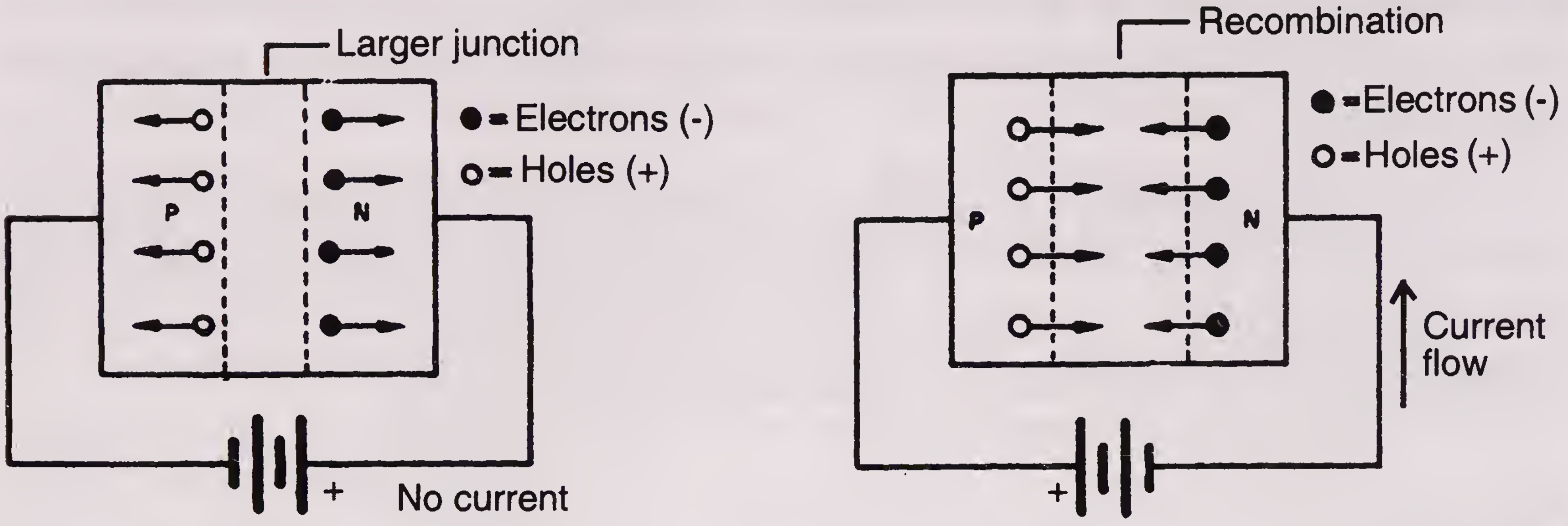
Larger junction
● = Electrons (-)
o = Holes (+)
P
N
+
No current
Recombination
● = Electrons (-)
o = Holes (+)
P
N
Current flow
+

(a) Reverse biased

(b) Direct biased

Figure 81

The practical semiconductor diode is formed by two pieces of semiconductor materials—n- and p-type germanium or silicon—placed inside an enclosure. The size of the materials is basically determined by the amount of current that they can conduct when forward-biased. Semiconductor diodes operate as one-way roads for the current and are found in a large number of electronic circuits.

Symbol and Types

The diodes are manufactured in different sizes and types according to their applications. Figure 82 shows common diodes and the symbol used to represent this component.

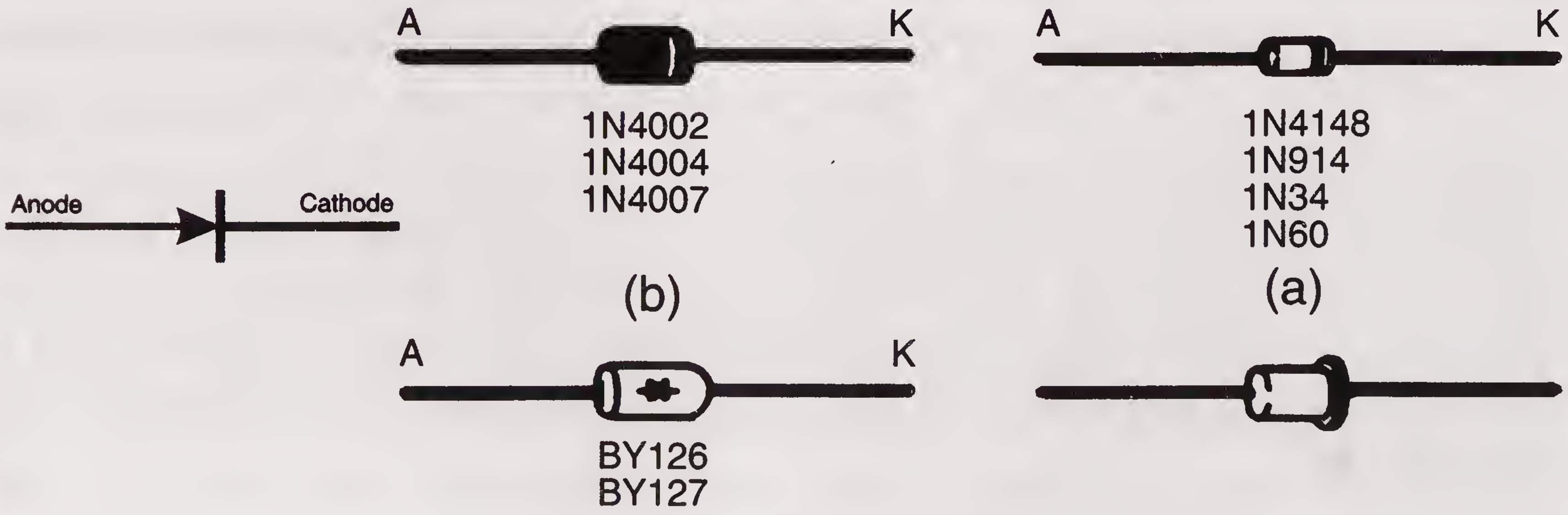
A
K
1N4002
1N4004
1N4007
A
K
1N4148
1N914
1N34
1N60
Anode
Cathode
(b)
(a)
A
K
BY126
BY127

Figure 82

In (a) are germanium and silicon small-signal diodes that are used to handle low currents. In (b) are some rectifier diodes used in circuits where large currents are found.

Diodes are polarized components. Their position in a circuit is important. The identification of the poles is made using several resources. One method is placing the symbol and the component to correspond with the position of the poles and other is to use a strip or ring at the side of the cathode terminal.

Specifications

A. Reverse voltage: Since a diode represents an open circuit when reverse-biased, across the diode appears all the voltage of a circuit. The maximum voltage that a diode can be submitted to without being burned under these conditions is an important parameter to be considered in projects. This voltage is also abbreviated by Vrrm or Vr in data sheets.

B. Direct current: This is the largest current that can flow through the diode when it is forward-biased. It is also abbreviated by If.

C. Part number: The most common in real components is the manufacturer specification by a part number. In most parts of America, specifications of diodes are identified by the group “1N” followed by a number. So, types such as 1N914, 1N4148, and 1N4002 are common. In European specifications diodes begin with an “A” (for germanium) or “B” (for silicon) followed by a letter. For example, “A” for general-purpose diodes, “Y” for rectifiers, etc. A European part number might look like these: AA115, BA315, BY127, etc. Other manufacturers use proper codes to identify their diodes. Some examples are MR751 and P600D.

Where they are found

The diodes are used basically in functions such as detectors or signal clippers in audio and RF circuits or as rectifiers in power supplies.

As detectors, the electrician will find small diodes (silicon or germanium) in radio receivers, TV, remote-control receivers, computers, interfaces, wireless telephones, etc. As signal clippers, the diodes are used to convert an AC into a DC as shown by the diagram in Figure 83.

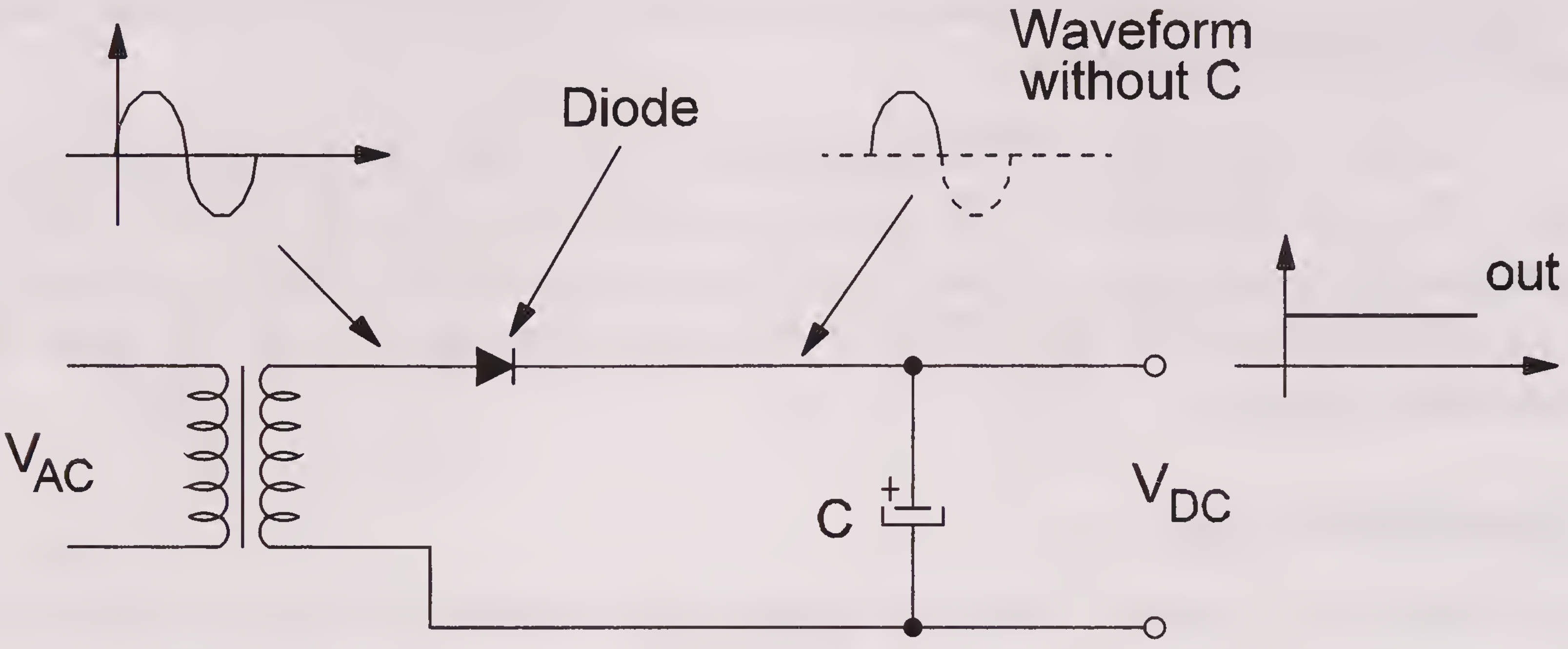

Figure 83

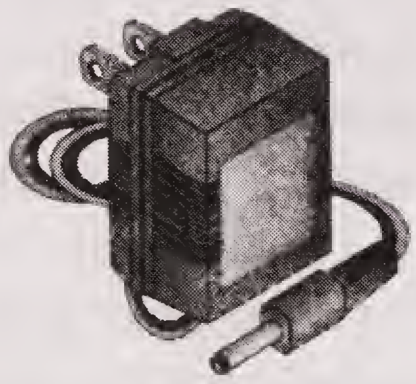

Figure 84

In this circuit, the diode conducts only the positive half-cycles of the AC voltage found in the secondary winding of the transformer. Filtering these positive pulses with a capacitor, a DC voltage can be placed at the output. Power supplies such as this are used to power small DC appliances like calculators, CD players, and transistor radios from the AC power line. Figure 84 shows an AC/DC adapter using an inside circuit like this.

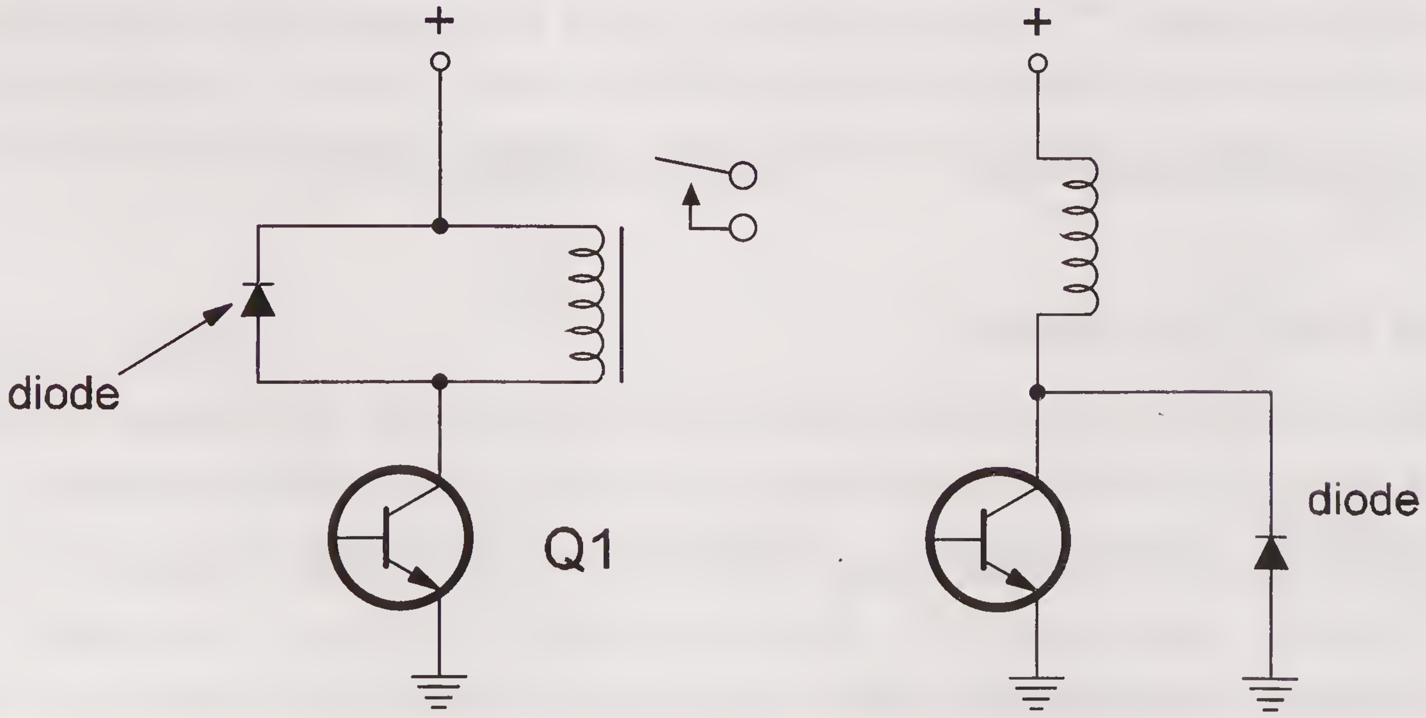

Figure 85

A third case to be included here, and one that is very important to electricians, is the use of a diode to protect semiconductors from voltage spikes generated when inductive loads are triggered. When a relay is open or a solenoid is turned off, it generates a voltage spike strong enough to damage the device used in the control circuit. The energy produced in this process can be absorbed by a diode wired in parallel with the load or with the control device as shown in Figure 85.

Testing

A diode can be tested using a common multimeter or any continuity tester. Adjust the multimeter to a medium resistance scale and place the probes in the diode's terminals. When biased forward the diode must present a low resistance (between 100 and 5000 ohms) and when reverse-biased a high resistance is read (above 1,000,000 ohms). A diode presenting low resistance when reversed and forward-biased is shorted, and the one that presents a high resistance of the two measurements is open. An intermediate read when reverse-biased (between 100,000 and 500,000 ohms) indicates a diode presenting losses.

Zener Diodes

When a diode is reverse-biased a limit value to the applied voltage exists. Beyond this limit, called breakdown voltage, the diode becomes a conductor and is destroyed as shown by the graphics in Figure 86. As shown by the curve, the diode, if not destroyed, tends to keep the voltage constant across its terminal in a range of currents.

Zener diodes are special diodes that are processed to have a nondestructive breakdown at a specified value of the reverse voltage. They are also called avalanche breakdown diodes or breakdown diodes.

The action of a zener diode can be used in many electronic applications. In fact, if you apply a reverse voltage to a zener diode its reverse resistance changes in a way to maintain constant voltage across it. This action can be used in voltage regulation, to clip signals, and many other applications.

Zener diodes are found in many types and sizes in electronic equipment. For the electrician, it is important to know how a common diode can

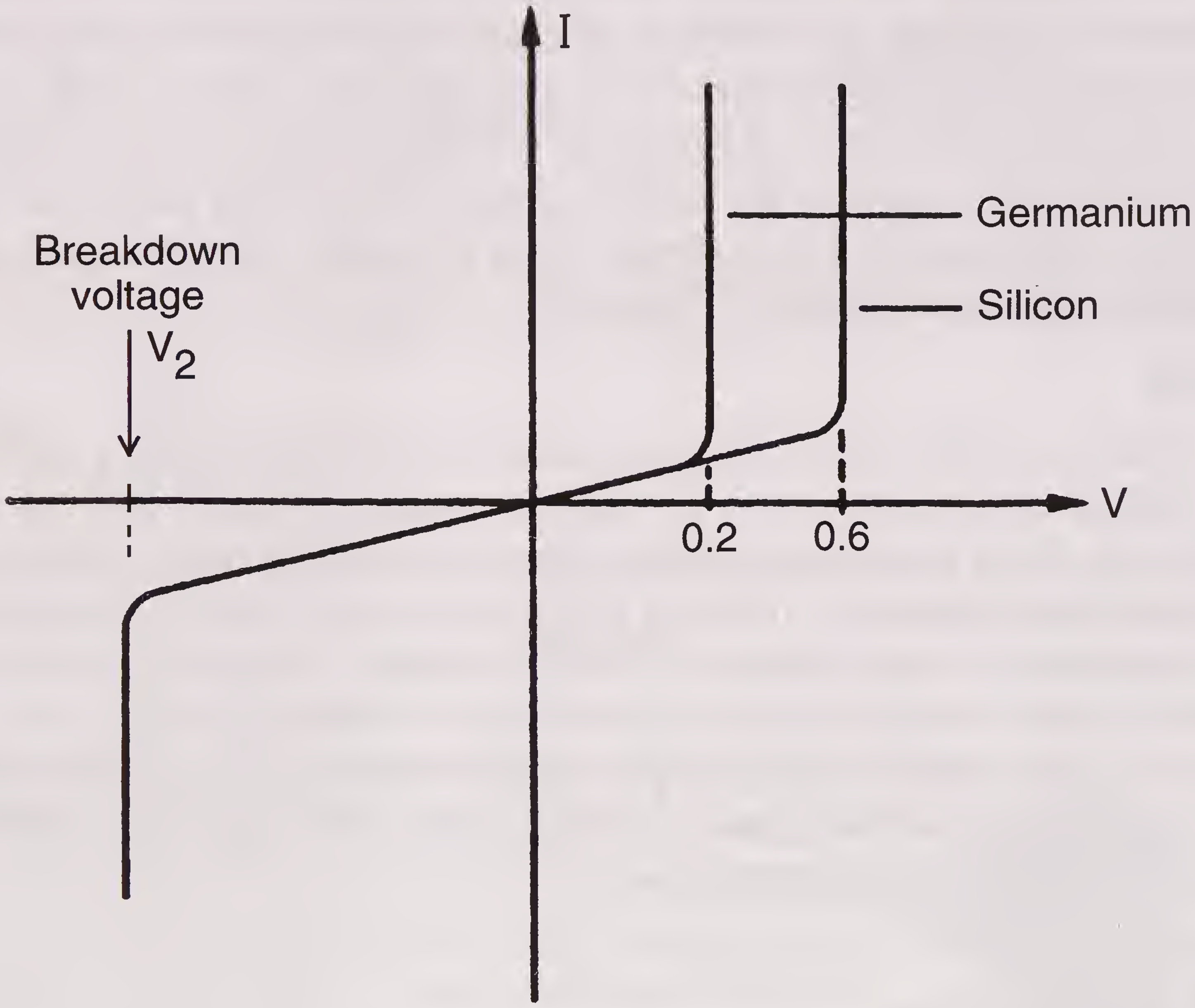

Figure 86

differentiate from a zener diode. Although they look the same, the way they are positioned in a circuit, the identification number, the symbol in the schematics, or any other specific information may differ.

Symbol and View

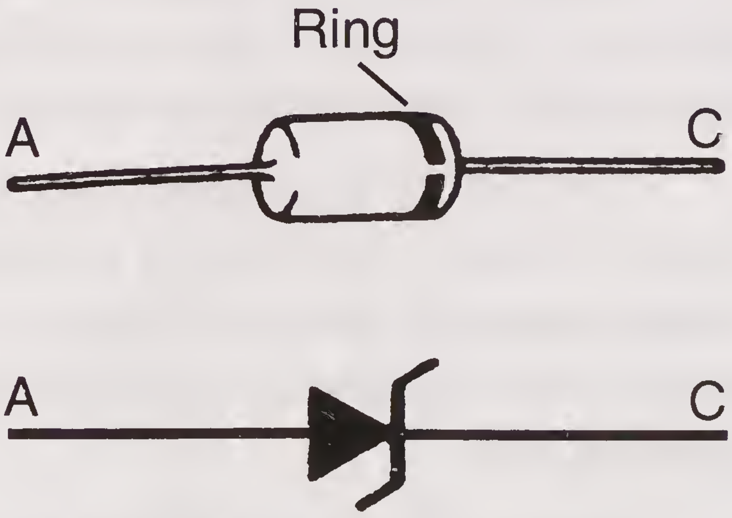

Figure 87

Figure 87 shows the symbol and a view of a common zener diode as the electrician will find them in electronic circuits.

Since zener diodes are polarized components, there must exist some identification for the poles in the component. It is usual to place a ring or strip at the cathode side of the component.

The size of the zener diode is determined by the amount of current it can control or the range of currents in which the diode can keep the voltage constant.

Specifications

The zener diode manufacturers used to indicate their products by a part number. So, a handbook with the main types or another kind listing of their characteristics is fundamental in order to know what they are. When replacing a zener diode without having the part number it is important to know two specifications:

1. Zener Voltage: the breakdown voltage or the voltage the device is designed to keep constant when reverse-biased. Zener diodes are made with voltages in the range between 1 and 2 volts to more than 100 volts for typical types found in domestic and industrial appliances.
2. Dissipation Power: the maximum amount of current flowing across a zener diode during operation. The zener voltage gives the dissipation power or the amount of electric power that the device can convert in heat. The larger the size of the device, the larger the dissipation power. Typical values are between 400 mW and 10 W.

Where they are found

Figure 88 shows some circuits where zener diodes are found. In (a) is a simple voltage divider in a simple power supply where the zener diode maintains a constant voltage in the load plugged into the output. A more

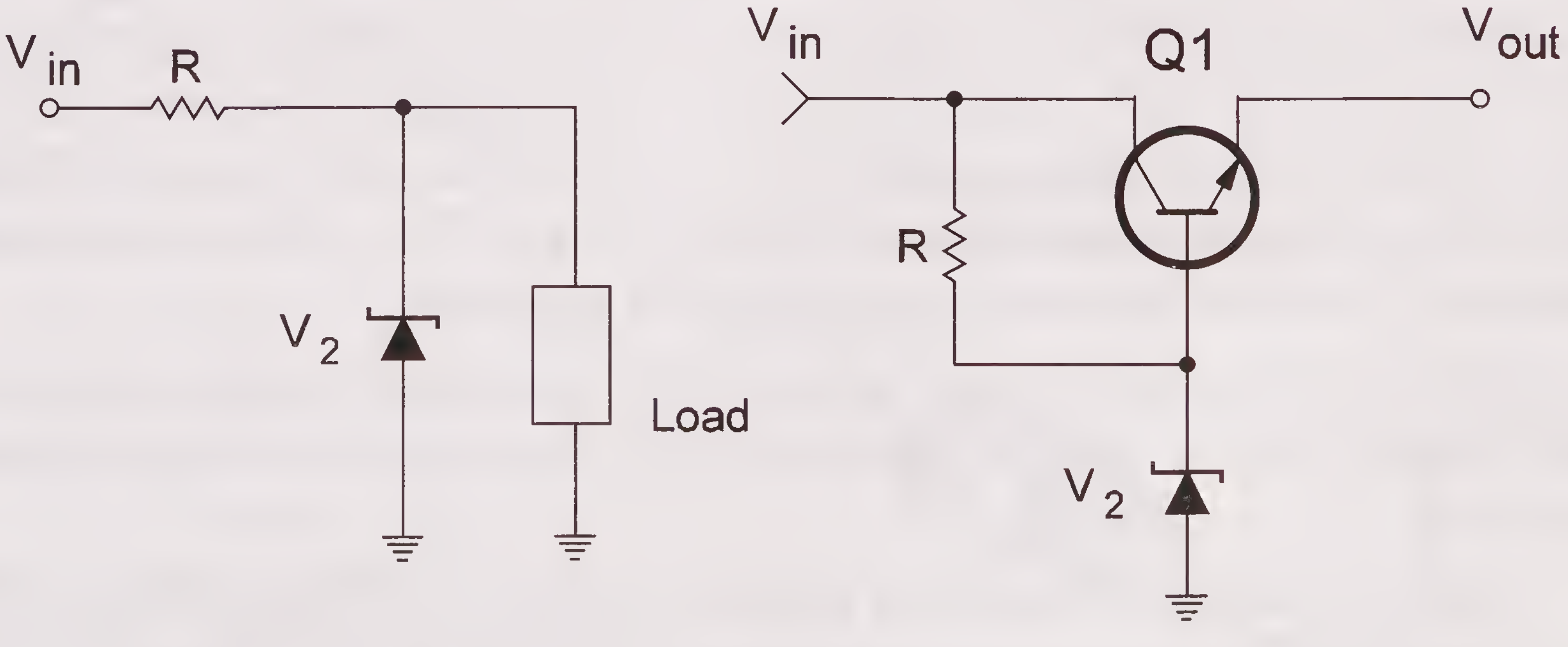

(a) (b)

Figure 88

complex power supply is shown in (b) where the transistor controls the amount of voltage applied to the load. This voltage is kept constant by the action of the zener diode placed at its base. How the transistor works in a circuit like this is discussed when explaining how transistors work.

Other applications of zener diodes include voltage reference voltage in instruments and signal clipping circuits.

Testing

The simplest way to test a zener diode uses a common multimeter on the lowest resistance scale. The zener diode is tested as a common diode; when forward-biased it must present a low resistance (100 to 10,000 ohms), and when reverse-biased it must present a high resistance (above 1,000,000 ohms). This test doesn't show anything about whether the zener voltage is altered or not. It is important to remember that this test is valid only if the zener voltage of the tested device is lower than the voltage used by the multimeter in the probe.

LEDs

Light Emitting Diodes or LEDs are special diodes that produce light when forward-biased. The current flowing to their PN junction causes the emission of visible light or infrared radiation. Researchers discovered that any semiconductor diode produces infrared light when conducting a current, the beginning of the LED's principle of operation. Advancing with the research, they found that special semiconductor materials, such as gallium when doped with special impurities like those of arsenium and indium, could result in diodes that emitted visible light when forward-biased. From this appeared a new class of diodes—the LEDs—that were primarily designed to act as light sources or infrared sources.

LEDs are monochromatic light sources because the light produced in their junction has a narrow spectral band as shown by the graphics in Figure 89.

LEDs are made of materials such as Gallium Arsenide (GaAs) doped with impurities like indium or phosphorus, in a degree that gives them their light emission characteristics or other spectral characteristics. The impuri-

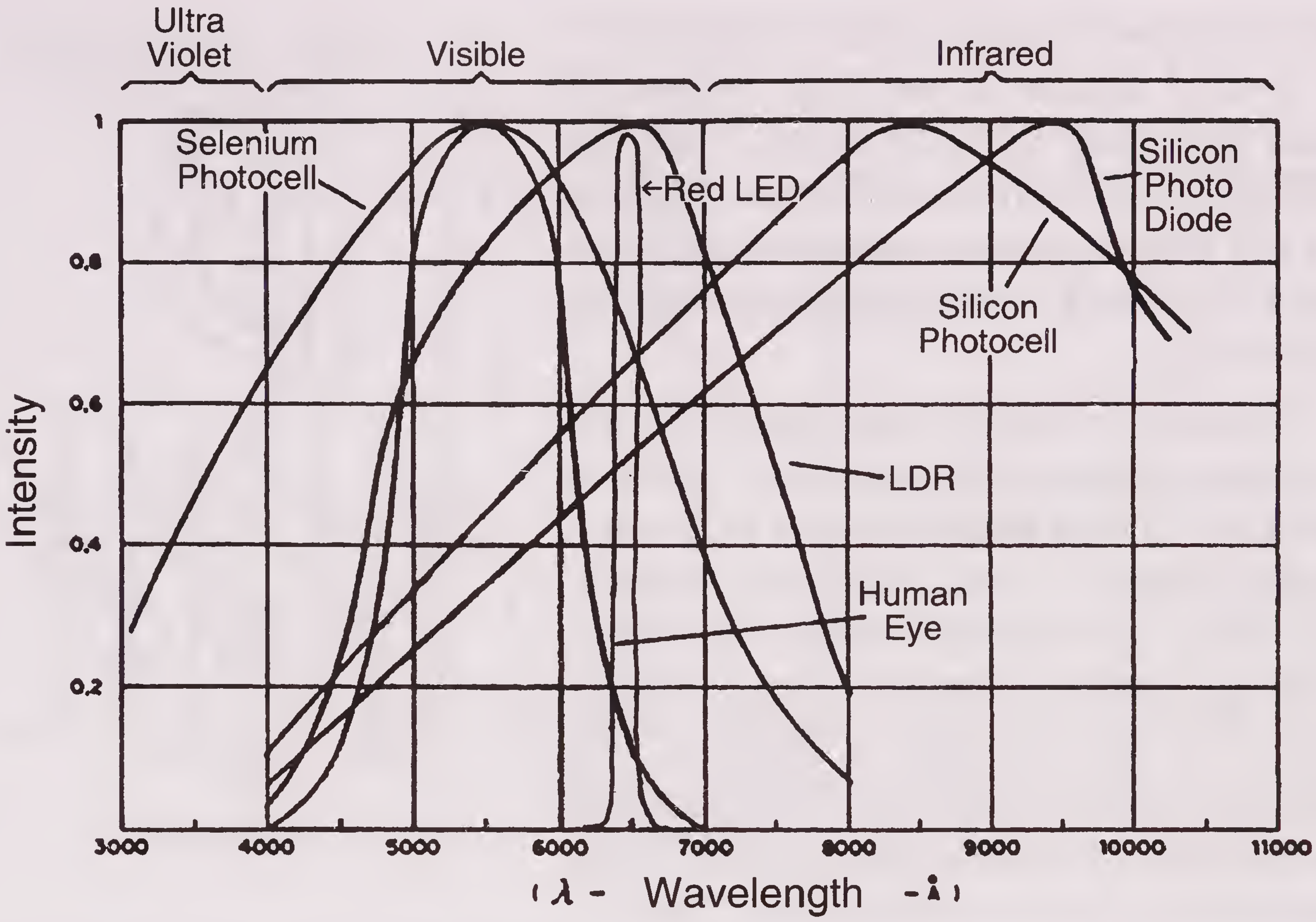

Figure 89

ties also determine the minimum voltage for forward-biasing an LED that makes it conductive. Red LEDs become conductive with 1.6 V; yellow LEDs need about 1.8 V, and blue and green LEDs need more than 2.0 volts.

Symbol and View

Figure 90 shows the symbol used to represent LEDs in schematic diagrams and in the same figure is the main type of LED.

The different types and sizes depend on the color of the light and also the application. IR (Infrared) LEDs are used in remote controls and as sensors. LEDs can also be placed to form an eight in a

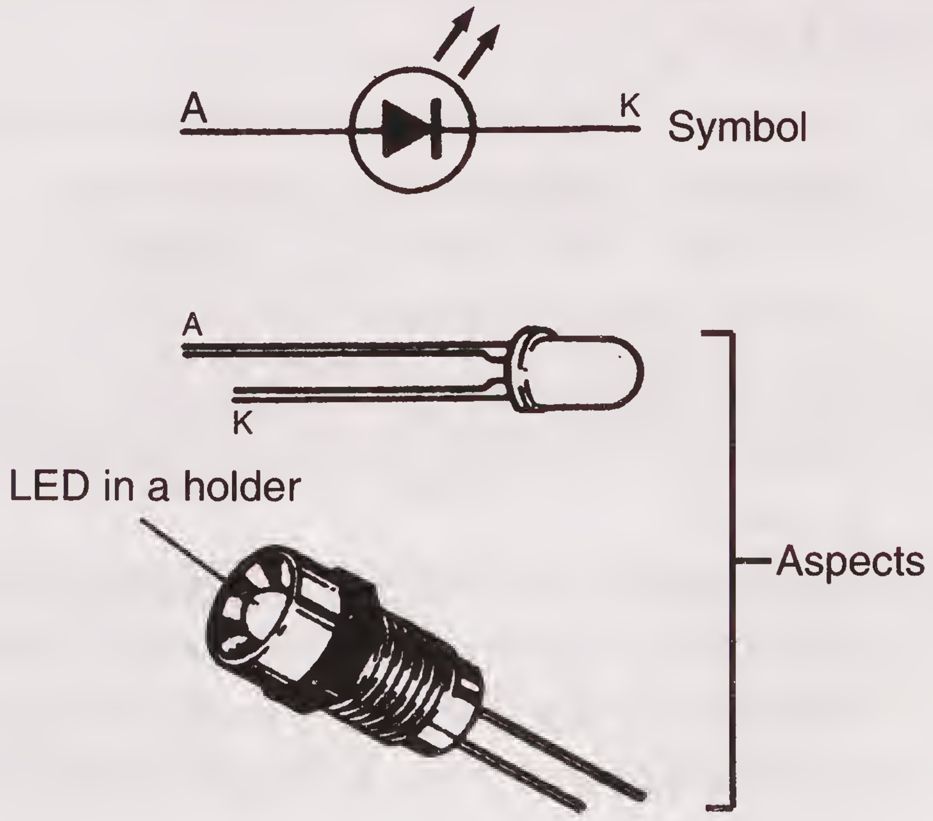

Figure 90

seven-segment display. Each segment in this kind of display is an LED. Figure 91 shows how they are connected. If all the LEDs are wired with the cathodes together they are called common-cathode displays. Figure 91 shows a seven-segment display using LEDs.

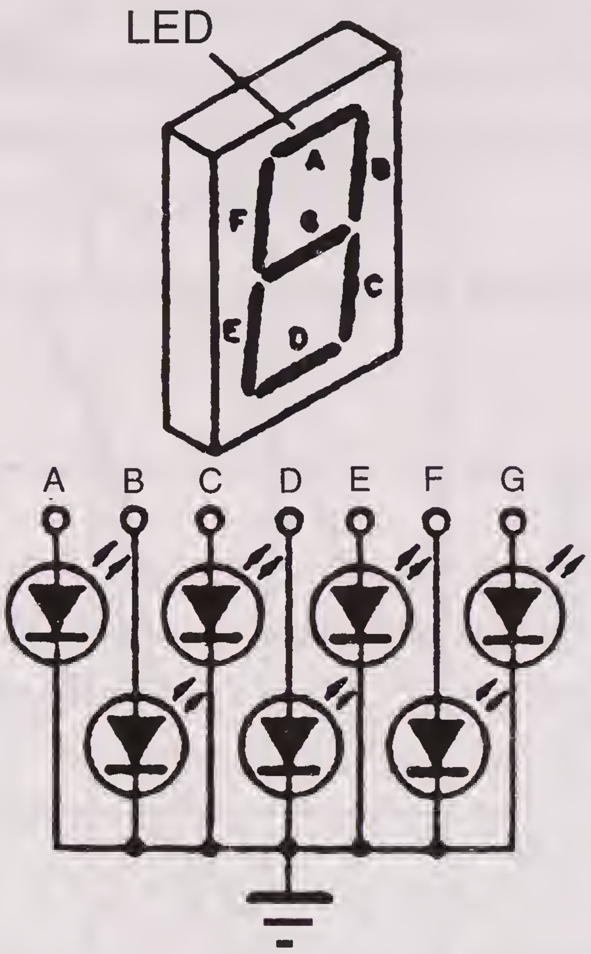

Figure 91

It is also possible to make a white LED by putting three LEDs on a chip with the basic colors (Red-Blue-Green or RGB) and powering them in a way that the combination of their light results in white. According to the amount of current in each of these LEDs, it is also possible to produce any color of the visible spectrum. The LEDs, like the diodes, are polarized components. Normally they are encapsulated in clear or colored plastic with a small, flat space or notch near the cathode lead.

Specifications

LEDs are specified by a part number provided by the manufacturer. But, catalogs also indicate other characteristics including the size, forward-bias voltage, maximum current, and the wavelength or color of the produced light.

A. Current: The amount of current flowing through an LED determines how much light or IR it can produce. Typical values are between 5 and 100 mA. This current is important to help the designer calculate the current limiting resistor (see how to use an LED).

B. The forward voltage: This is the voltage at which the LED becomes a conductor. It is the minimum voltage that can be applied to the LED in a circuit.

C. Wavelength of the light: The wavelength can be expressed in angstroms (Å) or nanometers (nm). 1000 A = 100 nm. The visible spectrum extends from about 4000 to 7000 angstroms or 400 to 700 nm. Figure 92 shows the placement of the emission curve of some common LEDs according to their colors.

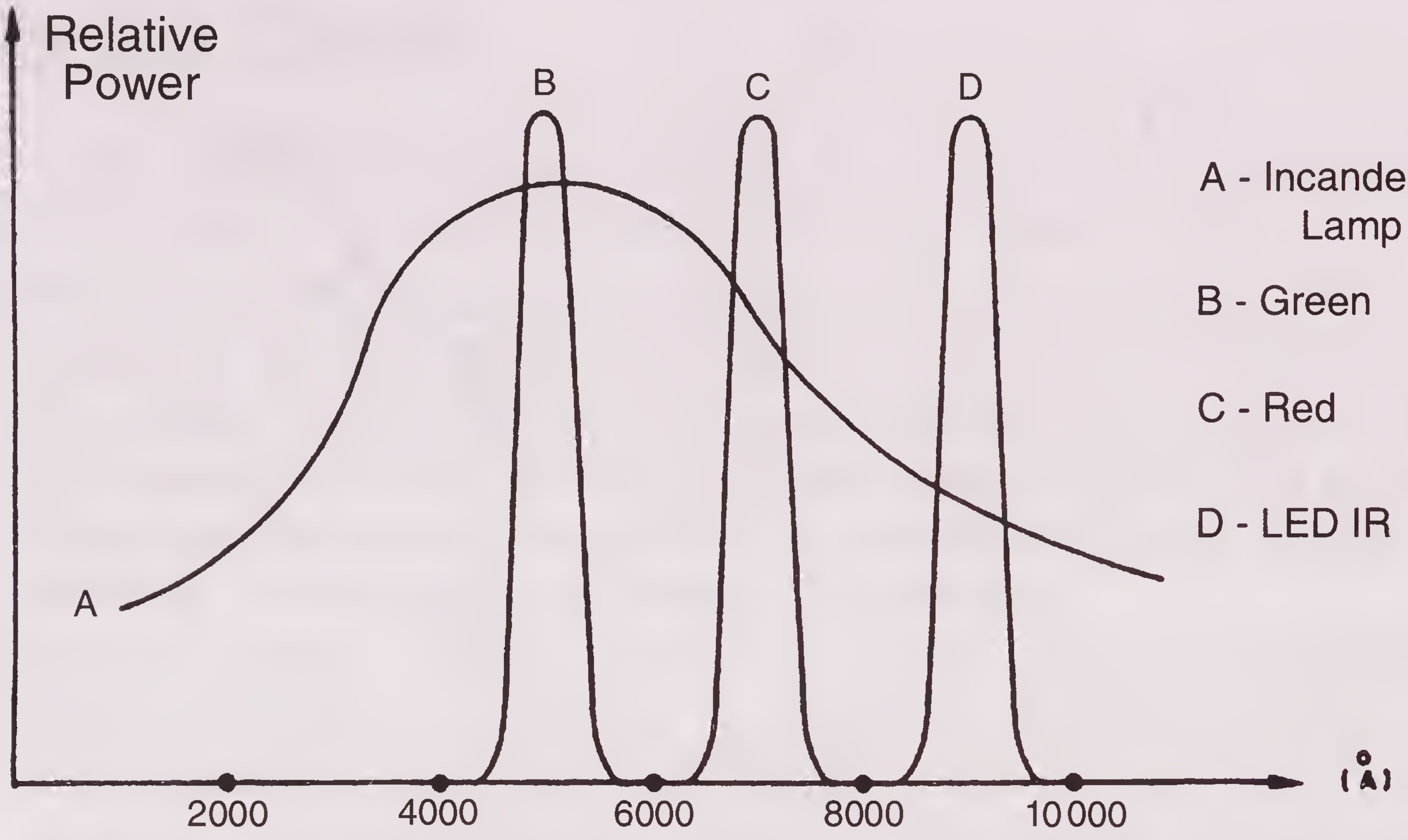

Figure 92

Where they are found and used

The main applications of LEDs in electronic equipment are as indicators. The electrician will find LEDs in panels of electronic, and even electric, devices indicating a function or displaying a number or even a figure as in the seven-segment types. The seven-segment displays also indicate the result of operations in calculators, the hours and minutes in clocks, time, and other information in microwave ovens, etc.

Infrared LEDs are found in remote controls of a variety of equipment like TVs, VCRs, automatic doors, car alarms, etc.

When using LEDs it is important not to exceed the maximum allowable forward current. Current is best limited by a resistor placed in series with the LED and the power supply as shown in Figure 93.

The resistances of the series resistor can be calculated by subtracting the LED's forward voltage from the power supply voltage and dividing the result by the desired current in amperes. For example, if we want to power a red LED at 20 mA from a 6 V power supply the math is as follows:

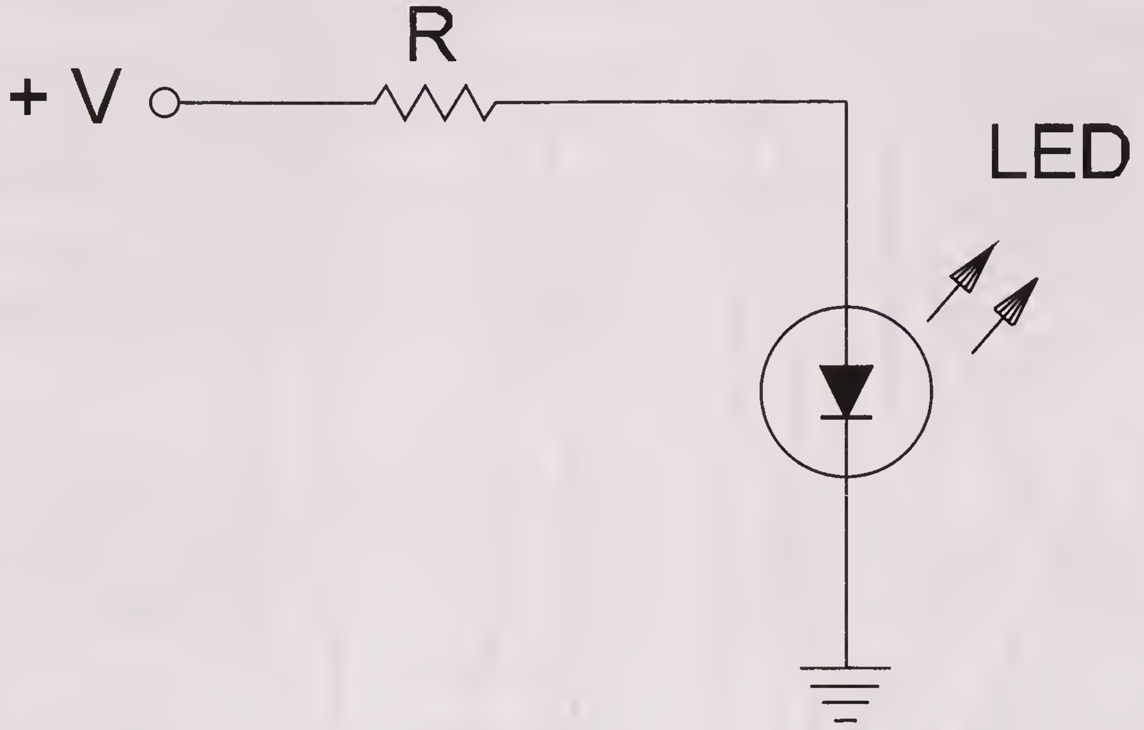

Figure 93

Subtracting 1.6 V (the forward voltage of a red LED) from 6 V results in 4.4 V. Dividing the result (4.4 V) by the current 20 mA (0.02 A) results in 220 ohms.

Testing

Never test an LED by plugging it into a power supply without a series current-limiting resistor. It can be burned. The best way to test is by plugging it into a 3 V supply (2 AA cells) in series with a 100 ohms to 270 ohms resistor as shown in Figure 94. The LED will glow if it is good. To check whether an infrared LED is producing radiation, some kind of detector must be used.

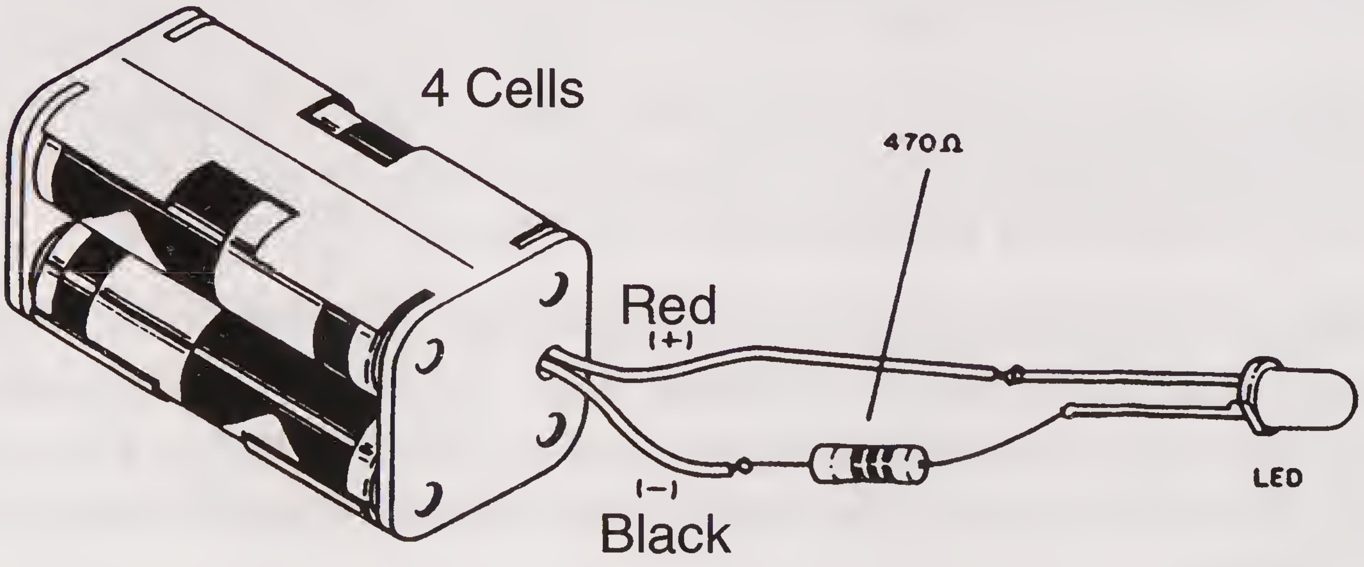

Figure 94

Special Diodes

The electric properties of a PN junction can be used in devices other than the LEDs. Many special diodes can be included in a list:

A. Photodiodes: Exposing the PN junction of a diode to light and reverse-biasing it, the current flow will depend on the amount of light falling on the junction region. Charge carriers are liberated by the light, increasing the reverse current. The semiconductor material with a PN junction is then manufactured in transparent enclosures and used as light or IR sensors. These devices are called photodiodes.

B. Tunnel diodes: Tunnel diodes are special devices presenting the characteristic of a curve in a negative resistance region. This characteristic makes the device ideal to produce very high-frequency signals. The tunnel diode is used as UHF (Ultrahigh Frequency) and SHF (Superhigh Frequency) oscillators and detectors in many applications like radar detectors.

C. Varicaps or varactors: When reverse-biased, a diode can be compared with a capacitor. The junction region is the dielectric and the PN materials with the plates as shown in Figure 95.

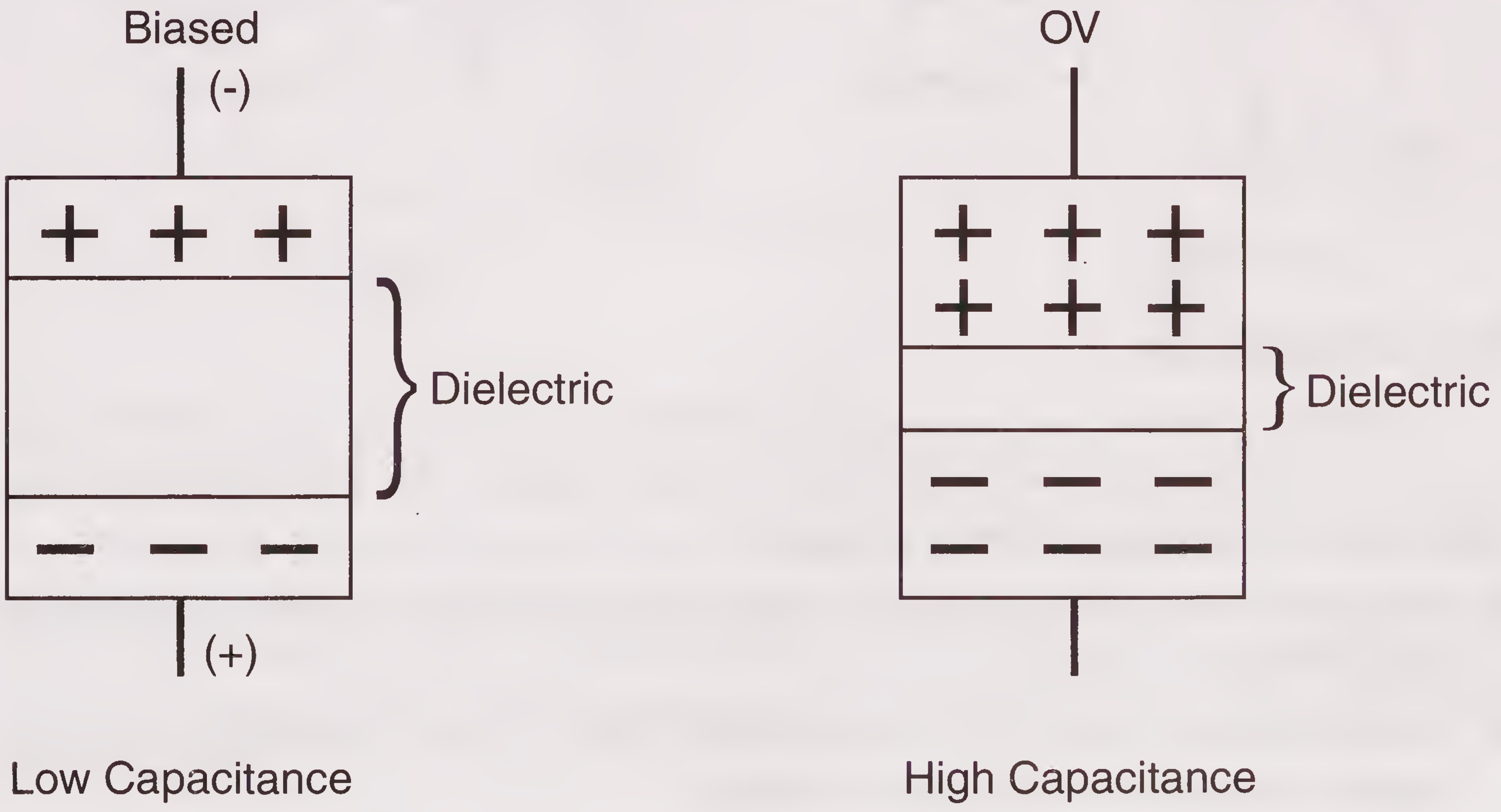

Figure 95

Depending on the applied voltage, the size of the junction region can be altered. If a low voltage is applied, the charges keep the junction region thinner. But, by increasing the voltage, the charges make the junction region fatter. This means that by applying a voltage in a reverse-biased diode we can control its capacitance.

Special diodes with large PN junctions (to get more capacitance) are fabricated to be used as variable capacitors controlled by voltages. They are called variable capacitance diodes or Varicaps. These capacitors are found in tuning circuits of TVs and radios.

Symbols and Types

Figure 96 shows the symbols and types of some special diodes. Remember that these diodes are also polarized components and some mark is used to the identification of their poles.

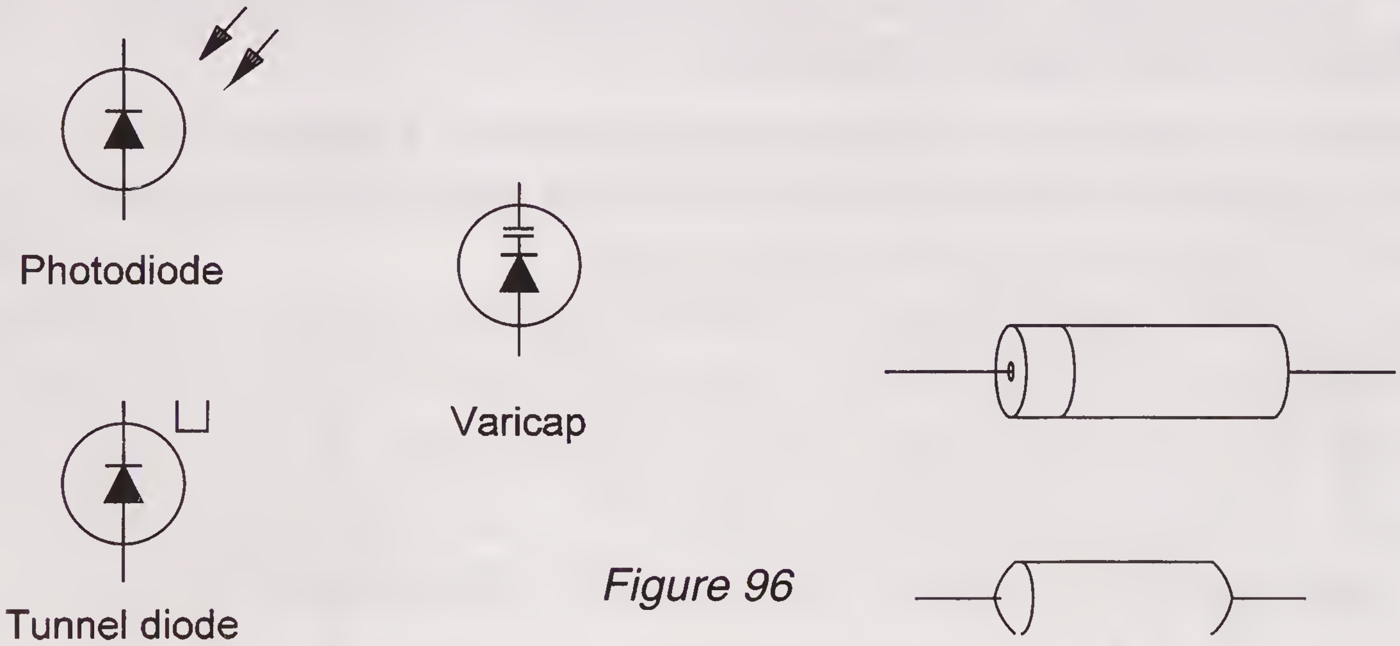

Figure 96

Specifications

Special diodes are generally indicated by a part number determined by the manufacturer. Depending on the application, it is important to know some specifications of these diodes.

- Photodiodes have, as a specification, the response curve (the wavelength of light they can "see").
- Tunnel diodes have, as a specification, the highest frequency they can operate and also the tunnel voltage.
- Varicaps can be specified by the capacitance range they can present between zero and a nominal voltage.

Testing

The test of many special diodes cannot be done using simple procedures. We can detect if a photodiode, tunnel diode, or varicap diode is shorted or open, but not determine any other important characteristics of these components, such as whether they are operating, for example.

Bipolar Transistors

Bipolar transistors are the most important of the semiconductor devices. They are formed by three pieces of semiconductor materials, forming two junctions in the same crystal with a structure such as the one shown in Figure 97.

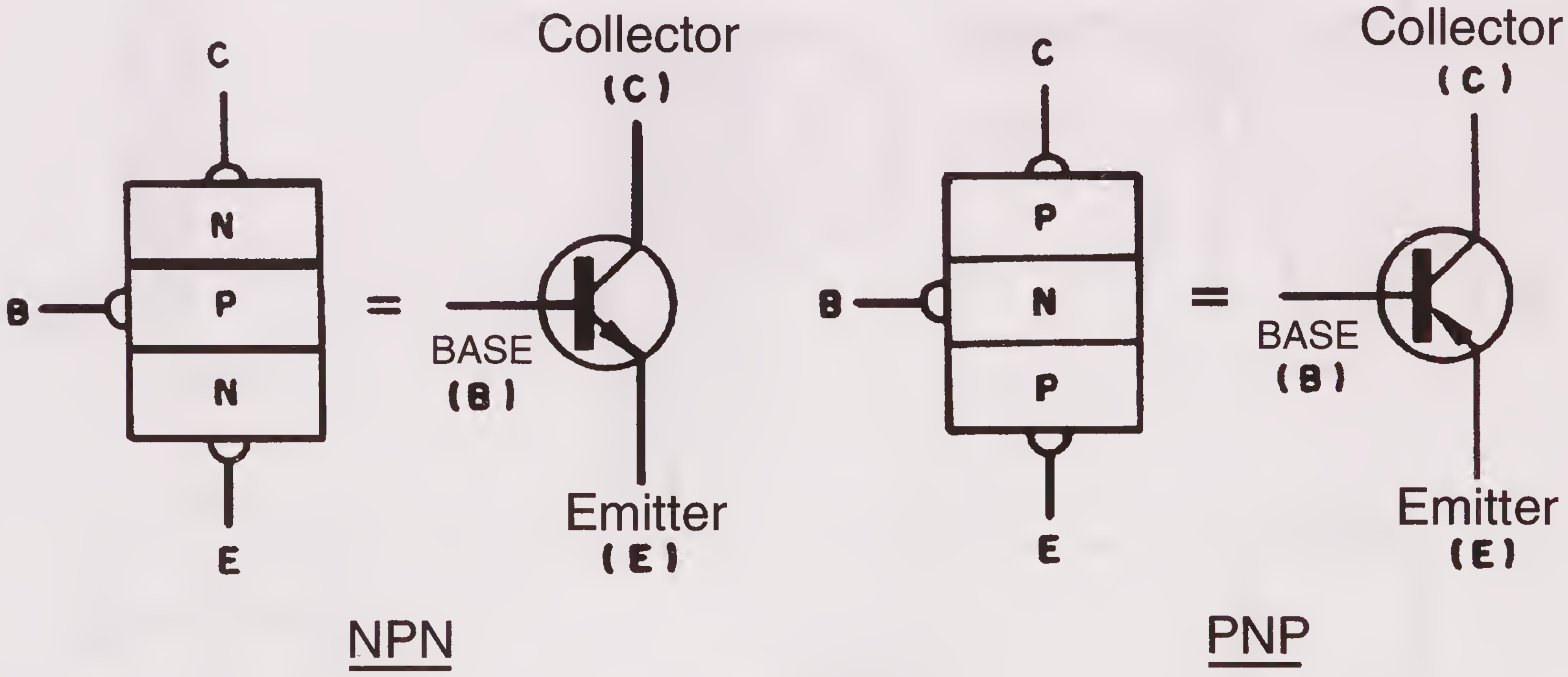

Figure 97

Although the structure can be compared with two diodes placed back-to-back, the fact that they are in the same crystal means the events occurring at one junction can influence those at the other. The possibility of control makes the transistor one of the most important of all the electronic components.

According to the type of semiconductor, the structures result in two types of transistors—NPN and PNP.

To each piece of semiconductor material, terminals or leads are placed corresponding to the three regions: base (B), emitter (E), and collector (C). The transistor forms a structure with special electric properties; a current flowing between emitter and collector can be controlled by a current flowing through the base. When connecting the positive pole of a battery to the emitter and the negative to the collector of an NPN transistor, because the two diodes are reverse-biased, no current can flow.

But, if the base-emitter junction is now forward-biased in a manner that a small current can flow, this current will cause a high current to flow between collector and emitter as shown in Figure 98.

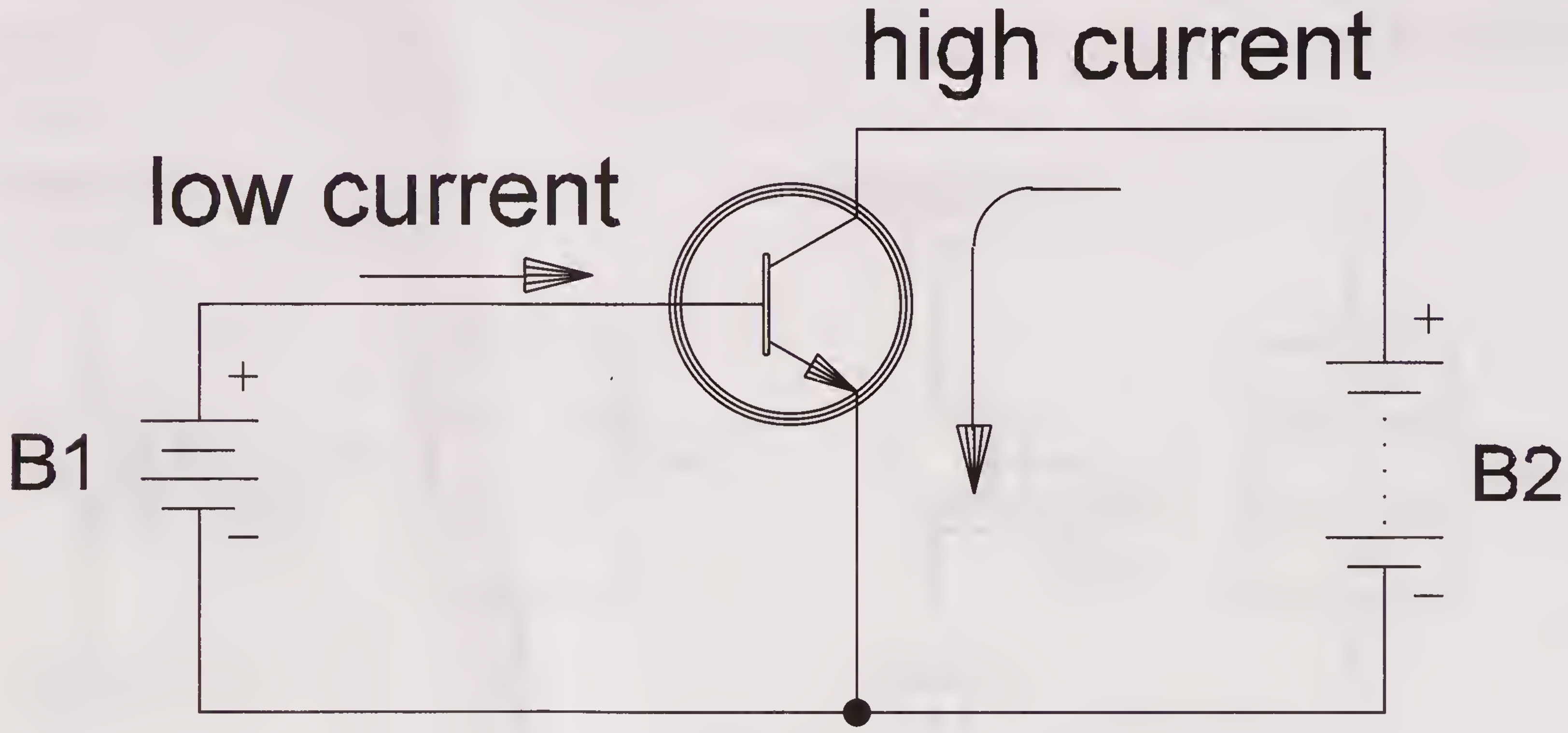

Figure 98

In other words, a small current flowing between the base and the emitter of a transistor can control a large current between collector and emitter. If a 1 mA current through the base causes a collector current of 100 mA, the transistor has a "gain" or "amplification factor" (also called Beta) of 100. This means that the transistor can be used to amplify signals or to act as an electric switch.

Based on the applications, we can find transistors of different sizes and formats with gains ranging from 5 to 10,000. Today there are millions of types of transistors identified by a manufacturer part number. Only with heavy, large “databooks” or “transistor manuals,” could the characteristics of all the main transistors be found.

In applications involving transistors, the use of about 100 or 200 main types with characteristics that are near most of other existing transistors is normal. These transistors can be found easily at a dealer, and in emergency cases, can be replaced with any other.

Symbol and Type

Figure 99 shows the symbols used to represent the two types of bipolar transistors and the common types of these devices. Notice the arrow in the emitter points outside in the NPN type and inside in the PNP type.

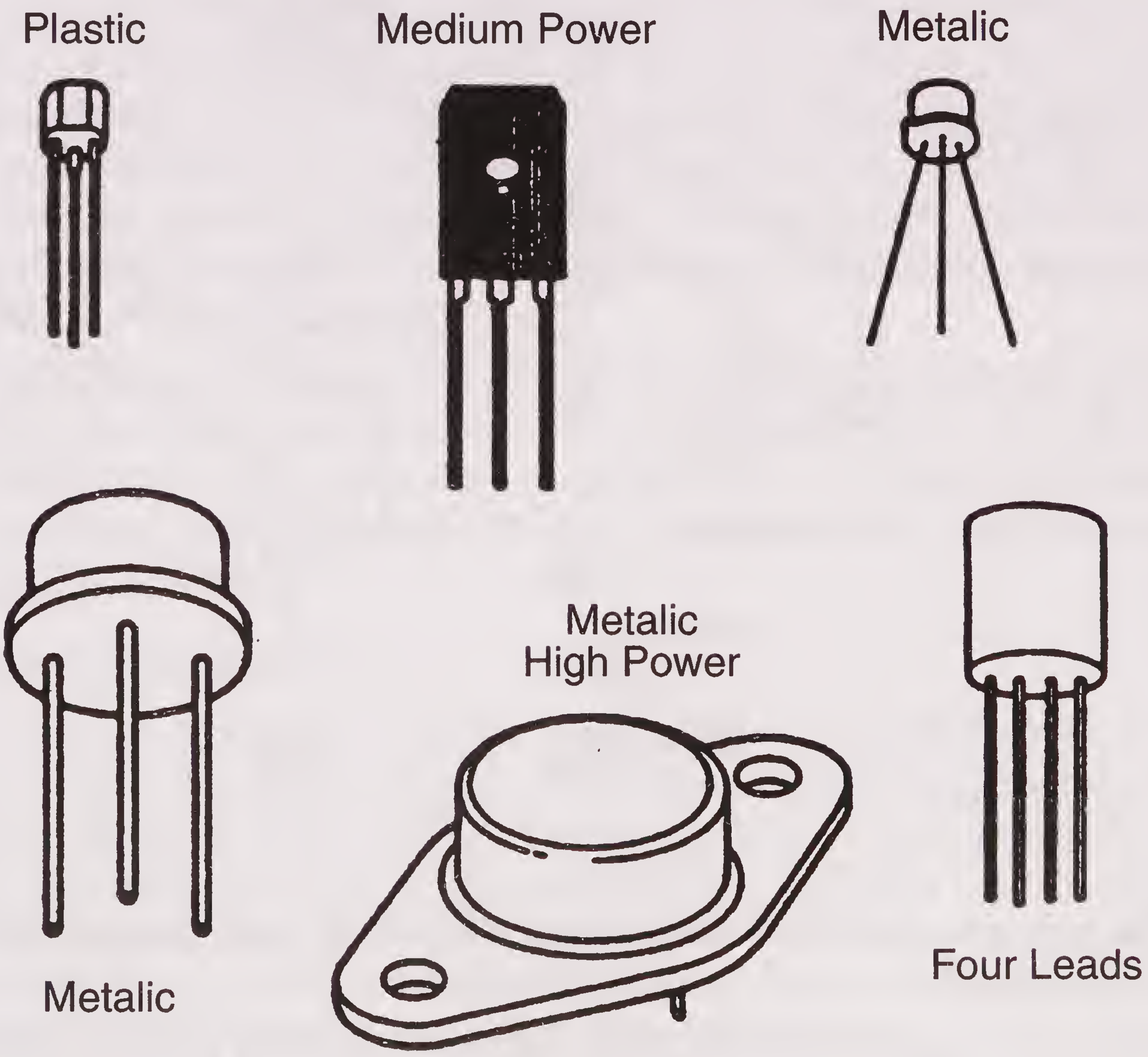

Figure 99

The size of a transistor is determined by the amount of current it can control. Transistors called general-purpose transistors or low-power transistors are small and encapsulated in plastic as shown in (a). Medium-power transistors are shown in (b). These transistors have a hole or another way to place them in heat sinks. In (e) is a high-power transistor that is mounted in a large heat sink. In (f) is a four-lead transistor. The fourth lead is the case acting as screen.

Figure 100 shows a high-power transistor like the one shown in (c), mounted on a heat sink.

Figure 100

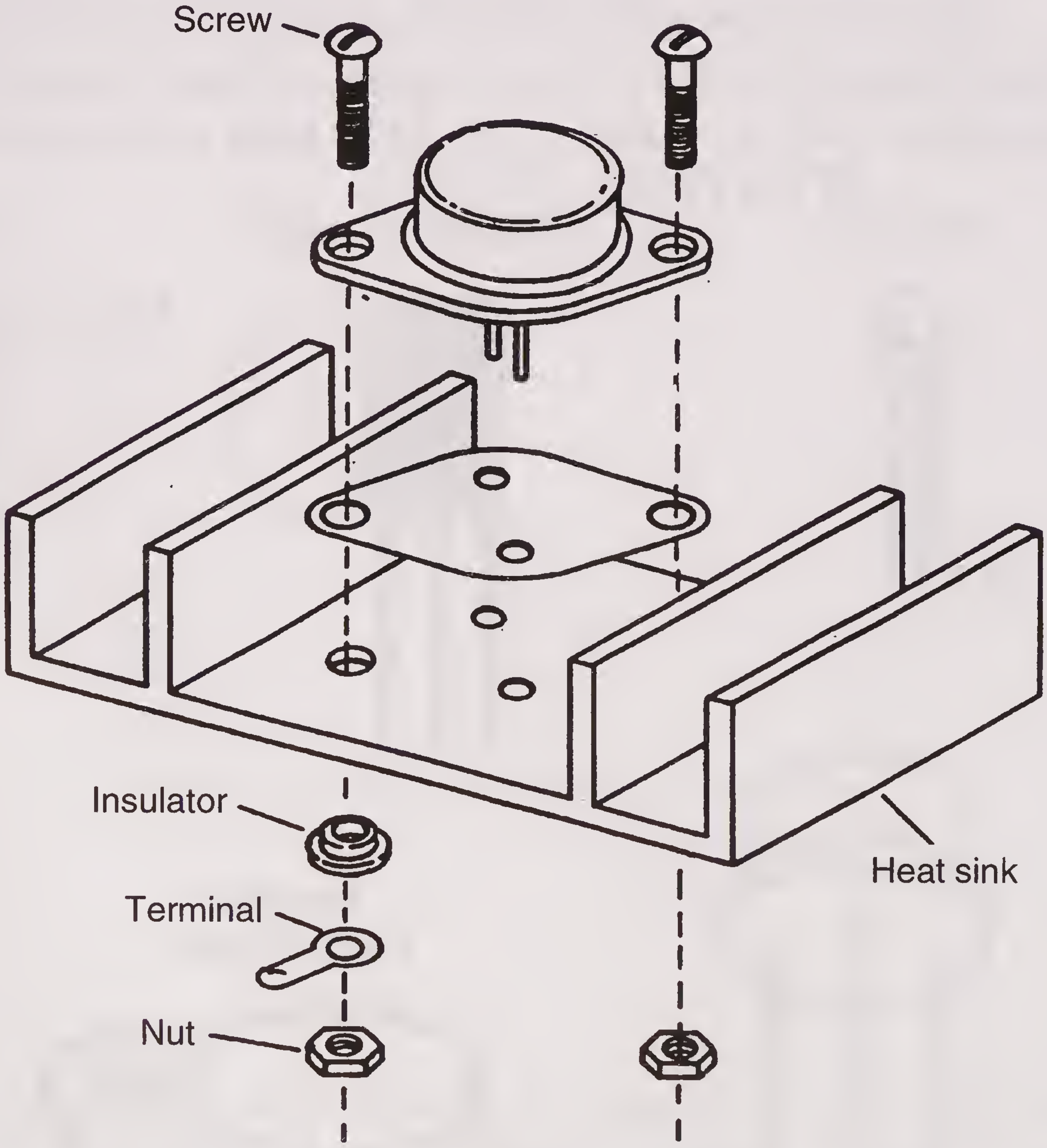

These transistors produce a large amount of heat when operating with large currents.

The main problem the electrician has when working with transistors is the identification of the terminals or leads (emitter, collector, and base). Many special codes and marks are used for this task such as flat spaces, shafts, marks, color points, and others. With the specific information about one transistor, it is possible to know which terminal does what. Otherwise, using the multimeter is also a possibility to make identification by measuring the resistance between the terminals.

Specifications

Besides the size and type (NPN or PNP), the bipolar transistor can also be classified according to the application. The basic groups according to use are:

A. Low-frequency or audio: Low-frequency or audio transistors are components intended for use in applications where the signals are in the audio band or below some hundreds of kilohertz as in audio amplifiers, power supplies, etc. They can handle low-frequency signals such as the ones found in small amplifiers and the audio stage of radios.

B. High-frequency or RF (Radio Frequency): High-frequency transistors are devices specially intended to operate with signals ranging from some megahertz to gigahertz. They are found in radios, TVs, remote controls, wireless telephones, etc., in the stages where the high-frequency signals are present. They also can be found in sizes according the power to be handled.

C. Switching: Transistors that perform this action are used as switches in high-speed circuits. They are used to turn devices on and off in many applications. They can also appear in low-, medium-, or high-power versions. They are found in inverters, logic circuits, digital instruments, digital controls, etc.

Other specifications

The manufacturers indicate the transistor by a part number. With the part number, an electrician can find all the electric characteristics of the device needed to know what it does in a circuit. This is especially important when replacing a transistor because it is possible to find an equivalent or replacement type. The equivalent is a transistor with another part number, either from the same or another manufacturer, but with all the characteristics as the original to replace it in an application.

It is very important to remember that a transistor that is equivalent to another one in application can't be equivalent in another application or circuit!

The main electrical specifications of a transistor are:

A. Maximum voltages: Vce(max) is the maximum voltage that can be applied between collector and emitter. Vceo(max) the same, but with open base. Vcb(max) is the maximum voltage that can be applied between collector and base. Vcbo(max) the same, but with open emitter. Veb(max) is the maximum voltage that can be applied between base and emitter. Vebo(max) the same, but with open collector. In some case the absolute values can be specified as Vc, Vb and Ve.

B. Maximum currents: The most important current in a transistor is the collector current. It can be specified as Ic or If (max).

C. Maximum power: When conducting a current, the transistor generates heat that can be transferred to the air. The maximum amount of heat produced by a transistor is given by its maximum dissipation power. These specifications are given in watts. Small transistors dissipate power in the range between 50 and 500 mW. High-power transistors have power specification above 100 W. The parameter is indicated by P or Pd in the datasheets of transistors.

D. Gain: A small current forced to flow by the base element causes a larger current to flow between emitter and collector. The number of times the collector current is larger than the base current is the gain of the transistor or beta factor (β). If a 1 mA current in the base of a transistor causes a 100 mA current between emitter and collector the gain of this transistor is 100. Normally, the transistor has a specification range of gain or minimum gain. For example, the transistor 2N3905 has a gain range between 50 and 150. The gain also can be indicated by a hybrid parameter named hFE, which is equivalent to the beta factor.

E. Transition frequency: The gain of a transistor falls when working with high-speed signals. But, there is a frequency where the gain falls to unit. This means that in this frequency, a variation of 1 mA in base current causes a variation of 1 mA in the collector current, and the transistor cannot operate as an amplifier. The output current is the same as the input current. This frequency is called transition frequency or fT.

Working with transistors

The electrician who is not very experienced with electronics sometimes can be confused when working with transistors. Millions of types and too many characteristics to observe when trying to replace one are the main factor of this embarrassing situation. It is even more aggravating when one considers that they have three terminals, not just two as the others.

How are the transistors used? How do you test a transistor? How do you find an equivalent? How do you know how they work? How do you know which is the base, collector, and emitter leads?

As the transistor is one of the most important of the electronic components in any electronic device, more information will be provided on this component later.

How the transistors are used

Figure 101 shows how a transistor is used in the two basic operation modes (linear and saturated). In (a) the load is placed between the power supply and the collector of an NPN transistor. The resistor in the base determines the amount of base current. This resistor is called the biasing resistor. When closing S1 the base current causes the current to flow through the load.

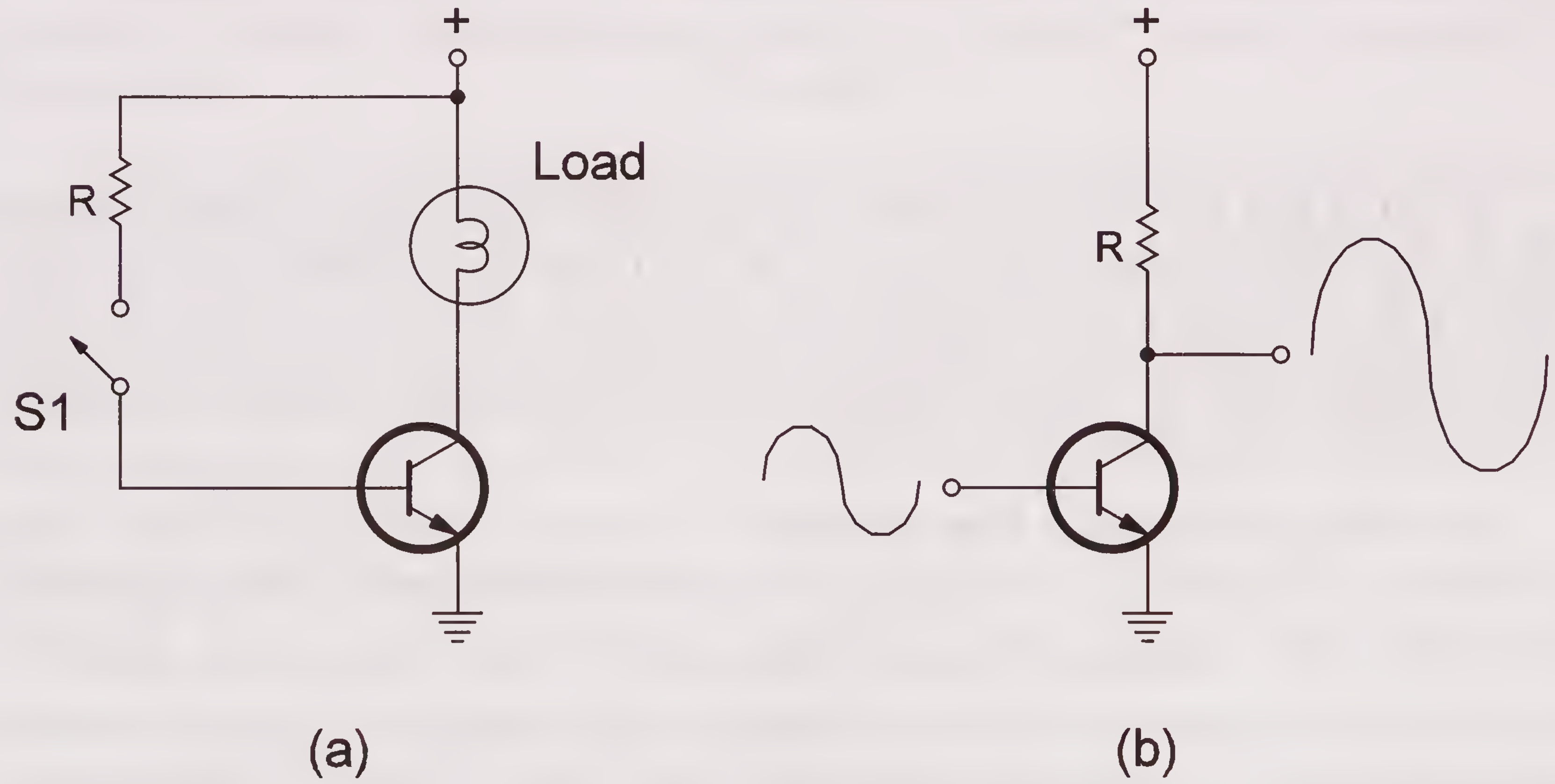

Figure 101

If as shown in (b), a variable current as the correspondent to an audio signal is applied to the base, the small variations of this current will produce large variations in the collector current.

The way the transistors are biased and how the signals to be amplified can be applied and taken off a circuit determine three basic configurations of the transistors. The three ways or basic configurations are shown in Figure 102.

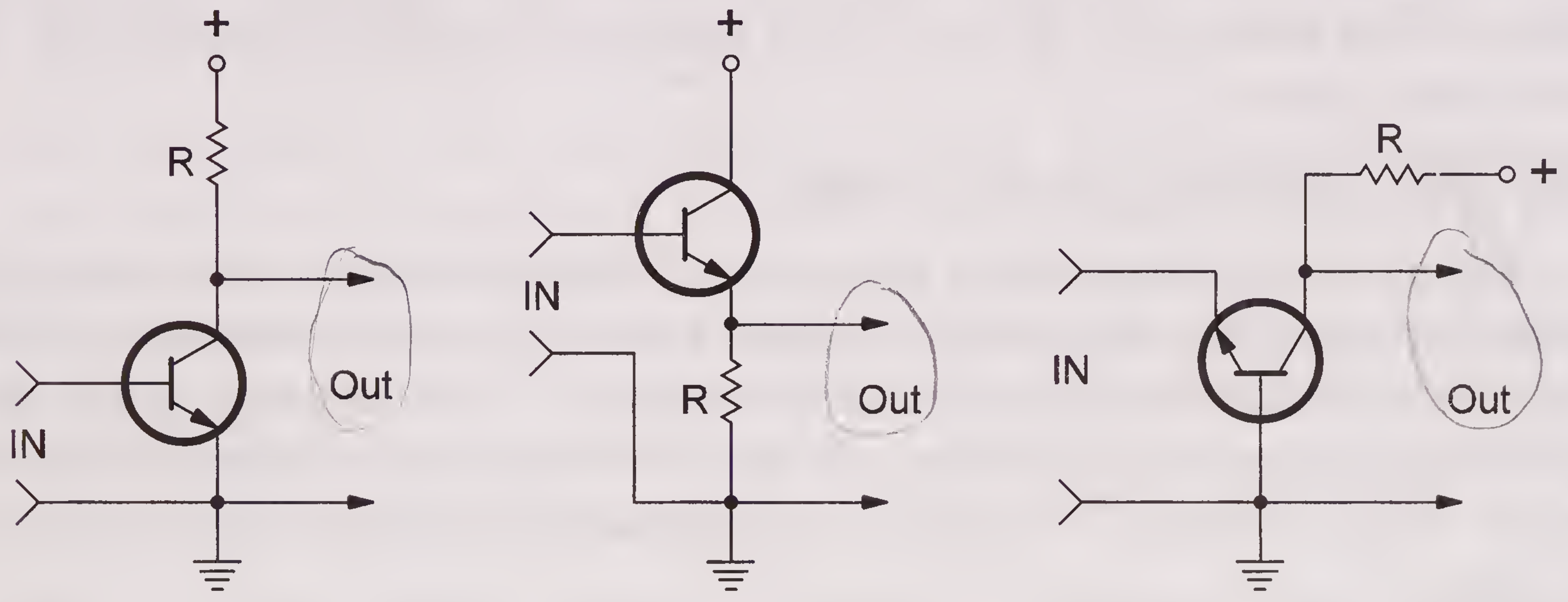

a) Common emitter b) Common collector c) Common base

Figure 102

In (a) is the common emitter configuration presenting larger voltage and current gain (power gain), in (b) is the common emitter configuration giving a high current gain but low-voltage gain, and in (c) the common base configuration is shown. Based on the application, one configuration can have the best performance and is preferred.

Normally, more than one transistor is used in any application. Each transistor is the center of a stage and the amplified signals must pass from one to the other stages being amplified until the desired level is reached.

An audio amplifier, for example, has many stages. Starting from the point where the preamplifier takes the weak signal from a microphone, passing it to intermediate amplification stages, and ending in a power output stage, it can source enough power to drive a loudspeaker reproducing sound with a desired volume. Figure 103 shows a typical circuit of an audio amplifier using two transistors.

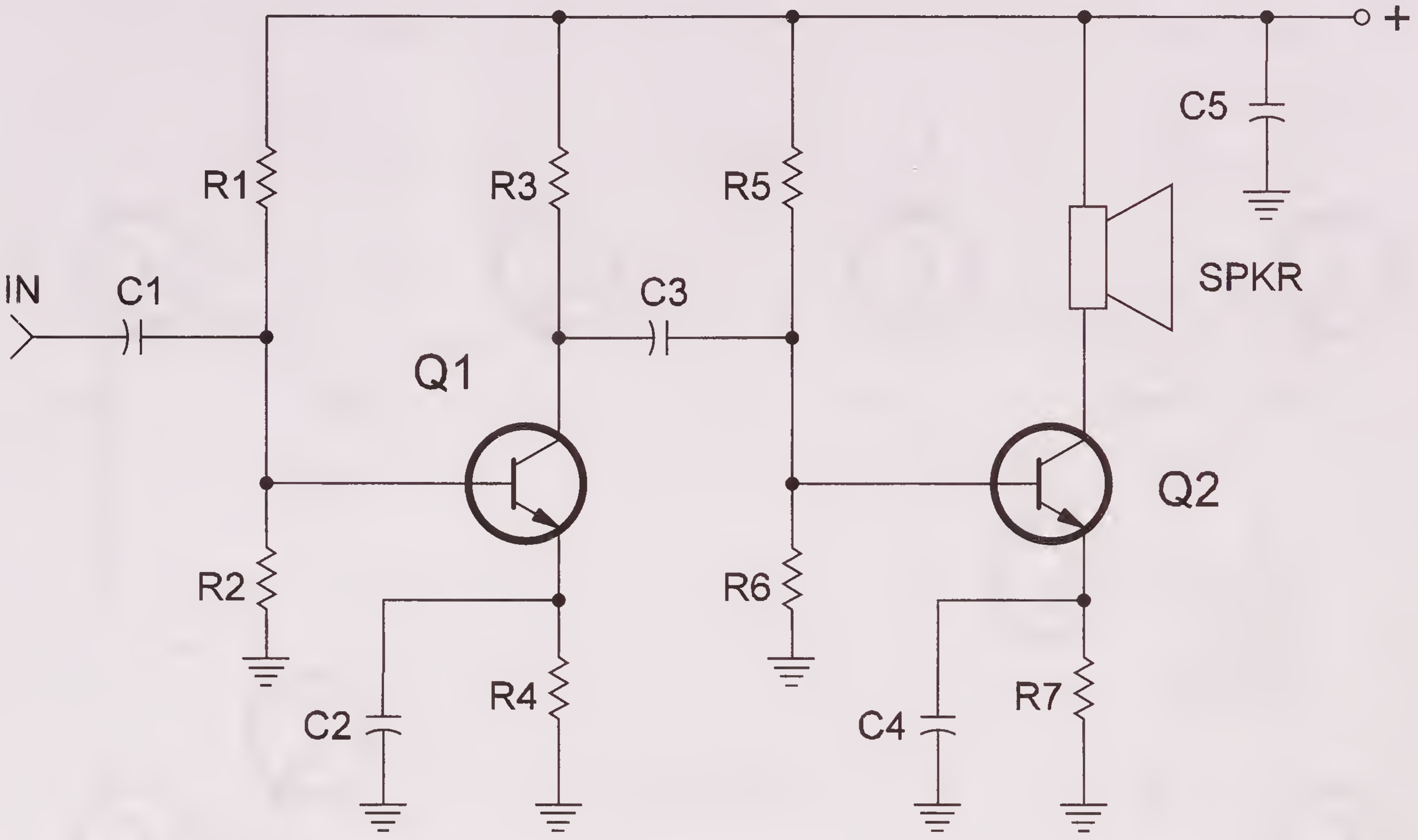

Figure 103

The way the transistor transfers the signals from one stage to another is called coupling, and there are several coupling methods as shown in Figure 104.

In (a) is an RC coupling. The signal is transferred from one transistor to another by a capacitor. In (b) is an LC coupling. The capacitor is also used to transfer the signal from one stage to another. In (c) is a transformer coupling, and in (d) is the direct coupling. A special method of direct coupling, called a Darlington coupling (not shown), is when two NPN or PNP transistors are used.

Testing and identification

The electrician who is just starting with electronics may find some appliances using transistors that are in trouble. How should one proceed in this case? Transistors are found practically in all equipment, forming the central element of many stages. Each stage can have as core one or more transistors in one of the configurations we have seen before. The main test that the electrician can do is measuring the voltages in the leads of the transistor using a multimeter.

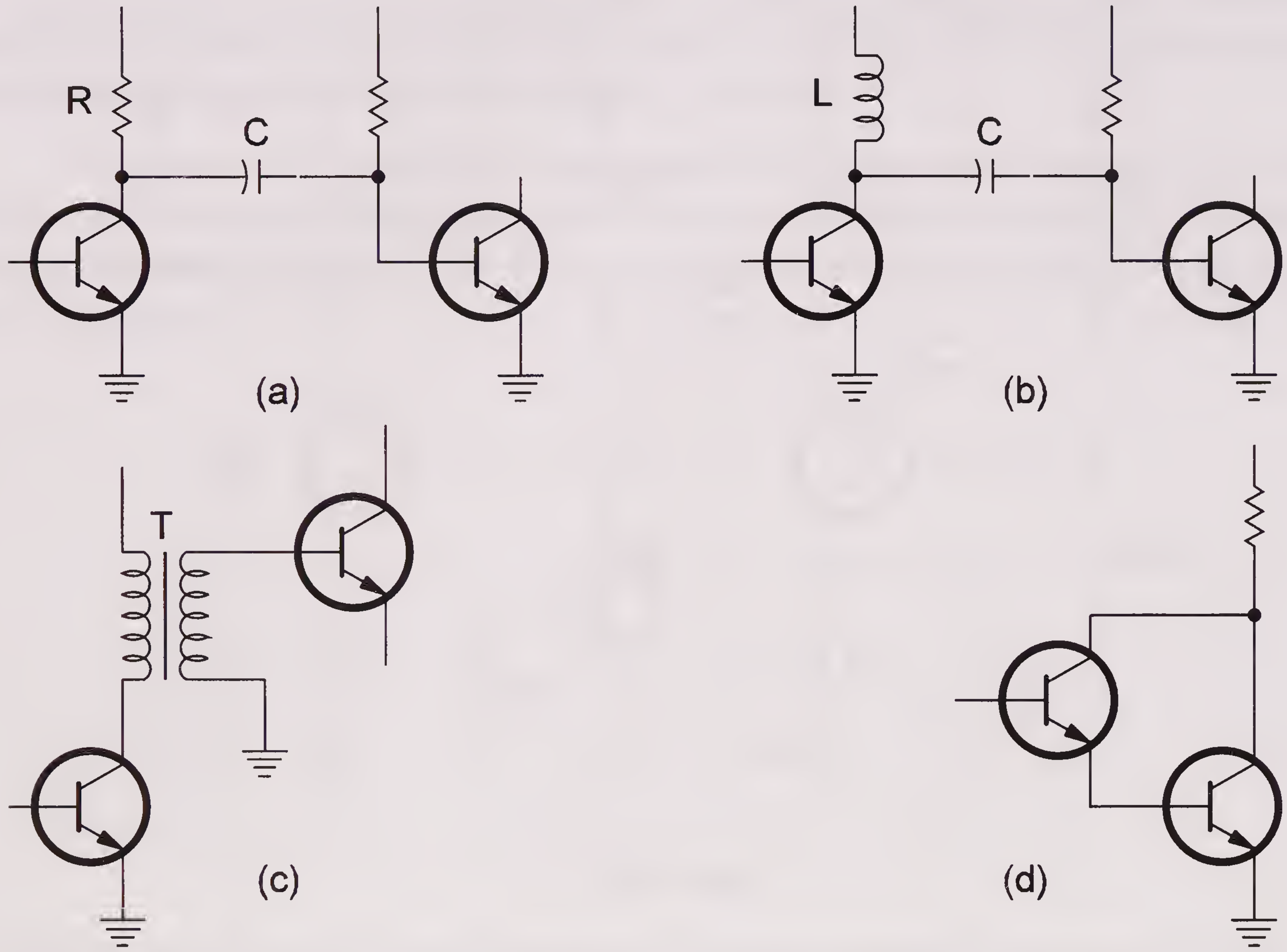

Figure 104

As a general rule, the collector voltage is higher than the emitter voltage in an NPN transistor and the inverse in a PNP transistor as shown in Figure 105.

The base voltage is 0.6 V to 0.7 V above the emitter voltage in an NPN silicon transistor and the same below the emitter voltage in a PNP silicon transistor. In old germanium transistors the difference is 0.2 V. If the voltages are different, the biasing resistor or other elements surrounding the transistors can have problems.

The static test of a transistor or the test out from a circuit, is done by measuring the resistance between the elements (EBC) using a multimeter. It is a "junction test" where it is verified whether the two junctions (base-emitter and base-collector) are okay.

The basic test consists of "seeing" if the two "virtual" diodes corresponding to the two junctions are good. Simply adjust a multimeter to a

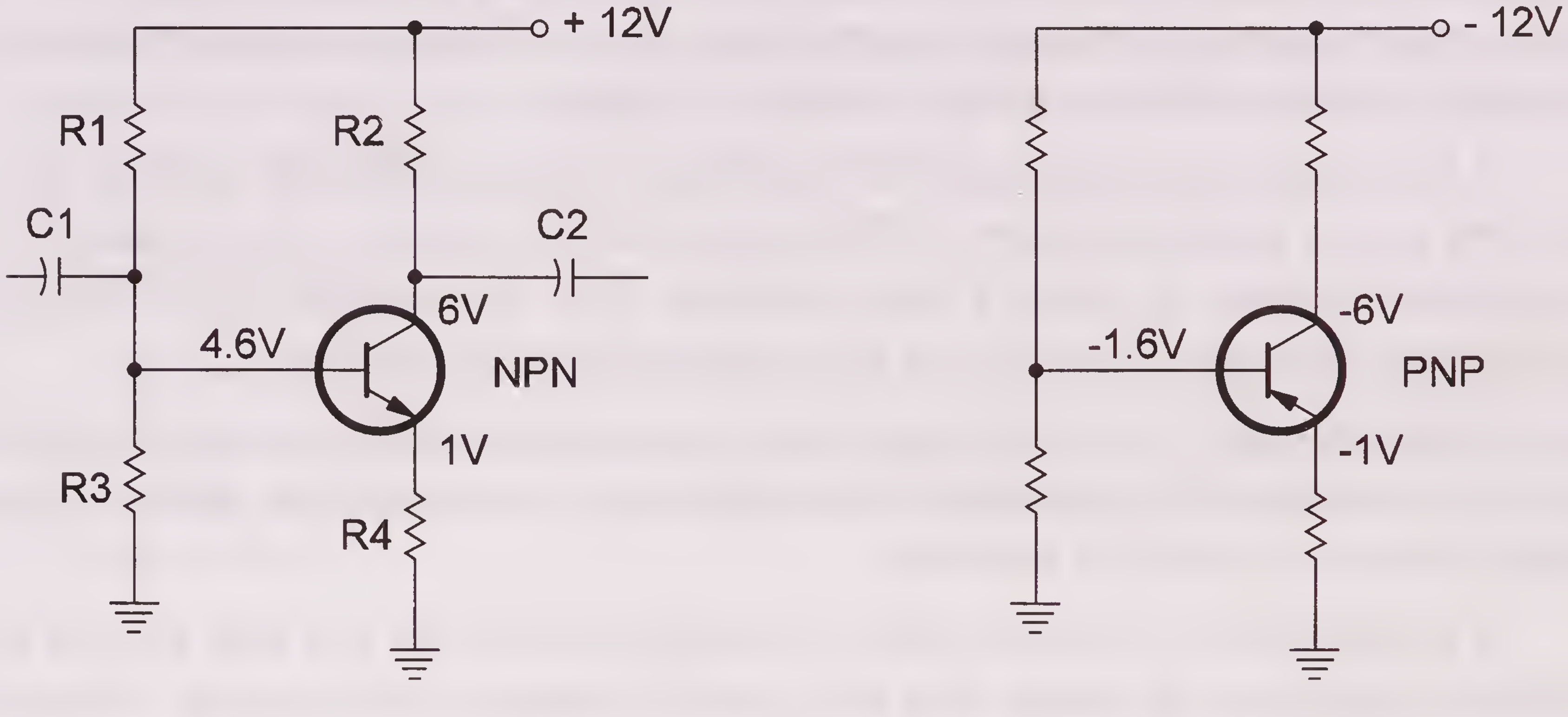
+ 12V
R1
R2
C1
C2
6V
4.6V
NPN
1V
R3
R4
- 12V
-6V
-1.6V
PNP
-1V

(a) (b)

Figure 105

low-resistance scale and make six measurements in on the transistor to test it. Figure 106 shows how to do these measurements.

When placing the probes between base/collector and base/emitter one measure must be of a high resistance and the other a low resistance.

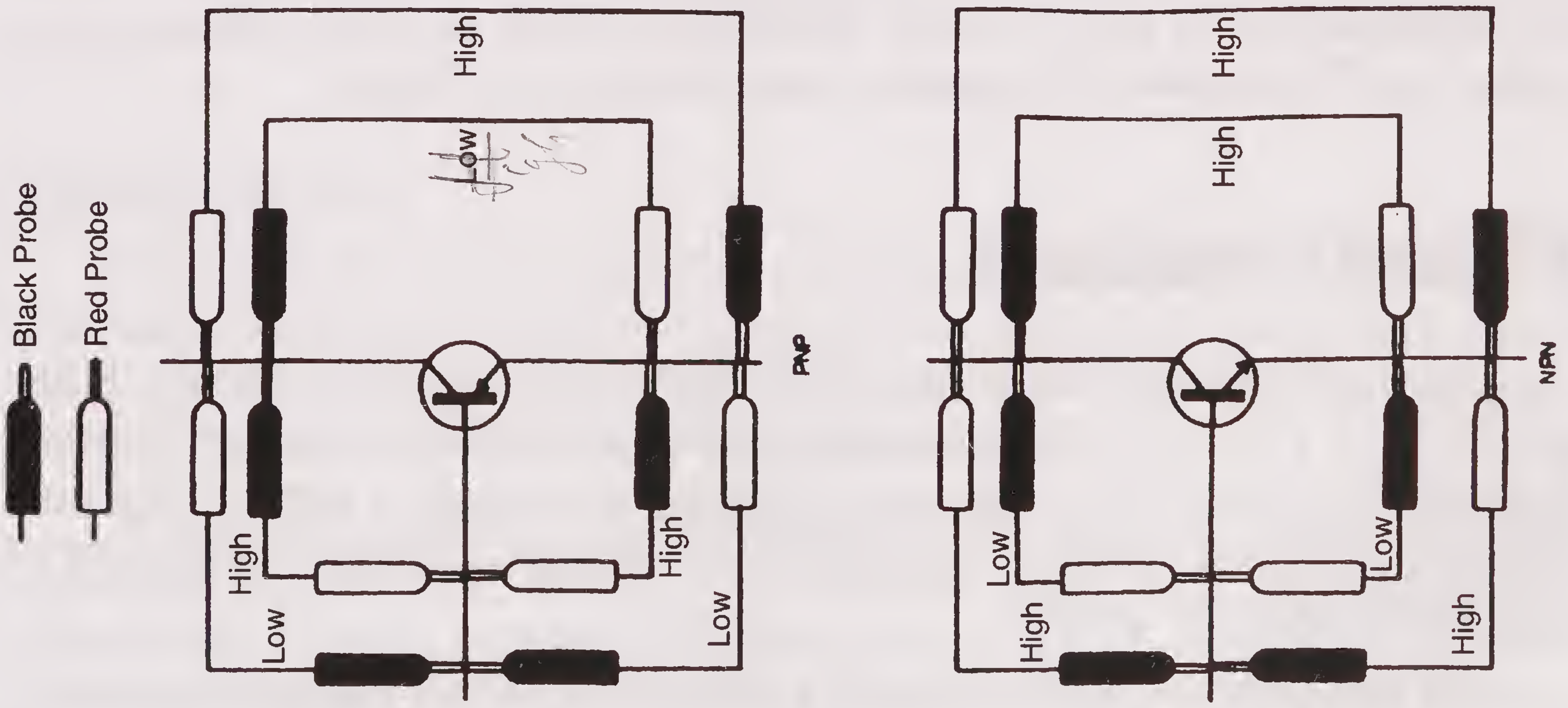
Black Probe
Red Probe
High
Low
High
Low
High
Low
PNP
High
High
Low
High
Low
High
NPN

Figure 106

Both measures, when placing the probes between collector and emitter, must result in a high-resistance reading. Any different result indicates a transistor with problems, either shorted or open.

If the expected resistance is high (more than 1,000,000 ohms), and we find some value between 20,000 and 500,000 ohms, the transistor is presenting losses. It is also a bad transistor. But, the main problem for the electrician who has to replace a transistor is how to identify it.

As this book is directed toward beginners, the indications given touch only the surface of the problem, although they are enough to give you an idea of how to solve the problem.

If a problem in a circuit using a transistor is found, the first step is to test the transistors to certify that they are the cause of the trouble. Always remember that a transistor can burn if associated components (capacitors and resistors) fail. A capacitor, in short, can cause high currents across a circuit burning a transistor. So, before replacing a burned transistor, test all the components near it as well.

For the replacement, if the original cannot be found, look in a handbook for equivalencies, replacement types, or the characteristics. As a practical rule, a transistor for the same function (small signal for instance) and the same type (NPN or PNP) with the same or greater gain, the same or greater Vce, and the same or greater collector current (Ic), can be used as replacement. In some critical applications such as high-frequency circuits, the fT (transition frequency) must be equal or higher.

Phototransistors

As in diodes, the junctions of a transistor are sensitive to light. If the junctions of a common transistor are exposed to light, the current flowing through the component changes and it can be used as a sensor. The current flow between collector and emitter can be controlled by the amount of light falling onto the device.

By mounting the silicon chip of a transistor inside a transparent enclosure or an enclosure with a window to let the light fall onto it, it becomes an important electronic device—the phototransistor.

Phototransistors are used the same way as the photodiodes—as sensors in many electronic applications. One important aspect of phototransistors when compared with LDRs is that they can be used to detect very fast changes in the light intensity.

Symbol and Types

Figure 107 shows the symbol adopted to represent the phototransistor and a few types. The phototransistor can "see" light invisible to the human eyes. They can be used to detect both IR (infrared) and, in some cases, UV (ultraviolet).

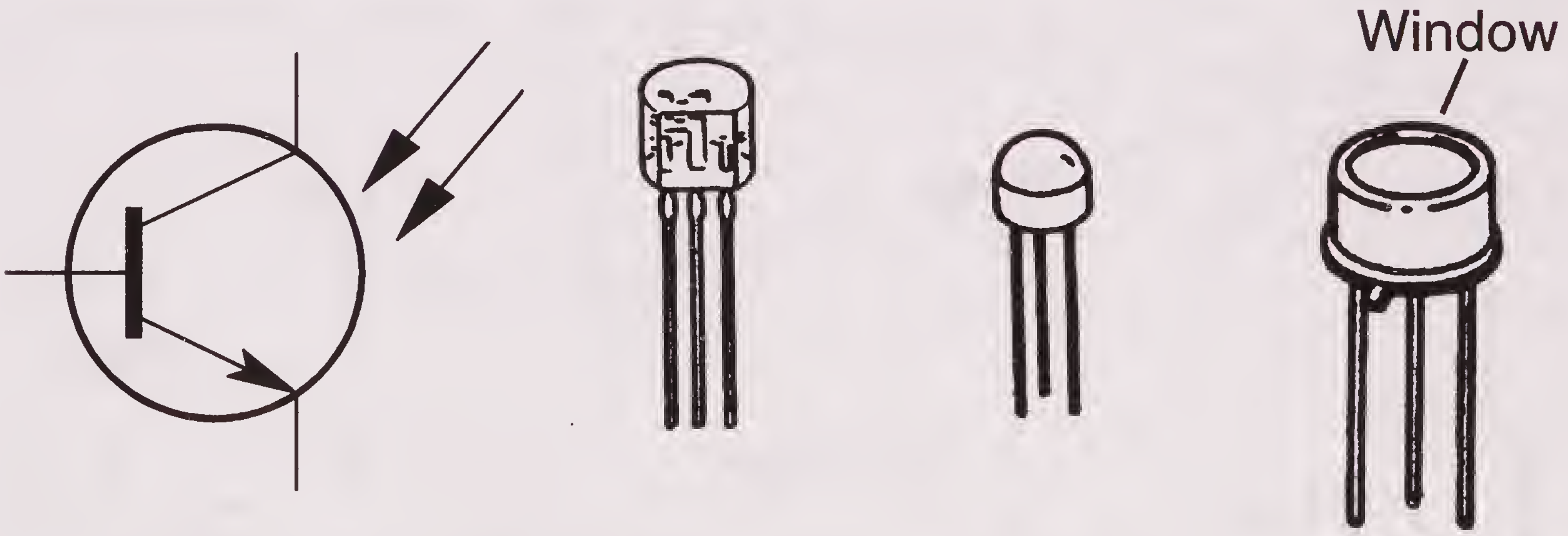

Figure 107

Specifications

The manufacturers mark their phototransistors with a part number. Characteristics such as wavelength response, gain, collector current, maximum operating voltage, and others must be found from using the part number to consult the manufacturer's documentation.

A common series of phototransistors are the one made by Texas Instruments. The first letters (TIL) indicate that it is an optoelectronic device and the next number indicates a special type. Examples of these part numbers are TIL81 and TIL411. The same initial letters are used to indicate other components working with light, such as LEDs and optocouplers.

Where and how they are used

The electrician will find phototransistors in many applications related to professional activities. In fact, alarms, motion detectors, remote controls, and many other applications use the phototransistor as a sensor.

Because the phototransistor is much faster than the LDRs, they have different uses. The electrician will find the LDR in light detectors where the changes or amount of light are slow as in emergency lights and automatic nightlights. In the applications where fast changes of light must be detected or where information must be sent through a light beam (visible or not) the photodiode and the phototransistor are preferred.

In a typical application the phototransistor is used as a diode, as shown in Figure 108.

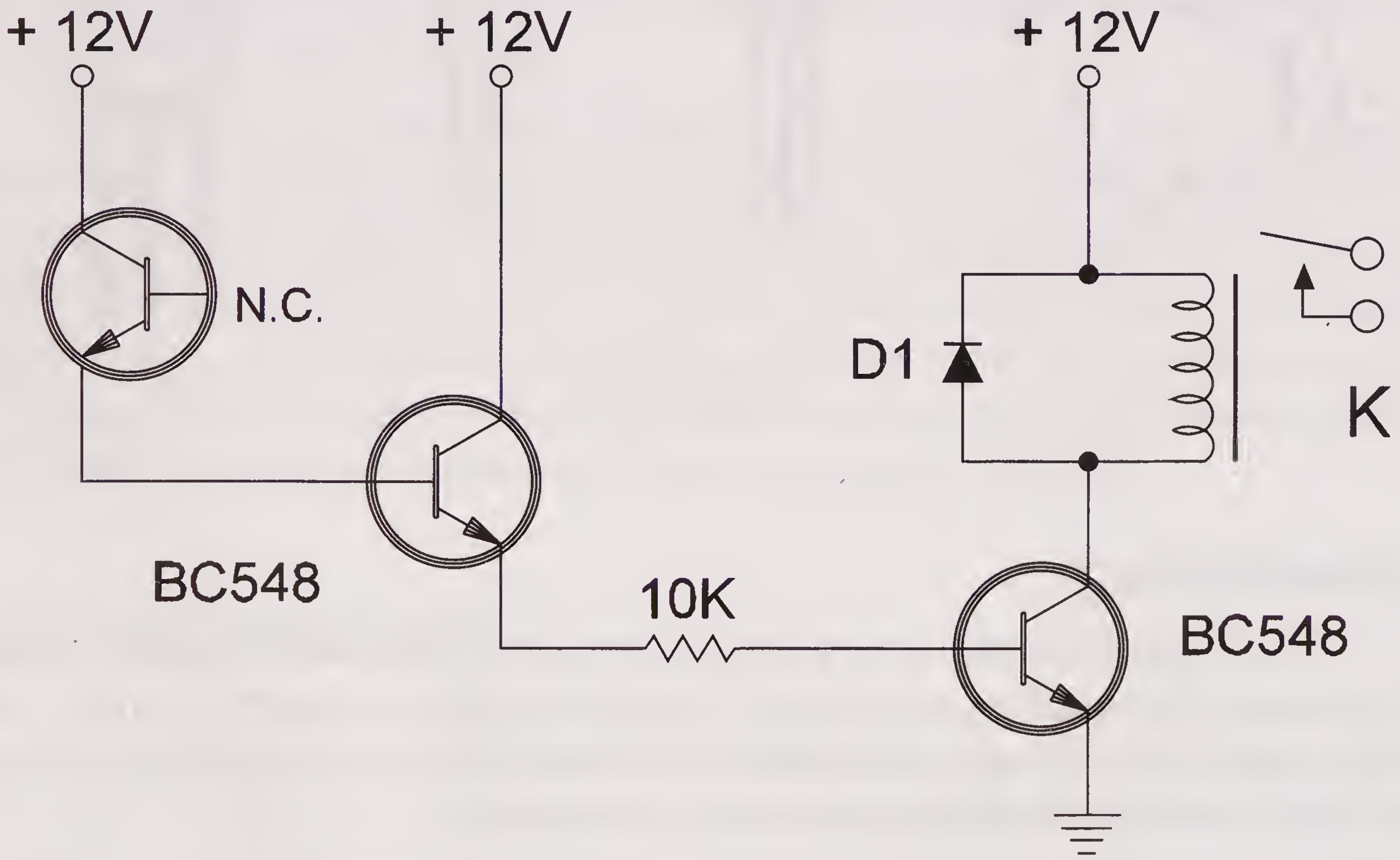

Figure 108

In the circuit the base terminal is let free and the component acts as a photodiode. The current flowing between emitter and collector depends on the light picked up by the device. In many cases, although the component is a transistor, the package doesn't present the base terminal.

In order to increase the amount of light picked up by a phototransistor a lens can be placed in front of the transistor. It is important to place the transistor in its focus as shown in Figure 109.

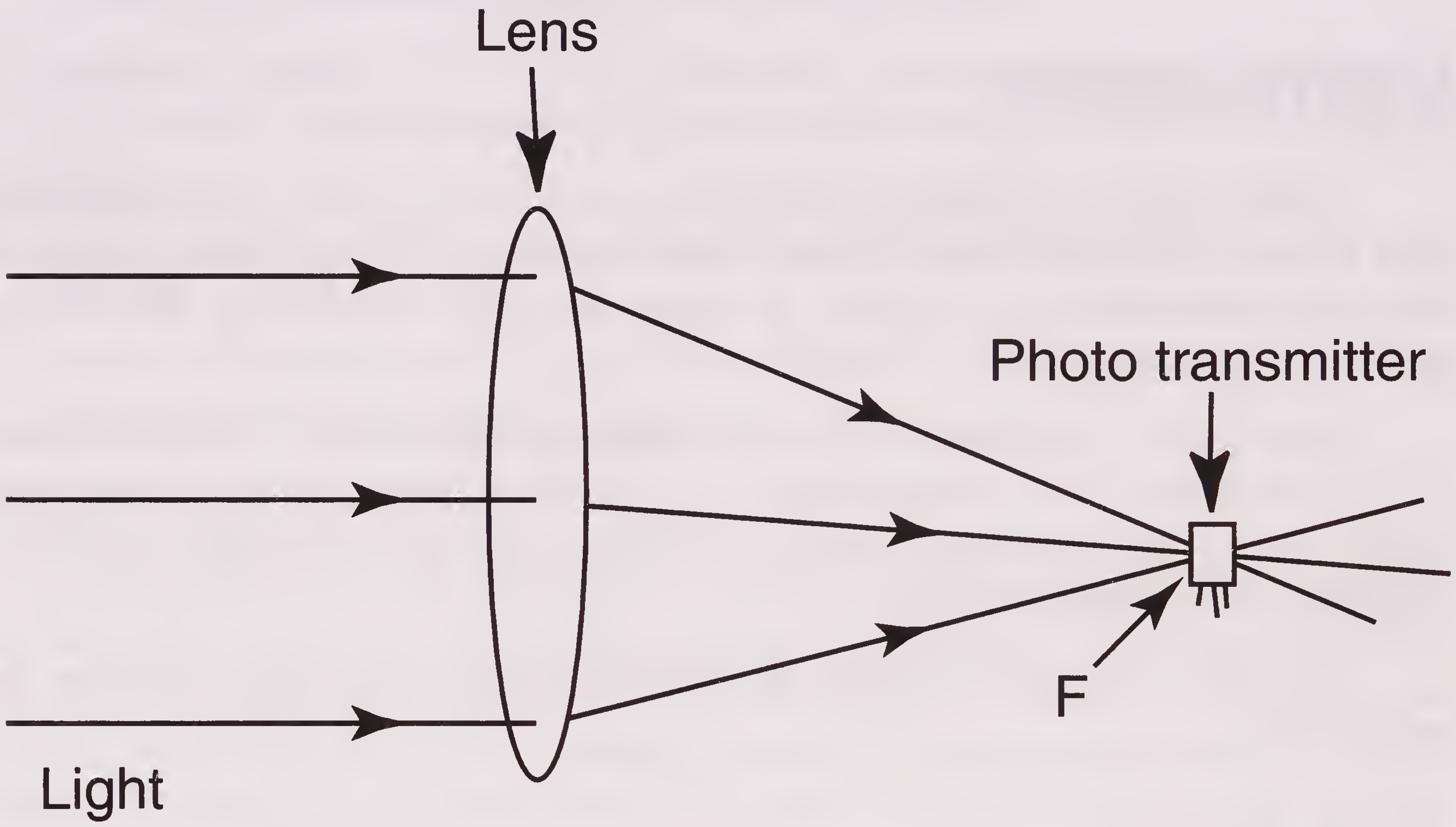
Lens
Photo transmitter
F
Light

Figure 109

A typical application of a phototransistor and an infrared LED is shown in Figure 110.

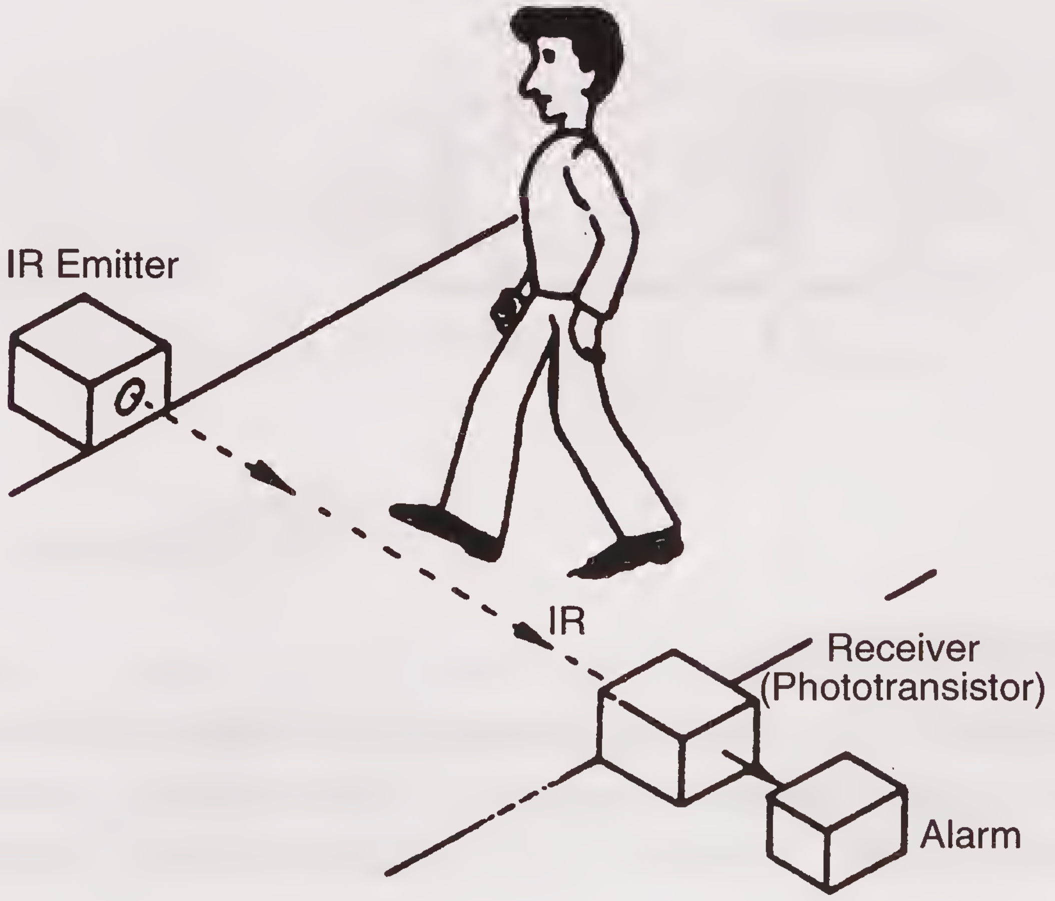
IR Emitter
IR
Receiver
(Phototransistor)
Alarm

Figure 110

The infrared LED sends a light beam that is focused on the phototransistor. If the light beam is interrupted (cut by an intruder) the circuit detects the cut and triggers an alarm.

Optocouplers

Optocouplers are devices formed by an LED (IR) and a phototransistor or a photodiode installed inside the same enclosure. The LED is pointed to the phototransistor in a manner in which the light emitted by one can be detected by the other.

If a signal is applied to the LED it can be transferred to the transistor by a light beam. This means that no electric connections exist between them, so the transmitting circuit and the receiving circuit are electrically isolated one from the other.

The use of an optocoupler is important when isolation problems are important.

Symbol and View

Figure 111 shows a view and the symbol used to represent an optocoupler.

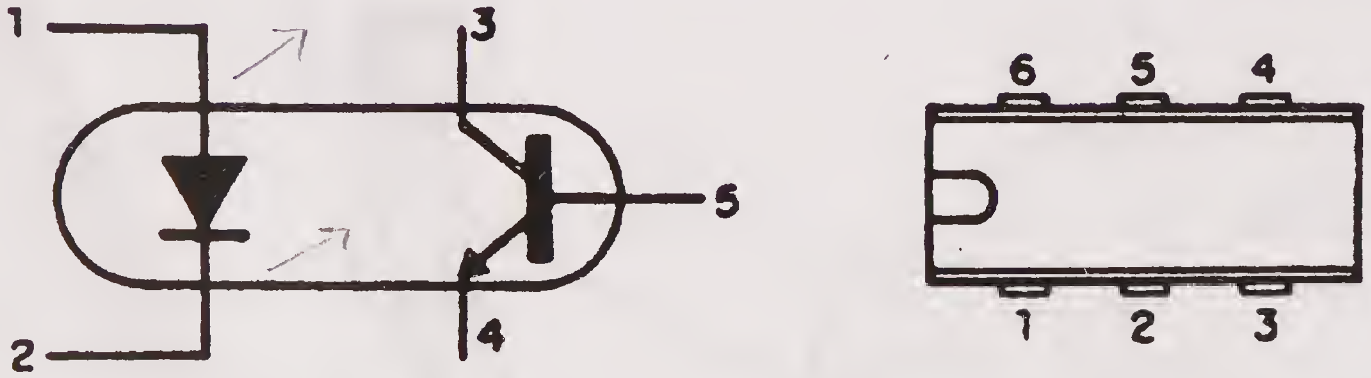

Symbol Aspect

Figure 111

Specifications

There are many important specifications for an optocoupler. The first thing it is important to know is if the sensor element is a photodiode, a phototransistor, or other light-sensitive device (there are many others). The part number is important, but gives no information on this topic. The electrician will need a handbook with the characteristics or documentation from the manufacturer.

The main characteristics and their meanings are:

A. Isolation voltage: This is the maximum applied voltage between the LED circuit and the sensor. Typical values range from 3000 to 7000 volts.
B. LED characteristics: These are the maximum forward current and voltage.
C. Sensor characteristics: This is the maximum voltage between collector and emitter if a transistor is used.

Where they are used

The electrician can find optocouplers in applications where a signal coming from a circuit or the power supply line must be isolated from other circuits. For example, a circuit that can be triggered from the AC power line voltage, but must not be connected directly to it by security requirements. Another application is interfacing PCs and microprocessors with circuits powered from the AC power line. Using an optocoupler, the delicate circuits of the PC are isolated from the AC power line. If something goes wrong with the circuit, the PC is protected against damage.

Testing

The optocouplers can be tested by independent verification of the LED and the phototransistor. Another way is to excite the LED from a power source (doen't forget the limiting resistor) and see if the resistance between collector and emitter of the internal transistor falls. A multimeter on the highest resistance scale can be used in this test.

Darlington Transistors

Two transistors wired as shown in Figure 112 form a Darlington stage or a Darlington Amplifier. This stage acts as a super transistor or a single transistor with gain corresponding to the total product of the gains of the individual transistors.

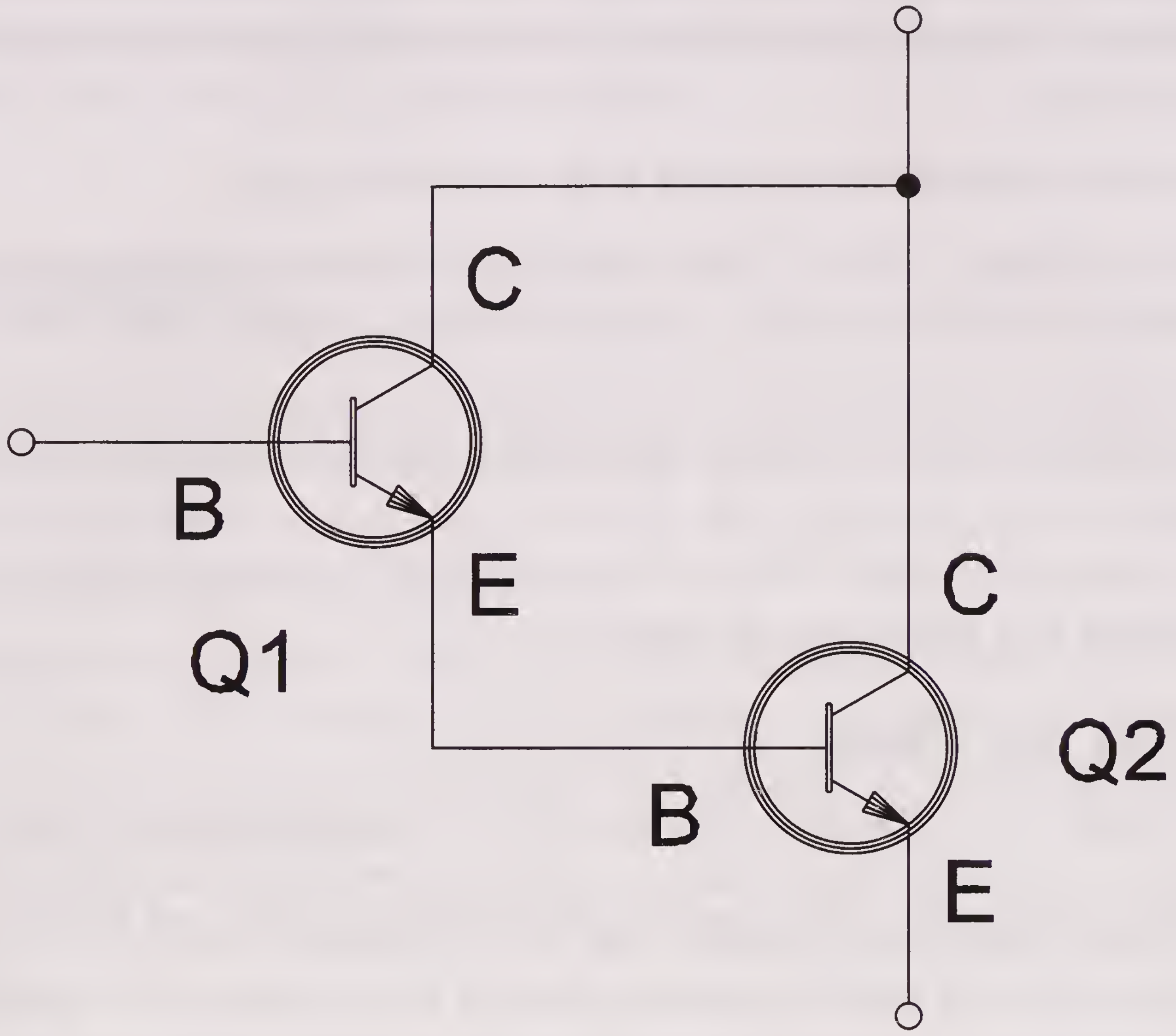

Figure 112

For example, if two transistors with gains equal to 100 ($\beta = 100$) are placed together to form a Darlington pair the final gain will be 100 x 100 or 10,000! They function as a single transistor with 10,000 of gain.

A Darlington transistor can be manufactured by placing two transistors in the same silicon chip, packing them inside a common enclosure forming a super transistor. Darlington transistors are useful where high gain is necessary, such as in solenoid, motor, relay, or lamp drivers as well as high-power audio amplifiers.

Symbols

The symbols for Darlington transistors are shown in Figure 113. As a common bipolar transistor, they are found in the NPN and PNP versions. In (a) is a low-power general purpose Darlington transistor and in (b) a high-power type that must be mounted in a heat sink.

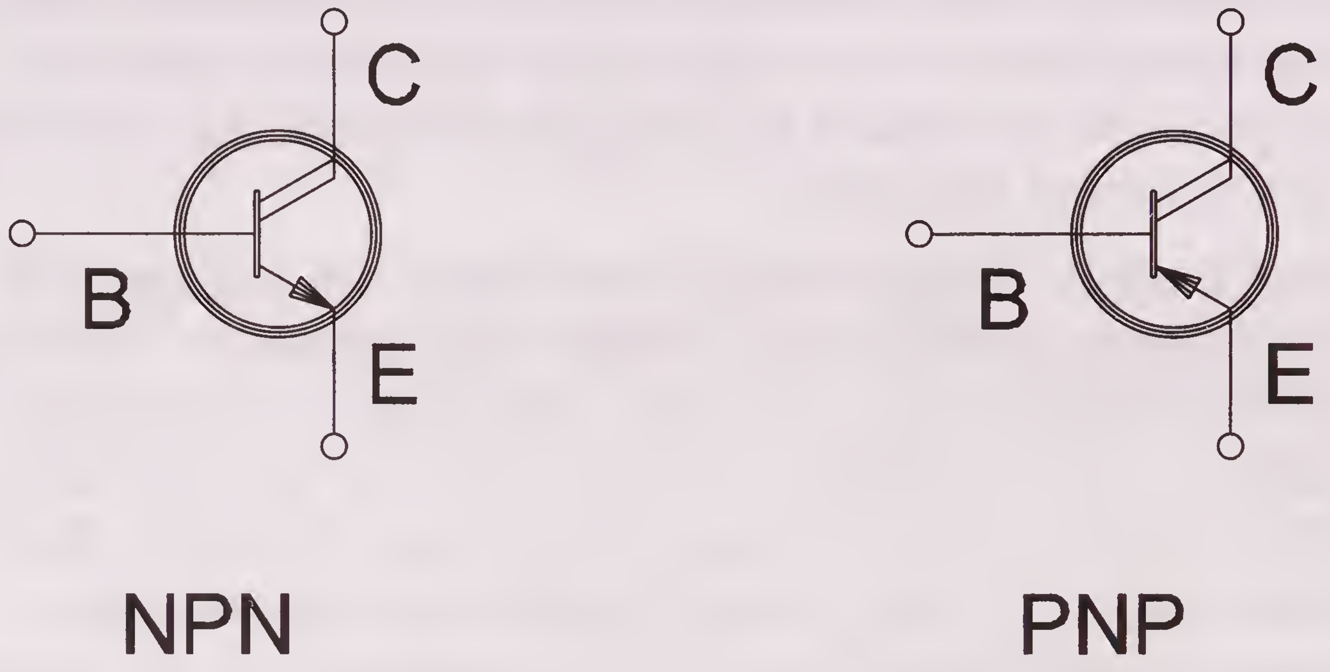

Figure 113

By the simple look of the component it isn't possible to know if it is a common transistor or a Darlington transistor, or any other semiconductor because they all use the same bodies. The identification is made by the type or part number that is normally a figure on the component's body.

Specifications

If you have a handbook of transistors or the specific data sheet (or diagram) for the equipment, it is possible to know if the component is a Darlington transistor.

A common series of Darlington transistors found in many electronic appliances and even in electric applications is the TIP series of Darlington transistors first manufactured by Texas Instruments and now by many other manufacturers. This series is formed by a number of Darlington NPN and PNP transistors designed to control currents from 1 A to more than 10 A and voltages up to 100 V and possibly even higher.

Some Darlington power NPN transistors in this series are TIP110, TIP111, and TIP112 for 2 A of collector currents. The complementary PNP types are the TIP115, TIP116, and TIP117.

Important: Not all transistors beginning with TIP are Darlington transistors!

The data sheets or data books normally have a lot of information about transistors including many curves and numerical information about performance. For electricians or even electronic technicians who only want to know enough to use or replace a Darlington transistor the following specifications are the most important.

A. Collector-Emitter Voltage (Vce)—the maximum voltage that can be applied between collector and emitter of a transistor. Typical values range from some volts to more than 400 volts. In some data books a third letter or group of letters is found after the "ce" indicating special conditions for the value. We can find a (max) indicating that it is the maximum value or a (typ) indicating that it is a typical value.

B. Collector Current (Ic)—the maximum current that can be controlled by the device. This current flows between collector and emitter when the transistor is in the conducting state and varies from some milliamperes to more than 20 amperes to even more heavy-duty types. The additional indication of (max) can also be found with the values.

C. Gain (hFE)—Gain for common Darlington transistors can vary from 1000 to more than 10,000. For what the gain means, refer to specifications in for bipolar transistors.

D. Dissipation Power (Pd)—The transistor generates heat when it conducts a current, which is transferred to the area surrounding the transistor. The Dissipation Power in watts gives the maximum amount of heat that can transfer to the surrounding area. Small-signal transistors can have dissipation in the range of milliwatts, but high-power types can dissipate powers up to 250 watts and more.

E. Transition Frequency (fT)—Darlington transistors are slow devices and can't operate in frequencies above some hundred kilohertz. The maximum theoretic operation frequency is given by the transition frequency of fT (the gain falls to 1) and is typically near 100 kHz for most common types.

Where they are found

Many common appliances use Darlington transistors to drive high current loads including solenoids, electromagnets, motors, and lamps. This is shown in the example circuit in Figure 114.

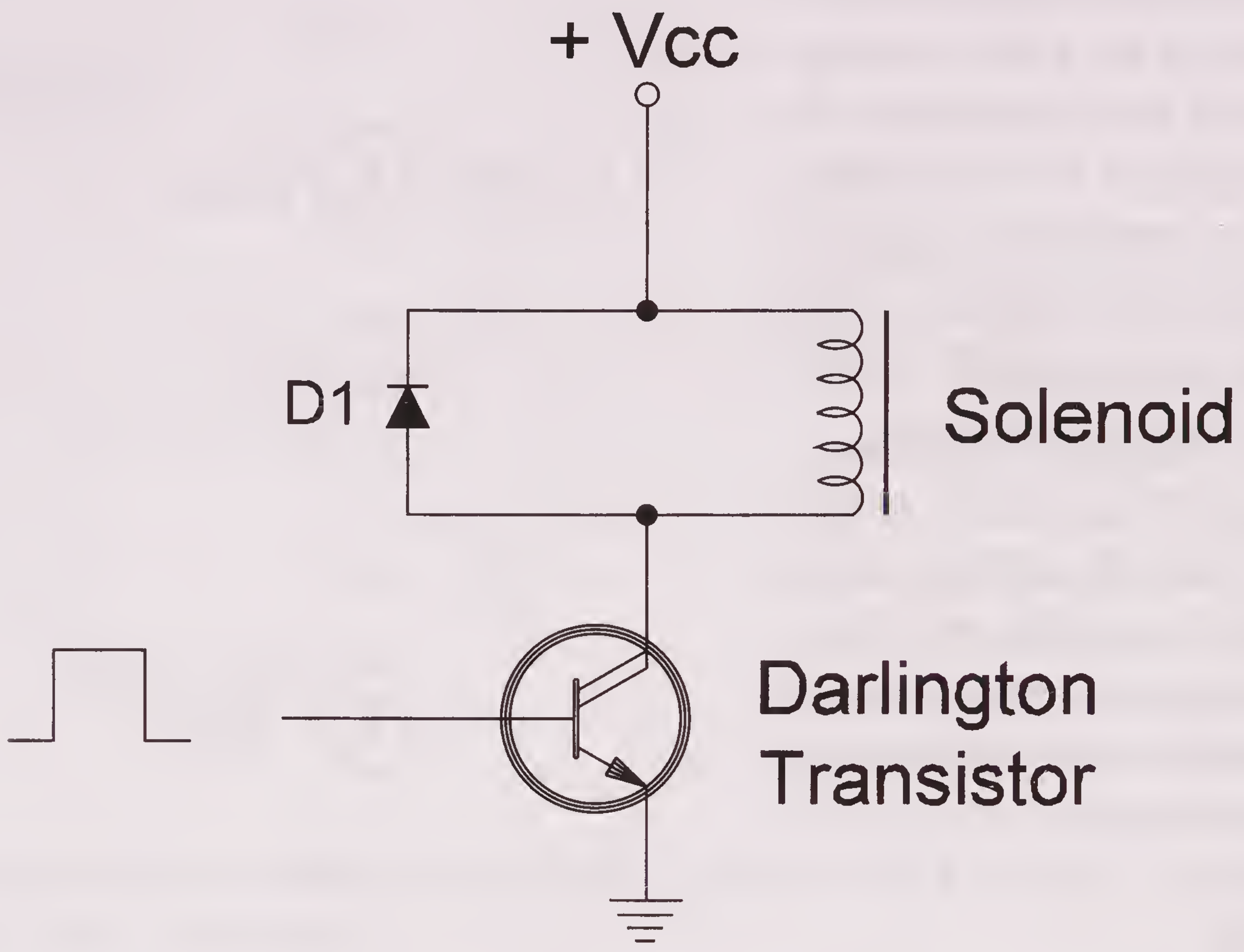
+ Vcc
D1
Solenoid
Darlington
Transistor

Figure 114

The load is connected between the power supply and the transistor's collector. When the transistor is biased it conducts the current energizing the load. In this configuration the transistor acts as a switch turning on and off the load according to the base current.

You can find transistors like these driving solenoids in electric appliances, electric doors, and in electronic equipment such as VCRs and DVD players. Darlington transistors can also be used as low-frequency amplifiers in other applications (Figure 115).

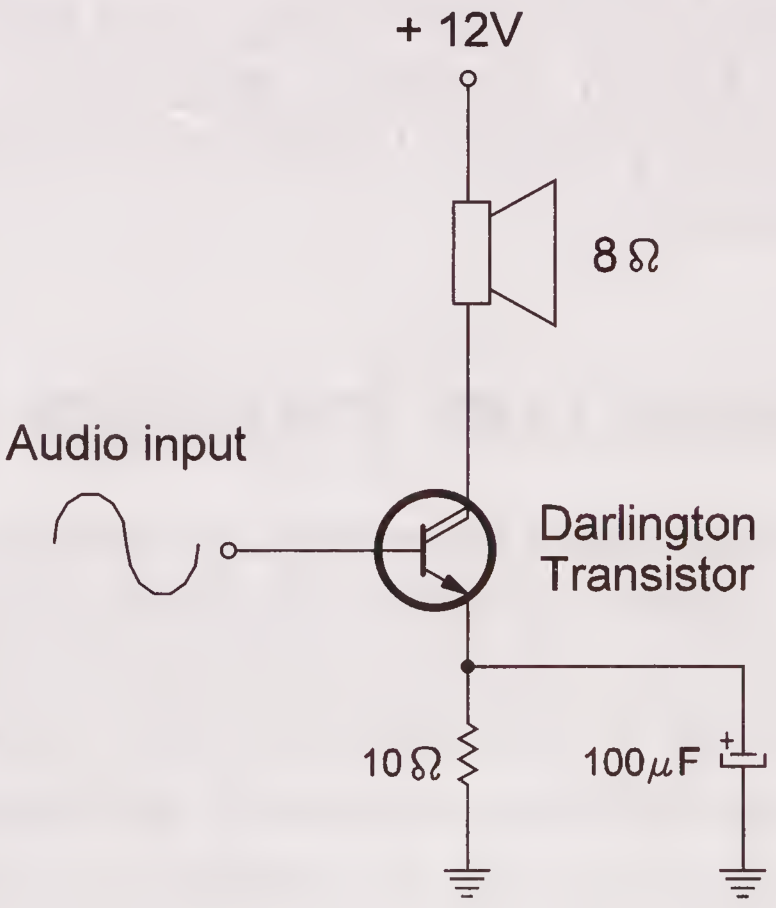
+ 12V
8 Ω
Audio input
Darlington
Transistor
10 Ω
100μF

Figure 115

The audio signal (corresponding to a sound) is applied to the transistor's base where it is amplified. Then the transistor triggers a loudspeaker where the signal is reproduced.

Two complementary Darlington transistors (Figure 116) are found in many hi-fi audio amplifiers. This kind of circuit can drive a loudspeaker with powers up to 250 watts.

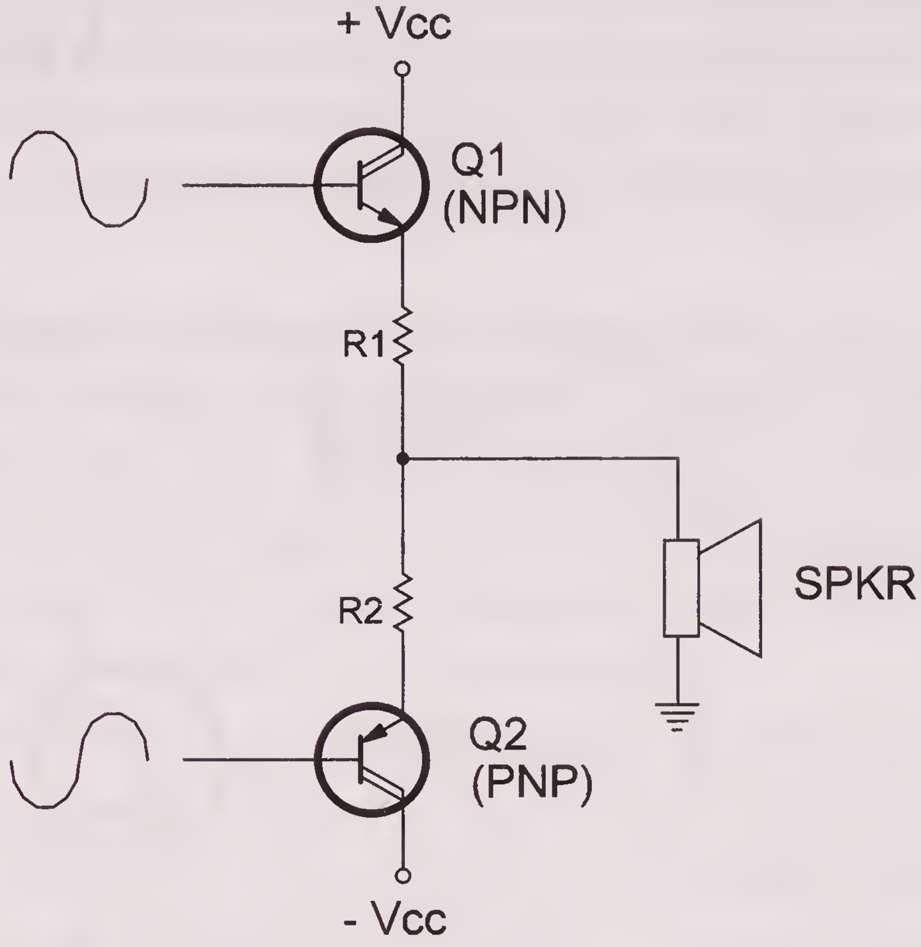

Figure 116

Testing

Darlington transistors can be tested the same way as common bipolar transistors. Using a multimeter, you can measure the resistances between electrodes (base, emitter, and collector) according to the type (NPN and PNP). Refer to the item about bipolar transistors and how to test them. Remember that some multimeters have the transistor test function that can be used with Darlington transistors.

Unijunction Transistor (UJT)

Unijunction transistors are three-terminal devices, but unlike the common bipolar transistors, they have only one junction (as the name suggests). The unijunction transistor or UJT exhibits a negative input resistance that is useful in a number of low-voltage circuits, such as low-frequency oscillators, saw-tooth generators, timers, etc.

As shown by the structure in Figure 117, an N-type silicon bar forms the UJT where two terminals called base 1 and base 2 (B1 and B2) are placed. Between these terminals the device presents an ohmic resistance.

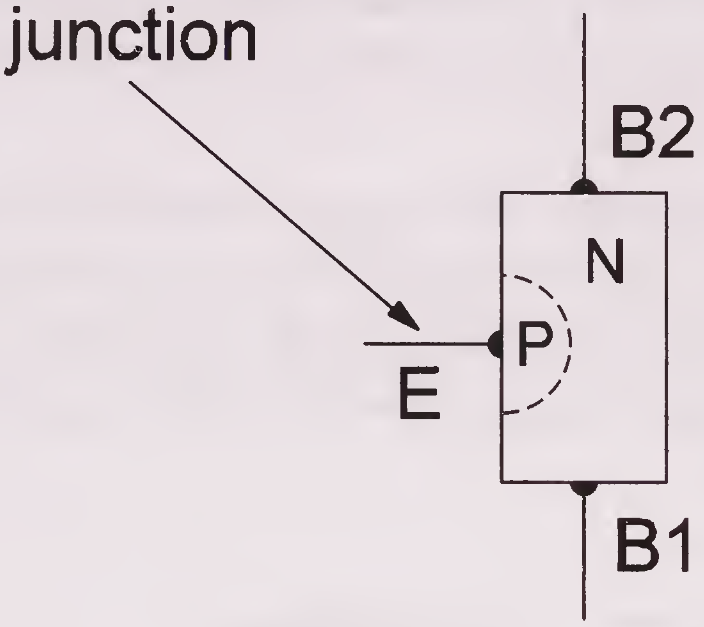

Figure 117

A PN junction is formed in the middle of the N-region where an emitter terminal is placed. In a typical circuit, such as an oscillator, the device is biased as shown in Figure 118.

The N-material placed at the middle of the bar acts as a voltage divider placing a certain voltage in the emitter region. The capacitor placed at the emitter of the transistor charges until the voltage between its plates rises to the value of the emitter voltage plus about 0.6 volts.

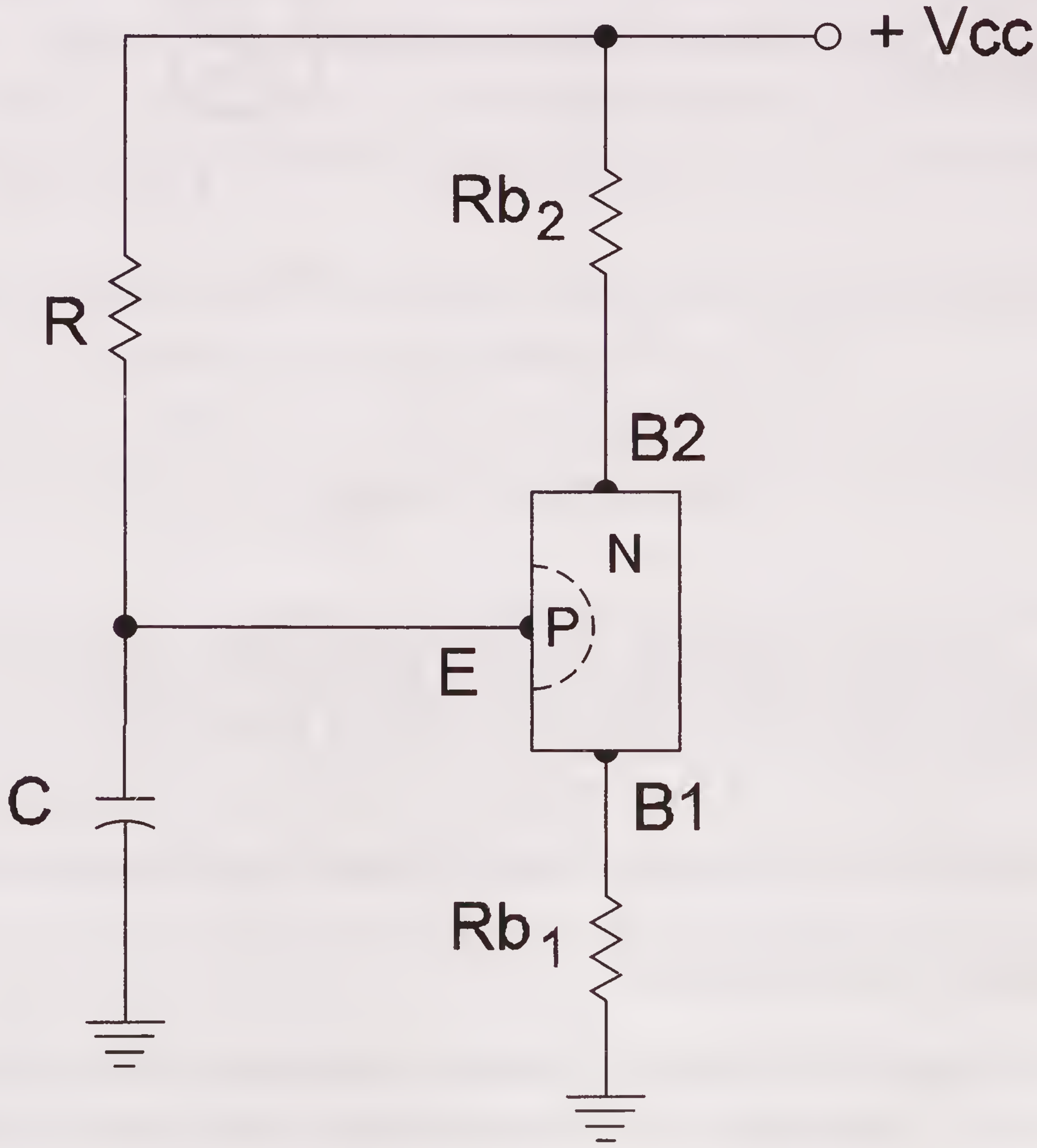

Figure 118

When this voltage is reached the device turns on and the resistance between the emitter to Base 1 falls to a very low value. The capacitor can then discharge through this circuit. A high current pulse is produced in this process.

As long as the charge in the capacitor is reduced, and the voltage across it falls to an appropriated value, the transistor turns off and returns to the nonconductive state. The charging process of the capacitor starts again and a new oscillation cycle begins. Figure 119 shows the waveforms of the signals produced by this circuit.

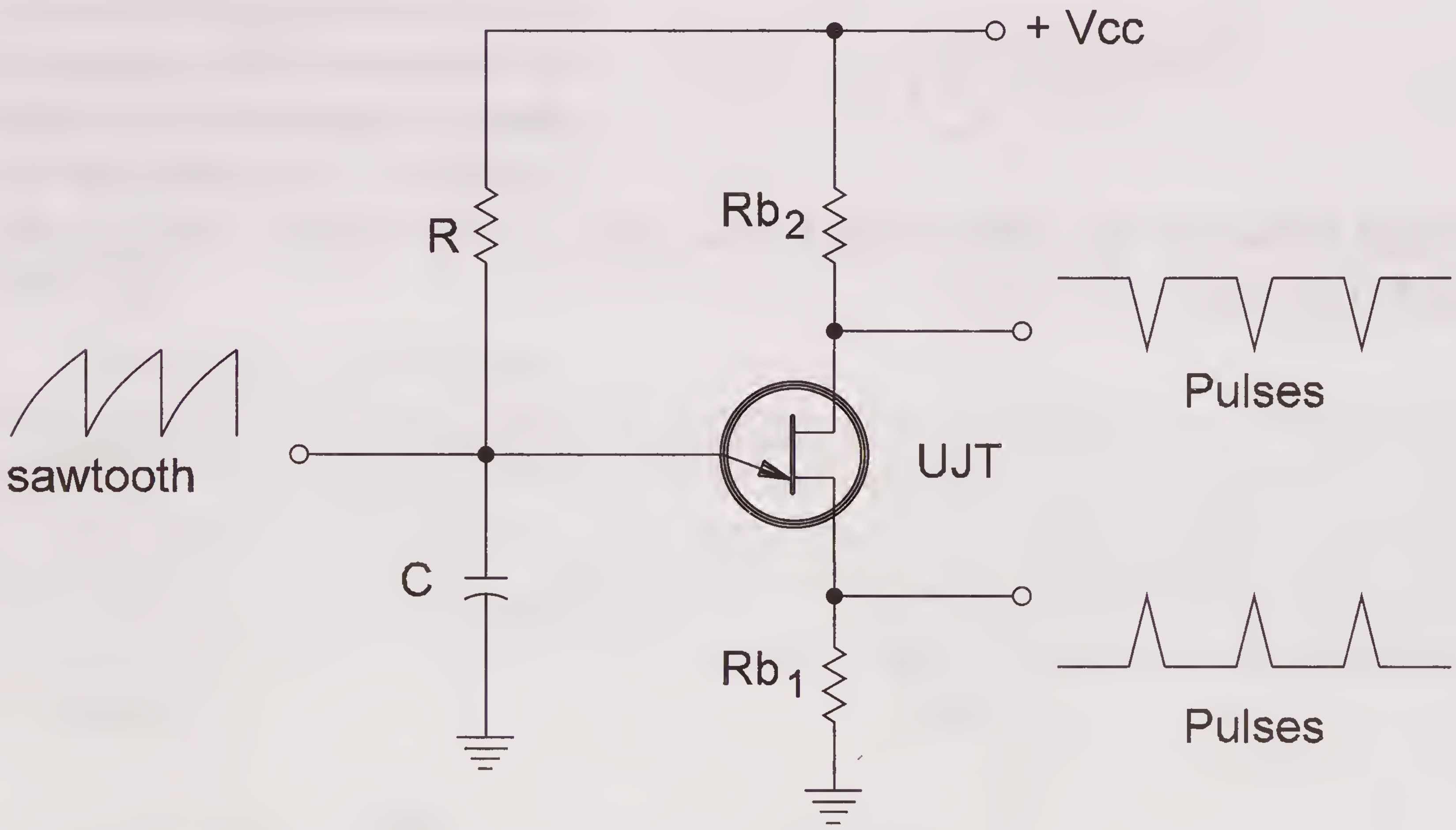

Figure 119

We can consider this the solid-state equivalent of the previously described relaxation oscillator using a neon lamp.

The UJT is a low-frequency device and signals of no more than some hundreds kilohertz can be produced by this basic configuration.

Symbol and Diagram

Figure 120 shows the symbol used to represent a UJT. In the same figure is a diagram, (b). One of the most popular UJT transistors, found in projects published by electronics magazines and even in some kits and appliances, is the 2N2646.

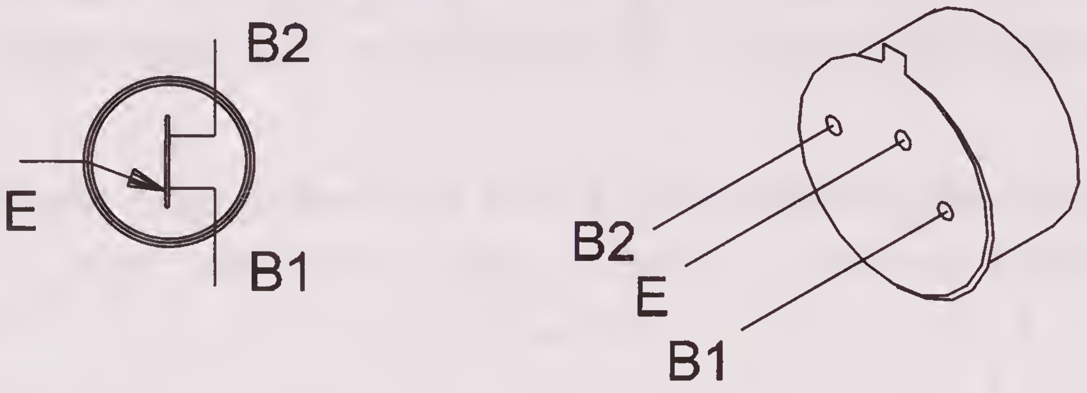

(a) Symbol (b)

Figure 120

Specifications

As in any other transistor, there is a part number figured on the device's body for its identification. But, when building a project and trying to understand how the circuit works or looking for a replacement type, it is important to know some electrical specifications of a UJT. The most important are listed here.

A. Voltage between bases (Vbb)—the maximum voltage that can be applied between bases. For the 2N2646 the specified value is 35 V.

B. Voltage between emitter and base 1 (Vb1e)—the highest voltage that can be applied between the two indicated electrodes. For the 2N2646 the value is 30 V.

C. Intrinsic standoff ratio (η)—the position of the emitter junction referring to the bases. This value is used to calculate the trigger point as a function of the applied voltages. For the 2N2646 this value is in the range between 0,56 and 0,75.

D. Resistance between bases (Rbb)—the ohmic resistance that can be measured between bases. Typical values for the 2N2646 are in the range between 4.7 and 9.1 k ohms.

E. Peak Pulse Emitter Current (Ie)—the maximum current that can flow between emitter and Base 1 when the transistor is triggered.

Where they are found

The UJT isn't a modern device—it was introduced on the market around 1960. Today equivalent circuits using modern devices are used to perform the same functions of a UJT. They are still found in many electronic and electric applications depending on the date they were made or their purpose.

The basic configuration of a UJT is as a relaxation oscillator. This type of circuit is found in timers, alarms, motor controls, toys, and others. For the electrician, it is important to know that the frequency of the signal produced by the UJT depends on R and C and can be calculated with a certain degree of precision by the formula:

$$f = \frac{1}{RxCx\ln\left[\frac{1}{1-\eta}\right]}$$

Where

f is the frequency in Hertz (Hz)

R is the resistance in ohms (Ω)

C is the capacitance in Farads (F)

1n is the natural logarithm

η is the intrinsic standoff ratio (about 0.6 for the 2N2646)

How to test

Using a multimeter you can make two tests in a UJT:

1. The resistance between the two bases in any direction is between 4.5 and 12 k ohms for a common UJT.
2. Placing the probes between the emitter and Base1, a high resistance can be measured in one direction and a low resistance in the opposite direction. This means that the junction (acting as a diode) is good.

The best way to test this device, however, is to mount a simple probe circuit like the one shown in Figure 121.

If the transistor is okay, the LED will flash at a rate that can be adjusted by the potentiometer.

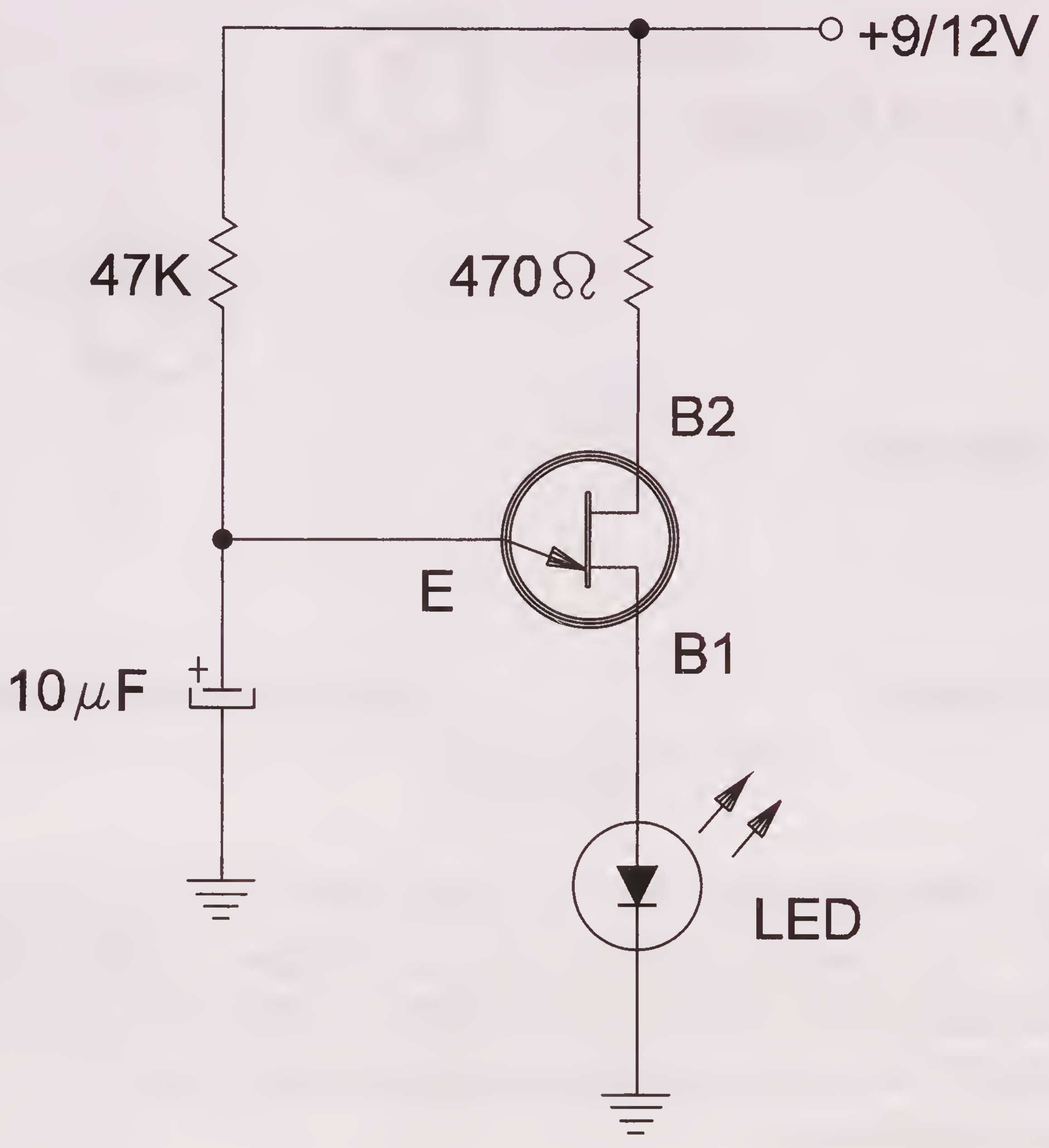

Figure 121

Programmable Unijunction Transistor—PUT

Although this device operates the same way as a UJT, its structure is different. It is a four-layer semiconductor placed in the category of components called thyristors. The structure of a PUT is like the one shown in Figure 122.

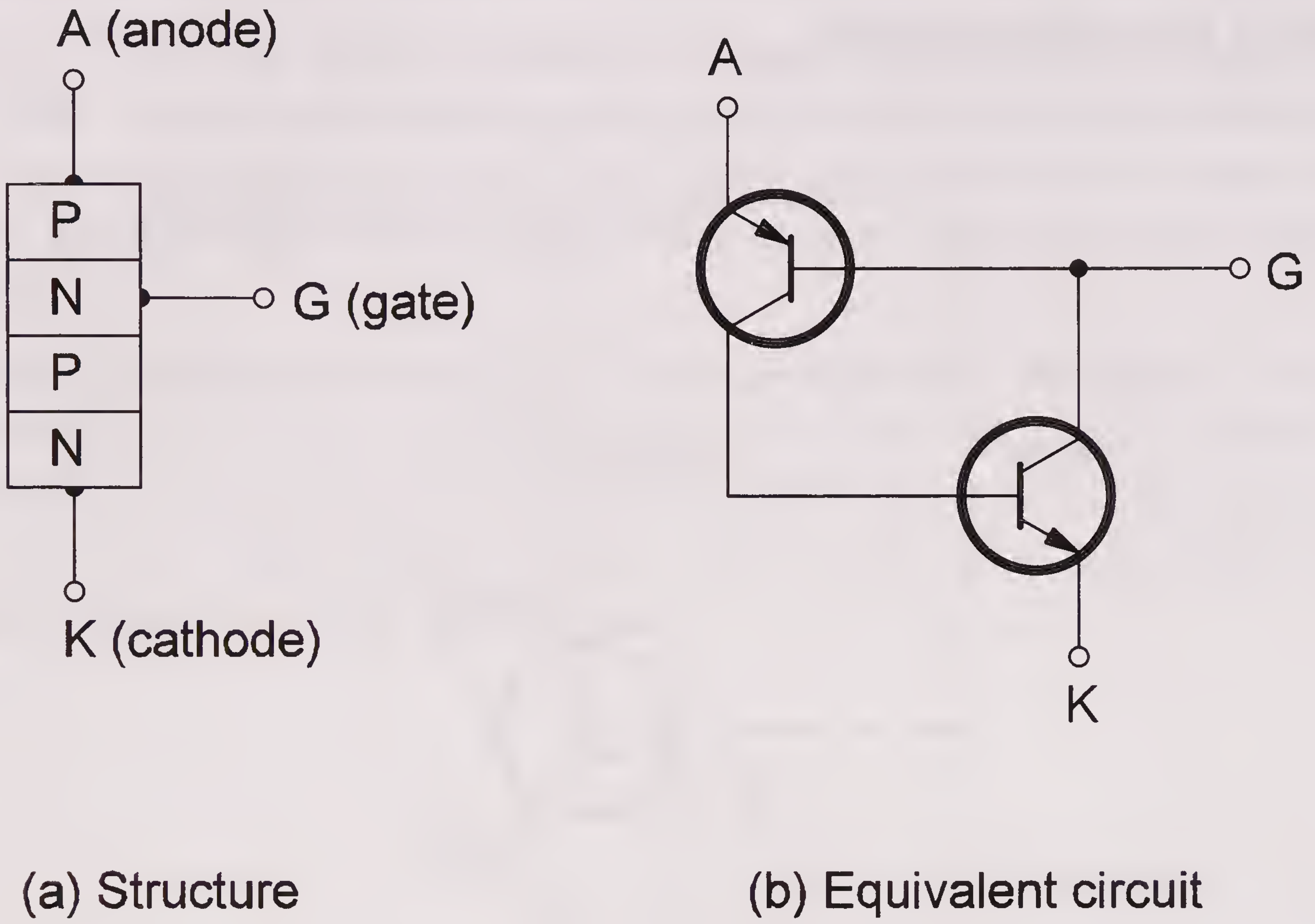
A (anode)
P
N
P
N
G (gate)
K (cathode)
A
G
K
(a) Structure
(b) Equivalent circuit

Figure 122

The basic difference between the UJT and the PUT is that the trigger point of a PUT can be programmed by a voltage divider placed in the anode terminal as shown in the circuit of Figure 123. The PUT is used the same way the UJT is—as a low-frequency oscillator and timer. It is found in the same applications.

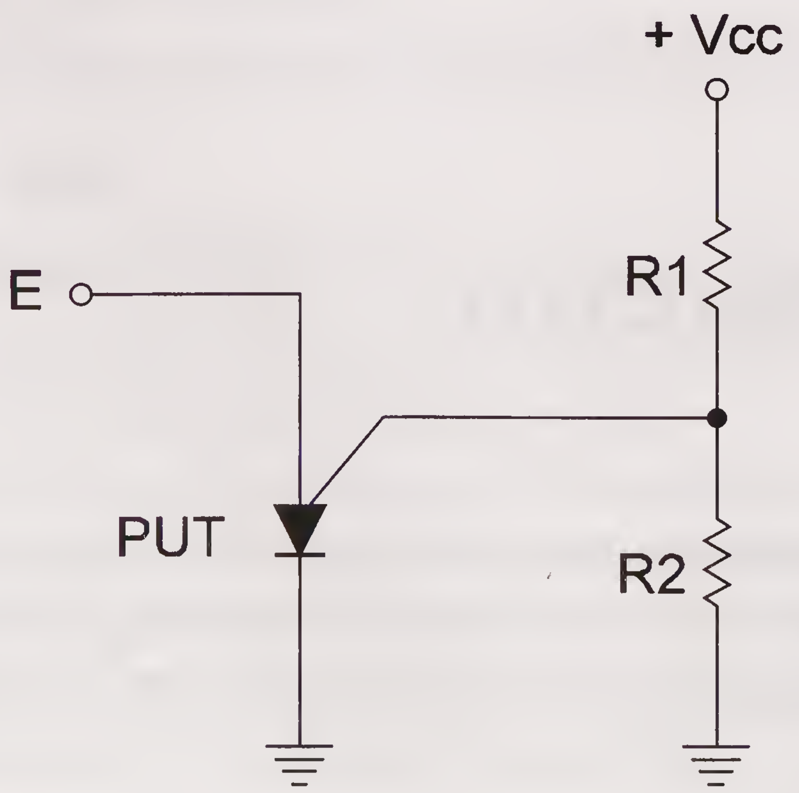
+ Vcc
E
R1
PUT
R2

Figure 123

Symbol

The symbol of a PUT is shown in Figure 124.

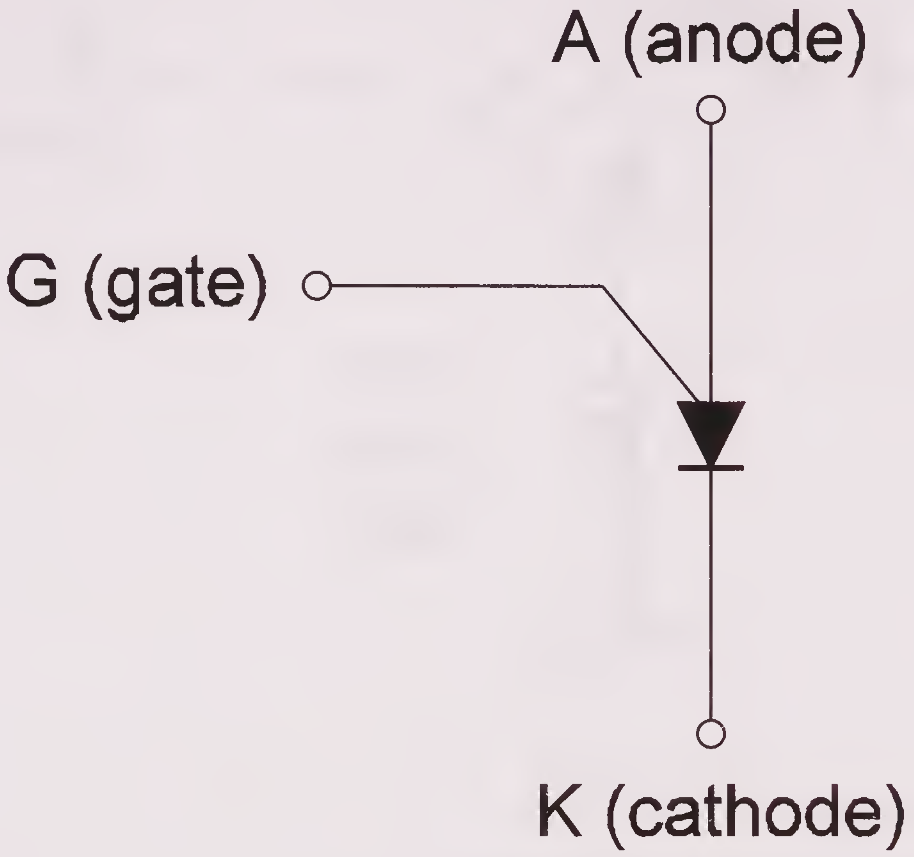

Figure 124

Specifications

The PUT is identified by a part number but there are electrical specifications that can be found in data sheets as the maximum voltage between anode and cathode, and the maximum operation frequency are all similar to the UJT.

Where they are found

The PUT is not a common device in modern applications, but the electrician working with electronics will find it in the same equipment where UJTs are found—timers, motor controls, audio oscillators in alarms, etc. The use of the PUT as a discrete device is not common because manufacturers use the same function in complex integrated circuits that include other functions. Knowing how it works is important because, although the device is not common, the function is.

Testing

The best way to test a PUT is using a circuit where it can drive an LED (Figure 125). If the PUT is okay the LED will flash.

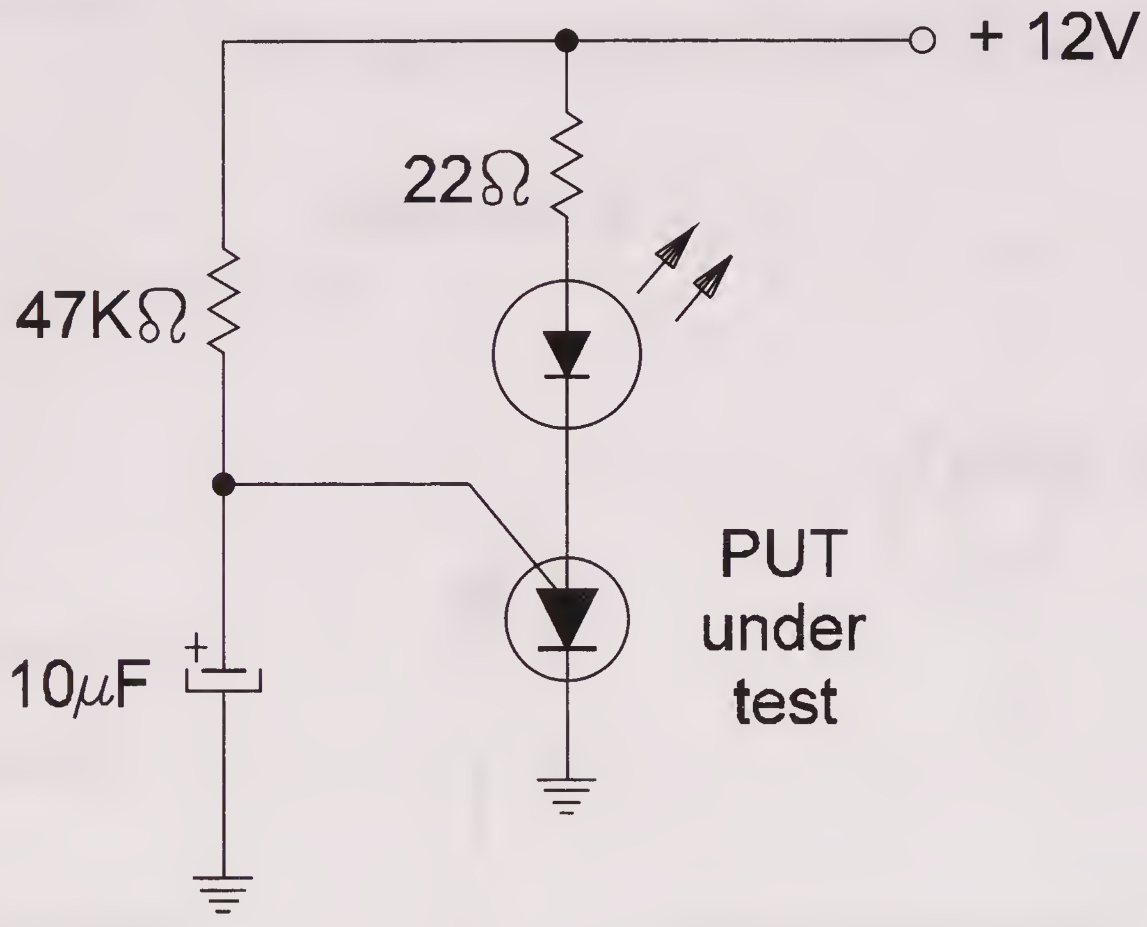

Figure 125

Field-Effect Transistor—FET

The Field-Effect Transistor (FET) is a semiconductor device that can be used in basically the same functions as the bipolar transistor: amplification, switching, or as an oscillator.

The basic FET is formed by a piece of N-type semiconductor material (silicon) where two P regions are formed (Figure 126).

The current through the channel of ~~P~~ N material, where the drain (d) and source (s) terminals are placed, can be controlled by a voltage applied to the gate (g) region. The basic difference between a bipolar transistor and the FET is that, unlike the bipolar transistor that controls a collector-emitter current from a base current, the FET controls a current between drain and source from a gate voltage. The bipolar transistor is said to be a typical current amplifier and the FET is a typical voltage amplifier.

Like the basic material where the drain and source are placed (N and P), the FETs can be found in two types: N channel and P channel. This

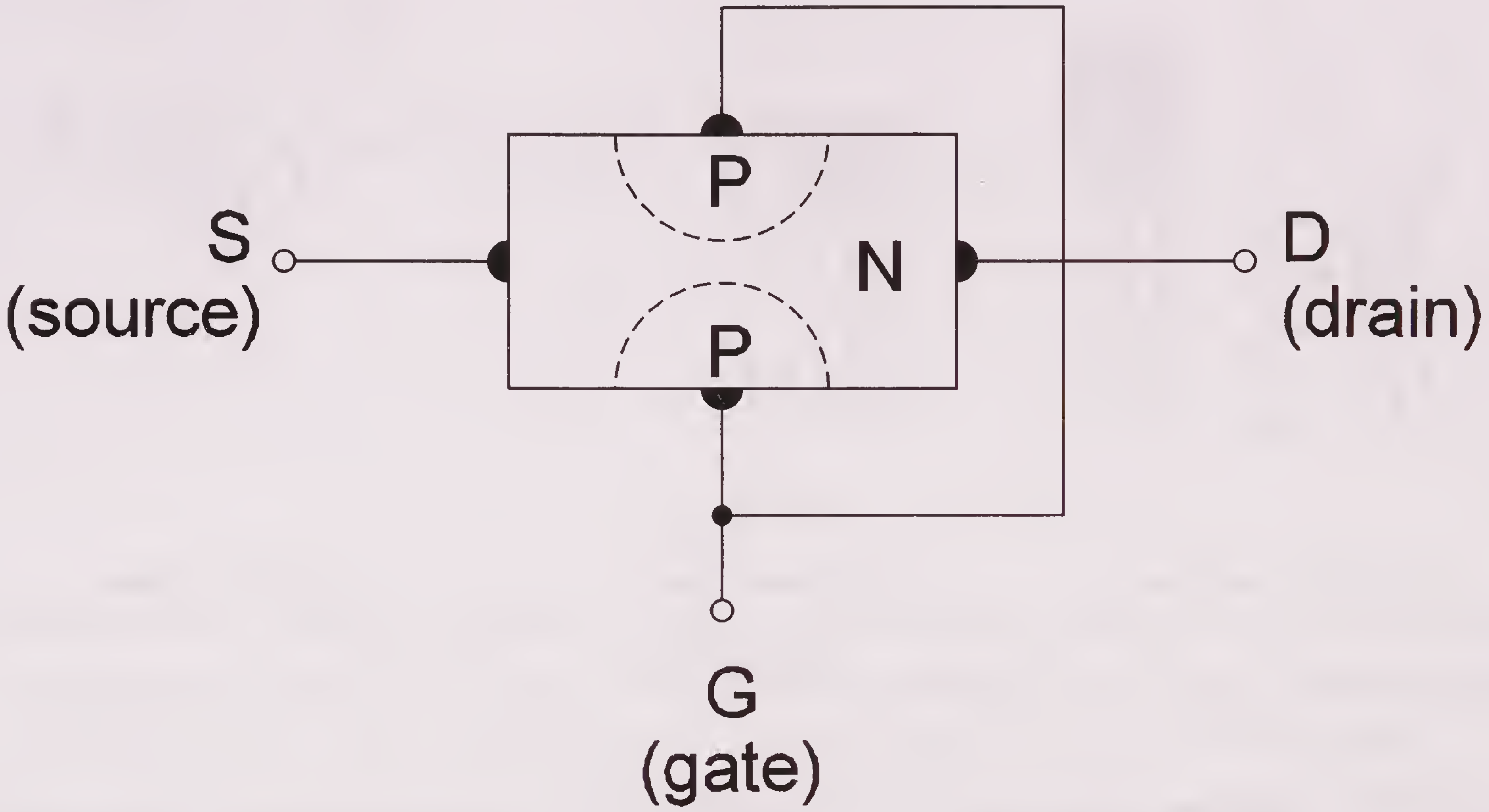
P
S
(source)
N
D
(drain)
P
G
(gate)

Figure 126

basic FET is referred to as JFET or Junction FET because, between the gate and the channel where the drain and the source are placed, exists a PN junction.

In the basic application the FET is wired as shown in Figure 127.

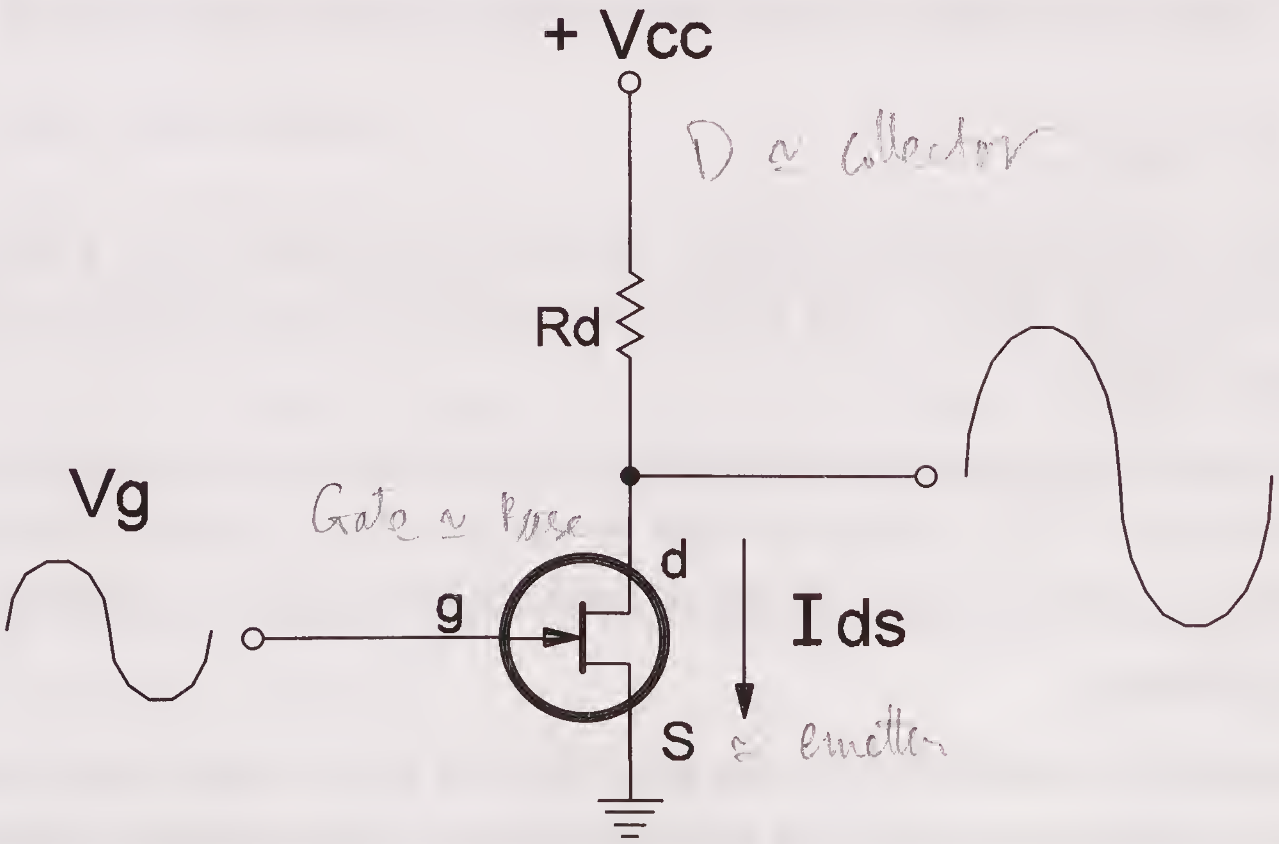
+ Vcc
D ≈ collector
Rd
Vg
Gate ≈ Base
d
g
Ids
S ≈ emitter

Figure 127

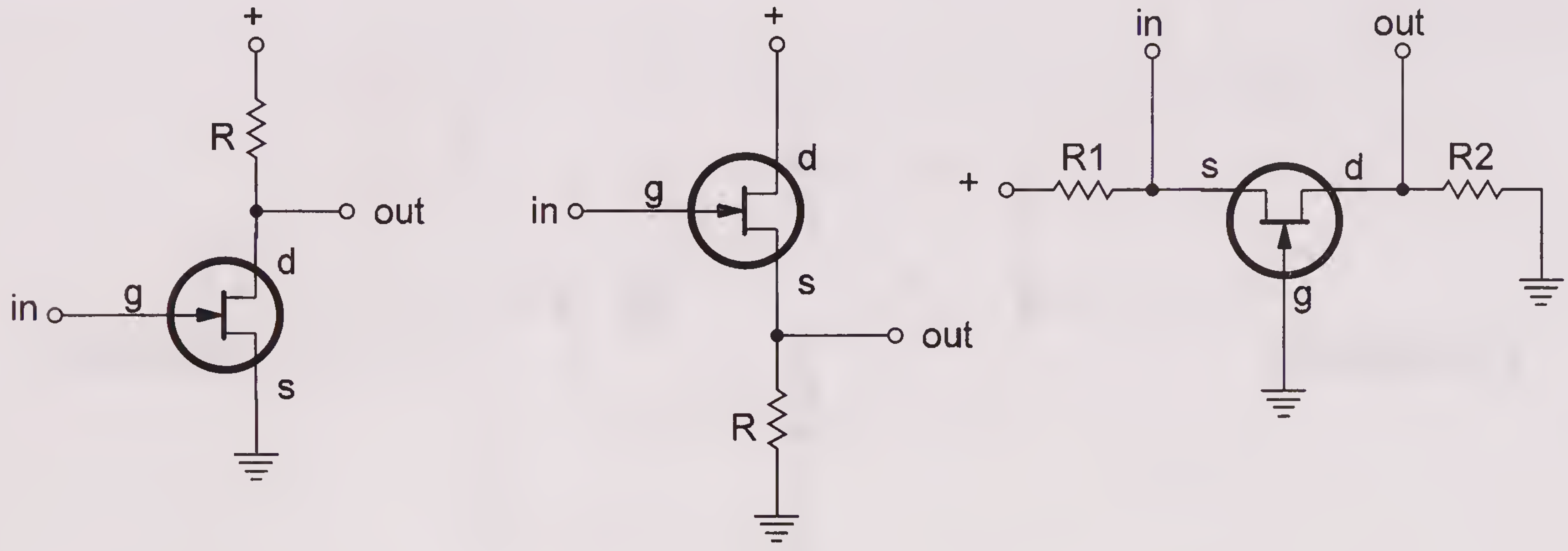

a) common source b) common drain c) common gate

Figure 128

The voltage changes in the gate will be transformed into current changes between drain and source. If a signal is applied to the gate, this signal can be taken already amplified from the drain or source of the transistor according to the configurations. The FET can be used in the same configurations as the ones found in bipolar transistor circuits. The common drain, common source, and common gate configurations are shown in Figure 128.

***Warning:** FETs are very sensitive devices. The electric charge of your body can destroy the device if you touch its terminals.*

Symbols and Types

Figure 129 shows the symbols used to represent both the P-channel and the N-channel JFET. The common types of these devices are shown in the same figure.

Observe that no external differences exist between a common bipolar transistor and an FET. The only way to know which one is which is by the part number or referring to the schematic diagram of the device.

Specifications

The electric specifications for an FET are important to help the electrician find a replacement type or to know if it can be used in a project. The main specifications for an FET are:

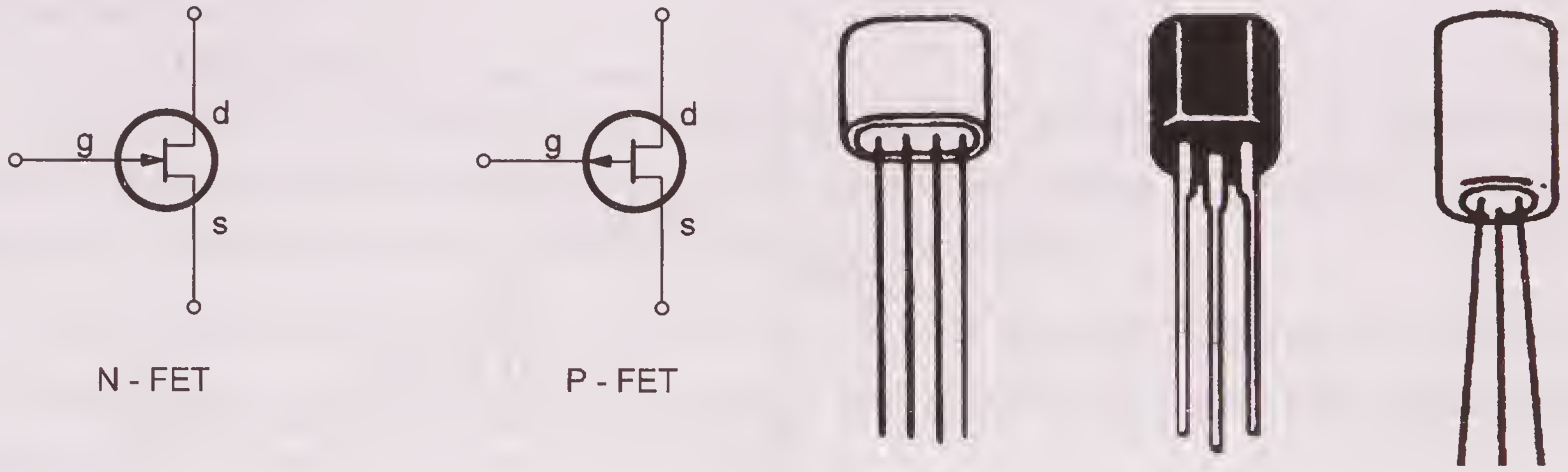

Figure 129

A. Maximum voltage between drain and source (Vds)—the maximum voltage that can be applied to the FET, typically in a range between 20 and 60 volts. This specification can be abbreviated as Vds (max).

B. Maximum drain current—the maximum current that can be controlled by a FET. A common abbreviation in the databooks is Id.

C. Transconductance—the equivalent of the gain of bipolar transistors. It measures the inverse of the resistance between drain and source under certain conditions. The transconductance is measured in Siemens (S). In old publications, the antiquated unit mho (the word "ohm" inverted) is used.

D. Dissipation power—the same as in bipolar transistors and is specified in watts. Small FETs have dissipation powers in a range from 50 to 500 mW.

Where they are found

The JFET or FET is found in circuits where signals must be generated or amplified. Audio circuits and sensor circuits in alarms can use an FET in their input. They can be used in the linear mode as amplifiers or switches like the common bipolar transistor. Figure 130 shows a typical stage where an FET is used as an amplifier for signals from a microphone or any other transducer.

Stages like this can be found in commercial equipment such as alarms, radios, and remote controls.

When replacing an FET it is important to know the type and, if a replacement is used, the placement of the terminals as they can change from type to type.

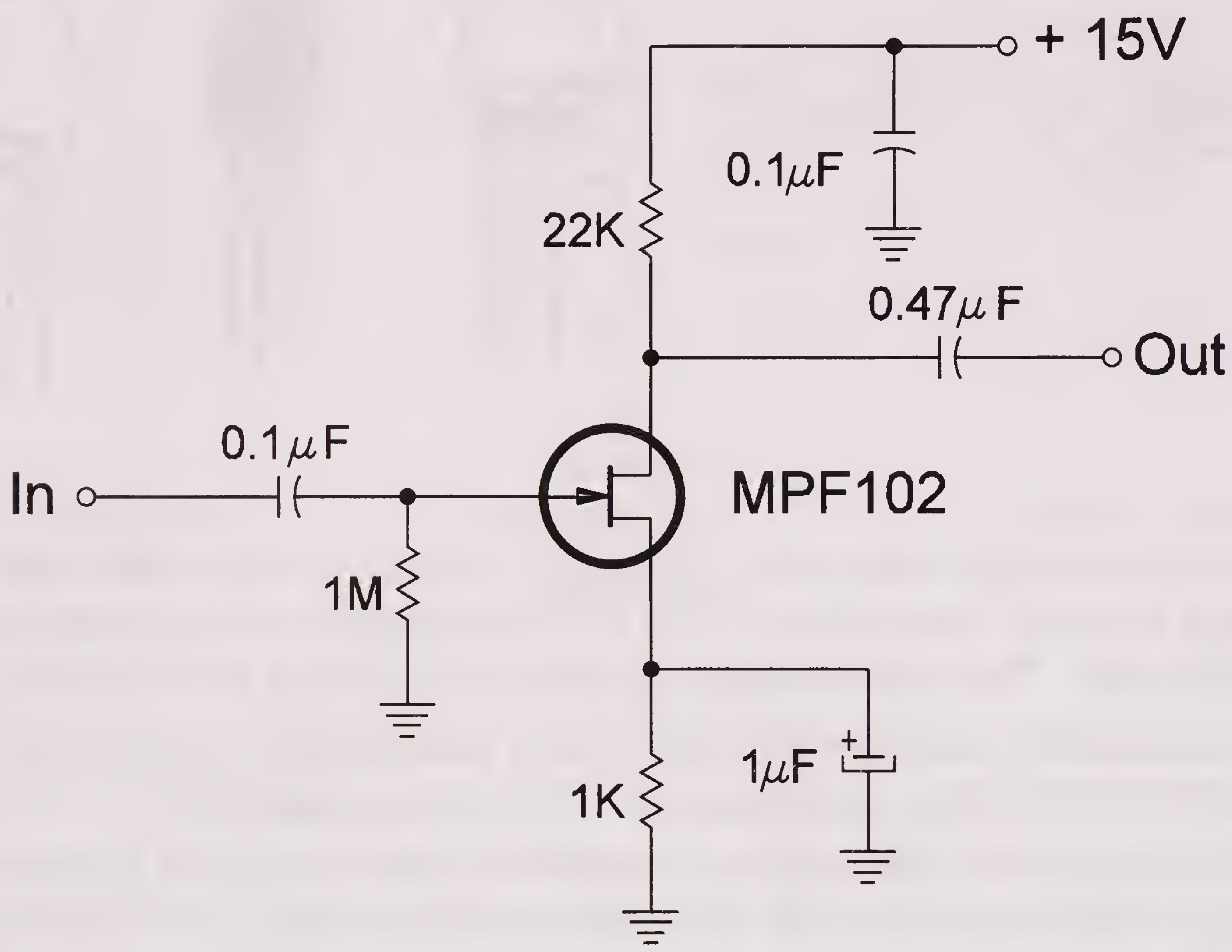
+ 15V
0.1μF
22K
0.47μ F
Out
0.1μF
In
MPF102
1M
1μF
+
1K

Figure 130

Testing

The ideal for testing an FET is a special meter dedicated to this task. But, even using a simple multimeter, it is possible to detect if an FET has problems. Figure 131 shows how to test an FET measuring the resistance between its terminals.

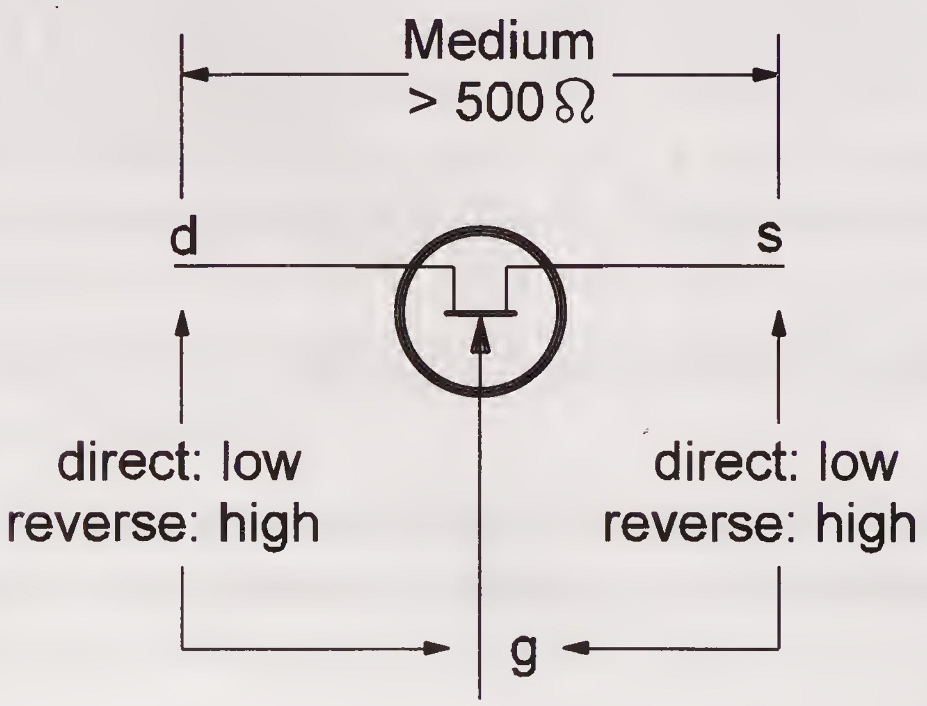
Medium
> 500 Ω
d
s
direct: low
reverse: high
direct: low
reverse: high
g

Figure 131

MOSFETS

MOSFET or Metal-Oxide Semiconductor Field-Effect Transistor is a device derived from the common FET but with some changes in the basic structure and also in the electrical characteristics.

As shown by the structure in Figure 132, MOSFETs have a thin layer of metal oxide (silicon oxide) isolating the substrate from the gate region, instead of a junction found in the JFET.

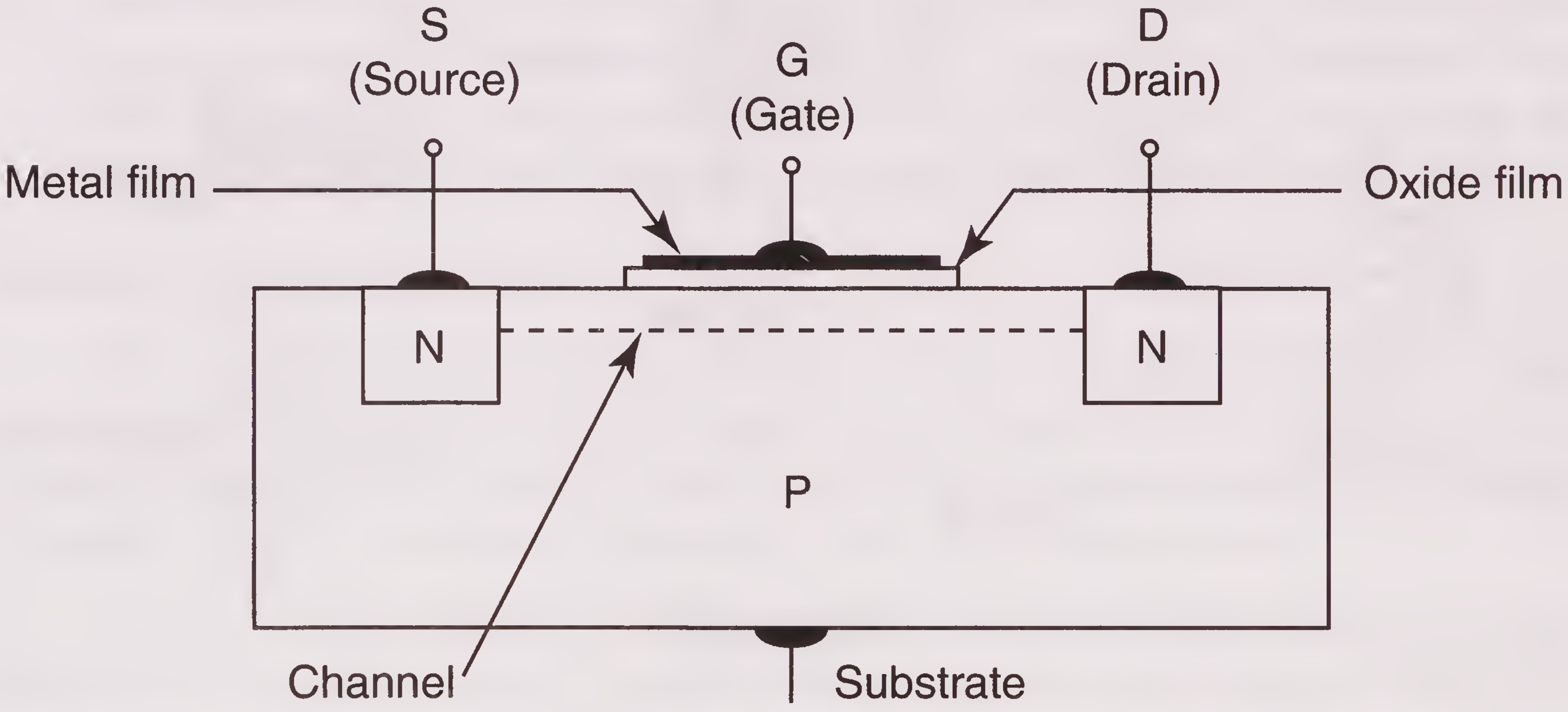

Figure 132

Although there are differences in the structure, the operation of a MOSFET is the same. A voltage applied in the gate can control the amount of current across the device or the current between drain and source. The MOSFET can be used in the same applications where the JFET is found with some advantages. It can be used in signal amplification, signal generation, and many other functions.

Warning: The layer that isolates the gate region from the channel is very thin, making the device extremely sensitive to static discharges. If you have a static charge in your body, touching any terminal of the device can produce a spark between the gate and channel, destroying the isolating layer and the device.

Symbols and Types

Figure 133 shows the symbols of the two types of MOSFETs commonly found in electronic circuits. This figure also shows types of these devices.

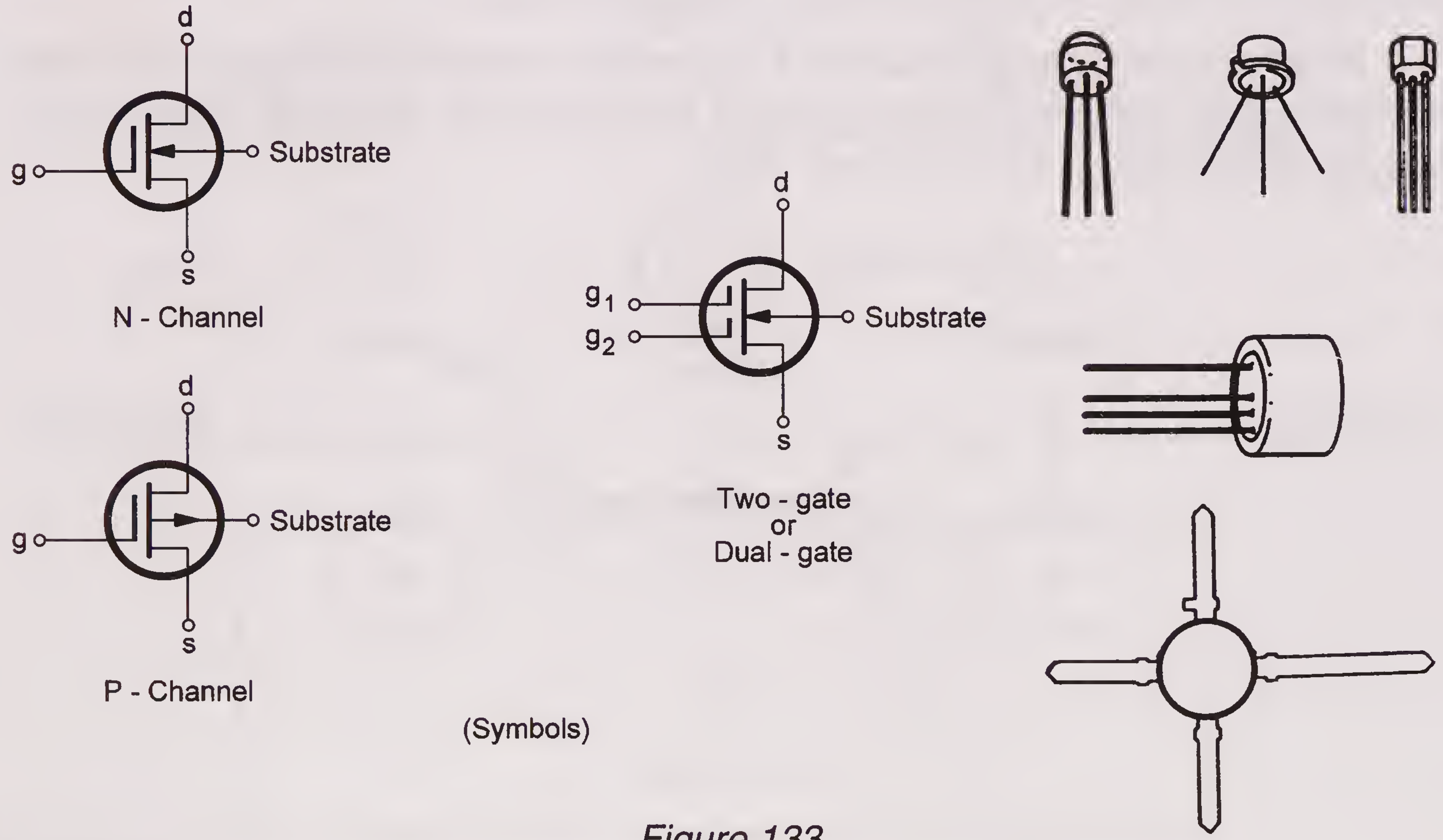

Figure 133

As we can see by this figure, the MOSFET can also be found in a two-gate version (g1 and g2). The amount of current across the device can be controlled independently by the voltages applied to each of the two gates. This device is suitable for applications where two signals must be mixed.

The common MOSFETs are low-power devices and look the same, making it difficult for an electrician to identify them from bipolar transistors. It is necessary to have the schematic diagram, a data sheet with information, or the type (part number on its body).

Specifications

A MOSFET is identified by a part number. By knowing this type number, consulting data sheets, or consulting handbooks it is possible to know its electric characteristics. These characteristics are important in order to find a replacement type. The main characteristics of a MOSFET are as follows:

A. Maximum voltage between drain and source—indicated as Vds (max) or only Vds, this is the maximum voltage that can be applied to the device without burning it. Values between 20 and 50 volts are common.

B. Maximum drain current—the maximum current controlled by the device. In common types this current ranges from 10 to 100 mA. It is also abbreviated by Id or Id (max).

C. Transconductance—in the common JFETs, it is the parameter that indicates the amplification capabilities of the device. It is measured in Siemens (S).

D. Dissipation Power (Pd)—as in any electronic device, the MOSFET converts electric energy into heat. The maximum amount of heat generated by the device and transferred to the area surrounding the device is the maximum dissipation power (Pd) given in watts (W).

Where they are found

The MOSFETs are mainly found in high- and low-frequency power circuits as in the input of audio amplifiers and as sensor amplifiers, or in high-frequency low-power circuits like RF, FI, mixer, and oscillator stages in radios, remote controls, receivers, TVs, walk-talkies, etc.

Today the functions of JFETs and MOSFETs can be included in other devices (integrated circuits) as we will see, but in some cases, depending on the equipment, the electrician may find these components.

Testing

The MOSFET can be tested by measuring the resistance between drain and source and also between the gates and the source (Figure 134). However, the best test can be made using a probe circuit that can easily be implemented on a solderless board (protoboard).

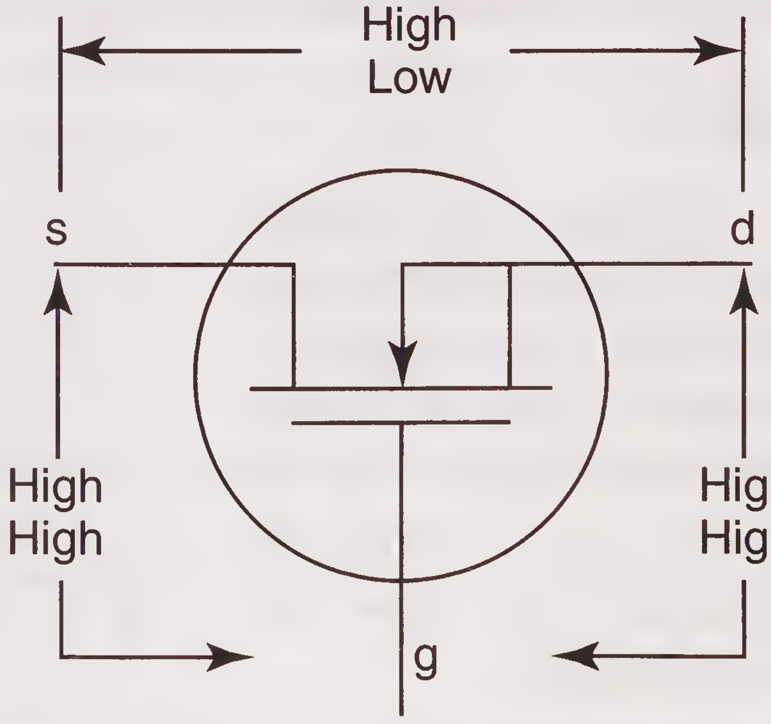

Figure 134

Power FETs

Power FETs are MOSFETs that can be used to control large amounts of current. They are very common today in digital and analog applications replacing the bipolar transistor. Power FETs, or power MOSFETs as they also called, can control currents as large as 50 amperes under voltages of more than 500 volts.

The structure and the operation principle of a power MOSFET are the same as the common MOSFET. The differences are in the size and shape of the silicon chip used when the device is manufactured.

Several technologies are used to produce these devices giving to the POWER-FET transistor names such as VFETs, power FETs, DFETs, and TMOS. The power FETs operate the same way the common MOSFET operates. A voltage applied to the gate can control the amount of current flowing across the device.

High-current devices such as solenoids, motors, electromagnets, relays, heaters, and lamps can be controlled by power FETs in the same kind of stages as the ones that use power bipolar transistors or Darlington transistors. The main advantage found in the use of power FETs is the low resistance presented when conducting a current (called Rds or drain source resistance). Because the amount of heat produced by the device depends on this resistance, power FETs can control much larger currents than the equivalent bipolar transistor while generating less heat.

Symbol

Figure 135 shows that the power FETs can be found in two types—N and P channel. They look the same as a power bipolar transistor because they use the same bodies.

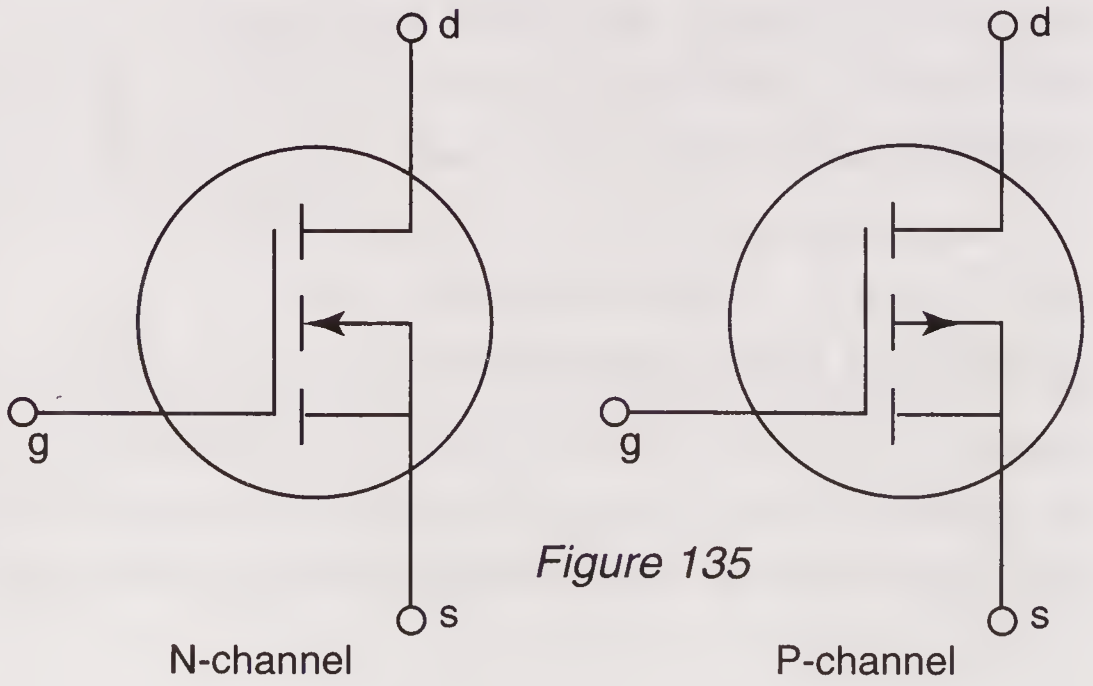

Figure 135

As with any other power transistor, the

power FETs must be mounted in heat sinks to transfer to the all the heat they generate when working.

Specifications

The power FETs are identified by a part number. A popular series of power MOSFETs and produced by many manufacturers including International Rectifier, Fairchild, and Motorola is the IRF series. All devices of this series have part designations beginning with IRF: IRF640, IR720, IRF630, IRF730, etc.

By knowing the name (part number), it is possible to find the electrical characteristics of a Power FET in handbooks or on data sheets. The principal characteristics are as follows.

A. Maximum voltage between drain and source—the maximum voltage that can be applied to the device without burning it. Common types have voltages ranging from 50 to 1000 volts. This specification can be indicated as Vds (max).

B. Maximum drain current—the highest current that can be controlled by the device and can range from 1 to 50 amperes. This specification is abbreviated by Id or Id (max).

C. Dissipation power—the amount of heat that the device can transfer to the surrounding area. It is measured in watts and can range from 10 to 250 watts.

D. Resistance between drain and source in the on state—the most important of the power FET's specifications. The amount of heat generated by the device depends on this resistance as it operates in more of the applications as an electronic switch. The product of this resistance multiplied by the square of the current gives the generated power. It is easy to see that the lower this resistance, the more efficient the device is when controlling high amounts of current and more heat is generated. This resistance used to be less than 1 ohm in more of the common devices. It is also indicated by rDS or Rds and is given in ohms (Ω).

Where they are found

Power FETs are slow devices used to control high current loads but in low- and medium-frequency circuits. Like the bipolar transistors, the power

FETs can be used as a switch (digital mode) or as a signal amplifier (linear mode).

As switches we can find the power FETs controlling solenoids, relays, step motors, and other loads in circuits like the one shown in Figure 136.

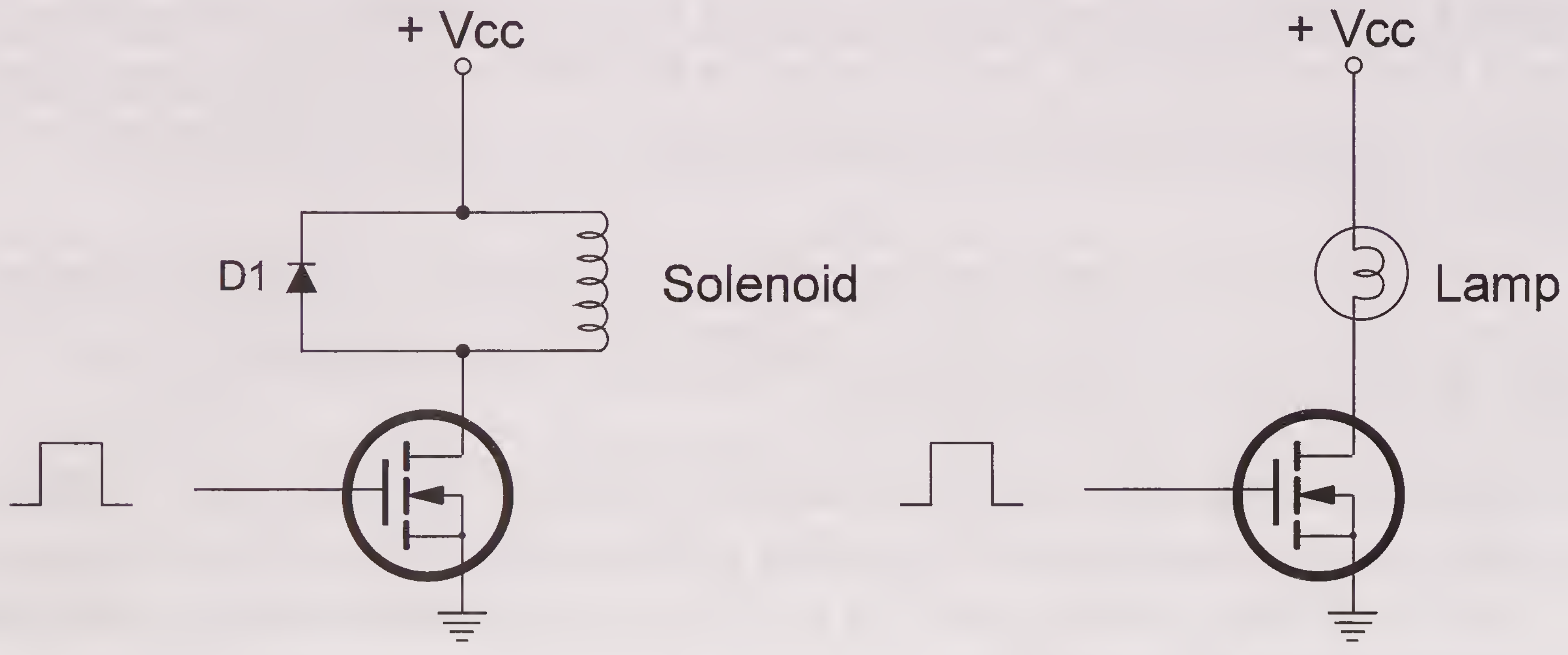
+ Vcc
D1
Solenoid
+ Vcc
Lamp

Figure 136

When a positive voltage is applied between gate and the ground, the power FET (P channel) turns on and the current flowing across the device powers on the load.

Another important application of power FETs is in Switch-Mode Power Supplies (SMPS) like the one shown in a block diagram in Figure 137.

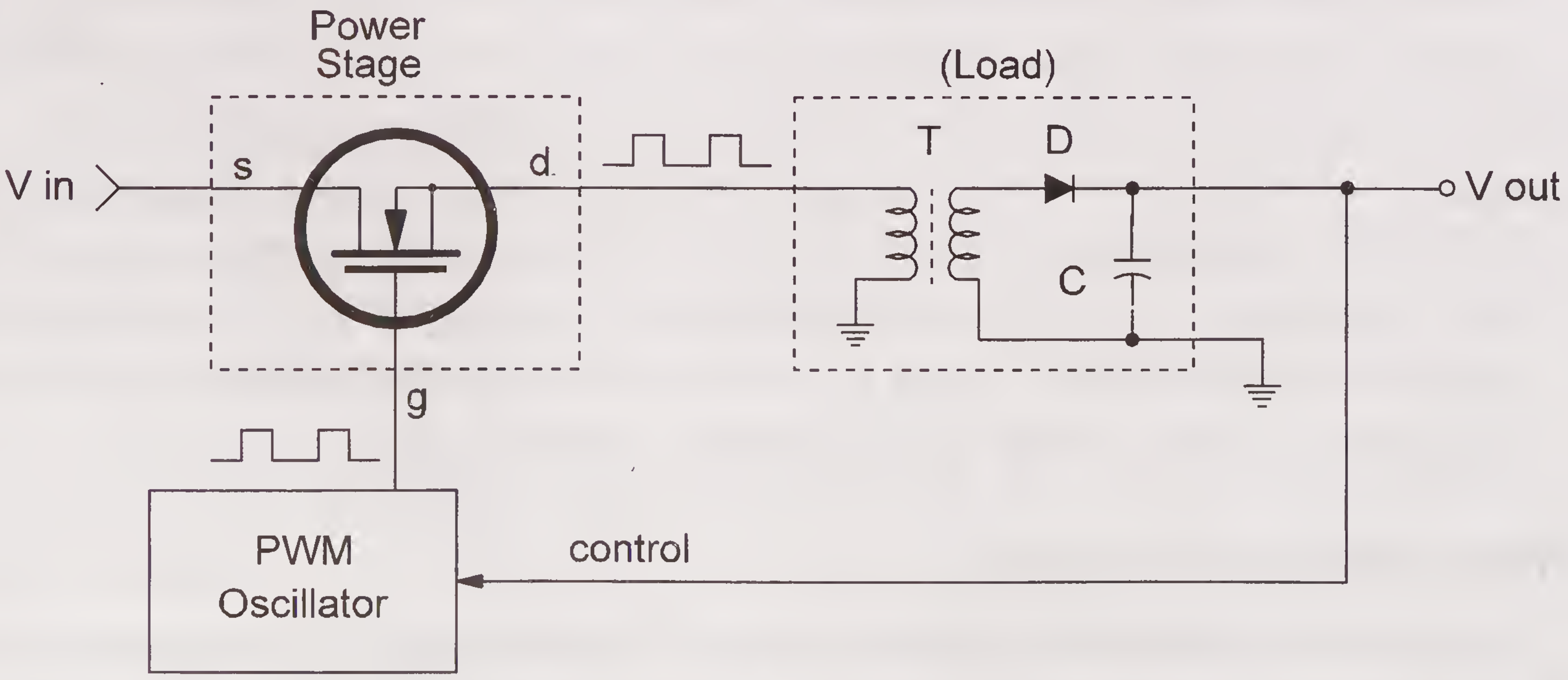
Power Stage
(Load)
V in
s
d
T
D
V out
C
g
PWM Oscillator
control

Figure 137

In this circuit the power FET acts as a switch that opens and closes very fast in a manner so that the pulses of voltage applied to the load can be controlled in frequency and duration.

If the interval or the width of the pulses changes, the amount of energy they transfer to the load also changes, and with them, the average voltage. By using a feedback circuit it is possible to control the width and interval between pulses to keep constant the average voltage applied to a load. So, in an SMPS exists a circuit that changes the voltage output, controlling the pulse width applied to the power FET to compensate the alterations and keep the output voltage constant. This kind of power supply is found in many types of electronic equipment, such as computers, video monitors, and TVs. In the linear mode, power FETs can be used as audio amplifiers or low-frequency signal amplifiers. A circuit where an audio signal is amplified by a power FET is shown in Figure 138.

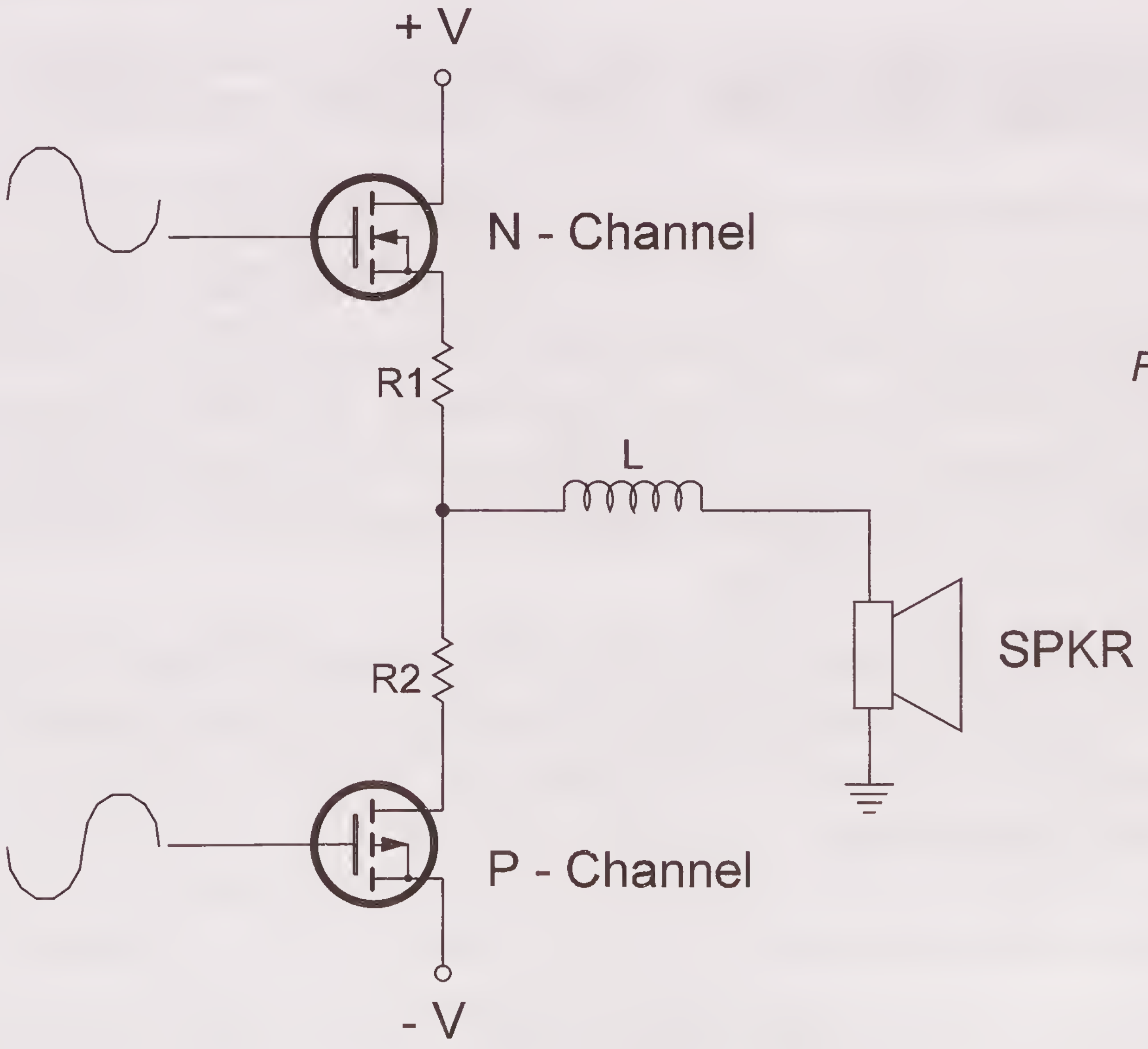

Figure 138

Alarms, which are a common type of audio equipment, can use this type of transistor in their output stages driving the loudspeaker with many watts.

High-fidelity (hi-fi) audio amplifiers that use power FETs are high-quality equipment presenting low distortion levels. The power FETs have characteristics similar to the ones found in the old tubes for amplifying low-frequency signals. Many enthusiasts of the bipolar transistor can't reproduce the sounds with the same quality found in old tube models. But, amplifiers using power FETs can present levels of fidelity closer to those found in the tube amplifiers better than the ones of bipolar transistors.

Testing

A simple multimeter can be used to test a power FET. The resistance between the gate and the other terminals must be very high in a good transistor. If a low resistance is measured the transistor has problems. A low resistance between gate and the other terminals indicates a transistor with losses or a short.

Isolated-Gate Bipolar Transistor—IGBT

The Isolated-Gate Bipolar Transistor or IGBT is a semiconductor device that combines the advantages of a bipolar transistor and a power FET. They are transistors where the main current or controlled current flows between a collector and an emitter as in a bipolar transistor, but is controlled by a gate voltage as in an FET. The advantage is that the device presents very high input impedance and the characteristics of commutation or control of a bipolar transistor.

IGBTs are used in the same functions where power bipolar transistors and power FETs are used, and in some cases they have advantages. They are used in applications where it is necessary to drive a high current load (solenoids, lamps, motors, etc).

Symbol

Figure 139 shows the symbol of an IGBT. As with other transistors, we can't identify an IGBT only by its appearance because it looks the same as bipolar transistors and power FETs. It is necessary have a part number or a schematic diagram of the equipment.

Specifications

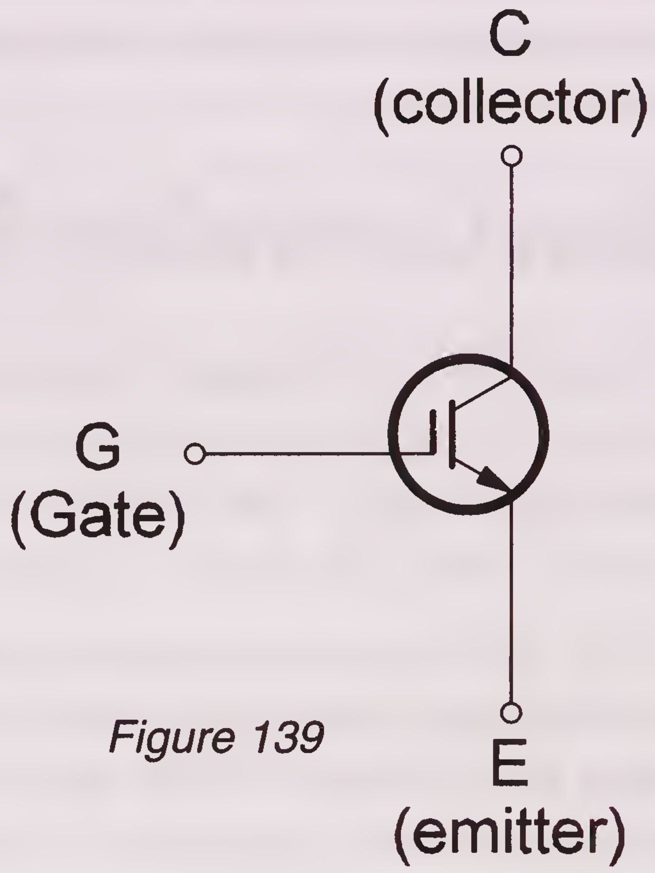

Figure 139

IGBTs are identified by a part number. From the part number, the electrical characteristics can be discerned. The main characteristics are as follows.

A. Maximum voltage between collector and emitter—the maximum voltage that can be applied to the device. It is also represented by Vce (max).

B. Maximum collector current—the maximum current that can be conducted by the device. It is also abbreviated by Ic (max).

C. Dissipation power—the maximum amount of power that the device can convert in heat and transfer to the surrounding area. It is given in watts and abbreviated as Pd (max) or Pd.

Where they are found

IGBTs are components widely found in industrial applications. The advantages of the use of this device in the control of high-current loads in industrial machines, such as motors, solenoids, electromagnets, and heaters, make them very commonly used in these types of equipment.

In many cases, the IGBT can directly replace a power FET or a Darlington bipolar transistor. The most the technician must do in many cases is change the value of some resistors in the circuit.

As IGBTs are new devices they are not very common in electric or electronic applications more than 10 years old. In home appliances, such as dishwashers, clothes washers, and others that use inductive loads, the IGBT can be found as the control element.

Testing

The simplest way to test an IGBT is measuring the resistance between the terminals. The resistance between the gate and the other elements in both directions must be very high. A low resistance in these measures indicates a shorted device or one with losses.

Silicon Controlled Rectifier—SCR

The SCR or Silicon Controlled Rectifier is one of the most important of the electronic devices the electrician will find in electric installations and appliances. In fact, the SCR is made to control devices plugged in to the AC power line in its basic configuration.

The SCR is a member of the thyristor family. Thyristors are four-layer semiconductor devices basically intended for applications of control and AC circuits. They are solid-state switches, can operate at voltages up to many hundreds of volts, and can handle currents up to hundreds of amperes.

As the SCR is very important to the electrician who wants to know how electronics devices are used in electric installations, more space than the usual dedicated to other components is given to their description and applications. So, what is an SCR, starting from its structure?

Figure 140 shows the structure of an SCR and the equivalent circuit formed by two complementary transistors. It is a silicon semiconductor with a PNPN structure.

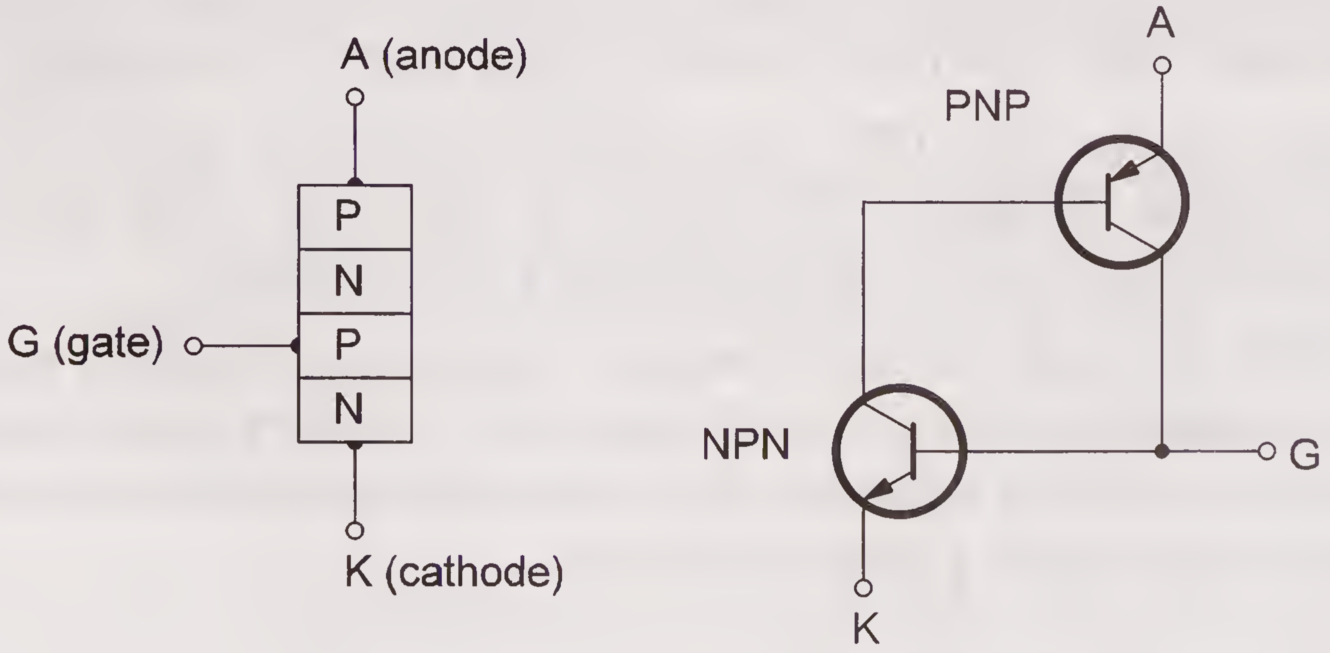

(a) Structure)

(b) Equivalent circuit

Figure 140

In the basic operation the anode (A) is made positive and the cathode (K or C) negative. The load is placed between anode and the power supply. If no voltage is applied to the gate (G) the two transistors remain off and no current can flow across the load. But, if enough voltage is applied to the gate, it can make the NPN transistor conductive and a current begins to flow between its collector and emitter.

This current is the one that comes from the base of the PNP transistor biasing on this component. The current that now begins to flow by the NPN transistor acts as a feedback to the NPN transistor as it also flows through its base. The feedback makes all the currents in the circuit increase speed and in a few microseconds the circuit reaches a saturated state. The current between the collector of the PNP transistor and the emitter of the NPN transistor is now at its maximum value. After completion of this very fast process, even if the initial voltage that started all this is no longer present, the device remains conductive.

In a real SCR the process is the same: a positive voltage applied in the gate triggers on the device that conducts the current from the anode to the cathode. Even if the voltage disappears once it is triggered, the device is kept in the on state. To trigger it off there are two possibilities:

1. Turn off the power supply for a moment. This makes the device current fall to zero.
2. Reducing the current flowing through the device to a value that cuts the feedback process. A short between anode and cathode, reducing the voltage to zero (Figure 141), can be created to trigger the SCR off.

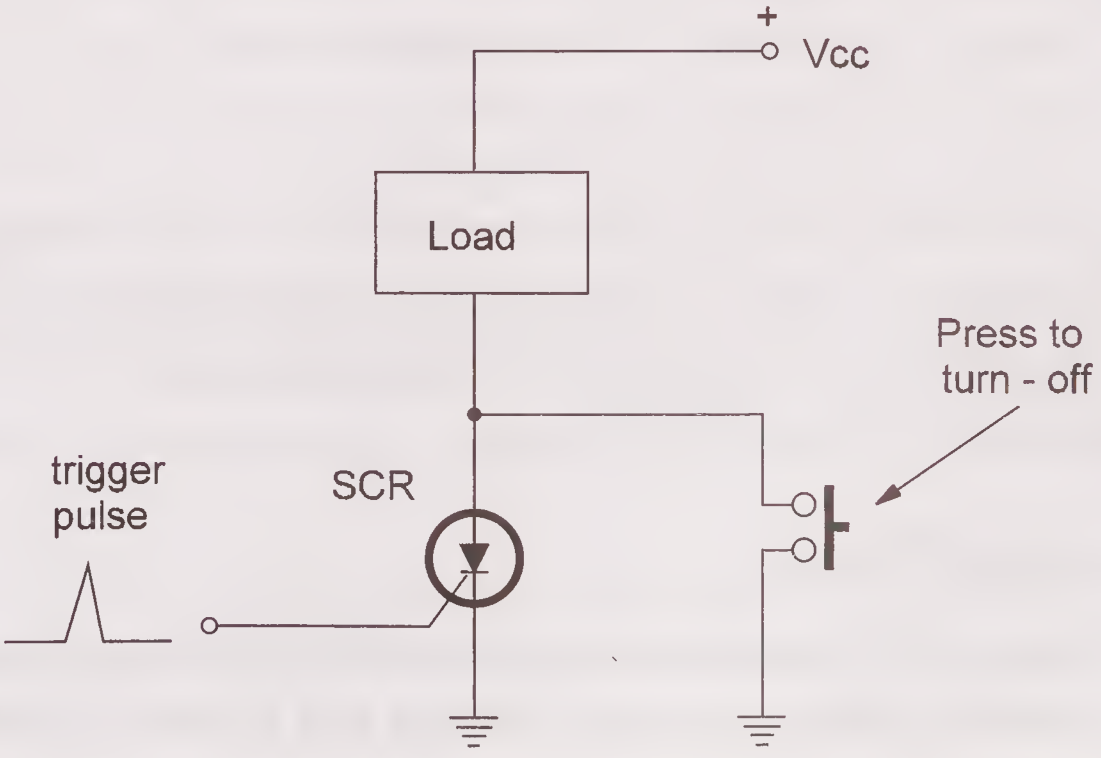

Figure 141

Notice that the SCR acts as a diode that can be made conductive or not conductive because the current flow only can flow in one direction. This explains its symbol—a diode with a gate.

Real SCRs can be triggered by voltages of about 1 volt and currents ranging from 100 microamperes for the most sensitive to some miliamperes for high current types and less sensitive types.

If placed in an AC circuit like the one shown in Figure 142, when turned on the SCR can conduct only the positive semicycles of the AC power line voltage.

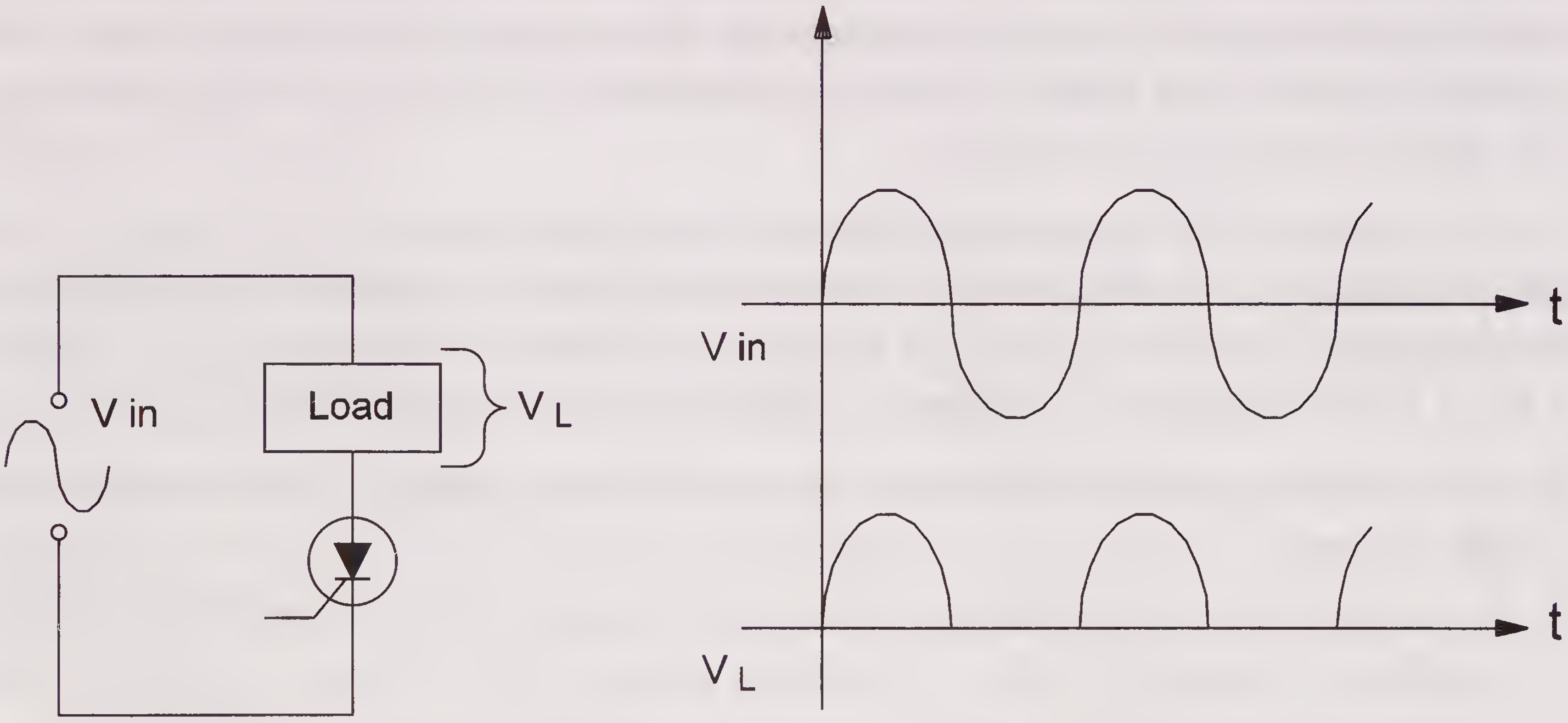

Figure 142

The SCRs can be used to control AC loads directly from the AC power line. The most common types can operate handling high voltages and high currents. This is because this device can be found in many applications related to the AC power line.

Symbol and Diagram

Figure 143 shows the symbol adopted to represent an SCR. The main types found in electronic equipment are shown in the same figure.

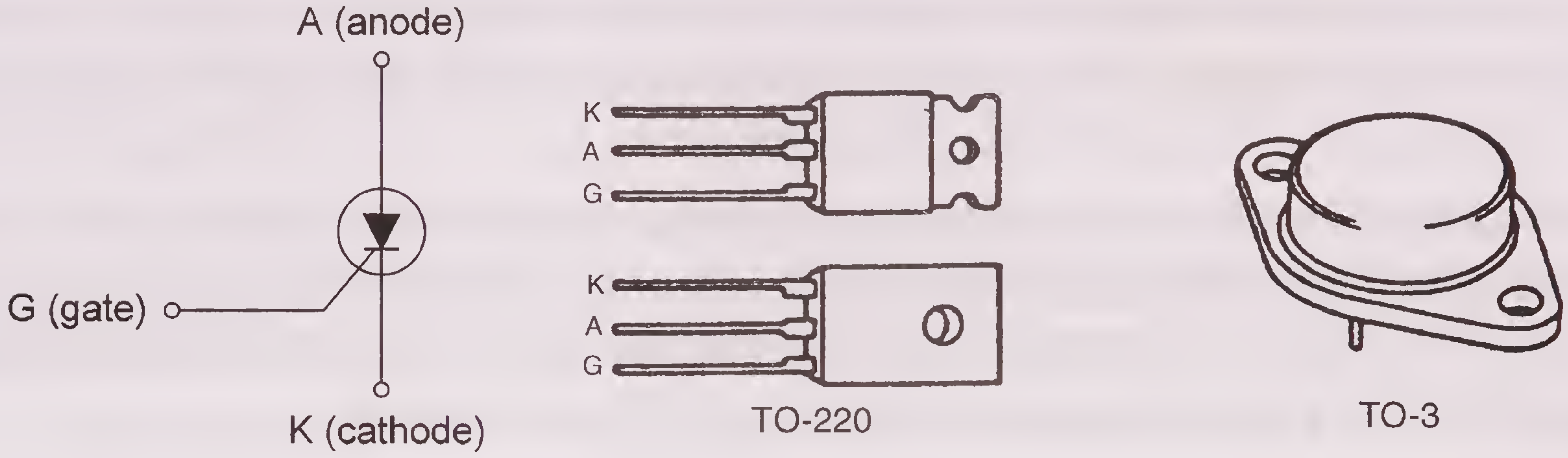

Figure 143

Although small SCRs designed to control low currents still exist, more of them are power devices including resources to place them on large heat sinks. The tab or the metal body of the SCRs used to be connected to the anode terminal.

Specifications

The SCRs are identified by a part number given by the manufacturer on the body. From this part number, looking in a handbook, in a databook, or on a data sheet it is possible to know the electrical characteristics.

When using any SCR it is important to know the following characteristics:

A. Maximum voltage between anode and cathode—the maximum voltage that can be applied to the SCR when it is in the off state. Typical values range from 50 to more than 1000 volts for common types. This specification can also be indicated as Repetitive Peak Off voltage and abbreviated by Vdrm or as Repetitive Peak Reverse Voltage and abbreviated by Vrrm.

B. Maximum current—the maximum current that the SCR can hold when in the conductive state. Common SCRs found in AC-powered appliances are specified to control currents in the range from a fraction of an ampere to more than 50 amperes.

C. Dissipation power—When an SCR is on, independently of the current it is conducting, the voltage between anode and cathode is about 2 volts. The product of this voltage by the current gives the dissipation of the

SCR. The maximum dissipation of an SCR is important to know in order to calculate the size and shape of the heat sink.

D. Holding current—the lowest current that the SCR can conduct without turning off. This current is abbreviated by I_H.

Other important specifications are the speed of the device given by the value referred to as Critical Rate of Rise of Off State Voltage and measured in volts by microseconds. This value shows how fast the voltage applied to the load increases when the SCR is turned on.

Where they are found

SCRs are found in several applications involving AC power control. Some devices where SCRs can be found are motor controls, timers, dimmers, alarms, solenoid drivers, and many home appliances.

The most common application of an SCR as a power control is the one shown in Figure 144.

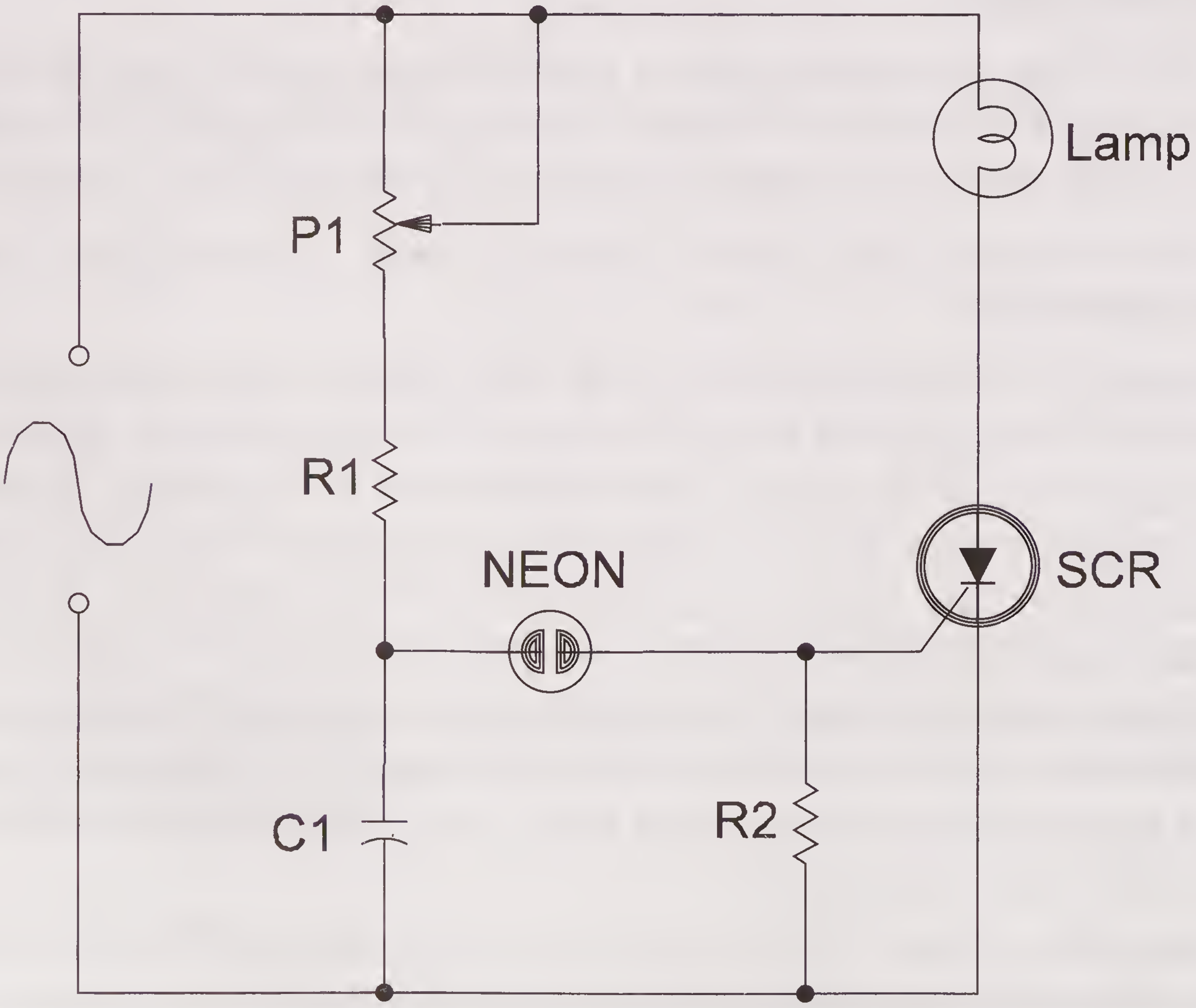

Figure 144

This circuit is found in commercial dimmers like those installed to replace common on/off switches. It is also used to control the brightness of incandescent lamps, in fans as a speed control, and in heaters as a temperature control.

It is very important for an electrician to know how this circuit works because its operation principle is applied to a large number of common electric appliances that have a power (speed, temperature, etc) control.

Look at the waveform of the AC power line voltage shown in Figure 145.

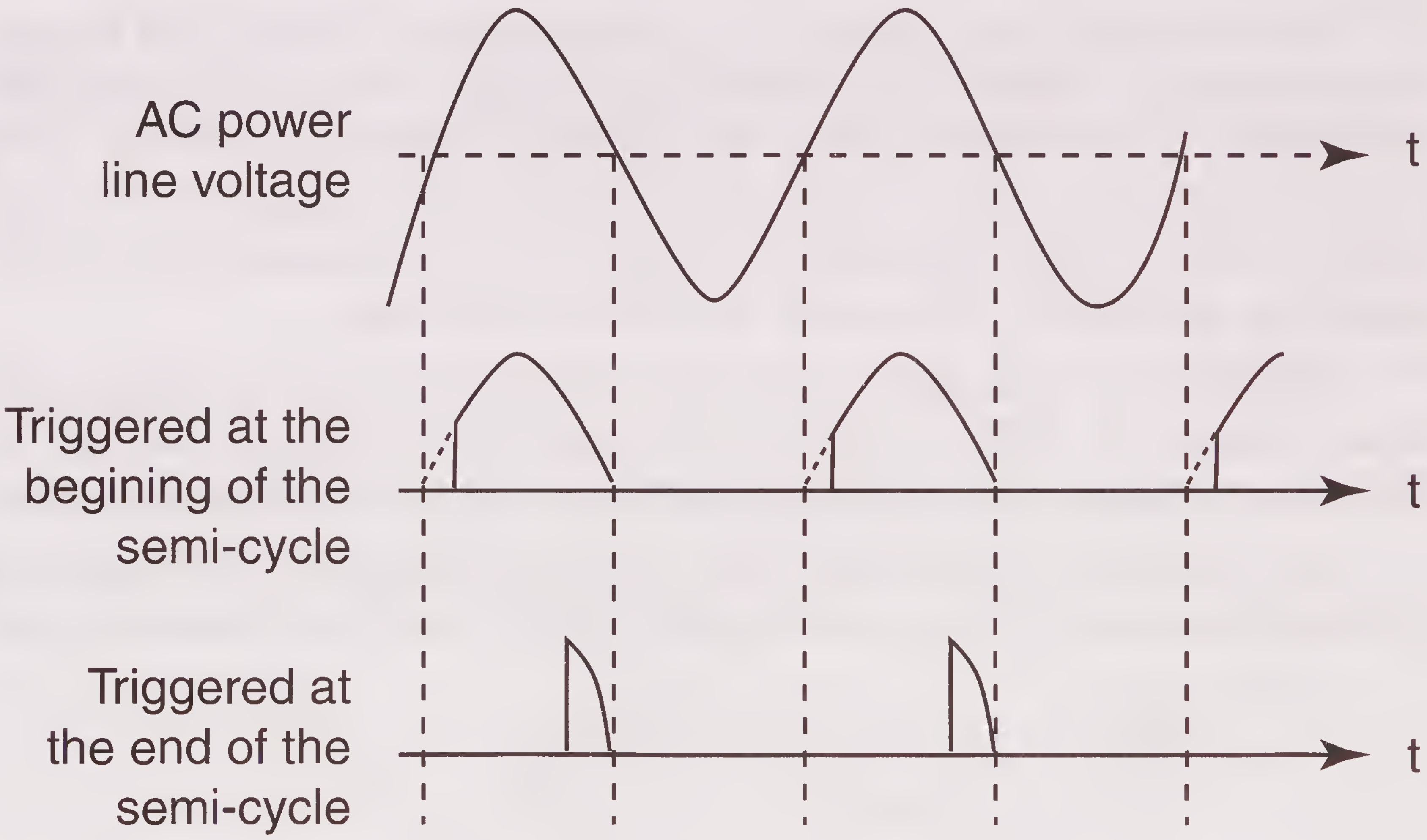
AC power line voltage
t
Triggered at the begining of the semi-cycle
t
Triggered at the end of the semi-cycle
t

Figure 145

The voltage changes make the current flow forward and backward across a load determining the amount of power absorbed by it. This can be reduced if we cut the semicycles of the AC power line voltage in different points.

For example, if the semicycles are cut at the end, only a small part of the AC power line voltage passes and is applied to the load. If it is a lamp it will present a low brightness. If it is a motor it runs at low speed. If it is a heater the temperature is lower. If the semicycle is now cut at the beginning, a larger part of the power passes and is applied to the load. If it is a lamp, the brightness is increased, a motor runs at a higher speed, and a heater is warmer.

Using an appropriate control, we can slide the point where the semicycles are cut from near zero to 100 percent and thus control, with a great precision, the amount of power applied to a load.

The circuit shown in Figure 144 makes this control. At the beginning of a semicycle of the AC power line the voltage applied to the capacitor C increases in speed depending on the value of the component and the value of the resistance adjusted in the potentiometer P1.

If the potentiometer is placed in a low-resistance position, the capacitor charges fast and the voltage necessary to trigger on the SCR is reached at the beginning of the semicycle. More power is applied to the load. If the potentiometer is now slid to a high-resistance position, the capacitor charges slowly and the voltage necessary to trigger the SCR is reached only at the end of the semicycle. Less power is applied to the load.

Notice that in both cases, when the semicycle ends, the voltage falls to zero letting the SCR turn off. As the capacitor discharges through the SCR when it triggers on, in the next semicycle, the process begins again.

By sliding the potentiometer from zero to 100 percent of its resistance it is possible to control the power applied to the load in a range that typi-

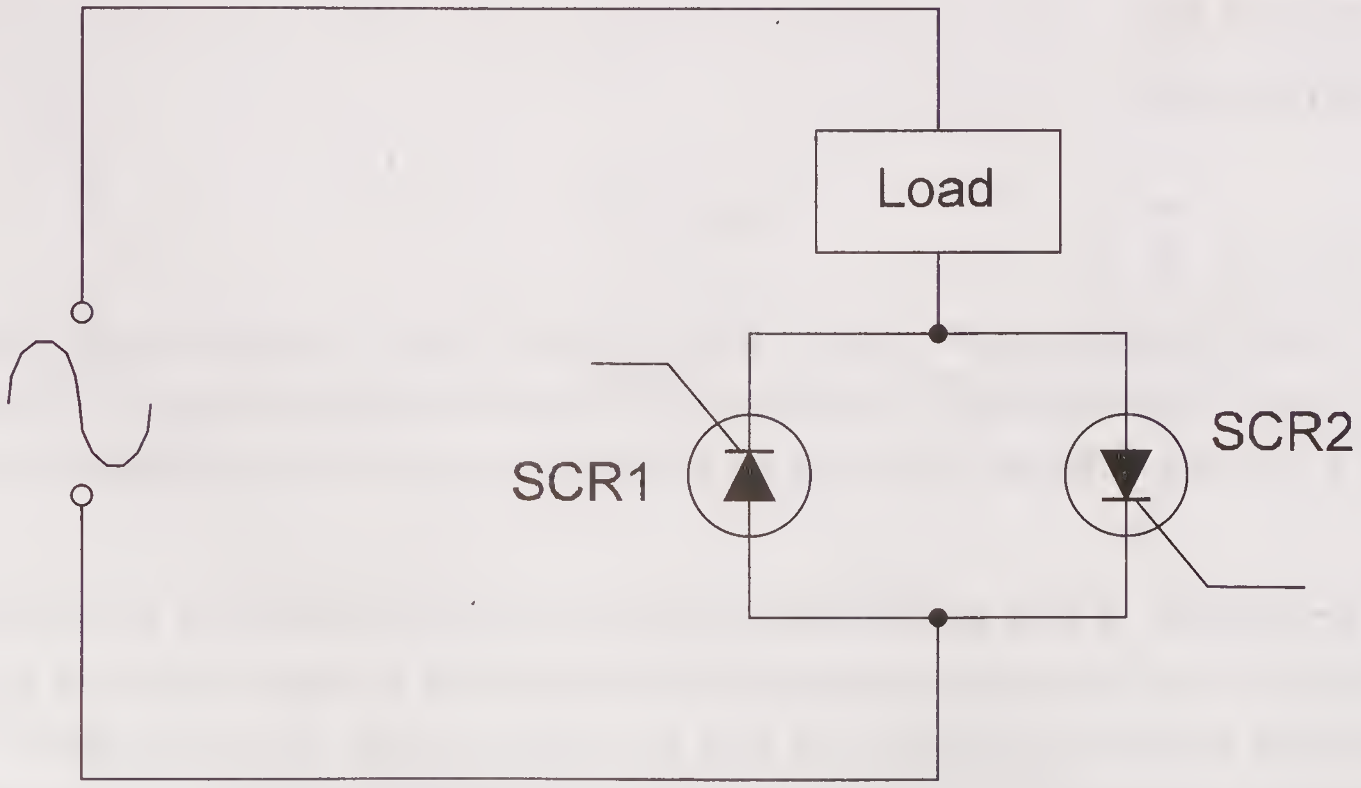

Figure 146

cally goes from near 0 to 95 percent of the maximum power. A little problem must be considered in this circuit: the SCR is a unilateral device and the AC loads must receive both semicycles to reach their normal maximum power. How can this problem be solved?

In some cases the circuit can operate only with the positive semicycles of the AC power supply line, and the SCR can be used. In other cases, two SCRs wired in an anti-parallel mode are used, as shown in Figure 146. One SCR conducts the positive semicycles and the other conducts the negative semicycles of the AC power line voltage.

Another solution to the problem is in the use of a diode bridge (rectifier) that converts the AC power line voltage into a DC pulsed voltage, or a voltage formed only by positive voltage semicycles as shown in Figure 147.

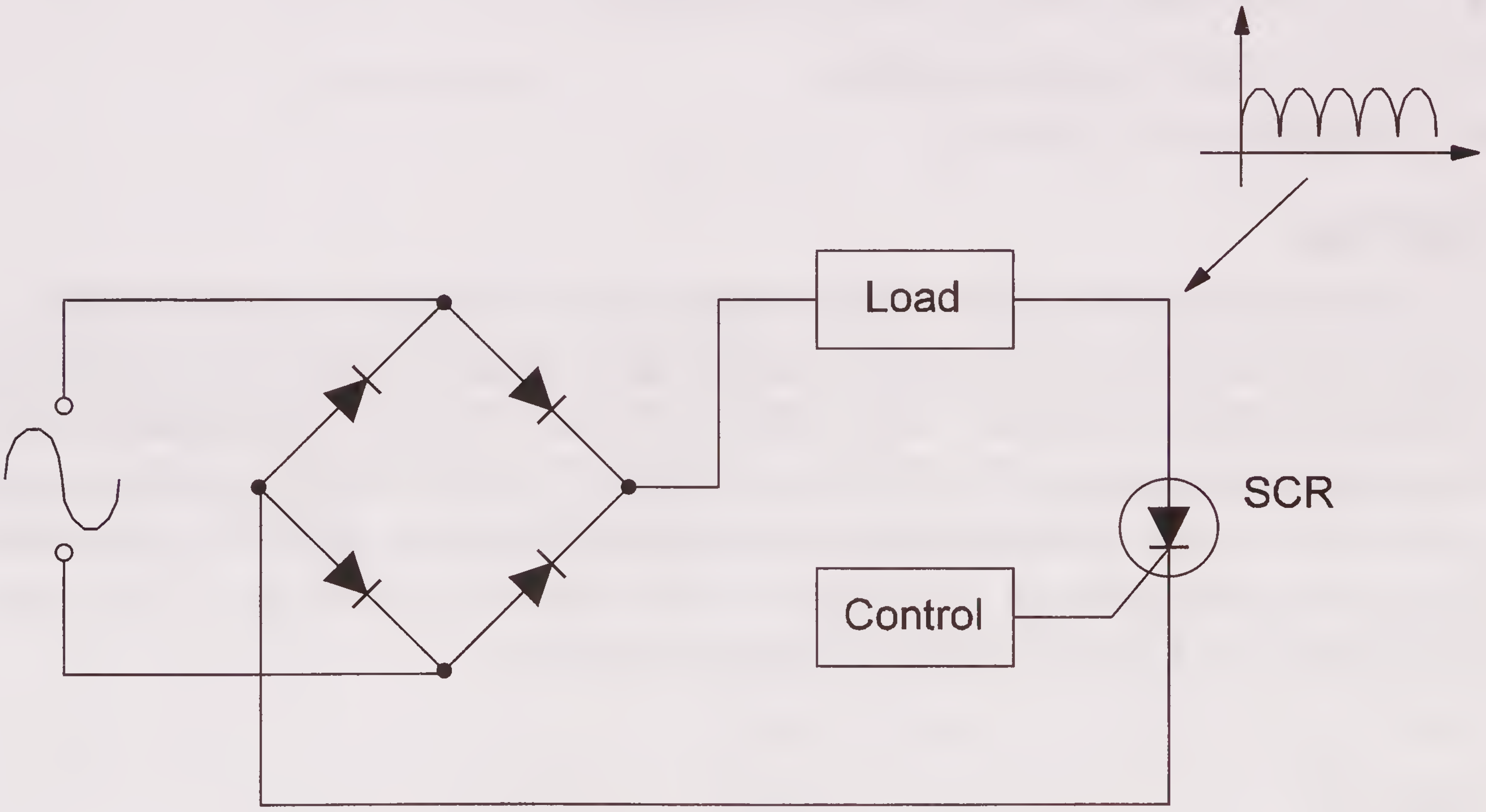

Figure 147

But, the best solution is in the use of another device of the thyristor family—the triac. The triac can be considered as two SCRs inside a unique case that can be used to control both semicycles of the AC power line voltage. The SCR is not used only as a power control. Many other configurations are possible including timers, oscillators, alarms, automatic switches, and protection circuits (crowbar).

The 106 Series

One of the most popular series of SCRs found in many electronic circuits is the 106 series. It is formed by low-cost general-purpose SCRs that can control currents up to 3 or 4 amperes (3.2 volts average for the TIC106) in circuits having voltages in the range between 50 and 1000 volts. The devices of this series are very sensitive; they can be triggered from currents as low as 60 uA (typical). The part number of the SCRs of this series begins with a group of letters identifying the manufacturer followed by 106, and a figure indicating the maximum operating voltage.

Some examples follow.

- MCR106-B (Motorola - the B indicates it is for 200 V)
- TIC106-6 (Texas Instruments for 400 V)
- IR106 (International Rectifier)
- C106 (General Electric)

Testing

SCRs can be tested with a multimeter. In normal conditions the device is an open circuit when the resistance between anode and cathode is measured (the multimeter will indicate infinite). If a low resistance is measured in this condition the SCR is shorted. The other test consists of the verification of the gate-cathode junction that acts as a diode. In one direction a low resistance is found and in the opposite direction a high resistance is found if the device is in good condition.

Triac

The Triac is another important device of Thyristor family. As demonstrated, SCRs are unidirectional thyristors meaning that they can control the current only in one direction—from anode to cathode. Unlike the SCR, triacs are bidirectional and can conduct the current in both directions. Compare a triac to two SCRs connected in inverse parallel (anti-parallel) and placed inside a single three-terminal case as shown in Figure 148.

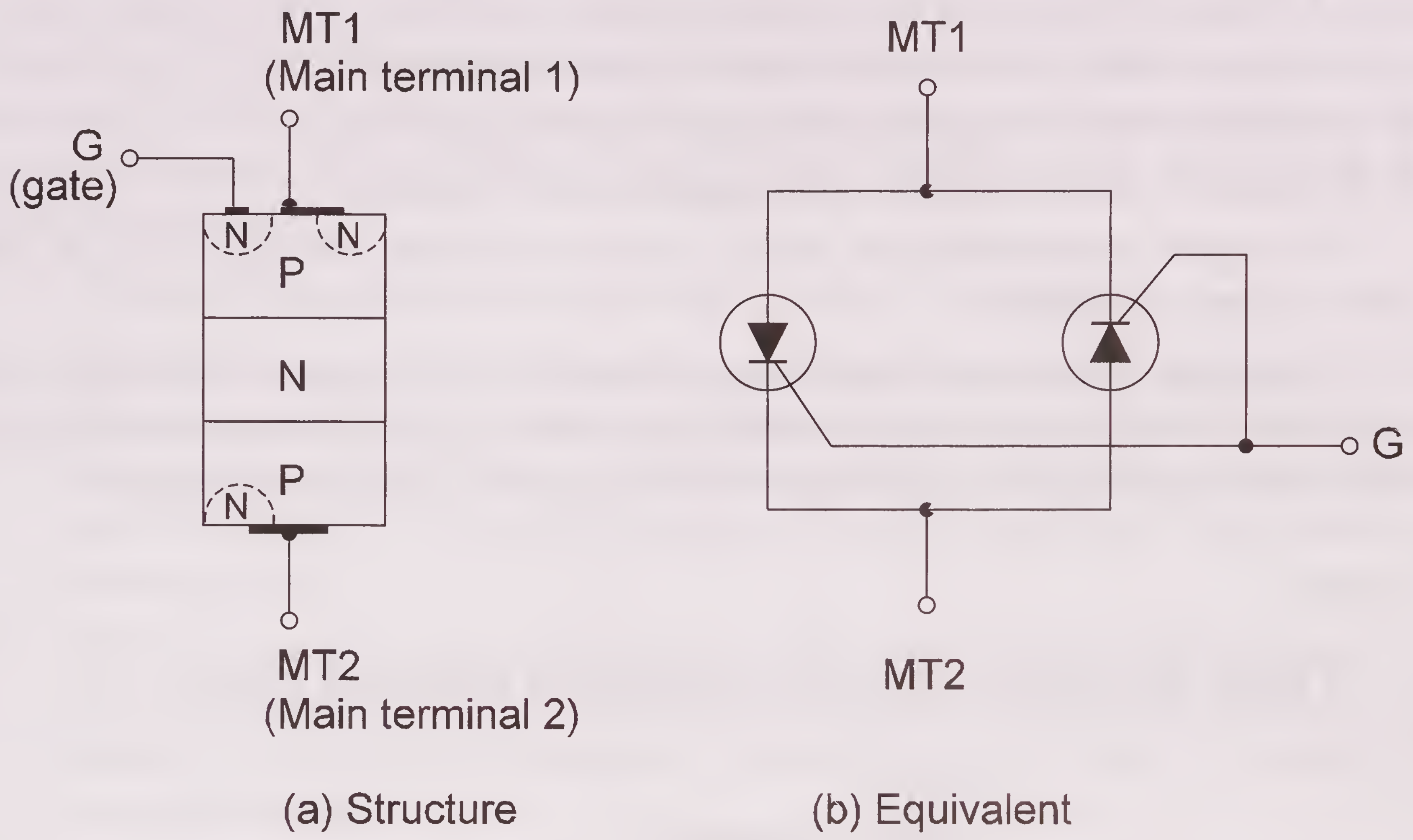
MT1
(Main terminal 1)
G
(gate)
N
N
P
N
P
N
MT2
(Main terminal 2)
(a) Structure
MT1
G
MT2
(b) Equivalent

Figure 148

The triac is indicated to applications where AC voltages must be controlled meaning that a large range of applications in domestic and industrial installations includes this component. So, it is another device with a large degree of importance for the electrician. In basic applications the triac is connected in a circuit as shown in Figure 149.

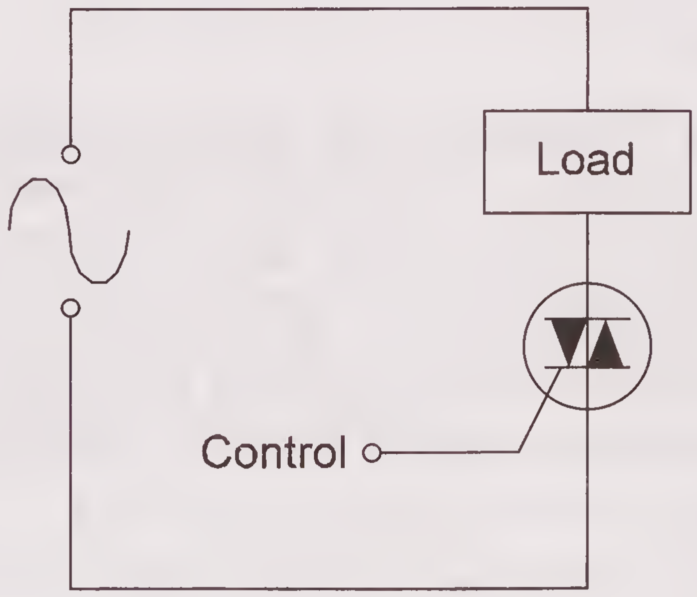
Load
Control

Figure 149

Like the SCR, the triac can be used as a power switch controlling a large amount of alternating current from the main terminals (MT1 and MT2) from a low voltage applied to the gate. Unlike the SCRs that need only positive voltages applied to the gate to be turned on, triacs can be triggered either by a positive or a negative gate signal, regardless of the polarities of the main terminal voltages. This means that the device can be triggered by four modes, indicated as follows:

- I+ Mode = MT2 positive and gate current positive
- I– Mode = MT2 positive and gate current negative
- III+ Mode = MT2 negative and gate current positive
- III– Mode = MT2 negative and gate current negative

The gate sensitivities in the I+ and III– modes are high and in the other modes are lower.

Common triacs need only some miliamperes of current in the gate to be turned on and once in the conductive state, high currents in the range from some amperes to more than 1000 amperes can be controlled.

Symbol

Figure 150 shows the symbol adopted to represent a triac.

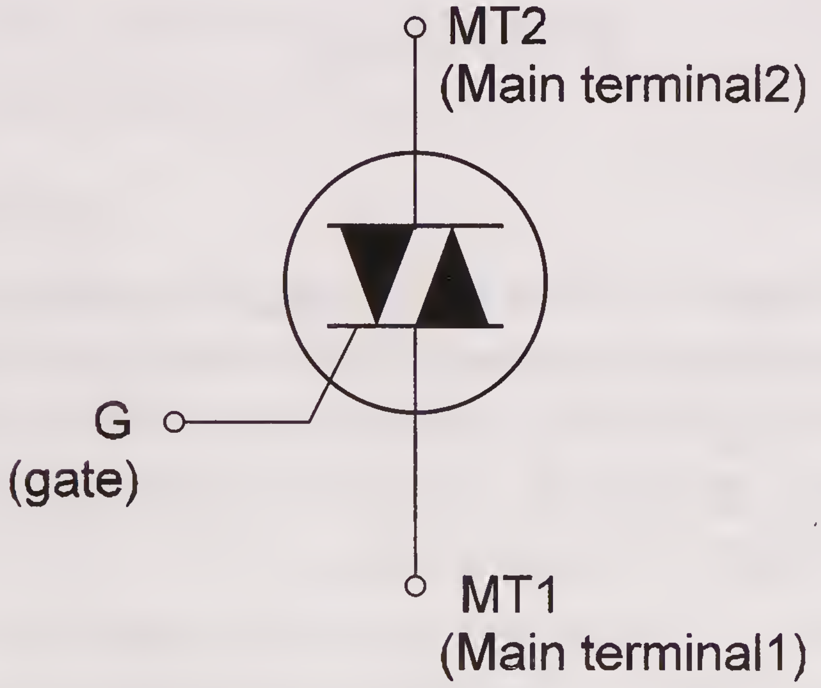

Figure 150

The body is the same as that used by high-power transistors, SCRs, and many other components. Notice that they are basically high-power solid-state switches and must be mounted in heat sinks when handling high currents. The identification of the terminals (MT1, MT2, and G) is made using the information of the data sheet or documents of the manufacturer as it changes from type to type. A general rule is that MT1 used to be connected to the case or to the tab.

Specifications

The triacs are identified by a part number displayed as a figure on the body, the same way the SCRs, transistors, and many other solid-state devices are. From the part number, using a data sheet or other document from the manufacturer, it is possible to find the electrical characteristics of the device. The inverse procedure is valid as well; from the characteristics, it is possible to find a triac suitable to an application by the part number.

The main specifications to be observed in a triac are:

A. Maximum voltage—the maximum voltage that can be applied to the triac in the off state. They are indicated also as the Repetitive Pulse Reverse Voltage (Vrrm) or Repetitive Peak Reverse Voltage (Vrrm) and can be typically in the range between 50 and more than 1000 volts for common types.

B. Maximum current—the maximum current that can be controlled by the device is indicated as the Continuous On State Current or average On State and can be in the range of 1 ampere to more than 100 amperes for common types.

C. Trigger current and voltage—the current that triggers the triac on is indicated as the Gate Trigger Current and abbreviated by Igt. Typically this current is in the range between some microamperes and 100 mA depending on the size of the device. The trigger voltage is also indicated as Vgt and is typically in the range between 0.8 and 1.5 volts for common types.

D. Dissipation power—When the triac conducts the voltage across, it is typically between 1.7 and 2.2 volts. This voltage times the controlled current or the load current gives the dissipation power of the device. The maximum dissipation depends on the device.

Where they are found

Triacs are found in all the modern applications where a speed, temperature, or power control must be made from an AC power line voltage. We can say that triacs are solid- state switches with a large range of applications in AC control.

Dimmers, speed motor control of electric appliances, heaters, and many other circuits use triacs as a central element.

Figure 151 shows a typical power control or dimmer for loads powered from the AC power line using a triac.

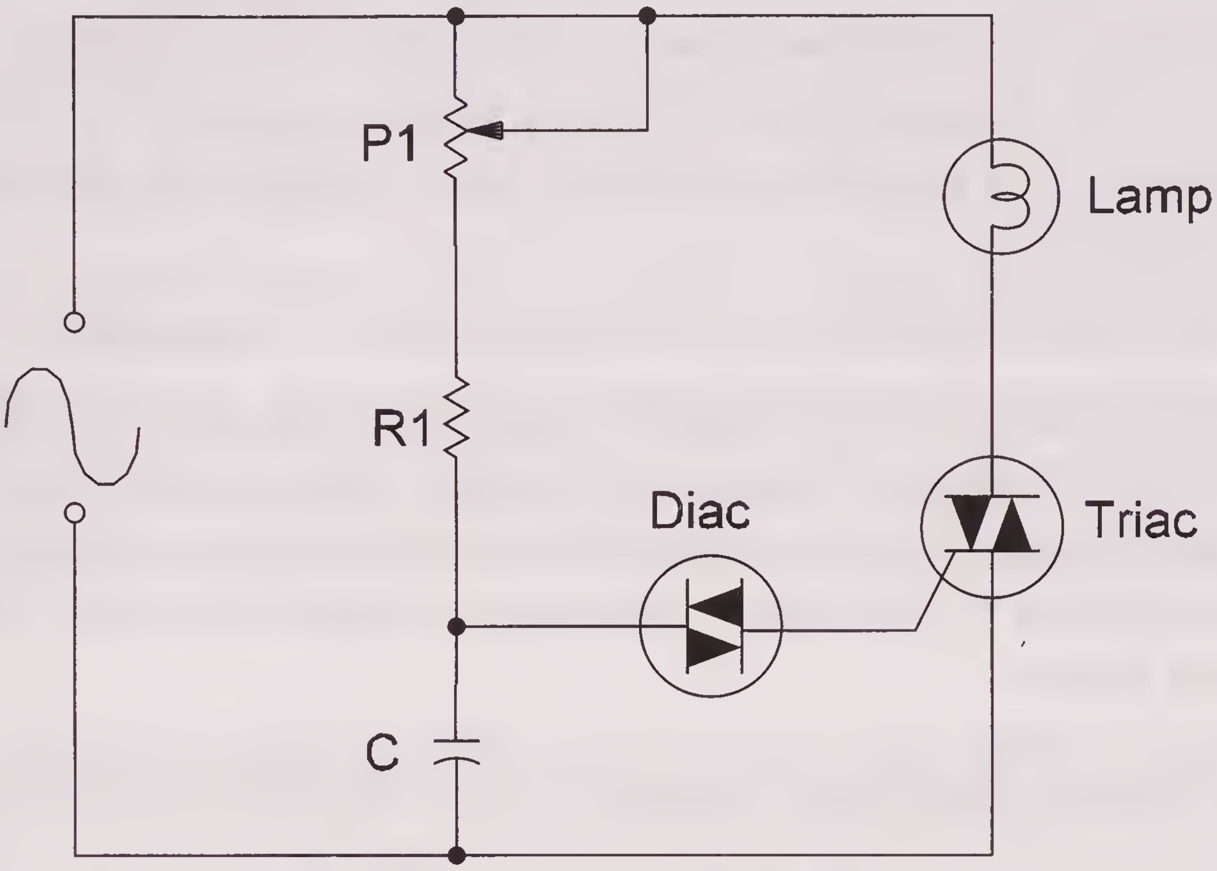

Figure 151

The operation principle is the same as the power control using SCRs previously described. The only difference is that the capacitor charges with both semicycles of the AC voltage and the triac triggers with them conducting both to the load.

The waveforms in the circuit are shown in Figure 152.

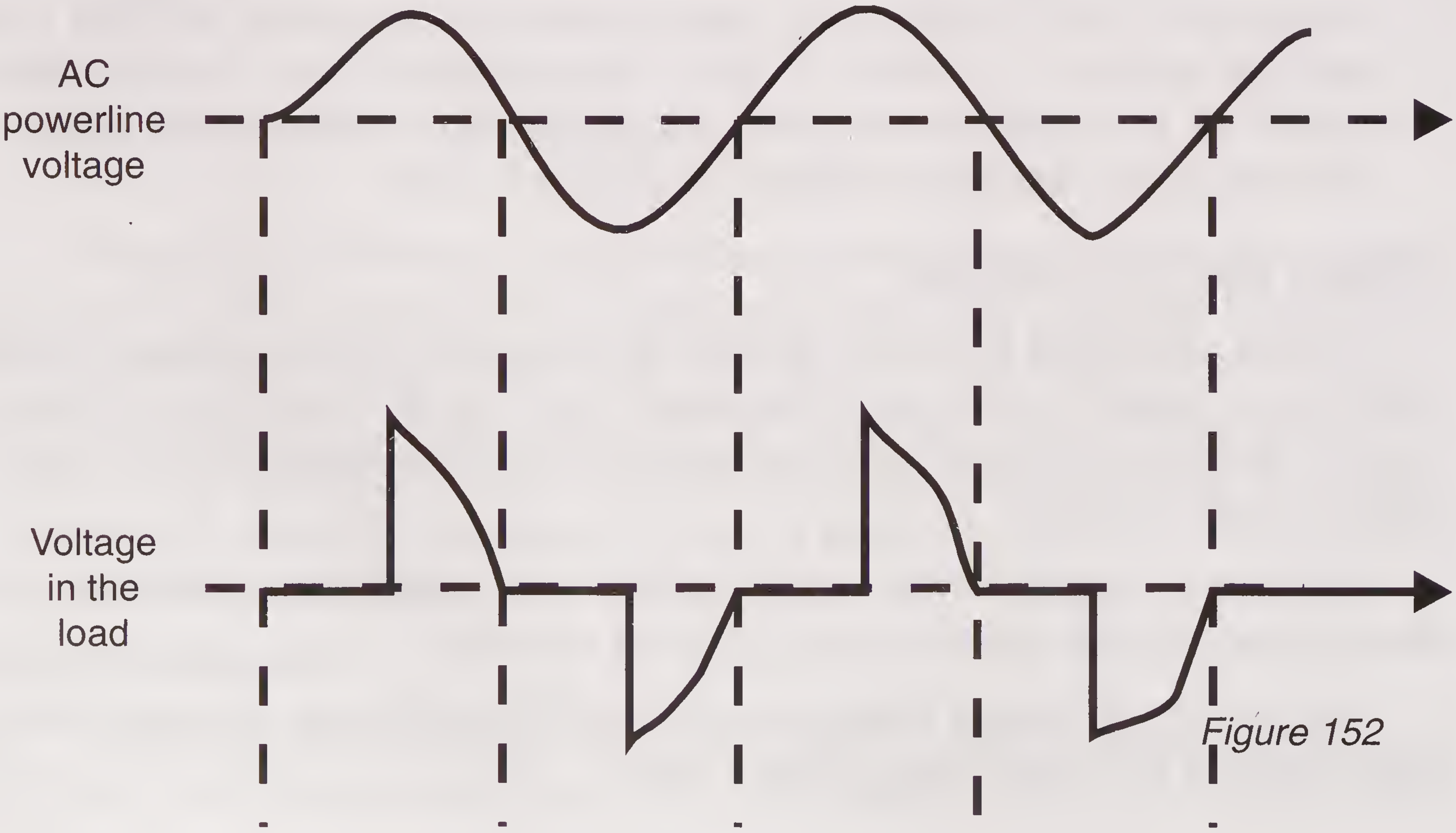

Figure 152

The diac is an auxiliary element used to help the triac trigger with the correct voltage adjusted by the potentiometer. Circuits like this are used in common incandescent lamp dimmers replacing common on-off switches. The speed control of motors used in electric appliances, such as drills and fans, use a circuit like this.

Testing

The best way to test a triac is using an appropriated circuit for this task. The multimeter can only reveal if a triac is shorted. If a low resistance is measured between MT1 and MT2 in a triac out of a circuit it is because it's shorted. When replacing a triac by an equivalent type, the specifications (voltage and current) must be equal or greater than the original.

Electromagnetic Interference (EMI)

Triacs and SCRs are high-speed power switches. Their turn-on and off times are very low, only a few microseconds, and this can cause problems with Electromagnetic Interference.

EMI or RFI (Radio Frequency Interference) is the interference caused by undesirable radio signals produced by devices like thyristors affecting the operation of many sensitive pieces of electronic equipment.

When an SCR or triac is used in a circuit as power control, the high-speed changes of the current in a load result in the generation of a series of harmonically related radio-frequency signals. The magnitude of the fundamental signal is proportional to the magnitude of the controlled current, and in a larger part of the cases, can be great enough to cause interference in AM radios and many other circuits that operate with low- and medium-frequency radio signals.

Specifically, the interference is higher when the SCR or triac controls inductive loads in motors, solenoids, magnets, etc. The snap action of the triac or SCR makes the current oscillate in the circuit generating a strong electromagnetic field.

A protection circuit that minimizes the interference is shown in Figure 153. The capacitor and the resistor form a damping circuit reducing the current oscillation and the electromagnetic interference.

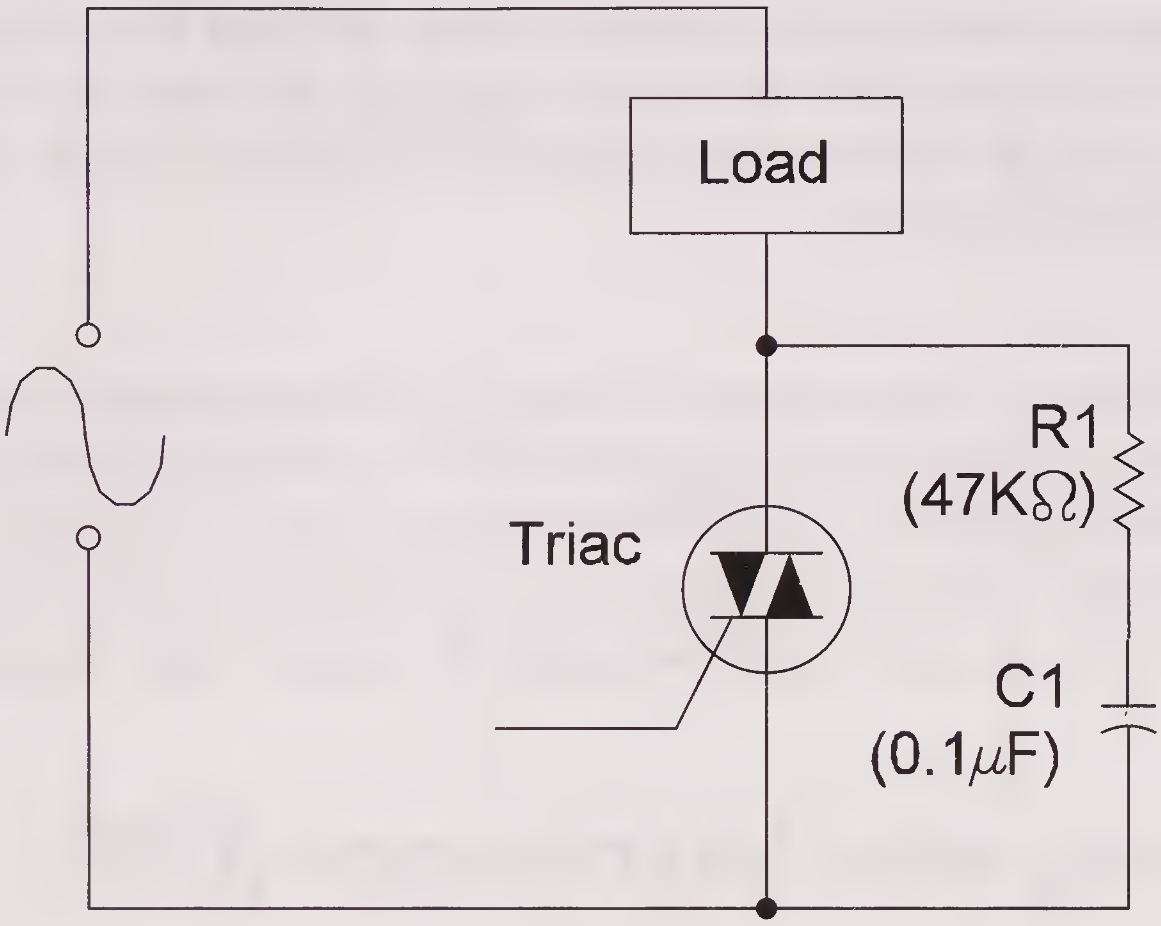
Load
R1
(47KΩ)
Triac
C1
(0.1μF)

Figure 153

The noise in an AM radio receiver produced when a circuit using triacs or SCRs is turned on or the small dots and points in a TV screen when the same device is turned on are examples of the interference process.

The interference can reach the radio receivers, TVs, or other sensitive devices two ways. The first way is for it to be irradiated directly into the air as a radio signal (Figure 154).

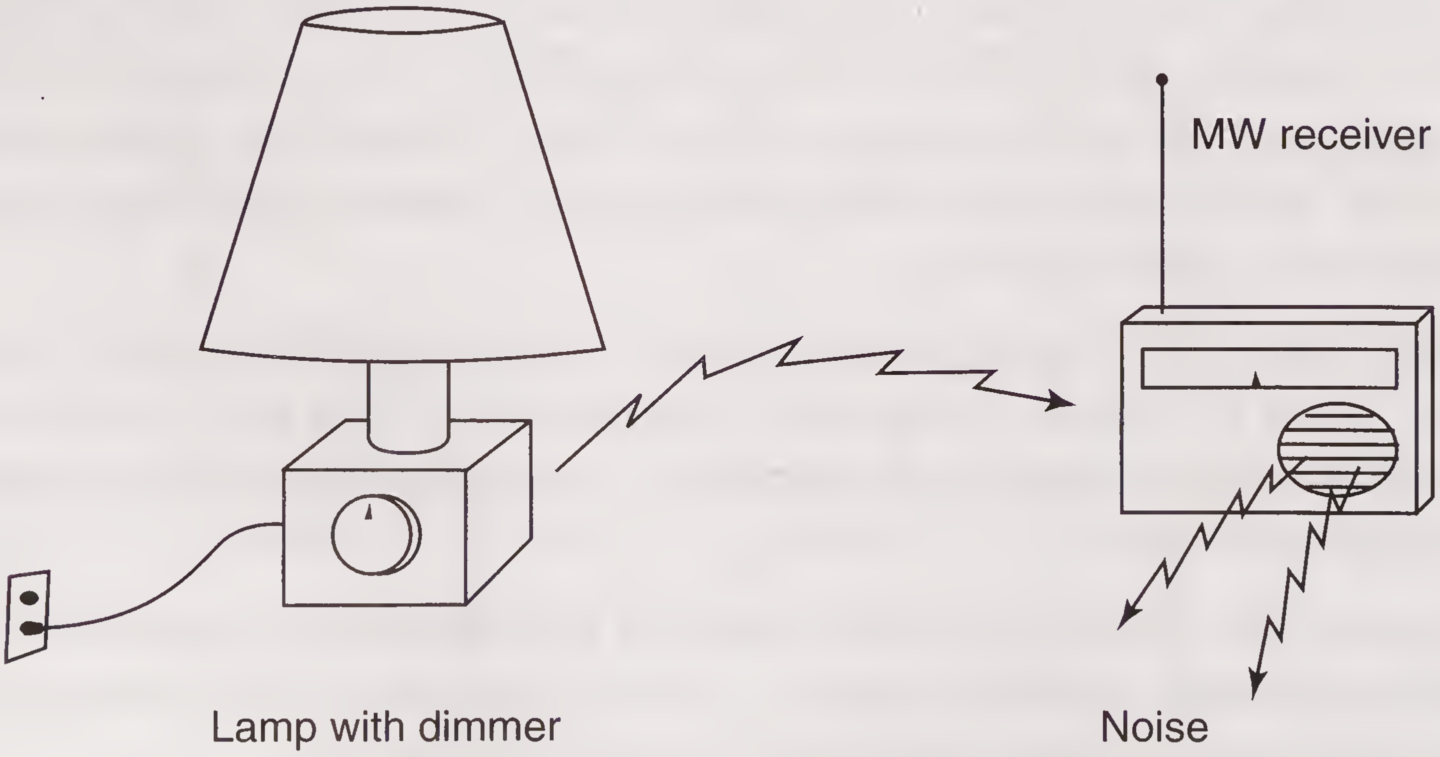
MW receiver
Lamp with dimmer
Noise

Figure 154

The undesirable signal, in general, is very weak and only causes problems if a radio or other type of sensitive equipment is placed near the source of interference. It is also concentrated in the low band of the radio spectrum as shown in Figure 155, interfering much more with power in the LF (low frequency) and MW (medium wave) band.

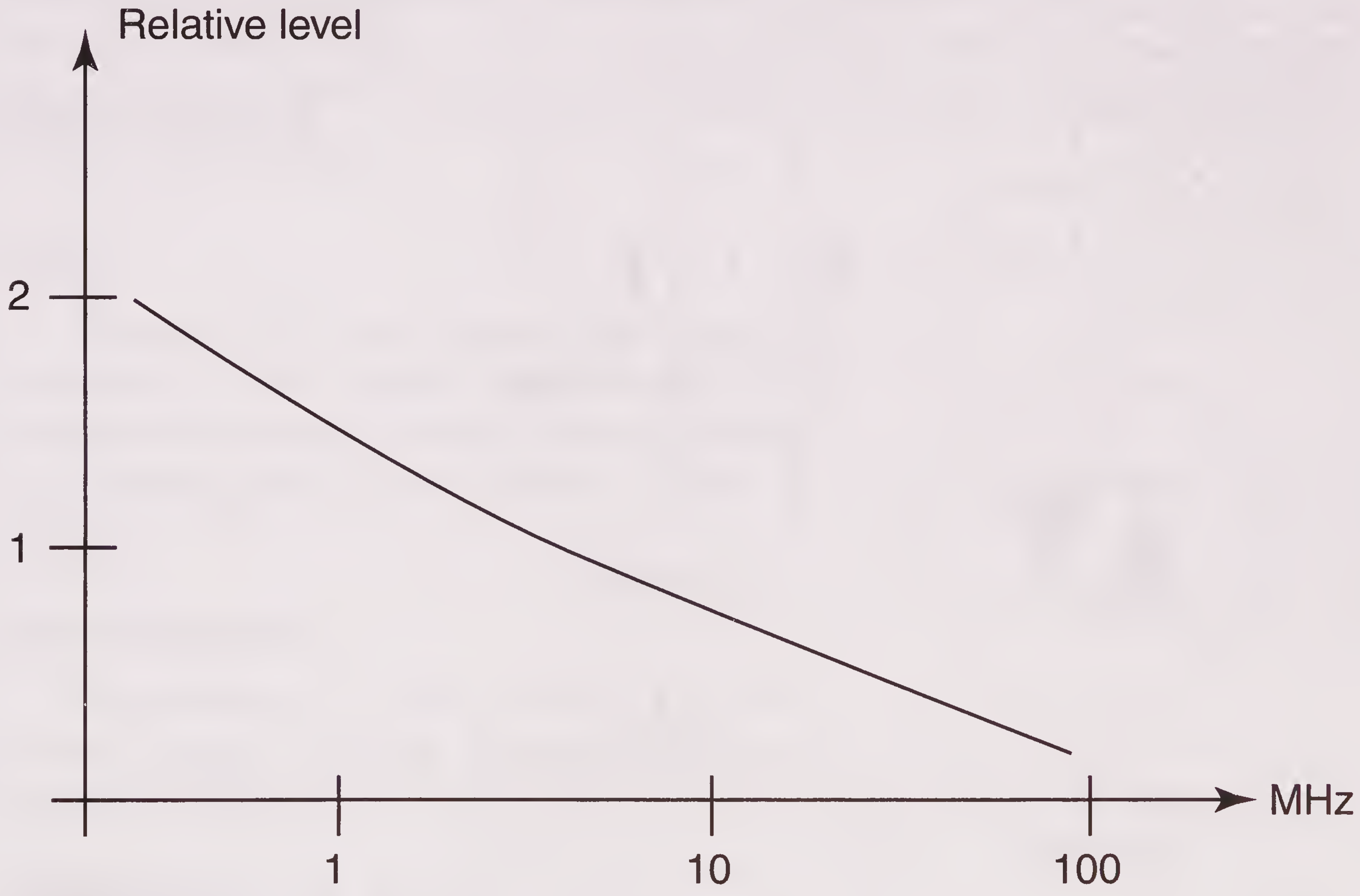

Figure 155

Radio receivers operating at higher frequencies, such as FM receivers, VHF receivers used in remote controls for garage doors, and cellular telephones aren't much affected by this interference not because of lower power concentrated in this band, but also because of their operation principles.

Secondly, the signal reaches the interfered equipment by the AC power line, as shown in Figure 156.

To avoid the interference in this case, use an L-C filter like the one shown in Figure 157.

The electrician can have this problem many times. When installing a dimmer or other device using SCRs and triacs in any domestic electric installation, a TV or an AM radio placed near it or powered from the same

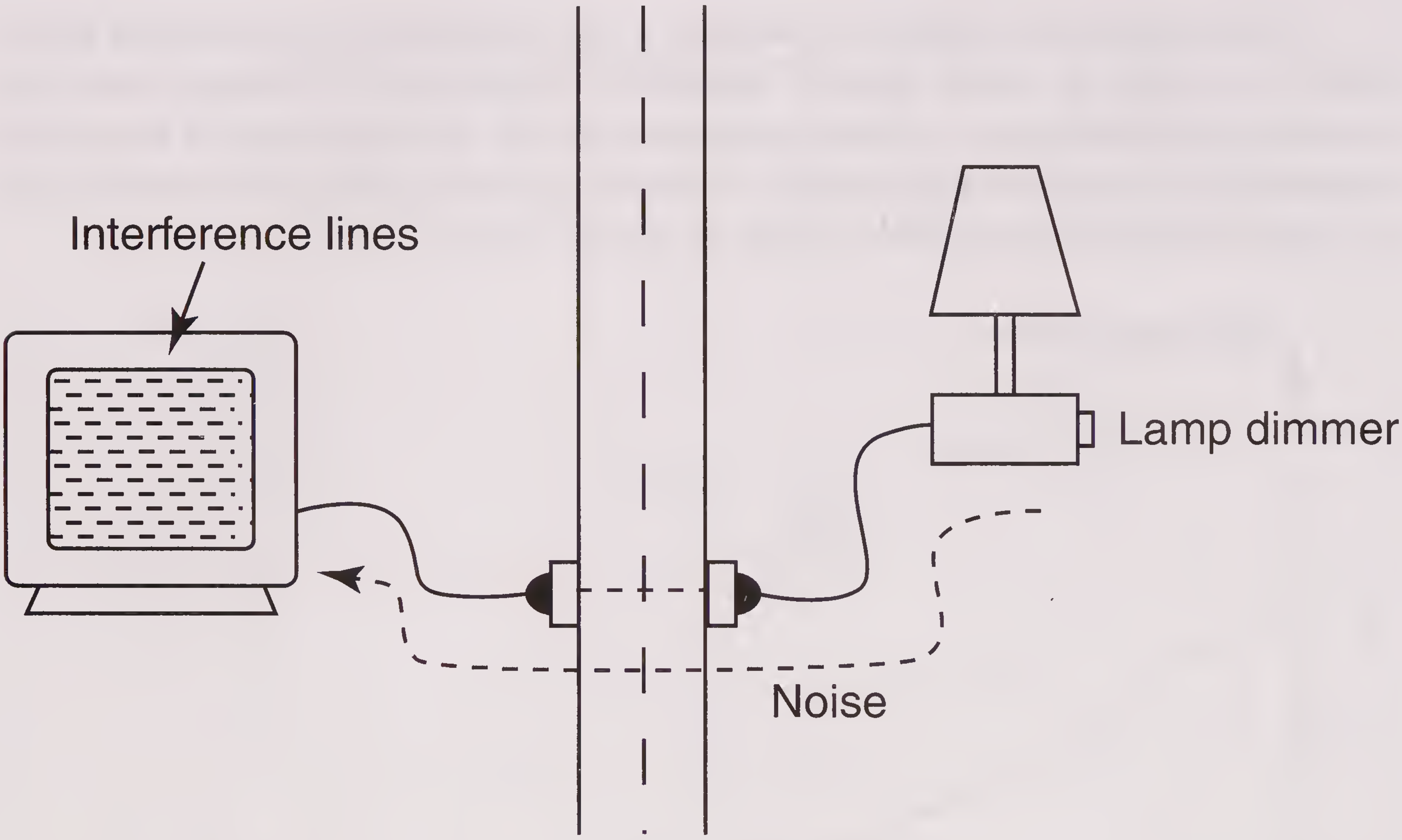
Interference lines
Lamp dimmer
Noise

Figure 156

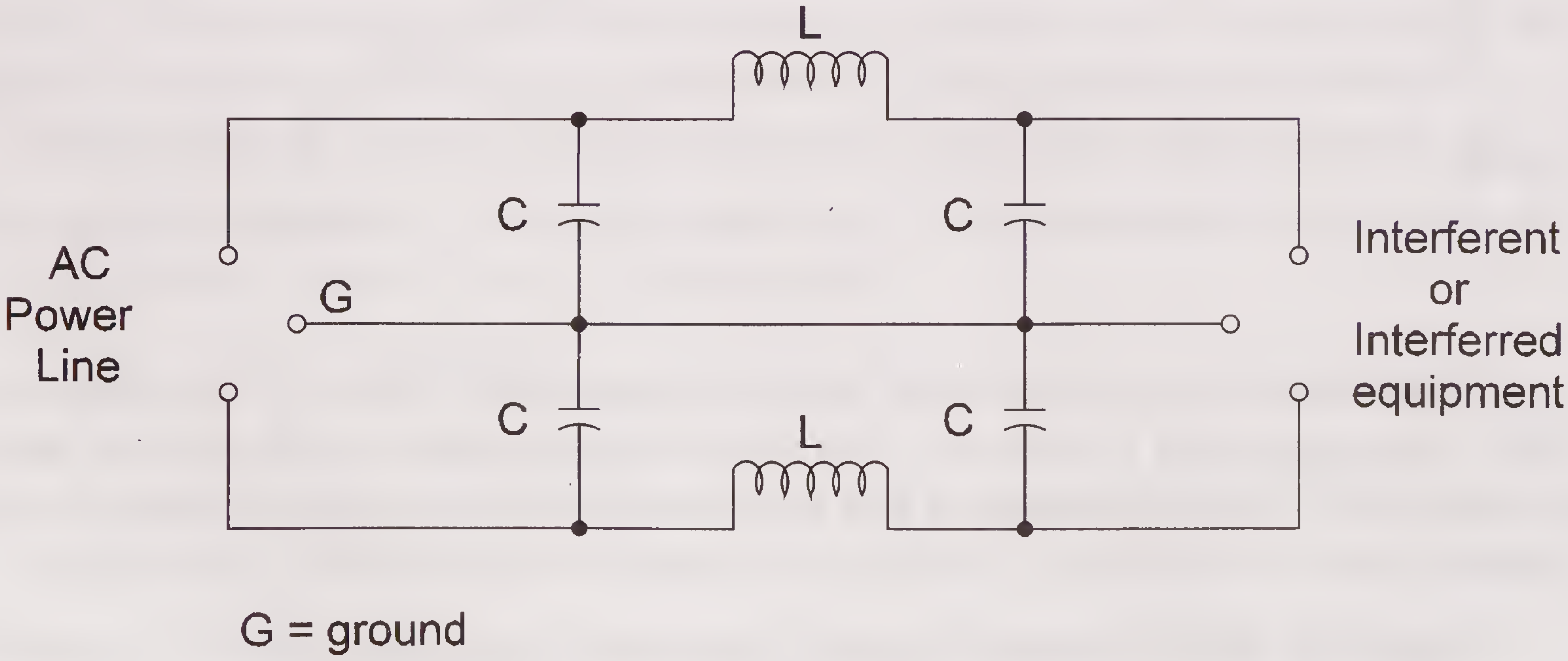
L
C
C
AC
Power
Line
G
Interferent
or
Interferred
equipment
C
C
L
G = ground

Figure 157

AC power line can interfere. A filter like the one shown in Figure 157 placed between the dimmer or the interfering equipment and the power source can solve the problem or reduce the interference to a comfortable level.

Diac

The DIAC is another semiconductor device of the thyristor family. It is also a four-layer component with special trigger characteristics. The diac is the semiconductor equivalent of the neon lamp. When the voltage across a diac reaches a certain value, the device triggers on, passing from a high-impedance state to a very low-impedance state. The high-speed triggering characteristics of this component make it ideal as a trigger element for circuits using triacs.

Symbol

Figure 158 shows the symbol used to represent a diac. Notice that the diac is a nonpolarized component because the triggering characteristics are the same in both directions.

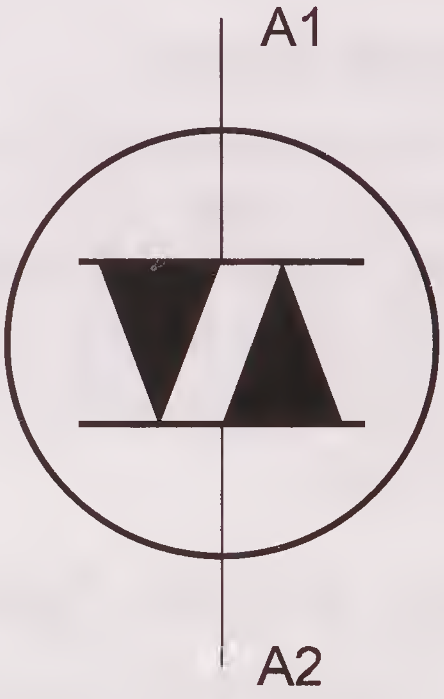

Figure 158

Specifications

Associated to a part number, the diacs have a trigger voltage. Typical values are around 35 volts.

Where they are found

Although many applications exist with diacs as a central element of a circuit, the most common is as a trigger element of a triac as in the dimmer. Speed controls for electric motors and dimmers are examples of circuits of where triacs are found, the triggering devices are the diacs.

Testing

The multimeter can't be used to test diacs as they don't have enough voltage to trigger the multimeter and when turned off, they act as an open circuit. The best way to test a diac is using a circuit. The multimeter can only reveal if the diac is shorted when presenting a very low resistance.

Quadrac

The quadrac is a device formed by a diac and a triac in the same body. It is also used as a power control for incandescent lamps, heaters, motors, and other inductive loads the same way the triacs and diacs are used.

Symbol

Figure 159 shows the symbol of a quadrac. They use the same body as power transistors.

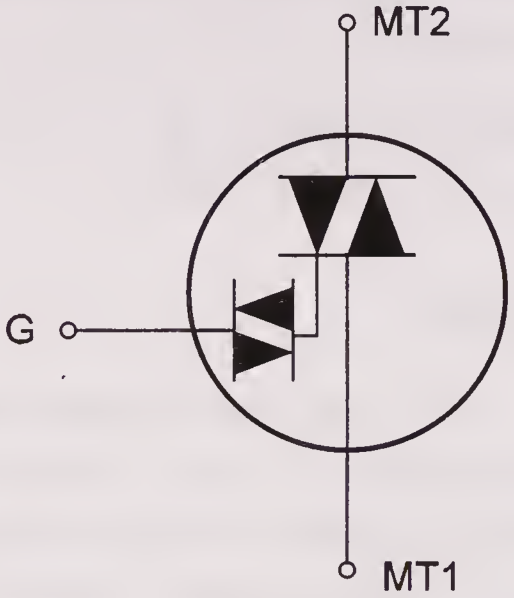

Figure 159

Specifications

The specifications of a quadrac are the same as a triac and a diac. The maximum voltage, maximum current, and trigger point are indicated by the manufacturer. A part number or type number is used to design the component.

Where they are found

Quadracs are not common today. They can be found in old motor controls, lamp dimmers, and some electric appliances where the speed of a motor, temperature, or any amount of power applied to an AC load must be controlled.

The basic application is the same as a triac in power controls as explained before.

Testing

The multimeter can only reveal if the quadrac is shorted. Low resistance between any of the three terminals indicates a problem.

Silicon Unilateral Switch—SUS

The Silicon Unilateral Switch is a device of the thyristor family and is used as a trigger element in circuits using SCRs. The SUS is an SCR with an anode gate and a built-in zener diode, as indicated by the equivalent circuit shown in Figure 160.

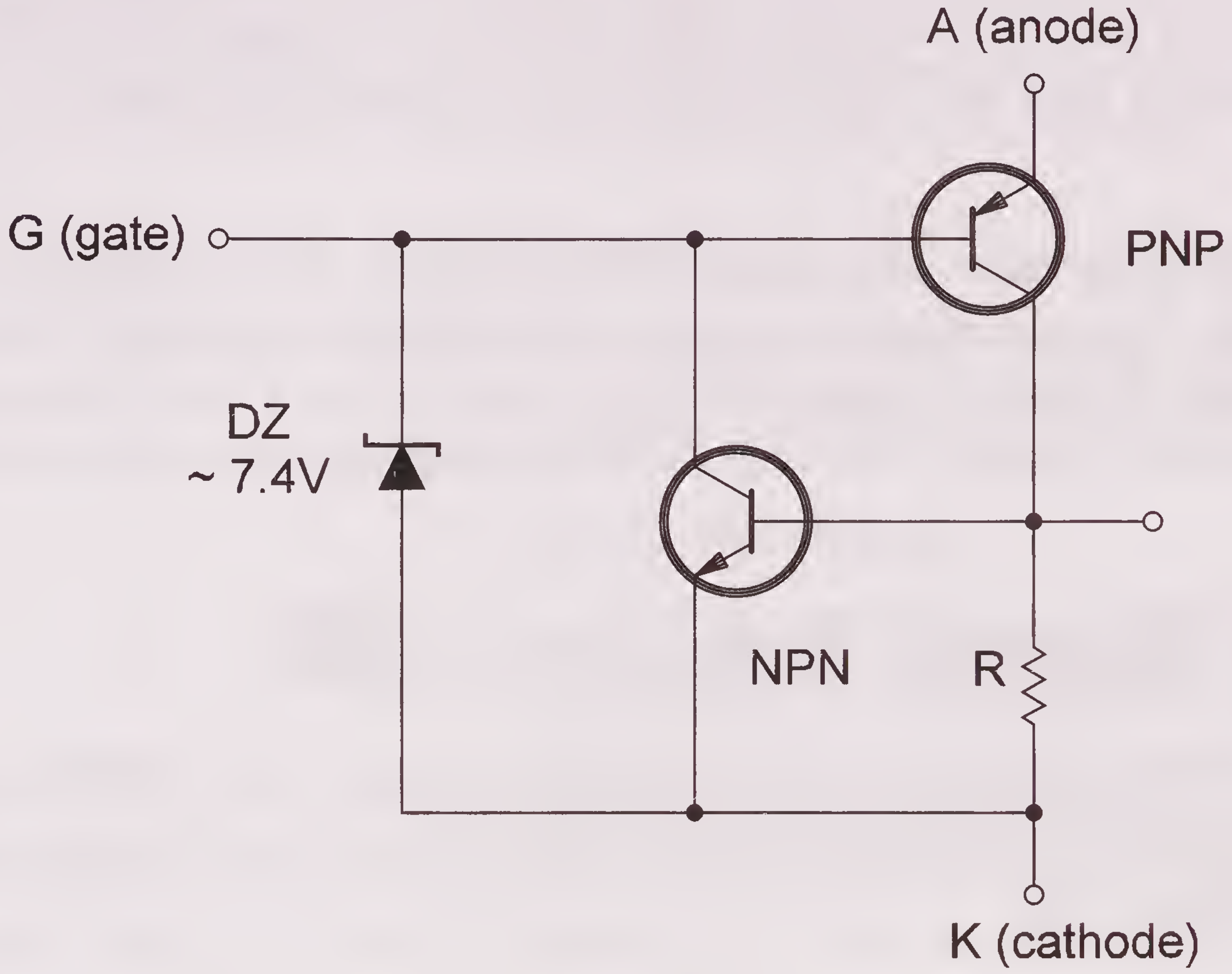

Figure 160

The zener diode determines the triggering voltage, which is normally around 8 V. But, by connecting a lower-voltage zener between the gate and the cathode it is possible to reduce the triggering voltage.

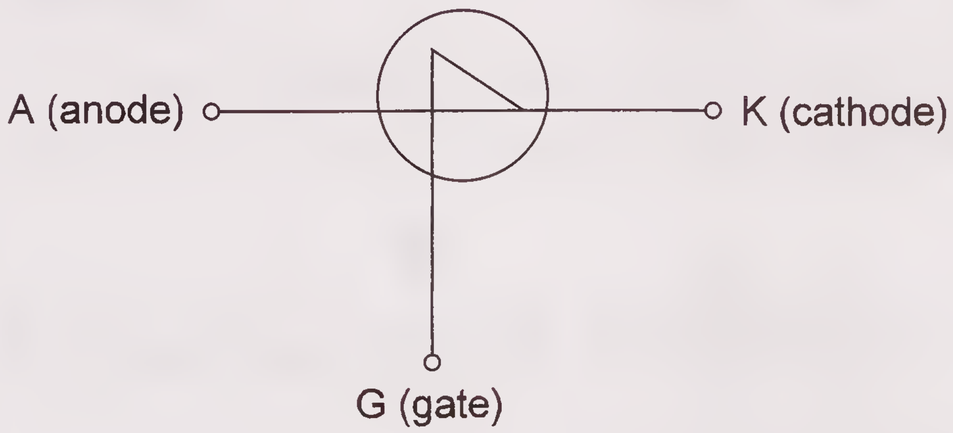

Figure 161

Symbol

Figure 161 shows the symbol of an SUS. It looks the same as a common low-power transistor because they use the same bodies.

Specifications

The main specification of an SUS is the triggering voltage or the internal zener voltage. Other specifications are the maximum current and the maximum voltage supported by the device. Manufacturers identify SUS by part numbers.

Where they are found

Silicon unilateral switches are not common devices today. They can sometimes be found in circuits using SCRs as a trigger element. The applications using only an SUS as core of any project are rare.

Testing

A multimeter can only reveal if the SUS is shorted between anode and cathode. The resistance between gate and anode is low in one direction and high in the opposite direction due to the PN junction between these two points. Ideally, the test must be made using a special circuit.

Silicon Bilateral Switches—SBS

Silicon Bilateral Switches (SBS) are devices of the thyristor family. It can be considered the same as two SBS wired in anti-parallel mode as shown in Figure 162.

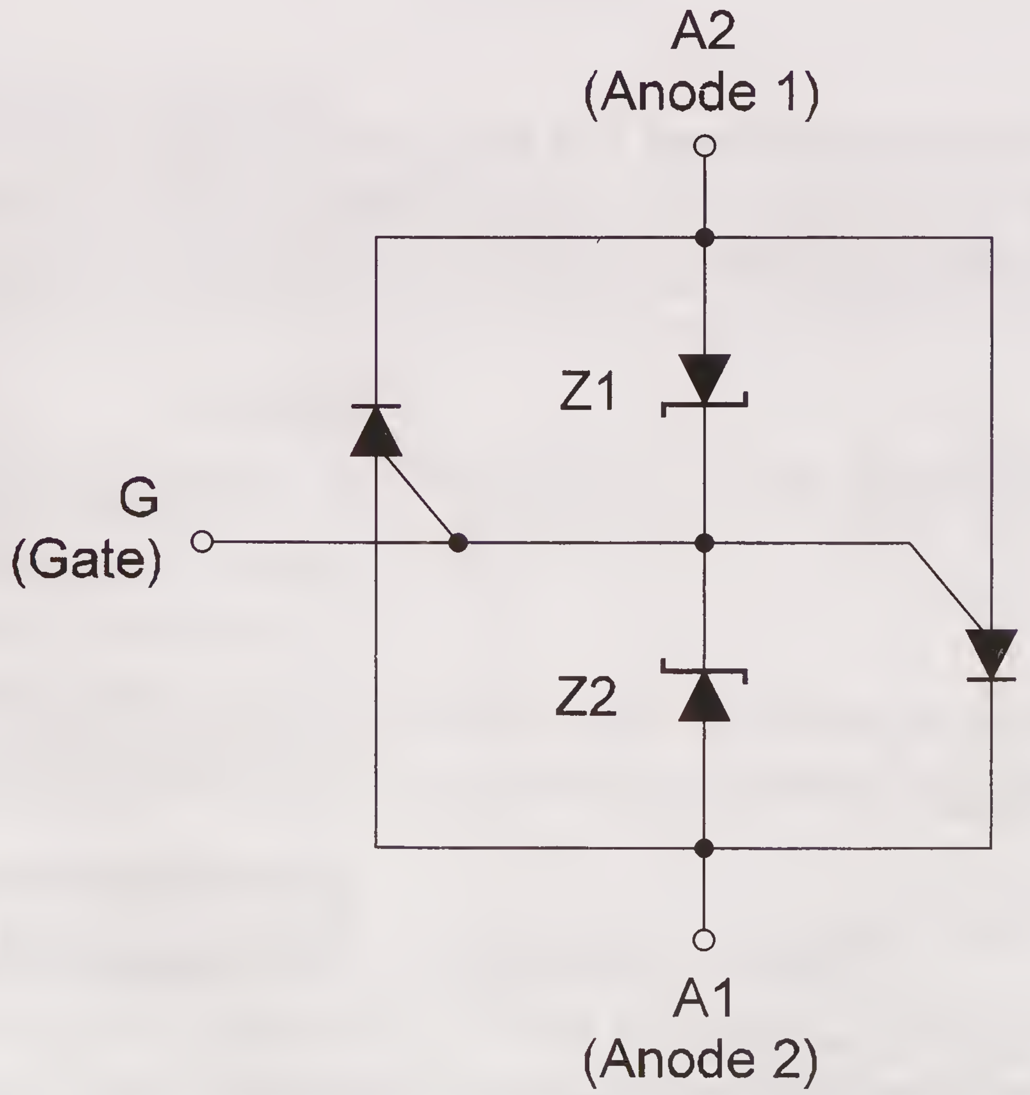

Figure 162

They are used as triggering elements for circuits using triacs. The operation is the same as the SUS; the internal zener diodes determine the trigger point of each SUS and the gate can be used to alter this value. External zener diodes can be wired between the gate and the two other terminals (A1 and A2) to lower the trigger point of the device.

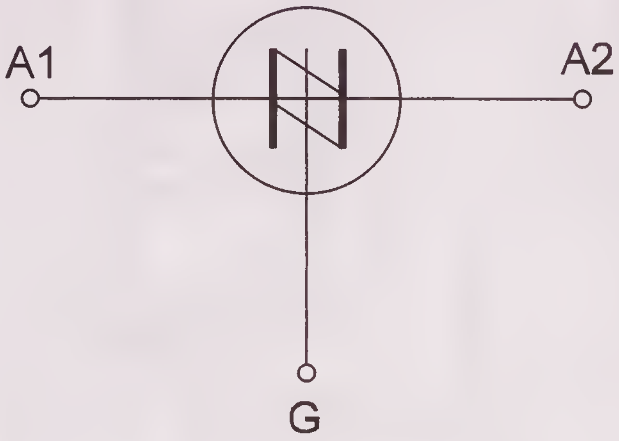

Figure 163

Symbol

Figure 163 shows the symbol adopted to represent the SBS.

Specifications

A part number is adopted by the manufacturer to indicate the device is an SBS. However, only with a data sheet or other manufacturer's document is it possible to access the electrical characteristics of the device, which are the maximum current and voltage between A1 and A2 and the voltage of the internal zener diodes.

Where they are found

The SBS isn't a common device in modern equipment. They can be found in some equipment using SCRs as motor controls, timers, alarms, sensor interfaces, etc.

Testing

The multimeter can only reveal if an SBS is shorted. The operation test must be done using an appropriated circuit.

Liquid Crystal Displays—LCD

Liquid Crystal Displays or LCDs are found in a lot of modern electronics. The liquid crystal displays can display information like the results of arithmetic operations in calculators; time in clocks, timers and watches; programmed operations in dishwashers, clothes washers, and microwave ovens, as well as many other modern applications.

The liquid crystal displays use the electrical properties of the molecules of certain substances that can change their positions when an electric field acts on them as shown in Figure 164.

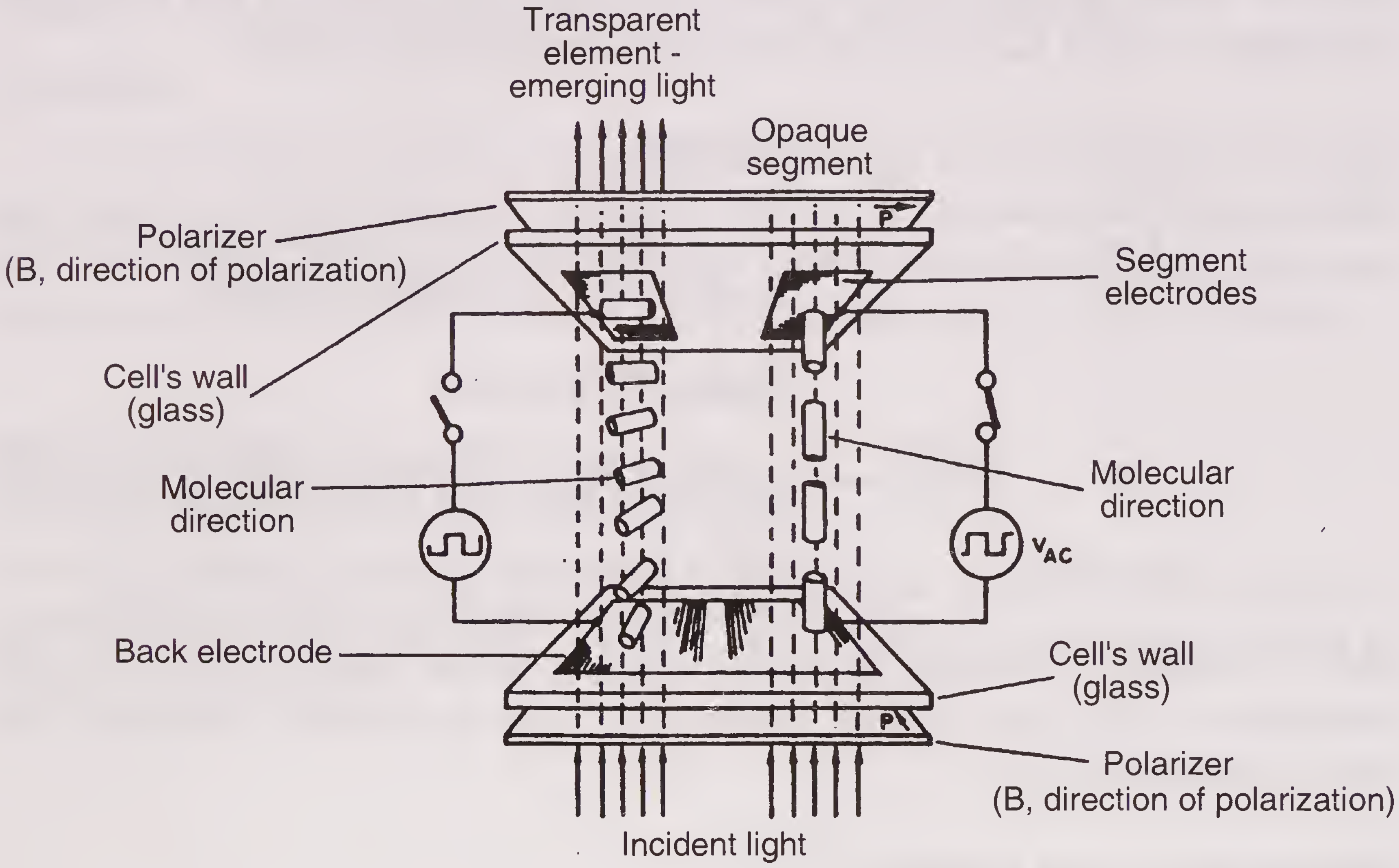
Transparent element - emerging light
Opaque segment
Polarizer (B, direction of polarization)
Segment electrodes
Cell's wall (glass)
Molecular direction
Molecular direction
V_{AC}
P
Back electrode
Cell's wall (glass)
P
Polarizer (B, direction of polarization)
Incident light

Figure 164

With no electric field present in the substance the molecules are oriented in a manner that lets the light pass through the liquid. The substance is now transparent. When the electric field is applied, the position of the molecules changes making them cut the light. The substance, under this condition, becomes opaque.

If transparent electrodes with the pattern to be displayed are placed inside a can filled with a liquid presenting this property, when a voltage is applied in any part or electrode, this part becomes opaque projecting the corresponding image as shown in Figure 165.

The simplest LCD display is the seven-segment type used to display numbers from zero to nine according to excited segments. This type of display is found in clocks, watches, calculators, and timers, just to name a few.

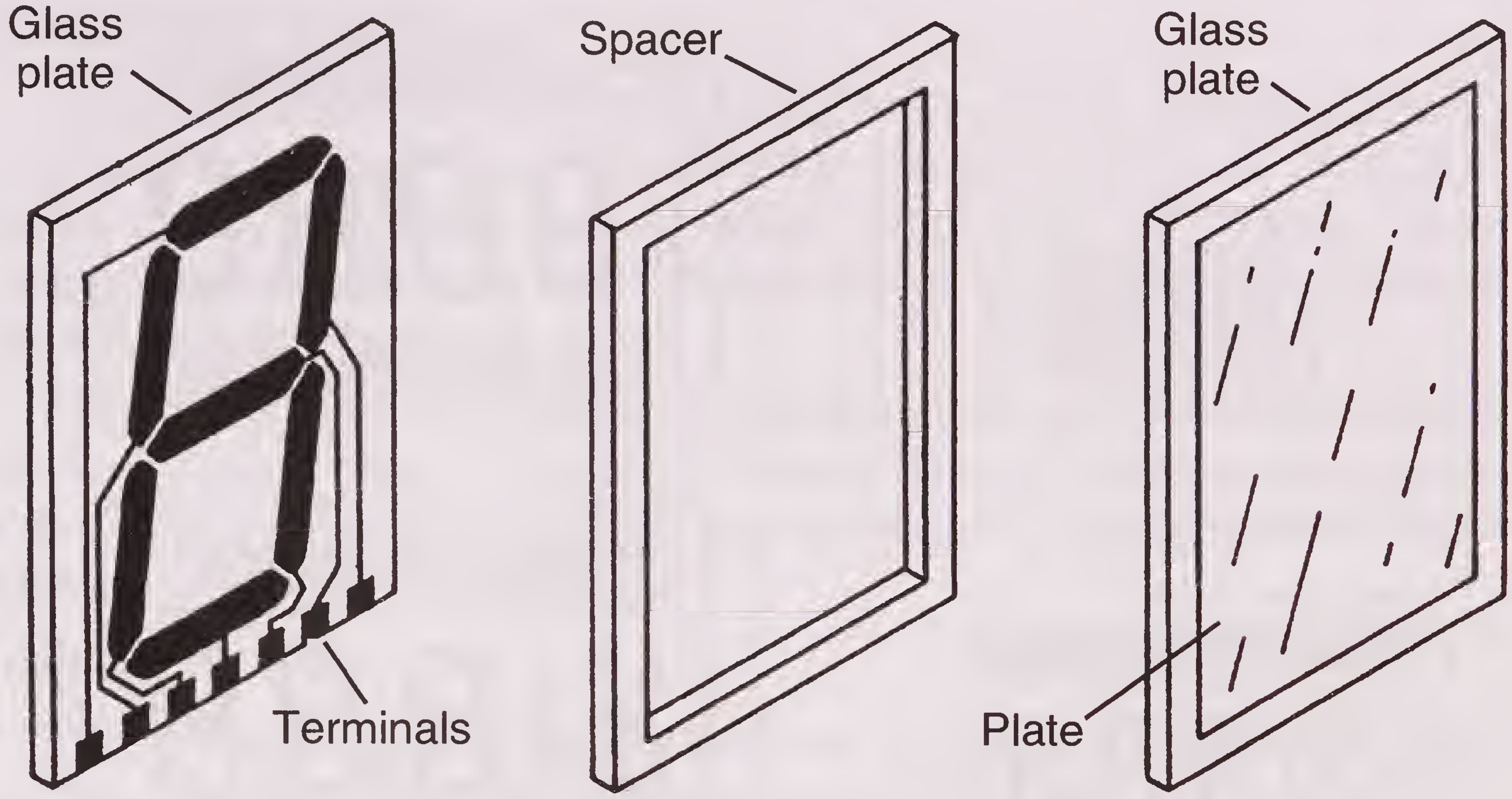

Figure 165

More complex displays include other symbols that can be energized with special signals such as figures in mini games, audio equipment, and radios.

The LCDs need special circuits to become active. The circuit not only recognizes what places of the display must be triggered to show the desired figure, but then also applies an AC voltage. If a DC voltage is used, an electrolysis phenomenon can occur causing the deterioration of the LCD.

The LCD drivers have special characteristics and are included in the group of the ICs (integrated circuits) described later.

Symbols and Views

Figure 166 shows the symbols used to represent an LCD. The same figure shows views of some of these displays.

There isn't any special symbol because many variations can be found in this component based on the application. Normally a representation that corresponds as near as possible from the final end is adopted and easily recognized by anyone who is reading a schematic diagram or working with the circuit.

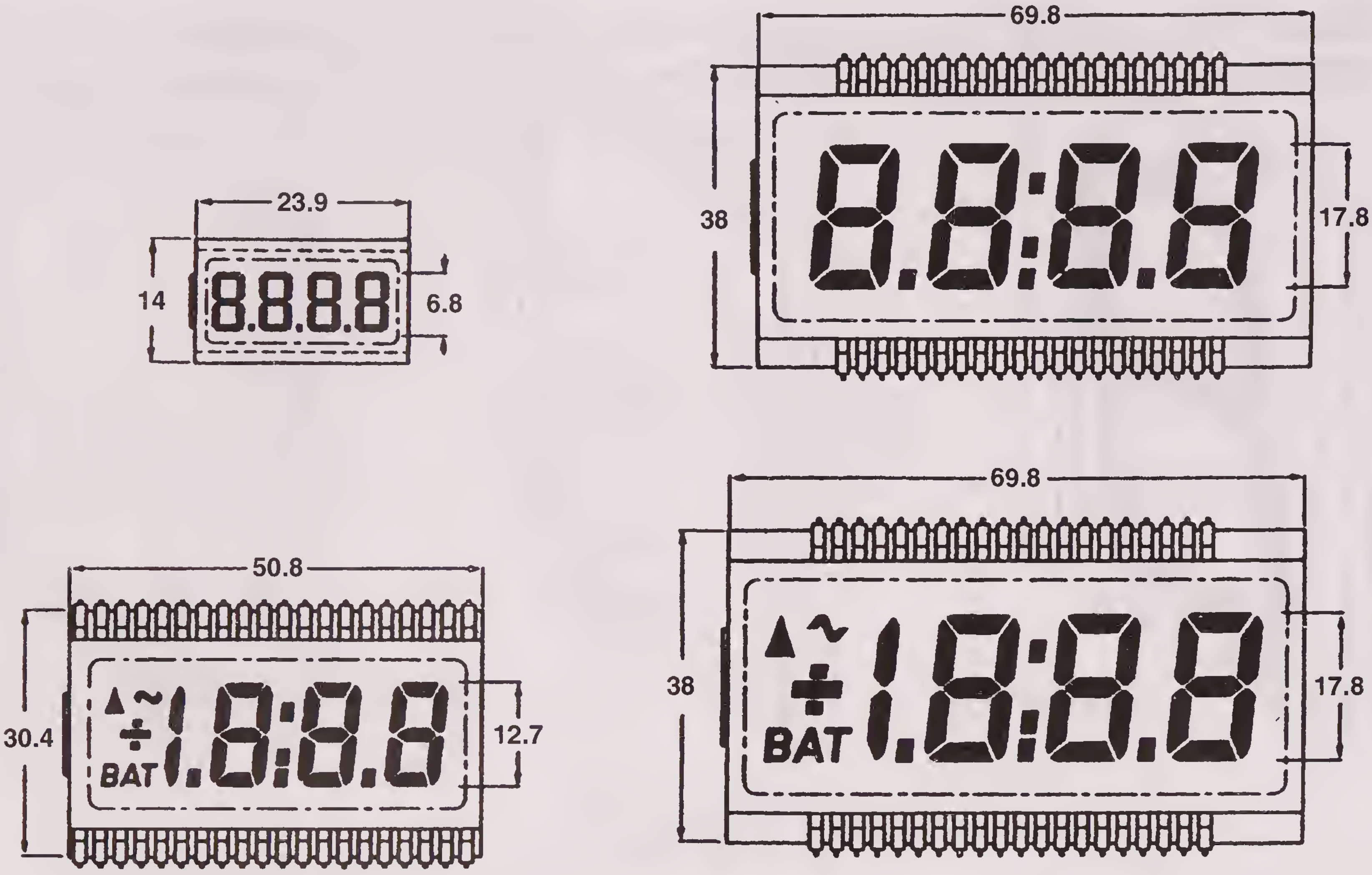

Dimensions in mm

Figure 166

Specification

The LCDs are indicated by a part number in most cases. We can separate LCDs in two groups: general-purpose types and special or custom types.

General purpose are those that present generic symbols (normally seven segments or matrix types) that can be used to present information for the projection. The same type of display can present several types of informational messages.

Specific types or custom types are those that are prepared for a determined manufacturer that creates a special configuration of information for a product. For example, the displays that have the symbol of a company or a figure representing the parts of equipment that are activated at any given time, are examples of these displays.

The electrical characteristics are not so important in this case as they always are driven by a special circuit (usually an integrated circuit).

Where they are found

LCDs are found in any application where numbers, alphanumeric information, or another type of information must be displayed. Watches, clocks, timers, microwave ovens, radios, telephones, car panels, industrial machine control panels, and facsimile machines are examples of places where LCDs can be found.

Some important information to know about the use of these devices is that they have a power consumption many times lower than the equivalents using LEDs so they are especially suitable for battery-powered applications.

Testing

No simple practical electrical test can provide any information about an LCD. The best way to test is using a specific circuit.

Tubes

Although integrated circuits (IC) and transistors are replacing the tubes in most modern applications, there is still old equipment in use today based on these components. To understand how the tubes work and are used, it is interesting to start from the simplest type: the diode.

A vacuum exists inside a glass tube with some electrodes. In the case of a diode, a tungsten filament is placed inside (in some cases near an electrode called a cathode). Also nearby (without touching) is another electrode called an anode. To put the tube to work a positive voltage is applied to the anode, the cathode is connected to a negative voltage, and the filament is heated by a low-voltage source as shown in Figure 167.

When the filament and cathode are heated, electrons are emitted causing current flows between the filament (cathode) and the anode. If the anode is negative and the filament is positive no current can flow across the tube. In a tube like this, the current can flow only in one direction, such as in a semiconductor diode.

If we now place a metallic grid (Figure 168) between the anode and the cathode we can observe new electrical properties in the device.

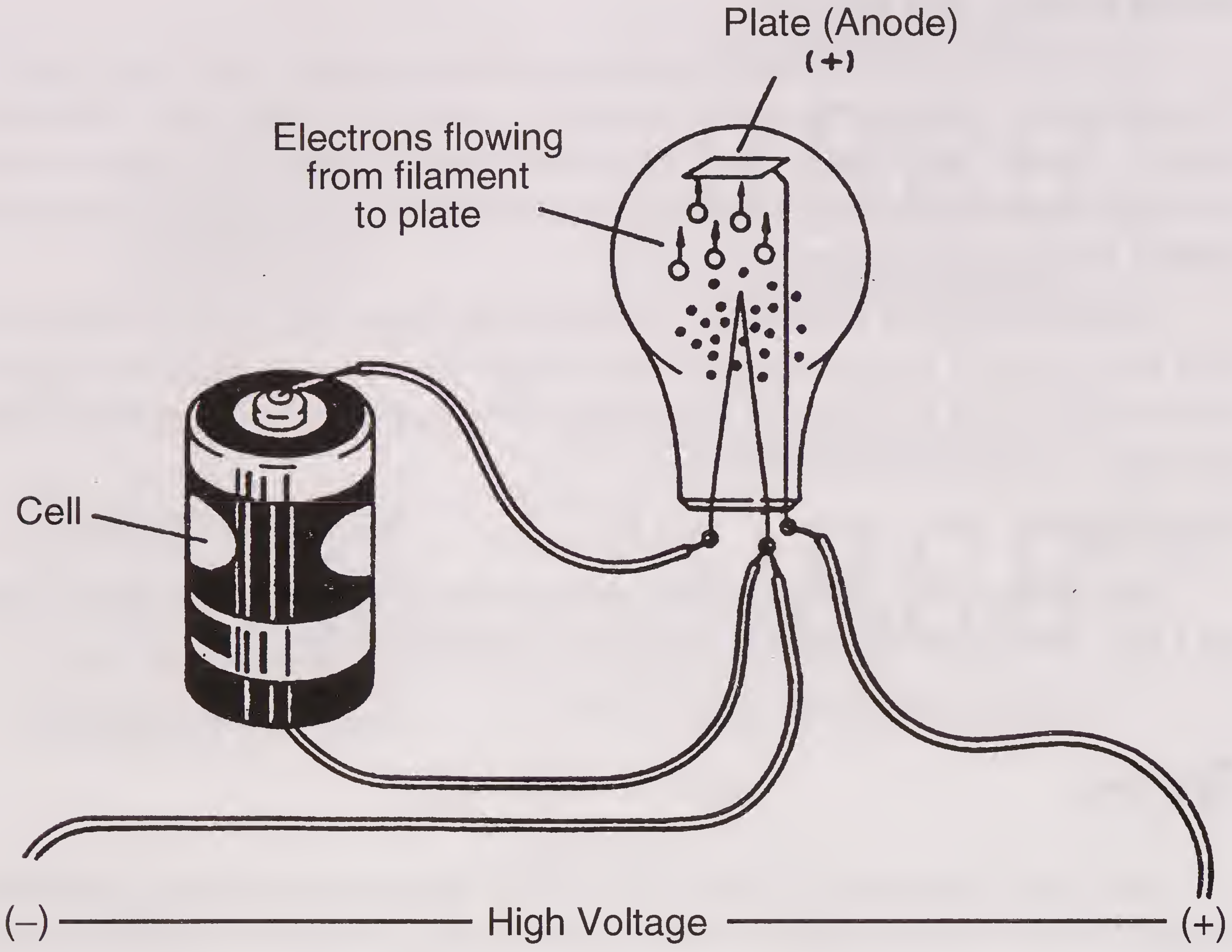
Plate (Anode)
(+)
Electrons flowing from filament to plate
Cell
(–) High Voltage (+)

Figure 167

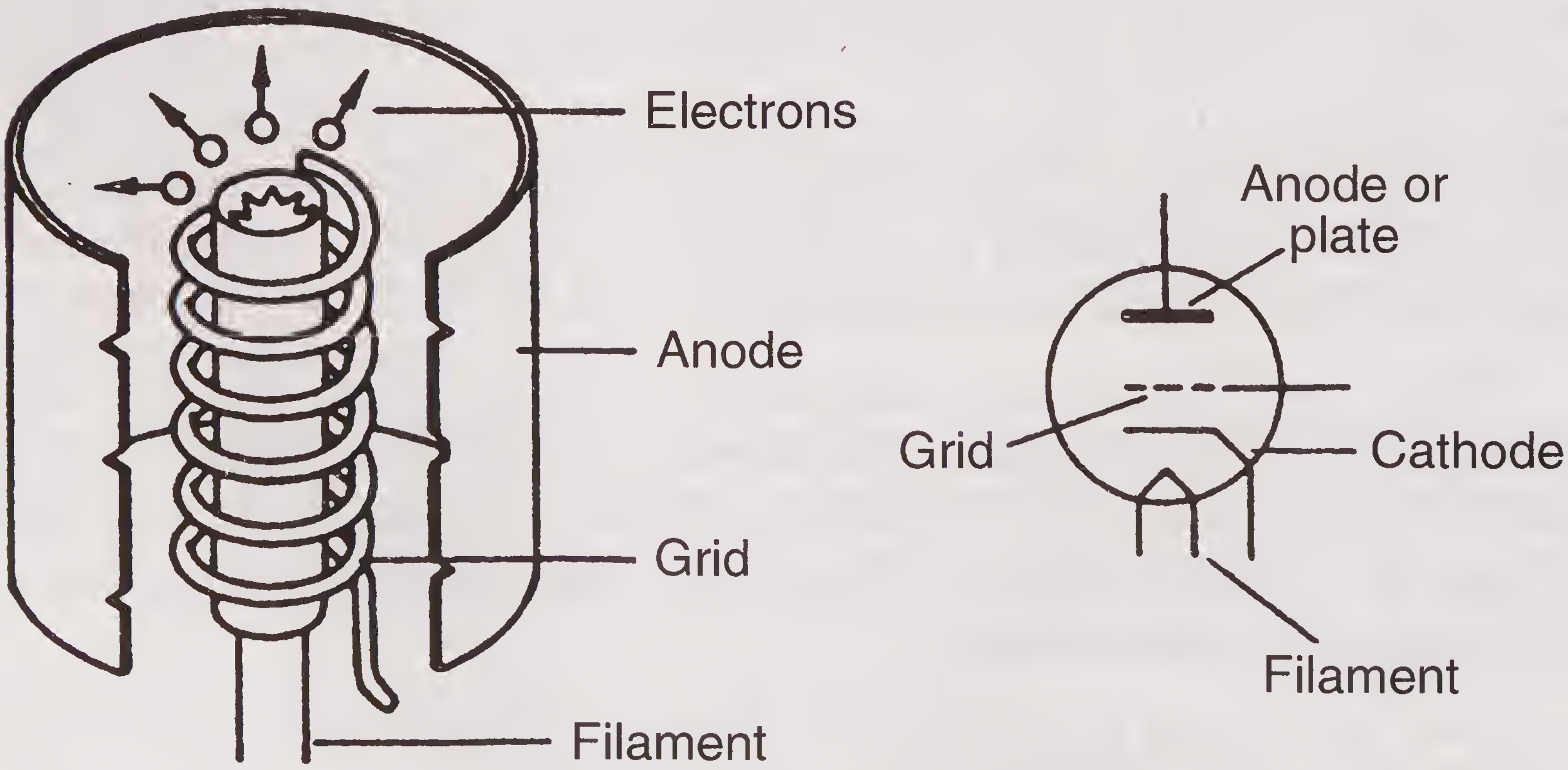
Electrons
Anode
Grid
Filament
Anode or plate
Grid
Cathode
Filament

Figure 168

A voltage applied to this grid can control the electron flow between cathode and anode. For example, if the grid is made negative the electrons are repelled and no current can flow. On the other hand, if the grid is made positive, the electrons are attracted and accelerate to the anode.

The amount of current between anode and cathode can be controlled with precision by the voltage applied to the grid. This device, called a triode, can be considered as an equivalent of the transistor and is used for the same purposes—as a switch or an amplifier.

New tubes were developed from the triode configuration by adding new internal elements. The tetrode (four elements), pentode (five elements), hexode (six elements), and heptode (seven elements) are all examples of tubes that can be found in old equipment like radio receivers, audio amplifiers and TVs.

Although the tubes are very efficient in some applications and even used in some of them until recently, they have some disadvantages when compared with the equivalent modern solid-state devices like transistors.

The tubes need to be heated to operate and this implies the consumption of large amounts of energy. Another problem is that all the heat can be transferred to the surrounding area. The tubes also need high voltage for their operation. Although common solid-state transistors operate with low voltages ranging from 1 to 30 volts in typical applications, in the same applications, the tubes need voltages in the typical range between 150 and 800 volts! Portable equipment using tubes is almost totally impractical.

Symbols and Examples

Figure 169 shows the symbols for many types of tubes found in old equipment (and even some modern) and two examples.

In the old types found in some military equipment, metal enclosures replaced the glass and the multiple tubes containing more than one configuration inside a body (a triode and a pentode, for example).

Specifications

The tubes are identified by a part number. There are two main codes used in their specifications. One of the codes uses a number followed by a

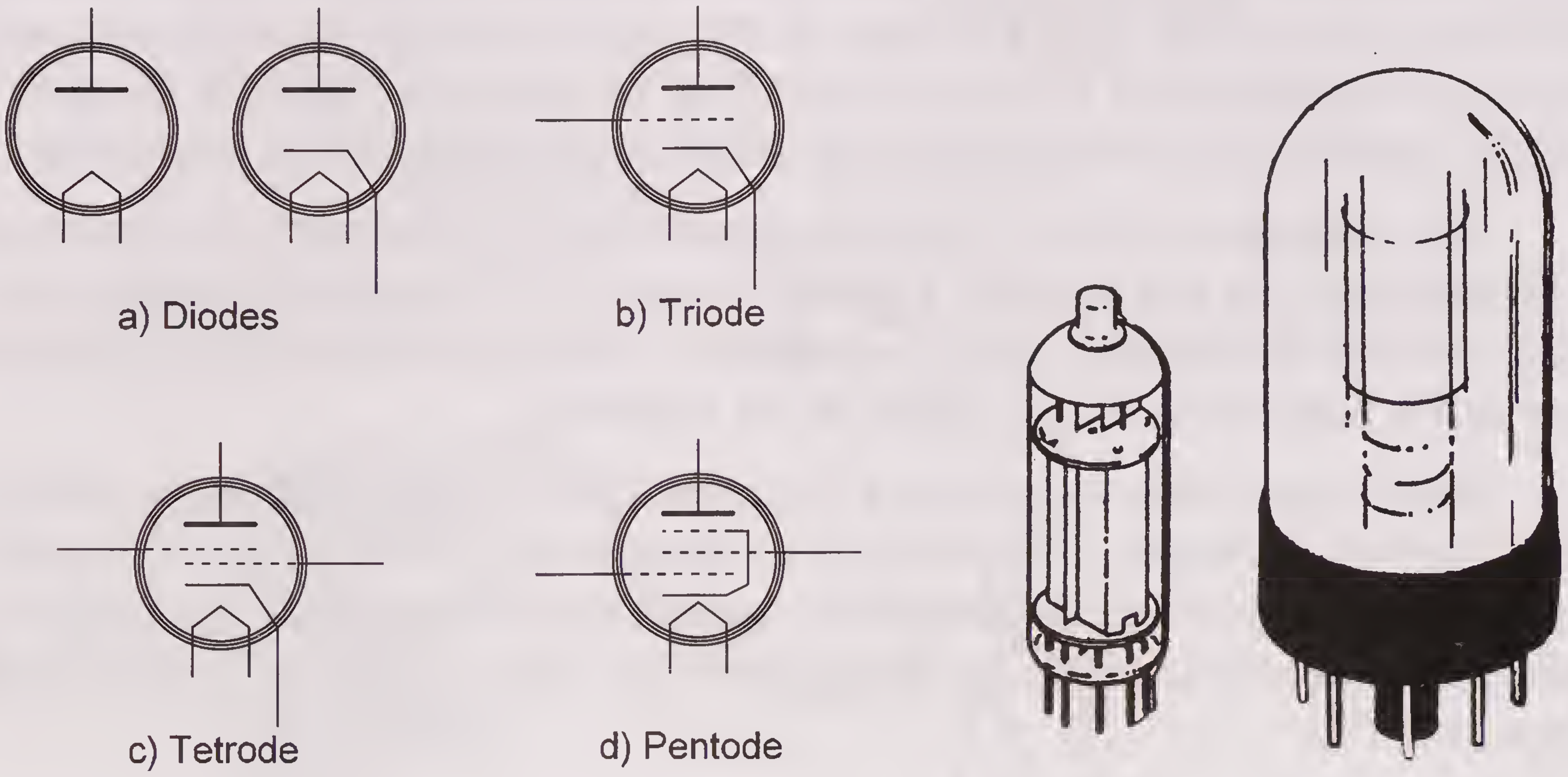
a) Diodes
b) Triode
c) Tetrode
d) Pentode

Figure 169

group of letters and another number. The first number is the filament voltage, or the voltage that must be applied to the filament to heat the tube. The second and third figures indicate the specific function. For example, a 12AX7 is a double triode tube that has a 12 V filament.

Another way of identification is the European code, where a group of letters indicates the function of the tube and the final number the specific type. For example, the tube ECC83 is a double triode (CC) (equivalent to the 12AX7!).

The best way to identify each tube is by consulting a manual. (Prompt® Publications has an excellent replacement guide for tubes, *Tube Substitution Handbook*, Sams 61036, ISBN 0-7906-1036-1.)

Where they are found

The electrician will not often find tubes in modern electric installations. The tubes can be found in older electronic equipment, especially radios and TVs. The tubes can be used in the same configurations as the transistors, as shown in Figure 170.

Notice the tubes need, in addition to the low voltage to heat the filament, a high-voltage source for the anode. This high voltage ranges from 150 to 800 volts in common equipment.

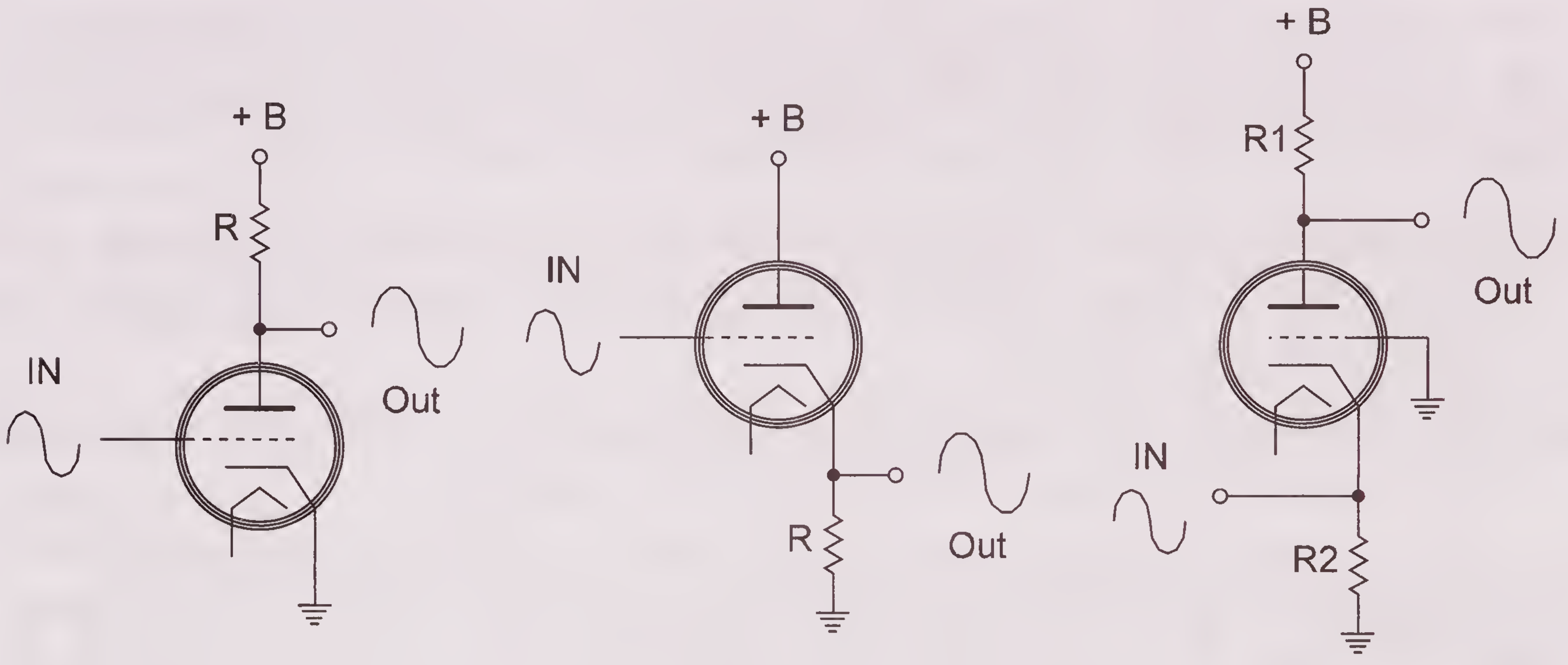

a) Common cathode b) Common anode c) Common grid

Figure 170

Although in some cases it is not a good idea to try to restore old tube equipment, if you like it (perhaps for sentimental value), many dealers have all types of old tubes for replacements.

Testing

When the filament is open (burned out), the tube doesn't function any more. In most of the cases, the malfunction of any piece of tube equipment is caused by this problem. The advantage of the tubes for the technicians is that it is easy to see if a tube is burned out because when the equipment is turned on, there is no orange or red light emitted by its filament. A visual inspection can reveal problems in this case. But, this is not conclusive; some equipment exists where the filaments are wired in series. So, if one of them burns out, all the others are not powered and all the tubes are off.

A multimeter can be used to test the filament of a tube. The operational test needs special circuits or test equipment.

Integrated Circuits

The Integrated Circuit can't be considered a component, but a collection of components housed in a single chip and installed inside a unique case. The basic idea originated from these questions:

1. Why do we have to manufacture all the components of a circuit in a separated process and then assemble them together to create the desired configuration?
2. Why not make them together in a unique silicon chip to perform the basic circuit for equipment using a single process?

The answer to these questions resulted in the integrated circuit or IC. The integrated circuit consists of a tiny silicon chip that houses a determined number of components, such as transistors, resistors, capacitors, and diodes, interconnected by conductive paths that act as wiring. The components and paths are formed by a complex process that includes diffusion, photo printing, and etching.

Figure 171 gives an idea how the many components can be created on a silicon chip.

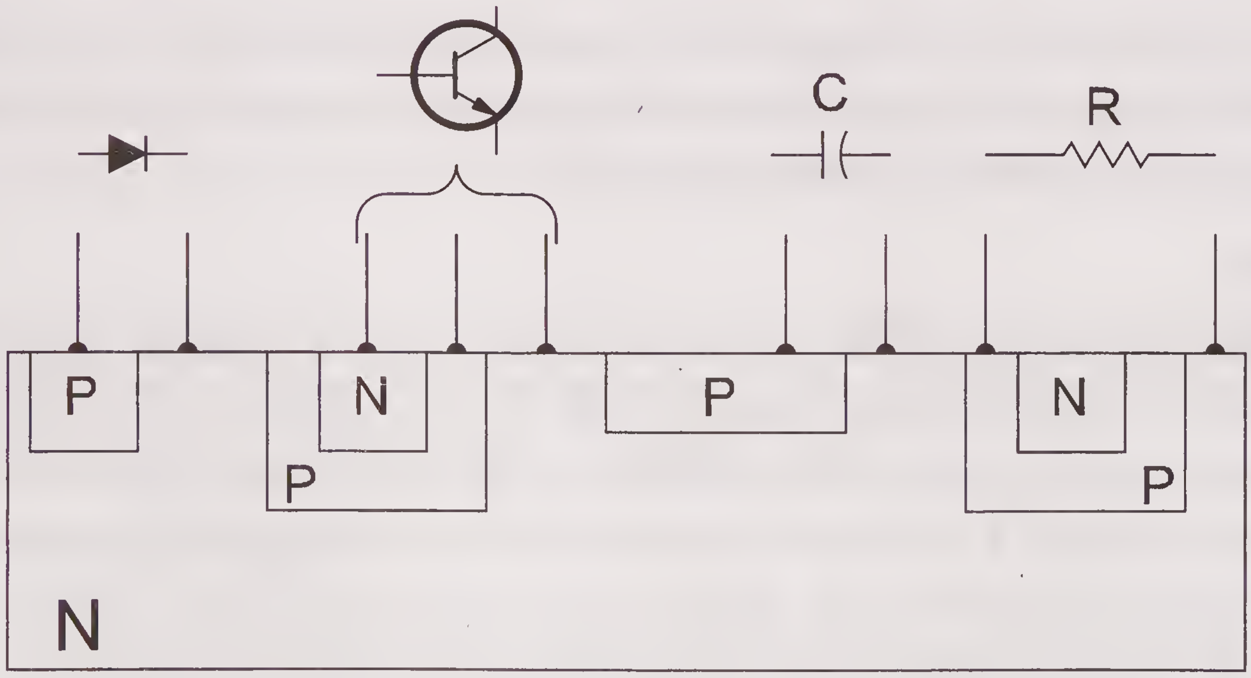

Figure 171

In an N substrate, P regions are created forming a diode. If in the P region of a diode, another N region is formed, a transistor will be created. A capacitor can be formed by a large P region. Having a reverse bias at the junction to create a capacitor, the junction will act as the dielectric and the PN regions as places similar to the varicaps. A resistor can be formed by a P channel formed in the N region. The width and length of the channel determine the resistance of this resistor. The number of components, their disposition in the chip, and the wiring are planned to perform as a complete circuit such as an amplifier, a voltage regulator, or even a computer.

It is possible to house in a single chip, circuits with few components for simple functions to circuits containing more than 7 million components, such as in modern microprocessors used in personal computers. Even a complex integrated circuit with millions of transistors has a tiny footprint measuring only in millimeters.

The principal advantage of the integrated circuit is that the circuit can be manufactured almost complete, needing only a few complementary support devices to perform in the desired equipment.

For example, it is possible to create an IC containing a complete audio amplifier that needs only the input jacks, the power supply, the volume control, and the loudspeaker to perform the complete equipment. The integration has some limits. Some components exist that can't be easily integrated on a silicon chip. For instance, large-value capacitors and inductors are components that can't be found in integrated circuits. So, when planning a complex integrated circuit, if large-value capacitors are needed, they must be external, and the common traditional discrete type must be used.

Although the integrated circuit is very important as a simplified solution for projects it can't form a complete piece of equipment or circuit in many cases, needing completion with external components as capacitors and inductors.

Since an IC can contain practically any configuration with any number of components, they exist in a very large number of types and functions.

Symbol and Types

Figure 172 shows the common symbols adopted to represent an integrated circuit and its various types.

At the side of the IC's symbol or inside, an identification figure allows the technician to identify it.

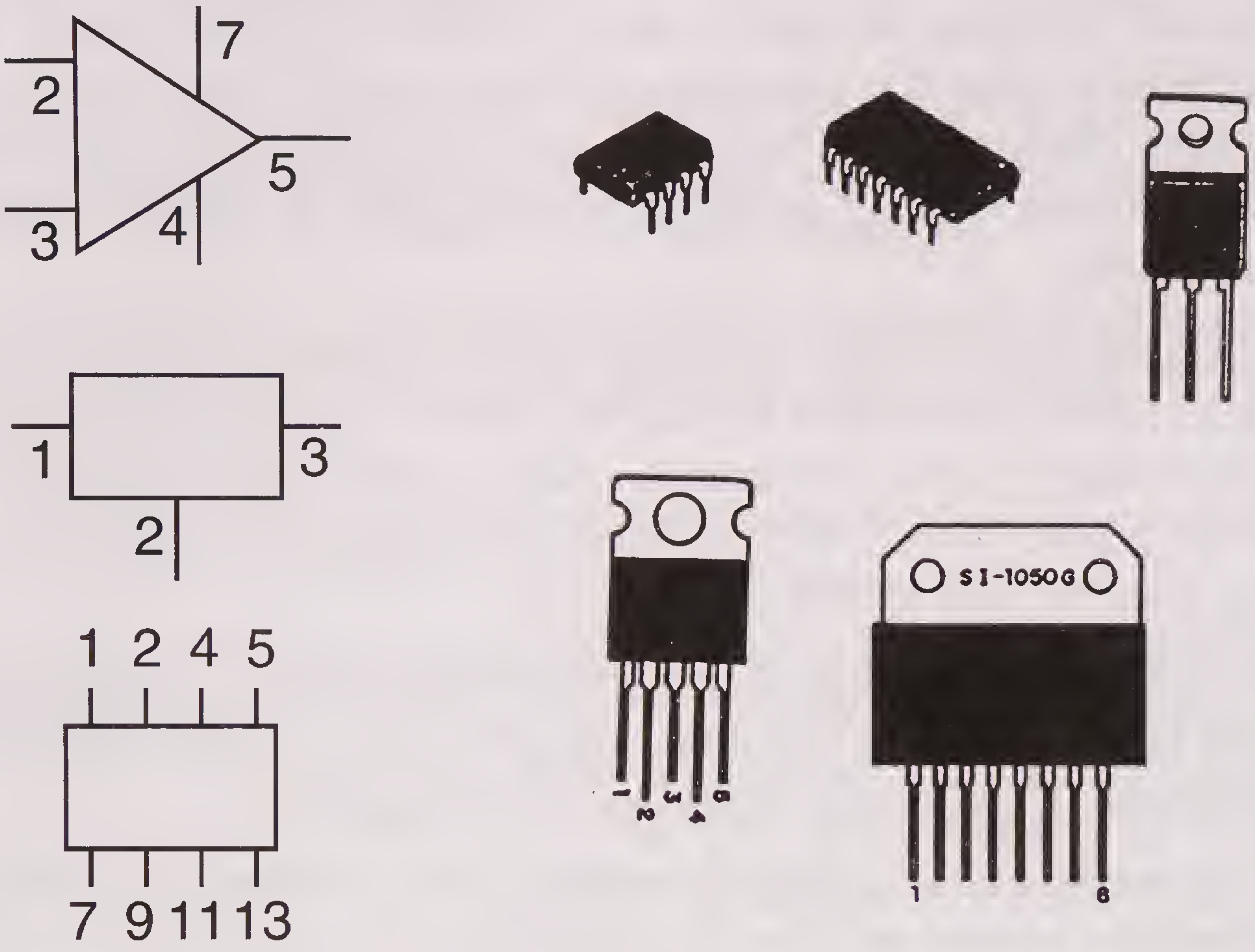

Figure 172

The number of terminals varies according the function. Types with number of terminals ranging from eight to more than 40 are used in common functions. Types with more than 200 terminals are found in special applications such as computers.

Figure 172 shows the most common types of ICs, found in many pieces of electronic equipment that use DIL (Dual In Line) and SIL (Single In Line) cases. See the number of the terminals starting from a mark or point in the case. The identification of the terminals is very important when replacing an IC with another or when trying to troubleshoot. Voltage measurements must be made in these components. Some integrated circuits that perform high-power functions like motor controls, voltage regulators, and audio amplifiers are equipped with tabs or resources to allow them to be on heat sinks.

Specifications and Types

Today we can say that there exist millions of different types of integrated circuits. They are identified by a part number. Only with specific information about the part number is it possible to know what an IC is and what it does. The electrician who wants to work with electronic components doesn't need to have complete handbooks or manuals of all the integrated circuit as it is impossible.

Common functions

Some integrated circuits contain a configuration simple enough and useful enough to allow them to be used in a large number of applications. So, these circuits act as basic blocks in many projects and can be found not only in one kind of application, but in a large number of applications. For example, an integrated circuit that contains a simple low-frequency stage with some transistors and some other components can be used in an audio amplifier, a low-frequency oscillator, an alarm, a motor, and many other circuits.

Many types of these general-purpose or common-function integrated circuits were created and many of them became popular and are used in all the applications we can imagine. Electronics magazines that give projects to hobbyists indicate those ICs in their projects as they can be found in any electronic parts shop and can be handled easily.

The common functions found in ICs are divided in two other groups:

Analog

Analog integrated circuits are the ones that work with DC and analog signals, such as audio, video, AC, and others. We define as an analog quantity, one that can assume any value between two limits. An AC voltage is an analog quantity as its amplitude can assume any value between zero and the maximum amplitude. On the other hand, a digital quantity can assume only determined values between two limits. A byte can assume only the integer values between 0000 0000 and 1111 1111.

In the group of analog ICs we can find some functioning as voltage regulators, operational amplifiers, comparators, audio amplifiers, video amplifiers, timers, PLLs, etc. It is interesting to give a brief description of each of these ICs as they are very common in many appliances and will be found by the electrician in electric installations and in many AC-powered appliances.

Voltage regulators

The purpose of a voltage regulator is to maintain constant the voltage in a circuit. The power supplies of many pieces of equipment use this type of IC to have a regulated output supply voltage. Today a three-terminal, 1-A voltage regulator, such as the 78XX and 79XX series or family (where the XX is replaced by a number indicating output voltage), is very common in appliances. For example, the 7806 furnishes a 6 V output under 1 A of current (maximum). Another type of voltage regulator is the one used in SMPS (Switch-Mode Power Supplies) and the ones found in computers, video monitors, and many other modern appliances.

The operation principles of these regulators are different from the linear regulators as they acts as oscillators, producing a voltage that depends of the frequency and pulse width of the output signal.

Symbols and Types

Figure 173 shows the symbols and some types of this IC. Notice that many of the ICs used as voltage regulators have resources to be mounted on a heat sink. Observe also that other ICs can have more than three terminals.

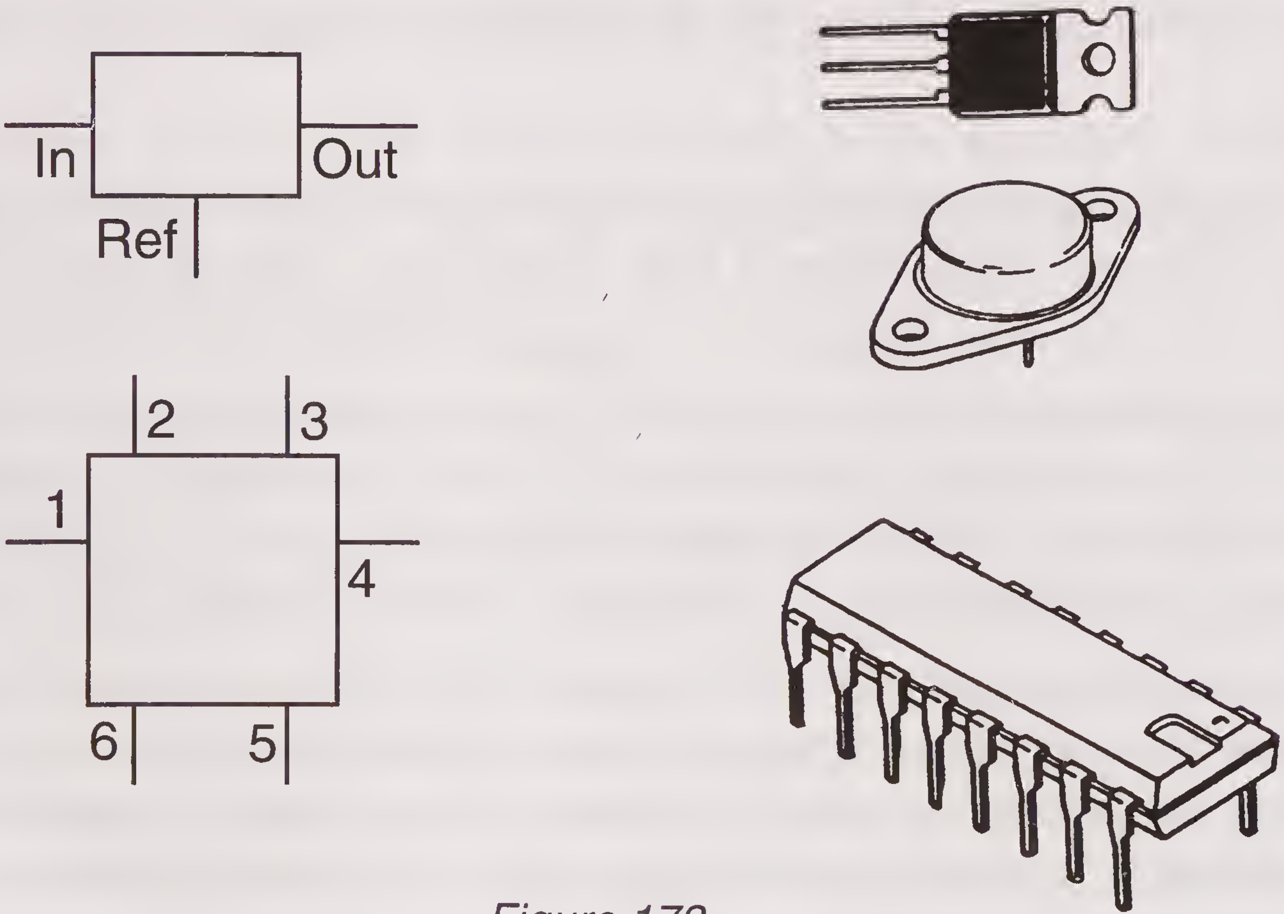

Figure 173

Specifications

The ICs used as voltage regulators, like any other IC, are specified by a part number. The electrical characteristics are found in data sourced by the manufacturer. The main specifications of these kind of IC are:

A. Input voltage range—the range of voltages that can be applied to the input. In the documentation it can be abbreviated by Vin.

B. Output voltage or range (Vout)—Some ICs are fixed voltage regulators. They have an internal zener that fixes the output voltage. Other have resources to be programmed by an external zener or resistors to a determined output voltage. Figure 174 shows a circuit, the LM350, which is an adjustable voltage regulator for a large range of voltages and output currents up to 3 A.

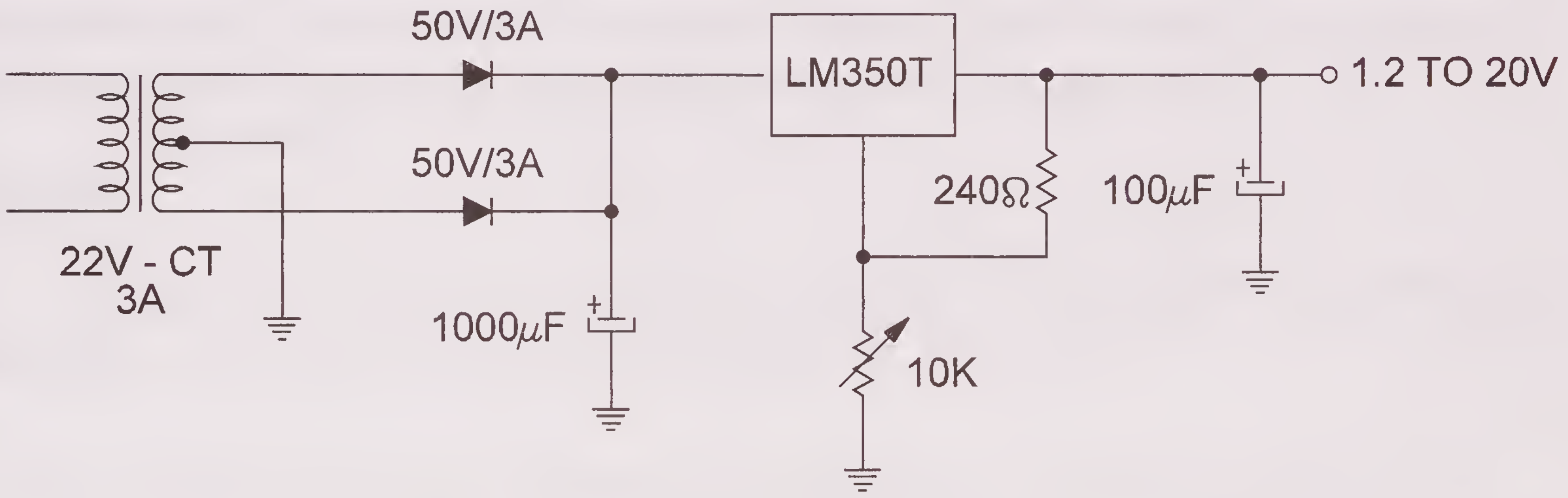

Figure 174

C. Output current—the maximum output current that can be regulated by the IC is an important parameter to be observed when using this kind of IC.

Where they are found

Voltage regulator ICs are present in almost all equipment plugged to the AC power line and even in battery-powered circuits acting as voltage reducers. After rectifying and filtering, they convert the AC power line voltage in one or more DC voltages necessary to power the electronic circuits. These ICs are present in the first stage of any equipment or "power supply" stage.

Testing

The simplest way to test this kind of circuit is measuring the input and the output voltages. The multimeter is the instrument recommended for this task.

Operational Amplifiers

Operational Amplifiers (AO) are general-purpose amplifiers originally designed to perform mathematical operations in analog computers. An operational amplifier has a noninverting input represented by a (+) and an inverting input indicated by a (–). The signals applied to the inverting input appear in the output with the phase shifted, as shown in Figure 175.

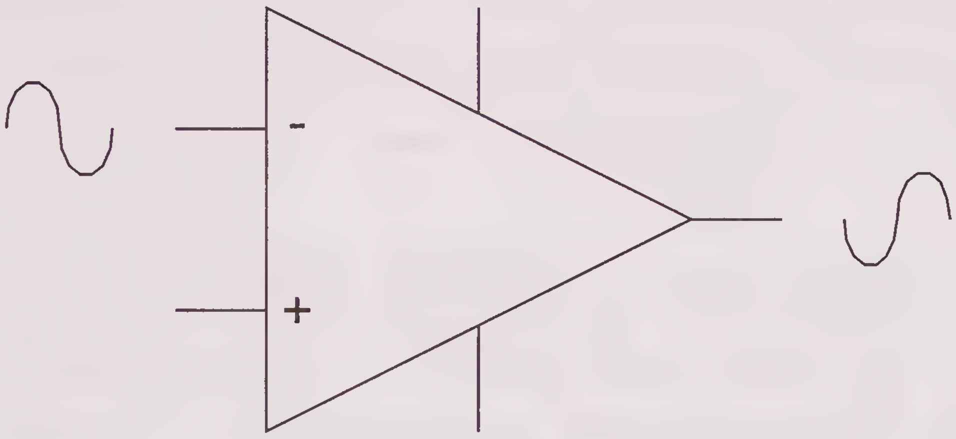

Figure 175

The ideal operational amplifier has an infinite input impedance and a null output impedance. The real types have a very high input impedance of many megohms and a very low output impedance (some ohms to hundreds of ohms).

The gain of an operational amplifier or the number of the times the output voltage is higher than the voltage applied to the input can be programmed by a feedback circuit as shown in Figure 176.

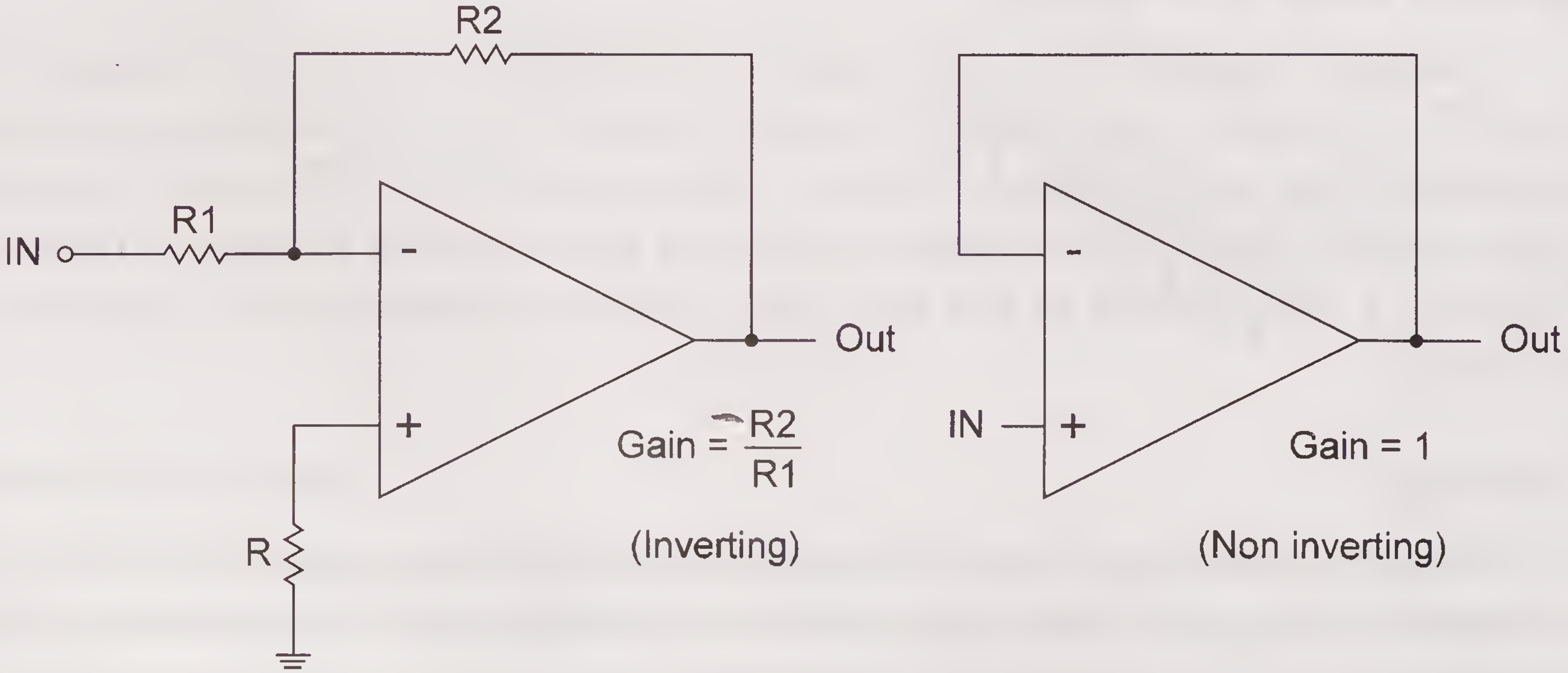

Figure 176

The higher gain is achieved when the feedback is infinite and typically ranges from 10,000 to 1,000,000 in the common types. The lowest gain is achieved when the feedback loop is a zero ohm resistance and is one.

The configuration with a voltage gain of one is called voltage follower. In a voltage follower the output voltage is the same as the input voltage. Despite this, the difference in impedance between the input and output makes the circuit an amplifier with a very high current gain.

Operational amplifiers can be found in a large number of types, sizes, and presenting a large range of electrical characteristics (input and output impedances, power supply voltages, gains, type of transistors found inside as FETs or bipolars, operation frequency range, etc.).

Symbol

Figure 177 shows the symbol adopted to represent an operational amplifier. They generally look the same as other ICs.

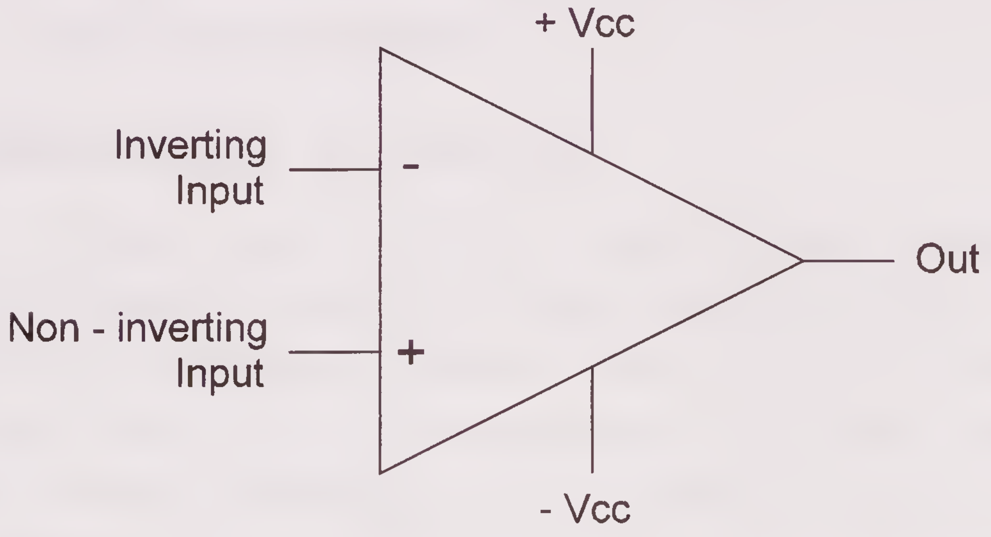

Figure 177

Notice that some ICs exist that contain two or more operational amplifiers. In some cases we can also find terminals where additional components are plugged. These external components are used to add special characteristics to the circuit such as frequency compensation.

Specifications

The operational amplifiers are indicated by a part number according to a code adopted by the manufacturer. Normally a group of letters is used to identify the manufacturer—MC = Motorola, SN = Texas Instruments, LM = National Semiconductor, etc.

From the part number or consulting data sheets and/or handbooks it is possible to access the electrical characteristics of each device. These characteristics are:

A. Open loop gain—The gain is determined by the internal circuit and is specified the maximum gain or the "open loop" gain where no feedback is used. The values can range from 1000 to 1,000,000 in common types. The gain of an AO can be abbreviated by G.

B. Power supply voltage range—This specification indicates the range of voltage that the operational amplifier needs for correct operation. It can range from 1.5 V to more than 40 V.

At this point it is interesting to observe that in many applications the AO needs a symmetrical or double power supply. It is a power supply where a positive and a negative voltage referred to the ground (0 V) must be generated as shown in Figure 178.

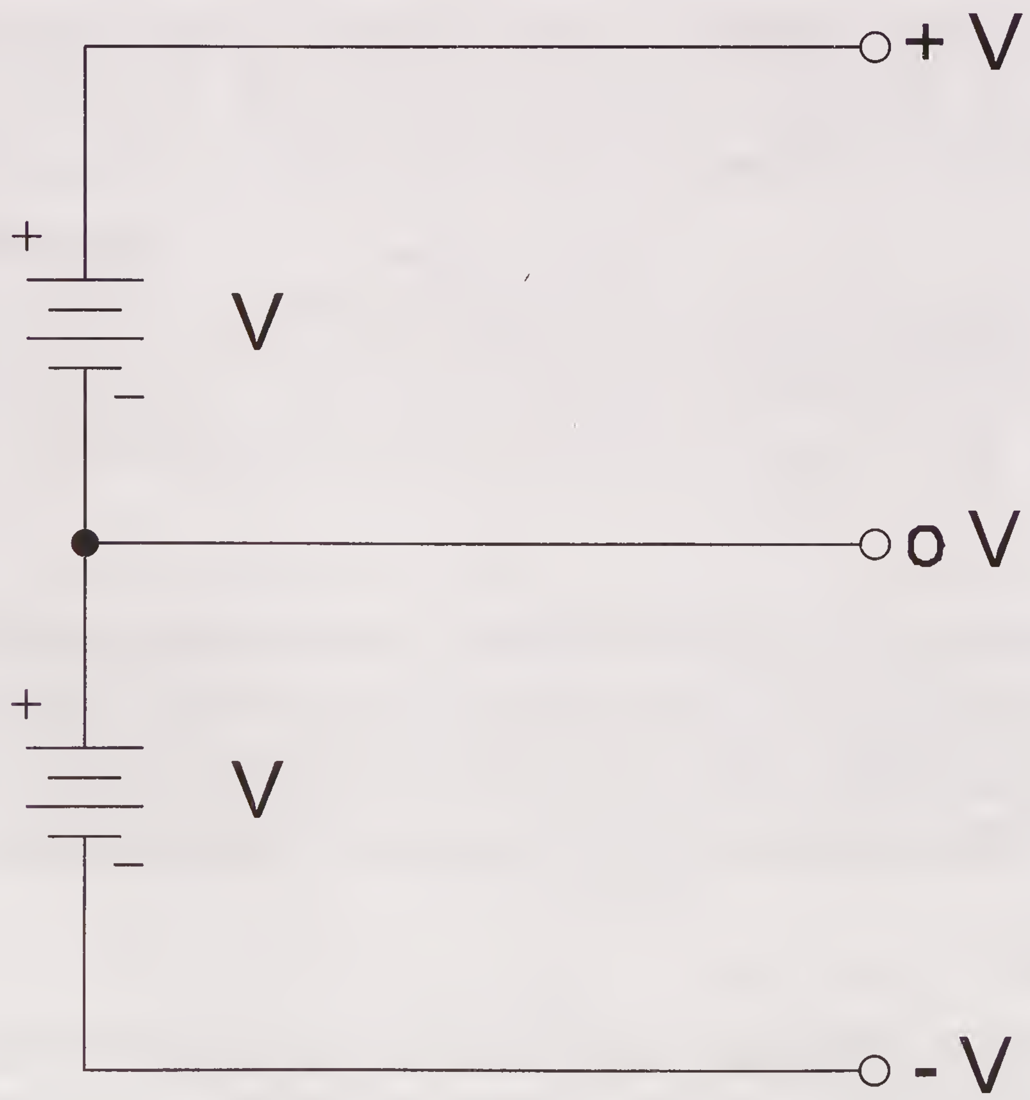

Figure 178

In some documents, data sheets, and diagrams it is common to refer the power supply voltage for an AO as 6-0-6 V, +6, -6, 0 V, +6/-6/0 V or 6+6 V if it needs a symmetrical supply as the one shown in Figure 178.

C. Gain x Bandwidth—The gain of an operational amplifier falls at the same time the signal frequency increases. The frequency band is limited by the point where the gain falls to one; this determines the gain bandwidth product. Common operational amplifiers can operate in frequencies from 1 or 2 MHz to up to some hundreds megahertz.

D. CMRR—If two signals with the same amplitude and phase are applied, one to the inverting input and the other to the noninverting input of an operational amplifier they must be cancelled and the output will be zero. In practice, due to small differences in the internal components, this is not possible and some part of the signal is not cancelled, which then appears in the output. How efficient is the AO when canceling a signal in this operation mode is given by a specification called Common Mode Rejection Ratio or CMRR. This specification must be as high as possible and is indicated in decibels (dB). Common types have CMMR as high as 90 dB and more.

Where they are found

Many appliances use operational amplifiers in the integrated circuit form. They normally are common types such as the 741, CA3140, TL080, TL081, and many others that are easily found in any electronic components shop. Special operational amplifiers can be found driving step motors or other devices, but in general, they can be found at a dealer by part numbers, too.

The basic function of an operational amplifier in many circuits is the amplification of signals coming from sensors, transducers, or other sources, driving them to powerful output stages or circuits that must work with signals having a larger amplitude.

Testing

In many applications the operational amplifier can be tested measuring the voltages in their pins. If the voltages are altered, and the components plugged into it are good, the problem will be certainly in the OA. Another way to test is using a special probe circuit.

Audio Amplifiers

Complete audio amplifiers with output powers ranging from some milliwatts to more than 100 watts can be found in the form of a linear integrated circuit. These integrated circuits consist, in some stages of amplification, of the preamplifier, driver, and output stages interconnected to perform as an audio amplifier.

The high-value capacitors necessary to many of the stages are connected to the circuit by external pins. In many cases, the manufacturers can use the circuit with a minimum of external capacitors.

Many audio amplifiers can have resources to program the gain or the frequency response by some external components. In the simplest types we have an input where the volume control is placed, an output where loudspeaker or headphones/earphones are connected, the power supply pins, and some other necessary parts, such as capacitors (coupling and decoupling) and resistors (to program gain). Many manufacturers have types containing two amplifiers in one case, making the project of stereo equipment easier.

Symbol and Example

Figure 179 shows the basic symbol adopted to represent an audio amplifier. The cases are SIL or DIL or another type based on the output power.

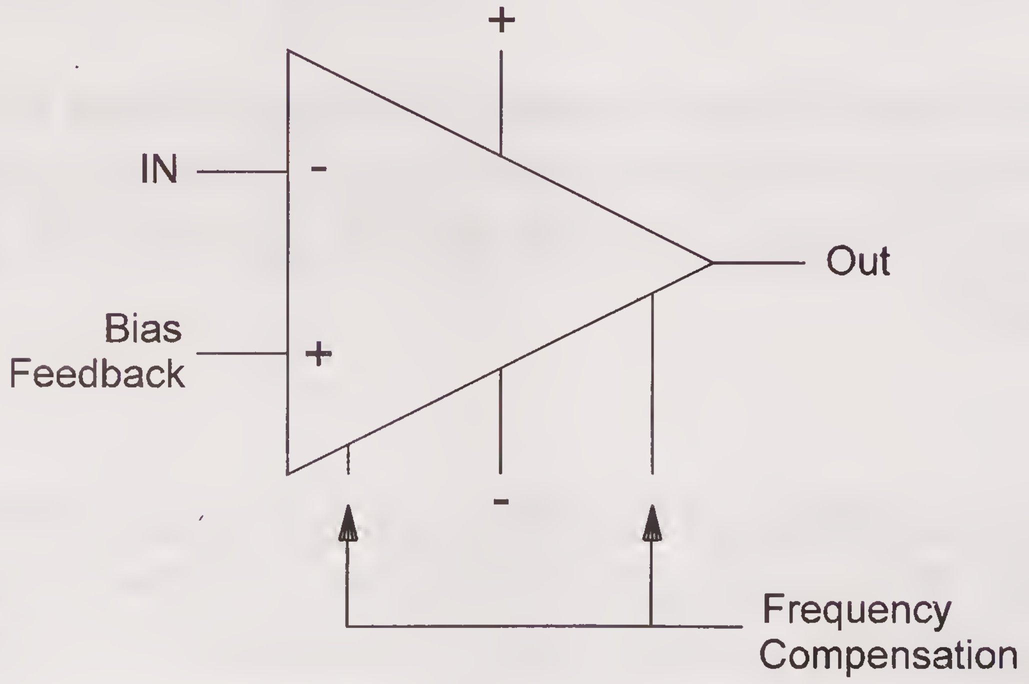

Figure 179

Specifications

The ICs of this group are also identified by a part number according to the manufacturer. The electric specifications of an audio amplifier in the form of an integrated circuit are the same for any other audio amplifier.

A. Output power—the amount of power that the circuit can drive to a loudspeaker. It is indicated in watts (W) and in some cases abbreviated by Po or Pout.

B. Power supply voltages and current—important when designing the power supply for the circuit. It is given in volts.

C. Output impedance—characteristics of the loudspeakers/earphones connected to the output of any audio amplifier are very important to not overload. Impedance is given in ohms and typically is between 2 and 16 ohms for types designed to drive loudspeakers and higher for earphone/headphones amplifiers.

D. Input impedance—given in ohms.

E. Gain—how many times the amplitude in volts of the output signal is higher than the amplitude of the input signal. This specification is important for the input impedance to determine the sensitivity of an amplifier.

F. Operation curves and other specifications—performance of an audio amplifier according to the signal frequency is given by curves. These response curves show how the fidelity of an amplifier is when working with signals of the audio spectrum. These specifications are important when making a project. For the technician who needs information to find a replacement, the part number is enough. It is important to call the reader's attention to the fact that amplifiers with the same power in the cases can be different in disposition of pins and the external components meaning that they are not interchangeable.

Where they are found

Audio amplifiers are found in every part and a large quantity of common AC-powered appliances, embedded in electric installations, or even powered from batteries, but all related to applications in the electric installation. For instance, intercoms, alarms, doorbells, wireless telephones, wireless intercoms, baby monitors, and ambient sound sources can use an audio amplifier (or several) in the form of an integrated circuit to drive an earphone or one or more loudspeakers.

When creating the equipment the designer must instead use a dedicated integrated circuit where all the functions are embedded in a unique device and can use some common integrated circuits in some functions as the audio output.

Testing

The best way to test a common integrated circuit, such as an audio amplifier, is by measuring the voltages in their pins. It is recommended to do this measurement only after verifying if the associated components (external capacitors, resistors, etc.) are good. For this, it is important to have an idea about the magnitude of voltages to be found or to have the schematic diagram of the equipment where the voltages are indicated.

Timers

Another important group of the linear integrated circuits family is the one formed by timers. These are circuits that can be used to generate a signal or to count a time interval. Basically, they are circuits that can be used in two operation modes: monostable and bistable.

In the monostable mode, once triggered, the circuit produces a signal during or after a time interval triggering something or controlling something as suggested in Figure 180.

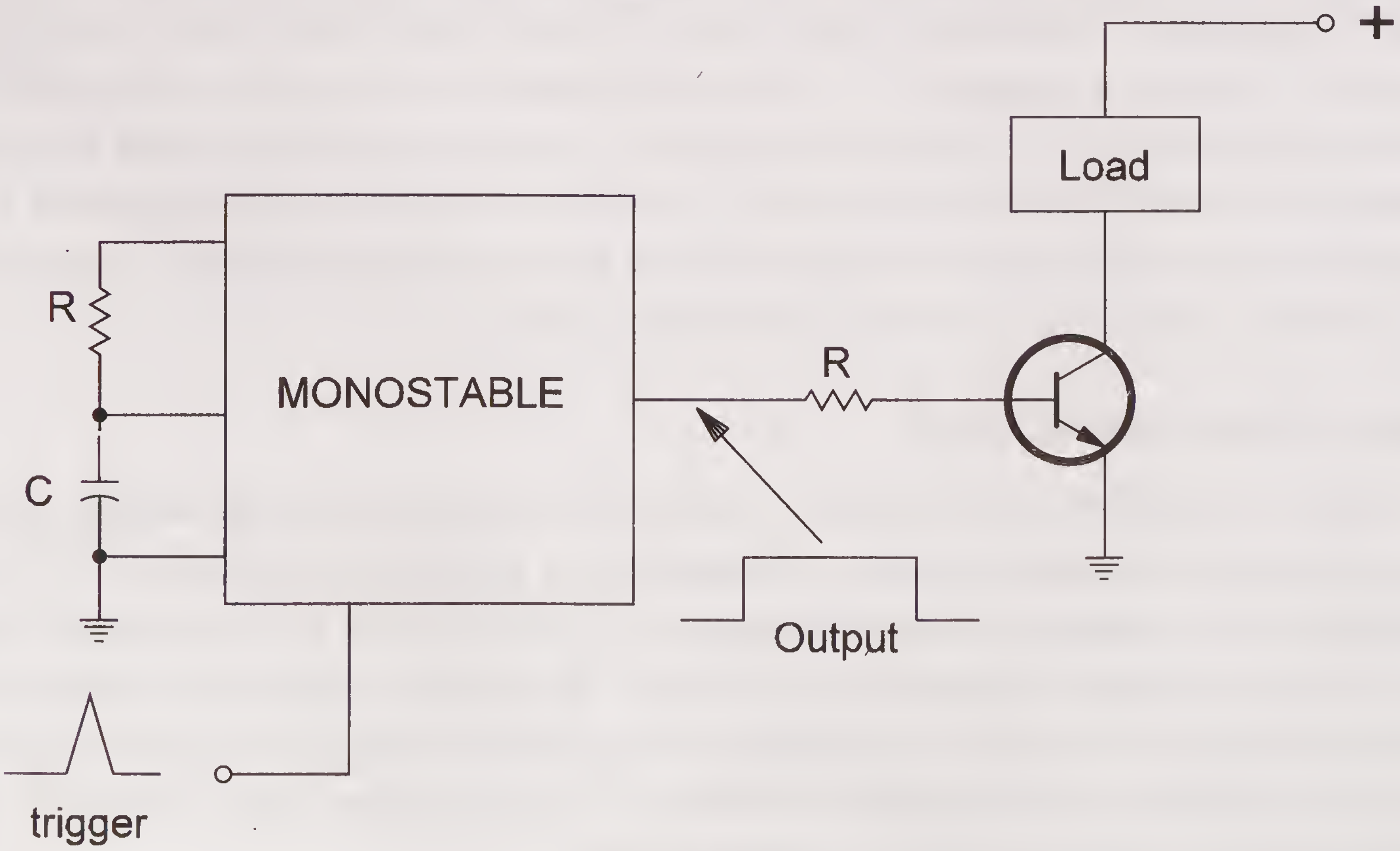

Figure 180

In the unstable mode, the circuit generates a square signal with a frequency dependent on some external components, normally an R-C network as shown in Figure 181.

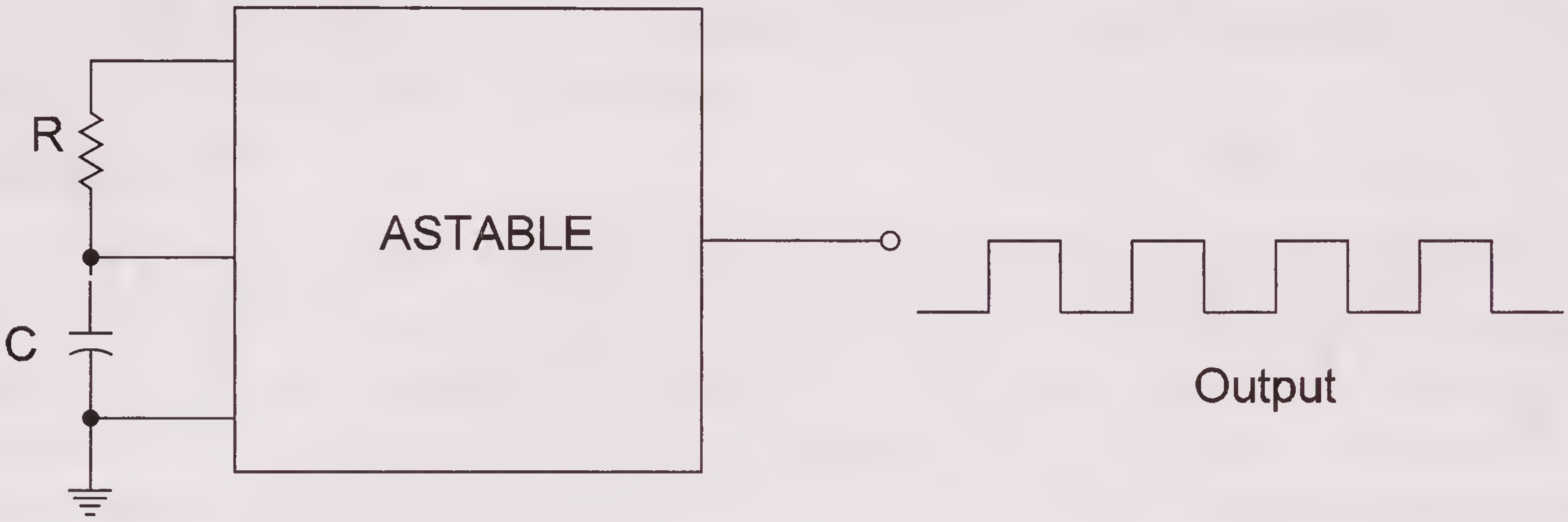

Figure 181

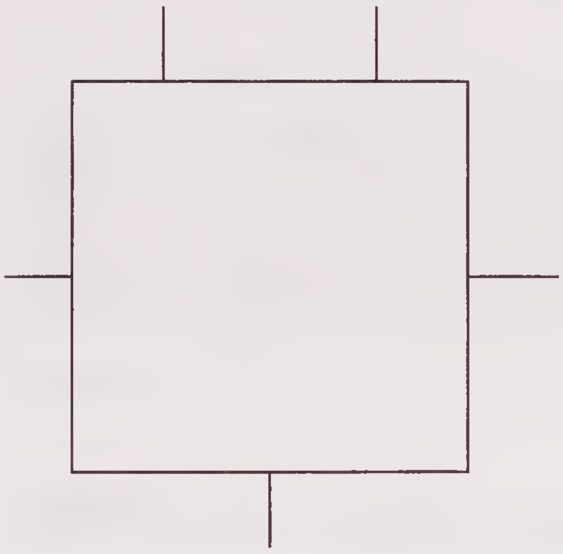

Figure 182

Symbol

Figure 182 shows the symbol used to represent this function. The ICs of this family are found in DIL cases and others.

The 555

The most common of the timers is the 555 that can be found with manufacturer designations such as NE555, TLC7555, LM555, and MC555.

This integrated circuit can operate in two modes: unstable and monostable depending only on the way some external components are connected as shown in Figure 183.

The bipolar version (where the internal transistors are bipolar types) can provide time intervals up to one hour and oscillate in frequencies up to 500 kHz. A version using JFET transistors (TLC7555) offers larger time intervals and produces higher frequencies with lower current consumption. The meaning of 555 to electronics can be evaluated by a recent report that said that more than 1 billion units of this device type were sold since it launched.

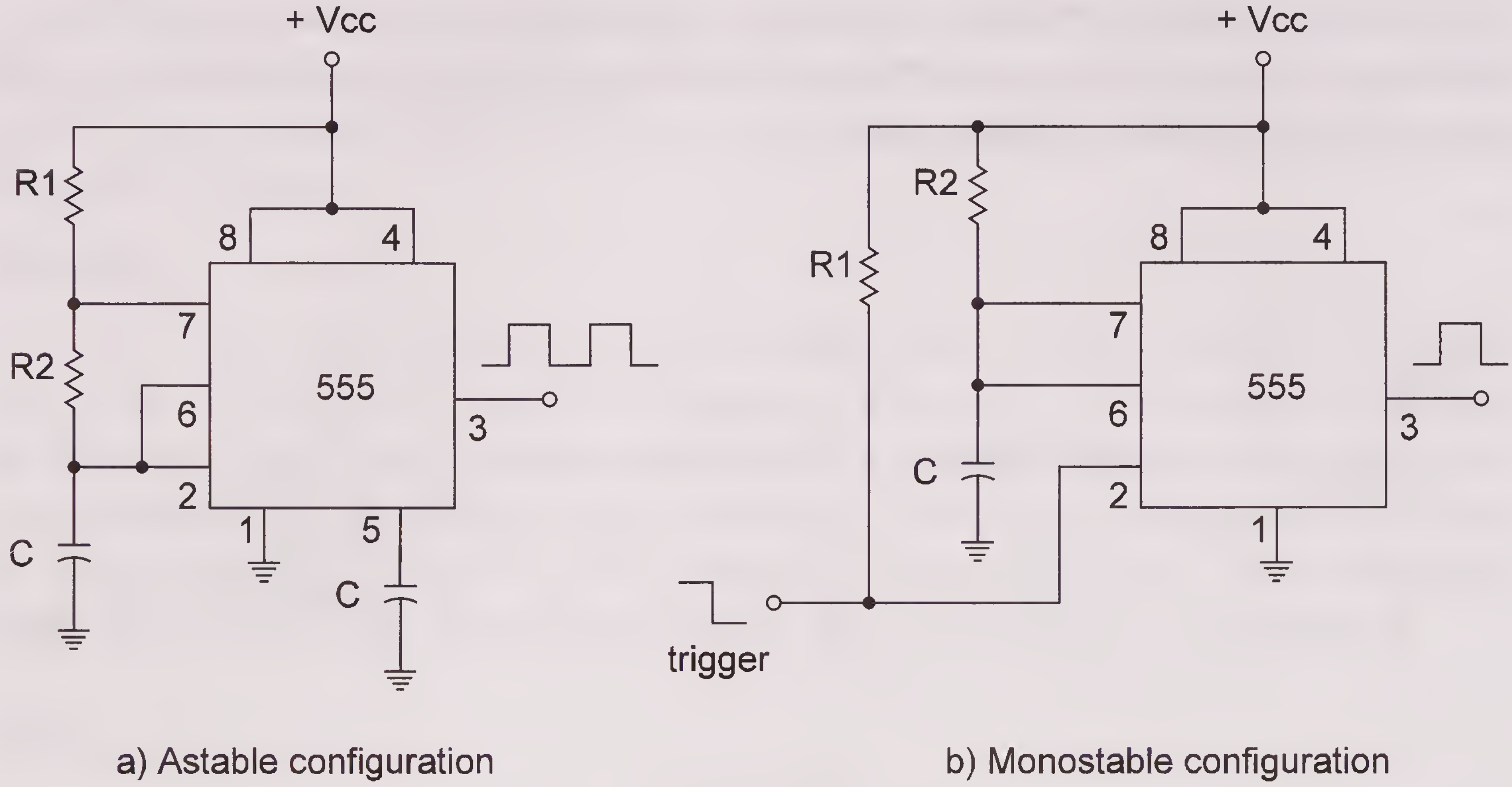

a) Astable configuration b) Monostable configuration

Figure 183

Specifications

The timers in the form of integrated circuits are specified by a part number just as others. The electrical characteristics can be found in the manufacturer's documentation or the schematics of the equipment where they are being used.

The main electrical specifications are:

A. Power supply voltage—the given power supply voltage range where the circuit can operate.

B. How to calculate the time interval or frequency—In the applications, the external components (R and C) determine the operation frequency. A formula or curve allows the designer to find the correct value for the components to any application.

C. Output current—This is an important piece of information, as it says what can be controlled from the IC and how it can be done. In some cases, a transistor stage must be used to control something with a large current consumption .

D. Operation modes—Many timers have different operation modes. They are given in the specifications with information about the way the external component must be connected to perform.

Where they are found

There are many small devices found in electric installations that are based in timers like the 555, TLC7555, and many others. When you press a pushbutton in the wall and the lamp stays on for a certain time interval, most likely the circuit controlling this includes a timer integrated circuit like the 555 or one of the same family.

Testing

As in many other integrated circuits, there isn't a special procedure to test them. Any integrated circuit is a different circuit and specific procedures should be developed to make an effective test. The best way to know if the IC is good is by reading the voltages at its pins. Start with the idea that all other components are good, and compare voltages with specifications or indications in a diagram.

Phase Locked Loop—PLL

Phase Locked Loop, or PLL, are circuits that can be used in many important applications related to electric installation, home appliances, and cars. A PLL is a circuit that can recognize a signal of a determined frequency acting on a relay or other device. If the signal is modulated in frequency, the PLL can also be used to extract the modulation signal. Many types of PLLs in the integrated circuit form are found.

Symbol

Figure 184 shows the symbol of a PLL. The body is not different from any other linear function described here. The only way to know if the IC is a PLL is by its part number displayed as a number on the component's body.

Specifications

As any other IC, the PLLs are indicated by a part number. Using the part number, it is important to know some electric specifications when using a PLL or when trying to find a replacement type. The main specifications are:

A. Power supply voltage range—the voltage range in which the device functions.

B. Frequency range—a very important specification as it indicates the maximum frequency of a signal that can be recognized by a PLL or generated by it when operating as oscillator. Common general purpose PLLs, such as the 556, can operate at frequencies up to 500 kHz.

C. Sensitivity—the minimum amplitude of a signal that can be recognized by a PLL.

D. Output current—this speci-fication indicates the amount of current the device can source or drain when driving a load.

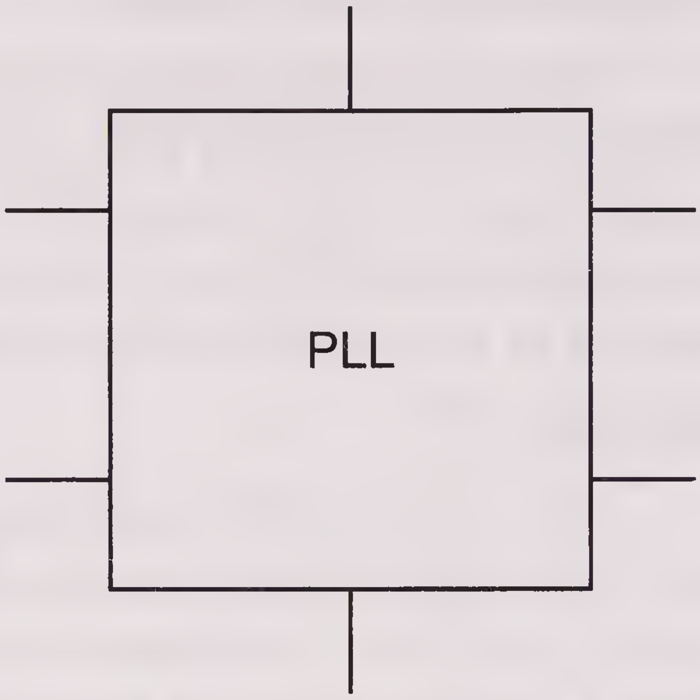

Figure 184

Where they are found

In many modern applications, the PLLs are included in more complex ICs. But, some exist that are for a simple application made by small manufacturers, or even made by amateurs that use the PLL as a single chip. In these case some common types, like the 4046, LM567, and others, can be found.

The next applications can give to the reader of how useful are the PLLs.

A. Remote control—used to recognize a tone sent by the transmitter (using a radio signal or IR) and act on a device (the motor of a garage door, or a function of an electric or electronic appliance)

B. Intercoms—Some types use the AC power line to send the sounds. The sound picked up by a microphone is used to modulate a circuit producing a frequency-modulated signal that is applied to the AC power line. A PLL in the receiver detects this signal and applies it to an amplifier.

C. Alarms—a tone is produced by a circuit and applied to a PLL via sensors. If any sensor is activated the PLL detects the absence of the tone and triggers the alarm.

Testing

The same procedure as recommended for others ICs is valid here: measure the voltage at the pins and compare it with the expected values.

Other Linear Functions

Many other functions can be found in linear ICs. Besides those described before, the reader can find ICs containing video amplifiers, comparators, voltage references, radio receivers, etc. There isn't a specific symbol for these functions and their aspects are the same found in other ICs.

Digital

The digital integrated circuits form a special group of devices with a large range of applications in modern electronics. The basic idea comes from digital electronics. Some fundamentals of digital electronics are important to know to expand information about the devices of this important group. Digital electronics start from the idea that any quantity can be represented by zeroes and ones (0 and 1).

Using only two digits to represent any quantity results in a numeration system named binary. Figure 185 shows how a number in the decimal system can be converted to the binary system.

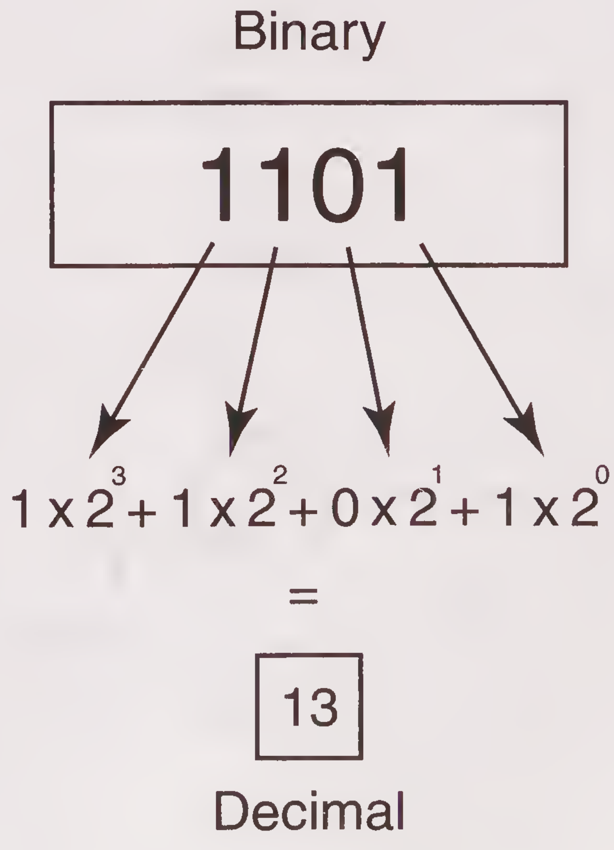

Figure 185

The advantage of the use of binary representation of any quantity is that it can more easily be handled by electronic circuits. In the simplest form, an open switch can be used to represent a zero and a closed switch can be used to represent a 1 as shown in Figure 186.

Working with opened and closed switches it is possible to make any mathematical or logic operations with numbers. The rules that command these operations were first established by Robert Boole two centuries ago. Moving now to the idea that a transistor (and some other components),

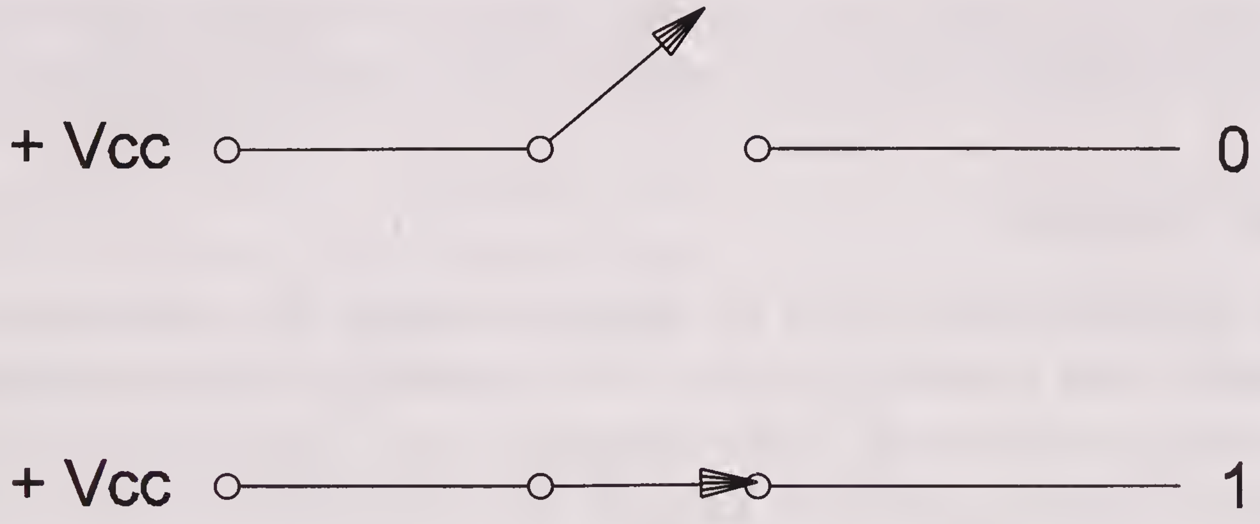

Figure 186

also act as a switch, it can be used not only to represent a zero or a 1 but to execute operations with binary numbers.

As shown in Figure 187, if a transistor is switched to a saturated state (conducting) the voltage in its collector falls to zero or a value near zero, allowing it to represent a low logic level or zero.

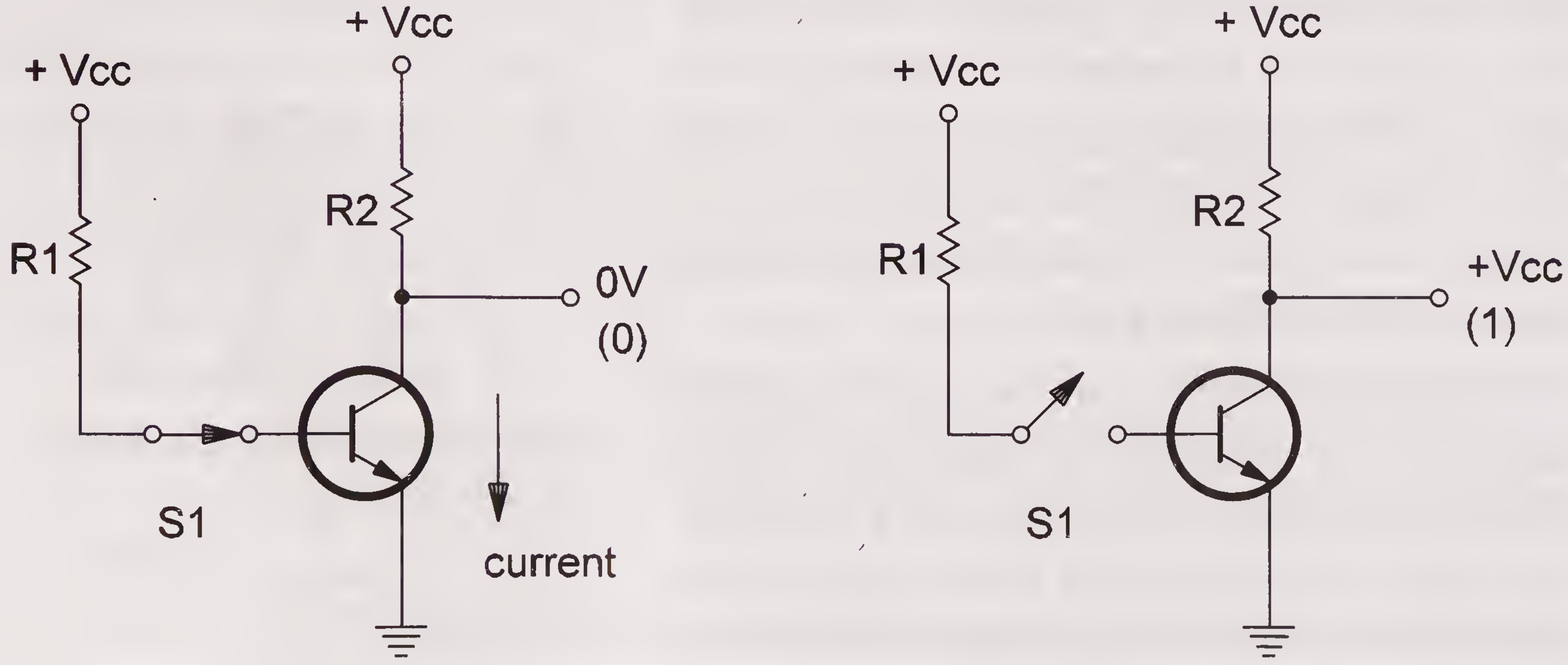
+ Vcc
+ Vcc
R1
R2
0V
(0)
S1
current
+ Vcc
+ Vcc
R1
R2
+Vcc
(1)
S1

Figure 187

On the other hand, if the transistor is switched off (cut), the voltage in its collector rises to a value near the power supply value. This voltage can be used to represent a high logic level or a 1.

Using many parallel transistors any complete number can be represented, and commanding them with appropriate signals, we can execute mathematics or logic operations.

A single digit, 1 or 0, that is used to represent a binary number, is called a bit and a group of eight bits is called a byte. A group of four bits is also called a nibble. A byte can be used to represent any quantity or decimal number between 0 (0000 0000) and 255 (1111 1111).

Thousands, and even millions, of transistors wired as switches can be placed in a chip in configurations that can perform any operation with binary numbers processing the bits to perform final bytes. They are digital circuits like the ones found in computers, microprocessors, microcontrollers, etc.

A certain number of basic functions are used in the projects of digital circuits. These functions are represented in Figure 188.

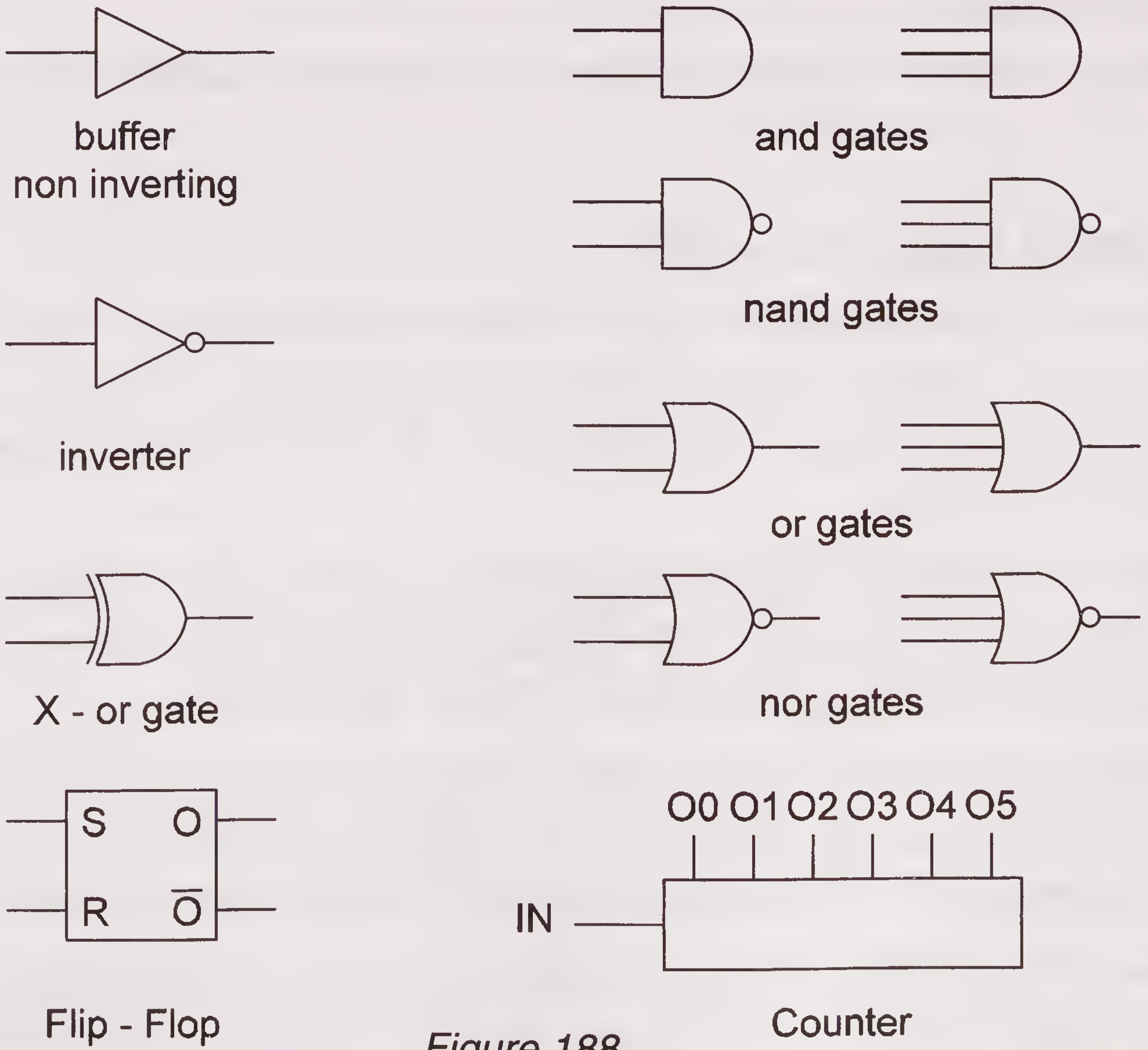

Figure 188

In the first group of basic functions are the gates. Gates are circuits that combine two or more logic levels or digital signals given as a result of logic level or signals determined by rules. The rules follow the mathematics created by Boole. In the second group are more complex functions such as flip-flops (that can store a bit), multiplexers, counters, and decoders. The readers who want to know how a computer and digital circuits work must study digital electronics and Boolean algebra.

In the first computers, like the ENIAC (1942), operations with the bits were done by circuits using tubes. Some time later, the tubes were replaced by transistors. The next step was the use of a large quantity of integrated circuits, each containing many transistors, placed in boards. Now almost all the circuits are inside a chip (one integrated circuit) forming processing units called microprocessors.

Although all the digital circuits are now inside the chip of a microcomputer or other advanced digital circuit, there are applications where integrated circuits performing simple operations or functions must be used. These circuits form families of digital integrated circuits were they all have matched characteristics allowing for interconnection directly to form more complex devices.

Today, a lot of equipment uses integrated circuits of two main families, the TTL and the CMOS.

Transistor Logic IC—TTL

The TTL or Transistor Logic IC family is formed by a large number of devices (near 1,000) with common characteristics. They are all powered from a 5 V power supply and the output of one can be plugged directly to the input of another.

The standard or normal family of TTL devices comprises a large number of functions as gates, flip-flops, counters, decoders, multiplexers, demultiplexer (MUX and DEMUX), PLLs, and timers.

Some subfamilies add to this group of ICs' special characteristics, such as the low-power Schottky (LS) that is compatible with the circuits found in computers adding to the devices' low consumption and high speed.

The TTL devices are easily recognized by the designation "74" beginning their part number. Sometimes the TTL family is also called 74 family or 74xx. Almost all devices in this family are designated by a 74 followed by a number according to their function (7400, 7492, 7191, etc.). The subfamilies are recognized by a letter or two after the 74. For example: 74LS04 (low-power Schottky), 74H93 (high power), 74S121 (Schottky), etc.

CMOS

The other important logic family of ICs is the one formed by the CMOS integrated circuits. These circuits perform the same functions found in the TTL but with different electric characteristics. They can be powered from supply voltages ranging from 3 to 15 volts or 3 to 18 volts. They are not as fast as the TTL, but their consumption is lower and their input impedance is very high.

The devices of this family are designated by a "40" followed by a number that says what the device does. So, we find designations like 4004, 4017, 4093, and 40121. Some CMOS subfamilies use the same designation or number of the equivalent TTL function. These devices are indicated by the 74 code but with a C to indicate it is a CMOS (i.e., 74C00).

When working with TTL or CMOS circuits, the technician needs a digital CMOS or TTL handbook or databook to reference what the function of the IC is. The important fact to be observed when working with logic circuits is that all of them work with pulses or logic levels. If the circuit operates with logic levels in the nonsynchronized mode, in the inputs and outputs you will find only two possible voltages: 0 V or the power supply voltage (5 V in TTL or between 3 and 15 V in CMOS circuits), as shown in Figure 189.

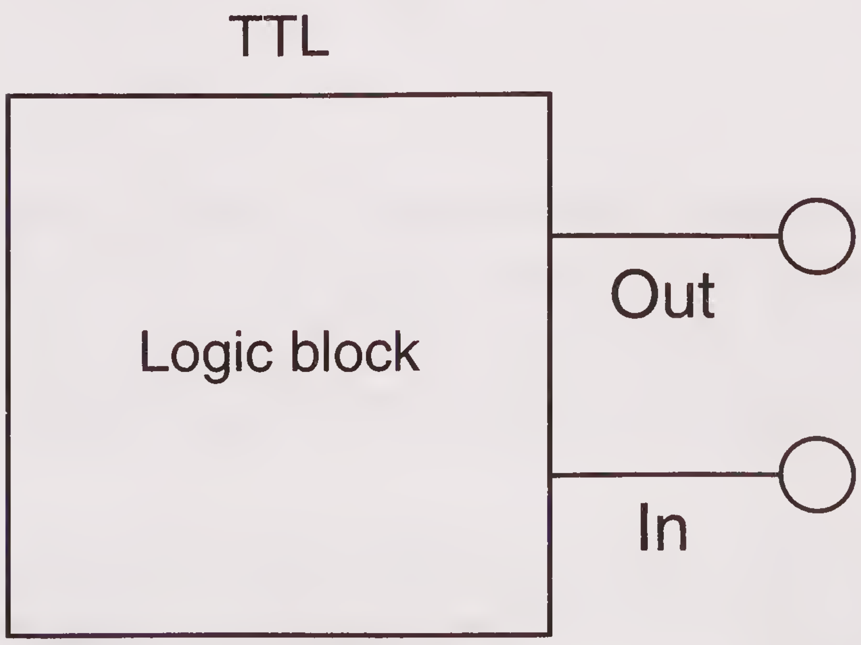

Logic level	Voltage
0	0
1	5v

Figure 189

If the circuit is synchronized by a clock, the signals are square pulses as shown in Figure 190.

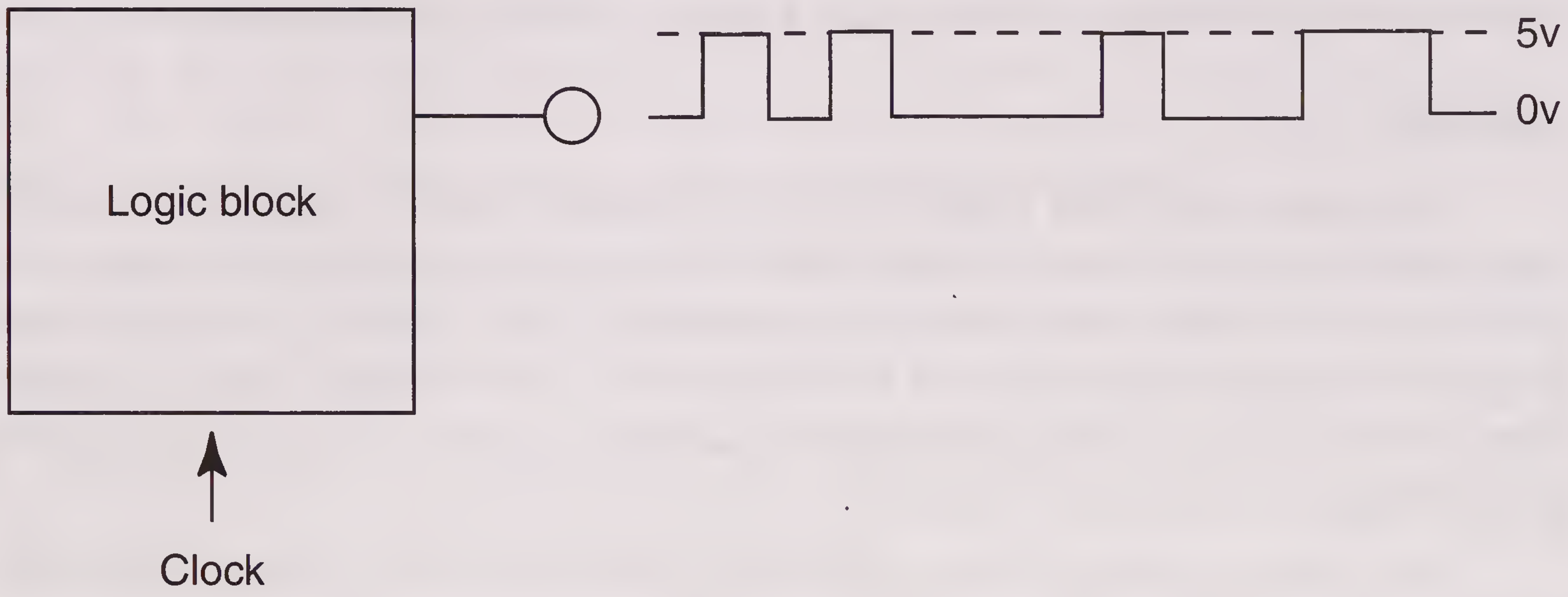

Figure 190

Specifications

The function of each digital integrated circuit is given by its part number. The basic electric characteristics are the same in the entire family. It is important to remember that TTLs always need 5 V to be powered and CMOS needs any voltage between 3 and 15 V.

WARNING—The integrated circuits of the CMOS family are very sensitive to electric charges stored in the body (static electricity) as are the CMOS transistors. Never touch the terminals of these devices directly. The discharge of the electric charge stored in your body can destroy the device.

Where they are found

As in other cases, the equipment where CMOS and TTL devices are found depends on several factors. Today, in modern equipment, the tendency is to place all the functions of a circuit inside one. This chip can contain the equivalent of hundreds or even thousands of TTL or CMOS-isolated ICs.

Although the operation principle is the same, you don't have access to each one and if something happens with one of them you must replace all of the circuits or the IC. On the other hand, the electrician, in his work with electronics, can find older equipment or equipment made by small manufacturers that doesn't use complex circuits, and also where isolated TTL or CMOS circuits are found.

Their circuits perform simple functions, such as those of timers, sequencers (light effects and signaling), alarms, remote controls, automatic lamps, night lamps, and inverters.

A basic circuit can show to the reader how they are used, such as the interesting circuit found in common applications that is shown in Figure 191.

This is a sequencer circuit using the common 4017. Each pulse you apply to the input makes one output of the circuit go to the high logic level, which means the voltage rises from zero to the power supply voltage. It is possible to select which output goes to the high logic level or is activated by the number of pulses applied to the input.

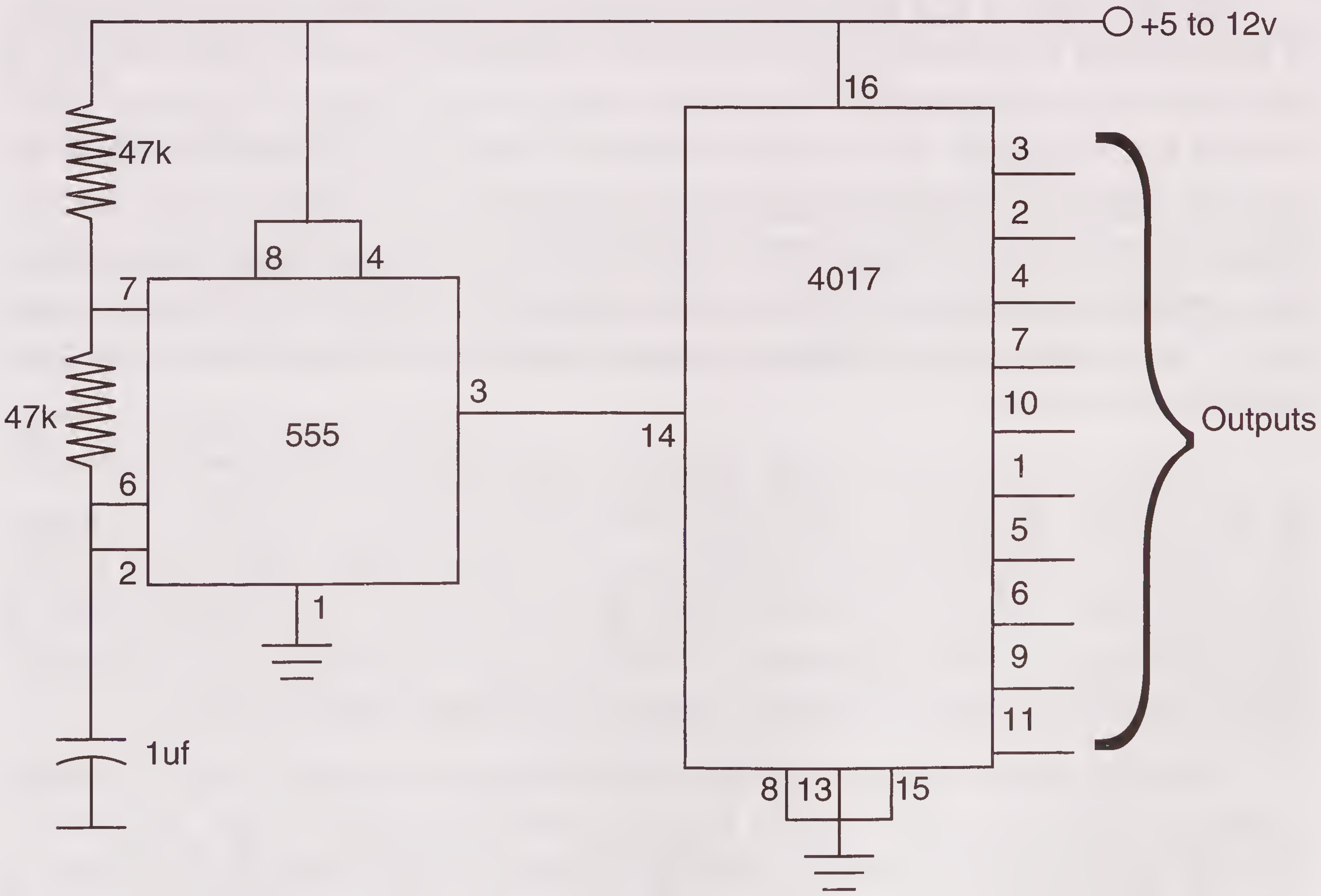

Figure 191

Testing

Testing of digital integrated circuits must be done using special instruments. The simplest way is measuring the voltages with a multimeter, but it is not conclusive in many cases (depending the way the circuit operates). The advanced test uses logic probes or oscilloscopes to monitor the digital signals in each IC.

Custom and Embedded

Today almost any function or circuit can be made in the form of an integrated circuit. To reduce costs and space the manufacturers can create their applications and put almost all the circuits in a single chip. This means that their products don't use many common integrated circuits, but instead use only one circuit dedicated to a specific use. In some equipment using these integrated circuits, functions like the power output can be performed by common transistors or SCRs, but the core of the circuit is a single chip with no equivalents.

As the cost of ICs for these applications falls, along with the quantity of equipment manufactured, the boards these ICs are installed on have become very inexpensive. This means that in the case of failures, entire boards are replaced. The main problem for the technician who makes repairs on these is that the boards are found only in authorized repair shops. Neither the board nor any of its components can be found at regular dealers. Another problem is that in many cases the equipment is a one-way type. This means once a failure occurs it is better to discard the equipment and buy a new unit.

Another kind of technology found in new equipment is one referred to as embedded. Embedded devices are used to control, monitor, or assist the operation of equipment, machinery, or entire areas, according to the IEE definition. These components are an integral part of the system where they operate. In a car, embedded circuits control all operations of the electrical system including electronic ignition and electronic injection.

They are basically formed by microprocessors and microcontrollers in the form of integrated circuits that are mounted inside black boxes. Access to embedded circuits is through cables. They allow the system to exchange control and sensor signals.

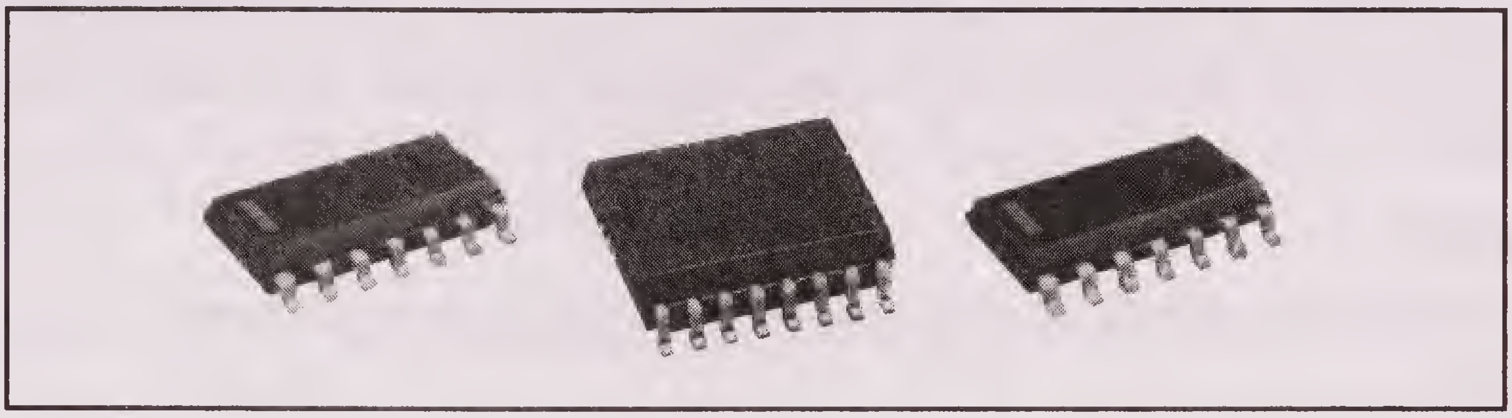

Figure 192

Types

No special symbol has been adopted to represent this kind of circuit. Figure 192 shows various types of these devices.

As the device is not intended to be handled by technicians, but only by machines, it is common to find them on boards in the SMD format, or a format called on-board. In the on-board format the chip, without case, is soldered directly on the printed circuit board and protected by a rosin or epoxy.

Specifications

In some cases the manufacturer indicates the IC by a part number, but in some one-way applications the device could only be identified by the manufacturer. In some cases, as in the black boxes of many embedded systems, the devices do not have any identification displayed.

Where they are found

The embedded functions are found in car equipment, home appliances, and toys just to name a few. For example, you can find these ICs in calculators where the manufacturer "creates" the circuit he needs to perform a function and places it directly on the board. Another example is electronic dolls that have a chip with voice-programmed functions so it can talk. The low cost of these specialized circuits and the form in which they are mounted make them impossible to be repaired or replaced.

Clocks and watches also use special chips containing the complete function they need in a single chip. Some of them can include the display. The clocks, counters, and other functions formed by a chip and the display are called modules. Many appliances use complete modules of counters, meters, clocks, and other basic functions. In some cases, if the function is not critical, the module can be replaced by another that has the same function.

Testing

There is no practical way to test them as each represents a separate function.

Microprocessors and Microcontrollers

Microprocessors and microcontrollers represent another great group of devices in the IC family. In a single chip thousands and even millions of components are used to perform a complex digital/analog function. Microcontrollers are intended to control applications, such as those in industrial machines and automatic devices. The microcontroller is formed by input/output blocks, a processing unit, and a task-programmed memory.

Figure 193 shows the structure of a typical microcontroller block.

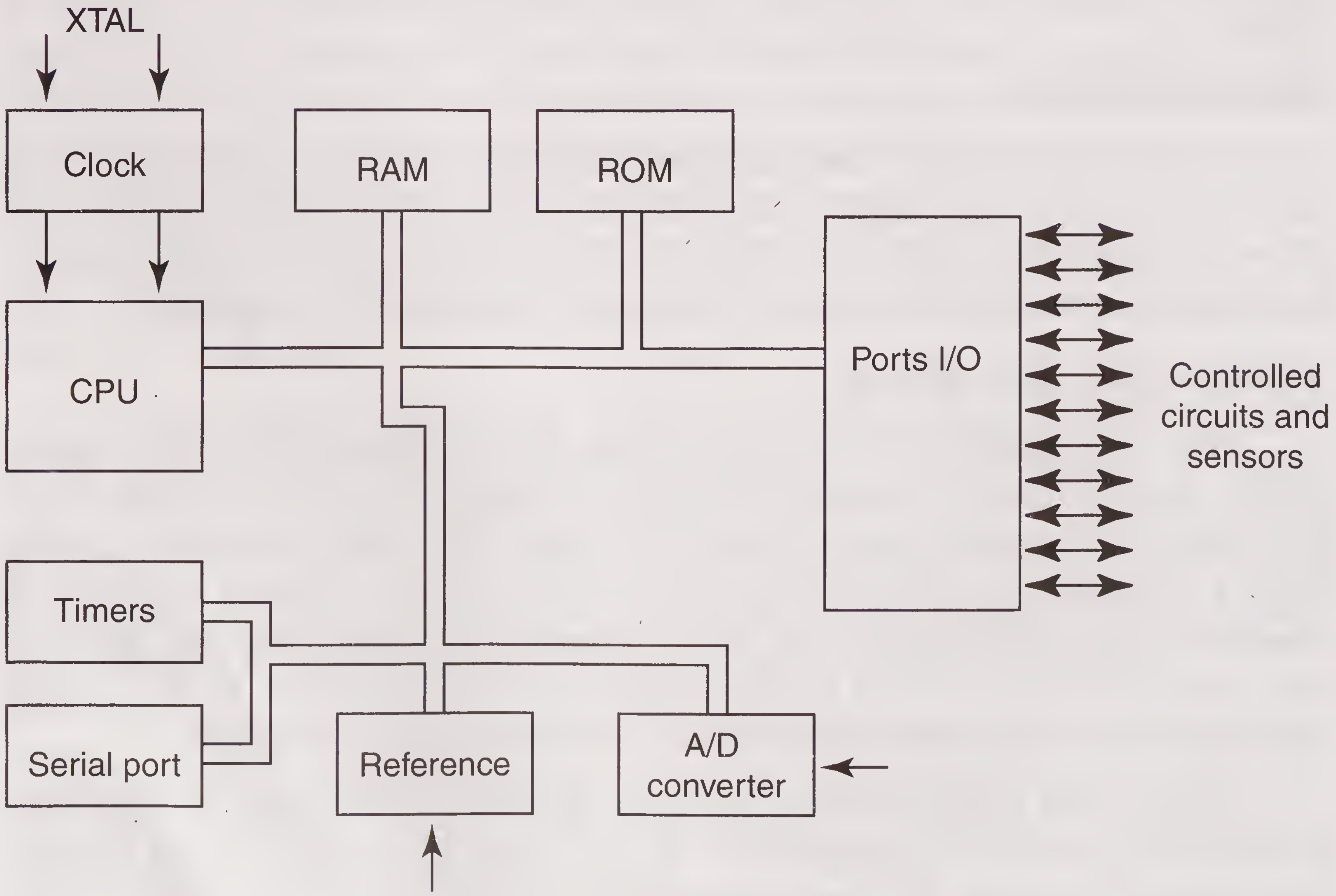
XTAL
Clock
RAM
ROM
CPU
Ports I/O
Controlled circuits and sensors
Timers
Serial port
Reference
A/D converter

Figure 193

Microprocessors are also used to perform complex mathematical and logic operations and are used in computers and other equipment. The microprocessor only works with digital data performing operations and returning with the results in the digital form. Figure 194 shows the block diagram of a typical microprocessor.

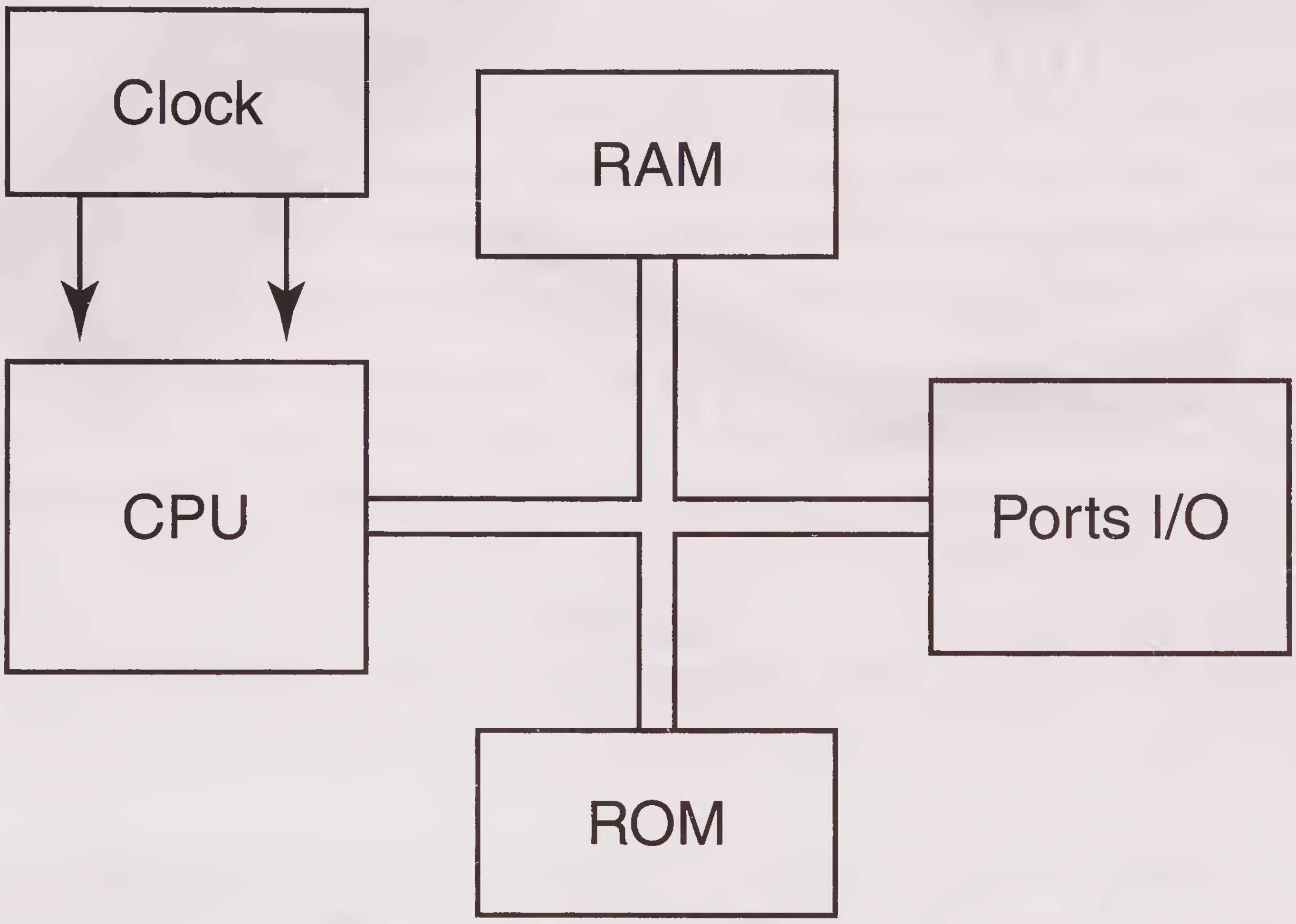

Figure 194

Microprocessors have input and output ports (I/O) where the information is placed and gathered, a CPU (central processing unit) where the mathematical and logic calculations are made, a memory, and other supporting circuits. Microprocessors are found in computers. Some families of microprocessors exist that are very common today.

Starting with the Z80, an 8080 is found in many old appliances as a "brain" where functions are calculated. More often now, we are using the microprocessors of the X86 family. The "X" is a number indicating the specific type. Several years ago, the 286 was common, now it is the 586. Variations with different names that include the Pentium K6, Athlon, and Celeron are prevalent. These microprocessors become famous through recognition in the personal computer (PC) field where they are used as the core.

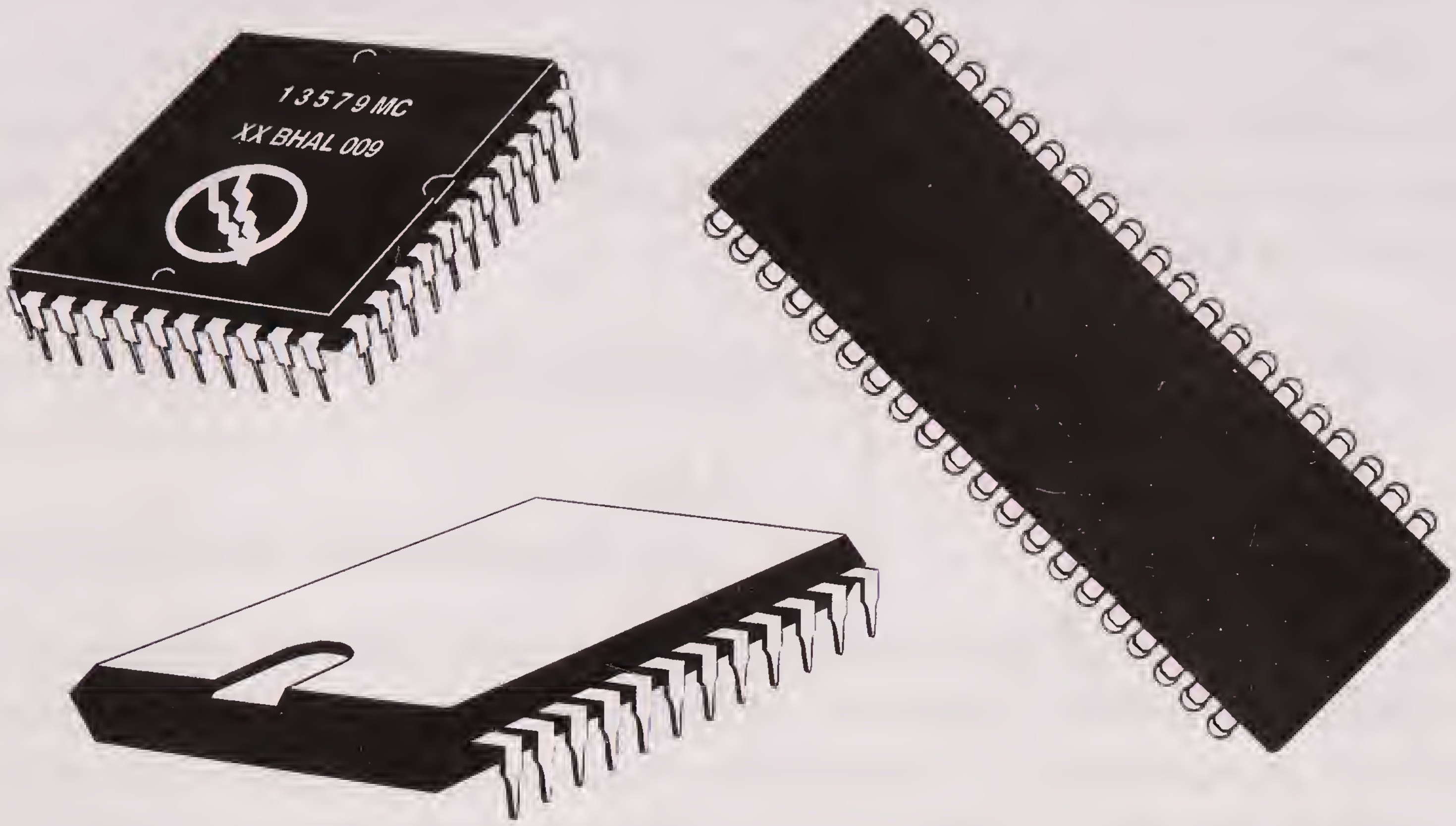

Figure 195
Chips and microprocessors similar to those used in PCs

The electrician who intends to work with computers will find a board inside, referred to as the mother board, where the microprocessor (also called CPU or central processing unit) is placed. The microprocessor is a very complex integrated circuit, containing millions of transistors and using a special body as shown in Figure 195.

Microcontrollers are found in industrial machines, appliances, cars, and other equipment performing control functions. They appear in the form of common integrated circuits in DIL bodies or more complex circuits (SMD and other) based on the number of pins and functions. Some families of microcontrollers are popular and are used in many applications. The 80C51, PIC, COP8, and others can be found in professional equipment.

The use of a microcontroller depends on some resources. The designer must have the tools to program them including a special board with a socket where the microcontroller is placed, a computer, and software to prepare the program. The designer should prepare the program that says what the microcontroller must do. Once prepared, the PC transfers the program to the microcontroller.

Specifications

Microcontrollers and microprocessors have large databooks to provide all their specifications. The technician wanting to use these devices needs special courses that involve a large amount of technical information.

Digital Signal Processor—DSP

Many other devices in the form of ICs can be found in electronic equipment including those related to electric installations and cars. New functions like the DSP are an example of devices that are designed to perform specific functions normally determined by the manufacturer (custom).

These devices are found in many kinds of electronic equipment, such as cellular telephones, computers, fax machines, industrial machine controls, dishwashers, clothes washers, remote controls, electronic injection, and ignition controls in cars, DVD players, TV satellite receivers, and videocassette players. They are not simple devices or components, but complete circuits programmed for a predetermined purpose, mounted on a chip, and installed inside a case like the ones used in any other integrated circuit.

What is a DSP?

The traditional analog circuit operates with analog signals like sounds, images, or voltage changes from a sensor, along with traditional components and circuits such as transistors and filters. The analog circuit uses the electric properties of the components to introduce some alterations in the signal's original form. For example, a bandpass filter lets only signals for a determined frequency band pass through it; an amplifier changes the amplitude of a signal.

Depending on what you want to do with a signal, the implementation of a network using common devices is not only very difficult, but is not very accurate. However, if we consider that any analog signal can be represented by number or a stream or numbers (converted to a digital form), why not perform the desired function with the signal working in the form of numbers? Instead of using a network that lets only a determined frequency band pass, convert the signal into a stream of numbers in a computer, apply an algorithm to the numbers that recognizes the frequencies and amplitudes, and the undesirables values are cut. (An algorithm is a sequence of mathematical and logic procedures.) After that, the remaining values are converted again to the analog form.

Figure 196 shows, by a block diagram, how the DSP works.

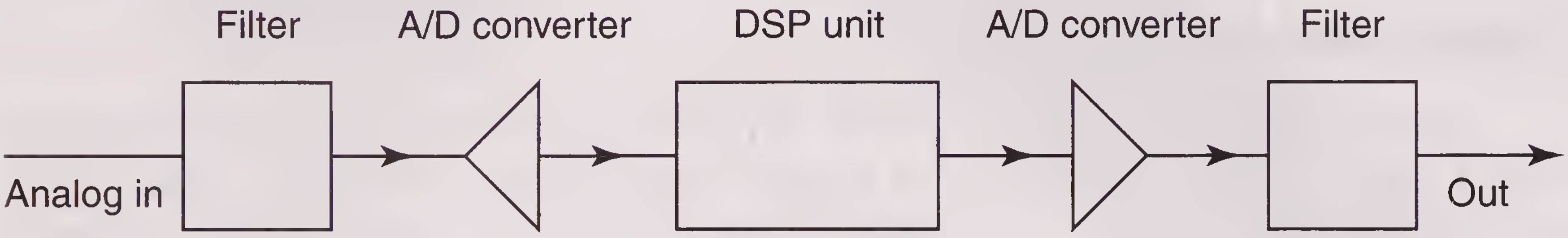

Figure 196

Since a computer is precise and fast, the results presented by this kind of circuit are fantastic—much better than a common analog circuit designed to do the same task. The processor inside a DSP is fast and can perform extremely complex operations, making it possible to make almost everything you want with a signal depending only on the software (algorithm) used. Using a DSP, what a circuit will do no longer depends only on the components used, but on a program!

Many DSPs are used today in common electric and electronic appliances in control functions to replace microcontrollers. For example, a manufacturer can buy an empty DSP and program it to perform all the functions needed to control the motor of a clothes washer, including functioning as the water-level sensor and the temperature sensor. The DSP processes this information and determines the voltage (speed) to be applied to the motor, the voltage applied to the water heater, etc.

To install it on a board, there are specific DSPs for this product without the need of common devices in a large board. The same DSP can be used by another manufacturer to control the motors, solenoid, and relays of an industrial machine. The only thing the other manufacturer has to do is plug the necessary peripherals into the DSP and install the appropriate software.

Of course, if a DSP fails in a machine, the technician has no other solution than to look for the original part from the manufacturer. Also, a single DSP (without the program) doesn't work in a machine if it isn't programmed with the original software. The DSP must be purchased with the original programming of the manufacturer. The problem is simpler if a peripheral component fails, such as a transistor, SCR, or relay. In this case, the component can be easily replaced.

It is important to know that the DSP is identified by a part number (Texas Instruments, Analog Devices, and Motorola are manufacturers of DSPs), but if you buy one at any dealer it comes without the programming. If you intend to build any electronic equipment, depending on its function, the use of a DSP should be considered.

Surface-Mounted Devices—SMDs

Many of the devices we have seen before can be found in an ultraminiaturized version mounted by machines in large manufacturers' assembly lines. Notice the majority of the devices are in cases. The active element of a resistor is only a thin layer of carbon; the active element of a transistor, SCR, or triac is a small silicon chip. The active element of an integrated circuit is also only a small silicon chip.

The body, for the most part, represents more than 95 percent of the size and weight of the component. It is clear that, if the body is reduced, we can gain space and profit, but on the other hand, we lose the ability to handle it. The basic idea of the SMD is to reduce the body to the minimum,

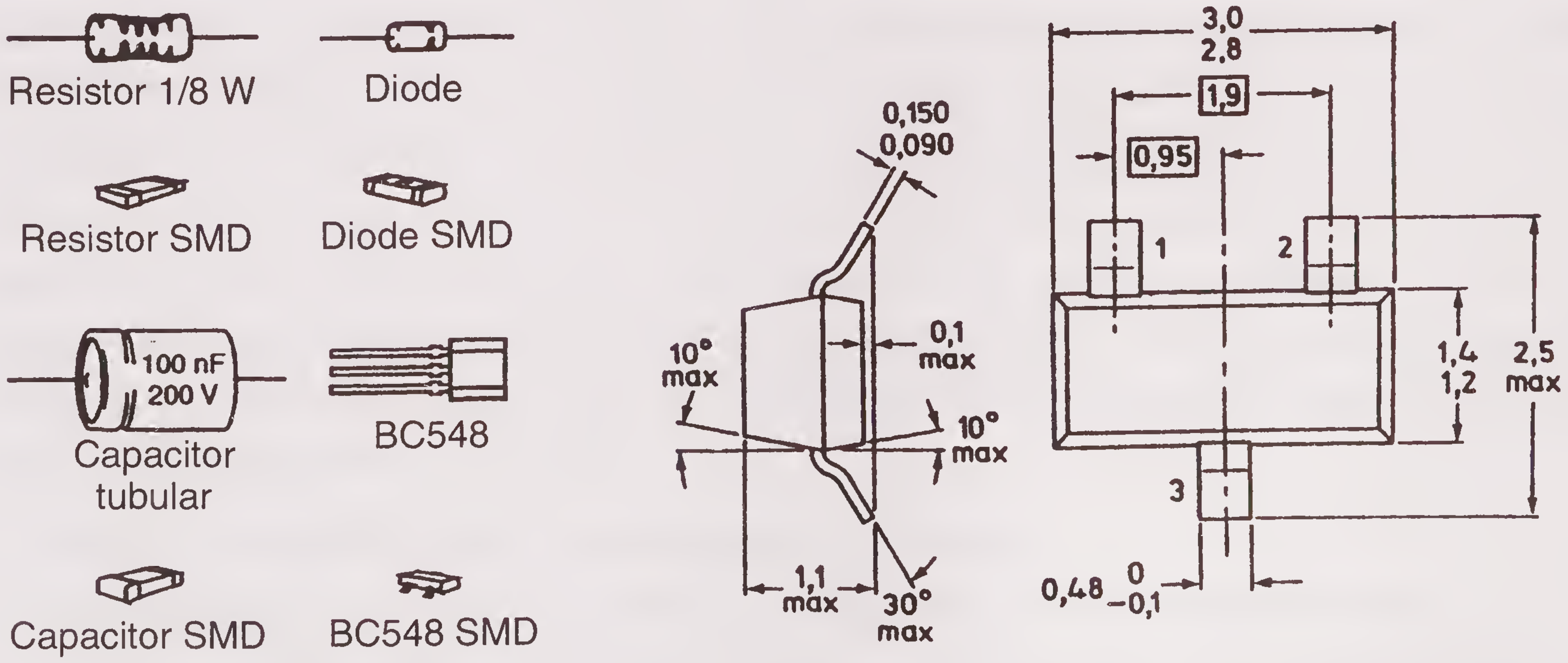
Resistor 1/8 W
Diode
Resistor SMD
Diode SMD
100 nF
200 V
Capacitor tubular
BC548
Capacitor SMD
BC548 SMD
3,0
2,8
1,9
0,95
0,150
0,090
1
2
3
0,1
max
10°
max
10°
max
1,4
1,2
2,5
max
1,1
max
30°
max
0,48
0
−0,1

(a) (b)

Figure 197

in a way that only machines can handle it. Reducing the size of the components in the equipment makes them cheaper, smaller, and lighter. Resistors, capacitors, diodes, transistors, and ICs can all be found in SMD versions as shown in Figure 197.

This figure shows the typical dimensions of these components and how they are mounted on a printed circuit board. Special machines place the components on a board using glue and solder. The working technology applied in this case is also called SMT (surface-mounting technology), so it is common to refer to devices used in this form as SMD and the technology as SMT.

It is important for the electrician to understand that the functions and the circuits are the same as those used in the common components. The difference is only in the size and shape of the components. The main problem is how to handle them if a repair is needed.

Special kits, consisting of small tweezers, screwdrivers, and other delicate tools are available to help technicians replace SMD devices when repairing equipment. The same dealers usually can furnish resistors, capacitors, and many other components in the SMD format as well.

Section 3

Troubleshooting and Repair

One of the purposes of this book is to give an electrician some basic knowledge about electronic components and circuits—knowledge that is necessary to make simple repairs and install electronic equipment coupled to electric installations, powered from electric installations, or even found in a car. It is not up to the electrician or professional of any field other than electronics to make complex repairs in electronic equipment. This task must be left to professionals. But, depending on the problem, repairing electronic equipment is not difficult if an electrician knows some basic procedures. There are two kinds of problems with electronic equipment the reader of this book can solve:

1. Simple problems with the equipment circuit.
2. Problems due to the installation or operation of the equipment.

This section presents a basic look at symptoms and causes of simple problems the reader, who may not yet be familiar with the operation of electronic circuits, can solve.

Because we can't see, hear, or smell electricity, knowing exactly what is wrong in a circuit only can be discovered with the aid of instruments or a procedure using symptoms as a basis for troubleshooting.

Of course, as this book is introductory, a reader wanting to go a step further will need more information. If you want to go into the profession, you will have to learn much more in a technical course as well as obtaining access to a variety of technical databases. It is also important and helpful to read technical periodicals because they provide useful procedures and ideas on new technologies.

The First Step

If a piece of electronic equipment doesn't work as expected what are the basic procedures to find the cause and how then to repair it? It is a false idea (coming from the tube age) that you should replace a given part when a certain symptom occurs. This is not valid for very complex equipment and circuits where hundreds or thousands of interdependent components exist. In complex circuits the same symptom can occur when any one of many components fails, and in some cases, many of them fail at the same time. You must know how the circuit works to form a hypothesis about what the cause of the trouble is.

For an electrician who is now starting with electronics, it is too early to try real troubleshooting and repair work in complex equipment. We can, however, provide some basic procedures to allow determination if the cause of a problem is inside the equipment (and requiring professional help), in the installation itself, the result of improper use, or some other external problem.

The basic procedures are also useful to determine the cause of simple problems that need no professional help to be solved, such as burned-out fuses, unplugged connectors, and broken wires.

Safety Rules of Troubleshooting

These are the most important items when working with any piece of AC-powered equipment, and even equipment powered by other sources can use high voltages. It is very important to know the hazards associated with the equipment you are troubleshooting. Electricity can kill and the electrician working with dangerous AC voltages must know that.

- Don't turn on the equipment without thinking about it first. Start with some analytical thinking. Be sure that you know how to operate the equipment you are troubleshooting before powering it on.
- Don't touch parts if you don't know what they are for or how they work. You can cause more damage to the equipment; a simple problem can be turned in an expensive problem.
- Look for visibly damaged parts such as burned-out resistors, interrupted cables, and bad connections. Many problems have simple solutions. The visual inspection is the starting point in any troubleshooting work.
- Learn from mistakes; incorrect procedures when troubleshooting can turn a simple problem into an expensive one. Let a tool fall inside equipment being tested causing a short and you will have a real example of what that means.
- Working with modular parts: modern electronic equipment often uses modular parts. If you can identify the cause of a problem in a module, it is easier to replace the complete module than trying to repair it by looking for burned-out components.
- Don't always trust your instruments. If you are using a multimeter to measure voltages or a resistance in a circuit and the reading is confusing, don't assume that is something wrong with the circuit. In many cases, the characteristic of the multimeter and the presence of other components near the one being tested can mask the results. Using the multimeter is an art and only with experience can you begin to trust all the measurement made by this kind of instrument.
- However, the multimeter is probably the most useful of all the instruments the electronics professional can have. It is very important for someone wanting to become an electronics professional to be familiar with all the uses of this instrument. The next section discusses some more of its uses.
- Film canisters, plastic ice cube trays, eggs trays, and pill bottles can be used for sorting and storing screws and other small parts when you disassemble equipment. Use a notebook to mark the position of any

screw in equipment tested if you have difficulty memorizing it. A leftover screw when reassembling a device can cause a big headache.

- Don't force any part of the enclosure in equipment when disassembling it. If you have to force it, it may be because you are not doing things correctly. The direction you are moving the part is not the correct way or there are more screws to be taken out.
- ESD (Electrostatic Discharge)—Some components, including the CMOS transistors and ICs, are very vulnerable to ESD. Don't touch their terminals when working with them.
- If possible, use a schematic. Many pieces of equipment have manuals or schematics that provide important information for the reader when troubleshooting. If you intend to go further in this field, knowing how to read a schematic is essential to the repair of any equipment.

Schematics

Sams Technical Publishing has circuit diagrams, schematics, or service information for many pieces of electronic equipment sold since the 1940s. You can access the company's service using this Internet URL: *http://www.samswebsite.com*.

You can also reach it at the following address and phone number:

Sams Technical Publishing
5436 W. 78th St.
Indianapolis, Indiana 46268
Customer Service: 1-800-428-7267

Many large public libraries may subscribe to Sams Photofact and have the documents available for photocopying.

Quick Tips for Troubleshooting and Repair

- A totally dead piece of equipment or one with many functions affected may have a defective power supply. Check whether the circuit is correctly powered. For example, see if there is voltage in the AC power line.
- Erratic or intermittent problems are almost always due to bad connections, such as cold solder joints or connectors that need to be cleaned.

- Gradually changing problems (problems that decrease or disappear completely when the equipment warms up) are often due to dried electrolytic capacitors.
- The majority of problems in equipment with many mechanical parts, such as VCRs, CD players, or DVDs, are due to mechanical or optical failures.

The Multimeter

The multimeter, VOM (Volt-Ohm-Milliamp meter), or Multitest is the most important instrument for an electrician who wants to work with electronics. Any test in electronic circuits or electronic components begins with the principle that when it is working, some measurable electric quantity exists. Here, three quantities are the most important: voltage, current, and resistance.

Measuring the current or the voltage in a circuit allows us to determine if the circuit is good or bad. In some cases, measurements can also determine the cause of a problem. Many electronic devices can be tested with some simple resistance measurements.

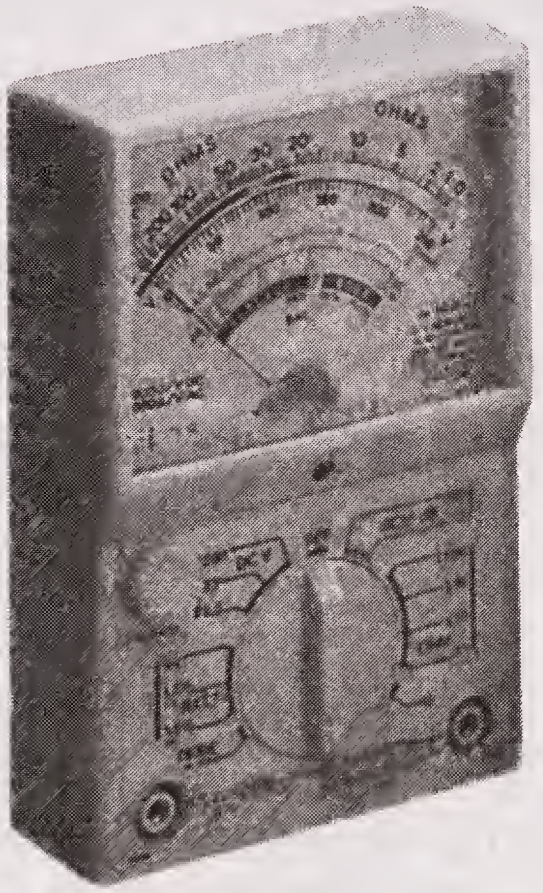

Figure 199
A low-cost analog multimeter

Analog Multimeters

There are two types of multimeters an electrician can purchase to work on not only electric installations, but also on analog and digital electronics. The analog multimeter is formed by a moving coil meter (microamperimeter) and circuitry with functions selected according the quantities to be measured.

For example, if an electrician must measure resistances he needs to select one of the resistance positions on the switch according to the expected value. If the position ohms x100 is selected, the values read in the scale must be multiplied

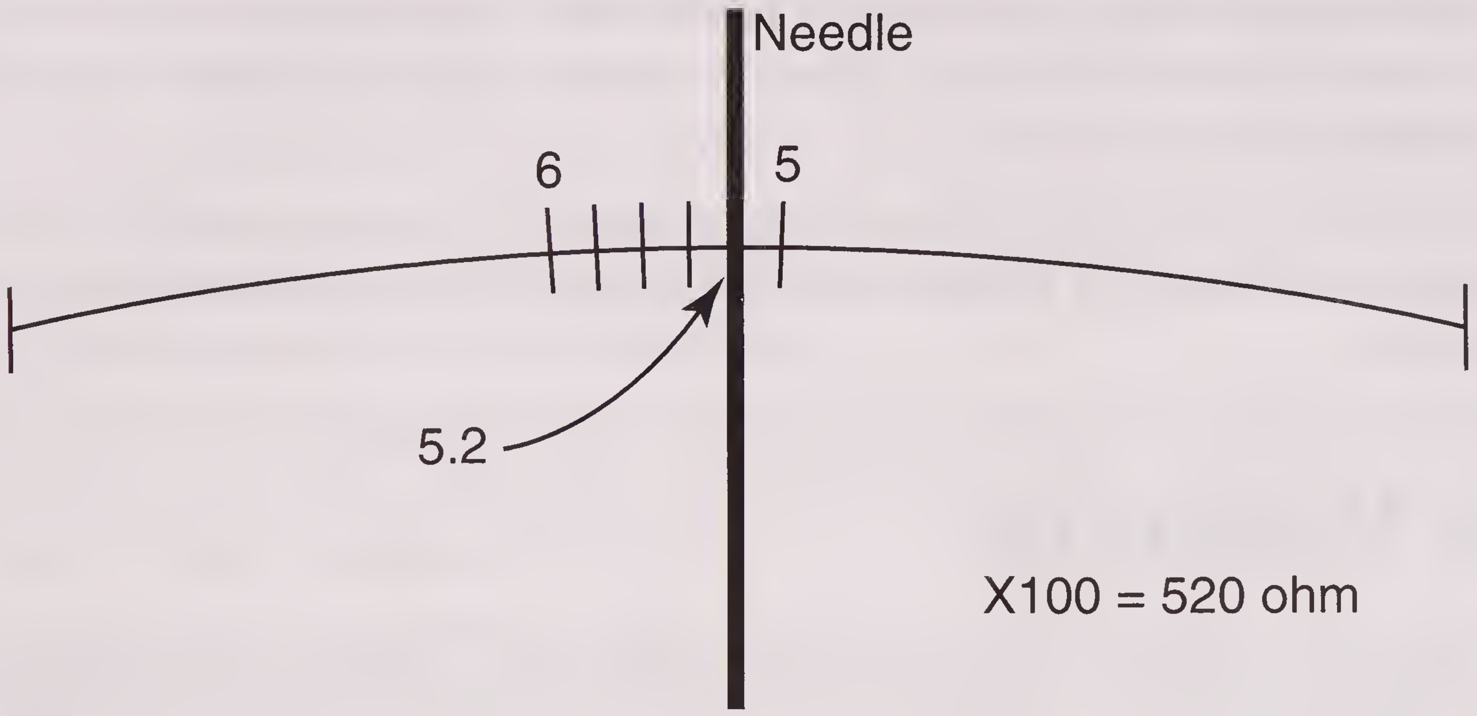

Figure 200

by 100. An indication of 5.2 as shown in Figure 200 indicates a resistance of 520 ohms.

The selection of the quantity to be measured is made by a switch or selected by jacks where the probes are connected. Figure 199 shows this kind of multimeter. The most important specification of an analog multimeter is its sensitivity. The analog multimeter, when measuring voltages, needs power to move its needle. The power comes from the circuit being measured. This power alters or "loads" the multimeter changing the value reading. The better the multimeter, the lower the load needed from the circuit to be measured, meaning a better multimeter can pick up lower voltages.

The sensitivity of an analog multimeter comes from its internal resistance when measuring a voltage (it is associated to the sensitivity of the moving coil meter). This value is indicated in ohms per volt (Ω / V) and should be as high as possible. Less expensive multimeters have a sensitivity of around 1000 ohms per volt. Better multimeters have a sensitivity of 10,000 ohms per volt or more.

When using a low-sensitivity multimeter, an electrician must take care to not be tricked by the voltages indicated when measuring high-resistance or high-impedance circuits. These are circuits where the presence of the instrument can cause great alterations in the voltage measured. This occurs because the presence of the multimeter loads the circuit .

Digital Multimeters

Many low-cost digital multimeters are available today. Although by many aspects they are better than the analog type, some functions or measurements exist where the analog multimeter is the better choice. For the electrician who is not familiar with multimeters, using this type of tool is not recommended as a starting point. When beginning, it is easier to use the analog type.

The digital multimeter is formed by an electronic circuit that presents the results of the measurements in a liquid crystal display (LCD) like the one shown in Figure 201. The functions for measurement using a digital multimeter are the same as analog multimeters.

Because the digital multimeter uses an internal power supply (normally a 9V battery), it doesn't need to extract large amounts of energy from the circuit where the voltages must be measured as the analog multimeter does. This means that the circuit being tested interprets it as a very high-impedance load not causing significant changes in the quantity to be measured. The problem of inaccuracy found in analog multimeters when measuring low voltages in high-resistance circuits doesn't exist in the digital types.

Figure 201
A low-cost digital multimeter

How to Use a Multimeter

Voltage measurements.

Electricians must take care not to overload the instrument when measuring voltages. This can cause damage including burning it. They must also differentiate the points where AC and DC voltages are found.

When measuring a DC voltage, place the function selection switch in the scale that gives you a comfortable reading of the expected voltage. For example, if you are expecting a reading of about 12V in a circuit, place the switch in the position 0-15VDC. The position 0-150VDC is not recommended because with 12V, you would have a small deflection of the needle.

Observing the poles, place the probes in the circuit to be measured. If you invert the probes or the poles are different from what is expected, the needle tends to move in the wrong direction, indicating something less than zero. If this happens, invert the probes to correct the reading. Some instruments have a pole switch that inverts the probe connections without moving them. Figure 202 shows how the multimeter is used to measure a DC voltage in a circuit. This process is for digital and analog meters.

Some digital multimeters indicate the value with a "–" sign if the voltage is negative or if the probes are inverted. If you have no idea about the voltage present in a circuit, begin by using a higher scale.

Current measurements

Current measurements are not very common when working with electronic circuits because the circuit where the current is measured must be interrupted to place the multimeter's probes (Figure 202).

When measuring DC currents, polarity must be observed when positioning the probes. It is very important to choose the correct scale when measuring currents. If a low-current scale is chosen in a high-current circuit, the meter can be damaged or the internal fuses can burn. Use the higher current if you have no idea about the value of a current to be measured in a circuit. The procedure is the same whether using a digital or analog multimeter. In some cases the multimeter has a self-adjusting scale and no switch is necessary for this task.

Resistance measurements

The resistance measurement and the continuity test are the most common in normal work with electronics. To run a resistance test, apply a voltage to the circuit or device to be measured. The magnitude of this current causes the multimeter to display the value of its resistance.

Analog multimeters use an internal cell (normally a 1.5V AA cell) and digital multimeters use a 9V power supply. As the batteries in a multimeter lose power, so does the multimeter. Before each use, analog types must be adjusted to zero. This adjustment is shown in Figure 203.

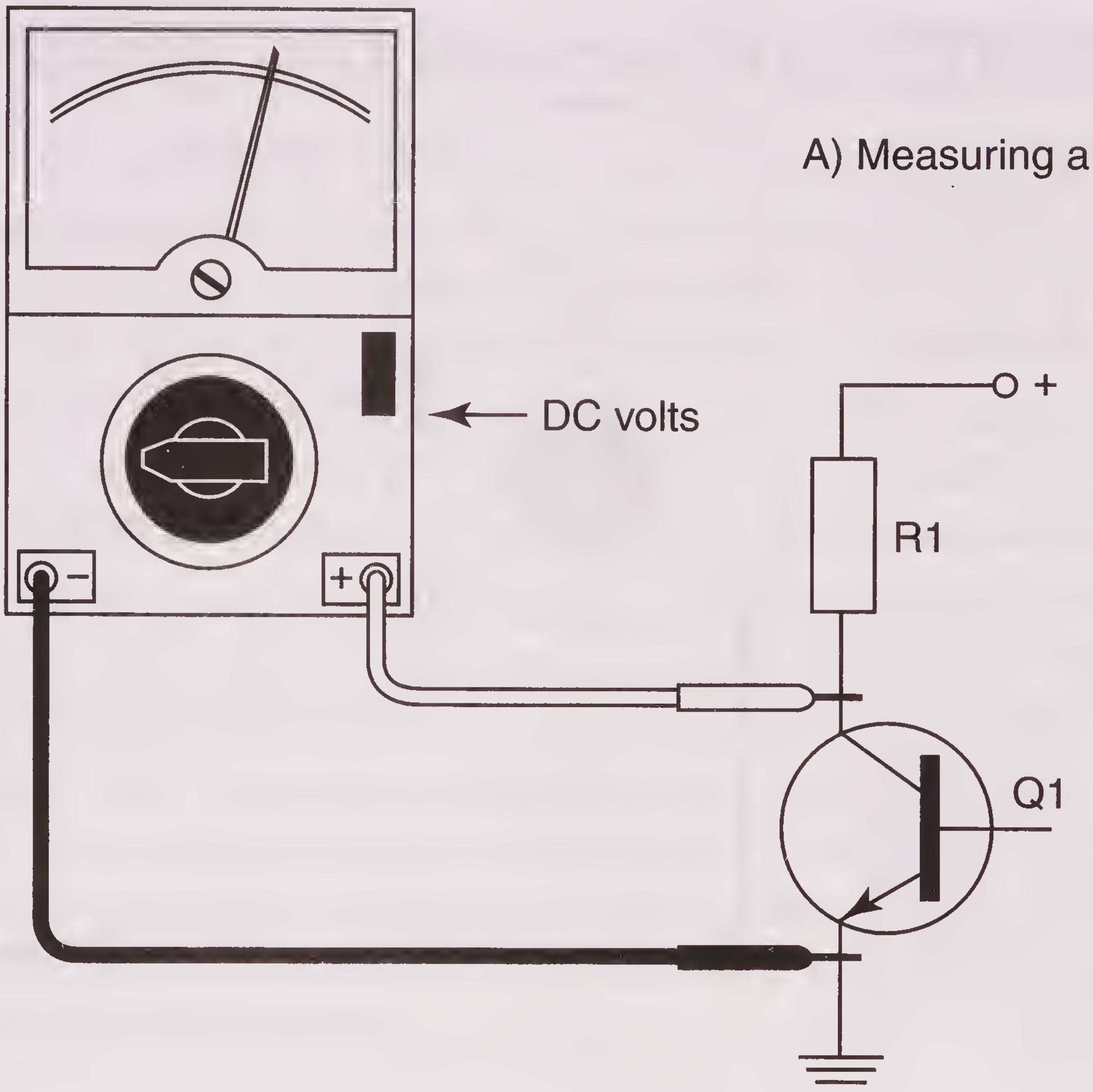

A) Measuring a voltage across Q1

DC volts
−
+
○ +
R1
Q1

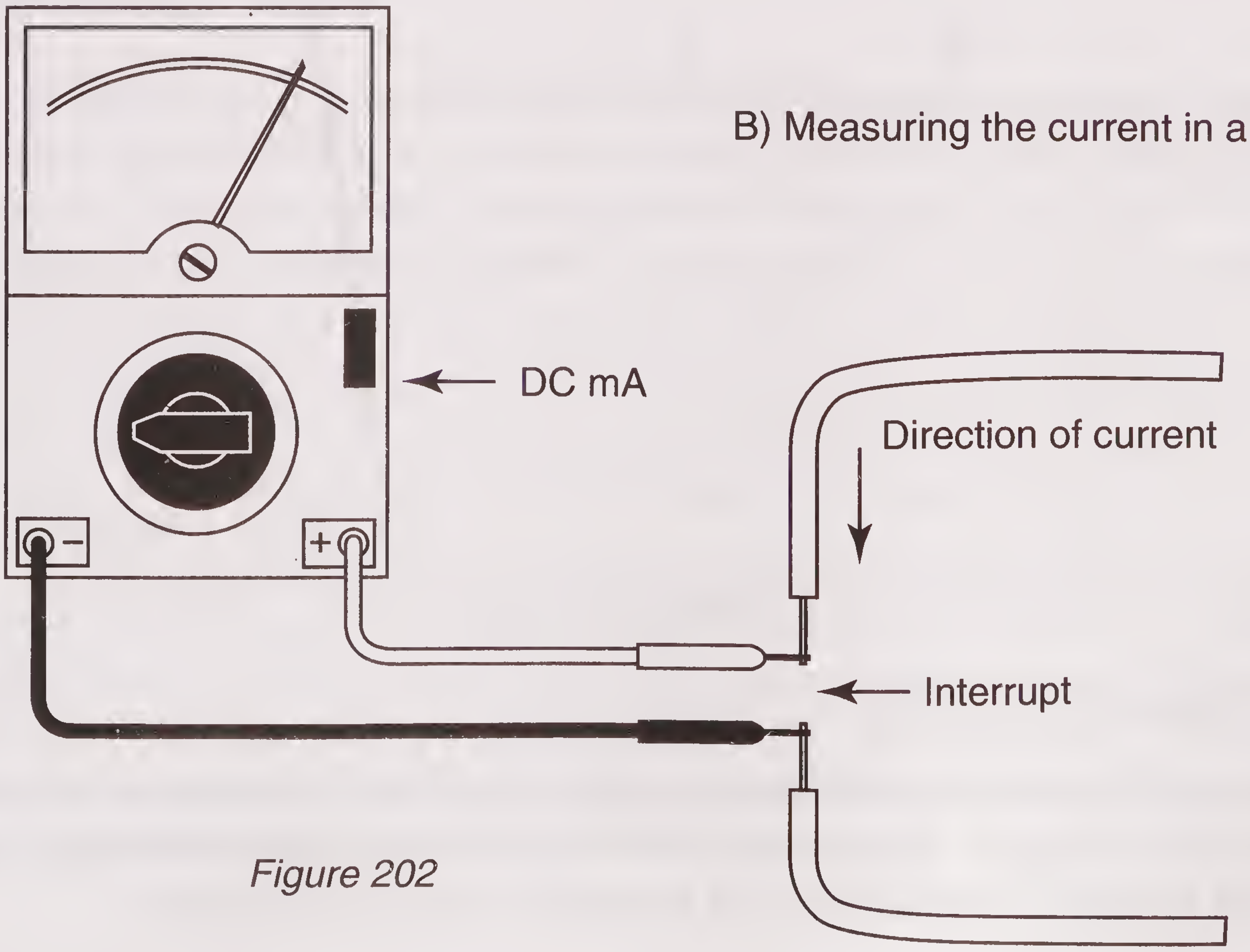

B) Measuring the current in a curcuit

DC mA
−
+
Direction of current
Interrupt

Figure 202

Place the probes together and adjust the "zero adj" potentiometer until the needle in the scale indicates zero. After this procedure you can use the multimeter to measure resistance.

Choose the appropriate scale for the resistance you expect to measure. A better reading is one made from the central region to the right extreme. For instance, if you place the switch in the ohms x 1k scale, it means you have to multiply the readings by 1000. Figure 204 shows that an indication of 4.3 in this scale means you are measuring 4300 ohms.

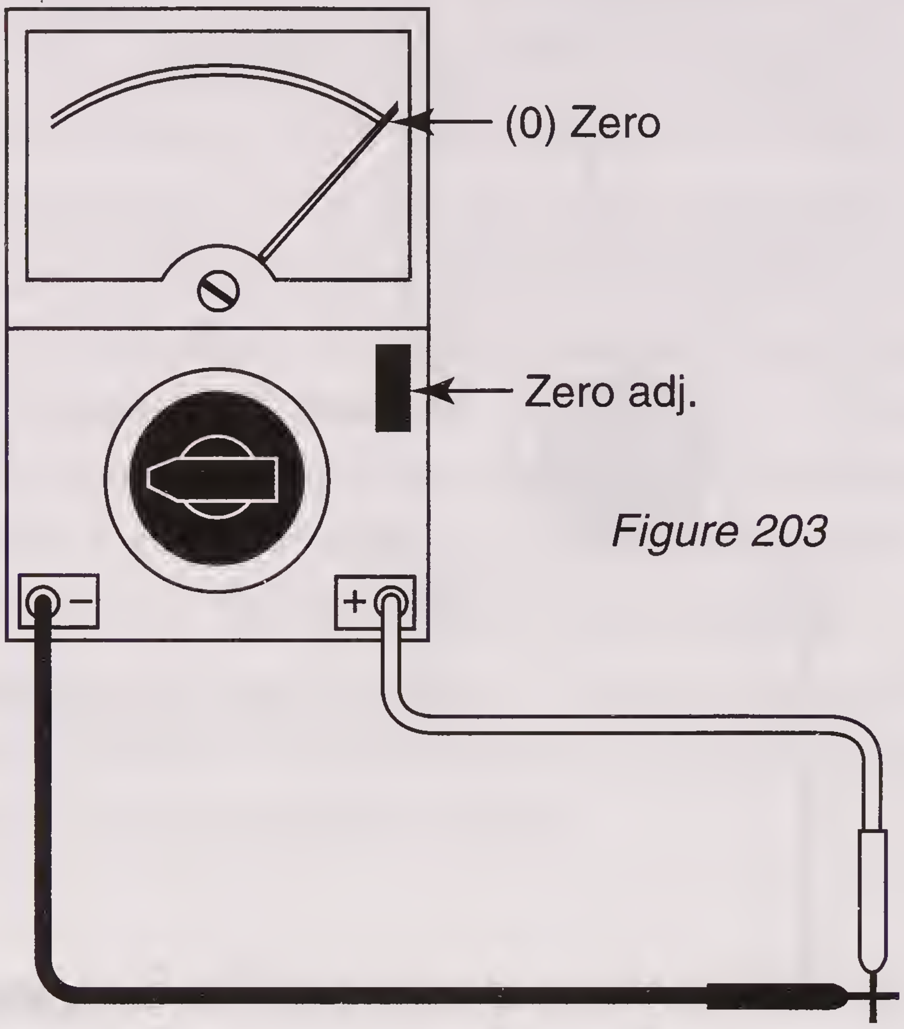

Figure 203

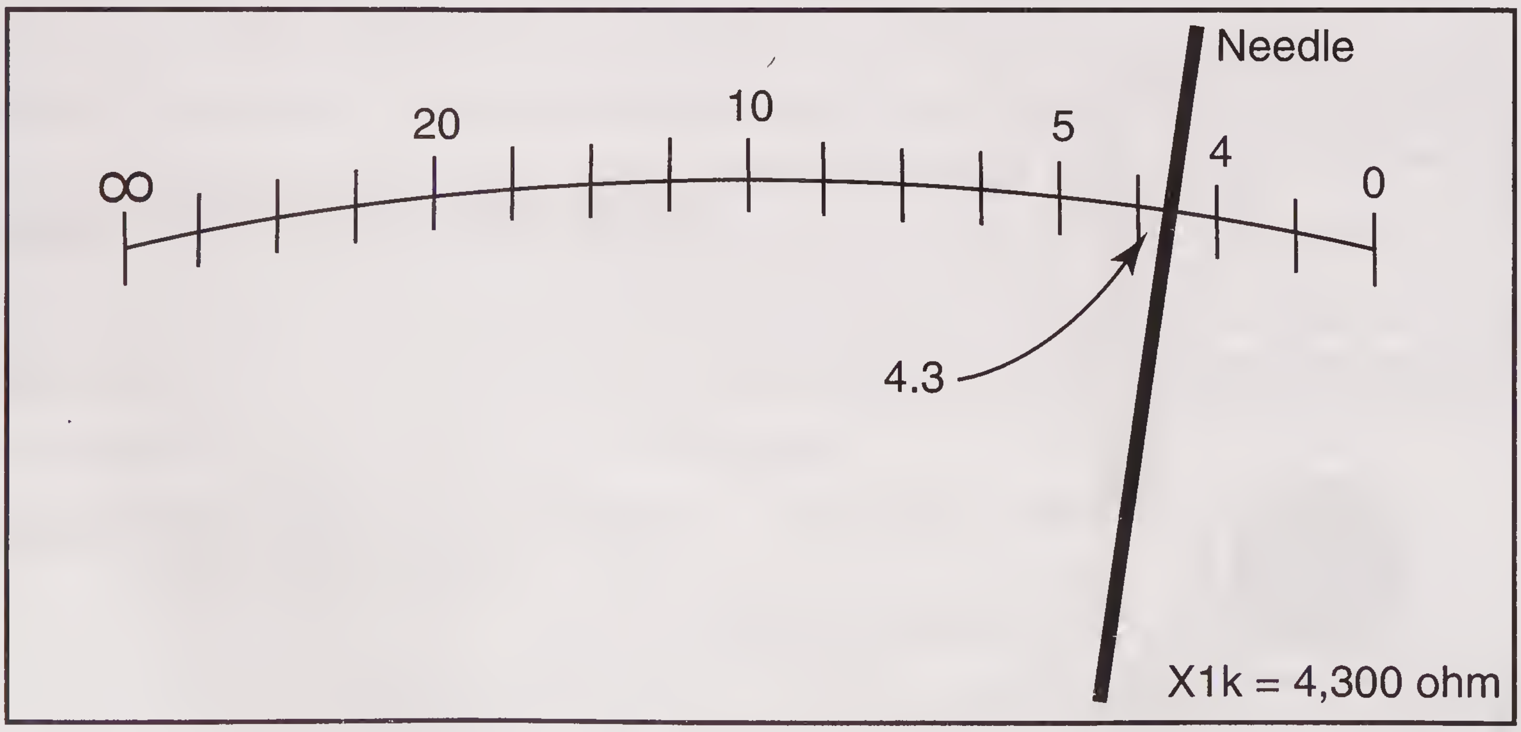

Figure 204

Other Multimeter Functions

Many multimeters can measure other quantities besides resistance, currents, and voltages. Multimeters with battery test scales, decibels, continuity test function, diode test, and transistor test are common.

Reading and Interpreting Multimeter Values

Knowing that placing the probes of a multimeter in the terminals of a component and having it give you a reading either with a needle or LCD is not enough. You need to know what the reading means to conclude something of value. Some components are good when they present high resistance (capacitors, reverse-biased diodes, open switches, etc.) and some components are good with the reverse indication—when they present low resistance (fuses, wires, direct-biased diodes, coils, etc.).

At the end of the description of each component, the text explained how to test many components using a multimeter. When measuring voltages in a circuit, you must take care with interpretation of the values. Although the schematics indicate some voltages, depending on the multimeter, the presence of the multimeter in the circuit can change the values and give a false reading (mainly in less expensive analog types). Don't trust your multimeter! You must be sure that your multimeter can measure the indicated voltage in a circuit.

Finding Parts

Electronics today often are made up of one-way circuits. Once they present any problem, the entire board—including other components and parts—is replaced. Few components in equipment such as this can be replaced when problems occur. In some cases, the cost of replacing the components is actually higher than the cost to replace the board. The price of a technician's hourly work also figures in here. But, repairing old equipment or equipment not designed to be discarded when something inside is burned out, is possible and a technician must know how to do it.

After pinpointing the location of the cause and the components involved in the failure, the next problem to solve is finding a replacement. Today technicians can count on many component dealers, many of whom work through the Internet and mail order. Once contacted, most dealers will send the component(s) directly to your home, often in a few days, and in some cases, a few hours!

For those with access to the Internet, some important dealer addresses in the United States and beyond are provided. Visit their web sites for more information about locations, shipping, and prices. Phone and fax numbers of many are also listed.

Dalbani
http://www.dalbani.com
U.S. Voice: 800-325-2264
U.S. Fax: 305-594-6588
Int. Voice: 305-716-0947
Int. Fax: 305-716-9719

MCM Electronics
http://www.mcmelectronics.com
U.S. Voice: 800-543-4330
U.S. Fax: 800-765-6960

Radio Shack
http://www.radioshack.com
U.S. Voice: 800-843-7422
U.S. Fax: 800-813-0087
Mail: Tech America, PO Box 1981,
Forth Worth, Texas 76101-1981

Parts Express
http://www.parts-express.com
U.S. Voice: 800-338-0531
U.S. Fax: 513-743-1677

All Electronics
http://www.allelectronics.com
U.S. Order: 888-826-5432
U.S. Fax: 818-781-2653
U.S. Voice: 818-904-0524

Alfa Electronics
http://www.alfaelectronics.com
U.S. Voice: 800-526-2532
U.S. Fax: 609-897-0206

Other Sources for Components

- Arrow Electronics
- Avnet
- CalSwitch
- Digi-Key
- Electro Sonic
- Gerber
- Avnet Electronics
- JACO
- Marshall
- Newark
- Nu Horizons
- Reptron
- NetBuy
- Partminer SDirect
- Powell Electronics

Another important resource for components is old equipment that no longer functions. This equipment often has many parts that can be used as replacements. If you like to repair things and are always trying to recover old equipment, it can be interesting to have old boards or even seemingly ancient tube radios and amplifiers as a source of electronic components.

Transistors, resistors, and some types of capacitors are hard to destroy and can be found in good condition even many years after the original equipment is abandoned. The more critical components are electrolytic capacitors that can become dry after long periods of time, losing capacitance or going to a low-resistance state (short or loss) and can't be used. It is recommended to test all components before using them.

Equivalents

A common problem facing an electrician working with damaged electronic equipment is finding replacement parts when the original is not available anymore. When this occurs, an equivalent part is needed. However, these parts are not always exactly the same as manufacturer originals, and the same one cannot necessarily replace the same device in other equipment.

The most problematic cases involve semiconductors. Diodes, bipolar transistors, FETs, power-FETs, SCRs, triacs, and ICs come in millions of different types. Many manufacturers create their own types or designations to replace those made by competitors. In some cases they even add internal alterations, changing the basic characteristics for better performance. For example, an operational amplifier, such as the popular 741, can be found with designations LM741, 1741, uA741, LH101, CA3056, MC1539, and many others. Theoretically, they are the same component, but small internal differences can alter the performance and make one type improper for replacement depending on the application.

Transistors suffer from this problem as well. Many manufacturers used to adopt their own code when placing a transistor in equipment. The common 2N3904 or the BC548 can be remarked as, for example, 3X456-V-334-bis or something like this by a manufacturer. This makes the work of a technician replacing this component difficult. In many cases the intention is to force the user to look for an authorized service center in the event of problems.

When the component is common, it can be easy to find a replacement beginning with its characteristics. Experienced technicians, by simple observation of a circuit, often "know" which transistor is the ideal replacement type. As a general rule, the replacement type must have characteristics equal or superior to the original. If you are just starting out with electronics, it is not a good idea to attempt replacing pieces with these kinds of problematic designations.

Practical Circuits: How They Work

As pointed out in the beginning of the book, electricians today are finding electric parts not only in domestic, commercial, or car installations, but many electronic circuits. Another problem concerning an electrician relates to the electronic devices and appliances that operate from electric power and are sometimes in building installations. Although a deep knowledge about how these things work is not necessary to repair or install circuits, an electrician should have basic information about the operation. This information can be very useful when installing or troubleshooting an electric installation with a problem.

Let's look at the operation of some electronic circuits and devices found in electric installations and AC powered in homes, commercial buildings, and even cars. How do they operate and how are they placed in electric installations?

AC/DC Adapters

Many small electric and electronic appliances operate from DC power supplied from cells and batteries. In some cases, it is more convenient to run these devices from AC power to preserve battery life. AC/DC adapters or DC power supplies convert the high AC voltage present in the AC power line to a low DC voltage (in the range between 1.5 and 12V).

AC/DC adapters are formed by a simple circuit with a transformer to isolate the AC power line, avoiding shock hazards and stepping down the voltage to the desired value. Current follows the transformer leads to a rectifier (diode) and a filter (capacitor) in a typical configuration like the one shown in Figure 205.

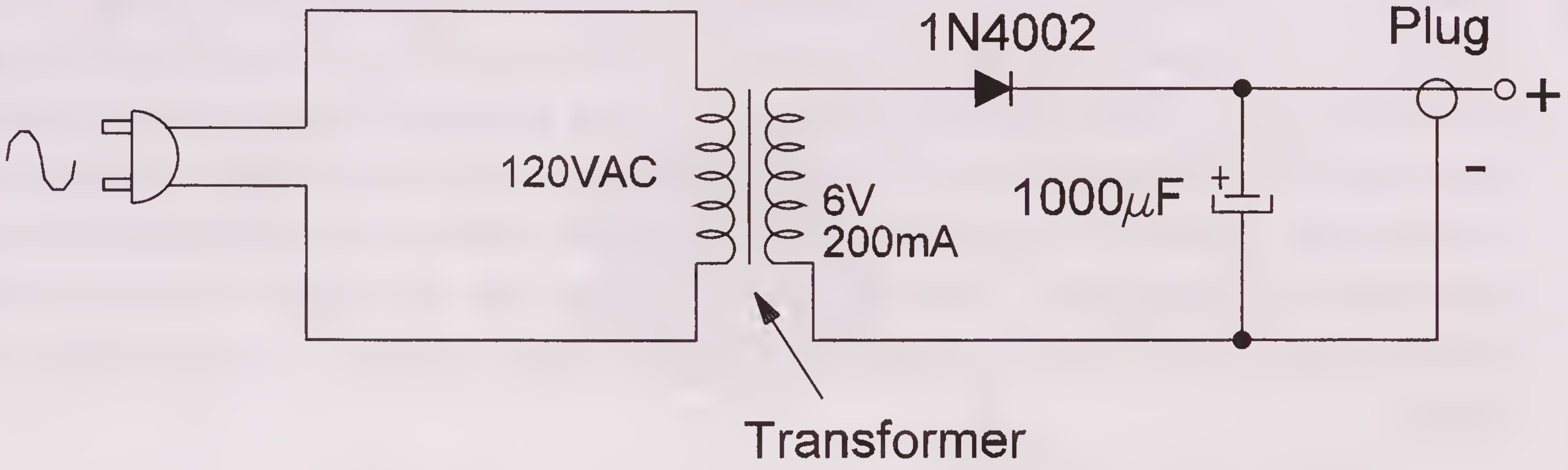

Figure 205

The output to the powered device is normally made by a cable with a plug as shown in the same figure. There are some details an electrician should pay attention to when using an AC/DC converter.

- Output voltage—output voltage can vary from 1.5V (one cell) to 12V (eight cells or a battery) for common AC/DC adapters. In some types, this voltage is fixed; you choose the voltage when buying the unit based on the voltage of the appliance. It is easy to identify this voltage from the number of cells. Each cell represents 1.5V, so if the powered device uses four cells it needs an AC/DC converter for 6V (4 x 1.5V). Some converters have a switch to choose the voltage. This switch can be adjusted based on the voltage of the powered device. It is important to make the correct adjustment because if it is a higher voltage than supported by the device, that device can be damaged.
- Output current—nominal output current is determined by the consumption of the appliance you intend to power. Normally, the consumption is evaluated by the size and type of the cells or batteries used to power it. In many of the cases, the current is indicated in the device, but if this is information is not available, use the next table to choose an AC/DC adapter.

Type of Cell	Current
AA	100 to 200mA
C	300 to 500mA
D	1 to 1.5A
9V cells	50 to 200mA

- Polarity of the plug—AC/DC adapters can be supplied by plugs with a positive or negative pole in the center. Be sure that the correct plug is used because if the polarity is inverted, the appliance can be damaged. Some AC/DC adapters have a switch to choose the polarity of the plug. Figure 206 shows a universal power adapter that can be plugged directly into the AC power line. This kind of adapter can be used to power small appliances, such as CD players, toys, video games, calculators, and radios.

An important point to observe is that common AC/DC adapters, in general, are very simple circuits with no voltage regulation stages. This means that the circuit presents a higher voltage in the output until the moment the appliance is plugged in, then it drops. This fact is important when testing an AC/DC adapter; the measured voltage in the output is normally higher than the adjusted or expected load until the device is powered on. When powering appliances where voltage is critical from the AC power line, it is important to use the specified types of adapters recommended by the manufacturer.

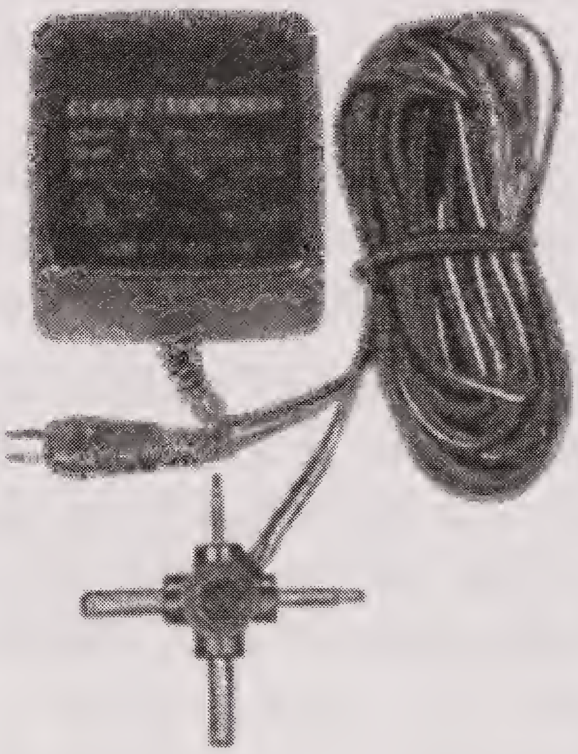

Figure 206
A Universal Power Adapter

Dimmers/Power Controls

Dimmers and power controls are devices used to control AC loads from an AC power line. They can be used to control the brightness of incandescent lamps, the temperature of heaters and small ovens, or the speed of motors including fans and pumps. The most common dimmer circuit is the one described in the Triac section starting on Page 162, where a sample configuration was analyzed.

This sort of circuit is usually small enough to be placed in the wall as a common switch, as shown in Figure 207. This dimmer can directly replace a wall switch. Equivalent dimmers using slide controls are also common and easily found for purchase.

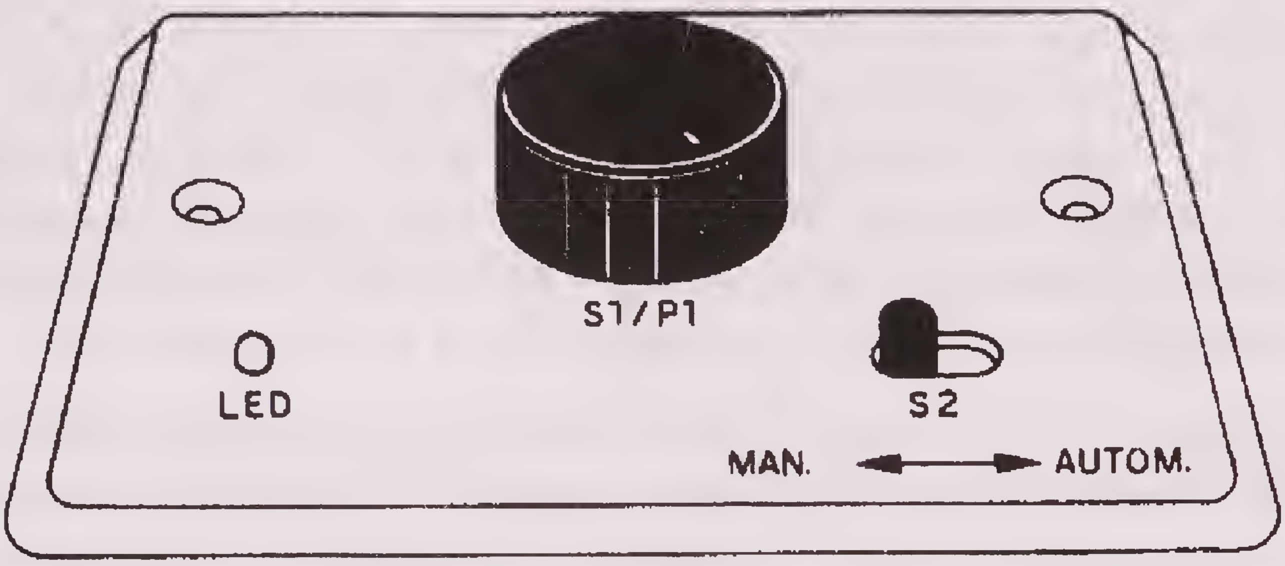

Figure 207

Another type of dimmer that is becoming popular is the touch dimmer. No mechanical parts are used to control the power applied to the load in this dimmer. The touch controlled dimmer uses a small metal plate where, by the touch of your fingers, the power applied to the load changes, slowly rising or falling to the desired level.

There are some special integrated circuits that have been designed for the task of operating a device with touch. One of them is the SLB0587 from Infineon, which contains all the processor circuits to slow up and slow down the power applied to the load from the touch in a sensor.

Dimmers can replace any wall switch in basic applications. When working with a dimmer an electrician must be attentive to some details.

A. Be sure the dimmer can control the load. Fluorescent lamps can't be controlled by common dimmers. Electronic appliances, such as audio amplifiers, TVs, VCRs, and some others can't be controlled by a dimmer.

B. Invert the positions of the wires in a dimmer if it doesn't control the load properly.

C. Use filters (see the sections on SCRs, triacs, and EMI for more information) if the circuit causes interference in other electronic equipment.

Inverters

Batteries are low-voltage DC power supplies and can't be used to power devices plugged into the AC power line. Inverters or DC/AC converters are designed to convert a low DC voltage from a battery or cells into a high AC voltage like the ones found in an AC outlet. An inverter is a complex circuit that not only has to step up the low voltage found in the battery or cell, but also convert it into an AC voltage with the same wave shape and frequency of the AC voltages found in the power line.

Keep in mind when using an inverter that it is not possible to create energy. This means that all the power needed by the load must be generated by the batteries or cells. Batteries and cells are low-power energy sources, which means that the amount of power sourced by an inverter can't be high, or under any circumstances, higher than the capabilities of the cells or battery. As a general rule, inverters powered from cells or batteries are only used to power low-consumption appliances, such as large fluorescent lamps in emergency systems, signaling devices, and xenon beacons. Inverters powered from car batteries or other heavy-duty batteries can source energy to medium-power appliances like small TVs, electric razors, small fluorescent lamps, transmitters, laptop computers, small xenon beacons, and incandescent lamps.

How they work

A typical inverter is formed by a low-frequency oscillator running on the AC power-line frequency and a high-power output stage using transistors (bipolar, power FETs, or IGBts) to drive a transformer, as shown in Figure 208.

The signal applied to the low-voltage winding of the transformer appears as an AC high voltage in the secondary winding. The quality of the transformer determines how much power can be transferred from the low-voltage to the high-voltage circuit, determining the efficiency of the converter. Values as high as 90 percent are common in commercial types, meaning that for each 10 watts of power drained from the battery, 9 watts are load and 1 watt is transformed into heat by the circuits.

Most of the heat is converted by the output transistor because they are mounted in large heat sinks. In commercial types, the metallic box is used as a heat sink.

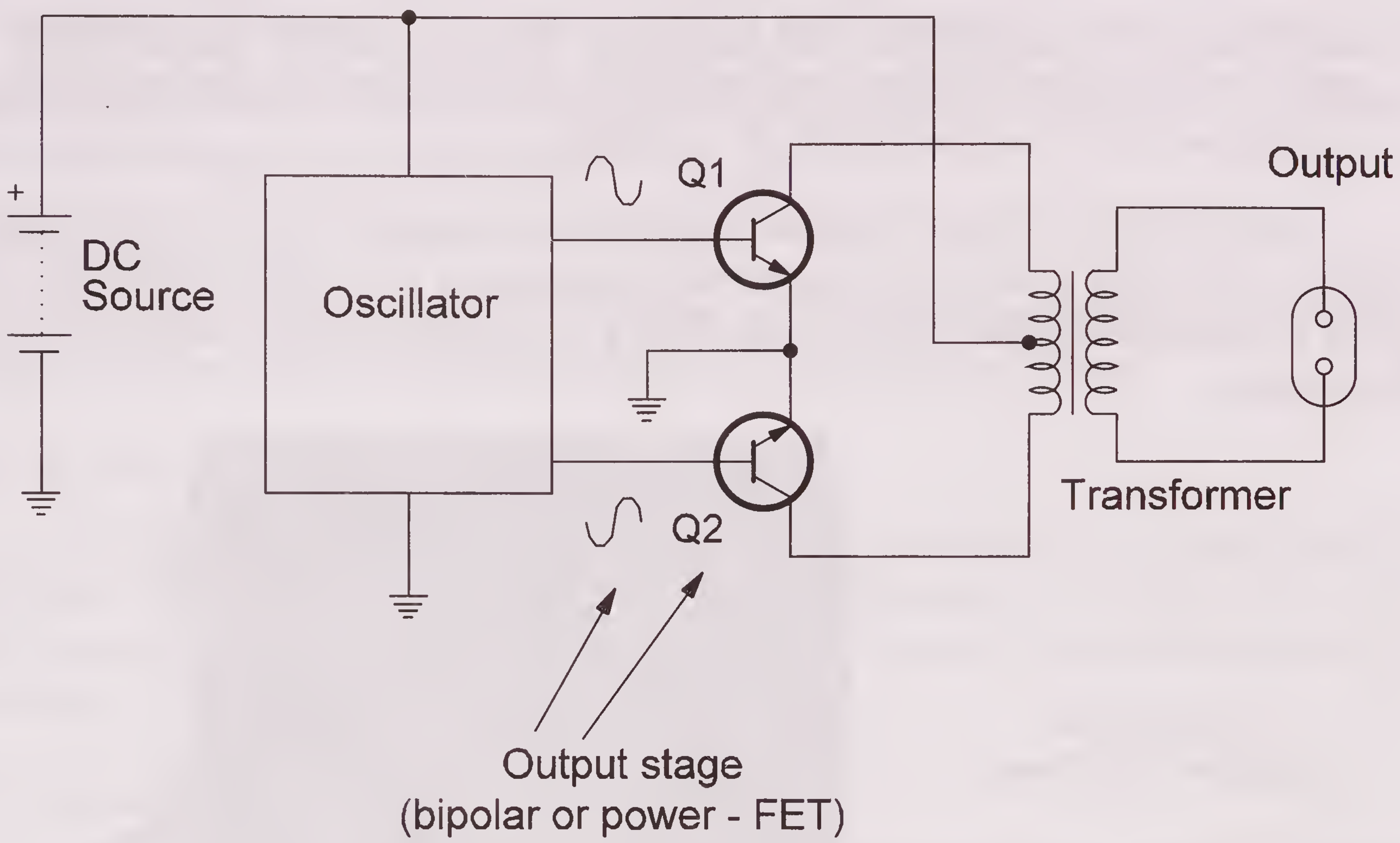

Figure 208

It is important to observe that in many types, the wave shape of the output voltage is not sinusoidal. The conversion made by circuits using pulses and the construction of the transformer can result in different wave shapes. In the applications here, the wave shape is critical because it is important to see if the inverter is working as needed. For an electrician, the inverters are important because they are found in emergency lighting systems or powering devices that are normally powered from the AC power line.

The following are some important applications for inverters.

- They can be used to power electric appliances from batteries and cells in emergency situations or in places where AC power is not available, such as cars, boats, and on camping trips.
- Inverters for fluorescent lamps are used in emergency lighting systems. The batteries used in these systems are charged by AC power-line voltages.
- They are used in no-break systems for computers, which supply energy for a computer if an AC power outage occurs.

- Xenon beacons and car ignition systems use high voltages that can be generated by inverters. For beacons the AC power-line voltage is stepped up by an inverter, and for ignition systems the car battery voltage is stepped up to values as high as 600V.

Figure 209 shows an inverter that can be plugged into an automobile's lighter outlet and used to power small appliances and other electronic devices by converting the 12V power from the battery of a car into 110-120V AC power.

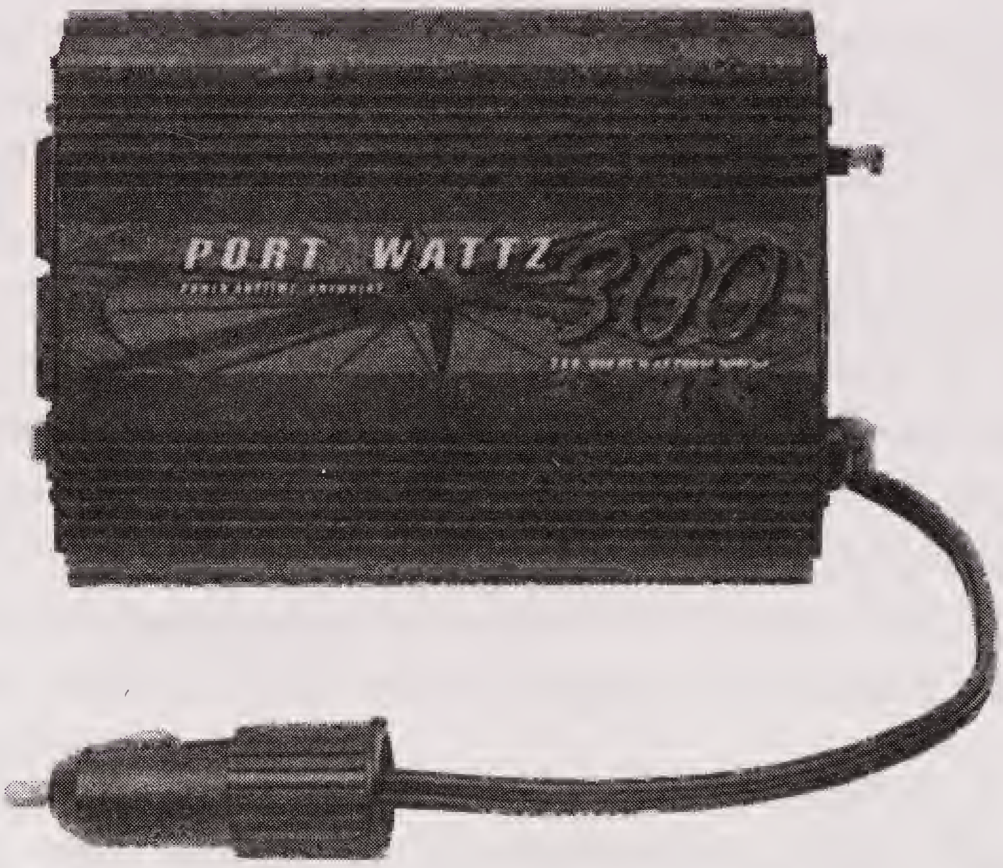

Figure 209
(Courtesy StatPower)

Working with inverters

It is very important for an electrician to know what kind of inverter is recommended for an application. Not only are inverters critical equipment, but also the powered circuits can be damaged if an incorrect choice is made. The following are main specifications an electrician must observe when working with inverters.

A. Output power

It is important to be sure that the inverter can supply enough power to the load. The power specification is given in watts (W). Take care with audio equipment where the output power doesn't reflect the consumed power; it can be given in a unit of watts, peak, or PMPO (instantaneous power peak).

B. Wave shape

The wave shape of many inverters isn't sinusoidal. This fact can be important in some applications. Lamps and heaters don't need a sinusoidal wave shape to be powered.

C. Performance

The higher the amount of energy converted, the better the inverter. Good inverters have a performance of 70 percent or more.

D. Isolation

The high voltages found in the output of inverters can cause severe shocks. Appropriate wires must be used to connect the loads if they are not plugged directly into an outlet, as in the case of fluorescent lamp inverters.

E. Battery placement

If not sealed, batteries used to power the inverter must be placed where the poisonous gases produced can be eliminated.

F. Wiring

Wiring to the battery can be done using wires based on the current drained by the inverter. A 100W inverter with an efficiency of 80 percent can drain a current as high as 12A from the battery. The wires must be able to handle this current.

Troubleshooting

The output transistor is the most fragile component in this type of circuit because it operates under limited conditions. When replacing these transistors, make sure that the type chosen supports the same voltage. The suffix indicates, in many cases, the operational voltage. For example, a TIP32A supports less voltage across it than the TIP32B.

Flashers

Lamp flashers are used in signaling. Garage doors, signal lights for customers in stores, and warning devices are some cases where flashers can be used. The simplest electrical type uses an incandescent lamp with some thermo-mechanics to open and close the circuit. But, with the aid of electronics circuits, the electrician can find many different configurations for lamp flashers.

A. Incandescent

Low-voltage flashers, such as the ones used in cars or powered from DC supplies, can be made with simple configurations like the one shown in Figure 210.

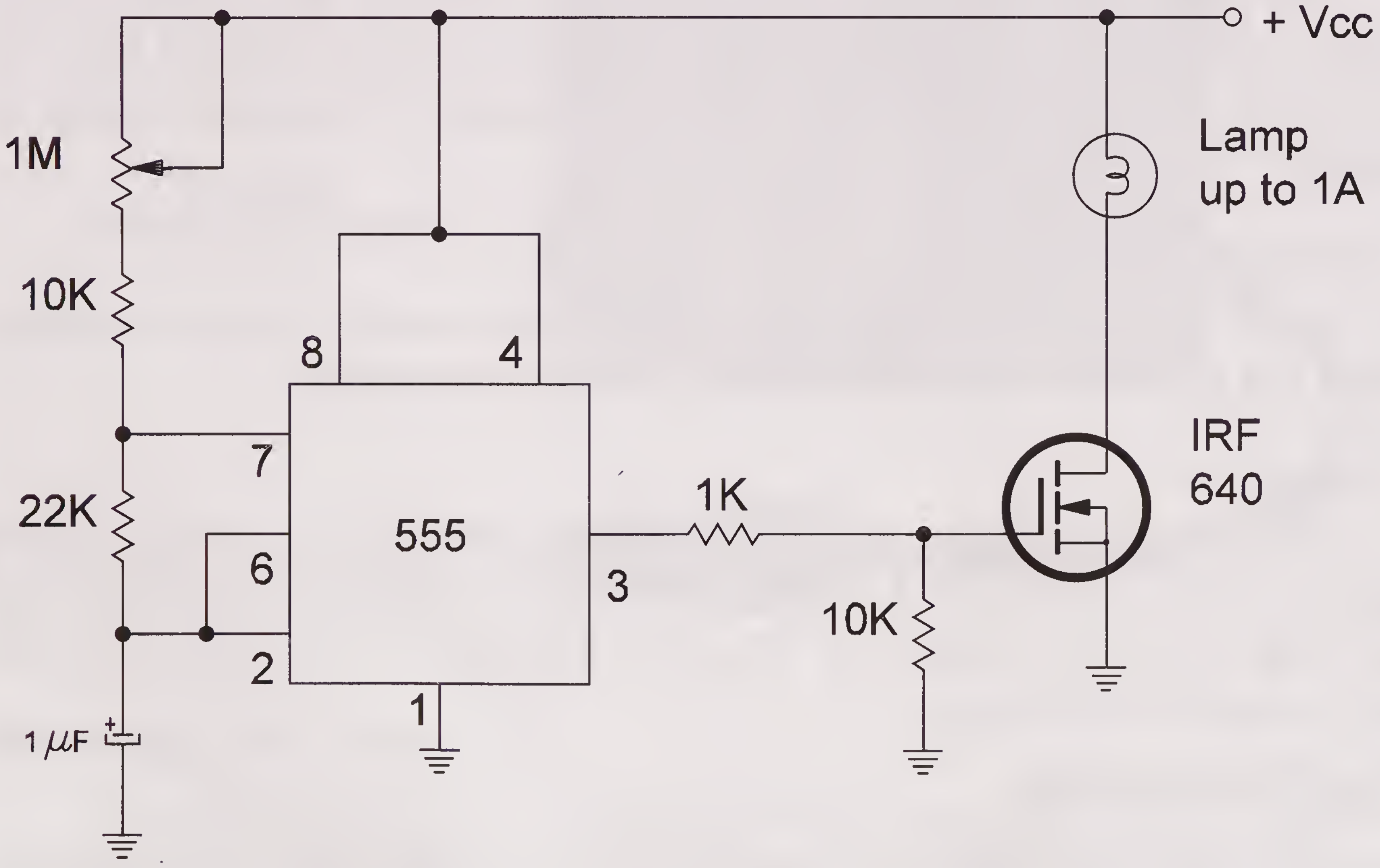

Figure 210

The power transistor (bipolar or FET) controls a lamp that flashes in a rate determined by the circuit adjustment. To flash a high-voltage, AC-powered incandescent lamp, a circuit using an SCR or a triac is more suitable. Figure 211 shows a circuit that can be found in garage doors using two lamps. The lamps flash in an alternating mode and the frequency is determined by an internal adjustment of the circuit.

In both cases the most sensitive of the components are the transistors and the SCRs (or triacs) that can open or short. When replacing these components, an electrician must observe the operating voltages.

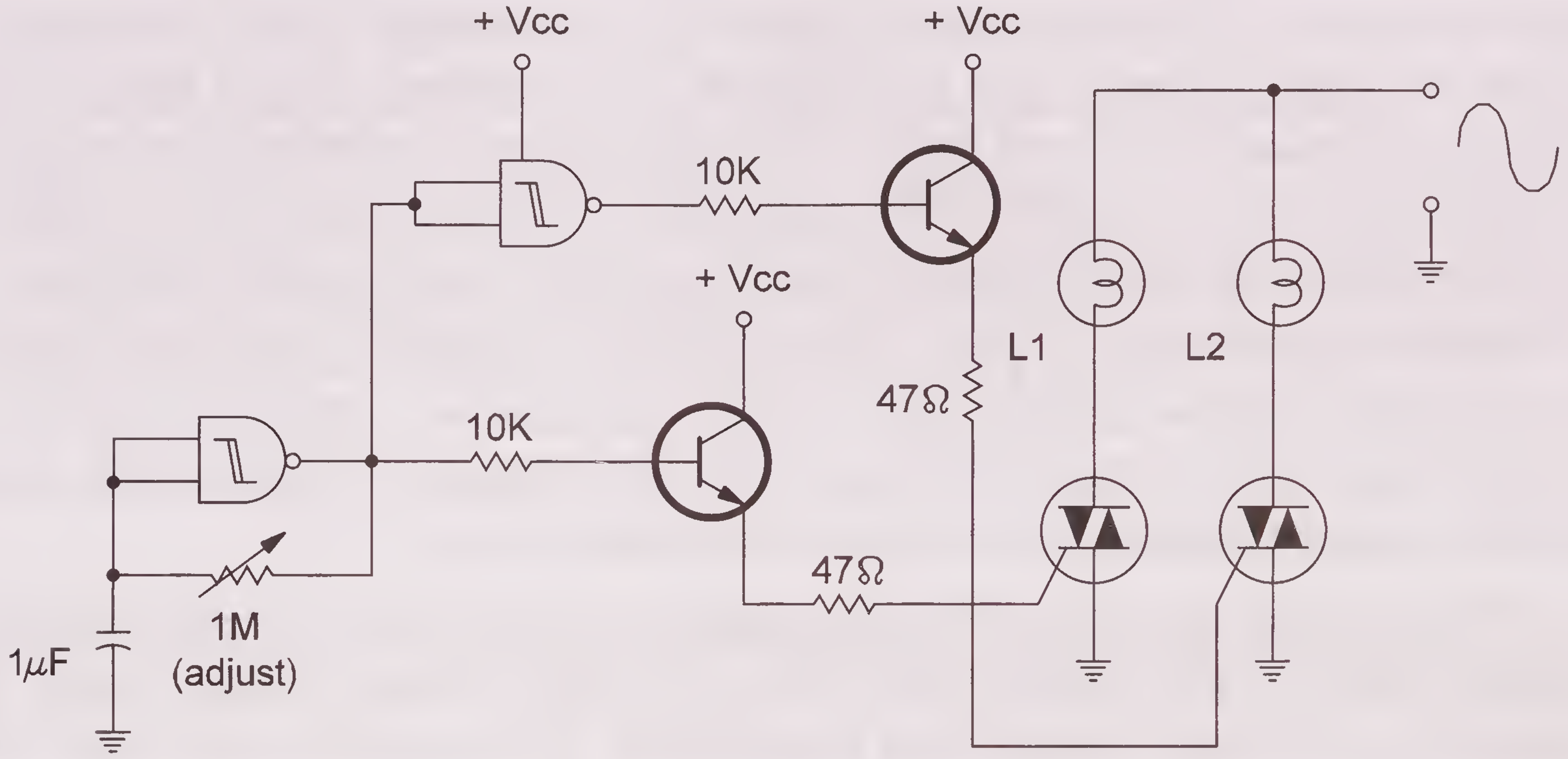
+ Vcc
+ Vcc
+ Vcc
10K
10K
47Ω
47Ω
L1
L2
1μF
1M
(adjust)

Figure 211

B. Xenon (strobe)

The xenon beacon or xenon flasher is used on towers and high buildings as a signaling device. Portable xenon beacons can also be used in emergency situations (powered from cells or batteries) or in police and rescue vehicles. The basic circuit of a battery-powered xenon beacon is shown in Figure 212.

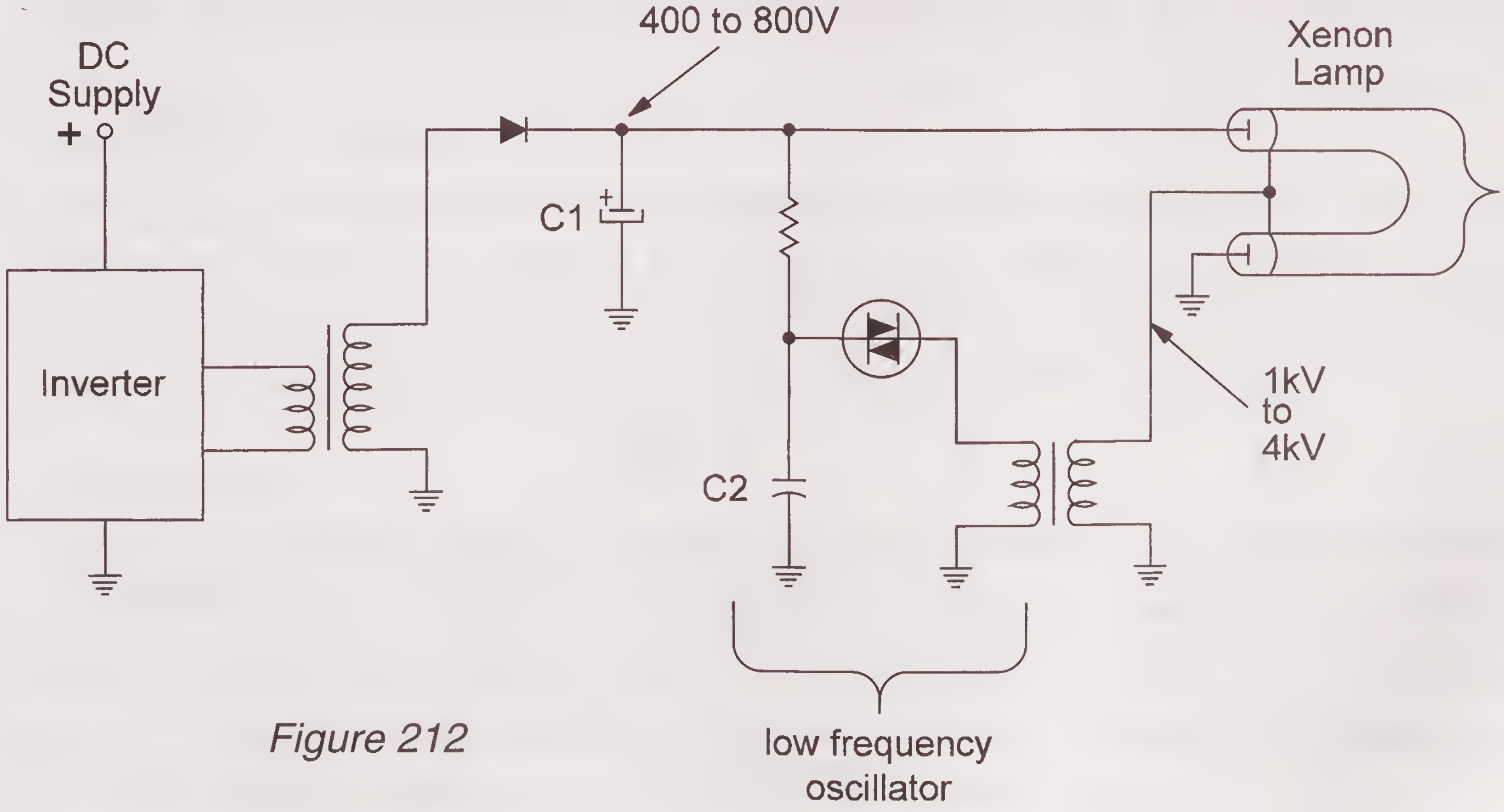
DC
Supply
+
400 to 800V
Xenon
Lamp
C1
Inverter
C2
1kV
to
4kV
low frequency
oscillator

Figure 212

A low-frequency oscillator operates as an inverter driving a high-voltage transformer. This transformer steps up the voltage to the necessary values in order to power the xenon lamp. For common xenon lamps the values are between 400 and 800V. The high voltage is rectified and used to charge a high-value capacitor. This capacitor (in some circuits many capacitors are wired in parallel forming a bank) is used to store electric energy.

At the same time, a very low-frequency oscillator produces pulses that are used to drive another high-voltage transformer. This second transformer has a secondary winding rated to voltages between 800 and 2000V—values necessary to trigger the xenon lamp.

The circuits can use SCRs, transistors, or even ICs to produce a low-voltage pulse to the transformer. When the high-voltage trigger pulse is applied to the lamp, it becomes conductive. The capacitor can then be discharged across it, transforming all the stored energy into a high-power, short light pulse. As long as the capacitor is discharged, the lamp becomes nonconductive again and a new operation cycle begins.

The output power, or luminous power, of a xenon beacon is measured in Joules (J), which is the energy stored by the capacitors. Typical values are in the range of millijoules (mJ).

The operation principle of these xenon flashers is the same as camera flashes. The difference is that, in the camera flash the circuit is triggered when the operator presses a button. In circuits operated from AC power, the inverter stage is not necessary. Figure 213 shows a circuit of this type.

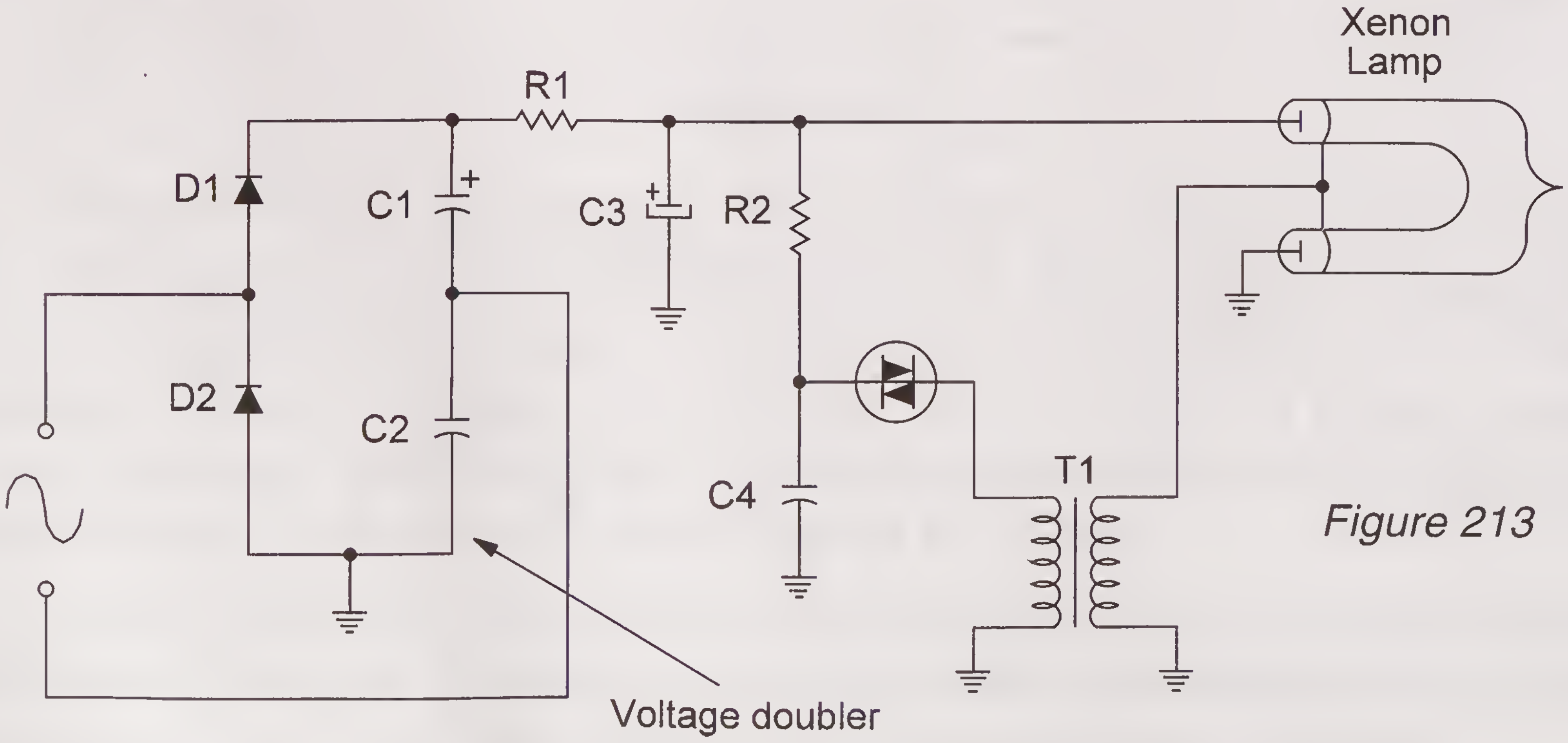

Figure 213

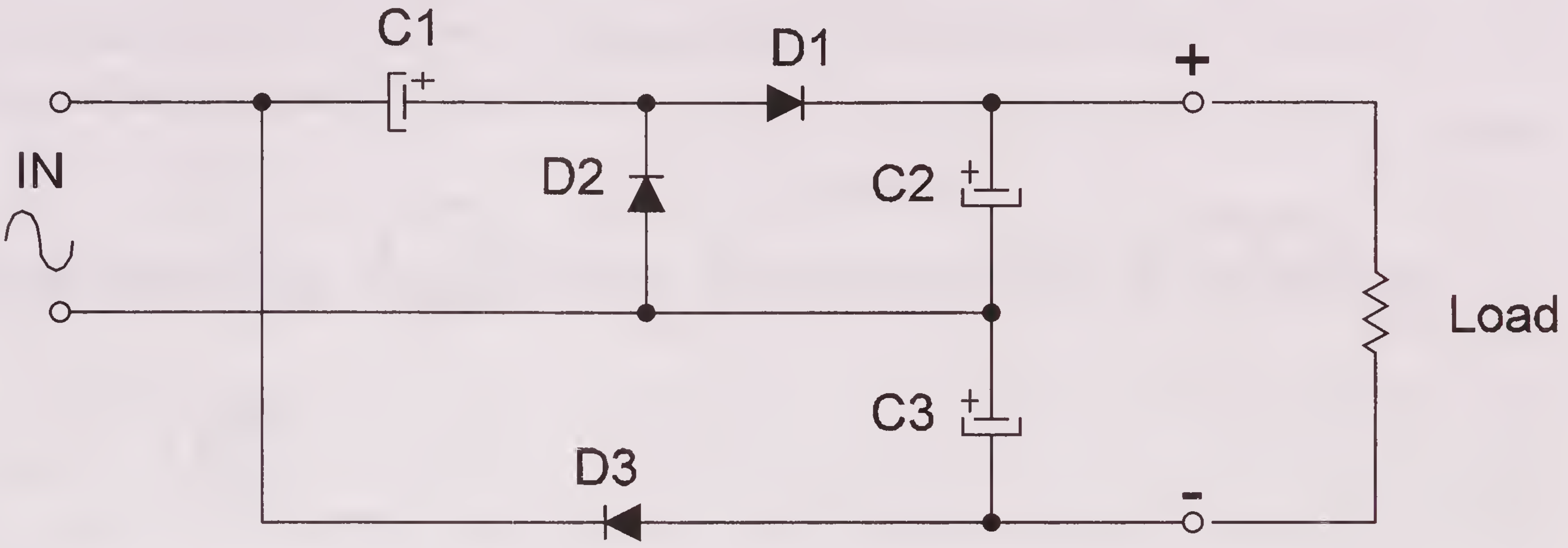

Figure 214

A voltage multiplier or a transformer is used to step up the AC voltage that can raise the necessary values for the lamp operation. A voltage multiplier is a network formed by diodes and capacitors that can step up the AC voltage converting it to a higher DC voltage as shown in Figure 214.

The important parts in this circuit are the diodes and capacitors that store energy. When working with these circuits an electrician must observe some specifications.

- When replacing the lamp, be sure that the replacement type has the same triggering voltage and the same power (mJ).
- Xenon flashers run with AC power are not isolated and can cause severe shocks if live parts are touched.
- In all devices, after the inverter stages, very high voltages can be found. These voltages are present even after the circuit is no longer plugged into the AC power source, as the voltage is stored in the capacitors. Do not touch the capacitors even when the device is unplugged.

C. Sequencers

Another type of luminous signaling device is the sequencer. This kind of device is found in the form of arrows indicating emergency doors or emergency exits and is even used in some types of Christmas tree lights. Figure 215 shows a typical sequencer circuit for incandescent lamps (or LEDs) that can be used in some signaling appliances.

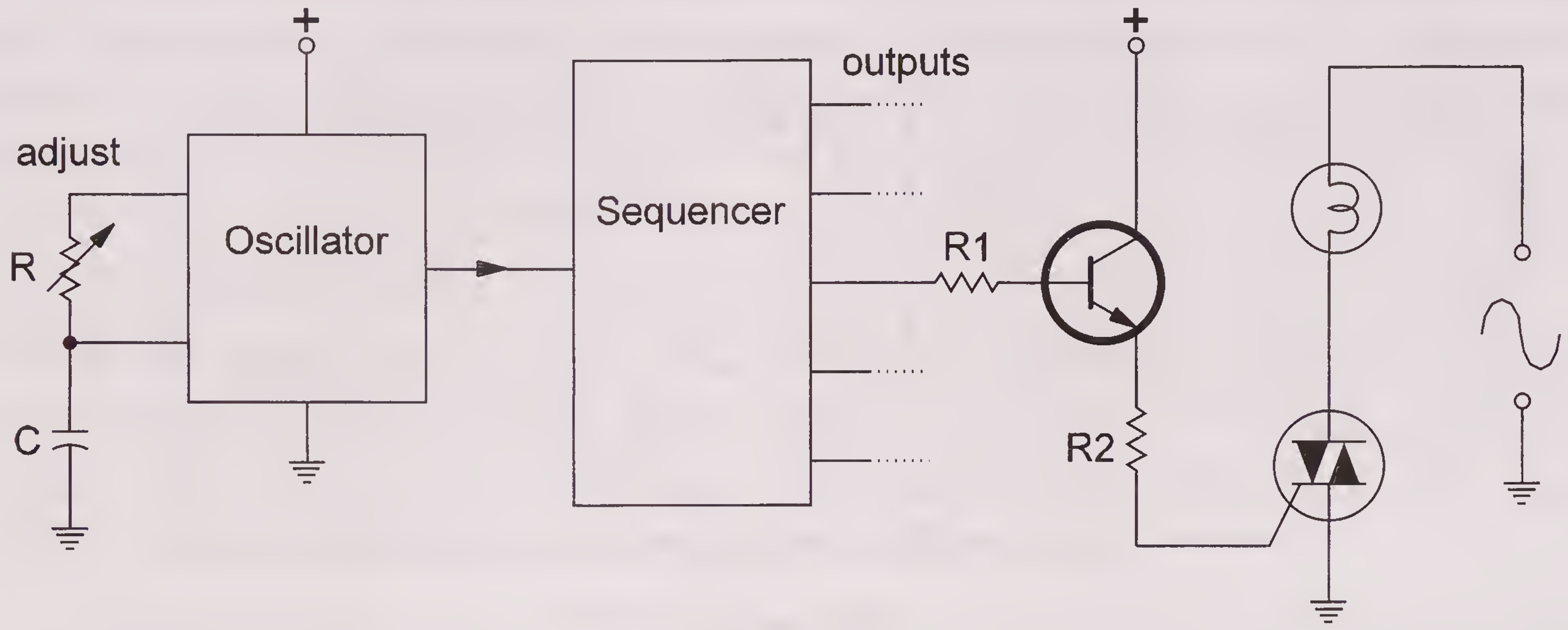

Figure 215

The clock circuit is an oscillator that determines the speed of the effect. The signals from this block drive the counter or another circuit that decodes the desired sequence or effect, such as a microprocessor or microcontroller. The decoded outputs of the sequencer circuit drive a transistor if low-voltage DC loads are controlled, such as LEDs or small incandescent lamps. If high-voltage AC loads are driven, SCRs or triacs are used in typical configurations like the ones shown in Figure 216.

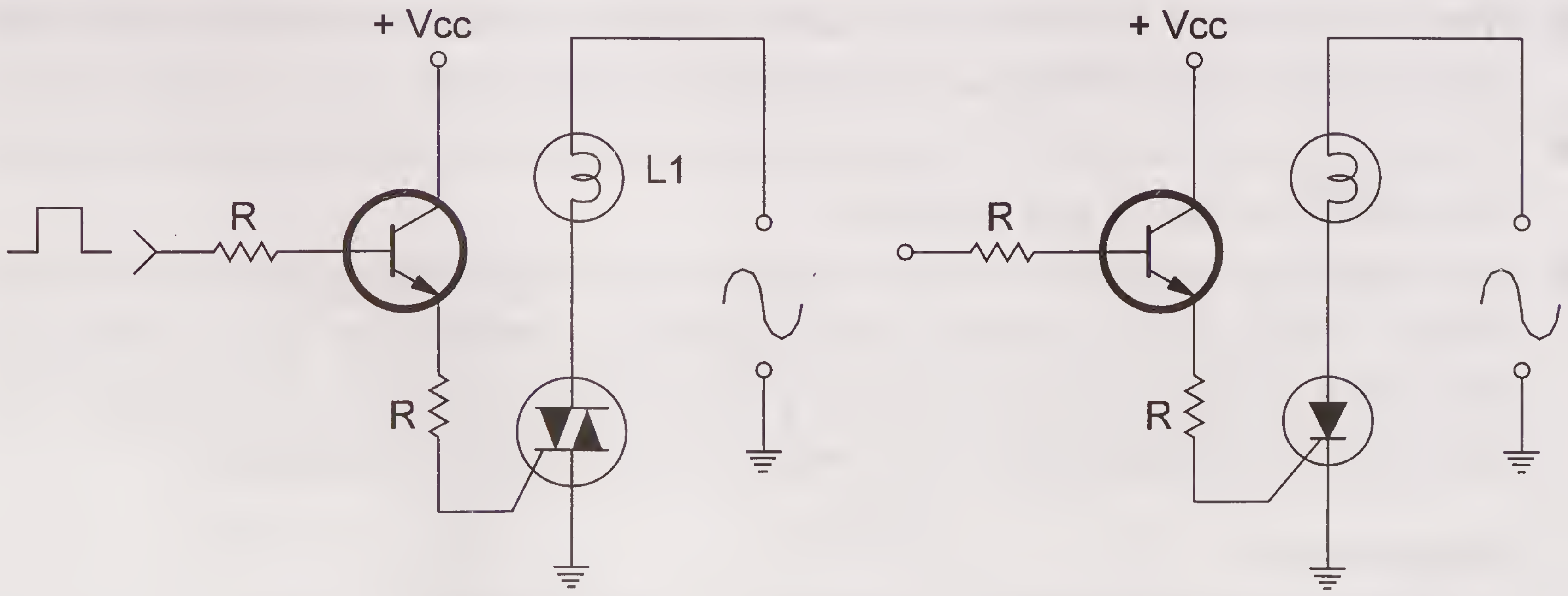

a) Driving a TRIAC

b) Driving a SCR

Figure 216

When using this type of circuit the electrician must keep in mind that the control block is a low-voltage circuit. The power circuit depends on the type of lamps used in the effect.

Automatic Lighting/Emergency Lighting

Some years ago, the common way to control a lamp—turning it on and off—was using a switch. With electronics today, many automatic ways to control lamps have appeared in electric installations. One of these controls is the automatic lighting system. This system is very useful in gardens, house entrances, and other areas where timed lighting is desirable. An automatic light turns on a lamp at dusk and turns it off at dawn. This kind of control is found in many formats from lamps plugged directly into the socket of a common incandescent lamp (Figure 217) to more complex systems that can control all the lamps in a garden or on a street.

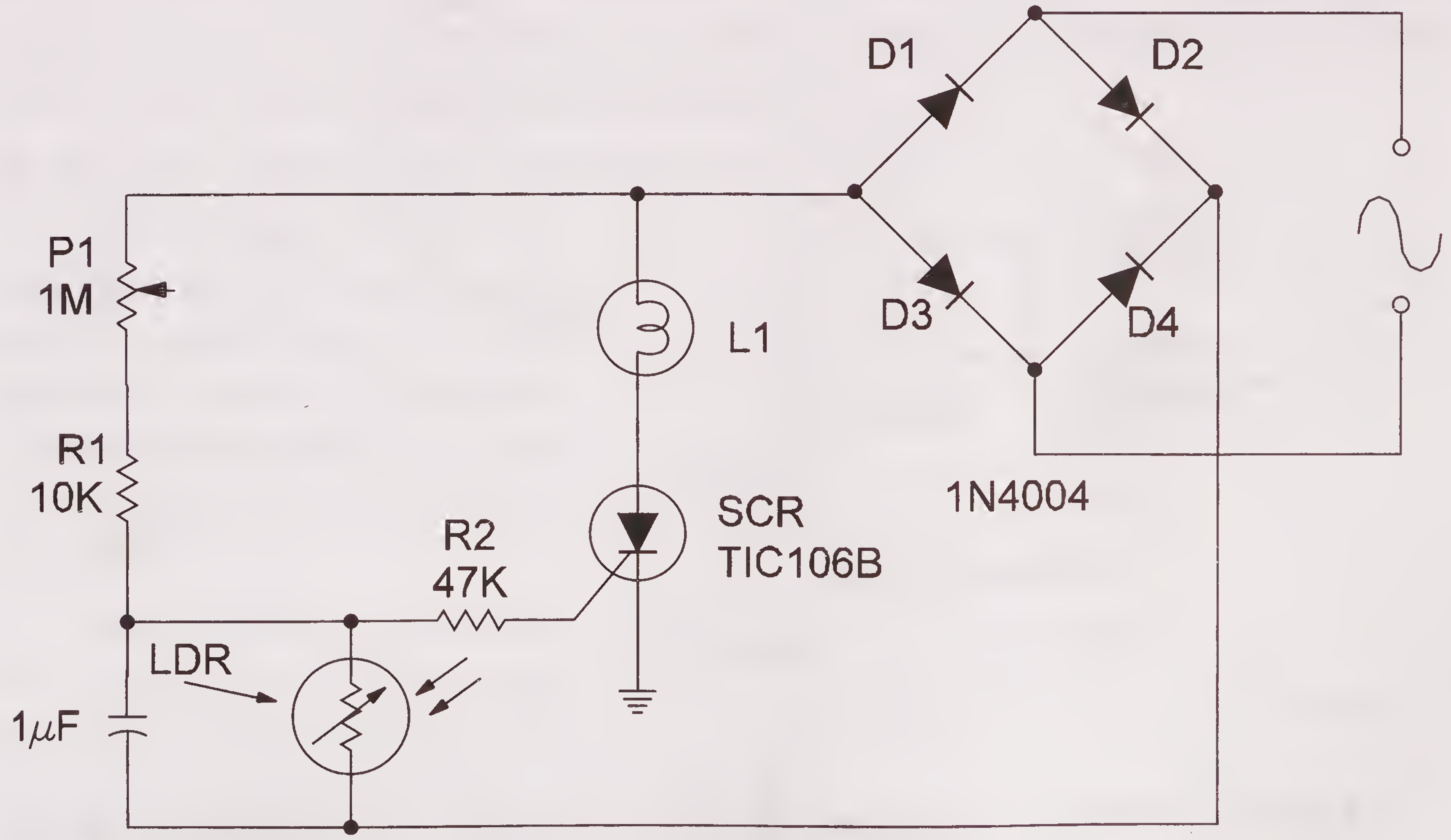

Figure 217

The sensor is an LDR or cadmium sulfide photocell that is mounted in such a way so that it receives only the ambient light and not the light from the lamp it controls, thus avoiding feedback. When the amount of light received by the sensor falls, the circuit triggers an SCR or triac that turns on the lamp. As with any circuit using SCRs or triacs, this kind of control can generate some radio interference (refer to the section on SCRs, triacs, and EMI).

Depending on the application, the electrician must install filters in the AC power line into which the device is plugged. When working with these devices, it is important to be sure that the sensor only receives ambient light. If light from the lamp is picked up by the sensor, a feedback process can make the circuit unstable.

Emergency lights turn on one or more lamps when the AC power fails. These lights are required equipment in public places, such as restaurants, movie theaters, schools, and many others. The basic emergency lighting circuit has an operation principle as shown in Figure 218.

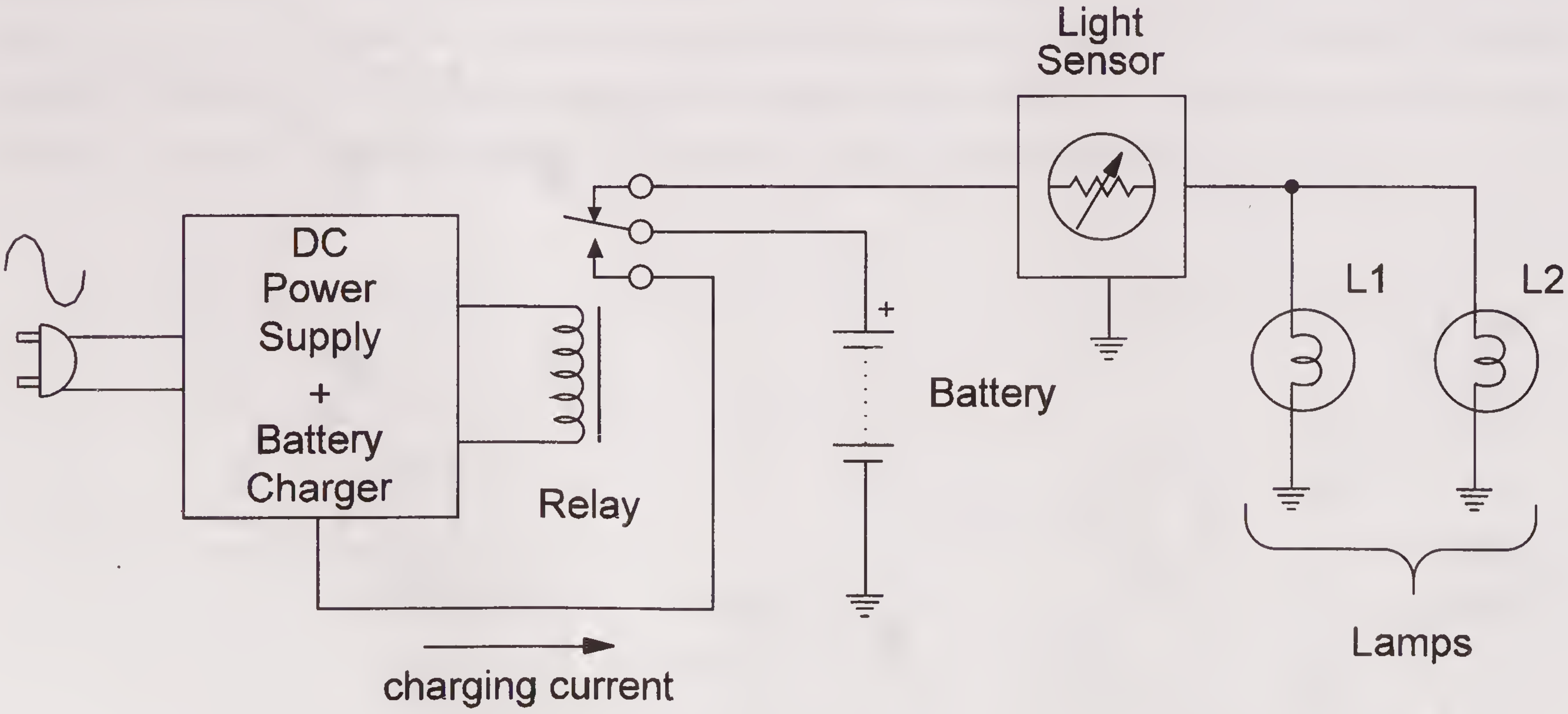

Figure 218

A battery charger is plugged into the AC power line, keeping a battery in a constant charging process. A relay or another switching circuit unplugs this block from the battery when the AC power fails and, at the same time, plugs the battery into the emergency lamps. A photocell is placed between the blocks to give the circuit some "intelligence." This block only turns on the lamps if the location is dark. It is not necessary to turn on the lamps during an emergency if the area is illuminated by natural light. In some systems there is an inverter between the battery and the lamps. This inverter is necessary to drive fluorescent lamps.

Alarms

Home security is increasingly becoming more important. High-tech equipment is available and can easily be installed in any home. The degree of sophistication of systems can vary from simple switches that trigger an alarm if a door or window is tampered with to complex microprocessed systems that dial a security provider's or the police's phone number if some disturbance is detected. Electricians are sometimes called to prepare a home to receive some kind of alarm system, which may include placement of sensors and wiring. They must also install the power supply for the system, which includes outlets, wires for sirens, and other alarm transducers. So, it is important they know how the basic system works to help in the installation or troubleshooting. The following are some of the basic alarm and security systems.

- Magnetic sensors and switches

The simplest alarm system uses a switch to trigger a siren or another sound or light source. Figure 219 shows a diagram of a system that uses a magnetic switch placed in a window.

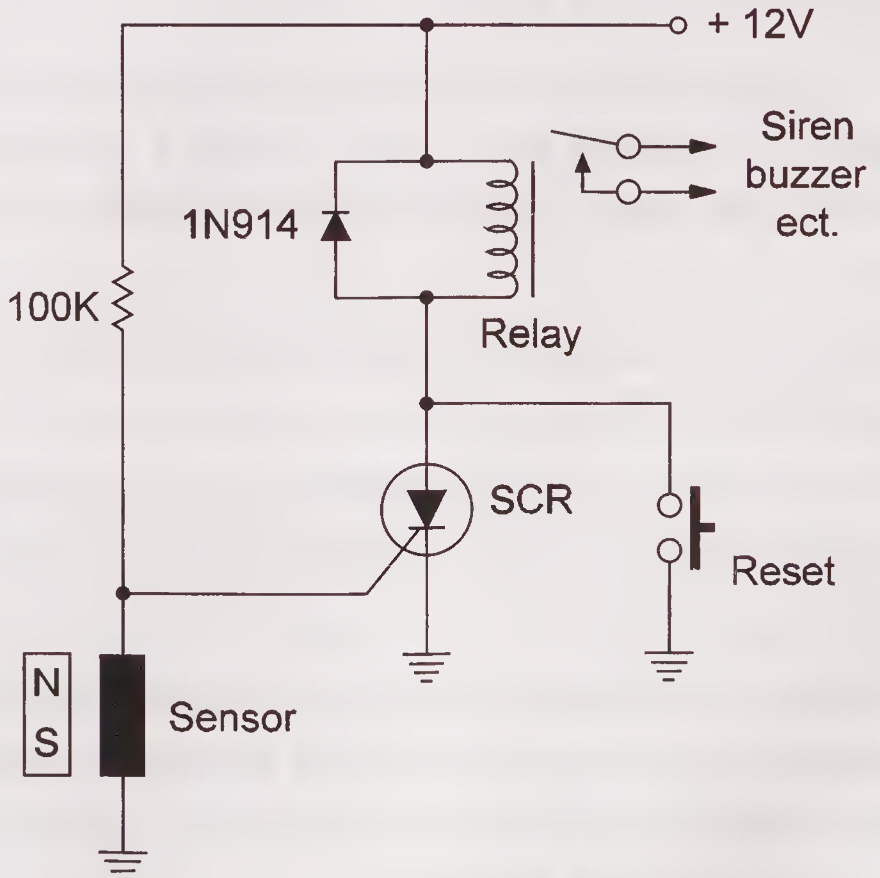

Figure 219

With the window closed, the magnet is in contact with the reed switch, keeping the contacts closed and the circuit off. If the magnet is moved away from the reed switch, the contacts open and the

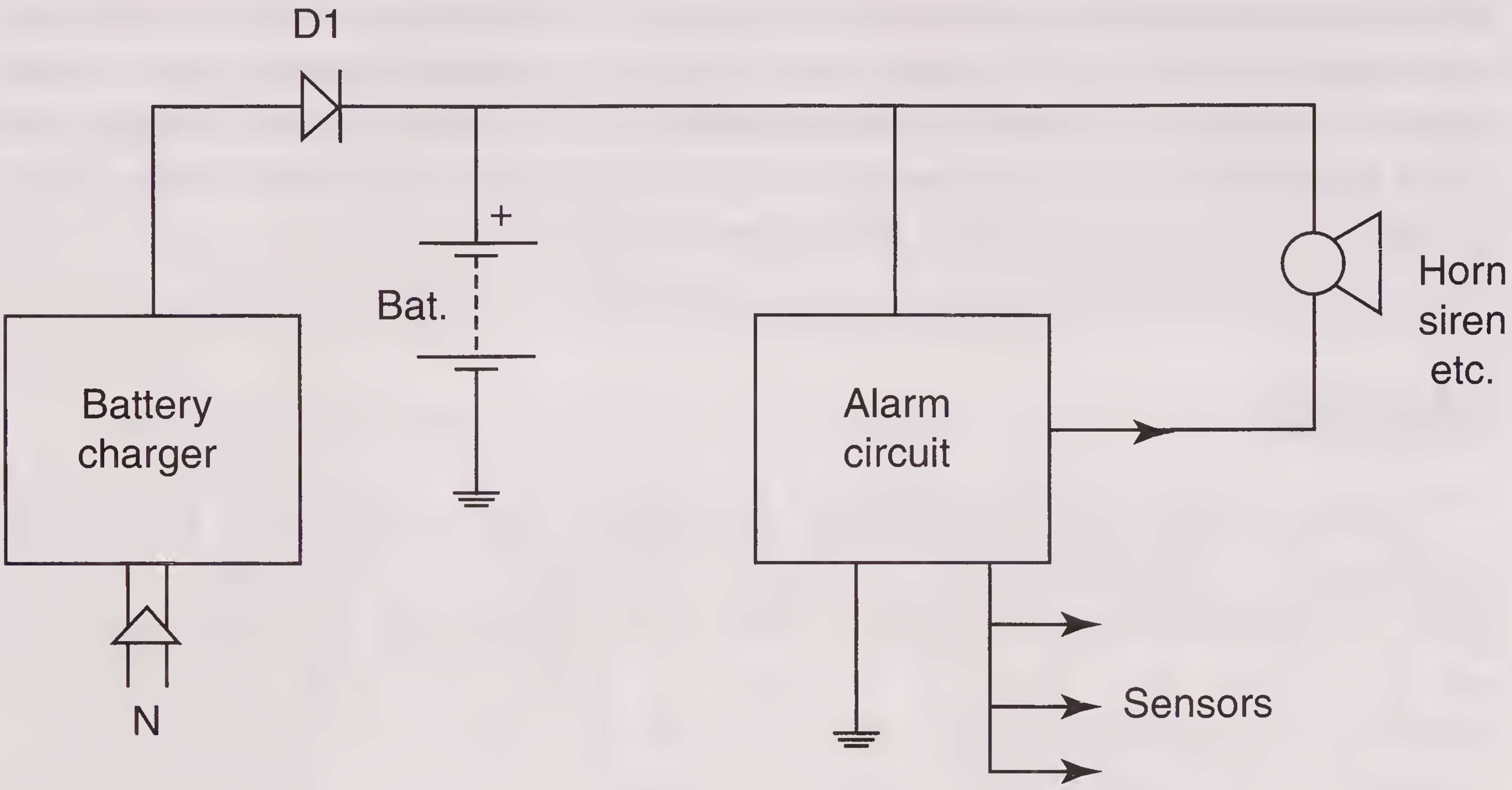

Figure 221

alarm is triggered. Many switches of this kind can be wired in series to protect all the windows and doors of a house or building. The relay is open until one of the sensors is activated. Once triggered, even if the window or door is closed again, the circuit remains on. To stop the alarm it is necessary to turn off the power supply, pressing S1 for a moment. Any number of sensors can be wired in series.

Notice that in this circuit, the current flowing across the sensor is very weak, allowing the use of cells or batteries for power. Radio Shack , Dimango, and other dealers offer the sensors for this type of alarm, and also the detectors, the sirens, and in some cases, complete kits to protect a home. A complete alarm system includes additional resources as shown in Figure 221.

In this system, a battery charger keeps the battery charged to power the alarm even if an intruder cuts the AC energy supply. The circuit also includes other sensors, such as microswitches (small switches activated when pressed or if an object is moved). When installing or repairing this kind of alarm an electrician doesn't need to be extremely experienced as many of the systems are simple, but must take care with the isolation of the wires to the sensors.

Another system includes in the sensor a radio transmitter that sends a signal to a receiver. If the sensor is activated, the signal changes and the alarm is triggered. This system is suitable for cases where there is no space to pass common wiring.

- Photoelectric

The use of photo sensors in alarms is common in many applications. Figure 222 shows a passage alarm or passage detector using a phototransistor as the sensor and an IR (infrared) LED as the radiation source.

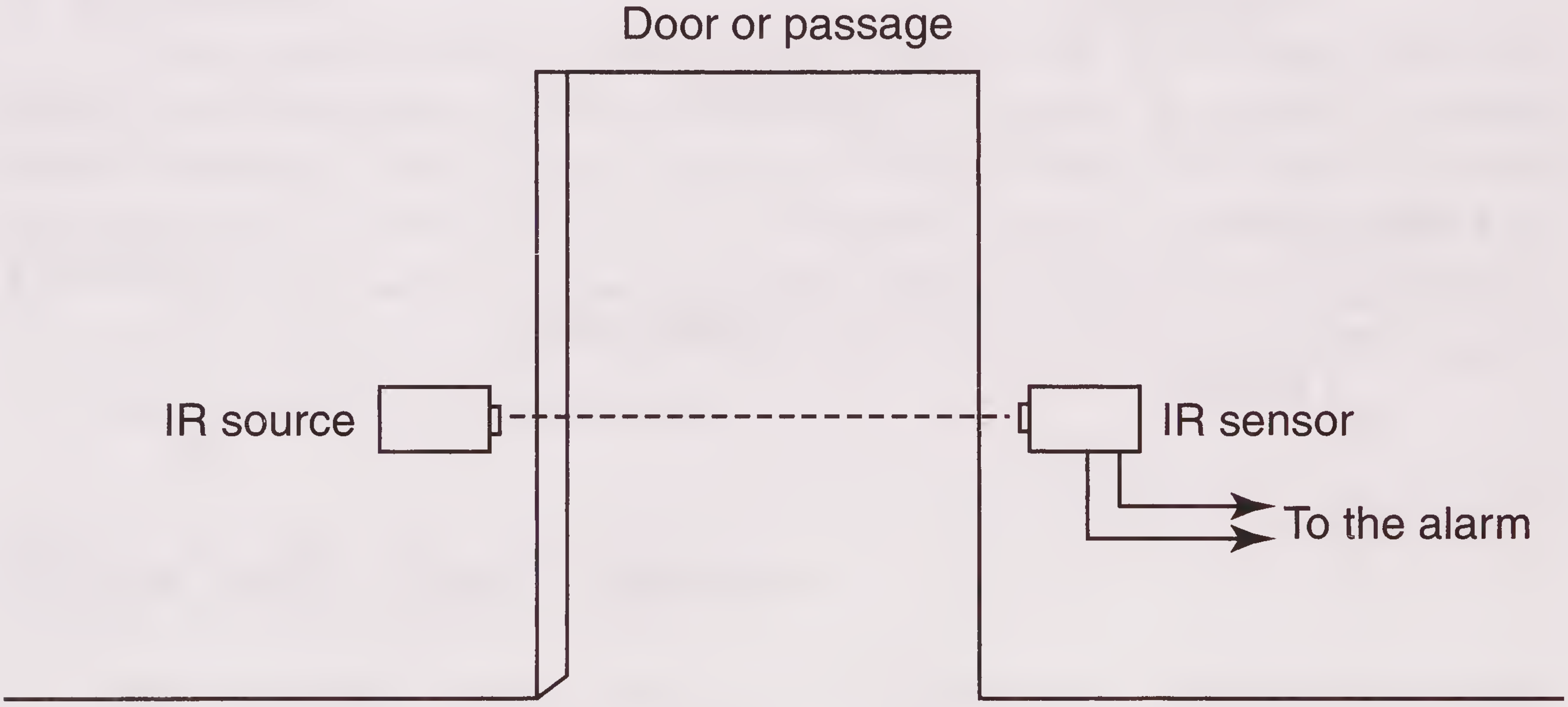

Figure 222

This circuit detects when a person or object cuts the radiation beam, even for a fraction of a second, in the sensor. The use of IR is recommended as the intruder can't see it, and therefore can't see where the system is placed.

Many circuits operate with modulated IR, meaning the circuit doesn't produce a continuous light beam, but produces short pulses in a certain frequency to be detected by the sensor. This procedure is used to avoid an intruder using a "false" IR source to stall the alarm as suggested in Figure 223.

When installing or working with this kind of device, an electrician must be careful with light sources that can cause interference in the system, such as fluorescent lamps. If large areas must be protected, the alarm can use a laser as its IR source (Figure 224).

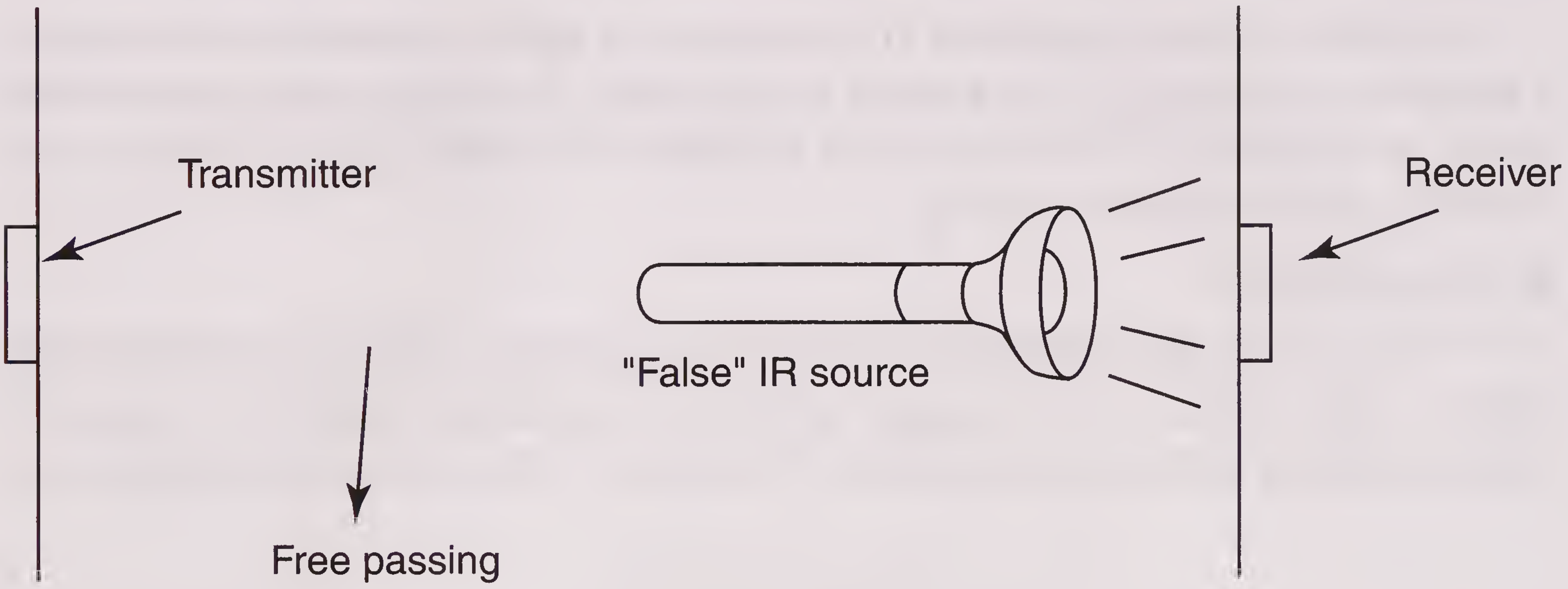
Transmitter
Receiver
"False" IR source
Free passing

Figure 223

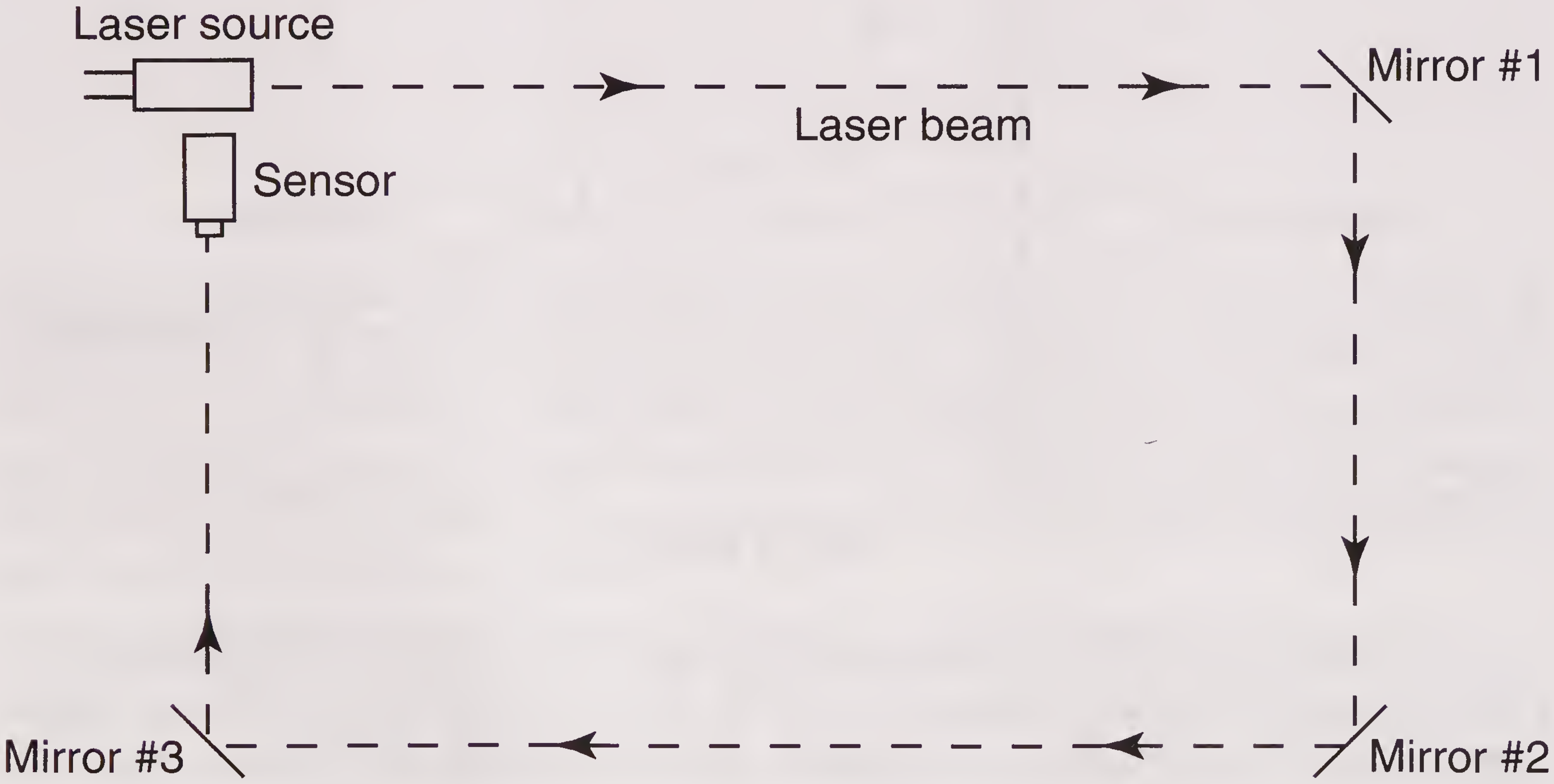
Laser source
Mirror #1
Laser beam
Sensor
Mirror #3
Mirror #2

Figure 224

Mirrors can bend the IR laser beam, fencing a closed area against any intruder. If the laser beam is disrupted on any side of the area, the alarm will be triggered.

- Motion Detectors

Motion detector alarms use a piroelectric sensor. It is formed by piezoelectric material that presents a surface electric charge that changes according to the amount of IR radiation falling onto it. Because a certain degree of IR radiation always exists, the circuits are designed to detect small changes in this radiation from the presence of a warm-blooded creature, such as a human. To detect the very small changes in the radiation, the circuit used is extremely sensitive and can be complex, involving special ICs.

- Vibration Alarms

Small vibration sensors are placed in home or shop windows to detect any type of impact or vibration. The circuit for this type of sensor is the same as in the magnetic switch.

- Ultrasound

In the ultrasound or super-sound system, the air is filled with inaudible sound vibrations (above 20,000 Hz). A sensor is placed to monitor these vibrations. The system makes an “image” from the pattern of direct and reflected sound waves and any alteration due to movement of an object—an intruder—when detected by the circuit. The sensors and the emitter must be placed in the positions indicated by the manufacturer as it is sensitive to external interference.

Doorbells And Chimes

The traditional electric doorbell is formed by a coil (electromagnet) and a metal plate that vibrates to produce sounds when powered by AC voltage. Modern doorbells and chimes use electronic circuits to produce musical sounds, such as notes in sequence, melodies, and even bird songs. In some cases, sequential systems act on solenoids that beat on metal tubes with varying lengths.

The basic circuit of an electronic doorbell in blocks is shown in Figure 225.

When triggered by the doorbell button, the circuit runs a program that produces a sequence of notes programmed in the chip. The signal corresponds to the notes and is amplified by a common low-power audio amplifier. Audio amplifiers with output powers in the range between 1 and 5 watts are common in commercial circuits.

The way this kind of circuit is installed is not much different from any common doorbell. The trigger button (pushbutton) is placed near the door.

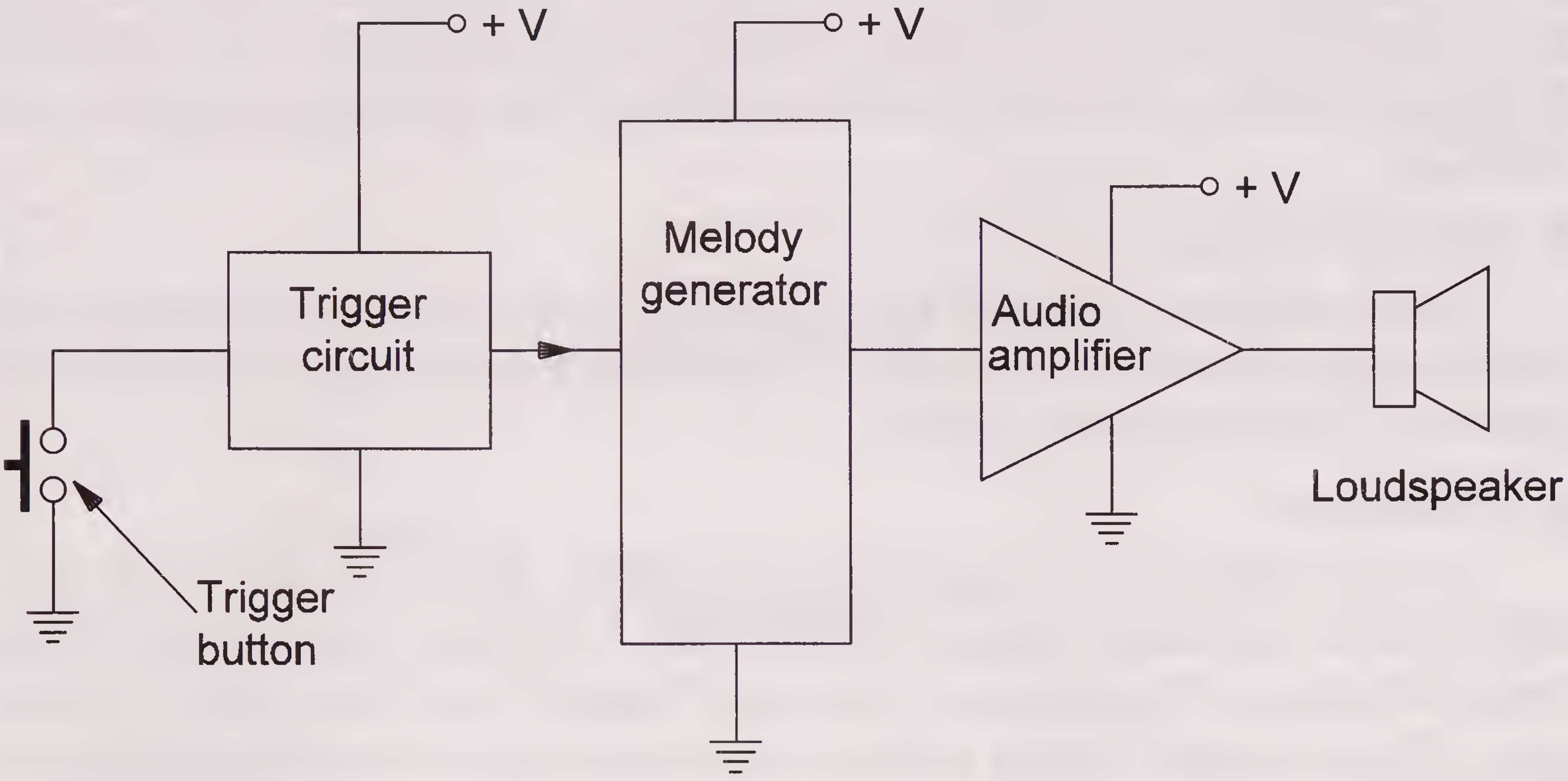

Figure 225

Many types allow more than one trigger point. The circuit is powered from the AC power line with a transformer to supply DC power. The power supply providing the low voltage to the DC circuit needs to remain on all the time. Some types of doorbells and chimes are coordinated with intercoms and video cameras.

Battery Chargers

Two types of battery chargers are used depending on the batteries to be charged. One is the small type used to charge Ni-Cad batteries and cellular telephone batteries or those embedded in wireless telephones and intercoms. This circuit is nothing more than a low-voltage DC power supply with a circuit that supplies constant current to the cells to be charged. The basic idea is that, by forcing a current to flow in the inverse direction of the normal current supplied to the load, a battery is recharged (Figure 226).

In this circuit, the resistors determine the amount of current used to charge a battery based on its type. This means that battery chargers have specifications indicating what type of battery they can accommodate. If the

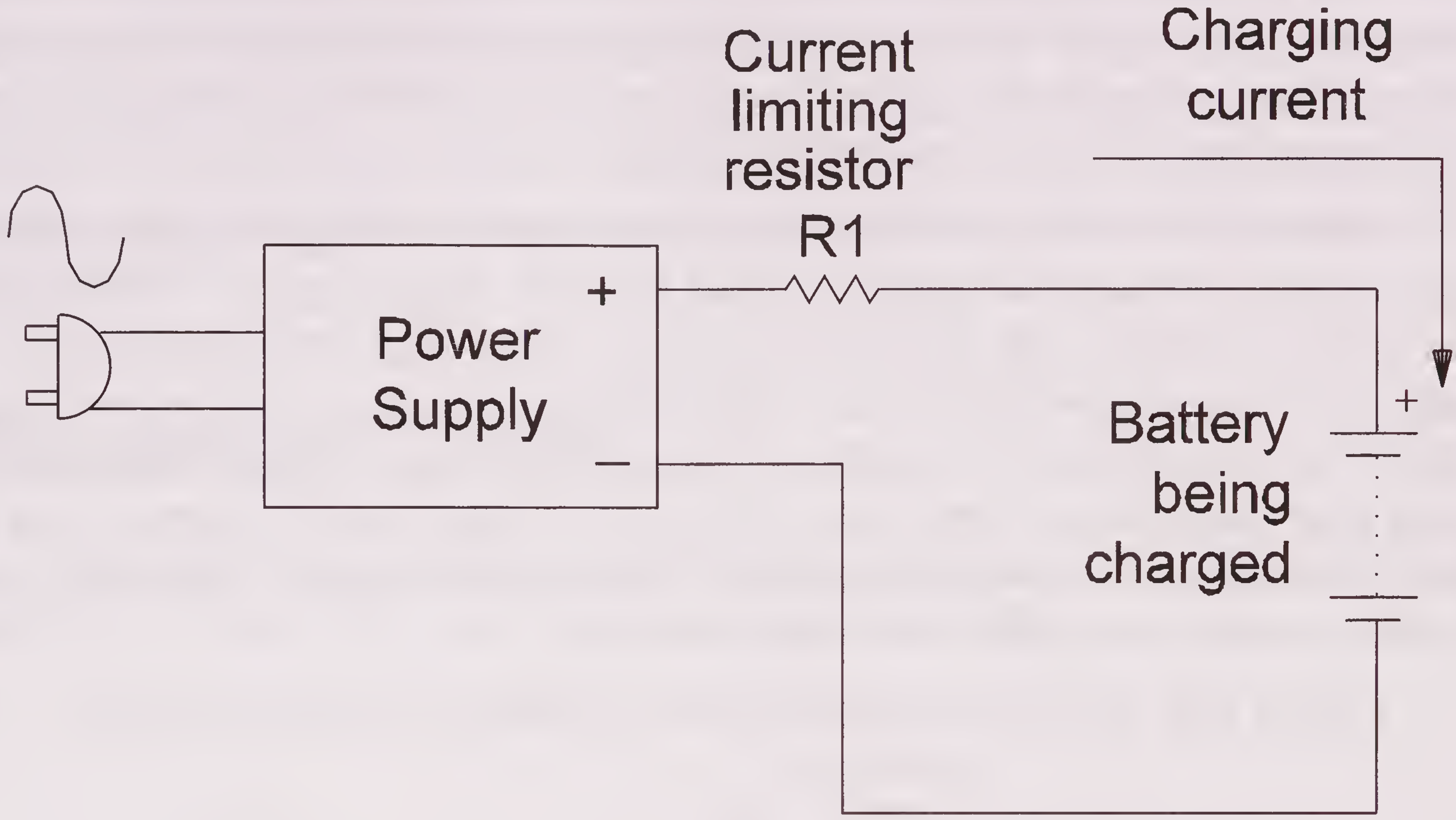
Current limiting resistor
R1
Charging current
Power Supply
+
–
+
Battery being charged

Figure 226

incorrect current is used, the battery life can be reduced. The most common types are the ones used to charge AA, C, D and 9V, and ni-cad batteries (Figure 227).

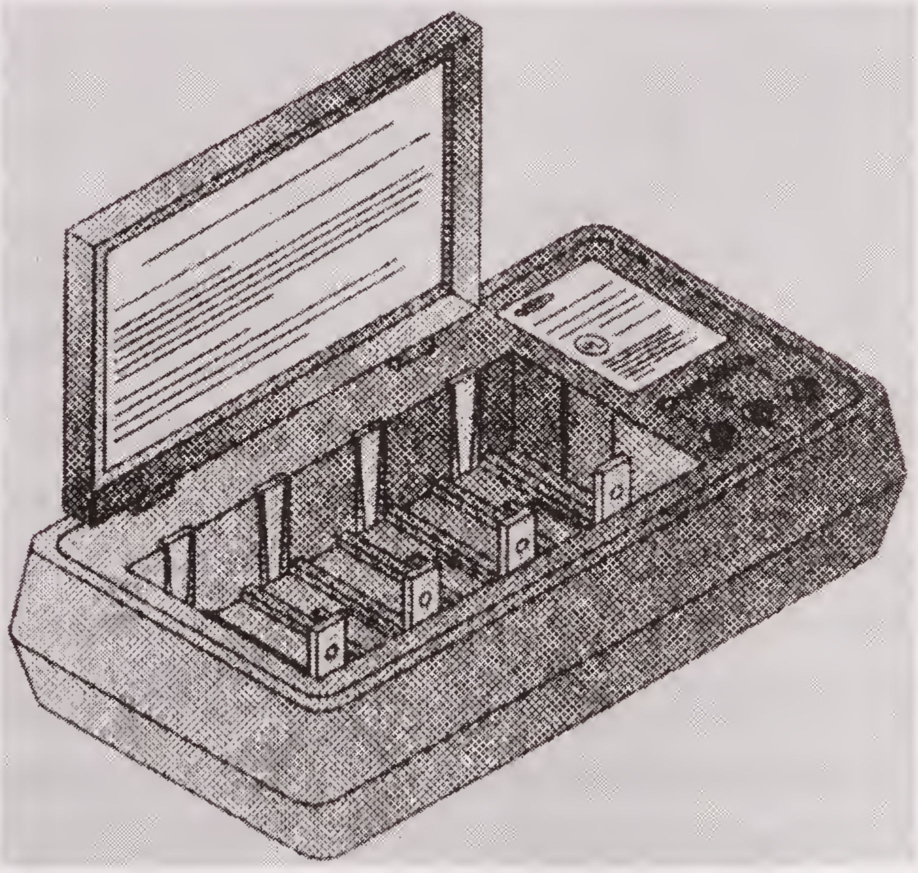

Figure 227

The second type is used to charge car and motorcycle batteries. These batteries need much more current usually over a larger time interval. Car battery chargers are also power supplies with resources to keep the current constant or to limit it to secure values.

Instead of the 100 to 500mA of current used to charge smaller batteries, car and motorcycle batteries need currents much higher. When using the battery charger, car batteries must be unplugged (at least one terminal) from the car's electric circuit. The voltage can rise to values not supported by the electronic circuits of the vehicle. The polarity of the cables must also be observed when using this kind of equipment. It is important to keep the battery in a ventilated place; the nonsealed types can leach poisonous gases during the recharging process.

Figure 228 shows a simple circuit of a battery charger for cars.

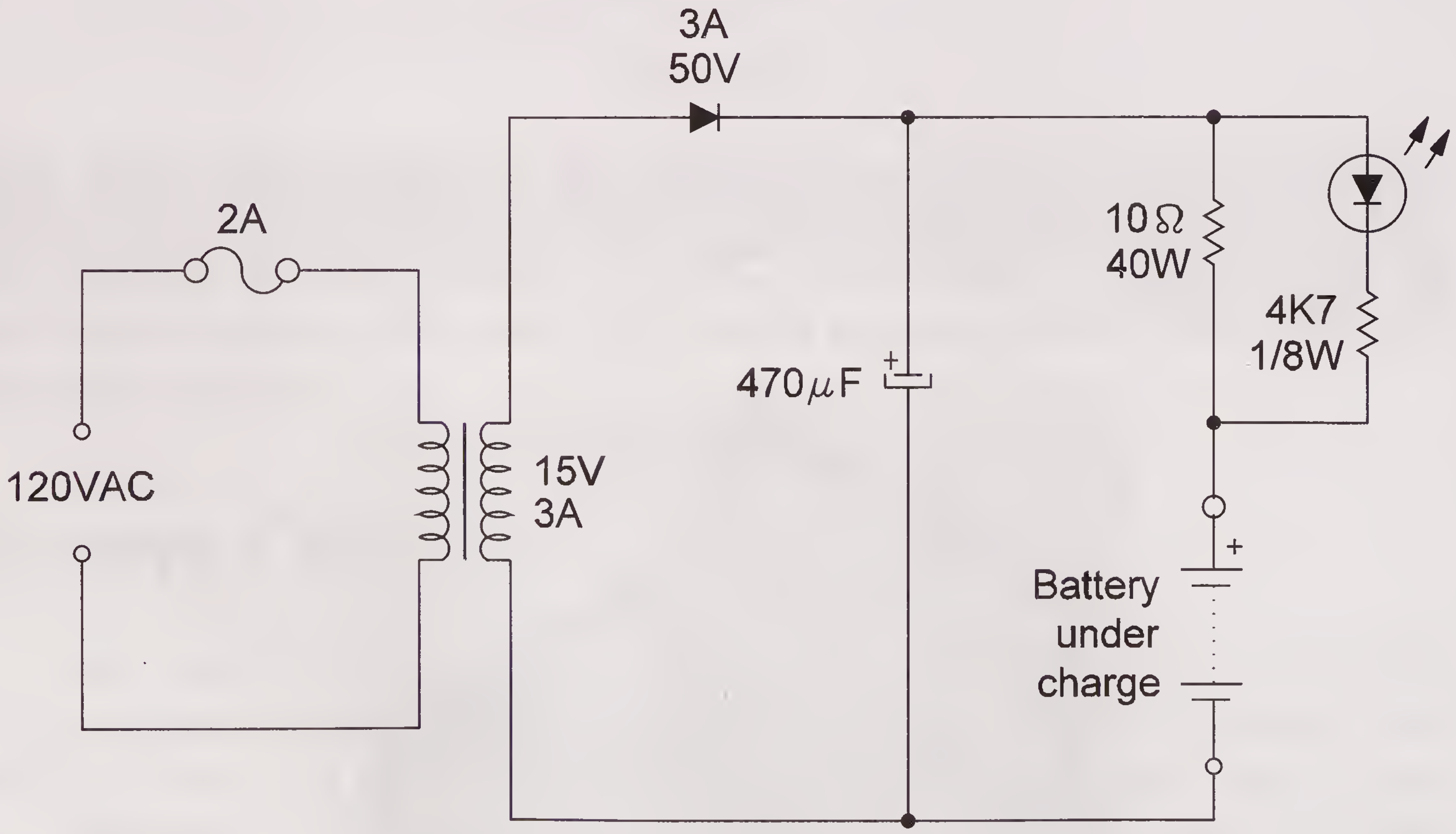

Figure 228

This circuit is a common power supply where the transformer steps down the AC power-line voltage to a value above the battery voltage. After rectification, the DC voltage is applied to the battery in the inverse direction of its normal operation. The resistor (R) limits the current to a secure value for the charge, and the LED monitors the charging process.

The amount of time to charge a common battery can vary from two to three hours to start an engine, and more than 12 hours for a complete charging process. If the battery doesn't retain the charge, it must be replaced. Sophisticated circuits can include resources to detect when the battery is completely charged and automatically interrupt the charging process.

Fluorescent Lamps

In traditional installations, fluorescent lamps use a starter and a ballast in a configuration, such as shown in Figure 229.

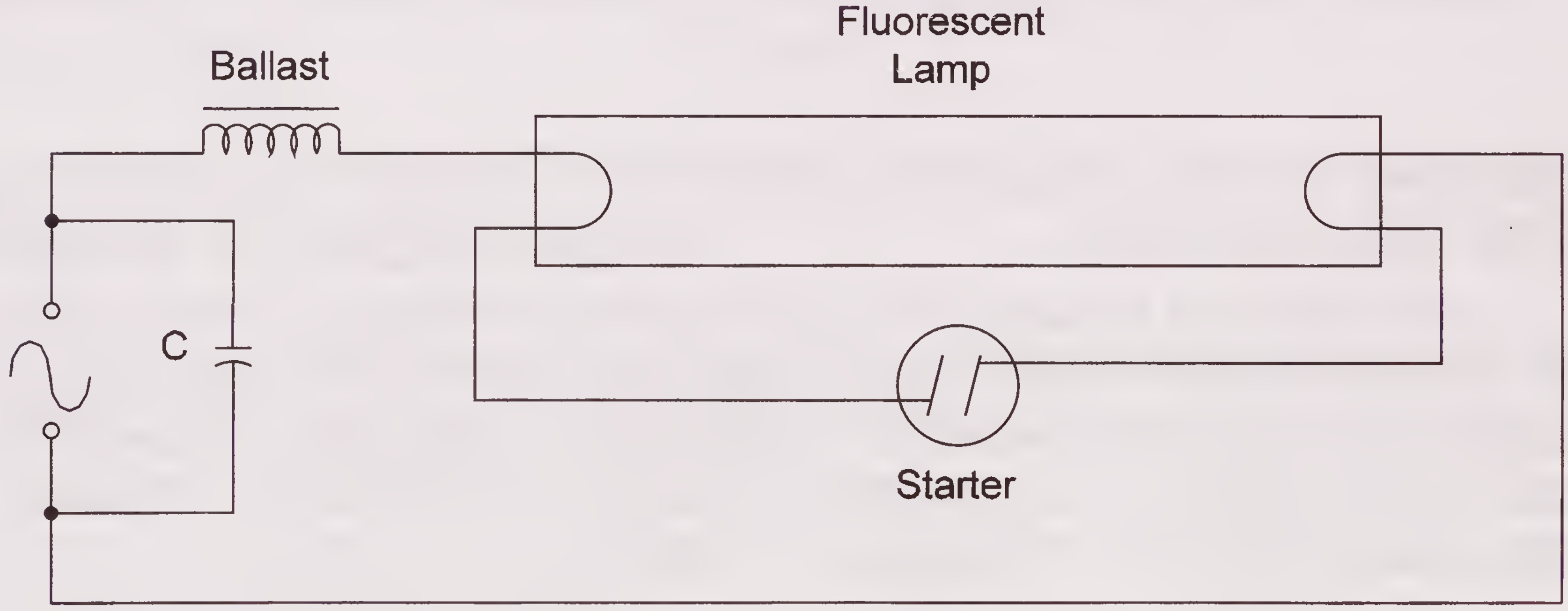

Figure 229

The starter acts as an automatic switch closing and opening the inductive circuit to produce high-voltage spikes. These spikes are necessary to ionize the gas inside the fluorescent lamp. As long as the gas is ionized, the lamp becomes conductive, and light is emitted. When the lamp is conductive, after the initial process, the starter is now inactive.

With electronics, the traditional ballast is replaced by high-voltage and high-frequency circuits that use modern components like power FETs and ICs (Figure 231). This circuit is really a high-frequency inverter operating directly from the AC power line. The AC is rectified and used to drive a

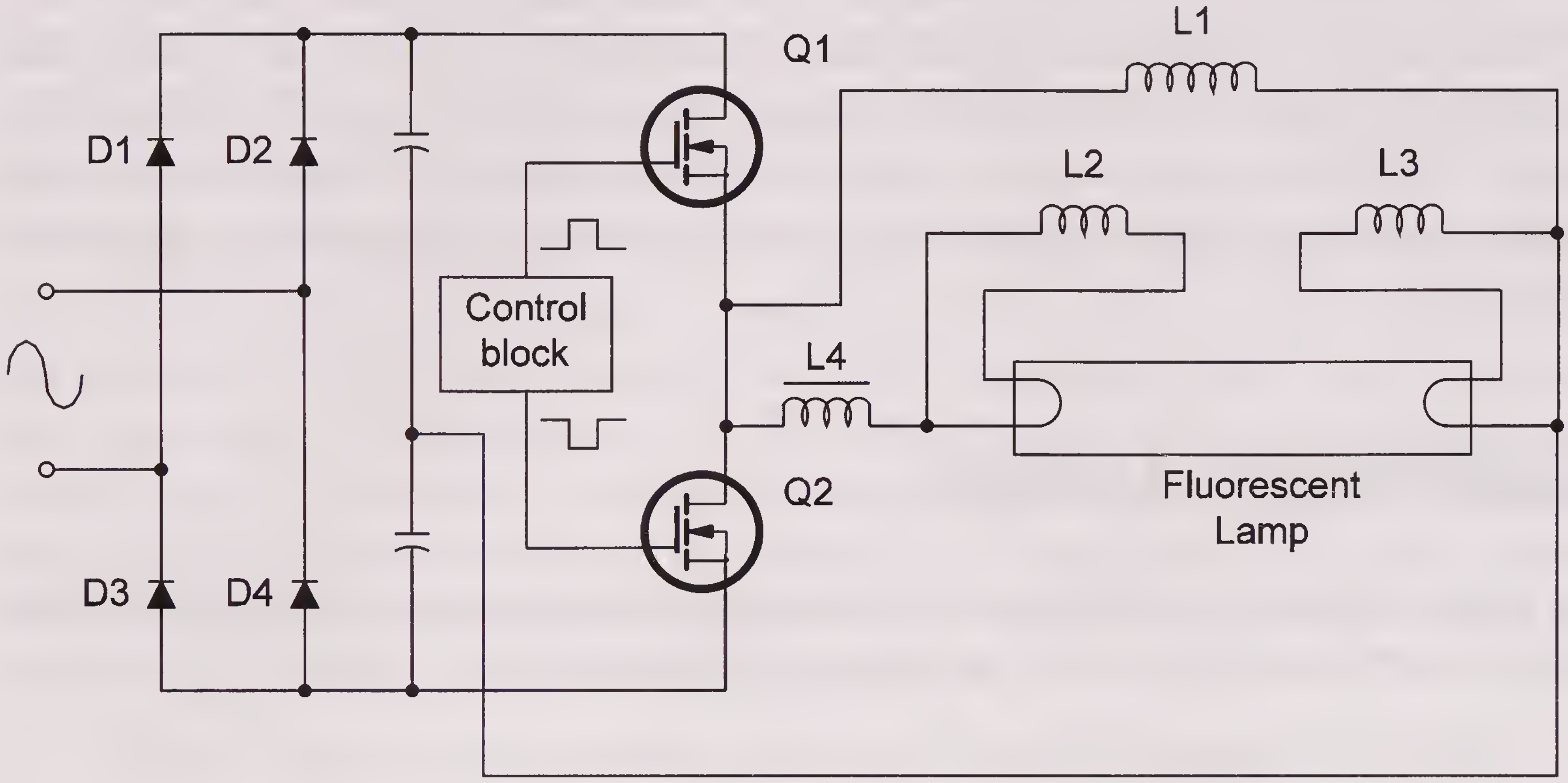

Figure 231

high-power oscillator stage with a transformer. The load of the transformer is the fluorescent lamp.

The difference between this circuit and the traditional configuration is the efficiency (more electric energy can be converted into light), and the oscillations in the AC power-line voltage don't influence or interfere in the circuit operation. But, the basic difference, for the electrician to observe, is the presence of an electronic circuit inside it.

When working with electronic ballast, watch out for EMI. The high-frequency operation of some circuits requires special care to avoid interference with nearby radio equipment. The basic procedures include the use of filters and grounding.

AC/AC Converters

The difference in AC voltages in many countries can be a problem when you have an appliance not able to operate on more than one specified voltage. If you purchased an appliance in another country, or if you are moving to another country, you must consider that your appliances may not be able to plug into an outlet.

The use of a step-up or step-down converter that can convert the 120VAC power-line voltage to 220 or 240VAC or a 220/240VAC to 120VAC is recommended. If you plug a 220/240VAC appliance into a 120VAC outlet, it won't work; however, if you plug a 120VAC appliance into a 220/240VAC outlet, the appliance may burn, have parts damaged , or have a fuse blown.

When recommending a step-up or step-down transformer, electricians must observe the power needs of the appliance. The power is measured in watts (W), meaning you have to use a transformer with more power than required by the load, with a certain tolerance margin to avoid overload. Many step-down transformers used to convert 220/240VAC into120VAC can also be used to convert 120VAC into 220/240VAC when reverse wired to the appliance. This is not a general rule because it is valid only for simple transformers without voltage regulation circuits included inside.

Power Conditioners

In theory, the AC power-line voltage must be kept constant with the 120VAC or another value specified by the energy company or according to the country. Unfortunately, local conditions, as consumption peak hours occur or long wires are used to power an appliance, the voltage can step down to unacceptable values.

Many modern appliances have circuits that include voltage regulators that let them operate in a large range of voltages without problems. But, if this type of resource is not found in the appliance, an external voltage regulator transformer or power conditioner is necessary. An external voltage regulator is an electric/electronic device connected between the appliance and the AC power-line outlet with the purpose of maintaining constant voltage. The most common types are automatic, using electronic circuits that sense the output voltage and correct it by acting on a transformer. When using this kind of device, it is important to check that the specified power is enough to provide the load with the necessary energy.

Intercoms

Many types of intercoms can be found in the electric installations of homes or buildings. The main type is the one associated with the doorbell and used to talk with a person at the front door. It is formed by a low-power audio amplifier where the loudspeaker is used to both reproduce the sounds and to act as a microphone. Figure 232 shows a simple circuit of an intercom using an IC from National Semiconductor.

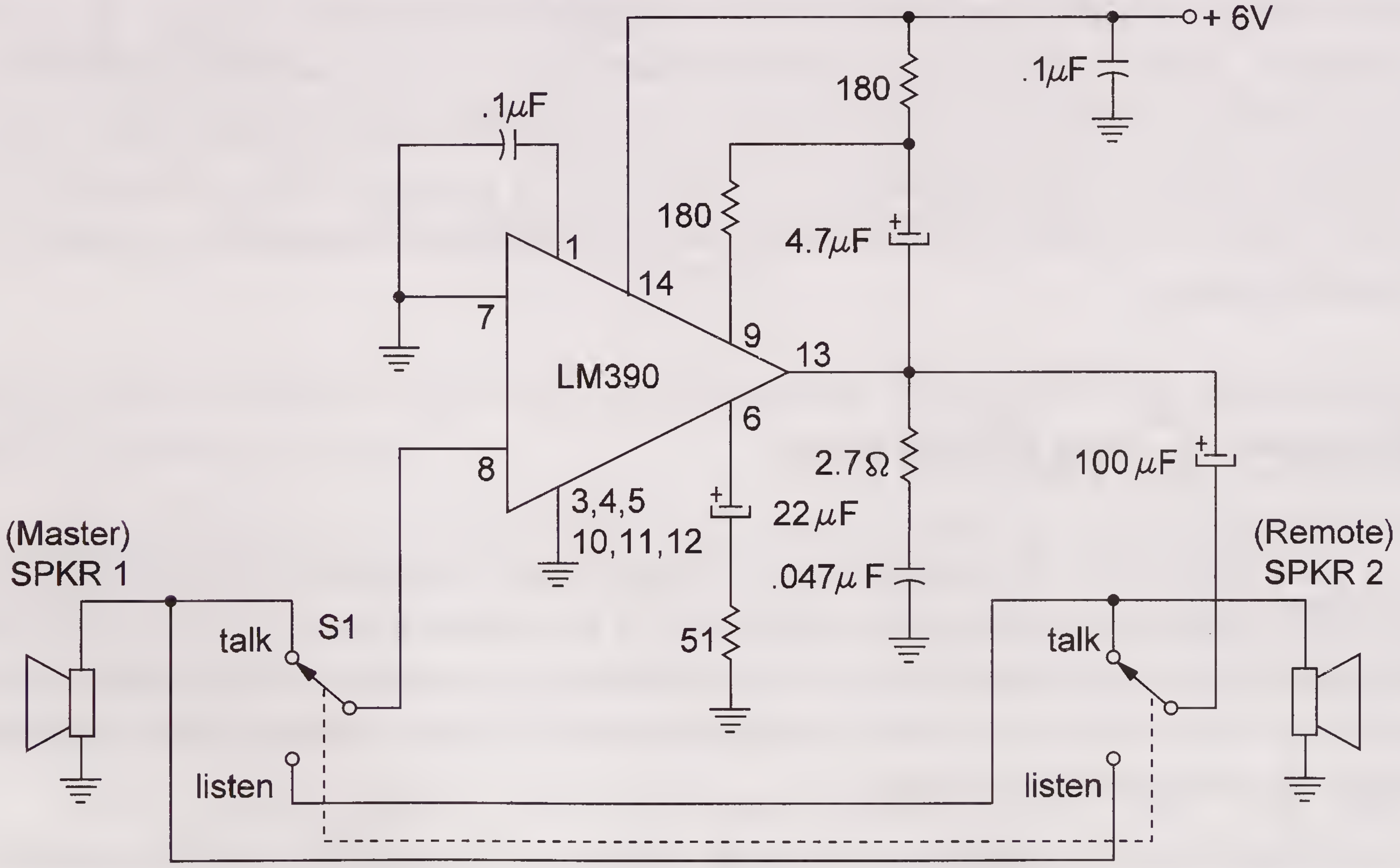
+ 6V
.1μF
180
.1μF
180
4.7μF
1
14
7
9
13
LM390
6
8
2.7Ω
100 μF
3,4,5
10,11,12
22 μF
(Master)
SPKR 1
.047μ F
(Remote)
SPKR 2
talk
S1
51
talk
listen
listen

Figure 232

The switch changes the function of the loudspeakers, plugging them into the input when they are used as microphones (to talk) and as loudspeakers (to hear). Other types can use separated elements for these functions, and electret microphones are used.

The same amplifier can also be used to reproduce the sound of the doorbell. In the basic installation, the central unit is placed inside the building and powered from the AC power line, cells, or batteries. The remote unit is placed at the front door or entrance of the building and two types of wires can be used. In some cases, separated wires for the bell and audio

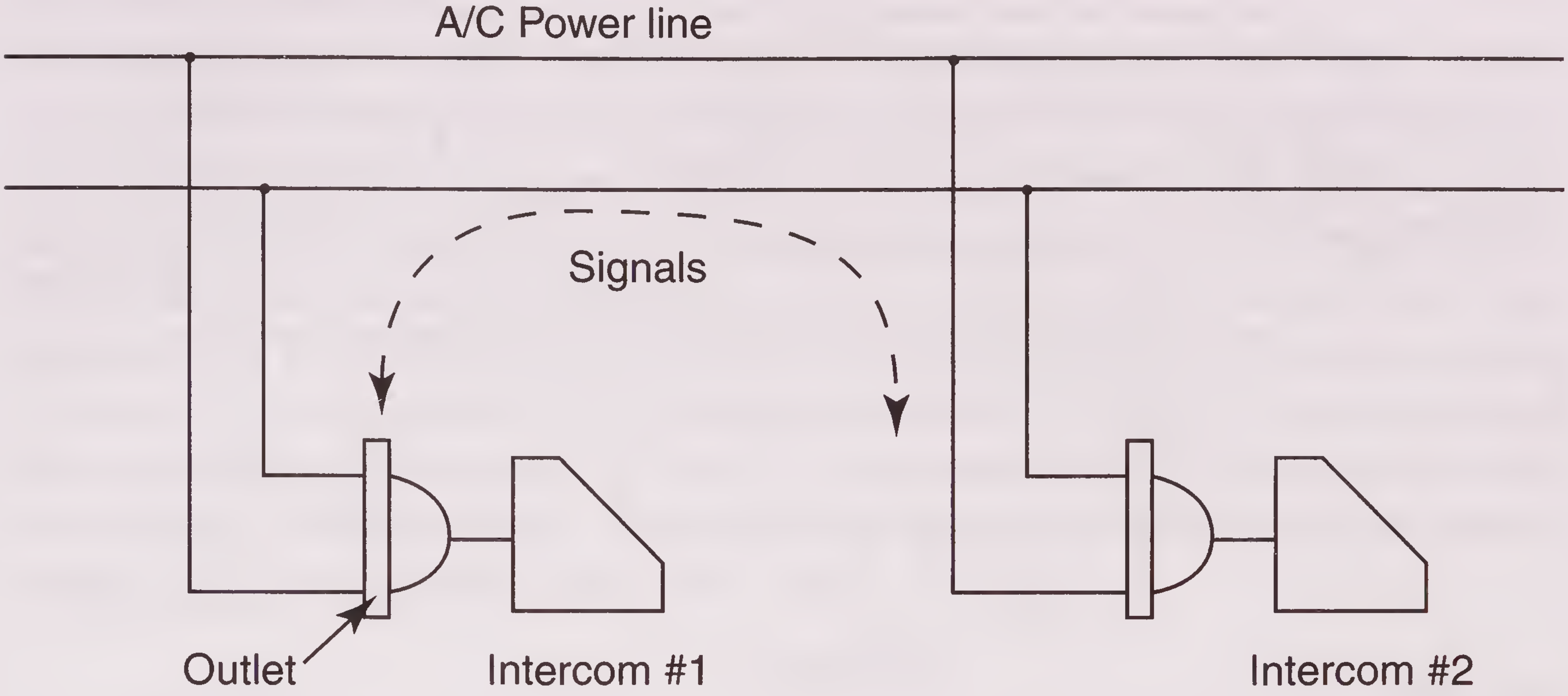

Figure 233

signals are used. The audio signals, in some cases, are conducted by screened cables.

Another type of intercom is one that uses the AC power line to transport the signal between the stations. This means that you only need to plug the units into outlets (Figure 233).

High-frequency signals are superposed to the voltage transporting the audio signals. Common types use PLLs to decode the signals in the range between 50 and 200kHz (refer to the section on PLLs). The principal problem found when working with this kind of intercom is that in some cases the units are plugged into outlets of different branches in an electric installation, as shown in Figure 234.

Even in adjacent rooms, the signals must travel a large distance. Sometimes they are unable to pass from one branch to another. The presence of inductive elements between the branches of the installations can stop the signals. Installing a 0.1 uF/400V paper or polyester capacitor between the branches where the signals can't pass will solve the problem.

Many versions of commercial intercoms can be purchased in many specialized stores and easily installed by an electrician, even one just beginning.

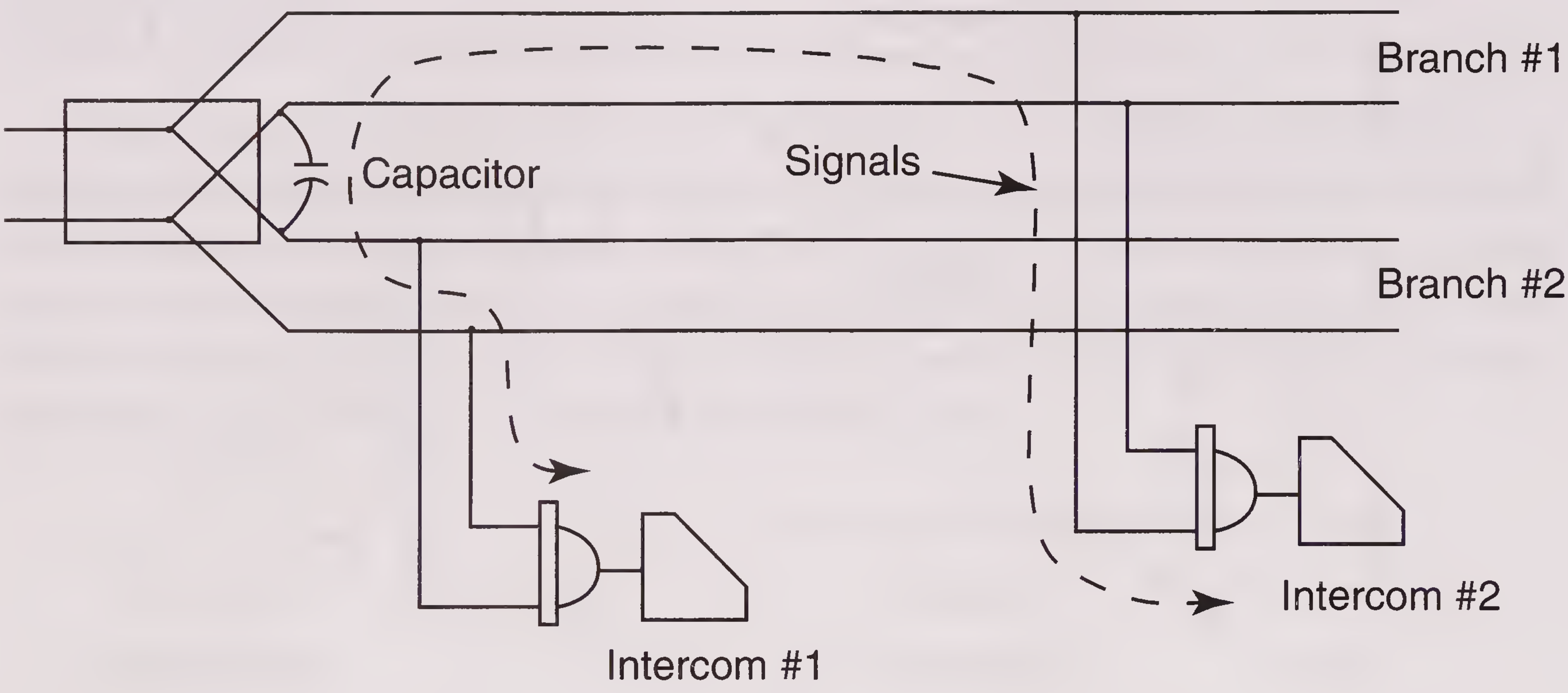

Figure 234

Wireless Systems

The use of the AC power-line voltage is the simplest solution for remote control of appliances and also the most traditional. By placing a switch at any point in the wiring, you can turn on and off any appliance plugged into that line. Today, with the use of complex devices—many of them as small as a matchbox—it is possible to control appliances without wires or the AC power line.

Many solutions have been found to make the control of electric and electronic appliances using electronic circuits easier. The electrician who installs appliances in any building that are AC powered should know how they work. This is especially important to avoid problems with correct installation, interference, or improper use. For example, a remote control for a garage door that uses radio waves in a location with a high level of interference doesn't work well. It can be replaced by an IR remote control with great success.

An electrician can find the following types of remote control for domestic appliances.

- Radio Frequency

This kind of remote control uses a high-frequency transmitter operating in the VHF or UHF band (between 30 and 600MHz) to send signals to a receiver placed at distances of up to 50 meters away, depending on the application. The signals are modulated, (i.e., transport information) communicating to the controlled device what it has to do. The simplest systems use a transistor as an oscillator to produce the high-frequency signals and an IC to produce the coded information to be sent. In some cases, the transistor is replaced by hybrid modules or a complete transmitter mounted in small modules. Figure 235 shows a typical transmitter used to open and close garage automatic doors.

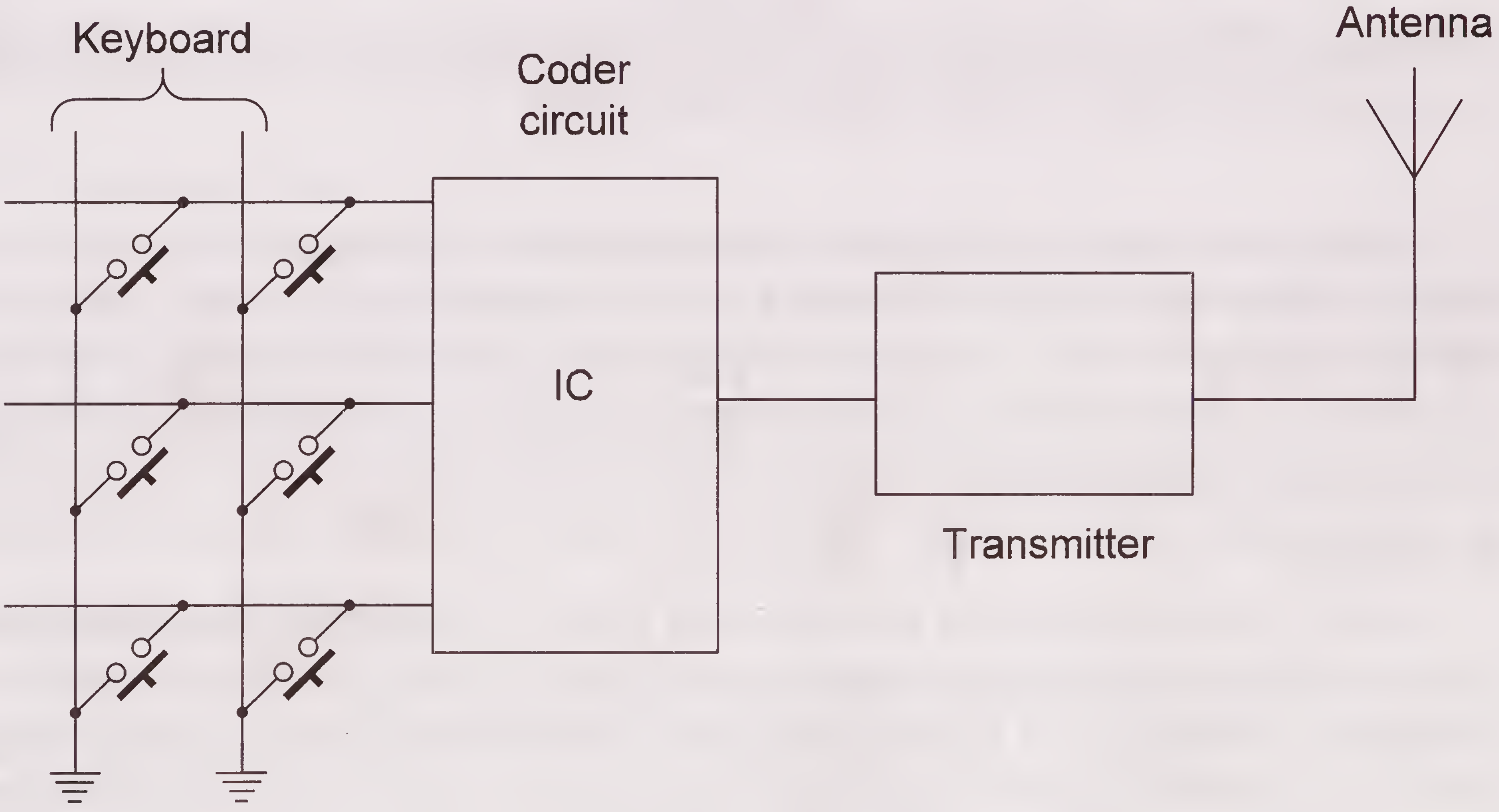

Figure 235

The IC can send a coded signal in a manner so that the receiver only recognizes the signals sent by that IC. The receiver decodes the signals, acting on relays, motors, and other devices to be activated. Normally, in the receiver we find the other member of the pair of ICs used to aid in decoding the signals. Small modules, or hybrid modules, forming pairs of transmitters and receivers are found in many remote control systems, such as the ones shown in the Figure 236.

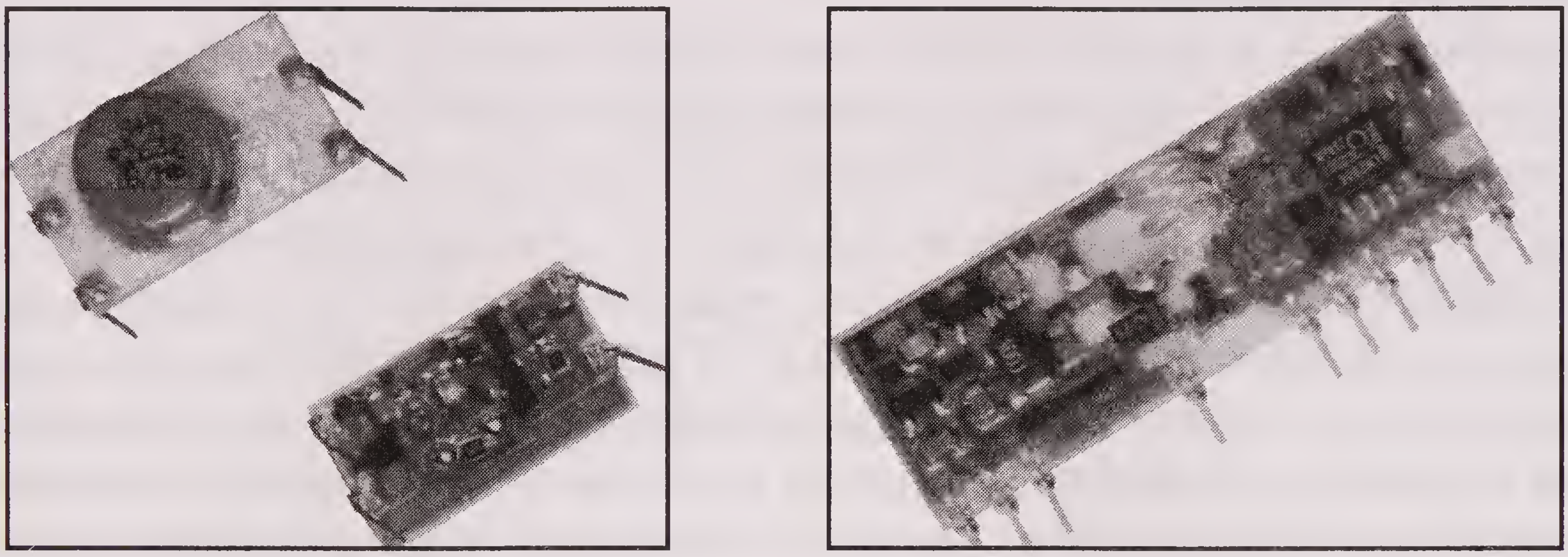

Figure 236
A transmitter hybrid module (left) and a receiver module (right) used in short-range communications and control. (Photos courtesy Tellecontrolli)

The set shown in Figure 236 operate in standard frequencies (315,418MHz and 43,392MHz) and are mounted using the thick film hybrid technology in the manufacturing process. One disadvantage in the use of this kind of remote control system is that it can fail with exposure to a high level of EMI or noise.

■ Infrared

In the remote control is a transmitter that sends IR modulated beams. The modulation can be produced by the same coded ICs found in radio-frequency systems. The information sent by the transmitter communicates what the controlled device has to do. To produce the IR beam, IR LEDs are used. To drive the LED, hybrid modules can be found.

In the receiver, a phototransistor or photodiode exists acting as a sensor for the IR beam. The signal produced by this receiver is applied to a decoder circuit (in some cases the same as found in radio systems) that decodes the information to the driver circuits. This sort of system is found in many electronic appliances, such as VCRs, TVs, and audio systems. It is also found in electric appliances, such as automatic garage doors, remote-controlled fans, and automatic doors.

Although the IR is not disrupted by interference by noise sources like RF or radio systems, its principal disadvantage is that the receiver must "see" the transmitter, meaning the units must have a direct line of communication open. Any solid obstacle can block the IR beam.

The range depends on local conditions, but normally doesn't exceed 50 meters.

- Ultrasound

Ultrasound vibrations can activate some appliances. This system is no longer in use, but it can be found in old appliances including TVs. The transmitter contained a small ultrasound transducer that sent vibrations in frequencies near 42kHz. These vibrations were picked up by a transducer tuned to the same frequency. The signal was also modulated by a circuit that sent information to the receiver, relaying what the appliance was to do. Due to its disadvantages, when compared with other modern systems, such as the IR and radio frequency systems, it is no longer used.

- AC power line

The AC power line can be used to send signals from a transmitter to a remote receiver connected to an appliance. The signals travel through building wiring, which transports information about what the appliance has to do. The operation principal is the same as found in wireless intercoms that use AC power to transport their signals. Although some appliances use this system, its performance is limited to short-range applications.

- Low-current wiring

Sometimes heavy-duty appliances, such as air conditioners, heaters, and high-power motors must be turned on and off remotely. If the control is part of the power cord, large-gauge wires will be needed because of the large current needs. Even the switches must be heavy-duty types, increasing the size of the control unit and adding a danger of shocks and shorts. A solution to this problem is the use of units having relays activated or controlled by low-current circuits like the one shown in Figure 237.

The current used to trigger the relay on and off is very low (milliamperes), allowing the use of low-gauge wires. And, as the control circuit is isolated from the AC power line and the voltage is very low, shock hazards are avoided.

On the Internet, an electrician can find many types of remote controls that can be adapted to common electric and electronic appliances. When

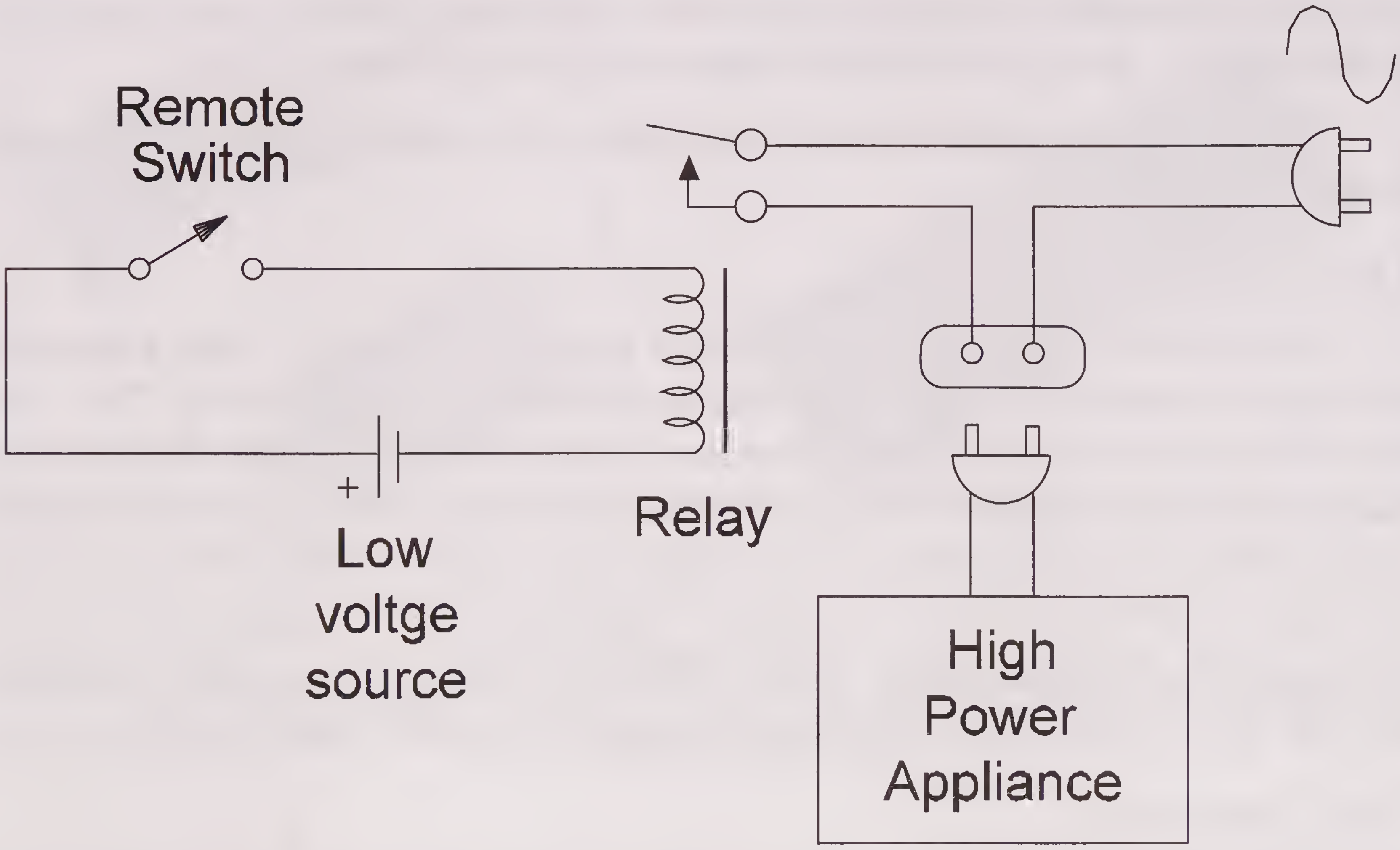

Figure 237

choosing one for an application, keep in mind if the area is noisy, with many interference sources. Avoid radio systems and use IR, or if there are obstacles choose a radio system over IR.

High-Frequency Appliances

Many electric and electronic appliances have high-frequency stages in their circuits. When installing or working with those appliances, it is important to know that radio frequency signals are critical and can be a source of problems for the less experienced electrician.

The first case to be considered is the installation of external antennas for TVs in the traditional VHF or UHF bands. Although the AC power-line voltage can be sent to every part using common wires, this is not a rule for high-frequency signals. High-frequency signals are attenuated by large lengths of wire and can be sensitive to bends or the close proximity of metallic objects.

A single curve in a wire can be sensed as a coil or inductance for a high-frequency signal. Two wires, one passing near the other, can represent a capacitor for high-frequency signals, causing alterations in its strength as shown in Figure 238.

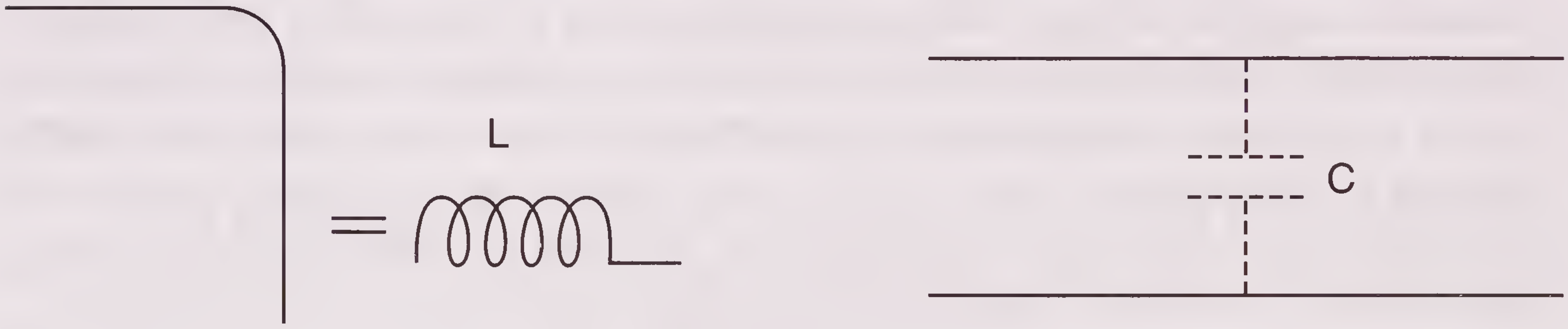

Figure 238

Another important point to consider is the fact that the wiring of a building acts as an antenna, radiating radio signals in the VLF (very low frequency band). These signals can be picked up by any nonshielded wires and, if amplified by audio circuits, can be reproduced as a hum in loudspeakers or phones. This problem can be noticed in audio amplifiers when you touch the input jack as shown in Figure 239.

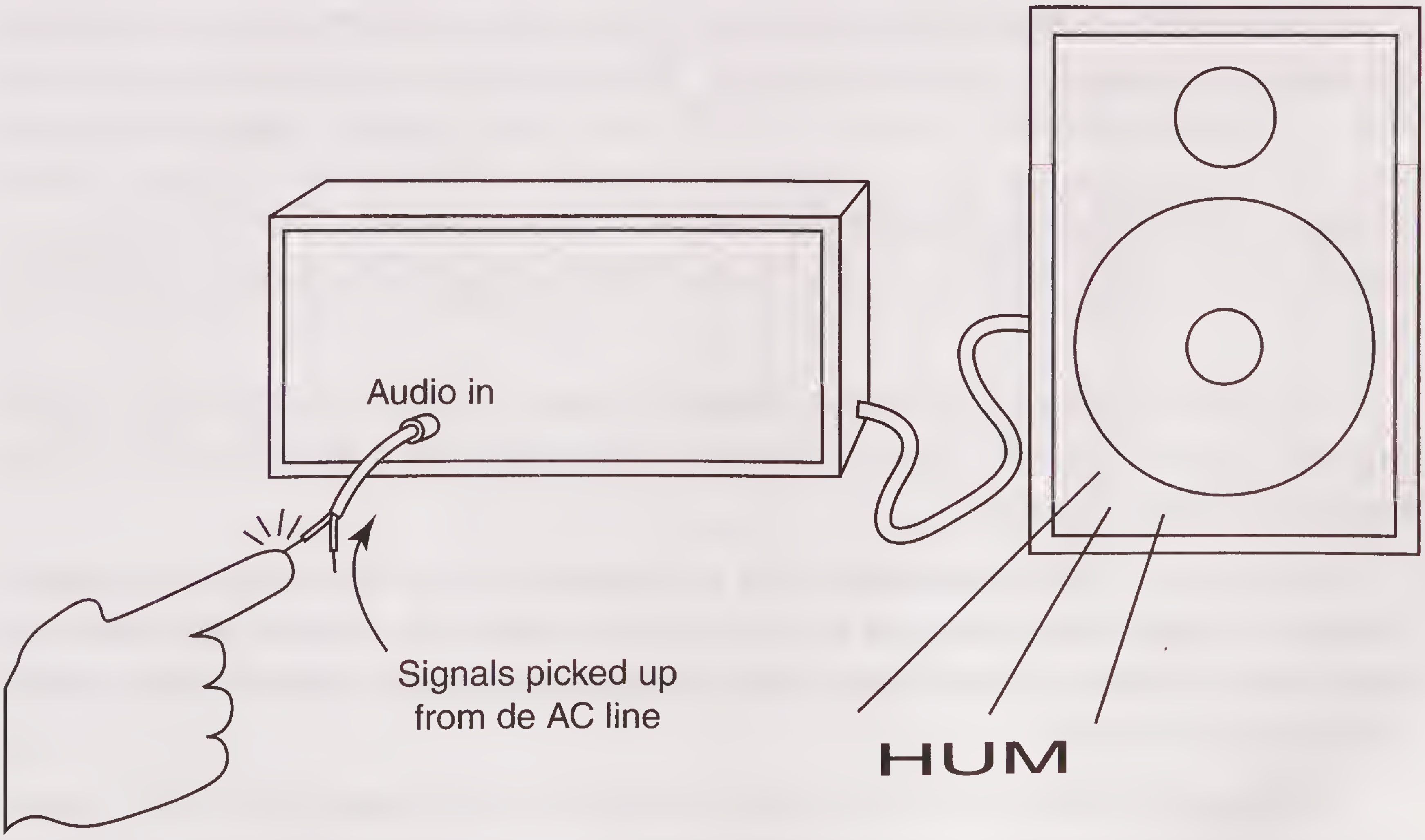
Audio in
Signals picked up from de AC line
HUM

Figure 239

In audio systems and intercoms, the hum can become a serious problem for an electrician trying to install this kind of equipment. Shielded cables only work if the screen is wired to a good ground. So, it is important to plan an installation for these kinds of equipment.

In video systems, an incorrect connection for the cables can result in noise affecting the image. Wavy images can be produced by noise picked up by cables with shield problems. Another problem are high-frequency signals produced by appliances connected to the power supply line being radiated by the wiring. Triacs, SCRs, and motors as explained previously (see EMI, triac, and SCR sections) are sources of these high-frequency signals that can cause interference in other high-frequency equipment.

Surge Protectors

Brief pulses of voltages that can exceed the normal 120 or 220/240V (depending on the voltage used in the AC power line) are called surges. These pulses can rise, in some cases, to volts as high as 1000 and can destroy equipment. Sometimes these pulses are produced by nearby lightning strikes.

To protect equipment against surges and voltage spikes, outlets equipped with surge protection exist. The protection essentially consists of a voltage-dependent resistor (VDR) placed in parallel with the line or two VDRs connected to a ground connection. When the voltage spike comes, the VDR becomes conductive for a brief instant. This is enough to absorb the energy carried by the spike that can potentially destroy equipment.

Surge protectors can be found inside small outlets that can be plugged directly to the AC power line or inside outlet strips. Not all outlet strips are surge protected, though.

Not all electric and electronic equipment is sensitive to surge and voltage spikes. Incandescent lamps, appliances with motors, and heaters are not sensitive as they have inertia enough to not respond to the rapid changes of voltage.

Otherwise, electronic equipment, mainly those that use CMOS components, is very sensitive to voltage spikes coming through the AC power line. This includes computers, electronic telephone answering machines, VCRs, TVs, and audio equipment. Telephone equipment and computers connected to the Internet by telephone lines are also sensitive to voltage spikes coming through the telephone wiring.

Although they have internal surge protection circuits, in some cases they are not fast enough to handle the spikes, or the spikes are strong enough to not be absorbed by them. So, it is common to find modems and other such equipment destroyed after a severe storm.

It is a good idea to connect all this type of equipment to the AC power line using protected outlets or to unplug them from the telephone line or AC power line during storms with severe lighting.

Conclusion

This book is only a brief introduction to electronics for the electrician. There are so many more things to learn and know about electronics as it relates to electric installation. The buildings are becoming even more sophisticated with electronic equipment being embedded in the electric installations. The electrician of the future will have to know not only about the traditional AC power-line voltages, but should have a broad base of knowledge in many complicated areas including digital electronics, microcontrollers, and DSPs.

The idea of wiring a building to connect with the Internet, to send information to a remote computer, or to have climate conditioning systems plugged into sophisticated computers that can sense humidity and temperature, are not old. The electrician of the future, the construction engineer, and even the architect without some knowledge of electronics will be in an unfortunate situation. Even if you are not directly related to a profession that works with electricity, electronics are present in almost all things we make. To know something about it is important to our comfort, our performance in any job, and even more, to our security. Don't stop here! Learn more about electronics. It is important for you, your job, and your future.

Index

A

AC 22
AC circuit 24
AC generator 24
AC power 230
AC/DC adapter 100, 230, 232
AC/DC converter 231
air core 78, 79
alkaline cell 48
alkaline-manganese cell 48
alternating current 22
AM 169
American Wire Gauge 42
ampere 12, 14, 45
amplification factor 112
amplifier, audio
    118, 194, 195, 196, 249, 256
amplifier, operational 229
analog multimeter 221, 223, 224
analog signal 187
anode gate 173
antimatter bodies 6
arsenium 104
atom 3
atomic particles 3
audio amplifier
    118, 194, 195, 196, 249, 256
audio frequency inductor 76
audio oscillator 139
AWG 42

B

ballast 253
bandpass filter 216
barium titanate 92
base 112
base-emitter junction 112
battery 47, 230
battery charger 250
battery test 226
bayonet base 59
Bipolar transistor 111
bipolar transistor
    115, 128, 140, 154, 229

boron 96
Brown and Sharpe (B&S) Wire Gauge 42

C

cadmium sulfide 63
cadmium sulfide photocell 244
capacitance 70, 71, 110
capacitor 37, 69, 100, 160, 167, 184, 229, 230, 241
capacitor, ceramic 69, 73
capacitor, Electrolytic 72
capacitor, electrolytic 18, 74, 229
capacitor, mica 69
capacitor, plastic film 70
capacitor, polyester 257
capacitor, polyester film 73
capacitor, Polystyrene film 73
capacitor, tantalum 72
capacitor, variable 74, 76
carbon electrode 47
carbon-zinc cell 48
cathode 106
CdS cell 63, 65
cell 47
cell, alkaline 48
ceramic capacitor 69, 73
Ceramic transducer 93
choke 76
circuits, electronic 1
closed circuit 9
closed loop 9
CMOS 208, 264
coil 18, 42, 76, 84, 88, 227
collector 112, 124
Collector Current 130
Collector-Emitter Voltage 130
Common Mode Rejection Ratio 193
compression wave 25
conductor 95
continuity test 226
cool solder 34
cosmic ray 29
coupling, Darlington 119
coupling, direct 119
coupling, LC 119
coupling, RC 119
coupling, transformer 119
Critical Rate of Rise 158
crystalline plate 93
current 14

D

Darlington coupling 119
Darlington transistor 128, 132
databook 113
DC 21
DC circuit 24
DC generator 22
DC motor 87
DC power 230
DC voltage 100
decibel 226
decompression wave 25
devices, electronic 2
diac 171, 172
dielectric 69
dielectric material 69
digital multimeter 223, 224
Digital Signal Processor 215

diode 95, 97, 100, 184, 230, 241
diode bridge 161
diode, direct-biased 227
diode test 226
diode, zener 101, 103, 173, 175
direct coupling 119
direct current 21
direct current generator 21
Dissipation Power 130
DPDT 46, 82
DSP 216
dynamo 9

E

earphone 42
electric circuit 8, 9, 14
electric current 5, 7
electric field 26
electric resistance 13
electric tension 12, 14
electric voltage 12
electricity 1
electrode 47, 69
electrolysis 18
electrolyte 47
electrolytic capacitor 18, 72, 74, 229
electromagnet 18, 81, 249
electromagnetic field 26
electromagnetic interference 167
electromagnetic switch 81
electromagnetic wave 17, 24, 26
electromotive force 12
electron 4, 23
electron flow 5, 12
electronic circuit 1, 24
electronic devices 2
electronics 1
Electrostatic Discharge 220
emf 12
EMI 167, 233, 244, 254, 264
emitter 112, 124
enameled wire 42
ENIAC 204
ESD 220

F

farad 70
ferrite 76
ferrite core 79
ferrite toroid 78
FET 140, 148, 229
Field-Effect Transistor 140
filament 58, 60
filament, tungsten 58
filter 230
filter choke 76
flange 59
fluorescent 17
fluorescent lamp 253
FM 169
frequency 26
fuse 43, 227
fusible link 44

G

GaAs 104
gain 130, 193
gallium 95, 104
Gallium Arsenide 104

gamma ray 28
gas, inert 83
gate 140, 145, 152, 153
Gate Trigger Current 165
gate voltage 140
gate-cathode junction 162
gauge 42
generator 8, 9
germanium 95, 97
germanium device 97
germanium transistor 120

H

heat sink 16, 114, 234
henry 76
hertz 24

I

IC 184, 185, 229
IGBT 154
impedance 80, 90, 91
impedance match 80
incandescent lamp 58
indium 96, 104
inductance coil 78
inductor 76
inert gas 83
infrared 17, 29, 107, 123, 247, 260
insulator 95
integrated circuit 179, 184, 187, 196, 216
inverter 234, 235, 245
ionization 17
iron core 78, 79
Isolated-Gate Bipolar Transistor 154

J

JFET 141
joule 15
Joule's Effect 15, 16
Joule's Law 15
Junction FET 141
junction test 120

L

lamp, incandescent 58
lamp, neon 60
laser 247
LC coupling 119
LCD 175, 178, 227
LCD driver 177
LDR 63, 64, 124, 244
LED 17, 60, 104, 105, 108, 123, 241, 252
Light Emitting Diodes 17, 104
light-dependent resistor 63
light-sensitive resistor 63
liquid crystal display 175
loudspeaker 42

M

magnetic field 18, 26, 86
magnetic sensor 42
magnetic transducer 91
match transformer 94
Maxwell, James Clerk 26

megahertz 24
Metal Oxide Varistor 67
Metal-Oxide Semiconductor Field-Effect Transistor 145
mica capacitor 69
microcontroller 213, 214
microfarad 70
microhenry 76
microprocessor 127, 213, 214
miliampere 45
millihenry 76
monochromatic light source 104
MOSFET 145, 146
Motion detector 248
motor 18, 42, 86
multimeter 45, 55, 65, 84, 101, 120, 162, 71, 219, 224, 227
multimeter, analog 221, 223, 224
multimeter, digital 223, 224

N

nanofarad 70
negative charge 7
negative pole 9
Negative Temperature Coefficient 65
neon 17
neon gas 60
neon lamp 60
neutron 3
ni-cad (nickel-cadmium) 48
nonferrous wire 42
NPN 111, 115, 128
NPN silicon transistor 120
NPN transistor 117
NTC 65, 67
nucleus 3

O

Off State Voltage 158
ohm 13, 14, 37, 50
ohmic resistance 91
Ohm's Law 13
operational amplifier 190, 193, 229
optocoupler 123, 127
orbital shell 3
oscillating electric charge 28
oscillator 94, 109, 136, 161, 234, 240, 242
oscillator, audio 139
output transistor 237

P

parallel 20
PCB 30, 46
Phase Locked Loop 199
phosphorus 96
photo sensor 247
photocell 244
photodiode 109, 111, 124, 126, 260
Photofact, Sams 220
phototransistor 122, 124, 126, 247, 260
picofarad 70
piezoelectric 248
piezoelectric ceramic 93
piezoelectric components 94
piezoelectric material 92

piezoelectric transducer 92
piezoelectric unit 89
piroelectric sensor 248
plastic film capacitor 70
PLL 199, 257
PN junction 104, 109, 110
PNP 111, 115, 128
PNP silicon transistor 120
PNP transistor 119
polarity 224, 232, 252
polyester capacitor 257
polyester film capacitor 73
Polystyrene film capacitor 73
positive charge 7
positive pole 9
Positive Temperature Coefficient 65
positron 6
potentiometer 56, 57, 160
potentiometer, slide 57
potentiometer, trimmer 56
power conditioner 255
Power FET 148
power MOSFET 148
preamplifier 118
primary winding 78, 80
printed circuit board 30, 46
probe connections 224
probes 224, 226
proton 3, 4
PTC 65, 67
PUT 139

Q

quadrac 172
quartz crystal 92, 93

R

Radio Frequency Interference 167
radio wave 29
RC coupling 119
rectifier 99, 161, 230
Red-Blue-Green 106
reed switch 83
relay 18, 42, 128, 152
repetitive peak reverse voltage 165
repetitive pulse reverse voltage 165
resistance 13, 14, 219
resistor 13, 37, 50, 229
resistor, fixed 50
resistor, Light-dependent 63
resistor, light-sensitive 63
resistor, variable 50, 56, 57, 74
Resistor, Voltage Dependent 67
reverse-biased diode 227
RF 260
RF inductor 76
RF transformer 80
RFI 167
RGB 106

S

Sams Photofact 220
Sams Technical Publishing 220
schematic 220
SCR 154, 157, 175, 216, 229, 233, 244, 264
secondary winding 78, 80

semiconductor 95, 109, 154, 229
sensor 18, 244, 246, 260
sensor, piroelectric 248
sensor, vibration 249
sensors, photo 247
sequencer 241
series 20
series-parallel associations 21
SHF 109
silicon 95
Silicon Bilateral Switch 174
silicon chip 185, 216
Silicon Controlled Rectifier 154
silicon device 97
Silicon Unilateral Switch 172
sine 24
sine curve 29
slide potentiometer 57
SMD 31
solder 32
soldering gun 34
soldering iron 32
solenoid 18, 42, 85, 101, 128, 150
solid-state electronics 96
SPDT 47, 82
SPST switch 44
static 5
static electricity 5, 6
step motor 86, 88, 152
Superhigh Frequency 109
surface-mounted devices 31
surge 264
surge protector 264
switch 45, 81
switch, electromagnetic 81
switch, reed 83
switch-mode power supplies 153, 188

T

tantalum capacitor 72
telephone slide 59
thyristor 137, 154, 161
timbre 26
transducer 18, 88, 91, 245
transducer, Ceramic 93
transducer, magnetic 91
transducer, piezoelectric 92
transformer 78, 80, 230, 241
transformer coupling 119
Transient Surge Absorber 67
Transistor 95
transistor 37, 104, 111, 124, 216, 234
transistor, Bipolar 111
transistor, bipolar effect 140
Transistor Logic IC 207
transistor manual 113
transistor, output 237
transistor test 226
transistor, unijunction 132
Transition Frequency 130
triac 161, 163, 165, 167, 172, 216, 229, 233, 244, 264
trigger voltage 171
trimmer potentiometer 56
TTL 207
tube 181
tungsten filament 58
tunnel diode 109, 111
tweeter 90, 93

U

UHF 109, 259, 262
UJT 132
Ultrahigh Frequency 109
ultrasonic sound 25
ultrasound 249
ultrasound transducer 261
ultraviolet 17, 123
unijunction transistor 132
universal power adapter 232

V

varactor 109
variable capacitance diode 110
variable capacitor 74, 76
variable resistor 50, 56, 57, 74
Varicap 109, 110
Varistor, Metal Oxide 67
Varistor, Zinc-Oxide 67
VDR 67, 69, 264
very low frequency 263
VHF 169, 259, 262
vibration sensor 249
visible light 29
VLF 263
volt 12, 14
Volt-Ohm-Milliamp meter 221
voltage 14, 219
Voltage Dependent Resistor 67
voltage multiplier 241
voltage regulator 255
voltage spike 264
voltage-dependent resistor 264
VOM 55, 221

W

watt 15, 90
waveform 26, 37
Winchester disk 91
wire 42
wire, enameled 42
wire, nonferrous 42
woofer 90
Working Voltage DC 71
WVDC 71

X

X-ray 28
xenon 17, 239

Z

zener diode 101, 103, 173, 175
Zinc-Oxide Varistor 67

PROMPT®
PUBLICATIONS

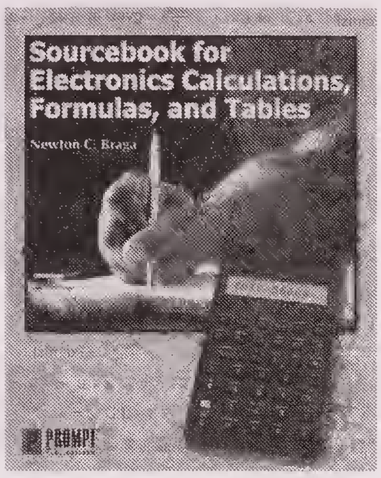
Sourcebook for Electronics Calculations, Formulas, and Tables
Newton C. Braga
PROMPT

Sourcebook for Electronics Calculations, Formulas, and Tables

Newton C. Braga

For the student, teacher, or experienced technician, this is your one-stop guide for formulas and calculations on nearly any electronics subject, including explanations of use, derivations, and practical applcation examples. Topics include cable resistance, electrolysis, basic capacitor forumlas, harmonics, and many more.

Professional Reference
440 pages • paperback • 8-3/8 x 10-7/8"
ISBN: 0-7906-1193-7 • Sams 61193
$34.95

FUN PROJECTS for the EXPERIMENTER
Newton C. Braga
PROMPT

Fun Projects for the Experimenter

Newton C. Braga

Author Newton C. Braga, whose works have appeared in electronics magazines for over 20 years, has collected fifty of his most fun, easy-to-build, and practical projects for your enjoyment. Basic electronic principles and fundamentals are stressed. These projects are primarily stand-alone, low-cost, and with few components, intended for one evening of work. The components needed are listed along with schematics, and hints and questions about the circuits are included to stimulate your imagination regarding possible modifications and alternate use.

Examples of the projects include an LED flasher, mini-metronome, electronic fishing lure, micro FM transmitter, touch switch, wireless beeper, and signal tracer. For hobbyists and students wanting to understand the basic principles of electronics, this book will provide answers to many of your questions.

Projects
328 pages • paperback • 7-3/8 x 9-1/4"
ISBN: 0-7906-1149-X • Sams 61149
$24.95

To order today or locate your nearest Prompt® Publications distributor at 1-800-428-7267 or www.samswebsite.com

Prices subject to change.

PROMPT®
PUBLICATIONS

SAMS Technical Publishing
Servicing Series
Zenith Televisions
by Bob Rose
PROMPT

HWS Servicing Series: Zenith Televisions

by Bob Rose

Expanding on the HWS Servicing Series, author Bob Rose takes an in-depth look at Zenith TVs, with coverage of manufacturer history, test equipment, literature, software, and parts. A variety of chassis are given a thorough analysis.

Troubleshooting & Repair
352 pages • paperback • 8-1/2" x 11"
ISBN 0-7906-1216-X • Sams 61216
$34.95

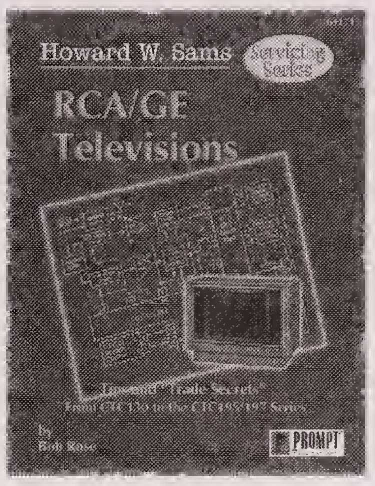

Servicing RCA/GE Televisions

Bob Rose

Let's assume you're a competent technician, with a good work ethic, knowledgeable, having the proper tools to do your job efficiently. You're hungry to learn, to find a way to get an edge on the competition—better yet, to increase your productivity, and hence, your compensation.

In *Servicing RCA/GE Televisions*, author Bob Rose has compiled years of personal experience to share his knowledge about the unique CTC chassis. From the early CTC130 through the CTC195/197 series, Bob reveals the most common faults and quickest ways to find them, as well as some not-so-common problems, quirks and oddities he's experienced along the way. From the RCA component numbering system to the infamous "tuner wrap" problem, Bob gives you all you need to make faster diagnoses and efficient repairs, with fewer call-backs—and that's money in the bank!

Troubleshooting & Repair
352 pages • paperback • 8-3/8" x 10-7/8"
ISBN 0-7906-1171-6 • Sams 61171
$34.95

To order today or locate your nearest Prompt® Publications distributor at 1-800-428-7267 or www.samswebsite.com

Prices subject to change.

PROMPT®
PUBLICATIONS

Exploring Solid-State Amplifiers

Joseph Carr

Exploring Solid-State Amplifiers is a complete and authoritative guide to the world of amplifiers. If you're a professional technician or a hobbyist interested in learning more about amplifiers, this is the book for you.

Beginning with amplifier electronics: overcoming the effects of noise, this book covers many useful and interesting topics. It includes helpful, detailed schematics and diagrams to guide you through the circuitry and construction of solid-state amplifiers, such as: Transistor Amplifiers; Junction Field-Effect Transistors and MOSFET Transistors; Operational Amplifiers; Audio Small Signal and Power Amplifiers; Solid-State Parametric Amplifiers; and Monolithic Microwave Integrated Circuits (MMICs). Two bonus chapters are devoted to troubleshooting circuits and selecting solid-state replacement parts.

Electronics Technology
240 pages • paperback • 7-3/8" x 9-1/4"
ISBN: 0-7906-1192-9 • Sams: 61192
$29.95

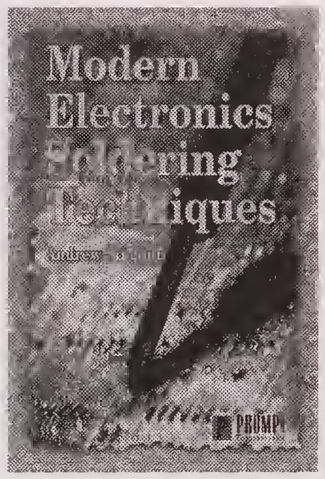

Modern Electronics Soldering Techniques

Andrew Singmin

The traditional notion of soldering no longer applies in the quickly changing world of technology. Having the skills to solder electronics devices helps to advance your career.

Modern Electronics Soldering Techniques is designed as a total learning package, providing an extensive electronics foundation that enhances your electronics capabilities. This book covers how to solder wires and components as well as how to read schematics. Also learn how to apply your newly learned knowledge by following step-by-step instructions to take simple circuits and convert them into prototype breadboard designs. Other tospic covered include troubleshooting, basic math principles used in electronics, simple test meters and instruments, surface-mount technology, safety, and much more!

Electronics Basics
304 pages • paperback • 6" x 9"
ISBN: 0-7906-1199-6 • Sams 61199
$24.95

To order today or locate your nearest Prompt® Publications distributor at 1-800-428-7267 or www.samswebsite.com

Prices subject to change.

PROMPT®
PUBLICATIONS

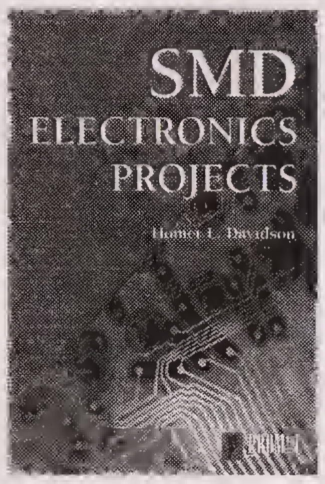

SMD Electronics Projects

Homer Davidson

SMD components have opened up a brand-new area of electronics project construction. These tiny components are now available and listed in many of the electronics mail-order catalogs for the electronics hobbyist. Projects include: earphone radio, shortwave receiver, baby monitor, cable checker, touch alarm and many more. Thirty projects in all!

Projects
320 pages • paperback • 7-3/8” x 9-1/4”
ISBN 0-7906-1211-9 • Sams 61211
$29.95

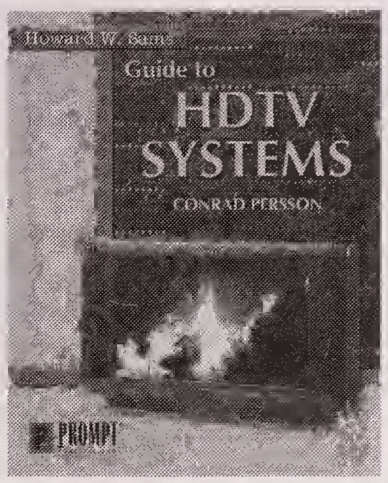

Guide to HDTV Systems

Conrad Persson

As HDTV is developed, refined, and becomes more available to the masses, technicians will be required to service them. Until now, precious little information has been available on the subject. This book provides a detailed background on what HDTV is, the technical standards involved, how HDTV signals are generated and transmitted, and a generalized description of the circuitry an HDTV set consists of. Some of the topics include the ATSC digital TV standard, receiver characteristics, NTSC/HDTV compatibility, scanning methods, test equipment, servicing considerations, and more.

Video Technology
256 pages • paperback • 7-3/8" x 9-1/4"
ISBN 0-7906-1166-X • Sams 61166
$34.95

To order today or locate your nearest Prompt® Publications distributor at 1-800-428-7267 or www.samswebsite.com

Prices subject to change.

PROMPT®
PUBLICATIONS

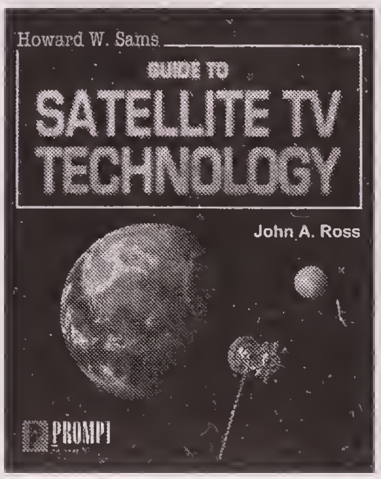

Howard W. Sams Guide to Satellite TV Technology

John A. Ross

This book covers all aspects of satellite television technology in a style that breaks "tech-talk" down into easily understood reading. It is intended to assist consumers with the installation, maintenance, and repair of their satellite systems. It also contains sufficient technical content to appeal to technicians as a reference.

Coverage includes C, Ku, and DBS signals. Chapters include How Satellite Television Technology Works, Parts of a Satellite Television Reception System, Installing the Hardware Portion of Your System, Installing the Electronics of Your System, Installing Your DSS, DBS, or Primestar System, Setting Up a Multi-Receiver Installation, Maintaining the System, Repairing Your System, and more.

Communications
464 pages • paperback • 7-3/8" x 9-1/4"
ISBN 0-7906-1176-7 • Sams 61176
$39.95

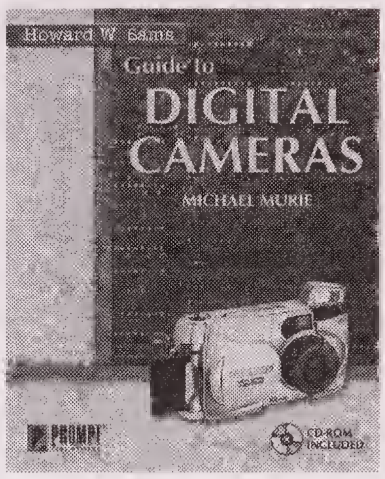

Complete Guide to Digital Cameras

Michael Murie

The *Complete Guide to Digital Cameras* will appeal to anyone who has recently purchased or is considering an investment in a digital camera. The first section introduces the reader to digital cameras, how they work, uses and features, and how to buy one. The second section gives tips on use, available options, and how to transfer images from camera to computer. The third section focuses on manipulating the images on computer in varying file formats, and looks at some color printers presently available. Along with model comparisons and index of currently available cameras, a CD-ROM contains sample images, trial software, and utilities.

The *Complete Guide to Digital Cameras* is the answer to all your questions. Author Michael Murie, avid photographer and multimedia developer, has compiled a comprehensive volume, covering all aspects of the digital camera world. As a bonus, we include a CD-ROM with sample images and software, as well as a comprehensive table of the latest digital cameras, comparing features and technical specifications.

Video Technology
536 pages • paperback • 7-3/8" x 9-1/4"
ISBN 0-7906-1175-9 • Sams 61175
$39.95

To order today or locate your nearest Prompt® Publications distributor at 1-800-428-7267 or www.samswebsite.com

Prices subject to change.

PROMPT®
PUBLICATIONS

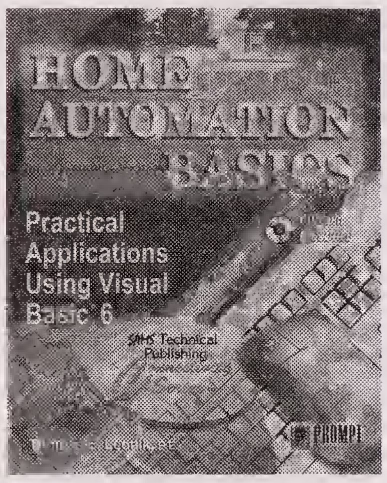

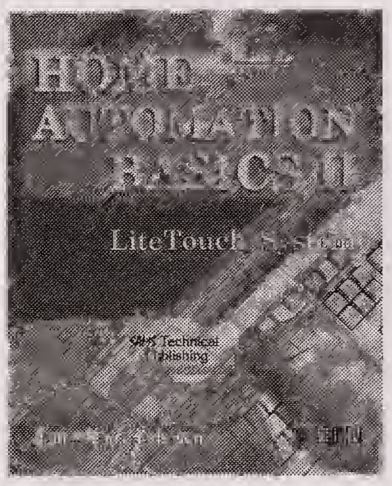

Home Automation Basics
Practical Applications Using Visual Basic 6
by Tom Leonik, P.E.

This book explores the world of Visual Basic 6 programming with respect to real-world interfacing, animation and control on a beginner to intermediate level.

Home Automation Basics demonstrates how to interface to a home automation system via the serial port on your PC. Using a programmable logic controller (PLC) as a home monitor, this book walks you through the process of developing the home monitor program using Visual Basic programming. After programming is complete the PLC will monitor the following digital inputs: front and rear doorbell pushbuttons, front and rear door open sensors, HVAC systems, water pumps, mail box, as well as temperature controls. The lessons learned in this book will be invaluable for future serial and animations projects!

Electronics Technology
386 pages • paperback • 7-3/8" x 9-1/4"
ISBN 0-7906-1214-3 • Sams 61214
$34.95

Home Automation Basics II: LiteTouch System
James Van Laarhoven

Daunted by the thought of installing a world-class home automation system? You shouldn't be! LiteTouch Systems has won numerous awards for its home automation systems, which combine flexibility with unlimited options. James Van Laarhoven takes LiteTouch to the next level with this text from Prompt® Publications. Van Laarhoven helps to make the installation, troubleshooting, and maintenance of LiteTouch 2000® a less daunting task for the installer, presenting information, examples, and situations in an easy-to-read format. LiteTouch 2000® is a true computer-control system that reduces unsightly switch banks and bulky high-voltage control wiring to a minimum.

Van Laarhoven's efforts make LiteTouch 2000® a product to be utilized by programmers of varying backgrounds and experience levels. *Home Automation Basics II* should be a part of every electrical and security professional's reference library.

Electronics Technology
304 pages • paperback • 7-3/8" x 9-1/4"
ISBN 0-7906-1226-7 • Sams 61226
$34.95

To order today or locate your nearest Prompt® Publications distributor at 1-800-428-7267 or www.samswebsite.com

Prices subject to change.

PROMPT®
PUBLICATIONS

Audio Systems Technology, Level I

Handbook for Installers and Engineers

NSCA

This book is a one-stop information source for today's audio technician. It can be used as a study guide to prepare for NICET audio technician certification exams, as well as a comprehensive reference for the installer of audio systems—both out in the field and at the bench.

Designed to correspond with Level I work elements on the NICET tests, this valuable handbook presents the basics of audio installation as it is practiced in the industry today. Topics include: Basic Electronic Circuits; Basic Math; Basics of Microphones & Loudspeakers; Basic Wiring; Switches and Connectors; Codes, Standards and Safety; and Reading Plans and Specifications.

Additionally, information about getting certified by NICET is included, with tips and strategies to help with your success on the NICET exams.

Audio
320 pages • paperback • 7-3/8 x 9-1/4"
ISBN: 0-7906-1162-7 • Sams 61162
$34.95

Audio Systems Technology, Level II

Handbook for Installers and Engineers

NSCA

Designed to correspond with Level II work elements on the NICET tests, this book presents intermediate level content on audio installation as it is practiced in the industry today. Some of the topics: Audio Calculations; Acoustical Measurements; Microphones, Loudspeakers and Mixers; Trigonometry and Geometry; Wiring and Cabling; Effective Business Communication; and Bench Test Equipment.

Information about getting certified by NICET is included, with tips and strategies for the test-taker.

Both Level I and Level II books are for you if:

- You want to be NICET-certified in audio systems.
- You are a developing audio installer or engineer.
- You want to enhance your knowledge of audio systems design and installation.
- You are experienced but want to brush up on intermediate-level audio.

Audio
432 pages • paperback • 7-3/8 x 9-1/4"
ISBN: 0-7906-1163-5 • Sams 61163
$39.95

To order today or locate your nearest Prompt® Publications distributor at 1-800-428-7267 or www.samswebsite.com

Prices subject to change.

PROMPT®
PUBLICATIONS

Audio Systems Technology
LEVEL III
HANDBOOK FOR INSTALLERS AND ENGINEERS
NSCA
PROMPT

Audio Systems Technology, Level III

Handbook for Installers and Engineers

NSCA

Audio Systems Technology Level III is an essential for the library of the advanced technician who has several years of job experience and an associate's degree or the equivalent. While each book in this series contains its own unique information, there is overlap from one level to the next, providing repetition on the most important, fundamental points. This intentional dovetailing also allows the entire series to be used as a systematic, progressive course of study from the basics through intermediate and advanced topics.

Audio
320 pages • paperback • 7-3/8 x 9-1/4"
ISBN: 0-7906-1178-3 • Sams 61178
$34.95

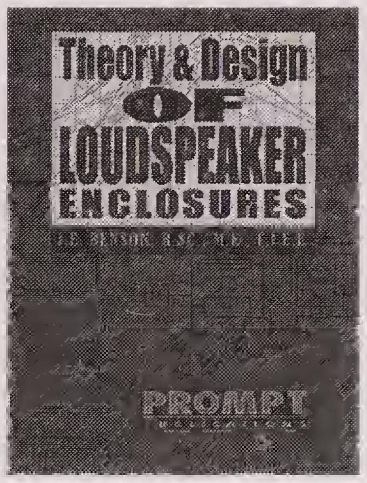

Theory and Design of Loudspeaker Enclosures

J.E. Benson

Written for design engineers and technicians, students, and intermediate-to-advanced level acoustics enthusiasts, *Theory & Design of Loudspeaker Enclosures* presents a general theory of loudspeaker enclosure systems. Full of illustrated and numerical examples, this book examines diverse developments in enclosure design, and studies the various types of enclosures as well as varying parameter values and performance optimization.

Topics examined in *Theory & Design of Loudspeaker Enclosures* include: The Synthesis of Vented Systems, Infinite-Baffle and Closed-Box Systems, Electro-Acoustical Relations, Reflex Response Relationships, System Response Formulae, Input Impedance, Circuit Parameters, System Parameters, Driver Parameters, Analogous Circuits, Terminology, And More.

Audio
244 pages • paperback • 6 x 9"
ISBN: 0-7906-1093-0 • Sams: 61093
$24.95

To order today or locate your nearest Prompt® Publications distributor at 1-800-428-7267 or www.samswebsite.com

Prices subject to change.

PROMPT®
PUBLICATIONS

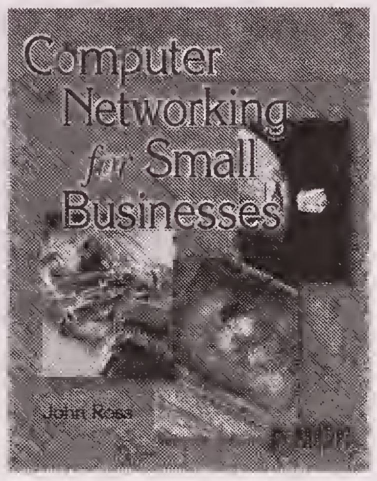

Computer Networking for Small Businesses

by John Ross

Small businesses, home offices, and satellite offices have flourished in recent years. These small and unique networks of two or more PCs can be a challenge for any technician. Small network systems are vastly different from their large-office counterparts. Connecting to multiple and off-site offices provides a unique set of challenges that are addressed in this book. Topics include installation, troubleshooting and repair, and common network applications relevant to the small-office environment.

Available in December

Communication
368 pages • paperback • 7-3/8" x 9-1/4"
ISBN 0-7906-1221-6 • Sams: 61221
$39.95

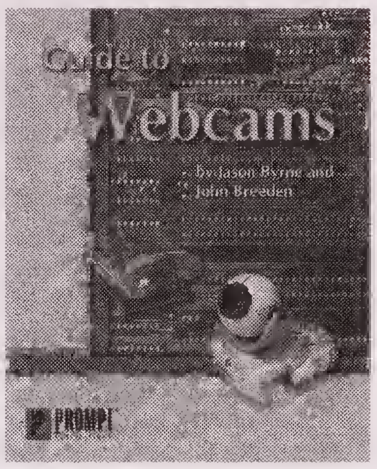

Guide to Webcams

by John Breeden & Jason Byre

Webcams are one of the hottest technologies on the market today. Their applications are endless and cross all demographics. Digital video cameras have become an increasingly popular method of communicating with others across the Internet. From video e-mail clips to web sites that broadcast people's lives for all to see, webcams have become a way for people to battle against the perceived threat of depersonalization caused by computers and to monitor areas from anywhere in the world. Projects and topics include picking a camera, avoiding obstacles, setup and installation, and applications.

Video Technology
320 pages • paperback • 7-3/8" x 9-1/4"
ISBN 0-7906-1220-8 • Sams: 61220
$29.95

To order today or locate your nearest Prompt® Publications distributor at 1-800-428-7267 or www.samswebsite.com

Prices subject to change.

PROMPT®
PUBLICATIONS

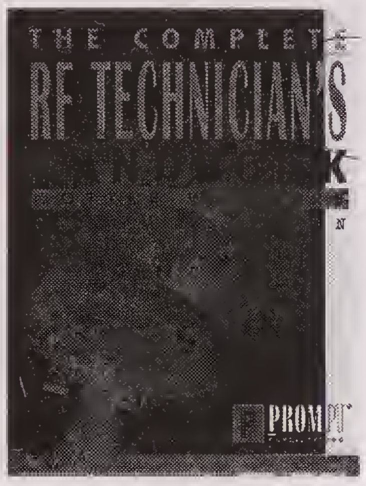

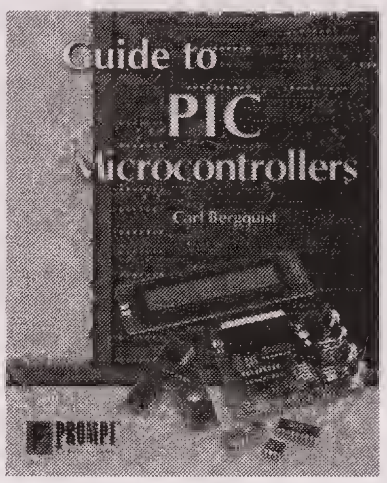

The Complete RF Technician's Handbook, *Second Edition*

Cotter W. Sayre

This is THE handbook for the RF or wireless communications beginner, student, experienced technician or ham radio operator, furnishing valuable information on the fundamental and advanced concepts of RF wireless communications.

Circuits found in the majority of modern RF devices are covered, along with current test equipment and their applications. Troubleshooting down to the component level is heavily stressed. RF voltage and power amplifiers, classes of operation, configurations, biasing, coupling techniques, frequency limitations, noise and distortion are all included, as well as LC, crystal and RC oscillators. RF modulation and detection methods are explained in detail – AM, FM, PM and SSB – and associated circuits, such as AGC, squelch, selective calling, frequency multipliers, speech processing, mixers, power supplies, AFC, multiplex-

Communications Technology
366 pages • paperback • 8-1/2 x 11"
ISBN: 0-7906-1147-3 • Sams: 61147
$34.95

Guide to PIC Microcontrollers

by Carl Bergquist

PICs, or peripheral interface controllers, easy to use and are the "chip of the '90s." Aimed at students and entry level technicians, Bergquist displays his expertise in the electronics field with this excellent guide on the use of PIC Microcontrollers. Projects and topics include:

- Electrical Structure
- Software Codes
- Prototype layout boards

Electronics Technology
336 pages • paperback • 7-3/8" x 9-1/4"
ISBN 0-7906-1217-8 • Sams 61217
$39.95

To order today or locate your nearest Prompt® Publications distributor at 1-800-428-7267 or www.samswebsite.com

Prices subject to change.

PROMPT®
PUBLICATIONS

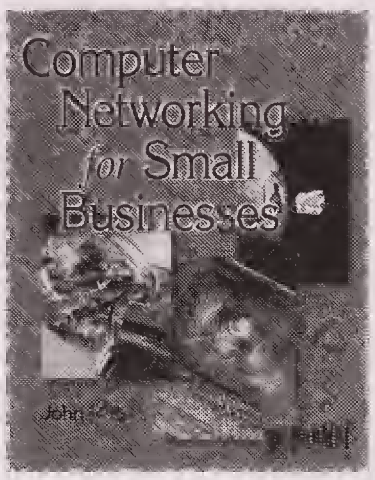
Computer Networking for Small Businesses
John Ross
PROMPT

Computer Networking for Small Businesses

by John Ross

Small businesses, home offices, and satellite offices have flourished in recent years. These small and unique networks of two or more PCs can be a challenge for any technician. Topics include:

- Installation
- Troubleshooting and repair
- Common network applications

Communication
368 pages • paperback • 7-3/8" x 9-1/4"
ISBN 0-7906-1221-6 • Sams 61221
$39.95

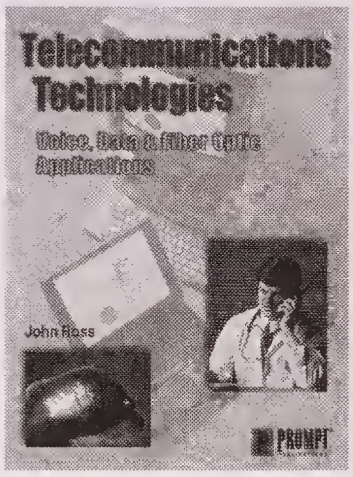
Telecommunications Technologies
Voice, Data & Fiber Optic Applications
John Ross
PROMPT

Telecommunications Technologies:

Voice, Data & Fiber Optic Applications

by John Ross

This book contains the information needed to develop a complete understanding of the technologies used within telephony, data and telecommunications networks. Projects and topics include:

- Equipment comparisons
- Business office applications
- Understanding the Technology

Communications
368 pages • paperback • 7-3/8" x 9-1/4"
ISBN 0-7906-1225-9 • Sams 61225
$39.95

To order today or locate your nearest Prompt® Publications distributor at 1-800-428-7267 or www.samswebsite.com

Prices subject to change.

PROMPT®
PUBLICATIONS

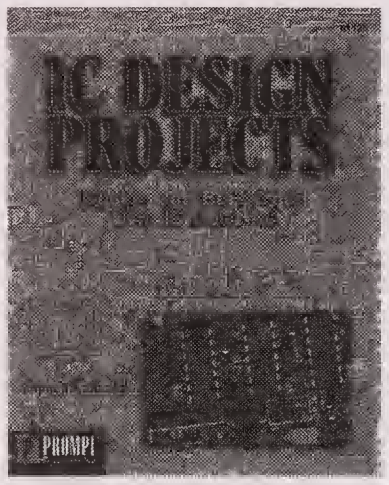

Carl J. Bergquist

IC Design Projects

Stephen Kamichik

IC Design Projects discusses some of the most popular and practical ICs, and offers you some projects in which you can learn to create useful and interesting devices with these ICs.

Once you have read through this book and completed some of its projects, you will have a real, working knowledge of ICs, enabling you to design and build you own projects!

Topics include: how power supplies operate, integrated circuit voltage regulators, TTL logic, CMOS logic, how operational amplifiers work, how phase-locked loops work, and more!

Projects include: battery charger, bipolar power supply, capacitance meter, stereo preamplifier, function generator, DC motor control, automatic light timer, darkroom timer, LM567 tone decoder IC, electronic organ, and more!

Projects
261 pages • paperback • 7-3/8 x 9-1/4"
ISBN: 0-7906-1135-X • Sams 61135
$24.95

Build Your Own Test Equipment

Carl J. Bergquist

Test equipment is among the most important tools that can be used by electronics hobbyists and professionals. Building your own test equipment can carry you a long way toward understanding electronics in general, as well as allowing you to customize the equipment to your actual needs.

Build Your Own Test Equipment contains information on how to build several pragmatic testing devices. Each and every device is designed to be highly practical and space conscious, using commonly-available components.

Projects include: Prototype Lab, Multi-Output Power Supply, Signal Generator and Tester, Logic Probe, Transistor Tester, IC Tester, Portable Digital Capacitance Meter, Four-Digit Counter, Digital Multimeter, Digital Function Generator, Eight-Digit Frequency Counter, Solid-State Oscilloscope, and more.

Professional Reference
267 pages • paperback • 7-3/8 x 9-1/4"
ISBN: 0-7906-1130-9 • Sams: 61130
$29.95

To order today or locate your nearest Prompt® Publications distributor at 1-800-428-7267 or www.samswebsite.com

Prices subject to change.

PROMPT®
PUBLICATIONS

Louis Columbus

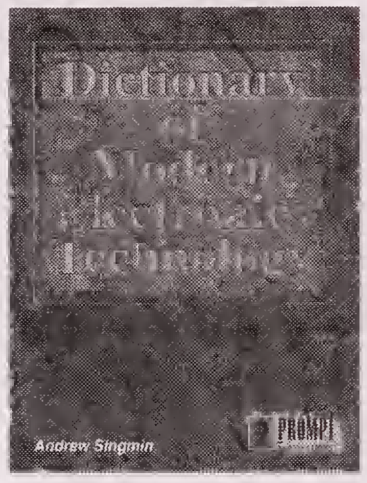

Exploring the World of SCSI

by Louis Columbus

Focusing on the needs of the hobbyist, PC enthusiast, as well as system administrator, *The World of SCSI* is a comprehensive book for anyone interested in learning the hands-on aspects of SCSI. It includes how to work with the Logical Unit Numbers (LUNs) within SCSI, how termination works, bus mastering, caching, and how the various levels of RAID provide varying levels of performance and reliability. This book provides the functionality that intermediate and advanced system users need for configuring SCSI on their systems, while at the same time providing the experienced professional with the necessary diagrams, descriptions, information sources, and guidance on how to implement SCSI-based solutions. Chapters include: How SCSI Works; Connecting with SCSI Devices; and many more.

Communication
500 pages • paperback • 7-3/8" x 9-1/4"
ISBN 0-7906-1210-0 • Sams 61210
$34.95

Dictionary of Modern Electronics Technology

Andrew Singmin

New technology overpowers the old everyday. One minute you're working with the quickest and most sophisticated electronic equipment, and the next minute you're working with a museum piece. The words that support your equipment change just as fast.

If you're looking for a dictionary that thoroughly defines the ever-changing and advancing world of electronic terminology, look no further than the Modern Dictionary of Electronics Technology. With up-to-date definitions and explanations, this dictionary sheds insightful light on words and terms used at the forefront of today's integrated circuit industry and surrounding electronic sectors.

Whether you're a device engineer, a specialist working in the semiconductor industry, or simply an electronics enthusiast, this dictionary is a necessary guide for your electronic endeavors.

Electronics Technology
220 pages • paperback • 7 3/8 x 9 1/4"
ISBN: 0-7906-1164-4 • Sams 61164
$34.95

To order today or locate your nearest Prompt® Publications distributor at 1-800-428-7267 or www.samswebsite.com

Prices subject to change.

PROMPT®
PUBLICATIONS

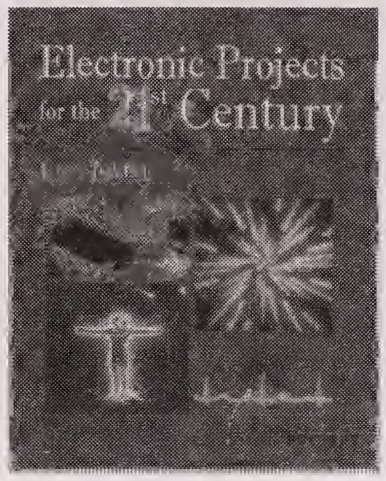
Electronic Projects for the 21st Century

Electronic Projects for the 21st Century

John Iovine

If you are an electronics hobbyist with an interest in science, or are fascinated by the technologies of the future, you'll find *Electronic Projects for the 21st Century* a welcome addition to your electronics library. It's filled with nearly two dozen fun and useful electronics projects designed to let you use and experiment with the latest innovations in science and technology. This book contains the expert, hands-on guidance and detailed instructions you need to perform experiments that involve genetics, lasers, holography, Kirlian photography, alternative energy sources and more. You will obtain all the information necessary to create the following: biofeedback/lie detector device, ELF monitor, Geiger counter, MHD generator, expansion cloud chamber, air pollution monitor, laser power supply for holography, pinhole camera, synthetic fuel from coal, and much more.

Projects
256 pages • paperback • 7-3/8 x 9-1/4"
ISBN: 0-7906-1103-1 • Sams: 61103
$24.95

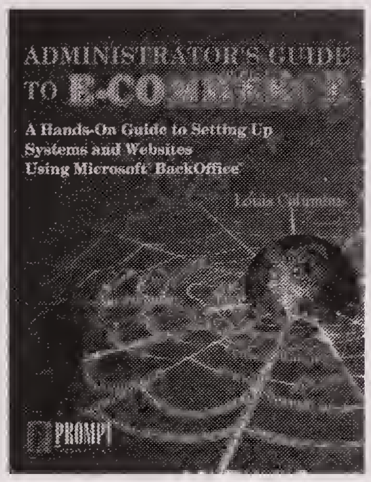
ADMINISTRATOR'S GUIDE TO E-COMMERCE
A Hands-On Guide to Setting Up Systems and Websites Using Microsoft BackOffice
Louis Columbus
PROMPT

Administrator's Guide to E-Commerce,

A Hands-On Guide to Setting Up Systems and Websites

by Louis Columbus

Unlike previous electronic commerce books which stress theory, the *Administrator's Guide to E-Commerce* is a hands-on guide to creating and managing Web sites using the Microsoft BackOffice product suite. A complete guide to setting up your own E-Commerce site.

Business Technology
416 pages • paperback • 7-3/8" x 9-1/4"
ISBN 0-7906-1187-2 • Sams 61187
$34.95

To order today or locate your nearest Prompt® Publications distributor at 1-800-428-7267 or www.samswebsite.com

Prices subject to change.

PROMPT®
PUBLICATIONS

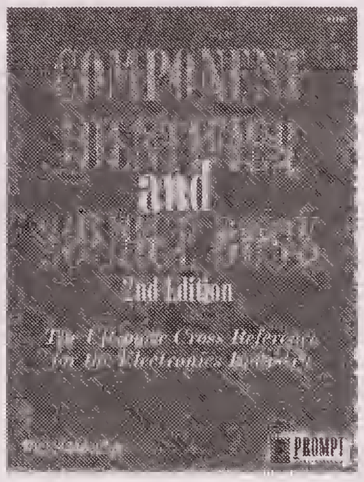

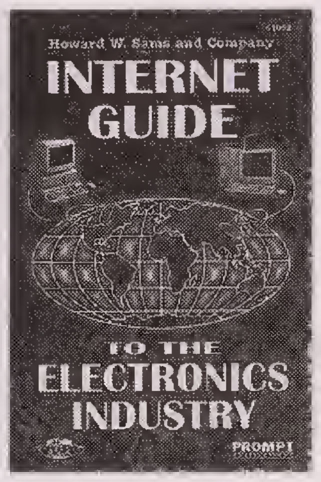

The Component Identifier and Source Book, *2nd Edition*

Victor Meeldijk

Because interface designs are often reverse engineered using component data or block diagrams that list only part numbers, technicians are often forced to search for replacement parts armed only with manufacturer logos and part numbers.

This book was written to assist technicians and designers in identifying components from prefixes and logos, as well as find sources for various types of microcircuits and other components. It will help readers cross reference component types to manufacturers, and also cross reference trade names, abbreviations, and part number prefixes to the manufacturer. Listings of the worldwide manufacturing and sales office addresses are provided, with a special listing of manufacturers who provide replacement devices and vendors who specialize in stocking and locating discontinued devices. There is not another book on the market that lists as many manufacturers of such diverse electronic components.

Professional Reference
425 pages • paperback • 8-1/2 x 11”
ISBN: 0-7906-1159-7 • Sams: 61159
$34.95

Internet Guide to the Electronics Industry

John J. Adams

Although the Internet pervades our lives, it would not have been possible without the growth of electronics. It is very fitting then that technical subjects, data sheets, parts houses, and of course manufacturers, are developing new and innovative ways to ride along the Information Superhighway. Whether it’s programs that calculate Ohm’s Law or a schematic of a satellite system, electronics hobbyists and technicians can find a wealth of knowledge and information on the Internet.

In fact, soon electronics hobbyists and professionals will be able to access on-line catalogs from manufacturers and distributors all over the world, and then order parts, schematics, and other merchandise without leaving home. The *Internet Guide to the Electronics Industry* serves mainly as a directory to the resources available to electronics professionals and hobbyists.

Professional Reference
192 pages • paperback • 5-1/2 x 8-1/2”
ISBN: 0-7906-1092-2 • Sams: 61092
$19.95

To order today or locate your nearest Prompt® Publications distributor at 1-800-428-7267 or www.samswebsite.com

Prices subject to change.

PROMPT®
PUBLICATIONS

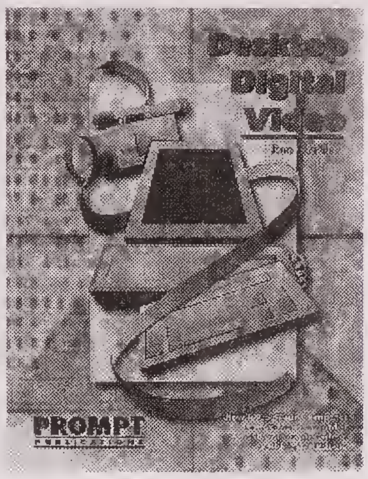

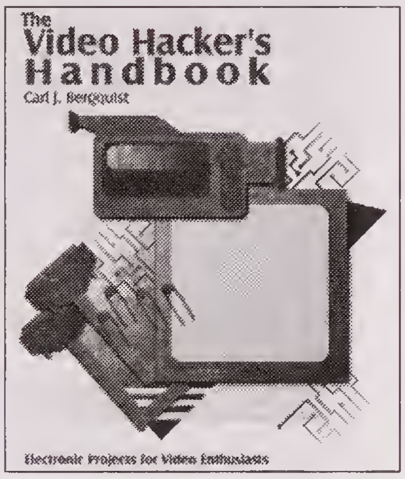
The Video Hacker's Handbook
Carl J. Bergquist
Electronic Projects for Video Enthusiasts

Desktop Digital Video

by Ron Grebler

Desktop Digital Video is for those people who have a good understanding of personal computers and want to learn how video (and digital video) fits into the bigger picture. This book will introduce you to the essentials of video engineering, and to the intricacies and intimacies of digital technology. It examines the hardware involved, then explores the variety of different software applications and how to utilize them practically. Best of all, *Desktop Digital Video* will guide you through the development of your own customized digital video system. Topics covered include the video signal, digital video theory, digital video editing programs, hardware, digital video software and much more.

Video Technology
225 pages • paperback • 7-3/8 x 9-1/4"
ISBN: 0-7906-1095-7 • Sams: 61095
$34.95

The Video Hacker's Handbook

by Carl Bergquist

Geared toward electronic hobbyists and technicians interested in experimenting with the multiple facets of video technology, *The Video Hacker's Handbook* features projects never seen before in book form. Video theory and project information is presented in a practical and easy-to-understand fashion, allowing you to not only learn how video technology came to be so important in today's world, but also how to incorporate this knowledge into projects of your own design. In addition to the hands-on construction projects, the text covers existing video devices useful in this area of technology plus a little history surrounding television and video relay systems.

Video Technology
336 pages • paperback • 7-3/8 x 9-1/4"
ISBN: 0-7906-1126-0 • Sams: 61126
$29.95

To order today or locate your nearest Prompt® Publications distributor at 1-800-428-7267 or www.samswebsite.com

Prices subject to change.

$34.95

1218

San Jose Public Library
3 1197 06095222 0

Gone are the days when electricians dealt only with wires, outlets, and switches. As new electronic devices extend their reach and influence into our daily lives, electricians are finding it necessary to adapt and improve their skills to meet the new demands. In *Electronics for the Electrician*, author Newton C. Braga gives electricians and other professionals a perfect way to do just that.

Electronic components in electric circuits and installations are becoming more common. Timers, dimmers, heater controls, intercoms, alarms, doorbells, battery chargers, emergency lights, illumination controls, air conditioners, washing machines, and electronic injection systems are all electric devices containing electronic circuits.

Until some years ago, electronics and electricity were independent fields. Electronics technicians fixed and repaired only electronic devices, concentrating their efforts in radio and TV, while electricians worked with cars, home and industrial installations, and energy distribution. They were entirely different fields. Now electricians must know how a timer or a dimmer works, because they need to know how to repair one if necessary.

Braga breaks his text into three sections:

- The Differences Between Electricity and Electronics
- Electronic Components
- Troubleshooting and Repair

Electronics for the Electrician isn't directed to all people; it's a book for electricians and other professionals—from construction engineers to architects. This book is intended to give professionals basic knowledge about electronics, in order to understand how electronic circuits and devices work in electric installations.

Newton C. Braga has written more than 50 books about electronics and electricity that have been published in his native Brazil, as well as Europe and the United States. The main training network of the Brazilian Federation of Industries (SENAI) recommended Braga's *Electric Installations For Beginners* and *The Basic Electronics Course*, published in Brazil, as basic texts in its courses.

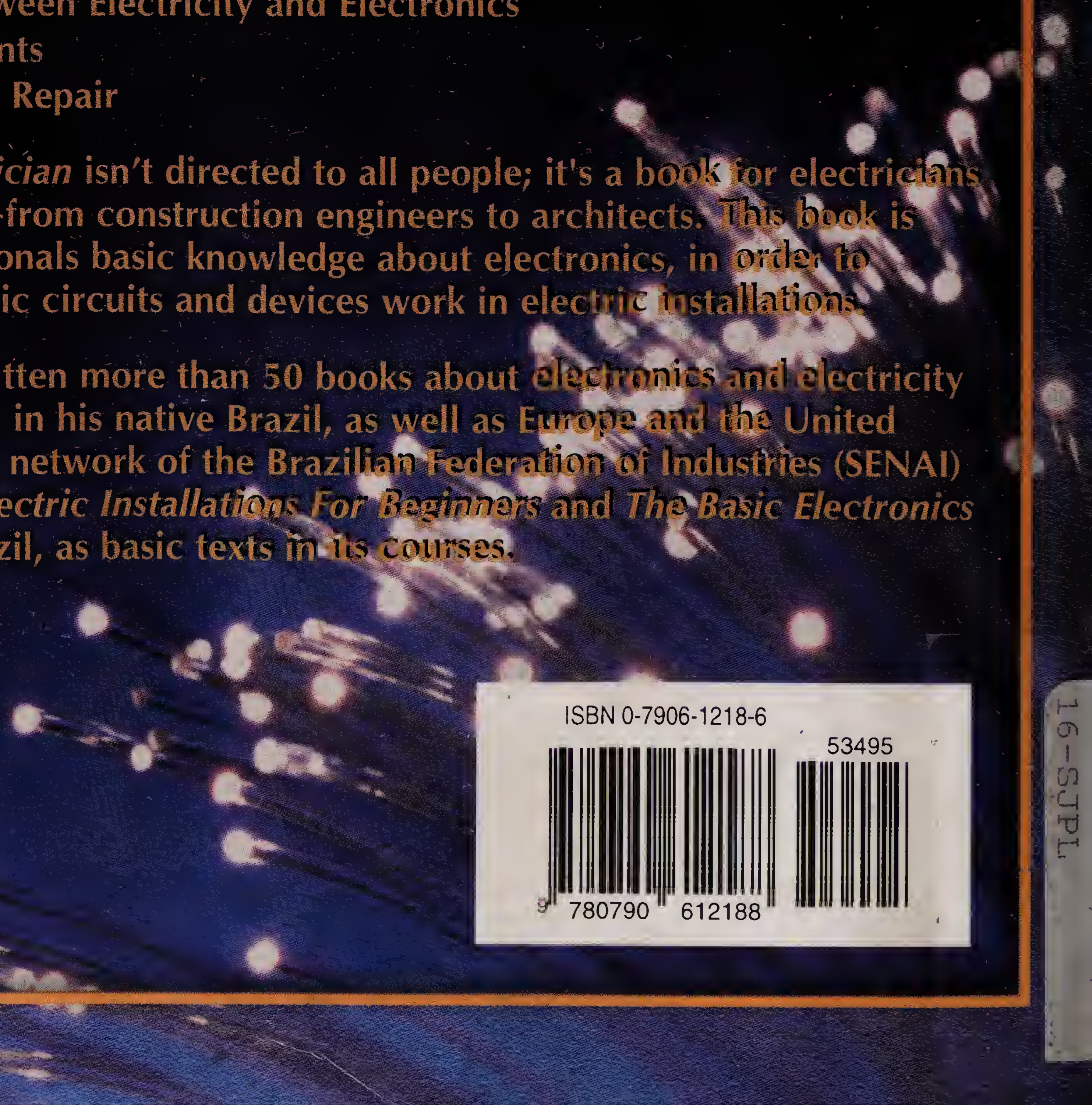

PROMPT®
PUBLICATIONS

SAMS Technical Publishing
Indianapolis, IN 46268-4149
800.428.SAMS
www.samswebsite.com

ISBN 0-7906-1218-6
53495
9 780790 612188

16-SJPL

61149

FUN PROJECTS *for the* EXPERIMENTER

Newton C. Braga

PROMPT®
PUBLICATIONS

During more than 20 years, the author, as a collaborator with American, European and Latin American electronics magazines*, has published a large assortment of practical circuits using common parts.

Now many of those projects are included in this volume, most of which you can build in one evening. The projects range from fun types through practical types to amusement types. Of course, there are other devices that can be used to teach you something about circuits or components.

An important feature of theses projects are the Ideas to Explore, intended for students looking for projects in science or to use in practical research.

We can consider this book as a source book of the easiest and fun-to-make of the hundreds of projects created and published by the author during these last years.

But, as the projects are in a wide range of types, we should separate the electronics experimenters into two groups: the ones who want to improve or expand some other area of their electronics interest, such as computers, radio, instrumentation, audio, security and even games; and the ones who want to learn something about electronic circuits and devices or want new ideas to use in science projects.

Most of the projects described herein can be stand-alone as individual accessories; wherever possible the circuits have been designed so that they can be ganged with one or more other projects. For example, many projects of audio effects or generators can be ganged with audio amplifiers or high-power output stages.

All the projects are simple, with few low-cost components that can be made in one evening of work.

To make it easy for the reader to choose the projects we added codes after each title to indicate the kind of experimenter for whom they are intended: the code "P" (Practical) indicates a project built for practical use. The code "E" (Experimental) indicates that the project is intended for the experimenter to teach something about circuits or devices.

Of course, you can also find projects with both codes (E and P)which can either be mounted to teach something or/and be used for a practical end.

The presentations of the projects are practical. Electronics components are listed with each circuit diagram. But secondary parts such as sockets, chassis, enclosures, miscellaneous hardware and so on, are not specified, since the reader is free to choose these non-critical items according to his preferences and demands.

The manner in which the circuits work and can be modified is explained in practical terms so the reader can acquire some knowledge of practical electronics as he progresses through the book.

Although many of the projects we enjoyed constructing as they are described here, you may think of possible modifications. We just recommend that you go ahead and modify the circuits to your personal ends. There is a wide latitude in circuit modifications and most of them will be of value to the experimenter who wants to see how things work, even though each project's primary intent is for the builder who desires a functional item of equipment as the result of his work.

As the book includes easy-to-build projects, the author hopes it will help you to learn many of the fundamentals of electronics in an easy and fun way, and, if you're a student, provide a source for school projects.

Newton C. Braga

* Mini-Projects is the title of a column published by the author during several years in a Brazilian Electronics Magazine named Eletrônica Total (Total Electronics).

MICRO-POWER LED FLASHER (E/P)

A simple LED flasher can be built with a TLC7555 Timer used as a low-frequency oscillator and drive to an LED. IC1 runs in a low frequency determined by R2/R3 and C1. Frequency is in the range between 0.1 and 5 Hz but you can alter it by changing C1. This capacitor can range from 1 to 10 uF.

Duty cycle is determined by R2/R3 ratio and is important in applications where you need reduction of energy consumed. Though Figure 1 shows fixed resistors for R2 and R3, you can use trimmer potentiometers if you prefer. The potentiometers, of course, are handy for altering the frequency and duty cycle of the oscillator.

R1 values depend upon the power supply voltage. With power supplies between 5 and 6 volts, R1 is 470 ohms. With power supplies between 6 and 9 volts, R1 is 560 or 680 ohms. With power supplies in the range between 9 and 12 volts, R1 must be 1,000 ohms. The circuit drains only 250 uA from a 6-volt power supply.

Figure 1 shows the schematic diagram for the Micro-Power LED Flasher. At the heart of the circuit is IC1, a 555 timer.

Use the parts placement diagram as a guide when assembling the Micro-Power LED Flasher. Components placement, using a terminal strip as chassis, is shown in ***Figure 2***.

Take special care with positions of polarized components, such as the LED and electrolytic capacitor.

After you are sure that you have wired all the components correctly, put the batteries in the battery holder. The LED should flash.

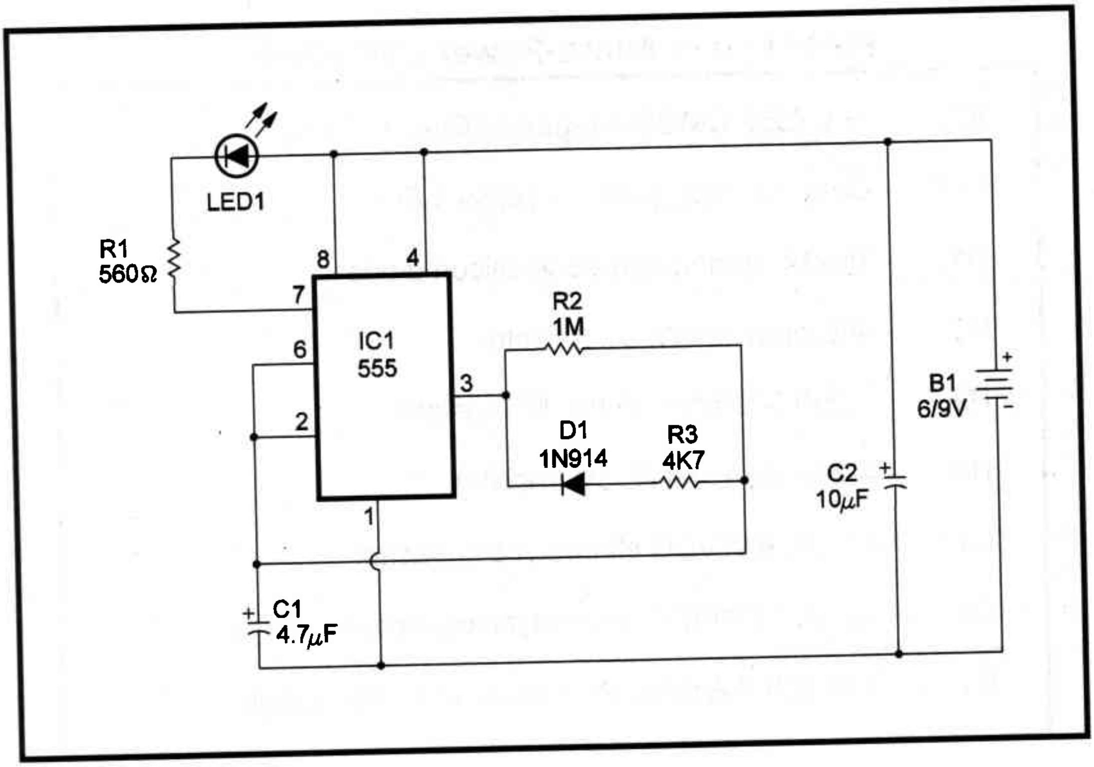
LED1
R1
560Ω
8
4
7
6
2
3
1
IC1
555
R2
1M
D1
1N914
R3
4K7
C2
10μF
+
B1
6/9V
+
-
+
C1
4.7μF

Figure 1

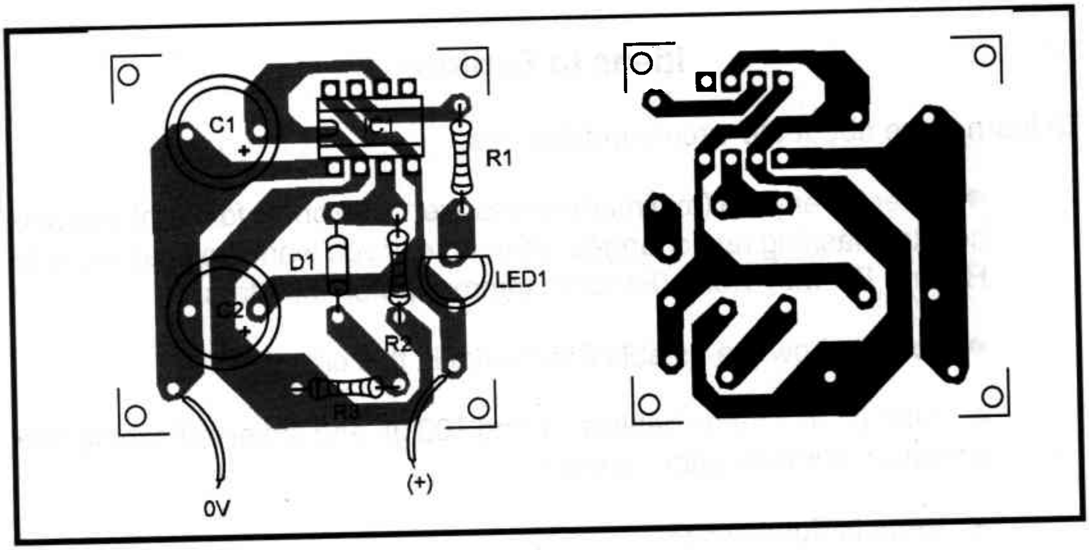
C1
+
IC1
R1
D1
LED1
C2
+
R2
R3
0V
(+)

Figure 2

Parts List — Micro-Power LED Flasher

IC1 - TLC7555 CMOS Integrated Circuit - Timer

LED1 - Common red, green or yellow LED

D1 - 1N914 general purpose silicon diode

R1 - 560 ohm, 1/4W, 5% resistor

R2 - 1,000,000 ohm, 1/4W, 5% resistor

R3 - 4,700 ohm, 1/4W, 5% resistor

C1 - 4,7 uF, 6 WVDC electrolytic capacitor

C2 - 10 uF, 12 WVDC electrolytic capacitor

B1 - 6 to 12V AA cells, 9V battery or power supply

Ideas to Explore

To learn more about the circuit and devices:

- Alter R2 and R3 (minimum values are 1,000 ohms for both) and see how the flashing rate changes. Of course, if you want, you can replace R2 and R3 with 1,000,000-ohm trimmer potentiometers.

- Explain how the astable 555 works in this circuit.

- Alter C1 in a range between 1 and 100 uF and observe flashing rate alteration. Explain what happens.

- What is duty-cycle?

Science experiments with the circuit:

- LED is a monochromatic light source. You can perform some experiments in optics using this kind of light source. Chromatic filters can be used in some experiments involving LEDs. Try to use LEDs of different colors (red, yellow, green and, if possible, blue).
- Light flashes, as the ones produced by this circuit, can be used in hypnotic experiments.
- The circuit can be used to indicate operation of other electronic devices. Wired to the power supply of a monitored device, the blinker LED will indicate its operation.

Remember that the theory in this book is specifically related to the associated experiments and projects. This book is designed as an introduction, not as an in-depth treatment of electronics. If you are interested in finding out more about the subjects touched on in this book, several texts can be found for further reading.

SET-RESET FLIP-FLOP (E)

With this project we can show how two NAND gates of a 7400 IC (TTL Integrated Circuit) can be wired to perform as an R-S flip-flop.

This circuit is a pulse-triggered flip-flop that needs negative-going pulses to be triggered. This circuit operates as follows: as we see, it has two outputs; a normal output called Q (LED1) and an inverted output called $\overline{Q}$ (LED2). When one output is 1 the other necessarily will be 0 and vice-versa because they are complementary.

The circuit has also two inputs named S (SET) and R (RESET) as shown in the schematic diagram, where the trigger signals are applied.

R input is wired to Q output and S input is wired to $\overline{Q}$ output, forming a closed-loop for the digital signals.

When a negative-going trigger pulse is applied to S input the output Q swings to the 1 state. As this output is wired to R input, the 1 state causes output $\overline{Q}$ to fall to a 0 level. But $\overline{Q}$ output is also wired to S input causing a feedback that makes its output remain at 1 also after the trigger pulse has disappeared. To trigger the flip-flop again, changing the output states, we should apply a negative-going pulse to R input. This pulse causes the output to go to 1, and as this output is wired to R input, the trigger pulse also causes the output Q to go to 0 level.

A zero in this output goes to R input, and also after the trigger pulse disappears the outputs remain in their states.

See that the circuit has two stable states, and we only can change these states with set or reset (S or R) negative-going pulses applied to its inputs.

Our project is a manually-triggered flip-flop and can be constructed wiring a switch to the R-S inputs as shown in ***Figure 1***.

Logic states are indicated by two LEDs. LED1 glows with a 1 at Q output and LED2 glows with a 1 at $\overline{Q}$ output.

The circuit must be powered by a 5-volt regulated power supply as it uses a TTL IC. Current requirements range typically from 5 to 15 mA.

Components placement on a small printed-circuit board is shown in ***Figure 2***.

Experiments like this can also be performed on what electronics experimenters call “breadboards”. These are boards on which the parts of experiments can be temporarily assembled. By using breadboards it is not necessary to

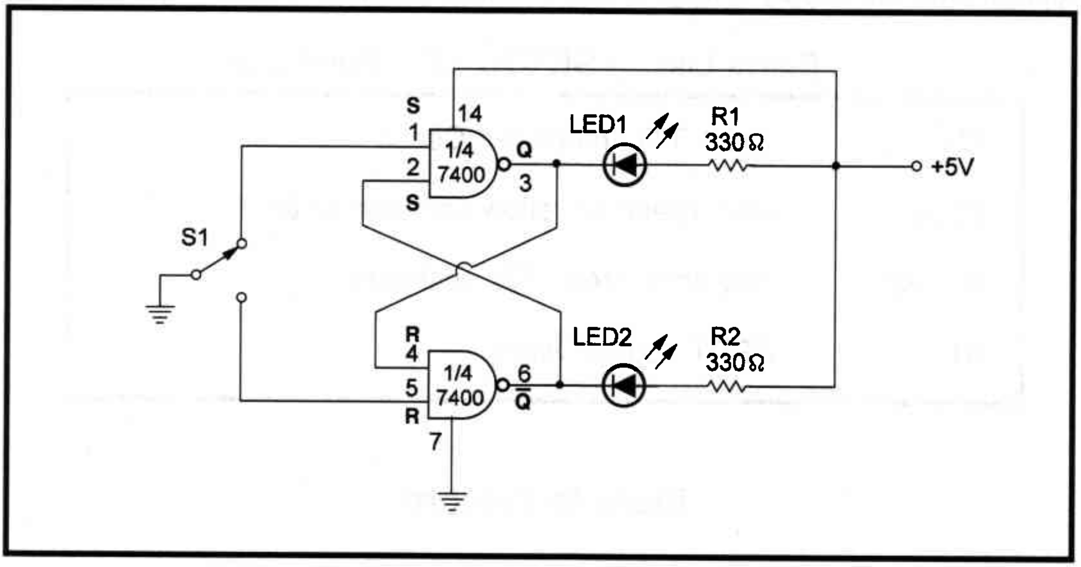
S
14
1
1/4
7400
Q
2
3
S
LED1
R1
330Ω
+5V
S1
R
4
1/4
7400
5
6
Q̄
R
7
LED2
R2
330Ω

Figure 1

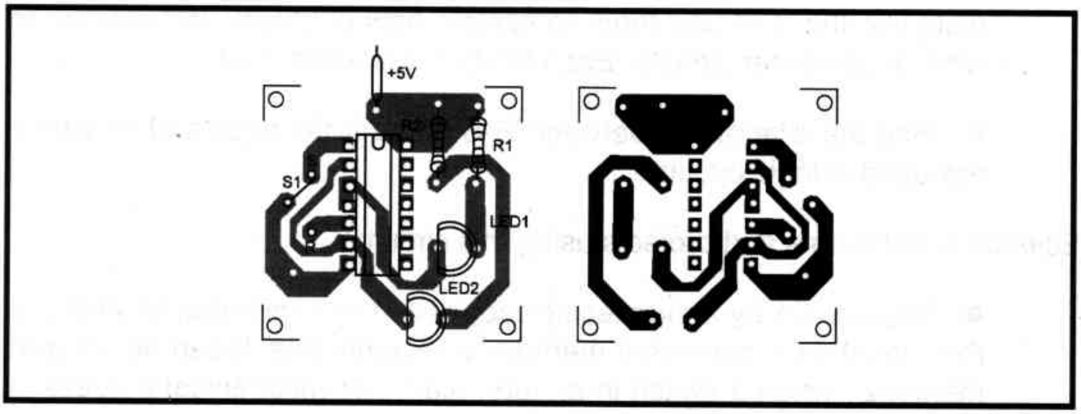
+5V
R2
R1
S1
LED1
LED2

Figure 2

solder the components. This is why they are also called "solderless boards"

The circuit can be used to teach much about flip-flops, used in computers as counters and in memories and many other applications.

Observation: In digital electronics an electrical signal is either high or low. These two states are used to represent binary bits 1 and 0. Since digital logic ICs operate from a single-ended power supply, a high state represents a voltage near the supply voltage and a low state represents a voltage near 0V.

PARTS LIST — SET-RESET FLIP-FLOP

IC1	- 7400 TTL Integrated Circuit
LEDs	- Red, green or yellow common LEDs
R1, R2	- 330 ohm, 1/4W, 5% resistors
S1	- SPDT toggle switch

Ideas to Explore

To learn more about the circuit and devices:

- The common flip-flop is a fundamental counter circuit. Mount other units like this and use them to explain how counters and memories work in computer circuits. Each flip-flop can store a bit.

- Find the differences between R-S and J-K flip-flops and how they are used in logic circuits.

Science experiments and projects using this circuit:

- Replace S1 by a magnetic (reed) switch or a microswitch and use this circuit as a one-event memory in experiments. It can detect and memorize when a switch in an lure, alarm, or mechanical process is closed for a moment.

MINI-METRONOME (P)

Here is a dandy two-transistor metronome which you can build into a small plastic case. You can use an old nonworking transistor radio to house this gadget. In most case you can merely remove the main circuitry but retain some parts, such as the on-off volume control potentiometer, speaker and battery holder. If you have done this, you have just eliminated the following items: SPKR, S1, and B1.

The circuit can be powered with two or four AA cells and drains about 10 mA.

As soon as S1 is closed, the metronome starts clacking. By adjusting potentiometer P1 you find the "beat" or frequency of clacking you wish to set.

To calibrate, you have only to synchronize your metronome with a commercial type or any known source.

Figure 1 shows the schematic diagram for the Mini-Metronome. At the heart of the circuit are Q1 and Q2, as a two-stage direct-coupled amplifier, which allows the circuit to generate intervalled pulses.

Figure 2 shows a mechanical view of the metronome, using a terminal strip as "chassis" for the small components.

Position of the polarized components, such as C1, C2 and the transistor, should be observed in the mounting.

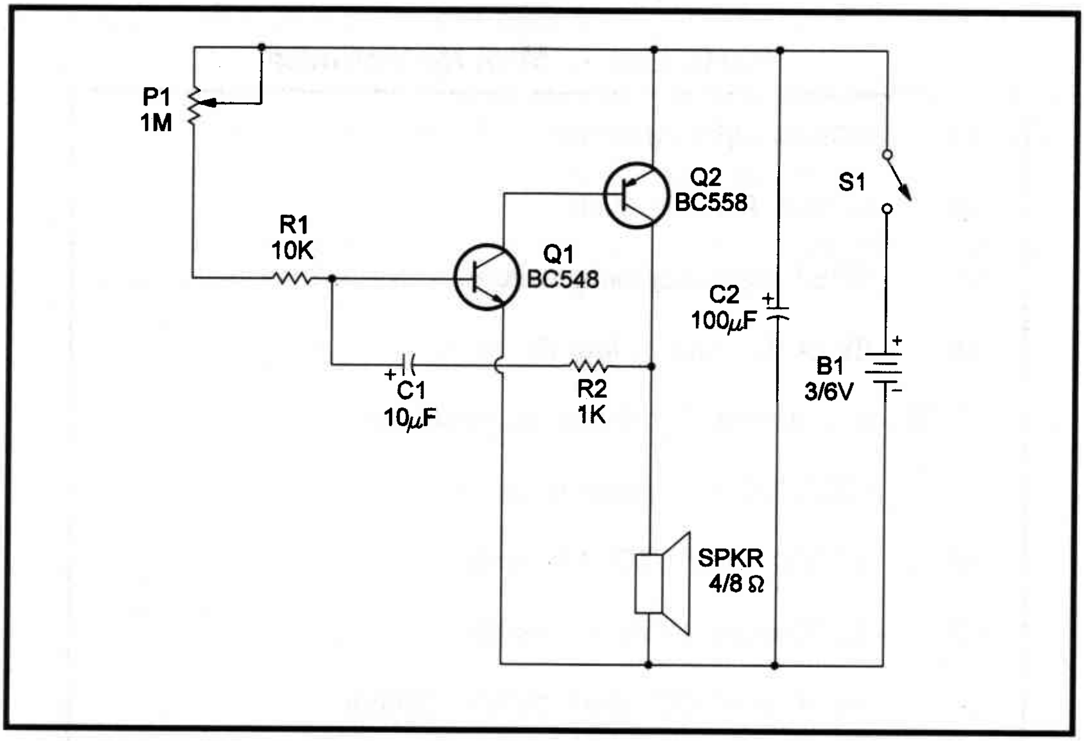
P1
1M
R1
10K
Q1
BC548
Q2
BC558
S1
C2
100µF
+
B1
3/6V
+
-
+
C1
10µF
R2
1K
SPKR
4/8 Ω

Figure 1

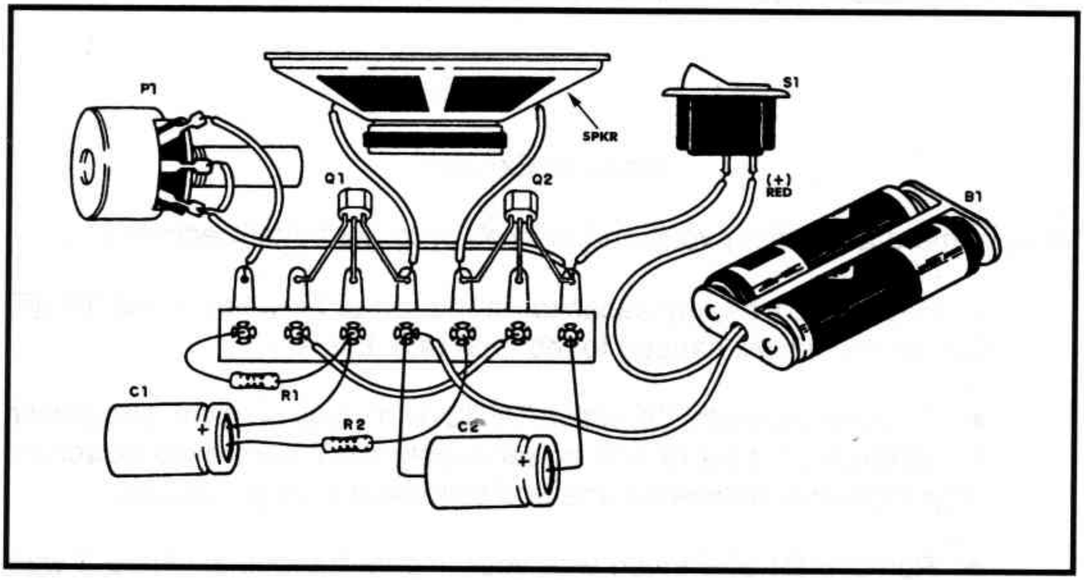
P1
SPKR
S1
Q1
Q2
(+)
RED
B1
C1
R1
+
R2
C2
+

Figure 2

Parts List — Mini Metronome

Q1 - BC548 NPN transistor

Q2 - BC558 PNP transistor

S1 - SPST toggle or slide switch

B1 - 3V or 6V - two or four AA cells

SPKR - 4 or 8 ohm, 2 to 4-inch loudspeaker

P1 - 1,000,000 ohm potentiometer

R1 - 10,000 ohm, 1/4W, 5% resistor

R2 - 1,000 ohm, 1/4W, 5% resistor

C1 - 10 μF, 6 WVDC electrolytic capacitor

C2 - 10 μF, 6 WVDC electrolytic capacitor

Ideas to Explore

To learn more about the circuit and devices or to get better performance:

- Replace C1 by any capacitor in the range between 1 and 10 μF. Observe what will happen to the beat rate. Explain it.

- Replace transistor Q2 with a TIP32 on a small heatsink and power the circuit from a 9V or 12V power supply. You'll have extra power for your experimental metronome and can use it in larger places.

- Remove P1 and touch with your fingers the points where it was wired. Will the circuit operate? Can you explain what is going on?

Science experiments and different uses for the circuit:

- The metronome can be used in gymnastic exercises or in running to find the correct rhythm. The metronome can be housed in a small plastic box, powered by cells and easily carried with the runner.
- Biological experiments with sounds can be made with this device; experiments in animal conditioning, for example.

PENDULUM ALARM (E/P)

Any movement that swings the pendulum sensor triggers this alarm, enabling a relay.

The circuit can be used to protect cars, large objects, homes and other places as the sensor can be used to detect any movement. This device can also be used for scientific experiments, as suggested at the end of the section.

The relay can drive powerful sound systems, such as sirens, horns, lamps, etc. The circuit can be powered by a 12-volt battery (when used in cars, for instance) or a power supply. Current drain is very low (about 100 uA) when the relay is deenergized, extending battery life.

The schematic circuit of the Pendulum Alarm is shown in ***Figure 1.*** The circuit has as its heart an SCR (Silicon Controlled Rectifier), which energizes the relay when the sensor detects any movement.

A mechanical view of the circuit mounting, using a terminal strip as chassis, is shown in ***Figure 2***.

Relay coil voltage is chosen according to the power supply voltage. You can use a mini DPDT 1A Mini Relay (Radio Shack 275-249) or, for heavy loads, a 10A mini SPDT relay (Radio Shack 275-248). These relays are rated to 12 VDC and drain only 38 mA (their coil resistance is about 320 ohms). As equivalent types can be used, you should take care with terminal layout.

Reset can be made with a momentary-type switch (Push-On) wired between the SCR anode and catode, or by using the on-off switch.

Positions of polarized components, such as the SCR and diode, should be observed.

Sensor details are also given in ***Figure 2***. Notice that operation occurs when the vertical (or horizontal) bare wire touches the ring during any movement.

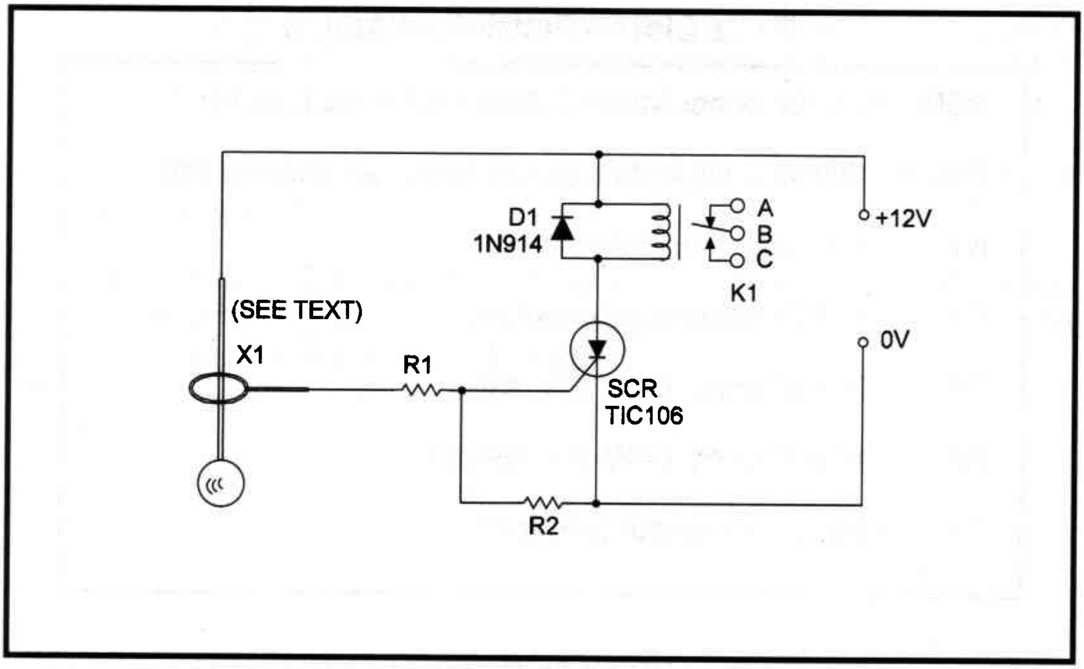
D1
1N914
A
B
C
K1
+12V
(SEE TEXT)
X1
R1
SCR
TIC106
0V
R2

Figure 1

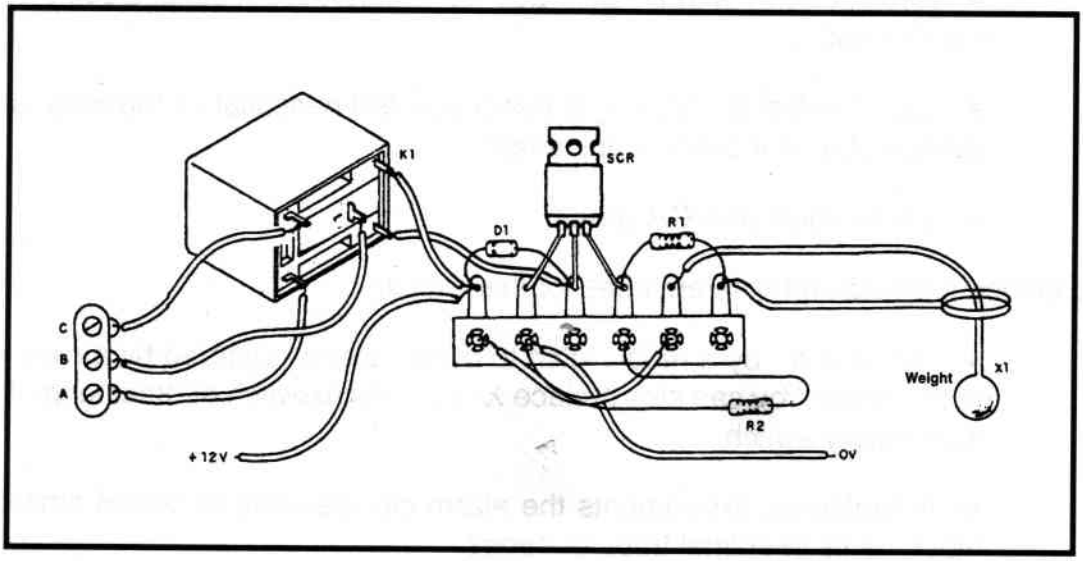
K1
SCR
D1
R1
C
B
A
Weight
X1
R2
+12V
0V

Figure 2

Parts List — Pendulum Alarm

SCR	- TIC106 or equivalent Silicon Controlled Rectifier
D1	- 1N914 or equivalent general-purpose silicon diode
K1	- 12V relay (see text)
S1	- SPST slide or toggle switch
R1	- 22,000 ohm, 1/4W, 5% resistor
R2	- 10,000 ohm, 1/4W, 5% resistor
X1	- Pendulum sensor (see text)

Ideas to Explore

To learn more about circuit and devices:

- Replace relay with a 12V x 200 mA incandescent lamp. Does the circuit work?
- Explain what D1 is doing in this circuit. Is it necessary if the relay is replaced by an incandescent lamp?
- Explain how an SCR works.

Science projects and different uses for this circuit:

- Replace X1 by a reed switch to get an alarm triggered by a magnetic sensor. You can also replace X1 by a microswitch or other kind of momentary switch.
- In biological experiments the alarm can be used to detect small movements in animal traps or cages.

- Replacing X1 by two metal wires enables the circuit to be triggered by water. Water level detectors can be made using this kind of sensor, and it can be used in many experiments and practical applications.

- Installed in your car, this circuit can be used as an alarm, driving the horn or a siren.

- By replacing X1 with a large metal plate, the circuit can be used to detect wind. Air flow can move the plate so that the bare wire touching the loop causes the alarm to trigger.

MEDIUM-WAVE RF PRESELECTOR (P)

The sensivity of most moderately-priced AM receivers is poor in the medium-wave band. Worse yet, selectivity also drops off sharply within the upper part of the band, making the receiver highly susceptible to images, noise and annoying heterodynes.

Even the best antenna and a good booster amplifier can't make up for these deficiencies in a common receiver's front end. But, using an RF preselector between the antenna and the receiver's antenna input terminals will give you real increase in the reception quality.

This circuit matches the antenna impedance with the receiver's antenna input according the frequency of the tuned station.

Figure 1 shows schematic diagram of the Medium-Wave Preselector.

Figure 2 shows an exploded view of the circuit, which can be housed in a small plastic box.

L1 is formed by turns of No. 28 enameled wire, wound on a 5/16-inch ferrite form. Taps are made each 15 or 20 turns and its number is not critical. The number of taps can be altered.

Attach your antenna and ground to the input terminals of the preselector. Tune to the desired station on your receiver, and, by choosing the best position of the alligator clip on the strip and by adjusting CV, find the best signal.

The required antenna will depend upon the station you want to tune. A random length of wire will give fairly good performance through the AM band. Antennas with lengths between 15 and 150 feet are recommended for best performance.

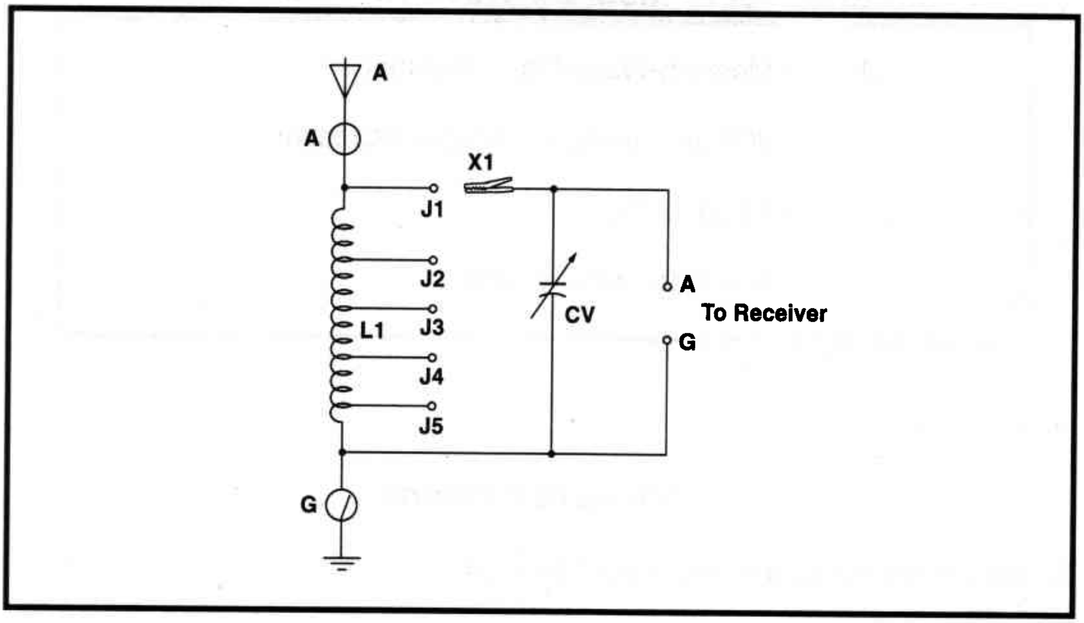
A
A
X1
J1
J2
J3
L1
J4
J5
CV
A
To Receiver
G
G

Figure 1

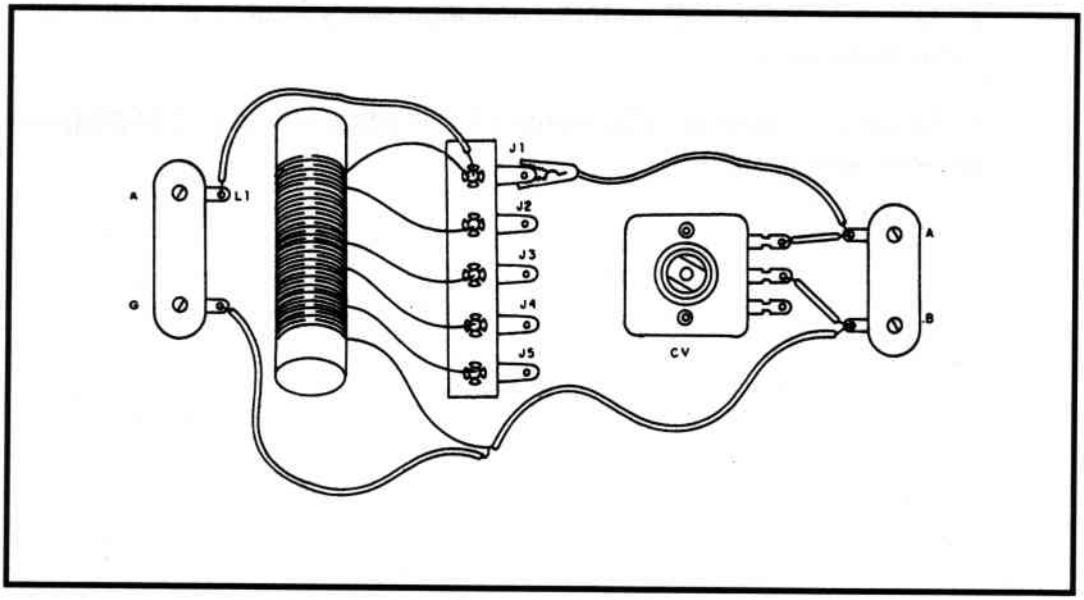
J1
A
L1
J2
J3
G
J4
J5
CV
A
B

Figure 2

Parts List — Medium-Wave RF Preselector

L1	- Medium-Wave Coil - see text
CV	- 365 pF miniature variable capacitor
X1	- Alligator clip
	- 2-position barrier strip

Ideas to Explore

To learn more about the circuit and devices:

- Explain the importance of a good ground connection in this circuit. What is the meaning of "impedance"?

- Use a new coil with 20 turns and taps every 3 turns to tune short-wave stations.

- Make a second coil (20 turns) on the same ferrite rod to plug the antenna and ground.

ELECTRONIC FISHING LURE (P)

This circuit imitates the sound of wet bugs milling about the surface to attract curious fish.

The circuit is basically a Hartley audio oscillator operating in a frequency between 200 and 2,000 Hertz. Powered by an AA cell, it can housed in a clean empty peanut butter jar with the buzzer circuitry mounted upside-down on the cover. You must be carefull not to drill in such a way as to allow water to leak through the cap. Silicon rubber can be used to close any water leak.

Capacitor C1 would give a resonant frequency of approximately 1,000 Hz, but P1 would give a wide variation in signal tone generated in the audio range.

You can also use a low-impedance earphone as a transducer, connecting it to the second coil (output) of the transformer.

To operate, unscrew the cap, throw S1 ON, put the cover back on, and drop the fish lure into the water in the vicinity of where you will be fishing. Allow 20 minutes before reeling them in. If necessary, adjust P1 to a new signal tone.

Figure 1 shows the schematic diagram for the Fish Lure. Observe that the secondary coil of the small audio transformer is not used in this circuit.

Construction details of the Electronic Fishing Lure, using a terminal strip as chassis, is shown in ***Figure 2***. The exact locations of components is not critical.

T1 is also not critical. Any 500:8 or 1,000:8 transistor transformer with center-tapped primary can be used. We can also suggest that any small transistor transformer can be experimented with.

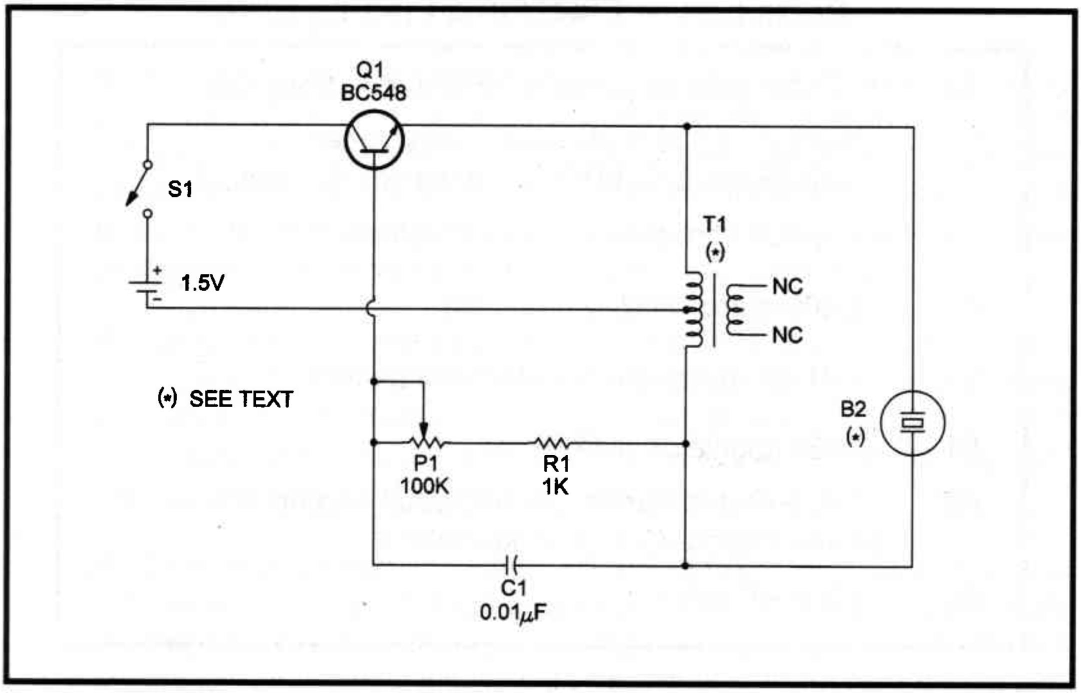
Q1
BC548
S1
1.5V
T1
(*)
NC
NC
(*) SEE TEXT
B2
(*)
P1
100K
R1
1K
C1
0.01μF

Figure 1

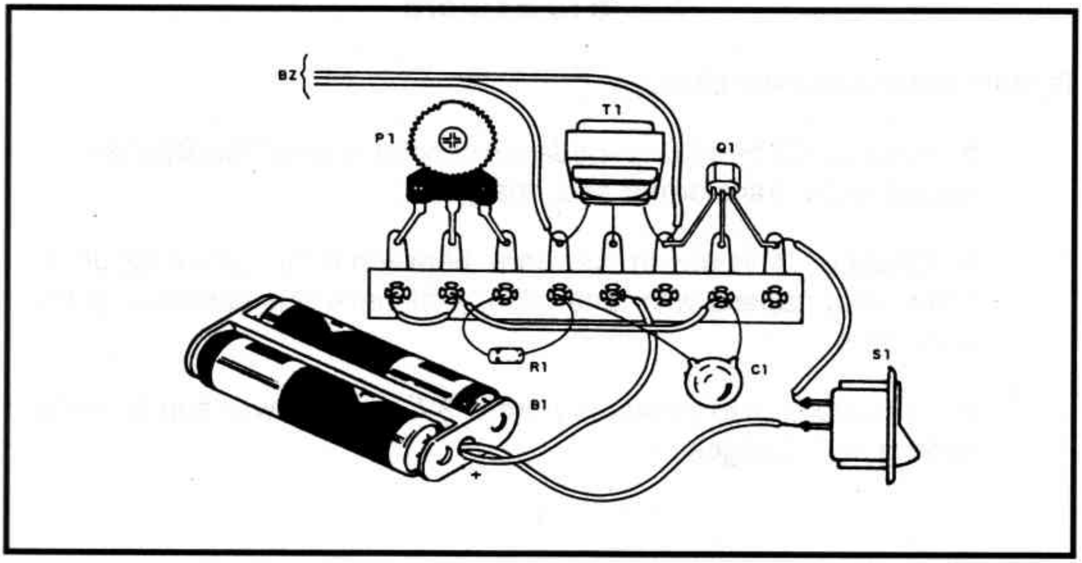
BZ
P1
T1
Q1
R1
C1
B1
S1
+

Figure 2

Parts List — Electronic Fishing Lure

Q1 - BC548 general-purpose NPN silicon transistor
T1 - 500:8 or 1,000:8 transistor transformer (RadioShack 273-1380 is suitable for this task)
P1 - 100,000 ohm trimmer potentiometer
R1 - 1,000 ohm, 1/4W, 5% resistor
C1 - 0.01 uF metal film or ceramic capacitor
S1 - SPST toggle or slide switch
BZ - Piezoelectric transducer or crystal earphone (Radio Shack273-073 or equivalent)
B1 - 1.5 V AA cell

Ideas to Explore

To learn more about the circuit:

- Replace BZ by a low-impedance tweeter or small loudspeaker connected to the transformer secondary coil.
- Change C1 values in the range between 0.001 and 0.22 uF. Observe what hapens with tone pitch. A higher-value capacitor gives a lower tone.
- The circuit also operates from a 3V or 6V power supply without making any changes.

Science and different uses for the device:

- You can make several experiments in animal conditioning, including the ones that live in the water. Try to alter the RC network (P1, R1 and C1) to get tones of different pitch. Try to use diodes, small transformer coils and other passive components in this experiment. Sounds like those produced by songbirds can be generated with a few components arranged in the correct way. Can you discover how to do it?

- Remove the piezoelectric transducer and wire electrodes in its place. The device can be use as a nerv stimulator. Experiments with animals and plants can be made. What physiological functions are affected? Can circadian rhytms or internal biological clocks be affected by continuous exposure to sounds?

- Electrodes placed into water and wired in place of the transducer will produce a current field. You can perform interesting scientific work relating how fish and other water beings react to the current field. Can the current fields affect physiological functions? If you experiment with plants, select a fast growing house plant.

FREE-POWER AM RECEIVER (P,E)

This project uses a strange way to get its power. It uses power "extracted" from the tuned station to give to the amplifier transistor. With a reasonably strong applied modulation envelope, this type of detector produces a strong demodulated output.

Naturally, the closer you are to a strong station, the more current your radio will be able to supply.

For best results, you must do everything possible to deliver a strong signal to the transistor detector. We recommend a good antenna and ground, the latter preferably being made to a water pipe or solid external ground composed of a pipe driven at least 5 feet into moist earth. This is important for ensuring maximum signal pickup.

If you have plenty of space available, the longer the antenna (up to about 100 feet), the better the results.

Figure 1 shows the schematic diagram of the Free-Power AM receiver. Observe that the heart of the circuit is a germanium transistor that works as a detector and audio amplifier.

Components placement on a terminal strip used as a chassis is shown in ***Figure 2.*** The terminal strip can be fixed on an experiment board that can be constructed using some common tools and materials.

Q1 is any germanium transistor, such as GE-2 or 2N107. You can find germanium transistors in old nonworking AM transistor radio receivers. Nonworking AM radios can also supply the variable capacitor and the loopstick (ferrite rod).

Earphone must be a high-impedance crystal type. Low-impedance earphones don't function in this circuit.

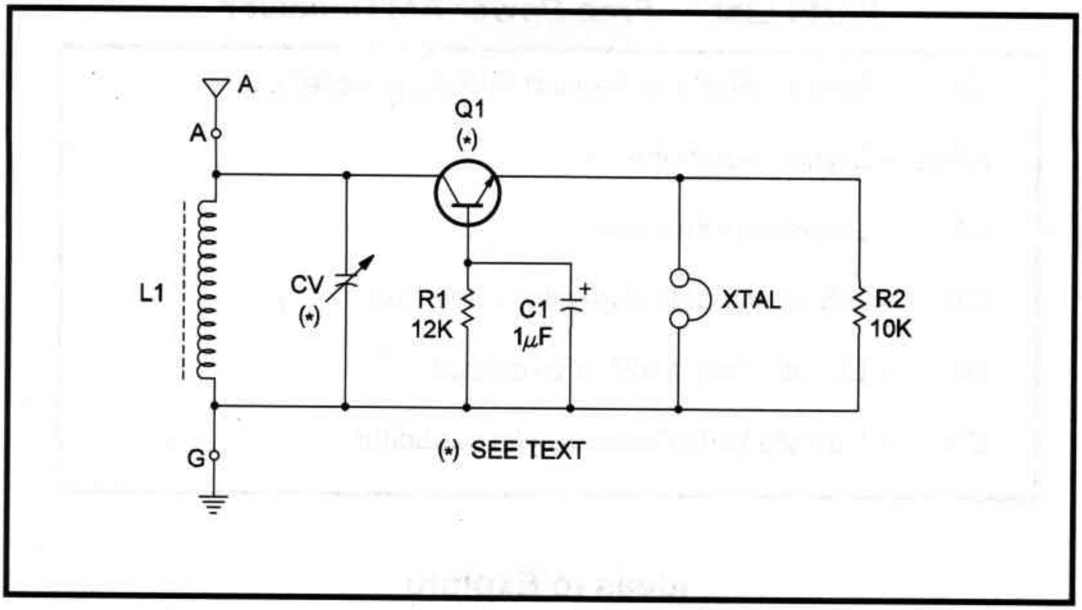
A
A
Q1
(*)
L1
CV
(*)
R1
12K
C1
1μF
+
XTAL
R2
10K
G
(*) SEE TEXT

Figure 1

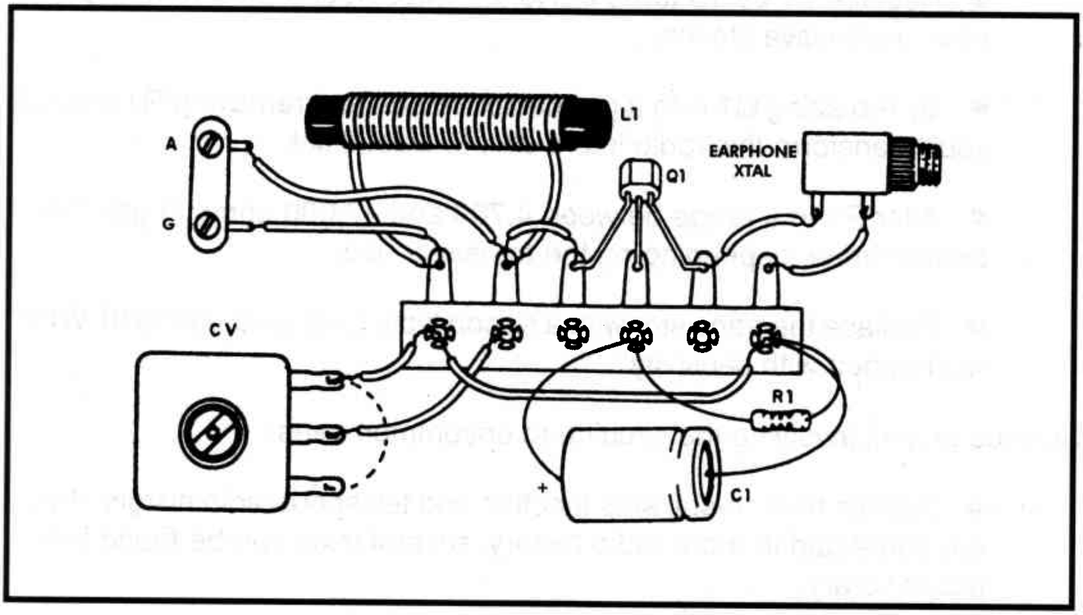
L1
A
G
EARPHONE
XTAL
Q1
CV
R1
+
C1

Figure 2

Parts List — Free-Power AM Receiver

Q1	- NPN or PNP germanium transistor - see text
XTAL	- Crystal earphone
L1	- Loopstick - see text
CV	- 365 pF variable capacitor - see text
R1	- 12,000 ohm, 1/4W, 5% resistor
C1	- 1 uF, 25 WVDC electrolytic capacitor

Ideas to Explore

To learn more about the circuit:

- Replace L1 by 15 turns of No. 28 wire, wound on a ferrite rod, to tune short-wave stations.
- By replacing Q1 with a germanium diode and removing R1 and C1 you'll transform this radio into a simple crystal set.
- Alter R1 in a range between 4,700 and 47,000 ohms to get better performance, depending on the transistor used.
- Replace the transistor with a silicon type, such as the BC548. What will happen with sensivity?

Science project involving the circuit and uncommon uses:

- Explain how crystal sets function and tell about radio history. If you are interested in more radio history, several texts can be found in the public library.
- If you want to learn more about radio receivers, try to find information about heterodyne and super-hereterodyne types.

DUAL-LED FLASHER (E,P)

This circuit uses a one-gate oscillator, with a 4093 CMOS IC, to produce a 50% duty-cycle flasher in two LEDs or a bi-color LED. The circuit can be powered by a 3 to 12-volt power supply and is useful in several experiments and practical applications. You can use it as part of alarms, warning systems, games, toys, and some scientific experiments, as we will describe here.

R2 and R3 values depend upon the power supply voltage, according to the following table:

Power Supply	R2, R3
3 Volts	220 ohms
5 Volts	330 ohms
6 Volts	470 ohms
9 Volts	820 ohms
12 Volts	1,000 ohms

For more output you can use a transistorized output stage driving incandescent lamps. Experiments to observe light effects on animals and plants can be conducted with this circuit.

P1 and C1 determine flash rate. Flash rate can be adjusted by P1, but C1 also can be altered in a large range of values, according to the application you intend for the device.

Schematic diagram for the Dual-Led Flasher is shown in ***Figure 1***. One gate of a 4093 IC is used as oscillator and the other two are used as inverters.

Components placement on a PC board is shown in ***Figure 2***. You can also use a breadboard or solderless board to mount it as an experimental circuit.

Position of the polarized pieces, such as LEDs, eletrolytics and IC, should be observed.

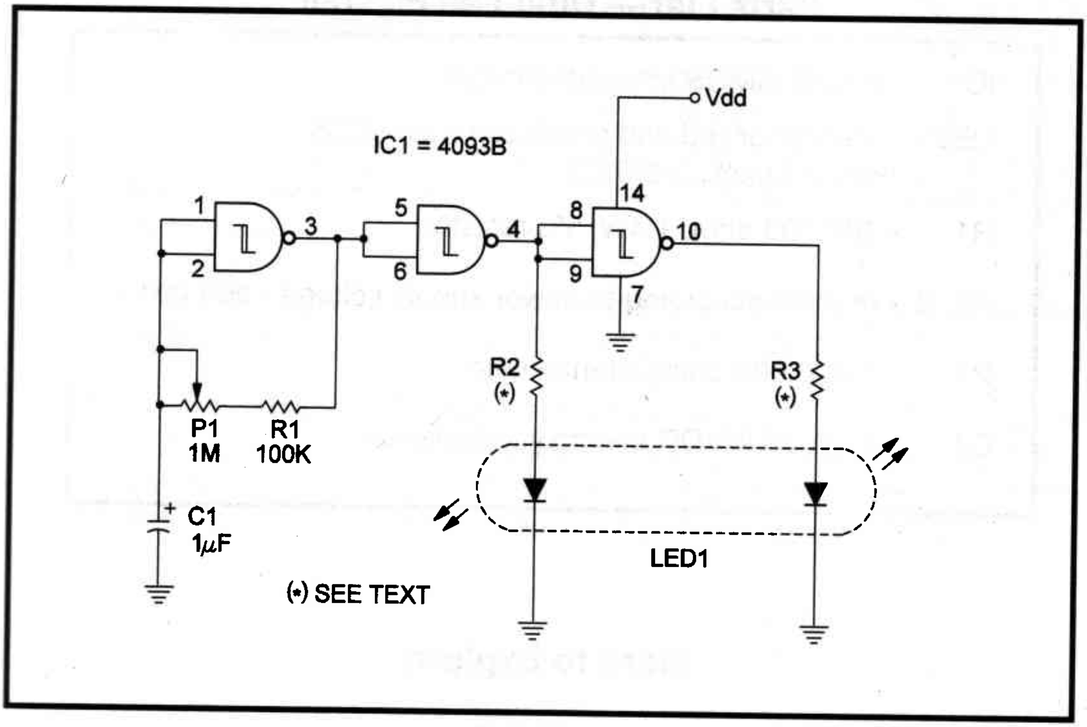
Vdd
IC1 = 4093B
1
2
3
5
6
4
8
9
14
10
7
R2
(*)
R3
(*)
P1
1M
R1
100K
+
C1
1μF
LED1
(*) SEE TEXT

Figure 1

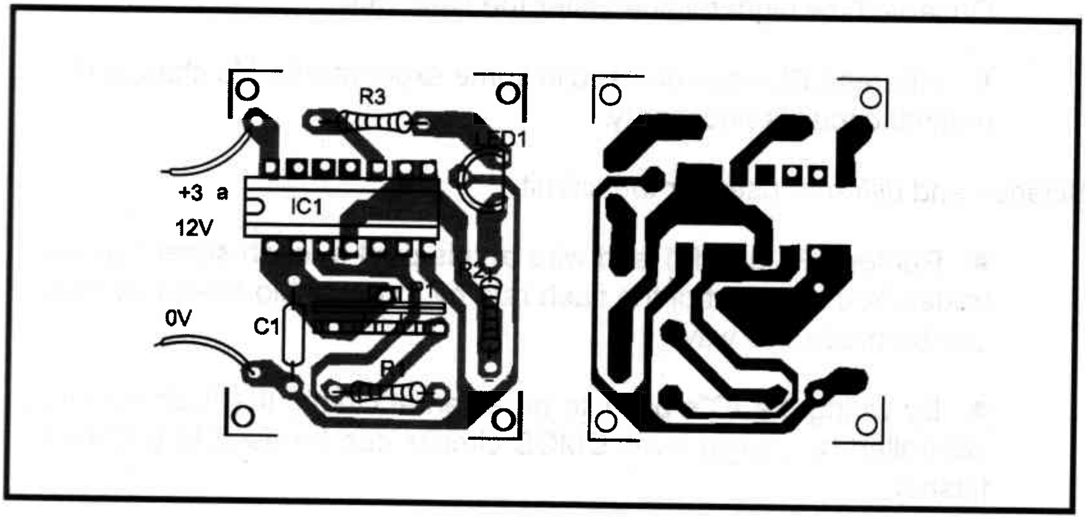
R3
LED1
+3 a
12V
IC1
R2
P1
0V
C1
R1

Figure 2

Parts List — Dual-Led Flasher

IC1	- 4093B CMOS integrated circuit
LED1	- bi-color or red and green common LEDs (Radio Shack 276-012)
R1	- 100,000 ohm, 1/4W, 5% resistor
R2, 3	- resistor according to power supply voltage - see text
P1	- 1,000,000 ohm potentiometer
C1	- 1 uF, 25 WVDC electrolytic capacitor

Ideas to Explore

To learn more about the circuit and devices:

- Replace C1 with capacitors in the range between 1 and 47 uF. Observe how higher values alter the flash rate.

- Infrared LEDs can be used in some experiments. No change in the original circuit is necessary.

Science and different uses for the circuit:

- Remove P1 and R1 and wire at this pont a touch sensor or electrodes. You can control the flash rate by touch. A bio-feedback device can be made this way.

- By wiring the IC”s pin 1 to an external circuit, the flasher can be controlled by a logic level. CMOS circuits can be used to control the flasher.

- Mount the LEDs in place of the lenses of a set of eyeglasses. The flash rate can be adjusted to make an hypnotic or relaxing effect. Combine this with a touch sensor and you have an uncommon bio-feedback device. You can also use this circuit to make experiments in circadian rhythms or internal biological clocks, determining how they can be altered by light pulses. Various theories have been proposed about this subject.

- Psychiatrists are using phototherapy to help some patients. Can you find some application for the circuit in this field?

- P1 also can be replaced by resistive transducers, such as LDRs or NTCs, to give a light- or temperature-dependent flash-rate circuit. Experiments can be made with conductivity of materials by connecting probes in series with R1. Replace P1 with the probes in this case.

6V X 1A POWER SUPPLY (P)

In order to perform experiments and projects involving electricity you will need a source of voltage. Voltage requirements for common projects, such as the ones in this book, depend upon their applications, and typically range from a few milliamperes to one or two amperes.

Instead of using a battery you can use a circuit called a *power supply*.

The power supply we show here can be used to provide 6V x 1A regulated output to the experimental and practical circuits described in this book. The output voltage is fixed by the IC.

Figure 1 shows the schematic diagram for this power supply. IC1 is a voltage regulator, #7806. The "06" indicates that it is a 6-volt regulator.

Figure 2 shows the components layout, using a terminal strip as chassis.

The complete project can be assembled in a small plastic box. The integrated circuit should be mounted on a heatsink. You can also add an LED in series with a 470 ohm resistor, wired in the output of the circuit to indicate that it is on.

External circuits can be connected to the power supply with alligator clips.

Position of the polarized pieces, such as diodes, the IC, and electrolytic capacitors, must be observed.

The transformer isn't a critical part of this project. Types with a secondary coil ranging from 7.5 to 12 volts and currents between 500 mA and 1A can be used.

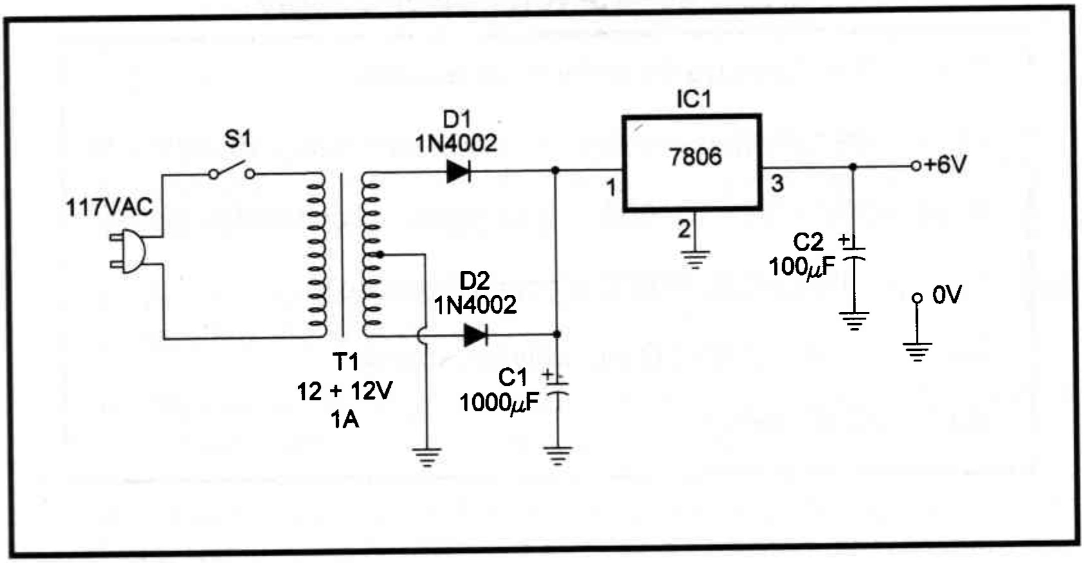
IC1
7806
D1
1N4002
S1
117VAC
1
3
2
+6V
C2
100µF
0V
D2
1N4002
T1
12 + 12V
1A
C1
1000µF

Figure 1

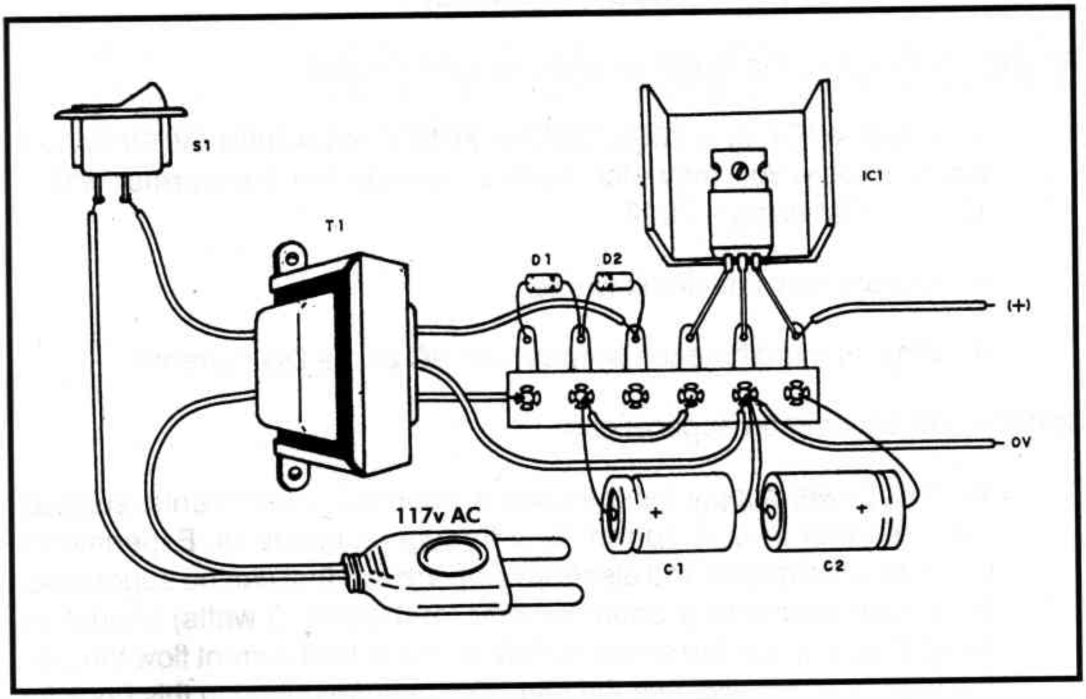
S1
IC1
T1
D1
D2
(+)
0V
117v AC
C1
C2

Figure 2

Parts List — 6-Volt Power Supply

IC1 - 7806 positive three-terminal regulator
T1 - Transformer: primary 117 VAC ; secondary 12+12V x 1A
D1, 2 - 50V x 1A - 1N4002 or equivalent silicon rectifier diodes
C1 - 1,000 uF, 25 WVDC electrolytic capacitor
C2 - 10 uF, 12 WVDC electrolytic capacitor
S1 - SPST switch

Ideas to Explore

To learn more about the circuit or alter the performance:

- Replace IC1 by a 7805, 7809 or 7812 if you want a 5V, 9V or 12V power supply. You may also have to change the transformer with a 117:15+15V using a 7812.
- Explain how the circuit works.
- What is the difference between an AC and a DC current?

Science and uncommon applications:

- This power supply can be used in chemical experiments, such as the ones that involve current flow through substances. Experiments involving electrolysis and electroplating are two that can be suggested. A variable wound-wire potentiometer (100 ohms, 5 watts) should be wired in series with the power supply output to limit current flow through the load. You can also use the Light Dimmer described in this book for this task.

- 6V, 200 mA small incandescent lamps can be supported by this power supply. These lamps are excellent light sources to work with microscopes.

- Small motors and lamps in experiments involving physics, mechanics and robotics can be powered from this power supply with advantages. You don't have to spend a lot of money on cells or batteries.

- All circuits described in this book which require a 6V supply can be powered by this power supply.

- The circuit can be used as an excellent power adapter for your 6V equipment, such as small radios, calculators, etc.

- 6V lamps can be used in experiments involving plant growth. Select a fast growing house plant and use a darkroom or cardboard box to cover them. Lamps with colored filters can be used in an experiment involving growth and development.

- Circadian rhythms can also be studied using animals or plants. Lamps can be powered on at night and off at day. What physiological functions could be affected?

INSECT REPELLENT (P)

Some continuous sounds can repel insects (and also attract). The frequency and intensity depends upon the application and type of insect and can be found through experimentation.

The circuit shown here generates a continuous sound that can be used to repel (or attract) some types of insects, or in scientific experiments involving animals.

Our insect repellent can be powered by 9V batteries, two AA cells or four AA cells, and its low current requirements will extend the life of those cells or batteries.

Figure 1 shows the schematic diagram of the Insect Repellent. The heart of the circuit is a 7555 IC, a CMOS timer wired as an audio oscillator and driving a piezoelectric transducer.

The components placement on a homemade printed-circuit board is shown in ***Figure 2***. Exact placement is not that critical.

All the components and the power supply can be housed in a small plastic box. Transducer BZ is a crystal earpiece or a piezoelectric transducer, such as the Radio Shack 273-073.

Position of the polarized pieces, such as C2 and the power supply, must be carefully observed.

Using the insect repellent is very easy. You only need to adjust the trimmer potentiometer P1 to produce a sound with the same pitch as the insect you intend to repel. Experimentation should be made until you find the best sound to repel a specific insect.

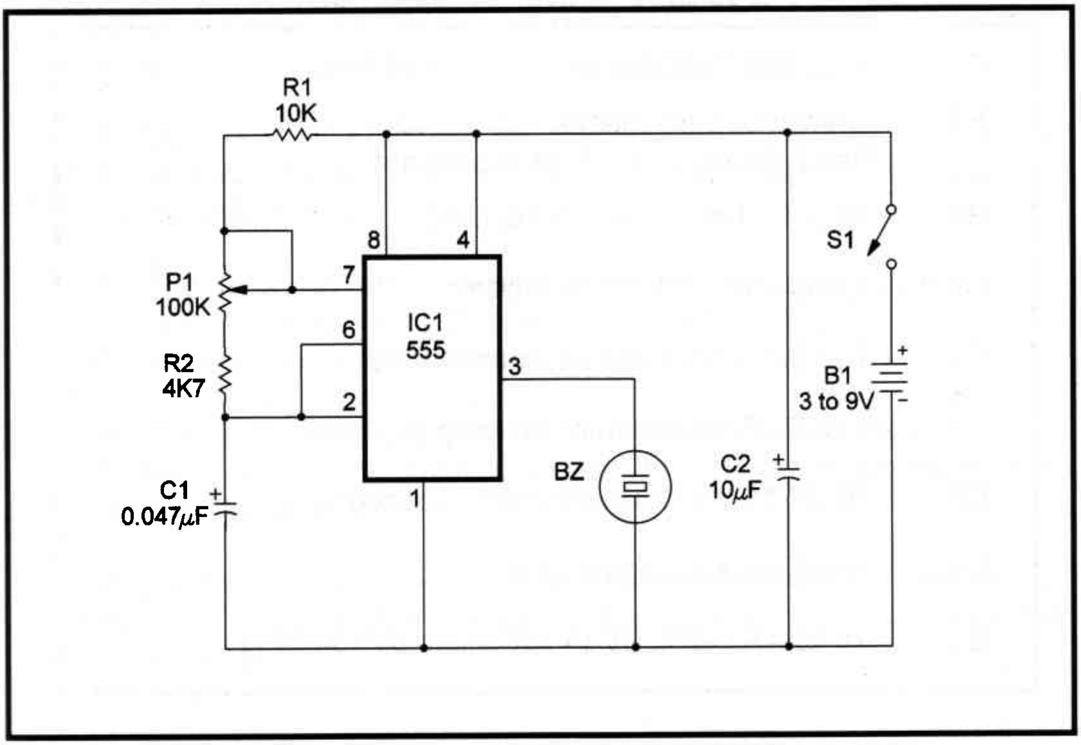
R1
10K
P1
100K
R2
4K7
C1
0.047μF
8
4
7
6
2
3
1
IC1
555
BZ
C2
10μF
S1
B1
3 to 9V
+
-

Figure 1

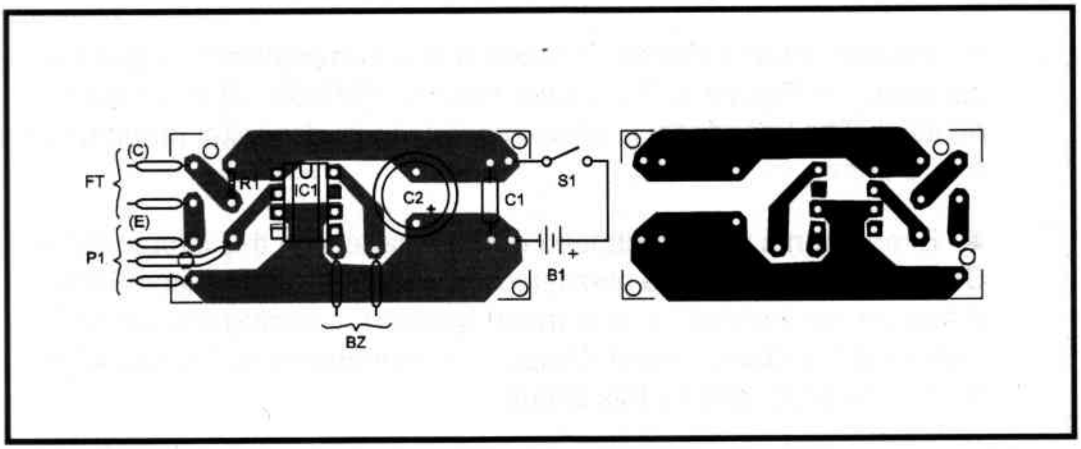
(C)
FT
(E)
P1
R1
IC1
C2
C1
S1
B1
BZ

Figure 2

Parts List — Insect Repellent

C1	- TLC7555 CMOS integrated-circuit timer
BZ	- Crystal or piezoelectric transducer Radio Shack 273-073 or equivalent
R1	- 10,000 ohm, 1/4W, 5% resistor
R2	- 4,700 ohm, 1/4W, 5% resistor
P1	- 100,000 ohm trimmer potentiometer
C1	- 0.047 uF metal film or ceramic capacitor
C2	- 10 uF, 12 WVDC electrolytic capacitor
S1	- SPST slide or toggle switch
B1	- 3V (2 AA cells), 6V (4 AA cells), 9V (battery)

Ideas to Explore

To learn more about the circuit or get better performance:

- Powerful output can be obtained with a transistorized output stage as shown in ***Figure 3***. The power transistor should be mounted on a heatsink. The transducer is a low-impedance loudspeaker ranging from 4 to 8 ohms.

- Experiments with ultrasound can be made with this circuit: reduce C1 to 1200 pF and use a piezoelectric tweeter as transducer. Remove the small transformer used in these tweeters and plug the transducer directly to the circuit output. Outputs in frequencies as high as 40,000 Hz can be produced by this circuit.

- Explain how sounds can be produced by electronic transducers such as the one used in this project.

- What are ultrasounds?

Science and different applications:

- Of course, you can test the real effects of this device on several kinds of insects. An attractive biological experiment can be conducted based on this circuit.

- Effects of sounds on plant growth or development can be conducted using this circuit. Can sounds (or ultrasounds) alter the circadian rhythms of plants and insects?

- The continuous tone produced by this circuit can be used in alarms or warning systems. Mice and other small animals are repelled by powerful ultrasounds.

- Replace R2 by an LDR. The circuit now becomes a light-dependent oscillator. Experiments with light and sounds can now be conducted using the device.

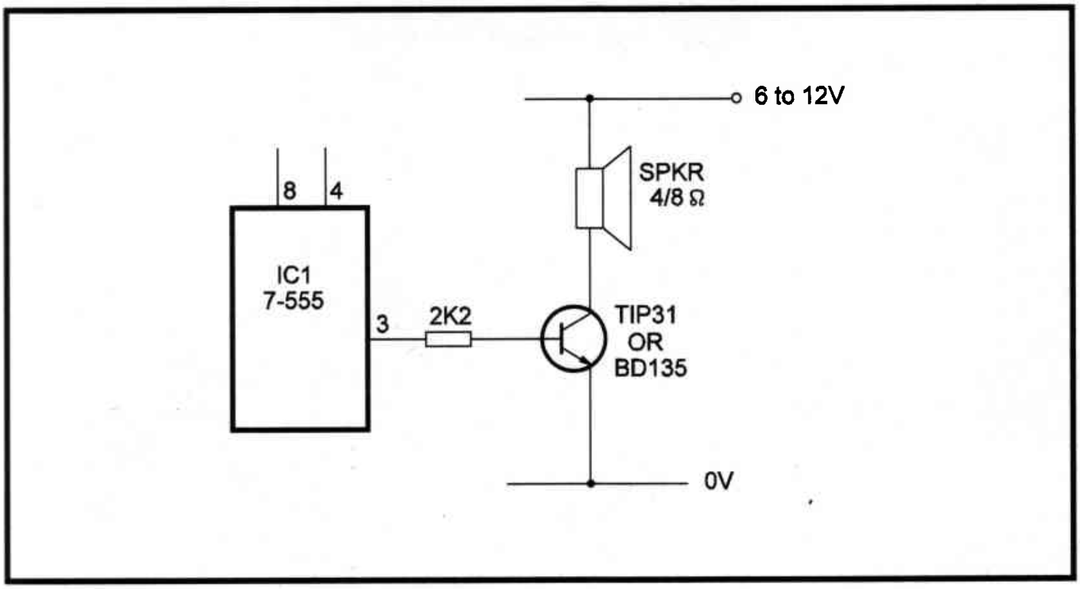

Figure 3

ULTRASONIC GENERATOR (E/P)

Dogs, mice, bats and other animals can hear sounds with frequencies up to 40,000 Hz. There are several kinds of insects that can also hear or react to these sounds. The circuit we propose here produces a continuous sound emission in frequencies above the human limit, or a range between 18,000 and 40,000 Hz. These sounds are called ultrasounds and can be used in several experimental and practical applications.

The device we describe here can be used to scare dogs and other animals, in biological experiments, and many other interesting applications.

The recommended piezoelectric transducer has its maximum output power between 700 and 3,000 Hz, but it also will operate at higher frequencies emitting less power.

The recommended power supply is four AA cells or a 9-volt battery.

Our project runs between approximately 18,000 and 40,000 Hz, but you can easily alter this range by changing C1, in a range between 470 pF and 0.001 uF. Frequency can be set by P1 in the range determined by C1.

Remember that the 4093B IC will oscillate in frequencies up to 500 kHz.

The complete circuit of the Ultrasonic Generator, using the four gates of a 4093B integrated circuit, is shown in ***Figure 1***.

Components placement on a printed-circuit board is shown in ***Figure 2***.

The circuit can be housed in a small plastic box and the transducer will be fixed in the front panel.

Take care with the position of the polarized parts, such as the transistor, electrolytic capacitor and power supply.

For continuous operation, Q1 should be mounted on a heatsink.

The transformer is not a critical part. Any transformer with a secondary coil ranging from 100 to 500 mA can be used in this project.

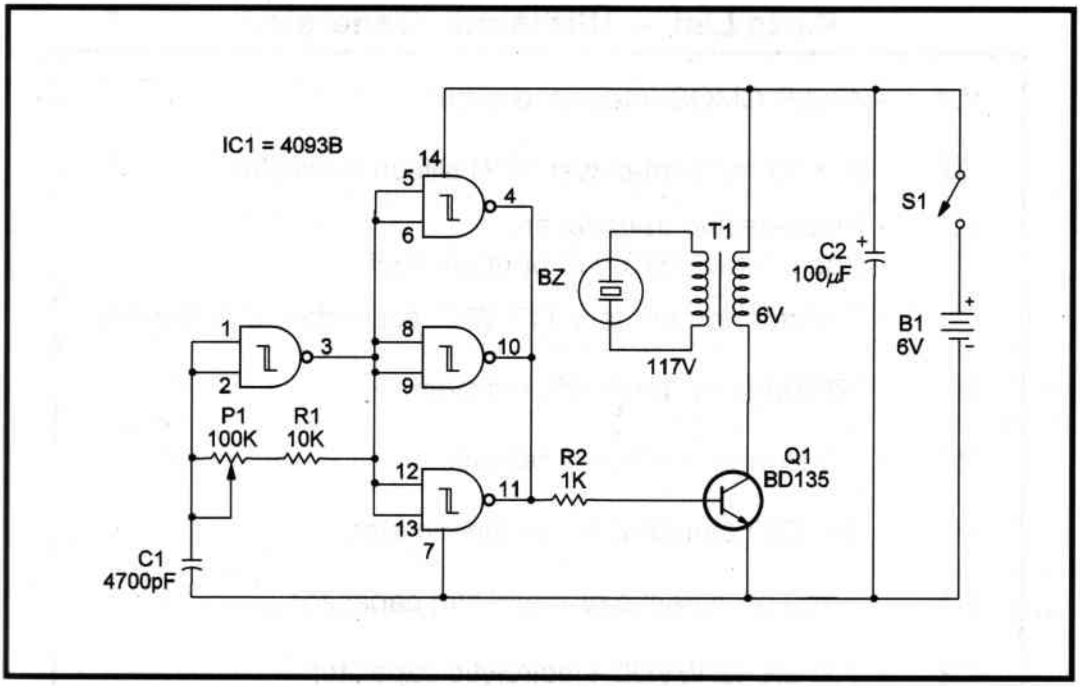
IC1 = 4093B
14
5
4
6
1
2
3
8
9
10
12
13
11
7
P1
100K
R1
10K
C1
4700pF
BZ
T1
117V
6V
C2
100µF
R2
1K
Q1
BD135
S1
B1
6V

Figure 1

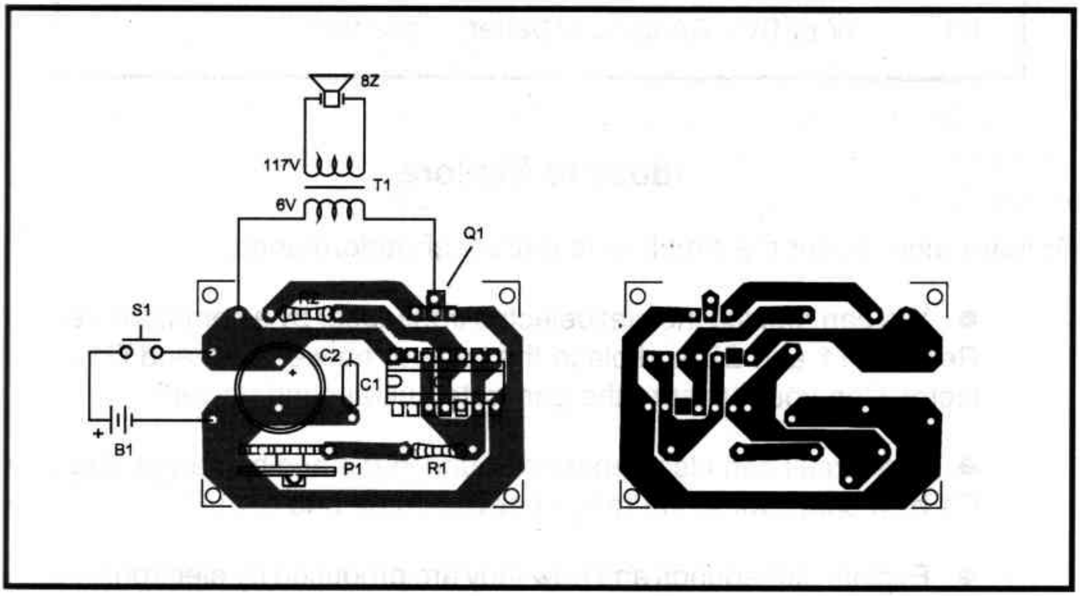
BZ
117V
T1
6V
Q1
S1
B1
R2
C2
C1
P1
R1

Figure 2

Parts List — Ultrasonic Generator

IC1	- 4093B CMOS Integrated circuit
Q1	- BD135 medium-power NPN silicon transistor
BZ	- Piezoelectric transducer Radio Shack 273-090 or equivalent
T1	- Transformer: primary 117 VAC; secondary 6Vx100 mA
R1	- 10,000 ohm, 1/4W, 5% resistor
R2	- 1,000 ohm, 1/4W, 5% resistor
P1	- 100,000 ohm trimmer potentiometer
C1	- 4,700 pF ceramic or metal film capacitor
C2	- 100 uF, 12 WVDC electrolytic capacitor
S1	- SPST toggle or slide switch
B1	- 6V or 9V - AA cells or battery - see text

Ideas to Explore

To learn more about the circuit or to get better performance:

- You can replace the piezoelectric transducer by a common tweeter. Remove T1 and BZ and place the tweeter between S1 and Q1's collector. Can you measure the generated ultrasound power?

- The circuit can also generate sound in the audible range. Replace C1 by a capacitor in the range between 0.02 and 0.1 uF.

- Explain ultrasounds and how they are produced by electronic transducers.

- What is a piezoelectric material?

Science and different applications:

- Several experiments can be made using ultrasounds. You can experiment with dog conditioning. Using this oscillator you can call a dog using an inaudible sound! (inaudible to humans)

- Choose a capacitor between 0.01 and 0.047 uF to produce sounds both in the audio range and ultrasound range. You can now conduct experiments to determine the exact upper-limit frequency heard by humans. You also can conduct experiments with animals. It will be important to have a frequency counter on hand to do this work.

- Replace BZ by electrodes. The circuit will generate a high-voltage output that can be used in nerve stimulation. Use a 10,000 ohm potentiometer to adjust output voltage at this point. Pulses can reach peaks as high as 400 volts.

- To get pulses (for nerve stimulation and other experiments) you only have to replace C1 with a 1 uF capacitor and P1 with a 1,000,000 ohm potentiometer. Pulses in the rate range between 0.1 per second to 10 per second can be generated with this change.

- This circuit can also be used as a fluorescent lamp inverter. A common fluorescent lamp (4 to 20 watts) in place of the transducer will glow, even one that is weak enough to not function on the AC power line.

DC LAMP DIMMER (E,P)

This circuit can be used to control the amount of current in an incandescent lamp and also a DC motor. If you change the amount of current through a lamp you change its brightness. And if you change the amount of current through a DC motor you change its speed. Input voltage can range from 6 to 12 volts and current drain in the output is up to 2A. You can also use this circuit in the output of a 6V or 12V fixed power supply, converting it into a variable 0 to 6 volt or 0 to 12 volt supply.

The power transistor should be mounted on a heatsink. The circuit can be housed in a small plastic box and the heatsink with the transistor fixed on the outside.

The circuit can be used to control brightness of a car's panel lamps and in robotics to control DC motor speed along a wide range of values. Also, several experiments involving current flow control can be performed using this circuit.

The circuit acts as a variable resistor or rheostat with the principal current flowing through the transistor and the control current flowing through the potentiometer. Current through the potentiometer is only a few milliamperes.

The schematic diagram of the DC dimmer is shown in ***Figure 1***. Q1 is an NPN power transistor and the principal component used in this project.

Components layout is shown in ***Figure 2***. The transistor does not show on the heatsink, however.

Equivalent transistors can be used, but take care with the maximum current they can control. Types such as TIP31 and TIP41 can be used to control lamps up to 2 amperes. Also, a further modification in R1 value must be made to ensure that, when P1 is set near its minimum value, that current is flowing through the lamp.

If equivalent transistors are used, take care with their terminal positions, as they can be different.

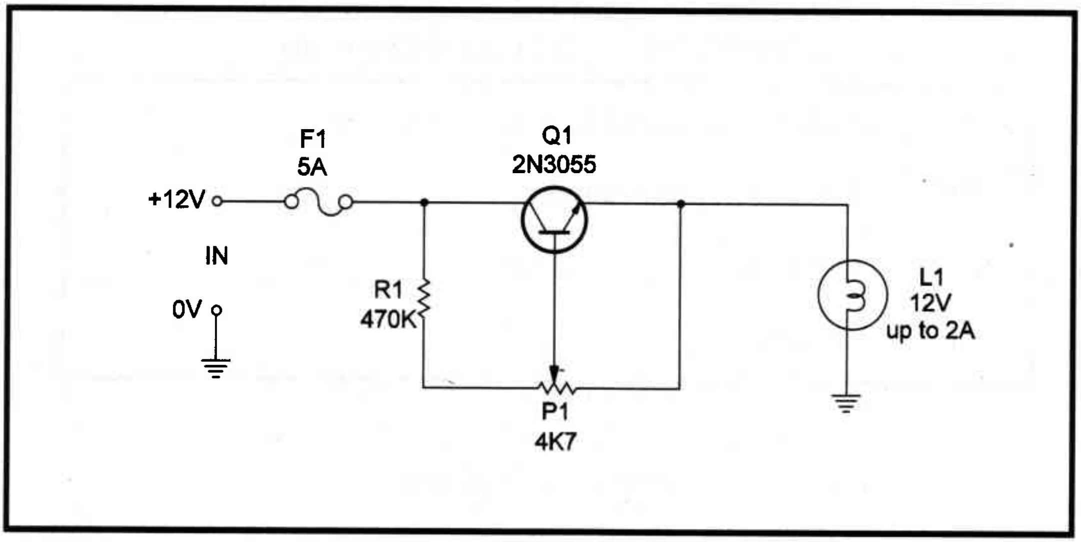
F1
5A
Q1
2N3055
+12V
IN
0V
R1
470K
P1
4K7
L1
12V
up to 2A

Figure 1

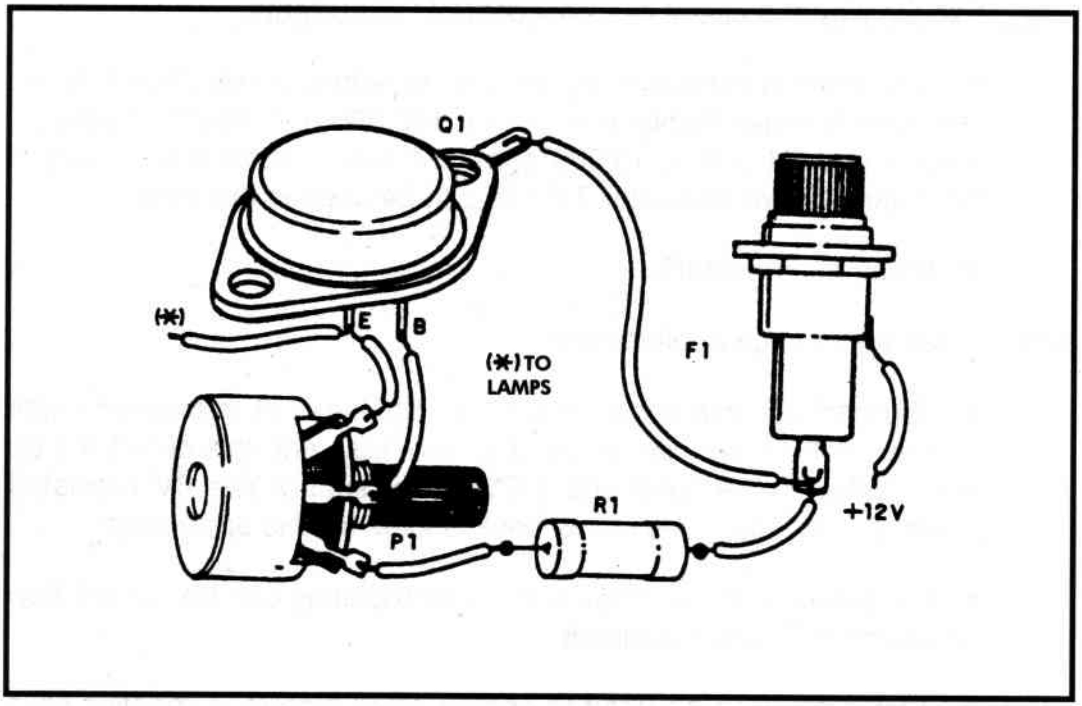
Q1
(✱)
E
B
(✱) TO
LAMPS
F1
R1
+12V
P1

Figure 2

Parts List — DC Lamp Dimmer

Q1	- 2N3055 Silicon NPN power transistor
P1	- 4,700 ohm potentiometer
R1	- 470 ohm, 2W, 5% resistor
F1	- 5A fuse

Ideas to Explore

To learn more about the circuit or to get better performance:

- You can use a PNP power transistor, inverting input polarization. Explain why this circuit can only control DC currents.

- A Darlington transistor can be used to reduce current flow through the potentiometer. Replace P1 with a 100,000 ohm common potentiometer and R1 with a 10,000 ohm, 1/4 watt resistor when using a Darlington power transistor. TIP110 can be used in this case.

- What is a rheostat?

Science and uncommon applications:

- You can use this circuit to control brightness of microscope light sources. A 6V lamp can be used for that task and this circuit will be wired between the lamp and a 6V power supply. The 6V regulated power supply described in this book is ideal for this application.

- Experiments in electrolysis and electroplating can be current-flow controlled by using this circuit.

- This circuit can be used to control small motors in robotics or in physics experiments.

- The amount of current in current fields can be controlled by using this circuit and a DC power supply. Experiments involving electrotropism can be attempted. Try to put plants within a coil to observe the influence of the magnetic field on their growth and development. Remember that there is a magnetic field around a coil of wire that is carrying a current.

Figure 3 shows how this experiment can be arranged.

The coil is formed by 10 to 50 turns of common No.18 to 22 wire around a cardboard box or wood form. An ammeter in series with the coil circuit can be used to control the amount of current used in the experiment.

Remember that too much current will cause the coil and the transistor to overheat. It is a good idea to add a lamp in series with the coil circuit. The lamp will add resistance to the coil circuit, reducing current flow to a safe level.

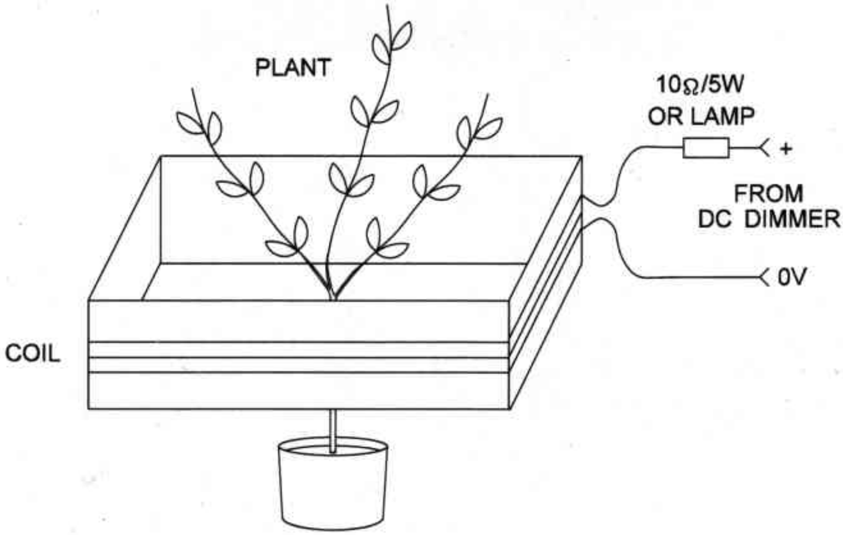

Figure 3

AUDIO OSCILLATOR (E)

This low-power experimental oscillator can generate audible signals in the range between 100 and 2,000 Hz, driving a small loudspeaker or a low-impedance earphone. The circuit can be powered by 2 two or four AA cells or a fixed 6 volt power supply. Current requirements for this circuit depend upon the power supply voltage and the loudspeaker impedance and range typically from 10 to 300 mA.

Potentiometer P1 adjusts the operating frequency and can be set within a large range of values. Potentiometers up to 1,000,000 ohms can be used, changing the frequency range lower limit to about 10 Hz.

C1 can also be altered, and values between 0.01 and 0.22 uF are suitable for experimentation. Large C1 values will produce frequencies in the lower part of the range.

The circuit can be used as part of alarms, games, toys and to learn more about transistorized oscillators.

Figure 1 shows the schematic diagram of the Audio-Oscillator. Q1 and Q2 form a directed-coupled amplifier and R1/C1 is the closed loop that takes the signal back from the output to the input.

Figure 2 shows the layout of the components, using a terminal strip as chassis. All the components and power supply can be housed in a small plastic box.

The key is a telegraph homemade-type switch, but you can also use an SPST toggle or slide switch for continuous operation. A key is recommended if you want to use the oscillator as a Morse Code sender in demonstrations or practice. You can wire the points A and B to the contacts of a relay to use the circuit as part of an alarm.

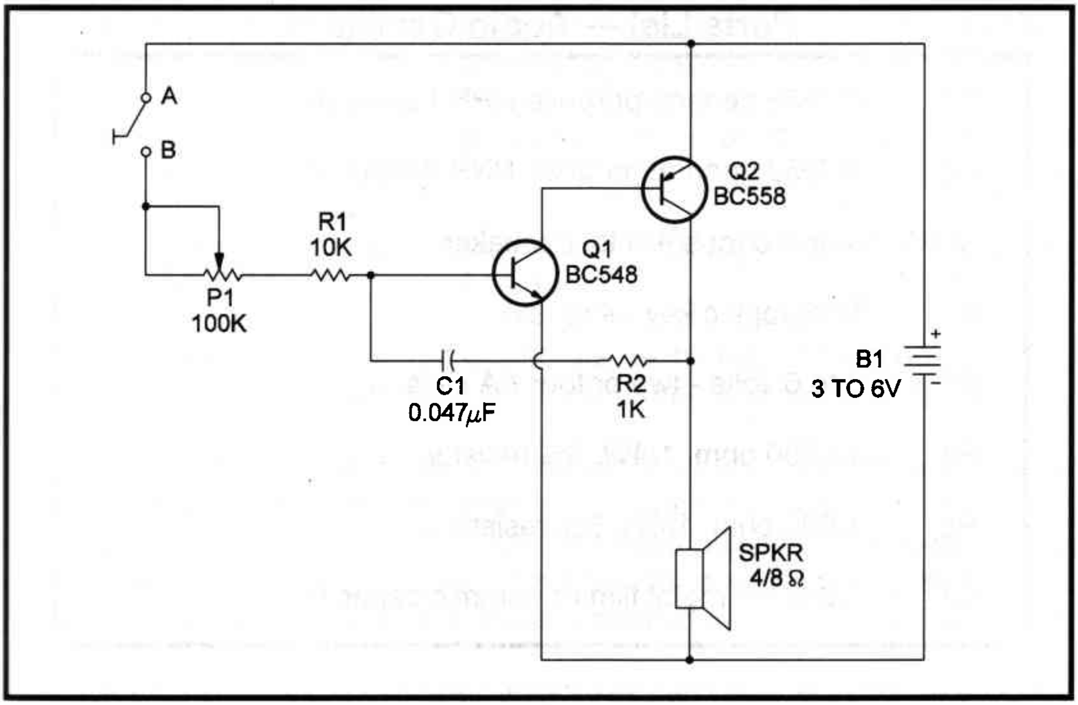
A
B
R1
10K
P1
100K
Q1
BC548
Q2
BC558
C1
0.047µF
R2
1K
B1
3 TO 6V
+
-
SPKR
4/8 Ω

Figure 1

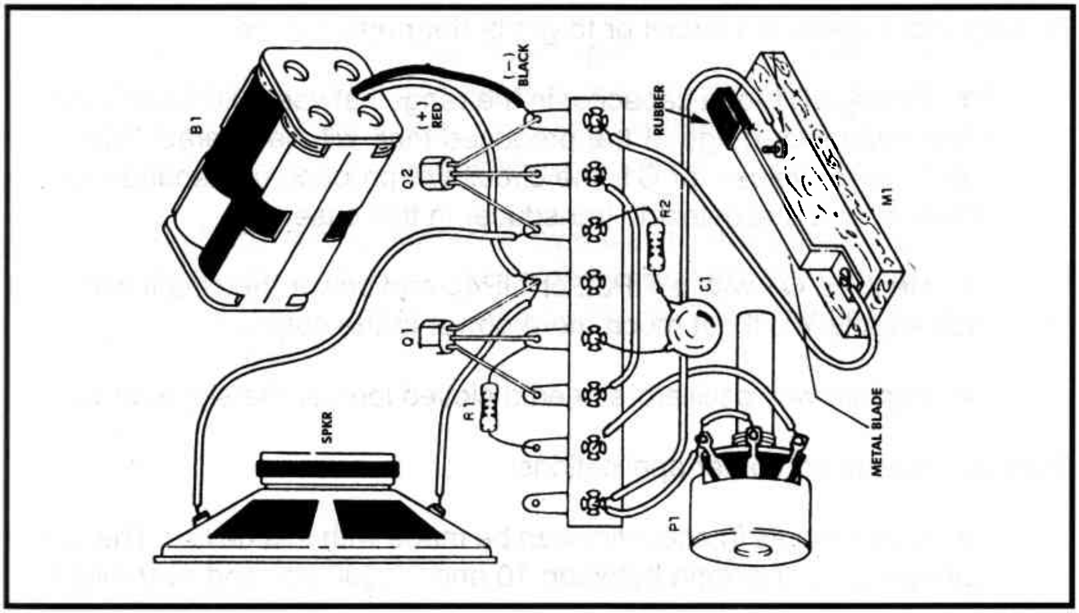
B1
(+)
RED
(−)
BLACK
Q2
RUBBER
R2
M1
C1
Q1
R1
SPKR
METAL BLADE
P1

Figure 2

Parts List — Audio Oscillator

Q1 - BC548 general purpose NPN transistor

Q2 - BC558 general purpose PNP transistor

SPKR - 4 or 8 ohm small loudspeaker

K - Telegraphic key - see text

B1 - 3 to 6 volts - two or four AA cells

R1 - 10,000 ohm, 1/4W, 5% resistor

R2 - 1,000 ohm, 1/4W, 5% resistor

C1 - 0.047 uF metal film or ceramic capacitor

Ideas to Explore

To learn more about the circuit or to get better performance:

- Change C1 for a capacitor in the range between 0.01 and 0.22 uF. The frequency range of the produced tone will be altered. With 0.01 uF or lower values for C1, the circuit will produce ultrasounds. Use a tweeter or a piezoelectric transducer in this case.

- Replace Q1 with a TIP32 or TIP42 and power the circuit with a 12 volt supply. You'll get much more power in the output.

- Explain why oscillators need a closed loop to the signal to work.

Science projects and other applications:

- Experiments in acoustics can be made with this device. The circuit can produce any tone between 10 and 20,000 Hz, and even higher.

- A telegraph station can be simulated with this circuit. You can find Morse Code in any book about telegraphy.

- By replacing K with a reed swich wired to points A and B, the circuit will operate as an alarm.

- Experiments involving conductivity of materials can be made with probes plugged to points A and B.

- Connect resistive transducers as LDRs and NTCs at points A and B to get a light-dependent or temperature-dependent oscillator.

- Using two metal rods as sensors, the circuit can be use as a lie detector or bio-feedback device. Replace C1 with a 0.22 or 0.47 uF capacitor and adjust P1 to get separated pulses at the speaker. The pulse rate will be altered when the skin conductivity of the interrogated person changes while in a state of strain due to a lie or stress.

- A 117V—6V x 250 mA transformer can replace SPKR to produce high-voltage pulses in nerve stimulation experiments. The generated pulses can reach voltages as high as 400Vpp.

- Experiments in animal physiology can be conducted using sounds or high voltage as described in other projects. Growth of plants submitted to high-power sounds or high-voltage fields can be observed using this circuit with practical experiments.

- AC fields can be generated by replacing the speaker with a 20-turn coil. The plants or animals can be positioned within the coil's electric field. See the same experiment using the DC Dimmer for coil and experiment details.

AUDIO SWITCH (E/P)

This audio controlled relay has a large number of applications and is interesting to the experimenter who wishes to explore the working universe of relays and audio or sound controlled circuits.

An audio signal in the amplitude range between 1 Vpp and 5 Vpp triggers the relay, powering on the device, wired between points A and B.

Relay value is determined by power supply voltage. For a 12 volt power supply you must use a 12 volt relay. Current drain depends upon the power supply voltage and also the relay used. Using a 12V x 38 mA (Radio Shack 275-248) relay, the current drain is low enough to allow the use of common cells in the power supply.

C1 can range between 0.01 and 1.0 uF, and determines the sensitivity of the circuit in the audio range. Large values give more sensitivity at low frequencies.

This circuit has a high-impedance input and needs a strong audio signal to operate.

Figure 1 shows the schematic diagram of the Audio Switch. Observe that the heart of the circuit is a common general-purpose NPN transistor that acts as a DC amplifier.

In ***Figure 2*** we show layout of components for this project. The components can be soldered on a terminal strip used as chassis. The terminal strip can be fixed on a wooden board.

Position of the polarized components, such as the diodes, power supply and capacitor C2, must be observed.

To use this audio relay you have to wire ponts A and B to the output of an audio amplifier (minimum power required is 5 watts) and adjust the volume control for best operation.

A small transistor output transformer should be used to operate the relay with low-impedance signals, such as the ones from the output of audio amplifiers.

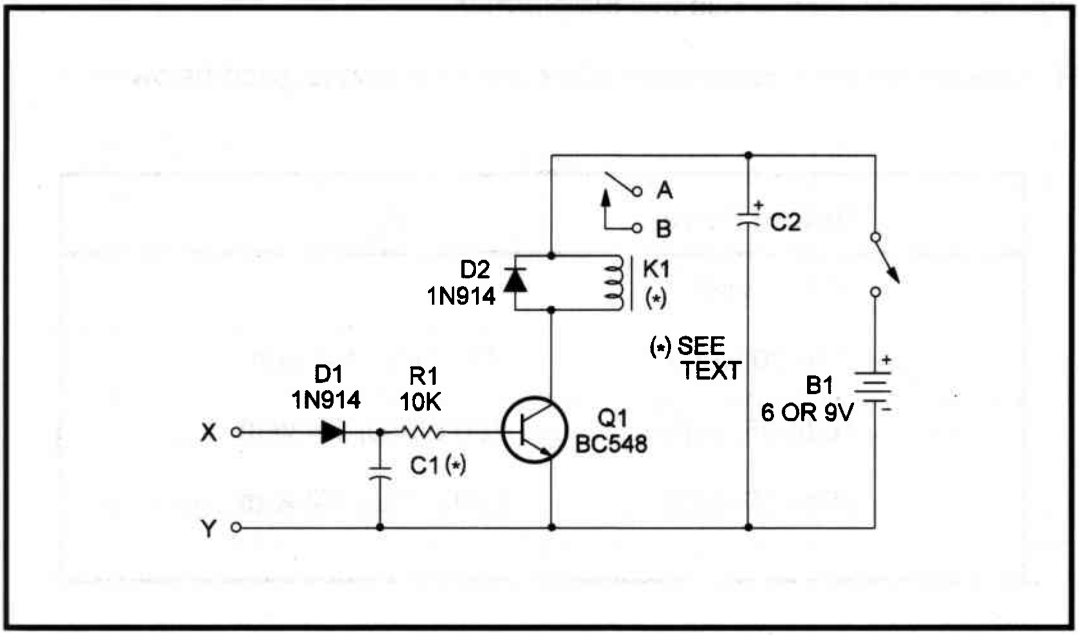
A
B
+
C2
D2
1N914
K1
(*)
(*) SEE TEXT
D1
1N914
R1
10K
X
C1 (*)
Q1
BC548
+
B1
6 OR 9V
-
Y

Figure 1

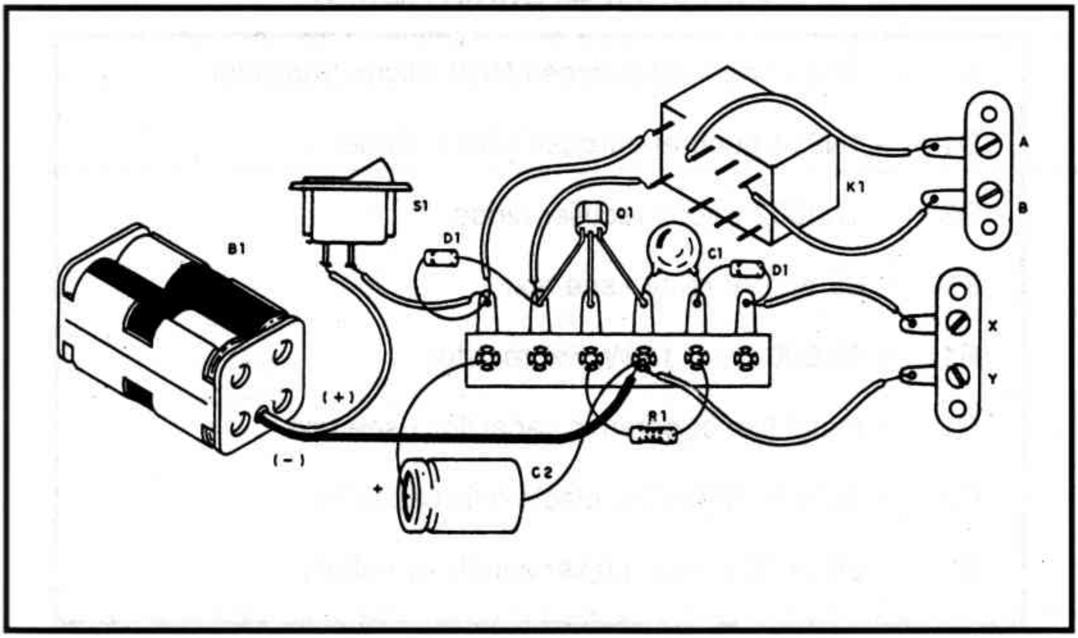
S1
B1
D1
Q1
C1
D1
K1
A
B
X
Y
(+)
(-)
R1
C2
+

Figure 2

Figure 3 shows how to use this transformer.

R_x depends on the output power of the amplifier and is given below:

Output Power	R_x
0 to 1 watt	-
1 to 10 watts	47 ohms, 1/4 watt
10 to 25 watts	100 ohms, 1/4 watt
25 to 50 watts	220 ohms, 1/2 watt

Parts List — Audio Switch

Q1	- BC548 general-purpose NPN silicon transistor
D1	- 1N914 general-purpose silicon diode
D2	- 1N4002 silicon rectifier diode
K1	- 6V or 12V relay - see text
R1	- 10,000 ohm, 1/4W, 5% resistor
C1	- metal film or ceramic capacitor - see text
C2	- 100 uF, 12 WVDC electrolytic capacitor
B1	- 6V or 12V cells, power supply or battery

Ideas to Explore

To learn more about the circuit and devices or to get better performance:

- Replace Q1 with a Darlington transistor (such as a BC517) and resistor R1 with 100,000 ohms. The circuit will increase in sensitivity.
- Explain how C1 functions in this circuit.

Science and different applications:

- Connect the circuit to an audio amplifier output, with a microphone plugged to the amplifier input. You'll have a sound-operated switch.
- This circuit with a microphone and an audio amplifier can be used to close a trap with sound activation or to activate a recording with sounds.

You can also use it to photograph an explosion or the crash of an object by triggering a flash with the produced sound.

- The audio output of a multimedia system in a computer can be used to control external devices with this circuit.

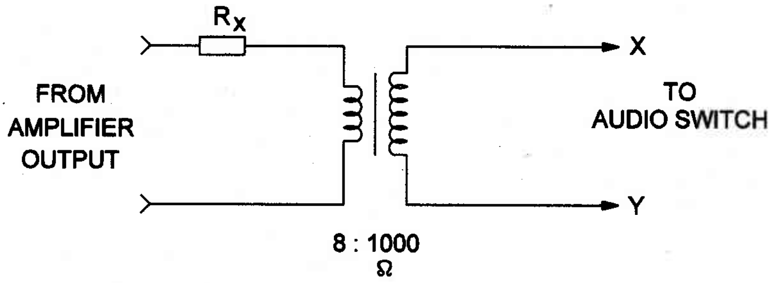

Figure 3

6 VOLTS FROM 12V BATTERIES (P)

With this adapter you can power 6 volt (up to 1A) devices, such as recorders, CD players, calculators, radios, battery chargers, etc., from your car battery. The current adapter is plugged into the vehicle lighter socket and is small enough to be housed in a very compact plastic box.

Only four components are used. The IC determines the output voltage and should be mounted on a heatsink. Capacitors are polarized components, so you must observe their position when mounting. The fuse is very important in order to protect the device and car wiring against shorts.

Output varies according to the device you are applying power to. DC extension cables with plugs to fit the device to be powered should be used. Take care with the output poles. Check if your device has a positive or negative pole in the central terminal.

The schematic diagram of the 12V to 6V x 1A converter is shown in ***Figure 1***. The principal component is IC1, a 6V voltage regulator that operates with input voltages ranging from 8 to 35 volts.

Components layout is shown in ***Figure 2***. All the components can easily be housed in a plastic box.

The integrated circuit can be changed to alter the output voltage. You can use a 7809 for 9V x 1A output or a 7805 for 5V x 1A output.

It is very important that no parts of the components used in this circuit touch one another. The IC has short-circuit and thermal protection.

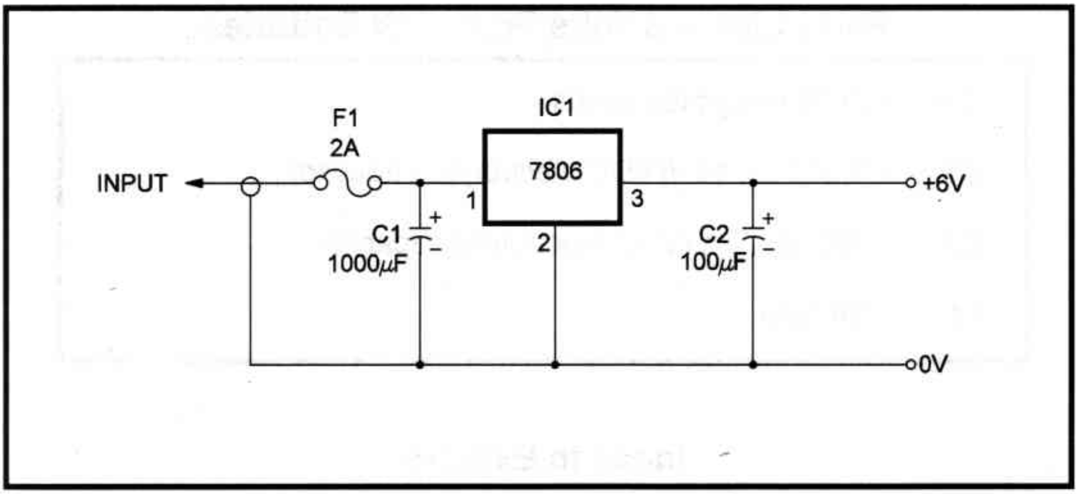
F1
2A
IC1
7806
INPUT
1
3
2
+6V
C1
1000μF
C2
100μF
+
−
0V

Figure 1

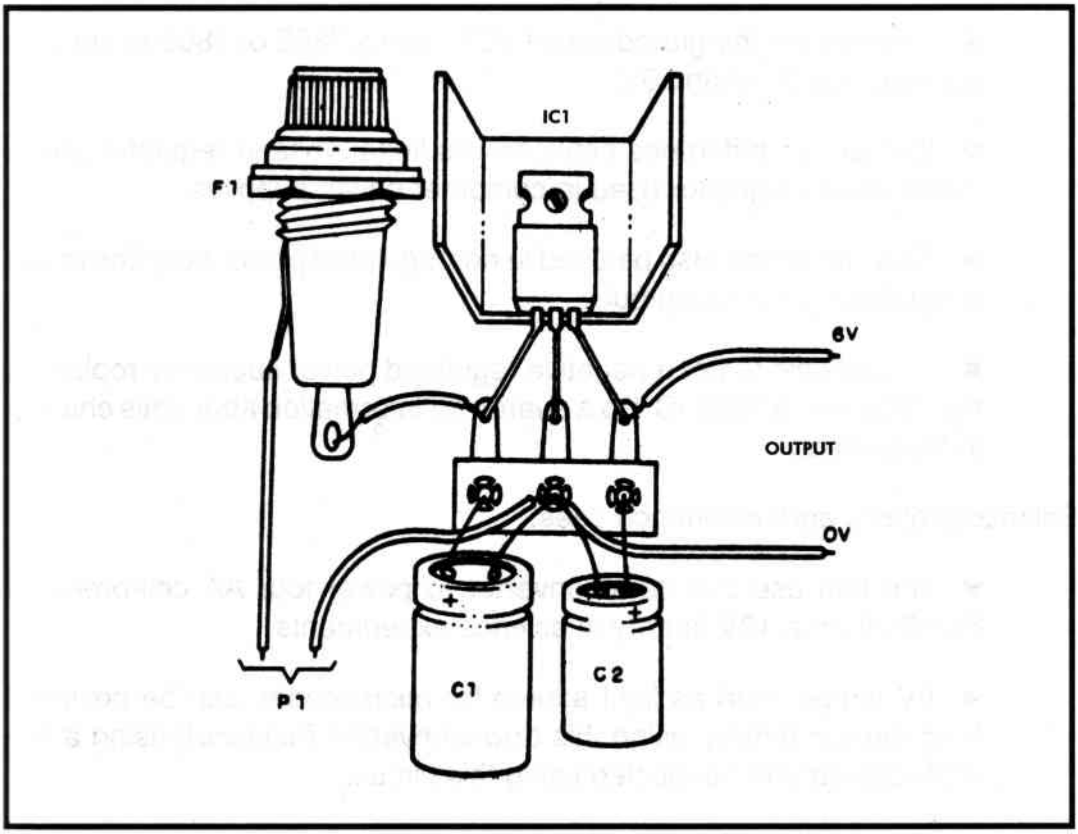
IC1
F1
6V
OUTPUT
0V
+
C1
C2
P1

Figure 2

Parts List — 6 Volts From 12V Batteries

IC1	- 7806 Integrated circuit
C1	- 1,000 uF, 16 WVDC electrolytic capacitor
C2	- 100 uF, 12 WVDC electrolytic capacitor
F1	- 2A fuse

Ideas to Explore

To learn more about the device or to get better performance:

- Replace the integrated circuit (IC1) with a 7805 or 7809 to get output voltages of 5V and 9V.
- Explain the difference between this linear voltage regulator and a switch-mode regulator used in computer power supplies.
- This circuit can also be used in nonregulated power supplies to get a regulated 6V x 1A output.
- It's possible to get a negative regulated power supply by replacing the 7806 with a 7906 IC. Do a search for information about this change in the project.

Science projects and uncommon uses:

- You can use this downconverter to power four AA cell-powered circuits from a 12V battery in science experiments.
- 6V lamps used as light source for microscopes can be powered from the car battery using this downconverter. Fieldwork using a microscope can be conducted using this circuit.

ONE-COMPONENT TRANSMITTER (E)

Here is a project that should be constructed by those with no patience for electronics building. Only one component is used to build a complete wireless transmitter!

Of course, this is the simplest way to generate radio signals. It is a CW (continuous wave) transmitter and is powerful enough to plunk down a signal anywhere on the 550 kHz to 7 MHz band.

The fundamental frequency depends upon power supply voltage and is in the range between 100 kHz and 1 MHz, but there are strong harmonics reaching up to 7 MHz.

Figure 1 shows the schematic diagram of this simple tramsmitter. As you can see, we have only one component: a 4049 CMOS IC that is formed by six inverters. In this project only five inverters are used.

The circuit is powered from AA cells or battery, and current drain is very low.

As an experimental project you can place the IC on a universal solderless board. The antenna is a single wire 2 to 5 feet long.

A common AM (medium wave or short wave) receiver is placed near the transmitter. A signal can be tuned in from distances up to some feet away.

There are two ways to tune the transmitter, shown in ***Figure 2***.

In the first case, as the frequency is voltage-dependent, we change the power supply voltage. Use a 12 volt power supply and adjust the voltage in the IC with P1. In the second case, we include a variable capacitor in the feedback loop to alter the frequency.

Parts List — One-Component Transmitter

IC	- 4069 or 4049 CMOS integrated circuit (only this!)
	(And a power supply, of course)

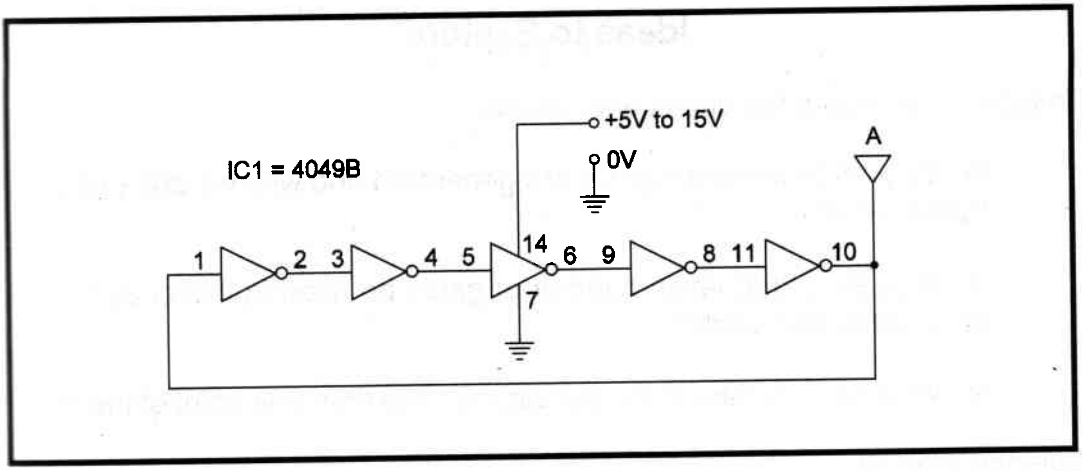
+5V to 15V
0V
IC1 = 4049B
A
1
2
3
4
5
14
6
9
8
11
10
7

Figure 1

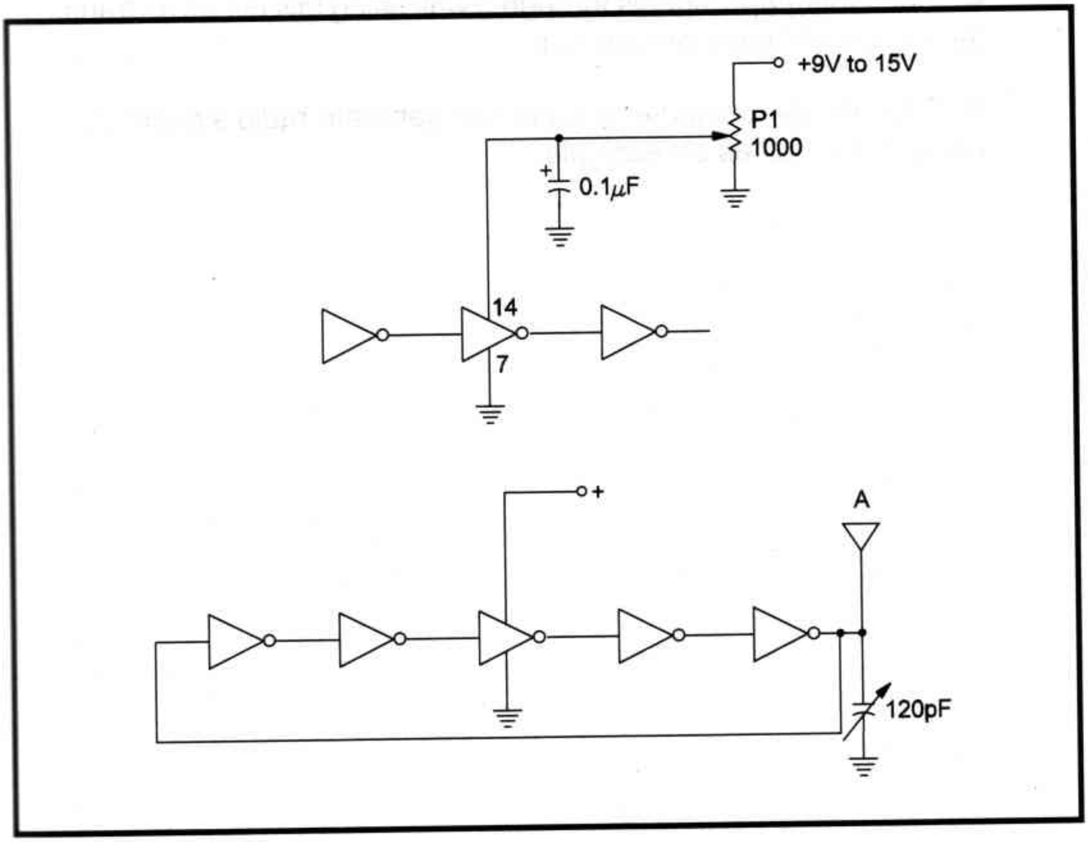
+9V to 15V
P1
1000
+
0.1μF
14
7
+
A
120pF

Figure 2

Ideas to Explore

To learn more about the circuit and device:

- Explain how radio signals are generated and why we don't need a tuned circuit.
- Why should an even number of gates be used in this circuit? Will an odd number work?
- Why can you tune in the signals on more than one point of the dial?

Science projects:

- Send telegraph signals through a wall using this circuit as transmitter and an AM radio as receiver.
- Explain how computer circuits can generate radio signals (noise), using this circuit as an example.

WIRELESS FM TRANSMITTER (E/P)

These circuits have always been popular for both utility and entertainment purposes.

The FM transmitter described here is built with few components to save space and cost. The circuit has only one transistor and is powered by two or four AA cells.

The broadcasting frequency of the unit, in the range between 88 and 108 MHz (FM band), is determined by L1 and the variable capacitor CV (a 20 or 30 pF trimmer capacitor). L1 consists of four turns of No. 22 wire, wound in a single close-spaced layer on a 1/2-inch form. The coil is center tapped where it is wired to the antenna terminal.

Any piece of stiff wire will serve as the antenna and will give good coverage. For better operation, antenna length should be between 5 and 15 inches.

Try out your unit at any convenient dead spot on the the FM band. Adjust CV to the best signal. The signals can be tuned in distances up to 150 feet.

A schematic diagram of the FM transmitter is shown in ***Figure 1***. Only one transistor as used as a high-frequency oscillator. Any small RF NPN transistor can be used for the task.

Components placement on a printed-circuit board is shown in ***Figure 2***. Capacitors are critical components and should be the types indicated in the parts list. Don't use metal film capacitors instead of the ceramic capacitors which are called for in the parts list.

You can also use a terminal strip as chassis as shown in ***Figure 3***, but keep all the component terminals and wiring short to avoid instability.

The circuit can be housed in a small plastic box along with the power supply. Avoid the use of metal boxes to house the circuit.

Take care with position of the polarized pieces, such as the electret microphone, electrolytic capacitor and power supply.

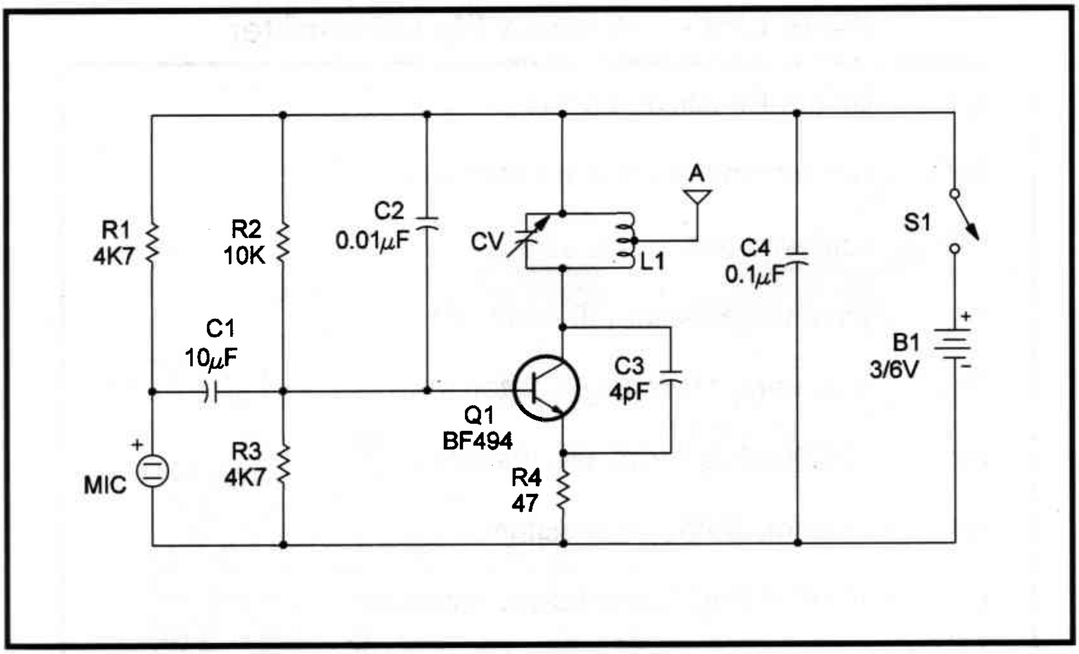
R1
4K7
R2
10K
C2
0.01μF
CV
A
L1
C4
0.1μF
S1
B1
3/6V
C1
10μF
C3
4pF
Q1
BF494
MIC
R3
4K7
R4
47

Figure 1

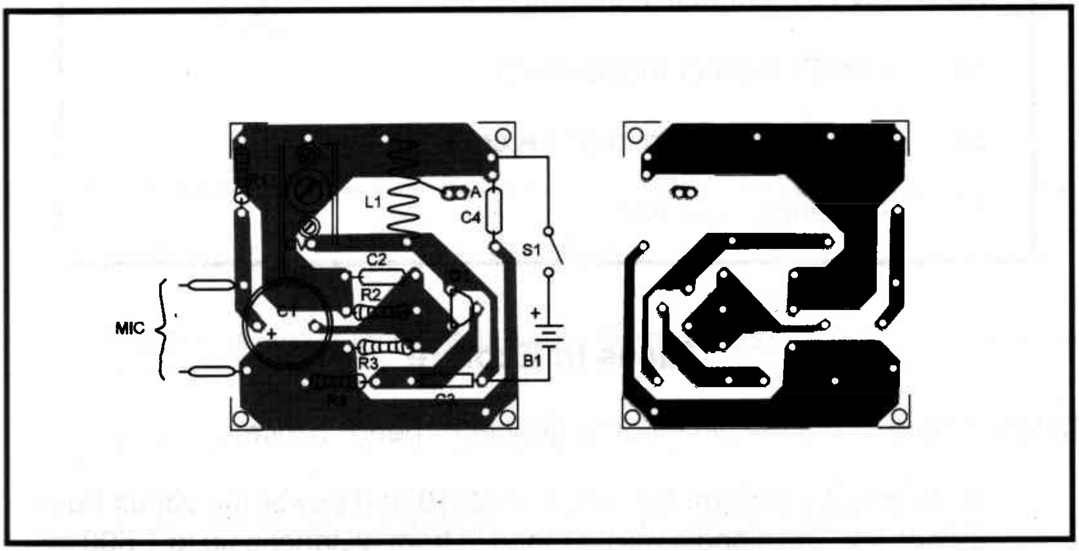
L1
A
C4
S1
C2
R2
C1
R3
B1
MIC
R1
CV
Q1
R4
C3

Figure 2

Parts List — Wireless FM Transmitter

Q1 - BF494 RF silicon transistor

MIC - two-terminal electret microphone

L1 - coil - see text

CV - trimmer capacitor (20 to 50 pF)

R1, 3 - 4,00 ohm, 1/4W, 5% resistor

R2 - 10,000 ohm, 1/4W, 5% resistor

R4 - 47 ohm, 1/4W, 5% resistor

C1 - 10 uF, 6 WVDC electrolytic capacitor

C2 - 0.01 uF ceramic capacitor

C3 - 4.7 pF ceramic capacitor

C4 - 0.1 uF ceramic capacitor

S1 - SPST slide or toggle switch

B1 - 3V or 6V - two or four AA cells

A - antenna - see text

Ideas to Explore

To learn more about the circuit or to get better performance:

- Replace transistor Q1 with a 2N2218 and power the circuit from a 9V battery. The signals will be tuned in from distances up to 1,500 feet.
- The circuit can be used as a “bug” to hear secret conversations.

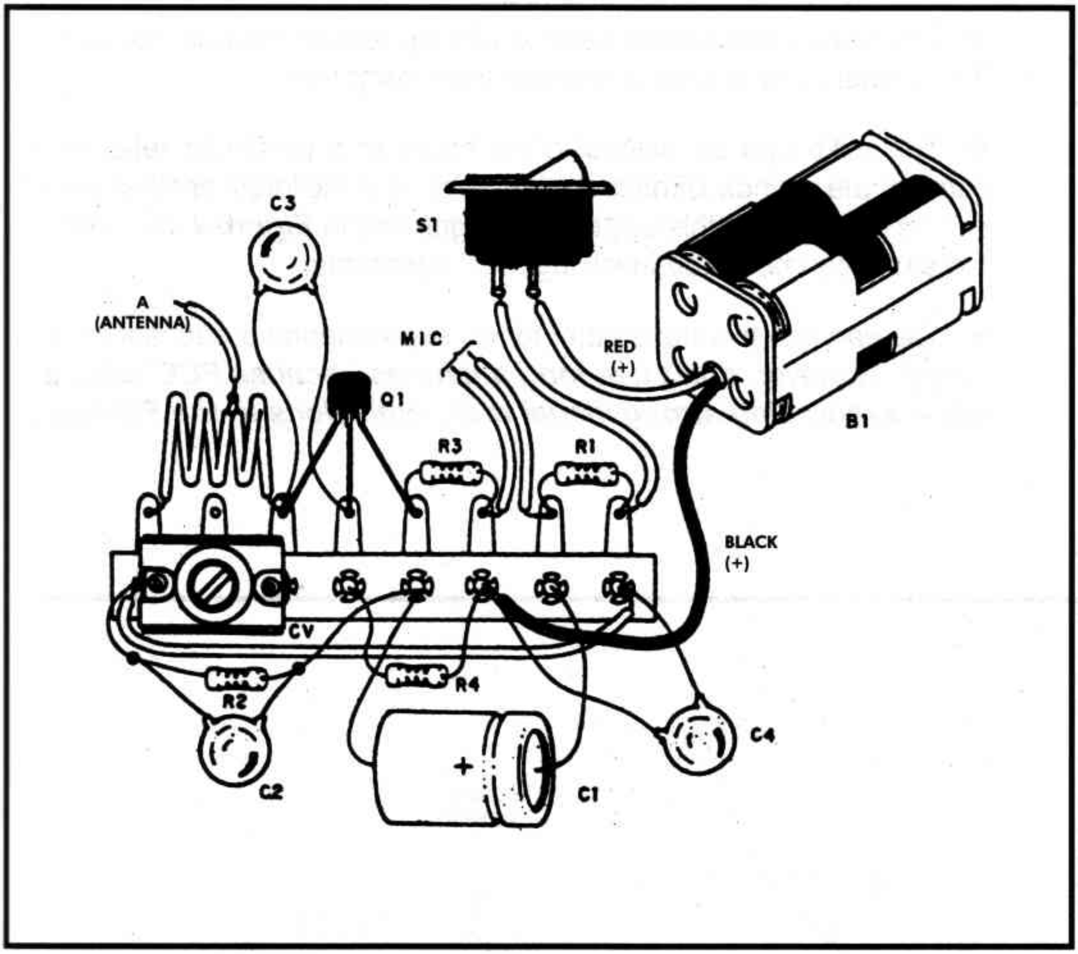
C3
S1
A
(ANTENNA)
MIC
RED
(+)
Q1
B1
R3
R1
BLACK
(+)
CV
R4
R2
C4
C2
C1
+

Figure 3

Place the circuit inside a false book on a desk to pick up conversations.

- Explain how the transmitter works.
- What are radio waves or electromagnetic waves?
- What is Frequency Modulation (FM) and Amplitude Modulation (AM)? Explain the differences.

Science projects using this transmitter:

- The transmitter can be used to pick up remote sounds. You can fix it in animal pens or jails to monitor them remotely.
- The mike can be placed in the focus of a parabolic reflector to concentrate sounds. Birds and other animals in the forest or other places can be heard using this system, as suggested in ***Figure 4***. An interesting study can be made involving local songbirds.
- You can use this transmitter for an experimental broadcast in your school. However, ***don't use long antennas because FCC rules are severe regarding strong illegal radio emissions in the FM band.***

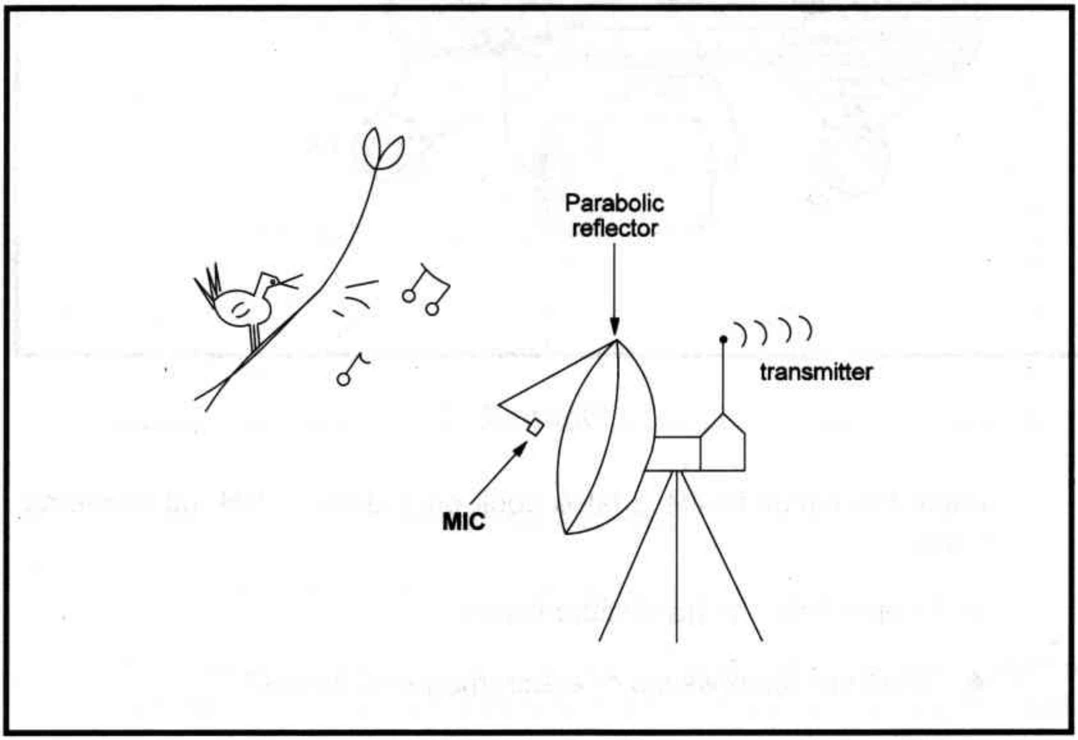

Figure 4

HIGH-POWER AUDIO OSCILLATOR (E)

This high-power version of an audio oscillator produces a strong audible tone using a piezoelectric transducer. All four gates existing in the 4093 IC are used in this project to drive the transducer with a good audio signal.

The circuit can be used in alarms, toys, as a standalone project to teach about oscillators specifically, or in experiments using continuous sounds in the frequency range between 100 and 10,000 Hz.

The recommended piezoelectric transducer emits maximum output power between 700 and 3,000 Hz, but it will also operate in other frequencies with less power.

Power supply voltage range is between 5 and 12 volts. Current requirements depend upon the power supply voltage, ranging from 10 to 50 mA typically.

P1 adjusts the tone frequency. C1 can be altered within a large range of values as indicated in the schematic diagram, shown in ***Figure 1***.

Components placement on a printed-circuit board is shown in ***Figure 2***. You can also mount the circuit using a solderless board for experimental applications.

All the components and the batteries (if used as power supply) can be housed in a small plastic box.

BZ can be replaced with a common piezoelectric tweeter. But, in this case, you should open the tweeter and remove the small transformer from inside it. Then the output of the circuit should be connected directly to the small piezoelectric transducer.

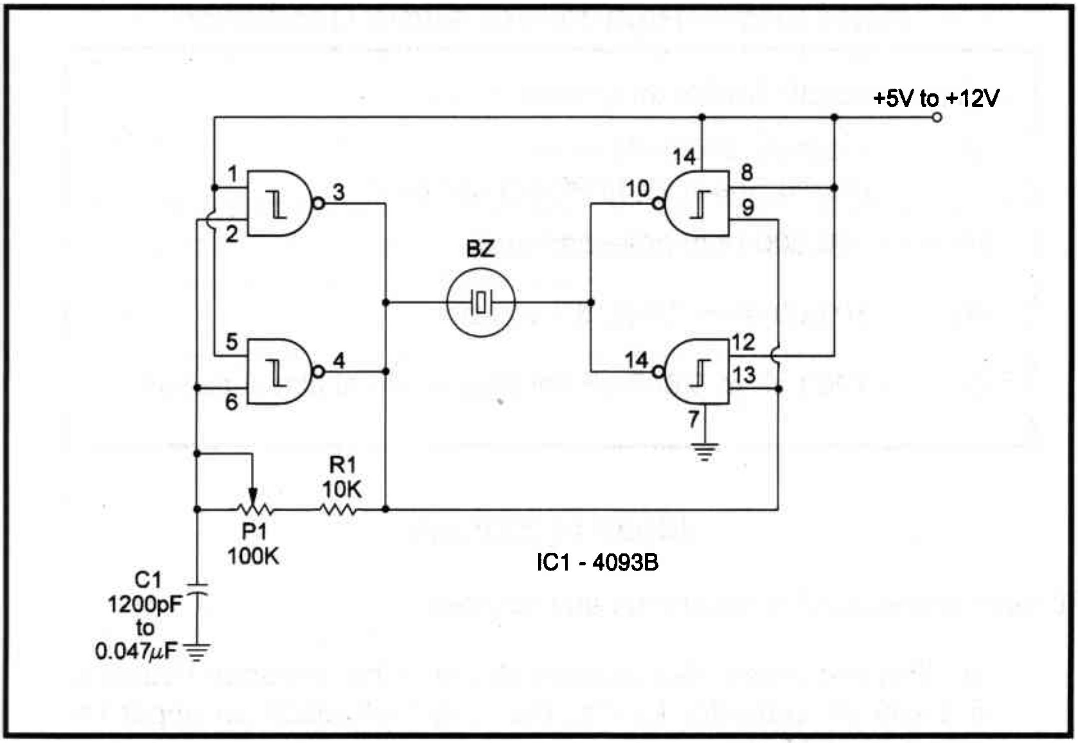
+5V to +12V
14
8
1
3
10
9
2
BZ
5
4
12
14
13
6
7
R1
10K
P1
100K
IC1 - 4093B
C1
1200pF
to
0.047µF

Figure 1

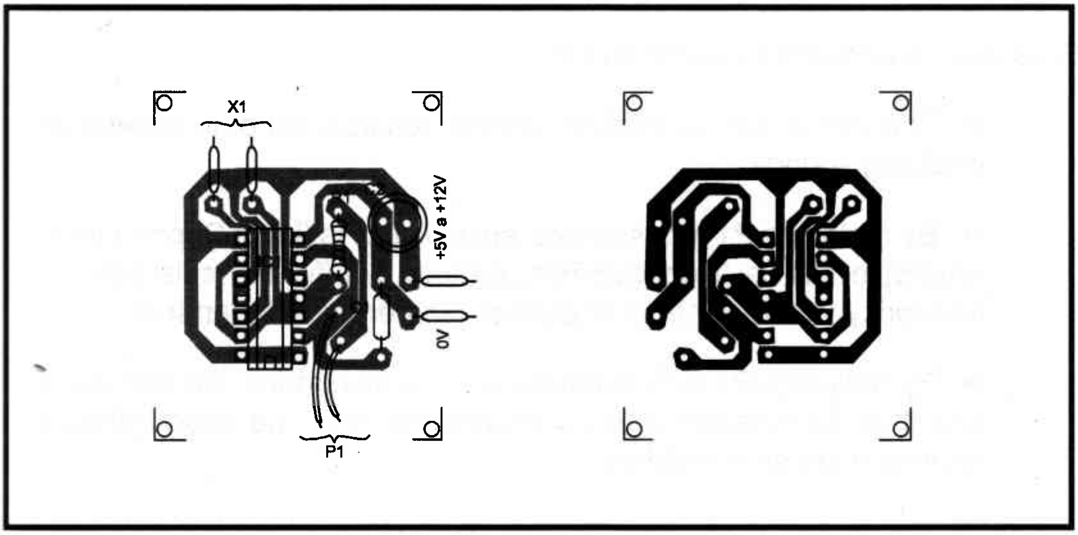
X1
+5V a +12V
0V
P1

Figure 2

Parts List — High-Power Audio Oscillator

IC1	- 4093B CMOS integrated circuit
BZ	- Piezoelectric transducer (Radio Shack 273-090 or equivalent)
P1	- 100,000 ohm potentiometer
R1	- 10,000 ohm, 1/4W, 5% resistor
C1	- 1200 pF to 0.047 uF ceramic or metal film capacitor

Ideas to Explore

To learn more about components and devices:

- This circuit can also produce sound in the ultrasound band. Using a 1,200 pF capacitor for C1, the range will reach an upper limit of about 100,000 Hz.

- Explain how a push-pull output stage operates.

Science and uncommon applications:

- This circuit can be used for animal conditioning or in experiments involving sounds.

- By replacing P1 with sensors, such as LDRs (Light Dependent Resistors) or NTCs (Negative-Temperature Coefficient resistors), tone frequency will be light-dependent or temperature-dependent.

- By replacing P1 with touch sensors or electrodes, we can use the circuit as lie detector or bio-feedback devices. The sound pitch will depend upon skin resistance.

- The circuit can also be used as a fish attractor or in experiments involving water animals and sounds. Replacing BZ with wires immersed into water will generate electric current fields for experimental works involving the influence of these fields on fish or other beings.

- Some animals, such as mice and rats, are scared off by ultrasounds. The circuit can be used in experiments to determine the frequencies that are more effective for this task.

See other projects involving sounds for new ideas to explore.

TIMER (E/P)

Small home appliances and circuits can be turned on after a time delay ranging from seconds to minutes using this simple timer. School and hobbyist experiments can be time-controlled by using this circuit.

The unit has a constant-current source to charge C1. C1 and P1 adjust the time range. You can alter C1 within a range of 100 and 1000 uF to change the time range.

Using a 1 ampere relay you can control AC loads up to 100 watts wired to the 117 VAC power line.

To operate the unit, connect the load to the relay contacts (points A and B, used as a switch) and adjust P1 to the desired time delay. Press S2 do discharge C1 completely . Then close S1 to start. After the adjusted time delay, the relay will close, and the load will be powered ON.

Conversely, by using the NC (Normally Closed) contacts of the relay you can turn a device OFF after the adjusted time.

To use the unit again you must press S2 to discharge C1 before starting.

Figure 1 shows the schematic diagram of the timer. Observe the different components used in this project: a unijunction transistor (UJT) and an SCR (Silicon Controlled Rectifier).

An mechanical view is shown in ***Figure 2,*** using a terminal strip as chassis. All the components and power supply can be housed in a small plastic box. Take care with positions of critical components, such as the UJT, SCR, diodes and electrolytic capacitors.

The relay depends upon the power supply voltage and voltage drop through the SCR. This drop is about 2 volts, so you need to use 6 to 7.5-volt relays with a 9-volt power supply.

The SCR needn't be mounted on a heatsink, as the current drain is low.

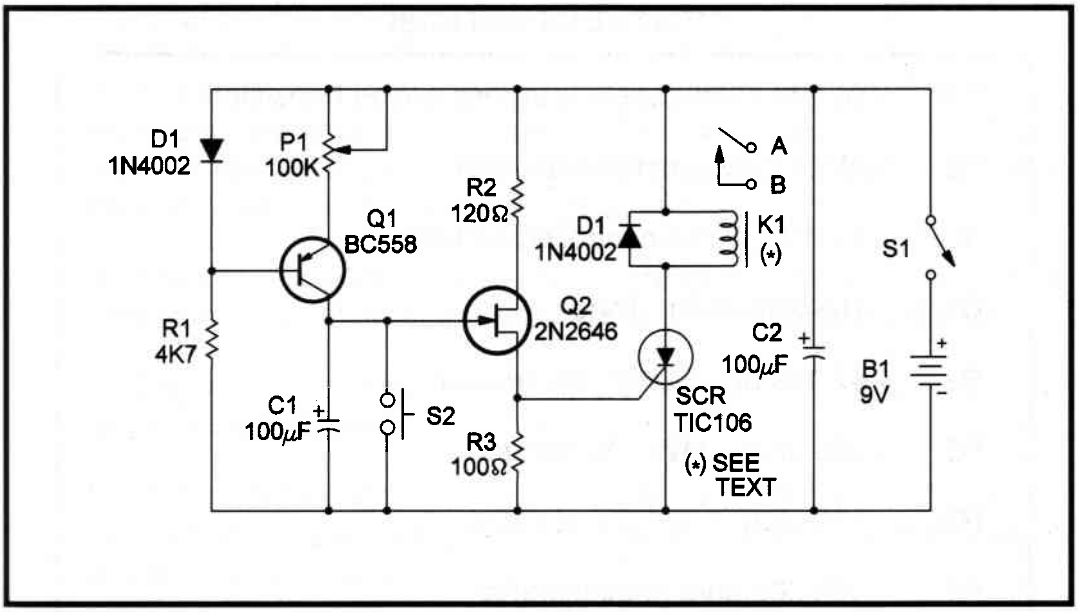
D1
1N4002
P1
100K
R2
120Ω
Q1
BC558
D1
1N4002
K1
(*)
A
B
S1
R1
4K7
Q2
2N2646
C2
100μF
B1
9V
C1
100μF
S2
R3
100Ω
SCR
TIC106
(*) SEE TEXT

Figure 1

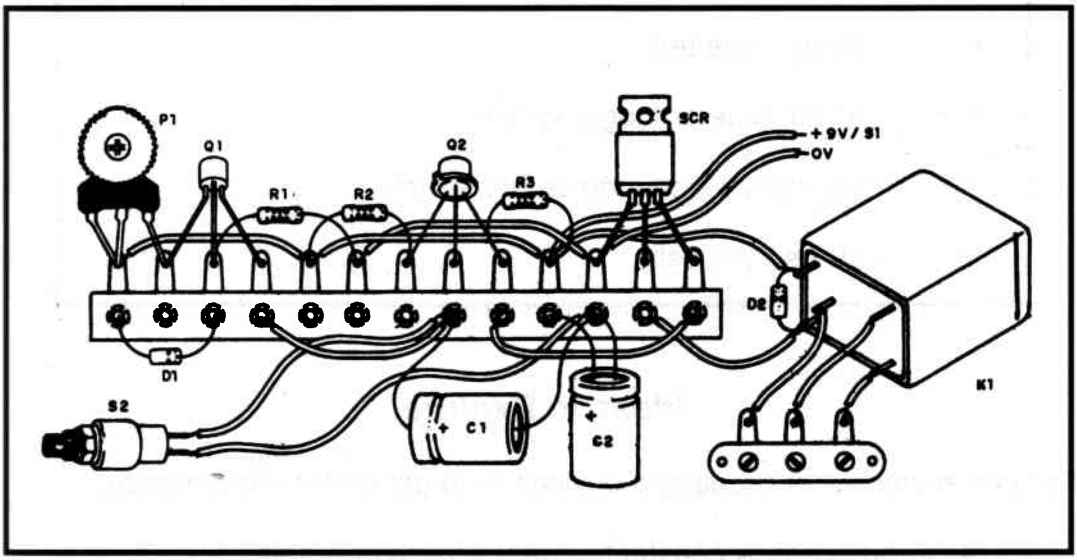
P1
Q1
Q2
R1
R2
R3
SCR
+9V / S1
0V
D2
K1
D1
S2
C1
C2

Figure 2

Use AA cells or a transformer to power this timer. Remember that current drain is low only when the relay is not energized.

Parts List — Timer

Q1 - BC548 NPN general-purpose silicon transistor

Q2 - 2N2646 unijunction transistor

SCR - TIC106 silicon controlled rectifier

D1, 2 - 1N4002 silicon diodes

R1 - 47,000 ohm, 1/4W, 5% resistor

R2 - 120 ohm, 1/4W, 5% resistor

R3 - 100 ohm, 1/4W, 5% resistor

P1 - 100,000 ohm potentiometer

C1, 2 - 100 uF, 12 WVDC electrolytic capacitors

K1 - Relay - see text

S1 - SPST slide or toggle switch

S2 - Normally Open momentary switch

B1 - 9V - six AA cells

Ideas to Explore

To learn much about circuit and devices or to get better performance:

- Explain how a constant-current source operates.
- Explain how a UJT and SCR operate.
- Alter C1 to get higher time delays. Try values up to 1,000 uF.

Science projects and uncommon uses for this device:

- All experiments that need time-delay control can use this device, as long as the time delay needed falls in the range of this timer (between a few seconds and minutes). Use a common watch as reference to make a scale for the potentiometer.
- The circuit can be used to activate dangerous experiments with a safe time-delay allowing the experimenter to reach safe distance.
- Traps can be armed after a time-delay, so the experimenter can avoid the trap himself.
- Chemical processes can be controlled using this timer. Any audio oscillator described in this book can be use as warning circuit when wired to the relay output.

MICRO AM RADIO (E/P)

This micro AM radio will tune strong stations in the frequency range between 530 and 1,600 kHz.

The circuit uses only one transistor as the amplifier and with a transformer drives a small loudspeaker. As the circuit is very poor in amplification, it needs a long wire antenna. The antenna should be from 15 to 50 feet long for best results. A good ground connection is also important.

Of course, this is a small inexpensive transistor radio with low listening volume, particularly on a weak station. If the station is too weak, replace the SPKR by a low-impedance earphone.

The power supply is formed by two or four AA cells and current drain is very low, extending their life. The circuit drains only few microamperes.

L1 is a tapped loopstick, connected to variable capacitor CV. These items can often be found in non-working transistor radios, as well as the speaker and transformer.

Figure 1 shows the schematic diagram for this project. Diode D1 acts as a detector and Q1 as an audio amplifier. T1 is a common transistor output transformer.

In ***Figure 2*** we show the components layout. All the components can be housed in a small plastic box. A terminal strip is used as chassis in an experimental version.

T1 is a transistor transformer (1,000:8) and the SPKR is a miniature type. But, if you have an unworking AM radio you can get these parts without spending any money.

To use, close S1 and tune to the desired station by adjusting CV. You can adjust R1 for the best results according to your transistor gain.

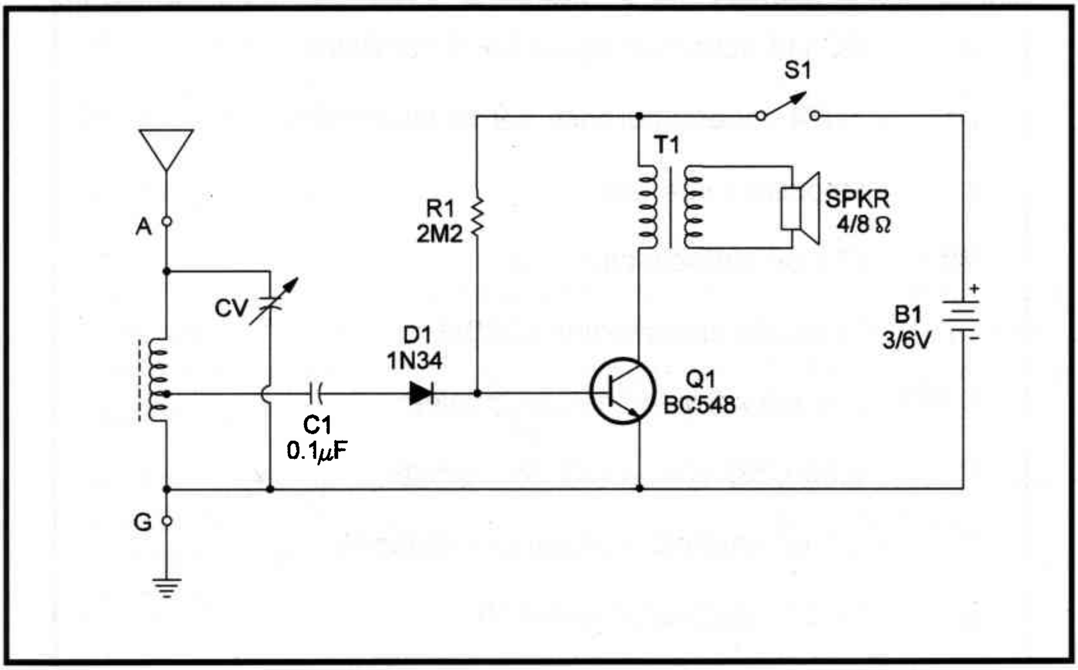
S1
T1
R1
2M2
SPKR
4/8 Ω
A
CV
D1
1N34
+
B1
3/6V
-
Q1
BC548
C1
0.1μF
G

Figure 1

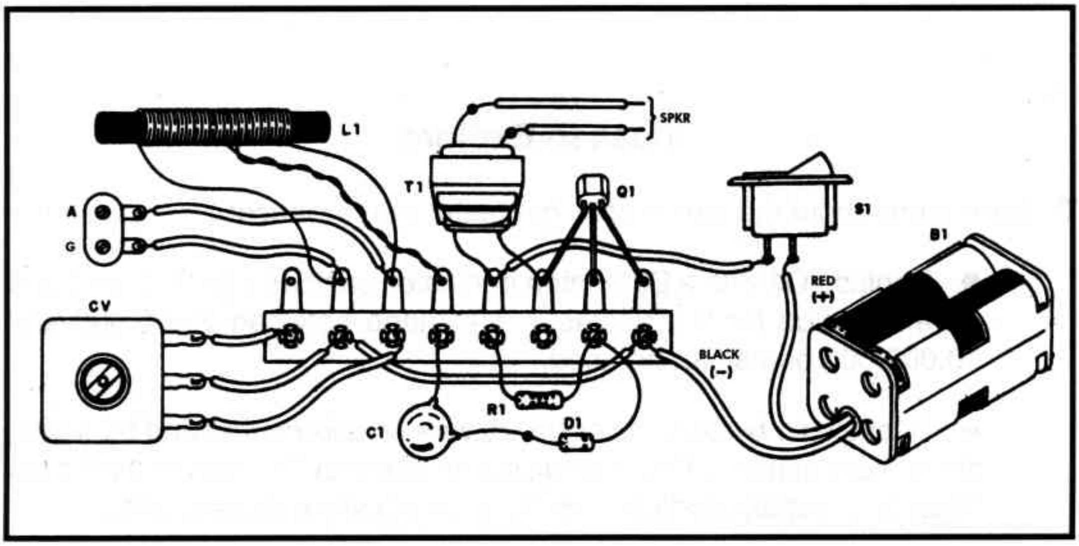
SPKR
L1
T1
Q1
S1
A
G
B1
RED
(+)
CV
BLACK
(−)
R1
C1
D1

Figure 2

Parts List — Micro AM Radio

Q1 - BC548 general-purpose NPN transistor

D1 - 1N34 general-purpose germanium diode

L1 - loopstick - see text

CV - 365 pF variable capacitor

T1 - Transistor transformer 1,000:8

SPKR - 8 ohms x 2 in. miniature speaker

R1 - 2,200,000 ohm, 1/4W, 5% resistor

C1 - 0.1 uF ceramic or metal film capacitor

S1 - SPST toggle or slide switch

B1 - 3 or 6V - two or four AA cells

Ideas to Explore

To learn more about the circuit or to get better performance:

- Replace Q1 with a Darlington transistor, such as a BC517, and find a better value for R1. Values in the range between 2,200,000 and 10,000,000 ohms can be used.

- Replace L1 by 20 turns of No. 28 wire on a ferrite rod and try tuning short-wave stations. Use a long wire as antenna. Position of the L1 tap must be carefully studied to get best selectivity and sensibility.

- Experiment with a crystal earpiece in the place of T1 and SPKR. A 10,000 ohm resistor should be wired in parallel with the earpiece.

- Explain how crystal sets or direct-detection receivers operate.
- Define the term "selectivity".
- What is a galena-crystal?

Science projects:

- This receiver can be used to demonstrate how antique radio receivers operated when radio was first invented. It can be presented as a "modern version" of a crystal or galena receiver in a demonstration. An interesting exhibit can be made with this set, the free-power AM receiver (described in this book), and a crystal set.
- The circuit can also be used as an RF "sniffer", detecting radio signals produced by experimental circuits in the AM or short-wave band. Replace L1 with a 1,000 uH choke and remove C1 to get a sensitive RF sniffer.

TOUCH SWITCH (E)

Small appliances, lamps and motors can be controlled by a simple touch with this experimental circuit. Loads are ON during the time adjusted by P1 for a few seconds to more than 10 minutes. Conversely, by using the NC (normally closed) contacts you can also turn off an appliance after the adjusted time.

The sensor is formed by two small metal plates, and shouldn't be placed far from the circuit. The metal plates are finger-touched at the same moment.

The circuit can be powered from a 6- or 12-volt power supply, depending on the relay used. Do not use transformeless power supplies with this project, to avoid shocks or dangerous shorts. Transformerless power supplies are not isolated from the AC power line and represent a dangerous shock hazard.

K1 is a relay with 6- or 12-volt coil and is energized by current up to 50 mA. With relays in this range you can power the touch switch from AA cells (6 volts) or small power supplies.

A schematic of the Touch Switch is shown in ***Figure 1***. The circuit uses a 555 IC timer as its heart, and time delay is determined by adjusting P1.

Circuit mounting, using a small homemade printed-circuit board, is shown in ***Figure 2***. If you'd like to etch your own board use this parts placement diagram as guide.

In ***Figure 3*** we show the sensor that can be made with two small metal plates on a plastic or wood base. Wires to the sensor shouldn't be long. A maximum lenght of 10 feet is recommended to avoid erratic operation.

Position of the polarized components, such as the electrolytic capacitors and diodes, should be observed. Load is connected to terminals A and B. Remember that the relay acts as a switch.

Don't power this circuit from transformerless power supplies to avoid dangerous shocks and shorts.

A sensitivity control can be added to this circuit in a simple manner. Use a 1,000,000 ohm potentiometer for this task.

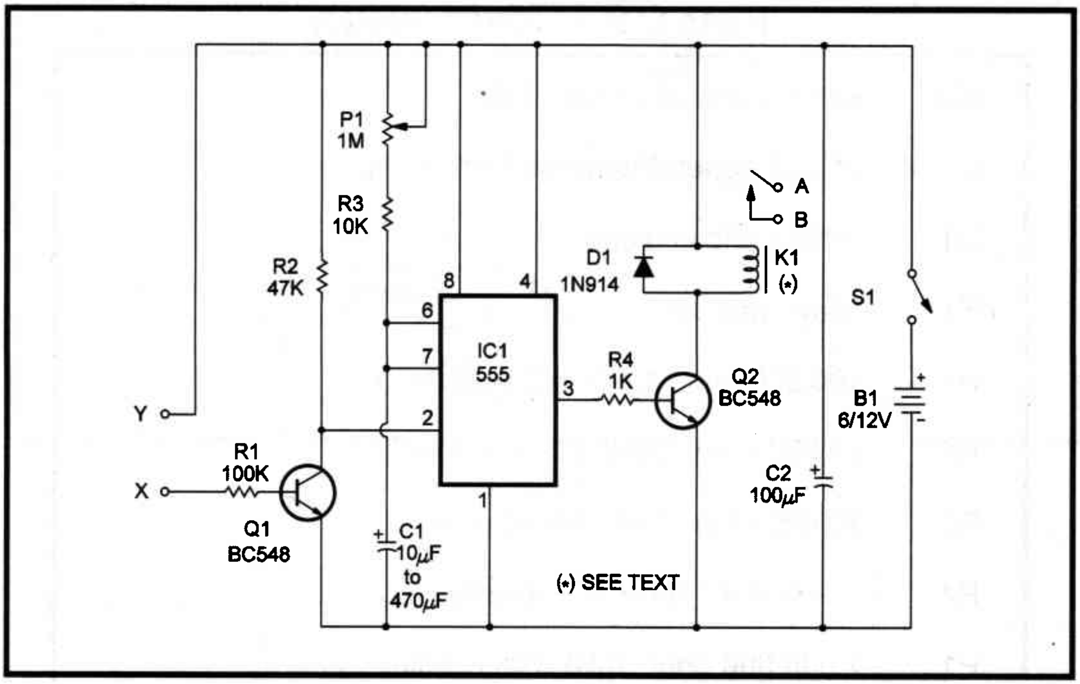
P1
1M
R3
10K
R2
47K
8
4
6
7
2
IC1
555
3
1
R4
1K
Q2
BC548
D1
1N914
K1
(*)
A
B
S1
B1
6/12V
+
-
Y
X
R1
100K
Q1
BC548
C1
10μF
to
470μF
C2
100μF
(*) SEE TEXT

Figure 1

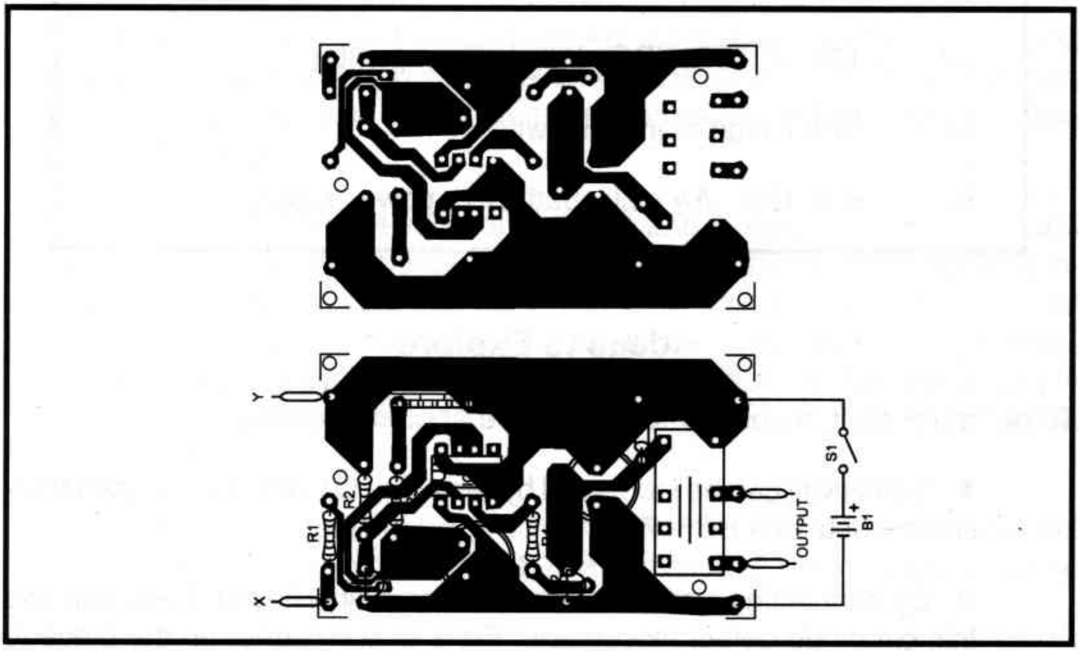
Y
X
R1
R2
S1
OUTPUT
B1
+

Figure 2

Parts List — Touch Switch

IC1	- 555 integrated circuit, timer
Q1, 2	- BC548 general-purpose NPN transistors
D1	- 1N914 silicon diode
K1	- relay - see text
R1	- 100,000 ohm, 1/4W, 5% resistor
R2	- 47,000 ohm, 1/4W, 5% resistor
R3	- 10,000 ohm, 1/4W, 5% resistor
R4	- 1,000 ohm, 1/4W, 5% resistor
P1	- 1,000,000 ohm, 1/4W, 5% resistor
C1	- 10 to 470 uF, 12 WVDC electrolytic capacitor - see text
C2	- 100 uF, 16 WVDC electrolytic capacitor
S1	- SPST toggle or slide switch
B1	- 6 or 12V - AA cells, battery or power supply

Ideas to Explore

To get more performance or to learn more about the circuit:

- Sensitivity can be increased by replacing Q1 with a Darlington transistor, such as a BC517.
- By connecting a reed switch between points Y and X you can use this circuit do detect movement. Fix a small magnet on the piece to act on the reed switch.

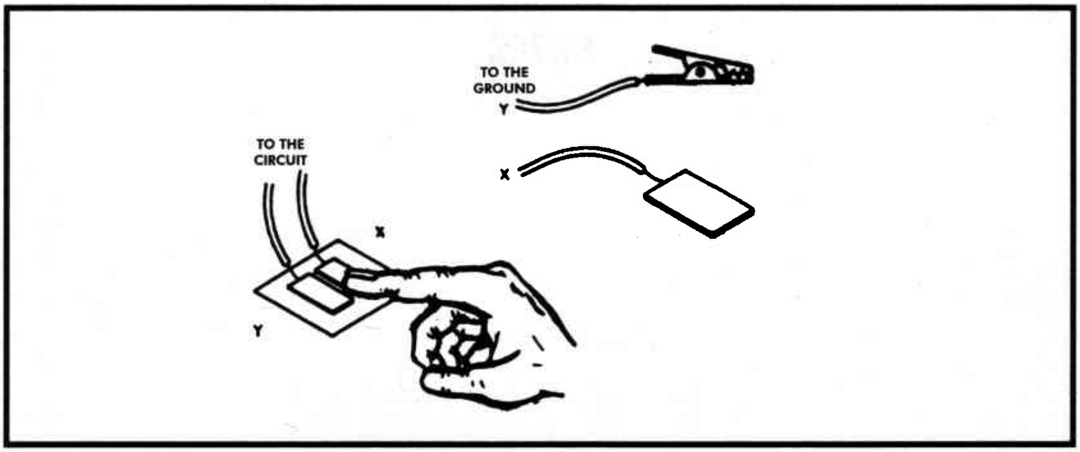

Figure 3

Science and uncommon applications:

- This circuit can be used to close a trap. The touch sensor will detect when the animal enters the trap.

- Projects that need a way to detect movements, switching any load, can be made with this circuit. Several kind of sensors, like microswitches, pendulum sensors or any other momentary-type switch can be used in this circuit.

- Connect an LDR between points A and B. Also connect a 100,000 ohm potentiometer between X and the negative pole of the power supply. You'll have a light-operated relay or remote control. A flash of light produced by a lantern or even a cigarette lighter can close the relay, powering the load. The circuit can be used to detect flashes of light in science experiments. Fix the LDR into a telescope and you can detect meteor fallings.

- Experiments involving thunderstorms at night can be conducted automatically using this circuit. The first flash of lightning can be used to power ON the equipment.

SIMPLE FLUORESCENT LAMP INVERTER (E/P)

This circuit will light a common fluorescent tube (from 7 to 40 watts) from a 6- or 12-volt power supply, such as ni-cad cells, D cells, car battery or other source, without need of a choke or starter. Even the old tubes which no longer function at all on AC power lines will light when used for this purpose. (Some of them, though, may be too weak to light when the circuit is powered from 6 or 12 volts. You should experiment beforehand.)

You can use the device as an emergency light, in trailers, signaling, etc.

By powering the fluorescent lamp with a transformer power supply you increase safety factor of an experiment, avoiding dangerous shocks or shorts.

Current drain, typically ranging from 100 to 800 mA, depends on the characteristics of the lamp used, the power supply voltage, and the transformer.

The lamp will brighten according to the drained current. Experiments should be done with some different transformers and oscillator frequencies to get better performance out of the circuit.

Basicaly, the circuit is a low-frequency oscillator driving a transformer. High voltage is obtained from the transformer to light a fluorescent light.

The trimmer potentiometer allows frequency adjustment to get better performance.

Schematic diagram of the Fluorescent Lamp Inverter is displayed in ***Figure 1.*** As you see, the circuit consists of a low-frequency oscillator with a pair of complementary transistors.

Components layout, using a terminal strip as chassis, is shown in ***Figure 2***. Components placement is not critical.

Transistor Q2 should be mounted on a large heatsink. T1 is a 6V x 450 mA or 12V x 450 mA transformer. If a CT (center-tapped secondary coil) transformer is you should use only two of the three secondary's wires.

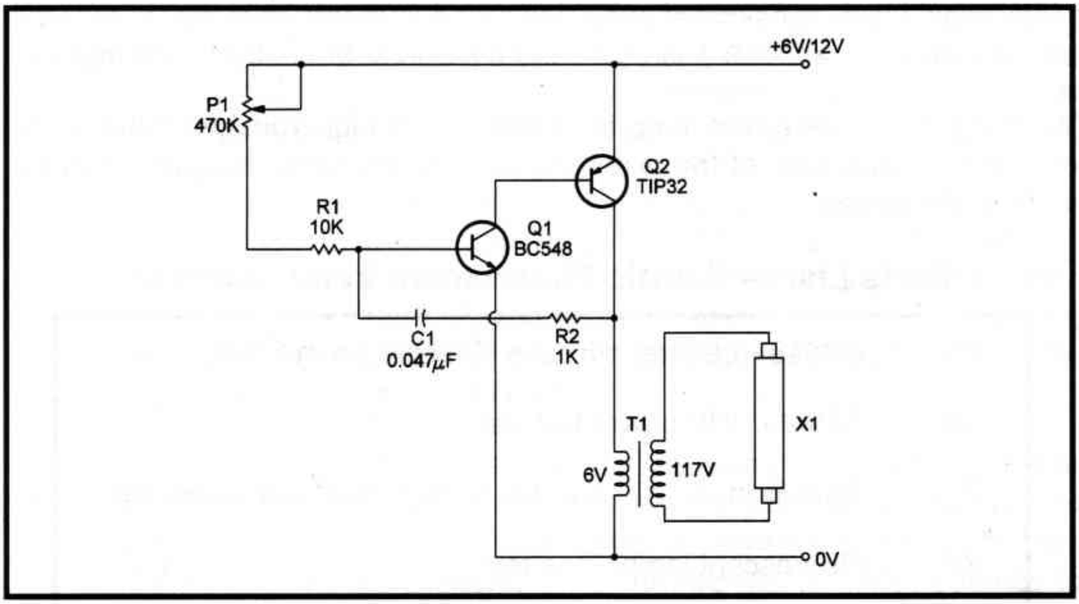
+6V/12V
P1
470K
Q2
TIP32
R1
10K
Q1
BC548
C1
0.047μF
R2
1K
T1
X1
6V
117V
0V

Figure 1

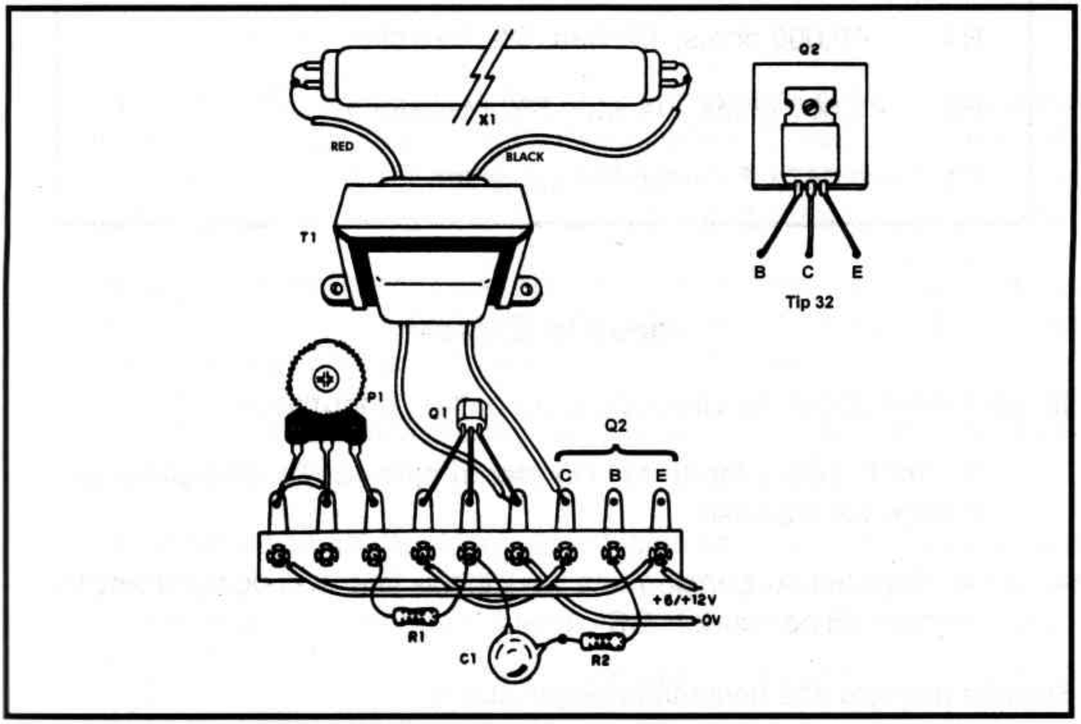
Q2
X1
RED
BLACK
T1
B
C
E
Tip 32
P1
Q1
Q2
C
B
E
+6/+12V
0V
R1
C1
R2

Figure 2

Experiments with transformers with secondary currents from 300 to 800 mA and voltages betweeen 9 and 15V can be made to get better performance.

Warning: The fluorescent lamp is powered with high voltages that can be dangerous. Take care of the wires' insulations and when handling connections to the device.

Parts List — Simple Fluorescent Lamp Inverter

Q1	- BC548 - general purpose NPN silicon transistor
Q2	- TIP42 - PNP power transistor
T1	- Transformer: 117 VAC; 6V x 450 mA or 12V x 450 mA
X1	- Fluorescent lamp - see text
P1	- 470,000 ohms - trimmer potentiometer
R1	- 10,000 ohms, 1/4-watt, 5% - resistor
R2	- 1,000 ohms, 1/4-watt, 5% - resistor
C1	- 0.047 uF - metal film capacitor

Ideas to Explore

To learn more about the circuit or to get better performance:

- Try to use a ferrite transformer in place of T1. The circuit performance will increase.
- Explain how transformers work in this kind of circuit and why transformers do not work in DC circuits.

Science projects and uncommon applications:

- Replace X1 by an small ultraviolet fluorescent lamp and you'll have a mineral or fluorescent materials searcher. Many materials, including the ones with a high percentage of calcium, will glow when illuminated by ultraviolet. Many experiments and demonstrations can be made involving minerals, plants and fluorescent materials .

- An interesting experiment with plants is to study the effect of ultraviolet radiation on different plants' growth or germination. Irradiate several batches of various seeds, such as tomato, been, squash, etc., with UV, and then plant. Because of increasing UV radiation reaching the earth due to depletion of the ozone layer, the effect of this radiation on plant life is of concern to us all.

- Compare effects of different light wavelengths to seed germination. The wavelenghts of incoming solar radiation range from 200 to 3,000 nanometers (nm). UV radiation is between 380 and 400 nm. Select seeds from common plants and expose differrent batches to UV and/or common light (red to blue). What happens with the diferent seeds during germination? You also can conduct the experiment not only with seeds, but plants and insects.

- High voltage present in the 117V transformer coil can be used in nerve stimulation experiments. *Please add a 10,000 ohm potentiometer to control the applied voltage if the circuit is intended to be used on humans! The voltage in the transformer has peaks as high as 400V!*

- Fluorescent light can also be use in experiments with insects. Some kinds of insects are attracted by light (visible and ultraviolet!). Insect traps can be constructed using this circuit as base.

- Fish and other animals can be attracted by fluorescent light. A circuit powered by cells or battery is important for fieldwork.

- Alternate methods of farm pest control can use light, visible as well as radiation in the ultraviolet part of the spectrum. Insect traps can be constructed with fluorescent lamp inverters. For further information:

The International Alliance for Sustainable Agriculture
1701 University Avenue SE
Minneapolis, MN 55414 (612) 331-1099

TOUCH ON/TOUCH OFF RELAY (P)

AC-powered appliances, such as lamps, small home appliances, tools and other items, can be controlled by your fingertips using this circuit. Also, test equipment and experiments can be controlled by this circuit. You can turn on any load by first touching time sensor X1, then turn it off by touching X1 again. There is no shock hazard, as the control circuit is completely isolated from the AC power line.

The control circuit is powered from four or eight AA cells, or, if you prefer, a power supply ranging from 6 to 12 volts according to the relay coil.

Don't use transformerless power supplies because they are not isolated from the AC power line and can cause severe shocks and shorts.

The schematic diagram of the Touch On/Touch Off Relay is given in ***Figure 1***. The circuit uses two ICs. The 555 timer works as a monostable, producing one control pulse when the sensor is touched. The 4013 is a flip-flop that controls the relay by a transistor.

Components placement on a homemade printed-circuit board is as shown in ***Figure 2***. Component layout should be altered if the relay used in the project is an equivalent of the type indicated in parts list.

Relay coil voltage is determined by the power supply voltage and the load current requirement. You can use a mini 1A DPDT (Radio Shack 275-249) 12V, 280 ohm, 43 mA, and wire it as shown in the figure, or use another type of relay, depending on the load requirements.

A 10A SPDT mini-relay (Radio Shack 275-278) is suitable for heavy duty appliances.

As a simple rule, you can use a 6- or 12-volt relay with coil currents ranging from 10 to 100 mA and contacts up to 10A, according to the task you have in mind.

Positions of the polarized components, such as the diode, electrolytic capacitors and transistor, should be observed.

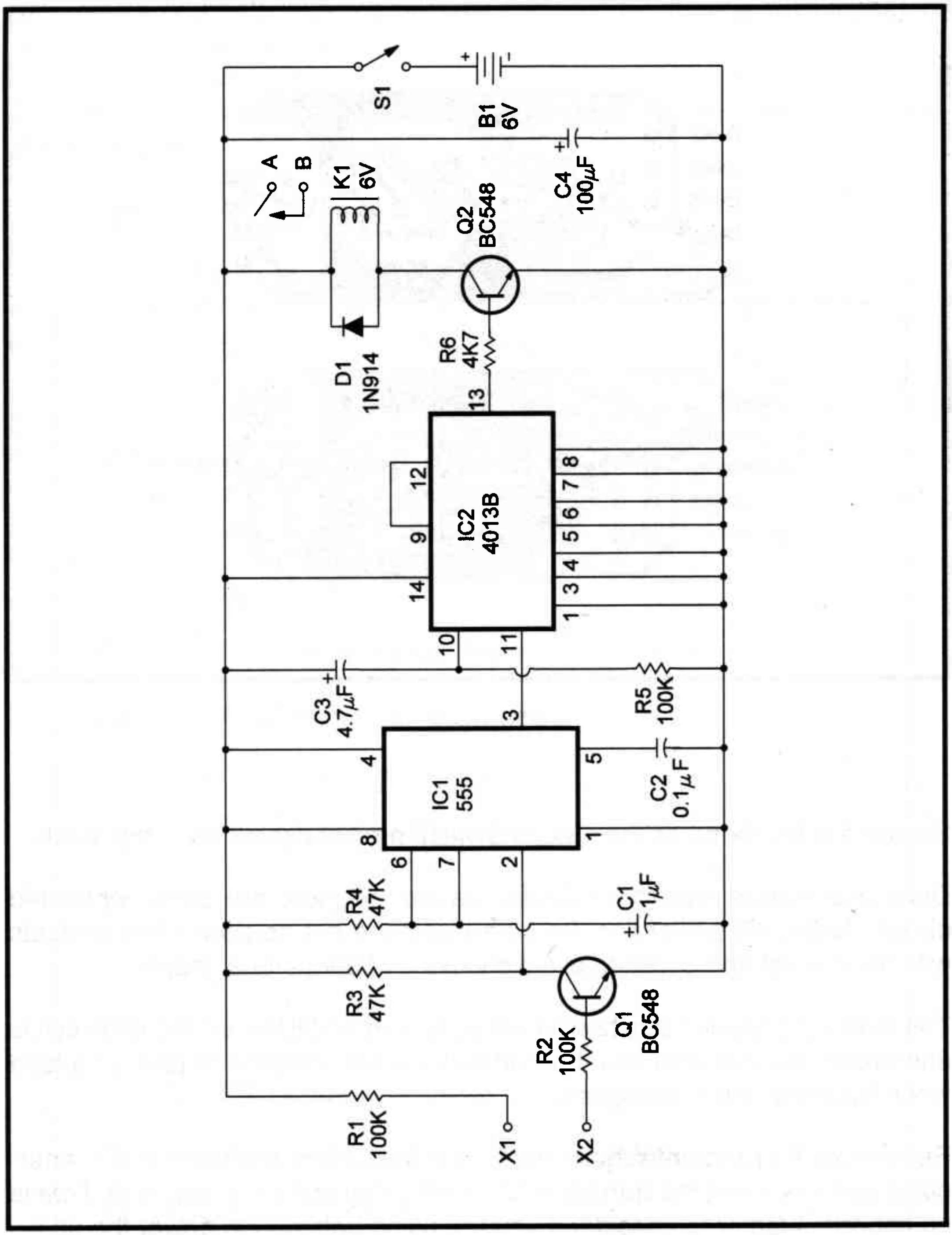
S1
B1 6V
C4 100μF
A
B
K1 6V
Q2 BC548
D1 1N914
R6 4K7
13
12
9
14
IC2 4013B
8
7
6
5
4
3
1
10
11
C3 4.7μF
R5 100K
3
4
IC1 555
5
C2 0.1μF
8
1
6
7
2
R4 47K
C1 1μF
R3 47K
Q1 BC548
R2 100K
R1 100K
X1
X2

Figure 1

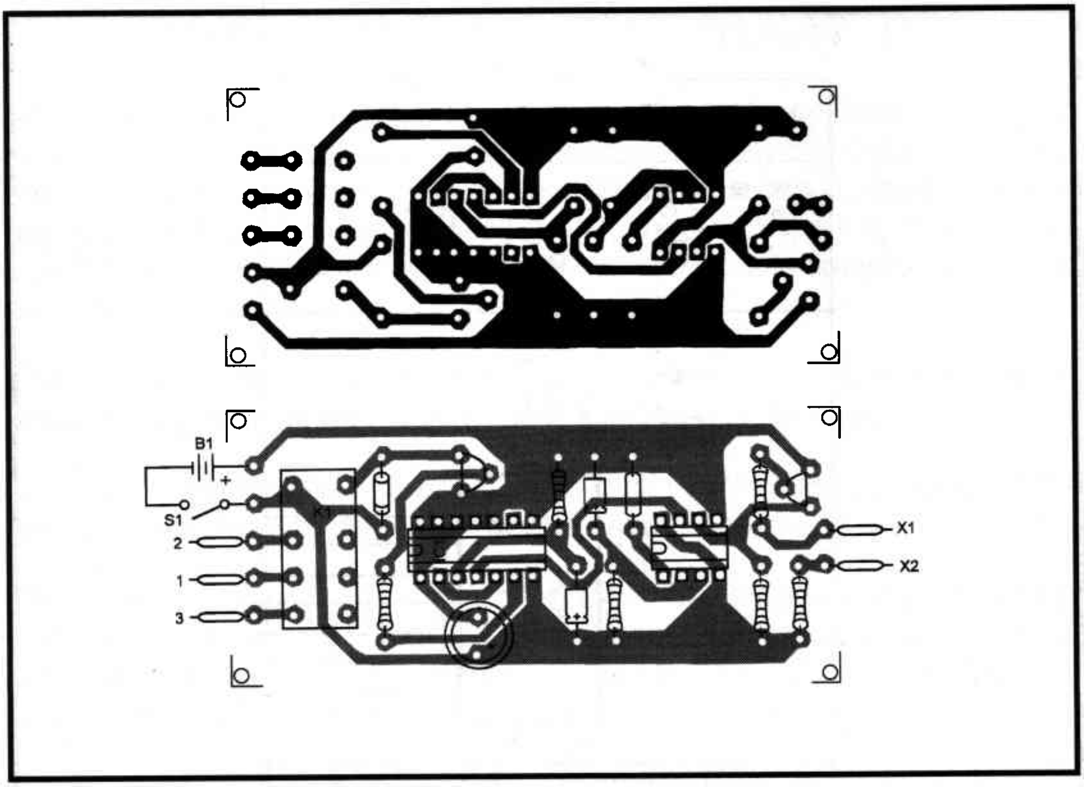

Figure 2

Sensor are the same as the "Touch Switch" project described in this book.

Don't use metallic box to house the device, as there are parts connected directly to the AC power line. Be sure that there are no power line contacts with the low voltage circuit to avoid shocks and dangerous shorts.

The load is connected as shown in ***Figure 3,*** to be ON when the relay coil is energized. You can also use the normally closed contacts to turn off a load when the relay coil is energized.

Remember that current requirements are high when the relay coil is energized and low when the transistor Q2 is off (relay coil nonenergized). That is an important factor to consider if you are using batteries to power the unit.

After you are sure that you have wired all parts of the circuit correctly, plug the power cord into the wall socket. You can now test the circuit.

Connect a load to the output (relay contacts) and touch the sensor. The load should be powered ON if all is OK.

Parts List — Touch On/Touch Off Relay

IC1 - 555 Integrated circuit - timer
IC2 - 4013B CMOS integrated circuit - dual flip-flop
Q1, 2 - BC548 General-purpose NPN silicon transistors
D1 - 1N914 General-purpose silicon diode
K1 - Relay - see text
R1, 2, 5 - 100,000 ohm, 1/4W, 5% resistors
R3, 4 - 47,000 ohm, 1/4W, 5% resistors
R6 - 1,000 ohm, 1/4W, 5% resistor
C1 - 1 uF, 25 WVDC electrolytic capacitor
C2 - 0.1 uF ceramic or metal film capacitor
C3 - 4.7 uF, 12 WVDC electrolytic capacitor
C4 - 100 uF, 16 WVDC electrolytic capacitor
S1 - SPST toggle or slide switch
B1 - 6 or 12V (according to relay) - see text
X1, 2 - sensor - see text

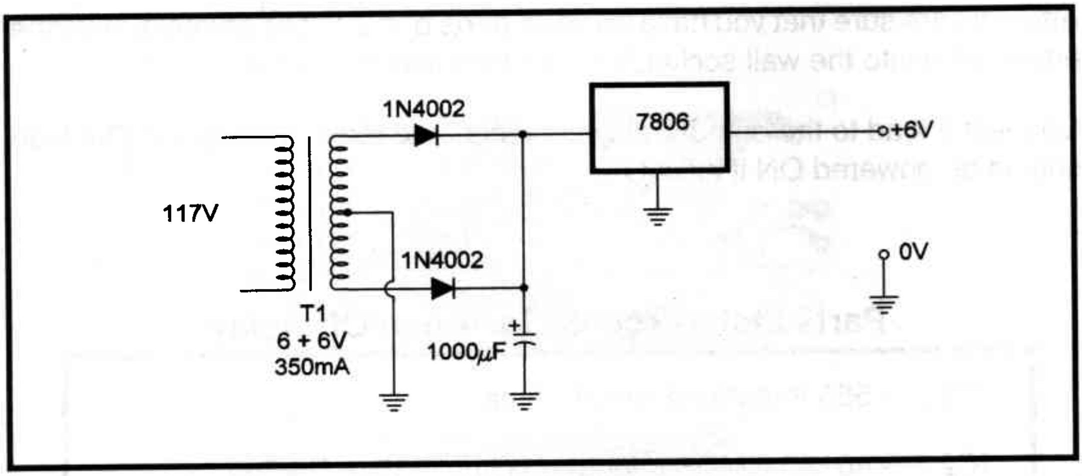

Figure 3

Ideas to Explore

To learn more about the circuit or to get better performance:

- Better performance can be obtained by replacing Q1 with a Darlington transistor, such as a BC517.
- A 100,000 ohm potentiometer placed between the base of Q1 and the negative pole of the power supply will give a sensitivity adjustment for the circuit.
- Explain how a flip-flop, as used in this circuit, works, and the function of the C3/R5 network in this project. Explain why the current flowing through the fingers when the sensor is touched doesn't cause any shock hazard.
- Depending upon the application, you can power ON the load when relay is not energized (to reduce power consumption). That is possible using the normally closed contacts of the relay.

Science projects and uncommon uses:

- By connecting a reed switch between points X1 and X2 you can use the circuit to detect movement of mechanical items in experiments or projects. Just affix a magnet on the moving item to act on the switch.
- Experiments with animals can be conducted with this circuit by using appropriate transducers connected between X1 and X2.
- An LDR can be used to control the circuit with a flash of light. A lantern, cigarette lighter, etc., can be used to turn appliances on or off.

EXPERIMENTAL GALVANOMETER (E)

The experimental galvanometer described here can detect currents as low as few microamperes and uses unusual parts. This galvanometer operates as common commercial galvanometers found in multimeters and several other analog instruments.

One purpose of this experiment can be to demonstrate that there is a magnetic field around a coil that that is carrying electricity (Oersted Experiment).

The operating principle is very simple to understand: an electric current flowing across a coil (or a wire) produces a magnetic field. The field can act on a mobile magnetic metal piece, such as a compass pointer or a blade, as shown in the figures herein. The compass or blade movement indicates the presence of a current.

Figure 1 shows the two basic versions of the experimental galvanometer, using a common compass and using a blade. Is very important to observe the correct path for the current flow in the coil. The coil should be wired exactly as shown in the figure.

In ***Figure 2*** we have a circuit to operate the galvanometer. The resistor can range from 100 to 100,000 ohms. The higher the resistor, the more sensitive is the galvanometer. Experiment to find the highest resistor value that can cause the pointer to move.

The coil is made with 50 to 200 turns of No. 28 to 32 wire on a paper form. Number of turns determines sensitivity. The higher the number of turns, the more sensitive is the galvanometer.

Currents as low as a few microamperes can be detected with a carefully made prototype.

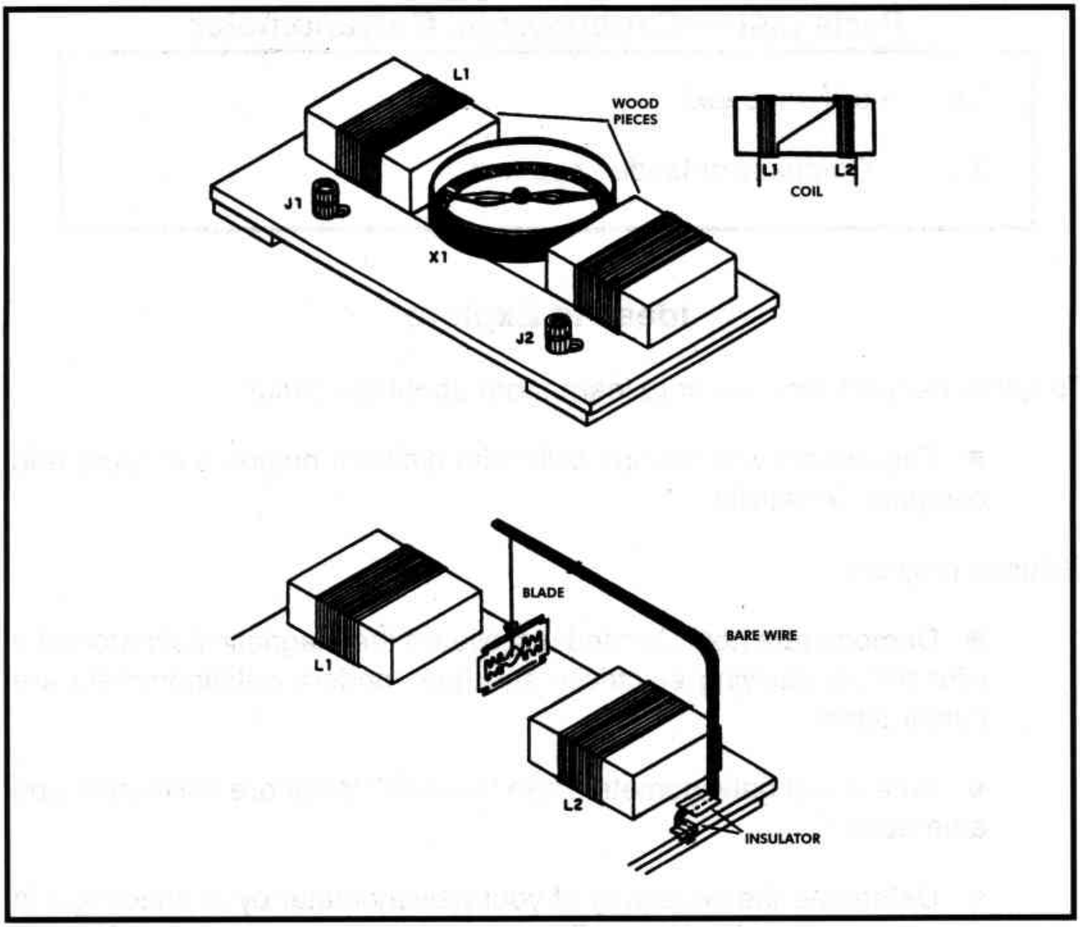
L1
WOOD PIECES
L1
L2
COIL
J1
X1
J2
BLADE
BARE WIRE
L1
L2
INSULATOR

Figure 1

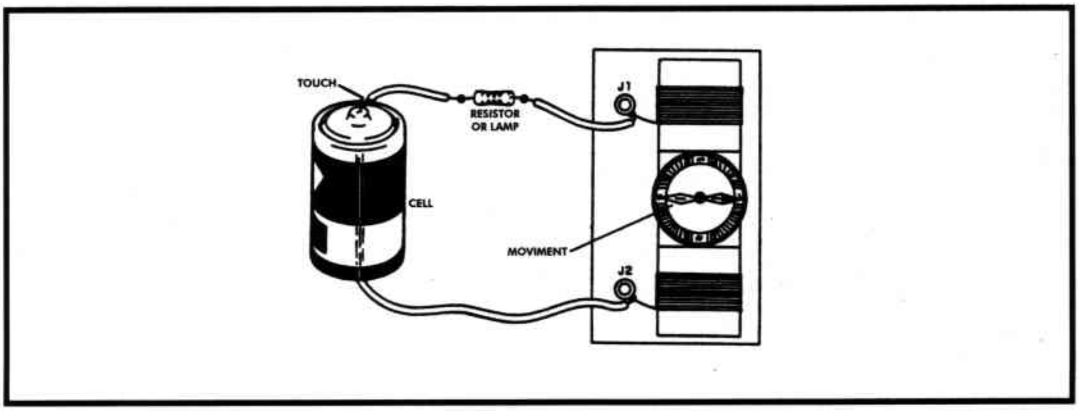
TOUCH
RESISTOR OR LAMP
J1
CELL
MOVIMENT
J2

Figure 2

Parts List — Experimental Galvanometer

L1	- coil - see text
X1	- Compass or blade - see text

Ideas to Explore

To get better performance or to learn more about the circuit:

- Experiment with several coils with different numbers of turns and compare sensitivity.

Science projects:

- Demonstrate how Oersted discovered the magnetic field around a wire that is carrying electricity and how modern galvanometers are constructed.
- Where can galvanometers can be used? What are voltmeters and ammeters?
- Determine the sensitivity of your galvanometer by connecting it in series with a multimeter (DC mA or uA scale) and a 100,000 ohm potentiometer.

OVER-TEMPERATURE ALARM (E/P)

This circuit will produce intermittent beeps by a piezoelectric transducer when the temperature rises over a preadjusted value. You can use this circuit in greenhouses, heaters, etc. The sensor is a common silicon diode but an NTC can also be used if you change some components' values.

Exchanging positions of R1 and P1 with D1 will cause the device to operate as an under-temperature alarm.

The intermittent signal is generated by two oscillators. IC1a (pins 1,2 and 3) is an inverter that controls IC1b and IC1c (oscillators). IC1c (pins 8, 9 and 10) is a very-low-frequency oscillator that provides the modulation rate for the second oscillator. The second oscillator (pins 5,6 and 4) has its frequency determined by R2 and C1. The produced audio tone can be altered by adjusting both C1 and R2 within a large range of values.

Power comes from a 3- to 15-volt power supply. For portable use a small 9-volt battery is recommended. Current drain is very low, about 0.5 mA when the tone is off. When the tone is on, current drain rises to about 5 mA.

A schematic of the Over-Temperature Alarm is shown in ***Figure 1***. IC1, a 4093B, has two gates operating as oscillators and two as buffer-inverters.

Components placement on a homemade printed-circuit board is shown in ***Figure 2***. You can made your prototype using a single-sided PC board.

Position of the polarized components, especially diode D1 (used as sensor), should be observed. This sensor can be placed far from the device by using common wiring. You must take some care to avoid humidity or water falling onto the sensor, which will cause adversely affect the alarm operation.

BZ is a piezoelectric transducer (Radio Shack 273-073 or equivalent) or a crystal earpiece.

Operation is adjusted by P1. Set this potentiometer to get sound at the desired temperature. To get a more precise adjustment you can replace the common trimmer potentiometer by a multi-turn potentiometer.

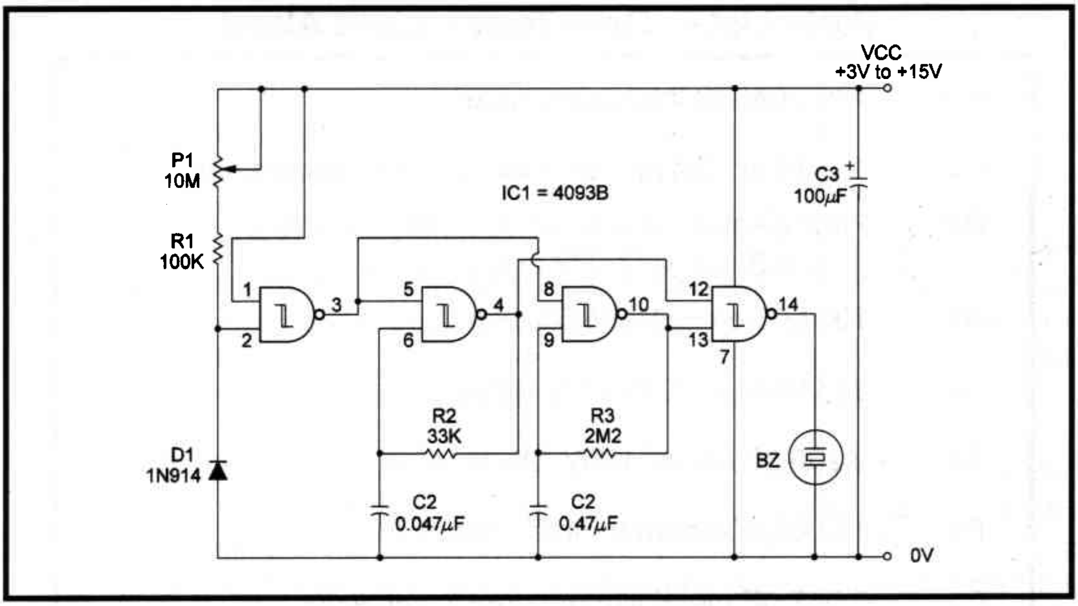
VCC
+3V to +15V
P1
10M
R1
100K
IC1 = 4093B
C3
100µF
1
2
3
5
6
4
8
9
10
12
13
14
7
R2
33K
R3
2M2
D1
1N914
BZ
C2
0.047µF
C2
0.47µF
0V

Figure 1

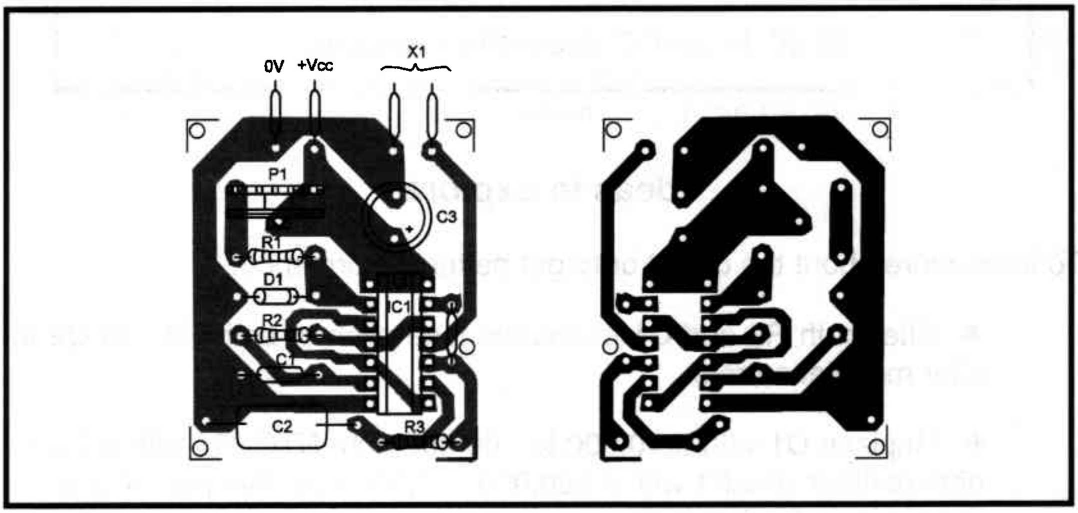
0V
+Vcc
X1
P1
C3
R1
D1
IC1
R2
C1
C2
R3

Figure 2

Get the sensor between your fingers to see how it works: when the temperature rises some degrees the circuit will produce a serie of beeps. Isolate the sensor, and after some seconds sound emission will stop.

Parts List — Over-Temperature Alarm

IC1	- 4093 CMOS integrated circuit
D1	- 1N914 or 1N4148 general purpose silicon diode
BZ	- Piezoelectric transducer or crystal earphone (Radio Shack 273-073 or equivalent)
R1	- 100,000 ohm, 1/4W, 5% resistor
R2	- 33,000 ohm, 1/4W, 5% resistor
R3	- 2,200,000 ohm, 1/4W, 5% resistor
P1	- 10,000,000 ohm trimmer potentiomer
C1	- 0.047 uF metal film or ceramic capacitor
C2	- 0.47 uF metal film or ceramic capacitor
C3	- 100 uF, 16 WVDC electrolytic capacitor

Ideas to Explore

To learn more about the circuit or to get better performance:

- Alter both R2 and C1 to change tone pitch or both R3 and C2 to alter modulation rate.

- Replace D1 with a 10,000 to 100,000 ohm NTC, R1 with a 2,200 ohm resistor, and P1 with a 100,000 to 1,000,000 ohm potentiometer, to get a new kind of sensor.

- Try to add a powerful output audio stage using a transistor, such as a BD135 and a small loudspeaker. See other projects in this book for suggestions about the configuration.

Science projects and other applications:

- Chemical processes that involve heat changes can be controlled using this alarm. The sensor should be protected against moisture and water in experiments involving liquids.

- Replace D1 by an LDR or photodiode. The circuit will operate as a light alarm or dark alarm, depending upon sensor position. With an LDR it would be better to replace P1 with a 1,000,000 ohm potentiometer and R1 with a 10,000 ohm resistor. Lenses placed in front of the LDR or phototransistor can be used to increase sensitivity.

- The circuit can also operate as a water-level alarm by using a two-wire sensor to replace D1. The sensor is made with two bare wires. The wires are separated by a distance of about 2 or 3 inches. The water completes the circuit while the wires are immersed. When the water level falls below the wires, the circuit is opened and the alarm is triggered.

- Temperature in greenhouses, animal pens, incubators and aquariums can be monitored using this device. And if working with an LDR as the sensor, experiments involving light can be conducted.

CAPACITOR TESTER (P)

This circuit can be used to test metal film ceramic capacitors and other types of capacitors in a range of values between 470 pF to 10 uF.

The device is formed by an audio-oscillator that produces a tone determined by the capacitor under test. Probes connect the capacitor to the circuit. The frequency of the produced tone is higher for lower-value capacitors. A falling tone indicates a bad capacitor.

Testing capacitors in the range between 1 and 10 uF, the oscillator will produce separated audio pulses, working like a metronome.

The circuit is powered by two or four AA cells, and current drain is very low. P1 is adjusted to set an audible tone at the speaker, according the value of the capacitor being tested.

You can also use this device to compare capacitor values. Two capacitors with the same value will give the same tone.

The schematic diagram of the Capacitor Tester is shown in ***Figure 1***. The circuit has as its heart two complementary transistors working as a direct-coupled amplifier. Feedback is provided by resistor R1 and the capacitor under test.

Components placement on a terminal strip is shown in ***Figure 2***. Use this parts placement as a guide when assembling the terminal strip and when housing the circuit into a box.

Termination components should be short as possible to avoid instability.

Position of the polarized components, such as the electrolytic capacitor and power supply, should be observed when mounting. The capacitor under test is placed in the circuit by two alligator clips as shown in Figure 2.

The circuit can be housed in a small plastic box.

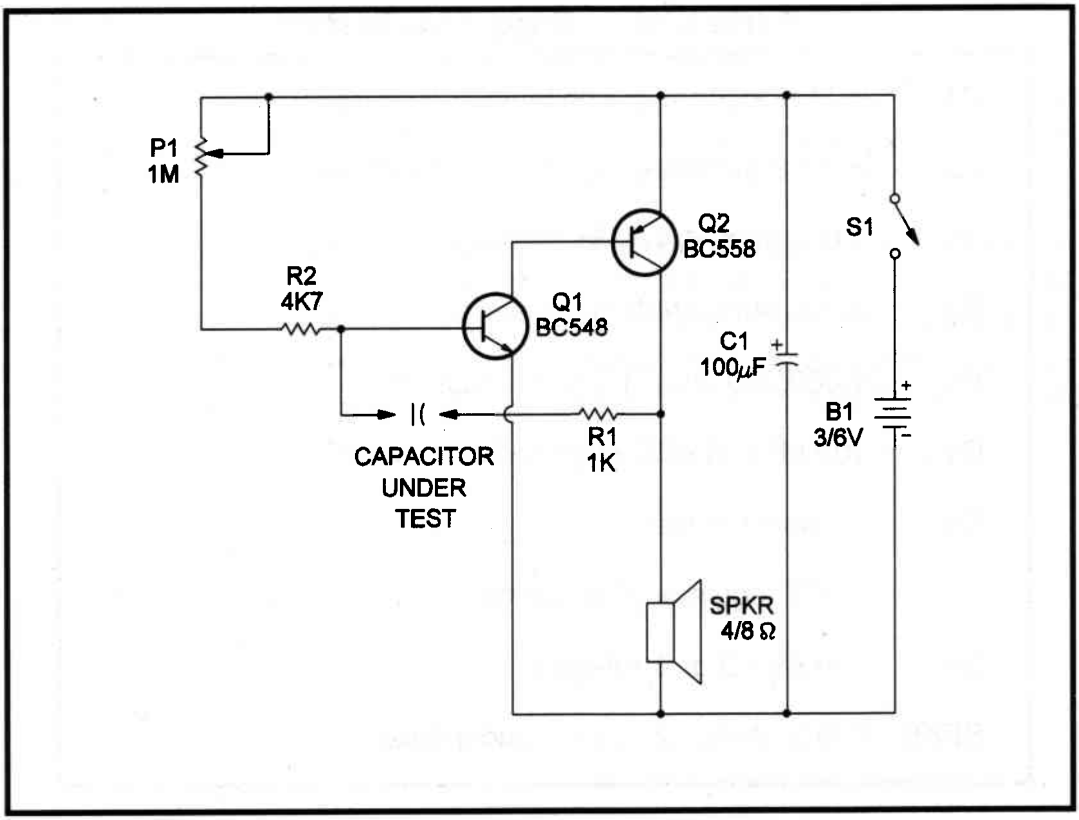
P1
1M
Q2
BC558
S1
R2
4K7
Q1
BC548
C1
+
100µF
B1
3/6V
+
-
CAPACITOR
UNDER
TEST
R1
1K
SPKR
4/8 Ω

Figure 1

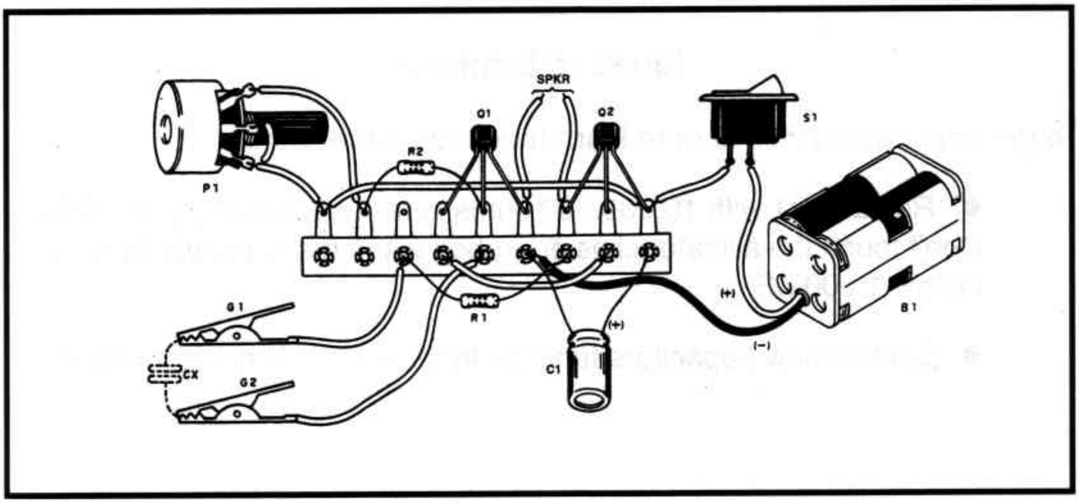
SPKR
Q1
Q2
S1
R2
P1
R1
(+)
B1
G1
CX
G2
(+)
(−)
C1

Figure 2

Parts List — Capacitor Tester

Q1 - BC548 general-purpose NPN transistor

Q2 - BC558 general-purpose PNP transistor

R1 - 1,000 ohm, 1/4W, 5% resistor

R2 - 4,700 ohm, 1/4W, 5% resistor

P1 - 1,000,000 ohm, 1/4W, 5% resistor

C1 - 100 uF, 6 WVDC electrolytic capacitor

Cx - capacitor in test

S1 - SPST toggle or slide switch

B1 - 3 or 6V - 2 or 4 AA-cells

SPKR - 4 to 8 ohms - 2 to 4 in. loudspeaker

Ideas to Explore

To get better performance or to learn more about the circuit:

- Replace P1 with 10,000,000 ohms to test capacitors under 470 pF. Don't touch the alligator clips or probes when testing small capacitors bellow 1,000 pF.
- Explain how capacitors function in this circuit. What is feedback?

Science projects and uncommon uses:

- Mount an experimental capacitor using aluminum foil and paper as dielectric. Test the capacitor using this device and compare the value with known capacitors.

- Experiments involving capacitive transducers can be conducted with this circuit. The circuit can be used as an analog-to-digital (A/D) converter. These circuits can convert capacitances (an analog quantity) into frequencies (a digital quantity). With an appropriate conversion scale it will be possible to use a frequency counter to measure capacitances.

CONTINUITY TESTER (E/P)

This circuit can be used to test electronic components such as diodes, resistors, coils, transformers, lamps, fuses, switches, loudspeakers and many others. The state of the component under test is indicated by an LED.

When the probes are separated or when the resistance between them is very high (more than 1,000,000 ohms) the LED is off. When the resistance is low (the transistor is in a conduction state) the LED is on.

The circuit is powered from a 6- or 9-volt supply (AA cells or battery). Current drain is less than 10 mA with the LED on. R2 value depends on the power supply voltage. For a 9V supply we recommend 820 or 1,000 ohms for R2.

A schematic diagram of the Continuity Tester is given in ***Figure 1***. Two transistor are wired as a Darlington pair. The configuration reduces the current flow between probes to microampere values.

Components placement on a terminal strip, used as chassis, is shown in ***Figure 2***. It is very important to observe positions of the transistors and LED.

House the device in a small plastic box and use probes or alligator clips to connect the components under test.

Parts List — Continuity Tester

Q1, 2	- BC558 general-purpose PNP silicon transistors
LED	- Common red, green or yellow LED
R1	- 100,000 ohm, 1/4W, 5% resistor
R2	- 470 ohm, 1/4W, 5% resistor - see text
S1	- SPST slide or toggle switch
B1	- 6V or 9V (four AA cells or battery)
PP1, 2	- Red and black probes

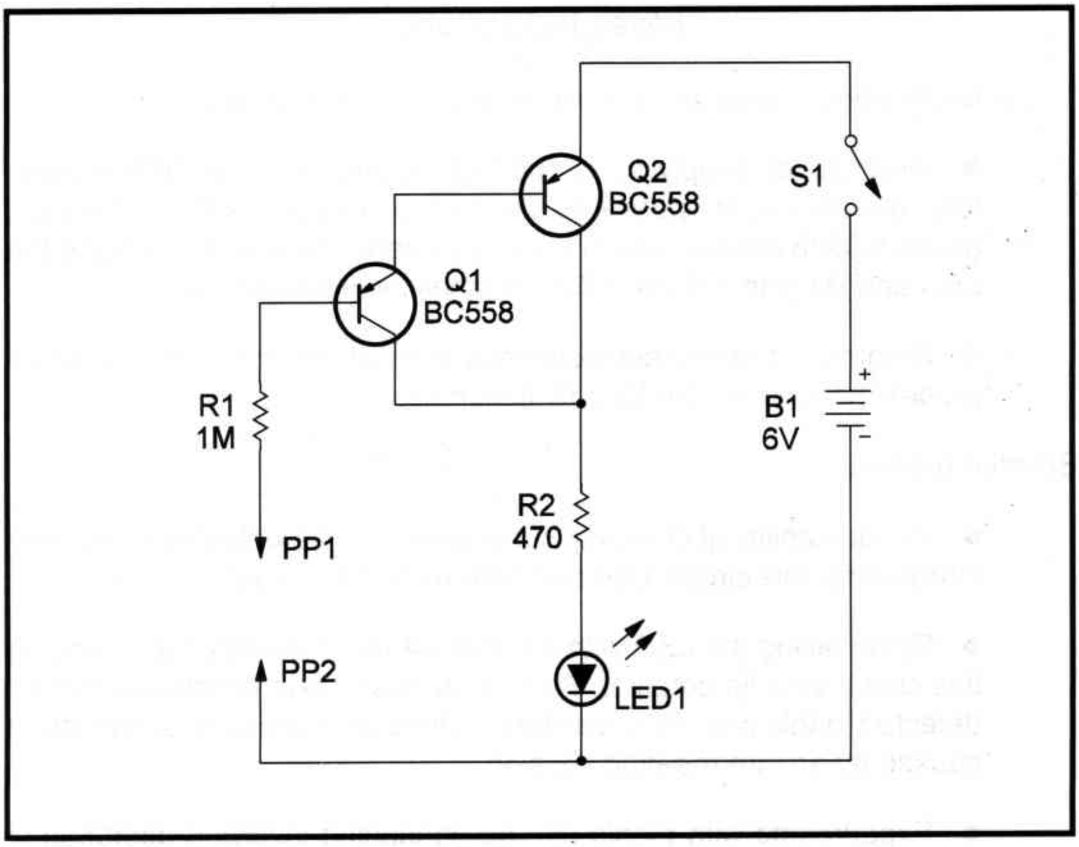
Q2
BC558
S1
Q1
BC558
R1
1M
+
B1
6V
-
R2
470
PP1
PP2
LED1

Figure 1

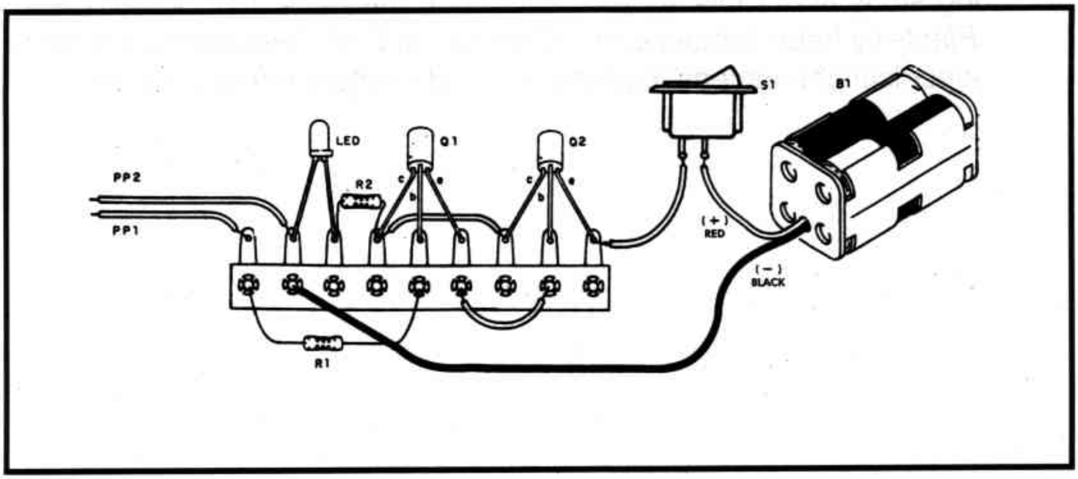
S1
B1
LED
Q1
Q2
PP2
R2
c
e
b
c
e
b
PP1
(+)
RED
(-)
BLACK
R1

Figure 2

Ideas to Explore

To get better performance or to learn more about circuit and devices:

- Invert power supply poles and LED position and use NPN transistors. Will the circuit operate as the original version? A transistor such as the BC548 can be used in this experiment. You can also replace the LED and R1 with a 6-volt x 50 mA small incandescent lamp.

- Determine the highest resistance that can be connected between probes that causes the LED to illuminate.

Science projects:

- Conductability of chemical substances can be detected or demonstrated with this circuit. Use two bare wires as probes.

- By replacing the LED with a 0-200 uA microammeter you can use this circuit as a lie detector. Small changes in skin resistance can be detected in this way. Skin resistance alters with stress or by the strain caused by an interrogation session.

- Experiments with plants can be conducted involving alteration in the resistance between electrodes placed on their leafs . An interesting study about this subject is found in the book *The Secret Life of Plants* by Peter Tompkins and Christopher Bird. The authors associate electrical phenomena in plants to a rudimentary nervous system.

LED BARGRAPH (E/P)

This circuit drives ten LEDs in response to an analog voltage applied to its input and can be used to provide a visual indication of the instantaneous power being developed by an audio amplifier. It can also be used to provide visual indication of variable events in science experiments.

You can extend the project by using two chips and you'll be able to continuously monitor the audio output power of both channels of your stereo system.

The circuit can be used with amplifiers ranging from 0.1 to 100 watts.

In some cases you should use a small transformer, wired as in ***Figure 1***, to isolate the input and match impedances. This is necessary when the amplifier has a low-output impedance. Resistor R_x is chosen according to the output power of the amplifier. Values are given in Table 1.

Table 1

Amplifier Output Power	Rx
0.1 to 1 watt	-
1 to 5 watts	10 ohms, 1/2 watt
5 to 20 watts	22 ohms, 1 watt
20 to 50 watts	47 ohms, 1 watt
50 to 100 watts	100 ohms, 2 watts

A schematic diagram of the LED bargraph is shown in ***Figure 2***. The circuit uses an LM3914 IC to drive the LEDs, converting voltages into one of ten outputs.

The components are placed on a printed-circuit board as shown in ***Figure 3***. If you'd like to build your circuit the same way, you can etch your own board using the solder and component-side templates shown in this figure.

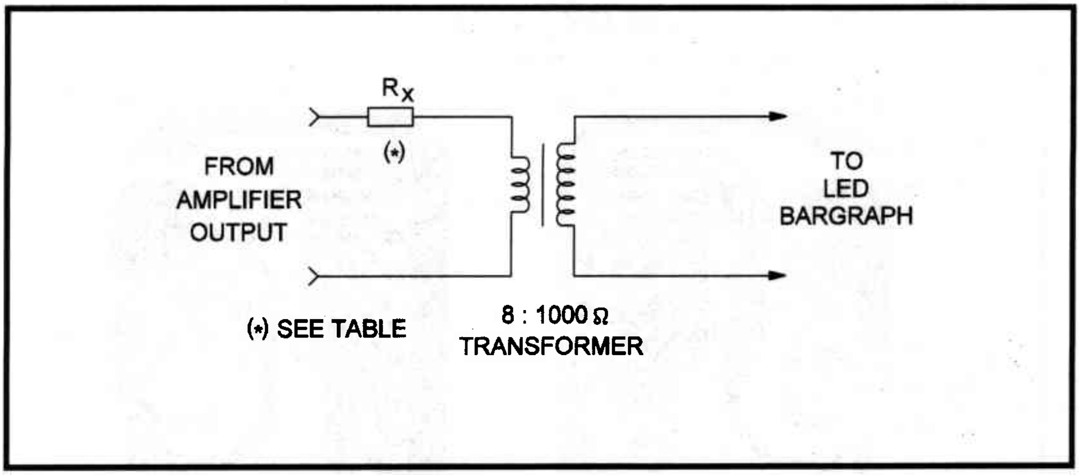
R_X
(*)
FROM
AMPLIFIER
OUTPUT
TO
LED
BARGRAPH
(*) SEE TABLE
8 : 1000 Ω
TRANSFORMER

Figure 1

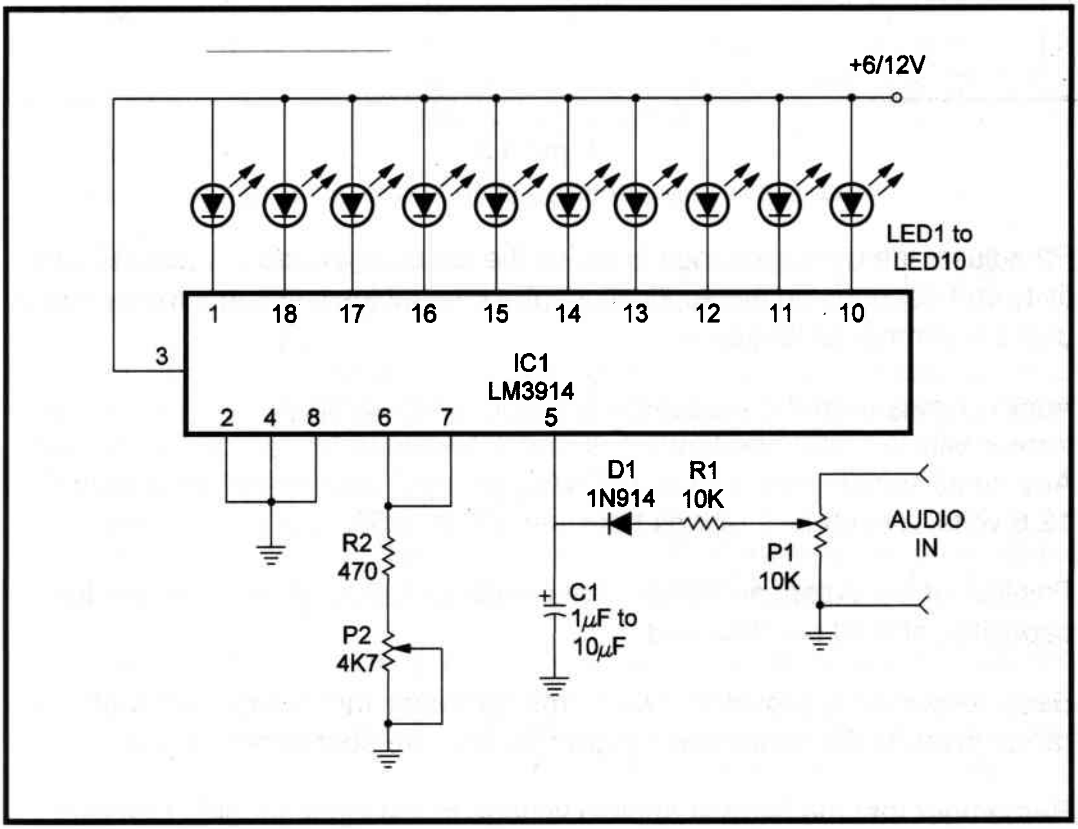
+6/12V
LED1 to
LED10
1
18
17
16
15
14
13
12
11
10
3
IC1
LM3914
2
4
8
6
7
5
D1
1N914
R1
10K
AUDIO
IN
P1
10K
R2
470
P2
4K7
C1
1µF to
10µF

Figure 2

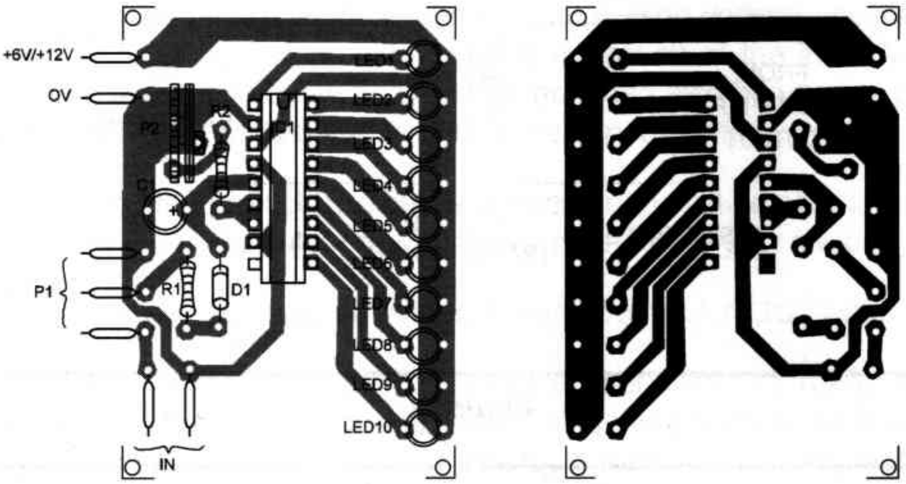

Figure 3

P2 adjusts the dynamic range in which the circuit operates. P1 adjusts sensitivity and depends on the application. You can set the value to a fixed point by using a trimmer potentiometer.

Audio comes from the loudspeaker output of the amplifier. Any small transformer with a high-impedance coil and a low-impedance coil can be used. Any small transformer with a 117 VAC primary, secondary rated from 5 to 12.6 volts, and current ratings between 100 and 500 mA can be used.

Position of the polarized components, such as LEDs, diode, and electrolytic capacitor, should be observed.

Bass response is provided by C1. You can alter this component within the range given in the schematic diagram to achieve best performance.

Remember that the highest applied voltage to the input shouldn't exceed the power supply voltage.

Parts List — LED Bargraph

IC1	- LM3914 Integrated circuit (National)
LEDs	- Common red, yellow or green
D1	- 1N914 or 1N4148 general purpose silicon diode
R1	- 10,000 ohm, 1/4W, 5% resistor
R2	- 470 ohm, 1/4W, 5% resistor
P1	- 10,000 ohm potentiometer
P2	- 4,700 ohm trimmer potentiometer
C1	- 1 to 10 uF, 12 WVDC electrolytic capacitor - see text

Ideas to Explore

To learn more about the circuit or to get better performance:

- Determine the voltage level that turns on each LED using a DC power supply plugged to the input of the device.
- By wiring probes at the positive power supply line and the input you can use this circuit as a digital resistance or conductance meter.

Science projects:

- This circuit can replace the common multimeter to indicate voltage variations in experiments. The visual effect is better.
- It can also be used in the output of a lie detector or bio-feedback.
- Conect an LDR between the input and (+) for a digital photometer. For a digital thermometer wire an NTC between (+) and the input.
- This circuit can also replace the LED and resistor in the Continuity Tester and the same experiments, with advantages, can be performed.

LIQUIDS SHAKER (P/E)

A toy's small DC motor can be used to mount a liquids shaker. The shaker can be used in chemical experiments to mix substances quickly and efficiently.

The circuit is powered from the power supply line (to avoid the expense of batteries), but there is no shock hazard because an isolation transformer is used. Of course, for fieldwork you can also power your mixer from common AA cells.

The transformer is chosen according to the motor voltage. A four-AA-cell DC motor requires a 6+6 volt x 500 mA transformer (a 6-volt center tapped transformer). A two-AA-cell DC motor requires a 3V x 500 mA transformer.

Secure the motor into a small cylindrical plastic box. Use a long shaft to attach a small plastic screw propeller.

Figure 1 shows the schematic diagram of the electronic part of the circuit: a power supply. The diodes are common 50V x 1A silicon rectifiers, and C1 is not critical. C1 should be in the range between 100 and 1,000 uF.

Figure 2 shows a mechanical view, using a small terminal strip as chassis. The fuse is important to avoid severe troubles if shorts occur.

R1 is a current-limiting resistor and can be altered according to the motor. Values between 4.7 and 100 ohms can be experimented with for better performance.

Position of the polarized parts, such as the diodes and electrolytic capacitor, should be observed. Remember that the motor rotates clockwise or not according to the voltage polarization.

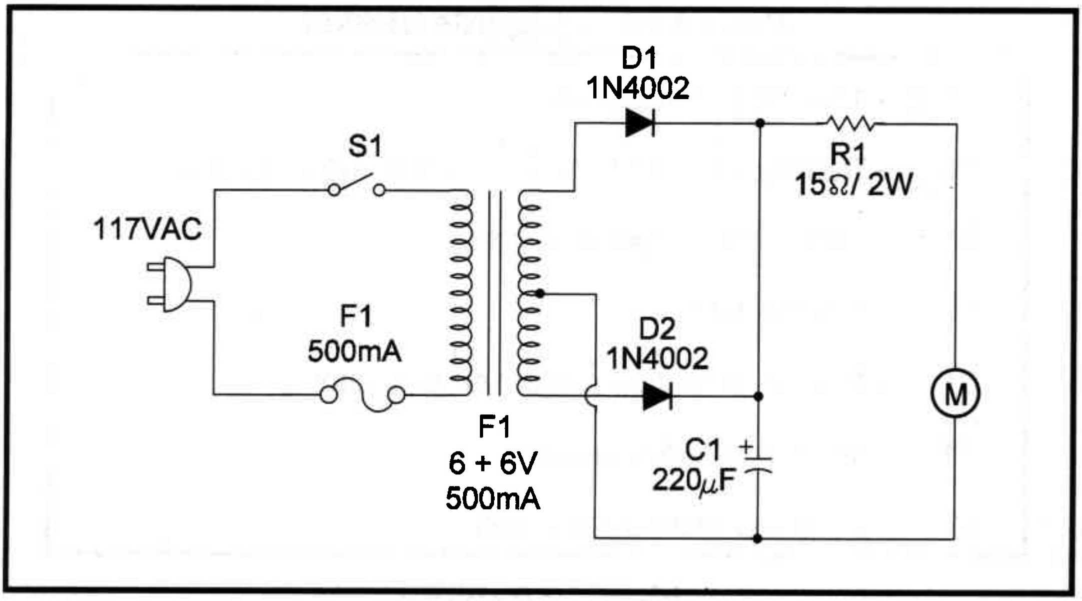
D1
1N4002
S1
117VAC
R1
15Ω/ 2W
F1
500mA
D2
1N4002
M
F1
6 + 6V
500mA
C1
220μF
+

Figure 1

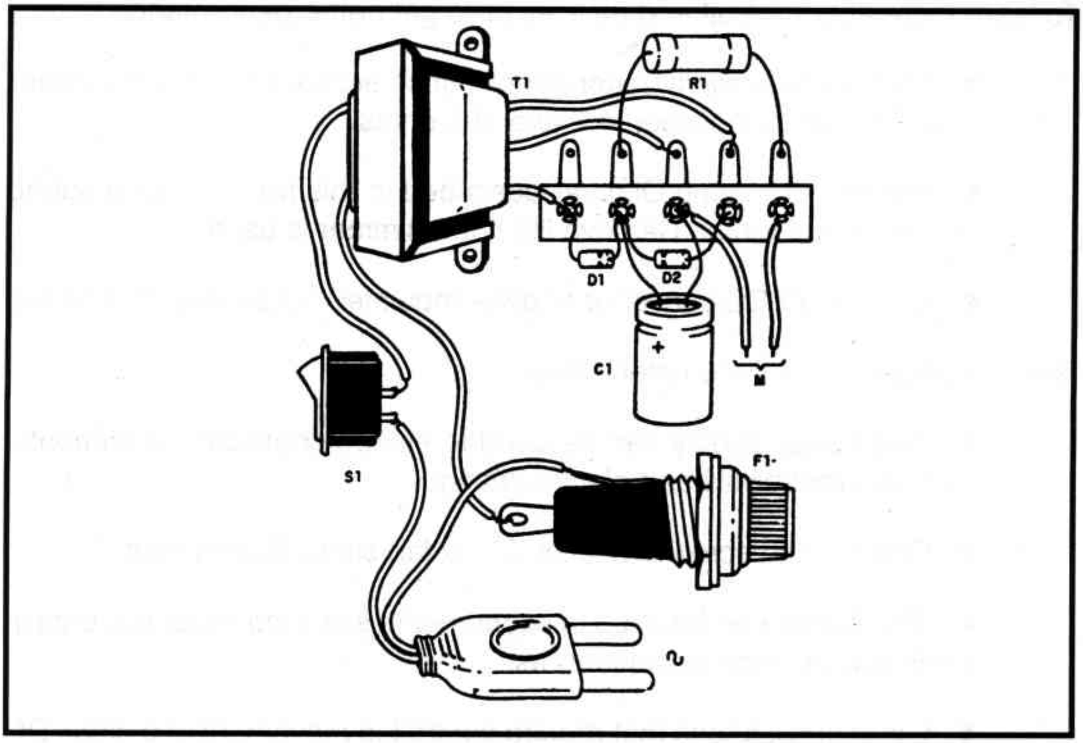
T1
R1
D1
D2
C1
+
M
S1
F1
∿

Figure 2

Parts List — Liquids Shaker

D1, 2	- 1N4002 silicon diodes
T1	- Transformer: 117 VAC: 6+6V x 500 mA - see text
S1	- SPST slide or toggle switch
F1	- 500 mA fuse
C1	- 220 uF, 16 WVDC electrolytic capacitor
R1	- 15 ohm, 2W, 5% resistor
M	- small DC motor (3 to 6 volt)

Ideas to Explore

To learn more about circuit and devices or to get better performance:

- Alter R1 to achieve better performance according to the DC motor used. Try values between 4.7 and 100 ohms.
- Use the DC Lamp Dimmer described in this book to get a speed control for the motor. Remove R1 if the dimmer is used.
- Use the 6V DC converter to get a regulated output from the circuit.

Science projects and uncommon uses:

- This power supply can be used in electro-chemical experiments, such as electrolysis and electroplating.
- Replace the screw propeller with a fan blade. Experiment.
- The device can be used in experiments involving water currents in small places, such as aquariums.
- Other applications that require controlled movement of a small DC motor can use this circuit.

SIGNAL TRACER (P)

Signal tracing by means of a signal tracer offers a valuable method for servicing a variety of electronic equipment. This method is often employed when a circuit is not completely dead, but rather malfunctioning. AM and FM receivers, CD players, multimedia amplifiers, TV sets, etc., can be serviced using this simple signal tracer.

The circuit we show here consists of a two-stage direct-coupled transistor amplifier. The circuit has two inputs. When the probe is used in the RF mode the detector diode is inserted into the circuit. When the probe is used in the audio mode the signal is applied directly to the base of the first transistor (Q1).

As the circuit has a low-power output, a gain control is not used.

The circuit is powered from two AA cells, and current drain is only few milliamperes. Since the circuit draws such low current, long battery life can be expected.

Figure 1 shows the schematic diagram of the signal tracer. R1 determines the gain of the amplifier stage and can be altered to get better performance depending on the transistors used.

Figure 2 shows the signal tracer mounted on a terminal strip, which is used as a chassis. Take care with the transistors; one is an NPN and the other is a PNP. If you switch one with the other the circuit will not work.

This project can be housed in a small plastic box. The size of the plastic box is determined by the speaker size. Position of the polarized pieces should be observed.

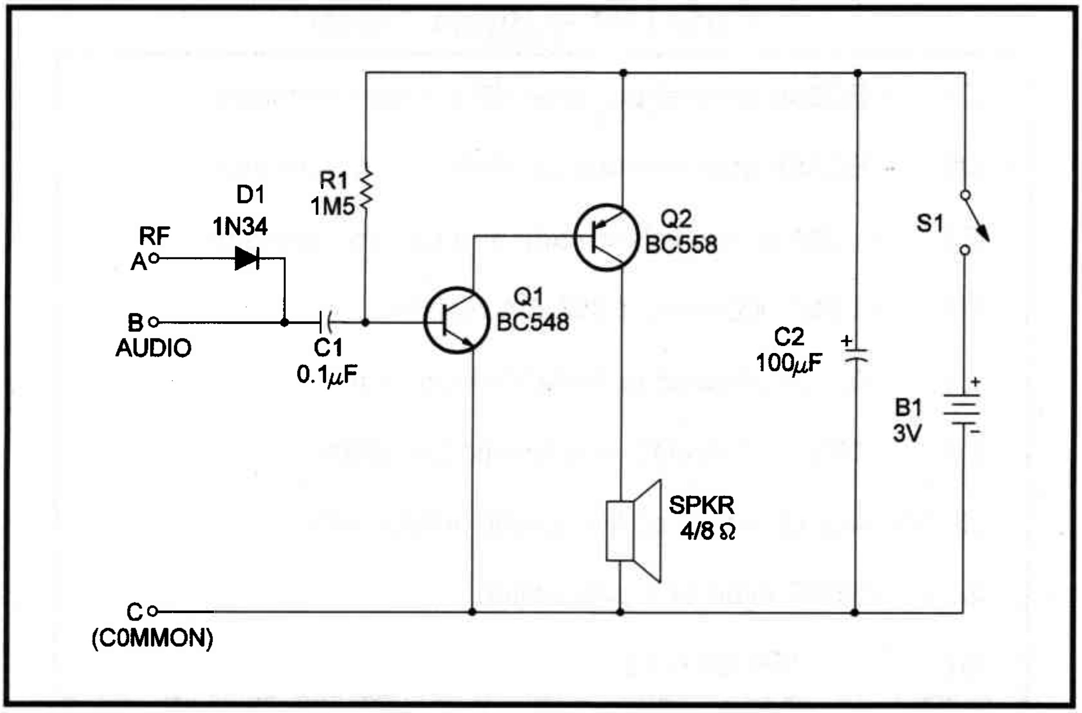
R1
1M5
D1
1N34
RF
A
B
AUDIO
C1
0.1μF
Q1
BC548
Q2
BC558
S1
C2
100μF
B1
3V
SPKR
4/8 Ω
C
(COMMON)

Figure 1

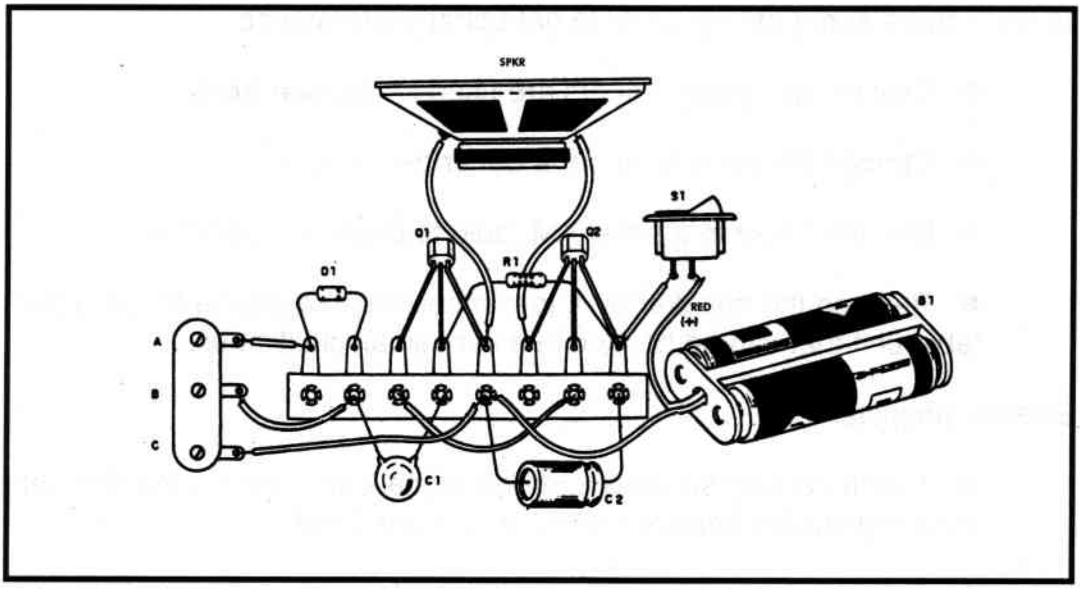
SPKR
S1
Q1
Q2
R1
D1
RED
(+)
B1
A
B
C
C1
C2

Figure 2

Parts List — Signal Tracer

Q1 - BC548 general-purpose NPN silicon transistor

Q2 - BC558 general-purpose PNP silicon transistor

D1 - 1N34 or equivalent - any germanium transistor

R1 - 1,500,000 ohm, 1/4W, 5% resistor

C1 - 0.1 uF ceramic or metal film capacitor

C2 - 100 uF, 6 WVDC electrolytic capacitor

SPKR - 4/8 ohms - 2 or 4 in. small loudspeaker

S1 - SPST slide or toggle switch

B1 - 3V - two AA cells

Ideas to Explore

To learn more about the circuit or to get better performance:

- Explain how signal tracers are used for service work.
- Change R1 value to achieve better performance.
- Use the tracer in a small AM radio to study its operation.
- Replace the speaker by a low-impedance earphone to get a portable unit that can be housed in a very small plastic box.

Science projects:

- The tracer can be used to detect signals produced in experiments involving current fields, as described in this book.
- The tracer can be used as an experimental audio amplifier with transducers such as piezoelectric pickups or microphones.

ONE-EVENT REGISTER (P/E)

This is a one-bit memory that can be used to store a simple bit of information, such as a remote power-failure or closed-trap detector, or a touch sensor.

The circuit operates as follows: When the resistance between the alligator clips falls, the SCR is triggered, powering ON the LED.

The SCR and LED stay in the ON state even after the applied pulse to the input is gone. So the device “memorizes” the event.

The sensor can be a magnetic switch (reed switch), a microswitch, or a touch switch. A simple touch switch is a metal plate wired to G2. To get best sensitivity, G1 needs a good ground. You can also use as sensor two metal plates that are touched at the same time to trigger the circuit ON.

The circuit is powered by four AA cells and current drain is very low when the LED is off. When the LED is on, current drain is less than 10 mA.

To reset the device you only have to toggle S1. Another way to reset the circuit is to wire a momentary switch between the anode and cathode of the SCR. Sensors can be wired to the device with long unshielded wires.

The schematic of the One-Event is shown in ***Figure 1.*** Though R2 and R3 are shown as fixed, you can use a 1,000,000 ohm potentiometer if you prefer. A sensitivity adjustment will get better performance from the register.

Figure 2 shows the One-event Register, on terminal strip mounting. As it is an experimental circuit you can also use a solderless board to install the components.

The circuit can easily be housed into a small plastic box. Observe the positions of all the polarized components, such as the LED, SCR and power supply.

Resistor R3 is only necessary if TIC106 is used. Other SCRs, such as the MCR106 and C106, don’t require R3.

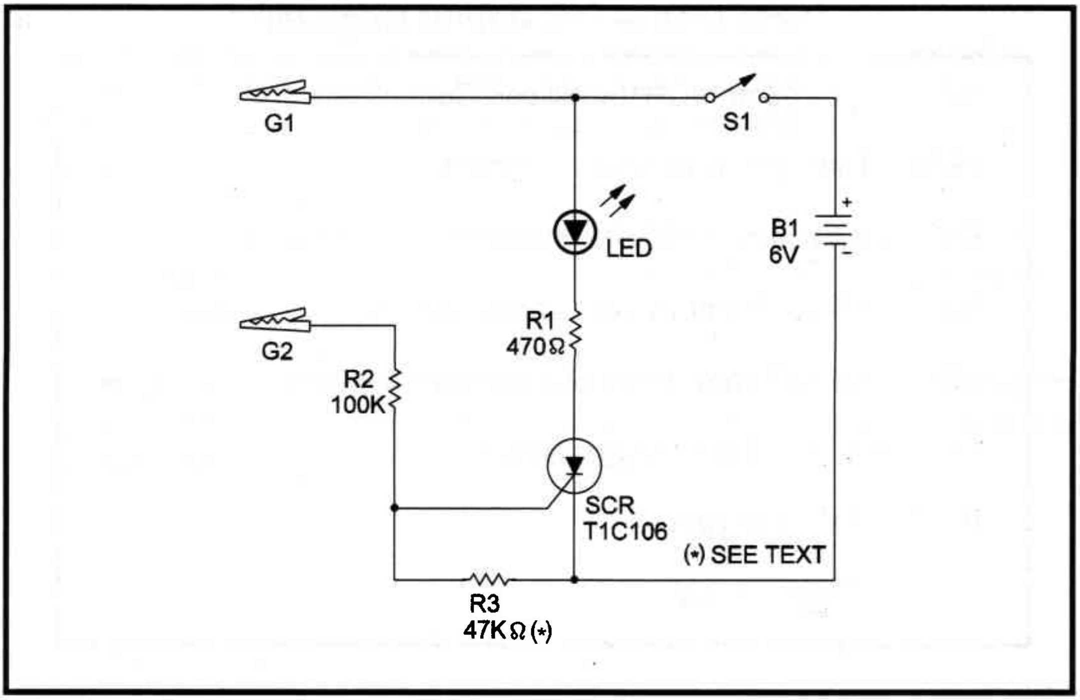
G1
S1
LED
B1
6V
+
-
G2
R1
470Ω
R2
100K
SCR
T1C106
(*) SEE TEXT
R3
47KΩ (*)

Figure 1

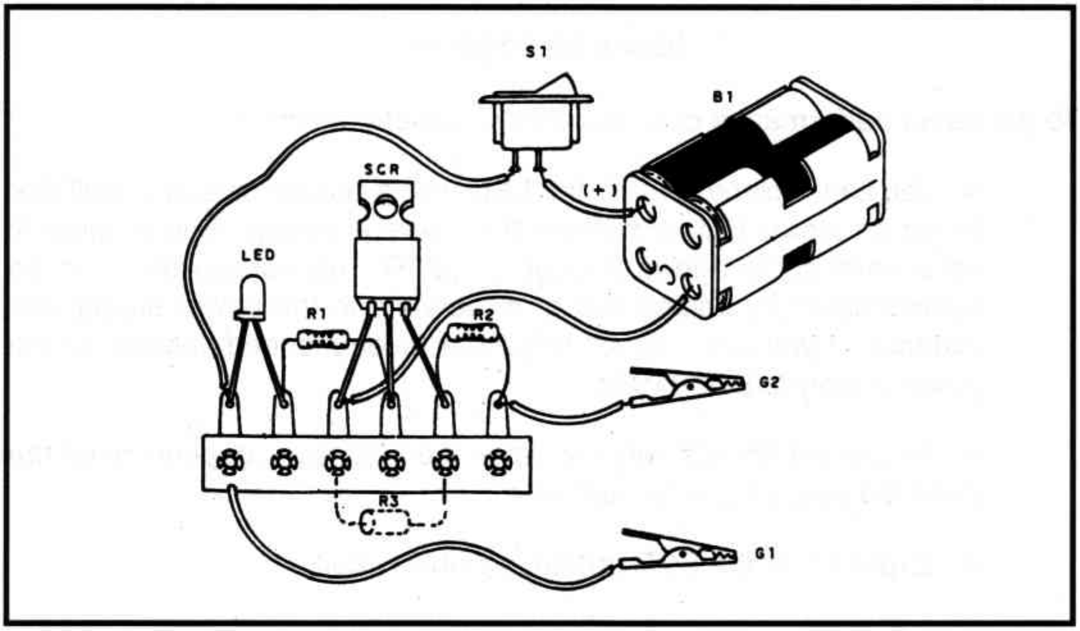
S1
B1
SCR
(+)
LED
R1
R2
G2
R3
G1

Figure 2

Parts List — One-Event Register

SCR	- TIC106 or MCR106 Silicon Controlled Rectifier
LEDs	- Red, yellow or green common
R1	- 470 ohm, 1/4W, 5% resistor
R2	- 100,000 ohm, 1/4W, 5% resistor
R3	- 47,000 ohm, 1/4W, 5% resistor - see text
S1	- SPST slide or toggle switch
B1	- 6 V - four AA cells
G1, 2	- Alligator clips

Ideas to Explore

To get better performance or to learn more about the circuit:

- You can replace R1 and the LED with a buzzer or audio oscillator to get an alarm. But, remember that there a voltage drop of about 2 volts when current flows through the SCR. This voltage drop can be compensated by adding two or three volts to the power supply. For instance, if you use a 6-volt relay, the recommended voltage for the power supply is 8 or 9 volts.

- As current drain is very low when the LED is off, depending on the intended use, S1 can be removed.

- Explain how the SCR operates in this circuit.

Science and uncommon applications:

- You can connect an LDR between G1 and G2 for a light-operated one-event memory or alarm. A single short pulse of light can trigger the circuit.
- Use this circuit to know when someone touches an object when you leave it unattended. You can also know if someone entered your room when you were out.
- By using two bare wire as sensors, you have a water-level alarm. The LED will glow when the water touches the two bare wires at the same time.

NOISE GENERATOR (E)

The purpose of this simple experiment is to produce noise in the AM radio band. The circuit can be used to show how brush motors and general appliances with motors can generate noise in the radio band, interfering in radio, TV, and radio communication services.

The circuit is very simple and the produced noise is too weak to cause any problem to radio receivers. (FCC rules about noise and radio interference are severe.) The noise produced by this experiment will interfere only in receivers just a few feet away.

The schematic diagram of the device is shown in ***Figure 1***. The heart of the circuit is a small DC motor that works as a automatic key, producing rapid variation in the current flow.

A mechanical view of the circuit, using a terminal strip as chassis, is shown in ***Figure 2***. All the components, including the power supply, can be fixed on a plastic board.

Any small DC motor obtained from a nonworking toy can be used in this project. The power supply depends upon the motor used.

CV and L1 can be obtained from any nonworking AM radio.

You can get more power for the RF noise by wiring a piece of wire to the antenna (A) terminal, but limit this wire to no longer than 6 feet. L1 is formed by 100 turns of No. 28 wire wound on a ferrite rod (diameter and length are not critical).

By adjusting CV, you can try to tune the signal to be stronger on a dead point of the AM radio band.

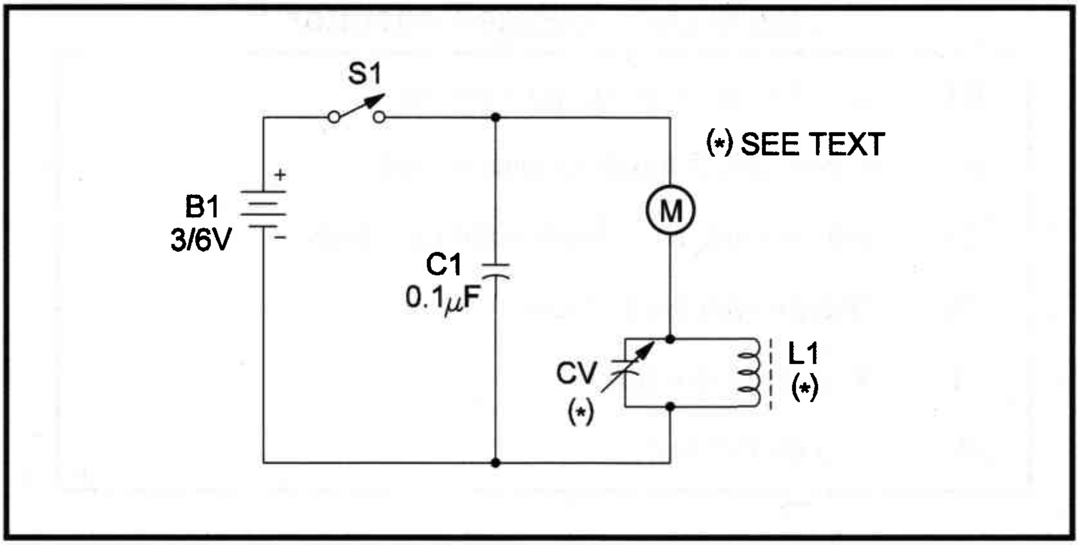
S1
B1
3/6V
+
-
C1
0.1μF
(*) SEE TEXT
M
CV
(*)
L1
(*)

Figure 1

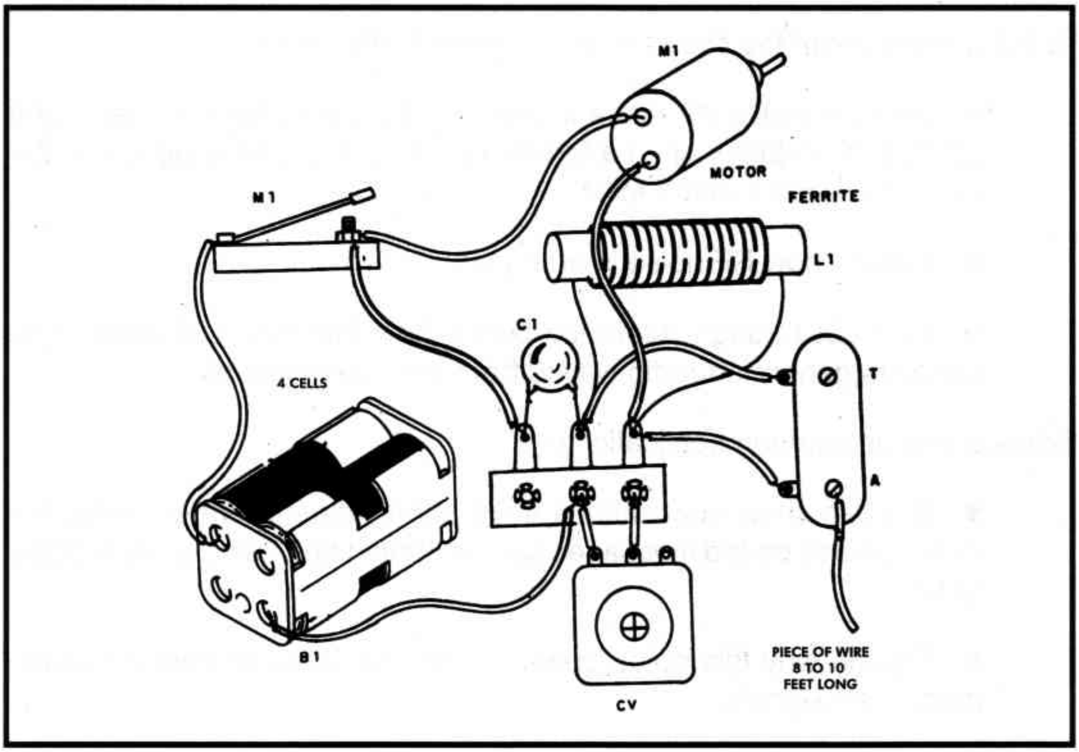
M1
MOTOR
FERRITE
M1
L1
C1
4 CELLS
T
A
B1
PIECE OF WIRE
8 TO 10
FEET LONG
CV

Figure 2

Parts List — Noise Generator

B1	- 3 to 6V - cells or battery - see text
T	- Key or SPST toggle or slide switch
C1	- 0.1 uF ceramic or metal plate capacitor
CV	- 365 pF variable capacitor
L1	- loopstick - see text
M	- 3 to 6V DC motor

Ideas to Explore

To learn more about the circuit or to get more performance:

- Try to produce RF noise in the high-frequency band by replacing L1. With a 30-turn coil on a ferrite rod you can produce signals in the range between 2 and 7 MHz.
- Explain how RF noise is produced.
- Find out through reference books how Marconi and other radio pioneers generated radio waves from their transmitters.

Science and uncommon applications:

- You can show how antique radio transmitters operated using this circuit. Morse coded messages can be transmitted using a key in place of S1.
- Explain why this noise generator can be used to transmit voice-modulated signals.

EXPERIMENTAL OHMMETER (E)

As commercial ohmmeters are cheap and easy to find, anyone who wants a good instrument for his own use will not build one. But you can build a simple ohmmeter to demonstrate how it works or study how the big ones are made.

The simple ohmmeter described here can easily be made from a milliammeter or microammeter, and can be used in experiments, demonstrations and many other applications.

Our ohmmeter is also a voltmeter. And since this ohmmeter can measure a large range of voltages and resistances we can say that it is, in practice, a multimeter. Voltages can be measured in the range between 0 and 12 volts (the range used to power almost all circuits in this book), and resistances between 0 and infinite.

If using a 200 uA microammeter, the sensitivity will be 5,000 ohms per volt. This value is typical of several commercial types of common multimeters.

But you can use any other ammeter with a full-scale value ranging from 100 uA to 1 mA. Simply change R1 to achieve best performance.

The schematic of the experimental ohmmeter is shown in ***Figure 1***. Note that we use only four components that can easily be housed in a small plastic box.

Layout is shown in ***Figure 2***. Components mounting is not critical, and you can replace the banana jacks by probes or any other method to make external connections.

Probes are connected to the external banana jacks depending on the measurement you intend to make. To measure voltage, the red probe is connected to J1 and the black probe is connected to J2. To measure resistance, connect the red probe to J1 and the black probe to J3. Put the probes together and adjust P1 to get an indication of zero at the instrument scale.

By using a series of known resistors, it is possible to calibrate the meter scale to read ohms directly. Start from 0 (with J1 and J2 shorted), and note the meter readings when 1k, 2k, etc, resistors are connected across J1 and J3.

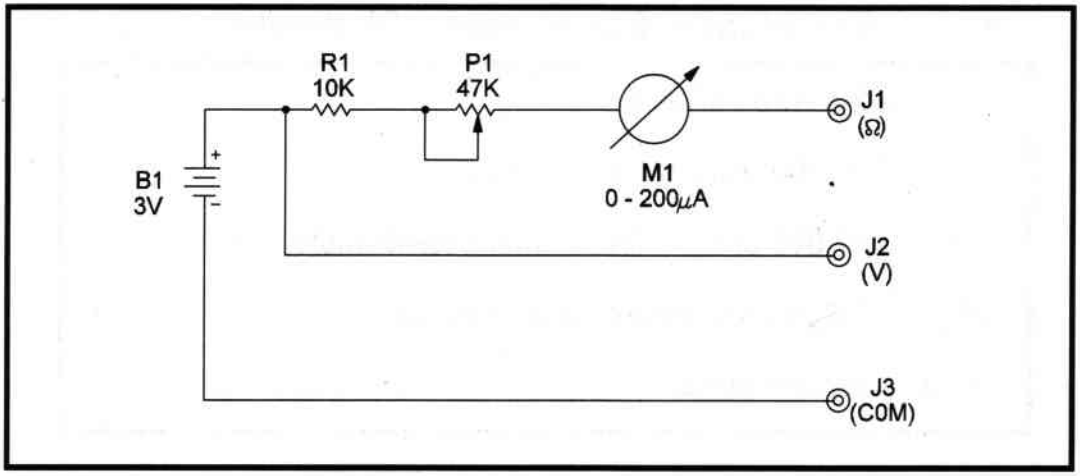
R1
10K
P1
47K
B1
3V
+
-
M1
0 - 200μA
J1
(Ω)
J2
(V)
J3
(COM)

Figure 1

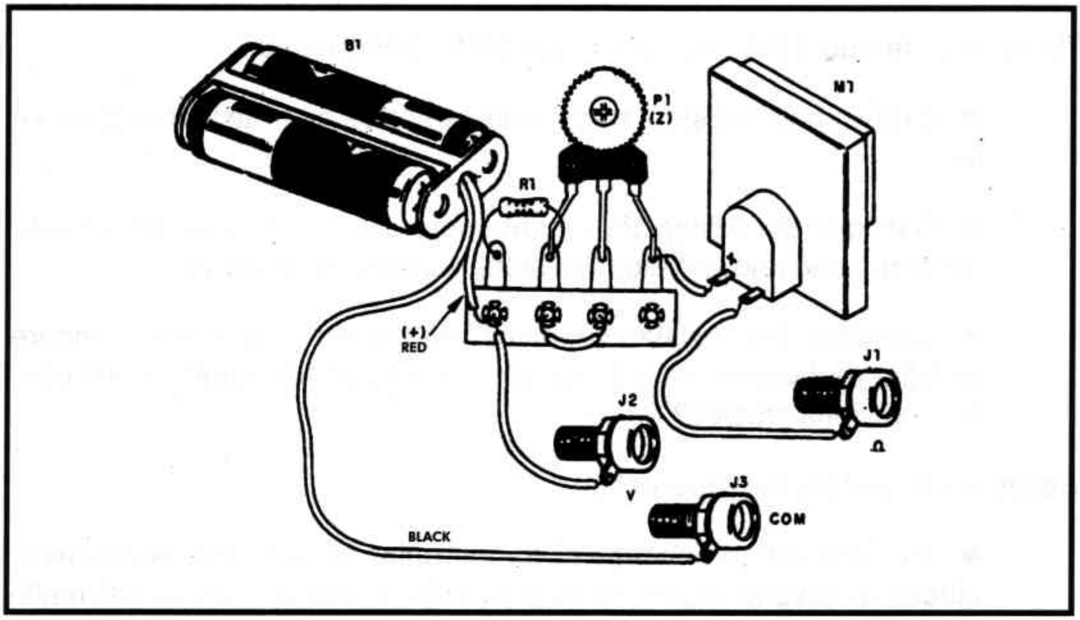
B1
P1
(Z)
M1
R1
(+)
RED
J1
Ω
J2
V
J3
COM
BLACK

Figure 2

After calibrating the device to read ohms, use a variable power supply to calibrate the voltage scale.

Position of the polarized components, such as the ammeter and power supply (cells), must be observed.

Parts List — Experimental Ohmmeter

B1	- 3V - two AA cells
R1	- 10,000 ohm, 1/4W, 5% resistor
P1	- 47,000 ohm, 1/4W trimmer potentiometer
M1	- 0-200 uA microammeter - see text
J1-3	- banana jacks

Ideas to Explore

To learn more about the circuit or to get better performance:

- Explain how resistance and voltage scales operate in this ohmmeter.
- Calibrate the scales of your ohmmeter using a variable power supply and some known-value resistors, as described above.
- Calculate the resistance values read in the meter and compare with the real values of used resistors. Do your calculated results confirm previous studies?

Science and uncommon applications:

- You can use the ohmmeter to determine sensor and media resistances in several experiments if you don't have a commercial multimeter.
- By wiring a 47,000 ohm potentiometer between J1 and J3 you can use this device as a position indicator. Cursor positions can be directly converted into current as indicated by the meter. A scale of angles can be arranged to read degrees directly.

TEST YOUR NERVES (P/E)

This circuit puts your manual skills to a lively test. As shown in the schematic diagram, the circuit has a small loop that you must navigate around a wire. The object of the game is to guide the loop over the weaving course without touching the wire.

A slight misdjudgement or quiver of the hand and the ring will contact the weaving wire, enabling the circuit that will produce a noticeable (but inoffensive) shock!

The skill required to play the game depends largely on the size of the loop and the degree of twist and turn in the wire.

Scoring is a matter of counting the number of times the player is affected by a shock! The person with the lowest total (or no touch) wins.

The circuit is powered from a D cell, as the current drain is high when the loop touches the wire. When not in use, don't leave the ring and wire together. This will cause the battery to run down in a short time.

A circuit diagram of the device is shown in ***Figure 1***. The heart of the project is the transformer that converts 1.5 VDC into high-voltage pulses up to 400 volts.

A mechanical view of the mounting is shown in ***Figure 2***.

All the components can be housed in a small plastic box. Wires to the loop and weaving wire should be 2 or 3 feet long to prevent pull-outs when the player is struck.

Observe the isolation between the points where the player touches and the weaving wire. The two wires that connect X2 should be twisted.

T1 is any small transformer with a 117 VAC primary, and secondary coil ranging between 3 and 9V. Current drain can range from 100 to 500 mA when the wire touches the loop.

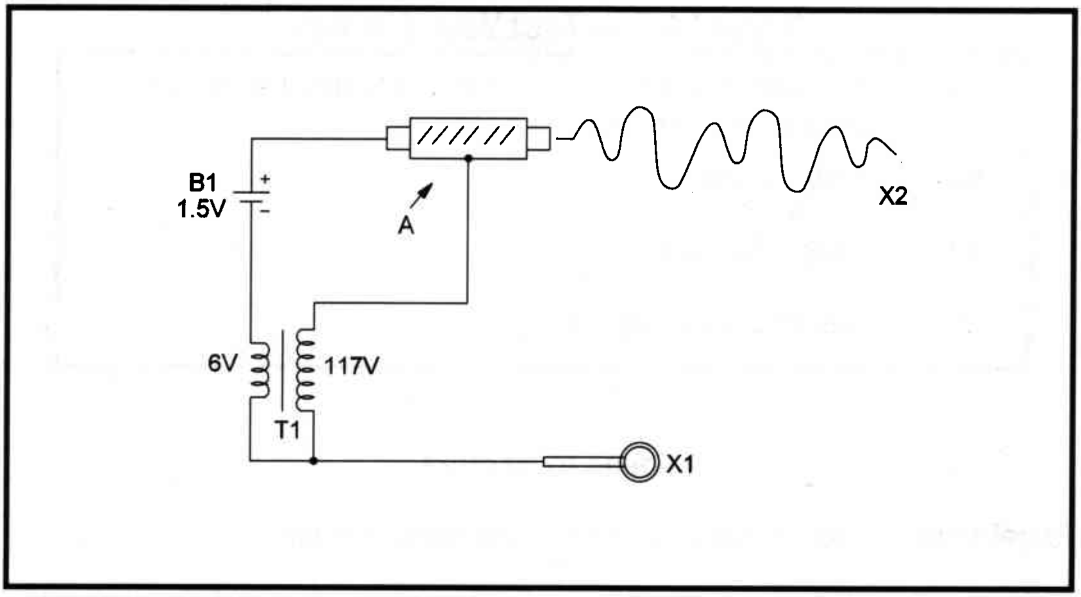
B1
1.5V
+
-
A
X2
6V
117V
T1
X1

Figure 1

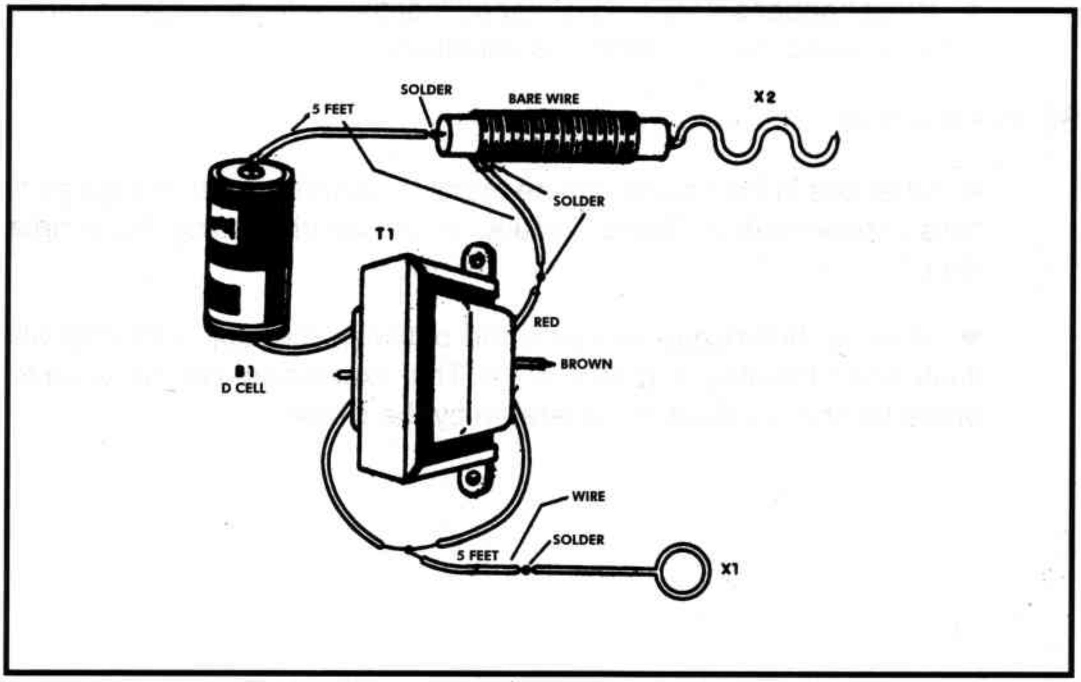
SOLDER
5 FEET
BARE WIRE
X2
SOLDER
T1
RED
BROWN
B1
D CELL
WIRE
SOLDER
5 FEET
X1

Figure 2

Parts List — Test Your Nerves

T1	- Any transformer with a 117 VAC coil and low-voltage secondary - see text
B1	- 1.5V - D cell
X1	- Loop - see text
X2	- Weaving wire - see text

Ideas to Explore

To get better performance or to learn more about the circuit:

- Explain why the transformer does not operate with DC but that this project is able to generate high voltage using that device.
- What happens if the loop is placed together with the weaving wire? Why it is necessary to avoid this condition?

Science projects:

- Changes in the project can be made to generate high voltage from cells in experiments. Nerve stimulation can be tried using this simple circuit.
- Neon or fluorescent lamps wired between the loop and wire will flash when the loop and wire touch. This experiment can be used to prove that high voltage is generated by the circuit.

WIRELESS BEEPER (E)

This FM wireless transmitter produces a beep that can be tuned on any dead point of the FM broadcast band. The circuit can be used as a localizer when attached to objects or persons. You also can use it as a wireless alarm.

The signal can be tuned on common FM radios in a range up to 150 feet. Alterations for a far-reaching version are given below.

The circuit produces an intermittent beep of about 1 Hz. An audio tone is generated by IC1b (pins 5, 6 and 4) and its frequency is determined by R2 and C2. You can easily alter this frequency by adjusting R2. Resistor values between 22,000 and 100,000 ohms can be experimented with or, if you prefer, 100,000 ohms in series with 10,000 ohms can replace R1.

Repetition rate is per R1 and C1. Values are variable. For C1, values between 0.15 and 1 uF can be tried. It is important to test for the ideal value of this component, depending on the intended application.

The schematic diagram of the wireless beeper is shown in ***Figure 1.*** Three gates of a 4093 IC are used to generate the low-frequency signals, and one RF transistor is used to generated the high-frequency signal.

Components placement on a homemade printed-circuit board is shown in ***Figure 2.*** Placement is critical for the high-frequency stage. Capacitors in this stage should be ceramic.

L1 consists of four turns of No. 18 or 20 wire on a 1/2 in. form without a ferrite core.

The antenna is formed by a piece of bare wire between 5 and 10 inches long. It can be connected to any tap on the coil. Experiment to get better performance.

The power supply current drain is about 10 mA. You can use four AA-cells or a 9V battery to power the transmitter.

For a wireless alarm, S1 can be replaced by the contacts of a relay. The relay is used to interface the alarm with the transmitter in this application.

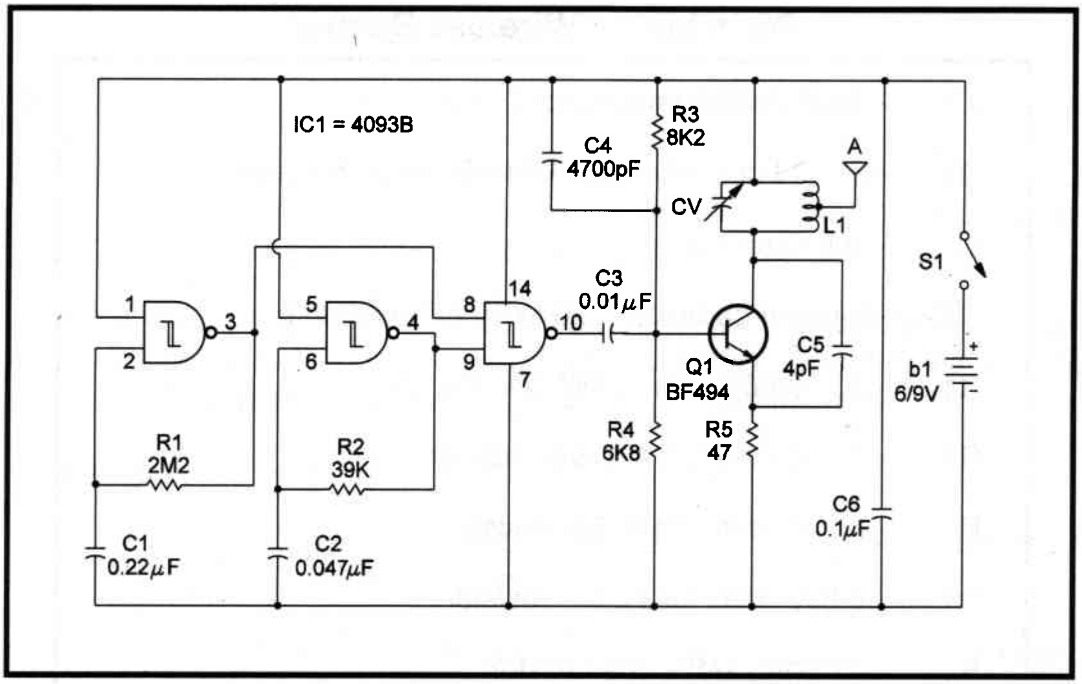
IC1 = 4093B
R1
2M2
R2
39K
C1
0.22µF
C2
0.047µF
C3
0.01µF
C4
4700pF
R3
8K2
R4
6K8
R5
47
CV
L1
A
Q1
BF494
C5
4pF
C6
0.1µF
S1
b1
6/9V
1
2
3
5
6
4
8
9
10
14
7

Figure 1

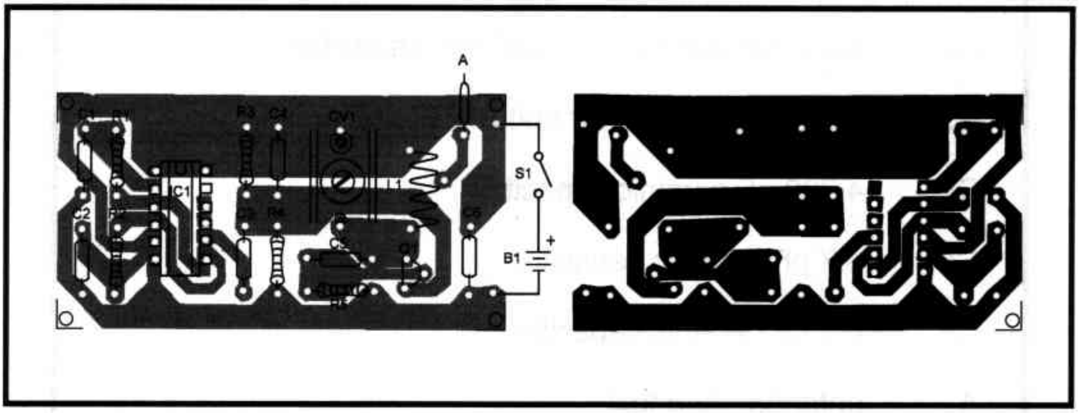

Figure 2

Another way to interface an alarm with the transmitter is to connect relay contacts in series with C3 or between points A and B in the schematic. See more detals in Ideas to Explore.

Parts List — Wireless Beeper

IC1 - 4093 CMOS Integrated Circuit

Q1 - BF494 or equivalent NPN RF silicon transistor

L1 - coil - see text

CV - trimmer capacitor - 20 to 40 pF maximum

R1 - 2,200,000 ohm, 1/4W, 5% resistor

R2 - 39,000 ohm, 1/4W, 5% resistor

R3 - 8,200 ohm, 1/4W, 5% resistor

R4 - 6,800 ohm, 1/4W, 5% resistor

R5 - 47 ohm, 1/4W, 5% resistor

C1 - 0.22 uF ceramic or metal film capacitor

C2 - 0.047 uF ceramic or metal film capacitor

C3 - 0.01 uF ceramic or metal film capacitor

C4 - 4,700 pF ceramic capacitor

C5 - 4.7 pF ceramic capacitor

C6 - 0.1 uF ceramic capacitor

A - antenna - see text

S1 - SPST slide or toggle switch

B1 - 6 to 9V - four AA cells or 9V battery

Ideas to Explore

To get better performance or to learn more about the circuit:

- You can control the beeper oscillator by IC1, pins 1 and 5. Put these pins together. When in a high logic level the circuit produces intervalled beeps. When in a low logic level the circuit stops.
- Replace Q1 by a 2N2218. Power the circuit with a 12V power supply. Signals can now be tuned from distances up to 1,500 feet. A 5- to 6-foot antenna will increase the range of the transmitter.

Science projects or uncommon applications:

- House this circuit in a box and tell someone to hide it. You'll find it by using a small FM receiver.
- The transmitter can be used to track animals by fixing it on a collar.
- You also can use the transmitter to monitor remote sensors in experiments. The circuits can power ON the transmiter or trigger the oscillator as indicated above.

ELECTROSCOPE (E/P)

An electroscope is an instrument that can be used for detecting the presence of an electric static charge. We describe here an electronic version of the traditional electroscope—one that is made with gold foils placed into a glass jar. This circuit can also be used for "sniffing" high voltage without actually making contact with dangerous circuitry.

A metal ring, which constitutes the sensor, is simply poked into the electric field. This operation must be done with caution to avoid contact with the conductors carrying the high voltage.

You can use the device in experiments with high voltage generators (Van de Graaff accelerator) and in many other applications at home or school.

The circuit is portable, powered from four AA cells or a 9V battery.

Presence of a static charge or a strong field created by high-voltage lines is indicated by an LED.

The schematic diagram of the Electroscope is shown in ***Figure 1***. The circuit uses four gates of a 4093 IC that work as buffers and inverters, driving the LED. The high-impedance input of the CMOS integrated circuits results in high sensivity to detect static charges.

All the components are mounted on a homemade printed-circuit board as shown in ***Figure 2***.

Sensor X1 is a small bare wire lop. The completedcircuit can be housed into a small plastic box with the batteries. The LED should be placed in a visible point of the box.

To use:

Bring a charged source close to the sensor. A piece of paper or plastic stroked with a rod of insulating material will do. Adjust P1 to get better sensitivity. The LED will glow according to the electric charge movement.

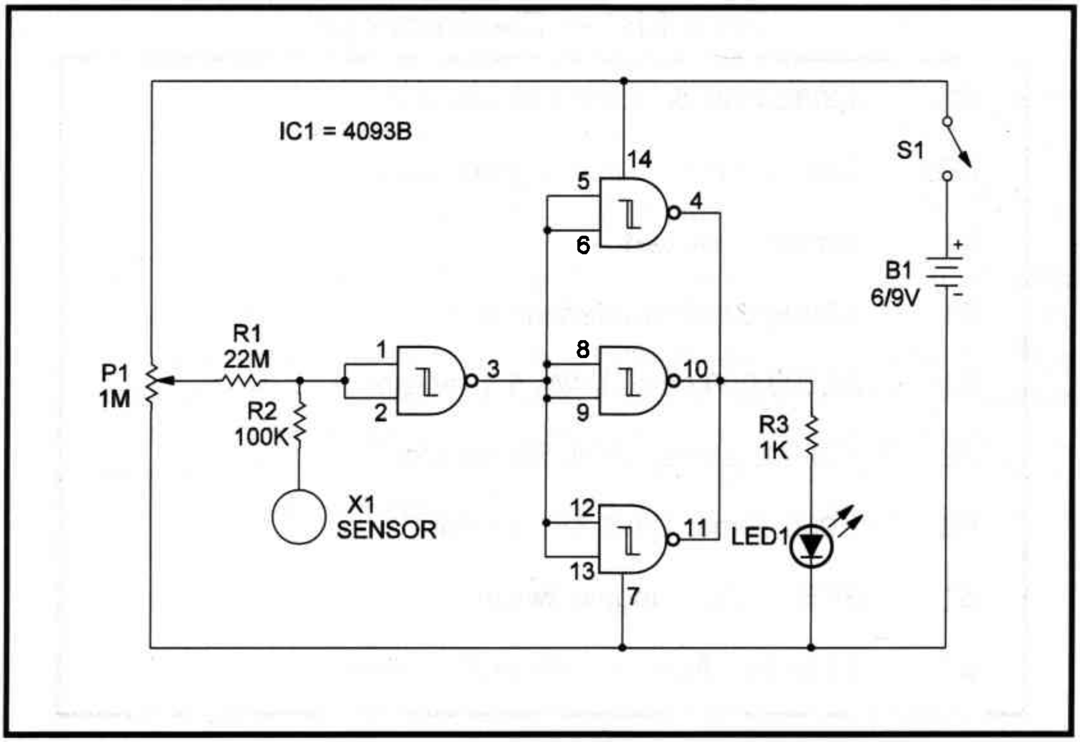
IC1 = 4093B
S1
14
5
4
6
+
B1
6/9V
-
R1
22M
P1
1M
1
3
8
10
R2
100K
2
9
R3
1K
X1
SENSOR
12
11
LED1
13
7

Figure 1

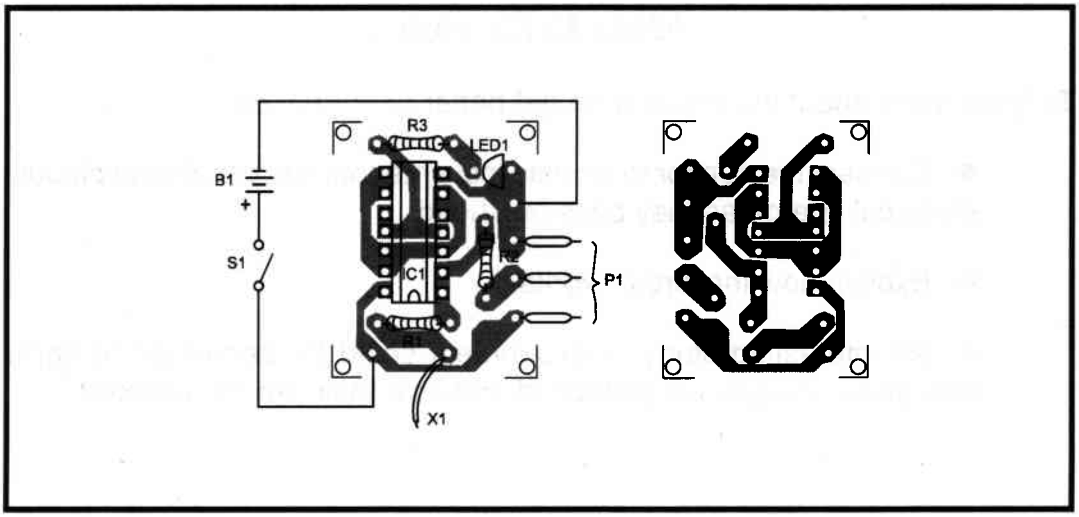
R3
LED1
B1
+
S1
R2
IC1
P1
R1
X1

Figure 2

Parts List — Electroscope

IC1 - 4093B CMOS Integrated Circuit

LED - Common red, yellow or green LED

X1 - Sensor - see text

P1 - 1,000,000 ohm potentiometer

R1 - 22,000,000 ohm, 1/4W, 5% resistor

R2 - 100,000 ohms, 1/4W, 5% resistor

R3 - 1,000 ohms, 1/4W, 5% resistor

S1 - SPST slide or toggle switch

B1 - 6V or 9V - four AA cells or 9V battery

Ideas to Explore

To learn more about the circuit or to get better performance:

- Connect the sensor to an external antenna. You can detect clouds' electrical charge as they pass overhead.
- Explain how the circuit works.
- An interesting study or experiment could be performed to show how static charges are produced and how they can be detected.

Science and uncommon applications:

- In all the experiments that can be conducted with a common electroscope, the electronic version can replace it.

- Determine the polarity of the charge that makes the LED to glow.

- Static charges could have some influence on animals and plants. An experiment to support this affirmation can be conducted. This project, of course, will involve some special equipment such as a high-voltage generator (Van de Graaff or other) and an electroscope. Select seeds from common plants such as tomatoes and beans, and expose several batches of seeds to a static high-voltage field. What are the effects on seed germination? Is plant growth also affected by electric fields?

UNIVERSAL PROBE (P)

This ultra-simple probe can be used to test almost every electronic component. Capacitors, resistors, fuses, switches, coils, transformers, diodes, wires, etc., can be tested with this simple circuit.

Only three components are used and they can easily be housed in a small plastic box, performing as a portable unit.

The circuit consists basically of a continuity tester with a visual indication by an LED. If current can flow between the probes the LED glows; if not, the LED remains off.

Current flowing between probes is very low to avoid overloading the tested components.

A schematic diagram of the Universal Probe is shown in ***Figure 1***. As you can see, only three electronic components are used in this project.

A components layout view is shown in ***Figure 2***. The components can be housed in a small box and no ON-OFF switch is necessary.

When mounting, take care with position of the two polarized components, the LED and power supply.

The power supply consists of two AA cells.

An ON-OFF switch is not necessary, as the circuit is off when the probes are separated.

Current through the component under test is determined by the voltage drop across the LED (about 1.6V for red LEDs and 1.8V for yellow LEDs), and resistance of R1.

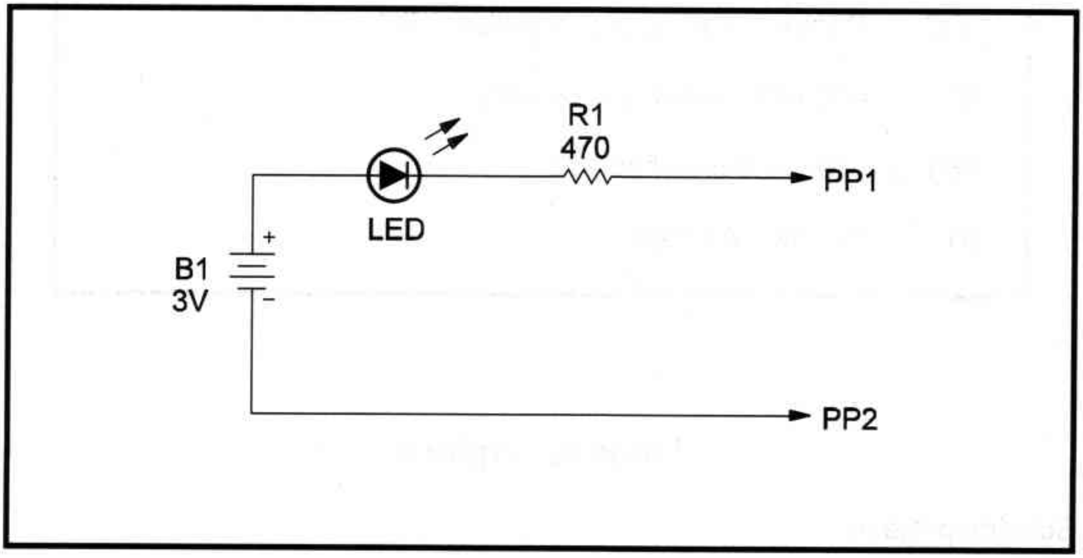
R1
470
LED
B1
3V
+
-
PP1
PP2

Figure 1

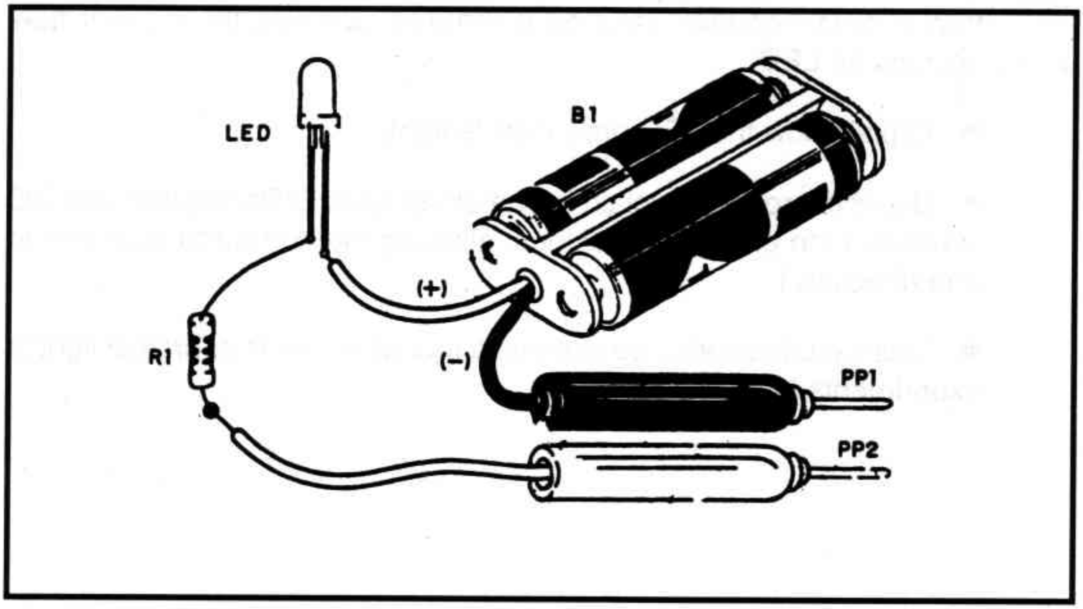
LED
B1
(+)
(−)
R1
PP1
PP2

Figure 2

Parts List — Universal Probe

LED - Common red, green or yellow LED

R1 - 470 ohm, 1/4W, 5% resistor

PP1, 2 - Red and black probes

B1 - 3V - two AA cells

Ideas to Explore

Science projects:

- Use this probe in to find materials that conduct or do not conduct electric current. Several other simple experiments in electricity can be conducted using this probe: for example, show how a switch works, or how a potentiometer used as a rheostat controls the current flow accross an LED.

- Explain how the LED produces its light.

- Use this probe to explain how a diode works. (Remember that DC is used in the probes and diodes, allowing the current to flow only in one direction.)

- This circuit can also be used as a monochromatic source of light in experiments involving optics.

TONE-ACTIVATED RELAY (E)

This circuit can be used to recognize a tone by its frequency. When this tone is present in a circuit the relay is energized, which could supply power to a selected appliance.

Wired to the output of a receiver, the circuit can be used as part of a tone-modulated wireless remote control. The receiver can be a common FM radio or an infrared receiver, depending on the intended application.

The circuit can be used with tones ranging up to 100,000 Hz. The circuit is tuned by P1.

Inputs of other units can be wired in parallel to get a multi-channel remote control system. Take care to avoid using harmonic frequencies.

The schematic diagram of the Tone Activated Relay is shown in ***Figure 1***. The heart of the circuit is the LM567 IC, a National Semiconductor PLL (Phase-Locked Loop) that drives the relay through transistor Q1.

Components placement on a homemade printed-circuit board is shown in ***Figure 2***. Placement is not critical, but when operating in frequencies above 50,000 Hz it is important to avoid long connections.

The relay is any type with a 6V coil and current rated for values between 10 and 100 mA.

Position of the polarized components, such as the diode and electrolytic capacitor, should be observed.

To use:

Wire the output of an audio generator or audio oscillator to points X and Y. The signal should have amplitude in a range between 100 mV and 1 V peak-to-peak for best results. Adjust P1 to trigger the relay. Reduce the signal amplitude and trim P1 for the best performance.

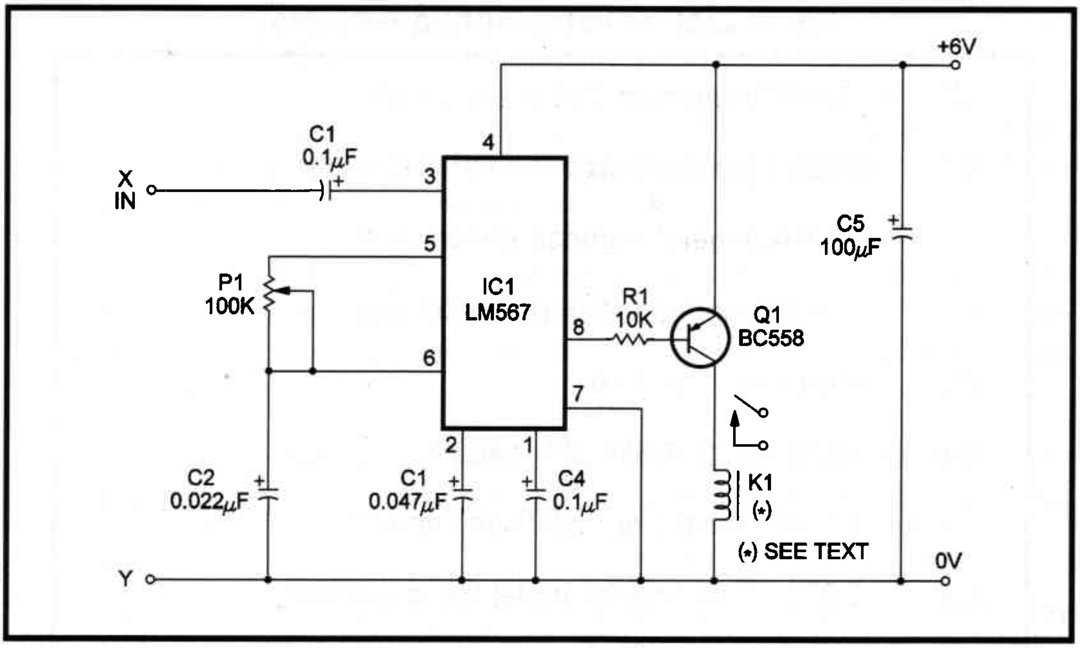
+6V
C1
0.1μF
X
IN
3
4
5
P1
100K
IC1
LM567
R1
10K
8
Q1
BC558
6
7
C5
100μF
2
1
C2
0.022μF
C1
0.047μF
C4
0.1μF
K1
(*)
(*) SEE TEXT
Y
0V

Figure 1

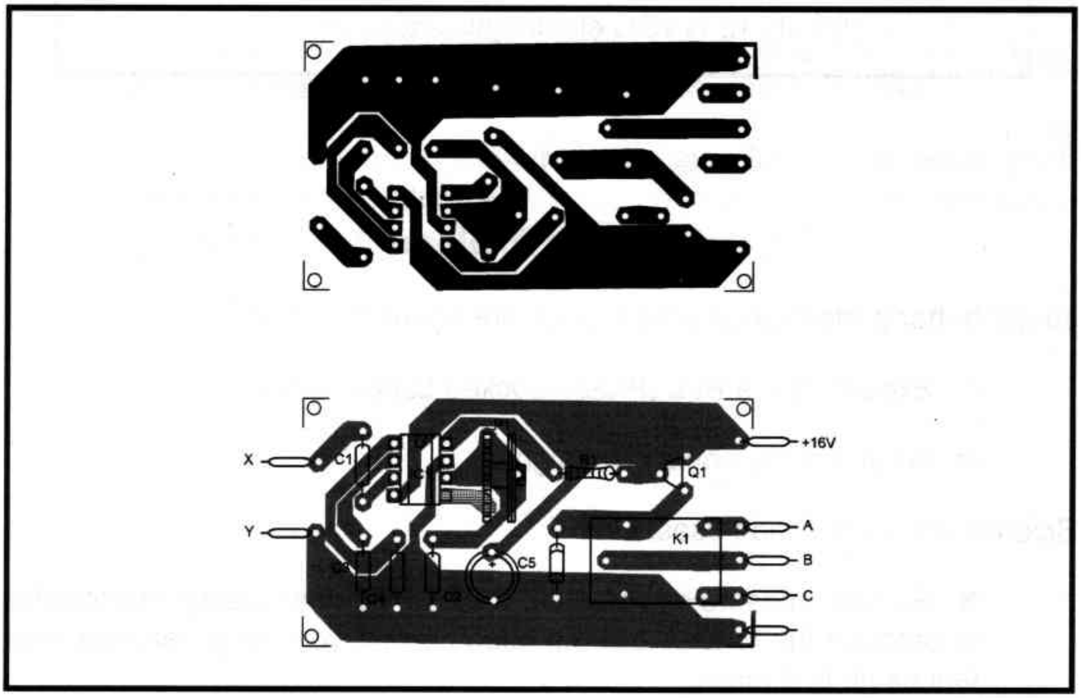
X
C1
R1
Q1
+16V
Y
K1
A
B
C
C5
C4
C2

Figure 2

Parts List — Tone-Activated Relay

IC1 - LM567 Integrated Circuit (National)
Q1 - BC558 general-purpose PNP silicon transistor
D1 - 1N914 general-purpose silicon diode
P1 - 100,000 ohm trimmer potentiometer
K1 - relay - 6V - see text
R1 - 10,000 ohm, 1/4W, 5% resistor
C1, 4 - 0.1 uF ceramic or metal film capacitor
C2 - 0.022 uF ceramic or metal film capacitor
C3 - 0.047 ceramic or metal film capacitor
C5 - 100 uF, 12 WVDC electrolytic capacitor

Ideas to Explore

To get better performance or to learn more about the circuit:

- Explain how a PLL (Phase-Locked Loop) works.
- What are harmonics of a signal?

Science and uncommon uses:

- Remote sensors can be used to close the relay. Using an oscillator to produce the tone, the wires used can be very long, reaching distances up to 5 miles.

- You can use a small oscillator as transmitter and this circuit as part of a receiver to activate devices by telephone line.

- Connect the input of this circuit to the output of an audio amplifier. Plug a microphone to the input of the amplifier. With some practice you can control the relay by whistling.

- You can also use this tone-activated relay to control appliances from your computer, using the multimedia audio output. Write a program to produce a tone when an icon on your Windows screen is clicked with the mouse.

- The tone can be used to turn on the relay if the circuit is plugged to the audio output of your computer. Three modes of operation are possible:

a) By adjusting the length of the tone pulse you can turn on any appliance connected to the relay during this time interval. The appliance will turn off after the pulse ends.

b) Using a monostable with a 555 timer, for instance, a short pulse can turn on the appliance after a preadjusted time interval.

c) Using the turn-on-and-off circuit, described in this book, you'll have a bistable operation: a click on the icon turns on the appliance controlled by the relay, and the next click turns it off.

LIGHT COMPARATOR (E)

This device can be used to compare two light sources by their intensities. Several optical experiments can be conducted using this. You can compare colors of surfaces by reflected light or compare colors using special filters.

The sensors used are CdS photoresistors or LDRs (Light Dependent Resistors) that have peak sensitivity in the red part of the spectrum (7350 angstroms or 735 nm).

As the name suggests, the LDR has a resistance which varies as light falls on its sensitive surface. As the light intensity increases, resistance falls. In total darkness resistance can be high as 1,000,000 ohms or more. Illuminated by direct solar light, the resistance falls to 100 ohms or less.

Our circuit is a simple Wheatstone Bridge equilibrated by two LDRs and an adjustable potentiometer (P1).

When the two LDRs are receiving the same amount of light, P1 can be adjusted to get a null indication on the ammeter (M1).

If we use one LDR as reference, we can compare another source of light with this by using the other LDR as sensor. A new adjustment of P1 can reveal the difference between the two light sources.

The schematic diagram of the light comparator is shown in ***Figure 1***. No transistors are used in this project—only resistors.

Components layout, using a terminal strip as chassis, is shown in ***Figure 2***. The terminal strip can be fixed in the plastic box using screws. Wires to the sensor can be as long as 6 feet without instability problems.

The LDRs should be mounted into two small opaque tubes to receive light from different sources without outside interference. M1 should be a zero-center microammeter. A 50-0-50 uA microammeter would be suitable.

The circuit can be powered from two or four AA cells, and current drain is very low, extending battery life to many months.

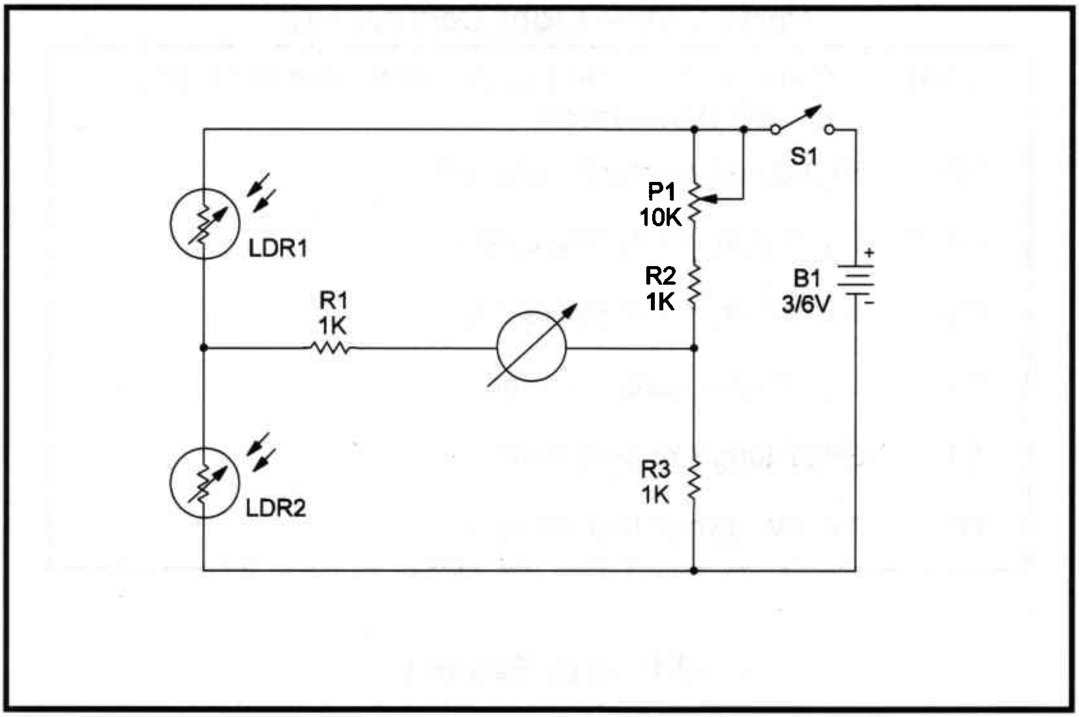
S1
P1
10K
LDR1
R2
1K
+
B1
3/6V
-
R1
1K
R3
1K
LDR2

Figure 1

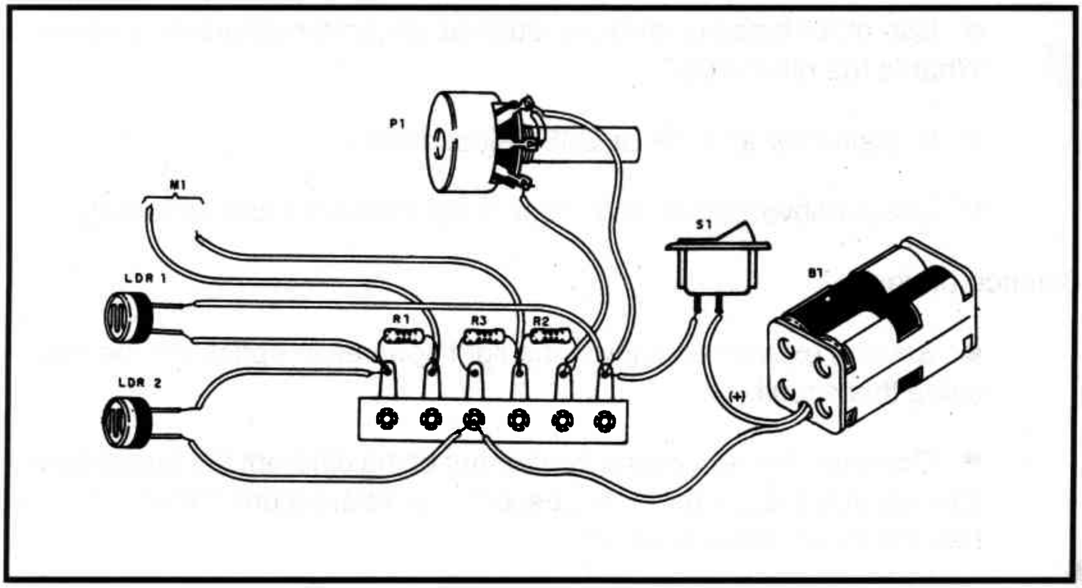
P1
M1
S1
B1
LDR 1
R1
R3
R2
LDR 2
(+)

Figure 2

Parts List — Light Comparator

LDR1, 2	- Common 1cm LDRs (Light Dependent Resistors) or CdS photosensors
M1	- 50-0-50 uA ammeter - see text
R1, 2	- 1,000 ohm, 1/4W, 5% resistor
R3	- 4,700 ohm, 1/4W, 5% resistor
P1	- 10,000 ohm potentiometer
S1	- SPST toggle or slide switch
B1	- 3 to 6V - two or four AA cells

Ideas to Explore

To learn more about the circuit or to get better performance:

- Use other types of sensors, such as phototransistors in the circuit. What is the difference?
- Explain how an LDR or CdS cell operates.
- Use a convergent lens in front of the cells for more sensitivity.

Science projects:

- Several experiments involving light sources in optics can be made using this circuit.
- Compare the reflections from paper using different but equal-power light sources, such as incandescent and fluorescent. Which of them has the better performance?
- It is possible to change this circuit into a temperature comparator. Just replace the LDRs with NTCs.

ELECTRONIC ORGAN (E/P)

This is an interesting experiment in audio which would make a nice toy for children.

This circuit produces a musical note by pressing a key on a keyboard. The sound is produced by a small speaker.

The electronic organ will play only one note at a time, but the number of notes is unlimited.

Different tones can be selected by pressing different keys, as in a common electronic organ. Tonal range is determined by C1, which can be altered within a large range of values. Values between 0.022 uF (for higher notes) and 1 uF (for lower notes) can be experimented.

You can also use this circuit as a multi-tone bell in your home or to monitor several places at the same time. Replacing the keys with different sensors (reed switches, for instance) will produce different tones when closed.

A schematic diagram of the simple Electronic Organ is given in ***Figure 1***. The circuit consistis of a simple two-transistor oscillator. The transistors are direct- coupled and drive a small loudspeaker.

Components layout, using a terminal strip as chassis is, shown in ***Figure 2***. The circuit can be housed in the same box as the keyboard.

In ***Figure 3*** we show the layout of the keyboard. Each trimmer potentiometer is used to adjust individual notes.

The circuit can be powered from AA cells or a power supply ranging from 3 to 6 volts.

The keyboard can also be made from a printed-circuit board or with small metal plates. When each metal plate is touched by a probe the corresponding circuit closes and a musical tone is produced.

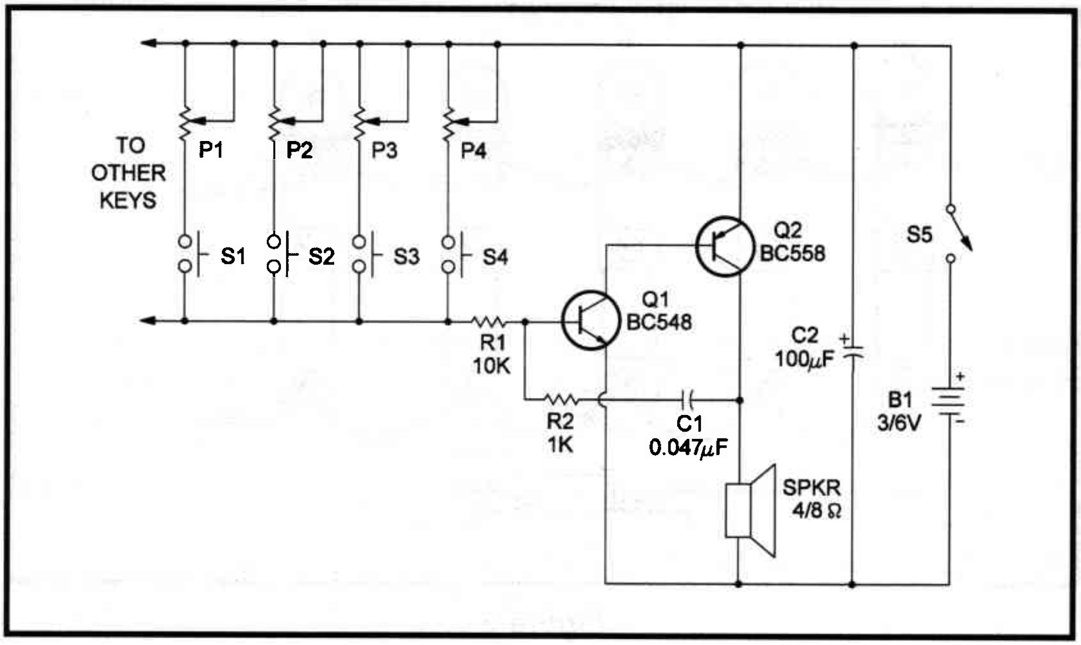
TO
OTHER
KEYS
P1
P2
P3
P4
S1
S2
S3
S4
Q2
BC558
S5
Q1
BC548
R1
10K
C2
100μF
+
B1
3/6V
+
-
R2
1K
C1
0.047μF
SPKR
4/8 Ω

Figure 1

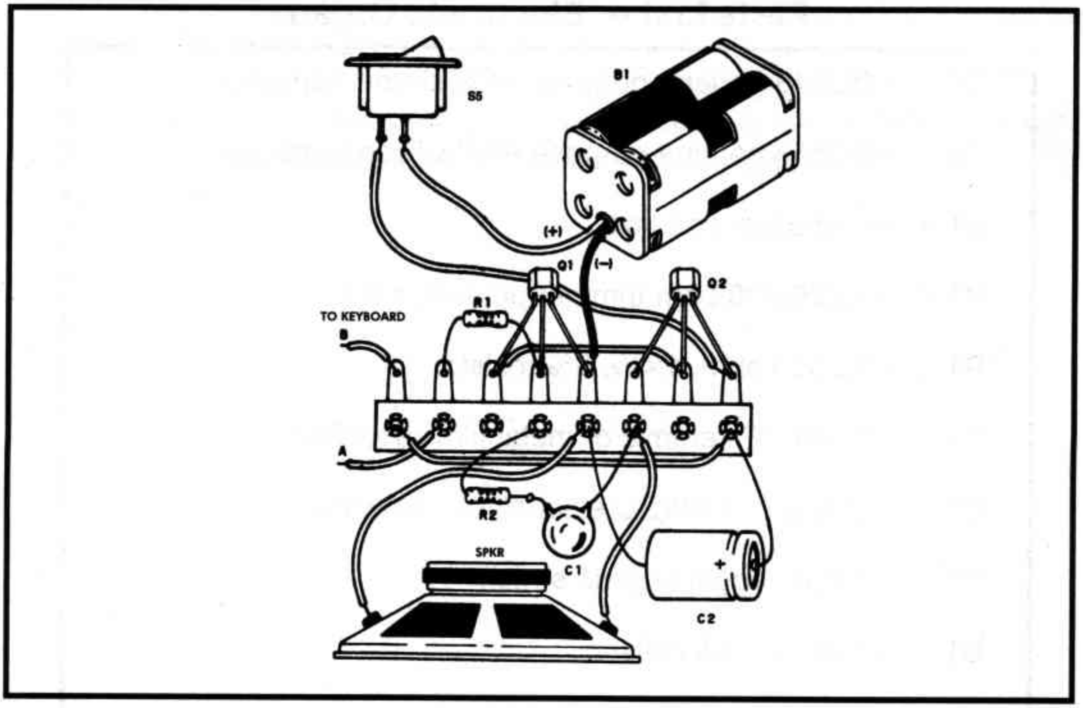
S5
B1
(+)
(−)
Q1
Q2
R1
TO KEYBOARD
B
A
R2
SPKR
C1
+
C2

Figure 2

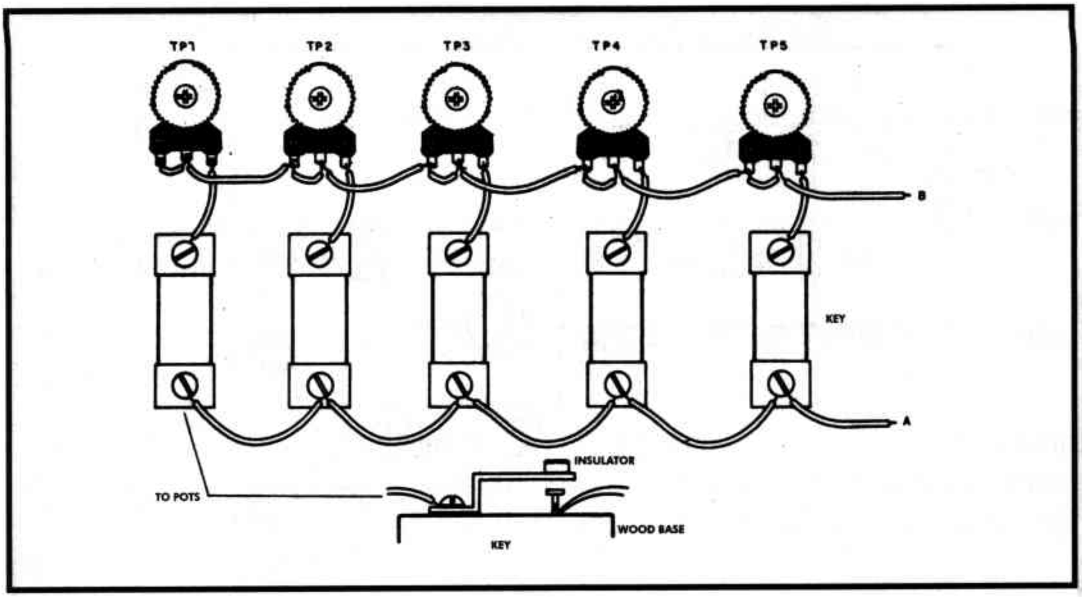
TP1
TP2
TP3
TP4
TP5
B
KEY
A
INSULATOR
TO POTS
WOOD BASE
KEY

Figure 3

Parts List — Electronic Organ

Q1 - BC548 general-purpose NPN silicon transistor

Q2 - BC558 general-purpose PNP silicon transistor

S1-4 - keyboard- see text

P1-4 - 1,000,000 ohm trimmer potentiometer

R1 - 10,000 ohm, 1/4W, 5% resistor

C1 - 0.047 uF ceramic or metal film capacitor

C2 - 100 uF, 12 WVDC electrolytic capacitor

S5 - SPST toggle or slide switch

B1 - 3 or 6V - AA cells

SPKR - 4/8 ohm, 2 to 4 in. - small loudspeaker

Ideas to Explore

To learn more about the circuit or to get better performance:

- Alter C1 in the range given in the text to produce a different range of sounds.
- Explain how the circuit works.
- Try to replace R1 with a network formed by small transformers, diodes, resistors, capacitors and other components. Modifications in the tone pitch could be found this way.

Science and uncommon applications:

- Replace S1, S2, etc., with different kinds of switches. You can identify the closed switch by the tone.
- Use only one potentiometer coupled to a cursor. You can produce music by rapidly changing cursor position and pressing the series switch.
- The same keyboard can be used in other audio oscillators, resulting in different configurations of electronic organ.

EXPERIMENTAL LIGHT-ALARM (E)

This circuit consists of a simple but effective light-activated alarm which uses only a half dozen components. A light falling on the LDR turns on the SCR. The lamp will remain on until the power is off or S1 is used to reset the circuit.

P1 adjusts sensitivity according to the application. To get better performance the LDR can be housed in an opaque tube with a convergent lens.

A 6-volt lamp can be used without problems since there is a voltage drop of about 2 volts when in the ON state.

The schematic diagram is shown in ***Figure 1***. S1 is optional, as you can reset the circuit by turning off the power supply.

Components layout, using a terminal strip, is shown in ***Figure 2***. The lamp can be replaced by an LED in series with a 470 ohm resistor or a buzzer. You also can control an audio oscillator, replacing the lamp with it.

The lamp isn't a critical component. Any 6V lamp ranging from 50 to 250 mA can be used.

The LDR is a common CdS photoresistor. The SCR needn't be mounted on a heatsink as the current is very low.

Parts List — Experimental Light-Alarm

SCR	- TIC106 Silicon Controlled Rectifier
LDR	- Common LDR - see text
P1	- 1,000,000 ohm potentiometer
X1	- 6V incandescent lamp - see text
B1	- 6-9V - four AA cells, D cells, battery or power supply
S1	- SPST momentary switch

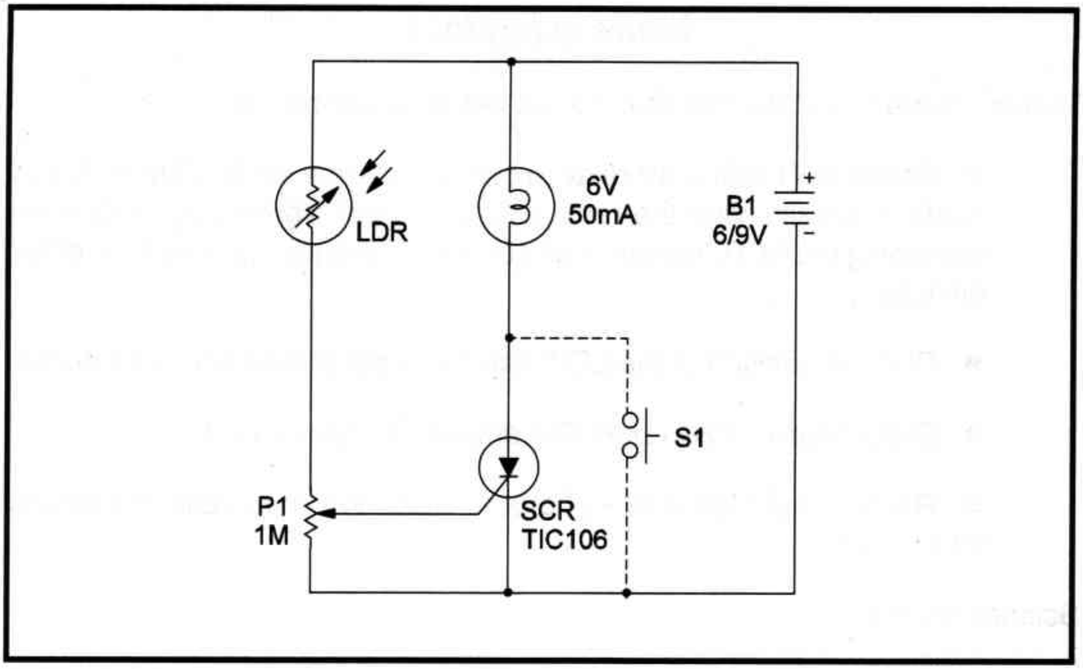
LDR
6V
50mA
B1
6/9V
+
-
S1
P1
1M
SCR
TIC106

Figure 1

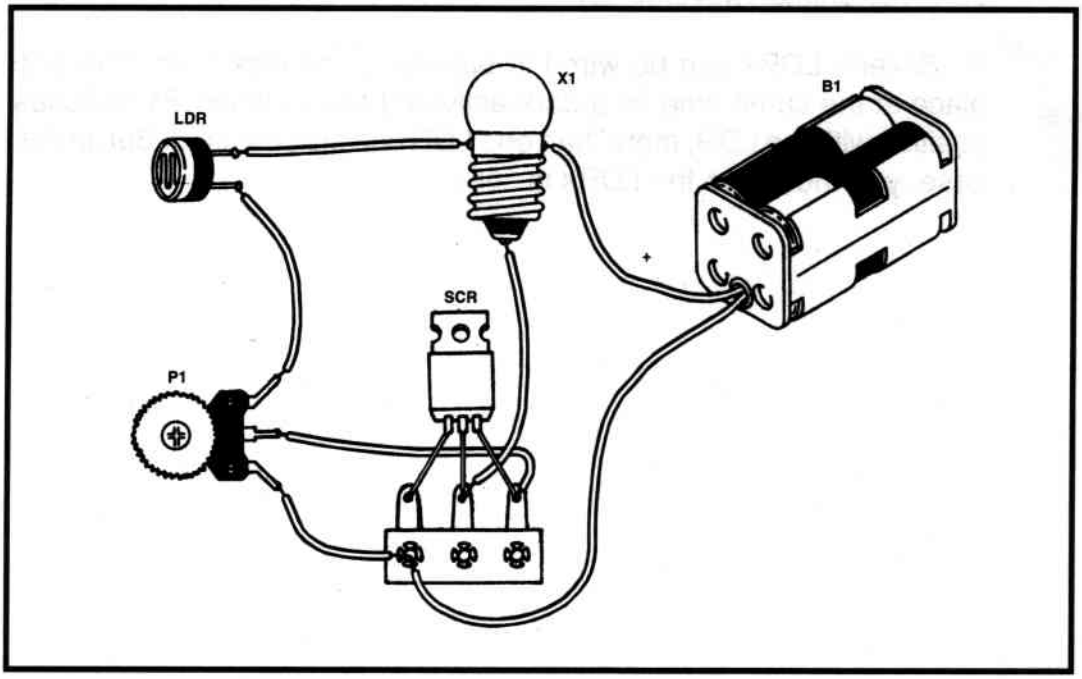
X1
B1
LDR
+
SCR
P1

Figure 2

Ideas to Explore

To learn more about the circuit or to get better performance:

- Replace X1 with a 6V relay to control external loads. Wire a 1N914 diode in parallel with the relay coil to prevent high-voltage kick from damaging the SCR. You can also control small audio oscillators or other devices.
- Change position of the LDR with P1 to get a dark-activated alarm.
- Explain how LDRs (Light Dependent Resistors) work.
- Replace the LDR with a phototransistor. Will the circuit operate as you want?

Science projects:

- This circuit can be used to monitor traps or events that involve object or animal movements.
- Several LDRs can be wired in parallel to monitor more than one place at the same time. In a dark-activated alarm, when P1 switches position with the LDR, more than one LDR can also be used. But, in this case, you must wire the LDRs in series.

PHOTOMETER (E/P)

Some of the most common applications for LDRs are as sensing elements in alarms, photo-relays, counters, light meters for photography and also experiments in optics.

The project described here is a simple light meter made from an LDR (Light Dependent Resistor or Photoresistor), a 0-200 uA microammeter, and a 3V battery formed by two AA cells.

Use any general-purpose LDR with a high dark-to-light resistance ratio.

Since the circuit is operating for short durations you don't need an ON-OFF switch. Just put the cells in the battery holder when using the unit.

P1 adjusts sensitivity to the range of your working light intensity.

Figure 1 shows the schematic of the Photometer. The ammeter is not critical. Any type with full scale ranging from 100 uA to 1 mA can be used.

Figure 2 shows a mounting view, using a terminal strip as chassis. The device can be housed in a small plastic box. The meter and LDR are placed on the front panel.

If the ammeter used is out of specification, you need to adjust P1. Try values between 10,000 and 100,000 ohms. The LDR can be housed in a small opaque tube, depending on the application.

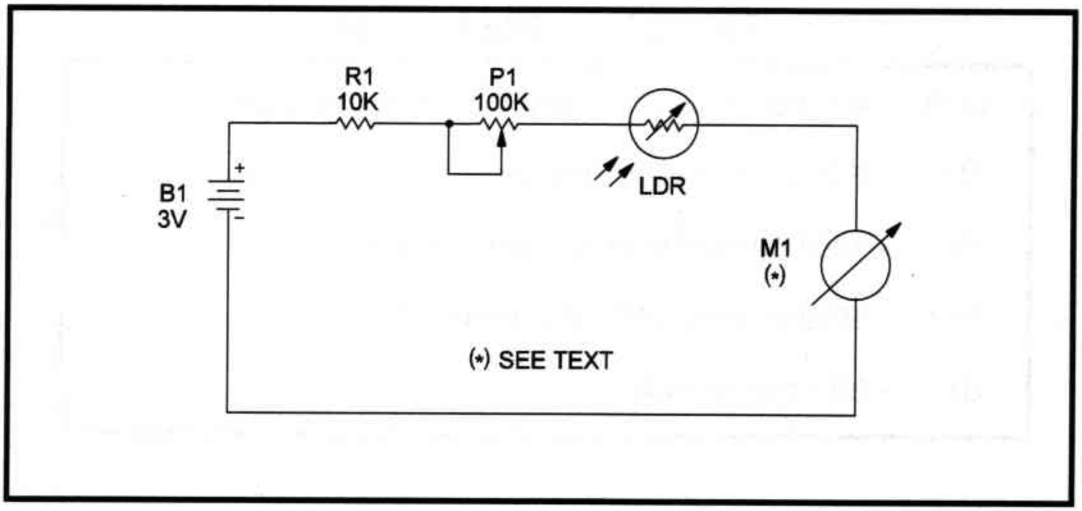
R1
10K
P1
100K
B1
3V
+
-
LDR
M1
(*)
(*) SEE TEXT

Figure 1

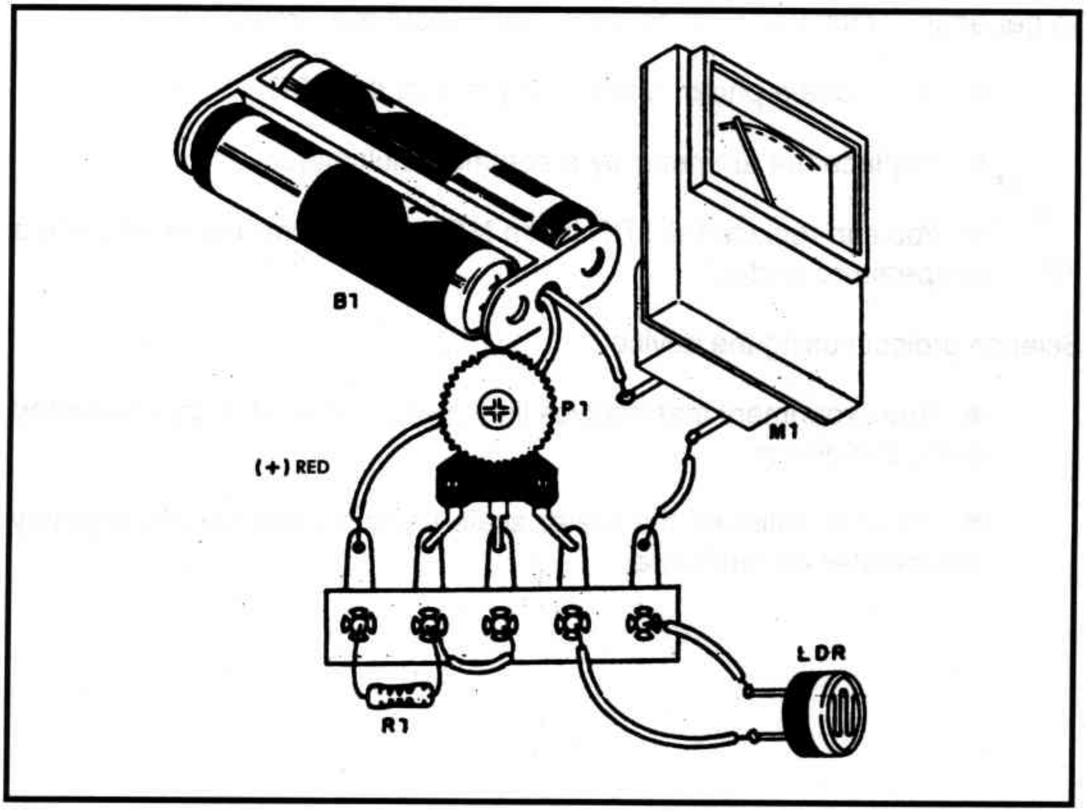
B1
P1
M1
(+) RED
LDR
R1

Figure 2

Parts List — Photometer

LDR - Any common LDR (Light Dependent Resistor)

M1 - 0-200 uA - meter - see text

P1 - 100,000 ohm trimmer potentiometer

R1 - 10,000 ohm, 1/4W, 5% resistor

B1 - 3V - two AA cells

Ideas to Explore

To get better performance or to learn more about the circuit:

- Try to use a phototransistor in place of the LDR.
- Replace the ammeter by a common multimeter.
- You can replace the LDR by an NTC and convert the device into a temperature meter.

Science projects using the device:

- Any experiment that involves light measurement can be conducted using this device.
- You can calibrate the meter scale using a common photography photometer as reference.

SEQUENCE GENERATOR (E/P)

An astable multivibrator, put to work by using it as a source of clock pulses for a 4017 counter, is the basis for this circuit.

We can use it to generate random numbers in a range of one-of-ten or as a 0-to-9 counter in science projects or in other applications.

You can also adjust the clock so that it supplies one pulse each second for using the circuit as a timer in scientific experiments or at home. If you alter the clock pulse frequency, time range can be changed so that it supplies one pulse each minute or ten minutes. In this case, the time range will rise to 9 minutes or 90 minutes (in 1- or 10-minute steps)!

As a sequence generator in logic circuits demonstrations, you can change the astable operation of the 555 to monostable, connecting a momentary switch or any sensor that can be used to trigger it.

The circuit can be powered from 6 to 12V supplies, and R3 depends upon the voltage of the power supply. Use a 470 ohm resistor if the supply voltage is 6V. Use a 1,000 ohm resistor if the power supply used is 12V.

Capacitor C1 determines the pulse rate range and can be altered. Values between 1 and 1,000 uF can be used. Remember that higher values mean lower pulse rate.

P1 adjusts the pulse rate within 100:1, depending on the capacitor used.

Figure 1 shows the schematic diagram of the Sequence Generator. The circuit uses a 555 IC timer to generate the clock pulses and a 4017 CMOS IC, a counter and 1-of-10 decoder to drive ten LEDs as indicators.

Components placement on a homemade printed-circuit board is shown in ***Figure 2***. You can also use a solderless board or breadboard to mount it.

Positions of the polarized pieces, such as the LEDs and electrolytic capacitor, should be observed.

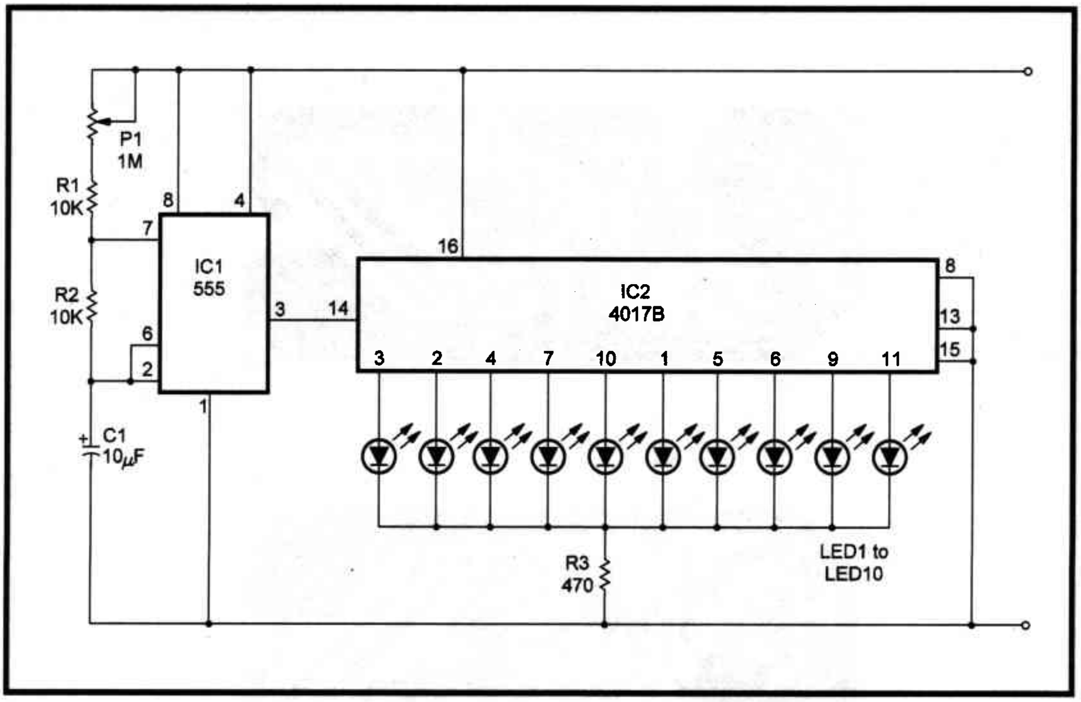
P1
1M
R1
10K
8
4
7
IC1
555
R2
10K
6
2
3
14
16
IC2
4017B
8
13
15
3
2
4
7
10
1
5
6
9
11
+
C1
10μF
1
R3
470
LED1 to
LED10

Figure 1

Parts List — Sequence Generator

IC1 - 555 - Integrated circuit, timer

IC2 - 4017 - CMOS Integrated circuit, counter

LEDs - Common red, yellow or green

P1 - 1,000,000 ohm potentiometer

R1, 2 - 10,000 ohm, 1/4W, 5% resistors

R3 - 470 ohm, 1/4W, 5% resistor

C1 - 10 uF/16 WVDC electrolytic capacitor

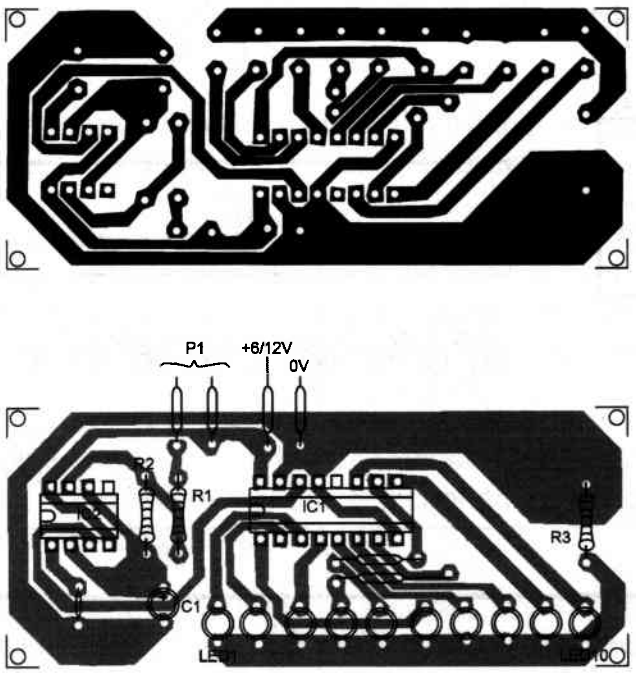

Figure 2

Ideas to Explore

To learn more about the circuit or to get better performance:

- Wire a momentary switch between IC1's pin 3 and IC2's pin 14 to control the counter. You can stop the counter any time with this switch.
- Explain how counters such as the 4017 work.
- Could you alter the circuit so that it counts to any value between 2 and 9?

Science and uncommon uses:

- Wire the 555 as a monostable multivibrator and use an LDR as sensor to trigger it. The circuit will operate as a pulse-of-light counter.
- Use this circuit in a project about computers. You can produce pulses to be counted using the experimental flip-flop described in this book.
- Experiments on ESP (Extra-Sensory Perception) can be conducted using this circuit. Random numbers in a scale of 1-to-10 can be generated using this circuit.

SIGNAL INJECTOR (E/P)

If you're interested in doing some servicing on your audio equipment (multimedia, for instance) you will undoubtedly find this square wave generator to be a handy tool. You can also use this circuit in RF stages in AM/FM receivers, since the oscillator harmonics are strong in frequencies as high as 100 MHz.

In this arrangment, frequency is determined by capacitor C2, resistors R1 and R2, and D1. The frequency can be altered by the values of resistors and/or capacitors selected. In our circuit the oscillator runs at about 1,000 Hz.

The output signal waveform is square and the output swings the full power supply voltage, which can be anything between 3 and 12 volts (we recommend a 3-volt power supply if you want a portable unit). Supply current is typicaly 10 uA, extending battery life.

Figure 1 shows the complete schematic diagram of the Signal Injector. IC1 is a 555 timer, wired as an audio oscillator. Frequency is determined by C2, R2 and R1. You can change C2 within a large range of values.

Figure 2 shows the homemade PC board where the components are mounted. A small 3V battery can also be used if you want a very compact mounting.

A signal injector is used from the "back" to the "front" of an audio (or RF) circuit. For instance, to use the signal injector with an AM receiver, apply the signal from the probe at the base of the output transistor. If that stage and everything after it operates correctly the signal will be heard in the speaker. If the output stage proves to be OK, move back to the base of the driver transistor. The output signal will be higher if everything is working. Then, apply the signal progressively towards the front of the circuit by injecting it at the volume control, detector stage, IF stages and the mixer.

The circuit can be housed in a small plastic box and connected to the external circuit through a probe and an alligator clip. ***Figure 3*** represents a possible housing arrangement.

To test the signal injector, apply the signal to the input of any audio amplifier.

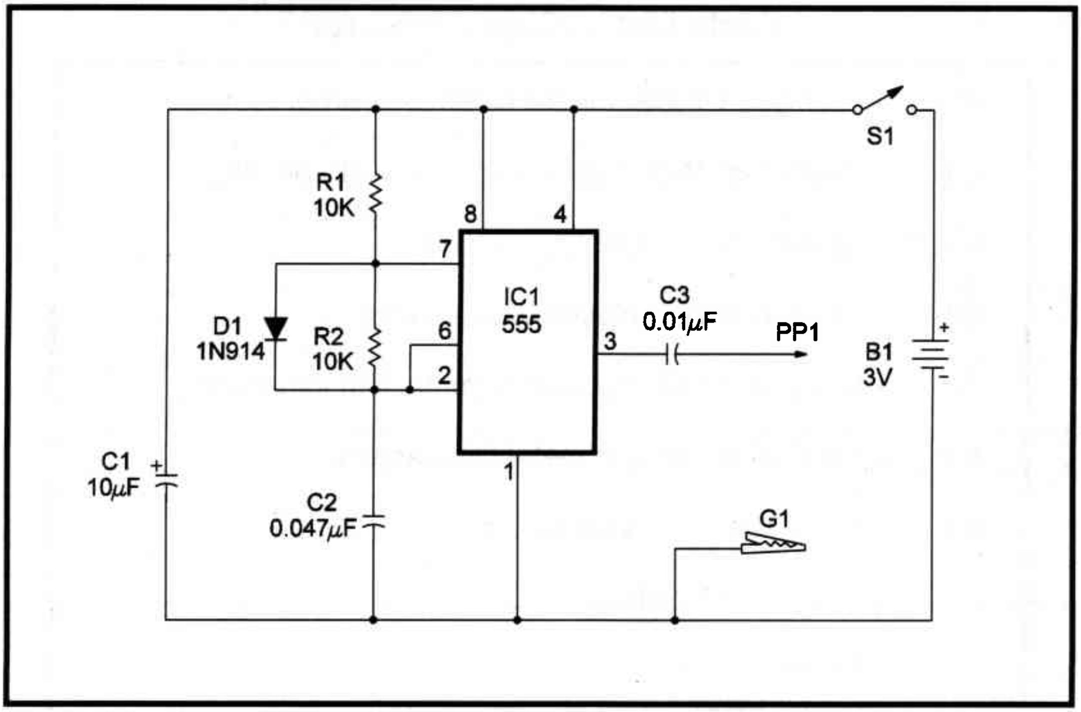
S1
R1
10K
8
4
7
D1
1N914
R2
10K
6
2
IC1
555
3
C3
0.01µF
PP1
+
B1
3V
-
C1
10µF
+
1
C2
0.047µF
G1

Figure 1

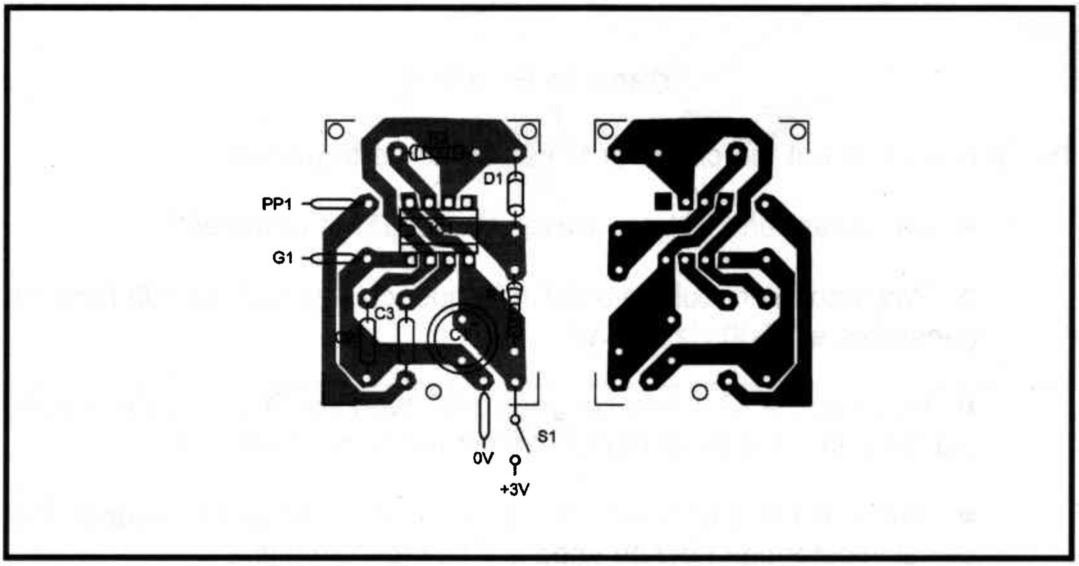
PP1
G1
D1
C3
C1
0V
S1
+3V

Figure 2

Parts List — Signal Injector

IC1	- TLC7555 CMOS integrated circuit - timer
D1	- 1N4148 or 1N914 general-purpose silicon diode
R1, 2	- 10,000 ohm, 1/4W, 5% resistor
C1	- 10 uF, 6 WVDC electrolytic capacitor
C2	- 0.047 to 0.1 uF ceramic or metal film capacitor
C3	- 0.01 uF ceramic or metal film capacitor
S1	- SPST toggle or slide switch
B1	- 3V - two AA cells
PP1	- Probe
G1	- Alligator clip

Ideas to Explore

To learn more about the circuit or to get better performance:

- What happens with the wave shape if D1 is removed?
- Why can the circuit be used in frequencies as high as 100 MHz if it generates a 1,000 Hz signal?
- Replace the IC with a common 555 (bipolar). The circuit will work, but the current drain is higher than in the original version.
- Remove C3 and power the circuit from a 5V power supply. The signal injector can now be used in TTL logic circuits.

- R1 and R2 are shown as fixed resistors in this project. You can change these resistors to a potentiometer (1,000,000 ohms, for instance) in series with a 1,000 ohm resistor to generate signals within a large range of frequencies and duty-cycles.

Science projects using the signal injector:

- The circuit can be used as a simple bio-stimulator in biological experiments. Stimulation levels as high as the power supply voltage can be achieved. These levels can be used in direct nerve stimulation.
- Wire a piezoelectric transducer or crystal earphone to the output of this signal injector. A tone of about 1,000 Hz will be produced and can be used in experiments.

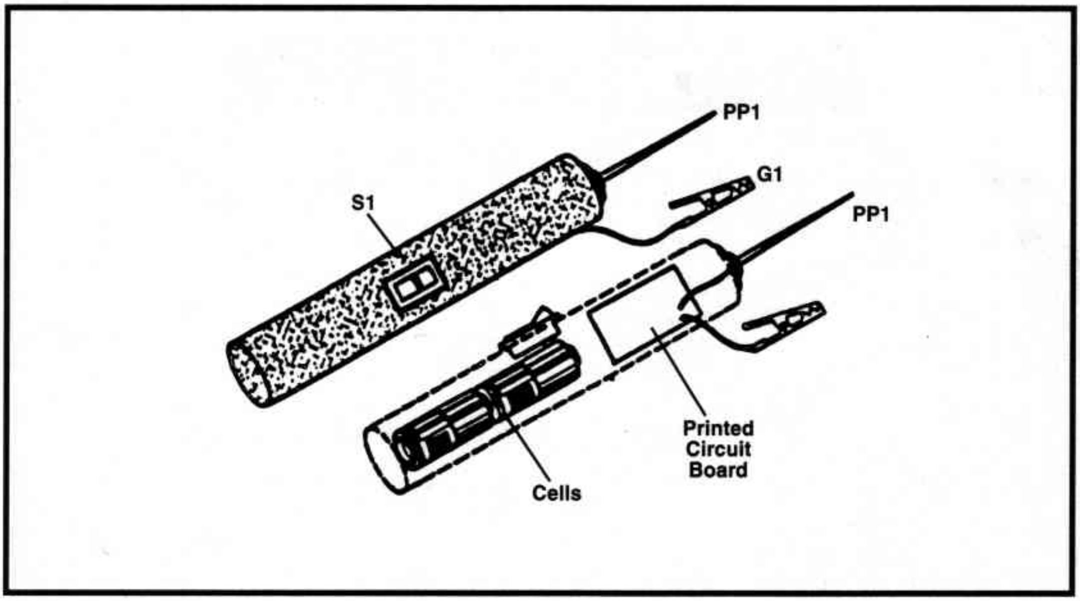

Figure 3

METAL DETECTOR (E)

You can use this simple metal detector to locate plumbing pipes, electrical wiring inside walls, find a lost watch or coin at the beach or even a treasure with gold coins buried years ago by pirates.

You can also use the circuit in experiments where you can show the magnetic properties of materials. The differences between paramagnetic, diamagnetic, and nonmagnetic materials can be shown. Interesting experiments can be conducted using the metal detector described here.

This circuit will detect the presence of metallic objects through any nonconductive material such as dirt, wood, stone, plaster or plastic.

With a little practice, you will be able to detect metallic objects to depths of over several inches and will give you a fairly good idea of their size.

Our circuit uses the beat-frequency method to detect an object.

The frequency of the search coil oscillator changes when the conductance in the field of the search coil changes. Therefore, when metals come within a short distance of the coil, the frequency is changed, which in turn changes the pitch of the audible sound.

The sound is produced by the speaker of a common AM receiver that tunes the coil oscillator.

A schematic diagram of the Metal Detector is shown in ***Figure 1***. Observe that the circuit is a very simple one-transistor oscillator. You can also use a general-purpose PNP silicon transistor, inverting the power supply polarity.

The physical arrangment of the electronic components, using a terminal strip as chassis, is shown in ***Figure 2***. The critical part of this project is the search coil.

The search coil is wound in a large 6-inch plastic Frisbee. Begin by securing one end of the enameled magnet wire form (18 to 22 AWG) with masking tape. On the tenth wrap (halfway) form a loop in the magnet wire for the center tap, then continue until you complete the 20 turns of wire.

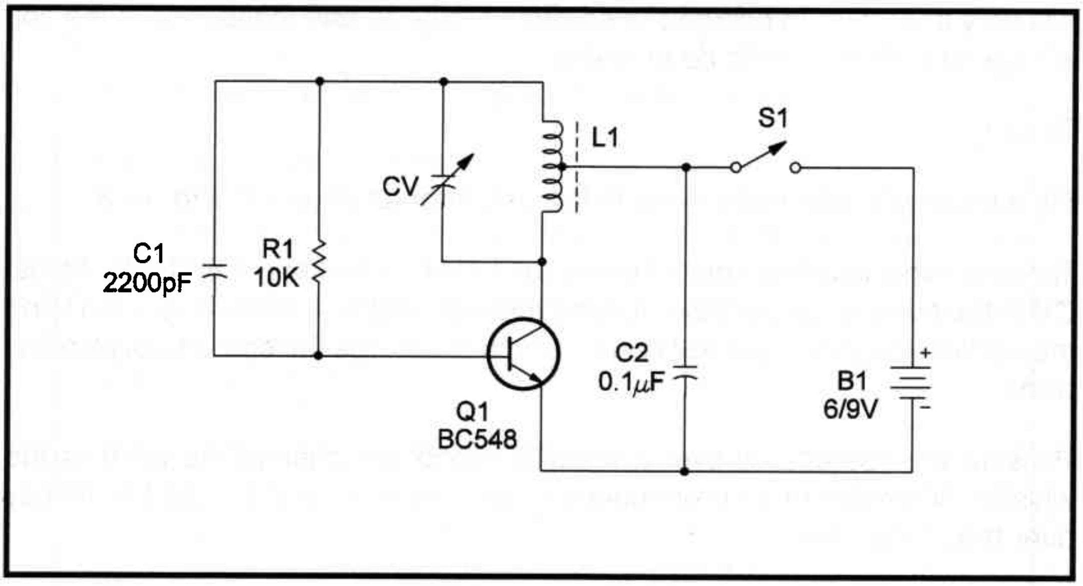

C1
2200pF
R1
10K
CV
L1
S1
Q1
BC548
C2
0.1µF
B1
6/9V
+
-

Figure 1

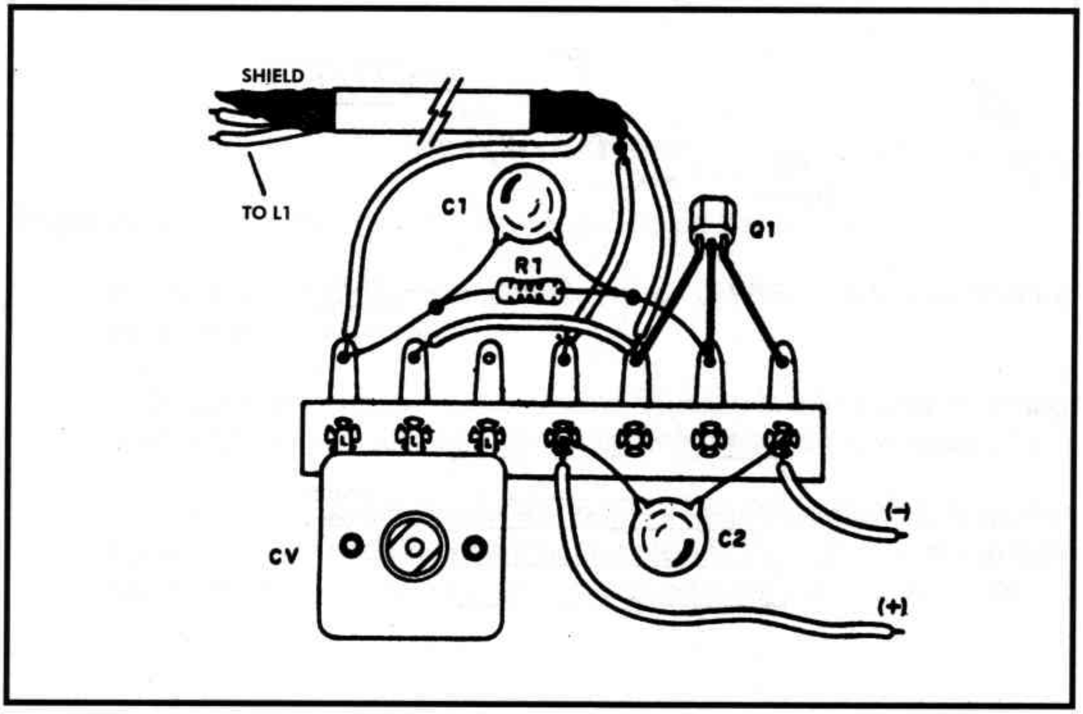

SHIELD
TO L1
C1
R1
Q1
CV
C2
(−)
(+)

Figure 2

It is very important to secure the coil wire firmly, as one loose loop in the coil will cause the detector to be unstable.

To use:

Fix a small portable radio close to the oscillator as shown in ***Figure 3***.

Tune the radio to a free point of the range between 550 and 1,600 kHz. Adjust CV in the search coil oscillator to tune this signal as a whistle. If you can't find the oscillator signal in the AM band, reduce or enlarge the search coil by a few turns.

Passing the search coil over a metallic object will change the pitch of the whistle. When the most pronounced change occurs, you should be directly over the metal object.

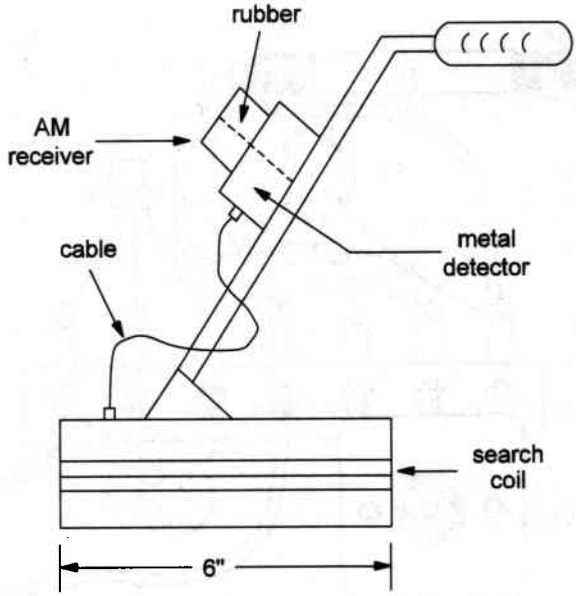

Figure 3

Parts List — Metal Detector

Q1 - BC548 general-purpose NPN silicon transistor

L1 - Search coil - see text

CV - 365 pF - any variable capacitor (from nonworking AM radio)

C1 - 2,200 pF ceramic capacitor

C2 - 0.1 uF ceramic capacitor

R1 - 10,000 ohm, 1/4W, 5% resistor

S1 - SPST slide or toggle switch

B1 - 6 or 9V - four AA cells or 9V battery

Ideas to Explore

To get better performance or to learn more about the circuit:

- Make experiments with different kinds of materials to see how the detector works with each one.
- Explain what happens when a metal object is introduced in a magnetic field. Is there any way to identify the metal being detected?
- The circuit can also be used as an RF signal generator. To get the signal from this oscillator, connect a probe to the collector of the transistor with a 1,000 pF capacitor and an alligator clip to the emitter.

Science projects and uncommon applications:

- Several experiments involving metal detection can be conducted using this simple metal detector. An example would be a work to show the differences between paramagnetic, diamagnetic and non-magnetic materials, as suggested earlier.

- The search coil can be modified to apply RF signals to plants or animals. Experiments to find the effects of RF (Radio Frequency) on seed germination or plant growth can be made with this circuit.

CRYSTAL SET (E)

A complete radio AM receiver which requires no battery power at all can be created with a few components. The classic circuit of the well-known crystal radio or crystal receiver can be based on a germanium diode as detector, or if you prefer, the original galena crystal.

Figure 1 shows the classic circuit. L1 is a proprietary antenna tuning coil wound on a paper tube and C1 is a matching variable capacitor (365 to 500 pF). A variable capacitor can be found in old non-working transistor radios. You can also find the diode in these sets and use it for this project.

Virtually any diode can be used, and the only other circuit component is capacitor C1, connected across the output. The earphones must be a high-impedance type (headphones).

An external aerial and good physical ground connection is essential for good reception.

The circuit will tune AM band radio stations not far from your home.

Component layout is shown in ***Figure 2***.

L1 is formed by 100 turns of No. 22 to 28 AWG enamelled wire on a 1-inch diameter paper tube.

Parts List — Crystal Radio

D1	- 1N34 or any germanium diode
L1	- Coil - see text
CV	- 365 to 500 pF variable capacitor
C1	- 470 pF ceramic capacitor
PHONE	- High-impedance (2,000 to 10,000 ohms) phone

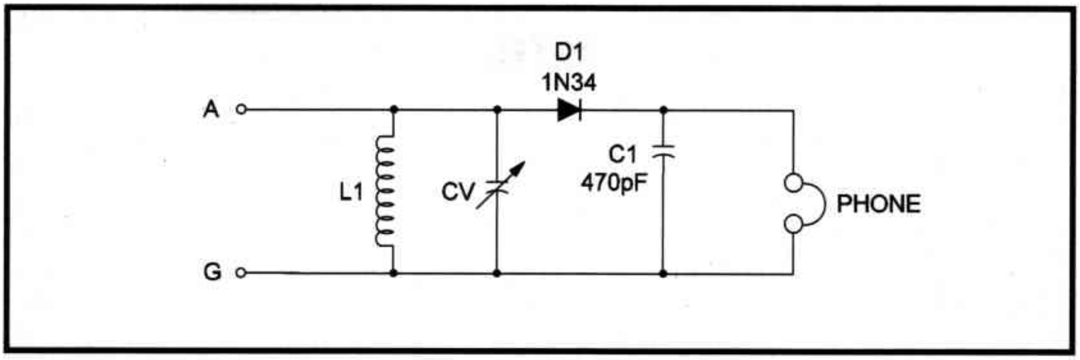
D1
1N34
A
G
L1
CV
C1
470pF
PHONE

Figure 1

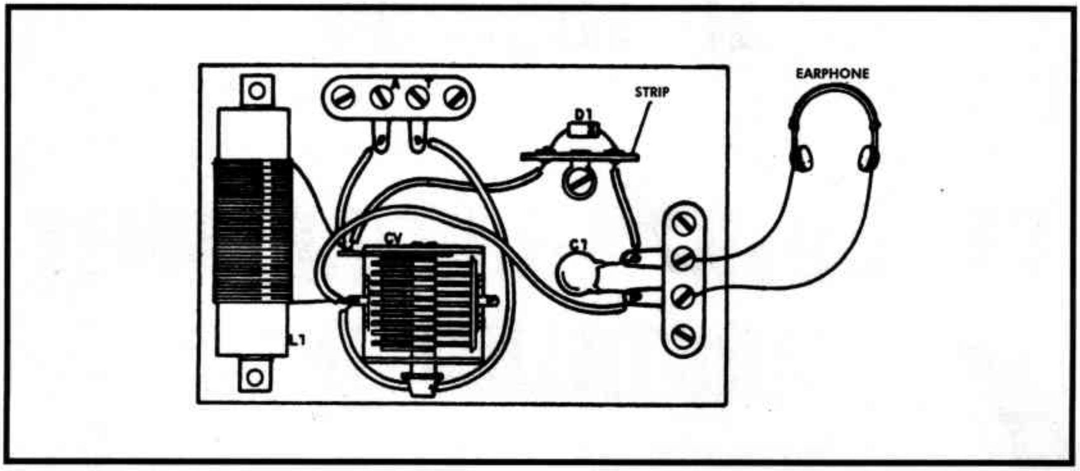
EARPHONE
STRIP
D1
CV
C1
L1

Figure 2

Ideas to Explore

To learn more about the circuit or to get better performance:

- Use an audio amplifier to get better reception.
- Replace the coil by 100 turns of No. 22 to 28 wire on a ferrite rod.

Science projects:

- This set can be used to demonstrate how old-time radio receivers operated. Explain how a galena detector works.

TEMPERATURE-CONTROLLED OSCILLATOR (E)

You can conduct interesting experiments with this temperature-dependent oscillator. Tone pitch depends upon the temperature of a sensor. The sensor is a common general-purpose silicon diode. The reverse resistance of this diode is slight, dependent on the junction temperature. The circuit can operate in a temperature range typically between -20 and +100 degrees Celsius.

Frequency range depends upon C1. This component can be altered in a large range of values. Values between 0.01 and 0.47 uF can be substituted to produce tones in the audio range.

If you want to produce “clicks” with a rate dependent upon the temperature of the sensor, try capacitors with values between 0.47 and 1 uF. Metal film or ceramic capacitors can be used in this circuit.

By adjusting P1 you can put the oscillator near the point where it begins to run. Then, any temperature rise will trigger the oscillator, making it operate as a temperature alarm.

The circuit can be powered from two or four AA cells, and current drain is low.

The sensor is any general-purpose silicon diode. The diode can be protected against moisture and water.

The schematic diagram of the Temperature-Controlled Oscillator is shown in ***Figure 1***. The sensor can be wired to the circuit by a long twisted-pair wire for remote operation.

Components layout, using a terminal strip as chassis, is shown in ***Figure 2***. As the circuit is intended for experimental purposes, you can also mount it on a solderless board.

Position of polarized components, such as the diode and power supply, should be observed.

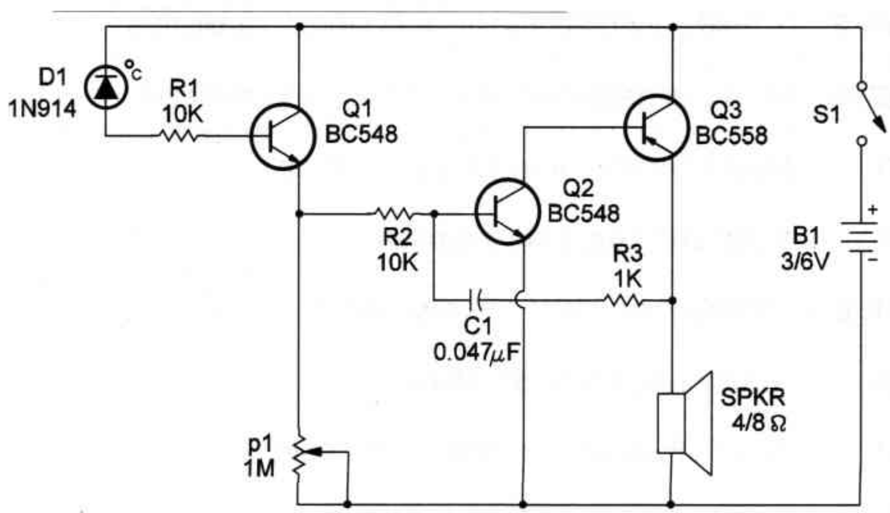
D1
1N914
°C
R1
10K
Q1
BC548
Q2
BC548
Q3
BC558
S1
R2
10K
R3
1K
C1
0.047µF
B1
3/6V
+
-
SPKR
4/8 Ω
p1
1M

Figure 1

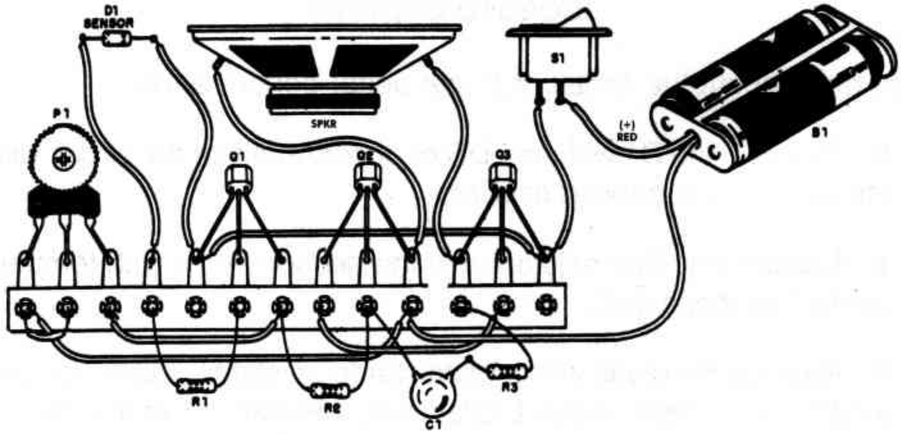
D1
SENSOR
P1
SPKR
S1
Q1
Q2
Q3
(+)
RED
B1
R1
R2
C1
R3

Figure 2

Parts List — Temperature-Controlled Oscillator

Q1, 2	- BC548 general-purpose NPN silicon transistors
Q3	- BC558 general-purpose PNP silicon transistor
D1	- 1N914 general-purpose silicon diode
P1	- 1,000,000 ohm potentiometer
R1, 2	- 10,000 ohm, 1/4W, 5% resistors
R3	- 1,000 ohm, 1/4W, 5% resistor
C1	- 0.047 uF ceramic or metal film capacitor
S1	- SPST toggle or slide switch
B1	- 3V or 6V - two or four AA cells
SPKR	- 4 or 8 ohms - 2 or 4 in. loudspeaker

Ideas to Explore

To learn more about the circuit or to get better performance:

- By replacing D1 with an LDR or phototransistor the circuit will operate as a light-controlled oscillator.
- Explain why the resistance of the sensor varies with temperature, controlling this circuit.
- Replace the diode with a transistor (conection is made using collector and emitter terminals). Explain why this device can also be used as a temperature sensor.

Science projects or uncommon uses:

- You can use this circuit as a biofeedback device by replacing the sensor with two electrodes. The subject, holding the sensor in hand, can learn how to control the tone through stress and breathing management. The circuit also operates as a lie detector using this configuration.

- Any experiment involving temperature variations can be conducted using this circuit.

FREQUENCY-MODULATION SIREN (E/P)

Acting on a capacitor charging, one oscillator (pins 1,2 and 3 of IC1) can modulate another oscillator in frequency, as described in this project. The produced tones thus run from high to low and vice-versa at a rate determined by the first oscillator frequency and also by P1 and P2 adjustments.

The circuit can be used as a siren, part of alarms, games and many other applications. With a 12V power supply the output power is up to a wattage representing an excellent audio level in a loudspeaker.

The basic circuit has tone, modulation rate and depth adjusted by trimmer potentiometer. There are also some components that can be changed to alter the performance of the siren. C1, for instance, determines modulation rate and can range from 1 to 47 uF. C2 determines modulation depth and rate and can also be altered within a large range of values. C3 determine tone frequency, ranging from 0.01 to 0.1 uF.

Power supply can range from 6 to 12V. With a 12V power supply current drain is as high as 2 or 3 amps. F1 is important to prevent problems with shorts.

Q1 is a power FET and equivalents can be used. You also can replace this transistor with a Darlington NPN power transistor such as TIP110, but you must also add a series resistor with base terminal. Use a 10,000 ohm resistor for this task.

The schematic for the Frequency-Modulation Siren is shown in ***Figure 1***. Observe that two gates are used as oscillators and the other two of a 4093B IC are used as buffer-inverters.

Components placement on a homemade printed-circuit board is shown in ***Figure 2***. For experimental purposes the circuit can also be mounted on a solderless board, and it is important to place the transistor on a large heatsink.

Position of the polarized components, such as electrolytic capacitors, should be observed.

The speaker should be a high-power type for better performance. Installing it in an enclosure will provide better sound level.

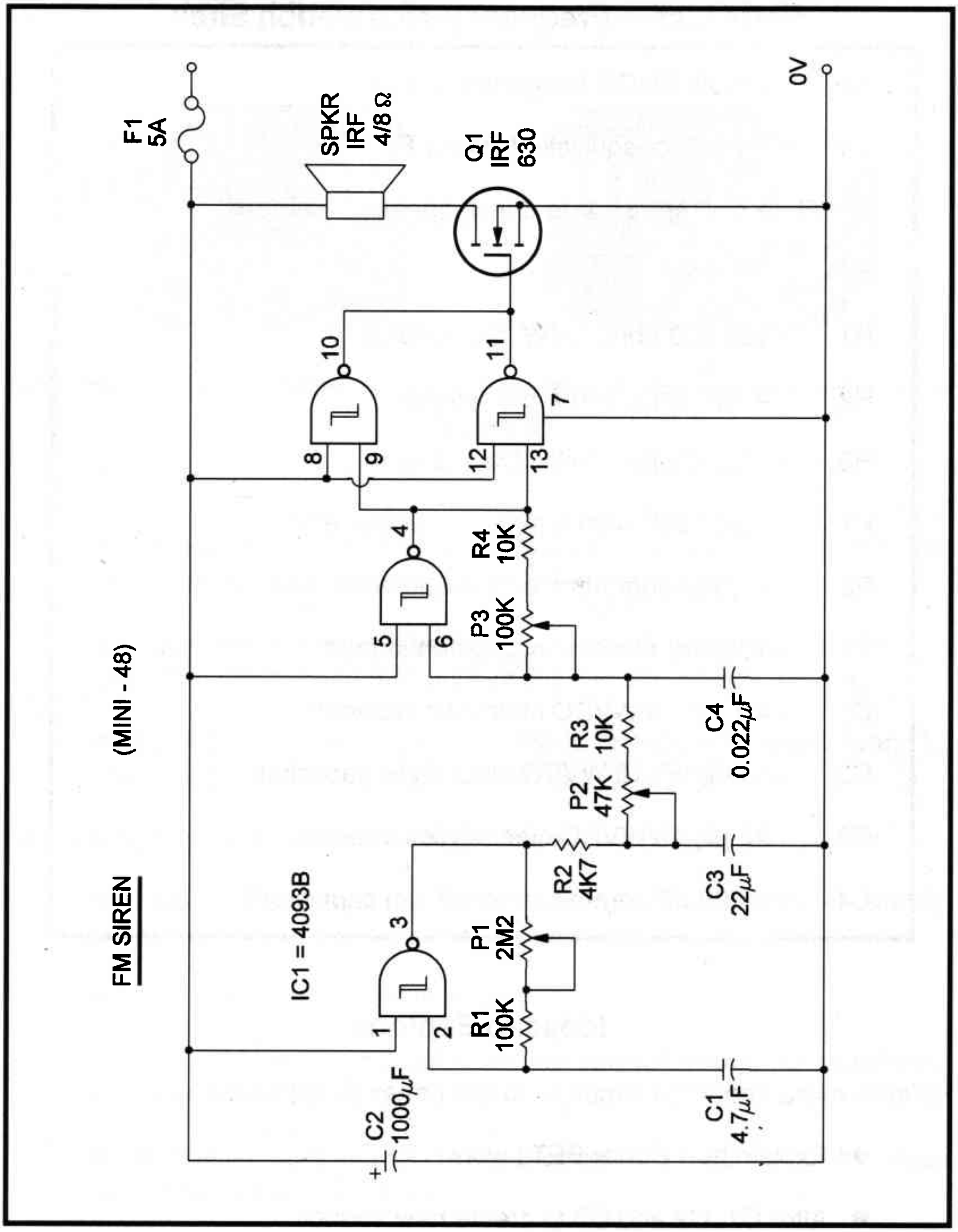
FM SIREN
(MINI - 48)
IC1 = 4093B
F1
5A
SPKR
IRF
4/8 Ω
Q1
IRF
630
0V
10
11
8
9
12
13
7
4
5
6
3
1
2
R4
10K
P3
100K
R3
10K
P2
47K
R2
4K7
P1
2M2
R1
100K
C4
0.022μF
C3
22μF
C1
4.7μF
C2
1000μF
+

Figure 1

Parts List — Frequency-Modulation Siren

IC1 - 4093B CMOS Integrated Circuit

Q1 - IRF630 or equivalent Power FET

SPKR - 4 or 8 ohms, 4- to 8-inch speaker - see text

F1 - 5A fuse

R1 - 100,000 ohm, 1/4W, 5% resistor

R2 - 4,700 ohm, 1/4W, 5% resistor

R3, 4 - 10,000 ohm, 1/4W, 5% resistors

P1 - 2,200,000 ohm trimmer potentiometer

P2 - 47,000 ohm trimmer potentiometer

P3 - 100,000 ohm trimmer potentiometer

C1 - 4.7 uF, 16 WVDC eletrolytic capacitor

C2 - 1,000 uF, 16 WVDC electrolytic capacitor

C3 - 22 uF, 16 WVDC electrolytic capacitor

C4 - 0.022 uF ceramic or metal film capacitor

Ideas to Explore

To learn more about the circuit or to get better performance:

- Explain how Power FETs work.
- Alter C1, C2 and C3 to create new sounds.

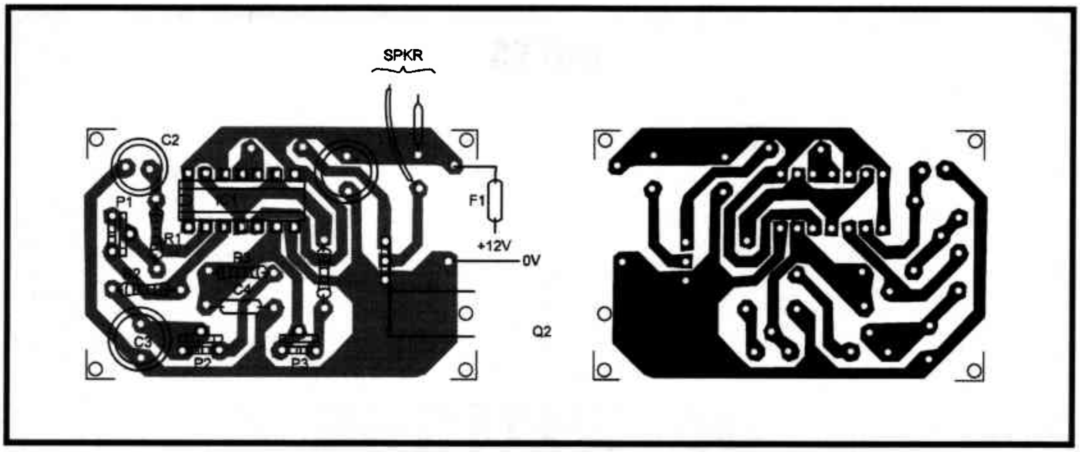

Figure 2

- Wire pins 1 and 5 together and to an external control circuit. The siren can be controlled by external logic circuits.

- Determine the logic level in pins 10 and 11 when pins 1 and 5 go low. What will happen with current drain in this case?

- Replace Q1 by a TIP31 or BD135 to operate with less power. The power supply should be reduced to 6 or 9V in this case.

Science projects and uncommon applications:

- Replace R1 with an LDR and experiment with a light-controlled siren.

- Use this siren as part of alarms.

- Using this siren you can generate special sounds to experiment with animal conditioning.

- Modulated ultrasounds can be produced by reducing C4 to values as low as 2,200 or 4,700 pF.

COIN TOSSER (E/P)

This circuit simulates the flipping of a coin by merely pressing S1. Of course, the electronic version, if used for important decisions (or experiments), can't be loaded or weighted and is 100% random.

The circuit has two LEDs that flick alternatingly, in a frequency rate determined by C1, when power is ON. When you press S1 the circuit stops immediately and only one LED remains ON.

Power supply can range from 5 to 12 volts, and this means that you can use AA cells in a portable version.

Some components can be altered according to the intended application: C1, for instance can be altered in a range from 0.01 to 0.47 uF. You can also use a touch sensor to stop the tosser or increase R1's value to up to 10,000,000 ohms.

The schematic diagram for the coin tosser is shown in ***Figure 1***. Only one gate a 4093 IC is used as oscillator to produce random pulses in this circuit.

Components placement on a homemade printed-circuit board is shown in ***Figure 2***. The circuit can easily be housed in a small plastic box.

Position of the polarized components, such as the LEDs and electrolytic capacitor, should be observed.

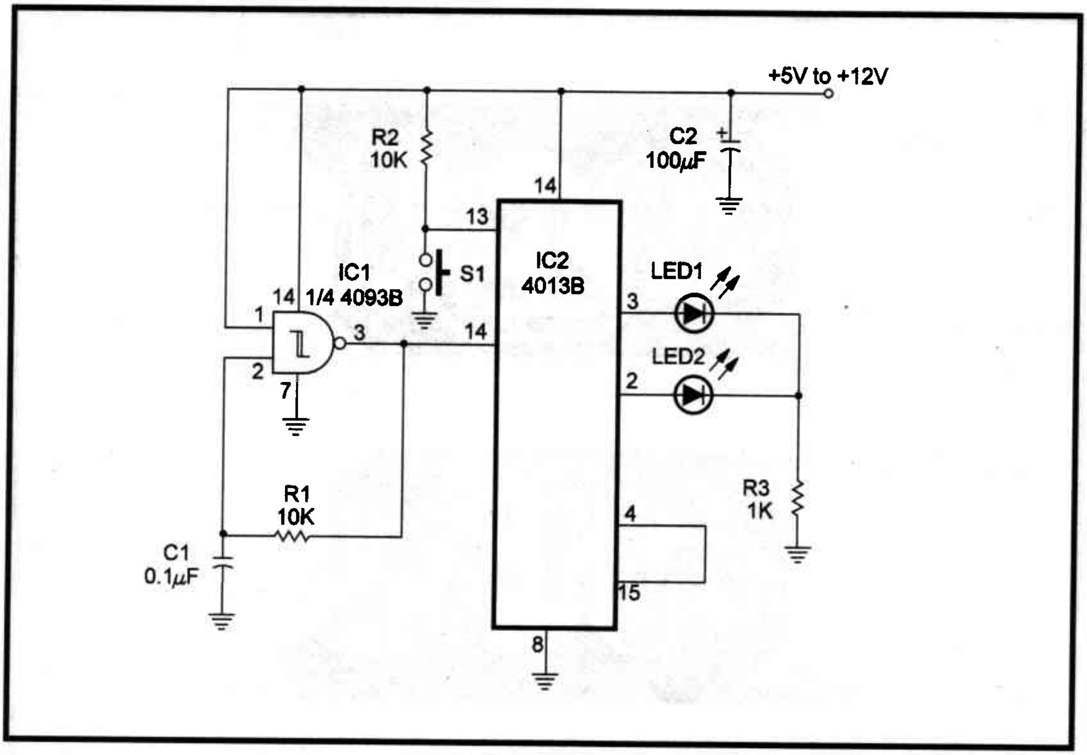
+5V to +12V
R2
10K
C2
100µF
14
13
IC2
4013B
IC1
14 1/4 4093B
S1
LED1
1
3
14
3
LED2
2
2
7
R1
10K
R3
1K
4
C1
0.1µF
15
8

Figure 1

Parts List — Coin Tosser

IC1	- 4093B CMOS Integrated Circuit
LEDs	- Common red, yellow or green
S1	- SPST Momentary switch
R1, 2	- 10,000 ohm, 1/4W, 5% resistors
R3	- 1,000 ohm, 1/4W, 5% resistor
C1	- 0.1 uF ceramic or metal film capacitor
C2	- 100 uF, 16 WVDC electrolytic capacitor

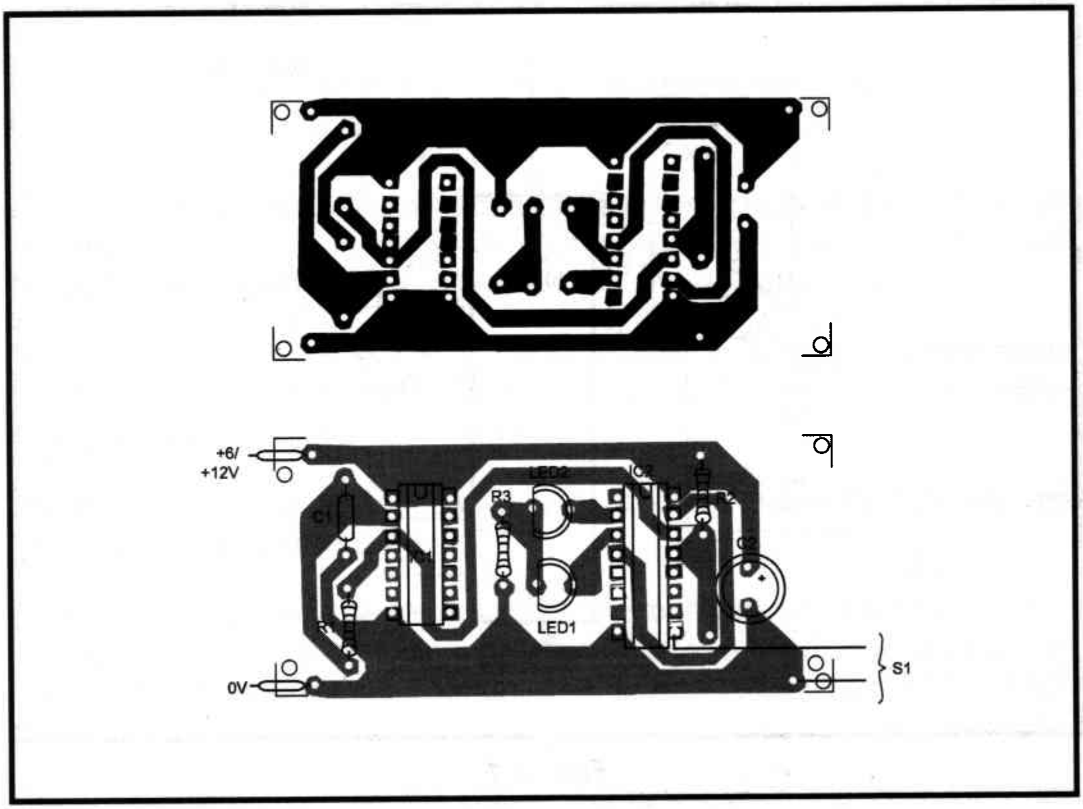

Figure 2

Ideas to Explore

To learn more about the circuit or to get better performance:

- Instead of using LEDs, transistors can be used to drive incandescent lamps.

- S1 can be placed as far as you want from the circuit. Common parallel wire can be used for this task.

- Explain how a flip-flop works and how you can guarantee that the circuit is 100% random.

Science and uncommon applications:

- You can conduct ESP (Extra-Sensory Perception) experiments using this circuit. S1 can be placed far from the LEDs for better results.
- Experiments with probability theory can be conducted using this circuit.
- Try to couple a sequence generator to the oscillator used in this circuit to get a 1-to-10 raffle circuit.

PLASMA OSCILLATOR (E)

A flame is a conductive medium that can be used in an uncommon feedback loop to control the frequency of an audio oscillator. This configuration can be used as a confirmation of the conductivity of a flame.

The audio oscillator described in this project is controlled by the "fourth" matter state, the "plasma" or an ionized gas, and can be used as part of an interesting experiment in physics.

The flame can be produced by a simple match or a candle and the flickering effect will modulate the generated sound.

The circuit runs in frequencies between 1 and 500 Hz, depending on the electrodes and flame positions.

A schematic diagram of the Plasma Oscillator is shown in ***Figure 1.*** The high input impedance of a 4093B is fundamental in this project. A flame typically has a resistance in the range of tenths of megohms.

The printed circuit board for the components positioning is shown in ***Figure 2***.

Sensor details are also given in that figure. Two wires, one placed near the other, with about 1 inch of bared length as shown in the figure, form the "plasma sensor". The flame should involve the two bared wires at the same time to allow the feedback current to flow and the oscillator to operate.

An output stage using a piezoelectric transducer can replace the transistor and loudspeaker. With a lower current drain this configuration can be powered from four AA cells or a 9V battery.

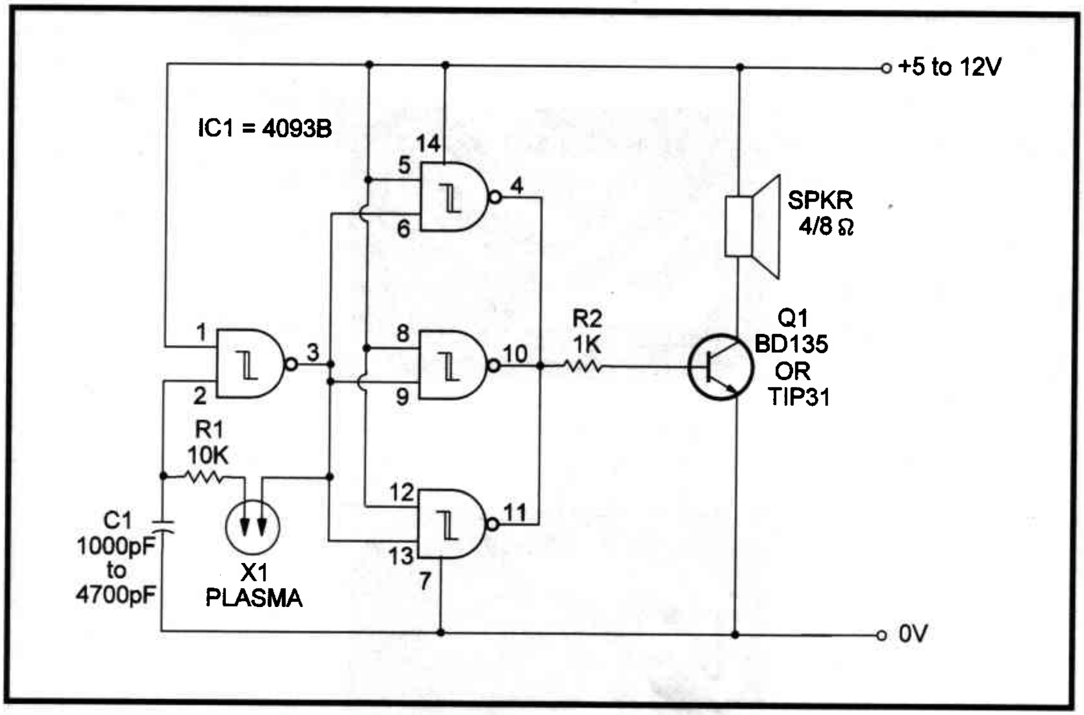
+5 to 12V
IC1 = 4093B
14
5
4
6
SPKR
4/8 Ω
1
3
8
10
R2
1K
Q1
BD135
OR
TIP31
2
9
R1
10K
12
11
C1
1000pF
to
4700pF
X1
PLASMA
13
7
0V

Figure 1

Parts List — Plasma Oscillator

IC1 - 4093B CMOS Integrated Circuit

SPKR - 4/8 ohm 4-inch loudspeaker

Q1 - BD135 or TIP31 Power NPN silicon transistor

R1 - 10,000 ohm, 1/4W, 5% resistor

R2 - 1,000 ohm, 1/4W, 5% resistor

X1 - Plasma sensor - see text

C1 - 1,000 to 4,700 pF ceramic or metal film capacitor

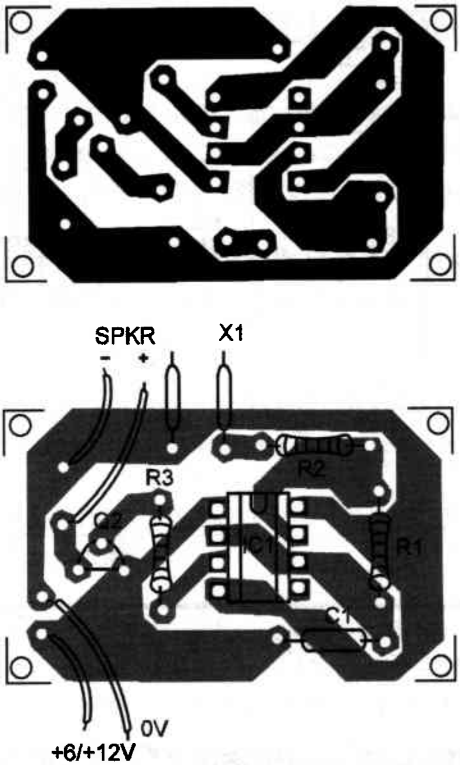

Figure 2

Ideas to Explore

To get better performance or to learn more about the circuit:

- Explain why the circuit doesn't operate using a neon or fluorescent lamp as a plasma medium.
- Explain what "plasma" is.
- You can alter this circuit to use it as a flame alarm.

Science applications:

- Off course, the basic idea of this project is to use it in an experiment involving plasma. Plasma is produced when an ionized gas loses electrons. The the gas converts in a "soup" of free electrons and ionized atoms.
- Use different kinds of flames, such as those produced by matches, gas, paper, etc., and compare their conductivity by the tone pitch.

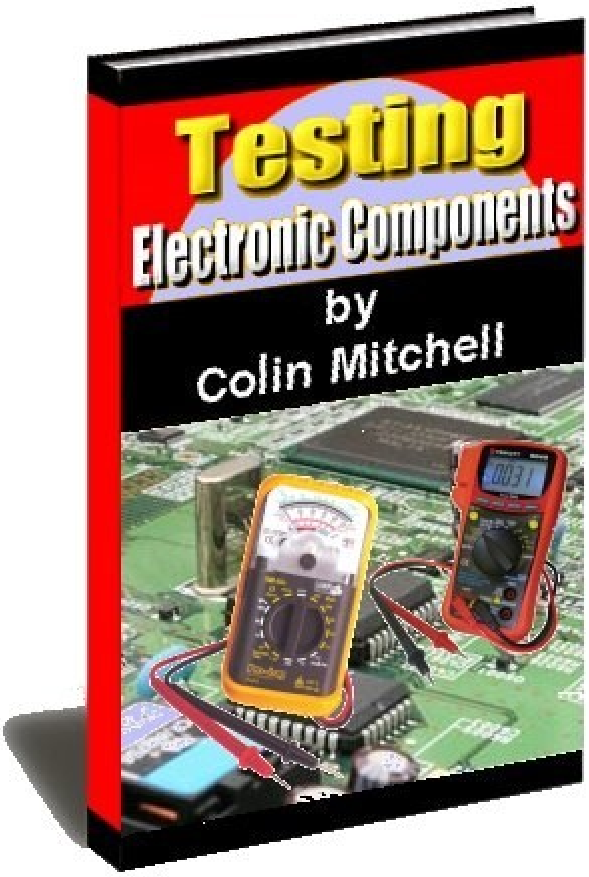

This eBook shows you how to TEST COMPONENTS.
To do this you need "TEST GEAR." The best item of Test Gear is a MULTIMETER. It can test almost 90% of all components. And that's what we will do in this eBook:

CONTENTS

Analogue Multimeter
Audio Stages
Batteries - testing
Burnt Resistor
Buying A Multimeter
Capacitors
Capacitors in Parallel
Cells - batteries
Circuit Symbols
Co-Ax Cables
Colour Code (Resistor)
Coils
Continuity
Creating any value of R
Current - measuring
Damper Diodes
Darlington Transistors
Digital Chips
Digital Multimeter
Diodes
Earth Leakage Detectors
Electrolytics
Multimeters
Non-polar Capacitor (electrolytic)
"Open" Resistor - damaged
Opto-couplers
Parallel - resistors
Piezo Diaphragms
Piezo Buzzers
Potentiometers
Pots - testing
Power Diodes
Relays
Remote Controls
Resistor Colour Code
Resistor Networks
Resistors - series
Schottky Diodes
SCRs
Symbols
Signal Injector
Silicon diodes
Soldering
Spark Gaps

FETs
Focus pots
Fuses
Germanium Diodes
Impedance - of a stage
IC's - also called Digital Chips
IC's - Analogue Chips
"In-Circuit" testing
Inductors
Inductors - measuring
Integrated Circuits
Isolation Transformer
LEDs
Logic Probe MkIIB
Logic Probe - Simple
Logic Probe - using CD4001
Logic Probe - using CD4011
Making your own components
Measuring Resistance
Measuring Voltage
Mica Washers and Insulators
Motor - testing
MOSFETs
Super Probe MkII
Surface Mount - Packs
Surface-Mount Resistors
Surface-Mount Resistor Markings
Switches
Testing A Circuit
Testing A Resistor
Testing Components "In-Circuit"
Transformers
Transistor Outlines
Transistors
Triacs
Unknown resistors - testing
Using A Multimeter
Voltage Regulators
Voltages on a circuit
Yokes
Zener Diodes
4-Band Resistors
5-Band Resistors

MULTIMETERS

There are two types:

DIGITAL and ANALOGUE

A **Digital Multimeter** has a set of digits on the display and an Analogue Multimeter has a scale with a pointer (or needle).

You really need both types to cover the number of tests needed for designing and repair-work. We will discuss how they work, how to use them and some of the differences between them.

024
UNI-T
6.43

DIGITAL AND ANALOGUE MULTIMETERS

BUYING A MULTIMETER

There are many different types on the market.
The cost is determined by the number of ranges and also the extra features such as diode tester, buzzer (continuity), transistor tester, high DC current and others.
Since most multimeters are reliable and accurate, buy one with the greatest number of ranges at the lowest cost.
This article explains the difference between a cheap analogue meter, an expensive analogue meter and a digital meter. You will then be able to work out which two meters you should buy.

Multimeters are sometimes called a "meter", a "VOM" (Volts-Ohms-Milliamps or Volt Ohm Meter) or "multi-tester" or even "a tester" - they are all the same.

USING A MULTIMETER

Analogue and digital multimeters have either a rotary selector switch or push buttons to select the appropriate function and range. Some Digital Multimeter (DMMs) are auto ranging; they automatically select the correct range of voltage, resistance, or current when doing a test. However you need to select the function.

Before making any measurement you need to know what you are checking. If you are measuring voltage, select the AC range (10v, 50v, 250v, or 1000v) or DC range (0.5v, 2.5v, 10v, 50v, 250v, or 1000v). If you are measuring resistance, select the Ohms range (x1, x10, x100, x1k, x10k). If you are measuring current, select the appropriate current range DCmA 0.5mA, 50mA, 500mA. Every multimeter is different however the photo below shows a low cost meter with the basic ranges.

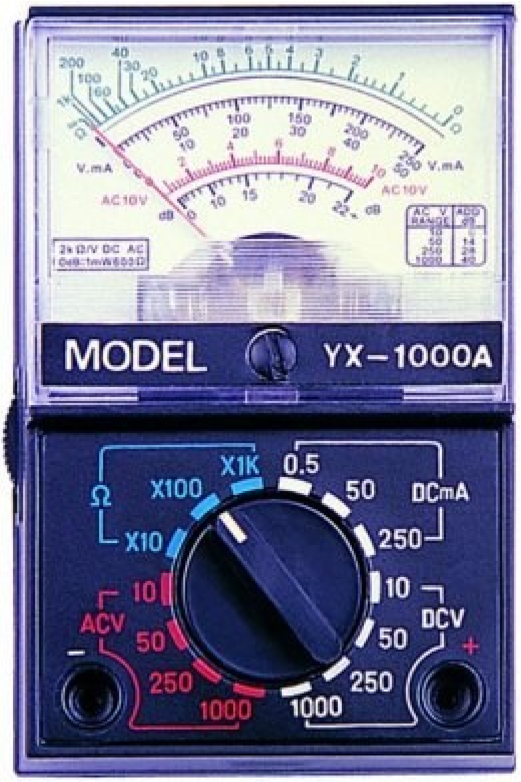
MODEL YX-1000A
Ω
X10
X100
X1K
0.5
50
DCmA
250
10
DCV
50
250
1000
ACV
10
50
250
1000
V.mA
AC10V
dB
2kΩ/V DC AC
0dB:1mW600Ω

The most important point to remember is this:
You must select a voltage or current range that is bigger or HIGHER than the maximum expected value, so the needle does not swing across the scale and hit the "end stop."
If you are using a DMM (Digital Multi Meter), the meter will indicate if the voltage or current is higher than the selected scale, by showing "OL" - this means "Overload." If you are measuring resistance such as 1M on the x10 range the "OL" means "Open Loop" and you will need to change the range. Some meters show "1' on the display when the measurement is higher than the display will indicate and some flash a set of digits to show over-voltage or over-current. A "-1" indicates the leads should be reversed for a "positive reading."
If it is an AUTO RANGING meter, it will automatically produce a reading, otherwise the selector switch must be changed to another range.

HV
000
DCV
OFF
ACV
DCA
hFE
DT-832
DIGITAL
MULTIMETER
10ADC
VΩmA
COM

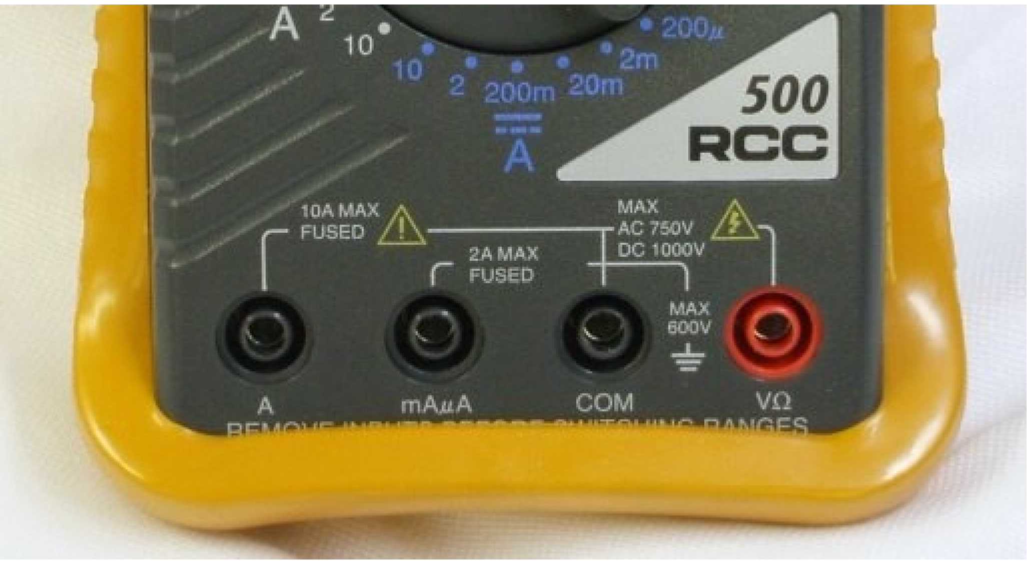
500
RCC
10A MAX
FUSED
2A MAX
FUSED
MAX
AC 750V
DC 1000V
MAX
600V
A
mAµA
COM
VΩ

The Common (negative) lead ALWAYS fits into
the "COM" socket. The red lead fits into the
red socket for Voltage and Resistance.
Place the red lead (red banana plug)
into "A" (for HIGH CURRENT "Amps")
or mA,uA for LOW CURRENT.

The black "test lead" plugs into the socket marked "-" "Common", or "Com," and the red "test lead" plugs into meter socket marked "+" or "V-W-mA." The third banana socket measures HIGH CURRENT and the positive (red lead) plugs into this. You DO NOT move the negative "-" lead at any time. The following two photos show the test leads fitted to a digital meter. The probes and plugs have "guards" surrounding the probe tips and also the plugs so you can measure high voltages without getting near the voltage-source.

FLUKE
600.0
COM
10ADC

Analogue meters have an "Ohms Adjustment" to allow for the change in voltage of the battery inside the meter (as it gets old).

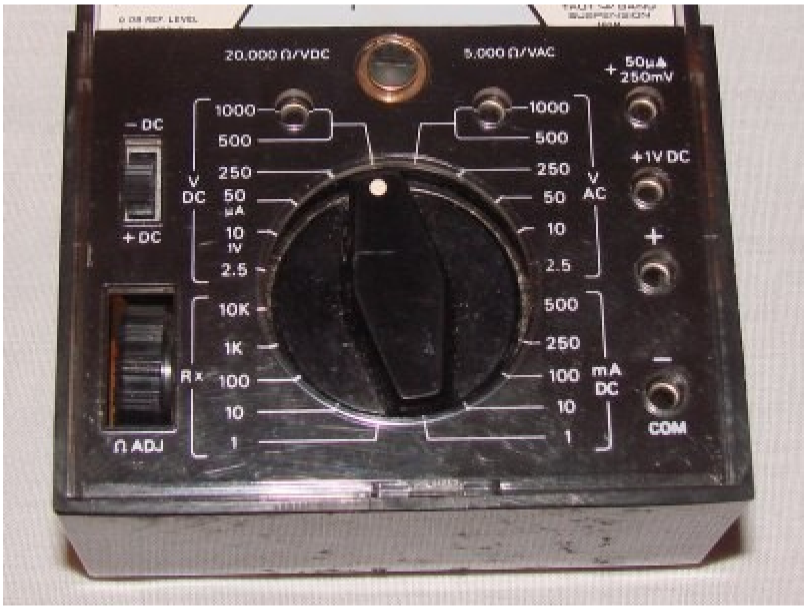

"Ohms Adjust" is also called "ZERO SET"
The sensitivity of this meter is 20,000ohms/volt
on the DC ranges and 5k/v on the AC ranges

Before taking a resistance reading (each time on any of the Ohms scales) you need to "ZERO SET" the scale, by touching the two probes together and adjust the pot until the needle reads "0" (swings FULL SCALE). If the pointer does not reach full scale, the batteries need replacing. Digital multimeters do not need "zero adjustment."

MEASURING VOLTAGE

Most of the readings taken with a multimeter will be VOLTAGE readings.
Before taking a reading, you should select the highest range and if the needle does not move up scale (to the right), you can select another range.
Always switch to the highest range before probing a circuit and keep your fingers away from the component being tested.
If the meter is Digital, select the highest range or use the auto-ranging feature, by selecting "V." The meter will automatically produce a result, even if the voltage is AC or DC.
If the meter is not auto-ranging, you will have to select V⎓ if the voltage is from a DC source or V~ if the voltage is from an AC source. DC means Direct Current and the voltage is coming from a battery or supply where the voltage is steady and not changing and AC means Alternating Current where the voltage is coming from a voltage that is rising and falling.
You can measure the voltage at different points in a circuit by connecting the black probe to chassis. This is the 0v reference and is commonly called "Chassis" or "Earth" or "Ground" or "0v."
The red lead is called the "measuring lead" or "measuring probe" and it can measure voltages at any point in a circuit. Sometimes there are "test points" on a circuit and these are wires or loops designed to hold the tip of the red probe (or a red probe fitted with a mini clip).
You can also measure voltages ACROSS A COMPONENT. In other words, the reading is taken in PARALLEL with the component. It may be the voltage across a transistor, resistor, capacitor, diode or coil. In most cases this voltage will be less than the supply voltage.
If you are measuring the voltage in a circuit that has a HIGH IMPEDANCE, the reading will be inaccurate, up to 90% !!!, if you use a cheap analogue meter.

Here's a simple case.

The circuit below consists of two 1M resistors in series. The voltage at the mid point will be 5v when nothing is connected to the mid point. But if we use a cheap analogue multimeter set to 10v, the resistance of the meter will be about 100k, if the meter has a sensitivity of 10k/v and the reading will be incorrect.
Here how it works:
Every meter has a sensitivity. The sensitivity of the meter is the sensitivity of the movement and is the amount of current required to deflect the needle FULL SCALE.
This current is very small, normally 1/10th of a milliamp and corresponds to a sensitivity of 10k/volt (or 1/30th mA, for a sensitivity of 30k/v).
If an analogue meter is set to 10v, the internal resistance of the meter will be 100k for a 10k/v movement.
If this multimeter is used to test the following circuit, the reading will be inaccurate.

The reading should be 5v as show in diagram **A**.
But the analogue multimeter has an internal resistance of 100k and it creates a circuit shown in **C**.
The top 1M and 100k from the meter create a combined PARALLEL resistance of 90k. This forms a series circuit with the lower 1M and the meter will read less than 1v
If we measure the voltage across the lower 1M, the 100k meter will form a value of resistance with the lower 1M and it will read less than 1v
If the multimeter is 30k/v, the readings will be 2v. See how easy it is to get a totally inaccurate reading.

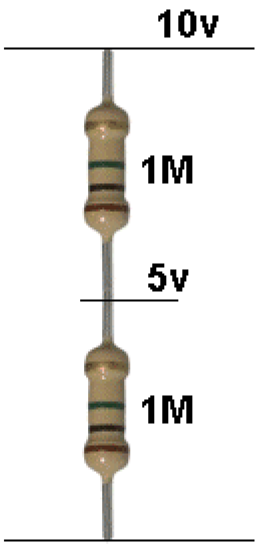

A

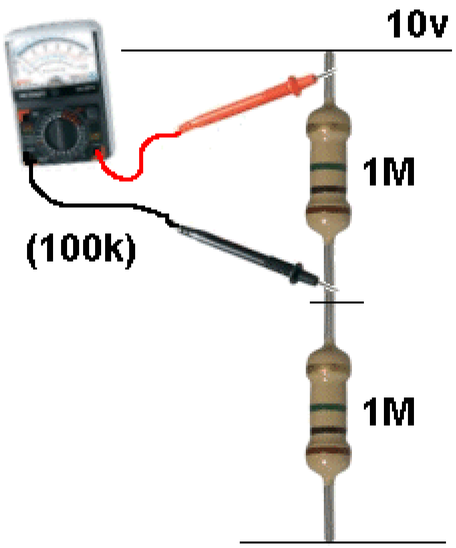

B

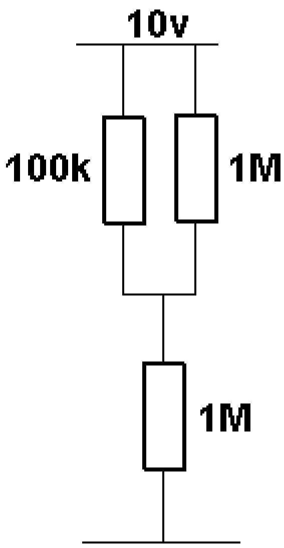

C

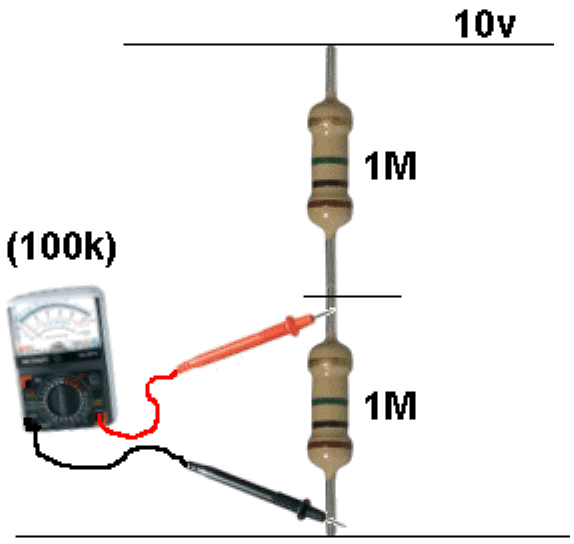

D

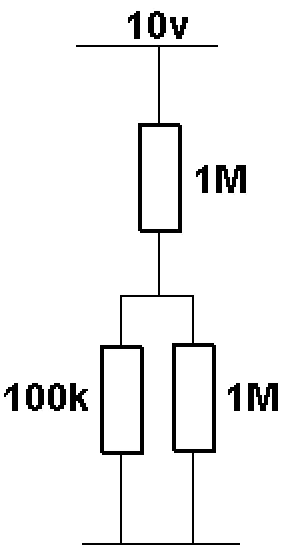

E

This introduces two new terms:
HIGH IMPEDANCE CIRCUIT and "RESISTORS in SERIES and PARALLEL."

If the reading is taken with a Digital Meter, it will be more accurate as a DMM does not take any current from the circuit (to activate the meter). In other words it has a very HIGH input impedance. Most Digital Multimeters have a fixed input resistance (impedance) of 10M - no matter what scale is selected. That's the reason for choosing a DMM for high impedance circuits. It also gives a reading that is accurate to about 1%.

MEASURING VOLTAGES IN A CIRCUIT

You can take many voltage-measurements in a circuit. You can measure "across" a component, or between any point in a circuit and either the positive rail or earth rail (0v rail). In the following circuit, the 5 most important voltage-measurements are shown. Voltage "A" is across the electret microphone. It should be between 20mV and 500mV. Voltage "B" should be about 0.6v. Voltage "C" should be about half-rail voltage. This allows the transistor to amplify both the positive and negative parts of the waveform. Voltage "D" should be about 1-3v. Voltage "E" should be the battery voltage of 12v.

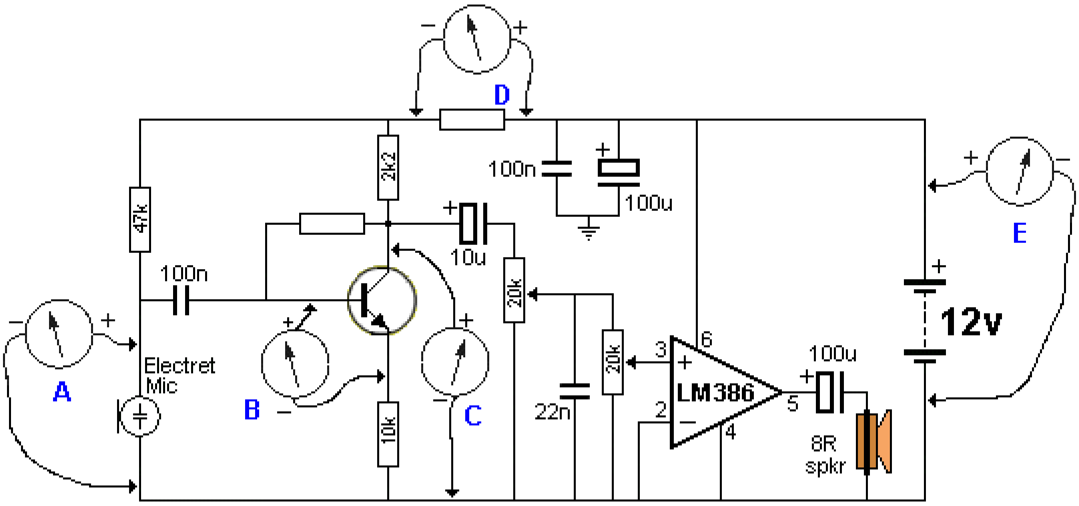

MEASURING VOLTAGES IN A CIRCUIT

MEASURING CURRENT

You will rarely need to take current measurements, however most multimeters have DC current ranges such as 0.5mA, 50mA, 500mA and 10Amp (via the extra banana socket) and some meters have AC current ranges. Measuring the current of a circuit will tell you a lot of things. If you know the normal current, a high or low current can let you know if the circuit is overloaded or not fully operational.

Current is always measured when the circuit is working (i.e: with power applied).
It is measured IN SERIES with the circuit or component under test.
The easiest way to measure current is to remove the fuse and take a reading across the fuse-holder.
Or remove one lead of the battery or turn the project off, and measure across the switch.
If this is not possible, you will need to remove one end of a component and measure with the two probes in the "opening."
Resistors are the easiest things to desolder, but you may have to cut a track in some circuits. You have to get an "opening" so that a current reading can be taken.
The following diagrams show how to connect the probes to take a CURRENT reading.
Do not measure the current ACROSS a component as this will create a "short-circuit."
The component is designed to drop a certain voltage and when you place the probes across this component, you are effectively adding a "link" or "jumper" and the voltage at the left-side of the component will appear on the right-side. This voltage may be too high for the circuit being supplied and the result will be damage.

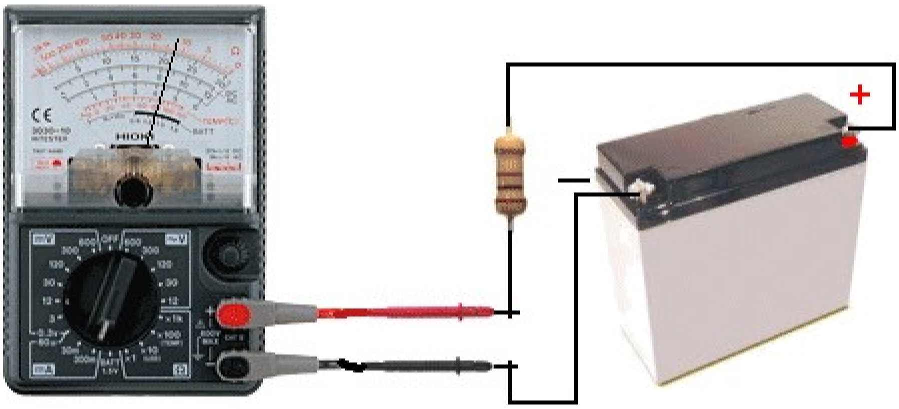

Measuring current through a resistor

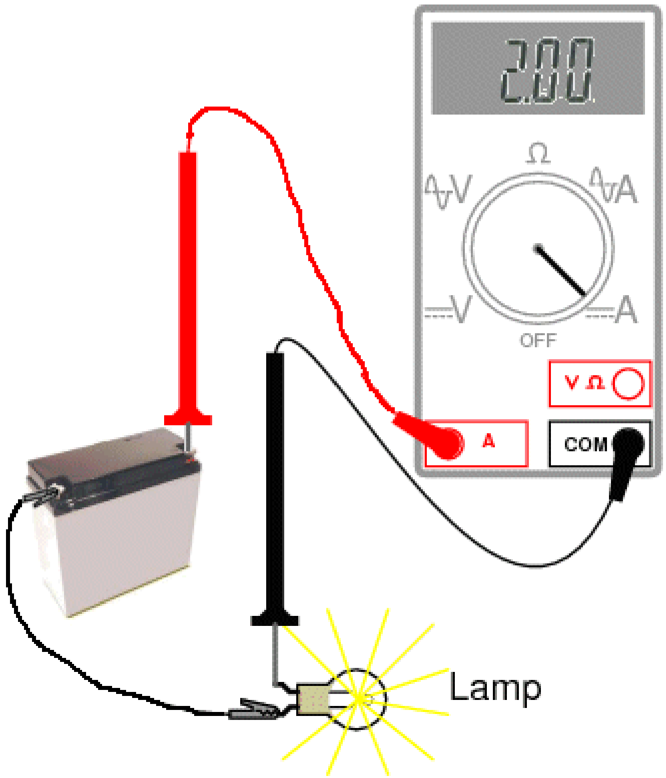

Measuring the current of a globe

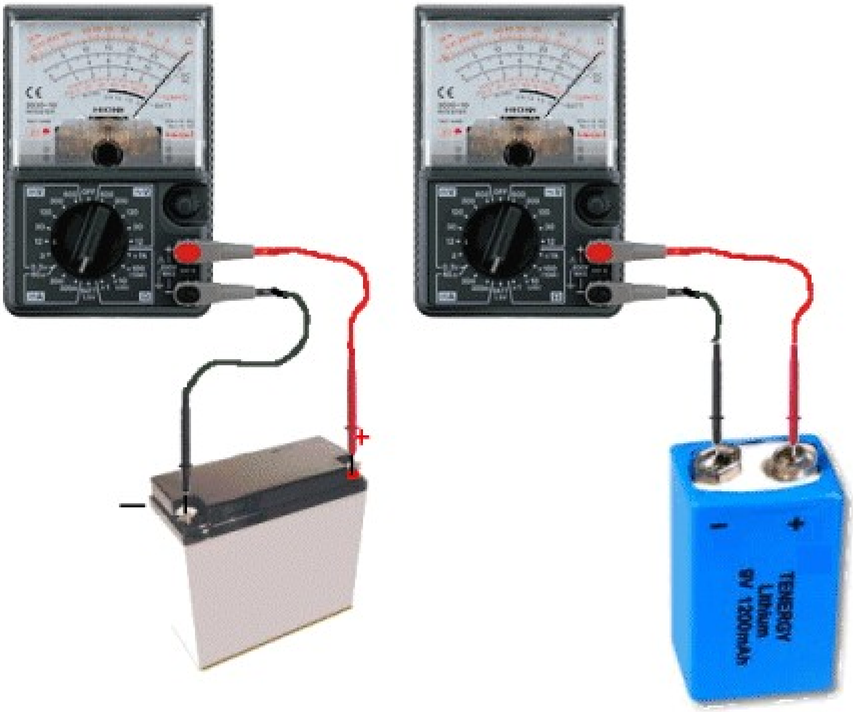

Do NOT measure the CURRENT of a battery
(by placing the meter directly across the terminals)
A battery will deliver a very HIGH current
and damage the meter

Do not measure the "current a battery will deliver" by placing the probes across the terminals. It will deliver a very high current and damage the meter instantly. There are special battery testing instruments for this purpose.
When measuring across an "opening" or "cut," place the red probe on the wire that supplies the voltage (and current) and the black probe on the other wire. This will produce a "POSITIVE" reading. A positive reading is an UPSCALE READING and the pointer will move across the scale - to the right. A "NEGATIVE READING" will make the pointer hit the "STOP" at the left of the scale and you will not get a reading. If you are using a Digital Meter, a negative sign "-" will appear on the screen to indicate the probes are around the wrong way. No damage will be caused. It just indicates the probes are

connected incorrectly.
If you want an accurate CURRENT MEASUREMENT, use a digital meter.

MEASURING RESISTANCE

Turn a circuit off before measuring resistance.
If any voltage is present, the value of resistance will be incorrect.
In most cases you cannot measure a component while it is in-circuit. This is because the meter is actually measuring a voltage across a component and calling it a "resistance." The voltage comes from the battery inside the meter. If any other voltage is present, the meter will produce a false reading.
If you are measuring the resistance of a component while still "in circuit," (with the power off) the reading will be lower than the true reading.

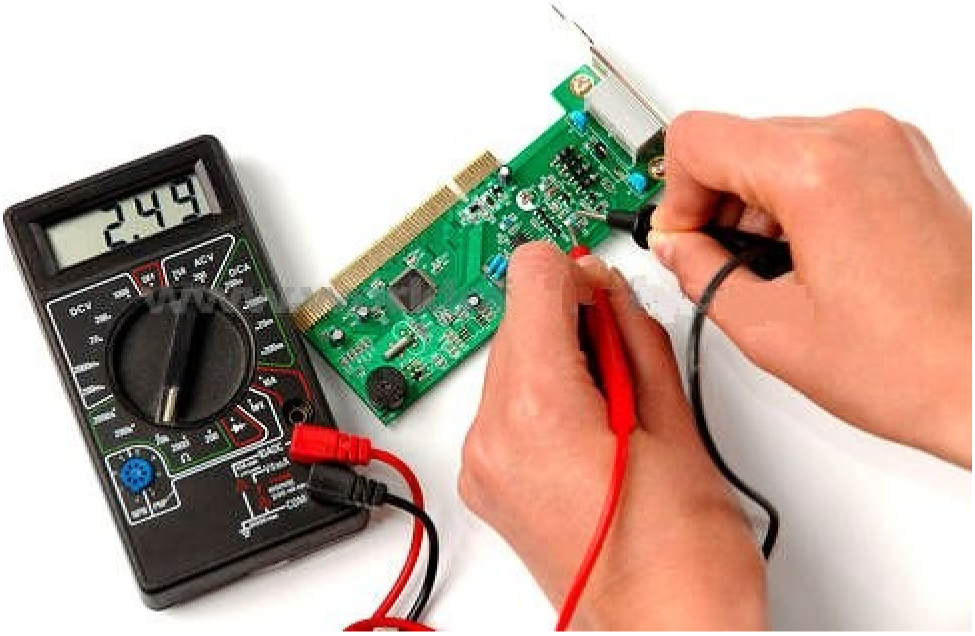

Measuring resistance

138
rolson
27259

Measuring resistance of a heater
(via the leads)

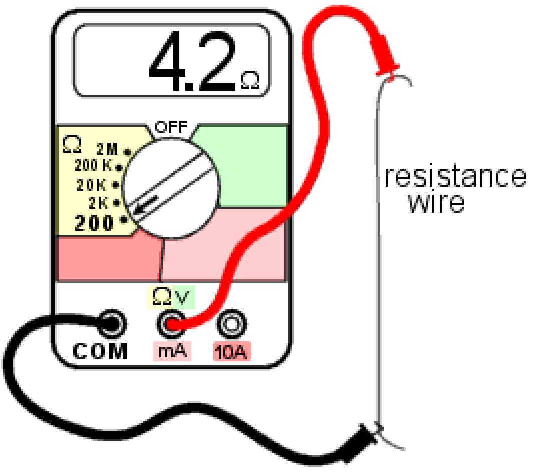
4.2 Ω
OFF
Ω
2M
200 K
20K
2K
200
ΩV
COM
mA
10A
resistance wire

Measuring the resistance of a piece of resistance-wire

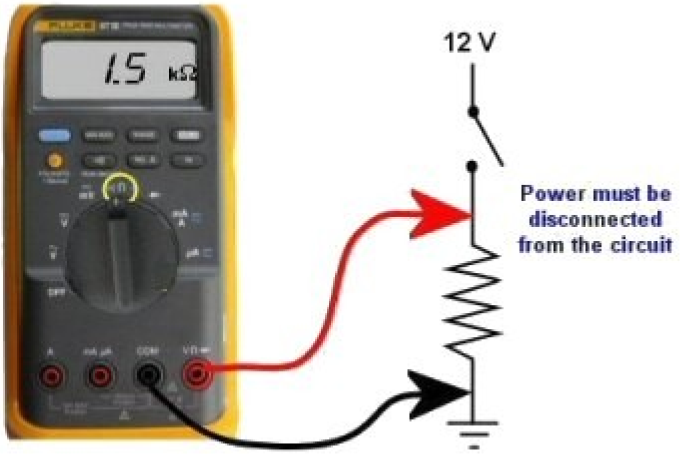
1.5 kΩ
12 V
Power must be disconnected from the circuit

Measuring the resistance of a resistor

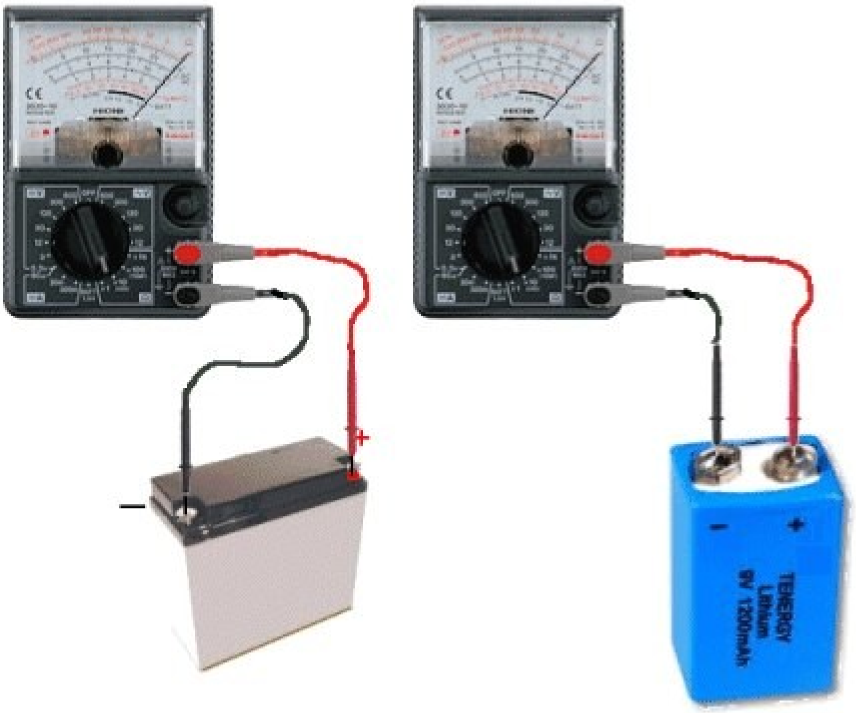
+
−
−
+
TENERGY
Lithium
9V 1200mAh

Do not measure the "Resistance of a Battery"

1. Do not measure the "resistance of a battery." The resistance of a battery (called the Internal impedance) is not measured as shown in the diagrams above. It is measured by creating a current-flow and measuring the voltage across the battery. Placing a multimeter set to **resistance** (across a battery) will destroy the meter.
2. Do not try to measure the resistance of any voltage or any "supply."

Resistance is measured in OHMs.
The resistance of a 1cm x 1cm bar, one metre long is 1 ohm.
If the bar is thinner, the resistance is higher. If the bar is longer, the resistance is higher.
If the material of the bar is changed, the resistance is higher.
When carbon is mixed with other elements, its resistance increases and this knowledge is used to make RESISTORS.
Resistors have RESISTANCE and the main purpose of a resistor is to reduce the CURRENT FLOW.
It's a bit like standing on a hose. The flow reduces.
When current flow is reduced, the output voltage is also reduced and that why the water does not spray up so high. Resistors are simple devices but they produce many different effects in a circuit.
A resistor of nearly pure carbon may be 1 ohm, but when non-conducting "impurities" are added, the same-size resistor may be 100 ohms, 1,000 ohms or 1 million ohms.
Circuits use values of less than 1 ohm to more than 22 million ohms.

Resistors are identified on a circuit with numbers and letters to show the exact value of resistance - such as 1k 2k2 4M7
The letter Ω (omega - a Greek symbol) is used to identify the word "Ohm."
but this symbol is not available on some word-processors, so the letter "R" is used. The letter "E" is also sometimes used and both mean "Ohms."

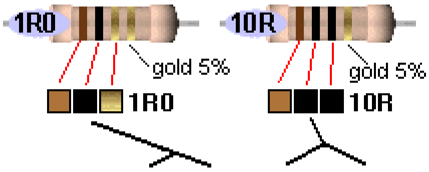

A one-ohm resistor is written "1R" or "1E." It can also be written "1R0" or "1E0."
A resistor of one-tenth of an ohm is written "0R1" or "0E1." The letter takes the place of the decimal point.
10 ohms = 10R
100 ohms = 100R
1,000 ohms = 1k (k= kilo = one thousand)
10,000 ohms = 10k
100,000 ohms = 100k
1,000,000 ohms = 1M (M = MEG = one million)

The size of a resistor has nothing to do with its resistance. The size determines the wattage of the resistor - how much heat it can dissipate without getting too hot.
Every resistor is identified by colour bands on the body, but when the resistor is a surface-mount device, numbers are used and sometimes letters.
You MUST learn the colour code for resistors and the following table shows all the colours for the most common resistors from 1/10th of an ohm to 22 Meg ohms for resistors with 5% and 10% tolerance.

If 3rd band is gold, Divide by 10
If 3rd band is silver, Divide by 100
(to get 0.22ohms etc)

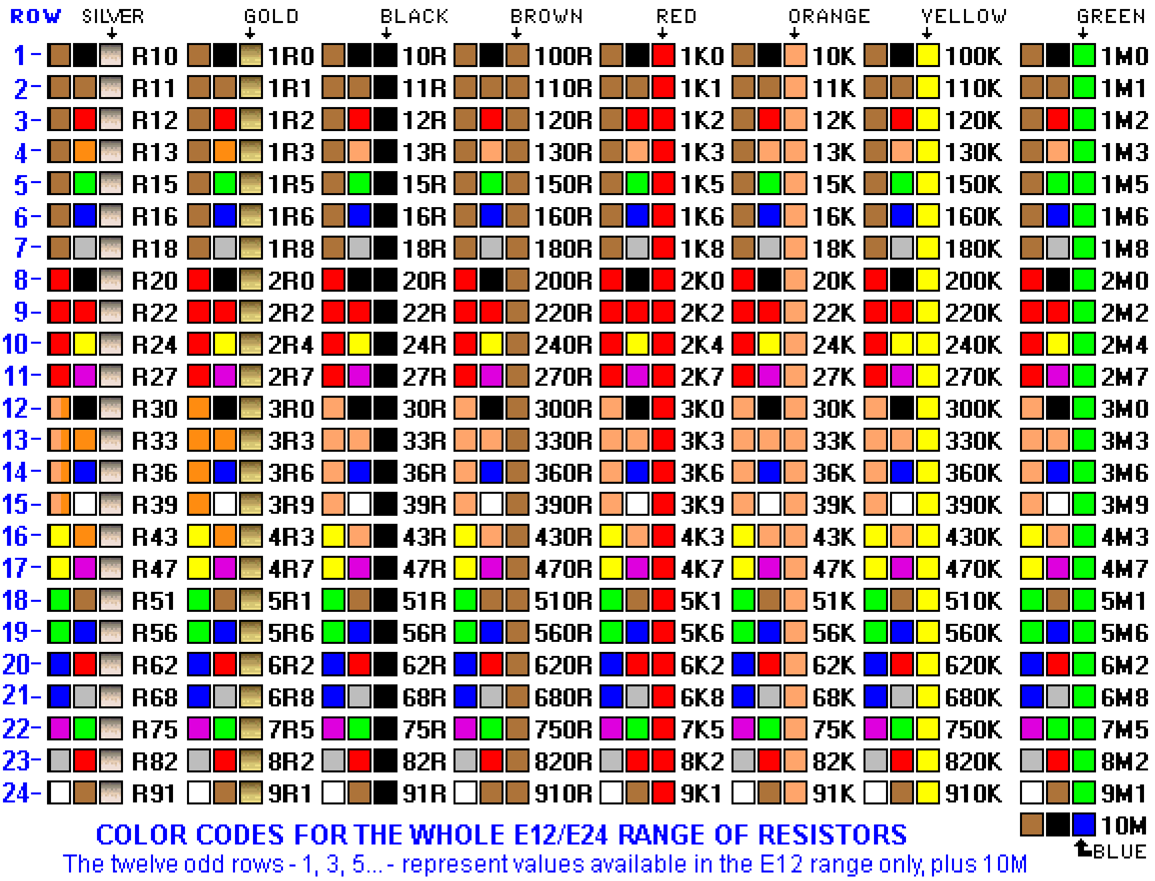

ROW	SILVER	GOLD	BLACK	BROWN	RED	ORANGE	YELLOW	GREEN
1-	R10	1R0	10R	100R	1K0	10K	100K	1M0
2-	R11	1R1	11R	110R	1K1	11K	110K	1M1
3-	R12	1R2	12R	120R	1K2	12K	120K	1M2
4-	R13	1R3	13R	130R	1K3	13K	130K	1M3
5-	R15	1R5	15R	150R	1K5	15K	150K	1M5
6-	R16	1R6	16R	160R	1K6	16K	160K	1M6
7-	R18	1R8	18R	180R	1K8	18K	180K	1M8
8-	R20	2R0	20R	200R	2K0	20K	200K	2M0
9-	R22	2R2	22R	220R	2K2	22K	220K	2M2
10-	R24	2R4	24R	240R	2K4	24K	240K	2M4
11-	R27	2R7	27R	270R	2K7	27K	270K	2M7
12-	R30	3R0	30R	300R	3K0	30K	300K	3M0
13-	R33	3R3	33R	330R	3K3	33K	330K	3M3
14-	R36	3R6	36R	360R	3K6	36K	360K	3M6
15-	R39	3R9	39R	390R	3K9	39K	390K	3M9
16-	R43	4R3	43R	430R	4K3	43K	430K	4M3
17-	R47	4R7	47R	470R	4K7	47K	470K	4M7
18-	R51	5R1	51R	510R	5K1	51K	510K	5M1
19-	R56	5R6	56R	560R	5K6	56K	560K	5M6
20-	R62	6R2	62R	620R	6K2	62K	620K	6M2
21-	R68	6R8	68R	680R	6K8	68K	680K	6M8
22-	R75	7R5	75R	750R	7K5	75K	750K	7M5
23-	R82	8R2	82R	820R	8K2	82K	820K	8M2
24-	R91	9R1	91R	910R	9K1	91K	910K	9M1
								10M BLUE

COLOR CODES FOR THE WHOLE E12/E24 RANGE OF RESISTORS

The twelve odd rows - 1, 3, 5... - represent values available in the E12 range only, plus 10M

Reading 4-band resistors

The most "common" type of resistor has 4 bands and is called the 10% resistor. It now has a tolerance of 5% but is still called the "10% type" as the colours increase by 20% so that a resistor can be 10% higher or 10% lower than a particular value and all the resistors produced in a batch can be used.
The first 3 bands produce the resistance and the fourth band is the "tolerance" band. Gold = 5% (Silver =10% but no modern resistors are 10%!! - they are 5% 2% or 1%)

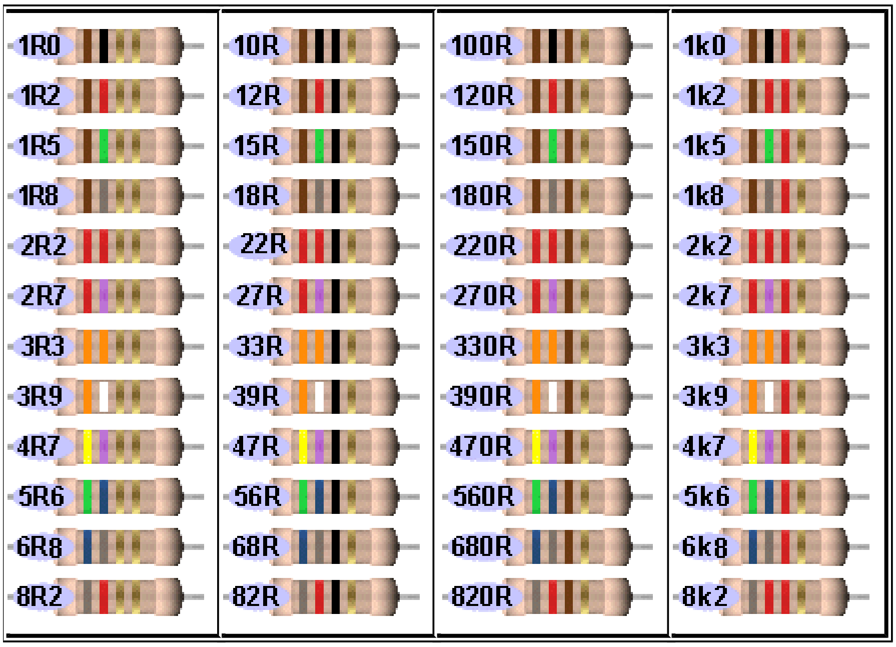
1R0
1R2
1R5
1R8
2R2
2R7
3R3
3R9
4R7
5R6
6R8
8R2
10R
12R
15R
18R
22R
27R
33R
39R
47R
56R
68R
82R
100R
120R
150R
180R
220R
270R
330R
390R
470R
560R
680R
820R
1k0
1k2
1k5
1k8
2k2
2k7
3k3
3k9
4k7
5k6
6k8
8k2

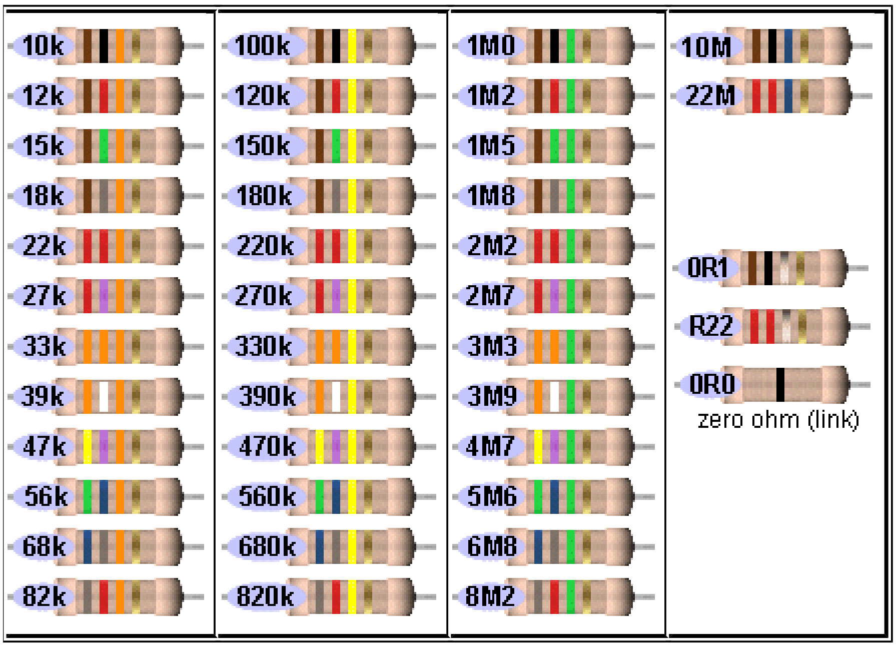
10k
12k
15k
18k
22k
27k
33k
39k
47k
56k
68k
82k
100k
120k
150k
180k
220k
270k
330k
390k
470k
560k
680k
820k
1M0
1M2
1M5
1M8
2M2
2M7
3M3
3M9
4M7
5M6
6M8
8M2
10M
22M
0R1
R22
0R0
zero ohm (link)

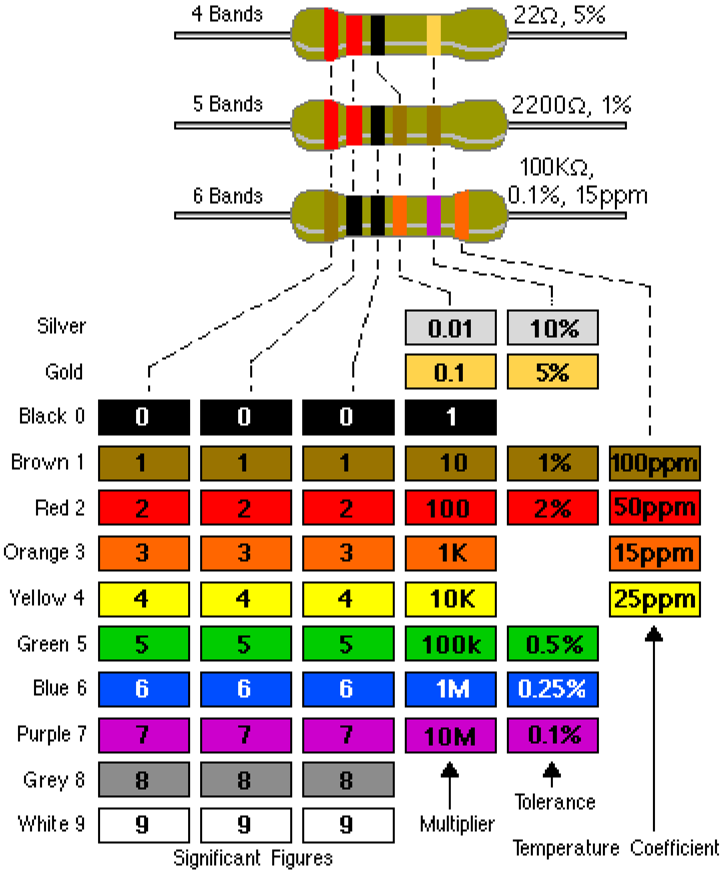
4 Bands
22Ω, 5%
5 Bands
2200Ω, 1%
6 Bands
100KΩ,
0.1%, 15ppm

	Significant Figures			Multiplier	Tolerance	Temperature Coefficient
Silver				0.01	10%	
Gold				0.1	5%	
Black 0	0	0	0	1		
Brown 1	1	1	1	10	1%	100ppm
Red 2	2	2	2	100	2%	50ppm
Orange 3	3	3	3	1K		15ppm
Yellow 4	4	4	4	10K		25ppm
Green 5	5	5	5	100k	0.5%	
Blue 6	6	6	6	1M	0.25%	
Purple 7	7	7	7	10M	0.1%	
Grey 8	8	8	8			
White 9	9	9	9			

Resistor Color Code System

Here is another well-designed resistor colour code chart:

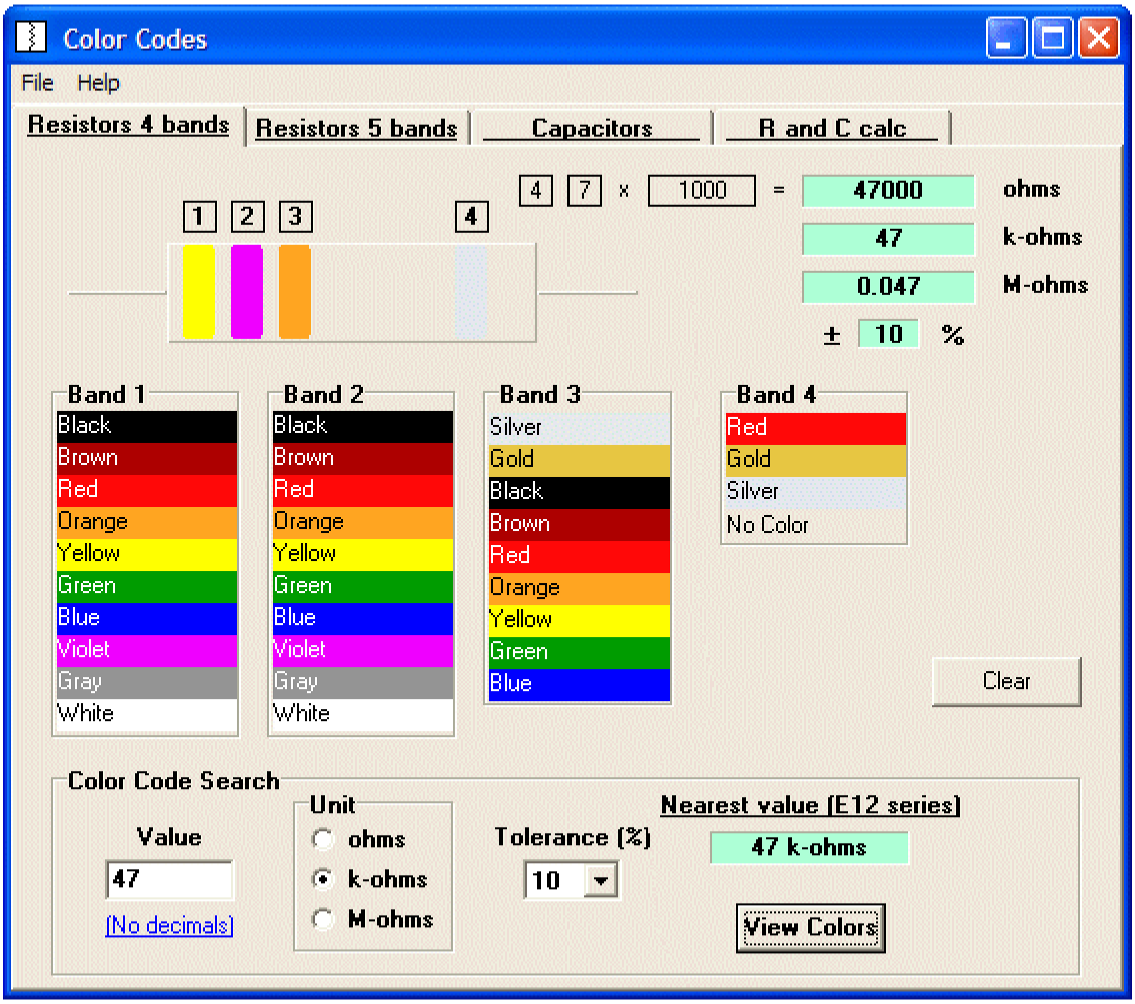
Color Codes
File
Help
Resistors 4 bands
Resistors 5 bands
Capacitors
R and C calc
1
2
3
4
Band 1
Black
Brown
Red
Orange
Yellow
Green
Blue
Violet
Gray
White
Band 2
Black
Brown
Red
Orange
Yellow
Green
Blue
Violet
Gray
White
Band 3
Silver
Gold
Black
Brown
Red
Orange
Yellow
Green
Blue
Band 4
Red
Gold
Silver
No Color
Clear
Color Code Search
(No decimals)
View Colors

4 7 x 1000 = 47000 ohms
47 k-ohms
0.047 M-ohms
± 10 %

Value: 47
Unit: ○ ohms ● k-ohms ○ M-ohms
Tolerance (%): 10
Nearest value (E12 series): 47 k-ohms

Download the program and save it on your desk-top for future reference:
ColourCode.exe (520KB)
ColourCode.zip (230KB)
ColourCode.rar (180KB)

RESISTORS LESS THAN 10 OHMS

When the **third** band is gold, it indicates the value of the "colors" must be divided by 10.
Gold = "divide by 10" to get values 1R0 to 8R2
When the **third** band is silver, it indicates the value of the "colors" must be divided by 100.
(Remember: more letters in the word "silver" thus the divisor is "a larger division.")
Silver = "divide by 100" to get values R1 to R82
e.g: 0R1 = 0.1 ohm   0R22 = point 22 ohms
See 4th Column above for examples.

The letters "R, k and M" take the place of a decimal point.
e.g: 1**R**0 = 1 ohm   2**R**2 = 2 point 2 ohms   22**R** = 22 ohms
2**k**2 = 2,200 ohms   100**k** = 100,000 ohms
2**M**2 = 2,200,000 ohms

HOW TO REMEMBER THE COLOUR CODE:

Each colour has a "number" (or divisor) corresponding to it.
Most of the colours are in the same order as in the spectrum. You can see the spectrum in a rainbow.
It is: ROY G BIV and the colours for resistors are in the same sequence.

black
brown - colour of increasing temperature
red
orange
yellow
green
blue
(indigo - that part of the spectrum between blue and violet)
violet

gray
white

colour	**value**	No of zero's
silver	-2	divide by 100
gold	-1	divide by 10
black	0	No zeros
brown	1	0
red	2	00
orange	3	,000 or k
yellow	4	0,000
green	5	00,000
blue	6	M
violet	7	
gray	8	
white	9	

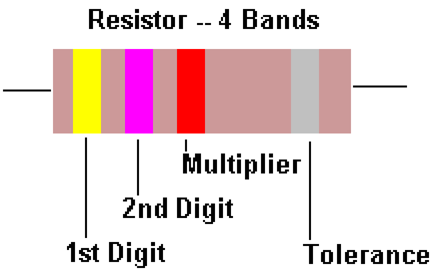
Resistor -- 4 Bands
Multiplier
2nd Digit
1st Digit
Tolerance

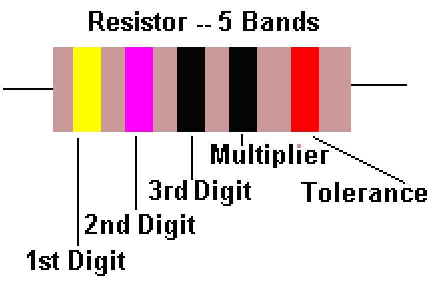
Resistor -- 5 Bands
Multiplier
3rd Digit
Tolerance
2nd Digit
1st Digit

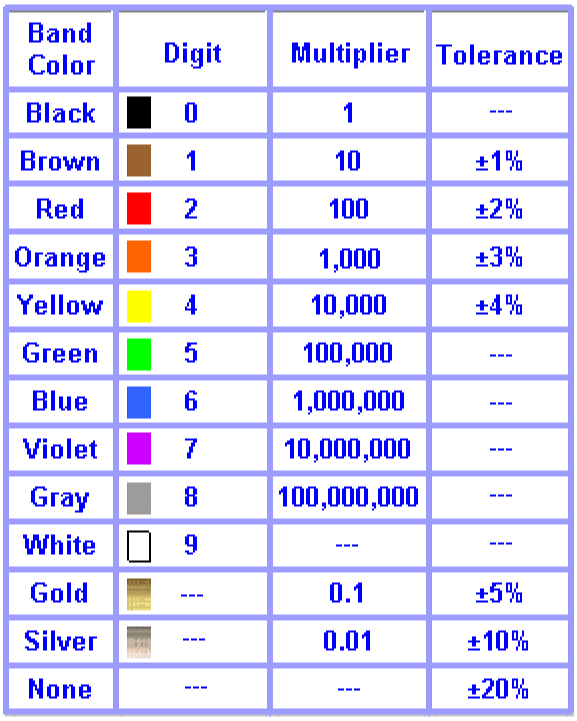

Band Color	Digit	Multiplier	Tolerance
Black	⬛ 0	1	---
Brown	🟫 1	10	±1%
Red	🟥 2	100	±2%
Orange	🟧 3	1,000	±3%
Yellow	🟨 4	10,000	±4%
Green	🟩 5	100,000	---
Blue	🟦 6	1,000,000	---
Violet	🟪 7	10,000,000	---
Gray	▨ 8	100,000,000	---
White	⬜ 9	---	---
Gold	▨ ---	0.1	±5%
Silver	▨ ---	0.01	±10%
None	---	---	±20%

Here are some common ways to remember the colour code:
Bad Beer Rots Our Young Guts, But Vodka Goes Well
Bright Boys Rave Over Young Girls But Violet Gets Wed
Bad Boys Rave Over Young Girls But Violet Gets Wed with Gold and Silver.

Reading 5-band resistors:

5-band resistors are easy to read if you remember two simple points. The first three bands provide the digits in the answer and the 4th band supplies the number of zero's.

Reading "STANDARD VALUES" (on 5-band resistors)

5-band resistors are also made in "Standard Values" but will have different colours to 4-band "common" resistors - and will be confusing if you are just starting out. For instance, a 47k 5% resistor

with 4-bands will be: yellow-purple-orange-gold. For a 47k 1% resistor the colours will be yellow-purple-black-red-brown. The brown colour-band represents 1%.
The first two colour-bands for a STANDARD VALUE or "common value" in 1% or 5% will be the SAME. These two bands provide the digits in the answer.
It's the 3rd band for a 5% resistor that is expanded into two bands in a 1% resistor. But it's easy to follow.
For a standard value, the 3rd band in a 1% resistor is BLACK. This represents a ZERO in the answer. (For 5-band resistors BLACK represents a ZERO when in the third band. This is different to 4-band resistors where black represents the word OHMS! If the third band is BROWN, the answer will be 1). So the 4th band has to represent one-less ZERO and is one colour UP THE COLOUR CHART! In other words the 3rd and 4th bands (combined) on a 1% resistor produces the same number of zero's as the 3rd band on a 5% resistor!

Resistors come in a range of values and the two most common are the E12 and E24 series. The E12 series comes in twelve values for each decade. The E24 series comes in twenty-four values per decade.

E12 series - 10, 12, 15, 18, 22, 27, 33, 39, 47, 56, 68, 82

E24 series - 10, 11, 12, 13, 15, 16, 18, 20, 22, 24, 27, 30, 33, 36, 39, 43, 47, 51, 56, 62, 68, 75, 82, 91

Here is the complete list of 1% 1/4watt resistors from:
CIRCUIT SPECIALISTS. The following list covers 10 ohms (10R) to 1M.
To buy 1% resistors from Circuit Specialists, click: HERE.

10R	121R	806R	3k83	7k15	14k7	39k2	121k
12R1	150R	825R	3k92	7k5	15k	40k2	147k
15R	182R	909R	4k02	7k87	15k8	44k2	150k
18R2	200R	1k0	4k22	71k5	16k9	46k4	182k
22R1	221R	1k21	4k64	8k06	17k4	47k	200k
27R4	240R	1k5	4k75	8k25	17k8	47k5	212k
30R1	249R	1k82	4k7	8k45	18k2	49k9	221k
33R2	274R	2k	4k87	8k66	20k	51k1	226k
36R5	301R	2k21	4k99	8k87	22k1	53k6	249k
39R2	332R	2k2	5k11	9k09	22k6	56k2	274k
47R5	348R	2k43	5k23	9k31	23k7	61k9	301k
49R9	392R	2k49	5k36	9k53	24k9	68k1	332k
51R1	402R	2k67	5k49	9k76	27k4	69k8	357k
56R2	475R	2k74	5k62	10k	29k4	75k0	392k
68R1	499R	3k01	5k76	11k	30k1	82k5	475k
75R	565R	3k32	5k9	12k	33k2	90k	487k
82R5	604R	3k48	6k04	12k1	34k8	90k9	499k
90R9	681R	3k57	6k19	12k4	36k5	95k3	562k
100R	750R	3k74	6k81	13k	38k3	100k	604k
							1M

Here is the list of 1% resistors from suppliers (such as Farnell):

1R0	13R	68R	360R	1k8	9k1	47k	240k
1R2	15R	75R	390R	2k0	10k	51k	270k
1R5	16R	82R	430R	2k2	11k	56k	300k
2R2	18R	91R	470R	2k4	12k	62k	330k
2R7	20R	100R	510R	2k7	13k	68k	360k
3R3	22R	110R	560R	3k	15k	75k	390k
3R9	24R	120R	620R	3k3	16k	82k	430k
4R7	27R	130R	680R	3k6	18k	91k	470k
5R6	30R	150R	750R	3k9	20k	100k	510k
6R2	33R	160R	820R	4k3	22k	110k	560k
6R8	36R	180R	910R	4k7	24k	120k	620k
7R5	39R	200R	1k	5k1	27k	130k	680k
8R2	43R	220R	1k1	5k6	30k	150k	750k
9R1	47R	240R	1k2	6k2	33k	160k	820k
10R	51R	270R	1k3	6k8	36k	180k	910k
11R	56R	300R	1k5	7k5	39k	200k	1M
12R	62R	330R	1k6	8k2	43k	220k	

Surface Mount Resistors

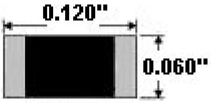

"120 thou"
x
"60 thou"
"one hundred and
twenty thousands
of an inch"

10R
10 ohm

0.047R

0.025R

sense resistor

121k

499R

240R

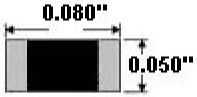

474
470 kΩ

11.8 ohm

10 ohm

M47
470 kΩ

R47	4R7	47R	K47	4K7	47K
0.47 Ω	4.7 Ω	47 Ω	470 Ω	4.7 kΩ	47 kΩ
R464	464R	4K64	471	472	473
0.464 Ω	464 Ω	4.64 kΩ	470 Ω	4.7 kΩ	47 kΩ

0 00 000 0000

SHORT-CIRCUITING "ZERO-OHM LINKS" OR "JUMPERS"

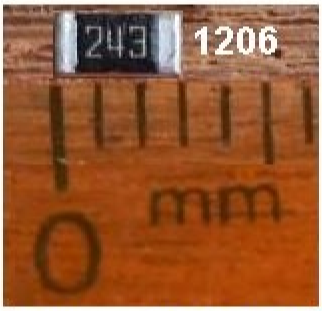
243
1206
0
mm

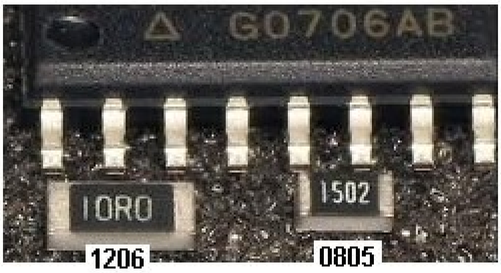
G0706AB
10R0
1502
1206
0805

1206 and 0805 SURFACE MOUNT RESISTORS

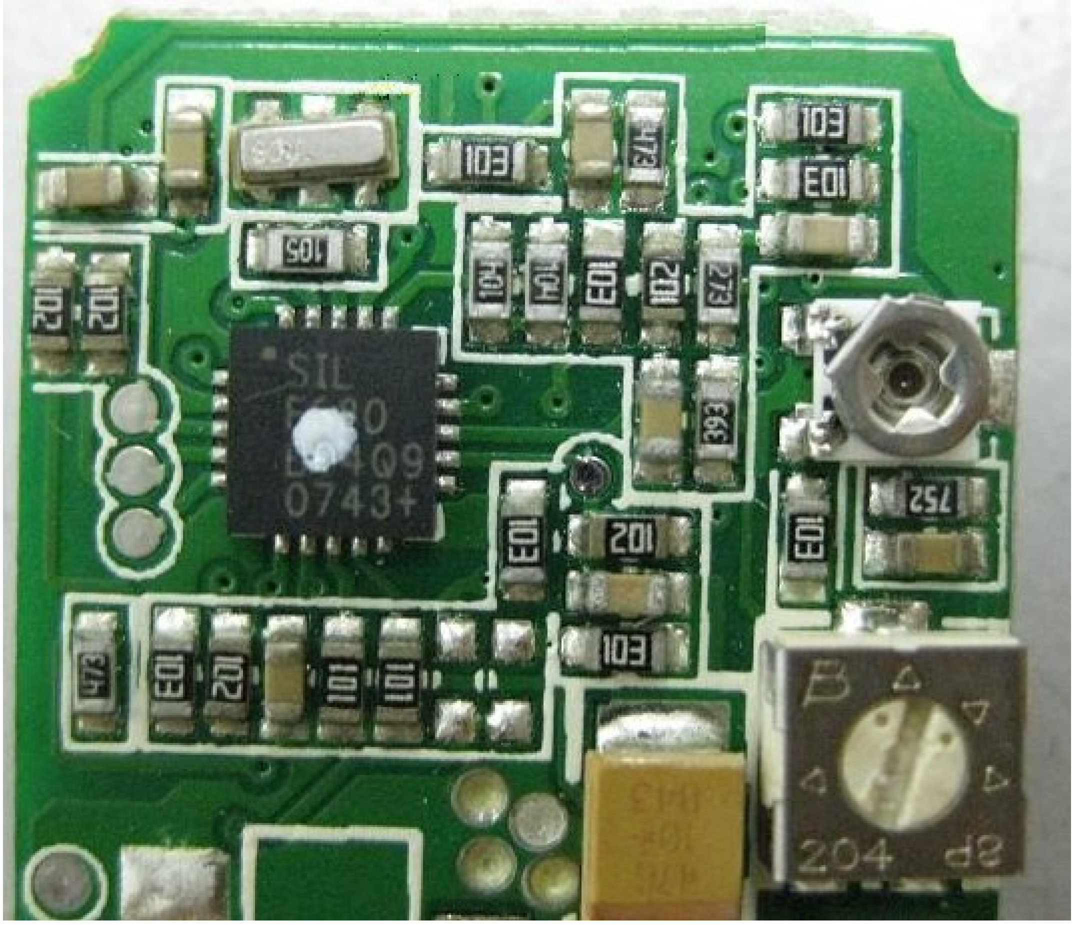

3-digit Surface Mount resistors on a PC board

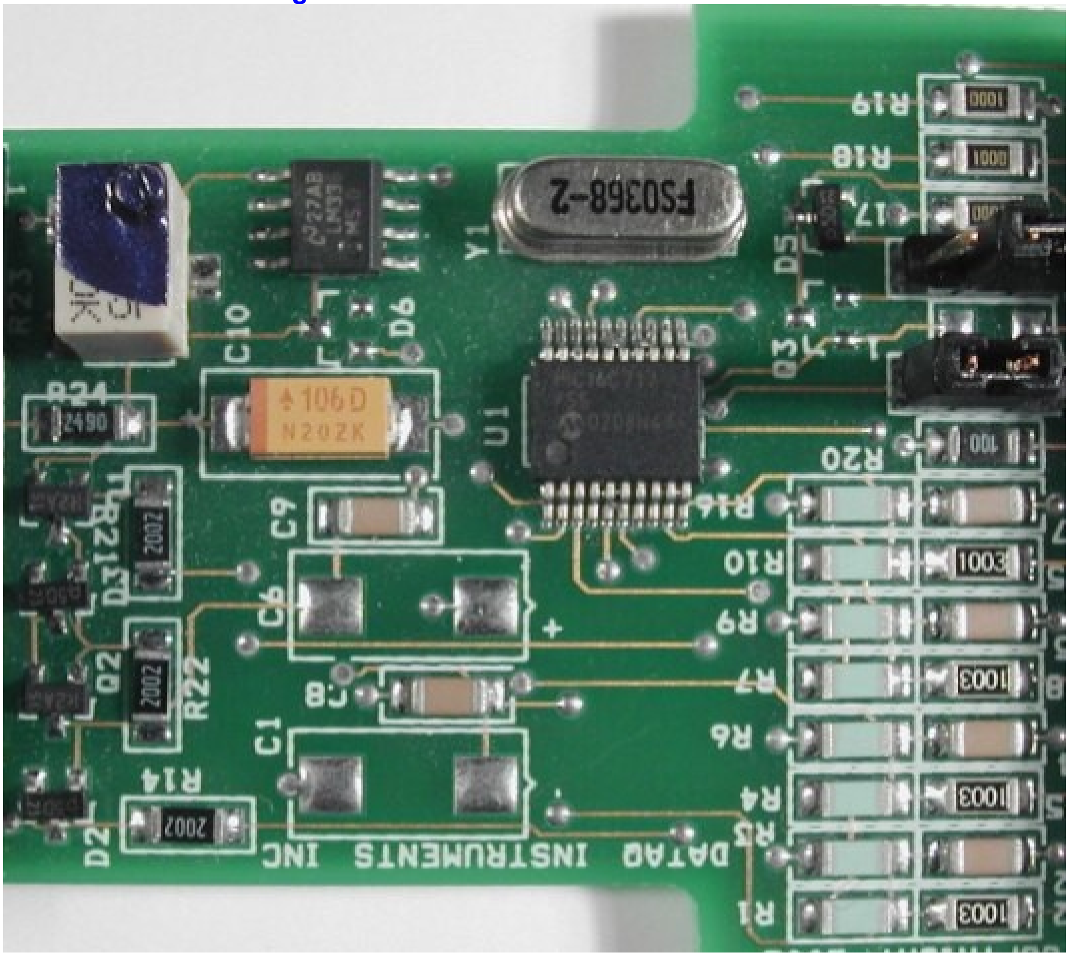

4-digit Surface Mount resistors on a PC board

The photo above shows surface mount resistors on a circuit board. The components that are not marked are capacitors (capacitors are NEVER marked).
All the SM resistors in the above photos conform to a 3-digit or 4-digit code. But there are a number of codes, and the 4-digit code caters for high tolerance resistors, so it's getting very complicated.

Here is a basic 3-digit SM resistor:

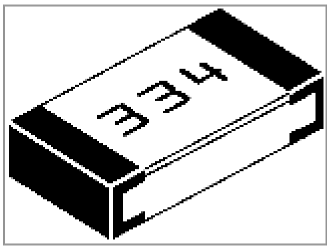

A 330k SM resistor

The first two digits represent the two digits in the answer. The third digit represents the number of zero's you must place after the two digits. The answer will be OHMS. For example: 334 is written 33 0 000. This is written 330,000 ohms. The comma can be replaced by the letter "k". The final answer is: 330k.
222 = 22 00 = 2,200 = 2k2
473 = 47 000 = 47,000 = 47k
474 = 47 0000 = 470,000 = 470k
105 = 10 00000 = 1,000,000 = 1M = one million ohms
There is one trick you have to remember. Resistances less than 100 ohms are written: 100, 220, 470. These are 10 and NO zero's = 10 ohms = 10R
or 22 and no zero's = 22R or 47 and no zero's = 47R. Sometimes the resistor is marked: 10, 22 and 47 to prevent a mistake.

Remember:
R = ohms
k = kilo ohms = 1,000 ohms
M = Meg = 1,000,000 ohms
The 3 letters (R, k and M) are put in place of the decimal point. This way you cannot make a mistake when reading a value of resistance.

THE COMPLETE RANGE OF SM RESISTOR MARKINGS

Click to see the complete range of SM resistor markings for 3-digit code:

Click to see the complete range of SM resistor markings for 4-digit code:

0000 is a value on a surface-mount resistor. It is a zero-ohm **LINK**!
Resistances less than 10 ohms have 'R' to indicate the position of the decimal point.
Here are some examples:

Three Digit Examples	Four Digit Examples
330 is 33 ohms *- not 330 ohms*	**1000 is 100 ohms *- not 1000 ohms***
221 is 220 ohms	**4992 is 49 900 ohms, or 49k9**
683 is 68 000 ohms, or 68k	**1623 is 162 000 ohms, or 162k**
105 is 1 000 000 ohms, or 1M	**0R56 or R56 is 0.56 ohms**
8R2 is 8.2 ohms	

A new coding system has appeared on **1% types**. This is known as the EIA-96 marking method. It consists of a three-character code. The first two digits signify the 3 significant digits of the resistor value, using the lookup table below. The third character - a letter - signifies the multiplier.

code	value	code	value	code	value	code	value	code	value	code	value
01	100	17	147	33	215	49	316	65	464	81	681
02	102	18	150	34	221	50	324	66	475	82	698
03	105	19	154	35	226	51	332	67	487	83	715
04	107	20	158	36	232	52	340	68	499	84	732
05	110	21	162	37	237	53	348	69	511	85	750
06	113	22	165	38	243	54	357	70	523	86	768
07	115	23	169	39	249	55	365	71	536	87	787
08	118	24	174	40	255	56	374	72	549	88	806
09	121	25	178	41	261	57	383	73	562	89	825
10	124	26	182	42	267	58	392	74	576	90	845
11	127	27	187	43	274	59	402	75	590	91	866
12	130	28	191	44	280	60	412	76	604	92	887
13	133	29	196	45	287	61	422	77	619	93	909
14	137	30	200	46	294	62	432	78	634	94	931
15	140	31	205	47	301	63	442	79	649	95	953
16	143	32	210	48	309	64	453	80	665	96	976

The **multiplier** letters are as follows:

letter	mult	letter	mult
F	100000	B	10
E	10000	A	1
D	1000	X or S	0.1
C	100	Y or R	0.01

22A is a 165 ohm resistor, **68C** is a 49900 ohm (49k9) and **43E** a 2740000 (2M74). This marking scheme applies to 1% resistors only.

A similar arrangement can be used for **2% and 5%** tolerance types. The multiplier letters are identical to 1% ones, but occur **before** the number code and the following **code** is used:

2%				5%			
code	value	code	value	code	value	code	value
01	100	13	330	25	100	37	330
02	110	14	360	26	110	38	360
03	120	15	390	27	120	39	390
04	130	16	430	28	130	40	430
05	150	17	470	29	150	41	470
06	160	18	510	30	160	42	510
07	180	19	560	31	180	43	560
08	200	20	620	32	200	44	620
09	220	21	680	33	220	45	680
10	240	22	750	34	240	46	750
11	270	23	820	35	270	47	820
12	300	24	910	36	300	48	910

With this arrangement, **C31** is 5%, 18000 ohm (18k), and **D18** is 510000 ohms (510k) 2% tolerance. Always check with an ohm-meter (a multimeter) to make sure.

Chip resistors come in the following styles and ratings:

Style: 0402, 0603, 0805, 1206, 1210, 2010, 2512, 3616, 4022

Power Rating: 0402(1/16W), 0603(1/10W), 0805(1/8W), 1206(1/4W), 1210(1/3W), 2010(3/4W), 2512(1W), 3616(2W), 4022(3W)

Tolerance: 0.1%, 0.5%, 1%, 5%

Temperature Coefficient: 25ppm 50ppm 100ppm

EIA marking code for surface mount (SMD) resistors

01S = 1R	01R = 10R	01A = 100R	01B = 1k	01C = 10k	01D = 100k	01E = 1M	01F = 10M
02S = 1R02	02R = 10R2	02A = 102R	02B = 1k02	02C = 10k2	02D = 102k	02E = 1M02	
03S = 1R05	03R = 10R5	03A = 105R	03B = 1k05	03C = 10k5	03D = 105k	03E = 1M05	18F = 15M
04S = 1R07	04R = 10R7	04A = 107R	04B = 1k07	04C = 10k7	04D = 107k	04E = 1M07	
05S = 1R1	05R = 11R	05A = 110R	05B = 1k1	05C = 11k	05D = 110k	05E = 1M1	30F = 20M
06S = 1R13	06R = 11R3	06A = 113R	06B = 1k13	06C = 11k3	06D = 113k	06E = 1M13	
07S = 1R15	07R = 11R5	07A = 115R	07B = 1k15	07C = 11k5	07D = 115k	07E = 1M15	
08S = 1R18	08R = 11R8	08A = 118R	08B = 1k18	08C = 11k8	08D = 118k	08E = 1M18	
09S = 1R21	09R = 12R1	09A = 121R	09B = 1k21	09C = 12k1	09D = 121k	09E = 1M21	
10S = 1R24	10R = 12R4	10A = 124R	10B = 1k24	10C = 12k4	10D = 124k	10E = 1M24	
11S = 1R27	11R = 12R7	11A = 127R	11B = 1k27	11C = 12k7	11D = 127k	11E = 1M27	
12S = 1R3	12R = 13R	12A = 130R	12B = 1k3	12C = 13k	12D = 130k	12E = 1M3	
13S = 1R33	13R = 13R3	13A = 133R	13B = 1k33	13C = 13k3	13D = 133k	13E = 1M33	
14S = 1R37	14R = 13R7	14A = 137R	14B = 1k37	14C = 13k7	14D = 137k	14E = 1M37	
15S = 1R4	15R = 14R	15A = 140R	15B = 1k4	15C = 14k	15D = 140k	15E = 1M4	
16S = 1R43	16R = 14R3	16A = 143R	16B = 1k43	16C = 14k3	16D = 143k	16E = 1M43	
17S = 1R47	17R = 14R7	17A = 147R	17B = 1k47	17C = 14k7	17D = 147k	17E = 1M47	
18S = 1R5	18R = 15R	18A = 150R	18B = 1k5	18C = 15k	18D = 15k	18E = 1M5	
19S = 1R54	19R = 15R4	19A = 154R	19B = 1k54	19C = 15k4	19D = 154k	19E = 1M54	
20S = 1R58	20R = 15R8	20A = 158R	20B = 1k58	20C = 15k8	20D = 158k	20E = 1M58	
21S = 1R62	21R = 16R2	21A = 162R	21B = 1k62	21C = 16k2	21D = 162k	21E = 1M62	
22S = 1R65	22R = 16R5	22A = 165R	22B = 1k65	22C = 16k5	22D = 165k	22E = 1M65	
23S = 1R69	23R = 16R9	23A = 169R	23B = 1k69	23C = 16k9	23D = 169k	23E = 1M69	
24S = 1R74	24R = 17R4	24A = 174R	24B = 1k74	24C = 17k4	24D = 174k	24E = 1M74	
25S = 1R78	25R = 17R8	25A = 178R	25B = 1k78	25C = 17k8	25D = 178k	25E = 1M78	
26S = 1R82	26R = 18R2	26A = 182R	26B = 1k82	26C = 18k2	26D = 182k	26E = 1M82	
27S = 1R87	27R = 18R7	27A = 187R	27B = 1k87	27C = 18k7	27D = 187k	27E = 1M87	
28S = 1R91	28R = 19R1	28A = 191R	28B = 1k91	28C = 19k1	28D = 191k	28E = 1M91	
29S = 1R96	29R = 19R6	29A = 196R	29B = 1k96	29C = 19k6	29D = 196k	29E = 1M96	
30S = 2R0	30R = 20R0	30A = 200R	30B = 2k0	30C = 20k0	30D = 200k	30E = 2M0	
31S = 2R05	31R = 20R5	31A = 205R	31B = 2k05	31C = 20k5	31D = 205k	31E = 2M05	
32S = 2R10	32R = 21R0	32A = 210R	32B = 2k10	32C = 21k0	32D = 210k	32E = 2M10	
33S = 2R15	33R = 21R5	33A = 215R	33B = 2k15	33C = 21k5	33D = 215k	33E = 2M15	
34S = 2R21	34R = 22R1	34A = 221R	34B = 2k21	34C = 22k1	34D = 221k	34E = 2M21	
35S = 2R26	35R = 22R6	35A = 226R	35B = 2k26	35C = 22k6	35D = 226k	35E = 2M26	
36S = 2R32	36R = 23R2	36A = 232R	36B = 2k32	36C = 23k2	36D = 232k	36E = 2M32	
37S = 2R37	37R = 23R7	37A = 237R	37B = 2k37	37C = 23k7	37D = 237k	37E = 2M37	
38S = 2R43	38R = 24R3	38A = 243R	38B = 2k43	38C = 24k3	38D = 243k	38E = 2M43	
39S = 2R49	39R = 24R9	39A = 249R	39B = 2k49	39C = 24k9	39D = 249k	39E = 2M49	
40S = 2R55	40R = 25R5	40A = 255R	40B = 2k55	40C = 25k5	40D = 255k	40E = 2M55	
41S = 2R61	41R = 26R1	41A = 261R	41B = 2k61	41C = 26k1	41D = 261k	41E = 2M61	
42S = 2R67	42R = 26R7	42A = 267R	42B = 2k67	42C = 26k7	42D = 267k	42E = 2M67	
43S = 2R74	43R = 27R4	43A = 274R	43B = 2k74	43C = 27k4	43D = 274k	43E = 2M74	
44S = 2R80	44R = 28R0	44A = 280R	44B = 2k80	44C = 28k0	44D = 280k	44E = 2M80	
45S = 2R87	45R = 28R7	45A = 287R	45B = 2k87	45C = 28k7	45D = 287k	45E = 2M87	
46S = 2R94	46R = 29R4	46A = 294R	46B = 2k94	46C = 29k4	46D = 294k	46E = 2M94	
47S = 3R01	47R = 30R1	47A = 301R	47B = 3k01	47C = 30k1	47D = 301k	47E = 3M01	
48S = 3R09	48R = 30R9	48A = 309R	48B = 3k09	48C = 30k9	48D = 309k	48E = 3M09	
49S = 3R16	49R = 31R6	49A = 316R	49B = 3k16	49C = 31k6	49D = 316k	49E = 3M16	
50S = 3R24	50R = 32R4	50A = 324R	50B = 3k24	50C = 32k4	50D = 324k	50E = 3M24	
51S = 3R32	51R = 33R2	51A = 332R	51B = 3k32	51C = 33k2	51D = 332k	51E = 3M32	
52S = 3R4	52R = 34R0	52A = 340R	52B = 3k4	52C = 34k0	52D = 340k	52E = 3M4	
53S = 3R48	53R = 34R8	53A = 348R	53B = 3k48	53C = 34k8	53D = 348k	53E = 3M48	
54S = 3R57	54R = 35R7	54A = 357R	54B = 3k57	54C = 35k7	54D = 357k	54E = 3M57	
55S = 3R65	55R = 36R5	55A = 365R	55B = 3k65	55C = 36k5	55D = 365k	55E = 3M65	
56S = 3R74	56R = 37R4	56A = 374R	56B = 3k74	56C = 37k4	56D = 374k	56E = 3M74	
57S = 3R83	57R = 38R3	57A = 383R	57B = 3k83	57C = 38k3	57D = 383k	57E = 3M83	
58S = 3R92	58R = 39R2	58A = 392R	58B = 3k92	58C = 39k2	58D = 392k	58E = 3M92	
59S = 4R02	59R = 40R2	59A = 402R	59B = 4k02	59C = 40k2	59D = 402k	59E = 4M02	
60S = 4R12	60R = 41R2	60A = 412R	60B = 4k12	60C = 41k2	60D = 412k	60E = 4M12	
61S = 4R22	61R = 42R2	61A = 422R	61B = 4k22	61C = 42k2	61D = 422k	61E = 4M22	
62S = 4R32	62R = 43R2	62A = 432R	62B = 4k32	62C = 43k2	62D = 432k	62E = 4M32	
63S = 4R42	63R = 44R2	63A = 442R	63B = 4k42	63C = 44k2	63D = 442k	63E = 4M42	
64S = 4R53	64R = 45R3	64A = 453R	64B = 4k53	64C = 45k3	64D = 453k	64E = 4M53	
65S = 4R64	65R = 46R4	65A = 464R	65B = 4k64	65C = 46k4	65D = 464k	65E = 4M64	
66S = 4R75	66R = 47R5	66A = 475R	66B = 4k75	66C = 47k5	66D = 475k	66E = 4M75	
67S = 4R87	67R = 48R7	67A = 487R	67B = 4k87	67C = 48k7	67D = 487k	67E = 4M87	
68S = 4R99	68R = 49R9	68A = 499R	68B = 4k99	68C = 49k9	68D = 499k	68E = 4M99	
69S = 5R11	69R = 51R1	69A = 511R	69B = 5k11	69C = 51k1	69D = 511k	69E = 5M11	
70S = 5R23	70R = 52R3	70A = 523R	70B = 5k23	70C = 52k3	70D = 523k	70E = 5M23	

71S = 5R36	71R = 53R6	71A = 536R	71B = 5k36	71C = 53k6	71D = 536k	71E = 5M36
72S = 5R49	72R = 54R9	72A = 549R	72B = 5k49	72C = 54k9	72D = 549k	72E = 5M49
73S = 5R62	73R = 56R2	73A = 562R	73B = 5k62	73C = 56k2	73D = 562k	73E = 5M62
74S = 5R76	74R = 57R6	74A = 576R	74B = 5k76	74C = 57k6	74D = 576k	74E = 5M76
75S = 5R9	75R = 59R0	75A = 590R	75B = 5k9	75C = 59k0	75D = 590k	75E = 5M9
76S = 6R04	76R = 60R4	76A = 604R	76B = 6k04	76C = 60k4	76D = 604k	76E = 6M04
77S = 6R19	77R = 61R9	77A = 619R	77B = 6k19	77C = 61k9	77D = 619k	77E = 6M19
78S = 6R34	78R = 63R4	78A = 634R	78B = 6k34	78C = 63k4	78D = 634k	78E = 6M34
79S = 6R49	79R = 64R9	79A = 649R	79B = 6k49	79C = 64k9	79D = 649k	79E = 6M49
80S = 6R65	80R = 66R5	80A = 665R	80B = 6k65	80C = 66k5	80D = 665k	80E = 6M65
81S = 6R81	81R = 68R1	81A = 681R	81B = 6k81	81C = 68k1	81D = 681k	81E = 6M81
82S = 6R98	82R = 69R8	82A = 698R	82B = 6k98	82C = 69k8	82D = 698k	82E = 6M98
83S = 7R15	83R = 71R5	83A = 715R	83B = 7k15	83C = 71k5	83D = 715k	83E = 7M15
84S = 7R32	84R = 73R2	84A = 732R	84B = 7k32	84C = 73k2	84D = 732k	84E = 7M32
85S = 7R5	85R = 75R0	85A = 750R	85B = 7k5	85C = 75k0	85D = 750k	85E = 7M5
86S = 7R68	86R = 76R8	86A = 768R	86B = 7k68	86C = 76k8	86D = 768k	86E = 7M68
87S = 7R87	87R = 78R7	87A = 787R	87B = 7k87	87C = 78k7	87D = 787k	87E = 7M87
88S = 8R06	88R = 80R6	88A = 806R	88B = 8k06	88C = 80k6	88D = 806k	88E = 8M06
89S = 8R25	89R = 82R5	89A = 825R	89B = 8k25	89C = 82k5	89D = 825k	89E = 8M25
90S = 8R45	90R = 84R5	90A = 845R	90B = 8k45	90C = 84k5	90D = 845k	90E = 8M45
91S = 8R66	91R = 86R6	91A = 866R	91B = 8k66	91C = 86k6	91D = 866k	91E = 8M66
92S = 8R87	92R = 88R7	92A = 887R	92B = 8k87	92C = 88k7	92D = 887k	92E = 8M87
93S = 9R09	93R = 90R9	93A = 909R	93B = 9k09	93C = 90k9	93D = 909k	93E = 9M09
94S = 9R31	94R = 93R1	94A = 931R	94B = 9k31	94C = 93k1	94D = 931k	94E = 9M31
95S = 9R53	95R = 95R3	95A = 953R	95B = 9k53	95C = 95k3	95D = 953k	95E = 9M53
96S = 9R76	96R = 97R6	96A = 976R	96B = 9k76	96C = 97k6	96D = 976k	96E = 9M76

If you want an accurate RESISTANCE measurement, remove the resistor from the circuit and use a Digital meter.

SURFACE MOUNT COMPONENTS - PACKS

Talking Electronics has packs of components for the repairman. The following packs are available:

SURFACE MOUNT RESISTOR PACK consists of 1 off each standard value
10 ohms to 1M & 2M2 (60 resistors)
$14.20   including pack and post

SURFACE MOUNT CAPACITOR PACK consists of:
2 - 10p   5 - 47p   5 - 100p   5 - 470p   5 - 1n   5 - 10n   5 - 22n   5 - 100n
5 - 1u 16v electrolytic   5 - 10u 16v electrolytic
(40 components)
$23.80 including pack and post

SURFACE MOUNT DIODE PACK consists of:
5 - 1N 4148 (marked as "A6")
$10.00   including pack and post

SURFACE MOUNT TRANSISTOR PACK consists of:
5 - BC 848 (marked as "1K") NPN
5 - BC858 PNP
$10.00   including pack and post

email **Colin Mitchell** for details on how to pay by credit card or PayPal.

CREATING ANY VALUE OF RESISTANCE

Any value of resistance can be created by connecting two resistors in PARALLEL or SERIES.
You can also create a higher wattage resistor by connecting them in SERIES OR PARALLEL.
We are only going to cover two EQUAL VALUE resistors in SERIES or in PARALLEL.
If you want to create a "Special Value," simply connect two resistors and read the value with a Digital Meter. Keep changing the values until you get the required value. We are not going into series or Parallel formulae. You can easily find a value with a multimeter.

TWO EQUAL-VALUE RESISTORS IN SERIES

Two equal-value resistors IN SERIES creates a value of DOUBLE. You simply ADD the values.
This can be done with any to two values as shown. Three equal-value resistors in series is three times the value.

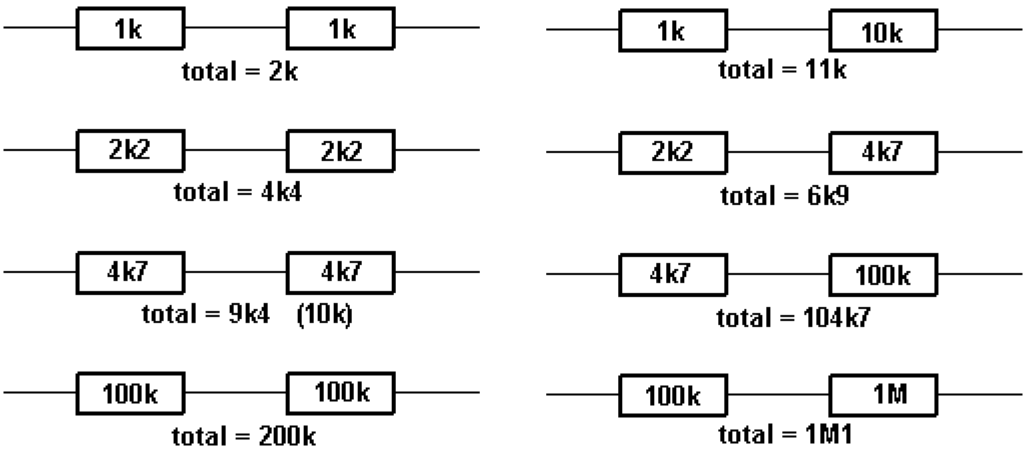
1k
1k
total = 2k
1k
10k
total = 11k
2k2
2k2
total = 4k4
2k2
4k7
total = 6k9
4k7
4k7
total = 9k4 (10k)
4k7
100k
total = 104k7
100k
100k
total = 200k
100k
1M
total = 1M1

TWO EQUAL-VALUE RESISTORS IN PARALLEL

Two equal-value resistors IN PARALLEL creates a value of HALF. Three equal-value resistors in parallel is equal to one-third the value.

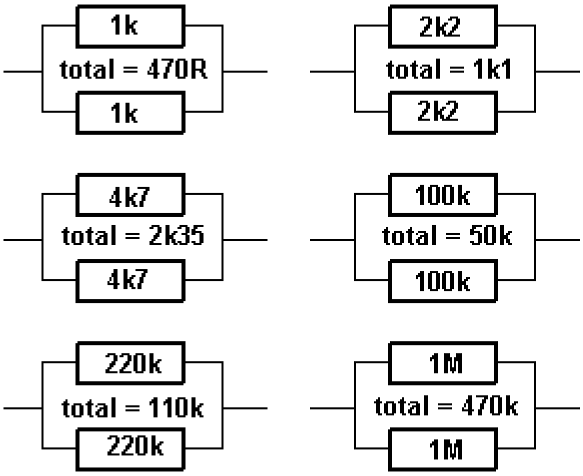
1k
total = 470R
1k
2k2
total = 1k1
2k2
4k7
total = 2k35
4k7
100k
total = 50k
100k
220k
total = 110k
220k
1M
total = 470k
1M

If you want a particular value and it is not available, here is a chart.
Use 2 resistors in series or parallel as shown:

Required Value	R1	Series/ Parallel	R2	Actual value:
10	4R7	S	4R7	9R4
12	10	S	2R2	12R2
15	22	P	47	14R9
18	22	P	100	18R
22	10	S	12	22
27	22	S	4R7	26R7
33	22	S	10	32R
39	220	P	47	38R7
47	22	S	27	49
56	47	S	10	57
68	33	S	33	66
82	27	S	56	83

There are other ways to combine 2 resistors in parallel or series to get a particular value. The examples above are just one way. 4R7 = 4.7 ohms

TESTING A RESISTOR

To check the value of a resistor, it should be removed from the circuit. The surrounding components can affect the

reading and make it lower.
Resistors **VERY RARELY** change value, but if it is overheated or damaged, the resistance can increase. You can take the reading of a resistor "in-circuit" in one direction then the other, as the surrounding components may have diodes and this will alter the reading.
You can also test a resistor by feeling its temperature-rise. It is getting too hot if you cannot hold your finger on it (some "metal film" resistors are designed to tolerate quite high temperatures).

TESTING AN "AC" RESISTOR

There is no such thing as an "AC" resistor. Resistors are just "resistors" and they can be in AC circuits or DC circuits. Resistors can be given names such as "Safety Resistor" "Ballast Resistor" "LOAD Resistor" "Feed Resistor" "Dropper Resistor" or "Supply Resistor." These are just normal resistors with a normal resistance - except a "Safety Resistor."
A safety resistor is made of a flame-proof material such as metal-oxide-film and not carbon-composition. It is designed to "burn out" when too much current flows BUT NOT CATCH FIRE.
It is a low-value resistor and has a voltage-drop across it but this is not intentional. The voltage-drop is to create a "heating-effect" to burn out the resistor. In all the other types of resistor, the voltage-drop is intentional.
A Ballast resistor is a normal resistor and can be called a Power resistor, Dropper resistor, Supply resistor or Feed resistor. It is designed to reduce the voltage from one source and deliver a lower voltage. It is a form of: "in-line" resistor.
A Load Resistor is generally connected across the output of a circuit and turns the energy it receives, into heat.

RESISTOR NETWORKS

To reduce the number of resistors in a circuit, some engineers use a set of identical resistors in a package called a Single-In-Line (SIL) resistor network. It is made with many resistors of the same value, all in one package. One end of each resistor is connected all the other resistors and this is the common pin, identified as pin 1 and has a dot on the package.
These packages are very reliable but to make sure all the resistors are as stated, you need to locate pin 1. All values will be identical when referenced to this pin.

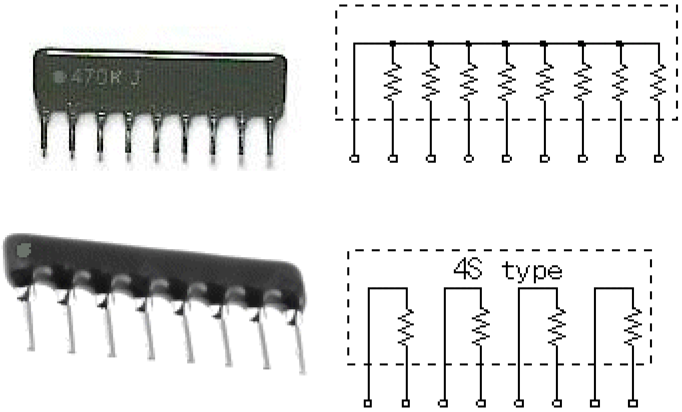

RESISTOR NETWORKS

Some resistor networks have a "4S" printed on the component. The 4S indicates the package contains 4 independent resistors that are not wired together inside. The housing has eight leads as shown in the second image.
Independent resistors have an even number of pins and measuring between each pair will produce identical values. Resistance between any pair will indicate leakage and may be a fault.

TESTING A POSISTOR

A Posistor is a resistor that connects in series with the degaussing coil around the picture tube or **Monitor.** When cold, it has a very low resistance and a large current flows when the monitor or TV is switched on. This current heats up the Posistor and the resistance increases. This causes the current to decrease and any magnetism in the shadow mask is removed. The posistor can one or two elements and it is kept warm so the resistance remains high. Many Posistors have a second element inside the case that connects directly to the supply to keep the Positive Temperature Coefficient resistor high so that the current through the degaussing coil falls to almost zero. This constant heat eventually destroys the package.
The heavy current that flows when a set is turned ON also causes the posistor to crack and break and this results in poor purity on the screen - as the shadow mask gradually becomes magnetic..
Posistors have different resistance values from different manufacturers and must be replaced with an identical type. They can be checked for very low resistance when cold but any loose pieces inside the case will indicate a damaged component.

A "BURNT" RESISTOR **- normally and technically called a "burnt-out" resistor.**

The resistance of a "burnt" resistor can sometimes be determined by scraping away the outer coating - if the resistor has a spiral of resistance-material. You may be able to find a spot where the spiral has been damaged.

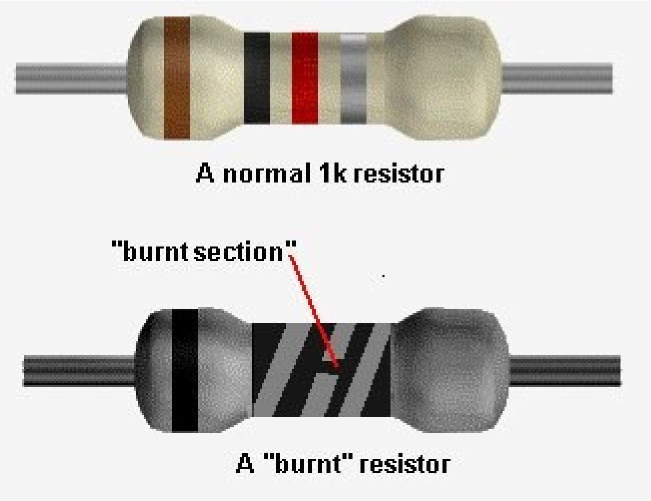

Clean the "spot" (burnt section of the spiral) very carefully and make sure you can get a good contact with the spiral and the tip of your probe. Measure from one lead of the resistor to the end of the damaged spiral. Then measure from the other lead to the other end of the spiral.
Add the two values and you have an approximate value for the resistor. You can add a small amount for the damaged section.
This process works very well for damaged wire-wound resistors. They can be pulled apart and each section of the resistance-wire (nichrome wire) measured and added to get the full resistance.

There is another way to determine the value of a damaged resistor.
Get a set of resistors of the same wattage as the damaged component and start with a high value. It's handy to know if the resistor is in the range: 10ohm to 100ohms or 1k to 10k etc, but this is not essential.
Start with a very high value and turn the circuit ON. You can perform voltage tests and if you know the expected output voltage, decrease the resistance until this voltage is obtained.
If you do not know the expected voltage, keep reducing the value of resistance until the circuit works

as designed.
This is the best advice in a situation where you do not know the value of a resistor.

There is a third way to determine the value and this requires measuring the voltage drop across the resistor and the current-flow. By multiplying the two you will get a wattage and this must be less than the wattage of the resistor being replaced.

TESTING POTENTIOMETERS (variable resistors)

To check the value of a variable resistor, it should be removed from circuit or at least 2 legs should be removed. A Rheostat is a variable resistor using only one end and the middle connected to a circuit.
The resistance between the two outside pins is the value marked on the component and the centre leg will change from nearly zero to the full resistance as the shaft is rotated.
"Pots" generally suffer from "crackle" when turned and this can be fixed by spraying up the shaft and into the pot via the shaft with a tube fixed to a can of "spray-lubricant" (contact cleaner).
"Pre-set pots" and "trim pots" are miniature versions of a potentiometer and they are all tested the same.

FOCUS POTS

Focus pots quite often get a spot of dirt where the wiper touches the track. Cleaning with spray fixes the bad focus but if the pot is leaking to chassis from inside the pot (due to the high voltage on the terminals) simply remove it from the chassis and leave it floating (this will restore the high voltage to the picture tube) or you can use one from an old chassis.

MAKING YOUR OWN RESISTOR, CAPACITOR, INDUCTOR or DIODE

Quite often you will not have the exact value of resistance or capacitance for a repair.
We have already covered placing resistors and capacitors in parallel and series:

Resistors in Parallel and/or Series
Capacitors in Parallel and/or Series

Here are some extras:

RESISTORS
Two 1k 0.5watt resistors in parallel produces a 470R 1watt resistor.
Two 1k 0.5watt resistors in series produces a 2k 1watt resistor.

CAPACITORS
Two 100n 100v capacitors in series produces a 50n capacitor @200v

INDUCTORS: Two inductors in series - **ADD THE VALUES**

DIODES: Two 1Amp 400v diodes in series produces a 1Amp 800v diode
Two 1Amp 400v diodes in parallel produces a 2Amp 400v diode

ZENER DIODES: Zener diodes can be connected in series to get a higher voltage.
Two 12v zener diodes in series produces a 24v zener.

CONTINUITY

Some multimeters have a "buzzer" that detects when the probes are touching each other or the resistance between the probes is very LOW. This is called a CONTINUITY TESTER.
You can use the resistance scale "x1" or "x10" to detect low values of resistance.
Set the pointer to "0" (right end of the scale) by touching the probes together and adjusting the "zero ohms" control.
When taking a reading, you will have to decide if a low value of resistance is a short-circuit or an "operating value."
For instance, the cold resistance of a 12v car globe is very low (about 2 ohms) and it increases (about 6 times) to 12 ohms when hot.
The "resistance of a circuit" may be very low as the electrolytics in the circuit are uncharged. This may not indicate a true "short-circuit."
The measurement across a diode is not a resistance-value but a "voltage-drop" and that is why the needle swings nearly full-scale.
Leads and wires and cords have a small resistance and depending on the length of the lead, this small

resistance may be affecting a circuit.
Remember this:
When a circuit takes 1 amp, and the resistance of the leads is 1 ohm, the voltage drop across the leads will be 1v.
That's why a 12v battery supplying a circuit with these leads will have 11v at the circuit.
Note:
Turn off the equipment before making any continuity tests. The presence of even a small voltage (from an electrolytic) can give a false reading.
You can determine the resistance of a lead very accurately by taking the example above and applying it to your circuit.
If the battery is 12.6v and the voltage across the circuit is 10v, when the current is 2.6 amps, the resistance of the "leads" is 12.6 - 10 = 2.6 R=V/I = 2.6/2.6 = 1ohm. By making the lead shorter or using thicker wire, the resistance will be less and the voltage on the project will increase.
When taking readings in a circuit that has a number of diodes built-into IC's (Integrated Circuits) and transistors, some Continuity Testers will beep and give a false reading.
The following circuit has the advantage of providing a beep when a short-circuit is detected but does not detect the small voltage drop across a diode. This is ideal when testing logic circuits as it is quick and you can listen for the beep while concentrating on the probe. Using a multimeter is much slower.

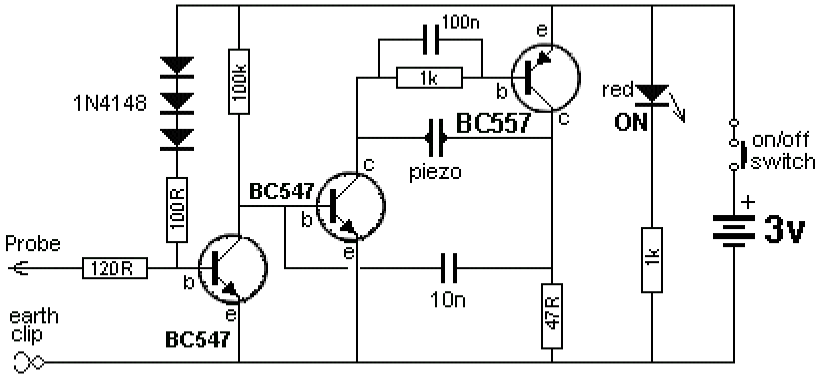

CONTINUITY TESTER

You can build the circuit on Matrix Board and add it to your Test Equipment.
You will need lots of "Test Equipment" and they can be built from circuits in this eBook.

TESTING FUSES, LEADS AND WIRES

All these components come under the heading TESTING for CONTINUITY. Turn off all power to the equipment before testing for shorts and continuity. Use the low resistance "Ohms Scale" or CONTINUITY range on your multimeter. All fuses, leads and wires should have a low, very low or zero resistance. This proves they are working.

A BLOWN FUSE

The appearance of a fuse after it has "blown" can tell you a lot about the fault in the circuit.
If the inside of the glass tube (of the fuse) is totally blackened, the fuse has been damaged very quickly. This indicates a very high current has passed through the fuse.
Depending on the rating of the fuse, (current rating) you will be able to look for components that can pass a high current when damaged - such as high power transistors, FETs, coils, electrolytics. Before re-connecting the supply, you should test the "SUPPLY RAILS" for resistance. This is done by measuring them on a low OHMs range in one direction then reverse the leads to see if the resistance is low in the other direction.
A reading can be very low at the start because electrolytics need time to charge-up and if the reading gradually increases, the power rail does not have a short. An overload can occur when the supply voltage rises to nearly full voltage, so you sometimes have to fit a fuse and see how long it takes to "blow."
If the fuse is just slightly damaged, you will need to read the next part of this eBook, to see how and why this happens:

FAST AND SLOW BLOW FUSES

There are many different sizes, shapes and ratings of a fuse. They are all current ratings as a fuse does not have a voltage rating. Some fuses are designed for cars as they fit into the special fuse holders. A fuse can be designed for 50mA, 100mA, 250mA, 315mA, 500mA, 1Amp, 1.5amp, 2amp, 3amp, 3.15amp 5amp, 10amp, 15amp, 20amp, 25amp, 30amp, 35amp, 50amp and higher.
Some fuses are fast-blow and some are slow-blow.

A "normal" fuse consists of a length of thin wire. Or it may be a loop of wire that is thin near the middle of the fuse. This is the section that will "burn-out."
A "normal" fuse is a fast-blow fuse. For instance, a 1amp fuse will remain intact when up to 1.25 amp flows. When a circuit is turned on, it may take 2-3 amps for a very short period of time and a normal 1 amp fuse will get very hot and the wire will stretch but not "burn-out." You can see the wire move when the supply turns on.
If the current increases to 2amps, the fuse will still remain intact. It needs about 3 amp to heat up the wire to red-hot and burn out.
If the current increases to 5 amp, the wire VOLATILISES (burns-out) and deposits carbon-black on the inside of the glass tube.
A slow-blow fuse uses a slightly thicker piece of wire and the fuse is made of two pieces of wire joined in the middle with a dob of low-temperature solder. Sometimes one of the pieces of wire is a spring and when the current rises to 2.5 amp, the heat generated in the wire melts the solder and the two pieces of wire "spring apart."
A slow-blow fuse will allow a higher current-surge to pass through the fuse and the wire will not heat up and sag.
Thus the fuse is not gradually being damaged and it will remain in a perfect state for a long period of time.
A fuse does not protect electronic equipment from failing. It acts AFTER the equipment has failed.
It will then protect a power supply from delivering a high current to a circuit that has failed.
If a slow-blow fuse has melted the solder, it could be due to a slight overload, slight weakening of the fuse over a period of time or the current-rating may be too low.
You can try another fuse to see what happens.
You can replace a fast-acting fuse (normal fuse) with a slow blow if the fast-acting fuse has been replaced a few times due to deterioration when the equipment is turned on.
But you cannot replace a slow-blow fuse with a fast acting fuse as it will be damaged slightly each time the equipment is turned on and eventually fail.

TESTING COILS, INDUCTORS and YOKES

Coils inductors and yokes are just an extension of a length of wire. The wire may be wrapped around a core made of iron or ferrite.
It is labeled "L" on a circuit board.
You can test this component for continuity between the ends of the winding and also make sure there is no continuity between the winding and the core.
The winding can be less than one ohm, or greater than 100 ohms, however a coil of wire is also called an INDUCTOR and it might look like a very simple component, but it can operate in a very complex way.
The way it works is a discussion for another eBook. It is important to understand the turns are insulated but a slight fracture in the insulation can cause two turns to touch each other and this is called a "SHORTED TURN" or you can say the inductor has "SHORTED TURNS."
When this happens, the inductor allows the circuit to draw MORE CURRENT. This causes the fuse to "blow."
The quickest way to check an inductor is to replace it, but if you want to measure the inductance, you can use an INDUCTANCE METER. You can then compare the inductance with a known good component.

An inductor with a shorted turn will have a very low or zero inductance, however you may not be able to detect the fault when it is not working in a circuit as the fault may be created by a high voltage generated between two of the turns.
Faulty yokes (both horizontal and vertical windings) can cause the picture to reduce in size and/or bend or produce a single horizontal line.
A TV or monitor screen is the best piece of Test Equipment as it has identified the fault. It is pointless trying to test the windings further as you will not be able to test them under full operating conditions.

MEASURING AND TESTING INDUCTORS

Inductors are measured with an INDUCTANCE METER but the value of some inductors is very small and some Inductance Meters do not give an accurate reading.
The solution is to measure a larger inductor and note the reading. Now put the two inductors in SERIES and the values ADD UP - just like resistors in SERIES. This way you can measure very small inductors. VERY CLEVER!

TESTING SWITCHES and RELAYS

Switches and relays have contacts that open and close mechanically and you can test them for CONTINUITY. However these components can become intermittent due to dirt or pitting of the surface of the contacts due to arcing as the switch is opened.

It is best to test these items when the operating voltage and current is present as they quite often fail due to the arcing. A switch can work 49 times then fail on each 50th operation. The same with a relay. It can fail one time in 50 due to CONTACT WEAR.
If the contacts do not touch each other with a large amount of force and with a large amount of the metal touching, the current flowing through the contacts will create HEAT and this will damage the metal and sometimes reduce the pressure holding the contact together.
This causes more arcing and eventually the switch heats up and starts to burn. Switches are the biggest causes of fire in electrical equipment and households.

A relay also has a set of contacts that can cause problems.
There are many different types of relays and basically they can be put into two groups.
1. An electromagnetic relay is a switch operated by magnetic force. This force is generated by current through a coil. The relay opens and closes a set of contacts.
The contacts allow a current to flow and this current can damage the contacts. Connect 5v or 12v to the coil (or 24v) and listen for the "click" of the points closing. Measure the resistance across the points to see if they are closing.
You really need to put a load on the points to see if they are clean and can carry a current.
The coil will work in either direction.
If not, the relay is possibly a CMOS relay or Solid State relay.

2. An electronic relay (Solid State Relay) does not have a winding. It works on the principle of an opto-coupler and uses a LED and Light Activated SCR or Opto-TRIAC to produce a low resistance on the output. The two pins that energise the relay (the two input pins) must be connected to 5v (or 12v) around the correct way as the voltage is driving a LED (with series resistor). The LED illuminates and activates a light-sensitive device.

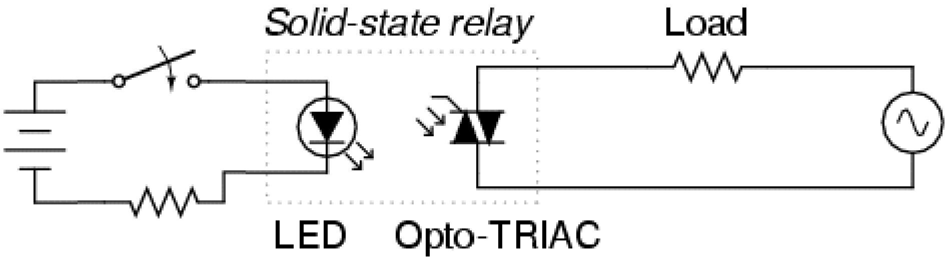

CAPACITORS

Capacitors are one of the most difficult things to test. That's because they don't give a reading on a multimeter and their value can range from 1p to 100,000u.
A faulty capacitor may be "open" when measured with a multimeter, and a good capacitor will also be "open."
You need a piece of test equipment called a CAPACITANCE METER to measure the value of a capacitor.

HOW A CAPACITOR WORKS

There are two ways to describe how a capacitor works. Both are correct and you have to combine them to get a full picture.
A capacitor has INFINITE resistance between one lead and the other.
This means no current flows **through** a capacitor. But it works in another way.
Suppose you have a strong magnet on one side of a door and a piece of metal on the other. By sliding the magnet up and down the door, the metal rises and falls.
The metal can be connected to a pump and you can pump water by sliding the magnet up and down.
A capacitor works in exactly the same way.
If you raise a voltage on one lead of a capacitor, the other lead will rise to the same voltage. This needs more explaining - we are keeping the discussion simple.
It works just like the magnetic field of the magnet through a door.
The next concept is this:
Capacitors are equivalent to a tiny rechargeable battery.
They store energy when the supply-voltage is present and release it when the supply drops.
These two concepts can be used in many ways and that's why capacitors perform tasks such as filtering, time-delays, passing a signal from one stage to another and create many different effects in a circuit.

CAPACITOR VALUES
The basic unit of capacitance is the FARAD. (C) This is the value used in all equations, but it is a very large value. A one FARAD capacitor would be the size of a car if made with plates and paper. Most electronic circuits use capacitors with smaller values such as 1p to 1,000u. 1p is about equal to two parallel wires 2cm long. 1p is one picofarad.

The easiest way to understand capacitor values is to start with a value of 1u. This is one microfarad and is one-millionth of a Farad. A 1 microfarad capacitor is about 1cm long and the diagram shows a 1u electrolytic.

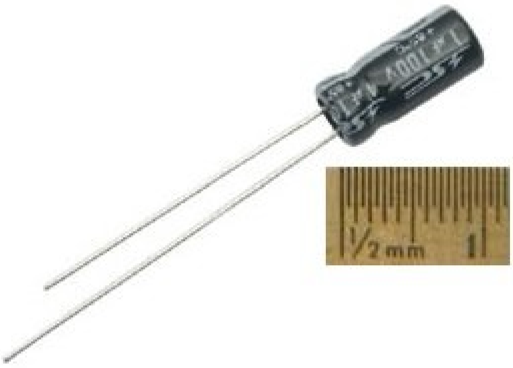

Smaller capacitors are ceramic and they look like the following. This is a 100n ceramic:

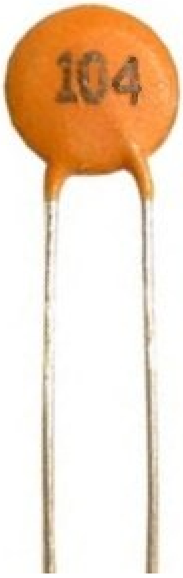

To read the value on a capacitor you need to know a few facts.

The basic value of capacitance is the FARAD.
1 microfarad is one millionth of 1 farad.
1 microfarad is divided into smaller parts called nanofarad.
1,000 nanofarad = 1 microfarad
Nanofarad is divided into small parts called picofarad
1,000 picofarad = 1 nanofarad.

Recapping:
1p = 1 picofarad. 1,000p = 1n (1 nanofarad)
1,000n = 1u (1 microfarad)
1,000u = 1millifarad
1,000,000u = 1 FARAD.

Examples:
All ceramic capacitors are marked in "p" (puff")
A ceramic with 22 is 22p = 22 picofarad
A ceramic with 47 is 47p = 47 picofarad
A ceramic with 470 is 470p = 470 picofarad

A ceramic with 471 is 470p = 470 picofarad
A ceramic with 102 is 1,000p = 1n
A ceramic with 223 is 22,000p = 22n
A ceramic with 104 is 100,000p = 100n = 0.1u

TYPES OF CAPACITOR

For testing purposes, there are two types of capacitor.
Capacitors from 1p to 100n are non-polar and can be inserted into a circuit around either way.
Capacitors from 1u to 100,000u are electrolytics and are polarised. They must be fitted so the positive lead goes to the supply voltage and the negative lead goes to ground (or earth).
There are many different sizes, shapes and types of capacitor. They are all the same. They consist of two plates with an insulating material between. The two plates can be stacked in layers or rolled together.
The important factor is the insulating material. It must be very thin to keep things small. This gives the capacitor its VOLTAGE RATING.
If a capacitor sees a voltage higher than its rating, the voltage will "jump through" the insulating material or around it.
If this happens, a carbon deposit is left behind and the capacitor becomes "leaky" or very low resistance, as carbon is conductive.

CERAMIC CAPACITORS

Nearly all small capacitors are **ceramic capacitors** as this material is cheap and the capacitor can be made in very thin layers to produced a high capacitance for the size of the component. This is especially true for surface-mount capacitors.
All capacitors are marked with a value and the basic unit is: "p" for "puff" However NO surface mount capacitors are marked and they are very difficult to test.

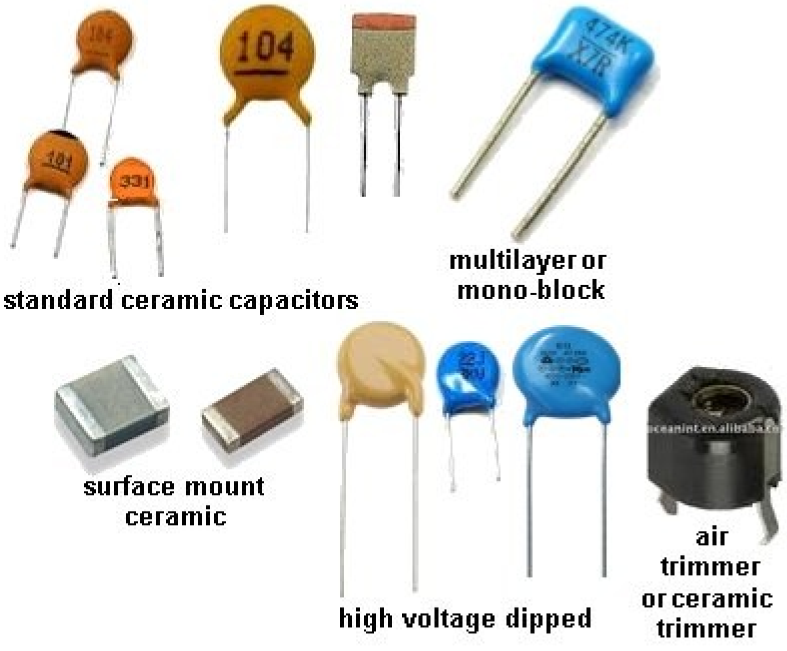

VALUE:	VALUE WRITTEN ON THE COMPONENT:
0.1p	0p1
0.22p	0p22
0.47p	0p47
1.0p	1p0
2.2p	2p2
4.7p	4p7
5.6p	5p6
8.2p	8p2
10p	10 or 10p
22p	22 or 22p
47p	47 or 47p

56p	56 or 56p
100p	100 on 101
220p	220 or 221
470p	470 or 471
560p	560 or 561
820p	820 or 821
1,000p (1n)	102
2200p (2n2)	222
4700p (4n7)	472
8200p (8n2)	822
10n	103
22n	223
47n	473
100n	104
220n	224
470n	474
1u	105

POLYESTER, POLYCARBONATE, POLYSTYRENE, MYLAR, METALLISED POLYESTER, ("POLY"), MICA and other types of CAPACITOR

There are many types of capacitor and they are chosen for their reliability, stability, temperate-range and cost.

For testing and repair work, they are all the same. Simply replace with exactly the same type and value.

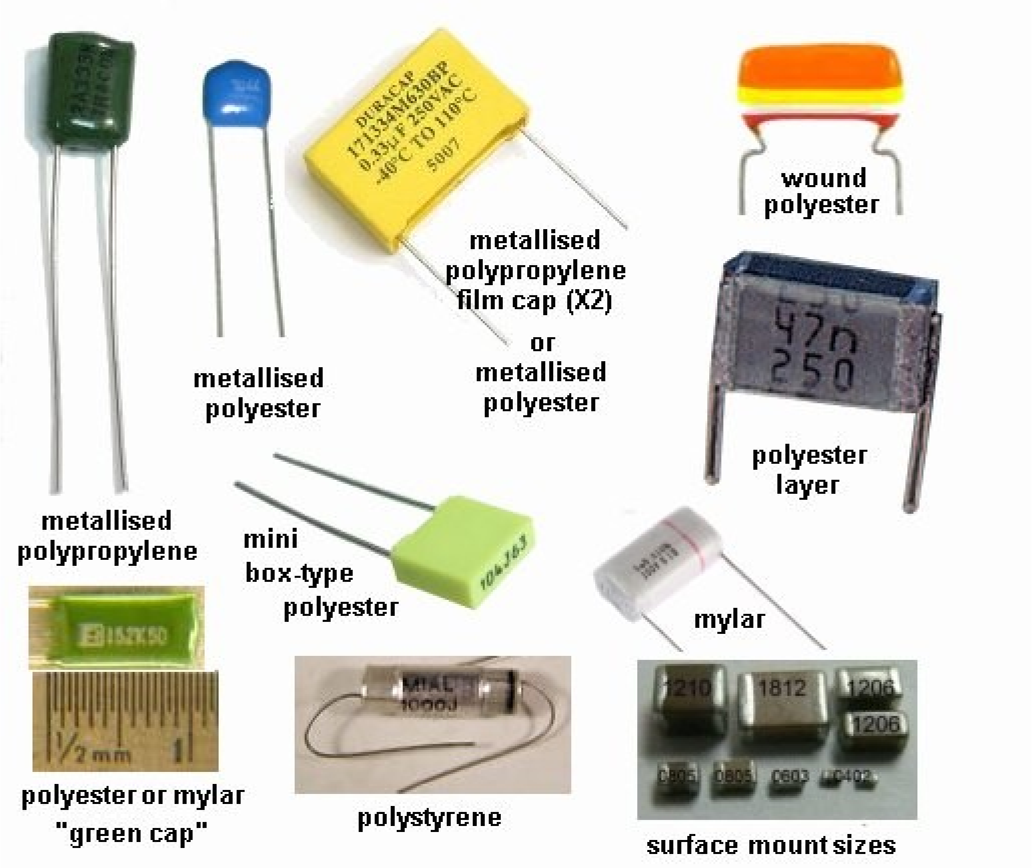
DURACAP 171334M630BP 0.33µF 250VAC -40°C TO 110°C 5007
wound polyester
metallised polypropylene film cap (X2) or metallised polyester
metallised polyester
47n 250
polyester layer
metallised polypropylene
mini box-type polyester
104J63
mylar
152K50
1/2 mm
polyester or mylar "green cap"
MIAL 1000J
polystyrene
1210
1812
1206
1206
0805
0805
0603
0402
surface mount sizes

Capacitor Colour Code Table

Colour	Digit A	Digit B	Multiplier D	Tolerance (T) > 10pf	Tolerance (T) < 10pf	Temperature Coefficient (TC)
Black	0	0	x1	± 20%	± 2.0pF	
Brown	1	1	x10	± 1%	± 0.1pF	-33×10^{-6}
Red	2	2	x100	± 2%	± 0.25pF	-75×10^{-6}

Orange	3	3	x1,000	± 3%		-150x10^{-6}
Yellow	4	4	x10,000	± 4%		-220x10^{-6}
Green	5	5	x100,000	± 5%	± 0.5pF	-330x10^{-6}
Blue	6	6	x1,000,000			-470x10^{-6}
Violet	7	7				-750x10^{-6}
Grey	8	8	x0.01	+80%,-20%		
White	9	9	x0.1	± 10%	± 1.0pF	
Gold			x0.1	± 5%		
Silver			x0.01	± 10%		

Pico Farads (pF)	Nano Farads (nF)	Micro Farads (µF)
1	0.001	0.000001
10	0.01	0.00001
100	0.1	0.0001
1,000	1	0.001
10,000	10	0.01
100,000	100	0.1
1,000,000	1,000	1
10,000,000	10,000	10
100,000,000	100,000	100

Type ⊕ = polarized	Pic	Cap Range
Ceramic		**pF - µF**
Mica **(silver mica)**		**pF - nF**
Plastic Film **(polyethylene polystyrene)**	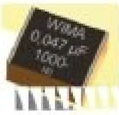	**few µFs**
Tantalum ⊕		**µFs**
OSCON ⊕		**µFs**
Aluminum Electrolytic ⊕		**high µFs**

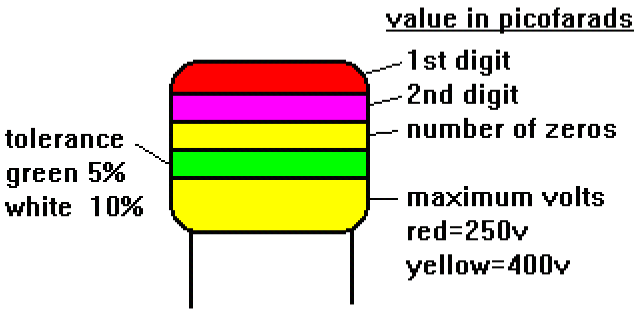
value in picofarads
1st digit
2nd digit
number of zeros
maximum volts
red=250v
yellow=400v
tolerance
green 5%
white 10%

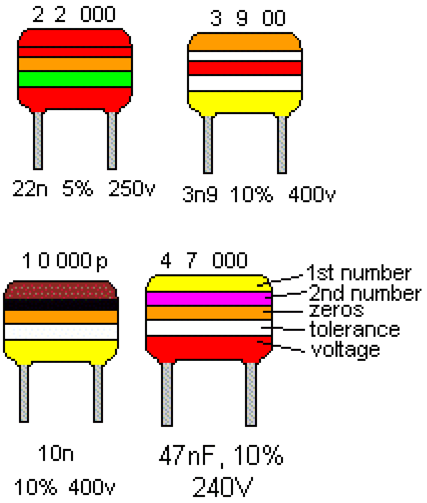
2 2 000
22n 5% 250v
3 9 00
3n9 10% 400v
1 0 000 p
10n
10% 400v
4 7 000
1st number
2nd number
zeros
tolerance
voltage
47nF, 10%
240V

Black	0
Brown	1
Red	2
Orange	3
Yellow	4
Green	5
Blue	6
Violet	7
Grey	8
White	9

ELECTROLYTIC and TANTALUM CAPACITORS

Electrolytics and Tantalums are the same for testing purposes but their performance is slightly different in some circuits. A tantalum is smaller for the same rating as an electrolytic and has a better ability at delivering a current. They are available up to about 1,000u, at about 50v but their cost is much higher than an electrolytic.

Electrolytics are available in 1u, 2u2 3u3 4u7 10u, 22u, 47u, 100u, 220u, 330u, 470u, 1,000u, 2,200u, 3,300u, 4,700u, 10,000u and higher.
The "voltage" or "working voltage" can be: 3.3v, 10v, 16v, 25v, 63v, 100v, 200v and higher.
There is also another important factor that is rarely covered in text books. It is RIPPLE FACTOR. This is the amount of current that can enter and leave an electrolytic. This current heats up the electrolytic and that is why some electrolytics are much larger than others, even though the capacitance and voltage-ratings are the same.
If you replace an electrolytic with a "miniature" version, it will heat up and have a very short life. This is especially important in power supplies where current (energy) is constantly entering and exiting the electrolytic as its main purpose is to provide a smooth output from a set of diodes that delivers "pulsing DC." (see "Power Diodes")

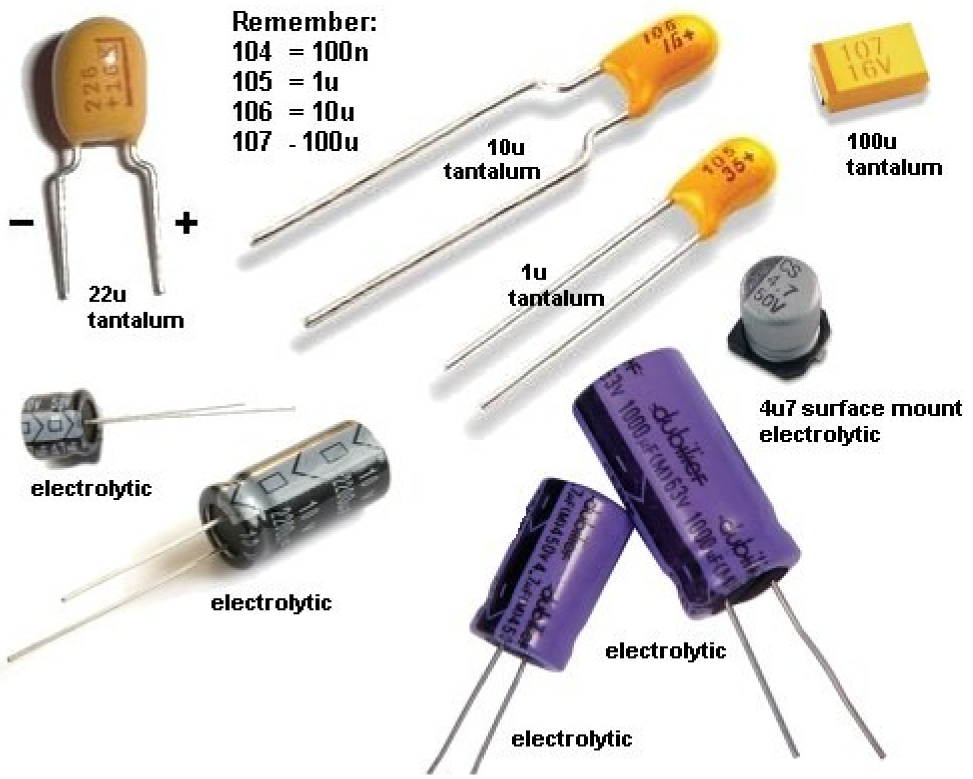
Remember:
104 = 100n
105 = 1u
106 = 10u
107 - 100u
22u tantalum
10u tantalum
100u tantalum
1u tantalum
4u7 surface mount electrolytic
electrolytic
electrolytic
electrolytic
electrolytic
226 +16
106 16+
107 16V
105 35+
CS 4.7 50V
1000µF(M)63V
dubilier

NON-POLAR CAPACITORS (ELECTROLYTICS)

Electrolytics are also available in non-polar values. It sometimes has the letters "NP" on the component. Sometimes the leads are not identified.
This is an electrolytic that does not have a positive and negative lead but two leads and either lead can be connected to the positive or negative of the circuit.
These electrolytics are usually connected to the output of an amplifier (such as in a filter near the speaker) where the signal is rising and falling.
A non-polar electrolytic can be created from two ordinary electrolytics by connecting the negative leads together and the two positive leads become the new leads.
For example: two 100u 63v electrolytics will produce a 47u 63v non-polar electrolytic.
In the circuit below, the non-polar capacitor is replaced with two electrolytics.

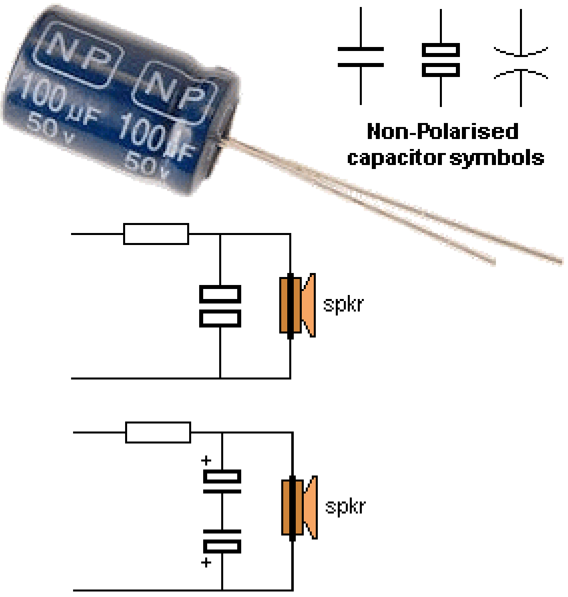
NP 100µF 50v
Non-Polarised capacitor symbols
spkr
+
+
spkr

PARALLEL and SERIES CAPACITORS

Capacitors can be connected in PARALLEL and/or SERIES for a number of reasons.
1. If you do not have the exact value, two or more connected in parallel or series can produce the value you need.

2. Capacitors connected in series will produce one with a higher voltage rating.
3. Capacitors connected in parallel will produce a larger-value capacitance.

Here are examples of two equal capacitors connected in series or parallel and the results they produce:

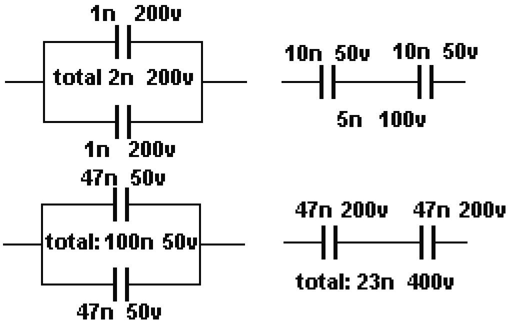

VOLTAGE RATING OF CAPACITOR

Capacitors have a voltage rating, stated as WV for working voltage, or WVDC. This specifies the maximum voltage that can be applied across the capacitor without puncturing the dielectric. Voltage ratings for "poly," mica and ceramic capacitors are typically 50v to 500 VDC. Ceramic capacitors with ratings of 1kv to 5kv are also available. Electrolytic capacitors are commonly available in 6v, 10v 16v, 25v, 50v, 100v, 150v, and 450v ratings.

CAUTION

If a capacitor has a voltage rating of 63v, do not put it in a 100v circuit as the insulation (called the dielectric) will be punctured and the capacitor will "short-circuit." It's ok to replace a 0.22uF 50WV capacitor with 0.22uF 250WVDC.

SAFETY

A capacitor can store a charge for a period of time after the equipment is turned off. High voltage electrolytic caps can pose a safety hazard. These capacitors are in power supplies and some have a resistor across them, called a bleed resistor, to discharge the cap after power is switched off.
If a bleed resistor is not present the cap can retain a charge after the equipment is unplugged.

How to discharge a capacitor

Do not use a screwdriver to short between the terminals as this will damage the capacitor internally and the screwdriver.
Use a 1k 3watt or 5watt resistor on jumper leads and keep them connected for a few seconds to fully discharge the electro.
Test it with a voltmeter to make sure all the energy has been removed.

Before testing any capacitors, especially electrolytics, you should look to see if any are damaged, overheated or leaking. Swelling at the top of an electrolytic indicates heating and pressure inside the case and will result in drying out of the electrolyte. Any hot or warm electrolytic indicates leakage and ceramic capacitors with portions missing indicates something has gone wrong.

TESTING A CAPACITOR

There are two things you can test with a multimeter:
1. A short-circuit within the capacitor
2. Capacitor values above 1u.

You can test capacitors in-circuit for short-circuits. Use the x1 ohms range.
To test a capacitor for leakage, you need to remove it or at least one lead must be removed. Use the x10k range on an analogue or digital multimeter.
For values above 1u you can determine if the capacitor is charging by using an analogue meter. The needle will initially move across the scale to indicate the cap is charging, then go to "no deflection." Any permanent deflection of the needle will indicate leakage.
You can reverse the probes to see if the needle moves in the opposite direction. This indicates it has been charged. Values below 1u will not respond to charging and the needle will not deflect.
This does not work with a digital meter as the resistance range does not output any current and the electrolytic does not charge.

Rather than spending money on a capacitance meter, it is cheaper to replace any suspect capacitor or electrolytic.
Capacitors can produce very unusual faults and no piece of test equipment is going to detect the problem.
In most cases, it is a simple matter to solder another capacitor across the suspect component and view or listen to the result.
This saves all the worry of removing the component and testing it with equipment that cannot possibly give you an accurate reading when the full voltage and current is not present.
It is complete madness to even think of testing critical components such as capacitors, with TEST EQUIPMENT. You are fooling yourself. If the Test Equipment says the component is ok, you will look somewhere else and waste a lot of time.

FINDING THE VALUE OF A CAPACITOR

If you want to find the value of a surface-mount capacitor or one where the markings have been removed, you will need a CAPACITANCE METER. Here is a simple circuit that can be added to your meter to read capacitor values from 10p to 10u.
The full article can be found HERE.

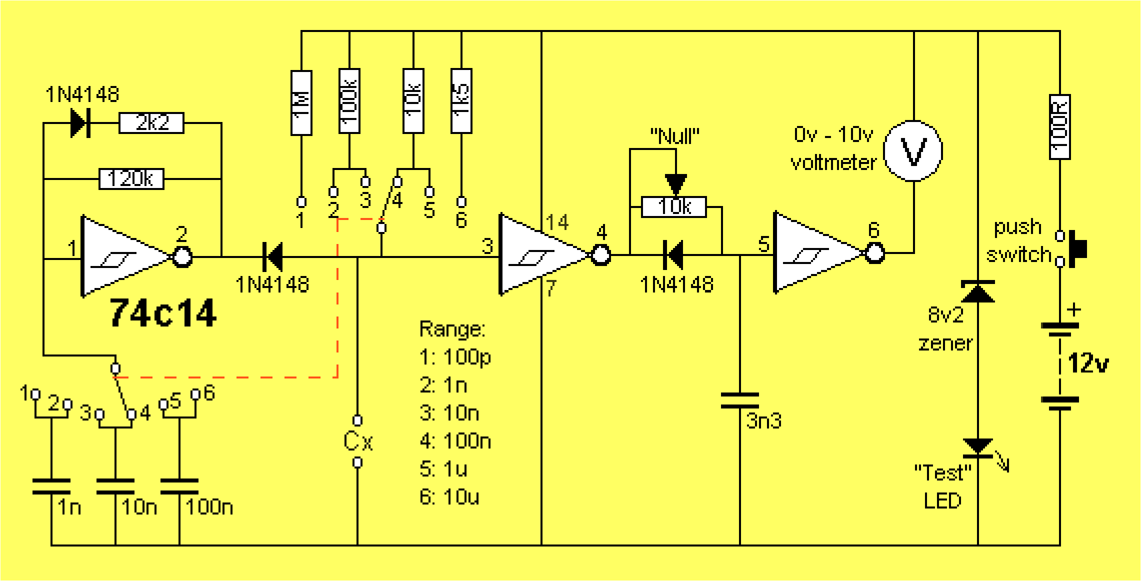

ADD-ON CAPACITANCE METER

REPLACING A CAPACITOR

Always replace a capacitor with the exact same type.
A capacitor may be slightly important in a circuit or it might be extremely critical.
A manufacturer may have taken years to select the right type of capacitor due to previous failures.
A capacitor just doesn't have a "value of capacitance."
It may also has an effect called "tightening of the rails."
In other words, a capacitor has the ability to react quickly and either absorb or deliver energy to prevent spikes or fluctuations on the rail.
This is due to the way it is constructed. Some capacitors are simply plates of metal film while others are wound in a coil. Some capacitors are large while others are small.
They all react differently when the voltage fluctuates.
Not only this, but some capacitors are very stable and all these features go into the decision for the type of capacitor to use.
You can completely destroy the operation of a circuit by selecting the wrong type of capacitor.
No capacitor is perfect and when it gets charged or discharged, it appears to have a small value of resistance in series with the value of capacitance. This is known as "ESR" and stands for EQUIVALENT SERIES RESISTANCE. This effectively makes the capacitor slightly slower to charge and discharge.
We cannot go into the theory on selecting a capacitor as it would be larger than this eBook so the only solution is to replace a capacitor with an identical type.
However if you get more than one repair with identical faults, you should ask other technicians if the original capacitor comes from a faulty batch.
The author has fixed TV's and fax machines where the capacitors have been inferior and alternate types have solved the problem.

Some capacitor are suitable for high frequencies, others for low frequencies.

TESTING DIODES

Diodes can have 4 different faults.
1. Open circuit in both directions.
2. Low resistance in both directions.
3. Leaky.
4. Breakdown under load.

TESTING A DIODE ON AN ANALOGUE METER

Testing a diode with an **Analogue Multimeter** can be done on any of the resistance ranges. [The high resistance range is best - it sometimes has a high voltage battery for this range but this does not affect our testing]
There are two things you must remember.
1. When the diode is measured in one direction, the needle will **not move at all**. The technical term for this is the diode is **reverse biased**. It will not allow any current to flow. Thus the needle will not move.
When the diode is connected around the other way, the needle will swing to the right (move up scale) to about 80% of the scale. This position represents the voltage drop across the junction of the diode and is NOT a resistance value. If you change the resistance range, the needle will move to a slightly different position due to the resistances inside the meter. The technical term for this is the diode is **forward biased**. This indicates the diode is not faulty.
The needle will swing to a slightly different position for a "normal diode" compared to a Schottky diode. This is due to the different junction voltage drops.
However we are only testing the diode at very low voltage and it may break-down when fitted to a circuit due to a higher voltage being present or due to a high current flowing.

2. The leads of an **Analogue Multimeter** have the positive of the battery connected to the black probe and the readings of a "good diode" are shown in the following two diagrams:

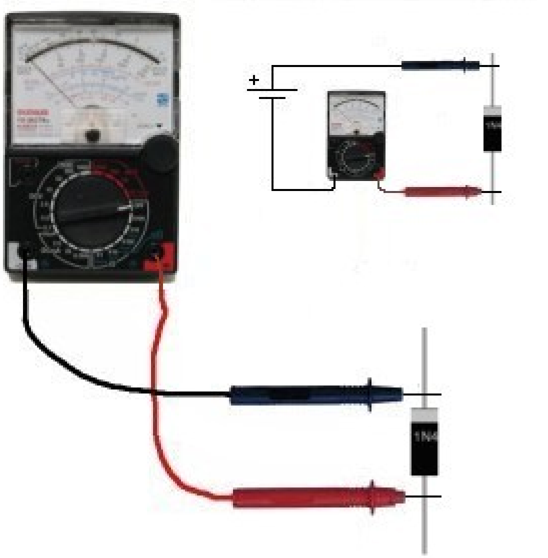

The diode is REVERSE BIASED in the diagram above and diodes not conduct.

Here is the equivalent circuit:

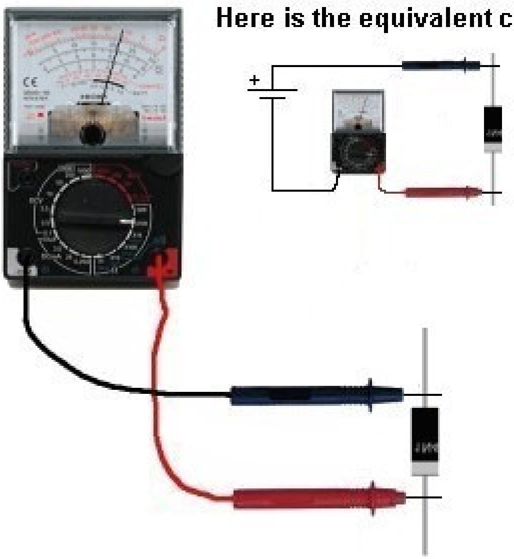

The diode is FORWARD BIASED in the diagram above and it conducts

TESTING A DIODE ON A DIGITAL METER

Testing a diode with a Digital Meter must be done on the "DIODE" setting as a digital meter does not deliver a current through the probes on some of the resistance settings and will not produce an accurate reading.

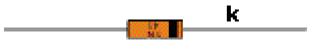

signal diode

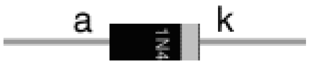

power diode

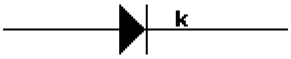

diode symbol

The best thing to do with a "suspect" diode is to replace it. This is because a diode has a number of characteristics that cannot be tested with simple equipment. Some diodes have a fast recovery for use in high frequency circuits. They conduct very quickly and turn off very quickly so the waveform is processed accurately and efficiently.
If the diode is replaced with an ordinary diode, it will heat up as does not have the high-speed characteristic.
Other diodes have a low drop across them and if an ordinary is used, it will heat up.
Most diodes fail by going: SHORT-CIRCUIT. This can be detected by a low resistance (x1 or x10 Ohms range) in both directions.
A diode can also go OPEN CIRCUIT. To locate this fault, place an identical diode across the diode being tested.
A leaky diode can be detected by a low reading in one direction and a slight reading the other direction.
However this type of fault can only be detected when the circuit is working. The output of the circuit will be low and sometimes the diode heats up (more than normal).
A diode can go open under full load conditions and perform intermittently.
Diodes come in pairs in surface-mount packages and 4 diodes can be found in a bridge.
They are also available in pairs that look like a 3-leaded transistor.
The line on the end of the body of a diode indicates the cathode and you cannot say "this is the positive lead." The correct way to describe the leads is to say the "cathode lead." The other lead is the anode. The cathode is defined as the electrode (or lead) through which an electric current flows out of a device.
The following diagrams show different types of diodes:

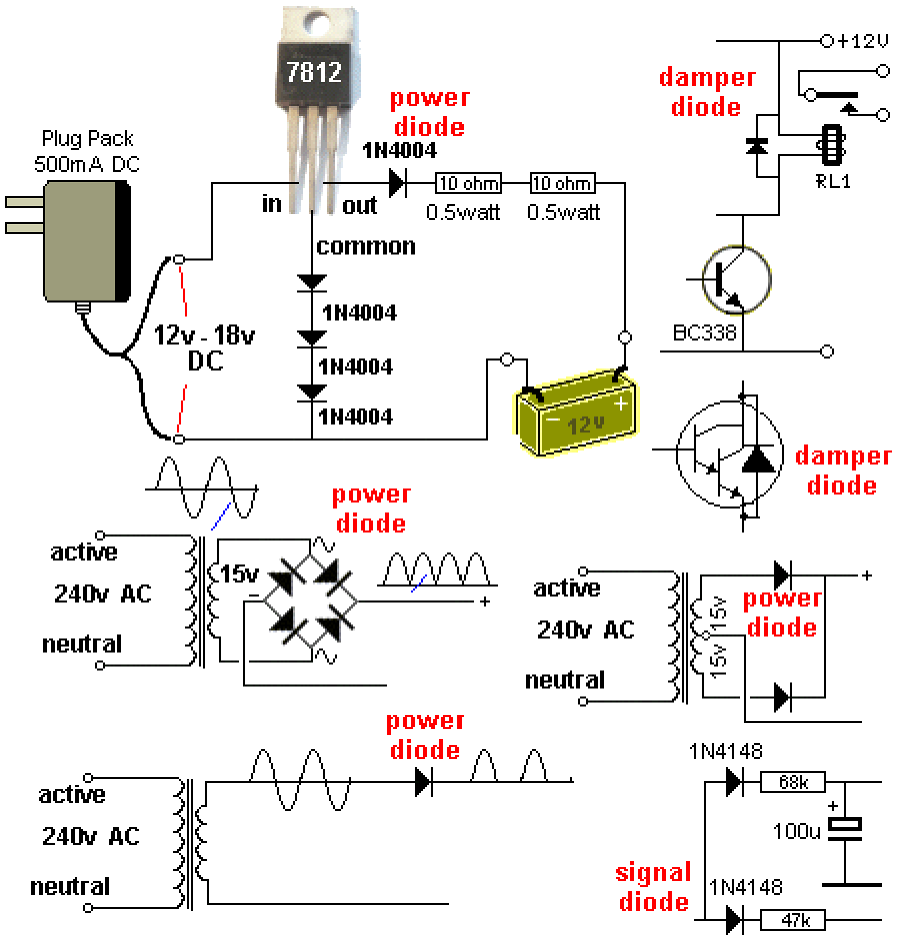
7812
power diode
1N4004
Plug Pack
500mA DC
in
out
common
10 ohm
0.5watt
10 ohm
0.5watt
1N4004
1N4004
1N4004
12v - 18v
DC
12v
damper diode
+12V
RL1
BC338
damper diode
power diode
active
240v AC
neutral
15v
active
240v AC
neutral
15v
15v
power diode
power diode
active
240v AC
neutral
1N4148
68k
100u
signal diode
1N4148
47k

POWER DIODES

To understand how a power diode works, we need to describe a few things. This has NEVER been described before, so read carefully.

The 240v AC (called the "mains") consists of two wires, one is called the ACTIVE and the other is NEUTRAL. Suppose you touch both wires. You will get a shock. The neutral is connected to an earth wire (or rod driven into the ground or connected to a water pipe) at the point where the electricity enters the premises and you do not get a shock from the NEUTRAL.

But the voltage on the active is rising to +345v then goes to -345v at the rate of 50 times per second (for a complete cycle).

345v is the peak voltage of 240v. You never get a 240v shock. (It is a 345v shock.)

In other words, if you touch the two wires at a particular instant, you would get a POSITIVE 345v shock and at another instant you would get a negative 345v shock. This is shown in the diagram below.

We now transfer this concept to the output of a transformer. The diagram shows an AC waveform on the output of the secondary.

This voltage is rising 15v higher than the bottom lead then it is 15v LOWER than the bottom lead. The bottom lead is called "zero volts." You have to say one lead or wire is not "rising and falling" as you need a "reference" or starting-point" or "zero point" for voltage measurements.

The diode only conducts when the voltage is "above zero" (actually when it is 0.7v above zero) and does not conduct (at all) when the voltage goes below zero.

This is shown on the output of the Power Diode. Only the positive peaks or the positive parts of the waveform appear on the output and this is called "pulsing DC." This is called "half-wave" and is not used in a power supply. We have used it to describe how the diode works. The electrolytics charge during the peaks and deliver energy when the diode is not delivering current. This is how the output becomes a steady DC voltage.

Power supplies use FULL WAVE rectification and the other half of the AC waveform is delivered to the output (and fills in the "gaps") and appears as shown in "A."

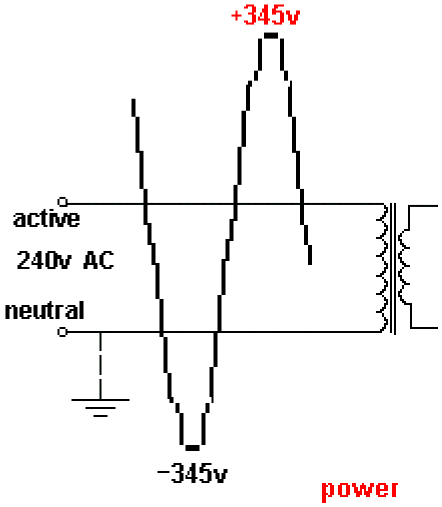

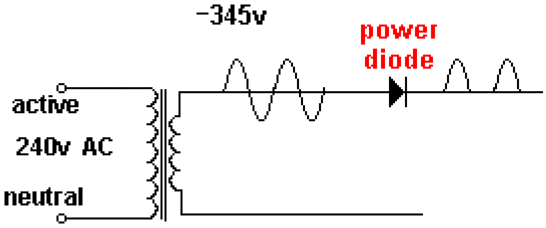

Half-wave rectifier

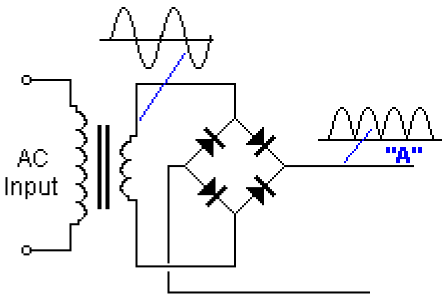

Full-wave rectifier

DAMPER DIODES

A damper diode is a diode that detects a high voltage and SQUELCHES IT (reduces it - removes it).
The signal that it squelches is a voltage that is in the opposite direction to the "supply voltage" and is produced by the collapsing of a magnetic field. Whenever a magnetic filed collapses, it produces a voltage in the winding that is opposite to the supply voltage and can be much higher. This is the principle of a flyback circuit or EHT circuit. The high voltage comes from the transformer.
The diode is placed so that the signal passes through it and less than 0.5v appears across it.
A damper diode can be placed across the coil of a relay, incorporated into a transistor or FET or placed across a winding of a flyback transformer to protect the driving transistor or FET.
It can also be called a "Reverse-Voltage Protection Diode," "Spike Suppression Diode," or "Voltage Clamp Diode."
The main characteristic of a Damper Diode is HIGH SPEED so it can detect the spike and absorb the energy.
It does not have to be a high-voltage diode as the high voltage in the circuit is being absorbed by the diode.

SILICON, GERMANIUM AND SCHOTTKY DIODES

When testing a diode with an analogue meter, you will get a low reading in one direction and a high (or NO READING) in the other direction. When reading in the LOW direction, the needle will swing nearly full scale and the reading is not a resistance-value but a reflection of the characteristic voltage drop across the junction of the diode. As we mentioned before, a resistance reading is really a voltage reading and the meter is measuring the voltage of the battery minus the voltage-drop across the diode.
Since Silicon, Germanium and Schottky Diodes have slightly different characteristic voltage drops across the junction, you will get a slightly different reading on the scale. This does not represent one diode being better than the other or capable of handling a higher current or any other feature.
The quickest, easiest and cheapest way to find, fix and solve a problem caused by a faulty diode is to replace it.
There is no piece of test equipment capable of testing a diode fully, and the circuit you are working on

is actually the best piece of test equipment as it is identifying the fault UNDER LOAD.

Only very simple tests can be done with a multimeter and it is best to check a diode with an ANALOGUE MULTIMETER as it outputs a higher current though the diode and produces a more-reliable result.
A Digital meter can produce false readings as it does not apply enough current to activate the junction.
Fortunately almost every digital multimeter has a **diode test mode.** Using this, a silicon diode should read a voltage drop between 0.5v to 0.8v in the forward direction and open in the reverse direction.
For a germanium diode, the reading will be lower, around 0.2v - 0.4v in the forward direction. A bad diode will read zero volts in both directions.

LIGHT EMITTING DIODES (LEDs)

Light Emitting Diodes (LEDs) are diodes that produce light when current flows from anode to cathode. The LED does not emit light when it is revered-biased. It is used as a low current indicator in many types of consumer and industrial equipment, such as monitors, TV’s, printers, hi-fi systems, machinery and control panels.
The light produced by a LED can be visible, such as red, green, yellow or white. It can also be invisible and these LEDs are called Infrared LEDs. They are used in remote controls and to see if they are working, you need to point a digital camera at the LED and view the picture on the camera screen.
An LED needs about 2v - 3.6v across its leads to make it emit light, but this voltage must be exact for the type and colour of the LED. The simplest way to deliver the exact voltage is to have a supply that is higher than needed and include a voltage-dropping resistor. The value of the resistor must be selected so the current is between 2mA and 25mA.
The cathode of the LED is identified by a flat on the side of the LED. The life expectancy of a LED is about 100,000 hours. LEDs rarely fail but they are very sensitive to heat and they must be soldered and de-soldered quickly. They are one of the most heat-sensitive components.
Light emitting diodes cannot be tested with most multimeters because the characteristic voltage across them is higher than the voltage of the battery in the meter.
However a simple tester can be made by joining 3 cells together with a 220R resistor and 2 alligator clips:

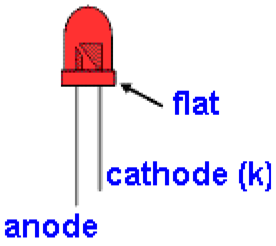

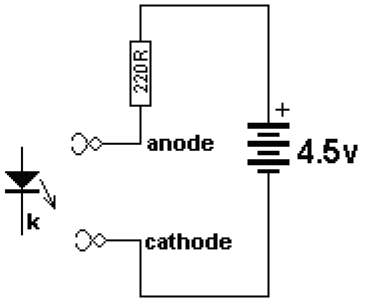

LED TESTER

Connect the clips to a LED and it will illuminate in only one direction.
The colour of the LED will determine the voltage across it. You can measure this voltage if you want to match two or more LEDs for identical operation.
Red LEDs are generally 1.7v to 1.9v. - depending on the quality such as "high-bright"
Green LEDs are 1.9v to 2.3v.
Orange LEDs are about 2.3v and
White LEDs and IR LEDs are about 3.3v to 3.6v.
The illumination produced by a LED is determined by the quality of the crystal. It is the crystal that produces the colour and you need to replace a LED with the same quality to achieve the same illumination.
Never connect a LED across a battery (such as 6v or 9v), as it will be instantly damaged. You must have a resistor in series with the LED to limit the current.

ZENER DIODES

All diodes are Zener diodes. For instance a 1N4148 is a 120v zener diode as this is its reverse breakdown voltage.
And a zener diode can be used as an ordinary diode in a circuit with a voltage that is below the zener value.
For instance, 20v zener diodes can be used in a 12v power supply as the voltage never reaches 20v, and the zener characteristic is never reached.

Most diodes have a reverse breakdown voltage above 100v, while most zeners are below 70v. A 24v zener can be created by using two 12v zeners in series and a normal diode has a characteristic voltage of 0.7v. This can be used to increase the voltage of a zener diode by 0.7v. See the diagram above. It uses 3 ordinary diodes to increase the output voltage of a 3-terminal regulator by 2.1v.
To tests a zener diode you need a power supply about 10v higher than the zener of the diode. Connect the zener across the supply with a 1k to 4k7 resistor and measure the voltage across the diode. If it measures less than 1v, reverse the zener.
If the reading is high or low in both directions, the zener is damaged.

Here is a zener diode tester. The circuit will test up to 56v zeners.

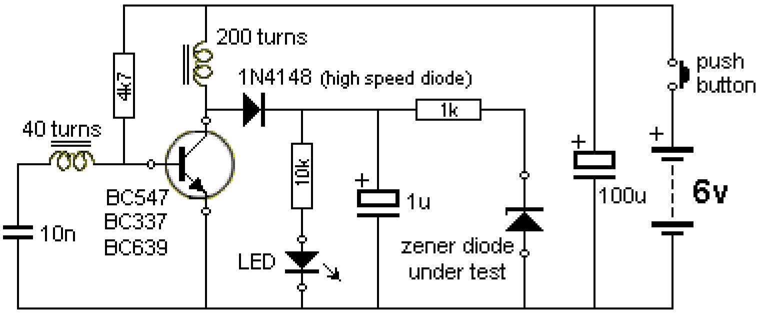

ZENER DIODE TESTER

TRANSFORMERLESS POWER SUPPLY

Here's a circuit that uses zener diodes in a power supply to show how they work. This clever design uses 4 diodes in a bridge to produce a fixed voltage power supply capable of supplying 35mA.
If we put 2 zener diodes in a bridge with two ordinary power diodes, the bridge will break-down at the voltage of the zener. This is what we have done. If we use 18v zeners, the output will be 17v4.

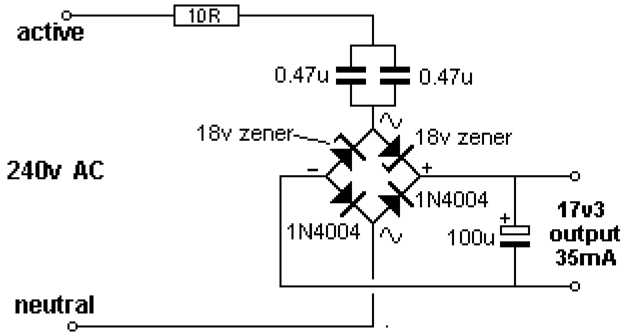

SUPPLY USING ZENER DIODES

When the incoming voltage is positive at the top, the left zener provides 18v limit (and the other zener produces a drop of 0.6v). This allows the right zener to pass current just like a normal diode. The output is 17v4. The same with the other half-cycle.
You cannot use this type of bridge in a normal power supply as the zener diode will "short" when the input voltage reaches the zener value. The concept only works in the circuit above.

VOLTAGE REGULATORS

A Voltage Regulator takes a high input voltage and delivers a fixed output voltage.
Providing the input voltage is 4v above the output voltage, the regulator will deliver a fixed output voltage with almost no ripple.
Voltage regulators are also called "3-TERMINAL REGULATORS" or "REGULATOR IC's" - although this name is not generally used.
In most cases, a voltage regulator gets quite hot and for this reason it has a high failure-rate.
If a regulator is not getting hot (or warm) it has either failed or the circuit is not operating.
A regulator can only decrease the voltage. It cannot increase the current. This means the current being supplied to a circuit must also be available from the circuit supplying the regulator.
All regulators have different pin-outs, so you need to find the input pin and output pin and make sure

the voltage-difference is at least 4v. Some regulators will work with a difference as low as 1v, so you need to read the specifications for the type you are servicing.
Some regulators are called "negative voltage regulators" and the input voltage will be negative and the output will be negative.
You need to test a voltage regulator with the power "ON".
Make sure you do not allow the probes to short any of the pins together as this will destroy the regulator or the circuit being supplied.
With the power turned off or the regulator removed from the circuit, you can test it with a multimeter set to resistance to see if it is ok. If any resistance readings are very low or zero ohms, the regulator is damaged.

TRANSFORMERS

All transformers and coils are tested the same way. This includes chokes, coils, inductors, yokes, power transformers, EHT transformers (flyback transformers), switch mode transformers, isolation transformers, IF transformers, baluns, and any device that has turns of wire around a former. All these devices can go faulty.
The coating on the wire is called insulation or "enamel" and this can crack or become overheated or damaged due to vibration or movement. When two turns touch each other, a very interesting thing happens. **The winding becomes two separate windings.**

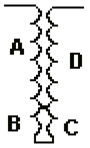

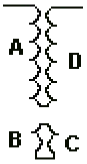

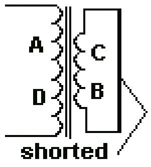

We will take the case of a single winding such as a coil. This is shown in the first diagram above and the winding is wound across a former and back again, making two layers. The bottom and top layers touch at the point shown in the diagram and the current that originally passed though A, B, C, D now passes though A & D.

Winding B C becomes a separate winding as shown in the second diagram.
In other words the coil becomes a TRANSFORMER with a SHORT CIRCUIT on the secondary winding as shown in the third diagram.
When the output wires of a transformer are shorted together, it delivers a very high current because you have created a SHORT-CIRCUIT. This short-circuit causes the transformer to get very hot.
That's exactly what happens when any coil or transformer gets a "shorted turn."
The shorted turns can be a single turn or many turns.
It is not possible to measure a fault like this with a multimeter as you don't know the exact resistance of a working coil or winding and the resistance of a faulty winding may be only 0.001 ohms less.
However when a transformer or coil is measured with an inductance meter, an oscillating voltage (or spike) is delivered into the core as magnetic flux, then the magnetic flux collapses and passes the energy into the winding to produce a waveform. The inductance meter reads this and produces a value of inductance in Henry (milliHenry or microHenry.)
This is done with the transformer removed from the circuit and this can be a very difficult thing to do, as most transformers have a number of connections.
If the coil or transformer has a shorted turn, the energy from the magnetic flux will pass into the turns that are shorted and produce a current. Almost no voltage will be detected from winding.
The reading from the inductance meter will be low or very low and you have to work out if it is correct.

However there is one major problem with measuring a faulty transformer or coil.
It may only become faulty when power is applied.
The voltage between the turns may be sparking or jumping a gap and creating a problem. A tester is not going to find this fault.
Secondly, an inductance meter may produce a reading but you do not know if the reading is correct.
An improved tester is a RING TESTER.
The circuit for a ring tester can be found here:

http://www.flippers.com/pdfs/k7205.pdf

It sends a pulse to the coil and counts the number of returning pulses or "rings." A faulty coil (or winding) may return one pulse but nearly all the energy will be passed to the shorted turns and you will be able to see this on the scale. You will only get one or two return pulses, whereas a good winding will return more pulses.

One way to detect a faulty power transformer is to connect it to the supply and feel the temperature-rise (when nothing is connected to the secondary).
It should NOT get hot.
Detecting shorted turns is not easy to diagnose as you really need another identical component to compare the results.
Most transformers get very hot when a shorted turn has developed. It may deliver a voltage but the heat generated and a smell from the transformer will indicate a fault.

ISOLATION TRANSFORMER

An isolation transformer is a piece of Test Equipment that provides "Mains Voltage" but the voltage is "floating." You will still get a shock if you touch the two output leads, but it has a special use when testing unknown equipment.
Many electrical appliances are fully insulated and only have two leads connected to the mains.
When you take these appliances apart, you do not know which end of say a heating element is connected to the "live" (active) side of the mains and which end connects to the neutral.
I am not suggesting you carry out the following tests, but they are described to show how an isolation transformer works.
If you touch a soldering iron on the "live" (active) end of the heating element it will cause a short-circuit.
However when the appliance is connected to the main via an **isolation transformer**, you can touch an earthed soldering iron on either end of the heater as both leads from the isolation transformer are "floating."
Note: As soon as you earth one lead of the output an isolation transformer, the other lead becomes "active."
You can make your own **Isolation Transformer** by connecting two identical transformers "back-to-back."
The following diagram shows how this is done:

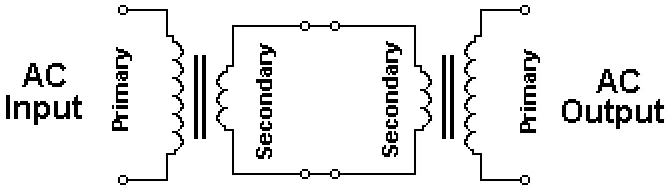

You can use any transformers providing the primary and secondary voltages are the same. The current capability of the secondary winding does not matter. However if you want a supply that has almost the same voltage as your "Mains," you need two transformers with the same voltages.
This handy isolation transformer will provide you with "Mains Voltage" but with a limited current.
In other words it will have a limited capability to supply "wattage." If you are using two 15VA transformers, you will only be able to test an appliance rated at 15 watts.
This has some advantages and some disadvantages.
If you are working on a project, and a short-circuit occurs, the damage will be limited to 15 watts.
If you are using two transformers with different VA ratings, the lower rating will be the capability of the combination.
If the secondaries are not equal, you will get a higher or lower "Mains Voltage."
If you get two old TV's or Monitors with a rating on the compliance plate of 45 watts, or 90 watts, you can assume the transformers are capable of delivering this wattage and making an isolation transformer will enable you to test similar items with the safety of being isolated from the mains.
Colin Mitchell designs a lot of "LED lighting lamps" that are connected directly to the mains. He always works with an isolating transformer, just to be safe. Working on exposed "mains" devices is extremely nerve-wracking and you have to very careful.

DETERMINING THE SPECS OF A TRANSFORMER

Suppose you have a "mains transformer" with unknown output voltages and unknown current capability.
You must be sure it is a mains transformer designed for operation on 50Hz or 60Hz. Switch-Mode transformers operate at frequencies 40kHz and higher and are not covered in this discussion.
To be on the safe-side, connect the unknown transformer to the output of your isolating transformer.
Since the transformer will take almost no current when not loaded, the output voltages it produces will be fairly accurate. Measure the input AC voltage and output AC voltage.

If the transformer has loaded your isolating transformer it will be faulty.
Mains transformers are approx 15 VA for 500gm, 30VA for 1kgm 50VA for 2kgm and and 100VA for 2.5kgm.
VA stands for Volts-Amps and is similar to saying watts. Watts is used for DC circuits, while VA refers to AC circuits.
Once you have the weight of the transformer and the output voltage, you can work out the current capability of the secondary.
For transformers up to 30vA, the output voltage on no-load is 30% higher than the final "loaded voltage."
This is due to the poor regulation of these small devices.
If the transformer is 15VA and the output voltage will be 15v AC, the current will be 1 amp AC.
You can check the "quality" of the transformer, (the regulation) by fully loading the output and measuring the final voltage. If the transformer has a number of secondaries, the VA rating must be divided between all the windings.

OPTO ISOLATORS and OPTO COUPLERS

Opto Isolators and Opto Couplers are the same thing. A common opto-coupler is 4N35. It is used to allow two circuits to exchange signals yet remain electrically isolated. The signal is applied to the LED, which shines on a silicon NPN photo-transistor in the IC.
The light is proportional to the signal, so the signal is transferred to the photo transistor to turn it on a proportional amount. Opto-couplers can have Light Activated SCR's, photodiodes, TRIAC's and other semiconductor devices as an output. The 4N35 opto-coupler schematic is shown below:

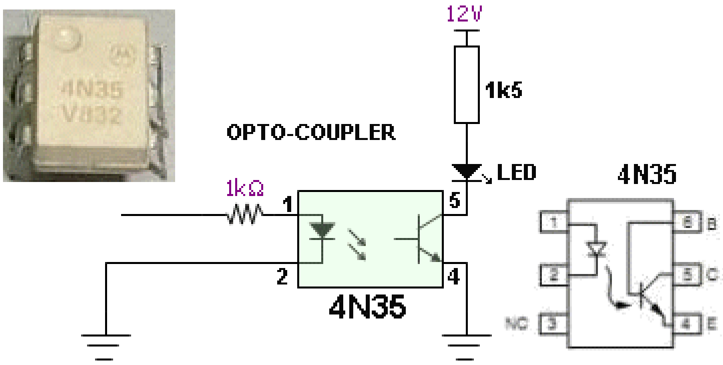

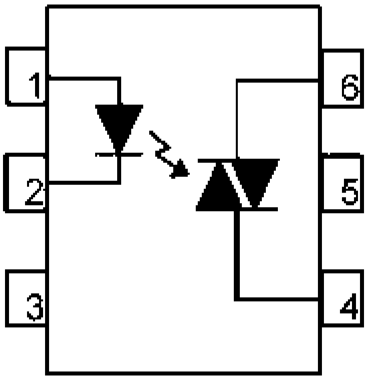

An opto-Coupler using a TRIAC

TESTING AN OPTO COUPLER

Most multimeters cannot test the LED on the input of an opto-coupler because the ohms range does not have a voltage high enough to activate the LED with at least 2mA.
You need to set-up the test-circuit shown above with a 1k resistor on the input and 1k5 on the output.
When the 1k is connected to 12v, the output LED will illuminate.
The opto-coupler should be removed from circuit to perform this test.

TRANSISTORS

Transistors are solid-state devices and although they operate completely differently to a diode, they appear as two back-to-back diodes when tested.
There are basically 2 types of transistor NPN and PNP.

A transistor is sometimes referred to as BJT (Bi-polar Junction Transistor) to distinguish it from other types of transistor such as Field Effect transistor, Programmable Unijunction Transistor and others.
In the following diagram, two diodes are connected together and although the construction of a transistor is more complex, we see the transistor as two diodes when testing it.

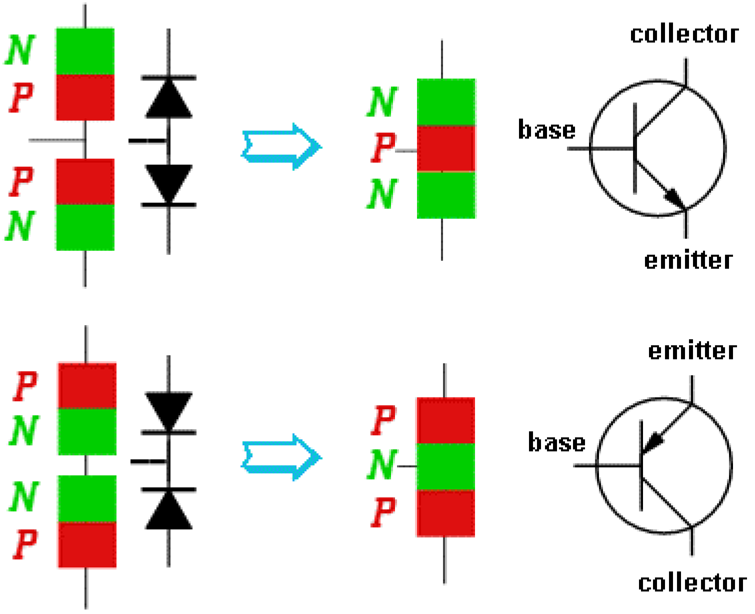

A TRANSISTOR APPEARS AS TWO DIODES WHEN TESTING IT

All transistors have three leads. Base (b), Collector (c), and Emitter (e).
For an NPN transistor, the arrow on the emitter points away from the base.
It is fortunate that the arrow on both symbols points in the direction of the flow of current (Conventional Current) and this makes it easy to describe testing methods using our simplified set of instructions. The symbols have been drawn exactly as they appear on a circuit diagram.

All transistors **are the same** but we talk about digital and analogue transistors. There is no difference between the two.
The difference is the circuit. And the only other slight difference between transistors is the fact that some have inbuilt diodes and resistors to simplify the rest of the circuit.
All transistors work the same way. The only difference is the amount of amplification they provide, the current and voltage they can withstand and the speed at which they work. For simple testing purposes, they are all the same.

NPN transistors are the most common and for an NPN transistor, the following applies.
(the opposite applies for PNP)
To test a transistor, there is **one thing** you have to know:
When the base voltage is higher than the emitter, current flows though the collector-emitter leads.
As the voltage is increased on the base, nothing happens until the voltage reaches 0.55v. At this point a very small current flows through the collector-emitter leads. As the voltage is increased, the current-flow increases. At about 0.75v, the current-flow is a MAXIMUM. (can be as high as 0.9v). That's how it works. A transistor also needs **current** to flow into the base to perform this amplifying function and this is the one feature that separates an ordinary transistor from a FET.

If the voltage on the base is 0v, then instantly goes to 0.75v, the transistor initially passes NO current, then FULL current. The transistor is said to be working in its two states: OFF then ON (sometimes called: "cut-off" and "saturation"). These are called digital states and the transistor is said to be a **DIGITAL TRANSISTOR** or a **SWITCHING TRANSISTOR** , working in **DIGITAL MODE.**

If the base is delivered 0.5v, then slowly rises to 0.75v and slowly to 0.65v, then 0.7v, then 0.56v etc, the transistor is said to be working in ANALOGUE MODE and the transistor is an **ANALOGUE TRANSISTOR**.
Since a transistor is capable of amplifying a signal, it is said to be an active device. Components such as resistors, capacitors, inductors and diodes are not able to amplify and are therefore known as passive components.

In the following tests, use your finger to provide the **TURN ON** voltage for the base (this is 0.55v to

0.7v) and as you press harder, more current flows into the base and thus more current flows through the collector-emitter terminals. As more current flows, the needle of the multimeter moves UP-SCALE.

TESTING A TRANSISTOR ON A DIGITAL METER

Testing a transistor with a **Digital Meter** must be done on the "DIODE" setting as a digital meter does not deliver a current through the probes on some of the resistance settings and will not produce an accurate reading.
The "DIODE" setting must be used for diodes and transistors. It should also be called a "TRANSISTOR" setting.

TESTING AN unknown TRANSISTOR

The first thing you may want to do is test an unknown transistor for COLLECTOR, BASE AND EMITTER.
You also want to perform a test to find out if it is NPN or PNP.
That's what this test will provide.
You need a cheap multimeter called an ANALOGUE METER - a multimeter with a scale and pointer (needle).
It will measure resistance values (normally used to test resistors) - (you can also test other components) and Voltage and Current. We use the resistance settings. It may have ranges such as "x10" "x100" "x1k" "x10"
Look at the resistance scale on the meter. It will be the top scale.
The scale starts at zero on the right and the high values are on the left. This is opposite to all the other scales.
When the two probes are touched together, the needle swings FULL SCALE and reads "ZERO." Adjust the pot on the side of the meter to make the pointer read exactly zero.

How to read: "x10" "x100" "x1k" "x10"

Up-scale from the zero mark is "1"
When the needle swings to this position on the "x10" setting, the value is 10 ohms.
When the needle swings to "1" on the "x100" setting, the value is 100 ohms.
When the needle swings to "1" on the "x1k" setting, the value is 1,000 ohms = 1k.
When the needle swings to "1" on the "x10k" setting, the value is 10,000 ohms = 10k.
Use this to work out all the other values on the scale.
Resistance values get very close-together (and very inaccurate) at the high end of the scale. [This is just a point to note and does not affect testing a transistor.]

Step 1 - FINDING THE BASE and determining NPN or PNP

Get an unknown transistor and test it with a multimeter set to "x10"
Try the 6 combinations and when you have the black probe on a pin and the red probe touches the other pins and the meter swings nearly full scale, you have an NPN transistor. The black probe is BASE
If the red probe touches a pin and the black probe produces a swing on the other two pins, you have a PNP transistor. The red probe is BASE
If the needle swings FULL SCALE or if it swings for more than 2 readings, the transistor is **FAULTY**.

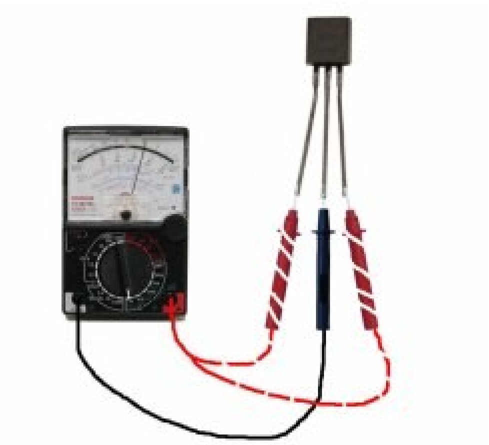

This is an NPN transistor
The black probe is the BASE

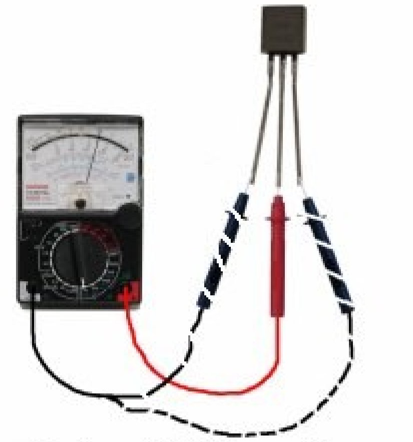

This is a PNP transistor
The red probe is the BASE

Step 2 - FINDING THE COLLECTOR and EMITTER

Set the meter to "x10k."

For an NPN transistor, place the leads on the transistor and when you press hard on the two leads shown in the diagram below, the needle will swing almost full scale.

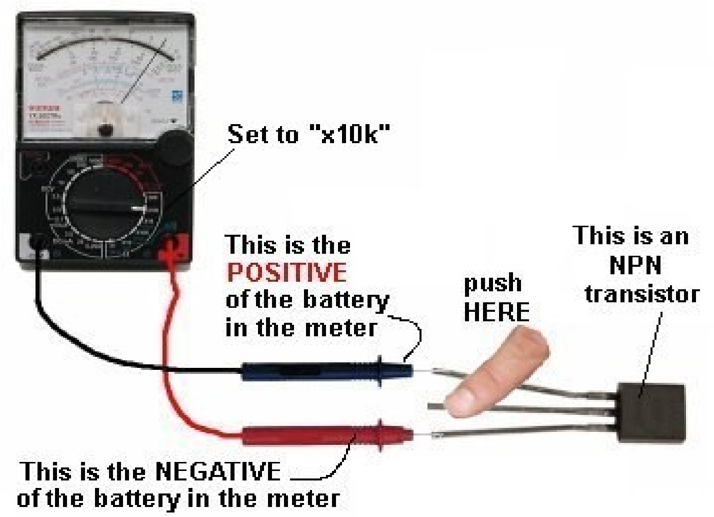
Set to "x10k"
This is the POSITIVE of the battery in the meter
push HERE
This is an NPN transistor
This is the NEGATIVE of the battery in the meter

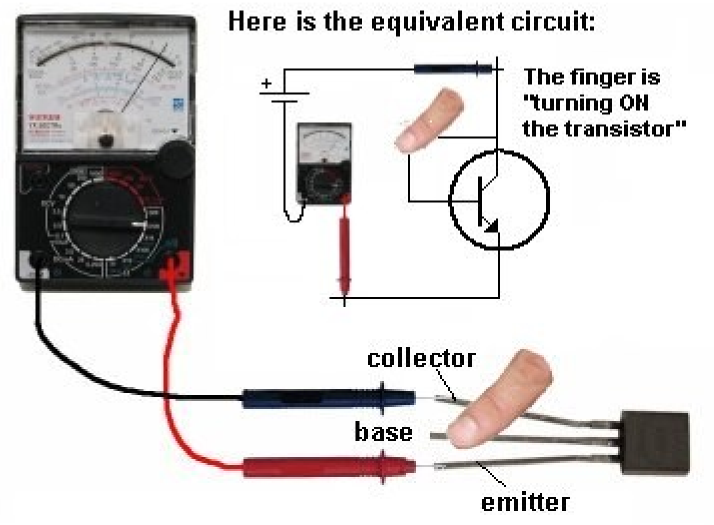
Here is the equivalent circuit:
+
The finger is "turning ON the transistor"
collector
base
emitter

You have found the Collector Base and Emitter

For a PNP transistor, set the meter to "x10k" place the leads on the transistor and when you press hard on the two leads shown in the diagram below, the needle will swing almost full scale.

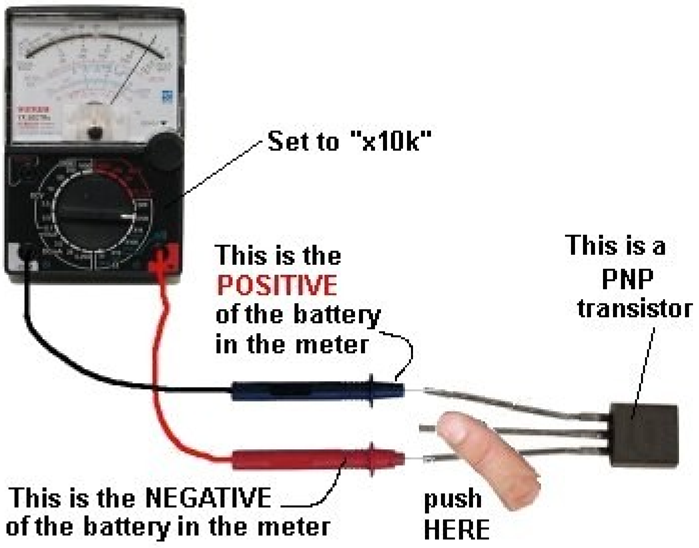
Set to "x10k"
This is the POSITIVE of the battery in the meter
This is a PNP transistor
This is the NEGATIVE of the battery in the meter
push HERE

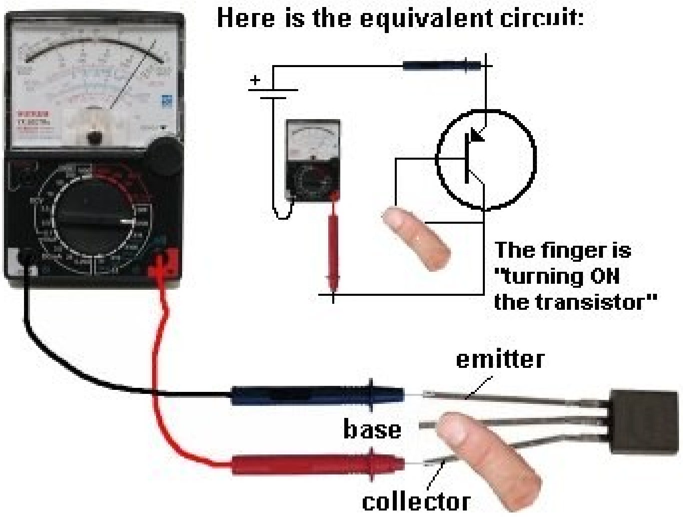
Here is the equivalent circuit:
+
The finger is
"turning ON
the transistor"
emitter
base
collector

You have found the Collector Base and Emitter

SIMPLEST TRANSISTOR TESTER

The simplest transistor tester uses a 9v battery, 1k resistor and a LED (any colour). Keep trying a transistor in all different combinations until you get one of the circuits below. When you push on the two leads, the LED will get brighter.

The transistor will be NPN or PNP and the leads will be identified:

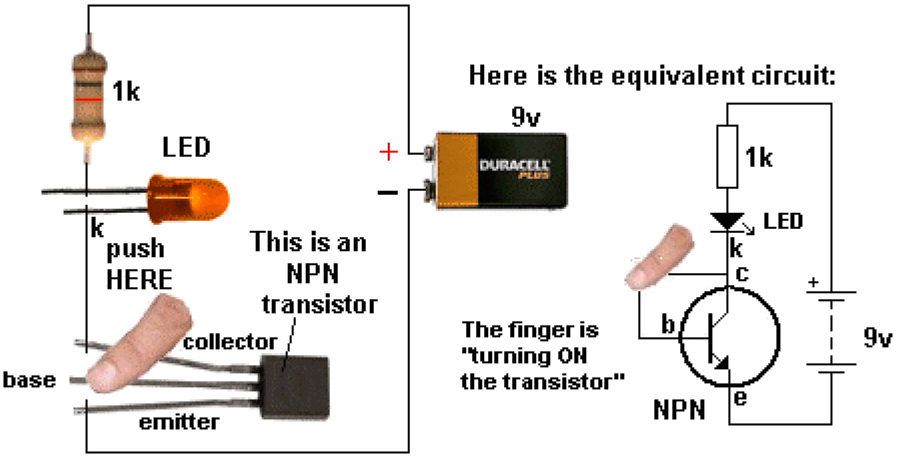
1k
LED
k
push
HERE
This is an
NPN
transistor
collector
base
emitter
+
–
Here is the equivalent circuit:
9v
DURACELL
PLUS
1k
LED
k
c
b
e
+
9v
The finger is
"turning ON
the transistor"
NPN

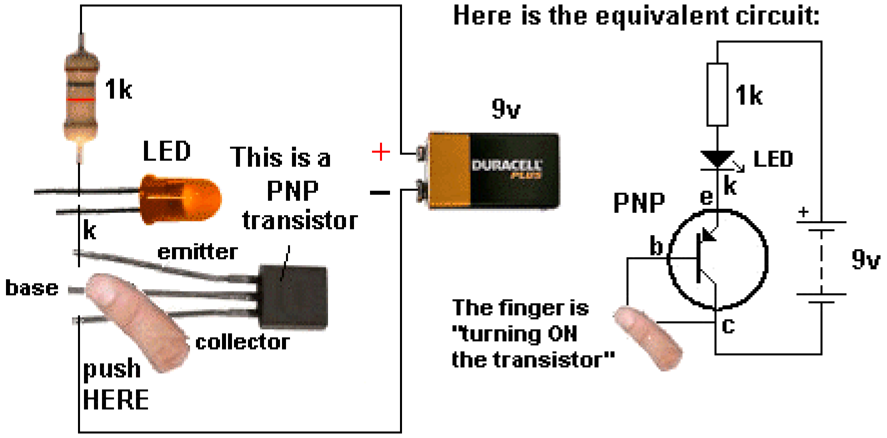
1k
LED
This is a
PNP
transistor
k
emitter
base
collector
push
HERE
+
–
Here is the equivalent circuit:
9v
DURACELL
PLUS
1k
LED
k
PNP
e
b
c
+
9v
The finger is
"turning ON
the transistor"

The leads of some transistors will need to be bent so the pins are in the same positions as shown in the diagrams. This helps you see how the transistor is being turned on. This works with NPN, PNP transistors and Darlington transistors.

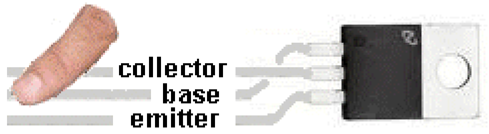

NPN transistor

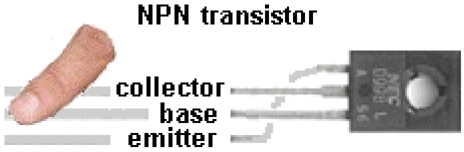

NPN transistor

HEATSINKING

Heat generated by current flowing between the collector and emitter leads of a transistor causes its temperature to rise. This heat must be conducted away from the transistor otherwise the rise may be high enough to damage the P-N junctions inside the device. Power transistors produce a lot of heat, and are therefore usually mounted on a piece of aluminium with fins, called a **HEATSINK**.
This draws heat away, allowing it to handle more current. Low-power signal transistors do not normally require heat sinking. Some transistors have a metal body or fin to connect to a larger heatsink. If the transistor is connected to a heatsink with a mica sheet (mica washer), it can be damaged or cracked and create a short-circuit. (See Testing Mica Washers). Or a small piece of metal may be puncturing the mica. Sometimes white compound called **Heatsink Compound** is used to conduct heat through the mica. This is very important as mica is a very poor conductor of heat and the compound is needed to provide maximum thermal conduction.

TRANSISTOR FAILURE

Transistor can fail in a number of ways. They have forward and reverse voltage ratings and once these are exceeded, the transistor will ZENER or conduct and may fail. In some cases a high voltage will "puncture" the transistor and it will fail instantly. In fact it will fail much faster via a voltage-spike than a current overload.

It may fail with a "short" between any leads, with a collector-emitter short being the most common. However failures will also create shorts between all three leads.
A shorted transistor will allow a large current to flow, and cause other components to heat up.
Transistors can also develop an open circuit between base and collector, base and emitter or collector and emitter.
The first step in identifying a faulty transistor is to check for signs of overheating. It may appear to be burnt, melted or exploded. When the equipment is switched off, you can touch the transistor to see if it feels unusually hot. The amount of heat you feel should be proportional to the size of the transistor's heat sink. If the transistor has no heat sink, yet is very hot, you can suspect a problem.
DO NOT TOUCH A TRANSISTOR IF IT IS PART OF A CIRCUIT THAT CARRIES 240VAC. Always switch off the equipment before touching any components.

TRANSISTOR REPLACEMENT

If you can't get an exact replacement, refer to a transistor substitution guide to identify a near equivalent.

The important parameters are:
- Voltage
- Current
- Wattage
- Maximum frequency of operation

The replacement part should have parameters equal to or higher than the original.

Points to remember:
- Polarity of the transistor i.e. PNP or NPN.
- At least the same voltage, current and wattage rating.
- Low frequency or high frequency type.
- Check the pinout of the replacement part
- Use a desoldering pump to remove the transistor to prevent damage to the printed circuit board.
- Fit the heat sink.

- Check the mica washer and use heat-sink compound
- Tighten the nut/bolt - not too tight or too loose.
- Horizontal output transistors with an integrated diode should be replaced with the same type.

DIGITAL TRANSISTORS

There is no such thing as a DIGITAL TRANSISTOR, however some transistors are available with built-in resistors between base and emitter (to save space on the board) and these transistors are often used in digital circuits. The transistor will amplify analogue signals but when the signal is 0v then immediately goes to a voltage above 0.7v, the transistor is in a DIGITAL CIRCUIT and the transistor is called a DIGITAL TRANSISTOR.   It is tested like an ordinary transistor but the low value resistor between base and emitter will produce a low reading in both directions.

DARLINGTON TRANSISTORS

A DARLINGTON TRANSISTOR is two transistors in a single package with three leads. They are internally connected in cascade so the gain of the pair is very high. This allows a very small input signal to produce a large signal at the output. They have three leads (Base, Collector and Emitter and can be PNP or NPN) and are equivalent to the leads of a standard individual transistor, but with a very high gain. The second advantage of a Darlington Transistor is its high input impedance. It puts very little load on the previous circuit.
Some Darlington transistors have a built-in diode and/or built-in resistor and this will produce a low reading in both directions between the base and emitter leads.

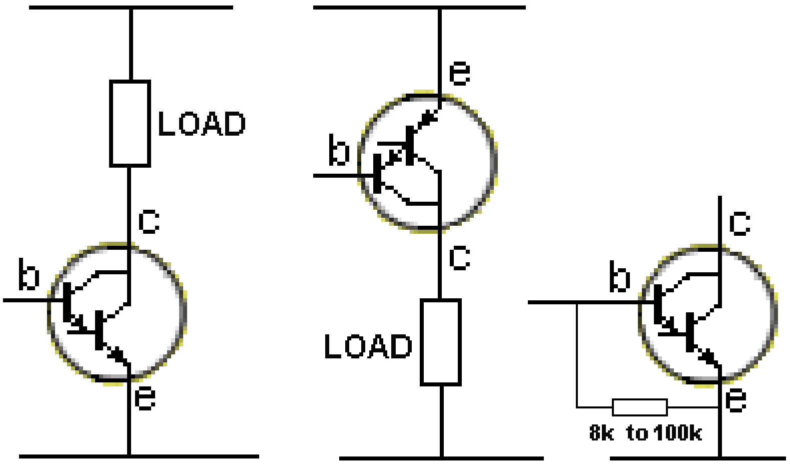

NPN Darlington   PNP Darlington

Darlington transistors are tested the same as an ordinary transistor and a multimeter will produce about the same deflection, even though you will be measuring across two junctions, (and a base-emitter resistor is present).

HORIZONTAL OUTPUT TRANSISTORS, SWITCH-MODE TRANSISTORS, FLYBACK TRANSISTORS, POWER TRANSISTORS, VERTICAL TRANSISTORS . . . .

These are all names given to a transistor when it is used in a particular circuit. ALL these transistors are the same for testing purposes.
We are not testing for gain, maximum voltage, speed of operation or any special feature. We are just testing to see if the transistor is completely faulty and SHORTED.
A transistor can have lots of other faults and the circuit **using the transistor** is the best piece of TEST EQUIPMENT as it is detecting the fault.

TESTING MOSFETs and FETs

MOSFETs and **JFETs** are all part of the **FET** family.
MOSFET stands for **Metal Oxide Semiconductor Field Effect Transistor.**
FETs operate exactly the same as a "normal" transistor except they have different names for the input and output leads and the voltage between the gate and the source has to between 2v to 5v for the device to turn on fully. A FET requires almost NO CURRENT into the Gate for it to turn on and when it

does, the voltage between drain and source is very low (only a few mV). This allows them to pass very high currents without getting hot. There is a point where they start to turn on and the input voltage must rise higher than this so the FET turns on FULLY and does not get hot.
Field Effect Transistors are difficult to test with a multimeter, but "fortunately" when a power **MOSFET** blows, it is completely damaged. All the leads will show a short circuit. 99% of bad **MOSFETs** will have GS, GD and DS shorted.
The following symbols show some of the different types of MOSFETs:

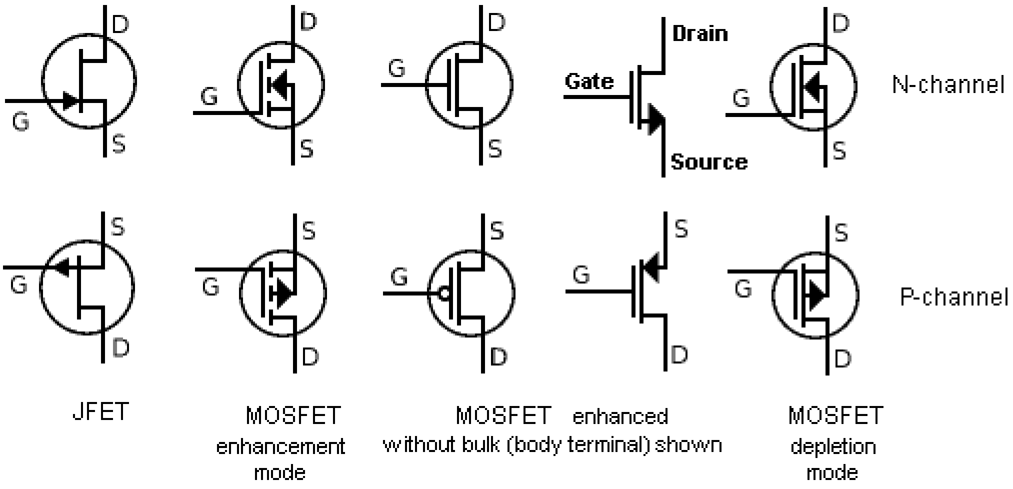
D
G
S
Drain
Gate
Source
N-channel
P-channel
JFET
MOSFET
enhancement mode
MOSFET enhanced
without bulk (body terminal) shown
MOSFET
depletion mode

Most **MOSFET** transistors cannot be tested with a multimeter. This due to the fact that the Gate needs 2v - 5v to turn on the device and this voltage is not present on the probes of either meter set to any of the ohms ranges.
You need to build the following Test Circuit:

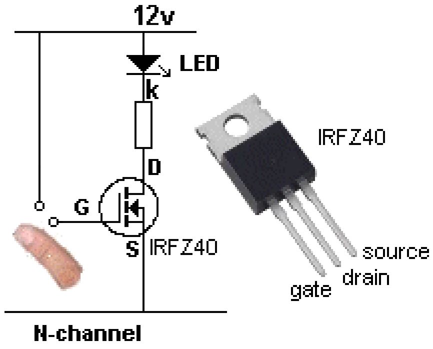
12v
LED
k
D
G
S
IRFZ40
IRFZ40
source
drain
gate
N-channel

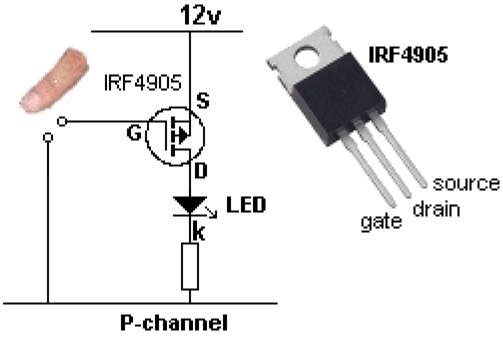
12v
IRF4905
S
G
D
LED
k
IRF4905
source
drain
gate
P-channel

Touching the Gate will increase the voltage on the Gate and the MOSFET will turn on and illuminate the LED. Removing your finger will turn the LED off.

SILICON CONTROLLED RECTIFIERs (SCR)

The **Silicon Controlled Rectifier** (SCR) is a semiconductor device that is a member of a family of control devices known as **Thyristors**. It is a 3-leaded device and when a small current enters the Gate, the **thyristor** turns on. AND STAYS ON. It only conducts current between Anode and Cathode in one direction and it is mainly only used in DC circuits. When it is used with AC, it will only conduct for a maximum of half the cycle.

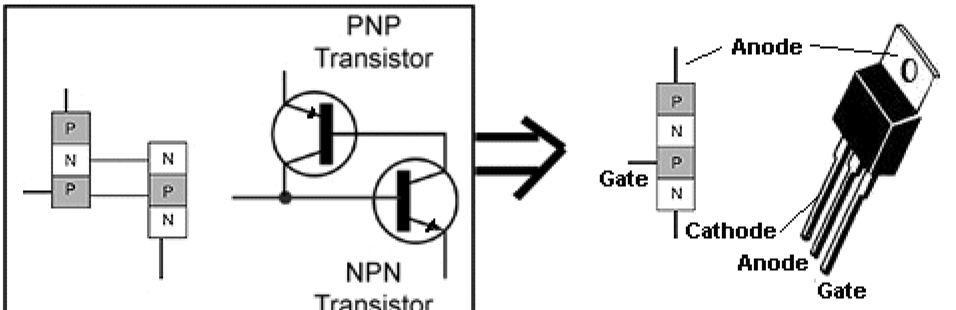
PNP
Transistor
P
N
P
N
P
N
NPN
Transistor
Anode
P
N
P
N
Gate
Cathode
Anode
Gate

THYRISTOR or SCR

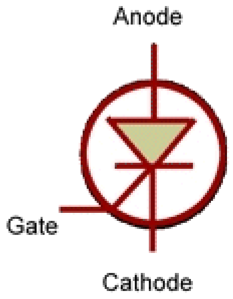
Anode
Gate
Cathode

SCR
Silicon Controlled Rectifier

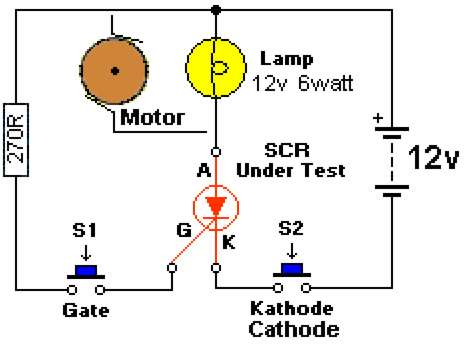
270R
Motor
Lamp
12v 6watt
+
12v
SCR
Under Test
A
G
K
S1
S2
Gate
Kathode
Cathode

SCR TESTER

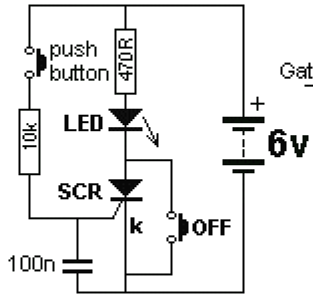
push
button
470R
10k
LED
SCR
k
OFF
100n
+
6v

A

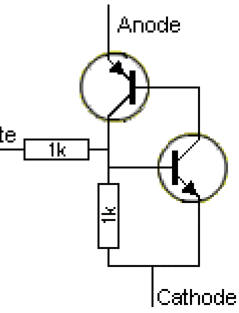
Anode
Gate
1k
1k
Cathode

SCR WITH TRANSISTORS

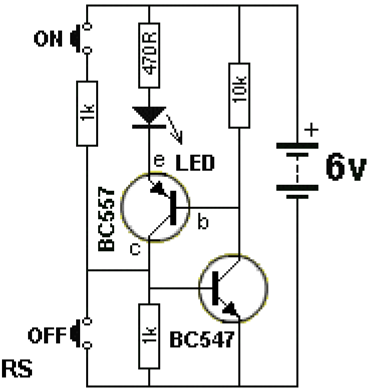
ON
470R
1k
10k
e
LED
BC557
b
c
+
6v
OFF
1k
BC547

B

To understand how an SCR "latches" when the gate is provided with a small current, we can replace it with two transistors as shown in diagram B above. When the ON button is pressed, the BC547 transistor turns on. This turns ON the BC557 and it takes over from the action of the switch.
To turn the circuit off, the OFF button removes the voltage from the base of the BC547.

Testing an SCR

An **SCR** can be tested with some multimeters but a minimum current Anode-to-Cathode is needed to keep the device turned on. Some multimeters do not provide this amount of current and the **SCR Tester** circuit above is the best way to test these devices.
Shorted SCRs can usually be detected with an ohmmeter check (SCRs usually fail shorted rather than open).
Measure the anode-to-cathode resistance in both the forward and reverse direction; a good SCR should measure near infinity in both directions.
Small and medium-size SCRs can also be gated ON with an ohmmeter (on a digital meter use the Diode Check Function). Forward bias the SCR with the ohmmeter by connecting the black (-) lead to the anode and the red (+) lead to the cathode (because the + of the battery is connected to the negative lead, in most analogue multimeters). Momentarily touch the gate lead to the anode while the probes are still touching both leads; this will provide a small positive turn-on voltage to the gate and the cathode-to-anode resistance reading will drop to a low value. Even after removing the gate voltage, the SCR will stay conducting. Disconnecting the meter leads from the anode or cathode will

cause the SCR to revert to its non-conducting state.
When making the above test, the meter impedance acts as the SCR load. On larger SCRs, it may not latch ON because the test current is not above the SCR holding current.

Using the SCR Tester

Connect an SCR and press Switch2. The lamp should not illuminate. If it illuminates, the SCR is around the wrong way or it is faulty.
Keep Switch 2 PRESSED. Press Sw1 very briefly. The lamp or motor will turn ON and remain ON.
Release Sw 2 and press it again. The Lamp or motor will be OFF.

TRIACs

A triac is a bidirectional, three-terminal dual, back-to-back thyristor (SCR) switch. This device will conduct current in both directions when a small current is constantly applied to the Gate.
If the gate is given a small, brief, current during any instant of a cycle, it will remain triggered during the completion of the cycle until the current though the Main Terminals drops to zero.
This means it will conduct both the positive and negative half-cycles of an AC waveform. If it is tuned on (with a brief pulse) half-way up the positive waveform, it will remain on until the wave rises and finally reaches zero. If it is then turned on (with a brief pulse) part-way on the negative wave, the result will be pulses of energy and the end result will be about 50% of the full-energy delivered at a rate of 100 times per second for a 50HZ supply.
TRIACs are particularly suited for AC power control applications such as motor speed control, light dimmers, temperature control and many others.

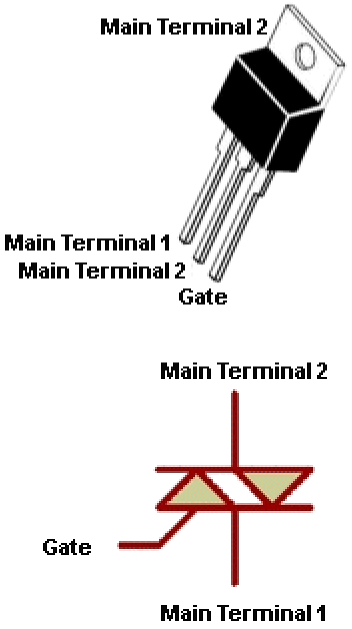

TRIAC (2 SCR's)
TRIode for Alternating Current

Using the TRIAC Tester

Connect a TRIAC and press Switch2. The lamp should not illuminate. If it illuminates, the TRIAC is faulty.
Keep Switch 2 PRESSED. Press Sw1 very briefly. The lamp or motor will turn ON and remain ON. If the lamp does not turn on, reverse the TRIAC as the current into the gate must produce a slight voltage between **Gate** and **Main Terminal 1**.
Release Sw 2 and press it again. The Lamp or motor will be OFF.

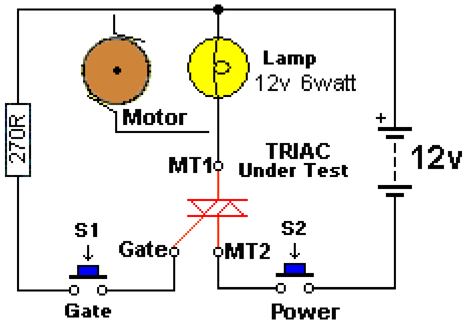

TRIAC TESTER

MICA WASHERS AND INSULATORS

Plastic insulating sheets (washers) between a transistor and heatsink are most often made from mica but some are plastic and these get damaged over a period of time, turn dark and become cracked. The plastic eventually becomes carbonized and conducts current and can affect the operation of the appliance. You can see the difference between a mica sheet (washer) and plastic by looking where it extends from under the transistor. Replace all plastic insulators as they eventually fail.

SPARK GAPS

Some TV's and monitors with a CRT (picture tube), have spark gaps either on the socket at the end of the tube or on the chassis.
These can consist of two wires inside a plastic holder or a glass tube or special resistive device.
The purpose of a spark gap is to take any flash-over (from inside the tube), to earth. This prevents damage to the rest of the circuit.
However if the tube constantly flashes over, a carbon track builds up between the wires and effectively reduces the screen voltage. This can cause brightness and/or focus problems. Removing the spark-gap will restore the voltage.
These are not available as a spare component and it's best to get one from a discarded chassis.

CO-AX CABLES

Co-Ax cables can produce very high losses and it seems impossible that a few metres of cable will reduce the signal. The author has had a 3 metre cable reduce the signal to "snow" so be aware that this can occur. Faults can also come from a splitter and/or balun as well as dirty plugs and sockets. This can result in very loud bangs in the sound on digital reception.

TESTING EARTH LEAKAGE DETECTORS or Residual Current Devices or Ground Fault Circuit Interrupters or GFCI

An Earth Leakage Detector or Sensor is a circuit designed to continuously monitor the imbalance in the current in a pair of load carrying conductors.
These two conductors are normally the Active and Neutral. Should the imbalance current reach 30mA the sensor will "trip" and remove the voltage (and current) from the line being monitored.
Some detectors will trip at 15mA.
You cannot alter the sensitivity of the device however there are a number of faults in these devices that can be fixed.
In some devices the contact pressure for the 10Amp or 15 Amp contacts is very weak and they arc and produce an open circuit. The result is this: When you press the rest button, power is not restored to the output.
Clean the contacts with a small file and bend the metal strips to the contacts so they make a very strong contact.
The other fault is the trip mechanism.
The magnetism from the coil does not allow the pin to move and "trip" the contacts. It may be due to a small metal filing or the pin not moving freely enough.
All good Earth Leakage Detectors have a TEST BUTTON. This connects a resistor between the active line and earth so that 15mA or 30mA flows.
The detector should trip immediately. Make sure the trigger mechanism trips when the test button is pressed.
None of the electronics in the detector can be replaced however you can test the mechanical operation and the pressure on the contacts when the unit is removed from the power. Do not work on the device

when it is connected to the mains.

TESTING CELLS AND BATTERIES

There is an enormous number of batteries and cells on the market and a number of "battery testers." Instead of buying a battery tester that may give you a false reading, here is a method of testing cells that is guaranteed to work.
There are two types of cell: a **rechargeable** cell and a non rechargeable cell.

The easiest way to test a **rechargeable** cell is to put a group of them in an appliance and use them until the appliance "runs down" or fails to work. If you consider the cells did not last very long, remove them and check the voltage of each cell. The cell or cells with the lowest voltage will be faulty. You can replace them with new cells or good cells you have in reserve.
There is no other simple way to test a rechargeable cell.
You cannot test the "current of a cell" by using an ammeter. A rechargeable cell can deliver 10 amps or more, even when nearly discharged and you cannot determine a good cell for a faulty cell.

Dry cells are classified as "non-rechargeable" cells.
DRY CELLS and MANGANESE CELLS are the same thing. These produce 1.5v per cell (manganese means the Manganese Dioxide depolariser inside the cell. All "dry cells" use manganese dioxide).
ALKALINE CELLS produce between 2 - 10 times more energy than a "dry cell" and produce 1.5v per cell.
Alkaline cells can fail for no reason at any stage in their life and are not recommended for emergency situations.
The output voltage of some Alkaline cells can fall to 0.7v or 0.9v for not apparent reason.
There are lots of other cells including "button cells," hearing-aid cells, air cells, and they produce from 1.2v to 3v per cell.

Note:
Lithium cells are also called "button cells" and they produce 3v per cell.
Lithium cells are non-rechargeable (they are generally called "button cells") but some Lithium cells can be recharged. These are Lithium-ion cells and generally have a voltage of 3.6v. Some Lithium-ion cells look exactly like 3v Lithium cells, so you have to read the data on the cell before charging.

You cannot test the voltage of a cell and come to any conclusion as to the age of the cell or how much energy remains. The voltage of a cell is characteristic to the chemicals used and the actual voltage does not tell you its condition.
Some "dry cells" deliver 1.5v up to the end of their life whereas others drop to about 1.1v very quickly.

Once you know the name of the cell that drops to 1.1v, avoid them as the operation of the equipment "drops off" very quickly.

However if you have a number of different cells and need to know which ones to keep, here's the solution:

1. Check the voltage and use those with a voltage above 1.1v
2. Next, select 500mA or 10A range on a meter and place the probes on a cell. For a AAA or AA cell, the current should be over 500mA and the needle will swing full scale very quickly.
Keep the testing short as you are short-circuiting the cell but it is the only way to determine the internal impedance of the cell and this has a lot to do with its stage-of-charge.
This will give you a cell with a good terminal voltage and a good current capability.

This also applies to button cells, but the maximum current they will deliver will be less.
If you want to get the last of the energy out of a group of cells they can be used in the following circuits:

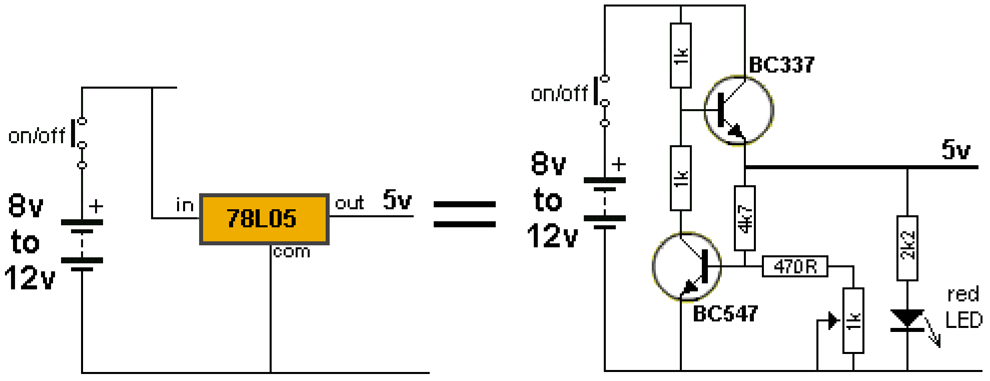

1mV drop @100mA @10v supply

10mV drop @100mA @10v supply

TESTING PIEZO DIAPHRAGMS and PIEZO BUZZERS

There are two types of piezo devices that produce a sound.
They are called **PIEZO DIAPHRAGMS** and **PIEZO BUZZERS**.
A **piezo diaphragm** consists of two metal plates with a ceramic material between. The ceramic expands and contracts when an alternating voltage is placed on the two plates and this causes the main plate to "dish" and "bow."
This creates a high-pitched sound. There are no other components inside the case and it requires an AC voltage of the appropriate frequency to produce a sound.
A **piezo buzzer** has a transistor and coil enclosed and when supplied with a DC voltage, the buzzer produces a sound.
Both devices can look exactly the same and the only way to tell them apart is by connecting a 9v battery. One device may have "+' and "-" on the case to indicate it is a piezo buzzer, but supplying 9v will make the buzzer produce a sound while the piezo diaphragm will only produce a "click."

PIEZO DIAPHRAGM
PIEZO BUZZER

A piezo diaphragm will produce a click
when connected to 9v DC.
A piezo buzzer will produce a tone when
connected to a DC voltage.

TESTING A CIRCUIT

Whenever you test a circuit, the TEST EQUIPMENT puts "a load" or "a change" on it.
It does not matter if the test equipment is a multimeter, Logic Probe, CRO, Tone Injector or simply a LED and resistor.
There are two things you need to know.
1. The IMPEDANCE of the circuit at the location you are testing, and
2. The amount of load you are adding to the circuit via the test equipment.

There is also one other hidden factor. The test equipment may be injecting "hum" due to its leads or the effect of your body at absorbing hum from the surroundings or the test equipment may be connected to the mains.
These will affect the reading on the test equipment and also any output of the circuit.
Sometimes the test equipment will prevent the circuit from working and sometimes it will just change the operating conditions slightly. You have to be aware of this.
The last section of this eBook covers High and Low Impedance and understanding impedance is something you need to know.
The point to note here is the fact that the equipment (and the reading) can be upset by hum and resistance/capacitance effects of test equipment. This is particularly critical in high impedance and high

frequency circuits.

TESTING INTEGRATED CIRCUITS (IC's)

Integrated Circuits can be tested with a LOGIC PROBE. A Logic Probe will tell you if a line is HIGH, LOW or PULSING.
Most logic circuits operate on 5v and a Logic Probe is connected to the 5v supply so the readings are accurate for the voltages being tested.
A Logic Probe can also be connected to a 12v CMOS circuit.
You can make your own Logic Probe and learn how to use it from the following link:

http://www.talkingelectronics.com/projects/LogicProbeMkIIB/LogicProbeMk-IIB.html

LOGIC PROBE MKII

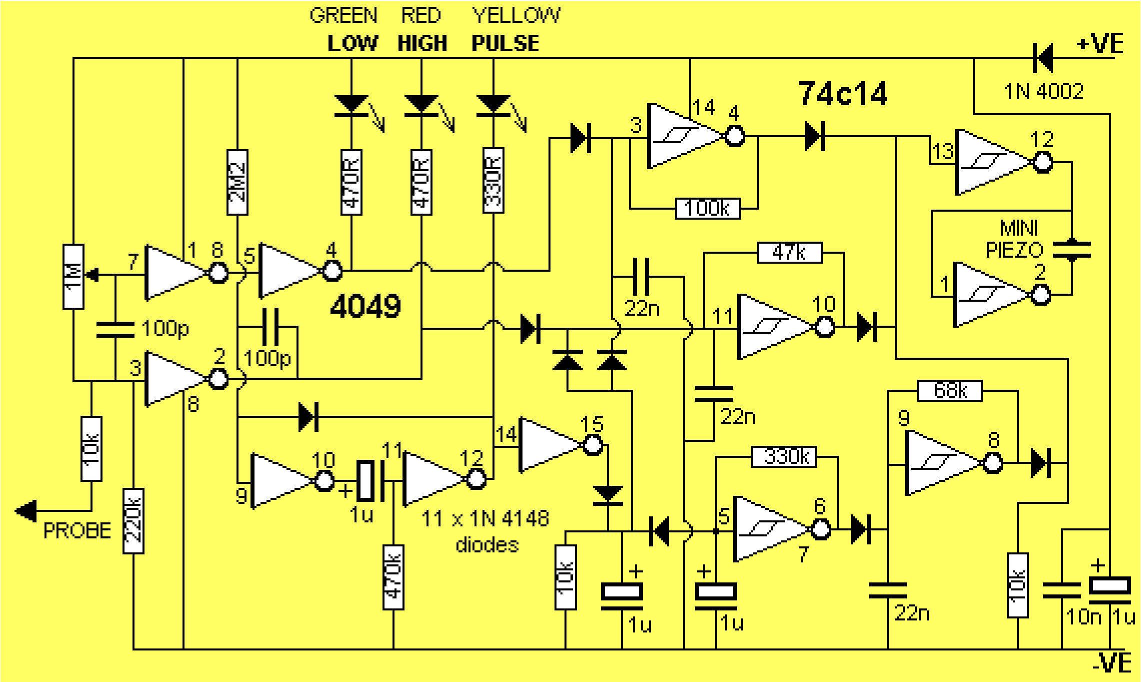
GREEN
RED
YELLOW
LOW
HIGH
PULSE
+VE
1N 4002
74c14
2M2
470R
470R
330R
1M
100p
4049
100p
100k
47k
22n
22n
MINI PIEZO
68k
330k
10k
PROBE
220k
1u
470k
11 x 1N 4148 diodes
10k
1u
1u
22n
10k
10n
1u
-VE

LOGIC PROBE with PULSE

This is a very simple transistor circuit to provide HIGH-LOW-PULSE indication for digital circuits. It can be built for less than $5.00 on a piece of matrix board or on a small strip of copper clad board if you are using surface mount components. The probe will detect a HIGH at 3v and thus the project can be used for 3v, 5v and CMOS circuits.

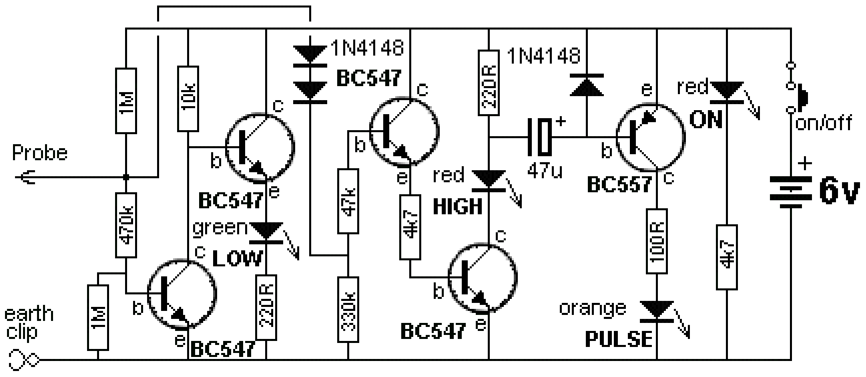
1N4148
BC547
1N4148
red
ON
on/off
6v
1M
10k
c
b
e
Probe
BC547
green
LOW
47k
220R
b
e
c
red
HIGH
47u
BC557
100R
4k7
4k7
470k
earth
clip
1M
BC547
220R
330k
BC547
orange
PULSE

LOGIC PROBE using CD4001 and CD4011

Here is a simple Logic Probe using a single chip. The circuits have been designed for the **CD4001** CMOS quad NOR gate and **CD4011** CMOS NAND gate. The output has an active buzzer that produces a beep when the pulse LED illuminates (the buzzer is not a piezo-diaphragm but an active buzzer containing components).

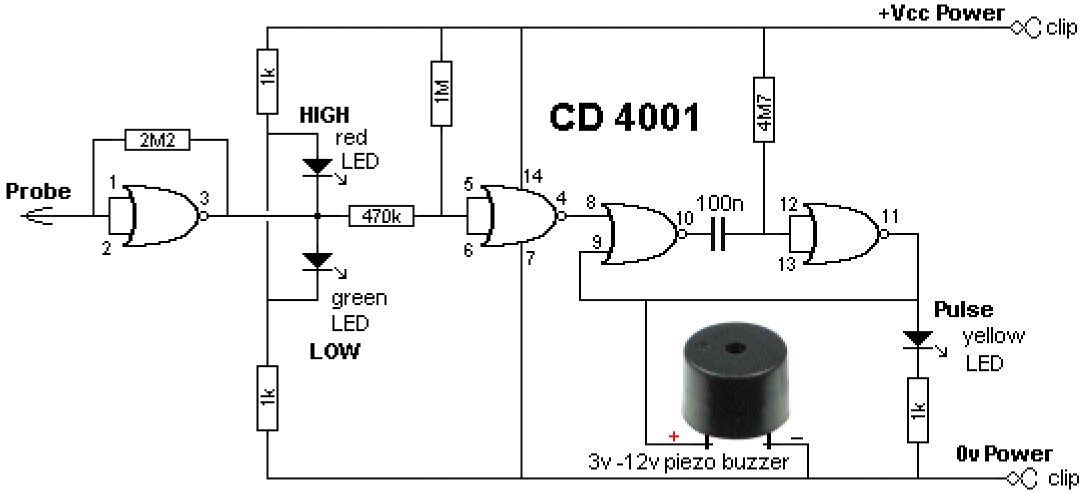
+Vcc Power
clip
1k
1M
4M7
CD 4001
2M2
HIGH
red
LED
Probe
1
2
3
470k
5
6
14
4
7
8
9
10
100n
12
13
11
green
LED
LOW
Pulse
yellow
LED
1k
1k
+
-
3v -12v piezo buzzer
0v Power
clip

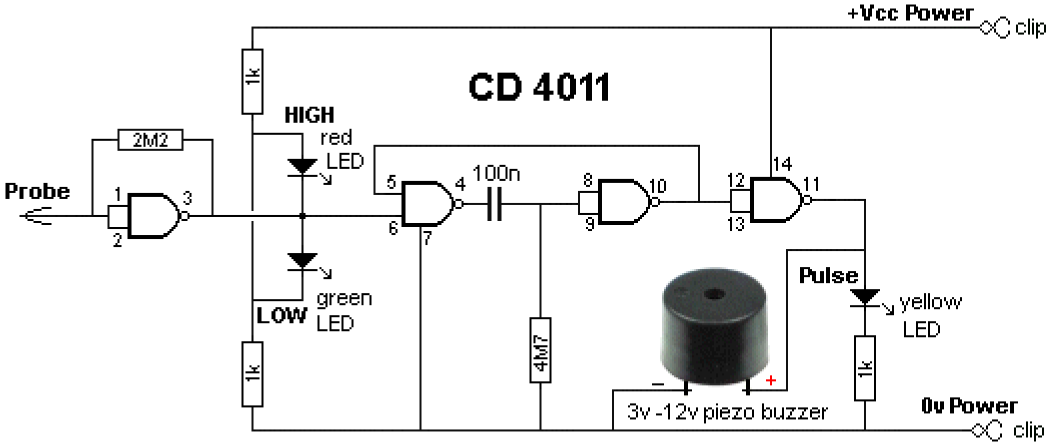
+Vcc Power
clip
1k
CD 4011
HIGH
red
LED
2M2
Probe
1
2
3
5
6
4
7
100n
8
9
10
12
13
14
11
green
LED
LOW
Pulse
yellow
LED
4M7
1k
1k
-
+
3v -12v piezo buzzer
0v Power
clip

SUPER PROBE MkII has 20 different features including a Logic Probe, capacitance tester, Inductance tester, and more.

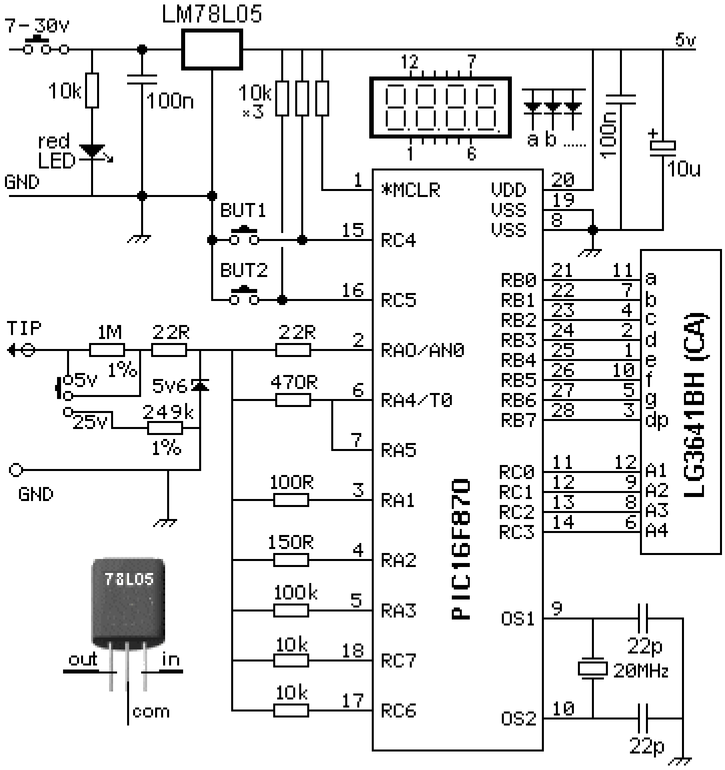
7-30v
LM78L05
5v
10k
100n
10k ×3
red LED
GND
BUT1
BUT2
*MCLR
RC4
RC5
RA0/AN0
RA4/T0
RA5
RA1
RA2
RA3
RC7
RC6
VDD
VSS
RB0
RB1
RB2
RB3
RB4
RB5
RB6
RB7
RC0
RC1
RC2
RC3
OS1
OS2
PIC16F870
LG3641BH (CA)
TIP
1M
1%
22R
5v
25v
5v6
249k
470R
100R
150R
100k
22p
20MHz
10u
78L05
out
in
com

SUPER PROBE MkII Circuit

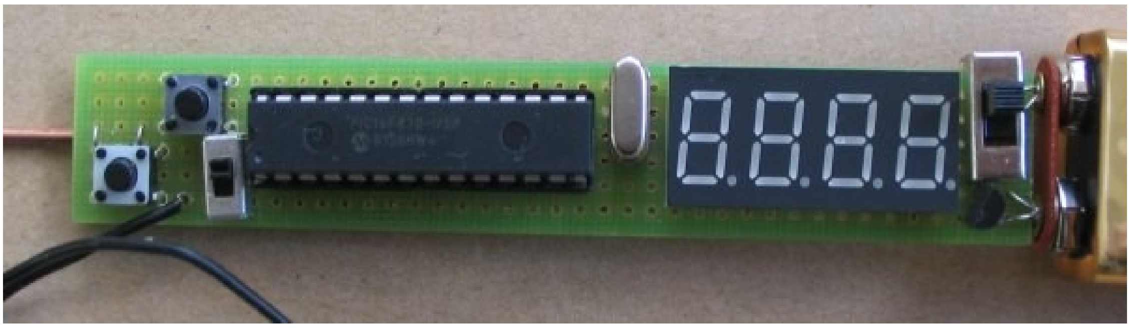

SUPER PROBE MkII

To test an IC, you need a circuit diagram with waveforms. These diagrams will show the signals and are very handy if a CRO (cathode ray Oscilloscope) is used to diagnose the problem. The CRO will reproduce the waveform and prove the circuit is functioning correctly.
A Logic Probe will just show activity and if an output is not producing a "pulse" or "activity," you should check the power to the IC and test the input line.
It is beyond the scope of this eBook to explain how to diagnose waveforms, however it is important to know if signals are entering and exiting an IC and a Logic Probe is designed for this.

SIGNAL INJECTOR

This circuit is rich in harmonics and is ideal for testing amplifier circuits. To find a fault in an amplifier, connect the earth clip to the 0v rail and move through each stage, starting at the speaker. An increase in volume should be heard at each preceding stage. This Injector will also go through the IF stages of radios and FM sound sections in TV's.

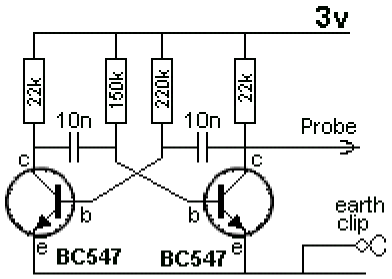
3v
22k
150k
220k
22k
10n
10n
Probe
c
b
e
BC547
BC547
earth clip

TESTING AUDIO AMPLIFIERS and AUDIO IC's

The **Super Probe MII** described above has a "noise" function and a tone function that allows you to inject a signal into an audio stage, amplifier (made from discrete components) or an audio chip, and detect the output on a speaker.
Audio stages are very difficult to work-with if you don't have a TONE GENERATOR or SIGNAL INJECTOR.
The signals are very small and not detected by a multimeter.
You can start anywhere in an amplifier and when a tone is heard, you can keep probing until the signal is not present or louder. From this you can work out which way the signal is travelling.
A Signal Injector is very handy for finding shorts and broken wires in switches, plugs, sockets and especially leads to headphones.
You can determine the gain of a stage (amplification) by probing before and after a chip or transistor and listen for the relative increase in volume from the speaker.
You can also use your finger to produce "hum" or "buzz" if a **Signal Injector** is not available.
Nearly all audio problems are plugs, sockets and cracks in the PC board, but finding them takes a lot of time and skill.

TESTING IC's - also called "CHIPS"

An Integrated Circuit is also called a "chip." It might have 8 pins or as many as 40.
Some chips are ANALOGUE. This means the input signal is rising and falling slowly and the output produces a larger version of the input.
Other chips are classified as DIGITAL and the input starts at 0v and rises to rail voltage very quickly. The output does exactly the same - it rises and falls very quickly.
You might think the chip performs no function, because the input and output voltage has the same value, but you will find the chip may have more than one output and the others only go high after a number of clock-pulses on the input, or the chip may be outputting when a combination of inputs is recognised or the output may go HIGH after a number of clock pulses.

ANALOGUE CHIPS (also see above)

Analogue chips are AUDIO chips or AMPLIFIER chips.
To test these chips you will need three pieces of test equipment:
1. A multimeter - this can be digital or analogue.
2. A Signal Injector
3. A Mini Bench Amplifier.

The Mini Bench Amplifier is available as a kit.

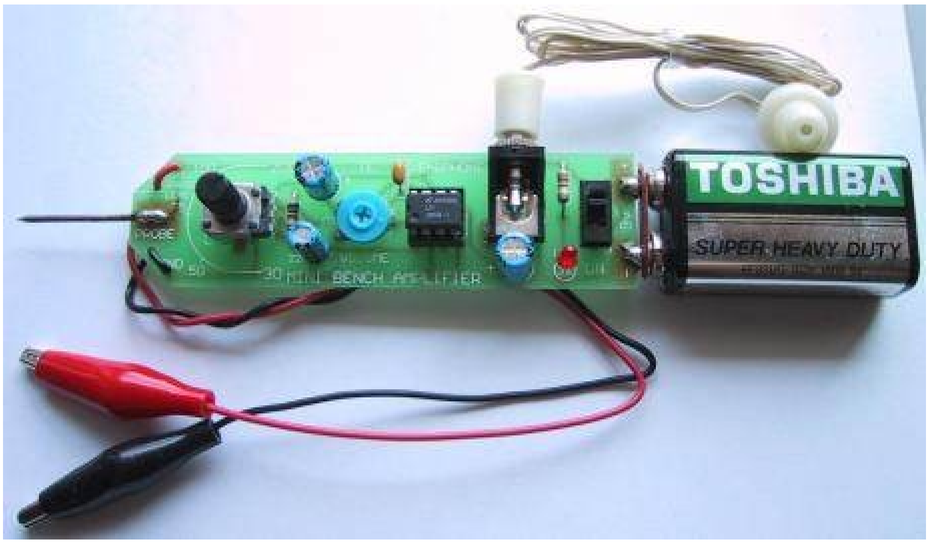

MINI BENCH AMPLIFIER

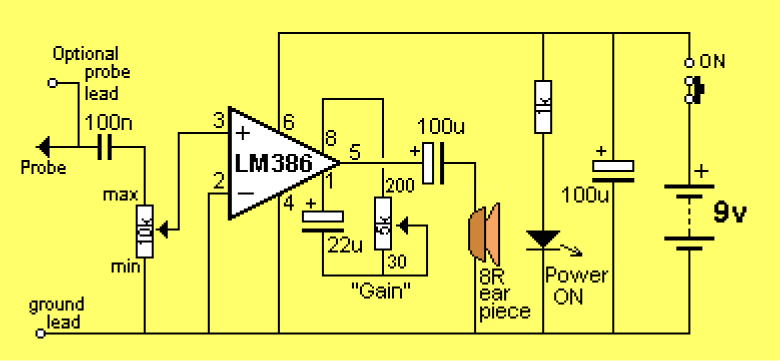

MINI BENCH AMPLIFIER CIRCUIT

Start by locating the power pin with a multimeter.
If the chip is receiving a voltage, you can use the Mini Bench Amplifier to detect an output.
Connect the Ground Lead of the Mini Bench Amplifier to 0v and touch the Probe tip on each of the pins.

You will hear faint audio on the Input pin and very loud audio on the Output pin.
If no input is detected, you can use a Signal Injector to produce a tone.
Connect the clip of the Signal Injector to 0v and the probe to the input pin of the amplifier chip. At the same time, connect the Mini Bench Amplifier to the output pin and you will hear a very loud tone.
These pieces of test equipment can also be used to diagnose an amplifier circuit constructed with individual components.
Amplifier circuits using discrete components are very hard to trouble-shoot and these pieces of test equipment make it very easy.

DIGITAL CHIPS

It is always best to have data on the chip you are testing, but if this is not available, you will need three pieces of equipment:
1. A multimeter - this can be digital or analogue.
2. A Logic Probe,
3. A logic Pulser.
Firstly test the chip to see if power is being delivered. This might be anything from 3v3 to 15v.
Place the negative lead of the multimeter on the earth rail of the project - this might be the chassis, or the track around the edge of the board or some point that is obviously 0v.
Try all the pins of the chip and if you get a reading, the chip will have "supply."
Identify pin 1 of the chip by looking for the "cut-out" at the end of the chip and you may find a small dimple below the cut-out (or notch). This is pin 1 and the "power pin" can be directly above or any of the other pins.
Next you need to now if a signal is entering the chip.
For this you will need a LOGIC PROBE.
A Logic Probe is connected to the same voltage as the chip, so it will detect a HIGH and illuminate a red LED.
Connect the Logic Probe and touch the tip of the probe on each pin.
You will not know if a signal is an input or output, however if you get two or more active pins, you can assume one is input and the other is output. If none of the pins are active, you can assume the signal is not reaching this IC.
If only one pin is active, you can assume the chip is called a CLOCK (or Clock Generator). This type of chip produces pulses. If more than two pins are active, you can assume the chip is performing its function and unless you can monitor all the pins at the same time, you don't know what is happening.
This is about all you can do without any data on the chip.

If you have data on the chip, you can identify the input(s) and output(s).
A Logic Probe on each of these pins will identify activity.
A Logic Probe has 3 LEDs. Red LED indicates a HIGH, Green indicates a LOW and Orange indicates a PULSE (activity).
Some Logic Probes include a piezo and you can hear what is happening, so you don't take your eyes off the probe-tip.
It is important not to let the probe tip slip between the pins and create a short-circuit.

LOGIC PULSER

If you have a board or a single chip and want to create activity (clock pulses), you can use a Logic Pulser. This piece of test equipment will produce a stream of pulses that can be injected into the clock-line (clock input) of a chip.

You can then use a Logic Probe at the same time on the outputs to observe the operation of the chip.

You can also use the Mini Bench Amplifier to detect "noise" or activity on the inputs and outputs of digital chips.
This only applies if the frequency is in the audio range such as scanning a keyboard or switches or a display.

This is how to approach servicing/testing in a general way. There are thousands of digital chips and if you want to test a specific chip for its exact performance, you will need to set-up a "test-bed."

REMOTE CONTROLS

There are two types of remote control - Infrared and RF. Infrared is used for short-range, line-of-sight for TV's DVD's etc.
A few faults can be fixed, but anything complex needs a new remote control.
Check the batteries and battery-contacts. See if the IR LED is illuminating by focusing it into a digital camera and looking on the screen for illumination.
The only other things are a sticky button, a worn-out button or a crack in the PC board. Water damage is generally too much work to repair.
RF remote controls for cars, garage doors etc need a second working unit to check the power output.
Here is a simple circuit that can be connected to an analog multimeter to detect the signal strength at a very close range:

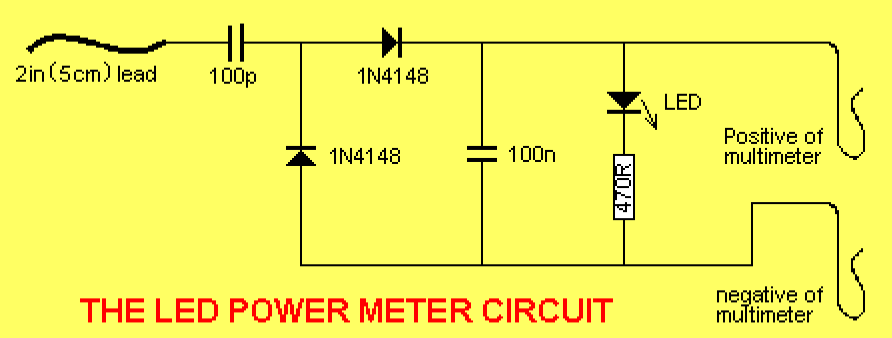
2in (5cm) lead
100p
1N4148
1N4148
100n
LED
470R
Positive of multimeter
negative of multimeter
THE LED POWER METER CIRCUIT

To hear the tone from a transmitter, the **Mini Bug Detector** circuit can be used:

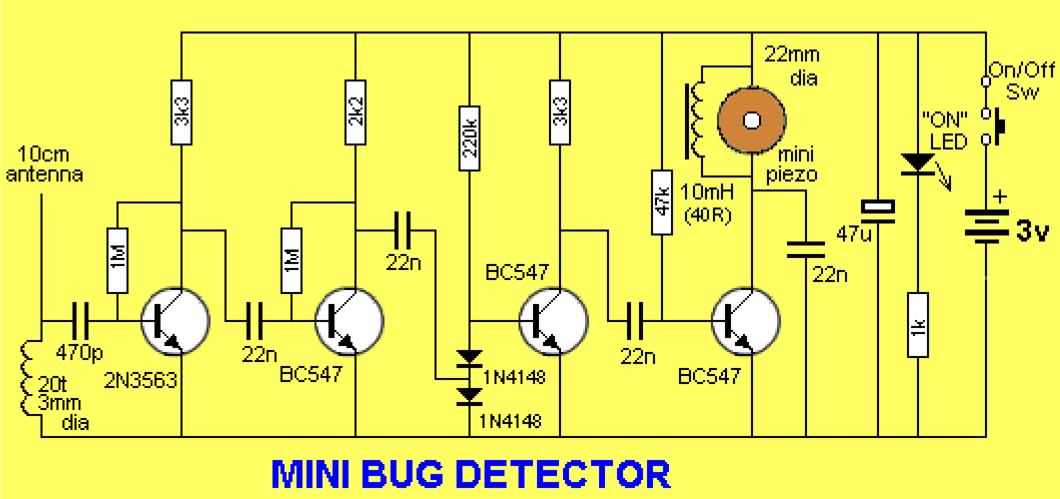
10cm antenna
3k3
1M
470p
20t 3mm dia
2N3563
2k2
1M
22n
BC547
22n
220k
1N4148
1N4148
3k3
BC547
22n
BC547
47k
10mH (40R)
22mm dia
mini piezo
22n
47u
"ON" LED
1k
On/Off Sw
+
3v
MINI BUG DETECTOR

Any further investigation requires a circuit diagram so you can work out what is actually being sent from the transmitter.
Most of the time it is a faulty switch, battery or contacts. Make sure the setting is correct on the "dip switches" and use a working unit to compare all your testing.

TESTING VOLTAGES ON (in) A CIRCUIT

There are basically two different types of circuit.
1. **ANALOGUE CIRCUIT**
An analogue circuit can also be called an AUDIO CIRCUIT and the voltages at different points in a circuit can be measured with a multimeter but the changes (the waveforms) will be quite small or changing at a rapid rate and cannot be detected by a multimeter.

You need a CRO to "see" the signals or a **Signal Injector** to inject a waveform into the circuit and hear the result on the circuit's speaker.

2. **DIGITAL CIRCUIT**
A digital circuit can also be called a "Computer Circuit" or "Logic Circuit" and some of the voltages can be measured with a multimeter (such as supply voltages) but the "signal lines" will be be changing from HIGH to LOW to HIGH very quickly and these signals are detected with a **Logic Probe**.

Here are some circuits with details of how to test the voltages.
Most circuits do not show voltages at various different points and we will explain what to expect on each "stage."

A "STAGE"

A stage is a set of components with an input and output. A "stage" can also be called a "Building Block."
Sometimes it has a capacitor on the input and one on the output.
This means the stage is completely isolated as far as DC is concerned.
The stage has a supply (a DC supply) and it is producing its own voltages on various points on the "stage." It can only process (amplify) "AC." (signals).
Sometimes the stage can be given a name, such as small-signal amplifier, push-pull amplifier or output.
If the stage has a link or resistor connected to a previous stage, the previous stage will have a "DC effect" on the stage. In other words it will be biasing or controlling the voltages on the stage. The stage may be called a "timer" or "delay" or "DC amplifier."

It is important to break every circuit into sections. This makes testing easy. If you have a capacitor at the input and output, you know all the problems lie within the two capacitors.
In a digital circuit (no capacitors) you need to work on each IC (integrated Circuit) and test the input for activity and all the outputs.

Once you have determined if the circuit is Analogue or Digital, or a combination of both, you have to look at the rail voltage and work out the size or amplitude of the voltage or waveform.
This is done before making a test, so your predictions are confirmed.
You will need a **multimeter** (either Digital or Analogue) a **Logic Probe** and a **Signal Injector** (**Tone Generator**). An analogue meter has the advantage that it will detect slight fluctuations of voltage at a test-point and its readings are faster than a digital meter. A digital meter will produce an accurate voltage-reading - so you should have both available.

HIGH IMPEDANCE AND LOW IMPEDANCE

Every point in a circuit has a characteristic called "IMPEDANCE." This has never been discussed before in any text book. That's why it will be new to you.
In other words, every point will be "sensitive to outside noise."
An audio amplifier is a good example. If you put your finger on the active input, it will produce hum or buzz in the speaker. This is because it is a HIGH IMPEDANCE line or high impedance section of the circuit.
The same applies to every part in a circuit and when you place Test Equipment on a line for testing purposes, the equipment will "upset" the line. It may be very slight but it can also alter the voltage on the point CONSIDERABLY.
We have already mentioned (above) how a cheap multimeter can produce a false reading when measuring across a 1M resistor. That's why you need high impedance test Equipment so you do not "load" the point you are testing and create an inaccurate reading.
The word **Impedance** really means resistance, but when you have surrounding components such as diodes, capacitors, transistors, coils, Integrated Circuits, supply-voltages and resistors, the combined effect is very difficult to work out as a "resistance" and that's why we call it "Impedance."
The term "**High and Low Impedance**" is a relative term and does not have any absolute values but we can mention a few points to help you decide.
In general, the base of a transistor, FET input of an IC are classified as HIGH IMPEDANCE.
The output of these devices are LOW IMPEDANCE.
Power rails are LOW IMPEDANCE.
An oscillator circuit and timing circuit are HIGH IMPEDANCE.
A LOAD is low impedance.
And it gets tricky: An input can be designed to accept a low-impedance device (called a transducer or pick-up) and when the device is connected, the circuit becomes LOW impedance, but the input circuitry is actually high impedance.
The impedance of a diode or LED is HIGH before the device sees a voltage higher than the junction

voltage and then it becomes LOW Impedance.
Impedance is one of the most complex topics however it all comes down to testing a circuit without loading it.
That's why test equipment should have an input impedance higher than 1M.

The first circuit we will investigate is the **Mini Bug Detector**, shown above and below. Points on the circuit have been labelled A, B, C etc:

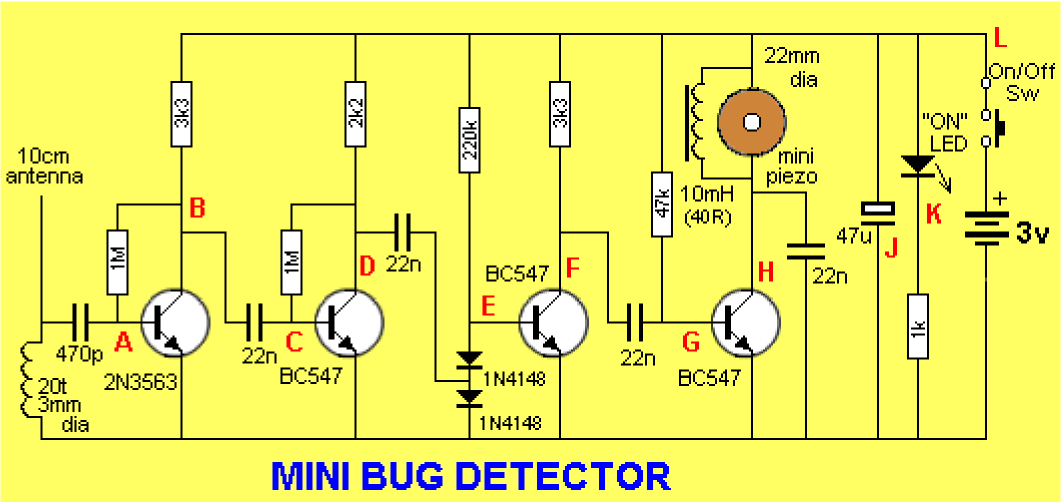

Point A - The first transistor is "self-biased" and will have 0.6v on the base. The antenna is connected to a 20 turn coil and you might think the coil will "short" the signals to earth.
But the coil and 470p capacitor form a circuit that oscillates at a high frequency when the antenna wire picks up stray signals. The coil and capacitor actually amplify the signals (see Talking Electronics website: Spy Circuits to see how a TANK CIRCUIT works) and these signals enter the base of the first transistor.
This is classified as a HIGH Impedance section because the signals are small and delicate and any loading via test equipment will kill them. The first transistor amplifies the signals about 70 times and they appear at **Point B**.

The signal passes though a 22n to **Point C** and the transistor amplifies the signal about 70 times to **point D**. **Point C** is classified as high impedance as any voltage measurement at this point will upset the biasing of the stage as a few millivolts change in base-voltage will alter the voltage on the collector considerably. **Point D** is classified as low impedance as any voltage-testing will not alter the voltage appreciably.
The output of the second stage passes through a capacitor to the join of two diodes. These two diodes are not turned on because the voltage at **Point E** can never rise above 0.7v as this is the voltage produced by the base-emitter of the third transistor.
The purpose of the two diodes is to remove background noise. Background noise is low amplitude waveforms and even though the transistor is turned on via the 220k, low amplitude signals will not be received. The third transistor works like this: It cannot be turned ON any more because any waveform from the 22n will be "clipped" by the bottom diode and it will never rise above 0.6v.
So, the only signal to affect the transistor is a negative signal - to turn it OFF.
Firstly we have to understand the voltage on the 22n. When the second transistor is sitting at mid-rail voltage, the 22n gets charged via the 2k2 and lower diode. When the transistor gets tuned ON, the collector voltage falls and the left side of the 22n drops. The right side of the 22n also drops and when it drops 0.6v, the top diode starts to conduct and when the voltage on the 22n drops more than 0.6v the third transistor starts to turn OFF. This effect is amplified by the transistor at least 100 times and appears at **Point F.** All the voltages around the two diodes are classified as HIGH Impedance as any piece of test equipment will upset the voltage and change the output.
There are some losses in amplitude of the signal as it passes through the 22n coupling capacitors but the end result is a very high strength signal at **point G.** The 4th transistor drives a 10mH choke and the mini piezo is effectively a 20n capacitor that detects the "ringing" of the inductor to produce a very loud output.
The 22n capacitor on the collector eliminates some of the background noise. The choke and piezo form an oscillatory circuit that can produce voltages above 15v, even though the supply is 3v.
The 47n capacitor at **Point J** is to keep the supply rails "tight" (to create a LOW Impedance) to allow weak cells to operate the circuit.
The "Power-ON" LED tells you to turn the device off when not being used and **Point L** is the power

supply - a low impedance line due to the 47u electrolytic.

Testing the Mini Bug Detector
To test the Mini Bug Detector, you will need a Signal Injector.
Place the Injector on **Point G** and you will hear a tone. Then go to **E**, **C** and **A**. The tone will increase in volume. If it does not increase, you have pin-pointed the faulty stage.

The next circuit is a combination of digital and analogue signals. It is a **Logic Probe**:

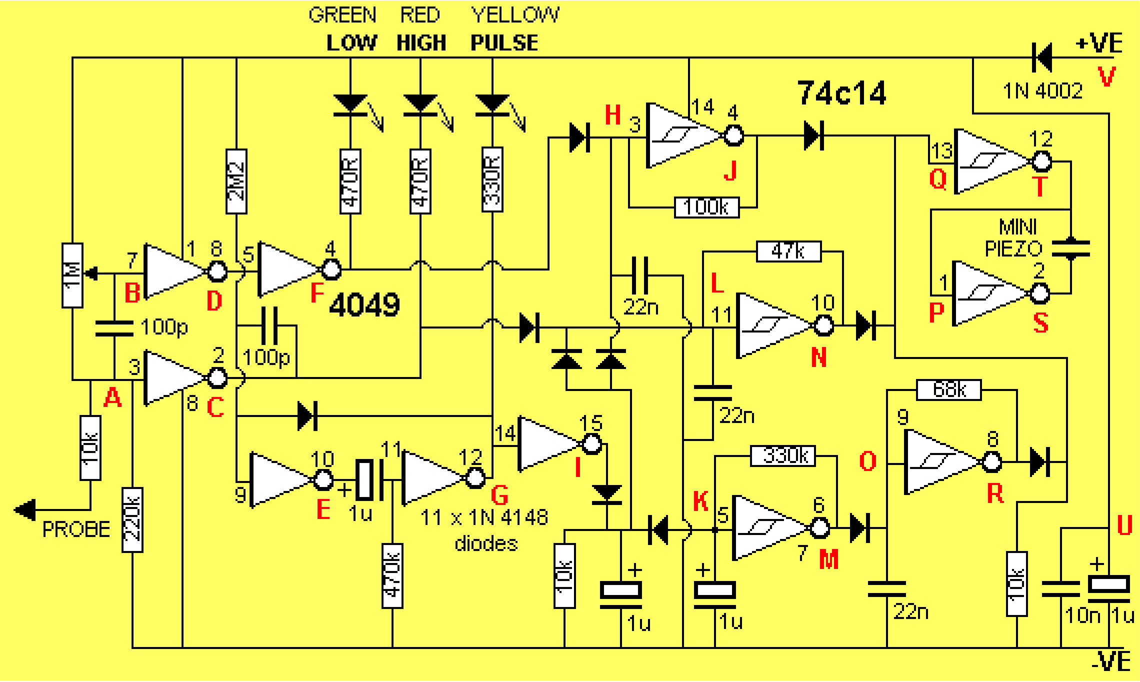

The voltage on a circuit (to be tested) is detected by the probe at **Point A** of the circuit above and the "tip" is classified as "reasonably high impedance" as it has a 220k resistor between the tip and 0v rail. The 1M reduces the impedance by about 20% but the inputs of the two inverters have no effect on the "tip" impedance as they are extremely high input-impedance devices.
The 1M trim pot is designed to put put a voltage on **point B** that is slightly higher than mid-rail so the green LED is turned off.
Point A will see a voltage below mid-rail and **point C** will be HIGH. **Point C** and **F** are low-impedance outputs.
When the tip of the probe is connected to a LOW voltage, Point B sees a LOW and Point F goes LOW to illuminate the green LED. At the same time it removes the "jamming voltage" produced by the diode between pin 4 of the 4049 and pin 3 of the 74C14 and the oscillator between **points H and J** produces a low-tone via the 100k resistor and 22n to indicate a LOW.
When the probe tip sees a HIGH, a lot more things happen.
Point C goes LOW and turns on the red LED. At the same time the 100p is in an uncharged state and the right lead goes LOW. This takes the left lead LOW as the left lead connects to a HIGH Impedance line and pin 9 goes LOW. This makes **point E** HIGH
and since the 1u is in an uncharged state, pin 11 goes HIGH. This makes **point G** LOW and the diode between pins 9 and 12 keeps pin 9 LOW and takes over from the pulse from the 100p. The yellow LED is illuminated. The 1u starts to charge via the 470k and when it is approx half-charged, pin 11 sees a HIGH and **point G** goes low. This creates the length of pulse for the yellow LED.
At the same time, **Point L** goes LOW because the "jamming diode" from pin 2 of the 4049 goes low and allows the inverter between point L and N to produce a tone for the piezo.
In addition, **Point I** goes HIGH and quickly charges a 1u electrolytic. This removes the effect of the jamming diode on pin 5 of the 74C14 and a low frequency oscillator made up of 68k and 1u between pins 5&6 turns on and off an oscillator between **points O** and **R** to get a beep. The mini piezo is driven n bridge mode via the two gates between **points QT** and **PS**.
Point U is a 1u electrolytic to reduce the impedance of the power rail and **Point V** is a protection diode to prevent damage if the probe is connected to the supply around the wrong way.

Testing the Logic Probe
You can test the Logic Probe with the simple Logic Probe with Pulse project described above. It will let you know if each point in the circuit is HIGH or LOW. You will also find out the difficulty in testing the points that are HIGH Impedance, as the Probe will upset the voltage levels and the reading may be

inaccurate.

More circuits will be added here in the future.

SOLDERING

Here are three 30-minute videos on soldering.
1. **TOOLS**
2. **Soldering components**
3. Soldering **SURFACE MOUNT** components

TESTING A MOTOR

Strictly speaking, a motor is not an electronic component, but since a website gave a useless description on testing motors, I have decided to supply the correct information.
The only REAL way to test a motor is to have two identical motors and check the torque by connecting them to a low voltage and trying to stop the shaft with your fingers. This will give you two results. Firstly it will let you know the torque of the motor.
This is the twisting effect of the shaft. There is no way to determine the torque by knowing the voltage or current.
The unknown factor is the strength of the field magnets (permanent magnets) and this determines the torque.
Secondly, feeling the shaft will let you know if the torque is even for a complete revolution.
By having two identical motors, you can see if one has a lower torque.
Almost nothing can go wrong with a motor except for the brushes. If the brushes wear out, additional resistance will be produced at the interface between the brush and commutator and this can be detected by allowing the shaft to rotate slowly and feeling the resistance as it revolves. A 3-pole motor will have three places where the strength is greatest and each should have the same feeling. A 5-pole motor will have five places of strength.
If the strength is weak or not uniform, the motor is faulty.
You cannot test a motor with a multimeter as the resistance of the armature winding is very low and if the motor is allowed to spin, the back voltage produced by the spinning, increases the reading on the meter and is false.
Micro motors have a coreless armature. This means the 3 windings for the armature are wound on a machine then bent slightly into shape and glued. A circular magnet with 3 poles is in the centre and the armature rotates around this.
This type of motor is reasonably efficient because the armature is the greatest distance from the point of rotation, and the motor reaches full RPM very quickly.
I have not heard of the armature-winding flying apart but if you hear any scraping noise, it may be the winding.
3-pole, 5-pole and micro motors can be found in printers, eject mechanisms of CD players, toys, RC helicopters, cars etc and rarely fail. This discussion has been presented to clarify the testing procedure and expose the absurd claims of another site.

TESTING COMPONENTS "IN-CIRCUIT"

You can test components while they are IN CIRCUIT, but the surrounding components will have an effect on the results.
You can get all sorts of "In-Circuit" testers. They are expensive and offer little more accuracy than a multimeter.
In-Circuit testing with a multimeter can give you the same results as a tester.
All you have to do is turn the project ON and use a multimeter (set to voltage) to determine the voltage at various points. It is best to have a circuit of the equipment so you can what to expect at each point.
Only major departures from the expected can be located in this way.
Obviously the first thing to look for is burnt-out components. Then feel components such as transistors for overheating.
The look for electrolytics that may be dry. Sometimes these have changed colour or are slightly swollen.
If they are near hot components, they will be dry.
For the cost of a few dollars I change ALL THE ELECTROLYTICS in some pieces of equipment, as a dry electrolytic is very difficult to detect.
Testing a transistor "in-circuit" is firstly done with the supply ON. That's because it is quicker.
Measure the voltage between ground and collector.
In most cases you should get a voltage of about half-rail. If it is zero, or close to rail voltage, you may

have a problem.
Turn off the supply and use the multimeter on low-ohms to measure all six resistances between the leads.
A low resistance in both directions on two leads will indicate a fault.
Resistors almost NEVER go "HIGH." For instance, a 22k will never go to 50k. However a low-value resistor will "burn-out" and you will read the value of the surrounding components.
Don't forget, some low-value resistors are designed to burn-out (called fusible resistors) and anytime you find a damaged low-value resistor, you will need to look for the associated semiconductor.
You can replace the resistor quickly and turn the circuit ON to see it burn out again.
Alternatively you can trace though the circuit and find the shorted semiconductor.
It's always nice to "see the fault" then "fix the fault."
Sometimes a transistor will only break-down when a voltage is present, or it may be influenced by other components.
When the piece of equipment is turned OFF, you can test for resistance values. The main thing you are looking for is "dry joints" and continuity. Dry joints occur around the termination of transformers and any components that get hot. Rather than wasting time checking for dry joints, it is better to simply go over the connections with a hot iron and fresh solder.
You may need to check the continuity of a track (trace) and it may go from one side of the PC board to the other.
Use a multimeter set to low-ohms and make sure the needle reads "zero-ohms."
It is very dangerous to do any testing on a project using a multimeter set to "amps" or "milliamps."
You cannot test "current flowing through a component" by placing the probes across a component. You will simply over-load the rest of the circuit and create a problem.
To find out if current is flowing though a circuit or a low-value resistor, turn the project ON and measure the voltage either across the component or the voltage on one end then the other.
A voltage-drop indicates current is flowing.
That's about it for testing "in-circuit." Use the rest of this eBook to help you with diagnosis.
Don't think an IN-CIRCUIT COMPONENT TESTER is going to find a fault any faster than a multimeter. They all use a multimeter principle.

THE END

This is not the full story to learning about servicing. It is just the beginning.
We have only covered the simplest tests and shown how 90% of faults can be found by checking voltages, waveforms and looking for obvious things such as burnt out components, cracks in PC boards.
The author has fixed over 35,000 TV's, radios, stereos, VCRs and all those things that were on the market 30 years ago.
Things have not changed. It's just that some repairs cost nearly as much as buying a new product and half the customers opt for dumping a faulty item and buying the latest "flat screen" version. That's why you have to get things through the workshop as fast and as cheaply as possible, to make a living.

If you want any more devices added to this list, email **Colin Mitchell**.

To help with understanding how a transistor circuit works, we have produced an eBook: **The Transistor Amplifier.** It covers a whole range of circuits using a transistor.

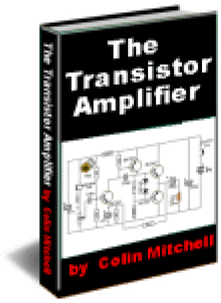

Not copyright by Colin Mitchell
You can use any of this material. Please pass this eBook to your friends and let them know that everything on the web is FREE. I have looked at all the "Pay Sites" and found the information they "sell" is available on the web at NO COST.
Nearly all text books are also downloadable for free on "Download.com" etc and when you see a used copy of a $74.00 textbook on Amazon for $12.00 you realise many users have already discarded their copy. A good textbook never gets thrown out.

See the enormous amount of information on Talking Electronics website.

16-10-2011 rev 17

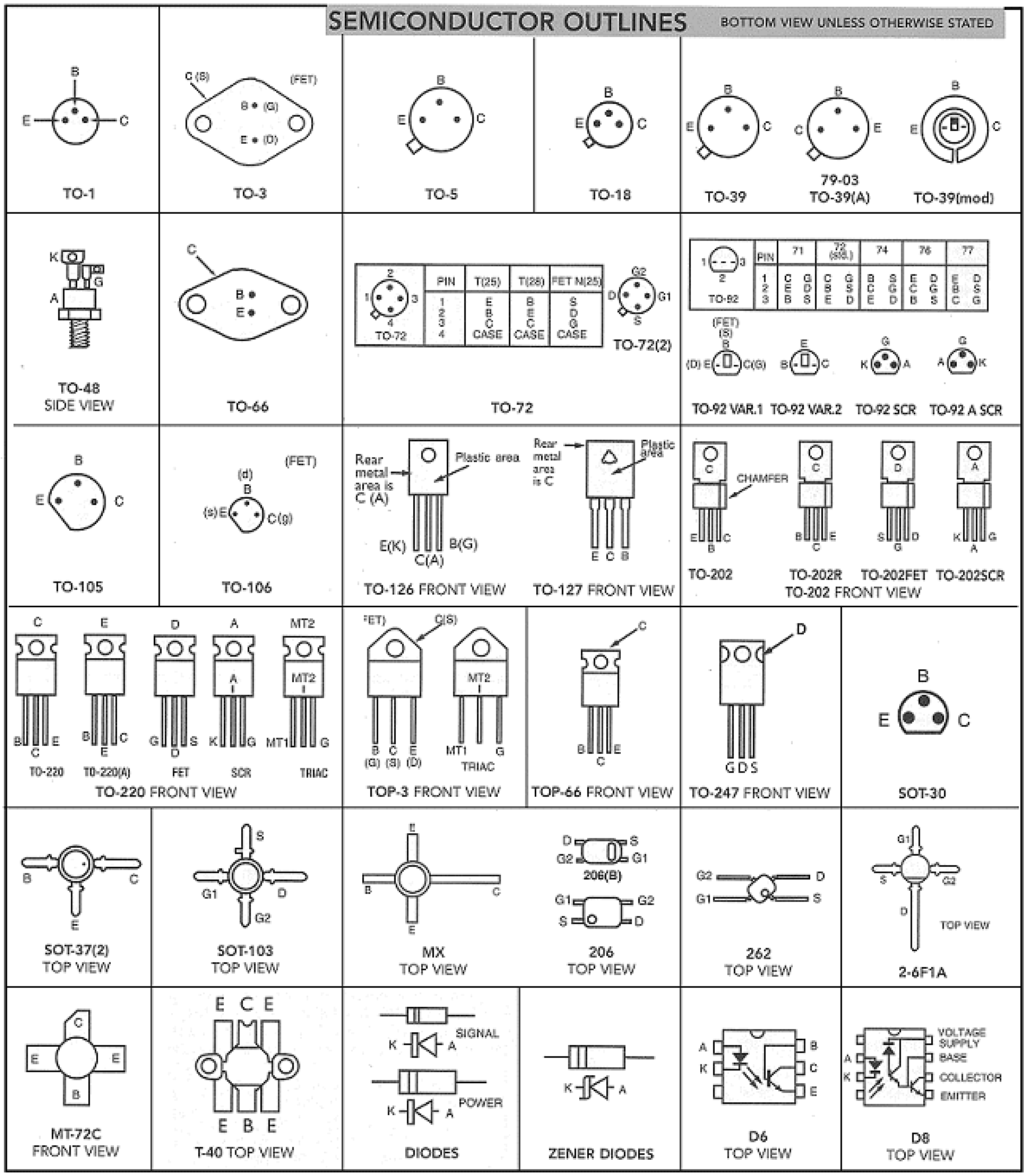

SEMICONDUCTOR OUTLINES BOTTOM VIEW UNLESS OTHERWISE STATED

TO-1
TO-3
C (S)
(FET)
B (G)
E (D)
TO-5
TO-18
TO-39
79-03
TO-39(A)
TO-39(mod)
TO-48
SIDE VIEW
TO-66
TO-72
PIN
T(25)
T(28)
FET N(25)
1 2 3 4
E B C CASE
B E C CASE
S D G CASE
TO-72(2)
G2 D G1 S
TO-92
PIN 71 72 (sel.) 74 76 77
1 2 3
C E B
G D S
C B E
G S D
B C E
S G D
E C B
D G S
E B C
D S G
(FET) (S) B
(D) E C(G)
TO-92 VAR.1
TO-92 VAR.2
TO-92 SCR
TO-92 A SCR
TO-105
TO-106
(d) B (s) E C(g)
Rear metal area is C (A)
Plastic area
E(K) C(A) B(G)
TO-126 FRONT VIEW
Rear metal area is C
Plastic area
E C B
TO-127 FRONT VIEW
CHAMFER
TO-202
TO-202R
TO-202FET
TO-202SCR
TO-202 FRONT VIEW
TO-220
TO-220(A)
FET
SCR
TRIAC
TO-220 FRONT VIEW
(FET) C(S)
B C E
(G) (S) (D)
MT2 MT1 G
TRIAC
TOP-3 FRONT VIEW
TOP-66 FRONT VIEW
D
G D S
TO-247 FRONT VIEW
SOT-30
SOT-37(2)
TOP VIEW
SOT-103
TOP VIEW
MX
TOP VIEW
206(B)
206
TOP VIEW
262
TOP VIEW
2-6F1A
TOP VIEW
MT-72C
FRONT VIEW
E C E
E B E
T-40 TOP VIEW
SIGNAL
POWER
DIODES
ZENER DIODES
D6
TOP VIEW
VOLTAGE SUPPLY
BASE
COLLECTOR
EMITTER
D8
TOP VIEW

CIRCUIT SYMBOLS by TALKING ELECTRONICS

AC current: 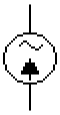 voltage: 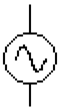	ALTERNISTOR TRIAC A TRIAC and 33 - 43v DIAC 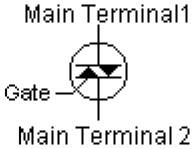 Main Terminal1 Gate Main Terminal 2	Ammeter (amp meter) 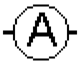
AND Gate 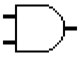	AND Gate 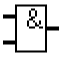	Antenna balanced 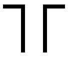
Antenna Loop, Shielded 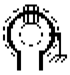	Antenna Loop, Unshielded 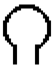	Antenna unbalanced 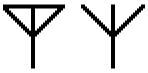
Attenuator, fixed (see Resistor) 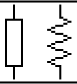	Attenuator, variable (see Resistor) 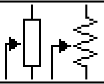	Battery 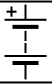
Bilateral Switch (DIAC) 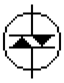	Bridge Rectifier (Diode Bridge) 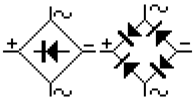	BUFFER (Amplifier Gate) 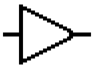
BUFFER (Amplifier Gate) 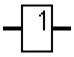	Buzzer 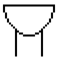	Capacitor feedthrough 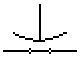
Capacitor non-polarised 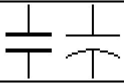	Capacitor polarised (see electrolytic) 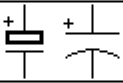	Capacitor Variable 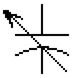
Cavity Resonator 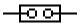	Cell 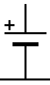	Circuit Breaker 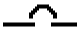
Coaxial Cable 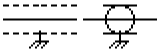	CRO – Cathode Ray Oscilloscope 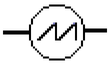	Crystal Microphone (Piezoelectric) 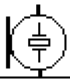
Connectors 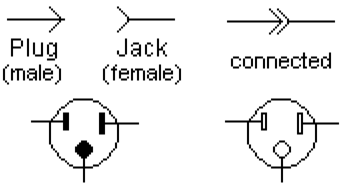 Plug (male) Jack (female) connected Plug (male) (female)	Crystal Piezoelectric 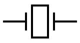 Darlington Transistor 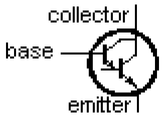 collector base emitter DIAC (Bilateral Switch) 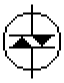	DC current: 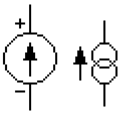 voltage: 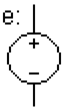 Delay Line 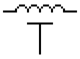 Diode 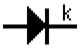
Diode - Gunn 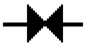	Diode - Light Emitting (LED) 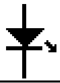	Diode Photo Sensitive 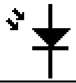
Diode Photovoltaic 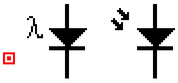	Diode Bridge (Bridge Rectifier) 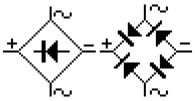	Diode - Pin 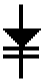
Diode - Varactor 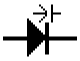	Diode - Zener 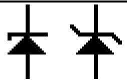	Earth Ground 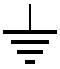
Earpiece (earphone, crystal earpiece) 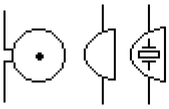	Electroluminescence 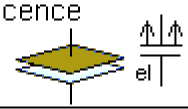	Electret Microphone (Condenser mic) 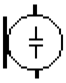
Electrolytic (Polarised Capacitor) 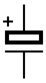 alternate symbols: (positive on top) 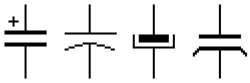	Electrolytic - Tanatalum 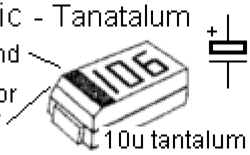 positive end black band or chamfer 10u tantalum 106	Exclusive-OR Gate (XOR Gate) 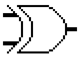 Exclusive-OR Gate (XOR Gate) 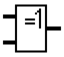
Field Effect Transistor (FET) n-channel also: N-Channel J FET 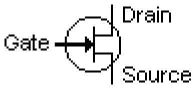 Drain Gate Source	Field Effect Transistor (FET) p-channel also: P-Channel J FET 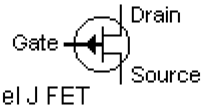 Drain Gate Source	Flashing LED (Light Emitting Diode) (Indicates chip inside LED) 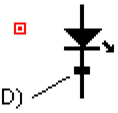

Ferrite Bead 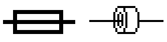	Fuse 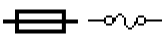	Galvanometer 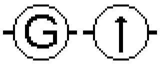
Globe 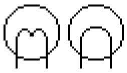	Ground Chassis 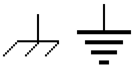	Ground Earth 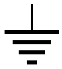
Heater (immersion heater) (cooker etc) 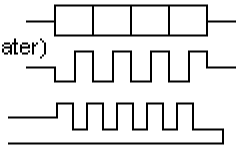	IC Integrated Circuit 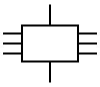 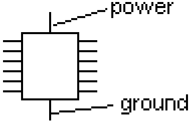	Inductor Air Core 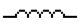
Headphone 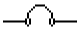		Inductor Iron Core or ferrite core 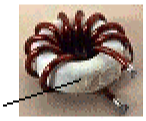 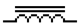
Inductor Tapped 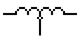	Inductor Variable 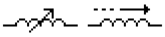	Integrated Circuit 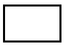
Inverter (NOT Gate) 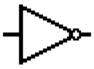	INVERTER (NOT Gate) 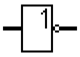	
Jack Co-axial 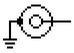	Jack Phone (Phone Jack) 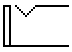	Jack Phone (Switched) 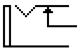
Jack Phone (3 conductor) 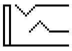	Key Telegraph (Morse Key) 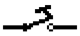	Lamp Incandescent 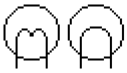
Lamp - Neon 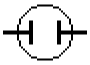	LASCR (Light Activated Silicon Controlled Rectifier) 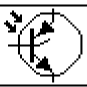	LDR (Light Dependent Resistor) 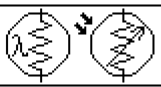
LASER diode 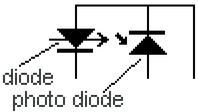	Light Emitting Diode (LED) 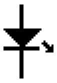	Light Emitting Diode (LED - flashing) (Indicates chip inside LED) 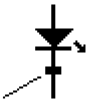
Mercury Switch 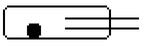	Micro-amp meter (micro-ammeter) 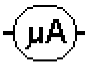	Microphone (see Electret Mic) 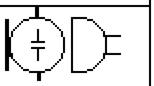
Microphone (Crystal - piezoelectric) 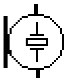	Milliamp meter (milli-ammeter) 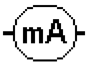	Motor 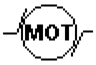
NAND Gate 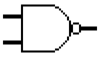	NAND Gate 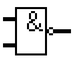	Nitinol wire "Muscle wire" 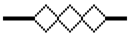
Negative Voltage Connection 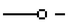	NOR Gate 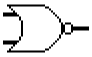	NOR Gate 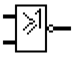
NOT Gate Inverter 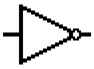	NOT Gate Inverter 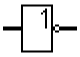	Ohm meter 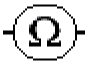
Operational Amplifier (Op Amp) 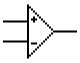	Optocoupler (Transistor output) 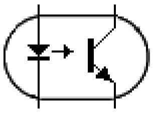	Opto Coupler (Opto-isolator) Photo-transistor output 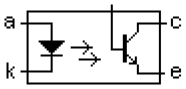
Optocoupler (Darlington output) 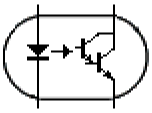	Opto Coupler (Opto-isolator) TRIAC output 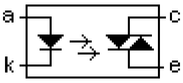	OR Gate 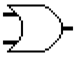
OR Gate 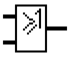	Oscilloscope see CRO 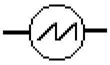	Outlet (Power Outlet) 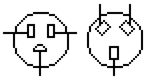
Piezo Diaphragm 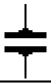	Photo Cell (photo sensitive resistor) 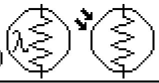	Photo Diode 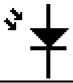
Photo Darlington Transistor 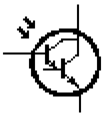	Photo FET (Field Effect Transistor) 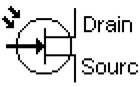	Photo Transistor 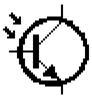

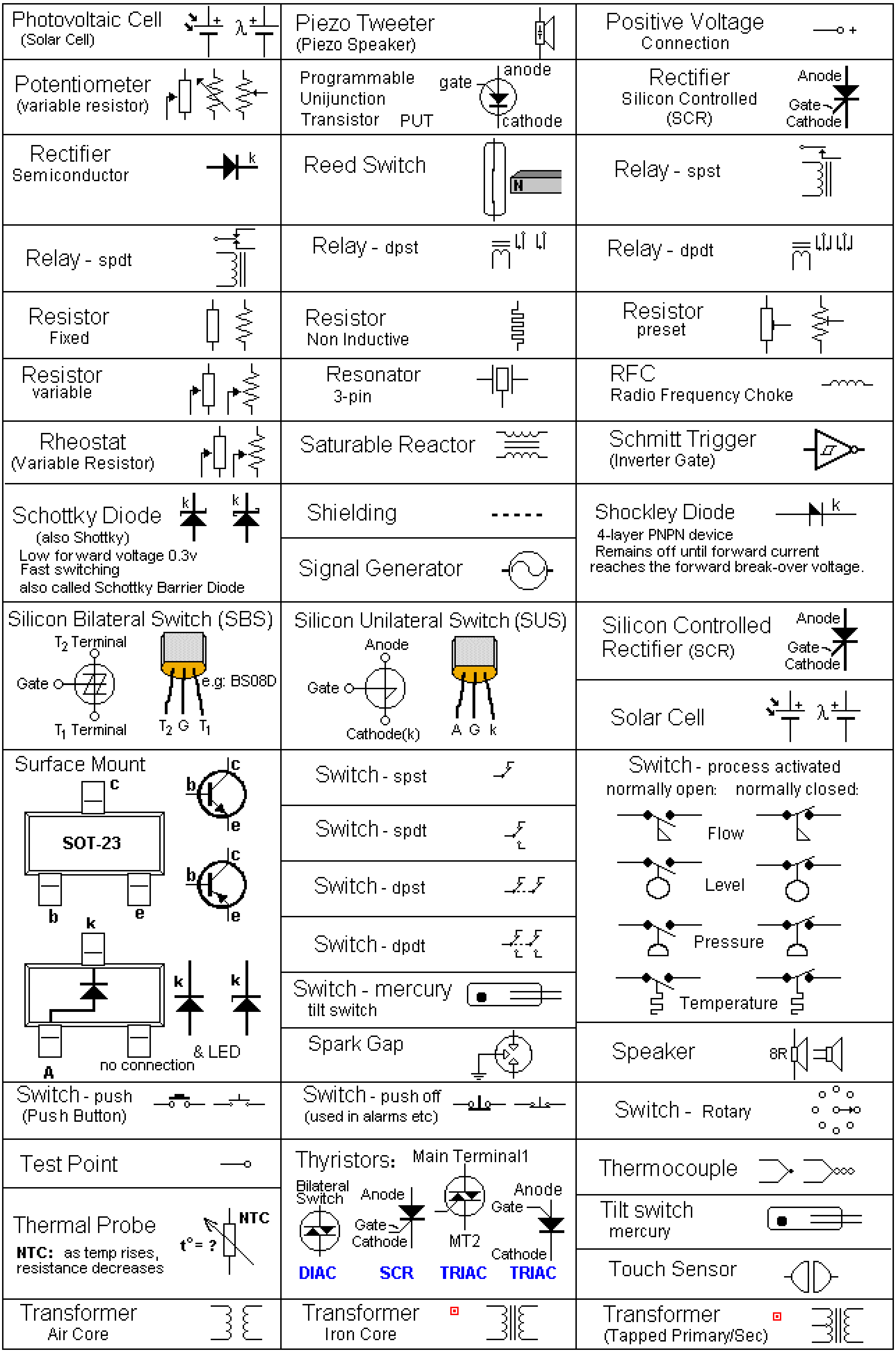
Photovoltaic Cell
(Solar Cell)
+
λ
Piezo Tweeter
(Piezo Speaker)
Positive Voltage
Connection
+
Potentiometer
(variable resistor)
Programmable
Unijunction
Transistor PUT
gate
anode
cathode
Rectifier
Silicon Controlled
(SCR)
Anode
Gate
Cathode
Rectifier
Semiconductor
k
Reed Switch
N
Relay - spst
Relay - spdt
Relay - dpst
Relay - dpdt
Resistor
Fixed
Resistor
Non Inductive
Resistor
preset
Resistor
variable
Resonator
3-pin
RFC
Radio Frequency Choke
Rheostat
(Variable Resistor)
Saturable Reactor
Schmitt Trigger
(Inverter Gate)
Schottky Diode
(also Shottky)
Low forward voltage 0.3v
Fast switching
also called Schottky Barrier Diode
k
k
Shielding
Signal Generator
Shockley Diode
4-layer PNPN device
Remains off until forward current
reaches the forward break-over voltage.
k
Silicon Bilateral Switch (SBS)
T2 Terminal
Gate
T1 Terminal
e.g: BS08D
T2 G T1
Silicon Unilateral Switch (SUS)
Anode
Gate
Cathode(k)
A G k
Silicon Controlled
Rectifier (SCR)
Anode
Gate
Cathode
Solar Cell
+
λ
Surface Mount
c
SOT-23
b
e
c
b
e
c
b
e
k
k
k
A
no connection
& LED
Switch - spst
Switch - spdt
Switch - dpst
Switch - dpdt
Switch - mercury
tilt switch
Spark Gap
Switch - process activated
normally open: normally closed:
Flow
Level
Pressure
Temperature
Speaker
8R
Switch - push
(Push Button)
Switch - push off
(used in alarms etc)
Switch - Rotary
Test Point
Thyristors: Main Terminal1
Bilateral
Switch
Anode
Gate
Cathode
MT2
Anode
Gate
Cathode
DIAC
SCR
TRIAC
TRIAC
Thermocouple
Tilt switch
mercury
Thermal Probe
NTC: as temp rises,
resistance decreases
t° = ?
NTC
Touch Sensor
Transformer
Air Core
Transformer
Iron Core
Transformer
(Tapped Primary/Sec)

Transistor Bipolar - NPN 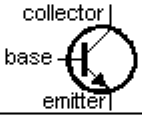 collector base emitter	Transistor Bipolar - PNP 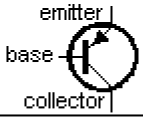 emitter base collector	Transistor n-channel Field Effect 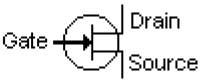 Gate Drain Source
Transistor p-channel Field Effect 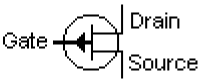 Gate Drain Source	Transistor Metal Oxide Single Gate 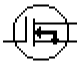	Transistor Metal Oxide Dual Gate 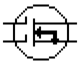
Transistor Photosensitive 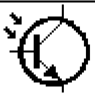	Transistor Schottky - NPN 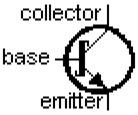 collector base emitter	Transistor Unijunction - UJT 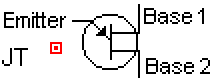 Emitter Base 1 Base 2 Unijunction Transistor (UJT) N-type
TRIAC 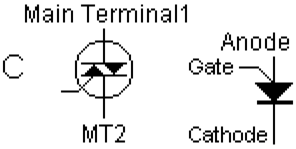 Main Terminal1 MT2 Anode Gate Cathode	Transistor Unijunction - UJT 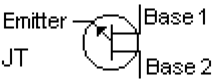 Emitter Base 1 Base 2 Unijunction Transistor (UJT) P-type	Tunnel Diode 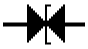 Unijunction Transistor - UJT 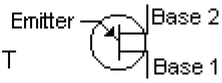 Emitter Base 2 Base 1
Varactor varactor diode 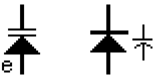	Voltage Regulator (7805 etc) 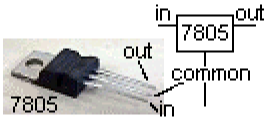 in 7805 out out common 7805 in	Voltmeter 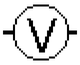
Wattmeter 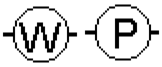	Wires 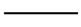	Wires Connected 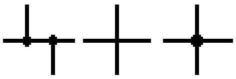
Wires Not Connected 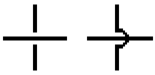	XOR Gate (exclusive OR) 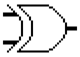	XOR Gate (exclusive OR) 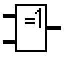
Zener Diode 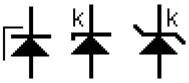	Learn **BASIC ELECTRONICS** Go to: http://www.talkingelectronics.com	

STARTING ELECTRONICS CONSTRUCTION

TECHNIQUES, EQUIPMENT AND PROJECTS

KEITH BRINDLEY

Newnes

Starting Electronics Construction

Techniques, Equipment and Projects

This Page Intentionally Left Blank

Starting Electronics Construction

Techniques, Equipment and Projects

Keith Brindley

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK •OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Newnes is an imprint of Elsevier

Newnes

Newnes
An imprint of Elsevier
Linacre House, Jordan Hill, Oxford OX2 8DP
30 Corporate Drive, Burlington, MA 01803, United States.
First edition 2005

Copyright © Keith Brindley 2005. All rights reserved.

The right of Keith Brindley to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

No part of this publication may be reproduced in any material form (including photocopying or storing in any medium by electronic means and whether or not transiently or incidentally to some other use of this publication) without the written permission of the copyright holders except in accordance with the provisions of the Copyright, Design and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London, England W1P 4LP. Applications for the copyright holders' written permission to reproduce any part of this publication should be addressed to the publishers.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN 0 7506 67362

For information on all Newnes publications
visit our website at www.newnespress.com

Typeset and produced by Co-publications, Loughborough

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER BOOK AID International Sabre Foundation

Contents

Preface vii

1. Practical electronics construction 1
2. Tools 5
3. Components 27
4. Printed circuit boards 73
5. Soldering 107
6. Housings 139
7. Fault-finding 177
8. Projects 193

Component suppliers 221

Index 225

This Page Intentionally Left Blank

Preface

This book is all about how to make electronics devices, from a constructional perspective. It is aimed at relative newcomers to electronics construction — that is, not the knowledgeable and trained electronics engineer, but those of us who want to find out about how to actually make up the circuits and projects that have already been designed.

It matters not whether the circuits you want to construct have been designed by others — perhaps they are presented in a magazine or a book as projects to build — or by yourself. It matters not whether you actually understand how that circuit works, or whether you haven't the foggiest idea of even its first principles — and maybe you don't even care! It matters not whether the intricate workings of a resistor; transistor; transformer; capacitor; or the latest wonder component is like Greek to double Dutch ears. There are other books that you can read which can give you all that understanding — and I'm bound to suggest my own book *Starting Electronics* as the ideal companion reference to this one, if that's the sort of understanding you are also looking for.

What *does* matter, on the other hand, is that *this* book (at least, that's my intention) shows everyone how to take an existing circuit and transform it into a complete and working device. This book starts from only the most limited knowledge of electronics processes, assuming very little — hopefully nothing at all, and introduces you to *all* the required concepts; the tools; the components; and the processes involved in turning a circuit into a finished electronic device. In fact, my intention in writing it is not only to make the journey from circuit to device as easy as possible to traverse, but to make it fun at the same time. I hope I've succeeded.

Running to keep up…

In electronics, as with most fast-moving technologies, there's always the risk that a book such as this one can become out-of-date very quickly. The way electronics moves these days, it's not unfair to say that a book like this is only up-to-date for as long as the ink is wet. So, at this point, I'd like to introduce readers to a central website:

http://keithbrindley.members.beeb.net

where I will try to keep relevant information about the methods and processes covered here up-to-date. The website also features relevant links to manufacturers, component stockists, and other links as I see fit. Projects (in Chapter 8) all feature printed circuit board copper foil track patterns, and there are links on the website to allow you to download these in electronic formats too. Indeed, I'd go as far as suggesting that if you intend constructing any of the projects here you should look on the website to see if there are any alterations or other suggestions regarding the projects before you start. Any corrections or alterations to the copper foil track patterns (or anything else to do with the projects) will be featured on the website as soon as I can get them there.

As you make it…

Electronics in general is a fascinating area. However, getting 'into it' in the first place is sometimes daunting. It is my aim that *Starting Electronics Construction* gives as many people as possible the means whereby they can do just that. In the meanwhile, if I can offer anyone a suggestion if they intend to try, it is this: make it fun!

That way, you'll learn so much more than you ever could from stuffy textbooks. I hope that by reading this book and getting into practical electronics construction that you do have fun.

Keith Brindley

1

Introduction

This book is a guide to anyone who wants to build an electronic device. It assumes very little prior knowledge, and so introduces you to the various practices that you need to follow — whether you want to build something for pleasure, or for education, or even for profit. Also, this book takes the stance that it matters little what the particular electronic product is you wish to build — the most important fact is that the practices you follow are similar (if not identical) whatever the electronic device happens to be.

Electronics is in part a theoretical entity, and you could go to textbooks or to magazines, schools, colleges or universities, or elsewhere to find and learn that theory. However, it is also in a very large part a practical entity; a fact that — particularly in recent years — seems almost to have been overlooked. The many students of electronics at universities who have little or no prior experience with the handling of components or the building of circuits attest to this.

So, although what you read on the pages of this book is theory, it is theory about the practical aspects of building electronic devices. As such, included in this book are (hopefully most of, possibly all) the practical things you need to know about and be able to do before you can make an electronic device.

What this book is *not*, on the other hand, is a book about how electronics circuits *work* (for that, see this book's sister book: *Starting Electronics*, which is written also by me). What you have in front of you now is a book only about how to *construct* electronic devices. I hope you enjoy it, and gain some knowledge about how to do just that.

What do you need to know about making an electronic device?

This section is intended as a summary of the whole book. It would be easy to miss this out altogether, but my feeling in writing it is that you need to see the *big picture* before you go on to look at things in more detail.

And the big picture in terms of building electronics devices is simply a quick overview of what an electronic device is, and also what parts of it should we understand in order that we can build it.

So, what is an electronic device?

We can summarize an electronic device (an example of which is given in Figure 1.1) as below.

An electronic device invariably has three main features:

NUMBER 1 Electronic components. An electronic device is anything that incorporates electronic components of any description.

NUMBER 2 A printed circuit board. The electronic components which make up an electronic device are mounted upon a *printed circuit board* — (often called just a *PCB*), which is a means of holding all the components securely while at the same time making the necessary electrical connections between them all.

NUMBER 3 A housing. The printed circuit board, together with all the electronic components that it holds, will be housed in a box, or a module, or a case. In this way the electronic components are protected from damage.

So, what's next?

Given that any electronic device you might choose to build has these three main features, it follows that you need to know as much as you can

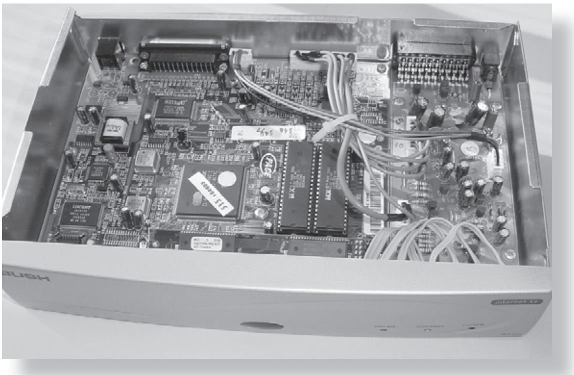

Figure 1.1 An example of an electronic device – note the electronic components, the printed circuit board, and the housing

about the features. And we'll look at them now, in brief, before devoting the rest of the book to that purpose.

Electronic components

There are many categories of electronic components, and each category is divided often into several sub-categories, and from there into a multitude of sub-sub-categories. You don't need to know about them *all* by any means, but you do need to know about the most common categories, and their most common sub-categories. You need to know how to handle these components so that you don't damage them, and so you need to know what damage *can* be done to them.

Some components are easily damaged just by holding them a little too roughly; some components are damaged by heat; and certainly many components are damaged if you insert them into a circuit incorrectly (the wrong way round, say, or too forcefully, thereby breaking their connection pins).

You also need to know about which tools you need to handle them and how to use these tools correctly.

Printed circuit boards

Making the printed circuit board onto which the electronic components of a device are mounted is not a trivial job. Indeed, the work involved in designing it, making it, then attaching the components to it represents a major proportion of the overall task. However, making a printed circuit board is not a job which should be rushed, for a poorly made printed circuit board is a recipe for disaster.

The printed circuit board is used for two main purposes: to support the electronics component; and to connect between the electronic components so that the circuit they are to be connected into can be formed. As a result, a printed circuit board that cannot support the components adequately, or makes incorrect connections, will prevent the electronic device from functioning correctly. So, you need to know about all aspects of printed circuit boards: how to design them; how to make them; and how to mount the components. You need to know what tools to use to do these tasks, and how to use the tools correctly.

I simply cannot stress enough how important the printed circuit board is. If the circuit itself is the brain of an electronic device — that monitors, controls and sends the signals that make the device work — then the printed circuit board is the heart that keeps the whole device going in the first place.

Housings

Once the printed circuit board is complete, with all electronic components in place, you have to consider how to house it. Housings are used to protect the components inside, as well as their printed circuit board. Housings also usually give a little extra space to allow batteries to be held, when needed, as well as providing mounting space for some components and controls not mounted on the board. They also protect users from potentially harmful voltages that may be present in the circuit — mains voltages, say. Labelling on a housing's facing panels can indicate what various control adjustments may be made by the user. As such, a housing is the interface — technically, visually, and aesthetically — with the outside world.

In short, if the circuit is the brain, and the printed circuit board is the heart, then the housing is the body, the skeleton, and the skin of an electronic device.

Housings done right can make a product a winner; done wrong the product is a duffer.

What have we learned?

To summarize this chapter we only need to remember one thing — an electronic device comprises three main parts. These main parts — the electronic components, the printed circuit board, and the housing — are highly dependent on each other. Designers and makers of *good* electronic devices understand this dependence, and use the main parts to the best advantage of their devices.

2

Tools

As with any practical technology, having the right tools for the task is important. Indeed, some of the tools are not just important — they are critical! You simply cannot do electronics construction without them. So, first off in this chapter is a hit parade of the tools you need, followed by another of the tools you should seriously consider (although you *could* do without them at a push!).

Tools you absolutely, positively, indisputably, and without a doubt need:

SOLDERING IRON — to solder components into place. The connecting lead of each component must be soldered into place on the printed circuit board. A specialized soldering iron must be used — one that is neither too big nor too powerful to harm either the component or the circuit board

MINIATURE SIDE-CUTTERS — to cut wire or component leads. While constructing an electronic device, there will be many wires to cut to size. Each component mounted on the printed circuit board needs its leads trimming close to the board after soldering

MINIATURE LONG-NOSED PLIERS — to form leads. Component leads are formed to the correct shape for mounting on a printed circuit board with small, long-nosed pliers. Larger components and many controls are attached with nuts and bolts, so a larger pair of long-nosed pliers is also needed

CONTINUED…

…CONTINUED

WIRE-STRIPPERS — to strip insulation from wires. Insulation must be removed from wires prior to soldering, and gripping the wire in your teeth before pulling the insulation off isn't the best method. A good pair of wire-strippers solves the problem and keeps your dental bills as low as possible

SOLDER REMOVAL TOOLS — to err… remove solder. At some point or other, you will need to remove solder from a soldered joint, usually to take out a component you have accidentally put in the wrong place, or to replace a faulty component. Some type of solder-removal tool is required, allowing you to extract the component's leads

DRILLS AND BITS — to err… drill holes. Every component lead goes through a hole in the printed circuit board. Each hole must be drilled. Components not mounted on the printed circuit board are usually mounted on the housing, usually with nuts and bolts, and holes for these must be drilled in the housing

SAFETY GOGGLES — to protect your eyes. Some of the processes involved in electronics construction may create potential hazards. You need to protect yourself! Goggles should be used when undertaking any process likely to create flying debris, such as drilling or using chemicals.

Tools you don't exactly *need* — but which make some jobs easier:

PLIERS — ordinary hand pliers

FILES — a small set of needle files, and a selection of larger files

PUNCHES — for making some larger holes in housings

VICE — to hold housings while working

A STAND — to hold printed circuit boards.

The Tools

We'll now take a closer look at the most important of those tools listed, and note any particular attributes they should have.

Soldering irons

Soldering in electronics does not have any particular requirements, other than the soldering iron should not be too powerful. On the face of it, this might sound odd to a newcomer given that the process of soldering requires solder to be melted thus forming a soldered joint, but it's quite logical when you consider exactly *what* you are soldering — electronic components. The problem is that electronic components can be damaged by too much heat — and the more powerful the soldering iron, the greater its capacity to deliver heat.

Soldering iron power is measured in watts, and a reasonable wattage for soldering electronic components is somewhere between 12 and 25 watts. Prices vary between models, but you should expect to pay between about £10 and £20 for a basic soldering iron, although more expensive soldering irons (with extra features) exist.

A typical soldering iron is shown in Figure 2.1, where you can see its main parts — a handle to hold it, a bit (the part that heats the joint and the solder), and a heating element (the part that heats the bit). Also shown in Figure 2.1 is a mains lead, although not all soldering irons require mains electricity to function, as we'll see soon.

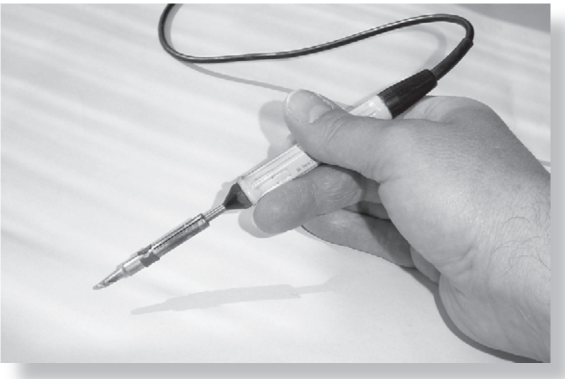

Figure 2.1 A typical soldering iron – the main parts of which are (from left to right): the bit; the heating element; the handle. This particular soldering iron is mains-powered, so you can also see the mains lead exiting the handle at the right.

Soldering iron types

There are several variants of soldering irons, in terms of heating methods, listed below.

Mains-powered irons — the most common. The advantage of these is that they work wherever a main outlet is available. The disadvantage is that they do have a mains lead, which some people find a little fiddly in use. Basic soldering irons of this type have no temperature control, and merely get as hot as they can. Soldering irons with thermostatic controls are available, and even complete soldering iron/base station combinations are common (see later). Your budget defines what you buy.

Rechargeable irons — with base station. Rechargeable irons have internal batteries that are used to heat the element and bit. They have a base station that is mains-powered, into which the soldering iron fits when not in use. Connections in the base station and soldering iron mate, and so the batteries are recharged whenever the iron is in the base station (and the base station is powered, of course).

The main advantage of rechargeable soldering irons is their lack of mains lead in use. The main disadvantage is that the batteries can't power the element for longer than just a few minutes at a time. By replacing the soldering iron into the base station after each joint is soldered though (so that its batteries recharge partially), this can be extended somewhat. Figure 2.2 shows a rechargeable soldering iron with its base station.

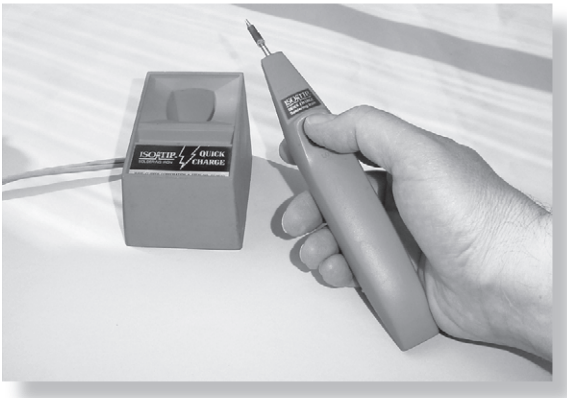

FIGURE 2.2 **Rechargeable soldering iron in use. To conserve batteries, the element only heats up when the button on the handle is pressed – heating rapidly to soldering temperature within seconds. In the background is the base station which recharges the soldering iron's batteries.**

Gas-powered irons — the ultimate in portability. Fitted with small refillable charges of butane gas which last up to around 30 minutes of constant use, these soldering irons require no electrical power. Figure 2.3 shows a gas soldering iron in use.

Priced between around £20 to £40 or so, these soldering irons are useful for certain awkward soldering jobs where electrical power is not available. They tend to be quite powerful, particularly the more expensive models, though with care can be used to good effect in electronics soldering.

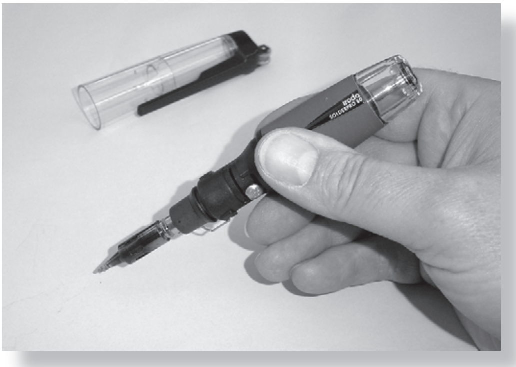

FIGURE 2.3 A gas-powered soldering iron. Small and portable, these can be recharged from lighter fuel butane gas containers.

Low-voltage irons — for battery-powered applications. Occasionally, when working on a car, say, there are applications where a low-voltage soldering iron might be useful. Some 12 V-powered soldering irons are available (an example is shown in Figure 2.4), usually having battery clips attached to the power lead.

Soldering iron stations

While we're on the subject of soldering iron types, it's worth taking a look at soldering iron stations. These are complete units, comprising a soldering iron, a stand to hold the iron when not in use, and a controlling circuit of some description. A typical soldering iron station is shown in Figure 2.5. The controlling circuit typically controls the power to the soldering iron element, so that the bit stays at a constant, definable temperature. While such stations are expensive, ranging from around £30 to over £100, they are excellent tools if you do a considerable amount of soldering.

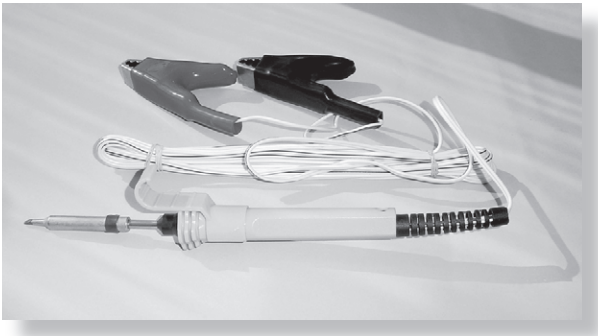

FIGURE 2.4 A 12 V-powered soldering iron. The battery clips on the lead attach to a car's battery for (relatively) portable power.

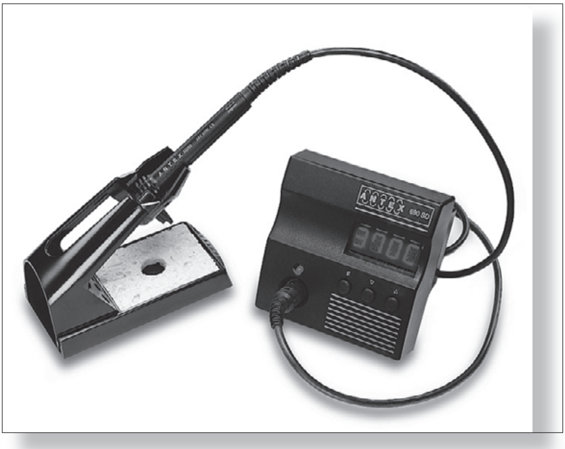

FIGURE 2.5 A soldering iron station, showing the iron itself in a stand, and the control knob to set bit temperature (Antex Electronics).

Soldering iron bits

Alongside the iron itself, you need to think about the bit. The bit is the part of the iron that comes into contact with the parts to be joined, so is a fairly critical part of the process.

A typical range of bit shapes is shown in Figure 2.6. Usually, for soldering electronic component joints an angled bit of about 2.4 mm diameter is ideal.

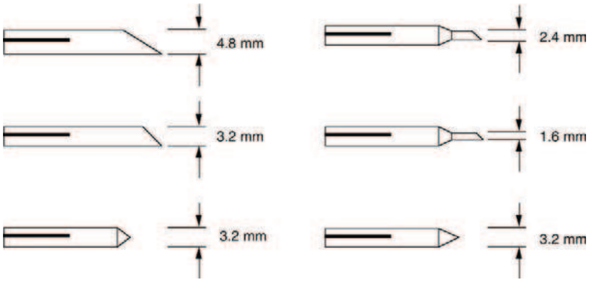

Figure 2.6 Typical sizes of soldering iron bits.

How the bit fits onto the soldering iron element is of concern, for two reasons:

- it determines how much of the heat generated by the iron's element passes to the bit's tip — and hence to the joint to be soldered
- it determines how easy it is to remove the bit — important if you are to solder different types of joints (for example, a bit with a smaller diameter would be used for finer joints), or for when you need to replace an old and worn bit.

In general, bits that slide over a soldering iron's element are most efficient, allowing much of the heat generated by the element to pass to the bit's tip. Figure 2.7 shows such a bit, fitted to its element. Also, this bit type is the easiest to remove.

Bits that fit inside an element, or screw onto it, are probably best avoided, if only because the very nature of the soldering process means that metal bits often corrode — even slightly — due to heat, and this may mean that the bit becomes hard to remove. Excessive pulling or twisting of the bit to remove it may then damage the element.

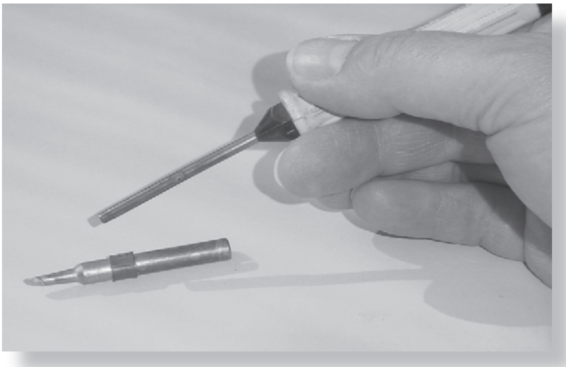

Figure 2.7 A soldering iron whose bit slides over the element. The bit is easily removed and/or replaced, and provides an efficient heat transfer between the element and the bit's tip.

Some soldering irons use a bit type that also comprises the element. The bit/element combination attaches to the soldering iron by two prongs that slide into the soldering iron body. Element and bit tip being unified in this way, there is a highly efficient heat transfer between element and tip, and this design has the great advantage that changing the bit effectively replaces the element in one operation. One of the best examples is the rechargeable soldering iron bit tip shown in Figure 2.8.

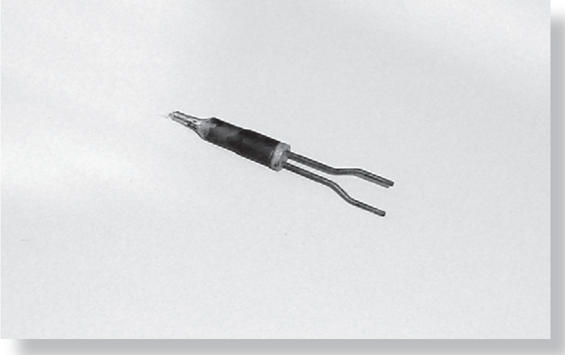

Figure 2.8 The unified element/bit of the rechargeable soldering iron shown in Figure 2.2 fits to the soldering iron's body by pushing in the two prongs.

Soldering iron accessories

Soldering iron stand

There's a few accessories for soldering irons that are worth buying. First — and most important — of these is a soldering iron stand (a typical stand is shown in Figure 2.9), which holds an iron while it's not in use. Soldering irons, of course, by their very nature are hot, so having a stand to put one in when not actually using it adds up to quite a safety feature. The stand in Figure 2.9 is around a fiver — so shouldn't break the bank.

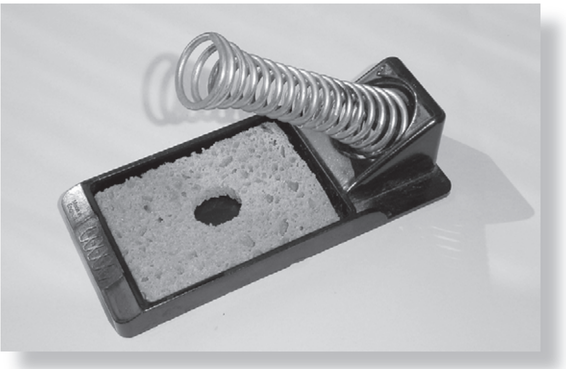

FIGURE 2.9 A soldering iron stand. This particular stand features a sponge.

Sponge

Talking of a soldering iron stand with a sponge, a sponge is actually a very handy accessory to have. Even if you have to resort to — as I have done on occasion — using a simple kitchen sponge, dampened with water, stood on a tea-plate, then so be it!

The £5 soldering iron stand of Figure 2.9, featuring an integrated sponge is a great way to combine the two accessories — not essential, just neat.

You use a dampened sponge to wipe your hot soldering iron bit tip on, immediately prior to tinning it (see page 117) and using it to solder joints. As you'll see later, a clean soldering iron bit tip is one of the prerequisites of good soldering.

Put another way: if your soldering iron bit tip is not clean, don't expect to make good soldered joints, and *do* expect your circuit not to work!

Tip tinning and cleaning block

Another way to keep your soldering iron bit's tip clean is to use a tip tinning and cleaning block. It's a thin disc-shaped block in a metal container that can be carried in a pocket, or mounted on a suitable surface close to the soldering iron. In use, you simply wipe the hot soldering iron bit's tip over the block's surface, simultaneously cleaning and tinning the tip, as shown in Figure 2.10. It's convenient, it's handy, it's cheap (around £4), and it saves having a wet sponge on your worktop.

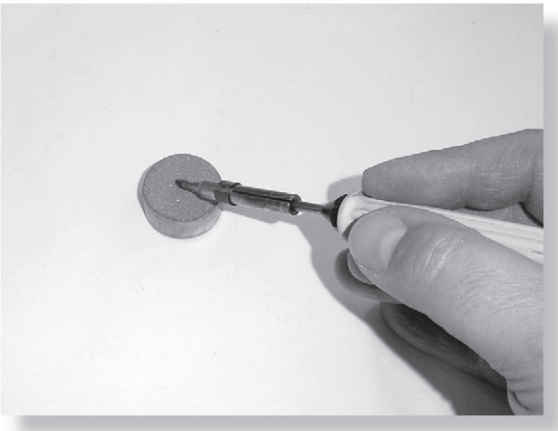

FIGURE 2.10 Using a tip tinning and cleaning block to err... clean and tin a hot soldering iron bit tip.

Abrasive fibreglass pencil

While we're on the subject of cleanliness (which we'll return to often throughout the book), the other components that go to make up a soldered joint have to be clean — just as much as the soldering iron bit tip. Fortunately, there are products that can help here too. The copper surfaces of the track on a printed circuit board should be clean and, as they oxidize when left for more than just a few hours (and copper oxide can prevent a good joint from being made), they should be cleaned more-or-less immediately prior to soldering the joints.

A good tool for this job is a fibreglass abrasive pencil. It's a propelling-pencil type of design, with an internal fibreglass stick instead of pencil lead — see Figure 2.11. In use, you rub the end of the fibreglass stick on the copper track a few times, removing any dirt, grease, or pre-finishes, leaving clean, shiny copper, ready for soldering.

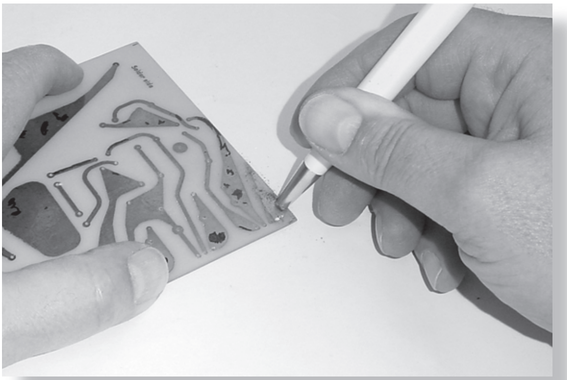

FIGURE 2.11 An abrasive fibreglass pencil in use – note the residue, which is of fibreglass origins.

After cleaning the copper surface, care should be taken when removing the residual dust created in the process — it *is* fibreglass-based after all, so may irritate skin, eyes, or internal organs.

Abrasive scrubbing block

An alternative to the fibreglass pencil is an abrasive scrubbing block, which performs the same function — scrub the copper track prior to soldering with the block, in an action reminiscent of using an eraser to rub out pencil marks. Figure 2.12 shows such a scrubbing block.

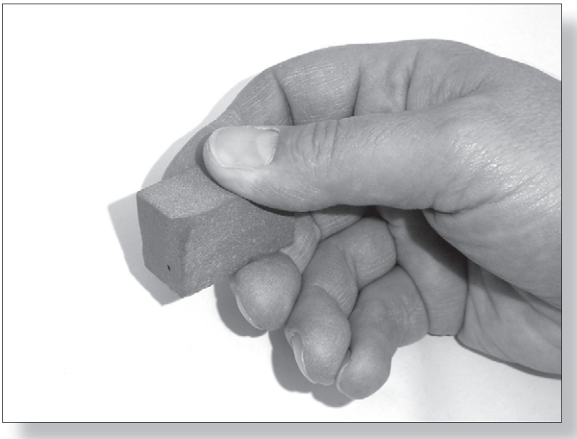

FIGURE 2.12 An abrasive scrubbing block.

Component leads — which form a third critical part of a soldered joint on a printed circuit board need to be clean every bit as much as the copper track and the soldering iron bit's tip. Most components are supplied with tinned leads (ie, they are coated with molten solder then cooled), fortunately, which means that a simple reheating to soldering temperature re-melts the solder which helps the process of soldering the joint. Even so, excess dirt — particularly grease — on a lead may still prevent a good joint being made, so if you suspect that to be the case you should clean the lead prior to soldering. It's much easier to do this before soldering, than trying to correct the effects of it afterwards.

Chemical solvent cleaners

Chemical cleaners such as isopropyl alcohol may be useful for cleaning printed circuit board tracks or component leads prior to soldering, as they dissolve grease, dirt, and some pre-finishes which may be present. They can be purchased in liquid form, although aerosol or pen applicators are often much more convenient.

Heat shunt

The final soldering accessory you might want to consider in your armoury is a heat shunt. Soldering — by its very nature — requires heat, but too much heat can damage some electronic components. Some people therefore like to use a heat shunt between the soldered joint and the component body to help ensure the heat at the joint is shunted away from the component itself.

The simplest form of heat shunt is a pair of locking tweezers which are clamped onto the component lead. As the joint is heated, heat is shunted into the tweezers rather than passing down the component lead into the component body.

Figure 2.13 shows such a pair of tweezers.

Take Note:

Having said all that, I am duty-bound to say that I have never owned a heat shunt, I have never even used a heat shunt, and I don't think I will ever use a heat shunt. All but a very few electronic components are fairly heat-resilient after all, and as long as the soldered joint is performed rapidly and without leaving the soldering iron in place on the joint too long, then excessive heat should not be a problem.

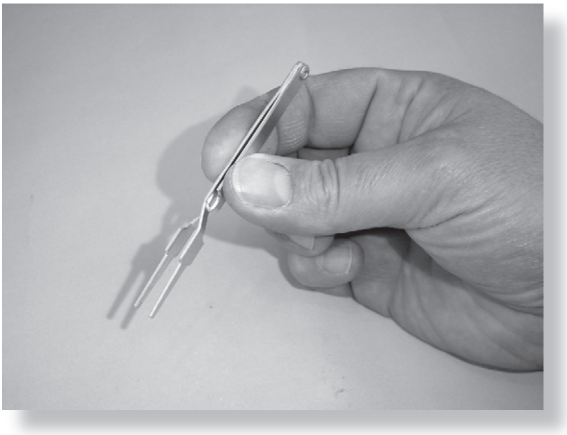

Figure 2.13 A heat shunt.

Side-cutters

After your soldering iron, the single most important tool in your toolbox is a decent-to-good quality pair of miniature side-cutters. As side-cutters *are* so important, it's best not to skimp on this tool — buy the best you can afford (decent quality side-cutters are available at around £10, while *good* quality side-cutters are priced between about £20 and £80…).

Before you baulk at the price, remember that the side-cutters you buy will be used every time you solder a printed circuit board, to trim the component leads after soldering joints. They'll be used every time you cut wires to length, for insertion into a printed circuit. They'll be used every time you trim your toenails. They should *not* be used for cutting thick wires or metal (or pruning the roses, or cutting toenails, if you hadn't already realized I wasn't being serious).

And finally, just to labour the point, a decent pair of side-cutters will last a lifetime in electronics, but a cheap pair will probably last till you need them.

Figure 2.14 shows a pair of side-cutters that have lasted a lifetime (mine). They were expensive (costing around £50 in 1976), and at the time I was young and foolish and thought it was a little over-the-top, but now that I've matured and I'm 21 again (err… yes, well…) I certainly realize their worth. In all seriousness though, the money you spend on this tool is worth it in the long run.

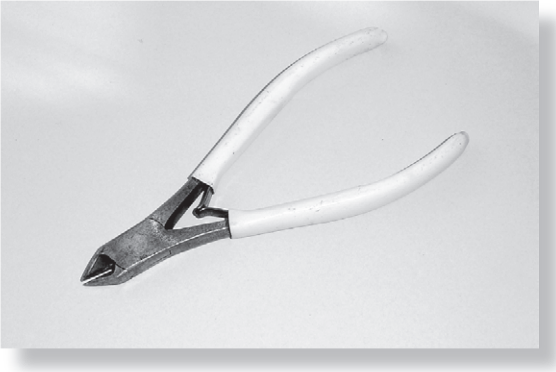

Figure 2.14 A good quality pair of miniature side-cutters. After your soldering iron, it's the tool you'll use the most in electronics.

Long-nosed pliers

Like side-cutters, a miniature pair of long-nosed pliers is an essential tool for the builder of electronics circuits. Like side-cutters also, it's pretty important to have a decent-to-good quality pair. And, if you haven't already guessed it, like side-cutters too, they may be expensive — decent quality long-nosed pliers will cost around £10, while a good quality pair will be priced between around £20 and £80). Again, get the best you can afford — they will be worth it. Figure 2.15 shows my long-nosed pliers (which I bought along with my side-cutters in 1976, for around £50). They're still working perfectly, which is more than I can say for myself.

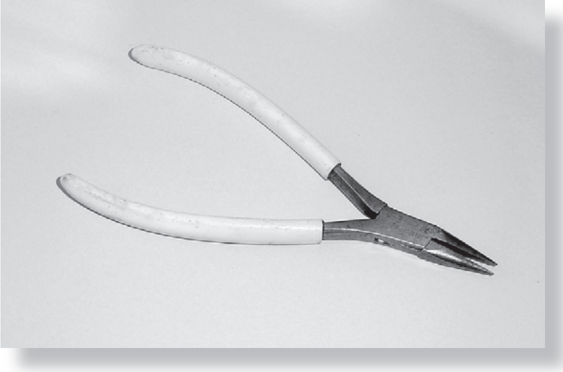

Figure 2.15 Miniature long-nosed pliers are extremely useful tools to keep in your electronics toolbox.

Long-nosed pliers are very useful for forming component leads prior to mounting the component in the printed circuit board. They should not be used for anything heavier than this — and should definitely *not* be used for tightening nuts onto large bolts, or working on the car. They are precision instruments which require a little forethought in handling.

Wire-strippers

Printed circuit boards are not usually complete entities, and they need to be connected to the outside world. The easiest and most popular way to do that is with interconnecting wires, which are normally insulation-covered to prevent short circuit with their neighbours. Each wire has to be soldered to the printed circuit board, so it's necessary to remove a piece of insulation at the wire end, leaving a short length of uninsulated wire that can be soldered. It's perfectly possible to strip a short piece of insulation away by putting the end of the wire into your mouth between your teeth, biting on the wire while pulling it back out of your mouth. Indeed, I used to do this very thing before I wore a hole in my front teeth and suffered many expensive dental bills. Now I use a pair of wire-strippers, and I wholeheartedly recommend you do the same.

Wire-strippers work in exactly the same way that front teeth do, in that you insert the wire into the wire-strippers jaws (and yes, they are called jaws) whereupon you squeeze the wire-strippers' handles together so that the wire-strippers' teeth catch the wire between them. Further squeezing of the handles forces the teeth into the insulation and simultaneously pulls the jaws back, stripping the insulation from the wire. Figure 2.16 shows a pair of wire-strippers in action.

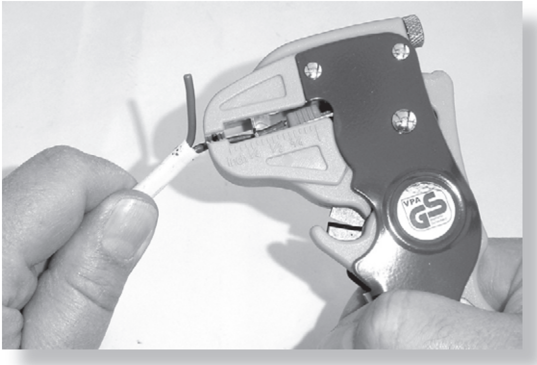

FIGURE 2.16 Using a pair of wire-strippers to strip the insulation from the end of a wire, prior to soldering to a printed circuit board.

Such wire-strippers aren't too expensive (around a tenner), although you can get much more expensive ones. As with the case of wire-cutters and long-nosed pliers, get the best you can afford. The best wire-strippers have replaceable jaws (unlike me) so can be always maintained with a good set of teeth. The cheaper ones are just throw-away tools (I've heard *that* said about me before, as well…).

Solder removal tools

If you solder electronic components into printed circuit boards, then sooner or later you will need to desolder them to remove them from printed circuit boards. It goes without saying that you will at some stage put a component in the wrong place and need to take it out, or you will find that you have a faulty component and need to replace it.

There are two tools that can help you do this and your selection of which you buy is merely a matter of choice. The cheapest is desoldering braid — lengths of copper braided wires that attract molten solder away from a soldered joint, simply because solder flows to cover the copper wire surfaces in the braid. Figure 2.17 shows desoldering braid in use — the braid is placed over the joint and a hot soldering iron bit presses on top. As the heat from the iron bit melts the solder it is attracted onto the braid away from the joint. Once the component lead becomes sufficiently free of solder to be removed, the braid and the soldering iron bit are removed, and the lead can be extracted from the joint hole.

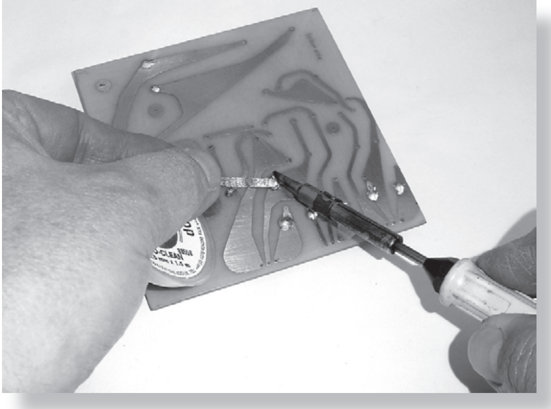

FIGURE 2.17 Desoldering braid in action. The braid is heated by the iron's bit, which in turn heats the solder on the joint to melting point, whereupon the solder is attracted to the copper wires in the braid.

Once the braid end has been used to desolder a joint, it becomes loaded with solder so must be cut off, leaving fresh braid for the next desoldering operation. As such, a length of desoldering braid only has a limited life (depending on how many joints you desolder). Desoldering braid costs £2 or £3 for a metre or so, although you can buy it in longer length reels if you need.

The alternative tool for desoldering purposes is a desoldering pump, commonly called a *solder sucker*. This works something like a bicycle pump in reverse — you operate the pump (which is spring-loaded) to push out air, then apply the pump to molten solder on a joint and release the plunger to literally suck the molten solder away from the joint into the pump where it rapidly hardens. The next operation of the pump forces out the hardened solder and primes the pump ready for the next desoldering operation. Figure 2.18 shows a desoldering pump in position over a joint as the joint is heated to become molten by a soldering iron bit tip. The pump has been primed (ie, its plunger has been pushed down the pump body), and as the solder becomes molten, the button catch on the desoldering pump body is pushed, thereby releasing the spring-loaded plunger which returns to the top of the pump body.

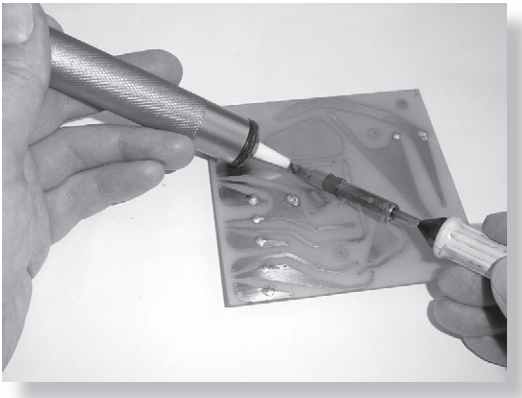

FIGURE 2.18 A desoldering pump in use. As the plunger returns to the top of the pump body, suction it creates sucks the molten solder on the joint inside the pump.

While a little more expensive than desoldering braid, desoldering pumps last considerably longer. Indeed, really the only part of the pump that may be worn by the desoldering process is the pump nozzle, due to the repetitive application of the heat of soldering, which nevertheless will have an operating life of many thousands of operations. Most desoldering pumps have a replaceable nozzle, so even if the nozzle *is* worn, the pump is not necessarily written off. Desoldering pumps are available from as little as £5 to £20.

Drills and bits

Each and every component mounted on a printed circuit board in the conventional way (note I say *conventional way* here: there *are* other methods which don't normally affect the individual electronic circuit builder, most notably the surface mounted component method — see page 36) has its component leads going through holes in the printed circuit board. Thus, the printed circuit board has to be drilled with holes for those very same component leads. The hole size is around 1 mm (give or take a very small percentage — depending on the actual component's leads), which means that your average power drill that you use to drill masonry or steel is not exactly the best tool to use. Power drills are good where the thing being drilled is hard, and the holes don't have to be too accurately positioned.

Printed circuit board, on the other hand, is made from a thin and fairly soft substrate, and the copper track is but a few micrometres thick, while the copper pads you'll be drilling through are not that much bigger than the holes required themselves. So, a drill for the task of drilling printed circuit board has few power requirements, but accuracy is of prime importance.

Small, hand-held, low-voltage drills (complete with mains transformer) are available from most do-it-yourself stores, and these are ideal for hand drilling of printed circuit boards. Figure 2.19 shows such a drill, being used to drill holes in a printed circuit board, prior to mounting and soldering of components.

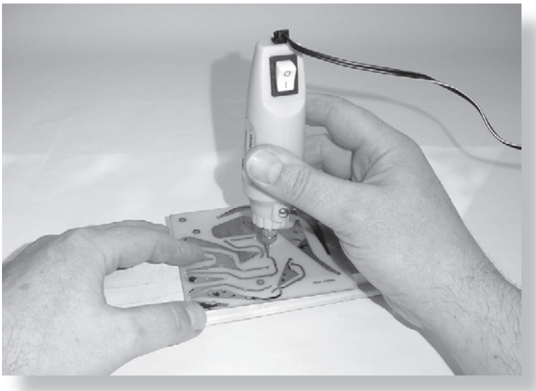

FYI

Make sure you keep a couple of 1 mm and 1.5 mm drills in your toolbox at all times – it's useful when one drill breaks (which they invariably do at the wrong time) to have another one ready to use. Oh, and make sure all your drills are sharp.

Figure 2.19 Using a hand-held low-voltage drill to drill holes in a printed circuit board.

Later in the book we'll see that larger power drills do actually have a place in the electronics circuit builder's toolbox, for drilling of large holes in housings for example. But you *do* need to have a low-voltage, hand-held drill for the finer tasks like that in Figure 2.19. They are not that expensive — starting at around £10 and going up to about £30. Even the cheapest will do the job effectively. Usually they are sold in kits, complete with a range of drill bits, together with small grinding wheels, polishing wheels and engraving tools. If you buy a kit, complete with a range of drills, then you will probably have all the drill sizes you need. If you buy the drill alone, make sure you also buy drills of 1 mm and 1.5 mm sizes — they are the most common ones you'll need.

Printed circuit board stand

When working on a printed circuit board by hand, it's not always easy to mount and solder components without the board moving along the worksurface as you work. If you have a hot soldering iron in one hand, and solder in the other hand, the last thing you want is your printed circuit board sliding along the bench and falling to the floor. Face it: you could do with a third hand! An alternative is to use a printed circuit board stand, into which your printed circuit board is clamped. The board can be repositioned easily, and even turned over to suit.

A typical stand is shown in Figure 2.20, which is a miniature vice with a lever-operated sucker arrangement on the bottom. The sucker allows the vice to be clamped to a work surface, while the vice holds a printed circuit board safely, leaving your hands free to insert and solder components.

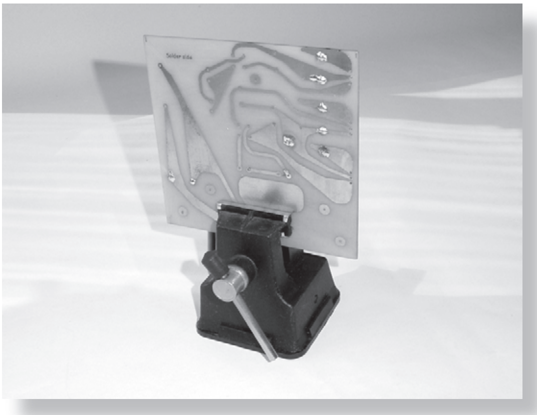

Figure 2.20 A printed circuit board stand.

Many other types of stand are available, ranging from a simple box that holds a printed circuit board, to a multi-armed arrangement complete with magnifying glass and clips to hold other components.

IC insertion tool

As you construct electronic circuits on printed circuit boards, you will come across integrated circuits (ICs — see page 53). Integrated circuits are components that contain many miniaturized components within. They are quite small, but usually have many relatively fragile connection pins (most commonly in two rows — the *dual-in-line* arrangement).

Inserting such integrated circuits into a printed circuit board is a little tricky, as the rows of connection pins are angled somewhat away from the vertical plane, so need to be sprung in as they enter their corresponding holes in the printed circuit board. It's perfectly possible to do this by hand with care, but an IC insertion tool (such as that shown in use in Figure 2.21) is well worth the cost (often less than £1) and makes the task a bit of a doddle.

In use, the integrated circuit is pushed into the tool (which keeps the connection pins at the correct angle and distance), then the tool is pushed down onto the printed circuit board to engage the integrated circuit into its correct position.

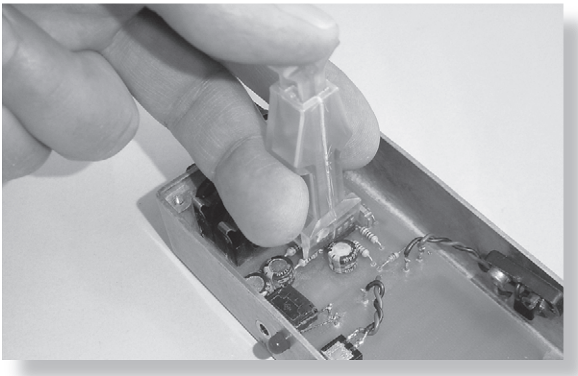

FIGURE 2.21 An integrated circuit insertion tool in operation.

IC extraction tool

Getting an integrated circuit back out of a printed circuit board is an equally tricky task. However, there are equally cheap extraction tools — one is shown in use in Figure 2.22. Just hook the tool at each end of the integrated circuit, and pull. Of course, if an integrated circuit is soldered directly into a printed circuit board, you'll need to desolder it first (see earlier). However, the integrated circuit shown in Figure 2.22 has been inserted into a holder. Integrated circuit holders provide a push-fit method of inserting integrated circuits (see Chapter 3).

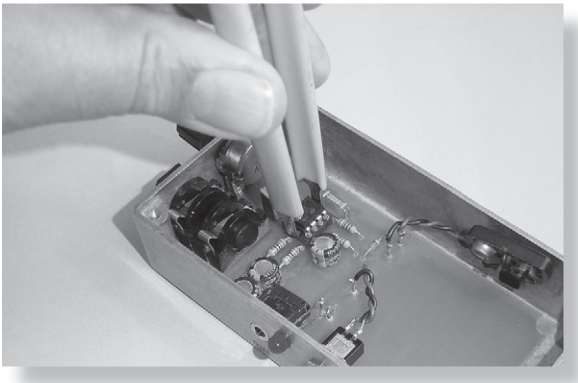

FIGURE 2.22 Using an extraction tool to pull out an integrated circuit – note the use of an integrated circuit holder, which means the integrated circuit itself is not soldered into the printed circuit board.

Terminal pin pusher

When connections are needed on a printed circuit board, it's common to use push-fit terminal pins (see Chapter 3) in the printed circuit board. It's then easy to solder a connecting lead to the pin (see Chapter 5) as required. However, while it's possible to push these pins into place in a printed circuit board using pliers, say, a good tool to buy is a terminal pin pusher — shown in use in Figure 2.23.

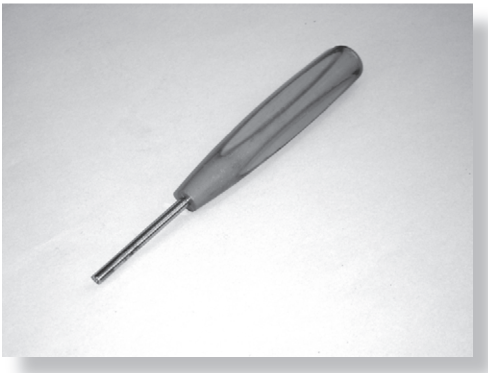

FIGURE 2.23 Terminal pin pusher – just put the pin loosely in place in the printed circuit board, then push home firmly with the pusher.

Spot-face cutter

When using stripboard to build a circuit on, it's inevitable that some of the strips need to be broken — particularly when siting a dual-in-line integrated circuit, which has two rows of parallel pins. In this case, if strips aren't broken between the rows of pins, short circuit between the pins will occur.

Breaking the strip can be done with a sharp craft knife, but this is messy, and a potentially dangerous method. Best to buy a spot-face cutter, which is a tool with a rotating cutting edges. Figure 2.24 shows a common type. Chapter 4 shows how to use one.

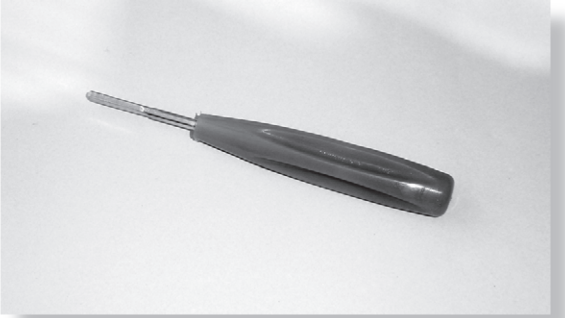

FYI

While stripboard is fine for prototype work, I would not condone it for use in a final project. Always use printed circuit board to produce best end results.

FIGURE 2.24 A spot-face cutter, used to break strips in stripboard.

3

Components

There are many types of components used in electronic devices — the main types are listed below. You need to know what each of them are, what their circuit symbols are, how they function, how to identify them, and what their units of measurement and values are.

Component types

- **RESISTORS**
- **CAPACITORS**
- **SEMICONDUCTORS**
- **SWITCHES**
- **INDUCTORS**
- **CONNECTORS**

There are other types of components, but those listed here and described in this chapter are the most important.

Resistors

The resistor is one of the most (if not *the* most) common yet important components used in electronics. Its primary function is to resist the flow of electricity in a defined manner.

To do this it has a resistance, specified in the unit of ohms (named after the physicist Ohm, who did considerable early work into the qualities of resistance) and given the abbreviation Ω. So, a resistor's value is given as a number followed by the ohms symbol eg, 100 Ω. Occasionally, the abbreviation Ω is changed to the abbreviation R — simply because the abbreviation Ω is not always easily produced in word processing applications on some personal computers. Nevertheless, you should know that the R stands for Ω.

Values of resistors range from a fraction of an ohm, through to many millions of ohms. Where thousands of ohms are specified, it is usual to use the unit of kilohms (a kilohm is 1000 ohms), so that say, 2 kΩ is used to specify 2000 ohms), or 470 kΩ specifies 470,000 ohms. Likewise millions of ohms are known as megohms, so that say, 2.2 MΩ is used to specify 2,200,000 ohms, and so on. With kilohms and megohms it has become usual to drop the superfluous Ω, thus values would be written 2.7 k, 470 k, 2.7 M and so on.

On some circuits (and even in some texts or lists), in an effort to decrease the digits and hence size of labels and complexity of the circuit, the ohmic unit is used to indicate the decimal point position — so 8.2 Ω would be written 8R2, 2.7 k would be written 2k7, and 3.3 M would be written 3M3.

Common symbols for a resistor in circuits are shown in Figure 3.1. The correct British standard is the rectangular box shape of Figure 3.1(a), though the zigzagged shape of Figure 3.1(b) is often used, so you must be aware of and be confident with both.

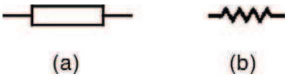

FIGURE 3.1 Resistor symbols commonly found in electronic circuits (a) correct British standard (b) zig-zag shape also used.

Apart from a resistor's value — which is by far the most important criterion — there are other important differences which define how well a given resistor works in a given circuit:

- **TOLERANCE — resistors are rarely the exact value specified.** Usually resistors are within a small percentage of the specified value — this small percentage is called their tolerance. So, a resistor of value 1000 Ω, which has a tolerance of ±5%, will have a value between 950 Ω and 1050 Ω

- **POWER RATING — as electricity flow causes heat.** If too much heat is generated in a resistor, the resistor may be damaged. Therefore resistors have a power rating (specified in units of watts, and given the abbreviation W) which must not be exceeded. A typical power rating for resistors used in common electronic devices is 0.25 W — in all but the most power hungry of circuits this will usually suffice. In fact, lower power resistors (eg, 0.125 W) will probably do — and are of course smaller

- **STABILITY — the ability of a resistor to keep its value over time.** Despite other changes (such as heat, humidity) that can occur around it, it should be able to maintain its resistance value as closely as possible.

Resistor types

There's a few important types of resistor, classified by how they are created.

Carbon resistors

There are actually two main types of carbon resistors, but they look similar if not identical — see Figures 3.2 and 3.3.

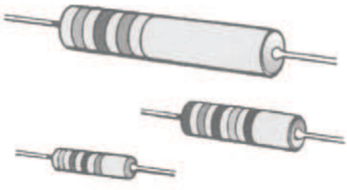

FIGURE 3.2 Carbon resistors – illustration.

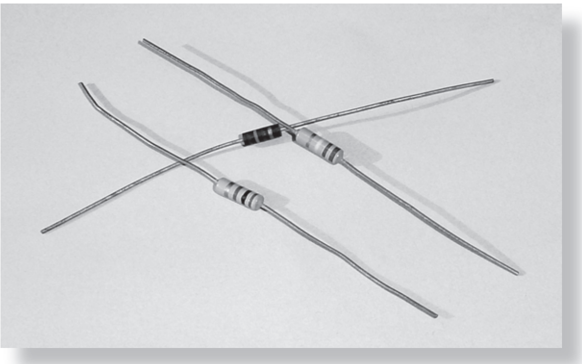

FIGURE 3.3 Photograph of carbon resistors.

Carbon composition resistors are made from a mixture of carbon and clay. Carbon conducts electricity but clay does not, so by mixing the two, resistors are formed. Stability and tolerance of resistors made this way are not particularly good, but they are cheap to make so are very common.

Carbon film resistors are made by moulding a layer of carbon material onto an insulating rod then etching a spiral track through the carbon till the resistance increases. Tolerance and stability are good.

Metal film resistors

These are made in much the same way as carbon film resistors, but tolerance and stability are much better so they are used in circuits which require great accuracy. They are however, more expensive. Figure 3.4 shows illustrations of some metal film resistors, while Figure 3.5 shows a photograph.

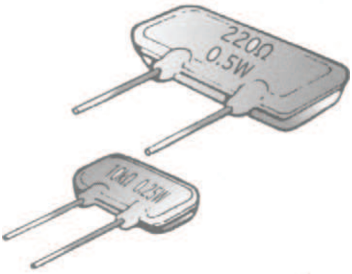

FIGURE 3.4 Metal film resistors – illustration.

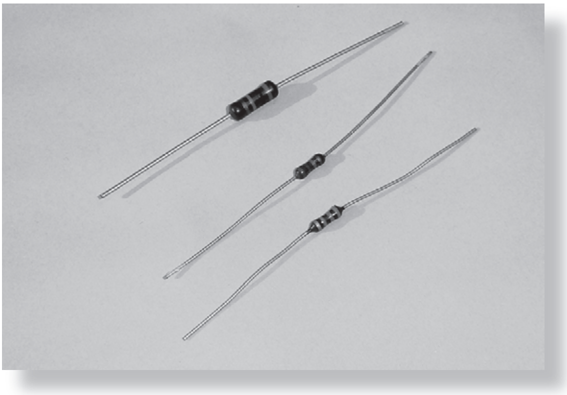

FIGURE 3.5 Photograph of metal film resistors.

Wirewound resistors
Wire is wound around and along a ceramic former to make these resistors. The thickness of wire and the size of the former define the resistance of the resistor. While being quite expensive, resistors made this way have the advantage of high power rating. Figures 3.6 and 3.7 show illustrations and a photograph of wirewound resistors.

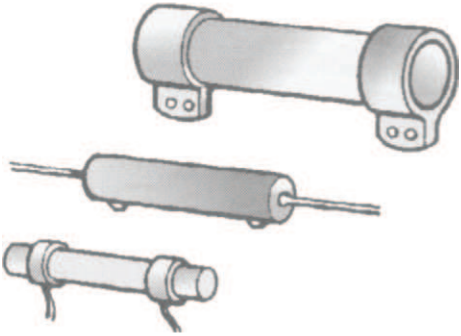

FIGURE 3.6 Wirewound resistors – illustration.

Variable resistors — potentiometers
Often in an electronic circuit there is a requirement for a control knob of some sort (volume, tone, voltage and so on). Variable resistors (usually called *potentiometers*, or just *pots*) can be used for this purpose. They feature a resistive track usually made from a carbon composition mixture in a circular format, with a rotating wiper arm whose position can be

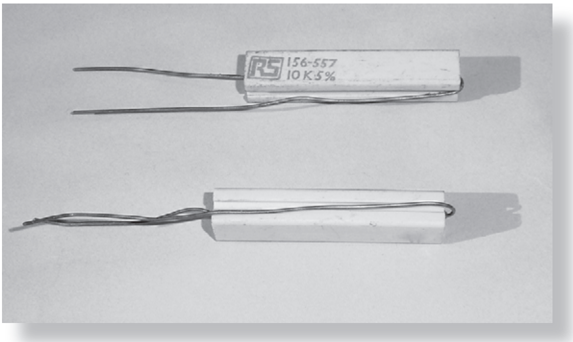

FIGURE 3.7 Photograph of wirewound resistors.

adjusted. There are two main types — standard potentiometers, featuring a central shaft that can be attached to a circuit's front panel for user-control, and *preset potentiometers* which are mounted on a printed circuit board and are adjusted with a screwdriver-type tool. Their various symbols are shown in Figure 3.8.

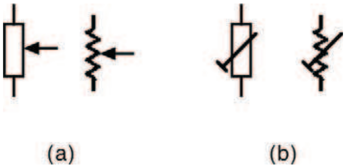

FIGURE 3.8 Symbols of (a) standard and (b) preset potentiometers.

Potentiometers can be of one of two main groups, *logarithmic* (or *log*) and *linear* (or *lin*) — which refers to whether the resistance along the track varies in a linear manner (ie, the resistance varies directly according to the position of the wiper arm) or in a logarithmic manner (ie, the resistance varies logarithmically according to the wiper arm position).

Mostly, linear potentiometers are used. However, in audio circuits in particular — where volume is being adjusted, say — then logarithmic potentiometers are used.

FYI

Volume is a logarithmic variant, in so far as a doubling in volume means a 3dB louder sound. So a sound of say 60dB is double the volume of a sound of 57dB. Similarly, a sound of 63dB is double the volume of the 60dB sound. To represent and control this with a potentiometer, the potentiometer must have a logarithmically scaled track. Logarithmic and linear scaled potentiometers can be used and will work in circuits requiring the use of the other type, but the characteristic of the potentiometer control as you turn it will not be correct.

Dual-gang potentiometers are common, particularly in audio circuits where both channels of a stereo amplifier, say, need to be adjusted simultaneously. These are quite literally made of two potentiometers coupled together with a single control that rotates both wipers together.

Most potentiometers are rotary, but sometimes you will see potentiometers that operate in a line. Such *slider potentiometers* work in exactly the same way — with a resistive track and a wiper arm.

Standard potentiometers are relatively large, and the shaft allows for a control knob to be fitted for user control. Switch mechanisms can also be incorporated into the back of the potentiometer body, which means that the control can be used as an on/off switch, say, at the same time as it being a volume control for example. Preset potentiometers are much smaller and get their name because they are typically set before the completed electronic device is first used, then rarely touched or adjusted again. All potentiometers — whether standard or preset — have three terminals: one for each end of the internal resistor, and one for the wiper arm. It's not always necessary to use all three terminals in a circuit though, and often one of the terminals is left unconnected.

Figure 3.9 shows illustrations of various potentiometers, while Figure 3.10 shows a photograph.

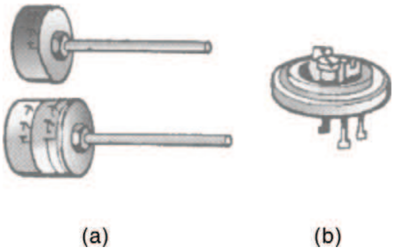

FIGURE 3.9 Potentiometers – illustrations: (a) front-panel fixing types – single- (top) and dual-ganged; (b) a preset potentiometer.

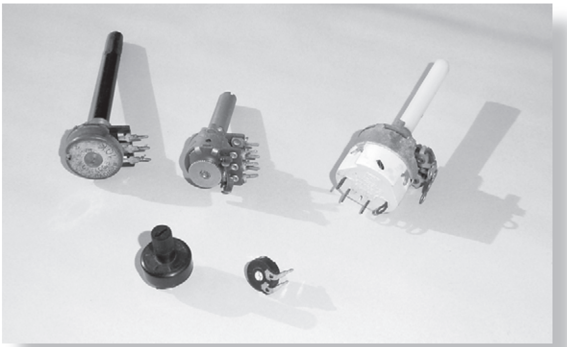

FIGURE 3.10 Photograph of various potentiometers: single-gang, dual-gang, switched single-gang, together with preset potentiometers.

Resistor colour code

Resistance value is occasionally printed numerically on a resistor's body, but normally it is designated in what is called the resistor colour code. The colour code is a collection of coloured bands (normally four bands, although five-band codes exist too) around the resistor body. Reading from left to right, with the first band being closest to the left-hand edge of a resistor, the bands signify:

- **BAND 1 — the first digit of the resistor's value**
- **BAND 2 — the second digit of the resistor's value**
- **BAND 3 — the figure to multiply the first two digits by**
- **BAND 4 — the resistor's tolerance.**

Figure 3.11 shows this graphically. Table 3.1 gives the bands' respective values.

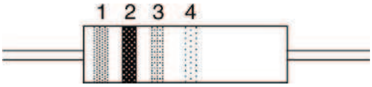

FIGURE 3.11 A resistor showing the four coloured bands that give its value according to the four-band resistor colour code detailed in Table 3.1.

Table 3.1 The four-band resistor colour code.

Colour	Band 1 1st figure	Band 2 2nd figure	Band 3 3rd figure	Band 4 tolerance
Black	0	0	x1	—
Brown	1	1	x10	1%
Red	2	2	x100	2%
Orange	3	3	x1000	—
Yellow	4	4	x10,000	—
Green	5	5	x100,000	—
Blue	6	6	x1,000,000	—
Violet	7	7	—	—
Grey	8	8	—	—
White	9	9	—	—
Gold	—	—	x0.1	5%
Silver	—	—	x0.01	10%
None	—	—	—	20%

Examples using the resistor colour code

Example 1 — let's say the four bands of a resistor are yellow, violet, orange, silver. The resistor value is 4 (yellow) 7 (violet) multiplied by 1000 (orange), which equals 47,000 ohms (more usually written 47 kΩ), with a tolerance of ±10% (silver).

Example 2 — if the four bands are grey, red, green, gold, then the value is 8 (grey) 2 (red) x100,000 (green), which is 8,200,000 ohms (more usually written 8.2 MΩ), with a tolerance of ±5% (gold)

Preferred values

Initially, when considering the resistor colour code, you may well think that it's a virtually impossible task to remember every single value of resistor together with its respective colour code. Calculating every resistor's value from the colour code under these circumstances will be a laborious task. However, it would be impossible for manufacturers to make, and for stockists to stock, resistors of all possible values — there would simply be too many to deal with effectively. As resistors are only made to certain tolerances anyway, and as circuit designers have a reasonable command over the components they specify for use within their circuits, it has become normal to use resistors of only certain values — these are called *preferred values*. The E12 standard range of preferred value resistances follows the series:

1, 1.2, 1.5, 1.8, 2.2, 2.7, 3.3, 3.9, 4.7, 5.6, 6.8, 8.2

As a result, your job of learning the colour code is made significantly easier — because these twelve values are the only ones you need to know. Any resistors you might use of higher values follow in this very same series merely by having a particular multiplying coloured band (eg, 1.8 x 1000 = 1k8, 4.7 x 1,000,000 = 4M7). After only a short while using resistors, you should find that you can almost instantly calculate any new resistor's value, and you won't even need to calculate resistor values you use most regularly — you'll know them!

Note though, that there are other preferred value ranges of resistors — notably the E24 range, which adds an extra twelve values above the E12 range (spaced in-between the E12 range values). You will rarely, if ever, encounter them however, and the twelve values of the E12 range will most likely be the only ones you need to know.

Special resistors

There are several other types of resistors you may encounter in your electronic travels. Some of these are covered elsewhere in the book, but for completeness the main ones are listed here:

Surface mount resistors — there has always been a trend to reduce electronic devices in size. The latest step in this trend is to use miniature components that are soldered directly to a printed circuit board's copper track ie, they are mounted on the board surface (see Figure 3.12), rather than with component leads that go through the board. As such, surface mount resistors are (by their nature) extremely small, so they are not as suited to hand assembly and soldering as conventional resistors with leads. However you may find yourself working with them occasionally.

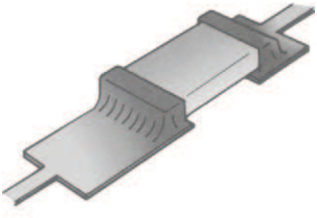

FIGURE 3.12 A typical surface mount resistor – colour coding or lettered markings may be used for component identification (but may not be...). Note that this is illustrated rather larger than full size – actual size is around 2 mm wide!

Thermistors — a thermistor is a resistor whose value changes according to heat. These are useful in applications that detect heat such as electronic thermometers. They vary in size and shape according to construction (they are actually made using semiconductor materials), and typical devices are shown in Figure 3.13. The symbol for a thermistor is shown in Figure 3.14.

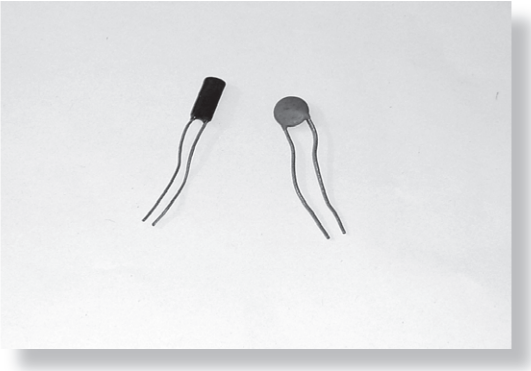

FIGURE 3.13 Typical thermistors.

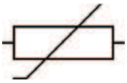

FIGURE 3.14 Thermistor circuit symbol.

Light dependent resistor — like the thermistor, the light dependent resistor (or LDR) is a resistor made from semiconductive materials (typically cadmium sulphide). Instead of its resistance varying with heat however, the light dependent resistor's resistance varies with applied light. Figure 3.15 shows a typical light dependent resistor — a device known as the ORP 12. The circuit symbol for a light dependent resistor is shown in Figure 3.16.

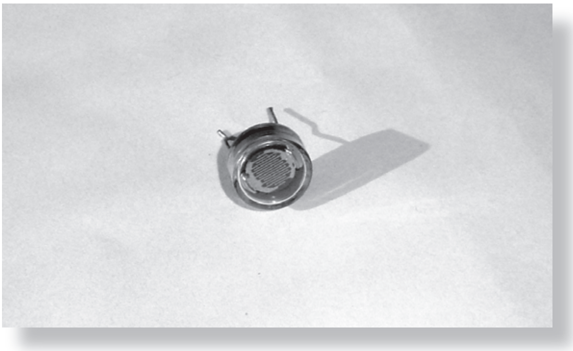

FIGURE 3.15 The ORP 12 light dependent resistor.

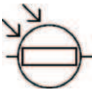

FIGURE 3.16 Circuit symbol of a light dependent resistor.

Capacitors

The second most common component used in electronic devices is probably the capacitor. The basic construction of any capacitor is of two metal (therefore conducting) plates, separated by an insulating layer which is called the *dielectric*. The size of the metal plates, together with the type and thickness of the dielectric, defines the capacitance of the component. Capacitance is measured in units of *farads*, named after the scientist Faraday, and given the abbreviation F. However, the farad is an extremely large unit, and capacitors used in modern electronic circuits are only minute fractions (millionths, thousand millionths, and million millionths) of a farad. One millionth of a farad is known as a microfarad (given the abbreviation µF — µ is the Greek letter *mu*, which is often difficult to produce on some computer systems, so the letter *u* is occasionally used instead), one thousand millionth of a farad is a nanofarad (nF), and one million millionth is a picofarad (pF).

It's useful to be able to switch between the three main sizes of capacitance units, as for example one thousand picofarad is the same as one nanofarad, and one thousand nanofarads is one microfarad. Table 3.2 shows the relationships between the three common units. You really need to be able to switch from one to the other easily, so it's worth getting to grips with it fairly early on in your electronics career. Fortunately, you'll get plenty of practice the moment you start to handle capacitors…

Table 3.2 Relationships between the three main capacitance units in modern electronics.

	microfarad (µF)	nanofarad (nF)	picofarad (pF)
microfarad (µF)	1	1000	1,000,000
nanofarad (nF)	0.001	1	1000
picofarad (pF)	0.000001	0.001	1

Electrolytic vs non-electrolytic

There are two main categories of capacitor — *electrolytic* and *non-electrolytic*. While this sounds tremendously complicated and worrying, all it actually means in practice is that electrolytic capacitors are polarized and must be inserted into a circuit the right way round. They will be marked with a positive and a negative terminal, and must be inserted so that the positive terminal is at a positive potential with respect to the negative terminal. (Note that neither the positive nor the negative terminal need to be at an *actual* positive or negative voltage respectively — they just have to be *more* positive or *more* negative than the other respective terminal!) If they are inserted the wrong way round the insulating layer forming the dielectric can be damaged hence won't work correctly, and the capacitor may even become a short circuit hence causing possible damage to the circuit in which it is installed. Non-electrolytic capacitors can be inserted into circuit either way round. Figure 3.17 shows some variations of electrolytic capacitors.

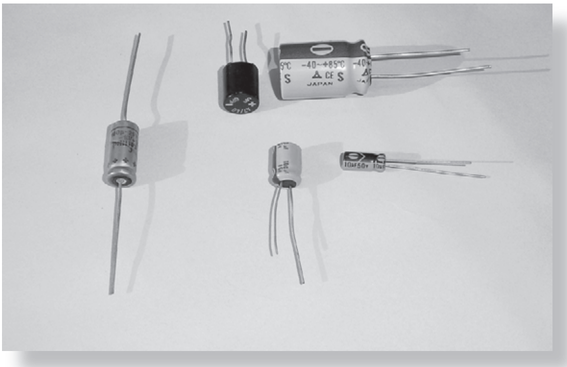

FIGURE 3.17 Electrolytic capacitors.

Electrolytic capacitors are always marked in some way — usually with a + symbol to indicate the positive component lead. Sometimes however, positive is marked with a ridged or ringed end on the component body.

Axial vs radial

Another way in which capacitors are categorized has to do with their shape, and is shown in Figure 3.17. Capacitors which are long, and have their component leads exiting from opposite ends of the body are known as *axial-leaded* capacitors. Capacitors where both component leads exit from one end of the capacitor body are known as *radial-leaded* capacitors.

Non-electrolytic capacitors

There are several categories of non-electrolytic capacitor, illustrated in Figure 3.18. Most of them are known simply by the substance that is used to make up their dielectrics. Common capacitors are *polycarbonate*, *polyester*, *polystyrene*, *ceramic*, *mica*, and *polypropylene*.

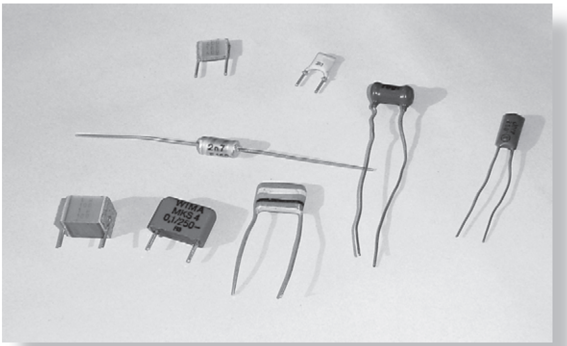

FIGURE 3.18 A variety of non-electrolytic capacitors.

Readers who have been paying attention might be wondering why it's necessary to use electrolytic capacitors at all if they're *that* awkward to use. Well, in theory, it doesn't matter which category of capacitor (electrolytic, non-electrolytic, or which type of dielectric) is used in any given circuit — a 47 nF polycarbonate capacitor has exactly the same capacitance as a 47 nF ceramic capacitor so, indeed, may technically replace it. However, the material used as a capacitor's dielectric has an effect on the capacitor's body size, shape, and component lead-out, so a capacitor of one dielectric simply may not fit into the space in a printed circuit board allocated to a capacitor of another dielectric. Also, in all but a few cases, it would simply not be possible to produce a large value capacitor from a non-electrolytic dielectric, without the resultant capacitor being huge.

Another thing to bear in mind is that the dielectric used (of both electrolytic and non-electrolytic capacitors) can have a significant effect on the component price. For example, *mica* capacitors (sometimes also known as *silvered mica*) are much more expensive to produce than, say, polycarbonate capacitors.

Finally, the dielectric has an effect on the maximum voltage a capacitor may be used at. Capacitors are rated (along with their actual capacitance) with a *working voltage* — which, fairly obviously I hope, should not be exceeded. It is up to the circuit designer to specify what the actual working voltage of the capacitor is. It is up to the constructor of the circuit to ensure that capacitors of at least this working voltage are used.

Variable capacitors

Like resistors and their variable counterparts (potentiometers), capacitors are available in variable forms. Usually though, variable capacitors are in preset form, set up when the electronic device is first made, then rarely touched afterwards. Figure 3.19 shows some examples.

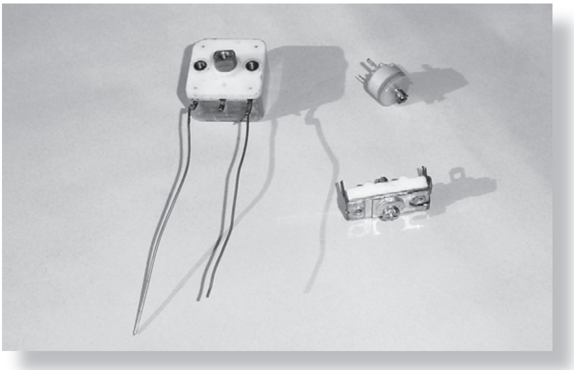

FIGURE 3.19 Variable capacitors.

Capacitor symbols

The symbols used for capacitors in circuits are shown in Figure 3.20, where Figure 3.20(a) shows the symbol for an electrolytic capacitor (note the positive lead is marked +), Figure 3.20(b) is the symbol of a non-electrolytic capacitor, and Figure 3.20(c) is that of a variable capacitor. They are fairly self-explanatory, graphically portraying the two plates of a capacitor separated by a dielectric.

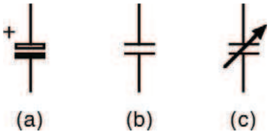

FIGURE 3.20 Symbols for capacitors: (a) electrolytic; (b) non-electrolytic; (c) variable.

Capacitor colour code

Very often, capacitors are marked with their value on the capacitor body. However, some capacitors are marked with a colour code in the same way that resistors are marked (see Figure 3.21), and the code (detailed in Tables 3.3 and 3.4) follows the same general method except that the value is marked in picofarads, not ohms.

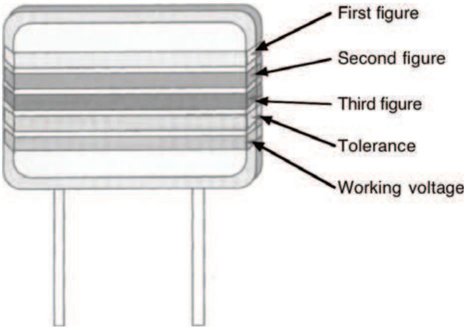

FIGURE 3.21 A capacitor marked with a colour code, showing the bands and their meanings according to the colour code of Tables 3.3 and 3.4.

Table 3.3 The capacitor colour code.

Colour	Band 1 1st figure	Band 2 2nd figure	Band 3 3rd figure	Band 4 tolerance	Band 5 working voltage (see below)
Black	0	0	x1	20%	
Brown	1	1	x10	1%	
Red	2	2	x100	2%	
Orange	3	3	x1000	—	
Yellow	4	4	x10,000	—	
Green	5	5	x100,000	5%	
Blue	6	6	x1,000,000	—	
Violet	7	7	—	—	
Grey	8	8	—	—	
White	9	9	—	10%	

Remember that to convert from picofarads to nanofarads simply divide by 1000. To convert from picofarads to microfarads divide by 1,000,000.

Band 5 of a colour coded capacitor gives its working voltage (ie, the highest voltage that should be applied to it). This does, however, vary with the type of capacitor. Table 3.4 lists capacitor types affected and gives the voltages according to the band colour.

Table 3.4 Capacitor types and working voltages according to the fifth band of the colour code of Table 3.3.

Capacitor type	White	Yellow	Black	Green	Blue	Grey	Pink	Red
Tantalum	3	6.3	10	16	20	25	35	
Polycarbonate		400			630			250
Polyester		400						250

Capacitor number/letter codes

Apart from colour coding, capacitors can also be marked with printed letters and numbers. Often these are fairly self-explanatory — a capacitor marked with the term 8p2 is obviously going to be an 8.2 pF device. However, there are some not-so-obvious variations. For example, a 680 pF capacitor is sometimes marked n68 — which means that it is 0.68 nF (which of course is the same as 680 pF). Sometimes a capacitor may be marked with just three numbers — say, 333. In such a case the first two digits indicate the first and second digits of the value, while the third digit indicates the multiplier (ie, number of zeroes to add at the end). So, the example of the capacitor printed with the number 333 would have a value of 33,000 pF, which is 33 nF.

Tolerances of capacitors are usually not very tight. Often 20% is an acceptable figure, although some capacitors with tolerances of 5% are available. Occasionally, tolerances are marked on a capacitor's body with a letter, such as J (which means 20%) or K (10%) somewhere alongside the value in numbers.

Semiconductors

There are many types of component which are grouped together and known as semiconductors. We'll take a look at some of them now, but it's important to remember that the category is huge, and there are many more semiconductors than we can possibly hope to cover here. However, there are some that are just *so* important that we can not afford to ignore them.

Diodes

Diodes are the first important component in the category known as semiconductors. Indeed they are the simplest semiconductor device. They get their name from the basic fact that they have two electrodes (di — ode, geddit?). One of these electrodes is known as the *anode*: the other is the *cathode*. Figure 3.22 shows the symbol for a diode, where the anode and cathode are marked. Figure 3.23 shows some typical diode body shapes — in which cathodes are identified as the banded or shaped ends.

The name *rectifier* is often used in place of *diode*. Strictly speaking, a rectifier is a diode that is used in high power applications, say, a mains power supply that converts mains alternating current to a low voltage direct current — the conversion from alternating current to direct current is called *rectification*. However, the two names are otherwise synonymous.

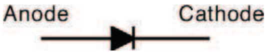

FIGURE 3.22 The circuit symbol for an ordinary diode.

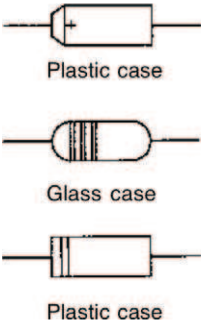

FIGURE 3.23 Some typical diode body shapes.

Figure 3.24(a) shows a diode whose anode is positive with respect to its cathode. Although we've shown the anode as positive with a + symbol, and the cathode as negative with a – symbol, they don't necessarily have to be positive and negative. The cathode could for example be at a voltage of +1000V if the anode was at a greater positive voltage of, say +1001V. All that needs to occur is that the anode is positive with respect to the cathode.

Under such a condition, the diode is said to be *forward biased* and current will flow, from anode to cathode.

When a diode is *reverse biased* ie, its cathode is positive with respect to the anode, no current flows, as shown in Figure 3.24(b). Obviously, something happens within the diode which we can't see, depending on the polarity of the applied voltage, to define whether current can flow or not. Just exactly what this something is, isn't necessary to understand here.

We needn't know any more about it here because we're only concerned with the practical aspects at the moment; and all we need to remember is that a forward biased diode conducts, allowing current to flow, while a reverse biased diode doesn't. Which way you should insert a diode into a particular circuit depends totally on the diode and the circuit. All you have to do is insert it the specified way round.

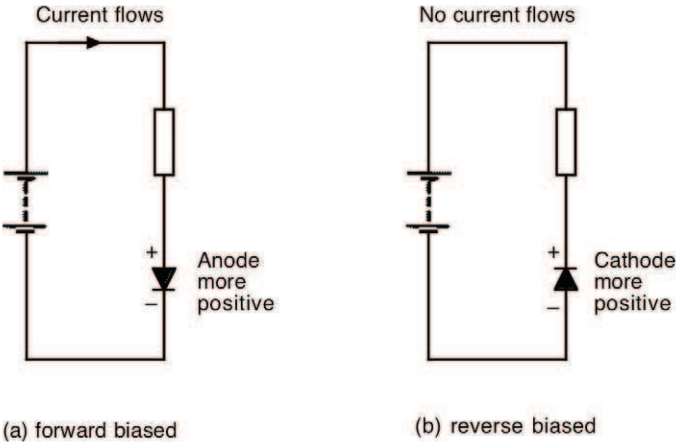

(a) forward biased

(b) reverse biased

Figure 3.24 Simple circuits for: (a) forward biased and (b) reverse biased diodes.

Unlike resistors and capacitors which have particular values, diodes vary only in parameters like the maximum current and maximum voltage they can be used with, and as such are usually specified by a type number — something like 1N4001 or 1N4148. Generally, diodes are marked with their type number, which makes it relatively easy to see what type a diode is. There is one important exception though, in that some diodes are colour banded in a similar way to resistors and capacitors, and use the same colours to indicate digits. However, the four colour bands merely indicate the four digits in the diode's type. Probably the only diode you'll see type-marked in this way will be the very commonly used diode — the 1N4148, which has bands coloured *yellow brown yellow grey* to indicate 4148.

Bridge rectifier

One use of diodes is as rectifiers — devices that convert alternating current to direct current — in power supplies. Usually (but not always), four diodes are connected together to create what is known as a *bridge rectifier*. These can be constructed from individual or discrete diodes, or obtained as a complete unit which mounts directly into the printed circuit board. Several varieties of these complete bridge rectifier exist, and a few are shown in Figure 3.25. Care must be taken when inserting a bridge rectifier into a printed circuit board that it is the correct way round. Markings on the device will indicate the two alternating current input connections (usually with the symbol ~), and positive (+) and negative (–) output connections. Like discrete diodes, the two important considerations when using bridge rectifiers are the maximum working voltage (actually the maximum alternating voltage applied across it) and the current required from it.

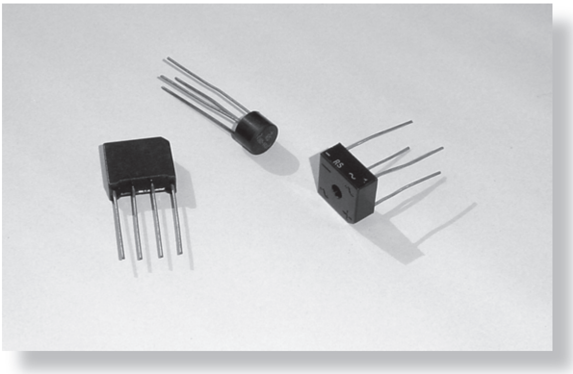

FIGURE 3.25 Various bridge rectifiers.

Transistors

The next important step up the semiconductor evolutionary chain is the *transistor*. Invented over 50 years or so ago, it has been the single most important electronic device ever to have been made. All subsequent electronic devices have been developed from the transistor, and even the most complex of devices such as a computer relies totally upon it still. Nowadays, however, individual transistors are rarely used. Instead several (hundreds, thousands, even millions!) are often combined into complete devices known as *integrated circuits* (we'll look at integrated circuits shortly) and used almost without knowledge or care of what is inside them.

Despite being used rarely, it's still just possible that you'll come across an individual transistor, so we need to look at them now, and fully understand them, so that you can also see how they are used in more complicated devices.

A transistor is shown in Figure 3.26. From the photograph, you'll see that it's pretty small and has three terminals. These terminals are called *base*, *collector* and *emitter* (often shortened to b, c and e). When you use transistors in electronic circuits it is essential that these three terminals go the correct way round.

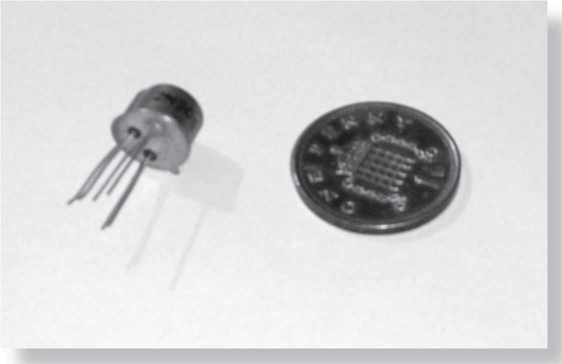

FIGURE 3.26 **A transistor next to a UK one penny coin.**

The problem arises in that different transistors have different body shapes (see Figure 3.27) and different terminal leadouts, so to make sure that the transistor is the correct way round, manufacturers supply them in a standard range of body shapes, and each body shape has a standard terminal lead-out. So, by knowing what body shape a transistor is you can work out its terminals from a chart. Common varieties you'll even learn by heart automatically, just by using them a couple of times.

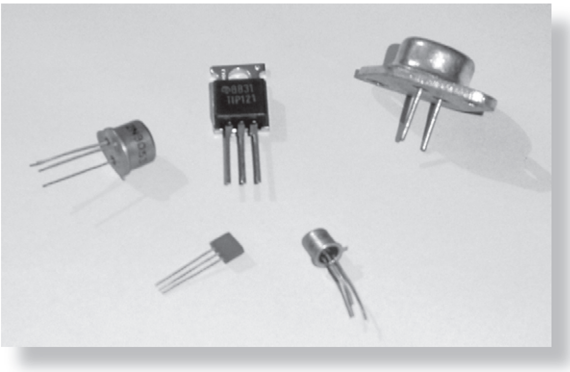

Figure 3.27 A selection of transistors, showing some of the variance in body shapes available.

Figure 3.28 illustrates leadout details for a selection of common transistors. From those shown, you can see that it's easy to work out terminals. For most transistors, leadouts are conventionally given as a plan view looking at the transistor's underside ie, with the transistor's leads pointing towards you. The exceptions are certain types of power transistors, where you view the transistor looking towards the metal pad of the transistor as an elevation view.

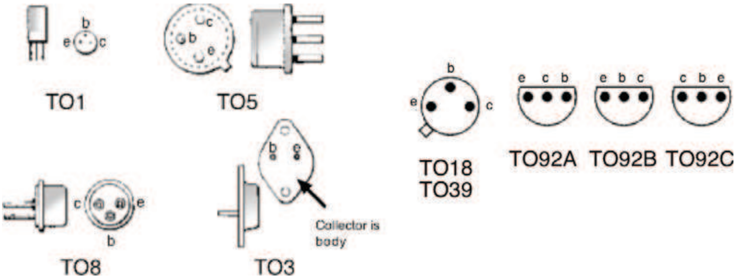

Figure 3.28 Leadouts of some commonly used transistors. Manufacturers of transistors can provide details of all their devices.

There are several different types of transistor, the most popular being what is called the *bipolar* transistor. These are constructed in two different ways, as three layers of semiconductor material — as a thin layer sandwiched between two thicker layers, as shown in Figure 3.29. The semiconductor layers are of two types — called *N-type*, and *P-type*. So the transistor can be made with a thin P-type layer sandwiched between two N-type layers, or with a thin N-type layer sandwiched between two P-type layers — which way defines the transistor's type as either an NPN transistor, or a PNP transistor.

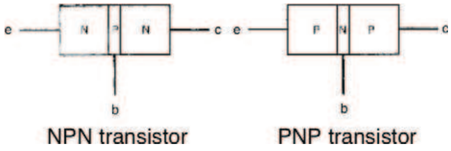

FIGURE 3.29 An NPN and a PNP transistor make-up.

The circuit symbols for the two types of bipolar transistor — NPN and PNP — are shown in Figure 3.30, along with their corresponding terminals. Usually, in circuit diagrams, terminals are not given, and it's taken as read that you will know which terminal is which.

Although the internal working of transistors is pretty complex, we can look on them quite simply — they are electronic switches, in which a small current through the base terminal creates a much larger current through the collector and emitter terminals. The small current through the base terminal is by convention called the *base current*. The much larger current through the collector and emitter terminals is called the *collector current*.

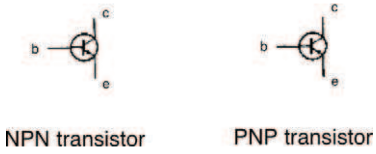

FIGURE 3.30 Symbols of NPN and PNP transistors.

The way bipolar transistors work is illustrated in Figure 3.31(a) for an NPN transistor, where the base current flows into the transistor's base causing a large collector current into the collector. Figure 3.31(b) shows a small base current out of a PNP transistor's base causing a large collector current out of the collector terminal.

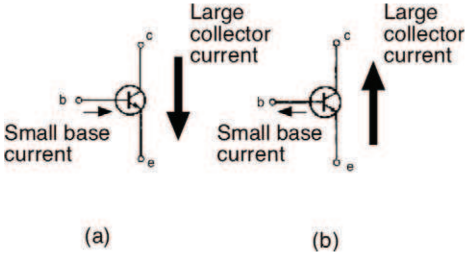
Large collector current
c
b
Small base current
e
Large collector current
c
b
Small base current
e
(a)
(b)

FIGURE 3.31 Showing how both NPN and PNP bipolar transistors work – a small current causes a large collector current.

Field effect transistors

Another major group of transistors are called *field effect transistors* (abbreviated to FET), of which there are two main types: the *junction gate field effect transistor* (JUGFET, or J-FET) and the *metal oxide semiconductor field effect transistor* (or MOSFET).

Like bipolar transistors, FETs can be of N- or P-types, but are actually known as N-channel or P-channel types. The three terminal connection leads are the *gate*, *drain* and *source*, which are roughly the equivalent of base, collector and emitter. MOSFETS are available with two gate terminals, and are called *dual-gate MOSFETS* (or DG MOSFETS).

Figure 3.32 shows the circuit symbols of a range of FETs.

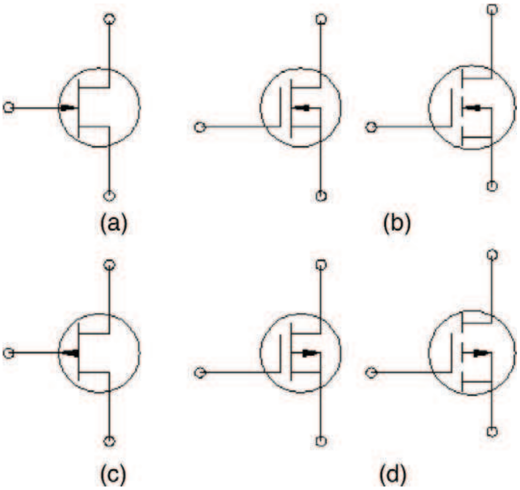

FIGURE 3.32 FET circuit symbols: (a) N-channel J-FET; (b) N-channel MOSFETs; (c) P-channel J-FET; (d) P-channel MOSFETs.

Thyristors

It's sometimes necessary to control current flow in circuits. A specific device to do this is the *silicon controlled rectifier* (or SCR), which also goes by the name *thyristor*. Much like a transistor in appearance, with the terminal *anode* (A), *cathode* (K) and *gate* (G), the thyristor operates like a rectifier or diode that can be switched on or off. Figure 3.33 shows the circuit symbol of a thyristor, while Figure 3.34 shows some common thyristor leadouts.

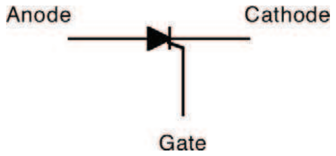

FIGURE 3.33 Thyristor circuit symbol.

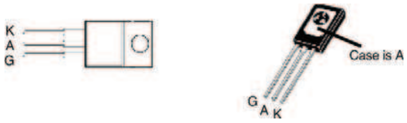

FIGURE 3.34 Some thyristor leadouts.

Triacs

Where thyristors are used to control direct current, *triacs* are used to control alternating current. Their terminals are known as *main terminal 1* (MT1), *main terminal 2* (MT2) and *gate* (G). Circuit symbol for a triac is shown in Figure 3.35, while Figure 3.36 shows some common triac leadouts.

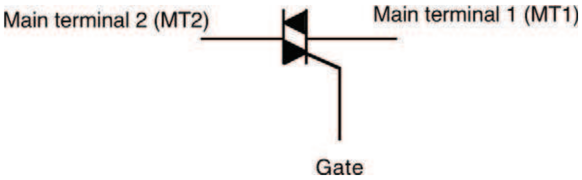

FIGURE 3.35 Triac circuit symbol.

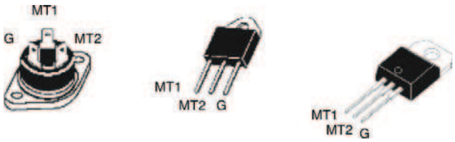

FIGURE 3.36 Some triac leadouts.

Integrated circuits

There are many, many different types of *integrated circuit* (abbreviated to IC) — far too many to describe in this book. Fortunately, at least for beginners to the subject of electronics, it's possible to generalize somewhat and identify perhaps a handful of integrated circuit types that are worthy of note.

For a start, most integrated circuits you'll come across are in a particular body shape — the *dual-in-line* integrated circuit (or DIL IC). These have two rows of terminal pins which are used to connect the internal circuit of the integrated circuit into an electronic circuit.

An example of a very common integrated circuit (the 555, as it happens) is shown in Figure 3.37, where you can see that it's got a total of eight terminal pins, ordered in two four of four. The rows of terminal pins of standard dual-in-line integrated circuits like the 555 are spaced 0.3 inch apart, and pins in the rows are 0.1 inch apart.

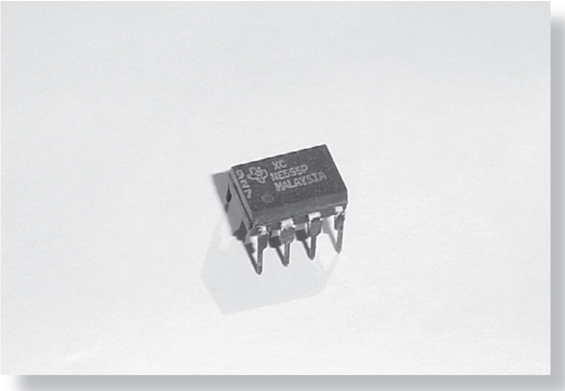

FIGURE 3.37 A 555 integrated circuit.

All dual-in-line integrated circuits like the 555 have their terminal pins numbered in the same way — pin 1 is top left, and all other pins follow in an anti-clockwise direction. All that's needed is to know which pin is number 1. Unfortunately, here, there is some difference between integrated circuits manufacturers. Figure 3.38 shows the most common methods of manufacturer markings for dual-in-line integrated circuits — note that (unlike transistors and their associated semiconductor components) integrated circuit pin diagrams are as viewed from the top of the integrated circuits.

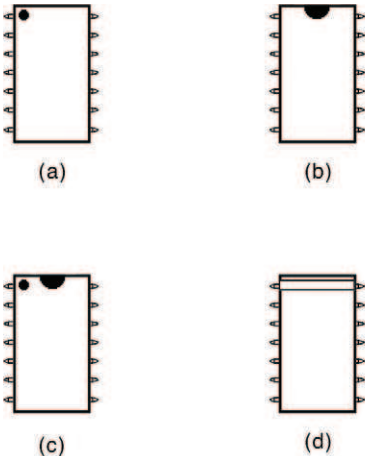
(a) (b) (c) (d)

FIGURE 3.38 Terminal pin 1 on dual-in-line integrated circuits is indicated by: (a) a dot; (b) a notch; (c) a dot and notch; (d) a line. Manufacturers vary, so you need to know the various methods.

Integrated circuit types

Integrated circuits contain highly miniaturized electronic circuits — they can, in fact, contain literally thousands of tiny transistors inside their bodies. It should come as no great surprise then to learn that there are literally thousands of different types of integrated circuits, each one designed to do a slightly different job to its neighbours. The exact job an integrated circuit does depends almost totally on the transistors that make up the circuit inside it. As such, it might seem difficult to find out what all these many thousands of types of integrated circuits do.

Fortunately though, integrated circuits fall very roughly into two highly generalized main categories — *analog* (sometimes known as *linear*) and *digital*. Analog integrated circuits are those that can be used in the likes of audio circuits, or some forms of control circuits for example. Digital integrated circuits are those that would be used in the likes of computing circuits, and also some forms of control circuits for example.

As you might expect, there is some overlap between the two categories, which blurs the distinction between analog and digital integrated circuits somewhat.

Typical analog integrated circuits

Perhaps the most common forms of analog integrated circuits are those that comprise the family of integrated circuits known as *operational amplifiers* or *op-amps*.

Operational amplifiers are often illustrated in an electronic circuit diagram with the symbol shown in Figure 3.39(a), where it can be seen that the device has two inputs (marked + and –), and an output.

It's also common to see an operational amplifier illustrated in an electronic circuit diagram inside its integrated circuit body, showing its terminal pins as part of the circuit, as shown in Figure 3.39(b).

Sometimes, the operational amplifier would be shown in a circuit diagram with its actual terminal pins numbered, but the integrated circuit body is not actually shown (Figure 3.39c). All three methods are strictly correct, and it really just depends on the level of detail required in a circuit diagram, along with personal preference, as to which method is used.

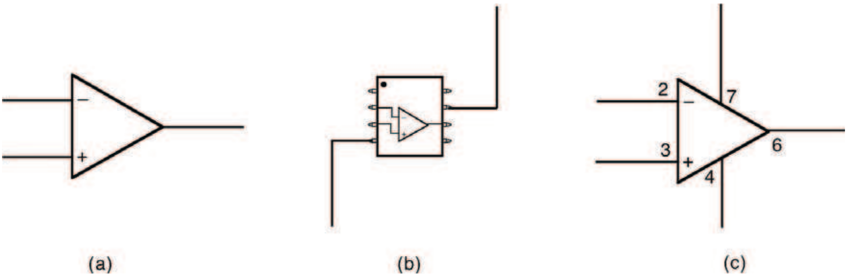

FIGURE 3.39 Circuit symbols for integrated circuits: (a) just the op-amp, with no pin numbers; (b) showing the op-amp complete with integrated circuit 'body' and pin numbers; (c) the operational amplifier with pin numbers.

FYI

While the symbol for an op-amp is relatively simple, you should realize that the op-amp itself may contain many hundreds of components inside the integrated circuit. Treating them collectively as a single component called an op-amp simply makes it easier for us to get our heads round it. An op-amp is a sort of 'black box' electronic component — we don't need to know how it's made up, just how it works. There are many other black box electronic components.

Many integrated circuit op-amps have identical terminal pins, so it's possible that one op-amp can be replaced by another in the same electronic circuit. But there are some differences, so before you use an op-amp, you should refer to manufacturers' data sheets for the op-amp to make sure it will fit in, and work in, the electronic circuit you are building. That being said, most integrated circuit manufacturers make a range of general-purpose op-amps that are sufficiently close in relevant characteristics that they may be classed as identical. For instance, most integrated circuit manufacturers make an op-amp called a 741. A 741 from one manufacturer should work in a 741 designed for another manufacturer's 741. For reference, Figure 3.40 (left) shows a 741 op-amp — it's in a standard 8-pin, dual-in-line integrated circuit format, while Figure 3.40 (right) shows the pin-out diagram for a 741.

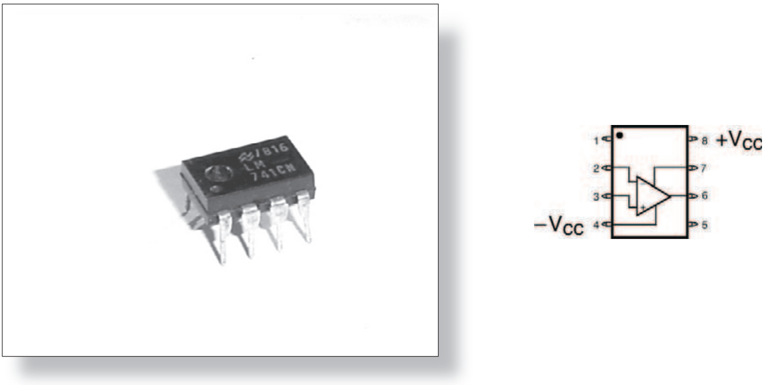

FIGURE 3.40 A 741 op-amp: (left) photograph; (right) pin-out diagram.

Typical digital integrated circuits

Perhaps the simplest forms of digital integrated circuits are those that make up digital gates. Like analog op-amps, these usually have a simple black box symbol, but in reality are made up from many individual transistors inside the integrated circuit. Common logic gates are shown in Figure 3.41. There are, however, many more types of logic circuits available in integrated circuit form. Also, like analog op-amps, logic gates are commonly shown in an electronic circuit with their pin numbers shown.

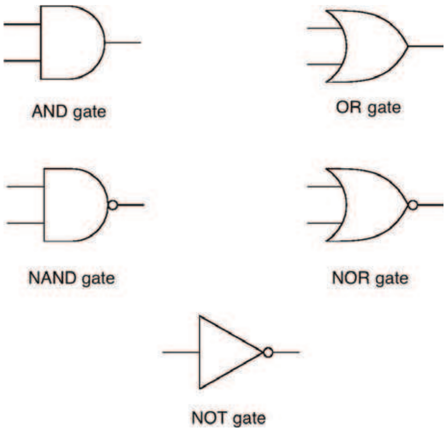

FIGURE 3.41 Common logic gate symbols.

There are several families of digital integrated circuits, generally based on their construction techniques. How the actual gates or logic circuits inside each integrated circuit in each family are connected through their respective terminal pins is defined in standard digital integrated circuit pin-out diagrams. Some digital integrated circuit pin-out diagrams are shown in Figure 3.42 (for digital integrated circuits in the 7400-series), and in Figure 3.43 (for digital integrated circuits in the 4000-series). These are the most common digital integrated circuit families.

Batteries

Electronic circuits need power of some sort to function. Power is sourced from either a power supply (usually mains-powered) or — particularly for portable equipment designs — batteries. Occasionally, a single cell might be sufficient to power an electronic circuit — the symbol is shown in Figure 3.44 (a) — but more usually multiple cells, or power pack batteries are used — the symbol of both is shown in Figure 3.44 (b). Note that the voltage should be shown beside either symbol.

Where cells or batteries are used to power electronic circuits, some form of connector must be used to clip the cell or battery into place. Various battery connectors are shown in Figure 3.45. The type used depends on the cells or batteries required, the circuit itself, and also the housing used to hold all the electronic circuit parts (including the cells or battery).

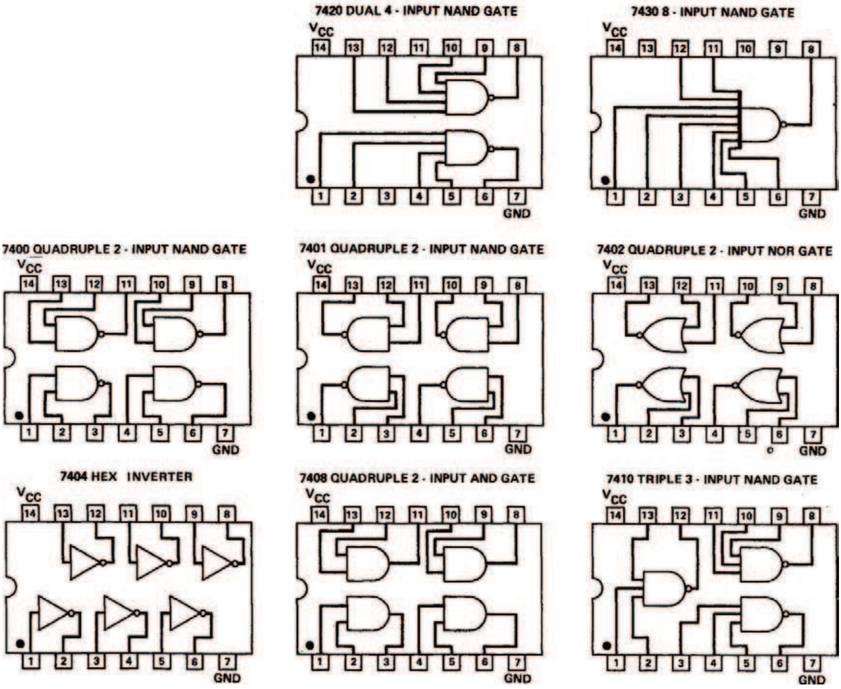
7420 DUAL 4 - INPUT NAND GATE
7430 8 - INPUT NAND GATE
7400 QUADRUPLE 2 - INPUT NAND GATE
7401 QUADRUPLE 2 - INPUT NAND GATE
7402 QUADRUPLE 2 - INPUT NOR GATE
7404 HEX INVERTER
7408 QUADRUPLE 2 - INPUT AND GATE
7410 TRIPLE 3 - INPUT NAND GATE
VCC
GND
14
13
12
11
10
9
8
1
2
3
4
5
6
7

FIGURE 3.42 Some pin-out diagrams for members of the 7400-series of digital integrated circuits.

Indicators

Often, it's useful or necessary to show when an electronic circuit is turned on. This is usually done with an indicator of some sort — either a bulb, a neon, or most often a LED.

Bulbs

Not used very often, low-voltage bulbs can give an effective indication of an electronic circuit's function. However, problems associated with mounting the bulb normally preclude their use. Some bulbs and mounting methods are shown in Figure 3.46.

Neons

Largely for the same reasons we don't often use bulbs, neons aren't often used either — with the possible exception of some mains switches (see later) that contain an integral neon bulb to show when an electronic circuit's power supply is turned on. Figure 3.47 shows an example of such a mains switch.

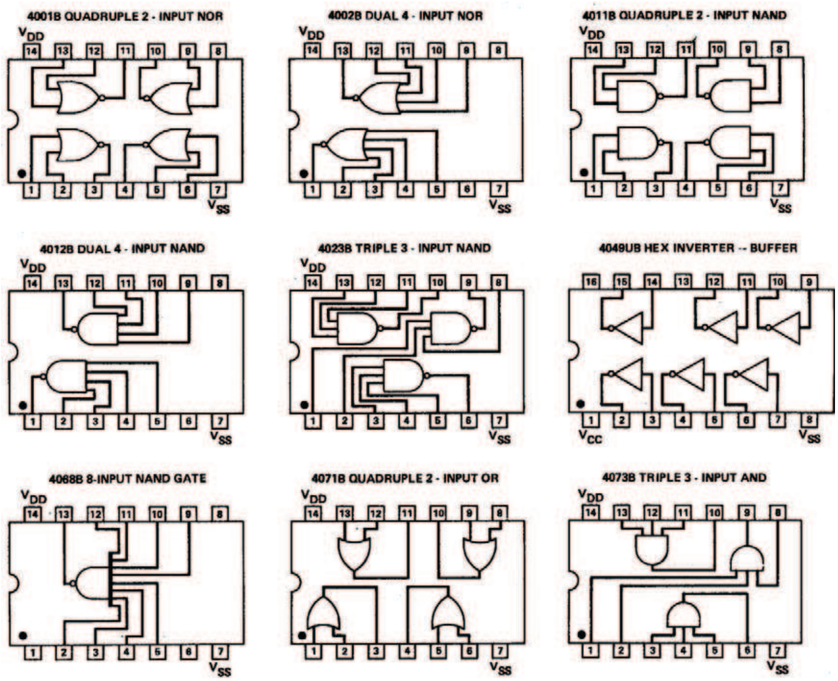
4001B QUADRUPLE 2 - INPUT NOR
4002B DUAL 4 - INPUT NOR
4011B QUADRUPLE 2 - INPUT NAND
4012B DUAL 4 - INPUT NAND
4023B TRIPLE 3 - INPUT NAND
4049UB HEX INVERTER -- BUFFER
4068B 8-INPUT NAND GATE
4071B QUADRUPLE 2 - INPUT OR
4073B TRIPLE 3 - INPUT AND
V_{DD}
V_{SS}
V_{CC}

Figure 3.43 Some pin-out diagrams for members of the 4000-series of digital integrated circuits.

LEDs

Light emitting diodes (LEDs) nowadays form the usual and best method of indication. They are easily used, and easily and cheaply mounted onto housings' front panels, so have become the preferred method. They are — as their name suggests — a particular form of diodes, and must be used with the same precautions as other diodes (see the section on diodes, earlier in this chapter), but when they are forward biased in a circuit emit light. Normally one side of the LED is marked in some way to indicate cathode. Different colours of LEDs are available (green, red, blue, white), and some forms emit light of different colours depending on connections. Figure 3.48 shows a typical round LED, although rectangular LEDs are also available.

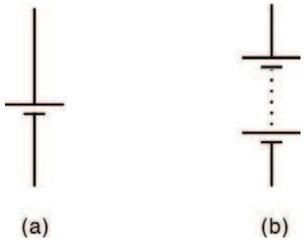

FIGURE 3.44 Battery symbols (a) single cell (b) multiple cell or powerpack.

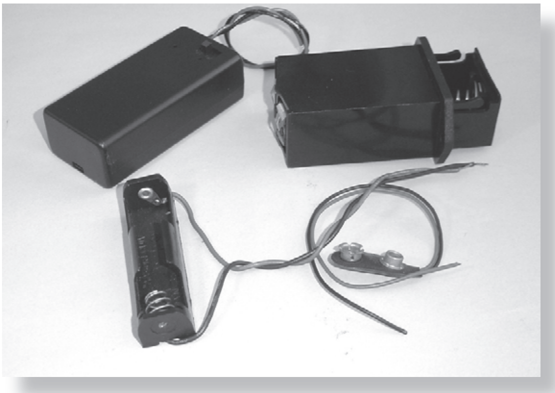

FIGURE 3.45 A selection of cell and battery connectors.

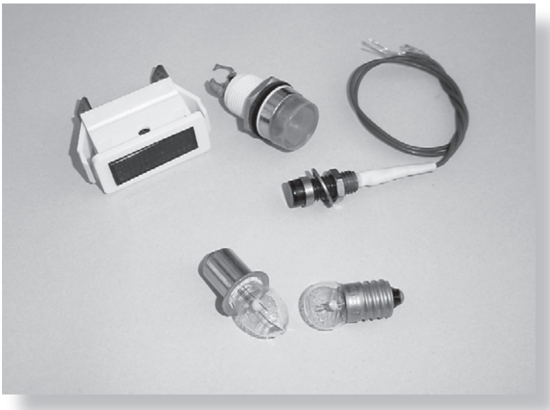

FIGURE 3.46 Bulbs and panel-mounted bulb holders.

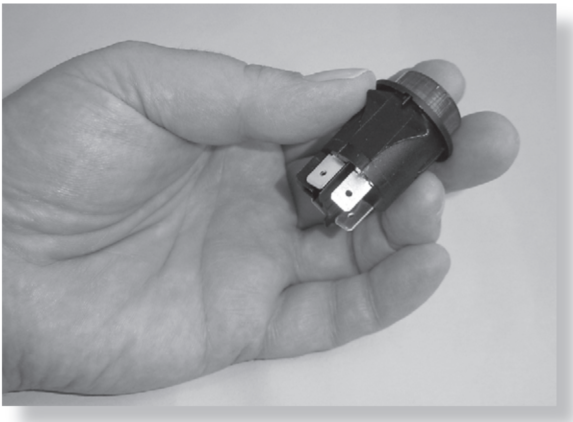

FIGURE 3.47 A mains switch with an integral neon bulb (this particular switch is used in the Bench Power Supply project in Chapter 8).

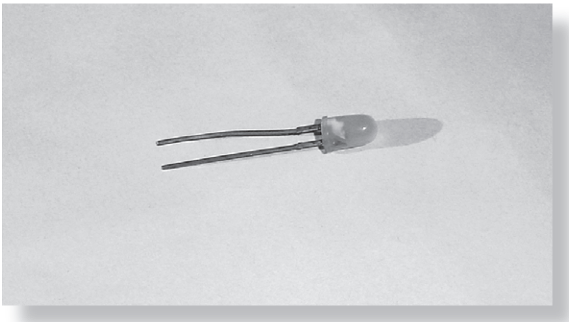

FIGURE 3.48 A typical LED.

Switches

There are many forms of switches used in electronic circuits, and they take many different shapes and sizes. They are all similar however, in that they are all mechanical devices which are physically operated by the user. Their main differences lie in how the internal contacts are arranged, and what form of mechanical device is operated by the user to switch contacts.

Contacts

The simplest switch has two contacts, and a simple on/off mechanism. Its circuit symbol is shown in Figure 3.49 (a). As it can be used in only one connection within a circuit, it is called *single-pole*. Also, as it can be only on or off, it is called *single-throw*. So, in full, it is known as a *single-pole, single-throw* switch (shortened to *SPST*).

Where the switch has a third contact, so the central contact can be switched to either one state or another, it is known as a *single-pole, double-throw* (*SPDT*) switch, and is occasionally called a *changeover switch*, and its circuit symbol is shown in Figure 3.49 (b).

Commonly, switches can be joined (or *ganged*) together in one housing, to allow more than one connection to be switched at the same time. The simplest of these switches would be a *double-pole, single-throw* (*DPST*) switch, the circuit symbol of which is shown in Figure 3.49 (c). The broken line between the two individual mechanisms of the circuit symbol indicates that both mechanisms are mechanically joined, and as one switch operates the other does too.

Circuit symbol of a *double-pole, double-throw* (*DPDT*) switch is shown in Figure 3.49 (d), where the two joined mechanisms are of double-throw format. Again the broken line joining the individual mechanisms indicates that both mechanisms operate in unison.

While these four switch types shown in Figure 3.49 (a) to (d) are obviously not the only switch types, they form by far the most common types you will come across in electronics.

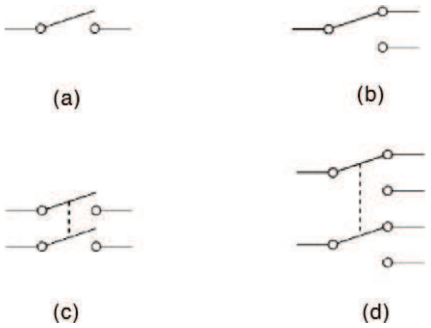

FYI

The switch types in Figure 3.49 are only the most common four. It's possible to get switches with many more poles, and many more throw mechanisms, for specific purposes.

Figure 3.49 Four main switch types: (a) single-pole, single-throw (SPST); (b) single-pole, double-throw (SPDT); (c) double-pole, single-throw (DPST); (d) double-pole-double-throw (DPDT).

Take Note

The contacts of switches of all types have only limited voltage and current ratings, and you must make sure that you do not exceed these contact ratings when using switches. If the circuit a switch is used in is battery-powered, it is unlikely that contact ratings will be exceeded, but if the switch is to be used to switch mains power then contact ratings should be carefully checked before use.

Mechanics

The most common mechanics used in switches in electronics are: slide, toggle, push, and rotary. Examples of all are shown in Figure 3.50. Each type may have a number of deviants, and is usually available in a range of sizes too. Switches may also be incorporated with other functions, such as a potentiometer (see page 31), or indicators (see page 58). Also, each mechanism type can have minor variants in operation — for example push switches can be *push-to-break* or *push-to-make*, and *switched* action or *momentary* action — so it's not hard to see that there are many hundreds of switch types to choose from.

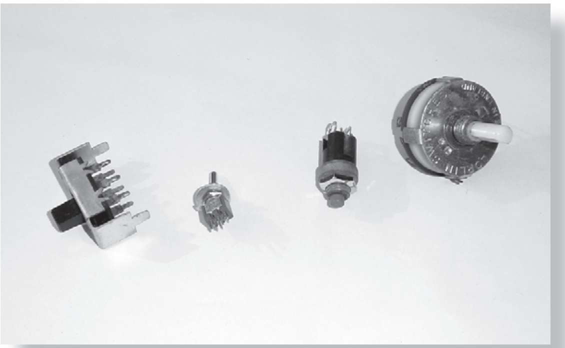

FIGURE 3.50 A range of switches, according to mechanism types, from left to right: slide, toggle, push, rotary.

Relays

Relays are just switches that are operated electro-mechanically. Thus, where an ordinary switch may be operated by you pushing a button, or flicking a toggle, a relay is operated by merely applying an electric current to an electro-magnetic coil. This current creates a magnetic field that flicks the relay's internal switch mechanism. The advantage is pretty obvious — an electronic circuit can operate the relay automatically, without human intervention. A circuit symbol for a single-pole, double-throw relay is shown in Figure 3.51.

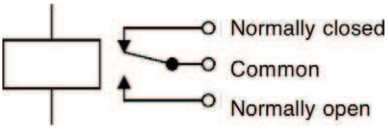

FIGURE 3.51 Circuit symbol of a single-pole, double-throw relay.

Generally, relays are mounted on an electronic circuit's printed circuit board, so are quite small. Several variants exist, and usually it is simply a matter of choosing the relay to suit in terms of number of switch poles and throws, contact ratings, coil operating voltage and current, and size. An example of a relay — a miniature, printed circuit board mounting relay — is shown in Figure 3.52.

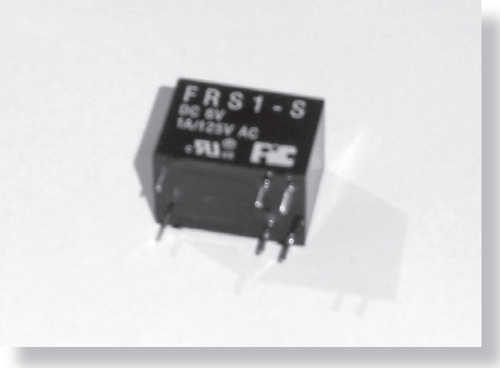

FIGURE 3.52 A miniature, printed circuit board mounting relay.

Loudspeakers and other sound output devices

In an audio circuit, it's often required that we can hear the audio signal. The usual way to do this is either with an in-built loudspeaker, or using earpieces.

Loudspeakers built in to electronic circuits are usually fastened to the housing of the circuit, and once in place are pretty robust. However, they do require some precautions in handling. They work because a cone of some material is vibrated electro-magnetically (due to an electric current passing through an attached coil of wire that is situated in a magnetic field). Problems can arise because the cone is easily damaged or split if mishandled. A couple of examples of loudspeakers for use with electronic circuits are shown in Figure 3.53.

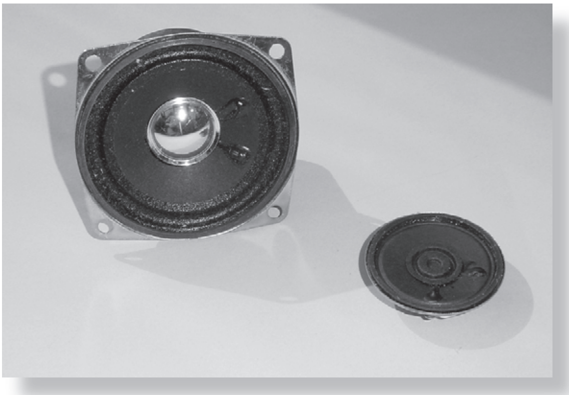

FIGURE 3.53 Loudspeakers typically used in electronic circuits.

Sometimes, while audio sound is not outputted from a circuit, some form of noise is — say, a buzz, or beep. There are several forms of devices for doing this, but only two main groups: *buzzers* and *sounders*.

Buzzers are generally electro-mechanical or piezo-electric sound sources which, when a voltage (generally of direct current) is applied across their terminals, vibrate at a fixed frequency, causing a sound. They come in various shapes and sizes and amplitude levels, some of which are shown in Figure 3.54. The circuit feeding the buzzer merely needs to switch a current to the buzzer to make it work.

Sounders are more like loudspeakers in operation (some are shown in Figure 3.54), and require some form of alternating current signal. Unlike loudspeakers, however (which can successfully operate over a range of frequencies — so can reproduce audio quite well), sounders require a fairly specific frequency signal to generate the loudest sound possible. The circuit feeding the sounder must produce an oscillation at the frequency required by the sounder, to make it work.

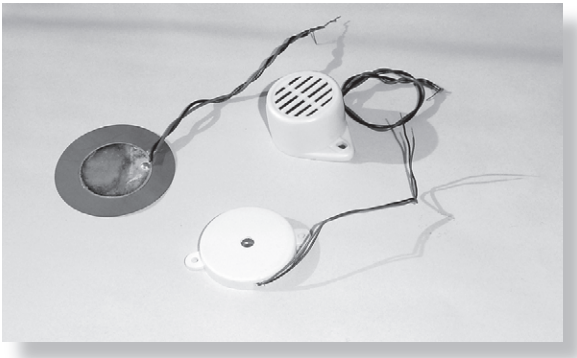

FIGURE 3.54 A buzzer with two types of sounders.

Microphones

Microphones are, in essence, the opposite of loudspeakers and sounders. They convert a sound into an electrical signal, for use by an electronic circuit. They are common in two main forms: *dynamic* and *electret*. Dynamic microphones are just like loudspeakers in reverse; they feature a diaphragm (that is much like a loudspeaker's cone) which vibrates as sound waves hit it, connected to a coil of wire that moves through a magnetic field, producing electric current.

Electret microphones (sometimes known as *condenser* microphones) operate in a vaguely similar way in that they have a polymer diaphragm that creates an electric signal as it vibrates due to impinging sound. However, this signal is too small to be used raw, so they also include an amplifier inside the microphone housing to amplify the signal. While this produces an extremely high quality and small microphone, the downside is that the amplifier must be powered somehow, but a simple low-voltage connection is all that is required (and, as most electronic circuits are powered by a low voltage, extremely easy to include).

The circuit symbol for a microphone is shown in Figure 3.55, while Figure 3.56 shows microphones of the two main types.

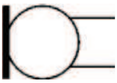

FIGURE 3.55 Circuit symbol for a microphone.

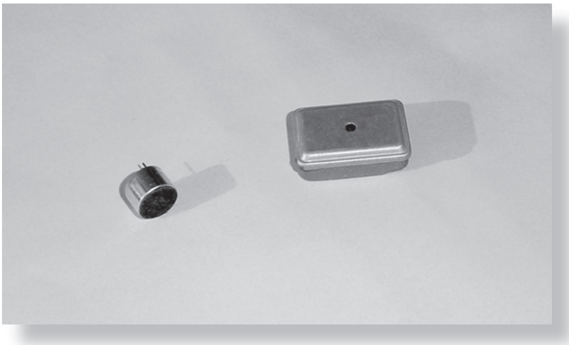

FIGURE 3.56 An electret microphone, and a dynamic microphone.

Coils and inductors

Technically, any coil of wire that has a current applied to it becomes an inductor. So, any electronic component that comprises a coil can also be classed as an inductor. That being said, it's not often that simple coils are used in an electronic circuit (normally only in radio frequency circuits). In most instances, you may even not realize that the component you use features an internal coil.

Examples of coils in common use that you may use are transformers (used, say, in mains power supplies), loudspeakers, and relays (see earlier).

Connectors

Any electronic circuit depends on connectors of many types: to get signals into and out of the circuit; to connect to controls that are on a housing's front panel; and to obtain power from a battery or power supply. The range of connector types is huge, and can't fully be covered here, but we can at least look at a range of common connectors you will almost certainly come across. Connectors are generally (as far as we're concerned here) used either on a printed circuit board, or on the housing used to house a printed circuit board.

Some typical plug and socket connectors for use on printed circuit boards are shown in Figure 3.57. Some typical connectors for mounting on a circuit's housing are shown in Figure 3.58.

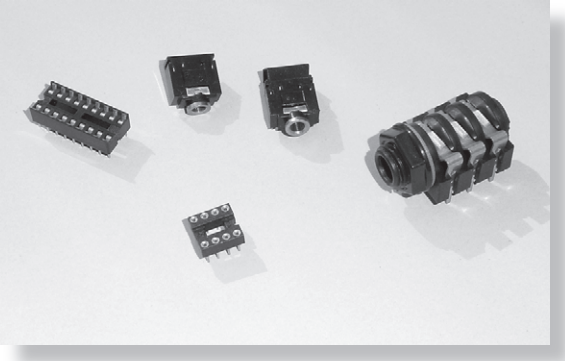

FIGURE 3.57 Typical printed circuit board connectors.

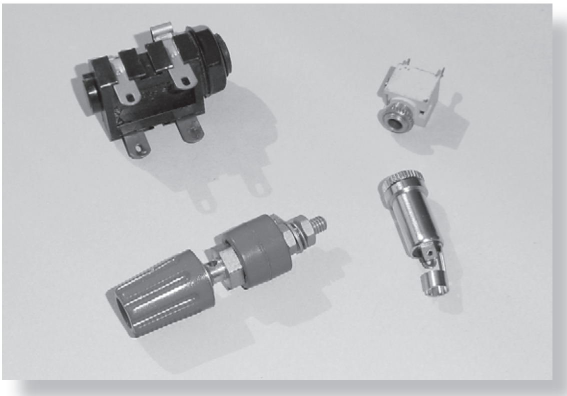

FIGURE 3.58 Connectors used on housings for front- or rear-panel mounting.

Integrated circuit holders

Where integrated circuits are to be fitted to printed circuit boards, it's common not to solder them in. Integrated circuits are, of course, complex, miniaturized devices. Excess heat when soldering may damage them (although some are less sensitive to heat damage than others), so an alternative is to use *integrated circuit holders* (sometimes called *integrated circuit sockets*), which *are* soldered into the printed circuit board. Integrated circuits are then push-fitted into the holders afterwards.

Figure 3.59 (left) shows an integrated circuit holder in place on a printed circuit board, while Figure 3.59 (right) shows the holder with an integrated circuit fitted. Integrated circuit holders are sized the same way that integrated circuits themselves are, so you would use an 8-pin holder for an 8-pin integrated circuit, a 14-pin holder for a 14-pin integrated circuit, and so on. Holders are not expensive — about 15p each — so it's worth buying a few of each common size (8-pin, 14-pin, 16-pin) for your component box.

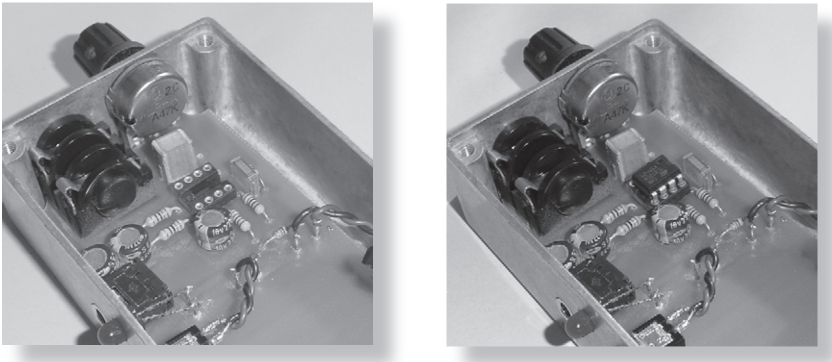

FIGURE 3.59 **Using integrated circuit holders: (left) the holder itself, in position in a printed circuit board; (right) showing an integrated circuit in the holder.**

Terminal pins

One of the main problems with a printed circuit is how to connect any external controls, switches or connection sockets. Often, it's possible and easy to use a connection lead, from the component side of the board to the track side, soldered into place at a junction in the copper foil track. Indeed, Chapter 5 shows in detail how to do this. However, if a printed circuit board has been mounted in a housing (see Chapter 6) then it's not easy to get to both sides of the board to do this.

It's therefore a good idea to use terminal pins. These push through from the copper track side of the printed circuit board to the component side. They are then soldered to the copper track. Thus, when the printed circuit board is mounted in its housing, it's still easy to solder a connecting lead to the pin connection (which is on the component side ie, the upper side, of the board). Chapter 5 shows how to solder a connecting lead to a terminal pin, but in order to do that, of course, you need to buy some terminal pins. They are usually sold in packs of 100 and cost only a few pounds a pack, so are well worth the investment. Various sizes are available, and various types, but the most common 1 mm diameter, single-sided, pre-tinned type is all you'll need, unless you are constructing a circuit that calls for anything more unusual.

Take Note

Don't buy gold-plated terminal pins – they are much more expensive, and in all but the most critical of situations, provide no benefit.

Figure 3.60 shows terminal pins, soldered in place in a printed circuit board.

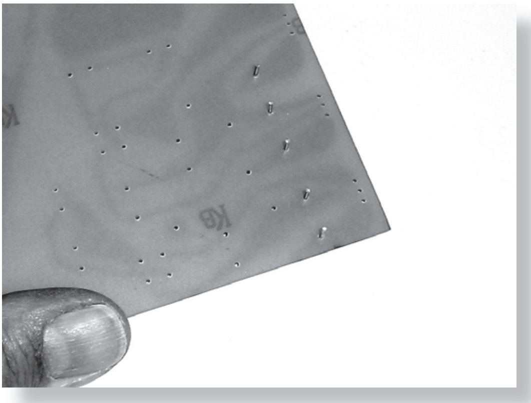

FIGURE 3.60 **Terminal pins in a printed circuit board – the pins are push-fitted into the board, then soldered in place.**

FYI

Terminal pins have a knurled ridge that holds the pin in place in the printed circuit board. The knurled ridge makes them a little hard to push into place however. Pliers, or a hammer, will do the job if you are careful. However, you may think it worth buying a terminal pin pushing tool to help with this job — see Chapter 2.

Printed circuit board links

While hardly being an individual component, it's best to take a look at the use of printed circuit board links. These are used on a printed circuit board if a connection in the circuit needs to be made, but the design of the copper foil track is such that it cannot be made using the copper foil.

In such cases, it's common to use a small link of wire that jumps from one connection point to another on the printed circuit board.

Many connections in the copper foil track can be run underneath the link, so it forms a useful method of bridging parts of the printed circuit board.

Figure 3.61 shows a printed circuit board with several links in place — the board shown is actually the Heart Flasher project printed circuit board, detailed in Chapter 8.

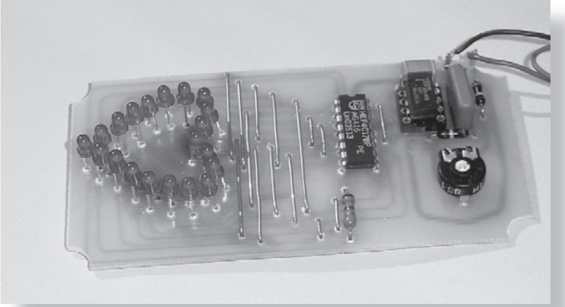

FIGURE 3.61 Links on a printed circuit board.

Printed circuit board links *can* be made with insulated connection lead, but this tends to make a rather messy board. Instead, it's best to use uninsulated single-strand wire. Being rigid, links don't bend and become untidy the way insulated wire would, and there is little risk of links touching other links or neighbouring components. Buy a reel of single-strand wire, then cut short pieces off to make links — see Chapter 5 for details.

Take Note

When designing a printed circuit board which requires links, make sure that links run parallel – in this way there is no danger they will touch other links close by.

4

Printed circuit boards

Most of the projects you'll ever want to make up from electronics magazines or electronics books like this one are constructed on printed circuit boards (PCBs). Without any doubt printed circuit boards help you make really neat and reliable projects, simply because they hold all the components of the projects in a fixed and long-standing way, with strong soldered joints giving the good electrical connections required in the circuits. You've only to take the back off any modern electronic appliance such as your television, radio, hifi amplifier, computer (carefully — and after you've unplugged it from the mains and assuming it's yours!) to see that it's not just book and magazine projects which use printed circuit boards, but manufacturers of electronic appliances do, too. Figure 4.l shows a printed circuit board in a typical modern appliance.

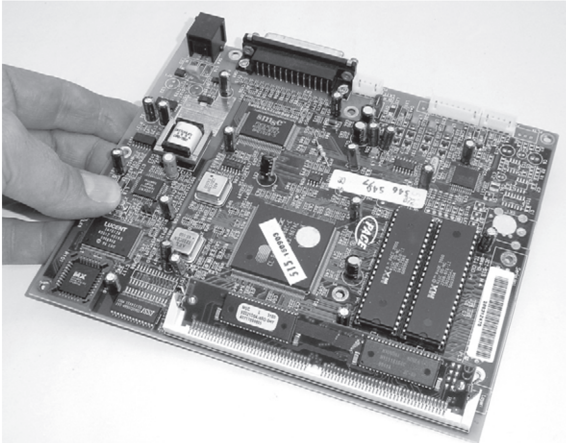

FIGURE 4.1 A printed circuit board - the type you might find in a television or other modern electronic appliance.

There are loads of benefits when using printed circuit boards:

NUMBER 1 Positioning components. Each component in the circuit has its own special little place on the printed circuit board — if you've forgotten to insert it, it's fairly obvious there's something missing because there is a space and a couple of holes

NUMBER 2 The components fit close up to the board. Component leads go through holes in the board to the other side, where they are soldered to strips of copper — in other words, the components are rigidly held so they can't easily work loose, and there's not much chance of one component lead shorting against another

NUMBER 3 Size. Yet another benefit is simply the sheer compactness of the completed assembly — as connections are made with copper tracks on the underside of the board (some printed circuit boards have copper tracks on both sides — and are called, naturally enough, double-sided printed circuit boards) and as components can be very close together without the risk of short circuits, the size of the printed circuit board need be very little more than just the total surface area of the components.

If you have an old valve radio or television around the place, you can compare equipment with modern printed circuit boards to the method used then; where all the components were simply strung between connection points on 'tag strips' underneath a metal chassis, in a seemingly haphazard arrangement. While such methods served their purpose in their day, no one can doubt that printed circuit boards are infinitely better.

Copper-clad printed circuit board

So what exactly is a printed circuit board, what's it made of, how's it made, and how do we set about getting it?

Well, the photograph in Figure 4.2 shows a small piece of printed circuit board, directly as it would be obtained from a printed circuit board manufacturer, or as you might buy from an electronics components shop. This small piece would have been cut from a much larger sheet (typically a few metres square). In its raw state like this, the printed circuit board is

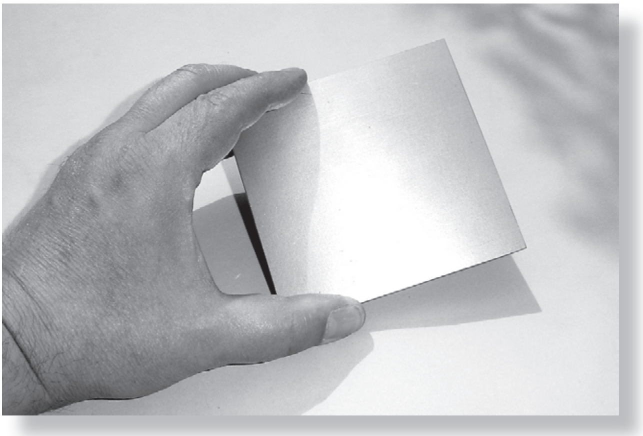

FIGURE 4.2 Raw printed circuit board, comprising a thin layer of copper on a insulating material baseboard.

called *copper-clad board* and it comprises a thin baseboard (about 1.5 mm) of insulating material such as resin-bonded paper or, more likely, fibreglass, with an even thinner layer of copper (about 0.2 mm) on one surface (or both, if it's double-sided copper-clad board). Figure 4.3 shows an enlarged cross-section of a typical piece of copper-clad board.

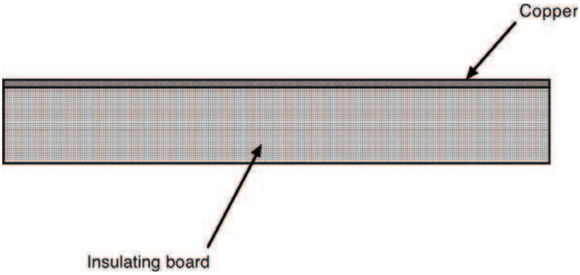

FIGURE 4.3 Enlarged cross-section of copper-clad printed circuit board

▷ FYI ◁

Apart from basic copper-clad printed circuit boards like this, industrial manufacturing often makes uses of double-sided copper-clad printed circuit boards (in which a circuit is made up from copper connections on both sides of the board) and multi-layered printed circuit boards (which comprise several internal layers of copper connections as well as those on one or both surface sides of the board) — typically these would be used in only the most complex of circuits, where space is at a premium!

This copper-clad board must now be turned into a usable printed circuit board, so some way must be found to remove excess areas of copper, leaving behind only those areas which we need to connect the components to. The areas left behind form what's called the copper track or the copper foil. Figure 4.4 shows such a printed circuit board where the copper track or foil is clearly visible.

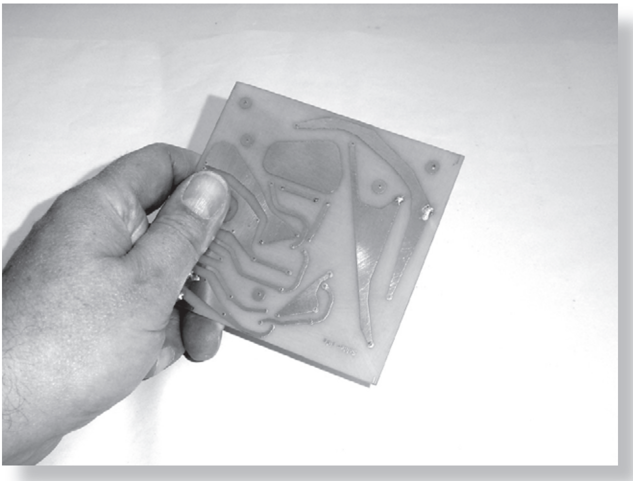

FIGURE 4.4 **Once the excess copper is removed from copper-clad board, the copper track or foil remains.**

Once a printed circuit board has been created, complete with a copper track, it is then a relatively simple task to drill the holes in the printed circuit board for the component leads to fit through, and then solder them in.

But we shouldn't go too far, too fast. We should really stop and think for a minute about what's just been said. Somehow a piece of copper-clad board (which consists of a thin layer of copper stuck onto the surface of a thin insulating board) must have areas of copper removed, leaving behind a copper foil which follows the foil pattern required on the printed circuit board for the project. Just exactly how do we remove the copper?

Don't worry, the answer's not quite as difficult as you might imagine — we simply dissolve it away using an acid-type etchant.

But doesn't the etchant dissolve away the copper foil, too? Well, no — because first we have to coat the areas of copper that are to make up the copper foil with something which will resist the etchant, preventing the copper underneath from being dissolved. Figure 4.5 shows the idea.

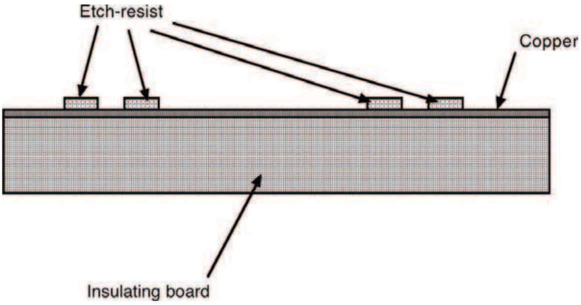

FIGURE 4.5 **Coating the copper surface with an etch-resistant material means that the copper under the material is protected from attack from etchant.**

After etching (Figure 4.6), the unwanted copper has been dissolved, but the copper foil under the etch-resist remains.

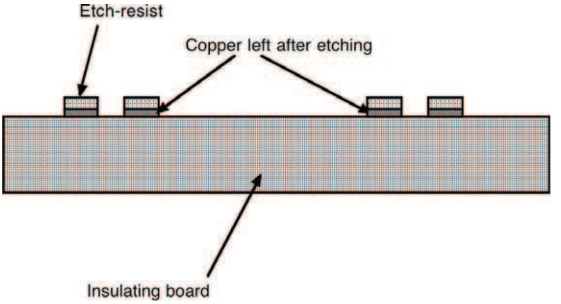

FIGURE 4.6 **Once the board of Figure 4.5 has been etched, only the copper underneath the etch-resistant material remains.**

After removal of the etch-resist, the printed circuit board is ready for drilling (Figure 4.7), component insertion (Figure 4.8), soldering (Figure 4.9), and finally component lead trimming (Figure 4.10).

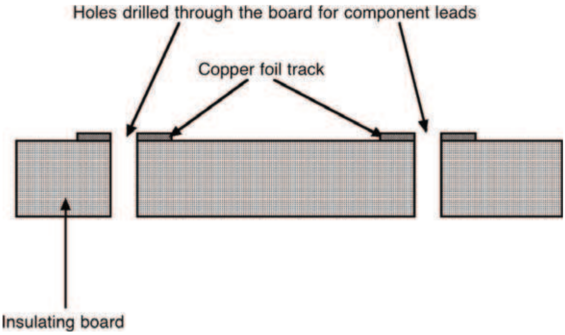
Holes drilled through the board for component leads
Copper foil track
Insulating board

FIGURE 4.7 Holes are drilled through the board at relevant points, after the etch-resistant material has been removed.

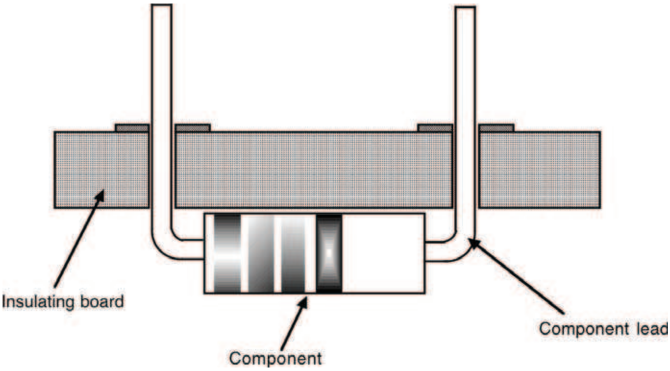
Insulating board
Component
Component lead

FIGURE 4.8 Component leads are inserted through the drilled holes.

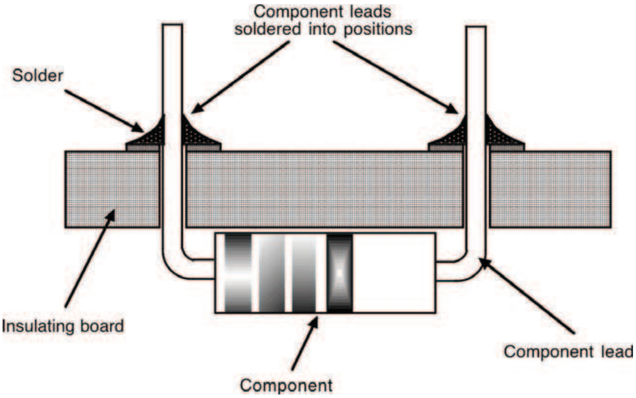
Component leads soldered into positions
Solder
Insulating board
Component
Component lead

FIGURE 4.9 Component leads are soldered into position.

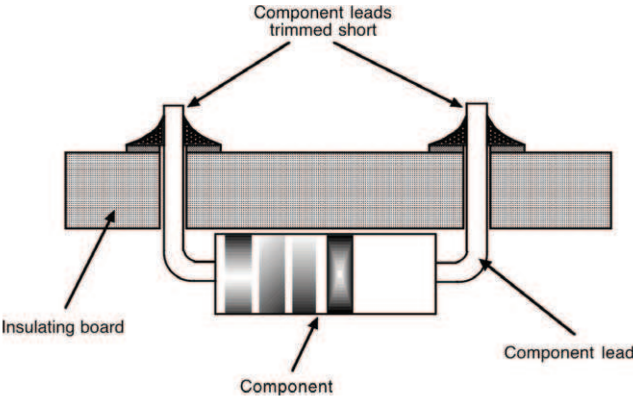
Component leads trimmed short
Insulating board
Component lead
Component

FIGURE 4.10 *Component leads are trimmed, leaving a neat soldered printed circuit board.*

In summary, the steps to making a printed circuit board are:

- **OBTAIN** or design the foil pattern for the circuit
- **APPLY** etch-resist to copper clad board in the shape of the foil pattern
- **ETCH** away the unwanted copper
- **REMOVE** the etch-resist and drill the component lead holes
- **INSERT** the component leads
- **SOLDER** the component leads to the board
- **TRIM** the excess component leads.

Later in this chapter these steps are described in full, and the various methods that you can use to create your own printed circuit boards are detailed.

Other boards

There's no doubt that printed circuit boards give a much better way to build circuits, and it's the way recommended wholeheartedly by this book and by anyone else who knows anything about basic electronics and who seriously wants to help people like you — the reader — to use the best methods.

However, printed circuit boards aren't the *only* method of building electronic circuits. So, we're going to take a short break from printed circuit boards, to look at a couple of other methods. That way, you'll see the differences, and see why printed circuit boards are best.

Matrix board

Made from a plain board of insulating material, with holes at 0.1 inch horizontal and vertical distances, matrix board is merely a way of holding components on a board. A plain piece of matrix board is shown in Figure 4.11.

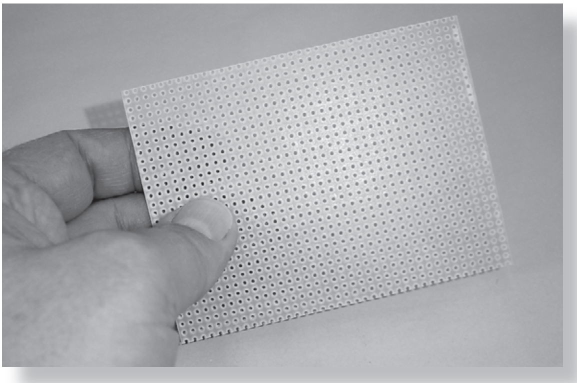

FIGURE 4.11 A plain piece of matrix board.

As with printed circuit boards, component leads are inserted through the holes on matrix board, but there the similarity ends. There is no copper on the underside of a matrix board as there is on a printed circuit board, so all connections between components must be made by either bending component leads to the next component or by adding extra wire to do the same. Where component leads or wires join, the joint must be soldered.

Overall, matrix board is useful for prototyping a simple circuit to make sure it works. It can also be useful where the circuit is being designed as it's built — it's quite easy to add extra components if a change is required, and it's also quite easy to adapt the circuit as you go. Indeed, when using matrix board *the whole layout* is usually designed as you go! However, it's not the ideal construction method for a finished electronic circuit as it is hard to replicate, and is tricky to use if the circuit is complex.

Stripboard

Stripboard is a sort of cross between matrix board and printed circuit board, so has advantages from both methods of construction. For a start, the underside (shown in Figure 4.12) has strips of copper to make connections between components. Also, the board is pre-drilled, with holes at 0.1 inch horizontal and vertical distances, just as with matrix board. As a result, it's quite easy to make a simple electronic circuit by inserting components, and soldering them in place. Also, as the copper strips are pre-defined, and unlike the earlier matrix board construction method, it's perfectly possible to design a layout on paper first, so that circuits can be replicated fairly easily. For this reason, stripboard circuit layouts are sometimes given in electronics magazines and books, particularly for simple circuits. They are not, however, ideal.

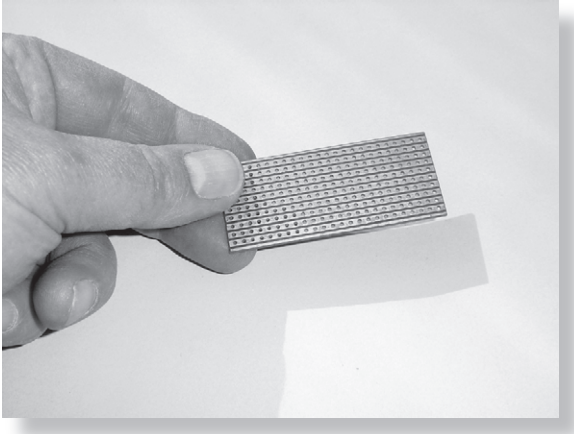

FIGURE 4.12 A plain piece of stripboard, showing the strips of copper. Pre-drilled holes in the strips mean that component insertion is straightforward, if defined.

Hint

When using stripboard it's often the case that only a couple of holes on a copper strip are used to connect components – the rest of the copper strip is then unusable. Also, where components like dual-in-line integrated circuits are to be used, the copper strips of plain stripboard would directly connect the integrated circuit's two rows of pins.

The answer is to cut the copper strip, effectively splitting it into parts. A sharp knife such as a craft knife could be used to split the copper strip, but a purpose-made tool – the spot-face cutter – is best used. To use the spot-face cutter just position it onto a hole in a copper strip, then twist it backwards and forwards two or three times, until the copper strip has been cut away (shown in Figure 4.13).

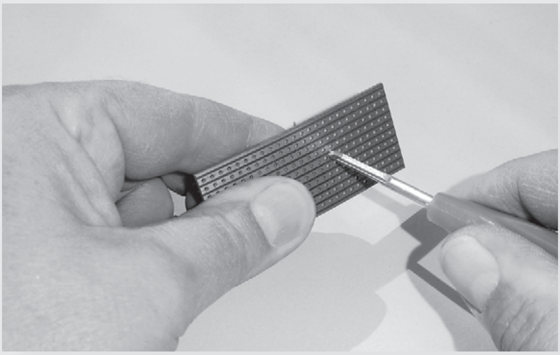

FIGURE 4.13 Using a spot-face cutter to split a copper strip on a stripboard.

There are variants of the plain stripboard, designed particularly for integrated circuits, in which the copper strips are organized to accommodate the rows of dual-in-line pins common to integrated circuits.

Nevertheless, other than being quite a useful method for the making of simple electronic circuits or for prototype circuits, stripboard is still not the ideal contructional method. Printed circuit boards remain the best method and should be the ideal you strive for.

Making your own printed circuit boards

Many people take advantage of the various printed circuit board services that electronics magazines offer, through which they can purchase ready-made printed circuit boards from the magazine publisher, complete and ready to build their projects onto. If you don't fancy making your own boards, these services are ideal.

Hint:

When electronics books and magazines give printed circuit boards for projects, they do so by printing what is called the foil patterns of the projects. It's the foil pattern of any particular project, of course, which is the shape of the copper to be left on the printed circuit board when the excess areas are removed. When you make a project given in a book or magazine you have the foil pattern to create the printed circuit board. If you want to design and build any other circuit, you must design the foil pattern yourself.

On the other hand, you may be trying to build a project taken from a book (like this one) or — simply — you may like to have a try making your own. Making your own printed circuit boards is certainly more rewarding than buying them ready-made — but it is a little more complex. However, it's nothing the average newcomer to electronics can't handle; as long as you follow the few simple rules which we'll now give you.

Right from the very beginning — the foil pattern

Let's assume that you have a circuit, which you may have designed yourself or somebody else may have designed it, but you have no printed circuit board foil pattern for it. Figure 4.14 is a typical circuit which you may recognize as being a multivibrator, and will do nicely for our purposes here.

At this point you need some basic tools — all of which you probably have at hand — paper (plain, tracing and graph), a pencil and a rubber, and a ruler. You'll also need a selection of one of each type of the components in the circuit eg, one resistor, one capacitor, one transistor, one integrated circuit etc, or their physical dimensions. This is because you need to know how big each component is so that when designing the foil pattern you know what distance to leave between component lead holes.

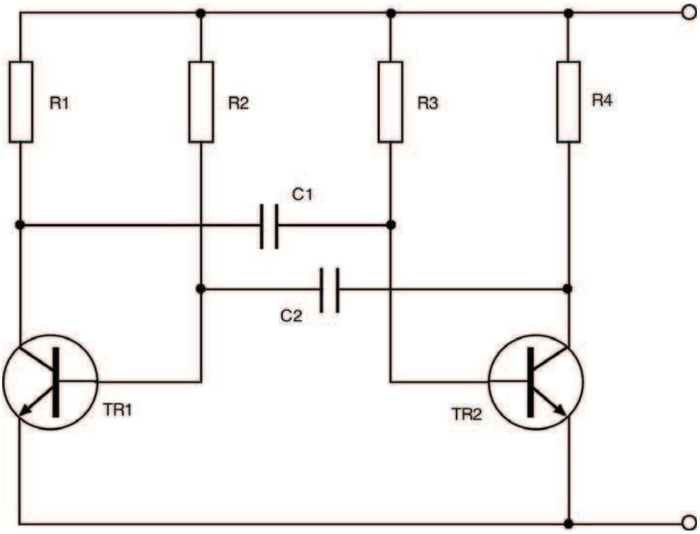

FIGURE 4.14 A simple circuit which we'll use to show how to design your own printed circuit boards.

Although this may seem daunting at first, don't worry, because if you design more than just a couple of foil patterns, you'll begin to remember the sizes of all but the most unusual components. Lay everything out on a desk or clean workbench.

Now the difficult bit (not really!): designing the foil pattern. What you have to remember here is that the components go into one side of the printed circuit board, but the copper foil making the connections is on the other side of the board.

Begin by choosing a component, fairly central to the circuit (we'll choose transistor TR1). Now draw its approximate shape (as viewed from above) onto a sheet of plain paper, as in Figure 4.15.

Just as we've done, mark where its connection leads will enter the paper 'board' and label it (TR1 in our example). Also, if it's anything other than a simple resistor or capacitor, you should mark the leads, too (*e*, *b*, and *c* if it's a transistor).

Now draw a few more components onto the paper, alongside the first, similarly marking them as you go (Figure 4.16). Don't be afraid to position them close together; a spacing of only 2 mm or so between components is normally ideal.

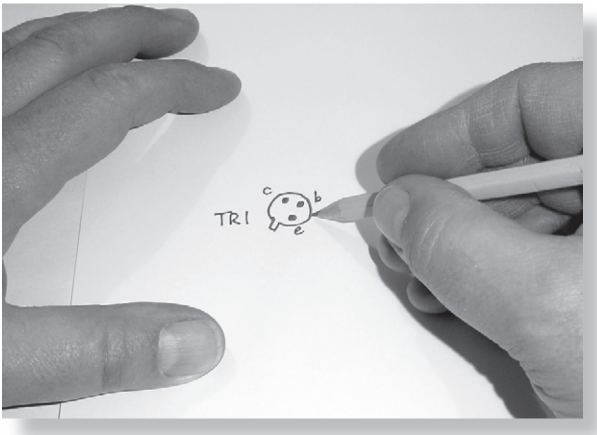

FIGURE 4.15 First steps: marking the first component onto plain paper. Also mark its leads – as it's a transistor, the leads should be named e, b and c to show the emitter, base and collector.

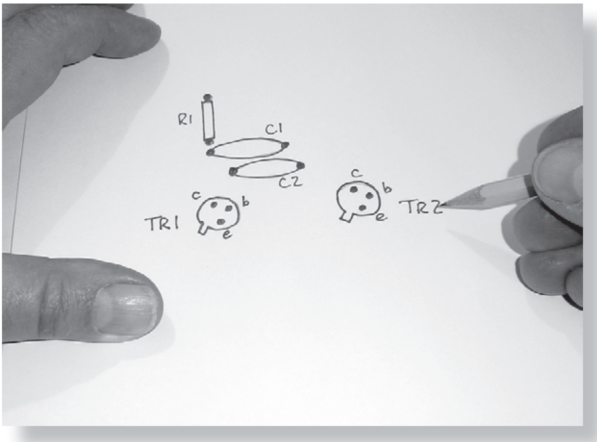

FIGURE 4.16 Add further components to the layout, by marking them on the paper – they can be quite close.

At this stage you may find it helpful to more or less follow the general pattern of the circuit diagram when positioning components, although this is by no means essential — it's probably just easier for a newcomer to do.

When you gain confidence in making printed circuit boards you probably won't follow the pattern of the circuit diagram at all — it really doesn't matter; as long as the connections follow the circuit diagram, the pattern can be like the Union Jack if that's what you want.

When you have drawn in a few components, begin to pencil in 'connections' corresponding to the copper track required, as in Figure 4.17, so that they will be electrically joined by these connections on the finished printed circuit board.

Don't worry if you get stuck, or even make a mistake, that's why you have the rubber! Simply rub out the error and start again.

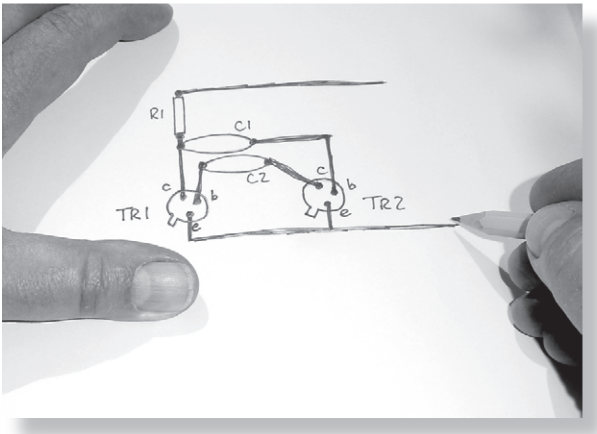

FIGURE 4.17 Connections between component leads should be drawn in now.

Build up the layout, putting in more and more components, and more and more 'connections'.

Remember that the finished printed circuit board will have its foil pattern on the other side to its components, so tracks can actually pass underneath components, as in Figure 4.18. There is also no reason at all why two, or even three, tracks can't pass underneath a component, so long as the tracks don't touch each other or touch the component lead holes.

If you were worried about your foil pattern being untidy, now's your opportunity to change all that. Place a sheet of tracing paper over a sheet of graph paper and tape the two together along the edges.

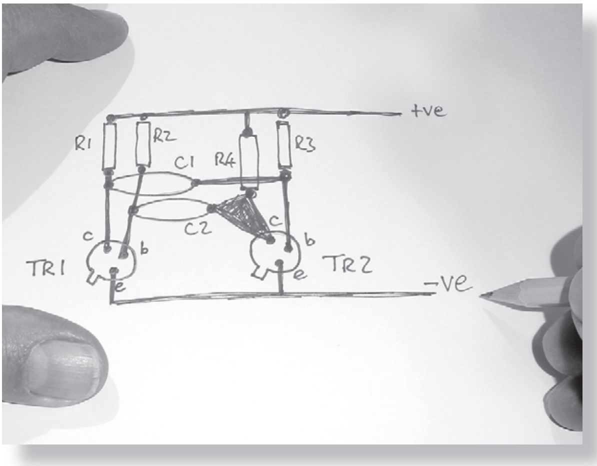

FIGURE 4.18 As the copper will be on the other side of the board to the components, there's nothing to stop you designing the printed circuit board so that copper tracks actually run beneath components.

Using the graph paper grid to align your components and connections accurately, redraw the layout on the tracing paper. This time, however, use two different colours of pencil, say, red for the component shapes and black for the connections. In this way you can instantly tell which is the 'top' and which is the 'bottom' of the board. The graph paper allows you to accurately get the sizes of components right, and position them as they will be on the final printed circuit board. Check, as you redraw the foil pattern, that it is correct with regard to the circuit diagram, and check you've remembered all components and got their polarities right.

And that's it! You've designed your own foil pattern.

You should now detach the tracing paper from the graph paper and keep it safely.

Pattern pending

This is the section which tells you how to transfer your pattern onto the copper-clad board, ready for etching.

As we saw before, the foil pattern must be put onto the board using etch-resistant materials. The form of this resist may be one of many: the simplest (and definitely the cheapest) can be ordinary household enamel or gloss paint, applied with a fine brush (messy); the most complex is photographic resist, developed on the circuit board much like your summer holiday piccies are developed on film and photographic paper (assuming you've not got a digital camera, of course). The one you use will depend on how much you're prepared to pay.

You can expect to pay about (gulp) £100 for a basic photographic system; this will let you make really ace printed circuit boards, but it won't do your pocket a lot of good — it's the sort of system which is likely to be used by schools and colleges where large numbers of one-off printed circuit boards are needed, so ask there first to see if they wouldn't mind you using theirs occasionally.

If you haven't got that sort of free cash around and can't find anyone willing to let you use their photographic system (ooh, the meanies) then you'll have to make do with a cheaper method.

Let's run through some of the methods now, so at least you know what the choice is.

Paint

The cheapest method of transferring your foil pattern onto the board is to use paint and a fine paint brush to hand-paint all the tracks and component holes. This is extremely messy, however, but will work quite well indeed, if a little care is taken.

A variant of this method is probably better though, where clear self-adhesive plastic has the image of the track pattern cut out of it (before the paper backing has been peeled off). The straight tracks can be cut out first, using a sharp craft knife and straight-edge, then the component holes can be punched out with a punching tool.

When the foil pattern is complete, you can peel off the paper backing and stick the plastic to the copper-clad board.

Now you effectively have a stencil, exactly the shape of your required foil pattern. There's no need to use a fine paint brush now; just paint over the whole plastic sheet with a much broader brush. After you've done that, carefully peel off the plastic and you're left with the foil pattern of painted resist.

Clever, eh?

Ink

Pens are available (ask at your local electronics store) which are loaded with special ink, resistant to the circuit board etchant.

With these pens it is quite an easy job (shown in Figure 4.19) to draw the foil pattern directly onto the copper-clad board. You won't make a perfectly neat printed circuit board using this method, but it does allow a quick and easy route to a semi-acceptable finished job. Final quality, however, depends totally on your artistic talent (or, at least, how steady your hand is!).

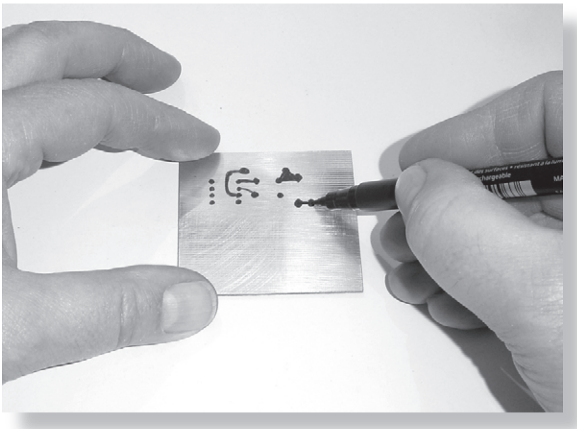

FIGURE 4.19 **Drawing copper foil pattern directly onto the copper surface of a copper-clad board using an etch-resistant ink pen.**

Hint:

In all these manual methods of applying resist it's sometimes a bit tricky to estimate where on the copper board the component leads go, before filling in the connecting etch-resistant bits. A useful trick is to put your tracing paper copy of the foil pattern upside-down onto the copper (the pencilled connections are to go on the bottom of the board, remember), and using a sharp pointed instrument such as a scriber or a compass point, prick through the tracing paper into the copper at every component hole location.

When you remove the tracing paper you will find a network of little holes, corresponding to each component lead hole.

Now you can put on the resist (whichever method you choose) knowing where the positions are.

Transfers

Probably the best method of printed circuit board manufacture for the amateur is shown in Figure 4.20, where you can see that rub-down transfers are being used as resist, for component holes. You can buy transfers for all kinds of shaped holes, too — for integrated circuits, or edge connectors, or even specially shaped ones for transistors.

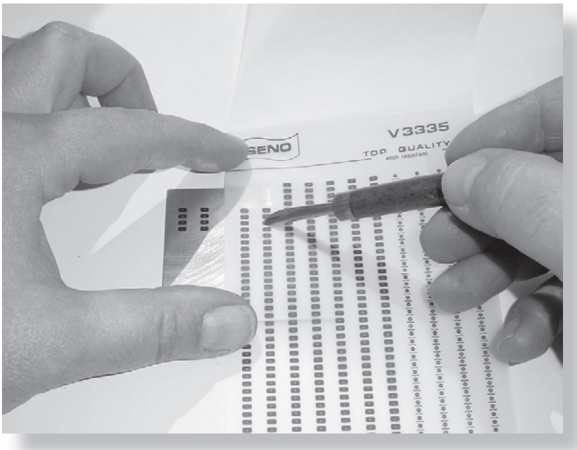

FIGURE 4.20 Rub-down transfers of all component shapes and sizes make an effective method of applying etch-resist.

When all the component locations are done, the connections can be made using straight-lined, or curved transfers, until the resist on the copper has the same shape as your foil pattern drawing.

Hint

A word of caution here: be extremely careful as you do this job, because any small cracks or breaks in the transfer resist will mean that a corresponding crack or break will occur in the copper when the board is etched.

▷ FYI ◁

You don't have to stick to just one manual method of resist application. It's perfectly possible to combine two or more of these methods to good effect. For instance, you can use rub-down transfers to apply the component lead positions (particularly useful for integrated circuit positions) but use a resist pen, or paintbrush and paint, to make the other connections.

Hint

If you make any mistakes at this stage they can be removed using the corner of a small piece of sticky tape – but once you have etched the board your mistakes are permanent.

One minor variant also in this method is to substitute reels of special, thin, self-adhesive tape for these straight-lined or curved rub-down transfers, as the tape can be stuck to the copper surface in either straight lines or formed quite easily around curves as you stick it down. For obvious reasons, this is often referred to as the 'dots-and-tape' method of making printed circuit boards.

Photographic method

Using the photographic method, the highest quality printed circuit boards can be made.

What's more, as it *is* a photographic method, you can make more than one printed circuit board from it (just as you can get reprints of your holiday snaps).

Firstly you need a full-sized master copy of the foil pattern, on film. There are two methods here: where the master copy is a negative (as from a camera film); or where it is a positive. Both methods are used in printed circuit board making, but the positive method is the most popular, for no other reason than because it's easier to make a positive master than a negative one, so we'll stick with that one.

Hint:

You can make a positive master from your foil pattern by laying a sheet of acrylic film (which you can get from drawing-office suppliers) over the tracing paper or magazine foil pattern.

Now, using the rub-down transfer or dots and tape method described earlier, make up the foil pattern directly onto the acrylic film. When complete, this film is your master copy – so keep it safe!

FYI

If your foil pattern is taken from a magazine or book, and you have a computer with a scanner and printer, you can make your master copy a little more accurately. Simply scan in the magazine foil pattern, then print it onto some acetate film. You'll need to check that the output size is the same as the source foil pattern and adjust as necessary. Note that some printers aren't capable of printing onto acetate film, however. Later there is more information about using computers to aid electronics construction.

Hint

If you are creating your own printed circuit board foil patterns for your own circuits, and you have a computer, then it might be worth looking at one or more of the many available programs for printed circuit layout. While these aren't always the easiest programs to use, they can give very good results, and if you intend making several printed circuit boards over the years, they are often worth the learning curve you have to climb. See later for details.

Resist-coated copper-clad board

The copper-clad circuit board you need now is not just ordinary board, but has a thin layer of *photographic resist* (sometimes shortened to *photo-resist*) over the copper, as illustrated in Figure 4.21, which — just like camera film — needs to be exposed to light. You can't just expose it to any old light, as you do with your camera film, though.

The only type of light which will affect the photo-resist is ultra-violet (UV) light.

Special light boxes (one is shown in Figure 4.22) are obtainable (it's these light boxes which cause this method to be expensive — they cost upwards of £80) in which the photo-resist coated board and the master foil pattern are placed, to expose the board, as illustrated in Figure 4.23.

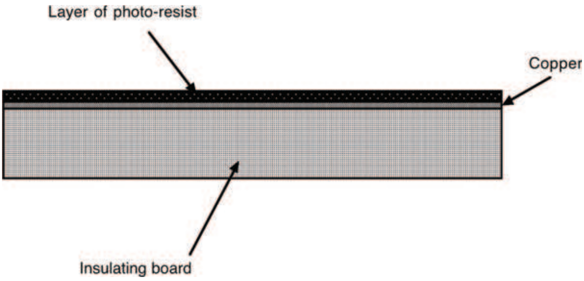

FIGURE 4.21 Photo-resist coated copper-clad board.

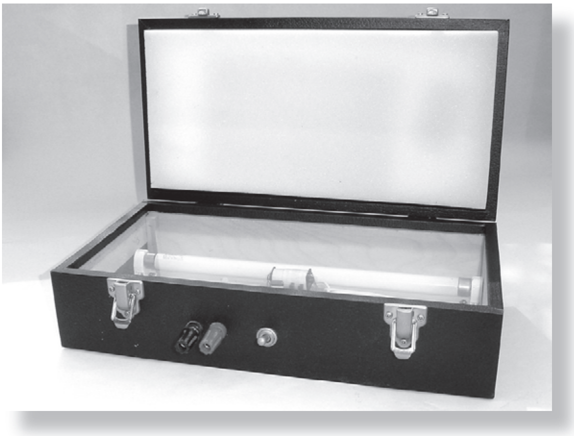

FIGURE 4.22 An ultra-violet light box.

FYI

UV light is harmful to the human eye, in quantity, so these light boxes typically have a lightproof seal, although the power of radiation emitted by the internal UV tubes is not high anyway.

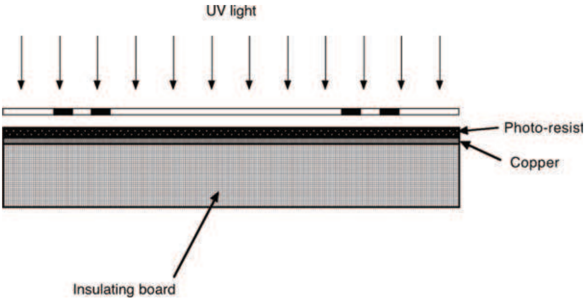

FIGURE 4.23 Exposing the photo-resist coated copper-clad board.

You must make sure that your foil pattern is the right way up here, because if it isn't your printed circuit board's track will be 'upside-down' or mirror-image, and the printed circuit board will be unusable!

After exposing the board through the master foil pattern, for the correct time (which depends on the strength of the UV tubes and the thickness of the layer of photo-resist, but is usually around five minutes or so), the board must now be developed.

The actual developing chemical depends on the photo-resist but is often a water solution of sodium hydroxide, and the exposed board is simply washed in a bath of it, as shown in Figure 4.24. As shown, the developing solution can easily be contained in a photographic-type developing tray, as it's not particularly hazardous.

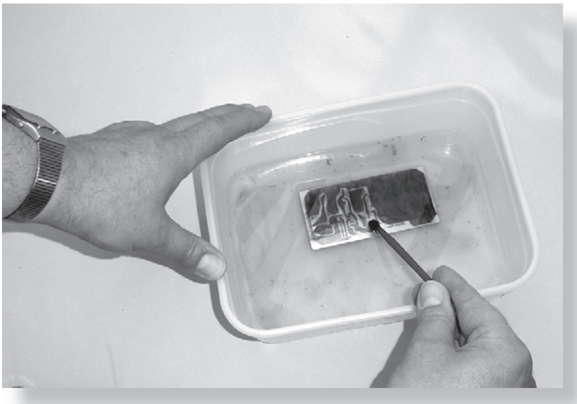

FIGURE 4.24 Developing an exposed, photo-resist coated, copper-clad board.

During development, the exposed areas of the board ie, the parts where no component holes or track connections were, soften and wash away. Meanwhile, the non-exposed areas corresponding to the foil pattern itself harden and remain in place, leaving the copper-clad board, complete with printed foil pattern.

Hint:

At this point in making a printed circuit board, it's always useful to check the foil pattern (which should be totally covered in etch-resist) to make sure that no breaks in coverage occur. If so, use a printed circuit board pen to patch broken areas.

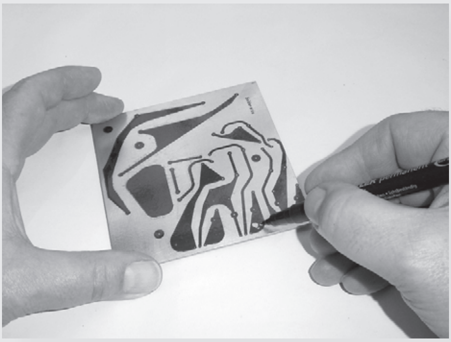

FIGURE 4.25 Touching up etch-resist with a printed circuit board pen.

If you wish to make more than one printed circuit board, you can use the acrylic film master foil pattern over and over again to expose other pieces of photo-resist coated board.

Using computers

The design of a printed circuit board copper foil track layout, and the later production of the printed circuit board from that copper foil are areas of electronics where accuracy is important. They are, indeed, almost sciences, rather than arts, although it can often help to have a steady hand.

As such, copper foil track layout is one particular area (there are others, some of which we'll see in this and other chapters) in electronics where computers can be put to good use, and there are several programs that can do this for us. Many of these programs form part of expensive industrial packages that take control of electronic device manufacture from initial marketing brief to late production stages. These packages are, needless to say, very expensive, and far more complex than we need to consider in this book.

On the other hand, there are several programs that are considerably cheaper than this, and even some that cost nothing at all to individuals.

Figure 4.26 shows a program called Crocodile Technology, which is an electronics simulator that allows us to build theoretical electronic circuits up on computer screen, testing them for function as required.

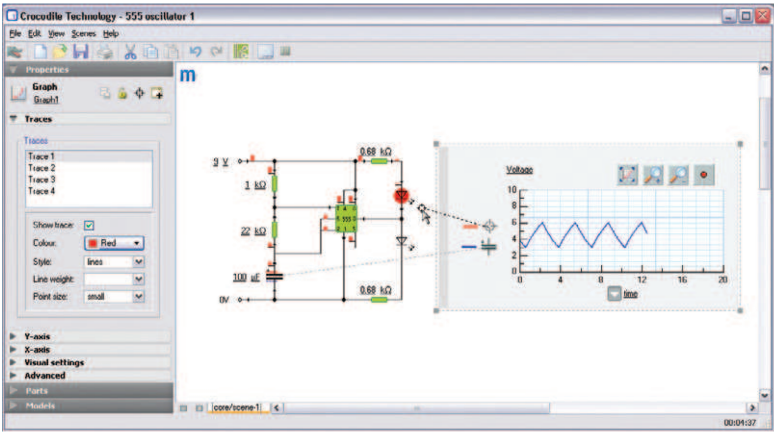
Crocodile Technology - 555 oscillator 1
File
Edit
View
Scenes
Help
Properties
Graph
Graph1
Traces
Traces
Trace 1
Trace 2
Trace 3
Trace 4
Y-axis
X-axis
Visual settings
Advanced
Parts
Models
m
9 V
1 kΩ
22 kΩ
100 µF
0V
0.68 kΩ
0.68 kΩ
555
Voltage
10
8
6
4
2
0
0
4
8
12
16
20
time
core/scene-1
00:04:37

Show trace: ☑
Colour: Red
Style: lines
Line weight:
Point size: small

FIGURE 4.26 Crocodile Technology – an electronics simulator (available for Windows and Mac OS) that allows us to develop and construct circuits, on computer screen (Crocodile Clips).

By itself, Crocodile Technology doesn't aid us in producing copper foil track layouts for a printed circuit board, until it is used with a sister computer program called realPCB, which does just that — producing printed circuit board copper foil track layouts from the simulated circuits within Crocodile Technology. Figure 4.27 shows a portion of a printed circuit board layout produced using realPCB.

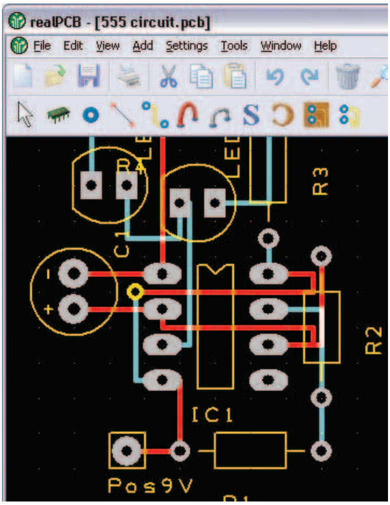

Figure 4.27 Part of a printed circuit board layout created using realPCB (Crocodile Clips).

While Crocodile Technology is available for Mac OS computer users, realPCB isn't available in Mac OS format. However, there are many alternatives for Mac users, one of which — Osmond — is a free program ideally suited for printed circuit board copper foil track layout.

Figure 4.28 shows a printed circuit board produced using Osmond — actually, the Bench Power Supply project's printed circuit board, in Chapter 8.

Programs such as realPCB and Osmond feature libraries of components, so that creating a printed circuit board layout is really just a matter of defining which components you want to use, then effectively joining the dots.

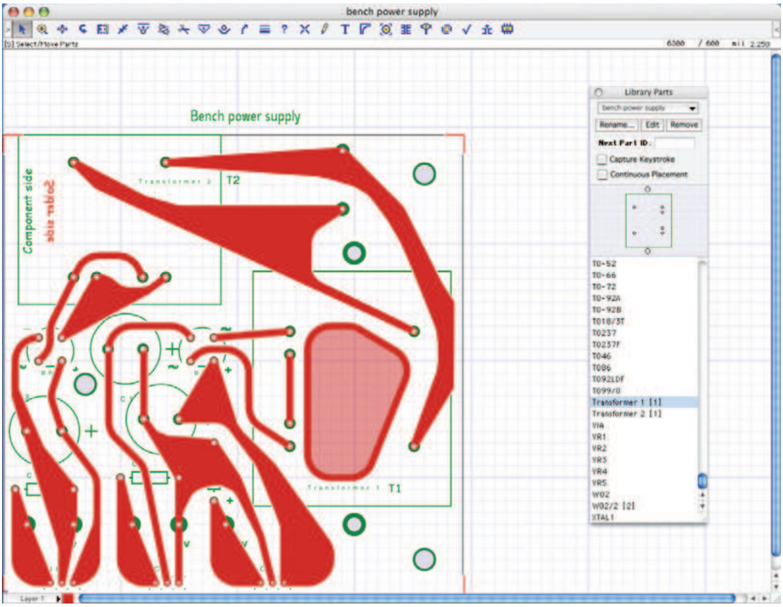
bench power supply
[S] Select/Move Parts
6300
/ 600
Bench power supply
Component side
Solder side
Transformer 2
T2
Transformer 1
T1
Library Parts
bench power supply
Rename...
Edit
Remove
Next Part ID:
☐ Capture Keystroke
☐ Continuous Placement
TO-52
TO-66
TO-72
TO-92A
TO-92B
TO18/3T
TO237
TO237F
TO46
TO86
TO92LDF
TO99/0
Transformer 1 [1]
Transformer 2 [1]
VIA
VR1
VR2
VR3
VR4
VR5
W02
W02/2 [2]
XTAL1
Layer 1

FIGURE 4.28 The Bench Power Supply copper foil track layout (see Chapter 8 for project details), produced using Osmond.

From screen to board

Producing a copper foil track layout on computer is one thing — but using that layout to make the actual printed circuit board is another! However, it is another area where computers can be helpful. As the computer can output to a printer, and as any printed circuit board layout program worth its salt can print its design, there are several ways that the computer can help us transfer the layout design onto the copper of copper-clad board, ready for etching.

Photographically

If you intend to use the photographic method detailed earlier, then it's a relatively simple process of printing the computerized copper foil track layout to film, ready for combination of the film with photographic resist-coated printed circuit board and an ultra-violet light box.

Purpose-produced translucent acetate sheet films are available for both laser and ink-jet printers (LaserStar for laser printers, JetStar for ink-jet printers). Figure 4.29 shows such a film that has been printed directly from a computer to a laser printer.

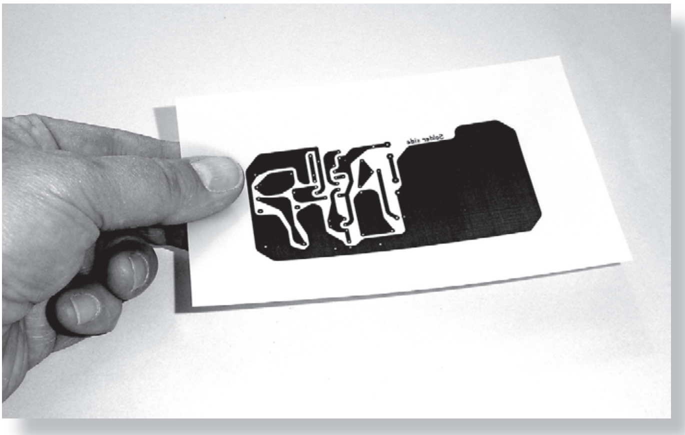

Figure 4.29 A printed circuit board design printed using a laser printer, on LaserStar translucent acetate sheet. JetStar sheets are used with ink-jet printers.

These translucent sheets are transparent to ultra-violet light (so are ideal for the purpose of ultra-violet light box exposure), but allow solid and dense areas representing the copper foil. Figure 4.30 shows the printed circuit board design of Figure 4.29 being used with an ultra-violet light box and photographic resist-coated printed circuit board.

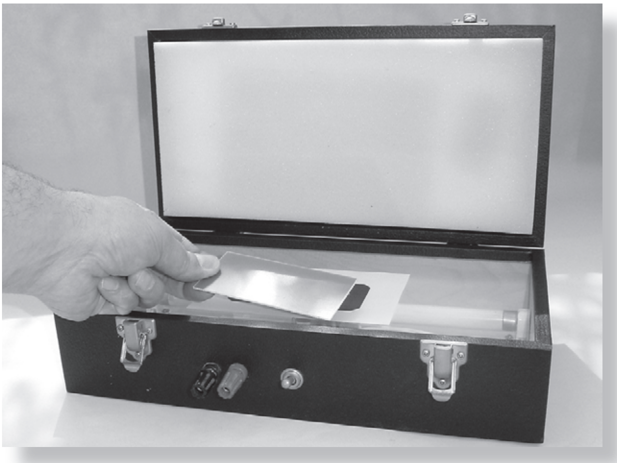

Figure 4.30 Placing the LaserStar film and photographic resist-coated printed circuit board into an ultra-violet light box, ready for exposure.

Printed circuit boards of the highest quality can be made using this method, direct from a computer.

Non-photographically

If you don't have access to the ultra-violet light box photographic method of making printed circuit boards, there is an excellent alternative. Like the photographic method, it comes in the form of a specially-formulated acetate film that can be used with a printer (there is, though, only one type, that must be used only with a laser printer — it will not work with an ink-jet printer). The method is called Press-n-Peel, for reasons which will become clear.

The system is very easy to use, albeit a little tricky, and may take a few practice runs to get right.

First, print the printed circuit board copper foil track layout design, from whichever computer program you use, onto a Press-n-Peel sheet in a laser printer. Next, invert the sheet and place it onto the copper side of a piece of ordinary (ie, non photographic resist-coated) copper-clad printed circuit board, as shown in Figure 4.31.

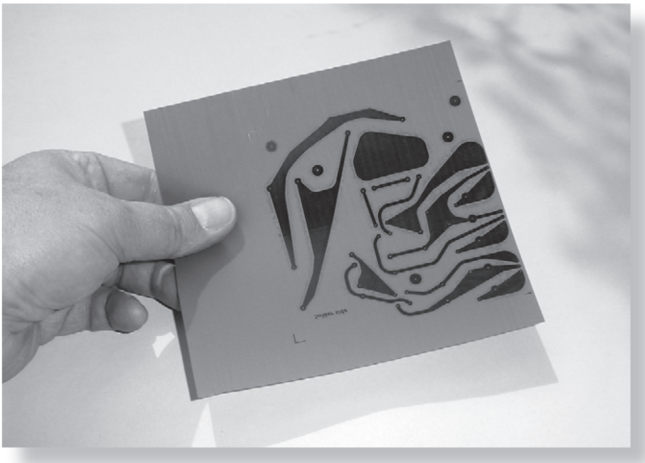

FIGURE 4.31 Lay the Press-n-Peel sheet over the plain copper surface of the copper-clad board.

Next, lay a piece of plain paper over the top. Using a hot iron, press onto the paper to transfer the copper foil track layout image from the Press-n-Peel sheet onto the printed circuit board copper surface, as shown in Figure 4.32.

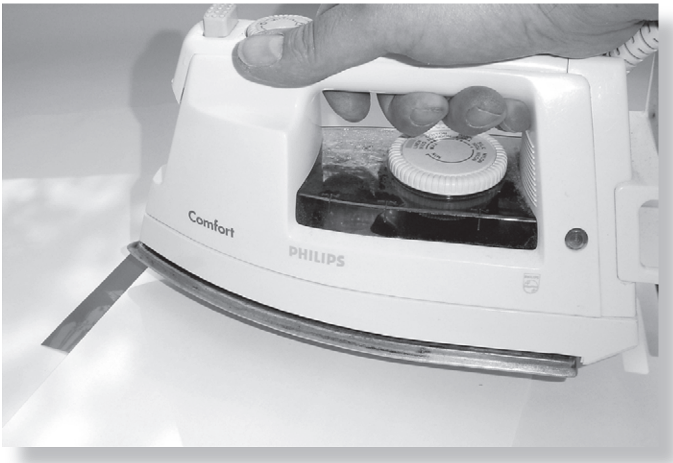

FIGURE 4.32 Using an iron to transfer the image from a printed Press-n-Peel sheet onto the copper surface of plain copper-clad printed circuit board.

The iron must be hot (275–325°F) and you must keep it moving at all times to prevent the sheet melting.

After a short while, remove the iron and paper, then cool the printed circuit board and Press-n-Peel sheet under running, cold water. When the board is cold, peel off the Press-n-Peel sheet (Figure 4.33).

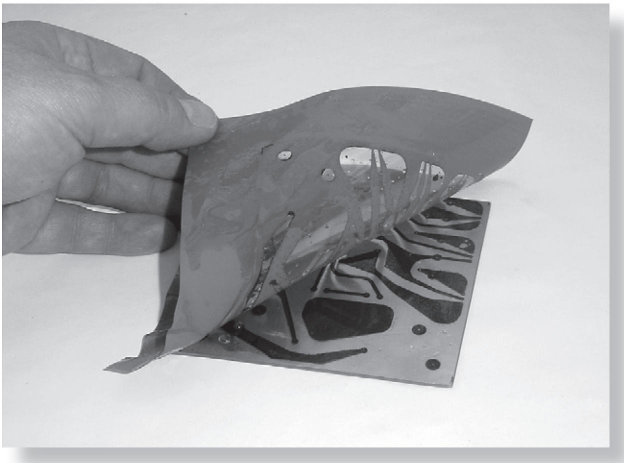

FIGURE 4.33 Peeling off the Press-n-Peel sheet, following ironing and cooling.

Come up and see my etchings...

Finally, it only remains (did that man say *only?*) to etch the board and it's then ready for drilling and component insertion.

The etchant used to etch printed circuit boards is not actually an acid as you might expect, but ferric chloride — a horribly messy fluid — but it does its job well. Most local electronic stores will sell it bagged as crystals or granules and you simply mix them with water to give an etchant solution. A note of caution before you touch this stuff however: it is a very dangerous fluid and a strong solution of it will merrily eat its way through your clothes, the carpet, or even the stainless steel kitchen sink. So, please be careful and don't say I didn't warn you.

The etching process takes place more rapidly when the ferric chloride solution is warm and constantly moving, and for this reason you can buy specially built containers called *etching baths* to hold the solution. These etching baths are fitted with heaters to maintain a high temperature, and with agitators to keep the solution moving (vital to stop the build-up of copper oxide on the surface of the copper, which prevents further etching). Some top-of-the-range etching baths use a spray technique to maintain a high level of agitation — an etching machine of this type (which can actually be used with developer, too) is shown in Figure 4.34.

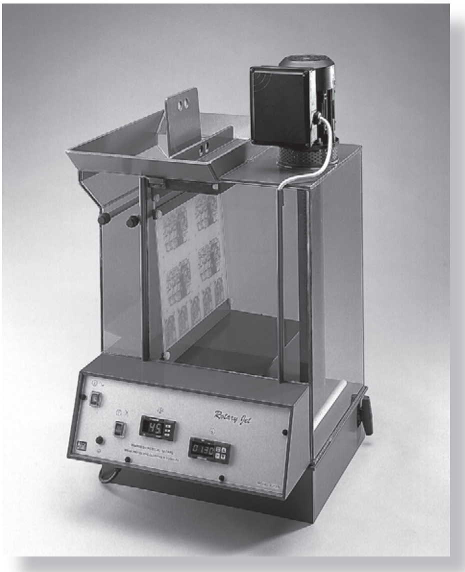

FYI

While historically ferric chloride has always been the etchant of choice in printed circuit board making, there are alternatives. While being somewhat less corrosive and nasty to use, these alternatives do tend to be a lot more expensive than ferric chloride however.

FIGURE 4.34 **The Rota-Spray etching/developing machine (Mega Electronics UK).**

After etching a number of printed circuit boards, the etchant will become exhausted and require renewal. You will notice this as a gradual increase in the time it takes to etch. When new, etchant may take only five or six minutes to etch a printed circuit board (particularly if the etchant is heated), whereas at the end of its working life etchant may take twenty minutes or more to do the job. The number of printed circuit boards a batch of etchant will successfully etch depends on the size of the printed circuit board and the area of copper foil to be etched, but you should realistically expect to be able to etch twenty or more printed circuit boards per batch.

Hint

Etching baths complete with heaters and agitators are (you guessed it!) expensive, but there's nothing to stop an enterprising reader making one, from nothing more complicated than a plastic container, aquarium heater, and a toy motor with plastic propellor. Just make sure that nothing metal is inserted into the etchant – for obvious reasons!

At the other end of the scale, a plastic photographic-type developing tray or similar can be adequately used as a makeshift etching bath, to hold the ferric chloride during etching, as shown in Figure 4.35. But as you'll have to keep rocking the bath gently, to agitate the etchant over the printed circuit board, this method is fraught with hazards — it's all too easy to splash some of the etchant.

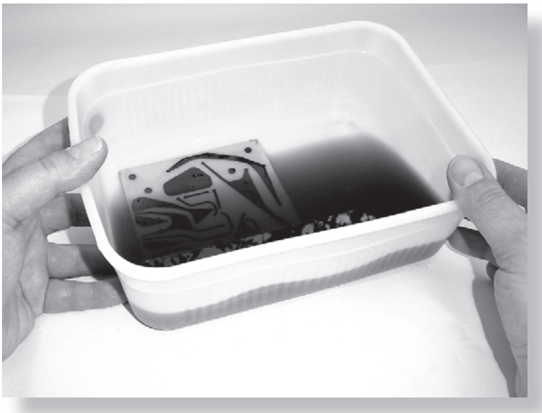

FIGURE 4.35 A makeshift etchant bath – an old ice-cream container, in which the etchant is manually agitated by rocking the tray gently. A good excuse to have some ice-cream!

However you choose to do this job, it's wise to wear old clothes or overalls, and do the job outside, or at least in an uncarpeted workshop. Also, wear rubber gloves because it can do some pretty ghastly things to fingers — it'll certainly stop anyone's nail-biting problem, as it'll leave no nails at all to bite!

After etching, wash the board thoroughly in water to get rid of all traces of the etchant. Also, make sure that any etching solution has been either washed away or properly returned to its lidded container, and is not left where anyone (or even your cat, for that matter) can accidentally spill it.

Drilling for oil

Now is the point where all the component holes in the printed circuit board should be drilled so that components can be inserted and soldered. You can do this with a hand drill, or even a standard electric drill, but neither of these are particularly accurate and so may be the cause of irregular or misshapen holes.

A far better type of drill is the hand-held electric rotary tool commonly found in DIY and hardware stores. Being of quite high speed and easy to use they can make an excellent job.

Follow standard drilling practice when drilling a printed circuit board — put a block of wood under the printed circuit board as shown in Figure 4.36, so that you drill through the printed circuit board into the wood. This helps prevent the drill tip breaking roughly through the printed circuit board, and also helps prevent any damage to the worksurface.

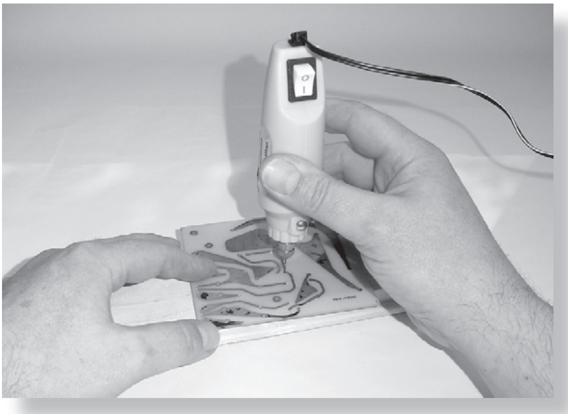

FYI

Make sure the drill bit (of whatever size you use) is perfectly sharp. A blunt drill increases drilling time, overworks your hand-held drill, and creates holes with burrs.

Figure 4.36 Hand-drilling a printed circuit board – rest the printed circuit board to be drilled on a block of wood. Hold the drill upright.

Irresistible

At this point, the etched and drilled board still has a coating of resist over the copper tracks of the foil pattern. Depending on your method of making the printed circuit board, you need to remove this resist prior to inserting the components and soldering. Resist removal can be done chemically using a solvent as shown in Figure 4.37.

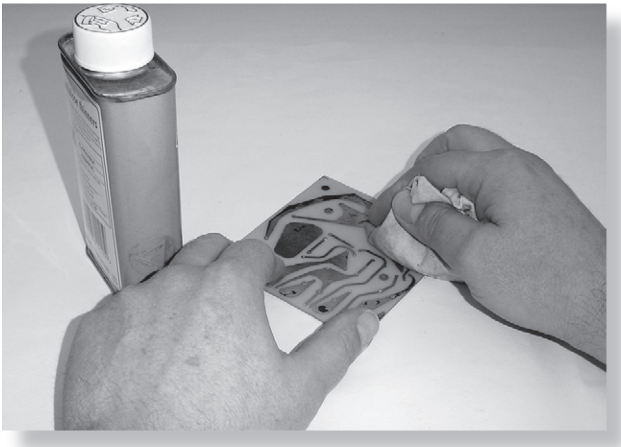

FIGURE 4.37 Removing etch-resist with a chemical.

Hint:

Some types of photo-resist coated boards actually allow soldering direct to the resist – in fact, the resist acts as a solder flux. If the resist-coated board you buy has such a resist, there is no requirement – other than personal preference – to remove it.

Such solvents are not cheap, so you can choose to remove the resist mechanically using an abrasive removal block (Figure 4.38) which you rub over the copper track to take off the resist. An alternative to a removal block is the fibreglass propelling pencil-type of abrasive cleaner, shown in Figure 4.39.

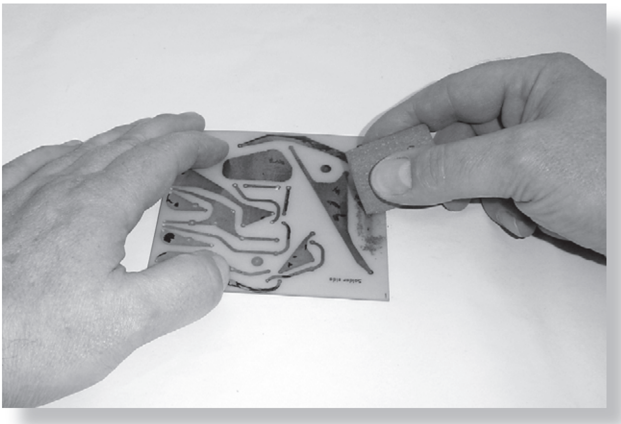

FIGURE 4.38 Removing resist using an abrasive block.

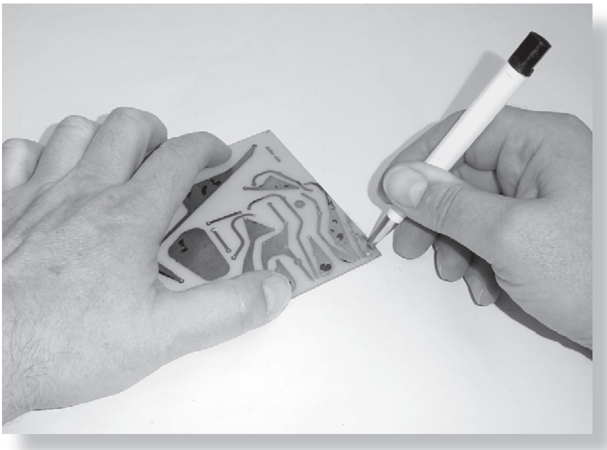

FIGURE 4.39 An abrasive cleaner in the form of a fibreglass propelling pencil-type of holder can also be used to remove resist.

Component insertion

As a rule of thumb, components should be inserted and soldered according to whether they can be damaged by heat. So, passive components like resistors and capacitors should be inserted and soldered first, leaving active components like integrated circuits until later.

Insert and solder only a few components at a time — no more than three or four at once for example. This means that components are less likely to slip out of their positions before solder is applied.

Once soldered into positions, trim all leads close to the soldered joint as shown in Figure 4.40. This prevents short circuits occurring between component leads.

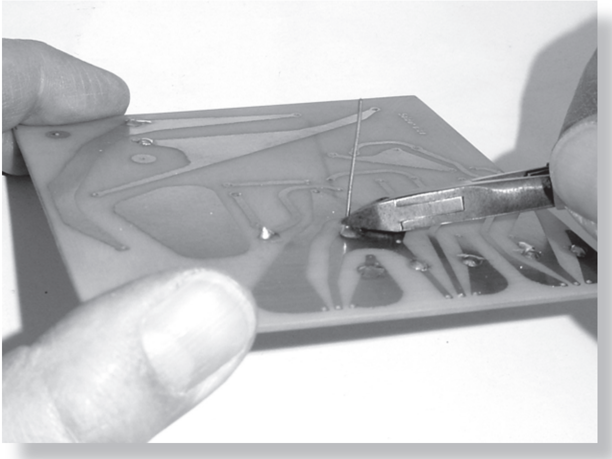

FYI

The process of inserting components into a printed circuit board is often — particularly in industrial electronics manufacturing — called (for fairly obvious reasons) stuffing.

FIGURE 4.40 Trim all component leads down after soldering. Use your side-cutters for this task.

Soldering

Soldering of components into a printed circuit board is actually quite an easy job, but there *is* a bit of a technique to it, and you need to be aware of the problems that incorrect soldering can create. Chapter 5 covers soldering in depth.

5

Soldering

The single most important process in electronics is soldering. It is — quite literally — the glue that holds all other processes together.

Soldering is the main means of:

- **HOLDING components in their positions on the printed circuit board**
- **CONNECTING components to other components in the circuit.**

What is solder?

Solder is a metallic compound which possesses a relatively low melting point. It has been used to join metals for thousands of years, since Roman times at least. Historically, one of the main constituents of solder has been lead. Indeed, the Latin name of lead — *plumbum* — gives a clue to one of solder's main uses, in plumbing, to join water pipes and fitments.

In electronics, solder's main use is to join component leads to printed circuit boards, in what is called a *soldered joint* — occurring between the component lead and the copper foil track of the printed circuit board. The copper foil track in turn allows that soldered joint to connect to other soldered joints (also formed by other component leads and the copper track), so that the complete electronic circuit can be made up on the printed circuit board. As a by-product, the rigidity of the soldered joints making up the circuit means that all components are also held securely in place.

The components of solder

Except in its very earliest uses, when pure lead was used, solder has always been a mixture of two or more substances. The Romans are known to have used a tin/lead alloy to joint lead water pipes. This is mainly because a tin/lead alloy has a lower melting point than lead itself. For example, pure lead has a melting point of 327°C, whereas an alloy of 62% tin and 38% lead has a melting point of 183°C — so it's possible to joint lead pipes, without actually melting the pipes themselves.

As soldering has evolved, and the uses to which it is put have changed, different solder alloys have been created for different purposes. In fact, soldering is still evolving — and is actually in the middle of quite a large evolutionary step — as we shall see here.

The 62% tin/38% lead solder alloy mentioned already, is known as the *eutectic composition*, and the melting point of 183°C (the *lowest* melting point of any tin/lead solder alloy) is correspondingly known as the *eutectic point*. It's not hard to understand that — purely in terms of soldering component leads to printed circuit board copper foil tracks in electronics — this has inevitably been viewed historically as a very useful, if not the ideal solder alloy.

However, times are changing. For a start, the world's resource of tin is depleting rapidly — only around 50 years' worth at current usage rates is left! More important though, is that lead is not exactly viewed as a safe substance, being a toxic, poisonous metal. For this toxicity reason, the world's electronics industry — along with other industries that currently use lead — is undergoing a revolution at present, to completely and permanently eliminate the use of lead in any of its processes and components.

Several new solder alloys have been developed to replace the old tin/lead alloys, in which various other elements — silver, bismuth, zinc, copper — have all been used with tin to form these new *lead-free* solder alloys.

While being inherently safer to both us *and* the environment, these lead-free solders all have a major disadvantage practically over the old tin/lead alloys in that the new alloys have higher melting points (around 30–40° higher!). This means that their very use may harm the electronic components being soldered, and they can be more difficult to use in general. Nevertheless, the electronics industry aims to be lead-free by 2006.

Effects of lead-free solder on us

The use of lead-free solder is mirrored in the non-industrial areas of electronics that readers of this book will be most interested in — in the home, school, college or university. Already lead-free solders are available off-the-shelf for personal use, and we can expect that tin/lead solders will be unavailable very much sooner rather than a little later.

The good soldered joint

Whatever the actual solder alloy used, the principles of making a soldered joint are the same. Solder is applied to a heated area of metal, whereupon the solder alloy melts and *wets* the metal to form a joint. Note that the term *wet* is used, and in some respects means much the same as a liquid such as water wetting a surface — it flows over the surface.

Wetting in soldering terms, on the other hand — unlike water wetting a surface — is a process where the solder comes into direct metallic contact with the metals to be soldered together into the joint. At the junction between the solder and the metal, a specific compound of both the solder and the metal is formed — known as an *intermetallic compound* — and occurs in a very thin layer between the two. Sometimes, the bond at that point is called the *intermetallic bond*. Interestingly, once this intermetallic compound between solder and metal has been created, it cannot be removed — you cannot remove solder from metal once it has been wetted!

Figure 5.1 shows a cross-section of a soldered joint, illustrating the main details. Figure 5.2 shows a micro-section of an actual soldered joint, in which can be seen the solder, the metal — copper — and the intermetallic compound layer between the two.

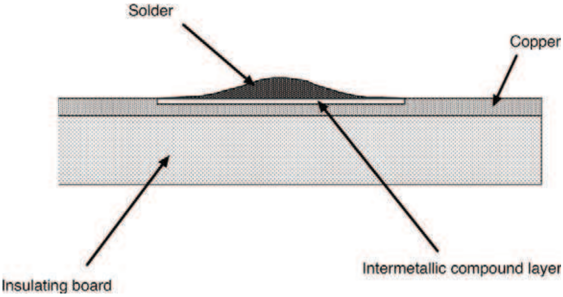

FYI

Note that we are talking about soldering joints with copper. Don't even think about trying to solder to aluminium — it just won't work — if you need to make an electrical connection to aluminium, do it with a nut and bolt!

FIGURE 5.1 Wetting takes place when solder comes into intimate contact with the surface metal, to form an intermetallic compound at the junction between solder and metal.

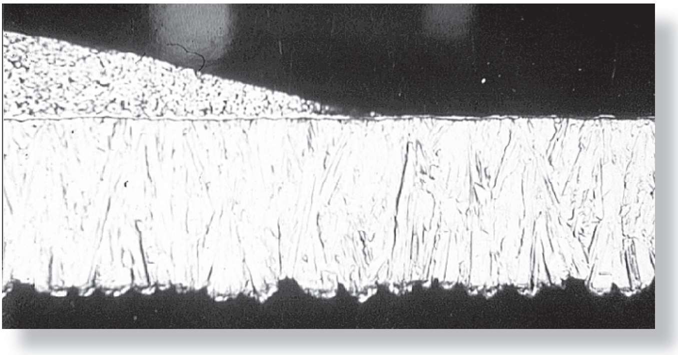

FIGURE 5.2 Micro-section through an actual soldered joint, where the intermetallic compound layer between solder on top and copper beneath is clearly visible.

The bad soldered joint

The condition for a good soldered joint to take place — illustrated previously in Figure 5.1 and shown photographically in Figure 5.2 — is that solder comes into contact with metal to wet it. Anything that *prevents* solder from coming into contact with metal, therefore, can work to prevent a good soldered joint from taking place.

The converse of the good soldered joint is one where the solder does not wet the joint, so does not come into contact with the metal. Such a bad soldered joint is shown in Figure 5.3.

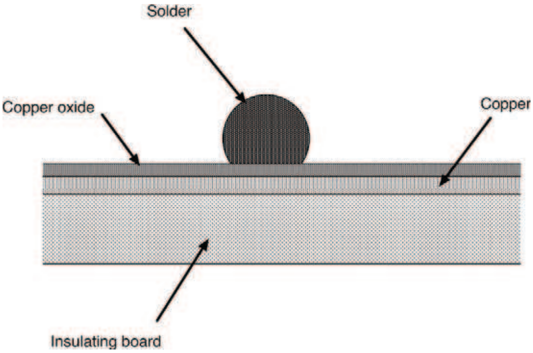

▷ FYI ◁

Many conditions can occur that create a bad soldered joint, and there are many types of bad soldered joints, so it is not always easy to specify exactly what has gone wrong in the soldering process. However, the usual problem is that the metal surface is not clean. Any number of contaminants may prevent a good soldered joint from taking place — dirt, grease, metal oxide, being the main ones.

FIGURE 5.3 A bad soldered joint – where a layer of metal oxide on the surface of the metal has prevented wetting of the solder to the metal taking place.

In practice, there are several levels of wetting which do not make a good joint, and these are worth considering as they give a clue as to the problems that we can experience in soldering electronic circuits.

All the various levels of wetting are described underneath, and Figure 5.4 shows the main levels by way of visual comparison.

- **NON-WETTING.** Pretty obvious to the eye — solder does not flow on to the metal at all. It will not adhere to the printed circuit board's copper foil track at all. A circuit built with no-wet joints will probably not work at all
- **PARTIAL WETTING.** While not flowing and completely wetting the metal surface, a partially wet joint will appear to be solid. However, it is unreliable, and may easily break off with time or vibration. Interestingly, the circuit itself may work initially when partially wet joints are present, but may not work for more than a short time
- **DE-WETTING.** Such joints may appear to have been created properly, but in the cooling phase after heat has been removed, the solder retracts — often leaving small solder balls around the joint. As in partial wetting, the joint may initially cause a circuit to work, but will break down fairly quickly
- **TOTAL WETTING.** Gives good joints, of course, which will create circuits that work and continue to do so. Note the difference between totally wet joints, and all the other wetting variants. A totally wet joint is concave in appearance (and will actually be quite shiny to the naked eye). All other wetting variants of joints will be convex (if not rounded) and may be dull in appearance to the naked eye.

Hint:

As you can see here, one of the main causes (if not the main cause) of faults in electronic circuits, is bad soldering. Therefore, it makes sense to be able to solder well. Practise it before you attempt to build an electronic circuit till you are able to make good joints every time. Don't bother using printed circuit boards for this, instead use a few bits of stripboard (see page 81 for details) – even just a couple of boards give you many holes and copper strips to practise with. By the time you've soldered a couple of handfuls of components and a few short lengths of lead to the stripboards, you'll be a past master at the art of soldering!

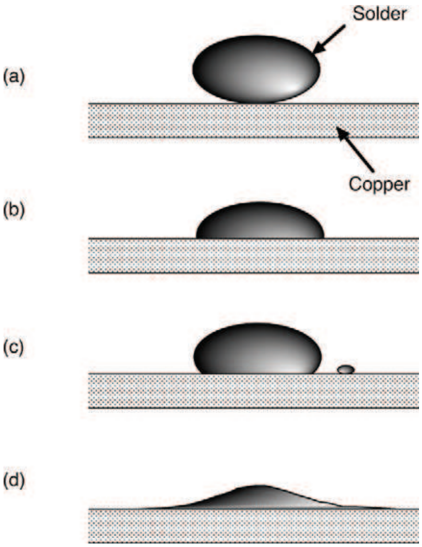

FIGURE 5.4 Four extremes of soldered joint: (a) non-wetting; (b) partial wetting; (c) de-wetting; (d) total wetting.

Flux

As we've seen, it's essential that the metal to be soldered is clean. The problem is that copper, the metal used to make copper-clad printed circuit boards that we use in electronics, oxidizes. A minutely thin surface layer of copper oxide forms on top of the copper-clad layer of the printed circuit board that makes up the copper foil track pattern very rapidly. So no matter how we clean the copper surface, the oxide layer builds up before we can get a soldering iron to it. And copper oxide cannot be soldered!

Further, every time a printed circuit board is handled it picks up dirt and grease, both of which work to prevent solder from wetting the board's copper track.

The key to making sure that the copper is returned to pristine condition in order that solder can wet it is a substance called *flux*.

Fluxes are chemically active — that is, they work on the surface contaminants (copper oxide, dirt, grease) on the copper track of the printed circuit board to dissolve them. This occurs when the copper is heated by a soldering iron, in the first stage of the soldering process. Once the copper surface is clean, solder can be applied.

▷ FYI ◁

There are many types of flux, judged usually by the level of chemical activity they have. Generally, the fluxes used in the solder used for hand soldering are of low activity. They are typically made by distilling the sap (or resin) from pine trees. The residue is then dissolved in a solvent, making it easier to apply.

At first sight, it might seem that co-ordinating all this (heating the copper, applying the flux, waiting till the copper is clean, applying the solder) is a tricky matter, best left to the professionals and their industrial soldering processes. However, the *real* trick is that it can all be done by hand in a single smooth process by anyone, as long as a few simple steps are followed.

Key to the whole process is that modern solder, used to solder printed circuit boards by hand, already contains the flux needed to clean each joint to be soldered.

The solder used has cores (usually four or six) throughout its length which are filled with flux — as illustrated in Figure 5.5. Diameter of the whole flux-cored solder lead is usually around 1 to 2 mm.

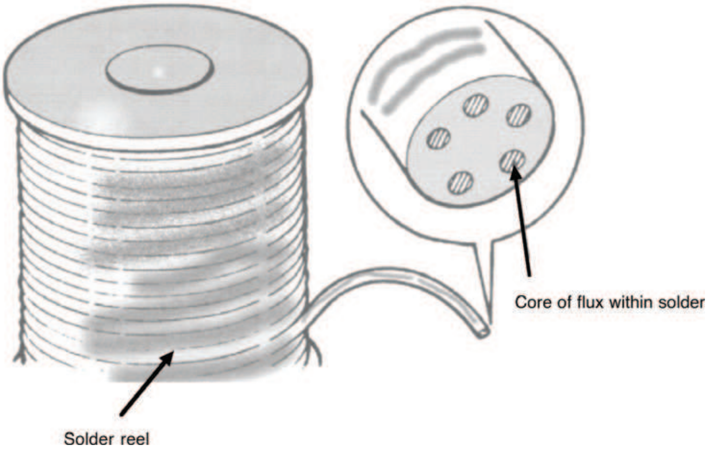
Core of flux within solder
Solder reel

FIGURE 5.5 **Cored solder, obtained in reels, appearing much like thin flexible wire. Flux is present in cores within the solder.**

So how does it work? After all, having the flux inside the solder like this means that the flux is presented to the joint to be soldered at the same time as the solder is, rather than before it — which would suggest it has no time to clean the joint. However, the flux has a much lower melting point than the solder has, so the flux melts and flows onto the joint first. It therefore has the time to coat the metal surfaces of the joint and clean them, before the solder melts and flows over the joint to make the soldered joint, as shown in Figure 5.6.

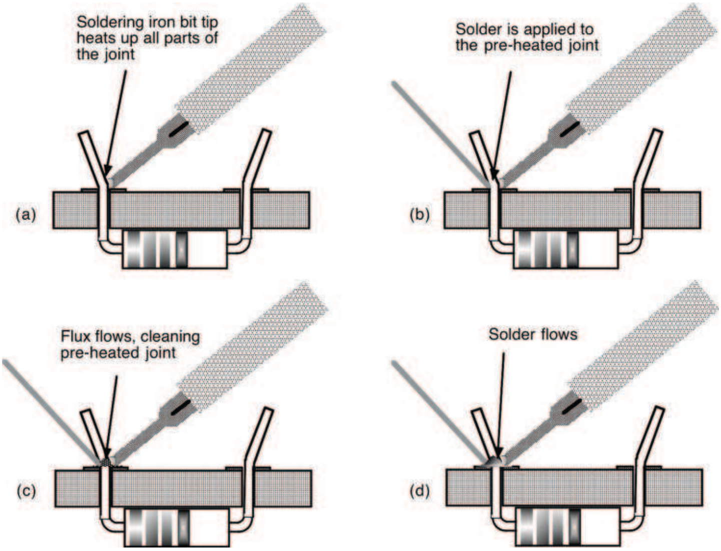
Soldering iron bit tip heats up all parts of the joint
Solder is applied to the pre-heated joint
(a)
(b)
Flux flows, cleaning pre-heated joint
Solder flows
(c)
(d)

FIGURE 5.6 The main stages of hand soldering a joint: (a) the soldering iron is applied to the joint, to heat it; (b) while maintaining the soldering iron's position, cored solder is applied to the joint; (c) flux melts quickly and flows from the cored solder onto the joint's metal surfaces, to clean them and protect them; (d) the solder melts, and flows over the joint's metal surfaces.

FYI

This whole process, from first applying the soldering iron to the solder flowing over the joint surfaces, takes only a few seconds (depending mainly on the size of the joint and how powerful the soldering iron is). If a joint takes longer than just a few seconds to make, there's a good chance the soldering iron isn't powerful enough.

Take Note

Throughout this chapter we look at processes involved in soldering various joints on printed circuit boards (ie, those circuit boards made from copper-clad board, with copper foil tracks). The processes – however – are actually identical whatever sort of circuit boards you use.

Key points when soldering

Being able to use a soldering iron isn't the only thing you need to know when soldering. The whole process depends on several aspects:

When soldering

- **CLEAN** all parts — copper printed circuit board track, component leads, soldering iron bit — *before* soldering
- **MAKE** a reliable mechanical joint — *before* soldering
- **HEAT** the joint sufficiently — *before* applying solder
- **APPLY** the solder — keeping the soldering iron on the joint while the solder melts
- **REMOVE** the soldering iron and allow the solder to solidify — *before* handling or moving the joint
- **TRIM** the excess component lead from the joint, once the joint has solidified and cooled sufficiently to allow handling.

We should now consider these six aspects in detail.

Cleaning all parts

We already know that flux is used automatically when soldering using cored solder to help clean the joint surfaces, as part of the soldering process. However, flux can only remove so much oxide, grease or dirt. It's best to improve your chances of making a good solder joint by manually cleaning the joint surfaces first. Use an abrasive fibreglass pencil (as shown in Figure 5.7), or an abrasive scrubbing block (as shown in Figure 5.8), rubbing them over both copper track and component leads to minimize dirt and grease.

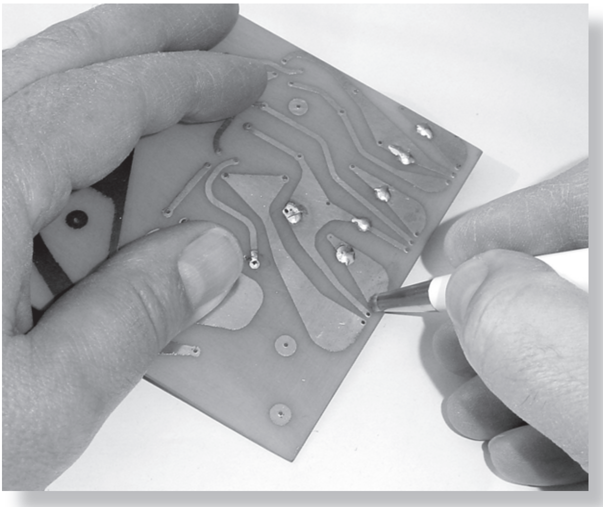

Figure 5.7 Cleaning copper track of a printed circuit board, using an abrasive fibreglass pencil – simply rub the pencil over the track, particularly over the component land areas to clean it. Brush aside the fibreglass residues before commencing soldering.

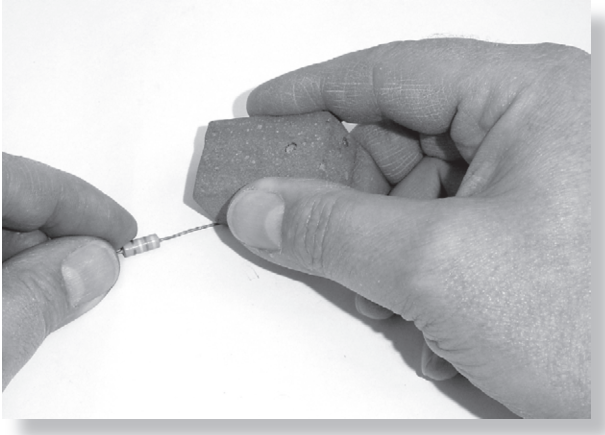

Figure 5.8 Using an abrasive scrubbing block to make sure a component's leads are clean, prior to soldering.

It's vitally important that the soldering iron bit tip be kept clean too, and the best way to do this is to *tin* it. Tinning is a process whereby fresh solder is regularly melted on to the heated bit tip, so that the tip remains coated with fresh solder. As flux-cored solder should be used, just as when soldering the actual joint, the flux within the cores cleans the tip simultaneously.

The process is a two-stage one: first wipe old excess solder off the soldering iron bit tip, then tin the tip by applying fresh solder. Figure 5.9 shows a wet sponge being used to wipe the soldering iron bit tip on. An alternative is shown in Figure 5.10, where a tip cleaning and tinning block is being used.

Hint

If you have bought yourself a new soldering iron and are preparing to tin it, you will find that when you first plug the soldering in to turn it on, smoke will be created and given off from the bit tip: this is because the bit is coated with a grease substance to prevent it oxidizing. This is useful, but also means that as the bit heats and burns off the grease, oxide very quickly forms, making the tip unusable. To prevent this, tin the bit tip as shown in Figure 5.11 immediately as it heats.

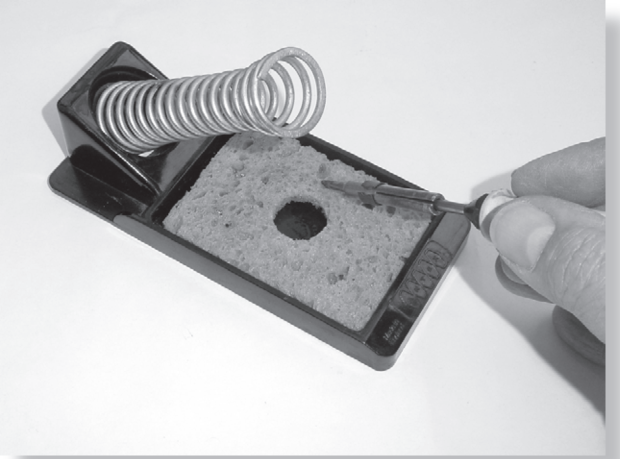

FIGURE 5.9 *Cleaning a soldering iron bit tip using a wet sponge. Simply wipe the tip over the sponge surface.*

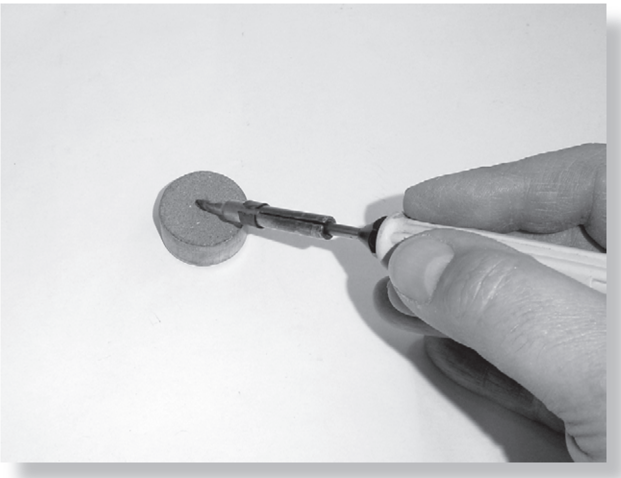

Figure 5.10 Using a tip cleaning and tinning block to clean a soldering iron bit tip – just wipe the tip over the block.

Immediately following removal of old excess solder, apply fresh solder to the soldering iron bit tip, as shown in Figure 5.11. Just a small amount of solder is all that's necessary to keep the tip in err... tip-top condition.

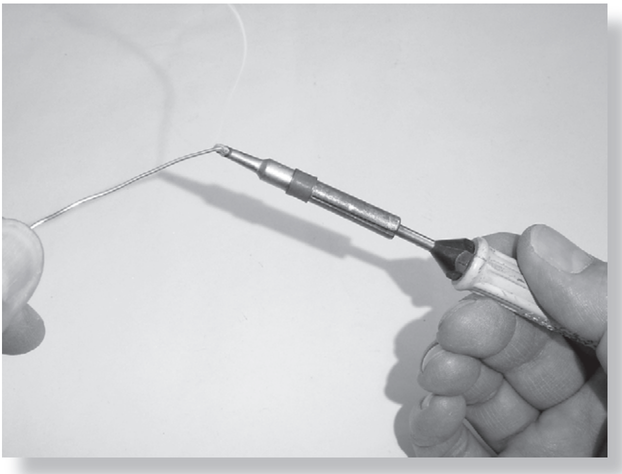

Figure 5.11 Tinning a soldering iron bit tip. Once the soldering iron bit tip has been cleaned as shown in either Figure 5.9 or Figure 5.10, quickly tin it using a small amount of ordinary flux-cored solder. The tip will have a shiny appearance if tinned properly.

Hint

Get into the habit of tinning the soldering iron bit tip every few minutes while it is on, to maintain it in peak condition and ready to solder joints.

Making a reliable joint

Components must not move while you are soldering a joint, otherwise the joint may be faulty. As a general rule, it's best to fix the components in place mechanically somehow before soldering. With axial- or radial-leaded components such as resistors and capacitors, the easiest way to do this is to bend the leads slightly after insertion into the printed circuit board, as shown in Figure 5.12. As a guide, the angle of the bend doesn't need to be (and *shouldn't* be) any more than just a few degrees — bend the leads only as far as necessary to hold the component in place.

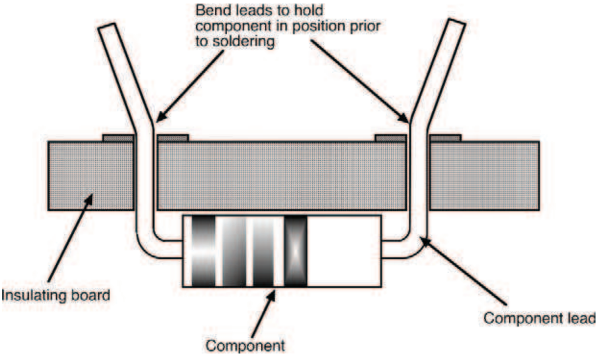

FIGURE 5.12 After inserting a component into the printed circuit board, bend its leads slightly to hold it in place ready for soldering.

Take Note

If you insert more than one component into a printed circuit board at a time before soldering, and bend their leads outwards to prevent them falling out, be careful that when you solder their joints solder doesn't bridge between them, as shown in Figure 5.13. When angled leads like this are soldered, the joint contains much more solder than a lead which is perpendicular to the printed circuit board.

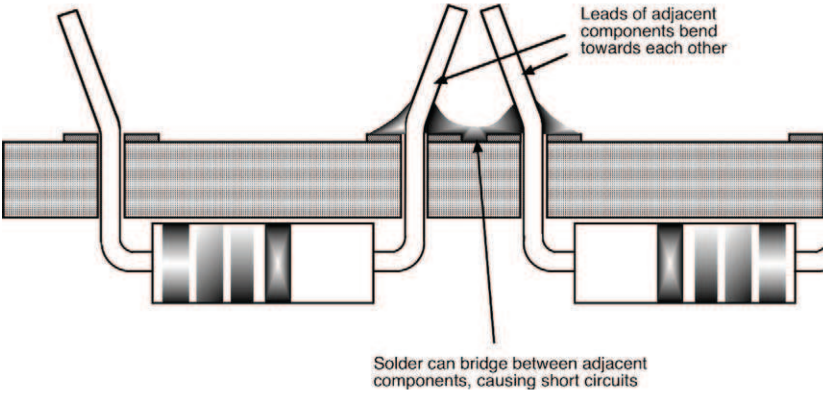
Leads of adjacent components bend towards each other
Solder can bridge between adjacent components, causing short circuits

FIGURE 5.13 Showing how solder can easily bridge between joints if angled component leads are used to hold components in place prior to soldering – take care when using this method to prevent such short circuits.

Heat the joint

For successful soldering, the metals to form the soldered joint should be pre-heated. Pre-heating is easily accomplished — apply the soldering iron bit tip to the joint — touching both the copper track and the component lead, as shown in Figure 5.14. The time you should pre-heat the joint depends on the size of metal to be heated and the power of the soldering iron, but generally this will be for no more than a few seconds — between about two and eight seconds, say. You will learn to judge the pre-heat time required by joints with practice, but in any case it's not normally *that* crucial.

The factors you should bear in mind are that:

- **IF you don't pre-heat the joint sufficiently, the molten solder cools too quickly and you run the risk of making a defective joint**
- **IF you pre-heat the joint too much, you may damage the component and may even cause the copper track to lift from the printed circuit board.**

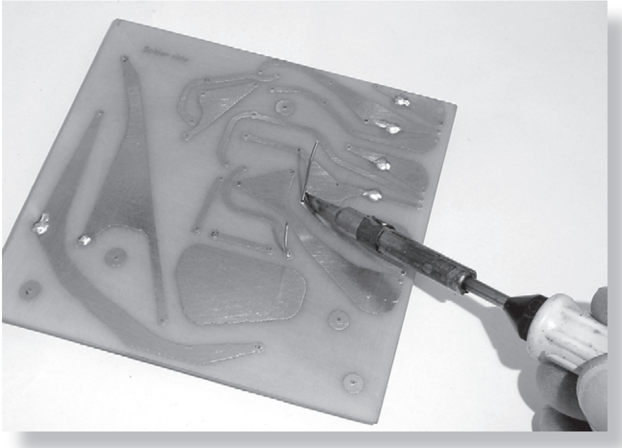

FIGURE 5.14 Pre-heating the joint to be soldered – apply the soldering iron bit tip to the joint's metal surfaces.

Apply the solder

With the soldering iron bit tip in position, apply the end of the flux-cored solder to the joint — touching it to both the joint and the soldering iron bit tip. If everything is as expected, the solder will melt and flow around the joint. Figure 5.15 shows this stage.

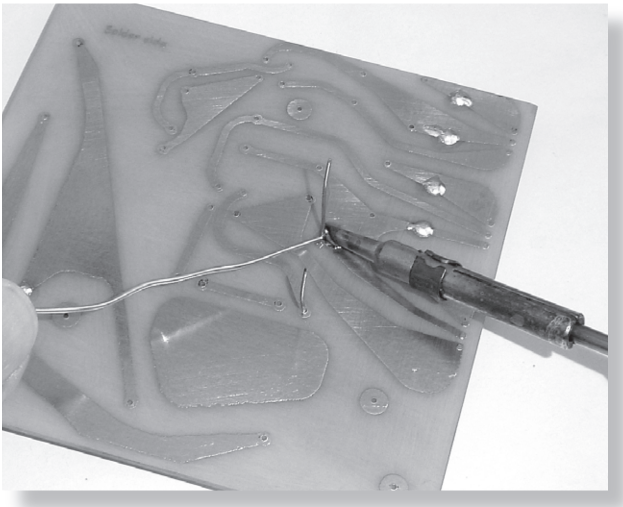

FIGURE 5.15 Solder melting and flowing around the pre-heated joint.

Take Note

Some beginners to soldering melt large amounts of solder on the soldering iron bit tip, then dab the soldering iron bit tip onto the joint they want to make – the soldering equivalent of putting glue onto a spatula then wiping the spatula onto something to spread the glue over it.

DO NOT UNDER ANY CIRCUMSTANCES DO THIS.

While the process might work well for glue, it certainly doesn't work for solder. Soldering is only effective when the joint is pre-heated, and fresh flux is applied at the same time as solder. If you put solder on the soldering iron bit tip, the flux burns away before applying it to the parts of the joint, and the joint is too cool anyway – causing only bad soldered joints.

The moral is: only use the method described earlier to solder joints!

Remove the soldering iron

Once the solder has flowed around the joint (not before), quickly remove the soldering iron. Do not move the printed circuit board, and do not move the component or its leads, until the solder has solidified. If you move anything before the solder solidifies, the joint may be damaged.

▷ FYI ◁

A good soldered joint has a concave, shiny appearance. If the joint is not heated sufficiently, or if you move the joint and damage it, you will probably notice that it is not concave, or that it is a dull, grey colour. Don't worry — just repair the fault by heating the joint again, until the solder flows properly.

Trim excess component leads

If the component leads were left as they are, and more than just a couple of components were inserted into the printed circuit board, you run the imminent risk of component leads touching — effectively forming short circuits. To prevent this you have to trim all excess component leads off immediately above the joint. Figure 5.16 shows this.

After all joints on the printed circuit board have been soldered and excess leads trimmed, it's a good time to check all the joints visually to see if they look 'good' or 'bad', repairing any that are suspect. Also check over all the copper foil track, to see that no solder bridges have formed between tracks or component joints. If you find any solder bridges, use desoldering tools — a solder sucker, or desolder braid — to remove excess solder, and resolder any joints that need it.

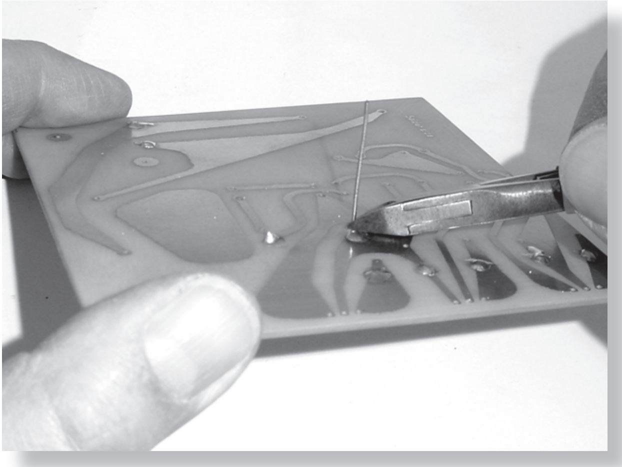

FIGURE 5.16 Trimming excess component leads off, close to the soldered joint, to prevent short circuits.

Connecting leads

Inevitably, any printed circuit board holding its electronic circuit is not isolated from the world around it! It must be connected — to switches, controls, off-board components, input and output sockets, and so on. So, it's common — indeed, normal — to need to solder connecting leads to the printed circuit board copper foil track.

To solder connecting leads to a printed circuit board, use the following procedure:

- **STRIP around 5 mm of insulation from the end of the lead**
- **IF the connecting lead is multi-stranded — twist the loose ends of the stripped lead.** If you solder the lead untwisted to the board, there's a possibility a loose strand may bridge across to another track
- **CLEAN and tin the soldering iron bit tip.** You should ALWAYS do this before any soldering operation

CONTINUED…

...CONTINUED

- **TIN the connecting lead.** This has two purposes: first, it removes any oxides, dirt and grease on the wires making up the leads, and; second, the solder you add in the tinning stage will be used when the joint is made — meaning you won't have to add much solder at that time. The soldering iron bit tip needs to be quite a broad one — around 3 mm — for this job. To tin the lead:

  - hold the flat part of the soldering iron bit tip upwards (as shown in Figure 5.17) and lay the stripped end of the connecting wire on the tip's flat surface
  - apply the end of a length of solder to the wire (not the soldering iron bit tip) — this means you can see when the wire is hot enough, as the solder begins to melt
  - when enough solder has flowed over the lead ends to fully cover the wire, remove the soldering iron and solder
  - allow the lead to cool before touching or moving it.

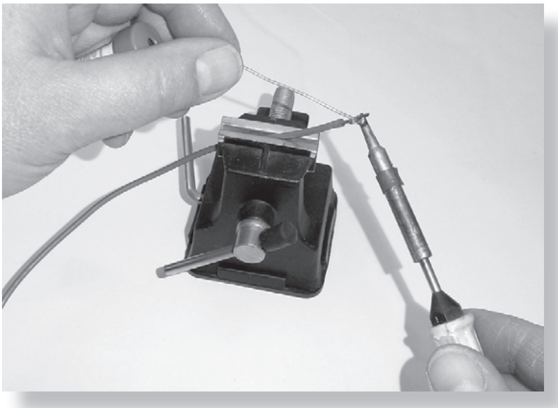

FIGURE 5.17 Tinning a connecting lead – do not heat the lead for too long, as insulation may melt or deteriorate.

CONTINUED...

…CONTINUED

- **TIN the printed circuit board.** At the point where the tinned lead is to be connected to the printed circuit board's copper foil track, the copper must be tinned, to remove oxide, dirt and grease, and to apply solder ready for the joint to be made between board and lead. To tin the printed circuit board track:

  - clean and tin the soldering iron bit tip
  - apply the soldering iron bit tip to the copper track of the printed circuit board, at the point where the lead is to be connected — as shown in Figure 5.18 — and wait for a few seconds

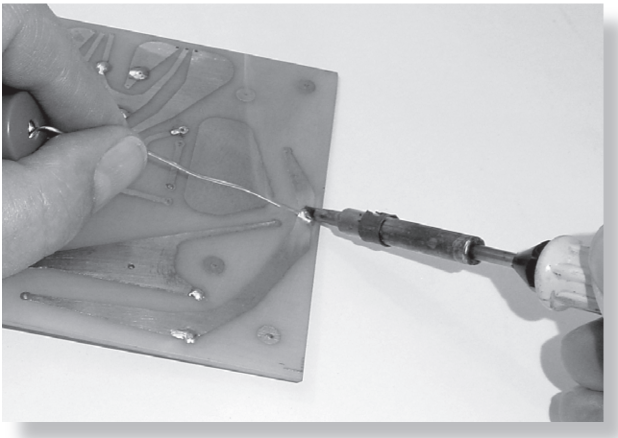

FIGURE 5.18 Tinning the copper foil track of a printed circuit board – apply the soldering iron bit tip first, and wait for a few seconds before applying solder.

  - apply the end of the flux-cored solder to the track, a small distance (about 2–3 mm) away from the soldering iron bit tip — wait for the solder to melt and flow towards the soldering iron bit tip
  - remove the soldering iron bit tip and the solder
  - allow the printed circuit board to cool before moving or touching the solder.

Take Note

While it's correct to tin the printed circuit board copper foil track at the points where connecting leads are to be jointed, this is NOT the procedure where components are to be fitted to the printed circuit board – if the copper foil track is tinned where a component lead is to be inserted, the hole for the lead may be closed with solder.

Once the lead is tinned, it can be soldered to the copper foil track of a printed circuit board. There are three main ways this can be done, soldering the lead:

- **1 directly to the printed circuit board copper foil track**
- **2 to a terminal pin, inserted through a hole in the copper foil track**
- **3 inserting the lead through a hole, purposely designed into the printed circuit board's copper foil track.**

The procedures you should follow for all three ways are detailed below.

1 Soldering a connecting lead directly to copper track

- **TIN the soldering iron bit tip, as described on page 116**
- **TIN the connecting lead, as described on page 124**
- **TIN the printed circuit board copper track, at the point where the lead is to be connected, as described on page 125**
- **HOLD the tinned end of the lead on the tinned copper track, as shown in Figure 5.19**

CONTINUED…

...CONTINUED

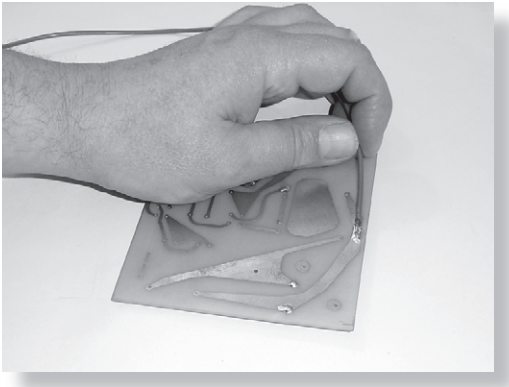

FIGURE 5.19 Position the tinned connecting lead end over the tinned copper track.

- **APPLY** the soldering iron bit tip to the tinned lead end (not the copper track), pressing down on the lead end to hold it in position, and wait for the solder on both the lead end and the copper track to melt, as shown in Figure 5.20

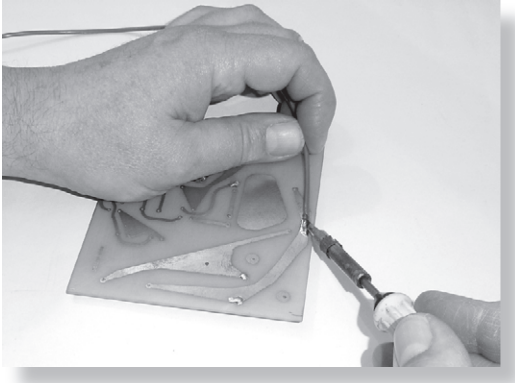

FIGURE 5.20 Heating the connecting lead end and copper track – the solder already present in the tinning on both the lead end and the track will be sufficient to make the soldered joint.

CONTINUED...

…CONTINUED

- **REMOVE** the soldering iron, but hold the lead without moving it

- **WAIT** for the solder to solidify before moving the printed circuit board or letting go the lead.

2 Soldering a connecting lead to a terminal pin

- **INSERT** a terminal pin into the printed circuit board — note that the pin goes through from the copper foil track side of the printed circuit board to the component side of the board ie, the connecting lead will be soldered to the component side NOT the copper track side (see Figure 5.21).

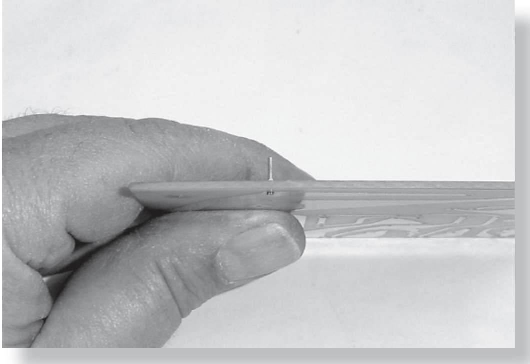

FIGURE 5.21 A terminal pin goes from the copper track side of a printed circuit board, to the component side.

- **CLEAN** and tin the soldering iron bit tip, as described on page 116

- **TIN** the connecting lead, as described on page 124

CONTINUED…

...CONTINUED

- **APPLY** the soldering iron bit tip to the terminal pin and wait for a few seconds (See Figure 5.22)

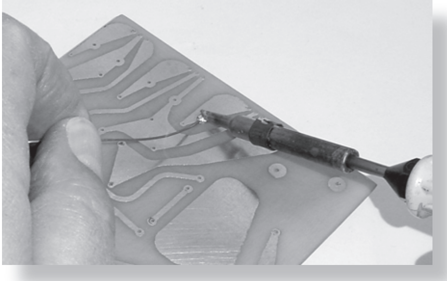

FIGURE 5.22 Soldering a terminal pin to a printed circuit board.

- **APPLY** solder to the terminal pin

- **WHEN** the solder melts and flows over the terminal pin and the copper track, remove the soldering iron and wait for the soldered joint to cool

- **TURN** the printed circuit board over, then tin the terminal pin

- **HOLD** the tinned end of the connecting lead against the tinned terminal pin, as shown in Figure 5.23

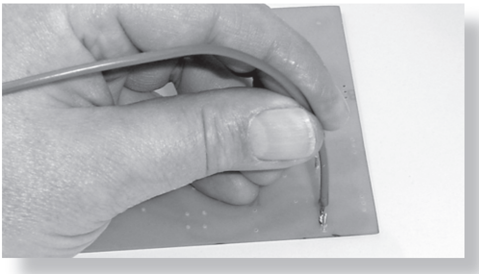

FIGURE 5.23 The connecting lead should be vertically aligned with the terminal pin, and touching it.

CONTINUED...

...CONTINUED

- **APPLY** the soldering iron bit tip to the connecting lead and terminal pin
- **WAIT** for a few seconds for the solder on both the lead and pin to melt, then remove the soldering iron
- **DO** not move either connecting lead or printed circuit board, until the solder has solidified.

Hint

For both these methods of soldering a connecting lead to a printed circuit board, be extra careful not to allow the connecting lead end to move at all once the joint has been soldered and during the cooling phase, or a defective joint may result.

3 Soldering a lead to a hole

- **CLEAN** and tin the soldering iron bit tip, as described on page 116
- **INSERT** the connecting lead through the hole in the printed circuit board — do not tin the connecting lead first, or it will not go through the hole!
- **APPLY** the soldering iron bit tip to the joint, touching both the copper foil track of the printed circuit board and the lead, as shown in Figure 5.24

CONTINUED...

...CONTINUED

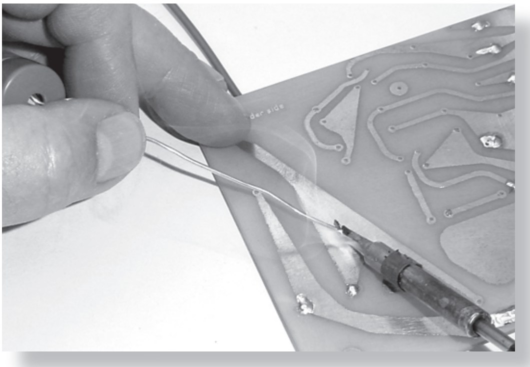

FIGURE 5.24 Soldering a connection lead to a printed circuit board.

- **APPLY** solder to the joint
- **WHEN** the solder melts and flows over the connection lead and the copper track, remove the soldering iron and wait for the soldered joint to cool
- **DO** not move either connecting lead or printed circuit board, until the solder has solidified.

Unsoldering

There are times when it's necessary to unsolder components from a printed circuit board. You may, perhaps, have inserted and soldered a component incorrectly into place, or a particular component may be faulty. In such cases, unsoldering is the only option.

There are two basic methods of unsoldering components (detailed overleaf), and both rely on having extra tools: either solder braid, or a desoldering pump.

Using solder braid

- **TIN** the soldering iron bit tip

- **PLACE** the end of the solder braid over the joint to be desoldered, as shown in Figure 5.25

- **APPLY the soldering iron bit tip onto the solder braid.** Press gently but firmly against the joint. After a few seconds or so, the solder will melt, and some will wick into the solder braid

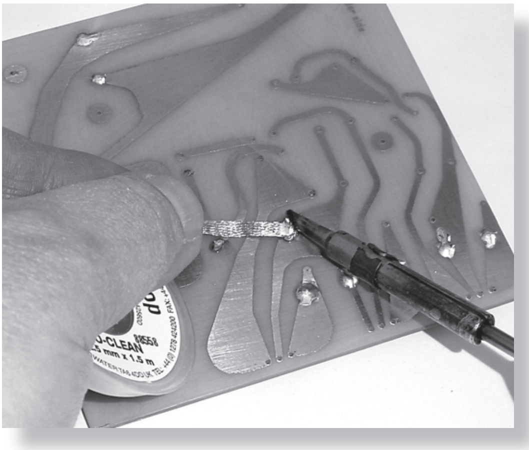

FIGURE 5.25 Desoldering using solder braid – the molten solder wicks into the strands of the solder braid away from the joint.

- **REMOVE the soldering iron bit tip.** Place a fresh portion of solder braid onto the solder joint and repeat the last step until all the solder forming the joint has been removed

- **REPEAT** the process for all component joints

- **REMOVE** the component.

Hint

As you use a portion of solder braid to unsolder a joint, it will become loaded with solder so cease to be usable. Using a pair of side-cutters, trim off the used portion to maintain a fresh solder braid end.

Hint

Sometimes not all solder can be removed from a joint (very often a small amount which cannot be removed will still hold the component lead to the copper track). If this occurs, you will find that you have to heat the joint with the soldering iron bit tip one last time, before quickly – before the solder solidifies – releasing the component.

Using a desoldering pump

- **TIN** the soldering iron bit tip
- **PRIME** the desoldering pump, by pressing the charger down, as shown in Figure 5.26

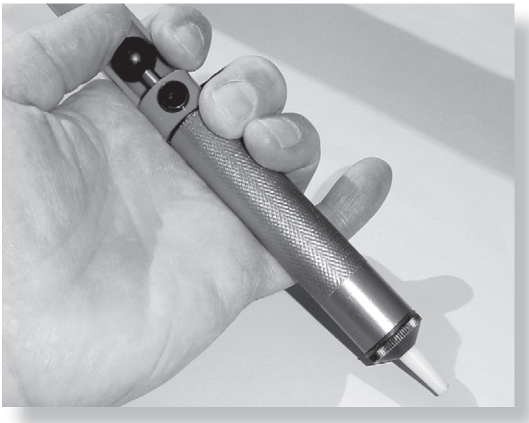

FIGURE 5.26 Priming the desolder pump by pressing down the charger.

- **APPLY** the soldering iron bit tip to the soldered joint and wait for the solder to melt

CONTINUED...

...CONTINUED

- **APPLY the nozzle of the desoldering pump to the molten solder, as shown in Figure 5.27**

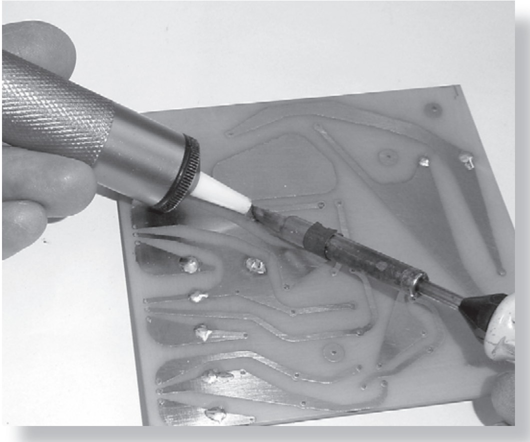

FIGURE 5.27 With the soldering iron bit tip heating the solder, apply the desoldering pump nozzle.

- **PUSH the pump's release button.** The vacuum will suck the molten solder away from the joint and into the desoldering pump chamber, where it will solidify

- **REPEAT the previous four steps as required, until all the solder has been removed from the joint.**

Take Note

Every now and again, the desoldering pump chamber will fill with solidified solder to the point that it doesn't work efficiently. Empty it by repeatedly priming and discharging the pump, so that the solder is forced out of the nozzle.

Take Note

Unsoldering usually requires considerably more application of heat than soldering – the soldering iron bit tip is applied to the joint more times and for longer periods than when soldering.

Remember that components may be damaged by excessive heat, so space the desoldering operations out to allow the components to cool down in between.

Care of your soldering iron

Your soldering iron is a tool — a vital one when it comes to building electronic circuits, of course — and like any tool needs looking after, to keep it in good working order throughout its potentially long life. Things are made a little complicated in that the soldering iron is — I'll state the obvious here — very hot, so you also need to take certain safety precautions while using it. But, fortunately, on the other hand it's a fairly simple tool, so the things you need to do to look after it and the precautions you need to take while you're using it aren't excessive.

Looking after your soldering iron

- **USE a stand.** When at rest, a soldering iron can be easily knocked or moved — it may fall off the work surface or simply touch against something else. Use a stand (as shown in Figure 5.28) when you are not actually soldering with it

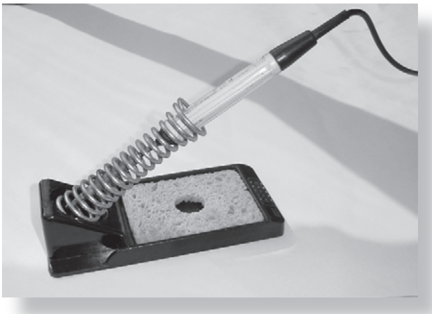

FIGURE 5.28 Use a soldering iron stand when not actually using your soldering iron. Keep the safety risks to a minimum.

CONTINUED...

...CONTINUED

- **TIN the soldering iron bit tip regularly — even at rest.** When you solder with a soldering iron it is regularly in contact with fresh flux-cored solder, so the bit tip is maintained in good condition. On the other hand, when you are not actually soldering and a soldering iron is standing unused but still turned on, the bit tip can become depleted of solder and flux due to the heat, so it oxidizes and become unusable. If you are not going to solder with the soldering iron for a while, remember still to tin it every few minutes

- **TURN the soldering iron off when unused.** To reduce safety risks to an absolute minimum, and to help prevent the soldering iron bit tip from becoming unusable, turn the soldering iron off if you are not going to use it in the next ten minutes or so

- **CLEAN the soldering iron bit tip when cold by lightly rubbing with a nylon pad.** This may help a soldering iron bit tip that will not wet — but do not use wire wool or emery paper to clean the soldering iron bit tip, as this will remove protective plating and shorten the bit life

- **NEVER put a soldering iron — hot or cold — into liquid**

- **CHECK the soldering iron's cable regularly for burns.** A soldering iron when on is, of course, very hot, and accidental burning on the mains cable is a possibility.

Printed circuit board links

If a circuit is complex, it's probable that the printed board copper foil track layout will be complicated too. Under these circumstances, it's often the case that it's simply not possible to design a copper foil layout in a single layer. One trick to get around this problem is to use the component side of the board as well as the copper track side to complete the circuit connections. This is done with the use of simple links of wire, which connect from one point on the copper foil track, go over the track on the component side, then connect to another point on the copper foil track. Figure 5.29 shows an example of a printed circuit board which uses links.

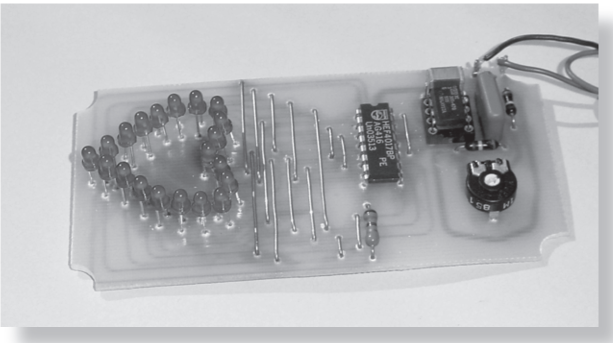

FIGURE 5.29 A printed circuit board that uses links to aid in copper foil track circuit layout. Note that many copper foil connections can take place under links, thereby easing a layout considerably. This particular printed circuit board is the Heart Flasher project, in Chapter 8.

Making a link:

- **CUT a short length of single-strand wire.** This should be a little longer than the distance the link needs to span between holes on the printed circuit board
- **BEND one end of the link around the nose of a pair of long-nosed pliers — see Figure 5.30.** The bend should be at 90°, and there should be sufficient wire to go through the board and be soldered to a copper track

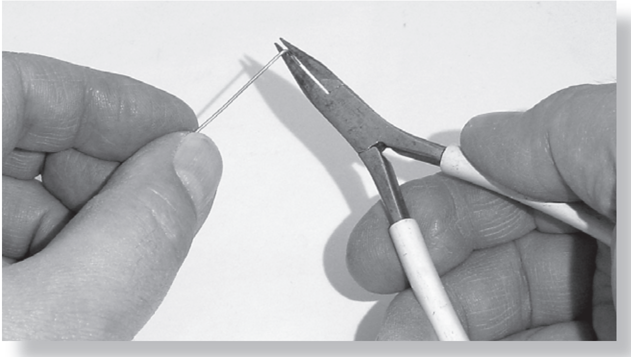

FIGURE 5.30 Bend one end of the link, using long-nosed pliers.

CONTINUED...

...CONTINUED

- **MEASURE the link length required**

- **BEND the other end of the link around the long-nosed pliers — see Figure 5.31.** The length of link between bends should be the same as that you've just measured

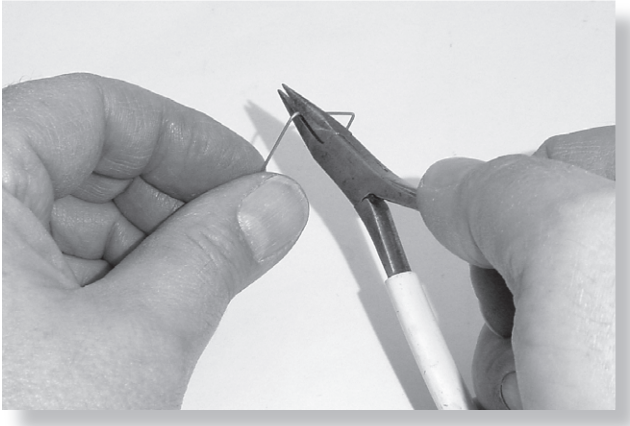

FIGURE 5.31 Bend the other end of the link.

- **INSERT and solder the link**

- **TRIM excess link wire off the soldered joints.**

6

Housings

After the printed circuit board, the next most important part of an electronic project is the *housing* (sometimes called an *enclosure* or *case*). OK, yes, while it's perfectly possible to use an electronic circuit just at its printed circuit board stage (complete with controls and connections hanging off…), it's certainly not the best method of ensuring your circuit has a long working life. Only the housing can 'finish' an electronic circuit off, and turn it into a complete design.

More than just the practical aspect of ensuring the internal electronic circuit is securely and safely held, however, in many respects the housing forms the main (if not the only) interface between the electronic circuit and the user. Yes, the electronic circuit may have connectors for inputs and outputs, and have controls that the user adjusts, but the housing is the means whereby all the controls and connectors are presented for use. How effectively this occurs depends very much on the housing, and how it has been designed. And there's a world of difference between a good housing and an acceptable one.

In fact, to emphasize how important an electronic circuit's housing is, we should look at a complete electronic circuit and see how it breaks down into a total of six levels of interconnection — as shown in Figure 6.1. In this look at the electronic circuit, we have to consider an integrated circuit as being a major component within the circuit itself, in order to understand the various levels and how they interact. Normally, of course, we would most likely use an integrated circuit, without understanding — or even caring — how it is constructed internally. This example, though, shows just how important each particular level is, with regard to the whole electronic circuit design.

The six levels are:

- **LEVEL 1 — on-the-device.** Where individual connections between the semiconductive component parts of integrated circuits are made as minute etched wires — shown in Figure 6.1(a)
- **LEVEL 2 — device-to-package.** Where the semiconductor die is permanently attached, with thin aluminium or gold wires, to the terminals of its package — shown in Figure 6.1(b)
- **LEVEL 3 — package-to-board.** Where flat, round or square package terminals are connected to printed circuit boards, as permanent (ie, soldered) or separable connections — shown in Figure 6.1(c)
- **LEVEL 4 — board-to-board and on-board.** Where components are interconnected with the conductive track of the printed circuit board, and where printed circuit boards are connected to each other by permanent or seperable connectors and cables — shown in Figure 6.1(d). The housing itself is directly involved at this level, simply because the types of connectors you use, and the printed circuit board positioning with respect to these connectors, may well be defined by the housing
- **LEVEL 5 — board-to-housing.** Where connections between a printed circuit board and its housing are made with permanent or separable connectors — shown in Figure 6.1(e). Sometimes, a distinct chassis may be involved between printed circuit board and housing
- **LEVEL 6 — housing-to-housing.** Where separable connections between housings are made with connecting cables — shown in Figure 6.1(f).

Figure 6.1 shows quite graphically what levels of an electronic circuit design the housing affects. Levels 1 to 3 (Figure 6.1(a) to (c)) have no involvement with the housing at all. They are merely concerned with connections between the individual components' innards and the printed circuit board. On the other hand, levels 4 to 6 (Figure 6.1(d) to (f)) are all indeed affected by the housing or housings you might choose for your electronic design.

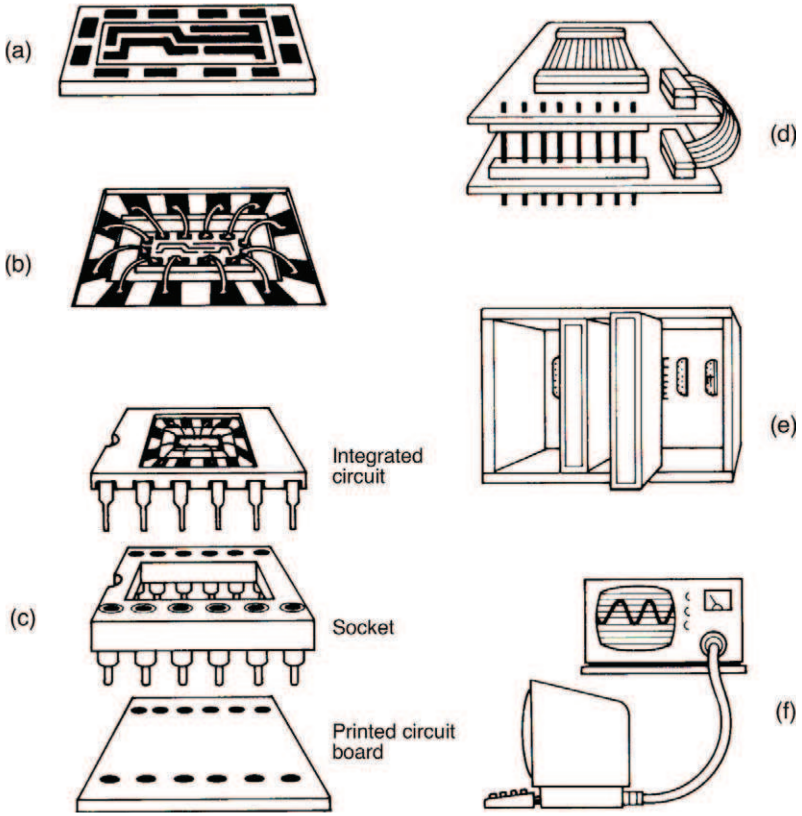

FIGURE 6.1 Six levels of interconnection between parts of an electronic design: (a) semiconductor die; (b) from the die to the terminals of the integrated circuit package; (c) from the integrated circuit to the printed circuit board; (d) between printed circuit boards; (e) between printed circuit boards and the housing; (f) between housings.

Yet, despite this fact that the housing affects or is affected by no less than half of the total electronic design, most magazines and books in general pay little attention to the housing, other than a few spartan words about putting the printed circuit board in to some sort of box. Indeed, only rarely is the actual housing design taken into consideration at all.

This chapter aims to put that right. Hopefully, by the end of reading it, you will see that the housing is as important a part of the total electronic device as the circuit design itself.

What's your housing made of?

There are three main types of housing available to the beginner in electronic construction:

- **PLASTIC**
- **METAL** diecast
- **METAL** folded chassis-type.

FYI

Actually there are more housing types, but these three types (and their derivatives) form the largest bulk of housings you will come across to begin with. As you progress in electronics you will undoubtedly meet other types.

Figure 6.2 shows examples of these housing types.

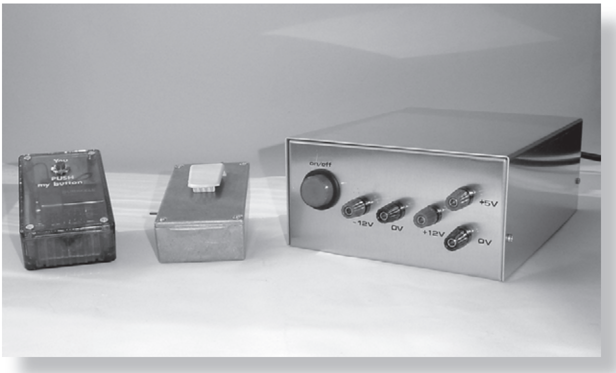

FIGURE 6.2 Housings – from left to right – plastic, metal diecast, metal folded chassis-type.

Plastic

Plastic housings come in several forms. They are relatively cheap to make in bulk, very easy to use, and provide a neat method of housing electronic circuits. Being usually moulded or extruded, some quite intricate and novel plastic housing shapes are available, which means that it's often possible to find a plastic housing that is a good match for an electronic circuit's function — for example, a plastic housing is shown in Figure 6.3 that is provided with a belt clip, and internal battery space, which would make an ideal housing for something like a music player or a pocket game.

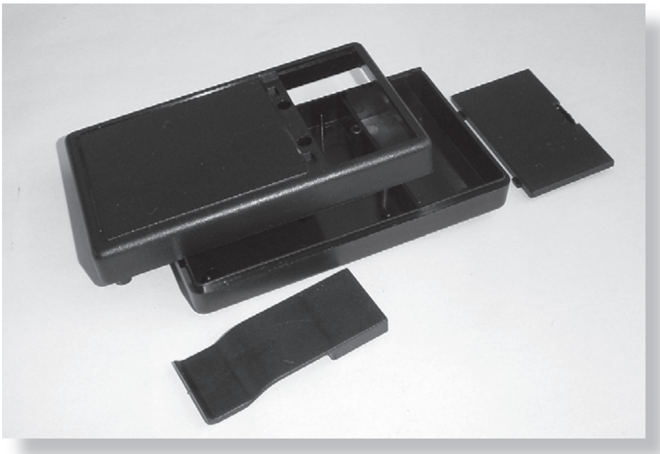

FIGURE 6.3 A small plastic housing, complete with belt clip and battery compartment.

Metal diecast

Made from cast aluminium, diecast housings are extremely tough, yet quite easy to work with. They are not, on the other hand, the most attractive of housings, so probably wouldn't be your first choice to house an electronic circuit that you wanted to use, say, on display in your living room, but a diecast box would be great for, say, a small pre-amplifier for use between an old record deck and a modern hi-fi amplifier.

Metal folded chassis-type

Usually made from just two thin sheets of aluminium, each folded to complement, metal folded chassis-type housings make very effective enclosures for many electronic circuits. The bottom half of a housing is where a printed circuit board and controls are fitted, while the top half of the housing forms a neat lid, allowing full access when removed, and protection when fitted. Often, the top half of such housings are plastic coated, or painted, which gives quite an attractive appearance to the finished electronic device.

Other housings

There are some other housing types that you should be aware off too, notably those made from extruded aluminium. Such housings are very tough, and have internal slots which are useful for the mounting of printed circuit boards — boards just slide into position from one end of the housing.

IP rating

In certain environments such as the great outdoors, it's necessary to use a housing that protects the electronic circuit inside from the elements to some degree or other. While all housings will protect their innards to some small measure, the problem is obviously magnified where the electronic device is to be used in such possibly detrimental environments as outside in the rain. For many years, there has existed a method of determining whether an enclosure is suitably protected against the environments in which it's expected to be used. This is a standard method, called the *ingress protection* (*IP*) rating system. It comprises two or three digits, and each digit refers to a particular level of protection, for a particular type of protection requirement. Thus, an enclosure might be given an IP rating of, say, IP63, or another might have an IP rating of IP485. Table 6.1 gives the IP ratings and numbers.

Where only two digits are given in an enclosure's IP rating, the third of these requirements (impact damage) is not relevant so should be ignored.

TABLE 6.1 Ingress Protection (IP) ratings by number, and their equivalent protection levels.

Value	1st digit	2nd digit	3rd digit
	Protection against ingress of solids	**Protection against ingress of liquids**	**Protection against mechanical impact damage**
0	No protection	No protection	No protection
1	Protected against solid objects over 50 mm eg, hands, large tools	Protected against vertically falling drops of water	Protected against 0.225 joule impact (150 g @ 15 cm)
2	Protected against solid objects over 12 mm eg, hands, large tools	Protected against direct sprays of water up to 15° from vertical	Protected against 0.375 joule impact (250 g @ 15 cm)
3	Protected against solid objects over 2.5 mm eg, wire, small tools	Protected against direct sprays of water up to 60° from vertical	Protected against 0.5 joule impact (250 g @ 20 cm)
4	Protected against solid objects over 1.0 mm eg, wires	Protected against water sprayed from any direction. Limited ingress permitted	
5	Limited protection against dust ingress (no harmful deposit)	Protected against low pressure water jets from any direction. Limited ingress permitted	Protected against 2.0 joule impact (500 g @ 40 cm)
6	Totally protected against dust ingress	Protected against high pressure water jets from any direction. Limited ingress permitted	
7		Protected against immersion between 15 cm and 1M	Protected against 6.0 joule impact (1.5 Kg @ 40 cm)
8		Protected against long periods of immersion under pressure	
9			Protected against 20 joule impact (5 Kg @ 40 cm)

Housings and their use

There are several tasks you will have to carry out which are common, whatever housing you opt to use for your electronic device.

The main tasks are:

- **MOUNTING the printed circuit board**
- **FITTING controls and switches**
- **FITTING connections between the printed circuit board and the outside world**
- **MOUNTING other internal components besides the printed circuit board**
- **WIRING of all internal parts**
- **LABELLING the front and rear panels of the housing.**

Mounting printed circuit boards

One of the primary purposes of a housing is to hold its internal electronic circuit's printed circuit board. It's not just a case of throwing the printed circuit board in, and closing the lid though. A printed circuit board should never be left loose inside any housing. Instead it should be properly and securely fixed, using one or more of a number of methods.

Often, the method you might choose to mount your printed circuit board inside your housing will depend on the type of housing you have opted to use *and* the size and shape of the printed circuit board. If you have a small printed circuit board and a plastic housing, it's perfectly possible to mount the printed circuit board to the inside of the housing using one or two pieces of double-sided adhesive sponge fixing pads, as shown in Figure 6.4.

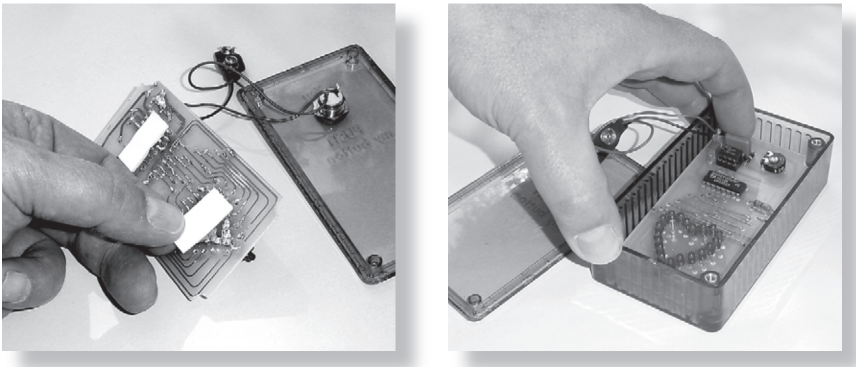

FIGURE 6.4 **Using double-sided adhesive sponge fixers to mount a printed circuit board inside a plastic housing: (left) applying the sponge fixing pads to the bottom of the printed circuit board; (right) applying the printed circuit board with sponge fixing pads to the inside of the plastic housing.**

This method should not be used to mount a printed circuit board inside a metal housing though — for the simple reason that sponge fixing pads are thin, and some of the soldered connections on the bottom of the printed circuit board may touch the metal housing base (as shown in Figure 6.5), causing short circuits and possibly damage.

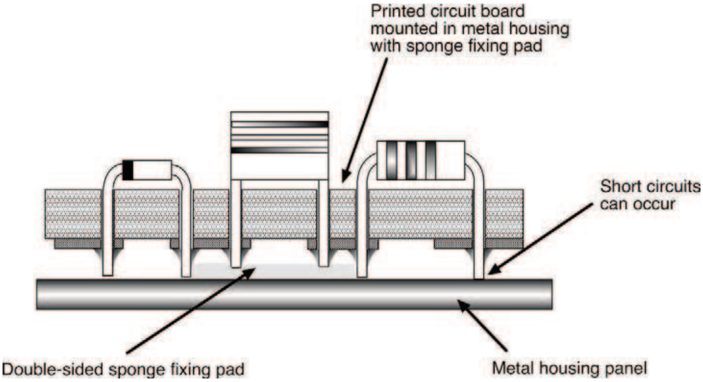
Printed circuit board mounted in metal housing with sponge fixing pad
Short circuits can occur
Double-sided sponge fixing pad
Metal housing panel

FIGURE 6.5 **A printed circuit board should not be mounted inside a metal housing using sponge fixing pads, or short circuits may result!**

Instead, when using a metal housing (or, indeed, if a larger printed circuit board is to be mounted even in a plastic housing) something more substantial must be used.

One of the best methods of mounting a printed circuit board in a housing is to use spacers. These are available in several formats. Some spacers use small bolts to hold them in place to the inside of a housing, as shown in Figure 6.6. Once the spacers are bolted to the housing, the printed circuit board is then bolted to the spacers. Obviously, the housing and the printed circuit board need to be drilled to accept the bolts prior to use.

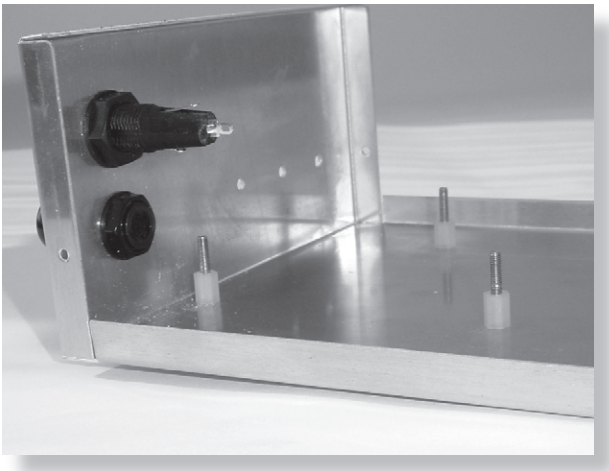

FIGURE 6.6 **Bolted spacers, showing the spacers and the bolts.**

Take Note

If you intend to use spacers to mount a printed circuit board to a housing, you must take into account where the spacers will be positioned on the printed circuit board. Component positions and the copper foil track may have to be adjusted to provide room – not just for the spacers themselves, but to allow access with tools to allow you to tighten the spacer bolts. You need to plan ahead, and decide which printed circuit board mounting method or methods you intend to use, as all the calculations regarding component positions and copper foil track layout need to be done in the very early stages of printed circuit board design – not at the last minute!

Other spacers are preformed, sprung-loaded plastic clips, which push into holes in the housing and the printed circuit board.

There are several variations of printed circuit board guides available. In use, these are fastened to the inside of the housing, and the printed printed circuit board is slid into position along them.

A particular variation of housing — extruded aluminium types — features in-built slots in internal surfaces. Conveniently, these are printed circuit board-sized, so provide an extremely easy method of mounting.

Take Note

In the printed circuit board design stages, make sure no components are positioned close to the edges of the printed circuit board if you plan to use guides of any sort. Further, if you plan to use an extruded aluminium housing, not only must there be no components fitted on the guide edge areas of the printed circuit board, but there must be no copper foil track in those areas.

If your printed circuit board has controls and connections (such as certain printed circuit board mounted potentiometers, switches and connectors) then you can use the controls and connectors to mount the printed circuit board inside the housing, with careful design. This is particularly so for small printed circuit boards.

Mounting the printed circuit board this way depends on using printed circuit board mounted controls and connectors. For example, Figure 6.7 shows a printed circuit board that has an on-board potentiometer and on-board connectors. These effectively hold the printed circuit board in place, as they are held to the housing sides.

Note though, that the actual printed circuit design is highly dependent on the actual controls or connectors. In other words, the housing you intend to use, the exact size and shape of the printed circuit board, *and* the actual controls or connectors all need to be taken fully into account during the very earliest printed circuit board design stages. Literally *everything* is dependent on each other. This method allows a highly integrated, a very professional looking, and often a much reduced size of device, so is often worth the extra time you might need to invest in the early design stages.

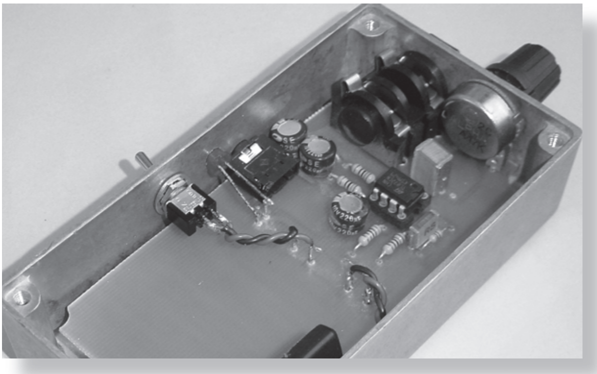

FIGURE 6.7 Using a printed circuit board mounted potentiometer and connectors to mount a printed circuit board inside its housing. This is the Guitar Headphone Amplifier project, in Chapter 8.

A common method of mounting small printed circuit boards inside housings is to make use of a particular type of connector known as an *edge connector*. Edge connectors are multi-pinned devices, comprising a row of contacts into which the edge of a printed circuit board fits tightly. A row of copper foil tracks on the edge of the printed circuit board mates with the edge connector contacts, as shown in Figure 6.8. The edge connector is bolted to the housing or to another — usually larger — printed circuit board. Thus, connections can be made to and from the printed circuit board, but just as importantly, the printed circuit board is held in position. A small printed circuit board mounted this way may need no other support.

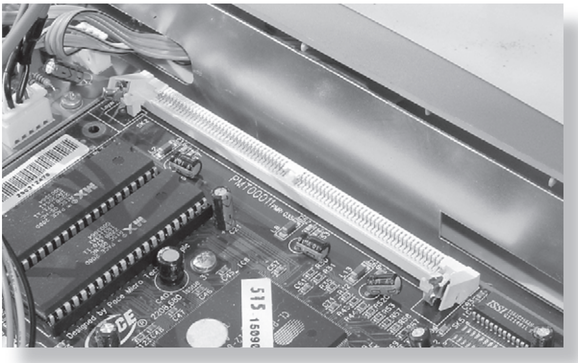

FIGURE 6.8 Using an edge connector to hold a printed circuit board in position. This particular connector (in white) is a RAM-card holder, common to computer and associated equipment.

Fitting controls and switches

While on the subject of controls and switches, we should include details on how they should be mounted to a housing. The methods of doing this are similar, whether they are special printed circuit board mounted controls and connectors, or ordinary, more usual, controls and connectors. The methods are also similar, whatever type of housing is used — metal or plastic, or even wood.

Controls are required in most electronic circuits. They can range from volume and tone controls in an audio amplifier, through time settings for a digital clock, simple on/off switches, or dials to tune in radio stations. The important point is that controls are used to provide an interface between an electronic circuit (in printed circuit board form, inside a housing) and the user of the electronic circuit (in human form — normally — outside the housing).

Potentiometers

Most potentiometers have a threaded stem, that allows them to be mounted to a housing using a supplied nut and washer. Figure 6.9 shows a typical potentiometer, complete with washer and nut. A hole — the diameter of the threaded stem — is required in the housing to insert the stem. Also visible in Figure 6.9 is a location stub that extends beyond the mounting point on the threaded stem. The purpose of the location stub is to fix the potentiometer in position and prevent its rotation — as might happen if the nut becomes loose, over time, say. The problem with the location stub is that a second small hole (the diameter of the location stub) is required in the housing alongside the hole for the threaded stem — the correct distance from the centre of the threaded stem's hole.

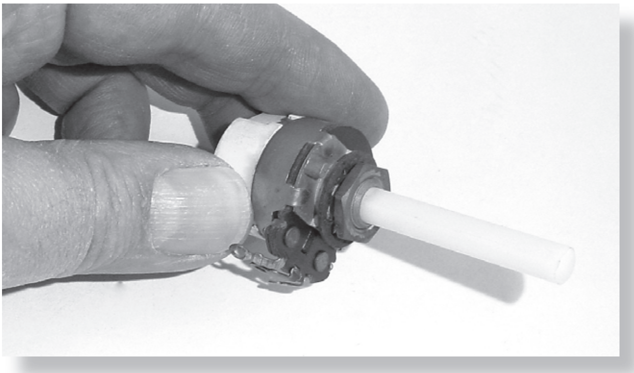

Figure 6.9 A potentiometer, showing its threaded stem, washer and nut, and the location stub.

Now, I should point out here that not all potentiometers have this location stub, and in which case the second hole is not required. And, indeed, I have been guilty myself of removing the location stub from a potentiometer in cases where I felt the nut wasn't likely to become loose, so didn't need the second hole.

Notwithstanding, the job to mount a potentiometer with such a stub is simple enough.

Mounting a potentiometer

- **MARK the centre of the hole where the potentiometer's threaded stem is to be, and mark the centre of the location stub's hole, as shown in Figure 6.10.** The distance between centres depends on the particular potentiometer, as does the angle at which you need to position the two holes relative to horizontal along the line of the housings panel.

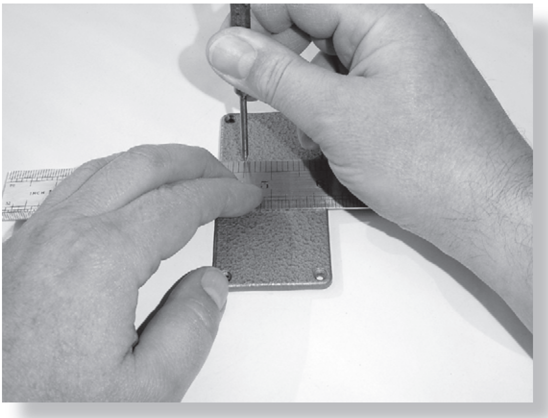

FIGURE 6.10 Marking the position of the two holes required for the potentiometer of Figure 6.9.

- **MARK both holes using a centre punch**
- **DRILL the holes using a pilot drill (no more than 2.5 mm).** Position the housing over a block of wood, so that the drill bit doesn't grab the metal and distort it as the drill bit's cutting edge exits the hole.

CONTINUED...

…CONTINUED

- **DRILL** both holes using a drill the correct size for the location stub
- **DRILL** the threaded stem's hole using the correct sized drill
- **POSITION** the potentiometer in place
- **PLACE** the washer, followed by the nut over the threaded stem.
  Tighten the nut to a finger tightness
- **TIGHTEN** the nut, using pliers.

Take Note

When you tighten a potentiometer nut (indeed, any nut which is located on a housing's panel like this), you should take extra care to make sure the pliers don't slip off the nut and scratch the panel. Figure 6.11 shows how to tighten the nut. Note how the washer underneath the nut protects the panel from the pliers.

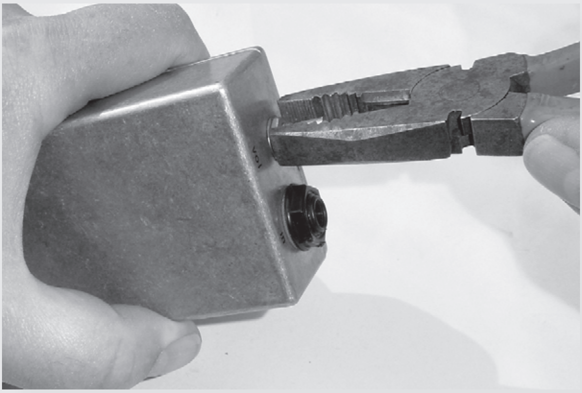

FIGURE 6.11 Tighten the potentiometer nut using pliers. Push the pliers against the washer under the nut to help prevent the pliers from slipping off, thereby protecting the housing panel.

Take Note

New potentiometers have long control stems, in order that you can cut them to the length you want. The final cut length is fixed however, by the type of control knob you want to use. In order that you can cut the stem to the required length, you therefore need to have the control knob to hand. Follow the instructions below to cut the stem to size.

Cutting a potentiometer stem to fit a control knob

- **PUSH** the control knob on the stem till it seats fully against the top of the stem
- **MARK** the stem directly in line with the bottom of the control knob (Figure 6.12)

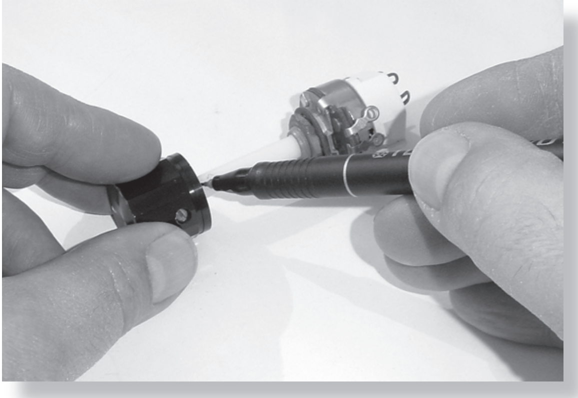

FIGURE 6.12 Push the knob on the potentiometer stem, and mark the stem.

- **REMOVE** the control knob, and measure the distance from the end of the stem to the mark (Figure 6.13)

CONTINUED…

...CONTINUED

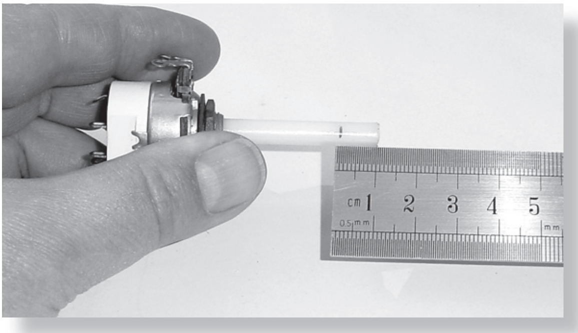

FIGURE 6.13 Measure from the stem end to the mark.

- **NOW make a second mark on the stem the same distance from the top of the potentiometer's threaded stem.** Actually, some knobs have an overhang specifically intended to cover the threaded stem, so this second mark can be a millimetre or two closer to the stem end if your knob features such an overhang

- **CUT the stem on this mark.** To do this, hold the end of the stem in a vice, with the stem horizontal. Hold the potentiometer body in one hand, while cutting the stem with a junior hacksaw in the other hand, as shown in Figure 6.14

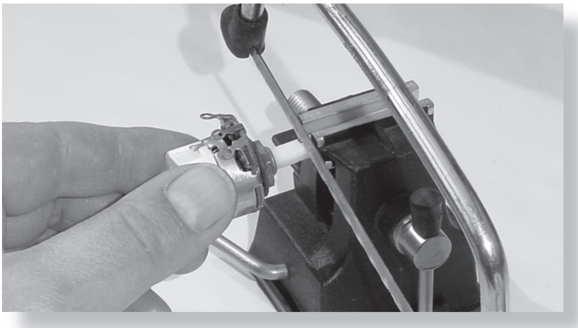

FIGURE 6.14 Cutting the stem of a potentiometer.

- **DEBURR the cut end of the stem with a fine file.** Your potentiometer is now ready to be fitted with its knob.

Take Note

Potentiometers are actually quite fragile. Do not hold the potentiometer body in a vice to cut off excess stem. Make sure you put the stem end in the vice!

Knobs

Many, many types of knobs are available, in all sorts of sizes and styles. Figure 6.15 shows a selection. Most are fitted to a potentiometer either with a push-fit, or (usually) with a small grub-screw through the side, which clamps onto the potentiometer stem.

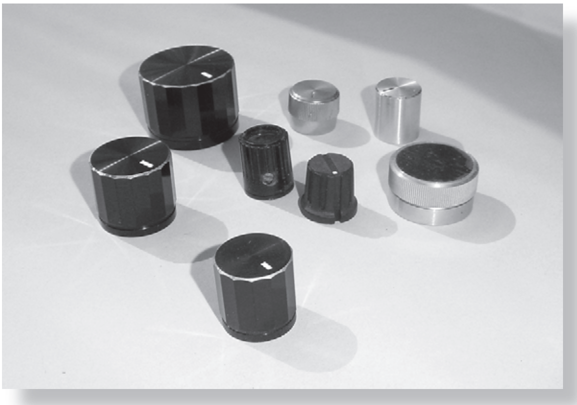

FIGURE 6.15 A selection of common control knobs.

Switches

Several variations exist in switches, with slide, toggle, push, and rotary being the most common (see Chapter 3). Within these variations, however, there are also variations in body shapes and sizes. As a result, it's not possible to include details of how to mount all switches here. Instead, I'm going to generalize, and show how the main variations should be mounted. These main variations are determined by body shape, not internal mechanics. Main body shapes are circular and rectangular, which narrows the field down somewhat.

Circular switches are rather like potentiometers in their fitting, and most of the details in the potentiometer section already covered are relevant here. Some circular switch types even have a location stub, usually in the form of a shaped washer, which requires a second hole (just as with potentiometers) to correctly mount the switch. Figure 6.16 shows a switch which mounts using this method. The method of mounting closely follows that of mounting potentiometers, earlier. Both holes should be marked out and drilled, then the switch is fitted using the location tab washer and a nut which fastens over the switch's threaded stem. As when tightening a potentiometer's mounting nut, take great care when tightening a switch's mounting nut, in case the pliers slip off and scratch the housing's panel.

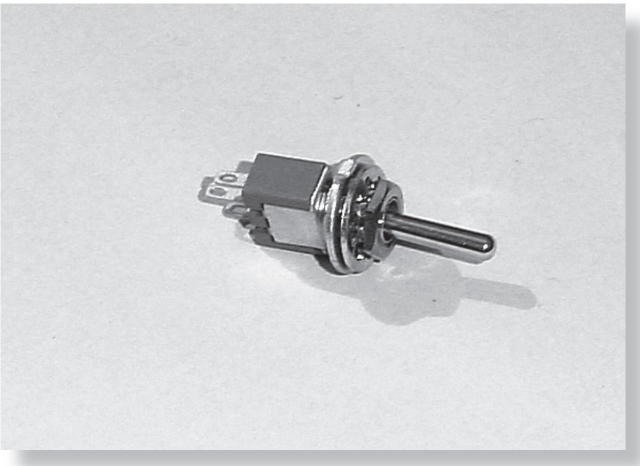

FIGURE 6.16 **A miniature circular toggle switch, complete with location washer having a location tab.**

Some circular switches are provided with two mounting nuts (you can see one on the switch shown in Figure 6.16). These have the purpose of allowing the switch to be adjusted in terms of position, sticking far out from the housing's panel, or closer to the panel. A benefit of this second nut is that the outside nut needn't be tightened with pliers (thus avoiding any risk of damage to the panel outside), and all tightening is done to the inside nut, using pliers.

Rectangular switches are rather harder to mount, simply because rectangular holes are very much more difficult to err… drill. Of course, in general, it's a question of drilling a hole in the panel that is no larger than (but only slightly smaller than) the switch itself, then carefully filing out the hole till it is big enough for the switch to fit. The process is actually easy enough, just somewhat laborious, and follows now.

Fitting a rectangular switch

- **MARK** out the rectangular hole to be made
- **MARK** the centre, using a centre punch
- **DRILL** out a hole (using a pilot drill, and subsequent larger drills), just slightly smaller than the required finished hole
- **FILE** the hole, using fine files to the exact size required.

Hint

When you file a hole in your housing's panel, you'll need to clamp the housing in a vice. To protect the housing from the vice jaws, you should get two small blocks of wood and clamp the housing between the blocks. Make sure the blocks are positioned close to the edge of the hole you are filing – this prevents the panel from flexing too much and being damaged while you file. This is particularly important with thin aluminium folded-metal chassis-type housings. Also, file from the outside in – this means that any burring you naturally get while filing forms on the inside of the housing, so is not visible from the outside when finished.

Some rectangular switches have particular hole requirements. For example, slide switches (other than needing a rectangular hole for the switch tab itself to protrude through) also require holes to suit screws which hold the switch body in place. There's nothing for it but to carefully mark out the holes to drill and file, and get on with it!

▷ FYI ◁

Holes for some common rectangular switch shapes (and some circular ones, for that matter) can be cut out using hole punches, purpose-designed for the task. A hole punch comprises three main parts — a cutting tool, a socket receptacle, and a bolt arrangement. Using them is easy: you drill a hole the size of the bolt diameter, then fit the punch together, with the cutting tool on the front of the housing panel, the socket on the back, and the bolt going through the panel. As the bolt is tightened, the cutting tool punches out the hole in the panel, leaving a very clean hole. Such hole punches are quite expensive, but possibly worth the cost if you intend to make many identical holes.

Hint:

There are so many switch variations that it's usually possible to choose a switch type that is easy to fit. In other words, unless you want to use a rectangular switch for a particular purpose (although personally, I can't think of any!), my advice is to use circular switches always. Remember that every operation you perform to your housing's panel is just another opportunity to scratch the panel.

Fitting connections

Many electronic circuits require some form of electrical signal as an input, or an output. An audio amplifer, say, has outputs to loudpseakers, or a multimeter has inputs from its probes. The connectors used for such connections varely widely, depending on what sort of signals are being connected. Figure 6.17 shows a selection.

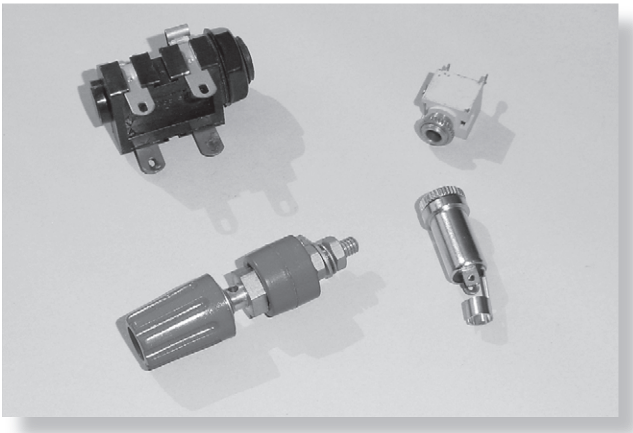

FIGURE 6.17 A variety of connectors used with electronic circuits and their housings.

The mounting of connectors to housings follows the same principles as that of mounting controls such as potentiometers and switches, with only a few minor additions or variations.

Often, circular connectors have a locating key of moulded plastic, which must mate with a notch cut in the panel as the connector is fitted to the housing. The purpose of the key and notch arrangement is simply to prevent the connector rotating. Figure 6.18 shows a connector with this sort of rotation prevention device. It all sounds good, but it does make fitting such connectors a little more tricky. After drilling a hole in the housing's panel to fit the connector, you then have to file out a notch in the panel to fit the key. A set of jewellers' files is best to do that.

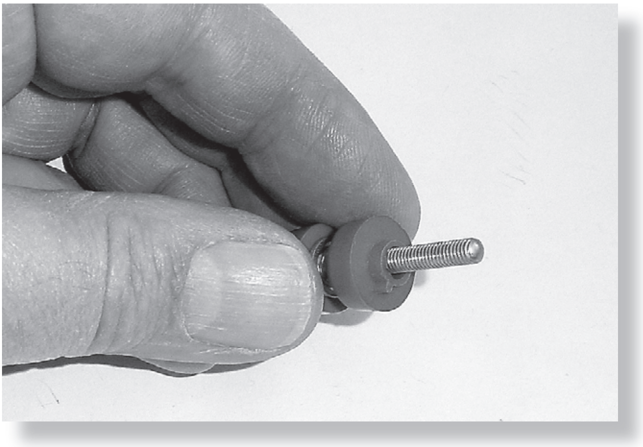

FIGURE 6.18 A connector with a location key – a corresponding notch in the housing panel must be filed (using jewellers' files).

Connectors need to be connected internally — usually to the printed circuit board, with a length of connecting lead. Some connectors may be multi-pole (ie, have two or more individual connections inside the one connector) while others will be single-pole (ie, only one individual connection). The type of connector defines what sort of connecting lead you need to use.

Often multi-pole connectors are audio in nature (microphone, earphones and so on) and they will usually require a special type of connecting lead known as screened cable. Screened cable is formed with a central insulated lead (some screened cable has more than one central insulated lead), wrapped with uninsulated wire, and the whole construction is insulated. Figure 6.19 shows some screened cable, in which you can see the central insulated lead and the uninsulated wire. The uninsulated wire around the central lead is intended to be earthed, so offering a measure of protection (or *screening* — hence the name) for the relatively small signals carried by the cable against electro-magnetic interference.

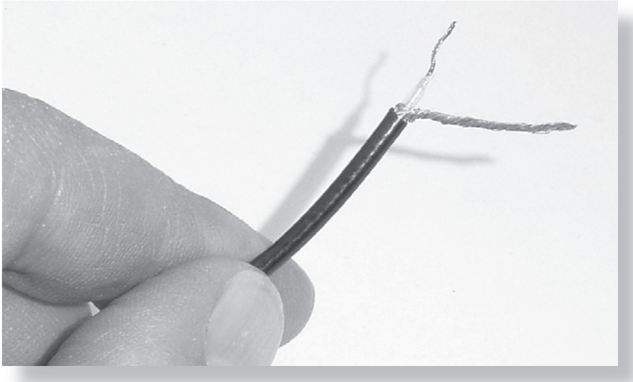

FIGURE 6.19 Screened cable, showing a central insulated lead, and the wire screening, all encased in insulation.

A variant of screened cable, for use with high frequency signals, is coaxial cable. Like screened cable, coaxial cable has a central signal lead, with an external screening. However, the screening is usually in a braided form.

Usually, connecting leads between the connectors and the printed circuit board are made by soldering the lead at each end. However, some connectors (notably power connectors) require some other joining method. The connector of Figure 6.18, for example, features a bolt forming the internal connection. Such a connector requires that solder tags be soldered to the ends of the connecting leads (Figure 6.20 shows a connecting lead with a solder tag connected to it). The solder tag can then be bolted to the connector.

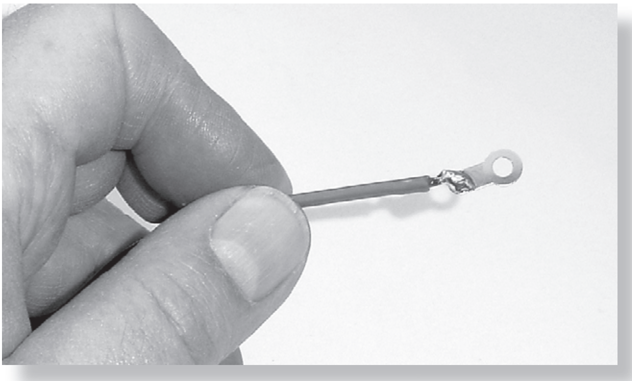

FIGURE 6.20 A solder tag, soldered to connecting lead, for use with bolted connectors.

Other power connectors — usually mains connectors, although the connection method is popular in car electrical circuits too — feature a spade male lug, as shown in Figure 6.21. A mating spade female connector must be crimped to the connecting lead, using a crimping tool. The result is a solid and secure connection.

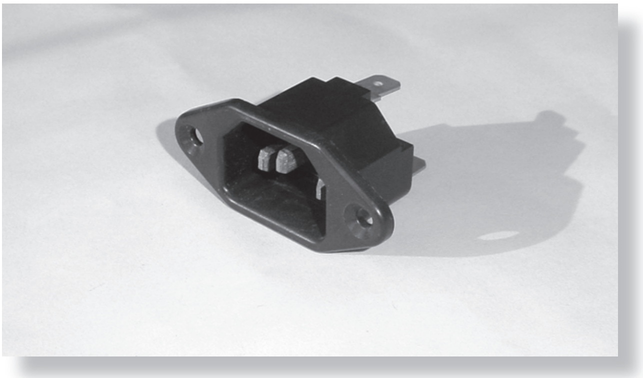

FIGURE 6.21 *Connector featuring spade lugs.*

Some connectors mount directly on to the printed circuit board, which negates the need for leads between the connector mounted on the front panel and the printed circuit board. As a benefit such connectors can be used, as shown on page 148, to aid when mounting the printed circuit board itself in the housing.

Mounting other internal components

A few other internal components that need to be mounted on a housing exist. The most common group is that of indicators — LEDs, bulbs and so on. Generally speaking, they have the same mounting conditions that connectors and switches have, with the exception of LEDs.

Several varieties of LEDs exist. Simplest to mount are the basic, round common LEDs, as shown in Figure 6.22. These *can* be mounted just by drilling a hole in the housing panel the exact diameter of the LED and pushing the LED in.

The problem with this method is that it's just as easy to push the LED back out, from the outside of the panel. The solution is a simple clip, shown in situ on a LED in Figure 6.23. In use, it's a simple matter of drilling a hole in the housing panel the correct size for the clip, inserting the clip into the hole from the panel front, then pushing the LED into the back of the clip, until it seats correctly with a click. Clips come in the same range of sizes that LEDs do — it's just a matter of making sure you have a clip the same size as the LED.

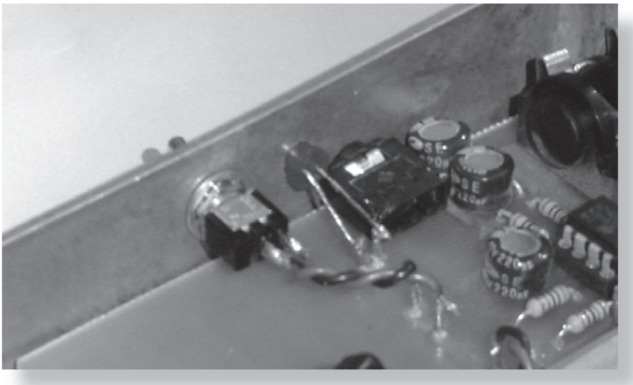

FIGURE 6.22 Simple mounting of a LED to a panel.

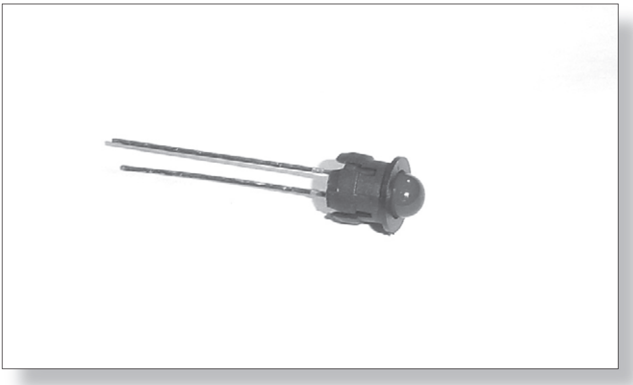

FIGURE 6.23 A LED, complete with panel clip.

Fuse holders

Mains-powered circuits usually require a fuse for safety purposes, and the most common type of fuse holder is chassis-mounted. These are fitted to a housing panel, so that the fuse is accessible from the outside of the housing, while all mains connections are made on the inside (see Figure 6.24). There are no particular considerations in the actual mounting requirements, but there may be wiring precautions (to make sure live terminals cannot be touched when removing or installing a fuse).

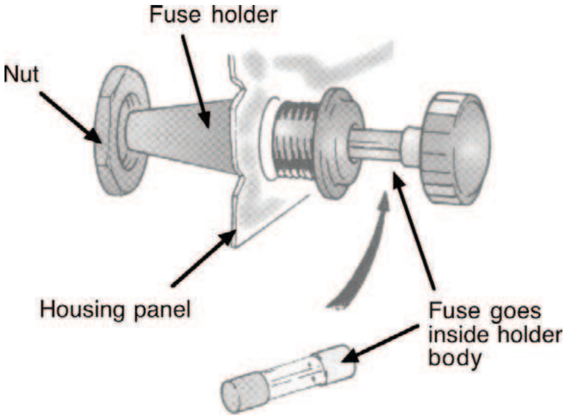

FIGURE 6.24 Mounting a fuse holder into a panel.

Heatsinks

All electronic components generate heat as electric current flows through them. Generally this is not a problem, as they naturally dissipate the heat so that the temperature doesn't get too high to damage them. However, large current flow through certain components such as certain transistors, generates heat to the point that such a component *could* be damaged if the heat was left to dissipate naturally, and some method of removing heat is required to maintain a component core temperature lower than that which would cause the damage.

The method of removing heat — known as *heatsinking* — usually relies on mechanically connecting the component to a large mass of metal. As metal is a heat conductor, heat from the component core is thus shunted away, preventing damage.

Inside an electronic device's housing, components which may heat up to the point of damage can be protected with readily available push-, clip-, or bolt-on heatsinks; a selection is shown in Figure 6.25. Typically, the larger metal surface of the heatsinks allows a degree of cooling just because of increased natural convection. To increase the surface area and air flow around heatsinks, they are often finned.

Most heatsinks of this type are simply fitted to the component in question and then forgotten about, so they provide quite an easy way of providing component cooling. They do, however, only have a limited ability to aid component cooling.

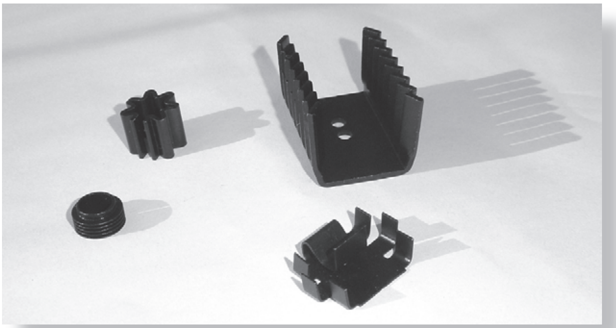

FIGURE 6.25 A selection of small heatsinks.

Where components are likely to get hotter than such heatsinks can cope with, a common trick is to mount the components on a panel of the housing (assuming the housing is metal, of course).

All these heatsinking methods rely on the fact that the medium directly in contact with the hot component will dissipate heat sufficiently to maintain an acceptably low temperature. When deciding which heatsink is required for which component, there are five main variables which need to be considered.

Five variables of heatsinks

- **TEMPERATURE of the component's surface.** What is the maximum temperature the component can reach when operating under circuit conditions?
- **TEMPERATURE of the surroundings.** What is the maximum ambient temperature?
- **THERMAL resistance of the junction between the component and the heatsink medium.** How easily does heat flow from one to the other?
- **ABILITY of the heatsink to dissipate heat**
- **ELECTRICAL conductance.** Components are metal, heatsinks are metal, metal conducts electricity — so is there likelihood of short circuits?

If you are designing an electronic circuit from scratch, then you need to know or at least be able to calculate all the above variables to ensure that a heatsink will do its job properly. If you are simply building a circuit from a magazine or book such as this one, then you don't need to do any calculatons — simply use the heatsink that is recommended.

Whether you design a circuit yourself, or follow one, there is a particular method when using a heatsink to make sure the best performance is obtained.

The problem is that the surface of a component, and the surface of a heatsink are not always perfectly flat. Small air pockets always exist which restrict heat flow from the component to the heatsink — known as the *thermal conductivity*.

Thermal conductivity can be improved by coating both surfaces with *thermal compound* (sometimes known as *thermal joint compound*, or *heatsink compound*, or *thermal grease*). This is a silicone oil based grease, impregnated with metal (usually zinc or silver) particles, which fills in air pockets and promotes heat flow. It's usually supplied in tubes or syringes (see Figure 6.26) for easy application.

Using it it quite easy: apply a small amount to your finger, then rub your finger and thumb together to thin the compound out, then apply some to each surface.

Finally, mate the surfaces.

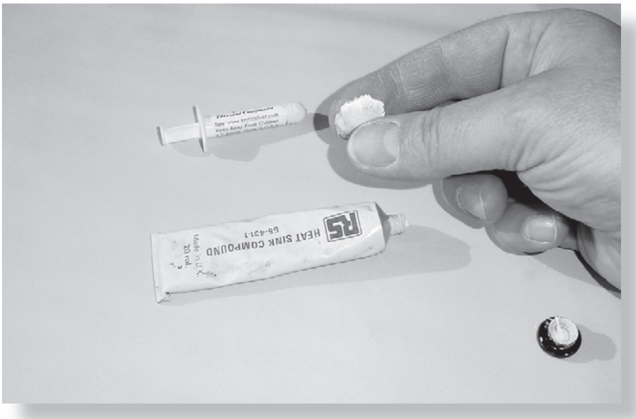

FIGURE 6.26 Thermal compound – supplied in various ways. Put a small amount on your finger and rub between finger and thumb before applying to a component or heatsink surfaces.

Take Note

It is important not to have too thick a layer of thermal compound between component and heatsink! Too much compound restricts heat flow just as much as too little compound.

FYI

Even mounting components on a metal housing's panel may not provide sufficient cooling for some applications. Better cooling methods need then to be adopted — the most common of which is the fan (seen most often in large computers) which helps to ensure that air flows round hot components. Other heatsinking methods exist, including heat exchange, refrigeration and water-cooled systems, but these are generally beyond the scope of the beginner to electronics construction.

To make sure no electrical connection occurs between a component or its heatsink, use an insulating washer and plastic mounting tab or bush. Traditionally, insulating washers are commonly made of the insulator mica, but increasingly popular are silcon rubber washers. Both are shown in Figure 6.27 along with an insulating mounting tab. Note that the size and shape of the insulating washer depends totally on the size and shape of component it is for. The washers shown in Figure 6.27 are meant to fit a TO220-type transistor shape component. Insulating washers — obviously — fit between the component and the heatsink, so introduce another two surfaces into the equation, so when applying thermal compound prior to fitting a heatsink, put thermal compound on all four surfaces.

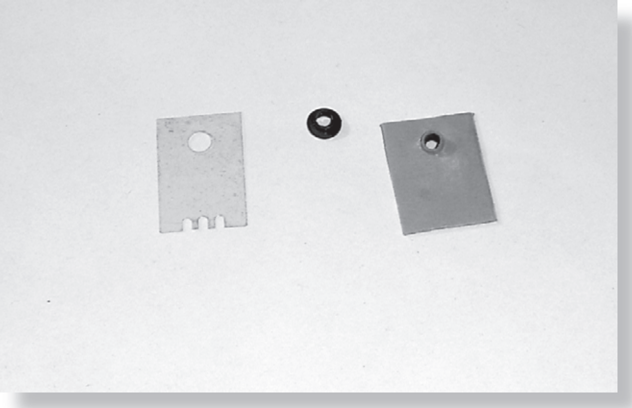

FIGURE 6.27 Insulating washers: left is a mica washer, right is a silcon rubber washer. In the middle is an insulating mounting tab.

The component needs to be fastened mechanically to the heatsink in some way. When fastening a component to a housing panel the usual method is with a small nut and bolt. The plastic insulating mounting tab ensures that the nut and bolt used do not cause a short circuit between the component and the housing panel. The way the arrangement should be fitted together is shown in Figure 6.28.

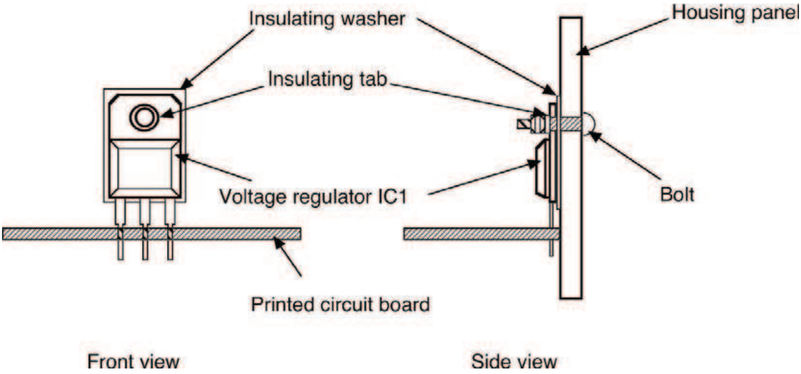
Insulating washer
Housing panel
Insulating tab
Voltage regulator IC1
Bolt
Printed circuit board
Front view
Side view

FIGURE 6.28 Bolting a component – in this case a voltage regulator integrated circuit – complete with insulating washer and mounting bush, to a heatsink.

Figure 6.29 shows an actual application of heatsinking components, in the Bench Power Supply project of Chapter 8. Here you can see the three voltage regulators that are heatsinked to the rear panel of the project housing.

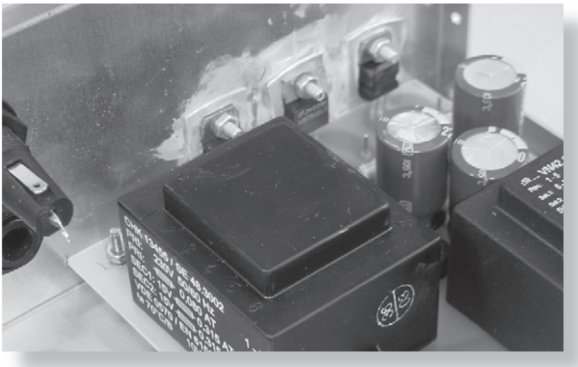

FIGURE 6.29 Showing three components, heatsinked to a metal housing rear panel.

Take Note

Electrical conductance does not need usually to be taken into consideration where a heatsink is a simple push-, clip-, or bolt-on type, that is fastened directly to an individual component on a printed circuit board. After all, as long as the heatsink touches nothing else metallic, there can be no danger of short circuits.

However, electrical conductance for a component mounted onto a metal housing's back panel for heatsinking purposes is of prime importance – several components of circuit parts (the printed circuit board, connectors, controls, power supply parts and so on) may also be mounted on the housing. Here, problems of possible short circuits need to be taken seriously. Some form of insulating washer should be used to prevent electrical connection between the component and the housing.

Wiring of all parts

In some instances, once a printed circuit board is mounted in its housing, the electronic device is complete, and it only requires the housing lid to be fastened on for final completion. In most instances though, there will remain some wiring of internal parts together.

If you are building an electronic device from a magazine or book design any internal wiring will usually be shown as a wiring diagram, although this is normally topographic — ie, like a railway map, it shows theoretical connections but not the actual routes. More often than not, how the wires themselves are connected between parts is left to you, the constructor to determine. With this in mind, there are some basic guidelines, but the rest is up to you.

Generally, use multi-stranded insulated wire for connections.

Multi-stranded wire types and uses

- **FOR low voltage, low power connections a suitable choice is known as 7/0.2 equipment wire (which means it has seven strands of wire, each one being 0.2 mm in diameter).** This is flexible and easy to route inside housings.

CONTINUED…

...CONTINUED

- **AN alternative to individual wires is ribbon cable (ie, several strands of insulated 7/0.2 wires, moulded together side-by-side).** Ribbon cable has an advantage over single equipment wire in that where multiple connection are made between internal parts, connections are grouped. Ribbon cable is available with each strand a different colour, which makes checking connections an easy job, too. Ribbon cable is available in 10-, 20-, and 30-way strips, but strips can be easily separated just by pulling, into the number of connected strands you want for each connection

- **MAINS or higher power connections must be made with thicker multi-stranded wire (eg, 16/0.2, 24/0.2, or 32/0.2).** Often the magazine or book will specify the exact type to use

As construction of the circuit and wiring goes on there may be instances where several lengths of wires follow the same route. As a matter of neatness, and to prevent wire movement (which could cause soldered joints to come apart) it's a good idea to tie the lengths together. Plastic cable ties are ideal for this job. An alternative is the cable clip, which comes with a double-sided adhesive pad to stick the clip onto a housing panel. Use ties or clips at least every 100 mm or so along panels, and at corners where the grouped wires go round.

▶ FYI ◀

In the 'good old days' tying of cables was done with special continuous cable tie cord, and was more a work of art than a technique. Fortunately, 'good old plastic' cable ties do an equally good job these days.

Where wires are grouped and tied or clipped together into cables, it can become tricky working out which wire is which. If you have a selection of coloured wires this can help, as can using coloured ribbon cable. But where you have more than just a few of these, you still might find it awkward. If this is the case, use cable markers. These are small plastic rings, placed over each end of wires before soldering. They can be coloured, and also numbered.

Mains wiring

The wiring of electronic circuits which use mains power bears particular mention. Mains power is, of course, dangerous — potentially lethal — and precautions must be taken when you intend to build a circuit which uses it.

General guidelines for mains wiring

- **IN general, use a metal housing for any electronic circuit you construct that has a mains power supply.** Use three-core mains cable (ie, with live, neutral, and earth wires), and make sure the cable enters the housing using a good quality cable entry clamp, properly tightened so that connections are not strained. A selection of cable entry clamps is shown in Figure 6.30. You must ensure the clamp is the correct size for the cable used.

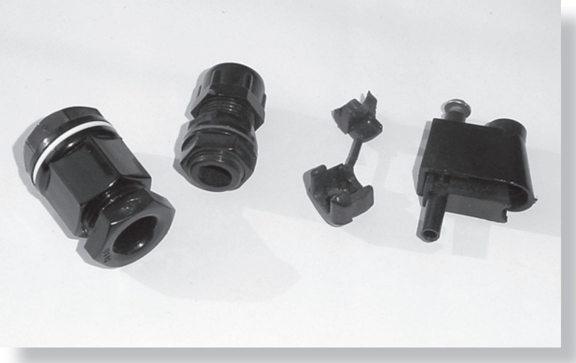

FIGURE 6.30 Cable entry clamps for use when mains cable enters a housing.

- **AN alternative to a cable entry clamp is a good quality plug and socket connector — such as that shown in Figure 6.21.**

- **USE insulating sleeving over all soldered connections, to ensure that no mains electrical parts are exposed.** Push the sleeving onto a wire prior to soldering it to a connection. After soldering push the sleeving back along the wire, and over the soldered joint, as in Figure 6.31.

CONTINUED…

...CONTINUED

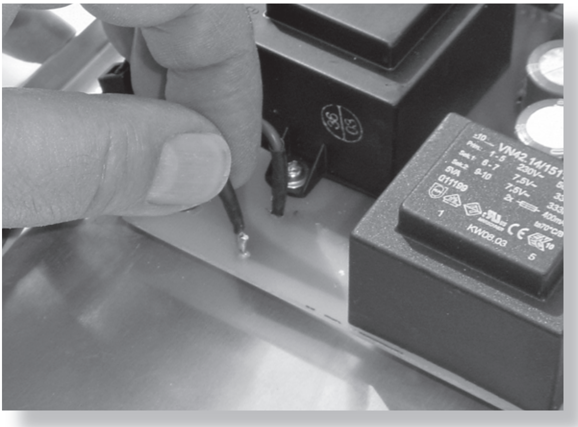

FIGURE 6.31 Using insulating sleeving over a mains circuit soldered joint. Here the joint has been made, and the sleeving is being pushed into position to cover the joint.

- **ALL soldered connections must be made good mechanically, prior to soldering.** Thus, if the soldered conection comes loose, the wire is still held in place.

- **MAINS live and neutral wires should go straight to an on/off switch.** Use a double-pole on-off switch for this purpose, and connect the live connection to one pole, and the neutral connection to the other. Such a switch ensures that the following circuitry is completely disconnected from mains when the switch is switched off.

- **TO protect the electronics circuit (and you, for that matter) from damage in case of an internal fault, use a panel-mounted fuse holder, with a fuse rating no greater than that specified in the magazine or book the electronic circuit comes from.** Mains live is taken directly from the other contact of the live pole of the on/off switch to the central connection of the fuse holder (thus, if the fuse is removed when the device is connected to mains, no contact can be made to a live part).

CONTINUED...

...CONTINUED

- **THE earth wire in the mains cable must go directly to a bolted connection on the housing chassis.** Use a solder tag on the earth wire, and shakeproof washers on the bolt, to ensure the bolted connection is reliable over the long term.

- **CHECK, double-check, and triple-check all connections and the circuit prior to connecting to mains power.**

Figure 6.32 shows a general wiring diagram for mains powered electronic circuits which follows these guidelines.

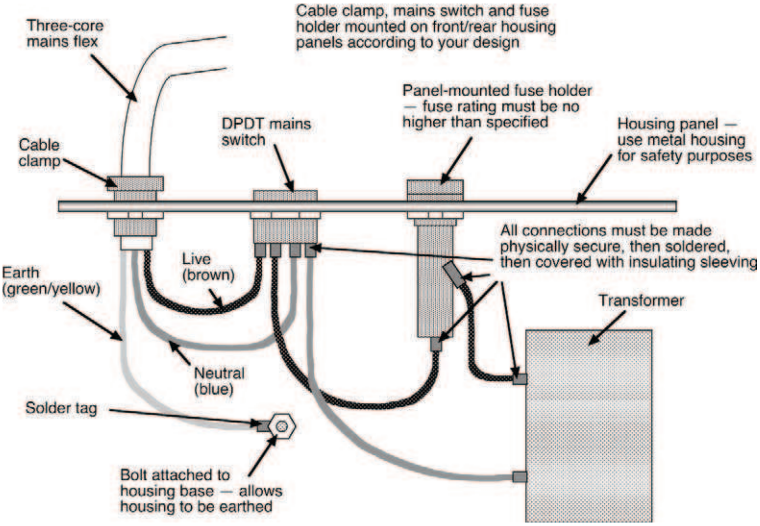
Three-core mains flex
Cable clamp, mains switch and fuse holder mounted on front/rear housing panels according to your design
Panel-mounted fuse holder — fuse rating must be no higher than specified
DPDT mains switch
Cable clamp
Housing panel — use metal housing for safety purposes
All connections must be made physically secure, then soldered, then covered with insulating sleeving
Live (brown)
Earth (green/yellow)
Transformer
Neutral (blue)
Solder tag
Bolt attached to housing base — allows housing to be earthed

FIGURE 6.32 A general wiring diagram for mains powered circuits.

Take Note

If a single-pole switch is used, even if it is in the live circuit on construction, at some stage somebody may rewire the mains plug incorrectly, at which point the switch is then in the neutral circuit and the remaining circuit will be live even if the switch is turned off! A double-pole switch prevents this situation.

Front and rear panel labelling

When an electronic circuit is mounted in a housing, and all connectors and controls are fitted, then the device is ready for use. Well, not quite!

The main interface between you and the electronic circuit you've just made is the housing's front panel and all its controls and connections. With maybe just a couple of these, you probably remember what each is, or does. But if there are more than just a couple, it becomes increasingly difficult to remember. And if anyone else needs to use the device, how will they even know at all?

The answer is that your front panel (and possibly the rear panel if there are controls and connections round the back) should be labelled.

Labelling is all part of the process which is often called *finishing off*. The point is though, that far from being a process you undertake at the finish of building a circuit, you need to be considering the front and rear panel and its layout right at the very earliest design stages.

For a start, labelling can't possibly be the last thing you do — to put labels on, all controls and connections need to be removed so that you can work on a clean, flat panel. So, when you think about it, Labelling is not actually a finishing off process, it has to be done quite a bit earlier on in the whole build.

Panel labelling

Once you know the layout of the panels, and have drilled holes for all control and connectors you can start labelling. There are several methods for this. The simplest would be merely writing the labels on with a permanent ink marker pen, but of course that would not make a particularly good-looking end result.

A much better method however is to use rub-down transfer letters. These are available in hundreds of typefaces (more commonly called fonts and sizes), and it is quite simple to create a good effect with a bit of practice.

Transfers such as these are available in sheets, and any good stationers will stock them or order them for you. Ask to see their catalogues to see the range of typefaces available.

Key points when labelling with rub-down transfers:

- **USE guidelines to ensure labels are level.** Rule faint lines in pencil, or stick lengths of tape along the length of the panel
- **TAKE care when centring labels around controls or connections — if they go too far to one side or the other they will look odd**
- **GROUP controls and connectors where you can.** Perhaps ring them using some rub-down transfer lines, so that they appear cohesive in one part of the panel
- **USE a typeface of only a few millimetres high for most labels.** This equates to a publishing measurement (in points), of around 12 to 16 points
- **USE a plain typeface for most labels.** In publishing-speak, this could be a sans serif font
- **USE a more ornate typeface if you want for the main descriptive label — eg, Audio Amplifier.**

Using a computer

If you have access to a computer, you can use it to aid when labelling a housing. First you can design the panel layout on screen, simply to give yourself a good idea of what looks good — even if you opt to use rub-down transfer lettering to actually do the labelling in the end. All you need is a fairly basic illustration application of some description on your computer. There are several such applications available, and even a fairly basic image manipulation application (such as you probably have bundled with your printer, for working with photographs, say) can often produce good results. All it needs to be able to do is to use the range of fonts and sizes your computer already has installed, and perhaps draw a few lines. Of course, a proper illustration application is best.

Designing a good-looking label is not rocket science — the key points above can help. The advantage of designing the panel on a computer screen, though, is that you can make adjustments until you get it just right. Figure 6.33 shows the sort of label you can hope to create using a computer. Note the outside marks, which act as registration marks to make sure the label will fit the panel exactly.

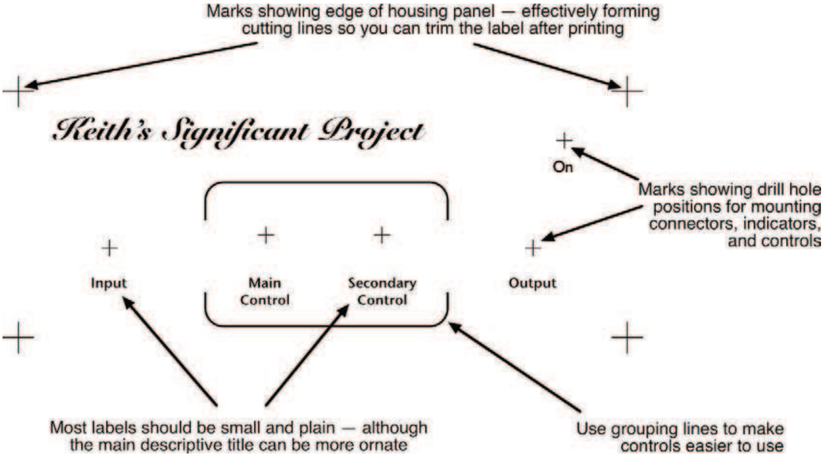
Marks showing edge of housing panel — effectively forming cutting lines so you can trim the label after printing
Keith's Significant Project
On
Marks showing drill hole positions for mounting connectors, indicators, and controls
Input
Main Control
Secondary Control
Output
Most labels should be small and plain — although the main descriptive title can be more ornate
Use grouping lines to make controls easier to use

FIGURE 6.33 A computer-generated front panel.

Better than this though, after you have designed the panel label on screen, you can print it out using a printer, then attach the whole label to the housing panel in one step. You don't even need a particularly expensive printer to do this — most modern inkjet printers produced over the last few years would work. As you are printing the label, you are not restricted to just monochrome, and judicial use of colours can be used to make the housing more attractive.

Simplest way here is to print the design onto photographic paper (glossy or matt, whichever you prefer). The paper can be trimmed to size, then stuck to the housing panel using adhesive (a spray adhesive is ideal for this task).

Alternatively, if your printer is capable, you can print the design onto acetate transparency. Print the design in mirror-image format on the back of the acetate sheet, then trim the acetate to fit. Stick the acetate onto paper (white or coloured, for the desired effect), then stick the whole thing onto the housing panel.

One of the best methods of labelling using a computer is with the Quick-Laser system, shown in Figure 6.34. In this process, special coloured polyester sheets with optional, die-cut and peel-off sections (similar in principle to address labels for envelopes) are used together with a laser printer to produce very high-quality self-adhesive labels.

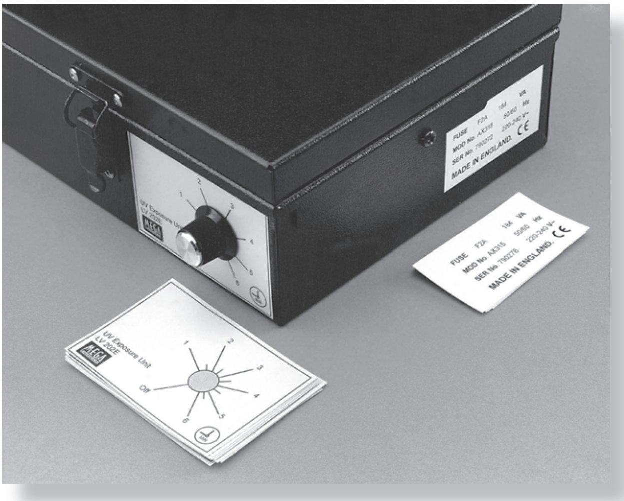
FUSE F2A 184 VA
MOD No. AX315 50/60 Hz
SER No. 790272 220-240 V~
MADE IN ENGLAND.
UV Exposure Unit
LV 202E
MEGA
1
2
3
4
5
6
FUSE F2A 184 VA
MOD No. AX315 50/60 Hz
SER No. 790276 220-240 V~
MADE IN ENGLAND.
UV Exposure Unit
LV 202E
MEGA
Off
1
2
3
4
5
6

FIGURE 6.34 The Quick-Laser system uses coloured polyester sheets to produce labels from a computer and laser printer (Mega Electronics UK).

House!

Finally, after you've labelled the housing, you can refit the controls and connectors, and your electronic device is complete. Congratulate yourself!

7

Fault-finding

At some time or another, anyone and everyone who has more than a passing interest in electronics has built a circuit up, perhaps spending a great deal of time not to mention money in doing so, only to find that it doesn't work. Even the professionals encounter this, so what chance have people like you or I got in producing a working and reliable circuit?

Well, fortunately, we have a remarkably good chance of doing so, particularly if we follow a fairly logical and straightforward approach to *building* electronic circuits in the first place. Fortunately also, in the vast majority of cases the reasons why circuits don't work is that there is some form of easily identifiable mechanical fault.

So, by a combined strategy of building electronics circuits properly in the first place (so that they *do* work first time usually), and being able to identify and correct the cause or causes of them *not* working in the second place, then we all stand an extremely good chance of being able to produce working circuits.

That is what this chapter, indeed, what this whole *book*, is all about — producing a combined strategy of proper building techniques so that (hopefully) circuits *will* work first time, together with the means of identifying the problems and faults in those circuits that do not work first time.

Hint

Remember: fault-finding is not just about locating what you've done wrong after you think you've finished – it's about locating what you did wrong BEFORE you think you'd finished...

A strategy of building electronic circuits

With all this in mind, it's obvious that to prevent faults in a circuit which you will have to spend time fault-finding later, it's better to take care when building your circuit in the first place so that the faults simply don't occur. Remember the old saying: a stitch in time saves nine.

As such, I'm going to give you, first, a fairly straightforward list of things that you should do when building an electronic circuit which, if you *do* follow will help to largely negate the need for fault-finding afterwards. This list is based on a knowledge of what faults typically do occur in electronic circuits, and what's more, why those faults occur. At the end of the chapter, I'll then give you a method of fault-finding which will see you in good stead for the rest of your electronics constructional life.

A Strategy To Follow

KEEP CLEAN — every part of the electronic circuit. Printed circuit board track, component leads and so on can all be affected by dirt and grease to the extent that soldered joints are not properly made. Worse, dirt or grease in a joint can affect its performance over time — it might work initially but stop working after a short while. Often, just cleaning the parts is enough to ensure a good joint.

CLOSELY INSPECT — the printed circuit board. Your complete electronic device depends totally on the printed circuit board. A minor imperfection (such as a crack in the copper foil track), or a hairline short circuit of copper between tracks which are not intended to be connected will prevent a circuit from working.

DOUBLE-CHECK — component values. Check once when you choose a component that it is the correct value, then check it again after it's been inserted in the board immediately prior to soldering.

DOUBLE-CHECK — polarized components. Electrolytic capacitors, semiconductors and so on must be inserted into a circuit the right way round (this is called *correct polarization*). Check their polarity is correct once when you insert the component in the board, then once more immediately prior to soldering.

TRIPLE-CHECK — component positioning. As you select a component for insertion into a printed circuit board, check its value on the circuit diagram. Next, check that the position on the printed circuit board is correct according to the layout diagram. Finally, after insertion, recheck both value and position again immediately prior to soldering.

DO NOT — insert more than a handful of components in at a time. If you insert too many components in to a printed circuit board before soldering them in, they may slip partially out of the circuit board. This may create short circuits that affect the circuit's operation. Only insert a few components before soldering them in, then moving on to the next batch.

WHEN SOLDERING — make sure each and every joint is properly made. Check each joint visually immediately after making the joint. Do not make hundreds of joints one after another without checking. Read Chapter 5 thoroughly, and practise making soldered joints until you can make a good joint *every* time.

AFTER SOLDERING — but before testing. Check for shorts between joints, and check for dry joints.

BUILD STAGES — and test as you go along. Divide the complete circuit into complete functional units (most — if not all — circuits can be thought of this way) and build each stage up in turn. As you finish one stage you can test it, before moving on to build up and test the next stage. In this way, if you *do* have a fault, you have isolated it in the production stage, and so can repair it more easily than if the complete circuit is built.

POWER ON — as a last resort only. Inevitably, when you have finished your printed circuit board you want to connect power and see if it works. The problem is — if you do and there is a fault such as a short circuit, or if polarized components have been inserted the wrong way round, or if components of the wrong value have been inserted, or for any number of other reasons — you could easily cause irreperable damage to components. Take time out for a few minutes, then inspect the printed circuit board carefully — use a magnifying glass if you have one, to see if you can find any possible faults. Remember: once you've connected the printed circuit board to its power supply it might be too late — a simple fault becomes a big-time problem…

Types of faults

If I was to be pinned down on this issue, I would have to say that in 99.9999999999999999 (OK, that's enough!) per cent of cases, the reason (or reasons) why a circuit does not work is due to a mechanical fault introduced by the person making the circuit. Very rarely — and I do mean *very* rarely — the reason why a circuit does not work properly first time is due to a faulty component. This is, quite simply, because electronic components are made to a pretty high standard, and are quality controlled so that you may rarely expect one to not work. Even more rare are those faults which occur due to a circuit (an incorrectly specified component, say, or an error in circuit design).

As a result, when a circuit you have built does not work, the first and most important thing to do is assume *you* are at fault — not the components. The second thing to do is look closely at what you have done — to find out where you have gone wrong.

With this in mind, you can now try to isolate the reason why the electronic circuit you've built doesn't work. Check things in the following order:

Finding faults? Walk this way:

- **SOLDERED joint faults**
- **PCB faults**
- **WIRING faults**

Once you've thoroughly checked, and are happy that your soldering, your pcb and your wiring is not at fault, then — and only then — you should begin to consider the remaining potential causes of faults:

- **CONNECTORS**
- **COMPONENTS.**

Soldered joint faults

Basically, there are two types of faults that can occur when soldering. The first is just a bad, or dry joint. The second is a solder bridge. Figure 7.1 compares a good joint with a bad joint and a solder bridge. Good joints tend to be concave and shiny in appearance. Bad joints tend to be blobby (to the point of being ball-shaped) and may be dull in appearance. Solder bridges form between copper tracks on the printed circuit board. In many cases of bad or bridged joints you can isolate each merely with a close visual inspection. It helps, though, if you have a magnifying glass and good lighting.

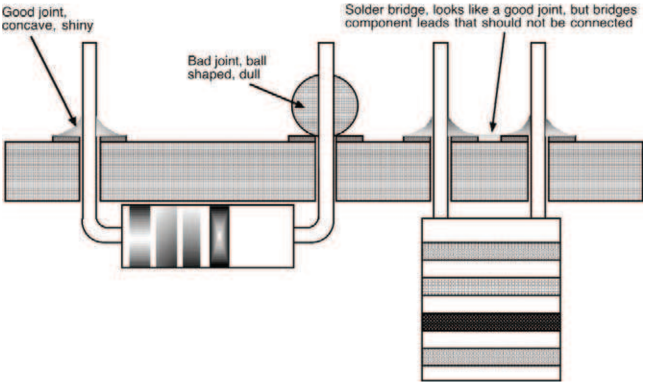
Good joint, concave, shiny
Bad joint, ball shaped, dull
Solder bridge, looks like a good joint, but bridges component leads that should not be connected

FIGURE 7.1 A good joint, a bad joint, and a solder bridge.

If you spot bad joints, they will need to be resoldered. Note that one of the main causes of a bad joint is that the joint was not hot enough before solder was applied, so when resoldering a bad joint, make sure it gets sufficiently hot first.

If you spot bridged joints, you need to reheat the joint and remove the excess solder. Use either desoldering braid or a desoldering pump for this task, but remember that once the solder has been removed you will need to reapply fresh solder to make the joint again.

PCB faults

Particularly if you make your own printed circuit boards — although even if you buy them ready-made — you should be aware that they can have faults. So, no matter how good your soldering or circuit building techniques are, they will never be able to produce a working electronic circuit. A good procedure would therefore be to check the printed circuit board, before you even start to solder components into it.

Look for three things:

- **1 Bridges between adjacent copper tracks.** Sometimes, if the etching process hasn't fully completed before the printed circuit board is removed from the etchant, or if small areas of etch-resist have not been developed fully, small bridges of copper remain between tracks. Sometimes these bridges can be extremely thin, but even a hairline bridge forms a short circuit. Thoroughly check all copper tracks of the printed circuit board — a magnifying glass and good lighting helps. If you spot what you think is a bridge, you can confirm it using a multimeter set to resistance measurement. Bear in mind that if you find one copper bridge, there is a good chance there will be more at other places on the printed circuit board

- **2 Broken tracks.** In the etching process, a small (even minute) break in the resist allows the etchant to remove the copper underneath. If the break in the resist cuts across a track, there will be a complete break in that track. The circuit built on that printed circuit board will not work correctly. Again, there is nothing for it but to first check the copper track visually. If you spot a broken track, clean the area to either side of the break with an abrasive scrubbing block or abrasive fibreglass pencil, then solder a joint which bridges the two sections of copper track

- **3 Incorrect or missing track.** The copper track on your printed circuit board is intended to represent and make all connections between components in the circuit diagram. If you have designed your own printed circuit board from a circuit diagram then made it yourself, there is the potential that somewhere in the copper foil track layout design process you have made an error. If the completed electronic circuit does not work, then as part of the fault-finding process you need to double-check your printed circuit board's copper foil track layout. Make sure that all connections on the circuit diagram have a corresponding connection on the copper foil track — if any connections are missing, then the circuit parts are effectively open circuit. If you find any connections missing, make them using a small length of 7/0.2 wire. Strip a very short amount of insulation from each end of the wire, then tin each end before soldering it into place to make the missing connection. Also make sure there are no *extra* connections (possibly forming what amounts to short circuits between circuit parts) that should not be in the copper foil track. If you find a connection present that should not be, cut through the copper track using a craft knife or similar to break it.

Wiring faults

When constructing any electronic device, mounting the printed circuit board in a housing, then connecting the printed circuit board to controls and connectors, wiring is always an issue and a potential area for faults to arise. If you suspect wiring to be the problem (or, at least, you have gone through the processes above and the circuit *still* does not work), check the wiring against the wiring diagram. Isolate each connection in turn between the printed circuit board and housing-mounted control and connectors, and make sure each end of it connects to its correct places. Look out for bad soldered joints at each end. Note that some connection terminals on connectors are relatively large metal terminals (at least in comparison with copper foil track connections), and soldering a connection to them requires significantly more heat than it does to printed circuit board joints. As a result, it’s easy to make a bad joint to such terminals. Resolder any joints that look in any way suspect.

Connectors

After soldering errors, printed circuit board faults, and wiring errors, the next most important cause of faults in electronic devices is that due to connector faults. Most of these occur after time and use, but you still must consider connectors as a potential cause for faults if you have worked through all the processes covered already in this chapter.

It is often easy to check that a connector is doing its job by using a multimeter set to resistance measurement mode. If resistance is infinite between one end of the connector and the other, the connection is open circuit. If two or more terminals on one side of the connector have zero resistance between them, the connections are short circuited.

Components

Finally, once you have gone through all the processes above to check soldered joints, printed circuit board copper foil tracks, wiring, and connectors, you can begin to suspect that one or more of the components is faulty.

Bear in mind though, that it may even be that one of the faults you have already found as you work through the processes in this chapter has actually caused one or more of the components on the printed circuit board to be damaged. A short circuit due to a solder bridge or copper track bridge, say, might cause sufficient current to pass through a component not designed to cope with the current to be damaged. So, even though you might have found one fault (or more) in your soldering or the printed circuit board, or the wiring, others may have been created by this.

Down with fault-finding

It's not the intention of this book to cover all fault-finding processes when all the processes have been followed and we are finally down to component level. There are, though, many other books on this topic. Just remember that fault-finding at component level must be the very last resort.

Instead, my intention in this book is to help the beginner to the construction of electronic circuits and devices see that fault-finding is not a process distinct and separate from the construction processes. Fault-finding should be, in fact, completely integral with the constructional processes.

Indeed, it should be *so* integral that checking for faults and correcting them as they occur should be natural to the constructor. So integral that you shouldn't even *think* of fault-finding as a distinct and separate process at all.

With this in mind, shown opposite is a method of fault-finding which comprises a general list of things to do and bear in mind as you construct any electronic circuit or device. As you become more experienced in electronics construction, the list should begin to become second nature to you.

Picture this!

So far, we've only looked at the theory behind fault-finding, trying to isolate the areas where faults can arise, what sorts of faults there may be, and what causes them. For the rest of this chapter we can take a look at what these faults actually look like.

Fortunately, several mechanical printed circuit board faults can be fairly readily identified with a close visual inspection. However, for the beginner to electronics construction (and, indeed, for even the not-such-a-beginnner!) it's not always *easy* to identify them, simply because finding something when you don't know what it looks like is like looking for a needle in a haystack when you're blindfolded.

So, for the remainder of this chapter, some photographs of the most common printed circuit board soldering faults are given, in Figures 7.2 to 7.14.

A total method of fault-finding:

- **KNOW the circuit, and how it works.** Break it up into logical *black box* sections. Often, the magazine or book giving the circuit will do this for you, in a separate 'How it Works' section, but if not, try to do it yourself. The more you understand how a circuit works, the easier it becomes to construct and get working!

- **CHECK the plain printed circuit board.** Check it for any short circuits. Check it for any broken tracks. Check that it follows the circuit, with copper foil track for every connection within the circuit, and with no extra connections

- **BUILD one section of the circuit at a time.** Insert and solder only those components within that section, then check the section works fully before you start the next section. If one section does not work, find and correct the fault before you start the next section. This way, any faults should lie only in the last section you built

- **ONCE the printed circuit board is complete, with all components 'stuffed', check that it works.** If not, as long as you have been following this method, the fault should again lie in the last section you built

- **HOUSE the printed circuit board, together with controls and connectors**

- **CHECK the complete and housed device.**

Blistering or lifting — where copper track lifts from the board surface, usually due to excessive soldering heat.

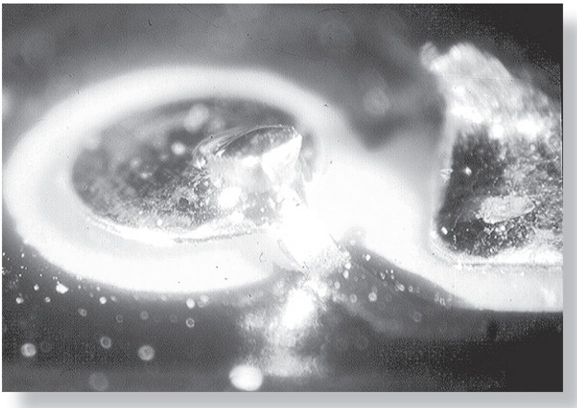

FIGURE 7.2 Lifted pads on a printed circuit board.

Bridging or shorts — where solder joins two or more conductive parts which are not meant to be connected.

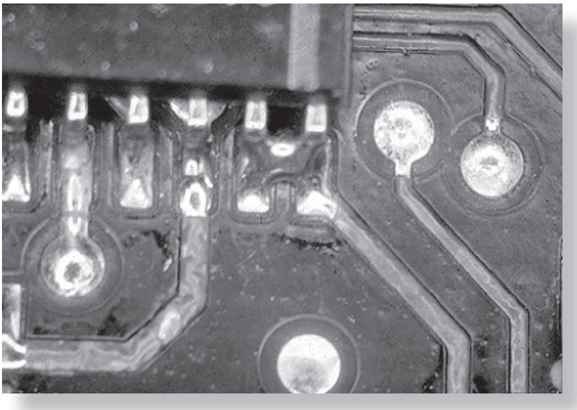

FIGURE 7.3 Solder bridging between connection pins of an integrated circuit.

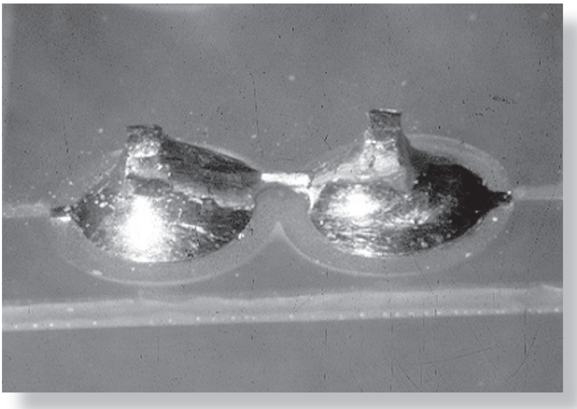

FIGURE 7.4 A minute solder bridge between two component pads.

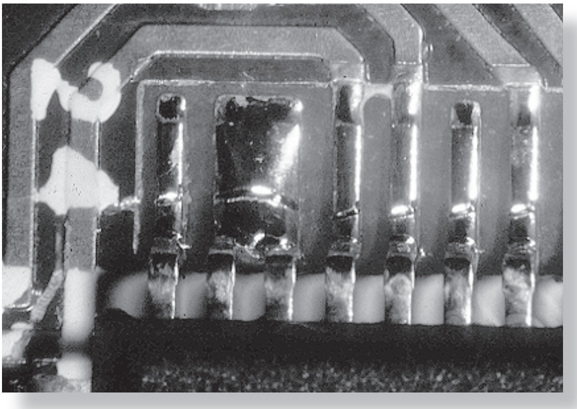

FIGURE 7.5 Solder bridging between adjacent printed circuit board tracks.

Components lifting — where components become displaced between stuffing and soldering. Faults like these don't necessarily cause the circuit to malfunction by themselves, but components that do lift may short to adjacent components.

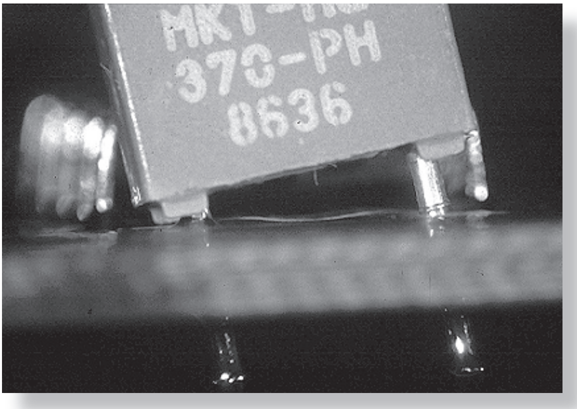

FIGURE 7.6 A capacitor that has lifted.

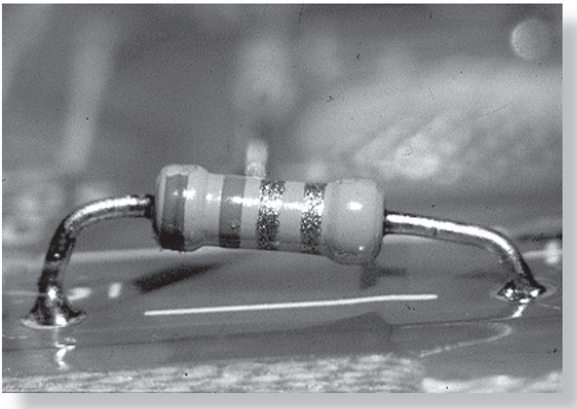

FIGURE 7.7 A lifted resistor.

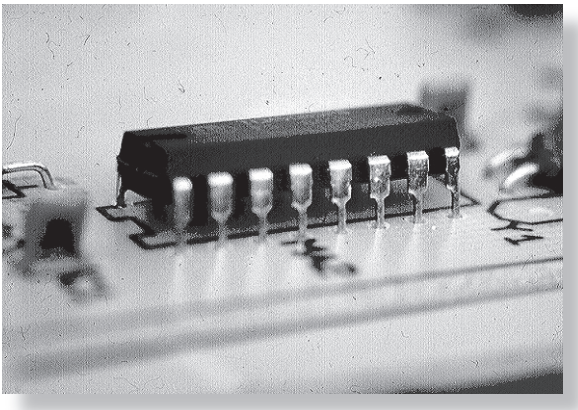

FIGURE 7.8 An integrated circuit that has lifted.

Cracked joint — if a printed circuit board is not adequately fixed in position in a housing, occasionally soldered joints that come under mechanical stress can crack, thereby creating a broken connection. The hairline cracks so caused are very easy to miss under normal visual inspection — a magnifying glass is usually required to see them.

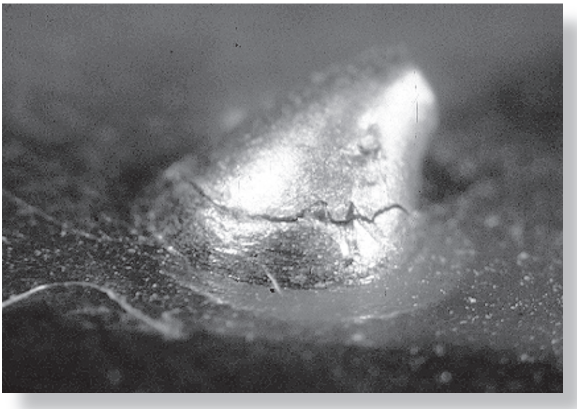

FIGURE 7.9 A cracked joint.

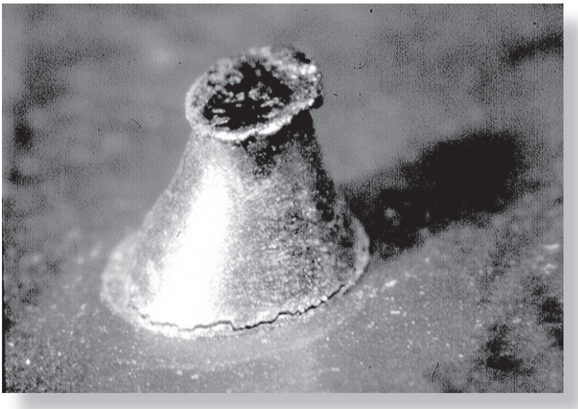

Figure 7.10 Another cracked joint.

Dewetting — where an initial bond is formed when soldering, followed by a withdrawal of solder from the joint, leaving irregular shaped mounds of solder separated by areas covered only with a thin solder film.

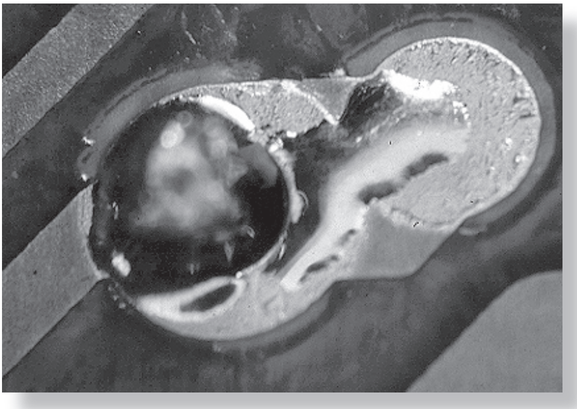

Figure 7.11 Dewetting between two connected component lands.

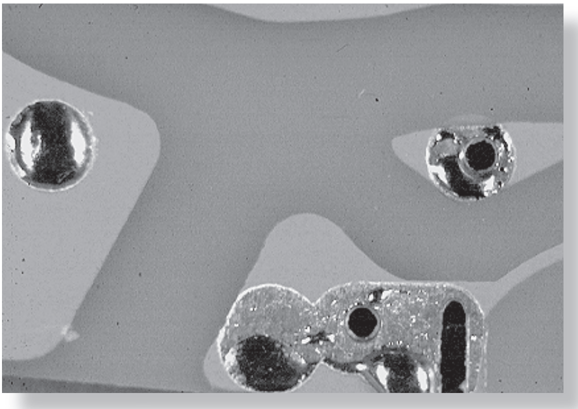

Figure 7.12 Dewetting of component lands.

Excessive solder — too much solder on a joint — this may also cause bridging with local lands.

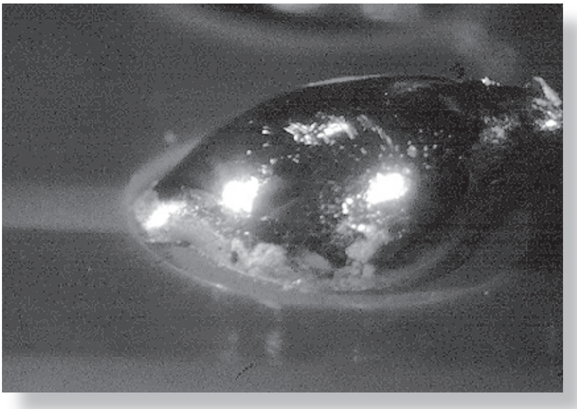

Figure 7.13 Excessive solder looks like a blob on the printed circuit board.

Solder voids — too little solder on a joint creates a potential weakness.

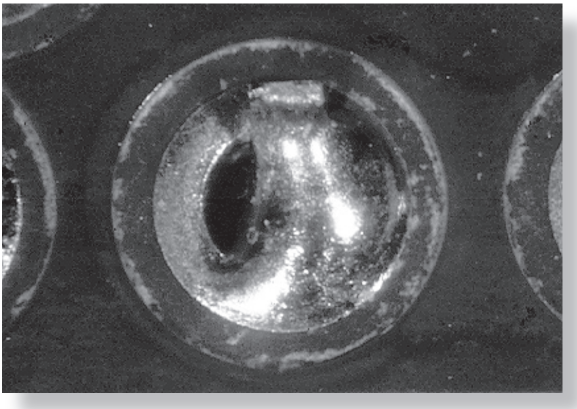

FIGURE 7.14 A solder void on a component land.

8

Projects

While this book so far has been quite theoretical in parts, I hope I haven’t lost you all to sleep…

To wake you up again if you did succumb to the effects of boredom while I’ve waffled on, I’ve included some projects in this chapter that you can choose to construct if you want.

Even if you choose *not* to construct them, the projects are still highly useful, in that they illustrate many of the processes that I have covered throughout the book. They are graded to a limited extent — the first ones being easiest to build and the later ones being harder. That’s not to say that if you want to construct one of the later circuits first you shouldn’t be put off by its greater practical requirements — with the exception of the last project — the mains-powered Bench Power Supply. The principles, after all, of contructing any electronics circuit are the same whatever the complexity, and are covered in detail throughout this book. So, if you decide to build any of the projects, hopefully all the information you need to do so is somewhere in these pages for you to read.

Take Note

The last project – a mains-powered power supply – should not be attempted unless you are confident of your skills. Mains voltages are dangerous – potentially lethal – and extreme care must be exercised when constructing the project to prevent a situation arising where a user comes into contact with those voltages.

In short, if you are not confident of your skills, and do not take requisite care – do not construct this project!

The projects

I thought long and hard about these projects, trying to weigh up complexity, ease-of-construction, and usefulness, aiming to get a balance between all three. What I decided early on was to promote what I feel is the *best* way of constructing projects, not necessarily the easiest. For instance, I have eschewed the use of stripboard construction, instead giving only printed circuit board descriptions and layouts. My reason in doing that is because stripboard is, in essence, merely a prototyping method of construction, not the sort of method I want to promote to my readers in the final projects in my book.

My feeling is though, that in a modern electronic circuit, the printed circuit board is as much a component of the final device as any other part. It is not merely a method of making connections between parts of a device — it is a part of the device in its own right! Indeed, in many circuits, how the printed circuit board is made, how it is designed, where all the components are physically placed in relation to each other, and in fact what it is made of, can all define how the circuit works, or even *whether* the circuit works.

Likewise, though not so important to the working of an electronic circuit, the housing used for a circuit is worth bearing in mind. A good housing can help to make a good electronic device; a poor housing inevitably makes a poor device. You can't make a silk purse from a sow's ear, and if your project is housed in a sow's ear of a housing, then no matter how well the circuit works, no matter how much hard graft you've put into it, your project simply can't be a silk purse.

So, on to the projects themselves. There are four, listed with brief details and grading for complexity below.

The four projects:

1. **Car-to-USB Power Supply.** Allows certain devices (digital audio players, mobile phones and so on) to be powered and recharged in your car. *Easy*.
2. **Heart Flasher.** Fun device to show your love. *Moderate*.
3. **Guitar Headphone Amplifier.** Don't wake the neighbours while fingering your frets. *Moderate*.
4. **Bench Power Supply.** Useful test-bench equipment you can use to power other circuits you build. *Hard*.

Note that even though I class them as being between easy and hard to build, none of them is beyond the scope of most competent DIY-ers and, with the exception of the Bench Power Pupply (which is mains-powered), you shouldn't worry too much about things going wrong, as the only damage you're likely to do even if you bodge up totally is to ruin a couple of cheap electronic components.

Furthermore, if you read this book first, and follow constructional details given here, things shouldn't go wrong at all.

Car-to-USB Power Supply

These days, rechargeable battery-powered consumer devices are becoming increasingly more complex. As many of them are now often linked with computers, they can often be recharged via the computer connection. *Digital audio players* (sometimes called *mp3 players* — a misnomer, as most play music files in other formats than just mp3) are the commonest such device, although some mobile phones and other devices seem to be going the same way.

A computer connection most commonly relies on a *universal serial bus* (*USB*) connection, from which the device can obtain its recharging power, as the USB interface allows a certain level of power (5 V, at up to 500 mA) through the connector. Away from a computer, however, users have to rely on another source or recharging power if the battery runs low on their devices. Typically, this is in the form of a mains-powered recharger. This is fine if you are within reach of a mains power socket. Not so fine though if you are further afield than that — in the car, say. A useful project if you have a USB-rechargeable device is therefore the Car-to-USB Power Supply — shown in Figure 8.1, and detailed first.

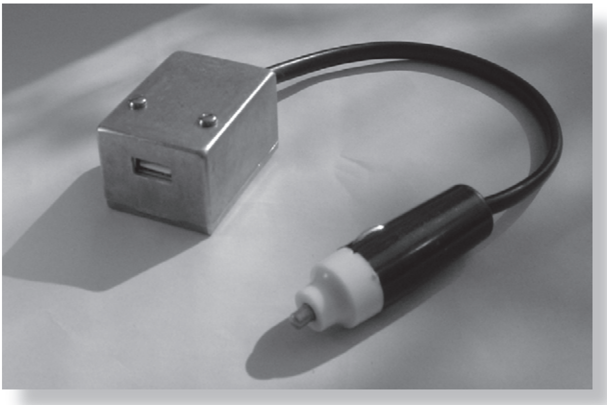

FIGURE 8.1 The Car-to-USB Power Supply.

Principle

The Car-to-USB Power Supply project here uses a car's internal low voltage electrical circuit to allow a USB-rechargeable device to be recharged. It does this by connecting to the car's electrical circuit using a cigar lighter plug that plugs into the car's cigar lighter socket. The low voltage (between about 11 V and 14 V DC) is regulated within the Car-to-USB Power Supply down to exactly 5 V DC, and supplied to a standard USB A-type socket (the same sort that's on any modern personal computer).

The principle of the Car-to-USB Power Supply is shown in Figure 8.2, where these three main stages are identified.

Take Note

The output from the Car-to-USB Power Supply project is not suitable for direct connection to rechargeable batteries, so you shouldn't use it to connect straight to, say, a battery pack. Such packs require some form of control circuitry to ensure only the correct recharge voltage and current is applied. The output of the Car-to-USB Power Supply is a basic 5 V DC regulated supply at up to 1 A, and control circuitry (always present within a rechargeable battery-powered consumer device) is still required.

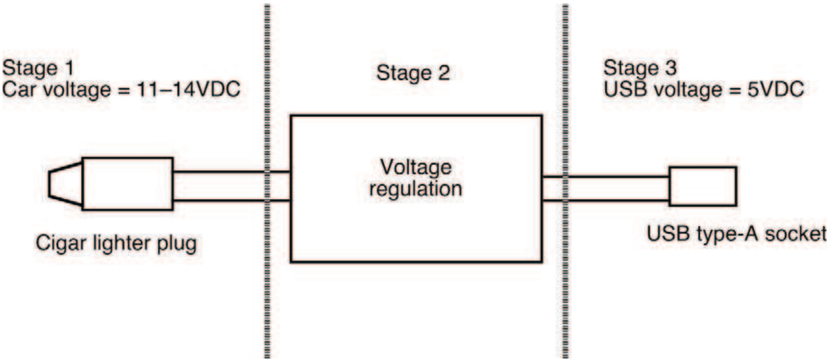

FIGURE 8.2 The Car-to-USB Power Supply principle.

Circuit

The Car-to-USB Power Supply circuit is shown in Figure 8.3. It's not the world's most complicated circuit as you'll appreciate, with the grand total of only five components. It does however, introduce an important principle — that of the integrated circuit voltage regulator (IC1). Voltage regulation is a process whereby an unregulated or varying DC voltage is changed to a regulated or fixed one. In this project a car's main voltage is around 12 V DC, but can vary between about 11 V and 14 V, depending on the state of the car battery and whether the engine is running, so is effectively unregulated. The USB specification, on the other hand, calls for a regulated 5 V DC supply. So what is required is a voltage regulator that takes the unregulated input of 11 to 14 V DC, and gives a regulated output of the required 5 V DC. Integrated circuit IC1 does just that.

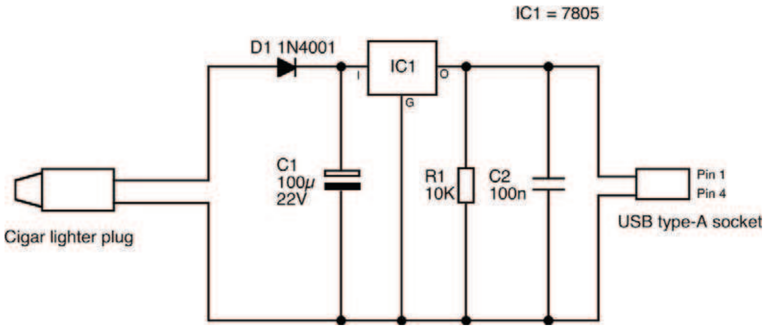

FIGURE 8.3 **The Car-to-USB Power Supply circuit.**

The remainder of the circuit merely aids this. Diode D1 protects the remainder of the circuit against reverse polarity connection (say, if you wired the project incorrectly so the input voltage was reversed). Capacitor C1 acts as a current reservoir, ensuring that sharp voltage peaks and troughs don't pass through the circuit to the output. Resistor R1 and capacitor C2 aid this.

Construction

Voltage regulator IC1 has a maximum voltage dropped across it of about 9 V (the maximum 14 V car voltage, minus the 5 V output voltage). At the USB specification of a maximum of 500 mA, the maximum power it dissipates is:

$$P = IV = 500 \text{ mA} \times 9 \text{ V} = 4.5 \text{ watts}$$

As a result, there is a potential heat problem if the voltage regulator is free-standing while operating in worst-case conditions, and it requires some form of heatsinking to aid heat dissipation and keep it cooler. The easiest way to provide heatsinking is — and that's the method used here — to bolt it to a metal housing. The metal housing conducts heat away from the voltage regulator quite efficiently.

FYI

Actually, voltage regulators like IC1 are protected from overheating by their internal circuitry. If such a voltage regulator gets too hot, an internal thermal cutout control turns the output supply off until it cools down again. Heatsinking merely aids in preventing the cutout situation happening.

None of the components are critical. Any of the 7805 variety of 1 A voltage regulators is usable, and neither the resistor nor the capacitors need to be closely matched. Capacitor C1 should have a voltage specification of at least the car's voltage, so 16 V is the next highest standard limit. The only component which needs matching in any way is the USB socket — as it's a board-mounting variety you should make sure you obtain the correct fitment. Alternatively, you could just adapt the printed circuit board copper foil track layout to suit the one you buy.

The printed circuit board layout is shown in Figure 8.4. Start construction by fitting and soldering the terminal pins, diode D1, resistor R1 and capacitors C1 and C2 into the printed circuit board, checking polarity of the electrolytic capacitor C1 prior to soldering.

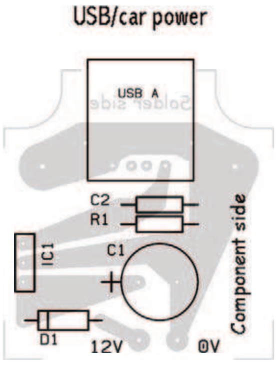

FIGURE 8.4 Printed circuit board layout of the Car-to-USB Power Supply (actual size).

Next up, fit the USB connector to the board. Note that it has four connection terminals, while two extra tabs are used to hold the whole socket in place. The tabs require larger holes (about 2 mm) to be drilled prior to mounting. Bend the lugs of the extra tabs over once the socket is in place on the printed circuit board, to provide extra strength. Solder the tabs to the printed circuit board along with the four terminals.

Take Note

The four connection terminals of a USB type-A socket are numbered as shown in Figure 8.5.

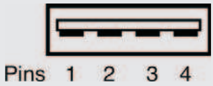

FIGURE 8.5 USB type-A socket, as viewed from the outside.

Now fit the voltage regulator IC1 — note though that if you intend using a small diecast metal housing like the one used in the project here, the regulator must be fitted very close to the printed circuit board, with only 1 or 2 mm clearance, in order that there is enough room for the completed printed circuit board to fit inside the housing. To allow the regulator to fit this closely to the printed circuit board, the component holes need to be larger than 1 mm to allow the wide top end of the regulator pins to go through the board — as shown in Figure 8.6. Use a 1.5 mm drill for these holes.

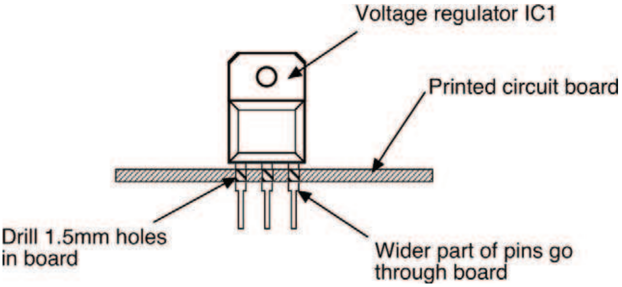
Voltage regulator IC1
Printed circuit board
Drill 1.5mm holes in board
Wider part of pins go through board

FIGURE 8.6 Fit the voltage regulator close to the board to reduce its height – this way, a smaller housing can be used.

Now solder the ends of the connection lead to the terminal pins. This completes the printed circuit board construction.

Drill the housing to take the cable clamp, the bolt to hold the voltage regulator, and two bolts for spacers. File a slot to take the USB connector. When complete, the housing should look something like that in Figure 8.7.

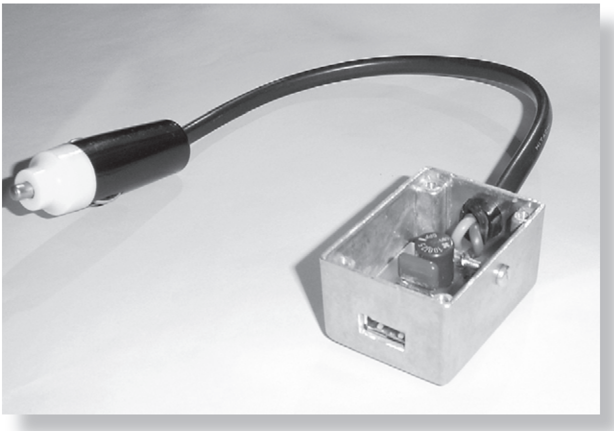

FIGURE 8.7 Showing holes and slot in the Car-to-USB Power Supply housing.

Fit the two printed circuit board mounting bolts to the inside of the housing, using washers and nuts to act as spacers (to prevent the soldered connections on the underside of the printed circuit from touching the metal housing). Next, feed the free end of the connection lead through the cable clamp hole. Rub some heatsink compound on the voltage regulator tab and on the inside of the housing where the regulator is to be bolted. Offer the complete printed circuit into place inside the housing onto the mounting bolts. Insert the voltage regulator bolt through from the outside, fitting a washer and nut on the inside. Fit washers and nuts to the mounting bolts to hold the printed circuit board in place. Tighten all bolts to hold the printed circuit board firmly.

Take Note

Note that no insulating washer is needed between the voltage regulator and the housing. The metal tab of the voltage regulator is at 0 V, so there is no risk of short circuiting between the voltage regulator and other components, or the housing and other parts within the car.

Finally, connect the cigar lighter plug to the free end of the connection lead. Make sure that polarity is correct — that is, the positive connection is the inner terminal of the plug.

Figure 8.8 shows the printed circuit board copper foil track layout.

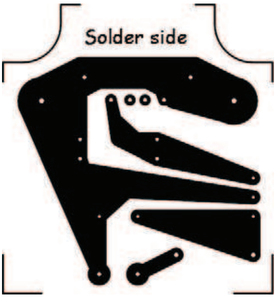

FIGURE 8.8 Printed circuit board copper foil track layout of the Car-to-USB Power Supply (actual size).

Parts list

Semiconductors

IC1	7805 voltage regulator
D1	IN4001 diode

Resistors

R1	10k 0.125 watt

Capacitors

C1	220µ, 16 V radial electrolytic
C2	100n polypropylene

Miscellaneous

Printed circuit board
Housing, diecast aluminium (approximate dimensions 52 by 38 by 32 mm)
USB type-A, printed circuit board-mounting socket
Car cigar lighter plug
Two-core insulated wire — to take at least 0.5 A
Cable clamp

Heart Flasher

OK, I admit it! This is just a fun thing. It's not a practical project at all really — scoring zero on a usefulness rating, it's just meant for a bit of a laugh. Nevertheless, the project does have some purpose, introducing readers to some solid electronic circuit concepts and providing a handy basis from which to practise some circuit building techniques.

The heart of the project — excuse the pun — is an array of light emitting diodes (LEDs), that flash on then off in an ordered and sequential manner, such that the physical shape of the array forms a heart appearance. Persistance of vision as the LEDs flash on and off means that to the observer the heart shape appears vibrant and pulsing (Figure 8.9). Just the thing to build and give to your loved one as, say a soppy Valentine's Day present, or just for fun.

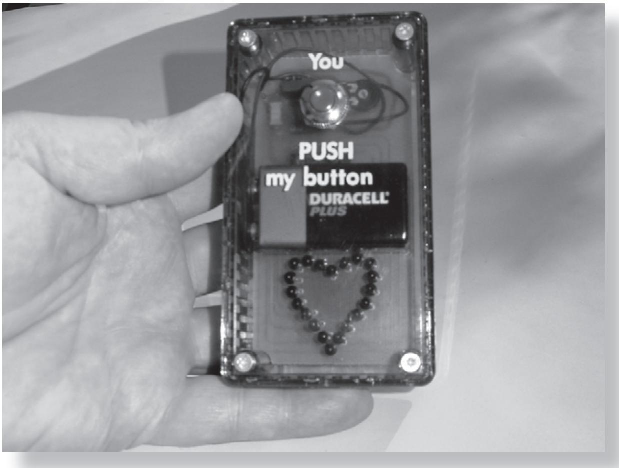

FIGURE 8.9 Wear your heart on your sleeve with the Heart Flasher.

Principle

The Heart Flasher project has three main sections, shown in Figure 8.10. Section one is formed around a clock circuit that provides a constant electrical pulsing at a rate of approximately 10 cycles a second (10 Hz).

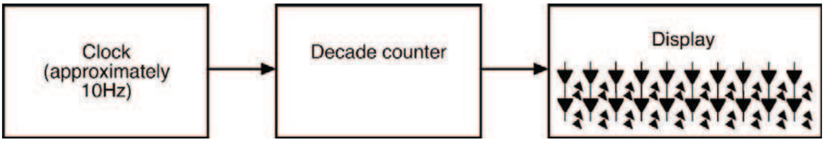
Clock
(approximately
10Hz)
Decade counter
Display

FIGURE 8.10 The principle of the Heart Flasher project.

The next section is a decade counter — a type of circuit that counts the incoming pulses from the clock, and issues output signals to each of ten outputs sequentially. Once the decade counter has counted ten input pulses from the clock, it cycles round to start at the first output again.

The final section of the circuit is an LED array. Some 20 LEDs are used, two per sequential output from the decade counter. The LEDs are formed into the shape of a heart, so that as they are cycled on and off, the visual impact is of a pulsing heart.

Circuit

The circuit of the Heart Flasher is given in Figure 8.11. Integrated circuit IC1 is a 555 timer, connected as an astable multivibrator, to provide a clock signal. Frequency of the astable multivibrator is given by the expression:

$$f = 1.44/C(R1 + 2[RV1 + R2])$$

which means that the frequency can be adjusted between about 50 Hz and 6 Hz.

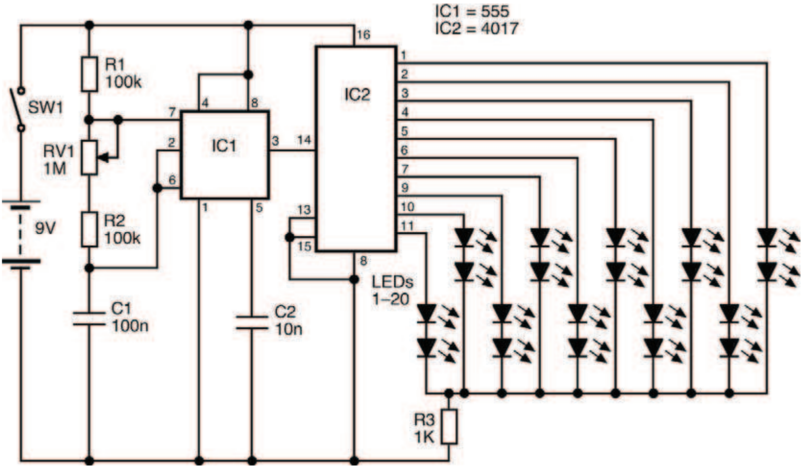
IC1 = 555
IC2 = 4017
R1
100k
SW1
RV1
1M
9V
R2
100k
C1
100n
IC1
IC2
C2
10n
LEDs
1–20
R3
1K
16
1
2
3
4
5
6
7
9
10
11
7
4
8
2
6
3
14
1
5
13
15
8

FIGURE 8.11 Circuit of the Heart Flasher project.

The output (from pin 3 of integrated circuit IC1) is taken to the input of integrated circuit IC2, which is a 4017, divide-by-10 Johnson counter with decoded outputs. Integrated circuit IC2 presents a high logic level at each of its decoded outputs sequentially. While the outputs are shown in the circuit as from pins 1–7 and 9–11 of the integrated circuit, the outputs are not sequentially in this order. Figure 8.12 shows how the actual integrated circuit pin numbers correlate to the decade counter outputs.

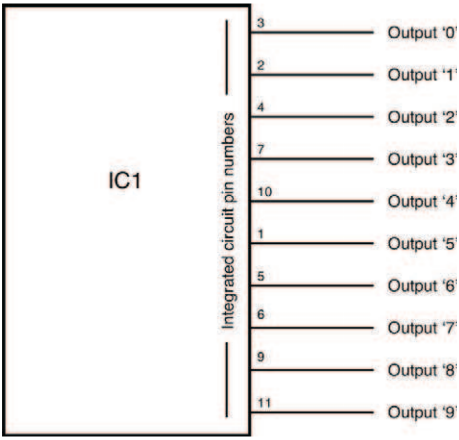

FIGURE 8.12 **Showing how the actual pin numbers of integrated circuit IC1 relate to the decade counter outputs.**

Light emitting diodes LEDs 1–20 form the heart-shaped array, driven directly by the decade counter outputs. Two LEDs are driven by each output, which means that the heart shape is actually made from ten pairs of LEDs, rather than twenty individual ones.

Construction

While construction of the project is not too complex, care must still be taken — particularly with the printed circuit board as it features many thin copper tracks and resultant small component pads. Soldering must therefore only be undertaken with a pretty fine soldering iron bit tip, and preferably only using no higher than a 15–20 watt soldering iron. Figure 8.13 shows the printed circuit layout and wiring connection details.

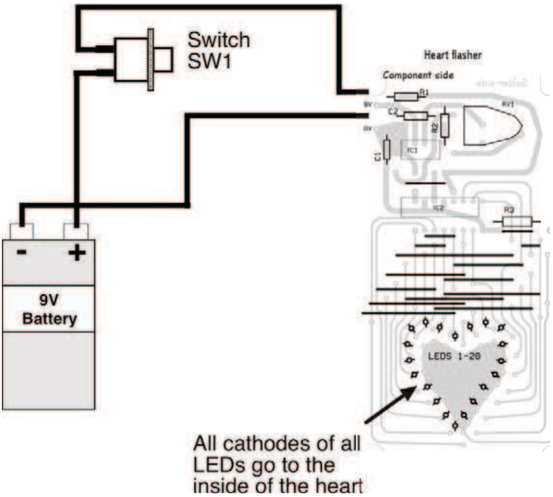

FIGURE 8.13 Printed circuit board layout and wiring connection details of the Heart Flasher.

First, make sure the printed circuit board fits snugly into the housing, filling out the corners to fit the housing corner posts. Insert and solder two terminal pins into the printed circuit board.

There are some fifteen wire links to be made on the printed circuit board, and these should be inserted and soldered next.

Insert and solder the remainder of the components in the following order: resistors, capacitors, integrated circuits, and finally the twenty LEDs. You can use an integrated circuit holder for integrated circuit IC1, but (if you use the same sized housing as used here) you should not use a holder for integrated circuit IC2 — the 9 V battery lies above this, and if IC2 is in a holder there will not be enough room. Make sure that all polarized components (ie, integrated circuits and LEDs) are the correct way round prior to soldering. Note that the cathodes of all LEDs go to the inside of the heart shape. Also, for visual purposes, make sure that all twenty LEDs are mounted at the same height above the board — approximately 10 mm.

Next, solder the battery connector and switch SW1 onto the printed circuit board.

Finally, mark and drill the housing to mount switch SW1, label the housing to suit, and complete the project.

Press switch SW1 to operate the Heart Flasher, and adjust the flash speed to suit by adjusting preset potentiometer RV1. While the flash speed is calculated to give a reasonable flash rate, you can adjust the overall speed by replacing resistors R1 and R2 with lower values to increase the flash rate, or higher values to slow it down.

Figure 8.14 shows the printed circuit board copper foil track layout.

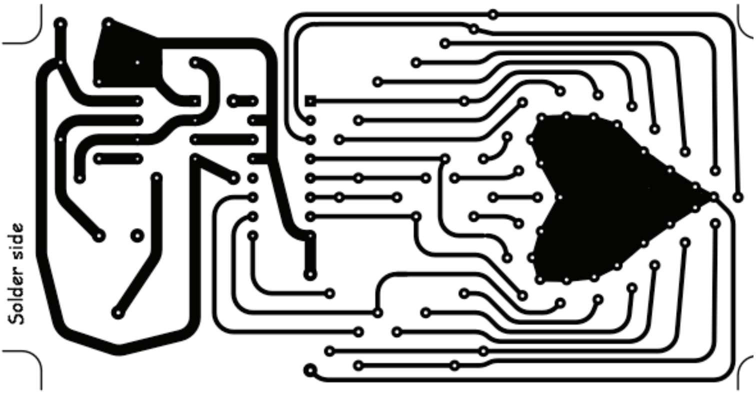

FIGURE 8.14 Printed circuit board copper foil track layout of the Heart Flasher (actual size).

Parts list

Semiconductors

IC1	555 multivibrator
IC2	4017 decade counter
LED1–20	3 mm light emitting diodes

Resistors

R1,2	100k 0.125 watt
R3	1k 0.125 watt
RV1	1M preset potentiometer

Capacitors

C1	100n polyester layer
C2	10n polypropylene

Miscellaneous

SW1 momentary action push button
Printed circuit board
PP3 battery clip
PP3 battery
Housing, transparent plastic (approximate dimensions 112 by 63 by 32 mm)
Terminal pins

Guitar Headphone Amplifier

Are the neighbours giving you grief? Are you fed up with the complaints over your racket? Do you wish they'd all go away and leave you to practise alone? Well, with this Guitar Headphone Amplifier your problems could all be solved. You could slash your axe at the volume you want, without disturbing anyone else's peace.

The Guitar Headphone Amplifier here is a neat project that amplifies your electric guitar's output to a suitably loud level for even the most headstrong head basher, but with ordinary earphones. So no one else hears your playing. The grief disappears, and everyone's happy.

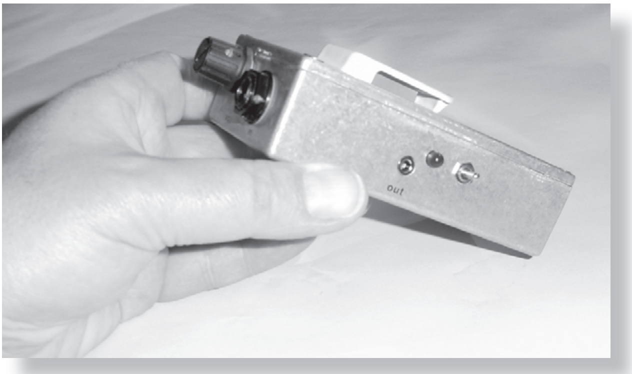

FIGURE 8.15 The Guitar Headphone Amplifier.

Principle

The Guitar Headphone Amplifier is based around two operational amplifiers, which together increase the signal created by an electric guitar when being played in two stages, up to the levels required by a pair of ordinary earphones.

Circuit

The circuit of the Guitar Headphone Amplifier is shown in Figure 8.16, where you can see the two operational amplifiers, which operate in series to amplify the small level signal at the circuit's input up to the levels necessary to drive earphones. The particular operational amplifiers used, while separate electrically, are not actually physically separate devices, and instead are combined in a single integrated circuit IC1. This reduces the size of the circuit, hence allowing a smaller printed circuit board to be used, in turn keeping the physical size of the completed project down to a pocket-sized format.

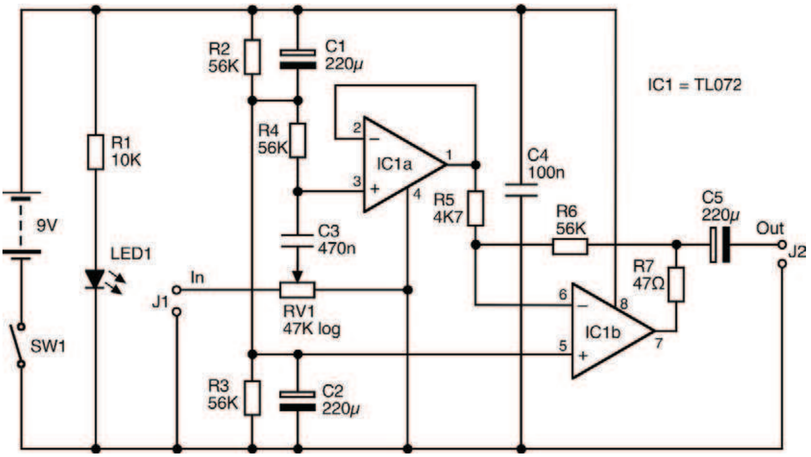

Figure 8.16 *Circuit of the Guitar Headphone Amplifier.*

Potentiometer RV1 allows control of the overall output volume, which in combination with the electric guitar's own volume control, can be configured to give clear sound levels, right up to severe distortion levels. The Guitar Headphone Amplifier should therefore suit most if not all kinds of electric guitar players from folk through to rock, from jazz through to heavy metal.

Operational amplifiers usually require a split-rail or three-level power supply (ie, a power supply with positive, negative, and ground voltages). Operating an operational amplifier from a single supply voltage requires that the voltage (9 V, provided by a PP3-sized battery) be split to give a mid-rail of around 4.5 V, thus emulating a split-rail supply. Capacitors C1 and C2, together with resistors R2 and R3 form a potential divider which does just that.

Construction

While the majority of component holes in the printed circuit board are of the standard 1 mm size, some component holes (for potentiometer RV1, and jack socket J1) should be drilled with a 1.5 mm diameter drill.

Also, if you use the same size of diecast housing as used here, you'll need to file the corner of the printed circuit board for the housing corner post. Also, cut the stem of potentiometer RV1 to the correct length to take its knob (see page 153). These jobs should be carried out before any components are soldered into the printed circuit board.

The printed circuit board layout for the Guitar Headphone Amplifier is shown in Figure 8.17, together with wiring details. Begin construction by inserting and soldering components in the following order: the six terminal pins, resistors R1–7, capacitors C1–5, followed by the larger components jack sockets J1 and J2, and potentiometer RV1.

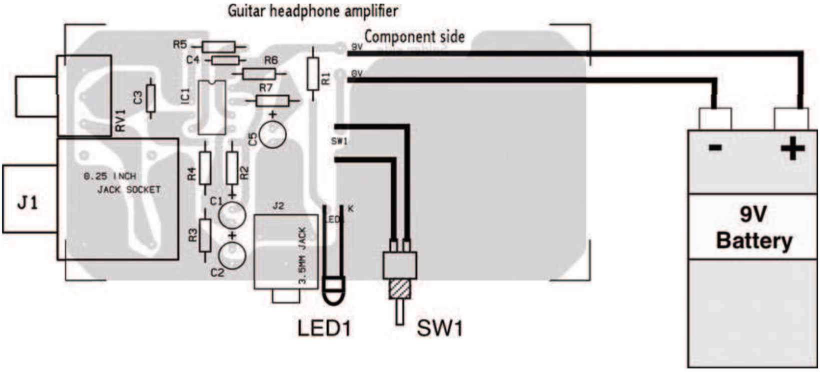

FIGURE 8.17 Guitar Headphone Amplifier printed circuit board layout and wiring diagram.

At this stage you can test the circuit by temporarily soldering the battery clip and switch SW1 to the circuit board. Plug in a guitar (using a standard guitar-style 0.25 inch jack lead) to jack socket J1, and plug in a pair of earphones to jack socket J2.

Once tested, it's time to fit the completed printed circuit board into a housing. Mark and drill holes in the housing for the two jack sockets, the potentiometer, light emitting diode LED1, and switch SW1. Make the holes for jack socket J1 and potentiometer RV1 slightly larger than their actual diameters (by a millimetre or so) then you will be able to wiggle the board into position as the tab of jack socket J2 pushes tightly into its hole in the housing. Using the washers and nuts provided with jack socket J1 and potentiometer RV1, clamp the components to the housing. The printed circuit board should now be held tight within the housing.

Making sure that light emitting diode LED1 is the right way round, push-fit it into the hole you've drilled, and solder its leads to the terminal pins on the printed circuit board. Mount switch SW1 into the housing, and solder its leads to the printed circuit board. Resolder the battery clip to the printed circuit board, and connect the battery.

Finally, mark and drill the holes on the housing lid to mount the belt clip, then bolt the clip to the lid.

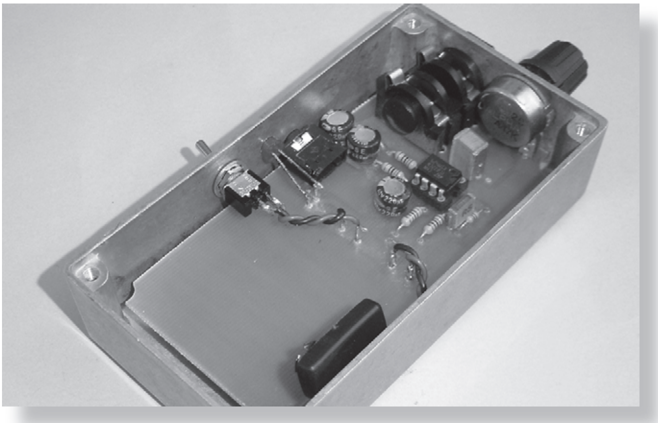

FIGURE 8.18 Showing how the printed circuit board fits inside the diecast housing.

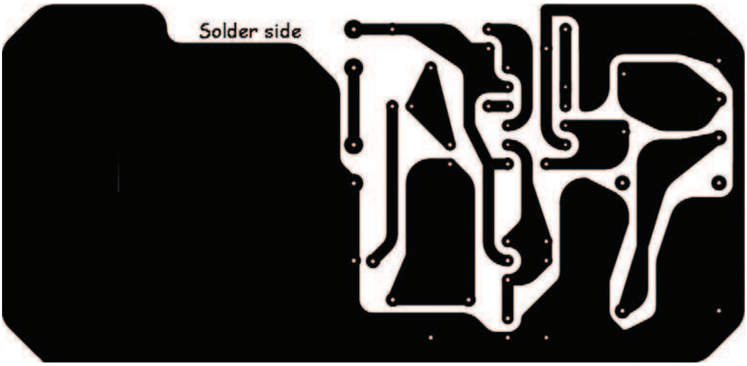

FIGURE 8.19 Printed circuit board copper foil track layout of the Guitar Headphone Amplifier (actual size).

Parts list

Semiconductors

IC1	TL072 dual operational amplifier
LED1	3 or 5 mm light emitting diode

Resistors

R1	10K 0.125W
R2,3,4,6	56K 0.125W
R5	4K7 0.125W
R7	47Ω 0.125W
RV1	47K log miniature printed circuit board mounting potentiometer

Capacitors

C1,2,5	220µ 10V radial electrolytic
C3	470n polyester layer
C4	100n polyester layer

Miscellaneous

Printed circuit board
PP3-sized battery
Battery clip
J1 0.25 inch jack socket, printed circuit board mounting
J2 3.5 mm stereo jack socket, printed circuit board mounting
SW1 SPST miniature toggle switch
Terminal pins
Housing, diecast aluminium (approximate dimensions 110 by 60 by 32 mm)
Belt clip, for housing
Knob (for potentiometer RV1)

Bench Power Supply

If you want to build electronics projects, work with electronic components, experiment with electronic circuits, and do any of the things covered in this book, you'll need some source of electric power. In many cases batteries can be used, of course, and they're fine for low-power circuit requirements, but batteries have several drawbacks. First, they can provide only limited current before their output voltage drops, so any circuit you work with that requires lots of current can't be powered from them. Second, batteries don't last forever and run down (usually when you least want them to!).

A much better source of power is a mains-powered bench power supply (Figure 8.20). These can give significantly greater current and, while connected to mains electricity, always provide the power needed.

To keep the project as simple as possible, the design used here gives a number of fixed voltages, rather than being variable. The voltages directly available are from a dual-rail (ie, positive and negative values) +12 V/0 V/–12 V supply at a maximum output current of about 0.3 A, together with a single-rail +5 V supply up to a maximum output current of about 0.5 A.

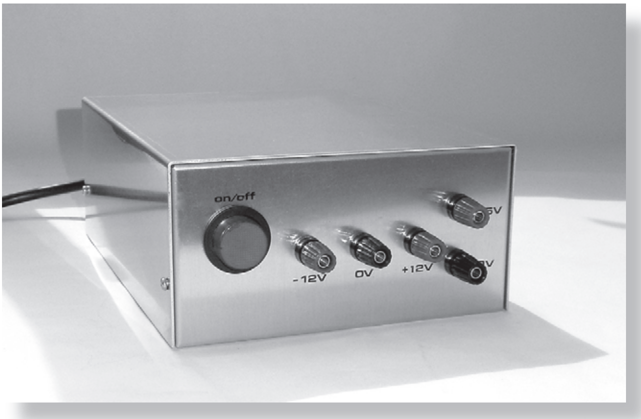

Figure 8.20 The Bench Power Supply.

These two separate power supplies are internally unconnected though it's possible to connect them externally as you wish, to add to or subtract from each other, this way providing a range of supply voltages.

As most circuits you'll build are powered by only a small number of voltage options anyway, the voltages provided by this power supply should be more than adequate. For example, most operational amplifier circuits require a dual-rail power supply of between about +/– 9 V to +/– 15 V, so the output of the power supply here is ideal. Likewise, much logic circuitry needs a single-rail power supply of between about 5 V and 10 V, so the single-rail power supply here is also ideal.

These standard voltages aside, it's also possible to combine the two separate internal supplies to create a variety of other voltages. For example, by connecting the 0 V terminal of the 5 V circuit to the +12 V of the other circuit, and using the –12 V connection as ground, a 29 V supply can be formed (ie, 12+12+5). Likewise, connecting the 5 V terminal to the +12 V terminal, then using the two 0 V terminals a 7 V supply can be formed (ie, 12–5). Other voltages can be formed similarly.

Principle

A mains power supply has the job of reducing and converting a high mains AC voltage into a low DC voltage. In a standard power supply this is undertaken in four main stages, illustrated in Figure 8.21 together with the waveforms at the various stages.

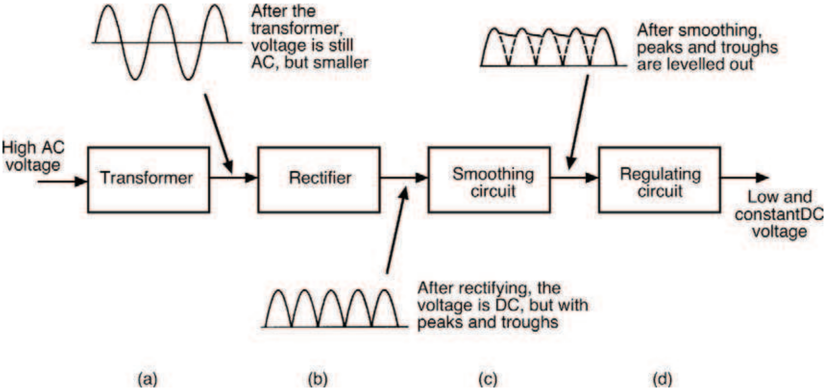
After the transformer, voltage is still AC, but smaller
After smoothing, peaks and troughs are levelled out
High AC voltage
Transformer
Rectifier
Smoothing circuit
Regulating circuit
Low and constantDC voltage
After rectifying, the voltage is DC, but with peaks and troughs
(a)
(b)
(c)
(d)

FIGURE 8.21 The four main stage waveforms in the mains-powered bench power supply.

First, the high mains AC voltage is transformed to a lower AC voltage, as shown at Figure 8.21 (a). After this the AC voltage is converted to a DC one by rectification — Figure 8.21 (b). Next, the voltage is smoothed to reduce the peaks and troughs of the basic rectified voltage — Figure 8.21 (c). Finally, as shown at Figure 8.21 (d), the output voltage is regulated to maintain a constant DC voltage.

The power supply here follows these basic stages, although there are in fact three individual power supply circuits making up the overall design. Two circuits combine together to give positive and negative +12 V/–12 V values, while the third 5 V circuit is unconnected. The main principle of each is the same however.

Circuit

The overall circuit of the power supply is shown in Figure 8.22, where the three individual circuits within the whole circuit can be identified. Transformers T1 and T2 transform 230 V ac mains voltage input to the lower voltages (15 V ac and 9 V ac) required by the rest of the circuit.

Bridge rectifiers BR1 and BR2 rectify the alternating voltages to direct voltages, while capacitors C1–3 smooth these voltages.

Integrated circuits IC1–3 are voltage regulators, which regulate the smoothed voltages into slightly lower, but fixed output voltages.

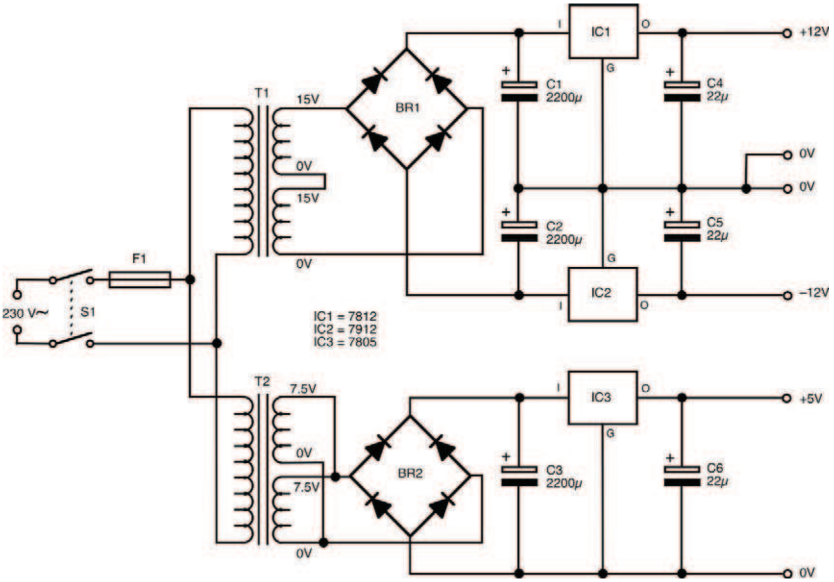
IC1
I
O
G
+12V
+
C1
2200µ
+
C4
22µ
T1
15V
BR1
0V
15V
0V
0V
0V
+
C2
2200µ
+
C5
22µ
G
F1
IC2
I
O
−12V
230 V~
S1
IC1 = 7812
IC2 = 7912
IC3 = 7805
T2
7.5V
IC3
I
O
G
+5V
0V
BR2
+
C3
2200µ
+
C6
22µ
7.5V
0V
0V

FIGURE 8.22 Mains-powered bench power supply circuit.

Construction

The printed circuit board contains most of the components for the Bench Power Supply, and is a compact yet easy-to-build design. Some holes on the printed circuit board should be drilled larger than 1 mm — the mounting holes for the housing and the mounting holes for transformer T1 are all 4 mm diameter, while the terminal holes for transformers T1 and T2 are 1.5 mm. Figure 8.23 shows the layout of the printed circuit board to locate the components.

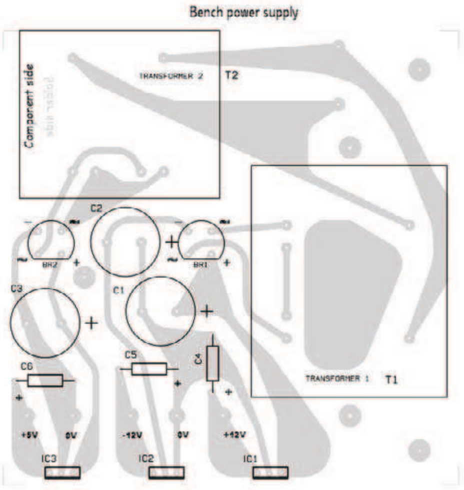

FIGURE 8.23 Layout of the Bench Power Supply printed circuit board.

Now position the printed circuit on the inside base of the housing where it is to be mounted — make sure the positions for the three integrated circuits IC1–3 are against the rear panel of the housing. Mark the holes in the housing base for the three mounting bolts. Remove the printed circuit board, and drill the holes in the housing base, using a 4 mm drill. Insert the 20 mm long M3 bolts and spacers, ready for the printed circuit board.

Insert and solder the seven terminal pins, and integrated circuits IC1–3 into the printed circuit board.

Mount the printed circuit board temporarily onto the three mounting bolts in the housing base. Mark the three integrated circuits IC1–3 mounting holes on the rear panel of the housing. Now remove the printed circuit board again, and drill the holes you just marked, using a 4 mm drill. Make sure the holes have no burrs on their edges, and if burrs exist use a large drill to hand ream the hole edges to remove them.

Now insert and solder all remaining components into the printed circuit board, in the following order: bridge rectifiers BR1 and BR2, capacitors C1–6, transformers T1 and T2. Now, using 6 mm long M3 bolts, bolt transformer T1 to the printed circuit board.

Mark and drill holes in the housing panels for the cable clamp, the panel-mounting fuse holder, and the on/off switch, then mount these components to the housing.

Position the printed circuit board onto its mounting bolts in the housing, but do not fasten it down with nuts and washers yet. Insulating washers must fit between the three integrated circuits IC1–3 and the housing panel, as the three integrated circuit tabs are at different voltages, and bolting them directly to the panel would create short circuits. Each integrated circuit must therefore be bolted to the panel as shown in Figure 8.24, with an insulating washer between the integrated circuit tab and the panel, an insulating bush through the integrated circuit tab hole, and a bolt through the bush. This arrangement ensures that the integrated circuit is electrically isolated from the housing, though rigidly held to it.

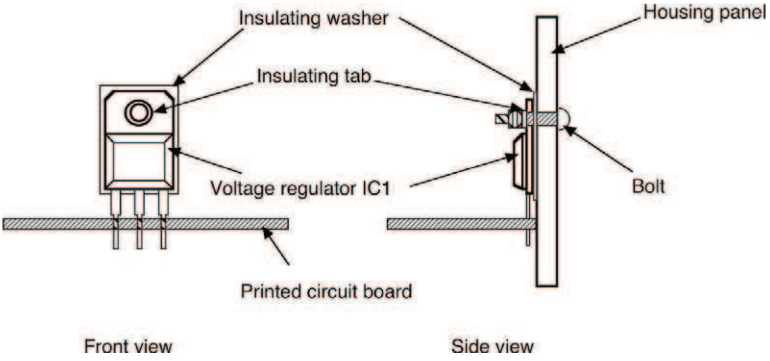

Figure 8.24 Mounting arrangement for integrated circuits IC1–3 to the housing rear panel.

The panel is also used as a heatsink, dissipating heat safely away from the integrated circuits, so heatsink compound must be applied to all mating surfaces prior to assembly. See page 165 for details of applying heatsink compound to surfaces.

Once heatsink compound is applied, the insulating washers and tabs are fitted, and the three integrated circuits are bolted to the housing rear panel. Complete the mechanical assembly by fastening the printed circuit board with nuts and washers onto the three mounting bolts.

Following the wiring diagram of Figure 8.25, wire up the project. For safety, and following the same sort of layout as in Figure 8.26, keep all mains wiring to one side of the housing, running it from the housing panel-mounted components (cable clamp, fuse holder, and on/off switch) down to the housing base, along the base and up the other panel. Take mains connections to the printed circuit board from the panel-mounted components down to the base, along the base with the other mains

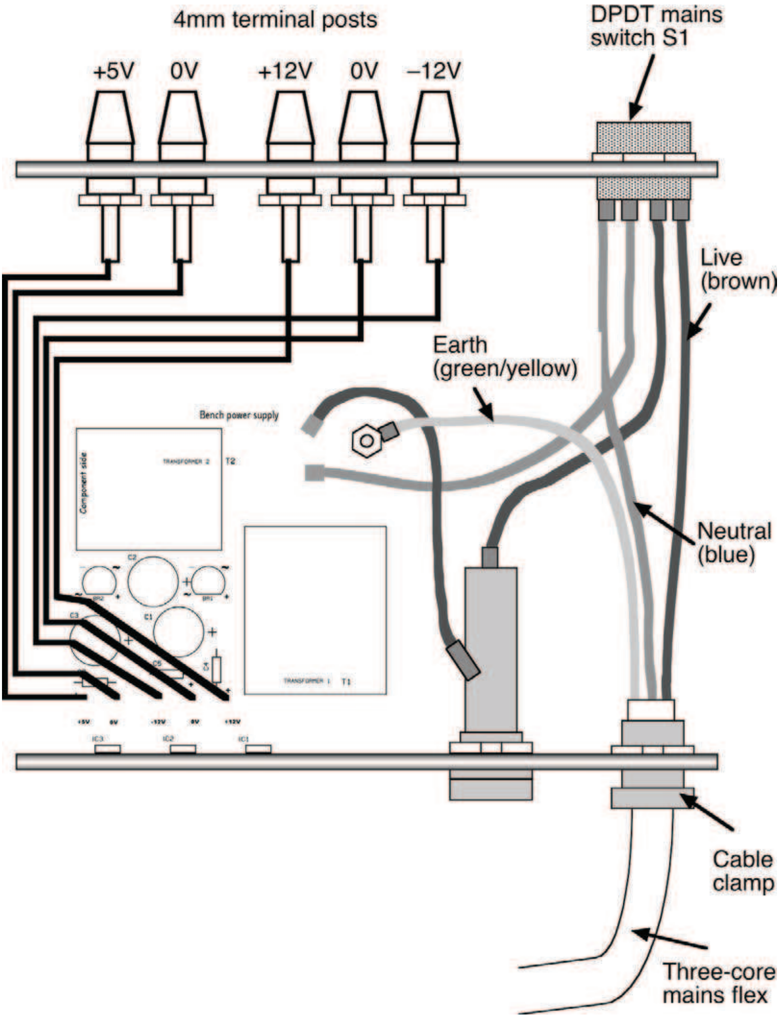
4mm terminal posts
DPDT mains switch S1
+5V
0V
+12V
0V
–12V
Live (brown)
Earth (green/yellow)
Bench power supply
Component side
TRANSFORMER 2
T2
Neutral (blue)
C2
BR2
BR1
C3
C1
C5
C4
TRANSFORMER 1
T1
+5V
0V
-12V
0V
+12V
IC3
IC2
IC1
Cable clamp
Three-core mains flex

FIGURE 8.25 Wiring diagram for the Bench Power Supply.

wires and directly from the wire group to the printed circuit board. Make doubly sure all mains connections are made tight. Use heatshrink insulated sleeving on all mains connections to make sure that no metal parts remain visible (or more important, make sure they are not touchable).

The five low voltage wires from the printed circuit board to the connectors should be taken to the other edge of the housing to the mains wires, running along the housing base to the front panel, then along the housing base to each connector.

Once wiring is completed, use cable ties or similar to tie the loose wires together into cables, as can be seen in Figure 8.26.

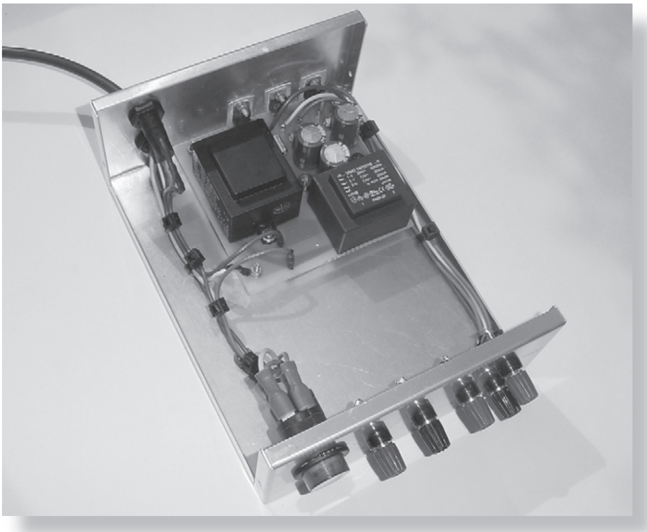

FIGURE 8.26 Showing the Bench Power Supply innards – note the use of cable ties to form the two groups of wires (mains wiring on the left, low voltage wires on the right).

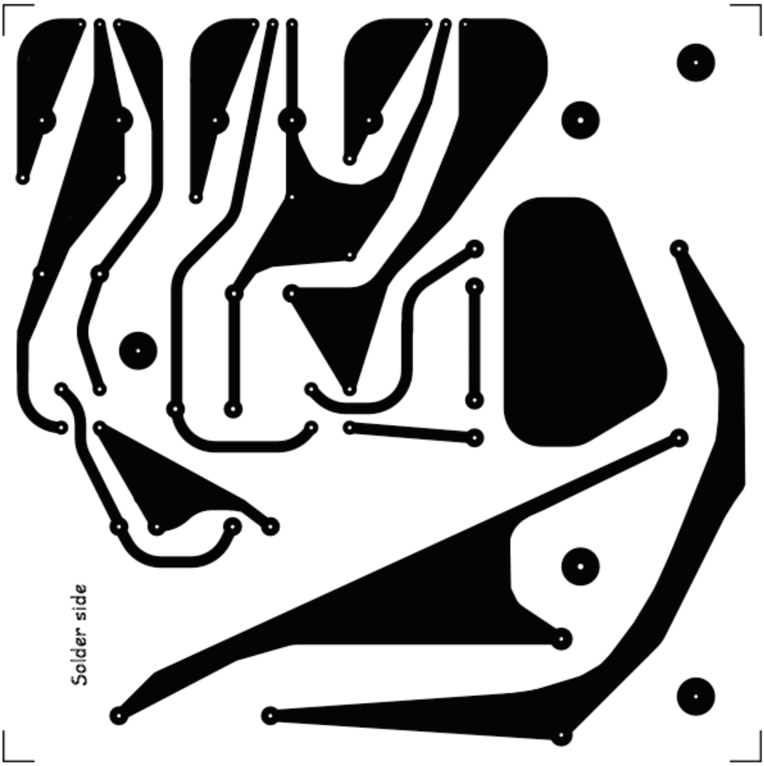
Solder side

FIGURE 8.27 Bench Power Supply printed circuit board copper track foil pattern (actual size).

Parts list

Semiconductors

IC1	7812 voltage regulator
IC2	7912 voltage regulator
IC3	7805 voltage regulator
BR1,2	W05 bridge rectifier (1.5A, 50Vac)

Capacitors

C1,2,3	2200µ, 25 V radial electrolytic
C4,5,6	22µ, 16V radial electrolytic

Miscellaneous

Printed circuit board	
T1	2 x 15 V, 10 VA printed circuit mounting transformer
T2	2 x 7.5 V, 5 VA printed circuit mounting transformer
S1	DPDT illuminated push button mains switch
Housing, folded aluminium (approximate dimensions 200 by 152 by 78 mm)	
5 x 4 mm terminal posts (2 black, 2 red, 1 green)	
3 insulating washers and tabs for TO-220 style components	
M3 (6 mm) bolts, nuts, washers	
M3 (20 mm) bolts, nuts, washers	
M3 10 mm spacers	
FS1	panel-mounting fuse holder
100 mA 20 mm fuse	
Length of 1 A mains flex and plug	
Cable clamp	
Terminal connection pins	
Solder tag (M3)	

Component suppliers

Over the next few pages is a list of some of the mail order electronic component suppliers in the UK. The list is just a sample of suppliers, and the presence of any supplier on the list is by no means an endorsement. Look in a telephone book for local suppliers. Also, check my website, at:

http://keithbrindley.members.beeb.net

where I'll try to keep an updated list of supplier links and any other relevant information.

2001 Electronic Components Ltd
Eastman Way
Stevenage Business Park
Pin Green
Stevenage
SG1 4SU

Telephone: 01438 742001
URL: http://www.2k1.co.uk

Bowood Electronics Limited
7 Bakewell Road
Baslow
Bakewell
Derbyshire
DE45 1RE

Telephone: 01246 200 222
URL: http://www.bowood-electronics.co.uk

Cricklewood Electronics
40-42 Cricklewood Broadway
London
NW2 3ET

Telephone: 020 8452 0161
URL: http://www.cricklewoodelectronics.co.uk

Dannell Electronics Limited
Unit 2 Funston's Commercial Centre
Arkesden Road
Clavering
Saffron Walden
Essex
CB11 4QU

Telephone: 0845 603 1509
URL: http://www.dannell.co.uk

Disponible
Highfield House
College Business Park
Ripon
North Yorkshire
HG4 2SG

URL: http://www.disponible.co.uk

ElectroValue Ltd
Unit 5
Beta Way
Thorpe Industrial Park
Egham
Surrey
TW20 8R

Telephone: 01784 433604
URL: http://www.electrovalue.co.uk

ESR Electronic Components
Station Road
Cullercoats
Tyne & Wear
NE30 4PQ

Telephone: 0191 251 4363
URL: http://www.users.zetnet.co.uk/esr

Farnell InOne,
Canal Road,
Leeds,
LS12 2TU

Telephone: 0871 200200
URL: http://uk.farnell.com

Fast Components Limited
Winchester House
Winchester Road
Walton on Thames
Surrey
KT12 2RH

Telephone: 0870 750 4468
URL: http://www.fastcomponents.co.uk

GreenWeld Limited
14 West Horndon Business Park
West Horndon
Brentwood
Essex
CM13 3XD

Telephone: 01277 811 042
URL: http://www.greenweld.co.uk

Henrys Electronics Ltd
404 Edgware Road
Paddington
London
W2 1ED

Telephone: 020 7258 1831
URL: http://www.henrys.co.uk

JPR Electronics Ltd
4 Circle Business Centre
Blackburn Road
Dunstable
LU5 5DD

Telephone: 01582 470000
URL: http://www.jprelec.co.uk

Magenta Electronics Ltd
135 Hunter Street
Burton-on-Trent
DE14 2ST

Telephone: 01283 565435
URL: http://www.magenta2000.co.uk

Mainline Surplus Sales
Unit 1A Cutters Close Industrial Estate
Cutters Close
Narborough
Leicester
LE19 2FZ

Telephone: 0870 241 0810
URL: http://www.mainlinegroup.co.uk

Maplin Electronics Ltd
National Distribution Centre
Valley Road
Wombwell
Barnsley
South Yorkshire
S73 0BS

Telephone: 0870 4296000
URL: http://www.maplin.co.uk

Mega Electronics
Mega House
Grip Industrial Estate
Linton
Cambridge
CB1 6NR

Telephone: 01223 893900
URL: http://www.megauk.com

Quasar Electronics Ltd
PO Box 6935
Bishops Stortford
CM23 4WP

Telephone: 08702 461 826
URL: http://www.quasarelectronics.com

Rapid Electronics Ltd
Severalls Lane
Colchester
Essex
CO4 5JS

Telephone: 01206 751166
URL: http://www.rapideducation.co.uk

RS Components Ltd
Birchington Road
Corby
Northants
NN17 9R

Telephone: 01536 444222
URL: http://rswww.com

SIR-KIT Electronics
57 Severn Road
Clacton-on-Sea
Essex
CO15 3RB

URL: http://www.geocities.com/sirkituk

Sycom
PO Box 148
Leatherhead
Surrey
KT22 9Y

Telephone: 01372 361421
URL: http://www.sycomcomp.co.uk

This Page Intentionally Left Blank

Index

Symbols

1N4001 46
1N4148 46
555 53
741 56

A

abrasive fibreglass pencil 14, 115
abrasive scrubbing block 15, 115
analog 54
    integrated circuits 54
axial-leaded 40
    insertion 119

B

batteries 57
battery connector 57
bipolar transistor 49
bridge rectifier 47
buzzer 65

C

cable entry clamp 170
cable markers 169
cable ties 169
CAD (computer aided design) 95
    Osmond 96
    panel labelling 174
    realPCB 96
capacitor 38
    axial-leaded 40
    colour code 42
    dielectric 41
    electrolytic 39
    non-electrolytic 39, 40
    number/letter codes 44
    radial-leaded 40
    variable 42
    working voltage 41
carbon resistors 29
circuit design
    simulator 95

coil 67
components 2, 3, 27
  resistor 28
  trimming 122
component insertion 77, 105
computer design:
  panel labelling 174
  printed circuit board 95
connecting lead 123, 124
  soldering to a hole 130
  soldering to a terminal pin 128
  soldering to board 126
connections, fitting 158
connectors 67
copper-clad 74
copper foil 76, 83
  computer design 95
    Osmond 96
    realPCB 96
cored solder 113
Crocodile Technology 95

D

de-wetting 111
desoldering 20, 131
  desoldering braid 20, 132
  desoldering pump 21, 133
    using 133
dielectric 41
digital 54
  integrated circuits 54
DIL IC 53
diodes 45
  bridge rectifier 47
  forward biased 46
  rectifier 45
  reverse biased 46
drill and bits 22
dual-in-line integrated circuit 53

E

electrolytic 39
electronics simulator 95
electronic:
  components 2
  device 2
etch-resist 77, 87
  ink 88
  paint 88
  photographic 91
  removal 104
    abrasive fibreglass pencil 104
    abrasive removal block 104
    solvent 104
  transfers 90
etchant 77, 101
etching printed circuit board 101

F

farad 39
fault-finding 177
  components 183
  connectors 183
  PCB faults 181

  - soldered joint faults 181
  - strategy 178
  - total method 185
  - types of faults 180
  - wiring faults 183
- ferric chloride 101
- fibreglass pencil 14
- field effect transistors 51
- flux 112
- flux-cored solder 113
- foil pattern 83
- forward biased 46
- fuse holder, chassis-mounted 162

H

- heatsink 163
  - compound 165
  - insulating washer 166
  - using thermal compound 165
  - variables 164
- heat shunt 16
- housing 2, 4, 139
  - IP rating 144
  - metal diecast 143
  - metal folded chassis-type 143
  - mounting printed circuit boards 145
  - panel labelling 173
  - plastic 142

I

- IC extraction tool 24
- IC insertion tool 24
- indicators 58
- inductor 67
- insulating:
  - sleeving 170
  - washer 166
- integrated circuit 53
  - analog 54
  - digital 54
  - holder 69
  - logic gate 57
- integrated circuit extraction tool 24
- integrated circuit insertion tool 24
- intermetallic:
  - bond 109
  - compound 109
- IP rating 144

J

- JetStar 97

K

- knobs 155

L

- labelling
  - key points 174
  - panel 174, 176
- LaserStar 97
- lead-free solder 108
- leadouts: transistor 49
- LED 59
  - fitting 161
- light dependent resistor 38

light emitting diode 59
links 71
  printed circuit board 136
    making 137
logic gate 57
loudspeaker 65

M

mains wiring: general guidelines 170
matrix board 80
metal diecast housing 143
metal film resistors 30
metal folded chassis-type housing 143
mica 40
microfarad 39
microphone 66
miniature long-nosed pliers 5, 18
miniature side-cutters 5, 17
multi-stranded wire types 168

N

N-channel 51
N-type 50
nanofarad 39
non-electrolytic 39, 40
non-wetting 111

O

Ohm 28
ohm 28
op-amp 55
operational amplifier 55
Osmond 96

P

P-channel 51
P-type 50
panel labelling 173
  key points 174
  Quick-Laser 176
partial wetting 111
PCB 73
photographic resist 92, 97
  developing board 93
    LaserStar & JetStar 97
picofarad 39
pin pusher 25
plastic housing 142
plumbum 107
polycarbonate 40
polyester 40
polypropylene 40
polystyrene 40
potentiometers 31, 150
  cutting stem 153
  dual-gang 33
  knobs 155
  linear 32
  logarithmic 32
  mounting 151
  preset 32
  slider 33
  tightening nut 152
power rating 29
pre-heating joints 120
preferred values 36
preset potentiometers 32
Press-n-Peel 99
printed circuit board 2, 3, 73
  component insertion 77

  component lead trimming 77
  computer design
    Osmond 96
  connecting leads 123
  copper-clad 74
  copper foil 76, 83
    LaserStar & JetStar 97
    Press-n-Peel 99
  drilling 77, 103
  etch-resist 77
  etchant 77
  etching 101
  links 71, 136
  main steps 79
  mounting, in housing 145
    double-sided adhesive sponge fixers 146
    spacers 147
  photographic resist 92, 93
  realPCB 96
  soldering 77
  stand 23
  tinning 125
  ultra-violet light box 92
projects 193
  bench power supply 212
    circuit 214
  car-to-USB power supply 195
    circuit 197
  guitar headphone amplifier 207
    circuit 207
  heart flasher 202
    circuit 203

Q

Quick-Laser 176

R

radial-leaded 40
  insertion 119
realPCB 96
rectifier 45
relay 64
resist:
  photographic 92
  ultra-violet light box 92
resistor 28
  colour code 34
  power rating 29
  preferred values 36
  stability 29
  tolerance 29
  types 29
    carbon resistors 29
    light dependent 38
    metal film resistors 30
    potentiometers 31
    surface mount 36
    thermistor 37
    wirewound resistors 31
resistor colour code 34
reverse biased 46
ribbon cable 169

S

screened cable 159
scrubbing block 15

semiconductors 44
  bridge rectifier 47
  diodes 45
  field effect transistors 51
  integrated circuit 53
  j-fet 51
  LED 59
  logic gate 57
  mosfet 51
  rectifier 45
  silicon controlled rectifier 52
  thyristor 52
  transistor 47
    leadouts 49
  triac 52
silicon controlled rectifier 52
simulator 95
SMC 36
sodium hydroxide 93
solder braid
  using 132
solder removal 6
solder sucker 21
solder tags 160
soldering 77, 107
  cleaning 115
  de-wetting 111
  flux 112
  lead-free 108
  non-wetting 111
  partial wetting 111
  pre-heating joints 120
  solder 108
  total wetting 111
  wetting 109
soldering iron 5, 7
  accessories 13
    abrasive fibreglass pencil 14
    abrasive scrubbing block 15
    heat shunt 16
    sponge 13
    tip tinning and cleaning block 14
    wet sponge 117
  bit shapes 10
  care of 135
  gas-powered 9
  low-voltage 9
  mains-powered 8
  power 7
  rechargeable 8
  station 9
  tinning 116, 117
sounder 65
spacers 147
sponge 13
spot-face cutter 26, 82
stability, of resistor 29
stripboard 81
stuffing 106
surface mount components 36
switches 61, 155
  fitting a rectangular switch 157

T

terminal pins 69
terminal pin pusher 25
thermal:
  compound 165
  conductivity 165
  grease 165
  joint compound 165
thermistor 37
thyristor 52
tinning:
  connecting lead 124
  printed circuit board 125
  soldering iron 116
tip tinning and cleaning block 14
tolerance 29
tools 5
  desoldering 20, 131
    braid 20, 131
    pump 21, 131
  drill and bits 22
  integrated circuit extraction tool 24
  integrated circuit insertion tool 24
  miniature long-nosed pliers 18
  miniature side-cutters 17
  printed circuit board stand 23
  solder:
    removal 20
    sucker 21
  spot-face cutter 26, 82
  terminal pin pusher 25
  wire-strippers 19
total wetting 111
transfers:
  etch-resist 90
transistor 47
  bipolar 49
  leadouts 49
  N-type 50
  P-type 50
triac 52
trimming 122

U

ultra-violet 92, 98
  light box 92
    developing board 93
      using LaserStar & JetStar 98
universal serial bus (USB) 195
unsoldering 131
  desoldering pump 133
  solder braid 132

W

wet sponge 117
wetting 109
wire-strippers 6, 19
working voltage 41

Copyrighted Material

RSGB RADIO SOCIETY OF GREAT BRITAIN

D-i-Y RADIO
Yearling
ANT. TUNE
BANDSPREAD
RF GAIN
OFF / AF GAIN
14.35
3.5
14.0
5.6

RADIO & ELECTRONICS COOKBOOK

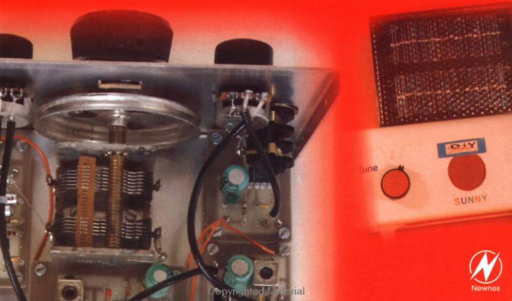
D-i-Y
Tune
SUNNY

Newnes

Copyrighted Material

Radio and Electronics Cookbook

Radio and Electronics Cookbook

Edited by
Dr George Brown, CEng, FIEE, M5ACN

RSGB

Newnes
OXFORD AUCKLAND BOSTON JOHANNESBURG MELBOURNE NEW DELHI

Newnes
An imprint of Butterworth-Heinemann
Linacre House, Jordan Hill, Oxford OX2 8DP
225 Wildwood Avenue, Woburn, MA 01801-2041
A division of Reed Educational and Professional Publishing Ltd

A member of the Reed Elsevier plc group

First published 2001

© Radio Society of Great Britain 2001

All rights reserved. No part of this publication may be reproduced in any material form (including photocopying or storing in any medium by electronic means and whether or not transiently or incidentally to some other use of this publication) without the written permission of the copyright holder except in accordance with the provisions of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London, England W1P 0LP. Applications for the copyright holder's written permission to reproduce any part of this publication should be addressed to the publishers

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN 0 7506 5214 4

RSGB
Lambda House
Cranborne Road
Potters Bar
Herts
EN6 3JE

Composition by Genesis Typesetting, Laser Quay, Rochester, Kent
Printed and bound in Great Britain

PLANT A TREE
BTCV
British Trust for Conservation Volunteers
FOR EVERY TITLE THAT WE PUBLISH, BUTTERWORTH-HEINEMANN WILL PAY FOR BTCV TO PLANT AND CARE FOR A TREE.

Contents

Preface ix

1. A medium-wave receiver 1
2. An audio-frequency amplifier 4
3. A medium-wave receiver using a ferrite-rod aerial 9
4. A simple electronic organ 12
5. Experiments with the NE555 timer 17
6. A simple metronome 21
7. What is a resistor? 24
8. Waves – Part 1 27
9. A beat-frequency oscillator 31
10. What is a capacitor? 34
11. Waves – Part 2 38
12. An LED flasher 41
13. Waves – Part 3 44
14. Choosing a switch 46
15. An aerial tuning unit for a receiver 49
16. A simple 2 m receiver preamplifier 52
17. Receiving aerials for amateur radio 54
18. The Colt 80 m receiver – Part 1 58
19. A crystal radio receiver 62
20. The varactor (or varicap) diode 64
21. A portable radio for medium waves 65
22. The Colt 80 m receiver – Part 2 70

23. A simple transistor tester 73
24. An introduction to transmitters 77
25. The Colt 80 m receiver – Part 3 81
26. A two-way Morse practice system 88
27. The Colt 80 m receiver – Part 4 91
28. A simple crystal set 95
29. A crystal calibrator 100
30. A simple short-wave receiver – Part 1 104
31. A fruit-powered medium-wave radio 106
32. A capacitance bridge 109
33. A simple short-wave receiver – Part 2 113
34. A basic continuity tester 117
35. A charger for NiCad batteries 119
36. An 80 metre crystal-controlled CW transmitter 123
37. A solar-powered MW radio 129
38. A receiver for the 7 MHz amateur band 133
39. Diodes for protection 137
40. An RF signal probe 140
41. An RF changeover circuit 142
42. A low-light indicator 146
43. A J-pole aerial for 50 MHz 149
44. Measuring light intensity – the photometer 153
45. A 70 cm Quad loop aerial 156
46. A UHF field strength meter 160
47. Christmas tree LEDs 162
48. An audio signal injector 166
49. Standing waves 168
50. A standing-wave indicator for HF 170
51. A moisture meter 174
52. Simple aerials 177
53. A breadboard 80 cm CW transmitter 182

54. A 7-element low-pass filter for transmitters 186
55. Radio-frequency mixing explained 189
56. A voltage monitor for a 12 V power supply 192
57. A 1750 Hz toneburst for repeater access 196
58. A circuit for flashing LEDs 201
59. Digital logic circuits 205
60. A resistive SWR indicator 210
61. An audio filter for CW 213
62. An electronic die 215
63. The absorption wavemeter 222
64. An HF absorption wavemeter 224
65. A vertical aerial for 70 cm 228
66. A UHF corner reflector aerial 230
67. A switched dummy load 234
68. A simple Morse oscillator 238
69. A bipolar transistor tester 240
70. The ‘Yearling’ 20 m receiver 245
71. Adding the 80 metre band to the Yearling receiver 251
72. How the Yearling works 255
73. A field strength meter 258
74. Preselector for a short-wave receiver 261
75. An audible continuity tester 265
76. An experimental 70 cm rhombic aerial 268
77. Water level alarm 272
78. A delta loop for 20 metres 275
79. A simple desk microphone 279
80. Morse oscillator 284
81. A simple 6 m beam 287
82. An integrated circuit amplifier 291
83. A novice ATU 293
84. CW QRP transmitter for 80 metres 297

85. An audio booster for your hand-held 303
86. A grid dip oscillator 306
87. A CW transmitter for 160 to 20 metres 312
88. Matching the end-fed random-wire aerial 315

Preface

Although we are surrounded by sophisticated computerised gadgets these days, there is still a fascination in putting together a few resistors, capacitors and the odd transistor to make a simple electronic circuit. It is really surprising how a handful of components can perform a useful function, and the satisfaction of having built it yourself is incalcuable.

This book aims to provide a wide variety of radio and electronic projects, from something that will take a few minutes to a more ambitious weekend's worth. Various construction techniques are described, the simplest requiring no more than a small screwdriver, the most complex involving printed circuit boards.

Originally published by the Radio Society of Great Britain, the projects were all chosen to be useful and straightforward, with the emphasis on practicality. In most cases the workings of the circuit are described, and the projects are backed up by small tutorials on the components and concepts employed. In the 21st century it may seem strange that few of the published circuits use integrated circuits (chips). This is intentional as it is much easier to understand how the circuit works when using discrete components.

Anyone buying the *Radio and Electronics Cookbook* will find that it will lead to hours of enjoyment, some very useful and entertaining gadgets, and increased knowledge of how and why electronics circuits work, and a great sense of satisfaction. Beware, electronic construction is addictive!

WARNING: This book contains construction details of transmitters. It is illegal to operate a transmitter without the appropriate licence. Information on how to obtain an Amateur Radio Licence can be obtained from the Radiocommunications Agency, tel. 020 7211 0160.

1 A medium-wave receiver

Introduction

Let us start off with something that is really quite simple and yet is capable of producing a sense of real satisfaction when complete – a real medium-wave (MW) radio receiver! It proves that receivers *can* be simple and, at the same time, be useful and enjoyable to make. To minimise the confusion to absolute beginners, *no* circuit diagram is given, only the constructional details. The circuits will come later, when you have become accustomed to the building process. In the true amateur spirit of ingenuity and inventiveness, the circuit is built on a terminal strip, the coil is wound on a toilet roll tube (as amateur MW coils have been for 100 years!), and the receiver is mounted on a piece of wood.

Putting it together

Start by mounting the components on the terminal strip as shown in **Figure 1**, carefully checking the position and value of each one. The three capacitors are all the same, and so present no problem. They (and the resistors) may be connected either way round, unlike the two semiconductors (see later). The resistors are coded by means of coloured bands. You can refer to Chapter 7 if you have difficulty remembering the colours and their values.

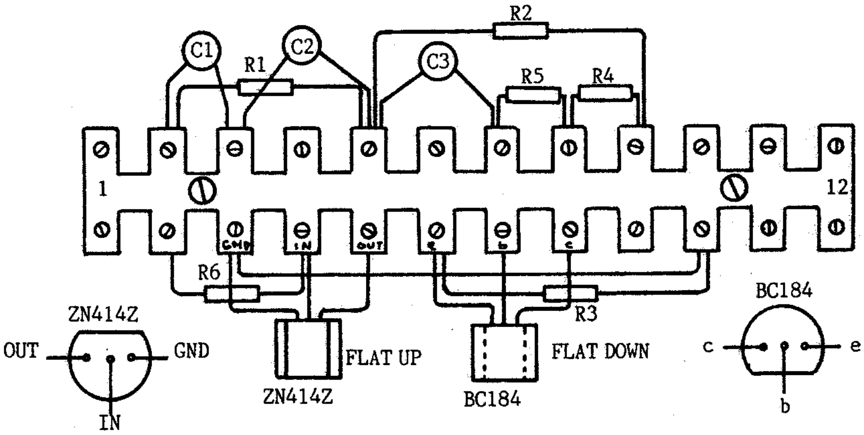

Figure 1 Terminal strip – position of components

1.	Brown, Black, Yellow	100 000 ohms	(R1, R5, R6)
2.	Green, Blue, Brown	560 ohms	(R2)
3.	Red, Violet, Brown	270 ohms	(R3)
4.	Brown, Black, Orange	10 000 ohms	(R4)

The integrated circuit (the ZN414Z) and the transistor (the BC184) *must* be connected correctly. Check Figure 1 carefully before fitting each device.

Now wind the coil. Most tubes are about 42 mm diameter and 110 mm long. Don't worry if your tube is slightly different; it shouldn't matter. Make two holes, about 3 mm apart, about 40 mm from one end, as shown in **Figure 2**. Loop your enamelled wire into one hole and out of the other, and draw about 100 mm through; loop this 100 mm through again, thus anchoring the wire firmly. Now wind on 80 turns, keeping the wire tight and the turns close together but not overlapping. After your 80th turn, make another two holes and anchor the wire in the same way as before. Again, leave about 100 mm free after anchoring. Using another piece of enamelled wire (with 100 mm ends as before), loop one end through the same two holes which contain the end anchor of the last winding, wind two turns and anchor the end of this short winding using another pair of holes. Figure 2 shows the layout.

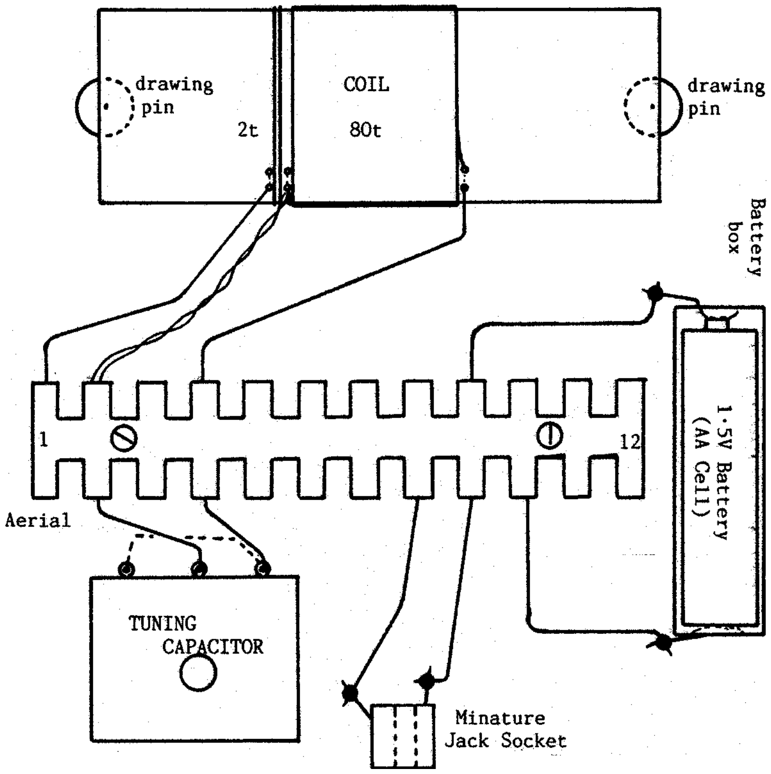

Figure 2 The layout of the parts on the wooden base

With some glass paper, remove the enamel from the ends of both pieces of wire which go through the same holes (i.e. the bottom of the large coil and the top of the small coil), then twist these bare ends together. Remove the enamel from the remaining ends of the coil. The coil is now finished!

The baseboard can be any piece of wood about 150 mm square. Fix the coil near the back edge using drawing pins and connect the wires from the coils to the terminal strip as shown in Figure 2. Using short pieces of PVC-insulated wire (and with assistance if you have never soldered before), solder one piece across the two outer tags of the variable capacitor, shown by the dotted line in Figure 2, and then two longer pieces to the centre tag and one outside tag. Connect these to the terminal strip. Then solder two more insulated wires on to the jack socket (into which you will plug your crystal earpiece), the other ends going to the terminal strip. The last two wires (one must be red) need to be soldered on to the battery box, their other ends going to the terminal strip also. Make sure the red wire goes to the *positive* terminal on the battery, and is connected to terminal 9. The other connection to the battery goes to terminal 10.

Attach the terminal strip to the baseboard with small screws or double-sided sticky tape. The other parts can be mounted the same way.

Listening is done ideally with the recommended crystal earpiece. Don't be tempted to use your Walkman earpieces; they are not the same and will not perform anything like as well. The receiver should work without an extra aerial, but one can be attached to terminal 1 if necessary. A long piece of wire mounted as high as possible is ideal. The *Audio-frequency Amplifier* project will enable you to use a loudspeaker with your receiver, using the signal from the jack socket. No circuit modifications will be needed!

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1, R5, R6	10 kilohms (kΩ)
R2	560 ohms
R3	270 ohms
R4	10 kilohms (kΩ)

Capacitors

C1, C2, C3	100 nanofarads (nF)
	500 picofarads (pF)

Semiconductors

ZN414Z, BC184

Additional items
- 12-way 2 A terminal strip
- 22 metres of 28 SWG enamelled copper wire
- A few short pieces of coloured PVC-insulated wire
- Crystal earpiece
- 3.5 mm jack socket
- 1.5 V AA-size battery and box
- Toilet roll tube
- Double-sided sticky tape or selection of screws

Tools required

Small screwdriver, soldering iron.

2 An audio-frequency amplifier

Introduction

This simple amplifier can be built by anyone who is able to solder reasonably well. It doesn't require any setting up and, provided our instructions are followed exactly, will work very well. The circuit diagram is included for the benefit of our more advanced readers, but it is not needed in the construction process. Please practise your soldering before you start, and don't use a printed circuit board (PCB) until you are confident that your soldering is up to scratch.

The amplifier can be used with other projects; it will provide plenty of sound from the MW Radio or from the Morse Sounder projects. It will usually be built into other pieces of equipment, so a box is not supplied with the kit. There is no reason why it shouldn't be put into a box and used as a general-purpose amplifier to help test other projects.

The components

Before you start, you should check that you have *all* the components to hand. A list and some helpful hints are given below.

1. PCB. The plain side is the *component side* and the soldered side is the *track side*. **Figure 1** shows the track side full size. Make the PCB from the pattern given in Figure 1. Otherwise, build the circuit on a matrix board.
2. Three resistors. Locate the gold or silver band around the resistor, and turn the resistor until this band is to the right. There are three coloured bands at the left-hand end of the resistor. Find the resistor whose colours are YELLOW, VIOLET, RED, and look at the resistor colour code chart which you will find in Chapter 7. From this, you will see that YELLOW indicates the value 4, VIOLET the value 7, and RED the value 2. The first two colours represent real numbers, and the last value is the number of zeros (noughts) which go *after* the two numbers. So, the value is 47 with two zeros, i.e. 4700 ohms. In this way, the resistor coloured BROWN, GREY, BROWN has a value of 180 ohms, and the last one, BROWN, RED, GREEN, has a value of 1 200 000 ohms. The ohm (often written as the Greek letter omega (Ω)) is the unit of resistance. If you do not yet feel confident in identifying resistors by their colours, use the Resistor Colour Codes given in Chapter 7.
3. Four capacitors. The two small 'beads' are tantalum capacitors and will be marked 4.7 μF or 4μ7, with a '+' above one lead. A tubular capacitor with wires coming from each end should be marked 220 μF, with one end marked '+' or '–'. This is called an *axial* capacitor because the wires lie on the axis of the cylinder. This is in contrast to the final capacitor, where both wires emerge from the same end. This is a *radial* capacitor, and will be marked 47 μF. Again, one lead will be marked '+' or '–'. Capacitors marked like this are said to be *polarised*, and it is vital that these are placed on the PCB the right way round, so take notice of those signs!
4. Two diodes. These are tiny glass cylinders with a band around one end, and may be marked 1N4148; this is their type number. Like polarised capacitors, they *must* be put on the PCB the correct way round!

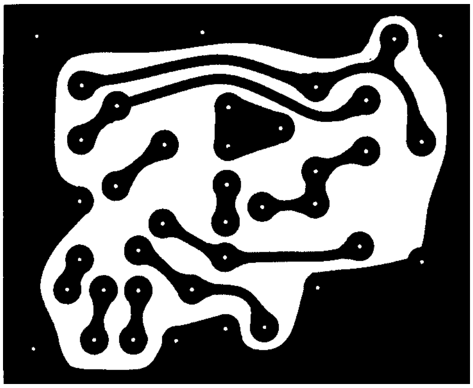

Figure 1 The toil pattern of the PCB – looking from the track side

5. Three transistors. One should be a BC548 (or a BC182), the other two should be BC558 (or BC212).
6. One volume control with internal switch.
7. One loudspeaker. This is quite fragile – don't let anything press against the cone.
8. One PP3 battery clip with red and black leads.

Putting it together

Lay the PCB on a flat, clean surface with the track side downwards. It is always useful to compare the layout with the circuit diagram, given here in **Figure 3**. Although you can't see it, the D-i-Y Radio sign should be at the top. Compare the hole positions with those shown in **Figure 2**. Bend the resistor wires at right angles to their bodies so that they fit cleanly into the holes in the PCB. Push each resistor towards the board so that it lies flat on the board. Then supporting each one, turn the board over and splay out the wires *just* enough to prevent the resistor falling out. Then, solder each wire to its pad on the PCB, and cut off the excess wire. When you have more confidence, you can cut of the excess wire *before* soldering; it often makes a tidier joint.

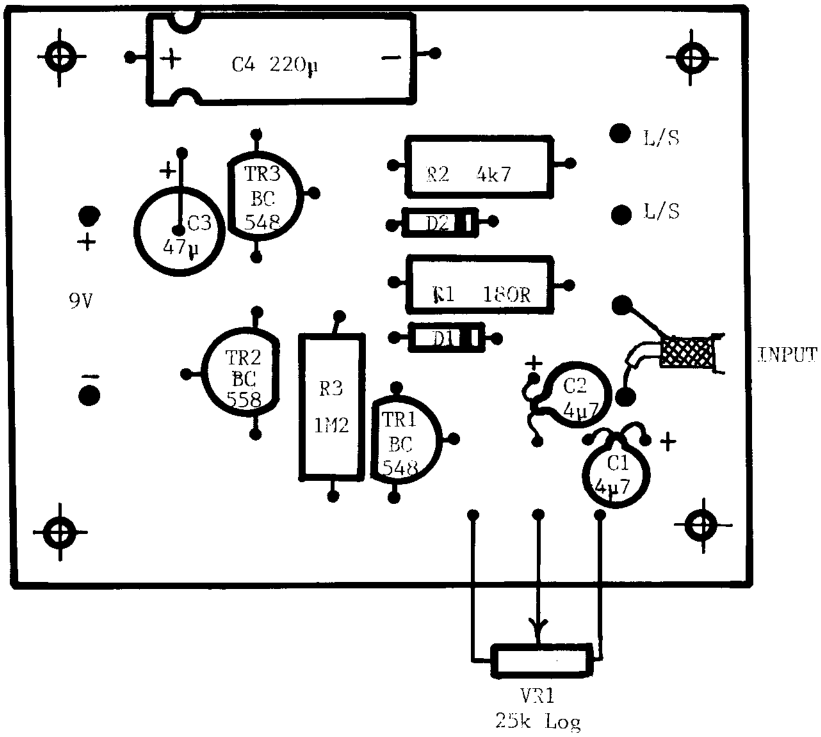

Figure 2 Positions of the components on the printed circuit board (PCB)

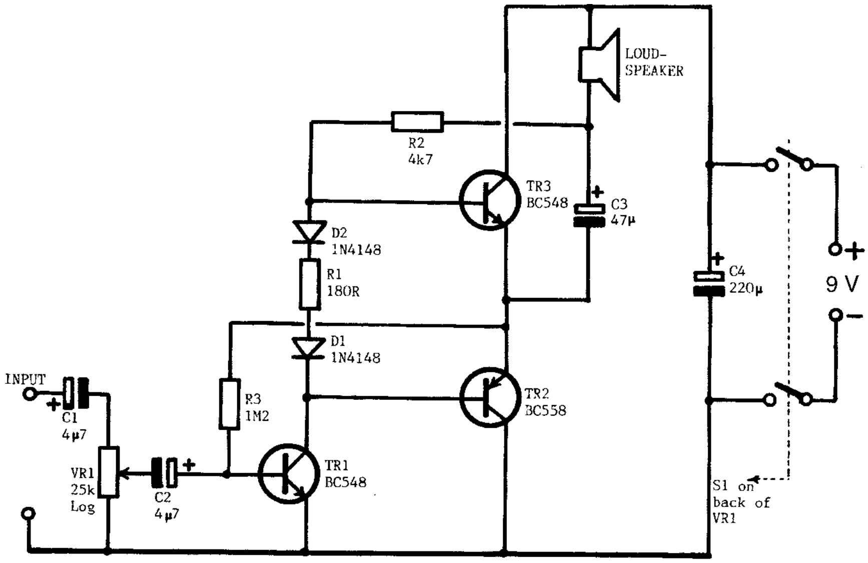
LOUD-
SPEAKER
R2
4k7
TR3
BC548
C3
47µ
D2
1N4148
R1
180R
C4
220µ
+
9 V
−
D1
1N4148
INPUT
C1
4µ7
R3
1M2
TR2
BC558
VR1
25k
Log
C2
4µ7
TR1
BC548
S1 on
back of
VR1

Figure 3 The amplifier's circuit diagram

Now fit the four capacitors. Each must be connected the right way round, so look at each component, match it up with the diagram of Figure 2, bend its wires carefully and repeat the soldering process you performed with the resistors, making sure that the components are close to the board and not up on stilts! Fit the two diodes the correct way round, and solder then as quickly as you can – they don't like to be fried!

Mount the transistors about 5 mm above the PCB. Make sure the correct transistors are in the correct places, and that the flats on the bodies match up with those shown in Figure 2.

Mount the volume control so that the spindle comes out from the front of the board. Use a piece of red insulated wire to the pad marked + on the PCB, and a black piece to the pad marked –, and solder these to the tags on the back of the control, as shown in **Figure 4**. Connect the two leads from the battery clip to the other tags on the switch; Figure 4 will help you. Finally, use two pieces of insulated wire about 100 mm long, twisted together, to connect the loudspeaker to the PCB.

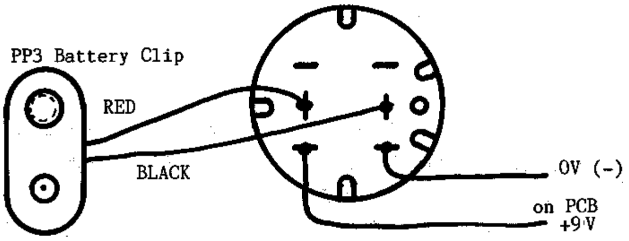
PP3 Battery Clip
RED
BLACK
−
−
+
+
0V (-)
on PCB
+9V

Figure 4 Connections to switch on back of VR1

Box clever!

If you wish to put the amplifier into a box, there is no problem; almost any box that is big enough will do. All that is needed is one hole big enough to accept the bush of the volume control; the PCB will be supported by the volume control. The prototype was not fitted into a box, but mounted on an odd piece of aluminium, bent into an L-shape and screwed on to a wooden base. The loudspeaker was mounted on the aluminium panel by two small pieces of aluminium with 3 mm holes drilled in them, which acted as clips around the edge of the speaker. Drill a few holes in the panel in the position of the speaker to let the sound get out!

Your input signal can be connected to the amplifier with two short pieces of wire, but if the connection needs to be long, use screened cable, with the braid connected as shown in Figure 2.

If you decide to use a different loudspeaker, make sure that its impedance (the resistance value marked on the back of the magnet) is at least 35 ohms. Anything lower may damage TR2 and TR3, and will certainly run down your battery very quickly. You will be surprised at the uses you can find for this little amplifier!

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1	180 ohms (Ω)
R2	4.7 kilohms (kΩ)
R3	1.2 megohms (MΩ)
VR1	25 kilohms (kΩ) log with DPST switch

Capacitors: all rated at 25 V minimum

C1, C2	4.7 microfarads (μF)
C3	47 microfarads (μF)
C4	220 microfarads (μF)

Semiconductors

TR1, TR3	BC548 npn
TR2	BC558 pnp
D1, D2	1N4148

Additional items

PCB
Speaker >35 ohms
PP3 battery clip and battery

3 A medium-wave receiver using a ferrite-rod aerial

Introduction

This design came from the Norfolk Amateur Radio Club, and enables you to build a simple Amplitude Modulation (AM) receiver for frequencies between 600 kHz and 1600 kHz. It should take you around 2 hours to build, and can be used with Walkman-type earpieces. **Figure 1** shows the circuit diagram.

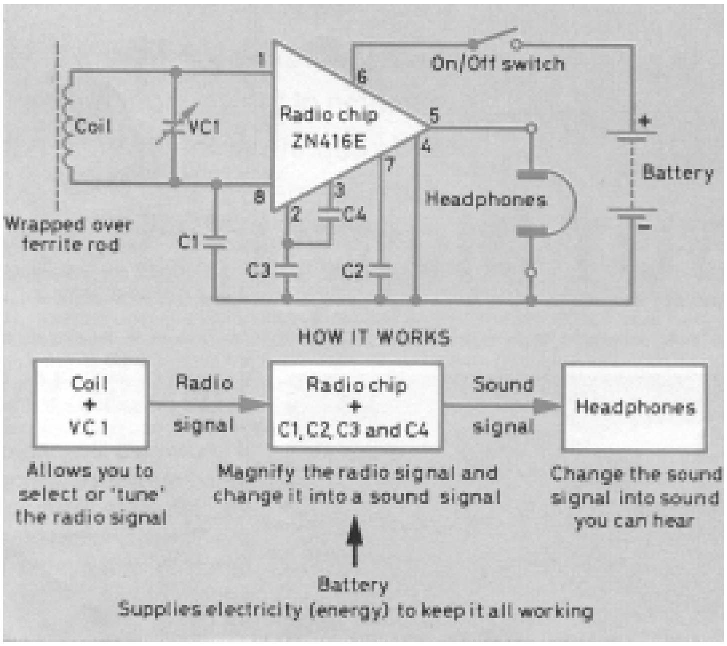
On/Off switch
Radio chip
ZN416E
Coil
VC1
1
2
3
4
5
6
7
8
C1
C2
C3
C4
Battery
+
−
Headphones
Wrapped over ferrite rod
HOW IT WORKS
Coil + VC 1
Radio signal
Radio chip + C1, C2, C3 and C4
Sound signal
Headphones
Allows you to select or 'tune' the radio signal
Magnify the radio signal and change it into a sound signal
Change the sound signal into sound you can hear
Battery
Supplies electricity (energy) to keep it all working

Figure 1 Circuit and block diagrams of the radio

Description

The whole circuit is built on a 50 mm by 50 mm printed circuit board (PCB) designed to fit on the inside of the lid of a plastic box, and is stuck there using sticky pads, the shaft of the variable capacitor going through a hole in

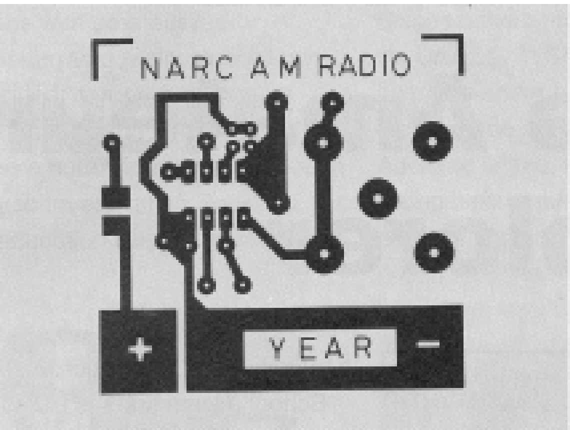

Figure 2a The PCB, solder side

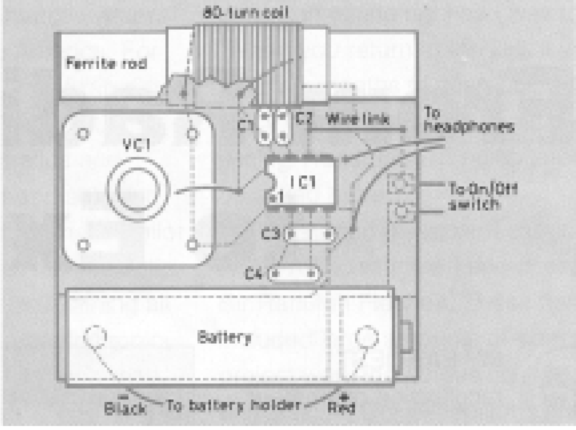

Figure 2b The PCB, component side

the lid. Only two pairs of leads are soldered to the board – one pair goes to the 1.5 V battery in its holder, and the other to the earphone socket. **Figures 2a** and **2b** show the printed circuit and the component layout double size for clarity. You are not obliged to build the circuit on a PCB.

Building it

1. **Check and identify components**. Tick the parts list.
2. **Carefully unwind the wire.** Use paper to make an insulating tube (called a 'former') around the centre of the ferrite rod and secure it with Sellotape. Now, close-wind *all* the wire (leave no gaps between adjacent turns) around the paper former. Secure the winding with more Sellotape, leaving 50 mm of wire free at each end for connection to the circuit. See **Figure 3a.**
3. **Solder in VC1.**
4. **Solder in the integrated circuit holder.** There is a notch in one end of the holder; this should face VC1. Solder also the wire link and the capacitors. Be careful to avoid solder 'bridges' between adjacent tracks on the PCB.
5. **Solder the battery leads.** These must be connected properly – the red battery lead to the + (positive) area and the black lead to the – (negative) area.
6. **Strip bare 1 cm of insulation from the ends of two wires**. Solder them between the PCB and the headphone socket (see **Figure 3b**). Use the end tabs on the socket. Using another pair of insulated wires connect the ON/OFF switch to the PCB tabs shown in Figure 2b.

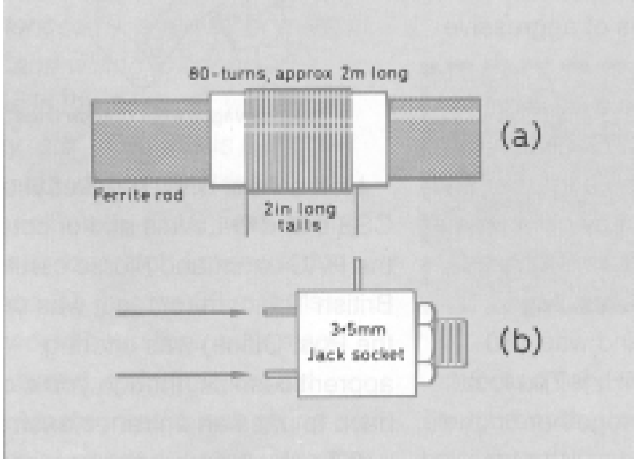

Figure 3 Details of coil and headphone socket

7. **Fix the elastic band.** This goes through the holes at the top of the PCB, with the ferrite rod being slipped through the two end loops. (Note: although the coating on the copper wire is designed to melt away during soldering, it is quite common for difficulty to be experienced in obtaining a good soldered joint; to be on the safe side, remove the coating *before* soldering (with a small piece of sandpaper).) Carefully place the wire ends of the coil through the PCB just above VC1, and solder on the track side.
8. **Fit IC1 into its holder.** This should be done carefully, making sure that *all* the pins are located above their respective clips *before* applying any pressure! Make sure also that the notch on the IC (as shown in Figure 2b) matches the notch in the holder, and faces VC1.
9. **Put battery in its holder.** Listen for some noise in the headphones as VC1 is rotated. Make sure the headphone plug is fully inserted into its socket.
10. **Fix the working board to the lid.** Use the sticky pads and apply *gentle* pressure. Fit the tuning knob, the ON/OFF switch and the earphone socket.
11. **Test again.** If all is still working, fit the lid screws and admire your completed radio!

Parts list

Capacitors

C1, C2	0.01 microfarad (μF)
C3, C4	0.1 microfarad (μF)
VC1	500 picofarads (pF)

Semiconductor

IC1	ZN416E

Additional items
- Plastic box (recommended size 76 × 64 × 50 mm internal)
- 8-pin DIL socket for IC1
- Printed circuit board
- Tuning knob for VC1
- Wire link for PCB
- 2 m of 30 SWG copper wire, self-fluxing
- Piece of paper 25 × 50 mm, to make the coil former
- Ferrite rod 70 mm long by 10 mm diameter, approximately
- Battery, AA size 1.5 V, with holder and attached wires
- Miniature earphone socket (3.5 mm stereo jack)
- ON/OFF switch (push-button SPST latched or slide switch)
- 4 off 100 mm insulated connecting wires, for jack socket and ON/OFF switch
- Pair Walkman-type earphones
- Elastic band, to attach ferrite rod to PCB
- 4 off sticky pads for securing PCB to box lid

Kits

Ready-made PCBs may be available from Alan J. Wright, G0KRU, Hewett School, Cecil Road, Norwich NR1 2PL.

4 A simple electronic organ

Introduction

This project has nothing to do with radio but, let's admit it, *any* electronics project is good experience! Why not build this little organ – it will keep the children amused at least! It uses the popular NE555 integrated circuit, which contains a circuit which will periodically switch the voltage on the output pin between the supply voltage and zero. Just how frequently this switching occurs depends upon the components external to the integrated circuit. If this switching occurs several hundred or thousand times a second, the change in voltage produced will generate a musical note when connected to a small loudspeaker. The circuit is shown in **Figure 1**.

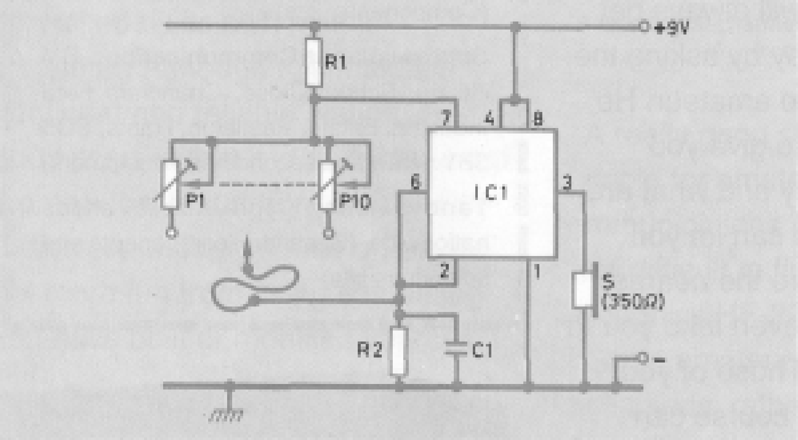
+9V
R1
7
4
8
P1
P10
6
IC1
3
2
1
S
(350Ω)
R2
C1
-

Figure 1 Circuit diagram

Putting it together

(a) Using a PCB. The job is very simple. The placement of components on the unsoldered side of the board is shown in **Figure 2** and the design on the copper track is illustrated in **Figure 3**. Put each component, in turn, on the board, making sure that it lies flat on the board with its tags or wires going cleanly through the holes provided for it; then, solder the wires to the board, cropping them before or after the soldering,

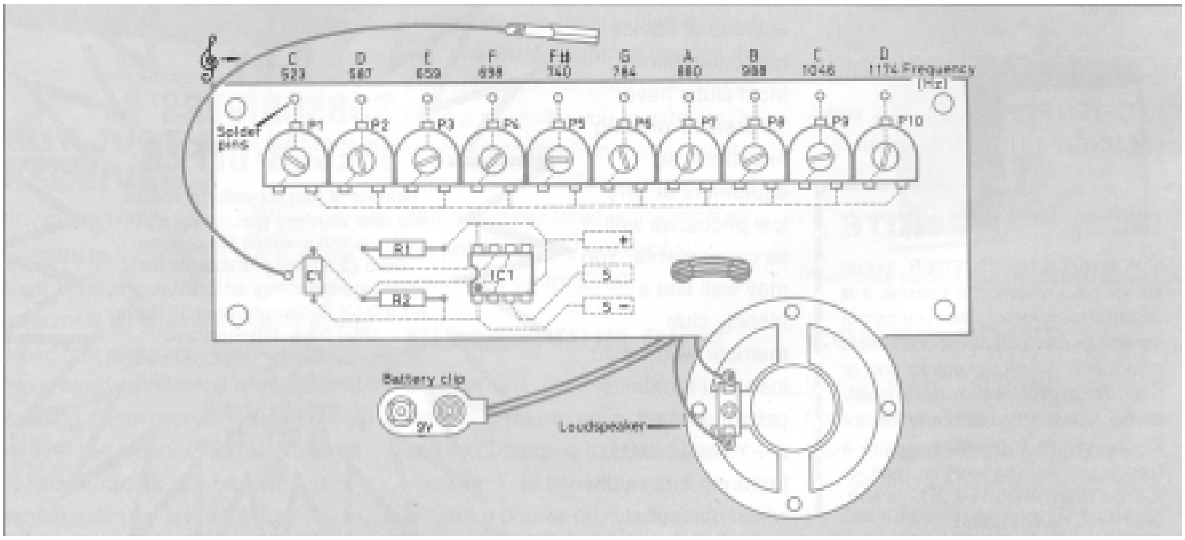
C 523
D 587
E 659
F 698
F# 740
G 784
A 880
B 988
C 1046
D 1174 Frequency (Hz)
Solder pins
P1
P2
P3
P4
P5
P6
P7
P8
P9
P10
R1
R2
IC1
C1
+
S
S -
Battery clip
9V
Loudspeaker

Figure 2 Position of components on the printed circuit board (PCB)

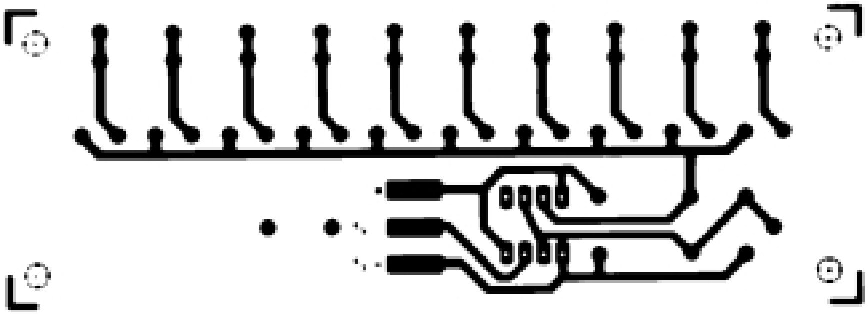

Figure 3 The connections

depending on your preference. If you choose to use a holder for your integrated circuit (highly recommended if your soldering is less than perfect), make sure that the end with a notch in it faces R1 and R2, as shown in Figure 2. Solder the two leads to the speaker to the tabs marked S (either way round), having looped them through the two holes to the right of the tabs in Figure 2. Looping them through the holes acts as a strain relief, ensuring that the soldered joints are not subjected to pulling and bending as you move the wires about. Do the same with the battery leads, the red lead going to the + tab and the negative lead to the – tab (which also has one speaker lead already attached to it). **Figure 4** shows this in detail. Treat the loudspeaker with care – the cone is quite fragile and must not be touched.

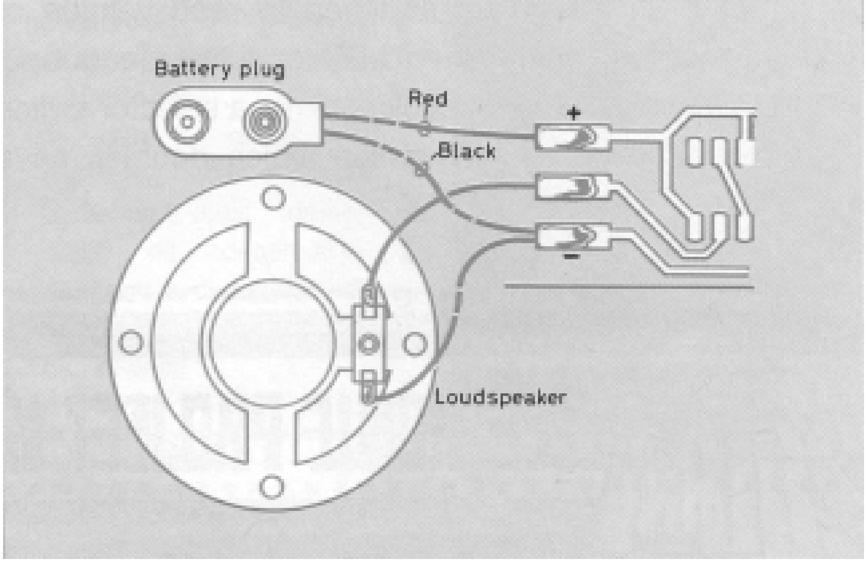

Figure 4 Battery plug and loudspeaker connections

(b) Without a PCB. This is more difficult, and you may need to enlist some help. Using some matrix board (such as Veroboard) is probably the best way of replacing the PCB. You could arrange your circuit in exactly the same way as in the PCB in Figure 3, using wires to replace the copper track.

A simple ‘keyboard’

The keyboard is a row of solder pins along the rear edge of the PCB, one for each note covering the range shown in Figure 2. A flying lead with a small spade on it is provided to touch any of the pins in turn, producing any one of ten different notes.

Testing

Check first that each component is in the correct place. When inserting the NE555 chip, first make sure that the end carrying the notch lies over the end of the holder with the notch; then, make sure each pin of the chip lies directly above the hole into which it fits, before pressing *gently* to insert the chip into the socket. Make sure the battery connections are correct, and insert the battery into the clip. Nothing should happen, except for a click from the loudspeaker; touching the spade on any of the pins should produce a coarse note from the speaker. If nothing happens, check everything again; don’t *assume* that wires go where you think they go!

After you get the first note, all the others should work, too, but they will sound off-tune at first. The organ needs tuning up by adjusting the 10 preset variable resistors P1 to P10. The approximate frequency to which each note should be tuned is given in Figure 2; if you can beg, borrow or steal a frequency counter, setting up is easy. If you have a piano, the organ can be tuned by comparison of the notes with those on the piano. The frequencies are given in Hertz (abbreviation Hz), and represent the number of times the IC switches on and off every second. If the sound coming from the loudspeaker is too loud or very distorted, then try putting an 330 Ω resistor (colour code orange, orange, brown) in series with the loudspeaker. This is done by taking the resistor and cutting its leads to about 5 mm; then, disconnect one speaker lead from the tab on the PCB (it doesn’t matter which). Solder one end of the resistor to the vacated speaker tab, and the free speaker lead to the other end of the resistor. This will limit the volume of sound from the speaker, and lengthen the life of your battery. If it is still too loud, try a resistor of a larger value, or use a smaller resistor to make it louder.

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1	2.7 kilohms (kΩ)
R2	1 Megohm (MΩ)
P1–P5	Preset resistor 100 kilohms (kΩ)
P6, P7	Preset resistor 50 kilohms (kΩ)
P8	Preset resistor 25 kilohms (kΩ)
P9, P10	Preset resistor 10 kilohms (kΩ)

Capacitor

C1	100 nanofarads (nF) or 0.1 microfarad (μF)

Integrated circuit

IC1	NE555 timer chip

Additional items

S	Loudspeaker >60 ohms (Maplin)
	1 off battery clip (for PP3 battery)
	1 off spade terminal
	12 off solder pins ‘Veropins’
	3 off 10 cm lengths of ‘hook-up’ wire

(This article is based on projects originally designed by Radio Scouting, Netherlands.)

5 Experiments with the NE555 timer

Introduction

Several of the projects in this book use the NE555 timer, an integrated circuit which is at the heart of many circuits whose processes are determined by time intervals. **Figure 1** shows the circuit diagram of an audio oscillator using the 555. The timing voltages (governing the frequency of oscillation) are produced by R1, R2 and C2; a voltage appears at pin 3 which 'switches' at this frequency between zero and a voltage close to the supply voltage, which in this case can be anywhere between 6 V and 14 V. The output current, when applied through R3 to a small loudspeaker, produces an audible tone, provided that there is a DC path between the two test leads.

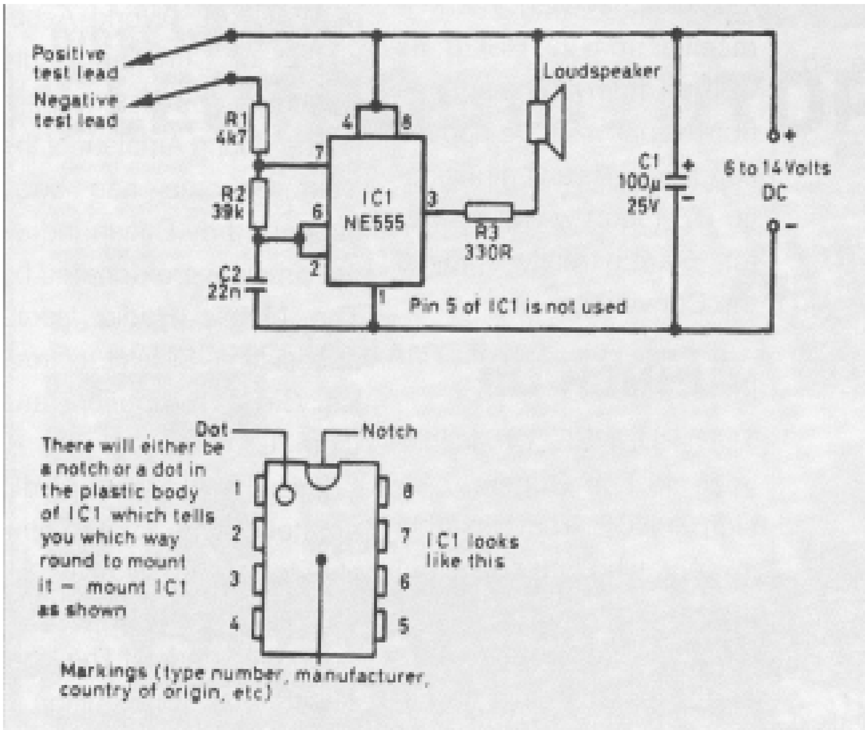
Positive test lead
Negative test lead
R1 4k7
R2 39k
C2 22n
4
8
7
6
2
1
3
IC1 NE555
R3 330R
Loudspeaker
C1 100µ 25V
+
−
6 to 14 Volts DC
Pin 5 of IC1 is not used
Dot
Notch
There will either be a notch or a dot in the plastic body of IC1 which tells you which way round to mount it – mount IC1 as shown
1
2
3
4
8
7
6
5
IC1 looks like this
Markings (type number, manufacturer, country of origin, etc)

Figure 1 The circuit diagram

Construction

The simplest way to mount the components is on a piece of matrix board (Veroboard), available from any of the good suppliers. The prototype of this circuit used the type of board with copper strips along the underside; these strips are used like the copper tracks on a PCB, to join components together. Firstly, cut the four strips between the positions of the pins of the IC socket, as shown in **Figure 2**. You can buy a tool for this purpose, but a small twist drill (about 3 mm diameter) is just as good. Turn it between your fingers – if you use a drill you will end up with holes right through the board! Then solder in the IC socket (with the notch in the position shown), followed by the four links made with single-conductor insulated wire. Put in each component as shown, ensuring that C1 (an *electrolytic* or *polarised* capacitor) is connected correctly. When all the components have been soldered in, take the 555 chip and lay it on its socket, with its own notch lying above that of the holder. Then, making sure that each pin lies *directly* above its corresponding socket, press down *gently* on the chip, with the board supported on a flat surface.

Testing

Connect the circuit to a battery or small power supply, ensuring that the positive and negative leads are the right way round. Always use red and black leads here, then you are less likely to get it wrong! Switch on. Nothing

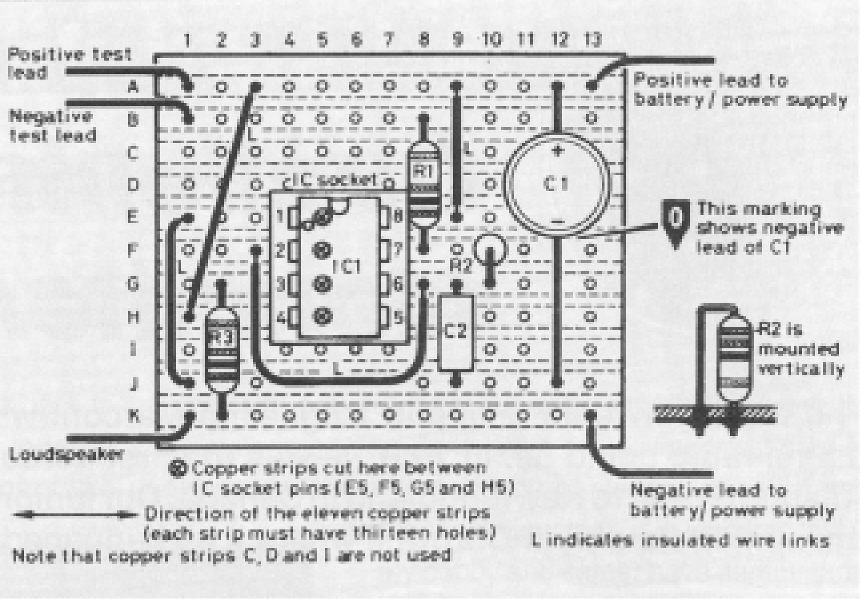
1 2 3 4 5 6 7 8 9 10 11 12 13
A
B
C
D
E
F
G
H
I
J
K
Positive test lead
Negative test lead
Loudspeaker
Positive lead to battery / power supply
Negative lead to battery/ power supply
IC socket
IC1
1
2
3
4
5
6
7
8
R1
R2
R3
C1
C2
+
−
L
This marking shows negative lead of C1
R2 is mounted vertically

⊗ Copper strips cut here between IC socket pins (E5, F5, G5 and H5)

←——→ Direction of the eleven copper strips (each strip must have thirteen holes)

Note that copper strips C, D and I are not used

L indicates insulated wire links

Figure 2 Veroboard layout. If you can read a circuit diagram, the project can be built using other methods

should happen until you short together the two test leads, when there should be a note from the loudspeaker. If this doesn't happen, switch off, disconnect the circuit and check your wiring and soldering. Is it exactly like Figure 2? Are the soldered joints round and shiny? If any are dull, then 'sweat' them briefly with a hot soldering iron until the solder runs, remove the iron, and check that they are as shiny as the rest. Check that there are no solder 'bridges' between adjacent tracks by holding your board up to a strong light. Then, reconnect, switch on and touch the test leads together. All should now work!

Uses of your circuit

1. As a Morse practice oscillator. Simply connect the two test leads to your key and, each time the key is pressed, you should hear a note from the speaker. The frequency of the note may be altered by putting a resistor in series with the key. To do this, remove one test lead from the key and select a resistor; connect one end of the resistor to the free test lead and the other to the empty terminal on your key. Selecting the value of resistor that you need will be a useful experiment in itself.
2. As a continuity tester. You can check fuses and lamp bulbs by connecting them across your test leads. If the speaker remains silent, the fuse or bulb has blown.
3. To indicate changes of resistance. Hold the ends of the test leads in each hand; you should hear a low note, because of the high resistance of your body. Squeeze the ends harder, and the frequency of the note should rise, because you are now making better contact. Repeat this with damp hands and the frequencies will be higher still.
4. As a thermometer. Connect the test leads to a thermistor (a device whose resistance changes with temperature) and warm it with a hair-dryer, or even in your hands, and you will hear the pitch changing with the temperature of the thermistor. A suitable 'bead' thermistor is available from Maplin (order code FX21).
5. As a diode tester. Use any diode, and connect the negative test lead to the end of the diode marked with the ring. This is the cathode of the diode. The other end, the anode, should be connected to the positive test lead, and a note should be heard from the speaker. This does not necessarily mean that the diode is working – yet. Reverse the connections and *nothing* should be heard. If this is the case, the diode *is* working.
6. As a light meter. Use a photoconductive cell (a device whose resistance changes with light intensity) connected between the test leads. A note should be heard. Shading the device with your hand will increase its resistance and the note should decrease in frequency. A suitable device is the ORP12 cell from Maplin (order code HB10).

There are many more applications. **Do not connect the test leads to other circuits that are switched on.** Your circuit, or the circuit you are connecting it to could be damaged. Think of a *passive* device or circuit (i.e. one *not* requiring a power supply or battery) where changes of resistance occur, and you have found another application!

Parts list

Resistors: all 0.25 watt carbon film types

R1	4.7 kilohms (kΩ) – yellow, violet, red
R2	39 kilohms (kΩ) – orange, white, orange
R3	330 ohms (Ω) – orange, orange, brown

Capacitors

C1	100 microfarads (μF) 25 V radial electrolytic
C2	22 nanofarads (nF) or 0.022 microfarad (μF) polyester type with 10 mm lead spacing

Integrated circuit

IC1	NE555

Additional items

- Miniature loudspeaker (preferably 35, 40 or 80 ohm)
- 8-pin DIL socket for IC1
- 0.1 inch Veroboard ('stripboard'), size 11 strips by 13 holes
- PVC-covered stranded wire for test leads, loudspeaker and battery connections
- PVC-covered solid wire for links on the board
- A power source of between 6 V and 14 V, such as a 9 V battery (PP3)

Component sources

Cirkit
Tandy – many high street shops

6 A simple metronome

Introduction

A metronome is a device used by musicians to indicate the tempo of a piece of music. Until electronics came on the scene, this 'beating of time' was achieved in much the same way as a clock keeps time, i.e. with a pendulum device, the clicking of the escapement indicating the beats of the music.

Those of you who have already built the Morse Key and Buzzer from the designs in this book, will recognise the circuit of this metronome – it is exactly the same as was used to produce the note of the buzzer. This circuit is shown in **Figure 1**.

The circuit

Three components determine the speed at which the circuit oscillates – the speaker (LS), the resistors (VR1 + R1) and the capacitor (C1). VR1 is a variable resistor, so that the speed at which the oscillator operates can be varied. Compared with the component values of the Morse Buzzer (which operated at around 800 Hz), these components now give an oscillation frequency of around 1.25 Hz, which is far too low to be heard as a note. What we *do* hear, however, is a series of clicks, as the voltage across the speaker changes quickly from 0 to 9 V and back again.

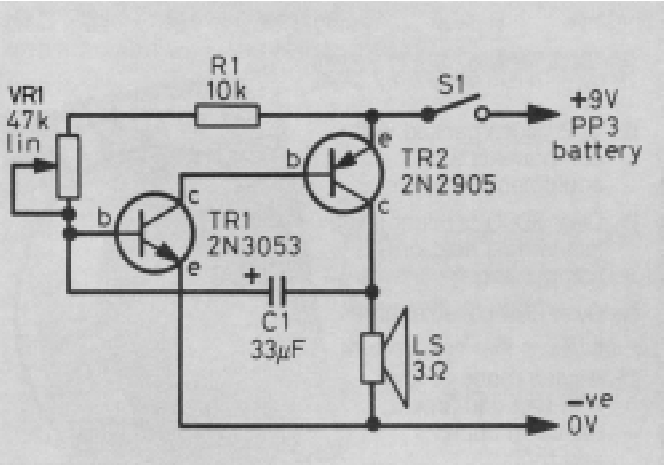

Figure 1 The metronome circuit is rather like the Morse oscillator

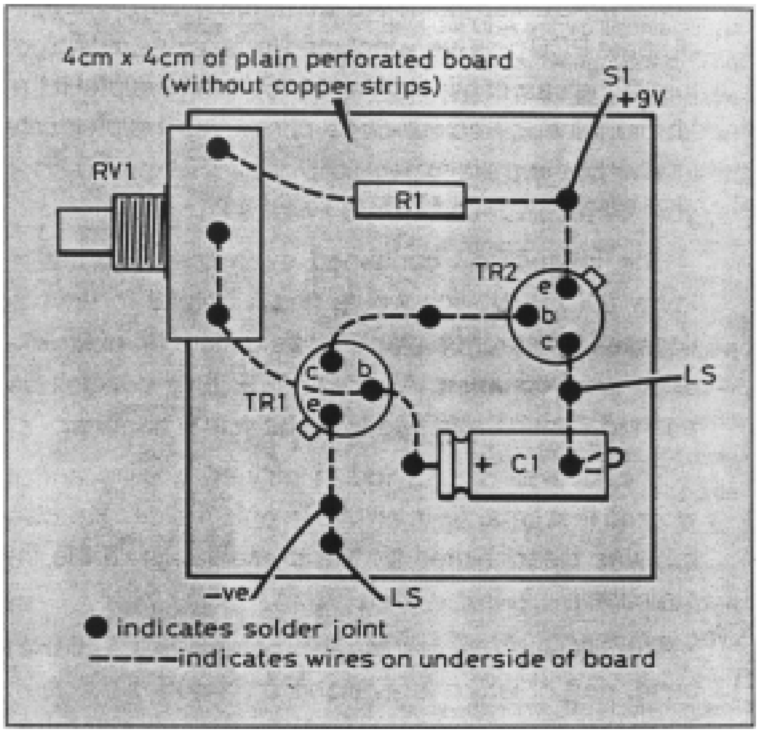
4cm x 4cm of plain perforated board
(without copper strips)
S1
+9V
RV1
R1
TR2
e
b
c
LS
c
b
TR1
e
+ C1
–ve
LS
● indicates solder joint
-----indicates wires on underside of board

Figure 2 The component wires are pushed through holes in the circuit board and joined together underneath

Variation of speed could be achieved by varying resistance or capacitance. However, as you may already know, variable capacitors have values in the picofarad range, not the tens of microfarads used here, so it is very simple to employ a variable resistor (potentiometer) to control the oscillator. You could use a multi-way switch to switch in one of several capacitors, as well as having the variable resistor, but this was found to be an unnecessary complication. This design operates between about 100 clicks per minute and 200 clicks per minute.

Making the prototype

A single piece of plain matrix board (no copper strips) measuring about 40 × 40 mm is sufficient to hold all the components except the potentiometer and switch (see later). The case can be plastic or aluminium, and one measuring 65 × 100 × 50 mm is about right. Make sure there are holes in the case beside the speaker cone to let the sound out, and larger holes for the potentiometer and switch. If a potentiometer is used with a combined ON/OFF switch, then the extra hole for the switch is not necessary! It is advisable to construct the circuit *before* putting it in the box, so that it can be tested to ensure that everything is working. If it is, then you can exercise your ingenuity in mounting the speaker, battery and board inside the box. A final test can be made before starting the calibration process.

Calibration

There is no ‘easy’ way to do this. The frequencies involved are too low to be measured with the average frequency counter, so you will need to resort to using a stopwatch and counting the number of clicks per minute.

Parts list

Resistors: 0.25 watt, 5% tolerance

R1	10 kilohms (kΩ)	
VR1	47 kilohms (kΩ) linear potentiometer	

Capacitor

C1	33 microfarads (μF) electrolytic	

Transistors

TR1	2N3053	npn
TR2	2N2905	pnp

Additional items

S1	SPST	ON/OFF switch
LS	3 ohms (Ω)	loudspeaker
	Knob with pointer for VR1	
	PP3 battery and connector	
	Aluminium case, 65 × 100 × 50 mm	
	Matrix board (plain), 40 × 40 mm	

7 What is a resistor?

Introduction

Materials that carry electricity easily are called **conductors**. They include all metals and salt water, for example. We use wire as a conductor, and the ease with which it passes an electric current depends upon the material, its thickness and its length. Silver (symbol Ag), gold (Au), copper (Cu) and aluminium (Al) are the best metallic conductors. Most wires are made of copper, although the best conductor, weight for weight, is aluminium.

Materials that *don't* carry current (or, at least, do so very badly) are called **insulators**, and they include dry wood, rubber, plastic and glass among their number. Wires are often coated with a layer of insulator to prevent adjacent wires touching and causing an accident.

Resistors

If there wasn't such a thing as resistance, the subject of electronics wouldn't exist; only infinite currents would flow and voltages wouldn't exist either! We need to reduce the flow of current if we are to make current do something useful for us. Components that resist the flow of current are called **resistors**, and they are said to have a **resistance** which is measured in ohms (Ω), named after Georg Ohm, who formulated the law (also named after him) by which the voltage and current through a conductor are related. His law gave rise to the formula *everyone* remembers:

$$I = \frac{V}{R},$$

where I is the current flowing, measured in amps,
V is the voltage across the conductor, and
R is the resistance of the conductor, measured in ohms.

From this equation, you can see that, for a constant value of voltage, V, if the resistance goes up, the current will go down, and vice versa. The circuit symbols for resistors are shown in **Figure 1**. You will find the upper symbol in older magazines; it is still preferred by many engineers. The lower symbol is the prevalent standard symbol.

Resistors are made in several ways, the cheapest using carbon; another type is usually made from a ceramic cylinder (used only as a support) on which

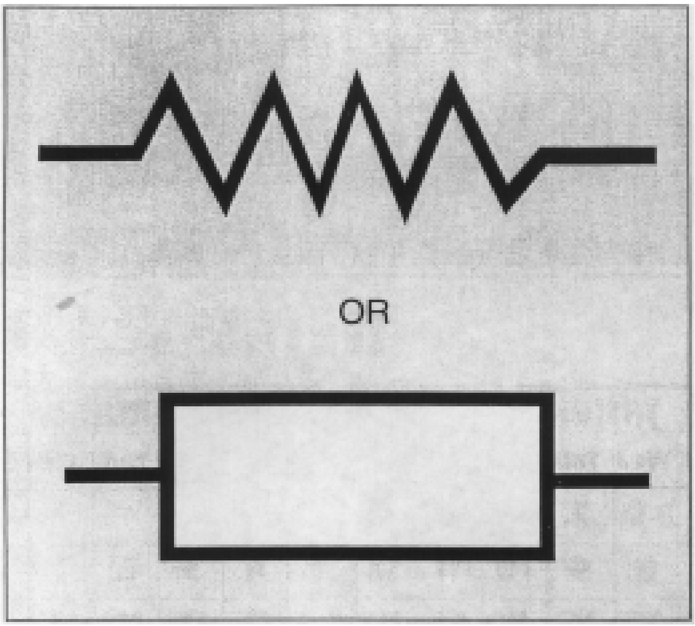

Figure 1 Circuit symbols for resistors

is placed a very thin film of metal – the thinner the film, the greater the resistance. All resistors are coated with a thin film of insulation, for the same reason we discussed earlier.

The colour code

Each resistor has coloured bands on it which enable us to see what value of resistance it has. There are normally three (but sometimes four) at one end, and a single one at the other (see **Figure 2**). The colours indicate figures, according to the list below.

Colour	*Value*	*Colour*	*Value*
Black	0	Green	5
Brown	1	Blue	6
Red	2	Violet	7
Orange	3	Grey	8
Yellow	4	White	9

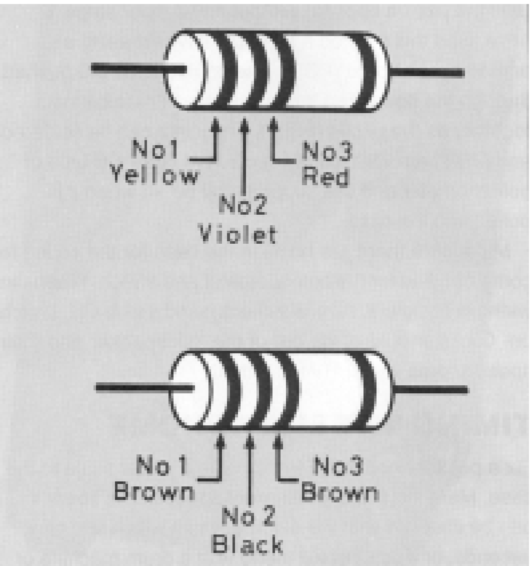

Figure 2 Some examples of resistor colour codes; top 4700 Ω (4.7 kΩ) and bottom 100 Ω

Using the colour codes is easy, once you see the logic behind it. Hold the resistor so that the single band is towards the right. The three colours on the left are read in the normal order from left to right. The first two bands *always* indicate numbers; the third band gives the number of zeros to add to the right of these two numbers. So, looking at the top resistor in Figure 2, yellow, violet, red means 4, 7, and two zeros, giving 4700 ohms! Looking at the lower resistor, brown, black, brown means 1, 0, and one zero, giving 100 ohms.

Remembering the order of the colours may be difficult at first. The colours from red to violet are the colours of the rainbow, in order, so if you know those, you're almost there! Around those colours are black and brown below the red, and grey and white above the violet, which you can imagine as getting brighter from black to white (well, almost!). It won't be long before you don't need to remember them at all.

The isolated band on the right-hand side is not part of the resistor's value; it indicates its *tolerance*, i.e. how close it might be to the indicated value. A brown band indicates ±1%, a red band ±2%, a gold band ±5% and a silver band ±10%. For example, a resistor marked as being 100 ohms with a ±5% tolerance will have an *actual* value somewhere between 95 ohms and 105 ohms.

8 Waves – Part 1

Introduction

Waves are responsible for most of the processes in life where energy is transferred from one place to another. Heat and light energy from the sun, for example, come to us as *electromagnetic waves*. Sound travels through the air as a wave; it is not the same sort of wave as light or heat, but it obeys many of the same properties. Damage is caused to coastal margins by the waves of the sea – again, another type of wave, but still obeying many of the same properties. Radio waves are of the same type as heat and light waves, as are gamma rays, X-rays, ultra-violet waves and infra-red waves. So, once we begin to understand what radio waves do, we are also learning about a huge chunk of physics at the same time! All these waves are part of the *electromagnetic spectrum*. The word ‘spectrum’ simply means a ‘range’, so what we have is a range of electromagnetic waves – that’s all!

Sensing things

Light waves are invisible, but our eyes can detect the *effect* they have on different materials because the waves produce an effect on the retina of the human eye which the brain can interpret. We cannot see heat waves either, but we can feel the *effect* they have on our skin. Gamma rays and X-rays are also invisible, but their detrimental *effects* on human tissue are well known. It is not surprising, then, that we cannot see radio waves. We cannot sense them, either, until we produce a device upon which they have an *effect*. That device is a **radio receiver**; it is able to process certain characteristics of radio waves, and make these characteristics audible by generating sound waves from the loudspeaker or headphones. Other characteristics of the same waves may be turned into light as a TV picture on a cathode-ray tube, or as a fax image on a sheet of paper.

Visible waves

Let’s start our description with some waves that we can actually see! When a small stone is thrown into a pond, we see circular water waves *radiating* from the point where the stone fell into the water, as **Figure 1** shows. (Notice that we use the word *radiating*, even with water waves; it is not a radio term, but one which describes any motion where the *radius* of a circle is increasing. In this case, it is the radius of the circular waves which is

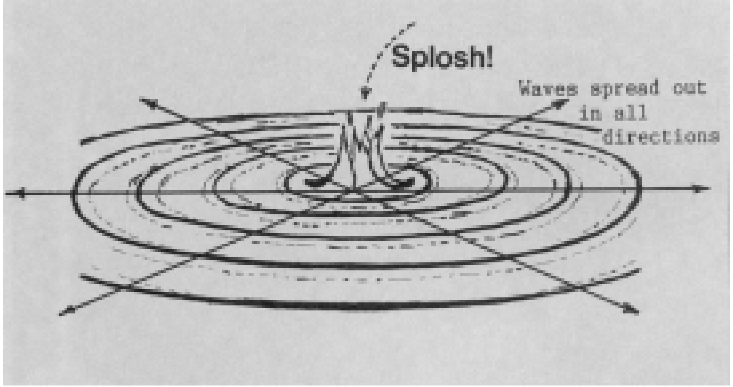
Splosh!
Waves spread out in all directions

Figure 1 The waves move out from the point where the stone landed

increasing.) If you were to watch the water waves down at the water level, perpendicular to the direction in which they are travelling, you would see something like the illustration in **Figure 2**.

The horizontal line represents the water level before the wave started, and the vertical line represents the direction in which the water is displaced at any instant.

All waves are described in the same way

'Freezing' the motion of the water in this way allows us to define two very important characteristics of a wave, characteristics which we talk about every day – *wavelength* and *amplitude*. The wavelength of a wave is simply the distance (measured in metres) from any point on one wave to the same point on the adjacent wave. Look at the diagram and you will see what is meant by 'the same point on the adjacent wave'. The amplitude of a wave

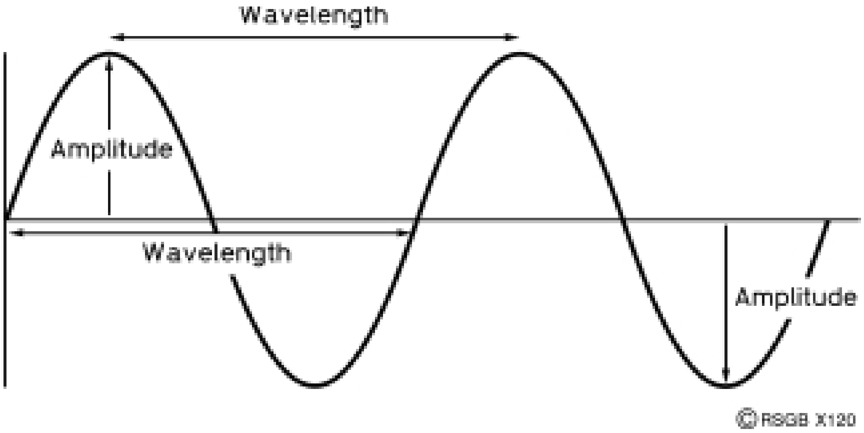

Figure 2 A water wave, viewed in cross-section

is always measured from the centre (undisturbed) position of a wave to the peak (or the trough) of the wave. Both these positions are shown, the arrow indicating that the measurement is taken *from* the centre *to* the peak or trough. The amplitude of a wave is defined as the maximum displacement of the wave from the centre position – the direction (up or down) of that displacement does not matter. Waves of greater amplitude carry more energy with them.

If we now ‘unfreeze’ the wave, we will see it travel from left to right (or right to left, depending on where we are looking). The speed at which it moves is called its *velocity*, and is measured in metres per second.

Another useful word is *propagate*; it means *travel*. We talk about radio waves propagating from a transmitter to a receiver. This velocity of propagation (for electromagnetic waves) is very fast indeed – they will cover 300 million metres in one second. This is virtually incomprehensible, so think of a radio wave travelling around the earth – it can travel 7½ times round the earth in one second! We use the symbol c for the velocity of radio waves (which is the same as the velocity of light, of course – all electromagnetic waves travel at this speed through air and space).

The last thing we need to know about the wave is its *frequency*. Imagine a cork floating on the water in the path of the wave; it will bob up and down. If we were able to count the number of times it went through its highest position in one second, then that number would be its frequency. Any periodic motion like this is said to go through one *cycle* each time one complete wave passes a point (in this case, our cork). We are thus counting the number of cycles per second of the cork’s motion. The unit of frequency is thus ‘cycles per second’; this unit is now named after Hertz, a radio pioneer, and is abbreviated to Hz.

Our description of the wave is now quite simple – we need only four quantities:

(a) **Frequency** symbol f – unit, hertz (Hz)
(b) **Wavelength** symbol λ (Greek letter lambda, pronounced ‘lamb-da’) – unit, metre (m)
(c) **Amplitude** symbol a – unit depends on application
(d) **Velocity** symbol c – unit, metres per second (m/s or $\mathrm{ms^{-1}}$)

The basic formula

Whatever may happen to a wave while it travels through different media (vacuum, air, brick, wood, etc.), one thing and **only** one thing remains constant – its frequency. Its wavelength, amplitude *and* velocity may change, but its frequency never does. Three of the four characteristics already identified are connected by the simple relationship

$$c = f \times \lambda.$$

Remember that c is constant if the wave travels in air or in a vacuum. This means that waves having higher frequencies (f large) must have smaller wavelengths (λ small) and vice versa. You can imagine frequency and wavelengths being on opposite ends of a see-saw!

Divisions of units

Because the frequencies of radio waves are so high (despite them having the lowest frequencies in the electromagnetic spectrum!) we have a problem with writing them down. Do you write in your log book that you have just heard a station on 14 100 000 Hz? Of course not, you write it as 14.1 MHz, knowing that the prefix mega (M) means ‘one million’. The prefixes which you need to know (when applied to frequency) are:

kHz	kilohertz	meaning	1000 Hz
MHz	megahertz	meaning	1 000 000 Hz
GHz	gigahertz	meaning	1 000 000 000 Hz.

Notice that the ‘k’ in kilohertz is a lower case letter. It is **incorrect** to write it as an upper case letter. ‘K’ is a computer-related prefix meaning **not** 1 000 but 1 024!

When we come on to discuss heat and light waves, we will use wavelengths rather than frequencies, because of the see-saw effect – as the frequencies get larger and larger, the wavelengths get smaller, and hence are numbers which are more manageable, both to talk about and to write down!

Bands

Gamma rays, X-rays, ultra-violet waves, light waves, infra-red waves are all part of the electromagnetic spectrum, but we divide them up because they have different properties. This is why we divide up our radio frequencies into different bands. The radio waves of top-band signals (around 2 MHz) have completely different properties compared with those in the 20 metre band, so we are dividing up the *radio spectrum* in the same way – by property.

9 A beat-frequency oscillator

Introduction

Many readers will know that, although they have a short-wave radio which covers at least one of the amateur bands (e.g. 7 MHz or 14 MHz), they are unable to listen to SSB or Morse signals. This is because the receiver lacks a Beat-Frequency Oscillator (BFO). We need the ‘carrier’ frequency of a BFO to replace the carrier that has been removed from the signal at the transmitter. When listening to Morse signals, the BFO signal ‘beats’ with the incoming signal to produce a note in the loudspeaker. If you are a musician, you will be familiar with the method of using ‘beats’ to tune one musical instrument from another; in the BFO, the beat frequency produced is the tone signal you hear.

In the more complex amateur radio receiver, a BFO is incorporated as part of the whole system. In our model, it is an external circuit that sits alongside your radio. The circuit diagram is shown in **Figure 1**.

Construction

Built on a small piece of matrix board about 80 × 50 mm, the circuit can be fitted inside a small plastic box. For once, we *don’t* want to screen the circuit to prevent it interfering with other equipment; we *want* it to interact

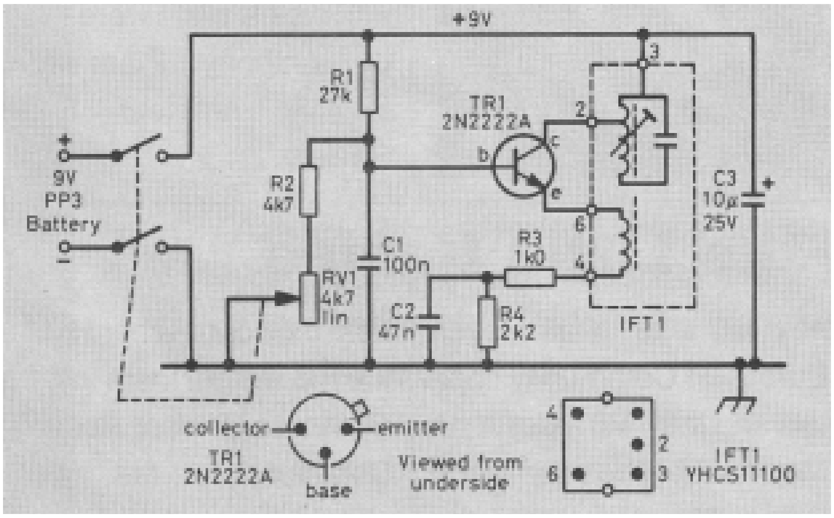
+9V
R1
27k
TR1
2N2222A
b
c
e
2
3
6
4
C3
10µ
25V
+
+
9V
PP3
Battery
−
R2
4k7
C1
100n
R3
1k0
RV1
4k7
lin
C2
47n
R4
2k2
IFT1
collector
emitter
TR1
2N2222A
base
Viewed from
underside
4
2
6
3
IFT1
YHCS11100

Figure 1 Circuit diagram of the BFO

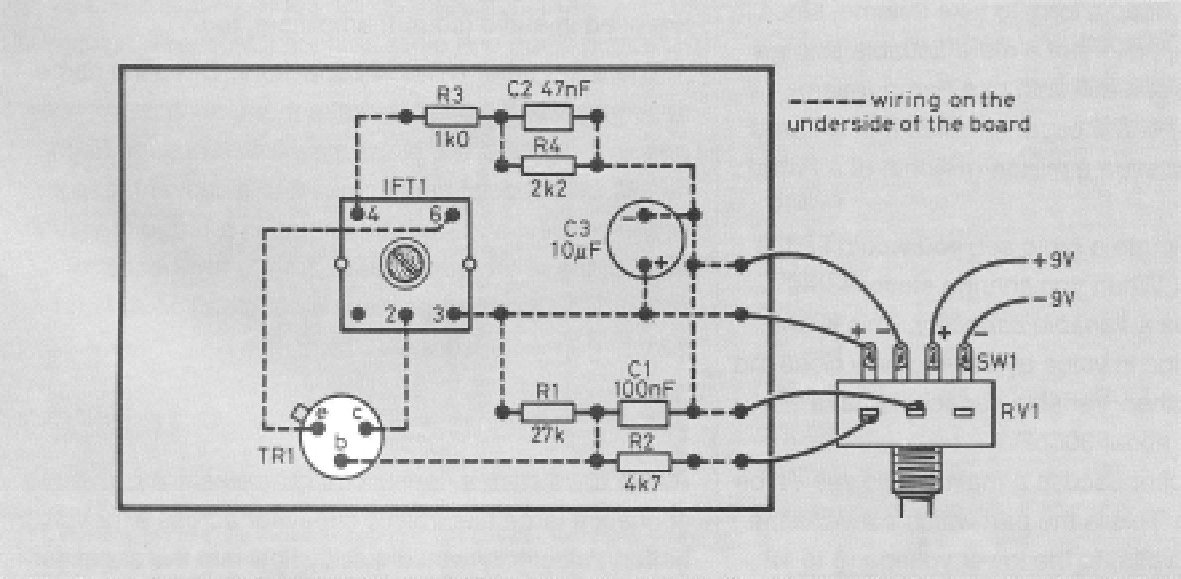

Figure 2 Matrix board layout shown from the component side. Adjust IFT1 carefully for the best results

with our receiver! This is why we use a plastic box. Maplin Electronic Components supply a suitable box, complete with the matrix board to fit inside (order code YU46).

Look at **Figure 2** carefully before you start to build the circuit, so that you can position the components correctly. Firstly, inspect the intermediate frequency (IF) transformer, IFT1, and remove its metal screening can very carefully. Again, this is to allow some signal to escape from our circuit and enter our radio. Having done this, solder the components, using the matrix board as a support. Underneath the board, the components are linked by single-conductor, insulated wire. Take particular care with the polarity of the electrolytic capacitor, C3, and the connections to the transistor, TR1 (see Figure 1).

The variable resistor, VR1, has a switch mounted behind the control itself, and the insulated leads to it from the battery should be about 10 cm long. Connect these before fitting VR1 into the case, so that the BFO can be calibrated (adjusted) correctly.

Calibration

After a final check that all the components have been fitted and soldered correctly, connect the battery, switch on, and hold the transistor between your fingers, to check that it is not getting hot. Place the circuit close to your receiver, and set RV1 to mid-position. Tune your receiver to find an amateur SSB transmission; the frequencies listed below will help you in knowing where to look. It may sound very strange, but don't worry. *Slowly* turn the core of IFT1 with a small, non-metallic screwdriver or with the correct 'trimming tool'. The core into which the blade fits is very fragile, so attempt this process with care. When the speech sounds as natural as you can get it, leave the core at this position, and use RV1 to make the speech sound natural.

Using the BFO

For best results, you may have to move the BFO nearer or further away from your radio. At the lower end of most bands (for instance just above 7.000 MHz or 14.000 MHz) you should be able to resolve Morse code (CW) signals. If you find that the BFO signal is a little weak, solder a 15 cm length of insulated wire to pin 2 of IFT1, and place it alongside your radio. This should improve signal intelligibility. When you are happy with the performance, switch off, drill a 10.5 mm hole in the box and fit RV1, followed by the matrix board assembly. Screw the base to the box, fit the knob, and the BFO is complete!

Where to listen

Band	**Frequencies (MHz)**
15 m	21.000–21.450
17 m	18.068–18.168
20 m	14.000–14.350
30 m	10.100–10.150
40 m	7.000–7.100
80 m	3.500–3.800

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1	27 kilohms (kΩ)
R2	4.7 kilohms (kΩ)
R3	1 kilohm (kΩ)
R4	2.2 kilohms (kΩ)
VR1	4.7 kilohms (kΩ), linear, with DPST switch

Capacitors

C1	100 nanofarads (nF) or 0.1 microfarad (μF), ceramic
C2	47 nanofarads (nF) or 0.047 microfarad (μF), ceramic
C3	10 microfarads (μF), 25 V radial, electrolytic

Additional items

TR1	2N2222A npn
IFT1	Toko type YHCS11100
Box	plastic, approximate size 100 × 70 × 45 mm
Board	matrix, to fit inside the box
Connector	for PP3 battery
Knob	for RV1

10 What is a capacitor?

Introduction

A capacitor is a device that stores energy in the form of electricity. Much less energy than a battery, and for a shorter time, however. The simplest form of capacitor takes the form of two flat metal sheets separated by air; connections are made to each plate, as **Figure 1** shows. If you imagine a pair of these plates, 30 cm square and separated from each other in air by 1 mm, the capacitance of this device would be almost exactly 80 picofarads (pF), i.e. 80 million-millionths of the unit of capacitance, the farad. Now this is quite a small value, you will agree, and it comes about because the farad is such a large unit. Nevertheless, as you will probably know, we may have capacitors of value 10 000 microfarads (μF) in our radio equipment, and they can be smaller than your little finger, so they are obviously not made the same way!

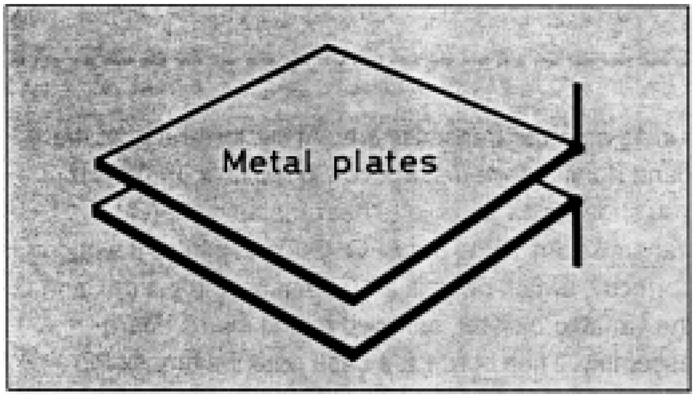

Figure 1

Large and small

We cannot go into the manufacture of capacitors here – after all, we are users of the devices, not the designers! First of all, beware of incorrect statements; the words ‘capacitor’ and ‘capacitance’ are *not* the same. For example, a large capacitor would be a description of one the size of a toilet roll. It need not have as large a capacitance as one the size of your little finger. A ‘large capacitor’ is one which is physically *big*, a ‘large capacitance’ refers to a capacitor which can store a larger amount of energy when a certain voltage is applied between its plates. The capacitors in a mains power supply are usually big **and** have large capacitances. High-power RF amplifiers may have large capacitors with small capacitances!

Electrolytics . . .

Electrolytic capacitors usually have capacitances of 1 µF or above. They differ from other capacitors in that they **must** be connected the right way round (i.e. they have positive and negative connections, just like a battery). They may explode if the connections are reversed!

. . . and the others

Other capacitors may be connected either way round, despite their names. We have *polystyrene*, *ceramic*, *silver-mica* and *tantalum*. Each has its own advantages and disadvantages, and the parts list for a project will always tell you which type is best.

Storing energy

If you were to connect a large capacitance across a 12 V power supply, nothing would appear to happen. Removing the capacitor from the supply and connecting it to a voltmeter would show that the capacitor has 12 V between its ends. This shows that, while the capacitor was connected to the supply, energy flowed from the supply into the capacitor. We say that the capacitor was *charged up* by the supply. If you are using an *analogue* voltmeter (i.e. one with a meter and pointer), you will notice that the indicated voltage slowly drops until, eventually, there is no voltage across the capacitor. This is because the capacitor has *discharged* its energy into the voltmeter. If you had used a smaller capacitance, the same would happen, except that the voltage would drop to zero more quickly – the capacitor stores a smaller amount of energy because its capacitance is smaller. Capacitors behave like other things in life – a small car can move more quickly than a large bus – a small piccolo emits a higher note than a flute – the voltages in a circuit containing a small capacitance will change more quickly than those in a circuit with large capacitance.

Varying the capacitance

Some capacitors are capable of having their capacitance changed manually; these are called *variable capacitors*. They work like the basic capacitor of Figure 1. Imagine moving the top plate of the pair a little to one side; the capacitance is determined, not just by the size of the two plates, but by their *area of overlap*. As this decreases, so does the capacitance. Such devices are limited in their capacitance, about 500 pF being the maximum value.

AC and DC

Because the plates of a capacitor do not touch each other, a direct current (DC) cannot pass between them. However, an alternating voltage on one plate can induce an identical alternating voltage on the opposite plate, and thus a capacitor appears to pass an alternating signal, even though currents as such, do not pass between the plates. This property of passing AC and not DC is very important, and a capacitor used in this way is called a *DC blocking capacitor* or, simply, a *blocking capacitor.* A blocking capacitor can be used at the same time, to couple a signal from one circuit to the next; here it would be known as a *coupling capacitor. Decoupling capacitors* are to be found where the capacitor is employed to remove an AC signal while retaining a DC component.

Finally. . .

Unlike resistors, the manufacture of capacitors renders them susceptible to excess voltage, so if you find a capacitor labelled 10 μF 16 V, it means that operating it above 16 V may fatally damage the device (and the circuit around it). This voltage is called the *working voltage* of the capacitor; on some electrolytics, you may find it expressed as *volts working* (i.e. 8 μF 450 V WKG).

Many smaller capacitors have their properties marked on them in a colour code, like resistors. **Figure 2** shows these codes, and their meaning, and the table below summarises the values of the colours.

Table 1

Colour	*Value*	*Voltage (tantalum capacitor)*	*Voltage (polyester capacitor)*
Black	0	10	–
Brown	1	–	100
Red	2	–	250
Orange	3	–	–
Yellow	4	6.3	400
Green	5	16	–
Blue	6	20	–
Violet	7	–	–
Grey	8	25	–
White	9	3	–

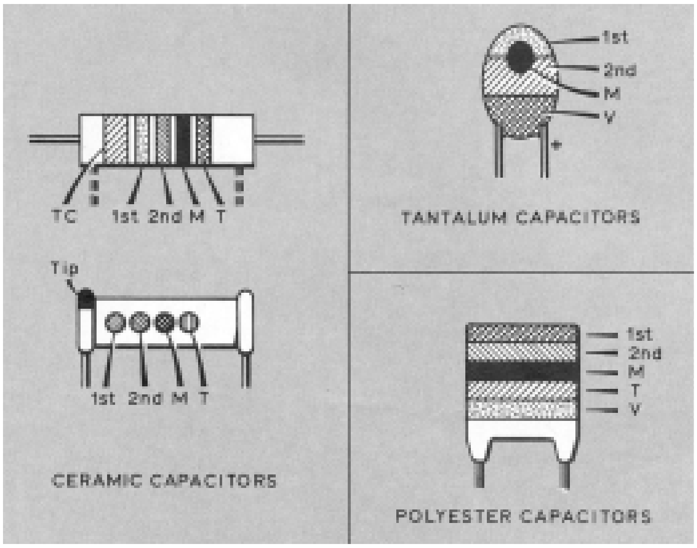
TC
1st
2nd
M
T
Tip
1st
2nd
M
T
CERAMIC CAPACITORS
1st
2nd
M
V
TANTALUM CAPACITORS
1st
2nd
M
T
V
POLYESTER CAPACITORS

Figure 2 Some capacitors have coloured bands or stripes, rather like resistors. The colour code, which is the same as the resistor code, is shown in Table 1. The band shown on the chart as '1st' is the first number of the capacitor's value in pico-Farads, '2nd' is the second number and 'M' is the Multiplier or number of noughts. For example, a capacitor reads from the top: Brown, Black, Yellow, Black, Red. Its value is One, then Nought, then Four more noughts = 100 000 pF (also referred to as 0.1 μF or 100 nF). Its tolerance (Black) is 20% and the working voltage (Red) is 250 V. The 'V' means the maximum working voltage. The band marked 'T' shows the tolerance, just like resistors, and the one marked 'TC' is only used on special capacitors designed to change their value with temperature

11 Waves – Part 2

Introduction

We left Part 1 with the concept that radio waves are divided up into bands which have different properties. Not *all* the properties are different, though. We need to discuss several wave properties, so we will start with what happens to waves as they propagate over long distances.

Getting weaker

Imagine a torch battery connected to a bulb by wires about 1 metre long. The bulb lights normally. If we now take the bulb 100 metres away from the battery and wire it up, we would expect the bulb to be somewhat dimmer, which is exactly what would happen. It happens because of the resistance of the wires – the wires do not form a perfect conductor. A similar situation occurs with radio waves.

All waves suffer from *attenuation* – they get weaker the further they travel. In cases of extreme attenuation, we need to apply some *amplification* before the attenuated wave can be used in a receiver.

Carrying information

When we speak over the telephone, the range of frequencies in our voices extends from very low frequencies up to about 15 or 20 kHz. In audio terms, this is quite a large *bandwidth* (meaning a wide band of frequencies). For communications purposes, however, most of this bandwidth is not needed, and in the telephone system (and in our transceivers), this is cut down so that it extends from about 200 Hz to 3 kHz, a reduction in bandwidth from 20 kHz to about 3 kHz. A bandwidth of 3 kHz has been found to be sufficient to convey speech intelligibly which, after all, is just what we need!

The radio waves coming from an amateur transmitter convey our speech signals over long distances. By themselves, the speech signals do not travel very far, so they have to be combined with a radio signal that *will* travel long distances. This extra signal is called the *carrier wave* (or just the *carrier*), because its job is to *carry* the speech signals along with it! The process of combining the speech (or Morse code) with the carrier is called *modulation*.

Wider and wider

Sending Morse code is achieved simply by switching (or *keying*) the carrier on and off. The bandwidth of the transmitted signal is only about 100 Hz. Speech, with its reduced bandwidth of 3 kHz, will produce a *single-sideband* (SSB) transmitted signal with a bandwidth of 3 kHz. If the same speech signal were used to produce an *amplitude-modulated* (AM) signal from the transmitter, it would have a bandwidth of about 6 kHz. Perhaps you can now understand why the bandwidths needed to produce hi-fi broadcasts need to be so large. TV signals need bandwidths running into tens of megahertz!

Waves need aerials

Radio waves are produced whenever changing currents flow through a wire, and when that wire is made in such a way as to maximise the *radiation* from the wire, it is called an *aerial* or *antenna*. The same piece of wire will receive radiation from other aerials; an aerial will transmit and receive. This is an important property of the aerial: when a current flows through it, electromagnetic waves are launched into the air; when electromagnetic waves in the air encounter the aerial, currents are produced in it.

From the simplest transmitting aerial, waves travel in all directions, like the waves on the pond that we considered in Part 1. They will travel a long way through air and space before they become too week to be received. They won't travel very far into the earth, however! The earth will *reflect* some of the wave and will *absorb* the rest. That portion of the wave which is reflected will again travel through air and space until it is totally attenuated.

Look at **Figure 1**; A represents a radio transmitter, with B1, B2 and B3 being receiving stations. The two arrows pointing 'downwards' from A represent two of the waves from A which just graze the earth's surface. Waves above

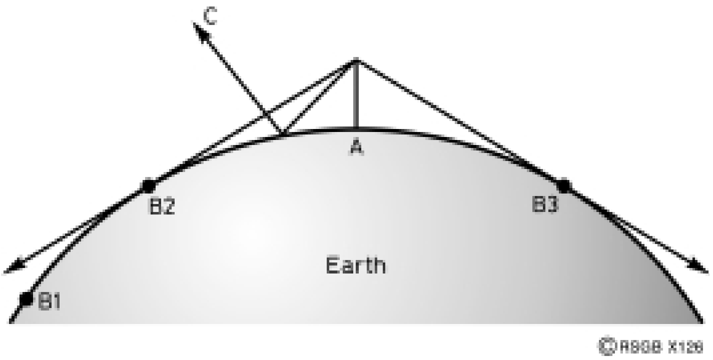

Figure 1 Stations between B2 and B3 receive the *ground wave*

these will travel on into space; waves below them will either be absorbed and reflected by the earth or received by aerials. B1 will not receive any signals from A, because it is below A's horizon. B2 and B3 can receive A's signals because they are *just* on A's horizon. Any stations between B2 and B3 will also receive A's signals, which are known as *ground-wave signals*. The wave at C represents one which is reflected by the ground and travels into space.

This description begs the question of how signals are received from stations well beyond the ground-wave range.

Mirrors in space

Suppose that there was something, out in space, that would reflect radio waves. Waves from A that travel out into space could be reflected off it and return to earth, enabling stations such as B1 to receive A's signals. The situation just described is illustrated in **Figure 2**.

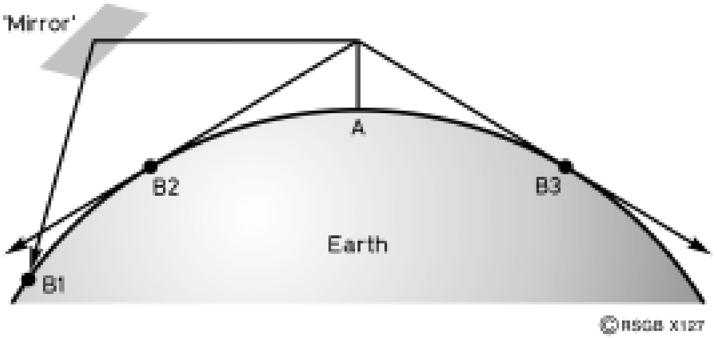

Figure 2 Reception of A's signals at B1 using the *sky wave*

Just such a mirror in the sky really *does* exist. It is not man-made, of course! The earth is surrounded by the atmosphere, a mixture of many gases, such as nitrogen and oxygen. The energy contained in the radiation from the sun is more than sufficient to *ionise* these gases, thus making them into electrical conductors. When a gas is ionised, some of its electrons are physically stripped out of the atoms and are free to move about, just as the electrons of a metal do in a wire. Consequently, we can regard this part of the atmosphere (the part illuminated by the sun) as acting like a sheet of metal, which *reflects* radio signals! It is not a perfect reflector, but is sufficient to produce long-range (DX) radio propagation via the *sky wave* under the right conditions. (A more down-to-earth example of ionised gases conducting electricity can be found in the fluorescent tube and the neon sign – many gases glow when they are continuously ionised.)

This conducting region at the extremity of the atmosphere is called the *ionosphere*, and it exists in layers between 60 km and 700 km above the earth's surface. When the ionosphere is sufficiently ionised, it glows; this is the natural phenomenon known as the *aurora borealis*, or the *northern lights*.

The property of the ionosphere that enables radio waves to be reflected does not act in a uniform way; it is very selective about which waves it reflects, and which waves go straight through it and into outer space. In general, it reflects only those waves with frequencies below about 30 MHz – the HF bands!

In Part 3 we will look at families of waves.

12 An LED flasher

Introduction

The LM3909 is an integrated circuit (IC) which will flash a light-emitting diode (LED). Using only two extra components and a battery, the circuit is cheap and has a very low current drain from a 1.5 V cell. The circuit can be used as a novelty flasher, an indicator for a dummy alarm bell box, or it could be attached to a torch so that it could be found easily in the dark! The simple circuit is shown in **Figure 1**.

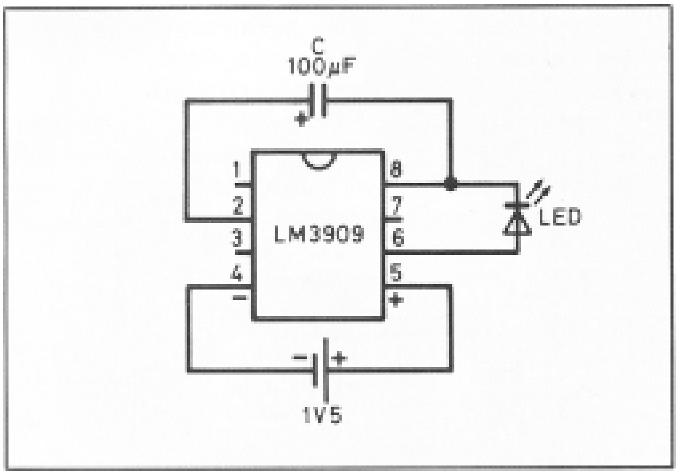

Figure 1 Circuit diagram of the LED Flasher. Pins 1, 3 and 7 of the IC are not used

Assembly

The circuit can be built on a small piece of Veroboard (the piece shown in **Figure 2** measures 15 holes by 10 strips). Using such a board, follow these instructions.

1. Depending on how far away you want the LED from the circuit board, solder a length of insulated wire to each lead of the LED. Use different colours of insulation – say, red and black, connecting the *red* lead to the *anode* (a) lead (the longer one) of the LED, and the black one to the *cathode* (k). Figure 2 shows these leads.
2. Cut the copper tracks as shown in Figure 2, using a 3 mm (⅛ inch) diameter drill, rotated between thumb and forefinger, or use the proper tool. Make *absolutely* sure that the tracks are completely broken!
3. Fit the IC holder in the correct position, using the cut tracks as guides, and make sure the small notch is facing towards the top of the board. Solder the pins to the copper tracks.
4. Mount the capacitor, positive end to the left, so that the positive lead is soldered to track F, which connects it to pin 2 of the IC; the negative lead is soldered to the right-hand side of track E, this being connected to pin 8 of the IC.
5. Solder on the battery leads, positive to the right, and the extended LED leads, positive downwards.

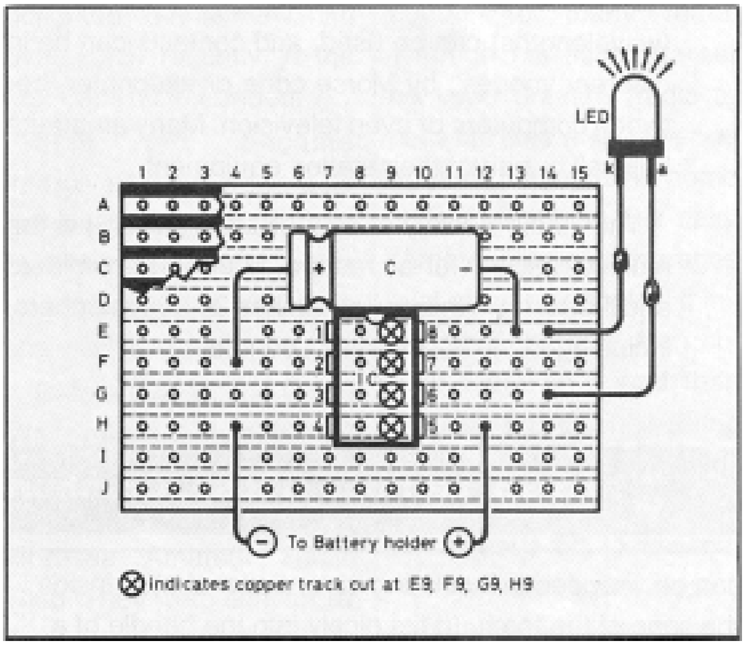

Figure 2 Board layout viewed from the component side. The tracks are cut under the board where shown

6. Check the circuit, and hold up the board to a bright light and look carefully for solder bridges between the tracks and pieces of copper swarf which may have escaped your inspection in 2 above! Remove whatever you find.
7. When all seems well, put the IC into the socket, ensuring that the notch or dot on the upper surface of the IC lines up with the notch on the holder. Line up each pin on the IC with the hole below it before pressing *gently* on the IC with the board supported on a firm surface.
8. Connect the battery; the LED should start to flash. The circuit is complete and working!

If you prefer, the whole circuit (battery included) can be mounted in a small plastic box, with the LED mounted on a clip and protruding through the panel. There are many other possibilities, and it is up to you to find an application for your own use.

Parts list

		Maplin code
LM3909	Integrated circuit	WQ39N
IC socket	8-pin DIL	BL17T
LED	5 mm diameter	WL27E
100 microfarad (µF)	Electrolytic capacitor (10 V)	FB48C
Battery holder	For AA-size cell	YR59P
Battery	1.5 V AA cell	

Small piece of Veroboard (15 holes by 10 strips)
Small plastic box (if required)
LED clip (if required)
Two lengths of coloured, insulated wire for LED (as required)

Availability

All parts can be obtained from Maplin Electronics Ltd.

13 Waves – Part 3

Introduction

When we talk about the spectrum being divided up into bands, this is just for our convenience; there are no natural divisions although, as we have seen, some of the properties of different bands really *are* different! Let's look at **Figure 1**, and see how the frequencies are divided up.

The divisions

- **Very low frequencies** (VLF) cover the range from a few kilohertz up to 30 kHz. Very long-range communication is possible, but at *very* small bandwidths. It is used for special purposes.
- **Long waves** (LW) are used for medium-distance commercial broadcasting and have frequencies from 30 kHz to 300 kHz.
- **Medium waves** (MW) are used for commercial broadcasting, and use frequencies from 300 kHz to about 1.5 MHz (1500 kHz). Typical range is about 200 km.
- **Short waves** (SW) encompass both the low-frequency (LF) and high-frequency (HF) amateur radio bands. There are nine narrow amateur bands in the SW spectrum between 1.8 MHz and 30 MHz. Some of these bands give round-the-world communication.
- **Very high frequencies** (VHF) span the range between 30 MHz and 300 MHz. Relatively short-range communication is possible. They were once used for broadcast TV before it moved to UHF. There are now three

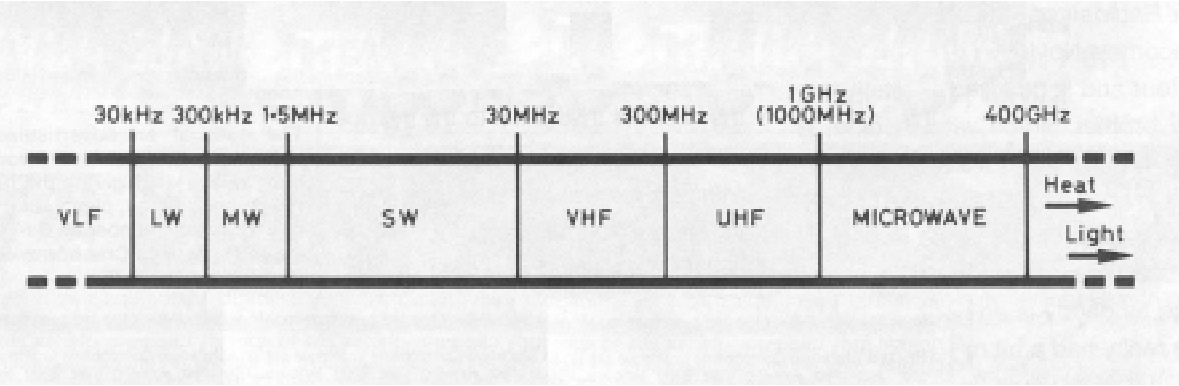

Figure 1 Diagram of radio frequency spectrum

amateur bands here – 6 m, 4 m, and 2 m. Repeaters are used to extend the usable range of mobile stations. VHF waves are not usually reflected by the ionosphere, but when they are, ranges of several thousand kilometres are possible. Weather affects these waves on a regular basis, however. In addition to amateur users, the VHF part of the spectrum is also used by the police, the fire and ambulance services, weather satellites, and many others,

- **Ultra-high frequencies**, sometimes called *centimetre waves*, cover the range from 300 MHz to 1000 MHz (or 1 GHz). The only amateur band in this range is the 70 cm band, and we share it with radar, TV and cellular telephone users as well.
- **Microwaves** begin at 1 GHz and extend to about 400 GHz. They are *never* reflected by the ionosphere, are partially attenuated by buildings, and are reflected from aircraft and cars. Microwave absorption in the atmosphere is quite significant, and rain and fog can attenuate microwaves quite heavily.
- **Heat, light** . . . Above 400 GHz we run into the infra-red bands and on into the visible light and ultra-violet bands. We generally take 400 GHz as being the limit of what we class as radio waves.

Bandwidth again

Complex signals need more bandwidth than simple signals. Even if it were possible, we would not be able to transmit a single television channel in the whole of the MW broadcast band! When TV used part of the VHF band, only five channels were possible in the range from 45 MHz to 68 MHz. By moving TV to the UHF band, we now have 47 channels between 470 MHz and 855 MHz!

It’s your choice!

The number of permutations you have amongst all the modes and all the bands is enormous! Only you can decide what you are interested in and what you want to learn about. That is the attraction of amateur radio!

14 Choosing a switch

Simple, but be careful!

A switch is the simplest electronic component, but care is needed in choosing the correct one for the job. Here is a list of the main characteristics of a switch, which will help you to select what you want.

1. **Rating**. This gives the maximum voltage and current that a switch can handle. For example, a 250 V 1.5 A switch will switch mains voltage at a current not exceeding 1.5 A. If the current is greater than 1.5 A, the switch may get hot and fail. If the voltage is too great, the switch may arc each time it is switched off, thus wearing away the contacts.
2. **Number of poles**. A single switch can control many circuits; the number of poles tells you *how many* different circuits it can handle. See **Figure 1**.
3. **Number of throws**. This tells you the number of positions each pole can have. This is best illustrated in Figure 1. The simplest ON/OFF switch is a Single-Pole, Single-Throw (SPST) switch. A Single-Pole,

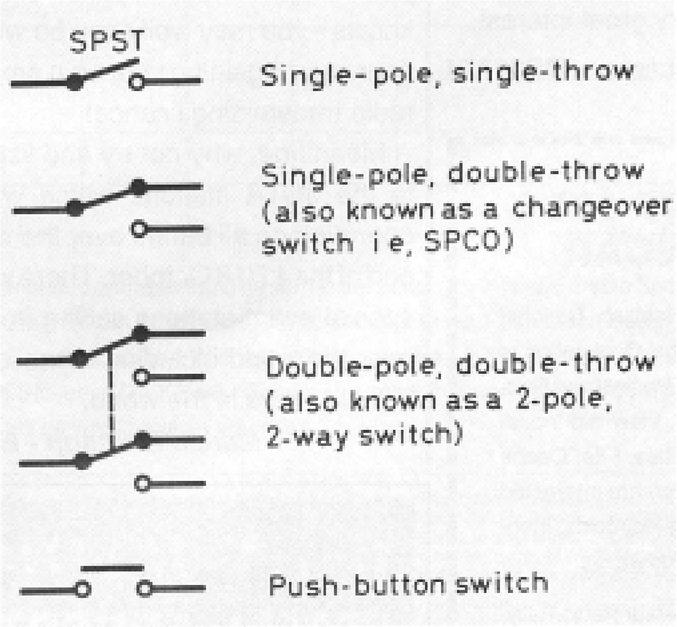
SPST
Single-pole, single-throw
Single-pole, double-throw (also known as a changeover switch i e, SPCO)
Double-pole, double-throw (also known as a 2-pole, 2-way switch)
Push-button switch

Figure 1 Some circuit symbols for different types of switch

Double-Throw (SPDT) switch may also be used as an ON/OFF switch, but is used mainly to change between two parts of a circuit, and is commonly known as a Single-Pole Change-Over (SPCO) switch. Two or more SPCO switches can be operated at once; Figure 1 shows an example. The two switches are said to be *ganged*. See also 'Types of switch' below.

4. **Number of ways**. When a switch has more than one throw, we tend to use the word 'ways' instead. This means that if a switch has one pole and six throws, we would normally call it a '1-pole 6-way' switch. Such switches tend to be rotary switches, as are described below.

Types of switch

(a) **Push-button**. These are found on calculators, telephones, electronic games and most equipment with a digital display.

(b) **Rotary**. These are switches controlled by a knob, and are turned instead of moved up and down. **Figure 2** shows the rear of such a switch and its circuit symbol. The commoner types of rotary switch are: 1-pole, 12-way; 2-pole, 6-way; 3-pole, 4-way; 4-pole, 3-way; 6-pole, 2-way. All these switches have 12 click positions, as you may have guessed, but each one comes with an adjustable end-stop so that you can set the correct number of ways according to the contacts on the switch.

(c) **Slide**. This switch is common on the cheaper types of radio, mainly as an ON/OFF or band-changing switch. They are not very rugged, but are small and cheap to produce. Very small types are manufactured for use on PCBs.

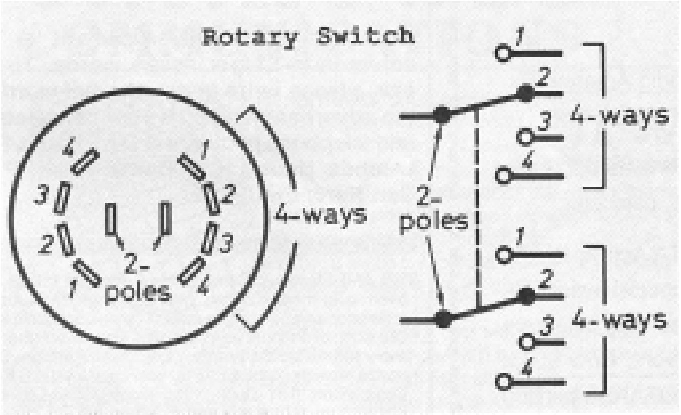

Figure 2 The rotary switch can select several different circuits at once

(d) **Toggle**. Available as two-way types or three-way, with a centre-off position. Can be manufactured to handle very high currents.

(e) **Micro-switch**. These are devices which are usually operated indirectly, such as when the cover is removed from a high-voltage power supply, or when the door of a fridge is opened. The ‘micro’ part of the name doesn’t refer to the size of the switch, but to the *small* movement that is required to activate it.

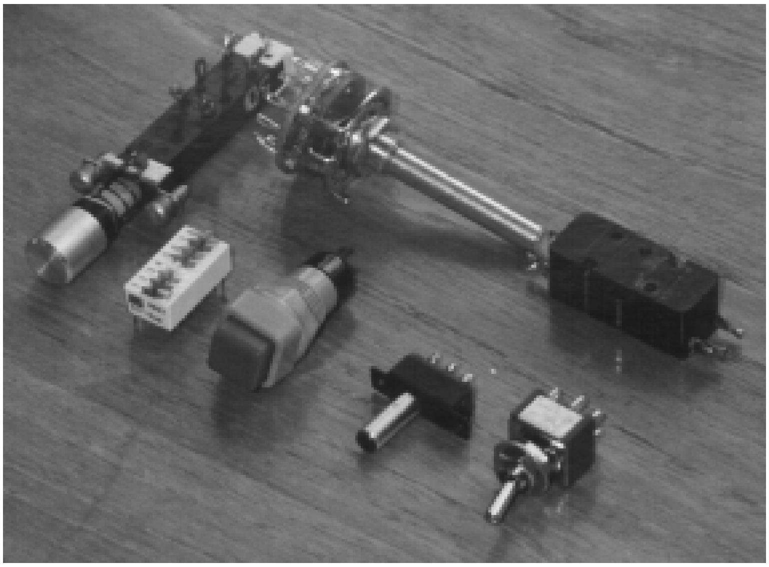

15 An aerial tuning unit for a receiver

Introduction

Any length of wire will act as an antenna (or aerial) but, to get the best results from a transceiver or receiver, an aerial tuning unit (ATU) is required. This matches the *impedance* of your aerial to the impedance at the aerial socket of your radio. Impedance is like resistance, and is measured in ohms, but it is used for alternating currents, and hence is common in audio and RF engineering. Most receivers have an impedance at the aerial socket of about 50 ohms (Ω); aerial impedance, on the other hand, can be anywhere between 20 Ω to over 1000 Ω, depending on its length and its height above ground.

Fortunately, you don't *need* to be able to calculate your aerial's impedance; all you need is a device that will perform the matching operation for aerials with a large range of impedances, and this is exactly what is described here! The subject of *matching* is a complex one, but all you need to know is that most signals will become clearer, and that there will be less noise and interference. Stations will become louder, so you will probably be able to reduce the setting of your RF gain control (always a good thing to do).

This design of ATU covers all amateur and broadcast bands from 10 m (28 MHz) to 80 m (3.5 MHz), and is very easy to build. The circuit is shown in **Figure 1**.

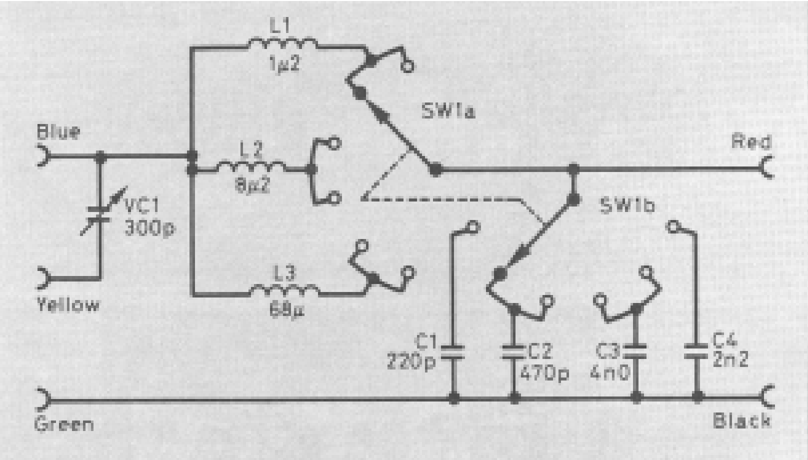

Figure 1 Circuit diagram of the Antenna Tuning Unit, showing the use of a 2-pole, 6-way rotary switch to select inductors (L) and capacitors (C)

Construction

1. Firstly, you will need a simple plastic case in which to house the ATU. The size should be approximately 85 × 145 × 50 mm.
2. Start by drilling two 10.5 mm holes in the front of the case; these are for the 6-way switch and the tuning capacitor.
3. Drill three 8 mm diameter holes in the left-hand side of the box, for the three sockets, coloured blue, yellow and green.
4. On the right-hand side of the box, drill two 8 mm diameter holes for the red and black sockets.
5. Now, fit the 6-way switch (SW1), the tuning capacitor (VC1), and all the sockets to the case. Check that the vanes of the capacitor rotate smoothly when the shaft is turned.
6. Wire up the inductors (coils). Figure 1 and the wiring diagram of **Figure 2**, will help with this. As you can see, each one side of each coil is connected to two switch connections, the other end going to VC1.
7. Solder in the fixed capacitors. One end of each goes directly to the ground socket (black), and the other end goes to the switch.
8. Solder a wire between the green and black sockets. The output from the ATU comes from the red socket, and this is connected to the two tags in the centre of the switch, as Figure. 2 shows clearly.
9. Finally, connect the blue and yellow sockets to the tuning capacitor, and the ATU construction is complete.

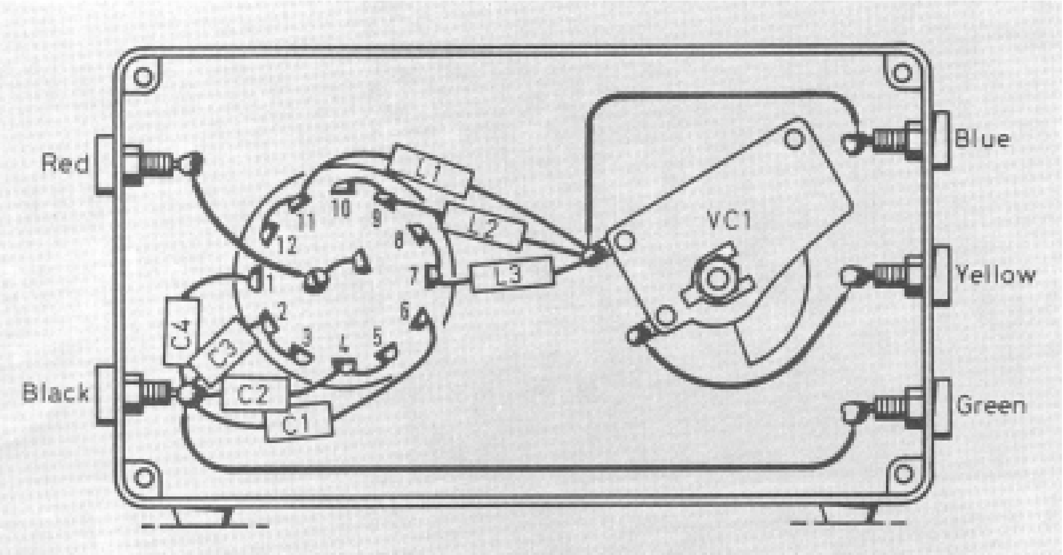

Figure 2 The internal view of the case shows the main tuning capacitor, VC1. This is a solid dielectric type, which has adjustable brass plates

In practice . . .

Figures 3 and **4** show two different ways of connecting your aerial to your ATU. In each case you will need to select each switch position in turn, and rotate the tuning capacitor through its full range while listening to a station. You should find that one switch position enables VC1 to produce a peak in the signal strength in the loudspeaker. At this point, your aerial and receiver are said to be *matched*. Stations in the same band will probably peak with VC1 at the same setting of SW1, but different bands will almost certainly require different positions of SW1.

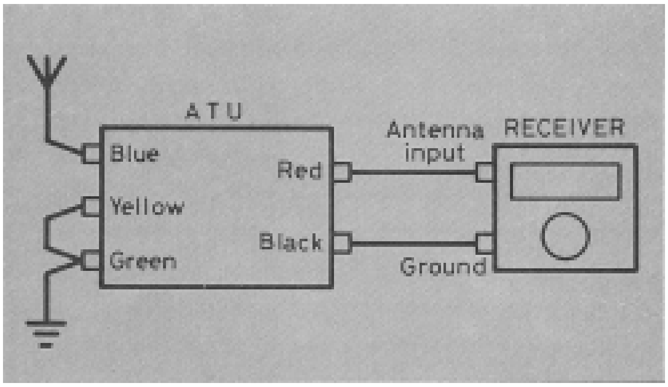
ATU
Blue
Yellow
Green
Red
Black
Antenna input
RECEIVER
Ground

Figure 3 You may find that parallel tuning gives best results with your antenna

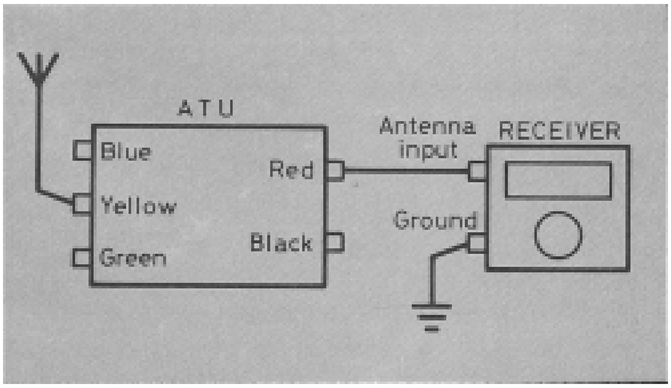
ATU
Blue
Yellow
Green
Red
Black
Antenna input
RECEIVER
Ground

Figure 4 In other cases, series tuning could be the most effective arrangement

Parts list

Capacitors (all rated at 16 V or more)

C1	220 picofarads (pF) polystyrene
C2	470 picofarads (pF) polystyrene
C3	1000 picofarads (pF) or 1 nanofarad (nF) polystyrene
C4	2200 picofarads (pF) or 2.2 nanofarads (nF) polystyrene

Inductors

L1	1.2 microhenries (μH)
L2	8.2 microhenries (μH)
L3	68 microhenries (μH)

Switch

SW1	2-pole 6-way rotary

Sockets

4 mm type, one each of red, black, yellow, blue, green

Additional items

Plastic or metal case, e.g. Maplin type YU54
Two large knobs for SW1 and VC1

16 A simple 2 m receiver preamplifier

Introduction

Designed specifically to complement the modified air-band portable (also described in this series), this can be used with some success on many receivers suffering from ‘deafness’ on VHF.

The circuit

An RF preamplifier is a device which improves the input signal to an existing receiver, enabling it to work more effectively. Because of the noise which is added to the signal by the preamplifier, *very* weak signals may not be usefully enhanced, but stronger signals will be improved considerably.

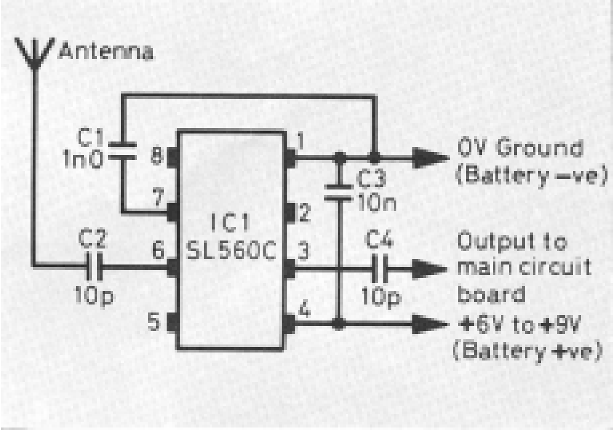

Figure 1 The integrated circuit is mounted upside down. Make sure you identify pin 1 which has a small dot next to it

This little circuit (shown in **Figure 1**) uses a GEC/Plessey integrated circuit type SL560C. With the addition of four capacitors, it is used between your aerial and the aerial input of the radio's PCB.

Putting it together

1. Use a small piece of prototype (matrix) board about 25 mm square. Use an 8-pin DIL socket for the integrated circuit (don't risk soldering the chip – it is seldom a risk worth taking). Figure 1 shows the connections to the socket, *looking from underneath*.
2. Make special note of the pin numbers, so that you know how to put the chip into the socket when you have finished. The positive and negative connections to the circuit are taken from the main PCB *after* the ON/OFF switch – so that the switch operates the preamplifier, too.
3. Unsolder the lead to the radio's telescopic aerial and connect it to the free end of C4, as shown. Then solder a short lead between the telescopic aerial and C2.
4. Sometimes it is possible to cajole your little preamplifier into the radio's plastic case, provided there is room and that you make sure that none of the soldered joints on your little board touch any of the metal inside the case when you replace the back and screw it on again.
5. If there is not enough room inside, then put the preamplifier into its own box, with battery and switch, and its own aerial. Keep the lead from the preamplifier to the aerial connection of the radio as short as possible – perhaps using screened cable.

Because your preamplifier is untuned, you will find not only that it helps with reception on 2 metres, but also that reception on the FM broadcast band is improved!

Parts list

Integrated circuit

IC1	GEC/Plessey SL560C

Capacitors

C1	1 nanofarad (nF) ceramic
C2, C4	10 picofarads (pF) ceramic
C3	10 nanofarads (nF) ceramic

Additional item

Prototype broad approx. 25 × 25 mm

17 Receiving aerials for amateur radio

Introduction

For any radio receiver to work well, it must have some form of antenna, or aerial. In almost all domestic transistor radios, the aerial is built into the set, either as a ferrite rod (which looks like a rod of dark grey metal) on which are wound coils of wire, or as a chromium-plated telescopic metal rod. Some radios have both forms of aerial, using the ferrite rod aerial for long waves (LW) and medium waves (MW) and the metal rod for very high-frequency (VHF) stations using frequency modulation.

There's broadcast reception . . .

No aerial is perfect, and these two types are *far* from perfect! As in most mass-produced equipment, *they serve their purpose*, which is not critical, and they allow the radio to be carried around easily, because they are not big and bulky. The broadcast stations are very powerful, usually quite close, and the circuits in the radio are quite sensitive, so the need for large aerials disappears.

. . . and there's amateur DX

Here we have an application which, in comparison with the broadcast situation, couldn't be much more different. The stations do not use much power, they may be half the world away, and the requirement for good aerials and very sensitive receivers is paramount. The aerials must be large and they must be as high as possible, away from buildings and trees, which cause reduction in signal strength, and away from man-made sources of interference.

You may not have thought about this but, in general, the larger an aerial becomes, the longer the cable (or 'feeder') must be in order to reach your shack and the receiver inside it. Cables reduce the received signal, so what your aerial gains by being large, the feeder (if you're not careful) will lose!

A simple aerial . . .

We now know that a 'good' aerial is essential. But what is a 'good' aerial? It depends on your purse, your property and your enthusiasm! One of the simplest (and, incidentally, one which is *not* subject to the cable loss problem discussed above) is the **Long Wire**, shown in **Figure 1**. This is, quite literally, a long piece of wire going from a chimney stack to a tree or pole

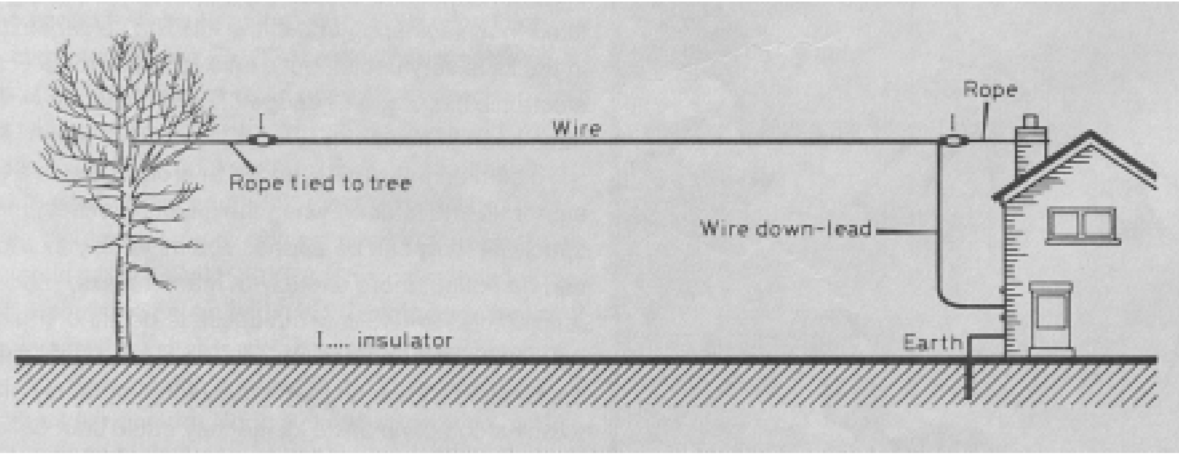

Figure 1 A tree at the bottom of the garden can provide a useful antenna support. Insulators can usually be obtained from Tandy stores or TV aerial suppliers

at the end of the garden. The longer it is, the better. Notice that the wire itself is *not* used to loop around the tree or chimney. Rope is used for both, and is secured to the end of the wire using an insulator of the 'dog bone' or 'egg' variety, to be found in profusion at rallies.

The wire is brought into the house through a window into the shack where the receiver is situated. The long wire aerial is best used with an aerial tuning unit (ATU), which is also described in this book. One of the advantages of the long wire is that it can be used on several frequency bands.

. . . and more complicated ones

Next up the ladder of complexity is the **dipole** (meaning two poles, or two elements). One form of dipole for lower frequencies would take the same basic form as the long wire, except that the feed to the receiver is taken, not from the end, but from the centre. The wire is essentially cut into two halves, and the two ends at the centre are connected to one end of a coaxial cable, which is then taken to the receiver. A smaller form of this is shown in **Figure 2**, which is conveniently mounted in a house loft. Aerials should always be mounted outside for best results, but will work when mounted inside, and the loft space is the logical situation.

The total length of the dipole should ideally be one-half of the wavelength of operation – hence the term 'half-wave dipole'. For example, a dipole for use on 20 m should be about 10 m long. Dipoles are thus 'single-band' aerials. They can be modified for use on several bands, and then become known as 'trapped dipoles', having coils and capacitors at certain points along their lengths.

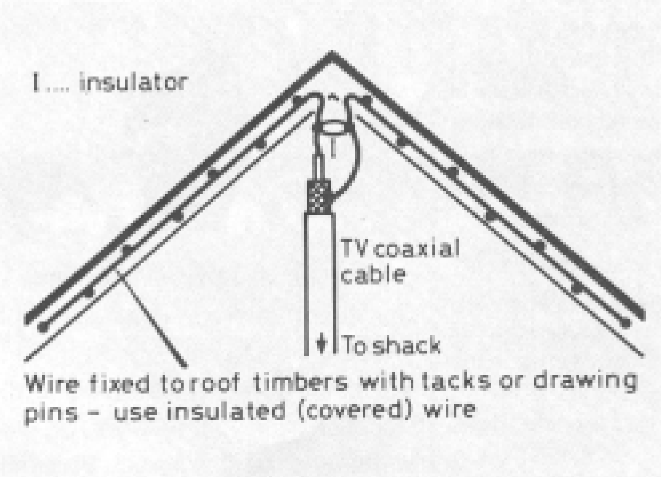
I.... insulator
I
TV coaxial cable
To shack
Wire fixed to roof timbers with tacks or drawing pins – use insulated (covered) wire

Figure 2 An indoor antenna should be mounted as high as possible

Indoor aerials

For flat- or apartment-dwellers, an indoor aerial is often the only solution, in which case the dipole of Figure 2 may be of great use. Other popular types include the **loop aerial**, of which several types are shown in **Figure 3**. The shape of the loop is of secondary importance, but it should be as large as possible.

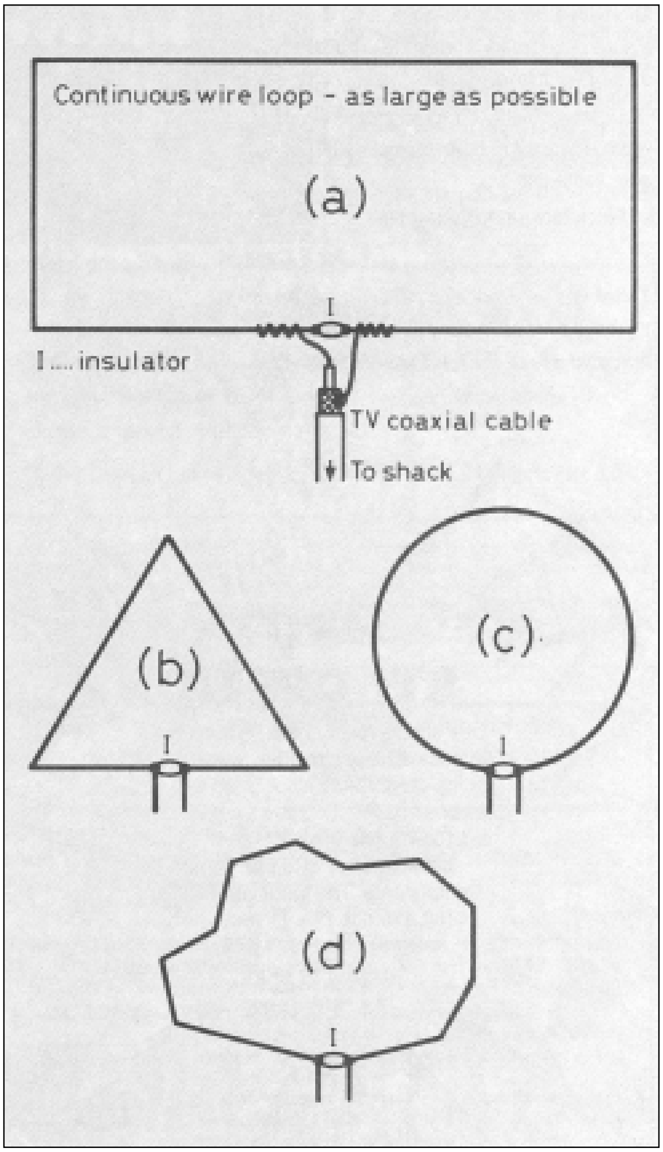
Continuous wire loop – as large as possible
(a)
I
I.... insulator
TV coaxial cable
To shack
(b)
I
(c)
I
(d)
I

Figure 3 Try different shapes of loop antenna to see which works best

Ingenuity is called for when considering indoor aerials. A dipole can be mounted along a picture rail or pelmet. A vertical loop of aluminium foil can be stuck to an inside wall and covered with wallpaper! The designs are limited only by your imagination, and will give you hours of listening pleasure

Outdoor aerials

If you want to venture beyond the simple long-wire discussed earlier, then the erection of outdoor aerials can be a bit tricky; they can be more than a one-man job, and can involve seeking planning permission. Get some help advice from someone who has done it before – what about your local radio club? Plenty of designs for aerials appear in *RadCom*, so you should never be short of ideas!

18 The Colt 80 m receiver – Part 1

Introduction

This is the first of four parts detailing the design and construction of a simple radio capable of good reception of amateur radio signals on the 80 metre band. If you can solder and have some basic hand tools, at the end of Part 4 you will have a working receiver, in which you can take pride, and start some serious listening on 80 m! For testing, all you will need is another receiver and a multimeter.

Description

The radio will be built in three modules, or stages, as illustrated in **Figure 1**. Each of these is built on a printed-circuit board (PCB) or matrix board and can be tested in its own right. The case will need some holes drilled, and care will be needed to mount the PCBs in the case.

Get to it!

Surprisingly enough, when you are building a radio section by section (which is *always* the best way), it is easiest to start at the output and work

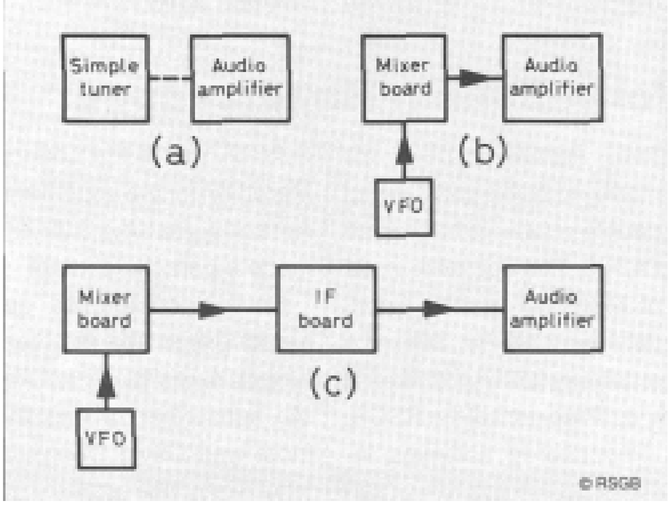
Simple tuner
Audio amplifier
(a)
Mixer board
Audio amplifier
VFO
(b)
Mixer board
IF board
Audio amplifier
VFO
(c)
©RSGB

Figure 1 Build your Colt and watch it grow. A simple crystal set (a) becomes a direct conversion receiver (b) and finally an 80 metre amateur band superhet (c)

towards the input, and you can test what you have done stage by stage. You will see what this means as you progress with your construction.

Every receiver needs some audio frequency (AF) amplification to make the sound signals big enough for you to hear. The circuit for the AF amplifier is shown in **Figure 2**. It uses the TDA7052 integrated circuit (IC) plus a handful of extra components. R1 in conjunction with C2 and C3 *decouple* the battery supply, preventing any audio signals getting through to it and affecting other parts of the radio, when they are connected. C1 acts to prevent high frequencies (above the 3 kHz bandwidth) going into the amplifier input. A volume control, VR1, is connected across the amplifier input, so that the amplifier can accept signal inputs over a wide range. **Figure 3c** shows the connections, which are made with screened cable. The

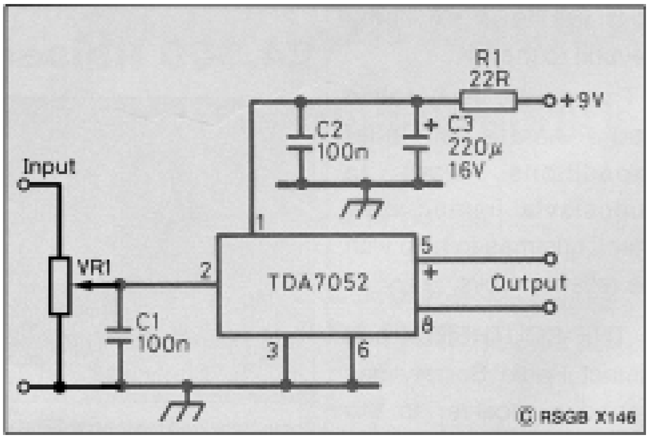
R1
22R
+9V
C2
100n
+ C3
220μ
16V
Input
1
5
VR1
2
TDA7052
+
Output
C1
100n
8
3
6
©RSGB X146

Figure 2 The Philips TDA7052 integrated circuit (IC) used in the audio amplifier needs very few extra components. It has a signal voltage gain of 100 times and the output is suitable for a loudspeaker or headphones

centre conductor of the left lead goes to the point marked 'input' in Figure 2, the braid being connected to the amplifier earth (0 V) tag. The right lead to VR1 goes to whatever signal source you have for testing – see later. VR1 is *not* mounted on the PCB.

The circuit is constructed on a small PCB or matrix board. Make sure that the electrolytic capacitor, C3, is soldered into the board the correct way – its positive and negative connections are shown in **Figure 3b** for reference. If you are **at all** concerned about soldering the IC into the board, enlist some help, *or*, obtain an 8-pin DIL socket, which you can solder in and then carefully insert the chip into the socket, making sure that is the correct way round. The markings on the chip are shown in **Figure 3a**.

The output leads from the amplifier go to a plastic 6.3 mm (¼ inch) mono jack socket, so that *neither* output lead is connected to the metal case. The amplifier will drive a pair of headphones or a small 8 Ω loudspeaker.

The battery leads must be the right way round also; the battery itself can be a PP3 or PP9, or you can use a small DC power supply.

Testing

First, check that all the components are in the right places, that your soldering is good, and that you have headphones or a loudspeaker connected. Set VR1 about halfway along its travel. Connect the battery.

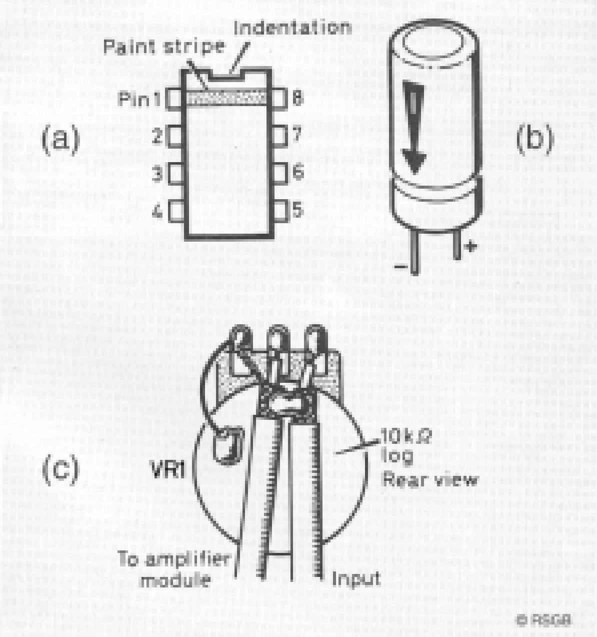

Figure 3 It is important to check the component connections carefully. The diagram shows (a) top view of the IC, (b) electrolytic capacitor and (c) volume control (VR1) connected across the input of the amplifier

A slight hissing noise should be heard; touching the input lead to the amplifier (the centre of the three connections on VR1) should produce a loud buzz. Touching the shaft at the same time will make the buzz quieter. This is the quickest way of confirming that your amplifier *seems* to be working. The only real test is to give it something meaningful to amplify! See the design of our *Crystal Radio Receiver* for full details.

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1	22 ohms (Ω)
VR1	10 kilohms (kΩ) log

Capacitors

C1, C2	0.1 microfarad (μF)
C3	220 microfarads (μF) electrolytic 16 V

Integrated circuit

IC1	TDA7052 audio amplifier

Additional item

PCB	(see below)

Component suppliers:
Maplin

The next part . . .

The metal case will be marked out ready to receive the completed modules.

19 A crystal radio receiver

Introduction

Here's a quick project to fill in a winter's evening! It was originally designed as a piece of test gear for the *Colt* receiver, but can be successfully used by anyone as a first radio project. Another use for it would be as test gear for any audio amplifier project requiring an audio input for test purposes; this is ideal, as it does not require any power supply!

Details

The initial rough-and-ready test for the audio amplifier of the *Colt* receiver will have whetted your appetite; you will want to prove more conclusively that your amplifier works, and in a way that others in the household will appreciate. Buzzing noises are not convincing in this respect! You are going to put together a simple crystal set – the simplest type of radio that there is – and use it as a signal injector for your amplifier. In this way, you build a real medium-wave (MW) receiver which drives a loudspeaker, as an intermediate product of the construction of an 80 metre amateur band receiver!

The circuit diagram is shown in **Figure 1**. It has a 60-turn coil mounted on a small piece of paper or card wrapped round a ferrite rod. The coil has a connection made to its centre-tap (the middle turn of the coil). The tuning capacitor, VC1, is the most expensive part of the circuit but don't worry, it will be used in the final design of the receiver also! Connect the diode, D1, from the centre-tap to the input to the potentiometer of the amplifier circuit of the *Colt*. Be careful to connect the aerial to the *vanes* of VC1, and not to its *frame*, or you will experience some strange effects when you are tuning. If your amplifier is working correctly, you should be able to receive local stations on medium-waves quite well.

If you think the crystal set will be of use to you in the future as a signal injector, all you will need will be another variable capacitor! If you do not intend to use the amplifier, a small crystal earpiece will allow you to listen. Walkman-type headphones will *not* work!

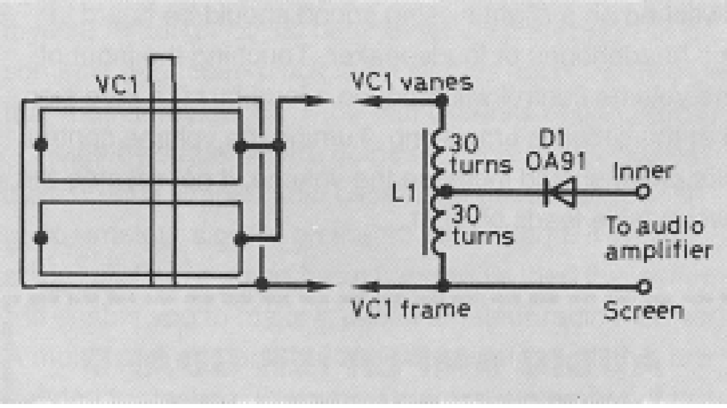
VC1
VC1 vanes
30 turns
D1
OA91
Inner
L1
30 turns
To audio amplifier
VC1 frame
Screen

Figure 1 Both sets of moving vanes are joined as shown. L1 is wound on a ferrite rod with 32 SWG wire and centre-tapped. A single winding (no tap) can be used, joining D1 to the top of the winding

Parts list

Variable capacitor	
VC1	125 + 125 picofarads (pF) twin-gang
Ferrite rod	
About 100 mm to 140 mm long	
Diode	
D1	OA91 germanium diode
Additional items	
Wire	2 m of enamelled copper wire, between 23 SWG and 32 SWG
Earpiece	High-impedance crystal type (only needed if you are not using the amplifier)

20 The varactor (or varicap) diode

Introduction

Many of the circuits for receivers and transmitters presented in this series rely upon the variable capacitor as a means of tuning. Another method of varying capacitance (without any moving parts) is provided by the *varactor diode*, sometimes called a varicap diode. This is a component which changes its capacitance as the voltage across it is varied.

The details

Figure 1 shows how a varactor diode might be connected to demonstrate its operation. Its symbol is that of an ordinary diode, with a capacitor symbol next to it. A variable voltage is applied across it in such a way that the diode is *reverse-biased*. This means that virtually no current passes through it – the positive voltage is applied to the cathode. Varactors are cheaper than variable capacitors, and they are tiny in comparison, very suitable for today's miniature circuits. If A and B were connected across the tuning coil in a simple receiver (with a series capacitor to block the DC from the battery reaching the coil), the tuning operation would be accomplished by turning the knob on the 10 kilohm potentiometer.

Varactors are available with different values, from less than 20 picofarad (pF) for VHF applications to 500 pF for medium-wave radios. They are

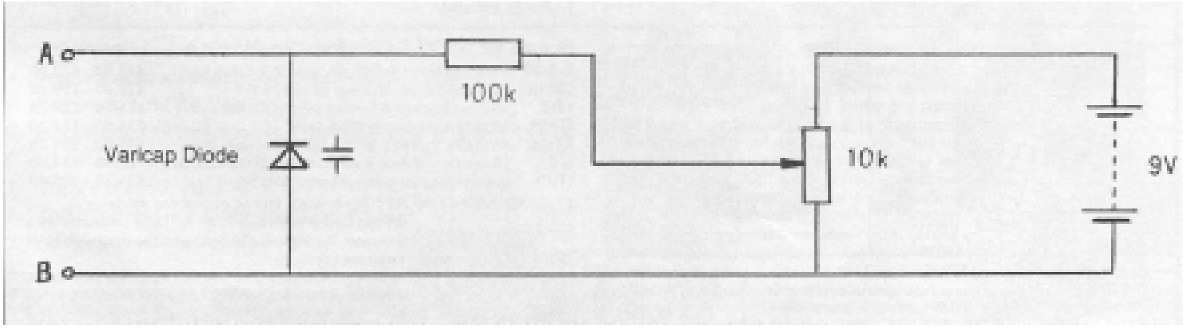

Figure 1 The capacitance of the varicap diode (between A and B) increases as the voltage is reduced, using the variable resistor

tuned usually by voltages between 2 V and 9 V. For a real application of varactors, you should consult the circuit diagram of the *Yearling* 20 metre receiver, elsewhere in this book.

In some circuit designs, several circuits are all tuned to the same frequency in order to improve the overall *selectivity* (the ability of the circuit to reject signals very close in frequency to the wanted signal). Special dual- and triple-varactors are available for circuits like this. Having been made at the same time from the same materials makes their individual characteristics virtually identical. Like all other diodes, they must be correctly wired into the circuit – their *polarity* is important.

Changes in temperature will cause the capacitance to change which, if it were part of an oscillator circuit, would cause the oscillator to drift – you would have to keep retuning the radio! This can be corrected by using a special integrated circuit called a *phase-locked loop* (PLL). Modern TV sets and satellite receivers use varactors and PLLs in this way.

Some useful varactor types

Type No.	*Tuning range*		*Description*
	pF/V	*pF/V*	
BB204B	42/2.0	15/12	Dual VHF
BB212	560/0.5	22/8	AM tuning
KV1235	450/2.0	30/8.5	Triple AM
KV1236	450/2.0	30/8.5	Dual AM
MV1404	120/2.0	9/10.0	HF tuning

21 A portable radio for medium waves

Introduction

The ZN415E integrated circuit (IC) can be used to make a very efficient AM portable MW broadcast radio with a built-in loudspeaker. Here's how!

The circuit

Figure 1 shows the circuit diagram of the portable radio. It's not as complicated as it may appear, especially after you have got started. L1 is a coil of wire mounted on a ferrite rod, acting as an aerial; VC1 is a variable capacitor which works, with L1, to tune in different stations. IC1 contains circuits of its own which boost the selected signal and it includes a *detector* which extracts the audio signal from the incoming RF signal. Earphones could be connected to the output of IC1 (between pins 4 and 5), but the output would not be powerful enough to drive a loudspeaker.

More sound

This is where IC2, an LM386 comes in. This is a small audio *power* amplifier which produces audio signals with enough power to drive a small

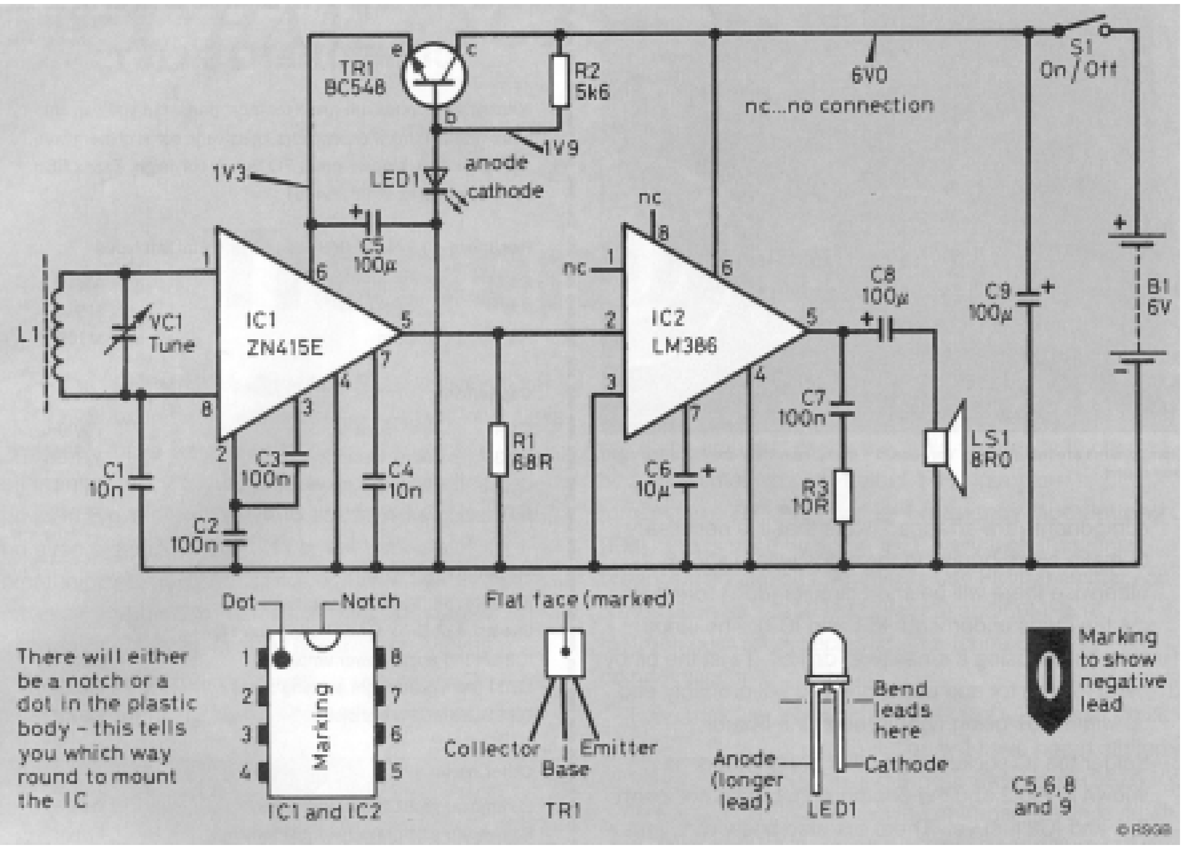
TR1
BC548
e
c
b
R2
5k6
6V0
nc...no connection
S1
On / Off
1V9
anode
cathode
1V3
LED1
nc
C5
100µ
nc
IC1
ZN415E
L1
VC1
Tune
IC2
LM386
C8
100µ
C9
100µ
B1
6V
C7
100n
R3
10R
LS1
8R0
R1
68R
C6
10µ
C1
10n
C3
100n
C4
10n
C2
100n
Dot
Notch
Flat face (marked)
There will either be a notch or a dot in the plastic body – this tells you which way round to mount the IC
Marking
IC1 and IC2
Collector
Base
Emitter
TR1
Anode (longer lead)
Bend leads here
Cathode
LED1
Marking to show negative lead
C5,6,8 and 9
©RSGB

Figure 1 The circuit diagram of our easy-to-build portable radio. Take care to mount the ICs and LED the correct way round

loudspeaker, LS1. The radio uses a 6 volt battery, which is made by connecting four 1.5 volt AA cells in series (4 × 1.5 V = 6 V) using a battery holder designed for this purpose. Although a 6 V supply is ideal for IC2, it is far too great for IC1, which needs only about 1.3 V. This lower voltage is provided from the 6 V supply by TR1 (an npn transistor), R2 and LED1 (a light-emitting diode). When current passes through an LED (see the description of the LED in this series) a reasonably constant voltage of 1.9 V appears between the anode and the cathode. Because of the voltage (0.6 V) that *always* exists between the base and emitter of a working transistor, the voltage on the emitter is about 1.9 V – 0.6 V = 1.3 V, and this is used as the power supply for IC1.

To keep the radio as simple as possible, no volume control has been fitted. Instead, you can use the directional properties of the ferrite rod aerial (see the information on ferrites in this book) to reduce the volume by rotating the set about a vertical axis using the handle provided.

Putting it all together

1. Start by covering the ferrite rod with Sellotape, or alternatively wrap a piece of paper tightly around it, and secure it with Sellotape. Then, with at least 2 metres of 24 SWG enamelled copper wire, wind 75 turns tightly around the rod. To be safe, leave about 50 mm of wire at the ends of the coil, then wrap the whole coil with Sellotape to hold the turns in place, leaving only the ends free. Then, using a small piece of sandpaper, remove the enamel from the last centimetre of each end of the coil.
2. Most of the components are mounted on a piece of Veroboard (the type with parallel copper strips on one side). The piece used on the prototype measured 32 holes by 10 strips, as **Figure 2** shows. Before you start fitting components, cut the copper strips as shown. It is easier to do it now than when the board is littered with components! The strips may be cut with a 3 mm (⅛ inch) twist drill rotated between thumb and forefinger. Resist the temptation to use a hand drill – the idea is just to cut the copper, **not** to drill right through the board!
3. Solder the IC sockets and the other components on the board as shown in Figure 2. Make sure that the IC sockets are fitted with their notches towards the *top* of the board, as viewed in Figure 2. Do not insert the chips yet. Always keep the wires left over from cropping resistors and capacitors, they will come in handy at times like this: make the wire links that are clearly shown in Figure 2. Connect the electrolytic capacitors (C5, C6, C8 and C9), the transistor and the LED the correct way round; then check it again when you have done it!
4. Finally, solder lengths of stranded insulated wire to act as 'flying leads' for future connection to L1, VC1, LS1, S1 and the battery connector.
5. Apart from the battery holder, everything is mounted on the case *lid*. This makes assembly and testing much easier, and eases fault-finding if

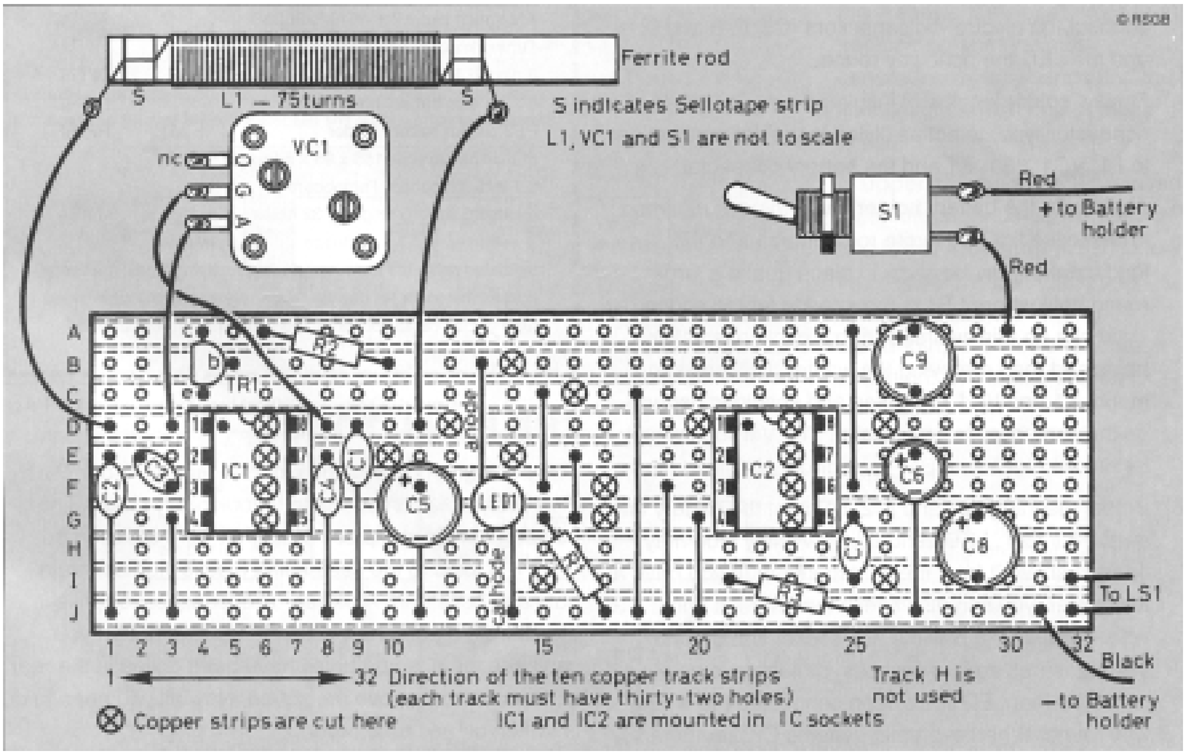
© R508
Ferrite rod
S
L1 — 75 turns
S
S indicates Sellotape strip
L1, VC1 and S1 are not to scale
VC1
nc
O
G
A
S1
Red
+ to Battery holder
Red
A
B
C
D
E
F
G
H
I
J
c
b
e
TR1
R2
IC1
C1
C2
C3
C4
C5
anode
cathode
LED1
R1
IC2
C6
C7
C8
C9
R3
To LS1
1 2 3 4 5 6 7 8 9 10 15 20 25 30 32
Black
− to Battery holder
1 ◄——► 32 Direction of the ten copper track strips (each track must have thirty-two holes)
Track H is not used
⊗ Copper strips are cut here
IC1 and IC2 are mounted in IC sockets

Figure 2 Veroboard layout for the Portable Radio. Make sure that all the wire links are included

the need arises! At the speaker position, make one large hole or a series of small holes to let the sound out. The ferrite rod may be stuck to the lid, as may the loudspeaker. Drill holes of the correct size to fit the particular types of variable capacitor (VC1) and switch (S1) that you are using. The Veroboard may be held in position by Blu-Tack or double-sided sticky tape.

6. Before inserting IC1 and IC2, connect the battery and switch S1 on. The LED should glow dimly (you may have to shield it with your hand in order to see it). If you have a test meter, check that there is about 1.3 V between pins 6 and 4 of IC1. If the reading is around 6 V or there is no glow, you may have connected the LED the wrong way round! When everything seems normal, switch off and disconnect the battery. Insert IC1 and IC2, making sure that the pins are straight and lie immediately above their corresponding holes in the sockets, and that the notches line up with the notches in the holders. Then push *gently* downwards on each IC in turn until the chip is firmly seated in its socket.
7. Switch on! By rotating the tuning capacitor, VC1, you should now be able to tune in many stations, rotating the radio to give you some volume control.

Final touches . . .

The handle was made with part of an old leather belt, secured to the case with 'number plate' nuts and bolts from Halfords. The loudspeaker grille is the lid from a *pot-pourri* container, and some extra holes were drilled in the back to improve the sound. See what you can find to finish off your radio!

Parts list

Maplin order codes are given for most of the parts, but you should get used to using the 'beg, borrow or steal' technique, or to use your ever-expanding junk box.

		Maplin code
Resistors: all 0.25 watt, 5% tolerance		
R1	68 ohms (Ω)	M68R
R2	5.6 kilohms (kΩ)	M5K6
R3	10 ohms (Ω)	M10R
Capacitors		
C1, C4	10 nanofarads (nF) or 0.01 microfarad (µF) ceramic	BX00A
C2, C3, C7	100 nanofarads (nF) or 0.1 microfarad (µF) ceramic	YR75S
C5, C8, C9	100 microfarads (µF) electrolytic, at least 10 V	FF10L
C6	10 microfarads (µF) electrolytic, at least 25 V	FF04E
Semiconductors		
IC1	ZN415E radio chip	
IC2	LM386 audio power amplifier	UJ37S
LED1	3 mm green LED	WL33L
TR1	BC548 npn transistor	QB73Q
Additional items		
LS1	Miniature 8 ohm loudspeaker	WB08J
S1	Miniature SPST toggle switch	FH97F
Ferrite rod	Length approx. 100 mm	YG20W
	24 SWG enamelled copper wire	BL28F

Additional items (*continued*)

VC1	Tuning capacitor 140 to 300 picofarads (pF)	FT78K
	Tuning knob	FK41U
	8-pin DIL IC sockets (two required)	BL17T
	4 × AA-size battery holder (long)	HF94C
	PP3-type clip for battery holder	HF28F
	Plastic box approx. 158 × 95 × 54 mm	LH51F
	0.1 inch Veroboard, min. size 32 holes × 10 strips	JP46A

Plus

- Stranded insulated conductor for flying leads
- Multicore solder
- Materials for handle and speaker grille
- Double-sided sticky tape or Blu-Tack
- Sellotape
- Glue
- Four AA-size 1.5 V batteries

22 The Colt 80 m receiver – Part 2

Introduction

In Part 1 we constructed the audio amplifier module for the system and tested it in a very simple way. If you did what was suggested and built the simple Crystal Set to use as a signal source, you will know just how well the amplifier works.

The case

Metal cases for the project are available from Maplin, telephone 01702 554 161 (code XB67). From the photograph on p. 72 you can see the way the components are mounted. The audio amplifier is seen at the top right of the base, to the right of the tuning capacitor VC1. The next in this series will deal with the variable-frequency oscillator (VFO) and VC1. The current part deals with preparing the case to receive the components.

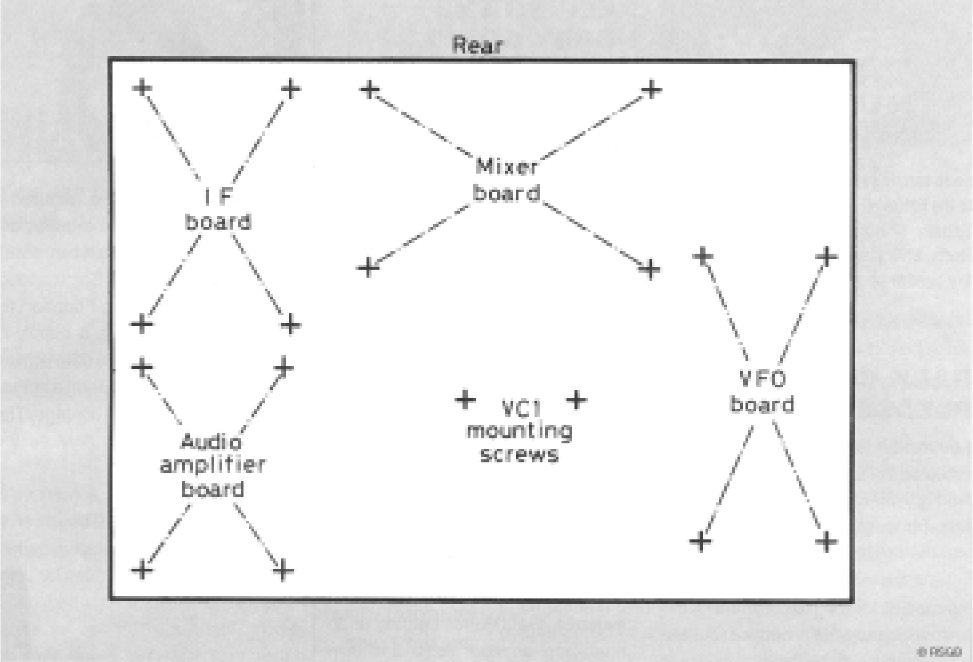
Rear
I F board
Mixer board
Audio amplifier board
VC1 mounting screws
VFO board
© RSGB

Figure 1 Fixing holes for each module are best measured from each printed circuit board or matrix board

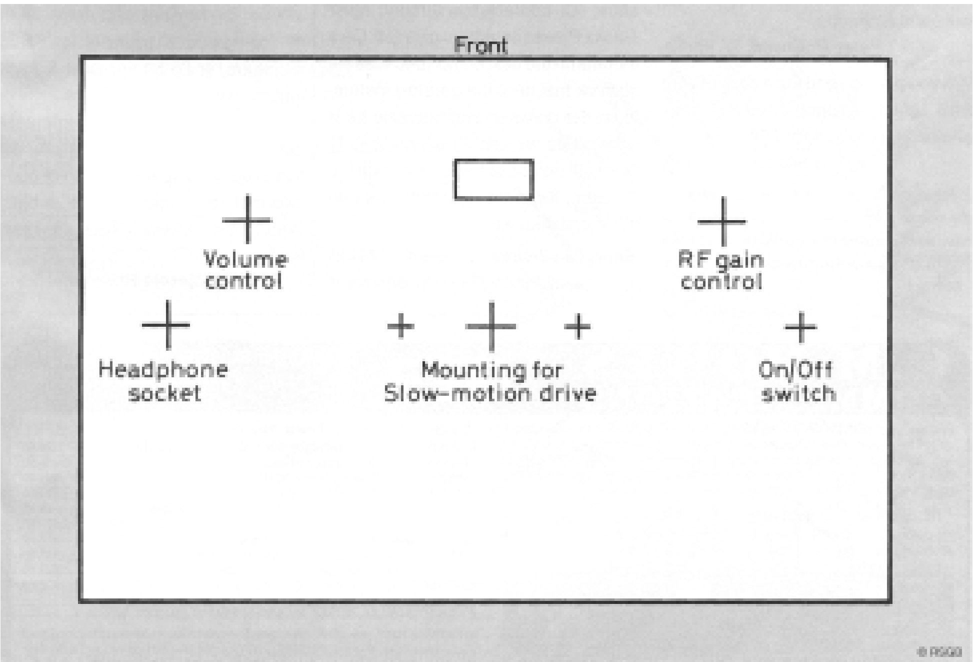
Front
Volume control
RF gain control
Headphone socket
Mounting for Slow-motion drive
On/Off switch
© RSGB

Figure 2 Position the slow motion drive to allow viewing of the tuning dial

Figures 1 and **2** show the markings for preparing the front panel and base. All the circuit boards and the tuning capacitor are mounted on the base using 10 mm stand-off pillars with 6BA bolts. The board locations are shown in Figure 1. The front panel control positions are shown in Figure 2, together with the small rectangular hole for viewing the tuning dial.

The best way to mark out the holes for the boards is to lie the boards flat on the base (before you've started soldering the components in) and marking the base through the holes in the boards. This minimises the scope for errors!

A *reduction drive* is used between the tuning knob and the capacitor shaft. This is simply a gear mechanism that slows down the capacitor shaft by a factor of six compared with the tuning knob, and makes tuning very much easier. The recommended variable capacitor also has a pulley wheel mounted on the shaft. Glued to this wheel will be a scale marked with frequency and is visible through the rectangular hole in the front panel.

The next part . . .

The variable-frequency oscillator and mixer will be added to the project.

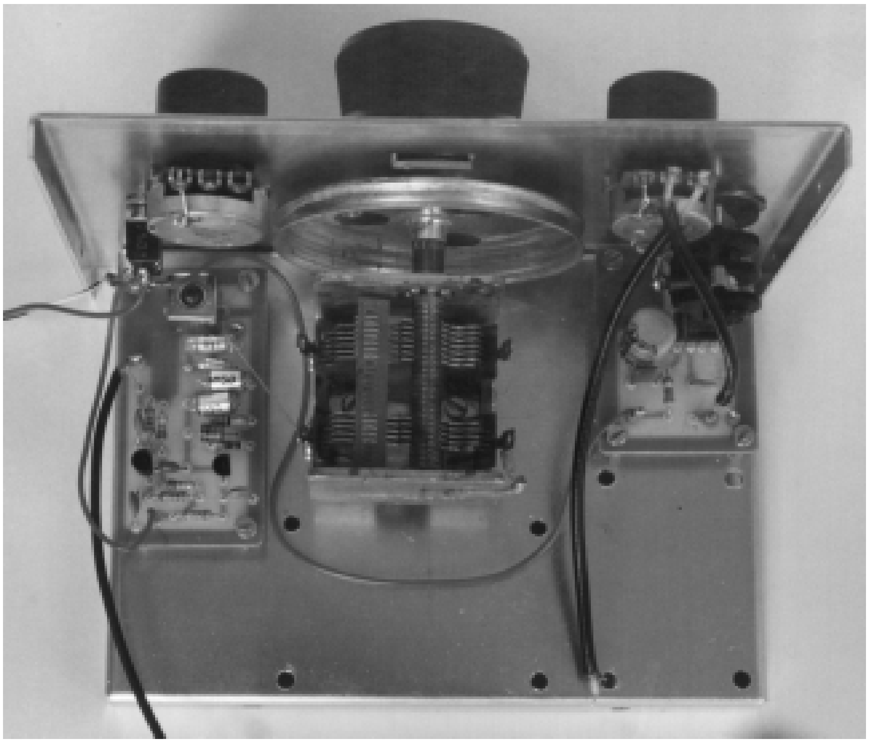

23 A simple transistor tester

Introduction

Although transistors aren't used as much as they were before integrated circuits came along, a transistor tester is still a useful piece of test equipment to have around the shack. This design is about the simplest possible and will produce an indication of whether a transistor is giving any current gain; this does not necessarily mean that the transistor is perfect but that it is working to a certain extent. This tester will **not** test field-effect transistors (FETs). If you buy a bag of transistors at a rally, this tester is useful for giving a yes/no indication of which ones go straight in the bin and which are kept for further use.

How it works

Figure 1 shows the simple wiring circuit. In order to explain the working of the circuit, a circuit diagram is shown in **Figure 2**, with the npn transistor, TR1, under test shown as part of the circuit. The pnp/npn selector switch, SW1, is omitted for clarity.

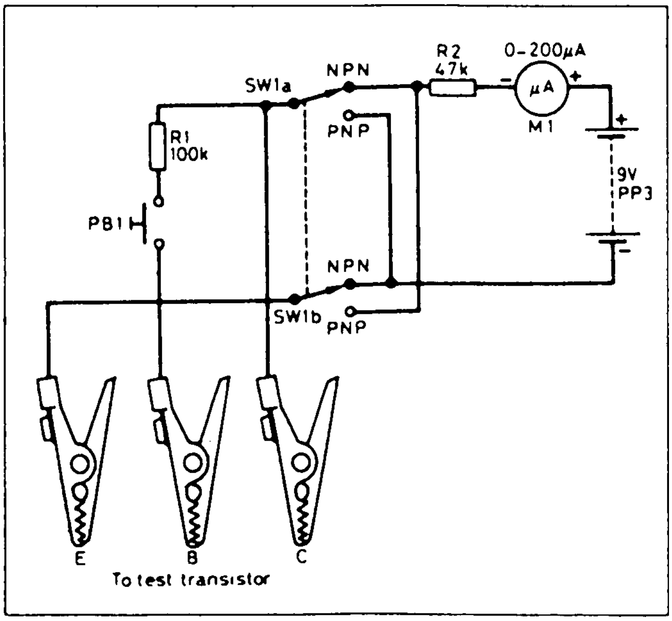

Figure 1 Circuit diagram of the transister tester

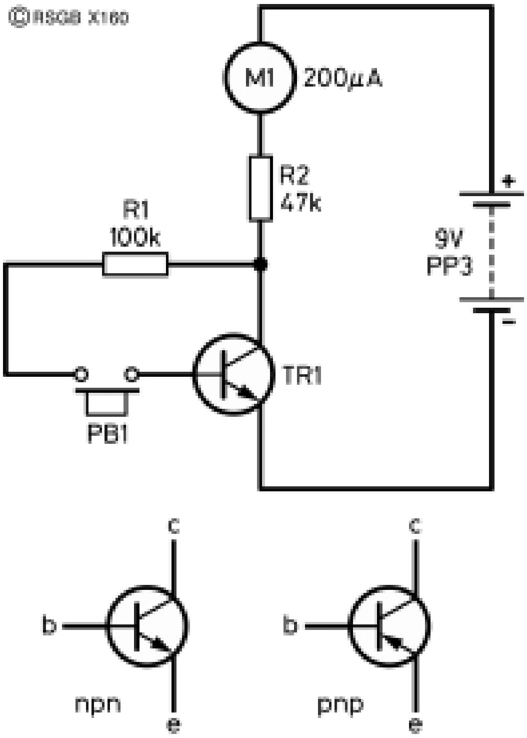

Figure 2 Testing an npn transistor

Any current that flows through the transistor, TR1, must flow from the battery, through the meter, M1, and through the protective resistor, R2. R2 prevents excessive current flowing through the meter and damaging it. Even if there is a short-circuit between emitter and collector, the maximum current that will flow is given by the simple equation

$$I = \frac{V}{R},$$

where I is the current flowing in amps,
V is the battery voltage, and
R is the total circuit resistance in ohms.

Putting in the correct values, gives

$$I = \frac{9}{47\ 000} = 0.000191 \text{ A, or } 191 \text{ microamps } (\mu\text{A}).$$

The resistance of the meter, M1, will cut this down a little more, but it is within the indicating range (200 µA) of the meter.

As shown, with the push-button switch, PB1, *open*, a good transistor will not draw any current from the battery, and M1 will thus remain at zero. When PB1 is *pressed*, a *very small* current is injected into the base of the transistor. If the transistor is working, it will produce a much larger current between the collector and emitter, and this current will also flow through M1 and R2, giving a significant reading on M1, showing that all appears to be well. If an appreciable current flows when PB1 is *open*, then your transistor is suspect.

Don't be put off by the apparently complicated switch, SW1. It is there to allow the other type of transistor, the pnp type, to be tested. All it does is reverse the battery connections, so that the emitter goes to the *negative* battery terminal for testing an npn transistor, and to the positive terminal for a pnp type!

Most transistors in common use are of the npn type, which is why Figure 2 shows the testing of an npn type. The connections to the different transistor *encapsulations* (shapes) are given in any good component catalogue. Avoid the trial-and-error method to discover the connections to a transistor. This is unscientific, and can be very frustrating, particularly if the transistor is faulty in the first place!

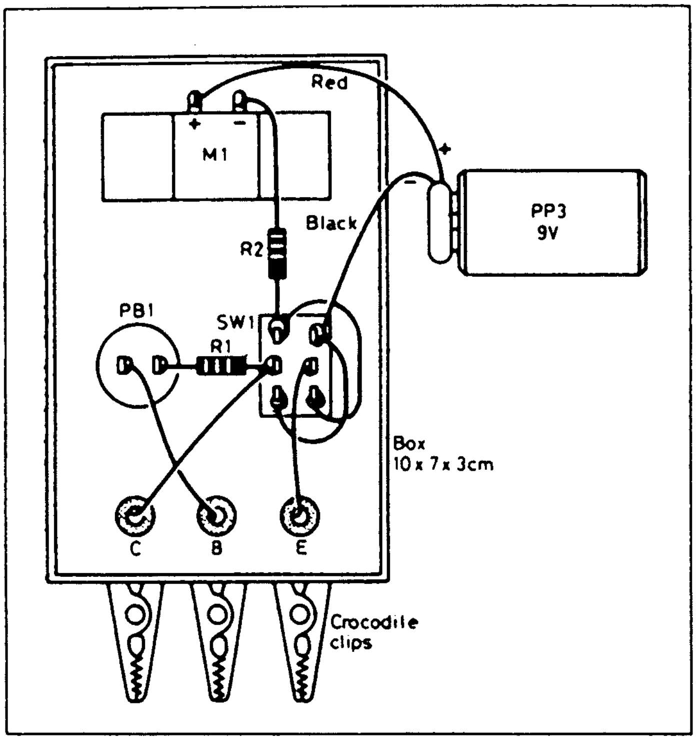

Figure 3 This shows how to wire up the tester

Construction

The unit can be built into a small plastic box, the components being soldered directly to the fixed terminals of the meter and the two switches; no circuit board is necessary! Use different colours of wire for the three test leads, and make sure you know which is which! Check the connections against the wiring diagram of **Figure 3**. When you are confident that all is correct, connect the battery and make sure there is no reading on the meter when nothing is connected to the crocodile clips!

Find any transistor for which you know the connections and the type (npn or pnp). Set the npn/pnp switch accordingly. Connect the three clips, making sure that they do not touch each other. A small reading on the meter at this stage means the transistor is suspect; a large reading means it is not working and should be thrown away! Press PB1 and watch the meter; a reading greater than half of full-scale indicates a good transistor. If it is less than half, you may have a transistor with 'low gain'; it may be usable for non-critical applications, but if you are in any doubt – throw it away!

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1	100 kilohms (kΩ)
R2	47 kilohms (kΩ)

Additional items

PB1	Push-button switch
SW1	DPDT (double-pole double-throw) switch
M1	Micro-ammeter – *not* less than 200 μA full-scale deflection
	Crocodile clips (3 needed)
	PP3 battery and connector
	Plastic box about 10 × 7 × 3 cm

Common types of transistor

BC108, BC109, 2N2369A

These are small-signal npn types, used in audio amplifiers. They are in metal cases (called TO18) and have a tab next to the emitter lead. Common types have a B or C suffix (e.g. BC109C). The C suffix indicates a higher current gain than those with a B suffix. The 2N2369A is specially designed for radio use at high frequencies.

2N3703, BC212L, BCY71
These are pnp transistors, and so must be used with the collector and base *negative* with respect to the emitter. These three types are used in small amplifiers and audio oscillators. The first two have plastic encapsulations (TO92), while the BCY71 has a metal case (TO18).

BFY50, BFY51, BFY52
For slightly higher powers, these are ideal. They have been used in novice transmitters up to 600 milliwatts (mW). The TO5 case is a scaled-up version of the TO18 case. They are all npn types.

2N3055, 2N3773, TIP35C
These are high-power transistors in bigger encapsulations. The thick metal TO3 case of the 2N3055 and 2N3773 is designed to bolt to a heat sink, a large piece of metal which conducts the heat into the air more rapidly than the transistor itself can. The TIP35C is made of plastic but has a thick metal tab by which it, too, can be bolted to a heat sink.

24 An introduction to transmitters

Introduction

We usually think of a transmitter as being a ‘black box’. However, that is the form a transmitter takes for our use on the amateur bands. Many electrical circuits are transmitters, even though transmitting may not be their primary function!

Anything that emits electromagnetic energy at *any* frequency is a transmitter, from radio at the low-frequency end of the spectrum, to gamma rays at the high-frequency end. We all know that a magnet will attract certain metals and that a comb rubbed on your coat sleeve will pick up small pieces of paper. The former is an example of the effect of a magnetic field, the latter of an electric field. Electromagnetic fields are combinations of both types of field, and are produced whenever an electromagnetic wave is transmitted.

What frequency?

Many everyday objects have a natural frequency of oscillation. This is called their *resonant frequency*. A wine glass will ring when struck gently; an empty wine bottle will sound if you blow across the top; a guitar string will vibrate when plucked. These are all examples of resonance, and the resonant frequencies will **not** change unless the objects themselves are changed physically in some way. These are resonances in *sound*; we are primarily interested in electrical resonances.

The basic electrical resonant circuit is the combination of an inductor (coil) and a capacitor, as shown in **Figure 1**.

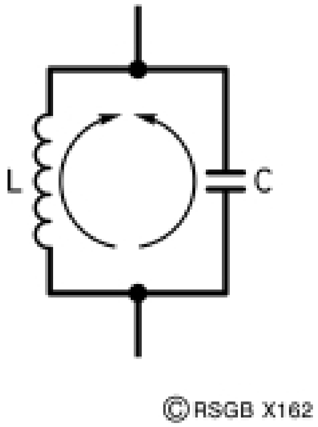

Figure 1 The basic electrical resonator. The energy in the circuit alternates between the inductor and the capacitor

A pulse of energy applied to this *tuned circuit* will make it *ring* (oscillate) at its resonant frequency. The energy in the circuit transfers between the inductor and the capacitor every cycle of the oscillation. Just like the wine glass, its oscillation dies away because it is losing some energy to its surroundings – it is transmitting! The frequency of the resonance depends on the values of L and C.

Keeping it going

If we want to keep the circuit oscillating, rather than having it die away, we must supply the circuit with just enough energy to replace the energy lost both by radiation and by losses in the circuit itself. Because of this, you will find in all oscillator circuits, a transistor, valve or FET working with the tuned circuit to provide this extra energy.

As it stands, of course, even with its transistor, our oscillator will not radiate very far. Connecting an aerial to it, and a Morse key to interrupt the power supply, it would become a very low-power CW transmitter. Add a couple more transistors to form a radio-frequency (RF) amplifier, and you have the basis of a simple low-power (QRP) transmitter.

Resonant circuits can also be made using quartz crystals; these work at the crystal frequency only, and this is marked on the crystal case.

A tiny spark transmitter

This is a simple piece of test gear that will increase your knowledge and understanding of resonance. You can use it to estimate the resonant frequency of most of the inductor/capacitor (LC) tuned circuits that you build. The circuit is shown in **Figure 2**. It operates around a relay. Any relay that operates from a 6 V to 9 V source and has contacts which are normally closed (i.e. closed when the battery is not connected). Fit the relay, a toggle switch and the battery in a metal box, and connected up as shown in the diagram. Some foam rubber inside the box may help to reduce the escaping noise of the relay. A small hole in the side of the box enables the 2-turn loop to emerge. This should be about 40 mm diameter, made with insulated wire. Switch on; there should be a loud buzzing noise from the relay. If not, you have probably chosen the wrong contacts on the relay!

When it is working, bring the loop close to the aerial of a radio – it should produce a loud noise from the speaker!

How it works

When you switch on, current flows through the relay contacts and through the relay coil. The relay operates and opens the contacts, causing the relay to 'drop out'. When it does, the circuit is completed again and the contacts are opened, and the cycle repeats. Each time the relay contacts open, there is a small spark between them, causing very rapid current surges through the wire loop. This makes the loop transmit RF energy, very briefly. In the early days of radio, this type of circuit was known as a *spark transmitter.*

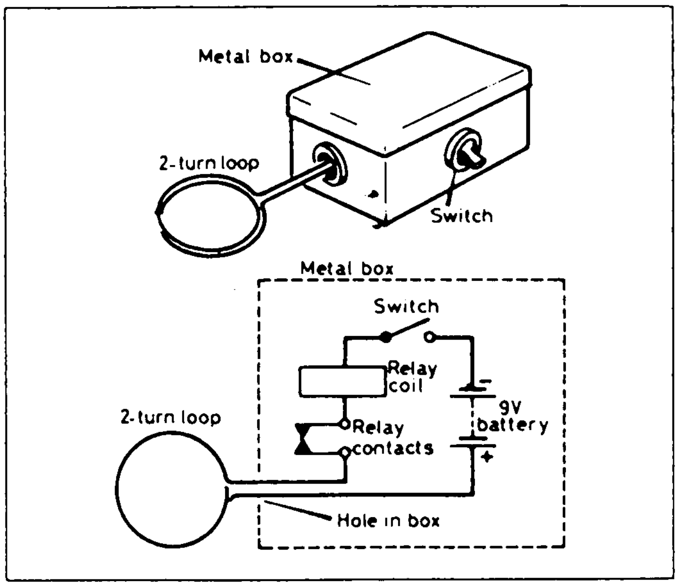

Figure 2 The current path is interrupted when relay is energised as shown above

Make a tuned circuit

Use a discarded toilet-roll centre, and wind about 10 turns of enamelled copper wire round it, keeping each turn close to the next. Scrape off the enamel for about 1 cm at each end, and solder a 100 picofarad (pF) capacitor (or a variable capacitor of about the same value) between the ends. The resonant frequency should be about 10 MHz. If you have used fewer turns or a smaller capacitor, the frequency will be higher.

Measuring the resonant frequency

Set up the buzzer as shown in **Figure 3**, with the loop around one end of your coil. Then make a similar loop, solder it to the end of a piece of coaxial cable going to the aerial socket of a calibrated receiver. Set the buzzer going, tune the receiver around 10 MHz, and search for the maximum noise level from the speaker. When you have found it, move the two loops as far away as possible from the main coil. This is called reducing the *coupling* between the coils, and it may result in a slightly different, but more accurate, resonant frequency.

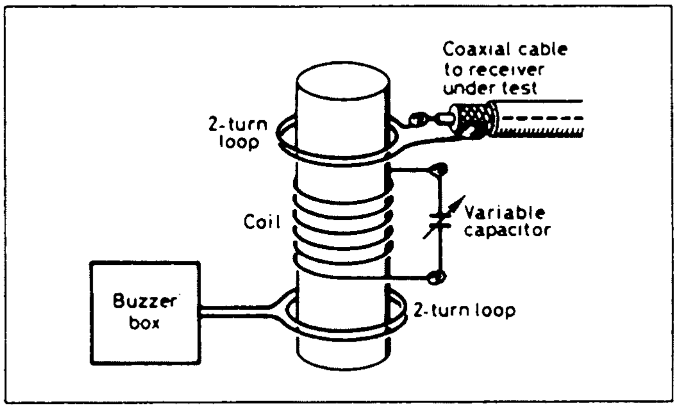

Figure 3 Experiment with the spacing between the loop and the tuned circuit

Parts list

Any small relay which operates between 6 V and 9 V
Metal box – do not use a plastic box!
9 volt battery and connector
On/off (SPST) toggle switch
Plastic foam, as required

25 The Colt 80 m receiver – Part 3

Introduction

In Part 2 we marked out the case ready for installation of the modules when they are completed. In this part, the building of the variable-frequency oscillator (VFO) and mixer will be described. This will produce a type of receiver known as *direct-conversion*, because it converts the radio-frequency (RF) signal directly into an audio-frequency (AF) signal which we can hear in a loudspeaker after amplification. A block diagram of the system is shown in **Figure 1**. The modifications needed to make a full superheterodyne receiver will be left until later.

The direct-conversion receiver covers the 80 metre amateur band and will receive both Morse (CW) and speech (SSB) signals. The audio amplifier was covered in Part 1, so your *Colt* is rapidly taking shape! By the time your construction has reached the end of this part, you will have a receiver ready to use, even if the project is not yet complete!

The direct conversion process

Like most things in radio, the principles of direct conversion are not difficult. From the aerial, the signal we want to hear is selected by the *tuned filter*, which rejects the signals we don't want. The signal then enters the

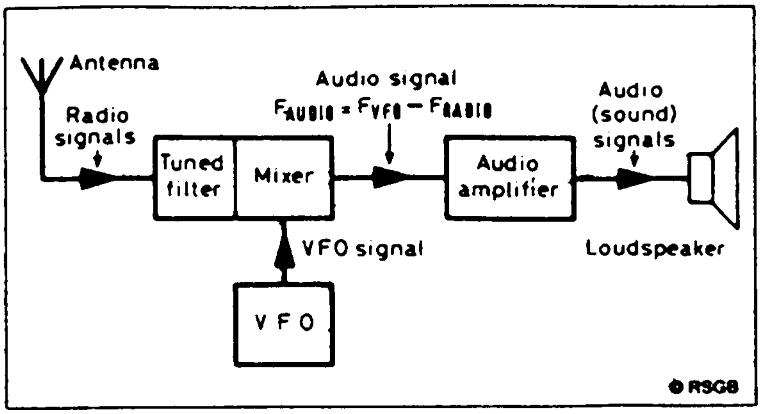
Antenna
Radio signals
Tuned filter
Mixer
Audio signal
F_AUDIO = F_VFO − F_RADIO
Audio amplifier
Audio (sound) signals
Loudspeaker
VFO signal
VFO
© RSGB

Figure 1 Stages of direct-conversion receiver

mixer, along with the signal from the *VFO*. The VFO produces a sine wave whose frequency can be varied across the whole of the 80 metre amateur band (3.5–3.8 MHz), by turning the knob on the tuning capacitor, VC1. The mixer produces, at its output, *two* signals; one signal is at the frequency of the **sum** of the signal and VFO frequencies, the other is at the **difference** of the two frequencies. It is the latter that we want. Let's look at the numbers involved. If the signal is at 3650 kHz and the VFO is at 3651 kHz, then the sum frequency is 7301 kHz, and the difference frequency is 1 kHz. If we feed the output of the mixer into our audio amplifier, the 7301 kHz signal is automatically removed (it is far too high to be considered an audio signal!) and the resulting 1 kHz signal is amplified and fed to the speaker, producing a note which we can hear!

Building the VFO

Figure 2 shows the circuit of the VFO. It is a tried and tested circuit, and should work first time. It uses a *field-effect transistor* (FET) for TR1, the oscillator itself. RFC is a *radio-frequency choke*, a coil of wire which will pass a direct current (DC) but which will prevent radio-frequency (RF) signals getting through.

All VFOs have a tuned circuit which, in this case, is formed by the coil L1 and the capacitors C1 and VC1. The frequency will also be affected to some extent by C2, C3, C4 and C5. Transistor TR2 is an *emitter follower*, a stage which gives no voltage gain but provides a good *buffer* stage, isolating the VFO from the effects of the stages that follow it. When building a VFO, the

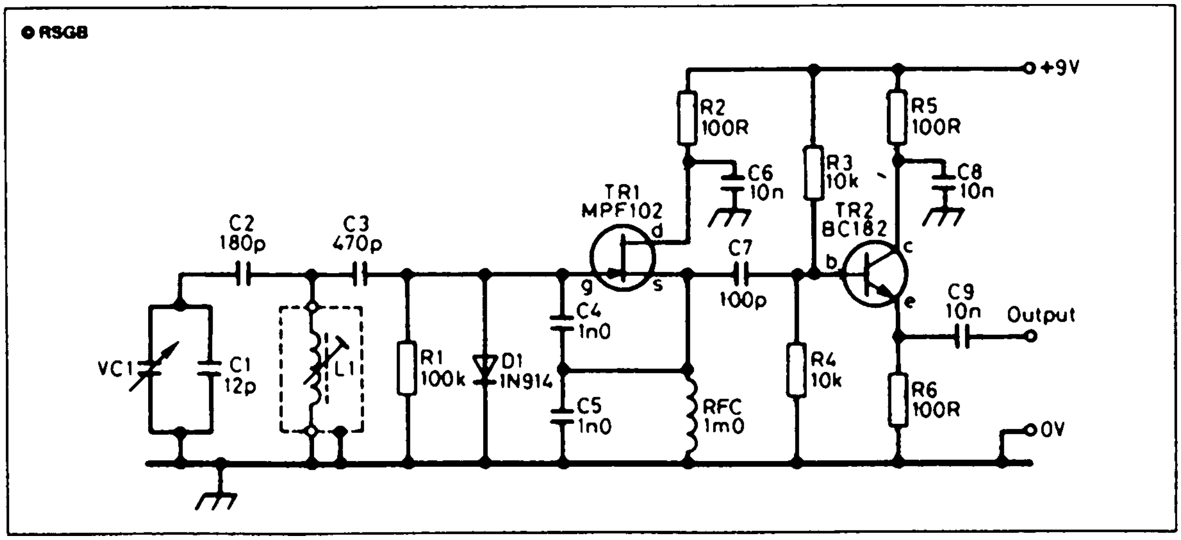

Figure 2 The variable-frequency oscillator uses a field-effect transistor (FET)

parts must be securely mounted. If components move, so does the frequency! At worst, the oscillation will become unstable and the VFO will be useless. Keep the component leads as short as possible – this improves their mechanical stability as well as their electrical stability!

Mount the parts on the printed-circuit board (PCB) or matrix board and, when completed, the VFO should look like the one shown on the left in the photograph on p. 72. Make sure that TR1, TR2 and D1 are the right way round.

On completion, check the component positions then mount it in the case as shown in the photograph, to the left of the tuning capacitor when viewed from the rear. The VFO coil, L1, will need some adjustment, but that will have to wait until the mixer is built. Connections to the other boards are made with screened cable.

The mixer board

So far, we have an audio-frequency (AF) amplifier and a VFO; the addition of a mixer board gives us a complete direct-conversion receiver for 80 m. The mixer circuit diagram is shown in **Figure 3**. Let's follow the signal path.

- From the aerial, the RF signal goes to the gain control potentiometer, RV1. This reduces very strong signals, to prevent them overloading the mixer.
- To select the required band of signals, a *bandpass* filter is made up of RF transformers T1 and T2; these are tuned by C1 and C3, and are coupled together by C2.

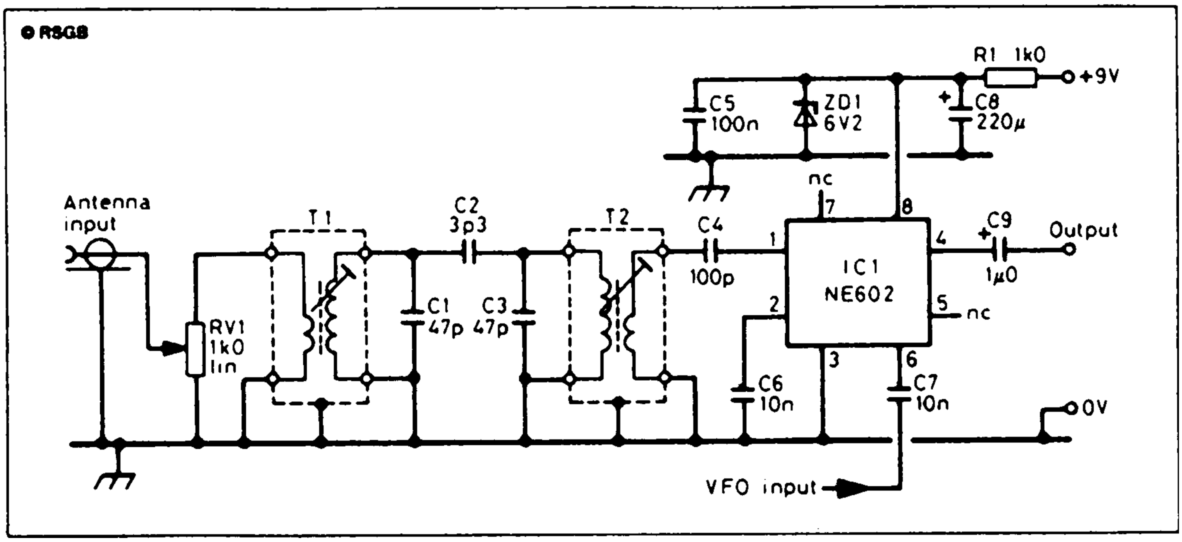

Figure 3 The mixer board has a bandpass filter and stabilised supply

- After the filter, the signal is coupled into the integrated circuit (IC) mixer type NE602, by the capacitor C4. Capacitors C5 and C8 *decouple* the supply line, to prevent unwanted signals on the supply from disturbing the VFO operation. The use of the term *decouple* is exactly the opposite of *couple*; when two circuits are *coupled* together, the signal passes from one to the other; when two circuits are *decoupled*, signals **cannot** pass from one to the other.
- The NE602 works with a 6 V supply; it is produced here from the 9 V supply by the *Zener diode* ZD1 and the resistor R1. ZD1 operates at 6.2 V, and gives a steady output for the mixer. The audio output from the mixer appears at pin 4 or IC1. This is taken to the audio amplifier board via C9 and the volume control (see Part 1).

Care must still be taken to insert some components the right way round. These are the electrolytic capacitors, C8 and C9, the Zener diode, ZD1, and the integrated circuit, IC1. Check all component positions and make sure all your soldered joints are bright and shiny.

Putting it together

The interboard wiring, shown in **Figure 4**, uses screened cable; ideally, this should be thin coaxial cable, but screened microphone cable is suitable. The diagram shows how the two controls, the RF Gain and Volume, are connected to the boards. The leads marked '+9 V' are all connected to the battery supply via a miniature on/off toggle switch. Double check all connections before connecting the battery.

Setting the VFO

Very little adjustment is needed to get the receiver going. Firstly, the VFO must be adjusted to cover the required band, in this case 3.5–3.8 MHz.

If you have a frequency counter, connect it to the output of the VFO. If you haven't, read this part anyway so you understand the process, then another means of setting the VFO will be given especially for you! Rotate VC1 until the vanes are fully meshed. Very carefully, adjust the core of the VFO coil (L1) with a plastic trimming tool so that the frequency approaches and settles at 3.500 MHz. When you rotate VC1, the frequency should increase to **at least** 3.800 MHz at the far end of its travel.

In the absence of a frequency counter, borrow a communications receiver, set it for SSB reception (USB or LSB) on exactly 3.500 MHz. Set VC1 with the vanes fully meshed and turn the core of L1 in both directions until you hear a whistle in the communications receiver. Rotate the core so that the whistle reduces in frequency. It will eventually fade out at around 200 Hz;

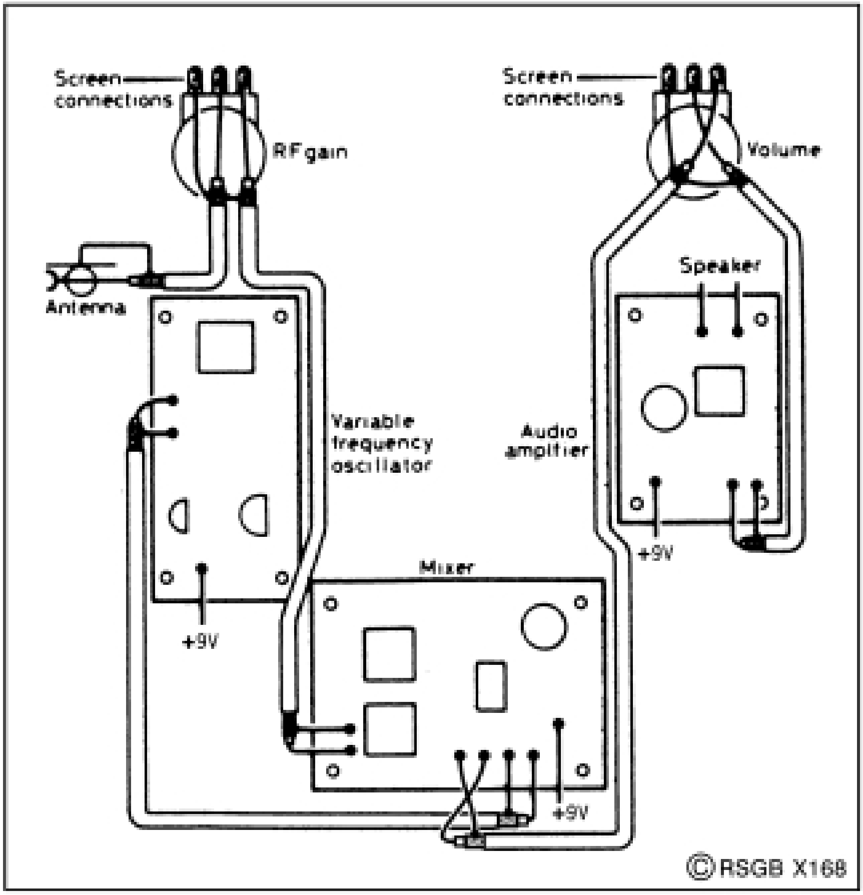
Screen connections
RF gain
Antenna
Variable frequency oscillator
+9V
Screen connections
Volume
Speaker
Audio amplifier
+9V
Mixer
+9V
©RSGB X168

Figure 4 Make sure that interconnections between the boards are correct, including cable screens

turn the core a little further and then leave it at that position. Rotate VC1 right to the other end of its travel, and search for the whistle with the tuning knob of the communications receiver. Check that the VFO frequency is at least 3.800 MHz.

Then, whichever method you are using for frequency measurement, mark the dial with frequency steps of 50 kHz. Setting and calibration are finished!

Setting the mixer

Again, there are various ways of doing this. If you have a signal generator, inject a signal at a frequency within the 80 m band, and adjust the cores of T1 and T2 sequentially for maximum output.

If you haven't a signal generator, connect an aerial to the mixer input, set the RF Gain to maximum (fully clockwise), and find a consistent signal. Adjust the volume control to a comfortable level. Rotating the core of T2 with your trimming tool, maximise the output. Then do the same with T1, although this will have much less effect. Find another station, and check that the positions of the cores aren't too different for a maximum signal.

You may find that your receiver benefits from the insertion of an aerial tuning unit (ATU) between the aerial and the input, to compensate for the impedance of your aerial not being 50 Ω. A design for such an ATU is presented in another part of this series. If the signals are still weak, connect the ATU to the junction of C1 and C2 via a 100 pF capacitor.

Try listening!

Remember that 80 m is a variable band. During daylight hours, your will hear Morse signals at the lower end of the band, and some British and closer continental stations between 3.7 and 3.8 MHz. In the evenings, stations up to 1000 miles away should be heard. Look for Novices around 3.7 MHz!

Parts list – VFO board

Resistors: all 0.25 watt, 5% tolerance		
R1	100 kilohms (kΩ)	
R2, R5, R6	100 ohms (Ω)	
R3, R4	10 kilohms (kΩ)	
RV1	1 kilohm (kΩ) linear	
Capacitors		
C1	12 picofarads (pF) polystyrene	
C2	100 picofarads (pF) polystyrene	
C3	470 picofarads (pF) polystyrene	
C4, C5	1 nanofarad (nF) polystyrene	
C6, C8, C9	10 nanofarads (nF) polystyrene	
C7	100 picofarads (pF) min. ceramic	
VC1	140 + 140 picofarads (pF) variable	
Semiconductors		
TR1	MPF102	FET
TR2	BC182	npn
D1	1N914	silicon
Inductors		
L1	Toko KANK3334	
RFC	1 mH	RF choke

Parts list – mixer board

Resistors: 0.25 watt, 5% tolerance

R1	1 kilohm (kΩ)
VR1	1 kilohm (kΩ) potentiometer (linear)

Capacitors

C1, C3	47 picofarads (pF) min. ceramic
C2	3.3 picofarads (pF) min. ceramic
C4	100 picofarads (pF) min. ceramic
C5	100 nanofarads (nF) min. ceramic
C6, C7	10 nanofarads (nF) min. ceramic
C8	220 microfarads (μF) electrolytic 16 V
C9	1 microfarad (μF) electrolytic 16 V

Integrated circuit

IC1	Philips NE602 or NE602A

Additional items

T1/T2	Toko KANK3333
	On/off switch
	Miniature toggle switch

The next part . . .

The IF amplifier and the Beat-Frequency Oscillator will be added to convert the *Colt* into a superheterodyne receiver.

26 A two-way Morse practice system

Introduction

It is said that the pleasure of sending and receiving Morse code more than compensates for the learning effort needed. Transmitters for Morse code can be much simpler than those needed for any of the speech modes.

Many people choose to learn sending and receiving on their own. It can be much more fun if you have someone to learn with you, and it is for this reason that the following project arose. It comprises just one small circuit, and you build an identical circuit for each of the people who want to learn with you. All the individual circuits are connected with their output leads in *parallel*, as the two-way circuit of **Figure 1** shows.

A simple circuit

This diagram shows two identical circuits, as would be used if two people wanted to learn together. The circuit centres around our old friend the NE555 integrated circuit (IC). As it is connected here, it works as an *astable multivibrator*, a daunting name for what is essentially an oscillator. Each circuit is self-contained, having its own battery, Morse key, sounder and

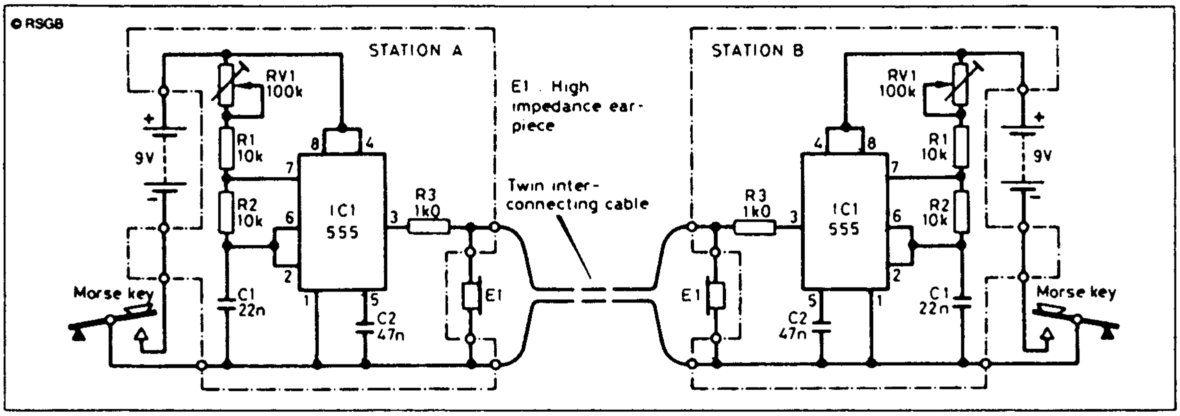
© RSGB
STATION A
RV1
100k
R1
10k
9V
+
R2
10k
8
4
7
6
2
1
5
3
IC1
555
R3
1kΩ
C1
22n
C2
47n
Morse key
E1
E1 . High impedance ear-piece
Twin inter-connecting cable
STATION B

Figure 1 The circuit diagram shows two stations, but you could connect several more if needed

plastic box. The Morse key makes and breaks the power supply to the circuit, thus turning on and off the note emitted by the earpiece. The frequency of the note is controlled by the variable resistor RV1, resistors R1 and R2, and the capacitor C1. The range given by RV1 is approximately from 500 Hz to 2200 Hz.

R3 protects the output of the IC against accidental shorting either to the 0 V or to the 9 V supply rails. For best results, a crystal earpiece or high-impedance headphones should be used. You could try Walkman-type headphones, but the volume may be too low for you.

The units may be interconnected with thin twin cable (the sort used for wiring doorbells) and can be comfortably separated by 25 m, more than enough to communicate between rooms, or even with a neighbour!

In use

When an operator presses his key, the note is heard by himself and by everyone else who is connected. If more than one key is pressed at the same time, two or more notes are heard by everyone! It helps if each person adjusts his own potentiometer, RV1, to give a note which is different from the others. When one operator stops sending, the other can start immediately, without pausing to change from transmit to receive. In practice, this technique is known as *full break-in*.

The current drawn by the circuit is only 10 milli-amps (mA) when the key is down, meaning that the life of a typical PP3 battery will be virtually its shelf-life! No switch is needed, because the key acts as the switch. The current drain will be even less if the NE555 is replaced by an ICM7555 IC. It has exactly the same pin connections as the NE555, so no circuit modifications are needed.

The circuit board

The circuit is constructed easily on a small piece of Veroboard of the copper-strip type. **Figure 2** shows the layout of the prototype, the board measuring 18 holes by 10 strips. **Be aware that there is no row 'I' when you are transferring mental images of where the parts are to real positions on the board!** Break the copper tracks where shown with a 3 mm (⅛ inch) twist drill, rotated carefully between thumb and forefinger. Hold the board up to a bright light to make sure that the tracks are completely broken and that there are no fragments of copper swarf shorting adjacent tracks. Then, solder in the wire links shown in Figure 2, followed by the resistors, RV1, and the capacitors, C1 and C2. Solder the battery connector leads to tracks A and G, G being positive. Use insulated wire to connect from the board to the jack sockets used as connectors for the key and the earphones. If you use

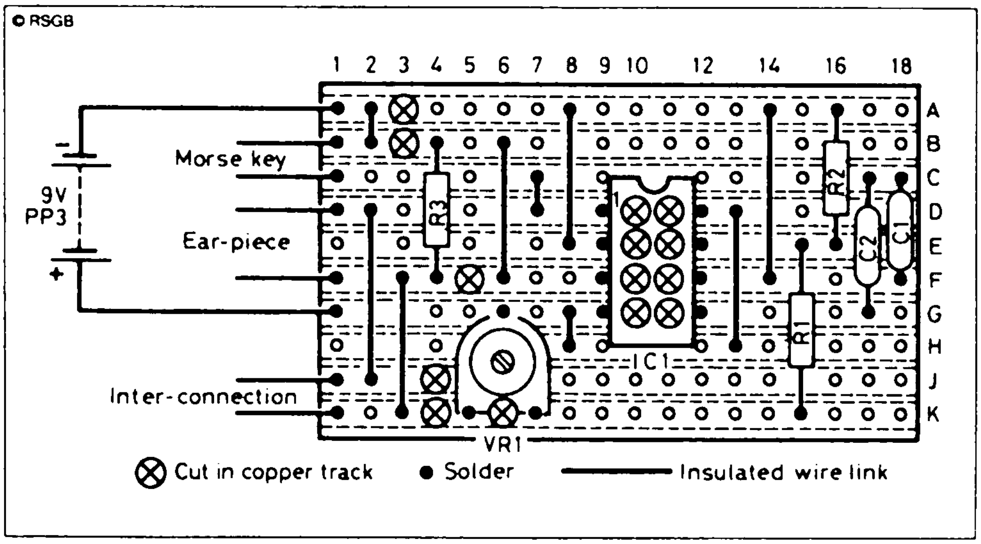
© RSGB
1 2 3 4 5 6 7 8 9 10 12 14 16 18
A B C D E F G H J K
9V PP3
–
+
Morse key
Ear-piece
Inter-connection
R3
R2
R1
C2
C1
IC1
1
VR1
Cut in copper track
Solder
Insulated wire link

Figure 2 The Morse Duet is an easy project to build on stripboard (Veroboard)

different sized jack sockets (i.e. 2.5 mm and 3.5 mm) you won't plug the earphones into the key socket! A two-screw terminal strip is useful to connect the cable runs between operators.

The key

The circuit is designed to be used with a *straight key* (one that moves up and down). It is a requirement of the UK Morse test that a straight key is used, so it is very sensible to learn sending with a straight key before you try anything more complex! You will find many to choose from at rallies.

Parts list

For each board, you will need:		**Maplin order codes**
Resistors: all 0.25 watt, 5% tolerance		
R1, R2	10 kilohms (kΩ)	M10K
R3	1 kilohm (kΩ)	M1K0
RV1	100 kilohm (kΩ) min. preset (horizontal mounting)	UH06G
Capacitors		
C1	22 nanofarads (nF) ceramic, 25 V	WX78K
C2	47 nanofarads (nF) ceramic, 25 V	RA47B

Integrated circuit	
IC1 NE555	QH66W
Additional items	
8-pin DIL socket	BL17T
Jack plug and socket 2.5 mm	HF76H, HF78K
Jack plug and socket 3.5 mm	HF80B, HF82D
Battery type PP3	
Battery connector	HF28F
Plastic box with lid 114 × 76 × 38 mm	LH14Q
Veroboard 0.1 inch with copper strips 18 holes × 10 strips	JP46A
Terminal strip (two section)	FE78K
Crystal earpiece	LB25C
Morse key	

27 The Colt 80 m receiver – Part 4

Introduction

In Parts 1 to 3 of this series, the design of an 80 metre direct-conversion receiver has been described. In this final part, we are going to change the circuit to operate as a superheterodyne receiver, or *superhet*. Most radio receivers are superhets, and they change the incoming signal to another frequency, known as the *intermediate frequency*, or IF, before producing an audio signal. The use of a superhet in a good receiver is mandated by the requirements for good *sensitivity* and *selectivity.*

Sensitivity and selectivity

Figure 1 shows the block diagram of the receiver. If you look closely, you will see that it is two similar circuits, one after the other. The incoming

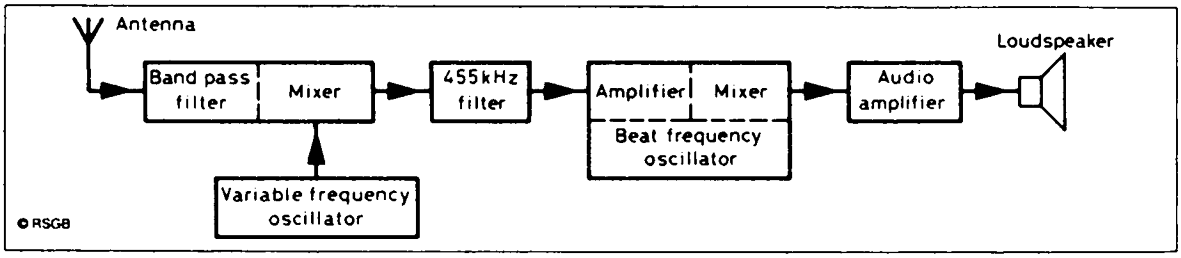

Figure 1 This is how the different stages of the Colt go together to make a superhet receiver

signal is filtered and fed to a mixer where it is combined (mixed) with the signal from the variable-frequency oscillator (VFO). This oscillator operates at a frequency which is 455 kHz higher than the incoming signal from the aerial, and the mixer output is therefore at a frequency of 455 kHz. If you are not sure about this, please refer to the section 'The direct conversion process' in Part 3 of this project. This new frequency is called the *intermediate frequency*, or IF. This frequency doesn't change; the tuning is accomplished by the VFO, and the mixer output is *always* at 455 kHz. The extraction of the audio signal from the IF signal is identical with the direct conversion process which is used in your existing receiver.

This may seem a long-winded way of doing things, but it has its advantages. A receiver must have a good *sensitivity*, or gain, so that it can receive very weak signals. In very general terms, it is easier to handle low-frequency signals than it is to handle high-frequency signals. We are changing our signal frequency from around 3.6 MHz down to 0.455 MHz (455 kHz), which is much lower and can be filtered and amplified relatively easily. A second advantage is that it is easier to provide gain at a fixed frequency than at a variable frequency. Remember that the IF is fixed, and providing gain is, again, relatively simple.

Our receiver also needs good *selectivity*, the ability to separate (or select) one station from another very close to it in frequency. This requirement is significantly simplified by the fact that the IF is fixed, and a good filter in the IF circuits can do wonders for the rejection of adjacent-frequency stations! Several stages of IF amplification and filtering are possible in more adventurous designs.

The filtered signal at 455 kHz passes to another mixer which has an associated oscillator, usually called a *beat frequency oscillator* (BFO). When receiving CW (Morse) signals, the BFO is usually tuned about 1 kHz above or below the IF (i.e. at 454 or 456 kHz) to produce a 1 kHz beat note as the audio signal. This signal is then amplified and fed to a loudspeaker or headphones.

The circuit

Figure 2 shows the circuit of the IF section shown in the photograph on p. 94; this section is added to the existing circuit to make it a superhet. The existing mixer board (described in Part 3) is used as the first mixer. Between it and the audio amplifier is connected the new IF board.

The signal output from the first mixer (at 455 kHz IF) is fed into a *crystal filter*, the most expensive part in the whole receiver. It provides the selectivity which makes the Colt such a good receiver. Another NE602 mixer/oscillator chip follows the filter. The oscillator section is controlled by the tuned circuit in T1, the frequency of which can be altered by rotating the core inside the coil. Once set, it remains fixed.

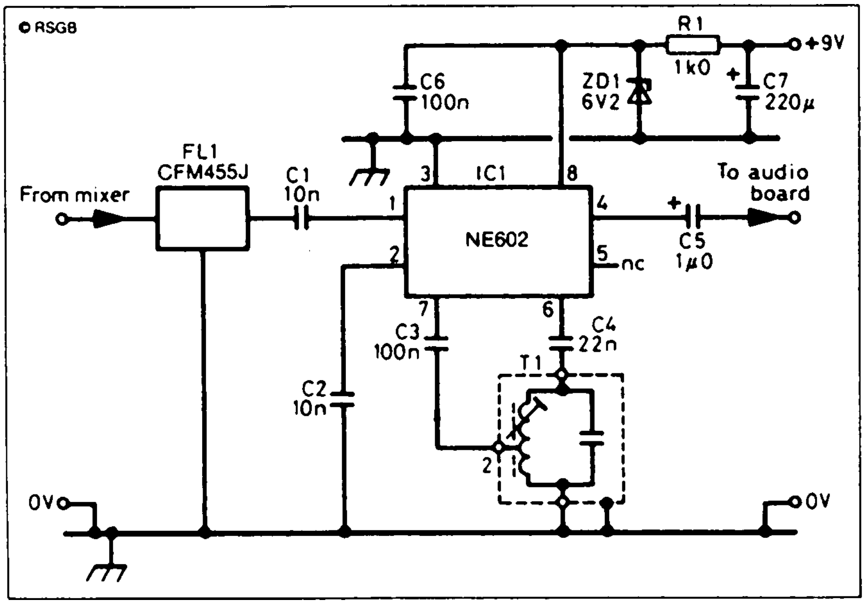

Figure 2 The IF board is connected between the mixer and audio amplifier

At this stage you should have built the first mixer and the audio amplifier, and proved that they *both* work by using the circuit as a direct-conversion receiver. When you have finished constructing and checking the IF board, you will need to add it to your existing circuits.

Adjusting and testing

There are three pairs of connections to the IF board – the 9 V supply leads, the input leads and the output leads. The IF board should be mounted on the metal baseplate, along with the other boards. **In making the following connections, make absolutely sure that the braid of each piece of coaxial**

cable is soldered to the correct connection on each board. The same applies to the polarity of the 9 V leads. Disconnect the screened lead which presently goes from the first mixer board to the AF amplifier at the amplifier end. Connect it instead to the *input* of the IF board. Connect a new piece of screened cable from the IF output to the vacated AF input. Now connect the 9 V supply leads to the same points as the supply leads from the other boards. Check these newly made connections.

You should have confidence at this point that things should be right. After all, you *have* tested the direct-conversion process and you *know* it works. All that you are now testing is the IF board. This is the attraction of building a receiver in modules, working from the speaker backwards, and testing as you go!

The VFO and BFO need to be correctly adjusted. Mesh the vanes of the tuning capacitor and, using the same frequency measurement method as you did originally, set the VFO frequency to 3.955 MHz (which is 3.500 MHz + 0.455 MHz, if you hadn't guessed!). The BFO can be set using a frequency counter, but it is just as good to set it by listening to SSB or CW signals. A high-pitched hissing sound should be heard in the speaker. As you rotate the core, the pitch should reduce, go through a minimum, then increase again. Set the core at the minimum pitch position. You may want to readjust the two cores a little as your listening skills improve but, once you are happy, they will never need to be altered again!

In conclusion . . .

You should now have built a superhet receiver capable of excellent results. It uses the same type of circuit as that found in far more advanced receivers.

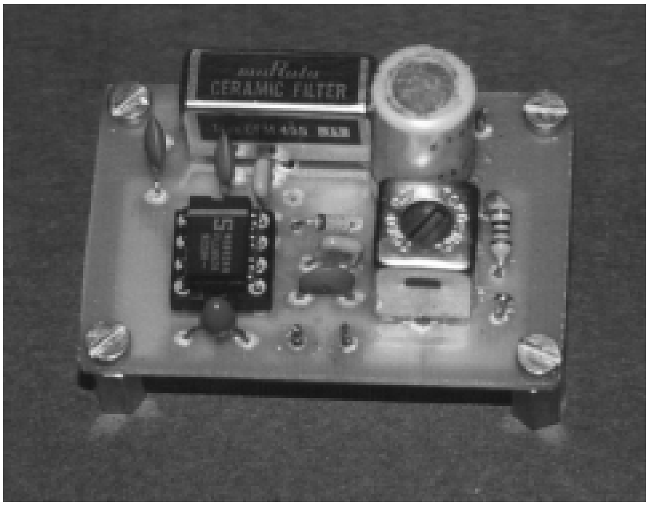

The superhet is far more sensitive than the direct-conversion type, and can weed out those elusive DX stations. You may have found that a station will appear at two places on the dial of the direct-conversion receiver; you will have no such problem with the superhet.

Parts list

Resistor

R1	1 kilohm (kΩ), 0.25 watt, 5% tolerance

Capacitors

C1, C2	10 nanofarads (nF) min. ceramic
C3, C6	100 nanofarads (nF) min. ceramic
C4	22 nanofarads (nF) min. ceramic
C5	1 microfarad (μF) 16 V electrolytic
C7	220 microfarad (μF) 16 V electrolytic

Integrated circuit

IC1	NE602

Additional items

ZD1	Zener, 6.2 V 0.5 watt
FL1	Crystal filter, Murata CFM455J
T1	Tuned inductor, Toko YHCS11100AC

28 A simple crystal set

Introduction

There are many designs of crystal set – they are all ‘simple’, and even the most seasoned radio amateur will build one of these every so often because it is something that never ceases to amaze! Although using modern components, the design is a period piece!

Design and construction

The circuit diagram of **Figure 1** shows the simplicity of the circuit. The prototype was constructed on a wooden baseboard with an aluminium front panel, although this could be made of wood if you prefer.

Winding the coil always causes the most groans but when finished, gives the most satisfaction. The construction of the coil is detailed in **Figure 2**. A discarded toilet roll centre is an ideal *former* on which to wind the coil. The coil will take up about 7 cm of the length of the tube, so assuming the tube is about 11 cm long, the ends of the coil will be about 2 cm in from each end of the tube.

The aerial coil, the *primary winding* of this radio-frequency (RF) transformer, is made from about 30 turns of 26 SWG enamelled copper wire, wound on the matchstick ribs. These ribs should be about 2 cm long, and are glued on **top** of the end of the secondary coil *after* the secondary coil has been wound. The *secondary winding* has 140 turns with *taps* every 10 turns for 70 turns. This coil covers most of the length of the cardboard tube. To make your coil-winding easier, here are some tips for completing the coil and still having some hair left at the end!

- Support the reel of wire on a dowel rod or pencil, clamped in a vice or held rigid in a vertical position by some other means. Don't leave it to trail around on the floor.
- Have a small piece of sandpaper handy to remove the enamel from the copper wire, starting with the end coming from the reel. Remove about 1 cm only, until the shiny copper is visible all round the end of the wire.

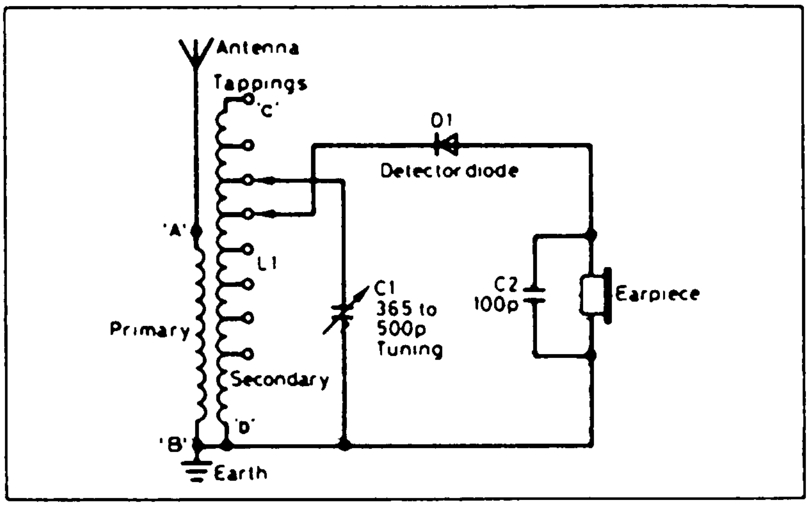

Figure 1 Circuit diagram of the crystal set

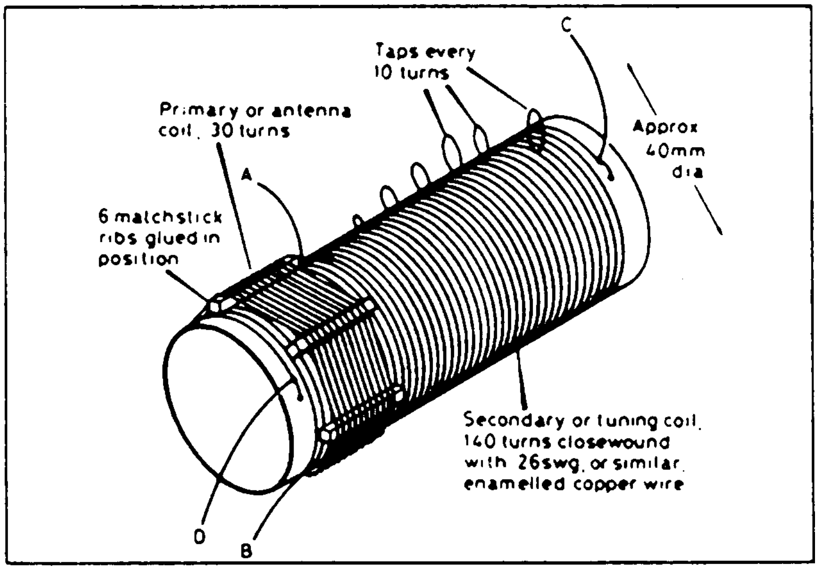
C
Taps every 10 turns
Primary or antenna coil, 30 turns
Approx 40mm dia
A
6 matchstick ribs glued in position
Secondary or tuning coil, 140 turns closewound with 26swg or similar enamelled copper wire
D
B

Figure 2 The coil is mounted on a cardboard tube as shown. Exact size is not too important. Try different taps for best results

- This part is the most difficult of all. Seasoned coil winders may have their own favourite ways of doing this, but for first-timers, play it safe and do it this way! Have a second bobbin (or a piece of stout card) ready, on which you will need to wind the wire loosely while you prepare the taps. Then, starting from the free end of the wire, measure 125 cm and, at this point, remove about 2 cm of the enamel with the sandpaper. Bend the wire firmly back on itself in the middle of the bared area, and *tin* (cover with solder) the exposed copper. This is your first tap. Then repeat this six more times at 125 cm intervals, thus giving you seven taps; wrap the wire as you go, on to the second bobbin or piece of card. After the seven taps, you are half-way down your coil so, to begin winding it in earnest, wind the wire carefully from the second bobbin or card **back** on to the original reel of wire.
- Looking at Figure 2, make two small holes about 3 mm apart, about 2 cm down from the top of the tube, where wire C will be entering. Poke the end of the wire into one hole, then bring it out again through the other; leave about 10 cm of wire on the free end. Loop this wire in and out once more, thus anchoring the wire firmly.
- Then, begin winding; keep the wire tight and make adjacent turns touch neatly. As you pass each tap, make sure it sticks outwards while avoiding flexing the wire too much; the wire is inherently weak at each tap.
- When you reach the last 10 cm of wire, stop. Put two more holes, like the first, beside your stopping point, and anchor the end, D, of the coil in the same way as you did with end C. The secondary winding is complete. Take a break *after* you have completed the next step!

- Prepare the six matchsticks for supporting the primary coil and glue these every 60 degrees around the end of the coil at D. Remove the enamel from the free end of the new copper wire on your reel and, using your favourite quick-drying cement, put a drop over the wire on one of the matchsticks, leaving 10 cm free as before. If you are using 26 SWG enamelled copper wire, its diameter is 0.46 mm, so 30 turns should occupy 14–15 mm of the 20 mm matchstick length. This will allow you to position the glued point such that the coil will be roughly in the centre of the matchsticks. When the glue has set, wind the 30 turns and anchor this end of the coil in the same way. Then remove the enamel at the end of the 10 cm lead.

Assembly

The coil former can be screwed to the baseboard at each end, *before* the front panel is screwed to the baseboard! As the photograph on the next page shows, the tuning capacitor is mounted in the centre of the front panel. Five solder tags (or drawing pins as a last resort) should be screwed into the baseboard. The two to the left of the photograph are the connections to the earphones, the two along the rear edge have the diode, D1, soldered to them. One end of the diode (and, for once, you can connect the diode either way round!) is connected to one earphone tag. The other end has a flying lead of about 10 cm attached to it and terminated in a crocodile clip which is used to connect to one of the taps on the coil. The fifth tag secures the end, A, of the primary coil, to which you will attach the aerial. The earth tag on the tuning capacitor, C1, serves as an anchor point for the ground connections to B and D and to the other earphone tag.

Use

The longest piece of wire you have available to use as an aerial should be connected to the aerial tag just mentioned. If you can connect the earth tag of the tuning capacitor to a good electrical earth, this will help also. You should be able to hear something as you turn the tuning knob. Try adjusting the tapping connections on the coil – change only one at a time, or you will never find the optimum positions. You now have a fully operational crystal set!

Simple it may be, but this circuit illustrates some important principles which are used even in the most expensive receivers. Firstly, the coil, L1, and the capacitor, C1, form a *tuned circuit* which *resonates* at the frequency of the station you have tuned in. This *selects* the signal you want to hear. The *modulation* on this signal is removed by the *detector*, D1, and fed to the earphones, which act as *transducers*, turning the electrical energy into sound energy which you can hear. It doesn't need a battery or other power supply either!

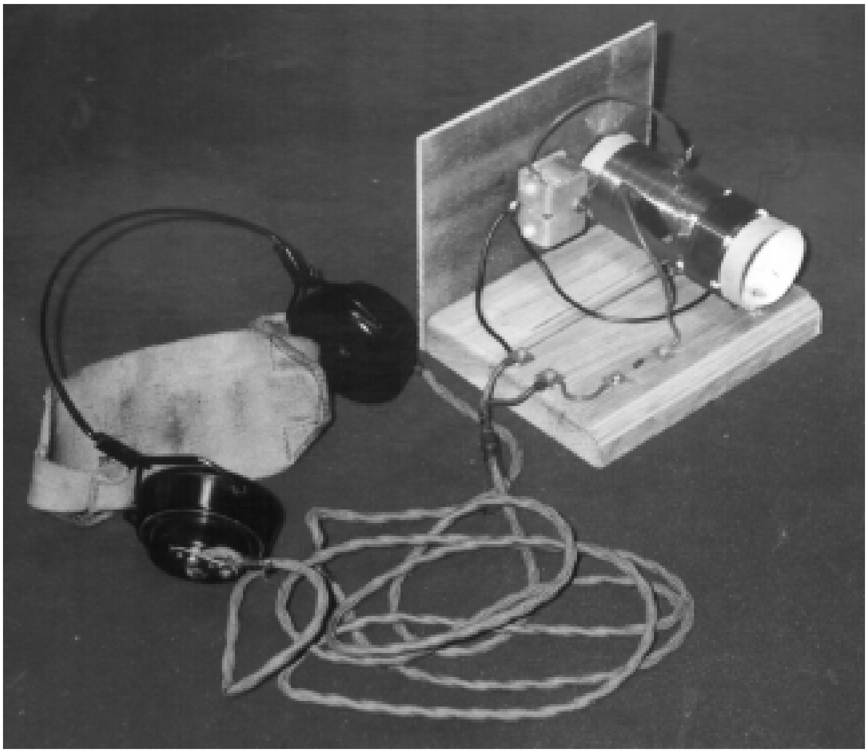

Parts list

Capacitors

C1	Variable capacitor of between 200 and 500 picofarads (pF) maximum
C2	100 picofarads (pF) min. ceramic

Additional items

D1	Germanium, type OA47 or OA90
	High impedance crystal earphone (not Walkman type)
	Tuning knob
	Wooden base, approx. 110 mm square
	Aluminium (or wood) front panel, approx. 80 × 100 mm
	Cardboard tube (toilet roll centre), about 110 mm long, 40 mm diameter
	Reel of 26 SWG enamelled copper wire for both windings of L1

29 A crystal calibrator

Introduction

A crystal calibrator is a device that produces oscillations of a precise frequency which are rich in *harmonics*. A harmonic is an integer multiple of the fundamental frequency; for example, if our fundamental frequency was 100 kHz (as it is in this circuit), there would be harmonics of this frequency at 200 kHz, 300 kHz, 400 kHz, and so on. These harmonics can be used, when picked up by any receiver, to calibrate that receiver, as the harmonics are quite accurate in frequency (see later for an assessment of accuracy). However, a gap of 100 kHz between harmonics is rather wide for most purposes, so we reduce this frequency to 25 kHz, so that the harmonics are then 25 kHz apart, thus producing a much more useful set of *marker points*. A calibrator such as this is often called a *crystal marker*, producing these marker points from 25 kHz to beyond 30 MHz.

The circuit

The circuit diagram is shown in **Figure 1**. The circuit around TR1 is the fundamental oscillator, and its frequency is controlled by the quartz crystal, X1. Even crystal oscillators are not 100% accurate, and the small trimmer capacitor, TC1, is able to ‘pull’ the frequency to one which is nearer the

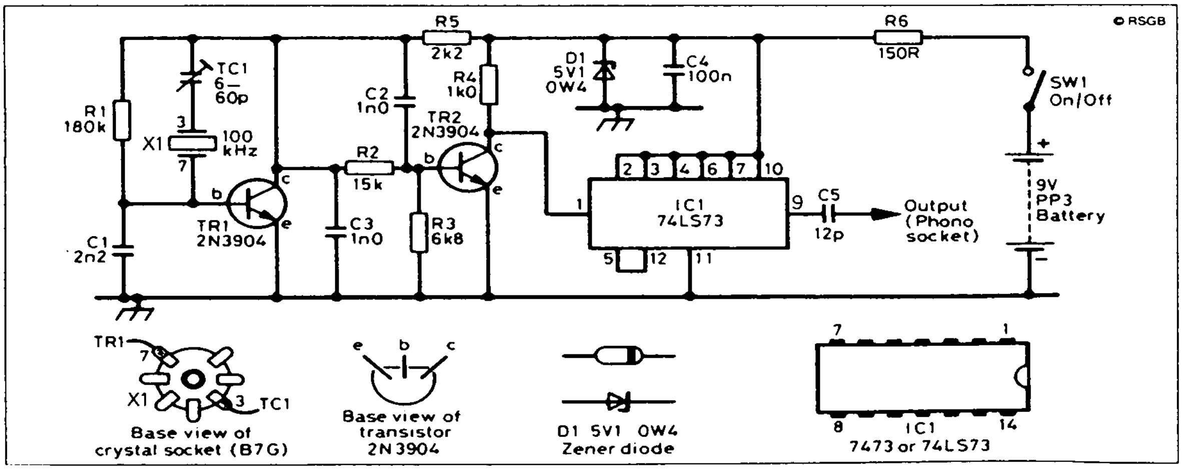
R5
2k2
R6
150R
© RSGB
TC1
6–
60p
R1
180k
X1
3
7
100
kHz
C2
1n0
R4
1k0
TR2
2N3904
D1
5V1
0W4
C4
100n
SW1
On/Off
R2
15k
b
c
e
TR1
2N3904
C1
2n2
C3
1n0
R3
6k8
2 3 4 6 7 10
1
IC1
74LS73
9
5 12 11
C5
12p
Output
(Phono
socket)
+
9V
PP3
Battery
−
TR1
7
X1
3
TC1
Base view of
crystal socket (B7G)
e b c
Base view of
transistor
2N3904
D1 5V1 0W4
Zener diode
7
1
8
14
IC1
7473 or 74LS73

Figure 1 Circuit diagram of the calibrator which uses easily obtainable components

correct one. TR2 is a *buffer* stage, which isolates the oscillator from the rest of the circuit. It acts as a switch, and applies a good signal (switching between 0 and 5 V) to the input of IC1.

IC1 is an integrated circuit which can be connected to do several things. It is connected here to divide the incoming frequency by a factor of 4, producing on pin 9 an output frequency of 25 kHz.

The combination of R6, C4 and D1 produces a supply of 5 V for IC1; it would be damaged if the battery voltage were applied to it. You will no doubt be ready to assemble the circuit, so here is some information for you.

Construction

If possible, always build a circuit in individual stages, which you can test as you go along. It is not always easy in small projects like this, but even the crystal calibrator can be split into two for construction and testing.

It is an ideal project for assembly on Veroboard of the copper-strip type; the prototype layout is shown in **Figure 2**, on a board measuring 11 holes by 24 strips. First of all, remove the copper strip at the locations shown, using a 3 mm drill rotated between the fingers. Hold the board up to the light to ensure that there are no pieces of copper swarf bridging adjacent strips, and that the copper is completely removed where it should be! Assemble the circuit from left to right, but do not wire up anything around IC1. Be aware that the diagram shows the board *from the component side*. Although you

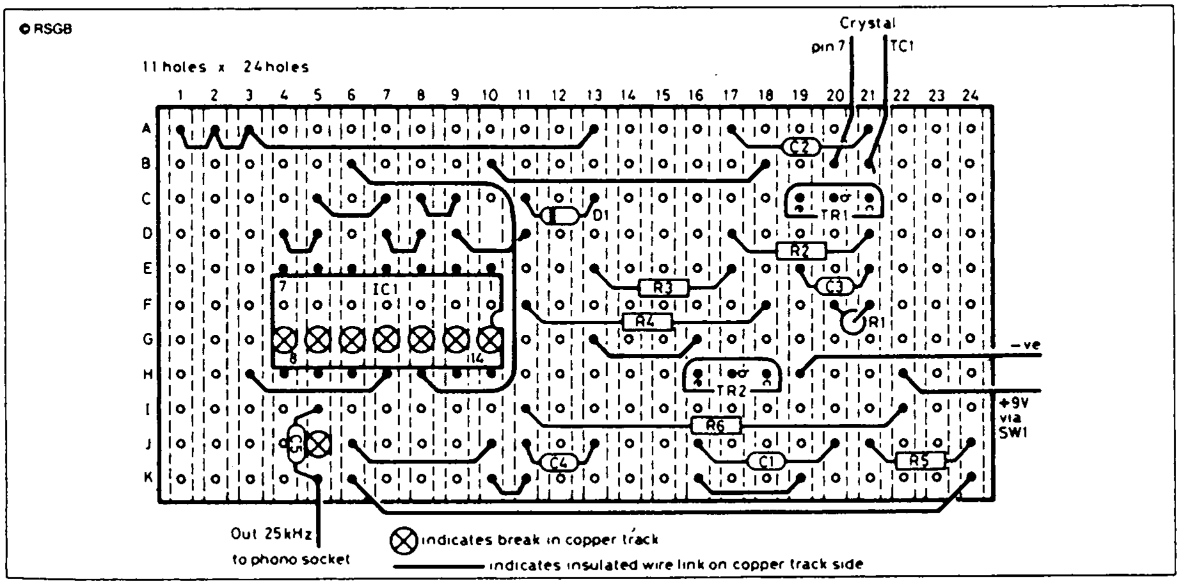
© RSGB
Crystal
pin 7
TC1
11 holes x 24 holes
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
A B C D E F G H I J K
C2
TR1
D1
R2
R3
C3
IC1
7
8
14
R4
R1
TR2
R6
C1
C4
R5
−ve
+9V via SW1
Out 25 kHz to phono socket
⊗ indicates break in copper track
indicates insulated wire link on copper track side

Figure 2 Stripboard, such as Veroboard, should be cut as shown to approx. 62 × 28 mm

can use virtually any kind of crystal, you will need to mount it firmly. If you use a small one, it can be soldered directly in position on the board. With one in a valve envelope, you will need a B7G valveholder. Be prepared to mount the circuit in a **metal** container; the prototype used an (empty) tin of tuna! The valveholder mounts well on any metal case.

Disconnect the aerial of your receiver, and replace it with about 30 cm of wire laid near your oscillator circuit. Connect the battery and switch on. Tune the receiver around until you can hear a whistle on SSB. Rotate the tuning knob to reduce the frequency of the whistle and, as the whistle becomes too low to hear, you have reached one of the calibration markers, and your frequency will be an integer multiple of 100 kHz. If you are already using a calibrated receiver, you will be able to verify this. Going up or down in frequency should locate another marker 100 kHz away, and so on right through the receiver's tuning range. If the signals coming from the circuit are very weak, switch off the calibrator and connect another 30 cm piece of insulated wire to the collector of TR2, and lay it close to (but *not* touching) the wire from the aerial connector. Switch on an try again; you should not now have a problem! If you have access to a multimeter, check that the voltage at the collector of TR1 is very close to 5 V. If it is close to 9 V, you have connected the diode, D1, the wrong way round!

Having verified that the oscillator is working, you can now wire up the integrated circuit socket, being careful to put the notched end of the socket pointing towards R4, the collector resistor for TR2.

Check your connections around IC1, and when you are satisfied that they are correct, line up IC1 with its socket, making sure that the notched ends are together, and press down *gently* to insert the IC into its socket. Insert the 30 cm piece of insulated wire into the output socket, connect the battery and switch on. You should still hear whistles in your receiver, but now they should be 25 kHz apart, rather than 100 kHz apart.

All that now remains to be done is to mount the circuit rigidly inside whatever casing your have chosen. Make sure that none of the connections under the board touch the metal case, and secure the valveholder, on/off switch and output connector to the case. You now have a completed crystal calibrator.

Calibration

The simplest way to calibrate your circuit is with a frequency counter. Most clubs will have one of these and, if not, will know someone who has! Connect it to the collector of TR2, where the frequency should be 100 kHz. Do *not* connect it to any part of the circuit around TR1, or you may alter the frequency you are trying to measure! If the frequency is not exactly 100.000 kHz, rotate TC1 until it is (or is as close as you can get it). Now your calibrator is as accurate as the counter with which you have calibrated it.

Accuracy

Despite your best efforts at calibration, by whatever means, your crystal will *never* have a constant frequency. Such a thing is a scientific impossibility. It is usual to express the accuracy of a crystal in parts-per-million (ppm), and it is governed by many things, principally its temperature. You will be *very* fortunate if your circuit maintains an accuracy better than about ±10 ppm. Expressed in figures, it means that the true frequency can be anywhere between 99.999 kHz and 100.001 kHz, i.e. within 1 Hz of the correct frequency.

Although this may seem more than adequate, it is as well to remember that the accuracy degrades as the frequency increases. At 1 MHz the error will be ±10 Hz and at 10 MHz it will be ±100 Hz. At 30 MHz it will be 300 Hz. Even so, this should be acceptable for most non-critical applications.

Parts list

Resistors: all 0.25 watt, 5% tolerance		**Maplin code**
R1	180 kilohms (kΩ)	M180K
R2	15 kilohms (kΩ)	M15K
R3	6.8 kilohms (kΩ)	M6K8
R4	1 kilohm (kΩ)	M1K0
R5	2 kilohms (kΩ)	M2K0
R6	150 ohms (Ω)	M150R
Capacitors		
C1, C2, C3	1 nanofarad (nF) or 1000 picofarads (pF)	WX68Y
C4	100 nanofarads or 0.1 microfarad (μF)	YR75S
C5	12 picofarads (pF)	WX45Y
TC1	60 picofarads (pF) trimmer	WL72P
Semiconductors		
TR1, TR2	2N3904 npn	QR28F
IC1	7473 or 74LS73N dual JK flipflop	YF30H
D1	5.1V 500 mW Zener	QH07H
Additional items		
100 kHz crystal		
B7G valveholder (found at most rallies)		
14-pin DIL IC socket		BL18U
Phono socket		YW06G
PP3 battery connector		HF28F
PP3 battery		
On/off switch		FH00A
Veroboard		JP47B
Metal case as required		

30 A simple short-wave receiver – Part 1

Introduction

No apologies for yet another design of short-wave radio. This is the beauty of our hobby – there is always something new to try. Not all circuits operate in the same way; not all circuits work equally well; sometimes a simple design suits the operator better than a complicated design. So here is another one for your consideration; it's a good project for a novice, *and* a good one-evening project for someone more experienced.

A basic description

Figure 1 gives the block diagram of the receiver, which employs a *regenerative detector*, one of the earliest techniques by which excellent *selectivity* (the ability to separate two stations very close in frequency) could be combined with good *sensitivity* (the ability to pick up very weak stations) using a very simple circuit. The receiver falls into the category known as *tuned radio frequency* (TRF), meaning that the whole circuit (prior to the extraction of the modulation from the carrier) operates at the incoming radio frequency. In other words, it is *not* a superhet.

As you can see, the regenerative detector has what is called a *feedback loop*, which feeds a small amount of the output signal back to the input. You have heard the effects of feedback with a public address system, when the microphone gain is too great – everything becomes very loud and then bursts into a deafening squeal! This is exactly what the feedback does here, except that it is carefully controlled, thus providing both gain and selectivity. The resulting audio is then amplified for use with headphones.

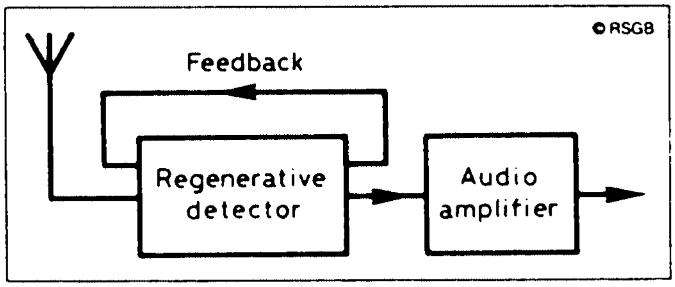

Figure 1 Just two simple stages make up the circuit

The circuit

The circuit of **Figure 2** shows the complete system. TR1 is an untuned *field-effect transistor* (FET) stage, and is used to match the aerial to the next stage. Occasionally, a regenerative detector produces unwanted signals, and TR1 also prevents them from reaching the aerial and being transmitted! The smaller of the two windings on T1 (the *primary* winding) will match a low-impedance aerial, the capacitor input matching a high-impedance aerial.

Ignore TR2 for the moment – the next stage in the signal path is TR3, the detector, using another FET. The tuned circuit is formed by T2, VC1 and VC2 (remember TR1 is untuned). The primary winding on T2 taking the output from TR1. The reason for having two variable tuning capacitors, one large, the other small, is that the large capacitor is the main tuning capacitor, while the small one is used as a *bandspread* control (for very fine tuning). The tuned signals are then detected (converted to audio frequencies) by TR3.

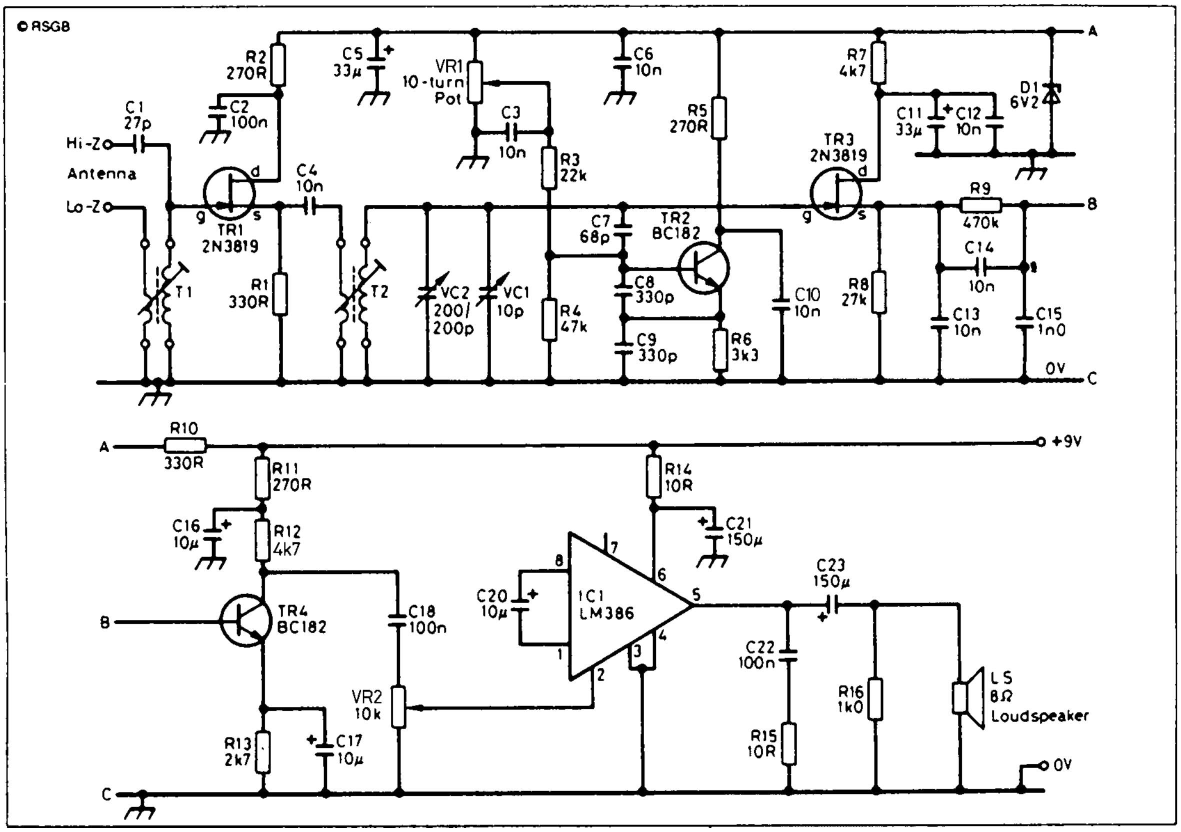
©RSGB
C1 27p
Hi-Z
Antenna
Lo-Z
T1
R2 270R
C2 100n
TR1 2N3819
d
g
s
R1 330R
C4 10n
T2
C5 33µ
VR1 10-turn Pot
C3 10n
R3 22k
VC2 200/200p
VC1 10p
R4 47k
C6 10n
C7 68p
TR2 BC182
C8 330p
C9 330p
R5 270R
R6 3k3
C10 10n
R7 4k7
TR3 2N3819
d
g
s
C11 33µ
C12 10n
D1 6V2
R8 27k
R9 470k
C14 10n
C13 10n
C15 1n0
A
B
0V
C
A
R10 330R
+9V
R11 270R
C16 10µ
R12 4k7
B
TR4 BC182
C18 100n
VR2 10k
R13 2k7
C17 10µ
C
C20 10µ
IC1 LM386
1
2
3
4
5
6
7
8
R14 10R
C21 150µ
C23 150µ
C22 100n
R15 10R
R16 1k0
LS 8Ω Loudspeaker
0V

Figure 2 The circuit diagram showing the 10-turn potentiometer VR1. This is the regeneration control for the circuit

TR2 is a *Q-multiplier* stage. It is a Colpitts-type oscillator which uses C8 and C9 to give feedback and produce the oscillation. C7 couples the tuned circuit (and the input to TR3) to TR2. The tuned circuit thus controls the frequency of oscillation of TR2 and the signal passed on to TR3.

The secret of the ease of operation of this receiver is VR2, a 10-turn potentiometer (or *helipot*). The resistance wire is wound in the form of a helix, giving a much greater wire length than in a normal potentiometer, and the shaft must be turned ten times to cover the whole length. Helipots are very useful when very fine adjustments have to be made. Here, VR2 sets the regeneration (or *reaction*) level, depending on the type of signal you are receiving, as will be discussed in Part 2.

TR4 provides the first stage of audio amplification and, after the volume control, VR2, the audio amplifier integrated circuit, IC1, will drive a small loudspeaker or headphones.

In Part 2, the construction will be discussed, together with the choice of aerial, the parts list, and advice on using the receiver.

31 A fruit-powered medium-wave radio

Introduction

This is a one-evening project that will result in a working medium-wave (MW) radio, and will also teach you a little about the way electricity can be generated from the right metals and a little (safe) acid. All you need are three lemons or other citrus fruit, three pieces of copper and three pieces of zinc (or galvanised metal) for your power supply.

Construction

Figure 1 shows the circuit and **Figure 2** its layout on a simple ‘plug-in’ prototype board. The six pieces of metal are connected as shown, to wire the three lemons in series; use ordinary wire between each lemon and the next. If you have a meter to measure the total voltage, it should be about 1.8 V. Use a standard ferrite rod, and wind on it about 40 turns of single-conductor PVC-insulated wire.

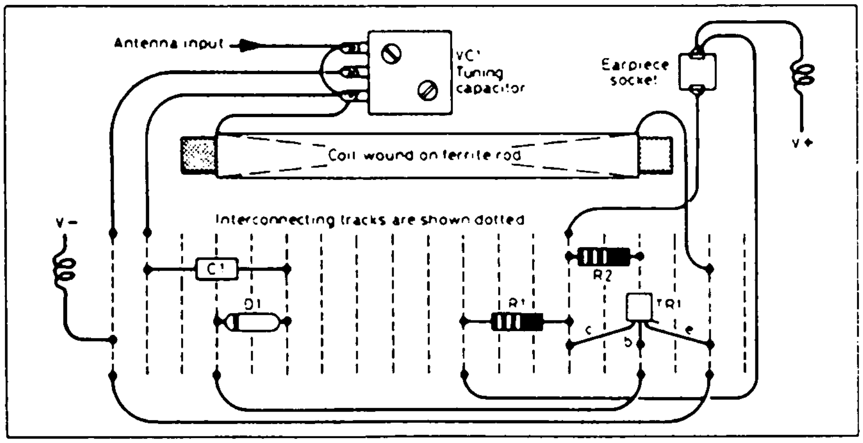
Antenna input
VC1
Tuning
capacitor
Earpiece
socket
V+
Coil wound on ferrite rod
Interconnecting tracks are shown dotted
V−
C1
D1
R1
R2
TR1
c
b
e

Figure 1 Most parts are plugged into the board as shown – soldering is not required

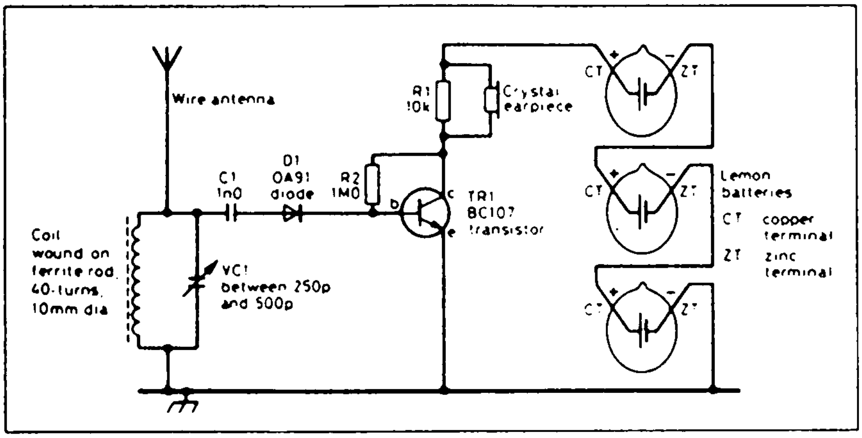
Wire antenna
R1
10k
Crystal
earpiece
CT
ZT
+
−
D1
C1
1n0
OA91
diode
R2
1M0
b
c
e
TR1
BC107
transistor
Lemon
batteries
CT copper terminal
ZT zinc terminal
Coil
wound on
ferrite rod,
40-turns,
10mm dia
VC1
between 250p
and 500p

Figure 2 Three lemons power the radio, which gives good results for such a simple circuit

Wire up the circuit on the board as illustrated in Figure 2. Soldering is not required with this type of board – just plug in the components and the wires. Only one transistor is needed. The tuning capacitor, VC1, selects the station you want to hear, and D1 helps to remove the carrier from the RF signal. The resulting audio signal is fed to TR1, a small transistor audio amplifier, which makes the signal big enough to drive a crystal earpiece comfortably. Walkman-type earphones will not work, so invest in a crystal earpiece which you can use in several other projects, too! If you use a smaller capacitor than that specified for VC1, you will need more turns on the aerial coil.

Tests on the prototype indicated that the radio will run for about a week on three lemons!

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1	1 megohm (MΩ)
R2	10 kilohm (kΩ)

Capacitors

C1	1 nanofarad (nF) min. ceramic
VC1	250–500 picofarad (pF) variable

Semiconductors

TR1	BC107 npn (or BC108, BC109C)
D1	OA90, OA91 germanium (not silicon)

Coil

L1	2 metres of single-conductor insulated wire on a standard ferrite rod

Additional items

- Plug-in prototype board, e.g. Maplin YR84F
- Wire aerial at least 3 m long
- Crystal earpiece
- Three juicy fruits

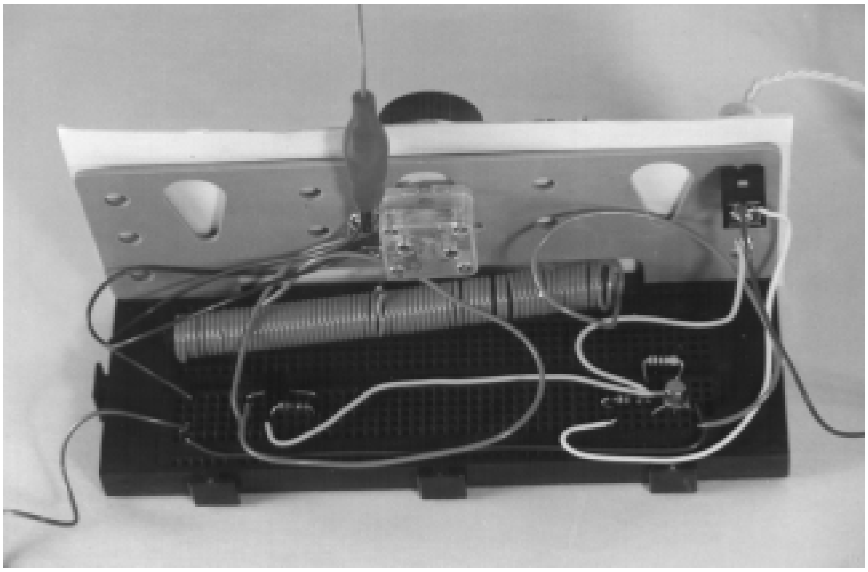

32 A capacitance bridge

Introduction

It is always valuable to have a means of measuring capacitors amongst your test gear. Universal LCR bridges (i.e. systems that can measure inductance, capacitance and resistance) can be found at most rallies, and you can work your way up to owning one of these. In the meantime, a piece of home-made equipment gives you experience as well as resulting in a useful measuring instrument.

The circuit described here is called a *capacitance bridge*, because it balances the effects of one resistor/capacitor pair against another; if one capacitor has an unknown value, then the other can be calculated. The basic bridge circuit is shown in **Figure 1**. To avoid having calculations to perform, this instrument will be calibrated by using capacitors of known values. The bridge is a useful way of performing measurements, because a knob is turned until there is a null in the signal from an earpiece or loudspeaker. The ear is very precise in being able to perceive nulls, which makes the bridge easy to use and reasonably accurate. At the null, the bridge is said to be *balanced*.

How does it work?

Figure 2 is the circuit diagram for this capacitance bridge. Transistors TR1 and TR2 form an oscillator. This is the audio oscillator shown in Figure 1, and produces an alternating voltage which is fed to the bridge. RV1 (in the collector lead of TR2) replaces both R1 and R2 in Figure 1 – that part of

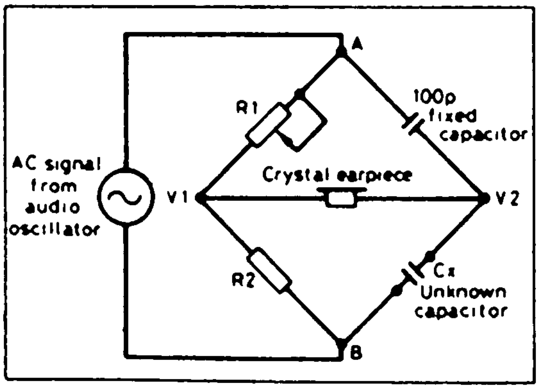

Figure 1 Simplified circuit of a capacitance bridge. R1 is adjusted for minimum sound

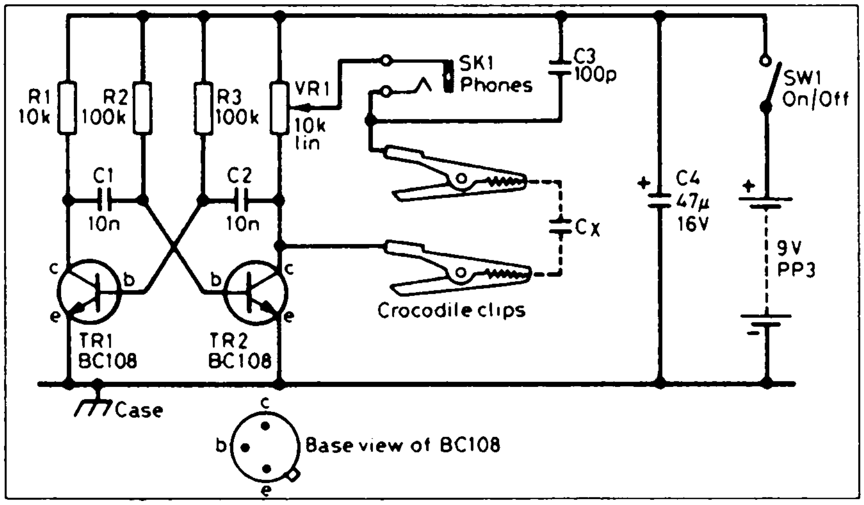
R1
10k
R2
100k
R3
100k
VR1
10k
lin
SK1
Phones
C3
100p
SW1
On/Off
C1
10n
C2
10n
Cx
C4
47µ
16V
9V
PP3
c
b
e
TR1
BC108
TR2
BC108
Crocodile clips
Case
Base view of BC108

Figure 2 Transistors TR1 and TR2 give an audio signal which is adjusted by variable resistor RV1

RV1 *above* the wiper represents R1, while that part *below* the wiper represents R2. The voltage on one side of the earpiece is determined by the ratio of these values, and is adjusted by rotating RV1. The voltage on the other side of the earpiece is determined by the ratio of C3 to Cx, where Cx is the unknown-valued capacitor. When these two voltages are the same, the bridge is balanced and there is no current through the earpiece.

Figure 3 shows the layout of the components on the matrix board of the prototype. It measures 10 holes by 14 holes, and is of the plain type, i.e. no copper strips. All earth connections are taken to a single solder tag. When

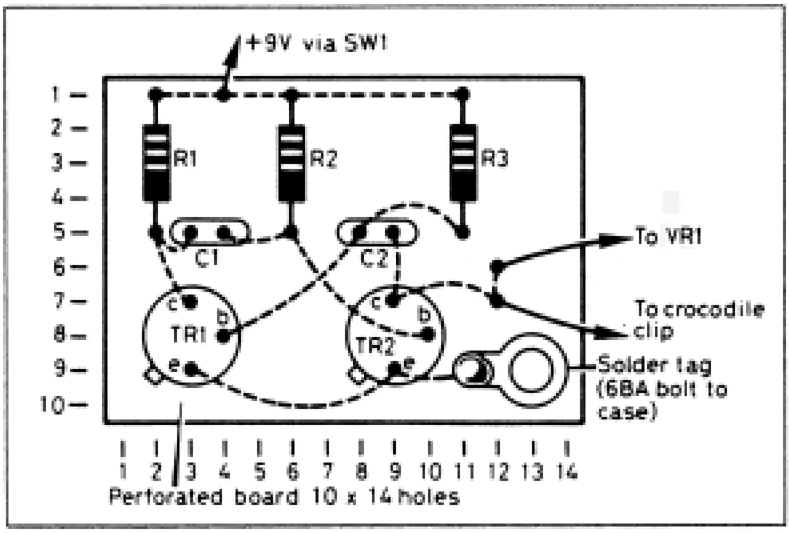
+9V via SW1
1
2
3
4
5
6
7
8
9
10
R1
R2
R3
C1
C2
TR1
TR2
c
b
e
To VR1
To crocodile clip
Solder tag (6BA bolt to case)
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Perforated board 10 x 14 holes

Figure 3 Components are soldered together on a small prototype board

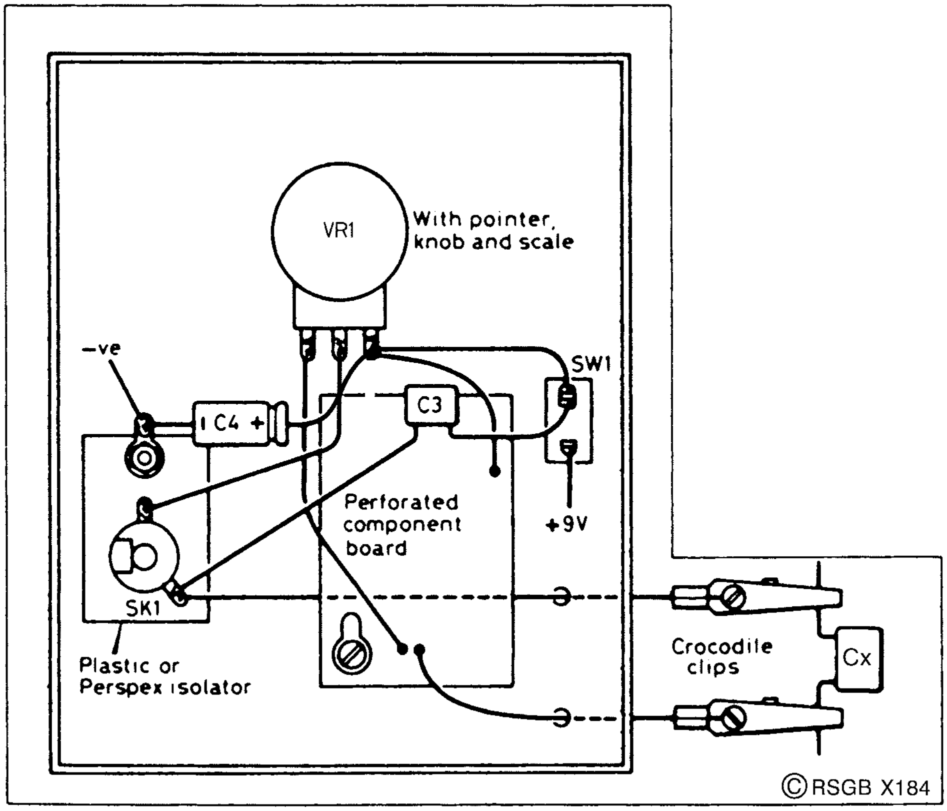
VR1
With pointer, knob and scale
−ve
C4 +
C3
SW1
Perforated component board
+9V
SK1
Plastic or Perspex isolator
Crocodile clips
Cx
©RSGB X184

Figure 4 Layout of the parts inside an aluminium box. The earpiece socket is insulated from the case

mounting the board inside a metal box, a long screw is used with extra nuts, to earth the board to the case while providing a stand-off, thus preventing any unwanted short-circuits between the board connections and the case. Be aware that a *crystal* earpiece must be used; low-impedance headphones of the Walkman type are not satisfactory. Solder two flying leads as shown in **Figure 4**, about 15 cm long, terminated in small crocodile clips for attaching to the unknown capacitor.

Calibration

After checking the circuit carefully, attach the battery and switch on; a buzzing noise should be heard in the earpiece. This is the first sign that everything should be OK. If there is no buzz, switch off and recheck the connections.

You will now need a range of close-tolerance (1%) silver-mica capacitors covering the range 10 to 1000 picofarads (pF). Arrange the capacitors in ascending order and connect them to your bridge in sequence. After having prepared a neat piece of card or paper mounted behind the knob on the front panel, mark the dial at the positions of the nulls for all the capacitors. You have now calibrated your capacitance bridge. If you are more likely to want to measure larger capacitors, replace C3 by a 1 nanofarad (nF) capacitor, and the bridge will measure up to 10 nF approximately.

Parts list

Resistors: all 0.25 watt, 5% tolerance	
R1	10 kilohms (kΩ)
R2, R3	100 kilohms (kΩ)
RV1	10 kilohms (kΩ) linear
Capacitors	
C1, C2	10 nanofarads (nF) ceramic
C3	100 picofarads (pF) silver mica or polystyrene
C4	47 microfarads (μF) 16 V electrolytic
Semiconductors	
TR1, TR2	BC108 npn
Additional items	
SW1	SPST on/off switch
	Battery connector, PP3 type
	Earpiece, crystal type
SK1	3.5 mm jack socket for earpiece
	Battery, 9 V PP3
	Matrix board, approx. 10 holes by 14 holes
	Aluminium case, approx. 10 × 8 × 5 cm

33 A simple short-wave receiver – Part 2

Introduction

In Part 1, the design of this receiver was discussed in some detail. Now we are going to put it all together and see how it works. The receiver is laid out on a printed-circuit board (PCB) or on a matrix board.

Construction

The layout of the components is shown in **Figure 1**. Identify each part separately, insert it into the holes in the board and solder carefully. Long leads may be cropped before or after soldering, depending on your skill and preference. All electrolytic capacitors, T1, T2 and IC1, must be connected correctly. The front-panel controls are connected to the PCB terminals shown in Figure 1.

The layout of the controls and the placing of the board inside the case are matters of personal preference. The size of the prototype front panel is shown in **Figure 2**. The prototype had a slow-motion drive fitted to VC2, the main tuning capacitor. This required the capacitor to be fitted on its own small panel. The bandspread control does not need any form of

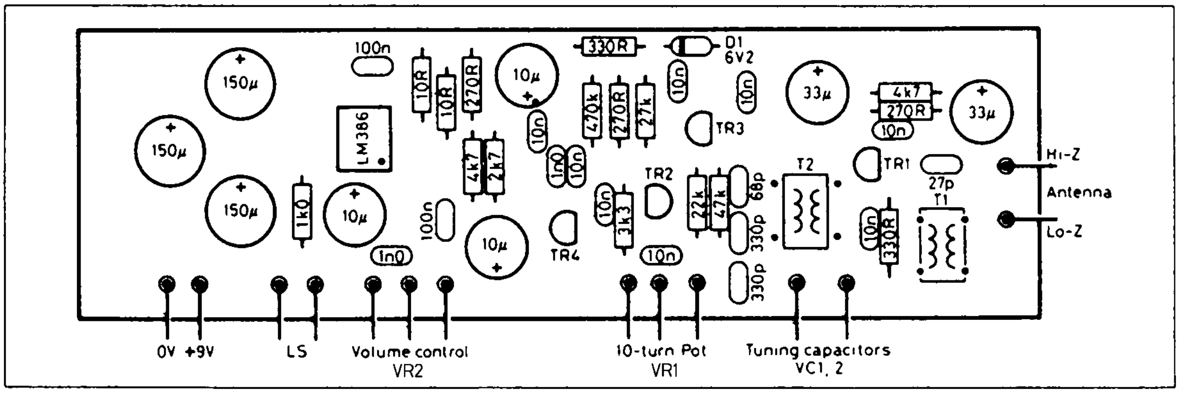
150µ
150µ
150µ
100n
LM386
1k0
10µ
10R
10R
270R
10µ
100n
1n0
4k7
2k7
10µ
10n
1n0
10n
TR4
330R
470k
270R
27k
10n
3k3
TR2
10n
D1
6V2
10n
10n
TR3
22k
47k
68p
330p
330p
T2
33µ
4k7
270R
10n
TR1
10n
330R
27p
T1
33µ
Hi-Z
Antenna
Lo-Z
0V
+9V
LS
Volume control
VR2
10-turn Pot
VR1
Tuning capacitors
VC1, 2

Figure 1 Layout of the receiver circuit board

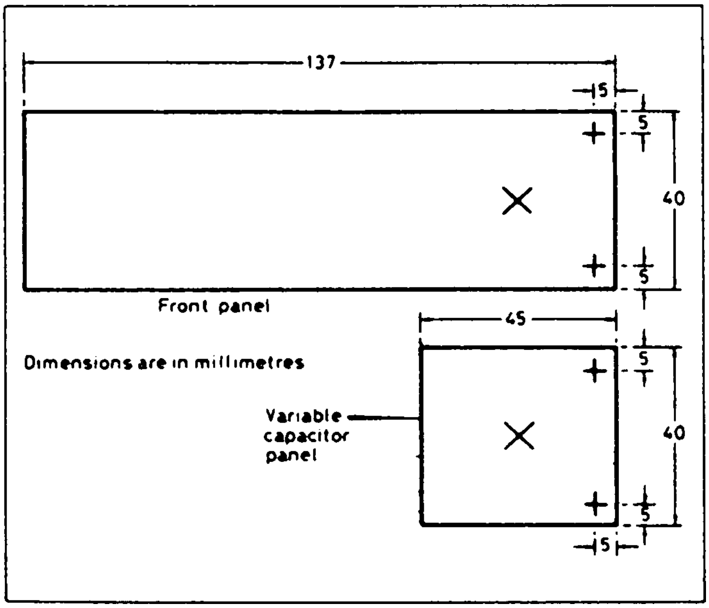

Figure 2 The case can be metal or made from PCB material

slow-motion drive. A tuning dial may be fitted behind the tuning knob on VC2 if required. The volume control, VR2, has its own on/off switch for the battery, which can be mounted behind the back panel, on which are mounted the aerial input sockets and the headphone/speaker socket. Check your soldering carefully, together with all the wiring between the board, the potentiometers and variable capacitors. When you are confident that everything is perfect, connect the battery and switch on.

The first tests

If you have a good aerial and are using an aerial tuning unit (ATU), use the low-impedance input. Excellent results are possible, though, with about 3 metres of wire connected to the high-impedance input. Set the volume control to give a gentle hiss in the headphones or speaker. Advance the reaction control to give a definite hiss. As you tune in an AM station, the hiss will change to a whistle; back off the reaction until the oscillation *just* stops. The receiver is now correctly set for AM reception.

On the amateur bands, the stations will be SSB or CW, and the reaction needs to be set *just above* the oscillation point, and will need slight adjustment from time to time for different qualities of signal. Juggling with the volume, reaction and tuning is part of the pleasure of using regenerative receivers!

In action

Practice is needed for best results. The regenerative receiver is renowned for its versatility in being able to be set up *exactly* for all types and strengths of signal. The basic receiver tunes from 6.5 MHz to 11 MHz approximately; this includes two amateur bands at 7.0 MHz and 10.1 MHz and two broadcast bands.

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1, R10	330 ohms (Ω)
R2, R5, R11	270 ohms (Ω)
R3	22 kilohms (kΩ)
R4	47 kilohms (kΩ)
R6	3.3 kilohms (kΩ)
R7, R12	4.7 kilohms (kΩ)
R8	27 kilohms (kΩ)
R9	470 kilohms (kΩ)
R13	2.7 kilohms (kΩ)
R14, R15	10 ohms (Ω)
R16	1 kilohm (kΩ)
VR1	10 kilohm (kΩ) linear 10-turn potentiometer
VR2	10 kilohm (kΩ) log potentiometer

Capacitors

C1	27 picofarads (pF)
C2, C18, C22	100 nanofarads (nF)
C3, C4, C6, C10, C13, C14	10 nanofarads (nF)
C5, C11	33 microfarads (μF) 16 V electrolytic
C7	68 picofarads (pF)
C8, C9	330 picofarads (pF)
C15	1 nanofarad (nF)
C16, C17, C20	10 microfarads (μF) 16 V electrolytic
C21, C23	150 microfarads (μF) 16 V electrolytic
VC1	10 picofarads (pF) variable
VC2	200 picofarads (pF) variable

Semiconductors

TR1, TR3	2N3819
TR2, TR4	BC182
IC1	LM386
D1	6.2 V 0.5 W Zener

Inductors

T1	3 turns primary, 15 turns secondary, wound on 2-hole ferrite bead, with 28 SWG wire
T2	2 turns primary, 17 turns secondary, wound on a T68-2 toroidal former, with 28 SWG wire

Additional items
Printed-circuit board
Battery connector
PP3 battery
8-pin DIL socket
8 ohm speaker or headphones (Walkman type)
Case to suit

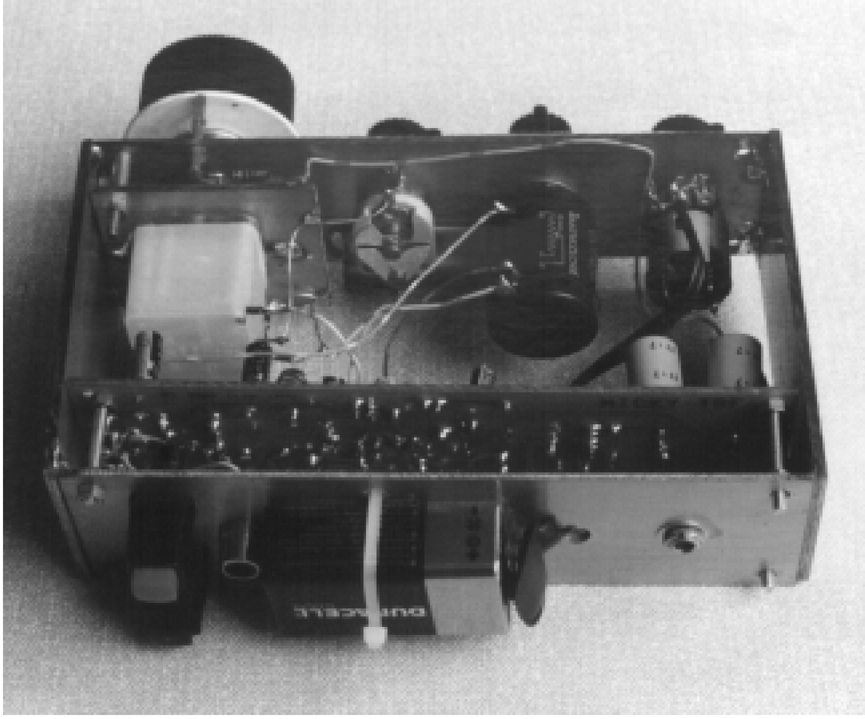

34 A basic continuity tester

Introduction

This little device can be built on a plug-in breadboard, and is ideal for testing fuses and cables, as well as doubling as a signal source for testing amplifiers, etc. It can even be used for testing npn transistors by replacing either of the transistors in the circuit, and seeing if the circuit still works!

Simple and quick to build

Using a plug-in breadboard, this circuit is so simple it could almost be built when you need it, and then dismantled again! If you want the circuit to make a different sound, the components to change are R2, R3, C1 and C2. Always make sure that R2 = R3 and C1 = C2, or the sound may be excessively 'edgy' and lacking in volume.

Basically, the circuit is an oscillator which drives a little loudspeaker directly. The fuse or cable you are testing is connected between the crocodile clips. If there is a current path between the two clips, the current also flows through the circuit, thus operating the oscillator and producing a sound from the loudspeaker.

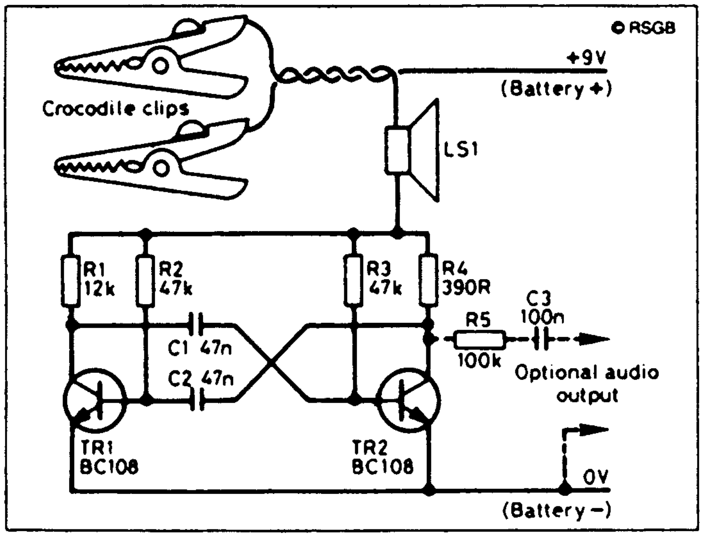
© RSGB
+9V
(Battery +)
Crocodile clips
LS1
R1
12k
R2
47k
R3
47k
R4
390R
C3
100n
R5
100k
C1 47n
C2 47n
Optional audio
output
TR1
BC108
TR2
BC108
0V
(Battery −)

Figure 1 The circuit diagram of the continuity tester

If you want to use the circuit to produce a signal to test an amplifier, for instance, connect the croc clip which comes directly from the battery to the junction of R1, R2, R3 and R4. This supplies current to the circuit while, at the same time, bypassing the loudspeaker. Use the two components R5 and C3 between the oscillator and your amplifier. The loudspeaker is bypassed because you will want to listen to the output from your amplifier, *not* from the oscillator!

Remember, NEVER test any equipment which is still connected to the mains electricity supply. Avoid testing anything which is switched on and has its own power supply. You may damage both your tester ***and*** the circuit you are 'testing'!

Parts list

Resistors: all 0.25 watt, 5% tolerance	
R1	12 kilohms (kΩ)
R2, R3	47 kilohms (kΩ)
R4	390 ohms (Ω)
R5	100 kilohms (kΩ) optional
Capacitors	
C1, C2	47 nanofarads (nF) ceramic
C3	100 nanofarads (nF), 0.1 microfarads (μF) ceramic optional
Semiconductors	
TR1, TR2	BC108 npn
Additional items	
LS1	Loudspeaker (8 Ω)
	2 crocodile clips
	Prototype circuit board
	PP3 battery connector
	PP3 battery

35 A charger for NiCad batteries

Introduction

The cost of replacing dry batteries can be alleviated to a great extent by the use of rechargeable batteries such as those containing nickel and cadmium (called *NiCads*). They are more expensive than the batteries they replace, but they can be charged hundreds of times and thus prove cheaper in the long run.

NiCads produce only 1.2 V per cell, compared with the 1.5 V of ordinary cells, so before you rush out and buy lots of these, please make sure that the equipment on which you plan to use them will work on the reduced voltage! For example, if you are using four cells to give you a 6 V supply, NiCads will give you only 4.8 V, which is quite a reduction. Using six cells to replace a 9 V battery will give you only 7.2 V. Not all equipment is happy with these reductions!

However, we all use them when we can, and they save substantial amounts of money. Here is the circuit of a charger to keep them in prime condition.

Charging NiCads – the ampère-hour

NiCads require charging at constant current, which means that connecting one across a normal power supply (constant *voltage*) is useless and can destroy it. They need pampering to the extent of needing a long charge (around 16 hours) at a rate dependent upon the *capacity* of the battery. By the capacity of a battery, we mean how much energy it can store. You will probably know that energy is measured in joules. For the purposes of storing energy in batteries, the joule is *not* the ideal unit, so we use one that is! This unit is the ampère-hour (Ah), and must be interpreted with some realism. For example, if the battery is rated at 2 Ah, it will deliver a current of 0.5 amp for 4 hours, or 0.25 amp for 8 hours. Provided the current is not too high, the product of the current (in amps) and the time for which it will flow (in hours) before the battery is flat will always be around 2 Ah. A workable ‘rule of thumb’ for calculating the charging current is that its value should be around one-tenth of the numerical value of the capacity; so, for our 2 Ah battery, a charging current of around 200 mA (2 ÷ 10 = 0.2 A or 200 mA) would be used.

Constant voltage to constant current

Many integrated circuit (IC) chips are available for use as voltage regulators, i.e. they supply a constant voltage. Most of these can be persuaded to become constant current supplies with one external resistor!

The voltage regulator IC usually has only three connections – ‘input’, ‘output’ and ‘common’. It is designed (in the case of the LM7805) to produce a constant 5 V output between the ‘output’ and the ‘common’ connections, at currents of up to 1 A. If a resistor is connected between these, the IC will maintain 5 V across it. If you look at **Figure 1**, you will see the circuit performing this conversion.

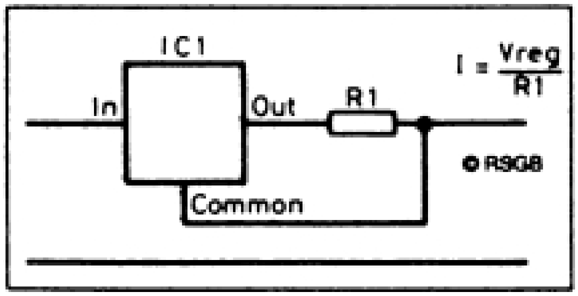

Figure 1 Voltage regulator arranged to produce a constant current

For the previous example, we derived a charging current of 200 mA, so we now need to calculate the value of resistor that will produce this current. Using the equation which is derived from Ohm’s law:

$$I = \frac{V}{R}, \text{ from which } R = \frac{V}{I},$$

where R = R1, the resistance in ohms that we are calculating,
V is the voltage across R1 (5 V), and
I is the current flowing (200 mA).

So,

$$\text{R1} = \frac{5}{0.2} = 25\,\Omega.$$

25 ohms is not a ‘common’ or ‘preferred’ resistor value, so we must choose the next *largest* value, which is 27 Ω. This reduces the current, but only slightly – it is now 185 mA. When calculating resistor values in power supply circuits, we must *always* check the power that they dissipate and make sure we specify and fit a suitable resistor.

Power (in watts) is the product of the voltage across and the current through a device, so in this case it is given by:

$$\text{Power} = V \times I = 5 \times 0.185 = 0.925\,\text{W}.$$

Rather than use a 1 watt resistor operating very near its limit, it is safer to use a 2 watt resistor operating well within its limits.

Looking again at Figure 1, we now have a constant-current source producing 185 mA, when R1 is a 27 Ω, 2 W resistor. For use with NiCads requiring charging currents other than 200 mA, you will need to repeat the two equations above, using a new value for *I*.

The full circuit and its assembly

This is shown in **Figure 2**, and can be broken down into two parts. The first is the AC to DC conversion produced by the mains transformer, T1, the bridge rectifier, BR1, and the smoothing capacitor, C1. The second is the constant-current section already discussed.

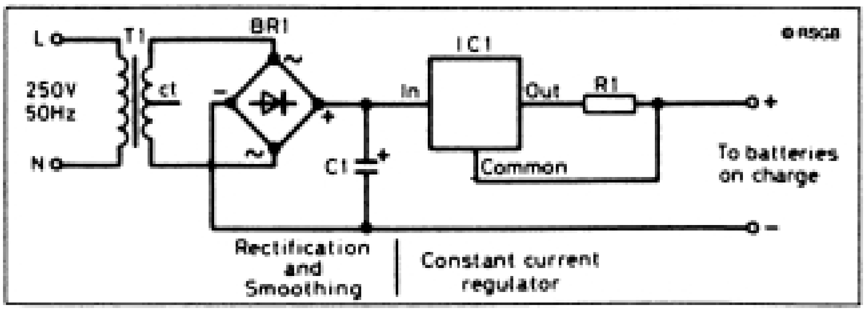

Figure 2 The complete circuit including mains components

The prototype was assembled on matrix board measuring 18 holes by 12 strips, although, as **Figure 3** shows, this is much larger than is strictly necessary. No strip cutting is needed, but make sure that IC1 and C1 are inserted correctly.

Warning! Before you attempt to wire up the transformer and the bridge rectifier, be aware that you will eventually be connecting the circuit to the mains supply, so there are three possibilities for you: (1) get a qualified friend to supervise your completion and testing of the circuit; (2) get your qualified friend to complete and test the circuit for you; (3) replace the transformer and bridge rectifier with a mains adapter.

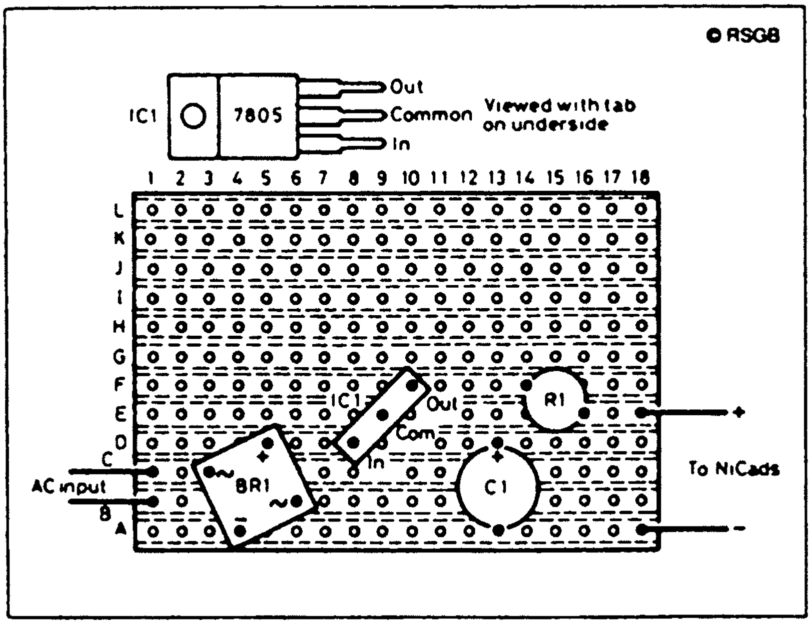

Figure 3 The charger can be built on Veroboard

If you decide to use the mains adapter, its output is connected directly across C1, because T1 and BR1 are now no longer needed. Make sure the polarity (positive and negative) is correct and that the adapter output is set to 12 V.

A quick test

If you have built the mains version, make sure all connections are correct, and that there are no soldered joints which will touch other parts of the circuit. The box must be securely closed before tests begin. The RSGB cannot be held responsible for damage to equipment or batteries! The version using the mains adapter need not be closed during tests.

Switch on. With nothing connected to the output, the unit should run cold. If this is not the case, switch off and recheck your circuit. If all is well, switch on again and connect a DC multimeter (on the current range) across the output. It should indicate only a slight difference from the calculated value of 185 mA. You can now charge your NiCads!

Other charging currents can be set by having different values for R1, perhaps selectable by a rotary switch. Remember to make sure that the values of both resistance *and* power dissipation are correct, and don't exceed the 250 mA rating of the transformer (or the 1 amp rating of the IC if you are using a bigger transformer).

Parts list

		Maplin code
Resistor		
R1	As required – see text	
Capacitor		
C1	1000 microfarads (μF) 35 V electrolytic	FF18U
Semiconductors		
IC1	LM7805 5 V, 1 A regulator	CR14Q
BR1	W005 50 V, 1 A full-wave rectifier	QL37S
Transformer		
T1	9–0–9 V, 250 mA sub-min. transformer	YN15R
Additional items		
Case		FG41U
Veroboard, 18 holes by 12 strips		
Plug to suit NiCads		
Double-sided sticky tape as required		
Insulated wire for battery connections		

36 An 80 metre crystal-controlled CW transmitter

Introduction

A simple transmitter is ideally suited to anyone venturing into our marvellous hobby for the first time. If you are put off by the complexities and prices of the 'black boxes' then this is the transmitter you've been looking for!

The circuit

The complete circuit is shown in **Figure 1**. It is a single-transistor crystal oscillator which is *keyed* (switched on and off) by the Morse key in the +12 V supply rail. The circuit first appeared in *QST* (the American equivalent of *RadCom*) in 1982, and has since appeared in a modified form in the *ARRL Handbook*. The circuit can produce about 2 watts on 80 m, and can be constructed on a piece of plain matrix board; the prototype board measured 7 cm by 3 cm, and its layout is shown in **Figure 2**.

Note that a bare copper wire runs along the bottom edge of the board to act as an earth wire for the relevant components. The only 'difficult' part of the construction is the winding and wiring of T1. The main winding is 38 turns of 26 SWG enamelled copper wire, and there are two link windings of four turns each. Make sure that all windings are wound the correct way round the toroid – Figure 1 shows this and should be studied carefully. If the windings do not have the correct *sense* (i.e. a clockwise coil has been wound anticlockwise, or vice versa), or have been connected incorrectly to the rest of the circuit, the oscillator will not work!

When putting components on the board, wire in the crystal socket *without the crystal in it*. Crystals do not like to be subjected to the horrors of a soldering iron, so keep your crystal to one side during the construction process!

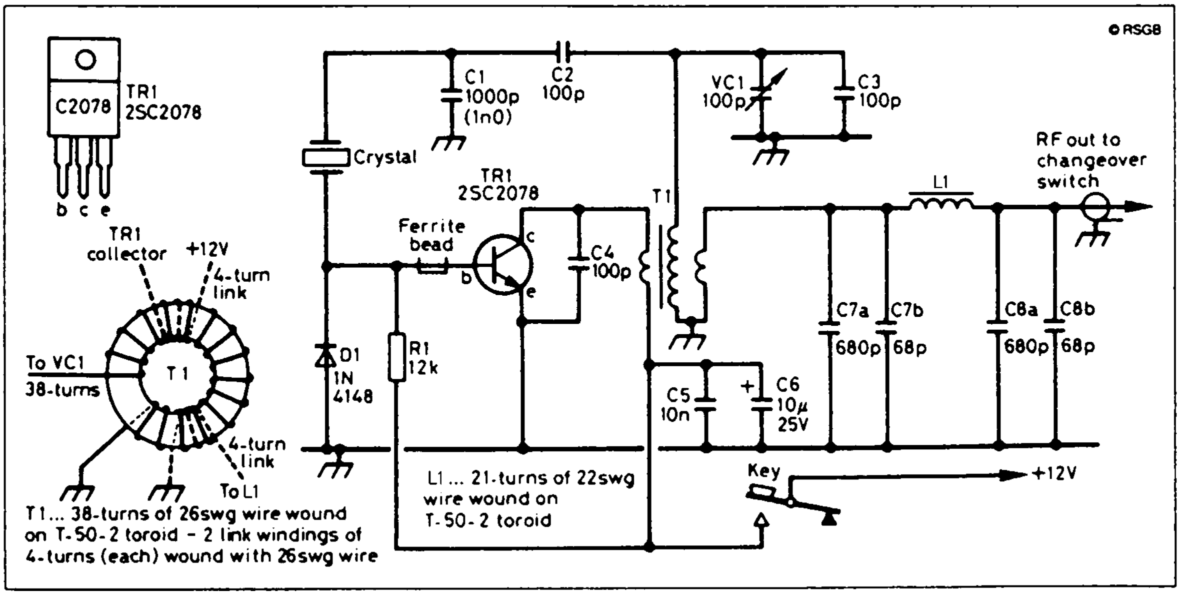
© RSGB
C2078
TR1
2SC2078
b c e
TR1
collector
+12V
4-turn
link
To VC1
38-turns
T1
4-turn
link
To L1
C1
1000p
(1n0)
C2
100p
Crystal
TR1
2SC2078
Ferrite
bead
b
c
e
C4
100p
D1
1N
4148
R1
12k
T1
VC1
100p
C3
100p
L1
RF out to
changeover
switch
C7a
680p
C7b
68p
C8a
680p
C8b
68p
C5
10n
+
C6
10µ
25V
Key
+12V

T1... 38-turns of 26swg wire wound on T-50-2 toroid – 2 link windings of 4-turns (each) wound with 26swg wire

L1 ... 21-turns of 22swg wire wound on T-50-2 toroid

Figure 1 The simplest form of transmitter is a keyed crystal oscillator. Note that L1, C7 and C8 make up a *low-pass filter* which reduces unwanted *harmonics* (outputs at the transmitter frequency multiplied by 2, 3, 4, etc.)

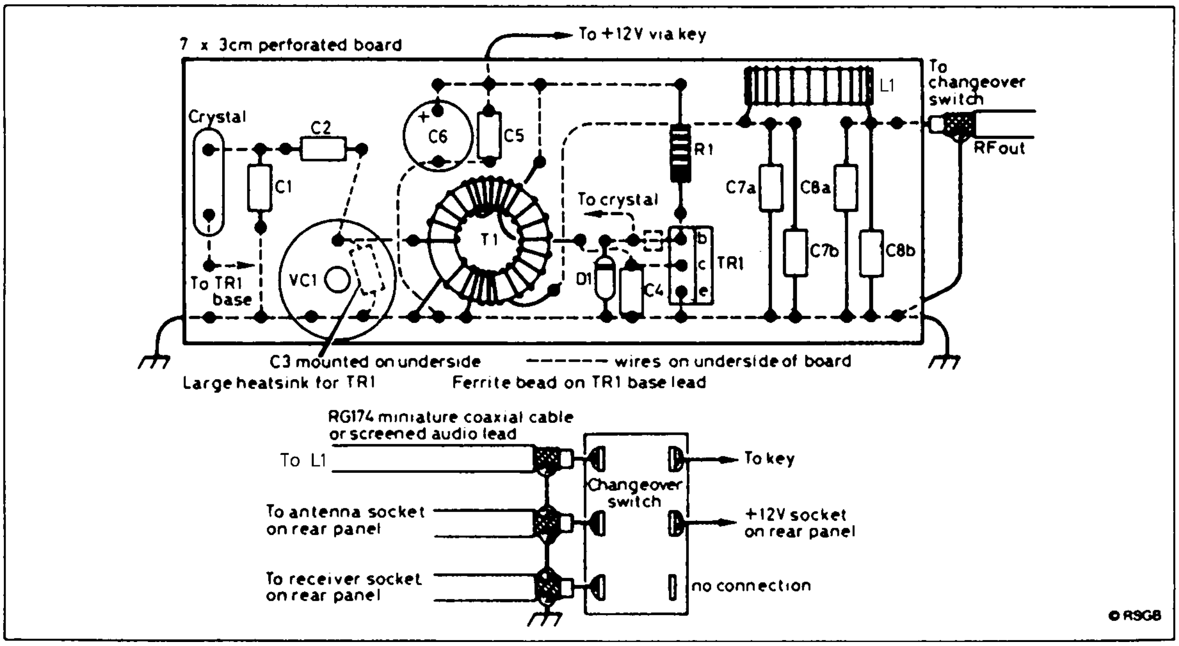
7 x 3cm perforated board
To +12V via key
Crystal
C2
C1
C6
C5
R1
L1
To changeover switch
RFout
C7a
C8a
C7b
C8b
To crystal
T1
TR1
b
c
e
D1
C4
VC1
To TR1 base
C3 mounted on underside
wires on underside of board
Large heatsink for TR1
Ferrite bead on TR1 base lead
RG174 miniature coaxial cable or screened audio lead
To L1
To antenna socket on rear panel
To receiver socket on rear panel
Changeover switch
To key
+12V socket on rear panel
no connection
© RSGB

Figure 2 Component layout is straightforward on a 'matrix board' (it has holes but no copper strips). The dotted lines are the connecting wires. The lower part shows the connections for the transmit/receive switch

The output filter, which comprises C7, C8 and L1, is a *low-pass* filter, which helps reduce any *harmonics* present in your signal. Harmonics are integral multiples of your transmitter frequency, so if you are transmitting on a frequency, *f*, harmonics will be present at frequencies 2*f*, 3*f*, 4*f*, . . . and so on. L2 is another inductance using 22 SWG enamelled copper on a ferrite toroid. The changeover switch is external to the transmitter board, and is used to switch your aerial between the transmitter and the receiver; its wiring is shown in Figure 2.

Use a dummy load

A dummy load enables you to test your circuit without actually transmitting a signal. If you haven't such a thing already, it is easy to construct one to use with this transmitter. Don't use it for transmitters of more than 2 watts output, though. Use two 100 ohm, 1 watt resistors, connected in parallel across the end of a short piece of coaxial cable, terminated in a BNC, PL259 or N-type free plug, as shown in **Figure 3**. Plug this into the aerial socket on your transmitter, plug in your crystal and connect the transmitter to a 12 V supply. Have another receiver switched on and tuned to the crystal frequency. Although the radiation from your dummy load is minimal, it will

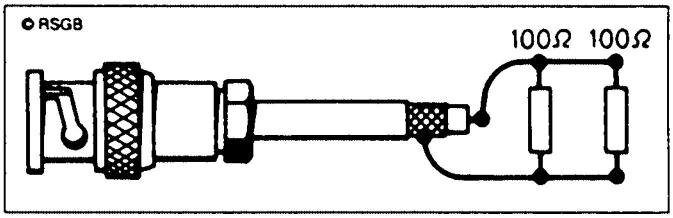

Figure 3 A simple 2 W dummy load can be made from two 100 Ω (ohm) resistors in parallel. The plug should match the socket on your transmitter

be enough to be picked up by a receiver in the same room. Send dashes with the Morse key, and adjust VC1 until the received note is *clean*. It should not sound rough, or have a *chirp* (change its frequency during a dash or dot). Avoid tuning for maximum power; this is seldom the correct setting!

You will need to put your completed transmitter in a metal box, using sockets for the power supply, aerial, receiver and Morse key. The sockets can be chosen to match your existing equipment.

Figures 4–7 are taken from the RSGB book *Practical Antennas for Novices*, and may give you some ideas on the type of aerial to be used with your transmitter.

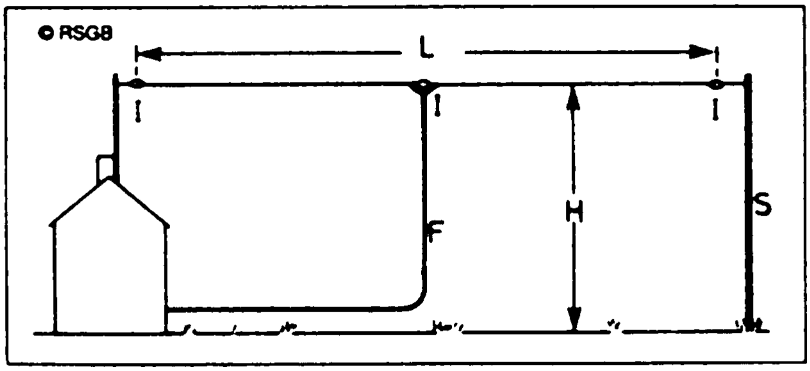

Figure 4 A simple dipole can be very effective. For the 3.5 MHz band, length L is 40 metres and height H should be as large as possible. The far support S can be a tree, pole or building. Insulators I may be home made from strong plastic and the feeder F should be 50 Ω (ohm) coax cable

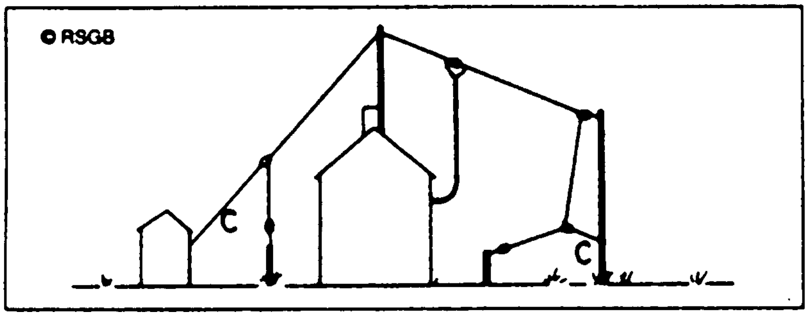

Figure 5 Your signal is radiated mostly from the centre of the dipole so the ends can droop or even be bent but the length may need shortening by a few centimetres because the ground and the bends will detune the dipole. Cords C are best made from strong plastic rope from a sailing or camping shop

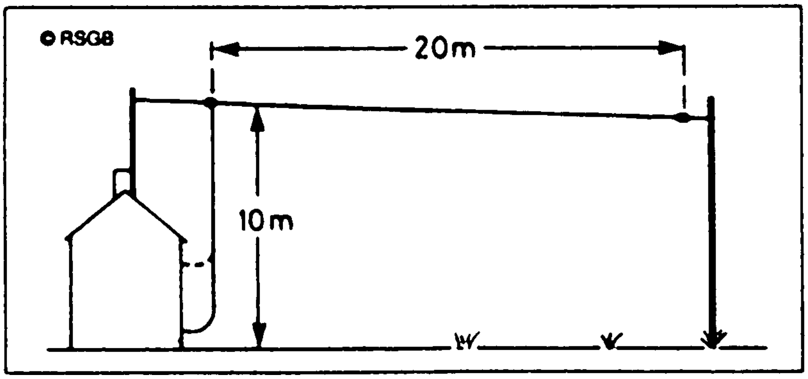
© RSGB
20m
10m

Figure 6 An ‘inverted-L’ takes up less space than a dipole and doesn’t need coax cable. Like the dipole, the end can droop or be bent to save space as in this case most of the radiation comes from the area around the top of the vertical part

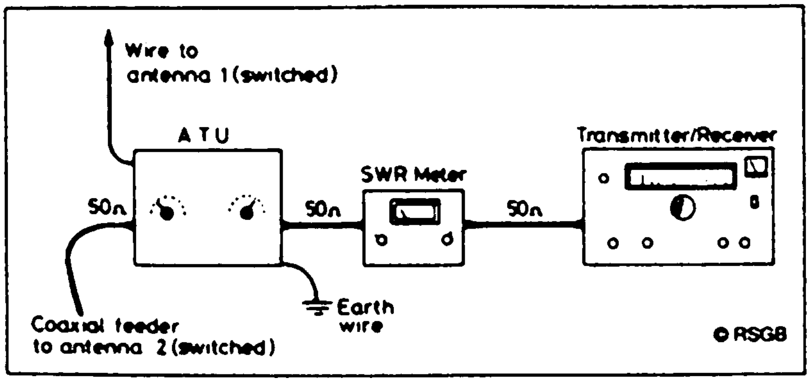
Wire to
antenna 1 (switched)
ATU
SWR Meter
Transmitter/Receiver
50Ω
50Ω
50Ω
Earth
wire
Coaxial feeder
to antenna 2 (switched)
© RSGB

Figure 7 Almost any length of wire more than 10 m or so long will work (though it will work better the longer and higher it is) but an Aerial System Tuning Unit (ASTU or ATU) will be needed

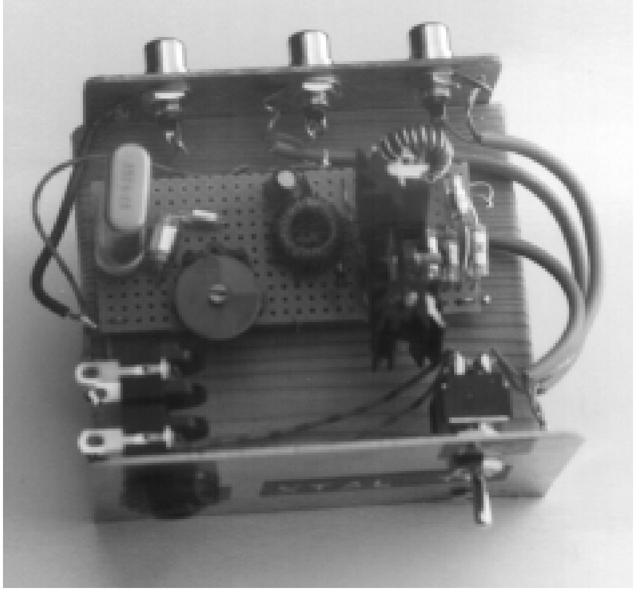

Parts list

Resistor

R1	12 kilohms (kΩ), 0.25 watt, 5% tolerance

Capacitors

C1	1000 picofarads (pF) polystyrene
C2	100 picofarads (pF) polystyrene
C7a, C8a	680 picofarads (pF) polystyrene
C7b, C8b	68 picofarads (pF) polystyrene
C3, C4	100 picofarads (pF) ceramic
C5	10 nanofarads (nF) polyester
C6	10 microfarads (μF) electrolytic 25 V
VC1	100 picofarads (pF) trimmer

Semiconductors

D1	1N4148
TR1	2SC2078 (see sources list)

Inductors

T1	38 turns 26 SWG enamelled copper on T–50–2 toroid, with two link windings of four turns
L1	21 turns 22 SWG enamelled copper on T–50–2 toroid

Additional items

- Ferrite bead
- Crystal (e.g. 3.579 MHz) and holder
- Metal box
- Socket for Morse key.
  This must be totally isolated from the metal of the box, as *both* connections can be at +12 V.
- Sockets for 12 V supply, aerial and receiver
- Switch – DPDT
- Heat sink for TR1
- RG174 miniature coaxial cable for signal leads (see Figure 2)

Component sources

Special components

2SC2078	Cricklewood Electronics Ltd, 40 Cricklewood Broadway, London, NW2 3ET.

37 A solar-powered MW radio

Introduction

What could be more ecologically friendly than a radio powered by the sun's energy? This design is quite standard, and if you have built any of the other medium-wave radios in this series, then this one should present few problems.

The solar panel

The solar panel is to the solar cell as the battery is to the cell; in other words a solar panel is several solar cells connected in series. The solar panel quoted for this radio will generate about 9 V at a current of around 30 mA on a sunny day. The circuit will operate on a supply of around 2 V, so bright sunshine is *not* necessary for satisfactory operation. The volume will be less, of course.

The circuit

The radio uses the ZN415E integrated circuit (IC), connected as shown in **Figure 1**, the complete circuit diagram. The signal is tuned in by the combination of L1 and VC1. L1 is made by winding about 35 turns of 24 SWG enamelled copper wire on a ferrite rod. A standard ferrite rod of 10 cm length and 1 cm diameter is used.

Signals selected by the tuned circuit are passed to IC1, which amplifies the signals and removes the audio component, which is then amplified further by IC2 for driving a small loudspeaker. The removal of the audio component is the process we call *detection*. In addition to this, IC1 provides *automatic gain control* (AGC), which helps to keep the audio signal constant, even when the incoming RF signal may vary due to fading.

The prototype board

Veroboard (also known as matrix board or stripboard) is ideal for the construction of the radio. The layout is shown in **Figure 2**. The board size is 11 strips by 30 holes (**please note that there is no row 'I', so take care with your counting!**). Using a 3 mm ($\frac{1}{8}$ inch) twist drill rotated between

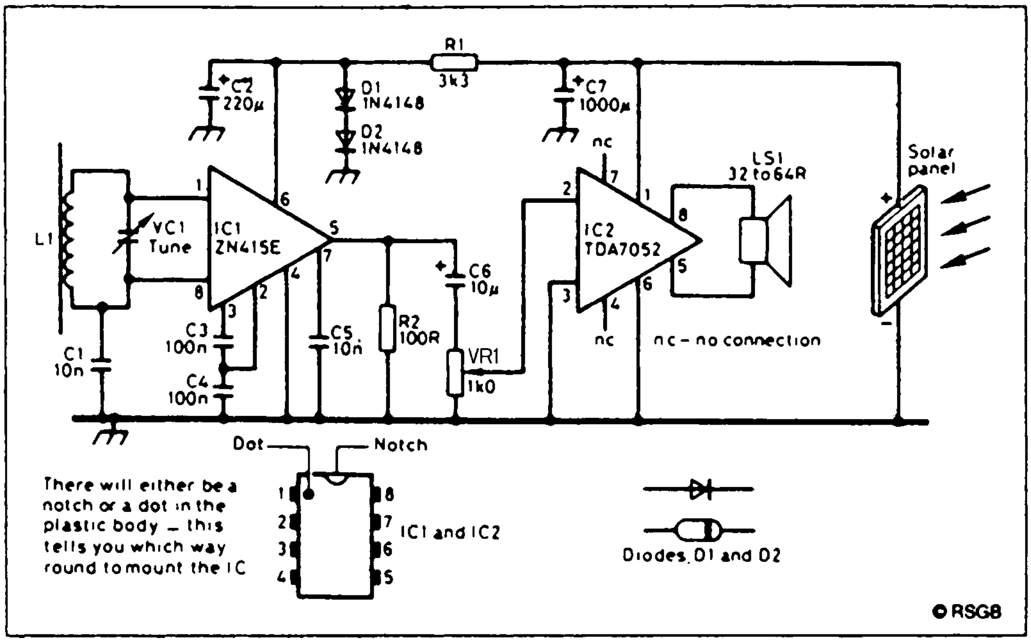
R1
3k3
C2
220μ
D1
1N4148
D2
1N4148
C7
1000μ
nc
LS1
32 to 64R
Solar panel
+
−
L1
VC1
Tune
IC1
ZN415E
IC2
TDA7052
C6
10μ
C1
10n
C3
100n
C4
100n
C5
10n
R2
100R
VR1
1k0
nc
nc – no connection
Dot
Notch
There will either be a notch or a dot in the plastic body – this tells you which way round to mount the IC
IC1 and IC2
Diodes D1 and D2
© RSGB

Figure 1 The radio uses just two integrated circuits (chips) and operates at any voltage from 2 to 9 V

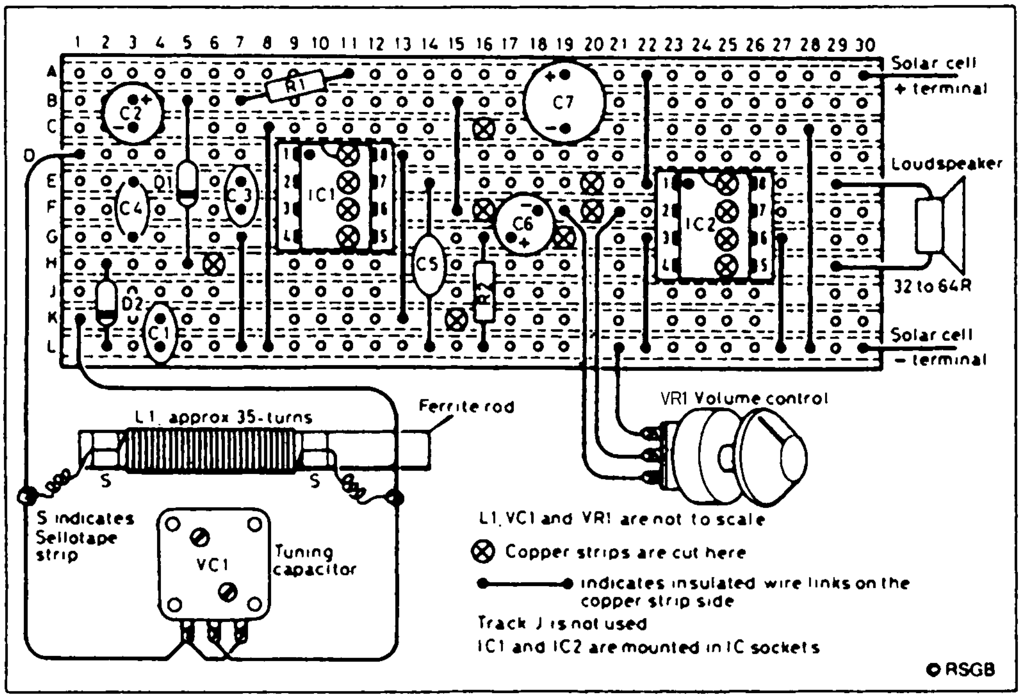
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
A B C D E F G H J K L
R1
C2
C7
D1
C4
C3
IC1
IC2
C6
C5
R2
D2
C1
Solar cell
+ terminal
Loudspeaker
32 to 64R
Solar cell
− terminal
VR1 Volume control
Ferrite rod
L1 approx 35-turns
S
S
S indicates Sellotape strip
VC1
Tuning capacitor
L1, VC1 and VR1 are not to scale
⊗ Copper strips are cut here
indicates insulated wire links on the copper strip side
Track J is not used
IC1 and IC2 are mounted in IC sockets
© RSGB

Figure 2 Follow the layout carefully, making sure that all connections are neatly soldered

your thumb and forefinger, break the strips at the points marked with a cross inside a circle. Hold the board up to the light to ensure that the tracks *have* been broken and that there is no copper swarf shorting tracks together.

Firstly, solder in the 8-pin DIL sockets for the ICs, making sure that the notches in the sockets are facing upwards, as shown on the diagram. Then, solder in the wire links, resistors, capacitors and diodes, making sure that the electrolytic capacitors (C2, C6 and C7) and the diodes (D1 and D2) are connected the correct way round. Use different colours of wire for the connections to the volume control, VR1, to avoid incorrect connections. Note the wiring of the tuning capacitor (VC1) shown in Figure 2; a two-section type is used, and both sections are wired in parallel to give twice the capacitance of a single section.

There is no on/off switch – just turn the volume down when you are finished using the set! The solar panel can be mounted parallel to the top of the case, or angled to receive the maximum energy from the sun, as shown in the photograph. You could have a battery available as a standby source to use the radio after dark; any battery of between 6 V and 9 V will do. Wire it with a simple changeover switch, so you can switch between solar and battery power. Ask a friend for help with this if you are not sure how to do it.

You may need to adjust the number of turns on L1 to get the best results, but it should be possible to receive at least five stations at good volume with your sun-powered radio!

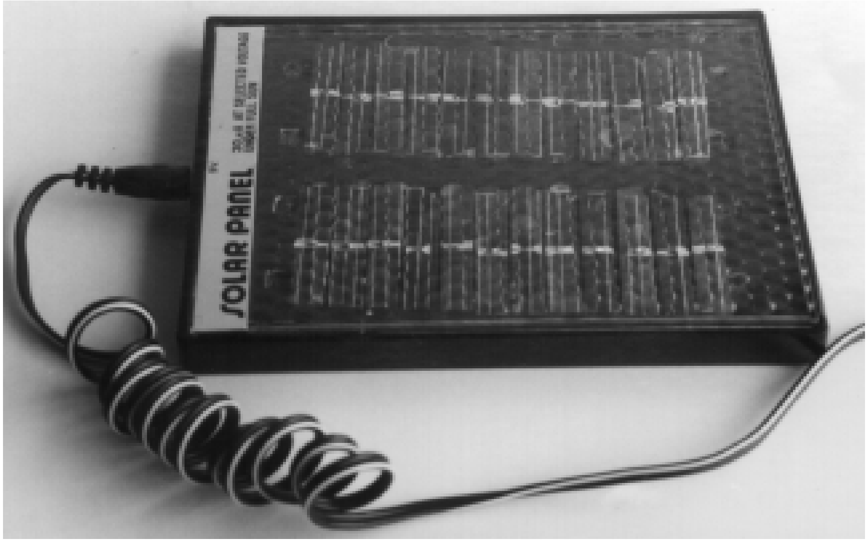

Parts list

		Maplin code
Resistors: 0.25 watts, 5% tolerance		
R1	3.3 kilohms (kΩ)	M3K3
R2	100 ohms (Ω)	M100R
VR1	1 kilohm (kΩ) log	FW21X
Capacitors		
C1, C5	10 nanofarads (nF) ceramic	BX00A
C2	220 microfarads (μF) electrolytic (10 V)	FB60Q
C3, C4	100 nanofarads ceramic	YR75S
C6	10 microfarads (μF) electrolytic (10 V)	FB22Y
C7	1000 microfarads (μF) electrolytic (10 V)	FB81C
VC1	140–300 picofarads (pF)	FT78K
Semiconductors		
IC1	ZN415E (or ZN416E)	UR70M
IC2	TDA7052	UK79L
D1, D2	1N4148	QL80B
Solar panel	9 V at 50 mA	RK23A
Additional items		
LS1	32–64 ohm miniature loudspeaker	YT28F
	Ferrite rod	YG20W
	24 SWG enamelled copper wire	BL28F
	Plastic box, approx. 220 × 140 × 70 mm	YN39N
	Veroboard, cut to size	JP47B
	8-pin DIL sockets, 2 required	BL17T
	Knobs, 2 required	FD67X
	Material for speaker grille	
	Connecting wire	

38 A receiver for the 7 MHz amateur band

Introduction

Listening on the 40 metre band (from 7.0 to 7.1 MHz) can be very rewarding – it is a popular haunt for HF Special Event stations, and at night there are signals to be heard from all over Europe. This receiver is designed purely for the 40 m band, and is ideal for those who have built the simpler receivers and are looking for something a little more challenging. The more experienced constructor may prefer to build this on prototype board.

The circuit and its construction

Figure 1 shows the circuit diagram. The receiver will work well with headphones or loudspeaker. Walkman-type headphones and speakers are ideal for use here.

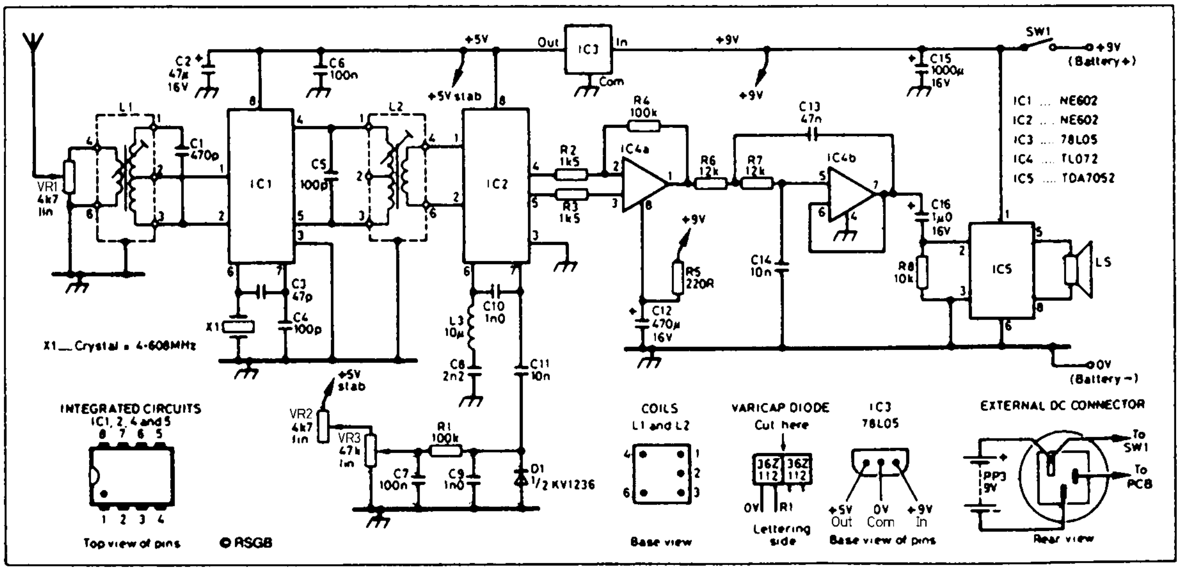
+5V
Out
IC3
In
Com
+9V
SW1
+9V
(Battery+)
C2
47µ
16V
C6
100n
+5V stab
+9V
C15
1000µ
16V
L1
L2
C1
470p
IC1
C5
100p
IC2
R4
100k
C13
47n
IC4a
IC4b
R2
1k5
R3
1k5
R6
12k
R7
12k
VR1
4k7
lin
+9V
R5
220R
C14
10n
C16
1µ0
16V
R8
10k
IC5
LS
C3
47p
C4
100p
X1
C10
1n0
L3
10µ
C8
2n2
C11
10n
C12
470µ
16V
0V
(Battery−)
X1 __ Crystal = 4·608MHz
IC1 ... NE602
IC2 ... NE602
IC3 .... 78L05
IC4 .... TL072
IC5 .... TDA7052
INTEGRATED CIRCUITS
IC1, 2, 4 and 5
Top view of pins
© RSGB
+5V stab
VR2
4k7
lin
VR3
47k
lin
R1
100k
C7
100n
C9
1n0
D1
1/2 KV1236
COILS
L1 and L2
Base view
VARICAP DIODE
Cut here
36Z 112
0V
R1
Lettering side
IC3
78L05
+5V Out
0V Com
+9V In
Base view of pins
EXTERNAL DC CONNECTOR
PP3 9V
To SW1
To PCB
Rear view

Figure 1 The receiver gives good performance on the 7 MHz amateur band as well as being simple to construct and align

Signals arriving at the aerial are coupled into IC1 via gain control VR1, which also functions as the on/off switch. Tuning is provided by varying the voltage on the *varactor diode* (or varicap), D1. VR3 is the main tuning control, and VR2 is the *bandspread* (fine tuning) control. The varactor diode is supplied as a dual device, which must be cut down the middle *carefully* with a sharp knife; with the lettering upwards, the ground lead (0 V) is on the left-hand side, as Figure 1 illustrates.

Solder in the IC sockets first, followed by the coils. After this come the links, resistors, capacitors and varactor diode. Ensure that IC3, the voltage regulator, is wired correctly, and check the polarity of the electrolytics. The crystal, X1, is very fragile, so take extra care with it. The wiring of the three controls is shown in **Figure 2**.

Before putting the ICs in their sockets, connect up the battery and check the following voltages with the negative voltmeter lead connected to the negative terminal of the battery:

Pin 8	IC1	5 V
Pin 8	IC2	5 V
Pin 8	IC4	9 V
Pin 1	IC5	9 V

When all these have been found to be correct, switch off and put the ICs carefully into their sockets. Use wire of different-coloured insulation to wire up the front-panel controls.

The case can be a small plastic box of size 22 cm by 15 cm by 8 cm, with three 10.5 mm holes drilled in the front and two 8 mm holes in the side for the aerial and earth connections. On one side are a 6 mm hole for the speaker socket and an 11 mm hole for the optional external power supply.

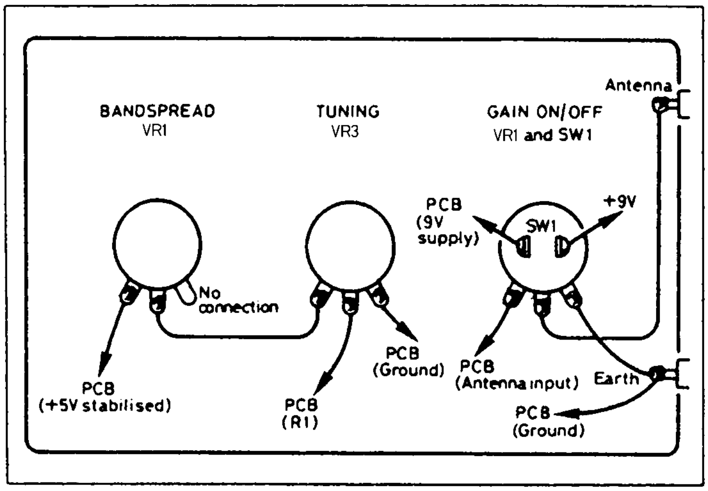

Figure 2 Rear view of the variable resistors. Check the connections carefully to make sure the wires are fitted correctly

Testing and tuning

The aerial for the receiver should be between 30 and 70 feet of wire, mounted as high as you can make it, away from trees and buildings if possible. Connect the battery and switch on. Adjust L1 and L2 for the best results. Tune slowly with VR3; you should find CW stations at the lower end of the band (anticlockwise) and SSB stations at the upper end (clockwise). You may find that it is easiest to make these adjustments *before* mounting the board in the case with double-sided sticky tape or pads. If you are planning to use an external DC supply, make sure it is a safety approved stabilised 9 V type, and **disconnect the battery before you use such a supply!**

If you suspect that the tuning doesn't quite cover the lower CW end of the band, try increasing C9 to 1200 pF. If it is the upper SSB end which is missing, decrease C9 to 820 pF.

It is always advisable to use an aerial tuning unit (ATU) between your aerial and the receiver. A suitable design of ATU is included as a project in this book.

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1, R4	100 kilohms (kΩ)
R2, R3	1.5 kilohms (kΩ)
R5	220 ohms (Ω)
R6, R7	12 kilohms (kΩ)
R8	10 kilohms (kΩ)
VR1	4.7 kilohms (kΩ) linear, with SPST switch
VR2	4.7 kilohms (kΩ) linear
VR3	47 kilohms (kΩ) linear

Capacitors: all rated 16 V or more

C1	470 picofarads (pF) polystyrene 5%
C2	47 microfarads (μF) electrolytic
C3	47 picofarads (pF) polystyrene 5%
C4, C5	100 picofarads (pF) polystyrene 5%
C6, C7	100 nanofarads (nF) ceramic
C8	2.2 nanofarads (nF) polystyrene 5%
C9, C10	1 nanofarad (nF) polystyrene 5%
C11, C14	10 nanofarads (nF) ceramic
C12	470 microfarads (μF) electrolytic
C13	47 nanofarads (nF) ceramic
C15	1000 microfarads (μF) electrolytic
C16	1 microfarad (μF) electrolytic

Inductors	
L1	Toko KANK3335R
L2	Toko KANK3333R
L3	10 μH 5%, e.g. Toko 283AS–100
Semiconductors	
IC1, IC2	NE602 or NE602A
IC3	78L05 5 V, 100 mA
IC4	TL072
IC5	TDA7052
Additional items	
D1	Toko KV1236 cut into two sections (one half used)
X1	4.608 MHz (available from Cirkit)
	3 × silver knobs, one with pointer
	Plastic case approx. 22 × 15 × 8 cm
	Speaker 8–32 Ω, or headphones
	4 × 8-pin DIL sockets for IC1, IC2, IC4, IC5
	2 × 4 mm sockets (red and black) for aerial and earth
	3.5 mm chassis-mounting jack socket for speaker
	DC power socket for external supply (if required)
	Prototype board

Kits

A complete kit is available from JAB Electronic Components.

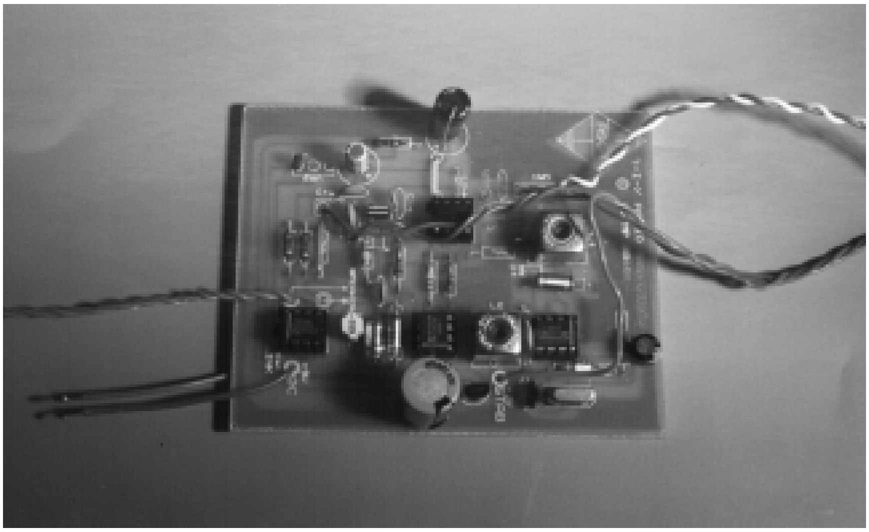

39 Diodes for protection

Introduction

Many semiconductor devices can be destroyed in an instant if their supply is reversed. With the use of batteries as power sources, it is quite common for the battery to be connected ‘the wrong way round’, and the scope for damaging equipment is significant and very real. Diodes can protect equipment in several ways, and you may do worse than to consider one of these approaches to protect your next expensive project.

Choose wisely!

All the diode circuits given here are so simple as to invite calamity. The circuits are not foolproof but, with a little care, will work first time. Remember that a semiconductor diode has a forward voltage drop of between 0.5 V and 0.7 V, depending on its type and the current flowing. This will be mentioned later.

The first thing you need to do is to insert a good multimeter *in series* with the circuit you want to protect; set the range to Amps DC, and switch on. Check that the circuit works properly. Then, decrease the current range on the meter until a good reading is indicated. Make a note of this current, as it is the normal running current of your circuit.

To choose a diode, you must consult the catalogues or data sheets and find one where the quoted *maximum forward current* exceeds the current you have measured; preferably it should be at least twice your measured current. Secondly, the diode will have a *peak inverse voltage* (PIV); this is the maximum voltage it can withstand when the cathode is made positive with respect to the anode, i.e when it is *reverse-biased* and not conducting. This voltage must be greater than your battery voltage, again by a factor of about 2. Except in the case of the bridge rectifier (see later), these criteria will enable the selection of a suitable diode to be made easily.

The series diode

The simplest and most obvious way to protect equipment is to insert a diode in the positive supply lead, as shown in **Figure 1**, with the diode passing current *only* when the supply is of the correct polarity. Because of the 0.6 V that exists across the diode, your equipment will normally operate on a

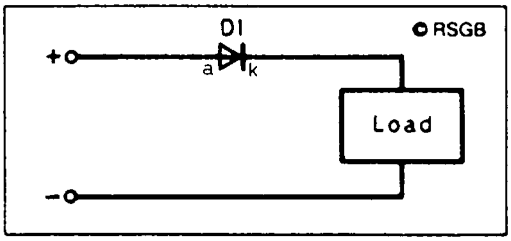

Figure 1 Series diode protection

slightly lower voltage. If you imagine reversing the supply to this circuit, you will see that the negative terminal of the battery is connected to the anode of the diode; the diode becomes reverse-biased and will not conduct any current, thus protecting the load. Your equipment will not operate when the battery connections are reversed.

The parallel diode

The circuit of **Figure 2** overcomes this voltage drop. It places the diode in parallel with the load (your equipment) but in a normally reverse-biased condition so that it draws no current when the battery is correctly connected. Reverse the battery connections, however, and a very large current will flow through the diode, thus blowing the fuse! For this technique to work successfully, the current drawn by the diode when the battery connections are reversed *must* be much greater than the maximum current drawn by the equipment, in order to blow the fuse. This is usually not a problem, however.

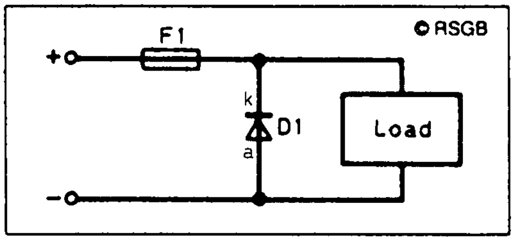

Figure 2 Parallel diode and fuse

The diode bridge

For sheer elegance, the circuit of **Figure 3** takes the biscuit! It uses four diodes connected in the form known as a *bridge rectifier*. Such rectifiers exist, and do not have to be made up from four discrete diodes.

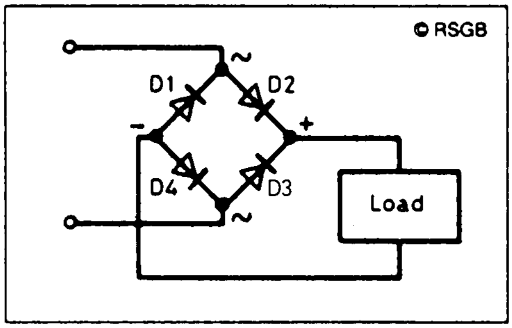

Figure 3 The diode bridge

Follow the current round the circuit from the supply, assuming initially that the top wire is positive and the bottom wire negative. It flows from the positive supply

(a) through D2
(b) through the load (top to bottom)
(c) through D4

and back to the negative of the supply.

Now assume that the bottom supply lead is positive and the top lead negative. The current flows from the positive supply

(a) through D3
(b) through the load (top to bottom)
(c) through D1

and back to the negative of the supply.

So, whichever way round the battery is connected, the current will *always* flow the same way through your equipment!

The circuit does have a drawback, however. Whichever way round the battery is connected, there are always **two** diodes conducting the current at any time. In the first case it is D2 and D4; in the second case, D1 and D3. This means that there is a total voltage drop of about 1.2 V. If your equipment can tolerate that reduction in voltage, then you will not have a problem.

Decoupling

Whenever the supply rail to a piece of equipment, or even to an individual stage of a circuit, is broken for the insertion of a device that will drop voltage, strange things can happen. This is because the supply for any circuit is assumed to have a low resistance to DC and a low *impedance* to AC. (Impedance is the AC equivalent of DC resistance.) These two are not the same, and the insertion of a diode or diodes is certain to make a big difference to them both. To overcome any irregularities in the operation of the circuit you are protecting in any of the ways described previously, the protected supply needs to be *decoupled* with a capacitor.

If the circuit handles audio frequencies only, placing an electrolytic capacitor directly across the load in all three circuits should solve the problem. The parallel circuit of Figure 2 is less at risk than the other circuits. The size of the capacitor will be determined by the current taken by the circuit, and may need to be chosen within the range 100 µF to 10 000 µF, with a working voltage greater than the supply voltage.

If the circuit is mainly handling RF currents, placing a capacitor of 0.01µF across the load should prevent any problems. A second capacitor, also across the load, of between 10 µF and 100 µF may be needed. Again, the parallel circuit of Figure 2 is less at risk than the other two.

Don't be afraid to experiment, but confine your experimenting (at first) to small equipment and low currents, until you get a 'feel' for the technique.

40 An RF signal probe

Introduction

A radio-frequency (RF) diode probe is a simple device which, when used with a conventional multimeter, enables the measurement of RF voltages in a circuit. When constructed, this will be one of the most useful pieces of test equipment for the experimenter who revels in the construction of transmitters and receivers.

The circuit

Figure 1 shows the simple circuit diagram. It is *almost* the same as a common diode rectifier circuit, but with a simple change to make it more sensitive. The circuit is known as a *voltage-doubler*, and is often found in high-voltage supplies, with beefier capacitors and diodes, of course! Because we are dealing with high frequencies and smaller voltages, the diodes and capacitors can be physically very small. The diode circuit of D1 and D2 *rectifies* or *detects* the RF from the probe, and any remaining AC is removed (short-circuited to ground) by C2. This produces, at the output, a constant voltage proportional to the peak-to-peak RF voltage present at the input; the output voltage is fed to an ordinary multimeter (on a voltage range).

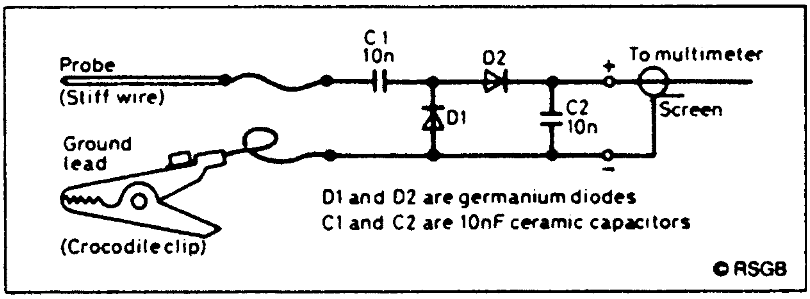
Probe
(Stiff wire)
C1
10n
D2
+
To multimeter
D1
C2
10n
Screen
Ground
lead
–
(Crocodile clip)
D1 and D2 are germanium diodes
C1 and C2 are 10nF ceramic capacitors
© RSGB

Figure 1 The diode probe can be made from four electronic components

Construction

Although the circuit layout is not critical, a description of the prototype is given here for your information. **Figure 2** has the details. The components are soldered to square pieces of copper-clad printed-circuit board (PCB) glued to a larger piece of PCB. The larger piece serves as a ground connection. Use a stiff copper wire as the probe, and an insulated flexible wire with a crocodile clip to connect the probe to the ground of the circuit under test.

Cut a piece of plain PCB, 30 mm by 45 mm and another of 15 mm by 45 mm. Cut the smaller piece into three measuring about 15 mm square. Stick the three small pieces to the larger piece, ensuring that there are small gaps between each, as Figure 2 shows. Solder the components in place. The connection to the multimeter should be thin coaxial cable or screened microphone cable. If you use unscreened cable, there may be RF pickup here which can lead to false readings.

Simple to use

Using the probe is simple. Connect it to the multimeter and set the meter to around 10 V DC – you may need to reduce this, depending upon the magnitude of the RF voltage you are trying to measure. Hold the probe by

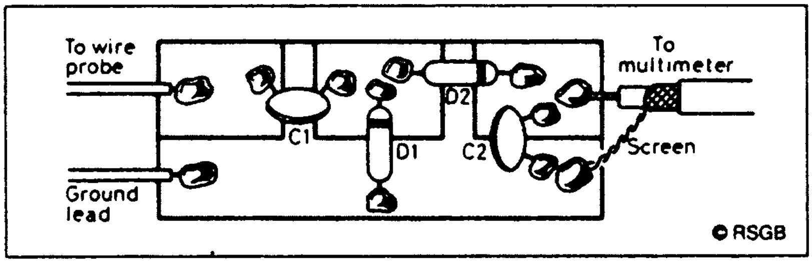
To wire
probe
To
multimeter
D2
C1
D1
C2
Screen
Ground
lead
© RSGB

Figure 2 The components are mounted on a copper-clad PCB

its base, being careful *not* to touch any of the components. Connect the croc clip to a ground point on the circuit and touch the probe on the circuit point to be tested. If an RF signal is present, there should be a reading on the multimeter. That's all there is to it!

Parts list

Capacitors

C1, C2	10 nanofarads (nF), or 0.01 microfarad (μF) ceramic

Semiconductors

D1, D2	Any *germanium* signal diode, e.g. OA91 or AA119

Additional items
- Stiff copper wire
- Insulated flexible wire
- Copper-clad PCB
- Crocodile clip
- Coaxial or screened cable

Source

All components are available from Maplin.

41 An RF changeover circuit

Introduction

The simplest entry route to operating an HF station is to buy or build a receiver, and to add a simple CW (Morse) transmitter. In theory, this sounds so simple, yet in practice there is one significant hurdle to be overcome. How can they both share the same aerial? A manually operated changeover switch is the obvious solution, but it takes time to perform the switching operation, by which time someone else has squeezed in before you and contacted the DX station! This small circuit accomplishes the changeover automatically, as soon as the transmitter is keyed.

How it’s done

The circuit is shown in **Figure 1**, and uses the signal from the transmitter to operate a relay. A relay is a switch which is operated electrically in the following way. A coil of wire with an iron core is used as an electromagnet. When current flows through the coil, it produces a magnetic field which, in turn, is used to pull a set of switch contacts. These contacts will be used to switch the aerial from transmitter to receiver and vice versa. The relay used here has a double-pole changeover (or DPDT, or DPCO) switch. If you are confused by the different types of switch, look at the basic descriptions elsewhere in this book. The *pole* of a switch is the part that doesn’t move; in Figure 1, one pole is connected to L1 and the other to the aerial output. In a circuit diagram, the switch contacts are always shown in their *normal* state, i.e. when *no* current flows in the relay coil. The changeover switch is normally in the receive position, so we say the receive switch is *normally closed*, and the transmit switch is *normally open*. We choose to have the circuit in the receive position normally, because everyone spends a lot more time receiving than transmitting. The relay is energised only when you are transmitting.

There are three RF sockets – one for the transmitter, one for the aerial, and one for the receiver. When the transmitter is not in use, there is a direct connection (via the Rx normally closed contacts of the relay) between the aerial and the receiver. The transmitter (which is not keyed at this time) is connected to a 50 ohm dummy load, R1.

When the transmitter is activated by pressing the Morse key, an RF signal appears at the transmitter socket. The wire carrying the signal to the relay

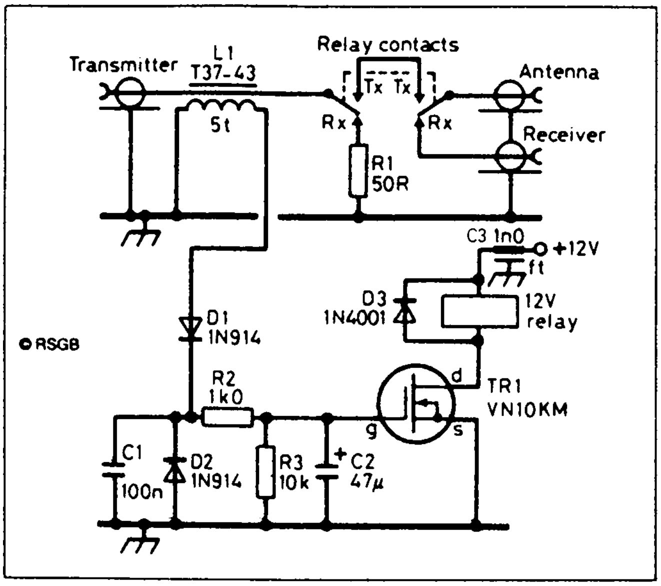

Figure 1 RF activated relay

Tx contacts passes through a toroid (a ferrite coil former, looking like a black Polo mint), which has a five-turn coil wrapped around it. This five-turn coil picks up the signal, which is then detected by D1 and D2, and converted to a steady voltage. C1 removes any remaining RF and C2 provides *hang*. Without it, the circuit would detect the gaps between every dot and dash being sent, and switch the aerial over very rapidly and very frequently! This is *not* what we want. We need the relay to remain in the transmit condition at normal Morse keying speeds, and then return to receive when the sending is complete. The combination of C2 and R3 achieves this.

The voltage appearing across C2 is fed to a VMOS field-effect transistor (FET), TR1, which acts as an electronic switch. When a voltage appears across C2, the FET switches on and passes a current through the relay coil, changing the contacts from receive to transmit. The diode, D3, across the relay winding protects the FET from being damaged by the large reverse-voltage spike which occurs across the coil when the FET switches off.

Construction

This is very simple – a circuit board is not required. The prototype was built into an old 50 g tobacco tin. All ground leads are soldered directly to the tin, thus supporting the components automatically. Phono sockets were used for the aerial, receiver and transmitter; you could use whatever connectors matched the rest of your station. The 12 V supply is fed through the tin using a 1000 pF feed-through capacitor, C3.

The 'dummy load', R1, must be able to withstand the RF output power of the transmitter. For novice use, 4 watts is adequate. Two suggestions for making up R2 from standard resistors are shown in **Figure 2**. Six 330 ohm, 1 watt resistors in parallel have a combined resistance of 330/6 = 55 Ω at 6 W. Two 100 ohm, 2 watt resistors in parallel have a combined resistance of 100/2 = 50 Ω at 4 W.

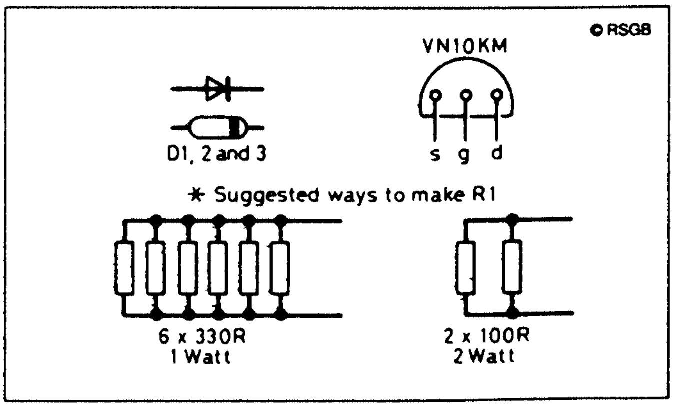

Figure 2 Component details

Make sure all the contacts to the relay are correct; also ensure that TR1, C2, D1, D2 and D3 are correctly wired.

Winding the toroid, L1, is very simple. It is made from PVC-insulated hook-up wire. Each time the wire passes through the centre of the toroid counts as one turn. The wire from the transmit socket to the relay simply passes through the centre! Use thin 50 Ω coax for the leads to the three sockets.

What happens to the receiver?

On transmit, the receiver is not connected to the aerial, but is only the separation of the switch contacts (about 0.5 mm!) away from the transmit lead. The receiver *will* pick up the transmitted signal, and there is a possibility that this signal will be enough to damage the receiver's sensitive input circuits. This can be prevented with the simple addition shown in **Figure 3**. Two diodes are connected back to back across the receiver socket. These act as a *limiter*, reducing the amount of signal that can enter the receiver. Solder these directly between the receiver socket and the tin.

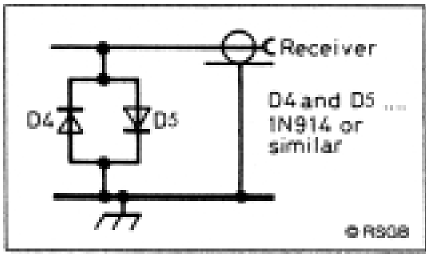

Figure 3 Receiver input protection

Warning

The wiring **must** be thoroughly checked. If the relay switch connections are wrong, there is a **great probability** that the full RF power will be applied to your receiver! You can check these connections as follows. With the 12 V supply connected, check that the relay energises (clicking sound heard) when the FET drain (d) is shorted to earth. Check the aerial and receiver connections with an ohmmeter or continuity tester. Disconnect the 12 V supply, and check that there is continuity between the centre pins of the aerial and receiver sockets. Reconnect the 12 V supply, short the FET drain to earth again (this will not damage the FET) and check that there is continuity between the centre pins of the aerial and transmitter sockets. If these tests show correct operation, disconnect the drain shorting wire (*most* important!) and your aerial auto-changeover is ready for use! Avoid using the device without the lid fitted.

Parts list

Resistors: 0.25 watt, 5% tolerance, unless otherwise stated

R1	50 ohms (Ω) – see text and Figure 2
R2	1 kilohm (kΩ)
R3	10 kilohms (kΩ)
Capacitors	
C1	100 nanofarads (nF), 0.1 microfarad (μF)
C2	47 microfarads (μF) 15 V electrolytic
C3	1000 picofarads (pF) feedthrough
Semiconductors	
D1, D2, D4, D5	1N914
D3	1N4001
TR1	VN10KM
Additional items	
L1	T37–43 toroid
	Relay 12 V DPDT (DPCO) relay
	Three sockets to suit (phono, SO239, etc.)
	Thin coax cable and hook-up wire

42 A low-light indicator

Introduction

This is a simple one-evening project that can be built for the pure fun of it, or to use as the basis of a more complex project to switch your shack lights on when it gets dark! In its prototype form, it simply flashes an LED when the ambient light level drops to a preset point.

Operation

The heart of the circuit shown in **Figure 1** is a *photo-conductive cell*, also called a *light-dependent resistor* (LDR), a device whose resistance changes according to the amount of light falling on it. In bright light, the resistance is low (about 1 kΩ), whereas in the dark, its resistance is very high (up to 10 MΩ). The cell is made from a semiconducting material known as cadmium sulphide (CdS), and is enclosed in a small plastic container. The semiconductor is laid on a flat insulating surface in the form of a small flat ribbon. The ribbon construction gives a good area of surface for a given

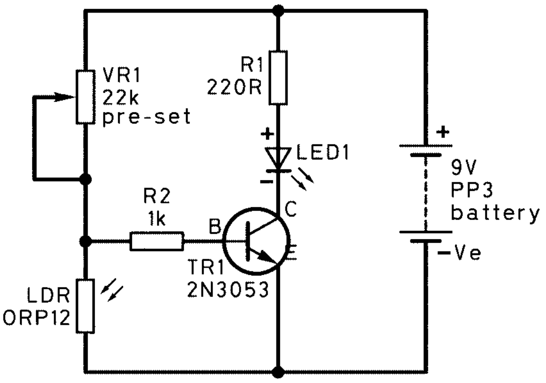
VR1
22k
pre-set
R1
220R
+
LED1
−
+
9V
PP3
battery
−Ve
R2
1k
B
C
E
TR1
2N3053
LDR
ORP12

Figure 1 Circuit diagram

©RSGB DY107

length of ribbon, and the length of the ribbon is maximised by laying it out in a zig-zag pattern, as can be seen diagrammatically in **Figure 2**. In the dark, CdS is an insulator; when light falls on it, electrons are released inside the CdS, making it conduct. The more light there is, the more electrons there are, and the resistance falls.

In this circuit, the LDR is connected across the 9 V supply in series with a variable resistor, VR1. In this arrangement, the voltage that exists across the LDR will be determined by the light level. As the light intensity increases, the resistance of the LDR falls, dropping a smaller voltage across it. The

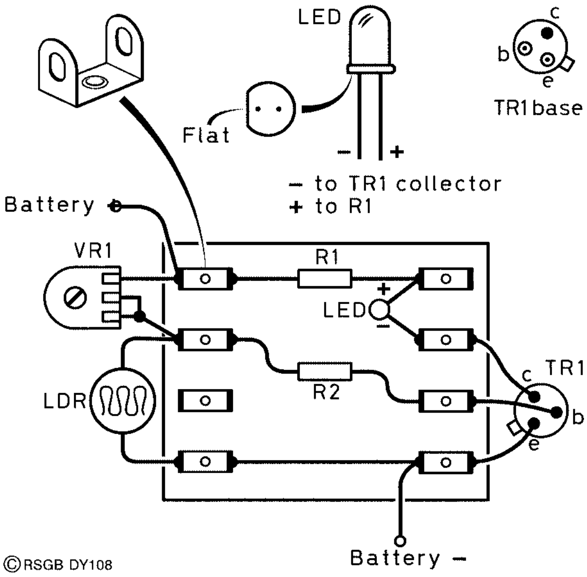
LED
Flat
− +
− to TR1 collector
+ to R1
c
b
e
TR1 base
Battery +
VR1
R1
+
LED
−
LDR
R2
c TR1
b
e
Battery −

Figure 2 Component layout on tagstrip

©RSGB DY108

reverse happens when the light intensity falls – the voltage across the LDR increases. This voltage is used to drive an npn transistor, TR1, connected as an electronic switch. As the voltage on the base (B) rises, it will reach a point where TR1 will suddenly start to pass current, just as an ordinary switch does when pressed. The current flowing through TR1 also flows through a flashing light-emitting diode (LED), D1 and its series resistor, R1. D1 can be a steadily glowing type, if preferred. If you want to make the circuit switch the LED on at a different ambient light level, adjust VR1.

Construction

The prototype was made on an eight-tag tagboard (Figure 2). The resistors R1 and R2 can be laid on the tagboard for soldering, the rest of the components lying above or to the side of the board. Check the circuit after you have soldered everything on, then connect the battery. The LED should flash if you put your hand over the LDR, and VR1 can be adjusted to vary the point at which the LED lights. The project can be housed in a small plastic box with holes provided for the LDR and LED and an on/off switch if you want one.

Parts list

Resistors: all 0.25 ohm, 5% tolerance

R1	220 ohms (Ω)
R2	1 kilohm (kΩ)
VR1	22 kilohms (kΩ) linear pre-set

Semiconductors

LDR	ORP12
TR1	2N3053 npn
LED1	Flashing LED

Additional items
Plastic box
Tagboard with 8 tags
PP3 battery connector

43 A J-pole aerial for 50 MHz

Introduction

For FM communication (i.e. voice and data) on the VHF bands, a vertical aerial is used to give all-round (non-directional) coverage. This is a half-wave aerial which can be fed at the end, thus removing the principal problem with the conventional vertical centre-fed half-wave dipole, which is that the feeder should leave the dipole at right angles. This is no problem when the dipole is horizontal, but can be difficult for the vertical dipole.

Basic facts

A feeder must be connected to an aerial at a point where the *impedance* (AC 'resistance') of the aerial closely matches that of the feeder. The difference between the two impedances gives rise to the voltage standing-wave ratio (VSWR), which is unity only when the two impedances are the same. With 50 Ω feeders, the feed point of a half-wave aerial is at the centre, where the aerial impedance is around the same value. At the end of a half-wave aerial, the impedance is high, so it is *not* a suitable point to connect a 50 Ω feeder.

Connection at this point can be effected using an RF transformer. RF transformers act in the same way as ordinary transformers, except that they are much smaller, and usually comprise wires of particular lengths adjacent to each other. **Figure 1** is a good starting point. It shows the aerial in its diagrammatic form. Notice that the aerial is in the form of an elongated letter 'J'; this shape gives rise to its nickname – the *J-pole*. The quarter-wave RF transformer is the lower 'U' section below the half-wave element. At the bottom of the U section, the impedance is zero (this may become clearer later) and at the top of the U section it is high, thus matching the aerial impedance. The coaxial feeder cable is connected part-way up the U section, where the impedance is around 50 Ω.

The practicalities

Figure 2 shows the aerial as constructed. As it is about 4.5 metres high, it may be too high for the average house loft, but is ideal for mounting outside, supported by a non-metallic pole or hung from a tree branch. The upper half-wave section is made from 1.5 mm insulated copper wire, as used in domestic mains wiring. The quarter-wave transformer below the half-wave section is made from 300 Ω balanced line ('ribbon cable'). The

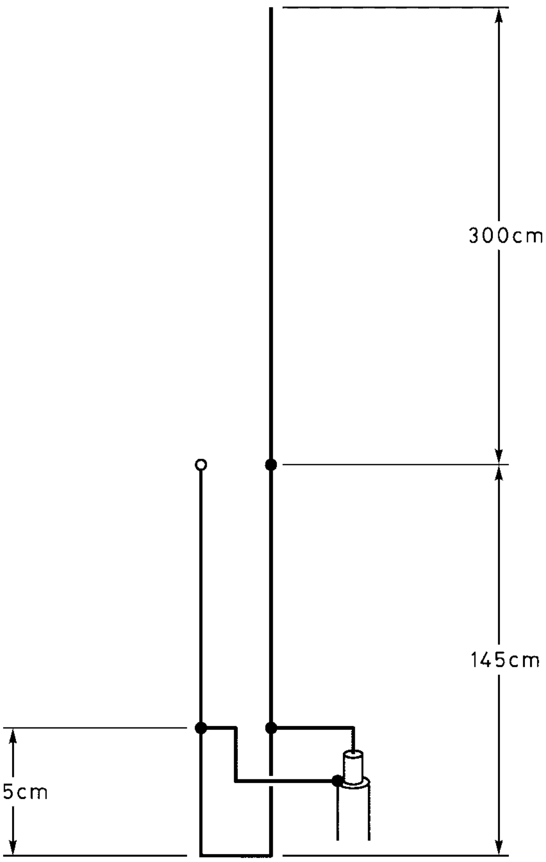

Figure 1 Overall dimensions of the 6 metre J-antenna

wires at the bottom of the transformer section are stripped of their insulation, twisted and soldered. At the upper end, only one wire of the balanced line is soldered to the bottom of the half-wave section. The other wire of the pair is not connected and is left insulated.

At the feed point of the transformer, the insulation needs to be carefully stripped from the balanced line. You will need a standing-wave meter (VSWR meter) in the coaxial line between your transmitter and the aerial, and you will need to adjust the position of the feed point. 45 cm from the bottom was the best point on the prototype, but this position is dependent upon the immediate surroundings of the aerial, and *must* be done when the aerial is in its final operating position. **Warning: *Never* make adjustments to the feed point when the transmitter is on. Make a VSWR measurement, switch off, move the feed point, switch on again, make another measurement, and so on.** You will need to aim for the lowest VSWR you can – certainly better than 2:1. Having found the best position, wrap all the exposed wires with self-amalgamating tape, to seal them against the ingress of moisture.

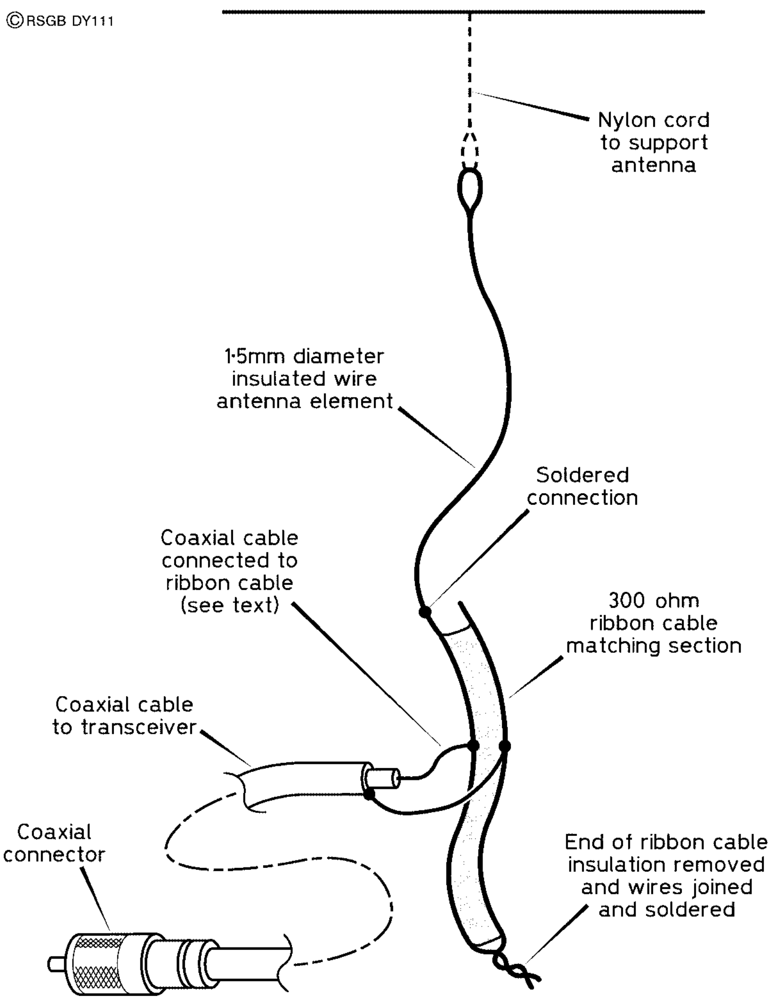
©RSGB DY111
Nylon cord to support antenna
1·5mm diameter insulated wire antenna element
Soldered connection
Coaxial cable connected to ribbon cable (see text)
300 ohm ribbon cable matching section
Coaxial cable to transceiver
Coaxial connector
End of ribbon cable insulation removed and wires joined and soldered

Figure 2 Construction details

How it performs

Figure 3 shows a computer prediction of how the J-pole radiates. It is called a *polar diagram*, and shows the distribution of your transmitted power when viewed 'from the end of your garden'. Most of your signal is sent at a fairly small angle to the horizontal; very little signal goes upwards, which is a good thing, of course. This also shows why the J-pole (or any other vertical aerial) should not be called 'omnidirectional', which means it radiates in all directions. It is omnidirectional *only* in the horizontal plane.

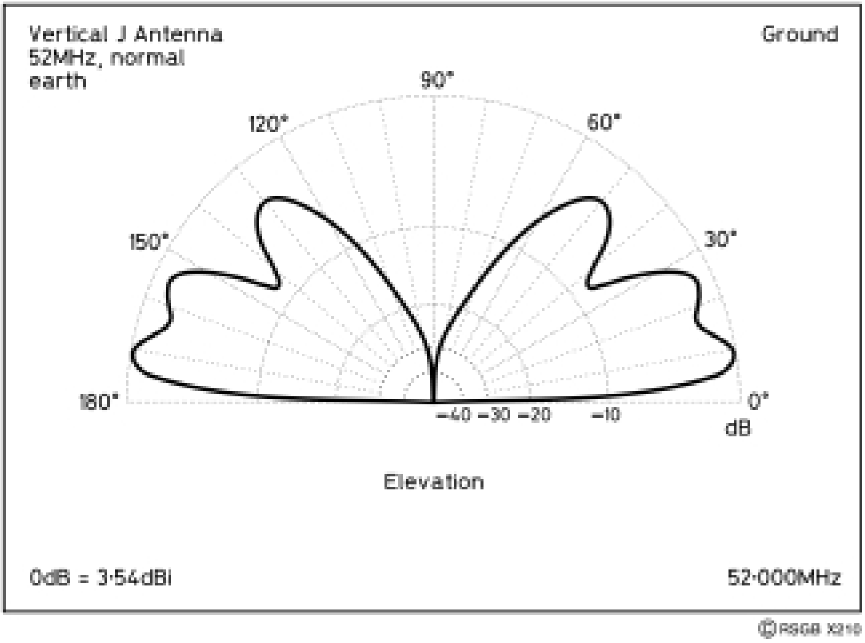

Figure 3 This polar diagram computed for the J-pole

Safety

Where you mount your aerial is a matter of personal preference and the restrictions of height and space, but the following safety rules *must* be applied.

1. Never fix an aerial where it may come into contact with power lines or telephone lines.
2. When climbing a ladder to put up an aerial outside, make sure the ladder is safe and that it is secured.
3. Don't do this alone. Preferably have someone with you. If this is not possible, make sure someone knows where you are.

Parts list

3.00 metres	1.5 mm insulated copper wire
1.50 metres	300 ohm balanced line
As required	50 ohm coaxial cable
As required	Self-amalgamating tape

44 Measuring light intensity – the photometer

Introduction

Before the days of automatic 'point-and-shoot' cameras, a photographer would use a light meter or *photometer* to measure the light level, then manually convert this reading into shutter speed and lens aperture settings to ensure a correctly exposed negative. Modern cameras have quite sophisticated photometers, which control the shutter speed and iris settings automatically.

A short explanation

A simple photometer circuit is shown in **Figure 1**, and is based on a device called a *light-dependent resistor*, or LDR. As its name tells us, its resistance depends upon the amount of light falling on it. In bright light, the resistance is relatively low (about 1 kΩ), whereas in the dark, its resistance is very high (up to 10 MΩ). The cell is made from a semiconducting substance known as cadmium sulphide (CdS), and is enclosed in a small plastic container. The semiconductor is laid on a flat insulating surface in the form of a small flat ribbon. The ribbon construction gives a good area of surface for a given length of ribbon, and the length of the ribbon is maximised by laying it out in a zig-zag pattern. In the dark, CdS is an insulator; when light falls on it, electrons are released inside the CdS, making it conduct. The more light there is, the more electrons there are, and the resistance falls.

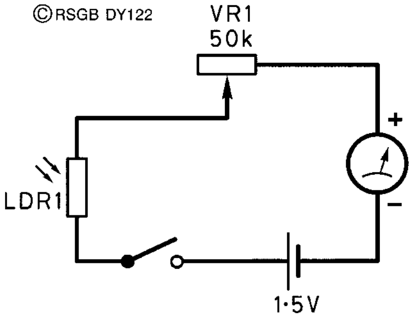

Figure 1 Circuit diagram of photometer

The photometer circuit

The circuit is simply a series connection of four things – the battery, the LDR, a variable resistor and a meter. A switch is also provided. The combination of the LDR resistance and that of VR1 determines the current flowing, which is indicated on the meter. Altering the resistance of VR1 sets the *sensitivity* of the photometer – you may want full-scale deflection of the meter needle for a bright light, or for a dim light.

It is simple to make, and a plug-in type matrix board is ideal to test the circuit, so that you can decide if you want to make a permanent version. Connect all the components in series; the only change you may want to make is the connection to VR1. If you find that the sensitivity control seems to work 'backwards', simply unsolder the wire from the end tag of VR1 and solder it to the opposite end tag. Problem solved!

In use

As soon as you connect up the battery, you will probably have a meter reading because of the daylight falling on the LDR. Shading it with your hand should reduce the reading. If the meter needle is hard over against the end-stop, turn VR1 until it indicates about half-scale. The LDR is very sensitive, and will read zero only in a dark room. If you put on a torch to see what the meter reading is, the LDR will detect the torch light, and will give a reading!

Here is a simple project where you can use the photometer in an experiment which has an analogy in radio. Draw a circle on a large (A3 or bigger) sheet of paper and divide it up into 30-degree sectors, as shown in **Figure 2**. Draw a series of smaller circles which divide the maximum radius into five. Look at the figure if you're not sure about this. Bring the LDR away from the circuit by using two long, flexible wires. The experiment must be performed in a darkened room (preferably in total darkness). Prepare a table with two columns, the left-hand one headed 'Angle (degrees)' and the right-hand one 'Meter reading'. Fill in the left-hand column 0, 30, 60 . . . and so on up to 360°.

Place the torch in the position shown, with its lens at the centre of the circle and pointing along the 0° line; switch it on. Place the LDR facing the torch and adjust VR1 until you have full-scale deflection of the meter needle. Suppose the meter indicates 10 units at this point. Enter this into your table in the 0° row. Keeping the torch the same distance from the circle centre, and pointing at it, move the LDR round all 30° positions and record the meter readings. Switch off the torch and take the sheet of paper and your tabulated results into daylight!

Lay the large sheet of paper on a table with your results beside it; then, at each 30° interval, plot the point along the radius corresponding to the meter

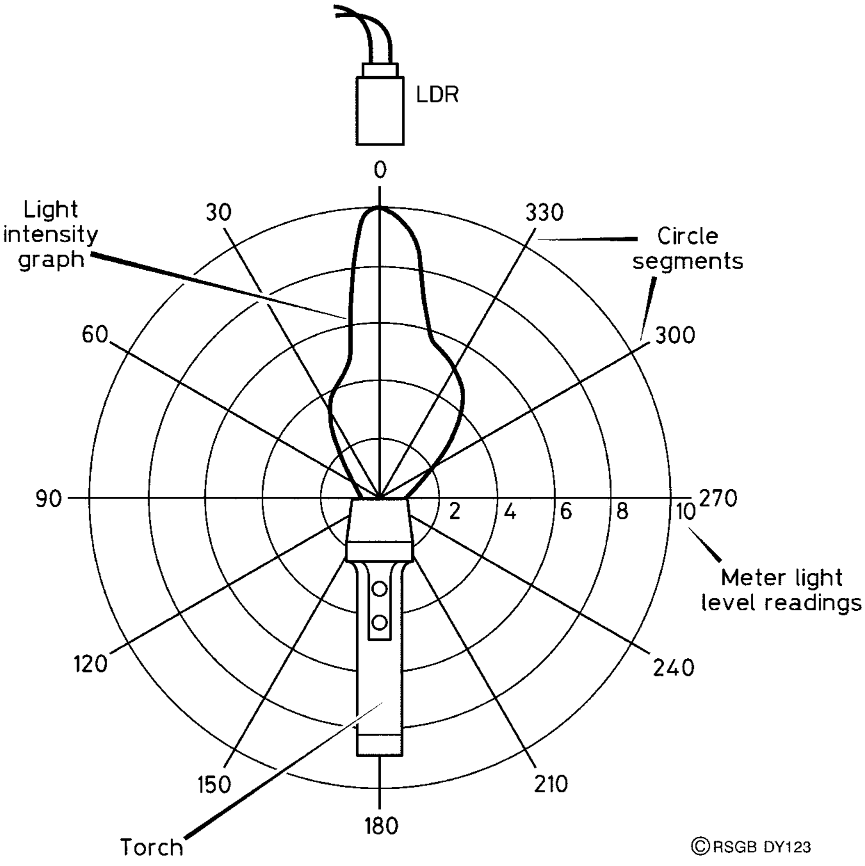

Figure 2 Torch light intensity pattern plotted at RSGB HQ

reading. Then, join up all the points and you have what is called a *polar diagram* of light intensity. The use of the word 'polar' implies that the readings have been taken in a circle and plotted that way.

Light waves and radio waves are both examples of *electromagnetic radiation*. The torch is designed to 'beam' its radiation in a particular direction, just like an aerial does. Hence the use of the word 'beam' for a directional aerial. If a similar polar diagram is drawn for a Yagi-type aerial, it will show the same general characteristics as does Figure 2, namely a main direction (or 'lobe') where most of the energy is concentrated, with evidence of *sidelobes*, indicated by 'lumps' on the otherwise smooth main lobe.

Also in this book you will find a project for the construction of a UHF field-strength meter, which you could use to carry out a measurement of the polar diagram of a UHF aerial.

Parts list

Resistor

VR1	50 kilohms (kΩ) linear

Semiconductor

LDR1	ORP12

Additional items

Meter, 50 or 100 μA
Battery clip for single AA cell
AA cell

45 A 70 cm Quad loop aerial

Introduction

This is a description of how to make an aerial for the 70 cm band which has gain compared with the 'rubber duck' or the dipole aerial. It can easily be dismantled and reassembled, making it ideal for contest use.

The principles

Most aerials which comprise several similar *elements* arranged along a *boom* are variations of the design originated by Yagi and Uda, and which takes the name (for historical reasons) of the former, and is know as the Yagi array. Let us suppose we have a Yagi aerial beaming left to right in front of us. The elements get progressively shorter from the left (look at almost any TV aerial to see this). All the elements on a Yagi aerial are classified as follows:

- The *reflector* – the leftmost element as we look at the array. It is the longest. Next to it is:
- The *driven element* – this is the element connected to the feeder, which in turn runs down to the transceiver. It is slightly shorter than the reflector.
- All the elements beyond the driven element are called *parasitic elements*, or *directors*. They are shorter than the driven element and usually get progressively shorter as we progress to the right along the boom. The directors are mainly responsible for the *directivity* (or *beamwidth*) of the array.

The progression from a simple dipole (a driven element in isolation) to a Yagi array is simple, but is nevertheless important. To make an aerial of two elements, a reflector (*not* a director) is added to the driven element. For three or more elements, directors are added to the two-element design. Adding more and more directors soon becomes impractical, the reduction in beamwidth (such as it is) does not warrant the extra expense, weight and wind resistance that is incurred.

Instead of using linear (straight) elements, as in the generic Yagi, this design uses loops. Designs using squares of wire instead of loops are known as *Quad* aerials, and HF designs require large X-shape frames to support the large squares of wire. At 70 cm, however, the use of wire loops is easier, and they are self-supporting.

Construction

This is quite simple. Any type of material (metal, plastic, wood) can be used for the *boom* (the support for the elements) and for the mast. The elements are made from 14 SWG enamelled copper wire. 16 SWG hard-drawn aerial wire, which is not enamelled, has also been used with success. Thinner wire might result in a rather ‘floppy’ aerial! The separate parts of the aerial are held together with *jubilee clips* (hose clips).

The driven element is secured to the boom with a jubilee clip and a three-connector plastic connector block as shown in detail in **Figure 1**, and in the photograph. Cut the wire for the driven element 70 mm longer than the 700 mm indicated in Figure 1. Then, using sandpaper, remove the enamel from one end to a distance of 20 mm, and from the other to a distance of 50 mm. After forming the loop of the driven element, bend both stripped ends through 90°, and insert them into the first to holes of the plastic connector block (Figure 1). Do not tighten the screws yet. Push both ends into the block as far as they will go, then bend the 50 mm end back on itself and pull the ends back through the connector block so that the end you have just bent goes into the third hole in the block. Now tighten the screws in the block and in the jubilee clip.

Each director and the reflector should be made 40 mm longer than the circumferences shown in Figure 1. Strip the enamel, as before, from the last 20 mm at each end. Form the wire into the loop, slip the ends under the jubilee clip (Figure 1) and tighten it. You may find that it helps to solder the stripped ends together before securing the jubilee clip.

The boom is fixed to the mast using jubilee clips and wire, as shown in Figure 1. Solder the feeder cable to the driven element, with the braid soldered to the end of the driven element which is connected to the boom (this applies to metal booms; with plastic or wooden designs, the feeder connections are not critical).

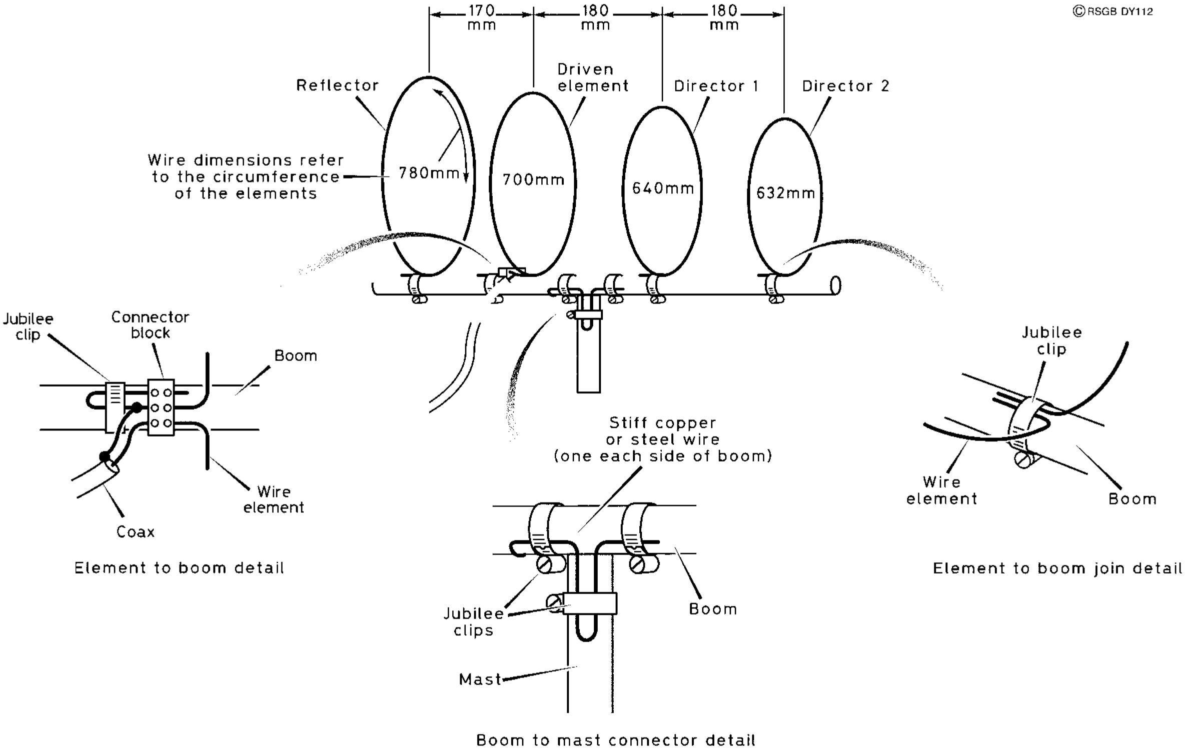
©RSGB DY112
170 mm
180 mm
180 mm
Reflector
Driven element
Director 1
Director 2
Wire dimensions refer to the circumference of the elements
780mm
700mm
640mm
632mm
Jubilee clip
Connector block
Boom
Wire element
Coax
Element to boom detail
Stiff copper or steel wire (one each side of boom)
Jubilee clips
Boom
Mast
Boom to mast connector detail
Jubilee clip
Wire element
Boom
Element to boom join detail

Figure 1 70 cm 4 element quad construction, with detail of how hose clamps are used in the construction

You will need a plug on the shack end of your feeder to suit the transceiver, aerial tuning unit (ATU) or the standing-wave meter (SWM) you are using. Always connect these in the order: transceiver – SWM – ATU – aerial.

Testing

Always test aerials outside and away from buildings (if possible!). This avoids getting misleading results.

Use a rubber duck, or whatever aerial you usually use, and tune around to find a repeater or beacon signal which is consistent. Note the reading on the S-meter. Then, connect your new Quad loop. Rotate it to give the strongest S-meter reading (don't forget it is directional). Verify that the meter reading varies as you rotate it. How does the S-meter reading compare with the original reading? Much depends on the siting of your original aerial; if you are comparing your Quad loop at ground level with a vertical on the chimney, you wouldn't expect your new aerial, even with its gain, to outperform a vertical which is well elevated!

Now a test on transmit is called for. The use of an SWR meter is essential here. Find a clear frequency and check that it *really is clear* before announcing your presence and carrying out the test. A reading of 1:1 is excellent, but any value less than about 1.8:1 would be acceptable. You can measure the directivity of your aerial using the field strength meter, also described in this series.

Materials

4 metres of 14 SWG enamelled copper wire
Material for boom and mast
15 amp connector block
7 Jubilee clips

The enamelled copper wire is available from AA&A Ltd, Sycamore House, Northwood, Wem, Shropshire SY4 5NN. Everything else is available from most hardware stores.

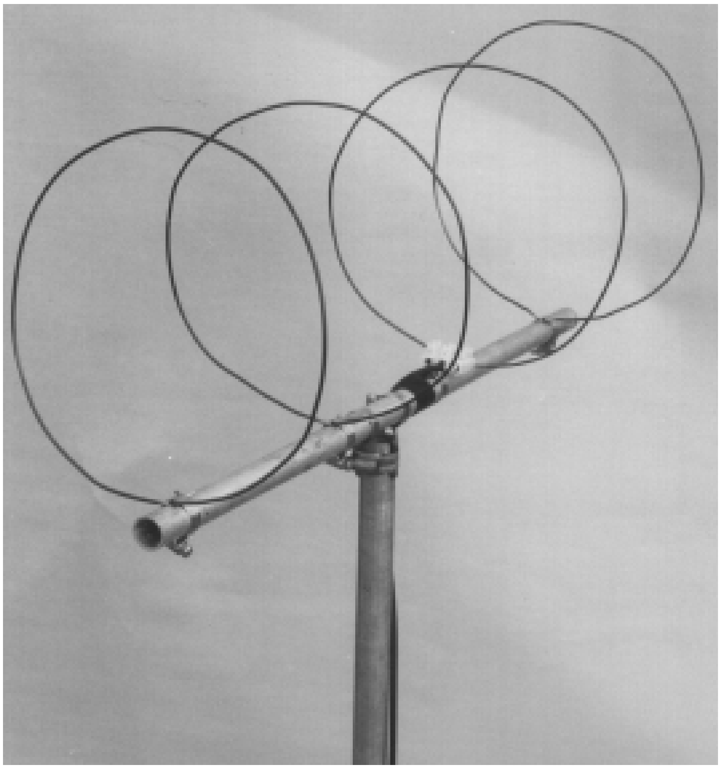

46 A UHF field strength meter

Introduction

It is always interesting, and often useful, to know just *where* the radiation is going from your aerial. How much of your radiated energy is going in the general direction of the station you are in contact with, and how much is being effectively wasted? Some of these questions can be answered with the use of a *field strength meter.* A field strength meter is simply a receiver, stripped down to its bare essentials, such that it responds only to the magnitude of the carrier. The use of a field strength meter assumes that the aerial under test is radiating a continuous carrier. Don't forget to find a clear frequency and identify your transmissions at least every quarter-hour, in order to comply with the terms of your licence.

Description

Two types of field strength meter are shown in **Figure 1**. You will recognise both circuits (Figure 1b particularly) as being types of 'crystal set' with a meter replacing the headphones. Figure 1a is a broad-band HF design (there is no tuning provision) and Figure 1b is tuned in the same way as the crystal set; with a loop of wire as an aerial, it will perform well in the VHF/UHF range.

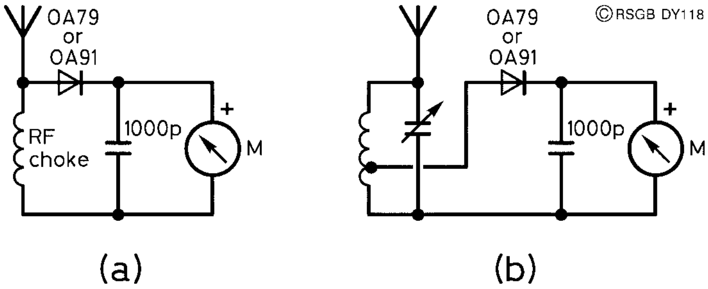

Figure 1 Construction of the UHF field strength meter

A tuned field strength meter also doubles as an *absorption wavemeter* if it is well calibrated. Such devices are useful for detecting transmitter harmonics also. This design is intended for use in the UHF band, so it will have to be sensitive around 432 MHz.

Construction

The field strength meter comprises a loop of wire, 600 mm long, which acts both as aerial and as the tuning inductance, a diode, a capacitor, a connector block, a meter and a length of twin wire. All the components, with the exception of the meter, are fixed to a pole with a jubilee clip, as shown in **Figure 2**. The meter should have a sensitivity of between 50 µA and 100 µA, or a multimeter can be used. The multimeter is more flexible, as you can select different current ranges, giving you a range of sensitivities.

Using the field strength meter

Connect your meter to the ends of the twin wire from the pole. Place a hand-held transceiver about 2 metres away, and press Transmit. If there is no meter reading switch off the transmitter and check the wiring. If the needle attempts to go negative, simply reverse the wires to the meter. If the reading is too high, either move the transmitter further away, or increase the current range on the multimeter. Try changing the orientation of the transmitter aerial, and note how the signal varies.

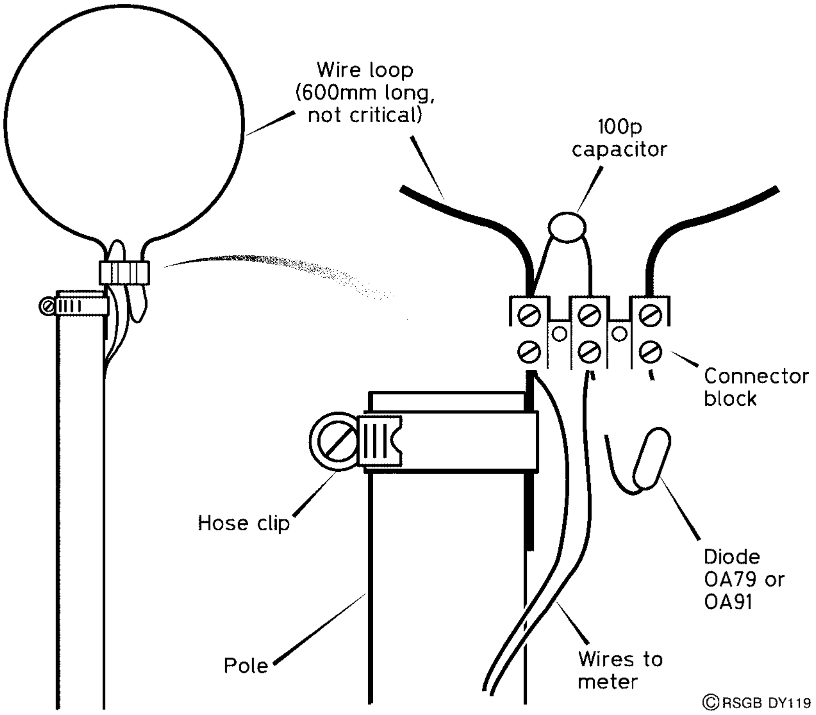
Wire loop
(600mm long,
not critical)
100p
capacitor
Connector
block
Hose clip
Diode
OA79 or
OA91
Pole
Wires to
meter
©RSGB DY119

Figure 2 Assembly of the field strength meter

To measure the *polar diagram* (a graph of the signal strength against the angle between the aerial boom and the field strength meter, plotted on polar axes) of a beam aerial, mount the loop as far away from the aerial as you can, preferably at the same height, and rotate the aerial, say 15 degrees at a time, and record the signal strength, until the aerial has been turned through 360 degrees. If you need help plotting the graph, enlist the help of a friend who has done it before, or one who knows about polar graphs! If you have already built the Photometer project, you will have measured and plotted the polar diagram of the light intensity from a torch. Now you will see the great similarity!

Parts list

Capacitor	100 picofarads (pF)
Diode	Germanium, OA79 or OA91
Connector block	10 A, 3-way
Wire	600 mm of 16 SWG enamelled copper Length of twin cable for meter connection
Clip	Jubilee (hose) clip

47 Christmas tree LEDs

Introduction

A novelty ideal for the festive season, this circuit causes one LED at a time to light up around a small cardboard Christmas tree.

Warning

This circuit uses members of the integrated circuit family known as CMOS (complementary metal-oxide semiconductor). These use very little current and can be completed destroyed if they come into contact with the magnitudes of static electricity that most of us carry about when we walk on carpets and wear rubber shoes. You will never know if this wanton destruction has happened – all you will discover is that your circuit doesn't work and that you have tested *everything*. To avoid this problem do the following things:

1. **Before you open the little packet in which each IC is supplied, touch something which you *know* to be earthed – the metalwork of any equipment which is mains-earthed, for example. Then open the packet.**
2. **Let the IC fall gently on the bench – don't pick it out with your fingers. Touch your earthed metalwork again. Pick up the IC and insert it gently into its holder.**
3. **Repeat the process with the second IC.**

The circuit is safe from destruction while it is connected to the battery. However, when the battery is removed, the same care should be exercised with its handling, because there are no supply decoupling capacitors across each IC.

Description

Figure 1 shows the layout, and how the wires are connected from the circuit board to the LEDs around the tree. The circuit, shown in **Figure 2**, is quite complicated, so you need to be confident in your logical approach to circuit-building before you attempt this one! It uses two common integrated circuits (ICs). IC1 is simply an oscillator which provides timing pulses for IC2, which 'counts' up to a maximum of 10. The outputs of the counter are indicated by light-emitting diodes (LEDs); red, orange, yellow and green are common colours which you can use.

The circuit is built on a single piece of Veroboard measuring 29 holes by 12 strips. **Be aware that there is no strip labelled 'I', so don't make mistakes in your counting!**

First of all, cut the tracks in the positions shown in Figure 1 using a 3 mm ($\frac{1}{8}$ inch) twist drill rotated between thumb and forefinger. Then, solder in the IC sockets, the notches facing row M. Wire up and solder in the links and then the Veropins for the connections to the LEDs and battery. Having done this, the resistors and the capacitor should be fitted. When wiring the LEDs, each cathode (the lead adjacent to the 'flat' on the LED encapsulation) is connected to R4, and the anodes go to separate pins on IC2.

Testing

Hold up your circuit board to the light and check carefully for solder bridges between adjacent tracks. Then check again that the wiring is correct. Place the ICs in their holders, with the notched ends lining up with those on the holders. Connect the battery, and the LEDs should illuminate in sequence.

If you have no success, you are now wishing you had checked the circuit more carefully! Learn something from your mistake and it will not have

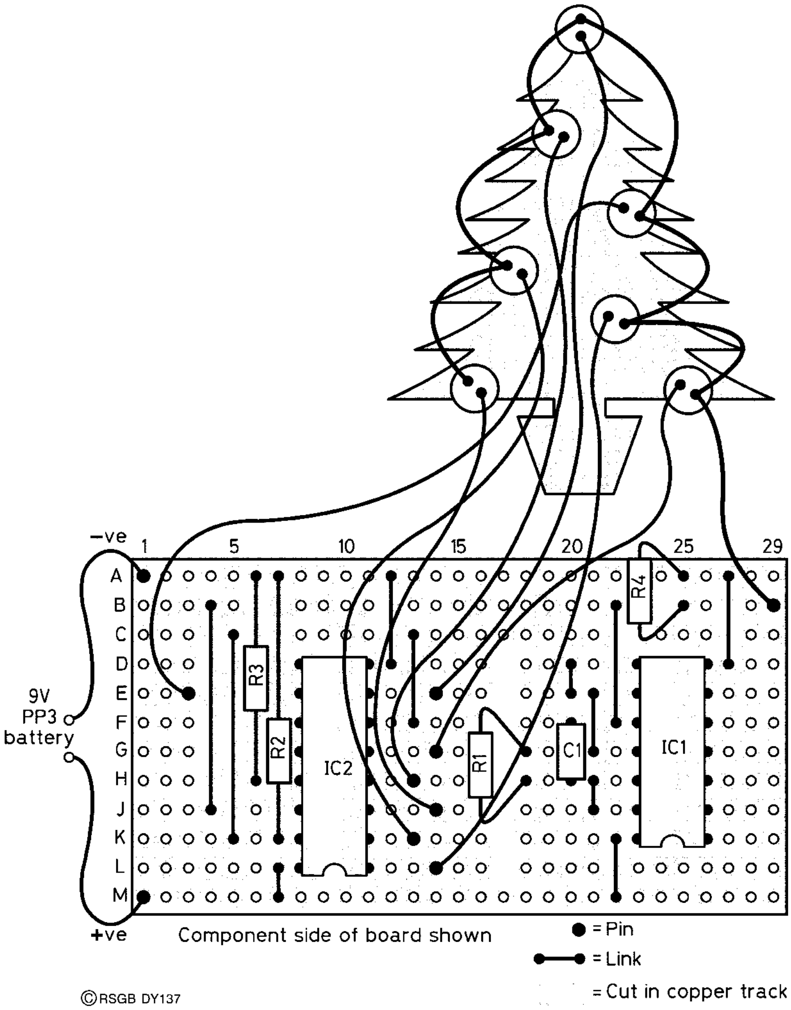
−ve
1
5
10
15
20
25
29
A
B
C
D
E
F
G
H
J
K
L
M
9V
PP3
battery
R3
R2
IC2
R1
C1
IC1
R4
+ve
Component side of board shown
● = Pin
●—● = Link
= Cut in copper track
©RSGB DY137

Figure 1 Christmas tree, component layout

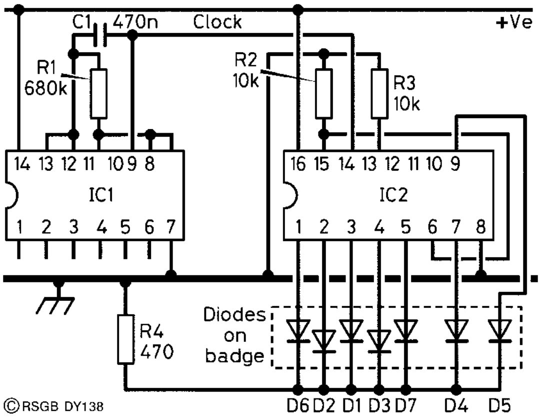
C1
470n
Clock
+Ve
R1
680k
R2
10k
R3
10k
14 13 12 11 10 9 8
IC1
1 2 3 4 5 6 7
16 15 14 13 12 11 10 9
IC2
1 2 3 4 5 6 7 8
R4
470
Diodes
on
badge
D6 D2 D1 D3 D7 D4 D5
©RSGB DY138

Figure 2 Christmas tree, circuit diagram

been in vain! The first thing to check is that the LEDs are the correct way round. If that is OK, then check the positions where the tracks are broken (intentionally!). After that, are the wire links in the right places – had you forgotten there is no row ‘I’? Check *all* the wiring, then check *again* for solder bridges, and switch on again. One of these tests should have revealed a fault. If it *still* refuses to work, perhaps one (or both) of your ICs were damaged by static electricity, despite your precautions – or did you choose to ignore them?

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1	680 kilohms (kΩ)
R2, R3	10 kilohms (kΩ)
R4	470 ohms (Ω)

Capacitor

C1	0.47 microfarad (μF) min. metallised polyester film

Semiconductors

IC1	4011
IC2	4017

Additional Items

D1–D7	5 mm LEDs in choice of colours
	14-pin DIL socket for IC1
	16-pin DIL socket for IC2
	PP3 battery clip and battery
	Veroboard, Veropins
	Insulated wire for links

48 An audio signal injector

Introduction

An audio signal injector is a device used to test audio frequency circuits. It is simply an oscillator running at a frequency in the audio range, so that when its output is fed to the input of an amplifier, it will produce a sound in the loudspeaker if the amplifier is working. The oscillation is so rich in harmonics that the signal can also be heard (sounding rather different) when injected into an RF circuit.

The design

The circuit is shown in **Figure 1**, and is a basic *astable multivibrator*, a free-running oscillator producing a roughly rectangular-wave output. The two transistors, TR1 and TR2, operating as switches, switch on and off alternately at a frequency around 500 Hz. The prototype was constructed on plain matrix board (no copper strips), as illustrated in **Figure 2**.

Both transistors are type BC108, which are only a few pence each new, and can be found at almost any rally. You can add an on/off switch, or simply disconnect the battery when you are finished using it. To make the unit in one piece, the battery can be taped to the board, as the diagram shows.

The probe itself is made from a short piece of stiff insulated wire, soldered to a tag on the board; an earth lead is also soldered to the board, and terminated in a crocodile clip to attach to the ground lead of the equipment under test.

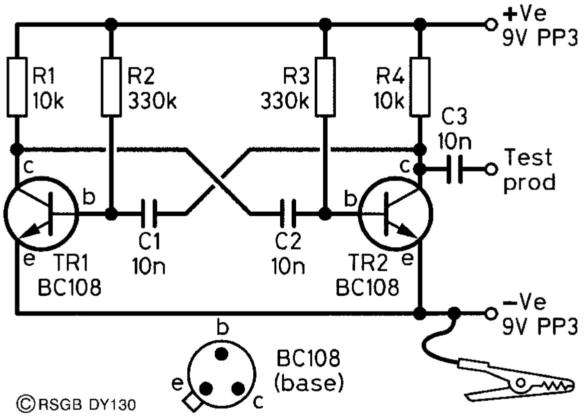

Figure 1 Circuit diagram of the signal injector

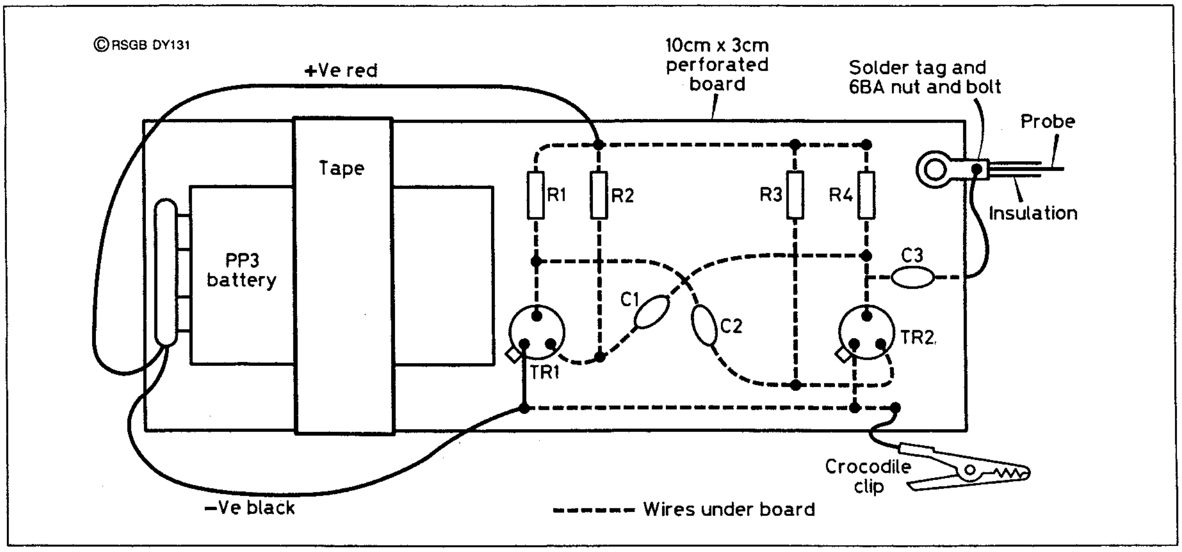
©RSGB DY131
+Ve red
10cm x 3cm perforated board
Solder tag and 6BA nut and bolt
Probe
Tape
PP3 battery
R1
R2
R3
R4
Insulation
C3
C1
C2
TR1
TR2
Crocodile clip
–Ve black
------ Wires under board

Figure 2 Component layout of the signal injector

Using it

After checking the wiring of the circuit, connect the battery. Find a transistor radio that is known to work. Carefully remove the case, switch on and apply the probe to the centre tag of the volume control. A *very* loud buzz should be heard from the speaker, indicating that the audio circuits of the radio are working. Using the injector to fault-find equipment you have made yourself is rather more instructive and rewarding, because you *know* where to inject the signal, and you should *know* what to expect when you do.

Warning

Do *not* work on any equipment connected to the AC mains. Work only on battery-powered circuits, for your own safety.

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1, R4	10 kilohms (kΩ)
R2, R3	330 kilohms (kΩ)

Capacitors

C1, C2, C3	10 nanofarads (nF) or 0.01 microfarad (µF) ceramic

Semiconductors
TR1, TR2 BC108

Additional items
PP3 battery and connector
Matrix board 10 cm by 3 cm
6BA solder tags
Thick insulated wire for probe
Crocodile clip

49 Standing waves

Introduction

Everyone has, or *should* have, a standing-wave ratio (SWR) meter as part of his/her array of test gear. Most people know how to use it, but what does it really do?

Before attempting to answer that question, we need to look at some aerial fundamentals. An aerial is a *transducer*, the word meaning 'to lead across'. We use it whenever one form of energy is converted into another form. A bulb is a transducer; it converts electrical energy in the filament to radiated heat and light energy. **Figure 1** shows the situation.

An aerial, or antenna, is also a transducer; it converts radio-frequency (RF) energy in the feeder (or *transmission line*, to give it its proper name) into radiated electromagnetic energy, in the manner shown in **Figure 2**. The aerial has resistance, just like the bulb filament, and if the filament had zero resistance, there would be no radiated energy. In the same way, if the aerial had no resistance, there would be no radiation from it.

Transmission lines should convey RF energy from transmitter to aerial with the minimum power loss. A common form of transmission line is the *coaxial*

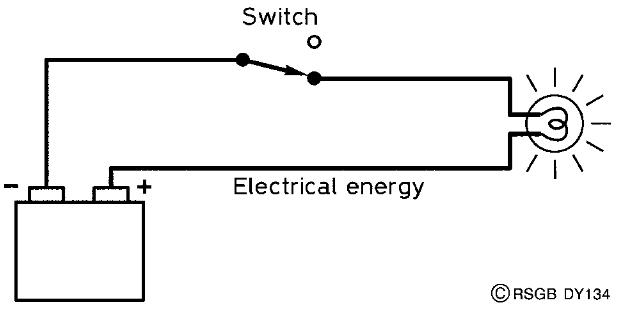

Figure 1 The light bulb converts electrical energy to electromagnetic light energy

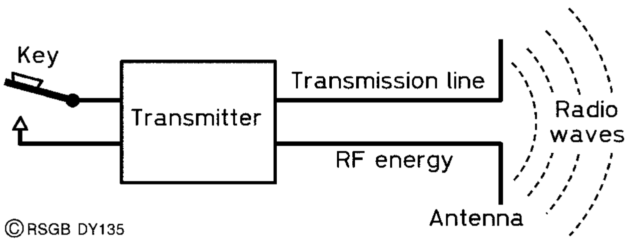

Figure 2 The antenna converts RF current to electromagnetic radio waves

cable. As you might expect, any feeder cable has a DC resistance caused by the resistance of the copper wire from which it is made. It also has an AC resistance, caused by the capacitance between the centre conductor and the braid, and by the inductance of the cable itself. This means that the feeder has an *impedance*, which is constant for the particular type of cable. This is what we call the *characteristic impedance* and, for the cables used in most amateur radio applications, it is 50 Ω.

If this impedance can be made the same as that of the aerial (it is already the same as the impedance of the transmitter output), then the transfer of energy will occur with minimum loss. If the aerial and cable impedances are *not* the same, then there is a *mismatch*, which causes some of the RF energy to be reflected back towards the transmitter. We now have a situation where RF energy is flowing along the cable from the transmitter to the aerial (the *forward* wave) and, at the same time, flowing from the aerial to the transmitter (the *reflected* wave). The two waves interact along the cable and form a stationary pattern of voltage and current. The pattern is known as a *standing wave*, and can be visualised from the waveforms in **Figure 3**. The ratio of the maximum voltage to the minimum voltage on a given wave defines the *voltage standing-wave ratio* (VSWR), or just *standing-wave ratio* (SWR) for short.

The SWR meter is easy to use. It is positioned in the feeder between the transmitter and the aerial tuning unit (ATU) if there is one, or between the transmitter at the aerial otherwise. Most meters have a single meter and a four-position switch. The transmitter is first keyed and, with the switch in

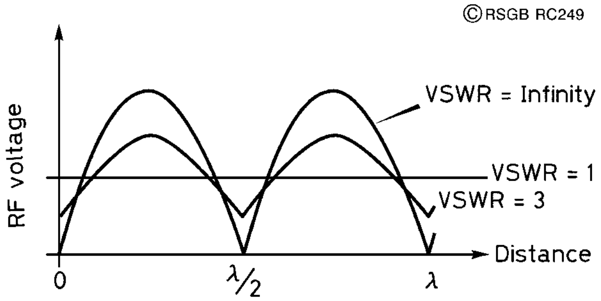

Figure 3 Standing waves on a transmission line

the 'calibrate' position, the sensitivity control is adjusted to give full-scale deflection on the meter. The switch is then changed to 'forward' to read the forward power, to 'reflected' to read the reflected power, and to 'SWR' to read the value of the standing-wave ratio. When there is no reflection (see Figure 3), the meter should read 1:1 or, simply, 1.

Most SWR meters remain in the feeder line while the transmitter is operating, so the condition of the aerial and feeder can be constantly monitored. Problems with the aerial (such as water entering the feeder at its junction with the driven element) are immediately shown up. Without the use of the SWR meter, the situation would slowly deteriorate over several months and you would be left wondering why so few stations were answering your calls!

50 A standing-wave indicator for HF

Introduction

The standing-wave ratio (SWR) meter shows how well the aerial system, including the feeder, is *matched* to the output of the transmitter. This design does not *measure* SWR, but it gives an indication of when the SWR is minimum for a given system of aerial and feeder. The design is usable on the HF bands from 1.8 to 28 MHz, and can be used at 50 MHz with reduced sensitivity.

How it works

There are two types of wave in any feeder: the *forward* wave, which travels from the transmitter to the aerial; the *reflected* wave, which travels back to the transmitter from the aerial. The presence of a reflected wave is evidence that some of your transmitted power is not being radiated, but is being returned to the transmitter to be lost as excess heat. When aerial and feeder are perfectly matched, there is *no* reflected wave, and all the power from the transmitter is radiated.

Referring to the circuit of **Figure 1**, a tiny fraction of the signal is removed by the transformer, T1, and by the capacitors, VC1 and C1. It is then detected by the germanium diodes, D1 and D2, and any residual RF removed by the capacitors, C2 and C3. The currents through the diode and meter (depending on the position of switch, S1) represent the forward and reflected signals. VR1 acts as a sensitivity control for the meter.

It pays to shop around for a suitable meter. Surplus types from tape recorders and hi-fi equipment are usually ideal for this purpose. A new one would cost several pounds. The more sensitive the meter, the more sensitive your indicator will be. Meter sensitivity is given by the current required to give full-scale deflection (FSD) of the pointer. One with an FSD of between 50 and 200 micro-amps (μA) is suitable for this circuit. The higher the FSD, the less sensitive the circuit.

Construction

The meter circuit and the sampling transformer (see **Figure 2**) are built and mounted on Veroboard of the copper-strip variety. It simplifies construction

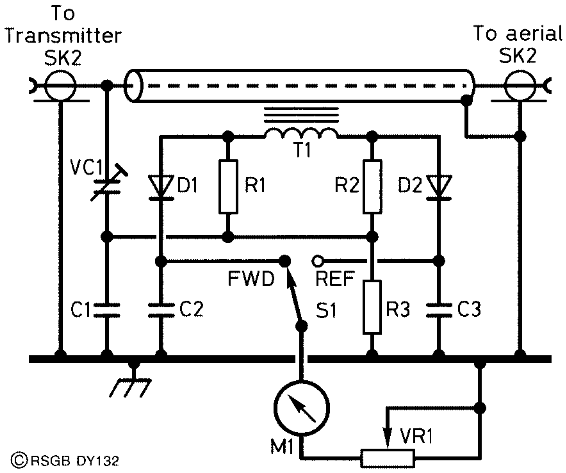

Figure 1 Circuit diagram of the SWR meter

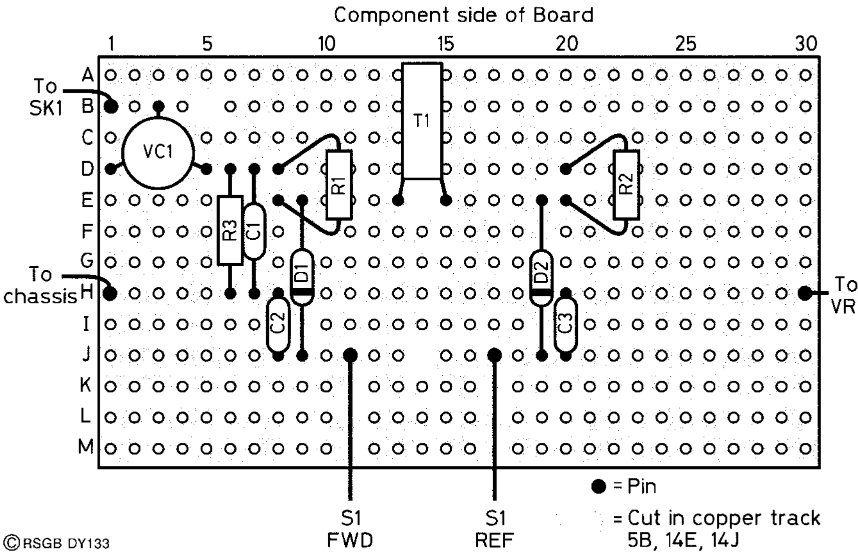

Figure 2 Component layout of the SWR meter

but reduces the operational range of the meter to below 30 MHz because of the capacitive coupling between strips. The board has 13 strips by 30 holes, although you can reduce this if you have a smaller case.

Firstly, cut the tracks at the three points shown. Then insert and solder Veropins for connections to the external components, the switch, variable resistor and the meter. Solder in the components starting with the resistors and followed by the capacitors and the diodes, ensuring that the diodes are inserted correctly.

Now you have to wind the transformer, T1, on a small toroidal ferrite core. Wind the secondary with 15 turns of 36 SWG enamelled copper wire, spaced evenly over about two-thirds of the former. The turns should not overlap, and considerable care must be taken; the wire is very thin, will kink easily and will break if you apply too much tension. The 'primary' is an 8 cm length of 50 Ω coaxial cable which passes through the toroid on its way between the input and output connectors. The braid of the cable is connected to the case at only *one* of the connectors (see Figure 1); this prevents the screen and the metal case between the two sockets forming a single, shorted turn.

The ends of the secondary winding must be carefully stripped of their enamel with sandpaper, before attaching the toroid to the board with cotton or nylon fishing line. On no account must wire be used for this!

Solder the secondary connections of T1 to the board and thread through the coaxial cable ready for soldering to the connectors.

The case used was an aluminium box (Maplin LF02C), but any suitable metal box could be used. Aluminium is preferable, as it is easily drilled with simple tools. Use standoff insulators to mount the board in the case. Once this has been done, the leads from the board to the chassis-mounted components can be soldered. So can the coaxial cable passing through the toroid. Make the lead from the input socket to VC1 as short as possible.

Setting up

You will need a 50 Ω dummy load and a transmitter to set up your indicator. Connect the transmitter to SK1 and the dummy load to SK2. Set the toggle switch, S1, to *forward* and the sensitivity control, VR1, to mid-travel. Switch on the transmitter, and set VR1 for maximum meter deflection. Switch to *reflected* and adjust VC1 until the reading is minimum (ideally zero). This completes the setting up!

Using the indicator

For setting up an aerial, connect your circuit between the transmitter and the cable leading to the aerial. With S1 in the *forward* position, key the transmitter and adjust VR1 for maximum reading on the meter. Switch to *reflected*, and then adjust your ATU to give minimum reflected power. If your adjustments are to be made to the aerial itself, to give minimum reflected power, you must make a note of the reflected reading, switch off the transmitter, make a change to the aerial, key the transmitter, and note whether the reflected power is greater or less than before. Then, make more changes to the aerial. **Never adjust your aerial with the transmitter on**. Make your adjustments on an unused frequency, and do it as quickly as possible, thus avoiding (or minimising) interference to other stations.

Parts list

Resistors: all 0.25 watt, carbon 5% tolerance
(or Maplin 0.6 watt metal film)

R1, R2	27 ohms (Ω)
R3	2.2 kilohms (kΩ)
VR1	10 kilohms (kΩ) linear

Capacitors

C1	220 picofarad (pF) disc ceramic 50 VDC
C2, C3	0.1 microfarad (μF) disc ceramic 50 VDC
VC1	20 picofarad (pF) trimmer

Semiconductors	
D1, D2	OA91 germanium
Additional items	
S1	Single-pole changeover (SPDT or SPCO)
SK1, SK2	Coaxial sockets to suit station standards
	Veroboard – 13 strips by 30 holes
	Veropins (7 off)
	Amidon FT 50–43 ferrite toroid
	Meter, less than 200 μA FSD
	36 SWG enamelled copper wire
	Short length of UR43 or RG58 coaxial cable
	Insulated stranded wire
	Aluminium box
	Standoff insulators for mounting the board
	Knob for the sensitivity control

51 A moisture meter

Introduction

Dry rot (*Merulius Lacrymans*) can strike havoc in buildings, causing the timbers to decay and crumble to dust – hence the term *dry rot*. Wood is attacked only if its moisture content rises above 20%.

Construction

The circuit of the moisture meter is shown in **Figure 1**. The two probes touch the wood, and the current that flows between them depends on the moisture content of the wood. If the moisture is sufficiently high, the current, after amplification, will be enough to light the LED.

The meter can be made on a piece of plain matrix board (no copper strips), as **Figure 2** shows. The board is big enough (10 cm by 2.5 cm) to accommodate the PP3 battery, taped on. No case is needed, unless you want

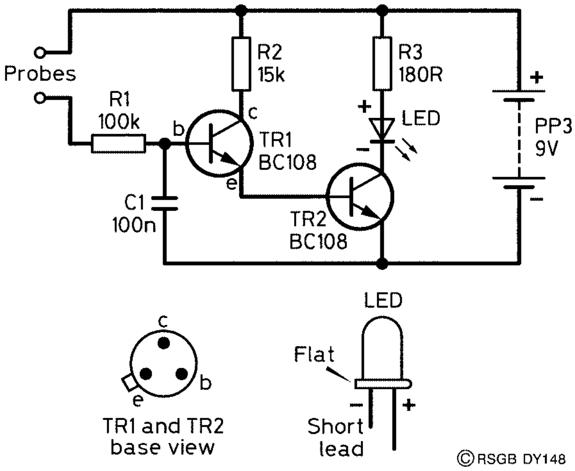
Probes
R1
100k
b
c
R2
15k
R3
180R
+
LED
−
TR1
BC108
e
TR2
BC108
C1
100n
+
PP3
9V
−
c
e
b
TR1 and TR2
base view
LED
Flat
−
+
Short
lead
©RSGB DY148

Figure 1 Circuit diagram of a moisture meter

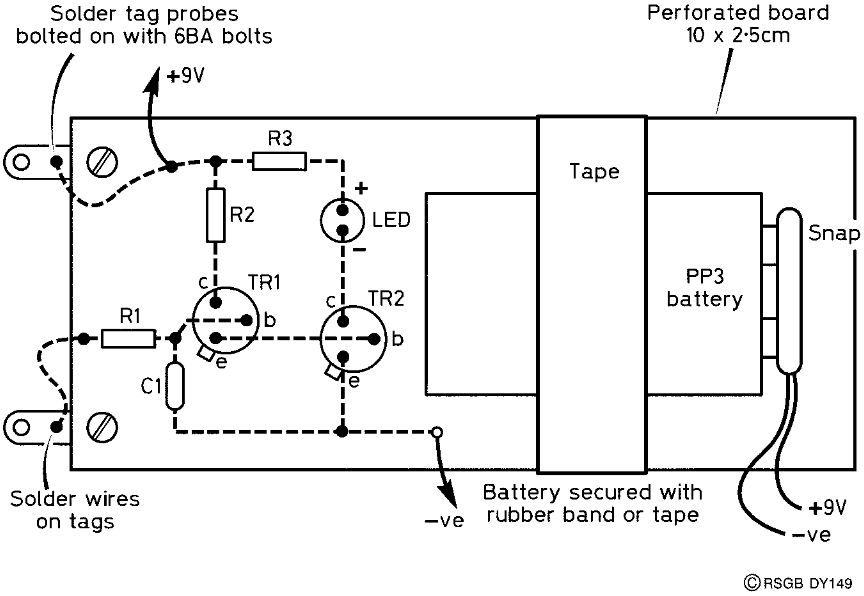
Solder tag probes
bolted on with 6BA bolts
+9V
R3
+
LED
−
R2
c
TR1
b
e
R1
C1
c
TR2
b
e
Perforated board
10 x 2·5cm
Tape
PP3
battery
Snap
Solder wires
on tags
−ve
Battery secured with
rubber band or tape
+9V
−ve
©RSGB DY149

Figure 2 Moisture meter, component layout

to leave the meter in a damp location for a protracted period! Make sure that the transistors and LED are mounted correctly. In Figure 2, the connections as shown to TR1 and TR2 are illustrated as if the transistors were transparent. An on/off switch is not really necessary, as only a very small current flows when the probes do not touch anything. Use solder tags, screwed to the board, to act as probes.

After the assembly is completed, check your circuit one final time, and then connect the battery. Nothing should happen at first. If you lick your forefinger and hold it across the probes, the LED should light.

Using it

The prototype was compared with a commercial moisture meter, and the LED lit when the moisture was around 20%. This was quite fortuitous, as the point at which the LED lights depends both on the separation of the probes and on the gain of the two transistors.

In addition to searching for dry rot, the instrument may be used to monitor the moisture in the soil of household plants. In this case, probes made of 16 SWG copper (*not* enamelled) should be soldered on to the two tags, and should penetrate the soil to a depth of several centimetres, and R1 may require adjusting so that the LED extinguishes if the soil is too dry, and lights if the soil is sufficiently moist. If you wanted to leave the meter with the probes in the soil, an on/off switch *would* be necessary.

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1	100 kilohms (kΩ)
R2	15 kilohms (kΩ)
R3	180 ohms (Ω)

Capacitor

C1	0.1 microfarad (μF) polyester

Semiconductors

TR1, TR2	BC108
LED	Any shape or colour will do

Additional items

PP3 battery and connector
Solder tags (2 off) for probes
Matrix board 10 cm by 2.5 cm

52 Simple aerials

Introduction

The performance of *any* receiver or transceiver, no matter how expensive it is, is limited by the aerial that feeds it. Two of the most frequently asked questions are:

- Which is the best sort of aerial to use?
- Where is the best place to locate an amateur radio aerial?

To answer these questions, you must ask yourself what sort of operation you want to do. Are you interested in local, chatty contacts on the lower bands or VHF, or are you more disposed towards long-distance (DX) contacts, and on what band?

A house with a moderately sized garden is assumed in the diagrams here, to illustrate the configurations of some simple aerials. You would not need all these aerials festooned around your house, because one or two would be sufficient for your needs. The problems incurred by properties with more restricted space will be covered later.

VHF aerials

For VHF operation, the aerial should be mounted as high as possible, either on a mast or on a chimney. For all-round coverage on FM and the local repeaters, a vertical colinear is a good choice. For SSB and CW DX operation, a horizontal rotatable beam is needed. If satellite working is envisaged, you will need to contemplate mounting an elevator on top of your rotator, so that your beam can point in *any* direction, including vertically upwards! An advantage of satellite working is that the aerials do not *necessarily* have to be up in the air, provided you have a relatively uncluttered site. Your rotator and elevator can be at ground level, which is good!

If the VHF aerial is mounted on the chimney, use a double mounting bracket, particularly if you have a beam and rotator. Keep the TV, broadcast FM and amateur aerials as far apart as possible, and keeping the feeders separated is also a good plan.

The dipole aerial

One of the simplest types of aerial for single-band operation is the half-wave dipole. (The name ‘dipole’ simply means ‘two poles’ or ‘two elements’, and

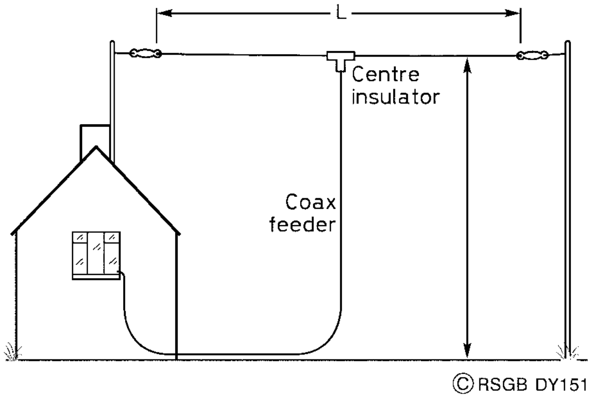

Figure 1 Layout for a dipole aerial

in this case the total length of the dipole is approximately half a wavelength at the operating frequency.) It is usually fed in the centre by coaxial cable as shown in **Figure 1**. The length of the dipole for the lowest frequency in each band is shown in **Table 1**. Normally, the length of the aerial will be 'trimmed' to be tuned to the centre frequency of the part of the band in which you will operate. This is done using the data in the right-hand column of Table 1. As an example, suppose you wanted your aerial to be resonant at 3.7 MHz. The table gives an overall dipole length of 42.86 m for 3.5 MHz. To resonate the aerial 200 kHz *higher*, then this length must be shortened by 2 × 0.595 m = 1.190 m. Your dipole would thus be 41.67 m long. Remember to allow extra wire for fixing the dipole ends to the insulators.

Table 1 Dipole lengths for lowest frequency of each band and the length to be trimmed from each to raise the resonant frequency by 100 kHz

Band (MHz)	*Dipole length (m)*	*Trim each end (mm/10 kHz)*
1.8	83.33	2190
3.5	42.86	595
7	21.43	150
10	14.85	70
14	10.71	35
18	8.33	20
21	7.14	15
24	6.03	12
28	5.36	10
50	3.00	6

On the lower-frequency bands, the lengths become rather large. In this case, you can 'bend' your dipole, as illustrated in **Figure 2**. The length of wire required to give an acceptable value of SWR (less than 2:1 on transmit) may need to be different from the calculated value, so be prepared to experiment!

Dipoles are single-band aerials, although they will often work acceptably on the third harmonic of their design frequency: a 7 MHz dipole often operates reasonably well on 21 MHz. It is possible to operate several dipoles in parallel, as **Figure 3** shows. Interaction between the elements can occur if the spacing between them is less than about 10 cm. A multi-band dipole, as shown in Figure 3, has the elements separated with plastic spacers, and drooping ends to produce maximum spacing between the elements' ends.

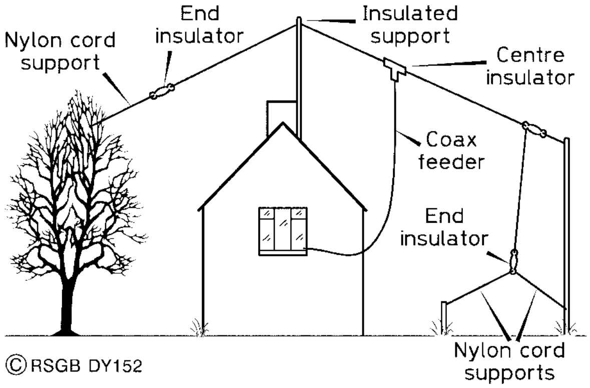
End insulator
Insulated support
Nylon cord support
Centre insulator
Coax feeder
End insulator
Nylon cord supports
©RSGB DY152

Figure 2 Possible layout for a dipole aerial in a confined space

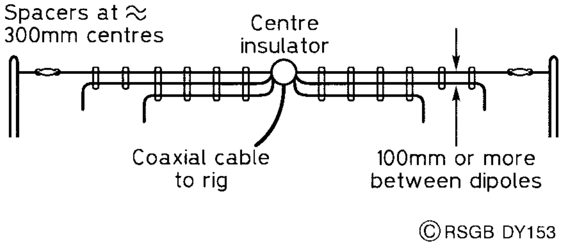
Spacers at ≈ 300mm centres
Centre insulator
Coaxial cable to rig
100mm or more between dipoles
©RSGB DY153

Figure 3 Multi-band dipole aerial

The long-wire aerial

This aerial is simple, cheap, easy to erect, and suits most houses and gardens, as **Figure 4** shows. Using an aerial tuning unit (ATU), an end-fed long wire can function on several bands when used with a set of radials or a counterpoise. **Figure 5** illustrates the setup. The length of the aerial will determine the bands which will be covered.

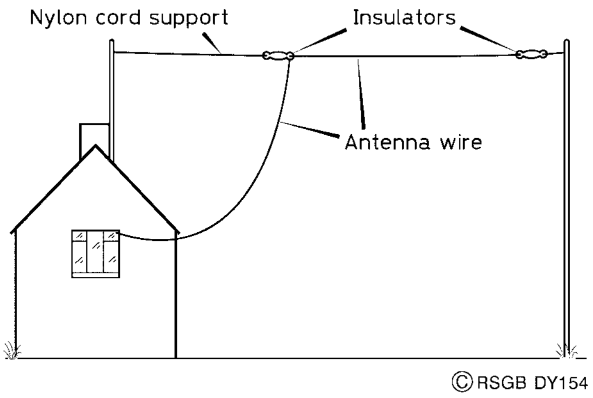
Nylon cord support
Insulators
Antenna wire
©RSGB DY154

Figure 4 Long-wire or inverted-L aerial

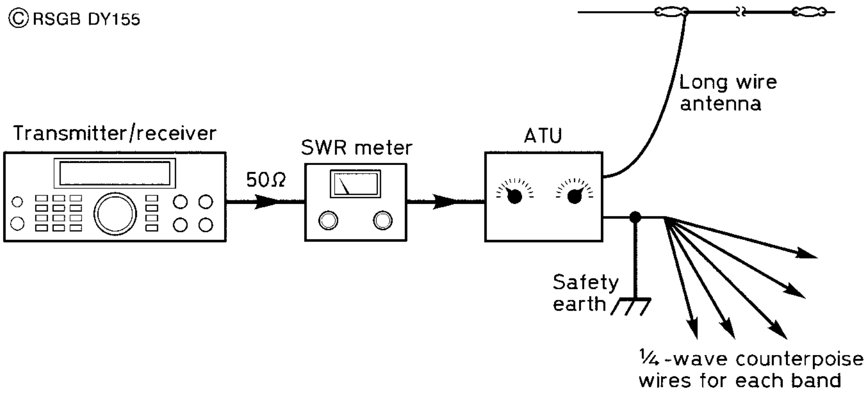
©RSGB DY155
Long wire antenna
Transmitter/receiver
SWR meter
ATU
50Ω
Safety earth
¼-wave counterpoise wires for each band

Figure 5 How to connect a radio to a long-wire aerial

A wire length of 10.5 m will work on the 40, 30, 17, 15 and 12 m bands.

A wire length of 15.5 m will work (with an ATU) on the 80, 40, 20 and 12 m bands and possibly (depending on your ATU) on the 17 and 15 m bands.

A wire length of 26.5 m will operate on all bands, but may be difficult to load on 10 m.

The wire lengths given here may need some adjustment because of the geometry of your particular house and garden. For receive-only purposes, the lengths are far less critical.

In general, you cannot get a good radio-frequency (RF) earth from a first-floor (or higher) shack. Unless a good RF earth exists within a small fraction of a wavelength of the transceiver, an *artificial ground* comprising a single $\frac{\lambda}{4}$ radial or counterpoise will be needed. You will need one counterpoise for each band you intend to use, and the wire can be concealed around the skirting-board of the shack, or under the carpet. Make sure that the free end

Table 2 Lengths of elements for vertical antennas, radials for verticals and counter-poises for end-fed long wire antennas

Band (MHz)	*Element length (m)*
1.8	39.66
3.5	20.40
7	10.20
10	7.14
14	5.1
18	3.96
21	3.4
24	2.95
28	2.55

of each counterpoise is well insulated; this point can carry a *very high voltage* when you transmit; anyone coming into contact with this can suffer very severe RF burns. Counterpoise lengths can be read from **Table 2**.

The vertical aerial

The single-band vertical aerial is sometimes used by DX operators because it has a low angle of radiation, which favours long-distance propagation. However, it must be sited clear of obstructions and must have a *good* counterpoise or radial system. Illustrations of the vertical aerial are shown, and the lengths of the vertical and radial sections are given in Table 2. The centre of the coaxial feeder is connected to the vertical section, and the braid to the counterpoise or radial system, which is made up of four or more wires buried just below the surface and joined together near the base of the aerial.

Cable entry to the house

Bringing coaxial cable into the house by an open window must be regarded as a temporary measure. Wooden window frames can be drilled, one hole for each feeder. Make the holes slope downwards from inside to out to prevent rain entering, and treat these with wood preservative. Leads from long-wire and inverted-L aerials should be kept separate from other cables.

Alternatively, a plastic pipe large enough to take all your feeders could be fitted into the brickwork (again, sloping downwards towards the outside). You may want to let a friendly builder do this for you.

53 A breadboard 80 m CW transmitter

Introduction

In the early days of radio, many circuits were built on a wooden baseboard, the parts being screwed down on the board. This was called breadboard construction, because it was a breadboard that was frequently commandeered for the process! Wives have always been generous in this respect, it appears!

This circuit was originally designed by GM3OXX, and became known as the *Oner*, because it was built on a circuit board one inch square! The circuit appeared in the G QRP Club journal *Sprat* and, since that time, many hundreds of *Oner* circuits have been built and used on the air. It is a well-proven circuit.

The transmitter has no tuned circuits in the power amplifier (PA) and thus has a rather high *harmonic* content. It **must** be used with the ***low-pass filter*** described elsewhere in this book. Without the low-pass filter, interference will be caused to other stations.

Simple aerial changeover switching is provided, which allows this circuit to be used with any of the 80 m receivers, such as the *Colt*, described in this book. It can also be used with any kit or commercial receiver for the 80 metre band.

The circuit

The transmitter circuit is shown in **Figure 1**. TR1 is a crystal oscillator, the frequency of which is controlled by crystal X1. A small trimmer capacitor, TC1, is added to allow the frequency of X1 to be varied by a small amount. If adjustment of this trimmer is made possible from the front panel, it is useful to adjust the transmit frequency to avoid other stations already on the crystal frequency. The collector load resistor, R2, of the oscillator transistor, TR1, determines the power output; a value of 3.3 kΩ seems to work well in producing an output of 3 watts.

TR1 is directly coupled to TR3, a VMOS transistor (a type of field-effect transistor (FET)). This acts as the power amplifier (PA) stage. TR3 should

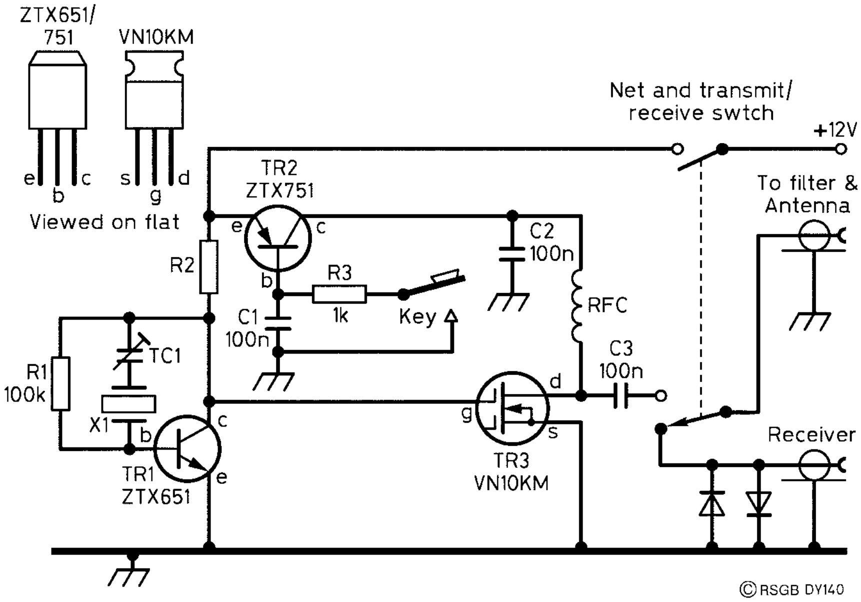
ZTX651/
751
VN10KM
e
b
c
s
g
d
Viewed on flat
Net and transmit/
receive swtch
+12V
TR2
ZTX751
e
c
b
To filter &
Antenna
C2
100n
R2
R3
1k
Key
RFC
C1
100n
C3
100n
R1
100k
TC1
X1
b
c
e
TR1
ZTX651
d
g
s
TR3
VN10KM
Receiver
©RSGB DY140

Figure 1 Circuit diagram of the breadboard transmitter

give about 3 W output, which is then coupled to the output by C3. The radio-frequency choke (RFC) providing the drain load of TR3 is simply a few turns of wire on a ferrite bead.

TR2 is an interesting addition to the circuit. It is used as a switch to ‘key’ the PA, TR3. The transmitter *could* be built without TR2, just placing the Morse key between the top of the RFC and the 12 volt supply. Adding TR2 is helpful, because it means that one side of the Morse key can be grounded (always a good thing), and some degree of *shaping* of the output RF waveform is provided by R3 and C1. This makes the transmission sound a little better and reduces the possibility of spurious frequencies being generated and transmitted. TR2 is a pnp transistor; note that it is the *emitter* of this transistor which is connected to the positive side of the supply.

Some form of changeover switching is needed for the aerial. A double-pole changeover toggle switch can be used. See the chapter on switches, later on in this book. One pole is used to switch the aerial between transmitter and receiver; the other pole is connected in the 12 volt supply line, and is labelled RECEIVE/TRANSMIT – NETTING. Its use will be described later. In this simple circuit, the PA cannot work when the key is open, because the key breaks its supply (via the RFC). When the key closes, TR2 switches on and applies the 12 volt supply to the top of the RFC. C2 is a *decoupling* capacitor, which prevents any residual RF signals at the top of the RFC reaching TR2.

Building

The prototype was put together as follows. Take a piece of plain printed-circuit board (PCB) measuring 5 cm by 4 cm. Then, with a *new*, *sharp blade* in a junior hacksaw, draw the blade horizontally across the surface of the copper in order to make a pattern of 6 squares along the 5 cm side and 5 squares along the 4 cm side. No more pressure should be applied than is necessary to cut through the copper! All the parts will be soldered on these pads in a form of surface-mount construction. To do this, each active pad (i.e. one that is going to have a component soldered to it) needs to be *tinned*. This means coating the pad's surface with solder, and is carried out as follows. Place the hot tip of the soldering iron on to the pad, and hold it there for a second or so. Then, with the tip still in place, touch the end of your reel of solder *on the pad*, not the tip of the iron. The solder should flow evenly all over the pad, and you can remove the iron. The solder should solidify in a rounded, shiny blob! This provides a good surface for making soldered joints.

To join component leads to the pads, cut each lead about 1 cm long, and then bend the last 2 mm at right angles to the rest of the lead. As you did before, tin the 2 mm length of each lead. Place the tinned portion on to the pad, and place the tip of your iron *on the pad*, close to the lead. The solder on the pad and on the lead will melt and run together; remove the iron and *hold the component still* until the solder solidifies. When the joint has cooled, give the lead a gentle tug to make sure you have a good joint (a good *mechanical* joint is usually a good *electrical* joint, too!). Each transistor straddles three pads, so the centre lead will need to be shorter than the other two. Take care here to get the lead lengths right – if you do, you will be surprised how much more firmly the transistor is held than if you just botched the lead lengths by bending them to fit! Make sure the connections to the transistors are correct.

Winding the RFC is quite simple. Seven (or more) turns of thin (32 SWG) enamelled copper wire are threaded through a small ferrite bead. This requires care, because the bead is small and the wire is thin. Trim the ends to within about 1 cm of the bead, remove the enamel carefully with sandpaper and tin the bare ends, prior to soldering the choke to the board.

After completion of the wiring, check the circuit against Figure 1. Bread-boarding a circuit like this has its advantages, but it can have disadvantages, too. One of these disadvantages is that it can make circuit checking difficult. For a simple circuit like this, it is not too bad! Check that no solder has run between the pads. Plug in your crystal for the 80 m band, and connect up the 12 V power supply. Do not connect the Morse key yet and do not switch on the power.

Testing and operating

Clear your workbench of all metallic objects, slivers of copper, bits of wire, etc., switch on your power supply. With an external receiver, listen on and around your crystal's frequency for a signal. Remember that the oscillator runs all the time and, because you haven't yet connected the Morse key, your receiver is close enough to pick up the signal from the oscillator. This confirms that your oscillator is running. Switch off.

Connect the station aerial to the transmitter's aerial socket, and the receiver to the transmitter's receiver socket. Connect the Morse key, put the Receive/Transmit–Netting switch in the receive position and switch on. You should be able to hear stations in the normal way. Now put the switch in the Transmit–Netting position. Signals in the receiver should almost disappear, as the circuit has disconnected the receiver's aerial.

Tune the receiver until you can hear your own crystal oscillator signal. This is known as *netting*, tuning your receiver and transmitter to the same frequency. Pressing the key will now transmit your signals when the switch is in the Transmit–Netting position; switch back to the receive position to listen for stations answering your call. As soon as you are happy that your circuit is functioning properly, **you must build the low-pass filter circuit** before using the transmitter regularly.

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1	100 kilohms (kΩ)
R2	See text
R3	1 kilohm (kΩ)

Capacitors

C1, C2, C3	100 nanofarads (nF), or 0.1 microfarad (μF)
TC1	3–60 picofarads (pF) trimmer

Semiconductors

TR1	ZTX651
TR2	ZTX751
TR3	VN10KM

Additional items

RFC	7 turns of 32 SWG enamelled copper on a ferrite bead
Switch	Double-pole changeover (DPDT or DPCO) toggle
Crystal	For 80 m band
Crystal holder	HC25 type
Sockets	According to station fittings

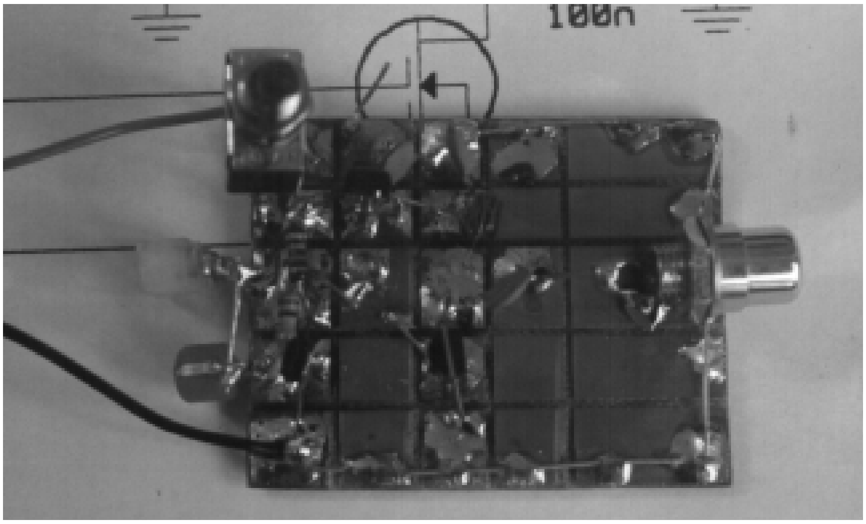

54 A 7-element low-pass filter for transmitters

Introduction

The simpler the transmitter, the more likely it is to radiate *harmonics* of its fundamental frequency. Harmonics are integral multiples of the frequency on which the transmitter is *designed* to operate. If you think you are transmitting on a frequency, *f*, for instance, you will also be radiating the harmonics of *2f*, *3f*, *4f*, . . . and so on. This results in your signals being heard on *several frequencies*, spread over a very wide frequency range. You are also contravening the terms of your licence. To avoid this, it is always advisable to use a *low-pass* filter between your transmitter and aerial. This is a filter which will *pass* your signal frequency, *f*, and all frequencies *below* it, but will *not pass* (attenuate) frequencies above *f* to any significant extent. The value of *f* is known as the *cutoff frequency* of the filter.

A design of 7-element low-pass filter

A 7-element low-pass filter (LPF) is so called because it has seven components, as the circuit diagram of **Figure 1** shows. Filters containing any odd number of elements are possible: a 3-element filter would

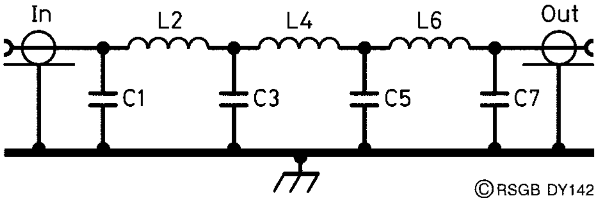

Figure 1 Circuit diagram of filter

comprise C1, L2 and C3 only, and is sometimes called a pi-network because the element disposition resembles the Greek letter pi (π); a 5-element filter would comprise C1, L2, C3, L4 and C5 only, and so on. In general, the more elements the filter has, the more effectively it attenuates signals above *f*.

The circuit of Figure 1 is designed to have an input and an output impedance of 50 Ω, which means that it can be placed in the aerial feed of any common transmitter. Filter design is a very complex business, and is best left to the experts. One such expert is W3NQN, who produced a number of computer designs of LPF using commonly available (*preferred value*) capacitors, and aimed specifically for use on amateur frequencies. The results of this work are condensed into **Table 1**. The inductors are wound on standard toroidal cores, and their details are included in the table.

Table 1 Filter component values for each brand

Band metres	*C1,7 pF*	*C3,5 pF*	*L2,6 turns*	*L4 turns*	*Core type*	*Wire SWG*
80	470	1200	25	27	T37-2	28
40	270	680	19	21	T37-2	26
30	270	560	19	20	T37-6	26
20	180	390	16	17	T37-6	24
15	82	220	12	14	T37-6	24
10	56	150	10	11	T37-6	22

Making the filter

The filter was made originally as an adjunct to the *Breadboard 80 m CW transmitter*, which you will also find in this book. It uses the same constructional technique, based on a single piece of plain, copper-clad

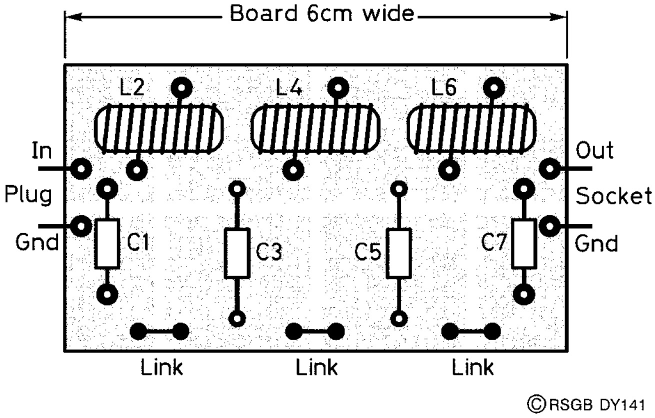

Figure 2 Layout and construction of filter

PCB, with 'pads' created by using a sharp blade in a junior hacksaw. The cuts in the copper are shown in the layout diagram of **Figure 2**. There are two ways of mounting the components: the first way is to drill small holes in each pad, as shown in Figure 2, and mount the components through the holes in the normal PCB manner; the second way is to solder the components directly to the pads, in the way that was described for the *Breadboard transmitter.*

Winding the inductors is quite simple. All you need to remember is that each time the wire passes through the core counts as one turn. Cut off the spare wire at each end of each coil to about 1 cm, scrape off the enamel with sandpaper, and tin the exposed copper. See the transmitter description if you are unsure of how to do this. Note the wire links between each of the lower pads, forming a solid 'ground' for the elements. The prototype had a plug and socket on the ends, to match the transmitter and aerial terminations.

The type of capacitor used in the design is not critical; the polystyrene type works well.

55 Radio-frequency mixing explained

Introduction

Mixers find widespread use in electronic circuitry. Many of the projects in this book, together with every TV set and radio in the home, contain mixer circuits – a good indication of their usefulness.

Confused?

Audio mixers (as used in recording studios and radio broadcast stations) are used to **add** or ‘balance’ the signals from various sources such as microphones, CD players, etc. These have nothing whatsoever to do with radio-frequency (RF) mixers, and should never be confused with them.

RF mixers and beat frequencies

Instead of **adding** signals (as in the audio mixer), the RF mixer **multiples them together.** As you might expect, this has an entirely different effect. The two signals entering the mixer *beat* or *heterodyne* with each other to produce *signals on other frequencies*. One example of this occurs in sound, when two musical notes of almost the same frequency are heard together. Instead of hearing two separate notes, the listener hears one note whose *intensity* (*loudness*) appears to increase and decrease. This intensity variation is called a *beat*, and its frequency is equal to the difference in the frequencies of the two original notes. The technique is used by musicians to tune their instruments. If one note is known to be a correct frequency, the other can be tuned to it by making the beat frequency as close to zero as is possible.

Multiplying together

The process of mixing presupposes that we have a device which will automatically multiply two signals together. Fortunately, this is easy; so easy, in fact, that it often occurs when we do not want it! Multiplying is achieved by any device which is *non-linear*; this means a device whose output is not a constant factor larger than its input, something that can be achieved by many electronic devices and circuits.

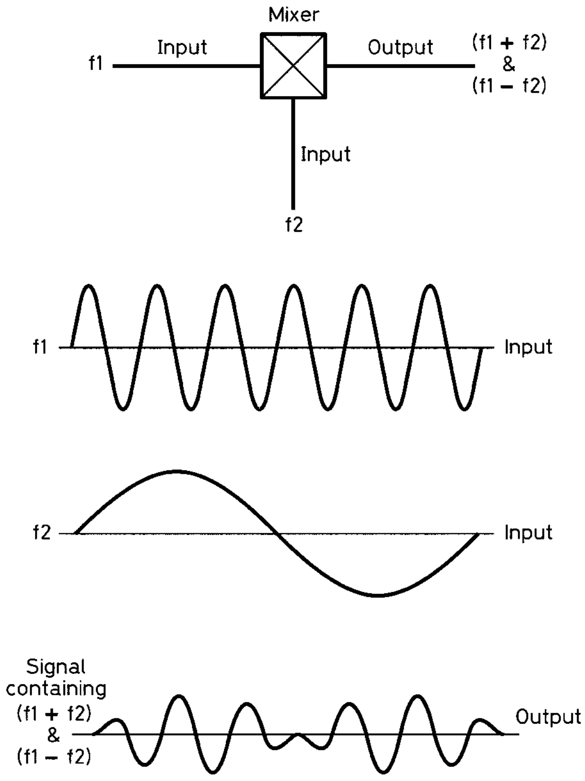

©RSGB DY145

Figure 1 The effect of multiplying (or mixing) two signals together

Let us look now at what a mixer does in concrete terms. Suppose two signals, of frequencies *f*1 and *f*2 go into our mixer. These signals are shown in **Figure 1**. Putting numbers in, to make the situation clearer, suppose *f*1 is 1.000 MHz and *f*2 is 160 kHz. The beat frequency is the *difference* of these: 1.000 MHz – 0.160 MHz = 0.840 MHz, or 840 kHz. A mixer *also* produces an output at the *sum* of these frequencies; in this case the new frequency would be 1.000 MHz + 0.160 MHz = 1.160 MHz.

Suppose you fed the output of your mixer, operating with these input frequencies, into a receiver and tuned around to find what frequencies were present. You would find two signals, one at 840 kHz and one at 1.160 MHz, showing that the two 'new' frequencies were very real!

In addition to drawing out the waveform of the resultant signal, as in Figure 1, we can draw the inputs and outputs on a frequency axis, to form a *spectrum* of the signal components. This is done in **Figure 2**. The top two diagrams show the input signals at *f*1 and *f*2. The bottom diagram shows the output signals in relation to the input signals. Depending on the type of mixer used, one or both of the input signals would be removed.

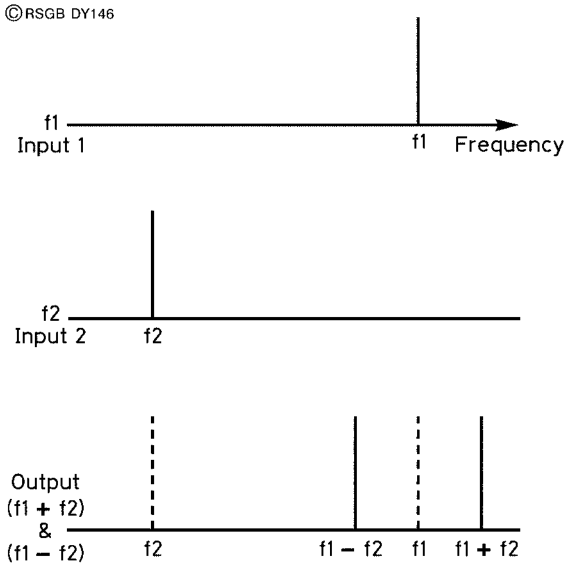
©RSGB DY146
f1
Input 1
f1
Frequency
f2
Input 2
f2
Output
(f1 + f2)
&
(f1 – f2)
f2
f1 – f2
f1
f1 + f2

Figure 2 The result of multiplying or mixing two signals together as seen on a spectrum analyser

A mixer in every radio

Basically, a mixer is used to change a signal from one frequency to another, something it does without altering the characteristics of the incoming signal. If the incoming signal is amplitude modulated (AM), then the frequency-changed signal would be AM also. The same applies to FM, SSB, CW, and all other modulation forms you can think of. This explains why mixers are often called *frequency changers*.

Frequency changing is the key process in the type of radio known as a *superheterodyne* (or *superhet*). By mixing the incoming signal with a variable-frequency *local oscillator* as **Figure 3**, shows, the signal can be converted to the fixed frequency of a filter and amplifier. This is useful

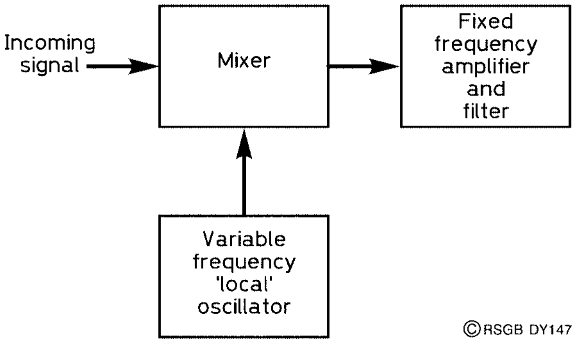
Incoming signal
Mixer
Fixed frequency amplifier and filter
Variable frequency 'local' oscillator
©RSGB DY147

Figure 3 The basic idea of a superhet receiver

because it is easier to make a very high-quality filter and amplifier at a single fixed frequency, than at a variable frequency.

All TV receivers and virtually all radio receivers (and transmitters) use mixers. Both the *Yearling* and *Colt* receivers (see the relevant projects) are superhets, and use mixers.

56 A voltage monitor for a 12 V power supply

Introduction

If for any reason, the stabilisation of your main 12 volt power supply unit (PSU) breaks down, it is possible that a voltage much higher then the nominal 13.8 V will be applied to your precious equipment. If you would like to know the instant that this occurred, and hence be able to switch things off before it was too late, then this circuit is what you need. It will give audible *and* visual indications if the voltage rises above 14.4 V, and a visual indication only if the voltage is reduced.

The circuit

The circuit uses three ICs and is shown in **Figure 1**. The circuit is powered by the PSU whose output is being monitored, and the circuit's immunity to supply line variations is secured by the 6 volt regulator, IC1. The heart of the circuit is IC2, an LM3914; it is a *bargraph driver*, which operates ten LEDs in a display resembling a thermometer – the string of lit LEDs increases in length as the voltage on pin 5 increases.

The input voltage range on pin 5 is 1.2 V maximum, making each LED correspond to one-tenth of this, which is 0.12 V, the *step size*. R1 and R2 act as a voltage divider, so that voltages of up to the maximum of 14.4 V may be applied to R1 without exceeding 1.2 V at pin 5. R3 sets the brightness of the LEDs and R4 determines the step size.

IC3 is an *opto-isolator*, a device containing an LED and a phototransistor in one package. This enables the piezoelectric sounder to operate without affecting the operation of the bargraph driver. The input to IC3 is provided by the voltage on D8, so that if any of the LEDs at or above D8 are lit, the

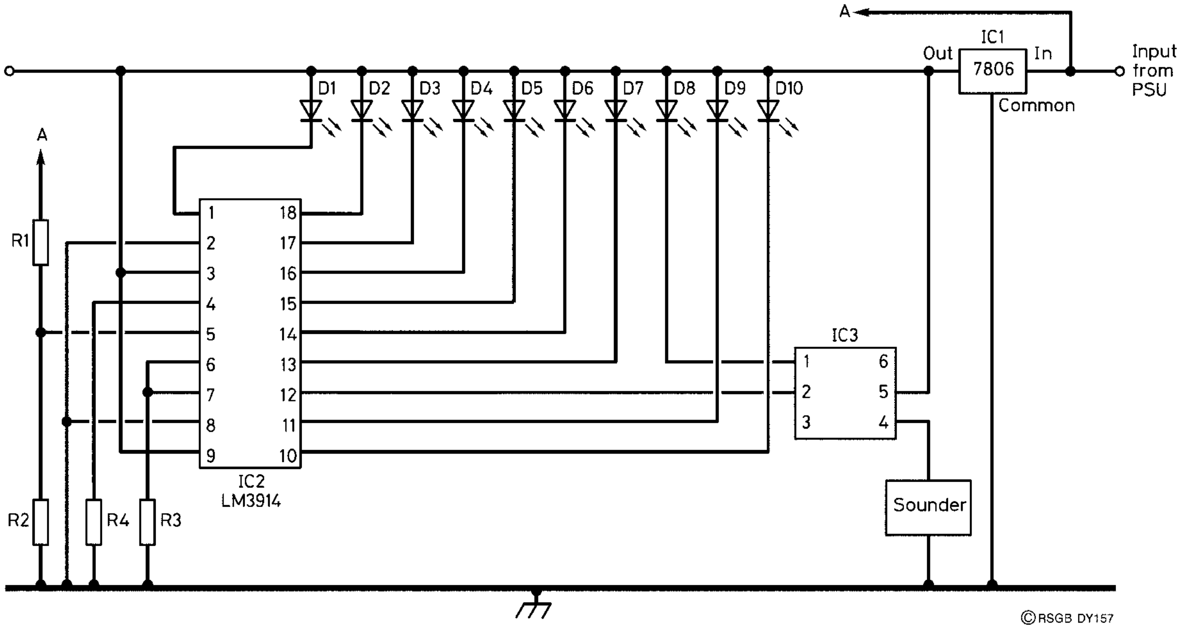
A
IC1
Out
7806
In
Input
from
PSU
Common
D1
D2
D3
D4
D5
D6
D7
D8
D9
D10
A
R1
R2
R4
R3
IC2
LM3914
1
2
3
4
5
6
7
8
9
18
17
16
15
14
13
12
11
10
IC3
1
2
3
6
5
4
Sounder
©RSGB DY157

Figure 1 PSU monitor, circuit diagram

sounder will operate, indicating overvoltage. Despite the fact that some of your other equipment might be damaged by this overvoltage, the monitor circuit itself is unaffected.

Construction

The prototype was built on two pieces of Veroboard of the copper-strip variety. The main (circuit) board measured 15 strips by 25 holes, and the display (LED) board measured 4 strips by 30 holes. Track cuts are necessary in each board. The correct places are shown for the main board in **Figure 2**, but there is sufficient flexibility in the layout of the display board for a prescriptive layout not to be needed. All the anodes of the LEDs are connected to the same strip, which makes things comparatively simple!

For the main board, insert the Veropins and the wire links first, and solder them to the copper strips. Then fit the IC holders and the resistors. Fit the IC holders with their notches towards the top of the board. When fitting the voltage regulator, IC1, note than the centre lead (the 'common' lead in Figure 1) does *not* go to position A3; the track should be cut at A3, and the common lead soldered at B3. This is indicated in Figure 2.

Wire up the LED board with its ten LEDs and 11 connecting wires, each about 10 cm long. This length depends on how far away from the main

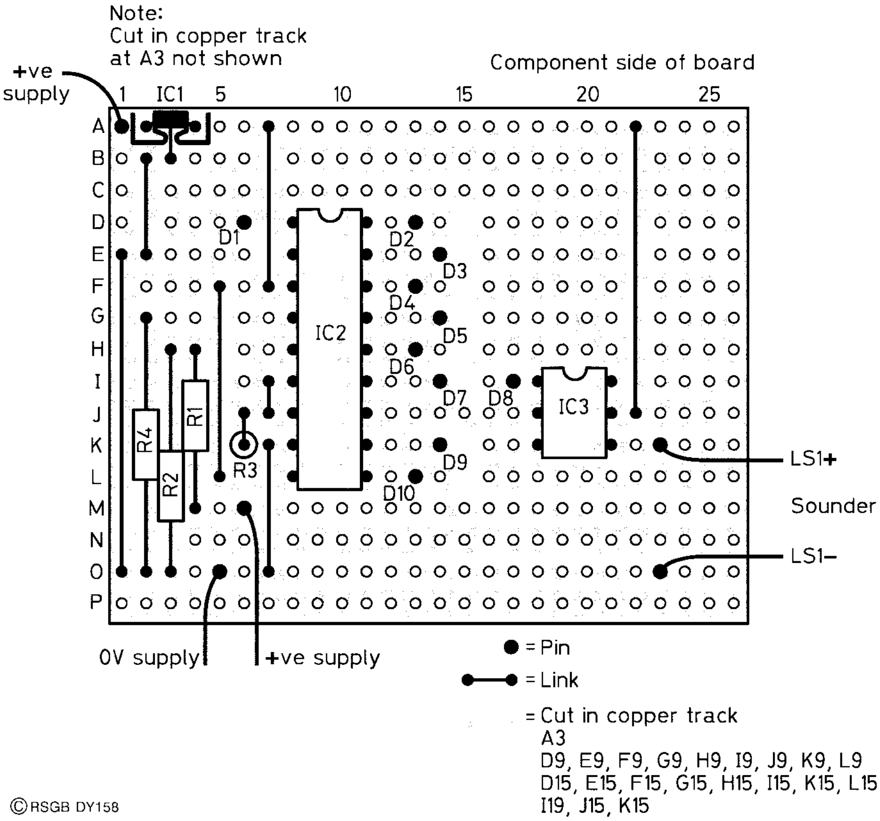
Note:
Cut in copper track
at A3 not shown
Component side of board
+ve supply
1 IC1 5 10 15 20 25
A B C D E F G H I J K L M N O P
D1 D2 D3 D4 D5 D6 D7 D8 D9 D10
IC2
IC3
R1 R2 R3 R4
LS1+
Sounder
LS1–
0V supply
+ve supply
● = Pin
●—● = Link
= Cut in copper track
A3
D9, E9, F9, G9, H9, I9, J9, K9, L9
D15, E15, F15, G15, H15, I15, K15, L15
I19, J15, K15
©RSGB DY158

Figure 2 PSU monitor, PCB layout

board you are planning to mount the display. The LEDs should have different colours: three orange, four green and three red, to indicate 'low', 'medium' and 'high' voltage.

Now solder the other ends of these leads to the Veropins on the main board, making sure that the order is correct. Finally, connect the piezoelectric sounder to the main board; the polarity is important, so make sure the red lead goes to pin 4 of IC3 and the black lead to the ground rail. If you were careful to fit the IC holders with their notches in the correct positions, match these up with the notches on the ICs before pushing home the ICs *gently*. Check your circuit for solder bridges and unwanted pieces of copper swarf before screwing the small heat sink to the voltage regulator, IC1.

Testing

(a) **With a variable-voltage PSU**. Set the PSU for 10 V, connect the circuit and switch on. Increase the voltage slowly, and check that each LED lights up after the one before it. As the voltage exceeds about 14.4 V, the first red LED should light and the sounder should operate. If an LED does not illuminate, you should immediately suspect either a dry joint or an incorrect LED polarity. The voltage at which the first red LED lights can be adjusted by varying R1; increase R1 if the LED comes on too early; decrease R1 if the LED comes on too late.

(b) **Without a variable-voltage PSU**. For your ‘variable supply’, you can use several AA-type 1.5 V cells (or 1.2 V NiCad cells) in series. The voltages produced by a range of cells is shown in **Table 1** – because of the lower voltage of NiCad cells, more of them are needed to produce a given voltage.

Once the operation of the circuit has been checked, it can be fitted into a plastic or metal box. Only two connections are needed for the PSU. You may want to drill some holes in the case to increase the apparent loudness of the sounder.

Table 1 Test voltages available from batteries in series

No. of batteries	*Voltage, nicads*	*Voltage, dry cells*
7	8.4	10.5
8	9.6	12
9	10.8	13.5
10	12	15
11	13.2	–
12	14.4	–
13	15.6	–

Parts list

Resistors: all 0.25 watt, 5% tolerance

R1	11 kilohms (kΩ) – see text
R2	1 kilohm (kΩ)
R3	1.2 kilohms (kΩ)
R4	18 kilohms (kΩ)

Semiconductors

IC1	L7806
IC2	LM3914
IC3	Opto-isolator – Maplin code WL35Q
D1–D3	3 mm LED, orange
D4–D7	3 mm LED, green
D8–D10	3 mm LED, red

Additional items

LS1	Piezoelectric sounder, wire leads
	6-pin DIL socket for IC3
	18-pin DIL socket for IC2
	Veroboard – two pieces for main and display boards, see text for sizes
	Veropins
	Heat sink for IC1
	Single-core insulated wire for links
	Insulated stranded wire for interconnecting the boards
	Case as required

57 A 1750 Hz toneburst for repeater access

Introduction

Repeaters across the UK and much of Europe need an access tone to switch the transmitter from standby ready for use. Commonly, this is a 1750 Hz tone of duration no greater than half a second. Although many UK repeaters may now be accessed using the *continuous tone-coded squelch system* (CTCSS – see the *RSGB Yearbook*), you may wish to access a repeater whose CTCSS frequency you don't know; in this case, using the universal 1750 Hz tone will gain you access. Commercial transceivers are usually fitted with an automatic toneburst, but if you are using a home-made design, then you may want to incorporate this little circuit.

Warning

This circuit uses a member of the integrated circuit family known as CMOS (complementary metal-oxide semiconductor). These use very little current

and can be completed destroyed if they come into contact with the magnitudes of static electricity that most of us carry about when we walk on carpets and wear rubber shoes. You will never know if this wanton destruction has happened – all you will discover is that your circuit doesn't work and that you have tested *everything*. To avoid this problem do the following things:

1. **Before you open the little packet in which the IC is supplied, touch something which you *know* to be earthed – the metalwork of any equipment which is mains earthed, for example. Then open the packet.**
2. **Let the IC fall gently on the bench – don't pick it out with your fingers. Touch your earthed metalwork again. Pick up the IC and insert it gently into its holder.**

The circuit is safe from destruction while it is connected to the battery.

Circuit description

The circuit is shown in **Figure 1**. The tone is generated by an integrated circuit oscillator (IC1), whose frequency is controlled by a *ceramic resonator*, XL1. Its frequency is very high, and is divided down to the 1750 Hz needed by the same chip. The ceramic resonator is designed to operate at 455 kHz, the intermediate frequency of many receivers. Because all divider circuits use powers of 2, we need the oscillator to run at 448 kHz so that when it is divided by 256 ($256 = 2^8$), we end up with 1750 Hz. Try it on your calculator:

$$\frac{448\ 000}{256} = 1750.$$

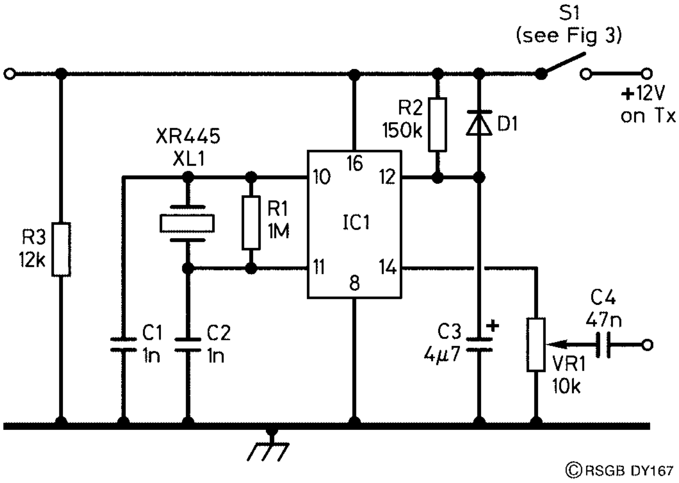

Figure 1 Toneburst module, circuit diagram

We use C1, C2 and R1 to *pull* the frequency of the oscillator away from 455 kHz to 448 kHz. The divider chain has eight *counters* in it, and each counter divides the frequency of the signal it sees by two, giving the final division of 256.

This counting process can be stopped at any time by taking the voltage on the *reset* pin (pin 12) up to the supply voltage. When the circuit is switched on by closing S1, pin 12 is at 0 V because C3 is discharged. The oscillator runs, producing the output frequency of 1750 Hz. As time progresses, C3 charges up through R2 and the voltage on pin 12 rises. When this has risen sufficiently, and in a time determined by the values of C3 and R2, the counter resets and stays in the reset state; no division takes place and there is no output. The duration of the toneburst is thus governed by C3 and R2.

When S1 is opened, the circuit is switched off, and C3 is discharged through D1 and R3, ready for the next toneburst. If you have used a repeater, you will know that a toneburst is needed **only** to activate a repeater in the standby condition; it is not needed once a contact has been established.

VR1 adjusts the amplitude of the tone fed to the microphone, and C4 prevents any voltage that may be present on your microphone connector from damaging the integrated circuit.

Construction

The prototype circuit was built on Veroboard of the copper-strip variety, measuring 20 holes by 14 strips. The layout is shown in **Figure 2**. Make the track cuts first, and check that there are no slivers of copper wedged between adjacent tracks. Then, solder in the IC socket (with the notched end

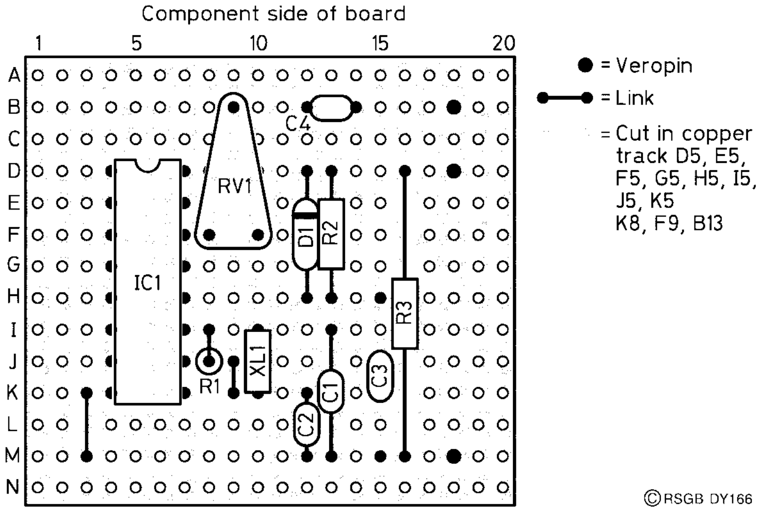

Figure 2 Veroboard component layout

facing towards track A), the wire links and the three Veropins. Having done this, solder in the resistors, capacitors and diode, making sure that D1 and C3 are the right way round! Using your best soldering technique, solder in the ceramic resonator quickly, to prevent heat damage. Recheck your circuit, check for solder splashes and bridges, and then gently insert IC1 into its socket, matching up its notch with that of the socket.

Testing

Set VR1 to half-way and connect a crystal earpiece to the output; apply power to the circuit. You should hear the tone, lasting for about half a second. If there is no tone, disconnect your circuit from the power supply, and check for dry joints in the vicinity of pin 12. Is the diode, D1, the correct way round? Is C3 the correct way round? Did you choose to ignore the CMOS safety precautions given earlier?

Once the circuit is working, you need to decide how you are going to connect it to your transmitter. Two options are shown in **Figure 3**. If you have access to a point in your transmitter circuit that has between 9 V and 12 V positive on it during transmit, you can use this to power your circuit. As the toneburst is needed only for repeaters, the switch, S1, disconnects it when not needed, as shown in Figure 3a. If you want the circuit to be self-powered, then a 9 volt PP3 battery may be used; Figure 3b shows this configuration.

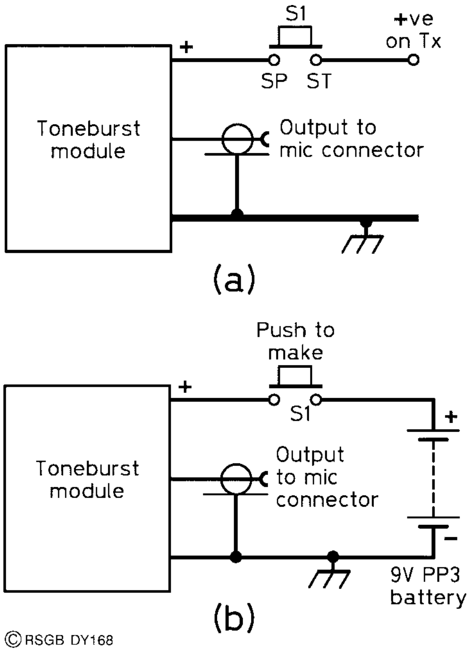

©RSGB DY168

Figure 3 Toneburst module: two alternative switching arrangements

The output from the circuit board is fed directly into the microphone socket, in parallel with the microphone itself; use thin coaxial or screened cable for this lead, or you may induce hum into the microphone circuit and suffer from RF breakthrough into the audio circuits. To adjust the setting of VR1, start with it at the zero output position and connect a dummy load to your transmitter. Slowly, increase the output while monitoring your transmitted signal on another nearby receiver. Make sure you do **not** increase the output so far that the signal sounds distorted. If you would prefer that the tone was on continuously while you made this adjustment, simply connect a wire across C3 remembering, of course, to remove it as soon as you have completed the test!

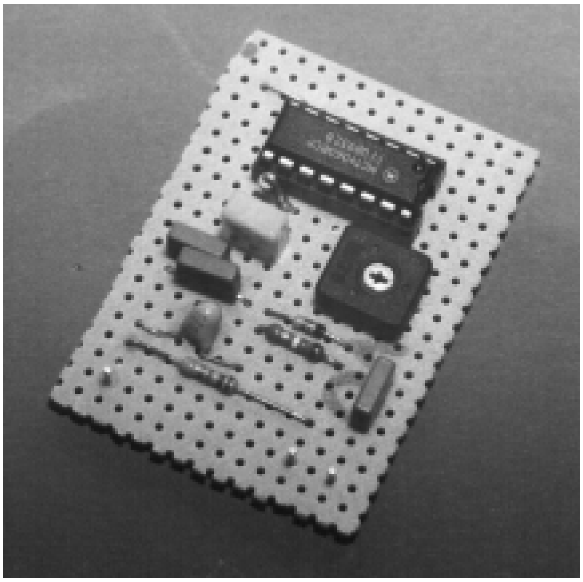

Parts list

Resistors: all 0.25 watt, 10% tolerance (or better)

R1	1 megohm (MΩ)
R2	150 kilohms (kΩ)
R3	12 kilohms (kΩ)
VR1	10 kilohms (kΩ) horizontal preset

Capacitors: all 16 V WKG or higher

C1, C2	1 nanofarad (nF) disc ceramic
C3	4.7 microfarad (μF) tantalum bead
C4	47 nanofarads (nF) disc ceramic

Semiconductors	
IC1	4060
D1	1N4148
Additional items	
XL1	XR455
	Veroboard (see text for size)
	Veropins (3)
S1	Switch (momentary action push-to-make SPST)
	16-pin DIL socket for IC1
	Single-core insulated wire for links
	Coaxial or screened cable for microphone connection

58 A circuit for flashing LEDs

Introduction

There are many occasions when one's attention needs drawing to the fact that something important has happened. A single red light coming on is seldom sufficient to attract attention, particularly if it is surrounded by other lights and indicators. The eye is known to be very sensitive to *changes* in its peripheral vision; such changes can be brought about by movement or by differences in light level – a flashing light, for example. So, a circuit that flashes a single LED or a pair of LEDs finds plenty of uses in the amateur station.

Warning

This circuit uses a member of the integrated circuit family known as CMOS (complementary metal-oxide semiconductor). These use very little current and can be completed destroyed if they come into contact with the magnitudes of static electricity that most of us carry about when we walk on carpets and wear rubber shoes. You will never know if this wanton

destruction has happened – all you will discover is that your circuit doesn't work and that you have tested *everything*. To avoid this problem do the following things:

1. **Before you open the little packet in which the IC is supplied, touch something which you *know* to be earthed – the metalwork of any equipment which is mains earthed, for example. Then open the packet.**
2. **Let the IC fall gently on the bench – don't pick it out with your fingers. Touch your earthed metalwork again. Pick up the IC and insert it gently into its holder.**

The circuit is safe from destruction while it is connected to the battery. However, when the battery is removed, the same care should be exercised with its handling, because there is no supply decoupling capacitor across the IC.

Basic description

LEDs can be made to flash (switch on and off) by driving them from sources that switch on and off. Such a source is an *astable multivibrator*. If you have built or read about *A basic continuity tester*, elsewhere in this book, you will have come across such a beast before. That circuit used an astable multivibrator made from two transistors. This new circuit achieves the same behaviour from a single integrated circuit, the CMOS 4011. To give it its full description, the 4011 is a *quad 2-input NAND gate*. Quite a mouthful, but all it means is that inside the chip are four NAND gates, each with two inputs.

A NAND gate needs a positive voltage (known as a *logic 1*) on *both* inputs in order to produce zero volts (known as *logic 0*) at the output. Two NAND gates can be connected, as are A and B in **Figure 1**, to make our astable multivibrator. The combination of A and B has been described as the most perverse circuit in electronics; as soon as the output goes to logic 1, the circuit decides that it would prefer to have a logic 0 there, and switches over. With logic 0 at the output, the circuit now prefers to have logic 1 there, and so it goes on! We are going to use this continuous switching backwards and forwards to flash two LEDs. The rate at which A and B 'change their minds' is the frequency at which our LEDs will flash, and is controlled by the charging and discharging times of C1 through R2 and C2 through R1. As the values of R1 and R2 are the same, and those of C1 and C2 are the same, the ON and OFF states of the circuit are the same.

Gates C and D do not contribute to the flashing action; they act as *buffers* to isolate the LEDs from the multivibrator circuit itself. You will find in electronics that an oscillator is seldom used to drive another device *directly*; there is usually a buffer between it and the stage it drives.

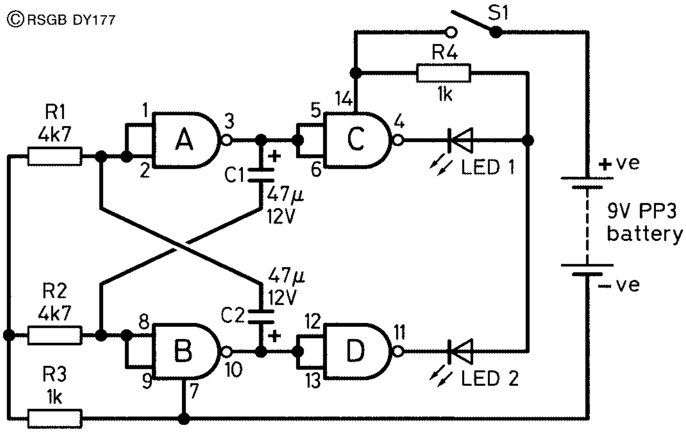

Figure 1 Flashing LEDs, circuit diagram

A characteristic of all multivibrators, astable or not, is that they have two outputs. In this case, those outputs are at pins 3 and 10, which are then buffered and appear at pins 4 and 11, respectively. When one output is at logic 1, the other is at logic 0, and vice versa. This means that LED1 is off when LED2 is on, and LED1 is on when LED2 is off, the two states switching backwards and forwards at the frequency of the oscillator.

Construction

Read the warning at the beginning of this article again. It is not intended to scare you off from building this, but is a genuine piece of advice which can save you time and irritation when all your labours result in a circuit that doesn't work! That extra bit of care can make all the difference!

Veroboard (the copper strip type) is used for the layout, shown in **Figure 2**. It measures 20 holes by 12 strips. **Be aware that there is no row 'I' in the layout, so don't miscount when you are placing components on the board!**

Firstly, cut the tracks using a 3 mm (⅛ inch) twist drill held between thumb and forefinger; check that there are no slivers of copper bridging any of the tracks, and that the tracks have been completely cut by the drill. Solder in the components carefully. Leave the IC in its carrier for the time being, and solder in the IC socket, with the notched end facing row A. On completion, check the circuit carefully. If you are happy that it is correct, follow the instructions given earlier and fit IC1 into its socket, matching up the two notches. Connect the battery and switch on. The two LEDs should flash on and off alternately. If only one LED flashes, you have probably connected the other one the wrong way round. Switch off, check and correct if necessary. If neither LED flashes, you must have a significant error in your circuit, which will need checking *again*! Or did you choose to ignore the handling precautions for the CMOS chip?

Figure 2 Flashing LEDs, component layout

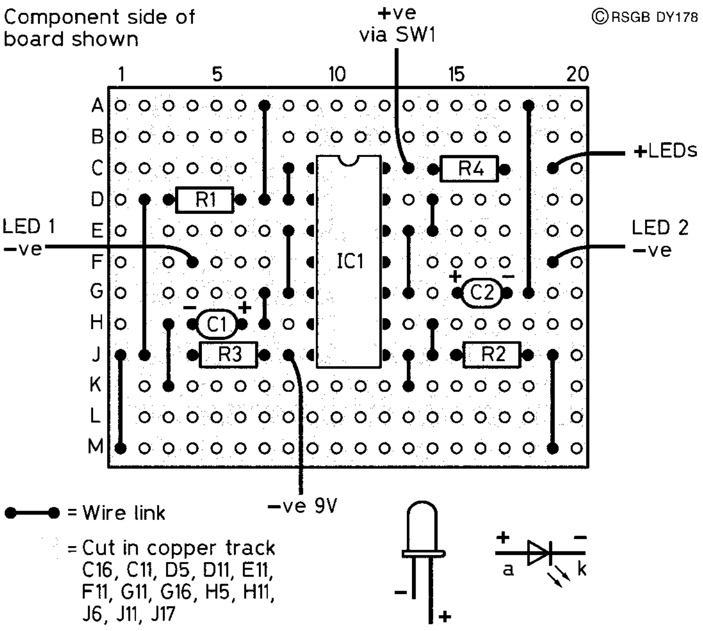
Component side of board shown
+ve via SW1
©RSGB DY178
1
5
10
15
20
A
B
C
D
E
F
G
H
J
K
L
M
R1
R3
R4
R2
C1
C2
IC1
+LEDs
LED 1 −ve
LED 2 −ve
−ve 9V
= Wire link
= Cut in copper track C16, C11, D5, D11, E11, F11, G11, G16, H5, H11, J6, J11, J17
a
k

The board can be mounted near to the point where you want your flashing LEDs to be seen, although long leads to the LEDs are acceptable. The LEDs can be different colours – it's all up to you now!

Parts list

Resistors: all 0.25 watt, 10% tolerance or better	
R1, R2	4700 ohms (Ω)
R3, R4	1 kilohm (kΩ)
Capacitors	
C1, C2	47 microfarads (μF) 12 V WKG
Semiconductors	
IC1	4011
Additional items	
LED1, LED2	Any size of LED, any colour
S1	SPST on/off
	Plastic box if needed, 8.5 by 5 by 2.5 cm

Source

Components are available from Maplin.

59 Digital logic circuits

Introduction

Logic circuits form the backbone of even the most advanced computer, yet their basic operation can be demonstrated by a couple of switches, a battery and a bulb.

Logic using switches

Everyone reading this article will look at **Figure 1** and know immediately how it works and be able to write down something like 'When switch A and switch B are closed, the light will come on'. Without knowing it, you have written down a logic statement involving the so-called **AND** operation; the light comes on only when switches A **AND** B are ON. Below the circuit in Figure 1, is a table showing the only possible positions of the two switches and the state of the bulb for each position. This is called a *truth table*, and is frequently used in logic analysis.

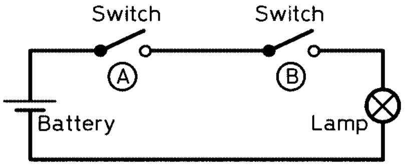

A	B	Lamp
Off	Off	Off
On	Off	Off
Off	On	Off
On	On	On

©RSGB DY182

Figure 1 Switches and lamp AND gate

Figure 2 shows a different circuit. Here, the two switches are in parallel rather than in series, as was the case in Figure 1. Again, if you analyse the circuit in words, you would say that the light will be on when switch A **OR** switch B is ON. This is an example of the **OR** operation, and its truth table is shown in Figure 2. The statement above is not complete, however; can you see why? The truth table will show you. The light comes on if A is ON, **OR** if B is ON, ***OR*** if A **AND** B are both ON. That third condition is easy to miss, but don't worry about it!

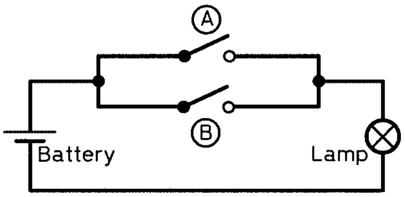

A	B	Lamp
Off	Off	Off
Off	On	On
On	Off	On
On	On	On

Figure 2 Switches and lamp OR gate

©RSGB DY184

Believe it or not, some very complicated logic is possible (in theory) using switches and lights, but it is highly impractical and would be very slow. This is where electronic logic circuits come in.

Switches with no moving parts

You may have come across projects in this book where a statement is made such as '. . . the transistor is being used as a switch . . .'. Transistors *can* be used as switches, as were thermionic valves in the world's first programmable computer *Colossus*, at Bletchley Park. However, technology has moved on from valves, through transistors to *logic gates*, combinations of electronic switches designed specifically to perform logic functions.

These act on voltage levels as their inputs and produce changes in voltage levels as their outputs. A positive voltage is called *logic 1*, and corresponds to a switch being ON in our previous descriptions; a zero voltage is called *logic 0*, and corresponds to a switch being OFF. The output from a logic gate (normally labelled Q) is also logic 1 or logic 0, corresponding to our light being ON or OFF, respectively, in our switch analogy.

Many logic devices operate from a stabilised 5 V supply, and this determines the *ideal* voltages corresponding to the two logic states:

logic 0 ≡ 0 V,

logic 1 ≡ 5 V.

The world isn't an ideal place, so the *real* voltage ranges used by the logic gates are:

logic 0 ≡ 0.0 to 0.4 V,

logic 1 ≡ 3.0 to 5.0 V.

The **AND** circuit of Figure 1 is now called an **AND gate**, and requires logic 1 inputs on A **AND** B to produce a logic 1 at the output. **Figure 3** shows this. The truth table is identical with that of Figure 1 – logic 1 replaces ON and logic 0 replaces OFF. Now compare Figure 2 with **Figure 4** – circuits and truth tables for the **OR** function. Again, we have exact similarity.

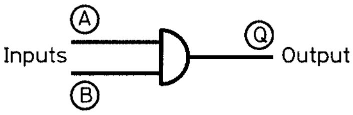

Inputs		Output
A	B	Q
0	0	0
0	1	0
1	0	0
1	1	1

©RSGB DY183

Figure 3 Electronic AND gate

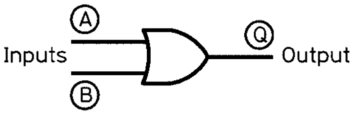

Inputs		Output
A	B	Q
0	0	0
0	1	1
1	0	1
1	1	1

©RSGB DY190

Figure 4 Electronic OR gate

There is another very common logic gate, which performs the **NOT** function. It is easy to understand. Just ask yourself the question 'What is **NOT** logic 0?', and the answer is obviously 'logic 1'. Similarly, logic 0 is **NOT** logic 1. A **NOT** gate simply changes the logic state of the input; it is also known (because of this behaviour) as an *inverter*. Its symbol and truth table can be found in **Fig 5**. Note the little circle on the output of the gate in Figure 5. In logic circuits, this symbol always implies inversion, or the presence of a **NOT** gate. Keep an eye open for it!

So far, the logic functions we have discussed have all been words which we use in everyday language, which has made the electronic interpretation of them relatively easy. Now we must introduce a function for which there is *no* analogy in normal speech – the **NAND** function. This means a

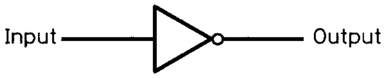

Input	Output
1	0
0	1

©RSGB DY185

Figure 5 Electronic NOT gate

combination of an **AND** gate and a **NOT** gate, and the sharp-eyed reader will have spotted the little circle added on to the normal **AND** gate symbol in **Figure 6**!

To make things easier to understand, the truth table in Figure 6 has four columns, not three as in previous tables. The third column is the standard **AND** output – compare it with the third column in Figure 3. That is the output from the **AND** gate *before* it encounters the little circle that inverts it, so the final output from the **NAND** gate is the output of the **AND** gate, inverted! The third and fourth columns are the inverse of each other.

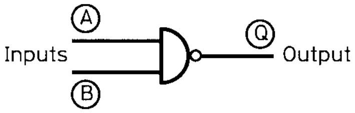

Inputs		and	Output
A	B		Q
0	0	0	1
0	1	0	1
1	0	0	1
1	1	1	0

©RSGB DY186

Figure 6 Electronic NAND gate

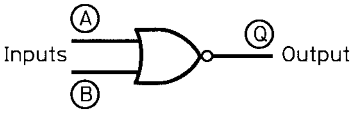

Inputs		or	Output
A	B		Q
0	0	0	1
0	1	1	0
1	0	1	0
1	1	1	0

©RSGB DY187

Figure 7 Electronic NOR gate

We can add the inverting operation to the output of an **OR** gate also, producing a **NOR** gate! The symbol and its truth table are shown in **Figure 7**. Again, notice that the final output is that of the ordinary **OR** gate, inverted!

These new functions of **NAND** and **NOR** are used more than the **AND** and **OR** functions because it makes other circuits easier to design using combinations of these gates.

A taste of Boolean algebra

The design of circuits using combinations of logic gates usually begins with a little mathematics, where the functions to be implemented are analysed. The mathematics used is surprisingly simple, and is a slightly changed version of ordinary algebra called *Boolean algebra*, which allows manipulation of logic functions to be made. Normal algebra has operations in it such as addition, subtraction and multiplication and division. The mathematician Boole found that the logical **AND** operation could be handled by the algebraic operation of *multiplication* (symbols × or •), and the **OR** operation by the algebraic operation of *addition* (symbol +). The **NOT** operation involved a new symbol, that of a bar over the input being inverted, such as $\overline{A}$.

So, our five basic logic operators can now be written in a mathematical form:

AND	$Q = A \times B$
OR	$Q = A + B$
NOT	$Q = \overline{A}$
NAND	$Q = \overline{A \times B}$
NOR	$Q = \overline{A + B}$.

Using logic operations in a mathematical form enables the most complex logic to be designed, simplified and converted into circuit diagram form in a very efficient and rapid way. There are more logic operators than the five we have considered here but, in general, they can all be broken down into combinations of these gates alone!

60 A resistive SWR indicator

Introduction

When a transmitter produces some output power, we want to make sure that as much as possible of this power is radiated by the aerial. This often requires the use of an *aerial tuning unit* (ATU), which matches the aerial impedance to that of the transmitter.

How do you know when this matching has been achieved? The most usual way is to use a *standing-wave ratio* (SWR) indicator. If the impedance of the aerial does not match that of the transmitter output, some of your transmitter power (also known as the *forward* power) is reflected back along the aerial feeder and back into the transmitter, where it causes excess heating. The forward and reflected waves interact along the feeder to produce a wave whose position remains constant, and which is therefore called a *standing wave* or a *stationary wave*. An SWR meter simply indicates forward power and reflected power, and adjustments are made to your ATU until the reflected power is as small as possible (ideally zero, of course). If there is no reflected power then, by a process of elimination, all your forward power is reaching the aerial!

Sampling the RF

Whatever type of SWR indicator you use, it must use some sort of sampling circuit to pick up the forward and reflected waves. The project *A standing-wave indicator for HF*, elsewhere in this book, uses a toroidal transformer to separate the readings for the forward and reverse waves. This design differs in that it measures the voltages across resistors through which the RF current is passing. Its advantages are:

(a) it uses cheap parts – four resistors, two capacitors and a diode, together with a rotary switch, a surplus meter, a preset potentiometer and two sockets;

(b) within this SWR indicator, there is always a resistive path for the RF current from the transmitter, formed by R1, R2 and R3; this can prevent damage to simpler home-made transmitters, which may be damaged during adjustment of the ATU when using more conventional SWR indicators.

The only disadvantage of this form of indicator is that it must be switched out of circuit once the ATU has been adjusted for a particular band.

Construction

The SWR indicator is very simple to build, as most of the components can be mounted on the back of the 3-way rotary switch. This is shown in **Figure 1**. The switch is a 4-pole, 3-way rotary type, of which only two poles are used.

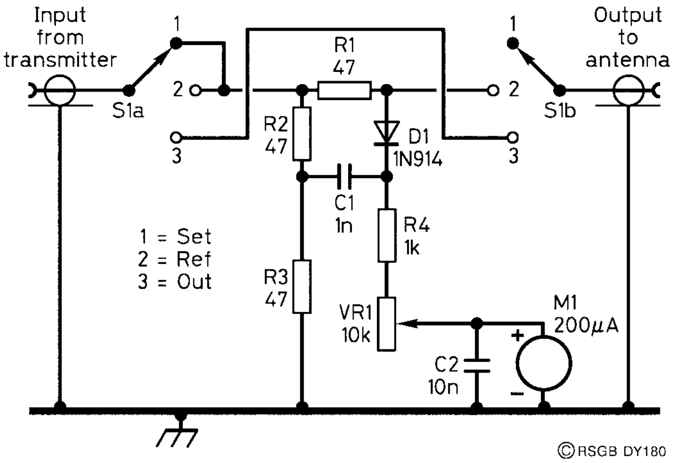

Figure 1 Resistive SWR meter, circuit diagram

Because the other switch contacts are not being used, they can be employed as support tags for other components. The ground wires are all soldered on to the metal frame of the switch. If your switch frame is of all-plastic construction, then a 12 SWG copper wire run around the switch will make a good earth connection to the metal case for the leads shown in Figure 1. The preset potentiometer used to control the sensitivity of the circuit can be mounted directly on the meter tag.

Resistors R1, R2 and R3 handle the RF power during the tuning-up process. If you have them, use 1 watt resistors; otherwise, you can use two 100 ohm half-watt resistors in parallel for each of R1, R2 and R3. The meter, M1, can be any DC type of sensitivity around 200 µA.

In use

First, find a clear frequency, and without the indicator in circuit, check that the frequency really *is* clear by asking and identifying yourself. If it is, connect the indicator between the transmitter and the ATU which, in turn, is connected to your aerial. Turn SW1 to the SET position and key the transmitter. Adjust VR1 until the meter reads full scale. Switch off the transmitter. Turn SW1 to the REF position and key the transmitter again. Adjust the ATU until the lowest reading is obtained on the meter. Switch off the transmitter. For the chosen frequency, you have adjusted your ATU for minimum reflected power and hence the lowest SWR. You will need to repeat the process when you change bands, and possibly when you change frequency within the same band. Switching SW1 to the OUT position, you are ready to transmit. You may have noticed that it is good practice to switch the transmitter off when operating SW1. Get into that habit!

Parts list

Resistors	
R1, R2, R3	47 ohm (Ω) 1 watt (or 2 × 100 Ω, ½ watt, see text)
R4	1 kilohm (kΩ) 0.25-watt
VR1	10 kilohms (kΩ) preset
Capacitors	
C1	1 nanofarad (nF)
C2	10 nanofarads (nF)
Semiconductors	
D1	1N914 or similar
Additional items	
SW1	4-pole 3-way rotary switch, of which only 2 poles are used
M1	200 μA DC
	Case

61 An audio filter for CW

Introduction

This is a simple passive circuit (it has no power supply) that adds some audio selectivity for Morse code reception and also includes a very simple noise limiter that gives a visual indication of when noise spikes are being removed!

The circuit

Figure 1 shows the complete circuit. The tuned circuit of C1 and L1 resonates very close to 800 Hz, so initially you will have to tune a signal in carefully until it sounds loudest – you will soon be able to do this without thinking. The two LEDs connected back to back across the signal path act as a noise limiter, reducing the amplitudes of static crashes and noise from car ignition systems, etc. The LEDs blink when they conduct – this is not necessary to the operation of the circuit, but adds a little colour to your listening! The noise limiter *does* make listening more comfortable, though.

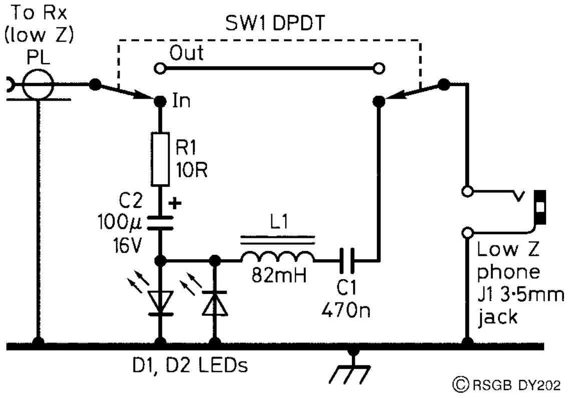

Figure 1 CW filter, circuit diagram

Construction

The circuit layout is shown in **Figure 2**. Point-to-point wiring is used, with a small tag-strip being the only item used for the extra support of C2 and R1. The LEDs and SW1 support the other components. An aluminium box (**Figure 3**) is used to make the circuit tidy and usable.

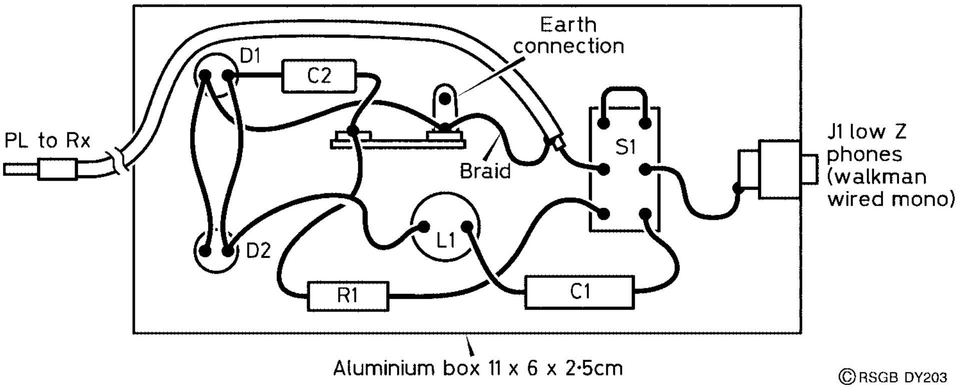
Earth
connection
D1
C2
PL to Rx
S1
Braid
J1 low Z
phones
(walkman
wired mono)
L1
D2
R1
C1
Aluminium box 11 x 6 x 2·5cm
©RSGB DY203

Figure 2 CW filter, component layout

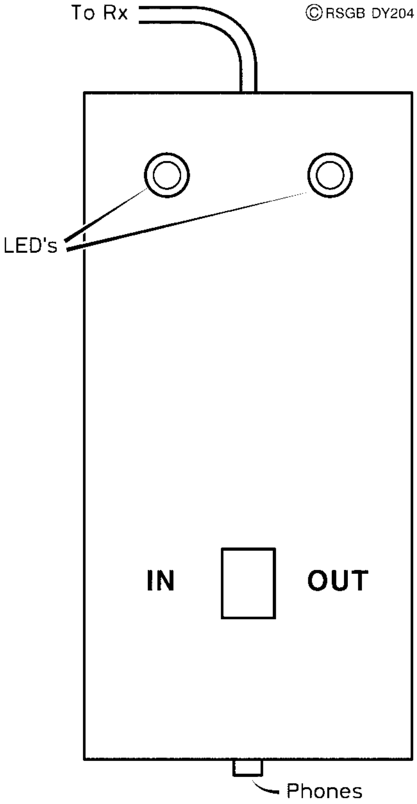
To Rx
©RSGB DY204
LED's
IN
OUT
Phones

Figure 3 CW filter, front panel

You must use low-impedance headphones for the circuit to perform properly. Plug the completed unit into the headphone socket of your receiver and adjust the volume so that the LEDs are just *not* blinking on normal audio. The unit should be switched out of circuit for speech listening; this is the purpose of the toggle switch, SW1.

Parts list

Resistor

R1	10 ohms (Ω), 0.25 watt, 10% tolerance

Capacitors

C1	470 nanofarads (nF)
C2	100 microfarads (μF) electrolytic, 16 V WKG

Inductor

L1	82 millihenries (mH)

Semiconductors

D1, D2	LEDs

Additional items

SW1	DPDT
J1	3.5 mm jack socket
PL	Jack plug to suit receiver
	Aluminium box, approx. 11 by 6 by 2.5 cm

Source

Components are available from Maplin.

62 An electronic die

Introduction

Throwing a die is a venerable way of generating a 'random' number between 1 and 6. A die is easy to lose, so here is an electronic die-throwing circuit which brings the technique up to date and serves as another application of logic circuits. If you need a reminder of the basics of logic, refer to *Digital logic circuits*.

Warning

This circuit uses members of the integrated circuit family known as CMOS (complementary metal-oxide semiconductor). These use very little current and can be completed destroyed if they come into contact with the

magnitudes of static electricity that most of us carry about when we walk on carpets and wear rubber shoes. You will never know if this wanton destruction has happened – all you will discover is that your circuit doesn't work and that you have tested *everything*. To avoid this problem do the following things:

1. **Before you open the little packet in which the IC is supplied, touch something which you *know* to be earthed – the metalwork of any equipment which is mains-earthed, for example. Then open the packet.**
2. **Let the IC fall gently on the bench – don't pick it out with your fingers. Touch your earthed metalwork again. Pick up the IC and insert it gently into its holder.**

The circuit is safe from damage while it is connected to the battery.

Description

The circuit is shown in **Figure 1**. IC1a and IC1b form an *astable multivibrator*, similar to that used in 'An LED Flasher' also in this book. It runs constantly, and its output is fed to IC2 via a single NAND gate, IC1c. The other input of IC1c is held at 0 V by R3 and C3. Whenever one input of a NAND gate is logic 0 (or 0 V), there can be *no output* from the gate, irrespective of what is happening at other inputs. So, despite the fact that the oscillator (or *clock*) is running all the time, its square-wave output never

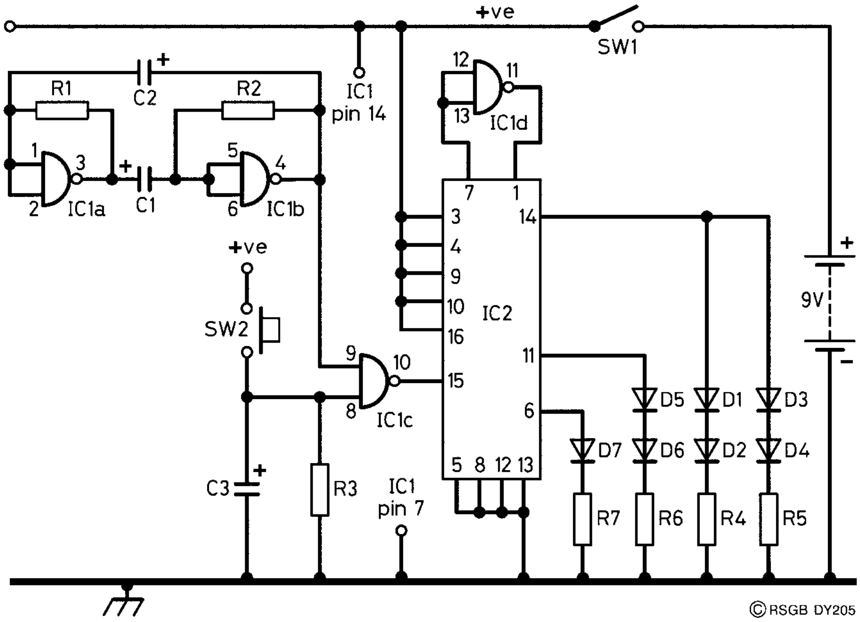
+ve
SW1
R1
C2
R2
IC1 pin 14
12
11
13
IC1d
1
3
2
IC1a
C1
5
4
6
IC1b
7
1
3
14
4
9
10
16
IC2
+ve
SW2
9
10
15
11
8
IC1c
6
D5
D1
D3
D7
D6
D2
D4
9V
C3
R3
IC1 pin 7
5 8 12 13
R7
R6
R4
R5
©RSGB DY205

Figure 1 Electronic dice, circuit diagram

reaches IC2 until the ‘other’ input of IC1c is brought away from logic 0. This happens when SW2 is pressed. Pressing SW2 momentarily, as is normally done, charges up C3 to 9 V, putting a logic 1 on the second input of IC1c and allowing the clock signal through to IC2. As the switch is released, C3 begins to discharge through R3, gradually lowering the voltage on pin 8 of IC1c. As this voltage crosses 4.5 V (half the supply voltage), IC1c then treats this input as logic 0, which cuts off the clock signal from IC2 again. Thus, when SW2 is momentarily pressed, the clock signal is fed to IC2 for a short interval of time, before being blocked again.

IC2 is a binary counter, and the outputs that concern us are from pins 14, 11 and 6. Pin 14 is the *most significant bit* and pin 6 the *least significant bit*. See the panel for an explanation of what happens here.

To understand these outputs, you should be able to count in binary, using three bits. At the start of the counting process, all the bits have zero values, i.e. 000. The left-most bit is called the *most significant bit*, and the right-most bit is called the *least significant bit*. (When we write numbers normally, a number 1 in the left-hand position represents one hundred, 1 in the middle column represents ten, while the number 1 in the right-hand position means one. One hundred is a more significant number than one, hence the nomenclature.)

The three bits of our binary number have values, from left to right, of 4 ($=2^2$), 2 ($=2^1$) and 1 ($=2^0$). This means that a number 1 in the left-hand column signifies the normal number four, in the middle column it would represent two, and in the right-hand column, it would represent one. This should help you understand the patterns of bits which emerge as the clock waveform is counted, as the next paragraph explains.

As each cycle of the clock enters IC2, it increments its internal counter and the values of that counter are shown by the states of pins 14, 11 and 6. After the first pulse, these three pins would have states corresponding to 001; after the second, 010; after the third, 011; after the fourth, 100; after the fifth, 101; after the sixth, 110. This sequence of 3-bit numbers represents a binary count from 1 to 6 in ‘normal’ parlance. These six states are used to illuminate the conventional pattern of dots on a die, using LEDs.

As the clock cycles are counted, the LEDs flicker as the die is ‘rolled’. Resistors R4 to R7 are used to limit the current through the LEDs. When the counter stops, it can be in any of the positions shown in **Table 1**. Because of

Table 1 Dice number relative to counter output

Step	*Pins*			*Dice number*
	14	*11*	*6*	
1	0	0	1	4
2	0	1	0	2
3	0	1	1	6
4	1	0	0	1
5	1	0	1	5
6	1	1	0	3

the way in which the dots are grouped on the faces of a die, the wiring of the LEDs is simpler than it might otherwise be. You can see from Figure 1 that there are really only three sets of connections to the LEDs – one from each bit of the counter output. When pin 14 is at logic 1, four LEDs are lit, corresponding to the number four. For the number five, pins 14 and 6 will be at logic 1. For the number one, only pin 6 is at logic 1. For three, pins 11 and 6 are at logic 1, for six, pins 14 and 11 are at logic 1, and for two, only pin 11 is at logic 1. These conditions are summarised in Table 1.

Construction

Two pieces of Veroboard (of the strip type) are needed – one for the main circuit (**Figure 2**) and the other for the display LEDs (**Figure 3**). The display board is easier to build, so we will do that first. It measures 20 holes by 18 strips.

Cut the tracks as shown in Figure 3, using a 3 mm (1⁄inch) twist drill rotated between thumb and forefinger. Insert and solder the Veropins, resistors and the wire links. Then solder the LEDs in place, making sure their polarities are correct. The LEDs are mounted proud of the board (**Figure 4**, which will help when you come to fix the board into a case.

The main board measures 32 holes by 18 strips. As before, remove the tracks in the places shown in Figure 2. Note that track 'L' is *not* broken under the position for IC2. Solder in the IC holders and the links. Solder in R1 and R2 vertically, and R3 horizontally. Then solder in C1, C2 and C3; these all being electrolytics, check their polarities. Solder in the Veropins and prepare the connections to those components not on the main board, using stranded, insulated wire. Check the circuit carefully, and read the handling precautions given earlier before carefully inserting the two ICs.

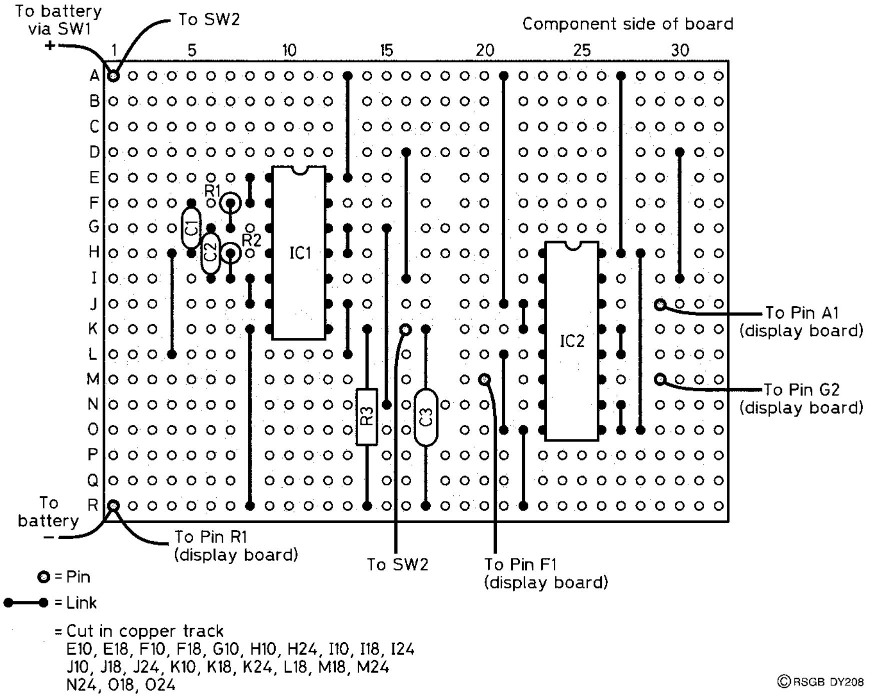
To battery via SW1
+
To SW2
Component side of board
1
5
10
15
20
25
30
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
R1
R2
C1
C2
IC1
IC2
R3
C3
To Pin A1 (display board)
To Pin G2 (display board)
To battery
−
To Pin R1 (display board)
To SW2
To Pin F1 (display board)
= Pin
= Link
= Cut in copper track
E10, E18, F10, F18, G10, H10, H24, I10, I18, I24
J10, J18, J24, K10, K18, K24, L18, M18, M24
N24, O18, O24
©RSGB DY208

Figure 2 Main board, component layout

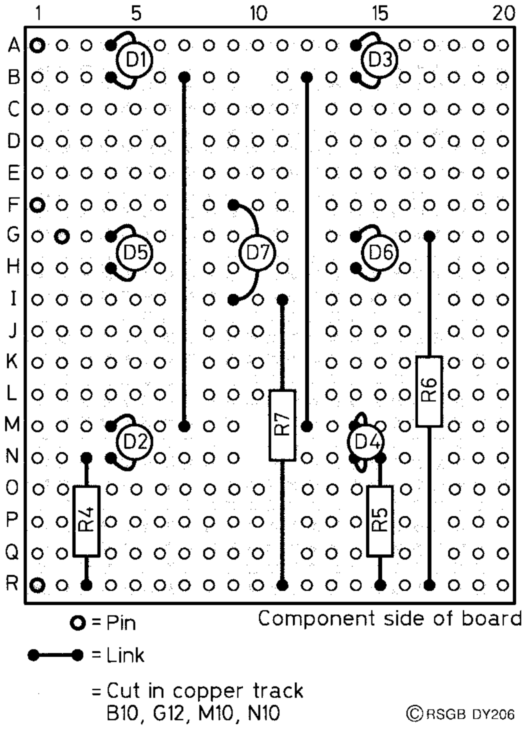
1
5
10
15
20
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
D1
D3
D5
D7
D6
D2
D4
R4
R5
R6
R7
= Pin
Component side of board
= Link
= Cut in copper track
B10, G12, M10, N10
©RSGB DY206

Figure 3 Board for the six LEDs

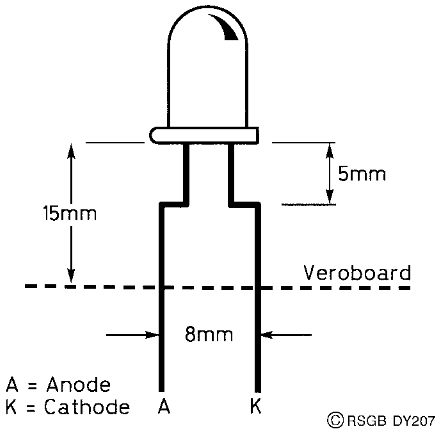

Figure 4 D7 lead bending

Testing and use

When the circuit is switched on, it is likely that the centre LED only will be lit. Press SW2, the 'throw' switch. All the LEDs should flicker, even after the switch is released, but only for a second or so. Then a standard pattern should show, signifying numbers between one and six. If this does not happen, switch off and check your circuit again. Are there any missing links? Are there any obvious dry joints (usually dull instead of shiny)? Can you see any solder bridges between tracks? Did you choose to ignore the handling precautions? If none of these results in a working circuit, try some fault-finding.

- Disconnect the lead to D7 on the main board (marked 'to pin F1' on Figure 2). Touch the lead on to the positive supply rail. D7 should light. Now transfer this lead on to pin 4 of IC1. If it flickers, then the oscillator is running (which it should).
- Transfer this lead to pin 10 of IC2. D7 should flicker when SW2 is pressed, but should be on permanently when it is released. This means that the correct signals are reaching IC2 from IC1.
- If one of the chains of LEDs (i.e. D1 and D2, D3 and D4, D5 and D6) does not light, you may have one *or both* LEDs the wrong way round.

When the die is working, you may care to experiment with the values of R1 and R2, but you should keep their values the same, i.e. R1 = R2. The larger the values of these, the more slowly the die will appear to 'roll'.

Finishing touches

On completion of the project, you will want to mount it in a smart case; any plastic box is suitable for this, with the display board mounted so that the

LEDs protrude through the top and are fixed to it using LED clips. The way in which you mount both boards to the case is entirely up to you.

Parts list

Resistors: all 0.25 watt, 10% tolerance, or better

R1, R2	5600 ohms (Ω)
R3	15 kilohms (kΩ)
R4–R7	470 ohms (Ω)

Capacitors

C1, C2	1 microfarad (μF) electrolytic, 16 V WKG
C3	68 microfarads (μF) electrolytic, 16 V WKG

Semiconductors

IC1	4011
IC2	4029
D1–D7	LEDs, any size and colour

Additional items

SW1	On-off switch SPST toggle
SW1	Push-to-make, non-latching
	Veroboard, 2 pieces, see text for sizes
	Veropins, 10
	Stranded insulated wire for general wiring
	Single-core insulated wire for links
	PP3 battery clip
	PP3 battery
	Plastic box to suit
	LED mounting clips (7)
	Means of mounting boards to box

Source

Components are available from Maplin.

63 The absorption wavemeter

Introduction

The purpose of the absorption wavemeter is to check that a transmitter is radiating within the correct waveband, and to detect any spurious *harmonic* emissions. A harmonic frequency is an integral multiple of the carrier frequency – e.g. if the carrier is at a frequency *f*, the harmonics are at frequencies 2*f*, 3*f*, 4*f* . . . and so on.

It can also be employed as a relative field-strength indicator, being used in experimentation with aerials. What it *cannot* do is to measure the transmitted frequency accurately; all it can do is to place the signal within a particular band, say the 7 MHz or the 18 MHz band (or their harmonics).

How it works

You may recognise the circuit of **Figure 1** as a crystal set with a meter replacing the headphones. It is simply the parallel combination of an inductance with a variable capacitance, which constitutes a resonant circuit with a variable frequency. A dial on the variable capacitor can be roughly calibrated with frequency. The fixed capacitor across the meter helps to improve the meter reading by making it read the peak value of the carrier wave rather than the average value.

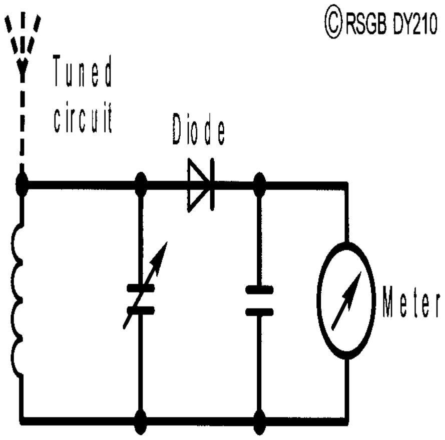

Figure 1 Absorption wavemeter circuit

When the wavemeter is situated in the vicinity of a relatively strong radio-frequency (RF) field, the inductor absorbs a small amount of energy from that field. If the circuit is resonant at the same frequency as that of the transmitter, an RF voltage is produced across the coil and is proportional to the strength of the RF field. The RF voltage is *rectified*, or *detected* (turned from AC into DC) by the diode and the second capacitor, and is displayed by the meter.

The tuned circuit

The formula for calculating the resonant frequency of a coil and capacitor is relatively simple. It is

$$f = \frac{1}{2\pi\sqrt{LC}},$$

where f is the resonant frequency (Hz),
L is the inductance (H), and
C is the capacitance (F).

If this formula puts you off, there is this slightly easier version:

$$f = \frac{159}{\sqrt{LC}},$$

where f is the frequency in **MHz**,
L is the inductance in **microhenries** (**μH**), and
C is the capacitance in **picofarads** (**pF**).

When a variable capacitor is specified, it is usual to quote its maximum capacitance, i.e. the value when the vanes are fully meshed. When the value of the capacitor is small, it is usual to quote its minimum capacitance also, the value when the vanes are fully open. If you use a variable capacitor of 10–100 pF with a 100 μH inductor, you will use the value of 10 pF to calculate the maximum resonant frequency, and the value of 100 pF to calculate the minimum resonant frequency, giving a tuning range of 1.6 MHz to 5 MHz. Check it yourself!

64 An HF absorption wavemeter

Introduction

An absorption wavemeter is basically a simple tunable detector circuit, such as would be used in a crystal set, but with the headphones replaced by a meter, in order to indicate the strength of the received signal.

The circuit

Figure 1 shows the circuit of the wavemeter. It is designed to cover all the HF bands from 1.8 MHz to 28 MHz in four switchable ranges. There is no built-in method of amplification, so a sensitive meter is needed in order to indicate sufficiently using the available absorbed energy. The meter does not need calibrating, it serves to produce only an indication of the absorbed power, *not* its absolute value.

An absorption wavemeter does not have outstanding selectivity, mainly because of the loading effect of the meter. This problem is ameliorated here (but only slightly) by the inclusion of R1 in series with the meter. It gives

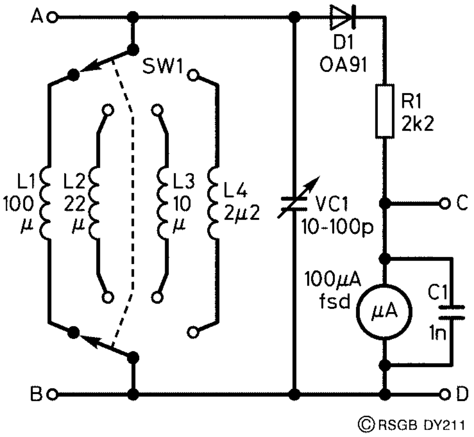

Figure 1 HF absorption wavemeter, circuit diagram

improved selectivity at the expense of a slight loss of sensitivity. The appearance of capacitor C1 across the meter is discussed in *The Absorption Wavemeter*, in the previous section.

Coils

These can be purchased (see the parts list) or hand-wound on short lengths of PVC water pipe or conduit, and the coil-winding details are given in **Table 1**. In theory, the 100 pF tuning capacitor would cover all the HF bands with only three coils, omitting the 22 μH coil. Sample calculations are explained in the previous article. However, its inclusion avoids the common problem of having some bands cramped at the extreme edges of the tuning scales.

Table 1 Close-wound coil values

L (μH)	No of turns	Span (mm)	SWG
PVC FORMER OF OUTSIDE DIAMETER 25 mm			
2.2	9	9	20
10	24	24	20
22	44	44	20
100	110	60	26
PVC FORMER OF OUTSIDE DIAMETER 20 mm			
2.2	10	10	20
2.2	9	5	26
10	32	32	20
10	24	12	26
22	63	62	20
22	41	21	26
100	134	70	26
PVC FORMER OF OUTSIDE DIAMETER 19 mm			
2.2	11	11	20
2.2	9	5	26
10	35	34	20
10	25	13	26
22	68	67	20
22	44	22	26
100	157	79	26

Construction

Hand-wound coils, using 19–25 mm PVC formers, have single-layer windings, with their ends secured by threading the wires through two small holes drilled in the formers at each end of each winding.

You may have wondered about the use of the two-pole switch SW1, as all the lower ends of the coils are grounded anyway. The answer lies in the ease of construction. Each coil can be mounted on the tags of the switch for support. If you are using the commercial coils, they are small enough to be accommodated on a miniature rotary switch; a larger rotary switch is required for hand-wound coils. The specification of a 6-way switch allows for the addition of extra frequency ranges at little extra cost. Rotary switches such as the 2-pole, 6-way variety have an adjustable end-stop which allows the number of ways to be set anywhere between 2-ways and 6-ways. Remove the nuts from the switch, and a ‘washer’ will fall out. It is not a washer, really, and has a little piece of metal bent over at right angles to the washer. This fits into one of 12 holes in the body of the switch, and prevents the shaft from turning through 360° and selecting the number of ways.

Calibration

This process needs a commercial amateur-bands transceiver, if it is to be done quickly and accurately.

- Connect the transceiver to a dummy load, and connect a piece of flexible wire from point A in Figure 1 and wrap two or three turns around the cable between the transceiver and the dummy load.
- Turn the transceiver power level control to minimum, and set the frequency at 1.81 MHz and the mode to FM or AM. Switch to transmit.
- With the wavemeter set to its lowest frequency range (i.e. 1.8–3.5 MHz), rotate the tuning capacitor until maximum deflection is obtained from the meter needle. Write the frequency on the inner ring of the wavemeter scale at that point.
- Switch off and retune the transmitter to 3.5 MHz. Repeat the above process.
- Repeat, using switch range 2 for 3.5, 7 and 10 MHz, range 3 for 7, 10 and 14 MHz, and range 4 for 14, 18, 21, 24 and 28 MHz.

An alternative source of calibration signal could be an HF signal generator, with a single-turn loop of wire at the remote end of its cable, and the wire from the wavemeter brought close to the loop.

Remember that the marks on the scale represent frequency *bands*, not precise frequencies.

Extending the range

If you would like to experiment with increasing the range to the lower VHF band, then 50 MHz should be achievable with an additional self-supporting coil (no former). Try three turns of 20 SWG enamelled copper wire of 25 mm ID (internal diameter). Wind it on some 25 mm PVC pipe, as before, then slide out the pipe! The same result could be achieved with four turns of 19 or 20 mm ID, or 12 turns of 7.1 mm ID.

Parts list

Resistor	
R1	2200 ohms (Ω) 0.25 watt, 10% tolerance
Capacitors	
C1	1 nanofarad (nF) min. ceramic
VC1	100 picofarad (pF) variable
Inductors	
L1	100 microhenries (μH)
L2	22 microhenries (μH)
L3	10 microhenries (μH)
L4	2.2 microhenries (μH)
Semiconductors	
D1	Germanium OA91 or 1N4148
Additional items	
SW1	Rotary 2-pole, 6-way
	Moving-coil meter, 50 or 100 μA FSD (any meter of this sensitivity will do)
	Plastic box, approx. 150 by 80 by 50 mm
	Connecting wire, coloured

65 A vertical aerial for 70 cm

Introduction

If the range of your hand-held transceiver is very limited, what you need is a vertical aerial mounted outside. This design is a half-wave dipole fed at the end instead of the middle. Because the impedance of the dipole is high at its ends (and low in the centre), a matching circuit is needed so that this high impedance can be matched to that of the low-impedance coaxial cable. All that is needed is a coil, which increases the *electrical length* of the aerial to ⅝-wavelength. (Note that the *electrical* length (i.e. the length as it appears to an RF signal) is not necessarily the same as the *actual* length.)

Construction

The aerial element and the coil are made from a single piece of 1.5 mm welding (brazing) rod, and the dimensions are given in **Figure 1.** Wind the coil around a 4 mm rod or the shank of a twist drill. The lower end is filed to a point and then soldered into the centre conductor of a 4-hole panel-mounting BNC socket. (Try to obtain a good-quality BNC socket with PTFE insulation – the insulation of cheaper sockets is easily damaged.) Trim the element to 427 mm (top of element to top of coil) *after* the wire has been soldered to the socket.

The base coil causes the aerial to be rather ‘whippy’, so a piece of 5 mm plastic knitting needle can be cut to the length of the coil and then forced into it.

The radials are made from four lengths of 3 mm welding rod. These are bent and soldered into the four mounting holes of the socket, and then cut to the lengths shown in Figure 1.

Testing

Using a standing-wave-ratio (SWR) meter connected between your aerial and the transceiver, measure the SWR (or obtain some indication of the reflected power) on transmit. If it is greater than 1.5, switch off the transmitter, trim about 3 mm off the end of the element, and try again. Repeat the process until the measured SWR (or the reflected power) is as low as you can get it.

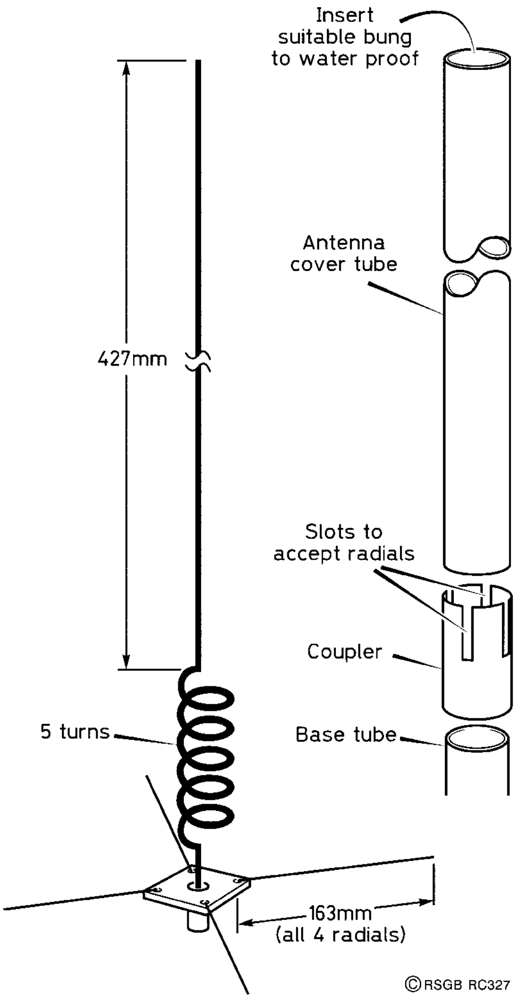
Insert suitable bung to water proof
Antenna cover tube
427mm
Slots to accept radials
Coupler
5 turns
Base tube
163mm (all 4 radials)
©RSGB RC327

Figure 1 Construction of the 70 cm antenna

Installation

The aerial is made waterproof by enclosing it in a 22 mm diameter PVC waste water pipe. It is 'weldable', and available at plumbers' merchants, usually by the metre. A coupler is slotted to take the radials (Figure 1). File the BNC socket as required, so that it slides inside the coupler until the radials poke out of the slots. Cut a length of plastic tubing which is 30 mm longer than the aerial, and push it into the coupler. You will then need a plastic bung or screw-top to waterproof the top end.

Plastic welding solution is now applied to the joints, sealing the aerial inside the tube. The coupler has been weakened as a result of making the slots, so it is worthwhile applying PVC tape around this joint until the welding solution sets.

A support for the aerial is made from an off-cut of tube pushed into the lower end of the coupler and held with a self-tapping screw.

Parts list

1	4-hole panel-mounting BNC socket
1 m	3 mm brazing rod (may be available from a small garage)
1 m	1.5 mm brazing rod
1 m	22 mm PVC waste water pipe
1	22 mm straight coupler
	Plastic welding solution

66 A UHF corner reflector aerial

Introduction

The corner reflector is a well-known design and is capable of good performance on the VHF and UHF bands. At UHF, the practical implementation of the corner reflector is an ideal constructional project.

Some details

Expressed quite simply, the aerial consists of a $\frac{\lambda}{2}$ dipole (where λ is the standard symbol for wavelength, making a '$\frac{\lambda}{2}$ dipole' a half-wave dipole).

Nothing new in that, you might say. However, the interesting feature is the reflector, which is not the usual single element, but a 90° metal 'corner', acting rather like a parabolic dish as used for satellite signal reception. The wind resistance of this type of reflector makes it impractical so, to reduce the 'windage' quite significantly, we make the 'corner' from closely spaced rods, as illustrated in **Figure 1**.

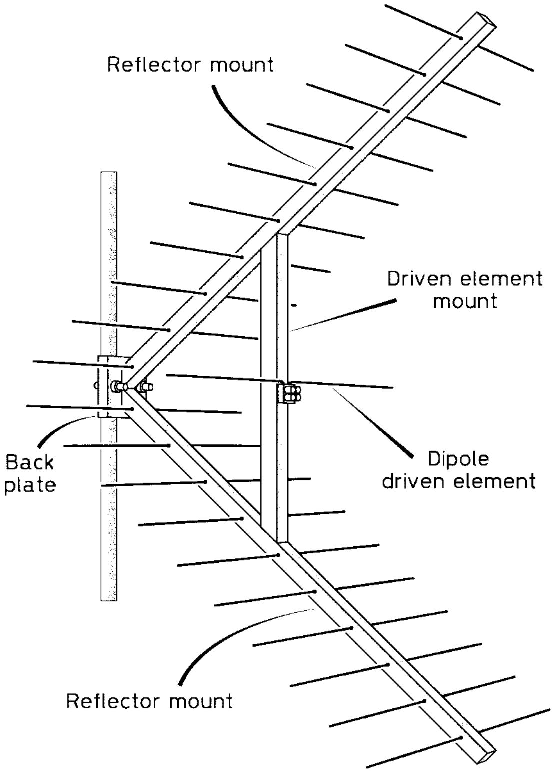

Figure 1 70 cm corner reflector antenna

©RSGB DY215

The reflector consists of a number of 0.6λ rods, spaced from each other by 0.1λ. The aerial frame can be made of metal or wood, but wood is easier to work with, and mounting the elements to the frame is simpler. The prototype was made with wood of 20 mm by 15 mm cross-section, as **Figure 2** shows. The wood was varnished for protection. The elements were made from 1.5 mm diameter copper wire, because a large reel of the wire happened to be available. The wire diameter is not critical; tubing could be used just as successfully. 14 SWG hard-drawn copper aerial wire would be even better than that used in the prototype.

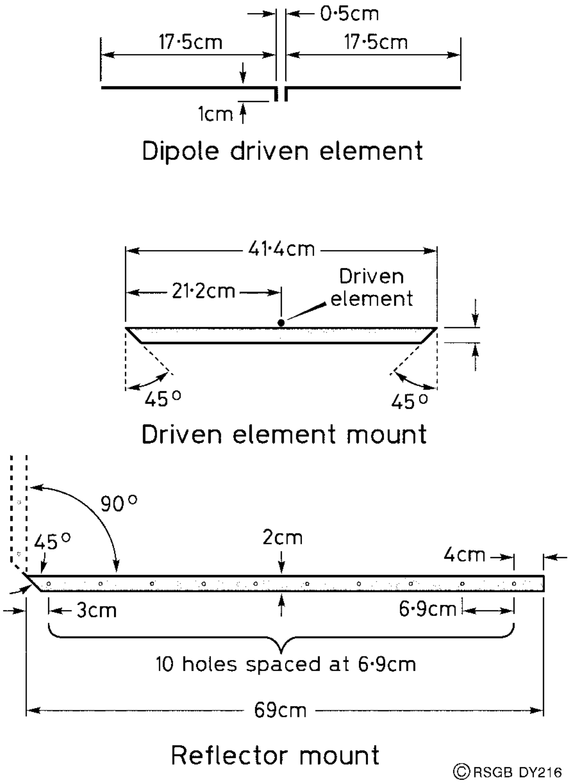
0·5cm
17·5cm
17·5cm
1cm
Dipole driven element
41·4cm
21·2cm
Driven element
45°
45°
Driven element mount
90°
45°
2cm
4cm
3cm
6·9cm
10 holes spaced at 6·9cm
69cm
Reflector mount
©RSGB DY216

Figure 2 Driven element dimensions, together with boom dimensions for driven element and reflectors

Construction

This project is just as much a woodworking project as a radio project! Follow the instructions carefully, and you should have little trouble.

- Cut the booms for the reflectors, as shown in Figure 2. A mitre block is invaluable here in producing the 45° corners.
- Using the dimensions given on the diagram, mark the hole positions for the reflector elements, and then drill holes of a size which holds the elements firmly.
- Cut the driven element boom according to the diagram, and mark the point midway along the longer side, which will assist you later in positioning the driven element.
- Cut the back plate to size (about 120 mm by 80 mm). You may need to alter this size depending on the size of the U-bolt you will be using to clamp the aerial to the mast.

- If you want to be extra cautious in your construction, use the belt-and-braces approach, commonly known as 'screw-and-glue' to fix the booms to each other and to the back plate.
- Fix the reflector booms to the back plate first, then slide in the driven-element boom until it will go no further, then apply the wood glue and screw the two ends tightly to the reflector booms. Leave for the period prescribed by the glue manufacturers for the glue to harden.
- Varnish the whole structure.
- Cut the driven element to the correct size plus a couple of centimetres (the reason for this will be evident in the *Testing* section), and fix it to the centre of its boom (at the position you marked earlier) with a 'chocolate block' connector to which the coaxial feeder cable will eventually be connected.
- Cut and fix the reflector elements in place. If you find that these are a loose fit in the holes then, *for each element*, drill a pilot hole through the boom to intersect the hole for the element. File off the point of a woodscrew, and screw it gently into the pilot hole until it meets the element and grips it in place. You will now see why the point was filed off! Alternatively, you can glue the elements in place.

Testing

Place the aerial on a mast, clear of obstructions. Connect it to a transceiver with a length of coaxial cable, with a standing-wave-ratio (SWR) meter in circuit. Find a clear frequency, identify your transmission and ask if the frequency *really is* clear. If so, key the transmitter again and note the SWR. **Do not stand in front of any aerial when it is radiating!** The length of the driven element must be adjusted to obtain an SWR of less than 2. If you have to shorten the dipole, bend the ends over rather than cut them off. That way, if you go too far, your can lengthen them again! The dipole was initially cut too long intentionally, to allow for adjustment here! Bending the ends over also reduces the risk of physical damage to clothing, skin and eyes. You may like to consider applying the same technique to the reflector elements for that reason alone.

Moving on . . .

Once you have warmed to the idea of the corner reflector as an aerial, you might like to ring the changes regarding the reflector. How about replacing the 20 reflector elements with a wire mesh, such as garden centres sell as 'chicken wire'? Choose the finest mesh if there is a choice. Some extra support may be needed around the edges of the mesh, but you could go on to make a comparison of aerial gain between the two types, using the *UHF Field Strength Meter* described elsewhere in this book.

Materials

Stiff wire or thin-walled tubing for dipole and reflector
Frame – wood, 15 mm by 20 mm cross-section, lengths given in text
Back plate – stout plywood, dimensions given in text
U-bolt to suit mast
Wood screws
Wood glue
50 W coaxial cable for feeder
2-terminal ‘chocolate block’ for dipole connection to feeder
Varnish

67 A switched dummy load

Introduction

A dummy load is a pure resistor of value 50 ohms which can replace your transmitting aerial and enable you to operate the transmitter for test purposes without radiating a signal. It sounds simple enough, but there are two main problems. Firstly, it is impossible to delve into your junk box and emerge with a resistor that will dissipate 100 W PEP and still retain its marked resistance value. Secondly, a ‘pure resistance’ is very difficult to achieve. A pure resistance is a device which has resistance but no reactance. All common resistors have significant reactance at radio frequencies, particularly the wire-wound varieties, which have a helical (i.e. wound like a coil) construction. This is particularly annoying, because wire-wound construction is normally used for large-wattage resistors.

Although all resistors have *some* reactance, not all are quite as bad as the wire-wound type. Carbon film resistors are made by depositing a thin film of carbon on the surface of a small, hollow ceramic cylinder, the thickness of the film of carbon determining the value of the resistor. Provided the lead lengths are kept short, these resistors have a tolerably small reactance, and will be used in this project.

Bearing the load

A 2 W carbon film resistor is hardly going to withstand our 100 W PEP of SSB, so it is obvious that the design of our dummy load must be a little more complex than a single resistor and a switch, despite what **Figure 1** might suggest! In fact, it uses 20 resistors, each of value 1000 Ω (1 kΩ). How does this solve our problem?

Perhaps a little theory is in order here, but no more than is required by the Radio Amateurs' Examination.

When two *equal* resistors of value r are combined in parallel (i.e. side by side), the total resistance, R_T, is given by:

$$\frac{1}{R_T} = \frac{1}{r} + \frac{1}{r}.$$

Adding $1/r$ to $1/r$ gives $2/r$, therefore:

$$\frac{1}{R_T} = \frac{2}{r}, \quad \text{i.e. } R_T = r/2.$$

So, by connecting *two* equal resistors in parallel, we get a combined resistance which is *half* the individual resistances. If we combine *three* in parallel, we get a *third* of the resistance, and so on.

Here, we are connecting *20* resistors of 1 kΩ in parallel, so we will produce an overall resistance of *one-twentieth* of the individual resistance, i.e. 1000/20 = 50 Ω, which is what we set out to achieve!

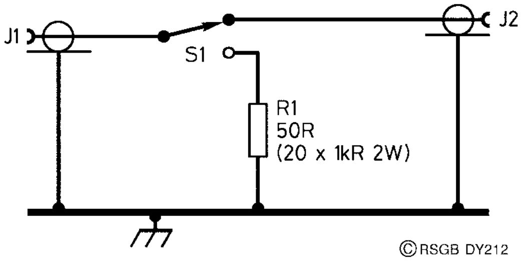

Figure 1 Basic circuit diagram

This is not the only advantage, though. Each resistor is capable of dissipating 2 W, so 20 of them will safely dissipate 40 W, for short periods at least. This power dissipation is approximately the same as 100 W PEP of normal speech, so the design should be capable of use in an 'average' amateur station. The power-handling ability of any dummy load can be improved by providing a 'heat sink' which helps to conduct the heat energy away from the resistors, thus lowering their temperature. One popular heat sink is a can of transformer oil, into which the resistors are immersed. This design uses a rather more mundane heat sink, but which is adequate for the job in hand.

Figure 2 PCB resistor supports

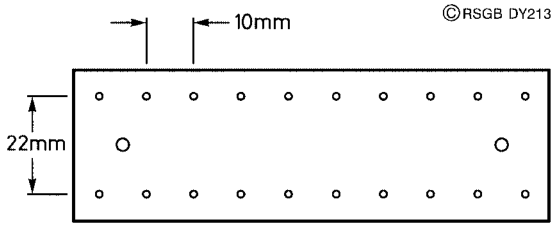
10mm
©RSGB DY213
22mm

2 PCBs (single sided) 100 x 35mm, 20 holes
1 'earth' PCB with two 3mm holes

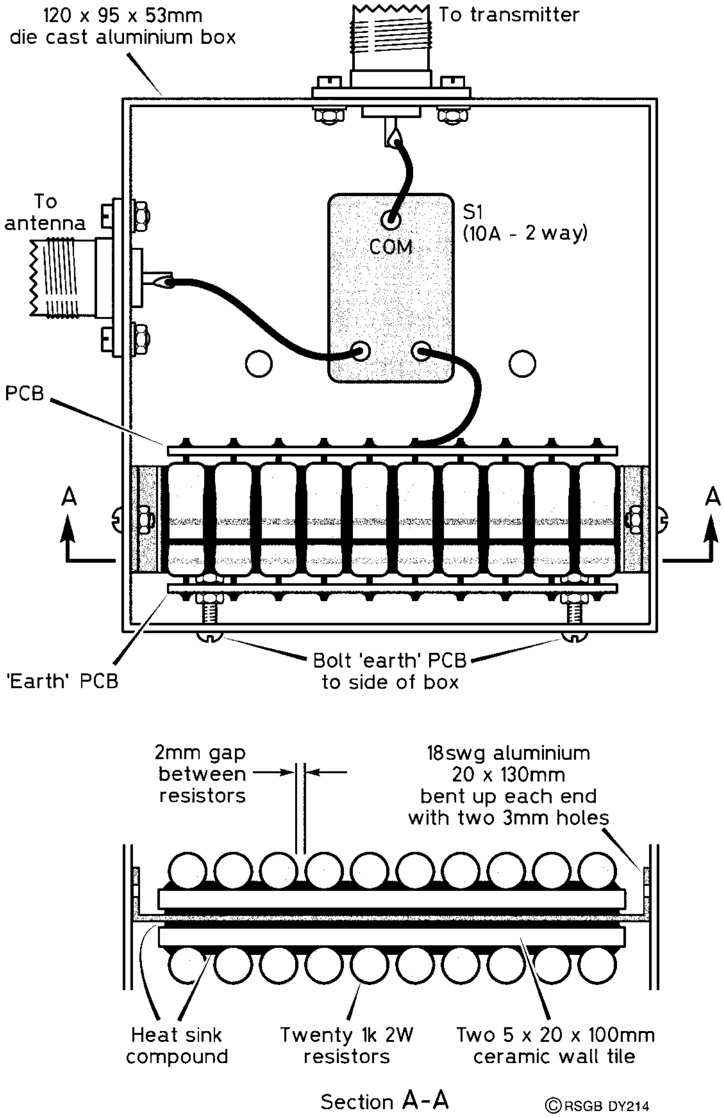
120 x 95 x 53mm
die cast aluminium box
To transmitter
To
antenna
S1
(10A - 2 way)
COM
PCB
A
A
'Earth' PCB
Bolt 'earth' PCB
to side of box
2mm gap
between
resistors
18swg aluminium
20 x 130mm
bent up each end
with two 3mm holes
Heat sink
compound
Twenty 1k 2W
resistors
Two 5 x 20 x 100mm
ceramic wall tile
Section A-A
©RSGB DY214

Figure 3 Switched dummy load, component layout

Pouring the heat away

Two basic PCBs are needed, the details being given in **Figure 2**. Both measure 11 × 35 mm. Each has a set of 20 holes of diameter about 1 mm to take the ends of the resistors. Only one of them has two 3 mm holes which are used for mounting the completed load. **Figure 3** shows how the heat sink is assembled. Solder one end of each resistor into the 'earth' PCB, and then mount this to the box using the two bolts as shown in the diagram. Between each rank of resistors is a 'sandwich' consisting of an aluminium strip and two pieces of ceramic wall tile, to act as a heat sink for the resistors. Heat sink compound is used to provide good thermal contact between the resistors and the tiles, and between the tiles and the aluminium strip. This is shown in the lower part of Figure 3. Thread the loose wires of the resistors through the holes in the second PCB, solder into place, and crop the protruding wires.

Switching

A changeover switch must be used so that the transceiver can be switched between the dummy load and the aerial without unscrewing connectors. For most purposes, an ordinary 10 A 230 V changeover switch will suffice, as found in many electrical shops and DIY stores.

Wire this into the circuit as shown in Figure 1 and Figure 3. Check your wiring. Put S1 in the 'dummy load' position. If you have a multimeter which includes an ohmmeter, measure the resistance across the socket, J1, before connecting it to the transceiver. It should be very near 50 Ω. With S1 in the 'aerial' position, there should be an infinite resistance across J1. Move your ohmmeter to read the resistance across J2. It should be infinite for both positions of S1. If all seems correct, close the box and your dummy load is ready for use!

Parts list

Resistors
  1000 Ω (1 kΩ) 2 W carbon film, 20 required
Additional items
  10 A 230 V changeover (SPDT) switch
  SO–239 sockets, 2 required
  Aluminium box 120 × 95 × 53 mm or similar
  Ceramic wall tile cut as required
  Aluminium strip, 18 SWG
  Nuts and bolts as required

68 A simple Morse oscillator

Introduction

This is an excellent project which uses the 'junk box' as its source of components. If you have trouble in finding the bits for this one, a good source of the components for this and many other similar projects is to be found with the parts list at the end of the project.

The circuit

This is shown in **Figure 1**, and uses an 'unknown' Plessey chip, which makes the overall circuit extremely easy to build. A 0.1 µF capacitor is connected between pins 7 and 8, a speaker (in the popular 8 to 25 Ω impedance range) is connected between pins 8 and 9. If a 9 V battery is connected with its positive terminal to pin 8 and its negative terminal to pin 1, 3 or 5, a tone will be produced in the speaker.

To make this circuit into a good Morse practice oscillator, it is necessary only to insert a Morse key into the supply rail from the battery.

However, there is another refinement which you may care to build into the circuit. The tone from the loudspeaker is different, depending upon which of pins 1, 3 or 5 you use. In the prototype, a single-pole changeover switch was used to select the tones from pin 1 or pin 3, and the Morse key would be connected to the circuit via a small jack socket. There is no need for an on/off switch, as the Morse key performs that function. The switch and the jack socket can be seen in the photograph.

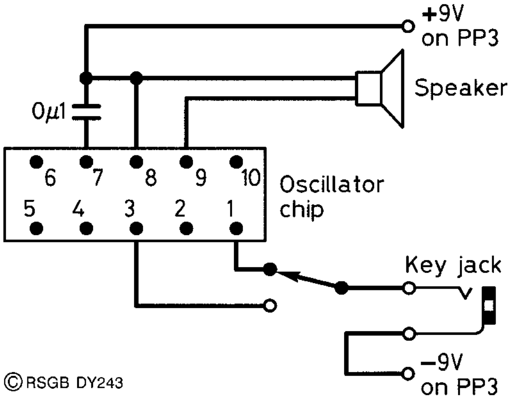

Figure 1 Morse oscillator, circuit diagram

The case

Almost any small loudspeaker will do; the higher the impedance the better. The ‘impedance’ is the figure usually printed on the rear of the speaker magnet. It is a value given in ohms (Ω). For simple circuits like this, you will usually find that the higher the impedance (within reason), the louder the sound it will produce. Speakers from old transistor radios will work, although their impedances can be rather low sometimes.

Any case big enough to house the components and the battery will be suitable. The prototype used a ‘Walkman’-type speaker and case, and is shown in the photograph.

To use the circuit, simply attach a 3.5 mm jack plug to your Morse key, and insert it into the socket. Nothing should happen until you press the key, when a tone should be heard from the loudspeaker.

Another application

Try soldering the two wires of a ‘twisted pair’ to the 3.5 mm jack plug. Touching the wires together produces a tone from the speaker. This simple circuit can then be used as a ‘continuity tester’. Touch the two wires to the ends of a fuse. If there is no sound, then the fuse is blown. There are many other such tests you could perform with this device – to check whether there really *is* a connection between one end of a wire and the other, for example.

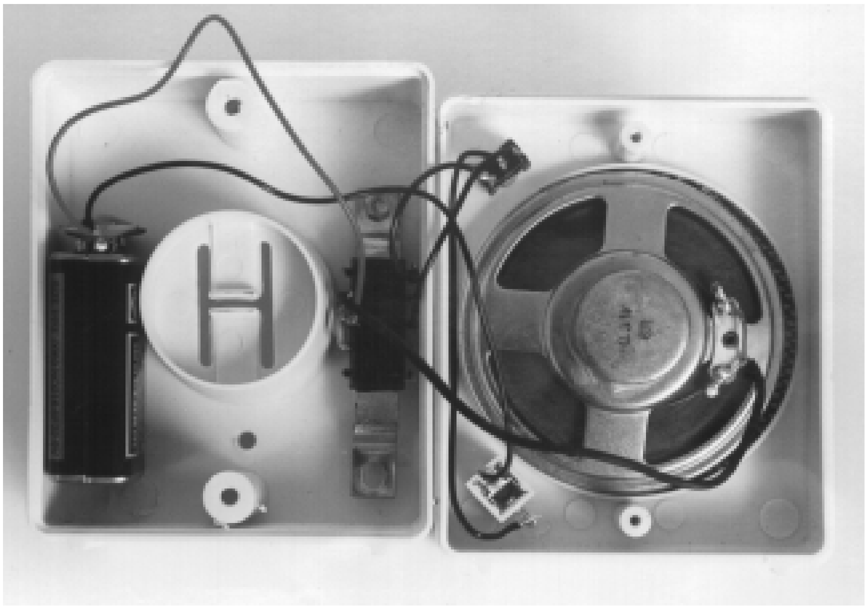

Warning

You must **never** make such tests on any equipment which is connected to the mains supply, even if it is switched off. If you want to make any such tests, make sure you are supervised by someone who understands what you are doing and is competent to advise and supervise you.

Parts list

Plessey oscillator chip
0.1 μF capacitor (disc ceramic)
Loudspeaker 8 to 25 Ω
3.5 mm jack socket (and 3.5 mm jack plug if needed)
Single-pole changeover switch

The Plessey oscillator chip is available from J. Birkett, 25 The Strait, Lincoln LN2 1JF, tel: 01522 520 767.

John Birkett may also provide a kit of parts (the chip, capacitor and loudspeaker).

69 A bipolar transistor tester

Introduction

This is a circuit which will test normal transistors, i.e. npn or pnp. It has the advantage of being able to test devices while they are still connected in their original circuits. However, when such tests are made, **the circuit containing the transistor under test must not be switched on**.

The circuit and how it works

The circuit runs from a 9 V battery such as a PP3 or six AA-type 1.5 V cells. Alkaline cells are to be preferred, as their electrolyte leakage properties are better. The circuit shown in **Figure 1** uses a single CMOS integrated circuit type 4001 or 4011. CMOS circuits require special handling precautions which are described in the project *Christmas Tree LEDs*, elsewhere in this book.

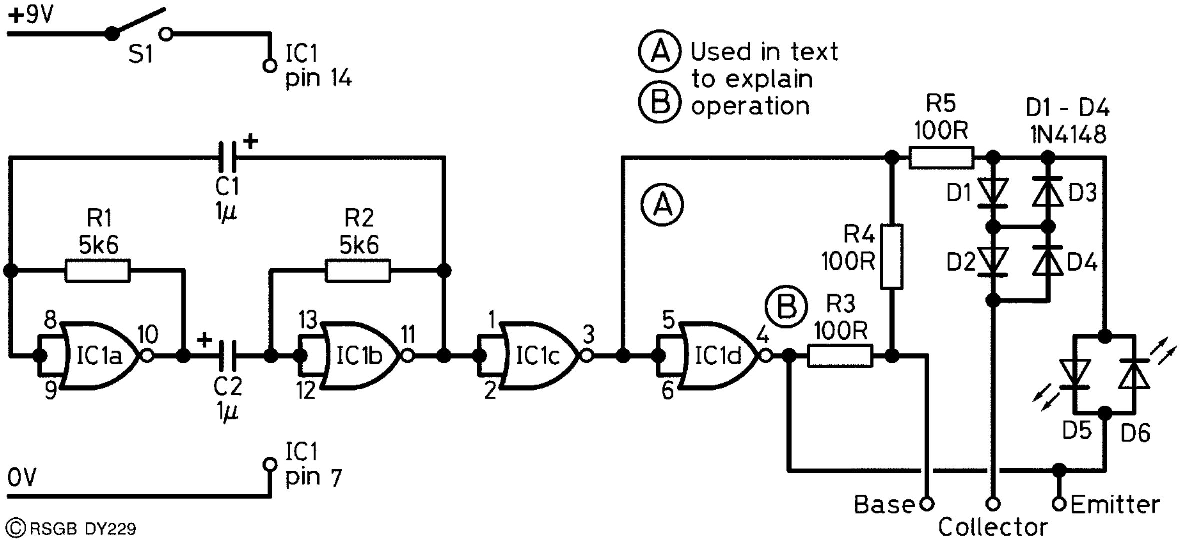
+9V
S1
IC1 pin 14
A Used in text
B to explain operation
C1 1µ
R1 5k6
R2 5k6
8
9
10
IC1a
C2 1µ
13
12
11
IC1b
1
2
3
IC1c
5
6
4
IC1d
A
B
R3 100R
R4 100R
R5 100R
D1 - D4 1N4148
D1
D2
D3
D4
D5
D6
0V
IC1 pin 7
Base
Collector
Emitter
©RSGB DY229

Figure 1 Transistor tester, circuit diagram

Inside the IC are four logic gates (see *Digital Logic Circuits*) which are all connected as inverters, which means that the output signal is always the logical 'opposite' of the input. The first two gates are connected as an oscillator; the circuit being the same as that used in *An Electronic Die*.

The output of the oscillator, at pin 11, is connected to the input of a buffer stage, IC1c, which helps to isolate the oscillator from the circuit that follows it. The buffer output appears on pin 3, which we shall label as *test point A* for future use. Another inverter, IC1d, follows this, its output at pin 4 being labelled *test point B*.

There are two LEDs connected back to back at the circuit output. These are D5 and D6, D5 being red and D6 being green. Whatever the output of the oscillator at any instant, one of the LEDs must be lit and the other unlit. With point A positive and point B zero, the red LED is lit, when A is zero and B is positive, the green LED is lit. Because the oscillator output is repeatedly switching from one polarity to the other, the lit LED is alternately red and green. They switch between the two colours much faster than we can see, so what we think we see are both LEDs lit together.

The two 100 ohm resistors, R3 and R4, provide the bias to the base of the transistor under test. When A is positive and B is zero, the base-emitter junction of the transistor will be forward-biased, and the transistor will switch on (if it is a working npn type). When the transistor is on, it effectively short-circuits D5 (the red LED) and it extinguishes. When point A is zero and B is positive, an npn transistor will be switched off and the green LED (D6) will light. Thus, for a working npn transistor, only the green LED is lit.

If the transistor under test is a pnp type, it will switch on when A is zero and B positive, thus short-circuiting the red LED (D6). When A is positive and B zero, the transistor is off and the red LED (D5) is lit. Thus for a working pnp transistor, only the red LED is lit.

To summarise, the states of the LEDs indicate the following conditions:

- Both LEDs apparently lit: no transistor connected, or transistor permanently open circuit.
- Neither LED lit: a collector-emitter short-circuit is almost certain.
- Red LED alone: pnp transistor in working order.
- Green LED alone: npn transistor in working order.

Construction

The prototype tester was built on a piece of Veroboard measuring 15 strips by 18 holes. Cut the tracks using a track-cutting tool or a 3 mm (⅛ inch) twist drill, as shown in **Figure 2**. Notice that, in this diagram, there is no

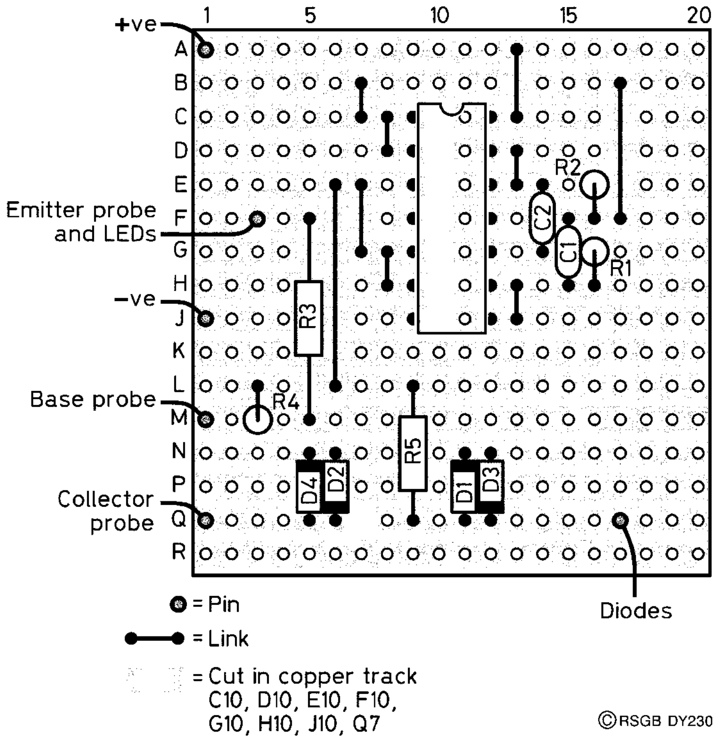
+ve
1
5
10
15
20
A
B
C
D
E
F
G
H
J
K
L
M
N
P
Q
R
Emitter probe and LEDs
−ve
Base probe
Collector probe
R1
R2
R3
R4
R5
C1
C2
D1
D2
D3
D4
Diodes
= Pin
= Link
= Cut in copper track C10, D10, E10, F10, G10, H10, J10, Q7
©RSGB DY230

Figure 2 Transistor tester, component layout

track ‘I’, so try to avoid miscounting when you translate diagram positions to real positions on the board. Hold the board up to a strong light to ensure that there is no copper swarf shorting adjacent tracks together, and that you have made the cuts in the correct places.

Having done this, insert and solder Veropins for all the connections to be made to components not on the board itself. Then solder in the IC holder and the wire links. Then insert the other components in the order resistors, capacitors and diodes D1–D4. Some resistors are mounted vertically so that their connections are on adjacent tracks. Double check the diode polarities – it is easy to make a mistake when wiring diodes in anti-parallel! Next, connect up the off–board components, again making sure that the LEDs have the correct polarity. The probe leads for the emitter, base and collector should be made from different colours of wire and terminated in probe clips (small insulated crocodile clips).

Check carefully for dry joints and errant blobs of solder. Plug the IC into its holder, ensuring that it is inserted the right way round, as shown in Figure 2.

Testing

Without a transistor in circuit, and the battery connected, both LEDs should be lit. Connect a known good npn transistor and verify that the green LED lights. Now simulate two transistor faults: disconnect the base lead and both LEDs should light; remove the transistor and connect the emitter and collector leads together. Neither LED should light.

Repeat the tests with a known good pnp transistor. The results should be the same, except that the correct indication should now be a lit red LED. On your computer, make a small label of the bulleted list above, to fit on your tester showing the states of the LEDs and what they mean. It will act as a useful *aide mémoire* when you use the tester in future.

Using

The circuit will test transistors in isolation or in an existing circuit, i.e. prior to use. You can check the lead identifications in component catalogues such as the Maplin catalogue. The tester is ideal for going through the large bags of unmarked transistors that you can buy for a song at rallies. You can sort them into three piles – npn, pnp and dud!

Parts list

Resistors: all 0.25 watt, 10% tolerance or better

R1, R2	5600 ohms (Ω)
R3, R4, R5	100 ohms (Ω)

Capacitors

C1, C2	1 microfarad (μF) electrolytic, 16 V WKG

Semiconductors

D1–D4	1N4148 general-purpose silicon diodes
D5	Red LED
D6	Green LED

Integrated circuit

IC1	CMOS 4001 or 4011

Additional items

Veroboard, 15 strips by 18 holes
Veropins
PP3 battery and connector (or 6 × AA cells in PP3 clip holder)
Switch, SPST
Connecting wire
One each of known working npn and pnp transistors for test purposes.

Source

Components are available from Maplin.

70 The ‘Yearling’ 20 m receiver

Introduction

Published to celebrate the first anniversary of *D-i-Y Radio,* this excellent receiver design forms a suitable ‘second receiver project’ for those who have successfully completed the MW receivers earlier in this series. The receiver is powered from a PP3 battery or from a mains adaptor, and can be built with the help of an experienced constructor, on a prototype board. The circuit diagram and some of the components used are shown in the separate diagram. Headphones or a loudspeaker can be used and, once the radio is completed, a few simple adjustments will make the *Yearling* spring to life!

Building the receiver

Before starting the constructional process, start by identifying all the parts. One by one, tick them off against the parts list. Are their values correct? The varactor diode is a twin type (see circuit diagram overleaf), and must be cut *carefully* down the middle, producing two devices, D1 and D2, with two wires each.

First, solder the IC sockets, followed by the coils (inductors); L1 is pink inside the top, and L2 is yellow inside. Then, solder in the varactors; the lettering on D1 should be next to coil L1, and the lettering on D2 should face resistor R7. After those, the capacitors, wire links and resistors should be soldered to the board. Take care to wire the voltage regulator, IC3, correctly. Solder in the crystal X1 as quickly and deftly as you can – crystals do *not* take kindly to having their leads bent and being fried with a soldering iron! Make sure that the electrolytic capacitors C2, C12, C15 and C16 are fitted the right way round. Most electrolytic capacitors have only the negative lead marked.

Figure 1 shows the rear of the front panel, illustrating the connections from the board and antenna socket to the controls. All normal connecting wires are 22 SWG or thereabouts, with insulation. Their lengths should be about 15 cm, except for the battery lead to the switch, which is about 8 cm. It is recommended that you use different-coloured wire for each connection to a control. **Figure 2** shows this. The variable resistor section of VR5, the AF gain control, uses single screened cable connected to 0 V (ground) at the

Amateur Radio

D-i-Y

Year

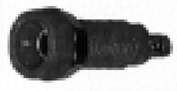

Antenna Socket

Variable Inductor

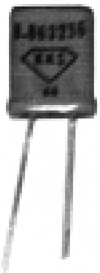

Crystal

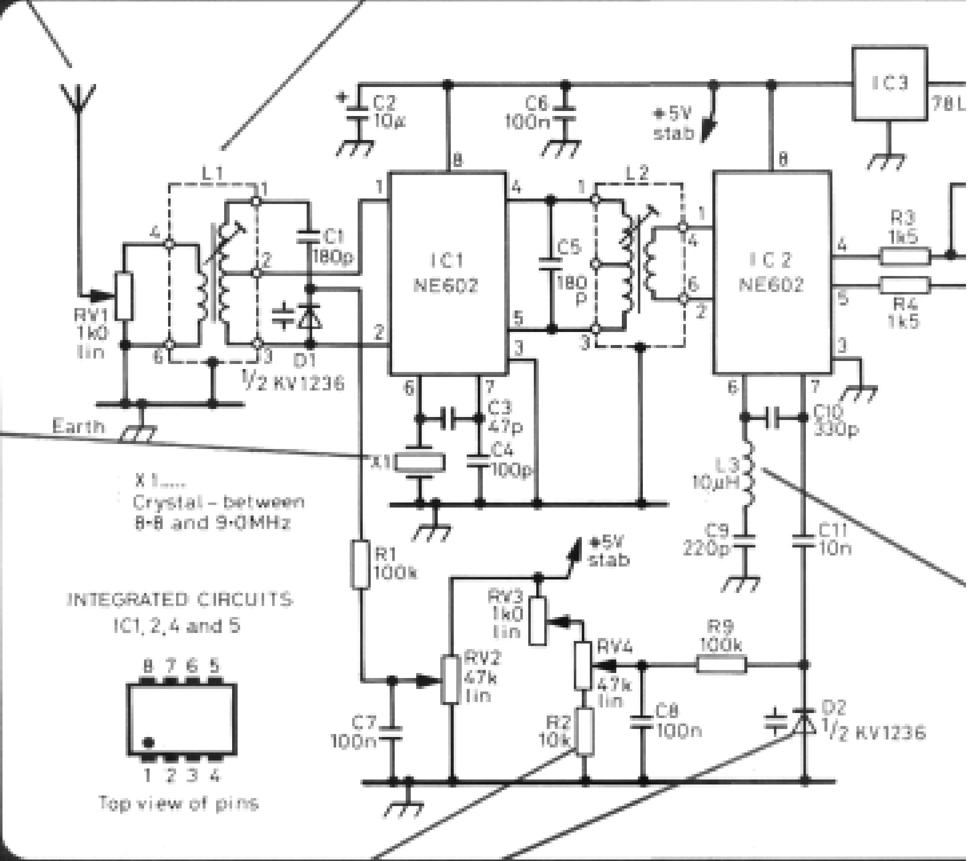
L1
RV1
1k0
lin
C1
180p
D1
1/2 KV1236
Earth
C2
10μ
C6
100n
+5V
stab
IC1
NE602
C5
180
P
L2
IC 2
NE602
IC3
78L
R3
1k5
R4
1k5
C3
47p
C4
100p
X1
C10
330p
L3
10μH
C9
220p
C11
10n
R1
100k
RV3
1k0
lin
RV4
47k
lin
R9
100k
RV2
47k
lin
C7
100n
R2
10k
C8
100n
D2
1/2 KV1236
X1.....
Crystal - between
8·8 and 9·0MHz
INTEGRATED CIRCUITS
IC1, 2, 4 and 5
8 7 6 5
1 2 3 4
Top view of pins

Resistor

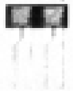

Varicap Diode

Most radio amateurs enjoy building part
is Do-it-Yourself (D-i-Y) Radio, and it is
successfully built yourself. Licensed
allowed (by the Government) to use ho
can see how radio components are dra
circuit diagram for the

D-i-Y Radio, RSGB, Lambda House, Cra

– Do It Yourself

RADIO

rling

Polyester Capacitor

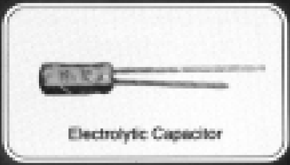
Electrolytic Capacitor

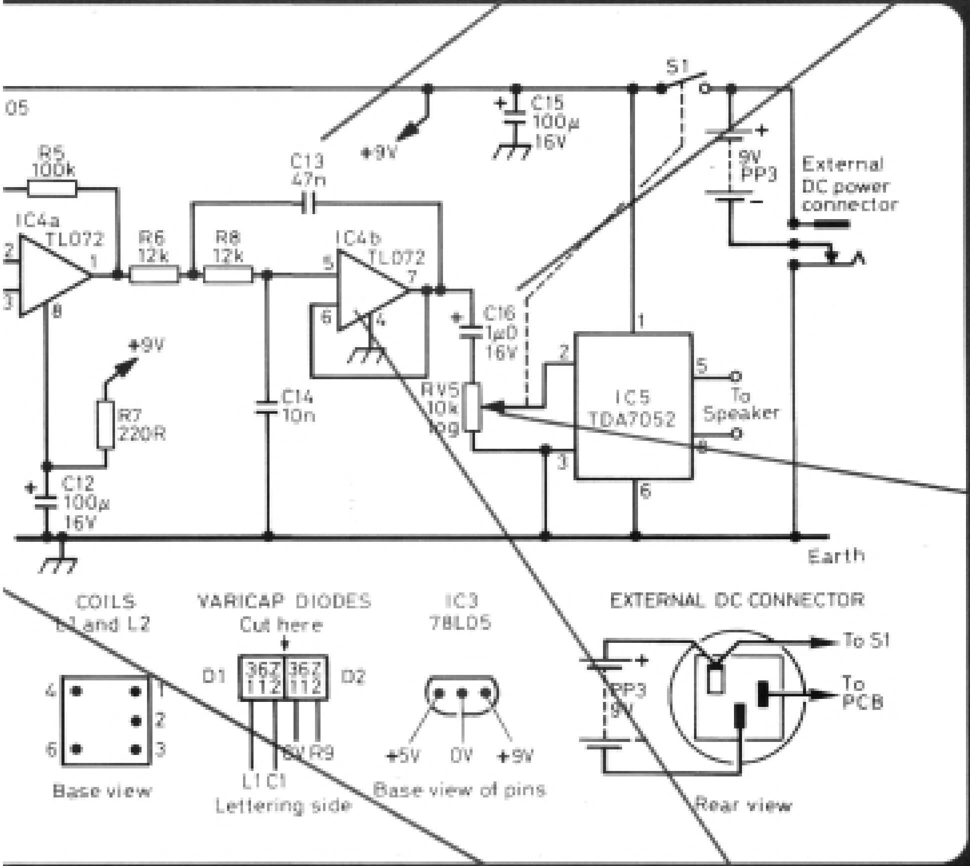
05
R5
100k
IC4a
TL072
2
3
1
8
R6
12k
R8
12k
C13
47n
+9V
C15
100µ
16V
S1
+
9V
PP3
–
External
DC power
connector
IC4b
TL072
5
6
7
4
C16
1µ0
16V
+9V
R7
220R
C14
10n
RV5
10k
log
1
2
3
5
8
6
IC5
TDA7052
To
Speaker
C12
100µ
16V
Earth
COILS
L1 and L2
4
1
2
6
3
Base view
VARICAP DIODES
Cut here
D1
367
112
367
112
D2
0V
R9
L1
C1
Lettering side
IC3
78L05
+5V
0V
+9V
Base view of pins
EXTERNAL DC CONNECTOR
+
PP3
9V
To S1
To
PCB
Rear view

AF Gain Control

of their radio stations themselves, this
very satisfying to use a radio you have
d radio amateurs are the only people
me-built transmitters. On this page you
wn on a circuit diagram. This is the full
D-i-Y Radio Yearling.

nborne Road, Potters Bar, Herts. EN6 3JE

Fixed Inductor

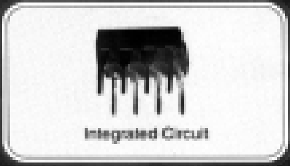
Integrated Circuit

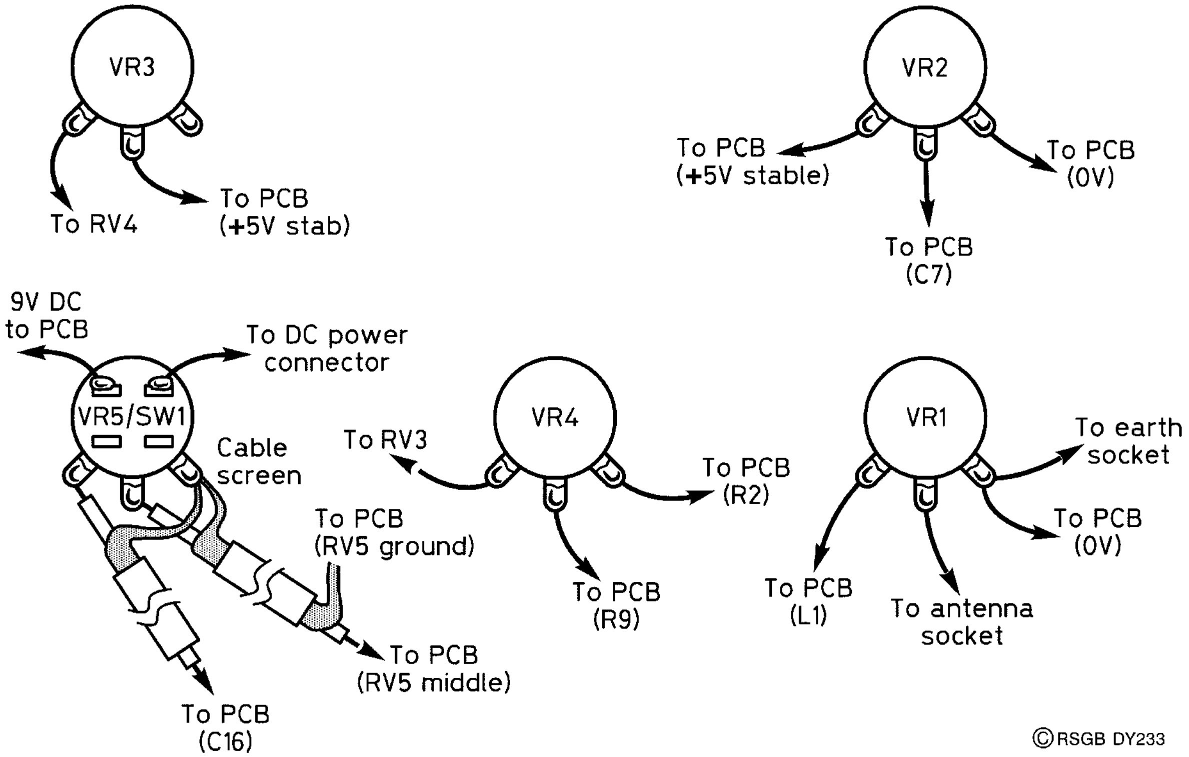
VR3
To RV4
To PCB
(+5V stab)
VR2
To PCB
(+5V stable)
To PCB
(0V)
To PCB
(C7)
9V DC
to PCB
To DC power
connector
VR5/SW1
Cable
screen
To PCB
(RV5 ground)
To PCB
(RV5 middle)
To PCB
(C16)
VR4
To RV3
To PCB
(R2)
To PCB
(R9)
VR1
To earth
socket
To PCB
(0V)
To PCB
(L1)
To antenna
socket
©RSGB DY233

Figure 1 Rear view of the variable resistors. Check the connections carefully to make sure the wires fit the correct holes

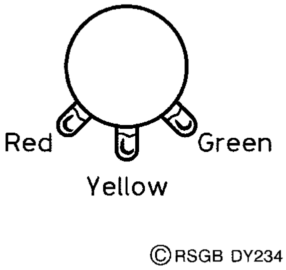
Red
Green
Yellow
©RSGB DY234

Figure 2 It is helpful to use different colour insulated connecting wires. Wires between each variable resistor and the board should be twisted together to give a neat wiring outfit.

board end. Now fit the ICs into their sockets. Make sure that they are the right way round (see circuit diagram) and that each pin lies directly above its corresponding socket before applying *gentle* pressure with the back of the board firmly supported.

Check that all the connections are correct (don’t *assume* this – check the ends of each wire against the circuit diagram) and that all your soldered joints are shiny. Lastly, drill the five 10.5 mm diameter holes for the main controls. On the side of the case, a 6.3 mm diameter hole is needed for the speaker socket, and 8 mm holes for the antenna and earth sockets. The external power socket requires an 11 mm diameter hole.

Adjusting it

Before you can make adjustments, and in order to hear *anything* on your *Yearling*, you will need to connect an antenna (aerial) to the antenna socket. About 8 metres of wire, preferably outdoors and as high as possible, is all you need to connect to the socket. Connect the 3.5 mm jack socket for the speaker, and a 3.5 mm jack to your speaker leads. Connect a battery and switch on.

1. Using a very small screwdriver or, better still, a non-metallic ‘trimming tool’, *gently* screw in the core of L2 as far as it will go, but *don’t force it*. Then unscrew it by three turns anticlockwise.
2. Set VR1, VR2 and VR4 to mid-position and rotate the core of L1 anticlockwise until the hissing noise you hear reaches a maximum intensity. Then adjust L2 for maximum noise.
3. If you now tune carefully with the main tuning control, VR4, you should hear some amateur Single-Sideband (SSB) speech signals. You may have to adjust the bandspread (fine tuning) control, VR3, to make the speech sound normal.
4. Having verified that everything is working, switch off and mount the controls on the front panel and the sockets on the side. To do this, it is much safer to disconnect all the controls and sockets, mount them in their final positions, and then wire them up again.
5. Fit the front panel knobs, connect your aerial and switch on again, checking that everything is working. Then, locate the cluster of CW (Morse) signals to be found at the bottom of the 20 metre band, set the bandspread control to mid-position and slacken off the main tuning knob. Turn the knob (but *not* the control!) until the pointer lies a little clockwise of 14.0 MHz. The SSB signals should now lie roughly between the dial centre and 14.35 MHz. Tighten up the knob.
6. Finally, fix the board to the rear panel, and secure the battery (if you are using one). Attach the rear panel to the back of the box, and you are finished!

Listening!

Make a habit of keeping the bandspread control in its centre position when searching for stations; then you can adjust it either way to make the signals readable. Unless you have a very big aerial, it is best to have the 'RF Gain' control, VR1, at maximum. Use the 'Antenna Tune' control, VR2, to give the best signal, and control the volume with the 'OFF/AF Gain' control. You will find some excellent DX stations with your *Yearling* receiver, and it will serve you well.

Parts list

Resistors (all 0.25 W, 5%)

R1, R5, R9	100 kilohms (kΩ)
R2	10 kilohms (kΩ)
R3, R4	1.5 kilohms (kΩ)
R6, R8	12 kilohms (kΩ)
R7	200 ohms (Ω)

Capacitors (all rated at 16 V or more, tolerance *at least* what is quoted)

C1, C5	180 picofarads (pF) polystyrene 5%
C2	10 microfarads (μF) electrolytic
C3	47 picofarads (pF) polystyrene 5%
C4	100 picofarads (pF) polystyrene 5%
C6, C7, C8	100 nanofarads (nF) or 0.1 microfarad (μF) ceramic
C9	220 picofarads (pF) polystyrene 2%
C10	330 picofarads (pF) polystyrene 2%
C11, C14	10 nanofarads (nF) or 0.01 microfarad (μF) ceramic
C12, C15	100 microfarads (μF) electrolytic
C13	47 nanofarads (nF) or 0.047 microfarad (μF) polyester, 5%
C16	1 microfarad (μF) electrolytic

Variable resistors

VR1, VR3	1 kilohm (kΩ) linear
VR2, VR4	47 kilohms (kΩ) linear
VR5	10 kilohms (kΩ) log with switch

Inductors

L1	Toko KANK3335R
L2	Toko KANK3334R
L3	10 microhenries (μH), 5%

Semiconductors

IC1, IC2	Philips/Signetics NE602 or NE602A
IC3	78L05 5 V 100 mA regulator
IC4	TL072 Dual Op-Amp
IC5	Philips TDA7052 audio amplifier

Additional items

D1, D2	Varactor diode Toko KV1236 (cut into two sections – see text)
X1	Crystal 8.86 MHz type (from Maplin, etc.)
4 off	8-pin DIL sockets for IC1, IC2, IC4 and IC5
2 off	4 mm sockets aerial (red) and earth (black)
1 off	3.5 mm chassis-mounting speaker jack socket
1 off	DC power socket for external supply (if required)
4 off	Red knobs with pointers
1 off	Tuning knob with pointer (e.g. 37 mm PK3 type)
1 off	Printed-circuit board or prototype board
1 off	Plastic case approx 170 × 110 × 6 mm (e.g. Tandy number 270-224)
1 off	Speaker 8–32 Ω impedance (or headphones)

71 Adding the 80 metre band to the Yearling receiver

Background

You will have noticed that your *Yearling* receiver has a dial which shows coverage of the 80-metre amateur band (3.5–3.8 MHz). This band is used for local contacts during the day, and contacts up to about 1600 miles in darkness. Longer distances are possible, particularly in the middle of winter.

The modifications

Only a few extra parts are required, as you may have noticed from the parts list. A low-pass filter, FL1 (one which passes low frequencies and rejects higher frequencies), is switched into the circuit on 80 m. The circuit of the

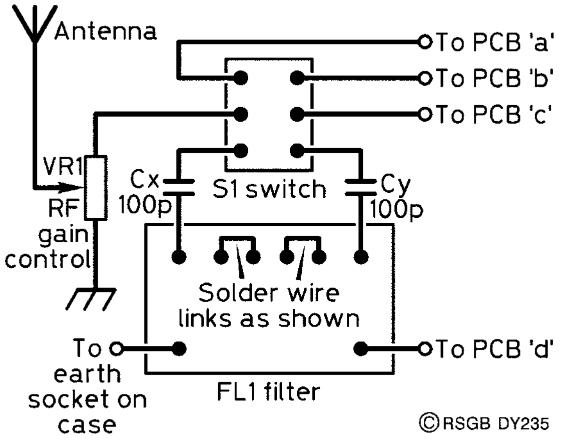
Antenna
To PCB 'a'
To PCB 'b'
To PCB 'c'
VR1
RF gain control
Cx 100p
S1 switch
Cy 100p
Solder wire links as shown
To earth socket on case
To PCB 'd'
FL1 filter
©RSGB DY235

Figure 1 The circuit diagram shows the extra components for 80 m operation. Note the connections to the PCB 'a', 'b', 'c' and 'd'

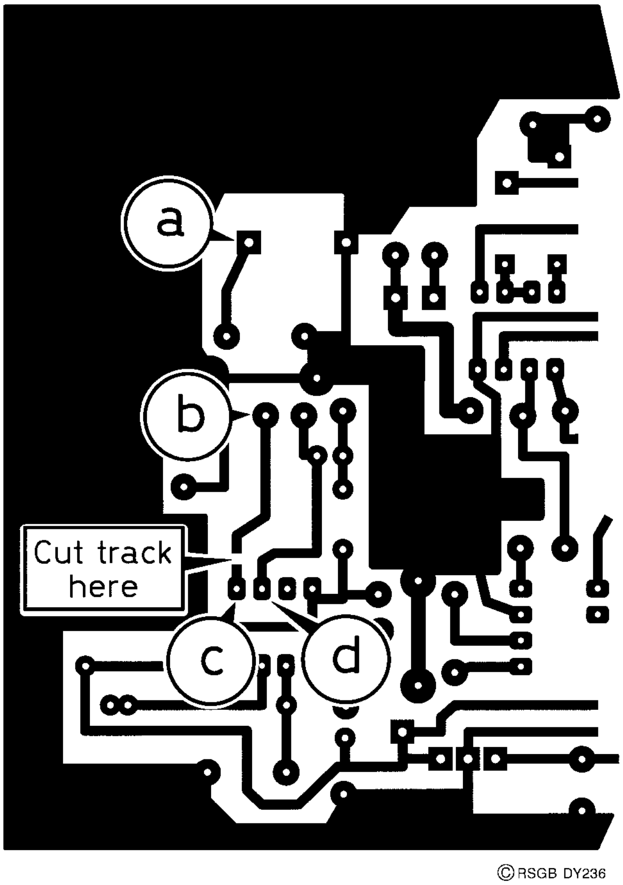
a
b
Cut track here
c
d
©RSGB DY236

Figure 2 The underside of the PCB. Wires are connected from the switch and filter as shown

switch and its connections is shown in **Figure 1**. Before making the modifications, dismantle the receiver so that you have easy (and safe) access to the case and the track side of the PCB.

1. The first thing to do is to drill a 6.5 mm diameter hole in the side of the case into which the switch fits.
2. Then, using a sharp Stanley knife or scalpel, carefully cut the track on the PCB as shown in **Figure 2**, making a gap about 1 mm wide.
3. Using 10 cm lengths of different-coloured insulated wire, make the four connections, a, b, c and d, to the PCB, as shown in Figure 2.
4. Solder the two links on the filter, and then make the connections to the switch, capacitors and PCB. You will have to disconnect the existing wire between the RF gain control, VR1, and pin 4 of L1 on the PCB.
5. Lastly, check your new connections carefully, then mount the switch in the new hole and fix the filter to the bottom of the case with a little glue, as shown in **Figure 3**. Reassemble the circuit, and replace the back of the case.

More testing!

Firstly, switch your new switch, S1, into the 20 m position, to check that the original circuit still works! If you find that the 'Antenna Tune' control peaks at a slightly different position, don't worry.

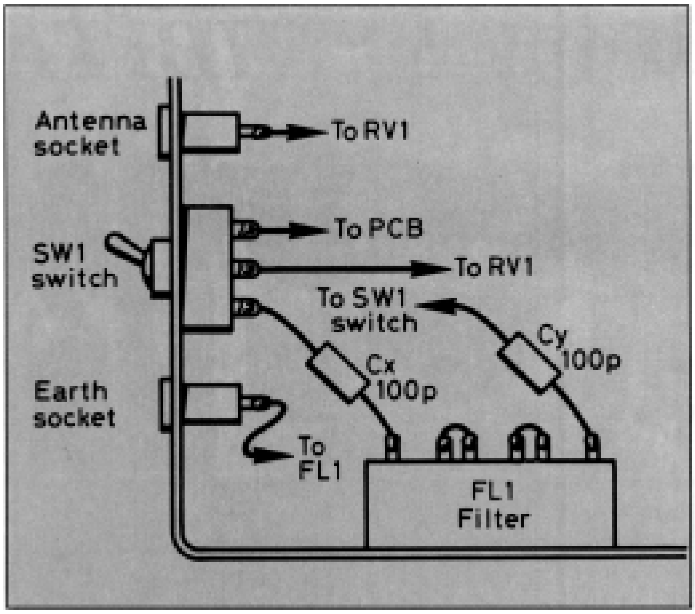

Figure 3 Internal view of the Yearling case. The filter FL1 is attached to the base with glue

Now switch to 80 m, and tune around the anticlockwise end of the dial; you should hear some SSB stations, particularly in the evenings and at weekends, when may people are on the air. At the other end of the travel of the tuning control, you should hear CW (Morse) stations.

The Radio Society of Great Britain broadcasts amateur radio news every Sunday morning on or about 3.65 MHz; the table below has the details. Finally, a good antenna is more important than ever for 80 metre reception – aim for more height and length, and then consider the project concerned with making an Antenna Tuning Unit (ATU)!

Parts list

Capacitors (all rated at 16 V or more, tolerance 10% or better)

Cx, Cy	100 picofarads (pF) polystyrene

Filter

FL1	Toko 237LVS1110 low-pass filter

Additional items

SW1	2-pole 2-way (changeover) toggle switch
7 off	Short lengths of insulated wire of different colours

The RSGB news broadcasts, GB2RS – Sunday mornings

Frequency (MHz)	*Local time*	*Reception area*
3.650	0900	SE England
3.650	0930	Midlands
3.650	1000	SW England
3.650	1100	Yorkshire
3.640	1130	Aberdeen
3.660	1130	Glasgow

The Midlands transmission is repeated at 1800 (6pm) local time on 3.650 MHz. All frequencies are approximate in order to avoid interference, and use lower sideband (LSB). If you also have a 40 metre receiver, there are GB2RS news broadcasts on 7.048 MHz at 0900 local time from Northern Ireland and from 1100 local time from the north midlands.

72 How the Yearling works

Introduction

The *Yearling* was designed to provide an introduction to *Amateur Radio* on the 20 m amateur band. Let's look at how the different sections (or 'stages') of a radio work, and how they fit together to form a complete receiver. **Figure 1** shows a block diagram which you can follow and compare with the circuit diagram of your *Yearling* receiver.

The antenna (or aerial)

Connected to your receiver, it will pick up not only amateur signals, but all other signals as well! This means that the receiver has to select the one signal that interests you, while rejecting all the others. The following stages do just that.

The RF filter

This stage (centred around L1) selects the band of radio frequencies (RF) containing the signal you want, in this case, those having wavelengths around 20 m. Signals from the 40 m band, for example, would not get through.

The crystal oscillator

This is an oscillator circuit designed around a quartz crystal (X1), and has a very stable frequency. It produces a single, very pure frequency to feed into the mixer. A crystal having a frequency between 8.800 MHz and 9.000 MHz is suitable for this circuit. The oscillator and mixer functions are both carried out inside IC1.

The first mixer

Yes, this stage 'mixes' two signals together. In this case, the two signals are (i) from the aerial via the RF Filter, and (ii) from the crystal oscillator. Two

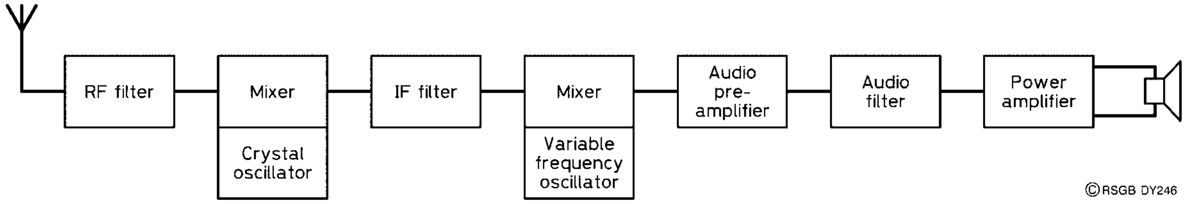

Figure 1 Block diagram of the Yearling, showing how the various stages fit together to make a complete radio receiver

bands of signals emerge from the mixer. The first is centred upon a frequency equal to the incoming signal frequency **added** to the crystal frequency, and the second is centred upon a frequency equal to the incoming signal frequency **subtracted** from the crystal frequency. Look at an example – if the signal is at 14 MHz and the oscillator at 9 MHz, then the mixer outputs will be 14 + 9 = 23 MHz and 14 – 9 = 5 MHz.

Intermediate frequency (IF) filter

It is the purpose of the IF filter (centred around L2) to select only one of these two bands of frequencies emerging from the mixer. In this case, it is the lower band of frequencies (around 5 MHz) which we select. This is because, in general, lower frequencies are easier to handle than higher ones.

Variable-frequency oscillator (VFO)

The VFO (part of IC2) enables us to tune into a particular station, and operates over a band of frequencies between 5 MHz and 5.35 MHz in this receiver. You will notice that IC1 and IC2 are the same type of chip, so that you will be expecting another mixer stage to be associated with the VFO. You are quite right!

The second mixer

This mixer obeys exactly the same rules as those of mixer 1. Sum and difference frequencies are produced, like this. Mixing is between the incoming IF signals (around 5 MHz) and the VFO signals (around 5 MHz), producing output frequency bands centred upon 10 MHz and 0 MHz. The use of the words ‘band of frequencies’ throughout this explanation is intentional. If all the signals were pure, there would be no bands; the bands are produced because of one thing – the modulation imposed on the pure frequencies at the transmitter. So, the ‘bands’ contain the one thing that we want to extract from the signal, and that is the speech or Morse code that the signal contains. The band of frequencies at 0 MHz is just that – the audio frequencies we want in the loudspeaker. Because of this, the audio frequency output of the second mixer is selected and passed on for amplification.

The audio preamplifier

Preliminary amplification of the minute audio signal which emerges from the second mixer is provided by IC4a, which will respond only to audio signals, automatically rejecting the 10 MHz signal.

The audio filter

The bandwidth of normal speech when transmitted by an amateur station is around 3 kHz, so there is no advantage to be gained in amplifying frequencies greater than this. IC4b is known as a low-pass filter, because it passes (lets through) lower frequencies and rejects higher ones.

The power amplifier

IC5 produces the final audio amplification and provides enough power (about 350 milliwatts (mW) to drive a small speaker.

How does it work on 80 m?

If you have fitted the 80 m modification to your receiver, you are probably wondering how the circuit works at this different frequency. Firstly, the filter which you fitted selects the 80 m band instead of the 20 m band. The only other slight difference lies in the way the first mixer stage works. Its job is to produce sum and difference frequencies from the incoming signal and crystal frequencies. On 20 m, it did this by subtracting the crystal frequency (9 MHz) from the incoming frequency (14 MHz) to produce an IF output of 5 MHz. On 80 m, the incoming frequency (3.5 MHz) is subtracted from the crystal frequency (9 MHz) to produce an IF output of 5.5 MHz, which is still within the tuning range of the VFO in the next stage.

73 A field strength meter

Introduction

Having a reliable field strength meter is always useful. There are two basic types – the tuned variety (also known as an absorption wavemeter) and the broad-band variety. A project centred on the tuned variety can be found in this book (*An HF absorption wavemeter*), and the way it works is also presented (*The absorption wavemeter*). This project describes an untuned or broad-band type which is every bit as useful as the tuned type, when it is the level of RF in which you are principally interested.

With the possible exception of the meter, you may have all the parts needed in your junk box. The prototype was distinguished by its meter, found at a rally, and marked 'Safe', 'Dangerous' and 'Explosive' on its scale.

How it works

The basic field strength meter uses the circuit of a crystal set, but with a meter replacing the headphones. A better design, which is used here, is that of a voltage doubler, giving more sensitivity. **Figure 1** illustrates the voltage doubler circuit.

The AC input shown will be our RF input, which will be explained soon. Diode D1 will pass the positive half of the signal and use it to charge up C1 to the peak value of the signal. D2 uses the negative half of the signal to charge up C2 to the same value. Because C1 and C2 are in series, the peak voltage appearing across *both* of them (which is the DC output voltage) is equal to *twice* the peak input voltage, hence the name voltage doubler. If you're wondering why DC is present at the output when AC comes in at the

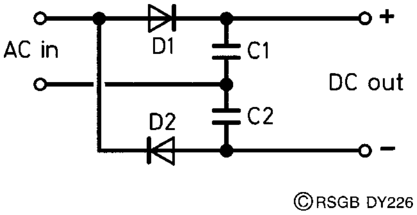

Figure 1 Voltage doubler circuit

input, remember that capacitors pass AC (RF) but block DC; thus the RF is shorted out through C1 and C2, but the steady voltage (DC) remains across the two capacitors. The voltage is thus proportional to the size of the RF signal applied at the input.

The circuit

The voltage doubler is converted into a field strength meter using the circuit of **Figure 2**. A piece of wire serves as the aerial to provide an RF signal across the radio-frequency choke (RFC). A choke is an inductor which is large enough to prevent the RF passing through it – it 'chokes' the RF. This produces the maximum RF signal at the input to the voltage doubler, and the DC output from it is measured on the meter.

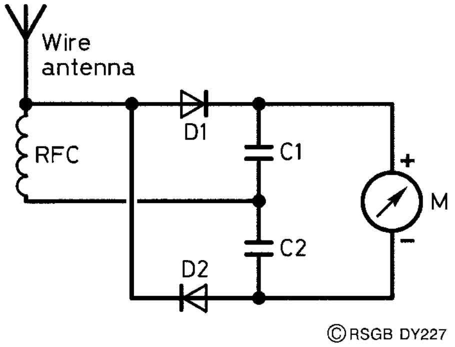

©RSGB DY227

Figure 2 Field strength meter

The parts used are not very critical. The RF choke can have any value between about 1 mH and 2.5 mH. Almost any common diode such as the 1N914 or 1N4148 can be used for D1 and D2. The two capacitors could be any value between 1 nF and 100 nF (0.001 μF and 0.1 μF). The meter should be reasonably sensitive, with a full-scale deflection (FSD) in the range 50 μA to 100 μA. Look for VU meters at rallies – these are ideal.

Construction

The prototype circuit was made on matrix board. If you don't want to use pins with the board, simply push the component leads through the holes and make the connections on the underside of the board, either with the excess component leads themselves, or with ordinary connecting wire.

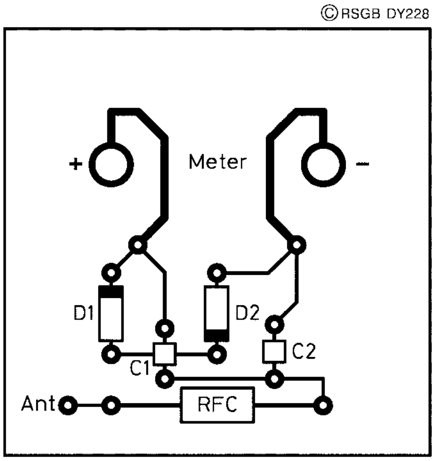

Figure 3 Board layout and interconnections

Component side of board shown.
All connections are under board

Figure 3 shows how the parts are placed and connected. The matrix board can be mounted directly on the meter using its terminal bolts!

In use

Needing no power supply other than an RF signal, just connect it all up and leave it to work! A short length of insulated wire is enough to pick up some RF and display it on the meter. Using a 200 μA meter, about 3 metres of wire gave a good deflection on the meter. To increase deflection, the wire can be wrapped around the aerial lead, provided that the wire is PVC covered and doesn't come into contact with the aerial wire.

This requires a little experimentation. Try a long piece of wire first and adjust its position until the meter needle kicks whenever there is a transmission. It is very reassuring to see the meter moving during a transmission. Although SWR meters also indicate power, they are usually set to read reverse power, and show little or no movement during transmission.

Parts list

D1, D2	1N914, 1N4148 or similar
C1, C2	10 nF disc ceramic
RFC	Miniature axial choke (1 mH)
Meter	Surplus VU meter or similar
Matrix board or similar	

Components are available from Maplin.

74 Preselector for a short-wave receiver

Introduction

A preselector is a simple RF tuned amplifier which is inserted between the aerial and the receiver. It provides some extra gain and may improve the overall performance of the receiver. This project uses a Field-Effect Transistor (FET) amplifier in grounded-gate mode.

The design has a tuned circuit at both the input and output which, with excessive gain and poor construction, would produce only one thing – oscillation! So, to avoid this happening, we will have only a low gain, and use a circuit which provides good isolation between input and output. The grounded-gate FET amplifier fulfils both these criteria. It will also cover a frequency range from about 7 to 30 MHz, which includes most of the HF amateur bands.

The circuit

This is shown in **Figure 1**. The signal from the aerial arrives at an RF transformer, the secondary of which is tuned with capacitor VC1a. The output from the tuned circuit is taken from a tap on the secondary to the source of the FET. The gate is grounded (earthed) and the amplified signal appears at the FET drain, which is then fed to the primary of another RF transformer, which is tuned by VC1b. The output to the receiver comes from the secondary of the RF transformer.

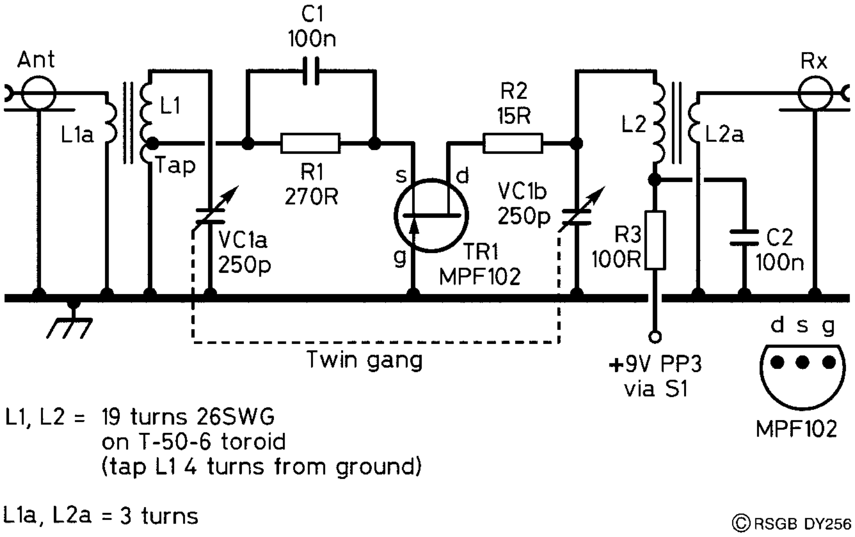
C1
100n
Ant
Rx
L1
L1a
Tap
R1
270R
R2
15R
L2
L2a
s
d
g
VC1a
250p
VC1b
250p
TR1
MPF102
R3
100R
C2
100n
Twin gang
+9V PP3
via S1
d s g
MPF102
L1, L2 = 19 turns 26SWG
on T-50-6 toroid
(tap L1 4 turns from ground)
L1a, L2a = 3 turns
©RSGB DY256

Figure 1 HF preselector, circuit diagram

Notice that the two RF transformers are identical, but they are used 'back to back', with the secondary of the first and the primary of the second being tuned. They are tuned with identical capacitors, fitted on the same shaft of a variable capacitor. We say that the two capacitors are 'ganged'. Because L1 and L2 are the same, and VC1a and VC1b are the same, both RF transformers should be resonant at the same frequency, no matter what that frequency is.

Construction

The final layout should look something like that shown in **Figure 2**. The external connectors and controls being two SO-239 sockets for connection to your receiver and aerial, a tuning control and its associated scale, and an on/off switch.

The circuit can be put together on a plain matrix board, using pins to anchor the components, or simply by pushing the component leads through the board and making connections on the underside. The layout of the prototype is shown in **Figure 3**. Mounting the board to the aluminium box is accomplished with bolts, solder tags and stand-off insulators.

Check your construction against the circuit diagram and against the layout diagram. Wire in the PP3 battery clip, put the switch in the 'off' position, and fit the battery. Testing can be carried out without fitting the top of the box.

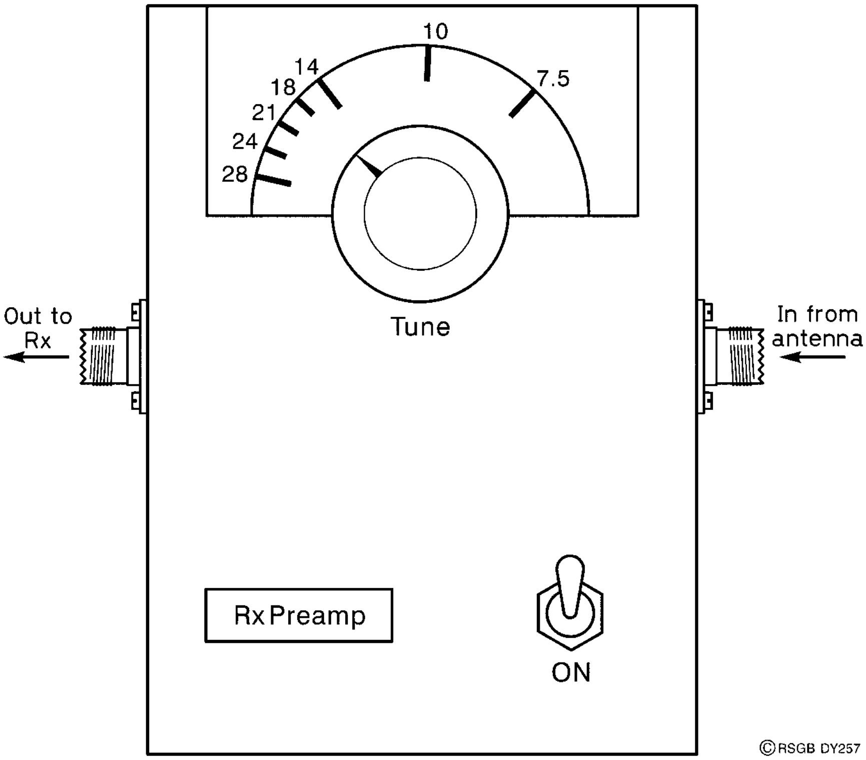

Figure 2 HF preselector, front panel layout

Testing

Don't fit your preselector yet. Tune your radio to a broadcast station, preferably a fairly weak one. Disconnect the aerial and fit your preselector between the aerial and receiver. Switch it on and rotate the tuning knob slowly. You should find a position where your original station is received more clearly than before. If it doesn't work at all, recheck your wiring. Is there a positive voltage on the drain of the FET? If not, work back towards the positive battery terminal. Is there a voltage at the junction of L2 and R2? Is there a voltage at the junction of L2 and R3? Is there a voltage at the junction of R3 and the battery lead? If there isn't a voltage at that point, then you have probably mounted your switch upside down, and it is off, not on! It's a common mistake.

Calibration

This is not obligatory, surprisingly enough. However, if having a frequency scale appeals to you, then using an RF signal generator (or using the services of a friend who has one) is the simplest solution. Feed in a weak modulated

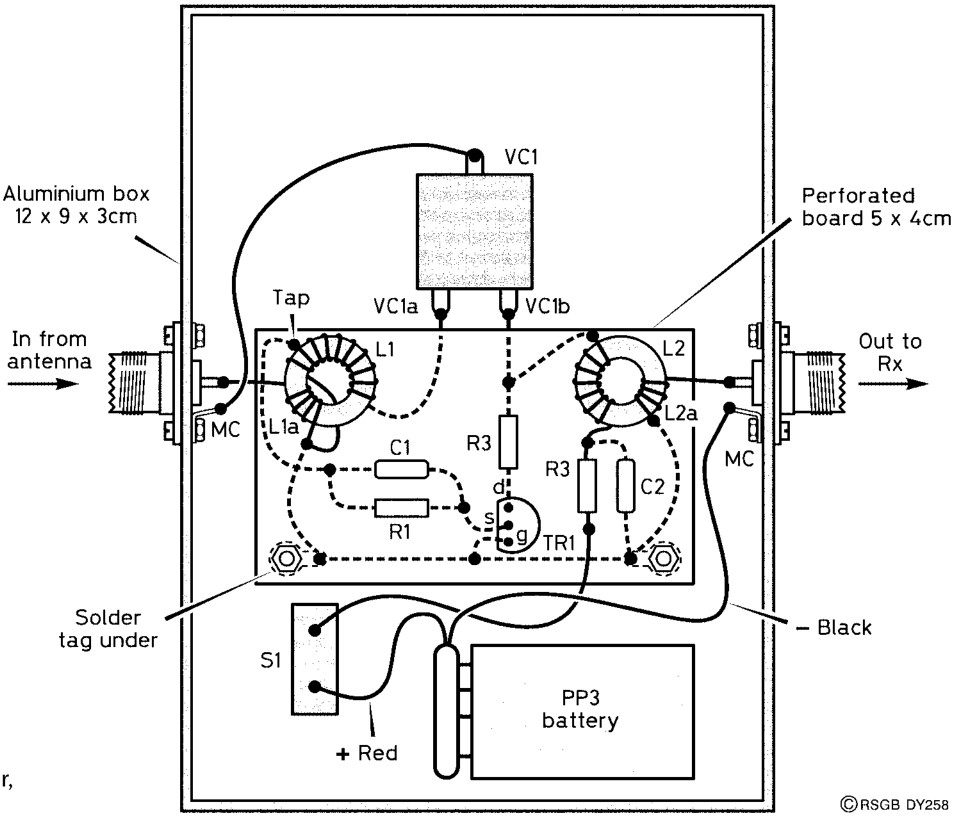
VC1
Aluminium box
12 x 9 x 3cm
Perforated
board 5 x 4cm
Tap
VC1a
VC1b
In from
antenna
L1
L2
Out to
Rx
MC
L1a
L2a
C1
R3
R3
C2
MC
d
s
g
R1
TR1
Solder
tag under
– Black
S1
PP3
battery
+ Red
©RSGB DY258

Figure 3 HF preselector, component layout

signal from the generator to the preselector, and rotate VC1 until the signal is maximum. Mark this frequency on your dial. Repeat the process for the frequencies shown on the dial in Figure 2.

If you have a commercial transceiver, feed its output into a dummy load (such as the type described in *A Switched Dummy Load*, elsewhere in this book) using a distinctive modulating signal such as an idling RTTY signal. Set the receiver to the same frequency with the preselector out of circuit. Insert it into the aerial lead, and search for the signal with VC1 until it gives the maximum deflection on your S-meter, then mark the frequency on your dial.

Parts list

Resistors: all 0.25 W carbon film or better

R1	270 Ω
R2	15 Ω
R3	100 Ω

Capacitors	
C1, C2	100 nF (0.1 μF) ceramic
VC1	250/250 pF polyvaricon
Inductors	
L1, L2	19 turns 26 SWG enamelled copper on T.50.6 toroid L1 has a tap 4 turns from ground end L1a and L2a – 3 turns wound over previous winding
Semiconductors	
TR1	MPF102 FET
Additional items	
Matrix board	To fit aluminium box (see Figure 3)
Aluminium box	12 × 9 × 3 cm
Battery and connector	PP3 9 V
S1	SPST on/off switch
Knob	As required
SO239	Sockets – 2 required

75 An audible continuity tester

Introduction

This is not the only continuity tester in this book. This alone attests to their use, so you may well want to experiment with several designs, then come up with one of your own! The very simplest form of continuity tester is probably a battery and a bulb in series, with the circuit being closed by connecting it to a fuse or other object being tested for continuity. The bulb could be replaced by a buzzer to give an audible indication. The current taken by the buzzer could damage some components, however. An ohmmeter can also be used, and is very popular for the purpose, as it indicates whether the circuit is low, medium or high resistance. This project has the advantages of indicating whether there is no continuity, some resistance or good continuity, and making an audible sound, so that you don't have to move your eyes from the circuit while making the test.

The circuit

Using only one integrated circuit, four components and a battery, this is a particularly simple circuit, as **Figure 1** shows. IC1 is usually used to flash an LED from a 1.5 V source, and to have a low current consumption. By changing the component values, IC1 is made to oscillate at audio frequencies, and we can hear these through the loudspeaker, LS1. Low-impedance speakers can be used, but result in an increased current drain, so the higher the impedance, the better.

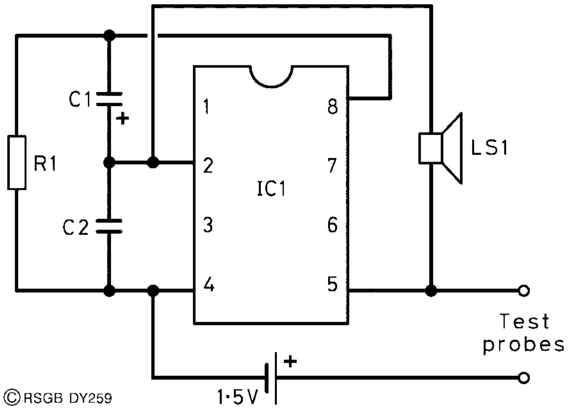

Figure 1 Continuity tester, circuit diagram

Construction

The layout of the circuit on Veroboard measuring 7 strips by 15 holes is shown in **Figure 2**. Start by making the four track cuts which will lie underneath the IC. Use a track cutter or a 3 mm ($\frac{1}{8}$ inch) twist drill rotated between thumb and forefinger for this. Solder the Veropins in place, followed by the wire link. Then solder in the IC holder, the resistor and the capacitors, making sure that the electrolytic capacitor, C1, is connected the correct way round.

Check, with the board against a bright light, that there are no shorted tracks, either by large blobs of solder or by copper swarf from the track-cutting process. Then insert IC1 into its holder the right way round. The probes can be ordinary connecting wire, the free ends being tinned with solder to prevent wire whiskers from touching components other than the one you are testing.

If you are happy that the circuit and the wiring appear to be correct, fit the battery into its clip. Nothing should happen until the probes are touched together, when you should hear a note from the loudspeaker. If nothing happens, all you can do is to recheck your circuit, as there is nothing else to go wrong!

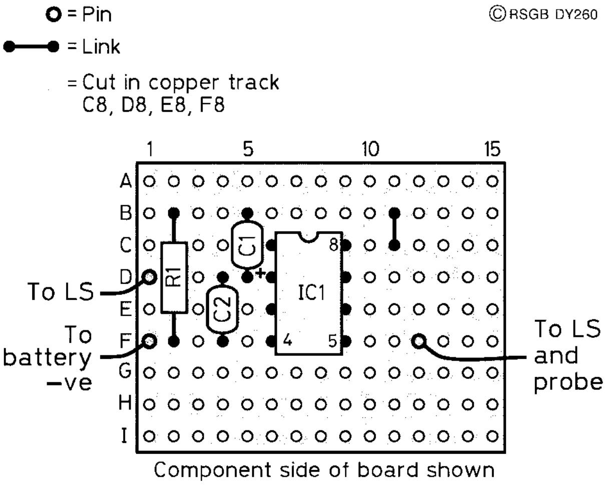
○ = Pin
●—● = Link
= Cut in copper track C8, D8, E8, F8
©RSGB DY260
1
5
10
15
A
B
C
D
E
F
G
H
I
R1
C1
C2
IC1
8
4
5
+
To LS
To battery −ve
To LS and probe
Component side of board shown

Figure 2 Continuity tester, component layout

Any box can be used; there is no justification for a metal case, unless you want to make use of something, which is to hand, such as a tobacco tin. Alternatively, any suitable plastic box will do.

Use

The tester will give different pitch notes for resistances of different values. The higher the resistance, the higher the pitch from the speaker. Try it with small inductors, and you will learn to recognise the different tones produced by the IC.

Safety notice

Using the continuity tester on components *in situ* is not advisable, as the results could be misleading. It can be dangerous to make measurements *in situ* on equipment, which is operating. If you must make such tests, always disconnect the equipment form its power source first.

Parts list

Resistor

R1	1000 Ω (1 kΩ), 0.25 W carbon, 10% or better

Capacitors

C1	10 μF electrolytic, 16 V
C2	0.1 μF subminiature polyester or ceramic

Integrated circuits

IC1	LM3909N

Additional items

	Veroboard 7 strips by 15 holes
LS1	64 Ω miniature loudspeaker
	IC holder
	Veropins
	Connecting wire
	Test leads
	1.5 V AA alkaline battery

76 An experimental 70 cm rhombic aerial

Introduction

Most commonly used aerials can be classed as *resonant* or *standing-wave* aerials. There is another class known as *non-resonant* or *travelling-wave*. Resonant aerials, such as the dipole, are narrow-band; this occurs because resonance occurs only over a narrow band of frequencies. Travelling-wave aerials, on the other hand, can operate over a wide band of frequencies.

The *rhombic* is an example of a non-resonant or travelling-wave aerial. It is often employed for fixed commercial and military short-wave radio links. Made with wire, it has a diamond shape when looking down on it from above. The four corners are supported on four masts. It is a very effective aerial, and has good gain, a quality which can be judged from the *polar diagram* shown in **Figure 1**. This has been obtained from a computer program, and so is the perfect shape for a rhombic aerial.

Theory (but only a little)

A polar diagram shows graphically the ability of an aerial to radiate (or receive) more effectively in one direction at the expensive of the radiation in other directions. Figure 1 shows the polar diagrams of two aerials, a simple 70 cm dipole and the rhombic described in this project. The dipole has the

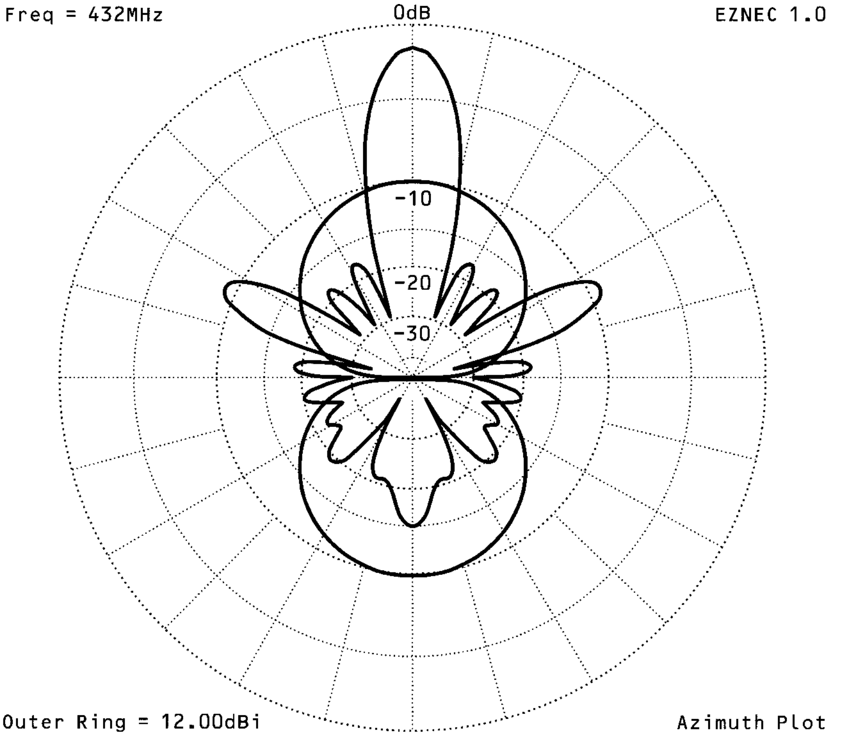

Figure 1 The horizontal polar diagrams of the small rhombic antenna and a dipole compared

©RSGB DY262

well-known figure-of-eight shape, showing a symmetrical angular behaviour about the direction of the aerial and about a direction perpendicular to the aerial. That of the rhombic, on the other hand, is quite irregular by comparison, but is still symmetrical about one axis only, not two. It's this asymmetry that gives the rhombic its gain, by virtue of its *front-to-back ratio*. This is the ratio of the power radiated forwards to that being radiated backwards. Notice the large *lobe* (lump) at the top of the polar diagram; this is the direction in which the aerial transmits best. The lobe in the opposite direction has been reduced significantly, allowing more power to be directed forwards, not backwards

Problems

The rhombic is not found in every amateur's back garden, despite its attractions. To work best each edge of the diamond shape should be about two wavelengths long. Hence, a rhombic for the 20 metre band could be about 80 m from tip to tip! Another disadvantage is that, because of its size, it cannot be rotated.

It will operate well over a range of frequencies. An HF rhombic could work on the 7, 10, 14, 18, 21 and 28 MHz bands. One for the lower VHF frequencies could operate on the 50, 70 and 144 MHz bands.

Its large size is less of a problem at UHF. A portable rhombic can be made for 70 cm which can be used with a hand-held transceiver. The aerial design to be described here will give a gain of up to 9 dB relative to a dipole. This is written as 9 dBd, the second ‘d’ meaning ‘relative to a dipole’. This is equivalent to improving your signal by 1.5 S-points at the receiver or (and you may be surprised by this) by fitting a linear amplifier to your transceiver that would take 5 W input and produce 40 W output to a normal dipole! Consider the relative costs of the two approaches to producing the same received signal. You will also receive everyone else’s signals 1.5 S-points better than before!

Unlike the popular Yagi aerial, this design has no critical dimensions and can be folded up for transport by car or bicycle. It has a fairly high input (or feed) impedance, being fed usually by balanced twin feeder. To feed it with standard 50 Ω coaxial cable, a matching transformer in the form of a *balun* is required. A balun will convert the *bal*anced (symmetrical) aerial impedance to the *un*balanced and lower impedance of the coaxial cable. The details of how to build the balun, which in this case is a half-wave transformer, are included in the constructional details.

Construction

The aerial frame is made up of 1 cm × 2 cm strips of wood fixed to a plywood centre using 30 mm long M4 bolts, as shown in **Figure 2a**. The outer bolts fixing the front and side supports can be removed for folding prior to transportation, as shown in **Figure 2b**.

The wires are fixed to the front and rear supports using screw connectors, sometimes called ‘chocolate block’ connectors. Detail X and Y of Figure 2 show how this is done. The side supports have holes in the ends through which the wire is threaded.

The aerial must be mounted in the horizontal plane using a small shelf bracket attached to the centre plate, using the same bolts that hold the rear support. The other half of the bracket may be mounted to a vertical mast using screws or jubilee clips (hose clips).

The balun

Cut a 23 cm length of coaxial cable; this is to be our half-wave transformer. Cut and remove 2 cm from each end of the sheath. Make the braid into a pigtail at each end. Cut and remove 1 cm of the inner insulator from each

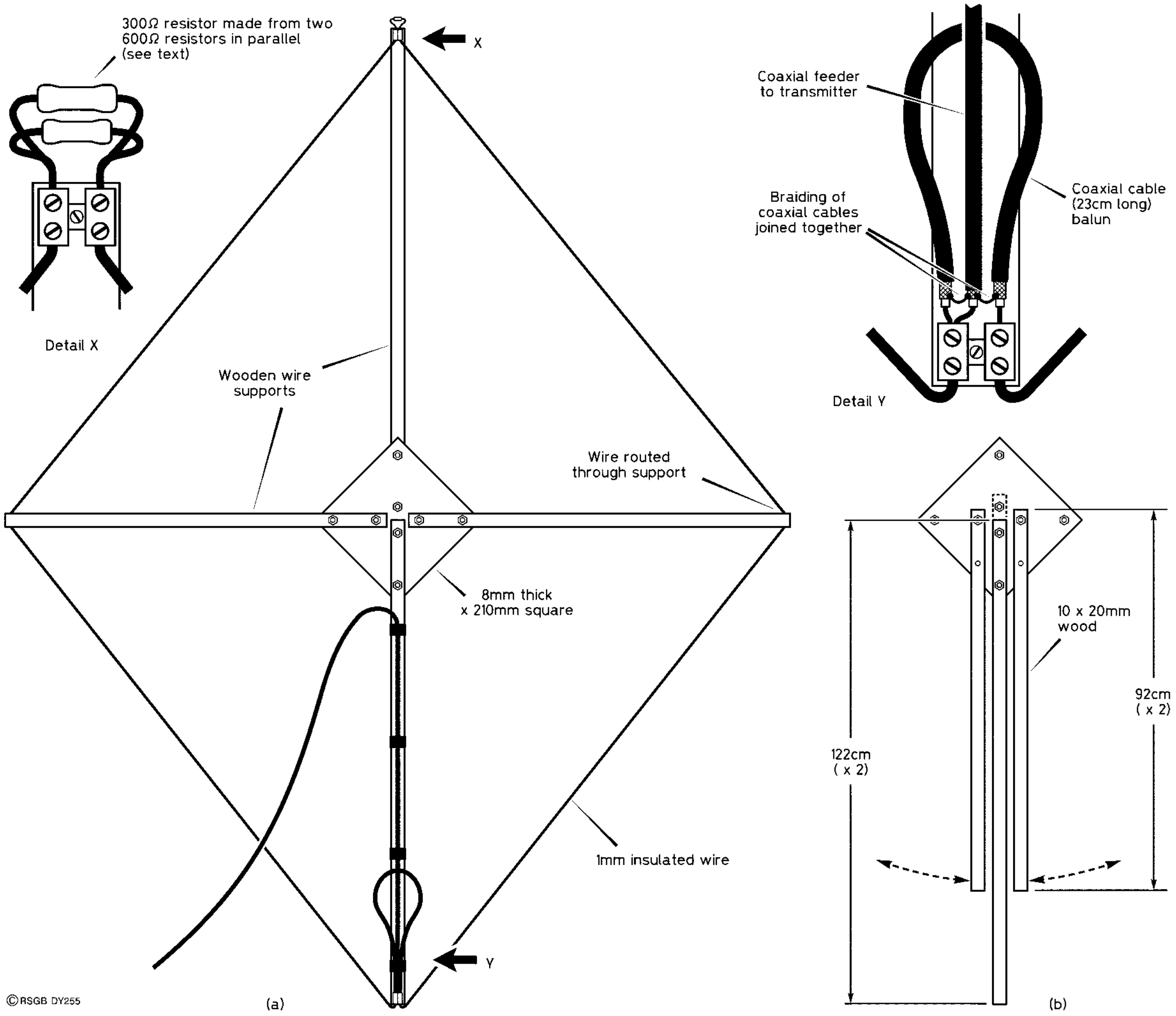
300Ω resistor made from two 600Ω resistors in parallel (see text)
X
Detail X
Wooden wire supports
Coaxial feeder to transmitter
Braiding of coaxial cables joined together
Coaxial cable (23cm long) balun
Detail Y
Wire routed through support
8mm thick x 210mm square
1mm insulated wire
Y
©RSGB DY255
(a)
10 x 20mm wood
92cm (x 2)
122cm (x 2)
(b)

Figure 2 Construction of the UHF rhombic antenna

end, leaving 1 cm of the centre conductor exposed. Fasten this piece of cable to the chocolate block connector together with the coaxial feeder cable as shown in Detail Y of Figure 2. The braids of all three prepared ends are soldered together, but are not connected to anything else. The inner conductor of the feeder from the transceiver is soldered to one end of the centre conductor of the 23 cm piece, and is connected to one end of the rhombic by one side of the chocolate block. The other end of the half-wave transformer also goes to the chocolate block, where it is connected to the other side of the rhombic loop.

The opposite corner of the rhombic, as shown in Detail X of Figure 2, shows that the loop is broken at the chocolate block, and is 'terminated' by two 600 Ω resistors in parallel. This makes the aerial a broad-band travelling-wave device, and gives it its directivity and gain. There is a rule of thumb, which includes a safety factor, which says that the terminating resistor must be able to absorb one-half of the maximum transmitter output power. So, if you use two 2 W 600 Ω carbon resistors (*not* wire-wound resistors) in parallel, you can use a transmitter with an output of 8 W, which is more than adequate for 70 cm hand-helds.

Using the rhombic

Fit the aerial to a pole or mast in the horizontal plane, if you intend to use CW or SSB, but in the vertical plane if you want to concentrate on FM work. In the latter case, a wood or fibreglass pole is mandatory. Tune to a local repeater, whose signal strength you know. Rotate the aerial to face the repeater, and you should see that the signal strength is greatly improved! Verify the directional properties of the aerial by rotating it and observing the changes in signal strength on your S-meter.

77 Water level alarm

Introduction

This is a simple device which, when you have built it, you begin to wonder how you ever managed without it! It will sound a buzzer when the level of water reaches a particular point, which you can set and vary at will.

Detecting water

Water is not too difficult to detect electrically, because it is a conductor of electricity. Not a good one, but sufficiently so to carry enough current to control a warning circuit.

The circuit is shown in **Figure 1**. If you imagine a resistor placed between the ends marked 'water detector', the circuit is then recognisable as a transistor used as an electronic switch. Whether the switch is off or on depends on the value of this resistor. If the resistance is low, the transistor switches on and the buzzer sounds. If the resistance is high, the transistor switches off and the buzzer stops.

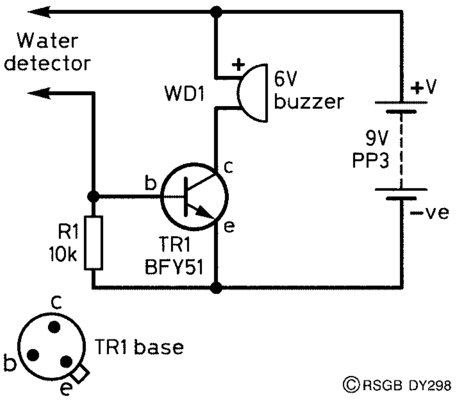
Water
detector
WD1
+
6V
buzzer
+V
9V
PP3
−ve
b
c
e
R1
10k
TR1
BFY51
c
b
e
TR1 base
©RSGB DY298

Figure 1 Circuit diagram of the water level detector

Construction

The unit is made in two separate sections – the detector board (**Figure 2**) and the main board (**Figure 3**). The main board is plain perforated board (without copper strips). The components are pushed through holes in the main board and connected together on the underside. The battery is held on with sticky tape. No on/off switch is used, as the circuit consumes virtually no current when it is not sounding the buzzer.

Unlike some buzzers, this type is polarised, and the red lead must go to the supply as shown in Figs 1 and 3.

The detector board is a piece of standard matrix board – the type with copper strips along one side. It measures 9 strips by 15 holes, and is wired in such a way that alternate strips are connected together, as Figure 2 shows. This gives a greater surface area for the water to touch, thus increasing the sensitivity of detection.

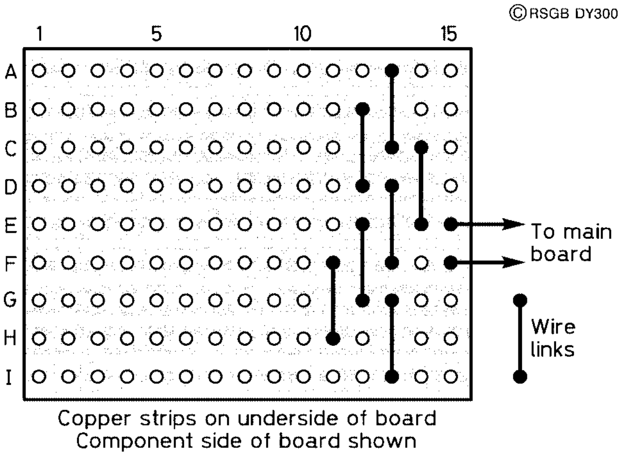
©RSGB DY300
1
5
10
15
A
B
C
D
E
F
G
H
I
To main
board
Wire
links
Copper strips on underside of board
Component side of board shown

Figure 2 Layout of the main board

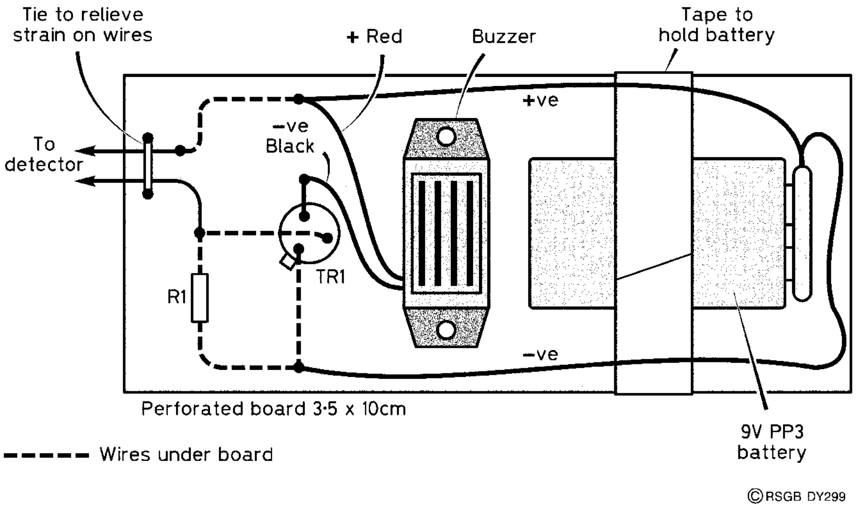
Tie to relieve strain on wires
+ Red
Buzzer
Tape to hold battery
To detector
−ve Black
+ve
TR1
R1
−ve
Perforated board 3·5 x 10cm
9V PP3 battery
----- Wires under board
©RSGB DY299

Figure 3 Water detector constructed from stripboard

The two boards are connected by ordinary twisted pair, enabling the detector to hang over the edge of the water container, with the buzzer circuit safely out of the way. When positioning the detector board, the left-hand edge (as shown in Figure 2) should face downwards, with the connecting-wire edge facing upwards. When the water level rises, the water causes current to flow between adjacent strips, and the higher the water, the lower the resistance between them. Eventually, the resistance will be low enough to turn on the transistor and sound the alarm.

Parts list

Resistor

R1	10 000 Ω (10 kΩ), 0.25 W

Semiconductor

TR1	BFY51

Additional items

WD1	6 V miniature buzzer (polarised)
	Battery PP3
	Battery connector
	Strip board 9 strips by 15 holes
	Perforated board 3.5 × 10 cm

78 A delta loop for 20 metres

Introduction

HF aerials take up a lot of room when they are straight. Space can be saved by bending them and, at the same time, giving them properties which are quite different from their original linear forms. The delta loops shown in the accompanying diagrams have the shape of the Greek upper-case letter delta (Δ) upside down. It can be upright, except that it is more practical to have the feed point at the bottom than at the top.

Putting the loop together

This is the design for a loop to be used on the 20 metre band. It is a very popular band and carries most of the amateur radio DX traffic. The aerial is light in weight and operates best if the top section is at least 30 ft above the ground, with the feeder point being about 6 ft above ground. All you need is some good wire, some polypropylene rope, some insulators and a method of matching the output of this 75 Ω balanced aerial to the 50 Ω unbalanced coaxial input of your transceiver or receiver.

Each of the three sides must be 7.2 m long. Using tent pegs or six-inch nails, mark out the three corners on the lawn, making sure that each pair of pegs or nails is exactly 7.2 m apart. Put the wire around the nails, together with two insulators on the wire for the two top corners. These insulators can be of the ‘dog bone’ variety, or of the home-made type, using a piece of flat plastic with holes in each end. The use of these is shown in **Figure 1**. Located at the positions of two of the tent pegs or nails, they must be secured at those points by any convenient means.

Now attach a support rope to the free end of each of the insulators, ready for suspension, as shown in the diagrams, from any convenient supports (house, tower, tree, etc.). The top section of the loop could be taped to a continuous piece of rope between the two supports. You might like to try this if you think it is easier. Then, when you find some dog-bone insulators, you can change the design and see if there is any noticeable change of performance.

Connecting to the radio

If you are the proud possessor of an aerial tuning unit (ATU) with a balanced output, all you will need is a length of twin cable soldered to the

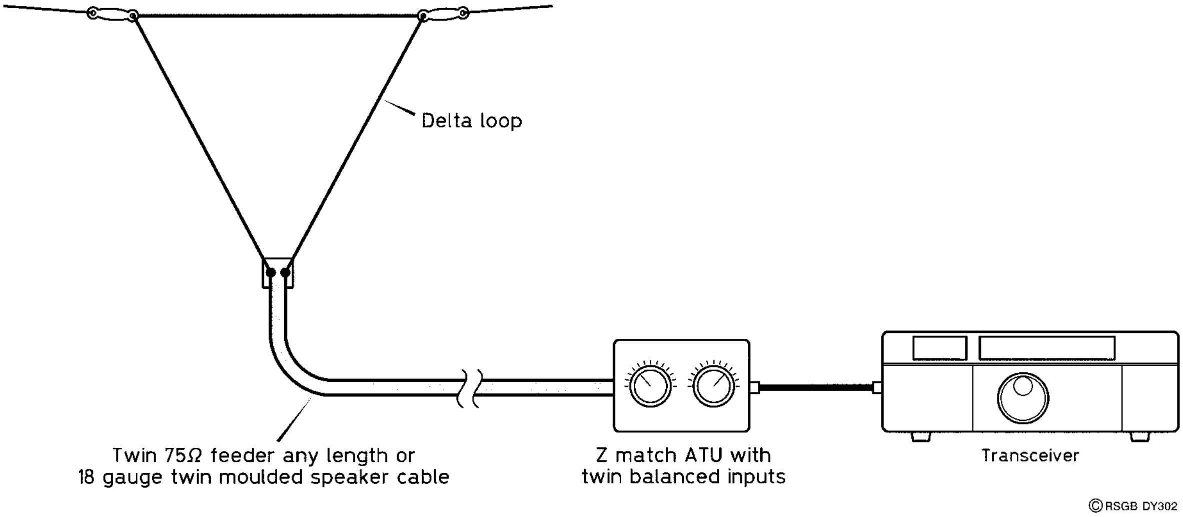
Delta loop
Twin 75Ω feeder any length or
18 gauge twin moulded speaker cable
Z match ATU with
twin balanced inputs
Transceiver
©RSGB DY302

Figure 1 Multiband version of the delta loop connected to the ASTU using 75 Ω twin feeder

ends of your delta loop and connected to the balanced output terminals on the ATU, as Figure 1 shows.

Another way is to buy a ferrite-cored 1:1 balun, and use it as indicated in **Figure 2**. This produces an aerial which will operate over the whole range of amateur and commercial short-wave frequencies, when used with an ATU.

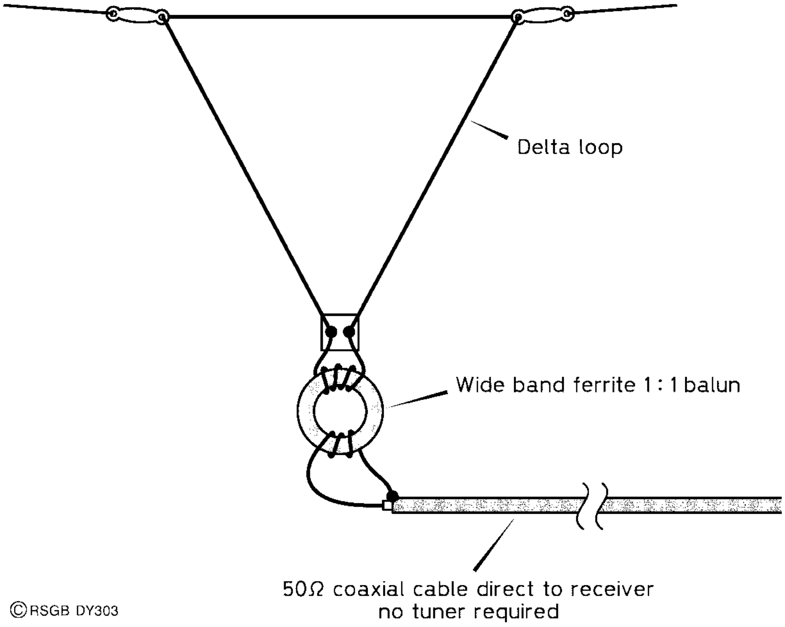
Delta loop
Wide band ferrite 1 : 1 balun
50Ω coaxial cable direct to receiver
no tuner required
©RSGB DY303

Figure 2 Multiband version of the delta loop using 75 Ω coaxial cable. This arrangement requires an ASTU if used with a multiband antenna

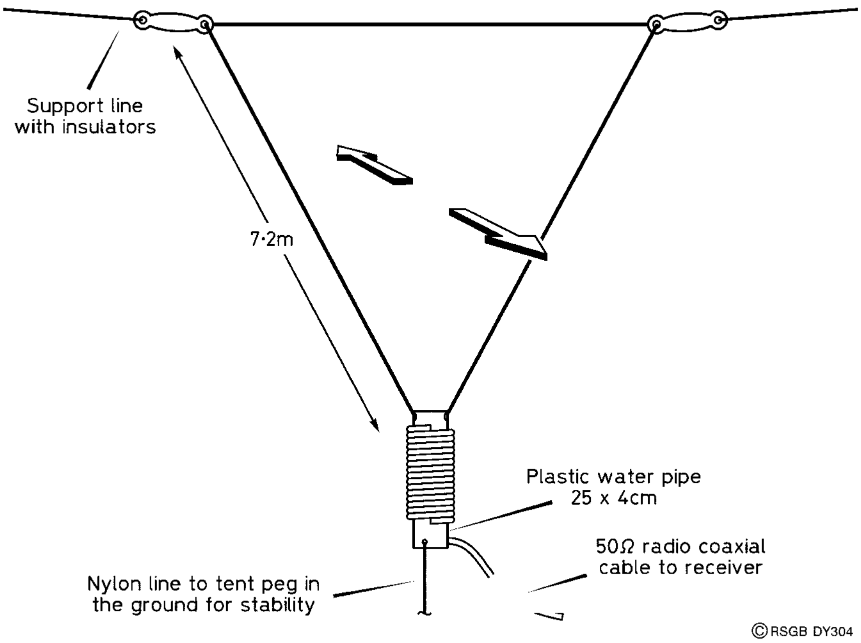

Figure 3 Single band version of the loop antenna using a coaxial balun. The direction of maximum signal strength is indicated by the arrows

However, since this aerial is designed to be a single-band type, there is a simpler answer to the problem of matching the aerial to your 50 Ω coaxial cable – **Figure 3** shows this. Make a tuned balun with a length of plastic water pipe, 25 cm long and 4 cm diameter and a piece of good-quality TV aerial cable. TV aerial cable has a characteristic impedance of 75 Ω, compared with the 50 Ω of the coaxial feeder that we usually use with amateur radio equipment. Use the type with a brown sheath and a closely knit earth braid, not the type having an earth foil inside the sheath.

The length we need for the balun is 3.8 m, but always allow between 3 and 4 cm extra for preparing the ends. Drill two small holes diametrically opposed in the top of the tube; these will be used to anchor the two ends of the delta loop, as shown in Figure 3. Drill another single hole in the bottom end of the tube, which will be used to anchor a nylon line going to the ground to add stability to the loop. Then, after drilling a pilot hole, drill a 5 or 6 mm hole near the top end of the tube, as shown in **Figure 4**. Prepare both ends of the coaxial cable, then feed one end into the tube, far enough for its ends to be soldered to the ends of the loop when the assembly reaches that stage.

Now, close-wrap the cable around the tube until only about 3 cm remain. Holding the cable tightly, drill another 6 mm hole beside the free end of the cable and feed it into this hole. At this point, feed the two ends of the loop into the top two holes and twist it back on itself. Figure 4 shows how this is done. Then, solder the ends of the coaxial balun to the ends of the loop.

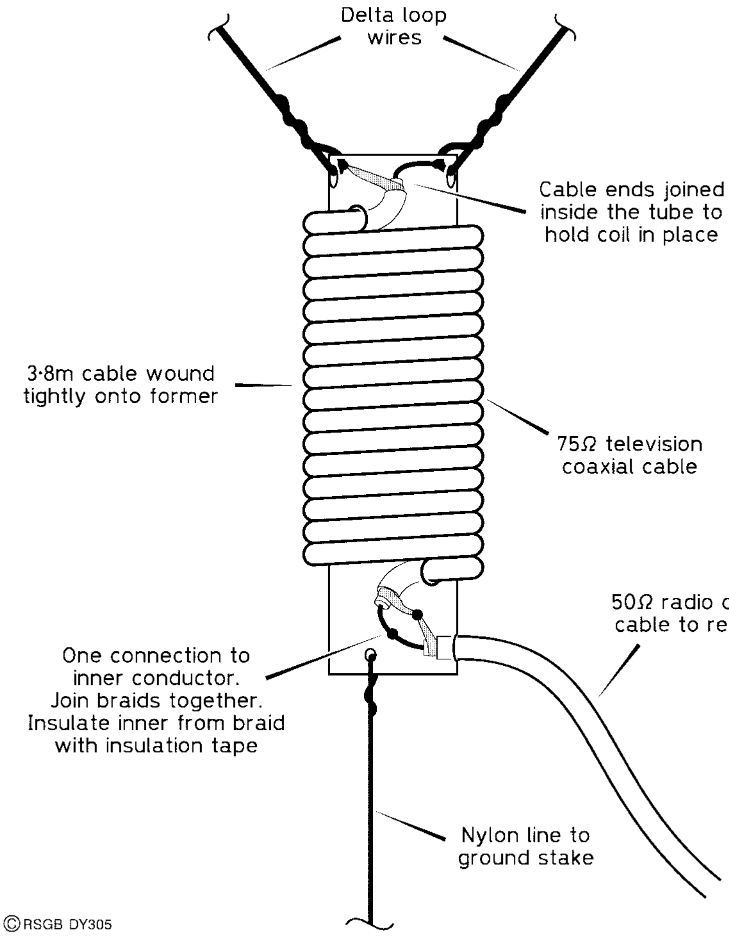
Delta loop wires
Cable ends joined inside the tube to hold coil in place
3·8m cable wound tightly onto former
75Ω television coaxial cable
50Ω radio coaxial cable to receiver
One connection to inner conductor. Join braids together. Insulate inner from braid with insulation tape
Nylon line to ground stake
©RSGB DY305

Figure 4 Construction of the balun

The bottom ends of the balun are then soldered to the 50 Ω coax which goes to your shack and to the transceiver.

Hoist the aerial into position carefully, being careful not to pull too hard on the support lines. Then, take the nylon line from the bottom of the balun to a peg in the ground. This adds stability to the aerial.

Using the delta loop

It is a directional aerial, as Figure 3 shows. It produces maximum power (and has maximum receive sensitivity) along a direction perpendicular to its own plane. Don't be too concerned with which direction it is pointing at first. Give it a try 'on the air' and see how it performs. Then you can contemplate how to point it in your favourite direction, to the USA, or Australia, for example.

Experimenting

You may want to enclose the balun in some sort of weatherproof container. Plastic ice cream containers are favourites for this sort of job. Seal all the holes where wires enter it with silicone sealant or self-amalgamating tape.

If you have used ordinary single-strand or multi-strand wire for your aerial, it will stretch over time under its own weight and that of the balun. Its operating frequency will fall slightly as a result. If you notice a significant difference, then dismantle it, remeasure and fix the sides and erect it again, perhaps facing a different direction. You can buy pre-stretched or hard-drawn wire for such purposes, if you feel that periodic tweaking of your aerial is a chore.

You can make a delta loop for different frequencies simply by scaling the lengths of wire for the loop and for the balun according to the design frequency. If your maths isn't quite up to this, enlist the help of someone well versed either in maths or aerial design!

79 A simple desk microphone

Introduction

A hand-held transceiver is usually the first type used by most novices. It is ideal when used out of doors, in the shack or mobile. However, using one in the shack with an outside aerial is not the best of solutions.

This article describes the design of a desk microphone which can be used either for transceivers using a common microphone and push-to-talk (PTT) line, or for those using a separate PTT line.

It is a simple design, using the barrels of two plastic pens, the lid of an old coffee jar, and the plastic sheath of a DIN plug.

Construction

The end caps and the ink tubes should be removed from two ball-point pens, leaving only the two plastic shells. One of them is used for the vertical microphone support, while the other is used to hold the microphone element.

The vertical support should have the pointed end shaped so that it holds the microphone support at a convenient angle, as **Figure 1** shows.

The base should be prepared next – see **Figure 2**. Four holes are required. Two, A and B, are small holes for the cable to be fed through. A third, C, is slightly larger and is used to take one of the plastic end caps. The fourth hole, D, is determined by the type of switch that you want to use for the PTT function.

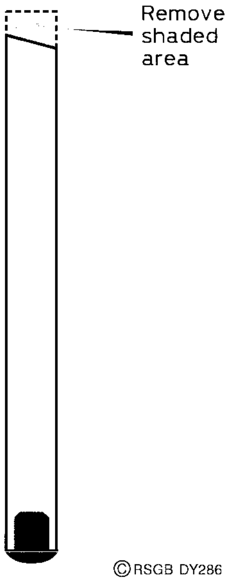
Remove
shaded
area
©RSGB DY286

Figure 1 Vertical support made from ball-point pen shell

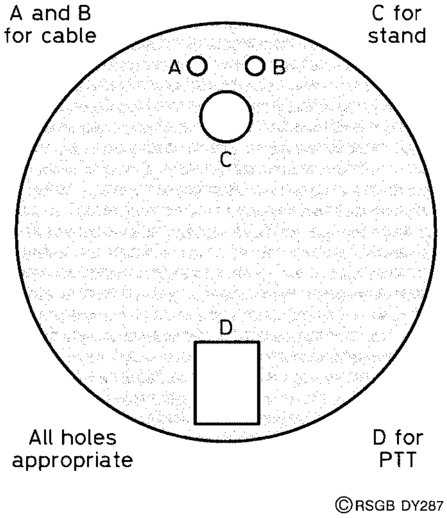
A and B
for cable
C for
stand
A
B
C
D
All holes
appropriate
D for
PTT
©RSGB DY287

Figure 2 The microphone stand base

Some epoxy resin such as Araldite Rapid is used to mount the separate parts. Mix only a small amount at a time, and glue each part of the base assembly and allow to harden before moving on. Be sure to read and follow the instructions for the use of the glue to prevent accidents.

Firstly, glue the end cap to the base, feeding it underneath and pushing it up through hole C. Then, the vertical pen body is placed over the protruding end cap and glued into place, making sure that the top (shaped) end is pointing in the right direction to hold the other plastic tube carrying the microphone. While the glue hardens, place the DIN plug sheath on the second pen body. Place and glue this pen body on the shaped top of the vertical support, making sure that the whole assembly is supported while the glue sets. What you should have now ought to resemble **Figure 3**, but without the microphone and wiring. Leave everything to set for 24 hours.

Any switch of the press-to-make variety can be used for your PTT switch, but choose one which does not apply too much force to operate! If you find the whole assembly is too light and slides over your table, some plasticine can be pushed into the base when all the wiring has been completed.

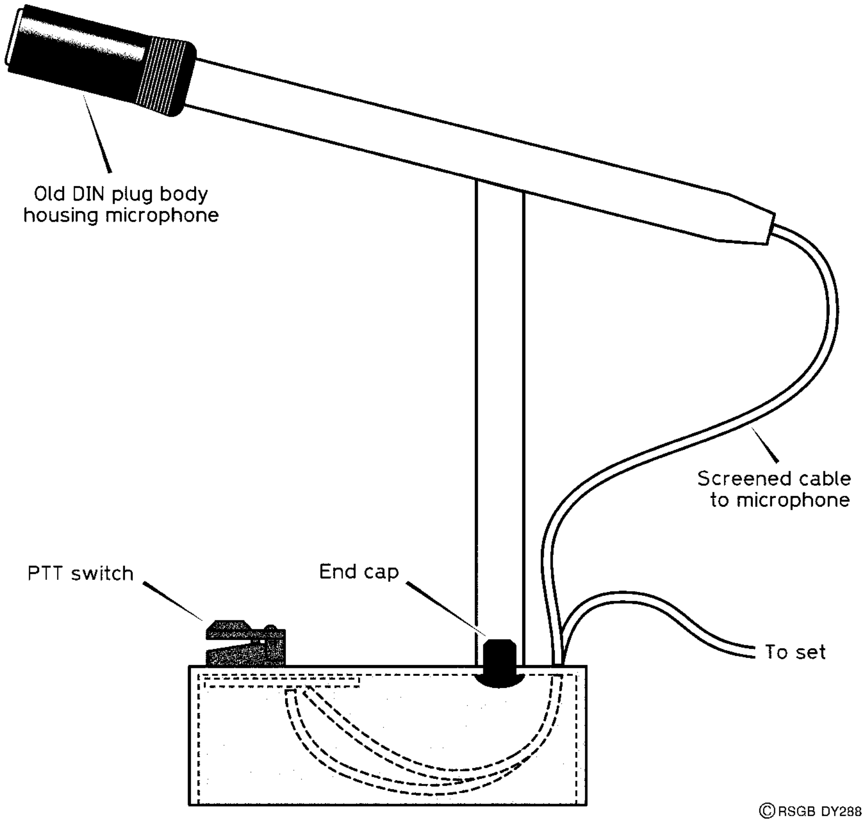
Old DIN plug body housing microphone
Screened cable to microphone
PTT switch
End cap
To set
©RSGB DY288

Figure 3 Complete microphone stand assembly

The circuit

A small electret condenser microphone is used, powered by a DC supply. Most hand-held radios use a single screened lead for both the PTT line and the microphone audio lead. In such cases, the circuit of **Figure 4** will operate the PTT function. Note that the PTT switch is in series with the audio lead from the microphone. If you find that the PTT switch does not switch your radio into transmit, reduce R1 from its normal value of 33 kΩ to 27 kΩ. You should find that the original value works with most hand-helds.

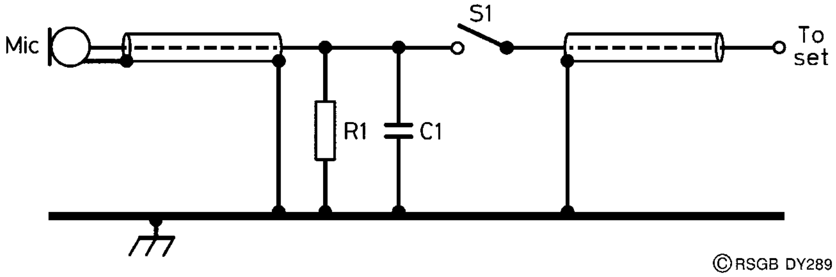

Figure 4 Wiring diagram for hand-held radios

For radios with a separate PTT lead, the circuit of **Figure 5** is used. However, a power supply is needed for the electret microphone. This can be a PP3 battery, or can (in most cases) be derived from the microphone socket on the transceiver. You will need to consult the makers' handbook for this information. Figure 5 shows the PTT switch wired in a 'ground-to-transmit' configuration, which is correct for most base-station transceivers. If you're in any doubt about what your radio needs, consult an experienced friend.

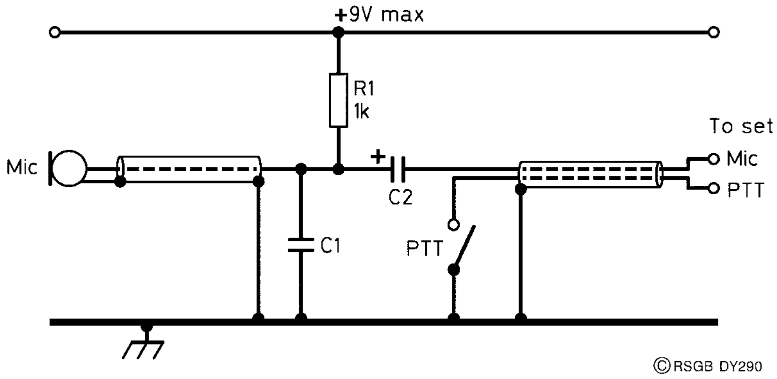

Figure 5 Wiring diagram for radios which use separate PTT lines

The last connections

The microphone element should be connected with screened cable. Tape (masking or insulating) should be wound around the electret insert until it is a snug fit inside the DIN plug sheath. Feed the screened cable through the barrel of the pen until it emerges from the far end. Poke it through hole B and make the connections under the coffee jar lid. The components can all be mounted on the PTT switch or on a small piece of Veroboard mounted under the top surface. The output lead emerges through hole A and is of sufficient length to reach your transceiver. A suitable plug needs to be fitted to it.

Parts list

Resistors: all 0.25 W carbon, 5% tolerance

R1 (Figure 4)	33 000 Ω (33 kΩ)
R1 (Figure 5)	1000 Ω (1 kΩ)

Capacitors

C1 (Figure 4 and 5)	0.1 microfarads (0.1 μF) disc
C2	10 microfarads (10 μF) electrolytic

Additional items

S1	Push-to-make
Mic	Electret microphone (Maplin type FS34W)
Screened cable	As required
Plug	As required by your radio
Veroboard	If required
PP3 battery and clip	If required

Two old plastic ball-point pens
Lid of coffee jar
Plastic sheath from DIN plug
Araldite or similar epoxy resin glue

80 Morse oscillator

Introduction

This is not the simplest Morse oscillator to build, but it differs from the simple circuits in that it produces a pure note, not a coarse, rasping sound. People who have practised Morse using a non-sinusoidal oscillator sometimes find that they have trouble copying Morse code with a pure tone. As the pure tone is the correct way to receive Morse code, it is important that you should learn to listen to the code from a pure oscillator – so here's one!

The twin-T

There are many oscillator circuits, and there are many variations of the twin-T oscillator that we are going to use. **Figure 1** shows one version of a very useful circuit. All oscillators must have positive feedback in order to work. The feedback determines the frequency of the note produced by the oscillator.

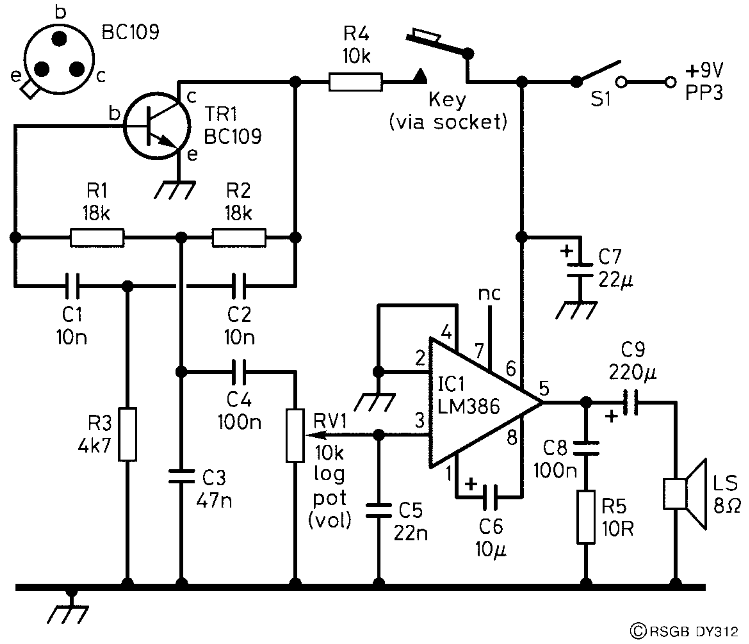
b
BC109
e
c
R4
10k
Key
(via socket)
+9V
PP3
S1
c
b
TR1
BC109
e
R1
18k
R2
18k
C7
22μ
C1
10n
C2
10n
nc
4
7
2
6
IC1
LM386
5
C9
220μ
C4
100n
RV1
10k
log
pot
(vol)
3
8
C8
100n
R3
4k7
C3
47n
1
C5
22n
C6
10μ
R5
10R
LS
8Ω

©RSGB DY312

Figure 1 Circuit diagram of the sinewave oscillator and amplifier

Here, the feedback circuit looks like two letter Ts. If you look at Figure 1, one T is formed by R1, R2 and C3, the other by C1, C2 and R3 – hence the name 'Twin-T'. Notice that the two Ts are connected in parallel between the collector and base of TR1, so any signal appearing at the collector is fed into both Ts. What emerges is then fed back into the base, producing in turn a signal at the collector. And so it goes on, producing a sine wave output.

The oscillator output is fed into an integrated circuit amplifier for output via a small loudspeaker.

Putting it together

The prototype was constructed on plain matrix board (the type *without* the copper strips), as shown in **Figure 2**. The components have their leads pushed through the holes in the board, and connections are made underneath.

Build the amplifier circuit first, using a socket for IC1. Connect the 9 V supply and touch pin 3 with an ordinary piece of wire. If a buzz is heard from the speaker, all should be well. If not, check your circuit and make changes until it does.

Build the oscillator circuit, and connect its output to the volume control VR1 via C4. Set VR1 half-way along its travel and switch on. A note should be heard from the speaker when the Morse key is depressed. The component values making up the twin-T determine the frequency of the note. Try varying them if you think your note is too high or too low. Whatever changes you make, either to the resistors or the capacitors, always ensure that R1 = R2 and that C1 = C2.

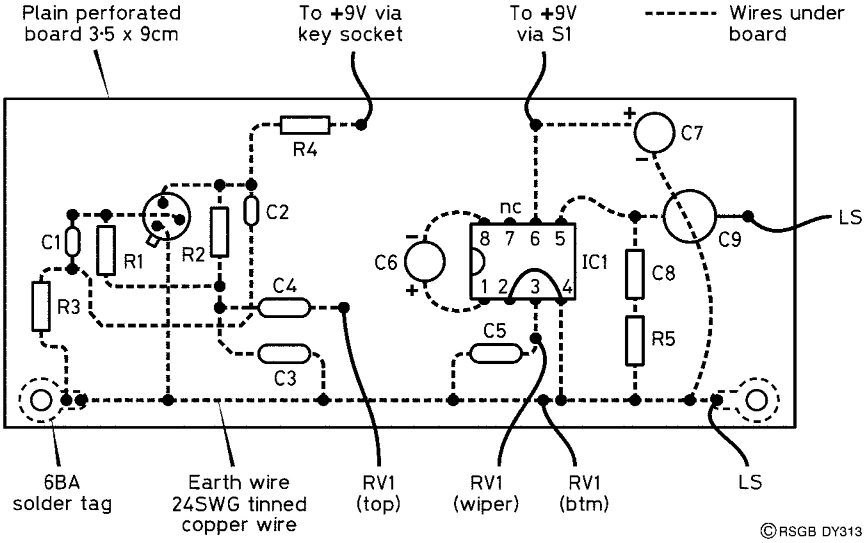
Plain perforated board 3·5 x 9cm
To +9V via key socket
To +9V via S1
------- Wires under board
R4
+
C7
−
C2
nc
8 7 6 5
IC1
LS
C1
R1
R2
C6
−
+
C9
C8
C4
1 2 3 4
R3
C5
R5
C3
6BA solder tag
Earth wire 24SWG tinned copper wire
RV1 (top)
RV1 (wiper)
RV1 (btm)
LS

Figure 2 Component layout and interconnection diagram

©RSGB DY313

Parts list

Resistors: all 0.25 W carbon film

R1, R2	18 000 ohms (18 kΩ)
R3	4700 ohms (4.7 kΩ)
R4	10 000 ohms (10 kΩ)
R5	10 ohms (10 Ω)

Capacitors

C1, C2	10 nF (0.01 μF)
C3, C8	47 nF (0.047 μF)
C4	100 nF (0.1 μF)
C5	22 nF (0.022 μF)
C6	10 μF electrolytic, 16 V WKG
C7	22 μF electrolytic, 16 V WKG
C9	220 μF electrolytic, 16 V WKG

Semiconductor

TR1	BC109

Integrated circuit

IC1	LM386

Additional items

VR1	10 kΩ log potentiometer with switch
LS	8 Ω loudspeaker

Matrix board 3.5 × 9 cm
Small jack socket for key input
Box
PP3 battery and clip

81 A simple 6 m beam

Introduction

The attraction of building your own aerials is an abiding feature of our hobby. You can *buy* almost any shape or size of aerial, but one you have made yourself can often work every bit as well as a commercial device costing ten times as much.

The design

This aerial, designed for use on the 6 m band, is essentially a two-element Yagi, with the elements bent in order to reduce the physical size. It is known

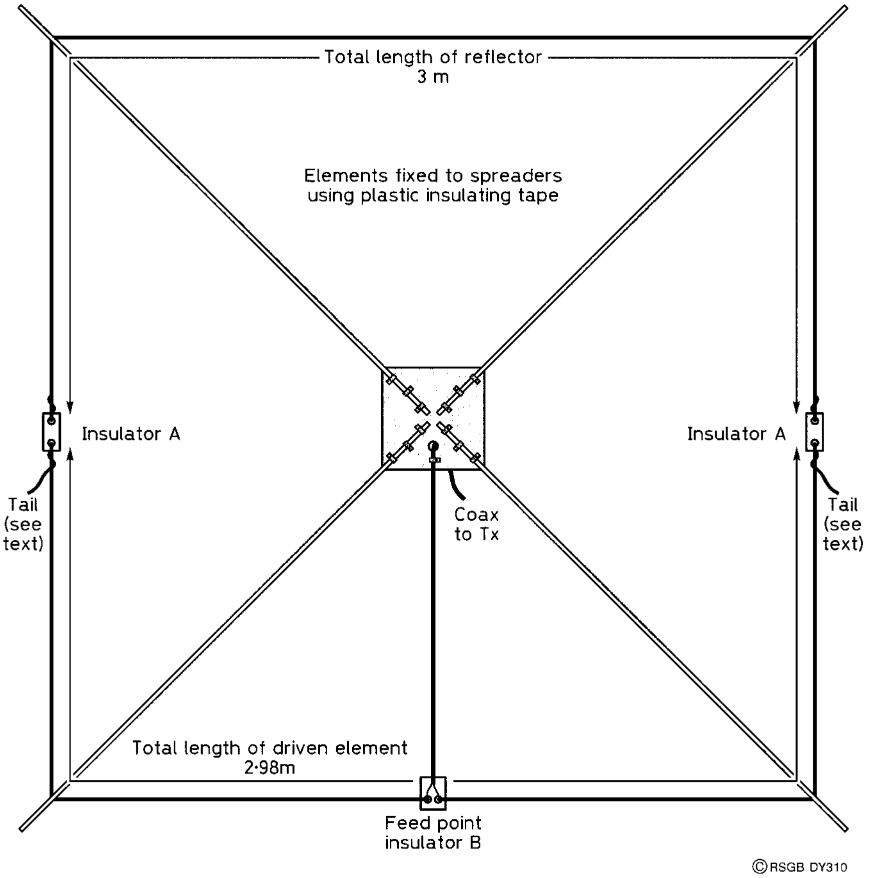
Total length of reflector
3 m
Elements fixed to spreaders
using plastic insulating tape
Insulator A
Insulator A
Tail
(see
text)
Coax
to Tx
Tail
(see
text)
Total length of driven element
2·98m
Feed point
insulator B
©RSGB DY310

Figure 1 Plan view of the complete 50 MHz VK2ABQ antenna

as the VK2ABQ beam, and was designed originally for the 20, 15 and 10 m bands, principally because of its space-saving qualities. It is made using a wooden frame and wire elements, and is ideal for portable operation.

Tools ready?

The beam is shown in **Figure 1**. The driven element is the one whose centre is fed by the coaxial cable, and lies between the two insulators marked A and the feedpoint at B. The reflector is also anchored at the points A, and lies over the upper half of the frame.

The wooden centrepiece is used to support the cross-pieces and to mount the aerial on the mast, using a common shelf bracket. The cross-pieces, known as *spreaders*, can be wooden canes or dowelling, and are mounted to the centrepiece using cable clips and adhesive. **Figure 2** shows how this is done.

If the aerial is to be a permanent installation, the spreaders should be weatherproofed using a good-quality exterior varnish. The wire elements are PVC covered and fed through holes in the spreaders.

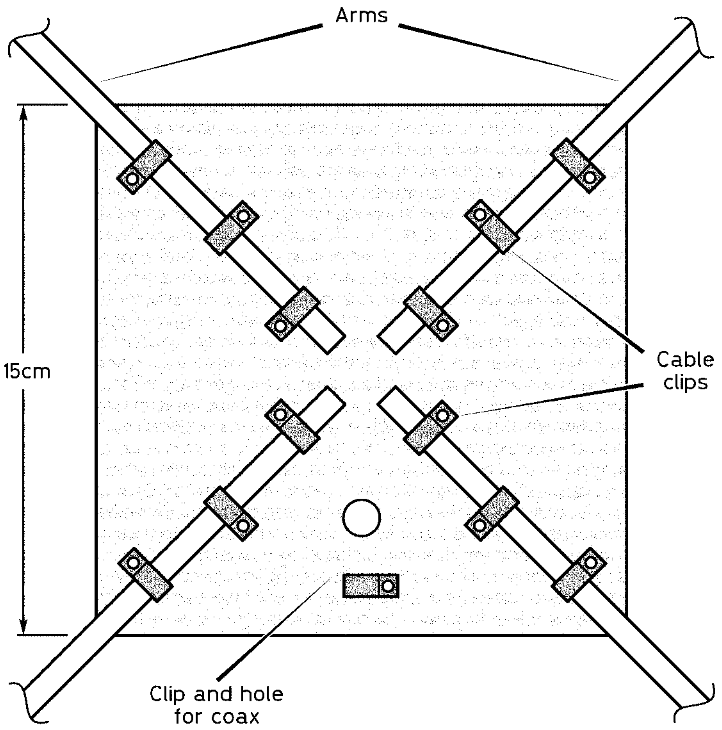

Figure 2 Centre support piece

©RSGB DY309

The end insulators are made of drilled perspex, and the wire passed through the two holes and twisted, as shown in **Figure 3**.

Note: if you have not drilled holes in perspex before, take care! The drill bit must be well-lubricated because it generates a lot of heat (enough to melt the perspex and jam the drill). Never turn the drill for more than a few seconds at a time, and moisten the bit in between. Start with a pilot hole and use bits of increasing size until the hole is the size you want.

Another perspex insulator is used to secure the feeder to the aerial, as shown in Figure 3. The feeder then passes directly to the centrepiece, where it is fastened with a cable clip and then passes down the mast.

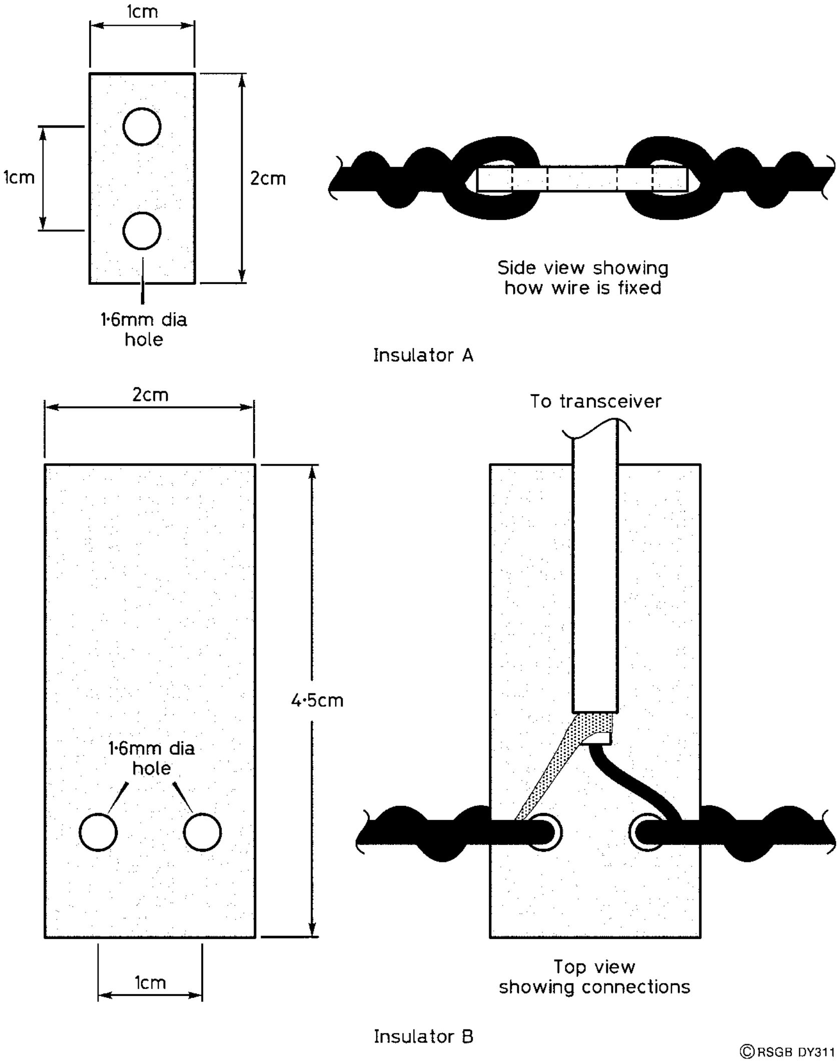
1cm
1cm
2cm
Side view showing how wire is fixed
1·6mm dia hole
Insulator A
2cm
To transceiver
4·5cm
1·6mm dia hole
1cm
Top view showing connections
Insulator B
©RSGB DY311

Figure 3 Details of insulators

Adjustment

The driven element (A–B–A) in Figure 1 is fixed to the end insulators in such a way as to have 'pigtails' which are about 10 cm long. Using an SWR meter between the aerial and the transmitter, trim the pigtails equally at each end for minimum SWR in that portion of the 6 m band in which you plan to work. Make sure that the transmitter is off when you trim the ends, as high voltages can be present there. Always listen on the frequency before you transmit and, when you do, ask if the frequency is in use and identify yourself. Use as little power as possible.

The prototype had an SWR of 1.2 at 50.2 MHz and performed well. If you look at Figure 1, which is a view of the aerial from above, you will immediately see that when the aerial is horizontal, it radiates with horizontal polarisation in a direction from the top of the page to the bottom. A metal pole or mast can be used for horizontal polarisation, but if you intend to use the aerial for vertical polarisation, it is much better to use a wooden or fibreglass pole.

Portable use

If you plan to operate portable with this aerial, the only real modifications you need are to the centrepiece and how it supports the spreaders. Instead of using glue and cable clips, nuts and bolts through the spreaders and centrepiece would allow the spreaders to be 'hinged' closed for transport.

Materials

Centrepiece	Hardwood, 15 × 15 × 25 mm
4 spreaders	110 cm long (cane or 6 mm dowel)
Wire	1.5 mm PVC-covered copper
50 Ω coaxial cable	RG58 or similar
Cable clips	6 mm (12 off)
Varnish	Polyurethane for waterproofing
Tape	Self-amalgamating, to waterproof all soldered joints
Insulators	See text

82 An integrated-circuit amplifier

Introduction

A simple audio-frequency amplifier is a very useful building-block in many more advanced electronic circuits. It is also a very useful piece of test equipment. To have one spare in the shack can be a life-saver at times.

Planning

Get into the habit of planning your project. How do you want it to look when it is finished. Do you want it in a box? Do you want it 'open-plan', with all the components on view?

The 'minimalist' approach to any project is simply a front panel and a baseboard, on to which all the components are fitted. In this case, the front panel would accommodate the loudspeaker, the volume control and the input and output jack sockets, and the baseboard would support the circuit board. Once you have decided these things, and know the size of the board, speaker and volume control, you can decided how big the panel and baseboard should be.

The amplifier

Instead of building the amplifier from discrete components (i.e. transistors), as was done in the project *An audio-frequency amplifier* (which you will find elsewhere in this book), we are going to use an integrated-circuit (IC) amplifier. External resistors and capacitors are still needed but, compared with the number of components *inside* the chip, these are very few indeed!

The IC we are going to use is the TBA820M. The circuit may look complicated, but with the use of a matrix circuit board it becomes quite simple. **Figure 1** shows the circuit diagram and **Figure 2** the layout on the board. The external connections to the PCB are shown in **Figure 3**.

To avoid having a separate switch to switch off the power supply when the amplifier is not being used, a volume control which incorporates a switch is used.

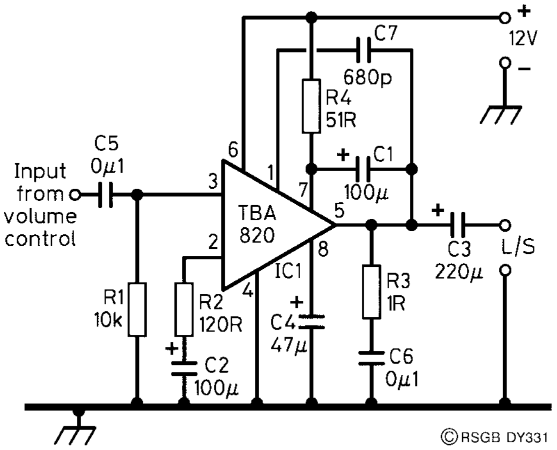
+
12V
−
C7
680p
R4
51R
C5
0µ1
Input from volume control
6
3
1
7
5
2
8
4
TBA 820
IC1
+ C1
100µ
+
C3
220µ
L/S
R1
10k
R2
120R
+ C2
100µ
+ C4
47µ
R3
1R
C6
0µ1
©RSGB DY331

Figure 1 Audio amplifier, circuit diagram

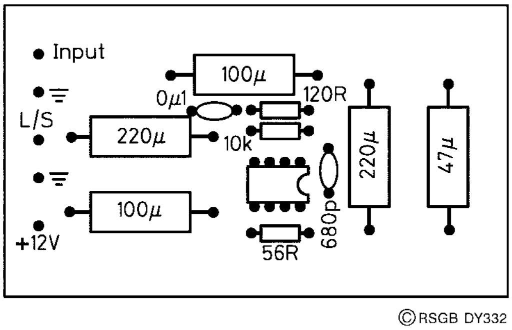
Input
L/S
+12V
100µ
0µ1
120R
220µ
10k
100µ
56R
680p
220µ
47µ
©RSGB DY332

Figure 2 Audio amplifier, component layout

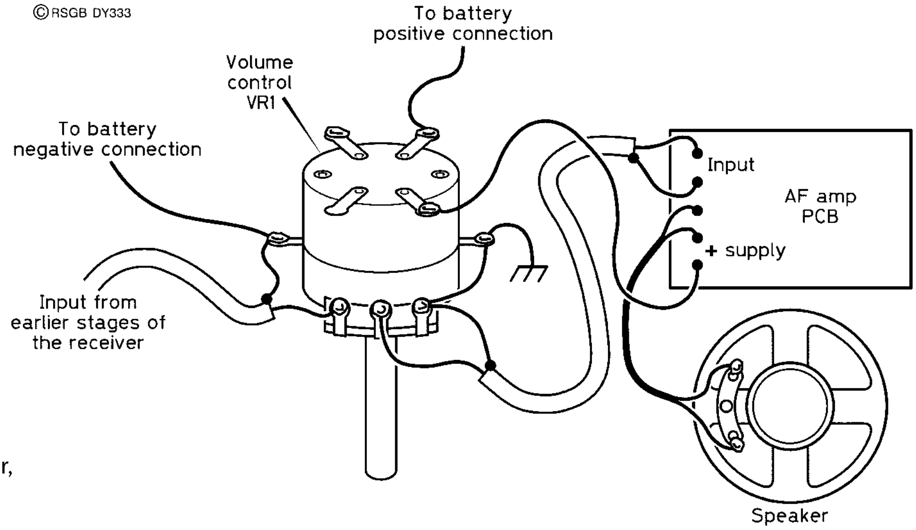
©RSGB DY333
To battery positive connection
Volume control VR1
To battery negative connection
Input
AF amp PCB
+ supply
Input from earlier stages of the receiver
Speaker

Figure 3 Connecting the circuit board to the speaker, volume control/switch and battery

What power supply?

If the amplifier is to be used only for short periods, as a test instrument, for example, then it can be run from a PP3 battery, which can be mounted on the baseboard. If you intend to use it often, then a connection to an external power supply is preferable. For this option, you may want to consider fixing two terminals to the baseboard for this connection.

83 A novice ATU

Introduction

Having an aerial tuning unit (ATU) is always useful. It is used for adjusting the aerial impedance, as 'seen' by the transceiver, to be the same as that of the transceiver itself, usually 50 Ω. This process is called *matching*, and ensures that the receiver and the power amplifier (PA) stage of the transmitter work efficiently.

This design is due to the late Doug DeMaw, W1FB, and uses readily available components. It will handle up to about 5 W, and operates over the frequency range 1.8 to 30 MHz.

Circuit evolution

The basic circuit of one type of ATU is shown in **Figure 1**. On transmit, the input signal at L1 is coupled into L2, which forms a resonant circuit with C1. The signal across the resonant circuit is fed to the aerial by C2. The combination of L2 and C1 helps to remove signal harmonics, because they are not at the resonant frequency and are shunted to earth. On receive, only those signals which are within the pass-band of L2 and C1 will pass into the receiver. This improves receiver performance by rejecting many out-of-band signals which the simpler receiver doesn't like.

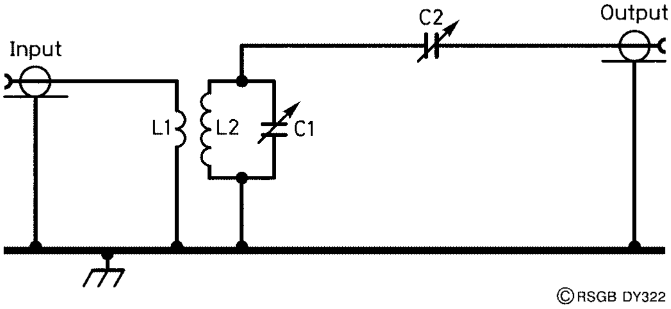

Figure 1 Basic circuit of the ATU

For this design, the all-important resonant frequency is chosen to be in the 80 metre band. If we want to make the circuit operate on other bands, we need to change either L2 or C1. It is easy to change L2 by using a set of switched inductors (L3–L5) as shown attached to S1 in **Figure 2**. Forget about S2 and L6 for the moment.

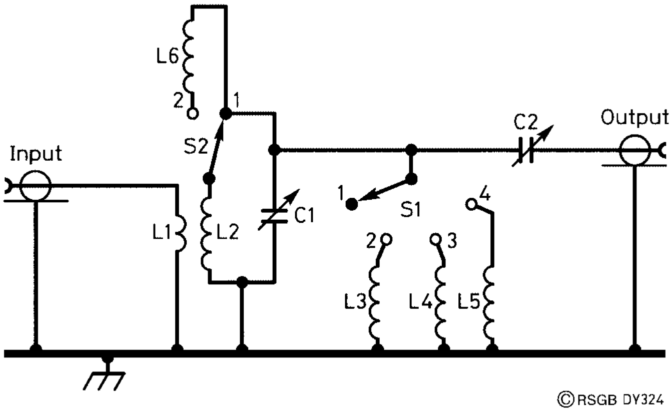

Figure 2 L6 added for 160 m

When S1 is in position 1, the circuit reverts to that of Figure 1. S2 should be brought into circuit only when S1 is in position 1. Looking at the way the circuit is drawn, you can see that when S2 is in position 2, we have two coils in series, which is equivalent to adding more turns to L2. More turns means more inductance, which lowers the resonant frequency still more, and gives coverage of the 160 m band.

Rather than having to remember to flick S2 to the correct position *and* move S1 to position 1 when we want to operate on 160 m, both functions can be combined into one rotary switch with two wafers. This circuit is shown in **Figure 3**.

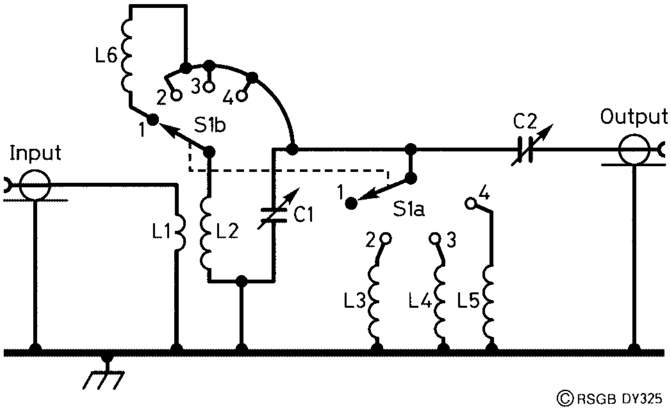
L6
2
3
4
1
S1b
C2
Output
Input
L1
L2
C1
1
S1a
4
2
3
L3
L4
L5
©RSGB DY325

Figure 3 Two-pole switch for all bands

In all positions of S1a *except* position 1, the extra coil, L6, is shorted out. The two halves of the switch, S1a and S1b, move together, as the shaft is turned, so both halves are in position 1 at the same time, position 2 at the same time, and so on. The two switches, S1 and S2, are said to be *ganged*.

Construction

L1 is formed by winding four turns of 22 SWG enamelled copper wire over L2, as in **Figure 4**. L2 already exists on the purchased former. After scraping the enamel off the ends of the wire (with a sharp knife or sandpaper), one end of L1 must be soldered to one end of L2, as shown in Figure 4. The free end of L1 goes to the transceiver aerial socket.

All components except the capacitors are assembled on the switch. Note that the rotor of C1 is earthed, but neither side of C2 is earthed. This means that the metal shaft of C2 is not earthed, and touching it will detune the ATU. Using a plastic knob for C2 will minimise this effect. If you decide to use a metal box, take precautions to ensure that no part of C2 is in electrical contact with the box.

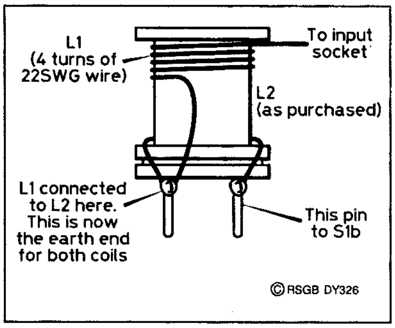
L1
(4 turns of
22SWG wire)
To input
socket
L2
(as purchased)
L1 connected
to L2 here.
This is now
the earth end
for both coils
This pin
to S1b
©RSGB DY326

Figure 4 L1 is wound over L2 as shown

The capacitors are mounted using M2.5 screws. Make sure that the screws do not foul the vanes of the capacitor. If your screws are too long, a few washers between the box and the capacitor will solve the problem!

In use

The best indication of a good match is obtained with a standing-wave meter between the ATU and the transceiver. The controls are adjusted alternately to ‘feel’ your way to a better and better match.

For receive-only use, the same alternate adjustments are used, watching for the maximum signal strength on the S-meter or, for a very weak signal, making the signal from the loudspeaker as large as possible.

Parts list

Capacitors	
C1, C2	350 pF variable
Inductors	
L1	See text
L2	27 µH
L3	10 µH
L4	2.2 µH
L5	1 µH
L6	65 µH
Additional items	
SO239 sockets (2 off)	
S1	2-pole 6-way rotary
Box	as required
Plastic knobs (2 off)	
Stick-on feet (4 off)	
M2.5 screws for capacitors	
Screws, nuts and washers for mounting the input and output sockets.	

84 CW QRP transmitter for 80 metres

Introduction

This is a relatively simple transmitter design having an output of 1 W. The design is not new, having been described before in other amateur radio publications. The components are all available new and the total cost should not exceed £15.

The circuit

Like other simple transmitters (see *An 80 Metre Crystal-Controlled CW Transmitter* and *A Breadboard 80 m CW Transmitter* elsewhere in this book) this one is crystal controlled. This assures frequency stability, but limits the usefulness of the transmitter. The key to increased frequency coverage without a conventional Variable Frequency Oscillator (VFO) is the use of a low-cost 3.58 MHz *ceramic resonator*. The 'pulling' range of a 3.58 MHz ceramic resonator covers the UK novice 80 m sub-band and some of the CW segment below 3.525 MHz.

A ceramic resonator is like a crystal, but not quite as stable in frequency. Its main advantage is its large pulling range.

The block diagram is shown in **Figure 1**. It is very similar to a crystal-controlled transmitter, and includes an oscillator, buffer and final amplifier.

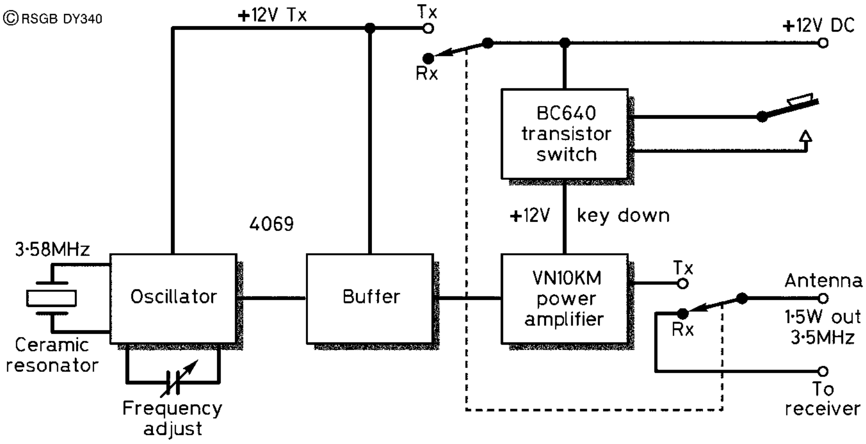
©RSGB DY340
+12V Tx
Tx
Rx
+12V DC
BC640 transistor switch
+12V
key down
4069
3·58MHz
Ceramic resonator
Oscillator
Buffer
VN10KM power amplifier
Tx
Rx
Antenna
1·5W out 3·5MHz
To receiver
Frequency adjust

Figure 1 Block diagram of the ceramic resonator controlled 80 m CW transmitter

This amplifier is keyed, the oscillator remaining running all the time. This improves frequency stability because the oscillator is not being continuously stopped and started by the key. It is switched off while receiving, though, to avoid interference with the received signal. Transmit/receive switching is accomplished by a panel-mounted switch controlling both the aerial, oscillator and buffer switching.

Figure 2 shows the transmitter circuit diagram. An unusual aspect of this transmitter is the use of a digital CMOS integrated circuit (IC) type 4069 for the buffer and oscillator stages. The IC houses six inverters, four of which are used in the circuit. One is used as the oscillator, two are used for the buffer stage, and the fourth provides an output for a direct-conversion receiver, should one be added at a later date.

The frequency of the oscillator is changed by varying the capacitance in the ceramic resonator circuit. This is provided by VC1.

The power amplifier (PA) is a small MOSFET (metal oxide semiconductor field-effect transistor), TR1. This is capable of providing an output power of 2 W but, in this circuit, it is run conservatively to give 1.5 W. The output can be varied by changing the resistance (R5 + R6) in the gate circuit. Attempts

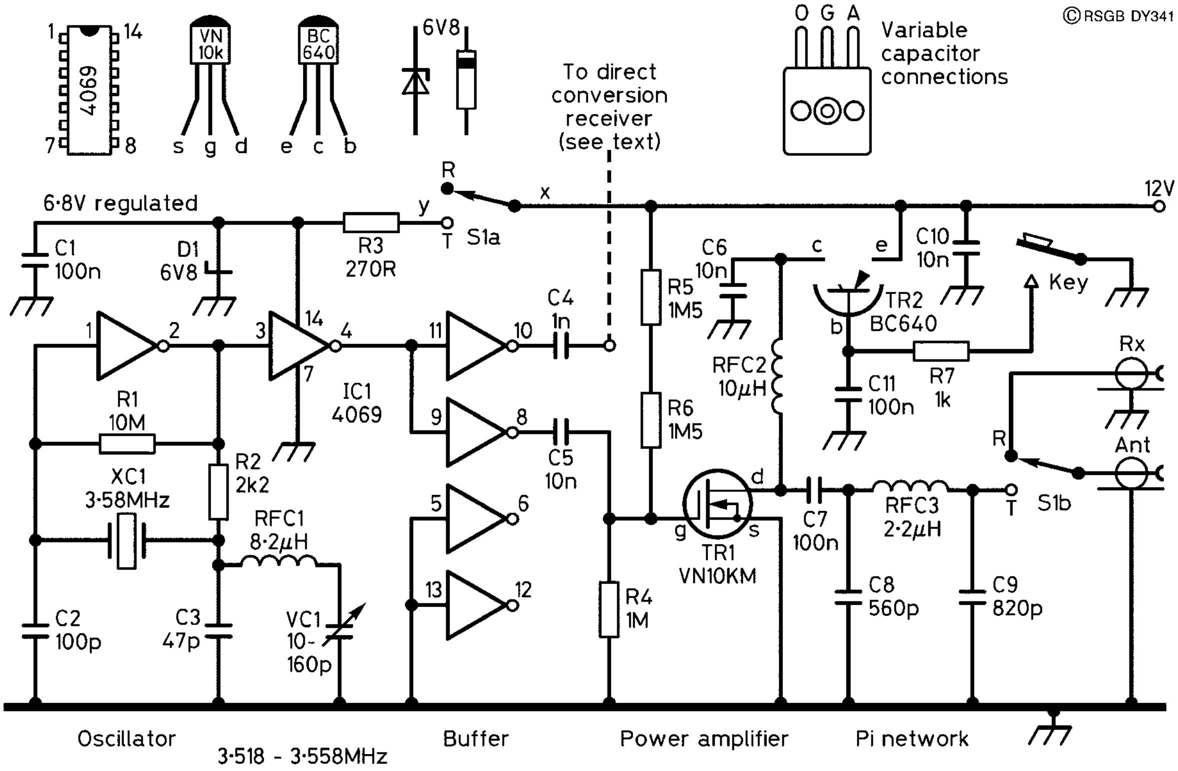
©RSGB DY341
1
14
4069
7
8
VN
10k
s g d
BC
640
e c b
6V8
To direct
conversion
receiver
(see text)
O G A
Variable
capacitor
connections
6·8V regulated
C1
100n
D1
6V8
R3
270R
y
R
x
T
S1a
12V
C6
10n
R5
1M5
c
e
C10
10n
TR2
BC640
b
Key
R7
1k
C11
100n
Rx
R
Ant
S1b
T
1
2
3
14
4
7
IC1
4069
11
10
C4
1n
9
8
C5
10n
5
6
13
12
R1
10M
XC1
3·58MHz
R2
2k2
RFC1
8·2μH
C2
100p
C3
47p
VC1
10-
160p
R6
1M5
RFC2
10μH
R4
1M
d
g
s
TR1
VN10KM
C7
100n
RFC3
2·2μH
C8
560p
C9
820p
Oscillator
3·518 - 3·558MHz
Buffer
Power amplifier
Pi network

Figure 2 Circuit diagram of 80 metre transmitter

to raise the output power by decreasing the values of these resistors may result in immediate MOSFET failure.

A pi-network (C8, RFC, C9) provides impedance matching to 50 Ω, together with harmonic suppression. Like all inductors in this transmitter, the pi-network inductor is a pre-wound RF choke. A pi-network is so called because the components are arranged in the shape of the Greek letter pi (π).

Keying is carried out by a pnp transistor switch, TR2. Closing the key earths the base and supplies 12 V to the collector of TR2 and to the drain of the MOSFET, TR1, allowing the PA to operate.

Construction

You must house your transmitter in a metal box, to avoid hand-capacity effects and the radiation of spurious frequencies. Size is not important, provided it is large enough to accommodate the transmitter without cramping the components. You may want to allow space for future additions such as a direct-conversion receiver, break-in keying, sidetone or a small power amplifier. A good size is 5 × 15 × 15 cm. You can make your own box, buy it, or even use a biscuit tin!

Front and rear panel connectors can be fitted first. The choice of these is a personal matter, but a good working choice would be:

(a) Power socket – 2.1 mm panel socket – centre pin positive.
(b) Key socket – ¼ inch jack socket.
(c) Aerial and receiver connectors – panel-mounting SO239 type.

Particular attention must be paid to the mounting of the variable capacitor, VC1. Make sure the hole for the shaft is amply big enough, and if you use screws to mount the capacitor on the front panel, then make sure they are not too long, otherwise they will touch the vanes of the capacitor! Mounting can be by means of glue, sparingly applied and kept well away from the shaft.

A board size of about 6 × 10 cm is adequate. Component layout on the board is suggested in **Figure 3**. The prototype used ordinary matrix board, which is preferable to stripboard for a design like this; stripboard has undue capacitance between adjacent strips. Component leads are fed through holes in the board and are connected underneath. Make sure that leads and connections are rigid because, if they can move, there is always the danger of short-circuiting, and capacitance changes.

To facilitate construction, servicing and testing, it is advisable to use Veropins for connections to the variable capacitor, transmit/receive switch,

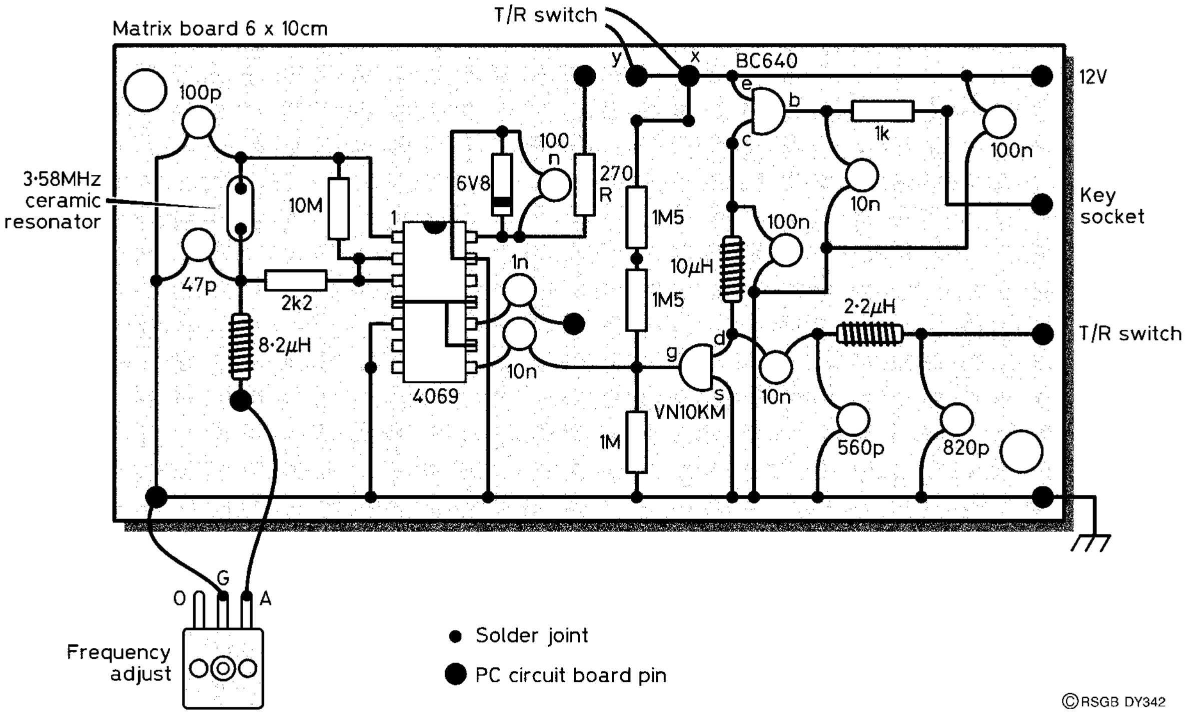
Matrix board 6 x 10cm
T/R switch
y
x
BC640
e
b
c
12V
100p
3·58MHz
ceramic
resonator
10M
1
6V8
100
n
270
R
1M5
1k
100n
10n
Key
socket
100n
10µH
47p
2k2
1n
1M5
2·2µH
T/R switch
8·2µH
g
d
4069
10n
s
10n
VN10KM
1M
560p
820p
G
O
A
Frequency
adjust
● Solder joint
● PC circuit board pin
©RSGB DY342

Figure 3 Component layout of the 80 metre transmitter. The transmit/receive switch is not mounted on the board and is not shown in this diagram

aerial and power sockets. Use screws and spacers to mount the circuit board to the box. Mounting the board horizontally assists troubleshooting.

Use a socket for the IC, and observe the CMOS handling precautions given in *An Electronic Die*, elsewhere in this book. When soldering the leads to the ceramic resonator, do it quickly – excessive heat damages the device. The earth lead running acoss the bottom of the board must be connected to the metal case by a short length of stout wire.

Testing

After carefully checking your wiring, both against the circuit diagram and the layout diagram, it is time to test your circuit. You will need a multimeter, an 80 m SSB receiver and a 50 Ω dummy load. A good design of dummy load can be found in the project *A Switched Dummy Load*, also in this book. An RF power meter and frequency counter will also be useful, although if your receiver has a digital frequency readout and S-meter, the latter two items are not really necessary. You will also need a 12 V 1 A power supply unit (PSU) to power the transmitter.

Switch the transmitter to receive and switch on the transmitter. No current should be consumed. Switch to transmit and check that pin 14 of IC1 is 6.8 V positive. With the dummy load connected to the aerial socket, press the key. The voltage on TR2 collector should now be 12 V, dropping to zero when the key is released.

Now check the operation of the oscillator. In transmit mode, you ought to be able to find a strong carrier signal with the receiver, even though the dummy load is connected. Adjusting the variable capacitor should change the frequency. At the lower end of the frequency range, you may find that the oscillator is unreliable in starting, because the circuit is attempting to pull the resonator too low in frequency. If this is the case, set the trimmer at the back of VC1 to minimum capacitance. If your version of VC1 has two trimmers, and you don't know which one to set, set them *both* to minimum capacitance. If there is still a problem, reduce the value of RFC1 to 6.8 or 4.7 μH.

In all probability, the unmodified circuit of Figure 2 will not require any of the changes outlined here. A coverage of 3.518 to 3.558 MHz should be possible, while preserving good frequency stability and reliable oscillation.

A signal probe (see *An RF Signal Probe*, elsewhere in this book) is useful for checking the operation of the oscillator and PA. Alternatively, an RF power meter or the receiver's S-meter can be used. With the PA running, the unit should draw between 200 and 300 mA. If TR1 becomes too hot to touch after a few seconds of transmitting, increase R5 or R6 to limit the transistor's heat dissipation. A small 6.3 V bulb connected across the aerial output is a simple way to check that the PA is working. An orange/white glow when the key is pressed is indicative of correct operation.

The final test is to monitor keying 'quality'. With your dummy load connected, press the key and listen to the note on the receiver's loudspeaker. Then operate the key, sending a string of dits, for example. What you hear should be free of chirps and clicks, as well as being stable in frequency. This test is sometimes better performed with no aerial connected to the receiver, thus preventing receiver overload and its associated plops. No problems should be encountered here.

Frequency tuning

A peculiarity of ceramic resonators is that, every now and again, their frequency changes abruptly by 100 Hz or so, then remains stable for some time. This is certainly noticeable in the received signal, but does not detract from the QSO and no characters are lost as a result. Try to keep the area around the ceramic resonator cool, to avoid temperature variations.

Parts list

Resistors: all 0.25 W, 5% tolerance

R1	10 megohms (MΩ)
R2	2200 ohms (2.2 kΩ)
R3	270 ohms (270 Ω)
R4	1 megohm (1 MΩ)
R5, R6	1.5 megohms (1.5 MΩ)
R7	1000 ohms (1 kΩ)

Capacitors

C1, C7	100 nF
C2	100 pF
C3	47 pF
C4	1 nF
C5, C6, C10	10 nF
C8	560 pF
C9	820 pF
VC1	10–160 pF variable

Inductors

RFC1	8.2 μH
RFC2	10 μH
RFC3	2.2 μH

Semiconductors

IC1	4069
TR1	VN10KM
TR2	BC640
D1	6V8 Zener

Additional items
- Matrix board (see text)
- 14-pin DIL socket
- Pointer knob
- Sockets (see text)

85 An audio booster for your hand-held

Introduction

The audio output from many hand-held transceivers and receivers usually leaves much to be desired, so this little amplifier was designed to increase the output at minimal expense.

All that is needed is a separate amplifier and bigger loudspeaker. This is accomplished using a single integrated circuit (IC), a few components, and a loudspeaker from the junk box. This circuit will enable the output from your hand-held to be heard easily in a car.

The circuit

This is shown in **Figure 1**. It uses only those components necessary to operate the IC amplifier. VR1 is the preset volume control, and varies the signal coming from the 'External speaker' jack on the hand-held before feeding it into the IC for amplification. C1 blocks any constant voltage present on the input.

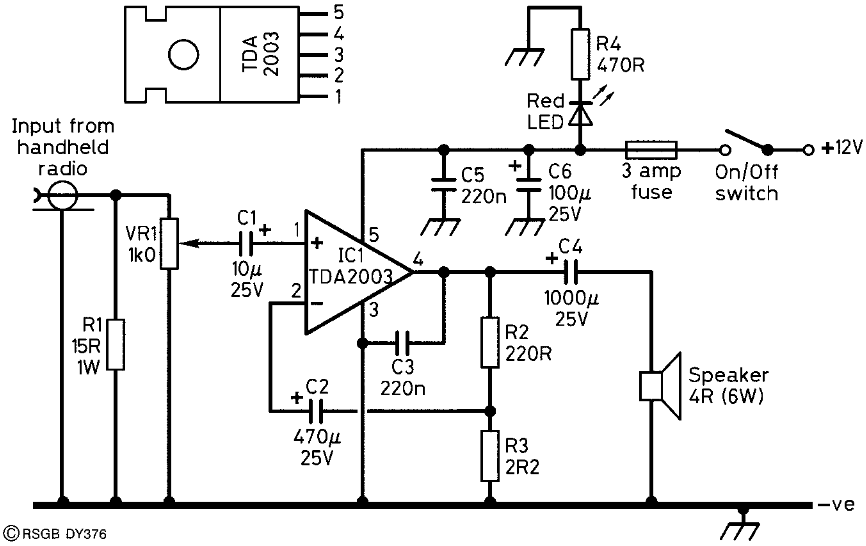
TDA 2003
5
4
3
2
1
R4
470R
Red LED
Input from handheld radio
C5 220n
C6 100µ 25V
3 amp fuse
On/Off switch
+12V
VR1 1k0
C1 10µ 25V
1
5
IC1 TDA2003
4
2
3
C4 1000µ 25V
R1 15R 1W
C3 220n
R2 220R
Speaker 4R (6W)
C2 470µ 25V
R3 2R2
−ve
©RSGB DY376

Figure 1 Circuit diagram of the audio amplifier. The power is derived from the cigarette lighter socket and the fuse is in-line with the lead

The IC output comes from pin 4 and is fed via the electrolytic capacitor, C4, to the loudspeaker. The circuit is provided with an on/off switch, fuse and LED to indicate when the circuit is switched on.

Construction

The box is made of aluminium. This is necessary to help to dissipate some of the heat generated by the IC. Do not build the circuit inside a plastic box unless you take special precautions! The IC has a metal mounting tab with a hole, specifically designed to be mounted to a metal box or other metal heat sink. Apply plenty of heat sink compound between the tab and the box, tighten the nut and bolt, and then wipe off any excess compound. The box will get slightly warm in operation.

The size of the box is not specified. You may want to decide on this when you find a loudspeaker. Choose one which will be able to handle 6 W

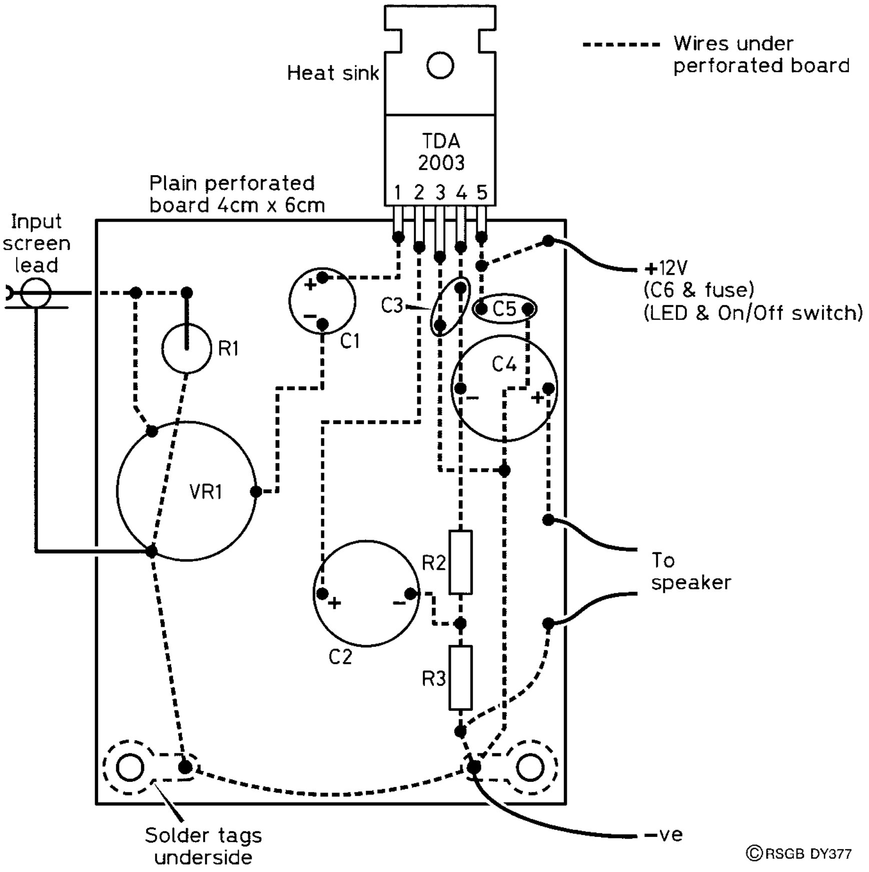
Heat sink
TDA 2003
1 2 3 4 5
------- Wires under perforated board
Plain perforated board 4cm x 6cm
Input screen lead
R1
C1
C3
C5
C4
VR1
C2
R2
R3
+12V
(C6 & fuse)
(LED & On/Off switch)
To speaker
Solder tags underside
−ve
©RSGB DY377

Figure 2 Layout of the components within the box

output. Drill all the holes in the box first. Holes for the speaker, input phono socket and the LED. The amplifier can be constructed on ordinary matrix board, which can be mounted inside the box with screws and spacers.

The layout is shown, for your guidance, in **Figure 2**. The components are mounted by pushing their leads through the holes in the board and making connections on the underside. The preset volume control, VR1, is set such that the hand-held's volume control is sufficient to control the final output over a good volume range. Use a screened lead from the 'External speaker' jack socket to the phono plug.

An external power supply is needed for this circuit. The normal dry battery which we usually use for small projects in this book will *not* work here, so you will need a proper mains power supply producing a stabilised 12 V. If you are going to use the unit principally in a car, then the cigar lighter socket can supply this voltage easily. Do make sure that the polarity is correct before you switch on!

When you plug the jack plug from your booster into the 'External speaker' socket on your hand-held, its internal speaker will be muted, so don't think that something dire has gone wrong! Adjust VR1 for a good volume range on your booster, when the volume control is turned on the hand-held.

Parts list

Resistors: all 0.25 W, 5% tolerance, unless otherwise stated

R1	15 ohms (15 Ω) 1 W
R2	220 ohms (220 Ω)
R3	2.2 ohms (2.2 Ω)
R4	470 ohms (470 Ω)
VR1	1000 ohms (1k Ω)

Capacitors

C1	10 μF 25 V
C2	470 μF 25 V
C3, C5	220 nF (0.22 μF) Mylar
C4	1000 μF 25 V
C6	100 μF 25 V

Semiconductor

LED	5 mm Red

Integrated circuit

IC1	TDA2003

Additional items
  Heat sink compound
  Nuts and bolts
  Loudspeaker 4 Ω 6 W
  3 A fuse
  On/off switch (SPST)
  Matrix board 4 × 6 cm
  Solder tags
  Plugs and screened cable for connecting lead
  Aluminium box

86 A grid dip oscillator

Introduction

When an inductor and a capacitor are connected, whether in series or parallel, they form a circuit with a natural (or *resonant*) frequency. The circuit stores energy, and this energy is being constantly shifted from the inductor to the capacitor and back again.

The dip oscillator is a simple instrument used to measure the resonant frequency of a tuned circuit without having to make any direct connection to the circuit. The circuit is more commonly known as the *grid dip oscillator* (GDO), from the days when the active device in the circuit was a valve. The FET or Field-Effect Transistor operates in a way which is very similar to that of the valve, so it is not quite a misnomer to call this instrument a grid dip oscillator, too.

The circuit

The GDO uses a calibrated, tunable FET oscillator in the circuit of **Figure 1**. It has a frequency range of 1.6 to 35.2 MHz in four ranges using a set of plug-in coils, shown in **Figure 2**. When the oscillator coil, L1, is placed near an external resonant circuit, some of its RF energy is coupled into the external circuit. A gain in energy of the external circuit must mean a loss of energy in the GDO circuit, resulting in a change of current through TR1, which is measured by the meter, M1.

The current through TR1 is of the order of 5 to 8 mA, but the change of current may be only a few microamps. To measure a very small

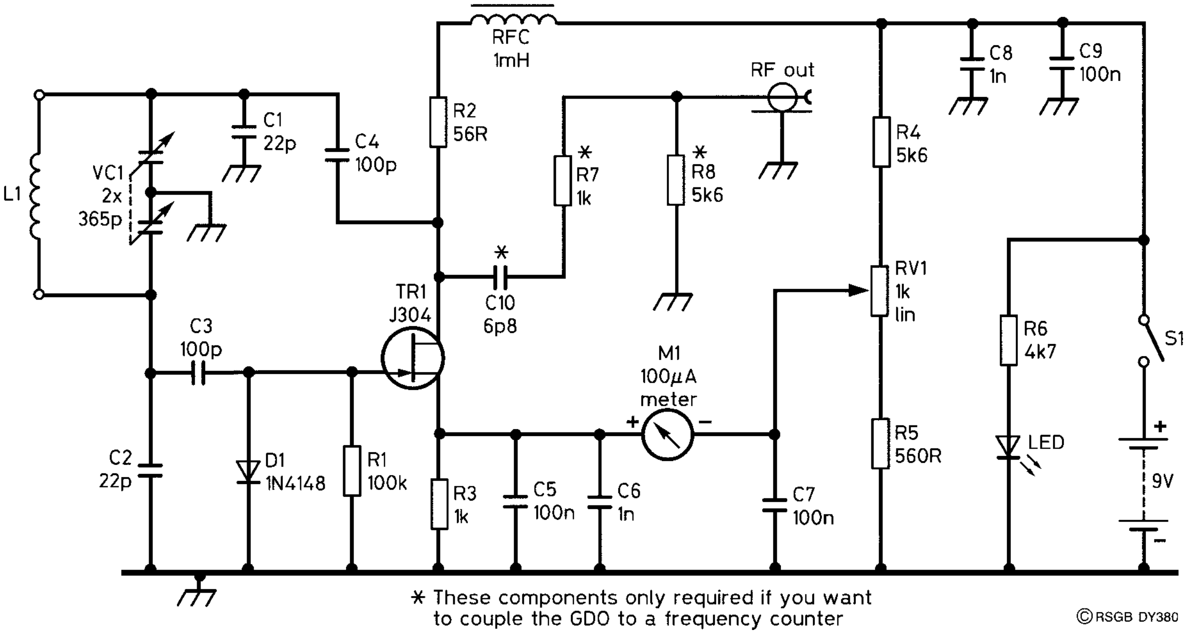
RFC
1mH
RF out
C8
1n
C9
100n
C1
22p
C4
100p
R2
56R
*
R7
1k
*
R8
5k6
R4
5k6
L1
VC1
2x
365p
*
C10
6p8
RV1
1k
lin
TR1
J304
R6
4k7
S1
C3
100p
M1
100μA
meter
+
−
LED
+
9V
−
C2
22p
D1
1N4148
R1
100k
R3
1k
C5
100n
C6
1n
C7
100n
R5
560R

* These components only required if you want to couple the GDO to a frequency counter

©RSGB DY380

Figure 1 Circuit of an FET GDO. The coils are wound on DIN speater plugs, which provide both a plug-in base and a coil former

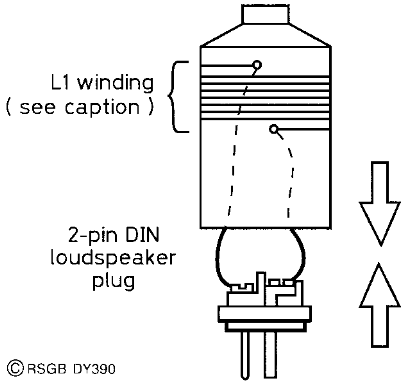
L1 winding
(see caption)
2-pin DIN
loudspeaker
plug

©RSGB DY390

Figure 2 Details of coil construction:
Range 1: 1.6–4.0 MHz 55 turns of 30 SWG
Range 2: 3.3–7.9 MHz 27 turns of 30 SWG
Range 3: 6.3–4.0 MHz 14 turns of 26 SWG
Range 4: 11.9–4.0 MHz 7 turns of 24 SWG.

change superimposed on a much larger standing current, the method of *offset* can be used.

One connection to the meter goes to the source of the FET, while the other goes to a variable offset voltage set by VR1. M1 has a full-scale deflection (FSD) of 100 μA. If the current through TR1 changes, the voltage across R3 changes. When there is no resonance, the voltage at the wiper of VR1 is set to be very slightly greater than that across R3, and there is a 75% FSD meter deflection. When the voltage across R3 decreases very slightly, due to external circuit resonance, a significant 'dip' in the meter deflection is produced, hence the name of the instrument.

The circuit is not difficult to make on standard matrix board. Provided you can follow a circuit and translate it into a good component layout, then this project is probably only an evening's work.

The most important part of the GDO is the tuning capacitor and its associated frequency-calibrated dial. New, air-spaced tuning capacitors can cost you up to £20, so it is worth delving around in junk boxes, or scouring the tables at a local boot sale or rally. The tuning capacitor from an old transistor radio should be perfect. It may even have a slow-motion drive and a dial which can be remarked for the project.

Choose a coil plug and socket arrangement that is practical. Think about crystal holders or phono plugs and sockets. The prototype shown in Figure 1 and Figure 2 used 2-pin DIN plugs, with the coil wound on the outside of the plastic plug cover. Figure 2 shows the coil construction and the winding details.

If you use a variable capacitor, VC1, with a value different from that shown in the parts list, then the frequency ranges will be different. This does not matter, as it will be taken into account during calibration.

Position VC1 so that the dial will be easy to see and to operate, while locating the coil socket as close to it as possible. **Figure 3** shows the traditional layout of the GDO.

Calibration

Because the GDO also radiates a very small amount of energy, a general coverage receiver can be used to calibrate the dial. Don't try to aim for great accuracy and clutter the dial with marks and figures! If you include C10, R7 and R8 in your circuit, you can connect a frequency counter directly to the GDO and leave it in circuit all the time.

The GDO in use

Always try to place the external tuned circuit with its coil coaxial with the plug-in coil, as shown in Figs 1 and 3. If the coils are at right angles, the GDO may not produce any resonance. Set the offset control to give about 75% FSD and slowly tune L1 through its whole range. If no dip occurs, you may have the wrong coil plugged in. When you eventually find a dip, move the external coil further away until only a minute dip is seen. You may have to retune the dip meter as you do this, but it gives a much more accurate reading of frequency.

Remember that you cannot 'dip' a coil by itself – there must always be a capacitor present.

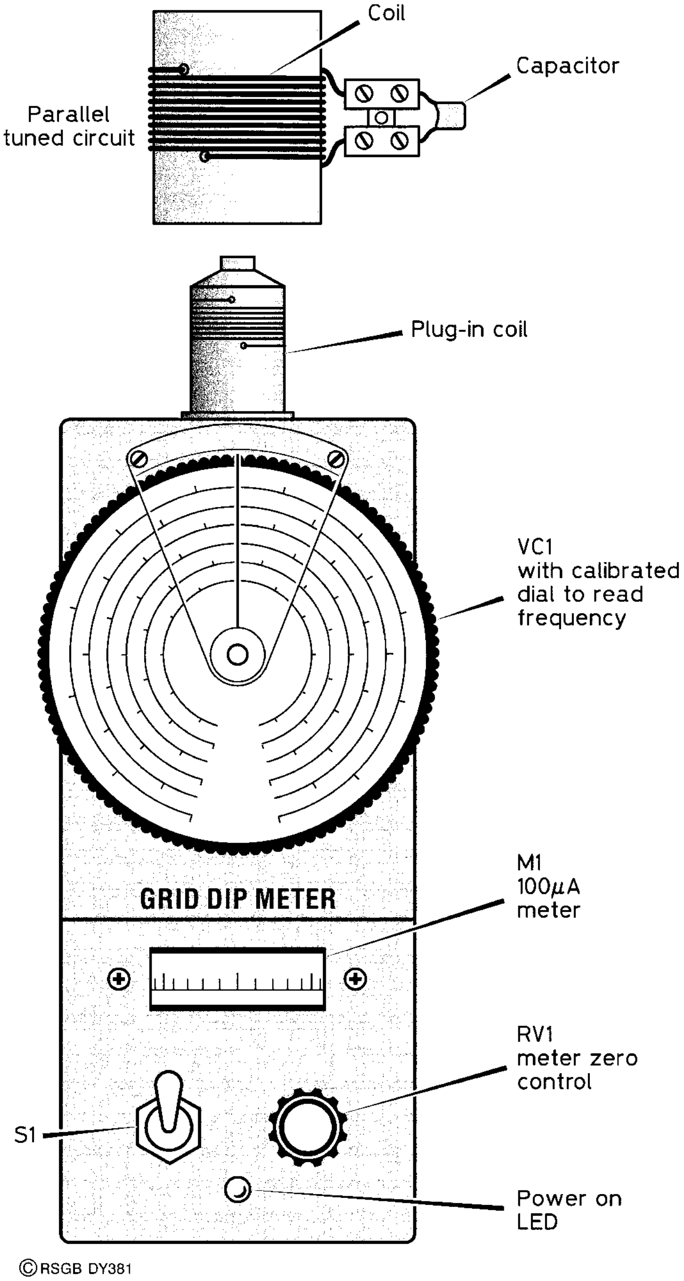
Coil
Capacitor
Parallel tuned circuit
Plug-in coil
VC1 with calibrated dial to read frequency
GRID DIP METER
M1 100μA meter
RV1 meter zero control
S1
Power on LED
©RSGB DY381

Figure 3 Layout of a typical GDO. The dial, meter and the location of the coil to the circuit under test can all be viewed at the same time. It is shown measuring resonance of a tuned circuit

Aerial resonance

Figure 4 shows how to check the resonance of a dipole aerial. Disconnect the coax feed at the aerial and place a short piece of wire, terminated with crocodile clips, across the centre insulator to short together the two ends of the aerial. By placing the GDO close to the shorting link, a dip should be seen on the meter while VC1 is turned. Alternatively, a loop in the element can be made around the coil, as in Figure 4, or the shorting link can be made long enough to loop over the coil. The latter method does not require tampering with the mounting and tensioning of the dipole wires.

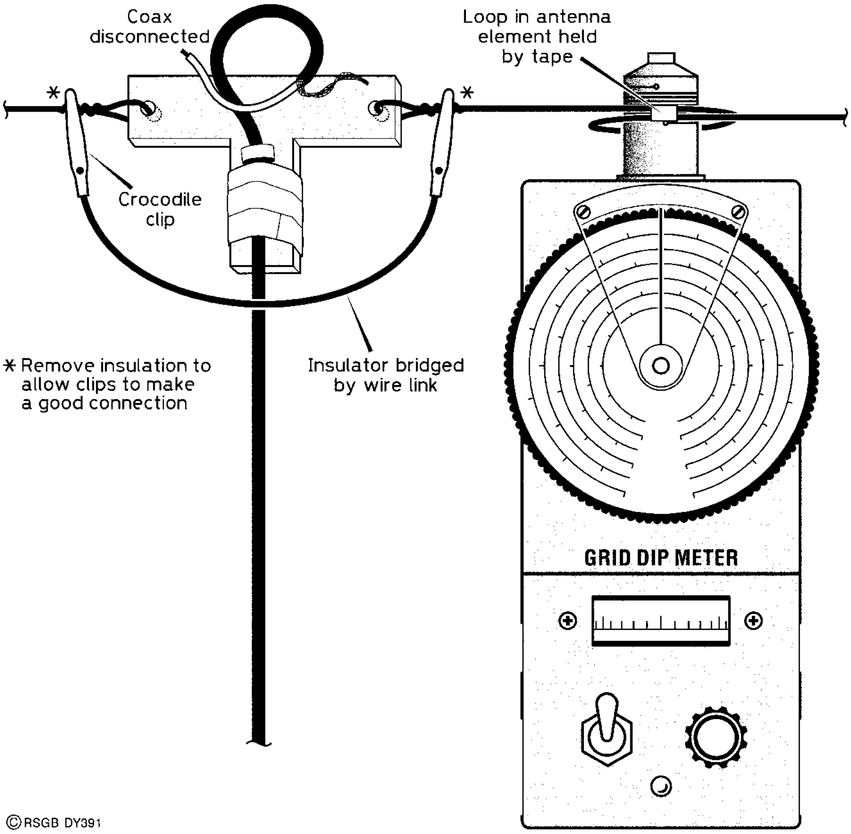
Coax
disconnected
Loop in antenna
element held
by tape
*
*
Crocodile
clip
* Remove insulation to
allow clips to make
a good connection
Insulator bridged
by wire link
GRID DIP METER
©RSGB DY391

Figure 4 A wire antenna element can be looped into a single turn coil for increased coupling to the GDO

Parts list

Resistors: all 0.25 W, 5% tolerance

R1	100 000 ohms (100 kΩ)
R2	56 ohms (56 Ω)
R3	1000 ohms (1 kΩ)
R4	5600 ohms (5.6 kΩ)
R5	560 ohms (560 Ω)
R6	4700 ohms (4.7 kΩ)
R7	1000 ohms (1 kΩ)
R8	5600 ohms (5.6 kΩ)
VR1	1000 (1 kΩ) linear

Capacitors

C1, C2	22 pF
C3, C4	100 pF
C5, C7, C9	100 nF (0.1 μF)
C6, C8	1 nF (1000 pF)
C10	6.8 pF
VC1	2 × 365 pF

Semiconductors

TR1	J304 or similar
D1	1N4148
LED	

Additional items

L1	See Figure 2
RFC	1 mH
S1	SPST
M1	100 μA

Source

Components are available from Maplin.

87 A CW transmitter for 160 to 20 metres

Introduction

This very small transmitter is designed to work on any band from top band (160 m) to 20 m, with an RF output of 1 W. It will work on higher frequencies but with a reduced output.

The circuit

The three-transistor circuit is shown in **Figure 1**. It comprises a crystal oscillator using a BC182 transistor which drives a 2N3866 power amplifier (PA) keyed by a ZTX750 PNP transistor. The oscillator and PA are coupled by a capacitor and resistor; this provides a very small amount of positive bias to the PA.

The oscillator can be used as a basic crystal oscillator but, by including a variable series capacitor as shown in Figure 1, the crystal frequency can be 'pulled' slightly, making the oscillator a *variable crystal oscillator* (VXO).

Construction

The PCB layout is shown in **Figure 2** and in the photograph. Although the prototype was built around a PCB, this circuit is equally amenable to

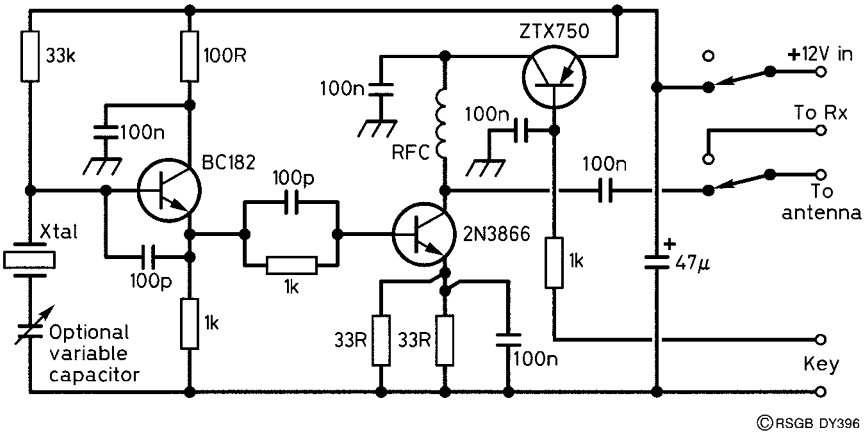
33k
100R
100n
BC182
Xtal
100p
100p
1k
1k
Optional
variable
capacitor
ZTX750
100n
RFC
100n
2N3866
33R
33R
100n
1k
100n
47µ
+12V in
To Rx
To
antenna
Key
©RSGB DY396

Figure 1 Transmitter circuit diagram

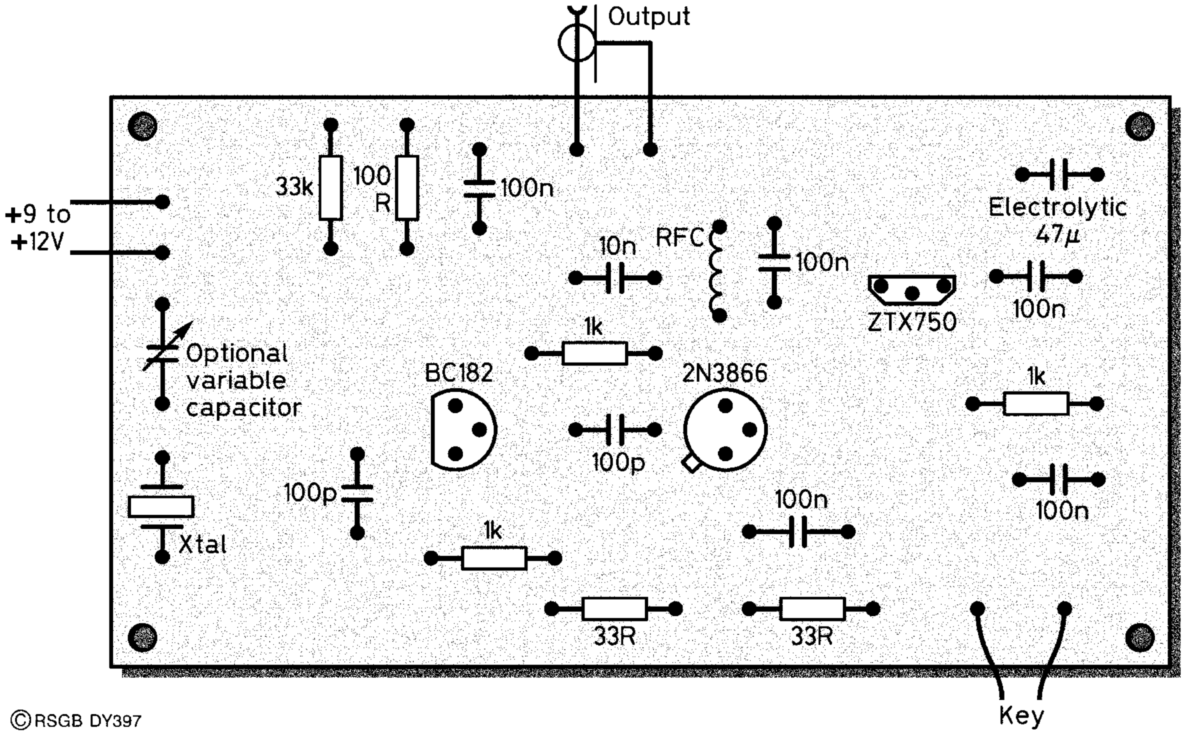
Output
+9 to
+12V
33k
100
R
100n
10n
RFC
100n
Electrolytic
47μ
100n
ZTX750
1k
Optional
variable
capacitor
BC182
2N3866
1k
100p
100p
100n
100n
Xtal
1k
33R
33R
Key
©RSGB DY397

Figure 2 Component placing on the PCB. The external connections are also shown

construction on a matrix board. Populating the PCB is very simple, and you can expect to be able to do this in about one hour. The radio-frequency choke (RFC) is made by winding 10 turns of 33 SWG wire on a ferrite bead. The enamel coating of the wire is intended to vaporise when soldered into the board, thus obviating the need to remove the enamel manually with a knife or sandpaper. However, if you do have problems with the PA either not working or keying intermittently, it is suggested that you investigate the RFC connections immediately!

If you decide to make the VXO version, you will have to cut the track between the crystal and earth, and connect the variable capacitor (250 pF) across the break.

In use

After performing the usual checks on the accuracy of your circuit building and the wiring of the external components, it is time to connect a 12 V battery between the points shown in Figure 2. Do not switch on yet. An aerial needs to be connected to the output via an ATU and a crystal, of frequency matching that of the aerial, fitted. The variable capacitor should give you a tuning range from about 14.058 to 14.064 MHz. ‘Netting’ (the process of tuning your transmitter to the same frequency as that of a received station) is achieved simply, because the oscillator is always running, and the leakage of the signal (despite the fact that the PA is not powered) is sufficient to bring the two frequencies to zero beat.

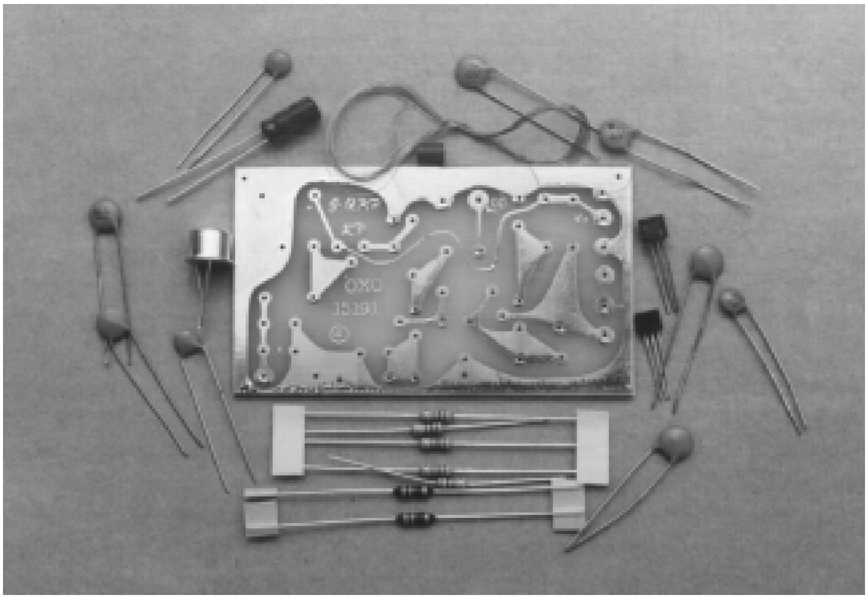

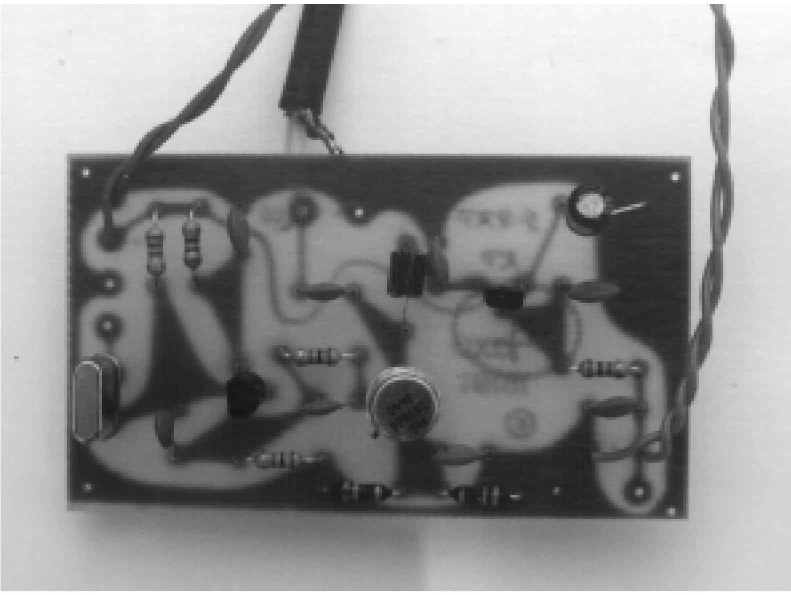

Warning

Note that the transmitter has no filtering; harmonics are not suppressed. It is strongly recommended that you use this transmitter in conjunction with the excellent low-pass filter described in the project *A 7-element low-pass filter for transmitters*, described elsewhere in this book.

88 Matching the end-fed random-wire aerial

Introduction

Many amateurs who do not have the space (or money) for a multi-band beam aerial, make use of the simplest possible alternative – the longest piece of wire that they can erect, with its end connected to the transceiver or receiver by an aerial tuning unit (ATU).

The length of the wire is not of major importance. Any length between 10 m and 80 m, with bends if necessary, will suffice. A good earth connection to the radio is just as important. Bends in the aerial wire can have some interesting effects on the directional properties of the aerial; V- or L-bends, or even a square shape are permitted. The only thing *not* permitted is to fold the wire back on itself in a tight hairpin bend!

Longish wire aerials

The term ‘long wire’ is usually used (incorrectly) to describe an end-fed aerial. How long is a piece of string? It depends what you mean by ‘long’. In aerial parlance, it means ‘long with respect to one wavelength’. Again, this depends on the band you are using. A long wire at 20 m is somewhat different from a long wire at 160 m. However, if you have sufficient real estate for a long wire on the 160 m band it must, by definition, be a long wire on all the other bands, too!

This should put you on your guard when analysing published data about feed-point impedance and the directional properties of a long wire aerial. Such theoretical data relate to a *real*, *ideal* long wire which is straight, horizontal, very high above perfect (conducting) ground, and not obstructed in any way. So your aerial doesn't quite match these criteria? Join the club!

Don't let this dampen your ardour when it comes to evaluating what the longish wire can do for you. The following should explain why.

Feed-point impedance

The impedance at the end of a longish-wire aerial can vary from a few tens of ohms to several thousand ohms, depending on the frequency in use and the wire length. It is also affected by factors such as bends, height above ground, proximity to buildings and wire diameter.

The actual value doesn't matter, provided we can make the aerial *appear* to have a 50 Ω impedance at the aerial socket of the transmitter. This process is what we call *impedance matching*, or simply *matching*. It maximises the power transfer from the transmitter to the aerial, and from the aerial to the receiver.

That is why an aerial tuning unit, or ATU, is almost (but not necessarily) obligatory.

The ATU

Many commercially produced HF receivers and transceivers have single 50 Ω coaxial sockets as their one and only means for connecting an external aerial. This means that an external aerial should have a 50 Ω feed impedance if it is to work efficiently, and it rules out most of the aerials being used by amateurs on the HF bands. Some means is necessary to change the aerial feed impedance to 'match' that of the transceiver. Such impedance-matching, or Z-matching (because Z is the symbol for impedance, just as L is the symbol for inductance) is the rôle of the ATU.

These can be bought and will accommodate either an end-fed or a coax-fed aerial. They can be bought ready for use or in kit form. Whether you want one for receiving only, or for use with a low-power (QRP) or high-power (QRO) transmitter, will determine what you need and how much you pay.

A simple single-band ATU

The simplest form of ATU is shown in **Figure 1**. It is simply a parallel LC (coil and capacitor) circuit, resonant at the chosen frequency, with taps on the coil for the aerial and the coaxial feed to the transceiver. If we assume the circuit is resonant, a high impedance exists at the top of the coil, and a low impedance at the bottom.

We said earlier that the end-fed longish wire presented an impedance which was high (or at least higher than 50 Ω). This explains why the aerial is tapped to the coil near the top, where the impedance is high, and the 50 Ω coax is tapped near the bottom, where the impedance is low.

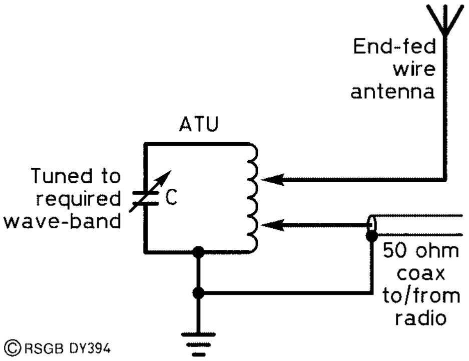

Figure 1 Parallel tuned circuit as single-band ATU for end-fed longish-wire antenna

Because the feed-point impedance of the aerial changes with frequency, so must the point at which the aerial is tapped to the coil to achieve impedance matching. The value of C must be changed also, to ensure that the circuit is resonant, and the 50 Ω tap will require tweaking also.

Setting up an ATU is quite simple. Make up an LC parallel-tuned circuit consisting of 50 turns of enamelled copper wire on an empty 35 mm film plastic container (or similar), tuned with a 500 pF variable capacitor. Make sure the enamel is removed from the ends of the wire before soldering.

Solder the inner wire of the coaxial cable from the radio to the first or second turn of the coil from its grounded end. Then solder the braid to the grounded end. Connect the aerial about one-third of the way down the coil from the top, removing the enamel at the connection point.

Adjust the variable capacitor for maximum noise or signal strength in the receiver. Then, try different tapping points from the aerial, to maximise the signal again. This matches the aerial impedance to that of the tuned circuit. Repeat the process with the coax tap, thus matching the impedance of the radio and the cable to that of the tuned circuit.

You will no doubt find that the tapping process on the coil was not easily accomplished, especially when the enamel must be removed at each tapping point without shorting adjacent turns together. It is therefore logical to produce a design where the taps have been prepared during the winding of the coil, and are selected with a rotary switch. To this end, the following multi-band design is described.

A multi-band ATU

The same type of coil design around a 35 mm film container is used (see **Figure 2**). The tapping points can be prepared in advance by a little judicious planning. Dismantle your original coil and measure the length of wire on it. It will be a little more than a calculation of $50 \times \pi D$ would suggest (where D is the diameter of the film container), due to the lack of tension in the coil and the wire diameter itself. You will need aerial tapping points at turns 1, 2, 3, 5, 10, 15, 20, 25, 30, 35, 40 and 45, corresponding to all the bands from 28 MHz to 1.8 MHz. The coaxial cable tap is fixed at turn 2 (an acceptable compromise). All the turns are counted from the earthy end.

Cut another piece of enamelled copper the same length as you used originally. Then, with the aid of an ordinary calculator, work out the positions of the points where the enamel must be removed for the taps. For example, turn 15 will have to be made $\frac{15}{50}$ of the way along the wire, turn 20 tap made at $\frac{20}{50}$ of the way along. So, if the length of wire is, say 4.7 metres, the two taps in question will be made at $\frac{15}{50} \times 4.7 = 1.41$ m and $\frac{20}{50} \times 4.7 = 1.88$ m from one end. This must be repeated for each of the tap positions, and the enamel removed ready for the wire to be soldered to it. With a

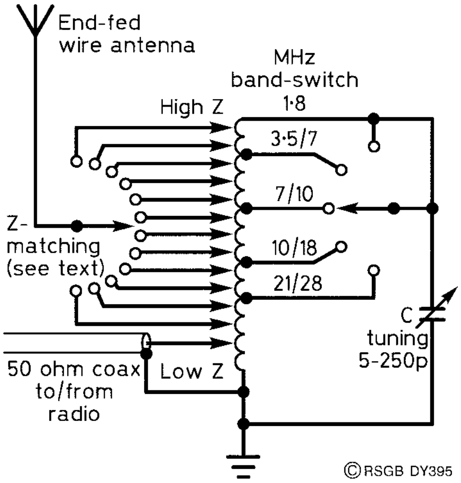

Figure 2 Simple multi-band ATU for end-fed longish-wire antennas

soldering iron, tin each tap point while the copper is shiny, thus ensuring a good, low-resistance connection. Do not solder on the taps yet.

The coil can now be wound as before. The taps can be soldered on, taking the lead from each one to the wafer of a single-pole 12-way rotary switch, the pole being connected to the aerial.

The tuning of the ATU is carried out by the same 250 pF capacitor, with a single-pole 5-way rotary switch used to select the band. Its tapping points will need to be chosen manually, using the method described earlier. Don't attempt to make new tapping points on the coil for this – use the taps available on the wafer of the other rotary switch, and find which is optimum for each band.

Notes

- For the aerial, use PVC-covered stranded tinned copper, of size 16/0.2 mm or 24/0.2 mm.
- Make the wire as long as possible, but anything over 10 m should be OK.
- Keep the wire as high as you can, in the clear and away from obstructions.
- Don't worry about bends, but don't use hairpin bends.
- Use a good insulator to attach your aerial – anything plastic will do.
- Anchor the wire near the point of entry to the building, but use a U-bend to prevent ingress of water.
- The wire can be brought in through the corner of a window, the PVC acting as an insulator. If you must drill a hole in the brickwork, make sure it slopes upwards from outside, so that water is deterred from entering.
- Use a good RF earth (as opposed to an electrical earth) such as half a dozen bare copper wires buried under the lawn in a fan shape. They should be joined together at the point of the fan and strapped to the earth connector of your ATU.

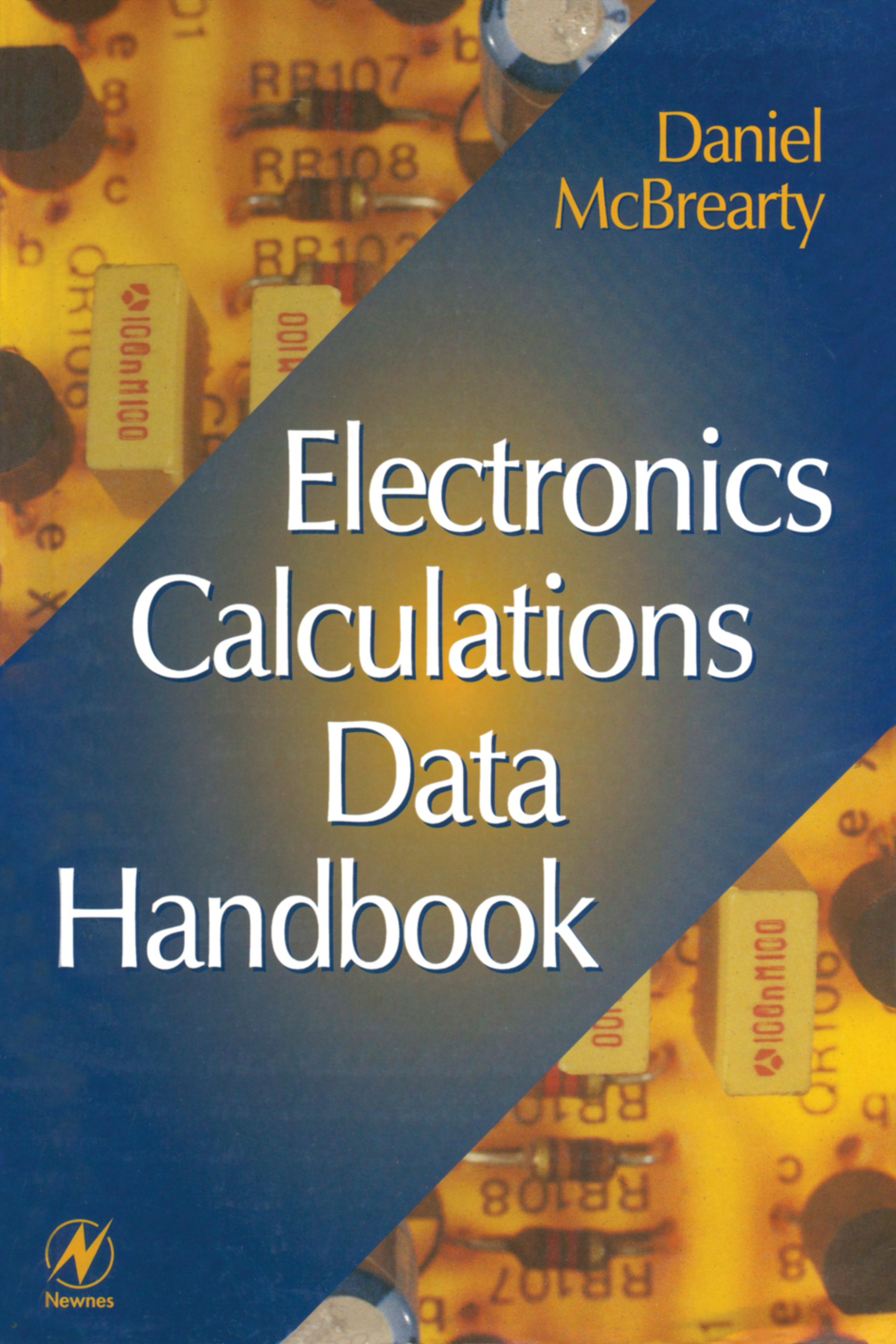

Daniel McBrearty

Electronics Calculations Data Handbook

Newnes

Electronics Calculations Data Handbook

About the author

Daniel McBrearty is a freelance electronics engineer specializing in audio systems. He was an RAF electronics technician for several years before studying for an HND at Luton College of Higher Education. Thereafter he worked for various companies including Soundcraft Electronics and Drake Electronics. He is currently involved in the redevelopment of the Royal Opera House, Covent Garden. His other interests include playing the saxophone.

Electronics Calculations Data Handbook

Daniel McBrearty

Newnes

OXFORD BOSTON JOHANNESBURG MELBOURNE NEW DELHI SINGAPORE

Newnes
An imprint of Butterworth-Heinemann
Linacre House, Jordan Hill, Oxford OX2 8DP
225 Wildwood Avenue, Woburn, MA 01801-2041
A division of Reed Educational and Professional Publishing Ltd

A member of the Reed Elsevier plc group

First published 1998

© Daniel McBrearty 1998

All rights reserved. No part of this publication may be reproduced in any material form (including photocopying or storing in any medium by electronic means and whether or not transiently or incidentally to some other use of this publication) without the written permission of the copyright holder except in accordance with the provisions of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Rd, London, England W1P 9HE. Applications for the copyright holder's written permission to reproduce any part of this publication should be addressed to the publishers.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library.

ISBN 07506 3744 7

Library of Congress Cataloguing in Publication Data
A catalogue record for this book is available from the Library of Congress.

PLANT A TREE
BTCV
British Trust for Conservation Volunteers
FOR EVERY TITLE THAT WE PUBLISH, BUTTERWORTH-HEINEMANN WILL PAY FOR BTCV TO PLANT AND CARE FOR A TREE.

Typeset by Jayvee Computer Services and Exports Pvt. Ltd., India
Printed and bound in Great Britain by Biddles Ltd., Guildford and Kings Lynn

Contents

Preface viii
Acknowledgements ix

Part One Basic Concepts 1

Introduction to Part One 3

1 Fundamentals 5
- To the beginner 5
- What is electricity? 6
- If we can't see electricity, how do we measure it? 6
- MMF and current 6
- EMF and voltage 8

2 Circuits and components 10
- Conductors and insulators 10
- Resistance 10
- Ohm's Law 10
- Circuits, diagrams and common expressions 11
- Resistors 12
- Are electrons psychic? 14
- Capacitance and capacitors 15
- Self-inductance and inductors 17
- Mutual-inductance and transformers 18

3 Direct current 20
- AC and DC signals 20
- Direct current 20
- Calculations on DC quantities 21
- Resistors in DC circuits 22
- Capacitors in DC circuits 22
- Inductors in DC circuits 22
- Power in DC circuits 23

4 Alternating current 26
  The sine wave 26
  Calculations on AC quantities 30
  Resistors and sine waves 31
  Capacitors and sine waves 31
  Inductors and sine waves 32
  Why are sine waves so important? 33
  More AC waveforms 35
  AC waveforms with a DC level 35
  Power calculations with AC waveforms 36

5 A circuit analysis toolkit 39
  Introduction 39
  Basic topologies 39
  The series circuit and Kirchhoff's voltage law 39
  The parallel circuit and Kirchhoff's current law 41
  Arrow directions for voltages and currents 42
  More about voltage and current sources 44
  Thévenin's and Norton's equivalent circuits 45

Part Two Resistive Circuits 53

6 Equivalent resistances 55
  Introduction 55
  Using the Tables for resistors 55
  Tolerance of components 56
  Using the Tables for capacitors and inductors 57
  Table 6.1 E6 equivalent resistances 58
  Table 6.2 E12 equivalent resistances 61
  Table 6.3 E24 equivalent resistances 70

7 Maximum powers for resistors 102
  Table 7.1 Maximum power for resistors up to 0.25W 103
  Table 7.2 Maximum power for resistors 0.33W to 1W 112
  Table 7.3 Maximum power for resistors 2W to 5W 121
  Table 7.4 Maximum power for resistors 6W to 10W 127
  Table 7.5 Maximum power for resistors 11W to 25W 132
  Table 7.6 Maximum power for resistors 50W to 300W 137

8 Voltage dividers 142
  Introduction 142
  Using Table 8.1 in cases where source and load impedances
    can be ignored 143

When source and load impedances cannot be ignored 143
Table 8.1 Voltage dividers 146

Part Three Reactive Circuits 203

9 First-order CR and LR circuits 205
Introduction 205
Time response 205
Frequency response 212
Table 9.1 Time and frequency constants for first-order CR circuits 216
Table 9.2 Time and frequency constants for first-order LR circuits 220
Table 9.3 Step response of first-order CR and LR circuits 221
Table 9.4 Frequency and phase response of first-order CR and LR circuits 222

10 LC circuits 223
Introduction 223
Series tuned circuit 223
Parallel tuned circuits 226
Table 10.1 Frequency constants for LC circuits 231

Part Four Operational Amplifier Circuits 257

11 Amplifier gains 259
Introduction 259
Inverting op-amp 259
Non-inverting op-amp 261
Table 11.1 Inverting op-amp gain 264

Appendix 1 Units, symbols and suffixes 285
Appendix 2 Colour codes 287
Appendix 3 The decibel 289
Appendix 4 Complex number analysis 291

Further reading 295

Index 297

Preface

This book began life as a selection of tables for electronics engineers. Dealing with such common requirements as setting op-amp gains, forming non-standard resistor values from standard pairs and setting CR time constants, it was my hope that such a book would find common favour.

In writing explanatory notes to accompany these tables, I found myself constantly wondering how much knowledge to assume on the part of the reader. After all, the book should be as useful to a beginner as to a professional with 20 years of experience.

The time came to make a decision, and Part One was born. In it I have attempted to go from first principles to the point where the reader can hope to understand the workings of the circuits discussed in Parts Two and Three with little prior knowledge. My intention has been to be non-mathematical, logical and concise. I have drawn a line at explaining the workings of electronic devices as there is not time or space, and this book only deals with them in the final chapter. An overview of the op-amp is given there. I hope that Part One provides a useful foundation for the beginner and a point of reference for the more experienced.

How well all this works as a whole will not be seen until the book is subjected to the final test by its readership. It is my hope that it will prove a valuable aid to you, whatever your level of experience. I will be happy to receive comments, criticism or suggestions for future improvement (and also to hear about any places where my spellchecker replaced 'squarewave' with 'scurvy' without my asking it to).

by mail:
Dan McBrearty
c/o Butterworth-Heinemann
Linacre House
Jordan Hill
Oxford OX2 8DP
UK

or by email:
dan@danmcb.demon.co.uk

Daniel McBrearty
December 1997

Acknowledgements

Thanks are due to Jim and Sybil McBrearty, for love and support in this; Eric Pressley and John Charlton, from whom I have learnt a great deal; all the staff at Butterworth-Heinemann for making the book a reality; William L. Bahn for helping me to convice myself that tuned circuits really should do what they do; and Hilary and Fi for putting up with my moods while I wrote!

This Page Intentionally Left Blank

Part One

Basic Concepts

Introduction to Part One

In this part of the book we take a look at the founding principles of electronics, from voltage, current and Ohm's Law to some basic ways to analyse circuits mathematically. If you know all of this then skip it, using it for reference as required. If not then read on.

I have tried to arrange ideas so that you are introduced to them sequentially, but this is a bit impractical in places. So if you are a real newcomer to electronics I'd like to suggest two readings; the first to get an overview of each topic, the second to cross-refer between them and to begin to see how they relate to each other.

This Page Intentionally Left Blank

1 Fundamentals

To the beginner

Before getting into the technical stuff, we might wish to ask ourselves what electricity is. My first recollection of thinking about it (though I did not know that I was) is of being a child and dismantling a prized radio to find out that no one was inside; just a mind-boggling collection of small coloured objects which evidently were not sweets, though some looked like them. My first direct experience of electricity was more sudden; an electric shock from a bar fire while trying to melt some plastic on it. I knew it was hot, but hadn't expected that! Later I came across the manual for a record player and amplifier in the house, which was old-fashioned enough to come with a circuit diagram. I was fascinated by it. It meant something to someone, but these hieroglyphics were like no language that I could figure out.

Later I took a technician's course, and a confusing set of concepts was presented to me as explanation for this unseen and magical force. Electricity, I had now realised, is used in a huge range of ways; recording sound, reproducing pictures, lighting our darkness and a lot else besides. All this was, they told me, due to unseen little balls which whiz around in some materials, creating equally unseen lines which can cause little balls in other places to whiz around as well. And this the mental territory of staid, rational looking people who would probably claim that they don't believe in magic.

I have to confess that, some years later, I have still not seen the little balls or the lines which they fling about, and I'm not very sure that I really understand them. But I have managed a reasonable career as a technician and an engineer, and I believe that I would have a fair chance of fixing my cassette machine, TV or house wiring should they misfunction. On the whole, I have not thought about the antics of little balls very much (though I suppose that it helps to know that they are there). Electricity looks, in my mind, more like water, wires like pipes, resistors like very thin pipes, capacitors like pairs of half-filled balloons which squash against each other, and inductors like – well something else. (I never said that the analogy was complete or accurate.)

I say all this as, I hope, a source of comfort to the new student of the subject who may be suffering the same pangs of insecurity about atoms, electrons and so on that I went through. Unless you feel strongly to do otherwise, let them live their own strange lives. You can be a very competent and effective user of the principles of electricity and electronics without needing to know much about them (though the same might not be said of completing your college course). Leave it to the physicists. What helps a great deal is having a good and clear grasp of the laws of electricity, and being able, as much as is possible, to prove them to yourself.

What is electricity?

We are never concerned with the behaviour of an individual atom in electronics; but with their movement in vast numbers, as they fling themselves, lemming-like, around the wires of your circuit. And yet we hardly think of them at all. We think a great deal about 'voltage', 'current' and about the golden rule of electronics, Ohm's Law. You will meet these soon.

Actually it is not atoms that move, but mostly bits of them, even tinier particles called electrons which separate themselves somehow from the larger atom which is their usual home. These electrons are now called 'free electrons'. Though unimaginably small, they are pretty powerful. They carry something which is called a 'negative charge' which is another way of saying that if, having been separated from the parent atom, they should ever encounter another with a vacancy then they will be in there like a shot and pretty difficult to move. And the feeling is mutual; once it has lost an electron the atom feels the loss and will try to bring close to it any waifs that it senses nearby. Being much bigger, it moves much less easily, or not at all. All of which physicists understate quite magnificently when they describe atoms which are an electron short as being 'holes'. Where a material has very few free electrons but many holes it is said to be 'positively charged'. The movement of charge is called 'electric current'.

If we can't see electricity, how do we measure it?

The answer is that we don't, directly. We measure voltage and current, we read the values written on the sides of components, and we work the rest out from there. But that's jumping the gun. Who built the meters that we use for measurement? What did they use?

Electricity is unseen. It is only through its ability to affect things physically that we know anything about it, be that people a hundred and more years ago, watching thunderstorms and making the legs of dead frogs jump, or us with our loudspeakers and computer screens. There are two ways of converting electrical energy to mechanical energy that we are concerned with.

MMF and current

MMF is most familiar to us as magnetism. The physical force that you feel between magnets is called MMF, or 'magnetomotive force'. A less familiar manifestation occurs when two wires with electric current passing through them are placed near to each other. A physical force exists between them. (It's usually too small to feel but the effect is exploited in electric motors.) An MMF is a result of a magnetic field, and the study of magnetic fields is called electromagnetics, which we won't get involved in here.

If we were to make the wires a given length and separate them by a given distance we could increase the current until we had some given measurable force. Then we would have a unit value of electrical current, the ampere or amp. The amp is abbreviated to A. The symbol which we use for electrical current in equations is I.

This in turn gives rise to a unit of measurement for electric charge, the coulomb. The coulomb is defined as the amount of charge passing a point in a wire that carries one amp of current in one second. This gives us our first equation:

$$I = \frac{\Delta Q}{\Delta t} \tag{1.1}$$

where I is current in amps, Q is charge in coulombs and t is time in seconds. (The symbol Δ means 'change in', so the term $\frac{\Delta Q}{\Delta t}$ means 'change in charge divided by change in time'–or the amount of charge which passed the point divided by the time period that charge was measured for.)

This tells us that a charge of one coulomb passing a point in one second equates to a current of one amp; two coulombs in a second, or half a coulomb in one second would be two amps; and so on. A coulomb is a lot of charge. Typically we deal in charges of thousandths of coulomb per second (mC/s), or mA. (The prefix 'm' denotes a thousandth. We will use these prefixes increasingly as we progress. They are described fully in Appendix 1.)

'Electron current' and 'conventional current' There is something of a mix-up which has taken place over history with regard to the direction of current flow, and no one has ever bothered to sort it out. You would think from what we've said above that, as the electrons do most or all of the running around in circuits, the current flow would be from the negative end of a battery to the positive end. It is, but we have a convention of always arrowing diagrams with the direction of current flow reversed. When people were discovering the laws of electricity they had some idea about things moving in wires but no way of seeing in which direction, so they had to guess. They guessed wrongly. Later this was discovered, but by this time people had been labelling currents on circuit diagrams for ages, quite happily and without any problem. The numbers all work out the same; its like deciding to rewrite your bank statement with the signs reversed. A minus sign means you are in credit or have deposited cash, and a plus that you took money out or are in debt. So long as you know what the symbols mean there is no problem. So it was with current flow, and they left it at that. It was too much hassle to change the labels on everything. If you hear talk about 'electron current' (which goes from negative to positive) and 'conventional current'

(positive to negative) this is what it means. Most people, and this book, use conventional current.

EMF and voltage

Electric charges of the same polarity repel and of opposite polarity attract. The attraction is the second physical force, called the 'electromotive force' or EMF. This is how electrons and holes know about each others' presence and are caused to move towards each other. EMF's are caused by electric fields. The study of electric fields is electrostatics and, like electromagnetics, we won't get into it here.

EMF is often called 'potential difference' or 'voltage'. It is measured in volts, abbreviated to V. When we need a symbol for voltage in equations we will also use V.

Voltage is a comparison between two places that are charged. We say that one place on a circuit is '10 V positive with respect to' another or that a potential difference of 10 V exists. The important thing to remember is that the number is always a difference between two places, or it means nothing. Generally we take one point, the metal chassis of equipment, mains earth or perhaps a piece of wire connected to the power supply, and we call it '0 V' or 'earth'. This becomes our reference for other measurements, unless we state otherwise.

The volt is defined by its ability to make charge do work (by pushing it more or less quickly through something and generating another form of energy, like heat or light). The amount of potential required to make one coulomb of charge do one joule of work is one volt:

$$V = \frac{J}{Q} \tag{1.2}$$

where V is voltage in volts, J is work in joules and Q is charge in coulombs. A joule is a measurement of energy or of work done which is derived from physical quantities. We will not actually use this but we will develop it into something more useful in Chapter 2. For now it is enough to know that a volt is a definite measurable quantity.

How fast does electricity move? When a current flows through a wire, the same electron doesn't go in one end and come out of the other; one bounces a little way along, until it is able to find a home, which it often does by dislodging another and sending it on its way down the wire. (In doing so it generates heat.) The average speed of electron movement in a wire is extremely low, fractions of millimetres per second, even for currents which are, by our standards, large.

However, what we generally mean by this question is something very different; when we apply a voltage to one end of a wire, how long is it before the voltage is felt at the other end of the wire? The answer is that the impulse travels at the speed of light – instantly for all our purposes.

An analogy is a string of marbles in a narrow tube. If you shove the end marble very hard, the impulse (the voltage) travels along the tube like a wave and reaches the other end quickly, say one metre in a second. In that time, any one marble may have only moved 1 cm. So the average marble speed (the current) is 1 cm/s while the impulse speed is 1 m/s, 100 times greater. The average marble speed changes at the other end of the tube just as quickly – they start falling out of the tube at a rate of 1 cm/s, a second later. With electrons the effect is the same, but unimaginably more exaggerated.

If that seems a bit confusing, please don't worry – it is to most people. The important thing is that changes in voltage or current travel instantly but electrons have different physical speeds. (The second fact is only relevant if you want to understand how power is dissipated in components. Even if you don't, you can get by quite well just knowing that it is. We look at this in Chapter 3.)

2 Circuits and components

Conductors and insulators

Why do some materials conduct electricity, while others don't? Some materials are populated by good and careful atoms which don't lose their electrons easily (and probably feel ashamed if they do). Hence the itinerant population is low. The few free electrons that exist tend to sit miserably in one place, having given up hope of ever finding a home. These are insulators, typically air, rubber or the fibreglass backing of a circuit board. In other materials atoms are positively feckless. There are many free electrons and should some charged material come near they move accordingly. How do they know? They feel the voltage. These are conductors. Most metals are conductors. We use copper a great deal, gold where corrosion could be a problem, and sometimes aluminium or other metals for housing equipment.

Resistance

Some materials fall into a kind of grey area between being conductors and insulators. We can use these to make resistors. It's as if the material was a kind of mud; if we make the mud thicker electrons move less easily. The resistance of a given piece of a material depends on 'how thick the mud is' (the 'resistivity' of the material) and its geometry. Making it thinner increases, and doubling or tripling length doubles or triples, its resistance.

Ohm's Law

The unit for resistance is the ohm, or , and the symbol in equations is R. Now we come to the most important equation in electronics:

$$V = IR \tag{2.1a}$$

with V in volts, I in amps, R in ohms. This is Ohm's Law. We will use it so much that it bears being rewritten in its two other forms:

$$I = \frac{V}{R} \tag{2.1b}$$

and:

$$R = \frac{V}{I} \tag{2.1c}$$

These are used so frequently that you need to know them by heart.

Ohm's Law tells us that placing a steady voltage across a material causes a proportional steady current to flow.

Even insulators and conductors have resistance. Insulators have very high resistance; tens or hundreds of MΩ or more, so that for the normal voltages that we use an unmeasurably small current flows. If the voltage is raised to sufficiently high levels, usually at least a kV, a condition called 'breakdown' occurs; electrons are blasted free from atoms within the structure of the material, and its essential nature is changed. You will know if this happens; there is a nasty burning smell, and things stop working. It is a condition which we try to avoid.

Conductors are the opposite; their resistance is very low, small fractions of an ohm typically. The commonest conductor that we use is copper. We cannot normally measure the voltage across a piece of copper wire, unless it is long. We can find out, from catalogues and so on, the resistance per metre of any given type of wire; this depends on the wire, but figures of around 0.1 Ω/m are typical. Occasionally we need to find this out and calculate the total resistance of a length of cable when our wires are very long, in a large system. The wire will also have a maximum current rating, which is what we can pass through it before the amount of heat generated in it becomes dangerous to its structure. The insulating material around it may melt, causing it to short to things that it shouldn't, or the conductor itself. We only need know that we should stay comfortably within the maximum rated current for a conductor. In essence, anything can be destroyed with too much voltage/current.

To summarize, we have three distinct classes of material; conductors, insulators and resistors. Conductors connect our circuit up and have very low resistance. We assume that the voltage is the same at both ends of them unless they are very long or current is very high. They are just vessels to carry current. Insulators are the separating material between different points in a circuit that need to be kept apart and have very high resistance. Resistors are electrical components having a known resistance. We can predict the current through them for any given voltage, or vice versa, using the all-powerful Ohm's Law, $V = IR$.

Circuits, diagrams and common expressions

A circuit is exactly what its name suggests; a loop around which electric current can flow. The loop is generally made up of conductors (bits of wire or otherwise), circuit elements or components (which we start to meet soon) and one or more sources of electrical energy – current or voltage.

Circuit diagrams, sometimes called schematic diagrams, are symbolic representations of the way that the components of a circuit are interconnected. They are the starting point for most designs, and invaluable aid in servicing or fault-finding equipment that already exists. Each element of the circuit, from a battery or a tiny resistor to a complex microprocessor chip, is represented by a symbol, with each connection to it represented by a line joining to it. Interconnections between elements are shown by lines drawn between components. In reality, these could be any form of conductor; a wire soldered directly to a leg of the component (not recommended for a microprocessor!), a copper track on a PCB, or a test lead with a crocodile clip at either end of it.

Sometimes a circuit diagram doesn't show all of the circuit, and you can't see all of the loop around which the current will flow. (For example, see Figure 2.2 later.) The inference here is that things start to happen when something else is connected to it, completing the circuit – in this case a DC voltage (as you will see).

'Block diagrams' are a special type of circuit diagram worthy of explanation. You generally find them with fairly complex electronic circuits which are either too large to see on one drawing, or more understandable by being broken into smaller parts. Each 'circuit block' is drawn as a box or some commonly understood symbol, with lines connecting the blocks. The lines represent 'signals' (in reality any means by which the individual circuits can communicate), generally one or more conductors carrying voltage or current. It is the job of the person drawing the block diagram to make it less vague than this definition!

Two more very common terms – 'short circuit' and 'open circuit'. A short circuit is what happens when you connect two points with a conductor. An open circuit between two points means that they are not connected at all (an insulator is between them).

Resistors

Figure 2.1 A resistor

We buy resistors as discrete electrical components. They come in values ranging from fractions of an ohm to tens of megohms, and are found everywhere in electronic circuits. We will see values between about 10 Ω and 1 M Ω most commonly in our circuits; values outside that do occur, but less commonly. They are identified by a series of coloured bands, each band signifying a number, or sometimes they just have the value written on. The colour code system is described in Appendix 2. The circuit symbol for a resistor is shown in Figure 2.1.

When we combine more than one resistor together in a circuit we can calculate an equivalent resistance for them, and treat them as a single component. There are two ways of doing this.

Resistors in series

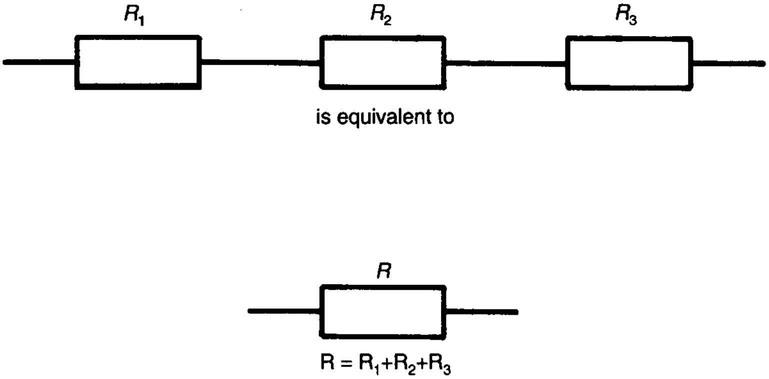

Figure 2.2 Resistors in series

Three resistors in series are shown in Figure 2.2. When we have this situation, we just add the values of the resistors to get their equivalent. We can do this for any number of resistors:

$$R = R_1 + R_2 + R_3 + \cdots + R_N \tag{2.2}$$

Example 2.1: Three resistors of 1k, 4k7 and 680 Ω in series have an equivalent resistance of 6380 Ω

Resistors in parallel

Three resistors in parallel are shown in Figure 2.3. Here we must add the reciprocal of each to get the reciprocal of the equivalent resistance. Again this applies for any number of them:

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots + \frac{1}{R_N} \tag{2.3a}$$

A special case of this worth remembering is when we have just two. Then we can calculate the equivalent in one step by dividing the product by the sum:

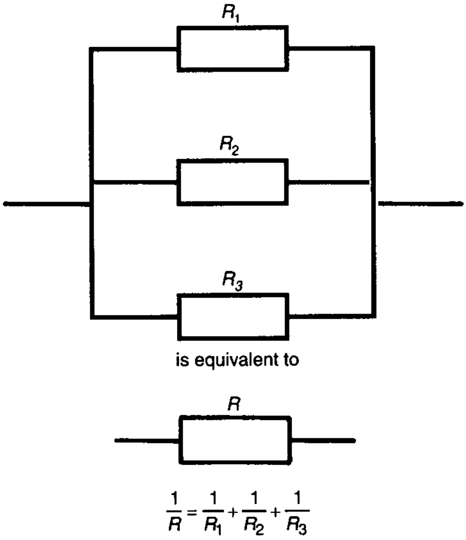

Figure 2.3 Resistors in parallel

Example 2.2: What is the equivalent parallel resistance of the three resistors from Example 2.1?

$$\frac{1}{R}=\frac{1}{1000}+\frac{1}{4700}+\frac{1}{680}=2.68\text{m}$$

$$R=\frac{1}{2.68\text{m}}=373\,\Omega$$

$$R=\frac{R_1.R_2}{R_1+R_2} \tag{2.3b}$$

We can always get some idea of equivalent resistance without calculating. Here are some rules of thumb:

1. For resistors in series the equivalent is always, obviously, larger than the largest resistor.
2. For resistors in parallel the equivalent is always smaller than the smallest.
3. Two same value resistors in parallel have a resistance which is half their value; three a third, and so on.

Are electrons psychic?

There are two ways in which electrons seem to know about what others are doing and be caused to move. One effect is due to EMF and the other to MMF. A full analysis requires excursions into the subjects of electrostatics and magnetics, but luckily we can get a good picture of these effects without all that. Both forces exist and have an application in electronics whether the charges are in the same conductor or two separate ones.

1. An EMF causes opposite charges to be attracted to each other, like ones to be repelled. If the two charges are in the same conductor this causes electrical current. If they are in two conductors separated by an insulator this gives rise to an effect known as capacitance. Capacitance is a property of a pair of isolated conductors which resists a change in the voltage between them.
2. Any electrical current creates a magnetic field, and any change in magnetic field creates a voltage in a conductor in the field. These two effects combine to create an effect known as inductance. There are two types of inductance. Self-inductance is a property of a conductor which opposes change in current flow. Mutual inductance is a property of isolated conductors by which a current flowing in one causes voltage in the other.

Capacitance and capacitors

When two conductors are separated by an insulator there is voltage between them if they have different amounts of charge. Changing that voltage requires charge to be moved, which requires energy. We can buy components specially made to have an appreciable amount of capacitance and use the effect in our circuits. Capacitors ('caps') consist of a pair of conductive plates separated by a special insulator called a 'dielectric', which magnifies the effect. When caps have no voltage across them we say that they are 'discharged' or, conversely, when they do that they are 'charged'.

The capacitance between the plates gets greater when:

1. they are moved closer together
2. they have greater surface area
3. the dielectric has a higher 'dielectric constant' – which, in English, means it is better at magnifying.

The unit of capacitance is the farad, or F, and the symbol used in equations is C. A farad is a lot of capacitance. A capacitor of one farad would have a potential of one volt across it when it had a charge of one volt on it:

$$Q = CV \quad \text{or} \quad C = \frac{Q}{V} \tag{2.4}$$

where Q is charge in coulombs, C is capacitance in farads and V is voltage in volts. This is more useful if we write it in terms of current, using Eqn(1.1):

$$I = \frac{\Delta Q}{\Delta t} = C\,\frac{\Delta V}{\Delta t} \tag{2.5}$$

This tells us that the rate of change of voltage across a capacitor is proportional to the current through it.

Capacitance exists everywhere, but is so small normally, say between two wires running side by side for a few feet, that we wouldn't bother much about it. (For cables containing more one conductor, separated by an insulator ('multicore cables') we can find out a value of 'typical capacitance between cores per metre', which we might take into account, if running signals through long cables, in the same way that we talked about for 'resistance/metre' earlier.)

Capacitors range in value from a few pF to 100 000 μF and more, with physical sizes getting larger with value. Figure 2.4 shows the circuit symbol for a capacitor.

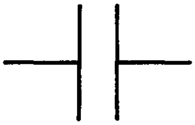

Figure 2.4 A capacitor

Capacitors come in a variety of types, with different characteristics. The most important of these are listed below:

1. Working voltage. If we put too much voltage across a capacitor, its dielectric, an insulator, will break down and current will flow from plate to plate, which it's not supposed to. The device will be destroyed. (They often, but not always, go short circuit.) The maximum working voltage tells us our safe limit. We try to design so that we never get too close to this.
2. Polarization. Large capacitors tend to be 'polarized' types – that is they use a special type of dielectric to squeeze more capacitance into a small volume, but the trade off is that you should only put voltage across them one way around. The most common type of polarized capacitor is the 'electrolytic'. They are marked with positive and negative terminals. If you disobey this by accident they go bang and bits of fluffy stuff come out of the top. Then you change them.
3. Tolerance. This varies greatly from type to type. Electrolytics have tolerances of something like −10%, +20%, because they are designed to be used in situations where you try to get as much capacitance into as small space as possible, and accuracy is not the prime consideration. At the other extreme, silvered mica types can be 1%, and give excellent stability. They are used in situations where the capacitor is used perhaps to set a time constant or a frequency (see Part Two) that must be accurately maintained.
4. Leakage current. Capacitors should be open circuit to a DC voltage. In practice you get a small current flowing, particularly for electrolytics.

There are other specifications which can come into play, but these are the most important. The value of a capacitor is usually written in numbers, in pF or in μF if no unit is specified. Unfortunately it is not always clear which, but it is usually possible to guess from the physical size, once you become familiar with them.

Capacitors in series and parallel

We can combine capacitors in series or parallel as we did with resistors. The law for caps in series is like that for resistors in parallel, and vice versa. Hence putting them in parallel creates a larger capacitor, and in series a smaller one. Here are the formulae:
In parallel:

$$C = C_1 + C_2 + C_3 + \ldots + C_N \tag{2.6}$$

In series:

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots + \frac{1}{C_N} \tag{2.7}$$

Self-inductance and inductors

Magnetic fields, when they move through a conductor, cause a voltage which can make electrons move. And electric current (or movement of electrons) causes a magnetic field. This adds up to an interaction of magnetic field, voltage and current, the net effect of which is this: if current flows in a wire there is a certain tendency for it to want to continue to do so, and not to change. (Sometimes we say this another way; when the current through an inductor changes, the inductor sets up a 'back EMF' which opposes the change in current.)

This effect is known as self-inductance (or just inductance). Like capacitance, it is not noticeable under normal circumstances. It can be made much stronger, however, if the wire is wound into a coil shape; the more turns the better. Inductance is also much increased by winding the coil on a core of some special material said to have a high 'permeability'; that is, it has the effect of concentrating the magnetic field. Such a device is called an inductor. Figure 2.5 is the circuit symbol.

The voltage across and current through an inductor are related as follows:

$$V = L\frac{\Delta I}{\Delta t} \tag{2.8}$$

where L is the inductance in its standard unit, the henry (H).

This says that the voltage across an inductor is proportional to the rate of change of current through it. We can see that this is not unlike Eqn (2.5), for a capacitor. And, indeed, capacitors and inductors are in many ways the converse of each other. One resists change in voltage, the other change in current. As we look at these components again in the following chapters, we will see that in each circumstance they do opposite things to each other.

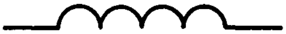

Figure 2.5 An inductor

Inductors are available with values from about 0.01 μH to 100 mH and more. Like capacitors, they get physically larger with value. As large value inductors consist of very many turns of wire on a heavy iron former, they also tend to get very heavy, one reason why they are not as popular as capacitors. They also tend to have a significant amount of resistance, which can be a disadvantage. The main specifications for an inductor are:

1. Internal resistance. This is caused by the large number of turns of wire, which can also be quite thin, to minimize the volume.
2. Maximum DC current. If too much DC current is passed through the wire of which the coil is made, it can overheat and melt, damaging or destroying the component. Even if the amount of DC is not so much as that, the core can 'saturate' – meaning that it doesn't work as well.

Despite their problems, they are quite common in some applications, and sometimes nothing else will do.

Inductors in series and parallel

Again, we can combine the inductance of series or parallel inductors and treat them as a single component. The laws are the same as those for resistors, and the inverse of those for capacitors.

In series:

$$L = L_1 + L_2 + L_3 + \ldots + L_N \tag{2.9}$$

In parallel:

$$\frac{1}{L} = \frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} + \ldots + \frac{1}{L_N} \tag{2.10}$$

Mutual-inductance and transformers

Magnetic fields also cause electrons in unconnected conductors to influence each other. This effect is called mutual inductance. To make the effect strong, coils of wire are wound together on a core (which again has the effect of concentrating the magnetic field). This is a transformer. The symbol is shown in Figure 2.6.

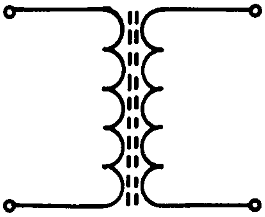

Figure 2.6 A transformer

The coil of a transformer into which you put current is called the 'primary'. The coil(s) from which you take current are the 'secondary(s)'. The 'turns ratio' of a transformer is the ratio of the number of turns on the primary to those on the secondary. When we apply AC voltage to the primary, we get AC from the secondary. And, conveniently:

$$\frac{T_p}{T_s} = \frac{V_p}{V_s} \tag{2.11}$$

where T_p and T_s are the number of turns on primary and secondary, respectively, and V_p and V_s the voltages.

Transformers are very useful in certain situations, the most common being the passing of a signal from one part of a circuit to another while keeping the two electrically isolated. Virtually all equipment connected to the AC mains supply is done so via a transformer. For safety reasons this is a special type, designed to have good electrical isolation. We *never* use anything but a proper mains transformer for this! Transformers are also very useful in many other circuits, and many other special types exist.

3 Direct current

AC and DC signals

We often hear electrical signals classified as 'AC' or 'DC'. These abbreviations stand for the terms 'alternating current' and 'direct current', but they are applied as much to voltages as to currents. In this chapter we will talk about DC. Given that we understand Ohm's Law, stated in Chapter 2, we have actually covered most of what ground there is. We will also introduce the term 'power' and consider its implications, which are important.

Direct current

A DC current is one which flows in one direction only. If we were to connect a resistor between the two terminals of a battery a DC current would result. For conventional current, it would flow from the positive terminal to the negative and have a magnitude, determined by Ohm's Law, of $I = V/R$. Similarly, we would refer to the voltage across that battery as a 'DC voltage'.

One application of DC is supplying power to an electronic circuit. The output from a power supply is often called the 'power rail' or just 'the rail'. The supply could be a separate circuit, operating from AC mains, or batteries. Virtually all circuits require one or more DC voltages to operate. A very common arrangement is to supply two rails with equal and opposite voltages, perhaps one of +15 V and one of −15 V. Then there will be three power rails; 0 V, used as a reference point for the whole circuit, and +/−15 V, used to supply power. Computer circuits often use a single rail of +5 V, perhaps with one or more other voltages used by the circuits which interface the computer to the outside world.

Circuit symbols for DC voltages and currents come in many flavours. Figure 3.1 shows some. You may just get lines, as in (a), with an indication of what the voltage or current is. (In a block diagram, like this example, the arrow may not follow the direction of current flow; it can just show that the supply is an 'input' to the circuit.) The symbol for a battery is shown at (b). The two general symbols for a voltage and current source are shown in (c). These would both indicate an unspecified device or circuit block; the intent is to indicate that a source exists at that point in the circuit, to help you see what is going on without actually telling you what physical devices to expect.

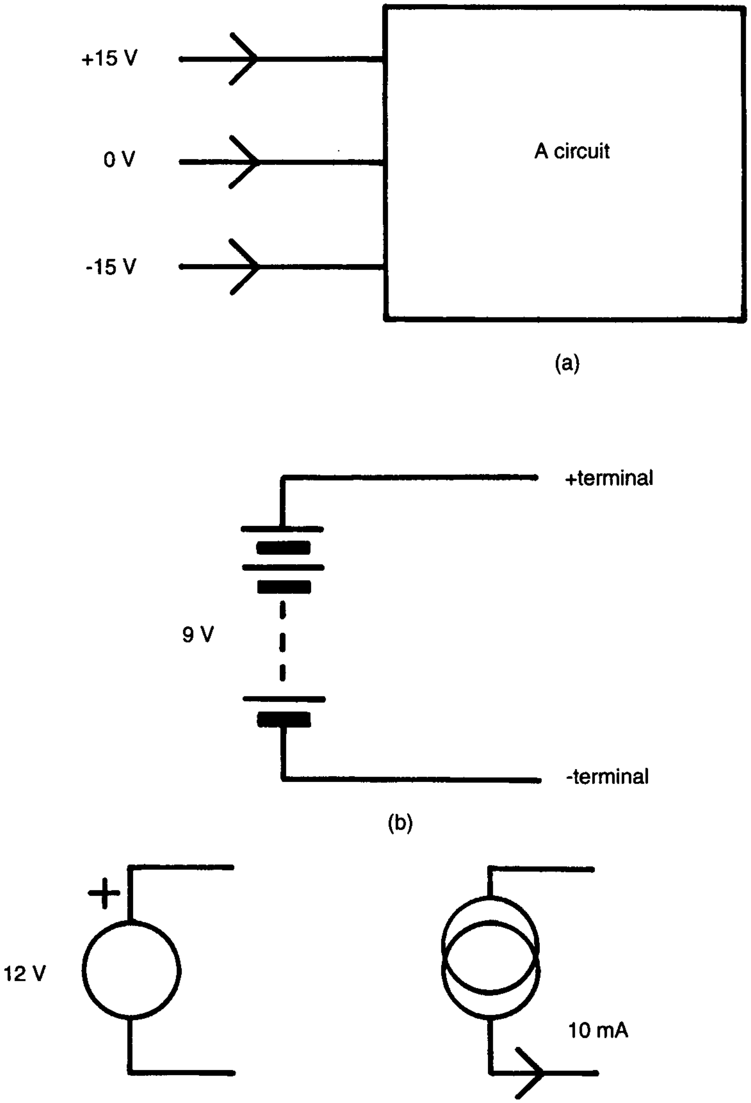
+15 V
0 V
-15 V
A circuit
+terminal
9 V
-terminal
+
12 V
10 mA

(a)

(b)

(c)

Figure 3.1 DC source symbols

Calculations on DC quantities

Addition, subtraction, multiplication and division can be performed on DC voltages and currents simply by performing the relevant operation on the numbers, paying due regard to their signs.

Resistors in DC circuits

There is hardly anything to say about this; they obey Ohm's Law, as already described. Also you should make sure that the rated power dissipation is not exceeded, as we discuss later.

Capacitors in DC circuits

If we put a DC voltage across a capacitor an instantaneous ('transient') current flows while it charges to the voltage. Thereafter no current flows as long as the voltage remains steady, with the exception of the leakage current that was discussed in Chapter 2. Capacitors are often used to isolate one part of a circuit from another to DC, while allowing AC signals to pass (as a 'DC block').

If we make a steady DC current flow into a capacitor (from a current source) the result is a ramp voltage across it. From Eqn (2.5), the ramp has a slope of:

$$\frac{\Delta V}{\Delta t} = \frac{I}{C} \tag{3.1}$$

$\frac{\Delta V}{\Delta t}$ is the slope of the ramp in volts per second, I is the current in amps and C the capacitance in farads. This is a common method of making ramp generator circuits.

Example 3.1: A steady DC current of 1 mA is fed into a 680 pF capacitor; what is the slope of the voltage across its terminals? How long does it take to change by 10 V?

$\text{slope} = \frac{I}{C} = \frac{1\text{m}}{680\,\text{p}} = 1.47\,\text{V}/\mu\text{s}$

Time taken to change 10 V is $\frac{10}{1.47} = 6.8\,\mu\text{s}$

Inductors in DC circuits

An ideal inductor would be a short to DC. A real one appears to be a resistor with the value of the internal resistance of the inductor (hopefully very low). When DC current is passed through the inductor we get a transient voltage (the back EMF mentioned in Chapter 2). Thereafter the voltage will be small. Because of this, inductors are sometimes used to block AC but pass DC (as a 'choke').

Placing a steady DC voltage across an inductor causes the current waveform to be a ramp (in the same way that a DC current in a capacitor causes a ramp voltage). The slope of the current waveform is, from Eqn (2.8):

$$\frac{\Delta I}{\Delta t} = \frac{V}{L} \tag{3.2}$$

(Did you notice how capacitors and inductors tend to do similar but opposite things to AC and DC voltages and currents?)

Power in DC circuits

Why do we worry about power dissipation?

As we saw in Chapter 2, connecting a voltage across a resistor causes a current to flow. The amount of the current depends on the value of the resistor; the lower the more current. In physical terms, this means that the average electron speed through the resistor is higher. Thus an electron making its way through strikes atoms in its path harder, and creates more heat. (We never have to worry about the electron speed, just the value of current; this is just a reminder of the physical process taking place.) The heat will be lost to the air, hopefully. How quickly the heat can dissipate depends mostly on physical structure of the resistor. If it cannot get away more quickly than it is generated, then the temperature will rise until it reaches a point where it destroys the component. Resistors go brown at this point, start to smell, and eventually go open circuit (sometimes catching fire in the process). Then you change them. Watching miniature resistors (1/8 W or 1/4 W types) catch fire isn't too bad, may even be mildly entertaining. (I have known heavy smoking engineers to use the trick to get a light in the middle of the night when the garage is shut, though I don't recommend the chemical content for your health.) Anything larger than that should definitely be avoided.

So that we can avoid this situation, resistors always have a 'maximum power rating'. Power is measured in watts (W). The power rating of a resistor might be anything from less than 0.1 W, for very small surface-mounted devices, to 100 W or more, for large, heavy devices intended to be bolted to 'heatsinks' (pieces of metal designed to radiate and make heat dissipate more quickly). On opening up pieces of electronic equipment we may sometimes see resistors which have obviously been overheating; the painted exterior may be brown and cracked. This may be the cause of the fault that we are investigating or an indication that some other fault is causing excessive current to flow. Sometimes it is an indication that the designer of the equipment wasn't as meticulous as he might have been. A resistor run continually at its maximum power may show signs of burning without misfunctioning, but the long-term reliability is likely to be reduced.

I have tried to follow a 'rule-of-thumb' which I learnt from an experienced engineer early in my career; to try not to run components continually at more than about one-third to one-half of their rated power. Running them at maximum for a short duration is generally OK; the device has time to cool off. If in doubt, prototype and use common sense. If it goes brown, you are being a bit ambitious.

Power calculation in DC circuits

How do we calculate the power dissipation in a resistor? Power is defined as the rate of energy dissipation, or:

$$P = \frac{J}{t} \tag{3.3a}$$

where P is the power in watts, J is the energy in joules and t is the time in seconds. From Eqn (1.2):

$$V = \frac{J}{Q} \quad \text{or} \quad J = VQ$$

Substituting into Eqn (3.3), we get:

$$P = V\frac{Q}{t} \tag{3.3b}$$

From Eqn (1.1):

$$I = \frac{Q}{t}$$

So:

$$P = VI \tag{3.4a}$$

Like Ohm's Law, this gets used a lot, and you need to know it. (The derivation is just for interest's sake.) The result is important enough to be restated, and learnt, in its two other forms:

$$I = \frac{P}{V} \tag{3.4b}$$

and:

$$V = \frac{P}{I}$$

Example 3.1: A resistor of 10 Ω has a voltage of 3 V across it. How much power does it consume?

Using Ohm's Law

$$I = \frac{V}{R} = 0.3 \text{ A}$$

and using Eqn (3.4a):

$$P = VI = 3 \times 0.3 = 0.9 \text{ W}$$

Often we use two equations which we can get from Ohm's Law and Eqn (3.4a) by a little algebraic substitution:

$$P = \frac{V^2}{R} \qquad (3.4d)$$

and

$$P = I^2 R \qquad (3.4e)$$

These enable us to answer questions like that posed in Example 3.1 in one step. Using Eqn (3.4d):

$$P = \frac{3^2}{10} = 0.9\,\text{W}$$

The business of using components within their maximum power rating quickly becomes fairly intuitive. For a circuit working from low voltages, say +/– 15 V DC, the maximum available voltage is 30 V. Our workshop resistor kit may contain miniature resistors all rated at 1/8 W. This enables us to calculate a value of resistor above which we cannot, no matter how we try, exceed the rated power.

Example 3.2: For the circumstances which we just described, what is the 'minimum' value, following the 'half-rated-power' rule-of-thumb?

$$P = \frac{V^2}{R} \quad \text{so} \quad R = \frac{V^2}{P}$$

P is half of 1/8:

$$\frac{0.125}{2} = 0.0625\,\text{W}$$

$V = 30$ V

Our threshold value of resistance is:

$$\frac{30^2}{0.0625} = 14400\,\Omega$$

– or say 15 kΩ.

This does not mean, of course, that we won't use any components of less than this value; but that above this value we can dot them around without worrying about power rating. At values around 15 kΩ, we only need worry if they spend a lot of time strapped across the power rails, which very few will. Even for low values, we will tend only to get the calculator out when we know that the resistor has a large voltage across it.

4 Alternating current

An AC voltage or current is any which is changing. It will very often be one which changes at a known rate. Where a DC voltage or current is easily described by its magnitude and sign ('3 V DC' or '–16.2 V DC' says it all, as long as we know where 0 V is), AC signals are a bit more complicated. We are going to start by considering the simplest type of AC signal. Then we will look at more complex ones.

The most ubiquitous AC signal is the sine wave. We will understand so much electronics in terms of how circuits respond to sine waves that it is worth our while spending some time describing and thinking about them.

The sine wave

If you haven't known about sine waves before, the word 'sine' probably brings back dimly remembered memories of school maths lessons, trigonometry, and all that goes with it. The sine is a mathematical function which describes the ratio of two sides of a right angled triangle in relation to one of the other angles, and a sine wave is a graph of the ratio against the angle. That's putting it mathematically. To many technicians and engineers, who try to do as much practical work as possible with the minimum use of mathematics, a sine wave is a very useful 'wiggle' which appears on the screens of an oscilloscope (a device which cleverly shows a graph of voltage against time) when they are working on equipment. We describe a great number of the essential properties of circuits in terms of what comes out when the input is a sine wave. The reasons for this are discussed later.

Figure 4.1 shows a sine wave. It is 2 V peak to peak (pk–pk or p–p) with a frequency of 1 Hz. (We'll define these terms soon.) The voltage begins at a level of 0 V, rises smoothly up to a level of 1 V, back down through 0 V to –1 V, and up to 0 V again. Each transition, 0 V to 1 V, 1 V to 0 V, 0 V to –1 V and –1 V to 0 V, takes the same amount of time, 0.25 s. The whole process is referred to as 'one cycle'.

We don't often come across sine waves which exist for one cycle only and then stop, however; usually the voltage oscillates continually while the circuit generating it is on. This may be for many hours. We call this a 'continuous signal'.

There are two things that we need to describe a sine wave unambiguously; how big it is, and how long it takes to complete one cycle. These two quantities we normally call 'amplitude' and either 'period' or 'frequency'.

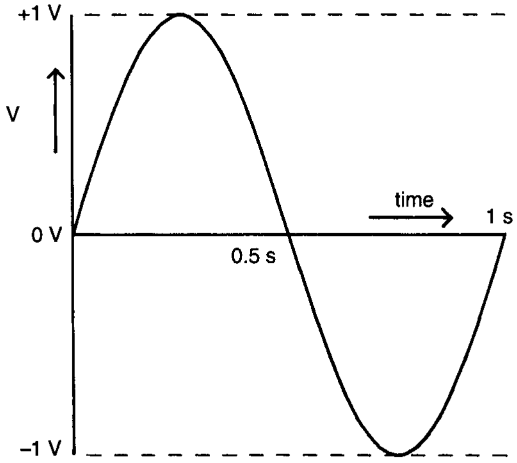

Figure 4.1 A sine wave

Amplitude

There are three ways of defining the amplitude of a sine wave:

1. Peak to peak amplitude is the difference between the positive and the negative peaks. Figure 4.1 is a sine wave of 2 V pk–pk. Pk–pk amplitude can be used to describe signals which are not sine waves as well.
2. Peak amplitude is the difference between the mid-point of the signal and the positive peak. Figure 4.1 is a sine wave of (1–0) = 1 V pk. We can say that:

$$\text{pk amplitude} = \frac{\text{pk} - \text{pk amplitude}}{2} \tag{4.1}$$

3. Root mean square (RMS) amplitude is the level of a DC signal which has the same energy content as the AC signal. This odd way of specifying amplitude exists because it is needed for power calculations (on which more later in this chapter). For a sine wave:

$$\text{RMS amplitude} = \frac{\text{pk amplitude}}{\sqrt{2}}$$

or:

$$\text{RMS amplitude} = \text{pk amplitude} \times 0.707 \tag{4.2a}$$

or again:

$$\text{pk amplitude} = \text{RMS amplitude} \times 1.414 \tag{4.2b}$$

Which of these three we use depends on the situation, and we often need to convert between them. For instance, we might see a sine wave displayed on the screen of our oscilloscope, and wish to calculate the power which it causes to be dissipated in a certain component. Here we might read the pk–pk amplitude from the display, divide by two to get pk amplitude and then multiply by 0.707 to get the RMS amplitude. Then we can make a power calculation. Conversely, we might measure the amplitude of a sine wave using an RMS multimeter. If we needed for some reason to have a rough idea of the pk–pk value of the voltage, we could divide by 0.7 (multiply by 1.4) and then multiply by 2 in our heads – multiply by 2.8.

Period and frequency

The period of a sine wave is the time which it takes to complete one cycle. Its symbol in equations is T, and its unit is that of time, the second (s). For Figure 4.1, T is one second.

It is more usual to refer to the 'frequency' of a sine wave. The symbol for frequency in equations is f, and its unit is the hertz, abbreviated to Hz, meaning cycles per second. The frequency is the number of cycles completed in one second. Converting from frequency to period is very easy:

$$f = \frac{1}{T} \quad \text{or} \quad T = \frac{1}{f} \tag{4.3}$$

with f in Hz and T in s.

So our signal of Figure 4.1 has a frequency of 1 Hz.

Amplitude and frequency are independent of each other; a sine wave of 3 kHz will oscillate 3000 times per second whether it has a level of 1 μV or 10 V. One of 10 V RMS will still be at 10 V RMS, and have the same heating effect through a given resistor whether it is at 10 Hz or 1 MHz.

(An interesting aside to this second case; if the frequency were 0.01 Hz, a period of 100 s, would this still be true? No; remember, from Chapter 3, the mechanism of heating and cooling of resistive components. In this case, the frequency is so low that we would do better to consider it as a slowly varying DC voltage. For a quarter of the period, 25 s, the resistor will have a voltage across it of 20 V or greater! This time is obviously long enough to destroy the resistor if it is not adequately rated. In our power calculations, we should budget for the pk amplitude of about 14 V. This is a very extreme example, and such slow moving signals are not often encountered. The statement in the above paragraph holds true for frequencies of a few Hz and above.)

Phase

We can describe a single sine wave completely by specifying its amplitude and frequency. When we are talking about two such signals the important question of phase arises.

Figure 4.2 shows the sine wave of Figure 4.1 again, labelled V_1, and a second nearly identical sine wave which starts its cycle at a later time t, labelled V_2. We say that there is a phase difference between V_1 and V_2.

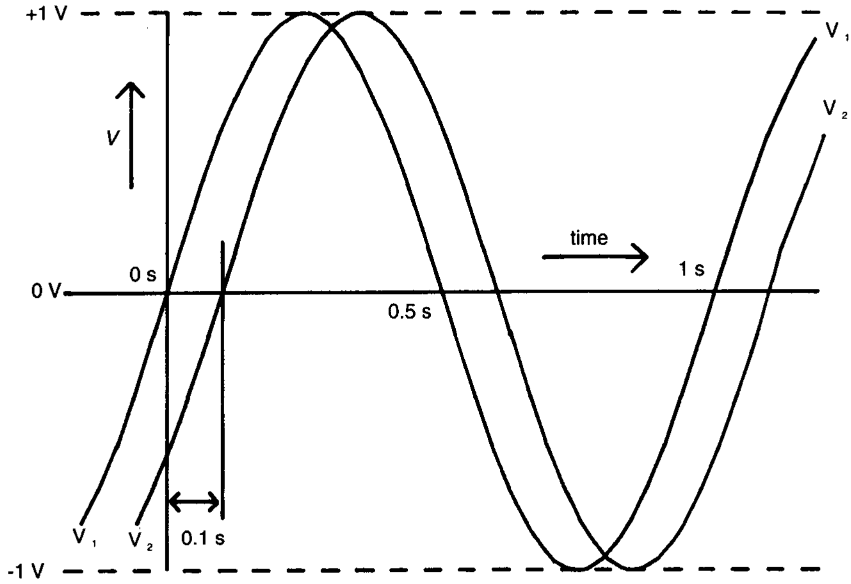

Figure 4.2 Illustration of phase

Phase is an angular quantity, and its units are degrees or radians. We use degrees in this book. There are two important points to bear in mind when talking about phase:

1. We only generally make phase comparisons between waveforms of the same frequency.
2. Phase, like voltage, is a relative quantity. It is meaningless to say that, for instance, 'the voltage has a phase angle of 30°' unless we specify that that 30° refers to some other waveform. A reference must be specified.

To calculate the phase difference between two waveforms, we use the following expression:

$$\phi = \frac{360t}{T} \tag{4.4a}$$

We can, of course, calculate t from ϕ and T:

$$t = \frac{T\phi}{360} \tag{4.4b}$$

where ϕ is the phase in degrees, t is the time difference in seconds and T is the period in seconds. For example, if t in Figure 3.2 is 100 ms, then $\phi = 36°$. V_2 starts 100 ms after V_1, so we can either say that V_2 lags V_1 by 36° or that V_1 leads V_2 by 36°. The two statements mean the same thing. We might, if we were being awkward, say that V_2 leads V_1 by (360–36)° or 324°. This statement means that V_2 starts its cycle 324°, or 900 ms, before V_1 – just as true if they are continuous, but not the easiest way to say it. Normally we specify phase using the smallest possible angle, which is always less than 180°. Sometimes calculations give us numbers larger than this, and we convert them to a more sensible form.

Waveforms which are 90° out of phase are said to be in quadrature, while ones which are 180° out of phase are said to be in antiphase. Putting a waveform in antiphase is 'inverting' it. Waveforms which have no phase difference are simply 'in phase'.

Calculations on AC quantities

We can do calculations on AC quantities of the same frequency to analyse the behaviour of a circuit at that frequency. Every quantity that we use will have two components; magnitude (amplitude) and phase.

Mathematical operations on waveforms of any phase are slightly tricky and require the use of something called 'phasor mathematics'. Ideally, a mathematical technique called 'complex algebra' is used which makes use of a notional quantity called j, equal to $\sqrt{(-1)}$ to denote a phase shift of 90°. This technique allows not only voltages and currents but also combinations of reactances, resistances and even amplification to be treated under one umbrella. The behaviour of circuits can then be predicted at a fixed frequency.

The algebra involved is actually not much more difficult than the more conventional techniques which the next chapter uses, but a complete explanation of it is somewhat beyond the scope of this book. Should you already be familiar with it then Appendix 4 gives a quick restatement of its principle techniques.

Even if you are not (or don't want to bother with it) its not really a problem. There are some simple rules and special cases of phasor mathematics which are presented below. In Part Three we will consider what some basic circuits do under these conditions and hence get a lot of insight into their workings. You could take these ideas on board now, or refer back to this point when you get there. Anyway, here goes:

Phase Rule 1:	For waveforms in phase, magnitudes of quantities can be used in calculations directly. The phase of the result is the same as that of the 'inputs' to the calculation.
Phase Rule 2:	A positive value of phase means 'leading', a negative one 'lagging'.
Phase Rule 3:	A negative value of magnitude means an inversion of the quantity

(i.e. it is in antiphase). So subtraction is addition with one waveform inverted. We could say that $-V$ is V at an angle of 180°.

Phase Rule 4: Adding two equal waveforms in antiphase gives a result of zero. This is called 'phase cancellation'.

Phase Rule 5: Multiplication of any pair of waveforms. The magnitudes are multiplied and the phase angles added together.

Phase Rule 6: Division of any pair of waveforms. The magnitudes are divided and the phase of the denominator subtracted from the phase of the numerator.

Resistors and sine waves

The terms pk, pk–pk and RMS can be used to refer to both voltages and currents. Provided that we are consistent we can do Ohm's Law calculations on resistors with any of them. So, for example:

1. A pk voltage of 10 V across a 1k resistor causes a pk current of $\frac{10}{1000} = 10\,\mathrm{mA}$ to flow.
2. A pk–pk current of 3 mA through a 4k7 resistor causes a pk–pk voltage of $3 \times 4\mathrm{k}7 = 14.1\,\mathrm{V}$ across it.
3. An RMS voltage of 150 V across a 1M resistor causes an RMS current of $\frac{150}{1\mathrm{M}} = 150\,\mu\mathrm{A}$ to flow.

Current and voltage through a resistor are always in phase. In fact, the current waveform is always an exact replica of the voltage waveform, its magnitude at every point being determined by Ohm's Law.

Capacitors and sine waves

A capacitor with a sinusoidal voltage of frequency f across it will have a sinusoidal current flowing through it. The ratio of the voltage to the current is known as the 'reactance' of the capacitor at frequency f. The situation is analogous to that with a resistor, and the unit of reactance is again ohms. And Ohm's Law again applies:

$$V_C = I_C X_C \tag{4.5}$$

where X_C is capacitive reactance, V_C is the voltage across the capacitor and I_C is the current through it. As with resistors, amplitudes of voltage and current can be specified using pk, pk–pk or RMS conventions, so long as the same is used for both.

We can calculate the reactance of a capacitor at any particular frequency using the expression:

$$X_C = \frac{1}{2\pi fC} \tag{4.6}$$

where C is the capacitance in farads and f is the frequency. We can see from this that the magnitude of the reactance of a capacitor decreases proportionally with frequency.

But hold on! Capacitors are more than 'frequency-dependent resistors'. They do something important to AC signals. The current through a capacitor always leads the voltage across it by 90°. This is the difference between a capacitor and a resistance with the same value as its reactance at that frequency.

We can often get a very good idea of what a circuit does by thinking of caps as 'frequency-dependent resistors', but we can't do calculations to solve CR networks (using techniques from Chapter 5) without taking account of phase. (A CR network is any consisting of capacitors and resistors – we'll meet CR, LR and LCR networks in Part Two.)

For calculation purposes, we consider that capacitive reactance has a phase angle of −90°, as passing a current with a phase angle of 0° through it results in a voltage of 90° lagging. (See Phase Rules 5 and 6 earlier in this chapter.) (It is this idea of assigning phase to reactance that makes otherwise incomprehensible complex numbers work.)

Now that we've met the terms resistance and reactance, this seems like the moment to mention the word 'impedance', which is used a lot. Strictly speaking, resistors have resistance, capacitors (and inductors) have reactance. Circuits have some combination of the two. If you want to be non-committal about which you mean, say 'impedance' – it can be either. And everyone should understand.

Inductors and sine waves

Like capacitors, the sinusoidal voltage across and current through an inductor are proportional at any given frequency. The ratio is again known as the reactance of the inductor, and here's Ohm's Law again:

$$V_L = I_L X_L \tag{4.7}$$

with X_L as inductive reactance, V_L voltage across the inductor and I_L current through it. Again we may use pk, pk–pk or RMS amplitudes but must be consistent. For an inductor voltage leads current by 90°, so we consider that the reactance has a phase angle of 90°. The inductive reactance can be calculated using the expression:

$$X_L = 2\pi fL \tag{4.8}$$

where L is the inductance in henries and f is the frequency in hertz. From this, inductive reactance increases proportionally with frequency.

(Again we see capacitors and inductors doing nice complementary things to each other, while resistors form a nice 'neutral background' against which to view it all.)

Why are sine waves so important?

In other words, why do we do so much of our analysis of AC circuits in terms of what they do when you put a sine wave through them?

All AC waveforms can be considered as being the sum of sine waves of various frequencies. The mathematical technique behind this is known as Fourier analysis. We are not going to get into the mathematics here, but we will consider the implications.

What the maths says is that we can treat any continuous periodic waveform (such as all the ones in Figure 4.3) as the sum of a number of sine waves of various frequencies. The frequencies will all be whole number multiples of the frequency of the repeating periodic waveform. So a 1 kHz triangle wave can be considered to be the sum of sine waves of 1 kHz, 2 kHz, 3 kHz . . . and so on. The amplitudes and phases of the sine waves will vary, generally getting less with frequency, but we could, if we knew them, and had the means to do it, reconstitute the waveform, like baking a cake to a recipe. For instance, the 'recipe' for a square wave of 1kHz is:

1 kHz sine wave, at 1 V RMS, plus ...
3 kHz sine wave, at $\frac{1}{3}$ V RMS, plus ...
5 kHz sine wave, at $\frac{1}{5}$ V RMS, plus ...
7 kHz sine wave, at $\frac{1}{7}$ V RMS, plus ...
. . .
n kHz sine wave, at $\frac{1}{n}V_{RMS}$.

'*n*' can be any whole number as high as you care to keep going. The higher in frequency you continue, the better the shape of the resulting wave – that is the closer to the ideal, with the voltage changing from one voltage to the other almost instantaneously. Real square waves are not ideal, having a finite 'risetime' – the time taken to get from 10 to 90% of the value of the step.

You will notice in our square wave 'recipe' (known more properly as a Fourier series) that we only have the odd number multiples of 1 kHz. The 1 kHz waveform is known as the 'fundamental' of the series (i.e. the sine wave at the same frequency as the periodic waveform being analysed). The sine waves at 2 kHz, 3 kHz, and so on, are known as the 'harmonics' of the fundamental. 2 kHz would be the 2nd harmonic, 3 kHz the third, and so on. So we can say that 'a square wave consists only of the fundamental plus odd harmonics'. It contains no even harmonics. The exact amplitude of the result can be foretold, but we are not often worried about it.

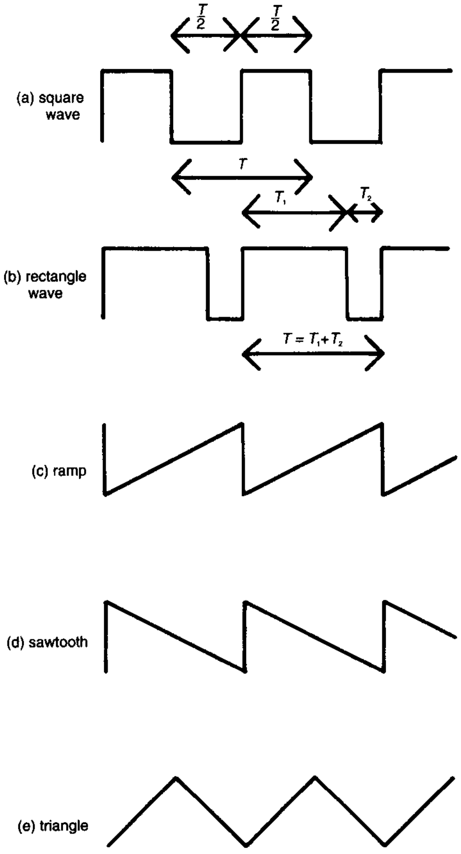

Figure 4.3 More AC waveforms

You can get tables showing the equations to write various waveforms in terms of their harmonics, but we rarely need them. Quite a lot can be done intuitively. Where a waveform has a sharp transition between two levels, that's where you get the high harmonics. The 'frequency response' of a circuit is the way that its output, relative to the input, varies with frequency. If a circuit, such as a low-pass filter (of which more in Part Two) tends to attenuate high frequencies then it will affect the edges of a square wave – they will be slower. If we know the frequency at which the filter begins to 'cut', then we can get an idea of how serious this is – for example if our square wave is at 1 kHz, and the filter starts to 'roll off' from 2 kHz, the effect will be quite noticeable. If the filter rolls off from 20 kHz, much less so. Conversely, the low frequencies of a waveform contribute most to the portions where it changes slowly, or not at all; the top and bottom of a square wave will tend to 'sag'. The time responses of 1st order circuits, which we look at in Chapter 9 are in fact extreme examples of this.

More AC waveforms

Figure 4.3 shows some types of AC signal which are commonly encountered. There is not too much to know about them, except to recognize their shapes and their names. With rectangle waves we sometimes hear talk of mark-to-space ratio. This is the ratio of duration of the more positive part to the more negative – $T_1 : T_2$ on the diagram. These waveforms are best specified, amplitude-wise, in terms of pk–pk level, again the difference between the positive and the negative voltage. Time-wise, we obviously refer to them by either their frequency or their period, with $f = 1/T$ once again.

AC waveforms with a DC level

It often happens that we come across signals which are a combination of AC and DC. Then we say that the waveform has a 'DC level'. The DC level of an AC waveform is just its average level. Hence if it has an equal area below 0 V as above it has a DC level of zero.

Sine waves, square waves, ramps, sawtooths and triangles have a DC level of zero if they move between equal positive and negative voltages. Rectangle waves with uneven mark-to-space ratios (if they didn't have they'd be square waves) will go more positive than negative (ratio less than 1:1) or more negative than positive (more than 1:1) to have no DC level. In fact changing their mark-to-space ratio changes their DC level too, assuming that they stay between the same two voltages.

Figure 4.4 shows some examples. The triangle wave goes between 1 and 9 V and so it has a DC level of 5 V. This is the case regardless of the duration of its ramps. (Sketch one of any shape on squared paper and then draw a line horizontally through its mid point. You should see that its area above the line is the same as that below. This is true for sines, squares, sawtooths and ramps as well.) So the DC level of any sine, square, ramp, sawtooth or triangle is:

$$V_{DC} = \frac{V_+ + V_-}{2} \tag{4.9}$$

For the triangle in Figure 4.4, $V_{DC} = (9 + 1)/2 = 5\text{V}$.

The rectangle wave goes between +5 V and −5 V. But it has a mark-to-space ratio of 3:1, and so has a positive DC level, as the area above 0 V is greater than that below. The DC level is given by:

$$V_{DC} = \frac{T_M}{T} V_M + \frac{T_S}{T} V_S \tag{4.10}$$

T_M is the duration of the mark, V_M its voltage, T_S and V_S the same for the space. T is the period, $T_M + T_S$. So for our example:

$$V_{DC} = 0.75 \times 5 - 0.25 \times 5 = 2.5\,\text{V}$$

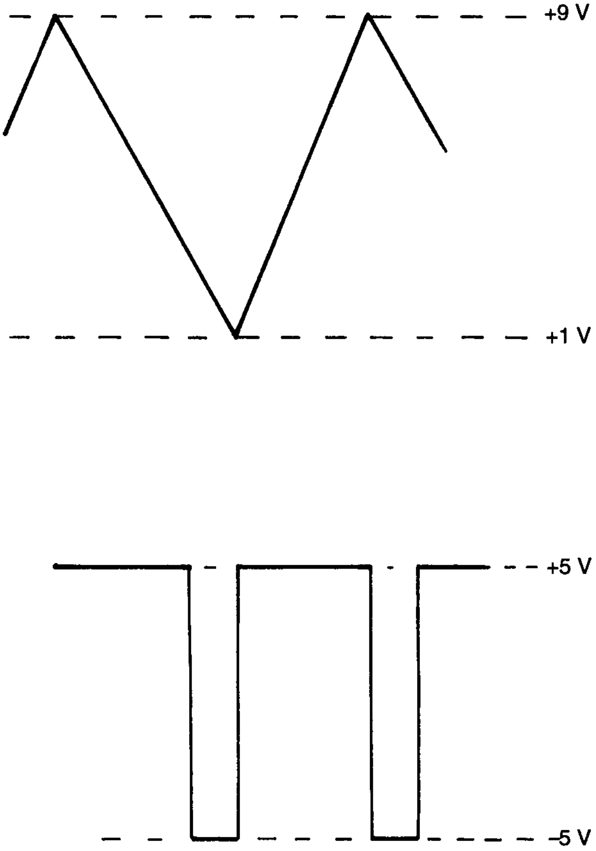

Figure 4.4 AC waveforms with DC level

Power calculations with AC waveforms

For power calculations involving AC waveforms, we must use the RMS value of the signal. RMS stands for 'root-mean-square'. It is what we get when we take the average of the square of the amplitude of a waveform over its cycle, and then take the square root of that. If that sounds like a rigmarole, it is, but we never actually have to do it.

Sine waves

We already know how to find the RMS value of a sine wave from our discussions on amplitude earlier in this chapter. When calculating the power that a sinusoidal signal develops in a load we must also take account of the phase angle between voltage across and current through that load.

The power dissipated by a circuit element which is passing a sinusoidal current (with no DC) is defined as:

$$P = VI \cos\phi \tag{4.11}$$

where P is the power in watts, V is the RMS voltage across the component in volts, I is the RMS current in amps and ϕ is the phase angle between voltage and current.

The $\cos\phi$ term varies between 1 when $\phi = 0°$ and 0 when $\phi = 90°$ (or $-90°$). Hence for resistors, for which $\phi = 0°$ always, we simply use the same power equations that were given for DC circuits in Chapter 3, of course using RMS values. Equally we can say that ideal capacitors and inductors, for which $\phi = 90°$ (or $-90°$), don't dissipate power. When we use circuits made up of combinations of these components, and ϕ is between 0 and 90°, the power will be somewhere between zero and 'volts-times-amps'. But all of that power will be dissipated in the resistive components of the circuit.

(An aside: the term 'peak power' is sometimes heard, but has no relationship to 'pk amplitude'. It is used in such things as sound amplification systems to describe the the maximum amount of electrical power delivered to a loudspeaker at any given moment. There is no such thing as pk–pk power. Thought I'd mention it.)

RMS values of other waveforms

Its not often that we have to calculate these, but if we do it's nice to know how. If you're a beginner you may well prefer to skip this. I just couldn't resist putting them in.

For any triangle, ramp or sawtooth wave with no DC level, and a positive excursion, V_{PK}:

$$RMS_{\text{triangle}} = \frac{V_{PK}}{\sqrt{3}} = 0.577\ V_{PK} \tag{4.12}$$

For any rectangle wave (of which a square wave is just a special case), if T_M is the duration of the mark, V_M its voltage, T_S the duration of the space, V_S its voltage and T the period $(T_M + T_S)$:

$$RMS_{\text{rectangle}} = \sqrt{\left(V_M^2 \frac{T_M}{T} + V_S^2 \frac{T_S}{T}\right)} \tag{4.13}$$

with V_M the voltage of the mark and V_S that of the space. This takes into account the DC level.

For a signal with AC and DC components:

$$RMS_{\text{AC+DC}} = \sqrt{\left(V_{\text{DC}}^2 + V_{\text{RMSAC}}^2\right)} \tag{4.14}$$

And finally, for a 'piecewise' signal with known RMS values over different parts of its period

$$RMS_{\text{total}} = \sqrt{\left(V1_{\text{RMS}}^2 \frac{t1}{T} + V2_{\text{RMS}}^2 \frac{t2}{T} + \ldots + Vn_{\text{RMS}}^2 \frac{tn}{T}\right)} \tag{4.15}$$

We can see that Eqn (4.13) is just a special case of this.

By way of illustration:

Example 4.1: The triangle wave of Figure 4.4 has an AC component that goes between +4 V and −4 V. Its RMS value is therefore 0.577 × 4 = 2.31 V. It has a DC component of 5 V, so the true RMS value is:

$$RMS_{\text{triangle}} = \sqrt{(5^2 + 2.31^2)} = 5.5\,\text{V}$$

Example 4.2: A signal consists of a high frequency sine wave of 5 V pk for 400 ms with no DC, followed by a DC voltage of 2 V for 300 ms. What is the RMS value?

RMS value of the sine wave is 5 × 0.707 = 3.54 V

$$RMS_{\text{total}} = \sqrt{\left(\frac{4}{7} \times 3.54^2 + \frac{3}{7} \times 2^2\right)} = 2.98\,\text{V}$$

5 A circuit analysis toolkit

Introduction

This chapter is intended as a quick guide to the basic techniques of network analysis. The mathematical requirements are minimal; you need to be able to transpose equations and solve simultaneous equations by substitution. We also get to consider some characteristics of real voltage and current sources, which ties in nicely with some of the analysis tools presented.

We will take DC circuits for all of our examples, as this is the easiest way of demonstrating the techniques. Everything said applies equally to AC circuits as well, but you generally need to use complex numbers (see Chapter 4 and Appendix 4) to take account of phase if there are capacitors and inductors present, or sources at different phases.

Basic topologies

There are two basic shapes that electric and electronic circuits come in; series and parallel. Each has associated with it one of two electrical laws which tell us how to go about calculating the voltages and currents in that circuit.

The series circuit and Kirchhoff's voltage law

A series circuit is represented in Figure 5.1. The components are an input voltage, V, 10 V DC, and three resistors: R_1, 1k; R_2, 2k2; and R_3, 4k7. We have also labelled on the drawing, for our convenience, the voltages across each resistor, V_1, V_2 and V_3. The voltages are labelled with arrows. With DC voltages we follow a convention of putting the head of the arrow to the more positive end of the component.

The key to analysing a series circuit is to remember that the current is identical through each component in the loop. (We can also say that connecting the components in a different order will not affect that current.) Hence the current in the loop is labelled just once, this time with an arrow on the loop, with the legend I next to it to identify it. The head of the arrow shows the direction of current flow.

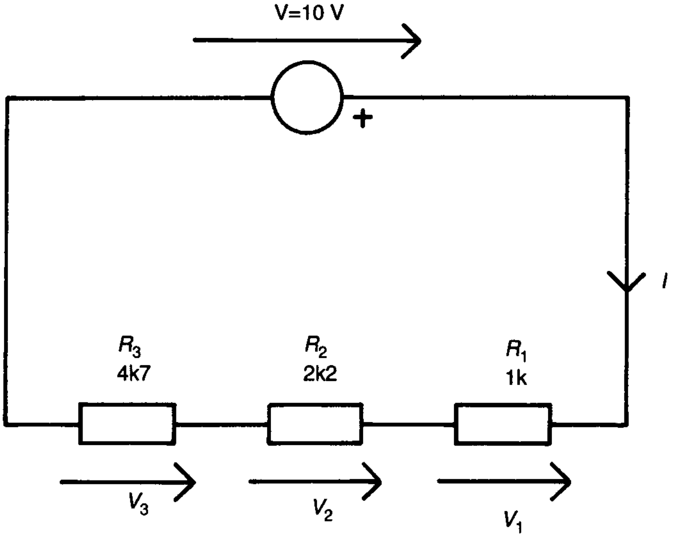

Figure 5.1 Series circuit

Example 5.1: Suppose that we wish to calculate V_1, V_2 and V_3. How do we do it? We need to find the value of current, I. Then we will be able to calculate each voltage using Ohm's Law.

We can find the total resistance as described in Chapter 2. In our case:

$$R_S = 1\text{k} + 2\text{k}2 + 4\text{k}7 = 7900\,\Omega$$

Now we can say that:

$$I = \frac{10}{7900} = 1.27\,\text{mA}$$

Now we can use Ohm's Law to find V_1, V_2 and V_3:

$$V_1 = 1.27\,\text{mA} \times 1\text{k} = 1.27\,\text{V}$$
$$V_2 = 1.27\,\text{mA} \times 2\text{k}2 = 2.78\,\text{V}$$
$$V_3 = 1.27\,\text{mA} \times 4\text{k}7 = 5.95\,\text{V}$$

Kirchhoff's voltage law states that, in any loop of voltages on a circuit, their sum will be zero if we take the clockwise ones as being positive, the anticlockwise ones as negative (or vice versa). Put another way, the clockwise ones equal the anticlockwise ones. This enables us to check our calculations: 1.27 + 2.78 + 5.95 = 10V – we did it right. In any mathematical analysis of a circuit, its always good to have a way to check our result, and using voltage loops is one way.

The parallel circuit and Kirchhoff's current law

Example 5.2: Figure 5.2 shows a basic parallel circuit made with the same components. With parallel circuits the rule is that the voltage across all the branches is identical. Hence there is only one voltage to be labelled, the source voltage. But there are four currents: I, the current leaving the voltage source, and I_1, I_2 and I_3, the currents through the resistors. We would like to calculate them.

We can see that the voltage across each resistor is the same; the 10 V source voltage. Therefore we can apply Ohm's Law directly to each branch of the circuit to find I_1, I_2 and I_3:

$$I_1 = \frac{10}{1k} = 10 \text{ mA}$$
$$I_2 = \frac{10}{2\text{k}2} = 4.55 \text{ mA}$$
$$I_3 = \frac{10}{4\text{k}7} = 2.13 \text{ mA}$$

Kirchhoff's current law states that the total of currents entering and leaving any 'node' (a node is any place where two or more conductors join) is zero. This enables us to calculate the current I:

$$I = I_1 + I_2 + I_3 = (10 + 4.55 + 2.13)\text{mA} = 16.68\,\text{mA}$$

We know that the resistance between any two points is the ratio of the voltage across them and the current passing between them. So in this case the resistance is:

$$R_\text{P} = \frac{10}{16.68\,\text{mA}} = 600\,\Omega$$

We could have calculated the parallel resistance directly, by using the formula from Chapter 2:

$$\frac{1}{R_\text{P}} = \frac{1}{1\text{k}} + \frac{1}{2\text{k}2} + \frac{1}{4\text{k}7} = 1.67\,\text{m}\,\Omega$$

So:

$$R_\text{P} = \frac{1}{1.67\,\text{mA}} = 600\,\Omega$$

Again we have cross-checked our result, and it's fine.

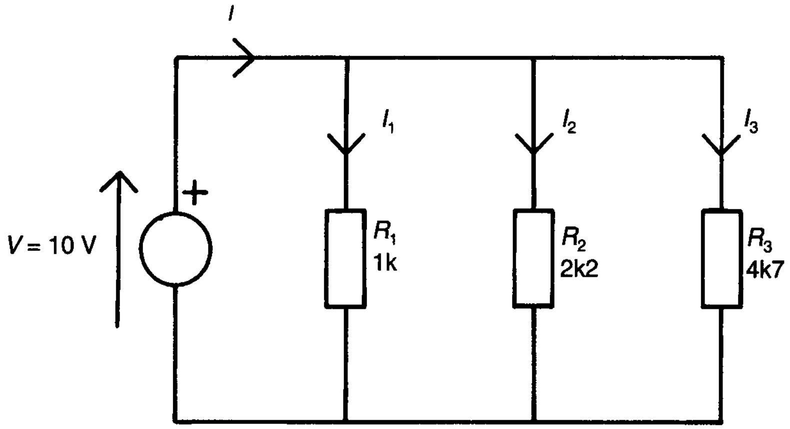

Figure 5.2 Parallel circuit

Arrow directions for voltages and currents

It is worth taking a moment to think about the rules that we apply when adding voltages and currents in using Kirchhoff's two theorems, as it can be surprisingly easy to confuse ourselves.

Here's the rules. If we follow these, things work:

1. The arrows for voltage across and current through a voltage or current source must point in the same direction.
2. The arrows for voltage across and current through a resistance must point in opposite directions.
3. For a voltage loop, the sum of arrows going clockwise equals the sum of arrows going anti-clockwise.
4. For a node, the sum of currents entering a node equals the sum of currents leaving the node, as denoted by the arrows' directions.

In the above examples, it was obvious which end of the resistors would be more positive and in which direction currents would be flowing. Hence drawing the arrows was easy.

In other instances it will not be obvious, when labelling currents and voltages at the start of the analysis, which end of the component is the more positive. Such a situation is shown in Figure 5.3. We wish to know the voltage V_2, but this voltage is affected by both voltage and current sources which have a tendency to create voltages of opposite polarity across R_2. The way to approach this sort of problem is not to worry about the polarity of the voltage; if we label it in the wrong direction, the voltage will come out as negative in solving the equations. We simply get on with labelling the voltages and currents (following rules 1 and 2), then we can write equations for voltage loops (using rule 3, the voltage law) or nodes (with rule 4, the current law).

Example 5.3: Solve Figure 5.3.

For node A, we can write:

$$I_1 = I_2 + 10\,\mathrm{mA}$$
$$I_2 = I_1 - 10\,\mathrm{mA} \quad \text{(call this Eqn (A))}$$

For the loop on the left we can write:

$$V_{IN} = V_1 + V_2$$
$$V_2 = V_{IN} - V_1$$

(but $V_1 = I_1.R_1 = 10\,\mathrm{k}.I_1$ and $V_2 = I_2.R_2 = 2\,\mathrm{k}2.I_2$)
so:

$$2\,\mathrm{k}2.I_2 = 5 - 10\,\mathrm{k}.I_1$$
$$I_2 = \left(\frac{5 - 10\,\mathrm{k}.I_1}{2\mathrm{k}2}\right)$$
$$I_2 = 2.27\,\mathrm{mA} - 4.55.I_1$$

We can substitute this back into Eqn (A), giving:

$$2.27\,\mathrm{mA} - 4.55.I_1 = I_1 - 10\,\mathrm{mA}$$
$$I_1(1 + 4.55) = 2.27\,\mathrm{mA} + 10\,\mathrm{mA}$$
$$I_1 = \frac{12.27\,\mathrm{mA}}{5.55} = 2.21\,\mathrm{mA}$$

So, again from Equation A:

$$I_2 = 2.21\,\mathrm{mA} - 10\,\mathrm{mA} = -7.79\,\mathrm{mA}$$

and therefore

$$V_2 = I_2.R_2 = -17.1\,\mathrm{V}$$

and

$$V_1 = 2.21\,\mathrm{mA} \times 10\mathrm{k} = 22.1\,\mathrm{V}$$

Our solution is: $V_1 = 22.1$ V, $I_1 = 2.21$ mA, $V_2 = -17.1$ V, $I_2 = -7.79$ mA. We can now redraw the circuit, reversing the direction of the arrows where our solution gave a minus sign (Figure 5.4). A few seconds of calculator-punching demonstrates that nothing went wrong with the number crunching, and the solution obeys the voltage and current laws. There are quicker ways to get to this result, but as a 'swiss army knife' technique, this is a good one to have. Soon we get some better ones still.

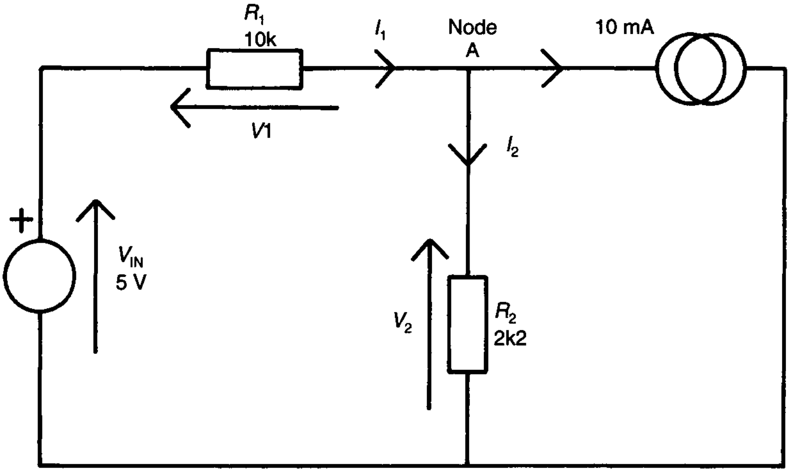
R_1
10k
I_1
Node
A
10 mA
V1
I_2
+
V_{IN}
5 V
V_2
R_2
2k2

Figure 5.3 Example 5.3

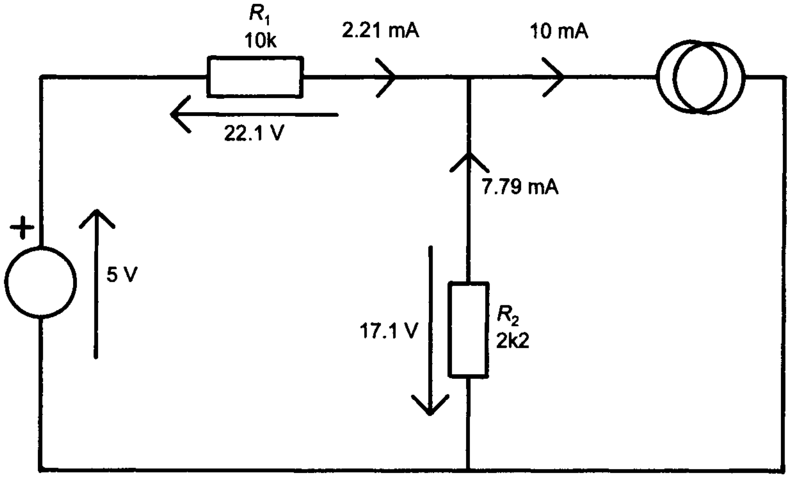
R_1
10k
2.21 mA
10 mA
22.1 V
7.79 mA
+
5 V
17.1 V
R_2
2k2

Figure 5.4 Solution to Example 5.3

More about voltage and current sources

The voltage and current sources that we considered above were 'ideal' – that is, we consider that they will supply their correct voltage or current, regardless of what load is placed on them. (As such, they have no real existence.) For the ideal voltage source, we could say that the change in voltage across it is zero for any change in output current. This means that a notional quantity which we call its 'internal resistance' is also zero, by Ohm's Law. For the current source we could say that the change in output current is zero for any change in output voltage. Hence its internal resistance is infinite. Real voltage and current sources are not ideal, and we have a way to represent this in our circuits (see Figure 5.5).

We give our ideal voltage source a series resistor, R_S. The closer this resistor is to a short circuit the closer the source is to ideal. If we connect a real voltage source to too low a load resistance the voltage across the load will drop (and the source could be damaged).

The current source is given a parallel resistor, R_P; the closer this is to an open circuit, the better the source. If we connect a real current source to too high a load resistance the current through the load will drop.

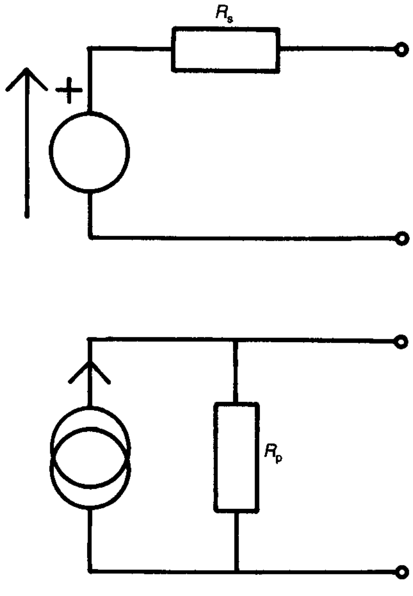

Figure 5.5 Representation of non-ideal sources

These circuit representations of real voltage and current sources are invaluable for circuit analysis, as we will now see. It is possible for any network of voltage sources, current sources and resistances to be replaced by either a single ideal voltage source with a series resistor, or a single ideal current source with a parallel resistor.

Thévenin's and Norton's equivalent circuits

The techniques that we used in the last example of circuit analysis are fine for instances where we can get a solution for the circuit by writing two or three simultaneous equations and solving them. If the network gets much more complicated than that, however, the amount of maths involved can get pretty tedious, and it can get time consuming to arrive at the correct result without making a numerical error. Also, we have only solved the network for one value of load resistance; if we want to see what happens with a different value, we must do the calculations again. This is where replacing whole lumps of circuit with an equivalent non-ideal current or voltage source is great.

Here's how we go about finding the right values for our equivalent circuit:

Replace with a voltage source (Thévenin's equivalent circuit):

1. Cut the network at two places to isolate the section to be replaced.
2. Calculate the voltage at the load terminals with no load connected (called the 'open circuit voltage' or V_{TH}).
3. Calculate the equivalent resistance looking into the load terminals when any voltage sources are replaced with a short circuit and any current sources with an open circuit (called the 'output resistance' or R_{TH}).

Example 5.4: What is the voltage across R_L in Figure 5.6?

First we separate out the 12 V and the 3 mA source, and combine then into a single voltage source (– Figure 5.7). The points where we cut are labelled A and B. Replacing the 12 V source with a short and the 3 mA one with an open, we get $R_{TH} = 1\text{k} + 1\text{k} = 2\text{k}$ (Figure 5.8). When we calculate the open circuit voltage, we get 3 mA × 1k = 3 V across R_2, and nothing across R_3, which has no current flowing in it. Hence V_{TH} is 12 – 3 = 9 V.

The equivalent circuit is replaced into the original in Figure 5.9. Things are getting easier! Now we can cut and replace at the load terminals themselves, L_1 and L_2. This time we'll replace with a current source.

Looking into L_1 and L_2 with both sources replaced by a short, we get 2k in parallel with 2k, so R_N is 1k. (Figure 5.10).

To calculate short circuit current, we replace R_L with a short, and then calculate the current through the two resistors; 2.5 mA and 4.5 mA as shown on the drawing, taking care to label directions correctly. I_N is then clearly the difference between the two (by Kirchhoff's current law), 2 mA. So Figure 5.11 shows the equivalent circuit with R_l now connected. If R_l is 1 kΩ, parallel resistance with R_N is 500 Ω, and voltage across R_l is 1 V, terminal L_1 being most positive.

Naturally we want to check, and the easiest way to do this is to redraw the original circuit, labelling in voltages and currents, using Ohm's Law and voltage and current laws, and working backwards from R_1. If there is a problem it will become apparent. We find we are OK (see Figure 5.12).

We now could find the current through any other load using our existing solution – no need to do all that number-crunching again.

The network can now be replaced with an ideal voltage source equivalent to V_{TH}, in series with a resistor equivalent to R_{TH}.

Replace with a current source (Norton's equivalent circuit):

1. Cut the network at two places to isolate the section to be replaced.
2. Calculate the current at the load terminals when they are short circuited (called the 'short circuit current' or I_N).
3. Calculate the equivalent resistance looking into the load terminals when any voltage sources are replaced with a short circuit and any current sources with an open circuit (called the 'output resistance' or R_N).

The network can now be replaced with an ideal current equivalent to I_N, in parallel with a resistor equivalent to R_N.

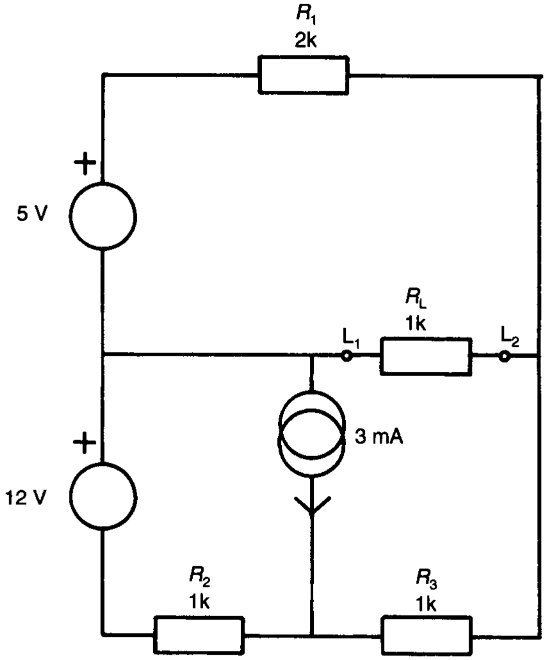

Figure 5.6 Example 5.4

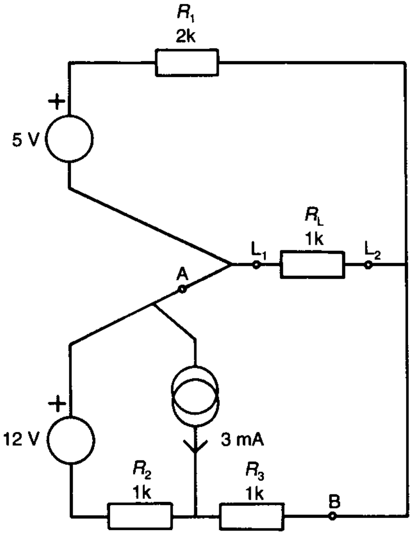
R_1
2k
5 V
+
R_L
1k
L_1
L_2
A
+
12 V
3 mA
R_2
1k
R_3
1k
B

Figure 5.7 Figure 5.6 redrawn before simplifying

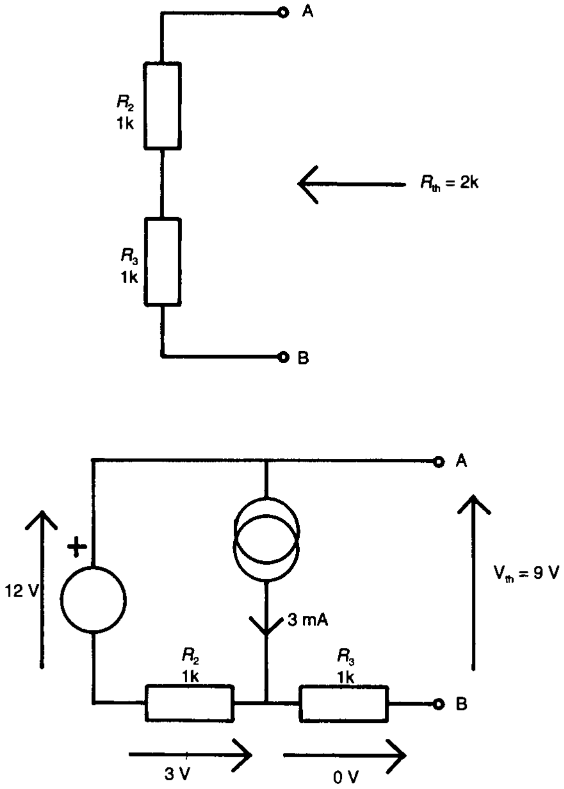
A
R_2
1k
R_3
1k
B
R_{th} = 2k
A
+
12 V
3 mA
V_{th} = 9 V
R_2
1k
R_3
1k
B
3 V
0 V

Figure 5.8 Finding R_{th} and V_{th}

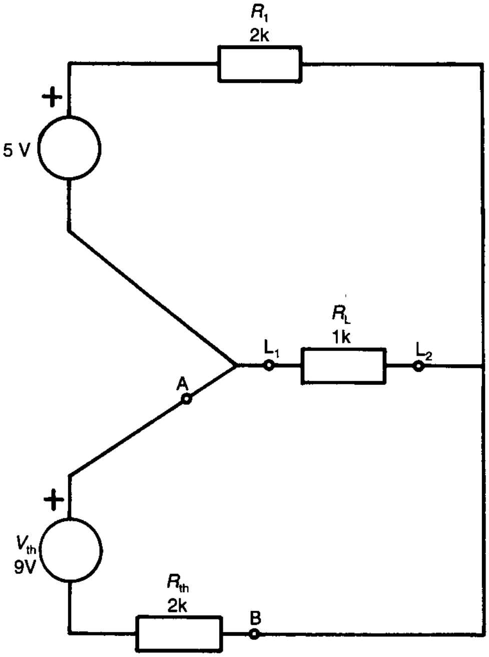

R_1
2k
5 V
R_L
1k
L_1
L_2
A
V_{th}
9V
R_{th}
2k
B

Figure 5.9 V_{th} and R_{th} replaced

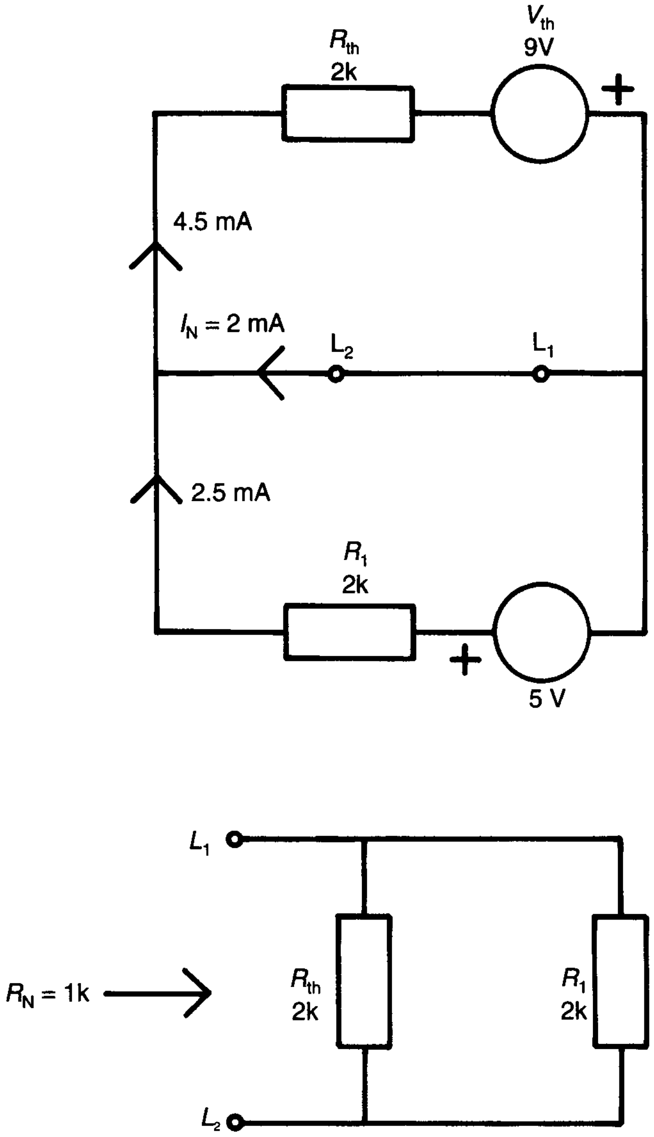
V_{th}
9V
R_{th}
2k
+
4.5 mA
I_N = 2 mA
L_2
L_1
2.5 mA
R_1
2k
+
5 V
L_1
R_N = 1k
R_{th}
2k
R_1
2k
L_2

Figure 5.10 Finding I_N and R_N

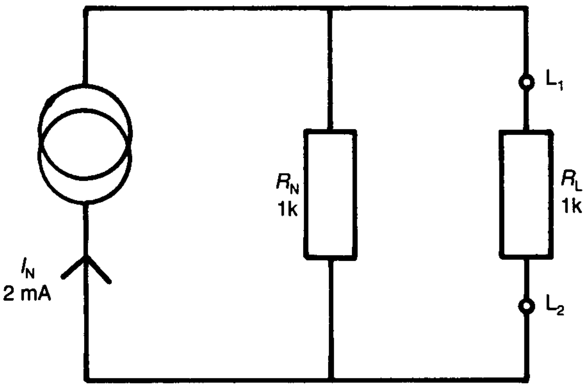
L1
RN
1k
RL
1k
IN
2 mA
L2

Figure 5.11 Equivalent circuit to Figure 5.6

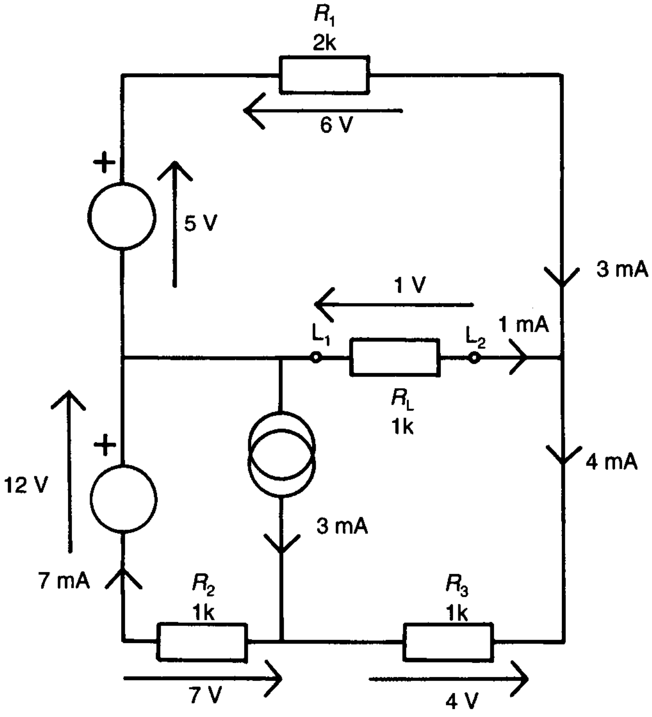
R1
2k
6 V
+
5 V
3 mA
1 V
L1
L2
1 mA
RL
1k
+
12 V
4 mA
3 mA
7 mA
R2
1k
R3
1k
7 V
4 V

Figure 5.12 Solution of Figure 5.6

Part Two

Resistive Circuits

This Page Intentionally Left Blank

6 Equivalent resistances

Introduction

Resistors can be bought in certain standard values. The ranges of values are commonly called the E6, the E12 and the E24 series, the numbers denoting the quantity of values per decade of resistance. For building or repairing low voltage electronic circuits most workshops will stock a full range of E12 or E24 miniature resistors (i.e. 1/8 W or 1/4 W power rating), usually 1 or 2% tolerance, plus a selection of resistors rated at higher powers, which are available in combinations of E6 and E12 values. An additional range, E96, is also available, but these are more expensive, 0.1% tolerance, and few workshops will stock them as standard due to the large number of values. Also they are hard to obtain below 100 Ω or over about 250k.

The standard values are almost always fine, but once in a while you need a value that lies between two of them. This can be achieved by combining two components in series or parallel, their values being chosen so that the overall equivalent is near enough.

It is evident that choosing the best pair of resistors for this purpose might involve a fair amount of trial and error with a calculator, and that several near solutions might exist. For this purpose, three 'Tables of equivalent resistance' are provided in the rest of this chapter. Table 6.1 is for E6, Table 6.2 for E12, and Table 6.3 for E24. You should use the appropriate one for the range of components which you intend to use.

Using the Tables for resistors

Each table shows all achievable values between 1k and 10k with two resistors. s denotes series connection and p denotes parallel. The tables are sorted in order of ascending equivalent resistance.

Example 6.1: Choose a pair of E24 resistors to give an equivalent resistance of 7k7.

From Table 6.1:

R_E	R_1	s/p	R_2	P_{REL}
7700	200R	s	7k5	1.03
7700	1k5	s	6k2	1.24
7700	3k	s	4k7	1.64

There are plenty of others that would be close enough too.
To get a value of less than 1k, or greater than 10k, you should scale all values by factors of ten as necessary.

Example 6.2: Choose a pair of E24 resistors to form a resistance of 125k.

R_E	R_1	s/p	R_2	P_{REL}
1250	1k5	p	7k5	1.2

125k is 1250 multiplied by 100, so resistors of 150k (1k5 times 100) and 750k (7k5 times 100), connected in parallel, form a resistance of 125k.

Maximum power rating of the pair

Each line of the table gives a value of P_{REL}. Where each resistor of the pair is of the same power rating, this can be used to find the overall power rating of the pair. It is always a number between 1 and 2. For example, if a pair of 1/4 W resistors are used and P_{REL} has a value of 1.2 then the power rating of the pair is equal to:

$$0.25 \times 1.2 = 0.3\,\text{W}$$

In this way, the pair of resistors can be lumped and considered as a single circuit element.

Tolerance of components

If two components of the same tolerance are used the tolerance of the equivalent resistance is the same.

Using the Tables for capacitors and inductors

Both inductors and capacitors are normally bought in combinations of the standard E6 and E12 ranges. Like resistors, they can be combined in series or parallel to obtain non-standard values. As was discussed in Chapter 2, inductors in series and parallel obey the same rules as resistors, while capacitors in series behave like resistors in parallel, and vice-versa.

Hence, the tables can be used to find equivalent capacitance or inductance. For capacitors, you should read p for s and s for p.

Example 6.3: To form a capacitor of 3n1 using a range of 1 %, E12 capacitors:

R_E	R_1	s/p	R_2	P_{REL}
3090	390R	s	2k7	1.14
3095	3k9	p	15k	1.26
3116	3k3	p	56k	1.06

This gives three alternatives: 2n7 p 390p, 3n9 s 15n and 3n3 s 56n. All three form a capacitor of 3n1, 1 %. However, as capacitors (and inductors) tend to increase in physical size with increasing value the first might be preferable.

Table 6.1 *E6 equivalent resistances*

R_E	R_1	s/p	R_2	P_{REL}
1010	10R	s	1k	1.01
1010	330R	s	680R	1.49
1015	15R	s	1k	1.02
1022	22R	s	1k	1.02
1031	1k5	p	3k3	1.45
1033	33R	s	1k	1.03
1047	47R	s	1k	1.05
1068	68R	s	1k	1.07
1100	100R	s	1k	1.1
1100	2k2	p	2k2	2
1137	1k5	p	4k7	1.32
1150	150R	s	1k	1.15
1150	470R	s	680R	1.69
1220	220R	s	1k	1.22
1229	1k5	p	6k8	1.22
1304	1k5	p	10k	1.15
1320	2k2	p	3k3	1.67
1330	330R	s	1k	1.33
1360	680R	s	680R	2
1364	1k5	p	15k	1.1
1404	1k5	p	22k	1.07
1435	1k5	p	33k	1.05
1454	1k5	p	47k	1.03
1468	1k5	p	68k	1.02
1470	470R	s	1k	1.47
1478	1k5	p	100k	1.02
1485	1k5	p	150k	1.01
1499	2k2	p	4k7	1.47
1515	15R	s	1k5	1.01
1522	22R	s	1k5	1.01
1533	33R	s	1k5	1.02
1547	47R	s	1k5	1.03
1568	68R	s	1k5	1.05
1600	100R	s	1k5	1.07
1650	150R	s	1k5	1.1
1650	3k3	p	3k3	2
1662	2k2	p	6k8	1.32
1680	680R	s	1k	1.68
1720	220R	s	1k5	1.15
1803	2k2	p	10k	1.22
1830	330R	s	1k5	1.22
1919	2k2	p	15k	1.15
1939	3k3	p	4k7	1.7
1970	470R	s	1k5	1.31
2000	1k	s	1k	2
2000	2k2	p	22k	1.1
2063	2k2	p	33k	1.07
2102	2k2	p	47k	1.05
2131	2k2	p	68k	1.03
2153	2k2	p	100k	1.02
2168	2k2	p	150k	1.01
2178	2k2	p	220k	1.01
2180	680R	s	1k5	1.45
2222	22R	s	2k2	1.01
2222	3k3	p	6k8	1.49
2233	33R	s	2k2	1.02
2247	47R	s	2k2	1.02
2268	68R	s	2k2	1.03
2300	100R	s	2k2	1.05
2350	150R	s	2k2	1.07
2350	4k7	p	4k7	2
2420	220R	s	2k2	1.1
2481	3k3	p	10k	1.33
2500	1k	s	1k5	1.67
2530	330R	s	2k2	1.15
2670	470R	s	2k2	1.21
2705	3k3	p	15k	1.22
2779	4k7	p	6k8	1.69
2870	3k3	p	22k	1.15
2880	680R	s	2k2	1.31
3000	1k5	s	1k5	2
3000	3k3	p	33k	1.1

R_E	R_1	s/p	R_2	P_{REL}
3083	3k3	p	47k	1.07
3147	3k3	p	68k	1.05
3195	3k3	p	100k	1.03
3197	4k7	p	10k	1.47
3200	1k	s	2k2	1.45
3229	3k3	p	150k	1.02
3251	3k3	p	220k	1.02
3267	3k3	p	330k	1.01
3333	33R	s	3k3	1.01
3347	47R	s	3k3	1.01
3368	68R	s	3k3	1.02
3400	100R	s	3k3	1.03
3400	6k8	p	6k8	2
3450	150R	s	3k3	1.05
3520	220R	s	3k3	1.07
3579	4k7	p	15k	1.31
3630	330R	s	3k3	1.1
3700	1k5	s	2k2	1.68
3770	470R	s	3k3	1.14
3873	4k7	p	22k	1.21
3980	680R	s	3k3	1.21
4048	6k8	p	10k	1.68
4114	4k7	p	33k	1.14
4273	4k7	p	47k	1.1
4300	1k	s	3k3	1.3
4396	4k7	p	68k	1.07
4400	2k2	s	2k2	2
4489	4k7	p	100k	1.05
4557	4k7	p	150k	1.03
4602	4k7	p	220k	1.02
4634	4k7	p	330k	1.01
4653	4k7	p	470k	1.01
4679	6k8	p	15k	1.45
4747	47R	s	4k7	1.01
4768	68R	s	4k7	1.01
4800	100R	s	4k7	1.02
4800	1k5	s	3k3	1.45

R_E	R_1	s/p	R_2	P_{REL}
4850	150R	s	4k7	1.03
4920	220R	s	4k7	1.05
5000	10k	p	10k	2
5030	330R	s	4k7	1.07
5170	470R	s	4k7	1.1
5194	6k8	p	22k	1.31
5380	680R	s	4k7	1.14
5500	2k2	s	3k3	1.67
5638	6k8	p	33k	1.21
5700	1k	s	4k7	1.21
5941	6k8	p	47k	1.14
6000	10k	p	15k	1.67
6182	6k8	p	68k	1.1
6200	1k5	s	4k7	1.32
6367	6k8	p	100k	1.07
6505	6k8	p	150k	1.05
6596	6k8	p	220k	1.03
6600	3k3	s	3k3	2
6663	6k8	p	330k	1.02
6703	6k8	p	470k	1.01
6733	6k8	p	680k	1.01
6868	68R	s	6k8	1.01
6875	10k	p	22k	1.45
6900	100R	s	6k8	1.01
6900	2k2	s	4k7	1.47
6950	150R	s	6k8	1.02
7020	220R	s	6k8	1.03
7130	330R	s	6k8	1.05
7270	470R	s	6k8	1.07
7480	680R	s	6k8	1.1
7500	15k	p	15k	2
7674	10k	p	33k	1.3
7800	1k	s	6k8	1.15
8000	3k3	s	4k7	1.7
8246	10k	p	47k	1.21
8300	1k5	s	6k8	1.22
8718	10k	p	68k	1.15

R_E	R_1	s/p	R_2	P_{REL}
8919	15k	p	22k	1.68
9000	2k2	s	6k8	1.32
9091	10k	p	100k	1.1
9375	10k	p	150k	1.07
9400	4k7	s	4k7	2

R_E	R_1	s/p	R_2	P_{REL}
9565	10k	p	220k	1.05
9706	10k	p	330k	1.03
9792	10k	p	470k	1.02
9855	10k	p	680k	1.01
9901	10k	p	1M	1.01

Table 6.2 *E12 equivalent resistances*

R_E	R_1	s/p	R_2	P_{REL}
1010	10R	s	1k	1.01
1010	330R	s	680R	1.49
1012	12R	s	1k	1.01
1015	15R	s	1k	1.02
1018	18R	s	1k	1.02
1020	1k2	p	6k8	1.18
1022	22R	s	1k	1.02
1027	27R	s	1k	1.03
1030	470R	s	560R	1.84
1031	1k5	p	3k3	1.45
1033	33R	s	1k	1.03
1039	39R	s	1k	1.04
1040	220R	s	820R	1.27
1047	47R	s	1k	1.05
1047	1k2	p	8k2	1.15
1056	56R	s	1k	1.06
1068	68R	s	1k	1.07
1070	390R	s	680R	1.57
1071	1k2	p	10k	1.12
1080	1k8	p	2k7	1.67
1082	82R	s	1k	1.08
1083	1k5	p	3k9	1.38
1090	270R	s	820R	1.33
1091	1k2	p	12k	1.1
1100	100R	s	1k	1.1
1100	2k2	p	2k2	2
1111	1k2	p	15k	1.08
1120	120R	s	1k	1.12
1120	560R	s	560R	2
1125	1k2	p	18k	1.07
1137	1k5	p	4k7	1.32
1138	1k2	p	22k	1.05
1149	1k2	p	27k	1.04
1150	150R	s	1k	1.15
1150	330R	s	820R	1.4
1150	470R	s	680R	1.69
1158	1k2	p	33k	1.04
1164	1k2	p	39k	1.03
1165	1k8	p	3k3	1.55
1170	1k2	p	47k	1.03
1175	1k2	p	56k	1.02
1179	1k2	p	68k	1.02
1180	180R	s	1k	1.18
1183	1k2	p	82k	1.01
1183	1k5	p	5k6	1.27
1186	1k2	p	100k	1.01
1188	1k2	p	120k	1.01
1210	390R	s	820R	1.48
1212	12R	s	1k2	1.01
1212	2k2	p	2k7	1.81
1215	15R	s	1k2	1.01
1218	18R	s	1k2	1.02
1220	220R	s	1k	1.22
1222	22R	s	1k2	1.02
1227	27R	s	1k2	1.02
1229	1k5	p	6k8	1.22
1232	1k8	p	3k9	1.46
1233	33R	s	1k2	1.03
1239	39R	s	1k2	1.03
1240	560R	s	680R	1.82
1247	47R	s	1k2	1.04
1256	56R	s	1k2	1.05
1268	68R	s	1k2	1.06
1268	1k5	p	8k2	1.18
1270	270R	s	1k	1.27
1282	82R	s	1k2	1.07
1290	470R	s	820R	1.57
1300	100R	s	1k2	1.08
1302	1k8	p	4k7	1.38
1304	1k5	p	10k	1.15
1320	120R	s	1k2	1.1
1320	2k2	p	3k3	1.67

R_E	R_1	s/p	R_2	P_{REL}
1330	330R	s	1k	1.33
1333	1k5	p	12k	1.13
1350	150R	s	1k2	1.13
1350	2k7	p	2k7	2
1360	680R	s	680R	2
1362	1k8	p	5k6	1.32
1364	1k5	p	15k	1.1
1380	180R	s	1k2	1.15
1380	560R	s	820R	1.68
1385	1k5	p	18k	1.08
1390	390R	s	1k	1.39
1404	1k5	p	22k	1.07
1407	2k2	p	3k9	1.56
1420	220R	s	1k2	1.18
1421	1k5	p	27k	1.06
1423	1k8	p	6k8	1.26
1435	1k5	p	33k	1.05
1444	1k5	p	39k	1.04
1454	1k5	p	47k	1.03
1461	1k5	p	56k	1.03
1468	1k5	p	68k	1.02
1470	270R	s	1k2	1.23
1470	470R	s	1k	1.47
1473	1k5	p	82k	1.02
1476	1k8	p	8k2	1.22
1478	1k5	p	100k	1.02
1481	1k5	p	120k	1.01
1485	1k5	p	150k	1.01
1485	2k7	p	3k3	1.82
1499	2k2	p	4k7	1.47
1500	680R	s	820R	1.83
1515	15R	s	1k5	1.01
1518	18R	s	1k5	1.01
1522	22R	s	1k5	1.01
1525	1k8	p	10k	1.18
1527	27R	s	1k5	1.02
1530	330R	s	1k2	1.28

R_E	R_1	s/p	R_2	P_{REL}
1533	33R	s	1k5	1.02
1539	39R	s	1k5	1.03
1547	47R	s	1k5	1.03
1556	56R	s	1k5	1.04
1560	560R	s	1k	1.56
1565	1k8	p	12k	1.15
1568	68R	s	1k5	1.05
1579	2k2	p	5k6	1.39
1582	82R	s	1k5	1.05
1590	390R	s	1k2	1.33
1595	2k7	p	3k9	1.69
1600	100R	s	1k5	1.07
1607	1k8	p	15k	1.12
1620	120R	s	1k5	1.08
1636	1k8	p	18k	1.1
1640	820R	s	820R	2
1650	150R	s	1k5	1.1
1650	3k3	p	3k3	2
1662	2k2	p	6k8	1.32
1664	1k8	p	22k	1.08
1670	470R	s	1k2	1.39
1680	180R	s	1k5	1.12
1680	680R	s	1k	1.68
1688	1k8	p	27k	1.07
1707	1k8	p	33k	1.05
1715	2k7	p	4k7	1.57
1720	220R	s	1k5	1.15
1721	1k8	p	39k	1.05
1734	1k8	p	47k	1.04
1735	2k2	p	8k2	1.27
1744	1k8	p	56k	1.03
1754	1k8	p	68k	1.03
1760	560R	s	1k2	1.47
1761	1k8	p	82k	1.02
1768	1k8	p	100k	1.02
1770	270R	s	1k5	1.18
1773	1k8	p	120k	1.02

R_E	R_1	s/p	R_2	P_{REL}
1779	1k8	p	150k	1.01
1782	1k8	p	180k	1.01
1788	3k3	p	3k9	1.85
1803	2k2	p	10k	1.22
1818	18R	s	1k8	1.01
1820	820R	s	1k	1.82
1822	22R	s	1k8	1.01
1822	2k7	p	5k6	1.48
1827	27R	s	1k8	1.02
1830	330R	s	1k5	1.22
1833	33R	s	1k8	1.02
1839	39R	s	1k8	1.02
1847	47R	s	1k8	1.03
1856	56R	s	1k8	1.03
1859	2k2	p	12k	1.18
1868	68R	s	1k8	1.04
1880	680R	s	1k2	1.57
1882	82R	s	1k8	1.05
1890	390R	s	1k5	1.26
1900	100R	s	1k8	1.06
1919	2k2	p	15k	1.15
1920	120R	s	1k8	1.07
1933	2k7	p	6k8	1.4
1939	3k3	p	4k7	1.7
1950	150R	s	1k8	1.08
1950	3k9	p	3k9	2
1960	2k2	p	18k	1.12
1970	470R	s	1k5	1.31
1980	180R	s	1k8	1.1
2000	1k	s	1k	2
2000	2k2	p	22k	1.1
2020	220R	s	1k8	1.12
2020	820R	s	1k2	1.68
2031	2k7	p	8k2	1.33
2034	2k2	p	27k	1.08
2060	560R	s	1k5	1.37
2063	2k2	p	33k	1.07
2070	270R	s	1k8	1.15
2076	3k3	p	5k6	1.59
2083	2k2	p	39k	1.06
2102	2k2	p	47k	1.05
2117	2k2	p	56k	1.04
2126	2k7	p	10k	1.27
2130	330R	s	1k8	1.18
2131	2k2	p	68k	1.03
2131	3k9	p	4k7	1.83
2143	2k2	p	82k	1.03
2153	2k2	p	100k	1.02
2160	2k2	p	120k	1.02
2168	2k2	p	150k	1.01
2173	2k2	p	180k	1.01
2178	2k2	p	220k	1.01
2180	680R	s	1k5	1.45
2190	390R	s	1k8	1.22
2200	1k	s	1k2	1.83
2204	2k7	p	12k	1.23
2222	22R	s	2k2	1.01
2222	3k3	p	6k8	1.49
2227	27R	s	2k2	1.01
2233	33R	s	2k2	1.02
2239	39R	s	2k2	1.02
2247	47R	s	2k2	1.02
2256	56R	s	2k2	1.03
2268	68R	s	2k2	1.03
2270	470R	s	1k8	1.26
2282	82R	s	2k2	1.04
2288	2k7	p	15k	1.18
2299	3k9	p	5k6	1.7
2300	100R	s	2k2	1.05
2320	120R	s	2k2	1.05
2320	820R	s	1k5	1.55
2348	2k7	p	18k	1.15
2350	150R	s	2k2	1.07
2350	4k7	p	4k7	2

R_E	R_1	s/p	R_2	P_{REL}
2353	3k3	p	8k2	1.4
2360	560R	s	1k8	1.31
2380	180R	s	2k2	1.08
2400	1k2	s	1k2	2
2405	2k7	p	22k	1.12
2420	220R	s	2k2	1.1
2455	2k7	p	27k	1.1
2470	270R	s	2k2	1.12
2479	3k9	p	6k8	1.57
2480	680R	s	1k8	1.38
2481	3k3	p	10k	1.33
2496	2k7	p	33k	1.08
2500	1k	s	1k5	1.67
2525	2k7	p	39k	1.07
2530	330R	s	2k2	1.15
2553	2k7	p	47k	1.06
2555	4k7	p	5k6	1.84
2576	2k7	p	56k	1.05
2588	3k3	p	12k	1.28
2590	390R	s	2k2	1.18
2597	2k7	p	68k	1.04
2614	2k7	p	82k	1.03
2620	820R	s	1k8	1.46
2629	2k7	p	100k	1.03
2641	2k7	p	120k	1.02
2643	3k9	p	8k2	1.48
2652	2k7	p	150k	1.02
2660	2k7	p	180k	1.02
2667	2k7	p	220k	1.01
2670	470R	s	2k2	1.21
2673	2k7	p	270k	1.01
2700	1k2	s	1k5	1.8
2705	3k3	p	15k	1.22
2727	27R	s	2k7	1.01
2733	33R	s	2k7	1.01
2739	39R	s	2k7	1.01
2747	47R	s	2k7	1.02
2756	56R	s	2k7	1.02
2760	560R	s	2k2	1.25
2768	68R	s	2k7	1.03
2779	4k7	p	6k8	1.69
2782	82R	s	2k7	1.03
2789	3k3	p	18k	1.18
2800	100R	s	2k7	1.04
2800	1k	s	1k8	1.56
2800	5k6	p	5k6	2
2806	3k9	p	10k	1.39
2820	120R	s	2k7	1.04
2850	150R	s	2k7	1.06
2870	3k3	p	22k	1.15
2880	180R	s	2k7	1.07
2880	680R	s	2k2	1.31
2920	220R	s	2k7	1.08
2941	3k3	p	27k	1.12
2943	3k9	p	12k	1.33
2970	270R	s	2k7	1.1
2988	4k7	p	8k2	1.57
3000	1k2	s	1k8	1.67
3000	1k5	s	1k5	2
3000	3k3	p	33k	1.1
3020	820R	s	2k2	1.37
3030	330R	s	2k7	1.12
3043	3k3	p	39k	1.08
3071	5k6	p	6k8	1.82
3083	3k3	p	47k	1.07
3090	390R	s	2k7	1.14
3095	3k9	p	15k	1.26
3116	3k3	p	56k	1.06
3147	3k3	p	68k	1.05
3170	470R	s	2k7	1.17
3172	3k3	p	82k	1.04
3195	3k3	p	100k	1.03
3197	4k7	p	10k	1.47
3200	1k	s	2k2	1.45

R_E	R_1	s/p	R_2	P_{REL}
3205	3k9	p	18k	1.22
3212	3k3	p	120k	1.03
3229	3k3	p	150k	1.02
3241	3k3	p	180k	1.02
3251	3k3	p	220k	1.02
3260	560R	s	2k7	1.21
3260	3k3	p	270k	1.01
3267	3k3	p	330k	1.01
3300	1k5	s	1k8	1.83
3313	3k9	p	22k	1.18
3328	5k6	p	8k2	1.68
3333	33R	s	3k3	1.01
3339	39R	s	3k3	1.01
3347	47R	s	3k3	1.01
3356	56R	s	3k3	1.02
3368	68R	s	3k3	1.02
3377	4k7	p	12k	1.39
3380	680R	s	2k7	1.25
3382	82R	s	3k3	1.02
3400	100R	s	3k3	1.03
3400	1k2	s	2k2	1.55
3400	6k8	p	6k8	2
3408	3k9	p	27k	1.14
3420	120R	s	3k3	1.04
3450	150R	s	3k3	1.05
3480	180R	s	3k3	1.05
3488	3k9	p	33k	1.12
3520	220R	s	3k3	1.07
3520	820R	s	2k7	1.3
3545	3k9	p	39k	1.1
3570	270R	s	3k3	1.08
3579	4k7	p	15k	1.31
3590	5k6	p	10k	1.56
3600	1k8	s	1k8	2
3601	3k9	p	47k	1.08
3630	330R	s	3k3	1.1
3646	3k9	p	56k	1.07
3688	3k9	p	68k	1.06
3690	390R	s	3k3	1.12
3700	1k	s	2k7	1.37
3700	1k5	s	2k2	1.68
3717	6k8	p	8k2	1.83
3723	3k9	p	82k	1.05
3727	4k7	p	18k	1.26
3754	3k9	p	100k	1.04
3770	470R	s	3k3	1.14
3777	3k9	p	120k	1.03
3801	3k9	p	150k	1.03
3817	3k9	p	180k	1.02
3818	5k6	p	12k	1.47
3832	3k9	p	220k	1.02
3844	3k9	p	270k	1.01
3854	3k9	p	330k	1.01
3860	560R	s	3k3	1.17
3861	3k9	p	390k	1.01
3873	4k7	p	22k	1.21
3900	1k2	s	2k7	1.44
3939	39R	s	3k9	1.01
3947	47R	s	3k9	1.01
3956	56R	s	3k9	1.01
3968	68R	s	3k9	1.02
3980	680R	s	3k3	1.21
3982	82R	s	3k9	1.02
4000	100R	s	3k9	1.03
4000	1k8	s	2k2	1.82
4003	4k7	p	27k	1.17
4020	120R	s	3k9	1.03
4048	6k8	p	10k	1.68
4050	150R	s	3k9	1.04
4078	5k6	p	15k	1.37
4080	180R	s	3k9	1.05
4100	8k2	p	8k2	2
4114	4k7	p	33k	1.14
4120	220R	s	3k9	1.06

R_E	R_1	s/p	R_2	P_{REL}
4120	820R	s	3k3	1.25
4170	270R	s	3k9	1.07
4195	4k7	p	39k	1.12
4200	1k5	s	2k7	1.56
4230	330R	s	3k9	1.08
4271	5k6	p	18k	1.31
4273	4k7	p	47k	1.1
4290	390R	s	3k9	1.1
4300	1k	s	3k3	1.3
4336	4k7	p	56k	1.08
4340	6k8	p	12k	1.57
4370	470R	s	3k9	1.12
4396	4k7	p	68k	1.07
4400	2k2	s	2k2	2
4445	4k7	p	82k	1.06
4460	560R	s	3k9	1.14
4464	5k6	p	22k	1.25
4489	4k7	p	100k	1.05
4500	1k2	s	3k3	1.36
4500	1k8	s	2k7	1.67
4505	8k2	p	10k	1.82
4523	4k7	p	120k	1.04
4557	4k7	p	150k	1.03
4580	680R	s	3k9	1.17
4580	4k7	p	180k	1.03
4602	4k7	p	220k	1.02
4620	4k7	p	270k	1.02
4634	4k7	p	330k	1.01
4638	5k6	p	27k	1.21
4644	4k7	p	390k	1.01
4653	4k7	p	470k	1.01
4679	6k8	p	15k	1.45
4720	820R	s	3k9	1.21
4747	47R	s	4k7	1.01
4756	56R	s	4k7	1.01
4768	68R	s	4k7	1.01
4782	82R	s	4k7	1.02
4788	5k6	p	33k	1.17
4800	100R	s	4k7	1.02
4800	1k5	s	3k3	1.45
4820	120R	s	4k7	1.03
4850	150R	s	4k7	1.03
4871	8k2	p	12k	1.68
4880	180R	s	4k7	1.04
4897	5k6	p	39k	1.14
4900	1k	s	3k9	1.26
4900	2k2	s	2k7	1.81
4920	220R	s	4k7	1.05
4935	6k8	p	18k	1.38
4970	270R	s	4k7	1.06
5000	10k	p	10k	2
5004	5k6	p	47k	1.12
5030	330R	s	4k7	1.07
5090	390R	s	4k7	1.08
5091	5k6	p	56k	1.1
5100	1k2	s	3k9	1.31
5100	1k8	s	3k3	1.55
5170	470R	s	4k7	1.1
5174	5k6	p	68k	1.08
5194	6k8	p	22k	1.31
5242	5k6	p	82k	1.07
5260	560R	s	4k7	1.12
5302	8k2	p	15k	1.55
5303	5k6	p	100k	1.06
5350	5k6	p	120k	1.05
5380	680R	s	4k7	1.14
5398	5k6	p	150k	1.04
5400	1k5	s	3k9	1.38
5400	2k7	s	2k7	2
5431	5k6	p	180k	1.03
5432	6k8	p	27k	1.25
5455	10k	p	12k	1.83
5461	5k6	p	220k	1.03
5486	5k6	p	270k	1.02

R_E	R_1	s/p	R_2	P_{REL}
5500	2k2	s	3k3	1.67
5507	5k6	p	330k	1.02
5520	820R	s	4k7	1.17
5521	5k6	p	390k	1.01
5534	5k6	p	470k	1.01
5545	5k6	p	560k	1.01
5634	8k2	p	18k	1.46
5638	6k8	p	33k	1.21
5656	56R	s	5k6	1.01
5668	68R	s	5k6	1.01
5682	82R	s	5k6	1.01
5700	100R	s	5k6	1.02
5700	1k	s	4k7	1.21
5700	1k8	s	3k9	1.46
5720	120R	s	5k6	1.02
5750	150R	s	5k6	1.03
5780	180R	s	5k6	1.03
5790	6k8	p	39k	1.17
5820	220R	s	5k6	1.04
5870	270R	s	5k6	1.05
5900	1k2	s	4k7	1.26
5930	330R	s	5k6	1.06
5941	6k8	p	47k	1.14
5974	8k2	p	22k	1.37
5990	390R	s	5k6	1.07
6000	2k7	s	3k3	1.82
6000	10k	p	15k	1.67
6000	12k	p	12k	2
6064	6k8	p	56k	1.12
6070	470R	s	5k6	1.08
6100	2k2	s	3k9	1.56
6160	560R	s	5k6	1.1
6182	6k8	p	68k	1.1
6200	1k5	s	4k7	1.32
6279	6k8	p	82k	1.08
6280	680R	s	5k6	1.12
6290	8k2	p	27k	1.3
6367	6k8	p	100k	1.07
6420	820R	s	5k6	1.15
6429	10k	p	18k	1.56
6435	6k8	p	120k	1.06
6500	1k8	s	4k7	1.38
6505	6k8	p	150k	1.05
6552	6k8	p	180k	1.04
6568	8k2	p	33k	1.25
6596	6k8	p	220k	1.03
6600	1k	s	5k6	1.18
6600	2k7	s	3k9	1.69
6600	3k3	s	3k3	2
6633	6k8	p	270k	1.03
6663	6k8	p	330k	1.02
6667	12k	p	15k	1.8
6683	6k8	p	390k	1.02
6703	6k8	p	470k	1.01
6718	6k8	p	560k	1.01
6733	6k8	p	680k	1.01
6775	8k2	p	39k	1.21
6800	1k2	s	5k6	1.21
6868	68R	s	6k8	1.01
6875	10k	p	22k	1.45
6882	82R	s	6k8	1.01
6900	100R	s	6k8	1.01
6900	2k2	s	4k7	1.47
6920	120R	s	6k8	1.02
6950	150R	s	6k8	1.02
6980	180R	s	6k8	1.03
6982	8k2	p	47k	1.17
7020	220R	s	6k8	1.03
7070	270R	s	6k8	1.04
7100	1k5	s	5k6	1.27
7130	330R	s	6k8	1.05
7153	8k2	p	56k	1.15
7190	390R	s	6k8	1.06
7200	3k3	s	3k9	1.85

R_E	R_1	s/p	R_2	P_{REL}
7200	12k	p	18k	1.67
7270	470R	s	6k8	1.07
7297	10k	p	27k	1.37
7318	8k2	p	68k	1.12
7360	560R	s	6k8	1.08
7400	1k8	s	5k6	1.32
7400	2k7	s	4k7	1.57
7455	8k2	p	82k	1.1
7480	680R	s	6k8	1.1
7500	15k	p	15k	2
7579	8k2	p	100k	1.08
7620	820R	s	6k8	1.12
7674	10k	p	33k	1.3
7676	8k2	p	120k	1.07
7765	12k	p	22k	1.55
7775	8k2	p	150k	1.05
7800	1k	s	6k8	1.15
7800	2k2	s	5k6	1.39
7800	3k9	s	3k9	2
7843	8k2	p	180k	1.05
7905	8k2	p	220k	1.04
7958	8k2	p	270k	1.03
7959	10k	p	39k	1.26
8000	1k2	s	6k8	1.18
8000	3k3	s	4k7	1.7
8001	8k2	p	330k	1.02
8031	8k2	p	390k	1.02
8059	8k2	p	470k	1.02
8082	8k2	p	560k	1.01
8102	8k2	p	680k	1.01
8119	8k2	p	820k	1.01
8182	15k	p	18k	1.83
8246	10k	p	47k	1.21
8282	82R	s	8k2	1.01
8300	100R	s	8k2	1.01
8300	1k5	s	6k8	1.22
8300	2k7	s	5k6	1.48

R_E	R_1	s/p	R_2	P_{REL}
8308	12k	p	27k	1.44
8320	120R	s	8k2	1.01
8350	150R	s	8k2	1.02
8380	180R	s	8k2	1.02
8420	220R	s	8k2	1.03
8470	270R	s	8k2	1.03
8485	10k	p	56k	1.18
8530	330R	s	8k2	1.04
8590	390R	s	8k2	1.05
8600	1k8	s	6k8	1.26
8600	3k9	s	4k7	1.83
8670	470R	s	8k2	1.06
8718	10k	p	68k	1.15
8760	560R	s	8k2	1.07
8800	12k	p	33k	1.36
8880	680R	s	8k2	1.08
8900	3k3	s	5k6	1.59
8913	10k	p	82k	1.12
8919	15k	p	22k	1.68
9000	2k2	s	6k8	1.32
9000	18k	p	18k	2
9020	820R	s	8k2	1.1
9091	10k	p	100k	1.1
9176	12k	p	39k	1.31
9200	1k	s	8k2	1.12
9231	10k	p	120k	1.08
9375	10k	p	150k	1.07
9400	1k2	s	8k2	1.15
9400	4k7	s	4k7	2
9474	10k	p	180k	1.06
9500	2k7	s	6k8	1.4
9500	3k9	s	5k6	1.7
9559	12k	p	47k	1.26
9565	10k	p	220k	1.05
9643	10k	p	270k	1.04
9643	15k	p	27k	1.56
9700	1k5	s	8k2	1.18

R_E	R_1	s/p	R_2	P_{REL}
9706	10k	p	330k	1.03
9750	10k	p	390k	1.03
9792	10k	p	470k	1.02
9825	10k	p	560k	1.02
9855	10k	p	680k	1.01
9880	10k	p	820k	1.01
9882	12k	p	56k	1.21
9900	18k	p	22k	1.82
9901	10k	p	1M	1.01

Table 6.3 *E24 equivalent resistances*

R_E	R_1	s/p	R_2	P_{REL}
1001	91R	s	910R	1.1
1005	1k2	p	6k2	1.19
1005	1k6	p	2k7	1.59
1008	1k1	p	12k	1.09
1010	10R	s	1k	1.01
1010	100R	s	910R	1.11
1010	330R	s	680R	1.49
1010	390R	s	620R	1.63
1011	11R	s	1k	1.01
1012	12R	s	1k	1.01
1013	13R	s	1k	1.01
1014	1k1	p	13k	1.08
1015	15R	s	1k	1.02
1016	16R	s	1k	1.02
1018	18R	s	1k	1.02
1018	1k3	p	4k7	1.28
1020	20R	s	1k	1.02
1020	110R	s	910R	1.12
1020	200R	s	820R	1.24
1020	270R	s	750R	1.36
1020	510R	s	510R	2
1020	1k2	p	6k8	1.18
1022	22R	s	1k	1.02
1024	24R	s	1k	1.02
1025	1k1	p	15k	1.07
1027	27R	s	1k	1.03
1029	1k1	p	16k	1.07
1029	1k8	p	2k4	1.75
1030	30R	s	1k	1.03
1030	120R	s	910R	1.13
1030	470R	s	560R	1.84
1031	1k5	p	3k3	1.45
1033	33R	s	1k	1.03
1034	1k2	p	7k5	1.16
1036	36R	s	1k	1.04
1036	1k3	p	5k1	1.25
1037	1k1	p	18k	1.06
1039	39R	s	1k	1.04
1040	130R	s	910R	1.14
1040	220R	s	820R	1.27
1040	360R	s	680R	1.53
1043	43R	s	1k	1.04
1043	1k1	p	20k	1.06
1043	1k6	p	3k	1.53
1047	47R	s	1k	1.05
1047	1k2	p	8k2	1.15
1048	1k1	p	22k	1.05
1048	2k	p	2k2	1.91
1050	300R	s	750R	1.4
1050	430R	s	620R	1.69
1051	51R	s	1k	1.05
1052	1k1	p	24k	1.05
1055	1k3	p	5k6	1.23
1056	56R	s	1k	1.06
1057	1k1	p	27k	1.04
1059	1k5	p	3k6	1.42
1060	150R	s	910R	1.16
1060	240R	s	820R	1.29
1060	1k2	p	9k1	1.13
1061	1k1	p	30k	1.04
1062	62R	s	1k	1.06
1065	1k1	p	33k	1.03
1067	1k1	p	36k	1.03
1068	68R	s	1k	1.07
1070	160R	s	910R	1.18
1070	390R	s	680R	1.57
1070	510R	s	560R	1.91
1070	1k1	p	39k	1.03
1071	1k2	p	10k	1.12
1073	1k1	p	43k	1.03
1075	75R	s	1k	1.08
1075	1k1	p	47k	1.02

R_E	R_1	s/p	R_2	P_{REL}
1075	1k3	p	6k2	1.21
1077	1k1	p	51k	1.02
1078	1k6	p	3k3	1.48
1079	1k1	p	56k	1.02
1080	330R	s	750R	1.44
1080	1k8	p	2k7	1.67
1081	1k1	p	62k	1.02
1082	82R	s	1k	1.08
1082	1k1	p	68k	1.02
1082	1k2	p	11k	1.11
1083	1k5	p	3k9	1.38
1084	1k1	p	75k	1.01
1085	1k1	p	82k	1.01
1087	1k1	p	91k	1.01
1088	1k1	p	100k	1.01
1089	1k1	p	110k	1.01
1090	180R	s	910R	1.2
1090	270R	s	820R	1.33
1090	470R	s	620R	1.76
1091	91R	s	1k	1.09
1091	1k2	p	12k	1.1
1091	1k3	p	6k8	1.19
1091	2k	p	2k4	1.83
1099	1k2	p	13k	1.09
1100	100R	s	1k	1.1
1100	2k2	p	2k2	2
1108	1k3	p	7k5	1.17
1108	1k6	p	3k6	1.44
1110	110R	s	1k	1.11
1110	200R	s	910R	1.22
1110	360R	s	750R	1.48
1110	430R	s	680R	1.63
1111	11R	s	1k1	1.01
1111	1k2	p	15k	1.08
1112	12R	s	1k1	1.01
1112	1k5	p	4k3	1.35
1113	13R	s	1k1	1.01
1115	15R	s	1k1	1.01
1116	16R	s	1k1	1.01
1116	1k2	p	16k	1.08
1118	18R	s	1k1	1.02
1120	20R	s	1k1	1.02
1120	120R	s	1k	1.12
1120	300R	s	820R	1.37
1120	560R	s	560R	2
1122	22R	s	1k1	1.02
1122	1k3	p	8k2	1.16
1124	24R	s	1k1	1.02
1125	1k2	p	18k	1.07
1125	1k8	p	3k	1.6
1127	27R	s	1k1	1.02
1130	30R	s	1k1	1.03
1130	130R	s	1k	1.13
1130	220R	s	910R	1.24
1130	510R	s	620R	1.82
1132	1k2	p	20k	1.06
1133	33R	s	1k1	1.03
1135	1k6	p	3k9	1.41
1136	36R	s	1k1	1.03
1137	1k5	p	4k7	1.32
1138	1k2	p	22k	1.05
1138	1k3	p	9k1	1.14
1139	39R	s	1k1	1.04
1140	390R	s	750R	1.52
1143	43R	s	1k1	1.04
1143	1k2	p	24k	1.05
1147	47R	s	1k1	1.04
1148	2k2	p	2k4	1.92
1149	1k2	p	27k	1.04
1149	2k	p	2k7	1.74
1150	150R	s	1k	1.15
1150	240R	s	910R	1.26
1150	330R	s	820R	1.4
1150	470R	s	680R	1.69

R_E	R_1	s/p	R_2	P_{REL}
1150	1k3	p	10k	1.13
1151	51R	s	1k1	1.05
1154	1k2	p	30k	1.04
1156	56R	s	1k1	1.05
1158	1k2	p	33k	1.04
1159	1k5	p	5k1	1.29
1160	160R	s	1k	1.16
1161	1k2	p	36k	1.03
1162	62R	s	1k1	1.06
1163	1k3	p	11k	1.12
1164	1k2	p	39k	1.03
1165	1k8	p	3k3	1.55
1166	1k6	p	4k3	1.37
1167	1k2	p	43k	1.03
1168	68R	s	1k1	1.06
1170	1k2	p	47k	1.03
1172	1k2	p	51k	1.02
1173	1k3	p	12k	1.11
1175	75R	s	1k1	1.07
1175	1k2	p	56k	1.02
1177	1k2	p	62k	1.02
1179	1k2	p	68k	1.02
1180	180R	s	1k	1.18
1180	270R	s	910R	1.3
1180	360R	s	820R	1.44
1180	430R	s	750R	1.57
1180	560R	s	620R	1.9
1181	1k2	p	75k	1.02
1182	82R	s	1k1	1.07
1182	1k3	p	13k	1.1
1183	1k2	p	82k	1.01
1183	1k5	p	5k6	1.27
1184	1k2	p	91k	1.01
1186	1k2	p	100k	1.01
1187	1k2	p	110k	1.01
1188	1k2	p	120k	1.01
1190	510R	s	680R	1.75

R_E	R_1	s/p	R_2	P_{REL}
1191	91R	s	1k1	1.08
1194	1k6	p	4k7	1.34
1196	1k3	p	15k	1.09
1200	100R	s	1k1	1.09
1200	200R	s	1k	1.2
1200	1k8	p	3k6	1.5
1200	2k	p	3k	1.67
1200	2k4	p	2k4	2
1202	1k3	p	16k	1.08
1208	1k5	p	6k2	1.24
1210	110R	s	1k1	1.1
1210	300R	s	910R	1.33
1210	390R	s	820R	1.48
1212	12R	s	1k2	1.01
1212	1k3	p	18k	1.07
1212	2k2	p	2k7	1.81
1213	13R	s	1k2	1.01
1215	15R	s	1k2	1.01
1216	16R	s	1k2	1.01
1218	18R	s	1k2	1.02
1218	1k6	p	5k1	1.31
1220	20R	s	1k2	1.02
1220	120R	s	1k1	1.11
1220	220R	s	1k	1.22
1220	470R	s	750R	1.63
1221	1k3	p	20k	1.07
1222	22R	s	1k2	1.02
1224	24R	s	1k2	1.02
1227	27R	s	1k2	1.02
1227	1k3	p	22k	1.06
1229	1k5	p	6k8	1.22
1230	30R	s	1k2	1.03
1230	130R	s	1k1	1.12
1232	1k8	p	3k9	1.46
1233	33R	s	1k2	1.03
1233	1k3	p	24k	1.05
1236	36R	s	1k2	1.03

R_E	R_1	s/p	R_2	P_{REL}
1239	39R	s	1k2	1.03
1240	240R	s	1k	1.24
1240	330R	s	910R	1.36
1240	560R	s	680R	1.82
1240	620R	s	620R	2
1240	1k3	p	27k	1.05
1243	43R	s	1k2	1.04
1244	1k6	p	5k6	1.29
1245	2k	p	3k3	1.61
1246	1k3	p	30k	1.04
1247	47R	s	1k2	1.04
1250	150R	s	1k1	1.14
1250	430R	s	820R	1.52
1250	1k5	p	7k5	1.2
1251	51R	s	1k2	1.04
1251	1k3	p	33k	1.04
1255	1k3	p	36k	1.04
1256	56R	s	1k2	1.05
1258	1k3	p	39k	1.03
1260	160R	s	1k1	1.15
1260	510R	s	750R	1.68
1262	62R	s	1k2	1.05
1262	1k3	p	43k	1.03
1265	1k3	p	47k	1.03
1268	68R	s	1k2	1.06
1268	1k3	p	51k	1.03
1268	1k5	p	8k2	1.18
1269	1k8	p	4k3	1.42
1269	2k2	p	3k	1.73
1270	270R	s	1k	1.27
1270	360R	s	910R	1.4
1271	1k3	p	56k	1.02
1271	2k4	p	2k7	1.89
1272	1k6	p	6k2	1.26
1273	1k3	p	62k	1.02
1275	75R	s	1k2	1.06
1276	1k3	p	68k	1.02
1278	1k3	p	75k	1.02
1280	180R	s	1k1	1.16
1280	1k3	p	82k	1.02
1282	82R	s	1k2	1.07
1282	1k3	p	91k	1.01
1283	1k3	p	100k	1.01
1285	1k3	p	110k	1.01
1286	1k3	p	120k	1.01
1286	2k	p	3k6	1.56
1287	1k3	p	130k	1.01
1288	1k5	p	9k1	1.16
1290	470R	s	820R	1.57
1291	91R	s	1k2	1.08
1295	1k6	p	6k8	1.24
1300	100R	s	1k2	1.08
1300	200R	s	1k1	1.18
1300	300R	s	1k	1.3
1300	390R	s	910R	1.43
1300	620R	s	680R	1.91
1302	1k8	p	4k7	1.38
1304	1k5	p	10k	1.15
1310	110R	s	1k2	1.09
1310	560R	s	750R	1.75
1313	13R	s	1k3	1.01
1315	15R	s	1k3	1.01
1316	16R	s	1k3	1.01
1318	18R	s	1k3	1.01
1319	1k6	p	7k5	1.21
1320	20R	s	1k3	1.02
1320	120R	s	1k2	1.1
1320	220R	s	1k1	1.2
1320	1k5	p	11k	1.14
1320	2k2	p	3k3	1.67
1322	22R	s	1k3	1.02
1322	2k	p	3k9	1.51
1324	24R	s	1k3	1.02
1327	27R	s	1k3	1.02

R_E	R_1	s/p	R_2	P_{REL}
1330	30R	s	1k3	1.02
1330	130R	s	1k2	1.11
1330	330R	s	1k	1.33
1330	510R	s	820R	1.62
1330	1k8	p	5k1	1.35
1333	33R	s	1k3	1.03
1333	1k5	p	12k	1.13
1333	2k4	p	3k	1.8
1336	36R	s	1k3	1.03
1339	39R	s	1k3	1.03
1339	1k6	p	8k2	1.2
1340	240R	s	1k1	1.22
1340	430R	s	910R	1.47
1343	43R	s	1k3	1.03
1345	1k5	p	13k	1.12
1347	47R	s	1k3	1.04
1350	150R	s	1k2	1.13
1350	2k7	p	2k7	2
1351	51R	s	1k3	1.04
1356	56R	s	1k3	1.04
1360	160R	s	1k2	1.13
1360	360R	s	1k	1.36
1360	680R	s	680R	2
1361	1k6	p	9k1	1.18
1362	62R	s	1k3	1.05
1362	1k8	p	5k6	1.32
1364	1k5	p	15k	1.1
1365	2k	p	4k3	1.47
1366	2k2	p	3k6	1.61
1368	68R	s	1k3	1.05
1370	270R	s	1k1	1.25
1370	620R	s	750R	1.83
1371	1k5	p	16k	1.09
1375	75R	s	1k3	1.06
1379	1k6	p	10k	1.16
1380	180R	s	1k2	1.15
1380	470R	s	910R	1.52

R_E	R_1	s/p	R_2	P_{REL}
1380	560R	s	820R	1.68
1382	82R	s	1k3	1.06
1385	1k5	p	18k	1.08
1389	2k4	p	3k3	1.73
1390	390R	s	1k	1.39
1391	91R	s	1k3	1.07
1395	1k5	p	20k	1.08
1395	1k8	p	6k2	1.29
1397	1k6	p	11k	1.15
1400	100R	s	1k3	1.08
1400	200R	s	1k2	1.17
1400	300R	s	1k1	1.27
1403	2k	p	4k7	1.43
1404	1k5	p	22k	1.07
1407	2k2	p	3k9	1.56
1410	110R	s	1k3	1.08
1412	1k5	p	24k	1.06
1412	1k6	p	12k	1.13
1420	120R	s	1k3	1.09
1420	220R	s	1k2	1.18
1420	510R	s	910R	1.56
1421	1k5	p	27k	1.06
1421	2k7	p	3k	1.9
1423	1k8	p	6k8	1.26
1425	1k6	p	13k	1.12
1429	1k5	p	30k	1.05
1430	130R	s	1k3	1.1
1430	330R	s	1k1	1.3
1430	430R	s	1k	1.43
1430	680R	s	750R	1.91
1435	1k5	p	33k	1.05
1437	2k	p	5k1	1.39
1440	240R	s	1k2	1.2
1440	620R	s	820R	1.76
1440	1k5	p	36k	1.04
1440	2k4	p	3k6	1.67
1444	1k5	p	39k	1.04

R_E	R_1	s/p	R_2	P_{REL}
1446	1k6	p	15k	1.11
1449	1k5	p	43k	1.03
1450	150R	s	1k3	1.12
1452	1k8	p	7k5	1.24
1454	1k5	p	47k	1.03
1455	1k6	p	16k	1.1
1455	2k2	p	4k3	1.51
1457	1k5	p	51k	1.03
1460	160R	s	1k3	1.12
1460	360R	s	1k1	1.33
1461	1k5	p	56k	1.03
1465	1k5	p	62k	1.02
1468	1k5	p	68k	1.02
1469	1k6	p	18k	1.09
1470	270R	s	1k2	1.23
1470	470R	s	1k	1.47
1470	560R	s	910R	1.62
1471	1k5	p	75k	1.02
1473	1k5	p	82k	1.02
1474	2k	p	5k6	1.36
1476	1k5	p	91k	1.02
1476	1k8	p	8k2	1.22
1478	1k5	p	100k	1.02
1480	180R	s	1k3	1.14
1480	1k5	p	110k	1.01
1481	1k5	p	120k	1.01
1481	1k6	p	20k	1.08
1483	1k5	p	130k	1.01
1485	1k5	p	150k	1.01
1485	2k7	p	3k3	1.82
1486	2k4	p	3k9	1.62
1490	390R	s	1k1	1.35
1492	1k6	p	22k	1.07
1499	2k2	p	4k7	1.47
1500	200R	s	1k3	1.15
1500	300R	s	1k2	1.25
1500	680R	s	820R	1.83
1500	750R	s	750R	2
1500	1k6	p	24k	1.07
1500	3k	p	3k	2
1503	1k8	p	9k1	1.2
1510	510R	s	1k	1.51
1510	1k6	p	27k	1.06
1512	2k	p	6k2	1.32
1515	15R	s	1k5	1.01
1516	16R	s	1k5	1.01
1518	18R	s	1k5	1.01
1519	1k6	p	30k	1.05
1520	20R	s	1k5	1.01
1520	220R	s	1k3	1.17
1522	22R	s	1k5	1.01
1524	24R	s	1k5	1.02
1525	1k8	p	10k	1.18
1526	1k6	p	33k	1.05
1527	27R	s	1k5	1.02
1530	30R	s	1k5	1.02
1530	330R	s	1k2	1.28
1530	430R	s	1k1	1.39
1530	620R	s	910R	1.68
1532	1k6	p	36k	1.04
1533	33R	s	1k5	1.02
1536	36R	s	1k5	1.02
1537	1k6	p	39k	1.04
1537	2k2	p	5k1	1.43
1539	39R	s	1k5	1.03
1540	240R	s	1k3	1.18
1540	2k4	p	4k3	1.56
1543	43R	s	1k5	1.03
1543	1k6	p	43k	1.04
1543	2k7	p	3k6	1.75
1545	2k	p	6k8	1.29
1547	47R	s	1k5	1.03
1547	1k6	p	47k	1.03
1547	1k8	p	11k	1.16

R_E	R_1	s/p	R_2	P_{REL}
1551	51R	s	1k5	1.03
1551	1k6	p	51k	1.03
1556	56R	s	1k5	1.04
1556	1k6	p	56k	1.03
1560	360R	s	1k2	1.3
1560	560R	s	1k	1.56
1560	1k6	p	62k	1.03
1562	62R	s	1k5	1.04
1563	1k6	p	68k	1.02
1565	1k8	p	12k	1.15
1567	1k6	p	75k	1.02
1568	68R	s	1k5	1.05
1569	1k6	p	82k	1.02
1570	270R	s	1k3	1.21
1570	470R	s	1k1	1.43
1570	750R	s	820R	1.91
1571	3k	p	3k3	1.91
1572	1k6	p	91k	1.02
1575	75R	s	1k5	1.05
1575	1k6	p	100k	1.02
1577	1k6	p	110k	1.01
1579	1k6	p	120k	1.01
1579	2k	p	7k5	1.27
1579	2k2	p	5k6	1.39
1581	1k6	p	130k	1.01
1581	1k8	p	13k	1.14
1582	82R	s	1k5	1.05
1583	1k6	p	150k	1.01
1584	1k6	p	160k	1.01
1589	2k4	p	4k7	1.51
1590	390R	s	1k2	1.33
1590	680R	s	910R	1.75
1591	91R	s	1k5	1.06
1595	2k7	p	3k9	1.69
1600	100R	s	1k5	1.07
1600	300R	s	1k3	1.23
1607	1k8	p	15k	1.12

R_E	R_1	s/p	R_2	P_{REL}
1608	2k	p	8k2	1.24
1610	110R	s	1k5	1.07
1610	510R	s	1k1	1.46
1616	16R	s	1k6	1.01
1618	18R	s	1k6	1.01
1618	1k8	p	16k	1.11
1620	20R	s	1k6	1.01
1620	120R	s	1k5	1.08
1620	620R	s	1k	1.62
1622	22R	s	1k6	1.01
1624	24R	s	1k6	1.02
1624	2k2	p	6k2	1.35
1627	27R	s	1k6	1.02
1630	30R	s	1k6	1.02
1630	130R	s	1k5	1.09
1630	330R	s	1k3	1.25
1630	430R	s	1k2	1.36
1632	2k4	p	5k1	1.47
1633	33R	s	1k6	1.02
1636	36R	s	1k6	1.02
1636	1k8	p	18k	1.1
1636	3k	p	3k6	1.83
1639	39R	s	1k6	1.02
1640	820R	s	820R	2
1640	2k	p	9k1	1.22
1643	43R	s	1k6	1.03
1647	47R	s	1k6	1.03
1650	150R	s	1k5	1.1
1650	3k3	p	3k3	2
1651	51R	s	1k6	1.03
1651	1k8	p	20k	1.09
1656	56R	s	1k6	1.03
1659	2k7	p	4k3	1.63
1660	160R	s	1k5	1.11
1660	360R	s	1k3	1.28
1660	560R	s	1k1	1.51
1660	750R	s	910R	1.82

R_E	R_1	s/p	R_2	P_{REL}
1662	62R	s	1k6	1.04
1662	2k2	p	6k8	1.32
1664	1k8	p	22k	1.08
1667	2k	p	10k	1.2
1668	68R	s	1k6	1.04
1670	470R	s	1k2	1.39
1674	1k8	p	24k	1.08
1675	75R	s	1k6	1.05
1680	180R	s	1k5	1.12
1680	680R	s	1k	1.68
1680	2k4	p	5k6	1.43
1682	82R	s	1k6	1.05
1688	1k8	p	27k	1.07
1690	390R	s	1k3	1.3
1691	91R	s	1k6	1.06
1692	2k	p	11k	1.18
1696	3k	p	3k9	1.77
1698	1k8	p	30k	1.06
1700	100R	s	1k6	1.06
1700	200R	s	1k5	1.13
1701	2k2	p	7k5	1.29
1707	1k8	p	33k	1.05
1710	110R	s	1k6	1.07
1710	510R	s	1k2	1.43
1714	1k8	p	36k	1.05
1714	2k	p	12k	1.17
1715	2k7	p	4k7	1.57
1720	120R	s	1k6	1.08
1720	220R	s	1k5	1.15
1720	620R	s	1k1	1.56
1721	1k8	p	39k	1.05
1722	3k3	p	3k6	1.92
1728	1k8	p	43k	1.04
1730	130R	s	1k6	1.08
1730	430R	s	1k3	1.33
1730	820R	s	910R	1.9
1730	2k4	p	6k2	1.39
1733	2k	p	13k	1.15
1734	1k8	p	47k	1.04
1735	2k2	p	8k2	1.27
1739	1k8	p	51k	1.04
1740	240R	s	1k5	1.16
1744	1k8	p	56k	1.03
1749	1k8	p	62k	1.03
1750	150R	s	1k6	1.09
1750	750R	s	1k	1.75
1754	1k8	p	68k	1.03
1758	1k8	p	75k	1.02
1760	160R	s	1k6	1.1
1760	560R	s	1k2	1.47
1761	1k8	p	82k	1.02
1765	1k8	p	91k	1.02
1765	2k	p	15k	1.13
1765	2k7	p	5k1	1.53
1767	3k	p	4k3	1.7
1768	1k8	p	100k	1.02
1770	270R	s	1k5	1.18
1770	470R	s	1k3	1.36
1771	1k8	p	110k	1.02
1772	2k2	p	9k1	1.24
1773	1k8	p	120k	1.02
1774	2k4	p	6k8	1.35
1775	1k8	p	130k	1.01
1778	2k	p	16k	1.13
1779	1k8	p	150k	1.01
1780	180R	s	1k6	1.11
1780	680R	s	1k1	1.62
1780	1k8	p	160k	1.01
1782	1k8	p	180k	1.01
1788	3k3	p	3k9	1.85
1800	200R	s	1k6	1.13
1800	300R	s	1k5	1.2
1800	2k	p	18k	1.11
1800	3k6	p	3k6	2

R_E	R_1	s/p	R_2	P_{REL}
1803	2k2	p	10k	1.22
1810	510R	s	1k3	1.39
1818	18R	s	1k8	1.01
1818	2k	p	20k	1.1
1818	2k4	p	7k5	1.32
1820	20R	s	1k8	1.01
1820	220R	s	1k6	1.14
1820	620R	s	1k2	1.52
1820	820R	s	1k	1.82
1820	910R	s	910R	2
1822	22R	s	1k8	1.01
1822	2k7	p	5k6	1.48
1824	24R	s	1k8	1.01
1827	27R	s	1k8	1.02
1830	30R	s	1k8	1.02
1830	330R	s	1k5	1.22
1831	3k	p	4k7	1.64
1833	33R	s	1k8	1.02
1833	2k	p	22k	1.09
1833	2k2	p	11k	1.2
1836	36R	s	1k8	1.02
1839	39R	s	1k8	1.02
1840	240R	s	1k6	1.15
1843	43R	s	1k8	1.02
1846	2k	p	24k	1.08
1847	47R	s	1k8	1.03
1850	750R	s	1k1	1.68
1851	51R	s	1k8	1.03
1856	56R	s	1k8	1.03
1857	2k4	p	8k2	1.29
1859	2k2	p	12k	1.18
1860	360R	s	1k5	1.24
1860	560R	s	1k3	1.43
1862	62R	s	1k8	1.03
1862	2k	p	27k	1.07
1867	3k3	p	4k3	1.77
1868	68R	s	1k8	1.04

R_E	R_1	s/p	R_2	P_{REL}
1870	270R	s	1k6	1.17
1872	3k6	p	3k9	1.92
1875	75R	s	1k8	1.04
1875	2k	p	30k	1.07
1880	680R	s	1k2	1.57
1881	2k7	p	6k2	1.44
1882	82R	s	1k8	1.05
1882	2k2	p	13k	1.17
1886	2k	p	33k	1.06
1889	3k	p	5k1	1.59
1890	390R	s	1k5	1.26
1891	91R	s	1k8	1.05
1895	2k	p	36k	1.06
1899	2k4	p	9k1	1.26
1900	100R	s	1k8	1.06
1900	300R	s	1k6	1.19
1902	2k	p	39k	1.05
1910	110R	s	1k8	1.06
1910	910R	s	1k	1.91
1911	2k	p	43k	1.05
1918	2k	p	47k	1.04
1919	2k2	p	15k	1.15
1920	120R	s	1k8	1.07
1920	620R	s	1k3	1.48
1920	820R	s	1k1	1.75
1925	2k	p	51k	1.04
1930	130R	s	1k8	1.07
1930	330R	s	1k6	1.21
1930	430R	s	1k5	1.29
1931	2k	p	56k	1.04
1933	2k7	p	6k8	1.4
1934	2k2	p	16k	1.14
1935	2k4	p	10k	1.24
1938	2k	p	62k	1.03
1939	3k3	p	4k7	1.7
1943	2k	p	68k	1.03
1948	2k	p	75k	1.03

R_E	R_1	s/p	R_2	P_{REL}
1950	150R	s	1k8	1.08
1950	750R	s	1k2	1.63
1950	3k9	p	3k9	2
1952	2k	p	82k	1.02
1953	3k	p	5k6	1.54
1957	2k	p	91k	1.02
1959	3k6	p	4k3	1.84
1960	160R	s	1k8	1.09
1960	360R	s	1k6	1.23
1960	2k2	p	18k	1.12
1961	2k	p	100k	1.02
1964	2k	p	110k	1.02
1967	2k	p	120k	1.02
1970	470R	s	1k5	1.31
1970	2k	p	130k	1.02
1970	2k4	p	11k	1.22
1974	2k	p	150k	1.01
1975	2k	p	160k	1.01
1978	2k	p	180k	1.01
1980	180R	s	1k8	1.1
1980	680R	s	1k3	1.52
1980	2k	p	200k	1.01
1982	2k2	p	20k	1.11
1985	2k7	p	7k5	1.36
1990	390R	s	1k6	1.24
2000	200R	s	1k8	1.11
2000	1k	s	1k	2
2000	2k2	p	22k	1.1
2000	2k4	p	12k	1.2
2004	3k3	p	5k1	1.65
2010	510R	s	1k5	1.34
2010	910R	s	1k1	1.83
2015	2k2	p	24k	1.09
2020	20R	s	2k	1.01
2020	220R	s	1k8	1.12
2020	820R	s	1k2	1.68
2022	22R	s	2k	1.01

R_E	R_1	s/p	R_2	P_{REL}
2022	3k	p	6k2	1.48
2024	24R	s	2k	1.01
2026	2k4	p	13k	1.18
2027	27R	s	2k	1.01
2030	30R	s	2k	1.02
2030	430R	s	1k6	1.27
2031	2k7	p	8k2	1.33
2033	33R	s	2k	1.02
2034	2k2	p	27k	1.08
2036	36R	s	2k	1.02
2039	39R	s	2k	1.02
2039	3k6	p	4k7	1.77
2040	240R	s	1k8	1.13
2043	43R	s	2k	1.02
2045	3k9	p	4k3	1.91
2047	47R	s	2k	1.02
2050	750R	s	1k3	1.58
2050	2k2	p	30k	1.07
2051	51R	s	2k	1.03
2056	56R	s	2k	1.03
2060	560R	s	1k5	1.37
2062	62R	s	2k	1.03
2063	2k2	p	33k	1.07
2068	68R	s	2k	1.03
2069	2k4	p	15k	1.16
2070	270R	s	1k8	1.15
2070	470R	s	1k6	1.29
2073	2k2	p	36k	1.06
2075	75R	s	2k	1.04
2076	3k3	p	5k6	1.59
2082	82R	s	2k	1.04
2082	2k7	p	9k1	1.3
2082	3k	p	6k8	1.44
2083	2k2	p	39k	1.06
2087	2k4	p	16k	1.15
2091	91R	s	2k	1.05
2093	2k2	p	43k	1.05

R_E	R_1	s/p	R_2	P_{REL}
2100	100R	s	2k	1.05
2100	300R	s	1k8	1.17
2100	1k	s	1k1	1.91
2102	2k2	p	47k	1.05
2109	2k2	p	51k	1.04
2110	110R	s	2k	1.06
2110	510R	s	1k6	1.32
2110	910R	s	1k2	1.76
2110	3k6	p	5k1	1.71
2117	2k2	p	56k	1.04
2118	2k4	p	18k	1.13
2120	120R	s	2k	1.06
2120	620R	s	1k5	1.41
2120	820R	s	1k3	1.63
2125	2k2	p	62k	1.04
2126	2k7	p	10k	1.27
2130	130R	s	2k	1.07
2130	330R	s	1k8	1.18
2131	2k2	p	68k	1.03
2131	3k9	p	4k7	1.83
2137	2k2	p	75k	1.03
2143	2k2	p	82k	1.03
2143	2k4	p	20k	1.12
2143	3k	p	7k5	1.4
2148	2k2	p	91k	1.02
2150	150R	s	2k	1.08
2150	4k3	p	4k3	2
2153	2k2	p	100k	1.02
2154	3k3	p	6k2	1.53
2157	2k2	p	110k	1.02
2160	160R	s	2k	1.08
2160	360R	s	1k8	1.2
2160	560R	s	1k6	1.35
2160	2k2	p	120k	1.02
2163	2k2	p	130k	1.02
2164	2k4	p	22k	1.11
2168	2k2	p	150k	1.01

R_E	R_1	s/p	R_2	P_{REL}
2168	2k7	p	11k	1.25
2170	2k2	p	160k	1.01
2173	2k2	p	180k	1.01
2176	2k2	p	200k	1.01
2178	2k2	p	220k	1.01
2180	180R	s	2k	1.09
2180	680R	s	1k5	1.45
2182	2k4	p	24k	1.1
2190	390R	s	1k8	1.22
2191	3k6	p	5k6	1.64
2196	3k	p	8k2	1.37
2200	200R	s	2k	1.1
2200	1k	s	1k2	1.83
2200	1k1	s	1k1	2
2204	2k4	p	27k	1.09
2204	2k7	p	12k	1.23
2210	910R	s	1k3	1.7
2210	3k9	p	5k1	1.76
2220	220R	s	2k	1.11
2220	620R	s	1k6	1.39
2222	22R	s	2k2	1.01
2222	2k4	p	30k	1.08
2222	3k3	p	6k8	1.49
2224	24R	s	2k2	1.01
2227	27R	s	2k2	1.01
2230	30R	s	2k2	1.01
2230	430R	s	1k8	1.24
2233	33R	s	2k2	1.02
2236	36R	s	2k2	1.02
2236	2k7	p	13k	1.21
2237	2k4	p	33k	1.07
2239	39R	s	2k2	1.02
2240	240R	s	2k	1.12
2243	43R	s	2k2	1.02
2246	4k3	p	4k7	1.91
2247	47R	s	2k2	1.02
2250	750R	s	1k5	1.5

R_E	R_1	s/p	R_2	P_{REL}
2250	2k4	p	36k	1.07
2251	51R	s	2k2	1.02
2256	56R	s	2k2	1.03
2256	3k	p	9k1	1.33
2261	2k4	p	39k	1.06
2262	62R	s	2k2	1.03
2268	68R	s	2k2	1.03
2270	270R	s	2k	1.14
2270	470R	s	1k8	1.26
2273	2k4	p	43k	1.06
2275	75R	s	2k2	1.03
2278	3k6	p	6k2	1.58
2280	680R	s	1k6	1.43
2282	82R	s	2k2	1.04
2283	2k4	p	47k	1.05
2288	2k7	p	15k	1.18
2291	91R	s	2k2	1.04
2292	2k4	p	51k	1.05
2292	3k3	p	7k5	1.44
2299	3k9	p	5k6	1.7
2300	100R	s	2k2	1.05
2300	300R	s	2k	1.15
2300	1k	s	1k3	1.77
2300	1k1	s	1k2	1.92
2301	2k4	p	56k	1.04
2308	3k	p	10k	1.3
2310	110R	s	2k2	1.05
2310	510R	s	1k8	1.28
2310	2k7	p	16k	1.17
2311	2k4	p	62k	1.04
2318	2k4	p	68k	1.04
2320	120R	s	2k2	1.05
2320	820R	s	1k5	1.55
2326	2k4	p	75k	1.03
2330	130R	s	2k2	1.06
2330	330R	s	2k	1.17
2332	2k4	p	82k	1.03
2333	4k3	p	5k1	1.84
2338	2k4	p	91k	1.03
2344	2k4	p	100k	1.02
2348	2k7	p	18k	1.15
2349	2k4	p	110k	1.02
2350	150R	s	2k2	1.07
2350	750R	s	1k6	1.47
2350	4k7	p	4k7	2
2353	2k4	p	120k	1.02
2353	3k3	p	8k2	1.4
2354	3k6	p	6k8	1.53
2356	2k4	p	130k	1.02
2357	3k	p	11k	1.27
2360	160R	s	2k2	1.07
2360	360R	s	2k	1.18
2360	560R	s	1k8	1.31
2362	2k4	p	150k	1.02
2365	2k4	p	160k	1.02
2368	2k4	p	180k	1.01
2372	2k4	p	200k	1.01
2374	2k4	p	220k	1.01
2376	2k4	p	240k	1.01
2379	2k7	p	20k	1.14
2380	180R	s	2k2	1.08
2390	390R	s	2k	1.2
2394	3k9	p	6k2	1.63
2400	200R	s	2k2	1.09
2400	1k1	s	1k3	1.85
2400	1k2	s	1k2	2
2400	3k	p	12k	1.25
2405	2k7	p	22k	1.12
2410	910R	s	1k5	1.61
2420	220R	s	2k2	1.1
2420	620R	s	1k8	1.34
2420	820R	s	1k6	1.51
2422	3k3	p	9k1	1.36
2424	24R	s	2k4	1.01

R_E	R_1	s/p	R_2	P_{REL}
2427	27R	s	2k4	1.01
2427	2k7	p	24k	1.11
2430	30R	s	2k4	1.01
2430	430R	s	2k	1.22
2432	3k6	p	7k5	1.48
2432	4k3	p	5k6	1.77
2433	33R	s	2k4	1.01
2436	36R	s	2k4	1.02
2438	3k	p	13k	1.23
2439	39R	s	2k4	1.02
2440	240R	s	2k2	1.11
2443	43R	s	2k4	1.02
2446	4k7	p	5k1	1.92
2447	47R	s	2k4	1.02
2451	51R	s	2k4	1.02
2455	2k7	p	27k	1.1
2456	56R	s	2k4	1.02
2462	62R	s	2k4	1.03
2468	68R	s	2k4	1.03
2470	270R	s	2k2	1.12
2470	470R	s	2k	1.24
2475	75R	s	2k4	1.03
2477	2k7	p	30k	1.09
2479	3k9	p	6k8	1.57
2480	680R	s	1k8	1.38
2481	3k3	p	10k	1.33
2482	82R	s	2k4	1.03
2491	91R	s	2k4	1.04
2496	2k7	p	33k	1.08
2500	100R	s	2k4	1.04
2500	300R	s	2k2	1.14
2500	1k	s	1k5	1.67
2500	1k2	s	1k3	1.92
2500	3k	p	15k	1.2
2502	3k6	p	8k2	1.44
2510	110R	s	2k4	1.05
2510	510R	s	2k	1.26

R_E	R_1	s/p	R_2	P_{REL}
2510	910R	s	1k6	1.57
2512	2k7	p	36k	1.08
2520	120R	s	2k4	1.05
2525	2k7	p	39k	1.07
2526	3k	p	16k	1.19
2530	130R	s	2k4	1.05
2530	330R	s	2k2	1.15
2538	3k3	p	11k	1.3
2539	4k3	p	6k2	1.69
2540	2k7	p	43k	1.06
2550	150R	s	2k4	1.06
2550	750R	s	1k8	1.42
2550	5k1	p	5k1	2
2553	2k7	p	47k	1.06
2555	4k7	p	5k6	1.84
2560	160R	s	2k4	1.07
2560	360R	s	2k2	1.16
2560	560R	s	2k	1.28
2564	2k7	p	51k	1.05
2566	3k9	p	7k5	1.52
2571	3k	p	18k	1.17
2576	2k7	p	56k	1.05
2580	180R	s	2k4	1.08
2580	3k6	p	9k1	1.4
2587	2k7	p	62k	1.04
2588	3k3	p	12k	1.28
2590	390R	s	2k2	1.18
2597	2k7	p	68k	1.04
2600	200R	s	2k4	1.08
2600	1k	s	1k6	1.63
2600	1k1	s	1k5	1.73
2600	1k3	s	1k3	2
2606	2k7	p	75k	1.04
2609	3k	p	20k	1.15
2614	2k7	p	82k	1.03
2620	220R	s	2k4	1.09
2620	620R	s	2k	1.31

R_E	R_1	s/p	R_2	P_{REL}
2620	820R	s	1k8	1.46
2622	2k7	p	91k	1.03
2629	2k7	p	100k	1.03
2630	430R	s	2k2	1.2
2632	3k3	p	13k	1.25
2634	4k3	p	6k8	1.63
2635	2k7	p	110k	1.02
2640	240R	s	2k4	1.1
2640	3k	p	22k	1.14
2641	2k7	p	120k	1.02
2643	3k9	p	8k2	1.48
2645	2k7	p	130k	1.02
2647	3k6	p	10k	1.36
2652	2k7	p	150k	1.02
2655	2k7	p	160k	1.02
2660	2k7	p	180k	1.02
2664	2k7	p	200k	1.01
2667	2k7	p	220k	1.01
2667	3k	p	24k	1.13
2669	5k1	p	5k6	1.91
2670	270R	s	2k4	1.11
2670	470R	s	2k2	1.21
2670	2k7	p	240k	1.01
2673	2k7	p	270k	1.01
2673	4k7	p	6k2	1.76
2680	680R	s	2k	1.34
2700	300R	s	2k4	1.13
2700	1k1	s	1k6	1.69
2700	1k2	s	1k5	1.8
2700	3k	p	27k	1.11
2705	3k3	p	15k	1.22
2710	510R	s	2k2	1.23
2710	910R	s	1k8	1.51
2712	3k6	p	11k	1.33
2727	27R	s	2k7	1.01
2727	3k	p	30k	1.1
2730	30R	s	2k7	1.01
2730	330R	s	2k4	1.14
2730	3k9	p	9k1	1.43
2733	33R	s	2k7	1.01
2733	4k3	p	7k5	1.57
2736	36R	s	2k7	1.01
2736	3k3	p	16k	1.21
2739	39R	s	2k7	1.01
2743	43R	s	2k7	1.02
2747	47R	s	2k7	1.02
2750	750R	s	2k	1.38
2750	3k	p	33k	1.09
2751	51R	s	2k7	1.02
2756	56R	s	2k7	1.02
2760	360R	s	2k4	1.15
2760	560R	s	2k2	1.25
2762	62R	s	2k7	1.02
2768	68R	s	2k7	1.03
2769	3k	p	36k	1.08
2769	3k6	p	12k	1.3
2775	75R	s	2k7	1.03
2779	4k7	p	6k8	1.69
2782	82R	s	2k7	1.03
2786	3k	p	39k	1.08
2789	3k3	p	18k	1.18
2790	390R	s	2k4	1.16
2791	91R	s	2k7	1.03
2798	5k1	p	6k2	1.82
2800	100R	s	2k7	1.04
2800	1k	s	1k8	1.56
2800	1k2	s	1k6	1.75
2800	1k3	s	1k5	1.87
2800	5k6	p	5k6	2
2804	3k	p	43k	1.07
2806	3k9	p	10k	1.39
2810	110R	s	2k7	1.04
2819	3k6	p	13k	1.28
2820	120R	s	2k7	1.04

R_E	R_1	s/p	R_2	P_{REL}
2820	620R	s	2k2	1.28
2820	820R	s	2k	1.41
2820	3k	p	47k	1.06
2821	4k3	p	8k2	1.52
2830	130R	s	2k7	1.05
2830	430R	s	2k4	1.18
2833	3k	p	51k	1.06
2833	3k3	p	20k	1.17
2847	3k	p	56k	1.05
2850	150R	s	2k7	1.06
2860	160R	s	2k7	1.06
2862	3k	p	62k	1.05
2870	470R	s	2k4	1.2
2870	3k3	p	22k	1.15
2873	3k	p	68k	1.04
2879	3k9	p	11k	1.35
2880	180R	s	2k7	1.07
2880	680R	s	2k2	1.31
2885	3k	p	75k	1.04
2889	4k7	p	7k5	1.63
2894	3k	p	82k	1.04
2900	200R	s	2k7	1.07
2900	1k1	s	1k8	1.61
2900	1k3	s	1k6	1.81
2901	3k3	p	24k	1.14
2903	3k6	p	15k	1.24
2904	3k	p	91k	1.03
2910	510R	s	2k4	1.21
2910	910R	s	2k	1.46
2913	3k	p	100k	1.03
2914	5k1	p	6k8	1.75
2920	220R	s	2k7	1.08
2920	3k	p	110k	1.03
2920	4k3	p	9k1	1.47
2927	3k	p	120k	1.03
2932	3k	p	130k	1.02
2939	3k6	p	16k	1.23

R_E	R_1	s/p	R_2	P_{REL}
2940	240R	s	2k7	1.09
2941	3k	p	150k	1.02
2941	3k3	p	27k	1.12
2942	5k6	p	6k2	1.9
2943	3k9	p	12k	1.33
2945	3k	p	160k	1.02
2950	750R	s	2k2	1.34
2951	3k	p	180k	1.02
2956	3k	p	200k	1.02
2960	560R	s	2k4	1.23
2960	3k	p	220k	1.01
2963	3k	p	240k	1.01
2967	3k	p	270k	1.01
2970	270R	s	2k7	1.1
2970	3k	p	300k	1.01
2973	3k3	p	30k	1.11
2988	4k7	p	8k2	1.57
3000	300R	s	2k7	1.11
3000	1k	s	2k	1.5
3000	1k2	s	1k8	1.67
3000	1k5	s	1k5	2
3000	3k3	p	33k	1.1
3000	3k6	p	18k	1.2
3000	3k9	p	13k	1.3
3007	4k3	p	10k	1.43
3020	620R	s	2k4	1.26
3020	820R	s	2k2	1.37
3023	3k3	p	36k	1.09
3030	30R	s	3k	1.01
3030	330R	s	2k7	1.12
3033	33R	s	3k	1.01
3036	36R	s	3k	1.01
3036	5k1	p	7k5	1.68
3039	39R	s	3k	1.01
3043	43R	s	3k	1.01
3043	3k3	p	39k	1.08
3047	47R	s	3k	1.02

R_E	R_1	s/p	R_2	P_{REL}
3051	51R	s	3k	1.02
3051	3k6	p	20k	1.18
3056	56R	s	3k	1.02
3060	360R	s	2k7	1.13
3062	62R	s	3k	1.02
3065	3k3	p	43k	1.08
3068	68R	s	3k	1.02
3071	5k6	p	6k8	1.82
3075	75R	s	3k	1.03
3080	680R	s	2k4	1.28
3082	82R	s	3k	1.03
3083	3k3	p	47k	1.07
3090	390R	s	2k7	1.14
3091	91R	s	3k	1.03
3092	4k3	p	11k	1.39
3094	3k6	p	22k	1.16
3095	3k9	p	15k	1.26
3099	3k3	p	51k	1.06
3099	4k7	p	9k1	1.52
3100	100R	s	3k	1.03
3100	1k1	s	2k	1.55
3100	1k3	s	1k8	1.72
3100	1k5	s	1k6	1.94
3100	6k2	p	6k2	2
3110	110R	s	3k	1.04
3110	910R	s	2k2	1.41
3116	3k3	p	56k	1.06
3120	120R	s	3k	1.04
3130	130R	s	3k	1.04
3130	430R	s	2k7	1.16
3130	3k6	p	24k	1.15
3133	3k3	p	62k	1.05
3136	3k9	p	16k	1.24
3144	5k1	p	8k2	1.62
3147	3k3	p	68k	1.05
3150	150R	s	3k	1.05
3150	750R	s	2k4	1.31
3160	160R	s	3k	1.05
3161	3k3	p	75k	1.04
3166	4k3	p	12k	1.36
3170	470R	s	2k7	1.17
3172	3k3	p	82k	1.04
3176	3k6	p	27k	1.13
3180	180R	s	3k	1.06
3185	3k3	p	91k	1.04
3195	3k3	p	100k	1.03
3197	4k7	p	10k	1.47
3200	200R	s	3k	1.07
3200	1k	s	2k2	1.45
3200	1k2	s	2k	1.6
3200	1k6	s	1k6	2
3204	3k3	p	110k	1.03
3205	3k9	p	18k	1.22
3206	5k6	p	7k5	1.75
3210	510R	s	2k7	1.19
3212	3k3	p	120k	1.03
3214	3k6	p	30k	1.12
3218	3k3	p	130k	1.03
3220	220R	s	3k	1.07
3220	820R	s	2k4	1.34
3229	3k3	p	150k	1.02
3231	4k3	p	13k	1.33
3233	3k3	p	160k	1.02
3240	240R	s	3k	1.08
3241	3k3	p	180k	1.02
3243	6k2	p	6k8	1.91
3246	3k3	p	200k	1.02
3246	3k6	p	33k	1.11
3251	3k3	p	220k	1.02
3255	3k3	p	240k	1.01
3260	560R	s	2k7	1.21
3260	3k3	p	270k	1.01
3264	3k3	p	300k	1.01
3264	3k9	p	20k	1.2

R_E	R_1	s/p	R_2	P_{REL}
3267	3k3	p	330k	1.01
3268	5k1	p	9k1	1.56
3270	270R	s	3k	1.09
3273	3k6	p	36k	1.1
3293	4k7	p	11k	1.43
3296	3k6	p	39k	1.09
3300	300R	s	3k	1.1
3300	1k1	s	2k2	1.5
3300	1k3	s	2k	1.65
3300	1k5	s	1k8	1.83
3310	910R	s	2k4	1.38
3313	3k9	p	22k	1.18
3320	620R	s	2k7	1.23
3322	3k6	p	43k	1.08
3328	5k6	p	8k2	1.68
3330	330R	s	3k	1.11
3333	33R	s	3k3	1.01
3336	36R	s	3k3	1.01
3339	39R	s	3k3	1.01
3342	4k3	p	15k	1.29
3343	43R	s	3k3	1.01
3344	3k6	p	47k	1.08
3347	47R	s	3k3	1.01
3351	51R	s	3k3	1.02
3355	3k9	p	24k	1.16
3356	56R	s	3k3	1.02
3360	360R	s	3k	1.12
3362	62R	s	3k3	1.02
3363	3k6	p	51k	1.07
3368	68R	s	3k3	1.02
3375	75R	s	3k3	1.02
3377	4k7	p	12k	1.39
3377	5k1	p	10k	1.51
3380	680R	s	2k7	1.25
3382	82R	s	3k3	1.02
3383	3k6	p	56k	1.06
3389	4k3	p	16k	1.27
3390	390R	s	3k	1.13
3391	91R	s	3k3	1.03
3394	6k2	p	7k5	1.83
3400	100R	s	3k3	1.03
3400	1k	s	2k4	1.42
3400	1k2	s	2k2	1.55
3400	1k6	s	1k8	1.89
3400	6k8	p	6k8	2
3402	3k6	p	62k	1.06
3408	3k9	p	27k	1.14
3410	110R	s	3k3	1.03
3419	3k6	p	68k	1.05
3420	120R	s	3k3	1.04
3430	130R	s	3k3	1.04
3430	430R	s	3k	1.14
3435	3k6	p	75k	1.05
3449	3k6	p	82k	1.04
3450	150R	s	3k3	1.05
3450	750R	s	2k7	1.28
3451	3k9	p	30k	1.13
3452	4k7	p	13k	1.36
3460	160R	s	3k3	1.05
3463	3k6	p	91k	1.04
3467	5k6	p	9k1	1.62
3470	470R	s	3k	1.16
3471	4k3	p	18k	1.24
3475	3k6	p	100k	1.04
3480	180R	s	3k3	1.05
3484	5k1	p	11k	1.46
3486	3k6	p	110k	1.03
3488	3k9	p	33k	1.12
3495	3k6	p	120k	1.03
3500	200R	s	3k3	1.06
3500	1k1	s	2k4	1.46
3500	1k3	s	2k2	1.59
3500	1k5	s	2k	1.75
3503	3k6	p	130k	1.03

R_E	R_1	s/p	R_2	P_{REL}
3510	510R	s	3k	1.17
3516	3k6	p	150k	1.02
3519	3k9	p	36k	1.11
3520	220R	s	3k3	1.07
3520	820R	s	2k7	1.3
3521	3k6	p	160k	1.02
3529	3k6	p	180k	1.02
3531	6k2	p	8k2	1.76
3536	3k6	p	200k	1.02
3539	4k3	p	20k	1.22
3540	240R	s	3k3	1.07
3542	3k6	p	220k	1.02
3545	3k9	p	39k	1.1
3547	3k6	p	240k	1.02
3553	3k6	p	270k	1.01
3557	3k6	p	300k	1.01
3560	560R	s	3k	1.19
3561	3k6	p	330k	1.01
3564	3k6	p	360k	1.01
3566	6k8	p	7k5	1.91
3570	270R	s	3k3	1.08
3576	3k9	p	43k	1.09
3579	4k7	p	15k	1.31
3579	5k1	p	12k	1.43
3590	5k6	p	10k	1.56
3597	4k3	p	22k	1.2
3600	300R	s	3k3	1.09
3600	1k2	s	2k4	1.5
3600	1k6	s	2k	1.8
3600	1k8	s	1k8	2
3601	3k9	p	47k	1.08
3610	910R	s	2k7	1.34
3620	620R	s	3k	1.21
3623	3k9	p	51k	1.08
3630	330R	s	3k3	1.1
3633	4k7	p	16k	1.29
3636	36R	s	3k6	1.01

R_E	R_1	s/p	R_2	P_{REL}
3639	39R	s	3k6	1.01
3643	43R	s	3k6	1.01
3646	3k9	p	56k	1.07
3647	47R	s	3k6	1.01
3647	4k3	p	24k	1.18
3651	51R	s	3k6	1.01
3656	56R	s	3k6	1.02
3660	360R	s	3k3	1.11
3662	62R	s	3k6	1.02
3663	5k1	p	13k	1.39
3668	68R	s	3k6	1.02
3669	3k9	p	62k	1.06
3675	75R	s	3k6	1.02
3680	680R	s	3k	1.23
3682	82R	s	3k6	1.02
3688	3k9	p	68k	1.06
3688	6k2	p	9k1	1.68
3690	390R	s	3k3	1.12
3691	91R	s	3k6	1.03
3700	100R	s	3k6	1.03
3700	1k	s	2k7	1.37
3700	1k3	s	2k4	1.54
3700	1k5	s	2k2	1.68
3707	3k9	p	75k	1.05
3709	4k3	p	27k	1.16
3710	110R	s	3k6	1.03
3711	5k6	p	11k	1.51
3717	6k8	p	8k2	1.83
3720	120R	s	3k6	1.03
3723	3k9	p	82k	1.05
3727	4k7	p	18k	1.26
3730	130R	s	3k6	1.04
3730	430R	s	3k3	1.13
3740	3k9	p	91k	1.04
3750	150R	s	3k6	1.04
3750	750R	s	3k	1.25
3750	7k5	p	7k5	2

R_E	R_1	s/p	R_2	P_{REL}
3754	3k9	p	100k	1.04
3760	160R	s	3k6	1.04
3761	4k3	p	30k	1.14
3766	3k9	p	110k	1.04
3770	470R	s	3k3	1.14
3777	3k9	p	120k	1.03
3780	180R	s	3k6	1.05
3786	3k9	p	130k	1.03
3800	200R	s	3k6	1.06
3800	1k1	s	2k7	1.41
3800	1k6	s	2k2	1.73
3800	1k8	s	2k	1.9
3801	3k9	p	150k	1.03
3804	4k3	p	33k	1.13
3806	4k7	p	20k	1.24
3806	5k1	p	15k	1.34
3807	3k9	p	160k	1.02
3810	510R	s	3k3	1.15
3817	3k9	p	180k	1.02
3818	5k6	p	12k	1.47
3820	220R	s	3k6	1.06
3820	820R	s	3k	1.27
3825	3k9	p	200k	1.02
3827	6k2	p	10k	1.62
3832	3k9	p	220k	1.02
3838	3k9	p	240k	1.02
3840	240R	s	3k6	1.07
3841	4k3	p	36k	1.12
3844	3k9	p	270k	1.01
3850	3k9	p	300k	1.01
3854	3k9	p	330k	1.01
3858	3k9	p	360k	1.01
3860	560R	s	3k3	1.17
3861	3k9	p	390k	1.01
3867	5k1	p	16k	1.32
3870	270R	s	3k6	1.08
3873	4k3	p	39k	1.11

R_E	R_1	s/p	R_2	P_{REL}
3873	4k7	p	22k	1.21
3892	6k8	p	9k1	1.75
3900	300R	s	3k6	1.08
3900	1k2	s	2k7	1.44
3900	1k5	s	2k4	1.63
3909	4k3	p	43k	1.1
3910	910R	s	3k	1.3
3914	5k6	p	13k	1.43
3917	7k5	p	8k2	1.91
3920	620R	s	3k3	1.19
3930	330R	s	3k6	1.09
3930	4k7	p	24k	1.2
3939	39R	s	3k9	1.01
3940	4k3	p	47k	1.09
3943	43R	s	3k9	1.01
3947	47R	s	3k9	1.01
3951	51R	s	3k9	1.01
3956	56R	s	3k9	1.01
3960	360R	s	3k6	1.1
3962	62R	s	3k9	1.02
3965	6k2	p	11k	1.56
3966	4k3	p	51k	1.08
3968	68R	s	3k9	1.02
3974	5k1	p	18k	1.28
3975	75R	s	3k9	1.02
3980	680R	s	3k3	1.21
3982	82R	s	3k9	1.02
3990	390R	s	3k6	1.11
3991	91R	s	3k9	1.02
3993	4k3	p	56k	1.08
4000	100R	s	3k9	1.03
4000	1k	s	3k	1.33
4000	1k3	s	2k7	1.48
4000	1k6	s	2k4	1.67
4000	1k8	s	2k2	1.82
4000	2k	s	2k	2
4003	4k7	p	27k	1.17

R_E	R_1	s/p	R_2	P_{REL}
4010	110R	s	3k9	1.03
4020	120R	s	3k9	1.03
4021	4k3	p	62k	1.07
4030	130R	s	3k9	1.03
4030	430R	s	3k6	1.12
4044	4k3	p	68k	1.06
4048	6k8	p	10k	1.68
4050	150R	s	3k9	1.04
4050	750R	s	3k3	1.23
4060	160R	s	3k9	1.04
4063	4k7	p	30k	1.16
4064	5k1	p	20k	1.26
4067	4k3	p	75k	1.06
4070	470R	s	3k6	1.13
4078	5k6	p	15k	1.37
4080	180R	s	3k9	1.05
4086	4k3	p	82k	1.05
4088	6k2	p	12k	1.52
4100	200R	s	3k9	1.05
4100	1k1	s	3k	1.37
4100	8k2	p	8k2	2
4106	4k3	p	91k	1.05
4110	510R	s	3k6	1.14
4111	7k5	p	9k1	1.82
4114	4k7	p	33k	1.14
4120	220R	s	3k9	1.06
4120	820R	s	3k3	1.25
4123	4k3	p	100k	1.04
4138	4k3	p	110k	1.04
4140	240R	s	3k9	1.06
4140	5k1	p	22k	1.23
4148	5k6	p	16k	1.35
4151	4k3	p	120k	1.04
4157	4k7	p	36k	1.13
4160	560R	s	3k6	1.16
4162	4k3	p	130k	1.03
4170	270R	s	3k9	1.07
4180	4k3	p	150k	1.03
4187	4k3	p	160k	1.03
4195	4k7	p	39k	1.12
4198	6k2	p	13k	1.48
4200	300R	s	3k9	1.08
4200	1k2	s	3k	1.4
4200	1k5	s	2k7	1.56
4200	1k8	s	2k4	1.75
4200	2k	s	2k2	1.91
4200	4k3	p	180k	1.02
4202	6k8	p	11k	1.62
4206	5k1	p	24k	1.21
4209	4k3	p	200k	1.02
4210	910R	s	3k3	1.28
4218	4k3	p	220k	1.02
4220	620R	s	3k6	1.17
4224	4k3	p	240k	1.02
4230	330R	s	3k9	1.08
4233	4k3	p	270k	1.02
4237	4k7	p	43k	1.11
4239	4k3	p	300k	1.01
4245	4k3	p	330k	1.01
4249	4k3	p	360k	1.01
4253	4k3	p	390k	1.01
4257	4k3	p	430k	1.01
4260	360R	s	3k9	1.09
4271	5k6	p	18k	1.31
4273	4k7	p	47k	1.1
4280	680R	s	3k6	1.19
4286	7k5	p	10k	1.75
4290	390R	s	3k9	1.1
4290	5k1	p	27k	1.19
4300	1k	s	3k3	1.3
4300	1k3	s	3k	1.43
4300	1k6	s	2k7	1.59
4303	4k7	p	51k	1.09
4313	8k2	p	9k1	1.9

R_E	R_1	s/p	R_2	P_{REL}
4330	430R	s	3k9	1.11
4336	4k7	p	56k	1.08
4340	6k8	p	12k	1.57
4343	43R	s	4k3	1.01
4347	47R	s	4k3	1.01
4350	750R	s	3k6	1.21
4351	51R	s	4k3	1.01
4356	56R	s	4k3	1.01
4359	5k1	p	30k	1.17
4362	62R	s	4k3	1.01
4368	68R	s	4k3	1.02
4369	4k7	p	62k	1.08
4370	470R	s	3k9	1.12
4375	75R	s	4k3	1.02
4375	5k6	p	20k	1.28
4382	82R	s	4k3	1.02
4387	6k2	p	15k	1.41
4391	91R	s	4k3	1.02
4396	4k7	p	68k	1.07
4400	100R	s	4k3	1.02
4400	1k1	s	3k3	1.33
4400	2k	s	2k4	1.83
4400	2k2	s	2k2	2
4410	110R	s	4k3	1.03
4410	510R	s	3k9	1.13
4417	5k1	p	33k	1.15
4420	120R	s	4k3	1.03
4420	820R	s	3k6	1.23
4423	4k7	p	75k	1.06
4430	130R	s	4k3	1.03
4445	4k7	p	82k	1.06
4450	150R	s	4k3	1.03
4459	7k5	p	11k	1.68
4460	160R	s	4k3	1.04
4460	560R	s	3k9	1.14
4464	5k6	p	22k	1.25
4465	6k8	p	13k	1.52
4467	5k1	p	36k	1.14
4468	6k2	p	16k	1.39
4469	4k7	p	91k	1.05
4480	180R	s	4k3	1.04
4489	4k7	p	100k	1.05
4500	200R	s	4k3	1.05
4500	1k2	s	3k3	1.36
4500	1k5	s	3k	1.5
4500	1k8	s	2k7	1.67
4505	8k2	p	10k	1.82
4507	4k7	p	110k	1.04
4510	910R	s	3k6	1.25
4510	5k1	p	39k	1.13
4520	220R	s	4k3	1.05
4520	620R	s	3k9	1.16
4523	4k7	p	120k	1.04
4536	4k7	p	130k	1.04
4540	240R	s	4k3	1.06
4541	5k6	p	24k	1.23
4550	9k1	p	9k1	2
4557	4k7	p	150k	1.03
4559	5k1	p	43k	1.12
4566	4k7	p	160k	1.03
4570	270R	s	4k3	1.06
4580	680R	s	3k9	1.17
4580	4k7	p	180k	1.03
4592	4k7	p	200k	1.02
4600	300R	s	4k3	1.07
4600	1k	s	3k6	1.28
4600	1k3	s	3k3	1.39
4600	1k6	s	3k	1.53
4600	2k2	s	2k4	1.92
4601	5k1	p	47k	1.11
4602	4k7	p	220k	1.02
4610	4k7	p	240k	1.02
4612	6k2	p	18k	1.34
4615	7k5	p	12k	1.63

R_E	R_1	s/p	R_2	P_{REL}
4620	4k7	p	270k	1.02
4628	4k7	p	300k	1.02
4630	330R	s	4k3	1.08
4634	4k7	p	330k	1.01
4636	5k1	p	51k	1.1
4638	5k6	p	27k	1.21
4639	4k7	p	360k	1.01
4644	4k7	p	390k	1.01
4649	4k7	p	430k	1.01
4650	750R	s	3k9	1.19
4653	4k7	p	470k	1.01
4660	360R	s	4k3	1.08
4674	5k1	p	56k	1.09
4679	6k8	p	15k	1.45
4690	390R	s	4k3	1.09
4698	8k2	p	11k	1.75
4700	1k1	s	3k6	1.31
4700	2k	s	2k7	1.74
4712	5k1	p	62k	1.08
4719	5k6	p	30k	1.19
4720	820R	s	3k9	1.21
4730	430R	s	4k3	1.1
4733	6k2	p	20k	1.31
4744	5k1	p	68k	1.08
4747	47R	s	4k7	1.01
4751	51R	s	4k7	1.01
4756	56R	s	4k7	1.01
4756	7k5	p	13k	1.58
4762	62R	s	4k7	1.01
4764	9k1	p	10k	1.91
4768	68R	s	4k7	1.01
4770	470R	s	4k3	1.11
4772	6k8	p	16k	1.43
4775	75R	s	4k7	1.02
4775	5k1	p	75k	1.07
4782	82R	s	4k7	1.02
4788	5k6	p	33k	1.17
4791	91R	s	4k7	1.02
4800	100R	s	4k7	1.02
4800	1k2	s	3k6	1.33
4800	1k5	s	3k3	1.45
4800	1k8	s	3k	1.6
4800	2k4	s	2k4	2
4801	5k1	p	82k	1.06
4810	110R	s	4k7	1.02
4810	510R	s	4k3	1.12
4810	910R	s	3k9	1.23
4820	120R	s	4k7	1.03
4829	5k1	p	91k	1.06
4830	130R	s	4k7	1.03
4837	6k2	p	22k	1.28
4846	5k6	p	36k	1.16
4850	150R	s	4k7	1.03
4853	5k1	p	100k	1.05
4860	160R	s	4k7	1.03
4860	560R	s	4k3	1.13
4871	8k2	p	12k	1.68
4874	5k1	p	110k	1.05
4880	180R	s	4k7	1.04
4892	5k1	p	120k	1.04
4897	5k6	p	39k	1.14
4900	200R	s	4k7	1.04
4900	1k	s	3k9	1.26
4900	1k3	s	3k6	1.36
4900	1k6	s	3k3	1.48
4900	2k2	s	2k7	1.81
4907	5k1	p	130k	1.04
4920	220R	s	4k7	1.05
4920	620R	s	4k3	1.14
4927	6k2	p	24k	1.26
4932	5k1	p	150k	1.03
4935	6k8	p	18k	1.38
4940	240R	s	4k7	1.05
4942	5k1	p	160k	1.03

R_E	R_1	s/p	R_2	P_{REL}
4955	5k6	p	43k	1.13
4959	5k1	p	180k	1.03
4970	270R	s	4k7	1.06
4973	5k1	p	200k	1.03
4980	680R	s	4k3	1.16
4980	9k1	p	11k	1.83
4984	5k1	p	220k	1.02
4994	5k1	p	240k	1.02
5000	300R	s	4k7	1.06
5000	1k1	s	3k9	1.28
5000	2k	s	3k	1.67
5000	7k5	p	15k	1.5
5000	10k	p	10k	2
5004	5k6	p	47k	1.12
5005	5k1	p	270k	1.02
5015	5k1	p	300k	1.02
5022	5k1	p	330k	1.02
5028	8k2	p	13k	1.63
5029	5k1	p	360k	1.01
5030	330R	s	4k7	1.07
5034	5k1	p	390k	1.01
5040	5k1	p	430k	1.01
5042	6k2	p	27k	1.23
5045	5k1	p	470k	1.01
5046	5k6	p	51k	1.11
5050	750R	s	4k3	1.17
5050	5k1	p	510k	1.01
5060	360R	s	4k7	1.08
5075	6k8	p	20k	1.34
5090	390R	s	4k7	1.08
5091	5k6	p	56k	1.1
5100	1k2	s	3k9	1.31
5100	1k5	s	3k6	1.42
5100	1k8	s	3k3	1.55
5100	2k4	s	2k7	1.89
5106	7k5	p	16k	1.47
5120	820R	s	4k3	1.19

R_E	R_1	s/p	R_2	P_{REL}
5130	430R	s	4k7	1.09
5136	5k6	p	62k	1.09
5138	6k2	p	30k	1.21
5151	51R	s	5k1	1.01
5156	56R	s	5k1	1.01
5162	62R	s	5k1	1.01
5168	68R	s	5k1	1.01
5170	470R	s	4k7	1.1
5174	5k6	p	68k	1.08
5175	75R	s	5k1	1.01
5175	9k1	p	12k	1.76
5182	82R	s	5k1	1.02
5191	91R	s	5k1	1.02
5194	6k8	p	22k	1.31
5200	100R	s	5k1	1.02
5200	1k3	s	3k9	1.33
5200	1k6	s	3k6	1.44
5200	2k2	s	3k	1.73
5210	110R	s	5k1	1.02
5210	510R	s	4k7	1.11
5210	910R	s	4k3	1.21
5211	5k6	p	75k	1.07
5219	6k2	p	33k	1.19
5220	120R	s	5k1	1.02
5230	130R	s	5k1	1.03
5238	10k	p	11k	1.91
5242	5k6	p	82k	1.07
5250	150R	s	5k1	1.03
5260	160R	s	5k1	1.03
5260	560R	s	4k7	1.12
5275	5k6	p	91k	1.06
5280	180R	s	5k1	1.04
5289	6k2	p	36k	1.17
5294	7k5	p	18k	1.42
5299	6k8	p	24k	1.28
5300	200R	s	5k1	1.04
5300	1k	s	4k3	1.23

R_E	R_1	s/p	R_2	P_{REL}
5300	2k	s	3k3	1.61
5302	8k2	p	15k	1.55
5303	5k6	p	100k	1.06
5320	220R	s	5k1	1.04
5320	620R	s	4k7	1.13
5329	5k6	p	110k	1.05
5340	240R	s	5k1	1.05
5350	5k6	p	120k	1.05
5350	6k2	p	39k	1.16
5353	9k1	p	13k	1.7
5369	5k6	p	130k	1.04
5370	270R	s	5k1	1.05
5380	680R	s	4k7	1.14
5398	5k6	p	150k	1.04
5400	300R	s	5k1	1.06
5400	1k1	s	4k3	1.26
5400	1k5	s	3k9	1.38
5400	1k8	s	3k6	1.5
5400	2k4	s	3k	1.8
5400	2k7	s	2k7	2
5411	5k6	p	160k	1.03
5419	6k2	p	43k	1.14
5421	8k2	p	16k	1.51
5430	330R	s	5k1	1.06
5431	5k6	p	180k	1.03
5432	6k8	p	27k	1.25
5447	5k6	p	200k	1.03
5450	750R	s	4k7	1.16
5455	7k5	p	20k	1.38
5455	10k	p	12k	1.83
5460	360R	s	5k1	1.07
5461	5k6	p	220k	1.03
5472	5k6	p	240k	1.02
5477	6k2	p	47k	1.13
5486	5k6	p	270k	1.02
5490	390R	s	5k1	1.08
5497	5k6	p	300k	1.02
5500	1k2	s	4k3	1.28
5500	1k6	s	3k9	1.41
5500	2k2	s	3k3	1.67
5500	11k	p	11k	2
5507	5k6	p	330k	1.02
5514	5k6	p	360k	1.02
5520	820R	s	4k7	1.17
5521	5k6	p	390k	1.01
5528	5k6	p	430k	1.01
5528	6k2	p	51k	1.12
5530	430R	s	5k1	1.08
5534	5k6	p	470k	1.01
5539	5k6	p	510k	1.01
5543	6k8	p	30k	1.23
5545	5k6	p	560k	1.01
5570	470R	s	5k1	1.09
5582	6k2	p	56k	1.11
5593	7k5	p	22k	1.34
5600	1k3	s	4k3	1.3
5600	2k	s	3k6	1.56
5610	510R	s	5k1	1.1
5610	910R	s	4k7	1.19
5634	8k2	p	18k	1.46
5636	6k2	p	62k	1.1
5638	6k8	p	33k	1.21
5652	10k	p	13k	1.77
5656	56R	s	5k6	1.01
5660	560R	s	5k1	1.11
5662	62R	s	5k6	1.01
5664	9k1	p	15k	1.61
5668	68R	s	5k6	1.01
5675	75R	s	5k6	1.01
5682	82R	s	5k6	1.01
5682	6k2	p	68k	1.09
5691	91R	s	5k6	1.02
5700	100R	s	5k6	1.02
5700	1k	s	4k7	1.21

R_E	R_1	s/p	R_2	P_{REL}
5700	1k8	s	3k9	1.46
5700	2k4	s	3k3	1.73
5700	2k7	s	3k	1.9
5710	110R	s	5k6	1.02
5714	7k5	p	24k	1.31
5720	120R	s	5k6	1.02
5720	620R	s	5k1	1.12
5720	6k8	p	36k	1.19
5727	6k2	p	75k	1.08
5730	130R	s	5k6	1.02
5739	11k	p	12k	1.92
5750	150R	s	5k6	1.03
5760	160R	s	5k6	1.03
5764	6k2	p	82k	1.08
5780	180R	s	5k6	1.03
5780	680R	s	5k1	1.13
5790	6k8	p	39k	1.17
5800	200R	s	5k6	1.04
5800	1k1	s	4k7	1.23
5800	1k5	s	4k3	1.35
5800	2k2	s	3k6	1.61
5801	9k1	p	16k	1.57
5805	6k2	p	91k	1.07
5816	8k2	p	20k	1.41
5820	220R	s	5k6	1.04
5838	6k2	p	100k	1.06
5840	240R	s	5k6	1.04
5850	750R	s	5k1	1.15
5869	6k2	p	110k	1.06
5870	270R	s	5k6	1.05
5870	7k5	p	27k	1.28
5871	6k8	p	43k	1.16
5895	6k2	p	120k	1.05
5900	300R	s	5k6	1.05
5900	1k2	s	4k7	1.26
5900	1k6	s	4k3	1.37
5900	2k	s	3k9	1.51
5918	6k2	p	130k	1.05
5920	820R	s	5k1	1.16
5930	330R	s	5k6	1.06
5941	6k8	p	47k	1.14
5954	6k2	p	150k	1.04
5958	11k	p	13k	1.85
5960	360R	s	5k6	1.06
5969	6k2	p	160k	1.04
5974	8k2	p	22k	1.37
5990	390R	s	5k6	1.07
5994	6k2	p	180k	1.03
6000	1k3	s	4k7	1.28
6000	2k4	s	3k6	1.67
6000	2k7	s	3k3	1.82
6000	3k	s	3k	2
6000	6k8	p	51k	1.13
6000	7k5	p	30k	1.25
6000	10k	p	15k	1.67
6000	12k	p	12k	2
6010	910R	s	5k1	1.18
6014	6k2	p	200k	1.03
6030	430R	s	5k6	1.08
6030	6k2	p	220k	1.03
6044	6k2	p	240k	1.03
6044	9k1	p	18k	1.51
6061	6k2	p	270k	1.02
6064	6k8	p	56k	1.12
6070	470R	s	5k6	1.08
6074	6k2	p	300k	1.02
6086	6k2	p	330k	1.02
6095	6k2	p	360k	1.02
6100	1k	s	5k1	1.2
6100	1k8	s	4k3	1.42
6100	2k2	s	3k9	1.56
6103	6k2	p	390k	1.02
6110	510R	s	5k6	1.09
6111	7k5	p	33k	1.23

R_E	R_1	s/p	R_2	P_{REL}
6112	6k2	p	430k	1.01
6112	8k2	p	24k	1.34
6119	6k2	p	470k	1.01
6126	6k2	p	510k	1.01
6128	6k8	p	62k	1.11
6132	6k2	p	560k	1.01
6139	6k2	p	620k	1.01
6154	10k	p	16k	1.63
6160	560R	s	5k6	1.1
6182	6k8	p	68k	1.1
6200	1k1	s	5k1	1.22
6200	1k5	s	4k7	1.32
6207	7k5	p	36k	1.21
6220	620R	s	5k6	1.11
6235	6k8	p	75k	1.09
6240	12k	p	13k	1.92
6254	9k1	p	20k	1.46
6262	62R	s	6k2	1.01
6268	68R	s	6k2	1.01
6275	75R	s	6k2	1.01
6279	6k8	p	82k	1.08
6280	680R	s	5k6	1.12
6282	82R	s	6k2	1.01
6290	7k5	p	39k	1.19
6290	8k2	p	27k	1.3
6291	91R	s	6k2	1.01
6300	100R	s	6k2	1.02
6300	1k2	s	5k1	1.24
6300	1k6	s	4k7	1.34
6300	2k	s	4k3	1.47
6300	2k4	s	3k9	1.62
6300	2k7	s	3k6	1.75
6300	3k	s	3k3	1.91
6310	110R	s	6k2	1.02
6320	120R	s	6k2	1.02
6327	6k8	p	91k	1.07
6330	130R	s	6k2	1.02
6346	11k	p	15k	1.73
6350	150R	s	6k2	1.02
6350	750R	s	5k6	1.13
6360	160R	s	6k2	1.03
6367	6k8	p	100k	1.07
6380	180R	s	6k2	1.03
6386	7k5	p	43k	1.17
6400	200R	s	6k2	1.03
6400	1k3	s	5k1	1.25
6404	6k8	p	110k	1.06
6420	220R	s	6k2	1.04
6420	820R	s	5k6	1.15
6429	10k	p	18k	1.56
6435	6k8	p	120k	1.06
6437	9k1	p	22k	1.41
6440	240R	s	6k2	1.04
6440	8k2	p	30k	1.27
6462	6k8	p	130k	1.05
6468	7k5	p	47k	1.16
6470	270R	s	6k2	1.04
6500	300R	s	6k2	1.05
6500	1k8	s	4k7	1.38
6500	2k2	s	4k3	1.51
6500	13k	p	13k	2
6505	6k8	p	150k	1.05
6510	910R	s	5k6	1.16
6519	11k	p	16k	1.69
6523	6k8	p	160k	1.04
6530	330R	s	6k2	1.05
6538	7k5	p	51k	1.15
6552	6k8	p	180k	1.04
6560	360R	s	6k2	1.06
6568	8k2	p	33k	1.25
6576	6k8	p	200k	1.03
6590	390R	s	6k2	1.06
6596	6k8	p	220k	1.03
6598	9k1	p	24k	1.38

R_E	R_1	s/p	R_2	P_{REL}
6600	1k	s	5k6	1.18
6600	1k5	s	5k1	1.29
6600	2k7	s	3k9	1.69
6600	3k	s	3k6	1.83
6600	3k3	s	3k3	2
6613	6k8	p	240k	1.03
6614	7k5	p	56k	1.13
6630	430R	s	6k2	1.07
6633	6k8	p	270k	1.03
6649	6k8	p	300k	1.02
6663	6k8	p	330k	1.02
6667	10k	p	20k	1.5
6667	12k	p	15k	1.8
6670	470R	s	6k2	1.08
6674	6k8	p	360k	1.02
6679	8k2	p	36k	1.23
6683	6k8	p	390k	1.02
6691	7k5	p	62k	1.12
6694	6k8	p	430k	1.02
6700	1k1	s	5k6	1.2
6700	1k6	s	5k1	1.31
6700	2k	s	4k7	1.43
6700	2k4	s	4k3	1.56
6703	6k8	p	470k	1.01
6710	510R	s	6k2	1.08
6711	6k8	p	510k	1.01
6718	6k8	p	560k	1.01
6726	6k8	p	620k	1.01
6733	6k8	p	680k	1.01
6755	7k5	p	68k	1.11
6760	560R	s	6k2	1.09
6775	8k2	p	39k	1.21
6800	1k2	s	5k6	1.21
6806	9k1	p	27k	1.34
6818	7k5	p	75k	1.1
6820	620R	s	6k2	1.1
6828	11k	p	18k	1.61

R_E	R_1	s/p	R_2	P_{REL}
6857	12k	p	16k	1.75
6868	68R	s	6k8	1.01
6872	7k5	p	82k	1.09
6875	75R	s	6k8	1.01
6875	10k	p	22k	1.45
6880	680R	s	6k2	1.11
6882	82R	s	6k8	1.01
6887	8k2	p	43k	1.19
6891	91R	s	6k8	1.01
6900	100R	s	6k8	1.01
6900	1k3	s	5k6	1.23
6900	1k8	s	5k1	1.35
6900	2k2	s	4k7	1.47
6900	3k	s	3k9	1.77
6900	3k3	s	3k6	1.92
6910	110R	s	6k8	1.02
6920	120R	s	6k8	1.02
6929	7k5	p	91k	1.08
6930	130R	s	6k8	1.02
6950	150R	s	6k8	1.02
6950	750R	s	6k2	1.12
6960	160R	s	6k8	1.02
6964	13k	p	15k	1.87
6977	7k5	p	100k	1.08
6980	180R	s	6k8	1.03
6982	8k2	p	47k	1.17
6982	9k1	p	30k	1.3
7000	200R	s	6k8	1.03
7000	2k7	s	4k3	1.63
7020	220R	s	6k8	1.03
7020	820R	s	6k2	1.13
7021	7k5	p	110k	1.07
7040	240R	s	6k8	1.04
7059	7k5	p	120k	1.06
7059	10k	p	24k	1.42
7064	8k2	p	51k	1.16
7070	270R	s	6k8	1.04

R_E	R_1	s/p	R_2	P_{REL}
7091	7k5	p	130k	1.06
7097	11k	p	20k	1.55
7100	300R	s	6k8	1.04
7100	1k5	s	5k6	1.27
7100	2k	s	5k1	1.39
7100	2k4	s	4k7	1.51
7110	910R	s	6k2	1.15
7130	330R	s	6k8	1.05
7133	9k1	p	33k	1.28
7143	7k5	p	150k	1.05
7153	8k2	p	56k	1.15
7160	360R	s	6k8	1.05
7164	7k5	p	160k	1.05
7172	13k	p	16k	1.81
7190	390R	s	6k8	1.06
7200	1k	s	6k2	1.16
7200	1k6	s	5k6	1.29
7200	3k3	s	3k9	1.85
7200	3k6	s	3k6	2
7200	7k5	p	180k	1.04
7200	12k	p	18k	1.67
7229	7k5	p	200k	1.04
7230	430R	s	6k8	1.06
7242	8k2	p	62k	1.13
7253	7k5	p	220k	1.03
7264	9k1	p	36k	1.25
7270	470R	s	6k8	1.07
7273	7k5	p	240k	1.03
7297	7k5	p	270k	1.03
7297	10k	p	27k	1.37
7300	1k1	s	6k2	1.18
7300	2k2	s	5k1	1.43
7300	3k	s	4k3	1.7
7310	510R	s	6k8	1.08
7317	7k5	p	300k	1.03
7318	8k2	p	68k	1.12
7333	7k5	p	330k	1.02
7333	11k	p	22k	1.5
7347	7k5	p	360k	1.02
7358	7k5	p	390k	1.02
7360	560R	s	6k8	1.08
7371	7k5	p	430k	1.02
7378	9k1	p	39k	1.23
7382	7k5	p	470k	1.02
7391	7k5	p	510k	1.01
7392	8k2	p	75k	1.11
7400	1k2	s	6k2	1.19
7400	1k8	s	5k6	1.32
7400	2k7	s	4k7	1.57
7401	7k5	p	560k	1.01
7410	7k5	p	620k	1.01
7418	7k5	p	680k	1.01
7420	620R	s	6k8	1.09
7426	7k5	p	750k	1.01
7455	8k2	p	82k	1.1
7480	680R	s	6k8	1.1
7500	1k3	s	6k2	1.21
7500	2k4	s	5k1	1.47
7500	3k6	s	3k9	1.92
7500	10k	p	30k	1.33
7500	12k	p	20k	1.6
7500	15k	p	15k	2
7511	9k1	p	43k	1.21
7522	8k2	p	91k	1.09
7543	11k	p	24k	1.46
7548	13k	p	18k	1.72
7550	750R	s	6k8	1.11
7575	75R	s	7k5	1.01
7579	8k2	p	100k	1.08
7582	82R	s	7k5	1.01
7591	91R	s	7k5	1.01
7600	100R	s	7k5	1.01
7600	2k	s	5k6	1.36
7600	3k3	s	4k3	1.77

R_E	R_1	s/p	R_2	P_{REL}
7610	110R	s	7k5	1.01
7620	120R	s	7k5	1.02
7620	820R	s	6k8	1.12
7624	9k1	p	47k	1.19
7630	130R	s	7k5	1.02
7631	8k2	p	110k	1.07
7650	150R	s	7k5	1.02
7660	160R	s	7k5	1.02
7674	10k	p	33k	1.3
7676	8k2	p	120k	1.07
7680	180R	s	7k5	1.02
7700	200R	s	7k5	1.03
7700	1k5	s	6k2	1.24
7700	3k	s	4k7	1.64
7710	910R	s	6k8	1.13
7713	8k2	p	130k	1.06
7720	220R	s	7k5	1.03
7722	9k1	p	51k	1.18
7740	240R	s	7k5	1.03
7742	15k	p	16k	1.94
7765	12k	p	22k	1.55
7770	270R	s	7k5	1.04
7775	8k2	p	150k	1.05
7800	300R	s	7k5	1.04
7800	1k	s	6k8	1.15
7800	1k6	s	6k2	1.26
7800	2k2	s	5k6	1.39
7800	2k7	s	5k1	1.53
7800	3k9	s	3k9	2
7800	8k2	p	160k	1.05
7816	11k	p	27k	1.41
7826	10k	p	36k	1.28
7828	9k1	p	56k	1.16
7830	330R	s	7k5	1.04
7843	8k2	p	180k	1.05
7860	360R	s	7k5	1.05
7877	8k2	p	200k	1.04

R_E	R_1	s/p	R_2	P_{REL}
7879	13k	p	20k	1.65
7890	390R	s	7k5	1.05
7900	1k1	s	6k8	1.16
7900	3k6	s	4k3	1.84
7905	8k2	p	220k	1.04
7929	8k2	p	240k	1.03
7930	430R	s	7k5	1.06
7935	9k1	p	62k	1.15
7958	8k2	p	270k	1.03
7959	10k	p	39k	1.26
7970	470R	s	7k5	1.06
7982	8k2	p	300k	1.03
8000	1k2	s	6k8	1.18
8000	1k8	s	6k2	1.29
8000	2k4	s	5k6	1.43
8000	3k3	s	4k7	1.7
8000	12k	p	24k	1.5
8000	16k	p	16k	2
8001	8k2	p	330k	1.02
8010	510R	s	7k5	1.07
8017	8k2	p	360k	1.02
8026	9k1	p	68k	1.13
8031	8k2	p	390k	1.02
8047	8k2	p	430k	1.02
8049	11k	p	30k	1.37
8059	8k2	p	470k	1.02
8060	560R	s	7k5	1.07
8070	8k2	p	510k	1.02
8082	8k2	p	560k	1.01
8093	8k2	p	620k	1.01
8100	1k3	s	6k8	1.19
8100	3k	s	5k1	1.59
8102	8k2	p	680k	1.01
8111	8k2	p	750k	1.01
8113	10k	p	43k	1.23
8115	9k1	p	75k	1.12
8119	8k2	p	820k	1.01

R_E	R_1	s/p	R_2	P_{REL}
8120	620R	s	7k5	1.08
8171	13k	p	22k	1.59
8180	680R	s	7k5	1.09
8182	15k	p	18k	1.83
8191	9k1	p	82k	1.11
8200	2k	s	6k2	1.32
8200	3k9	s	4k3	1.91
8246	10k	p	47k	1.21
8250	750R	s	7k5	1.1
8250	11k	p	33k	1.33
8273	9k1	p	91k	1.1
8282	82R	s	8k2	1.01
8291	91R	s	8k2	1.01
8300	100R	s	8k2	1.01
8300	1k5	s	6k8	1.22
8300	2k7	s	5k6	1.48
8300	3k6	s	4k7	1.77
8308	12k	p	27k	1.44
8310	110R	s	8k2	1.01
8320	120R	s	8k2	1.01
8320	820R	s	7k5	1.11
8330	130R	s	8k2	1.02
8341	9k1	p	100k	1.09
8350	150R	s	8k2	1.02
8360	160R	s	8k2	1.02
8361	10k	p	51k	1.2
8380	180R	s	8k2	1.02
8400	200R	s	8k2	1.02
8400	1k6	s	6k8	1.24
8400	2k2	s	6k2	1.35
8400	3k3	s	5k1	1.65
8405	9k1	p	110k	1.08
8410	910R	s	7k5	1.12
8420	220R	s	8k2	1.03
8426	11k	p	36k	1.31
8432	13k	p	24k	1.54
8440	240R	s	8k2	1.03

R_E	R_1	s/p	R_2	P_{REL}
8459	9k1	p	120k	1.08
8470	270R	s	8k2	1.03
8471	16k	p	18k	1.89
8485	10k	p	56k	1.18
8500	300R	s	8k2	1.04
8500	1k	s	7k5	1.13
8505	9k1	p	130k	1.07
8530	330R	s	8k2	1.04
8560	360R	s	8k2	1.04
8571	12k	p	30k	1.4
8571	15k	p	20k	1.75
8580	9k1	p	150k	1.06
8580	11k	p	39k	1.28
8590	390R	s	8k2	1.05
8600	1k1	s	7k5	1.15
8600	1k8	s	6k8	1.26
8600	2k4	s	6k2	1.39
8600	3k	s	5k6	1.54
8600	3k9	s	4k7	1.83
8600	4k3	s	4k3	2
8610	9k1	p	160k	1.06
8611	10k	p	62k	1.16
8630	430R	s	8k2	1.05
8662	9k1	p	180k	1.05
8670	470R	s	8k2	1.06
8700	1k2	s	7k5	1.16
8700	3k6	s	5k1	1.71
8704	9k1	p	200k	1.05
8710	510R	s	8k2	1.06
8718	10k	p	68k	1.15
8739	9k1	p	220k	1.04
8759	11k	p	43k	1.26
8760	560R	s	8k2	1.07
8768	9k1	p	240k	1.04
8775	13k	p	27k	1.48
8800	1k3	s	7k5	1.17
8800	2k	s	6k8	1.29

R_E	R_1	s/p	R_2	P_{REL}
8800	12k	p	33k	1.36
8803	9k1	p	270k	1.03
8820	620R	s	8k2	1.08
8824	10k	p	75k	1.13
8832	9k1	p	300k	1.03
8856	9k1	p	330k	1.03
8876	9k1	p	360k	1.03
8880	680R	s	8k2	1.08
8889	16k	p	20k	1.8
8893	9k1	p	390k	1.02
8900	2k7	s	6k2	1.44
8900	3k3	s	5k6	1.59
8911	9k1	p	430k	1.02
8913	10k	p	82k	1.12
8914	11k	p	47k	1.23
8919	15k	p	22k	1.68
8927	9k1	p	470k	1.02
8940	9k1	p	510k	1.02
8950	750R	s	8k2	1.09
8954	9k1	p	560k	1.02
8968	9k1	p	620k	1.01
8980	9k1	p	680k	1.01
8991	9k1	p	750k	1.01
9000	1k5	s	7k5	1.2
9000	2k2	s	6k8	1.32
9000	3k9	s	5k1	1.76
9000	4k3	s	4k7	1.91
9000	9k1	p	820k	1.01
9000	12k	p	36k	1.33
9000	18k	p	18k	2
9010	9k1	p	910k	1.01
9010	10k	p	91k	1.11
9020	820R	s	8k2	1.1
9048	11k	p	51k	1.22
9070	13k	p	30k	1.43
9091	10k	p	100k	1.1
9100	1k6	s	7k5	1.21

R_E	R_1	s/p	R_2	P_{REL}
9110	910R	s	8k2	1.11
9167	10k	p	110k	1.09
9176	12k	p	39k	1.31
9191	91R	s	9k1	1.01
9194	11k	p	56k	1.2
9200	100R	s	9k1	1.01
9200	1k	s	8k2	1.12
9200	2k4	s	6k8	1.35
9200	3k	s	6k2	1.48
9200	3k6	s	5k6	1.64
9210	110R	s	9k1	1.01
9220	120R	s	9k1	1.01
9230	130R	s	9k1	1.01
9231	10k	p	120k	1.08
9231	15k	p	24k	1.63
9250	150R	s	9k1	1.02
9260	160R	s	9k1	1.02
9263	16k	p	22k	1.73
9280	180R	s	9k1	1.02
9286	10k	p	130k	1.08
9300	200R	s	9k1	1.02
9300	1k1	s	8k2	1.13
9300	1k8	s	7k5	1.24
9320	220R	s	9k1	1.02
9326	13k	p	33k	1.39
9340	240R	s	9k1	1.03
9342	11k	p	62k	1.18
9370	270R	s	9k1	1.03
9375	10k	p	150k	1.07
9382	12k	p	43k	1.28
9400	300R	s	9k1	1.03
9400	1k2	s	8k2	1.15
9400	4k3	s	5k1	1.84
9400	4k7	s	4k7	2
9412	10k	p	160k	1.06
9430	330R	s	9k1	1.04
9460	360R	s	9k1	1.04

R_E	R_1	s/p	R_2	P_{REL}
9468	11k	p	68k	1.16
9474	10k	p	180k	1.06
9474	18k	p	20k	1.9
9490	390R	s	9k1	1.04
9500	1k3	s	8k2	1.16
9500	2k	s	7k5	1.27
9500	2k7	s	6k8	1.4
9500	3k3	s	6k2	1.53
9500	3k9	s	5k6	1.7
9524	10k	p	200k	1.05
9530	430R	s	9k1	1.05
9551	13k	p	36k	1.36
9559	12k	p	47k	1.26
9565	10k	p	220k	1.05
9570	470R	s	9k1	1.05
9593	11k	p	75k	1.15
9600	10k	p	240k	1.04
9600	16k	p	24k	1.67
9610	510R	s	9k1	1.06
9643	10k	p	270k	1.04
9643	15k	p	27k	1.56
9660	560R	s	9k1	1.06
9677	10k	p	300k	1.03
9699	11k	p	82k	1.13
9700	1k5	s	8k2	1.18
9700	2k2	s	7k5	1.29
9706	10k	p	330k	1.03
9714	12k	p	51k	1.24
9720	620R	s	9k1	1.07
9730	10k	p	360k	1.03
9750	10k	p	390k	1.03
9750	13k	p	39k	1.33
9773	10k	p	430k	1.02
9780	680R	s	9k1	1.07
9792	10k	p	470k	1.02
9800	1k6	s	8k2	1.2
9800	3k	s	6k8	1.44
9800	3k6	s	6k2	1.58
9800	4k7	s	5k1	1.92
9808	10k	p	510k	1.02
9814	11k	p	91k	1.12
9825	10k	p	560k	1.02
9841	10k	p	620k	1.02
9850	750R	s	9k1	1.08
9855	10k	p	680k	1.01
9868	10k	p	750k	1.01
9880	10k	p	820k	1.01
9882	12k	p	56k	1.21
9891	10k	p	910k	1.01
9900	2k4	s	7k5	1.32
9900	4k3	s	5k6	1.77
9900	18k	p	22k	1.82
9901	10k	p	1M	1.01
9910	11k	p	100k	1.11
9920	820R	s	9k1	1.09
9982	13k	p	43k	1.3

7 Maximum powers for resistors

Tables 7.1–7.6 show the voltage across or current through standard value resistors which are equivalent to a selection of common rated maximum powers.

As discussed in Chapter 3, it is wise to avoid running components continuously at full rated power. A good rule of thumb is not to exceed half of rated power; that is roughly 0.7 times the value of voltage or current presented here.

Table 7.1 *Maximum power for resistors up to 0.25 W*

R	0.063 W	0.1 W	0.125 W	0.155 W	0.25 W	
0R10	79.4m	100m	112m	124m	158m	V
	794m	1.00	1.12	1.24	1.58	A
0R12	86.9m	110m	122m	136m	173m	V
	725m	913m	1.02	1.14	1.44	A
0R15	97.2m	122m	137m	152m	194m	V
	648m	816m	913m	1.02	1.29	A
0R18	106m	134m	150m	167m	212m	V
	592m	745m	833m	928m	1.18	A
0R22	118m	148m	166m	185m	235m	V
	535m	674m	754m	839m	1.07	A
0R27	130m	164m	184m	205m	260m	V
	483m	609m	680m	758m	962m	A
0R33	144m	182m	203m	226m	287m	V
	437m	550m	615m	685m	870m	A
0R39	157m	197m	221m	246m	312m	V
	402m	506m	566m	630m	801m	A
0R47	172m	217m	242m	270m	343m	V
	366m	461m	516m	574m	729m	A
0R56	188m	237m	265m	295m	374m	V
	335m	423m	472m	526m	668m	A
0R68	207m	261m	292m	325m	412m	V
	304m	383m	429m	477m	606m	A
0R82	227m	286m	320m	357m	453m	V
	277m	349m	390m	435m	552m	A
1R	251m	316m	354m	394m	500m	V
	251m	316m	354m	394m	500m	A
1R1	263m	332m	371m	413m	524m	V
	239m	302m	337m	375m	477m	A
1R2	275m	346m	387m	431m	548m	V
	229m	289m	323m	359m	456m	A
1R3	286m	361m	403m	449m	570m	V
	220m	277m	310m	345m	439m	A
1R5	307m	387m	433m	482m	612m	V
	205m	258m	289m	321m	408m	A
1R6	317m	400m	447m	498m	632m	V
	198m	250m	280m	311m	395m	A

R	0.063 W	0.1 W	0.125 W	0.155 W	0.25 W	
1R8	337m	424m	474m	528m	671m	V
	187m	236m	264m	293m	373m	A
2R	355m	447m	500m	557m	707m	V
	177m	224m	250m	278m	354m	A
2R2	372m	469m	524m	584m	742m	V
	169m	213m	238m	265m	337m	A
2R4	389m	490m	548m	610m	775m	V
	162m	204m	228m	254m	323m	A
2R7	412m	520m	581m	647m	822m	V
	153m	192m	215m	240m	304m	A
3R	435m	548m	612m	682m	866m	V
	145m	183m	204m	227m	289m	A
3R3	456m	574m	642m	715m	908m	V
	138m	174m	195m	217m	275m	A
3R6	476m	600m	671m	747m	949m	V
	132m	167m	186m	207m	264m	A
3R9	496m	624m	698m	777m	987m	V
	127m	160m	179m	199m	253m	A
4R3	520m	656m	733m	816m	1.04	V
	121m	152m	170m	190m	241m	A
4R7	544m	686m	766m	854m	1.08	V
	116m	146m	163m	182m	231m	A
5R1	567m	714m	798m	889m	1.13	V
	111m	140m	157m	174m	221m	A
5R6	594m	748m	837m	932m	1.18	V
	106m	134m	149m	166m	211m	A
6R2	625m	787m	880m	980m	1.24	V
	101m	127m	142m	158m	201m	A
6R8	655m	825m	922m	1.03	1.30	V
	96.3m	121m	136m	151m	192m	A
7R5	687m	866m	968m	1.08	1.37	V
	91.7m	115m	129m	144m	183m	A
8R2	719m	906m	1.01	1.13	1.43	V
	87.7m	110m	123m	137m	175m	A
9R1	757m	954m	1.07	1.19	1.51	V
	83.2m	105m	117m	131m	166m	A
10R	794m	1.00	1.12	1.24	1.58	V
	79.4m	100m	112m	124m	158m	A

R	0.063 W	0.1 W	0.125 W	0.155 W	0.25 W	
11R	832m	1.05	1.17	1.31	1.66	V
	75.7m	95.3m	107m	119m	151m	A
12R	869m	1.10	1.22	1.36	1.73	V
	72.5m	91.3m	102m	114m	144m	A
13R	905m	1.14	1.27	1.42	1.80	V
	69.6m	87.7m	98.1m	109m	139m	A
15R	972m	1.22	1.37	1.52	1.94	V
	64.8m	81.6m	91.3m	102m	129m	A
16R	1.00	1.26	1.41	1.57	2.00	V
	62.7m	79.1m	88.4m	98.4m	125m	A
18R	1.06	1.34	1.50	1.67	2.12	V
	59.2m	74.5m	83.3m	92.8m	118m	A
20R	1.12	1.41	1.58	1.76	2.24	V
	56.1m	70.7m	79.1m	88.0m	112m	A
22R	1.18	1.48	1.66	1.85	2.35	V
	53.5m	67.4m	75.4m	83.9m	107m	A
24R	1.23	1.55	1.73	1.93	2.45	V
	51.2m	64.5m	72.2m	80.4m	102m	A
27R	1.30	1.64	1.84	2.05	2.60	V
	48.3m	60.9m	68.0m	75.8m	96.2m	A
30R	1.37	1.73	1.94	2.16	2.74	V
	45.8m	57.7m	64.5m	71.9m	91.3m	A
33R	1.44	1.82	2.03	2.26	2.87	V
	43.7m	55.0m	61.5m	68.5m	87.0m	A
36R	1.51	1.90	2.12	2.36	3.00	V
	41.8m	52.7m	58.9m	65.6m	83.3m	A
39R	1.57	1.97	2.21	2.46	3.12	V
	40.2m	50.6m	56.6m	63.0m	80.1m	A
43R	1.65	2.07	2.32	2.58	3.28	V
	38.3m	48.2m	53.9m	60.0m	76.2m	A
47R	1.72	2.17	2.42	2.70	3.43	V
	36.6m	46.1m	51.6m	57.4m	72.9m	A
51R	1.79	2.26	2.52	2.81	3.57	V
	35.1m	44.3m	49.5m	55.1m	70.0m	A
56R	1.88	2.37	2.65	2.95	3.74	V
	33.5m	42.3m	47.2m	52.6m	66.8m	A
62R	1.98	2.49	2.78	3.10	3.94	V
	31.9m	40.2m	44.9m	50.0m	63.5m	A

R	0.063 W	0.1 W	0.125 W	0.155 W	0.25 W	
68R	2.07	2.61	2.92	3.25	4.12	V
	30.4m	38.3m	42.9m	47.7m	60.6m	A
75R	2.17	2.74	3.06	3.41	4.33	V
	29.0m	36.5m	40.8m	45.5m	57.7m	A
82R	2.27	2.86	3.20	3.57	4.53	V
	27.7m	34.9m	39.0m	43.5m	55.2m	A
91R	2.39	3.02	3.37	3.76	4.77	V
	26.3m	33.1m	37.1m	41.3m	52.4m	A
100R	2.51	3.16	3.54	3.94	5.00	V
	25.1m	31.6m	35.4m	39.4m	50.0m	A
110R	2.63	3.32	3.71	4.13	5.24	V
	23.9m	30.2m	33.7m	37.5m	47.7m	A
120R	2.75	3.46	3.87	4.31	5.48	V
	22.9m	28.9m	32.3m	35.9m	45.6m	A
130R	2.86	3.61	4.03	4.49	5.70	V
	22.0m	27.7m	31.0m	34.5m	43.9m	A
150R	3.07	3.87	4.33	4.82	6.12	V
	20.5m	25.8m	28.9m	32.1m	40.8m	A
160R	3.17	4.00	4.47	4.98	6.32	V
	19.8m	25.0m	28.0m	31.1m	39.5m	A
180R	3.37	4.24	4.74	5.28	6.71	V
	18.7m	23.6m	26.4m	29.3m	37.3m	A
200R	3.55	4.47	5.00	5.57	7.07	V
	17.7m	22.4m	25.0m	27.8m	35.4m	A
220R	3.72	4.69	5.24	5.84	7.42	V
	16.9m	21.3m	23.8m	26.5m	33.7m	A
240R	3.89	4.90	5.48	6.10	7.75	V
	16.2m	20.4m	22.8m	25.4m	32.3m	A
270R	4.12	5.20	5.81	6.47	8.22	V
	15.3m	19.2m	21.5m	24.0m	30.4m	A
300R	4.35	5.48	6.12	6.82	8.66	V
	14.5m	18.3m	20.4m	22.7m	28.9m	A
330R	4.56	5.74	6.42	7.15	9.08	V
	13.8m	17.4m	19.5m	21.7m	27.5m	A
360R	4.76	6.00	6.71	7.47	9.49	V
	13.2m	16.7m	18.6m	20.7m	26.4m	A
390R	4.96	6.24	6.98	7.77	9.87	V
	12.7m	16.0m	17.9m	19.9m	25.3m	A

R	0.063 W	0.1 W	0.125 W	0.155 W	0.25 W	
430R	5.20	6.56	7.33	8.16	10.4	V
	12.1m	15.2m	17.0m	19.0m	24.1m	A
470R	5.44	6.86	7.66	8.54	10.8	V
	11.6m	14.6m	16.3m	18.2m	23.1m	A
510R	5.67	7.14	7.98	8.89	11.3	V
	11.1m	14.0m	15.7m	17.4m	22.1m	A
560R	5.94	7.48	8.37	9.32	11.8	V
	10.6m	13.4m	14.9m	16.6m	21.1m	A
620R	6.25	7.87	8.80	9.80	12.4	V
	10.1m	12.7m	14.2m	15.8m	20.1m	A
680R	6.55	8.25	9.22	10.3	13.0	V
	9.63m	12.1m	13.6m	15.1m	19.2m	A
750R	6.87	8.66	9.68	10.8	13.7	V
	9.17m	11.5m	12.9m	14.4m	18.3m	A
820R	7.19	9.06	10.1	11.3	14.3	V
	8.77m	11.0m	12.3m	13.7m	17.5m	A
910R	7.57	9.54	10.7	11.9	15.1	V
	8.32m	10.5m	11.7m	13.1m	16.6m	A
1k	7.94	10.0	11.2	12.4	15.8	V
	7.94m	10.0m	11.2m	12.4m	15.8m	A
1k1	8.32	10.5	11.7	13.1	16.6	V
	7.57m	9.53m	10.7m	11.9m	15.1m	A
1k2	8.69	11.0	12.2	13.6	17.3	V
	7.25m	9.13m	10.2m	11.4m	14.4m	A
1k3	9.05	11.4	12.7	14.2	18.0	V
	6.96m	8.77m	9.81m	10.9m	13.9m	A
1k5	9.72	12.2	13.7	15.2	19.4	V
	6.48m	8.16m	9.13m	10.2m	12.9m	A
1k6	10.0	12.6	14.1	15.7	20.0	V
	6.27m	7.91m	8.84m	9.84m	12.5m	A
1k8	10.6	13.4	15.0	16.7	21.2	V
	5.92m	7.45m	8.33m	9.28m	11.8m	A
2k	11.2	14.1	15.8	17.6	22.4	V
	5.61m	7.07m	7.91m	8.80m	11.2m	A
2k2	11.8	14.8	16.6	18.5	23.5	V
	5.35m	6.74m	7.54m	8.39m	10.7m	A
2k4	12.3	15.5	17.3	19.3	24.5	V
	5.12m	6.45m	7.22m	8.04m	10.2m	A

R	0.063 W	0.1 W	0.125 W	0.155 W	0.25 W	
2k7	13.0	16.4	18.4	20.5	26.0	V
	4.83m	6.09m	6.80m	7.58m	9.62m	A
3k	13.7	17.3	19.4	21.6	27.4	V
	4.58m	5.77m	6.45m	7.19m	9.13m	A
3k3	14.4	18.2	20.3	22.6	28.7	V
	4.37m	5.50m	6.15m	6.85m	8.70m	A
3k6	15.1	19.0	21.2	23.6	30.0	V
	4.18m	5.27m	5.89m	6.56m	8.33m	A
3k9	15.7	19.7	22.1	24.6	31.2	V
	4.02m	5.06m	5.66m	6.30m	8.01m	A
4k3	16.5	20.7	23.2	25.8	32.8	V
	3.83m	4.82m	5.39m	6.00m	7.62m	A
4k7	17.2	21.7	24.2	27.0	34.3	V
	3.66m	4.61m	5.16m	5.74m	7.29m	A
5k1	17.9	22.6	25.2	28.1	35.7	V
	3.51m	4.43m	4.95m	5.51m	7.00m	A
5k6	18.8	23.7	26.5	29.5	37.4	V
	3.35m	4.23m	4.72m	5.26m	6.68m	A
6k2	19.8	24.9	27.8	31.0	39.4	V
	3.19m	4.02m	4.49m	5.00m	6.35m	A
6k8	20.7	26.1	29.2	32.5	41.2	V
	3.04m	3.83m	4.29m	4.77m	6.06m	A
7k5	21.7	27.4	30.6	34.1	43.3	V
	2.90m	3.65m	4.08m	4.55m	5.77m	A
8k2	22.7	28.6	32.0	35.7	45.3	V
	2.77m	3.49m	3.90m	4.35m	5.52m	A
9k1	23.9	30.2	33.7	37.6	47.7	V
	2.63m	3.31m	3.71m	4.13m	5.24m	A
10k	25.1	31.6	35.4	39.4	50.0	V
	2.51m	3.16m	3.54m	3.94m	5.00m	A
11k	26.3	33.2	37.1	41.3	52.4	V
	2.39m	3.02m	3.37m	3.75m	4.77m	A
12k	27.5	34.6	38.7	43.1	54.8	V
	2.29m	2.89m	3.23m	3.59m	4.56m	A
13k	28.6	36.1	40.3	44.9	57.0	V
	2.20m	2.77m	3.10m	3.45m	4.39m	A
15k	30.7	38.7	43.3	48.2	61.2	V
	2.05m	2.58m	2.89m	3.21m	4.08m	A

R	0.063 W	0.1 W	0.125 W	0.155 W	0.25 W	
16k	31.7	40.0	44.7	49.8	63.2	V
	1.98m	2.50m	2.80m	3.11m	3.95m	A
18k	33.7	42.4	47.4	52.8	67.1	V
	1.87m	2.36m	2.64m	2.93m	3.73m	A
20k	35.5	44.7	50.0	55.7	70.7	V
	1.77m	2.24m	2.50m	2.78m	3.54m	A
22k	37.2	46.9	52.4	58.4	74.2	V
	1.69m	2.13m	2.38m	2.65m	3.37m	A
24k	38.9	49.0	54.8	61.0	77.5	V
	1.62m	2.04m	2.28m	2.54m	3.23m	A
27k	41.2	52.0	58.1	64.7	82.2	V
	1.53m	1.92m	2.15m	2.40m	3.04m	A
30k	43.5	54.8	61.2	68.2	86.6	V
	1.45m	1.83m	2.04m	2.27m	2.89m	A
33k	45.6	57.4	64.2	71.5	90.8	V
	1.38m	1.74m	1.95m	2.17m	2.75m	A
36k	47.6	60.0	67.1	74.7	94.9	V
	1.32m	1.67m	1.86m	2.07m	2.64m	A
39k	49.6	62.4	69.8	77.7	98.7	V
	1.27m	1.60m	1.79m	1.99m	2.53m	A
43k	52.0	65.6	73.3	81.6	104	V
	1.21m	1.52m	1.70m	1.90m	2.41m	A
47k	54.4	68.6	76.6	85.4	108	V
	1.16m	1.46m	1.63m	1.82m	2.31m	A
51k	56.7	71.4	79.8	88.9	113	V
	1.11m	1.40m	1.57m	1.74m	2.21m	A
56k	59.4	74.8	83.7	93.2	118	V
	1.06m	1.34m	1.49m	1.66m	2.11m	A
62k	62.5	78.7	88.0	98.0	124	V
	1.01m	1.27m	1.42m	1.58m	2.01m	A
68k	65.5	82.5	92.2	103	130	V
	963u	1.21m	1.36m	1.51m	1.92m	A
75k	68.7	86.6	96.8	108	137	V
	917u	1.15m	1.29m	1.44m	1.83m	A
82k	71.9	90.6	101	113	143	V
	877u	1.10m	1.23m	1.37m	1.75m	A
91k	75.7	95.4	107	119	151	V
	832u	1.05m	1.17m	1.31m	1.66m	A

R	0.063 W	0.1 W	0.125 W	0.155 W	0.25 W	
100k	79.4	100	112	124	158	V
	794u	1.00m	1.12m	1.24m	1.58m	A
110k	83.2	105	117	131	166	V
	757u	953u	1.07m	1.19m	1.51m	A
120k	86.9	110	122	136	173	V
	725u	913u	1.02m	1.14m	1.44m	A
130k	90.5	114	127	142	180	V
	696u	877u	981u	1.09m	1.39m	A
150k	97.2	122	137	152	194	V
	648u	816u	913u	1.02m	1.29m	A
160k	100	126	141	157	200	V
	627u	791u	884u	984u	1.25m	A
180k	106	134	150	167	212	V
	592u	745u	833u	928u	1.18m	A
200k	112	141	158	176	224	V
	561u	707u	791u	880u	1.12m	A
220k	118	148	166	185	235	V
	535u	674u	754u	839u	1.07m	A
240k	123	155	173	193	245	V
	512u	645u	722u	804u	1.02m	A
270k	130	164	184	205	260	V
	483u	609u	680u	758u	962u	A
300k	137	173	194	216	274	V
	458u	577u	645u	719u	913u	A
330k	144	182	203	226	287	V
	437u	550u	615u	685u	870u	A
360k	151	190	212	236	300	V
	418u	527u	589u	656u	833u	A
390k	157	197	221	246	312	V
	402u	506u	566u	630u	801u	A
430k	165	207	232	258	328	V
	383u	482u	539u	600u	762u	A
470k	172	217	242	270	343	V
	366u	461u	516u	574u	729u	A
510k	179	226	252	281	357	V
	351u	443u	495u	551u	700u	A
560k	188	237	265	295	374	V
	335u	423u	472u	526u	668u	A

R	0.063 W	0.1 W	0.125 W	0.155 W	0.25 W	
620k	198	249	278	310	394	V
	319u	402u	449u	500u	635u	A
680k	207	261	292	325	412	V
	304u	383u	429u	477u	606u	A
750k	217	274	306	341	433	V
	290u	365u	408u	455u	577u	A
820k	227	286	320	357	453	V
	277u	349u	390u	435u	552u	A
910k	239	302	337	376	477	V
	263u	331u	371u	413u	524u	A
1M	251	316	354	394	500	V
	251u	316u	354u	394u	500u	A
1M2	275	346	387	431	548	V
	229u	289u	323u	359u	456u	A
1M5	307	387	433	482	612	V
	205u	258u	289u	321u	408u	A
1M8	337	424	474	528	671	V
	187u	236u	264u	293u	373u	A
2M2	372	469	524	584	742	V
	169u	213u	238u	265u	337u	A
2M7	412	520	581	647	822	V
	153u	192u	215u	240u	304u	A
3M3	456	574	642	715	908	V
	138u	174u	195u	217u	275u	A
3M9	496	624	698	777	987	V
	127u	160u	179u	199u	253u	A
4M7	544	686	766	854	1.08k	V
	116u	146u	163u	182u	231u	A
5M6	594	748	837	932	1.18k	V
	106u	134u	149u	166u	211u	A
6M8	655	825	922	1.03k	1.30k	V
	96.3u	121u	136u	151u	192u	A
8M2	719	906	1.01k	1.13k	1.43k	V
	87.7u	110u	123u	137u	175u	A
10M	794	1.00k	1.12k	1.24k	1.58k	V
	79.4u	100u	112u	124u	158u	A

Table 7.2 *Maximum power for resistors between 0.33 and 1 W*

R	0.33 W	0.4 W	0.5 W	0.66 W	1 W	
0R10	182m	200m	224m	257m	316m	V
	1.82	2.00	2.24	2.57	3.16	A
0R12	199m	219m	245m	281m	346m	V
	1.66	1.83	2.04	2.35	2.89	A
0R15	222m	245m	274m	315m	387m	V
	1.48	1.63	1.83	2.10	2.58	A
0R18	244m	268m	300m	345m	424m	V
	1.35	1.49	1.67	1.91	2.36	A
0R22	269m	297m	332m	381m	469m	V
	1.22	1.35	1.51	1.73	2.13	A
0R27	298m	329m	367m	422m	520m	V
	1.11	1.22	1.36	1.56	1.92	A
0R33	330m	363m	406m	467m	574m	V
	1.00	1.10	1.23	1.41	1.74	A
0R39	359m	395m	442m	507m	624m	V
	920m	1.01	1.13	1.30	1.60	A
0R47	394m	434m	485m	557m	686m	V
	838m	923m	1.03	1.19	1.46	A
0R56	430m	473m	529m	608m	748m	V
	768m	845m	945m	1.09	1.34	A
0R68	474m	522m	583m	670m	825m	V
	697m	767m	857m	985m	1.21	A
0R82	520m	573m	640m	736m	906m	V
	634m	698m	781m	897m	1.10	A
1R	574m	632m	707m	812m	1.00	V
	574m	632m	707m	812m	1.00	A
1R1	602m	663m	742m	852m	1.05	V
	548m	603m	674m	775m	953m	A
1R2	629m	693m	775m	890m	1.10	V
	524m	577m	645m	742m	913m	A
1R3	655m	721m	806m	926m	1.14	V
	504m	555m	620m	713m	877m	A
1R5	704m	775m	866m	995m	1.22	V
	469m	516m	577m	663m	816m	A
1R6	727m	800m	894m	1.03	1.26	V
	454m	500m	559m	642m	791m	A

R	0.33 W	0.4 W	0.5 W	0.66 W	1 W	
1R8	771m	849m	949m	1.09	1.34	V
	428m	471m	527m	606m	745m	A
2R	812m	894m	1.00	1.15	1.41	V
	406m	447m	500m	574m	707m	A
2R2	852m	938m	1.05	1.20	1.48	V
	387m	426m	477m	548m	674m	A
2R4	890m	980m	1.10	1.26	1.55	V
	371m	408m	456m	524m	645m	A
2R7	944m	1.04	1.16	1.33	1.64	V
	350m	385m	430m	494m	609m	A
3R	995m	1.10	1.22	1.41	1.73	V
	332m	365m	408m	469m	577m	A
3R3	1.04	1.15	1.28	1.48	1.82	V
	316m	348m	389m	447m	550m	A
3R6	1.09	1.20	1.34	1.54	1.90	V
	303m	333m	373m	428m	527m	A
3R9	1.13	1.25	1.40	1.60	1.97	V
	291m	320m	358m	411m	506m	A
4R3	1.19	1.31	1.47	1.68	2.07	V
	277m	305m	341m	392m	482m	A
4R7	1.25	1.37	1.53	1.76	2.17	V
	265m	292m	326m	375m	461m	A
5R1	1.30	1.43	1.60	1.83	2.26	V
	254m	280m	313m	360m	443m	A
5R6	1.36	1.50	1.67	1.92	2.37	V
	243m	267m	299m	343m	423m	A
6R2	1.43	1.57	1.76	2.02	2.49	V
	231m	254m	284m	326m	402m	A
6R8	1.50	1.65	1.84	2.12	2.61	V
	220m	243m	271m	312m	383m	A
7R5	1.57	1.73	1.94	2.22	2.74	V
	210m	231m	258m	297m	365m	A
8R2	1.64	1.81	2.02	2.33	2.86	V
	201m	221m	247m	284m	349m	A
9R1	1.73	1.91	2.13	2.45	3.02	V
	190m	210m	234m	269m	331m	A
10R	1.82	2.00	2.24	2.57	3.16	V
	182m	200m	224m	257m	316m	A

R	0.33 W	0.4 W	0.5 W	0.66 W	1 W	
11R	1.91	2.10	2.35	2.69	3.32	V
	173m	191m	213m	245m	302m	A
12R	1.99	2.19	2.45	2.81	3.46	V
	166m	183m	204m	235m	289m	A
13R	2.07	2.28	2.55	2.93	3.61	V
	159m	175m	196m	225m	277m	A
15R	2.22	2.45	2.74	3.15	3.87	V
	148m	163m	183m	210m	258m	A
16R	2.30	2.53	2.83	3.25	4.00	V
	144m	158m	177m	203m	250m	A
18R	2.44	2.68	3.00	3.45	4.24	V
	135m	149m	167m	191m	236m	A
20R	2.57	2.83	3.16	3.63	4.47	V
	128m	141m	158m	182m	224m	A
22R	2.69	2.97	3.32	3.81	4.69	V
	122m	135m	151m	173m	213m	A
24R	2.81	3.10	3.46	3.98	4.90	V
	117m	129m	144m	166m	204m	A
27R	2.98	3.29	3.67	4.22	5.20	V
	111m	122m	136m	156m	192m	A
30R	3.15	3.46	3.87	4.45	5.48	V
	105m	115m	129m	148m	183m	A
33R	3.30	3.63	4.06	4.67	5.74	V
	100m	110m	123m	141m	174m	A
36R	3.45	3.79	4.24	4.87	6.00	V
	95.7m	105m	118m	135m	167m	A
39R	3.59	3.95	4.42	5.07	6.24	V
	92.0m	101m	113m	130m	160m	A
43R	3.77	4.15	4.64	5.33	6.56	V
	87.6m	96.4m	108m	124m	152m	A
47R	3.94	4.34	4.85	5.57	6.86	V
	83.8m	92.3m	103m	119m	146m	A
51R	4.10	4.52	5.05	5.80	7.14	V
	80.4m	88.6m	99.0m	114m	140m	A
56R	4.30	4.73	5.29	6.08	7.48	V
	76.8m	84.5m	94.5m	109m	134m	A
62R	4.52	4.98	5.57	6.40	7.87	V
	73.0m	80.3m	89.8m	103m	127m	A

R	0.33 W	0.4 W	0.5 W	0.66 W	1 W	
68R	4.74	5.22	5.83	6.70	8.25	V
	69.7m	76.7m	85.7m	98.5m	121m	A
75R	4.97	5.48	6.12	7.04	8.66	V
	66.3m	73.0m	81.6m	93.8m	115m	A
82R	5.20	5.73	6.40	7.36	9.06	V
	63.4m	69.8m	78.1m	89.7m	110m	A
91R	5.48	6.03	6.75	7.75	9.54	V
	60.2m	66.3m	74.1m	85.2m	105m	A
100R	5.74	6.32	7.07	8.12	10.0	V
	57.4m	63.2m	70.7m	81.2m	100m	A
110R	6.02	6.63	7.42	8.52	10.5	V
	54.8m	60.3m	67.4m	77.5m	95.3m	A
120R	6.29	6.93	7.75	8.90	11.0	V
	52.4m	57.7m	64.5m	74.2m	91.3m	A
130R	6.55	7.21	8.06	9.26	11.4	V
	50.4m	55.5m	62.0m	71.3m	87.7m	A
150R	7.04	7.75	8.66	9.95	12.2	V
	46.9m	51.6m	57.7m	66.3m	81.6m	A
160R	7.27	8.00	8.94	10.3	12.6	V
	45.4m	50.0m	55.9m	64.2m	79.1m	A
180R	7.71	8.49	9.49	10.9	13.4	V
	42.8m	47.1m	52.7m	60.6m	74.5m	A
200R	8.12	8.94	10.0	11.5	14.1	V
	40.6m	44.7m	50.0m	57.4m	70.7m	A
220R	8.52	9.38	10.5	12.0	14.8	V
	38.7m	42.6m	47.7m	54.8m	67.4m	A
240R	8.90	9.80	11.0	12.6	15.5	V
	37.1m	40.8m	45.6m	52.4m	64.5m	A
270R	9.44	10.4	11.6	13.3	16.4	V
	35.0m	38.5m	43.0m	49.4m	60.9m	A
300R	9.95	11.0	12.2	14.1	17.3	V
	33.2m	36.5m	40.8m	46.9m	57.7m	A
330R	10.4	11.5	12.8	14.8	18.2	V
	31.6m	34.8m	38.9m	44.7m	55.0m	A
360R	10.9	12.0	13.4	15.4	19.0	V
	30.3m	33.3m	37.3m	42.8m	52.7m	A
390R	11.3	12.5	14.0	16.0	19.7	V
	29.1m	32.0m	35.8m	41.1m	50.6m	A

R	0.33 W	0.4 W	0.5 W	0.66 W	1 W	
430R	11.9	13.1	14.7	16.8	20.7	V
	27.7m	30.5m	34.1m	39.2m	48.2m	A
470R	12.5	13.7	15.3	17.6	21.7	V
	26.5m	29.2m	32.6m	37.5m	46.1m	A
510R	13.0	14.3	16.0	18.3	22.6	V
	25.4m	28.0m	31.3m	36.0m	44.3m	A
560R	13.6	15.0	16.7	19.2	23.7	V
	24.3m	26.7m	29.9m	34.3m	42.3m	A
620R	14.3	15.7	17.6	20.2	24.9	V
	23.1m	25.4m	28.4m	32.6m	40.2m	A
680R	15.0	16.5	18.4	21.2	26.1	V
	22.0m	24.3m	27.1m	31.2m	38.3m	A
750R	15.7	17.3	19.4	22.2	27.4	V
	21.0m	23.1m	25.8m	29.7m	36.5m	A
820R	16.4	18.1	20.2	23.3	28.6	V
	20.1m	22.1m	24.7m	28.4m	34.9m	A
910R	17.3	19.1	21.3	24.5	30.2	V
	19.0m	21.0m	23.4m	26.9m	33.1m	A
1k	18.2	20.0	22.4	25.7	31.6	V
	18.2m	20.0m	22.4m	25.7m	31.6m	A
1k1	19.1	21.0	23.5	26.9	33.2	V
	17.3m	19.1m	21.3m	24.5m	30.2m	A
1k2	19.9	21.9	24.5	28.1	34.6	V
	16.6m	18.3m	20.4m	23.5m	28.9m	A
1k3	20.7	22.8	25.5	29.3	36.1	V
	15.9m	17.5m	19.6m	22.5m	27.7m	A
1k5	22.2	24.5	27.4	31.5	38.7	V
	14.8m	16.3m	18.3m	21.0m	25.8m	A
1k6	23.0	25.3	28.3	32.5	40.0	V
	14.4m	15.8m	17.7m	20.3m	25.0m	A
1k8	24.4	26.8	30.0	34.5	42.4	V
	13.5m	14.9m	16.7m	19.1m	23.6m	A
2k	25.7	28.3	31.6	36.3	44.7	V
	12.8m	14.1m	15.8m	18.2m	22.4m	A
2k2	26.9	29.7	33.2	38.1	46.9	V
	12.2m	13.5m	15.1m	17.3m	21.3m	A
2k4	28.1	31.0	34.6	39.8	49.0	V
	11.7m	12.9m	14.4m	16.6m	20.4m	A

R	0.33 W	0.4 W	0.5 W	0.66 W	1 W	
2k7	29.8	32.9	36.7	42.2	52.0	V
	11.1m	12.2m	13.6m	15.6m	19.2m	A
3k	31.5	34.6	38.7	44.5	54.8	V
	10.5m	11.5m	12.9m	14.8m	18.3m	A
3k3	33.0	36.3	40.6	46.7	57.4	V
	10.0m	11.0m	12.3m	14.1m	17.4m	A
3k6	34.5	37.9	42.4	48.7	60.0	V
	9.57m	10.5m	11.8m	13.5m	16.7m	A
3k9	35.9	39.5	44.2	50.7	62.4	V
	9.20m	10.1m	11.3m	13.0m	16.0m	A
4k3	37.7	41.5	46.4	53.3	65.6	V
	8.76m	9.64m	10.8m	12.4m	15.2m	A
4k7	39.4	43.4	48.5	55.7	68.6	V
	8.38m	9.23m	10.3m	11.9m	14.6m	A
5k1	41.0	45.2	50.5	58.0	71.4	V
	8.04m	8.86m	9.90m	11.4m	14.0m	A
5k6	43.0	47.3	52.9	60.8	74.8	V
	7.68m	8.45m	9.45m	10.9m	13.4m	A
6k2	45.2	49.8	55.7	64.0	78.7	V
	7.30m	8.03m	8.98m	10.3m	12.7m	A
6k8	47.4	52.2	58.3	67.0	82.5	V
	6.97m	7.67m	8.57m	9.85m	12.1m	A
7k5	49.7	54.8	61.2	70.4	86.6	V
	6.63m	7.30m	8.16m	9.38m	11.5m	A
8k2	52.0	57.3	64.0	73.6	90.6	V
	6.34m	6.98m	7.81m	8.97m	11.0m	A
9k1	54.8	60.3	67.5	77.5	95.4	V
	6.02m	6.63m	7.41m	8.52m	10.5m	A
10k	57.4	63.2	70.7	81.2	100	V
	5.74m	6.32m	7.07m	8.12m	10.0m	A
11k	60.2	66.3	74.2	85.2	105	V
	5.48m	6.03m	6.74m	7.75m	9.53m	A
12k	62.9	69.3	77.5	89.0	110	V
	5.24m	5.77m	6.45m	7.42m	9.13m	A
13k	65.5	72.1	80.6	92.6	114	V
	5.04m	5.55m	6.20m	7.13m	8.77m	A
15k	70.4	77.5	86.6	99.5	122	V
	4.69m	5.16m	5.77m	6.63m	8.16m	A

R	0.33 W	0.4 W	0.5 W	0.66 W	1 W	
16k	72.7	80.0	89.4	103	126	V
	4.54m	5.00m	5.59m	6.42m	7.91m	A
18k	77.1	84.9	94.9	109	134	V
	4.28m	4.71m	5.27m	6.06m	7.45m	A
20k	81.2	89.4	100	115	141	V
	4.06m	4.47m	5.00m	5.74m	7.07m	A
22k	85.2	93.8	105	120	148	V
	3.87m	4.26m	4.77m	5.48m	6.74m	A
24k	89.0	98.0	110	126	155	V
	3.71m	4.08m	4.56m	5.24m	6.45m	A
27k	94.4	104	116	133	164	V
	3.50m	3.85m	4.30m	4.94m	6.09m	A
30k	99.5	110	122	141	173	V
	3.32m	3.65m	4.08m	4.69m	5.77m	A
33k	104	115	128	148	182	V
	3.16m	3.48m	3.89m	4.47m	5.50m	A
36k	109	120	134	154	190	V
	3.03m	3.33m	3.73m	4.28m	5.27m	A
39k	113	125	140	160	197	V
	2.91m	3.20m	3.58m	4.11m	5.06m	A
43k	119	131	147	168	207	V
	2.77m	3.05m	3.41m	3.92m	4.82m	A
47k	125	137	153	176	217	V
	2.65m	2.92m	3.26m	3.75m	4.61m	A
51k	130	143	160	183	226	V
	2.54m	2.80m	3.13m	3.60m	4.43m	A
56k	136	150	167	192	237	V
	2.43m	2.67m	2.99m	3.43m	4.23m	A
62k	143	157	176	202	249	V
	2.31m	2.54m	2.84m	3.26m	4.02m	A
68k	150	165	184	212	261	V
	2.20m	2.43m	2.71m	3.12m	3.83m	A
75k	157	173	194	222	274	V
	2.10m	2.31m	2.58m	2.97m	3.65m	A
82k	164	181	202	233	286	V
	2.01m	2.21m	2.47m	2.84m	3.49m	A
91k	173	191	213	245	302	V
	1.90m	2.10m	2.34m	2.69m	3.31m	A

R	0.33 W	0.4 W	0.5 W	0.66 W	1 W	
100k	182	200	224	257	316	V
	1.82m	2.00m	2.24m	2.57m	3.16m	A
110k	191	210	235	269	332	V
	1.73m	1.91m	2.13m	2.45m	3.02m	A
120k	199	219	245	281	346	V
	1.66m	1.83m	2.04m	2.35m	2.89m	A
130k	207	228	255	293	361	V
	1.59m	1.75m	1.96m	2.25m	2.77m	A
150k	222	245	274	315	387	V
	1.48m	1.63m	1.83m	2.10m	2.58m	A
160k	230	253	283	325	400	V
	1.44m	1.58m	1.77m	2.03m	2.50m	A
180k	244	268	300	345	424	V
	1.35m	1.49m	1.67m	1.91m	2.36m	A
200k	257	283	316	363	447	V
	1.28m	1.41m	1.58m	1.82m	2.24m	A
220k	269	297	332	381	469	V
	1.22m	1.35m	1.51m	1.73m	2.13m	A
240k	281	310	346	398	490	V
	1.17m	1.29m	1.44m	1.66m	2.04m	A
270k	298	329	367	422	520	V
	1.11m	1.22m	1.36m	1.56m	1.92m	A
300k	315	346	387	445	548	V
	1.05m	1.15m	1.29m	1.48m	1.83m	A
330k	330	363	406	467	574	V
	1.00m	1.10m	1.23m	1.41m	1.74m	A
360k	345	379	424	487	600	V
	957u	1.05m	1.18m	1.35m	1.67m	A
390k	359	395	442	507	624	V
	920u	1.01m	1.13m	1.30m	1.60m	A
430k	377	415	464	533	656	V
	876u	964u	1.08m	1.24m	1.52m	A
470k	394	434	485	557	686	V
	838u	923u	1.03m	1.19m	1.46m	A
510k	410	452	505	580	714	V
	804u	886u	990u	1.14m	1.40m	A
560k	430	473	529	608	748	V
	768u	845u	945u	1.09m	1.34m	A

R	0.33 W	0.4 W	0.5 W	0.66 W	1 W	
620k	452	498	557	640	787	V
	730u	803u	898u	1.03m	1.27m	A
680k	474	522	583	670	825	V
	697u	767u	857u	985u	1.21m	A
750k	497	548	612	704	866	V
	663u	730u	816u	938u	1.15m	A
820k	520	573	640	736	906	V
	634u	698u	781u	897u	1.10m	A
910k	548	603	675	775	954	V
	602u	663u	741u	852u	1.05m	A
1M	574	632	707	812	1.00k	V
	574u	632u	707u	812u	1.00m	A
1M2	629	693	775	890	1.10k	V
	524u	577u	645u	742u	913u	A
1M5	704	775	866	995	1.22k	V
	469u	516u	577u	663u	816u	A
1M8	771	849	949	1.09k	1.34k	V
	428u	471u	527u	606u	745u	A
2M2	852	938	1.05k	1.20k	1.48k	V
	387u	426u	477u	548u	674u	A
2M7	944	1.04k	1.16k	1.33k	1.64k	V
	350u	385u	430u	494u	609u	A
3M3	1.04k	1.15k	1.28k	1.48k	1.82k	V
	316u	348u	389u	447u	550u	A
3M9	1.13k	1.25k	1.40k	1.60k	1.97k	V
	291u	320u	358u	411u	506u	A
4M7	1.25k	1.37k	1.53k	1.76k	2.17k	V
	265u	292u	326u	375u	461u	A
5M6	1.36k	1.50k	1.67k	1.92k	2.37k	V
	243u	267u	299u	343u	423u	A
6M8	1.50k	1.65k	1.84k	2.12k	2.61k	V
	220u	243u	271u	312u	383u	A
8M2	1.64k	1.81k	2.02k	2.33k	2.86k	V
	201u	221u	247u	284u	349u	A
10M	1.82k	2.00k	2.24k	2.57k	3.16k	V
	182u	200u	224u	257u	316u	A

Table 7.3 *Maximum power for resistors between 2 and 5 W*

R	2 W	2.5 W	3 W	4 W	5 W	
0R01	141m	158m	173m	200m	224m	V
	14.1	15.8	17.3	20.0	22.4	A
0R02	200m	224m	245m	283m	316m	V
	10.0	11.2	12.2	14.1	15.8	A
0R03	245m	274m	300m	346m	387m	V
	8.16	9.13	10.0	11.5	12.9	A
0R033	257m	287m	315m	363m	406m	V
	7.78	8.70	9.53	11.0	12.3	A
0R047	307m	343m	375m	434m	485m	V
	6.52	7.29	7.99	9.23	10.3	A
0R05	316m	354m	387m	447m	500m	V
	6.32	7.07	7.75	8.94	10.0	A
0R056	335m	374m	410m	473m	529m	V
	5.98	6.68	7.32	8.45	9.45	A
0R068	369m	412m	452m	522m	583m	V
	5.42	6.06	6.64	7.67	8.57	A
0R07	374m	418m	458m	529m	592m	V
	5.35	5.98	6.55	7.56	8.45	A
0R10	447m	500m	548m	632m	707m	V
	4.47	5.00	5.48	6.32	7.07	A
0R12	490m	548m	600m	693m	775m	V
	4.08	4.56	5.00	5.77	6.45	A
0R15	548m	612m	671m	775m	866m	V
	3.65	4.08	4.47	5.16	5.77	A
0R18	600m	671m	735m	849m	949m	V
	3.33	3.73	4.08	4.71	5.27	A
0R22	663m	742m	812m	938m	1.05	V
	3.02	3.37	3.69	4.26	4.77	A
0R27	735m	822m	900m	1.04	1.16	V
	2.72	3.04	3.33	3.85	4.30	A
0R33	812m	908m	995m	1.15	1.28	V
	2.46	2.75	3.02	3.48	3.89	A
0R39	883m	987m	1.08	1.25	1.40	V
	2.26	2.53	2.77	3.20	3.58	A
0R47	970m	1.08	1.19	1.37	1.53	V
	2.06	2.31	2.53	2.92	3.26	A

R	2 W	2.5 W	3 W	4 W	5 W	
0R56	1.06	1.18	1.30	1.50	1.67	V
	1.89	2.11	2.31	2.67	2.99	A
0R68	1.17	1.30	1.43	1.65	1.84	V
	1.71	1.92	2.10	2.43	2.71	A
0R82	1.28	1.43	1.57	1.81	2.02	V
	1.56	1.75	1.91	2.21	2.47	A
1R	1.41	1.58	1.73	2.00	2.24	V
	1.41	1.58	1.73	2.00	2.24	A
1R2	1.55	1.73	1.90	2.19	2.45	V
	1.29	1.44	1.58	1.83	2.04	A
1R5	1.73	1.94	2.12	2.45	2.74	V
	1.15	1.29	1.41	1.63	1.83	A
1R8	1.90	2.12	2.32	2.68	3.00	V
	1.05	1.18	1.29	1.49	1.67	A
2R2	2.10	2.35	2.57	2.97	3.32	V
	953m	1.07	1.17	1.35	1.51	A
2R7	2.32	2.60	2.85	3.29	3.67	V
	861m	962m	1.05	1.22	1.36	A
3R3	2.57	2.87	3.15	3.63	4.06	V
	778m	870m	953m	1.10	1.23	A
3R9	2.79	3.12	3.42	3.95	4.42	V
	716m	801m	877m	1.01	1.13	A
4R7	3.07	3.43	3.75	4.34	4.85	V
	652m	729m	799m	923m	1.03	A
5R6	3.35	3.74	4.10	4.73	5.29	V
	598m	668m	732m	845m	945m	A
6R8	3.69	4.12	4.52	5.22	5.83	V
	542m	606m	664m	767m	857m	A
8R2	4.05	4.53	4.96	5.73	6.40	V
	494m	552m	605m	698m	781m	A
10R	4.47	5.00	5.48	6.32	7.07	V
	447m	500m	548m	632m	707m	A
12R	4.90	5.48	6.00	6.93	7.75	V
	408m	456m	500m	577m	645m	A
15R	5.48	6.12	6.71	7.75	8.66	V
	365m	408m	447m	516m	577m	A
18R	6.00	6.71	7.35	8.49	9.49	V

R	2 W	2.5 W	3 W	4 W	5 W	
	333m	373m	408m	471m	527m	A
22R	6.63	7.42	8.12	9.38	10.5	V
	302m	337m	369m	426m	477m	A
27R	7.35	8.22	9.00	10.4	11.6	V
	272m	304m	333m	385m	430m	A
33R	8.12	9.08	9.95	11.5	12.8	V
	246m	275m	302m	348m	389m	A
39R	8.83	9.87	10.8	12.5	14.0	V
	226m	253m	277m	320m	358m	A
47R	9.70	10.8	11.9	13.7	15.3	V
	206m	231m	253m	292m	326m	A
56R	10.6	11.8	13.0	15.0	16.7	V
	189m	211m	231m	267m	299m	A
68R	11.7	13.0	14.3	16.5	18.4	V
	171m	192m	210m	243m	271m	A
82R	12.8	14.3	15.7	18.1	20.2	V
	156m	175m	191m	221m	247m	A
100R	14.1	15.8	17.3	20.0	22.4	V
	141m	158m	173m	200m	224m	A
120R	15.5	17.3	19.0	21.9	24.5	V
	129m	144m	158m	183m	204m	A
150R	17.3	19.4	21.2	24.5	27.4	V
	115m	129m	141m	163m	183m	A
180R	19.0	21.2	23.2	26.8	30.0	V
	105m	118m	129m	149m	167m	A
220R	21.0	23.5	25.7	29.7	33.2	V
	95.3m	107m	117m	135m	151m	A
270R	23.2	26.0	28.5	32.9	36.7	V
	86.1m	96.2m	105m	122m	136m	A
330R	25.7	28.7	31.5	36.3	40.6	V
	77.8m	87.0m	95.3m	110m	123m	A
390R	27.9	31.2	34.2	39.5	44.2	V
	71.6m	80.1m	87.7m	101m	113m	A
470R	30.7	34.3	37.5	43.4	48.5	V
	65.2m	72.9m	79.9m	92.3m	103m	A
560R	33.5	37.4	41.0	47.3	52.9	V
	59.8m	66.8m	73.2m	84.5m	94.5m	A

R	2 W	2.5 W	3 W	4 W	5 W	
680R	36.9	41.2	45.2	52.2	58.3	V
	54.2m	60.6m	66.4m	76.7m	85.7m	A
820R	40.5	45.3	49.6	57.3	64.0	V
	49.4m	55.2m	60.5m	69.8m	78.1m	A
1k	44.7	50.0	54.8	63.2	70.7	V
	44.7m	50.0m	54.8m	63.2m	70.7m	A
1k2	49.0	54.8	60.0	69.3	77.5	V
	40.8m	45.6m	50.0m	57.7m	64.5m	A
1k5	54.8	61.2	67.1	77.5	86.6	V
	36.5m	40.8m	44.7m	51.6m	57.7m	A
1k8	60.0	67.1	73.5	84.9	94.9	V
	33.3m	37.3m	40.8m	47.1m	52.7m	A
2k2	66.3	74.2	81.2	93.8	105	V
	30.2m	33.7m	36.9m	42.6m	47.7m	A
2k7	73.5	82.2	90.0	104	116	V
	27.2m	30.4m	33.3m	38.5m	43.0m	A
3k3	81.2	90.8	99.5	115	128	V
	24.6m	27.5m	30.2m	34.8m	38.9m	A
3k9	88.3	98.7	108	125	140	V
	22.6m	25.3m	27.7m	32.0m	35.8m	A
4k7	97.0	108	119	137	153	V
	20.6m	23.1m	25.3m	29.2m	32.6m	A
5k6	106	118	130	150	167	V
	18.9m	21.1m	23.1m	26.7m	29.9m	A
6k8	117	130	143	165	184	V
	17.1m	19.2m	21.0m	24.3m	27.1m	A
8k2	128	143	157	181	202	V
	15.6m	17.5m	19.1m	22.1m	24.7m	A
10k	141	158	173	200	224	V
	14.1m	15.8m	17.3m	20.0m	22.4m	A
12k	155	173	190	219	245	V
	12.9m	14.4m	15.8m	18.3m	20.4m	A
15k	173	194	212	245	274	V
	11.5m	12.9m	14.1m	16.3m	18.3m	A
18k	190	212	232	268	300	V
	10.5m	11.8m	12.9m	14.9m	16.7m	A
22k	210	235	257	297	332	V

R	2 W	2.5 W	3 W	4 W	5 W	
	9.53m	10.7m	11.7m	13.5m	15.1m	A
27k	232	260	285	329	367	V
	8.61m	9.62m	10.5m	12.2m	13.6m	A
33k	257	287	315	363	406	V
	7.78m	8.70m	9.53m	11.0m	12.3m	A
39k	279	312	342	395	442	V
	7.16m	8.01m	8.77m	10.1m	11.3m	A
47k	307	343	375	434	485	V
	6.52m	7.29m	7.99m	9.23m	10.3m	A
56k	335	374	410	473	529	V
	5.98m	6.68m	7.32m	8.45m	9.45m	A
68k	369	412	452	522	583	V
	5.42m	6.06m	6.64m	7.67m	8.57m	A
82k	405	453	496	573	640	V
	4.94m	5.52m	6.05m	6.98m	7.81m	A
100k	447	500	548	632	707	V
	4.47m	5.00m	5.48m	6.32m	7.07m	A
120k	490	548	600	693	775	V
	4.08m	4.56m	5.00m	5.77m	6.45m	A
150k	548	612	671	775	866	V
	3.65m	4.08m	4.47m	5.16m	5.77m	A
180k	600	671	735	849	949	V
	3.33m	3.73m	4.08m	4.71m	5.27m	A
220k	663	742	812	938	1.05k	V
	3.02m	3.37m	3.69m	4.26m	4.77m	A
270k	735	822	900	1.04k	1.16k	V
	2.72m	3.04m	3.33m	3.85m	4.30m	A
330k	812	908	995	1.15k	1.28k	V
	2.46m	2.75m	3.02m	3.48m	3.89m	A
390k	883	987	1.08k	1.25k	1.40k	V
	2.26m	2.53m	2.77m	3.20m	3.58m	A
470k	970	1.08k	1.19k	1.37k	1.53k	V
	2.06m	2.31m	2.53m	2.92m	3.26m	A
560k	1.06k	1.18k	1.30k	1.50k	1.67k	V
	1.89m	2.11m	2.31m	2.67m	2.99m	A
680k	1.17k	1.30k	1.43k	1.65k	1.84k	V
	1.71m	1.92m	2.10m	2.43m	2.71m	A

R	2 W	2.5 W	3 W	4 W	5 W	
820k	1.28k	1.43k	1.57k	1.81k	2.02k	V
	1.56m	1.75m	1.91m	2.21m	2.47m	A
1M	1.41k	1.58k	1.73k	2.00k	2.24k	V
	1.41m	1.58m	1.73m	2.00m	2.24m	A

Table 7.4 *Maximum power for resistors between 6 and 10 W*

R	6 W	7 W	7.5 W	9 W	10 W	
0R01	245m	265m	274m	300m	316m	V
	24.5	26.5	27.4	30.0	31.6	A
0R02	346m	374m	387m	424m	447m	V
	17.3	18.7	19.4	21.2	22.4	A
0R03	424m	458m	474m	520m	548m	V
	14.1	15.3	15.8	17.3	18.3	A
0R033	445m	481m	497m	545m	574m	V
	13.5	14.6	15.1	16.5	17.4	A
0R047	531m	574m	594m	650m	686m	V
	11.3	12.2	12.6	13.8	14.6	A
0R05	548m	592m	612m	671m	707m	V
	11.0	11.8	12.2	13.4	14.1	A
0R056	580m	626m	648m	710m	748m	V
	10.4	11.2	11.6	12.7	13.4	A
0R068	639m	690m	714m	782m	825m	V
	9.39	10.1	10.5	11.5	12.1	A
0R07	648m	700m	725m	794m	837m	V
	9.26	10.0	10.4	11.3	12.0	A
0R10	775m	837m	866m	949m	1.00	V
	7.75	8.37	8.66	9.49	10.0	A
0R12	849m	917m	949m	1.04	1.10	V
	7.07	7.64	7.91	8.66	9.13	A
0R15	949m	1.02	1.06	1.16	1.22	V
	6.32	6.83	7.07	7.75	8.16	A
0R18	1.04	1.12	1.16	1.27	1.34	V
	5.77	6.24	6.45	7.07	7.45	A
0R22	1.15	1.24	1.28	1.41	1.48	V
	5.22	5.64	5.84	6.40	6.74	A
0R27	1.27	1.37	1.42	1.56	1.64	V
	4.71	5.09	5.27	5.77	6.09	A
0R33	1.41	1.52	1.57	1.72	1.82	V
	4.26	4.61	4.77	5.22	5.50	A
0R39	1.53	1.65	1.71	1.87	1.97	V
	3.92	4.24	4.39	4.80	5.06	A
0R47	1.68	1.81	1.88	2.06	2.17	V
	3.57	3.86	3.99	4.38	4.61	A

R	6 W	7 W	7.5 W	9 W	10 W	
0R56	1.83	1.98	2.05	2.24	2.37	V
	3.27	3.54	3.66	4.01	4.23	A
0R68	2.02	2.18	2.26	2.47	2.61	V
	2.97	3.21	3.32	3.64	3.83	A
0R82	2.22	2.40	2.48	2.72	2.86	V
	2.71	2.92	3.02	3.31	3.49	A
1R	2.45	2.65	2.74	3.00	3.16	V
	2.45	2.65	2.74	3.00	3.16	A
1R2	2.68	2.90	3.00	3.29	3.46	V
	2.24	2.42	2.50	2.74	2.89	A
1R5	3.00	3.24	3.35	3.67	3.87	V
	2.00	2.16	2.24	2.45	2.58	A
1R8	3.29	3.55	3.67	4.02	4.24	V
	1.83	1.97	2.04	2.24	2.36	A
2R2	3.63	3.92	4.06	4.45	4.69	V
	1.65	1.78	1.85	2.02	2.13	A
2R7	4.02	4.35	4.50	4.93	5.20	V
	1.49	1.61	1.67	1.83	1.92	A
3R3	4.45	4.81	4.97	5.45	5.74	V
	1.35	1.46	1.51	1.65	1.74	A
3R9	4.84	5.22	5.41	5.92	6.24	V
	1.24	1.34	1.39	1.52	1.60	A
4R7	5.31	5.74	5.94	6.50	6.86	V
	1.13	1.22	1.26	1.38	1.46	A
5R6	5.80	6.26	6.48	7.10	7.48	V
	1.04	1.12	1.16	1.27	1.34	A
6R8	6.39	6.90	7.14	7.82	8.25	V
	939m	1.01	1.05	1.15	1.21	A
8R2	7.01	7.58	7.84	8.59	9.06	V
	855m	924m	956m	1.05	1.10	A
10R	7.75	8.37	8.66	9.49	10.0	V
	775m	837m	866m	949m	1.00	A
12R	8.49	9.17	9.49	10.4	11.0	V
	707m	764m	791m	866m	913m	A
15R	9.49	10.2	10.6	11.6	12.2	V
	632m	683m	707m	775m	816m	A

R	6 W	7 W	7.5 W	9 W	10 W	
18R	10.4	11.2	11.6	12.7	13.4	V
	577m	624m	645m	707m	745m	A
22R	11.5	12.4	12.8	14.1	14.8	V
	522m	564m	584m	640m	674m	A
27R	12.7	13.7	14.2	15.6	16.4	V
	471m	509m	527m	577m	609m	A
33R	14.1	15.2	15.7	17.2	18.2	V
	426m	461m	477m	522m	550m	A
39R	15.3	16.5	17.1	18.7	19.7	V
	392m	424m	439m	480m	506m	A
47R	16.8	18.1	18.8	20.6	21.7	V
	357m	386m	399m	438m	461m	A
56R	18.3	19.8	20.5	22.4	23.7	V
	327m	354m	366m	401m	423m	A
68R	20.2	21.8	22.6	24.7	26.1	V
	297m	321m	332m	364m	383m	A
82R	22.2	24.0	24.8	27.2	28.6	V
	271m	292m	302m	331m	349m	A
100R	24.5	26.5	27.4	30.0	31.6	V
	245m	265m	274m	300m	316m	A
120R	26.8	29.0	30.0	32.9	34.6	V
	224m	242m	250m	274m	289m	A
150R	30.0	32.4	33.5	36.7	38.7	V
	200m	216m	224m	245m	258m	A
180R	32.9	35.5	36.7	40.2	42.4	V
	183m	197m	204m	224m	236m	A
220R	36.3	39.2	40.6	44.5	46.9	V
	165m	178m	185m	202m	213m	A
270R	40.2	43.5	45.0	49.3	52.0	V
	149m	161m	167m	183m	192m	A
330R	44.5	48.1	49.7	54.5	57.4	V
	135m	146m	151m	165m	174m	A
390R	48.4	52.2	54.1	59.2	62.4	V
	124m	134m	139m	152m	160m	A
470R	53.1	57.4	59.4	65.0	68.6	V
	113m	122m	126m	138m	146m	A

R	6 W	7 W	7.5 W	9 W	10 W	
560R	58.0	62.6	64.8	71.0	74.8	V
	104m	112m	116m	127m	134m	A
680R	63.9	69.0	71.4	78.2	82.5	V
	93.9m	101m	105m	115m	121m	A
820R	70.1	75.8	78.4	85.9	90.6	V
	85.5m	92.4m	95.6m	105m	110m	A
1k	77.5	83.7	86.6	94.9	100	V
	77.5m	83.7m	86.6m	94.9m	100m	A
1k2	84.9	91.7	94.9	104	110	V
	70.7m	76.4m	79.1m	86.6m	91.3m	A
1k5	94.9	102	106	116	122	V
	63.2m	68.3m	70.7m	77.5m	81.6m	A
1k8	104	112	116	127	134	V
	57.7m	62.4m	64.5m	70.7m	74.5m	A
2k2	115	124	128	141	148	V
	52.2m	56.4m	58.4m	64.0m	67.4m	A
2k7	127	137	142	156	164	V
	47.1m	50.9m	52.7m	57.7m	60.9m	A
3k3	141	152	157	172	182	V
	42.6m	46.1m	47.7m	52.2m	55.0m	A
3k9	153	165	171	187	197	V
	39.2m	42.4m	43.9m	48.0m	50.6m	A
4k7	168	181	188	206	217	V
	35.7m	38.6m	39.9m	43.8m	46.1m	A
5k6	183	198	205	224	237	V
	32.7m	35.4m	36.6m	40.1m	42.3m	A
6k8	202	218	226	247	261	V
	29.7m	32.1m	33.2m	36.4m	38.3m	A
8k2	222	240	248	272	286	V
	27.1m	29.2m	30.2m	33.1m	34.9m	A
10k	245	265	274	300	316	V
	24.5m	26.5m	27.4m	30.0m	31.6m	A
12k	268	290	300	329	346	V
	22.4m	24.2m	25.0m	27.4m	28.9m	A
15k	300	324	335	367	387	V
	20.0m	21.6m	22.4m	24.5m	25.8m	A

R	6 W	7 W	7.5 W	9 W	10 W	
18k	329	355	367	402	424	V
	18.3m	19.7m	20.4m	22.4m	23.6m	A
22k	363	392	406	445	469	V
	16.5m	17.8m	18.5m	20.2m	21.3m	A
27k	402	435	450	493	520	V
	14.9m	16.1m	16.7m	18.3m	19.2m	A
33k	445	481	497	545	574	V
	13.5m	14.6m	15.1m	16.5m	17.4m	A
39k	484	522	541	592	624	V
	12.4m	13.4m	13.9m	15.2m	16.0m	A
47k	531	574	594	650	686	V
	11.3m	12.2m	12.6m	13.8m	14.6m	A
56k	580	626	648	710	748	V
	10.4m	11.2m	11.6m	12.7m	13.4m	A
68k	639	690	714	782	825	V
	9.39m	10.1m	10.5m	11.5m	12.1m	A
82k	701	758	784	859	906	V
	8.55m	9.24m	9.56m	10.5m	11.0m	A
100k	775	837	866	949	1.00k	V
	7.75m	8.37m	8.66m	9.49m	10.0m	A

Table 7.5 *Maximum power for resistors between 11 and 25 W*

R	11 W	12 W	15 W	17 W	25 W	
0R01	332m	346m	387m	412m	500m	V
	33.2	34.6	38.7	41.2	50.0	A
0R02	469m	490m	548m	583m	707m	V
	23.5	24.5	27.4	29.2	35.4	A
0R03	574m	600m	671m	714m	866m	V
	19.1	20.0	22.4	23.8	28.9	A
0R033	602m	629m	704m	749m	908m	V
	18.3	19.1	21.3	22.7	27.5	A
0R047	719m	751m	840m	894m	1.08	V
	15.3	16.0	17.9	19.0	23.1	A
0R05	742m	775m	866m	922m	1.12	V
	14.8	15.5	17.3	18.4	22.4	A
0R056	785m	820m	917m	976m	1.18	V
	14.0	14.6	16.4	17.4	21.1	A
0R068	865m	903m	1.01	1.08	1.30	V
	12.7	13.3	14.9	15.8	19.2	A
0R07	877m	917m	1.02	1.09	1.32	V
	12.5	13.1	14.6	15.6	18.9	A
0R10	1.05	1.10	1.22	1.30	1.58	V
	10.5	11.0	12.2	13.0	15.8	A
0R12	1.15	1.20	1.34	1.43	1.73	V
	9.57	10.0	11.2	11.9	14.4	A
0R15	1.28	1.34	1.50	1.60	1.94	V
	8.56	8.94	10.0	10.6	12.9	A
0R18	1.41	1.47	1.64	1.75	2.12	V
	7.82	8.16	9.13	9.72	11.8	A
0R22	1.56	1.62	1.82	1.93	2.35	V
	7.07	7.39	8.26	8.79	10.7	A
0R27	1.72	1.80	2.01	2.14	2.60	V
	6.38	6.67	7.45	7.93	9.62	A
0R33	1.91	1.99	2.22	2.37	2.87	V
	5.77	6.03	6.74	7.18	8.70	A
0R39	2.07	2.16	2.42	2.57	3.12	V
	5.31	5.55	6.20	6.60	8.01	A
0R47	2.27	2.37	2.66	2.83	3.43	V
	4.84	5.05	5.65	6.01	7.29	A

R	11 W	12 W	15 W	17 W	25 W	
0R56	2.48	2.59	2.90	3.09	3.74	V
	4.43	4.63	5.18	5.51	6.68	A
0R68	2.73	2.86	3.19	3.40	4.12	V
	4.02	4.20	4.70	5.00	6.06	A
0R82	3.00	3.14	3.51	3.73	4.53	V
	3.66	3.83	4.28	4.55	5.52	A
1R	3.32	3.46	3.87	4.12	5.00	V
	3.32	3.46	3.87	4.12	5.00	A
1R2	3.63	3.79	4.24	4.52	5.48	V
	3.03	3.16	3.54	3.76	4.56	A
1R5	4.06	4.24	4.74	5.05	6.12	V
	2.71	2.83	3.16	3.37	4.08	A
1R8	4.45	4.65	5.20	5.53	6.71	V
	2.47	2.58	2.89	3.07	3.73	A
2R2	4.92	5.14	5.74	6.12	7.42	V
	2.24	2.34	2.61	2.78	3.37	A
2R7	5.45	5.69	6.36	6.77	8.22	V
	2.02	2.11	2.36	2.51	3.04	A
3R3	6.02	6.29	7.04	7.49	9.08	V
	1.83	1.91	2.13	2.27	2.75	A
3R9	6.55	6.84	7.65	8.14	9.87	V
	1.68	1.75	1.96	2.09	2.53	A
4R7	7.19	7.51	8.40	8.94	10.8	V
	1.53	1.60	1.79	1.90	2.31	A
5R6	7.85	8.20	9.17	9.76	11.8	V
	1.40	1.46	1.64	1.74	2.11	A
6R8	8.65	9.03	10.1	10.8	13.0	V
	1.27	1.33	1.49	1.58	1.92	A
8R2	9.50	9.92	11.1	11.8	14.3	V
	1.16	1.21	1.35	1.44	1.75	A
10R	10.5	11.0	12.2	13.0	15.8	V
	1.05	1.10	1.22	1.30	1.58	A
12R	11.5	12.0	13.4	14.3	17.3	V
	957m	1.00	1.12	1.19	1.44	A
15R	12.8	13.4	15.0	16.0	19.4	V
	856m	894m	1.00	1.06	1.29	A

R	11 W	12 W	15 W	17 W	25 W	
18R	14.1	14.7	16.4	17.5	21.2	V
	782m	816m	913m	972m	1.18	A
22R	15.6	16.2	18.2	19.3	23.5	V
	707m	739m	826m	879m	1.07	A
27R	17.2	18.0	20.1	21.4	26.0	V
	638m	667m	745m	793m	962m	A
33R	19.1	19.9	22.2	23.7	28.7	V
	577m	603m	674m	718m	870m	A
39R	20.7	21.6	24.2	25.7	31.2	V
	531m	555m	620m	660m	801m	A
47R	22.7	23.7	26.6	28.3	34.3	V
	484m	505m	565m	601m	729m	A
56R	24.8	25.9	29.0	30.9	37.4	V
	443m	463m	518m	551m	668m	A
68R	27.3	28.6	31.9	34.0	41.2	V
	402m	420m	470m	500m	606m	A
82R	30.0	31.4	35.1	37.3	45.3	V
	366m	383m	428m	455m	552m	A
100R	33.2	34.6	38.7	41.2	50.0	V
	332m	346m	387m	412m	500m	A
120R	36.3	37.9	42.4	45.2	54.8	V
	303m	316m	354m	376m	456m	A
150R	40.6	42.4	47.4	50.5	61.2	V
	271m	283m	316m	337m	408m	A
180R	44.5	46.5	52.0	55.3	67.1	V
	247m	258m	289m	307m	373m	A
220R	49.2	51.4	57.4	61.2	74.2	V
	224m	234m	261m	278m	337m	A
270R	54.5	56.9	63.6	67.7	82.2	V
	202m	211m	236m	251m	304m	A
330R	60.2	62.9	70.4	74.9	90.8	V
	183m	191m	213m	227m	275m	A
390R	65.5	68.4	76.5	81.4	98.7	V
	168m	175m	196m	209m	253m	A
470R	71.9	75.1	84.0	89.4	108	V
	153m	160m	179m	190m	231m	A

R	11 W	12 W	15 W	17 W	25 W	
560R	78.5	82.0	91.7	97.6	118	V
	140m	146m	164m	174m	211m	A
680R	86.5	90.3	101	108	130	V
	127m	133m	149m	158m	192m	A
820R	95.0	99.2	111	118	143	V
	116m	121m	135m	144m	175m	A
1k	105	110	122	130	158	V
	105m	110m	122m	130m	158m	A
1k2	115	120	134	143	173	V
	95.7m	100m	112m	119m	144m	A
1k5	128	134	150	160	194	V
	85.6m	89.4m	100m	106m	129m	A
1k8	141	147	164	175	212	V
	78.2m	81.6m	91.3m	97.2m	118m	A
2k2	156	162	182	193	235	V
	70.7m	73.9m	82.6m	87.9m	107m	A
2k7	172	180	201	214	260	V
	63.8m	66.7m	74.5m	79.3m	96.2m	A
3k3	191	199	222	237	287	V
	57.7m	60.3m	67.4m	71.8m	87.0m	A
3k9	207	216	242	257	312	V
	53.1m	55.5m	62.0m	66.0m	80.1m	A
4k7	227	237	266	283	343	V
	48.4m	50.5m	56.5m	60.1m	72.9m	A
5k6	248	259	290	309	374	V
	44.3m	46.3m	51.8m	55.1m	66.8m	A
6k8	273	286	319	340	412	V
	40.2m	42.0m	47.0m	50.0m	60.6m	A
8k2	300	314	351	373	453	V
	36.6m	38.3m	42.8m	45.5m	55.2m	A
10k	332	346	387	412	500	V
	33.2m	34.6m	38.7m	41.2m	50.0m	A
12k	363	379	424	452	548	V
	30.3m	31.6m	35.4m	37.6m	45.6m	A
15k	406	424	474	505	612	V
	27.1m	28.3m	31.6m	33.7m	40.8m	A

R	11 W	12 W	15 W	17 W	25 W	
18k	445	465	520	553	671	V
	24.7m	25.8m	28.9m	30.7m	37.3m	A
22k	492	514	574	612	742	V
	22.4m	23.4m	26.1m	27.8m	33.7m	A
27k	545	569	636	677	822	V
	20.2m	21.1m	23.6m	25.1m	30.4m	A
33k	602	629	704	749	908	V
	18.3m	19.1m	21.3m	22.7m	27.5m	A
39k	655	684	765	814	987	V
	16.8m	17.5m	19.6m	20.9m	25.3m	A
47k	719	751	840	894	1.08k	V
	15.3m	16.0m	17.9m	19.0m	23.1m	A
56k	785	820	917	976	1.18k	V
	14.0m	14.6m	16.4m	17.4m	21.1m	A
68k	865	903	1.01k	1.08k	1.30k	V
	12.7m	13.3m	14.9m	15.8m	19.2m	A
82k	950	992	1.11k	1.18k	1.43k	V
	11.6m	12.1m	13.5m	14.4m	17.5m	A
100k	1.05k	1.10k	1.22k	1.30k	1.58k	V
	10.5m	11.0m	12.2m	13.0m	15.8m	A

Table 7.6 *Maximum power for resistors between 50 and 300 W*

R	50 W	100 W	200 W	250 W	300 W	
0R01	707m	1.00	1.41	1.58	1.73	V
	70.7	100	141	158	173	A
0R02	1.00	1.41	2.00	2.24	2.45	V
	50.0	70.7	100	112	122	A
0R03	1.22	1.73	2.45	2.74	3.00	V
	40.8	57.7	81.6	91.3	100	A
0R033	1.28	1.82	2.57	2.87	3.15	V
	38.9	55.0	77.8	87.0	95.3	A
0R047	1.53	2.17	3.07	3.43	3.75	V
	32.6	46.1	65.2	72.9	79.9	A
0R05	1.58	2.24	3.16	3.54	3.87	V
	31.6	44.7	63.2	70.7	77.5	A
0R056	1.67	2.37	3.35	3.74	4.10	V
	29.9	42.3	59.8	66.8	73.2	A
0R068	1.84	2.61	3.69	4.12	4.52	V
	27.1	38.3	54.2	60.6	66.4	A
0R07	1.87	2.65	3.74	4.18	4.58	V
	26.7	37.8	53.5	59.8	65.5	A
0R10	2.24	3.16	4.47	5.00	5.48	V
	22.4	31.6	44.7	50.0	54.8	A
0R12	2.45	3.46	4.90	5.48	6.00	V
	20.4	28.9	40.8	45.6	50.0	A
0R15	2.74	3.87	5.48	6.12	6.71	V
	18.3	25.8	36.5	40.8	44.7	A
0R18	3.00	4.24	6.00	6.71	7.35	V
	16.7	23.6	33.3	37.3	40.8	A
0R22	3.32	4.69	6.63	7.42	8.12	V
	15.1	21.3	30.2	33.7	36.9	A
0R27	3.67	5.20	7.35	8.22	9.00	V
	13.6	19.2	27.2	30.4	33.3	A
0R33	4.06	5.74	8.12	9.08	9.95	V
	12.3	17.4	24.6	27.5	30.2	A
0R39	4.42	6.24	8.83	9.87	10.8	V
	11.3	16.0	22.6	25.3	27.7	A
0R47	4.85	6.86	9.70	10.8	11.9	V
	10.3	14.6	20.6	23.1	25.3	A

R	50 W	100 W	200 W	250 W	300 W	
0R56	5.29	7.48	10.6	11.8	13.0	V
	9.45	13.4	18.9	21.1	23.1	A
0R68	5.83	8.25	11.7	13.0	14.3	V
	8.57	12.1	17.1	19.2	21.0	A
0R82	6.40	9.06	12.8	14.3	15.7	V
	7.81	11.0	15.6	17.5	19.1	A
1R	7.07	10.0	14.1	15.8	17.3	V
	7.07	10.0	14.1	15.8	17.3	A
1R2	7.75	11.0	15.5	17.3	19.0	V
	6.45	9.13	12.9	14.4	15.8	A
1R5	8.66	12.2	17.3	19.4	21.2	V
	5.77	8.16	11.5	12.9	14.1	A
1R8	9.49	13.4	19.0	21.2	23.2	V
	5.27	7.45	10.5	11.8	12.9	A
2R2	10.5	14.8	21.0	23.5	25.7	V
	4.77	6.74	9.53	10.7	11.7	A
2R7	11.6	16.4	23.2	26.0	28.5	V
	4.30	6.09	8.61	9.62	10.5	A
3R3	12.8	18.2	25.7	28.7	31.5	V
	3.89	5.50	7.78	8.70	9.53	A
3R9	14.0	19.7	27.9	31.2	34.2	V
	3.58	5.06	7.16	8.01	8.77	A
4R7	15.3	21.7	30.7	34.3	37.5	V
	3.26	4.61	6.52	7.29	7.99	A
5R6	16.7	23.7	33.5	37.4	41.0	V
	2.99	4.23	5.98	6.68	7.32	A
6R8	18.4	26.1	36.9	41.2	45.2	V
	2.71	3.83	5.42	6.06	6.64	A
8R2	20.2	28.6	40.5	45.3	49.6	V
	2.47	3.49	4.94	5.52	6.05	A
10R	22.4	31.6	44.7	50.0	54.8	V
	2.24	3.16	4.47	5.00	5.48	A
12R	24.5	34.6	49.0	54.8	60.0	V
	2.04	2.89	4.08	4.56	5.00	A
15R	27.4	38.7	54.8	61.2	67.1	V
	1.83	2.58	3.65	4.08	4.47	A

R	50 W	100 W	200 W	250 W	300 W	
18R	30.0	42.4	60.0	67.1	73.5	V
	1.67	2.36	3.33	3.73	4.08	A
22R	33.2	46.9	66.3	74.2	81.2	V
	1.51	2.13	3.02	3.37	3.69	A
27R	36.7	52.0	73.5	82.2	90.0	V
	1.36	1.92	2.72	3.04	3.33	A
33R	40.6	57.4	81.2	90.8	99.5	V
	1.23	1.74	2.46	2.75	3.02	A
39R	44.2	62.4	88.3	98.7	108	V
	1.13	1.60	2.26	2.53	2.77	A
47R	48.5	68.6	97.0	108	119	V
	1.03	1.46	2.06	2.31	2.53	A
56R	52.9	74.8	106	118	130	V
	945m	1.34	1.89	2.11	2.31	A
68R	58.3	82.5	117	130	143	V
	857m	1.21	1.71	1.92	2.10	A
82R	64.0	90.6	128	143	157	V
	781m	1.10	1.56	1.75	1.91	A
100R	70.7	100	141	158	173	V
	707m	1.00	1.41	1.58	1.73	A
120R	77.5	110	155	173	190	V
	645m	913m	1.29	1.44	1.58	A
150R	86.6	122	173	194	212	V
	577m	816m	1.15	1.29	1.41	A
180R	94.9	134	190	212	232	V
	527m	745m	1.05	1.18	1.29	A
220R	105	148	210	235	257	V
	477m	674m	953m	1.07	1.17	A
270R	116	164	232	260	285	V
	430m	609m	861m	962m	1.05	A
330R	128	182	257	287	315	V
	389m	550m	778m	870m	953m	A
390R	140	197	279	312	342	V
	358m	506m	716m	801m	877m	A
470R	153	217	307	343	375	V
	326m	461m	652m	729m	799m	A

R	50 W	100 W	200 W	250 W	300 W	
560R	167	237	335	374	410	V
	299m	423m	598m	668m	732m	A
680R	184	261	369	412	452	V
	271m	383m	542m	606m	664m	A
820R	202	286	405	453	496	V
	247m	349m	494m	552m	605m	A
1k	224	316	447	500	548	V
	224m	316m	447m	500m	548m	A
1k2	245	346	490	548	600	V
	204m	289m	408m	456m	500m	A
1k5	274	387	548	612	671	V
	183m	258m	365m	408m	447m	A
1k8	300	424	600	671	735	V
	167m	236m	333m	373m	408m	A
2k2	332	469	663	742	812	V
	151m	213m	302m	337m	369m	A
2k7	367	520	735	822	900	V
	136m	192m	272m	304m	333m	A
3k3	406	574	812	908	995	V
	123m	174m	246m	275m	302m	A
3k9	442	624	883	987	1.08k	V
	113m	160m	226m	253m	277m	A
4k7	485	686	970	1.08k	1.19k	V
	103m	146m	206m	231m	253m	A
5k6	529	748	1.06k	1.18k	1.30k	V
	94.5m	134m	189m	211m	231m	A
6k8	583	825	1.17k	1.30k	1.43k	V
	85.7m	121m	171m	192m	210m	A
8k2	640	906	1.28k	1.43k	1.57k	V
	78.1m	110m	156m	175m	191m	A
10k	707	1.00k	1.41k	1.58k	1.73k	V
	70.7m	100m	141m	158m	173m	A
12k	775	1.10k	1.55k	1.73k	1.90k	V
	64.5m	91.3m	129m	144m	158m	A
15k	866	1.22k	1.73k	1.94k	2.12k	V
	57.7m	81.6m	115m	129m	141m	A

R	50 W	100 W	200 W	250 W	300 W	
18k	949	1.34k	1.90k	2.12k	2.32k	V
	52.7m	74.5m	105m	118m	129m	A
22k	1.05k	1.48k	2.10k	2.35k	2.57k	V
	47.7m	67.4m	95.3m	107m	117m	A
27k	1.16k	1.64k	2.32k	2.60k	2.85k	V
	43.0m	60.9m	86.1m	96.2m	105m	A
33k	1.28k	1.82k	2.57k	2.87k	3.15k	V
	38.9m	55.0m	77.8m	87.0m	95.3m	A
39k	1.40k	1.97k	2.79k	3.12k	3.42k	V
	35.8m	50.6m	71.6m	80.1m	87.7m	A
47k	1.53k	2.17k	3.07k	3.43k	3.75k	V
	32.6m	46.1m	65.2m	72.9m	79.9m	A
56k	1.67k	2.37k	3.35k	3.74k	4.10k	V
	29.9m	42.3m	59.8m	66.8m	73.2m	A
68k	1.84k	2.61k	3.69k	4.12k	4.52k	V
	27.1m	38.3m	54.2m	60.6m	66.4m	A
82k	2.02k	2.86k	4.05k	4.53k	4.96k	V
	24.7m	34.9m	49.4m	55.2m	60.5m	A
100k	2.24k	3.16k	4.47k	5.00k	5.48k	V
	22.4m	31.6m	44.7m	50.0m	54.8m	A

8 Voltage dividers

Introduction

Figure 8.1 shows a voltage divider. It is used everywhere in electronics to reduce the level of AC and/or DC voltages. The attenuation, input resistance and output resistance are as follows:

$$A = \frac{R_2}{R_1 + R_2} \tag{8.1}$$

$$\mathrm{dB} = -20\,\log_{10}(A) \tag{8.2}$$

$$R_I = R_1 + R_2 \tag{8.3}$$

$$R_O = \frac{R_1 R_2}{R_1 + R_2} \tag{8.4}$$

All of these parameters are given in Table 8.1 for various combinations of E24 resistor, and sorted in order of increasing attenuation. The Table starts at 1 dB and ends at 60 dB.

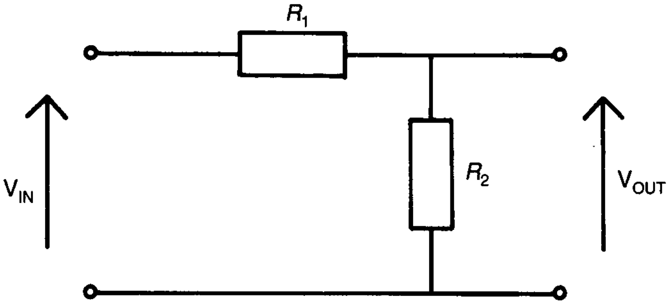

Figure 8.1 A voltage divider

The equations assume that the circuit is driven by a perfect voltage source, i.e. a circuit with zero output impedance, and drives a circuit of infinite input impedance. If the source is the output pin of an op-amp, and the load a non-inverting op-amp circuit, this is as true as makes no difference. If source resistance is a tenth of R_I, or load impedance ten times R_O, then the error is just under a dB; if the ratios are 100:1, then it is less than 0.1 dB.

Using Table 8.1 in cases where source and load impedances can be ignored

We can simply look up the desired attenuation in column A or dB as appropriate, and choose a pair.

Example 8.1: A voltage divider of 4 dB could be formed by resistors $R_1, R_2 = 3k6, 6k2$ and 3k, 5k1 respectively, amongst many.

When source and load impedances cannot be ignored

Selecting values for a voltage divider with ideal source and load impedances is pretty straightforward, as we saw above, but what happens when source and load begin affecting the nominal attenuation noticeably? Let us consider what happens.

We'll call source impedance R_S, input and output impedances of the attenuator R_I and R_O, as before, and load impedance R_L. Ideally, we want R_I to be much greater than R_S, and R_L much greater than R_O. This implies that (as R_O cannot be greater than R_I) R_L should be much greater than R_S.

(Of course, the situation can arise where the loss due to R_S and R_L is actually greater than the desired attenuation. If we are in doubt, we can check this by calculating a value for $\frac{R_L}{(R_S+R_L)}$; if this gives a greater attenuation than that desired, we need a buffer amplifier, rather than an attenuator.)

In Table 8.1, the columns R_{SMAX} and R_{LMIN} give suggested values for, as their names suggest, a maximum value of R_S and a minimum of R_L. This is calculated so that, with both in place, they cause the circuit to have an attenuation of 1 dB more than the value given in the dB column. This is true so long as R_S is still less than or equal to R_{SMAX}, and R_L greater than or equal to R_{LMIN}. The column R_{LS}(1 dB) is the ratio of them.

(If we wish, we can scale both R_1 and R_2 by factors of 10 in the same sense. Then R_I, R_O, R_{SMAX} and R_{LMIN} will also be scaled. A, dB and R_{LS}(1 dB) will be unaffected.)

If we set out to design an attenuator to work in a circuit where R_S and R_L are significant, we can start out by calculating their ratio, $\frac{R_L}{R_S}$. I'm going to call this R_{LS} (real). Two possible situations now arise:

1. If R_{LS} (real) is greater than or equal to R_{LS}(1 dB) for the desired value of dB in

Table 8.1, the attenuation with R_L and R_S in place should be within 1 dB of nominal. (Provided that we can find a pair where $R_S \le R_{SMAX}$ and $R_L \ge R_{LMIN}$.)

2. If R_{LS} (real) is less than R_{LS}(1 dB) the error will be more than 1 dB. In this situation the problem could be approached by other methods, or the values in the table used as a starting point.

Having selected a pair of values, we can check them numerically, if we are concerned about the accuracy. First we calculate the parallel resistance of R_L and R_2:

$$R_P = \frac{R_L R_2}{R_L + R_2} \tag{8.5}$$

Then:

$$\text{attenuation} = \frac{R_P}{R_S + R_1 + R_P} \tag{8.6}$$

The best way to see how things work is to look at some examples.

Example 8.2: Form a 10 dB attenuator to go between a source of 100 Ω and a load of 10k.

Our ratio of load to source is 100. We look down the dB column at values around 10 dB. R_{LS}(1 dB) is given as 63 or so here, so our error is less than 1 dB. Now look through the available pairs, to find one where we will be able to make R_S and R_L in range. We can see that we will need to divide by 10 to make $R_{SMAX} \ge 100\,\Omega$ and $R_{LMIN} \le$ 10k. We might choose:

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS} (1 dB)
12k	5k6	318m	09.95	17.6k	3.82k	1.03k	65.1k	63.1

This would yield $R_{SMAX} = 103\,\Omega$, $R_{LMIN} =$ 6.51k. We conclude that the error will be around 0.7 dB. (R_S is about equal to R_{SMAX}, so say 0.5 dB, R_L is about 1.5 times R_{LMIN}, we could guess at 0.2 dB. Or we might calculate the actual attenuation using Eqns(3.5) and (3.6). So we aim up the table to lines where dB = 9.3 or thereabouts. We find (again selecting for $R_{SMAX} \ge 100\,\Omega$ and $R_{LMIN} \le$ 10k):

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS} (1 dB)
13k	6k8	343m	09.28	19.8k	4.46k	1.16k	76.1k	65.6

and decide to try $R_1 =$ 1k3, $R_2 = 680\,\Omega$, as $R_{SMAX} = 116\,\Omega$, $R_{LMIN} =$ 7.6k then. To check, we work backwards and calculate that $R_P = 680\,\Omega$ in parallel with 10k, or 634 Ω, and we get:

$$\text{attenuation} = \frac{634}{100 + 1300 + 634} = 0.312 \quad \text{or} \quad 10.1\,\text{dB}$$

which is pretty good.

Here's another example, where the values of R_L and R_S make things tougher.

Example 8.3: An attenuator of 3 dB, with a source of 400 Ω, and a load of 3k.

Here we see that our ratio is a lot lower, $\frac{3000}{400} = 7.5$. If we go to the part of the table where dB = 3, we see that the R_{LS}(1 dB) column gives figures of about 67, telling us that we will lose a lot more than 1 dB due to R_S and R_L. Obviously we need less nominal attenuation; we want the error due to R_S and R_L and the nominal attenuation of the attenuator chosen to total 3 dB.

A possible solution is to use the load resistance itself as R_2. The load is 3k; looking up the table, we see:

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
1k2	3k	714m	02.92	4.20k	857	235	15.3k	65.4

which is, by good luck, ideal. R_1 would ideally be 1k2, but we already have a source of 400 Ω, so we want the nearest value to 800 Ω, which is 820 Ω. If we wanted to check, we would calculate $\frac{3000}{3000+400+820} = 0.711 = 2.96$ dB, very close.

Example 8.4: R_S = 1k, R_L = 5k and the attenuation desired is 8 dB.

Here our ratio is again low, 5. Our loss will be too great again, so we look for values of R_2 near to 5k. We find:

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS} (1 dB)
7k5	5k1	405m	07.86	12.6k	3.04k	732	52.3k	71.5

suggesting that we need a series resistor between source and load of (7k5 − 1k), or 6k5. The nearest preferred value is 6k8, giving an attenuation of $\frac{5}{5+6.8+1} = 0.391$ or 8.16 dB, very close.

Note that in cases where the load and source resistances form a part of the attenuator itself, we should check that they have a suitable power rating.

Table 8.1 *Voltage dividers*

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
6k8	56k	892m	01.00	62.8k	6.06k	3.13k	122k	38.9
6k2	51k	892m	01.00	57.2k	5.53k	2.85k	111k	38.9
1k	8k2	891m	01.00	9.20k	891	459	17.9k	39.0
10k	82k	891m	01.00	92.0k	8.91k	4.59k	179k	39.0
2k2	18k	891m	01.00	20.2k	1.96k	1.01k	39.3k	39.1
3k3	27k	891m	01.00	30.3k	2.94k	1.51k	59.0k	39.1
2k7	22k	891m	01.01	24.7k	2.40k	1.23k	48.2k	39.2
1k6	13k	890m	01.01	14.6k	1.42k	728	28.6k	39.3
2k	16k	889m	01.02	18.0k	1.78k	895	35.7k	39.9
1k5	12k	889m	01.02	13.5k	1.33k	672	26.8k	39.9
3k	24k	889m	01.02	27.0k	2.67k	1.34k	53.6k	39.9
1k3	10k	885m	01.06	11.3k	1.15k	568	22.9k	40.3
3k9	30k	885m	01.06	33.9k	3.45k	1.70k	68.7k	40.3
5k6	43k	885m	01.06	48.6k	4.95k	2.44k	98.6k	40.3
4k3	33k	885m	01.06	37.3k	3.80k	1.87k	75.7k	40.4
4k7	36k	885m	01.07	40.7k	4.16k	2.05k	82.7k	40.4
5k1	39k	884m	01.07	44.1k	4.51k	2.22k	89.7k	40.5
1k2	9k1	883m	01.08	10.3k	1.06k	517	21.1k	40.9
6k2	47k	883m	01.08	53.2k	5.48k	2.67k	109k	40.9
8k2	62k	883m	01.08	70.2k	7.24k	3.52k	144k	41.0
1k	7k5	882m	01.09	8.50k	882	426	17.6k	41.2
1k6	12k	882m	01.09	13.6k	1.41k	682	28.1k	41.2
2k	15k	882m	01.09	17.0k	1.76k	853	35.2k	41.2
2k4	18k	882m	01.09	20.4k	2.12k	1.02k	42.2k	41.2
3k6	27k	882m	01.09	30.6k	3.18k	1.54k	63.3k	41.2
6k8	51k	882m	01.09	57.8k	6.00k	2.90k	120k	41.2
10k	75k	882m	01.09	85.0k	8.82k	4.26k	176k	41.2
9k1	68k	882m	01.09	77.1k	8.03k	3.87k	160k	41.3
7k5	56k	882m	01.09	63.5k	6.61k	3.19k	132k	41.4
1k1	8k2	882m	01.09	9.30k	970	467	19.3k	41.4
2k7	20k	881m	01.10	22.7k	2.38k	1.14k	47.4k	41.6
1k5	11k	880m	01.11	12.5k	1.32k	627	26.3k	41.9
3k	22k	880m	01.11	25.0k	2.64k	1.25k	52.6k	41.9
2k2	16k	879m	01.12	18.2k	1.93k	923	38.1k	41.3
3k3	24k	879m	01.12	27.3k	2.90k	1.38k	57.2k	41.3
1k8	13k	878m	01.13	14.8k	1.58k	749	31.2k	41.7

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
5k1	36k	876m	01.15	41.1k	4.47k	2.08k	88.3k	42.5
4k7	33k	875m	01.16	37.7k	4.11k	1.91k	81.3k	42.6
1k3	9k1	875m	01.16	10.4k	1.14k	526	22.5k	42.7
4k3	30k	875m	01.16	34.3k	3.76k	1.74k	74.3k	42.8
5k6	39k	874m	01.17	44.6k	4.90k	2.26k	96.7k	42.9
6k2	43k	874m	01.17	49.2k	5.42k	2.49k	107k	43.0
3k9	27k	874m	01.17	30.9k	3.41k	1.56k	67.3k	43.0
6k8	47k	874m	01.17	53.8k	5.94k	2.72k	117k	43.1
1k6	11k	873m	01.18	12.6k	1.40k	643	27.4k	42.6
1k2	8k2	872m	01.19	9.40k	1.05k	480	20.5k	42.7
8k2	56k	872m	01.19	64.2k	7.15k	3.28k	140k	42.8
1k1	7k5	872m	01.19	8.60k	959	439	18.8k	42.8
2k2	15k	872m	01.19	17.2k	1.92k	878	37.6k	42.8
9k1	62k	872m	01.19	71.1k	7.94k	3.63k	155k	42.8
1k	6k8	872m	01.19	7.80k	872	398	17.1k	42.9
7k5	51k	872m	01.19	58.5k	6.54k	2.99k	128k	42.9
10k	68k	872m	01.19	78.0k	8.72k	3.98k	171k	42.9
2k4	16k	870m	01.21	18.4k	2.09k	939	40.9k	43.5
1k5	10k	870m	01.21	11.5k	1.30k	587	25.6k	43.5
1k8	12k	870m	01.21	13.8k	1.57k	704	30.7k	43.5
2k7	18k	870m	01.21	20.7k	2.35k	1.06k	46.0k	43.5
3k	20k	870m	01.21	23.0k	2.61k	1.17k	51.1k	43.5
3k3	22k	870m	01.21	25.3k	2.87k	1.29k	56.2k	43.5
3k6	24k	870m	01.21	27.6k	3.13k	1.41k	61.3k	43.5
2k	13k	867m	01.24	15.0k	1.73k	764	34.0k	44.5
5k1	33k	866m	01.25	38.1k	4.42k	1.94k	86.6k	44.6
5k6	36k	865m	01.26	41.6k	4.85k	2.14k	94.1k	43.9
4k7	30k	865m	01.26	34.7k	4.06k	1.79k	78.9k	44.2
6k8	43k	863m	01.28	49.8k	5.87k	2.56k	114k	44.5
1k3	8k2	863m	01.28	9.50k	1.12k	489	21.8k	44.6
6k2	39k	863m	01.28	45.2k	5.35k	2.33k	104k	44.6
4k3	27k	863m	01.28	31.3k	3.71k	1.61k	72.0k	44.7
7k5	47k	862m	01.29	54.5k	6.47k	2.81k	126k	44.8
1k2	7k5	862m	01.29	8.70k	1.03k	448	20.1k	44.9
2k4	15k	862m	01.29	17.4k	2.07k	896	40.2k	44.9
1k6	10k	862m	01.29	11.6k	1.38k	597	26.8k	44.9
8k2	51k	861m	01.30	59.2k	7.06k	3.05k	137k	45.0

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
1k	6k2	861m	01.30	7.20k	861	371	16.7k	45.1
10k	62k	861m	01.30	72.0k	8.61k	3.71k	167k	45.1
1k1	6k8	861m	01.30	7.90k	947	407	18.4k	45.2
3k9	24k	860m	01.31	27.9k	3.35k	1.44k	65.2k	45.4
9k1	56k	860m	01.31	65.1k	7.83k	3.35k	152k	45.4
1k8	11k	859m	01.32	12.8k	1.55k	658	30.1k	45.7
3k6	22k	859m	01.32	25.6k	3.09k	1.32k	60.2k	45.7
1k5	9k1	858m	01.33	10.6k	1.29k	545	25.0k	46.0
3k3	20k	858m	01.33	23.3k	2.83k	1.20k	55.1k	46.0
2k	12k	857m	01.34	14.0k	1.71k	720	33.3k	46.3
3k	18k	857m	01.34	21.0k	2.57k	1.08k	50.0k	46.3
2k7	16k	856m	01.35	18.7k	2.31k	971	44.5k	45.8
2k2	13k	855m	01.36	15.2k	1.88k	790	36.2k	45.9
5k6	33k	855m	01.36	38.6k	4.79k	2.00k	92.2k	46.0
5k1	30k	855m	01.36	35.1k	4.36k	1.82k	83.9k	46.0
6k2	36k	853m	01.38	42.2k	5.29k	2.19k	102k	46.5
1k3	7k5	852m	01.39	8.80k	1.11k	457	21.3k	46.7
4k7	27k	852m	01.39	31.7k	4.00k	1.64k	77.2k	47.0
6k8	39k	852m	01.40	45.8k	5.79k	2.38k	112k	47.0
7k5	43k	851m	01.40	50.5k	6.39k	2.62k	123k	47.0
8k2	47k	851m	01.40	55.2k	6.98k	2.86k	135k	47.0
1k6	9k1	850m	01.41	10.7k	1.36k	555	26.2k	47.3
1k2	6k8	850m	01.41	8.00k	1.02k	415	19.7k	47.4
3k9	22k	849m	01.42	25.9k	3.31k	1.34k	63.9k	47.6
1k1	6k2	849m	01.42	7.30k	934	379	18.0k	47.6
9k1	51k	849m	01.43	60.1k	7.72k	3.12k	149k	47.8
1k	5k6	848m	01.43	6.60k	848	342	16.4k	47.8
10k	56k	848m	01.43	66.0k	8.48k	3.42k	164k	47.8
4k3	24k	848m	01.43	28.3k	3.65k	1.47k	70.3k	47.9
1k8	10k	847m	01.44	11.8k	1.53k	612	29.4k	48.0
2k7	15k	847m	01.44	17.7k	2.29k	918	44.1k	48.0
3k6	20k	847m	01.44	23.6k	3.05k	1.22k	58.8k	48.0
2k	11k	846m	01.45	13.0k	1.69k	681	32.3k	47.4
1k5	8k2	845m	01.46	9.70k	1.27k	508	24.2k	47.6
2k2	12k	845m	01.46	14.2k	1.86k	744	35.5k	47.7
3k3	18k	845m	01.46	21.3k	2.79k	1.12k	53.2k	47.7
2k4	13k	844m	01.47	15.4k	2.03k	807	38.7k	47.9

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
5k6	30k	843m	01.49	35.6k	4.72k	1.86k	90.2k	48.4
3k	16k	842m	01.49	19.0k	2.53k	994	48.3k	48.6
6k2	33k	842m	01.50	39.2k	5.22k	2.05k	99.7k	48.6
5k1	27k	841m	01.50	32.1k	4.29k	1.68k	82.0k	48.8
6k8	36k	841m	01.50	42.8k	5.72k	2.24k	109k	48.8
8k2	43k	840m	01.52	51.2k	6.89k	2.68k	132k	49.1
1k3	6k8	840m	01.52	8.10k	1.09k	424	20.9k	49.2
7k5	39k	839m	01.53	46.5k	6.29k	2.43k	120k	49.4
1k2	6k2	838m	01.54	7.40k	1.01k	387	19.2k	49.6
9k1	47k	838m	01.54	56.1k	7.62k	2.94k	146k	49.6
3k9	20k	837m	01.55	23.9k	3.26k	1.25k	62.4k	49.8
1k6	8k2	837m	01.55	9.80k	1.34k	513	25.6k	49.9
4k3	22k	837m	01.55	26.3k	3.60k	1.38k	68.7k	49.9
4k7	24k	836m	01.55	28.7k	3.93k	1.50k	75.1k	50.0
1k	5k1	836m	01.56	6.10k	836	319	16.0k	50.0
10k	51k	836m	01.56	61.0k	8.36k	3.19k	160k	50.0
1k1	5k6	836m	01.56	6.70k	919	351	17.6k	50.1
1k8	9k1	835m	01.57	10.9k	1.50k	571	28.7k	50.2
1k5	7k5	833m	01.58	9.00k	1.25k	475	23.7k	49.8
2k	10k	833m	01.58	12.0k	1.67k	634	31.6k	49.8
2k2	11k	833m	01.58	13.2k	1.83k	697	34.7k	49.8
2k4	12k	833m	01.58	14.4k	2.00k	760	37.9k	49.8
3k	15k	833m	01.58	18.0k	2.50k	950	47.3k	49.8
3k6	18k	833m	01.58	21.6k	3.00k	1.14k	56.8k	49.8
6k8	33k	829m	01.63	39.8k	5.64k	2.10k	107k	50.8
3k3	16k	829m	01.63	19.3k	2.74k	1.02k	51.8k	50.8
6k2	30k	829m	01.63	36.2k	5.14k	1.91k	97.3k	50.9
5k6	27k	828m	01.64	32.6k	4.64k	1.72k	87.8k	51.0
2k7	13k	828m	01.64	15.7k	2.24k	829	42.3k	51.1
7k5	36k	828m	01.64	43.5k	6.21k	2.30k	118k	51.2
1k3	6k2	827m	01.65	7.50k	1.07k	396	20.4k	51.4
8k2	39k	826m	01.66	47.2k	6.78k	2.49k	128k	51.5
9k1	43k	825m	01.67	52.1k	7.51k	2.75k	142k	51.7
5k1	24k	825m	01.67	29.1k	4.21k	1.54k	79.7k	51.8
1k	4k7	825m	01.68	5.70k	825	301	15.6k	51.9
10k	47k	825m	01.68	57.0k	8.25k	3.01k	156k	51.9
1k6	7k5	824m	01.68	9.10k	1.32k	481	25.0k	52.0

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
4k7	22k	824m	01.68	26.7k	3.87k	1.41k	73.3k	52.0
1k2	5k6	824m	01.69	6.80k	988	359	18.7k	52.1
4k3	20k	823m	01.69	24.3k	3.54k	1.28k	67.0k	52.2
1k1	5k1	823m	01.70	6.20k	905	328	17.1k	52.3
3k9	18k	822m	01.70	21.9k	3.21k	1.17k	60.2k	51.5
2k4	11k	821m	01.71	13.4k	1.97k	714	37.0k	51.8
1k8	8k2	820m	01.72	10.0k	1.48k	533	27.7k	52.0
2k	9k1	820m	01.73	11.1k	1.64k	591	30.8k	52.0
3k3	15k	820m	01.73	18.3k	2.70k	975	50.8k	52.1
2k2	10k	820m	01.73	12.2k	1.80k	650	33.8k	52.1
1k5	6k8	819m	01.73	8.30k	1.23k	442	23.1k	52.1
2k7	12k	816m	01.76	14.7k	2.20k	783	41.4k	52.8
3k6	16k	816m	01.76	19.6k	2.94k	1.04k	55.1k	52.8
6k8	30k	815m	01.77	36.8k	5.54k	1.96k	104k	53.0
7k5	33k	815m	01.78	40.5k	6.11k	2.16k	115k	53.1
8k2	36k	814m	01.78	44.2k	6.68k	2.36k	125k	53.2
6k2	27k	813m	01.80	33.2k	5.04k	1.77k	94.6k	53.5
3k	13k	813m	01.80	16.0k	2.44k	853	45.7k	53.7
5k1	22k	812m	01.81	27.1k	4.14k	1.44k	77.7k	53.8
1k3	5k6	812m	01.81	6.90k	1.06k	368	19.8k	53.8
1k	4k3	811m	01.82	5.30k	811	282	15.2k	53.9
10k	43k	811m	01.82	53.0k	8.11k	2.82k	152k	53.9
5k6	24k	811m	01.82	29.6k	4.54k	1.58k	85.2k	54.0
9k1	39k	811m	01.82	48.1k	7.38k	2.56k	138k	54.0
1k1	4k7	810m	01.83	5.80k	891	309	16.7k	54.1
4k7	20k	810m	01.83	24.7k	3.81k	1.32k	71.5k	54.4
1k6	6k8	810m	01.84	8.40k	1.30k	447	24.3k	54.4
1k2	5k1	810m	01.84	6.30k	971	335	18.2k	54.4
4k3	18k	807m	01.86	22.3k	3.47k	1.20k	64.5k	53.8
1k8	7k5	806m	01.87	9.30k	1.45k	500	27.0k	54.0
2k4	10k	806m	01.87	12.4k	1.94k	667	36.0k	54.0
3k6	15k	806m	01.87	18.6k	2.90k	1.00k	54.0k	54.0
2k2	9k1	805m	01.88	11.3k	1.77k	608	32.9k	54.2
1k5	6k2	805m	01.88	7.70k	1.21k	414	22.5k	54.2
3k9	16k	804m	01.89	19.9k	3.14k	1.07k	58.3k	54.5
2k	8k2	804m	01.90	10.2k	1.61k	549	29.9k	54.5
2k7	11k	803m	01.91	13.7k	2.17k	737	40.3k	54.7

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
8k2	33k	801m	01.93	41.2k	6.57k	2.22k	122k	55.1
7k5	30k	800m	01.94	37.5k	6.00k	2.02k	112k	55.3
3k	12k	800m	01.94	15.0k	2.40k	807	44.6k	55.3
6k8	27k	799m	01.95	33.8k	5.43k	1.82k	101k	55.6
9k1	36k	798m	01.96	45.1k	7.26k	2.43k	135k	55.7
3k3	13k	798m	01.96	16.3k	2.63k	877	48.9k	55.8
5k6	22k	797m	01.97	27.6k	4.46k	1.48k	83.0k	55.9
1k3	5k1	797m	01.97	6.40k	1.04k	344	19.3k	56.0
5k1	20k	797m	01.97	25.1k	4.06k	1.35k	75.6k	56.0
1k2	4k7	797m	01.98	5.90k	956	317	17.8k	56.0
1k1	4k3	796m	01.98	5.40k	876	290	16.3k	56.1
1k	3k9	796m	01.98	4.90k	796	264	14.8k	56.2
10k	39k	796m	01.98	49.0k	7.96k	2.64k	148k	56.2
1k6	6k2	795m	01.99	7.80k	1.27k	419	23.7k	56.5
6k2	24k	795m	02.00	30.2k	4.93k	1.62k	91.7k	56.5
3k9	15k	794m	02.01	18.9k	3.10k	1.02k	57.6k	56.7
4k7	18k	793m	02.02	22.7k	3.73k	1.22k	69.4k	56.9
2k4	9k1	791m	02.03	11.5k	1.90k	618	35.3k	57.1
1k8	6k8	791m	02.04	8.60k	1.42k	467	26.2k	56.2
2k	7k5	789m	02.05	9.50k	1.58k	516	29.1k	56.4
1k5	5k6	789m	02.06	7.10k	1.18k	385	21.8k	56.6
2k2	8k2	788m	02.06	10.4k	1.73k	565	32.0k	56.6
4k3	16k	788m	02.07	20.3k	3.39k	1.10k	62.4k	56.7
2k7	10k	787m	02.08	12.7k	2.13k	689	39.2k	56.8
3k	11k	786m	02.09	14.0k	2.36k	760	43.4k	57.1
8k2	30k	785m	02.10	38.2k	6.44k	2.07k	119k	57.2
3k3	12k	784m	02.11	15.3k	2.59k	831	47.7k	57.4
9k1	33k	784m	02.12	42.1k	7.13k	2.29k	131k	57.5
1k3	4k7	783m	02.12	6.00k	1.02k	326	18.8k	57.6
3k6	13k	783m	02.12	16.6k	2.82k	901	51.9k	57.6
1k	3k6	783m	02.13	4.60k	783	250	14.4k	57.7
7k5	27k	783m	02.13	34.5k	5.87k	1.87k	108k	57.7
10k	36k	783m	02.13	46.0k	7.83k	2.50k	144k	57.7
1k2	4k3	782m	02.14	5.50k	938	299	17.3k	57.9
5k6	20k	781m	02.14	25.6k	4.38k	1.39k	80.6k	58.0
6k2	22k	780m	02.16	28.2k	4.84k	1.53k	89.1k	58.2
1k1	3k9	780m	02.16	5.00k	858	271	15.8k	58.2

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
5k1	18k	779m	02.17	23.1k	3.97k	1.25k	73.2k	58.4
6k8	24k	779m	02.17	30.8k	5.30k	1.67k	97.6k	58.4
1k6	5k6	778m	02.18	7.20k	1.24k	390	22.9k	58.8
4k3	15k	777m	02.19	19.3k	3.34k	1.05k	61.6k	58.9
1k8	6k2	775m	02.21	8.00k	1.40k	434	25.7k	59.3
2k4	8k2	774m	02.23	10.6k	1.86k	575	34.2k	59.6
2k2	7k5	773m	02.23	9.70k	1.70k	526	31.4k	59.6
4k7	16k	773m	02.24	20.7k	3.63k	1.12k	67.0k	59.7
2k	6k8	773m	02.24	8.80k	1.55k	477	28.5k	59.7
1k5	5k1	773m	02.24	6.60k	1.16k	358	21.4k	59.7
2k7	9k1	771m	02.26	11.8k	2.08k	640	38.4k	59.9
3k9	13k	769m	02.28	16.9k	3.00k	926	54.8k	59.1
3k	10k	769m	02.28	13.0k	2.31k	712	42.1k	59.1
3k3	11k	769m	02.28	14.3k	2.54k	784	46.3k	59.1
3k6	12k	769m	02.28	15.6k	2.77k	855	50.5k	59.1
1k3	4k3	768m	02.29	5.60k	998	307	18.2k	59.4
1k	3k3	767m	02.30	4.30k	767	236	14.0k	59.4
10k	33k	767m	02.30	43.0k	7.67k	2.36k	140k	59.4
9k1	30k	767m	02.30	39.1k	6.98k	2.14k	127k	59.5
8k2	27k	767m	02.30	35.2k	6.29k	1.93k	115k	59.5
1k1	3k6	766m	02.32	4.70k	843	258	15.4k	59.7
1k2	3k9	765m	02.33	5.10k	918	279	16.7k	59.9
6k8	22k	764m	02.34	28.8k	5.19k	1.58k	94.8k	60.1
6k2	20k	763m	02.35	26.2k	4.73k	1.44k	86.4k	60.2
5k6	18k	763m	02.35	23.6k	4.27k	1.29k	78.0k	60.3
7k5	24k	762m	02.36	31.5k	5.71k	1.73k	104k	60.4
4k7	15k	761m	02.37	19.7k	3.58k	1.08k	65.3k	60.5
1k6	5k1	761m	02.37	6.70k	1.22k	367	22.2k	60.5
5k1	16k	758m	02.40	21.1k	3.87k	1.16k	70.6k	61.2
1k5	4k7	758m	02.41	6.20k	1.14k	339	20.8k	61.2
2k4	7k5	758m	02.41	9.90k	1.82k	542	33.2k	61.3
1k8	5k6	757m	02.42	7.40k	1.36k	405	24.9k	61.4
2k	6k2	756m	02.43	8.20k	1.51k	449	27.6k	61.5
2k2	6k8	756m	02.43	9.00k	1.66k	493	30.4k	61.6
3k9	12k	755m	02.44	15.9k	2.94k	870	53.8k	61.8
3k6	11k	753m	02.46	14.6k	2.71k	799	49.5k	62.0
2k7	8k2	752m	02.47	10.9k	2.03k	597	37.1k	62.2

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
3k	9k1	752m	02.47	12.1k	2.26k	662	41.2k	62.2
3k3	10k	752m	02.48	13.3k	2.48k	728	45.3k	62.2
4k3	13k	751m	02.48	17.3k	3.23k	947	59.0k	62.3
1k	3k	750m	02.50	4.00k	750	219	13.7k	62.6
1k1	3k3	750m	02.50	4.40k	825	241	15.1k	62.6
1k2	3k6	750m	02.50	4.80k	900	263	16.4k	62.6
1k3	3k9	750m	02.50	5.20k	975	285	17.8k	62.6
10k	30k	750m	02.50	40.0k	7.50k	2.19k	137k	62.6
9k1	27k	748m	02.52	36.1k	6.81k	1.98k	124k	62.9
5k1	15k	746m	02.54	20.1k	3.81k	1.10k	69.5k	63.1
6k8	20k	746m	02.54	26.8k	5.07k	1.47k	92.6k	63.1
1k6	4k7	746m	02.54	6.30k	1.19k	345	21.8k	63.1
7k5	22k	746m	02.55	29.5k	5.59k	1.63k	101k	62.0
8k2	24k	745m	02.55	32.2k	6.11k	1.78k	110k	62.0
6k2	18k	744m	02.57	24.2k	4.61k	1.34k	83.4k	62.3
1k5	4k3	741m	02.60	5.80k	1.11k	321	20.1k	62.7
5k6	16k	741m	02.61	21.6k	4.15k	1.19k	75.0k	62.8
2k4	6k8	739m	02.63	9.20k	1.77k	509	32.1k	63.0
1k8	5k1	739m	02.63	6.90k	1.33k	382	24.1k	63.0
3k9	11k	738m	02.64	14.9k	2.88k	824	52.1k	63.2
2k2	6k2	738m	02.64	8.40k	1.62k	465	29.4k	63.2
2k	5k6	737m	02.65	7.60k	1.47k	420	26.6k	63.4
4k3	12k	736m	02.66	16.3k	3.17k	902	57.2k	63.5
2k7	7k5	735m	02.67	10.2k	1.99k	564	35.9k	63.6
3k6	10k	735m	02.67	13.6k	2.65k	752	47.9k	63.6
1k3	3k6	735m	02.68	4.90k	955	271	17.3k	63.8
4k7	13k	734m	02.68	17.7k	3.45k	978	62.5k	63.8
3k3	9k1	734m	02.69	12.4k	2.42k	685	43.8k	63.9
1k2	3k3	733m	02.69	4.50k	880	249	15.9k	64.0
3k	8k2	732m	02.71	11.2k	2.20k	619	39.7k	64.2
1k1	3k	732m	02.71	4.10k	805	227	14.6k	64.3
10k	27k	730m	02.74	37.0k	7.30k	2.05k	132k	64.6
1k	2k7	730m	02.74	3.70k	730	205	13.2k	64.6
1k6	4k3	729m	02.75	5.90k	1.17k	326	21.1k	64.7
8k2	22k	728m	02.75	30.2k	5.97k	1.67k	108k	64.7
5k6	15k	728m	02.76	20.6k	4.08k	1.14k	73.8k	64.8
7k5	20k	727m	02.77	27.5k	5.45k	1.52k	98.7k	64.9

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
6k8	18k	726m	02.78	24.8k	4.94k	1.37k	89.3k	65.1
9k1	24k	725m	02.79	33.1k	6.60k	1.83k	119k	65.3
1k8	4k7	723m	02.82	6.50k	1.30k	359	23.5k	65.5
1k5	3k9	722m	02.83	5.40k	1.08k	298	19.6k	65.7
2k4	6k2	721m	02.84	8.60k	1.73k	475	31.3k	65.9
6k2	16k	721m	02.84	22.2k	4.47k	1.23k	80.8k	65.9
3k9	10k	719m	02.86	13.9k	2.81k	768	50.8k	66.1
4k3	11k	719m	02.87	15.3k	3.09k	846	55.9k	66.1
4k7	12k	719m	02.87	16.7k	3.38k	923	61.1k	66.2
2k	5k1	718m	02.87	7.10k	1.44k	392	26.0k	66.2
5k1	13k	718m	02.87	18.1k	3.66k	1.00k	66.3k	66.2
2k2	5k6	718m	02.88	7.80k	1.58k	431	28.6k	66.3
1k3	3k3	717m	02.88	4.60k	933	254	16.9k	66.3
3k6	9k1	717m	02.90	12.7k	2.58k	702	46.7k	66.4
2k7	6k8	716m	02.90	9.50k	1.93k	525	34.9k	66.5
1k2	3k	714m	02.92	4.20k	857	235	15.3k	65.4
3k	7k5	714m	02.92	10.5k	2.14k	586	38.4k	65.4
3k3	8k2	713m	02.94	11.5k	2.35k	642	42.1k	65.6
1k1	2k7	711m	02.97	3.80k	782	212	14.0k	65.9
8k2	20k	709m	02.98	28.2k	5.82k	1.57k	104k	66.1
1k6	3k9	709m	02.99	5.50k	1.13k	307	20.3k	66.1
6k2	15k	708m	03.00	21.2k	4.39k	1.18k	78.6k	66.3
9k1	22k	707m	03.01	31.1k	6.44k	1.74k	115k	66.4
1k5	3k6	706m	03.03	5.10k	1.06k	285	19.0k	66.7
10k	24k	706m	03.03	34.0k	7.06k	1.90k	126k	66.7
1k	2k4	706m	03.03	3.40k	706	190	12.6k	66.7
7k5	18k	706m	03.03	25.5k	5.29k	1.42k	94.9k	66.7
1k8	4k3	705m	03.04	6.10k	1.27k	340	22.7k	66.8
5k1	12k	702m	03.08	17.1k	3.58k	954	64.1k	67.2
6k8	16k	702m	03.08	22.8k	4.77k	1.27k	85.5k	67.2
2k	4k7	701m	03.08	6.70k	1.40k	374	25.1k	67.2
4k7	11k	701m	03.09	15.7k	3.29k	876	59.0k	67.3
2k4	5k6	700m	03.10	8.00k	1.68k	446	30.1k	67.4
3k9	9k1	700m	03.10	13.0k	2.73k	726	48.9k	67.4
4k3	10k	699m	03.11	14.3k	3.01k	798	53.9k	67.5
5k6	13k	699m	03.11	18.6k	3.91k	1.04k	70.1k	67.6
2k2	5k1	699m	03.12	7.30k	1.54k	407	27.5k	67.6

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
1k3	3k	698m	03.13	4.30k	907	240	16.3k	67.7
2k7	6k2	697m	03.14	8.90k	1.88k	497	33.7k	67.9
3k6	8k2	695m	03.16	11.8k	2.50k	659	44.8k	68.1
3k3	7k5	694m	03.17	10.8k	2.29k	603	41.1k	68.1
3k	6k8	694m	03.17	9.80k	2.08k	547	37.3k	68.2
1k6	3k6	692m	03.19	5.20k	1.11k	290	19.8k	68.4
1k2	2k7	692m	03.19	3.90k	831	218	14.9k	68.4
6k8	15k	688m	03.25	21.8k	4.68k	1.22k	83.8k	68.9
1k	2k2	688m	03.25	3.20k	688	179	12.3k	69.0
1k5	3k3	688m	03.25	4.80k	1.03k	268	18.5k	69.0
10k	22k	688m	03.25	32.0k	6.88k	1.79k	123k	69.0
9k1	20k	687m	03.26	29.1k	6.25k	1.62k	112k	69.0
8k2	18k	687m	03.26	26.2k	5.63k	1.46k	101k	69.0
1k1	2k4	686m	03.28	3.50k	754	195	13.5k	69.2
1k8	3k9	684m	03.30	5.70k	1.23k	318	22.1k	69.4
5k1	11k	683m	03.31	16.1k	3.48k	899	62.4k	69.5
2k	4k3	683m	03.32	6.30k	1.37k	352	24.5k	69.6
5k6	12k	682m	03.33	17.6k	3.82k	982	68.4k	69.7
2k2	4k7	681m	03.34	6.90k	1.50k	385	26.9k	69.7
7k5	16k	681m	03.34	23.5k	5.11k	1.31k	91.5k	69.8
4k7	10k	680m	03.35	14.7k	3.20k	821	57.3k	69.8
2k4	5k1	680m	03.35	7.50k	1.63k	419	29.2k	69.8
4k3	9k1	679m	03.36	13.4k	2.92k	748	52.3k	69.9
3k9	8k2	678m	03.38	12.1k	2.64k	682	46.9k	68.7
6k2	13k	677m	03.39	19.2k	4.20k	1.08k	74.4k	68.8
3k6	7k5	676m	03.41	11.1k	2.43k	626	43.1k	68.9
1k3	2k7	675m	03.41	4.00k	878	226	15.6k	69.0
2k7	5k6	675m	03.42	8.30k	1.82k	468	32.3k	69.0
3k	6k2	674m	03.43	9.20k	2.02k	519	35.9k	69.1
1k6	3k3	673m	03.43	4.90k	1.08k	276	19.1k	69.2
3k3	6k8	673m	03.44	10.1k	2.22k	570	39.4k	69.2
1k	2k	667m	03.52	3.00k	667	169	11.8k	70.0
1k5	3k	667m	03.52	4.50k	1.00k	254	17.7k	70.0
1k8	3k6	667m	03.52	5.40k	1.20k	304	21.3k	70.0
7k5	15k	667m	03.52	22.5k	5.00k	1.27k	88.7k	70.0
1k1	2k2	667m	03.52	3.30k	733	186	13.0k	70.0
1k2	2k4	667m	03.52	3.60k	800	203	14.2k	70.0

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
10k	20k	667m	03.52	30.0k	6.67k	1.69k	118k	70.0
9k1	18k	664m	03.55	27.1k	6.04k	1.53k	107k	70.2
5k6	11k	663m	03.57	16.6k	3.71k	935	65.8k	70.4
5k1	10k	662m	03.58	15.1k	3.38k	851	59.9k	70.4
2k4	4k7	662m	03.58	7.10k	1.59k	400	28.2k	70.5
2k2	4k3	662m	03.59	6.50k	1.46k	366	25.8k	70.5
8k2	16k	661m	03.59	24.2k	5.42k	1.36k	96.2k	70.5
2k	3k9	661m	03.60	5.90k	1.32k	332	23.5k	70.6
4k7	9k1	659m	03.62	13.8k	3.10k	778	55.0k	70.7
6k2	12k	659m	03.62	18.2k	4.09k	1.03k	72.5k	70.7
3k9	7k5	658m	03.64	11.4k	2.57k	642	45.5k	70.9
6k8	13k	657m	03.65	19.8k	4.46k	1.12k	79.2k	71.0
4k3	8k2	656m	03.66	12.5k	2.82k	704	50.1k	71.1
2k7	5k1	654m	03.69	7.80k	1.77k	440	31.3k	71.3
3k6	6k8	654m	03.69	10.4k	2.35k	586	41.8k	71.3
3k3	6k2	653m	03.71	9.50k	2.15k	535	38.2k	71.4
1k6	3k	652m	03.71	4.60k	1.04k	259	18.5k	71.4
3k	5k6	651m	03.73	8.60k	1.95k	485	34.7k	71.5
1k3	2k4	649m	03.76	3.70k	843	209	15.0k	71.8
1k8	3k3	647m	03.78	5.10k	1.16k	287	20.7k	71.9
1k2	2k2	647m	03.78	3.40k	776	192	13.8k	71.9
8k2	15k	647m	03.79	23.2k	5.30k	1.31k	94.1k	72.0
1k1	2k	645m	03.81	3.10k	710	175	12.6k	72.1
1k	1k8	643m	03.84	2.80k	643	158	11.4k	72.3
1k5	2k7	643m	03.84	4.20k	964	237	17.1k	72.3
2k	3k6	643m	03.84	5.60k	1.29k	316	22.8k	72.3
10k	18k	643m	03.84	28.0k	6.43k	1.58k	114k	72.3
2k4	4k3	642m	03.85	6.70k	1.54k	378	27.3k	72.4
5k6	10k	641m	03.86	15.6k	3.59k	879	63.7k	72.5
5k1	9k1	641m	03.86	14.2k	3.27k	800	58.0k	72.5
6k2	11k	640m	03.88	17.2k	3.97k	969	70.4k	72.6
2k2	3k9	639m	03.89	6.10k	1.41k	344	25.0k	72.6
6k8	12k	638m	03.90	18.8k	4.34k	1.06k	77.0k	72.7
9k1	16k	637m	03.91	25.1k	5.80k	1.41k	103k	72.8
4k7	8k2	636m	03.94	12.9k	2.99k	727	53.0k	72.9
4k3	7k5	636m	03.94	11.8k	2.73k	665	48.5k	72.9
3k9	6k8	636m	03.94	10.7k	2.48k	603	44.0k	72.9

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
2k7	4k7	635m	03.94	7.40k	1.71k	417	30.4k	73.0
7k5	13k	634m	03.96	20.5k	4.76k	1.16k	84.4k	73.1
3k6	6k2	633m	03.98	9.80k	2.28k	552	40.4k	73.1
3k	5k1	630m	04.02	8.10k	1.89k	461	33.2k	71.9
3k3	5k6	629m	04.02	8.90k	2.08k	507	36.5k	72.0
1k3	2k2	629m	04.03	3.50k	817	199	14.4k	72.0
1k6	2k7	628m	04.04	4.30k	1.00k	245	17.6k	72.1
1k2	2k	625m	04.08	3.20k	750	182	13.2k	72.3
1k8	3k	625m	04.08	4.80k	1.13k	273	19.8k	72.3
2k	3k3	623m	04.12	5.30k	1.25k	302	21.9k	72.5
9k1	15k	622m	04.12	24.1k	5.66k	1.37k	99.5k	72.6
1k1	1k8	621m	04.14	2.90k	683	165	12.0k	72.7
2k2	3k6	621m	04.14	5.80k	1.37k	330	24.0k	72.7
2k4	3k9	619m	04.17	6.30k	1.49k	359	26.1k	72.8
5k6	9k1	619m	04.17	14.7k	3.47k	837	60.9k	72.8
6k8	11k	618m	04.18	17.8k	4.20k	1.01k	73.8k	72.9
6k2	10k	617m	04.19	16.2k	3.83k	922	67.2k	72.9
5k1	8k2	617m	04.20	13.3k	3.14k	757	55.3k	73.0
1k	1k6	615m	04.22	2.60k	615	148	10.8k	73.1
7k5	12k	615m	04.22	19.5k	4.62k	1.11k	81.1k	73.1
10k	16k	615m	04.22	26.0k	6.15k	1.48k	108k	73.1
1k5	2k4	615m	04.22	3.90k	923	222	16.2k	73.1
4k7	7k5	615m	04.23	12.2k	2.89k	694	50.8k	73.1
2k7	4k3	614m	04.23	7.00k	1.66k	398	29.1k	73.2
3k9	6k2	614m	04.24	10.1k	2.39k	575	42.1k	73.2
8k2	13k	613m	04.25	21.2k	5.03k	1.21k	88.4k	73.2
4k3	6k8	613m	04.26	11.1k	2.63k	632	46.3k	73.3
3k	4k7	610m	04.29	7.70k	1.83k	438	32.2k	73.4
3k6	5k6	609m	04.31	9.20k	2.19k	524	38.5k	73.5
3k3	5k1	607m	04.33	8.40k	2.00k	478	35.2k	73.6
1k3	2k	606m	04.35	3.30k	788	188	13.8k	73.7
1k	1k5	600m	04.44	2.50k	600	142	10.5k	74.1
1k2	1k8	600m	04.44	3.00k	720	171	12.7k	74.1
1k8	2k7	600m	04.44	4.50k	1.08k	256	19.0k	74.1
2k	3k	600m	04.44	5.00k	1.20k	285	21.1k	74.1
2k2	3k3	600m	04.44	5.50k	1.32k	313	23.2k	74.1
2k4	3k6	600m	04.44	6.00k	1.44k	341	25.3k	74.1

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
10k	15k	600m	04.44	25.0k	6.00k	1.42k	105k	74.1
1k6	2k4	600m	04.44	4.00k	960	228	16.9k	74.1
5k1	7k5	595m	04.51	12.6k	3.04k	717	53.3k	74.4
6k8	10k	595m	04.51	16.8k	4.05k	956	71.1k	74.4
6k2	9k1	595m	04.51	15.3k	3.69k	871	64.8k	74.4
1k5	2k2	595m	04.52	3.70k	892	211	15.7k	74.4
7k5	11k	595m	04.52	18.5k	4.46k	1.05k	78.4k	74.4
5k6	8k2	594m	04.52	13.8k	3.33k	785	58.5k	74.4
8k2	12k	594m	04.52	20.2k	4.87k	1.15k	85.6k	74.5
1k1	1k6	593m	04.54	2.70k	652	154	11.5k	74.5
4k7	6k8	591m	04.56	11.5k	2.78k	654	48.8k	74.6
2k7	3k9	591m	04.57	6.60k	1.60k	376	28.0k	74.6
4k3	6k2	590m	04.58	10.5k	2.54k	598	44.6k	74.7
3k9	5k6	589m	04.59	9.50k	2.30k	541	40.4k	74.7
3k	4k3	589m	04.60	7.30k	1.77k	415	31.1k	74.7
9k1	13k	588m	04.61	22.1k	5.35k	1.26k	94.1k	74.8
3k3	4k7	588m	04.62	8.00k	1.94k	455	34.1k	74.8
3k6	5k1	586m	04.64	8.70k	2.11k	495	37.1k	74.9
1k3	1k8	581m	04.72	3.10k	755	176	13.3k	75.2
1k6	2k2	579m	04.75	3.80k	926	216	16.3k	75.3
2k4	3k3	579m	04.75	5.70k	1.39k	324	24.4k	75.3
1k1	1k5	577m	04.78	2.60k	635	148	11.2k	75.4
2k2	3k	577m	04.78	5.20k	1.27k	296	22.3k	75.4
2k	2k7	574m	04.81	4.70k	1.15k	267	20.2k	75.5
8k2	11k	573m	04.84	19.2k	4.70k	1.09k	82.5k	75.5
5k6	7k5	573m	04.84	13.1k	3.21k	746	56.3k	75.6
6k8	9k1	572m	04.85	15.9k	3.89k	905	68.4k	75.6
1k2	1k6	571m	04.86	2.80k	686	159	12.0k	75.6
1k5	2k	571m	04.86	3.50k	857	199	15.1k	75.6
1k8	2k4	571m	04.86	4.20k	1.03k	239	18.1k	75.6
2k7	3k6	571m	04.86	6.30k	1.54k	359	27.1k	75.6
5k1	6k8	571m	04.86	11.9k	2.91k	677	51.2k	75.6
7k5	10k	571m	04.86	17.5k	4.29k	996	75.3k	75.6
6k2	8k2	569m	04.89	14.4k	3.53k	820	62.0k	75.7
4k7	6k2	569m	04.90	10.9k	2.67k	621	47.0k	75.7
9k1	12k	569m	04.90	21.1k	5.18k	1.20k	90.9k	75.7
3k9	5k1	567m	04.93	9.00k	2.21k	518	38.4k	74.3

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
3k6	4k7	566m	04.94	8.30k	2.04k	477	35.5k	74.3
3k3	4k3	566m	04.95	7.60k	1.87k	437	32.5k	74.3
4k3	5k6	566m	04.95	9.90k	2.43k	569	42.3k	74.3
1k	1k3	565m	04.96	2.30k	565	132	9.83k	74.3
3k	3k9	565m	04.96	6.90k	1.70k	397	29.5k	74.3
10k	13k	565m	04.96	23.0k	5.65k	1.32k	98.3k	74.3
1k2	1k5	556m	05.11	2.70k	667	155	11.6k	74.7
1k6	2k	556m	05.11	3.60k	889	207	15.5k	74.7
2k4	3k	556m	05.11	5.40k	1.33k	310	23.2k	74.7
1k3	1k6	552m	05.17	2.90k	717	167	12.5k	74.9
2k2	2k7	551m	05.18	4.90k	1.21k	282	21.1k	74.9
2k7	3k3	550m	05.19	6.00k	1.49k	345	25.8k	74.9
1k8	2k2	550m	05.19	4.00k	990	230	17.2k	74.9
8k2	10k	549m	05.20	18.2k	4.51k	1.05k	78.4k	74.9
5k1	6k2	549m	05.21	11.3k	2.80k	650	48.7k	75.0
5k6	6k8	548m	05.22	12.4k	3.07k	713	53.4k	75.0
7k5	9k1	548m	05.22	16.6k	4.11k	954	71.5k	75.0
6k2	7k5	547m	05.23	13.7k	3.39k	787	59.1k	75.0
9k1	11k	547m	05.24	20.1k	4.98k	1.16k	86.6k	75.0
6k8	8k2	547m	05.25	15.0k	3.72k	862	64.7k	75.0
3k9	4k7	547m	05.25	8.60k	2.13k	494	37.1k	75.0
1k	1k2	545m	05.26	2.20k	545	126	9.49k	75.0
1k5	1k8	545m	05.26	3.30k	818	190	14.2k	75.0
2k	2k4	545m	05.26	4.40k	1.09k	253	19.0k	75.0
3k	3k6	545m	05.26	6.60k	1.64k	379	28.5k	75.0
10k	12k	545m	05.26	22.0k	5.45k	1.26k	94.9k	75.0
3k6	4k3	544m	05.28	7.90k	1.96k	454	34.1k	75.1
4k7	5k6	544m	05.29	10.3k	2.56k	592	44.5k	75.1
4k3	5k1	543m	05.31	9.40k	2.33k	540	40.6k	75.1
1k1	1k3	542m	05.33	2.40k	596	138	10.4k	75.1
3k3	3k9	542m	05.33	7.20k	1.79k	414	31.1k	75.1
1k3	1k5	536m	05.42	2.80k	696	161	12.1k	75.3
1k6	1k8	529m	05.52	3.40k	847	195	14.7k	75.4
2k4	2k7	529m	05.52	5.10k	1.27k	293	22.1k	75.4
1k8	2k	526m	05.58	3.80k	947	218	16.5k	75.5
2k7	3k	526m	05.58	5.70k	1.42k	328	24.7k	75.5
8k2	9k1	526m	05.58	17.3k	4.31k	994	75.0k	75.5

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
5k6	6k2	525m	05.59	11.8k	2.94k	678	51.2k	75.5
6k8	7k5	524m	05.61	14.3k	3.57k	822	62.0k	75.5
3k9	4k3	524m	05.61	8.20k	2.05k	471	35.6k	75.5
1k	1k1	524m	05.62	2.10k	524	121	9.11k	75.5
2k	2k2	524m	05.62	4.20k	1.05k	241	18.2k	75.5
3k	3k3	524m	05.62	6.30k	1.57k	362	27.3k	75.5
10k	11k	524m	05.62	21.0k	5.24k	1.21k	91.1k	75.5
9k1	10k	524m	05.62	19.1k	4.76k	1.10k	82.9k	75.5
5k1	5k6	523m	05.62	10.7k	2.67k	615	46.4k	75.5
6k2	6k8	523m	05.63	13.0k	3.24k	747	56.4k	75.5
7k5	8k2	522m	05.64	15.7k	3.92k	902	68.1k	75.5
4k3	4k7	522m	05.64	9.00k	2.25k	517	39.1k	75.5
1k1	1k2	522m	05.65	2.30k	574	132	9.98k	75.5
2k2	2k4	522m	05.65	4.60k	1.15k	264	20.0k	75.5
3k3	3k6	522m	05.65	6.90k	1.72k	397	30.0k	75.5
4k7	5k1	520m	05.67	9.80k	2.45k	563	42.6k	75.5
1k2	1k3	520m	05.68	2.50k	624	144	10.9k	75.5
3k6	3k9	520m	05.68	7.50k	1.87k	431	32.6k	75.5
1k5	1k6	516m	05.74	3.10k	774	178	13.5k	75.6
1k	1k	500m	06.02	2.00k	500	115	8.70k	75.7
1k1	1k1	500m	06.02	2.20k	550	126	9.57k	75.7
1k2	1k2	500m	06.02	2.40k	600	138	10.4k	75.7
1k3	1k3	500m	06.02	2.60k	650	149	11.3k	75.7
1k5	1k5	500m	06.02	3.00k	750	172	13.0k	75.7
1k6	1k6	500m	06.02	3.20k	800	184	13.9k	75.7
1k8	1k8	500m	06.02	3.60k	900	207	15.7k	75.7
2k	2k	500m	06.02	4.00k	1.00k	230	17.4k	75.7
2k2	2k2	500m	06.02	4.40k	1.10k	253	19.1k	75.7
2k4	2k4	500m	06.02	4.80k	1.20k	276	20.9k	75.7
2k7	2k7	500m	06.02	5.40k	1.35k	310	23.5k	75.7
3k	3k	500m	06.02	6.00k	1.50k	345	26.1k	75.7
3k3	3k3	500m	06.02	6.60k	1.65k	379	28.7k	75.7
3k6	3k6	500m	06.02	7.20k	1.80k	414	31.3k	75.7
3k9	3k9	500m	06.02	7.80k	1.95k	448	33.9k	75.7
4k3	4k3	500m	06.02	8.60k	2.15k	494	37.4k	75.7
4k7	4k7	500m	06.02	9.40k	2.35k	540	40.9k	75.7
5k1	5k1	500m	06.02	10.2k	2.55k	586	44.4k	75.7

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
5k6	5k6	500m	06.02	11.2k	2.80k	644	48.7k	75.7
6k2	6k2	500m	06.02	12.4k	3.10k	713	53.9k	75.7
6k8	6k8	500m	06.02	13.6k	3.40k	782	59.2k	75.7
7k5	7k5	500m	06.02	15.0k	3.75k	862	65.2k	75.7
8k2	8k2	500m	06.02	16.4k	4.10k	943	71.3k	75.7
9k1	9k1	500m	06.02	18.2k	4.55k	1.05k	79.2k	75.7
10k	10k	500m	06.02	20.0k	5.00k	1.15k	87.0k	75.7
1k6	1k5	484m	06.31	3.10k	774	178	13.5k	75.6
1k3	1k2	480m	06.38	2.50k	624	145	10.7k	74.0
3k9	3k6	480m	06.38	7.50k	1.87k	436	32.2k	74.0
5k1	4k7	480m	06.38	9.80k	2.45k	569	42.1k	74.0
1k2	1k1	478m	06.41	2.30k	574	134	9.89k	74.0
2k4	2k2	478m	06.41	4.60k	1.15k	267	19.8k	74.0
3k6	3k3	478m	06.41	6.90k	1.72k	401	29.7k	74.0
4k7	4k3	478m	06.42	9.00k	2.25k	523	38.7k	74.0
8k2	7k5	478m	06.42	15.7k	3.92k	912	67.5k	74.0
6k8	6k2	477m	06.43	13.0k	3.24k	755	55.9k	74.0
5k6	5k1	477m	06.44	10.7k	2.67k	621	46.0k	74.0
10k	9k1	476m	06.44	19.1k	4.76k	1.11k	82.1k	74.0
1k1	1k	476m	06.44	2.10k	524	122	9.02k	74.0
2k2	2k	476m	06.44	4.20k	1.05k	244	18.0k	74.0
11k	10k	476m	06.44	21.0k	5.24k	1.22k	90.2k	74.0
3k3	3k	476m	06.44	6.30k	1.57k	366	27.1k	74.0
4k3	3k9	476m	06.45	8.20k	2.05k	476	35.2k	74.0
7k5	6k8	476m	06.46	14.3k	3.57k	830	61.4k	74.0
6k2	5k6	475m	06.47	11.8k	2.94k	685	50.7k	74.0
9k1	8k2	474m	06.48	17.3k	4.31k	1.00k	74.3k	74.0
2k	1k8	474m	06.49	3.80k	947	221	16.3k	74.0
3k	2k7	474m	06.49	5.70k	1.42k	331	24.5k	74.0
1k8	1k6	471m	06.55	3.40k	847	197	14.6k	73.9
2k7	2k4	471m	06.55	5.10k	1.27k	296	21.9k	73.9
1k5	1k3	464m	06.66	2.80k	696	163	12.0k	73.8
1k3	1k1	458m	06.78	2.40k	596	139	10.3k	73.7
3k9	3k3	458m	06.78	7.20k	1.79k	418	30.8k	73.7
5k1	4k3	457m	06.79	9.40k	2.33k	546	40.2k	73.6
5k6	4k7	456m	06.81	10.3k	2.56k	598	44.0k	73.6
4k3	3k6	456m	06.83	7.90k	1.96k	459	33.8k	73.6

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
1k2	1k	455m	06.85	2.20k	545	128	9.39k	73.6
2k4	2k	455m	06.85	4.40k	1.09k	255	18.8k	73.6
12k	10k	455m	06.85	22.0k	5.45k	1.28k	93.9k	73.6
1k8	1k5	455m	06.85	3.30k	818	192	14.1k	73.6
3k6	3k	455m	06.85	6.60k	1.64k	383	28.2k	73.6
4k7	3k9	453m	06.87	8.60k	2.13k	499	36.7k	73.5
8k2	6k8	453m	06.87	15.0k	3.72k	871	64.0k	73.5
11k	9k1	453m	06.88	20.1k	4.98k	1.17k	85.8k	73.5
7k5	6k2	453m	06.89	13.7k	3.39k	795	58.5k	73.5
9k1	7k5	452m	06.90	16.6k	4.11k	964	70.8k	73.5
6k8	5k6	452m	06.90	12.4k	3.07k	720	52.9k	73.5
6k2	5k1	451m	06.91	11.3k	2.80k	656	48.2k	73.5
10k	8k2	451m	06.93	18.2k	4.51k	1.06k	77.6k	73.4
2k2	1k8	450m	06.94	4.00k	990	232	17.1k	73.4
3k3	2k7	450m	06.94	6.00k	1.49k	348	25.6k	73.4
2k7	2k2	449m	06.96	4.90k	1.21k	284	20.9k	73.4
1k6	1k3	448m	06.97	2.90k	717	168	12.4k	73.4
1k5	1k2	444m	07.04	2.70k	667	157	11.5k	73.3
2k	1k6	444m	07.04	3.60k	889	209	15.3k	73.3
3k	2k4	444m	07.04	5.40k	1.33k	314	23.0k	73.3
1k3	1k	435m	07.23	2.30k	565	134	9.74k	72.9
13k	10k	435m	07.23	23.0k	5.65k	1.34k	97.4k	72.9
3k9	3k	435m	07.23	6.90k	1.70k	401	29.2k	72.9
5k6	4k3	434m	07.24	9.90k	2.43k	575	41.9k	72.9
4k3	3k3	434m	07.25	7.60k	1.87k	441	32.2k	72.9
4k7	3k6	434m	07.26	8.30k	2.04k	482	35.1k	72.9
5k1	3k9	433m	07.26	9.00k	2.21k	523	38.1k	72.8
12k	9k1	431m	07.30	21.1k	5.18k	1.23k	89.1k	72.8
6k2	4k7	431m	07.31	10.9k	2.67k	633	46.0k	72.8
8k2	6k2	431m	07.32	14.4k	3.53k	836	60.8k	72.7
2k4	1k8	429m	07.36	4.20k	1.03k	244	17.7k	72.7
10k	7k5	429m	07.36	17.5k	4.29k	1.02k	73.8k	72.7
1k6	1k2	429m	07.36	2.80k	686	163	11.8k	72.7
2k	1k5	429m	07.36	3.50k	857	203	14.8k	72.7
3k6	2k7	429m	07.36	6.30k	1.54k	366	26.6k	72.7
6k8	5k1	429m	07.36	11.9k	2.91k	691	50.2k	72.7
9k1	6k8	428m	07.38	15.9k	3.89k	923	67.0k	72.6

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
7k5	5k6	427m	07.38	13.1k	3.21k	761	55.2k	72.6
11k	8k2	427m	07.39	19.2k	4.70k	1.11k	80.9k	72.6
2k7	2k	426m	07.42	4.70k	1.15k	273	19.8k	72.5
1k5	1k1	423m	07.47	2.60k	635	151	10.9k	72.4
3k	2k2	423m	07.47	5.20k	1.27k	302	21.9k	72.4
2k2	1k6	421m	07.51	3.80k	926	221	16.0k	72.3
3k3	2k4	421m	07.51	5.70k	1.39k	331	23.9k	72.3
1k8	1k3	419m	07.55	3.10k	755	180	13.0k	72.2
5k1	3k6	414m	07.66	8.70k	2.11k	505	36.3k	72.0
4k7	3k3	413m	07.69	8.00k	1.94k	464	33.4k	71.9
13k	9k1	412m	07.71	22.1k	5.35k	1.28k	92.2k	71.9
4k3	3k	411m	07.72	7.30k	1.77k	424	30.4k	71.8
5k6	3k9	411m	07.73	9.50k	2.30k	552	39.6k	71.8
6k2	4k3	410m	07.75	10.5k	2.54k	610	43.7k	71.7
3k9	2k7	409m	07.76	6.60k	1.60k	383	27.5k	71.7
6k8	4k7	409m	07.77	11.5k	2.78k	668	47.9k	71.7
1k6	1k1	407m	07.80	2.70k	652	157	11.2k	71.6
12k	8k2	406m	07.83	20.2k	4.87k	1.17k	83.9k	71.5
8k2	5k6	406m	07.83	13.8k	3.33k	801	57.3k	71.5
11k	7k5	405m	07.84	18.5k	4.46k	1.07k	76.8k	71.5
2k2	1k5	405m	07.84	3.70k	892	215	15.4k	71.5
9k1	6k2	405m	07.85	15.3k	3.69k	888	63.5k	71.5
10k	6k8	405m	07.86	16.8k	4.05k	975	69.7k	71.5
7k5	5k1	405m	07.86	12.6k	3.04k	732	52.3k	71.5
15k	10k	400m	07.96	25.0k	6.00k	1.45k	103k	71.2
1k5	1k	400m	07.96	2.50k	600	145	10.3k	71.2
1k8	1k2	400m	07.96	3.00k	720	174	12.4k	71.2
2k4	1k6	400m	07.96	4.00k	960	232	16.5k	71.2
2k7	1k8	400m	07.96	4.50k	1.08k	261	18.6k	71.2
3k	2k	400m	07.96	5.00k	1.20k	290	20.7k	71.2
3k3	2k2	400m	07.96	5.50k	1.32k	319	22.7k	71.2
3k6	2k4	400m	07.96	6.00k	1.44k	348	24.8k	71.2
2k	1k3	394m	08.09	3.30k	788	192	13.6k	70.8
5k1	3k3	393m	08.12	8.40k	2.00k	488	34.5k	70.8
5k6	3k6	391m	08.15	9.20k	2.19k	534	37.7k	70.7
4k7	3k	390m	08.19	7.70k	1.83k	447	31.5k	70.6
6k8	4k3	387m	08.24	11.1k	2.63k	644	45.4k	70.4

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
13k	8k2	387m	08.25	21.2k	5.03k	1.23k	86.6k	70.4
6k2	3k9	386m	08.27	10.1k	2.39k	586	41.2k	70.3
4k3	2k7	386m	08.27	7.00k	1.66k	406	28.6k	70.3
7k5	4k7	385m	08.29	12.2k	2.89k	708	49.8k	70.3
12k	7k5	385m	08.30	19.5k	4.62k	1.13k	79.5k	70.2
1k6	1k	385m	08.30	2.60k	615	151	10.6k	70.2
16k	10k	385m	08.30	26.0k	6.15k	1.51k	106k	70.2
2k4	1k5	385m	08.30	3.90k	923	226	15.9k	70.2
8k2	5k1	383m	08.33	13.3k	3.14k	772	54.2k	70.1
10k	6k2	383m	08.34	16.2k	3.83k	941	65.9k	70.1
11k	6k8	382m	08.36	17.8k	4.20k	1.03k	72.4k	70.0
3k9	2k4	381m	08.38	6.30k	1.49k	366	25.6k	70.0
9k1	5k6	381m	08.38	14.7k	3.47k	853	59.7k	70.0
1k8	1k1	379m	08.42	2.90k	683	168	11.8k	69.8
3k6	2k2	379m	08.42	5.80k	1.37k	337	23.5k	69.8
15k	9k1	378m	08.46	24.1k	5.66k	1.40k	97.6k	69.7
3k3	2k	377m	08.46	5.30k	1.25k	308	21.4k	69.7
2k	1k2	375m	08.52	3.20k	750	186	12.9k	69.5
3k	1k8	375m	08.52	4.80k	1.13k	279	19.4k	69.5
2k7	1k6	372m	08.59	4.30k	1.00k	250	17.3k	69.3
2k2	1k3	371m	08.60	3.50k	817	203	14.1k	69.3
5k6	3k3	371m	08.62	8.90k	2.08k	517	35.8k	69.2
5k1	3k	370m	08.63	8.10k	1.89k	470	32.5k	69.2
6k2	3k6	367m	08.70	9.80k	2.28k	569	39.2k	68.9
13k	7k5	366m	08.73	20.5k	4.76k	1.19k	81.9k	68.8
4k7	2k7	365m	08.76	7.40k	1.71k	430	29.5k	68.7
6k8	3k9	364m	08.77	10.7k	2.48k	621	42.7k	68.7
7k5	4k3	364m	08.77	11.8k	2.73k	685	47.1k	68.7
8k2	4k7	364m	08.77	12.9k	2.99k	749	51.5k	68.7
16k	9k1	363m	08.81	25.1k	5.80k	1.46k	99.9k	68.6
12k	6k8	362m	08.83	18.8k	4.34k	1.09k	74.8k	68.5
3k9	2k2	361m	08.86	6.10k	1.41k	354	24.2k	68.4
11k	6k2	360m	08.86	17.2k	3.97k	999	68.3k	68.4
9k1	5k1	359m	08.89	14.2k	3.27k	824	56.3k	68.3
10k	5k6	359m	08.90	15.6k	3.59k	906	61.8k	68.3
4k3	2k4	358m	08.92	6.70k	1.54k	389	26.5k	68.2
18k	10k	357m	08.94	28.0k	6.43k	1.63k	111k	68.1

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
1k8	1k	357m	08.94	2.80k	643	163	11.1k	68.1
2k7	1k5	357m	08.94	4.20k	964	244	16.6k	68.1
3k6	2k	357m	08.94	5.60k	1.29k	325	22.1k	68.1
2k	1k1	355m	09.00	3.10k	710	182	12.1k	66.6
15k	8k2	353m	09.03	23.2k	5.30k	1.36k	90.4k	66.4
3k3	1k8	353m	09.05	5.10k	1.16k	299	19.9k	66.4
2k2	1k2	353m	09.05	3.40k	776	199	13.2k	66.4
2k4	1k3	351m	09.09	3.70k	843	217	14.4k	66.3
5k6	3k	349m	09.15	8.60k	1.95k	504	33.3k	66.0
3k	1k6	348m	09.17	4.60k	1.04k	270	17.8k	65.9
6k2	3k3	347m	09.18	9.50k	2.15k	557	36.7k	65.9
6k8	3k6	346m	09.21	10.4k	2.35k	610	40.1k	65.8
5k1	2k7	346m	09.21	7.80k	1.77k	457	30.1k	65.8
8k2	4k3	344m	09.27	12.5k	2.82k	733	48.1k	65.6
13k	6k8	343m	09.28	19.8k	4.46k	1.16k	76.1k	65.6
7k5	3k9	342m	09.32	11.4k	2.57k	669	43.7k	65.4
12k	6k2	341m	09.35	18.2k	4.09k	1.07k	69.7k	65.3
9k1	4k7	341m	09.36	13.8k	3.10k	809	52.8k	65.3
3k9	2k	339m	09.40	5.90k	1.32k	346	22.5k	65.1
16k	8k2	339m	09.40	24.2k	5.42k	1.42k	92.4k	65.1
4k3	2k2	338m	09.41	6.50k	1.46k	381	24.8k	65.1
4k7	2k4	338m	09.42	7.10k	1.59k	416	27.1k	65.1
10k	5k1	338m	09.43	15.1k	3.38k	886	57.6k	65.0
11k	5k6	337m	09.44	16.6k	3.71k	974	63.3k	65.0
18k	9k1	336m	09.48	27.1k	6.04k	1.59k	103k	64.8
15k	7k5	333m	09.54	22.5k	5.00k	1.32k	85.3k	64.6
2k	1k	333m	09.54	3.00k	667	176	11.4k	64.6
3k	1k5	333m	09.54	4.50k	1.00k	264	17.1k	64.6
3k6	1k8	333m	09.54	5.40k	1.20k	317	20.5k	64.6
20k	10k	333m	09.54	30.0k	6.67k	1.76k	114k	64.6
2k2	1k1	333m	09.54	3.30k	733	194	12.5k	64.6
2k4	1k2	333m	09.54	3.60k	800	211	13.6k	64.6
6k8	3k3	327m	09.72	10.1k	2.22k	592	37.9k	64.0
3k3	1k6	327m	09.72	4.90k	1.08k	287	18.4k	63.9
6k2	3k	326m	09.73	9.20k	2.02k	540	34.5k	63.9
5k6	2k7	325m	09.75	8.30k	1.82k	487	31.1k	63.8
2k7	1k3	325m	09.76	4.00k	878	235	15.0k	63.8

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
7k5	3k6	324m	09.78	11.1k	2.43k	651	41.5k	63.7
13k	6k2	323m	09.82	19.2k	4.20k	1.13k	71.6k	63.6
8k2	3k9	322m	09.83	12.1k	2.64k	710	45.1k	63.5
9k1	4k3	321m	09.87	13.4k	2.92k	786	49.8k	63.4
5k1	2k4	320m	09.90	7.50k	1.63k	440	27.8k	63.3
10k	4k7	320m	09.90	14.7k	3.20k	862	54.5k	63.2
16k	7k5	319m	09.92	23.5k	5.11k	1.38k	87.1k	63.2
4k7	2k2	319m	09.93	6.90k	1.50k	405	25.6k	63.1
12k	5k6	318m	09.95	17.6k	3.82k	1.03k	65.1k	63.1
4k3	2k	317m	09.97	6.30k	1.37k	369	23.3k	63.0
11k	5k1	317m	09.99	16.1k	3.48k	944	59.4k	62.9
3k9	1k8	316m	10.01	5.70k	1.23k	334	21.0k	62.8
2k4	1k1	314m	10.05	3.50k	754	205	12.9k	62.7
18k	8k2	313m	10.09	26.2k	5.63k	1.54k	96.1k	62.5
20k	9k1	313m	10.10	29.1k	6.25k	1.71k	107k	62.5
22k	10k	313m	10.10	32.0k	6.88k	1.88k	117k	62.5
2k2	1k	313m	10.10	3.20k	688	188	11.7k	62.5
3k3	1k5	313m	10.10	4.80k	1.03k	282	17.6k	62.5
15k	6k8	312m	10.12	21.8k	4.68k	1.28k	79.8k	62.4
2k7	1k2	308m	10.24	3.90k	831	229	14.2k	61.9
3k6	1k6	308m	10.24	5.20k	1.11k	305	18.9k	61.9
6k8	3k	306m	10.28	9.80k	2.08k	575	35.5k	61.8
7k5	3k3	306m	10.30	10.8k	2.29k	633	39.1k	61.7
8k2	3k6	305m	10.31	11.8k	2.50k	692	42.7k	61.6
6k2	2k7	303m	10.36	8.90k	1.88k	522	32.1k	61.4
3k	1k3	302m	10.39	4.30k	907	252	15.5k	61.3
5k1	2k2	301m	10.42	7.30k	1.54k	428	26.2k	61.2
13k	5k6	301m	10.43	18.6k	3.91k	1.09k	66.7k	61.2
10k	4k3	301m	10.44	14.3k	3.01k	839	51.3k	61.1
9k1	3k9	300m	10.46	13.0k	2.73k	762	46.5k	61.0
5k6	2k4	300m	10.46	8.00k	1.68k	469	28.6k	61.0
11k	4k7	299m	10.48	15.7k	3.29k	921	56.1k	61.0
4k7	2k	299m	10.50	6.70k	1.40k	393	23.9k	60.9
12k	5k1	298m	10.51	17.1k	3.58k	1.00k	61.0k	60.8
16k	6k8	298m	10.51	22.8k	4.77k	1.34k	81.4k	60.8
4k3	1k8	295m	10.60	6.10k	1.27k	358	21.6k	60.5
18k	7k5	294m	10.63	25.5k	5.29k	1.50k	90.3k	60.4

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
24k	10k	294m	10.63	34.0k	7.06k	1.99k	120k	60.4
2k4	1k	294m	10.63	3.40k	706	199	12.0k	60.4
3k6	1k5	294m	10.63	5.10k	1.06k	299	18.1k	60.4
22k	9k1	293m	10.67	31.1k	6.44k	1.82k	110k	60.2
15k	6k2	292m	10.68	21.2k	4.39k	1.24k	74.8k	60.2
3k9	1k6	291m	10.72	5.50k	1.13k	323	19.3k	60.0
20k	8k2	291m	10.73	28.2k	5.82k	1.65k	99.2k	60.0
2k7	1k1	289m	10.77	3.80k	782	223	13.3k	59.8
8k2	3k3	287m	10.84	11.5k	2.35k	674	40.1k	59.5
3k	1k2	286m	10.88	4.20k	857	246	14.6k	59.3
7k5	3k	286m	10.88	10.5k	2.14k	616	36.5k	59.3
6k8	2k7	284m	10.93	9.50k	1.93k	557	33.0k	59.1
9k1	3k6	283m	10.95	12.7k	2.58k	745	44.0k	59.0
3k3	1k3	283m	10.98	4.60k	933	270	15.9k	58.9
5k6	2k2	282m	10.99	7.80k	1.58k	457	26.9k	58.9
13k	5k1	282m	11.00	18.1k	3.66k	1.06k	62.5k	58.8
5k1	2k	282m	11.00	7.10k	1.44k	416	24.5k	58.8
12k	4k7	281m	11.01	16.7k	3.38k	979	57.6k	58.8
11k	4k3	281m	11.02	15.3k	3.09k	897	52.7k	58.7
10k	3k9	281m	11.04	13.9k	2.81k	815	47.8k	58.7
16k	6k2	279m	11.08	22.2k	4.47k	1.30k	76.2k	58.5
6k2	2k4	279m	11.09	8.60k	1.73k	504	29.5k	58.5
3k9	1k5	278m	11.13	5.40k	1.08k	317	18.5k	58.3
4k7	1k8	277m	11.15	6.50k	1.30k	381	22.2k	58.2
24k	9k1	275m	11.22	33.1k	6.60k	1.94k	113k	58.0
18k	6k8	274m	11.24	24.8k	4.94k	1.45k	84.2k	57.9
20k	7k5	273m	11.29	27.5k	5.45k	1.61k	93.0k	57.7
15k	5k6	272m	11.31	20.6k	4.08k	1.21k	69.5k	57.5
22k	8k2	272m	11.32	30.2k	5.97k	1.77k	102k	57.5
4k3	1k6	271m	11.33	5.90k	1.17k	346	19.9k	57.5
27k	10k	270m	11.36	37.0k	7.30k	2.17k	124k	57.3
2k7	1k	270m	11.36	3.70k	730	217	12.4k	57.3
3k	1k1	268m	11.43	4.10k	805	240	13.7k	57.1
8k2	3k	268m	11.44	11.2k	2.20k	657	37.4k	57.0
3k3	1k2	267m	11.48	4.50k	880	264	15.0k	56.9
9k1	3k3	266m	11.50	12.4k	2.42k	727	41.3k	56.8
13k	4k7	266m	11.52	17.7k	3.45k	1.04k	58.9k	56.7

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
3k6	1k3	265m	11.53	4.90k	955	287	16.3k	56.7
7k5	2k7	265m	11.54	10.2k	1.99k	598	33.8k	56.6
10k	3k6	265m	11.54	13.6k	2.65k	798	45.1k	56.6
12k	4k3	264m	11.57	16.3k	3.17k	956	54.0k	56.5
5k6	2k	263m	11.60	7.60k	1.47k	446	25.1k	56.4
6k2	2k2	262m	11.64	8.40k	1.62k	493	27.7k	56.2
11k	3k9	262m	11.64	14.9k	2.88k	874	49.1k	56.2
5k1	1k8	261m	11.67	6.90k	1.33k	405	22.7k	56.1
6k8	2k4	261m	11.67	9.20k	1.77k	540	30.2k	56.1
16k	5k6	259m	11.73	21.6k	4.15k	1.27k	70.7k	55.8
4k3	1k5	259m	11.75	5.80k	1.11k	340	19.0k	55.7
18k	6k2	256m	11.83	24.2k	4.61k	1.42k	78.6k	55.4
24k	8k2	255m	11.88	32.2k	6.11k	1.89k	104k	55.2
22k	7k5	254m	11.90	29.5k	5.59k	1.73k	95.4k	55.1
4k7	1k6	254m	11.90	6.30k	1.19k	369	20.4k	55.1
15k	5k1	254m	11.91	20.1k	3.81k	1.18k	64.9k	55.0
20k	6k8	254m	11.91	26.8k	5.07k	1.57k	86.5k	55.0
27k	9k1	252m	11.97	36.1k	6.81k	2.12k	116k	54.8
30k	10k	250m	12.04	40.0k	7.50k	2.35k	128k	54.5
3k	1k	250m	12.04	4.00k	750	235	12.8k	54.5
3k3	1k1	250m	12.04	4.40k	825	258	14.1k	54.5
3k6	1k2	250m	12.04	4.80k	900	282	15.3k	54.5
3k9	1k3	250m	12.04	5.20k	975	305	16.6k	54.5
13k	4k3	249m	12.09	17.3k	3.23k	1.01k	55.1k	54.3
10k	3k3	248m	12.11	13.3k	2.48k	780	42.3k	54.2
9k1	3k	248m	12.11	12.1k	2.26k	710	38.5k	54.2
8k2	2k7	248m	12.12	10.9k	2.03k	639	34.6k	54.2
11k	3k6	247m	12.16	14.6k	2.71k	856	46.2k	54.0
12k	3k9	245m	12.21	15.9k	2.94k	933	50.2k	53.8
6k8	2k2	244m	12.24	9.00k	1.66k	528	28.3k	53.7
6k2	2k	244m	12.26	8.20k	1.51k	481	25.8k	53.6
5k6	1k8	243m	12.28	7.40k	1.36k	434	23.2k	53.5
7k5	2k4	242m	12.31	9.90k	1.82k	581	31.0k	53.4
4k7	1k5	242m	12.33	6.20k	1.14k	364	19.4k	53.3
16k	5k1	242m	12.33	21.1k	3.87k	1.24k	65.9k	53.3
5k1	1k6	239m	12.44	6.70k	1.22k	393	20.8k	52.8
15k	4k7	239m	12.45	19.7k	3.58k	1.16k	61.0k	52.8

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
24k	7k5	238m	12.46	31.5k	5.71k	1.85k	97.4k	52.7
18k	5k6	237m	12.49	23.6k	4.27k	1.38k	72.8k	52.6
20k	6k2	237m	12.52	26.2k	4.73k	1.54k	80.7k	52.5
22k	6k8	236m	12.54	28.8k	5.19k	1.69k	88.6k	52.4
3k9	1k2	235m	12.57	5.10k	918	299	15.6k	52.3
3k6	1k1	234m	12.61	4.70k	843	276	14.4k	52.1
27k	8k2	233m	12.65	35.2k	6.29k	2.06k	107k	51.9
30k	9k1	233m	12.66	39.1k	6.98k	2.29k	119k	51.9
3k3	1k	233m	12.67	4.30k	767	252	13.1k	51.9
33k	10k	233m	12.67	43.0k	7.67k	2.52k	131k	51.9
4k3	1k3	232m	12.68	5.60k	998	328	17.0k	51.8
11k	3k3	231m	12.74	14.3k	2.54k	839	43.3k	51.6
12k	3k6	231m	12.74	15.6k	2.77k	915	47.2k	51.6
13k	3k9	231m	12.74	16.9k	3.00k	991	51.2k	51.6
10k	3k	231m	12.74	13.0k	2.31k	762	39.3k	51.6
9k1	2k7	229m	12.81	11.8k	2.08k	692	35.5k	51.3
6k8	2k	227m	12.87	8.80k	1.55k	516	26.4k	51.1
5k1	1k5	227m	12.87	6.60k	1.16k	387	19.8k	51.1
16k	4k7	227m	12.88	20.7k	3.63k	1.21k	61.9k	51.0
7k5	2k2	227m	12.89	9.70k	1.70k	569	29.0k	51.0
8k2	2k4	226m	12.90	10.6k	1.86k	622	31.7k	50.9
6k2	1k8	225m	12.96	8.00k	1.40k	469	23.8k	50.7
15k	4k3	223m	13.04	19.3k	3.34k	1.13k	57.0k	50.3
5k6	1k6	222m	13.06	7.20k	1.24k	422	21.2k	50.2
18k	5k1	221m	13.12	23.1k	3.97k	1.35k	67.8k	50.0
24k	6k8	221m	13.12	30.8k	5.30k	1.81k	90.3k	50.0
3k9	1k1	220m	13.15	5.00k	858	293	14.6k	49.9
22k	6k2	220m	13.16	28.2k	4.84k	1.65k	82.5k	49.9
20k	5k6	219m	13.20	25.6k	4.38k	1.50k	74.6k	49.7
4k3	1k2	218m	13.22	5.50k	938	323	16.0k	49.6
27k	7k5	217m	13.26	34.5k	5.87k	2.02k	100k	49.5
36k	10k	217m	13.26	46.0k	7.83k	2.70k	133k	49.5
3k6	1k	217m	13.26	4.60k	783	270	13.3k	49.5
13k	3k6	217m	13.28	16.6k	2.82k	974	48.1k	49.4
4k7	1k3	217m	13.28	6.00k	1.02k	352	17.4k	49.3
33k	9k1	216m	13.30	42.1k	7.13k	2.47k	122k	49.3
12k	3k3	216m	13.32	15.3k	2.59k	897	44.1k	49.2

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
30k	8k2	215m	13.36	38.2k	6.44k	2.24k	110k	49.0
11k	3k	214m	13.38	14.0k	2.36k	821	40.2k	48.9
10k	2k7	213m	13.45	12.7k	2.13k	745	36.2k	48.7
16k	4k3	212m	13.48	20.3k	3.39k	1.19k	57.8k	48.5
8k2	2k2	212m	13.49	10.4k	1.73k	610	29.6k	48.5
5k6	1k5	211m	13.50	7.10k	1.18k	416	20.2k	48.4
7k5	2k	211m	13.53	9.50k	1.58k	557	26.9k	48.3
6k8	1k8	209m	13.58	8.60k	1.42k	504	24.3k	48.1
9k1	2k4	209m	13.61	11.5k	1.90k	674	32.4k	48.0
18k	4k7	207m	13.68	22.7k	3.73k	1.33k	63.5k	47.7
15k	3k9	206m	13.71	18.9k	3.10k	1.11k	52.8k	47.6
24k	6k2	205m	13.75	30.2k	4.93k	1.77k	84.0k	47.4
6k2	1k6	205m	13.76	7.80k	1.27k	457	21.7k	47.4
39k	10k	204m	13.80	49.0k	7.96k	2.87k	136k	47.2
3k9	1k	204m	13.80	4.90k	796	287	13.6k	47.2
4k3	1k1	204m	13.82	5.40k	876	317	14.9k	47.2
4k7	1k2	203m	13.83	5.90k	956	346	16.3k	47.1
20k	5k1	203m	13.84	25.1k	4.06k	1.47k	69.3k	47.1
5k1	1k3	203m	13.84	6.40k	1.04k	375	17.7k	47.1
22k	5k6	203m	13.85	27.6k	4.46k	1.62k	76.1k	47.0
13k	3k3	202m	13.87	16.3k	2.63k	956	44.9k	46.9
36k	9k1	202m	13.90	45.1k	7.26k	2.65k	124k	46.8
27k	6k8	201m	13.93	33.8k	5.43k	1.98k	92.6k	46.7
30k	7k5	200m	13.98	37.5k	6.00k	2.20k	102k	46.5
12k	3k	200m	13.98	15.0k	2.40k	880	40.9k	46.5
33k	8k2	199m	14.02	41.2k	6.57k	2.42k	112k	46.3
11k	2k7	197m	14.11	13.7k	2.17k	804	37.0k	46.0
8k2	2k	196m	14.15	10.2k	1.61k	598	27.4k	45.8
16k	3k9	196m	14.16	19.9k	3.14k	1.17k	53.5k	45.8
6k2	1k5	195m	14.21	7.70k	1.21k	452	20.6k	45.6
9k1	2k2	195m	14.21	11.3k	1.77k	663	30.2k	45.6
15k	3k6	194m	14.26	18.6k	2.90k	1.09k	49.5k	45.4
7k5	1k8	194m	14.26	9.30k	1.45k	545	24.8k	45.4
10k	2k4	194m	14.26	12.4k	1.94k	727	33.0k	45.4
18k	4k3	193m	14.30	22.3k	3.47k	1.31k	59.2k	45.2
6k8	1k6	190m	14.40	8.40k	1.30k	493	22.1k	44.8
5k1	1k2	190m	14.40	6.30k	971	369	16.6k	44.8

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
20k	4k7	190m	14.41	24.7k	3.81k	1.45k	64.9k	44.8
4k7	1k1	190m	14.44	5.80k	891	340	15.2k	44.7
24k	5k6	189m	14.46	29.6k	4.54k	1.74k	77.4k	44.6
39k	9k1	189m	14.46	48.1k	7.38k	2.82k	126k	44.6
4k3	1k	189m	14.49	5.30k	811	311	13.8k	44.5
43k	10k	189m	14.49	53.0k	8.11k	3.11k	138k	44.5
5k6	1k3	188m	14.50	6.90k	1.06k	405	18.0k	44.5
22k	5k1	188m	14.51	27.1k	4.14k	1.59k	70.6k	44.4
13k	3k	188m	14.54	16.0k	2.44k	938	41.6k	44.3
27k	6k2	187m	14.57	33.2k	5.04k	1.95k	86.0k	44.2
36k	8k2	186m	14.63	44.2k	6.68k	2.59k	114k	43.9
33k	7k5	185m	14.65	40.5k	6.11k	2.38k	104k	43.9
30k	6k8	185m	14.67	36.8k	5.54k	2.16k	94.5k	43.8
12k	2k7	184m	14.72	14.7k	2.20k	862	37.6k	43.6
16k	3k6	184m	14.72	19.6k	2.94k	1.15k	50.1k	43.6
6k8	1k5	181m	14.86	8.30k	1.23k	487	21.0k	43.0
15k	3k3	180m	14.88	18.3k	2.70k	1.07k	46.1k	43.0
10k	2k2	180m	14.88	12.2k	1.80k	716	30.7k	43.0
9k1	2k	180m	14.89	11.1k	1.64k	651	28.0k	42.9
8k2	1k8	180m	14.89	10.0k	1.48k	587	25.2k	42.9
11k	2k4	179m	14.94	13.4k	1.97k	786	33.6k	42.7
18k	3k9	178m	14.99	21.9k	3.21k	1.28k	54.7k	42.6
5k1	1k1	177m	15.02	6.20k	905	364	15.4k	42.4
20k	4k3	177m	15.04	24.3k	3.54k	1.43k	60.3k	42.3
5k6	1k2	176m	15.07	6.80k	988	399	16.8k	42.2
22k	4k7	176m	15.09	26.7k	3.87k	1.57k	66.0k	42.2
7k5	1k6	176m	15.10	9.10k	1.32k	534	22.5k	42.1
47k	10k	175m	15.12	57.0k	8.25k	3.34k	141k	42.1
4k7	1k	175m	15.12	5.70k	825	334	14.1k	42.1
24k	5k1	175m	15.13	29.1k	4.21k	1.71k	71.7k	42.0
43k	9k1	175m	15.16	52.1k	7.51k	3.06k	128k	41.9
39k	8k2	174m	15.20	47.2k	6.78k	2.77k	116k	41.7
6k2	1k3	173m	15.22	7.50k	1.07k	440	18.3k	41.7
36k	7k5	172m	15.27	43.5k	6.21k	2.55k	106k	41.5
13k	2k7	172m	15.29	15.7k	2.24k	921	38.1k	41.4
27k	5k6	172m	15.30	32.6k	4.64k	1.91k	79.1k	41.4
30k	6k2	171m	15.33	36.2k	5.14k	2.12k	87.6k	41.3

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
16k	3k3	171m	15.34	19.3k	2.74k	1.13k	46.6k	41.2
33k	6k8	171m	15.35	39.8k	5.64k	2.33k	96.1k	41.2
11k	2k2	167m	15.56	13.2k	1.83k	774	31.3k	40.4
12k	2k4	167m	15.56	14.4k	2.00k	845	34.1k	40.4
15k	3k	167m	15.56	18.0k	2.50k	1.06k	42.6k	40.4
18k	3k6	167m	15.56	21.6k	3.00k	1.27k	51.2k	40.4
7k5	1k5	167m	15.56	9.00k	1.25k	528	21.3k	40.4
10k	2k	167m	15.56	12.0k	1.67k	704	28.4k	40.4
9k1	1k8	165m	15.64	10.9k	1.50k	639	25.6k	40.1
5k6	1k1	164m	15.69	6.70k	919	393	15.7k	39.9
51k	10k	164m	15.71	61.0k	8.36k	3.58k	143k	39.8
5k1	1k	164m	15.71	6.10k	836	358	14.3k	39.8
24k	4k7	164m	15.72	28.7k	3.93k	1.68k	67.0k	39.8
22k	4k3	163m	15.73	26.3k	3.60k	1.54k	61.3k	39.8
8k2	1k6	163m	15.74	9.80k	1.34k	575	22.8k	39.7
20k	3k9	163m	15.75	23.9k	3.26k	1.40k	55.6k	39.7
47k	9k1	162m	15.80	56.1k	7.62k	3.29k	130k	39.5
6k2	1k2	162m	15.80	7.40k	1.01k	434	17.1k	39.5
39k	7k5	161m	15.85	46.5k	6.29k	2.73k	107k	39.3
6k8	1k3	160m	15.89	8.10k	1.09k	475	18.6k	39.2
43k	8k2	160m	15.91	51.2k	6.89k	3.00k	117k	39.1
27k	5k1	159m	15.98	32.1k	4.29k	1.88k	73.1k	38.8
36k	6k8	159m	15.98	42.8k	5.72k	2.51k	97.5k	38.8
33k	6k2	158m	16.02	39.2k	5.22k	2.30k	89.0k	38.7
16k	3k	158m	16.03	19.0k	2.53k	1.11k	43.1k	38.7
30k	5k6	157m	16.07	35.6k	4.72k	2.09k	80.5k	38.5
13k	2k4	156m	16.15	15.4k	2.03k	903	34.5k	38.2
18k	3k3	155m	16.20	21.3k	2.79k	1.26k	47.1k	37.3
12k	2k2	155m	16.20	14.2k	1.86k	841	31.4k	37.3
8k2	1k5	155m	16.21	9.70k	1.27k	575	21.4k	37.2
11k	2k	154m	16.26	13.0k	1.69k	770	28.6k	37.1
15k	2k7	153m	16.33	17.7k	2.29k	1.05k	38.6k	36.8
20k	3k6	153m	16.33	23.6k	3.05k	1.40k	51.5k	36.8
10k	1k8	153m	16.33	11.8k	1.53k	699	25.7k	36.8
24k	4k3	152m	16.37	28.3k	3.65k	1.68k	61.5k	36.7
56k	10k	152m	16.39	66.0k	8.48k	3.91k	143k	36.6
5k6	1k	152m	16.39	6.60k	848	391	14.3k	36.6

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
51k	9k1	151m	16.40	60.1k	7.72k	3.56k	130k	36.6
6k2	1k1	151m	16.44	7.30k	934	433	15.8k	36.5
22k	3k9	151m	16.44	25.9k	3.31k	1.53k	55.9k	36.4
6k8	1k2	150m	16.48	8.00k	1.02k	474	17.2k	36.3
9k1	1k6	150m	16.51	10.7k	1.36k	634	23.0k	36.2
47k	8k2	149m	16.56	55.2k	6.98k	3.27k	118k	36.0
43k	7k5	149m	16.56	50.5k	6.39k	2.99k	108k	36.0
39k	6k8	148m	16.57	45.8k	5.79k	2.71k	97.7k	36.0
27k	4k7	148m	16.58	31.7k	4.00k	1.88k	67.6k	36.0
7k5	1k3	148m	16.61	8.80k	1.11k	521	18.7k	35.9
36k	6k2	147m	16.66	42.2k	5.29k	2.50k	89.3k	35.7
30k	5k1	145m	16.75	35.1k	4.36k	2.08k	73.6k	35.4
33k	5k6	145m	16.77	38.6k	4.79k	2.29k	80.8k	35.3
13k	2k2	145m	16.79	15.2k	1.88k	901	31.8k	35.3
16k	2k7	144m	16.81	18.7k	2.31k	1.11k	39.0k	35.2
12k	2k	143m	16.90	14.0k	1.71k	830	28.9k	34.9
18k	3k	143m	16.90	21.0k	2.57k	1.24k	43.4k	34.9
20k	3k3	142m	16.98	23.3k	2.83k	1.38k	47.8k	34.6
9k1	1k5	142m	16.98	10.6k	1.29k	628	21.7k	34.6
11k	1k8	141m	17.04	12.8k	1.55k	758	26.1k	34.4
22k	3k6	141m	17.04	25.6k	3.09k	1.52k	52.2k	34.4
24k	3k9	140m	17.09	27.9k	3.35k	1.65k	56.6k	34.2
56k	9k1	140m	17.09	65.1k	7.83k	3.86k	132k	34.2
6k8	1k1	139m	17.12	7.90k	947	468	16.0k	34.1
62k	10k	139m	17.15	72.0k	8.61k	4.27k	145k	34.1
6k2	1k	139m	17.15	7.20k	861	427	14.5k	34.1
51k	8k2	139m	17.17	59.2k	7.06k	3.51k	119k	34.0
15k	2k4	138m	17.21	17.4k	2.07k	1.03k	34.9k	33.9
7k5	1k2	138m	17.21	8.70k	1.03k	516	17.5k	33.9
10k	1k6	138m	17.21	11.6k	1.38k	687	23.3k	33.9
47k	7k5	138m	17.23	54.5k	6.47k	3.23k	109k	33.8
27k	4k3	137m	17.24	31.3k	3.71k	1.85k	62.6k	33.8
39k	6k2	137m	17.25	45.2k	5.35k	2.68k	90.3k	33.7
8k2	1k3	137m	17.28	9.50k	1.12k	563	18.9k	33.6
43k	6k8	137m	17.29	49.8k	5.87k	2.95k	99.1k	33.6
30k	4k7	135m	17.36	34.7k	4.06k	2.06k	68.6k	33.4
36k	5k6	135m	17.42	41.6k	4.85k	2.46k	81.8k	33.2

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
33k	5k1	134m	17.47	38.1k	4.42k	2.26k	74.5k	33.0
13k	2k	133m	17.50	15.0k	1.73k	889	29.3k	32.9
12k	1k8	130m	17.69	13.8k	1.57k	818	26.4k	32.3
16k	2k4	130m	17.69	18.4k	2.09k	1.09k	35.2k	32.3
18k	2k7	130m	17.69	20.7k	2.35k	1.23k	39.6k	32.3
20k	3k	130m	17.69	23.0k	2.61k	1.36k	44.0k	32.3
22k	3k3	130m	17.69	25.3k	2.87k	1.50k	48.4k	32.3
24k	3k6	130m	17.69	27.6k	3.13k	1.64k	52.8k	32.3
10k	1k5	130m	17.69	11.5k	1.30k	681	22.0k	32.3
51k	7k5	128m	17.84	58.5k	6.54k	3.47k	110k	31.8
68k	10k	128m	17.84	78.0k	8.72k	4.62k	147k	31.8
6k8	1k	128m	17.84	7.80k	872	462	14.7k	31.8
62k	9k1	128m	17.86	71.1k	7.94k	4.21k	134k	31.8
15k	2k2	128m	17.86	17.2k	1.92k	1.02k	32.4k	31.8
7k5	1k1	128m	17.86	8.60k	959	510	16.2k	31.8
56k	8k2	128m	17.87	64.2k	7.15k	3.80k	121k	31.7
8k2	1k2	128m	17.88	9.40k	1.05k	557	17.7k	31.7
11k	1k6	127m	17.93	12.6k	1.40k	747	23.6k	31.6
47k	6k8	126m	17.97	53.8k	5.94k	3.19k	100k	31.4
27k	3k9	126m	17.98	30.9k	3.41k	1.83k	57.5k	31.4
43k	6k2	126m	17.99	49.2k	5.42k	2.92k	91.4k	31.4
39k	5k6	126m	18.02	44.6k	4.90k	2.64k	82.6k	31.3
30k	4k3	125m	18.04	34.3k	3.76k	2.03k	63.5k	31.2
9k1	1k3	125m	18.06	10.4k	1.14k	616	19.2k	31.2
33k	4k7	125m	18.08	37.7k	4.11k	2.23k	69.4k	31.1
36k	5k1	124m	18.13	41.1k	4.47k	2.44k	75.4k	31.0
13k	1k8	122m	18.30	14.8k	1.58k	877	26.7k	30.4
16k	2k2	121m	18.35	18.2k	1.93k	1.08k	32.6k	30.3
24k	3k3	121m	18.35	27.3k	2.90k	1.62k	49.0k	30.3
11k	1k5	120m	18.42	12.5k	1.32k	741	22.3k	30.1
22k	3k	120m	18.42	25.0k	2.64k	1.48k	44.6k	30.1
20k	2k7	119m	18.49	22.7k	2.38k	1.35k	40.1k	29.8
8k2	1k1	118m	18.54	9.30k	970	551	16.4k	29.7
56k	7k5	118m	18.55	63.5k	6.61k	3.76k	112k	29.7
68k	9k1	118m	18.56	77.1k	8.03k	4.57k	135k	29.6
15k	2k	118m	18.59	17.0k	1.76k	1.01k	29.8k	29.6
7k5	1k	118m	18.59	8.50k	882	504	14.9k	29.6

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
12k	1k6	118m	18.59	13.6k	1.41k	806	23.8k	29.6
18k	2k4	118m	18.59	20.4k	2.12k	1.21k	35.7k	29.6
27k	3k6	118m	18.59	30.6k	3.18k	1.81k	53.6k	29.6
51k	6k8	118m	18.59	57.8k	6.00k	3.42k	101k	29.6
75k	10k	118m	18.59	85.0k	8.82k	5.04k	149k	29.6
62k	8k2	117m	18.65	70.2k	7.24k	4.16k	122k	29.4
47k	6k2	117m	18.67	53.2k	5.48k	3.15k	92.4k	29.3
9k1	1k2	117m	18.67	10.3k	1.06k	610	17.9k	29.3
39k	5k1	116m	18.74	44.1k	4.51k	2.61k	76.1k	29.1
36k	4k7	115m	18.75	40.7k	4.16k	2.41k	70.2k	29.1
33k	4k3	115m	18.76	37.3k	3.80k	2.21k	64.2k	29.0
43k	5k6	115m	18.77	48.6k	4.95k	2.88k	83.6k	29.0
30k	3k9	115m	18.78	33.9k	3.45k	2.01k	58.2k	29.0
10k	1k3	115m	18.78	11.3k	1.15k	670	19.4k	29.0
12k	1k5	111m	19.08	13.5k	1.33k	800	22.5k	28.1
16k	2k	111m	19.08	18.0k	1.78k	1.07k	30.0k	28.1
24k	3k	111m	19.08	27.0k	2.67k	1.60k	45.0k	28.1
13k	1k6	110m	19.20	14.6k	1.42k	865	24.0k	27.8
22k	2k7	109m	19.23	24.7k	2.40k	1.46k	40.6k	27.7
18k	2k2	109m	19.26	20.2k	1.96k	1.20k	33.1k	27.6
27k	3k3	109m	19.26	30.3k	2.94k	1.80k	49.6k	27.6
82k	10k	109m	19.28	92.0k	8.91k	5.45k	150k	27.6
8k2	1k	109m	19.28	9.20k	891	545	15.0k	27.6
51k	6k2	108m	19.30	57.2k	5.53k	3.39k	93.3k	27.5
56k	6k8	108m	19.31	62.8k	6.06k	3.72k	102k	27.5
75k	9k1	108m	19.32	84.1k	8.12k	4.98k	137k	27.5
62k	7k5	108m	19.34	69.5k	6.69k	4.12k	113k	27.4
9k1	1k1	108m	19.34	10.2k	981	604	16.6k	27.4
68k	8k2	108m	19.36	76.2k	7.32k	4.52k	123k	27.4
39k	4k7	108m	19.37	43.7k	4.19k	2.59k	70.8k	27.3
15k	1k8	107m	19.40	16.8k	1.61k	995	27.1k	27.2
20k	2k4	107m	19.40	22.4k	2.14k	1.33k	36.2k	27.2
30k	3k6	107m	19.40	33.6k	3.21k	1.99k	54.2k	27.2
10k	1k2	107m	19.40	11.2k	1.07k	664	18.1k	27.2
36k	4k3	107m	19.44	40.3k	3.84k	2.39k	64.8k	27.1
47k	5k6	106m	19.46	52.6k	5.00k	3.12k	84.4k	27.1
43k	5k1	106m	19.49	48.1k	4.56k	2.85k	76.9k	27.0

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
11k	1k3	106m	19.52	12.3k	1.16k	729	19.6k	26.9
33k	3k9	106m	19.52	36.9k	3.49k	2.19k	58.9k	26.9
13k	1k5	103m	19.71	14.5k	1.34k	859	22.7k	26.4
16k	1k8	101m	19.90	17.8k	1.62k	1.05k	27.3k	25.9
24k	2k7	101m	19.90	26.7k	2.43k	1.58k	41.0k	25.9
18k	2k	100m	20.00	20.0k	1.80k	1.19k	30.4k	25.6
27k	3k	100m	20.00	30.0k	2.70k	1.78k	45.6k	25.6
82k	9k1	99.9m	20.01	91.1k	8.19k	5.40k	138k	25.6
56k	6k2	99.7m	20.03	62.2k	5.58k	3.69k	94.2k	25.6
68k	7k5	99.3m	20.06	75.5k	6.75k	4.47k	114k	25.5
39k	4k3	99.3m	20.06	43.3k	3.87k	2.57k	65.4k	25.5
20k	2k2	99.1m	20.08	22.2k	1.98k	1.32k	33.4k	25.4
30k	3k3	99.1m	20.08	33.3k	2.97k	1.97k	50.2k	25.4
10k	1k1	99.1m	20.08	11.1k	991	658	16.7k	25.4
9k1	1k	99.0m	20.09	10.1k	901	598	15.2k	25.4
91k	10k	99.0m	20.09	10.1k	9.01k	5.98k	152k	25.4
51k	5k6	98.9m	20.09	56.6k	5.05k	3.35k	85.2k	25.4
62k	6k8	98.8m	20.10	68.8k	6.13k	4.08k	103k	25.4
75k	8k2	98.6m	20.13	83.2k	7.39k	4.93k	125k	25.3
43k	4k7	98.5m	20.13	47.7k	4.24k	2.83k	71.5k	25.3
11k	1k2	98.4m	20.14	12.2k	1.08k	723	18.3k	25.3
22k	2k4	98.4m	20.14	24.4k	2.16k	1.45k	36.5k	25.3
33k	3k6	98.4m	20.14	36.6k	3.25k	2.17k	54.8k	25.3
47k	5k1	97.9m	20.19	52.1k	4.60k	3.09k	77.6k	25.2
12k	1k3	97.7m	20.20	13.3k	1.17k	788	19.8k	25.1
36k	3k9	97.7m	20.20	39.9k	3.52k	2.36k	59.4k	25.1
15k	1k6	96.4m	20.32	16.6k	1.45k	984	24.4k	24.8
11k	1k1	90.9m	20.83	12.1k	1.00k	717	16.9k	23.5
12k	1k2	90.9m	20.83	13.2k	1.09k	782	18.4k	23.5
13k	1k3	90.9m	20.83	14.3k	1.18k	847	19.9k	23.5
15k	1k5	90.9m	20.83	16.5k	1.36k	978	23.0k	23.5
16k	1k6	90.9m	20.83	17.6k	1.45k	1.04k	24.5k	23.5
18k	1k8	90.9m	20.83	19.8k	1.64k	1.17k	27.6k	23.5
20k	2k	90.9m	20.83	22.0k	1.82k	1.30k	30.7k	23.5
22k	2k2	90.9m	20.83	24.2k	2.00k	1.43k	33.8k	23.5
24k	2k4	90.9m	20.83	26.4k	2.18k	1.56k	36.8k	23.5
27k	2k7	90.9m	20.83	29.7k	2.45k	1.76k	41.4k	23.5

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
30k	3k	90.9m	20.83	33.0k	2.73k	1.96k	46.0k	23.5
33k	3k3	90.9m	20.83	36.3k	3.00k	2.15k	50.6k	23.5
36k	3k6	90.9m	20.83	39.6k	3.27k	2.35k	55.2k	23.5
39k	3k9	90.9m	20.83	42.9k	3.55k	2.54k	59.8k	23.5
43k	4k3	90.9m	20.83	47.3k	3.91k	2.80k	66.0k	23.5
47k	4k7	90.9m	20.83	51.7k	4.27k	3.06k	72.1k	23.5
51k	5k1	90.9m	20.83	56.1k	4.64k	3.32k	78.2k	23.5
56k	5k6	90.9m	20.83	61.6k	5.09k	3.65k	85.9k	23.5
62k	6k2	90.9m	20.83	68.2k	5.64k	4.04k	95.1k	23.5
68k	6k8	90.9m	20.83	74.8k	6.18k	4.43k	104k	23.5
75k	7k5	90.9m	20.83	82.5k	6.82k	4.89k	115k	23.5
82k	8k2	90.9m	20.83	90.2k	7.45k	5.34k	126k	23.5
91k	9k1	90.9m	20.83	100.0k	8.27k	5.93k	140k	23.5
100k	10k	90.9m	20.83	110.0k	9.09k	6.52k	153k	23.5
10k	1k	90.9m	20.83	11.0k	909	652	15.3k	23.5
16k	1k5	85.7m	21.34	17.5k	1.37k	1.04k	23.1k	22.3
13k	1k2	84.5m	21.46	14.2k	1.10k	841	18.5k	22.0
39k	3k6	84.5m	21.46	42.6k	3.30k	2.52k	55.6k	22.0
51k	4k7	84.4m	21.48	55.7k	4.30k	3.30k	72.6k	22.0
12k	1k1	84.0m	21.52	13.1k	1.01k	776	17.0k	21.9
24k	2k2	84.0m	21.52	26.2k	2.02k	1.55k	34.0k	21.9
36k	3k3	84.0m	21.52	39.3k	3.02k	2.33k	51.0k	21.9
47k	4k3	83.8m	21.53	51.3k	3.94k	3.04k	66.5k	21.9
82k	7k5	83.8m	21.54	89.5k	6.87k	5.30k	116k	21.9
68k	6k2	83.6m	21.56	74.2k	5.68k	4.40k	95.9k	21.8
56k	5k1	83.5m	21.57	61.1k	4.67k	3.62k	78.9k	21.8
100k	9k1	83.4m	21.58	109.0k	8.34k	6.46k	141k	21.8
11k	1k	83.3m	21.58	12.0k	917	711	15.5k	21.8
22k	2k	83.3m	21.58	24.0k	1.83k	1.42k	30.9k	21.8
33k	3k	83.3m	21.58	36.0k	2.75k	2.13k	46.4k	21.8
43k	3k9	83.2m	21.60	46.9k	3.58k	2.78k	60.3k	21.7
75k	6k8	83.1m	21.60	81.8k	6.23k	4.85k	105k	21.7
62k	5k6	82.8m	21.64	67.6k	5.14k	4.01k	86.7k	21.6
91k	8k2	82.7m	21.65	99.2k	7.52k	5.88k	127k	21.6
20k	1k8	82.6m	21.66	21.8k	1.65k	1.29k	27.9k	21.6
30k	2k7	82.6m	21.66	32.7k	2.48k	1.94k	41.8k	21.6
18k	1k6	81.6m	21.76	19.6k	1.47k	1.16k	24.8k	21.4

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
27k	2k4	81.6m	21.76	29.4k	2.20k	1.74k	37.2k	21.4
15k	1k3	79.8m	21.96	16.3k	1.20k	966	20.2k	20.9
13k	1k1	78.0m	22.16	14.1k	1.01k	835	17.1k	20.5
39k	3k3	78.0m	22.16	42.3k	3.04k	2.51k	51.3k	20.5
51k	4k3	77.8m	22.19	55.3k	3.97k	3.28k	66.9k	20.4
56k	4k7	77.4m	22.22	60.7k	4.34k	3.60k	73.2k	20.3
43k	3k6	77.3m	22.24	46.6k	3.32k	2.76k	56.1k	20.3
12k	1k	76.9m	22.28	13.0k	923	770	15.6k	20.2
18k	1k5	76.9m	22.28	19.5k	1.38k	1.16k	23.4k	20.2
24k	2k	76.9m	22.28	26.0k	1.85k	1.54k	31.2k	20.2
36k	3k	76.9m	22.28	39.0k	2.77k	2.31k	46.7k	20.2
47k	3k9	76.6m	22.31	50.9k	3.60k	3.02k	60.8k	20.2
82k	6k8	76.6m	22.32	88.8k	6.28k	5.26k	106k	20.1
11k	910R	76.4m	22.34	11.9k	840	706	14.2k	20.1
75k	6k2	76.4m	22.34	81.2k	5.73k	4.81k	96.6k	20.1
91k	7k5	76.1m	22.37	98.5k	6.93k	5.84k	117k	20.0
68k	5k6	76.1m	22.37	73.6k	5.17k	4.36k	87.3k	20.0
62k	5k1	76.0m	22.38	67.1k	4.71k	3.98k	79.5k	20.0
100k	8k2	75.8m	22.41	108.0k	7.58k	6.41k	128k	19.9
22k	1k8	75.6m	22.43	23.8k	1.66k	1.41k	28.1k	19.9
33k	2k7	75.6m	22.43	35.7k	2.50k	2.12k	42.1k	19.9
27k	2k2	75.3m	22.46	29.2k	2.03k	1.73k	34.3k	19.8
16k	1k3	75.1m	22.48	17.3k	1.20k	1.03k	20.3k	19.8
15k	1k2	74.1m	22.61	16.2k	1.11k	960	18.8k	19.5
20k	1k6	74.1m	22.61	21.6k	1.48k	1.28k	25.0k	19.5
30k	2k4	74.1m	22.61	32.4k	2.22k	1.92k	37.5k	19.5
13k	1k	71.4m	22.92	14.0k	929	830	15.7k	18.9
39k	3k	71.4m	22.92	42.0k	2.79k	2.49k	47.0k	18.9
56k	4k3	71.3m	22.94	60.3k	3.99k	3.57k	67.4k	18.9
43k	3k3	71.3m	22.94	46.3k	3.06k	2.74k	51.7k	18.9
47k	3k6	71.1m	22.96	50.6k	3.34k	3.00k	56.4k	18.8
51k	3k9	71.0m	22.97	54.9k	3.62k	3.25k	61.1k	18.8
12k	910R	70.5m	23.04	12.9k	846	765	14.3k	18.7
62k	4k7	70.5m	23.04	66.7k	4.37k	3.95k	73.7k	18.7
82k	6k2	70.3m	23.06	88.2k	5.76k	5.23k	97.3k	18.6
16k	1k2	69.8m	23.13	17.2k	1.12k	1.02k	18.8k	18.5
20k	1k5	69.8m	23.13	21.5k	1.40k	1.27k	23.5k	18.5

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
24k	1k8	69.8m	23.13	25.8k	1.67k	1.53k	28.3k	18.5
36k	2k7	69.8m	23.13	38.7k	2.51k	2.29k	42.4k	18.5
68k	5k1	69.8m	23.13	73.1k	4.74k	4.33k	80.1k	18.5
100k	7k5	69.8m	23.13	108.0k	6.98k	6.37k	118k	18.5
91k	6k8	69.5m	23.16	97.8k	6.33k	5.80k	107k	18.4
75k	5k6	69.5m	23.16	80.6k	5.21k	4.78k	87.9k	18.4
11k	820R	69.4m	23.18	11.8k	763	700	12.9k	18.4
27k	2k	69.0m	23.23	29.0k	1.86k	1.72k	31.4k	18.3
15k	1k1	68.3m	23.31	16.1k	1.02k	954	17.3k	18.1
30k	2k2	68.3m	23.31	32.2k	2.05k	1.91k	34.6k	18.1
22k	1k6	67.8m	23.38	23.6k	1.49k	1.40k	25.2k	18.0
33k	2k4	67.8m	23.38	35.4k	2.24k	2.10k	37.8k	18.0
18k	1k3	67.4m	23.43	19.3k	1.21k	1.14k	20.5k	17.9
51k	3k6	65.9m	23.62	54.6k	3.36k	3.24k	56.7k	17.5
47k	3k3	65.6m	23.66	50.3k	3.08k	2.98k	52.0k	17.5
13k	910R	65.4m	23.69	13.9k	850	824	14.4k	17.4
43k	3k	65.2m	23.71	46.0k	2.80k	2.73k	47.3k	17.4
56k	3k9	65.1m	23.73	59.9k	3.65k	3.55k	61.5k	17.3
62k	4k3	64.9m	23.76	66.3k	4.02k	3.93k	67.9k	17.3
39k	2k7	64.7m	23.78	41.7k	2.53k	2.47k	42.6k	17.2
68k	4k7	64.6m	23.79	72.7k	4.40k	4.31k	74.2k	17.2
16k	1k1	64.3m	23.83	17.1k	1.03k	1.01k	17.4k	17.1
12k	820R	64.0m	23.88	12.8k	768	760	13.0k	17.1
82k	5k6	63.9m	23.89	87.6k	5.24k	5.19k	88.5k	17.0
11k	750R	63.8m	23.90	11.8k	702	696	11.8k	17.0
22k	1k5	63.8m	23.90	23.5k	1.40k	1.39k	23.7k1	17.0
91k	6k2	63.8m	23.91	97.2k	5.80k	5.76k	98.0k	17.0
75k	5k1	63.7m	23.92	80.1k	4.78k	4.75k	80.6k	17.0
100k	6k8	63.7m	23.92	107.0k	6.37k	6.33k	107k	17.0
15k	1k	62.5m	24.08	16.0k	938	948	15.8k	16.7
18k	1k2	62.5m	24.08	19.2k	1.13k	1.14k	19.0k	16.7
24k	1k6	62.5m	24.08	25.6k	1.50k	1.52k	25.3k	16.7
27k	1k8	62.5m	24.08	28.8k	1.69k	1.71k	28.5k	16.7
30k	2k	62.5m	24.08	32.0k	1.88k	1.90k	31.6k	16.7
33k	2k2	62.5m	24.08	35.2k	2.06k	2.09k	34.8k	16.7
36k	2k4	62.5m	24.08	38.4k	2.25k	2.28k	38.0k	16.7
20k	1k3	61.0m	24.29	21.3k	1.22k	1.26k	20.6k	16.3

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
51k	3k3	60.8m	24.33	54.3k	3.10k	3.22k	52.3k	16.3
56k	3k6	60.4m	24.38	59.6k	3.38k	3.53k	57.1k	16.2
47k	3k	60.0m	24.44	50.0k	2.82k	2.96k	47.6k	16.1
68k	4k3	59.5m	24.51	72.3k	4.04k	4.28k	68.3k	15.9
13k	820R	59.3m	24.53	13.8k	771	819	13.0k	15.9
62k	3k9	59.2m	24.56	65.9k	3.67k	3.90k	61.9k	15.9
43k	2k7	59.1m	24.57	45.7k	2.54k	2.71k	42.9k	15.8
75k	4k7	59.0m	24.59	79.7k	4.42k	4.72k	74.6k	15.8
16k	1k	58.8m	24.61	17.0k	941	1.01k	15.9k	15.8
12k	750R	58.8m	24.61	12.8k	706	755	11.9k	15.8
24k	1k5	58.8m	24.61	25.5k	1.41k	1.51k	23.8k	15.8
82k	5k1	58.6m	24.65	87.1k	4.80k	5.16k	81.0k	15.7
100k	6k2	58.4m	24.67	106.0k	5.84k	6.29k	98.5k	15.7
11k	680R	58.2m	24.70	11.7k	640	692	10.8k	15.6
39k	2k4	58.0m	24.74	41.4k	2.26k	2.45k	38.2k	15.6
91k	5k6	58.0m	24.74	96.6k	5.28k	5.72k	89.0k	15.6
18k	1k1	57.6m	24.79	19.1k	1.04k	1.13k	17.5k	15.5
36k	2k2	57.6m	24.79	38.2k	2.07k	2.26k	35.0k	15.5
15k	910R	57.2m	24.85	15.9k	858	943	14.5k	15.4
33k	2k	57.1m	24.86	35.0k	1.89k	2.07k	31.8k	15.3
20k	1k2	56.6m	24.94	21.2k	1.13k	1.26k	19.1k	15.2
30k	1k8	56.6m	24.94	31.8k	1.70k	1.88k	28.7k	15.2
27k	1k6	55.9m	25.04	28.6k	1.51k	1.69k	25.5k	15.0
22k	1k3	55.8m	25.07	23.3k	1.23k	1.38k	20.7k	15.0
56k	3k3	55.6m	25.09	59.3k	3.12k	3.51k	52.6k	15.0
51k	3k	55.6m	25.11	54.0k	2.83k	3.20k	47.8k	14.9
62k	3k6	54.9m	25.21	65.6k	3.40k	3.89k	57.4k	14.8
13k	750R	54.5m	25.26	13.8k	709	815	12.0k	14.7
47k	2k7	54.3m	25.30	49.7k	2.55k	2.94k	43.1k	14.6
68k	3k9	54.2m	25.31	71.9k	3.69k	4.26k	62.2k	14.6
75k	4k3	54.2m	25.32	79.3k	4.07k	4.70k	68.6k	14.6
82k	4k7	54.2m	25.32	86.7k	4.45k	5.14k	75.0k	14.6
16k	910R	53.8m	25.38	16.9k	861	1.00k	14.5k	14.5
12k	680R	53.6m	25.41	12.7k	644	751	10.9k	14.5
39k	2k2	53.4m	25.45	41.2k	2.08k	2.44k	35.1k	14.4
11k	620R	53.4m	25.46	11.6k	587	689	9.91k	14.4
91k	5k1	53.1m	25.50	96.1k	4.83k	5.69k	81.5k	14.3

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
100k	5k6	53.0m	25.51	106.0k	5.30k	6.26k	89.5k	14.3
43k	2k4	52.9m	25.54	45.4k	2.27k	2.69k	38.4k	14.3
18k	1k	52.6m	25.58	19.0k	947	1.13k	16.0k	14.2
27k	1k5	52.6m	25.58	28.5k	1.42k	1.69k	24.0k	14.2
36k	2k	52.6m	25.58	38.0k	1.89k	2.25k	32.0k	14.2
20k	1k1	52.1m	25.66	21.1k	1.04k	1.25k	17.6k	14.1
15k	820R	51.8m	25.71	15.8k	777	937	13.1k	14.0
22k	1k2	51.7m	25.73	23.2k	1.14k	1.37k	19.2k	14.0
33k	1k8	51.7m	25.73	34.8k	1.71k	2.06k	28.8k	14.0
24k	1k3	51.4m	25.78	25.3k	1.23k	1.50k	20.8k	13.9
56k	3k	50.8m	25.87	59.0k	2.85k	3.50k	48.1k	13.7
30k	1k6	50.6m	25.91	31.6k	1.52k	1.87k	25.6k	13.7
62k	3k3	50.5m	25.93	65.3k	3.13k	3.87k	52.9k	13.7
51k	2k7	50.3m	25.97	53.7k	2.56k	3.18k	43.3k	13.6
68k	3k6	50.3m	25.97	71.6k	3.42k	4.24k	57.7k	13.6
82k	4k3	49.8m	26.05	86.3k	4.09k	5.11k	69.0k	13.5
13k	680R	49.7m	26.07	13.7k	646	811	10.9k	13.5
75k	3k9	49.4m	26.12	78.9k	3.71k	4.68k	62.6k	13.4
12k	620R	49.1m	26.17	12.6k	590	748	9.95k	13.3
91k	4k7	49.1m	26.18	95.7k	4.47k	5.67k	75.4k	13.3
39k	2k	48.8m	26.24	41.0k	1.90k	2.43k	32.1k	13.2
16k	820R	48.8m	26.24	16.8k	780	997	13.2k	13.2
43k	2k2	48.7m	26.25	45.2k	2.09k	2.68k	35.3k	13.2
47k	2k4	48.6m	26.27	49.4k	2.28k	2.93k	38.5k	13.2
100k	5k1	48.5m	26.28	105.0k	4.85k	6.23k	81.9k	13.2
11k	560R	48.4m	26.30	11.6k	533	685	8.99k	13.1
18k	910R	48.1m	26.35	18.9k	866	1.12k	14.6k	13.0
15k	750R	47.6m	26.44	15.8k	714	933	12.1k	12.9
20k	1k	47.6m	26.44	21.0k	952	1.24k	16.1k	12.9
22k	1k1	47.6m	26.44	23.1k	1.05k	1.37k	17.7k	12.9
24k	1k2	47.6m	26.44	25.2k	1.14k	1.49k	19.3k	12.9
30k	1k5	47.6m	26.44	31.5k	1.43k	1.87k	24.1k	12.9
36k	1k8	47.6m	26.44	37.8k	1.71k	2.24k	28.9k	12.9
68k	3k3	46.3m	26.69	71.3k	3.15k	4.22k	53.1k	12.6
33k	1k6	46.2m	26.70	34.6k	1.53k	2.05k	25.8k	12.6
62k	3k	46.2m	26.72	65.0k	2.86k	3.85k	48.3k	12.5
56k	2k7	46.0m	26.75	58.7k	2.58k	3.48k	43.5k	12.5

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
27k	1k3	45.9m	26.76	28.3k	1.24k	1.68k	20.9k	12.5
75k	3k6	45.8m	26.78	78.6k	3.44k	4.66k	58.0k	12.4
13k	620R	45.5m	26.84	13.6k	592	807	9.99k	12.4
82k	3k9	45.4m	26.86	85.9k	3.72k	5.09k	62.8k	12.3
91k	4k3	45.1m	26.91	95.3k	4.11k	5.65k	69.3k	12.3
51k	2k4	44.9m	26.95	53.4k	2.29k	3.16k	38.7k	12.2
100k	4k7	44.9m	26.96	105.0k	4.49k	6.20k	75.8k	12.2
16k	750R	44.8m	26.98	16.8k	716	992	12.1k	12.2
47k	2k2	44.7m	26.99	49.2k	2.10k	2.92k	35.5k	12.2
12k	560R	44.6m	27.02	12.6k	535	744	9.03k	12.1
43k	2k	44.4m	27.04	45.0k	1.91k	2.67k	32.3k	12.1
11k	510R	44.3m	27.07	11.5k	487	682	8.23k	12.1
39k	1k8	44.1m	27.11	40.8k	1.72k	2.42k	29.0k	12.0
24k	1k1	43.8m	27.17	25.1k	1.05k	1.49k	17.8k	11.9
18k	820R	43.6m	27.22	18.8k	784	1.12k	13.2k	11.9
20k	910R	43.5m	27.23	20.9k	870	1.24k	14.7k	11.9
22k	1k	43.5m	27.23	23.0k	957	1.36k	16.1k	11.8
33k	1k5	43.5m	27.23	34.5k	1.43k	2.04k	24.2k	11.8
15k	680R	43.4m	27.26	15.7k	651	929	11.0k	11.8
36k	1k6	42.6m	27.42	37.6k	1.53k	2.23k	25.9k	11.6
27k	1k2	42.6m	27.42	28.2k	1.15k	1.67k	19.4k	11.6
68k	3k	42.3m	27.48	71.0k	2.87k	4.21k	48.5k	11.5
75k	3k3	42.1m	27.50	78.3k	3.16k	4.64k	53.3k	11.5
82k	3k6	42.1m	27.52	85.6k	3.45k	5.07k	58.2k	11.5
62k	2k7	41.7m	27.59	64.7k	2.59k	3.83k	43.7k	11.4
30k	1k3	41.5m	27.63	31.3k	1.25k	1.85k	21.0k	11.3
51k	2k2	41.4m	27.67	53.2k	2.11k	3.15k	35.6k	11.3
13k	560R	41.3m	27.68	13.6k	537	803	9.06k	11.3
100k	4k3	41.2m	27.70	104.0k	4.12k	6.18k	69.6k	11.3
56k	2k4	41.1m	27.72	58.4k	2.30k	3.46k	38.8k	11.2
91k	3k9	41.1m	27.72	94.9k	3.74k	5.62k	63.1k	11.2
11k	470R	41.0m	27.75	11.5k	451	680	7.61k	11.2
47k	2k	40.8m	27.78	49.0k	1.92k	2.90k	32.4k	11.2
12k	510R	40.8m	27.79	12.5k	489	741	8.26k	11.1
16k	680R	40.8m	27.79	16.7k	652	988	11.0k	11.1
43k	1k8	40.2m	27.92	44.8k	1.73k	2.65k	29.2k	11.0
18k	750R	40.0m	27.96	18.8k	720	1.11k	12.2k	10.9

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
24k	1k	40.0m	27.96	25.0k	960	1.48k	16.2k	10.9
36k	1k5	40.0m	27.96	37.5k	1.44k	2.22k	24.3k	10.9
22k	910R	39.7m	28.02	22.9k	874	1.36k	14.7k	10.9
15k	620R	39.7m	28.03	15.6k	595	926	10.0k	10.9
39k	1k6	39.4m	28.09	40.6k	1.54k	2.41k	25.9k	10.8
20k	820R	39.4m	28.09	20.8k	788	1.23k	13.3k	10.8
27k	1k1	39.1m	28.15	28.1k	1.06k	1.67k	17.8k	10.7
82k	3k3	38.7m	28.25	85.3k	3.17k	5.05k	53.5k	10.6
30k	1k2	38.5m	28.30	31.2k	1.15k	1.85k	19.5k	10.5
75k	3k	38.5m	28.30	78.0k	2.88k	4.62k	48.7k	10.5
68k	2k7	38.2m	28.36	70.7k	2.60k	4.19k	43.8k	10.5
91k	3k6	38.1m	28.39	94.6k	3.46k	5.61k	58.4k	10.4
33k	1k3	37.9m	28.43	34.3k	1.25k	2.03k	21.1k	10.4
56k	2k2	37.8m	28.45	58.2k	2.12k	3.45k	35.7k	10.4
13k	510R	37.7m	28.46	13.5k	491	801	8.28k	10.3
51k	2k	37.7m	28.46	53.0k	1.92k	3.14k	32.5k	10.3
12k	470R	37.7m	28.48	12.5k	452	739	7.63k	10.3
11k	430R	37.6m	28.49	11.4k	414	677	6.98k	10.3
100k	3k9	37.5m	28.51	104.0k	3.75k	6.16k	63.3k	10.3
16k	620R	37.3m	28.56	16.6k	597	985	10.1k	10.2
62k	2k4	37.3m	28.57	64.4k	2.31k	3.82k	39.0k	10.2
39k	1k5	37.0m	28.63	40.5k	1.44k	2.40k	24.4k	10.2
47k	1k8	36.9m	28.66	48.8k	1.73k	2.89k	29.3k	10.1
24k	910R	36.5m	28.75	24.9k	877	1.48k	14.8k	10.0
18k	680R	36.4m	28.78	18.7k	655	1.11k	11.1k	9.99
20k	750R	36.1m	28.84	20.8k	723	1.23k	12.2k	9.92
15k	560R	36.0m	28.88	15.6k	540	922	9.11k	9.88
22k	820R	35.9m	28.89	22.8k	791	1.35k	13.3k	9.87
43k	1k6	35.9m	28.90	44.6k	1.54k	2.64k	26.0k	9.85
27k	1k	35.7m	28.94	28.0k	964	1.66k	16.3k	9.81
30k	1k1	35.4m	29.03	31.1k	1.06k	1.84k	17.9k	9.72
82k	3k	35.3m	29.05	85.0k	2.89k	5.04k	48.8k	9.70
33k	1k2	35.1m	29.10	34.2k	1.16k	2.03k	19.5k	9.64
91k	3k3	35.0m	29.12	94.3k	3.18k	5.59k	53.7k	9.62
13k	470R	34.9m	29.15	13.5k	454	798	7.66k	9.59
36k	1k3	34.9m	29.16	37.3k	1.25k	2.21k	21.2k	9.58
75k	2k7	34.7m	29.18	77.7k	2.61k	4.60k	44.0k	9.55

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
100k	3k6	34.7m	29.18	104.0k	3.47k	6.14k	58.6k	9.55
12k	430R	34.6m	29.22	12.4k	415	737	7.01k	9.51
56k	2k	34.5m	29.25	58.0k	1.93k	3.44k	32.6k	9.48
62k	2k2	34.3m	29.30	64.2k	2.12k	3.80k	35.9k	9.43
11k	390R	34.2m	29.31	11.4k	377	675	6.36k	9.42
51k	1k8	34.1m	29.35	52.8k	1.74k	3.13k	29.3k	9.38
68k	2k4	34.1m	29.35	70.4k	2.32k	4.17k	39.1k	9.38
16k	560R	33.8m	29.42	16.6k	541	981	9.13k	9.31
43k	1k5	33.7m	29.45	44.5k	1.45k	2.64k	24.5k	9.28
18k	620R	33.3m	29.55	18.6k	599	1.10k	10.1k	9.17
24k	820R	33.0m	29.62	24.8k	793	1.47k	13.4k	9.10
22k	750R	33.0m	29.64	22.8k	725	1.35k	12.2k	9.08
47k	1k6	32.9m	29.65	48.6k	1.55k	2.88k	26.1k	9.07
15k	510R	32.9m	29.66	15.5k	493	919	8.32k	9.06
20k	680R	32.9m	29.66	20.7k	658	1.23k	11.1k	9.06
27k	910R	32.6m	29.73	27.9k	880	1.65k	14.9k	8.98
30k	1k	32.3m	29.83	31.0k	968	1.84k	16.3k	8.89
33k	1k1	32.3m	29.83	34.1k	1.06k	2.02k	18.0k	8.89
36k	1k2	32.3m	29.83	37.2k	1.16k	2.20k	19.6k	8.89
39k	1k3	32.3m	29.83	40.3k	1.26k	2.39k	21.2k	8.89
13k	430R	32.0m	29.89	13.4k	416	796	7.02k	8.83
100k	3k3	31.9m	29.91	103.0k	3.19k	6.12k	53.9k	8.81
91k	3k	31.9m	29.92	94.0k	2.90k	5.57k	49.0k	8.80
82k	2k7	31.9m	29.93	84.7k	2.61k	5.02k	44.1k	8.79
11k	360R	31.7m	29.98	11.4k	349	673	5.88k	8.74
12k	390R	31.5m	30.04	12.4k	378	734	6.37k	8.68
68k	2k2	31.3m	30.08	70.2k	2.13k	4.16k	36.0k	8.65
62k	2k	31.3m	30.10	64.0k	1.94k	3.79k	32.7k	8.62
56k	1k8	31.1m	30.13	57.8k	1.74k	3.42k	29.4k	8.59
75k	2k4	31.0m	30.17	77.4k	2.33k	4.59k	39.2k	8.56
47k	1k5	30.9m	30.19	48.5k	1.45k	2.87k	24.5k	8.54
16k	510R	30.9m	30.20	16.5k	494	978	8.34k	8.53
51k	1k6	30.4m	30.34	52.6k	1.55k	3.12k	26.2k	8.40
15k	470R	30.4m	30.35	15.5k	456	917	7.69k	8.39
24k	750R	30.3m	30.37	24.8k	727	1.47k	12.3k	8.37
18k	560R	30.2m	30.41	18.6k	543	1.10k	9.17k	8.33
20k	620R	30.1m	30.44	20.6k	601	1.22k	10.1k	8.31

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
22k	680R	30.0m	30.46	22.7k	660	1.34k	11.1k	8.28
39k	1k2	29.9m	30.50	40.2k	1.16k	2.38k	19.6k	8.25
36k	1k1	29.6m	30.56	37.1k	1.07k	2.20k	18.0k	8.19
27k	820R	29.5m	30.61	27.8k	796	1.65k	13.4k	8.15
30k	910R	29.4m	30.62	30.9k	883	1.83k	14.9k	8.14
33k	1k	29.4m	30.63	34.0k	971	2.01k	16.4k	8.13
43k	1k3	29.3m	30.65	44.3k	1.26k	2.62k	21.3k	8.11
11k	330R	29.1m	30.71	11.3k	320	671	5.41k	8.05
12k	360R	29.1m	30.71	12.4k	350	732	5.90k	8.05
13k	390R	29.1m	30.71	13.4k	379	793	6.39k	8.05
100k	3k	29.1m	30.71	103.0k	2.91k	6.10k	49.2k	8.05
91k	2k7	28.8m	30.81	93.7k	2.62k	5.55k	44.3k	7.97
51k	1k5	28.6m	30.88	52.5k	1.46k	3.11k	24.6k	7.91
68k	2k	28.6m	30.88	70.0k	1.94k	4.15k	32.8k	7.91
16k	470R	28.5m	30.89	16.5k	457	976	7.71k	7.90
75k	2k2	28.5m	30.90	77.2k	2.14k	4.57k	36.1k	7.89
82k	2k4	28.4m	30.92	84.4k	2.33k	5.00k	39.4k	7.87
62k	1k8	28.2m	30.99	63.8k	1.75k	3.78k	29.5k	7.81
15k	430R	27.9m	31.10	15.4k	418	914	7.05k	7.72
56k	1k6	27.8m	31.13	57.6k	1.56k	3.41k	26.3k	7.69
18k	510R	27.6m	31.20	18.5k	496	1.10k	8.37k	7.63
24k	680R	27.6m	31.20	24.7k	661	1.46k	11.2k	7.63
39k	1k1	27.4m	31.24	40.1k	1.07k	2.38k	18.1k	7.60
22k	620R	27.4m	31.24	22.6k	603	1.34k	10.2k	7.59
20k	560R	27.2m	31.30	20.6k	545	1.22k	9.19k	7.55
43k	1k2	27.1m	31.32	44.2k	1.17k	2.62k	19.7k	7.52
27k	750R	27.0m	31.36	27.8k	730	1.64k	12.3k	7.49
36k	1k	27.0m	31.36	37.0k	973	2.19k	16.4k	7.49
13k	360R	26.9m	31.39	13.4k	350	792	5.91k	7.47
47k	1k3	26.9m	31.40	48.3k	1.27k	2.86k	21.3k	7.46
33k	910R	26.8m	31.43	33.9k	886	2.01k	14.9k	7.44
12k	330R	26.8m	31.45	12.3k	321	731	5.42k	7.42
30k	820R	26.6m	31.50	30.8k	798	1.83k	13.5k	7.38
11k	300R	26.5m	31.52	11.3k	292	670	4.93k	7.36
100k	2k7	26.3m	31.60	103.0k	2.63k	6.09k	44.4k	7.29
16k	430R	26.2m	31.64	16.4k	419	974	7.07k	7.26
82k	2k2	26.1m	31.66	84.2k	2.14k	4.99k	36.2k	7.25

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
56k	1k5	26.1m	31.67	57.5k	1.46k	3.41k	24.7k	7.24
75k	2k	26.0m	31.71	77.0k	1.95k	4.56k	32.9k	7.21
68k	1k8	25.8m	31.77	69.8k	1.75k	4.14k	29.6k	7.16
91k	2k4	25.7m	31.80	93.4k	2.34k	5.53k	39.5k	7.13
18k	470R	25.4m	31.89	18.5k	458	1.09k	7.73k	7.06
15k	390R	25.3m	31.92	15.4k	380	912	6.42k	7.03
24k	620R	25.2m	31.98	24.6k	604	1.46k	10.2k	6.99
62k	1k6	25.2m	31.99	63.6k	1.56k	3.77k	26.3k	6.98
39k	1k	25.0m	32.04	40.0k	975	2.37k	16.5k	6.94
43k	1k1	24.9m	32.06	44.1k	1.07k	2.61k	18.1k	6.93
47k	1k2	24.9m	32.08	48.2k	1.17k	2.86k	19.7k	6.91
20k	510R	24.9m	32.09	20.5k	497	1.22k	8.39k	6.91
51k	1k3	24.9m	32.09	52.3k	1.27k	3.10k	21.4k	6.90
22k	560R	24.8m	32.10	22.6k	546	1.34k	9.22k	6.89
13k	330R	24.8m	32.13	13.3k	322	790	5.43k	6.88
36k	910R	24.7m	32.16	36.9k	888	2.19k	15.0k	6.85
27k	680R	24.6m	32.19	27.7k	663	1.64k	11.2k	6.83
12k	300R	24.4m	32.26	12.3k	293	729	4.94k	6.78
30k	750R	24.4m	32.26	30.8k	732	1.82k	12.3k	6.78
33k	820R	24.2m	32.31	33.8k	800	2.00k	13.5k	6.74
11k	270R	24.0m	32.41	11.3k	264	668	4.45k	6.66
82k	2k	23.8m	32.46	84.0k	1.95k	4.98k	32.9k	6.62
16k	390R	23.8m	32.47	16.4k	381	971	6.43k	6.62
62k	1k5	23.6m	32.53	63.5k	1.46k	3.76k	24.7k	6.57
91k	2k2	23.6m	32.54	93.2k	2.15k	5.52k	36.3k	6.56
15k	360R	23.4m	32.60	15.4k	352	910	5.93k	6.52
75k	1k8	23.4m	32.60	76.8k	1.76k	4.55k	29.7k	6.52
100k	2k4	23.4m	32.60	102.0k	2.34k	6.07k	39.6k	6.52
18k	430R	23.3m	32.64	18.4k	420	1.09k	7.09k	6.49
51k	1k2	23.0m	32.77	52.2k	1.17k	3.09k	19.8k	6.40
68k	1k6	23.0m	32.77	69.6k	1.56k	4.12k	26.4k	6.40
20k	470R	23.0m	32.78	20.5k	459	1.21k	7.75k	6.39
47k	1k1	22.9m	32.82	48.1k	1.07k	2.85k	18.1k	6.36
39k	910R	22.8m	32.84	39.9k	889	2.36k	15.0k	6.35
24k	560R	22.8m	32.84	24.6k	547	1.46k	9.24k	6.35
43k	1k	22.7m	32.87	44.0k	977	2.61k	16.5k	6.33
56k	1k3	22.7m	32.88	57.3k	1.27k	3.40k	21.4k	6.32

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
22k	510R	22.7m	32.90	22.5k	498	1.33k	8.41k	6.31
13k	300R	22.6m	32.93	13.3k	293	788	4.95k	6.28
27k	620R	22.4m	32.98	27.6k	606	1.64k	10.2k	6.25
36k	820R	22.3m	33.05	36.8k	802	2.18k	13.5k	6.20
33k	750R	22.2m	33.06	33.8k	733	2.00k	12.4k	6.19
30k	680R	22.2m	33.09	30.7k	665	1.82k	11.2k	6.17
12k	270R	22.0m	33.15	12.3k	264	727	4.46k	6.13
16k	360R	22.0m	33.15	16.4k	352	969	5.94k	6.13
68k	1k5	21.6m	33.32	69.5k	1.47k	4.12k	24.8k	6.01
15k	330R	21.5m	33.34	15.3k	323	908	5.45k	6.00
100k	2k2	21.5m	33.34	102.0k	2.15k	6.06k	36.3k	6.00
91k	2k	21.5m	33.35	93.0k	1.96k	5.51k	33.0k	5.99
82k	1k8	21.5m	33.36	83.8k	1.76k	4.97k	29.7k	5.99
11k	240R	21.4m	33.41	11.2k	235	666	3.96k	5.95
18k	390R	21.2m	33.47	18.4k	382	1.09k	6.44k	5.91
51k	1k1	21.1m	33.51	52.1k	1.08k	3.09k	18.2k	5.89
20k	430R	21.0m	33.54	20.4k	421	1.21k	7.10k	5.87
56k	1k2	21.0m	33.56	57.2k	1.17k	3.39k	19.8k	5.85
22k	470R	20.9m	33.59	22.5k	460	1.33k	7.77k	5.83
75k	1k6	20.9m	33.60	76.6k	1.57k	4.54k	26.4k	5.82
47k	1k	20.8m	33.62	48.0k	979	2.84k	16.5k	5.81
24k	510R	20.8m	33.64	24.5k	499	1.45k	8.43k	5.80
43k	910R	20.7m	33.67	43.9k	891	2.60k	15.0k	5.78
39k	820R	20.6m	33.73	39.8k	803	2.36k	13.6k	5.74
62k	1k3	20.5m	33.75	63.3k	1.27k	3.75k	21.5k	5.73
36k	750R	20.4m	33.80	36.8k	735	2.18k	12.4k	5.69
13k	270R	20.3m	33.83	13.3k	265	786	4.46k	5.68
27k	560R	20.3m	33.84	27.6k	549	1.63k	9.26k	5.67
30k	620R	20.2m	33.87	30.6k	607	1.81k	10.3k	5.65
16k	330R	20.2m	33.89	16.3k	323	968	5.46k	5.64
33k	680R	20.2m	33.90	33.7k	666	2.00k	11.2k	5.63
11k	220R	19.6m	34.15	11.2k	216	665	3.64k	5.48
12k	240R	19.6m	34.15	12.2k	235	725	3.97k	5.48
15k	300R	19.6m	34.15	15.3k	294	907	4.96k	5.48
18k	360R	19.6m	34.15	18.4k	353	1.09k	5.96k	5.48
75k	1k5	19.6m	34.15	76.5k	1.47k	4.53k	24.8k	5.48
100k	2k	19.6m	34.15	102.0k	1.96k	6.04k	33.1k	5.48

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
91k	1k8	19.4m	34.25	92.8k	1.77k	5.50k	29.8k	5.42
56k	1k1	19.3m	34.30	57.1k	1.08k	3.38k	18.2k	5.38
51k	1k	19.2m	34.32	52.0k	981	3.08k	16.6k	5.37
24k	470R	19.2m	34.33	24.5k	461	1.45k	7.78k	5.37
22k	430R	19.2m	34.35	22.4k	422	1.33k	7.12k	5.36
82k	1k6	19.1m	34.36	83.6k	1.57k	4.95k	26.5k	5.35
20k	390R	19.1m	34.37	20.4k	383	1.21k	6.46k	5.34
47k	910R	19.0m	34.43	47.9k	893	2.84k	15.1k	5.31
62k	1k2	19.0m	34.43	63.2k	1.18k	3.74k	19.9k	5.31
39k	750R	18.9m	34.49	39.8k	736	2.36k	12.4k	5.27
68k	1k3	18.8m	34.54	69.3k	1.28k	4.11k	21.5k	5.24
43k	820R	18.7m	34.56	43.8k	805	2.60k	13.6k	5.23
27k	510R	18.5m	34.64	27.5k	501	1.63k	8.45k	5.18
36k	680R	18.5m	34.64	36.7k	667	2.17k	11.3k	5.18
33k	620R	18.4m	34.68	33.6k	609	1.99k	10.3k	5.16
16k	300R	18.4m	34.70	16.3k	294	966	4.97k	5.15
30k	560R	18.3m	34.74	30.6k	550	1.81k	9.28k	5.12
13k	240R	18.1m	34.83	13.2k	236	785	3.98k	5.07
12k	220R	18.0m	34.89	12.2k	216	724	3.65k	5.04
18k	330R	18.0m	34.89	18.3k	324	1.09k	5.47k	5.04
82k	1k5	18.0m	34.91	83.5k	1.47k	4.95k	24.9k	5.02
11k	200R	17.9m	34.96	11.2k	196	664	3.32k	5.00
20k	360R	17.7m	35.05	20.4k	354	1.21k	5.97k	4.95
100k	1k8	17.7m	35.05	102.0k	1.77k	6.03k	29.8k	4.95
15k	270R	17.7m	35.05	15.3k	265	905	4.48k	4.95
24k	430R	17.6m	35.09	24.4k	422	1.45k	7.13k	4.92
56k	1k	17.5m	35.12	57.0k	982	3.38k	16.6k	4.91
51k	910R	17.5m	35.12	51.9k	894	3.08k	15.1k	4.91
62k	1k1	17.4m	35.17	63.1k	1.08k	3.74k	18.2k	4.88
22k	390R	17.4m	35.18	22.4k	383	1.33k	6.47k	4.87
68k	1k2	17.3m	35.22	69.2k	1.18k	4.10k	19.9k	4.85
91k	1k6	17.3m	35.25	92.6k	1.57k	5.49k	26.5k	4.84
47k	820R	17.1m	35.32	47.8k	806	2.83k	13.6k	4.80
43k	750R	17.1m	35.32	43.8k	737	2.59k	12.4k	4.80
39k	680R	17.1m	35.32	39.7k	668	2.35k	11.3k	4.80
27k	470R	17.1m	35.34	27.5k	462	1.63k	7.80k	4.79
75k	1k3	17.0m	35.37	76.3k	1.28k	4.52k	21.6k	4.77

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
36k	620R	16.9m	35.43	36.6k	610	2.17k	10.3k	4.74
30k	510R	16.7m	35.54	30.5k	501	1.81k	8.46k	4.68
33k	560R	16.7m	35.55	33.6k	551	1.99k	9.29k	4.67
13k	220R	16.6m	35.58	13.2k	216	783	3.65k	4.66
16k	270R	16.6m	35.60	16.3k	266	964	4.48k	4.65
12k	200R	16.4m	35.71	12.2k	197	723	3.32k	4.59
18k	300R	16.4m	35.71	18.3k	295	1.08k	4.98k	4.59
20k	330R	16.2m	35.79	20.3k	325	1.20k	5.48k	4.55
91k	1k5	16.2m	35.80	92.5k	1.48k	5.48k	24.9k	4.54
11k	180R	16.1m	35.86	11.2k	177	662	2.99k	4.51
22k	360R	16.1m	35.86	22.4k	354	1.32k	5.98k	4.51
24k	390R	16.0m	35.92	24.4k	384	1.45k	6.48k	4.48
56k	910R	16.0m	35.92	56.9k	895	3.37k	15.1k	4.48
68k	1k1	15.9m	35.96	69.1k	1.08k	4.09k	18.3k	4.46
62k	1k	15.9m	35.99	63.0k	984	3.73k	16.6k	4.45
51k	820R	15.8m	36.01	51.8k	807	3.07k	13.6k	4.44
15k	240R	15.7m	36.06	15.2k	236	903	3.99k	4.41
75k	1k2	15.7m	36.06	76.2k	1.18k	4.52k	19.9k	4.41
100k	1k6	15.7m	36.06	102.0k	1.57k	6.02k	26.6k	4.41
47k	750R	15.7m	36.08	47.8k	738	2.83k	12.5k	4.40
27k	430R	15.7m	36.10	27.4k	423	1.63k	7.14k	4.39
39k	620R	15.6m	36.11	39.6k	610	2.35k	10.3k	4.39
82k	1k3	15.6m	36.13	83.3k	1.28k	4.94k	21.6k	4.38
43k	680R	15.6m	36.16	43.7k	669	2.59k	11.3k	4.36
30k	470R	15.4m	36.24	30.5k	463	1.81k	7.81k	4.33
36k	560R	15.3m	36.30	36.6k	551	2.17k	9.31k	4.30
33k	510R	15.2m	36.35	33.5k	502	1.99k	8.48k	4.27
13k	200R	15.2m	36.39	13.2k	197	782	3.32k	4.25
12k	180R	14.8m	36.61	12.2k	177	722	2.99k	4.15
16k	240R	14.8m	36.61	16.2k	236	962	3.99k	4.15
18k	270R	14.8m	36.61	18.3k	266	1.08k	4.49k	4.15
20k	300R	14.8m	36.61	20.3k	296	1.20k	4.99k	4.15
22k	330R	14.8m	36.61	22.3k	325	1.32k	5.49k	4.15
24k	360R	14.8m	36.61	24.4k	355	1.44k	5.99k	4.15
100k	1k5	14.8m	36.61	102.0k	1.48k	6.01k	24.9k	4.15
51k	750R	14.5m	36.78	51.8k	739	3.07k	12.5k	4.07
68k	1k	14.5m	36.78	69.0k	986	4.09k	16.6k	4.07

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
62k	910R	14.5m	36.79	62.9k	897	3.73k	15.1k	4.06
15k	220R	14.5m	36.80	15.2k	217	902	3.66k	4.06
75k	1k1	14.5m	36.80	76.1k	1.08k	4.51k	18.3k	4.06
56k	820R	14.4m	36.81	56.8k	808	3.37k	13.6k	4.05
82k	1k2	14.4m	36.82	83.2k	1.18k	4.93k	20.0k	4.05
11k	160R	14.3m	36.87	11.2k	158	661	2.66k	4.02
47k	680R	14.3m	36.92	47.7k	670	2.83k	11.3k	4.00
27k	390R	14.2m	36.93	27.4k	384	1.62k	6.49k	4.00
43k	620R	14.2m	36.95	43.6k	611	2.58k	10.3k	3.99
39k	560R	14.2m	36.98	39.6k	552	2.34k	9.32k	3.97
30k	430R	14.1m	37.00	30.4k	424	1.80k	7.15k	3.97
91k	1k3	14.1m	37.03	92.3k	1.28k	5.47k	21.6k	3.96
33k	470R	14.0m	37.05	33.5k	463	1.98k	7.82k	3.94
36k	510R	14.0m	37.10	36.5k	503	2.16k	8.49k	3.92
13k	180R	13.7m	37.29	13.2k	178	781	3.00k	3.84
16k	220R	13.6m	37.35	16.2k	217	961	3.66k	3.81
24k	330R	13.6m	37.35	24.3k	326	1.44k	5.49k	3.81
11k	150R	13.5m	37.42	11.2k	148	661	2.50k	3.78
22k	300R	13.5m	37.42	22.3k	296	1.32k	4.99k	3.78
20k	270R	13.3m	37.51	20.3k	266	1.20k	4.50k	3.74
82k	1k1	13.2m	37.56	83.1k	1.09k	4.92k	18.3k	3.72
56k	750R	13.2m	37.58	56.8k	740	3.36k	12.5k	3.71
68k	910R	13.2m	37.58	68.9k	898	4.08k	15.2k	3.71
12k	160R	13.2m	37.62	12.2k	158	721	2.66k	3.70
15k	200R	13.2m	37.62	15.2k	197	901	3.33k	3.70
18k	240R	13.2m	37.62	18.2k	237	1.08k	4.00k	3.70
27k	360R	13.2m	37.62	27.4k	355	1.62k	6.00k	3.70
51k	680R	13.2m	37.62	51.7k	671	3.06k	11.3k	3.70
75k	1k	13.2m	37.62	76.0k	987	4.50k	16.7k	3.70
62k	820R	13.1m	37.69	62.8k	809	3.72k	13.7k	3.67
47k	620R	13.0m	37.71	47.6k	612	2.82k	10.3k	3.66
91k	1k2	13.0m	37.71	92.2k	1.18k	5.46k	20.0k	3.66
39k	510R	12.9m	37.78	39.5k	503	2.34k	8.50k	3.63
36k	470R	12.9m	37.80	36.5k	464	2.16k	7.83k	3.62
33k	430R	12.9m	37.81	33.4k	424	1.98k	7.16k	3.62
43k	560R	12.9m	37.82	43.6k	553	2.58k	9.33k	3.61
30k	390R	12.8m	37.83	30.4k	385	1.80k	6.50k	3.61

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
100k	1k3	12.8m	37.83	101.0k	1.28k	6.00k	21.7k	3.61
12k	150R	12.3m	38.17	12.2k	148	720	2.50k	3.47
16k	200R	12.3m	38.17	16.2k	198	960	3.33k	3.47
24k	300R	12.3m	38.17	24.3k	296	1.44k	5.00k	3.47
13k	160R	12.2m	38.30	13.2k	158	780	2.67k	3.42
22k	270R	12.1m	38.33	22.3k	267	1.32k	4.50k	3.41
18k	220R	12.1m	38.36	18.2k	217	1.08k	3.67k	3.40
27k	330R	12.1m	38.36	27.3k	326	1.62k	5.50k	3.40
82k	1k	12.0m	38.38	83.0k	988	4.92k	16.7k	3.39
51k	620R	12.0m	38.41	51.6k	613	3.06k	10.3k	3.38
56k	680R	12.0m	38.42	56.7k	672	3.36k	11.3k	3.38
75k	910R	12.0m	38.43	75.9k	899	4.50k	15.2k	3.37
62k	750R	12.0m	38.45	62.8k	741	3.72k	12.5k	3.36
91k	1k1	11.9m	38.46	92.1k	1.09k	5.46k	18.3k	3.36
68k	820R	11.9m	38.48	68.8k	810	4.08k	13.7k	3.35
39k	470R	11.9m	38.48	39.5k	464	2.34k	7.84k	3.35
15k	180R	11.9m	38.52	15.2k	178	899	3.00k	3.34
20k	240R	11.9m	38.52	20.2k	237	1.20k	4.00k	3.34
30k	360R	11.9m	38.52	30.4k	356	1.80k	6.00k	3.34
100k	1k2	11.9m	38.52	101.0k	1.19k	6.00k	20.0k	3.34
36k	430R	11.8m	38.56	36.4k	425	2.16k	7.17k	3.32
47k	560R	11.8m	38.58	47.6k	553	2.82k	9.34k	3.31
43k	510R	11.7m	38.62	43.5k	504	2.58k	8.51k	3.30
11k	130R	11.7m	38.65	11.1k	128	659	2.17k	3.29
33k	390R	11.7m	38.65	33.4k	385	1.98k	6.50k	3.29
13k	150R	11.4m	38.86	13.2k	148	779	2.50k	3.21
16k	180R	11.1m	39.07	16.2k	178	959	3.00k	3.13
24k	270R	11.1m	39.07	24.3k	267	1.44k	4.51k	3.13
18k	200R	11.0m	39.18	18.2k	198	1.08k	3.34k	3.10
27k	300R	11.0m	39.18	27.3k	297	1.62k	5.01k	3.10
82k	910R	11.0m	39.19	82.9k	900	4.91k	15.2k	3.09
56k	620R	11.0m	39.21	56.6k	613	3.35k	10.3k	3.08
68k	750R	10.9m	39.24	68.8k	742	4.07k	12.5k	3.07
39k	430R	10.9m	39.25	39.4k	425	2.34k	7.18k	3.07
20k	220R	10.9m	39.27	20.2k	218	1.20k	3.67k	3.07
30k	330R	10.9m	39.27	30.3k	326	1.80k	5.51k	3.07
100k	1k1	10.9m	39.27	101.0k	1.09k	5.99k	18.4k	3.07

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
91k	1k	10.9m	39.28	92.0k	989	5.45k	16.7k	3.06
51k	560R	10.9m	39.28	51.6k	554	3.06k	9.35k	3.06
62k	680R	10.8m	39.29	62.7k	673	3.71k	11.4k	3.06
75k	820R	10.8m	39.32	75.8k	811	4.49k	13.7k	3.05
43k	470R	10.8m	39.32	43.5k	465	2.58k	7.85k	3.05
11k	120R	10.8m	39.34	11.1k	119	659	2.00k	3.04
22k	240R	10.8m	39.34	22.2k	237	1.32k	4.01k	3.04
33k	360R	10.8m	39.34	33.4k	356	1.98k	6.01k	3.04
47k	510R	10.7m	39.38	47.5k	505	2.82k	8.51k	3.02
36k	390R	10.7m	39.40	36.4k	386	2.16k	6.51k	3.02
12k	130R	10.7m	39.40	12.1k	129	719	2.17k	3.02
15k	160R	10.6m	39.53	15.2k	158	898	2.67k	2.97
11k	110R	9.90m	40.09	11.1k	109	658	1.84k	2.79
12k	120R	9.90m	40.09	12.1k	119	718	2.01k	2.79
13k	130R	9.90m	40.09	13.1k	129	778	2.17k	2.79
15k	150R	9.90m	40.09	15.2k	149	898	2.51k	2.79
16k	160R	9.90m	40.09	16.2k	158	958	2.67k	2.79
18k	180R	9.90m	40.09	18.2k	178	1.08k	3.01k	2.79
20k	200R	9.90m	40.09	20.2k	198	1.20k	3.34k	2.79
22k	220R	9.90m	40.09	22.2k	218	1.32k	3.68k	2.79
24k	240R	9.90m	40.09	24.2k	238	1.44k	4.01k	2.79
27k	270R	9.90m	40.09	27.3k	267	1.62k	4.51k	2.79
30k	300R	9.90m	40.09	30.3k	297	1.80k	5.01k	2.79
33k	330R	9.90m	40.09	33.3k	327	1.97k	5.51k	2.79
36k	360R	9.90m	40.09	36.4k	356	2.15k	6.02k	2.79
39k	390R	9.90m	40.09	39.4k	386	2.33k	6.52k	2.79
43k	430R	9.90m	40.09	43.4k	426	2.57k	7.19k	2.79
47k	470R	9.90m	40.09	47.5k	465	2.81k	7.85k	2.79
51k	510R	9.90m	40.09	51.5k	505	3.05k	8.52k	2.79
56k	560R	9.90m	40.09	56.6k	554	3.35k	9.36k	2.79
62k	620R	9.90m	40.09	62.6k	614	3.71k	10.4k	2.79
68k	680R	9.90m	40.09	68.7k	673	4.07k	11.4k	2.79
75k	750R	9.90m	40.09	75.8k	743	4.49k	12.5k	2.79
82k	820R	9.90m	40.09	82.8k	812	4.91k	13.7k	2.79
91k	910R	9.90m	40.09	91.9k	901	5.45k	15.2k	2.79
100k	1k	9.90m	40.09	101.0k	990	5.98k	16.7k	2.79
16k	150R	9.29m	40.64	16.2k	149	957	2.51k	2.62

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
13k	120R	9.15m	40.78	13.1k	119	777	2.01k	2.58
39k	360R	9.15m	40.78	39.4k	357	2.33k	6.02k	2.58
51k	470R	9.13m	40.79	51.5k	466	3.05k	7.86k	2.58
12k	110R	9.08m	40.84	12.1k	109	718	1.84k	2.56
24k	220R	9.08m	40.84	24.2k	218	1.44k	3.68k	2.56
36k	330R	9.08m	40.84	36.3k	327	2.15k	5.52k	2.56
47k	430R	9.07m	40.85	47.4k	426	2.81k	7.19k	2.56
82k	750R	9.06m	40.85	82.8k	743	4.90k	12.5k	2.56
68k	620R	9.04m	40.88	68.6k	614	4.07k	10.4k	2.55
56k	510R	9.02m	40.89	56.5k	505	3.35k	8.53k	2.55
100k	910R	9.02m	40.90	101.0k	902	5.98k	15.2k	2.55
11k	100R	9.01m	40.91	11.1k	99.1	658	1.67k	2.54
22k	200R	9.01m	40.91	22.2k	198	1.32k	3.34k	2.54
33k	300R	9.01m	40.91	33.3k	297	1.97k	5.02k	2.54
43k	390R	8.99m	40.93	43.4k	386	2.57k	6.52k	2.54
75k	680R	8.99m	40.93	75.7k	674	4.48k	11.4k	2.54
62k	560R	8.95m	40.96	62.6k	555	3.71k	9.37k	2.53
91k	820R	8.93m	40.98	91.8k	813	5.44k	13.7k	2.52
20k	180R	8.92m	40.99	20.2k	178	1.20k	3.01k	2.52
30k	270R	8.92m	40.99	30.3k	268	1.79k	4.52k	2.52
18k	160R	8.81m	41.10	18.2k	159	1.08k	2.68k	2.49
27k	240R	8.81m	41.10	27.2k	238	1.61k	4.01k	2.49
15k	130R	8.59m	41.32	15.1k	129	897	2.18k	2.43
13k	110R	8.39m	41.52	13.1k	109	777	1.84k	2.37
39k	330R	8.39m	41.52	39.3k	327	2.33k	5.52k	2.37
51k	430R	8.36m	41.55	51.4k	426	3.05k	7.20k	2.36
56k	470R	8.32m	41.59	56.5k	466	3.35k	7.87k	2.35
43k	360R	8.30m	41.62	43.4k	357	2.57k	6.03k	2.35
12k	100R	8.26m	41.66	12.1k	99.2	717	1.67k	2.33
18k	150R	8.26m	41.66	18.2k	149	1.08k	2.51k	2.33
24k	200R	8.26m	41.66	24.2k	198	1.43k	3.35k	2.33
36k	300R	8.26m	41.66	36.3k	298	2.15k	5.02k	2.33
47k	390R	8.23m	41.69	47.4k	387	2.81k	6.53k	2.32
82k	680R	8.22m	41.70	82.7k	674	4.90k	11.4k	2.32
75k	620R	8.20m	41.72	75.6k	615	4.48k	10.4k	2.32
91k	750R	8.17m	41.75	91.8k	744	5.44k	12.6k	2.31
68k	560R	8.17m	41.76	68.6k	555	4.06k	9.37k	2.31

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
62k	510R	8.16m	41.77	62.5k	506	3.70k	8.54k	2.30
100k	820R	8.13m	41.79	101.0k	813	5.97k	13.7k	2.30
22k	180R	8.12m	41.81	22.2k	179	1.31k	3.01k	2.29
33k	270R	8.12m	41.81	33.3k	268	1.97k	4.52k	2.29
27k	220R	8.08m	41.85	27.2k	218	1.61k	3.68k	2.28
16k	130R	8.06m	41.87	16.1k	129	956	2.18k	2.28
15k	120R	7.94m	42.01	15.1k	119	896	2.01k	2.24
20k	160R	7.94m	42.01	20.2k	159	1.19k	2.68k	2.24
30k	240R	7.94m	42.01	30.2k	238	1.79k	4.02k	2.24
13k	100R	7.63m	42.35	13.1k	99.2	776	1.67k	2.16
39k	300R	7.63m	42.35	39.3k	298	2.33k	5.02k	2.16
56k	430R	7.62m	42.36	56.4k	427	3.34k	7.20k	2.15
43k	330R	7.62m	42.37	43.3k	327	2.57k	5.53k	2.15
47k	360R	7.60m	42.38	47.4k	357	2.81k	6.03k	2.15
51k	390R	7.59m	42.40	51.4k	387	3.05k	6.53k	2.15
62k	470R	7.52m	42.47	62.5k	466	3.70k	7.87k	2.13
82k	620R	7.50m	42.49	82.6k	615	4.90k	10.4k	2.12
16k	120R	7.44m	42.56	16.1k	119	955	2.01k	2.10
20k	150R	7.44m	42.56	20.2k	149	1.19k	2.51k	2.10
24k	180R	7.44m	42.56	24.2k	179	1.43k	3.02k	2.10
36k	270R	7.44m	42.56	36.3k	268	2.15k	4.52k	2.10
68k	510R	7.44m	42.56	68.5k	506	4.06k	8.54k	2.10
100k	750R	7.44m	42.56	101.0k	744	5.97k	12.6k	2.10
91k	680R	7.42m	42.60	91.7k	675	5.43k	11.4k	2.10
75k	560R	7.41m	42.60	75.6k	556	4.48k	9.38k	2.10
27k	200R	7.35m	42.67	27.2k	199	1.61k	3.35k	2.08
15k	110R	7.28m	42.76	15.1k	109	895	1.84k	2.06
30k	220R	7.28m	42.76	30.2k	218	1.79k	3.69k	2.06
22k	160R	7.22m	42.83	22.2k	159	1.31k	2.68k	2.04
33k	240R	7.22m	42.83	33.2k	238	1.97k	4.02k	2.04
18k	130R	7.17m	42.89	18.1k	129	1.07k	2.18k	2.03
51k	360R	7.01m	43.09	51.4k	357	3.04k	6.03k	1.98
47k	330R	6.97m	43.13	47.3k	328	2.80k	5.53k	1.97
43k	300R	6.93m	43.19	43.3k	298	2.57k	5.03k	1.96
56k	390R	6.92m	43.20	56.4k	387	3.34k	6.54k	1.96
62k	430R	6.89m	43.24	62.4k	427	3.70k	7.21k	1.95
39k	270R	6.88m	43.25	39.3k	268	2.33k	4.53k	1.94

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
68k	470R	6.86m	43.27	68.5k	467	4.06k	7.88k	1.94
16k	110R	6.83m	43.31	16.1k	109	955	1.84k	1.93
82k	560R	6.78m	43.37	82.6k	556	4.89k	9.39k	1.92
22k	150R	6.77m	43.39	22.2k	149	1.31k	2.51k	1.92
91k	620R	6.77m	43.39	91.6k	616	5.43k	10.4k	1.91
75k	510R	6.75m	43.41	75.5k	507	4.47k	8.55k	1.91
100k	680R	6.75m	43.41	101.0k	675	5.97k	11.4k	1.91
15k	100R	6.62m	43.58	15.1k	99.3	895	1.68k	1.87
18k	120R	6.62m	43.58	18.1k	119	1.07k	2.01k	1.87
24k	160R	6.62m	43.58	24.2k	159	1.43k	2.68k	1.87
27k	180R	6.62m	43.58	27.2k	179	1.61k	3.02k	1.87
30k	200R	6.62m	43.58	30.2k	199	1.79k	3.35k	1.87
33k	220R	6.62m	43.58	33.2k	219	1.97k	3.69k	1.87
36k	240R	6.62m	43.58	36.2k	238	2.15k	4.02k	1.87
20k	130R	6.46m	43.80	20.1k	129	1.19k	2.18k	1.83
51k	330R	6.43m	43.84	51.3k	328	3.04k	5.53k	1.82
56k	360R	6.39m	43.89	56.4k	358	3.34k	6.04k	1.81
47k	300R	6.34m	43.95	47.3k	298	2.80k	5.03k	1.80
68k	430R	6.28m	44.04	68.4k	427	4.05k	7.21k	1.78
62k	390R	6.25m	44.08	62.4k	388	3.70k	6.54k	1.77
43k	270R	6.24m	44.10	43.3k	268	2.56k	4.53k	1.77
75k	470R	6.23m	44.11	75.5k	467	4.47k	7.88k	1.76
16k	100R	6.21m	44.14	16.1k	99.4	954	1.68k	1.76
24k	150R	6.21m	44.14	24.2k	149	1.43k	2.52k	1.76
82k	510R	6.18m	44.18	82.5k	507	4.89k	8.55k	1.75
100k	620R	6.16m	44.21	101.0k	616	5.96k	10.4k	1.74
39k	240R	6.12m	44.27	39.2k	239	2.33k	4.03k	1.73
91k	560R	6.12m	44.27	91.6k	557	5.43k	9.39k	1.73
18k	110R	6.07m	44.33	18.1k	109	1.07k	1.85k	1.72
36k	220R	6.07m	44.33	36.2k	219	2.15k	3.69k	1.72
33k	200R	6.02m	44.40	33.2k	199	1.97k	3.35k	1.71
20k	120R	5.96m	44.49	20.1k	119	1.19k	2.01k	1.69
30k	180R	5.96m	44.49	30.2k	179	1.79k	3.02k	1.69
27k	160R	5.89m	44.60	27.2k	159	1.61k	2.68k	1.67
22k	130R	5.87m	44.62	22.1k	129	1.31k	2.18k	1.66
56k	330R	5.86m	44.64	56.3k	328	3.34k	5.54k	1.66
51k	300R	5.85m	44.66	51.3k	298	3.04k	5.03k	1.66

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
62k	360R	5.77m	44.77	62.4k	358	3.70k	6.04k	1.63
47k	270R	5.71m	44.86	47.3k	268	2.80k	4.53k	1.62
68k	390R	5.70m	44.88	68.4k	388	4.05k	6.54k	1.61
75k	430R	5.70m	44.88	75.4k	428	4.47k	7.22k	1.61
82k	470R	5.70m	44.88	82.5k	467	4.89k	7.89k	1.61
39k	220R	5.61m	45.02	39.2k	219	2.32k	3.69k	1.59
91k	510R	5.57m	45.08	91.5k	507	5.42k	8.56k	1.58
100k	560R	5.57m	45.08	101.0k	557	5.96k	9.40k	1.58
43k	240R	5.55m	45.11	43.2k	239	2.56k	4.03k	1.57
18k	100R	5.52m	45.15	18.1k	99.4	1.07k	1.68k	1.56
27k	150R	5.52m	45.15	27.2k	149	1.61k	2.52k	1.56
36k	200R	5.52m	45.15	36.2k	199	2.14k	3.36k	1.56
20k	110R	5.47m	45.24	20.1k	109	1.19k	1.85k	1.55
22k	120R	5.42m	45.31	22.1k	119	1.31k	2.01k	1.54
33k	180R	5.42m	45.31	33.2k	179	1.97k	3.02k	1.54
24k	130R	5.39m	45.37	24.1k	129	1.43k	2.18k	1.53
56k	300R	5.33m	45.47	56.3k	298	3.34k	5.04k	1.51
30k	160R	5.31m	45.51	30.2k	159	1.79k	2.69k	1.50
62k	330R	5.29m	45.52	62.3k	328	3.69k	5.54k	1.50
51k	270R	5.27m	45.57	51.3k	269	3.04k	4.53k	1.49
68k	360R	5.27m	45.57	68.4k	358	4.05k	6.04k	1.49
82k	430R	5.22m	45.65	82.4k	428	4.88k	7.22k	1.48
75k	390R	5.17m	45.72	75.4k	388	4.47k	6.55k	1.47
91k	470R	5.14m	45.78	91.5k	468	5.42k	7.89k	1.46
39k	200R	5.10m	45.85	39.2k	199	2.32k	3.36k	1.45
43k	220R	5.09m	45.87	43.2k	219	2.56k	3.69k	1.44
47k	240R	5.08m	45.88	47.2k	239	2.80k	4.03k	1.44
100k	510R	5.07m	45.89	101.0k	507	5.96k	8.56k	1.44
20k	100R	4.98m	46.06	20.1k	99.5	1.19k	1.68k	1.41
22k	110R	4.98m	46.06	22.1k	109	1.31k	1.85k	1.41
24k	120R	4.98m	46.06	24.1k	119	1.43k	2.02k	1.41
30k	150R	4.98m	46.06	30.2k	149	1.79k	2.52k	1.41
36k	180R	4.98m	46.06	36.2k	179	2.14k	3.02k	1.41
68k	330R	4.83m	46.32	68.3k	328	4.05k	5.54k	1.37
33k	160R	4.83m	46.33	33.2k	159	1.96k	2.69k	1.37
62k	300R	4.82m	46.35	62.3k	299	3.69k	5.04k	1.36
56k	270R	4.80m	46.38	56.3k	269	3.33k	4.53k	1.36

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
27k	130R	4.79m	46.39	27.1k	129	1.61k	2.18k	1.36
75k	360R	4.78m	46.42	75.4k	358	4.47k	6.05k	1.35
82k	390R	4.73m	46.50	82.4k	388	4.88k	6.55k	1.34
91k	430R	4.70m	46.55	91.4k	428	5.42k	7.22k	1.33
51k	240R	4.68m	46.59	51.2k	239	3.04k	4.03k	1.33
100k	470R	4.68m	46.60	100.0k	468	5.95k	7.89k	1.33
47k	220R	4.66m	46.63	47.2k	219	2.80k	3.70k	1.32
43k	200R	4.63m	46.69	43.2k	199	2.56k	3.36k	1.31
39k	180R	4.59m	46.76	39.2k	179	2.32k	3.02k	1.30
24k	110R	4.56m	46.82	24.1k	109	1.43k	1.85k	1.29
22k	100R	4.52m	46.89	22.1k	99.5	1.31k	1.68k	1.28
33k	150R	4.52m	46.89	33.2k	149	1.96k	2.52k	1.28
27k	120R	4.42m	47.08	27.1k	119	1.61k	2.02k	1.25
36k	160R	4.42m	47.08	36.2k	159	2.14k	2.69k	1.25
68k	300R	4.39m	47.15	68.3k	299	4.05k	5.04k	1.25
75k	330R	4.38m	47.17	75.3k	329	4.46k	5.54k	1.24
82k	360R	4.37m	47.19	82.4k	358	4.88k	6.05k	1.24
62k	270R	4.34m	47.26	62.3k	269	3.69k	4.54k	1.23
30k	130R	4.31m	47.30	30.1k	129	1.79k	2.18k	1.22
51k	220R	4.30m	47.34	51.2k	219	3.03k	3.70k	1.22
100k	430R	4.28m	47.37	100.0k	428	5.95k	7.23k	1.21
56k	240R	4.27m	47.40	56.2k	239	3.33k	4.03k	1.21
91k	390R	4.27m	47.40	91.4k	388	5.42k	6.55k	1.21
47k	200R	4.24m	47.46	47.2k	199	2.80k	3.36k	1.20
43k	180R	4.17m	47.60	43.2k	179	2.56k	3.03k	1.18
24k	100R	4.15m	47.64	24.1k	99.6	1.43k	1.68k	1.18
36k	150R	4.15m	47.64	36.2k	149	2.14k	2.52k	1.18
39k	160R	4.09m	47.77	39.2k	159	2.32k	2.69k	1.16
27k	110R	4.06m	47.83	27.1k	110	1.61k	1.85k	1.15
82k	330R	4.01m	47.94	82.3k	329	4.88k	5.55k	1.14
30k	120R	3.98m	47.99	30.1k	120	1.78k	2.02k	1.13
75k	300R	3.98m	47.99	75.3k	299	4.46k	5.04k	1.13
68k	270R	3.95m	48.06	68.3k	269	4.05k	4.54k	1.12
91k	360R	3.94m	48.09	91.4k	359	5.41k	6.05k	1.12
33k	130R	3.92m	48.13	33.1k	129	1.96k	2.19k	1.11
56k	220R	3.91m	48.15	56.2k	219	3.33k	3.70k	1.11
51k	200R	3.91m	48.16	51.2k	199	3.03k	3.36k	1.11

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
100k	390R	3.88m	48.21	100.0k	388	5.95k	6.56k	1.10
62k	240R	3.86m	48.28	62.2k	239	3.69k	4.03k	1.09
39k	150R	3.83m	48.33	39.2k	149	2.32k	2.52k	1.09
47k	180R	3.82m	48.37	47.2k	179	2.80k	3.03k	1.08
43k	160R	3.71m	48.62	43.2k	159	2.56k	2.69k	1.05
27k	100R	3.69m	48.66	27.1k	99.6	1.61k	1.68k	1.05
30k	110R	3.65m	48.75	30.1k	110	1.78k	1.85k	1.04
82k	300R	3.65m	48.77	82.3k	299	4.88k	5.04k	1.03
33k	120R	3.62m	48.82	33.1k	120	1.96k	2.02k	1.03
91k	330R	3.61m	48.84	91.3k	329	5.41k	5.55k	1.03
36k	130R	3.60m	48.88	36.1k	130	2.14k	2.19k	1.02
75k	270R	3.59m	48.91	75.3k	269	4.46k	4.54k	1.02
100k	360R	3.59m	48.91	100.0k	359	5.95k	6.05k	1.02
56k	200R	3.56m	48.97	56.2k	199	3.33k	3.36k	1.01
62k	220R	3.54m	49.03	62.2k	219	3.69k	3.70k	1.00
51k	180R	3.52m	49.08	51.2k	179	3.03k	3.03k	998m
68k	240R	3.52m	49.08	68.2k	239	4.04k	4.04k	998m
43k	150R	3.48m	49.18	43.2k	149	2.56k	2.52k	987m
47k	160R	3.39m	49.39	47.2k	159	2.79k	2.69k	963m
30k	100R	3.32m	49.57	30.1k	99.7	1.78k	1.68k	943m
33k	110R	3.32m	49.57	33.1k	110	1.96k	1.85k	943m
36k	120R	3.32m	49.57	36.1k	120	2.14k	2.02k	943m
39k	130R	3.32m	49.57	39.1k	130	2.32k	2.19k	943m
100k	330R	3.29m	49.66	100.0k	329	5.94k	5.55k	934m
91k	300R	3.29m	49.67	91.3k	299	5.41k	5.05k	933m
82k	270R	3.28m	49.68	82.3k	269	4.87k	4.54k	932m
68k	220R	3.22m	49.83	68.2k	219	4.04k	3.70k	916m
62k	200R	3.22m	49.86	62.2k	199	3.69k	3.36k	913m
56k	180R	3.20m	49.89	56.2k	179	3.33k	3.03k	910m
75k	240R	3.19m	49.92	75.2k	239	4.46k	4.04k	906m
47k	150R	3.18m	49.95	47.2k	150	2.79k	2.52k	903m
51k	160R	3.13m	50.10	51.2k	159	3.03k	2.69k	888m
39k	120R	3.07m	50.26	39.1k	120	2.32k	2.02k	871m
36k	110R	3.05m	50.32	36.1k	110	2.14k	1.85k	865m
33k	100R	3.02m	50.40	33.1k	99.7	1.96k	1.68k	858m
43k	130R	3.01m	50.42	43.1k	130	2.56k	2.19k	856m
100k	300R	2.99m	50.48	100.0k	299	5.94k	5.05k	849m

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
91k	270R	2.96m	50.58	91.3k	269	5.41k	4.54k	840m
51k	150R	2.93m	50.66	51.2k	150	3.03k	2.52k	833m
68k	200R	2.93m	50.66	68.2k	199	4.04k	3.37k	833m
75k	220R	2.92m	50.68	75.2k	219	4.46k	3.70k	831m
82k	240R	2.92m	50.70	82.2k	239	4.87k	4.04k	829m
62k	180R	2.89m	50.77	62.2k	179	3.68k	3.03k	822m
56k	160R	2.85m	50.91	56.2k	160	3.33k	2.69k	809m
39k	110R	2.81m	51.02	39.1k	110	2.32k	1.85k	799m
43k	120R	2.78m	51.11	43.1k	120	2.56k	2.02k	790m
36k	100R	2.77m	51.15	36.1k	99.7	2.14k	1.68k	787m
47k	130R	2.76m	51.19	47.1k	130	2.79k	2.19k	783m
100k	270R	2.69m	51.40	100.0k	269	5.94k	4.54k	765m
82k	220R	2.68m	51.45	82.2k	219	4.87k	3.70k	760m
56k	150R	2.67m	51.47	56.2k	150	3.33k	2.52k	759m
75k	200R	2.66m	51.50	75.2k	199	4.46k	3.37k	755m
68k	180R	2.64m	51.57	68.2k	180	4.04k	3.03k	750m
91k	240R	2.63m	51.60	91.2k	239	5.41k	4.04k	747m
62k	160R	2.57m	51.79	62.2k	160	3.68k	2.69k	731m
39k	100R	2.56m	51.84	39.1k	99.7	2.32k	1.68k	727m
43k	110R	2.55m	51.86	43.1k	110	2.55k	1.85k	725m
47k	120R	2.55m	51.88	47.1k	120	2.79k	2.02k	723m
51k	130R	2.54m	51.89	51.1k	130	3.03k	2.19k	722m
82k	200R	2.43m	52.28	82.2k	200	4.87k	3.37k	691m
62k	150R	2.41m	52.35	62.2k	150	3.68k	2.53k	686m
91k	220R	2.41m	52.35	91.2k	219	5.41k	3.70k	685m
75k	180R	2.39m	52.42	75.2k	180	4.45k	3.03k	680m
100k	240R	2.39m	52.42	100.0k	239	5.94k	4.04k	680m
51k	120R	2.35m	52.59	51.1k	120	3.03k	2.02k	667m
68k	160R	2.35m	52.59	68.2k	160	4.04k	2.69k	667m
47k	110R	2.33m	52.63	47.1k	110	2.79k	1.85k	663m
43k	100R	2.32m	52.69	43.1k	99.8	2.55k	1.68k	659m
56k	130R	2.32m	52.71	56.1k	130	3.33k	2.19k	658m
68k	150R	2.20m	53.15	68.2k	150	4.04k	2.53k	626m
100k	220R	2.20m	53.17	100.0k	220	5.94k	3.70k	624m
91k	200R	2.19m	53.18	91.2k	200	5.40k	3.37k	623m
82k	180R	2.19m	53.19	82.2k	180	4.87k	3.03k	622m
51k	110R	2.15m	53.34	51.1k	110	3.03k	1.85k	612m

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
56k	120R	2.14m	53.40	56.1k	120	3.33k	2.02k	608m
75k	160R	2.13m	53.44	75.2k	160	4.45k	2.69k	605m
47k	100R	2.12m	53.46	47.1k	99.8	2.79k	1.68k	603m
62k	130R	2.09m	53.59	62.1k	130	3.68k	2.19k	595m
75k	150R	2.00m	54.00	75.2k	150	4.45k	2.53k	567m
100k	200R	2.00m	54.00	100.0k	200	5.94k	3.37k	567m
91k	180R	1.97m	54.09	91.2k	180	5.40k	3.03k	561m
56k	110R	1.96m	54.15	56.1k	110	3.32k	1.85k	557m
51k	100R	1.96m	54.17	51.1k	99.8	3.03k	1.68k	556m
82k	160R	1.95m	54.21	82.2k	160	4.87k	2.69k	554m
62k	120R	1.93m	54.28	62.1k	120	3.68k	2.02k	549m
68k	130R	1.91m	54.39	68.1k	130	4.04k	2.19k	542m
82k	150R	1.83m	54.77	82.2k	150	4.87k	2.53k	519m
100k	180R	1.80m	54.91	100.0k	180	5.94k	3.03k	511m
56k	100R	1.78m	54.98	56.1k	99.8	3.32k	1.68k	507m
62k	110R	1.77m	55.04	62.1k	110	3.68k	1.85k	504m
68k	120R	1.76m	55.08	68.1k	120	4.04k	2.02k	501m
91k	160R	1.76m	55.11	91.2k	160	5.40k	2.70k	499m
75k	130R	1.73m	55.24	75.1k	130	4.45k	2.19k	492m
91k	150R	1.65m	55.67	91.2k	150	5.40k	2.53k	468m
68k	110R	1.62m	55.84	68.1k	110	4.04k	1.85k	459m
62k	100R	1.61m	55.86	62.1k	99.8	3.68k	1.68k	458m
75k	120R	1.60m	55.93	75.1k	120	4.45k	2.02k	454m
100k	160R	1.60m	55.93	100.0k	160	5.93k	2.70k	454m
82k	130R	1.58m	56.01	82.1k	130	4.87k	2.19k	450m
100k	150R	1.50m	56.49	100.0k	150	5.93k	2.53k	426m
68k	100R	1.47m	56.66	68.1k	99.9	4.04k	1.69k	418m
75k	110R	1.46m	56.69	75.1k	110	4.45k	1.85k	417m
82k	120R	1.46m	56.71	82.1k	120	4.87k	2.02k	416m
91k	130R	1.43m	56.91	91.1k	130	5.40k	2.19k	406m
82k	110R	1.34m	57.46	82.1k	110	4.87k	1.85k	381m
75k	100R	1.33m	57.51	75.1k	99.9	4.45k	1.69k	379m
91k	120R	1.32m	57.61	91.1k	120	5.40k	2.02k	375m
100k	130R	1.30m	57.73	100.0k	130	5.93k	2.19k	369m
82k	100R	1.22m	58.29	82.1k	99.9	4.86k	1.69k	346m
91k	110R	1.21m	58.36	91.1k	110	5.40k	1.85k	343m
100k	120R	1.20m	58.43	100.0k	120	5.93k	2.02k	341m

R_1	R_2	A	dB	R_I	R_O	R_{SMAX}	R_{LMIN}	R_{LS}(1 dB)
100k	110R	1.10m	59.18	100.0k	110	5.93k	1.85k	313m
91k	100R	1.10m	59.19	91.1k	99.9	5.40k	1.69k	312m
100k	100R	999u	60.01	100.0k	99.9	5.93k	1.69k	284m

This Page Intentionally Left Blank

Part Three

Reactive Circuits

This Page Intentionally Left Blank

9 First-order CR and LR circuits

Introduction

There are two first-order CR and two first-order LR circuits which we will consider here. One circuit of each type forms an integrator, or a low-pass filter, and one a differentiator, or a high-pass filter. All four circuits can be considered to be variations on the voltage dividers which we looked at in Chapter 8. In each case, either the input resistor or the output resistor is replaced with either a capacitor or an inductor.

Where we call them integrators and differentiators we will look at their output waveforms when their input is a pulse. This is called the time response. Then we will consider their action as filters; that is, we will look at the amplitude and phase of their outputs when the input is a sine wave of varying frequency. This is the frequency response.

The two approaches are in fact different ways of looking at the same thing, which is their transfer function – in other words their output for any given input. Both methods can be both understood and quantified without much mathematics, and one, other or both is usually very helpful when we need to use them in a circuit, or understand their use.

Time response

A quick explanation of CR and LR circuit waveforms

The four circuits, with their associated waveforms, are shown in Figure 9.1.

We assume that the resistance of the source is negligible, and the load on the circuit is high enough to be ignored. Both CR circuits have the same current waveform, and so do both LR circuits. (As discussed in Chapter 5, the current in a series circuit is the same through each of its components, and putting them in a different order makes no difference to that current.) Also, the sum of the capacitor's and the resistor's, or the inductor's and the resistor's, voltages always equals the input voltage (by Kirchhoff's voltage law).

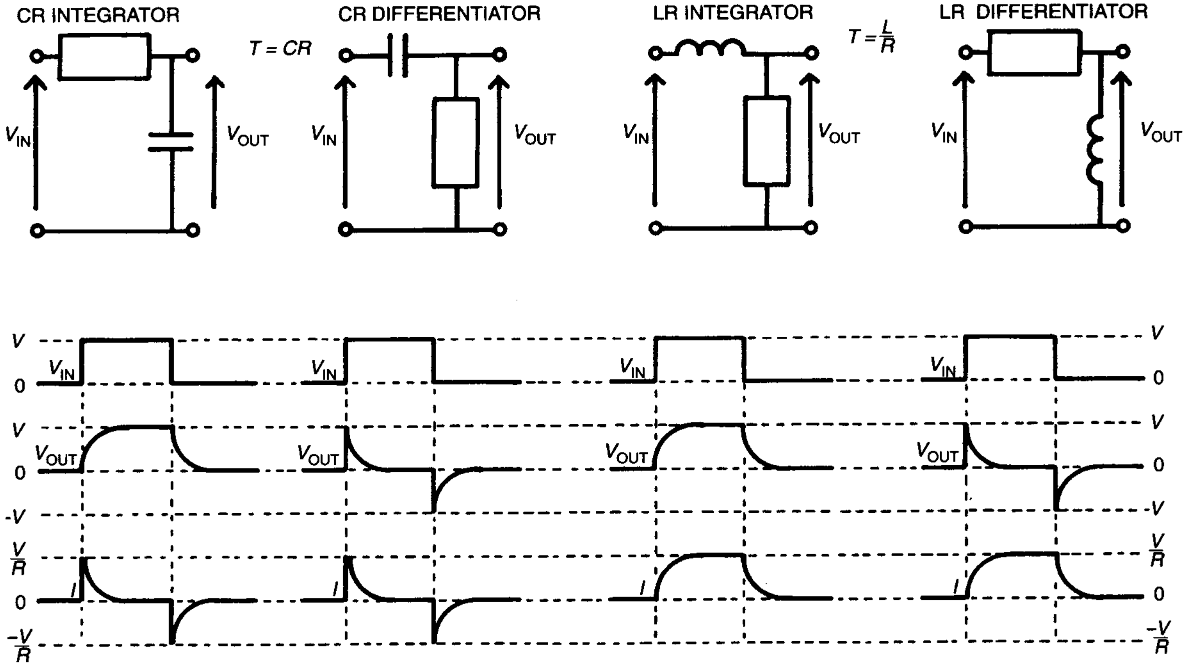

CR INTEGRATOR
CR DIFFERENTIATOR
LR INTEGRATOR
LR DIFFERENTIATOR
T = CR
T = L/R
V_{IN}
V_{OUT}
V
0
-V
V/R
-V/R
I

Figure 9.1 First-order circuits – time response

A capacitor will always resist a sudden change in the voltage across its terminals. Hence, when the leading edge of the pulse arrives, its voltage remains, initially, at zero. All of the available voltage is, at that moment, across the resistor. Therefore the current at that moment is $I = \frac{V}{R}$.

The current charges the capacitor, causing its voltage to increase from zero. As this happens the voltage across the resistor decreases. The current (proportional always to the voltage across the resistor) falls accordingly. This in turn slows down the charge rate of the capacitor. Hence the rate of change of the voltage across the capacitor decreases as that voltage approaches the input voltage. The current in the circuit falls from $I = \frac{V}{R}$ towards zero, and the voltage across the resistor falls from V towards zero.

Eventually, the capacitor is fully charged, and the voltage across it is equal to the input voltage. Things will stay like this until the input changes again. When it falls to zero the voltage across the cap cannot change instantly; it must now discharge, back through the resistor, and into the source (the only possible path – remember that the source has a low, and the load a high, resistance). Hence a current of $\frac{-V}{R}$ now flows for an instant (the minus sign indicating that the direction has reversed). This current decreases towards zero again as the voltage which created it, that across the plates of the capacitor, falls. Eventually, there is no charge left in the capacitor, no voltage across its terminals and no current in the circuit. The circuit is once again at rest, as it was in the beginning.

Because an inductor resists a sudden change in current, an LR circuit's current is at zero when the leading edge of the pulse arrives. So then is the voltage across the resistor at that moment. For the sum of the resistor's and the inductor's voltage to equal the input voltage, as it must, the inductor sets up a back EMF across its terminals, opposing the input voltage. The back EMF decays with time, and current begins to flow in the circuit, creating a voltage across the resistor. Eventually there is no back EMF, and the current arrives at a final value of $\frac{V}{R}$. The voltage across the inductor is now zero, and the voltage across the resistor is V.

When the trailing edge of the pulse arrives at the input (i.e. the input voltage falls towards zero again) the current in the circuit also will try to fall to zero. Again the inductor will resist change. To do this it will set up another back EMF, opposing the voltage across the resistor now. This back EMF will also decay with time, allowing the current in the circuit and the voltage across the resistor to fall towards zero.

The only difference between the two CR (or the two LR) circuits is which device, resistor or capacitor (inductor), is on the output side. The voltages across the components, and the currents through them, are identical.

Time constants It can be seen that, essentially, only two waveform shapes exist for both output voltage and for current in all four circuits. These two waveforms are shown in more detail in Figure 9.2.

All of the curves are exponential in shape. How long they take to change from initial to final values is determined by the circuit time constant T (not to be confused with the T used for period of an AC waveform), which is in turn a function of component values in the circuit.

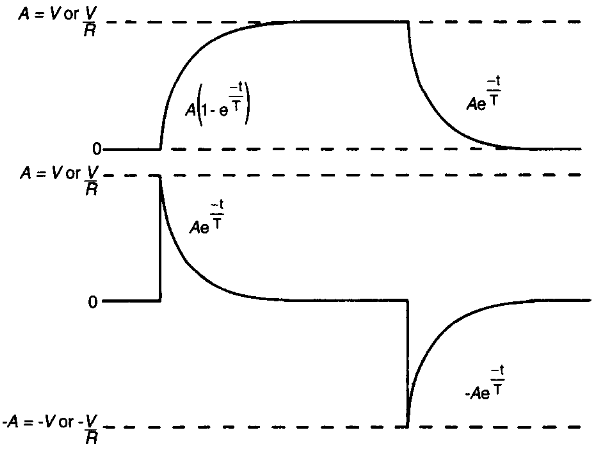

Figure 9.2 Detail of Figure 9.1

The formulae for T are:

for a CR circuit:

$$T = CR \tag{9.1}$$

and for an LR circuit:

$$T = \frac{L}{R} \tag{9.2}$$

T is tabulated for various values of C and R in Table 9.1, and for L and R in Table 9.2. These tables also give values of frequency constants which we will look at later under 'Frequency response'.

The tables are sorted in order of T, increasing for CR and decreasing for LR. T is given in μs. Should we want to obtain a value of T which is outside the range of the table we should do it by scaling one or both components by factors of 10 appropriately. As always, some examples:

Example 9.1: Find values for a CR circuit with T = 56 ns.

From Table 9.1:

f(Hz)	T(μs)	R	C
2842	56.00	1k	56n
2842	56.00	5k6	10n

We want T = 56 ns, so we need to make R and C between them 1000 times smaller. We could for example make $R = 100\ \Omega$ and C = 560p, or $R = 56\ \Omega$ and C = 1n.

Example 9.2: Find values for an LR circuit with $T = 4\,\mu$s.

From Table 9.2:

f(Hz)	$T(\mu$s)	R	L
4063	39.17	1k2	47m

To make T 10 times smaller we could either make R 10 times larger or L 10 times smaller. So R = 12k and L = 47m or R = 1k2 and L = 4.7m would both work.

Values of voltage and current as a function of t and T Once we have determined a value of T, we can find the value of V_t or I_t, that is output voltage for any moment in time, t, after the rising or falling edge of the input waveform. The formulae are as follows.

For a curve going from zero to a final value:

$$V_t \text{ or } I_t = \left(V \text{ or } \frac{V}{R}\right)\left(1 - e^{\frac{-t}{T}}\right) \tag{9.3}$$

and for a curve going back down to zero:

$$V_t \text{ or } I_t = (-)\left(V \text{ or } \frac{V}{R}\right)e^{\frac{-t}{T}} \tag{9.4}$$

Table 9.3 gives values of Eqns (9.3) and (9.4) for various values of $\frac{t}{T}$.

Example 9.3: Find out to what the voltage out from a 2k2, 1μF differentiator will be after 6 ms, when the input is a rising edge of 6 V:

First we calculate T, from Eqn(9.1).

$T = CR = 2\text{k}2 \times 10\text{u} = 22\,\text{ms}$

Now we evaluate $\frac{t}{T}$. If t = 6 ms, $\frac{t}{T} = \frac{6}{22} = 0.27$. Referring to Figures 9.1 and 9.2, we see that the waveform is a rising edge, governed by the $1 - e^{\frac{-t}{T}}$ function. So we use Table 9.1 to find a value for that function where $\frac{t}{T} = 0.27$. We find:

$\frac{t}{T}$	$1 - e^{\frac{-t}{T}}$	$e^{\frac{-t}{T}}$
0.3	0.7408	0.2592

which is the nearest to 0.27. The final value of the voltage will be 6 V. So its value after 6 ms will be 0.741 × 6 = 4.45 V.

Example 9.4: Design a circuit whose output voltage will be −7 V 220 ms after the falling edge of a + 10 V pulse input. Give a CR and an LR alternative.

By referring to Figure 9.1, we see that the circuit must be a differentiator, as that circuit has an output voltage which goes from $-V$ to 0 after the falling edge of the input. The equation for that curve is $-Ve^{\frac{-t}{T}}$, where $V = 10$ V in this case. So we can write:

$-7 = -10\,e^{\frac{-t}{T}}$, or $e^{\frac{-t}{T}} = 0.7$

and we can use Table 9.1 in reverse to find:

$\frac{t}{T}$	$1 - e^{\frac{-t}{T}}$	$e^{\frac{-t}{T}}$
1.2	0.3012	0.6988

Now we know that $t = 220$ ms, and that $\frac{t}{T} = 1.2$, so we know that $T = \frac{220}{1.2} = 183.3$ ms. Now we can use Table 9.3 to find suitable C and R values, and Table 9.4 to find L and R.

From Table 9.3:

f(Hz)	T(μs)	R	C
8682	18.33	4k7	3n9

T is in μs, so we need to scale up by $\frac{183.3\,\text{m}}{18.33\,\mu} = 10\,000$. We could do this by multiplying both C and R by 100. So we get $R = 470$ k, $C = 390$ n.

From Table 9.4:

f(Hz)	T(μs)	R	L
8681	18.33	1k8	33m

Here we need to scale by making R smaller and L larger. We could say theoretical values of $R = 18\ \Omega$ and $L = 3.3$ H. A 3.3 H inductor would be huge (though not unknown!), which illustrates why CR circuits tend to be more common than LR ones.

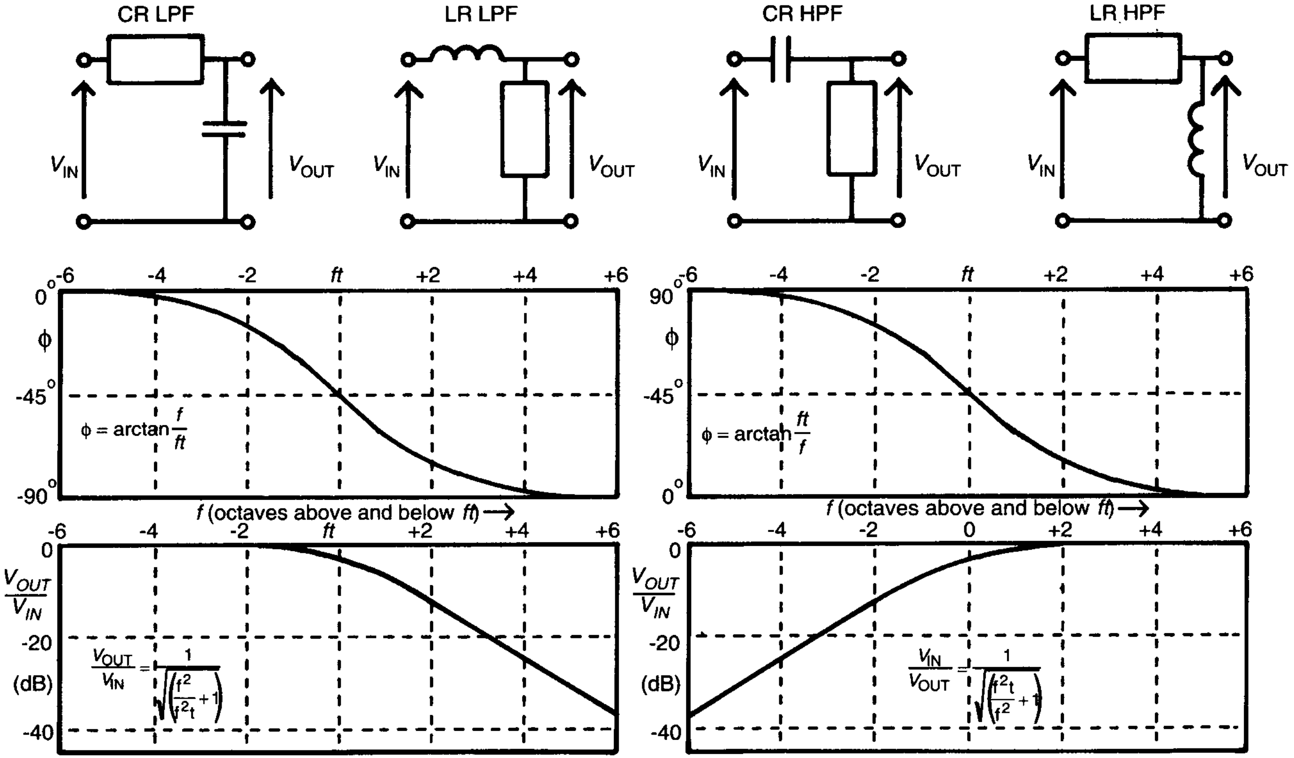

Figure 9.3 First-order circuits – frequency response

Frequency response

Now we look at the behaviour of the same circuits when their input is a continuous sine wave. We are interested in what happens as the frequency of the input is varied.

How first-order filters work

As we saw in Chapter 4, capacitive reactance falls with increasing frequency:

$$X_{\mathrm{C}} = \frac{1}{2\pi fC} \tag{9.5}$$

Inductive reactance increases with frequency:

$$X_{\mathrm{L}} = 2\pi fL \tag{9.6}$$

Each obeys Ohm's Law, in that $V_{\mathrm{C}} = IX_{\mathrm{C}}$, or $X_{\mathrm{L}} = IX_{\mathrm{L}}$ – that is the voltage across a capacitor or an inductor is the current through it multiplied by its reactance, like a resistor. However, whereas voltage and current are in phase with a resistor, they are 90° out of phase with inductors and capacitors.

The four basic filter circuits are shown in Figure 9.3. We can view them all as being similar to the voltage divider which we looked at in Chapter 6, except that one component in each has a 'resistance' which varies with frequency. Hence the overall attenuation of each varies with frequency too.

Where the component on the input side of the circuit has higher impedance at higher frequencies, the circuit is a low-pass filter. These are the two circuits on the left hand side of Figure 9.3, a CR with the capacitor across the output or an LR with the inductor on the input. Conversely, a high-pass filter is formed by a CR with the capacitor on the input side or by an LR with the inductor across the output.

Frequency constants The frequency at which the reactance of the capacitor or inductor is the equal to the resistor is known as the 'turnover frequency' (f_{t}). The attenuation at this frequency is not 6 dB, as it would be if both components were resistive. Because of the fact that the voltage across the reactive component is 90° out of phase with the voltage across the resistor the attenuation is only 3 dB. The turnover frequency can be calculated from the component values (or component values selected to give a desired f_{t}). The governing formulae are:

For a CR circuit:

$$R = \frac{1}{2\pi f_{\mathrm{t}} C}$$

which becomes

$$f_t = \frac{1}{2\pi RC} \tag{9.7}$$

For an LR circuit:

$$R = 2\pi f_t L$$

or

$$f_t = \frac{R}{2\pi L} \tag{9.8}$$

Values for f_t are given in Table 9.1 for CR circuits and in Table 9.2 for LR circuits. As before, component values can be scaled to get frequencies which are outside the range of the table.

Example 9.5: Give component values for a CR filter with $f_t = 12\,\text{kHz}$.

From Table 9.1:

f(Hz)	$T(\mu s)$	R	C
1205	132.00	1k1	120n

We need to multiply the frequency by 10, so we could divide either C or R by 10 to achieve this. So both $R = 110\,\Omega, C = 120\text{n}$ and $R = 1\text{k1}$, $C = 12\text{n}$ would do it.

Example 9.6: Give component values for an LR filter with $f_t = 500\text{kHz}$.

From Table 9.2:

f(Hz)	$T(\mu s)$	R	L
5079	31.33	1k5	47m

We need f_t 100 times larger, so we can make L smaller and/or R larger. We could choose, for instance, $R = 15\text{k}$, $L = 4.7\text{m}$.

Determining the level and phase at any frequency The level of the output voltage, relative to the input, is determined by the ratio of the signal frequency and f_t:

For a LPF:

$$\frac{V_{OUT}}{V_{IN}} = \frac{1}{\sqrt{\left(\frac{f}{f_t}\right)^2 + 1}} \tag{9.9}$$

We can see that when f is much less than f_t, the value of $\frac{V_{OUT}}{V_{IN}}$ is very close to 1, and hence the signal is not much attenuated by the circuit. When $\frac{f}{f_t}$ is much greater than 1, the value of $\frac{V_{OUT}}{V_{IN}}$ approximates to:

$$\frac{V_{OUT}}{V_{IN}} = \frac{1}{\frac{f}{f_t}} = \frac{f_t}{f} \tag{9.10}$$

From this simplified expression, it can be seen that a doubling in frequency will lead to a halving in output signal, or multiplying the input frequency by 10 will reduce its output level to one-tenth. Hence it is customary to say that the filter slope, above the turnover frequency, is −6 db/octave, or −20 dB/decade.

The equation for the attenuation of a HPF is:

$$\frac{V_{OUT}}{V_{IN}} = \frac{1}{\sqrt{\left(\frac{f_t}{f}\right)^2 + 1}} \tag{9.11}$$

and a similar argument applies; this time the slope is −6 dB/octave or −20 dB/decade below the turnover frequency.

These filters affect the relative phase (ϕ), as well as the level, of the signal. The LPF causes the output signal to be between 0 and 90° lagging the input signal, the HPF 0° to 90° leading. In both cases the phase angle is 45° at f_t. These functions are again determined by the ratio of f and f_t:

For a HPF:

$$\phi\,(\text{leading}) = \arctan\left(\frac{f_t}{f}\right) \tag{9.12}$$

For a LPF:

$$\phi\,(\text{lagging}) = \arctan\left(\frac{f}{f_t}\right) \tag{9.13}$$

Hence it is possible to predict both attenuation and phase shift through a filter for any given frequency, provided that f_t is known. Table 9.4 gives both at intervals of a third of an octave, for six octaves either side of f_t.

The first column of Table 9.4 gives the frequency in terms of the number of octaves away from f_t. (The minus sign means higher in frequency for an HPF, lower for an LPF.) The second gives the ratio of frequencies, which we call F (f_t/f for an HPF, f/f_t for an LPF). The third column is the attenuation as the ratio V_{OUT}/V_{IN}, the fourth is attenuation in dBs. The fifth column is the phase shift in degrees.

Example 9.6: What is the approximate attenuation and phase shift at a frequency of 5 kHz for a LPF consisting of a 3n3 capacitor and a 15k resistor?

$$f_t = \frac{1}{2\pi \times 15k \times 3n3} = 3215\,\text{Hz}$$

$$F = \frac{f}{f_t} = \frac{5000}{3215} = 1.56$$

From Table 9.4:

Octaves	F	Att	dB	$\phi°$
0 2/3	01.59	0.533	05.47	57.8

Nearest value in the table is 1.59 (2/3 octave) where attenuation is 5.5 dB, phase shift is 58°.

Example 9.7: A circuit is needed to low-pass filter a signal such that the level at 100 kHz is 15 dB down on the signal below f_t. What would f_t be (using the table to get the nearest approximation)? What will the attenuation be at 50 kHz?

From Table 9.4:

Octaves	F	Att	dB	$\phi°$
2 1/3	05.04	0.195	14.22	78.8

This tells us that for a 14 dB attenuation at 100 kHz, $F = \frac{f}{f_t} = 5$, so $f_t = \frac{100\,\text{kHz}}{5}$, or 20 kHz. At 50 kHz $F = \frac{50}{20}$ or 2.5, and we get:

Octaves	F	Att	dB	$\phi°$
0 2/3	01.59	0.533	05.47	57.8

and the attenuation is 5.5 dB.

Table 9.1 *Time and frequency constants for first-order CR circuits*

f(Hz)	T(μs)	R	C
1020	156.00	1k3	120n
1026	155.10	3k3	47n
1026	155.10	4k7	33n
1046	152.10	3k9	39n
1053	151.20	2k7	56n
1053	151.20	5k6	27n
1061	150.00	1k	150n
1061	150.00	1k5	100n
1064	149.60	2k2	68n
1064	149.60	6k8	22n
1078	147.60	1k8	82n
1078	147.60	8k2	18n
1105	144.00	1k2	120n
1122	141.90	4k3	33n
1129	141.00	3k	47n
1134	140.40	3k6	39n
1156	137.70	5k1	27n
1166	136.50	9k1	15n
1167	136.40	6k2	22n
1170	136.00	2k	68n
1179	135.00	7k5	18n
1184	134.40	2k4	56n
1206	132.00	1k1	120n
1213	131.20	1k6	82n
1224	130.00	1k3	100n
1237	128.70	3k3	39n
1237	128.70	3k9	33n
1254	126.90	2k7	47n
1254	126.90	4k7	27n
1292	123.20	2k2	56n
1292	123.20	5k6	22n
1294	123.00	1k5	82n
1294	123.00	8k2	15n
1300	122.40	6k8	18n
1300	122.40	1k8	68n
1326	120.00	1k	120n
1326	120.00	1k2	100n
1340	118.80	3k6	33n
1360	117.00	3k	39n
1371	116.10	4k3	27n
1411	112.80	2k4	47n
1415	112.50	7k5	15n
1418	112.20	5k1	22n
1421	112.00	2k	56n
1426	111.60	6k2	18n
1447	110.00	1k1	100n
1457	109.20	9k1	12n
1461	108.90	3k3	33n
1463	108.80	1k6	68n
1493	106.60	1k3	82n
1511	105.30	3k9	27n
1511	105.30	2k7	39n
1539	103.40	2k2	47n
1539	103.40	4k7	22n
1560	102.00	6k8	15n
1560	102.00	1k5	68n
1579	100.80	1k8	56n
1579	100.80	5k6	18n
1592	100.00	1k	100n
1608	99.00	3k	33n
1617	98.40	8k2	12n
1617	98.40	1k2	82n
1637	97.20	3k6	27n
1682	94.60	4k3	22n
1693	94.00	2k	47n
1700	93.60	2k4	39n
1711	93.00	6k2	15n
1734	91.80	5k1	18n
1749	91.00	9k1	10n
1764	90.20	1k1	82n
1768	90.00	7k5	12n
1776	89.60	1k6	56n

f(Hz)	T(μs)	R	C
1786	89.10	2k7	33n
1786	89.10	3k3	27n
1800	88.40	1k3	68n
1855	85.80	2k2	39n
1855	85.80	3k9	22n
1881	84.60	1k8	47n
1881	84.60	4k7	18n
1895	84.00	1k5	56n
1895	84.00	5k6	15n
1941	82.00	1k	82n
1941	82.00	8k2	10n
1950	81.60	6k8	12n
1950	81.60	1k2	68n
1965	81.00	3k	27n
2010	79.20	2k4	33n
2010	79.20	3k6	22n
2040	78.00	2k	39n
2056	77.40	4k3	18n
2080	76.50	5k1	15n
2116	75.20	1k6	47n
2122	75.00	7k5	10n
2128	74.80	1k1	68n
2133	74.62	9k1	8n2
2139	74.40	6k2	12n
2183	72.90	2k7	27n
2186	72.80	1k3	56n
2192	72.60	2k2	33n
2192	72.60	3k3	22n
2258	70.50	1k5	47n
2258	70.50	4k7	15n
2267	70.20	1k8	39n
2267	70.20	3k9	18n
2341	68.00	1k	68n
2341	68.00	6k8	10n
2367	67.24	8k2	8n2
2368	67.20	1k2	56n
2368	67.20	5k6	12n
2411	66.00	2k	33n
2411	66.00	3k	22n
2456	64.80	2k4	27n
2456	64.80	3k6	18n
2468	64.50	4k3	15n
2551	62.40	1k6	39n
2567	62.00	6k2	10n
2572	61.88	9k1	6n8
2584	61.60	1k1	56n
2588	61.50	7k5	8n2
2601	61.20	5k1	12n
2605	61.10	1k3	47n
2679	59.40	2k2	27n
2679	59.40	3k3	18n
2679	59.40	1k8	33n
2679	59.40	2k7	22n
2721	58.50	3k9	15n
2721	58.50	1k5	39n
2822	56.40	4k7	12n
2822	56.40	1k2	47n
2842	56.00	1k	56n
2842	56.00	5k6	10n
2854	55.76	6k8	8n2
2854	55.76	8k2	6n8
2947	54.00	2k	27n
2947	54.00	3k6	15n
2947	54.00	3k	18n
3014	52.80	1k6	33n
3014	52.80	2k4	22n
3078	51.70	1k1	47n
3084	51.60	4k3	12n
3121	51.00	7k5	6n8
3121	51.00	5k1	10n
3123	50.96	9k1	5n6
3131	50.84	6k2	8n2
3139	50.70	1k3	39n
3215	49.50	1k5	33n

f(Hz)	T(μs)	R	C
3215	49.50	3k3	15n
3275	48.60	1k8	27n
3275	48.60	2k7	18n
3288	48.40	2k2	22n
3386	47.00	1k	47n
3386	47.00	4k7	10n
3401	46.80	1k2	39n
3401	46.80	3k9	12n
3442	46.24	6k8	6n8
3466	45.92	8k2	5n6
3466	45.92	5k6	8n2
3537	45.00	3k	15n
3617	44.00	2k	22n
3684	43.20	1k6	27n
3684	43.20	2k4	18n
3684	43.20	3k6	12n
3701	43.00	4k3	10n
3710	42.90	1k1	39n
3710	42.90	1k3	33n
3721	42.77	9k1	4n7
3775	42.16	6k2	6n8
3789	42.00	7k5	5n6
3806	41.82	5k1	8n2
3930	40.50	1k5	27n
3930	40.50	2k7	15n
4019	39.60	3k3	12n
4019	39.60	1k2	33n
4019	39.60	1k8	22n
4019	39.60	2k2	18n
4081	39.00	1k	39n
4081	39.00	3k9	10n
4130	38.54	4k7	8n2
4130	38.54	8k2	4n7
4179	38.08	6k8	5n6
4179	38.08	5k6	6n8
4384	36.30	1k1	33n
4421	36.00	2k4	15n
4421	36.00	3k	12n
4421	36.00	2k	18n
4421	36.00	3k6	10n
4485	35.49	9k1	3n9
4514	35.26	4k3	8n2
4515	35.25	7k5	4n7
4521	35.20	1k6	22n
4534	35.10	1k3	27n
4584	34.72	6k2	5n6
4589	34.68	5k1	6n8
4823	33.00	1k	33n
4823	33.00	2k2	15n
4823	33.00	1k5	22n
4823	33.00	3k3	10n
4912	32.40	1k2	27n
4912	32.40	2k7	12n
4912	32.40	1k8	18n
4977	31.98	3k9	8n2
4977	31.98	8k2	3n9
4980	31.96	4k7	6n8
4980	31.96	6k8	4n7
5075	31.36	5k6	5n6
5300	30.03	9k1	3n3
5305	30.00	2k	15n
5305	30.00	3k	10n
5359	29.70	1k1	27n
5391	29.52	3k6	8n2
5441	29.25	7k5	3n9
5443	29.24	4k3	6n8
5462	29.14	6k2	4n7
5526	28.80	2k4	12n
5526	28.80	1k6	18n
5565	28.60	1k3	22n
5573	28.56	5k1	5n6
5882	27.06	3k3	8n2
5882	27.06	8k2	3n3
5895	27.00	1k	27n

f(Hz)	T(μs)	R	C
5895	27.00	1k8	15n
5895	27.00	1k5	18n
5895	27.00	2k7	10n
6001	26.52	3k9	6n8
6001	26.52	6k8	3n9
6029	26.40	2k2	12n
6029	26.40	1k2	22n
6047	26.32	4k7	5n6
6047	26.32	5k6	4n7
6431	24.75	7k5	3n3
6470	24.60	3k	8n2
6478	24.57	9k1	2n7
6501	24.48	3k6	6n8
6577	24.20	1k1	22n
6582	24.18	6k2	3n9
6609	24.08	4k3	5n6
6631	24.00	1k6	15n
6631	24.00	2k	12n
6631	24.00	2k4	10n
6640	23.97	5k1	4n7
6801	23.40	1k3	18n
7074	22.50	1k5	15n
7092	22.44	3k3	6n8
7092	22.44	6k8	3n3
7189	22.14	2k7	8n2
7189	22.14	8k2	2n7
7205	22.09	4k7	4n7
7234	22.00	1k	22n
7234	22.00	2k2	10n
7287	21.84	3k9	5n6
7287	21.84	5k6	3n9
7368	21.60	1k2	18n
7368	21.60	1k8	12n
7779	20.46	6k2	3n3
7802	20.40	3k	6n8
7860	20.25	7k5	2n7
7875	20.21	4k3	4n7
7895	20.16	3k6	5n6
7950	20.02	9k1	2n2
7958	20.00	2k	10n
8002	19.89	5k1	3n9
8038	19.80	1k1	18n
8087	19.68	2k4	8n2
8162	19.50	1k3	15n
8289	19.20	1k6	12n
8612	18.48	3k3	5n6
8612	18.48	5k6	3n3
8669	18.36	2k7	6n8
8669	18.36	6k8	2n7
8683	18.33	3k9	4n7
8683	18.33	4k7	3n9
8822	18.04	2k2	8n2
8822	18.04	8k2	2n2
8842	18.00	1k2	15n
8842	18.00	1k5	12n
8842	18.00	1k	18n
8842	18.00	1k8	10n
9406	16.92	3k6	4n7
9457	16.83	5k1	3n3
9474	16.80	3k	5n6
9490	16.77	4k3	3n9
9507	16.74	6k2	2n7
9646	16.50	7k5	2n2
9646	16.50	1k1	15n
9705	16.40	2k	8n2
9716	16.38	9k1	1n8
9752	16.32	2k4	6n8
9947	16.00	1k6	10n

Table 9.2 *Time and frequency constants for first-order LR circuits*

f(Hz)	T(μs)	R	L
1592	100.00	1k	100m
1751	90.91	1k1	100m
1910	83.33	1k2	100m
2069	76.92	1k3	100m
2341	68.00	1k	68m
2387	66.67	1k5	100m
2546	62.50	1k6	100m
2575	61.82	1k1	68m
2809	56.67	1k2	68m
2865	55.56	1k8	100m
3043	52.31	1k3	68m
3183	50.00	2k	100m
3386	47.00	1k	47m
3501	45.45	2k2	100m
3511	45.33	1k5	68m
3725	42.73	1k1	47m
3745	42.50	1k6	68m
3820	41.67	2k4	100m
4064	39.17	1k2	47m
4213	37.78	1k8	68m
4297	37.04	2k7	100m
4402	36.15	1k3	47m
4681	34.00	2k	68m
4775	33.33	3k	100m
4823	33.00	1k	33m
5079	31.33	1k5	47m
5149	30.91	2k2	68m
5252	30.30	3k3	100m
5305	30.00	1k1	33m
5418	29.37	1k6	47m
5617	28.33	2k4	68m
5730	27.78	3k6	100m
5787	27.50	1k2	33m
6095	26.11	1k8	47m
6207	25.64	3k9	100m
6270	25.38	1k3	33m
6319	25.19	2k7	68m
6773	23.50	2k	47m
6844	23.26	4k3	100m
7022	22.67	3k	68m
7234	22.00	1k	22m
7234	22.00	1k5	33m
7450	21.36	2k2	47m
7480	21.28	4k7	100m
7717	20.62	1k6	33m
7724	20.61	3k3	68m
7958	20.00	1k1	22m
8117	19.61	5k1	100m
8127	19.58	2k4	47m
8426	18.89	3k6	68m
8681	18.33	1k2	22m
8681	18.33	1k8	33m
8913	17.86	5k6	100m
9128	17.44	3k9	68m
9143	17.41	2k7	47m
9405	16.92	1k3	22m
9646	16.50	2k	33m
9868	16.13	6k2	100m

Table 9.3 *Step response of first-order CR and LR circuits*

$\frac{t}{T}$	$e^{\frac{-t}{T}}$	$1 - e^{\frac{-t}{T}}$
0.1	0.9048	0.0952
0.2	0.8187	0.1813
0.3	0.7408	0.2592
0.4	0.6703	0.3297
0.5	0.6065	0.3935
0.6	0.5488	0.4512
0.7	0.4966	0.5034
0.8	0.4493	0.5507
0.9	0.4066	0.5934
1.0	0.3679	0.6321
1.1	0.3329	0.6671
1.2	0.3012	0.6988
1.3	0.2725	0.7275
1.4	0.2466	0.7534
1.5	0.2231	0.7769
1.6	0.2019	0.7981
1.7	0.1827	0.8173
1.8	0.1653	0.8347
1.9	0.1496	0.8504
2.0	0.1353	0.8647
2.1	0.1225	0.8775
2.2	0.1108	0.8892
2.3	0.1003	0.8997
2.4	0.0907	0.9093
2.5	0.0821	0.9179
2.6	0.0743	0.9257
2.7	0.0672	0.9328
2.8	0.0608	0.9392
2.9	0.0550	0.9450
3.0	0.0498	0.9502
3.1	0.0450	0.9550
3.2	0.0408	0.9592
3.3	0.0369	0.9631
3.4	0.0334	0.9666
3.5	0.0302	0.9698
3.6	0.0273	0.9727
3.7	0.0247	0.9753
3.8	0.0224	0.9776
3.9	0.0202	0.9798
4.0	0.0183	0.9817
4.1	0.0166	0.9834
4.2	0.0150	0.9850
4.3	0.0136	0.9864
4.4	0.0123	0.9877
4.5	0.0111	0.9889
4.6	0.0101	0.9899
4.7	0.0091	0.9909
4.8	0.0082	0.9918
4.9	0.0074	0.9926
5.0	0.0067	0.9933

Table 9.4 *Frequency and phase response of first-order CR and LR circuits*

Octaves	F	Att	dB	$\phi(^\circ)$
−6	00.02	1.000	00.00	00.9
−6 1/3	00.02	1.000	00.00	01.1
−6 2/3	00.02	1.000	00.00	01.4
−5	00.03	1.000	00.00	01.8
−5 1/3	00.04	0.999	00.01	02.3
−5 2/3	00.05	0.999	00.01	02.8
−4	00.06	0.998	00.02	03.6
−4 1/3	00.08	0.997	00.03	04.5
−4 2/3	00.10	0.995	00.04	05.7
−3	00.12	0.992	00.07	07.1
−3 1/3	00.16	0.988	00.11	09.0
−3 2/3	00.20	0.981	00.17	11.2
−2	00.25	0.970	00.26	14.0
−2 1/3	00.31	0.954	00.41	17.5
−2 2/3	00.40	0.929	00.64	21.6
−1	00.50	0.894	00.97	26.6
−1 1/3	00.63	0.846	01.45	32.2
−1 2/3	00.79	0.783	02.12	38.4
0	01.00	0.707	03.01	45.0
0 1/3	01.26	0.622	04.13	51.6
0 2/3	01.59	0.533	05.47	57.8
1	02.00	0.447	06.99	63.4
1 1/3	02.52	0.369	08.66	68.4
1 2/3	03.17	0.300	10.45	72.5
2	04.00	0.243	12.30	76.0
2 1/3	05.04	0.195	14.22	78.8
2 2/3	06.35	0.156	16.16	81.0
3	08.00	0.124	18.13	82.9
3 1/3	10.08	0.099	20.11	84.3
3 2/3	12.70	0.079	22.10	85.5
4	16.00	0.062	24.10	86.4
4 1/3	20.16	0.050	26.10	87.2
4 2/3	25.40	0.039	28.10	87.7
5	32.00	0.031	30.11	88.2
5 1/3	40.32	0.025	32.11	88.6
5 2/3	50.80	0.020	34.12	88.9
6	64.00	0.016	36.12	89.1

10 LC tuned circuits

Introduction

When inductors and capacitors are combined the circuit is called a tuned circuit. Where the circuits in Chapter 9 were termed first-order, we call these second-order, because there are two frequency-dependent elements rather than one. The capacitor and inductor can be combined in either series or parallel, giving rise to two types which we shall consider separately. These circuits both display an effect known as resonance, by which we mean that their impedance is at a minimum or a maximum at one specific frequency.

Series tuned circuit

A series tuned circuit is an inductor and a capacitor in series. A resistive component is also assumed to exist which may be considered to consist solely of the circuit's residual resistances (the inductor's internal resistance being, usually, the only significant one) or perhaps of residual resistances plus a physical resistor. The circuit is shown in Figure 10.1.

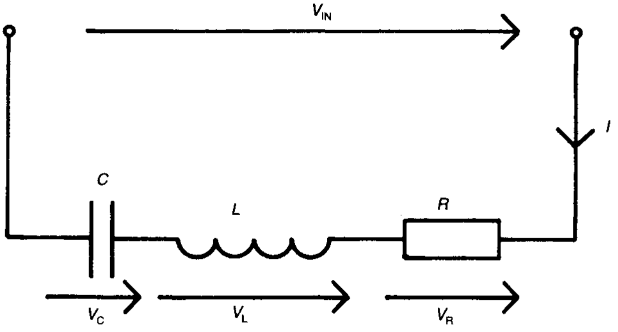

Figure 10.1 Series tuned circuit

What the circuit does

The reactance of the capacitor decreases with frequency, while the reactance of the inductor increases. As we know, we cannot just add the impedance of each component at some frequency together because the phase angles of the voltages will be different. But we can deduce what is happening.

We will take the current in the circuit as being our phase reference, as it is common to all components. The voltage across the capacitor lags it by 90°, the voltage across the inductor leads it by 90° and the voltage across the resistor is in phase with it. The input voltage, which we would divide by input current to find the impedance, will be the phasor sum of these. (If you're a beginner, and this is losing you, look back at 'Calculations on AC quantities' in Chapter 4.)

If X_C decreases with increasing frequency, and X_L increases, then they must be equal at some point. At this frequency the capacitor's and the inductor's voltage are equal, but opposing each other in phase. (Their magnitudes can be calculated by using IX_L and IX_C respectively, and I is the same for each.) This means that they cancel each other – their phasor sum is zero. So the input voltage is equal to the voltage across the resistor – the resistor has the maximum amount of voltage across it that it can have.

This causes the current in the circuit to be at a maximum, and hence impedance is at a minimum, and in phase (resistive). We call this frequency the resonant frequency of the circuit (f_O). Below resonance X_C is greater than X_L, and the impedance is overall capacitive. This means that the current leads the input voltage by some small angle when the impedances are almost equal, and by nearly 90° when the frequency is very low. Above resonance X_L is more significant and the impedance is inductive. Then input voltage leads current by between 0 and 90°.

Calculating resonant frequency

We can deduce the frequency where the two impedances are equal:

$$\frac{1}{2\pi f_O C} = 2\pi f_O L$$

After a little algebra this becomes:

$$f_O = \frac{1}{2\pi\sqrt{LC}} \tag{10.1}$$

The impedance at resonance, which we term Zo is simply:

$$Z_O = R \tag{10.2}$$

Q factor

Although V_C and V_L cancel at resonance, they are still definitely there. In fact they can be larger than the input voltage. This is because the current can be large, given by $\frac{V_{IN}}{R}$, if R is small. The voltage across, say, the capacitor, will be IX_C, or $\frac{V_{IN}}{R} X_C$. We call the ratio of this magnified voltages to the input voltage the Q, or quality factor, of the circuit. Q can be calculated:

$$V_C = \frac{V_{IN}}{R} X_C$$

(because the impedance at resonance is simply R). So:

$$Q = \frac{V_C}{V_{IN}} = \frac{X_C}{R} = \frac{1}{2\pi f_O CR} \tag{10.3}$$

or

$$V_L = IX_L = \frac{V_{IN}}{R} X_L$$

so

$$Q = \frac{V_L}{V_{IN}} = \frac{X_L}{R} = \frac{2\pi f_O L}{R} \tag{10.4}$$

The result should be the same whichever equation we use to calculate Q. From these equations it can be seen that Q can be decreased quite easily by simply adding more resistance in series with the circuit. Conversely the maximum obtainable Q is set by the leftover resistance in the circuit when we have no physical resistor present. The inductor will normally possess far more resistance than the capacitor, so it is common to talk about the Q of an inductor at some specified frequency – this is the highest Q obtainable with that inductor in a tuned circuit with that f_O.

Q as an indication of bandwidth

The fact that the impedance of a series tuned circuit drops to a minimum at resonance makes it very useful for applications where we need some mechanism for picking a certain group of frequencies out of a wideband signal – for instance radio tuners, or audio filters. In these situations we need to know how selective the circuit actually is. The Q factor, conveniently, tells us this – the circuit's bandwidth, or BW. The bandwidth is the difference, in Hz, between the frequencies above and below resonance where the current is 3 dB less than at resonance, for the same level of input voltage – i.e. where the impedance drops by 3 dB. The formula we need is this:

$$BW = \frac{f_O}{Q} \tag{10.5}$$

For circuits of high selectivity (low BW), the requirement is that Q be high, and hence R low. This will influence the selection of components used in the design, particularly that an inductor with low internal resistance be found relative to its impedance at f_O. Often we design the circuit with Q higher than needed, and then add a suitable series resistor, to reduce the bandwidth to the desired value. This is called a 'damping' resistor.

Table 10.1 lists the resonant frequencies from 500 Hz to 50 MHz obtainable with E6 inductors and E12 capacitors. A column is also provided, X, showing the value of X_C or X_L at resonance. This can be used to find the highest possible Q, provided that the resistance of the inductor is known.

Example 10.1: (a) What values of C and L can we use to give a series resonant circuit with $f_O = 8\,\text{kHz}$?

From Table 10.1:

f_O	C	L	X
8k	12n	33m	1.66k
8k	18n	22m	1.11k
8k	120n	3m3	166
8k	180n	2m2	111

This gives us a choice of four pairs.

(b) By consulting suppliers catalogues, we source a 3.3 mH inductor with $R = 6.6\,\Omega$, and a 2.2 mH inductor with resistance, R_L, of 4.1 Ω, What is the best possible Q with each of these? What will the bandwidth be?

With 3.3 mH, X is given as 166 Ω, so Q will be $\frac{166}{6.6} = 25.2$ at best. The bandwidth will be $\frac{8000}{25.2} = 318$ Hz

With 2.2 mH, X is given as 111 Ω, so Q will be $\frac{111}{4.1} = 27.1$ at best. The bandwidth will be $\frac{8000}{27.1} = 295$ Hz

(c) Now calculate values for a series resistor to reduce Q to around 3 for each.

For 3.3 mH: If $X = 166\,\Omega$, and $Q = 3$, then $R = \frac{X}{Q} = \frac{166}{3} = 55.3\,\Omega$. As the inductor already has 6.6 Ω of resistance, we should add a resistor of $55.3 - 6.6 = 48.7\,\Omega$. (with typical inductor tolerances of 10%, 47 Ω would be fine).

For 2.2 mH: $R = \frac{X}{Q} = \frac{111}{3} = 37\,\Omega$. Resistor should be $37 - 4.1 = 32.9\,\Omega$ – say a standard value of 33 Ω.

Parallel tuned circuits

A parallel tuned circuit consists of a capacitor and an inductor in parallel. The circuit can be understood by using a similar approach to the one which we took for the series tuned circuit. However, the practical circuit is slightly complicated by the fact that the inductive branch possesses more resistance than the capacitive branch, which makes it somewhat asymmetrical. In many cases this effect can be ignored, but not all, so we need to be able to predict whether it will be significant or not. But to make the analysis one step at a time, it makes sense to start by considering an ideal version of the circuit, where the effect does not come into play at all.

Example 10.2: If the 2.2 mH inductor of Example 10.1 has a 10% tolerance, how much might f_O and Q vary in practice, assuming that both inductance and resistance increase together?

If the inductance and resistance increase by 10%, we would have $L(\text{max}) = 2.2\,\text{mH} \times 1.1 = 2.42\,\text{mH}$, and $R_{L(\text{max})} = 4.1 \times 1.1 = 4.51\,\Omega$. This would give, using Eqn (6.1):

$$f_O = \frac{1}{2\pi\sqrt{2.42\text{m} \times 180\text{n}}} = 7.63\,\text{kHz}$$

and a corresponding $Q = \frac{2\pi \times 7.63\text{k} \times 2.42\text{m}}{33 + 4.51} = 3.1$
Decreasing by 10% gives $L_{(\text{min})} = 2.2\text{m} \times 0.9 = 1.98\,\text{mH}$ and $\text{R}_\text{L}(\text{min}) = 4.1 \times 0.9 = 3.69\,\Omega$

$$f_O = \frac{1}{2\pi\sqrt{1.98\text{m} \times 180\text{n}}} = 8.43\,\text{kHz}$$

with

$$Q = \frac{2\pi \times 8.43\text{k} \times 1.98\text{m}}{33 + 3.69} = 2.9$$

The ideal parallel tuned circuit

This consists of an ideal inductor and capacitor in parallel – each has no resistance. In this circuit we take the voltage across the circuit, V_{IN}, as our phase reference, as it is common to both components, and look at the current that each branch draws from the source at different frequencies. These currents will always be in antiphase – the capacitor's current, I_C, leading V_{IN} by 90°, the inductor's, I_L, lagging. The input current, I_{IN}, will be their phasor sum. As they are in antiphase, their phasor sum is simply the difference between them.

We can once again say that X_C will fall with increasing frequency, and X_L will rise. And once again, X_C will equal X_L at some frequency, that given by Eqn (10.1). At this frequency, again f_O, the current in each branch will be equal, and hence the input current will be zero. The impedance is $\frac{V_{IN}}{I_{IN}}$, and is infinite. The circuit looks like an open circuit to the source at resonance, and its resonant frequency is calculated identically to that of the series tuned circuit.

Below resonance, the capacitor will have the greater impedance of the two devices, and the impedance will appear inductive with current lagging voltage by 90°. Above, the inductor possesses more impedance and the impedance is capacitive, with V_{IN} in lagging I_{IN}.

At resonance, the input current is zero, while I_C and I_L are $\frac{V_{IN}}{X_C}$ and $\frac{V_{IN}}{X_L}$ respectively. The Q factor, $\frac{I_C}{I_{IN}}$ or $\frac{I_L}{I_{IN}}$ in would therefore be infinite.

The practical parallel tuned circuit

The practical circuit differs from the ideal mostly in that the inductor possesses resistance. Such a circuit is shown in Figure 10.2, with R_L representing the internal resistance of the inductor.

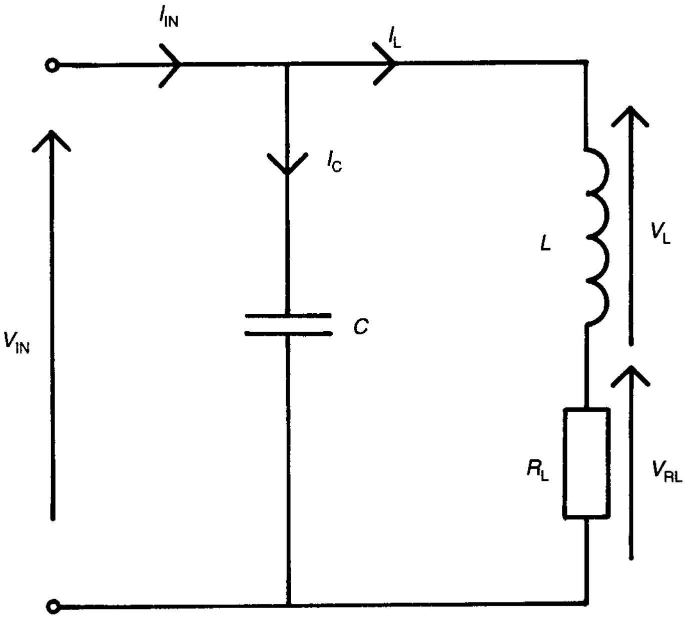

Figure 10.2 Practical parallel tuned circuit

The addition of the resistive element in series with the inductor causes I_L to be less than 90° lagging V_{IN}. In turn, this causes I_{IN}, at the frequency at which $X_C = X_L$, to be leading V_{IN} by some angle, rather than in phase, as it was with the ideal circuit—the circuit is overall capacitive, rather than resistive. To compensate, and bring I_{IN} back in phase with V_{IN}, making the circuit truly resonant, the frequency of V_{IN} must be lowered. This has the effect of making I_C smaller and I_L larger. It can be seen that the effect becomes more pronounced the greater the inductor's resistance is, compared to its reactance.

To what degree all this affects Eqn (10.1) depends on the Q of the inductor at the resonant frequency, as defined in Eqn (10.4).

Resonant frequency error

The resonant frequency as found from Eqn (10.1) must be multiplied by the term:

$$k = \sqrt{1 - \frac{CR^2}{L}} = \sqrt{1 - \frac{1}{Q^2}} \qquad (10.6)$$

to find the true resonant frequency. In practice, the Q has to be pretty low to cause much deviation; a Q of 4 gives $k = 0.97$, a Q of 2 gives $k = 0.87$. Eqn (10.1) is often used as it stands.

Dynamic impedance

The practical circuit no longer looks like an open circuit to the source at resonance. Its impedance is purely resistive however, and is often called the dynamic impedance. It is given by:

$$Z_O = \frac{L}{CR_L} \qquad (10.7)$$

From this we can see that dynamic impedance is inversely proportional to the resistance of the inductor.

Q factor

With the practical parallel tuned circuit the Q is, as with the series resonant circuit, the same as the Q of the inductor at the resonant frequency. We might ask ourselves what happens when we wish to deliberately reduce the circuit's Q; for if we placed our damping resistor in series with the inductor, it would also affect our resonant frequency. The solution is that the damping resistor is placed across the whole tuned circuit, as shown in Figure 10.3, where it has the desired effect without affecting f_O. The value of the damping resistor can be calculated as follows:

$$R_D = Z_O \frac{Q_D}{Q_I - Q_D} \qquad (10.8)$$

where R_D is the damping resistor value, Q_D is the desired Q(*Q damped*), $Z_O = \frac{L}{CR}$ as previously, and Q_I is the Q of the inductor at resonance, or the Q of the circuit with no damping resistor. Q_D must obviously be less than Q_I.

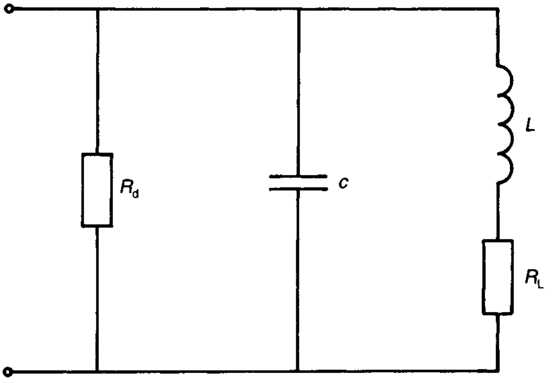

Figure 10.3 Practical parallel tuned circuit with damping resistor

Maximum capacitor value

We can see that as R becomes larger, k Eqn (10.6) goes from 1 towards zero, and the resonant frequency falls, also towards zero. When:

$$\frac{CR_L^2}{L} = 1 \text{ or } C = \frac{L}{R_L^2} \tag{10.9}$$

then $k = 0$, and the circuit ceases to resonate. This tells us that for any practical inductor, there is a maximum value of capacitor which can be placed across it for it to remain resonant. This value is easily calculated from Eqn (10.9).

Example 10.3: design a parallel tuned circuit with $f_O = 120\,\text{kHz} +/-2\,\text{kHz}$, $Q = 20$ and the highest feasible dynamic impedance, with the restriction that C is greater than 100 pF. What is the dynamic impedance, with and without the damping resistor? (Assume that the resistance of the inductor is 5 Ω.)

From Table 10.1, we choose $C = 120\,\text{pF}$, $L = 15\,\text{mH}$, giving a nominal frequency of 119 kHz. We can calculate the Q of the inductor by taking X from the table – 11.2k – and dividing by 5 Ω. This gives $Q = 2240$. Now we calculate Z_O, using Eqn (10.7):

$$Z_O = \frac{15\text{m}}{120\text{p} \times 5} = 25\,\text{M}\Omega$$

and so

$$R_D = 25\text{M} \times \frac{20}{2240 - 20} = 225\,\text{k}\Omega$$

As Z_O is so much larger than R_D, we can take the dynamic impedance with R_D fitted to be equal to R_D, 225 k Ω.

Table 10.1 *Frequency constants for LC circuits*

f_O	C	L	X
503	1u	100m	316
556	820n	100m	349
610	680n	100m	383
610	1u	68m	261
673	560n	100m	423
674	820n	68m	288
734	470n	100m	461
734	1u	47m	217
740	680n	68m	316
806	390n	100m	506
811	820n	47m	239
816	560n	68m	348
876	330n	100m	550
876	1u	33m	182
890	680n	47m	263
890	470n	68m	380
968	820n	33m	201
969	270n	100m	609
977	390n	68m	418
981	560n	47m	290
1.06k	330n	68m	454
1.06k	680n	33m	220
1.07k	470n	47m	316
1.07k	220n	100m	674
1.07k	1u	22m	148
1.17k	560n	33m	243
1.17k	270n	68m	502
1.18k	390n	47m	347
1.18k	820n	22m	164
1.19k	180n	100m	745
1.28k	470n	33m	265
1.28k	330n	47m	377
1.30k	150n	100m	816
1.30k	1u	15m	122
1.30k	220n	68m	556
1.30k	680n	22m	180
1.40k	390n	33m	291
1.41k	270n	47m	417
1.43k	560n	22m	198
1.44k	820n	15m	135
1.44k	180n	68m	615
1.45k	120n	100m	913
1.53k	330n	33m	316
1.57k	220n	47m	462
1.57k	470n	22m	216
1.58k	680n	15m	149
1.58k	150n	68m	673
1.59k	100n	100m	1.00k
1.59k	1u	10m	100
1.69k	270n	33m	350
1.72k	390n	22m	238
1.73k	180n	47m	511
1.74k	560n	15m	164
1.76k	820n	10m	110
1.76k	82n	100m	1.10k
1.76k	120n	68m	753
1.87k	220n	33m	387
1.87k	330n	22m	258
1.90k	150n	47m	560
1.90k	470n	15m	179
1.93k	68n	100m	1.21k
1.93k	100n	68m	825
1.93k	680n	10m	121
1.93k	1u	6m8	82.5
2.07k	180n	33m	428
2.07k	270n	22m	285
2.08k	390n	15m	196
2.12k	120n	47m	626
2.13k	56n	100m	1.34k
2.13k	560n	10m	134
2.13k	82n	68m	911
2.13k	820n	6m8	91.1

f_O	C	L	X
2.26k	150n	33m	469
2.26k	330n	15m	213
2.29k	220n	22m	316
2.32k	47n	100m	1.46k
2.32k	100n	47m	686
2.32k	470n	10m	146
2.32k	1u	4m7	68.6
2.34k	68n	68m	1.00k
2.34k	680n	6m8	100
2.50k	270n	15m	236
2.53k	120n	33m	524
2.53k	180n	22m	350
2.55k	39n	100m	1.60k
2.55k	390n	10m	160
2.56k	82n	47m	757
2.56k	820n	4m7	75.7
2.58k	56n	68m	1.10k
2.58k	560n	6m8	110
2.77k	33n	100m	1.74k
2.77k	100n	33m	574
2.77k	150n	22m	383
2.77k	220n	15m	261
2.77k	330n	10m	174
2.77k	1u	3m3	57.4
2.82k	680n	4m7	83.1
2.82k	47n	68m	1.20k
2.82k	68n	47m	831
2.82k	470n	6m8	120
3.06k	82n	33m	634
3.06k	820n	3m3	63.4
3.06k	27n	100m	1.92k
3.06k	180n	15m	289
3.06k	270n	10m	192
3.09k	390n	6m8	132
3.09k	39n	68m	1.32k
3.10k	120n	22m	428
3.10k	56n	47m	916
3.10k	560n	4m7	91.6
3.36k	150n	15m	316
3.36k	33n	68m	1.44k
3.36k	68n	33m	697
3.36k	330n	6m8	144
3.36k	680n	3m3	69.7
3.39k	47n	47m	1.00k
3.39k	470n	4m7	100
3.39k	22n	100m	2.13k
3.39k	100n	22m	469
3.39k	220n	10m	213
3.39k	1u	2m2	46.9
3.70k	56n	33m	768
3.70k	560n	3m3	76.8
3.71k	27n	68m	1.59k
3.71k	270n	6m8	159
3.72k	39n	47m	1.10k
3.72k	390n	4m7	110
3.75k	82n	22m	518
3.75k	820n	2m2	51.8
3.75k	18n	100m	2.36k
3.75k	120n	15m	354
3.75k	180n	10m	236
4.04k	33n	47m	1.19k
4.04k	47n	33m	838
4.04k	330n	4m7	119
4.04k	470n	3m3	83.8
4.11k	15n	100m	2.58k
4.11k	100n	15m	387
4.11k	150n	10m	258
4.11k	1u	1m5	38.7
4.11k	22n	68m	1.76k
4.11k	68n	22m	569
4.11k	220n	6m8	176
4.11k	680n	2m2	56.9
4.44k	390n	3m3	92.0
4.44k	39n	33m	920

f_O	C	L	X
4.47k	27n	47m	1.32k
4.47k	270n	4m7	132
4.53k	56n	22m	627
4.53k	560n	2m2	62.7
4.54k	820n	1m5	42.8
4.54k	82n	15m	428
4.55k	18n	68m	1.94k
4.55k	180n	6m8	194
4.59k	12n	100m	2.89k
4.59k	120n	10m	289
4.82k	33n	33m	1.00k
4.82k	330n	3m3	100
4.95k	22n	47m	1.46k
4.95k	47n	22m	684
4.95k	220n	4m7	146
4.95k	470n	2m2	68.4
4.98k	680n	1m5	47.0
4.98k	15n	68m	2.13k
4.98k	68n	15m	470
4.98k	150n	6m8	213
5.03k	10n	100m	3.16k
5.03k	100n	10m	316
5.03k	1u	1m	31.6
5.33k	27n	33m	1.11k
5.33k	270n	3m3	111
5.43k	39n	22m	751
5.43k	390n	2m2	75.1
5.47k	180n	4m7	162
5.47k	18n	47m	1.62k
5.49k	56n	15m	518
5.49k	560n	1m5	51.8
5.56k	8n2	100m	3.49k
5.56k	82n	10m	349
5.56k	820n	1m	34.9
5.57k	12n	68m	2.38k
5.57k	120n	6m8	238
5.91k	22n	33m	1.22k

f_O	C	L	X
5.91k	33n	22m	816
5.91k	220n	3m3	122
5.91k	330n	2m2	81.6
5.99k	15n	47m	1.77k
5.99k	47n	15m	565
5.99k	150n	4m7	177
5.99k	470n	1m5	56.5
6.10k	6n8	100m	3.83k
6.10k	100n	6m8	261
6.10k	680n	1m	38.3
6.10k	10n	68m	2.61k
6.10k	68n	10m	383
6.10k	1u	680u	26.1
6.53k	18n	33m	1.35k
6.53k	180n	3m3	135
6.53k	27n	22m	903
6.53k	270n	2m2	90.3
6.58k	390n	1m5	62.0
6.58k	39n	15m	620
6.70k	12n	47m	1.98k
6.70k	120n	4m7	198
6.73k	5n6	100m	4.23k
6.73k	56n	10m	423
6.73k	560n	1m	42.3
6.74k	8n2	68m	2.88k
6.74k	82n	6m8	288
6.74k	820n	680u	28.8
7.15k	15n	33m	1.48k
7.15k	33n	15m	674
7.15k	150n	3m3	148
7.15k	330n	1m5	67.4
7.23k	22n	22m	1.00k
7.23k	220n	2m2	100
7.34k	4n7	100m	4.61k
7.34k	10n	47m	2.17k
7.34k	47n	10m	461
7.34k	100n	4m7	217

f_O	C	L	X
7.34k	470n	1m	46.1
7.34k	1u	470u	21.7
7.40k	6n8	68m	3.16k
7.40k	68n	6m8	316
7.40k	680n	680u	31.6
7.91k	27n	15m	745
7.91k	270n	1m5	74.5
8.00k	12n	33m	1.66k
8.00k	18n	22m	1.11k
8.00k	120n	3m3	166
8.00k	180n	2m2	111
8.06k	3n9	100m	5.06k
8.06k	39n	10m	506
8.06k	390n	1m	50.6
8.11k	8n2	47m	2.39k
8.11k	82n	4m7	239
8.11k	820n	470u	23.9
8.16k	5n6	68m	3.48k
8.16k	56n	6m8	348
8.16k	560n	680u	34.8
8.76k	3n3	100m	5.50k
8.76k	10n	33m	1.82k
8.76k	15n	22m	1.21k
8.76k	22n	15m	826
8.76k	33n	10m	550
8.76k	100n	3m3	182
8.76k	150n	2m2	121
8.76k	220n	1m5	82.6
8.76k	330n	1m	55.0
8.76k	1u	330u	18.2
8.90k	6n8	47m	2.63k
8.90k	68n	4m7	263
8.90k	680n	470u	26.3
8.90k	4n7	68m	3.80k
8.90k	47n	6m8	380
8.90k	470n	680u	38.0
9.68k	8n2	33m	2.01k
9.68k	82n	3m3	201
9.68k	820n	330u	20.1
9.69k	2n7	100m	6.09k
9.69k	18n	15m	913
9.69k	180n	1m5	91.3
9.69k	27n	10m	609
9.69k	270n	1m	60.9
9.77k	3n9	68m	4.18k
9.77k	39n	6m8	418
9.77k	390n	680u	41.8
9.80k	12n	22m	1.35k
9.80k	120n	2m2	135
9.81k	5n6	47m	2.90k
9.81k	56n	4m7	290
9.81k	560n	470u	29.0
10.6k	15n	15m	1.00k
10.6k	150n	1m5	100
10.6k	3n3	68m	4.54k
10.6k	6n8	33m	2.20k
10.6k	33n	6m8	454
10.6k	68n	3m3	220
10.6k	330n	680u	45.4
10.6k	680n	330u	22.0
10.7k	4n7	47m	3.16k
10.7k	47n	4m7	316
10.7k	470n	470u	31.6
10.7k	2n2	100m	6.74k
10.7k	10n	22m	1.48k
10.7k	22n	10m	674
10.7k	100n	2m2	148
10.7k	220n	1m	67.4
10.7k	1u	220u	14.8
11.7k	5n6	33m	2.43k
11.7k	56n	3m3	243
11.7k	560n	330u	24.3
11.7k	2n7	68m	5.02k
11.7k	27n	6m8	502

f_O	C	L	X
11.7k	270n	680u	50.2
11.8k	3n9	47m	3.47k
11.8k	39n	4m7	347
11.8k	390n	470u	34.7
11.8k	8n2	22m	1.64k
11.8k	82n	2m2	164
11.8k	820n	220u	16.4
11.9k	1n8	100m	7.45k
11.9k	12n	15m	1.12k
11.9k	18n	10m	745
11.9k	120n	1m5	112
11.9k	180n	1m	74.5
12.8k	3n3	47m	3.77k
12.8k	4n7	33m	2.65k
12.8k	33n	4m7	377
12.8k	47n	3m3	265
12.8k	330n	470u	37.7
12.8k	470n	330u	26.5
13.0k	1n5	100m	8.16k
13.0k	10n	15m	1.22k
13.0k	15n	10m	816
13.0k	100n	1m5	122
13.0k	150n	1m	81.6
13.0k	1u	150u	12.2
13.0k	2n2	68m	5.56k
13.0k	6n8	22m	1.80k
13.0k	22n	6m8	556
13.0k	220n	680u	55.6
13.0k	680n	220u	18.0
13.0k	68n	2m2	180
14.0k	3n9	33m	2.91k
14.0k	390n	330u	29.1
14.0k	39n	3m3	291
14.1k	2n7	47m	4.17k
14.1k	27n	4m7	417
14.1k	270n	470u	41.7
14.3k	5n6	22m	1.98k
14.3k	56n	2m2	198
14.3k	560n	220u	19.8
14.4k	8n2	15m	1.35k
14.4k	82n	1m5	135
14.4k	820n	150u	13.5
14.4k	1n8	68m	6.15k
14.4k	18n	6m8	615
14.4k	180n	680u	61.5
14.5k	12n	10m	913
14.5k	1n2	100m	9.13k
14.5k	120n	1m	91.3
15.3k	3n3	33m	3.16k
15.3k	33n	3m3	316
15.3k	330n	330u	31.6
15.7k	2n2	47m	4.62k
15.7k	4n7	22m	2.16k
15.7k	22n	4m7	462
15.7k	47n	2m2	216
15.7k	220n	470u	46.2
15.7k	470n	220u	21.6
15.8k	1n5	68m	6.73k
15.8k	6n8	15m	1.49k
15.8k	15n	6m8	673
15.8k	68n	1m5	149
15.8k	150n	680u	67.3
15.8k	680n	150u	14.9
15.9k	1n	100m	10.0k
15.9k	10n	10m	1.00k
15.9k	100n	1m	100
15.9k	1u	100u	10.0
16.9k	2n7	33m	3.50k
16.9k	27n	3m3	350
16.9k	270n	330u	35.0
17.2k	3n9	22m	2.38k
17.2k	39n	2m2	238
17.2k	390n	220u	23.8
17.3k	1n8	47m	5.11k

f_O	C	L	X
17.3k	18n	4m7	511
17.3k	180n	470u	51.1
17.4k	5n6	15m	1.64k
17.4k	56n	1m5	164
17.4k	560n	150u	16.4
17.6k	820p	100m	11.0k
17.6k	8n2	10m	1.10k
17.6k	82n	1m	110
17.6k	820n	100u	11.0
17.6k	1n2	68m	7.53k
17.6k	12n	6m8	753
17.6k	120n	680u	75.3
18.7k	2n2	33m	3.87k
18.7k	3n3	22m	2.58k
18.7k	22n	3m3	387
18.7k	33n	2m2	258
18.7k	220n	330u	38.7
18.7k	330n	220u	25.8
19.0k	1n5	47m	5.60k
19.0k	4n7	15m	1.79k
19.0k	15n	4m7	560
19.0k	47n	1m5	179
19.0k	150n	470u	56.0
19.0k	470n	150u	17.9
19.3k	680p	100m	12.1k
19.3k	1n	68m	8.25k
19.3k	6n8	10m	1.21k
19.3k	10n	6m8	825
19.3k	100n	680u	82.5
19.3k	680n	100u	12.1
19.3k	1u	68u	8.25
19.3k	68n	1m	121
20.7k	1n8	33m	4.28k
20.7k	2n7	22m	2.85k
20.7k	18n	3m3	428
20.7k	27n	2m2	285
20.7k	180n	330u	42.8

f_O	C	L	X
20.7k	270n	220u	28.5
20.8k	3n9	15m	1.96k
20.8k	39n	1m5	196
20.8k	390n	150u	19.6
21.2k	1n2	47m	6.26k
21.2k	12n	4m7	626
21.2k	120n	470u	62.6
21.3k	560p	100m	13.4k
21.3k	5n6	10m	1.34k
21.3k	56n	1m	134
21.3k	560n	100u	13.4
21.3k	8n2	6m8	911
21.3k	820n	68u	9.11
21.3k	820p	68m	9.11k
21.3k	82n	680u	91.1
22.6k	1n5	33m	4.69k
22.6k	3n3	15m	2.13k
22.6k	15n	3m3	469
22.6k	33n	1m5	213
22.6k	150n	330u	46.9
22.6k	330n	150u	21.3
22.9k	2n2	22m	3.16k
22.9k	22n	2m2	316
22.9k	220n	220u	31.6
23.2k	470p	100m	14.6k
23.2k	1n	47m	6.86k
23.2k	4n7	10m	1.46k
23.2k	10n	4m7	686
23.2k	47n	1m	146
23.2k	100n	470u	68.6
23.2k	470n	100u	14.6
23.2k	1u	47u	6.86
23.4k	6n8	6m8	1.00k
23.4k	680n	68u	10.0
23.4k	680p	68m	10.0k
23.4k	68n	680u	100
25.0k	2n7	15m	2.36k

f_O	C	L	X
25.0k	27n	1m5	236
25.0k	270n	150u	23.6
25.3k	1n2	33m	5.24k
25.3k	1n8	22m	3.50k
25.3k	12n	3m3	524
25.3k	18n	2m2	350
25.3k	120n	330u	52.4
25.3k	180n	220u	35.0
25.5k	3n9	10m	1.60k
25.5k	390n	100u	16.0
25.5k	390p	100m	16.0k
25.5k	39n	1m	160
25.6k	820p	47m	7.57k
25.6k	8n2	4m7	757
25.6k	82n	470u	75.7
25.6k	820n	47u	7.57
25.8k	560p	68m	11.0k
25.8k	5n6	6m8	1.10k
25.8k	56n	680u	110
25.8k	560n	68u	11.0
27.7k	330p	100m	17.4k
27.7k	1n	33m	5.74k
27.7k	1n5	22m	3.83k
27.7k	2n2	15m	2.61k
27.7k	3n3	10m	1.74k
27.7k	10n	3m3	574
27.7k	15n	2m2	383
27.7k	22n	1m5	261
27.7k	33n	1m	174
27.7k	100n	330u	57.4
27.7k	150n	220u	38.3
27.7k	220n	150u	26.1
27.7k	330n	100u	17.4
27.7k	1u	33u	5.74
28.2k	470p	68m	12.0k
28.2k	680p	47m	8.31k
28.2k	4n7	6m8	1.20k
28.2k	6n8	4m7	831
28.2k	47n	680u	120
28.2k	68n	470u	83.1
28.2k	470n	68u	12.0
28.2k	680n	47u	8.31
30.6k	820p	33m	6.34k
30.6k	8n2	3m3	634
30.6k	82n	330u	63.4
30.6k	820n	33u	6.34
30.6k	270p	100m	19.2k
30.6k	1n8	15m	2.89k
30.6k	2n7	10m	1.92k
30.6k	18n	1m5	289
30.6k	27n	1m	192
30.6k	180n	150u	28.9
30.6k	270n	100u	19.2
30.9k	390p	68m	13.2k
30.9k	3n9	6m8	1.32k
30.9k	39n	680u	132
30.9k	390n	68u	13.2
31.0k	1n2	22m	4.28k
31.0k	12n	2m2	428
31.0k	120n	220u	42.8
31.0k	560p	47m	9.16k
31.0k	5n6	4m7	916
31.0k	56n	470u	91.6
31.0k	560n	47u	9.16
33.6k	1n5	15m	3.16k
33.6k	15n	1m5	316
33.6k	150n	150u	31.6
33.6k	330p	68m	14.4k
33.6k	680p	33m	6.97k
33.6k	3n3	6m8	1.44k
33.6k	6n8	3m3	697
33.6k	33n	680u	144
33.6k	330n	68u	14.4
33.6k	680n	33u	6.97

f_O	C	L	X
33.6k	68n	330u	69.7
33.9k	470p	47m	10.0k
33.9k	4n7	4m7	1.00k
33.9k	47n	470u	100
33.9k	470n	47u	10.0
33.9k	220p	100m	21.3k
33.9k	1n	22m	4.69k
33.9k	2n2	10m	2.13k
33.9k	10n	2m2	469
33.9k	22n	1m	213
33.9k	100n	220u	46.9
33.9k	220n	100u	21.3
33.9k	1u	22u	4.69
37.0k	5n6	3m3	768
37.0k	560p	33m	7.68k
37.0k	56n	330u	76.8
37.0k	560n	33u	7.68
37.1k	2n7	6m8	1.59k
37.1k	270n	68u	15.9
37.1k	270p	68m	15.9k
37.1k	27n	680u	159
37.2k	3n9	4m7	1.10k
37.2k	390n	47u	11.0
37.2k	390p	47m	11.0k
37.2k	39n	470u	110
37.5k	820p	22m	5.18k
37.5k	8n2	2m2	518
37.5k	82n	220u	51.8
37.5k	820n	22u	5.18
37.5k	180p	100m	23.6k
37.5k	1n2	15m	3.54k
37.5k	1n8	10m	2.36k
37.5k	12n	1m5	354
37.5k	18n	1m	236
37.5k	120n	150u	35.4
37.5k	180n	100u	23.6
40.4k	470p	33m	8.38k

f_O	C	L	X
40.4k	3n3	4m7	1.19k
40.4k	330p	47m	11.9k
40.4k	4n7	3m3	838
40.4k	33n	470u	119
40.4k	47n	330u	83.8
40.4k	330n	47u	11.9
40.4k	470n	33u	8.38
41.1k	150p	100m	25.8k
41.1k	1n	15m	3.87k
41.1k	1n5	10m	2.58k
41.1k	10n	1m5	387
41.1k	15n	1m	258
41.1k	100n	150u	38.7
41.1k	150n	100u	25.8
41.1k	1u	15u	3.87
41.1k	220p	68m	17.6k
41.1k	680p	22m	5.69k
41.1k	2n2	6m8	1.76k
41.1k	6n8	2m2	569
41.1k	22n	680u	176
41.1k	68n	220u	56.9
41.1k	220n	68u	17.6
41.1k	680n	22u	5.69
44.4k	390p	33m	9.20k
44.4k	3n9	3m3	920
44.4k	39n	330u	92.0
44.4k	390n	33u	9.20
44.7k	270p	47m	13.2k
44.7k	2n7	4m7	1.32k
44.7k	27n	470u	132
44.7k	270n	47u	13.2
45.3k	560p	22m	6.27k
45.3k	5n6	2m2	627
45.3k	56n	220u	62.7
45.3k	560n	22u	6.27
45.4k	820p	15m	4.28k
45.4k	8n2	1m5	428

f_O	C	L	X
45.4k	82n	150u	42.8
45.4k	820n	15u	4.28
45.5k	1n8	6m8	1.94k
45.5k	180n	68u	19.4
45.5k	180p	68m	19.4k
45.5k	18n	680u	194
45.9k	120p	100m	28.9k
45.9k	1n2	10m	2.89k
45.9k	12n	1m	289
45.9k	120n	100u	28.9
48.2k	330p	33m	10.0k
48.2k	3n3	3m3	1.00k
48.2k	33n	330u	100
48.2k	330n	33u	10.0
49.5k	2n2	4m7	1.46k
49.5k	220n	47u	14.6
49.5k	220p	47m	14.6k
49.5k	470p	22m	6.84k
49.5k	22n	470u	146
49.5k	47n	220u	68.4
49.5k	470n	22u	6.84
49.5k	4n7	2m2	684
49.8k	680p	15m	4.70k
49.8k	1n5	6m8	2.13k
49.8k	6n8	1m5	470
49.8k	680n	15u	4.70
49.8k	150p	68m	21.3k
49.8k	15n	680u	213
49.8k	68n	150u	47.0
49.8k	150n	68u	21.3
50.3k	100n	100u	31.6
50.3k	100p	100m	31.6k
50.3k	1n	10m	3.16k
50.3k	10n	1m	316
50.3k	1u	10u	3.16
53.3k	270p	33m	11.1k
53.3k	2n7	3m3	1.11k
53.3k	27n	330u	111
53.3k	270n	33u	11.1
54.3k	390n	22u	7.51
54.3k	390p	22m	7.51k
54.3k	3n9	2m2	751
54.3k	39n	220u	75.1
54.7k	180p	47m	16.2k
54.7k	1n8	4m7	1.62k
54.7k	18n	470u	162
54.7k	180n	47u	16.2
54.9k	560p	15m	5.18k
54.9k	5n6	1m5	518
54.9k	56n	150u	51.8
54.9k	560n	15u	5.18
55.6k	82p	100m	34.9k
55.6k	820p	10m	3.49k
55.6k	8n2	1m	349
55.6k	82n	100u	34.9
55.6k	820n	10u	3.49
55.7k	120p	68m	23.8k
55.7k	1n2	6m8	2.38k
55.7k	12n	680u	238
55.7k	120n	68u	23.8
59.1k	220p	33m	12.2k
59.1k	2n2	3m3	1.22k
59.1k	3n3	2m2	816
59.1k	330p	22m	8.16k
59.1k	22n	330u	122
59.1k	33n	220u	81.6
59.1k	220n	33u	12.2
59.1k	330n	22u	8.16
59.9k	150p	47m	17.7k
59.9k	470p	15m	5.65k
59.9k	1n5	4m7	1.77k
59.9k	4n7	1m5	565
59.9k	15n	470u	177
59.9k	47n	150u	56.5

f_O	C	L	X
59.9k	470n	15u	5.65
59.9k	150n	47u	17.7
61.0k	100p	68m	26.1k
61.0k	680p	10m	3.83k
61.0k	1n	6m8	2.61k
61.0k	6n8	1m	383
61.0k	68n	100u	38.3
61.0k	100n	68u	26.1
61.0k	680n	10u	3.83
61.0k	1u	6u8	2.61
61.0k	68p	100m	38.3k
61.0k	10n	680u	261
65.3k	180p	33m	13.5k
65.3k	270p	22m	9.03k
65.3k	1n8	3m3	1.35k
65.3k	2n7	2m2	903
65.3k	18n	330u	135
65.3k	27n	220u	90.3
65.3k	180n	33u	13.5
65.3k	270n	22u	9.03
65.8k	3n9	1m5	620
65.8k	390p	15m	6.20k
65.8k	39n	150u	62.0
65.8k	390n	15u	6.20
67.0k	1n2	4m7	1.98k
67.0k	12n	470u	198
67.0k	120p	47m	19.8k
67.0k	120n	47u	19.8
67.3k	56p	100m	42.3k
67.3k	560p	10m	4.23k
67.3k	5n6	1m	423
67.3k	56n	100u	42.3
67.3k	560n	10u	4.23
67.4k	820n	6u8	2.88
67.4k	82p	68m	28.8k
67.4k	820p	6m8	2.88k
67.4k	8n2	680u	288

f_O	C	L	X
67.4k	82n	68u	28.8
71.5k	150p	33m	14.8k
71.5k	330p	15m	6.74k
71.5k	1n5	3m3	1.48k
71.5k	3n3	1m5	674
71.5k	15n	330u	148
71.5k	33n	150u	67.4
71.5k	330n	15u	6.74
71.5k	150n	33u	14.8
72.3k	220p	22m	10.0k
72.3k	2n2	2m2	1.00k
72.3k	22n	220u	100
72.3k	220n	22u	10.0
73.4k	100p	47m	21.7k
73.4k	470p	10m	4.61k
73.4k	1n	4m7	2.17k
73.4k	10n	470u	217
73.4k	47n	100u	46.1
73.4k	100n	47u	21.7
73.4k	470n	10u	4.61
73.4k	1u	4u7	2.17
73.4k	47p	100m	46.1k
73.4k	4n7	1m	461
74.0k	6n8	680u	316
74.0k	680n	6u8	3.16
74.0k	68p	68m	31.6k
74.0k	680p	6m8	3.16k
74.0k	68n	68u	31.6
79.1k	2n7	1m5	745
79.1k	270p	15m	7.45k
79.1k	27n	150u	74.5
79.1k	270n	15u	7.45
80.0k	120p	33m	16.6k
80.0k	180p	22m	11.1k
80.0k	1n2	3m3	1.66k
80.0k	1n8	2m2	1.11k
80.0k	12n	330u	166

f_O	C	L	X
80.0k	18n	220u	111
80.0k	120n	33u	16.6
80.0k	180n	22u	11.1
80.6k	39p	100m	50.6k
80.6k	390p	10m	5.06k
80.6k	3n9	1m	506
80.6k	39n	100u	50.6
80.6k	390n	10u	5.06
81.1k	820p	4m7	2.39k
81.1k	8n2	470u	239
81.1k	820n	4u7	2.39
81.1k	82p	47m	23.9k
81.1k	82n	47u	23.9
81.6k	5n6	680u	348
81.6k	56n	68u	34.8
81.6k	56p	68m	34.8k
81.6k	560p	6m8	3.48k
81.6k	560n	6u8	3.48
87.6k	2n2	1m5	826
87.6k	100p	33m	18.2k
87.6k	150p	22m	12.1k
87.6k	220p	15m	8.26k
87.6k	1n	3m3	1.82k
87.6k	1n5	2m2	1.21k
87.6k	15n	220u	121
87.6k	22n	150u	82.6
87.6k	220n	15u	8.26
87.6k	330n	10u	5.50
87.6k	330p	10m	5.50k
87.6k	3n3	1m	550
87.6k	33n	100u	55.0
87.6k	150n	22u	12.1
87.6k	1u	3u3	1.82
87.6k	33p	100m	55.0k
87.6k	10n	330u	182
87.6k	100n	33u	18.2
89.0k	680p	4m7	2.63k
89.0k	6n8	470u	263
89.0k	47n	68u	38.0
89.0k	470n	6u8	3.80
89.0k	680n	4u7	2.63
89.0k	47p	68m	38.0k
89.0k	68p	47m	26.3k
89.0k	470p	6m8	3.80k
89.0k	4n7	680u	380
89.0k	68n	47u	26.3
96.8k	82p	33m	20.1k
96.8k	820p	3m3	2.01k
96.8k	8n2	330u	201
96.8k	820n	3u3	2.01
96.8k	82n	33u	20.1
96.9k	1n8	1m5	913
96.9k	180p	15m	9.13k
96.9k	270p	10m	6.09k
96.9k	18n	150u	91.3
96.9k	180n	15u	9.13
96.9k	27p	100m	60.9k
96.9k	2n7	1m	609
96.9k	27n	100u	60.9
96.9k	270n	10u	6.09
97.7k	3n9	680u	418
97.7k	39n	68u	41.8
97.7k	390n	6u8	4.18
97.7k	39p	68m	41.8k
97.7k	390p	6m8	4.18k
98.0k	120p	22m	13.5k
98.0k	1n2	2m2	1.35k
98.0k	12n	220u	135
98.0k	120n	22u	13.5
98.1k	56p	47m	29.0k
98.1k	560p	4m7	2.90k
98.1k	5n6	470u	290
98.1k	56n	47u	29.0
98.1k	560n	4u7	2.90

f_O	C	L	X
106k	1n5	1m5	1.00k
106k	150p	15m	10.0k
106k	15n	150u	100
106k	150n	15u	10.0
106k	33p	68m	45.4k
106k	68p	33m	22.0k
106k	330p	6m8	4.54k
106k	680p	3m3	2.20k
106k	3n3	680u	454
106k	6n8	330u	220
106k	33n	68u	45.4
106k	330n	6u8	4.54
106k	680n	3u3	2.20
106k	68n	33u	22.0
107k	4n7	470u	316
107k	47p	47m	31.6k
107k	470p	4m7	3.16k
107k	47n	47u	31.6
107k	470n	4u7	3.16
107k	22p	100m	67.4k
107k	100p	22m	14.8k
107k	220p	10m	6.74k
107k	1n	2m2	1.48k
107k	2n2	1m	674
107k	10n	220u	148
107k	22n	100u	67.4
107k	100n	22u	14.8
107k	220n	10u	6.74
107k	1u	2u2	1.48
117k	56p	33m	24.3k
117k	560p	3m3	2.43k
117k	5n6	330u	243
117k	56n	33u	24.3
117k	560n	3u3	2.43
117k	27n	68u	50.2
117k	27p	68m	50.2k
117k	270p	6m8	5.02k

f_O	C	L	X
117k	2n7	680u	502
117k	270n	6u8	5.02
118k	39p	47m	34.7k
118k	390p	4m7	3.47k
118k	3n9	470u	347
118k	39n	47u	34.7
118k	390n	4u7	3.47
118k	82p	22m	16.4k
118k	820p	2m2	1.64k
118k	8n2	220u	164
118k	82n	22u	16.4
118k	820n	2u2	1.64
119k	1n2	1m5	1.12k
119k	18p	100m	74.5k
119k	120p	15m	11.2k
119k	180p	10m	7.45k
119k	1n8	1m	745
119k	12n	150u	112
119k	18n	100u	74.5
119k	120n	15u	11.2
119k	180n	10u	7.45
128k	33p	47m	37.7k
128k	47p	33m	26.5k
128k	330p	4m7	3.77k
128k	470p	3m3	2.65k
128k	3n3	470u	377
128k	330n	4u7	3.77
128k	470n	3u3	2.65
128k	4n7	330u	265
128k	33n	47u	37.7
128k	47n	33u	26.5
130k	100p	15m	12.2k
130k	15p	100m	81.6k
130k	150p	10m	8.16k
130k	1n	1m5	1.22k
130k	1n5	1m	816
130k	10n	150u	122

f_O	C	L	X
130k	15n	100u	81.6
130k	100n	15u	12.2
130k	150n	10u	8.16
130k	1u	1u5	1.22
130k	2n2	680u	556
130k	6n8	220u	180
130k	22n	68u	55.6
130k	220n	6u8	5.56
130k	680n	2u2	1.80
130k	22p	68m	55.6k
130k	68p	22m	18.0k
130k	220p	6m8	5.56k
130k	680p	2m2	1.80k
130k	68n	22u	18.0
140k	39p	33m	29.1k
140k	390p	3m3	2.91k
140k	3n9	330u	291
140k	390n	3u3	2.91
140k	39n	33u	29.1
141k	27p	47m	41.7k
141k	270p	4m7	4.17k
141k	2n7	470u	417
141k	27n	47u	41.7
141k	270n	4u7	4.17
143k	5n6	220u	198
143k	56n	22u	19.8
143k	56p	22m	19.8k
143k	560p	2m2	1.98k
143k	560n	2u2	1.98
144k	82p	15m	13.5k
144k	820p	1m5	1.35k
144k	8n2	150u	135
144k	82n	15u	13.5
144k	820n	1u5	1.35
144k	18p	68m	61.5k
144k	180p	6m8	6.15k
144k	1n8	680u	615

f_O	C	L	X
144k	18n	68u	61.5
144k	180n	6u8	6.15
145k	12p	100m	91.3k
145k	120p	10m	9.13k
145k	1n2	1m	913
145k	12n	100u	91.3
145k	120n	10u	9.13
153k	33p	33m	31.6k
153k	330p	3m3	3.16k
153k	3n3	330u	316
153k	33n	33u	31.6
153k	330n	3u3	3.16
157k	22p	47m	46.2k
157k	47p	22m	21.6k
157k	220p	4m7	4.62k
157k	470p	2m2	2.16k
157k	2n2	470u	462
157k	4n7	220u	216
157k	22n	47u	46.2
157k	47n	22u	21.6
157k	220n	4u7	4.62
157k	470n	2u2	2.16
158k	15p	68m	67.3k
158k	68p	15m	14.9k
158k	150p	6m8	6.73k
158k	680p	1m5	1.49k
158k	1n5	680u	673
158k	6n8	150u	149
158k	15n	68u	67.3
158k	150n	6u8	6.73
158k	680n	1u5	1.49
158k	68n	15u	14.9
159k	10p	100m	100k
159k	100p	10m	10.0k
159k	1n	1m	1.00k
159k	10n	100u	100
159k	100n	10u	10.0

f_O	C	L	X
159k	1u	1u	1.00
169k	27p	33m	35.0k
169k	270p	3m3	3.50k
169k	2n7	330u	350
169k	270n	3u3	3.50
169k	27n	33u	35.0
172k	3n9	220u	238
172k	390n	2u2	2.38
172k	39p	22m	23.8k
172k	390p	2m2	2.38k
172k	39n	22u	23.8
173k	18p	47m	51.1k
173k	180p	4m7	5.11k
173k	1n8	470u	511
173k	180n	4u7	5.11
173k	18n	47u	51.1
174k	5n6	150u	164
174k	56p	15m	16.4k
174k	560p	1m5	1.64k
174k	56n	15u	16.4
174k	560n	1u5	1.64
176k	8n2	100u	110
176k	82p	10m	11.0k
176k	820p	1m	1.10k
176k	82n	10u	11.0
176k	820n	1u	1.10
176k	120p	6m8	7.53k
176k	1n2	680u	753
176k	12n	68u	75.3
176k	120n	6u8	7.53
176k	12p	68m	75.3k
187k	22p	33m	38.7k
187k	220p	3m3	3.87k
187k	330p	2m2	2.58k
187k	2n2	330u	387
187k	3n3	220u	258
187k	220n	3u3	3.87

f_O	C	L	X
187k	330n	2u2	2.58
187k	33p	22m	25.8k
187k	22n	33u	38.7
187k	33n	22u	25.8
190k	15p	47m	56.0k
190k	47p	15m	17.9k
190k	150p	4m7	5.60k
190k	470p	1m5	1.79k
190k	1n5	470u	560
190k	4n7	150u	179
190k	15n	47u	56.0
190k	47n	15u	17.9
190k	150n	4u7	5.60
190k	470n	1u5	1.79
193k	10p	68m	82.5k
193k	100p	6m8	8.25k
193k	6n8	100u	121
193k	10n	68u	82.5
193k	100n	6u8	8.25
193k	680n	1u	1.21
193k	68p	10m	12.1k
193k	680p	1m	1.21k
193k	1n	680u	825
193k	68n	10u	12.1
207k	18p	33m	42.8k
207k	27p	22m	28.5k
207k	180p	3m3	4.28k
207k	1n8	330u	428
207k	2n7	220u	285
207k	27n	22u	28.5
207k	180n	3u3	4.28
207k	270p	2m2	2.85k
207k	18n	33u	42.8
207k	270n	2u2	2.85
208k	390n	1u5	1.96
208k	39p	15m	19.6k
208k	390p	1m5	1.96k

f_O	C	L	X
208k	3n9	150u	196
208k	39n	15u	19.6
212k	12p	47m	62.6k
212k	120p	4m7	6.26k
212k	1n2	470u	626
212k	12n	47u	62.6
212k	120n	4u7	6.26
213k	56p	10m	13.4k
213k	5n6	100u	134
213k	56n	10u	13.4
213k	560p	1m	1.34k
213k	560n	1u	1.34
213k	8n2	68u	91.1
213k	82p	6m8	9.11k
213k	820p	680u	911
213k	82n	6u8	9.11
226k	15p	33m	46.9k
226k	33p	15m	21.3k
226k	150p	3m3	4.69k
226k	330p	1m5	2.13k
226k	1n5	330u	469
226k	3n3	150u	213
226k	33n	15u	21.3
226k	150n	3u3	4.69
226k	330n	1u5	2.13
226k	15n	33u	46.9
229k	22p	22m	31.6k
229k	220p	2m2	3.16k
229k	2n2	220u	316
229k	22n	22u	31.6
229k	220n	2u2	3.16
232k	10p	47m	68.6k
232k	100p	4m7	6.86k
232k	4n7	100u	146
232k	100n	4u7	6.86
232k	47p	10m	14.6k
232k	470p	1m	1.46k
232k	1n	470u	686
232k	10n	47u	68.6
232k	47n	10u	14.6
232k	470n	1u	1.46
234k	6n8	68u	100
234k	68p	6m8	10.0k
234k	680p	680u	1.00k
234k	68n	6u8	10.0
250k	27p	15m	23.6k
250k	270p	1m5	2.36k
250k	2n7	150u	236
250k	27n	15u	23.6
250k	270n	1u5	2.36
253k	12p	33m	52.4k
253k	18p	22m	35.0k
253k	120p	3m3	5.24k
253k	180p	2m2	3.50k
253k	1n2	330u	524
253k	1n8	220u	350
253k	12n	33u	52.4
253k	18n	22u	35.0
253k	120n	3u3	5.24
253k	180n	2u2	3.50
255k	39p	10m	16.0k
255k	390p	1m	1.60k
255k	3n9	100u	160
255k	39n	10u	16.0
255k	390n	1u	1.60
256k	82p	4m7	7.57k
256k	820p	470u	757
256k	8n2	47u	75.7
256k	82n	4u7	7.57
258k	5n6	68u	110
258k	56n	6u8	11.0
258k	56p	6m8	11.0k
258k	560p	680u	1.10k
277k	10p	33m	57.4k

f_O	C	L	X
277k	22p	15m	26.1k
277k	150p	2m2	3.83k
277k	220p	1m5	2.61k
277k	1n5	220u	383
277k	2n2	150u	261
277k	3n3	100u	174
277k	100n	3u3	5.74
277k	220n	1u5	2.61
277k	15p	22m	38.3k
277k	33p	10m	17.4k
277k	100p	3m3	5.74k
277k	330p	1m	1.74k
277k	1n	330u	574
277k	10n	33u	57.4
277k	15n	22u	38.3
277k	22n	15u	26.1
277k	33n	10u	17.4
277k	150n	2u2	3.83
277k	330n	1u	1.74
282k	47p	6m8	12.0k
282k	68p	4m7	8.31k
282k	470p	680u	1.20k
282k	680p	470u	831
282k	4n7	68u	120
282k	6n8	47u	83.1
282k	47n	6u8	12.0
282k	68n	4u7	8.31
306k	82p	3m3	6.34k
306k	820p	330u	634
306k	8n2	33u	63.4
306k	82n	3u3	6.34
306k	18p	15m	28.9k
306k	180p	1m5	2.89k
306k	1n8	150u	289
306k	180n	1u5	2.89
306k	27p	10m	19.2k
306k	270p	1m	1.92k

f_O	C	L	X
306k	2n7	100u	192
306k	18n	15u	28.9
306k	27n	10u	19.2
306k	270n	1u	1.92
309k	3n9	68u	132
309k	39n	6u8	13.2
309k	39p	6m8	13.2k
309k	390p	680u	1.32k
310k	12p	22m	42.8k
310k	120p	2m2	4.28k
310k	1n2	220u	428
310k	12n	22u	42.8
310k	120n	2u2	4.28
310k	56p	4m7	9.16k
310k	560p	470u	916
310k	5n6	47u	91.6
310k	56n	4u7	9.16
336k	15p	15m	31.6k
336k	150p	1m5	3.16k
336k	1n5	150u	316
336k	15n	15u	31.6
336k	150n	1u5	3.16
336k	3n3	68u	144
336k	33p	6m8	14.4k
336k	68p	3m3	6.97k
336k	330p	680u	1.44k
336k	680p	330u	697
336k	6n8	33u	69.7
336k	33n	6u8	14.4
336k	68n	3u3	6.97
339k	47p	4m7	10.0k
339k	470p	470u	1.00k
339k	4n7	47u	100
339k	47n	4u7	10.0
339k	10p	22m	46.9k
339k	22p	10m	21.3k
339k	100p	2m2	4.69k

f_O	C	L	X
339k	2n2	100u	213
339k	100n	2u2	4.69
339k	220n	1u	2.13
339k	220p	1m	2.13k
339k	1n	220u	469
339k	10n	22u	46.9
339k	22n	10u	21.3
370k	56p	3m3	7.68k
370k	5n6	33u	76.8
370k	56n	3u3	7.68
370k	560p	330u	768
371k	27p	6m8	15.9k
371k	270p	680u	1.59k
371k	2n7	68u	159
371k	27n	6u8	15.9
372k	39p	4m7	11.0k
372k	390p	470u	1.10k
372k	3n9	47u	110
372k	39n	4u7	11.0
375k	82p	2m2	5.18k
375k	820p	220u	518
375k	8n2	22u	51.8
375k	82n	2u2	5.18
375k	12p	15m	35.4k
375k	18p	10m	23.6k
375k	120p	1m5	3.54k
375k	1n2	150u	354
375k	1n8	100u	236
375k	12n	15u	35.4
375k	18n	10u	23.6
375k	120n	1u5	3.54
375k	180n	1u	2.36
375k	180p	1m	2.36k
404k	33p	4m7	11.9k
404k	47p	3m3	8.38k
404k	330p	470u	1.19k
404k	3n3	47u	119
404k	33n	4u7	11.9
404k	47n	3u3	8.38
404k	470p	330u	838
404k	4n7	33u	83.8
411k	10p	15m	38.7k
411k	15p	10m	25.8k
411k	100p	1m5	3.87k
411k	150p	1m	2.58k
411k	1n	150u	387
411k	1n5	100u	258
411k	10n	15u	38.7
411k	15n	10u	25.8
411k	100n	1u5	3.87
411k	150n	1u	2.58
411k	2n2	68u	176
411k	22p	6m8	17.6k
411k	68p	2m2	5.69k
411k	680p	220u	569
411k	6n8	22u	56.9
411k	22n	6u8	17.6
411k	220p	680u	1.76k
411k	68n	2u2	5.69
444k	39p	3m3	9.20k
444k	390p	330u	920
444k	39n	3u3	9.20
444k	3n9	33u	92.0
447k	27p	4m7	13.2k
447k	270p	470u	1.32k
447k	2n7	47u	132
447k	27n	4u7	13.2
453k	5n6	22u	62.7
453k	56n	2u2	6.27
453k	56p	2m2	6.27k
453k	560p	220u	627
454k	82p	1m5	4.28k
454k	820p	150u	428
454k	8n2	15u	42.8

f_O	C	L	X
454k	82n	1u5	4.28
455k	18p	6m8	19.4k
455k	180p	680u	1.94k
455k	1n8	68u	194
455k	18n	6u8	19.4
459k	1n2	100u	289
459k	12n	10u	28.9
459k	12p	10m	28.9k
459k	120n	1u	2.89
459k	120p	1m	2.89k
482k	33p	3m3	10.0k
482k	330p	330u	1.00k
482k	3n3	33u	100
482k	33n	3u3	10.0
495k	22p	4m7	14.6k
495k	220p	470u	1.46k
495k	2n2	47u	146
495k	47p	2m2	6.84k
495k	470p	220u	684
495k	4n7	22u	68.4
495k	22n	4u7	14.6
495k	47n	2u2	6.84
498k	1n5	68u	213
498k	15p	6m8	21.3k
498k	68p	1m5	4.70k
498k	150p	680u	2.13k
498k	680p	150u	470
498k	6n8	15u	47.0
498k	15n	6u8	21.3
498k	68n	1u5	4.70
503k	10p	10m	31.6k
503k	100p	1m	3.16k
503k	1n	100u	316
503k	10n	10u	31.6
503k	100n	1u	3.16
533k	27p	3m3	11.1k
533k	27n	3u3	11.1

f_O	C	L	X
533k	270p	330u	1.11k
533k	2n7	33u	111
543k	39p	2m2	7.51k
543k	390p	220u	751
543k	3n9	22u	75.1
543k	39n	2u2	7.51
547k	18p	4m7	16.2k
547k	180p	470u	1.62k
547k	1n8	47u	162
547k	18n	4u7	16.2
549k	56p	1m5	5.18k
549k	560p	150u	518
549k	5n6	15u	51.8
549k	56n	1u5	5.18
556k	82p	1m	3.49k
556k	820p	100u	349
556k	8n2	10u	34.9
556k	82n	1u	3.49
557k	120p	680u	2.38k
557k	1n2	68u	238
557k	12n	6u8	23.8
557k	12p	6m8	23.8k
591k	3n3	22u	81.6
591k	22p	3m3	12.2k
591k	33p	2m2	8.16k
591k	220p	330u	1.22k
591k	330p	220u	816
591k	2n2	33u	122
591k	22n	3u3	12.2
591k	33n	2u2	8.16
599k	15p	4m7	17.7k
599k	47p	1m5	5.65k
599k	150p	470u	1.77k
599k	470p	150u	565
599k	1n5	47u	177
599k	4n7	15u	56.5
599k	15n	4u7	17.7

f_O	C	L	X
599k	47n	1u5	5.65
610k	680p	100u	383
610k	1n	68u	261
610k	6n8	10u	38.3
610k	10n	6u8	26.1
610k	10p	6m8	26.1k
610k	68p	1m	3.83k
610k	100p	680u	2.61k
610k	68n	1u	3.83
653k	18p	3m3	13.5k
653k	27p	2m2	9.03k
653k	180p	330u	1.35k
653k	270p	220u	903
653k	1n8	33u	135
653k	2n7	22u	90.3
653k	18n	3u3	13.5
653k	27n	2u2	9.03
658k	39p	1m5	6.20k
658k	390p	150u	620
658k	3n9	15u	62.0
658k	39n	1u5	6.20
670k	12p	4m7	19.8k
670k	120p	470u	1.98k
670k	1n2	47u	198
670k	12n	4u7	19.8
673k	56p	1m	4.23k
673k	560p	100u	423
673k	5n6	10u	42.3
673k	56n	1u	4.23
674k	820p	68u	288
674k	8n2	6u8	28.8
674k	82p	680u	2.88k
715k	33p	1m5	6.74k
715k	150p	330u	1.48k
715k	330p	150u	674
715k	3n3	15u	67.4
715k	15n	3u3	14.8
715k	33n	1u5	6.74
715k	15p	3m3	14.8k
715k	1n5	33u	148
723k	22p	2m2	10.0k
723k	220p	220u	1.00k
723k	2n2	22u	100
723k	22n	2u2	10.0
734k	10p	4m7	21.7k
734k	47p	1m	4.61k
734k	100p	470u	2.17k
734k	470p	100u	461
734k	1n	47u	217
734k	4n7	10u	46.1
734k	10n	4u7	21.7
734k	47n	1u	4.61
740k	680p	68u	316
740k	6n8	6u8	31.6
740k	68p	680u	3.16k
791k	27p	1m5	7.45k
791k	270p	150u	745
791k	2n7	15u	74.5
791k	27n	1u5	7.45
800k	12p	3m3	16.6k
800k	18p	2m2	11.1k
800k	180p	220u	1.11k
800k	1n8	22u	111
800k	12n	3u3	16.6
800k	18n	2u2	11.1
800k	120p	330u	1.66k
800k	1n2	33u	166
806k	3n9	10u	50.6
806k	39p	1m	5.06k
806k	390p	100u	506
806k	39n	1u	5.06
811k	82p	470u	2.39k
811k	820p	47u	239
811k	8n2	4u7	23.9

f_O	C	L	X
816k	560p	68u	348
816k	5n6	6u8	34.8
816k	56p	680u	3.48k
876k	10p	3m3	18.2k
876k	15p	2m2	12.1k
876k	22p	1m5	8.26k
876k	33p	1m	5.50k
876k	100p	330u	1.82k
876k	150p	220u	1.21k
876k	220p	150u	826
876k	330p	100u	550
876k	1n	33u	182
876k	1n5	22u	121
876k	2n2	15u	82.6
876k	3n3	10u	55.0
876k	10n	3u3	18.2
876k	15n	2u2	12.1
876k	22n	1u5	8.26
876k	33n	1u	5.50
890k	47p	680u	3.80k
890k	68p	470u	2.63k
890k	470p	68u	380
890k	680p	47u	263
890k	4n7	6u8	38.0
890k	6n8	4u7	26.3
968k	82p	330u	2.01k
968k	820p	33u	201
968k	8n2	3u3	20.1
969k	18p	1m5	9.13k
969k	180p	150u	913
969k	270p	100u	609
969k	1n8	15u	91.3
969k	2n7	10u	60.9
969k	18n	1u5	9.13
969k	27p	1m	6.09k
969k	27n	1u	6.09
977k	3n9	6u8	41.8

f_O	C	L	X
977k	39p	680u	4.18k
977k	390p	68u	418
980k	12p	2m2	13.5k
980k	120p	220u	1.35k
980k	1n2	22u	135
980k	12n	2u2	13.5
981k	56p	470u	2.90k
981k	5n6	4u7	29.0
981k	560p	47u	290
1.06M	15p	1m5	10.0k
1.06M	150p	150u	1.00k
1.06M	1n5	15u	100
1.06M	15n	1u5	10.0
1.06M	33p	680u	4.54k
1.06M	68p	330u	2.20k
1.06M	330p	68u	454
1.06M	680p	33u	220
1.06M	3n3	6u8	45.4
1.06M	6n8	3u3	22.0
1.07M	47p	470u	3.16k
1.07M	4n7	4u7	31.6
1.07M	470p	47u	316
1.07M	220p	100u	674
1.07M	2n2	10u	67.4
1.07M	10p	2m2	14.8k
1.07M	22p	1m	6.74k
1.07M	100p	220u	1.48k
1.07M	1n	22u	148
1.07M	10n	2u2	14.8
1.07M	22n	1u	6.74
1.17M	5n6	3u3	24.3
1.17M	56p	330u	2.43k
1.17M	560p	33u	243
1.17M	270p	68u	502
1.17M	2n7	6u8	50.2
1.17M	27p	680u	5.02k
1.18M	39p	470u	3.47k

f_O	C	L	X
1.18M	390p	47u	347
1.18M	3n9	4u7	34.7
1.18M	82p	220u	1.64k
1.18M	820p	22u	164
1.18M	8n2	2u2	16.4
1.19M	12p	1m5	11.2k
1.19M	18p	1m	7.45k
1.19M	120p	150u	1.12k
1.19M	180p	100u	745
1.19M	1n2	15u	112
1.19M	1n8	10u	74.5
1.19M	12n	1u5	11.2
1.19M	18n	1u	7.45
1.28M	3n3	4u7	37.7
1.28M	33p	470u	3.77k
1.28M	47p	330u	2.65k
1.28M	330p	47u	377
1.28M	470p	33u	265
1.28M	4n7	3u3	26.5
1.30M	10p	1m5	12.2k
1.30M	100p	150u	1.22k
1.30M	150p	100u	816
1.30M	1n	15u	122
1.30M	1n5	10u	81.6
1.30M	10n	1u5	12.2
1.30M	15n	1u	8.16
1.30M	15p	1m	8.16k
1.30M	2n2	6u8	55.6
1.30M	22p	680u	5.56k
1.30M	68p	220u	1.80k
1.30M	220p	68u	556
1.30M	680p	22u	180
1.30M	6n8	2u2	18.0
1.40M	39p	330u	2.91k
1.40M	3n9	3u3	29.1
1.40M	390p	33u	291
1.41M	2n7	4u7	41.7
1.41M	27p	470u	4.17k
1.41M	270p	47u	417
1.43M	56p	220u	1.98k
1.43M	560p	22u	198
1.43M	5n6	2u2	19.8
1.44M	820p	15u	135
1.44M	8n2	1u5	13.5
1.44M	82p	150u	1.35k
1.44M	18p	680u	6.15k
1.44M	180p	68u	615
1.44M	1n8	6u8	61.5
1.45M	120p	100u	913
1.45M	1n2	10u	91.3
1.45M	12p	1m	9.13k
1.45M	12n	1u	9.13
1.53M	3n3	3u3	31.6
1.53M	33p	330u	3.16k
1.53M	330p	33u	316
1.57M	2n2	4u7	46.2
1.57M	22p	470u	4.62k
1.57M	47p	220u	2.16k
1.57M	220p	47u	462
1.57M	470p	22u	216
1.57M	4n7	2u2	21.6
1.58M	1n5	6u8	67.3
1.58M	6n8	1u5	14.9
1.58M	15p	680u	6.73k
1.58M	68p	150u	1.49k
1.58M	150p	68u	673
1.58M	680p	15u	149
1.59M	10p	1m	10.0k
1.59M	100p	100u	1.00k
1.59M	1n	10u	100
1.59M	10n	1u	10.0
1.69M	27p	330u	3.50k
1.69M	270p	33u	350
1.69M	2n7	3u3	35.0

f_O	C	L	X
1.72M	3n9	2u2	23.8
1.72M	39p	220u	2.38k
1.72M	390p	22u	238
1.73M	18p	470u	5.11k
1.73M	1n8	4u7	51.1
1.73M	180p	47u	511
1.74M	56p	150u	1.64k
1.74M	560p	15u	164
1.74M	5n6	1u5	16.4
1.76M	82p	100u	1.10k
1.76M	820p	10u	110
1.76M	8n2	1u	11.0
1.76M	12p	680u	7.53k
1.76M	120p	68u	753
1.76M	1n2	6u8	75.3
1.87M	330p	22u	258
1.87M	2n2	3u3	38.7
1.87M	3n3	2u2	25.8
1.87M	22p	330u	3.87k
1.87M	33p	220u	2.58k
1.87M	220p	33u	387
1.90M	15p	470u	5.60k
1.90M	150p	47u	560
1.90M	470p	15u	179
1.90M	1n5	4u7	56.0
1.90M	4n7	1u5	17.9
1.90M	47p	150u	1.79k
1.93M	10p	680u	8.25k
1.93M	100p	68u	825
1.93M	680p	10u	121
1.93M	1n	6u8	82.5
1.93M	6n8	1u	12.1
1.93M	68p	100u	1.21k
2.07M	1n8	3u3	42.8
2.07M	2n7	2u2	28.5
2.07M	18p	330u	4.28k
2.07M	27p	220u	2.85k
2.07M	270p	22u	285
2.07M	180p	33u	428
2.08M	3n9	1u5	19.6
2.08M	39p	150u	1.96k
2.08M	390p	15u	196
2.12M	12p	470u	6.26k
2.12M	120p	47u	626
2.12M	1n2	4u7	62.6
2.13M	56p	100u	1.34k
2.13M	560p	10u	134
2.13M	5n6	1u	13.4
2.13M	82p	68u	911
2.13M	820p	6u8	91.1
2.26M	33p	150u	2.13k
2.26M	330p	15u	213
2.26M	1n5	3u3	46.9
2.26M	3n3	1u5	21.3
2.26M	15p	330u	4.69k
2.26M	150p	33u	469
2.29M	220p	22u	316
2.29M	2n2	2u2	31.6
2.29M	22p	220u	3.16k
2.32M	10p	470u	6.86k
2.32M	47p	100u	1.46k
2.32M	100p	47u	686
2.32M	470p	10u	146
2.32M	1n	4u7	68.6
2.32M	4n7	1u	14.6
2.34M	68p	68u	1.00k
2.34M	680p	6u8	100
2.50M	27p	150u	2.36k
2.50M	270p	15u	236
2.50M	2n7	1u5	23.6
2.53M	12p	330u	5.24k
2.53M	18p	220u	3.50k
2.53M	180p	22u	350
2.53M	1n2	3u3	52.4

f_0	C	L	X
2.53M	1n8	2u2	35.0
2.53M	120p	33u	524
2.55M	39p	100u	1.60k
2.55M	390p	10u	160
2.55M	3n9	1u	16.0
2.56M	82p	47u	757
2.56M	820p	4u7	75.7
2.58M	56p	68u	1.10k
2.58M	560p	6u8	110
2.77M	2n2	1u5	26.1
2.77M	22p	150u	2.61k
2.77M	150p	22u	383
2.77M	220p	15u	261
2.77M	1n5	2u2	38.3
2.77M	3n3	1u	17.4
2.77M	10p	330u	5.74k
2.77M	15p	220u	3.83k
2.77M	33p	100u	1.74k
2.77M	100p	33u	574
2.77M	330p	10u	174
2.77M	1n	3u3	57.4
2.82M	47p	68u	1.20k
2.82M	68p	47u	831
2.82M	470p	6u8	120
2.82M	680p	4u7	83.1
3.06M	820p	3u3	63.4
3.06M	82p	33u	634
3.06M	18p	150u	2.89k
3.06M	27p	100u	1.92k
3.06M	180p	15u	289
3.06M	270p	10u	192
3.06M	1n8	1u5	28.9
3.06M	2n7	1u	19.2
3.09M	39p	68u	1.32k
3.09M	390p	6u8	132
3.10M	12p	220u	4.28k
3.10M	1n2	2u2	42.8

f_0	C	L	X
3.10M	120p	22u	428
3.10M	56p	47u	916
3.10M	560p	4u7	91.6
3.36M	150p	15u	316
3.36M	1n5	1u5	31.6
3.36M	15p	150u	3.16k
3.36M	33p	68u	1.44k
3.36M	330p	6u8	144
3.36M	680p	3u3	69.7
3.36M	68p	33u	697
3.39M	47p	47u	1.00k
3.39M	470p	4u7	100
3.39M	10p	220u	4.69k
3.39M	22p	100u	2.13k
3.39M	100p	22u	469
3.39M	220p	10u	213
3.39M	1n	2u2	46.9
3.39M	2n2	1u	21.3
3.70M	56p	33u	768
3.70M	560p	3u3	76.8
3.71M	27p	68u	1.59k
3.71M	270p	6u8	159
3.72M	39p	47u	1.10k
3.72M	390p	4u7	110
3.75M	820p	2u2	51.8
3.75M	82p	22u	518
3.75M	1n2	1u5	35.4
3.75M	12p	150u	3.54k
3.75M	18p	100u	2.36k
3.75M	120p	15u	354
3.75M	180p	10u	236
3.75M	1n8	1u	23.6
4.04M	33p	47u	1.19k
4.04M	330p	4u7	119
4.04M	470p	3u3	83.8
4.04M	47p	33u	838
4.11M	10p	150u	3.87k

f_O	C	L	X
4.11M	15p	100u	2.58k
4.11M	100p	15u	387
4.11M	150p	10u	258
4.11M	1n	1u5	38.7
4.11M	1n5	1u	25.8
4.11M	22p	68u	1.76k
4.11M	68p	22u	569
4.11M	220p	6u8	176
4.11M	680p	2u2	56.9
4.44M	39p	33u	920
4.44M	390p	3u3	92.0
4.47M	27p	47u	1.32k
4.47M	270p	4u7	132
4.53M	56p	22u	627
4.53M	560p	2u2	62.7
4.54M	82p	15u	428
4.54M	820p	1u5	42.8
4.55M	18p	68u	1.94k
4.55M	180p	6u8	194
4.59M	12p	100u	2.89k
4.59M	120p	10u	289
4.59M	1n2	1u	28.9
4.82M	33p	33u	1.00k
4.82M	330p	3u3	100
4.95M	22p	47u	1.46k
4.95M	47p	22u	684
4.95M	220p	4u7	146
4.95M	470p	2u2	68.4
4.98M	15p	68u	2.13k
4.98M	68p	15u	470
4.98M	150p	6u8	213
4.98M	680p	1u5	47.0
5.03M	10p	100u	3.16k
5.03M	100p	10u	316
5.03M	1n	1u	31.6
5.33M	270p	3u3	111
5.33M	27p	33u	1.11k

f_O	C	L	X
5.43M	39p	22u	751
5.43M	390p	2u2	75.1
5.47M	18p	47u	1.62k
5.47M	180p	4u7	162
5.49M	56p	15u	518
5.49M	560p	1u5	51.8
5.56M	82p	10u	349
5.56M	820p	1u	34.9
5.57M	12p	68u	2.38k
5.57M	120p	6u8	238
5.91M	22p	33u	1.22k
5.91M	33p	22u	816
5.91M	220p	3u3	122
5.91M	330p	2u2	81.6
5.99M	15p	47u	1.77k
5.99M	47p	15u	565
5.99M	150p	4u7	177
5.99M	470p	1u5	56.5
6.10M	10p	68u	2.61k
6.10M	100p	6u8	261
6.10M	68p	10u	383
6.10M	680p	1u	38.3
6.53M	27p	22u	903
6.53M	180p	3u3	135
6.53M	270p	2u2	90.3
6.53M	18p	33u	1.35k
6.58M	39p	15u	620
6.58M	390p	1u5	62.0
6.70M	120p	4u7	198
6.70M	12p	47u	1.98k
6.73M	56p	10u	423
6.73M	560p	1u	42.3
6.74M	82p	6u8	288
7.15M	150p	3u3	148
7.15M	330p	1u5	67.4
7.15M	33p	15u	674
7.15M	15p	33u	1.48k

f_O	C	L	X
7.23M	22p	22u	1.00k
7.23M	220p	2u2	100
7.34M	10p	47u	2.17k
7.34M	47p	10u	461
7.34M	100p	4u7	217
7.34M	470p	1u	46.1
7.40M	68p	6u8	316
7.91M	27p	15u	745
7.91M	270p	1u5	74.5
8.00M	18p	22u	1.11k
8.00M	120p	3u3	166
8.00M	180p	2u2	111
8.00M	12p	33u	1.66k
8.06M	39p	10u	506
8.06M	390p	1u	50.6
8.11M	82p	4u7	239
8.16M	56p	6u8	348
8.76M	22p	15u	826
8.76M	100p	3u3	182
8.76M	150p	2u2	121
8.76M	220p	1u5	82.6
8.76M	10p	33u	1.82k
8.76M	15p	22u	1.21k
8.76M	33p	10u	550
8.76M	330p	1u	55.0
8.90M	47p	6u8	380
8.90M	68p	4u7	263
9.68M	82p	3u3	201
9.69M	18p	15u	913
9.69M	180p	1u5	91.3
9.69M	27p	10u	609
9.69M	270p	1u	60.9
9.77M	39p	6u8	418
9.80M	12p	22u	1.35k
9.80M	120p	2u2	135
9.81M	56p	4u7	290
10.6M	150p	1u5	100
10.6M	15p	15u	1.00k
10.6M	33p	6u8	454
10.6M	68p	3u3	220
10.7M	47p	4u7	316
10.7M	10p	22u	1.48k
10.7M	22p	10u	674
10.7M	100p	2u2	148
10.7M	220p	1u	67.4
11.7M	56p	3u3	243
11.7M	27p	6u8	502
11.8M	39p	4u7	347
11.8M	82p	2u2	164
11.9M	12p	15u	1.12k
11.9M	18p	10u	745
11.9M	120p	1u5	112
11.9M	180p	1u	74.5
12.8M	33p	4u7	377
12.8M	47p	3u3	265
13.0M	10p	15u	1.22k
13.0M	15p	10u	816
13.0M	100p	1u5	122
13.0M	150p	1u	81.6
13.0M	22p	6u8	556
13.0M	68p	2u2	180
14.0M	39p	3u3	291
14.1M	27p	4u7	417
14.3M	56p	2u2	198
14.4M	82p	1u5	135
14.4M	18p	6u8	615
14.5M	12p	10u	913
14.5M	120p	1u	91.3
15.3M	33p	3u3	316
15.7M	22p	4u7	462
15.7M	47p	2u2	216
15.8M	15p	6u8	673
15.8M	68p	1u5	149
15.9M	10p	10u	1.00k

f_O	C	L	X
15.9M	100p	1u	100
16.9M	27p	3u3	350
17.2M	39p	2u2	238
17.3M	18p	4u7	511
17.4M	56p	1u5	164
17.6M	82p	1u	110
17.6M	12p	6u8	753
18.7M	22p	3u3	387
18.7M	33p	2u2	258
19.0M	15p	4u7	560
19.0M	47p	1u5	179
19.3M	10p	6u8	825
19.3M	68p	1u	121
20.7M	18p	3u3	428
20.7M	27p	2u2	285
20.8M	39p	1u5	196
21.2M	12p	4u7	626
21.3M	56p	1u	134
22.6M	15p	3u3	469
22.6M	33p	1u5	213
22.9M	22p	2u2	316
23.2M	10p	4u7	686
23.2M	47p	1u	146
25.0M	27p	1u5	236
25.3M	12p	3u3	524
25.3M	18p	2u2	350
25.5M	39p	1u	160
27.7M	10p	3u3	574
27.7M	15p	2u2	383
27.7M	22p	1u5	261
27.7M	33p	1u	174
30.6M	18p	1u5	289
30.6M	27p	1u	192
31.0M	12p	2u2	428
33.6M	15p	1u5	316
33.9M	10p	2u2	469
33.9M	22p	1u	213
37.5M	12p	1u5	354
37.5M	18p	1u	236
41.1M	10p	1u5	387
41.1M	15p	1u	258
45.9M	12p	1u	289
50.3M	10p	1u	316

Part Four

Operational Amplifier Circuits

This Page Intentionally Left Blank

11 Amplifier gains

Introduction

Operational amplifiers, or op-amps, are perhaps the most common building blocks found in analogue circuits. They can be used to create amplification stages very easily, and many other types of circuit besides. There are very many 'op-amp cookbooks' around that give lots of these circuits, so we won't add to them here. The main purpose of this chapter is to present you with a pair of tables to enable easy selection of resistor values for any desired gain.

We'll take a moment, however, to state the two rules that govern an op-amp's behaviour for the purpose of achieving a first-off analysis of their circuits:

1. The input terminals of op-amps draw no current.
2. The output of the op-amp does whatever it can to keep the voltage between the input terminals at zero.

These two statements are something of a generalization, but they are fine if you want to work out what an op-amp circuit is doing.

There are two common ways to use op-amps as amplifier stages. We look at them next.

Inverting op-amp

Figure 11.1 shows a basic inverting op-amp circuit. This circuit can be used to give either attenuation or gain with two resistors. The output is in antiphase with the input. The gain is given by:

$$A_V = \frac{R_2}{R_1} \tag{11.1}$$

Values of R_1 and R_2 can be found from Table 11.1, where gain is given both as a ratio and in dBs. R_1 and R_2 may, of course, be scaled by factors of 10 without affecting the gain. The Table only shows values above unity gain (0 dB). If you need to select a pair of resistors to give attenuation, then the Table can still be used, but the positions of R_1 and R_2 should be switched. The dB column then reads attenuation directly, or the A_V column the reciprocal of gain expressed as a ratio.

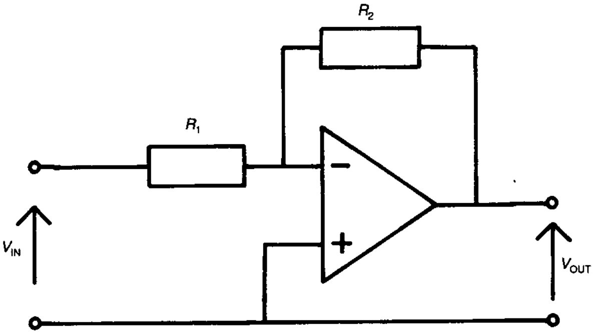

Figure 11.1 Inverting op-amp

The input resistance of the circuit is equal to the value of the input resistor, R_1. The output resistance is, for practical purposes, negligible. (Limitations on its ability to drive a load will be dependent on maximum values of output current, voltage swing and slew rate, for which the manufacturer's data on the chip used should be consulted.)

It is sometimes desirable to limit the frequency response of the circuit. The standard method is to add suitable value capacitors, as in Figure 11.2. C_1, in series with the input resistor, limits low frequency response, while C_2, in parallel with R_2, limits high frequency response. Each gives a 6 dB/octave slope, as for the first order LPF and HPF described in Chapter 9. The equations for the lower −3 dB point, f_{t-}, and the upper −3 dB point, f_{t+}, are also similar:

$$f_{t-} = \frac{1}{2\pi R_1 C_1} \quad \text{and} \quad f_{t+} = \frac{1}{2\pi R_2 C_2} \tag{11.2a}$$

Example 11.1: give suitable values for an inverting op-amp circuit with a gain of 15 dB, and an input resistance of between 20k and 30k. Determine suitable values for C_1 and C_2 to make the upper and lower −3 dB points as near as possible to 100 Hz and 15 kHz respectively, using standard E12 capacitors.

Table 11.1 gives many combinations, close to 15 dB, but to meet the requirement for input resistance R_1 should be between 20k and 30k. We take the pair 2k7/15k, and multiply by 10, so that R_1 = 27k and R_2 = 150k.

To find values for C_1 and C_2, we transpose Eqn (11.2a):

$$C = \frac{1}{2\pi f_t R} \tag{11.2b}$$

and get:

1. C_1 = 59 nF (ideal) – nearest is 56 nF, substitute back into Eqn (11.2) to get f_{t-} = 105 Hz
2. C_2 = 71 pF (ideal) – nearest is 68 pF, giving f_{t+} = 15.6 kHz

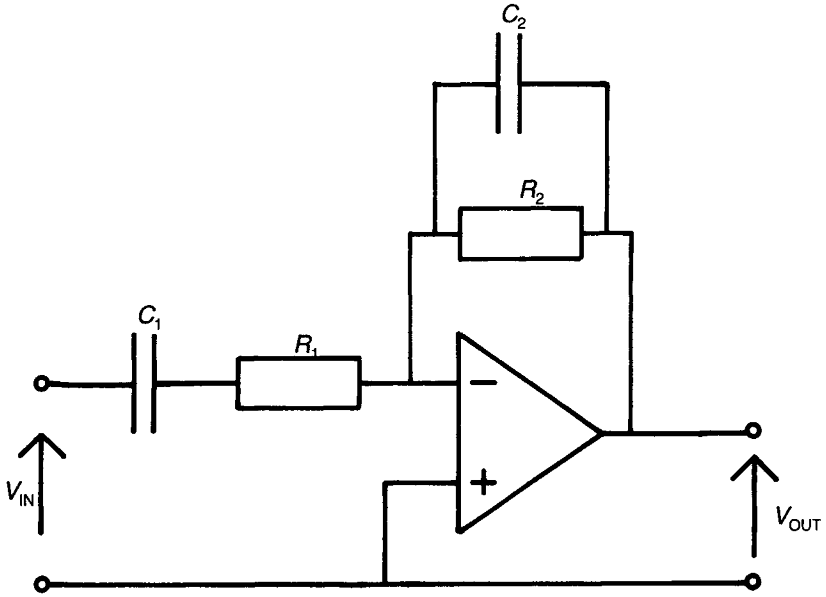

Figure 11.2 Inverting op-amp with restricted bandwidth

Non-inverting op-amp

Figure 11.3 shows a non-inverting op-amp circuit, in its most basic configuration. This circuit has a gain as follows:

$$A_V = \frac{R_1 + R_2}{R_1} \qquad (11.3)$$

This means that the circuit can only provide gain of greater than 1. Its input impedance is very large, and is not usually taken into account. As with the inverting op-amp, its output impedance is very low, with the same restrictions on output drive applying (Figure 11.4).

Where it is desired to restrict the bandwidth of the stage, capacitors C_1 and C_2 are added. The calculations are identical to those for the inverting stage. Note that with C_1 in this position the input signal must have some defined DC level. If the input is connected using a DC block capacitor then C_1 must be omitted so that the op-amp has a DC reference. If it is desired to limit the low frequency response of the stage it is sometimes better to connect the input via a suitable high-pass filter, as described in Chapter 9.

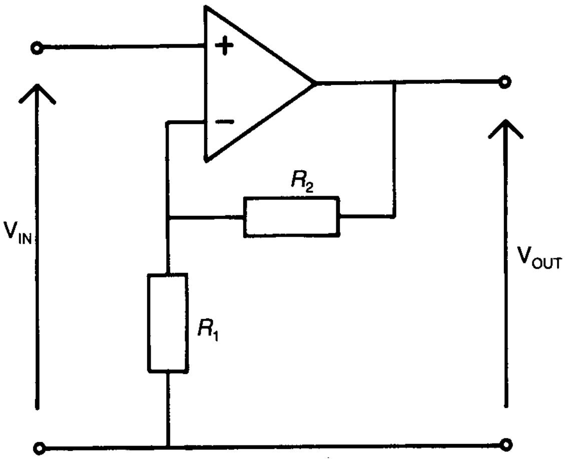

Figure 11.3 Non-inverting op-amp

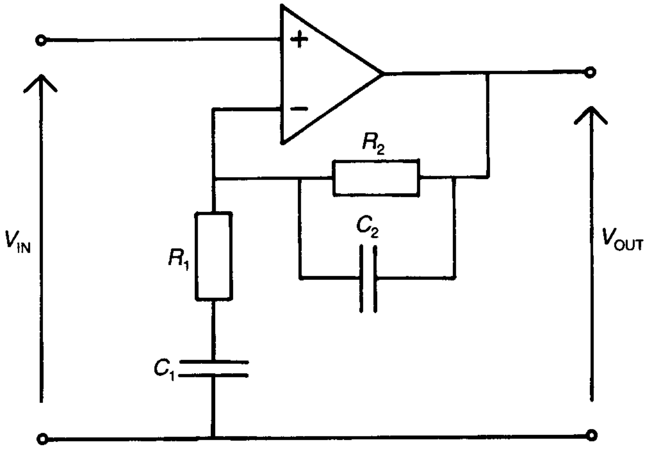

Figure 11.4 Non-inverting op-amp with restricted bandwidth

The main advantage of this stage (other than its being non-inverting) is that of its input impedance. It makes an ideal buffer stage, after circuits which are sensitive to load impedance (such as the reactive circuits described in Part Two).

Where it is desired to use this circuit without gain, purely as a buffer stage, it is customary to omit R_1 and make R_2 short circuit – i.e. to link the output terminal to the inverting input terminal, as in Figure 11.5. This is a 'voltage follower'. The output is identical to the input, but comes from a very low source impedance.

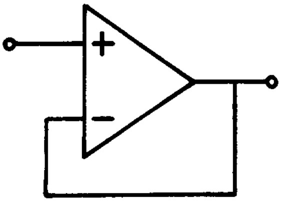

Figure 11.5 A voltage follower

Values for R_1 and R_2 for the non-inverting amplifier can be selected using Table 8.1, in the 'Voltage dividers' chapter. The dB column gives values of gain directly, and the A column values of the reciprocal of gain as a ratio. R_1 from the table should be fitted as position R_2 from Figure 11.3 and R_2 as R_1. As usual, R_1 and R_2 can be scaled by factors of 10 if desired.

Table 11.1 *Inverting op-amp gains*

A_V	dB	R1	R2
1.07	00.56	1k5	1k6
1.08	00.70	1k2	1k3
1.08	00.70	3k6	3k9
1.09	00.71	4k7	5k1
1.09	00.76	1k1	1k2
1.09	00.76	2k2	2k4
1.09	00.76	3k3	3k6
1.09	00.77	4k3	4k7
1.09	00.78	7k5	8k2
1.10	00.80	6k2	6k8
1.10	00.81	5k1	5k6
1.10	00.82	9k1	10k
1.10	00.83	1k	1k1
1.10	00.83	2k	2k2
1.10	00.83	3k	3k3
1.10	00.85	3k9	4k3
1.10	00.85	6k8	7k5
1.11	00.88	5k6	6k2
1.11	00.90	8k2	9k1
1.11	00.92	1k8	2k
1.11	00.92	2k7	3k
1.13	01.02	1k6	1k8
1.13	01.02	2k4	2k7
1.15	01.24	1k3	1k5
1.18	01.45	1k1	1k3
1.18	01.45	3k3	3k9
1.19	01.48	4k3	5k1
1.19	01.52	4k7	5k6
1.19	01.54	3k6	4k3
1.20	01.58	1k	1k2
1.20	01.58	2k	2k4
1.20	01.58	1k5	1k8
1.20	01.58	3k	3k6
1.21	01.62	3k9	4k7
1.21	01.63	6k8	8k2
1.21	01.65	9k1	11k
1.21	01.65	6k2	7k5
1.21	01.68	7k5	9k1
1.21	01.69	5k6	6k8
1.22	01.70	5k1	6k2
1.22	01.72	8k2	10k
1.22	01.74	1k8	2k2
1.22	01.74	2k7	3k3
1.23	01.78	2k2	2k7
1.23	01.80	1k3	1k6
1.25	01.94	1k2	1k5
1.25	01.94	1k6	2k
1.25	01.94	2k4	3k
1.30	02.28	1k	1k3
1.30	02.28	3k	3k9
1.30	02.29	4k3	5k6
1.30	02.30	3k3	4k3
1.31	02.32	3k6	4k7
1.31	02.33	3k9	5k1
1.32	02.40	9k1	12k
1.32	02.41	4k7	6k2
1.32	02.43	6k2	8k2
1.33	02.50	1k8	2k4
1.33	02.50	7k5	10k
1.33	02.50	1k2	1k6
1.33	02.50	1k5	2k
1.33	02.50	2k7	3k6
1.33	02.50	5k1	6k8
1.34	02.53	6k8	9k1
1.34	02.54	5k6	7k5
1.34	02.55	8k2	11k
1.35	02.61	2k	2k7
1.36	02.69	1k1	1k5
1.36	02.69	2k2	3k
1.38	02.77	1k6	2k2
1.38	02.77	2k4	3k3
1.38	02.83	1k3	1k8

A_V	dB	R1	R2
1.42	03.03	3k6	5k1
1.42	03.07	3k3	4k7
1.43	03.10	9k1	13k
1.43	03.13	3k	4k3
1.44	03.14	3k9	5k6
1.44	03.18	4k3	6k2
1.44	03.19	2k7	3k9
1.45	03.21	4k7	6k8
1.45	03.25	1k1	1k6
1.46	03.31	8k2	12k
1.46	03.31	5k6	8k2
1.47	03.33	1k5	2k2
1.47	03.33	7k5	11k
1.47	03.33	6k2	9k1
1.47	03.35	5k1	7k5
1.47	03.35	6k8	10k
1.50	03.52	1k	1k5
1.50	03.52	1k2	1k8
1.50	03.52	1k6	2k4
1.50	03.52	1k8	2k7
1.50	03.52	2k	3k
1.50	03.52	2k2	3k3
1.50	03.52	2k4	3k6
1.54	03.74	1k3	2k
1.55	03.78	3k3	5k1
1.56	03.84	3k6	5k6
1.57	03.90	3k	4k7
1.58	03.98	4k3	6k8
1.59	04.00	8k2	13k
1.59	04.03	3k9	6k2
1.59	04.04	2k7	4k3
1.60	04.06	4k7	7k5
1.60	04.08	1k	1k6
1.60	04.08	1k5	2k4
1.60	04.08	7k5	12k
1.61	04.12	5k1	8k2
1.61	04.15	6k2	10k

A_V	dB	R1	R2
1.62	04.18	6k8	11k
1.63	04.22	2k4	3k9
1.63	04.22	5k6	9k1
1.64	04.28	1k1	1k8
1.64	04.28	2k2	3k6
1.65	04.34	9k1	15k
1.65	04.35	2k	3k3
1.67	04.44	1k2	2k
1.67	04.44	1k8	3k
1.69	04.54	1k6	2k7
1.69	04.57	1k3	2k2
1.70	04.59	3k3	5k6
1.70	04.61	3k	5k1
1.72	04.72	3k6	6k2
1.73	04.78	7k5	13k
1.74	04.81	2k7	4k7
1.74	04.83	3k9	6k8
1.74	04.83	4k3	7k5
1.74	04.83	4k7	8k2
1.76	04.90	9k1	16k
1.76	04.93	6k8	12k
1.77	04.97	2k2	3k9
1.77	04.98	6k2	11k
1.78	05.03	5k1	9k1
1.79	05.04	5k6	10k
1.79	05.07	2k4	4k3
1.80	05.11	1k	1k8
1.80	05.11	1k5	2k7
1.80	05.11	2k	3k6
1.82	05.19	1k1	2k
1.83	05.25	8k2	15k
1.83	05.26	1k2	2k2
1.83	05.26	1k8	3k3
1.85	05.33	1k3	2k4
1.87	05.42	3k	5k6
1.88	05.46	1k6	3k
1.88	05.48	3k3	6k2

A_V	dB	R1	R2
1.89	05.52	2k7	5k1
1.89	05.52	3k6	6k8
1.91	05.61	4k3	8k2
1.91	05.63	6k8	13k
1.92	05.68	3k9	7k5
1.94	05.74	6k2	12k
1.94	05.74	4k7	9k1
1.95	05.80	2k	3k9
1.95	05.81	8k2	16k
1.95	05.82	2k2	4k3
1.96	05.84	2k4	4k7
1.96	05.85	5k1	10k
1.96	05.86	5k6	11k
1.98	05.92	9k1	18k
2.00	06.02	1k	2k
2.00	06.02	1k1	2k2
2.00	06.02	1k2	2k4
2.00	06.02	1k5	3k
2.00	06.02	1k8	3k6
2.00	06.02	7k5	15k
2.06	06.28	3k3	6k8
2.06	06.29	1k6	3k3
2.07	06.31	3k	6k2
2.07	06.34	2k7	5k6
2.08	06.35	1k3	2k7
2.08	06.38	3k6	7k5
2.10	06.43	6k2	13k
2.10	06.45	3k9	8k2
2.12	06.51	4k3	9k1
2.13	06.55	2k4	5k1
2.13	06.56	4k7	10k
2.13	06.58	7k5	16k
2.14	06.59	2k2	4k7
2.14	06.62	5k6	12k
2.15	06.65	2k	4k3
2.16	06.68	5k1	11k
2.17	06.72	1k8	3k9
2.18	06.78	1k1	2k4
2.20	06.83	8k2	18k
2.20	06.84	9k1	20k
2.20	06.85	1k	2k2
2.20	06.85	1k5	3k3
2.21	06.87	6k8	15k
2.25	07.04	1k2	2k7
2.25	07.04	1k6	3k6
2.27	07.11	3k	6k8
2.27	07.13	3k3	7k5
2.28	07.15	3k6	8k2
2.30	07.22	2k7	6k2
2.31	07.26	1k3	3k
2.32	07.30	2k2	5k1
2.32	07.32	5k6	13k
2.33	07.33	4k3	10k
2.33	07.36	2k4	5k6
2.33	07.36	3k9	9k1
2.34	07.39	4k7	11k
2.35	07.42	2k	4k7
2.35	07.43	5k1	12k
2.35	07.43	6k8	16k
2.39	07.56	1k8	4k3
2.40	07.60	1k	2k4
2.40	07.60	7k5	18k
2.40	07.60	1k5	3k6
2.42	07.67	9k1	22k
2.42	07.67	6k2	15k
2.44	07.74	1k6	3k9
2.44	07.74	8k2	20k
2.45	07.80	1k1	2k7
2.48	07.91	3k3	8k2
2.50	07.96	1k2	3k
2.50	07.96	3k	7k5
2.52	08.02	2k7	6k8
2.53	08.05	3k6	9k1
2.54	08.09	1k3	3k3

A_V	dB	R1	R2
2.55	08.12	2k2	5k6
2.55	08.13	5k1	13k
2.55	08.13	2k	5k1
2.55	08.14	4k7	12k
2.56	08.16	4k3	11k
2.56	08.18	3k9	10k
2.58	08.23	6k2	16k
2.58	08.24	2k4	6k2
2.60	08.30	1k5	3k9
2.61	08.34	1k8	4k7
2.64	08.42	9k1	24k
2.65	08.46	6k8	18k
2.67	08.52	7k5	20k
2.68	08.56	5k6	15k
2.68	08.57	8k2	22k
2.69	08.59	1k6	4k3
2.70	08.63	1k	2k7
2.73	08.71	1k1	3k
2.73	08.73	3k	8k2
2.75	08.79	1k2	3k3
2.76	08.81	3k3	9k1
2.77	08.84	4k7	13k
2.77	08.85	1k3	3k6
2.78	08.87	2k7	7k5
2.78	08.87	3k6	10k
2.79	08.91	4k3	12k
2.80	08.94	2k	5k6
2.82	09.00	2k2	6k2
2.82	09.01	3k9	11k
2.83	09.05	1k8	5k1
2.83	09.05	2k4	6k8
2.86	09.12	5k6	16k
2.87	09.15	1k5	4k3
2.90	09.26	6k2	18k
2.93	09.33	8k2	24k
2.93	09.35	7k5	22k
2.94	09.36	1k6	4k7
2.94	09.37	5k1	15k
2.94	09.37	6k8	20k
2.97	09.45	9k1	27k
3.00	09.54	1k	3k
3.00	09.54	1k1	3k3
3.00	09.54	1k2	3k6
3.00	09.54	1k3	3k9
3.02	09.61	4k3	13k
3.03	09.63	3k3	10k
3.03	09.64	3k	9k1
3.04	09.65	2k7	8k2
3.06	09.70	3k6	11k
3.08	09.76	3k9	12k
3.09	09.80	2k2	6k8
3.10	09.83	2k	6k2
3.11	09.86	1k8	5k6
3.13	09.90	2k4	7k5
3.13	09.92	1k5	4k7
3.14	09.93	5k1	16k
3.19	10.07	1k6	5k1
3.19	10.08	4k7	15k
3.20	10.10	7k5	24k
3.21	10.14	5k6	18k
3.23	10.17	6k2	20k
3.24	10.20	6k8	22k
3.25	10.24	1k2	3k9
3.27	10.30	1k1	3k6
3.29	10.35	8k2	27k
3.30	10.36	9k1	30k
3.30	10.37	1k	3k3
3.31	10.39	1k3	4k3
3.33	10.46	3k	10k
3.33	10.46	3k3	11k
3.33	10.46	3k6	12k
3.33	10.46	3k9	13k
3.37	10.55	2k7	9k1
3.40	10.63	1k5	5k1

A_V	dB	R1	R2
3.40	10.63	2k	6k8
3.40	10.64	4k7	16k
3.41	10.65	2k2	7k5
3.42	10.67	2k4	8k2
3.44	10.74	1k8	6k2
3.49	10.85	4k3	15k
3.50	10.88	1k6	5k6
3.53	10.95	5k1	18k
3.53	10.95	6k8	24k
3.55	10.99	1k1	3k9
3.55	11.00	6k2	22k
3.57	11.06	5k6	20k
3.58	11.09	1k2	4k3
3.60	11.13	1k	3k6
3.60	11.13	7k5	27k
3.61	11.15	3k6	13k
3.62	11.16	1k3	4k7
3.63	11.19	9k1	33k
3.64	11.21	3k3	12k
3.66	11.27	8k2	30k
3.67	11.29	3k	11k
3.70	11.37	2k7	10k
3.72	11.41	4k3	16k
3.73	11.43	2k2	8k2
3.73	11.44	1k5	5k6
3.75	11.48	2k	7k5
3.78	11.54	1k8	6k8
3.79	11.58	2k4	9k1
3.83	11.66	4k7	18k
3.85	11.70	3k9	15k
3.87	11.76	6k2	24k
3.88	11.77	1k6	6k2
3.90	11.82	1k	3k9
3.91	11.84	1k1	4k3
3.92	11.86	1k2	4k7
3.92	11.87	5k1	20k
3.92	11.87	1k3	5k1

A_V	dB	R1	R2
3.93	11.88	5k6	22k
3.94	11.91	3k3	13k
3.96	11.95	9k1	36k
3.97	11.98	6k8	27k
4.00	12.04	3k	12k
4.00	12.04	7k5	30k
4.02	12.09	8k2	33k
4.07	12.20	2k7	11k
4.10	12.26	2k	8k2
4.10	12.26	3k9	16k
4.13	12.33	1k5	6k2
4.14	12.33	2k2	9k1
4.17	12.40	1k8	7k5
4.17	12.40	2k4	10k
4.17	12.40	3k6	15k
4.19	12.44	4k3	18k
4.25	12.57	1k2	5k1
4.25	12.57	1k6	6k8
4.26	12.58	4k7	20k
4.27	12.61	1k1	4k7
4.29	12.64	5k6	24k
4.29	12.64	9k1	39k
4.30	12.67	1k	4k3
4.31	12.68	1k3	5k6
4.31	12.70	5k1	22k
4.33	12.74	3k	13k
4.35	12.78	6k2	27k
4.39	12.85	8k2	36k
4.40	12.87	7k5	33k
4.41	12.89	6k8	30k
4.44	12.96	2k7	12k
4.44	12.96	3k6	16k
4.53	13.13	1k5	6k8
4.55	13.15	2k2	10k
4.55	13.15	3k3	15k
4.55	13.16	2k	9k1
4.56	13.17	1k8	8k2

A_V	dB	R1	R2
4.58	13.22	2k4	11k
4.62	13.28	3k9	18k
4.64	13.32	1k1	5k1
4.65	13.35	4k3	20k
4.67	13.38	1k2	5k6
4.68	13.41	4k7	22k
4.69	13.42	1k6	7k5
4.70	13.44	1k	4k7
4.71	13.45	5k1	24k
4.73	13.49	9k1	43k
4.76	13.55	8k2	39k
4.77	13.57	1k3	6k2
4.80	13.62	7k5	36k
4.81	13.65	2k7	13k
4.82	13.66	5k6	27k
4.84	13.69	6k2	30k
4.85	13.71	3k3	16k
4.85	13.72	6k8	33k
5.00	13.98	1k5	7k5
5.00	13.98	2k	10k
5.00	13.98	2k2	11k
5.00	13.98	2k4	12k
5.00	13.98	3k	15k
5.00	13.98	3k6	18k
5.06	14.08	1k8	9k1
5.09	14.14	1k1	5k6
5.10	14.15	1k	5k1
5.11	14.16	4k7	24k
5.12	14.18	4k3	22k
5.13	14.19	1k6	8k2
5.13	14.20	3k9	20k
5.16	14.26	9k1	47k
5.17	14.26	1k2	6k2
5.20	14.32	7k5	39k
5.23	14.37	1k3	6k8
5.24	14.39	8k2	43k
5.29	14.48	5k1	27k
5.29	14.48	6k8	36k
5.32	14.52	6k2	33k
5.33	14.54	3k	16k
5.36	14.58	5k6	30k
5.42	14.67	2k4	13k
5.45	14.74	3k3	18k
5.45	14.74	2k2	12k
5.47	14.75	1k5	8k2
5.50	14.81	2k	11k
5.56	14.89	1k8	10k
5.56	14.89	2k7	15k
5.56	14.89	3k6	20k
5.58	14.93	4k3	24k
5.60	14.96	1k	5k6
5.60	14.97	9k1	51k
5.64	15.02	1k1	6k2
5.64	15.03	3k9	22k
5.67	15.07	1k2	6k8
5.69	15.10	1k6	9k1
5.73	15.17	8k2	47k
5.73	15.17	7k5	43k
5.74	15.17	6k8	39k
5.74	15.19	4k7	27k
5.77	15.22	1k3	7k5
5.81	15.28	6k2	36k
5.88	15.39	5k1	30k
5.89	15.41	5k6	33k
5.91	15.43	2k2	13k
5.93	15.46	2k7	16k
6.00	15.56	2k	12k
6.00	15.56	3k	18k
6.06	15.65	3k3	20k
6.07	15.66	1k5	9k1
6.11	15.72	1k8	11k
6.11	15.72	3k6	22k
6.15	15.78	3k9	24k
6.15	15.78	9k1	56k

A_V	dB	R1	R2
6.18	15.82	1k1	6k8
6.20	15.85	1k	6k2
6.22	15.88	8k2	51k
6.25	15.92	1k2	7k5
6.25	15.92	1k6	10k
6.25	15.92	2k4	15k
6.27	15.94	7k5	47k
6.28	15.96	4k3	27k
6.29	15.97	6k2	39k
6.31	16.00	1k3	8k2
6.32	16.02	6k8	43k
6.38	16.10	4k7	30k
6.43	16.16	5k6	36k
6.47	16.22	5k1	33k
6.50	16.26	2k	13k
6.67	16.48	1k5	10k
6.67	16.48	1k8	12k
6.67	16.48	2k4	16k
6.67	16.48	2k7	18k
6.67	16.48	3k	20k
6.67	16.48	3k3	22k
6.67	16.48	3k6	24k
6.80	16.65	1k	6k8
6.80	16.65	7k5	51k
6.81	16.67	9k1	62k
6.82	16.67	1k1	7k5
6.82	16.67	2k2	15k
6.83	16.69	8k2	56k
6.83	16.69	1k2	8k2
6.88	16.75	1k6	11k
6.91	16.79	6k8	47k
6.92	16.81	3k9	27k
6.94	16.82	6k2	43k
6.96	16.86	5k6	39k
6.98	16.87	4k3	30k
7.00	16.90	1k3	9k1
7.02	16.93	4k7	33k

A_V	dB	R1	R2
7.06	16.97	5k1	36k
7.22	17.17	1k8	13k
7.27	17.23	3k3	24k
7.27	17.23	2k2	16k
7.33	17.31	1k5	11k
7.33	17.31	3k	22k
7.41	17.39	2k7	20k
7.45	17.45	1k1	8k2
7.47	17.46	7k5	56k
7.47	17.47	9k1	68k
7.50	17.50	1k	7k5
7.50	17.50	1k6	12k
7.50	17.50	2k	15k
7.50	17.50	2k4	18k
7.50	17.50	3k6	27k
7.50	17.50	6k8	51k
7.56	17.57	8k2	62k
7.58	17.59	6k2	47k
7.58	17.60	1k2	9k1
7.65	17.67	5k1	39k
7.66	17.68	4k7	36k
7.67	17.70	4k3	33k
7.68	17.71	5k6	43k
7.69	17.72	1k3	10k
7.69	17.72	3k9	30k
8.00	18.06	1k5	12k
8.00	18.06	2k	16k
8.00	18.06	3k	24k
8.13	18.20	1k6	13k
8.15	18.22	2k7	22k
8.18	18.26	2k2	18k
8.18	18.26	3k3	27k
8.20	18.28	1k	8k2
8.23	18.30	6k2	51k
8.24	18.31	6k8	56k
8.24	18.32	9k1	75k
8.27	18.35	7k5	62k

A_V	dB	R1	R2
8.27	18.35	1k1	9k1
8.29	18.37	8k2	68k
8.30	18.38	4k7	39k
8.33	18.42	1k2	10k
8.33	18.42	1k8	15k
8.33	18.42	2k4	20k
8.33	18.42	3k6	30k
8.37	18.46	4k3	36k
8.39	18.48	5k6	47k
8.43	18.52	5k1	43k
8.46	18.55	1k3	11k
8.46	18.55	3k9	33k
8.67	18.76	1k5	13k
8.89	18.98	1k8	16k
8.89	18.98	2k7	24k
9.00	19.08	2k	18k
9.00	19.08	3k	27k
9.01	19.10	9k1	82k
9.03	19.12	6k2	56k
9.07	19.15	7k5	68k
9.07	19.15	4k3	39k
9.09	19.17	1k1	10k
9.09	19.17	2k2	20k
9.09	19.17	3k3	30k
9.10	19.18	1k	9k1
9.11	19.19	5k6	51k
9.12	19.20	6k8	62k
9.15	19.22	8k2	75k
9.15	19.23	4k7	43k
9.17	19.24	3k6	33k
9.17	19.24	1k2	11k
9.17	19.24	2k4	22k
9.22	19.29	5k1	47k
9.23	19.30	1k3	12k
9.23	19.30	3k9	36k
9.38	19.44	1k6	15k
10.0	20.00	1k	10k
10.0	20.00	1k1	11k
10.0	20.00	1k2	12k
10.0	20.00	1k3	13k
10.0	20.00	1k5	15k
10.0	20.00	1k6	16k
10.0	20.00	1k8	18k
10.0	20.00	2k	20k
10.0	20.00	2k2	22k
10.0	20.00	2k4	24k
10.0	20.00	2k7	27k
10.0	20.00	3k	30k
10.0	20.00	3k3	33k
10.0	20.00	3k6	36k
10.0	20.00	3k9	39k
10.0	20.00	4k3	43k
10.0	20.00	4k7	47k
10.0	20.00	5k1	51k
10.0	20.00	5k6	56k
10.0	20.00	6k2	62k
10.0	20.00	6k8	68k
10.0	20.00	7k5	75k
10.0	20.00	8k2	82k
10.0	20.00	9k1	91k
10.7	20.56	1k5	16k
10.8	20.70	1k2	13k
10.8	20.70	3k6	39k
10.9	20.71	4k7	51k
10.9	20.76	3k3	36k
10.9	20.76	1k1	12k
10.9	20.76	2k2	24k
10.9	20.77	4k3	47k
10.9	20.78	7k5	82k
11.0	20.80	6k2	68k
11.0	20.81	5k1	56k
11.0	20.82	9k1	100k
11.0	20.82	910R	10k
11.0	20.83	1k	11k

A_V	dB	R1	R2
11.0	20.83	2k	22k
11.0	20.83	3k	33k
11.0	20.85	3k9	43k
11.0	20.85	6k8	75k
11.1	20.88	5k6	62k
11.1	20.90	8k2	91k
11.1	20.92	1k8	20k
11.1	20.92	2k7	30k
11.3	21.02	1k6	18k
11.3	21.02	2k4	27k
11.5	21.24	1k3	15k
11.8	21.45	1k1	13k
11.8	21.45	3k3	39k
11.9	21.48	4k3	51k
11.9	21.52	4k7	56k
11.9	21.54	3k6	43k
12.0	21.58	1k	12k
12.0	21.58	1k5	18k
12.0	21.58	2k	24k
12.0	21.58	3k	36k
12.1	21.62	3k9	47k
12.1	21.63	6k8	82k
12.1	21.65	910R	11k
12.1	21.65	6k2	75k
12.1	21.68	7k5	91k
12.1	21.69	5k6	68k
12.2	21.70	5k1	62k
12.2	21.72	8k2	100k
12.2	21.72	820R	10k
12.2	21.74	1k8	22k
12.2	21.74	2k7	33k
12.3	21.78	2k2	27k
12.3	21.80	1k3	16k
12.5	21.94	1k2	15k
12.5	21.94	1k6	20k
12.5	21.94	2k4	30k
13.0	22.28	1k	13k
13.0	22.28	3k	39k
13.0	22.29	4k3	56k
13.0	22.30	3k3	43k
13.1	22.32	3k6	47k
13.1	22.33	3k9	51k
13.2	22.40	910R	12k
13.2	22.41	4k7	62k
13.2	22.43	6k2	82k
13.3	22.50	1k2	16k
13.3	22.50	1k5	20k
13.3	22.50	1k8	24k
13.3	22.50	2k7	36k
13.3	22.50	5k1	68k
13.3	22.50	7k5	100k
13.3	22.50	750R	10k
13.4	22.53	6k8	91k
13.4	22.54	5k6	75k
13.4	22.55	820R	11k
13.5	22.61	2k	27k
13.6	22.69	1k1	15k
13.6	22.69	2k2	30k
13.8	22.77	1k6	22k
13.8	22.77	2k4	33k
13.8	22.83	1k3	18k
14.2	23.03	3k6	51k
14.2	23.07	3k3	47k
14.3	23.10	910R	13k
14.3	23.13	3k	43k
14.4	23.14	3k9	56k
14.4	23.18	4k3	62k
14.4	23.19	2k7	39k
14.5	23.21	4k7	68k
14.5	23.25	1k1	16k
14.6	23.31	820R	12k
14.6	23.31	5k6	82k
14.7	23.33	1k5	22k
14.7	23.33	750R	11k

A_V	dB	R1	R2
14.7	23.33	6k2	91k
14.7	23.35	5k1	75k
14.7	23.35	6k8	100k
14.7	23.35	680R	10k
15.0	23.52	1k	15k
15.0	23.52	1k2	18k
15.0	23.52	1k6	24k
15.0	23.52	1k8	27k
15.0	23.52	2k	30k
15.0	23.52	2k2	33k
15.0	23.52	2k4	36k
15.4	23.74	1k3	20k
15.5	23.78	3k3	51k
15.6	23.84	3k6	56k
15.7	23.90	3k	47k
15.8	23.98	4k3	68k
15.9	24.00	820R	13k
15.9	24.03	3k9	62k
15.9	24.04	2k7	43k
16.0	24.06	4k7	75k
16.0	24.08	1k	16k
16.0	24.08	1k5	24k
16.0	24.08	750R	12k
16.1	24.12	5k1	82k
16.1	24.15	6k2	100k
16.1	24.15	620R	10k
16.2	24.18	680R	11k
16.3	24.22	2k4	39k
16.3	24.22	5k6	91k
16.4	24.28	1k1	18k
16.4	24.28	2k2	36k
16.5	24.34	910R	15k
16.5	24.35	2k	33k
16.7	24.44	1k2	20k
16.7	24.44	1k8	30k
16.9	24.54	1k6	27k
16.9	24.57	1k3	22k

A_V	dB	R1	R2
17.0	24.59	3k3	56k
17.0	24.61	3k	51k
17.2	24.72	3k6	62k
17.3	24.78	750R	13k
17.4	24.81	2k7	47k
17.4	24.83	3k9	68k
17.4	24.83	4k3	75k
17.4	24.83	4k7	82k
17.6	24.90	910R	16k
17.6	24.93	680R	12k
17.7	24.97	2k2	39k
17.7	24.98	620R	11k
17.8	25.03	5k1	91k
17.9	25.04	5k6	100k
17.9	25.04	560R	10k
17.9	25.07	2k4	43k
18.0	25.11	1k	18k
18.0	25.11	1k5	27k
18.0	25.11	2k	36k
18.2	25.19	1k1	20k
18.3	25.25	820R	15k
18.3	25.26	1k8	33k
18.3	25.26	1k2	22k
18.5	25.33	1k3	24k
18.7	25.42	3k	56k
18.8	25.46	1k6	30k
18.8	25.48	3k3	62k
18.9	25.52	3k6	68k
18.9	25.52	2k7	51k
19.1	25.61	4k3	82k
19.1	25.63	680R	13k
19.2	25.68	3k9	75k
19.4	25.74	620R	12k
19.4	25.74	4k7	91k
19.5	25.80	2k	39k
19.5	25.81	820R	16k
19.5	25.82	2k2	43k

A_V	dB	R1	R2
19.6	25.84	2k4	47k
19.6	25.85	510R	10k
19.6	25.85	5k1	100k
19.6	25.86	560R	11k
19.8	25.92	910R	18k
20.0	26.02	1k	20k
20.0	26.02	1k1	22k
20.0	26.02	1k2	24k
20.0	26.02	1k5	30k
20.0	26.02	1k8	36k
20.0	26.02	750R	15k
20.6	26.28	3k3	68k
20.6	26.29	1k6	33k
20.7	26.31	3k	62k
20.7	26.34	2k7	56k
20.8	26.35	1k3	27k
20.8	26.38	3k6	75k
21.0	26.43	620R	13k
21.0	26.45	3k9	82k
21.2	26.51	4k3	91k
21.3	26.55	2k4	51k
21.3	26.56	470R	10k
21.3	26.56	4k7	100k
21.3	26.58	750R	16k
21.4	26.59	2k2	47k
21.4	26.62	560R	12k
21.5	26.65	2k	43k
21.6	26.68	510R	11k
21.7	26.72	1k8	39k
21.8	26.78	1k1	24k
22.0	26.83	820R	18k
22.0	26.84	910R	20k
22.0	26.85	1k	22k
22.0	26.85	1k5	33k
22.1	26.87	680R	15k
22.5	27.04	1k2	27k
22.5	27.04	1k6	36k
22.7	27.11	3k	68k
22.7	27.13	3k3	75k
22.8	27.15	3k6	82k
23.0	27.22	2k7	62k
23.1	27.26	1k3	30k
23.2	27.30	2k2	51k
23.2	27.32	560R	13k
23.3	27.33	430R	10k
23.3	27.33	4k3	100k
23.3	27.36	3k9	91k
23.3	27.36	2k4	56k
23.4	27.39	470R	11k
23.5	27.42	2k	47k
23.5	27.43	510R	12k
23.5	27.43	680R	16k
23.9	27.56	1k8	43k
24.0	27.60	1k	24k
24.0	27.60	1k5	36k
24.0	27.60	750R	18k
24.2	27.67	910R	22k
24.2	27.67	620R	15k
24.4	27.74	1k6	39k
24.4	27.74	820R	20k
24.5	27.80	1k1	27k
24.8	27.91	3k3	82k
25.0	27.96	1k2	30k
25.0	27.96	3k	75k
25.2	28.02	2k7	68k
25.3	28.05	3k6	91k
25.4	28.09	1k3	33k
25.5	28.12	2k2	56k
25.5	28.13	510R	13k
25.5	28.13	2k	51k
25.5	28.14	470R	12k
25.6	28.16	430R	11k
25.6	28.18	3k9	100k
25.6	28.18	390R	10k

A_V	dB	R1	R2
25.8	28.23	620R	16k
25.8	28.24	2k4	62k
26.0	28.30	1k5	39k
26.1	28.34	1k8	47k
26.4	28.42	910R	24k
26.5	28.46	680R	18k
26.7	28.52	750R	20k
26.8	28.56	560R	15k
26.8	28.57	820R	22k
26.9	28.59	1k6	43k
27.0	28.63	1k	27k
27.3	28.71	1k1	30k
27.3	28.73	3k	82k
27.5	28.79	1k2	33k
27.6	28.81	3k3	91k
27.7	28.84	470R	13k
27.7	28.85	1k3	36k
27.8	28.87	2k7	75k
27.8	28.87	3k6	100k
27.8	28.87	360R	10k
27.9	28.91	430R	12k
28.0	28.94	2k	56k
28.2	29.00	2k2	62k
28.2	29.01	390R	11k
28.3	29.05	1k8	51k
28.3	29.05	2k4	68k
28.6	29.12	560R	16k
28.7	29.15	1k5	43k
29.0	29.26	620R	18k
29.3	29.33	820R	24k
29.3	29.35	750R	22k
29.4	29.36	1k6	47k
29.4	29.37	510R	15k
29.4	29.37	680R	20k
29.7	29.45	910R	27k
30.0	29.54	1k	30k
30.0	29.54	1k1	33k
30.0	29.54	1k2	36k
30.0	29.54	1k3	39k
30.2	29.61	430R	13k
30.3	29.63	3k3	100k
30.3	29.63	330R	10k
30.3	29.64	3k	91k
30.4	29.65	2k7	82k
30.6	29.70	360R	11k
30.8	29.76	390R	12k
30.9	29.80	2k2	68k
31.0	29.83	2k	62k
31.1	29.86	1k8	56k
31.3	29.90	2k4	75k
31.3	29.92	1k5	47k
31.4	29.93	510R	16k
31.9	30.07	1k6	51k
31.9	30.08	470R	15k
32.0	30.10	750R	24k
32.1	30.14	560R	18k
32.3	30.17	620R	20k
32.4	30.20	680R	22k
32.5	30.24	1k2	39k
32.7	30.30	1k1	36k
32.9	30.35	820R	27k
33.0	30.36	910R	30k
33.0	30.37	1k	33k
33.1	30.39	1k3	43k
33.3	30.46	3k	100k
33.3	30.46	300R	10k
33.3	30.46	330R	11k
33.3	30.46	360R	12k
33.3	30.46	390R	13k
33.7	30.55	2k7	91k
34.0	30.63	1k5	51k
34.0	30.63	2k	68k
34.0	30.64	470R	16k
34.1	30.65	2k2	75k

A_V	dB	R1	R2
34.2	30.67	2k4	82k
34.4	30.74	1k8	62k
34.9	30.85	430R	15k
35.0	30.88	1k6	56k
35.3	30.95	510R	18k
35.3	30.95	680R	24k
35.5	30.99	1k1	39k
35.5	31.00	620R	22k
35.7	31.06	560R	20k
35.8	31.09	1k2	43k
36.0	31.13	1k	36k
36.0	31.13	750R	27k
36.1	31.15	360R	13k
36.2	31.16	1k3	47k
36.3	31.19	910R	33k
36.4	31.21	330R	12k
36.6	31.27	820R	30k
36.7	31.29	300R	11k
37.0	31.37	2k7	100k
37.0	31.37	270R	10k
37.2	31.41	430R	16k
37.3	31.43	2k2	82k
37.3	31.44	1k5	56k
37.5	31.48	2k	75k
37.8	31.54	1k8	68k
37.9	31.58	2k4	91k
38.3	31.66	470R	18k
38.5	31.70	390R	15k
38.7	31.76	620R	24k
38.8	31.77	1k6	62k
39.0	31.82	1k	39k
39.1	31.84	1k1	43k
39.2	31.86	1k2	47k
39.2	31.87	510R	20k
39.2	31.87	1k3	51k
39.3	31.88	560R	22k
39.4	31.91	330R	13k

A_V	dB	R1	R2
39.6	31.95	910R	36k
39.7	31.98	680R	27k
40.0	32.04	300R	12k
40.0	32.04	750R	30k
40.2	32.09	820R	33k
40.7	32.20	270R	11k
41.0	32.26	2k	82k
41.0	32.26	390R	16k
41.3	32.33	1k5	62k
41.4	32.33	2k2	91k
41.7	32.40	1k8	75k
41.7	32.40	240R	10k
41.7	32.40	2k4	100k
41.7	32.40	360R	15k
41.9	32.44	430R	18k
42.5	32.57	1k2	51k
42.5	32.57	1k6	68k
42.6	32.58	470R	20k
42.7	32.61	1k1	47k
42.9	32.64	910R	39k
42.9	32.64	560R	24k
43.0	32.67	1k	43k
43.1	32.68	1k3	56k
43.1	32.70	510R	22k
43.3	32.74	300R	13k
43.5	32.78	620R	27k
43.9	32.85	820R	36k
44.0	32.87	750R	33k
44.1	32.89	680R	30k
44.4	32.96	270R	12k
44.4	32.96	360R	16k
45.3	33.13	1k5	68k
45.5	33.15	220R	10k
45.5	33.15	2k2	100k
45.5	33.15	330R	15k
45.5	33.16	2k	91k
45.6	33.17	1k8	82k

A_V	dB	R1	R2
45.8	33.22	240R	11k
46.2	33.28	390R	18k
46.4	33.32	1k1	51k
46.5	33.35	430R	20k
46.7	33.38	1k2	56k
46.8	33.41	470R	22k
46.9	33.42	1k6	75k
47.0	33.44	1k	47k
47.1	33.45	510R	24k
47.3	33.49	910R	43k
47.6	33.55	820R	39k
47.7	33.57	1k3	62k
48.0	33.62	750R	36k
48.1	33.65	270R	13k
48.2	33.66	560R	27k
48.4	33.69	620R	30k
48.5	33.71	330R	16k
48.5	33.72	680R	33k
50.0	33.98	1k5	75k
50.0	33.98	2k	100k
50.0	33.98	200R	10k
50.0	33.98	220R	11k
50.0	33.98	240R	12k
50.0	33.98	300R	15k
50.0	33.98	360R	18k
50.6	34.08	1k8	91k
50.9	34.14	1k1	56k
51.0	34.15	1k	51k
51.1	34.16	470R	24k
51.2	34.18	430R	22k
51.3	34.19	1k6	82k
51.3	34.20	390R	20k
51.6	34.26	910R	47k
51.7	34.26	1k2	62k
52.0	34.32	750R	39k
52.3	34.37	1k3	68k
52.4	34.39	820R	43k
52.9	34.48	510R	27k
52.9	34.48	680R	36k
53.2	34.52	620R	33k
53.3	34.54	300R	16k
53.6	34.58	560R	30k
54.2	34.67	240R	13k
54.5	34.74	220R	12k
54.5	34.74	330R	18k
54.7	34.75	1k5	82k
55.0	34.81	200R	11k
55.6	34.89	1k8	100k
55.6	34.89	180R	10k
55.6	34.89	360R	20k
55.6	34.89	270R	15k
55.8	34.93	430R	24k
56.0	34.96	1k	56k
56.0	34.97	910R	51k
56.4	35.02	1k1	62k
56.4	35.03	390R	22k
56.7	35.07	1k2	68k
56.9	35.10	1k6	91k
57.3	35.17	820R	47k
57.3	35.17	750R	43k
57.4	35.17	680R	39k
57.4	35.19	470R	27k
57.7	35.22	1k3	75k
58.1	35.28	620R	36k
58.8	35.39	510R	30k
58.9	35.41	560R	33k
59.1	35.43	220R	13k
59.3	35.46	270R	16k
60.0	35.56	200R	12k
60.0	35.56	300R	18k
60.6	35.65	330R	20k
60.7	35.66	1k5	91k
61.1	35.72	180R	11k
61.1	35.72	360R	22k

A_V	dB	R1	R2
61.5	35.78	390R	24k
61.5	35.78	910R	56k
61.8	35.82	1k1	68k
62.0	35.85	1k	62k
62.2	35.88	820R	51k
62.5	35.92	1k2	75k
62.5	35.92	1k6	100k
62.5	35.92	160R	10k
62.5	35.92	240R	15k
62.7	35.94	750R	47k
62.8	35.96	430R	27k
62.9	35.97	620R	39k
63.1	36.00	1k3	82k
63.2	36.02	680R	43k
63.8	36.10	470R	30k
64.3	36.16	560R	36k
64.7	36.22	510R	33k
65.0	36.26	200R	13k
66.7	36.48	1k5	100k
66.7	36.48	150R	10k
66.7	36.48	180R	12k
66.7	36.48	240R	16k
66.7	36.48	270R	18k
66.7	36.48	300R	20k
66.7	36.48	330R	22k
66.7	36.48	360R	24k
68.0	36.65	1k	68k
68.0	36.65	750R	51k
68.1	36.67	910R	62k
68.2	36.67	1k1	75k
68.2	36.67	220R	15k
68.3	36.69	820R	56k
68.3	36.69	1k2	82k
68.8	36.75	160R	11k
69.1	36.79	680R	47k
69.2	36.81	390R	27k
69.4	36.82	620R	43k

A_V	dB	R1	R2
69.6	36.86	560R	39k
69.8	36.87	430R	30k
70.0	36.90	1k3	91k
70.2	36.93	470R	33k
70.6	36.97	510R	36k
72.2	37.17	180R	13k
72.7	37.23	220R	16k
72.7	37.23	330R	24k
73.3	37.31	150R	11k
73.3	37.31	300R	22k
74.1	37.39	270R	20k
74.5	37.45	1k1	82k
74.7	37.46	750R	56k
74.7	37.47	910R	68k
75.0	37.50	1k	75k
75.0	37.50	160R	12k
75.0	37.50	200R	15k
75.0	37.50	240R	18k
75.0	37.50	360R	27k
75.0	37.50	680R	51k
75.6	37.57	820R	62k
75.8	37.59	620R	47k
75.8	37.60	1k2	91k
76.5	37.67	510R	39k
76.6	37.68	470R	36k
76.7	37.70	430R	33k
76.8	37.71	560R	43k
76.9	37.72	1k3	100k
76.9	37.72	130R	10k
76.9	37.72	390R	30k
80.0	38.06	150R	12k
80.0	38.06	200R	16k
80.0	38.06	300R	24k
81.3	38.20	160R	13k
81.5	38.22	270R	22k
81.8	38.26	220R	18k
81.8	38.26	330R	27k

A_V	dB	R1	R2
82.0	38.28	1k	82k
82.3	38.30	620R	51k
82.4	38.31	680R	56k
82.4	38.32	910R	75k
82.7	38.35	750R	62k
82.7	38.35	1k1	91k
82.9	38.37	820R	68k
83.0	38.38	470R	39k
83.3	38.42	120R	10k
83.3	38.42	240R	20k
83.3	38.42	1k2	100k
83.3	38.42	180R	15k
83.3	38.42	360R	30k
83.7	38.46	430R	36k
83.9	38.48	560R	47k
84.3	38.52	510R	43k
84.6	38.55	390R	33k
84.6	38.55	130R	11k
86.7	38.76	150R	13k
88.9	38.98	180R	16k
88.9	38.98	270R	24k
90.0	39.08	200R	18k
90.0	39.08	300R	27k
90.1	39.10	910R	82k
90.3	39.12	620R	56k
90.7	39.15	750R	68k
90.7	39.15	430R	39k
90.9	39.17	110R	10k
90.9	39.17	220R	20k
90.9	39.17	1k1	100k
90.9	39.17	330R	30k
91.0	39.18	1k	91k
91.1	39.19	560R	51k
91.2	39.20	680R	62k
91.5	39.22	820R	75k
91.5	39.23	470R	43k
91.7	39.24	120R	11k
91.7	39.24	240R	22k
91.7	39.24	360R	33k
92.2	39.29	510R	47k
92.3	39.30	390R	36k
92.3	39.30	130R	12k
93.8	39.44	160R	15k
100	40.00	1k	100k
100	40.00	100R	10k
100	40.00	110R	11k
100	40.00	120R	12k
100	40.00	130R	13k
100	40.00	150R	15k
100	40.00	160R	16k
100	40.00	180R	18k
100	40.00	200R	20k
100	40.00	220R	22k
100	40.00	240R	24k
100	40.00	270R	27k
100	40.00	300R	30k
100	40.00	330R	33k
100	40.00	360R	36k
100	40.00	390R	39k
100	40.00	430R	43k
100	40.00	470R	47k
100	40.00	510R	51k
100	40.00	560R	56k
100	40.00	620R	62k
100	40.00	680R	68k
100	40.00	750R	75k
100	40.00	820R	82k
100	40.00	910R	91k
107	40.56	150R	16k
108	40.70	120R	13k
108	40.70	360R	39k
109	40.71	470R	51k
109	40.76	110R	12k
109	40.76	220R	24k

A_V	dB	R1	R2
109	40.76	330R	36k
109	40.77	430R	47k
109	40.78	750R	82k
110	40.80	620R	68k
110	40.81	510R	56k
110	40.82	910R	100k
110	40.83	100R	11k
110	40.83	200R	22k
110	40.83	300R	33k
110	40.85	390R	43k
110	40.85	680R	75k
111	40.88	560R	62k
111	40.90	820R	91k
111	40.92	180R	20k
111	40.92	270R	30k
113	41.02	160R	18k
113	41.02	240R	27k
115	41.24	130R	15k
118	41.45	110R	13k
118	41.45	330R	39k
119	41.48	430R	51k
119	41.52	470R	56k
119	41.54	360R	43k
120	41.58	100R	12k
120	41.58	150R	18k
120	41.58	200R	24k
120	41.58	300R	36k
121	41.62	390R	47k
121	41.63	680R	82k
121	41.65	620R	75k
121	41.68	750R	91k
121	41.69	560R	68k
122	41.70	510R	62k
122	41.72	820R	100k
122	41.74	180R	22k
122	41.74	270R	33k
123	41.78	220R	27k

A_V	dB	R1	R2
123	41.80	130R	16k
125	41.94	120R	15k
125	41.94	160R	20k
125	41.94	240R	30k
130	42.28	100R	13k
130	42.28	300R	39k
130	42.29	430R	56k
130	42.30	330R	43k
131	42.32	360R	47k
131	42.33	390R	51k
132	42.41	470R	62k
132	42.43	620R	82k
133	42.50	120R	16k
133	42.50	150R	20k
133	42.50	180R	24k
133	42.50	270R	36k
133	42.50	510R	68k
133	42.50	750R	100k
134	42.53	680R	91k
134	42.54	560R	75k
135	42.61	200R	27k
136	42.69	110R	15k
136	42.69	220R	30k
138	42.77	160R	22k
138	42.77	240R	33k
138	42.83	130R	18k
142	43.03	360R	51k
142	43.07	330R	47k
143	43.13	300R	43k
144	43.14	390R	56k
144	43.18	430R	62k
144	43.19	270R	39k
145	43.21	470R	68k
145	43.25	110R	16k
146	43.31	560R	82k
147	43.33	150R	22k
147	43.33	620R	91k

A_V	dB	R1	R2
147	43.35	510R	75k
147	43.35	680R	100k
150	43.52	100R	15k
150	43.52	120R	18k
150	43.52	160R	24k
150	43.52	180R	27k
150	43.52	200R	30k
150	43.52	220R	33k
150	43.52	240R	36k
154	43.74	130R	20k
155	43.78	330R	51k
156	43.84	360R	56k
157	43.90	300R	47k
158	43.98	430R	68k
159	44.03	390R	62k
159	44.04	270R	43k
160	44.06	470R	75k
160	44.08	100R	16k
160	44.08	150R	24k
161	44.12	510R	82k
161	44.15	620R	100k
163	44.22	240R	39k
163	44.22	560R	91k
164	44.28	110R	18k
164	44.28	220R	36k
165	44.35	200R	33k
167	44.44	120R	20k
167	44.44	180R	30k
169	44.54	160R	27k
169	44.57	130R	22k
170	44.59	330R	56k
170	44.61	300R	51k
172	44.72	360R	62k
174	44.81	270R	47k
174	44.83	390R	68k
174	44.83	430R	75k
174	44.83	470R	82k
177	44.97	220R	39k
178	45.03	510R	91k
179	45.04	560R	100k
179	45.07	240R	43k
180	45.11	100R	18k
180	45.11	150R	27k
180	45.11	200R	36k
182	45.19	110R	20k
183	45.26	120R	22k
183	45.26	180R	33k
185	45.33	130R	24k
187	45.42	300R	56k
188	45.46	160R	30k
188	45.48	330R	62k
189	45.52	270R	51k
189	45.52	360R	68k
191	45.61	430R	82k
192	45.68	390R	75k
194	45.74	470R	91k
195	45.80	200R	39k
195	45.82	220R	43k
196	45.84	240R	47k
196	45.85	510R	100k
200	46.02	100R	20k
200	46.02	110R	22k
200	46.02	120R	24k
200	46.02	150R	30k
200	46.02	180R	36k
206	46.28	330R	68k
206	46.29	160R	33k
207	46.31	300R	62k
207	46.34	270R	56k
208	46.35	130R	27k
208	46.38	360R	75k
210	46.45	390R	82k
212	46.51	430R	91k
213	46.55	240R	51k

A_V	dB	R1	R2
213	46.56	470R	100k
214	46.59	220R	47k
215	46.65	200R	43k
217	46.72	180R	39k
218	46.78	110R	24k
220	46.85	100R	22k
220	46.85	150R	33k
225	47.04	120R	27k
225	47.04	160R	36k
227	47.11	300R	68k
227	47.13	330R	75k
228	47.15	360R	82k
230	47.22	270R	62k
231	47.26	130R	30k
232	47.30	220R	51k
233	47.33	430R	100k
233	47.36	240R	56k
233	47.36	390R	91k
235	47.42	200R	47k
239	47.56	180R	43k
240	47.60	100R	24k
240	47.60	150R	36k
244	47.74	160R	39k
245	47.80	110R	27k
248	47.91	330R	82k
250	47.96	120R	30k
250	47.96	300R	75k
252	48.02	270R	68k
253	48.05	360R	91k
254	48.09	130R	33k
255	48.12	220R	56k
255	48.13	200R	51k
256	48.18	390R	100k
258	48.24	240R	62k
260	48.30	150R	39k
261	48.34	180R	47k
269	48.59	160R	43k

A_V	dB	R1	R2
270	48.63	100R	27k
273	48.71	110R	30k
273	48.73	300R	82k
275	48.79	120R	33k
276	48.81	330R	91k
277	48.85	130R	36k
278	48.87	270R	75k
278	48.87	360R	100k
280	48.94	200R	56k
282	49.00	220R	62k
283	49.05	240R	68k
283	49.05	180R	51k
287	49.15	150R	43k
294	49.36	160R	47k
300	49.54	100R	30k
300	49.54	110R	33k
300	49.54	120R	36k
300	49.54	130R	39k
303	49.63	330R	100k
303	49.64	300R	91k
304	49.65	270R	82k
309	49.80	220R	68k
310	49.83	200R	62k
311	49.86	180R	56k
313	49.90	240R	75k
313	49.92	150R	47k
319	50.07	160R	51k
325	50.24	120R	39k
327	50.30	110R	36k
330	50.37	100R	33k
331	50.39	130R	43k
333	50.46	300R	100k
337	50.55	270R	91k
340	50.63	150R	51k
340	50.63	200R	68k
341	50.65	220R	75k
342	50.67	240R	82k

A_V	dB	R1	R2
344	50.74	180R	62k
350	50.88	160R	56k
355	50.99	110R	39k
358	51.09	120R	43k
360	51.13	100R	36k
362	51.16	130R	47k
370	51.37	270R	100k
373	51.43	220R	82k
373	51.44	150R	56k
375	51.48	200R	75k
378	51.54	180R	68k
379	51.58	240R	91k
388	51.77	160R	62k
390	51.82	100R	39k
391	51.84	110R	43k
392	51.86	120R	47k
392	51.87	130R	51k
410	52.26	200R	82k
413	52.33	150R	62k
414	52.33	220R	91k
417	52.40	180R	75k
417	52.40	240R	100k
425	52.57	120R	51k
425	52.57	160R	68k
427	52.61	110R	47k
430	52.67	100R	43k
431	52.68	130R	56k
453	53.13	150R	68k
455	53.15	220R	100k
455	53.16	200R	91k
456	53.17	180R	82k
464	53.32	110R	51k
467	53.38	120R	56k
469	53.42	160R	75k
470	53.44	100R	47k
477	53.57	130R	62k

A_V	dB	R1	R2
500	53.98	150R	75k
500	53.98	200R	100k
506	54.08	180R	91k
509	54.14	110R	56k
510	54.15	100R	51k
513	54.19	160R	82k
517	54.26	120R	62k
523	54.37	130R	68k
547	54.75	150R	82k
556	54.89	180R	100k
560	54.96	100R	56k
564	55.02	110R	62k
567	55.07	120R	68k
569	55.10	160R	91k
577	55.22	130R	75k
607	55.66	150R	91k
618	55.82	110R	68k
620	55.85	100R	62k
625	55.92	120R	75k
625	55.92	160R	100k
631	56.00	130R	82k
667	56.48	150R	100k
680	56.65	100R	68k
682	56.67	110R	75k
683	56.69	120R	82k
700	56.90	130R	91k
745	57.45	110R	82k
750	57.50	100R	75k
758	57.60	120R	91k
769	57.72	130R	100k
820	58.28	100R	82k
827	58.35	110R	91k
833	58.42	120R	100k
909	59.17	110R	100k
910	59.18	100R	91k
1.00k	60.00	100R	100k

This Page Intentionally Left Blank

Appendix 1

Units, symbols and suffixes

I have used standard units, symbols and suffixes throughout this work. Here they are:

Units
Ω or R: ohms (resistance)
F: farads (capacitance)
H: henries (inductance)
s: seconds (time)
Hz: hertz (frequency)
°: degrees (phase angle)
V: volts (electrical potential)
A: amperes (electrical current)
W: watts (power)
J: joules (energy or work)

Symbols
R: resistance
C: capacitance
L: inductance
V: voltage
I: current
X_C: capacitive reactance
X_L: inductive reactance
Z: impedance
f: frequency
ϕ: phase angle
$\mathrm{f_O}$: resonant frequency
$\mathrm{Z_O}$: dynamic (resonant) impedance

Electronics engineers tend to deal in a very wide range of numbers, from million-millionths of farads to millions of ohms. Hence it is usual to use a range of letter suffixes to indicate the order of magnitude of a number.
These are the standard suffixes used in this book:
p: $\times 10^{-12}$
n: $\times 10^{-9}$
μ : $\times 10^{-6}$
m: $\times 10^{-3}$
k: $\times 10^{3}$
M: $\times 10^{6}$

So, for instance, $\frac{1\text{V}}{1\,\text{k}\Omega} = 1\text{m}A$; $1\text{M} \times 1\,\text{nF} = 1\text{ms}$; and so on.

Appendix 2

Colour codes

The following colour codes are used on resistors, and sometimes on capacitors and inductors, to indicate value:
Number bands
Black 0
Brown 1
Red 2
Orange 3
Yellow 4
Green 5
Blue 6
Violet 7
Grey 8
White 9

Tolerance bands
Brown 1%
Red 2%
Gold 5%
Silver 10%

Figure A2 shows the significance of the bands for a resistor.

For example; a four band resistor with colours yellow–violet–red–red (the last one thick) is $47 \times 10^2 = 4\text{k}7$, 2%. A five band with brown–grey–black–brown–brown (again the last thick) is $180 \times 10 = 1\text{k}8$, 1%.

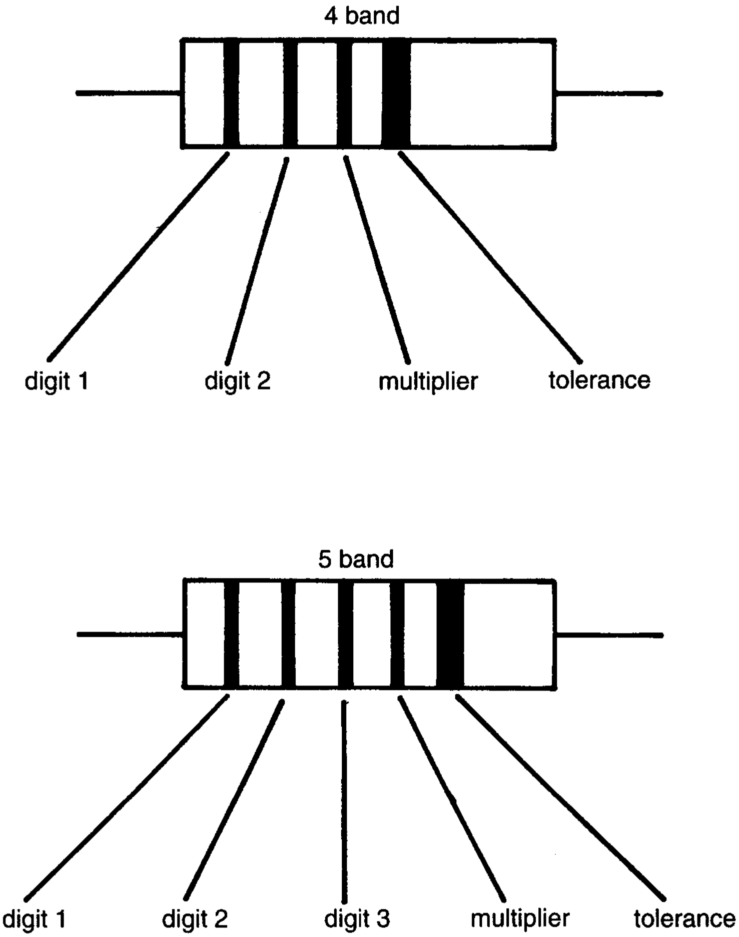

4 band

digit 1
digit 2
multiplier
tolerance

5 band

digit 1
digit 2
digit 3
multiplier
tolerance

Figure A2 Resistor colour codes

Appendix 3

The decibel

The decibel, or dB, is a logarithmic unit for expressing ratios, much beloved of engineers. Its popularity is due to the fact that it enables a wide range of numbers to be expressed conveniently.

Where we are talking about a ratio of two voltages then:

$$dB = 20\,log_{10}\left(\frac{V1}{V2}\right) \tag{A3.1}$$

which means that 'A is twice B' is often expressed as 'A is 6dB greater than B', and 'A is ten times B' becomes 'A is 20dB up on B'. Gain or attenuation of circuit blocks are very well expressed like this (a negative gain is an attenuation, or a ratio of less than one). The ratio is 'output/input'.

dB's are also used to express signal levels a lot. Here some reference level should be stated to make the situation clear. The most common are:

$$\mathrm{dBV} = 20\log_{10}(\mathrm{V_{rms}}) - \text{reference to } 1\mathrm{V_{rms}}$$

$$\mathrm{dBu} = 20\log_{10}\left(\frac{\mathrm{V_{rms}}}{0.775}\right) - \text{reference to 1mW into 600 } \Omega$$

though you may very well come across others. Where there is doubt it is often wise to try to find out what reference was used, as it can be a source of confusion.

The beauty of this system is that the level of a signal at various point in a chain can be found just by adding the dB values together. See Figure A. 3. It is apparent that the calculation in dB's is trivial compared to that using V_{RMS} and gain/attenuation as ratios.

dB's are also used sometimes to express relative power. The equation then used is:

$$\mathrm{dB} = 10\log_{10}\left(\frac{P1}{P2}\right) \tag{A3.2}$$

This is because power varies proportionally to the square of voltage, as implied by Eqs 2.4d and 2.4e.

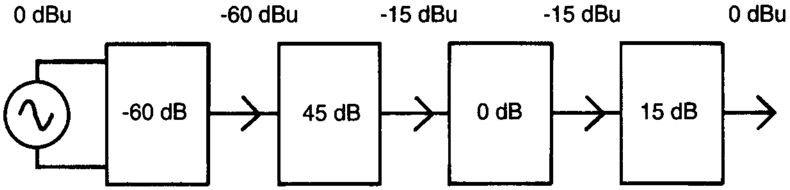

$$0.775\ \mathrm{V} \times \frac{1}{1000} = 0.775\ \mathrm{mV} \times 177.8 = 138\ \mathrm{mV} \times 1 = 138\ \mathrm{mV} \times 5.62 = 0.775\ \mathrm{V_{RMS}}$$

Figure A3 Illustration of signal level calculations using dB's

Appendix 4

Complex number analysis

The standard way to analyse a circuit containing AC quantities at a fixed frequency is to use complex numbers. A complex number has two parts, real and imaginary. The imaginary part is some function of the quantity j (i to pure mathematicians, but that would be confused with current by us), where j represents $\sqrt{(-1)}$. This is clearly a 'nonsense number', or at least unquantifiable, yet it has some useful properties:

$$j^2 = -1 \tag{A4.1a}$$

$$j^3 = -j \tag{A4.1b}$$

$$j^4 = 1 \tag{A4.1c}$$

We use multiplication by j to signify advancing its phase by 90°. Complex numbers owe their usefulness to the fact that they can be seen as representations of phasors. For example:

$$V = a + jb$$

is a voltage which can be represented in the j-plane as in Figure A. 4.

This number could also be represented as 'magnitude and phase' (as in Chapter 4) by using some trigonometry:

$$M = \sqrt{a^2 + b^2} \tag{A4.2}$$

$$\phi = \arctan\left(\frac{b}{a}\right) \tag{A4.3}$$

Then we are able to write the phasor in polar form:

$$V = M < \phi$$

Should we wish to convert a quantity from polar to complex form we need a little more trigonometry:

$$a = M cos\phi \tag{A4.4}$$

$$b = M sin\phi \tag{A4.5}$$

Mathematical operations with complex numbers

We treat complex numbers like any other algebraic quantity, keeping the j part separate as we would with x or y, but also applying Eqs A4.1 as necessary. So:

$$\text{Addition} : (a + jb) + (c + jd) = (a + b) + j(c + d) \tag{A4.6}$$

$$\text{Subtraction} : (a + jb) - (c + jd) = (a - c) + j(b - d) \tag{A4.7}$$

$$\text{Multiplication} : (a + jb)(c + jd) = (ac - bd) + j(bc + ad) \tag{A4.8}$$

Division: a useful tool for separating fractions is multiplication by the 'complex conjugate', which makes the denominator wholly real:
as:

$$(c + jd)(c - jd) = c^2 + jcd - jcd - j^2d^2 = c^2 + d^2$$

then:

$$\frac{a + jb}{c + jd} = \frac{(a + jb)(c - jd)}{c^2 + d^2} \tag{A4.9}$$

However, multiplication and division are sometimes more easily performed by converting to polar form and then:

Multiplication in polar form:

$$M_1 < \phi_1 . M_2 < \phi_2 = M_1 M_2 < (\phi_1 + \phi_2) \tag{A4.10}$$

Division in polar form:

$$\frac{M_1 < \phi_1}{M_2 < \phi_2} = \frac{M_1}{M_2} < (\phi_1 - \phi_2) \tag{A4.11}$$

Representation of capacitors and inductors

Complex numbers come into their own when we use them to represent reactive components in equations. The impedance of a capacitor becomes:

$$X_C = \frac{-j}{2\pi f C} \tag{A4.12}$$

and an inductor:

$$X_L = j2\pi fL \tag{A4.13}$$

and all of the analysis techniques of Chapter 5 can be used, following the algebraic rules given above.

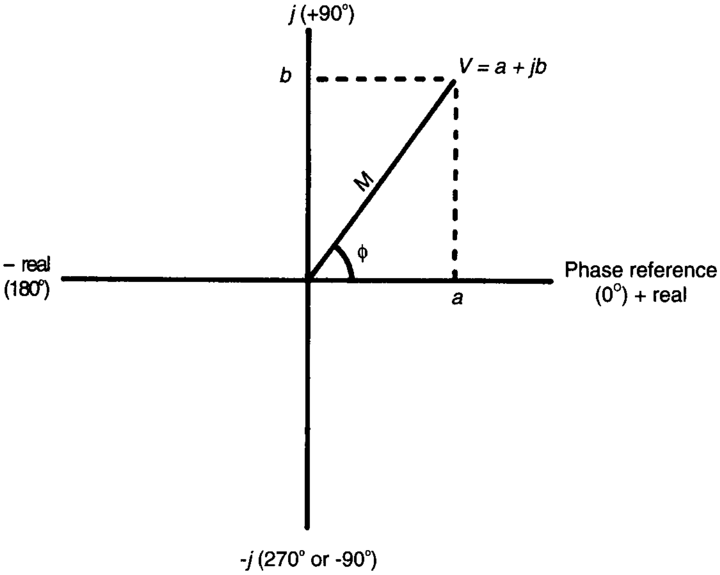
j (+90°)
b
$V = a + jb$
M
ϕ
– real
(180°)
a
Phase reference
(0°) + real
$-j$ (270° or -90°)

Figure A4 The j-plane

This Page Intentionally Left Blank

Further reading

Should the reader wish to gain greater understanding of any of the topics which this book touches on, I have found the following to be excellent:

Schuler, C. A. and Fowler, R. J. (1993) *Electric Circuit Analysis*, McGraw-Hill, London.
A clear and thorough explanation of circuit analysis techniques which also contains good chapters on electromagnetism and electrostatics.

Marston, R. (1993) *Electronic Circuits Pocket Book*, Vols One and Two, Butterworth-Heinemann, Oxford.
A comprehensive and informed circuit sourcebook that also analyses the ideas used in sufficient depth to make them really useful. Contains many useful circuits.

Horowitz, P. and Hill, W. (1989) *The Art of Electronics*, Cambridge University Press, Cambridge.
Deservedly considered to be the classic all-round electronics text. Clear, concise and entertaining too.

Phillips, G. (1993) *Newnes Electronics Toolkit*, Butterworth-Heinemann, Oxford.
A great deal of useful information in a small space; an ideal source of reference for the workshop or lab.

This Page Intentionally Left Blank

Index

AC, defined, 26
alternating current, definition of, 26
ampere, definition of, 7
antiphase, 30
atoms, fanciful behaviour of, 5, 14
  insecurity about, 5
bad smells, 23
bandwidth, see series and parallel tuned circuit
cable, capacitance of, 16
  current rating of, 11
  long, 11, 16
  resistance of, 11
calculations, checking, 40, 41, 46
  involving phase, 30
  labelling circuits for, 42
capacitance, definition of, 15
  explanation of, 15
capacitors, 15
  as DC block, 22
  as ramp generators, 22
  effect on phase, 32
  in series and parallel, 17
  non-standard values, 57
  polarized, 16
  power in, 37
  reactance of, 31
  similarity with inductors, 18, 23, 32
  working voltage, 16
charge, definition of, 6
  free electrons, 6, 10
  holes, 6
  on capacitor, 15
circuit, definition of, 11
open, definition, 12
  short, definition, 12
  labelling quantities, 42
complex numbers, 30, 39, 291
components, standard values, 55
conductors, 10
  voltage across, 11
coulomb, definition of, 7
CR circuit, see first-order circuits
current, and magnetic field, 7, 15
  conventional, 7
  definition of, 6
  destruction due to, 11
  direction of, 7
  electron, 7
  ideal and real source, 44
  leakage, 16
  speed of, 8
dB, 289
DC, defined, 20
  level on signal, 35
decibels, 289
diagrams, block, 12
  circuit or schematic, 12
differentiator, see first-order circuits
direct current, definition, 20
earth, as reference for voltage, 8
electrons, free electrons, 6, 10
  holes and, 6
  speed of, 8, 23
EMF, 8
equivalent circuits, Thévenin's and Norton's, 45
farad, definition of, 15
filter, high-pass, see first-order circuits
filter, low-pass, see first-order circuits
first-order circuits, 205
  frequency response, 212, 213
  instantaneous voltage and current, 209
  phase response, 213
  time constants, 207
  time response, 205
  turnover frequency, 212
  and capacitance, 32
  back EMF in LR, 207
Fourier analysis, 33
frequency, definition of, 28
  very low, 28
harmonics, 33
heat, generation by current, 23

henry, defined, 17
hertz, definition of, 28
holes, 6
impedance, 32
inductance, definitions of, 15, 17
  explanation of, 17
  mutual, 18
  self, 17
inductors, 17
  Q of, 78
  DC current rating, 18
  as chokes, 22
  current ramp in, 22
  in series and parallel, 18
  non-standard values, 57
  power in, 37
  reactance of, 32
  resistance of, 18
  saturation of, 18
  similarity with capacitors, 18, 23, 32
insulators, 10
  breakdown of, 23
integrator, see first-order circuits
Kirchhoff, current law, 41
  voltage law, 40
LR circuit, see first-order circuits
mark-to-space ratio, 35
MMF, 6
Norton's equivalent, 45
ohm, definition of, 10
Ohms Law, 10
  defined, 10
  in capacitors, 31
  in inductors, 32
op-amps, 259
  inverting, 259
  non-inverting, 261
  easy rules for, 259
open circuit, definition, 12
parallel circuits, rules for, 41
parallel tuned circuit, 226
  conditions for resonance, 230
  dynamic impedance, 229
  ideal, 227
  practical, 228
  Q factor, 229
  resonant frequency for practical, 229
peak amplitude, 27
peak to peak amplitude, 27
period, definition of, 28
permeability, 17
phase cancellation, 31
phase, definition of, 29
  effect on power, 37
  reference, 29
  rules for calculation, 30
phasors, 30, 291
potential difference (also see voltage), 8
power supplies, 8, 20
power, at DC, 23
  definition of, 24
  explanation of, 23
  for AC, 36
Q factor, see series and parallel tuned circuit
quadrature, 30
ramp generator, 22
ramp wave, 34
  RMS value of, 37
rectangle wave, 34
  RMS value of, 37
resistance, 10
resistivity, 10
resistors, 12
  colour codes, 287
  in parallel, 13
  in parallel, rules of thumb, 14
  in series, 13
  non-standard values, 55
  power rating of, 23
  tables for power, 102
resonance, see series and parallel tuned circuit
    risetime, see square wave
RMS, defined, 36
  for sine waves, 27
  value for misc signals, 37
rules of thumb, resistors in parallel, 14
  power, 23, 102
sawtooth wave, 34
  RMS value, 37
series circuits, rules for, 39
series tuned circuit, 223
  Q factor, 224
  bandwidth, 225
  explanation of, 224
  impedance at resonance, 224
  resonant frequency, 224
short circuit, definition, 12
signals, DC level of, 35
  fundamental of, 33
  vague definition of, 12
  very low frequency, 28
sine waves, amplitude of, 27
  definition of, 26
  importance of, 33

square wave, 34
  risetime of, 33
  RMS value of, 37
suffixes, 286
Thévenin's equivalent, 45
transformers, 18
triangle wave, 34
  RMS value of, 37
units, 285
voltage divider, 142
  effect of source and load on, 143
voltage, definition of, 8
  ideal and real source, 45
  reference for, 8
  speed of, 8
  work done by, 8

This Page Intentionally Left Blank

This Page Intentionally Left Blank

rojects you can build! **GERNSBACK SPECIALTY SERIES** 49604

996 Electronics EXPERIMENTER'S handbook

lectronics NOW

The Project Anthology

An Exclusive Collection of 18 Fun Projects!

NCLUDING: Computer Voice • Vol-Expander • Twinkle Tree • Sound Switch • Audible Timer • Minefields Game • Audio Power Booster • Light Alarm • Sig-Tracer

...AND MUCH MORE

PRECISION Audio Signals from your PC

ISOLATION Transformer you can build

MILLIOHM Tester adapts to any VOM

CUSTOM Meter Faces your PC designs

PLUS!

PROTOTYPING STATION

Build a full-featured Breadboard laboratory

See Page 20

61 >
0 09128 49604 7

$3.95 U.S.
$4.50 CAN.

A GERNSBACK PUBLICATION

CALL TOLL FREE
(800) 292-7711
Se Habla Español

C&S SALES
EXCELLENCE IN SERVICE

WRITE FOR A FREE 60 PAGE CATALOG!

AFFORDABLE, HIGH QUALITY ELENCO OSCILLOSCOPES
2 YEAR WARRANTY

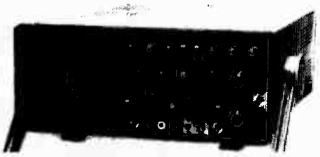

STANDARD SERIES

S-1325 25MHz $335
S-1340 40MHz $489

Features:
- TV Sync
- 1mV Sensitivity
- X - Y Operation
- High Luminance 6" CRT
- Complete Schematic
- Plus much, much more!!

2 FREE probes included with each scope!!

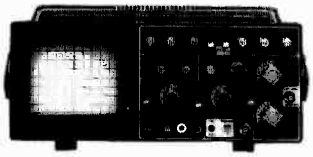

DELUXE SERIES

S-1330 25MHz $439
S-1345 40MHz $569
S-1360 60MHz $759

Features:
- Delayed Sweep
- Automatic Beam Finder
- Z Axis Modulation
- Built-in Component Test
- Dual time base
- Illuminated internal gradicule
- Plus all of the features of the "affordable" series!!

MX-9300

Four Functions In One Instrument

Features:
- One instrument with four test and measuring systems:
  - 1.3GHz Frequency Counter
  - 2MHz Sweep Function Generator
  - Digital Multimeter
  - Digital Triple Power Supply
- 0-30V @ 3A, 15V @ 1A, 5V @ 2A

NEW!

$479.95

B&K 2MHz Function Generator Model 4010

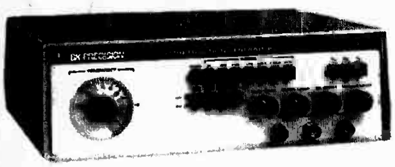

$195.95

- 0.2Hz to 2MHz
- Sine, Square, Triangle, Pulse and ramp output
- Variable duty cycle
- Variable DC offset

Model XP-581

4 Fully Regulated Power Supplies In One Unit

4 DC voltages: 3 fixed - +5V @ 3A, +12V @ 1A, -12V @ 1A • 1 Variable - 2.5 - 20V @ 2A

Ideal for laboratories, service shops and hobbyists.

$85

Fluke Multimeters

70 Series	
Model 70II	$69.95
Model 73II	$97.50
Model 75II	$135.00
Model 77II	$154.95
Model 79II	$175.00
80 Series	
Model 83	$235.00
Model 85	$269.00
Model 87	$289.00

Model M-6000

Features Computer Interface and Software

- Frequency to 4MHz
- Capacitance to 40nF
- Large 3 3/4 LCD Display
- Temperature to 1999°F
- TTL logic
- Data/Peak Hold
- Audible continuity
- Auto power off
- Unit indicator
- Diode test
- w/ RS232 Cable

$99.95

B&K Model 1688

High Current (25 amp) Power Supply

Variable 3-14VDC

$249.95

1.3GHz Universal Counter
Model F-1300

- Period
- Frequency
& 8 other functions

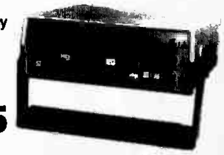

$225

M-1700
Digital Multimeter

11 functions including freq to 20MHz, cap to 20μF. Meets UL-1244 safety specs.

$75

XK-550 Digital / Analog Trainer

Elenco's advanced designed Digital / Analog Trainer is specially designed for school projects. It is built on a single PC board for maximum reliability. It includes 5 built-in power supplies, a function generator with continuously sine, triangular and square wave forms. 1560 tie point breadboard area.

XK-550
Assembled and Tested
$169.95

XK-550K - Kit
$139.95

Tools and meter shown optional

Kit Corner over 100 kits available

Model AR-2N6K

2 meter / 6 meter Amateur Radio Kit
$34.95

Model AM/FM-108K

AM/FM Transistor Radio Kit
$29.95

M-2665K
Digital Multimeter Kit
$54.95

Full 34 Ranges w/ cap & trans tests

TT-400K Telephone Analyzer Kit

$19.95

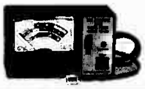

WE WILL NOT BE UNDERSOLD

UPS SHIPPING: 48 STATES 5%
OTHERS CALL ($3 min. / $10 max.)
IL RES. 8% TAX

C&S SALES, INC.
150 W. CARPENTER AVENUE
WHEELING, IL 60090
FAX: (708*) 520-9904 (708*) 541-0710
* NOTE: Effective January 20, 1996, our area code will be changed to 847.

15 DAY MONEY BACK GUARANTEE
FULL FACTORY WARRANTY
PRICES SUBJECT TO CHANGE WITHOUT NOTICE

Electronics NOW®

ELECTRONICS EXPERIMENTER'S handbook™

1996

Hugo Gernsback (1884-1967) founder

Larry Steckler, EHF, CET,
editor-in-chief and publisher

Julian S. Martin
handbook editor

EDITORIAL DEPARTMENT
Brian C. Fenton, editor
Neil Sclater, associate editor
Teri Scaduto, assistant editor
Jeffrey K. Holtzman
computer editor
Dan Karagiannis, assistant editor
Robert Grossblatt, circuits editor
Larry Klein, audio editor
David Lachenbruch
contributing editor
Don Lancaster
contributing editor
Evelyn Rose, editorial assistant
Andre Duzant, art director
Russell C. Truelson, illustrator

PRODUCTION DEPARTMENT
Ruby M. Yee, production director
Kathy R. Campbell,
production assistant

CIRCULATION DEPARTMENT
Jacqueline P. Cheeseboro
circulation director
Theresa Lombardo, circulation assistant

REPRINT BOOKSTORE
Michele Torrillo
manager

Typography by **Mates Graphics**
Cover design by **David Loewy**

Advertising Sales Offices listed on page 111.

**Gernsback Publications, Inc.,
500 Bi-County Blvd., Farmingdale, NY 11735. 516-293-3000**

As a service to readers, *1996 Electonics Experimenter's Handbook* publishes available plans or information relating to newsworthy products, techniques and scientific and technological developments. Because of possible variances in the quality and condition of materials and workmanship used by readers, we disclaim any responsibility for the safe and proper functioning of reader-built projects based upon or from plans or information published in this magazine.

Since some of the equipment and circuitry described in *1996 Electronics Experimenter's Handbook* may relate to or be covered by U.S. patents, we disclaim any liability for the infringement of such patents by the making, using, or selling of any such equipment or circuitry, and suggest that anyone interested in such projects consult a patent attorney.

Electronics Experimenter's Handbook is published annually by Gernsback Publications Inc. All rights reserved. Printed in U.S.A. Single copy price $3.95, Canada $4.50. Canadian GST Registration No. R125166280. © Gernsback Publications Inc., 1995, 1996.

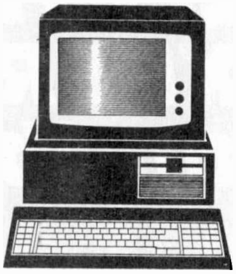

Editorial

There are many and varied good reasons for building a project from plans published in a magazine. I won't go into listing all these reasons because I would probably leave your motivation off the list. Now then, can you think of one bad reason for building a project from a magazine plans? Think about it.

I find that too many project builders follow the plans scrupulously, not varying from the schematic diagram. They even produce a project that mirrors the photographs supplied in the construction article. Admittedly, editors prefer that readers build projects than not. Albeit, why not show some creativity. Every project can be improved upon in some way. A cosmetic change to suit your fancy may not be an electronic-design advancement, but it does show some individualism.

To encourage the builder's development of published instruction articles, the editors decided to publish The Project Anthology that appears in the center of this magazine. This Anthology is a Lilliputian compilation of varied projects that are bound to excite the builder into assembling more than one of them. Since the complete plans are not provided (in most cases only schematic diagrams and parts lists are provided), the plans leave the final assembly design up to the builder. Also, the circuits themselves are simple enough so that the introduction of new electronic design features can be added.

What the editors are attempting to do is promote your sculpturing of the project to suit on-site needs. We want you to build what you need from our plans. The foray into designing circuits eventually becomes habitual and rewarding.

Who knows! You may become one of our authors when your projects are designed from your original plans!

Julian S. Martin

Julian S. Martin,
Editor

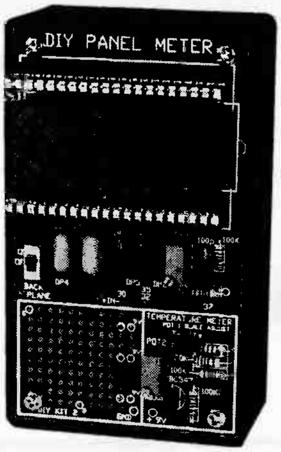

Solid-State Thermometer
See page 39

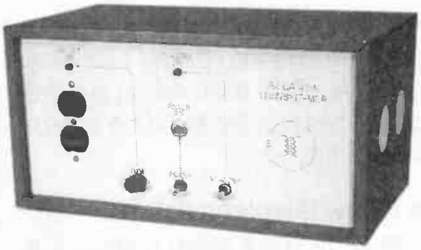

Isolation Transformer
See page 43

Put that Phone on Hold
See page 36

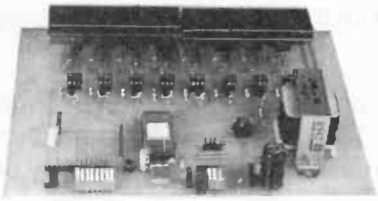

Power Controller
See page 100

MILLIOHM
ADAPTER

Milliohm Tester
See page 82

ELECTRONICS EXPERIMENTER'S handbook™ 1996

CONTENTS

Custom Meter Faces 16
Fabricate pro-like meter scales with your PC and laser printer.

Prototyping Station 20
Construct a full-featured breadboard laboratory with custom features to suit your taste.

WWV Receiver 27
A simple superheterodyne receiver that pulls in atomic time and other useful signals.

Sinewave Doubler 32
Double the frequency of any sinewave from 10 Hz to more than 50 MHz.

Put that Phone on Hold 36
A novel telephone circuit that works from any phone in your houseóeven cordless.

Solid-State Thermometer 39
An accurate handheld device you can build from scratch or from a kit.

Isolation Transformer 43
Build an isolation transformer to protect yourself—and your equipmentówhile you service electronic appliances.

Precision Audio Signals from Your PC 46
You can have an accurate computer-controlled audio generator for under ten bucks.

Milliohm Tester 82
Measure low resistances with this super simple circuit that connects to your digital multimeter.

Voltage Cursor Adapter 84
Measure voltage values of waveforms accurately and easily on your oscilloscope.

An
Electronics NOW®
Publication

Introducing. . .
The Projects Anthology
18 Electronic Gadgets You Can Build!

The Project Anthology

A collection of eighteen projects that will add spice to your project building pleasure. .... pages 49-80

1—**Twinkle Tree**, LED light jollies, 2—**One-Second Flasher**, for the shutter-buff bug, 3—**5-Volt Power Supply**, to build more projects, 4—**Sound-Switch**, a must for voice actuated action, 5—**Light Alarm**, for dark-only areas, 6—**Sig-Tracer**, discover where the signal disappears, 7—**CW Filter**, for better code reception, 8—**Vol-Expander,** squeezes sound to tape, 9—**Computer Voice**, a fun voice effect, 10—**Audio Power Booster**, double your power, double your listening pleasure, 11—**Pit Stop**, a racy game, 12—**Minefields,** a boomy game, 13—**Kapellmeister,** a simple transmission-line loudspeaker system, 14—**Helix**, apartment-dweller's hidden antenna, 15—**Inverted V Antenna**, a modified half-wave antenna that pulls in the signals, 16—**Audible Timer**, this beep is for you, 17—**Long Interval Timer**, up to two weeks long, 18—**Twangy Distortion Unit**, in tribute to Jerry Garcia.

Mini Logic Analyzer 87
With about $30 worth of parts and an IBM/compatible computer you are ready for the troubleshooting big leagues

Magnetic Field Meter 95
Determine your exposure to line-frequency magnetic-fields with this portable ELF gaussmeter.

Power Controller 100
Control eight channels of 120-volt AC with a 68705 microcontroller.

Departments

Editorial 1
New Products 6
New Literature 13
Classified Advertising 111
Advertising Index 111
Advertising Sales Offices 112
Free Information Coupon 15, 108, 111

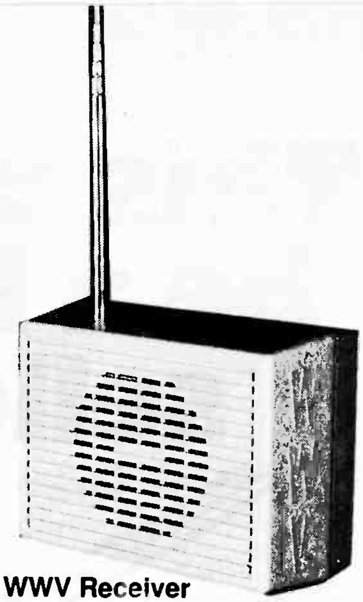

WWV Receiver
See page 27

Audio Signals from Your PC
See page 46

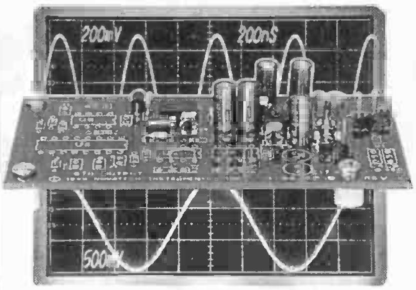

Sinewave Doubler
See page 32

220
ON
OFF
ATL
φ-METER

Magnetic Field Meter
See page 95

This could be a best seller.

But it's free.

You can't buy the Consumer Information Catalog anywhere. But you can send for it, free!

It's your guide to more than 200 free or low-cost government publications about getting federal benefits, finding jobs, staying healthy, buying a home, educating your children, saving and investing, and more.

Send today for your latest free Consumer Information Catalog. The Catalog is free. The information is priceless.

Send your name and address to:

Consumer Information Center
Department BEST
Pueblo, Colorado 81009

Starting off with a few chuckles!

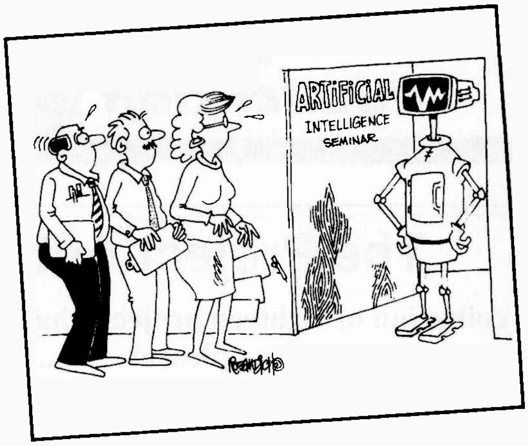
ARTIFICIAL
INTELLIGENCE
SEMINAR

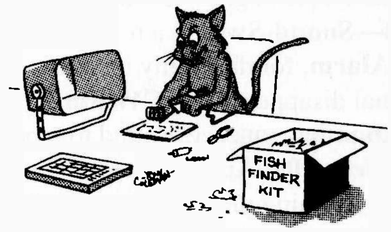
FISH
FINDER
KIT

"Meow!"

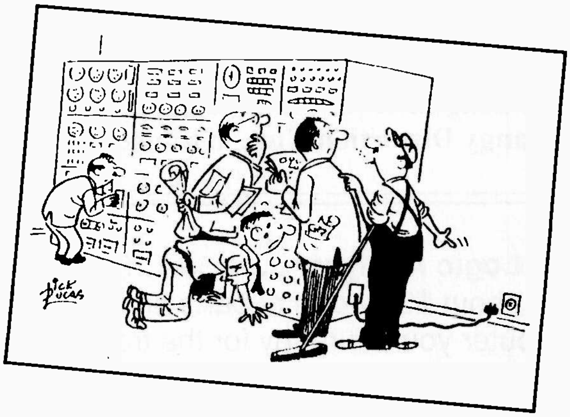

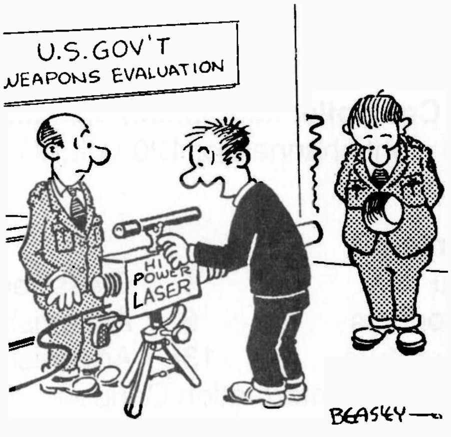
U.S. GOV'T
WEAPONS EVALUATION
HI POWER LASER

"Now I remember—first you put the safety on, then you plug the unit in!"

New and Pre-Owned Test Equipment

Substantial SAVINGS on New & Pre-Owned B+K Precision, Fluke, Hewlett-Packard, Goldstar, Leader, Tektronix, and more ...

LG Precision
New Oscilloscope Specials

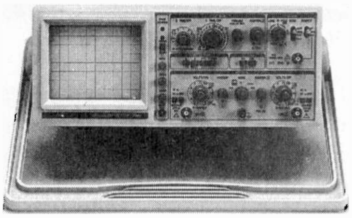

FREE SHIPPING!
ON ALL LG PRECISION EQUIPMENT ANYWHERE IN THE U.S.
Excluding AK & HI

Model OS-9100P → **$999.00**
SUPER SALE!

Full 100 MHz Bandwidth!

- Dual-Channel, High Sensitivity
- TV Synchronization Trigger
- Calibrated Delayed Sweep
- Includes Two Probes, 2 Year Warranty

Professional Quality Digital Storage Oscilloscopes at Affordable Prices!

- Transmit and Receive Data with Standard RS-232C Interface
- Includes FREE Comprehensive Communication Software
- Waveform Save, Calculation, and Print
- Dual Channel. Delayed Sweep, TV Synchronization Trigger
- 20 MS/s Sampling Rate, Two Save Memories
- Switch Between Analog and Digital Modes
- CRT Readout Including Measurement Cursors

Bandwidth	20 MHz	40 MHz	60 MHz
Model	OS-3020	OS-3040	OS-3060
SALE Price	$1,199.00	$1,599.00	$1,899.00

Pre-Owned Oscilloscope Specials

Tektronix 465	100 MHz	$589.00
Tektronix 465B	100 MHz	$689.00
Tektronix 475	200 MHz	$749.00
Tektronix 475A	250 MHz	$849.00

- Professionally Refurbished
- Calibrated to Original Specifications
- Dual Channel, Calibrated Delayed Sweep
- 90 Day Warranty

BK PRECISION
MAXTEC INTERNATIONAL CORP.

Professional Computer Monitor Pattern Generator

Model 1280 $429.00 **NEW!**

- Tests CGA, EGA, VGA, SVGA, and MAC II Monitors
- Outputs Interface to DB 9, HD 15, DB 15 and BNC
- Interlaced or Non-Interlaced
- Color Bars, Raster, and Crosshatch Patterns

20 MHz Sweep/Function Generator

Model 4040 $499.00 **NEW!**

- 0.2 Hz to 20 MHz, 5 digit LED Display
- AM & FM Internal or External Modulation
- Sine, Square, Triangle, TTL, CMOS Outputs
- Burst Operation
- External 30 MHz Frequency Counter

Don't Leave Your Computer Unprotected!

Fairstone Uninterruptible Power System...... Reliable Blackout, Brownout, Spike and Surge Protection.

VA Rating	500	600	750
Model	P500	P600	P750
SALE Price	$239.00	$269.00	$329.00

- Sleek Good Looks and Small Footprint Design
- Cold Start Function, LCD Status Display

Automatic Monitoring and Shutdown Software for Windows with cable .... **Only $79.00**

Full line of Oscilloscopes, RF, Video & Audio Test Equipment, Power Supplies, Meters, Probes and Accessories.

VISA MasterCard AMEX C.O.D.

TOLL FREE 1-800-99-METER
1-800-996-3837

Availability of Equipment Subject to Prior Sale
Prices Subject to Change Without Notice

FOTRONIC
CORPORATION
P.O. BOX 708
Medford, MA 02155
(617) 665-1400
FAX (617) 665-0780

NEW PRODUCTS

USE THE FREE INFORMATION CARD FOR FAST RESPONSE

OSCILLOSCOPE TRIGGER DEVICES

THE SUPERPROBE OSCILLOSCOPE triggering devices from Programmable Designs are instruments that can trigger complex events in digital circuitry and simplify circuit debugging and testing. They will work with any oscilloscope to provide a wide range of event-triggering options. Four different models are available.

The SuperProbe II has 18 digital logic inputs and a clock input. It supports three clocked triggering modes and one combinatorial (pattern-match) triggering mode. The device works with flexible logic combinations for specifying triggering events. These include no-match triggering for triggering when certain unexpected events occur. This model has separate "pattern select" and "don't care" configuration DIP switches.

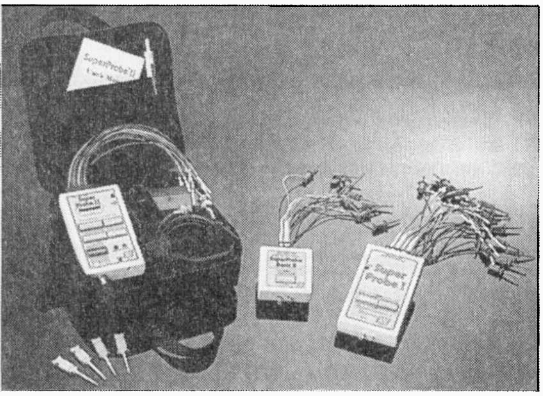

CIRCLE 60 ON FREE INFORMATION CARD

Status LEDs permit the monitoring of power, input signals, triggering activity, and clock activity. This instrument includes an interface cable assembly with removable grabbers and gold-plated machined contacts that work with IC clips and 0.025-inch posts.

SuperProbe I is a pattern-match or word-recognition version with 17 digital inputs. It allows any logical combination with up to 17 signals to be specified for the trigger event. This model is offered in two versions. The standard version includes an interface cable with permanently attached grabbers for through-hole component leads. The deluxe version has a cable identical to the one included with the SuperProbe II.

The SuperProbe Basic-8 permits pattern-match triggering with up to eight signal inputs. It includes a cable with permanently attached through-hole component grabbers.

SuperProbe pricing ranges from $99.00 to $549.00, depending on the model.

PROGRAMMABLE DESIGNS INC.

41 Enterprise Drive
Ann Arbor, MI 48103-9503
Phone: 313-769-7540
Fax: 313-769-7242
E-mail: design@prog-designs.com

INDUSTRIAL CONTROL SYSTEM

THE QED INDUSTRIAL Control System (ICS) from Mosaic Industries is a single-board computer with either a character or graphics LCD display, and a keypad or touchscreen. It has a battery-backed memory, real time clock, analog and digital I/O, board area for prototyping, and provision for serial communications. The system is packaged in an industrial-style enclosure with screw-terminal connections.

This packaged system is intended for many different applications including machine automation, data acquisition, and robotics. The ICS can interpret data from temperature, pressure, and optical sensors; process the input data; and control stepper motors, valves, relays, and other actuators.

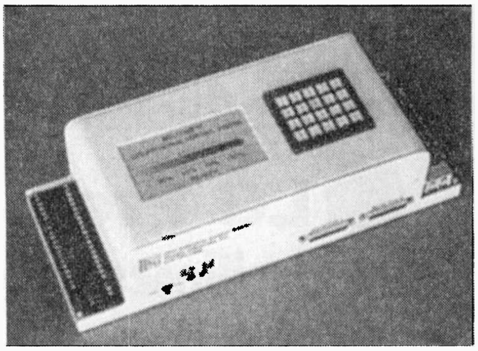

CIRCLE 37 ON FREE INFORMATION CARD

Palm-sized plug-in cards introduce isolated high-current outputs and high-voltage inputs, 4 to 20-milliampere signalling, and analog-to-digital or digital-to-analog conversions. The system modules can be programmed in C or Forth from a personal computer with on-board multitasking software.

The QED Industrial Control System with character display and keypad (ICS-AK) is priced at $995.00. With a graphics display and keypad (ICS-GK), it is priced at $1190.00.

MOSAIC INDUSTRIES, INC.

5437 Central Avenue, Suite 1
Newark, CA 94560
hone: 510-790-1255
Fax: 510-790-0925

MONOLITHIC CRT DRIVER

THE LM2406T FROM NATional Semiconductor is advertised as the industry's first monolithic IC cathode-ray tube driver. This triple-channel, 40-MHz CRT driver, designed for SVGA and XGA monitors. The bipolar device is capable of supporting scan frequencies up to 58 kHz/72 Hz.

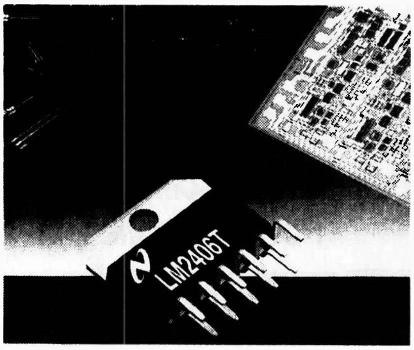

CIRCLE 42 ON FREE INFORMATION CARD

AMAZING Electronic and Scientific Products

Mystery Levitating Device!

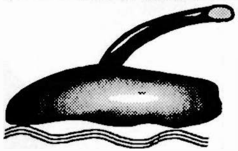

Remember War of the Worlds? Objects float in air and move to the touch. Defies gravity, amazing gift, conversation piece, magic trick or great science project.

ANT1K Easy to Assemble Kit / Plans **$19.50**

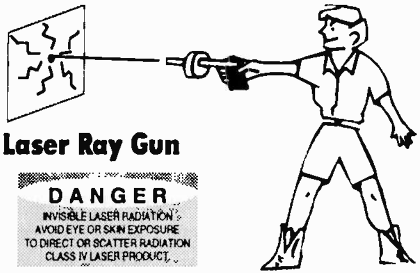
DANGER
INVISIBLE LASER RADIATION
AVOID EYE OR SKIN EXPOSURE
TO DIRECT OR SCATTER RADIATION
CLASS IV LASER PRODUCT

Laser Ray Gun

Advanced project produces a burst of light energy capable of burning holes in most materials. Hand-held device uses rechargeable batteries. 500 joules of flash energy excite either a neodymium glass, yag or other suitable 3" laser rod. This is a dangerous CLASS IV project (individual parts/assemblies available). **LAGUN1 Plans** **$20.00**

LAGUN1K Kit / Plans ....**Price on Request**

Extended Play Telephone Recording System

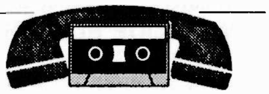

READY TO USE! Automatically controls and records on our X-4 extended play recorder, taping both sides of a telephone conversation. Intended for order entry verification. Check your local laws as some states may require an alerting beeper.

TAP20X Ready to Use System.................................. **$129.50**

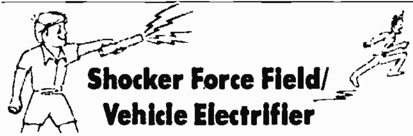

Shocker Force Field/ Vehicle Electrifier

Neat little device allows you to make hand and shock balls, shock wands and electrify objects, charge capacitors. Great payback for those wise guys who have wronged you!

SHK1K Easy to Assemble Electronic Kit............ **$29.50**

Electric Charge Gun

All New Technology!

Stuns/immobilizes attackers up to 15 feet away! •Legal in most state (not in NY, NJ, MA, WI) • More knock-down power than most handguns • No permanent injury • ID coded • Free 80KV stun gun with every purchase.

ECG1 Data Packet, Creditable toward purcase............ **$10.00**

ECG10 Charge Gun, Ready to Use, w/Free 80KV Gun **$249.50**

Homing / Tracking Transmitter

Beeper device, 3 mile range.

HOD1 Plans ....**$10.00** **HOD1K Kit / Plans** **$49.50**

Listen Through Walls, Floors

Highly sensitive stethoscope mike.

STETH1 Plans.........**$8.00** **STETH1K Kit/Plans**......... **$44.50**

Infinity Transmitter ++

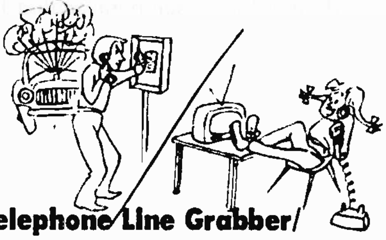

Telephone Line Grabber/ Room Monitor / Controller

All New - The Ultimate in Home/Office Security & Safety!

Simple to use! Call your home or office, push a secret tone on your telephone keypad to access: • On premises sounds and voices • Ongoing phone conversation w/break-in capability • Up to 10 external electrical functions, lights, TV, alarms, coffee pots, heater, etc. CAUTION! Check legality with your state's attorney general's office before use for monitoring of voices.

TELECON1K Kit, Includes PC board **$149.50**

TELECON10 Ready to Use **$199.50**

Visible Beam Laser

Easy to build, RED Beam, visible for miles. Use for light shows, window bounce holography, cloud illumination and much more! **LAS1KM Kit w/1mw Laser Tube, Class II.** .... **.$69.50**

LAS3KM Kit w/2.5mw Laser Tube, Class IIIA **$99.50**

Life is Precious – *PROTECT IT!*

Hard hitting, crackling, sizzling HIGH VOLTAGE plasma! Stuns and immobilizes most attackers

STUN200 Ready to Use, 200,000V **$69.50**

STUN120 Smaller Unit, 120,000V **$39.50**

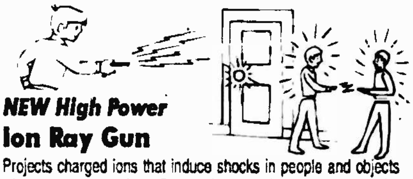

NEW High Power Ion Ray Gun

Projects charged ions that induce shocks in people and objects without any connection! Great science project as well as a high tech party prank. **IOG7 Plans** **$15.00**

IOG7K Kit/Plans **$99.50**

Invisible Pain Field Generator

Shirt pocket size electronic device produces time variant complex shock waves of intense directional acoustic energy, capable of warding off aggressive animals, etc.

IPG7 Plans ....**$8.00** **IPG7K Kit/Plans** ...**$49.50**

IPG70 Assembled.. **$74.50**

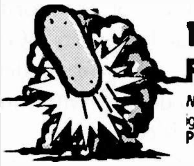

1000 Ft++ Potato Cannon

NOT A TOY. Uses electronic or piezo ignition. CAUTION REQUIRED!

POT1 Plans......
(Dangerous Product)............**$10.00**

FireBall Gun

Shoots flaming ball - two shot capacity Great for special effects and remote fire starting. CAUTION REQUIRED!

FIREBALL Plans (Dangerous Product).....**$10.00**

TV & FM Joker / Jammer

Shirt pocket device allows you to totally control and remotely disrupt TV or radio reception. Great gag to play on family or friends. Discretion required.

EJK1KM Easy to Assemble Electronic Kit **$24.50**

ATTENTION: High Voltage Fans!

4,000 volts in the palm of your hand! Experiment with anit-gravity, hovercraft, ion guns, force fields, plasma guns, shock devices, wireless energy and electrical pyrotechnics. Input: 9-14VDC.

MINIMAX4 Ready to Use .... **$ 19.50**

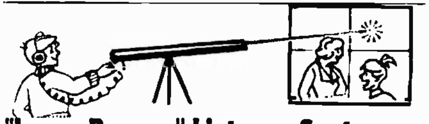

"Laser Bounce" Listener System

NEW - Latest Technology! Allows you to hear sounds from a premises without gaining access. Aim at room window and listen to sounds from within via reflected laser light. Not for illegal use. Requires video tripods.

LWB3K 5mw Laser and Receiver Kit **$149.50**

LWB30 Ready to Use **$199.50**

5mw Visible Red Pocket Laser

Utilizes our touch power control!

VRL5KM Kit / Plans....................................**$ 74.50**

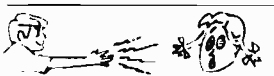

Electronic Hypnotism

Puts subjects under control using highly effective electronic stimuli. Intended for parties and entertainment but must be used with caution. Includes valuable text book reference and plans.

EH2 Plans and Text Book.. **$19.50**

Automotive NEON

Easy-to-Install 4-Tube Kit for Cars, Trucks, Vans!

Available in Pink, Purple, Blue or Green - please specify color when ordering.

RG4K (Specify Color) ....**$ 129.50**

Flash-To-Music Option for above kit FMU1......................**$ 29.50**

License Frame Kit (Specify Color)

LIC1K ....**$ 24.50**

3 Mi FM Wireless Microphone

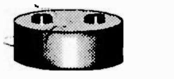

Subminiature! Crystal clear, ultra sensitive pickup transmits voices and sounds to FM radio. Excellent for security, monitoring of children or invalids. Become the neighborhood disk jockey!

FMV1 Plans......**$7.00** **FMV1K Kit and Plans** **$39.50**

Telephone Transmitter – 3 Mi

Automatically transmits both sides of a telephone conversation to an FM radio. • Tunable Frequency • Undetectable on Phone • Easy to Build and Use • Up to 3 Mile Range • Only transmits during phone use. **VWPM7 Plans** **$7.00**

VWPM7K Kit/Plans **$39.50**

INFORMATION UNLIMITED

Dept PEM-19, Box 716, Amherst, NH 03031
Phone: 603-673-4730 FAX: 603-672-5406
MC, VISA, COD, Checks accepted Please add $5.00 Shipping & Handling

CATALOG
With many more items!
Free with Order or send $1 P&H

Order by Mail or by 24 Hour Orders-Only Phone
800-221-1705

The LM2406T can drive a cathode capacitance of 8 pF with a voltage swing of 40 volts peak-to-peak, and it has a rise and fall time of 9 nanoseconds. When paired with National Semiconductor's LM1203N or LM1207N video amplifiers, the combination forms a complete video channel for SVGA and XGA computer monitors.

The monolithic IC includes internal protection to prevent momentary short circuiting. It reduces monitor component count, simplifies system design and manufacture, and increases system reliability .It is offered in a reduced-size TO-220 package.

The LM2406T monolithic CRT driver is priced at $2.50 in quantities of 1000.

NATIONAL SEMICONDUCTOR CORPORATION
2900 Semiconductor Drive
Santa Clara, CA 95052-8090
Phone: 1-800-272-9959

MULTIMETER WITH CLAMP-ON ADAPTER

THE AMPROBE MODEL PMM-1 is designed so that a clamp-on adapter can be connected, allowing it to measure up to 300 amperes. The basic DMM is a 3200-count instrument that measures AC and DC voltage to 600 volts and ohms to 30 megohms. It has an annunciated LCD readoutand a 65-segment analog bargraph,. The DMM offers auto and manual ranging, audible continuity testing and data hold. It also performs diode checks The PMM-1 has 600-volt AC and DC protection on all ranges.

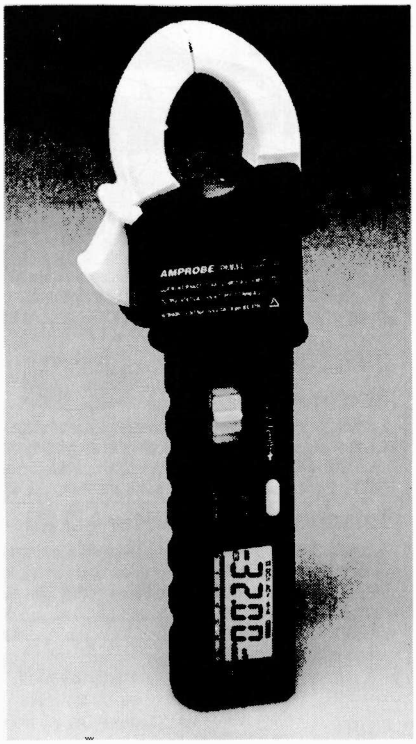
CIRCLE 56 ON FREE INFORMATION CARD

With the addition of the PMM-C clamp-on adapter, the PMM-1 can measure up to 300 amperes AC. The adapter jaws will close over conductors up to 1.14-inch in diameter. The combination has an output of 1 millivolt AC per 0.1-ampere AC. With additional optional accessories, thePMM can do infrared scanning and measure up to 1000 amperes AC, 600 amperes DC, temperature, and up to 400 kilowatts.

The PMM-1 digital multimeter with carrying case, test leads, batteries, and a user's manual is priced at $69.85. The PMM-C clamp-on accessory is priced at $49.85.

AMPROBE INSTRUMENT
630 Merrick Road
Lynbrook, NY 11563
Phone: 1-800-477-VOLT
or 516-593-5600
Fax: 516-593-5682

COMPUTER CARD FOR MULTIMEDIA PRESENTATIONS.

THE MODEL ME1000 (VGA2-NTSC) plug-in computer card from Methode Electronics allows multimedia presentations to be displayed on a standard television receiver or recorded on a VCR. The card converts VGA computer graphics, computer-aided designs, and multimedia presentations to standard NTSC video output.

The card automatically detects and switches between 640×480 pixel, 320×200 pixel, and 80-column text modes. The plug-in card works with any IBM or compatible personal computer that has an Intel 386, 486, or Pentium processor , an ISA bus, and a graphics card with the appropriate connector.

The ME1000 also provides flicker-free images, CAD-drawing output with no missing lines, and sharp 80-column text mode. The card conforms to multimedia standards, and it is compatible with AutoCAD and 3D Studio CAD software. It is also compatible with DOS and Windows operating systems. The cards can display 256 colors and are easily installed with the included internal cable.

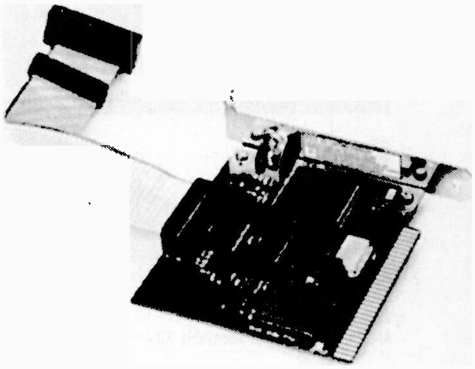
CIRCLE 33 ON FREE INFORMATION CARD

The ME1000 (VGA2NTSC) plug-in card is priced at $525.00

METHODE ELECTRONICS, INC.
dataMate Division
7444 West Wilson Avenue
Chicago, IL 60656
Phone: 708-867-9600
or 1-800-323-6858
Fax: 708-867-3149

DESKTOP MACHINING CENTER

KEPRO CIRCUIT SYSTEMS has introduced the AccuMILL/Plus, a miniature automated X-Y-Z desktop machining center. It can drill, mill, route, and engraves circuit boards. The machine can be fitted with eight different tools. It can drill through both single- and double-sided circuit

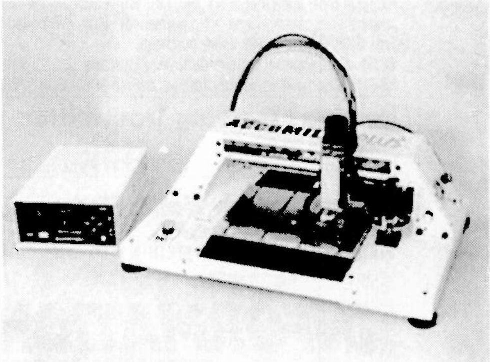
CIRCLE 50 ON FREE INFORMATION CARD

Thousands of Successful Graduates Are Proof That NRI Computer Repair Training Works!

NEW! Course Includes Powerful Multimedia Computer and Windows 95!

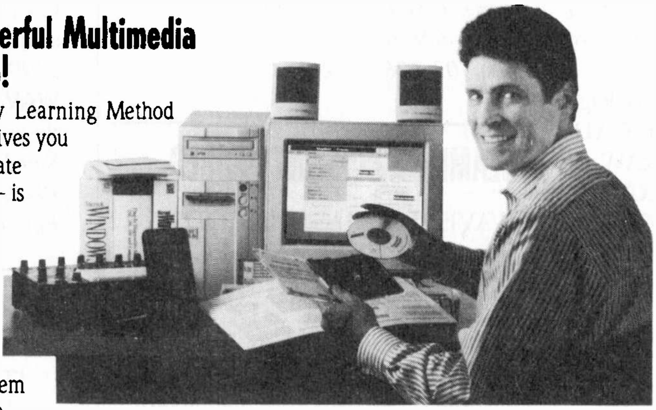

NRI's highly acclaimed Discovery Learning Method – the learn-by-doing approach that gives you first-hand experience with the intricate electronics behind microcomputers – is the most effective way to learn a new skill. You start with NRI's Discovery Lab – a complete breadboarding system that lets you design circuits, modify them, and then diagnose and repair faults.

You then go on to explore and test a state-of-the-art 486DX4/100 MHz computer. This powerful MPC2 system now boasts these exciting features:

- 8 meg RAM
- 128K cache memory
- 540 meg hard drive
- Super VGA color monitor
- 14,400 baud fax/modem
- Quad-speed CD-ROM drive
- 16-bit sound card

> ***"Just two years after finishing the NRI training in microcomputer servicing, I set up my own company. I contract myself out to several computer retailers as their service technician. I have control over my future and total confidence in my business."***
>
> **V. Bender, Calgary, Alberta**

What's more, you perform diagnostic tests using the professional-level Ultra-X R.A.C.E.R. II plug-in card and QuickTech-PRO software, all included in your course.

Plus, NRI training now includes Windows 95, the graphical user interface that's taken the computing world by storm.

Train in the Privacy of Your Own Home

No prior experience necessary – with your personal instructor and NRI's team of technical specialists behind you every step of the way, you learn at your own pace. No classroom pressures, no night school, no need to quit your present job until you're ready to make your move.

Take Advantage of Today's Money-Making Opportunities in Computer Repair

The Department of Labor forecasts over 220,000 jobs for computer service technicians by the year 2005. With the right training and skills, you can cash in on this wide-open opportunity. Only NRI prepares you so thoroughly for a full- or part-time job as a computer service technician – even a business of your own!

> ***"I recommend NRI to anyone looking for a change, success, and real challenge. Hands-on projects allow you to build on what you studied in the lessons – no better way exists to learn how to troubleshoot and service computers!"***
>
> **R. Buda, Tenley Park, IL**

FREE Catalog Gives You the Whole Story—Call or Write NRI Today

If the coupon is missing, write to NRI Schools, 4401 Connecticut Avenue, NW, Washington, DC 20008.

Send today for FREE catalog or call 1-800-321-4634, Ext.1421

NRI Schools
4401 Connecticut Avenue, NW, Washington, DC 20008

For career courses approved under GI Bill, ☐ check here for details.

☑ **Check one FREE career catalog only**
☐ **MICROCOMPUTER SERVICING**
☐ TV/Video/Audio Servicing
☐ Visual Programming in C++
☐ Networking with Windows NT
☐ Basic Electronics
☐ Desktop Publishing with PageMaker
☐ Computer Programming

OR GET YOUR DEGREE!
☐ **NEW!** Associate Degree in Accounting
☐ **NEW!** Associate Degree in Business Management

Name____________________ (please print) Age______

Address____________________

City______________ State_____ Zip________

A Division of The ***McGraw-Hill*** *Companies*

Accredited Member, Distance Education and Training Council

0032-0296

boards to 0.00025-inch resolution for short-production runs or prototyping. It can also mill circuit patterns and machine complex three-dimensional objects from plastic, aluminum, brass, wood, or modeling wax.

The worktable for the AccuMILL/Plus measures 10×14-inches. Its automated tool carousel has eight tool motors. These include one fully automatic, fixed/floating position for accurate milling of irregular surfaces. Adjustments can be made to 0.001 inch. The machine has three different motor speeds: 7000, 28,000, or 50,000 rpm. Three-axis mounting of the tools permits the fabrication of three-dimensional parts.

The AccuMILL software package is compatible with AutoCAD, AutoSKETCH, and CorelDRAW. It can interpret ASCII and HPLG plotter-files running under MS-Windows. The AccuSTEPPER controller is an automated computer interface that controls all three axes.

The AccuMILL/Plus desktop machining center is priced at costs xx.

KEPRO CIRCUIT SYSTEMS, INC.
630 Axminister Drive
Fenton, MO 63026-2992
Phone: 1-800-325-3878
Fax: 314-343-0668

MEMORY TEST PACKAGE

THE SIMCHECK PLUS MEMory tester from Aristo Computers combines the SIMCHECK memory module tester with a 40-bit port adapter in a single package. Together, they can do complete testing of both 30-pin and 72-pin single-in-line memory modules (SIMM). The enhanced memory tester includes a universal power supply that can be plugged directly into the tester or into the 40-bit port.

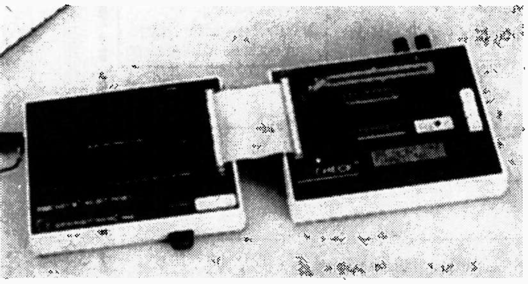

CIRCLE 31 ON FREE INFORMATION CARD

The basic SIMCHECK memory tester tests all standard SIMM and single-in-line (SIP) memory modules with eight or nine bits of 64- and 256-kilobit and 1-, 4-, and 16-megabit memory. The 40-bit port enhances the tester's ability to test 72-pin SIMM modules that have a data bus with up to 40 bits and a maximum memory capacity of 128 megabytes.

SIMCHECK Plus is priced at $1490.00

ARISTO COMPUTERS INC.
6700 SW 105th Avenue, Suite 300
Beaverton, OR 97008
Phone: 1-800-3-ARISTO
or 503-626-6333
Fax: 503-626-6492

CLAMP-ON CURRENT PROBES

WAVETEK HAS INTRODUCed its CT235, CT237, and CT238 AC/DC clamp-on current probes that will measure AC or DC current noninvasively from 5 milliamperes to 1000 amperes peak with 1% accuracy. The broad AC bandwidth and 1% accuracy of the probes permit the precise measurement of complex waveforms distorted by harmonics or power-line anomalies.

The current probes can be used with different test instruments including analog or digital multimeters, dataloggers, chart recorders, and power analyzers. Current measurements are obtained with Hall-effect sensing, and the zero-adjust feature cancels residual magnetic fields, so DC measurement accuracy is unaffected.

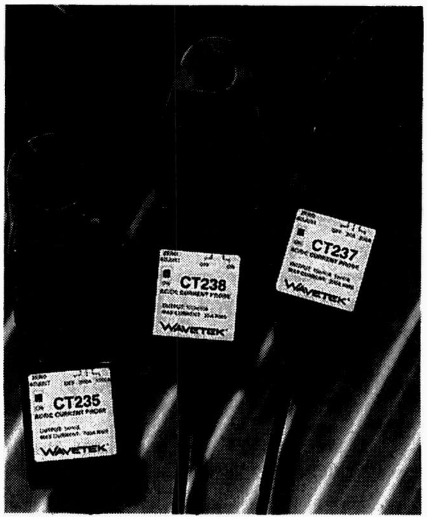

CIRCLE 46 ON FREE INFORMATION CARD

Applications for the probes include the analysis of current waveforms for harmonics, the testing of banks of batteries, the analysis of inverter power output, the recording of UPS charge and discharge rates, and the diagnosis of automotive electrical systems. The probes have IEC1010-1, Category III, 300-volt, pollution degree 2 safety approval.

The pricing on the probes is: CT235 (1000-ampere)—$179.00; CT237 (200-ampere)—$249.00, and the CT238 (20-ampere)—$299.00

WAVETEK CORPORATION
9045 Balboa Avenue
San Diego, CA 92123
Phone: 619-279-2200
Fax: 619-565-9558

LOW-COST PIC PROGRAMMER

ITU TECHNOLOGIES MOdel PIC-1 PIC programmer will program the Microchip family of PIC16C6, 16C7 and 16C8 microcontrollers in conjunction with an IBM or compatible personal computer host. An 18-pin socket is provided on board for programming 18-pin PIC devices; and 28/40-pin adapters are available from the manufacturer. An expansion header is provided for adapters and in-system programming of PICs with onchip EEPROM.

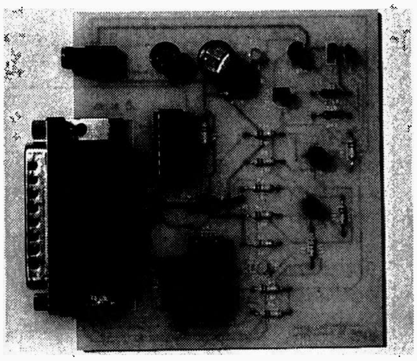

CIRCLE 52 ON FREE INFORMATION CARD

The Model PIC-1 includes all accessories needed to do microcontroller programming. It is recommend for students, hobbyists, and novice programmers. It is supplied with a cable for interfacing to the personal computer's parallel port, a power supply, programming software for the computer, and a user's manual.

The PIC-1 is available as a kit or fully assembled and tested. The kit, including a PC board and construction manual, is priced at $29.00 The assembled programmer is priced at $49.00.

ITU TECHNOLOGIES
3477 Westport Court
Cincinnati, OH 45248-3026

DIRECT DIGITAL SYNTHESIZER MODULE

THE MODEL DDS4M DIRECT digital synthesizer (DDS) module from Novatech Instruments generates a precise sinewave and an accurate TTL clock signal. It is packaged on a 3.5×4.5-inch circuit board. The output frequency is programmable from 1 Hz up to a maximum of 34 MHz in steps as small as 0.02 Hz.

The typical phase noise for the module is −i90 dBc at 1-kHz offset from carrier. The desired frequency can be selected by setting a 31-bit binary number, either manually with DIP switches or remotely with HCMOS-compatible parallel input lines. The DDS4m contains a quartz crystal oscillator that provides stability of 10-ppm/year.

The DDS4m direct digital synthesizer module is priced at $395.00

NOVATECH INSTRUMENTS, INC.
1530 Eastlake Avenue East, No. 303
Seattle, WA 98102
Phone: 206-322-1562
Fax: 206-328-6904

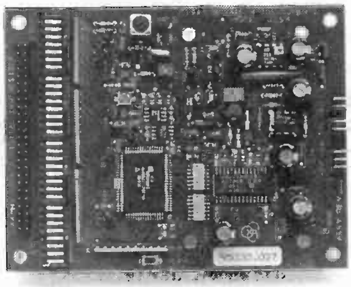

CIRCLE 40 ON FREE INFORMATION CARD

NEGATIVE-VOLTAGE CIRCUIT BREAKERS

THE UCC3913 FAMILY OF BiCMOS electronic circuit breakers from Unitrode is intended for negative power supplies. The circuit breakers are intended for mainframe computer systems, telecommunications equipment, fault-tolerant and point-of-sale computers, banking ATMs, mass data storage equipment, and industrial-control systems.

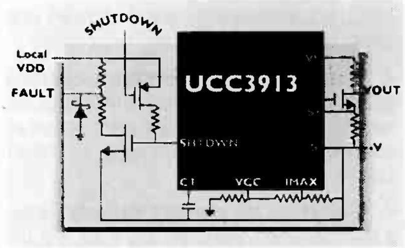

CIRCLE 58 ON FREE INFORMATION CARD

According to Unitrode, the electronic circuit breakers react faster and have more accurate thresholds than the large bus fuses it is designed to replace. Their protection levels

TURN YOUR PC INTO A DIGITAL STORAGE OSCILLOSCOPE

mission technology, Inc.'s DS Scope now sold by Amaze Electronics Corp.

FEATURES:
1. Dual channel; external triggered
2. Digital storage; MS Windows graphical user interface
3. Connects to PC parallel port
4. Programmable sampling intervals from 1 sample/8hrs to 20 million samples/1 second
5. Programmable gain from 5v/div to 5mv/div
6. Integrated 16 bit output

$459 + shpg. & hndlg.

For industrial, hobbyist, educational, automotive, audio, and medical uses

FINALLY! A PLCC prototype adapter that fits on a single breadboard strip!

PLCC socket mounted on solder masked PCB with .25" gold plated leads

Item	Price
28 Pin PLCC	$ 6.95 ea
44 Pin PLCC	$ 9.49 ea
52 Pin PLCC	$13.49 ea
68 Pin PLCC	$14.95 ea
84 Pin PLCC	$18.95 ea

plus shipping and handling

Streamlined, patent-pending design for breadboarding, wire wrap, and design prototyping

AMAZE ELECTRONICS CORPORATION
10395 Mary Avenue, Cupertino CA 95014 - VISA/MC/Check
Phone: 408-996-8943 Fax: 408-996-0366

CIRCLE 13 ON FREE INFORMATION CARD

Tube & General Purpose Projects

NEW

POTS & SOCKETS ALSO AVAILABLE

LC SERIES
0.080" ANODIZED ALUMINUM
TOP AND BOTTOM PANELS ARE FIELD REMOVABLE
FRONT, REAR 0.063" GOLD ANODIZED; OTHER PARTS BLACK ANODIZED VENTED CAGES AND OAK END PANELS AVAILABLE.

Audio Transformers for Tube Construction

Inputs – Line/Mic

MODEL	PRI Ω	SEC Ω	CASE STYLE	USE	PRICE $ A	B
AT-1	150	60k	A	MIC INPUT	40.50	42.50
AT-2	15k	15k	A	LINE INPUT	40.50	42.50
AT-3	15k	600	B	LINE OUTPUT	46.75	48.50

Tube Outputs

MODEL	PRI Ω	SEC Ω	CASE STYLE	USE	PRICE $ A	B
OT-1	5k	8	B	EL84 3 WATT	9.00	11.00
OT-2	6k	4.8.16	B	EL84 10 WATT	12.00	16.00
OT-3	7k	4.8.16	B	EL34 20 WATT	18.00	24.00

ULTRA LINEAR DESIGN

OUTPUT TRANSFORMERS DESIGNED FOR SIMPLE OUTPUT STAGES AND FEATURE FREQUENCY RESPONSE ±2 dB 40-15k Hz

Power Supply

MODEL	SECONDARY #1	#2	#3	CASE STYLE	USE	PRICE $ A	B
PT-1	125V/ 25mA	6V3/ 3A		B	PRE AMP	9.00	11.00
PT-2	250-0-300 40mA	5 2A	6V3 2A	B	5W AMP	18.00	24.00
PT-3	300-0-300 70mA	5 2A	6V3 2A	B	10W AMP	20.00	26.00
PT-4	400-0-400 180mA	5 3A	6V3 4A	B	20W AMP	24.00	30.00

MENTION AD AND TAKE 5% OFF YOUR FIRST ORDER!

FOR MORE INFO REQUEST CATALOG TB-1
WRITE OR FAX FOR OTHER CATALOGS;
AUDIO SOLUTIONS, CONSTRUCTOR'S HARDWARE, AUDIO SIPS and AUDIO TRANSFORMERS

LC Series

MODEL	W x D x H (mm)	PRICE $ A	B
LC-1	100 x 100 x 50	45.00	47.50
LC-2	150 x 100 x 50	50.00	52.50
LC-3	200 x 100 x 50	55.00	57.50
LC-4	300 x 100 x 50	60.00	62.50
LC-5	150 x 150 x 75	50.00	54.00
LC-6	200 x 150 x 75	55.00	59.00
LC-7	300 x 150 x 75	60.00	64.00
LC-8	400 x 150 x 75	65.00	69.00
LC-9	200 x 200 x 75	60.00	64.00
LC-10	250 x 200 x 75	65.00	69.00
LC-11	300 x 200 x 75	70.00	74.00
LC-12	425 x 200 x 75	75.00	79.00
LC-13	250 x 250 x 100	70.00	76.00
LC-14	300 x 250 x 100	75.00	81.00
LC-15	350 x 200 x 100	60.00	86.00
LC-16	425 x 250 x 100	85.00	91.00

ALL HARDWARE SUPPLIED.

NEW LARGER SIZES

METAL CABINET
PANELS ARE .063" ALUMINUM;
BLACK OR GOLD FRONT & REAR PANELS
AVAILABLE FOR ADDITIONAL $5.00

Mini Project Boxes

MODEL	W x D x H (IN)	PRICE $ A	B
MPB-1	1 x 2 x 1	1.95	2.25
MPB-2	1 x 4 x 1	2.40	2.75
MPB-3	1 x 6 x 1	2.85	3.25
MPB-4	1.5 x 2 x 1.5	2.06	2.25
MPB-5	1.5 x 4 x 1.5	2.55	2.85
MPB-6	1.5 x 6 x 1.5	3.00	3.60
MPB-7	1.5 x 2 x 3	2.25	2.55
MPB-8	1.5 x 4 x 3	2.75	3.15
MPB-9	1.5 x 6 x 3	4.00	4.50
MPB-10	1.5 x 2 x 5	2.50	3.00
MPB-11	1.5 x 4 x 5	2.95	3.55
MPB-12	1.5 x 6 x 5	4.25	5.05
MPB-13	2 x 2 x 2	2.20	2.60
MPB-14	2 x 4 x 2	2.65	3.15
MPB-15	2 x 6 x 2	4.40	5.00
MPB-16	2 x 8 x 2	6.65	7.35
MPB-17	3 x 2 x 3	2.45	2.95
MPB-18	3 x 4 x 3	2.90	3.50
MPB-19	3 x 6 x 3	4.20	4.90
MPB-20	3 x 8 x 3	4.65	5.45
MPB-21	4 x 6 x 3	4.55	5.15
MPB-22	4 x 10 x 3	4.75	5.45
MPB-23	4 x 12 x 3	5.70	6.30
MPB-24	4 x 14 x 3	6.10	6.80

PERFORATED OR PLEXIGLASS PANELS ALSO AVAILABLE

GREAT LITTLE BOXES
FOR LOW COST CONSTRUCTION
DESIGNED TO GET THE JOB DONE
WITHOUT THE FRILLS!

NEW

MINI PROJECT BOX
NOT SPECIFICALLY DESIGNED FOR TUBE CONSTRUCTION, BUT THEY'RE SUPER FOR SMALL PROJECTS.
0.40" BARE ALUMINUM PANELS

Metal Cabinets

MODEL	W x D x H (IN)	PRICE $ A	B
MC-1A	4 x 3 x 2	12.00	14.00
MC-2A	6 x 3 x 2	14.00	16.00
MC-3A	8 x 3 x 2	16.00	18.00
MC-4A	4 x 5 x 3	14.00	16.00
MC-5A	6 x 5 x 3	16.00	18.00
MC-6A	8 x 5 x 3	18.00	20.00
MC-7A	4 x 7 x 4	16.00	18.00
MC-8A	6 x 7 x 4	18.00	20.00
MC-9A	8 x 7 x 4	20.00	22.00
*MC-10A	8.5 x 7 x 1.75	32.00	34.50
*MC-11A	8.5 x 10 x 1.75	34.00	36.50
*MC-12A	8.5 x 14 x 1.75	36.00	38.50
*MC-13A	8.5 x 7 x 3.5	33.50	36.00
*MC-14A	8.5 x 10 x 3.5	35.50	38.00
*MC-15A	8.5 x 14 x 3.5	37.50	40.00
*MC-16A	17 x 7 x 1.75	34.50	38.00
*MC-17A	17 x 10 x 1.75	36.75	40.25
*MC-18A	17 x 14 x 1.75	42.75	46.25
*MC-19A	17 x 7 x 3.5	36.75	40.25
*MC-20A	17 x 10 x 3.5	39.25	42.75
*MC-21A	17 x 14 x 3.5	45.25	48.75
EXTRA PAIR FRONT/REAR PANELS		5.00	7.50

FRONT/REAR PANELS CLEAR FINISH STANDARD

ORDER TODAY! ($30.00 MINIMUM ORDER)
ALL PRICES INCLUDE WORLDWIDE SHIPPING! **A** IS USA, CANADA & MEXICO; **B** IS REST OF WORLD

ORDERS (800) 634-3457 ■ FAX (800) 551-2749
OFFICE (702) 565-3400 ■ FAX (702) 565-4828
SESCOM, INC. 2100 WARD DRIVE, HENDERSON, NV 89015 U.S.A.
SESCOM INC is not responsible for inadvertent typographical errors and prices and specifications are subject to change without notice

CIRCLE 12 ON FREE INFORMATION CARD

TIPS FOR MAIL ORDER PURCHASE

It is impossible for us to verify the claims of advertisers, including but not limited to product availability, credibility, reliability and existence of warranties. The following information is provided as a service for your protection. It is not intended to constitute legal advice and readers are advised to obtain independent advice on how to best protect their own interests based upon their individual circumstances and jurisdictions.

1. **Confirm price and merchandise information** with the seller, including brand, model, color or finish, accessories and rebates included in the price.

2. **Understand the seller's return and/or refund policy**, including the allowable return period, who pays the postage for returned merchandise and whether there is any "restocking" or "return" charge.

3. **Understand the product's warranty.** Is there a manufacturer's warranty, and if so, is it for a U.S. or foreign manufacturer? Note that many manufacturers assert that, even if the product comes with a U.S. manufacturer's warranty, if you purchase from an unauthorized dealer, you are not covered by the manufacturer's warranty. If in doubt, contact the manufacturer directly. In addition to, or instead of the manufacturer's warranty, the seller may offer its own warranty. In either case, what is covered by warranty, how long is the warranty period, where will the product be serviced, is there a charge for service, what do you have to do to obtain service and will the product be repaired or replaced? You may want to receive a copy of the written warranty before placing your order.

4. **Keep a copy of all transactions**, including but not limited to cancelled check, receipt and correspondence. For phone orders, make a note of the order including merchandise ordered, price, order date, expected delivery date and salesperson's name.

5. **If the merchandise is not shipped within the promised time**, or if no time was promised, within 30 days of receipt of the order, you generally have the right to cancel the order and get a refund.

6. **Merchandise substitution** without your express prior consent is generally not allowed.

7. **If you have a problem with your order or the merchandise**, write a letter to the seller with all the pertinent information and keep a copy.

8. **If you are unable to obtain satisfaction from the seller**, contact the consumer protection agency in the seller's state and your local Post Office.

If, after following the guidelines, you experience a problem with a mail order advertiser that you are unable to resolve, please let us know. Write to Advertising Department, Gernsback Publications Inc., 500B Bi-County Blvd. Farmingdale, NY 11735.

Be sure to include copies of all correspondence.

exceed those of fuses. Under fault conditions, the UCC3913 devices quickly limit overcurrent, preventing damage to host circuitry. They open the circuit automatically and reset with a 2% duty cycle to limit power dissipation during extended-duration faults.

An automatic reset feature reduces field maintenance. The auto-reset feature also prevents overheating that could lead to circuit failure. The UCC3913 devices manage power by permitting full logic control of each module's power. Their programmable start delays are especially important in "hot-swap" situations such as when power supply modules are being replaced in fault-tolerant computer systems.

The electronic circuit breakers keep the power off, then "soft-start" the module in hot-swap applications. That controlled start prevents the module from excessively loading the bus as it powers up, preventing power and data glitches. The UCC3913 family is offered in eight-pin SOIC and DIP packages in three temperature ranges.

The UCC3913 is priced at $1.40 in 1000-piece quantities.EP

UNITRODE CORPORATION
7 Continental Blvd.
Merrimack, NH 03056-0399
Phone: 603-424-2410
Fax: 603-424-3460

INSULATED-GATE BIPOLAR TRANSISTORS

A NEW SERIES OF 600- AND 1200-volt insulated-gate bipolar transistors (IGBTs) for high-voltage applications has been announced by Motorola. The high-voltage IGBTs are optimized to operate efficiently from full-wave rectified 230- or 460-volt AC with minimum conduction and turn-off losses. The IGBTs will withstand a short circuit for a minimum duration of 10 microseconds.

The IGBTs are suitable as inverters for motor drives. They perform the power-switching function that converts rectified DC line voltage to pulses that permit the control of various kinds of motors. IGBTs can also function in uninterruptible power supplies and conventional switching power supplies operating directly off line at moderate frequencies.

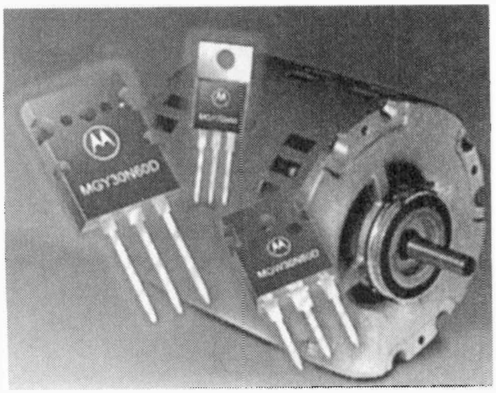

CIRCLE 41 ON FREE INFORMATION CARD

For applications where the stored energy of an inductive load is returned back to the power source through a diode, IGBTs and fast-recovery diodes have been co-packaged. The fast-recovery diode is connected anti-parallel to the IGBT to allow reverse conduction. That permits the return of energy stored in an inductive load to the power source in bridge circuits.

The reverse characteristics of the anti-parallel diodes have been designed to achieve "soft recovery" to minimize transients and EMI. The co-packaged diode minimizes transient voltages that are generated by stray circuit inductance. For applications where the anti-parallel diode is not needed, IGBTs without the diode are available at a slightly lower price.

The high-voltage IGBTs are priced in the range of $4.50 to $6.00 each in 100 to 999 quantity.

MOTOROLA, INC.
MD Z-301
5005 East McDowell Road
Phoenix, AZ 85018
Phone: 602-244-4911
Fax: 602-944-4015

PORTABLE PROTOCOL ANALYZER

THE MODEL 904 PC COMscope from Telebyte Technology is a portable protocol analyzer that interfaces with a host computer to access its resources. It can function as a high-performance communications data line monitor, emulator, and bit error

(Continued on page 109)

NEW LITERATURE

USE THE FREE INFORMATION COUPON FOR FAST RESPONSE

Prototype to Production Catalog No. 955

Digi-Key Corporation
701 Brooks Avenue South
P. O. Box 677
Thief River Falls, MN 56701-0677
Phone: 1-800-344-4539
free

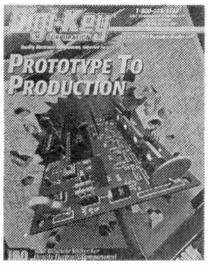

CIRCLE 35 ON FREE INFORMATION CARD

Mail-order distributor Digi-Key offers more than 1400 new products from 15 different manufacturers in its latest catalog. Newly introduced products include ribbon-cable and D-type subminiature connectors from AMP, Inc., power supplies from Power-One, microcontrollers from Microchip, and development systems from Picmaster.

A Quick Index on the first page will speed up finding references to the products being offered. This is supplemented by detailed manufacturer and product indexes. Digi-Key sells just about everything a manufacturer or serious electronic hobbyist needs and its catalog is a valuable reference for the professional designer as well as all advanced hobbyists.

The catalog is arranged alphabetically by manufacturer with alphabetical tabs on the outside edges of the pages for quick reference. A comprehensive index both by product and by manufacturer makes it easy to find what you are looking for in this huge 368-page compendium of electronic parts.

Digi-Key's product line card includes all the basic passive electronic components, discrete and integrated semiconductor devices, and connectors and cable assemblies and much more. The company offers hand tools, soldering materials, power supplies, and a wide variety of test equipment in addition to all the active and passive component lines that it lists.

Sendmail: Theory and Practice

by Frederick M. Avolio and Paul A. Vixie.
Digital Press
313 Washington Street
Newton, MA 02158-1626
Phone: 617-928-2500
Fax: 617-928-2620
$29.95

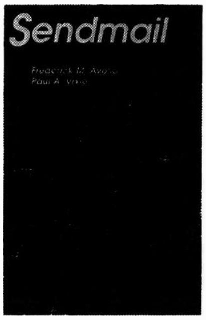

CIRCLE 38 ON FREE INFORMATION CARD

This book explains the history, architecture, configuration, and maintenance of Sendmail, a standard mail deliverer on Berkeley-Unix systems. Widely used on the Internet, many consider Sendmail impossible to understand. The authors have tried to correct that.

Avolio and Vixie explain how and why Sendmail works and offer practical advice on designing and maintaining an electronic mail system. There are "cookbook recipes" and simplified explanations of how to manage a mail system. It compares E-mail and Sendmail and defines key terms. This book is great for computer system managers and readers who want to

BREAST CANCER BEGINS EVEN SMALLER THAN THIS. THAT'S WHY YOU NEED A YEARLY MAMMOGRAM, ESPECIALLY AS YOU GET OLDER. MAMMOGRAMS CAN DETECT LUMPS TOO SMALL FOR YOU TO FEEL AND EARLY DETECTION MAY SAVE YOUR LIFE, SO CALL 1-800-ACS-2345.

GET A MAMMOGRAM. EARLY DETECTION IS THE BEST PROTECTION.

American Cancer Society

Become a computer repair expert!

Career-level home study course. Learn all about PC repairs, troubleshooting, installation, upgrading, servicing, and more.

Open your own business; become a more valuable employee. **Free literature.**

Call now: 800-223-4542

Name ____________ Age ____
Address ____________
City/State ____________
Zip ________ Phone (   ) ________

The School of PC Repair
PCDI, 6065 Roswell Road
Dept. JJA351, Atlanta, Georgia 30328

Learn VCR repair!

Professional-level home study course. Master easy-to-learn, high-profit repairs without investing in costly high-tech instruments. **Send or call for free career literature.**

VHS

Call now: 800-223-4542

Name ____________ Age ____
Address ____________
City/State ____________
Zip ________ Phone (   ) ________

The School of VCR Repair, Dept. VRA351
PCDI, 6065 Roswell Rd., Atlanta, GA 30328

Whaddya Say To A Guy Who's Had The Same Job For 50 Years, Has Never Called In Sick Or Showed Up Late, Never Taken A Vacation Or A Holiday, Never Asked For A Raise Or Griped About His Bonus And, Believe It Or Not, Has No Plans For Retirement?

Thanks.

Remember - only you can prevent forest fires.

set up and maintain a mail system for computers.

Hipercomm Databook (BR1334/D)

Motorola Inc.
Literature Distribution Center
P. O. Box 20924
Phoenix, AZ 85063
Phone: 1-800-441-2447
Fax: 602-994-6430
free

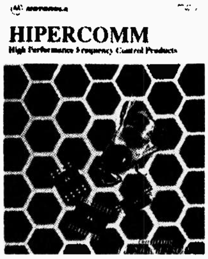

CIRCLE 44 ON FREE INFORMATION CARD

This is a revised edition of Motorola's Hipercomm Databook. It contains device specifications in the form of data sheets and applications information for Motorola's frequency-control products. It contains technical data on a broad line of integrated circuits for phase-locked loop circuits. An introductory section simplifies the selection of the right components for a given set of applications requirements.

Motorola's line of low-power prescalers includes products with more than 40 different combinations of frequency, divide ratio, and power to meet the needs of most designs. Information is also included on Motorola's integrated, single-chip synthesizers that conserve power, making them suitable for handheld, battery-operated communications products.

Low Power Communications

Volume 3: QRP Hardware
by Richard Arland, K7YHA.
Tiare Publications
P. O. Box 493
Lake Geneva, WI 53147
Phone: 1-800-420-0579
$14.95

This is the third volume of three in a series on QRP, an international radio Q signal that means "reduce your power". Previous titles were "QRP Basics" and "Advanced QRP Operating." This book on low-power communications examines the new and used equipment available for the radio amateur who is planning to upgrade a QRP station that might be communicating with microwatts of power. It includes tips on buying and trading used equipment, and discusses operating accessories of value to the QRP operator.

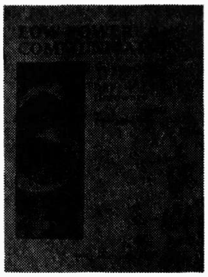

CIRCLE 36 ON FREE INFORMATION CARD

Arland's book includes a chapter on the planning and erection of antenna towers and another on how to start and run a QRP club. The addresses of QRP clubs and suppliers and manufacturers of QRP equipment are listed in the appendix.

ESD Control Products.

Charleswater
90 Hudson Road
Canton, MA 02021-1407
Phone: 617-821-8370
Fax: 617-575-0172
free

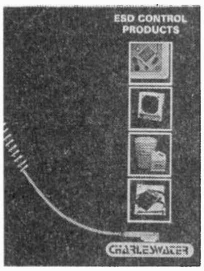

CIRCLE 48 ON FREE INFORMATION CARD

This is the latest catalog from Charleswater on its products for the control of electrostatic discharge (ESD). Many new ESD control products and training aids are included. It also provides up-to-date information on existing product categories.

Charleswater's product line does include all of the standard accessories necessary to EDS-proof a designated ESD-protected workstation: conductive workplace mats, grounded wrist straps, conductive floor mats, and instruments for testing the integrity of the ground. Also included are conductive tote boxes, protective bags and packaging for vulnerable components and circuits, air ionizers, and conductive sprays.

This catalog also includes a basic tutorial on electrostatic discharge, explaining what it is and how it is controlled by grounding, shielding, and moving-air neutralization.

HF Filter Design and Computer Simulation

by Randall W. Rhea.
McGraw-Hill Book Company
11 West 19th Street
New York, NY 10011
Phone: 1-800-2-MCGRAW.

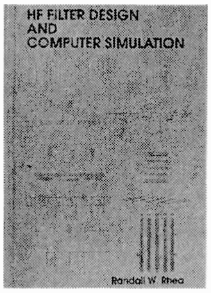

CIRCLE 54 ON FREE INFORMATION CARD

This book is a complete design guide for printed and LC filters for RF and microwave applications. It includes easy to understand discussions of high-frequency filter theory and accurate computer simulation models for many different filters. Rhea's book includes a review of the classic design formulas as well as the more recent filter design concepts developments.

The first chapters are devoted to fundamental concepts, with greater emphasis given to lumped-elements than distributed elements. Those chapters are followed by a review of available computer-aided filter-design techniques, including both simulation and synthesis.

The author goes on to discuss distributed low-pass, bandpass, high-pass, and bandstop filters, and he explains how to select the appropriate design for specific applications.

Expanding Your BBS

Volume 3: QRP Hardware
by Richard Arland, K7YHA.
Tiare Publications
P. O. Box 493
Lake Geneva, WI 53147
Phone: 1-800-420-0579
$14.95

This book and an included CD-ROM gives bulletin board system (BBS) administrators all the information they need to maintain an active, attractive bulletin board. There are an estimated 10-million BBS users around the world whose interests are as diverse as scuba diving to gourmet cooking.

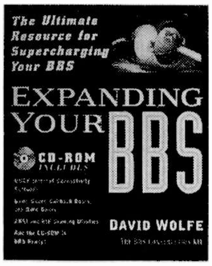

CIRCLE 32 ON FREE NFORMATION CARD

Wolfe's book offers the software and expert advise needed to add the most wanted features to your BBS. For example, Wolfe explains how to add UUCP Internet connectivity, fax capability, and QWK networking and mail. Wolfe also explains how to add improved graphics and speed, and multimedia applications. The topics covered include the integration of high-speed modems, digiboards, intelligent I/O boards, and telephone caller ID.

The companion BBS-ready CD-ROM contains game doors such as "Tradewars 2002, Baron Realms Elite, Land of Devastation, Battle Grid, Murder Motel," and "Yankee Trader." Call back doors increase BBS security and automate the registration process, while Waffle and Waffle utilities are for Internet UUCP connectivity.

The CD-ROM also includes upload checkers that scan for viruses, purge old files, insert BBS ads, and check file integrity. Batch file power is increased with four DOS utilities. ANSI and RIP screen-drawing utilities include "The Draw, Dead Paint," and "Tombstone Artist." Other BBS utilities included are Fossil drivers, external protocols, QWK mail doors, hard-to-find compression utilities, and port and modem diagnostics.

Zworykin, Pioneer of Television

by Albert Abramson.
University of Illinois Press
1325 South Oak Street
Champaign, IL 61820
$36.95

Vladimir Kosma Zworykin, the Russian-born American scientist, made a significant contribution to the electronically-scanned television system that we know today with his invention

(Continued on page 109)

FREE

VOID AFTER JULY 1, 1996

FREE READERS SERVICE: To receive more information about products and services mentioned in this issue, please circle the applicable Reader Service Number, and mail today!

EEH	11	21	31	41	51
1996	12	22	32	42	52
	13	23	33	43	53
	14	24	34	44	54
5	15	15	35	45	55
6	16	26	36	46	56
7	17	27	37	47	57
8	18	28	38	48	58
9	19	29	39	49	59
10	20	30	40	50	60

Additional Coupons on pages 108 and 111.

1 ☐ Please send me 12 issues of POPULAR ELECTRONICS® for $18.95 and bill me. (Canada $25.63 U.S. FUNDS — includes G.S.T.)
2 ☐ Please send me 12 issues of ELECTRONICS NOW® for $17.97 and bill me. (Canada $25.65 U.S. FUNDS — includes G.S.T.)

Mail to: ELECTRONICS EXPERIMENTER'S HANDBOOK, Reader Service Management Dept.,
PO Box 5192, Pittsfield, MA 01203-9989

Name ________________ Please Print
Company Name (if applicable) ________________
Address ________________
City ________ State ________ Zip +4 ________

Allow 6-8 weeks for delivery of first issue. OFFER VALID U.S. AND CANADA ONLY.

FREE

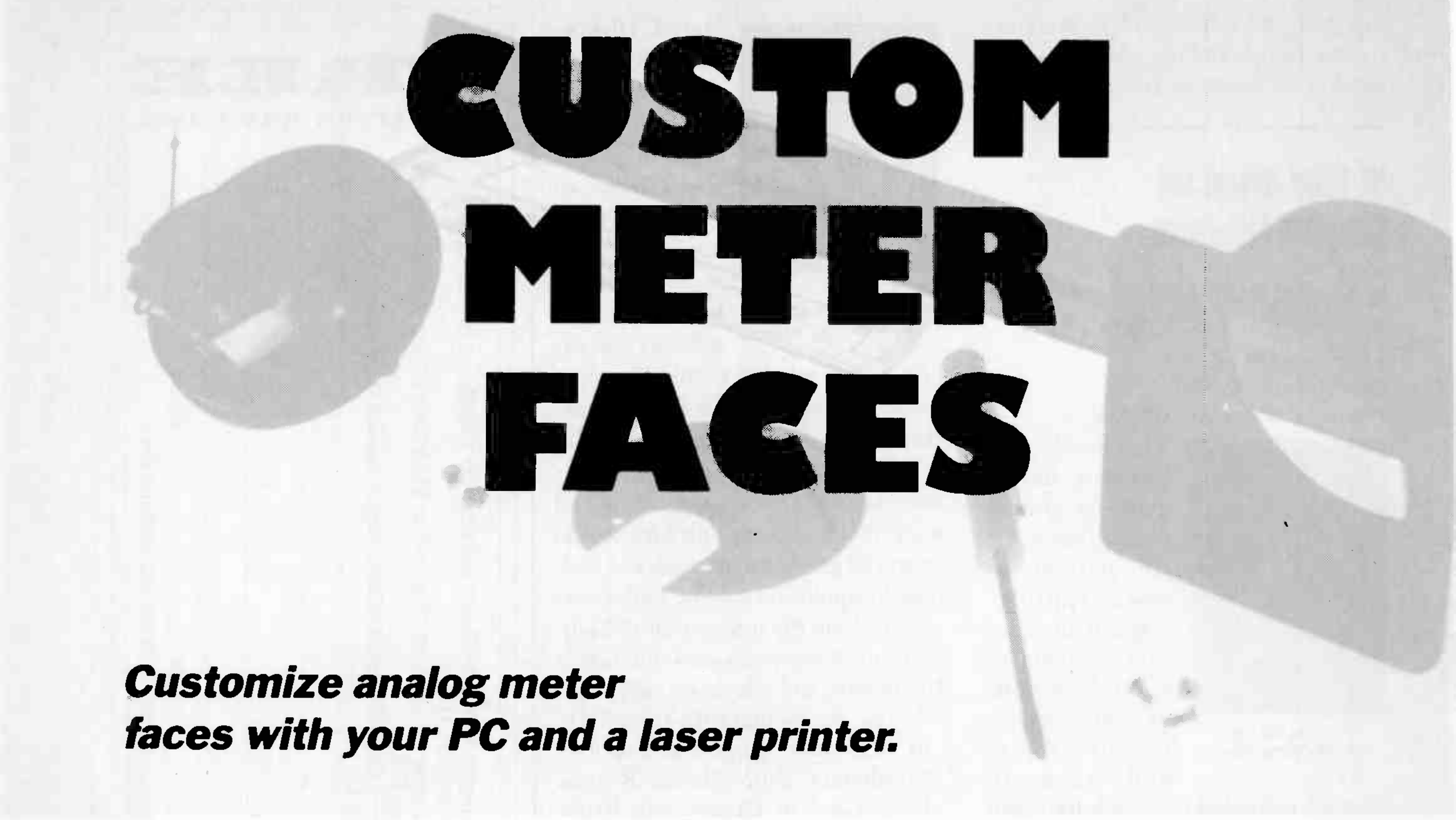

CUSTOM METER FACES

Customize analog meter faces with your PC and a laser printer.

SKIP WITHROW, WB0BBE

EVEN IN TODAY'S WORLD OF DIGITAL electronics there are still times when an analog meter is the better choice for displaying voltage, current, or other signal levels. With the use of a personal computer and a laser printer, this article shows how you can customize analog meters from your junk box or local surplus dealer to display almost any scale and units. Gone are the days of trying to draw meter scales by hand; you can make your computer and laser printer do the work for you.

Any laser printer that understands HPGL (Hewlett-Packard Graphics Language) commands is compatible with the techniques presented here. The BASIC programming language is used to assemble the commands which are then sent to the printer. The printer prints the new scale on a whole-page laser label and then the label is applied to the old meter. Figure 1 shows a meter, purchased for $2 at a surplus electronics outlet, before and after modification.

Finding the right meter

The first step to successfully making a newly labeled meter is to start with the right meter. Although many meters can be found for very reasonable prices in the surplus market, not every meter is a candidate for a new scale. Several specifications, such as the full-scale current and the meter movement's resistance, are helpful in determining whether the meter is a suitable choice.

To change the scale of a meter, the face must be accessible. Some inexpensive meter movements have cases that can't be opened, and these should be avoided. Generally the cover of a meter face can either be removed by simply prying it off, or by removing several small screws on the outside of the meter. The second task is to remove the meter face itself. Again, some inexpensive meters might have their faces mounted permanently, but the majority of meters have several small screws that hold the meter face in place. The ability to remove the meter face is essential.

Other things that must be known about the meter before it can be used in a circuit properly are its full-scale current and the resistance of its movement coil. For a new meter, these are easy to determine from the meter's specification sheet. Sometimes you will be lucky enough to find information printed on an old meter face that will help determine the full-scale reading.

If you know nothing about a meter's electrical specifications, they can be determined rather easily. Generally most meters have a full-scale current between 100 microamperes and 1 milliampere. Coil resistance generally varies between about 40 and 1000 ohms. The setup to measure full scale current and coil resistance is simple. The value of the power supply is not critical. A potentiometer between the supply and the unknown meter should be large enough to limit current to about 50 microamperes (about 200K). First set the potentiometer to its maximum value, turn on the power supply, then decrease the resistance until the meter reads full scale. Measure

Just like these Fully Trained Electronics Professionals

"Thanks to CIE I have tripled my previous salary, and I am now in a challenging and rewarding new field where only the sky is the limit."
Daniel Wade Reynolds
Industrial Electrician
Ore-Ida Foods

"CIE was recommended to me by my boss. It was appealing since I could study at my own pace at home and during business travel."
Dan Parks
Marketing Manager/Consumer Products
Analog Devices, Inc.

"I loved the flexibility CIE offered. It was the only way I could continue both school and my demanding job."
Britt A. Hanks
Director of Engineering
Petroleum Helicopters, Inc.

"I liked the way the school was set up with laboratory assignments to enforce conceptual learning. The thing which impressed me the most about CIE's curriculum is the way they show application for all the theory that is presented."
Daniel N. Parkman
Missile Electro-Mechanical Technician
U.S. Air Force

"Completing the course gave me the ability to efficiently troubleshoot modern microprocessor based audio and video systems and enjoy a sense of job security."
Tony Reynolds
Service Manager/Technician
Threshold Audio & Video

Graduate with an Associate Degree from CIE!

CIE INSTITUTE OF ELECTRONICS · 1934 ·

CIE is the best educational value you can receive if you want to learn about electronics, and earn a good income with that knowledge. CIE's reputation as the world leader in home study electronics is based solely on the success of our graduates. And we've earned our reputation with an unconditional commitment to provide our students with the very best electronics training.

Just ask any of the 150,000-plus graduates of the Cleveland Institute of Electronics who are working in high-paying positions with aerospace, computer, medical, automotive and communications firms throughout the world. They'll tell you success didn't come easy...but it did come...thanks to their CIE training. And today, a career in electronics offers more rewards than ever before.

CIE'S COMMITTED TO BEING THE BEST...IN ONE AREA...ELECTRONICS.
CIE isn't another be-everything-to-everyone school. CIE teaches only one subject and we believe we're the best at what we do. Also, CIE is accredited by the National Home Study Council. And with more than 1,000 graduates each year, we're the largest home study school specializing exclusively in electronics. CIE has been training career-minded students for nearly sixty years and we're the best at our subject... ELECTRONICS... IT'S THE ONLY SUBJECT WE TEACH!

CIE PROVIDES A LEARNING METHOD SO GOOD IT'S PATENTED.
CIE's AUTO-PROGRAMMED® lessons are a proven learning method for building valuable electronics career skills. Each lesson is designed to take you step-by-step and principle-by-principle. And while all of CIE's lessons are designed for independent study, CIE's instructors are personally available to assist you with just a toll free call. The result is practical training... the kind of experience you can put to work in today's marketplace.

LEARN BY DOING...WITH STATE-OF-THE-ART EQUIPMENT AND TRAINING.
CIE pioneered the first Electronics Laboratory Course and the first Microprocessor • Course. Today, no other home study school can match CIE's state-of-the-art equipment and training. And all your laboratory equipment, books and lessons are included in your tuition. It's all yours to use while you study and for on-the-job after you graduate.

PERSONALIZED TRAINING....TO MATCH YOUR BACKGROUND.
While some of our students have a working knowledge of electronics others are just starting out. That's why CIE has developed twelve career courses and an A.A.S. Degree program to choose from. So, even if you're not sure which electronics career is best for you, CIE can get you started with core lessons applicable to all areas in electronics. And every CIE Course earns credit towards the completion of your Associate in Applied Science Degree. So you can work toward your degree in stages or as fast as you wish. In fact, CIE is the only school that actually rewards you for fast study, which can save you money.

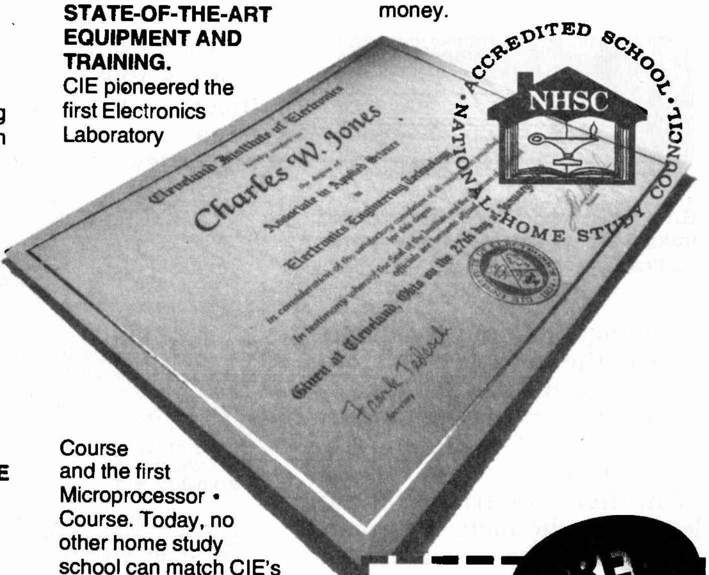
ACCREDITED SCHOOL · NATIONAL HOME STUDY COUNCIL
NHSC
Cleveland Institute of Electronics
Charles W. Jones
Associate in Applied Science
Electronics Engineering Technology
Given at Cleveland, Ohio on the 27th day

Send for CIE's FREE Course Catalog and See How We Can Help Your Career Too!

YES! I want to get started. Send me my CIE course catalog including details about the Associate Degree Program. (For your convenience, CIE will have a representative contact you - there is no obligation.)

Please Print Clearly

Name ____________________

Address ____________________

City ____________________

State _____ Zip __________ Age _____

Phone No. ____________________

Check box for G.I. Bill Benefits.
☐ Veteran
☐ Active Duty

AHV05

Cleveland Institute of Electronics, Inc.
1776 East 17th Street
Cleveland, OH 44114

A School of Thousands.
A Class of One. Since 1934.

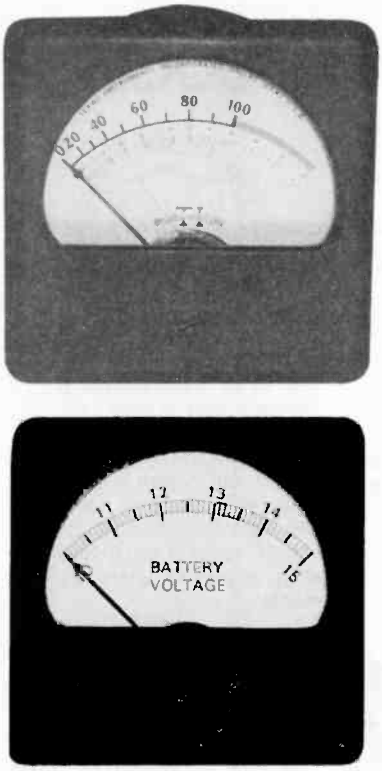

FIG. 1—A VOLTMETER, purchased for $2 surplus, shown before (top) and after modification (bottom).

the voltage of the supply and the value of the variable resistance. The voltage divided by the resistance will then give you the current required to deflect the meter full scale.

Another potentiometer can determine the meter coil resistance. With the meter indicating full-scale current, place a 1K potentiometer across the meter and adjust until the meter reads exactly ½ scale. Remove this potentiometer and measure its value. This will be the value of the coil resistance of the meter. The meter in Fig. 1 had a full-scale current reading of 229 microamperes and a coil resistance of 819 ohms.

Making faces

A protractor is the best tool for measuring the total angle of movement of the meter needle, which generally runs from 90 to 120 degrees. Also measure the distance from the pivot point to the scale of the old meter face and the location of the faceplate mounting holes with respect to the pivot point. Those measurements will be input to the program that draws the new meter face.

The next step in putting a new face on your meter is to carefully remove the old meter face plate. After opening up the meter case it is usually just a matter of removing the small screws holding the face plage on. A set of jeweler's screw drivers can sometimes be helpful for particularly small meters. Once the face plate is loose it can be removed by very carefully sliding it from under the needle of the meter. Be careful not to bend or bump the needle.

Using HPGL

HPGL is a graphics language of short, simple statements that allows lines, arcs, and text to be drawn. Originally developed for plotters, this language has been incorporated into many of today's laser printers. By using just a few of the many HPGL statements, professional looking meter scales can be created with customized features.

The BASIC language can send HPGL commands to the printer for drawing the various features of the scale. Using a computer language to send the commands to the printer is helpful because many commands can be generated using program loops with an incrementing variable. This lets rather short programs generate many different individual graphic elements to compose the new meter face plate.

Listing 1 is the BASIC program that draws the meter face in Fig. 2. Only a few of the many HPGL commands are used in the program. Most are two-letter commands followed by arguments. In Listing 1 many of the arguments are calculated by the BASIC program. The program plots points on a graph that has

LISTING 1

```
REM - PUT THE ORIGIN IN THE MIDDLE OF THE PAGE.
REM - THIS IS THE NEEDLE PIVOT POINT
XORG = 4000
YORG = 5500

REM - CONSTANTS USED FOR PLOTTING
IRAD = 1.4 * 1016
ORAD = 1.5 * 1016
IIRAD = 1.33 * 1016
ANGLE = 90
K = .0174533
MAJORDIV = 5
MINORDIV = 50

REM - CLEAR THE PRINTER
LPRINT CHR$(27); "E"
REM - PUT THE PRINTER INTO HPGL MODE
LPRINT CHR$(27); "%0B"
REM - INITALIZE HPGL MODE
LPRINT "IN;"

REM - SELECT A PEN SO IT WILL PRINT AND SET THE WIDTH
LPRINT "SP1;"
LPRINT "PW0;"

REM - DRAW THE UPER AND LOWER ARCS FOR THE METER SCALE
LPRINT "PA"; XORG - SIN(ANGLE / 2 * K) * ORAD; ","; YORG + COS(ANGLE / 2 * K)
* ORAD; ";PD;"
LPRINT "AA"; XORG; ","; YORG; ","; -ANGLE; ";"
LPRINT "PU;"
LPRINT "PA"; XORG - SIN(ANGLE / 2 * K) * IRAD; ","; YORG + COS(ANGLE / 2 * K)
* IRAD; ";PD;"
LPRINT "AA"; XORG; ","; YORG; ","; -ANGLE; ";"

REM - PUT IN THE 50 TICK MARKS FOR TENTHS OF VOLTS
FOR I = -ANGLE / 2 TO ANGLE / 2 STEP ANGLE / MINORDIV
LPRINT "PU;"
LPRINT "PA"; XORG + SIN(I * K) * IRAD; ","; YORG + COS(I * K) * IRAD; ";"
LPRINT "PD;"
LPRINT "PA"; XORG + SIN(I * K) * ORAD; ","; YORG + COS(I * K) * ORAD; ";"
LPRINT "PU;"
NEXT I

REM - PRINT THE VOLT MARKS USING HEAVIER LINES AND EXTENDING BELOW THE ARC
LPRINT "PW.5;"
FOR I = -ANGLE / 2 TO ANGLE / 2 STEP ANGLE / MAJORDIV
LPRINT "PU;"
LPRINT "PA"; XORG + SIN(I * K) * IIRAD; ","; YORG + COS(I * K) * IIRAD; ";"
LPRINT "PD;"
LPRINT "PA"; XORG + SIN(I * K) * ORAD; ","; YORG + COS(I * K) * ORAD; ";"
LPRINT "PU;"
NEXT I
```

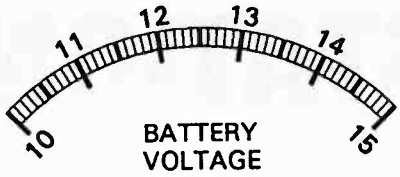

FIG. 2—THIS METER FACE is the result of the BASIC program in Listing 1.

1016 points (pixels) per inch in both the X and Y directions.

Here's a list of the commands that are used and short descriptions:

IN—Initialize
SP—Select Pen
PW—Set Pen Width
PA—Plot Absolute
AA—Arc Absolute
PU—Pen Up
PD—Pen Down
SD—Standard Font Definition
DT—Define Label Terminator
SS—Select Standard Font
LO—Label Origin
LB—Label

The syntax of those commands is explained in the technical reference section for the HPGL language in your printer manual. Since some commands take X and Y coordinate locations, some simple trigonometry is used to calculate the arguments from the index angle in the program of Listing 1.

Developing your own meter face is a simple matter of running the program and printing the resulting graphics on the printer. You probably won't get it right the first time, but making modifications is easy. Writing the program in blocks that add a new graphic after each iteration works best. Printing locating marks on the new face for the mounting holes helps to align the new face properly on the old face plate.

Once everything looks the way you want it to on paper, print the results on a full-page laser label. It's best to apply the new label on the reverse side of the old face plate since it's usually a nice clean surface. Punch pinholes in the location marks to help locate them from the back side of the label. Peel off the backing of the new scale label and place the label face down. Then stick the face plate to the label, lining up the locator holes in the center of the face plate mounting holes. After applying the label use a modeling knife to trim the excess paper from around the meter face plate and mounting holes. Figure 3 shows the new label applied to the old meter.

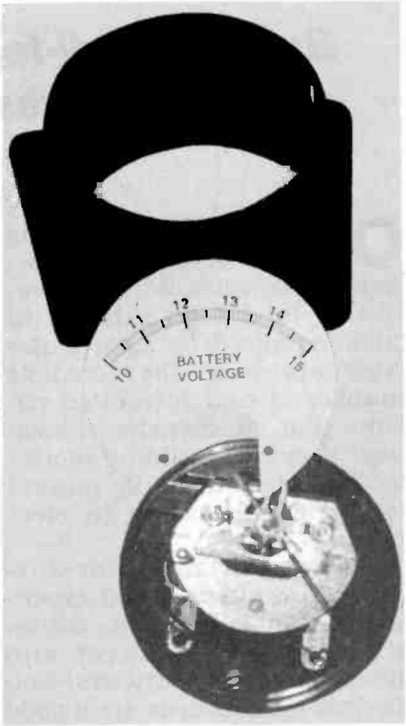

FIG.3—HERE IS THE NEW LABEL pasted onto the surplus meter.

Reassembly finishes the job. Carefully slide the face plate under the meter needle and replace the mounting screws. Next, replace the meter cover.

Creativity is the key to adding an exciting meter display to your next project. Expanded-scale voltmeters are an ideal application for this technique. Power supply current meters are also nice to have. This technique can also create non-linear meter scales such as decibel meters for audio or RF applications. Meter recycling can enhance your projects with just a small investment in time. Ω

```
REM - PRINT THE HALF VOLT MARKS USING A HEAVY LINE BETWEEN THE TWO ARC LINES
LPRINT "PW.5;"
FOR I = -ANGLE / 2 TO ANGLE / 2 STEP ANGLE / MAJORDIV / 2
LPRINT "PU;"
LPRINT "PA"; XORG + SIN(I * K) * IRAD; ","; YORG + COS(I * K) * IRAD; ";"
LPRINT "PD;"
LPRINT "PA"; XORG + SIN(I * K) * ORAD; ","; YORG + COS(I * K) * ORAD; ";"
LPRINT "PU;"
NEXT I

REM - SET UP TO LABEL THE METER AND PRINT "BATTERY VOLTAGE"
LPRINT "SD1,21,2,1,4,10,7,4148;"
LPRINT "DT*,1;"
LPRINT "SS;"
LPRINT "PA"; XORG; ","; YORG + 900; ";"
LPRINT "LO4;"
LPRINT "LBBATTERY*;"
LPRINT "PA"; XORG; ","; YORG + 750; ";"
LPRINT "LBVOLTAGE*;"

REM - PUT THE LOW AND HIGH VALUE BELOW THE SCALE ARC
LPRINT "LO16;"
LPRINT "PA"; XORG + SIN((-ANGLE / 2) * K) * IIRAD; ","; YORG + COS((-ANGLE /
2) * K) * IIRAD; ";"
LPRINT "DI"; COS(ANGLE / 2 * K); ","; SIN(ANGLE / 2 * K); ";"
LPRINT "LB10*;"
LPRINT "PA"; XORG + SIN(ANGLE / 2 * K) * IIRAD; ","; YORG + COS(ANGLE / 2 * K)
* IIRAD; ";"
LPRINT "DI"; COS((-ANGLE / 2) * K); ","; SIN((-ANGLE / 2) * K); ";"
LPRINT "LB15*;"

REM - PUT THE INTERMEDIATE VALUES ABOVE THE SCALE ARC
LPRINT "LO14;"
FOR I = -(ANGLE / 2) + ANGLE / MAJORDIV TO ANGLE / 2 - ANGLE / MAJORDIV STEP
ANGLE / MAJORDIV
LPRINT "PA"; XORG + SIN(I * K) * ORAD; ","; YORG + COS(I * K) * ORAD; ";"
LPRINT "DI"; COS(-I * K); ","; SIN(-I * K); ";"
LPRINT "LB"; INT(I / (ANGLE / MAJORDIV) + 12.5); "*;"
NEXT I

REM - PUT IN THE LOCATION DOTS FOR THE MOUNTING HOLES
LPRINT "PA"; XORG - 508; ","; YORG; ";"
LPRINT "PD;"
LPRINT "PU;"
LPRINT "PA"; XORG + 508; ","; YORG; ";"
LPRINT "PD;"
LPRINT "PU;"

REM - RESET THE PRINTER
LPRINT CHR$(27); "%0A"
LPRINT CHR$(27); "E"
END
```

PROTOTYPING STATION

Build this full-featured breadboard laboratory, and add custom features to suit your own needs.

One of the most gratifying benefits of being an electronics hobbyist is the opportunity to design and build custom projects for a particular need or purpose. The incredible number of new integrated circuits that are introduced each uear provides building blocks for circuits and fertile ground for experimentation in electronics.

One requirement for successful prototyping and experimentation is a quick, convenient way to connect and disconnect components. Solderless breadboards are a good start, but all too often one is likely to end up with the board in the center surrounded by a maze of wires that connect switches, potentiometers, meters, power supplies, and any number of other components that dangle in all directions. The arrangement of components can be both frustrating and irritating. The solution is to combine the breadboard, power supply, and other commonly used parts into a self-contained unit.

Factory-made laboratories like that are available, but they can be quite expensive. The lab described in this article, however, combines economy, expansion, and easy customization—you can use many of the parts you might already have on hand in your junkbox.

Design

The prototyping station base consists of a 20-inch length of 1- by 12-inch pine board that is sanded, stained and finished. The various power supplies are located to the right; the center section is reserved for three breadboards and the main equipment backboard. The left side is left open to accommodate additional modular backboards, or any other peripherals such as keypads or data-entry terminals. The base can be raised to make room for plastic drawers, power transformers, or other equipment. Rubber feet can be used to ensure that the lab is stable on the bench. That basic layout is shown in Fig. 1.

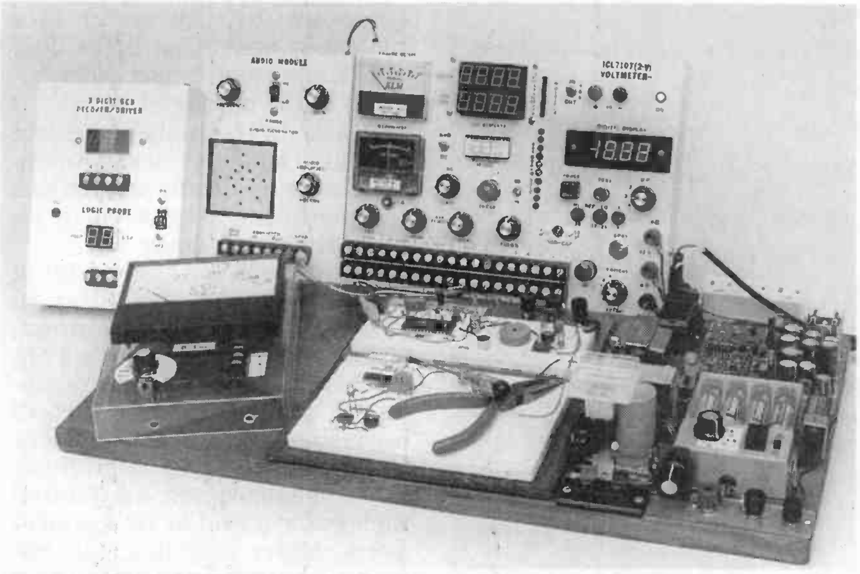

FIG. 1—BASIC PROTOTYPING LAYOUT. The base can be raised to make room for plastic drawers, power transformers, or other equipment.

Power supplies

Since the availability of power is always a requirement for successful prototyping, a variety of different power sources is incorporated on the right-hand section of the station (Fig. 2 shows a close-up view of the power section). The first power supply on the upper left is a surplus Texas Instruments computer board, bought at a clearance sale for $5. With the addition of an LED indicator, it provides a good clean source of ±5 volts and +12 volts. Being a computer-grade supply, the outputs are well-regulated, well-filtered, and can output about 1 ampere each. Power for this supply comes from an off-board wall-outlet transformer, and it is input to the supply via a two-position pushbutton terminal (the kind you might see on the back of a speaker).

Since a standard dry cell is appropriate to power many circuits, the second supply shown on the upper right of Fig. 2 consists of four AA cells and one 9-volt battery. The case from an old nickel-cadmium battery charger holds the AA cells, and a rotary switch and power indicator were added. The battery holder is tapped at each of the four positive battery pads, and each pad is wired to one pole of the rotary switch. The 9-volt battery is wired to a fifth pole on the rotary switch. This allows the selection of 1.5, 3, 4.5, or 6 volts DC from the AA pack, and

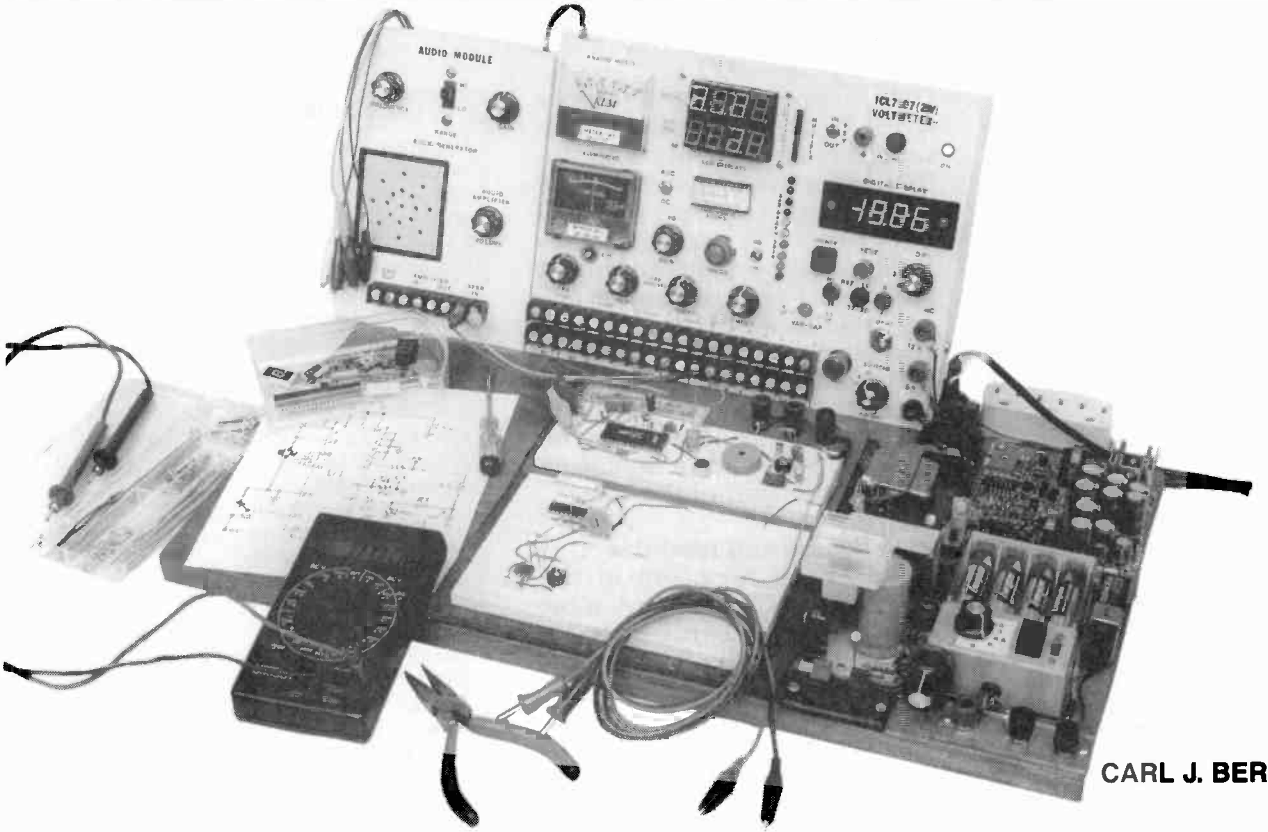

CARL J. BERGQUIST

9 volts from the 9-volt battery. The LED indicator is a 0.3-inch high, seven-segment display, with all segments turned on to form an "8," which is a more noticeable reminder of the "on" status of the battery pack than a single LED. The output from this supply is connected to two of three binding posts for easy access. The third binding post is the negative 5-volt supply from the TI computer supply.

Another voltage source is provided by a +1.2- to 20-volt DC variable power supply that was built from a kit some years ago. The potentiometer next to the battery pack allows for adjustment of the output voltage, which is available at another two-position pushbutton terminal. An analog meter indicates the voltage at the pushbutton terminals.

Next to the variable supply is a 12.6-volt, center-tapped transformer, complete with a switch and indicator LED, that provides 6- and 12-volts AC. An SPST slide switch is connected to one side of the 120-volt AC input to the transformer primary. The outputs are wired to multipurpose posts on the main backboard.

AC power is made available at a triple-outlet AC strip, which gets its power via a heavy-duty line cord. The power inputs for the AC and variable supplies are also connected to this line cord.

The five separate power supplies can be customized to suit your needs. If you decide to enclose the power supplies, be sure to provide for adequate ventilation to allow heat to escape. The main 120-volt AC input should be protected against surges and spikes if you anticipate prototyping computer circuits.

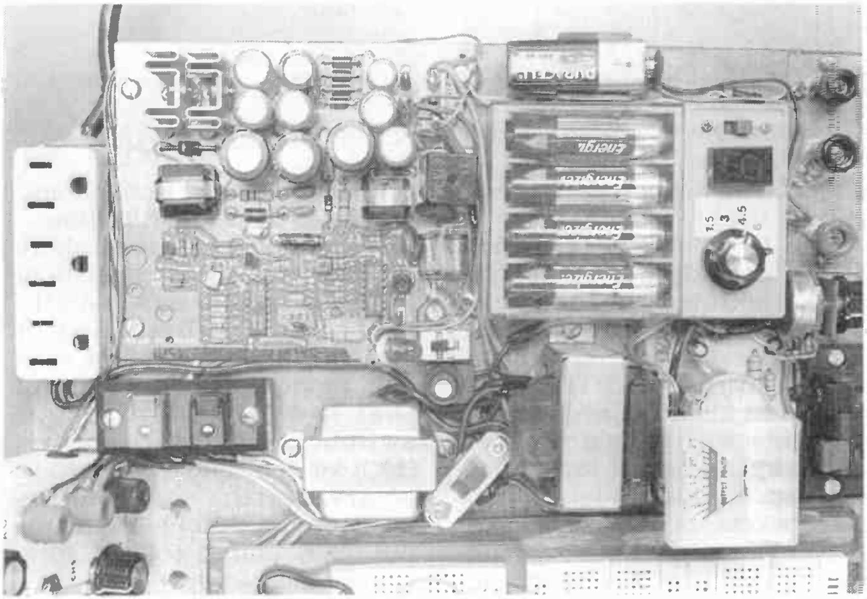

FIG. 2—A VARIETY OF POWER SOURCES is incorporated into the lab.

Breadboards

In the center of the base board are three 640-point solderless breadboards, laid out side-by-side (see Fig. 3). The top breadboard is a Radio Shack unit on a metal base with three multipurpose posts. The other two breadboards are plug-in panels mounted on a plywood base. All three breadboards have dual two-connection bus lines at their tops and bottoms, and will accommodate SIPs, DIPs, dis-

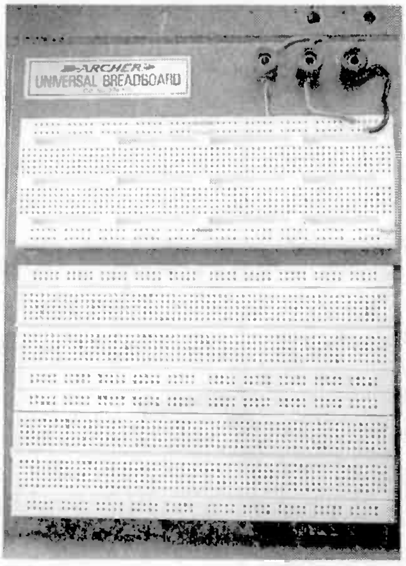
ARCHER
UNIVERSAL BREADBOARD

FIG. 3—THREE 640-POINT breadboards are laid out side-by-side. The top breadboard is on a metal base with three multipurpose posts, and the other two are plug-in panels mounted on a plywood base.

crete components, and jumper wires. While this breadboard arrangement has provided the author with enough space and has been versatile enough for all of his projects to date, it is certainly not the only arrangement possible. The breadboards are available in a variety of shapes and sizes, with from 250 contact points to 2500 or more. Some thought about the types of circuits you'll be working on will help you make a final decision on this part of the lab.

Backboards and modules

The main function of the backboard is to support more cumbersome components, modules, and any other circuits that the builder wants to add—items such as switches, potentiometers, meters, displays. The main backboard occupies about 10 inches across the back edge of the base (see Fig. 4).

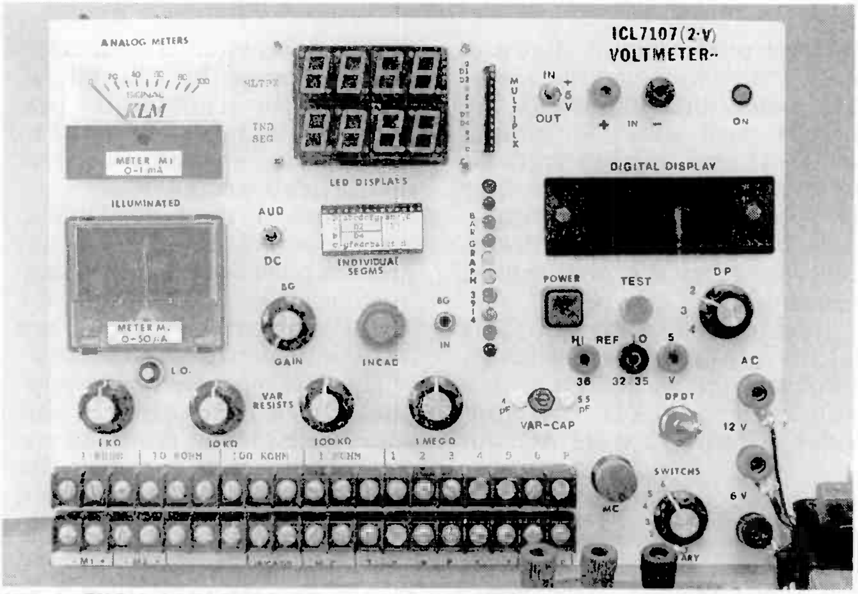
ANALOG METERS
KLM
METER M1 0-1 mA
ILLUMINATED
METER M2 0-50µA
TND SEG
LED DISPLAYS
INDIVIDUAL SEGMS
AUD
DC
BG
GAIN
INCAD
BG
IN
BARGRAPH 3914
MULTIPLEX
LO
VAR RESISTS
1KΩ
10KΩ
100KΩ
1MEGΩ
ICL7107(2-V) VOLTMETER
IN
+5V
OUT
+ IN −
ON
DIGITAL DISPLAY
POWER
TEST
DP
HI REF LO
36
32 35
5
V
AC
4 pF
55 pF
VAR-CAP
DPDT
12 V
SWITCHS
MC
6 V

FIG. 4—THE MAIN BACKBOARD occupies about 10 inches across the back edge of the base. The backboard holds many of the accessories.

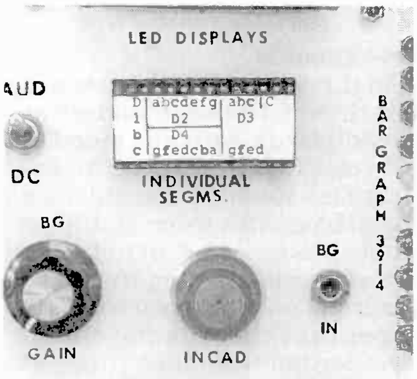
LED DISPLAYS
AUD
D abcdefg abc C
1 D2 D3
b D4
c gfedcba gfed
INDIVIDUAL SEGMS
DC
BG
BG
GAIN
INCAD
IN
BARGRAPH 3914

FIG. 5—A BARGRAPH is mounted in the center of the backboard.

PARTS LIST—MAIN UNIT

Resistors
1K panel-mount potentiometer
10K panel-mount potentiometer
100K panel-mount potentiometer
1 megohm panel-mount potentiometer
Capacitors
4 to 55 pF variable capacitor
Semiconductors
One 0.3-inch 7-segment display, common anode
Four 0.56-inch 7-segment displays, common anode (MAN6710) or four 0.56-inch 7-segment displays, common cathode (MAN6740)
Power supplies
Texas Instrument surplus computer power supply, or equivalent
12.6-volt AC transformer
4-AA battery holder
9-volt battery connector
1.2 to 20 volt variable DC supply
120 VAC power strip
Other components
Two spring-loaded speaker-type terminals, six binding posts, 6-foot AC linecord, plastic project cases, three 640-point breadboards, two 19-point terminal strips, heavy-duty DPDT toggle switch, momentary pushbutton switch, 6-position rotary switch, 0 to 1 mA analog meter movement, small push-on push-off switch, 11-pin SIP socket, 28-pin DIP socket, bayonet lamp socket with lens and bulb assortment, hookup wire and zip cord, hardware, seven panel knobs, TO-220 10-watt heatsink, 4-contact tie point.

PARTS LIST—BARGRAPH

IC1—LM3914 bargraph driver and 18-pin DIP socket, National Semiconductor
R1—1000 ohms, ¼-watt resistor
R2—5000 ohms, panel-mount potentiometer with switch
T1—audio transformer, 1K primary, 8-ohm secondary (Radio Shack 273-1380)
J1—⅛-inch phono jack and plug
D1—1N4001 diode
LED1–LED10—light-emitting diode, any color
S1—DPDT switch

Two 19-position terminal strips provide access to the parts mounted on the backboard. The strips on the prototype were removed from a surplus alarm board. The screw terminals allow connections to be made with alligator clips, bare wire, and spade lugs. However, spring-loaded terminals, SIP sockets, or any other connector you might have on hand can be used here. Each point on the terminal strips is wired to a connection on the component, and labeled for identification.

The row of components above the terminal strips consists of four potentiometers (1-, 10-, and 100-kilohms and 1 megohm) and a 4- to 55-pF variable capacitor. To the right of that row is a DPDT toggle switch, momentary control (MC) switch, and six-position rotary switch. The binding posts on the extreme right output the 6- and 12-volts AC, previously mentioned.

On the left side are two analog

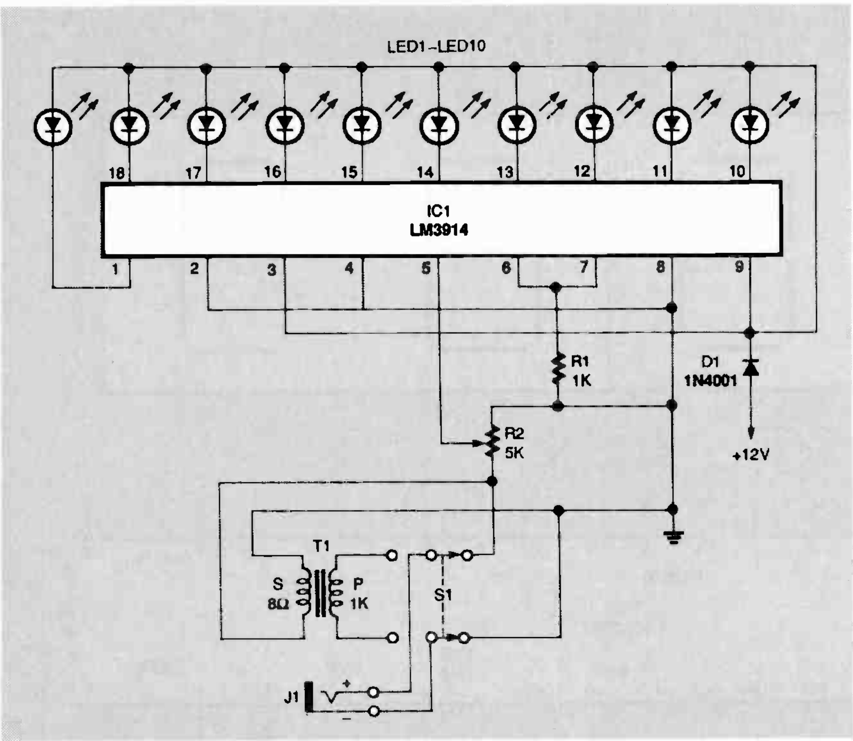

FIG. 6—LED BARGRAPH SCHEMATIC. The circuit consists mainly of an LM3914 bargraph display driver IC.

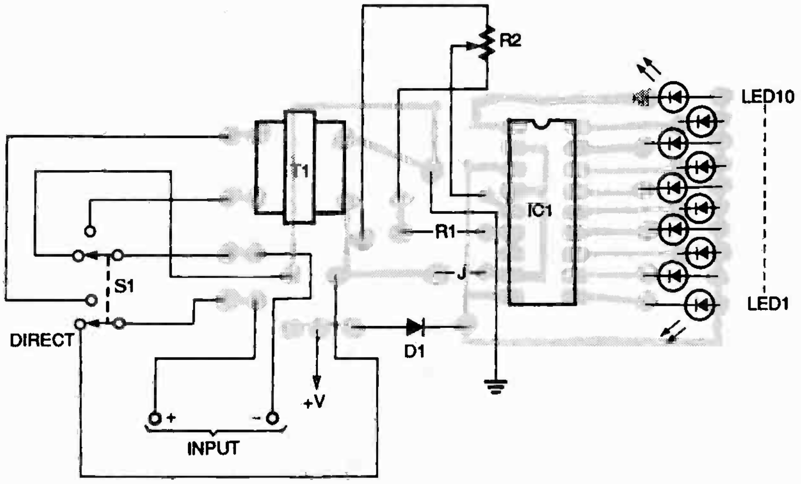

FIG. 7—PARTS-PLACEMENT DIAGRAM for the bargraph circuit.

meters (M1 and M2). Meter M1 is rated for 0 to 1 milliampere, which is a useful value for many circuit designs, and M2 is 500 microampere, ± center scale unit. This is essential when designing a circuit where the output can swing either positive or negative. Meter M2 is illuminated, and the small push-button switch labeled L.O. activates the 6-volt lamp.

A bank of seven-segment LED displays is mounted in the top center of the backboard. The first row of four is in a multiplexed arrangement, with all like segments tied together. Connection to the multiplexed displays is accomplished via the SIP socket to the right, which is labeled accordingly. The second row of seven-segment displays is wired as a 3½-digit display that allows individual access to each of the 24 segments and the common anode. Connections to this display row are made via the DIP socket below it, which is marked appropriately for use with digital circuits having that type of output.

Below the DIP socket is an incandescent lamp indicator with a bayonet-type socket. Bulbs are easily changed, and they are available in 6-, 12-, 14-, and 28-volt versions.

The author installed a ten-step LED bargraph and a 3½-digit voltmeter on the backboard, both as a matter of personal preference. The bargraph is mounted in the center of the backboard as shown in Fig. 5. A schematic of the LED bargraph circuit is shown in Fig. 6. The circuit consists mainly of an LM3914 bargraph display driver IC. Switch S1 allows signals to be input to the driver chip

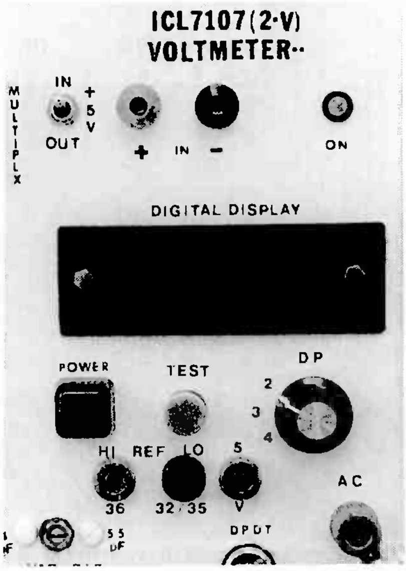

FIG. 8—A 3½-DIGIT VOLTMETER occupies the upper right hand corner of the backboard.

PARTS LIST—VOLTMETER

IC1—ICL7107CPL A/D converter/LED display driver, Harris
DISP1—two 2-digit common-anode 7-segment displays (MAN6710 or equiv.)
C1—0.22 µF, mica capacitor
C2—0.047 µF, mica capacitor
C3—0.01 µF, mica capacitor
C4—0.1 µF, mica capacitor
C1—100 pF, ceramic disc capacitor
R1—470,000 ohms, ¼-watt, 5% resistor
R2—1 megohm, ¼-watt, 5% resistor
R3—25,000 ohms, PC-mount potentiometer
R4—22,000 ohms, ¼-watt, 5% resistor
R5—100,000 ohms, ¼-watt, 5% resistor
S1—SPDT toggle switch
Miscellaneous: pushbutton switch, 3-position rotary switch, five banana jacks, green LED

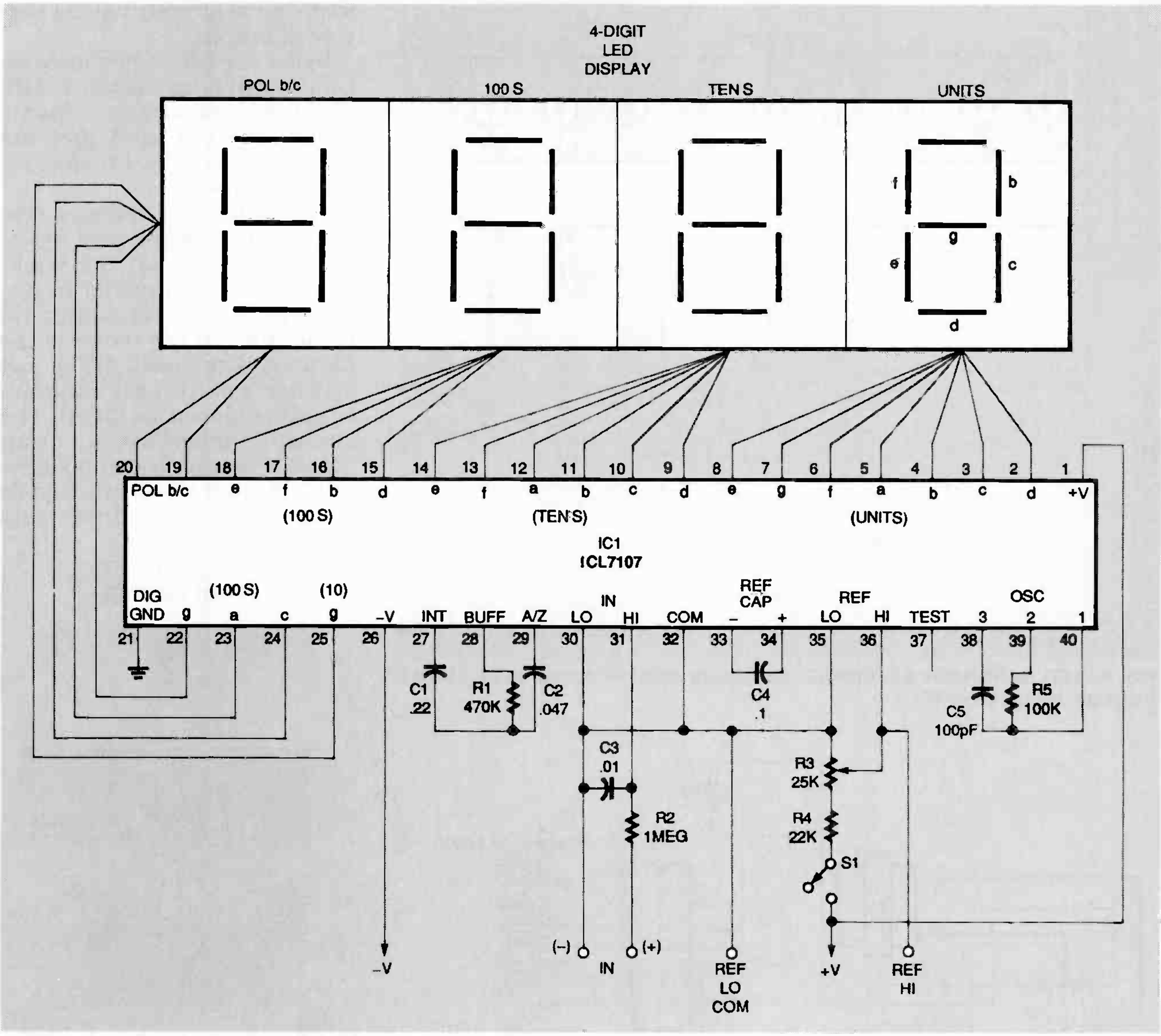
4-DIGIT LED DISPLAY
POL b/c
100 S
TENS
UNITS
f
b
g
e
c
d
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
POL b/c
e
f
b
d
e
f
a
b
c
d
e
g
f
a
b
c
d
+V
(100 S)
(TENS)
(UNITS)
IC1
ICL7107
DIG GND
g
(100 S) a
c
(10) g
-V
INT
BUFF
A/Z
IN LO
HI
COM
REF CAP
-
+
REF LO
HI
TEST
OSC 3
2
1
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
C1 .22
R1 470K
C2 .047
C4 .1
C5 100pF
R5 100K
C3 .01
R2 1MEG
R3 25K
R4 22K
S1
-V
(-)
IN
(+)
REF LO COM
+V
REF HI

FIG. 9—THE SCHEMATIC DIAGRAM of the voltmeter circuit.

PARTS LIST—BCD DECODER

- IC1—74LS48 BCD to 7-segment decoder
- IC2—CD4553 3-digit BCD counter, Harris
- Q1–Q3—2N3906 PNP transistors
- R1–R7—220 ohms, ¼-watt, 5% resistor
- R8–R10—1000 ohms, ¼-watt, 5% resistor
- C1—0.001 µF, ceramic disc capacitor
- DISP1—three common-cathode 7-segment displays

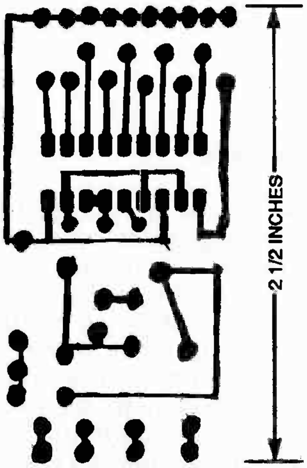

BARGRAPH FOIL PATTERN.

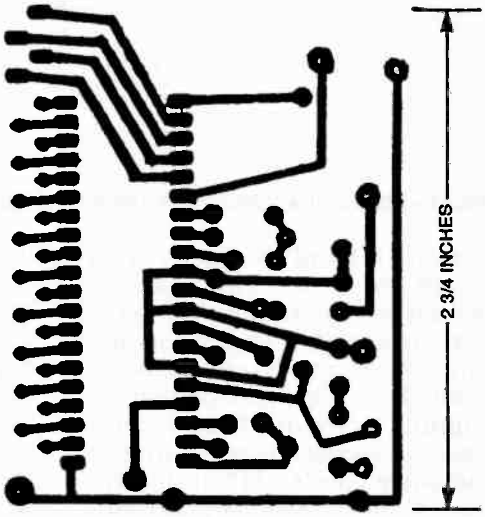

VOLTMETER FOIL PATTERN.

directly, or through a matching transformer T1 for audio.

Signals are input to the bargraph circuit at J1, a ⅛-inch mini phono jack. The 5-kilohm potentiometer (R2) controls gain. The author used a potentiometer with a built-in on/off switch so that power to the bargraph display could be turned on and off independently. You can either do the same, or use a separate switch for power. Power for the bargraph circuit is provided by the TI computer supply.

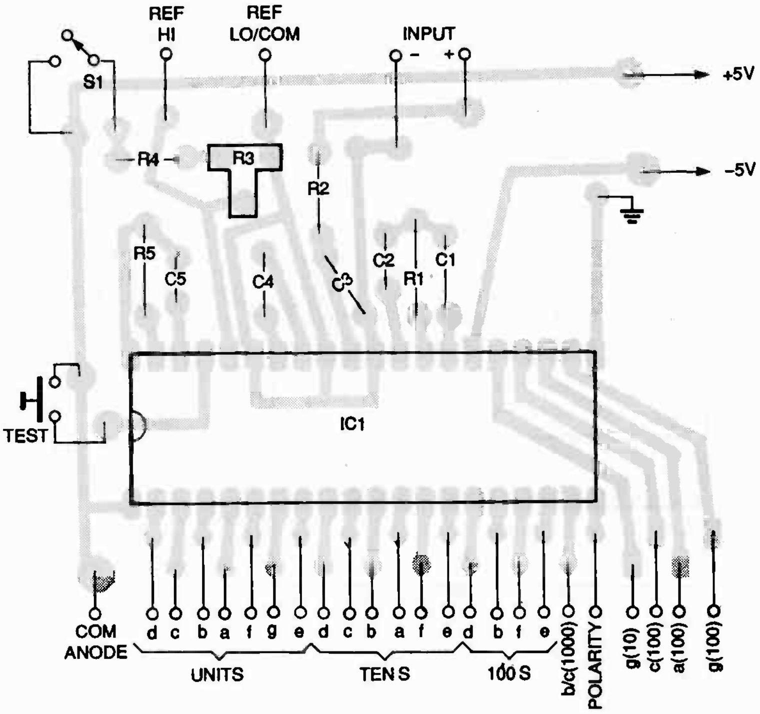
REF HI
REF LO/COM
INPUT - +
S1
+5V
-5V
R4
R3
R2
R5
C5
C4
C3
C2
R1
C1
IC1
TEST
COM ANODE
d c b a f g e d c b a f e d b f e
UNITS
TENS
100S
b/c(1000)
POLARITY
g(10)
c(100)
a(100)
g(100)

FIG. 10—PARTS-PLACEMENT DIAGRAM for the voltmeter circuit.

PARTS LIST AUDIO AMPLIFIER

IC1—LM386 audio amplifier, National Semiconductor

R1—10,000 ohms, panel-mount potentiometer with switch (or use separate potentiometer and switch)

C1—20 µF, 25 volts, electrolytic

8-position terminal strip

1½-inch, 8-ohm speaker

PARTS LIST—METER

M1—4- to 6-inch, 0 to 1 mA analog-meter movement
S1—8-position rotary switch
D1—1N4002 diode
R1—10 ohms, ¼-watt, 5% resistor
R2—100 ohms, ¼-watt, 5% resistor
R3—1000 ohms, ¼-watt, 5% resistor
R4—10,000 ohms, ¼-watt, 5% resistor
R5—100,000 ohms, ¼-watt, 5% resistor
R6—1 megohm, ¼-watt, 5% resistor
R6—10 megohms, ¼-watt, 5% resistor
Two banana jacks

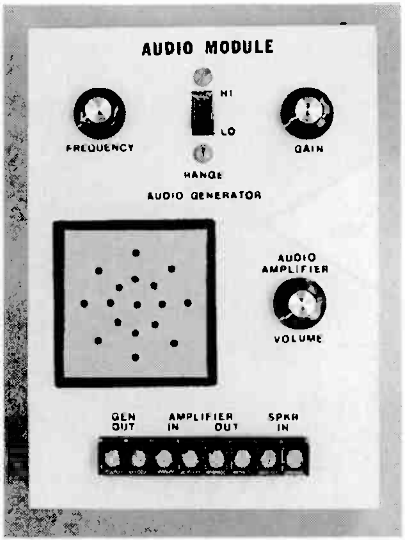
AUDIO MODULE
FREQUENCY
HI
LO
RANGE
GAIN
AUDIO GENERATOR
AUDIO AMPLIFIER
VOLUME
GEN OUT
AMPLIFIER IN OUT
SPKR IN

FIG. 11—THE AUDIO MODULE consists of an audio generator, an amplifier, a speaker, and a terminal strip.

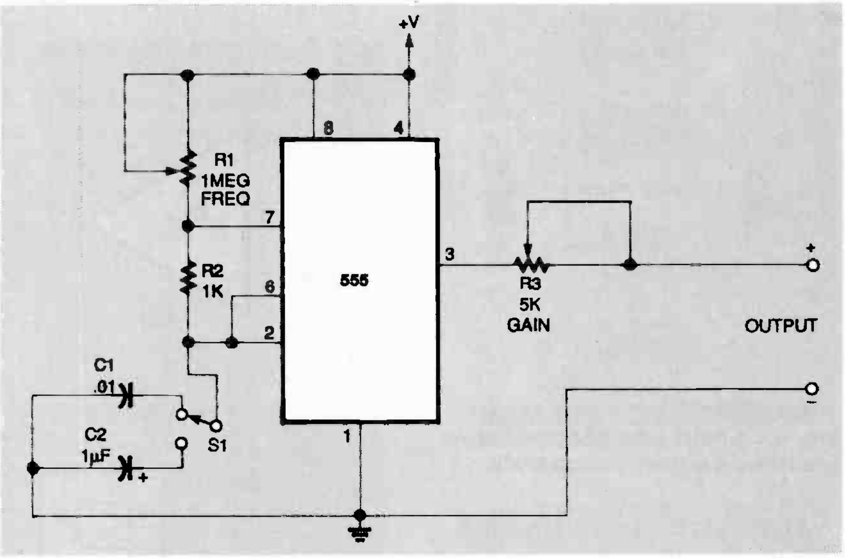
+V
8
4
R1
1MEG
FREQ
7
R2
1K
6
2
555
3
R3
5K
GAIN
+
OUTPUT
-
C1
.01
C2
1µF
S1
1

FIG. 12—SIGNAL GENERATOR. Potentiometer R1 sets the output frequency.

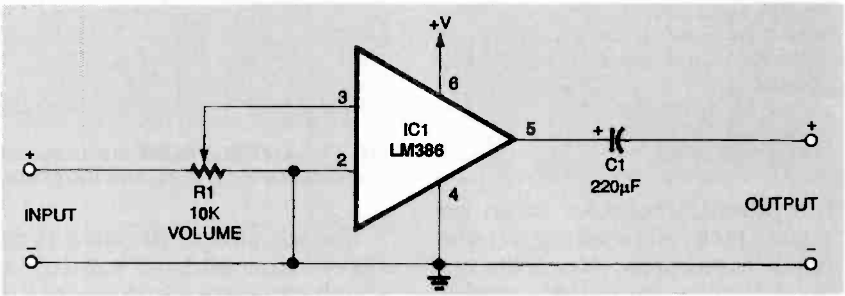
+V
6
3
2
IC1
LM386
4
5
+
C1
220µF
+
INPUT
-
R1
10K
VOLUME
OUTPUT
+
-

FIG. 13—AUDIO AMPLIFIER. This circuit has a gain of about 20.

A foil pattern is provided for the bargraph circuit, and the parts-placement diagram for it is shown in Fig. 7.

The 3½-digit voltmeter occupies the upper right hand corner of the backboard, as shown in Fig. 8. Figure 9 is the schematic diagram of the voltmeter circuit. A foil pattern is provided for this circuit, and Fig. 10 is its parts-placement diagram. This circuit can also be point-to-point wired. The circuit contains a Harris ICL 7107 analog-to-digital converter/ display driver. A standard 2-volt configuration was chosen for this circuit. Switch S1 allows the +5-volt source to be routed either to the on-board 25-kilo-

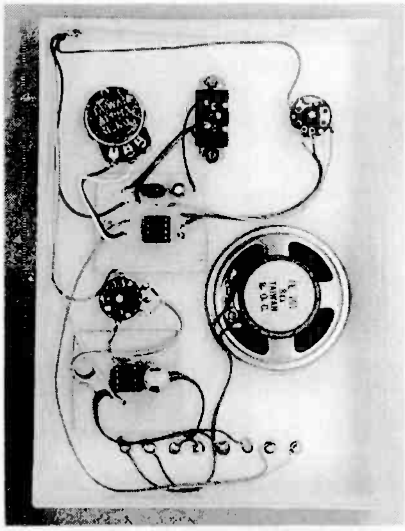

FIG. 14—THE AUDIO MODULE is too simple to call for a PC board, so point-to-point wiring was used.

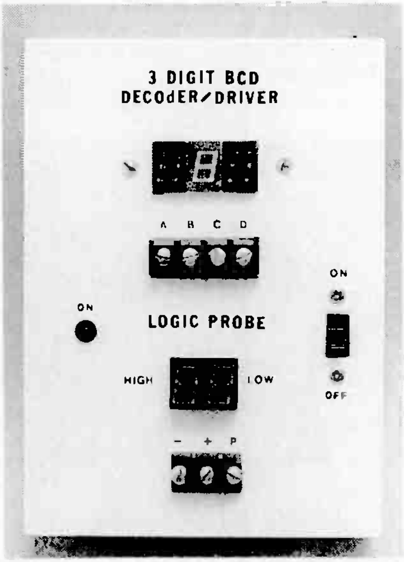
3 DIGIT BCD
DECOdER/DRIVER
A B C D
ON
LOGIC PROBE
HIGH
LOW
OFF
- + P

FIG. 15—3-DIGIT BCD DECODER-driver is another practical function block.

PARTS LIST—AUDIO MODULE

- IC1—LM555 timer
- R1—1 megohm, panel-mount potentiometer with switch
- R2—1000 ohms, ¼-watt, 5% resistor
- R3—5000 ohms, panel-mount potentiometer
- C1—0.01 µF, Tantalum
- C2—1 µF, Tantalum
- S1—SPDT switch

hm potentiometer, or to an external jack, depending on the need. To calibrate the meter, apply 1.2 volts from the variable power supply to the input, and adjust R3 until the display reads 1.200. Then, switch the input leads, and the display should read −1.200.

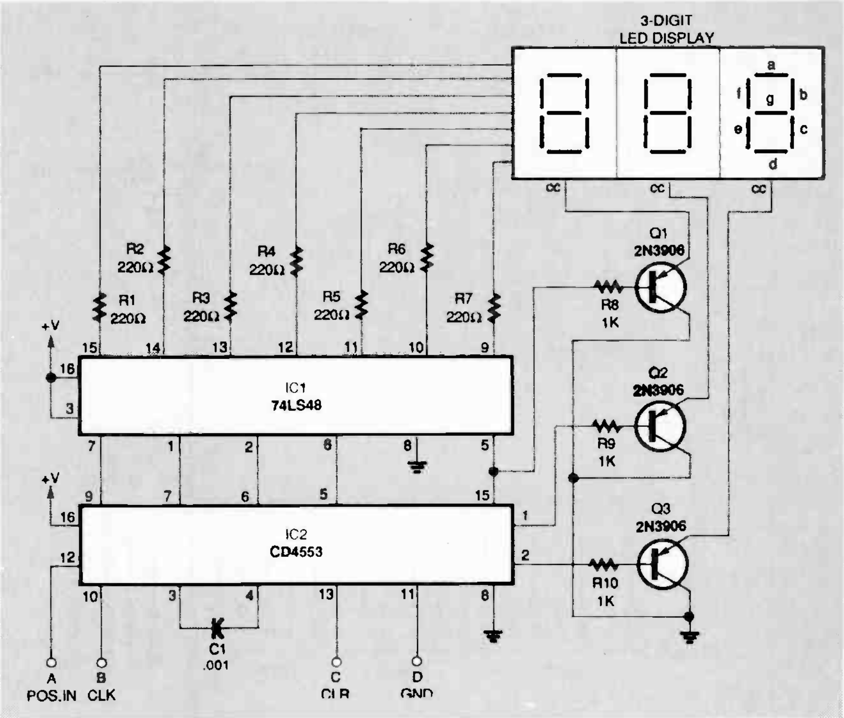
3-DIGIT
LED DISPLAY
a
b
c
d
e
f
g
cc
cc
cc
Q1
2N3906
Q2
2N3906
Q3
2N3906
R8
1K
R9
1K
R10
1K
R1
220Ω
R2
220Ω
R3
220Ω
R4
220Ω
R5
220Ω
R6
220Ω
R7
220Ω
+V
IC1
74LS48
IC2
CD4553
C1
.001
A
POS.IN
B
CLK
C
CLR
D
GND

FIG. 16—THE BCD DECODER-DRIVER circuit will interface with any standard BCD output to produce a digital display.

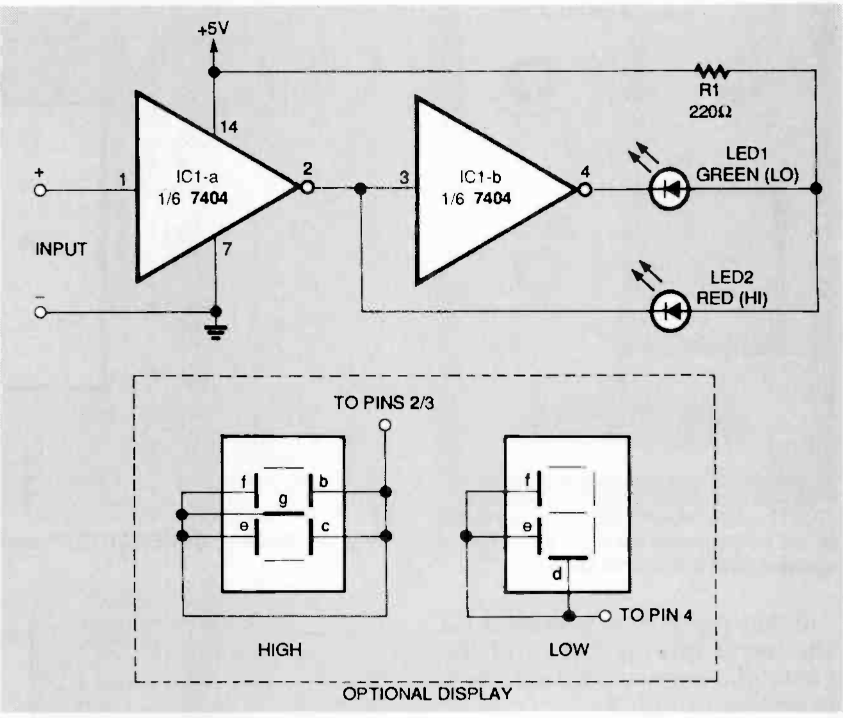
+5V
R1
220Ω
IC1-a
1/6 7404
IC1-b
1/6 7404
+
INPUT
-
LED1
GREEN (LO)
LED2
RED (HI)
TO PINS 2/3
TO PIN 4
HIGH
LOW
OPTIONAL DISPLAY

FIG. 17—A LOGIC PROBE is also included in BCD decoder module. The red LED lights to indicate a logic high, and the green LED lights to indicate a logic low.

The left side of the base is reserved for add-on modules, which can have a number of different functions depending on what is needed. The audio module shown in Fig. 11 consists of an audio signal generator, an

(Continued on page 112)

PARTS LIST—LOGIC PROBE

- IC1—7404 hex inverter
- R1—220 ohms, ¼-watt, 5% resistor
- LED1—green light-emitting diode
- LED2—red light-emitting diode
- Two common-anode 7-segment displays (optional, see text) source

Build this simple superheterodyne receiver to receive time and other useful signals from NIST's WWV and WWVH.

WWV RECEIVER

NEIL HECKT

BROADCASTING THE EXACT TIME OF day is just one of the many services provided by radio stations WWV and WWVH operated by the National Institute of Standards and Technology (NIST). Their ultra-accurate 10-megahertz carrier frequencies can be used to calibrate signal generators in a process called "zero beating" against the carrier frequency. The stations also transmit many other accurate frequencies, storm warnings for mariners at sea, and GPS (Global Positioning Satellite) position data.

This article explains how to build a simple, inexpensive superheterodyne radio receiver for WWV and WWVH signals. The objective in the design of this receiver was to obtain reliable reception with an antenna only 3- to 6-feet long. The sensitivity obtained with this receiver equals that obtainable from a $1000 communications receiver operating from a 60-foot antenna.

Can the performance of a receiver made from $30 worth of parts equal the performance of a $1000 factory-built receiver? The answer is no because sensitivity is only one measure of receiver performance. The most significant difference between receivers is in their overall signal-to-noise ratios. However, the WWV/WWVH receiver described here has sufficient sensitivity to achieve its dedicated function.

This receiver economizes in the quality of its filters. Most factory-made communications receivers contain expensive precision crystal intermediate frequency (IF) filters which greatly improve the signal-to-noise ratio by screening out most of the noise in the radio-frequency spectrum. Moreover, those communications receivers also have superior intermodulation distortion (IMD) and dynamic range not required for WWV/WWVH reception.

WWV/WWVH transmissions

The WWV transmitter is located in Fort Collins, Colorado, and its signal can be identified by male voice announcements.

TABLE 1—UTC TIME ZONE CONVERSION

UTC	Eastern	Central	Mountain	Pacific Alaska	Alaska Hawaii
0000	7:00PM	6:00PM	5:00PM	4:00PM	2:00PM
0100	8:00PM	7:00PM	6:00PM	5:00PM	3:00PM
0200	9:00PM	8:00PM	7:00PM	6:00PM	4:00PM
0300	10:00PM	9:00PM	8:00PM	7:00PM	5:00PM
0400	11:00PM	10:00PM	9:00PM	8:00PM	6:00PM
0500	Midnight	11:00PM	10:00PM	9:00PM	7:00PM
0600	1:00AM	Midnight	11:00PM	10:00PM	8:00PM
0700	2:00AM	1:00AM	Midnight	11:00PM	9:00PM
0800	3:00AM	2:00AM	1:00AM	Midnight	10:00PM
0900	4:00AM	3:00AM	2:00AM	1:00AM	11:00PM
1000	5:00AM	4:00AM	3:00AM	2:00AM	Midnight
1100	6:00AM	5:00AM	4:00AM	3:00AM	1:00AM
1200	7:00AM	6:00AM	5:00AM	4:00AM	2:00AM
1300	8:00AM	7:00AM	6:00AM	5:00AM	3:00AM
1400	9:00AM	8:00AM	7:00AM	6:00AM	4:00AM
1500	10:00AM	9:00AM	8:00AM	7:00AM	5:00AM
1600	11:00AM	10:00AM	9:00AM	8:00AM	6:00AM
1700	Noon	11:00AM	10:00AM	9:00AM	7:00AM
1800	1:00PM	Noon	11:00AM	10:00AM	8:00AM
1900	2:00PM	1:00PM	Noon	11:00AM	9:00AM
2000	3:00PM	2:00PM	1:00PM	Noon	10:00AM
2100	4:00PM	3:00PM	2:00PM	1:00PM	11:00AM
2200	5:00PM	4:00PM	3:00PM	2:00PM	Noon
2300	6:00PM	5:00PM	4:00PM	3:00PM	1:00PM

(add one hour during daylight savings time)

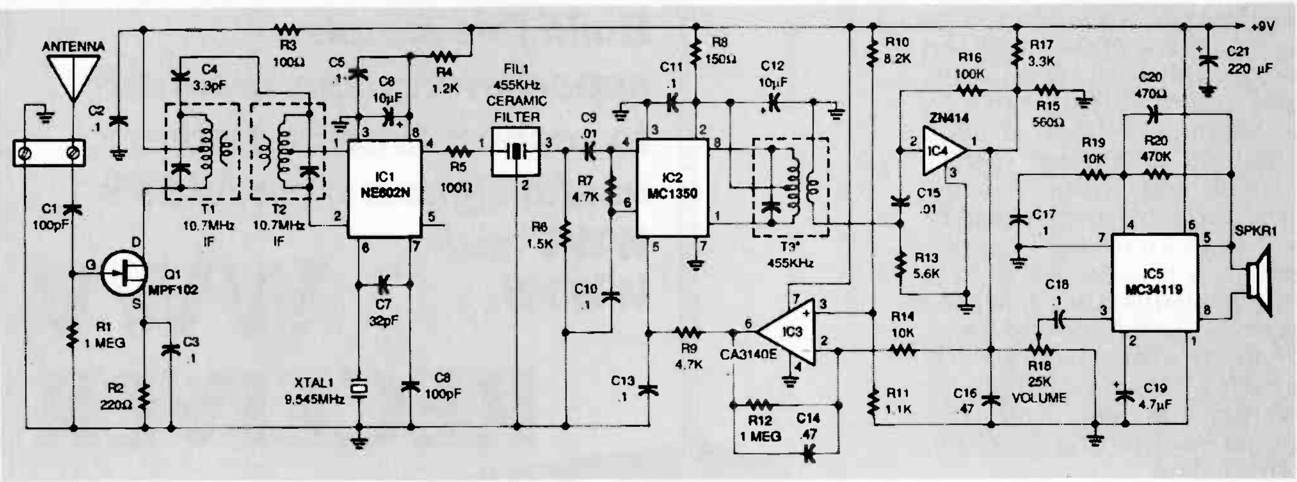
ANTENNA
C2 .1
C1 100pF
R3 100Ω
C4 3.3pF
C5 .1
C6 10µF
R4 1.2K
T1 10.7MHz IF
T2 10.7MHz IF
IC1 NE602N
Q1 MPF102
G
D
S
R1 1 MEG
R2 220Ω
C3 .1
C7 32pF
XTAL1 9.545MHz
C8 100pF
R5 100Ω
FIL1 455KHz CERAMIC FILTER
C9 .01
R6 1.5K
R7 4.7K
C10 .1
C11 .1
R8 150Ω
C12 10µF
IC2 MC1350
T3* 455KHz
C13 .1
R9 4.7K
IC3 CA3140E
R12 1 MEG
C14 .47
R10 8.2K
R16 100K
R17 3.3K
ZN414
IC4
C15 .01
R13 5.6K
R14 10K
R11 1.1K
C16 .47
R15 560Ω
R19 10K
R20 470K
C20 470Ω
C17 .1
C18 .1
R18 25K VOLUME
C19 4.7µF
IC5 MC34119
SPKR1
C21 220 µF
+9V

FIG. 1—SCHEMATIC OF THE WWV RECEIVER. A short antenna from three to six feet long is sufficient.

WWVH's transmitter is located in Kauai, Hawaii, and its signal can be identified by female voice announcements. Both stations broadcast on exactly the same frequencies, but they do not interfere with each other because their bands are so narrow they sound like a single station when both are received with equal signal strength.

Stations WWV and WWVH operate on a carrier frequency of 2.5 MHz. Station WWV broadcasts with 2.5 kilowatts (kW) of power, while WWVH broadcasts with 5 kW of power. Other frequency-power relationships are 5, 10, and 15 MHz at 10 kW and 20 and 25 MHz at 2.5 kW. The receiver described in this article was designed to receive 10-MHz signals because the author has found that they provide the most reliable reception.

Carrier frequency and all other time-related data is derived from cesium-controlled oscillators that are accurate to within ±1 part in 10^{11}. Daily deviations are less than one part in 10^{12} from day-to-day. Phenomena such as Doppler and diurnal shifts can degrade the accuracy at the receiver, but long-term accuracy of one part in 10^9 can easily be obtained.

There are many different time zones around the world, so it would be impractical to broadcast time for each time zone. To simplify matters, WWV/WWVH transmit time data in *Coordinated Universal Time* (UTC), also known as Greenwich Mean Time (GMT), World Time and Zulu Time.

What is UTC? It is the time in England, uncorrected for daylight savings time, and it is always stated in a 24-hour format. It is called GMT because the zero or Greenwich meridian passes through England, west of London. Table 1 provides information that will permit you to determine the local time-of-day from the UTC announced by WWV and WWVH.

Refer to the sidebar for more technical details on the WWV and WWVH transmission formats. More information on the services of WWV and WWVH can be obtained by writing Frequency-Time Broadcast Services Section, Time and Frequency Division, National Institute of Standards and Technology, Boulder, CO 80302.

Circuit description

A schematic of the WWV receiver is shown in Fig. 5. A short antenna (three to six feet) represents a high-impedance source. Transistor Q1, a Motorola

WWV and WWVH minute signals

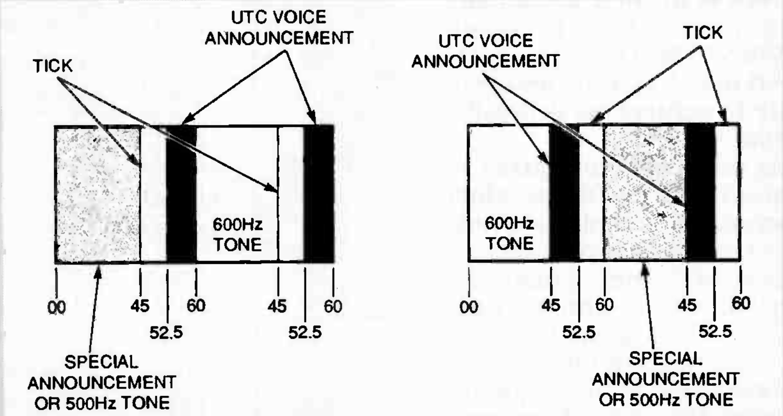
TICK
UTC VOICE ANNOUNCEMENT
600Hz TONE
00
45
52.5
60
45
52.5
60
SPECIAL ANNOUNCEMENT OR 500Hz TONE
UTC VOICE ANNOUNCEMENT
TICK
600Hz TONE
00
45
52.5
60
45
52.5
60
SPECIAL ANNOUNCEMENT OR 500Hz TONE

WWV and WWVH each have different one minute increments of time, as shown. Minute zero, and all even numbered minutes of WWV begin with a 45-second interval containing either a special announcement or a 500-Hz tone. This is followed by 7.5 seconds of silence except for the one-second *tick*.

The last 7.5 seconds contains the voice UTC time announcement. Minute 1, and all odd numbered minutes, have 45 seconds of a 600-Hz tone followed by the 7.5 seconds of silence except for tick and the 7.5-second UTC time announcement.

WWVH broadcasts with a similar format except even and odd minutes are reversed, and the silent and UTC time announcement intervals are reversed. This prevents interference if both stations are received simultaneously (this occurs on the West Coast of the U.S.). The 45-second interval of each minute might contain variations.

MPF102 JFET, provides input impedance matching and some gain. A double-tuned input filter, consisting of T1, T2, and C4, provides adequate selectivity to attenuate the image frequency of 9.09 MHz.

A Philips NE602N mixer (IC1) contains a double-balanced mixer and oscillator which supplies approximately 14 dB of conversion gain. The oscillator section is crystal controlled to provide stability and eliminate the need for tuning. The output of the NE602N is coupled to a 455-kHz ceramic IF filter (FIL1) with about a 4-kHz bandwidth. The filter's 1500-ohm input and output impedance is matched by resistors R5 and R7. The output impedance of IC1 is 1500 ohms which, with R5, equals about 1500 ohms.

A Motorola MC1350P IF amplifier (IC2) provides about 60 dB of gain with about 80 dB of automatic gain control (AGC) control range. The output is coupled through 455-kHz IF transformer T3. The output impedance of IC2 at 455 kHz is about 200 kilohms. Transformer T3 has a turns ratio of 6:1, and it provides an impedance transformation of 36:1.

Resistor R13 terminates the secondary of T3 with 5600 ohms, to provide the proper primary matching impedance. Other turns ratios can be used for T3 provided that R13 is also changed to maintain the proper impedance match. The input impedance of IC4 is high and it does not significantly affect the impedance match.

The GEC Plessey ZN414Z, an amplifier detector, is packaged in a typical plastic transistor case, but it is actually a ten-transistor IC which provides about 70 dB of gain, detection and some AGC response. It is designed to be powered from a 1.5-volt DC power supply with a 500-ohm load. This is emulated by resistors R15 and R17 whose equivalent circuit is 1.5-volts DC flowing through 500 ohms. The DC level at the output of IC4 is about 1.1-volts DC with no signal, and 0.9-volt DC with a strong signal.

The output signal is amplified, inverted, and filtered by IC3. Capacitor C14 removes any audio from the signal before it is applied to the gain control input

WWV hour signals

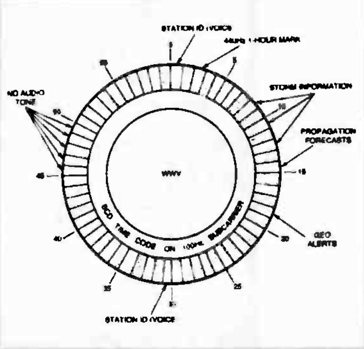

The figure shows how the WWV hour is divided. The beginning of each hour is identified by a 0.8-second,1500-Hz tone. The beginning of each minute is identified by a 0.8 second,1200-Hz tone. The 29th and 59th second pulse of each minute is omitted.

WWVH hour signals

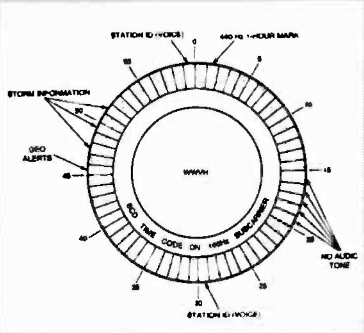

The figure illustrates how the WWVH hour is divided. The beginning of each hour is identified by a 0.8-second, 1500-Hz tone. The beginning of each minute is identified by a 0.8-second, 1200-Hz tone. The 29th and 59th second pulse of each minute is omitted.

WWV and WWVH broadcasts are approximate mirror images of each other. This prevents the two stations from creating interference mutual interference during voice announcements.

PARTS LIST

All resistors are ½-watt, 5%, unless otherwise noted.
R1, R12—1 megohm
R2—220 ohms
R3—100 ohms
R4—1200 ohms
R5—100 ohms
R6, R17—1500 ohms
R7, R9—4700 ohms
R8—150 ohms
R10—8200 ohms
R11—1100 ohms
R13—5600 ohms
R14, R19—10,000 ohms
R15—560 ohms
R16—100,000 ohms
R18—25,000 ohms, potentiometer PCB-mount
R20—470,000 ohms

Capacitors
C1, C8—100 pF, ceramic
C2, C3, C5, C10, C11, C13, C17, C18—0.1 µF, ceramic
C4—3.3 pF, ceramic
C6, C12—10 µF, 10 volts, aluminum electrolytic
C7—33 pF, ceramic
C9, C15—0.01 µF, ceramic
C14, C16—0.47 µF, ceramic
C19—4.7 µF, 10 volts, aluminum electrolytic
C20—470 pF, ceramic
C21—220 µF, 10 volts, aluminum electrolytic

Semiconductors
IC1—NE602AN mixer, Philips or equiv.
IC2—MC1350P IF amplifier, Motorola or equiv.
IC3—CA3140E operational amplifier, Harris or equiv.
IC4—ZN414Z amplifier/detector (GEC Plessey)
IC5—MC34119P power amplifier, Motorola or equiv.
Q1—MPF102 JFET, Motorola or equiv.

Other components
FIL1—455-kHz ceramic filter, Tobo HCFM2-455A or equiv.
T1, T2—10.7-MHz IF transformer (Toko or equivalent)
T3—455-kHz IF transformer, 6:1 ratio (Toko or equivalent)
XTAL1—9.545-MHz crystal
SPKR1—speaker, square, 2.5-inch, 8-ohms
S1—SPDT switch, PCB mount
Miscellaneous: 5.3 × 4 × 2 -inch plastic project case, PC board, 9-volt battery, battery clip, hookup wire, solder.

Note: The following items are available from Almost All Digital Electronics, 1412 Elm Street S.E., Auburn, WA 98092:

- **Complete kit—$39.95 + $4.00 S&H**
- **Partial kit (includes all electronic components, PC board, and crystal—does not include project case, speaker, or antenna)—$29.95 + $4.00 S&H**
- **PC board and crystal, only—$10.00 + $2.90 S&H**

Send check or money order. Washington State residents add local sales tax.

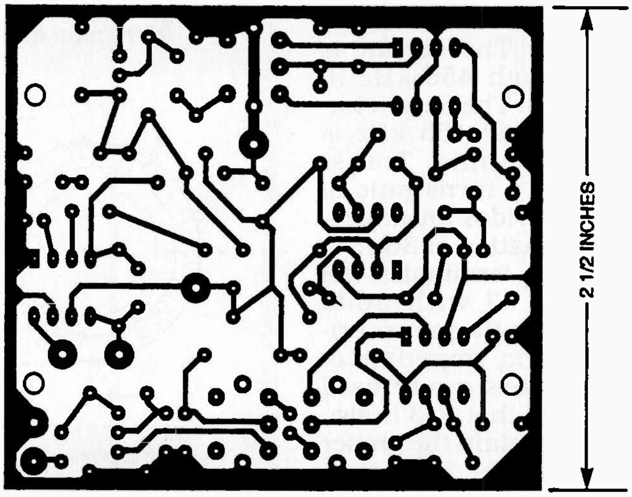

WWV RECEIVER FOIL PATTERN.

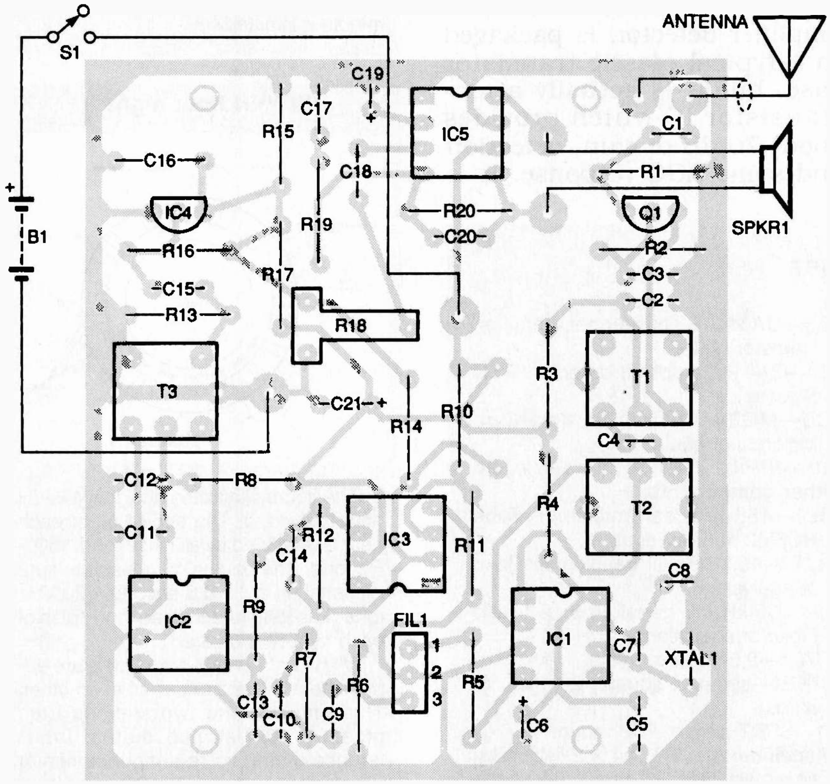

FIG. 2—PARTS-PLACEMENT DIAGRAM. The volume control is mounted on the PC board and adjusted before closing up the case.

pin 5 of the Motorola MC1350P IF amplifier (IC2). When the voltage at the output of IC4 exceeds the 1-volt DC reference level set by R10 and R11, the gain of IC2 starts to roll off. As a result, the AGC response of IC2 and IC3 tries to maintain the signal level so that the DC level at the output of IC4 is 1-volt DC.

The gain of the Motorola MC34119P audio amplifier (IC5) is set by the ratio of R20 to R19, which equals 47 (about 16 dB). This audio amplifier was selected because it is a bridge amplifier and does not require a large coupling capacitor to drive the speaker.

The overall maximum gain of the receiver is about 144 dB. This is the sum of a −10 dB in the input filter, +14 dB from IC1, −6 dB in FIL1, +60 dB from IC2, +70 dB from IC3, and +16 dB from IC4. Ordinarily a gain of 144 dB is too much. Normal operation can be expected to be with about a 120 dB maximum gain, which is typical of a high-quality communications receiver. However, the gain is usually less than that, as set by the volume-control potentiometer R18 and AGC loop.

Building the receiver

You can build the receiver on a single-sided PC board that you make yourself from the foil pattern provided here. Alternatively, partial and complete kits containing a finished PC board, as well as complete receivers, are available from the source given in the Parts List.

Figure 6 is a parts-placement diagram. The order of assembling the components to the circuit board and soldering them is not critical. Pay particular attention to the orientation of the semiconductors, that is the location of pin 1. Mount the volume control on the PC board. Keep in mind that it cannot be adjusted after it is installed in the case.

The author designed the circuit assuming that the receiver would be turned on only for brief periods to obtain the correct time, and that convenient access to the volume control would not be necessary. However, if you want an external volume control, one can be mounted on the case and wired directly to the pads intended for the PCB-mounted control.

After all the electronic components are inserted and soldered to the circuit board, wire the speaker, battery clip, power switch, and antenna. Install the completed circuit board in the case as shown in Fig. 7. Install a 9-volt battery.

Circuit alignment

To align the circuit, adjust transformers T1, T2, and T3 for peak AGC voltage at pin 6 of IC3. The desired tuning is broad for receiving a strong signal. Attach an antenna that is as short as practical to pick up an extremely weak signal for the alignment procedure. Even with those precautions, the tuning range of the IF transformers will appear to be quite broad.

Dividing Up Time For Broadcast

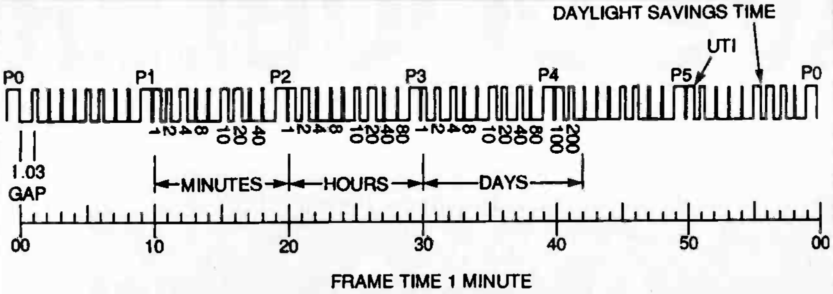

The figure illustrates the pulse format for the BCD time code (100-Hz subcarrier). During the passage of a minute, both stations broadcast the time of year on a pulse-width modulated 100-Hz subcarrier. Three different pulse widths are used:

1. P0-P5—position identifiers—0.8 seconds (80 cycles)
2. Logic 1—data bit—0.5 seconds (50 cycles)
3. Logic 0—data bit— 0.2 seconds (20 cycles)

In addition there is a 1.03 second (103 cycles) "hole" in the code for synchronization at the beginning of each minute signal. The minute is divided into six parts of 10 seconds duration. Each part consists of nine data bits and a position identifier pulse. The following listing is the contents of the various sections.

1.—Synchronization "hole" and eight logic zeros.

2.—Four data bits encoding the unit minutes in binary-coded decimal (BCD), a logic 0, three data bits encoding the tens of minutes in BCD and another logic 0.

3.—Four data bits encoding the unit hours in BCD, a logic 0, and two data bits encoding the tens of hours in BCD and two logic 0's.

4.—Four data bits encoding the unit days in BCD, a logic 0, and four data bits encoding the tens of days in BCD.

5.—Two data bits encoding the hundreds of days in BCD and seven logic 0's.

6.—One data bit if UT1 correction should be applied, four logic 0's, a data bit which is logic 1 if daylight saving time and three data bits with the UT1 correction in tenths of a second. The UT1 correction accounts for the Earth's rotation which varies slightly over time.

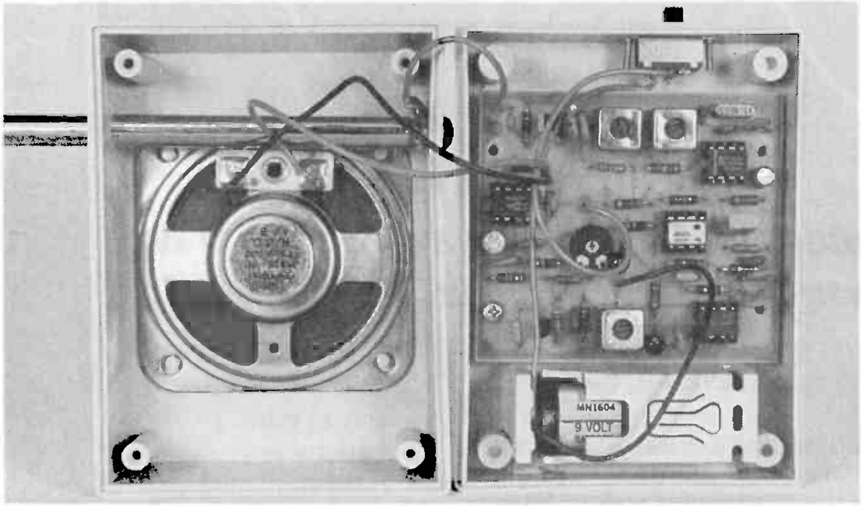

FIG. 3—THE FINAL ASSEMBLY consists of wiring the speaker, battery clip, power switch, and antenna. The author's prototype is shown here.

As an alternative to "on-air tuning," set up a 10-megahertz signal source. It will make the task of circuit alignment easier by providing a constant, local, easy to control signal. The author set a frequency counter at 10-megahertz and its internal oscillator provided sufficient ambient signal leakage to act as an excellent signal source for alignment.

The most difficult step is the alignment of the double-tuned input filter, T1/T2. To overcome that obstacle, all of the transformers that are supplied with the kit from the source listed in the Parts List are pre-aligned. Only slight adjustments are then necessary for the transformers after assembly.

Antenna considerations

The author lives near Seattle, Washington—about 1000 miles from Fort Collins, Colorado. Moreover, there are several mountain ranges between Seattle and Fort Collins. The quality of reception is usually excellent during the day with an antenna consisting of a 3- to 6-foot length of wire. When the signal is particularly strong, no antenna is needed.

The prototype receiver shown in the photographs has a 30-inch telescopic rod antenna that works quite well, but this a convenience feature rather than a requirement. There are two pads on the board for the connecting the antenna: one is connected to the antenna and the other is the ground connection. While a single wire on the antenna terminal usually works well, reception can generally be improved with a simple dipole antenna formed by a second wire connected to ground. Ω

It keeps more than memories alive.

THE AMERICAN HEART ASSOCIATION MEMORIAL PROGRAM®

1-800-242-8721

American Heart Association

This space provided as a public service.

SINEWAVE DOUBLER

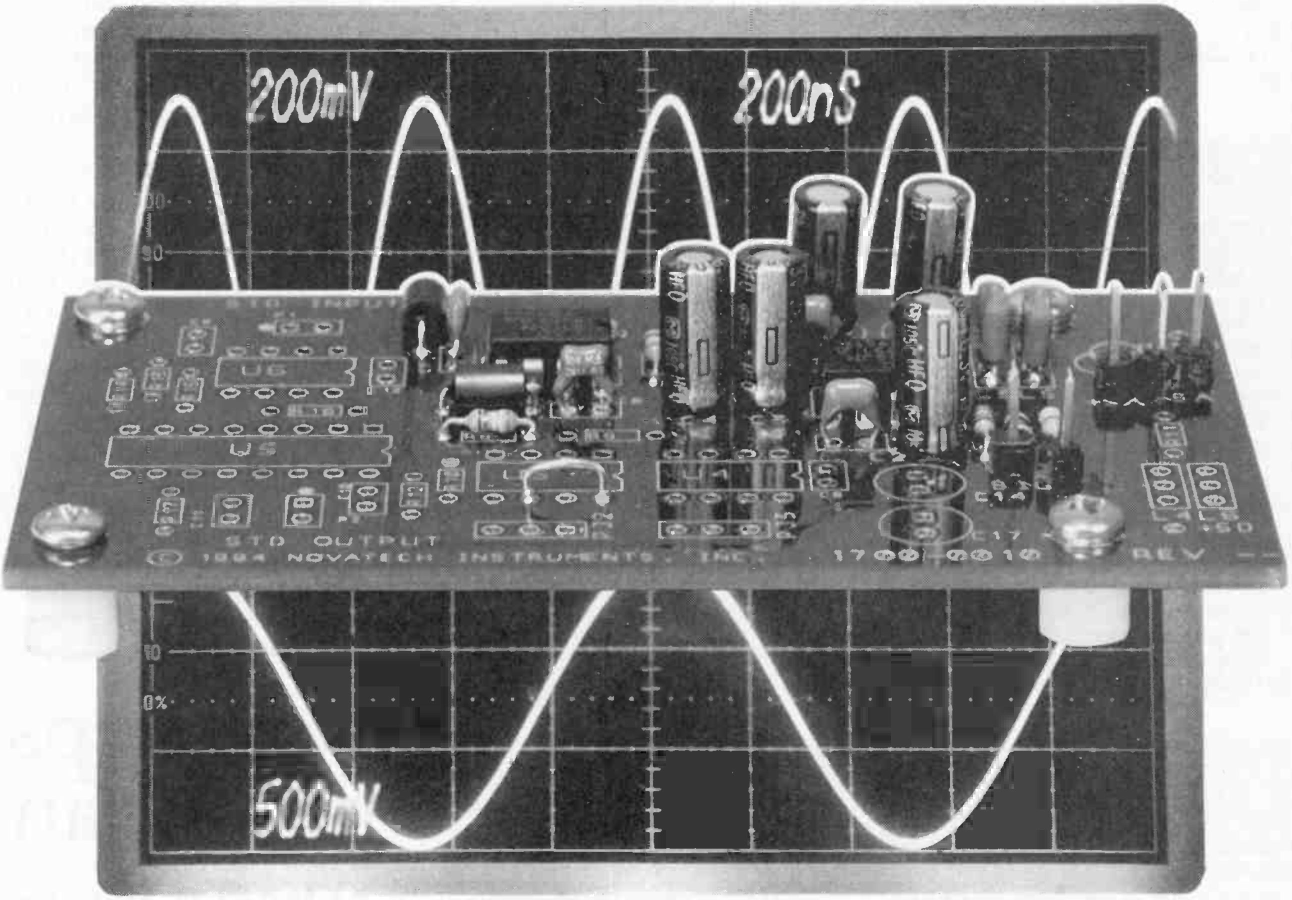

Double the frequency of any sinewave from 10 Hz to more than 50 MHz.

STEVEN D. SWIFT

HOW OFTEN HAVE YOU WISHED THAT your signal generator would generate a higher frequency? The small module described in this article doubles the frequency of any sinewave input, providing outputs from 10 hertz up to and beyond 50 megahertz. It measures less than 2 × 3.5-inches, and consumes about one-third of a watt. It can be permanently installed in almost any signal generator or used as an outboard module.

Theory of operation

A schematic of the frequency doubler module is shown in Fig. 1. The core of the module is IC1, an Analog Devices AD834 wideband four-quadrant multiplier. This IC provides wideband analog multiplication of two input signals, regardless of their polarity. In this circuit both X and Y inputs are tied together to provide a squaring function.

The operation of IC1 is best understood with the aid of simple mathematics. If the input is assumed to be a normalized (amplitude of 1) sinewave of frequency f, then the input can be written as:

$$\text{Input} = \sin(2\pi f)$$

where

$\pi = 3.14159\ldots$

f = frequency

t = time

The AD834 performs the following function:

$$XY = \sin(2\pi f) \times \sin(2\pi f)$$
$$= (\sin(2\pi f))^2$$

A fundamental trigonometric identity gives:

$$(\sin(2\pi f))^2 = (1 - \cos(2(2\pi f)))/2$$
$$= (1 - \cos(2\pi 2f))/2$$

The last term shows a DC offset as well as a cosine term with twice the frequency as the input. This performs the frequen-

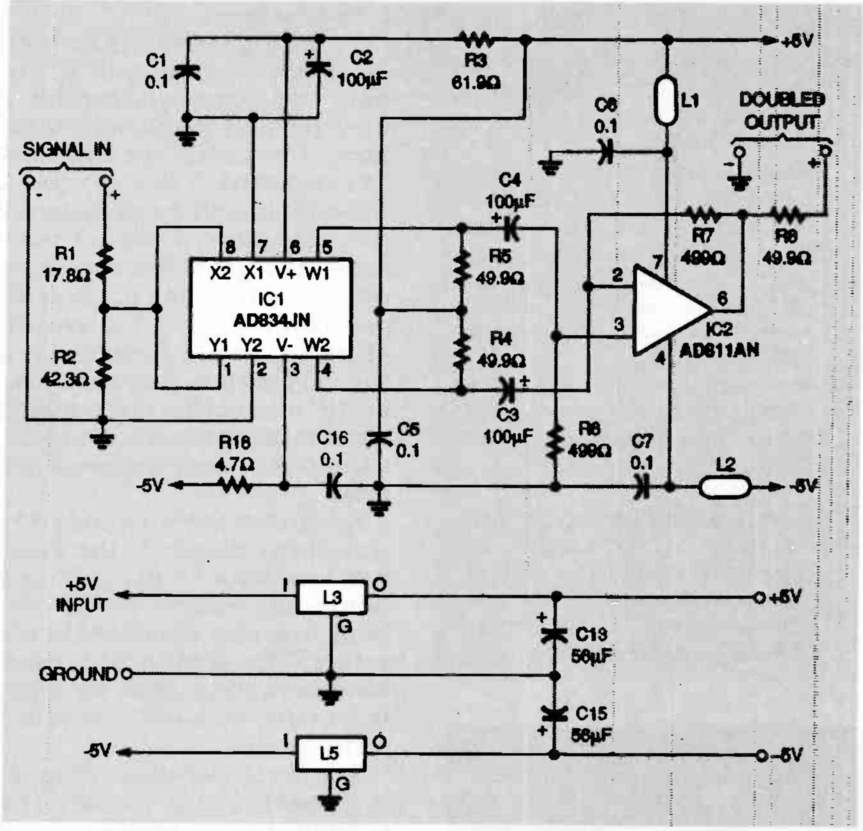
C1
0.1
C2
100µF
R3
61.9Ω
+5V
C6
0.1
L1
DOUBLED
OUTPUT
SIGNAL IN
R1
17.8Ω
R2
42.3Ω
C4
100µF
R5
49.9Ω
R4
49.9Ω
C3
100µF
X2 X1 V+ W1
IC1
AD834JN
Y1 Y2 V- W2
R7
499Ω
R8
49.9Ω
IC2
AD811AN
R18
4.7Ω
C16
0.1
C5
0.1
R6
499Ω
C7
0.1
L2
-5V
+5V
INPUT
L3
GROUND
C13
56µF
C15
56µF
L5
-5V

FIG. 1—SCHEMATIC OF THE FREQUENCY DOUBLER. An Analog Devices AD834 wideband, four-quadrant multiplier (IC1) provides wideband analog multiplication of two input signals.

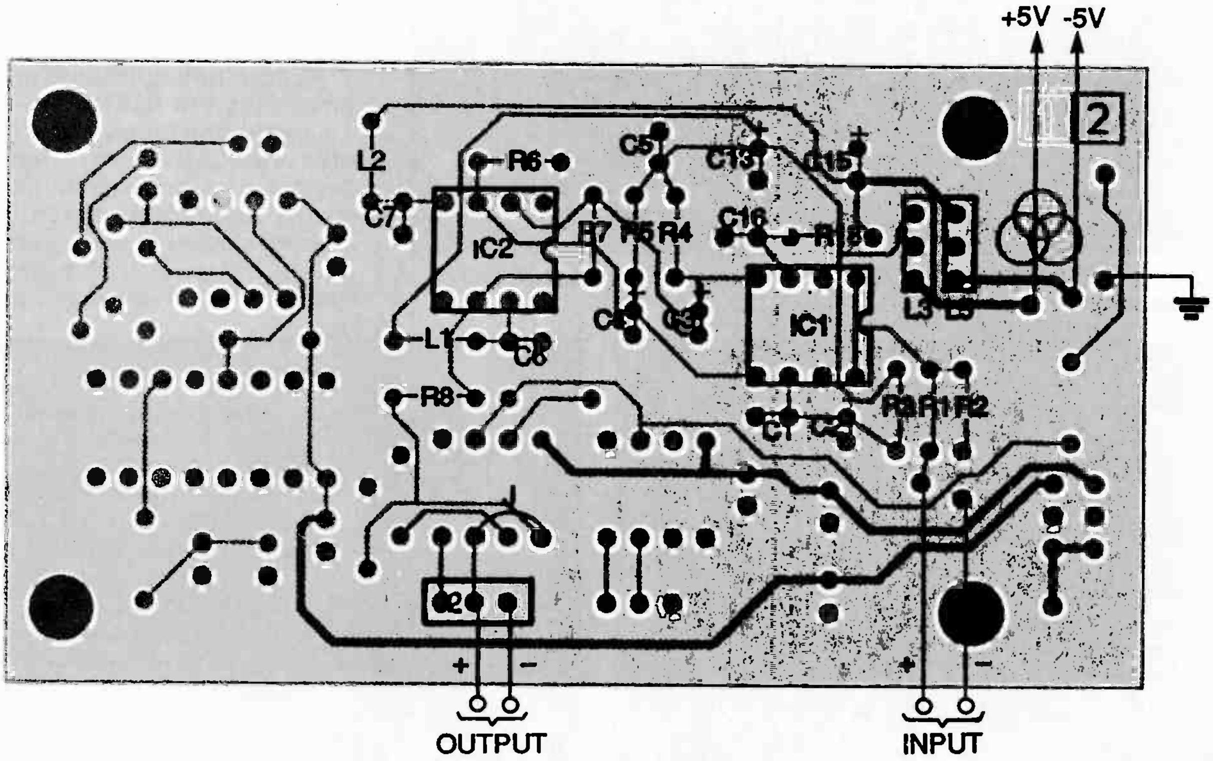
+5V -5V
L2
R6
C5
C13
C15
C7
IC2
R7
R5
R4
C16
R18
L3
L5
L1
C8
C4
C3
IC1
R8
R3
R1
R2
C1
C2
OUTPUT
INPUT

FIG. 2—PARTS-PLACEMENT DIAGRAM. Square pads are on pin 1 of all ICs and on the positive leads of all polarized capacitors.

cy doubling function. The cosine term exhibits a 90° phase shift from the original input, but phase is of no concern in this circuit. Note, however, that the amplitude is cut in half.

Since the AD834 provides differential current-mode outputs, load resistors R4 and R5, gain resistors R6 and R7, and IC2 are added to convert the signal to an amplified and buffered single-ended output. The AD811 wideband current-mode feedback amplifier (IC2) is configured as a differential amplifier. Capacitors C3 and C4 remove the DC offset from the output. Resistor R8 provides a nominal 50-ohm output impedance, and R1 and R2 attenuate the input signal and provide a low-impedance source to IC1.

Inductors L1, L2, L3, and L5 and capacitors C5, C6, C7, C13, and C15 perform high-frequency filtering and decoupling. Resistors R3 and R18 and capacitors C1, C2, and C16 provide decoupling and biasing for IC1.

The circuit requires a +5-volt supply at about 30 milliamperes and −5 volts at about 40 milliamperes. Higher voltages, up to ±9 volts, can be used if the bypass capacitor voltage ratings are increased. Higher voltages provide higher bandwidths with higher power dissipation.

PARTS LIST

All resistors are ⅛-watt, 1%, metal film, unless noted.
R1—17.8 ohms
R2—42.3 ohms
R3—61.9 ohms
R4, R5, R8—49.9 ohms
R6, R7—499 ohms
R18—4.7 ohms, ⅛-watt, 5%

Capacitors
C1, C5–C7, C16—0.1 µF, ceramic
C2–C4—100 µF, 6.3 volts, high-frequency aluminum electrolytic
C13, C15—56 µF, 16 volts, high-frequency aluminum electrolytic

Semiconductors
IC1—AD834JN wideband multiplier (Analog Devices)
IC2—AD811AN current mode amplifier (Analog Devices)

Other components
L1, L2—Leaded EMI bead (Panasonic EXC-ELSA35, Digi-Key P9820BK-ND)
L3, L5—3-terminal EMI filter (Panasonic EXC-EMT103DT, Digi-Key 9809CT-ND)
L4—not used

Miscellaneous: PC board, wire jumpers, seven 0.025-inch square posts (if desired), solder.

Note: The following items are available from Novatech Instruments, Inc., 1530 Eastlake Ave. East, Suite 303, Seattle, WA 98102:

- **Complete kit of all parts (Model DOUB-1, includes PC board and documentation)—$50.00**

Please add $5 shipping and handling for US and Canada. $10 overseas. Add $10 for COD orders. Washington State residents must add 8.2% sales tax. Check or money order only.

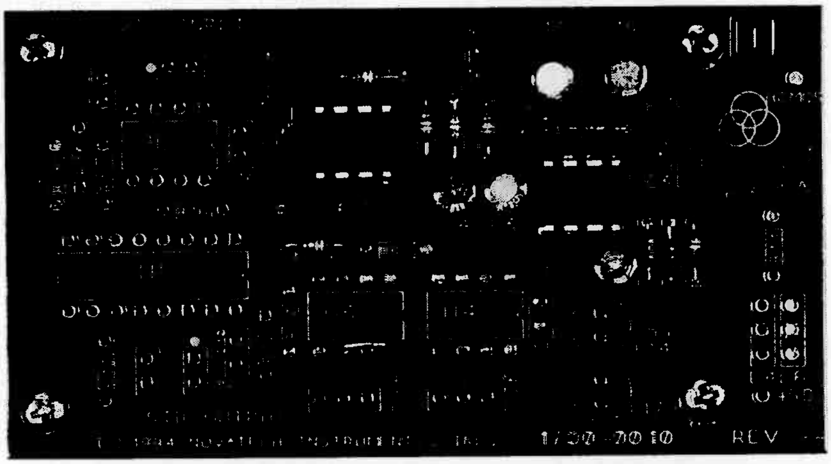

FIG. 3—THE COMPLETED BOARD. The locations for extra components are for other board applications.

Construction

Because IC1 and IC2 are high-frequency components, you must be extremely careful to prevent high-frequency oscillations if you build the circuit on a breadboard. It is preferable to use a PC board. A foil pattern is provided here. Keep all wires short and place the bypass capacitors as close to IC1 and IC2 as possible. A solid ground plane on the component side of the PC board is recommended, and IC sockets are discouraged. If you must use sockets, use machined-pin, open-frame sockets only.

As can be seen from the parts-placement diagram, the board has locations for extra components not related to the doubling function discussed in this article. The additional features were included to allow the board to be used with other Novatech synthesizers.

The parts-placement diagram is shown in Fig. 2. Install all the components as shown, taking care to observe polarity on diodes and capacitors. Square pads indicate pin 1 of all ICs and the positive leads of all polarized capacitors.

All resistors specified for this circuit (except R18) are ⅛-watt, 1% metal-film types. Their small size and tight tolerance improves high-frequency performance and circuit matching. If resistors other than metal-film are substituted, the performance of the frequency doubler

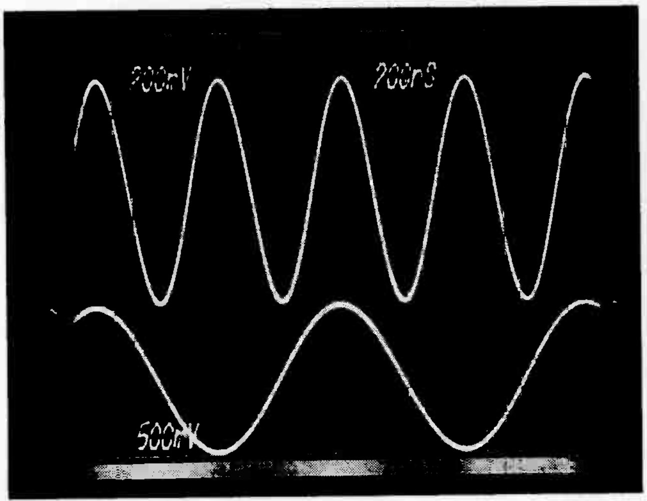

FIG. 4—INPUT AND OUTPUT. The top trace is the output of the doubler with a 1-megahertz, 1.5-volt peak-peak input signal, and the bottom trace is the input signal.

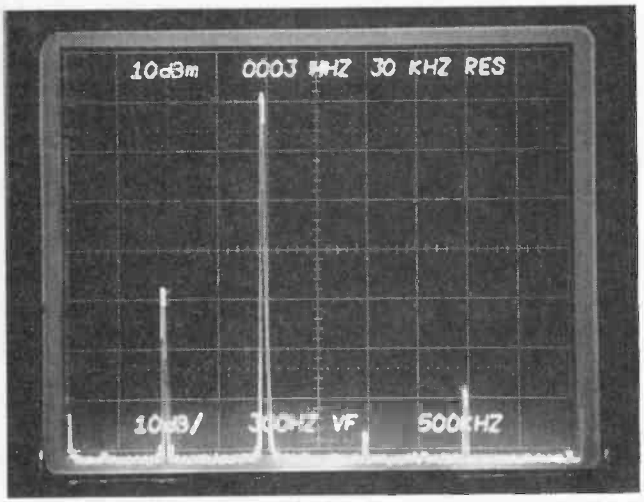

FIG. 5—FREQUENCY SPECTRUM of the output signal in Fig. 4. The fundamental frequency is about 40 dB down from the doubled frequency.

will be degraded. In particular, its distortion will increase, and its bandwidth will decrease.

As mentioned before, C3 and C4 provide DC blocking. The values specified allow operation down to about 10 hertz. If you intend to use this circuit exclusively above 5 kilohertz, better high-frequency performance can be obtained by substituting 0.1 μF ceramic capacitors. The extra capacitors are included with the kit for this purpose. Figure 3 shows the board.

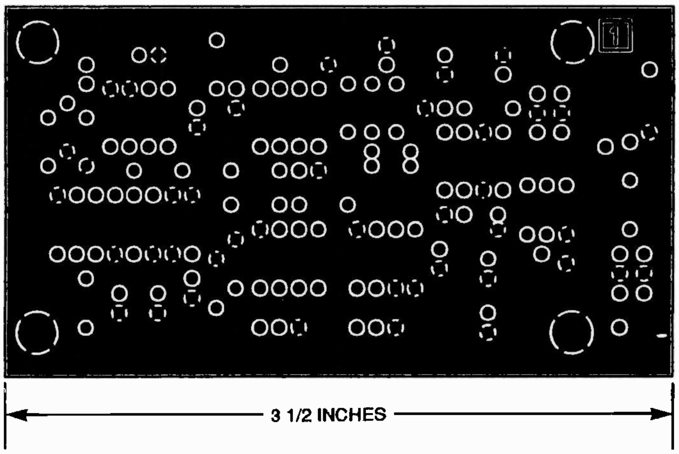

COMPONENT SIDE FOIL PATTERN.

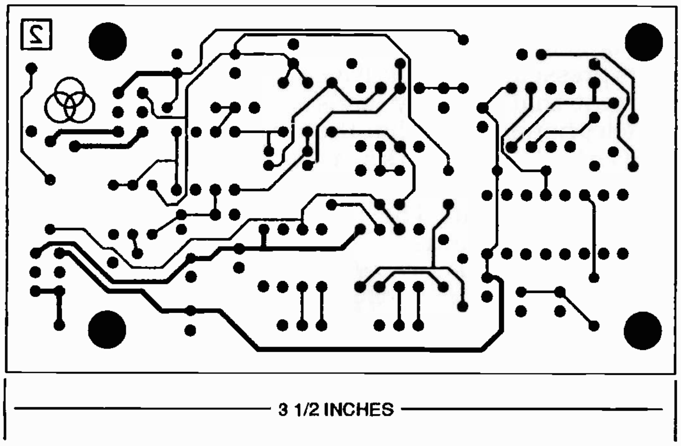

SOLDER SIDE FOIL PATTERN.

Checkout

Verify that all components are installed properly and that there are no solder bridges. Apply power to the board as indicated in Fig. 2 (plus and minus 5 volts DC, both ± 0.25 volt). Also connect a function generator set to approximately 1 volt RMS to the signal input as shown. Connect an oscilloscope or frequency counter to the output pin (the center pin of J2).

Observe that the output is double the frequency of the input signal as the input is varied. The typical 3 dB point of the doubler is greater than 30 megahertz at the output. You can adjust the input to vary the output, but with an input above about 1.5 volts RMS, the output might be clipped.

The top trace in Fig. 4 is the output of the doubler with a 1-megahertz, 1.5-volt peak-to-peak input signal, shown at 200 millivolts per division. The bottom trace is the input signal at 500 millivolts per division. Figure 5 shows the frequency spectrum of the output signal in Fig. 4. The fundamental frequency is about 40 dB down from the doubled frequency, the second harmonic is more than 65 dB down, and the third is about 60 dB down. This fundamental frequency feedthrough increases to about −20 dB below the carrier as the output is increased to 50 megahertz. Ω

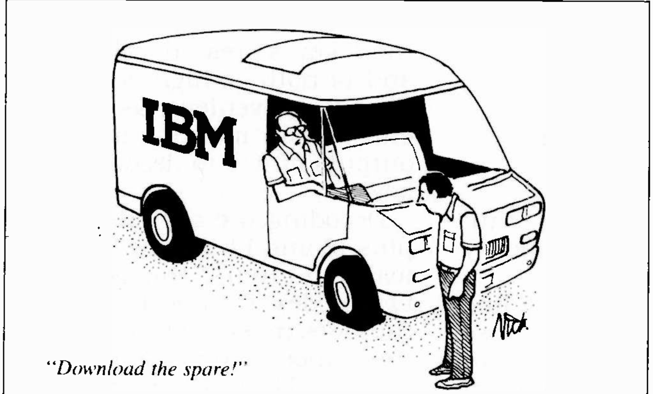

"Download the spare!"

IN STOCK

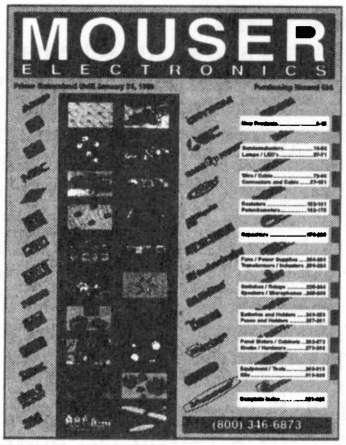

FREE 324 page catalog

- Passive Components
- Electromechanical
- Connectors/Cable
- Semiconductors
- Surface Mount and Through Hole

59,955 Products
120 Manufacturers
All Orders Ship Same Day

MOUSER® ELECTRONICS

800-992-9943
FAX: 817-483-0931

958 North Main St., Mansfield, TX 76063

PUT THAT PHONE ON HOLD!

Build a telephone hold circuit that works from any phone in your home—even cordless!

FRANK MONTEGARI

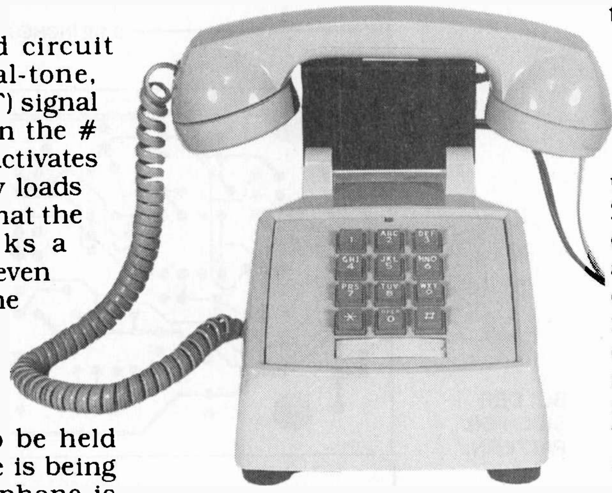

IF YOU HAVE TOUCH-TONE TELEphone service, you can now put a call on hold from any phone in your house by plugging this simple device into any telephone jack. The universal hold circuit works with any phone that has a keypad with a # key. To put a call on hold, press the # key and hang the phone up. A timer extends the #-key function while you hang up phones that have a keypad built into the handset.

The universal hold circuit first detects the dual-tone, multifrequency (DTMF) signal that is generated when the # key is pressed. It then activates a circuit that partially loads the telephone line so that the central office thinks a phone is still off-hook even after it is hung up. The hold circuit remains active for five seconds after the # key is released, so the key does not have to be held down while the phone is being hung up. When any phone is again picked up, the hold function is canceled.

How it works

Figure 1 shows the schematic diagram of the universal hold circuit. The telephone line is connected to the hold components through bridge rectifier BR1 so that the input is not polarity sensitive. The positive side of the line is always connected to Zener diode D2, and the negative side of the line is always connected to IC2. The telephone line also connects to tone-decoder IC4 through C4 and T1. Power is supplied to the circuit by a 12-volt DC power adapter and by IC5, a 78L05 5-volt regulator, and is filtered by C2 and C3.

Optoisolator IC1 contains a silicon-controlled rectifier (SCR) that latches on after it's turned on by current flowing through an optically coupled LED within the IC. The SCR continues to conduct even after the LED current is removed, providing that enough anode current is available to sustain it. When the SCR anode voltage is removed and conduction stops, it will not resume until the SCR is again triggered by the internal LED.

When all phones on the line are on-hook, the voltage across the telephone line is about 48 volts. When a phone is taken off-hook, it places a load on the line and a current of about 20 milliamps flows through the phone, which causes the line voltage to drop to about 3 to 8 volts, depending on the telephone. Current also flows through circuits at the central office, indicating that a phone has been picked up.

A call is put on hold when all phones are hung up and the SCR in IC1 is triggered on by circuits driving its LED. When the SCR conducts, current flows through BR1, D2, LED1, IC1, IC2 and back through BR1, placing a load on the line that keeps the central-office circuits active. A 15-volt drop across D2, a 2-volt drop across D1, a 1.5-volt drop across the LED in IC2, and about a 1.5-volt drop across the diodes in BR1 result in the normal 48-volt line voltage being clamped to about 20 volts. That allows line current to flow, but at a higher voltage than normal. The LED lights to indicate that a call is on hold. When a phone is picked up and the line voltage drops to 8 volts or less, there is no longer enough voltage to keep current flowing through D2, D1 IC1, IC2, and BR1, and the hold function is canceled.

The telephone line also connects through capacitor C4 and transformer T1 to capacitor C8, which couples the DTMF signal to the input of IC4, a Motorola MC145436 DTMF decoder used to detect the #-key tone. Zener diode D6 clamps IC4's input voltage to 9.1 volts to prevent damage from transients. The decoder has four outputs that produce a hexadecimal code corresponding to the tones it receives when a key is pressed (see Table 1). Notice that only when the # key is pressed will pins 13 and 14 both go high (logic "1"). That is convenient, as the circuit has to monitor only two outputs of IC4 to decode the # key.

Decoding the outputs of IC4 pins 13 and 14 is done by a logical AND circuit consisting of two diodes, D4 and D5, whose cathodes must both be high for their anodes to be pulled high by R5. That discharges C1 through R4 into the positive

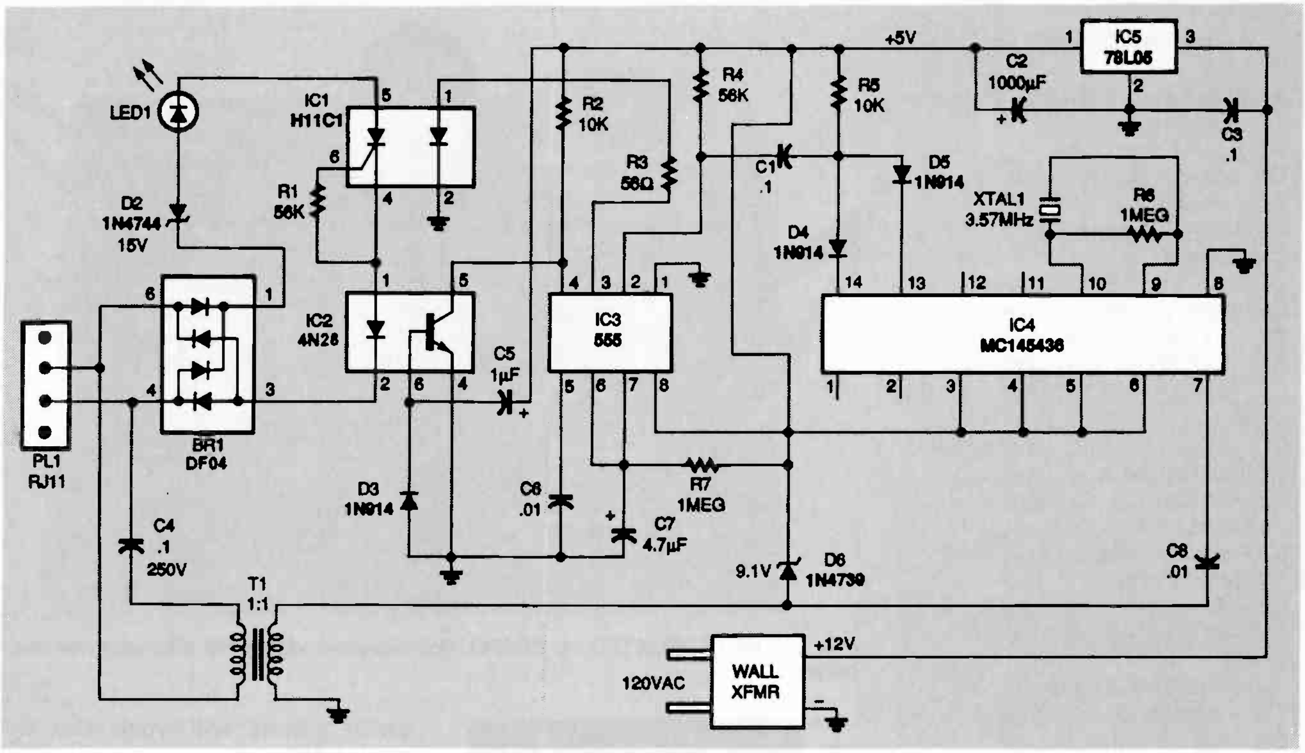
+5V
IC5
78L05
1
2
3
C2
1000µF
C3
.1
LED1
IC1
H11C1
R2
10K
R4
56K
R5
10K
D5
1N914
C1
.1
R3
56Ω
XTAL1
3.57MHz
R6
1MEG
D2
1N4744
15V
R1
56K
D4
1N914
IC2
4N28
IC3
555
IC4
MC145436
C5
1µF
PL1
RJ11
BR1
DF04
D3
1N914
C6
.01
C7
4.7µF
R7
1MEG
C4
.1
250V
9.1V
D6
1N4739
C8
.01
T1
1:1
+12V
120VAC
WALL
XFMR

FIG. 1—SCHEMATIC DIAGRAM of the universal hold circuit. The telephone line is connected to the hold components through bridge rectifier BR1, so that the input is not polarity sensitive.

supply. When the # key is released, pins 13 and 14 go low again and a negative pulse is produced by C1 and R4 at the trigger input (pin 2) of 555 timer IC3. That starts IC3's timing cycle where pin 3 goes high for about five seconds. That drives current through R3 into the LED within IC1, latching the internal SCR. The five-second output from IC3 keeps current flowing long enough to hang up the phone. The 5-second duration is controlled by R7 and C7 and may be altered by changing the value of R7.

Optoisolator IC2 has a transistor output. After the # key is pressed and the phone is hung up, telephone line current flowing through the LED in IC2 turns its transistor on, which resets timer IC3.

The reset function removes current from the LED in IC1 and allows a hold to be canceled immediately after initiating it without having to wait for the five-second cycle to complete. The transistor within IC2 also performs a power-on reset by having its base coupled through C5 to the positive supply. When

TABLE 1—IC4 OUTPUTS

Key	IC pin 13	IC pin 14	IC pin 1	IC pin 2
1	0	0	0	1
2	0	0	1	0
3	0	0	1	1
4	0	1	0	0
5	0	1	0	1
6	0	1	1	0
7	0	1	1	1
8	1	0	0	0
9	1	0	0	1
0	1	0	1	0
*	1	0	1	1
#	1	1	0	0

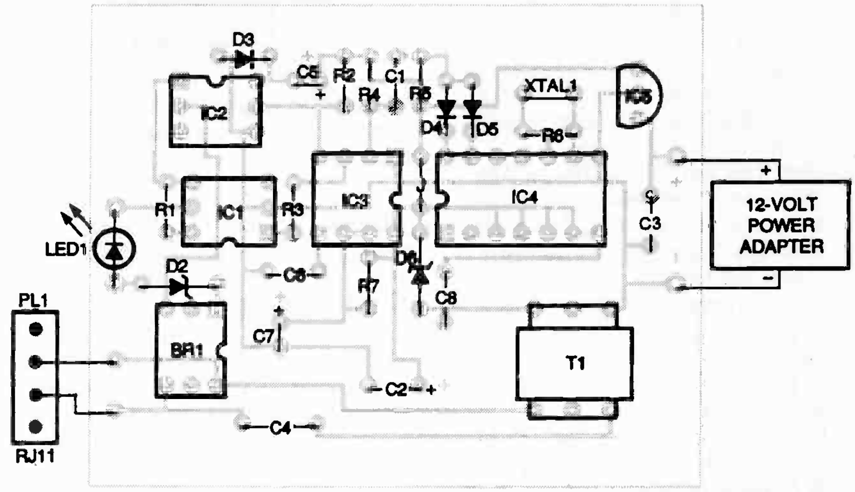
D3
C5
R2
C1
R4
R5
D4
D5
XTAL1
IC5
R6
IC2
IC1
IC3
IC4
R1
R3
C3
LED1
D2
C6
D6
R7
C8
12-VOLT
POWER
ADAPTER
PL1
BR1
C7
T1
C2
C4
RJ11

FIG. 2—PARTS-PLACEMENT DIAGRAM. Leave LED1's leads long and bend them in an "S" shape so its position can be adjusted later to fit through a hole in the top of the case.

PARTS LIST

All resistors are ⅛-watt, 5%, unless noted.
R1, R4—56,000 ohms
R2, R5—10,000 ohms
R3—56 ohms
R6, R7—1 megohm

Capacitors
C1, C3—0.1 µF, 50 volts, metal film
C2—1000 µF, 16 volts, electrolytic
C4—0.1 µF, 250 volts, polyester
C5—1 µF, 50 volts, electrolytic
C6, C8—0.01 µF, 50 volts, metal film
C7—4.7 µF, 35 volts, electrolytic

Semiconductors
IC1—H11C1 SCR optoisolator (Harris or equiv.)
IC2—4N28 transistor optoisolator (Harris or equiv.)
IC3—555 timer
IC4—MC145436 (Motorola or equiv.) or CD22204 (Harris or equiv.) DTMF decoder
IC5—78L05 voltage regulator
BR1—DF04 1-ampere, 400 volt, DIP bridge rectifier (Allied Electronics or equiv.)
D1—not used
D2—1N4744 Zener diode, 15-volts, 1 watt
D3–D5—1N914 silicon diode
D6—1N4739 Zener diode, 9.1 volts, 1 watt
LED1—light-emitting diode, any color, T1-¾

Other components
XTAL—3.579545 MHz crystal
T1—matching transformer, 600-ohm primary, 600-ohm secondary (Jameco P/N 105902)
PL1—RJ11 modular plug attached to telephone wire

Miscellaneous: three 6-pin IC sockets, one 8-pin IC socket, one 14-pin IC socket, 120-volt AC to 12-volt DC, 100 milliamp, power adapter, 2- × 3- × 1-inch plastic project box (Jameco P/N 18921), two cable ties, PC board, wire solder

Note: The following items are available from American Electromechanical Inc., 134 Van Voorhis Terrace, Wappingers Falls, NY 12590:

- **Complete kit of parts including PC board and pre-drilled enclosure—$49.95**
- **Etched and drilled PC board only—$6.95**

Check or money order, only. Add $4.00 shipping and handling. NY State residents must add sales tax

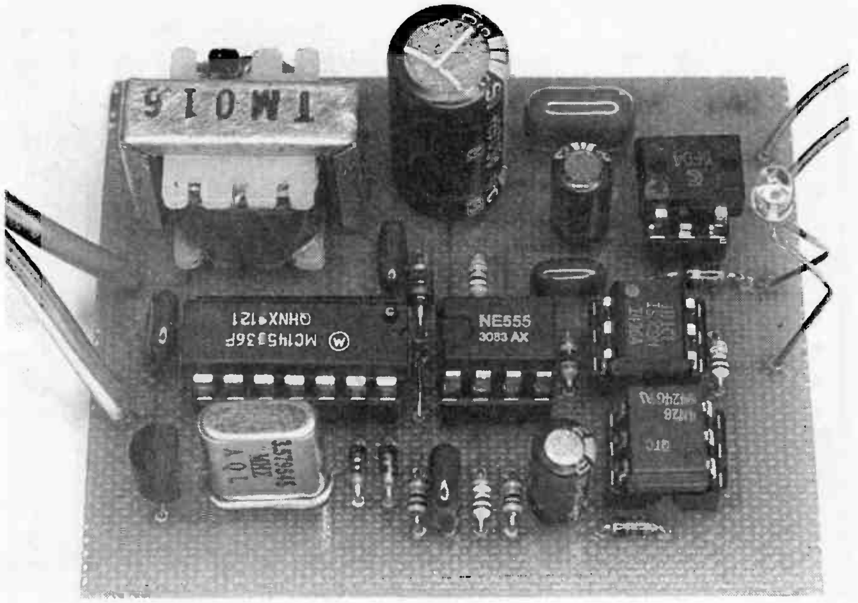

FIG. 3—THE COMPLETED PC BOARD. Point-to-point wiring will also work for this project.

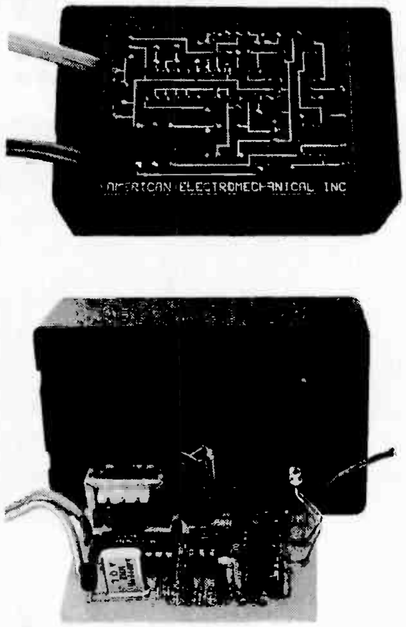

FIG. 4—THE BOARD FITS tightly in a 2- × 3- × 1-inch plastic case, and is held securely without any hardware.

power is turned off, C5 discharges through D3.

Construction

Since this is a low-frequency circuit, parts layout and wiring are not at all critical, parts tolerances are not critical, and equivalent types may be substituted.

The prototype was built on a small PC board for which we've provided a foil pattern. Point-to-point wiring will work just as well, however.

If the printed-circuit board is used, use Fig. 2 as a guide to install the parts. It's always best to use IC sockets for all ICs, although they aren't absolutely necessary.

When installing LED1, leave its leads long and bend them in an S shape so its position can be adjusted later on to fit through a hole in the top of the case. After all parts are installed, plug the ICs into their sockets. Keep in mind that IC4 is a CMOS device that is sensitive to ESD or electrostatic discharge. Figure 3 shows the completed PC board.

The author installed the board in a 2- × 3- × 1-inch plastic case (see Parts List), in which the board fits perfectly and is held securely without any hardware. You are free to use any case you like, but the installation details are up to you if you decide to go that route.

Place a nylon tie around the power leads and the telephone wire to act as strain reliefs, or tie knots in them to serve the same purpose. Cut and file slots in the edge of the case to pass the wires through and drill a hole in the top of the case directly above where the LED will be once the board is installed. Figure 4 shows how the board fits into

(Continued on page 110)

SOLID STATE THERMOMETER

Build an accurate, 3½-digit, handheld thermometer from a kit or from scratch.

MARC SPIWAK

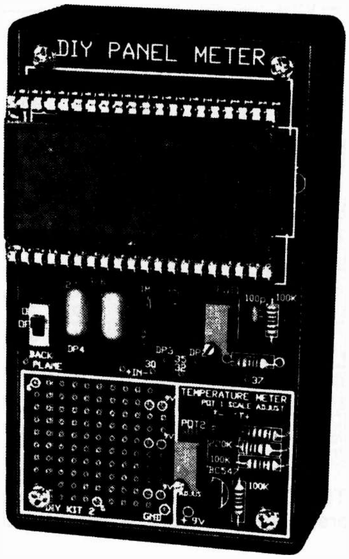

THE LIQUID-CRYSTAL DISplay (LCD) and the single-chip analog-to-digital converter (ADC) have made possible an almost endless list of battery-powered, handheld test instruments. The most prominent of these are the digital multimeter (DMM) and the digital panel meter (DPM), but others include digital capacitance meters, thermometers, timers, sound-level meters, and light meters.

The single chip ADC eliminated a slew of discrete components and increased circuit reliability, while the LCD presents large characters that are easily seen in daylight. But of more importance in battery-powered instruments, both of these components are miserly in power consumption. This is especially important for battery conservation.

Within recent years the prices of single-chip ADC ICs and large-character LCDs have fallen, and they are now readily available as low-cost, off-the-shelf items from electronics stores and distributors.

The digital thermometer described here can be made from an available kit or the parts can be obtained from most electronics distributors. By building the thermometer, the builder will gain experience in working with advanced ICs and display modules, and at the same time gain insight into circuitry common to many different instruments.

The digital thermometer can make accurate, reliable, and repeatable temperature measurements in the laboratory or in the field over a wide range of temperatures. It will save money over the price of a factory-built instrument and, as a bonus, it provides circuit board space for adding extra components for experiments.

The basic digital thermometer circuit displays temperature in degrees Celsius to a an accuracy of a tenth of a degree on its 3½-digit LCD in half-inch high characters. The PC board can be purchased as part of the kit or made with the foil pattern included here. The electronic components are visible on the top surface of the instrument case, as shown in the illustration of the meter.

How does it work?

Figure 1 is a schematic for the digital thermometer. The key semiconductor device in the circuit is IC1, an ICL7106CPL single-chip analog-to-digital converter from Harris Semiconductor. It is a 3½-digit ADC with a built-in LCD display driver, BCD to seven-segment decoders, clock and voltage reference. The ADC is packaged in a 40-pin plastic DIP. In the digital thermometer described in this article, IC1 will be set to display 200 millivolts full scale (199.9), plus or minus.

The temperature sensor for this digital thermometer is an NPN transistor modified to act like a silicon diode. The base of transistor Q1 is short-circuited to the collector, so it functions as a diode. However, it has a response curve that is more linear over a wider temperature range than a diode.

Transistor Q1 provides a variable input voltage to IC1. The voltage drop across the effective diode, typically about 0.7 volt, depends on the temperature of the diode's junction and the current flowing through it. The voltage drop is nearly linear, and will typically vary by 2.2 millivolts per degree Celsius.

A silicon diode has a negative temperature coefficient. Thus, in this application, as the temperature rises, voltage falls, and as the temperature falls the voltage rises. For example, if the ambient temperature rises by 2° Celsius, the voltage drop across the diode will decrease by about 4.4 millivolts. This linear voltage drop is measured to determine temperature.

The kit for the digital thermometer includes two transistors (Q1 and Q2) identified as BC-547s (but marked C547B). Pin identification is shown in Fig. 1. However, more readily available, industry-standard 2N2222 transistors can be substituted, although they have different pinout arrangements. Refer to the instructions given later if you want to make this substitution.

Trimmer potentiometers R9

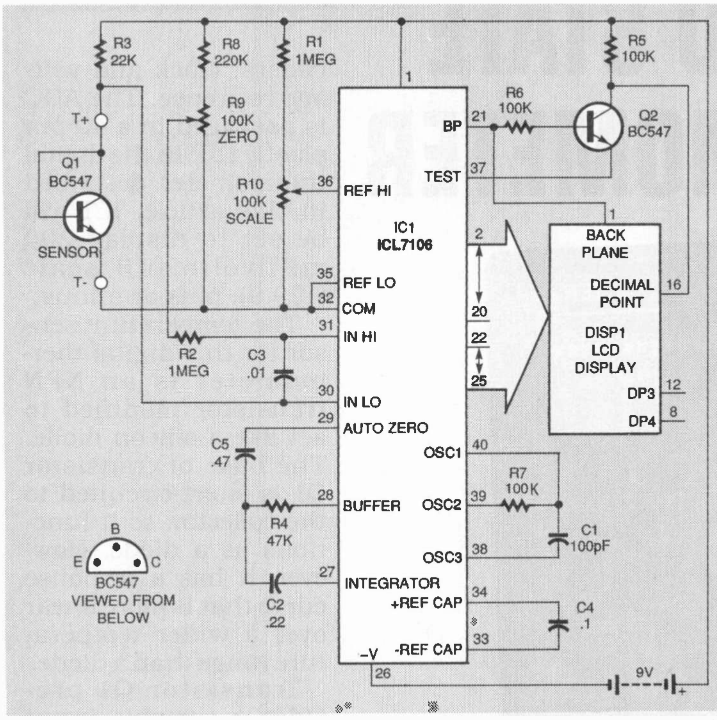
R3
22K
R8
220K
R1
1MEG
R5
100K
R9
100K
ZERO
R10
100K
SCALE
T+
T-
Q1
BC547
SENSOR
R6
100K
Q2
BC547
BP
TEST
REF HI
IC1
ICL7106
REF LO
COM
IN HI
IN LO
AUTO ZERO
OSC1
BUFFER
OSC2
OSC3
INTEGRATOR
+REF CAP
-REF CAP
-V
BACK
PLANE
DECIMAL
POINT
DISP1
LCD
DISPLAY
DP3
DP4
R2
1MEG
C3
.01
C5
.47
R4
47K
R7
100K
C1
100pF
C2
.22
C4
.1
9V
B
E
C
BC547
VIEWED FROM
BELOW

IC1	DISP1
2	18
3	19
4	20
5	21
6	22
7	23
8	17
9	14
10	15
11	24
12	25
13	26
14	13
15	10
16	29
17	31
18	9
19	3
20	2
22	32
23	30
24	11
25	27

FIG. 1—SOLID-STATE THERMOMETER SCHEMATIC. The ICL7106 contains an analog-to-digital converter, BCD to 7-segment decoders, display drivers, a clock, and a reference voltage source.

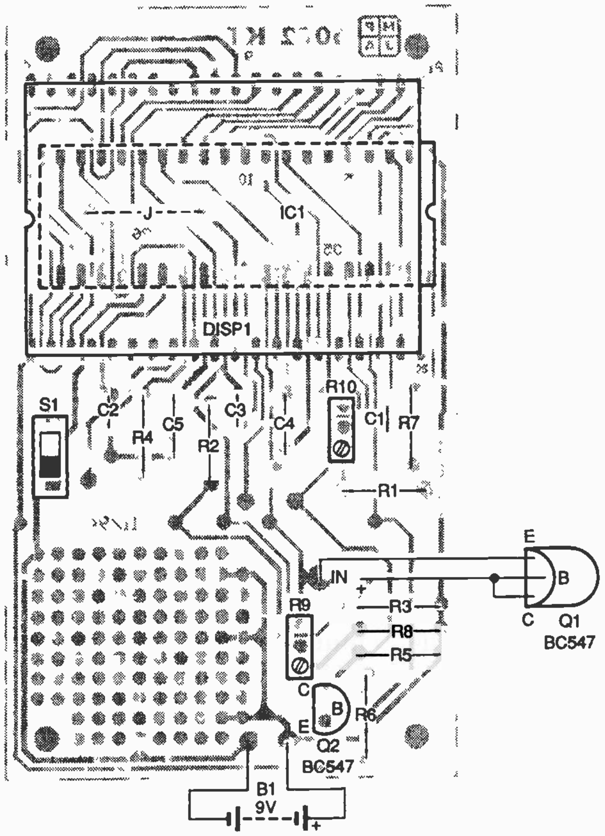
IC1
DISP1
J
S1
C2
R4
C5
C3
R2
C4
R10
C1
R7
R1
IN
R9
R3
R8
R5
E
B
C
Q1
BC547
C
B
E
R6
Q2
BC547
B1
9V

FIG. 2—PARTS-PLACEMENT DIAGRAM. IC1 is located beneath the liquid-crystal display module.

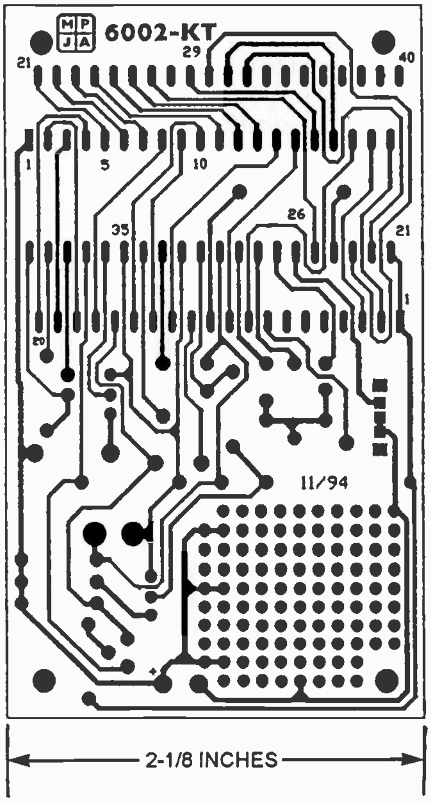
6002-KT
MP
JA
21
29
40
1
5
10
35
26
21
1
20
11/94
2-1/8 INCHES

THERMOMETER FOIL PATTERN.

(ZERO control) and R10 (SCALE control) can be set to zero the meter and make scale adjustments in the input voltage to IC1. A display of 00.0 must correspond to zero degrees Celsius and the display temperature must increment in degrees Celsius. The multiturn potentiometers permit precise settings.

Because IC1 can also indicate the polarity of the input voltage, the digital thermometer is theoretically capable of displaying temperature measurements from −200 to +200 degrees Celsius. However, the transistor used as a temperature sensor and its leads could be damaged by those temperature extremes.

The digital thermometer is effectively a solid-state voltmeter that could be adapted for making other measurements such as voltage, current, power, light level, and sound intensity, and more. If the output of the sensor for the variable you want to measure can be scaled between 0 and 200 millivolts, the circuit presented here can display that variable in appropriate standard units with little or no circuit modification.

The PC board for this project contains a small space where additional components can be

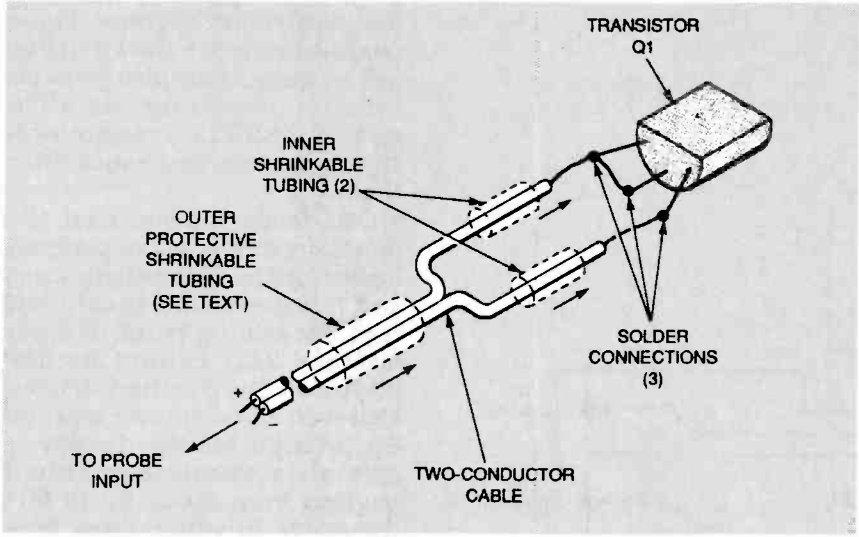
TRANSISTOR Q1
INNER SHRINKABLE TUBING (2)
OUTER PROTECTIVE SHRINKABLE TUBING (SEE TEXT)
SOLDER CONNECTIONS (3)
+
−
TO PROBE INPUT
TWO-CONDUCTOR CABLE

FIG. 3—THERMAL PROBE. The base and collector leads of transistor Q1 are bent together and soldered near the transistor's base and a two-conductor cable is attached.

PARTS LIST

All resistors are ¼-watt, 1%, unless otherwise noted.
R1, R2—1 megohm
R3—22,000 ohms
R4—47,000 ohms
R5–R7—100,000 ohms
R8—220,000 ohms
R9, R10—100,000 ohms, potentiometer 10-turn PCB

Capacitors
C1—100 pF, polyester
C2—0.22 µF, polyester
C3—0.01 µF, polyester
C4—0.1 µF, polyester
C5—0.47 µF, polyester

Semiconductors
IC1— ICL7106CPL ADC, Harris or equiv.
Q1, Q2—BC547 NPN transistor (or 2N2222, see text and Fig. 5)

Other components
S1—slide switch, SPST, PCB-mount
DISP1—VI-302-DP-RC LCD module, 3½-digit, 40-pin package, Varitronix or equivalent)

Miscellaneous: PC board; three 40-pin IC sockets (see text); five machined pin sockets (optional); insulated hookup wire, 9-volt battery, battery clip with wires, project case with cover; miscellaneous hardware; solder.

Note: A kit for the digital thermometer (No. 6002-KT) can be purchased for $29.95 plus $4.00 for shipping and handling from Marlin P. Jones & Associates, Inc., P.O. Box 12685, Lake Park, FL 33403-0685, Phone: 407-848-8236, Fax: 407-844-8764. Florida residents please add local sales tax.

located and powered to perform experiments and make other kinds of measurements.

Figure 2 is the parts placement diagram for this meter. The circuit could be wired point-to-point, but that would call for a lot of wiring, especially in the display section. Consequently, the use of a PC board is recommended.

Insert all leaded components and sockets, and solder them in place on the solder side of the board. Insert resistor R1 and trimmer potentiometer R10 in individual machined pin sockets (two for R1 and three for R10) if you might want to replace those components with others having different values for different applications.

However, if you have no intention of experimenting with the finished meter, solder all resistors directly to the board. If you intend to substitute 2N2222 transistors, insert Q2 with its flat side opposite that shown in Fig. 2. More instructions on the substitution of 2N2222s will be given later.

Notice that analog-to-digital converter, IC1, is located under the LCD display, DISP1. When IC1 is installed, mount the LCD directly above it while orienting it to the opposite polarity (the notches are on opposite sides of the board). First install a socket for IC1. Then cut two 40-pin sockets apart along their long axes to provide four single in-line (SIP) sockets. Insert one SIP socket within another to form two double-height SIP sockets. These will leave enough "headroom" for positioning IC1.

Solder the two double-deck SIP sockets to the board for DISP1, as shown in Fig. 2, and then install IC1. Then insert DISP1 in the double-height SIP sockets. This arrangement makes it easy to remove DISP1 by leaving the upper socket strips attached to the LCD and prying them away from the two lower socket strips that are soldered to the board. CAUTION: Do not attempt to remove the LCD from the upper sockets because the display pins can bend, and the glass body can break.

Cut a scrap of tinned lead wire to form jumper "J," insert and solder it. If you buy the digital thermometer kit, a case is included. The PC board fits securely in the top of the case to form a cover. Therefore, the inside the case provides enough space for the 9-volt battery. Solder the battery clip leads to the solder side of the board at the points shown in Fig. 2.

Refer to Fig. 3 and short-circuit QI's base lead to its collector lead near the body of the transistor by bending them together and soldering them. Slip short lengths of heat shrinkable tubing over each input end of a two-conductor cable. Solder the two shorted transistor leads to the positive sensor input wire and

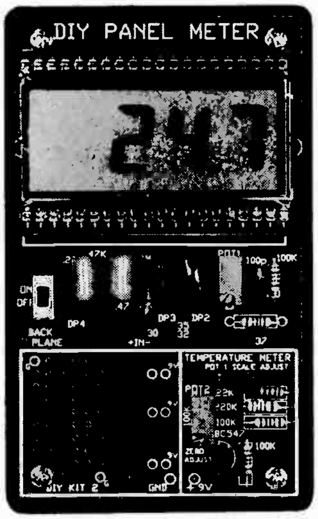

FIG. 4—THE TEMPERATURE-SENSING transistor can be located on the PC board, but leads must be attached for calibration.

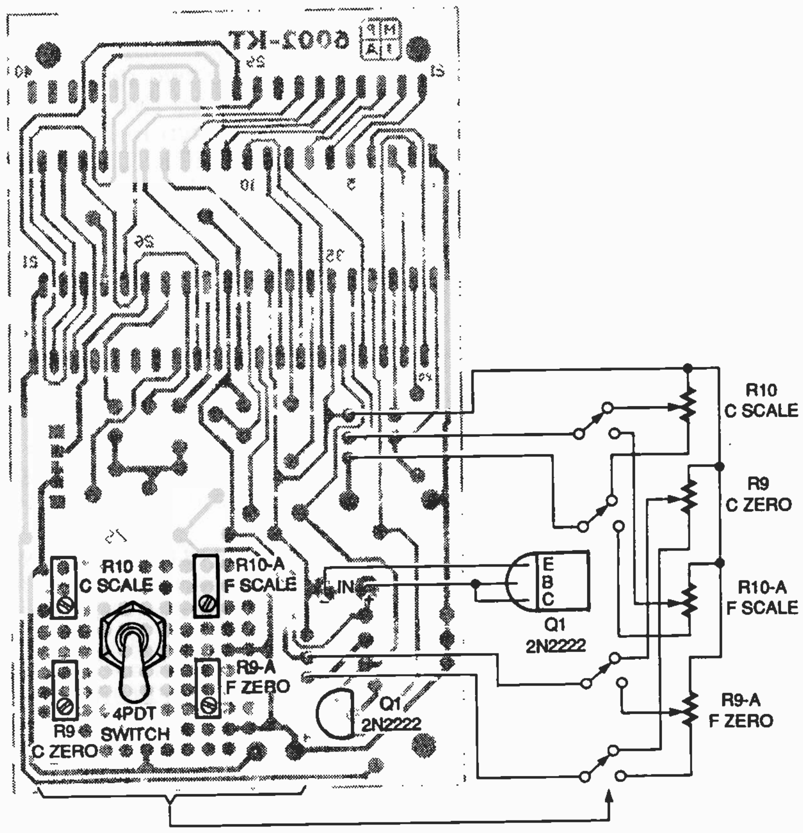

FIG. 5—THE CIRCUIT CAN BE MODIFIED as shown here to display part of the Fahrenheit scale in addition to the Celsius scale. Alternative 2N2222 transistors are installed as shown.

the emitter lead to the negative input wire.

Slip a short length of heat-shrinkable tubing on both leads and pull it over the end of the transistor to form an outer protective jacket, as shown in Fig. 3. Heat the tubing so that it forms a tight shrink fit around both the end of the transistor and the cable.

Standard insulated hookup wire leads are suitable for the temperature probe up to about 6 inches long. However, if the probe is to be mounted on leads longer than 6 inches, use thin shielded cable.

The temperature probe can also be mounted on the PC board directly with leads soldered to the "+" and "−" probe inputs. This arrangement will, however, make thermometer calibration more difficult. If you elect this option, temporarily attach the probe to the board with leads for calibration and mount it permanently to the board afterward. Figure 4 shows the completed digital thermometer.

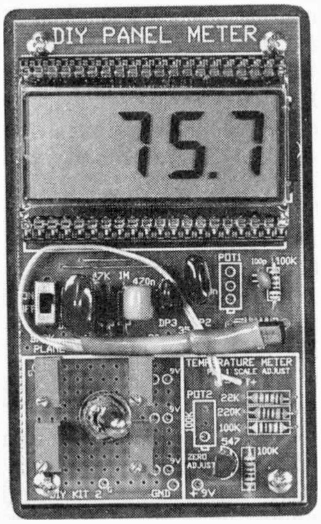

FIG. 6—MODIFIED THERMOMETER can display the full Celsius range and Fahrenheit from about 60° to about 80°.

Modifying the meter

As mentioned earlier, a small area on the PC board is set aside to accommodate components for experiments. We added two more potentiometers and a 4PDT switch to the circuit so the digital thermometer could be calibrated in Fahrenheit as well as Celsius degrees. Those modifications are shown in Fig. 5. The same figure also gives details for installing the alternative 2N2222 transistors, should you decide to substitute them.

Our tests showed that the meter's response is not perfectly linear for the Fahrenheit scale, and it is impossible to calibrate it at the boiling point of water because 212° cannot be displayed on the 3½-digit display. However, the response was linear enough for the display to give an accurate Fahrenheit reading from about 60 to 80°, the range in which most temperature readings are likely to be taken.

Calibrating the meter

Carefully inspect the circuit board for soldering errors and correct them before proceeding. If the board passes visual inspection, connect a 9-volt battery to it and switch on power. The display will either indicate a "1" or some other reading. If the display shows a "1," adjust potentiometers R9 and R10 to obtain a reading other than a "1."

Fill a container with ice cubes and add a small amount of water. Turn on the digital thermometer and submerge its temperature probe in the ice water, holding it in close contact with an ice cube. Wait until the display stabilizes and then adjust ZERO control potentiometer R9 until the display reads 00.0.

Boil water in a kettle and carefully hold the probe over the kettle spout in the steam with tongs to prevent accidentally scalding yourself. When the display reading stabilizes, adjust SCALE control potentiometer R10 until the display reads 100. If you have access to a digital multimeter with a temperature probe that is known to be calibrated, compare readings and adjust trimmers R9 and R10 more precisely.

If you added the Fahrenheit potentiometers, as shown in Fig. 5, calibrate the Fahrenheit scale next, bearing in mind the limitations discussed earlier. It is recommended that the tem-

(Continued on page 109)

Home-brew Isolation Transformer

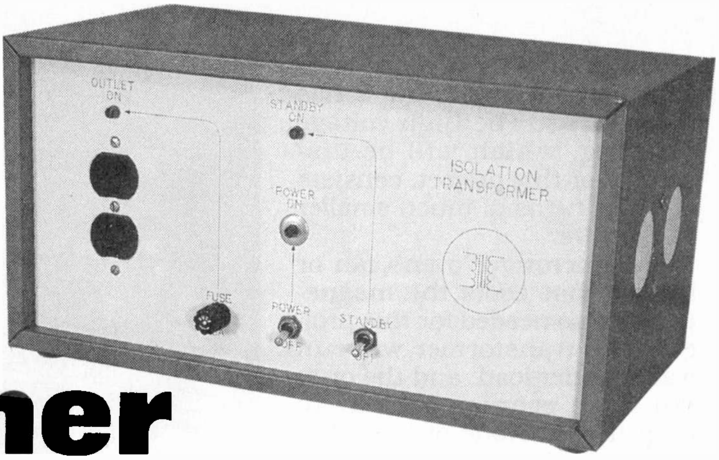

Buils this inexpensive transformer to protect yourself—and your test equipment— while you service and repair electronic appliances.

DOYLE WHISENANT

EVERYONE WHO REGULARLY SERvices television sets really should have an isolation transformer. One side of the chassis of most line-powered TV receivers and tube-type radios is connected directly to the AC line. Therefore, servicing this electronic equipment can be very dangerous because the chassis can be "hot." An isolation transformer isolates the chassis of the equipment being serviced from the AC line. This article explains how to build an isolation transformer for a fraction of the price of a commercial product. It will even supply more current than most commercial units.

The heart of this project is the transformer. The author obtained the two transformers required from two discarded microwave ovens. Buying a new transformer for this project will not save you much money over buying a commercial isolation transformer because the transformer required can cost more than $150. For example, B&K Precision sells an isolation transformer for $189.00. Although electronics distributor Mouser Electronics sells one for $63.50, it will supply only 2.17 amperes, or about 250 watts. Even this modest price is almost three times what our complete project will cost!

Operation

Figure 1 is the schematic for the isolation transformer. Fuse-protected AC line power is first applied to power switch S1. When S1 is closed, neon power indicator NE1 lights up, and power is applied to a cooling fan and to standby switch S2. This arrangement allows transformer power to be turned off, while permitting the fan to continue cooling the transformer.

Neon standby indicator NE2 is wired across switch S2 so that it lights when S2 is open. When S2 is closed, power is applied to the primary side of 1:1 isolation transformer T1. The secondary side of T1 supplies isolated AC power to receptacle SO1. Neon indicator NE3 lights when power is applied to the receptacle SO1.

The isolation transformer will supply 1000 watts for short periods of time, and 500 watts or less continuously. However, it will run too hot if 500 watts is exceeded for more than about half an hour.

The transformer for this project was made from the windings of two microwave-oven transformers. Microwave ovens have heavy-duty transformers (see Fig. 2) that consist of three windings: a 120-volt AC input coil, a 2000- to 3000-volt AC output coil, and a 3- to 5-volt AC coil that serves as the filament winding for the oven's magnetron. The magnetron is the vacuum tube that produces the microwave energy for the oven.

The filament winding is easily identified; it consists of three to five single turns of 12- or 13-gauge enamel-covered wire. This winding is usually wound directly over the 120-volt AC winding which is wound from

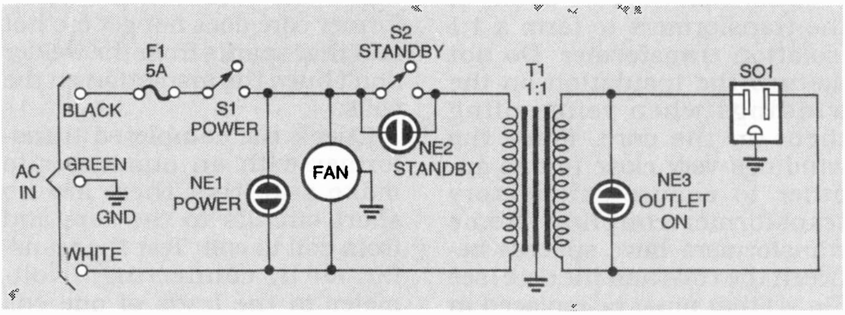

FIG. 1—ISOLATION TRANSFORMER SCHEMATIC. Switch S2 allows power to the transformer to be turned off while the fan continues to cool the transformer.

approximately the same size wire as the filament, but it has more turns. The high-voltage winding, which will be discarded for this project, consists of many turns of much smaller gauge wire.

The microwave oven's fan or blower that cools the magnetron is also needed for this project. The transformer will run warm under load, and the oven fan is an economical way to keep it cool. None of the other parts (see the Parts List), while commonly available, will be found in a microwave oven.

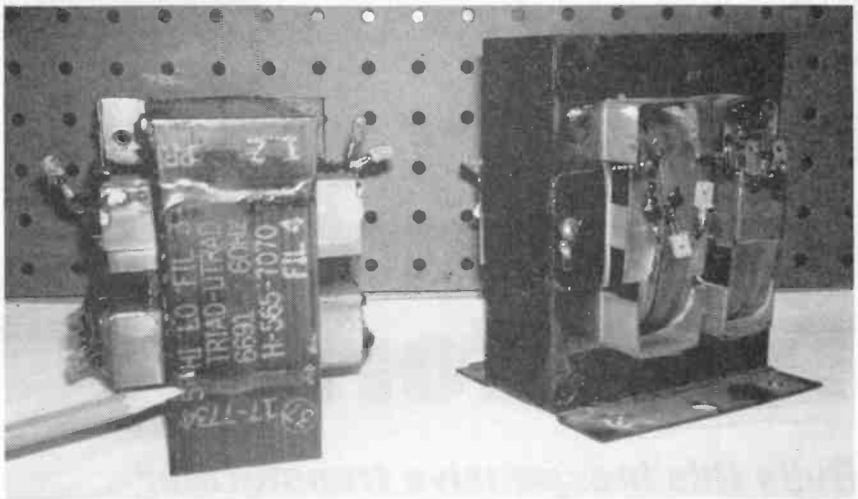

FIG. 2—THE TRANSFORMER for this project was made from two discarded microwave-oven transformers.

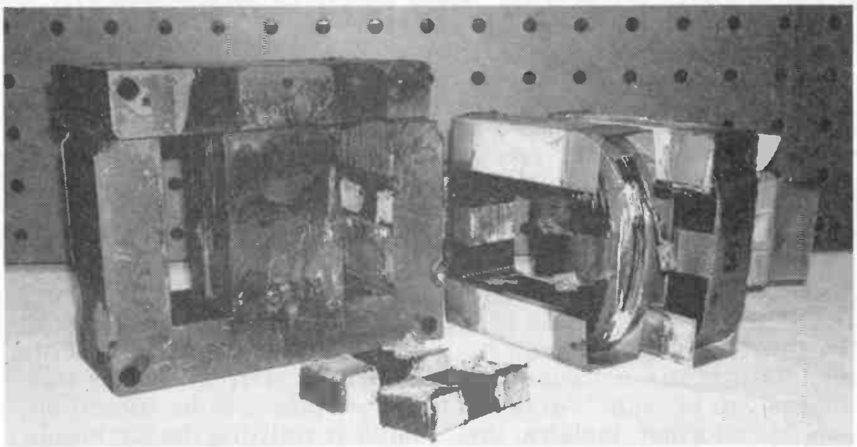

FIG. 3—DISASSEMBLED TRANSFORMER. Remove the windings from the transformer core, being careful not to disturb the insulation on the windings. Discard the high-voltage windings.

Modifying the transformer

Locate two surplus or discarded microwave ovens that are identical, or as similar as possible. As stated earlier, the transformers from two ovens are needed to build the isolation transformer. These transformers usually remain in good working condition. As an alternative, two surplus transformers might be easier for you to obtain than two scrap microwave ovens.

The pencil in Fig. 2 points to one of the welds that must be removed from the microwave-oven transformer. With a handheld grinder, very carefully grind away the welds on both sides of the transformer and separate the top core from the base. Figure 3 shows a transformer that has been disassembled. Remove the windings from the core, being careful not to disturb the insulation on the windings. Discard the high-voltage windings.

The 120-volt AC input coils from the two transformers will be used to make the isolation transformer. Place both 120-volt coils back on the core of one of the transformers to form a 1:1 isolation transformer. Do not disturb the insulation on the windings when reinstalling them on the core. Place the windings very close to one another to ensure satisfactory transformer coupling. Some transformers have spacers between the coils and the core (see Fig. 4) that must be replaced in the correct locations.

Once the windings are correctly positioned back on the core, reinstall the base of the transformer. To do this, weld the base back onto the transformer. If you have access to an electric welder and know how to use it, do this yourself. Otherwise, take the transformer to a welding shop and let a skilled welder do the job for you. Regardless of who does the welding, make sure that the transformer core does not get too hot and that sparks from the welder don't burn the insulation on the coils.

Check the completed transformer with an ohmmeter to make sure that there are no short circuits to the core and from coil to coil. Test the transformer by connecting a voltmeter to the leads of one coil designated as the output coil. Connect the leads from the other coil to an AC line cord through a 5-ampere fuse. Insulate all exposed connections.

Power up the transformer just long enough to get a reading on the voltmeter. Record the output reading, disconnect the AC power, and reverse the input and output coil connections. Repeat the procedure and check the output voltage again. Determine which output coil voltage measures closest to 120-volts AC; it will be the output side. The prototype transformer had a lower voltage reading in one direction than in the other.

Construction

This project requires a sturdy metal case that measures approximately 14- × 7- × 9-inches to hold both the fan and the transformer, and still have room for all connections.

This project consists of only a few parts, so point-to-point wiring can be used throughout. Figure 5 shows the inside of the prototype unit. Test fit the fan and the transformer before drilling any holes in the case. The fan should be mounted so that its air stream is directed at the transformer. Once the correct positions are determined for the fan and transformer, mark the locations and drill the mounting holes, but don't mount anything yet.

PARTS LIST

T1—Modified microwave oven transformer (see text)
F1—5-ampere fuse and panel-mount fuse holder
S1, S2—SPST 15-ampere toggle switch
NE1–NE3—120-volt AC neon indicator lamp
SO1—Standard duplex AC receptacle
Metal case (Mendleson Electronics, 1-800-344-4465, Part No. 160-1782F or other suitable enclosure), grounded AC line cord, cooling fan from old microwave oven (see text), 14- to 16-gauge wire, terminal blocks, solderless connectors, wire ties, hardware

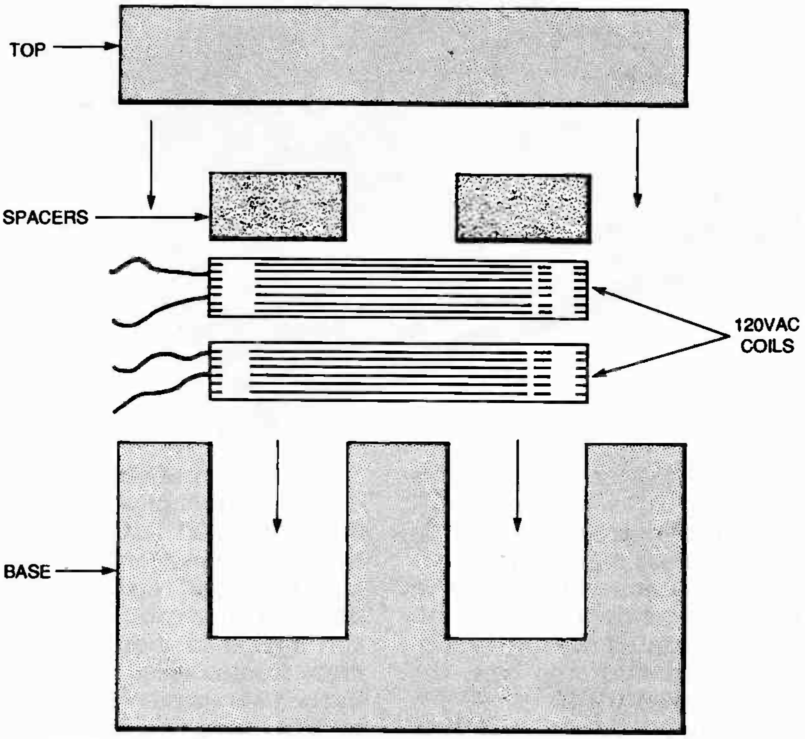

FIG. 4—SOME TRANSFORMERS have spacers between their coils and the core.

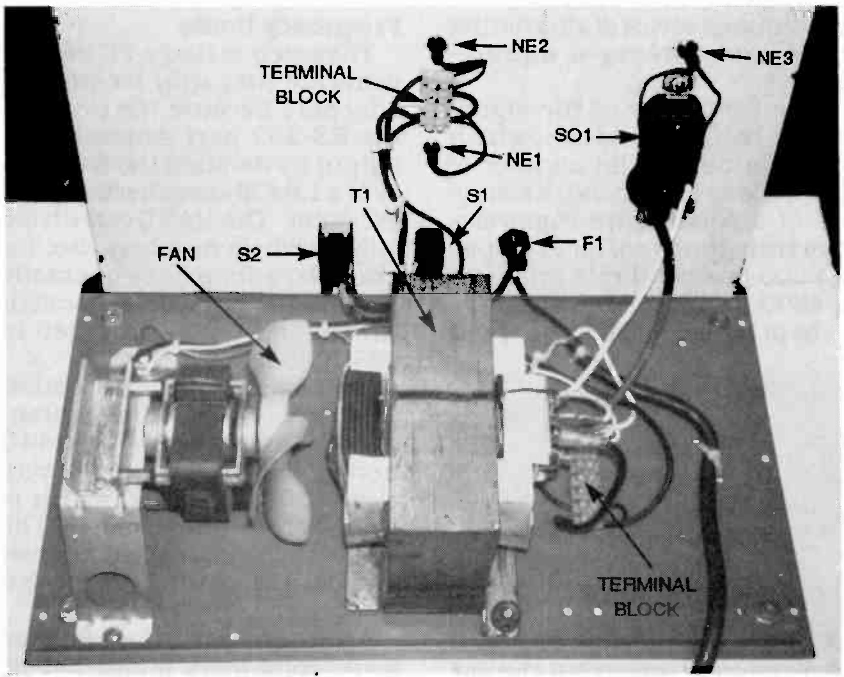

FIG. 5—POINT-TO-POINT WIRING is used throughout this project.

If you purchase the case specified in the Parts List, you will find that many of the necessary holes are already punched out. Otherwise, prepare the enclosure by drilling a hole and mounting the line cord to the back of the enclosure. Prepare the front of the enclosure by drilling holes to mount the neon indicators, NE1, NE2, and NE3. Then drill holes for switches S1 and S2. The prototype unit has two large holes punched in each end to allow the fan to draw in cool air and exhaust hot air. Place a screen over those holes to prevent debris from being drawn inside the unit. Prepare an opening for the duplex receptacle on the front panel with a nibbling tool.

Once all the metal working on the enclosure is complete, it can be painted. Next label the enclosure; the labels can be protected with clear spray enamel.

When the enclosure is completely dry, the transformer and fan can be mounted. Start with the transformer, mounting it to the base of the enclosure with rubber washers to reduce vibration. Mount the fan in the direction so that the cooling air flow will be blowing over the transformer. Next install the fuse holder, the line cord, the three neon indicators, switches S1 and S2, and the duplex receptacle SO1.

The isolation transformer can produce several amperes of current, so use 14- to 16-gauge wire for all connections to and from the transformer. However, smaller gauge wire can be used for connecting the fan and neon indicators.

Refer to the schematic di-

(Continued on page 110)

PRECISION AUDIO SIGNALS FROM YOUR PC

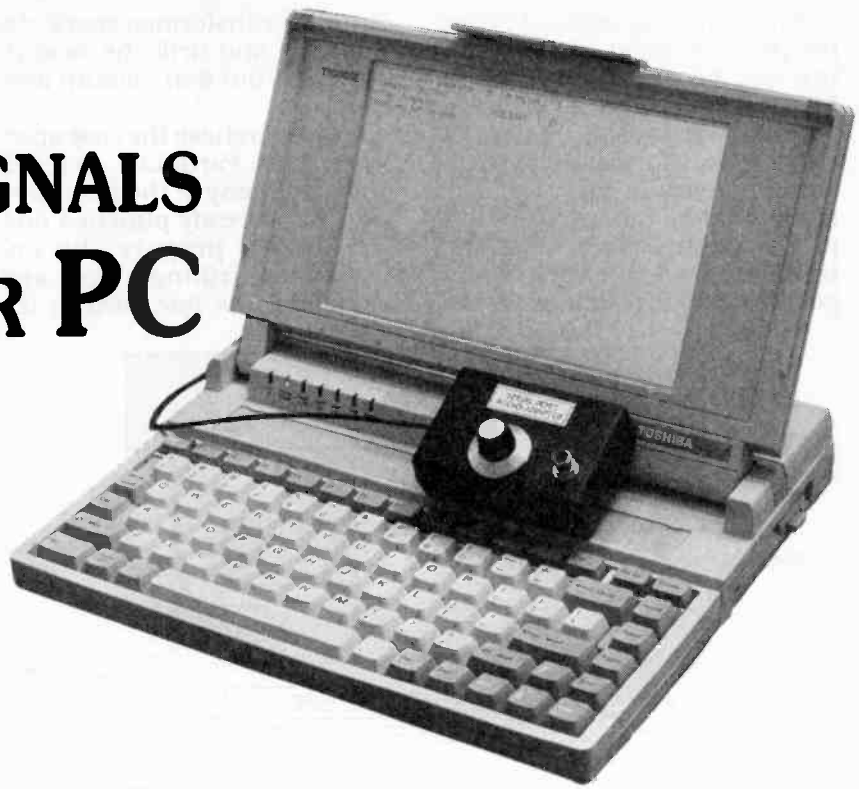

Build this computer-controlled audio generator—for under ten bucks.

MICHAEL A. COVINGTON

DO YOU NEED A PRECISION SQUARE-wave generator for audio testing? If so, look no further than your computer's serial port, a handful of passive components, and a very simple 30-line BASIC program.

That combination can deliver signals with frequencies as high as 4800 hertz, and with crystal-controlled accuracy of 0.1% or better. The software was developed on an IBM-compatible PC, but it should run on just about any computer.

How it works

The trick is U, the ASCII character "U," that is. The hexadecimal value of "U" is 55, which in binary is 01010101 (with eight data bits and no parity, or seven data bits and even parity).

The RS-232 protocol specifies that the bits of an ASCII character are transmitted from least to most significant, preceded by a start bit (always 0) and followed by a stop bit (always 1). So, after adding the requisite start and stop bits, the result is 1010101010.

Now suppose a string of U's is generated at the serial port at some steady rate. The result is a continuous series of alternating ones and zeroes—a squarewave.

The frequency of the signal will be half the baud rate, which by definition is the number of transitions per second. Each cycle of a squarewave comprises two transitions so, for example, a 9600-bps baud rate produces a 4800-hertz squarewave.

In practical terms, just about any computer should be able to deliver frequencies of 55, 150, 300, 600, 1200, 2400, and 4800 hertz, corresponding to the standard baud rates from 110 to 9600. In addition, the signal can appear at many other discrete frequencies, limited only by the CPU speed of the computer. But there is a catch to this scheme.

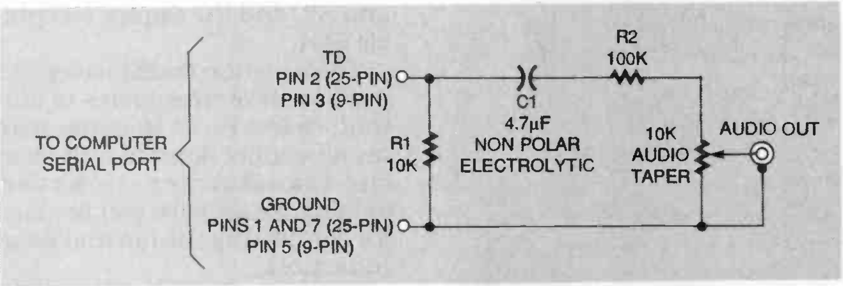

FIG. 1—COMPLETE CIRCUIT of the precision audio generator requires just a handful of passive components.

Frequency limits

The catch is that a PC cannot generate just *any* frequency. Why not? Because the UART in the RS-232 port generates its output by dividing the frequency of a 1.8432-megahertz crystal oscillator. The UART can divide only by whole numbers. So, for example, a frequency of exactly 1000 hertz can't be generated; the nearest you can get is 993.1035 hertz.

If you want to tune a guitar, you can't quite produce a standard concert-pitch "A" (440 hertz), but you can get a very clsoe 439.6947 hertz, which is off by only 1/80 of a semitone. The software will display the nearest standard value to any requested frequency.

Another limitation is that some computers might not be able to deliver high frequencies, because they can't output U's fast enough. In that case, squarewave bursts, with silence in between, is generated. If you could hear that, it might sound like buzzing, flapping, or clicking superimposed on the high-pitched tone. The best way to detect this kind of problem is with an oscilloscope or frequency counter.

Even if the computer is fast, there can be breaks in the squarewave. That can happen if the computer is doing task-switching (e.g., under Windows), or if it is heavily loaded with terminate and stay resident programs (TSRs). But under DOS, the author had good results up to 4800 hertz with an old Toshiba laptop.

Hardware and software

There's not much to the circuit, which is shown in Fig. 1. The output of a serial port is nominally 24 volts peak-to-peak, which is much too high a voltage to feed to the input of an audio amplifier. The circuit attenuates the signal to a more useful level, a variable 2-volts peak-to-peak. The circuit also protects the computer from static electricity and voltage surges. Capacitor C1, a non-polarized unit, blocks DC because the serial port, when idling, outputs approximately −12 volts.

The attenuator consists of only four components, so it does not need a PC board. The circuit was built in a small plastic case by mounting the resistors and capacitors directly to the potentiometer and output jack, as shown in Fig. 2. Figure 3 shows the assembled unit.

PARTS LIST

R1—10,000 ohms, ⅛-watt
R2—100,000 ohms, ⅛-watt
R3—10,000 ohms, audio-taper potentiometer
C1—4.7 µF, 50-volts, non-polarized electrolytic capacitor (Radio Shack 272-998 or equivalent)
Enclosure, cables, and connectors to suit your equipment and needs.

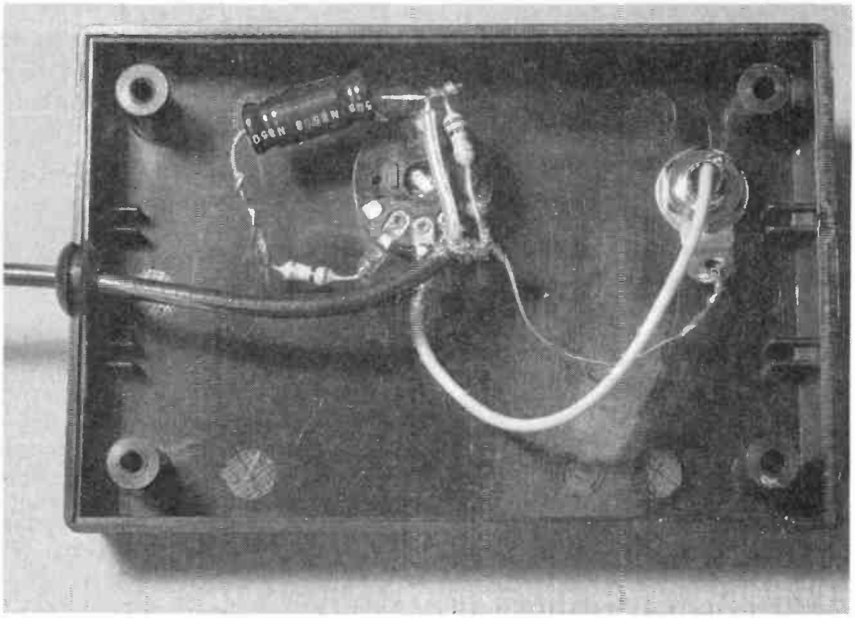

FIG. 2—MOUNT ALL COMPONENTS directly on the potentiometer, as shown here. This eliminates the need for any kind of circuit board.

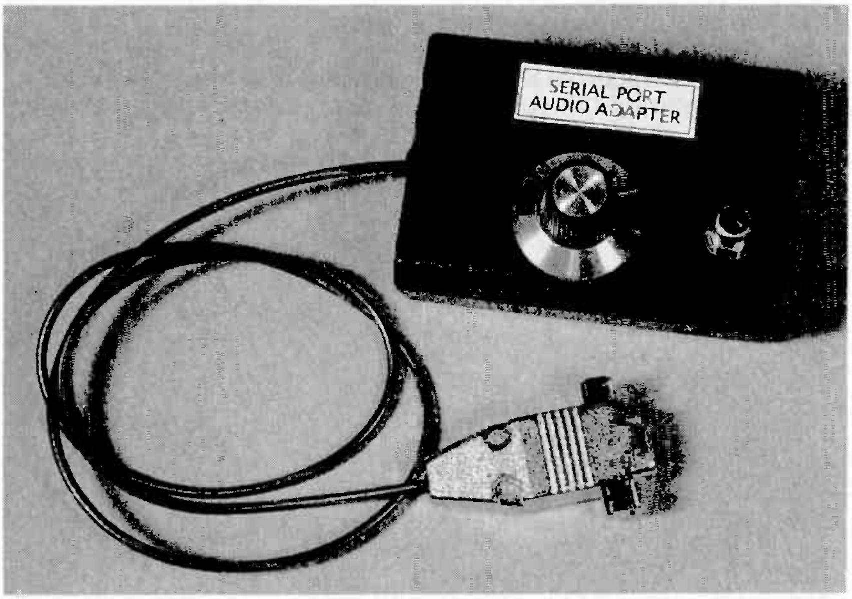

FIG. 3—THE ASSEMBLED UNIT is compact and attractive. An even smaller case than the one shown here could also be used.

If you want the circuit to deliver signals for testing, just run an appropriate cable from J1 to your equipment. For audio output, say for tuning musical instruments or playing audible tones, the circuit can drive a speaker directly, as shown in Fig. 4.

A word of caution before discussing applications: RS-232 ports are supposed to be tolerant of static charges, short circuits, and extraneous voltages, but be aware that some ports are not. Use extra care if your port is part of a multifunction card that also includes a disk controller.

Many of these cards include both the serial port and the disk controller in a single, fragile VLSI device. If part of the circuit fails, you're likely to lose the whole thing—including access to your disk drives. Generally speaking, when experimenting with accessories connected to a serial port, it's safer to use a card with discrete RS-232 transmitter and receiver ICs.

As for the software, Listing 1

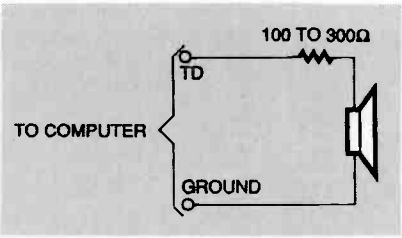

FIG. 4—TO DRIVE A SPEAKER from the generator, use this circuit.

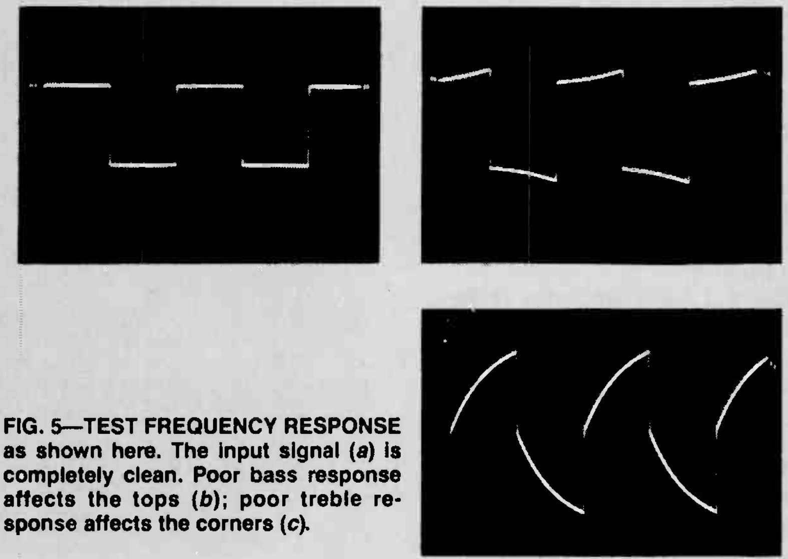

FIG. 5—TEST FREQUENCY RESPONSE as shown here. The input signal (*a*) is completely clean. Poor bass response affects the tops (*b*); poor treble response affects the corners (*c*).

shows the complete program. After setting up several constants, the program requests an output frequency, displays the nearest attainable value, and then starts pumping U's from the serial port. That continues until the user presses a key and then it stops.

The constants defined in lines 120 and 130 specify values for the PC's standard COM1 serial port. To use a different port, make the appropriate modifications. Also note line 190, which specifies the dividend in the frequency calculation. To run the program on another computer, that value might have to be adjusted.

LISTING 1—BASIC PROGRAM

```
100 PRINT "PC Square Wave Generator - M. Covington 1994"
110 INPUT "Frequency (Hz)? ", FREQ
120 ADDR% = &H3F8                      ' 2F8 for COM2
130 PARM$ = "COM1:4800,N,8,1,cs0,ds0" ' or "COM2..." etc.
140 '
150 ' Open the serial port
160 OPEN PARM$ FOR OUTPUT AS #1
170 '
180 ' Choose customized baud rate
190 DIVIDEND = 115200                  ' 111850 on PCjr
200 DIVISOR% = INT(.5 + DIVIDEND / (FREQ * 2))
210 FREQ = DIVIDEND / (DIVISOR% * 2)
220 PRINT "Actual frequency: "; FREQ; " Hz"
230 '
240 ' Set serial port to new baud rate
250 WAIT ADDR% + 5, &H20
260 OUT ADDR% + 3, INP(ADDR% + 3) OR &H80
270 OUT ADDR%, DIVISOR% MOD 256
280 OUT ADDR% + 1, DIVISOR% / 256
290 OUT ADDR% + 3, INP(ADDR% + 3) AND &H7F
300 '
310 ' Transmit square wave until told to stop
320 PRINT "Press any key to stop."
330 WHILE INKEY$ <> "": WEND           ' clear kbd buffer
340 WHILE INKEY$ = ""
350   PRINT #1, "UUUU";
360 WEND
370 PRINT "Emptying buffer..."
380 CLOSE #1
390 PRINT "All done."
400 END
```

Putting it to use

A precision squarewave generator has many applications. For example, you can use it to test the frequency response of an amplifier, as shown in Fig. 5. The input signal to the amplifier appears in Fig. 5-*a*; note the signal's sharp corners and flat tops. If the amplifier circuit has poor bass response, the tops of the waveform won't be flat (Fig. 5-*b*); if the treble is weak, the corners of the waveform won't be square (Fig. 5-*c*).

In general, a clean looking squarewave at the output of an amplifier indicates a good frequency response over a 100-to-1 range. For example, an amplifier that cleanly reproduces a 1-kilohertz squarewave should provide good performance from 100 hertz to 10 kilohertz. Be aware, however, that a squarewave test won't detect clipping in the amplifier.

I have already mentioned tuning musical instruments. Another application is testing wow and flutter in tape recorders. Generate a constant frequency and record it; then play it back while comparing it to the same frequency coming directly from the computer. You can do the comparison by ear, but it's better to use an oscilloscope, with one squarewave going to the external sync input and the other to the vertical input. Then look closely to see how much the waveform jiggles from side to side. Ω

The Project Anthology

A collection of eighteen projects that will add spice to your project building pleasure!

CONTENTS

1 TWINKLE TREE
2 ONE-SECOND FLASHER
3 5-VOLT POWER SUPPLY
4 SOUND-SWITCH
5 LIGHT ALARM
6 SIG-TRACER
7 CW FILTER
8 VOL-EXPANDER
9 COMPUTER VOICE
10 AUDIO POWER BOOSTER
11 PIT STOP
12 MINEFIELDS
13 KAPELLMEISTER
14 HELIX
15 INVERTED V ANTENNA
16 AUDIBLE TIMER
17 LONG INTERVAL TIMER
18 TWANGY DISTORTION UNIT

INTRODUCTION

Building projects is fun! Exactly where the most fun is can be debated. Is it in the actual assembly of the gadget? What about designing added features or molding it to particular needs? How about watching friends and family "pop eyes" as you exhibit its operation? Who knows? Never examine a pleasure, just enjoy it! With that admonition, inspect the plans of the seventeen projects presented in this Project Anthology while you heat up your soldering iron.

The experience level required of the project builder is minimal. Of course you require some level of project building experience but every effort has been made to keep the level of difficulty down. In some projects the illustrations go beyond the schematic diagram to include mechanical layout information. Although your electronic theory understanding may be scant, you will require higher levels should you indulge in designing additional functions and features into the projects presented herein. Here are general notes about the parts used in projects.

Resistors. Fixed-valued resistors are rated at either 1/4-watt or 1/2-watt, within 5 percent of their designated value. One-half watt is the preferred rating for resistors because of their physical size and ease of handling on circuit boards. Hobbyists with construction experience should select the wattage size that they deem suitable for the project. Resistors that are specified with wattage ratings may be substituted for other resistors of higher wattage rating. For example: you could replace a 1-watt resistor with a 3-watt resistor provided that the replacement is not of wire-wound construction. Wire-wound resistors have a measure of inductance in them and they may upset the operation of the circuit, especially high-frequency audio and radio circuits. The physical size of the resistor increases significantly with its power rating so common sense directs that you do not use outlandish substitutions. Only composition resistors are used in the following projects. Do not substitute different ohmic values unless the text discusses the possibility, or you are thoroughly qualified to make ohmic value changes by adequate academic knowledge and/or experience.

Capacitors. The voltage ratings of capacitors in the projects should be equal to or above that specified in the Parts List. Electrolytic capacitors should be reasonably close to the specified value in microFarads. The working-voltage-DC (WVDC) rating must never be less than specified, and as a rule never more than twice the specified rating. Two capacitors connected in parallel add their values. Two capacitors, say .10-µF and .25-µF in rating, when connected in parallel will be equal to a single capacitor rated at .35 µF. When paralleling capacitors to obtain useable values, be sure that the WVDC ratings of both capacitors are at least equal to the specified value.

Semiconductors. Solid-state parts such as transistors, integrated circuit chips, diodes, LEDs, etc., cause the most problems for new project builders. It is recommended that they obtain catalogs from retail and mail-order outlets. Radio Shack is an excellent retail source—there's one in every town in North America that has more than two traffic lights, or so it seems. There are a host of parts advertisers who offer free 800 telephone numbers so you can do your shopping at home. Some readers resort to substitution listings for what appears to be hard-to-get parts. This could be risky for beginners because pin connections vary with so-called identical units. If you are calling an 800 telephone number parts outlet, insist upon the exact part. There are enough 800 electronic parts suppliers in North America to call and get the parts you want.

Light-emitting diodes (LEDs) usually have their cathodes identified by the shorter lead which is closest to the flat edge of the circular rim at the base of the part. There are variances, and for bi-color LEDs, this is true for one of the colors only. If you have any doubts, do not hard solder the LED into the circuit at first. Run a power-on test and if the LED does not light, or the wrong color comes up, reverse the leads and try again. The LED should light or the correct color appear. Now you can solder the LED into the circuit. Solderless circuit boards can be used for testing mockups.

Not every construction or parts problem can be forecasted, however, experience is a great teacher and you will learn much by jumping into the construction of a simple project. Experience grows with practice. The fantasies of the Project Anthology are now all yours to enjoy. Have fun!

THE PROJECTS

1 Twinkle Tree

Twinkle Tree is an easy project for beginners to build, and its basic circuit has a number of useful applications. The circuit's visible action appears as a string of ten LEDs (light-emitting diodes) flashing on one at a time in sequence, this being repeated so long as the circuit is powered.

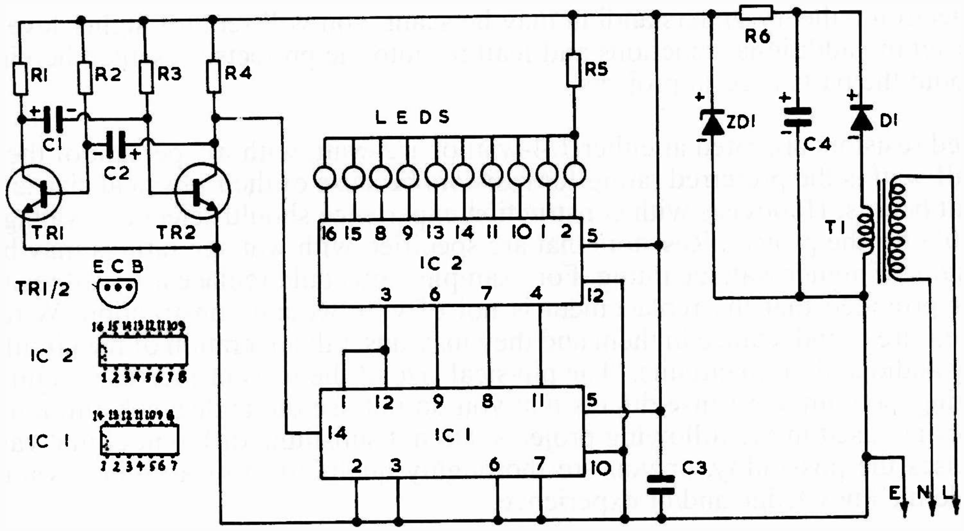

Fig. 1– Circuit diagram for the Twinkle Tree project.

The LEDs can be used on a small table decoration, in the form of a Christmas tree (we call it Twinkle Tree), or around a picture frame; or they may be placed at various points on a hanging decoration, such as a bunch of mistletoe, or can be incorporated in other decorations or models. The LED light display provides an interesting and novel twinkling effect.

The same circuit can be used with the LEDs placed round a small display, so that they flash on in turn with a revolving or racing effect.

How It Works. Figure 1 is the schematic diagram of the circuit for Twinkle Tree which includes a small power supply for operation from AC lines. If the circuit is to be powered from an existing supply, or batteries, omit circuit components ZD1, D1, R6, C4 and T1. The external power supply negative line goes to the "E" line in Fig. 1, and the positive line to the junction of resistors R1 through R5. The voltage supplied should be approximately 5-volts DC. You may want to connect an AC wall-plug power pack rated at 5-volts DC as the power source. Do not use 6-volt DC from dry cells or else the ICs will be damaged. 4.5-volt DC from three 1.5-volt dry cells may be used, however the LEDs will appear to be a bit dimmer.

The multivibrator circuit consists of transistors Tr1 and Tr2 which produces pulses that form

PARTS LIST FOR TWINKLE TREE

Semiconductors
D1—1N4001 rectifier
IC1—7490, decade counter (divide by 2 and 5)
IC2—7441, 1BCD-to-decimal decoder
LED1-LED10—Light-emitting diode, color optional
Tr1, Tr2—2N3706, PNP
ZD1—5.1-volt, 400-µA Zener diode
Resistors
R1—2,200-ohm
R2, R3—33,000-ohm
R4—1,000-ohm
R5—220-ohm
R6—33-ohm, 1/2-watt
Capacitors
C1, C2—2-µF, 6-WVDC, electrolytic
C3—47-nF or .05-µF, metallized polyester
C4—1,000-µF, 12-WVDC, electrolytic
Miscellaneous
T1—Filament transformer: 117-volt AC primary winding, 6.3-volt AC secondary winding (obtain smallest, lowest power unit available)
On/Off switch optional, 14-pin and 16-pin IC sockets, perf-board, wire, battery clip (optional), solder.

the input to the binary-coded, decimal-decade counter IC1, at pin 14. This IC counts the pulses and provides a binary output at 1-12, 9, 8 and 11 pins. Pin 5 is its positive line, and 10 is its negative line.

With counters having two or more numerals, as shown later, the IC can pass on a pulse to the "tens" section. Here, it is used alone and repeats the same series of outputs over and over so long as the multivibrator pulses are applied to pin 14.

Chip IC2 is a decoder-driver. Its purpose is to receive the various inputs at 3, 6, 7 and 4, and decode them into outputs at 16, 15, 8, 9 and so on, along to 1 and 2. Thus each of these ten points provides a circuit to the appropriate LED in turn. The outputs connected to LEDs are located at the top of IC2 when Fig. 1 is held right-reading.

All the LEDs are returned through the common limiting resistor R5, to operate from the same 5-volt DC line. The effect is, that each LED is illuminated in turn, along the line, and when the last is reached, this is followed by the first, and the sequence is repeated. Check the LED polarity, and mark this with red sleeving material, or other means, if the LEDs have no means of identification (long lead or flat side).

For the Twinkle Tree the ten LEDs are scattered at random, and any LED could be connected to any of the outputs 16 to 2.

Only one LED is on at a time, so current drain is only some 50 mA or so, and this small current load makes battery operation feasible and long lasting.

Transformer T1 (6.3-volt AC secondary) provides about 9-volt peak across C4, and R6 drops this for the 5.1-volt DC Zener diode ZD1. A meter placed across this diode should read about this voltage.

AC line current is drawn from a 3-pin powerplug line cord with a 2A built-in fuse. The green lead ground circuit in the power cord provides safe grounding of the secondary and low-voltage circuit. A double-insulated transformer specified as requiring no grounding may be operated without grounding through the AC power cord's green lead. Nevertheless, always play safe by using a ground circuit whenever using the AC line. Naturally construction and wiring must assure that no AC voltages can reach the transformed secondary output or the low-voltage circuit.

Construction. This is where the fun in electronics begins. Design your own Twinkle Tree. Maybe you would want to add a few fixed (always-on) LEDs or build two circuit boards and have twice as many twinkling lights. It's up to you!

2 One-Second Flasher

When making a long exposure using a camera with the shutter set to "B" or "T", some sort of timer is needed in order to obtain accurate results. Unless very long exposures are required, a simple one-second indicator such as the One-Second Flasher is probably the ideal solution. The unit pulses on a LED (light-emitting diode) indicator at one second intervals. The shutter is opened as the LED flashes, and then closed after the appropriate number of flashes have been counted off. There are almost certainly a great many other uses for a simple timer of this type, and it could be used as a simple enlarger timer for example.

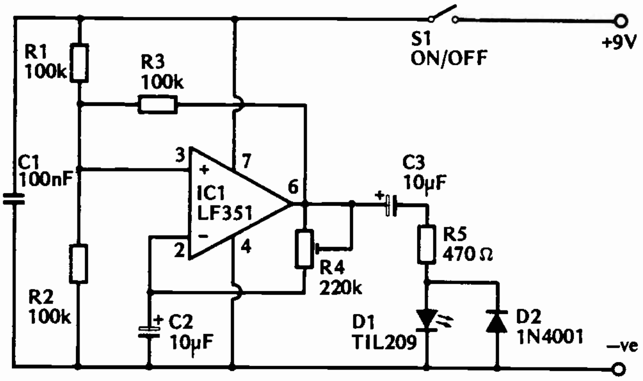

Fig. 2 –Circuit diagram for the One-Second Flasher.

How It Works. The circuit diagram for the One-Second Flasher appears in Fig. 2, and is based on an operational amplifier that is biased by R1, R2 and R3 to act as a form of Schmitt trig-

PARTS LIST FOR ONE-SECOND FLASHER

Semiconductors
D1—Light-emitting diode, red
D2—1N4001 rectifier
IC1—LF351 operational amplifier
Resistors
R1-R3—100,000-ohm
R4—220,000-ohm trimmer
R5—470-ohm
Capacitors
C1—100-nF, metallized polyester
C2-C3—10-μF, electrolytic
Additional Parts & Materials
S1—S.p.s.t. slide switch
Plastic cassette holder, 9-volt transistor-radio battery, wire, perfboard, solder, etc.

ger. The output goes to the low state if the inverting input is taken about 2/3V+, and high if it is taken below 1/3V+. The output therefore goes high initially, but C2 soon charges to 2/3V+ via R4, and then the output goes low. Capacitor C2 then discharges to 1/3V+ via R4, sending the output high again, and producing continuous oscillation.

Resistor R4 is adjusted to give an operating frequency of one Hertz. The One-Second Flasher can be calibrated against a watch or clock with a seconds hand by empirical means. The output of IC1 is coupled to the LED indicator, D1, by way of DC blocking capacitor C3 and current limiting resistor R5, and the LED is briefly pulsed on as the output voltage swings positive. Diode D2 ensures that there is both a charge and discharge path for C3 so that the output signal is properly coupled to D1.

The current consumption of the unit is about 2 mA.

As You Like It. Assembly and packaging of the unit is left to the builder. The circuit can be housed in a plastic cassette holder. The LED need not protrude because the case is clear. The circuit uses very little current so that a 9-volt transistor-radio battery can be used. The battery can be disconnected by unsnapping the clip connector or adding a single-pole, single-throw slide switch.

3 5-Volt Power Supply

The series of 7400 TTL integrated circuits are intended to operate from a 5-volt DC supply as are many other modern chips, and integrated circuit voltage regulators for this purpose are available.

The 5-Volt Power Supply shown here will provide voltage-regulated power for digital ICs

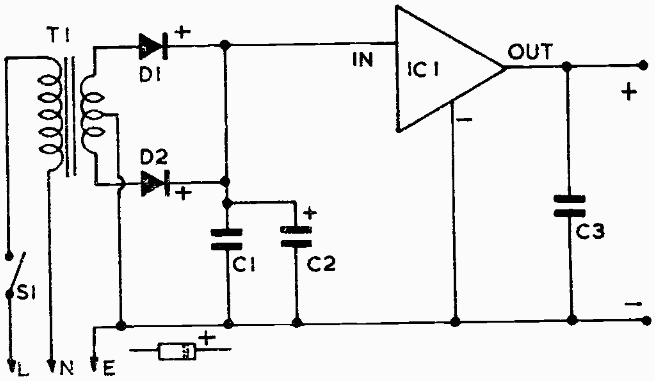

Fig. 3– Circiut diagram for the 5-Volt Power Supply.

needing up to 500 mA. By constructing the 5-Volt Power Supply as a separate unit, any circuit to be operated can be plugged in, and this allows it to provide power for various projects. It may also be used where a 5-volt DC, 500 mA supply is suitable, so it can in some cases be pressed into service where a 4.5-volt or 6-volt battery would otherwise be fitted.

The Circuit. Figure 3 shows the schematic diagram for the circuit. Transformer T1 provides isolation from the AC power line, and has a 9-0-9-volt, 500-mA secondary winding. After rectification, the DC input voltage for IC1 is something over 9-volts. A filament transformer rated at 12-0-12-volts AC may be substituted.

Diode rectifiers D1 and D2 are rated at 50 volts at 1 A: These diodes are readily available. Capacitor C1 helps suppress transients arising from the AC line, and C2 is the main smoothing

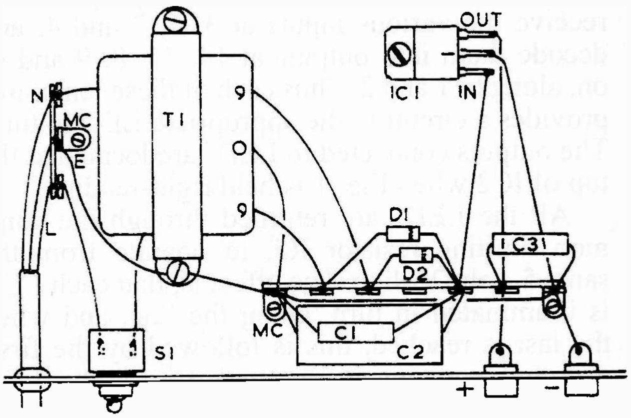

Fig. 4– Parts layout for the 5-Volt Power Supply. Observe markings on the diodes (D1 and D2) when soldering to the terminal strip

(filtering) capacitor. Here, a value of 1500-μF to 3500-μF may be used. Higher capacitances provide better voltage-ripple smoothing.

Chip IC1 is the LM341, a 5-volt, 500-mA positive regulator. Actual maximum current depends to some extent on other factors. Here, heat sinking is to a small metal chassis.

Figure 4 shows layout inside a small metal-

PARTS LIST FOR 5-VOLT POWER SUPPLY

Semiconductors
IC1—LM341 voltage regulator chip
D1, D2—1N4001 rectifier
Capacitors
C1—0.1—µF, 50-WVDC, electrolytic
C2—3500-µF, I6-WVDC, electrolytic
C3—0.1-µF, 50-WVDC, polyester
Additional Parts & Materials
S1—S.p.s.t., toggle switch
T1—120-volt/9-0-9-volt, 500-mA power transformer
Aluminum chassis box, 3-conductor power cord with molded plug and built-in fuse holder, 2 A fuse, 1 terminal strip with a minimum of 4 insulated ties and 2 grounded ties, 1 terminal strip with a minimum of 2 insulated ties and 1 grounded tie, wire, solder, 2 output connectors, red and black (binding posts preferred), etc.

box chassis. This can be a 4 X 6-in. chassis with enough height to clear the top of the transformer, T1. Follow the layout in the diagram when assembling the project to avoid errors.

Safety. The AC line power cord has three leads, green for ground or earth (E), white for neutral (N), and black for live (L) conductors. A thin power cord (#18 gauge) is adequate. Use a 3-pin fused plug, and place a low-rating fuse of 2A or 3A in this.

The cord passes through a rubber grommet, and its conductors are anchored on the 3-way terminal strip. Note that the switch S1 is placed in the live conductor (L) lead. The earth (E) lead (usually green in color) is connected to the chassis. This is also connected to the transformer core, and to the output negative line. If this lead is omitted and the negative return provided by the chassis is to be relied upon, be sure the attaching screws are very tight, and will not become loose.

The grounding is for safety. If a short circuit should arise in T1 or elsewhere, which could carry AC line voltages to the chassis and low-voltage output circuit, the low-rating fuse in the plug will blow.

The unit needs to be enclosed in an aluminum chassis box or case, so that no hot AC leads can be touched.

Safety, with this unit and similar AC line-operated power supplies, depends upon sound construction, correct wiring, fusing and grounding, and shielding of high-voltage points against accidental contact by the user. No high voltage must ever be able to arise on the low-voltage output sockets, or low-voltage wiring as equipment connected to the unit will no longer be safe to touch.

Testing. A DC voltmeter should show approximately 5-volts at the output terminals. Due to tolerances in IC1, and lack of exact accuracy of the meter, the reading might be very slightly above or below 5-volts.

To check for regulation, connect a load taking up to about 500 mA. A 6.3-volt, 0.3 A or 3-watt light bulb is suitable. The reading at the output sockets should remain virtually unchanged.

4 Sound-Switch

The Sound-Switch circuit of Fig. 5 is suitable for use in a sound-activated tape recorder and similar voice-activated applications. Provided a reasonably sensitive crystal microphone is used, the unit will operate from speech of normal volume at a range of several feet. To go beyond that range will invite the rustling of paper, coughs, heel clicks, crowd and traffic noises, air conditioner noise, and many other room and out-of-door noises that will obviate the need of Sound-Switch. A too sensitive sound activated switch would be on most of the time when placed in a noisy environment and a super sensitive microphone used to increase range would be worthless.

How It Works. The circuit uses IC1 as a low-noise preamplifier having a voltage gain of 270 times, and this is followed by a second stage of

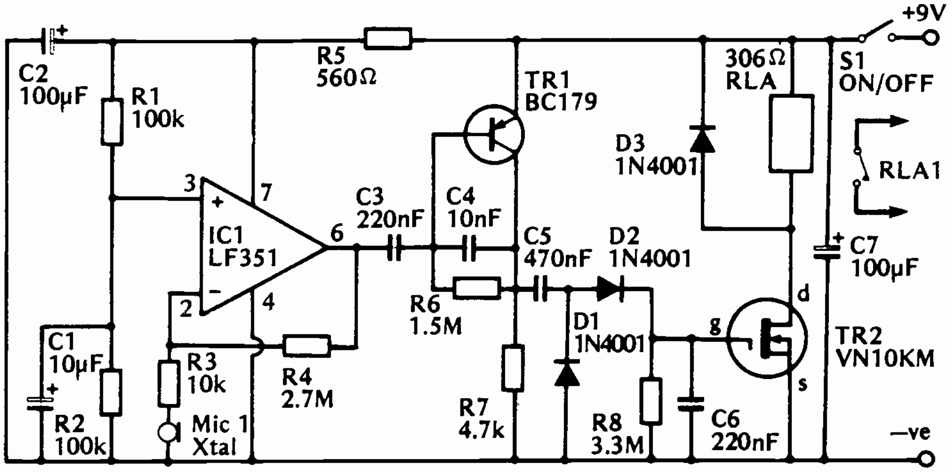

Fig. 5 –The circuit diagram for the Sound Switch project.

PARTS LIST FOR SOUND-SWITCH

Semiconductors
D1, D2—1N4001 rectifier
IC1—LF351 operational amplifier
Tr1—BC179 or 2N3965 , silicon PNP transistor
Tr2—VN10KM
Resistors
R1, R2—100,000-ohm
R3—10,000-ohm
R4—2.7-Megohm
R5—560-ohm
R6—1.5-Megohm
R7—4,700-ohm
R8—3.3-Megohm
Capacitors
C1—10-µF, electrolytic
C2, C7—100-µF, electrolytic
C3—220-nF, metallized polyester
C4—10-nF, metallized polyester
C5—470-nF, metallized polyester
C6—220-nF, metallized polyester
Additional Parts & Materials
Mic1—Crystal microphone
RLA—Relay with DC coil resistance of 306-ohms
Plastic or aluminum case, 9-volt battery (see text), wire, solder, etc.

amplification which gives a similar voltage gain. The second stage uses Tr1 in the common-emitter mode and this is a conventional arrangement except for the inclusion of C4. This capacitor provides a considerable amount of high-frequency attenuation, and this is necessary in order to prevent instability. Capacitor C5 couples the greatly amplified output of Tr1 to a rectifier and smoothing circuit which gives a positive DC output signal which is roughly proportional to the input signal level. If the input signal is sufficiently strong, the bias voltage at the gate of VMOS device Tr2 will be adequate to bias this transistor into conduction, and the relay which forms its drain load is then activated.

A pair of normally open relay contacts are used to control whatever item of equipment is operated by the unit. Of course, the voltage at the gate of Tr2 soon decays as C6 discharges through R8 if the input signal ceases, and the relay is then switched off. The decay time is roughly one second, which is about the optimum time for most applications. The attack time of the circuit is only a fraction of a second, and the unit responds almost immediately when a sound is initially picked up by the microphone.

Using It. You may want to record your voice giving instructions or reports, such as announcing the disassembly of a complicated component. As you speak, imitate the police and military who use such equipment. Begin each sentence with the word, "Kay." For example, "Kay. Removing the pawl from the gimbals." On playback, all your important words are not clipped and you have an accurate disassembly procedure that can be reversed for assembly purposes.

As the circuit has a fairly high level of gain the component layout must be designed with reasonable care, although the very limited bandwidth of the circuit greatly reduces the risk of instability. The current consumption of the unit is about 3 mA under quiescent conditions, but this increases considerably when the relay is activated. The circuit may be operated from a 9-volt transistor-radio battery, but you may want to use six C cells in series or a wall-plug power pack.

5 LIGHT ALARM

The Light Alarm unit produces a noticeable audio tone when the unit's photocell is subjected to a reasonably high level of light; normal daylight and artificial room lighting are sufficient to operate the circuit. The Light Alarm could be used as a simple burglar deterrent, and it would then be placed in a cabinet or a drawer, and would produce the alarm signal if the cabinet or drawer was opened by a burglar. The sudden noise would hopefully unnerve him or her sufficiently to make them leave the premises immediately. In the case of diet cheaters, the Light Alarm should be refrigerated. It could also be used in a medicine cabinet to sound a warning if a small child should somehow manage to open the cabinet. In this application the on/off switch should be fitted somewhere on the outside of the cabinet (preferably out of sight on top of the cabinet where a small child would be unable to reach it and probably be unaware of its existence) so that you could switch the alarm off before opening the cabinet, and switch on again after it has been closed.

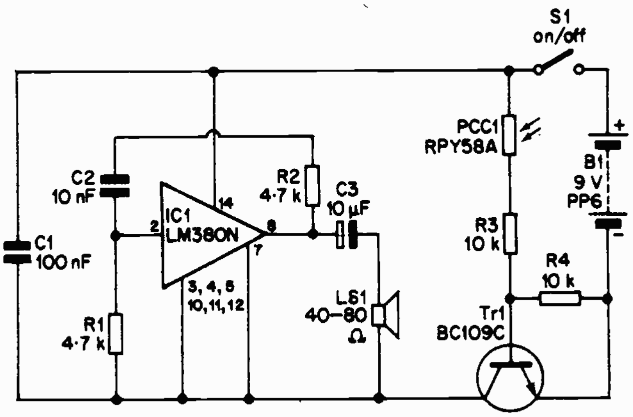

Fig. 6—The circuit diagram of the Light Alarm.

How It Works. The circuit is activated if Tr1 is switched on by a suitable base current and voltage. Refer to Fig. 6. The voltage and current available at the base of Tr1 is dependent on two main factors, the resistance provided by R4, and the setting of control VR1. If VR1 is set at maximum value, photocell PCC1 needs to have a resistance of about 10,000 ohms to bias Tr1 into conduction and activate the audio alarm circuit of which IC1 is a primary part. Fixed resistor R4 has been used across the base-emitter terminals of the switching transistor so that the sensitivity of the circuit is preset. However, R4 can be raised somewhat in value if increased sensitivity is required. Alternatively, it can be replaced with a 100,000-ohm linear potentiometer if a variable in sensitivity is required from time to time.

The audio-alarm generator uses an LM380N (IC1) in a simple audio-oscillator circuit, and drives high-impedance loudspeaker LS1 via coupling capacitor C3. It is a simple matter to get a non-inverting amplifier having a fairly high gain to oscillate, and it is merely necessary to provide AC coupling from the output to the input. Provided the losses through this coupling are less than the voltage gain provided by the amplifier this will give sufficient positive feedback to sustain oscillation.

The values for R1, R2, and C2 shown in the circuit diagram (Fig. 6) give considerably more feedback than is needed to just sustain oscillations and the circuit oscillates strongly producing a squarewave output at a frequency in the region of 1 kHz (1000 Hz).

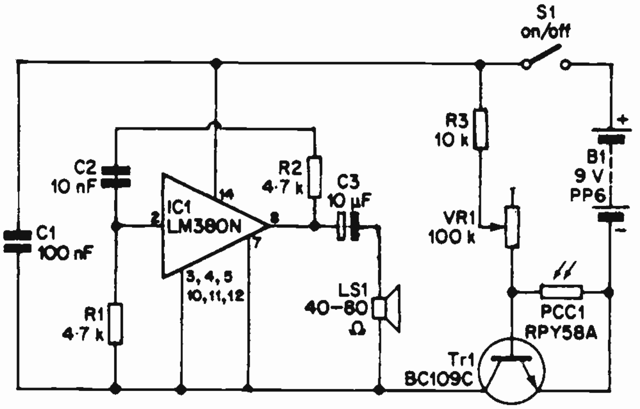

Fig.8–The circuit diagram of the Dark Alarm.

The breadboard component layout for the Light Alarm is shown in Fig. 7.

Incidentally, it is possible to alter the sound tone produced by the circuit; an increase in the value of R1, R2, or C2 (or more than one of these) produces a reduction in pitch, and vice versa.

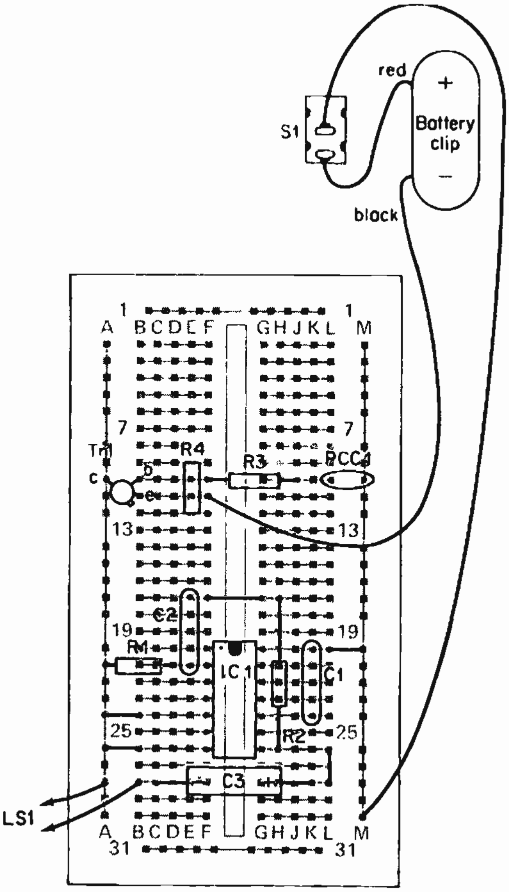

Fig. 7– Construction details of the Light alarm.

Dark Alarm. With a minor circuit change (compare Fig. 6 with Fig. 8) the Dark Alarm circuit is similar to the Light Alarm just described, but it operates an audible alarm if the light intensity received by the photocell falls below a certain threshold level, rather than if it exceeds the threshold level. A possible application for the unit is as a headlights-on reminder for the driver of a car. The circuit will operate properly from a

PARTS LIST FOR LIGHT ALARM

Semiconductors
IC1—LM380N, audio power amplifier
PCC1—Photo cell, type RPY58A or common silicon photo cells found in the hobbyist marketplace
Tr1—BC109C, ECG123A or 2N930, silicon NPN transistor

Resistors
R1, R2—4,700-ohm
R3, R4—10,000-ohm

Capacitors
C1—.1-µF, metallized polyester
C2—10-nF, metallized polyester
C3—10-µF, 25-WVDC, electrolytic

Additional Parts & Materials
B1—9-volt transistor-radio battery and connector to match
S1—S.p.s.t. miniature toggle
LS1—Loudspeaker, miniature, impedance in the range 40 to 80 ohms
Verobloc or similar solderless breadboard or mating/matching printed-circuit board, solder, wire, etc.

12-volt car battery instead of a 9-volt transistor-radio battery provided nose filtering is provided in the battery line.

Control VR1 is used to set the light threshold.

6

SIG-TRACER

The purpose of a Sig-Tracer is much the same as that of a signal injector. However, a signal tracer is, in a way, used in the opposite manner to a signal injector. Instead of injecting a signal at the output of the equipment under investigation and then working towards the input, a signal tracer is used to first verify that an input signal is present, and further tests are then made at strategic points working stage-by-stage towards the output of the equipment. When the signal can no longer be traced (or is obviously at too low a level, seriously distorted, or in some other way incorrect) the approximate area of the fault has been located. Normal voltage testing and similar checks are then used to precisely locate the fault.

A signal tracer for use on audio equipment simply consists of an audio amplifier driving a loudspeaker or an earphone, so that the traced signal, if present and correct, will give the appropriate audio output from the signal tracer. For RF or IF signal tracing the signal tracer must additionally have a suitable demodulator built-in so that an audio signal is produced from the tracer, provided a suitable input signal is present, of course.

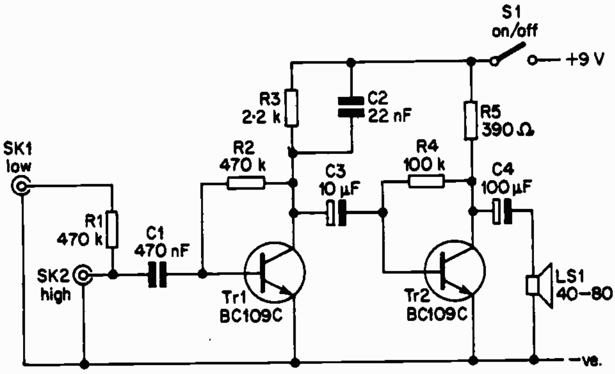

Fig. 9— The circuit diagram for Sig-Tracer.

The Sig-Tracer has good sensitivity so that it can be used for any normal type of audio signal tracing. The circuit provides AM demodulation, and the Sig-Tracer can therefore be used for IF and RF testing on AM radios provided a suitably strong antenna signal is available. The unit has an integral loudspeaker, and although the maximum output power is only a few tens of milliwatts, the volume level obtained is more than adequate for this application. If preferred though, the loudspeaker could be replaced by an earpiece or headphones. Low, medium and high impedance headphones are all suitable for use with the unit.

The Circuit. The circuit diagram for the Sig-Tracer (see Fig. 9) is basically just two common-emitter amplifiers with capacitive coupling. Transistor Tr1 is the input stage, and has C1 to provide DC blocking at the input, R2 as the bias resistor, and R3 as the collector-load resistor. An input signal from jack SK2 direct to the input of the amplifier (Tr1) may well be sufficient to overload the amplifier in many cases, and a second input having lower sensitivity is therefore provided (SK1), and the necessary attenuation (of nearly 40 dB) is provided by R1 and the input impedance of the amplifier. The input impedance of the amplifier at the high-sensitivity input (SK2) is several thousand ohms, but R1 provides a useful boost in input impedance so that the input impedance at SK1 is approaching 500,000 ohms.

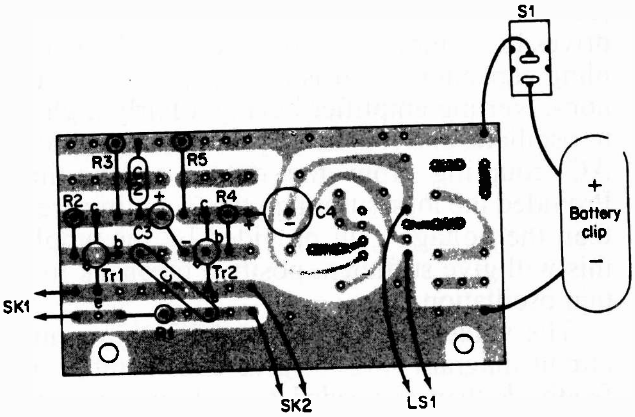

Fig. 10—The components layout for Sig-Tracer.

Capacitor C2 is used to roll-off the high-frequency response of the circuit in order to aid stability. It also gives the circuit a simple AM demodulator action by providing RF filtering. The rectification that is also needed in order to produce AM demodulation is provided by transistor Tr1 which, like any amplifier, does not give perfect linearity. It tends to amplify more during positive input half cycles when the collector current increases than it does on negative input half cycles when the collector current decreases. This form of non-linearity is typical of any normal bipolar transistor, and gives the required rectification. There is no need to have any AF/RF switching since the circuit will properly process either type of signal.

The output from the collector of Tr1 is coupled to the input of Tr2 by C3. Tr2 is used as a straightforward common emitter stage, but it is

PARTS LIST FOR SIG-TRACER

Semiconductors

Tr1, Tr2—BC109C, ECG123A or 2N930, silicon NPN transistor

Resistors

R1, R2—470,000-ohm
R3—2,200-ohm
R4—100,000-ohm
R5—390-ohm

Capacitors

C1—.470-µF, metallized polyester
C2—.022-µF, metallized polyester
C3—10-µF, 25-WVDC, electrolytic
C4—100-µF, 10-WVDC, electrolytic

Additional Parts & Materials

S1—Miniature s.p.s.t., toggle switch

Aluminum case, printed-circuit board, perf-board or solderless bread board, miniature loudspeaker having an impedance in the range 40 to 80 ohms (LS1), two 3.5-mm jacks (SK1 and SK2) 9-volt transistor-radio battery and battery clip connector, test leads, wire, solder, etc.

run at a fairly high collector current of around 15 mA so that a strong enough output signal to give reasonable volume from a high-impedance loudspeaker can be obtained. C4 couples the output signal to the loudspeaker.

Construction. Before you begin, remember that it is advisable to fit this project into a metal case grounded to the negative terminal of the power supply so that the circuit is screened from stray electrical signals. A suggested component layout on a printed-circuit board for the project is shown in Fig. 10. The project is simple enough to do without a printed-circuit board and may be assembled on a perf-board or some conventional pre-etched, general purpose, circuit board. As the input and output of the unit are in-phase, keep the speaker leads reasonably well separated from the input leads in order to minimize stray feedback and give good stability. It should not be necessary to use shielded leads to connect the board to jacks SK1 and SK2 in order to reduce feedback oscillations, but the external lead which connects to these sockets should be a shielded type.

The current consumption of the Sig-Tracer is about 17 mA, and, a standard 9V transistor-radio battery will not last very long on continuous service. This will not matter in many cases since the unit is likely to be used intermittently during troubleshooting procedures. However, if it is envisaged that the unit will be used quite frequently and for long periods it would be wiser to use a larger battery.

7 CW Filter

Although the IF filtering of shortwave receivers has tended to improve somewhat over recent years, few receivers have a very narrow bandwidth available, and this results in more adjacent-channel interference during CW (Morse code) reception than is really necessary. You have heard this interference: voice, music or heterodyne signals (even excessive static) riding over code. A simple way of obtaining improved results with a receiver that does not have a built-in CW filter is to use an add-on audio filter having a narrow bandwidth. If the receiver has a good filter for SSB reception, using an audio filter plus the SSB filter during CW reception should provide excellent results. We have just the unit for you: our CW Filter.

The CW filter is connected between the headphone or loudspeaker socket or terminal strip of your receiver, and either the headphones or an external loudspeaker. The output of the receiver should have an impedance of 8 ohms or more. As the circuit provides unity gain at pass frequencies and a low impedance output, there should be no problems with a mismatching when the filter is in use.

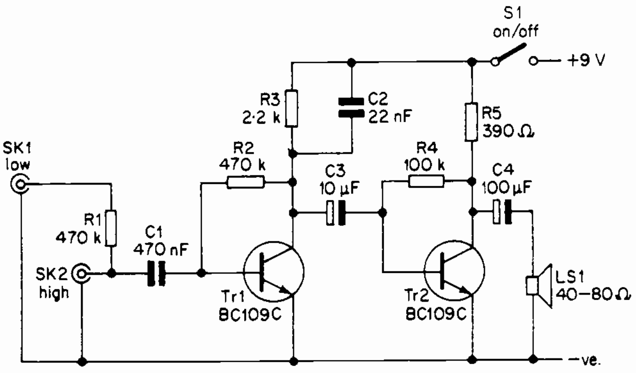

Fig. 11— Circuit diagram for the CW Filter.

The frequency response of the circuit peaks at approximately 800 Hertz, and the -6-dB bandwidth is about 300 Hertz or so. The -0 dB points occur at about 350 Hertz and 2 kHz. This is sufficient to normally give a substantial reduction in adjacent channel interference, but the response is not so narrow and peaky that using the receiver with the filter in circuit becomes difficult, with the wanted signal tending to drift out of the pass-band and become lost.

The Circuit. Resistors R1 and R2 (see Fig.

11) form a simple attenuator which is needed to counteract the gain of the circuitry that follows, and thus give unity voltage gain overall. The rest of the circuit is a conventional operational amplifier bandpass filter having the circuit values chosen for the desired center frequency and a high Q value (to give the required fairly sharp, peaky response).

An ordinary operational amplifier would not give good results in this circuit due to the limited current drive capability of typically just a few milliamps. A few high-current operational amplifiers are available, but even these can usually only provide a few tens of milliamps, and cannot provide a very high output power. This circuit is based on a TDA2006 device which is primarily intended for use as an audio power amplifier having an output power of up to about 10 watts, and the output overload protection circuitry within the device limits the output current to about 3 amps!

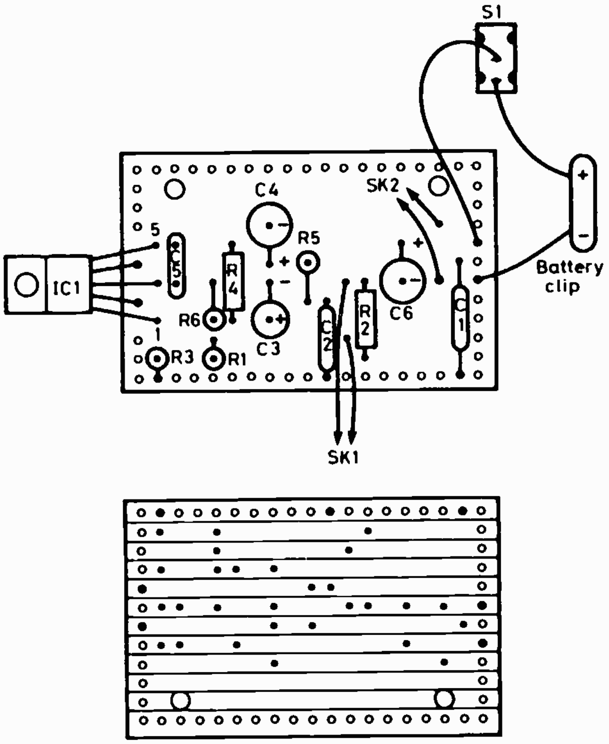

Fig. 12— Suggested parts layout on a stripboard for the CW Filter. Details are provided in text.

PARTS LIST FOR CW FILTER

Semiconductor
IC1—TDA2006, power amplifier
Resistors
R1, R3—2,700-ohm
R2—100-ohm
R4, R6—33,000-ohm
R5—180,000-ohm
Capacitors
C1, C2—10-nF, metallized polyester
C3—10-µF, 25-WVDC, electrolytic
C4-C6—100-µF, 10-WVDC, electrolytic
Additional Parts & Materials
S1—S.p.s.t. miniature, toggle-type switch
Suitable aluminum case, perfboard or 0.1-in matrix stripboard, 2 RCA phono jacks (SK1 and SK2), DC power supply (see text), wire, solder, etc.

Although intended for use as an audio power amplifier, the TDA2006 is a very adaptable device because it can really be regarded as an operational amplifier having a high-power, class B, output stage. The TDA2006 has inverting and non-inverting inputs with a high-input impedance (5 Megohms) at each of these, plus a high open-loop voltage gain over the audio frequency range. Unlike many audio ICs which are basically operational amplifiers, the TDA2006 does not have any internal biasing components, and it can therefore be used, as it is here, in standard operational amplifier circuit configurations.

The lowest recommended supply voltage for the TDA2006 is 12 volts, but it seems to operate well using a 9-volt battery as the power source. The current consumption is around 20 mA or so, and this can rise substantially if the circuit is used at high volume with an 8 ohm load. It is therefore advisable to use a fairly large battery pack or suitable 12-volt DC wall plug power pack. A supply voltage of 12 or 15 volts is perfectly acceptable provided C6 working DC voltage (WVDC) has a suitable high, maximum-voltage rating.

Construction. The component layout and wiring of the CW Filter are illustrated in Fig. 12. Construction of the component panel is very straightforward, and there are no breaks needed in the copper strips. IC1 has pre-formed leads and it is necessary to splay the readouts slightly in order to fit them onto a 0.1-in. matrix. IC1 should not need a heat sink if the circuit is battery powered, but this might be necessary if a 12 or 15 volt high-current battery supply or AC-line power supply is used.

If an operating frequency other than 800 Hertz is desired, the center frequency of the filter can be changed by altering the values of C1 and C2. The operating frequency is inversely proportional to the value of these capacitors, but note that reducing the center frequency also reduces the bandwidth of the filter, and increasing the center frequency gives an increase in bandwidth.

When first using the filter it may appear to give a large reduction in gain, and this is simply due to the fact that at most frequencies the filter does give a substantial amount of attenuation. The CW note from the receiver must be fairly accurately tuned to the center of the filter's response in order to obtain a strong output, and after a little experience with the unit, this tuning will be found to be quite easy.

8 Vol-Expander

Many types of live music, particularly classical music, have a very wide dynamic range. The dynamic range of music is simply the difference between the maximum and minimum sound levels, usually expressed in dBs. This dynamic range can be over 70 dB (although in most cases it is not quite this high).

To record this type of music requires a certain amount of compression being used in one form or another, especially in the case of a cassette recording where the available dynamic range of the recording system (tape and machine) is likely to be limited. Without some form of compression to restrict the dynamic range of the signal either loud passages of music overload the recording or transmission medium, or quiet passages fall into or below the noise level (or a combination of the two).

Music that has undergone even a small amount of carefully applied compression can be rather lacking in musical impact. You hear that kind of flat, background music in elevators and banks. An improvement in the playback can often be obtained by using a certain amount of expansion during playback or reception in order to boost the effective dynamic range of the signal. In order to give the desired effect an expander is required not to affect low-level signals, but produces a small boost in volume at middle dynamic levels, steadily rising to an increase of about 10 or 12 dB at the highest dynamic levels. In this way the lowest volume levels (and the background noise) remain unaltered, but the highest volume levels are substantially boosted. Of course, some noise is also boosted, but this is not noticeable as the wanted signal is strong enough to mask the noise.

Thus the use of a volume expander gives an apparent increase in signal to noise ratio, and can give more impact to music which requires a wide dynamic range. A volume expander cannot be expected to precisely counteract any compression used during the recording process or during transmission, but it can nevertheless produce a substantial subjective improvement when used with many program sources.

This all leads to the Vol-Expander that is our budget version of a practical volume expander hobbyists can build and use with rewarding music enjoyment.

The Circuit. An operational transconductance amplifier makes a good basis for our Vol-Expander since it is essential that the expansion is applied gradually, and is not virtually switched on and off. A voltage-controlled amplifier (VCA) using an operational transconductance amplifier tends to naturally give a suitable control characteristic. The circuit diagram of the Volume Expander is shown in Figure 13, and the operational transconductance amplifier used in the unit is a LM13600N device (IC1). This device has a number of additional features when compared to a basic device such as the CA3080E, the most obvious one being the fact that it is a dual device, and only one section of the device is used.

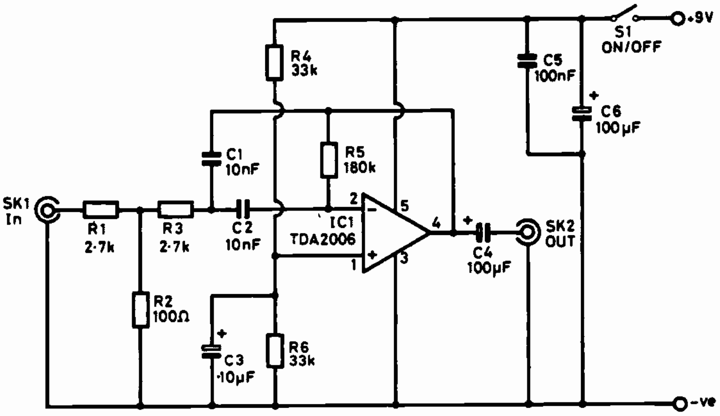

Fig. 13—Circuit diagram for Vol-Expander.

The other additional features of the LM13600N are the inclusion of a Darlington pair-emitter follower stage for use at the output of each section of the device, and linear diodes at the input of each section. The linear diodes can be fed with a small bias current, and this enables the device to handle a significantly higher signal level (about 10 dB in fact) before serious distortion occurs. This enables a higher overload margin and (or) increased signal-to-noise ratio to be achieved.

Resistor R8 is the discrete load resistor for the internal buffer stage of IC1, and R6 provides the bias current for the linear diodes. R7 sets the quiescent voltage gain of the circuit at a little under unity.

Some of the input signal is coupled by C4 to the input of a common-emitter amplifier which utilizes Trl. The amplified output from Trl is coupled by C5 to a rectifier and smoothing circuit which produces a positive bias voltage which is roughly proportional to the input signal level. This signal is applied to the amplifier bias input of IC1 via R12.

If only a low input-signal level is present there will be little or no current flow through R12 and into the amplifier bias input of IC1, and the voltage gain through the circuit will not be significantly changed. At higher signal levels the bias potential produced across C6 becomes large enough to produce a significant current flow through R12, and the gain of the circuit is boosted somewhat. At high signal levels of around 500 mV *rms*, the bias voltage becomes high enough to produce a boost in gain of around 12 dB. This is about the maximum expansion that can be used without the signal processing becoming obvious.

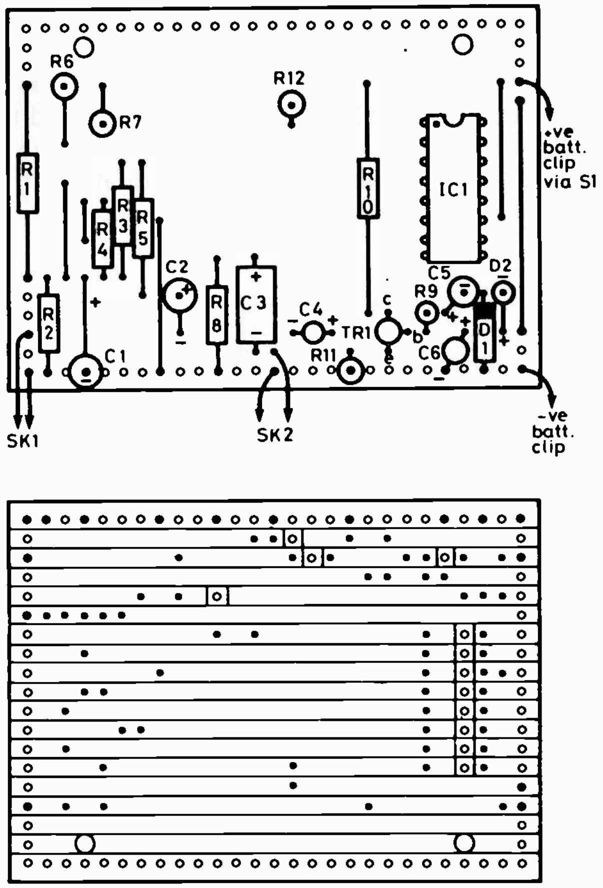

Fig. 14—Construction details of the Vol-Expander.

Capacitor C6 has been given a fairly low value so that the attack and decay times of the circuit are both fairly short, and the unit responds very rapidly to changes in the input level. However, the attack and decay times are made sufficiently long to prevent significant distortion from being produced.

The current consumption of the circuit is only about 6 mA, and a small 9-volt transistor-radio battery makes an economic power source.

PARTS LIST FOR VOL-EXPANDER

Semiconductors
IC1—LM13600N, transconductance linear opamp
Tr1—BC109C, ECG123A or 2N930, silicon NPN transistor
D1-D2—1N4148, silicon signal diode
Resistors (1/2- or 1/4-watt, 5%; 10% over 1 Megohm)
R1, R2—3,900-ohm
R3—220-ohm
R4—15,000-ohm
R5—10,000-ohm
R6—68,000-ohm
R7—39,000-ohm
R8—3,300-ohm
R9—1.8 Megohm
R10—4,700-ohm
R11—120-ohm
R12—10,000-ohm
Capacitors
C1—100-µF, 10-WVDC, electrolytic
C2—4.7-µF, 10-WVDC, electrolytic
C3—10-µF, 25-WVDC, electrolytic
C4—220-µF, 35-WVDC, tantalum
C5—1-µF, 10-WVDC, electrolytic
C6—330-µF, 35-WVDC, tantalum
Additional Parts & Materials
S1, S.p.s.t. miniature, toggle switch
Suitable aluminum cabinet or box, 0.1-in matrix stripboard or perf-board, two RCA phono jacks (SK1 and SK2), 9-volt transistor-radio and battery-clip connector, wire, solder, etc.

Construction. An 0.1-in. matrix stripboard panel for the Vol-Expander is detailed in Fig. 14, and a board having 19 copper strips by 27 holes is required. If you cannot obtain this board, a perf-board of the same size with flea clips inserted may be used. Interconnect the flea clips with No. 22 copper wire. Jacks SK1 and SK2 are the popular audio RCA phono types on the prototype.

If a stereo expander is required, it will be necessary to make up two boards, one to process each stereo channel. Alternatively, a longer piece of stripboard or perf-board could be used, and the otherwise unused section of IC1 could then be used as the basis of the additional expander circuit. The power supply center tap provided by R1, R2 and C1 could be used for the additional expander circuit, but all other components (apart from IC1 of course) would need to be duplicated in the extra expander circuit.

The Vol-Expander circuit should give good results if fed with the output of a tuner, cassette or tape deck, or a record deck via a suitable preamplifier. Some items of equipment might be found to have too little output to drive the circuit properly, and it will then be necessary to reduce the value of R11, or to even replace it with a shorting link. It is just possible that some signal sources will provide too strong a signal so that the expansion commences at a fairly low level, and in such cases R11 can be increased in value. If preferred, R11 could be replaced by a 1000-ohm linear potentiometer which would be adjusted for (subjectively)the best results.

9

COMPUTER VOICE

Need a mechanical-sounding voice, often called a "computer voice"? Well, here's a simple circuit that can be put to good use in amateur dramatic productions, telephone message tapes, etc. It's easy to produce synthetic- or computer-type voice effects from an ordinary voice-signal input. There are several ways to generate this effect, but probably the simplest method that gives good results is to amplitude modulate the ordinary voice signal by using a squarewave as the modulating signal. This is similar to the well-known tremolo effect which is sometimes used with guitars and other electronic instruments, but tremolo units normally use a triangular or sinewave modulation signal to give a smoother form of modulation. In this application a squarewave gives better results, and a sinewave or triangular waveform would be less effective. The unit switches the input signal between being passed straight through to the output, and being attenuated by around 20 dB, or so. A switching rate of a few Hertz is used.

The Circuit. The circuit diagram of the Computer Voice unit is given in Fig. 15. It consists of two sections: an oscillator to provide the modulation signal, and the modulator itself.

The oscillator uses a 555 timer chip in the astable multivibrator mode, and the frequency of operation has been set at about 10 Hertz by the values given to R1, R2 and Cl. The mark space ratio (duty cycle) is virtually one to one as the value of R2 has been made high in value relative to the value of R1.

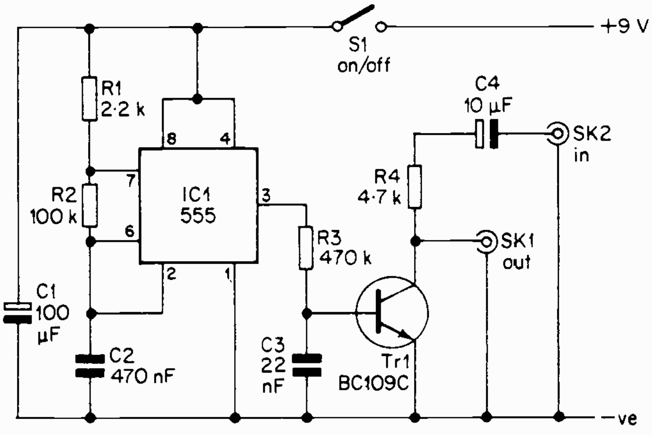

Fig. 15 —The circuit diagram for the Computer Voice unit.

PARTS LIST FOR COMPUTER VOICE

Semiconductors

IC1—555 timer chip

Tr1—BC109C, ECG123A or 2N930, silicon NPN transistor

Resistors

R1—2,200-ohm

R2—100,000-ohm

R3—470,000-ohm

R4—4,700-ohm

Capacitors

C1—100-µF, 10-WVDC electrolytic

C2—.47-µF, metallized polyester

C3—22-nF, metallized polyester

C4—10-µF, 25-WVDC electrolytic

Additional Parts & Materials

S1—S.p.d.t. miniature toggle

Aluminum case, solderless or perf-board circuit board, 9-volt transistor-radio battery and matching battery clip, two 3.5-mm jack sockets (SK1 and SK2) or reasonable substitute, wire, solder, etc.

A very simple modulator is used, but this is quite all right in this application where a small amount of distortion is quite acceptable. Distortion produces new frequencies that help to change the voice signal and make it sound less like the original. A large amount of distortion is obviously not desirable as it would severely impair the intelligibility of the output signal.

Transistor Tr1 is used as a sort of voltage controlled resistor: and, in conjunction with R4, it forms a voltage controlled attenuator. Capacitor C4 is merely a DC blocking capacitor at the input and R3 is used in the base circuit of Tr1 to effectively make Tr1 a voltage-controlled device rather than a current-controlled one. During the periods when IC1's output is high, Trl is biased into conduction by the base current it receives via R3, and effectively has a collector-to-emitter resistance of only a few hundred ohms. This gives losses of around 20 dB or so through R4, and attenuates the signal fed to the output by this amount. When IC1's output is low, Trl is switched off and the input signal can pass straight through R4 to the output. The only attenuation is due to the potential divider action between R4 and the input impedance of the equipment to which the output signal is fed. This should minimize losses.

Thus the output signal is switched up and down in amplitude, producing the required computer-type voice effect. Capacitor C3 is included in the circuit to slightly slow down the sharpness with which the signal is switched between the two levels of gain, and this helps to avoid "clicks" which can otherwise occur during transi-

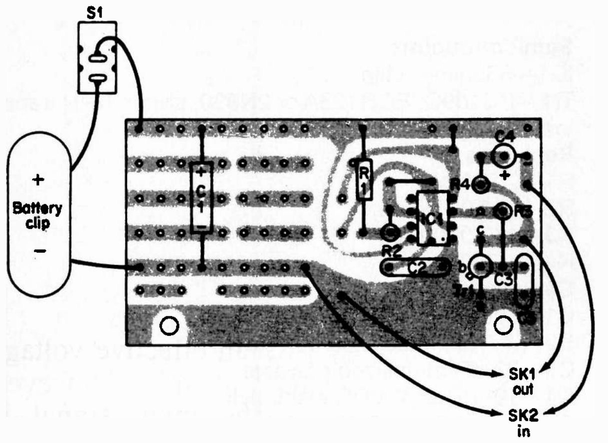

Fig. 16— The Computer Voice unit assembled on a multi-project board. The text advises using a solderless circuit board. Use that which is most comfortable for you.

tions from one gain level to the other. Capacitor Cl is a power-supply decoupling capacitor and this also helps to prevent switching "clicks" from being generated within the unit.

Construction. It is advisable to assemble the project on a solderless circuit board as illustrated in Fig. 16. Assembly is easy and experimental changes in component values will enable you to customize the output voice signals to your taste. Jack sockets (3.5-mm) were used for SK1 and SK2 on the prototype but these can be any two-way audio connectors. It is a good idea to house this project in an aluminum metal case (grounded to the negative supply rail of the circuit) so that the circuitry is screened from power-line hum, radio frequency signals, and other possible sources of electrical interference.

The unit can be used to process a signal directly from a microphone, but this may well cause problems with large losses through the unit even when Tr1 is in the off state. This loss is due to the fact that some microphones are designed to feed into quite a low impedance, and with such a load-impedance mismatch substantial losses occur. Some microphones are designed to feed into high impedance and will not produce this problem. Nevertheless, as microphones of this type have quite a high impedance, the losses provided by the unit when Trl is in the on state will be much higher than is really desirable. There is also the problem of the low-output level of a microphone that requires a high level of gain from the unit. This amplification will also boost the noise level to an uncomfortable listening level.

It is better to feed the microphone signal to the unit via a preamplifier, so that a fairly high-signal level is fed to the unit. A mere one stage of isolation with attending load matching is invaluable here.

There should be no problems if the unit is fed from a tape recorder or some piece of equipment which provides a similar signal level (a few hundred millivolts *rms*). It is not advisable to use the unit to process a signal having an amplitude of more than about 1.0 volt *rms* as quite severe distortion could then be produced.

10 Audio Power Booster

Despite the fact that there is a fair amount of noise inside most cars when they are traveling, many car radios and tape players have rather insufficient output powers to play over road noise. One reason for this is probably the fact that it is not possible to obtain a great deal of power using a 12-volt DC supply and a modern transformerless output stage into a 4- or 8-ohm loudspeaker. The maximum possible output swing is nominally 12 volts peak-to-peak. No output stage gives an output voltage swing equal to the supply voltage as there are inevitably voltage drops through the output transistors. On the other hand these are counteracted to some extent by the fact that a car battery usually has an actual voltage which is somewhat more than its nominal level. In terms of *rms* voltage this limits the maximum output to only about 4.25 volts, which gives only about 2.25-watts *rms* into an 8-ohm speaker, or 4.5-watts *rms* into a 4-ohm type.

There are ways of obtaining increased output, and the method used in the Audio Power Booster is to use a bridge-amplifier circuit. The Audio Power Booster is connected between the car radio or tape player and the loudspeaker, and it gives an output power boost of about 9-watts *rms* into an 8-ohm load or 18-watts *rms* into a 4-ohm load.

A Bridge Does It. A bridge-amplifier circuit consists of two power amplifiers with the loudspeaker being driven from the two outputs; neither speaker lead being connected to the car's chassis ground. The two amplifiers are arranged so that under quiescent conditions their outputs are at the usual level of about half the supply voltage, and there is no significant voltage present across the loudspeaker. The amplifiers are arranged to have output signals that are in antiphase, so that a positive change in the output potential of one is matched by a similar but negative change in the output voltage of the other. Thus when one amplifier has its output fully positive, the output of the other one is fully negative. This gives about 12 volts across the loudspeaker. On signal peaks of the opposite polarity the two output states are reversed so that 12 volts is again

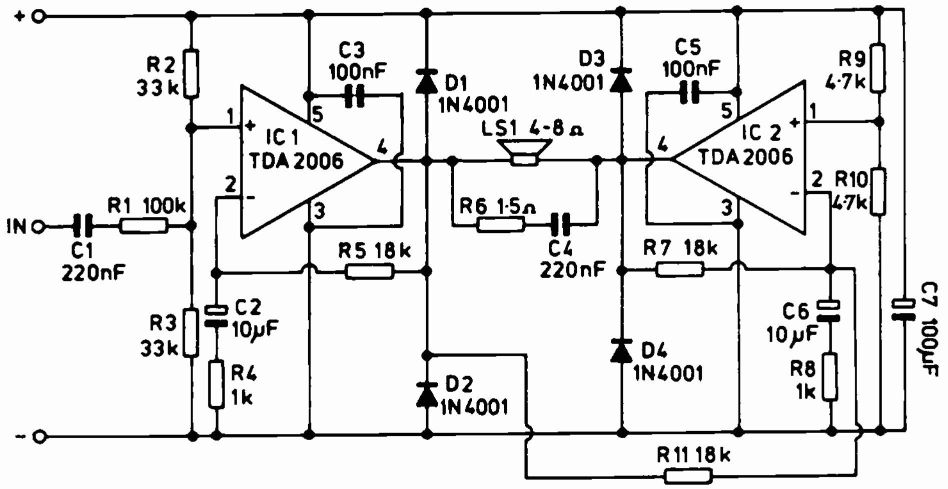

Fig. 17—The circuit diagram of the Audio Power Booster. Notice the almost mirror image of the circuit components surrounding the ICs.

developed across the loudspeaker, but the polarity of the output signal has changed. Thus the amplifier can give up to ±12 volts across the loudspeaker, or a peak-to-peak voltage swing of 24 volts. This is double that obtained with a normal amplifier, and gives a maximum output current that is also doubled; thus the output power for a given speaker impedance is boosted by a factor of four when compared to an ordinary audio power amplifier.

The Circuit. The Audio Power Booster is based on two TDA2006 audio IC power amplifiers, as can be seen from the circuit diagram of Fig. 17.

Power amplifier IC1 is used as what is virtually a non-inverting amplifier, with the non-inverting input of the device being biased to half the supply voltage by R2 and R3. R5 provides 100-percent negative feedback from the output to the inverting input of IC1 at DC so that the circuit has unity voltage gain and the output is biased to the required level of half the supply voltage. C2 and R4 remove some of the feedback at audio frequencies and this gives a voltage gain of about 18 times at these frequencies. This is far more than is actually needed, and a voltage gain of little more than unity is in fact sufficient. It is necessary to remove a substantial amount of feedback though as instability is otherwise likely to occur. R1 is used at the input of the amplifier to reduce the sensitivity to a more suitable level, and C1 simply provides DC blocking at the input.

IC2 is used in virtually the same configuration, but its non-inverting input is not fed with an audio signal, and only receives the DC bias signal from R9 and R10. Resistors R2, R3, R9 and R10 are all close tolerance components so as to ensure a reasonably small voltage difference across the outputs of the two amplifier stages under quiescent conditions; and, consequently, a small and insignificant quiescent current through the loudspeaker. Resistor R11 couples the output signal of IC1 to the inverting input of IC2, and the value of R11 is chosen to give IC2 an effective voltage gain of unity. However, as the input signal is coupled to IC2's inverting input there is a phase inversion through this section of the amplifier, giving the required anti-phase relationship at the two outputs.

Diodes D1 to D4 are protection diodes for the two ICs, while R6 plus C4 form a *Zobel* network which helps to prevent instability. Components C3, C5 and C7 are all supply decoupling capacitors.

Construction. A 0.1-in. matrix stripboard having 18 copper strips by 17 holes is used to accommodate most of the components, and the component layout of this board is given in Fig. 18. There are no breaks in any of the copper strips. The audio power ICs are modern types which physically resemble plastic power transistors, but they have five rather than three leads. These must be carefully bent apart slightly in order to fit these devices into this stripboard.

PARTS LIST FOR AUDIO POWER BOOSTER

Semiconductors
IC1, IC2—TDA2006 audio power amplifier chip
D1-D4, 1N4001 rectifier
Resistors
R1—100,000-ohm
R2, R3—33,000-ohm, 2%
R4, R7—100-ohm
R5, R11—18,000-ohm
R6—1.5-ohm
R8—1,000-ohm
R9, R10—4,700-ohm, 2%
Capacitors
C1,C4—.22-µF, Mylar
C2, C6—10-µF, 25-WVDC, electrolytic
C3,C5—.1-µF, metallized polyester
C7—100-µF, 25-WVDC, electrolytic
Additional Parts & Materials
Suitable aluminum chassis box, 0.1-in matrix stripboard or perf-board, transistor heat sinks (see text), s.p.s.t. switch (optional), sockets, wire, solder, etc.

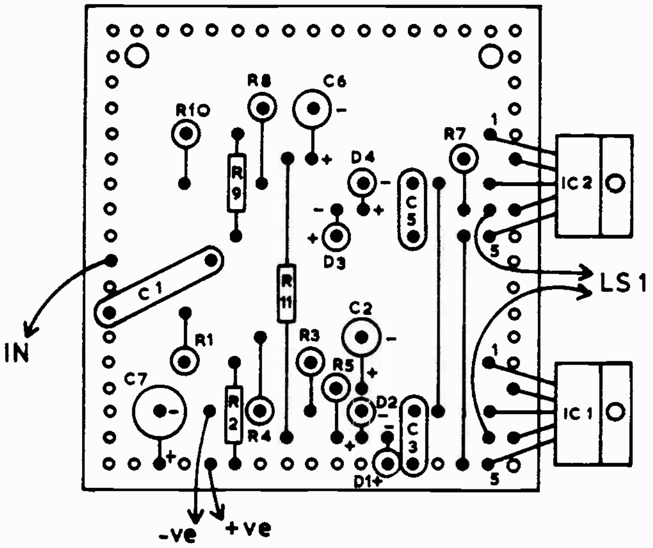

Fig. 18—Component layout for the Audio Power Booster as placed on a 0.1-in. matrix stripboard. The copper strips (on the reverse side of the board) run horizontally from left to right. If perfboard is used, interconnect all leads on the common horizontal row of holes for each strip.

Ensure that you do not omit the two jumper (link) wires. Be very careful to connect the polarized components with the correct polarity, especially the four rectifiers as these could be damaged or cause damage if connected incorrectly. Use the same care with electrolytic capacitors.

The two ICs should be bolted to a substantial heatsink. If the unit is used in a negative ground vehicle it can be fitted into a metal case which can be used as the heatsink, but it will probably be necessary to use an aluminum bracket to provide a thermal path from the ICs to the case. A suitable bracket can easily be produced from #18 *swg* aluminum that is commonly used in manufactured chassis boxes.

In positive ground cars this is not really advisable since the heat-tabs of the ICs connect internally to their negative supply readouts. This would result in the case connecting to the negative supply, and the likelihood of a short circuit between the case of the booster and the positive chassis of the car. Insulation sets supplied with the semiconductors can be used to insulate the heat-tabs from the case. Do a careful job since vibrations may shake the parts apart

Components R6 and C4 are not mounted on the component panel, but are wired across the output socket or terminals of the Audio Power Booster.

When installing the booster remember that neither output lead should be allowed to come into contact with the chassis of the vehicle. The TDA2006 incorporates output short-circuit protection circuitry and thermal overload shutdown circuitry, and accidental contacts between the output leads and the chassis (or each other) should not damage the unit.

Power. The power supply voltage for the unit should be taken from the supply line of the radio or tape player with which the booster is employed, as it will then be switched on and off with the main item of equipment. However, the unit can have its own on/off switch added into whichever supply lead is not grounded. The input for the booster is taken from whichever output lead of the radio or tape player is not grounded, and the correct lead can be found simply by connecting the two leads, in turn, to the input of the unit to see which one provides the desired result.

If the unit is used with a radio or tape player that has only a very limited output power it may be necessary to reduce R1 in value somewhat in order to fully drive the booster.

11 Pit-Stop—The Indy of Electronics

Pit-Stop is a motor-racing game that can be played with up to four competitors, the winner being the first driver past the checkered flag after a pre-arranged number of laps. The rules are simple and to the point. A driver can move away from the starting grid only when he has tanked up. Fueling and moves (distance round the track) are determined by an electronic indicator, the number of moves depending on skill in selecting the numbered LEDs that flash in sequence. Each move results in fuel being used up and the state of each fuel tank is monitored by a FUEL GAUGE. Lack of fuel is shown by the TANK EMPTY indicator light; drivers must then stop to TANK SELECT and refuel before any further movement. In addition to the DRIVE and REFUEL operations, there are hazards and advantages along the track dictated by game cards.

Pit stops can be enforced by game-card selection and, in a more realistic version of the game where refueling is confined to the pits, also made at a driver's discretion. If this rule applies, a TANK EMPTY light on when the car is out on the track puts the driver out of the race.

The Circuit. The circuit diagram for the Pit Stop game can be considered to consist of four basic blocks:

- counter;
- the decade counter and ten-LED flashing display that can be stopped momentarily to indicate either REFUEL or DRIVE distance outputs, selected by the players;
- four tank capacitors, chargeable when tank-up switches are pressed;

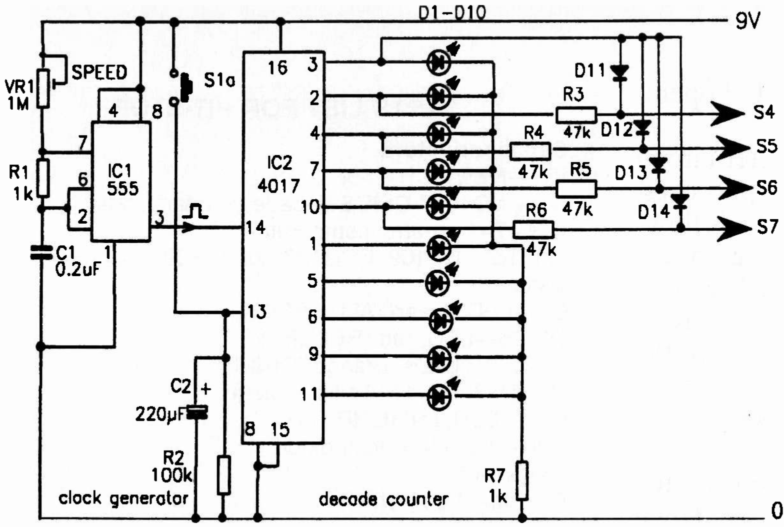
D1-D10
9V
VR1
1M
SPEED
S1a
4
8
7
6
2
3
1
IC1
555
R1
1k
C1
0.2uF
16
IC2
4017
14
13
8 15
3
2
4
7
10
1
5
6
9
11
D11
R3
47k
D12
R4
47k
R5
D13
R6
47k
D14
47k
S4
S5
S6
S7
C2
220µF
R2
100k
clock generator
decade counter
R7
1k
0V

Fig. 19—Circuit diagram for the Pit-Stop counter circuit.

- the comparator, controlling the TANK EMPTY light, and a meter that monitors fuel levels.

In Fig. 19 the clock generator, which provides the timing pulses for controlling the decade, the clock generator, IC1, is the popular CMOS 555 timer used in the astable multivibrator mode. The frequency of oscillation is controlled by the fixed capacitor Cl, and the variable resistor VR1 (SPEED). Output pulses are available at pin 3.

The 555 output of IC1 is directly coupled to the clock input of integrated circuit IC2, a 4017 decade divider/counter. With the reset output (pin 15) connected to the 0V line, the ten LEDs (D1-D10) connected in the divider outputs flash sequentially at a speed determined by the setting of VR1. Normally, VR1 is set so that the counter output LEDs flash at several times per second. During a turn by a player, when S1 is pressed, contact Sla freezes the display on one LED for a few seconds before flashing resumes. The rate of flashing should be fast enough so that some skill is required to anticipate the selection of a particular LED. However, for younger players, it may help to select a slower speed with VR1.

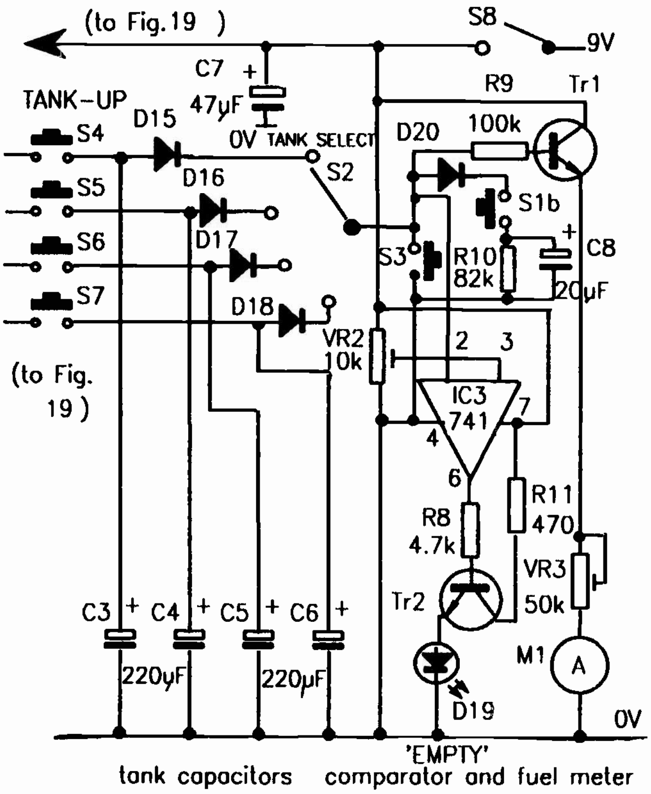
(to Fig.19)
S8
9V
C7
47µF
TANK-UP
S4
S5
S6
S7
D15
D16
D17
D18
0V
TANK SELECT
S2
R9
Tr1
D20
100k
S1b
S3
R10
82k
C8
20µF
VR2
10k
2
3
IC3
741
4
7
6
R11
470
R8
4.7k
VR3
50k
Tr2
M1
A
D19
(to Fig. 19)
C3
C4
C5
C6
220µF
220µF
0V
'EMPTY'
tank capacitors
comparator and fuel meter

Fig. 20—Circuit diagram for the Pit-Stop fuel-tank and comparator circuit.

As shown in Fig. 20, each player has a tank (electrolytic capacitors C3-C6) that can be charged via a TANK-UP switch (S4-S7). Four of the outputs of the counter are coupled via resistors (R3-R6 on Fig. 19) to the TANK-UP switches. During refueling, when a selected output is frozen for a few seconds, if the relevant TANK-UP switch is pressed, the positive voltage on the output pin is connected via the 47,000-ohm resistor (R9) and switched to charge the driver's tank capacitor. If by misjudgment another driver's LED is selected, then the other driver can press his or her TANK-UP button and take on some fuel.

Before pressing the GO switch S1, the TANK SELECT switch S2 must be set to the player's position. Diodes D15-D18 isolate the tank capacitors and prevent any short-circuiting of charges if the rotary contacts of S2 should make-before-break. Switch S2 connects the comparator (IC3) and fuel meter (M1) to monitor the state of the player's fuel tank capacitor. It also connects the tank capacitor to a discharge circuit, when S1 is pressed, formed by R10 and C8 connected in parallel. On each turn, a player's tank is discharged slightly by this circuit, i.e. fuel is used up on the drive, especially if short distances are selected. Switch S3, DRAIN TANK, serves two purposes: it can be used with a Hazard Card to simulate a loss of fuel situation, or (at all positions of the TANK SELECT switch) to drain tanks for the next game.

Diodes D11 to D14 enable tank capacitors to be charged via the TANK-UP pushbuttons if the ALL refuel light is on.

The comparator is a 741 operational amplifier, IC3, which monitors the voltage in a tank capacitor and compares it with a reference voltage set by potentiometer VR2. When the voltage level in the connected tank capacitor falls below the reference voltage, output pin 6 of the comparator goes high; this switches on transistor Tr2 and the

TANK EMPTY LED lights. A TANK-UP operation is then required.

Switch S2 also connects the fuel meter circuit to a driver's tank circuit for monitoring purposes. Transistor Trl is connected as an emitter-follower, the high-input impedance preventing excessive drain on the tank capacitor. The fuel reading on the micro-ammeter (M1) in the emitter circuit is set by the series preset resistor VR3. Leave the setting for VR3 in the maximum resistance position. This will prevent the meter from being damaged when the game is first turned on.

Construction. The prototype was constructed on a 15-in. square, 3-ply board, layout as shown in Fig. 21, with four 2-in, xxX 1/4-in strips for the sides. The clock generator and decade counter were mounted on a separate piece of 0.1-in stripboard or perf-board. A perf-board with flea clips may be used or a solderless breadboard and later converted to a matching pre-etched circuit board with components soldered in place. The wiring is not critical, but neatness is important should troubleshooting be required.

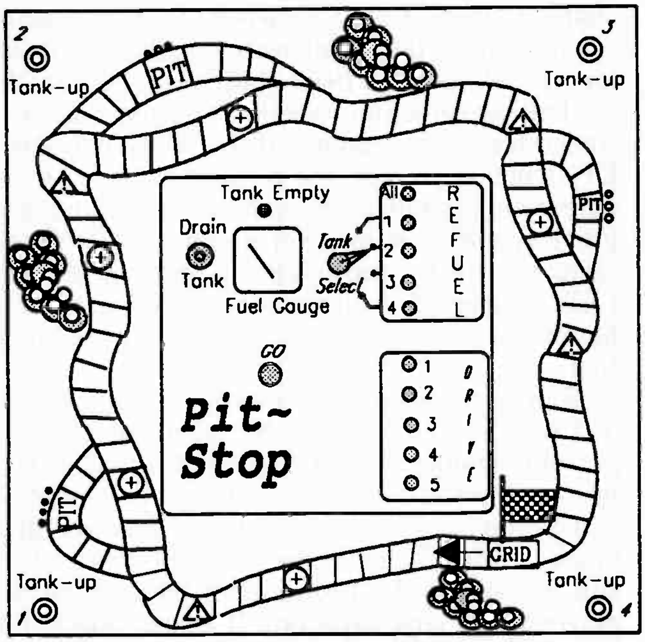

Fig. 21—Suggested Pit-Stop game board layout. The final layout is limited only by your imagination.

The track layout can be an enlarged replica of the one shown in Fig. 21, or can be modified to incorporate other ideas, especially if the constructor is a motor-racing enthusiast.

Playing. If you drive a car, a quick briefing on the play philosophy is all you will need to understand the game.

1. Before starting, set the battery switch S8 on and make sure that the four fuel tanks are discharged by pressing the DRAIN TANK switch S3 while rotating the TANK SELECT switch S2.

PARTS LIST FOR PIT-STOP

Semiconductors
IC1—NE555CP timer
IC2—CD4017 CMOS decade counter/divider
IC3—741 op-amp comparator
Tr1, Tr2—BC109, ECG123A or 2N929, silicon NPN transistor
D1—LED, green (ALL REFUEL)
D2-D5—LED, red (REFUEL)
D6-D10—LEDs, orange (DRIVE)
D11-D18—1N4148 silicon diode
D19—LED, red (EMPTY)
D20—1N4148 silicon diode
Resistors
R1, R7—1,000-ohm
R2—100,000-ohm
R3-R6—47,000-ohm
R8—4,700-ohm
R9—100,000-ohm
R10—82,000-ohm
R11—470-ohm
VR1—1,000,000-ohm potentiometer
VR2—10,000-ohm trimmer
VR3—50,000-ohm trimmer
Capacitors
C1—0.2-μF, metallized polyester
C2—C6—220-μF, 10-WVDC, electrolytic
C7—47-μF, 10-WVDC, electrolytic
C8—20-μF, 10-WVDC, electrolytic
Additional Parts Materials
M1—Microammeter, 50 or 100 microamperes
S1—2-pole, non-locking, push-button switch
S2—Single-pole, 4-way rotary switch
S3-S7—Non-locking, push-button switch
S8—S.p.s.t., toggle or slide (on/off) switch

2. Set the SPEED potentiometer VR1 to an agreed speed.
3. Select in turn to decide order of play by highest drive positions, and which two cars will be in pole (front) positions.

NOTE: Only two cars can be abreast at any time during a race, so a third car coming up fast behind two cars on the same patch of track will have to tuck in behind and forfeit any extra distance selected for the move.

4. The driver in pole position starts by switching the TANK SELECT switch to his position (RED, BLUE, YELLOW or GREEN). For instance, when the RED switch position is selected as shown, the red driver tries to anticipate when the LED for the red output is on and presses GO momentarily to select it. If successful, he then presses his TANK-UP switch to charge up his tank only while the LED remains on. Successful fueling is indicated if the TANK EMPTY lamp goes off, and a driver can select a drive distance when the next turn comes round, or elect to tank-up further.

5. The next driver in pole position, switches the TANK SELECT switch to his position and tanks-up in the same way.
6. The two drivers on the back row of the starting grid follow suit, but cannot drive off until at least one driver in pole position has moved (the no-passing two rule).

> **NOTE:** Holding down the TANK-UP switch after flashing resumes will result in some loss of fuel.

7. At the beginning of a turn, on TANK SELECT, if the TANK EMPTY light is on, a driver must tank-up. Any DRIVE moves selected inadvertently are forfeited.
8. If the ALL position is selected, all drivers can press their TANK-UP button to take on fuel.

Game Cards. Game cards are of two types—hazard and advantage. When a car lands on a hazard square (indicated by a triangle) the player draws a Hazard Card and obeys the instructions on the card. Likewise, an Advantage Card is selected whenever a car lands on an advantage square (indicated by a circle). Cut a number of cards to the size of business cards. Leave one side blank and on the other side type or write the following:

Hazard Cards
- Oil spill-lose a turn.
- Spin off track-lose a turn!
- Slow-down for chicane-reduce drive by 1.
- Tank holed-press DRAIN TANK.
- Aqua-planing-compulsory pit stop to fit wet tires. Lose a turn.
- Engine overheating-compulsory pit stop. Lose one turn.
- Flat tire-reduce drives by one and one compulsory pit stop.

Advantage Cards
- Super-fuel!-add 2 to drive score.
- Engine finely tuned-up-add 1 to drive score.
- Following wind-double drive score.
- Reserve tank-perform TANK-UP.
- Double accelerate-hold card until needed.
- Close gap-tuck in behind car immediately in front.
- Good road-holding-hold card and add 1 to drive score on next bend.

If you wish, let your imagination add additional hazards and advantages to the game cards you make.

The playing pieces for the board game may be pawns from a discarded chess set or some other easily gripped object. Coins are too difficult to manipulate. Paint the playing pieces appropriate colors to reduce confusion. Now then, drivers, man your cars!

12 Minefields

You are the captain of two ships that must navigate through magnetic minefields, racing against other ships in order to enter your safe port first. The game can be played by up to four players, each with two ships. In turn, under electronic control, the ships are steered past minefields in the direction of the home port. Refer to Fig. 22 for the board layout. The eight minefields normally flash sequentially (free-run) to warn ships of danger, until a mine is detonated (glows steadily) by a player.

Depending on the players' temperament or the state of the race, some players may take the longer safe channels (unless under other orders), while other captains may risk a winning dash for the home harbor via the minefields.

Movement. Players move in turn by pressing their DETONATE pushbuttons, which freeze the flashing minefield display for a few seconds. This results in one of two situations, as follows:

1. On detonate, one of the eight mines will glow for a few seconds together with one of the navigation lights. The player will observe which of the navigation lights is on, and also check if there is a ship lying on one of the eight shaded squares that surround the glowing mine. After the player has made his move, any victim in the minefield must either return his ship to base in the case of a direct hit (dark square) or take a NEAR MISS game card (gray square) and act upon instructions.

With regard to the direction of play, Fig. 23 shows the four moves. These are made in relation to a player's starting direction shown next to the DETONATE button:

- A green (starboard) light indicates a forward diagonal turn to the right.
- A red (port) light indicates a forward diagonal turn to the left.
- AHEAD indicates a forward move directly towards the harbor.
- ASTERN indicates a backwards move directly towards the base.
- A ship is only moved one square at a time, unless a game card instructs otherwise.
- Ships cannot be moved sideways.

2. If a mine does not remain on after a DETONATE, then one of the two center LEDs ACTION STATIONS or NEAR MISS will glow.

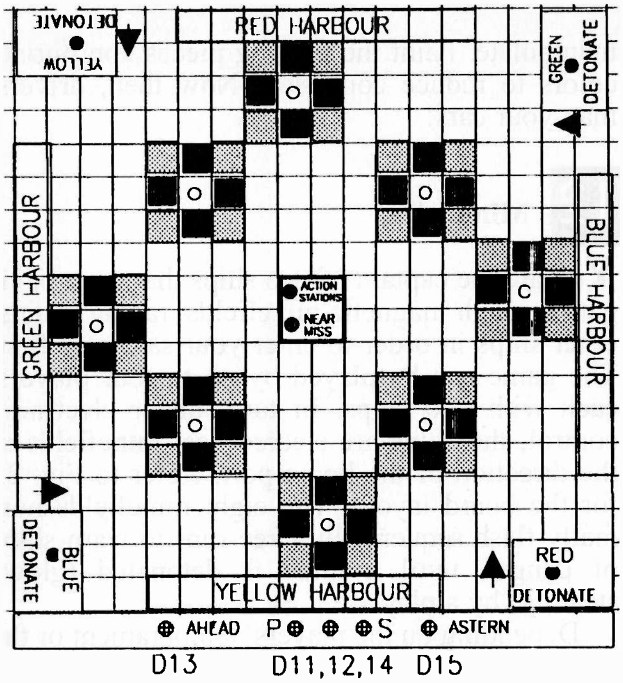
YELLOW
DETONATE
RED HARBOUR
GREEN
DETONATE
GREEN HARBOUR
BLUE HARBOUR
ACTION STATIONS
NEAR MISS
DETONATE
BLUE
YELLOW HARBOUR
RED
DETONATE
AHEAD
P
S
ASTERN
D13
D11,12,14
D15

Fig. 22— Minefields game board layout.

PARTS LIST FOR MINEFIELDS

Semiconductors
IC1—NE555CP timer chip
IC2—CD4017 CMOS decade counter/divider
D1-D8, D10—Red light-emitting diode
D9, D15—Orange light-emitting diode
D11, D12, D14—Tricolor (green/red) light-emitting diode
D13—Yellow light-emitting diode
Resistors
R1—2,200-ohm
R2—5,600-ohm
R3—100,000-ohm
R4—1,000-ohm
VR1—500,000-ohm potentiometer
Capacitors
C1—0.5-µF, metallized polyester
C2—220-µF, 10-WVDC, electrolytic
C3—4.7-µF, 10-WVDC, electrolytic
Switches
S1—S.p.s.t. slide or toggle (on/off)
S2-S5—Push-button, non-locking (DETONATE)
Additional Parts & Materials
Shallow box (can be made from plywood), 9V transistor-radio battery, plastic shapes for ships, stripboard or perfboard, wire, etc.

If it is a NEAR MISS light, then the player will take a NEAR MISS game card and act on it. However, in the event of ACTION STATIONS, all players are at liberty to press their DETONATE button as soon as the display returns to free-run, primarily to ward off any danger or to create havoc. The player will then take a game card from the ACTION STATIONS pile and act on the instructions. Game cards can sometimes be held for future use if not valid. In this case, they will take the place of a turn. After use they will be returned to the bottom of their respective piles.

The SPEED control (VR1) can be panel-mounted if desired and marked for half-speed and full-speed settings to cater for younger children, and for special instructions, e.g. on ACTION STATION game cards.

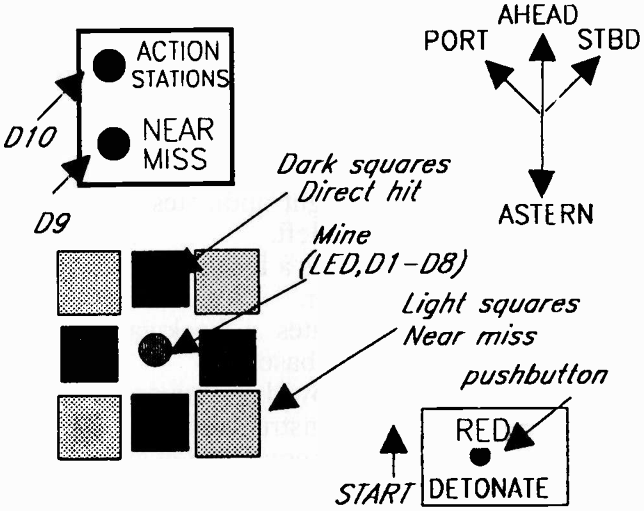
ACTION STATIONS
NEAR MISS
D10
D9
Dark squares
Direct hit
Mine
(LED,D1-D8)
Light squares
Near miss
pushbutton
RED
DETONATE
START
AHEAD
PORT
STBD
ASTERN

Fig. 23— Details of one minefield and related DETONATE switch and LEDs.

Depending on the agreed speed and a certain amount of anticipation, the display can easily be frozen on a desired mine, or navigation direction to shorten a game.

If a move cannot be made for any reason, e.g. obstacles such as coast lines, other ships, mines, then the move is forfeited. A move can be made by either of a player's two ships when thought advantageous, e.g. to hold direction or to avoid a minefield.

Hazards. One of the eight mines can be detonated (permanently lit): by a player's DETONATE switch as a normal turn, by detonating after the ACTION STATION light comes on, or if an ACTION STATION game card says DETONATE.

Any ships in a minefield when it is detonated (one mine permanently on) either take instructions from the top NEAR MISS game card if in the light-shaded square or if a DIRECT HIT is registered (on a dark square) go back to base for repairs.

The HALF SPEED and DETONATE game card introduces an element of skill. It allows more time for a player to anticipate which mine shall be detonated. The player operates the HALF SPEED button to slow down the display before detonating. This gives a chance to inflict a direct hit or a near miss on any rival ships already in a minefield.

The SPEED control can be reduced for the first few games, or to assist younger players; alternatively, younger players could use the

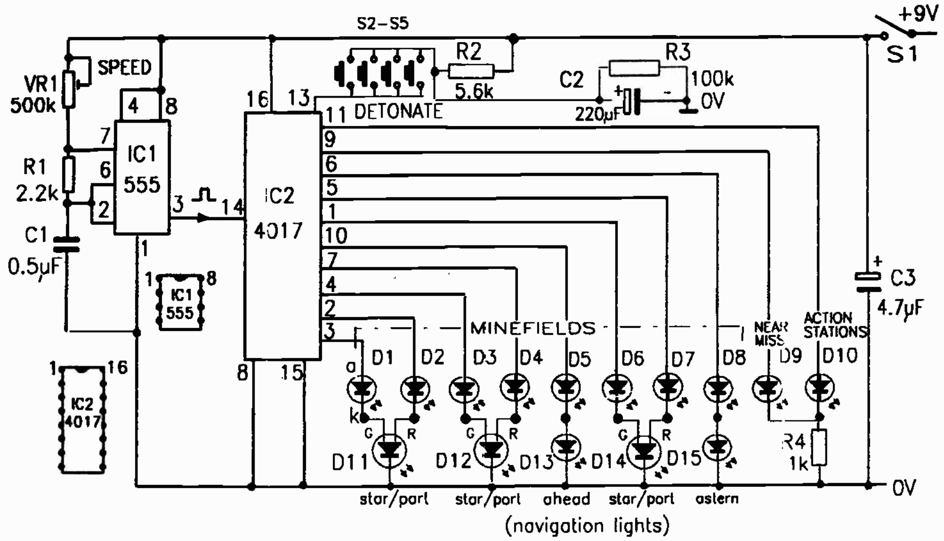

Fig. 24— Circuit diagram of the Minefields circuit.

HALF SPEED control for each move.

Ships may not enter an opponent's HARBOR.

Circuit. The 555 timer can be varied by VR1 to give a suitable speed for the free-run display for the mine and navigational LEDs. Refer to Fig. 24. The inclusion of three tricolor LEDs in series with the minefield LEDs gives green (starboard) and red (port) navigation signals. Also, the electrolytic capacitor C2 together with resistor R3 provides a time constant to hold the detonate and navigation information for a few seconds for scrutiny after a DETONATE push-button is released. This time can be lengthened or shortened as desired by altering the value of capacitor C2 (reduce capacitor value to shorten time and vice versa).

Layout is not critical, but requires a minimum of woodworking skill. The circuit elements can be soldered rats-nest style or assembled on pre-etched circuit boards and solderless boards. Neatness will make troubleshooting easier.

13 Building the Kapellmeisters

Most transmission-line loudspeaker enclosures involve a lot of tricky woodwork so that many speaker builders with only modest woodworking skills are deterred from building one. With the Kapellmeisters, although quite a lot of woodworking is needed, none of it is difficult, mostly it consists of cutting straight edges. These edges must be straight though, so if your saw cuts tend to wander, get the lumber yard to cut them for you. The measurements are uniform with many pieces being identical.

Preparation. Ceramic tiles used in the construction of the Kapellmeisters are standard 8- × 4-inch, cut down to 7 × 4 inches. That is just one straight cut per tile. The exception is the top tile over the speaker which because of its different angle must be narrower, 7 × 3.5 inches.

Triangular wood blocks are used to support the tiles. These are made by first cutting four 3-inch squares, then sawing diagonally to give eight triangles. The two wood blocks for the over-speaker tile are made by halving a 30- × 2-inch rectangle along the diagonal.

A standard 3 ×1 mix of sand and Portland cement (we will call it concrete) is used to fill the space behind the tiles. The concrete is applied between the blocks and the tile bedded on to it.

In all cases screw two or three stout wood flathead screws at random angles into the wood where the cement is to be laid leaving about an inch out of the wood, so that the protruding head of the screw will be buried in the concrete. These screws will then secure the concrete block in place when it is dry. Thoroughly wet the back of the tile before applying it to the concrete. In some cases the front of the tile may need to be held in place while the concrete sets, with wood panel nails knocked into the wooden sides. It does no harm to leave them there afterward.

All jointing is done by a strong wood glue. Elmer's wood glue was used for the prototypes which is very strong and convenient to apply; however there are many suitable wood glue alternatives on the market. The glue also fills any small gaps where the saw might have made a slight rough cut. Construction must proceed in numerous stages to allow the glue and concrete to set before continuing with the next stage, so some patience must be exercised. Make both speakers at the same time so that each stage can be completed on both and thus save time.

First Stage. Refer to the diagram of wood panels (Fig. 25), and buy sufficient plywood for all pieces. Remember that these pieces are for one speaker, so each unit will have to be duplicated. As so many are just 3-in. strips, it is likely that much can be obtained as scrap cuts. Most large lumber yards sell these at a reduced rate, and the color or grade doesn't matter as all are concealed except the top and bottom cheeks. The thickness should be as specified, 3/8 inch for the sides, baffles and back, 1/2 inch for the front, and

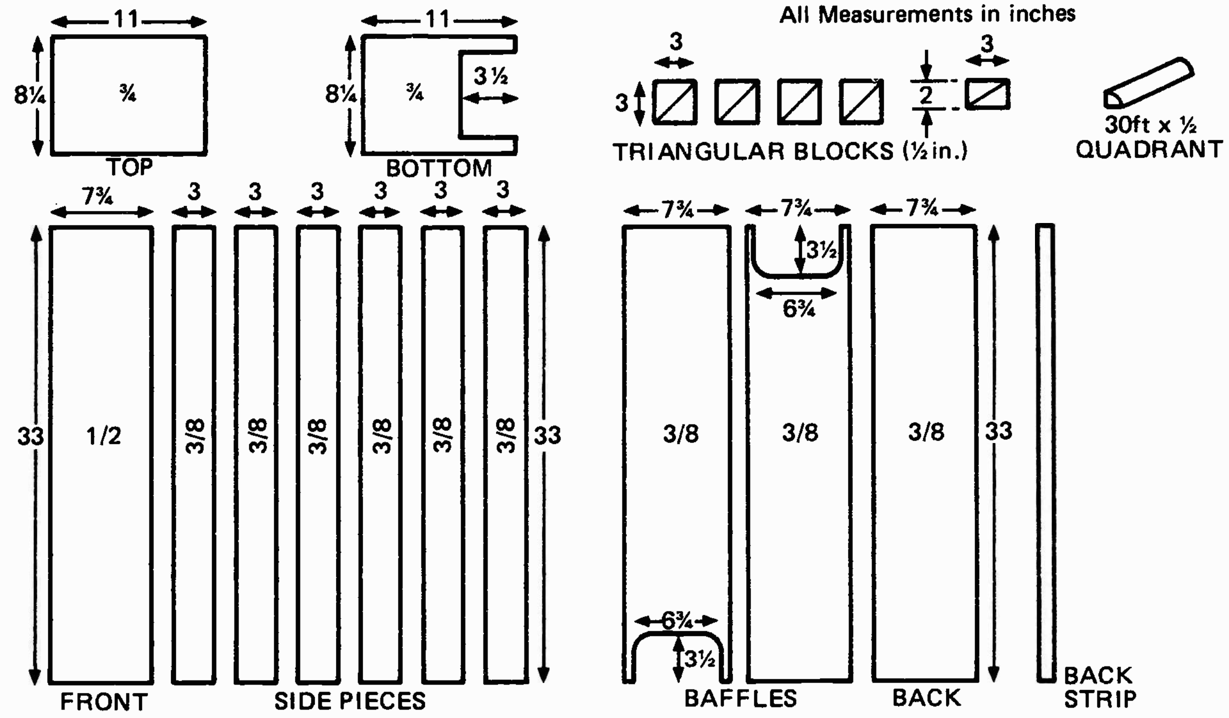

Fig 25 —View of all the wooden parts for one speaker cabinet including dimensions. Each side of the cabinet is made of three parts requiring a total of six side pieces for one speaker. All pieces should be plywood. See text for addition details on materials.

3/4 inch for the cheeks. It really will save a lot of time and energy if you get the lumberyard to cut the pieces with his machine saw. They usually charge slightly for each cut.

The lumber yard probably will not do the shaped ports at one end of the baffles and bottom cheek, but this can be managed with a fret or copping saw. A 7-1/2 × 4-1/4-in. elliptical speaker hole should be cut in the front panel starting 21 inches from the top. Before you cut the hole, check the physical size of the loudspeaker. It may be slightly different. It is desirable though not essential for the hole to be beveled outward.

Loudspeaker. A word about the loudspeaker that you should select. An oval speaker was selected for the original unit and it was rated at 8 watts with pretty good specifications at a low price. You should do the same; however, that would take away the fun of redesigning all project builders get involved with. Keep in mind that as the price goes up, the increase in performance lags behind so that dollars are wasted. Keep in mind that the Kapellmeister is designed to bring the best out of inexpensive loudspeakers.

Second Stage. Start with the front panel; lay it face down supported on some scrap quarter-inch ply or hardboard. Glue the top and bottom edges and fit the top and bottom boards. The front edges of the top and bottom cheeks should *not* rest on the ply supports but directly on the work surface; they will thus protrude a quarter of an inch beyond the panel. Refer to Fig. 26(a). The top and bottom cheeks should also be positioned to give an equal overlap at either side. The idea is for the top and bottom cheeks to overhang the front, sides and back 1/4-in. from all sides.

Weights should be applied to the free sides of the cheeks to hold them against the panel while drying. Measure the distance between the rear edges of the top and bottom cheeks to ensure that it is exactly 33 inches and therefore the top and bottom are parallel. Wait for glue to set and harden.

Third Stage. Fit the first pair of side pieces, gluing the ends and the edge contacting the back of the front panel; ensure the pieces are flush with the edge of the front panel. Measure across the upper edges to make sure they are 7-3/4 inches and so are true. Now glue the triangular blocks in place at the bottom and top as shown in (a) of Fig. 26, the top ones being the special sized ones. Glue the edges as well as the face that contacts the sides, but be careful in pressing them into place that you do not move the sides. Wait for the glue to set.

Fourth Stage. Now fit the speaker, screwing it in place over the aperture, and connect by soldering a pair of wires which are run down the panel to a hole drilled in the bottom. Leave a few inches of free wire outside the cabinet, and make sure both speakers are connected the same way to the color-coded wire. After the wires are installed, fill the drilled hole for the wires with a mixture of saw dust and wood glue to seal the hole airtight and secure the wire in place.

It is prudent to cover the front of the speaker with a piece of stiff cardboard secured by masking tape or staples to protect the cone from accidental damage.

Fit several screws to the bottom and top cheeks between the triangular blocks, leaving about an inch protruding at different angles. These will be embedded in the concrete when it is applied and so will hold the resulting concrete wedge in place. Next fill the space between the bottom blocks with cement (not too wet) and bed the trimmed tile onto it. Place the top narrow tile on the blocks and fill in behind it with cement.

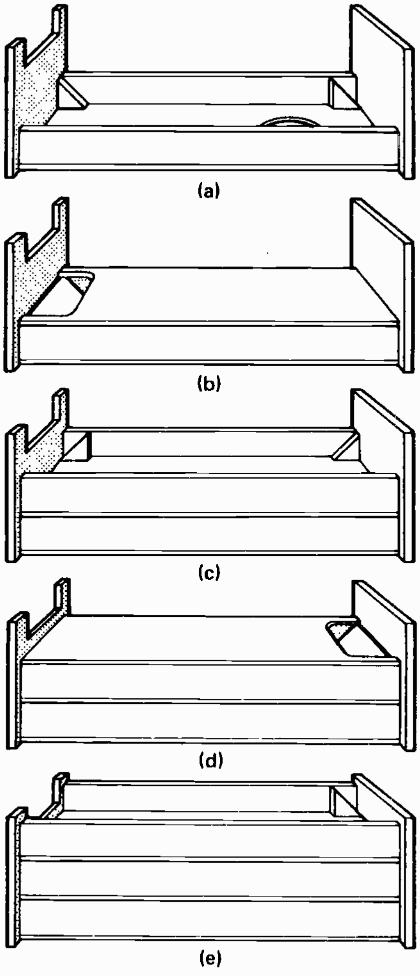

Fig. 26 — The vies a through e show the assembly procedure for the Kapellmeister. (a) First stage in construction. The front is glued to the top and bottom cheeks., using 1/4-in. supports to lift front panel to correct position. First pair of side pieces fitted and triangular blocks (small ones at the top) in position. Then apply concrete, fit tiles and lay wadding. (b) Fix in place first baffle, gluing down to insure close fit. (c) Fit second pair of side pieces and four triangular blocks (all blocks installed hereafter are regular size). Install tiles, concrete and wadding. Tuck extra lengths of wading through aperture into first channel. (d) Fit second baffle in place. Aperture goes at the top. (e) Glue in third pair of sides and last pair of triangular blocks. Fit tile and concrete and install wadding with extra length going into the aperture. Install the back panel (not shown). Refer to text for complete details.

Fifth Stage. Saw suitable lengths of 1-2-inch round molding (quadrant) and glue into the corners between the front panel and the sides. If they are warped they should be held in place with panel nails. Glue two additional strips of 1-2-inch round molding at the top inside edge of the sides and nail in place.

Cut three lengths of 1-inch thick acoustical wadding to size and lay them in the cabinet so that two start at the bottom of the speaker, and the third lies over it to the top of the case. (You can substitute roof/wall glass-foam 3-1/2-in. insulation for the acoustical wadding. Remove the insulation from its paper backing.) Fill the space above the speaker with a rolled up piece of wadding. Do not compress the wadding. Make the lengths a few inches longer so that they bend up at the bottom over the tile. Drops of wood glue or any quick-drying cement will hold the wadding in place.

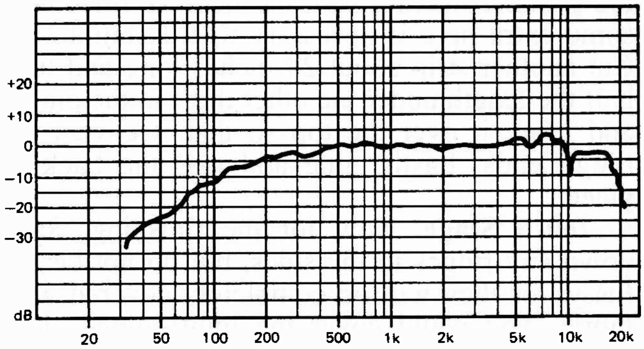

Fig. 27—Curve of the Kapellmeister speaker system frequency response. Some bass boost by the amplifier will straighten the curve at the low end.

Now fit the first baffle with the cut-out at the bottom, gluing to the top edge of both the sides and the upper 1/2-inch round molding surface. Also glue to the top and bottom cheeks. Secure with wood panel nails to ensure a close fit as (b) in Fig. 26. Next, fit the second pair of side pieces and two pairs of blocks top and bottom, (see c). Fit 1/2-inch round molding to corners as with the first channel, also to top edges of the sides; wait for all the glue to dry.

Sixth Stage. Fit the tiles and cement as with the first pair, not forgetting to use concrete-securing screws to top and bottom cheeks in the area to be filled with concrete. This time the top one will be bedded and the bottom one rear-filled. Wait for the concrete to set.

Seventh Stage. Cut two pieces of acoustical wadding about 18-inches long and push half the length of each up the lower channel through the cut-out, and lay the other half length back along the top channel. Now lay three full length strips over these along the complete upper channel. This gives the extra density at the first bend needed to dampen the third-harmonic antinode.

Next comes the second baffle which is glued and nailed as the first but with its cut-out at the top as shown at (d). Fit the third and final pair of sides, also the last pair of blocks plus the 1/2-inch round molding in the corners and top edges, see (e). Allow glue to harden,

Eighth Stage. Now for the last tile, again not forgetting the screws in the top to secure the concrete wedge. Mount the tile on the blocks using wood-panel nails to keep it in place; this will be easier if the enclosure is stood vertically upside down. Return to the assembly to the horizontal, and fill in rear with cement. Wait until set.

Ninth Stage. Lay three strips of wadding in the channel making sure the bend is filled. Put some extra here if necessary to fill completely. Cut another strip about 24-inches long and tuck half the length under the other three at the outlet, and bring it over the top so that it covers the rough ends. Lastly glue and nail the back in place.

Tenth Stage. Now for the finishing. Sand down any ridges in the sides, but do not be too fussy, for they will be completely covered with fabric. You could cover the multi-joined sides with a wood veneer, but it is not mandatory.

Check carefully for any cracks or crevices in the jointing and fill with the wood glue. Sand, then stain or varnish the top cheek and the edges of the bottom one; there is no need to do underneath unless you are fussy. Paint the body with flat black paint including the inside rim of the loudspeaker aperture, but be very careful not to get any paint on the loudspeaker cone as this would affect its flexibility. The painting ensures that the bare wood does not show through the grille fabric with which the whole body excluding the cheeks is covered.

The two back strips (one for each speaker) should now be cut to fit exactly between the top and bottom cheek rear overhangs. These strips conceal the join in the fabric covering so should be about one-inch wide preferably beveled at both edges to give a good finish. They should be stained or varnished the same color as the cheeks.

Eleventh Stage. Obtain sufficient black speaker grille fabric to cover both speakers. Any other color can of course be used if preferred and is obtainable.

Cut the fabric to the exact size to cover the body between the cheek overhangs, but leave a flap 4 inches longer and 8-inches wide, at the start. Secure the vertical starting edge at one edge of the back of the enclosure with tacks so that the flap hangs over the bottom. Then pull it around, keeping it taut, overlapping the start and securing it with one of the wooden strips down the middle of the back. This can be done with brass countersunk screws with cupped washers which give a professional-looking effect.

Next trim the flap to fit between the prongs of the bottom and fix it across the exit port with staples.

Twelfth Stage. Make or purchase two sets of 2-inch high, wood legs. A single rear one should be made to incline backward under the rear exit vent to give greater stability. They can be stained or varnished to match the rest. The Kapellmeisters are now complete and ready to go into action.

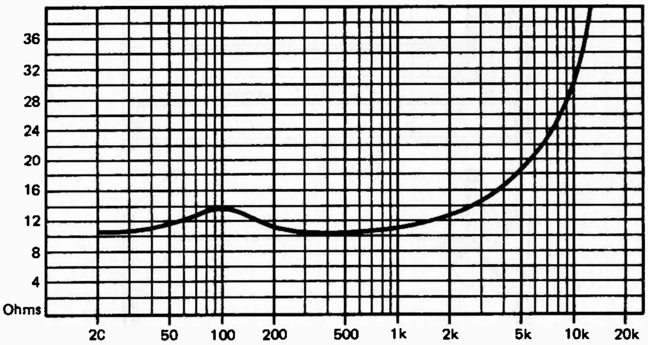

Fig. 28 — The Kapellmeister's impedance curve which is typical of projects using quality loudspeakers.

Performance. As an anechoic room was not available to test the Kapellmeisters, frequency response tests were made with the multiple-microphone-position technique. The test was repeated on the second speaker, and the results were very close, so the plot in Fig. 27 can be considered fairly accurate.

Surprisingly, the treble response is sustained beyond 16 kHz and actually continues up to 20 kHz, a remarkable achievement for a single-driver speaker with no tweeter. This is a tribute to the controlled flexure of the cone and the effectiveness of the high-frequency folded horn inside the cabinet. The undulations are fairly smooth, and some of the vicious peaks and dips encountered with certain multi-driver speakers are notably absent. The response is within 5 dB from 16 kHz to 200 Hz, apart from small deviations at 7 kHz and 10 kHz. There are no phase problems over any part of the range.

Of course the performance varies from loudspeaker type to speaker type, but when you start with a known-good loudspeaker the results will be good.

As expected, the bass is not sustained with a flat response to as low a frequency as would be obtained from a large infinite baffle or reflex enclosure. The response crosses the -5 dB level at 200 Hz and from there a very gentle descent, but audible output is maintained down to around 36 Hz. Actually, there is a 3-dB drop in the octave 500 to 250 Hz, a 6-dB drop from 250 to 125 Hz, and a 12-dB drop from 125 to 62 Hz and below. This gentle and gradually increasing slope results in a more natural and musical bass than when the roll-off is lower, but steeper. A further advantage is that this curve is ideal for applying a little bass boost at the amplifier. All bass boost controls hinge the response curve upward from a pivotal point at 1000 Hz. Frequencies just below 1000 Hz are hardly affected, but the amount of boost increases as the frequency drops. If bass is boosted with speakers having a steep roll-off, the frequencies just above the roll-off point are lifted to produce a hump, so resulting in a boomy effect. Here, bass boost will lift the curve to give a flatter response without boom.

Thus with a little boost in the bass, the single-driver loudspeaker in an enclosure only 8-inches wide can be made to give a response equivalent to that of a much larger multi-driver speaker, but without the phasing problems and distortions they and the crossover network produce.

The impedance curve is also of interest and is here shown (Fig. 28). There is no large peak in the bass, just a small rise at 100 Hz, which is at the third harmonic of the transmission line resonant frequency.

Bass peaks indicate a high back *emf* generated by a large cone movement at the resonant frequency, even though the acoustic output may be damped by the cabinet design. Often these excursions may be large enough to enter the non-linear region or even strain the cone suspension. In such cases bass boost should be applied cautiously and moderately if at all to avoid speaker damage, quite apart from causing a hump in the response as noted above. Here, having no impedance peak, there is no excessive cone excursion at any frequency and no such restraint is necessary. Bass boost can thus be applied to obtain a satisfactory balance without fear of damage, providing the speaker's power rating is not exceeded.

At the high-frequency end, the impedance curve gradually rises to 60 ohms at 20 kHz. This is not due to increasing cone movement but the increase of the coil reactance with frequency. This means that less power is being taken from the amplifier at these frequencies. As amplifier distortion often decreases with an increase of load impedance, this should present a very easy load for any amplifier, and there should be none of the unexpected problems often encountered when an amplifier takes a dislike to a particular speaker. Nowhere does the impedance fall below 10 ohms.

An amplifier of just a few watts may be used for full listening impact in a listening room. Five watts per channel should be sufficient and well within the, speaker's 8-watts rating. You can go the full power rating for a larger listening room. The efficiency afforded by the Kapellmeisters allows the opportunity to use a class A amplifier which can easily be designed for low powers such as these.

14 Window Helix Antenna

A low-cost and effective window-mount vertical antenna can be constructed from insulated hook-up wire and PVC (polyvinyl chloride) piping (Fig. 29). A 5- or 6-foot length of 1-inch PVC piping will do well. If you have a tall window you may be able to use a longer length with a secure attachment. A swivel mounting bracket which is attached to the window sill will permit you to tilt the vertical outward away from the window. When you do so be certain to use a guy arrangement at the top of the window (or even a long bracket) to prevent whipping in the wind.

Good performance on the 11- to 19-meter bands is obtained using a helical-wire length of approximately 15.5 feet which is laced through screw eyes turned into the PVC piping. Space the screw eyes to accommodate the overall length of

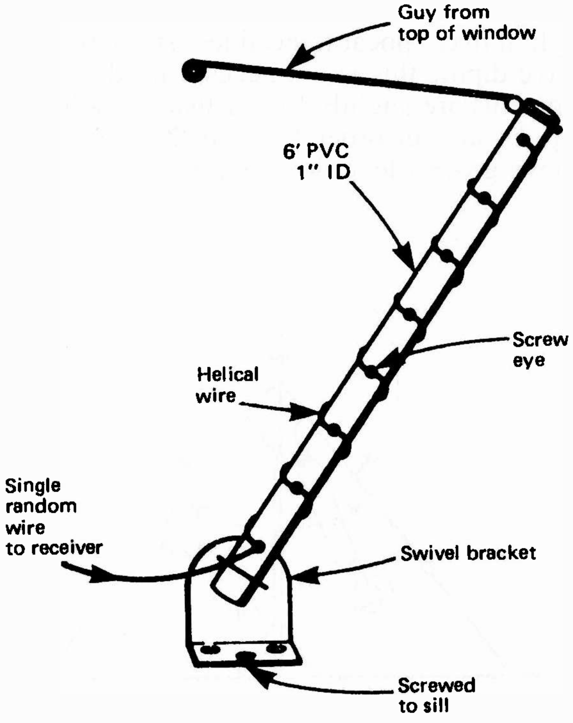

Fig. 29— Construction details for the Window Helix antenna.

wire you plan to use. In place of screw eyes, you may want to cement the wire to the pipe using the glue commonly used to cement PVC pipes together.

A 15.5-foot wire length also resulted in acceptable performance on the 22- to 31-meter bands as well. A random-wire feed was used permitting the position of the helical wire and the wire connecting from the base to the receiver to act as a long-wire so as to improve performance on the lower frequency shortwave bands. A tuner is helpful, though by no means mandatory.

Coaxial feed can also be used. When doing so, connect the outer braid to the swivel bracket.

If you live in an apartment complex where the management objects to window antennas, paint the pipe black so it will be less noticeable and attach a national flag to it. It would be un-American or un-Canadian to ask you to take it down, especially if you first raise the flag on a national holiday.

15 Inverted V Antenna

It may not appear so, but the Inverted V Antenna is really a modified version of the half-wave dipole. Looking at Fig. 30, it should be obvious how the Inverted V name is obtained and its relationship to the dipole antenna.

The Inverted V Antenna configuration has similar characteristics to a half-wave dipole and has the obvious advantage of needing only one mast or other central support. Although at first sight it may appear to need less space than a half-wave dipole this is not necessarily the case. The elements are slightly longer than for a half wave dipole, and in order to keep the elements well above ground level it is either necessary to have a tall mast (especially for an antennas cut for a low frequency band) or keep the angle between the elements quite large and have very long supporting lines. Angles of the Inverted V are typically 60 to 120 degrees. This type of antenna is probably most suitable for use in situations where suitable supports for the antennas are already available or are easily improvised. For example, antennas of this type are often installed over the pointed roof of a house (but must be kept reasonably well clear of the roof in order to obtain good results).

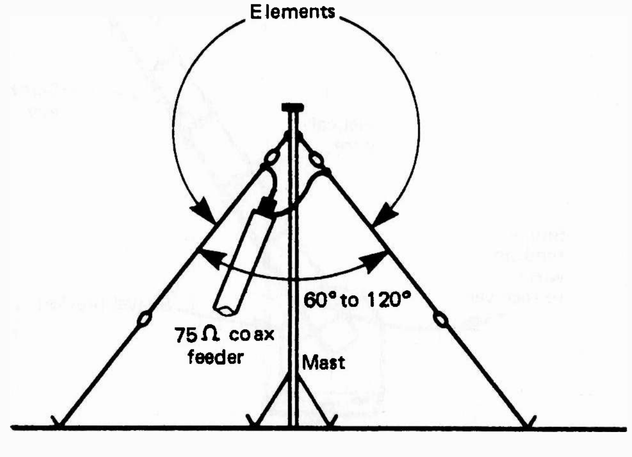

Fig. 30 —The arrangement used for the Inverted V Antenna. The length of one element of the "V" are given in the table.

The length of an Inverted V Antenna in feet is equal to 486 divided by the frequency in megahertz, or in meters it is equal to 148 divided by the frequency in megahertz. This is the total length of the two elements, and must be halved to give the length of each element.

The lists provided below give the element lengths (not the overall lengths) for Inverted V antennas for use on the shortwave, amateur and broadcast bands.

One-half Length (One Element) of Inverted V Antenna

Amateur Band (Meters)	Antenna Length (Feet)	Amateur Band (Meters)	Antenna Length (Feet)
10	8.40	29.5 (30)	23.98
12	9.74	40	34.45
15	11.45	80	66.50
16.5 (17)	13.39	160	127.79
20	17.13		

Broadcast Band (Meters)	Antenna Length (Feet)	Broadcast Bands (Meters)	Antenna Length (Feet)
11	9.38	41	33.73
13	11.22	49	39.96
16	13.68	60	49.24
19	15.81	75	61.09
22	17.72	90	73.56
25	20.47	120	101.21
31	25.03		

16 Audible Timer

This easy-to-build-timer circuit can be set for any period of minutes, from 1 minute to 7 minutes. A rotary switch is used to select the timing period.

The position to which the switch is set is indicated by numerals marked on the panel. A push-button resets the circuit and timing begins from the instant the button is released. At the end of the period there is a loud audible signal from the speaker. This is an intermittent note for maximum impact against a noisy background. The note continues for 1 minute. As well as the sound signal there is a high-intensity LED which comes on during the timing period and flashes when the period ends.

One of the advantages of this timer is that it can be set and run in the dark. This makes it suitable for dark-room timing. It is also ideal as a timer for use by blind or deaf persons.

How It Works. The Audible Timer relies on the principle of obtaining a relatively long timing period by dividing down the frequency of a high-frequency astable multivibrator. The chip used for the multivibrator is the well-known 555 timer in its CMOS form (Fig. 31). The frequency is independent of variations in the level of the supply voltage, so the circuit is ideal for battery operation. The preset resistor VR1 allows the frequency to be set to 273 Hz. This provides the alarm tone. The 273 Hz signal is divided by the 14-stage counter IC2. At pin 8 of the counter we obtain a signal at approximately 1 Hz which is used to make the note intermittent. The timing frequency comes from pin 14 of the IC, at which the astable frequency is divided by 2^{14}, or 16384, giving a frequency of 1/60 Hz, or 1 count per minute.

The output from pin 14 is inverted by transistor TR1 and fed to the clock input of a second counter IC3. This is a decade counter with decoded outputs. The outputs of this counter are normally at logical high, except that just one of the outputs is low at each stage of counting. When the counter is reset, output 0 goes high. On the next positive-going clock input, output 0 goes low and output 1 goes high. At each successive high-going clock input the outputs go high in turn from 0 to 7, repeating. The rotary switch S3 selects which of the outputs is to be used to indicate the termination of the timing period.

The circuit is reset by pressing button S2, which resets both counters and also the flip-flop formed by gates IC4a and IC4d. All outputs of IC2 go low. All outputs of IC3 go low, except for output 0, which goes high. Output 0 is not used in the circuit. A low output from one of the other pins of IC3 makes the output of the flip-flop (IC4 pin 3) go high. The output of IC4c goes low. This turns off the npn transistor TR2, which drives the loudspeaker, but turns on the pnp transistor TR3, causing the LED to light.

While the reset button is held, the output from IC2 pin 3 is low, turning off TR1. This means that the input to IC3 pin 14 is high. Counting begins when S2 is released. After 30 seconds, IC2 pin 3 goes high, and the input to IC3 goes low, but this has no effect on IC3. After 1 minute IC2 pin 3 goes low, and the input to IC3 goes high, causing

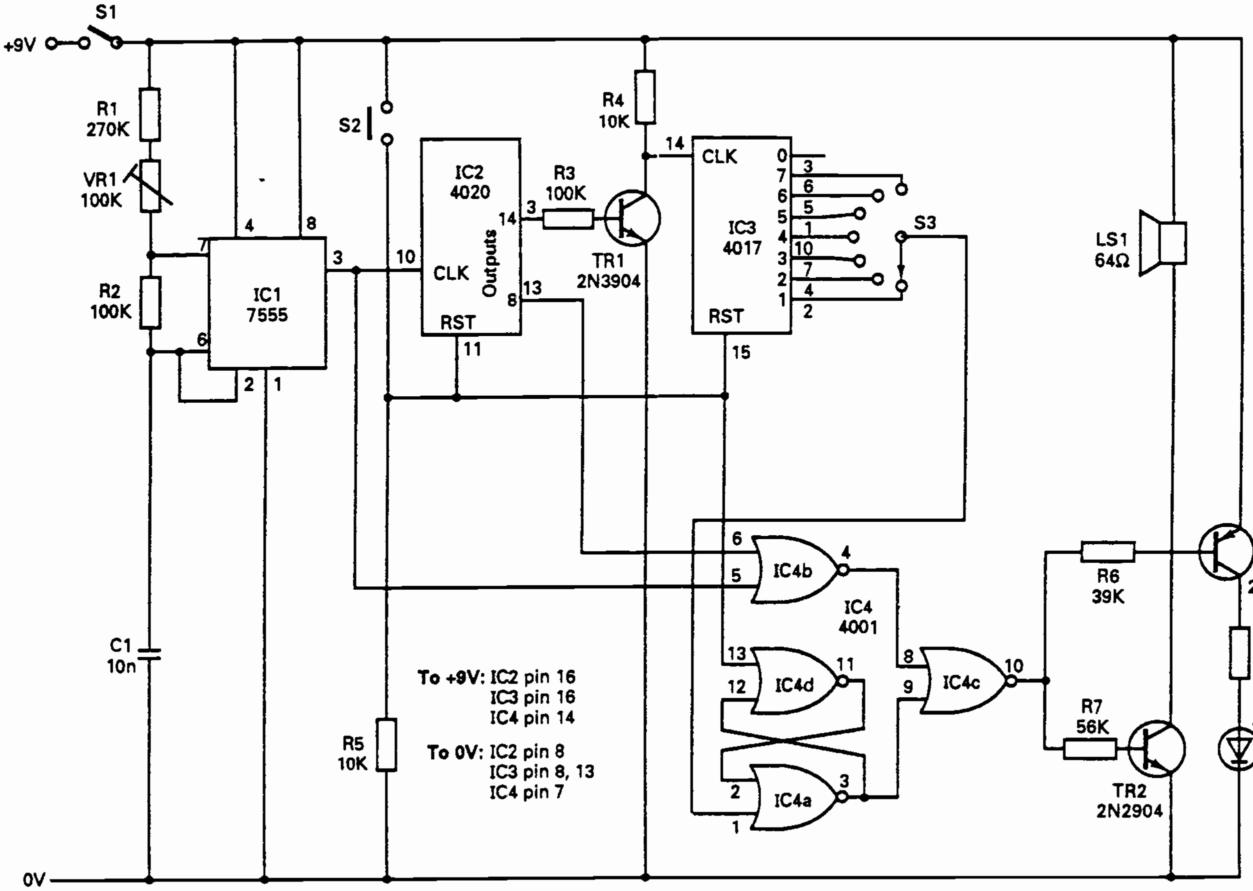
S1
+9V
R1
270K
VR1
100K
R2
100K
C1
10n
IC1
7555
4
8
7
6
3
2
1
S2
IC2
4020
CLK
RST
Outputs
10
11
14
3
13
8
R3
100K
TR1
2N3904
R4
10K
14
CLK
IC3
4017
RST
15
0
7
6
5
4
3
2
1
3
6
5
1
10
7
4
2
S3
LS1
64Ω
IC4b
IC4d
IC4c
IC4a
IC4
4001
6
5
4
13
12
11
8
9
10
2
1
3
R6
39K
R7
56K
TR2
2N2904
TR3
2N3906
R8
390
LED1
R5
10K
To +9V: IC2 pin 16
IC3 pin 16
IC4 pin 14
To 0V: IC2 pin 8
IC3 pin 8, 13
IC4 pin 7
0V

Fig. 31—Circuit diagram for the Audible Timer.

PARTS LIST FOR THE AUDIBLE TIMER

Semiconductors
IC1—7555 timer
IC2—4020 counter chip
IC3—4017 counter/divider, 10-line output chip
IC4—4001 quad 2-input
LED1—Light-emitting-diode, high-intensity red
TR1—2N3904 transistor, npn
TR2—2N2904 transistor, npn
TR3—2N3906 transistor, pnp
Resistors
R1—270,000-ohm, metal-film
R2—100,000-ohm, metal-film
R3—100,000-ohm
R4, R5—10,000-ohm
R6—39,000-ohm
R7—56,000-ohm
R8—390-ohm
VR1—100,000-ohm linear-taper potentiometer
Additional Parts & Materials
C1—1-nF, polystyrene capacitor
LS1—Loudspeaker, sub-miniature, 1.5-in. dia., 64-ohm
S1—Switch, s.p.s.t. (on/off), toggle or slide
S2—Switch, push-button, non-locking
S3—Switch, rotary, 7 positions minimum
9-volt transistor-radio battery, battery connector, plastic case, wire, perf-board, solder.

this counter to advance one count. Its output 1 goes high. The counter is incremented every minute until the output selected by S3 goes high.

The high level from the selected output sets the flip-flop. IC4 pin 3 goes low. This makes pin 9 low, allowing the signal reaching IC4c from IC4b to pass through to TR2 and TR3. This signal consists of the logical NOR of the astable signal (273 Hz) and the signal from output 8 of IC2 (1 Hz). The result is a note at 273 Hz, pulsing at the rate of 1 Hz. This is heard from the loudspeaker. The LED flashes on at the same time. Since the mark-space ratio of the astable output is high, the LED is turned on at almost full brightness during the on periods.

Construction. This circuit requires only a small current and is powered from a 9-volt transistor radio battery. Assuming that a small (1.5-in. diameter) loudspeaker is used, it can be housed in a small plastic case with the circuit. The parts can be assembled on a perf-board or some other pre-drilled board with holes on a 0.1-in grid. For stability of timing, and minimum effect of temperature, use metal-film resistors for R1 and R2 and a polystyrene capacitor for C1.

Assemble the astable circuit and the first counter (IC1 and IC2). Check that all outputs of IC2 go low when S2 is pressed. When S2 is released, a voltmeter shows that the output at pin 13 is approximately 1 Hz and that pin 3 changes from high to low or low to high every 30 seconds. Adjust VR1 to obtain this timing. While the remainder of the circuit is being built, it is convenient to use the output from IC2 pin 1 to drive TR1. The signal there runs at 1/15 Hz thus saving prolonged periods of waiting while testing the circuit.

Wire up TR1 and IC3, and test their operation, then complete and test the remainder of the circuit.

17 Long-Interval Alarm Timer

The Long-Interval Alarm Timer is capable of timing intervals of hours or even days. The maximum interval is two weeks. The timer sounds an alarm when the selected interval has elapsed. In addition, it sounds a different alarm if any attempt is made to turn it off before the end of the interval. In this way the unit can convey two messages: "Interval still in progress!", and "Interval finished!". While it is timing, the unit uses very little current (2.5 mA) so it is suitable for battery operation over prolonged periods.

The circuit was intended as a pill-box reminder, though it has other applications. If you have to take pills regularly, it is all too easy to forget exactly when the next dose is due. Conversely, it is easy to forget that a dose has been taken only recently and taking a second dose too soon could have undesirable effects, if nothing less than wasting money on expensive medicine.

The Long-Interval Alarm Timer can not prevent a person taking out too many pills when the pill-box is opened, neither can it make sure that the person actually swallows the pills removed, but, with intelligent use, it is a helpful personal prompter.

Pill Box. The circuit is housed in a small box serving as the pill-box. A micro-switch activated by the lid of the box lets the circuit know when the box is opened. The warnings are given by an audible sounder and by a pair of flashing LEDs. Shutting the box (presumably after having taken a pill) starts the long-period timer; both LEDs are out and there is no alarm sound. If the box is opened before the next time that a dose is due, an urgent high-pitched alarm sounds, bleeping rapidly, and the red LED flashes. This warns the person that it is not yet time for a pill. The person shuts the box, the sounder and LED cease their activity, and timing continues unaffected by the action of opening the box. When a dose is due, a slower, deeper bleeping note is heard and a green LED flashes. This continues until the box is opened.

The LEDs are of different shape, the red one

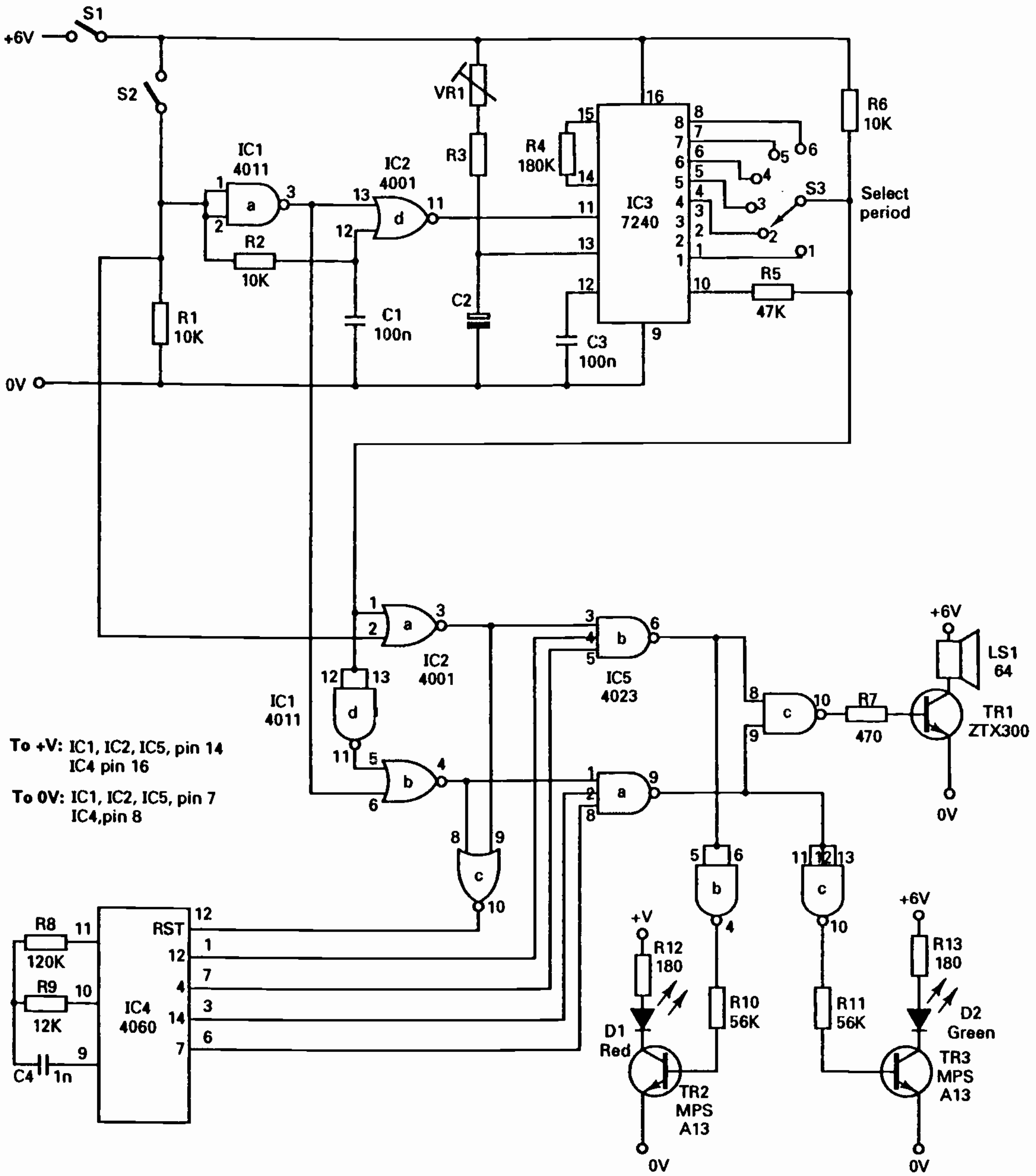
+6V
S1
S2
IC1
4011
a
1
2
3
R2
10K
IC2
4001
13
12
d
11
VR1
R3
C2
C1
100n
R1
10K
0V
R4
180K
15
14
11
13
12
16
IC3
7240
8
7
6
5
4
3
2
1
10
9
C3
100n
R5
47K
R6
10K
S3
Select period
IC2
4001
a
b
c
d
IC1
4011
IC5
4023
R8
120K
R9
12K
C4
1n
IC4
4060
RST
+6V
LS1
64
R7
470
TR1
ZTX300
0V
+V
R12
180
D1
Red
R10
56K
TR2
MPS
A13
0V
R11
56K
+6V
R13
180
D2
Green
TR3
MPS
A13
0V

To +V: IC1, IC2, IC5, pin 14
IC4 pin 16

To 0V: IC1, IC2, IC5, pin 7
IC4, pin 8

Fig.32—Circuit diagram for the Long-Interval Alarm Timer.

being triangular to reinforce the warning message. The difference of shape avoids the risk of confusion by color-blind persons.

How It Works. This project is based on the 7240 CMOS programmable timer chip (IC3), which has an accuracy of 0.5%. It contains a time-base generator, the frequency of which is decided by a resistor and capacitor (R3/VR1 and C2 in Fig. 32). The basic time period is RC seconds; so, given that the maximum value of R is 10 Megohms and the maximum value of C is 1000 -μF , the maximum time period is 10,000 seconds, or 2.8 hours. IC3 also has an 8-stage binary-divider chain. The total period available is 27 times the above, which is just over 14 days. But, as explained below, there may be practical problems in obtaining the very longest periods.

In Fig. 32, S1 is an optional power switch. S2

PARTS LIST FOR LONG-INTERVAL ALARM TIMER

Semiconductors
D1—Light-emitting diode, red (see text)
D2—Light-emitting diode, green (see text)
IC1—4011 2-input NAND chip
IC2—4001 2-input NOR chip
IC3—7240 **CMOS** programmable timer chip
IC4—4060 14-stage counter with oscillator chip
IC5—4023 3-input NAND chip
TR1—ZTX300, SK3854 or ECG123AP,
TR2, Tr3—MPSA13 high-gain Darlington
Resistors
R1, R2, R6—10,000-ohm
R3—See text to determine value
R4—180,000-ohm
R5—47,000-ohm
R7—470-ohm
R8—120,000-ohm
R9—12,000-ohm
R10, R11—56,000-ohm
R11—56,000-ohm
VR1—(See text)
Capacitors
C1, C3—.1-μF
C2—(See text)
C4—1-nF
Additional Parts & Materials
B1—Battery consisting of 4 AA alkaline cells and holder
LS1—Loudspeaker, 1.5-in. dia, 64-ohms (approximately)
S1—Microswitch, normally-open (See text)
S2—Rotary switch, miniature, non-shorting, 6 (or more) positions
Plastic case, 4- x 2.5-in. circuit board builder fabricates, knob, wire, solder, etc.

is the switch which is closed to initiate timing. In the pill-box application, this could be a microswitch, mounted so that it is closed when the lid of the pill-box is closed. Closing S2 generates a brief low pulse which goes to pin 11 of the timer IC3 and starts the timing. The output of the timer is normally high, but goes low for the whole of the timing interval. The length of the interval is selected by a rotary switch S3. If you have a dedicated version planned for the timer project, you could hard-wire the switch out of the circuit.

Pins 1 to 8 of IC3 are the outputs from the 8-stage divider chain. When the counter is reset they all go high, while timing they go through an inverted binary sequence. Thus pin 1 goes low after the basic time period, while pin 4, for example, goes low after 8 times the basic period. R5 connects the output to the reset terminal (pin 10) so that the counter is reset at the end of the interval. Figure 32 shows only 6 outputs being used; this is sufficient for the pill-box application, but there is no reason why all 8 outputs or any combination of fewer outputs should not be selected. For a single-period timer, omit S3 and wire the chosen output directly to the junction between R5 and R6.

The first stage of the logic consists of two NOR gates and a NAND gate wired as an inverter (part of IC1, 4001). These gates detect the two alarm states. Gate IC2a goes high (at pin 3) when S2 is open (box open) and the IC3 is timing (output low). This is the state of opening the box while the timing interval is in progress. Gate IC2b goes high (at pin 4) when S2 is closed and the output of IC3 is high. This is the state after the end of the interval if the box remains unopened.

A high output on either pin 8 or pin 9 causes a low output from Gate IC2c (pin 10). This makes the reset input of IC4 low. IC4 is a 14-stage counter with its own oscillator, which begins to oscillate when the reset is made low. The oscillator has a period of about 25 kHz, which is divided down to produce 1.6 kHz at pin 7 (high-pitched note), 200 Hz at pin 6 (low-pitched note), 6 Hz at pin 1 (fast bleeping), and 1.5 Hz at pin 3 (slow bleeping). The remainder of the logic consists of gates producing the fast high-pitched bleep signal, which goes to the red LED (D1) by way of transistor TR2, and the slower low-pitched bleep, which goes to the green LED (D2) by way of TR3. Both signals go to the speaker LS1 by way of TR1.

Putting It Together. The project is built on a board only 4 in. (10 cm) by 2.5 in. (6.5 cm). With a miniature loudspeaker, 1.5 in. (38 mm) in diameter, the unit is housed in a reasonably small plastic box. Since it requires only 2.5 mA, the timer runs for about a month on a set of four AA alkaline cells. There are only 3 logic gate ICs, IC1 (2-input NAND, 4011), IC2 (2-input NOR, 4001) and IC5 (3-input NAND, 4023) and all gates are used. The gates belonging to these three ICs are scattered in ones and twos all over the circuit diagram but can easily be identified by their symbols.

Begin with the trigger circuit IC1a/IC2d and timer 1C3, and their associated resistors and capacitors. For use as a pillbox timer, with a maximum interval of 8 hours (IC3, pin 8) values for the timing components are R3 = 3.3 Megohm, VR1 = 2.2 Megohm and C3 = 47 μF . Use a tantalum capacitor for C3. The tolerance of these is ±20%, but VR1 allows timing to be adjusted to accommodate this.

The main problem with tantalum and aluminum electrolytic capacitors is that they have an appreciable leakage current. The effect of this is that charging times tend to be longer than those calculated by multiplying R and C together. As a rule, tantalum capacitors have lower leakage than aluminum electrolytic capacitors. The most commonly available types have a leakage of 0.02 µA per volt per µF , with a minimum leakage of 1 µA. Low-leakage types are available with half this leakage, though such types are not widely stocked. The leakage quoted above is for a capacitor operating at its maximum working voltage; leakage is markedly reduced when operated at a lower voltage. The voltage across the capacitor in this project ranges from 1.6 to 4.2 volt, but the working voltage is likely to be 16 WVDC or even 20 WVDC for a 47-µF capacitor. This helps to keep leakage to an acceptable level. Using a 47-µF tantalum capacitor with resistances totaling 5 megohms or more is feasible. If really long intervals are required, make the resistor slightly less than the calculated value.

Having settled on a suitable value for R3, adjust VR1 until an interval of 225 seconds is obtained with S3 switched to the calibration position, position 1. This gives intervals of 8 hours, 4 hours, 2 hours , 1 hour and 30 minutes from pins 8 to 4. Other values of R3 and C3 may be substituted for other applications.

Next assemble the oscillator circuit, IC4 with resistors and capacitor. A voltmeter monitoring the output at pin 14 confirms that this works correctly when pin 12 is made low. Assemble the rest of the logic and check that the outputs at IC1 pins 4 and 10 and IC5 pin 10 behave as described earlier. Finally, add the transistor switches. TR2 and TR3 are shown in the schematic diagram as single NPN transistors, but the MPSA13 is a high-gain Darlington transistor. High-intensity LEDs are preferred but a red triangular one of normal intensity may be used for D1.

18 Twangy Distortion Unit

A conventional distortion unit has a major effect on the processed signal. Apart from adding distortion products, it also changes the input signal's envelope. In other words, it alters the way in which the volume of each note varies. Normally the output from a guitar-string pluck has a high initial level, but it rapidly falls away to a much lower level, and then decays at a more gradual rate. The envelope of a signal is the all-important parameter that governs its precise character and sound. In the case of a guitar it is the fast attack and quite rapid initial decay that gives the characteristic "twangy" guitar sound.

Clipping the signal from a guitar tends to give an output signal of uniform amplitude right up to the point where the input signal decays away to practically nothing. The fast attack of the input signal is retained, but the rapid fall in amplitude immediately after the attack period is not. This gives a sound which is largely twang free! Even with soft clipping, the amount of compression applied to the input signal is such that the original envelope of the signal is largely lost.

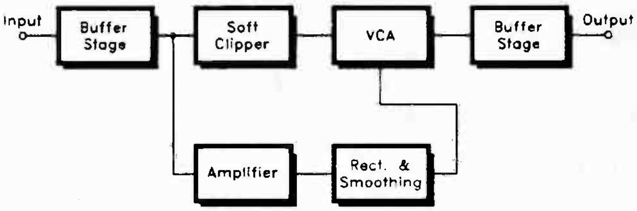

Fig. 33 —Block diagram for the Twangy Distortion Unit.

It is possible to produce a distortion unit that retains the original envelope of the input signal, or something close to it. It requires a rather more complex circuit than an ordinary distortion unit, but our Twangy Distortion Unit is reasonably

PARTS LIST FOR TWANGY DISTORTION UNIT

Semiconductors
IC1—NE5532N, very low-noise, dual-operational amplifier
IC2—LM13700N or LM13600N, very low-noise, dual-operational amplifier
IC3—LF351N JFET wide-band operational amplifier
D1-D8—OA91 diode

Resistors
R1, R2, R14—100,000-ohm
R3—1,000-ohm
R4, R10—10,000-ohm
R5, R6, R17—3,900-ohm
R7, R8—470-ohm
R9, R11, R13—18,000-ohm
R12—270,000-ohm
R15—27,000-ohm
R16—33,000-ohm
VR1—100,000-ohm, linear-taper, carbon
VR2—4,700-ohm, miniature, trimmer
VR3—470,000-ohm, miniature, trimmer

Capacitors
C1—10-nF ceramic
C2—1-µF, 50-WVDC, electrolytic
C3—22-µF , 16-WVDC, electrolytic
C4—470-µF , 10-WVDC, electrolytic
C5—10-µF , 25-WVDC, electrolytic
C6, C7—2.2-µF , 50-WVDC, electrolytic
C8—1-µF , 50-WVDC, electrolytic
C9—4.7-µF , 50-WVDC, electrolytic

Additonal Parts & Materials
B1—9-volt, transistor-radio battery and clip connector
JK1, JK2—Audio jack (select to match audio system cables)
S1—S.p.s.t., miniature, toggle switch
S2—S.p.s.t., push-button, locking switch
8-pin IC socket (2 required), 16-pin IC socket, control knob, aluminum case, circuit board, wire, solder, etc.

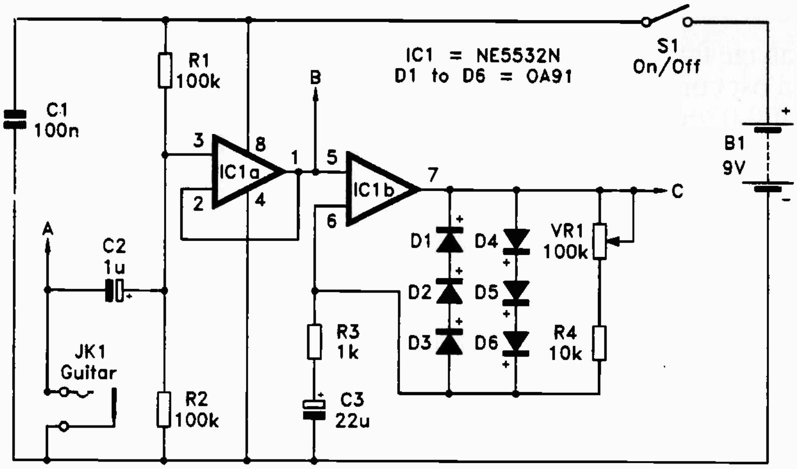

Fig. 34—Circuit diagram for the input buffer and clipping amplifier stages of the Twangy Distortion Unit.

simple and inexpensive to build.

Signal Flow. The block diagram of Fig. 33 shows the arrangement used in the Twangy Distortion Unit. The input signal is applied to a buffer stage, and the signal is then split two ways. The main route is via a soft clipping amplifier, or a hard clipping amplifier can be used here if preferred. The distorted and compressed output signal from the clipping circuit is fed to a voltage-controlled amplifier (VCA), and then to the output via a buffer amplifier. The amplitude of the output signal is governed by the control voltage fed to the VCA. The higher the control voltage to the VCA, the higher the amplitude of the output signal. By feeding the VCA with a suitable control voltage, the output signal can be molded into the required envelope shape.

The control voltage is produced by a side chain which first amplifies the input signal from the buffer stage. The amplified signal is fed to a rectifier and smoothing circuit. This produces a DC output voltage that is roughly proportional to the amplitude of the input signal. Via the VCA, this DC voltage modulates the output signal to give an output envelope shape that is reasonably close to the envelope shape of the input signal. The output envelope is not identical to the input envelope because a few small errors occur in the amplitude control circuits. However, the match is close enough to give the required twangy sound.

The Circuit. Figure 34 shows the circuit for the input buffer and clipping amplifier stages. The circuit for the VCA and other stages is shown in Fig. 35. Taking Fig. 34 first, the clipping amplifier is based on chip IC1b. If hard distortion is required, simply replace D1 to D6 with a pair of 1N4148 silicon diodes, one diode for each string of three (D1-D3 and D4-D6) connected as in Fig.34; IC1a acts as the input buffer amplifier, and this is a simple non-inverting type which provides an input impedance of 50,000 ohms. IC1 is an NE5532N, which is a very low-noise dual operational amplifier.

A transconductance operational amplifier (IC2) is used as the basis of the VCA. Although this is a form of operational amplifier, apart from differential inputs it has little in common with ordinary operational amplifiers such as the 741C and LF351N. A transconductance amplifier is current rather than voltage operated. The output current is controlled by the differential input current. In practical circuits, including the present one, series resistors at the inputs plus a load resistor at the output effectively convert the device to a form of voltage operation. R10 is the resistor in series with the input signal, and R9 is the output load resistor. C4 plus R5 to R8 are used to provide biasing to the inputs of IC2, and they also provide a center-tap on the supply lines which is used as a sort of central ground bus for the output load resistor. Unlike ordinary, operational amplifiers, transconductance amplifiers are often used "open loop". IC2 is certainly used in this manner, and it therefore lacks any form of negative feedback circuit.

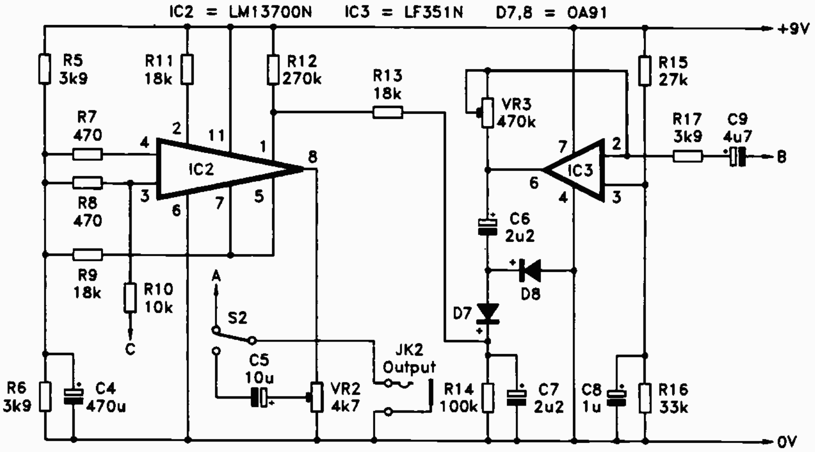

Fig. 35—The circuit for the voltage-controlled amplifier and rectifier stages of the Twangy Distortion Unit.

Chip IC2 has a built-in emitter-follower buffer stage which can be connected at the output of the transconductance amplifier. VR2 is the load resistor for this stage, and it also acts as a variable output attenuator. This enables the output level of

Continued on page 107

Many More Projects You Can Build

The project sources for *The Project Anthology* came from selected inexpensive books that are listed below. You can expand your personal *Project Anthology* by purchasing one or more of these easy-to-read books today! Make your selection by checking the applicable box(es) and filling out the coupon below. Get started today!

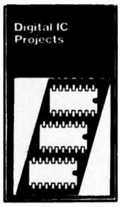

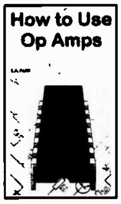

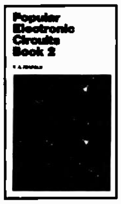

30 SOLDERLESS BREADBOARD PROJECTS
Book 1

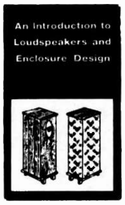

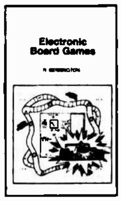

☐ **BP84—Digital IC Projects $5.50** Presents simple and more advanced projects that help the reader develop a working knowledge of digital circuits. To help the newcomer to the hobby, the author has included a number of board layouts and wiring diagrams.

☐ **BP88—How to Use Opamps $5.95** A project builder designer's guide that covers operational amplifiers. Serving as both a source book of circuits and a reference book. The approach has been made as non-mathematical as possible and should be easily understandable by hobbyists.

☐ **BP98—Popular Electronic Circuits-Book $5.95** Designs for electronic enthusiasts who are capable of producing working projects from just a circuit diagram without the aid of detailed construction information. However, where relevant, any special procedures are described.

☐ **BP103—Multi-Circuit Board Projects $5.50** Contains information on the construction of 21 simple electronic projects, all of which may be constructed on the same specially designed printed circuit board. Many components may be used several times, reducing the cost of each project.

☐ **BP105—Aerial (Antenna) Projects $5.50** Your radio's performance ultimately depends on the antenna. This book covers practical antenna designs, including active, loop and ferrite antennas which perform well and are relatively simple and inexpensive to build.

☐ **BP106—Modern Opamp Projects $5.75** Construction projects that make use of many types of operational amplifiers available to hobbyists, including low-noise, low-distortion, ultra-high input impedance, low slew-rate and high-output current types.

☐ **BP107—30 Solderless Breadboard Projects $5.95** Each project is designed to be built on a "Verobloc" breadboard and presented with a brief circuit description, circuit diagram, component layout diagram, components list. Notes on construction and use are provided where necessary.

☐ **BP136—25 Simple Indoor and Window Antennas $5.50** Written for those people who live in apartments, have no garden or yard areas, or have other space-limiting restrictions that prevent them from constructing a conventional antenna system. The antennas have been designed, built and tested.

☐ **BP256—An Introduction To Loudspeakers and Enclosure Design $5.95** All you need to know about the theory and operation of loudspeakers and the cabinets that house them. Book includes the complete construction details of how to make an inexpensive, high-quality enclosure called the "Kapellmeister."

☐ **BP317—Practical Electronic Timing $6.95** Time measurement projects are among the most constructed gadgets by hobbyists. This book provides the theory and backs it with a wide range of practical construction projects. Each project has how-it-works theory and how to check it for correct operation.

☐ **BP350—Electronic Board Games $6.00** Here are twenty novel electronic board games that you can build. Whether you are interested in motor racing, searching for buried treasure on a desert island or for gold in Fort Knox, spinning the wheel of fortune, a musical quiz—there is something for you to build and enjoy!

☐ **BP368—Practical Electronics Musical Effecst Units $5.95** There is a constant clamor for musical effects projects by the hobbyist community. This book provides practical circuits for several projects that range in complexity and are sure to work. All the circuits are easy to build and use readily-available parts.

SHIPPING COSTS

$5.01 to $10.00	$2.50
$10.01 to $20.00	$3.50
$20.01 to $30.00	$4.50
$30.01 to $40.00	$5.50
$40.01 to $50.00	$6.50
$50.01 and above	$8.00

Telephone Orders: If you wish to place your credit card order by phone, call 516/293-0467. Automated order taking system functions 24 hours a day. Have your credit card ready. Minimum order is $25.00. Sorry, No orders accepted outside of U.S.A. and Canada. New York state residents must add applicable sales tax.

Please return this order form to:

ELECTRONIC TECHNOLOGY TODAY, INC.
P.O. BOX 240
MASSAPEQUA PARK, NY 11762-0240

To purchase one or more books, check the applicable box(es) and add up the cost(s) for the book(s). You may pay by check, money order or credit card in US funds. Do not send cash. Sorry, US and Canada orders only.

Total Amount $ ________
Add shipping cost (see table) $ ________
Add Local NY State Sales Tax $ ________
TOTAL COST (US Funds Only) $ ________

Name ____________________________

Address ____________________________

City ____________ State ________ ZIP ________

If you wish to use a Credit Card: ☐ MasterCard ☐ VISA

Card No. ____________________ Expire Date ___/___

SIGNATURE ____________________
(Do not print)

Allow 6-8 weeks for order to be fulfilled.

MILLIOHM
ADAPTER

MILLIOHM TESTER

Measure low resistances with this super simple circuit that connects to your DMM.

ROBERT A. HEIL

THE MILLIOHM ADAPTER ALLOWS resistances from 1 milliohm to 1 ohm to be measured with a high degree of accuracy on any digital multimeter. The circuit loads the device under test with a current of 100 milliamperes at 5 to 6 volts. The adapter connects to a DMM that is set on its millivolt- or 2-volt scale.

Ohm's law says that resistance equals voltage divided by current, or R = V/I. Thus, a DMM reading of 5.7 millivolts would correspond to 0.057 ohms. (5.7 mV/100 mA = 57 milliohms or .057 ohms.)

Circuit description

The milliohm adapter circuit, shown in Fig. 1, is powered from a 9-volt battery. A resistor to be tested (R_X) is connected across banana jacks J1 and J2, and a pair of banana plugs, connected directly to J1 and J2, plugs into the voltage input jacks of a DMM.

Switch S1 applies battery power to 7806 voltage regulator IC1. Capacitor C1 removes voltage transients. Resistors R1 and R2 form a voltage divider for the ground pin of IC1. Potentiometer R2 trims IC1's output voltage to exactly 6-volts DC. Potentiometer R3 sets the output current through R_X to 100 milliamperes. Because R3 is a relatively large resistance compared to R_X, the error introduced by different values of R_X (1 milliohm to 1 ohm), or the effect it will have on the 100-milliampere current source, is below 2%.

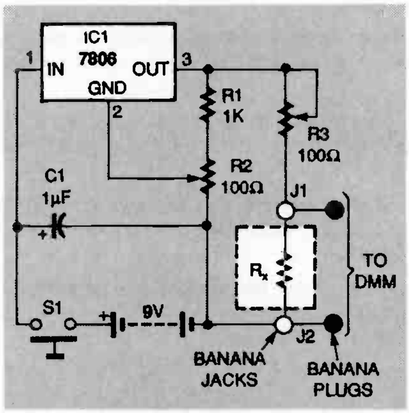

FIG. 1—MILLIOHM ADAPTER CIRCUIT. A resistor to be tested (R_X) is connected across banana jacks J1 and J2, and a pair of banana plugs, connected directly to J1 and J2, plugs into the voltage input jacks of a DMM.

Construction

A PC board is available from the source given in the Parts List, but the project is also easy to breadboard. You must select a case for the project before beginning the assembly. The prototype's case measures approximately 2 by 3¼ inches and is about 1 inch deep. The case has an aluminum cover.

A parts-placement diagram is shown in Fig. 2. Stuff the board as indicated, and check your work before continuing. Remove the covers from the banana plugs and place them into the common and the volt/ohms terminals of your DMM. Dab some petroleum jelly or other similar substance on the ends of the jacks protruding from the DMM. With the lid attached to the project case and facing up, press the upper left back of the plastic case onto the ends of the plugs stuck in the DMM. The petroleum jelly will transfer onto the project case and will mark the hole locations for drilling, allowing the adapter to plug directly into the DMM. Permanently mark those locations before continuing. Then turn the case over and similarly mark

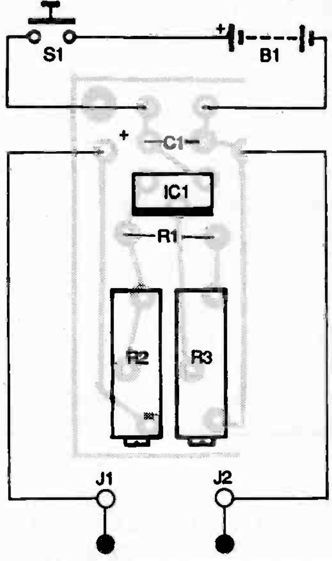

FIG. 2—PARTS-PLACEMENT DIAGRAM. This circuit is simple enough to point-to-point wire.

the locations on the lid of the case for the two banana jacks directly above the holes for the plugs.

Drill appropriately sized holes in the case bottom and lid for the banana plugs and jacks. Be sure to use insulated jacks if your case has an aluminum nal plug covers. Next install the

0.00
Ω
V
A~
A—
OFF
WAVETEK
2010
10A
100m
MILLIOHM ADAPTER

THE ADAPTER looks like it was custom made for this meter.

PARTS LIST

- R1—1000 ohms
- R2, R3—100 ohms, 20-turn potientiometer
- C1—1 µF, disc capacitor
- IC1—7806 6-volt regulator
- B1—9-volt battery
- S1—SPST momentary switch
- J1, J2—banana jack/plug combo
- Project case (Radio Shack No. 270-230 or similar unit), PC board, wire, solder.

Note: the following items are available from RAH Projects, P.O. Box 15904, N.B., CA 92659:

- **Etched and drilled PC board—$3.95 plus return postage**
- **Parts kit including PC board (no case)—$10.95 plus $2.50 S&H**

Check or money order, only. California residents please add sales tax. Allow 4 to 6 weeks for delivery. Personal check orders will be shipped after the funds have been cleared.

cover. Solder a 3-inch piece of test lead wire to each banana plug and install them in the bottom of the plastic case with No. 10-32 nuts instead of the origi-

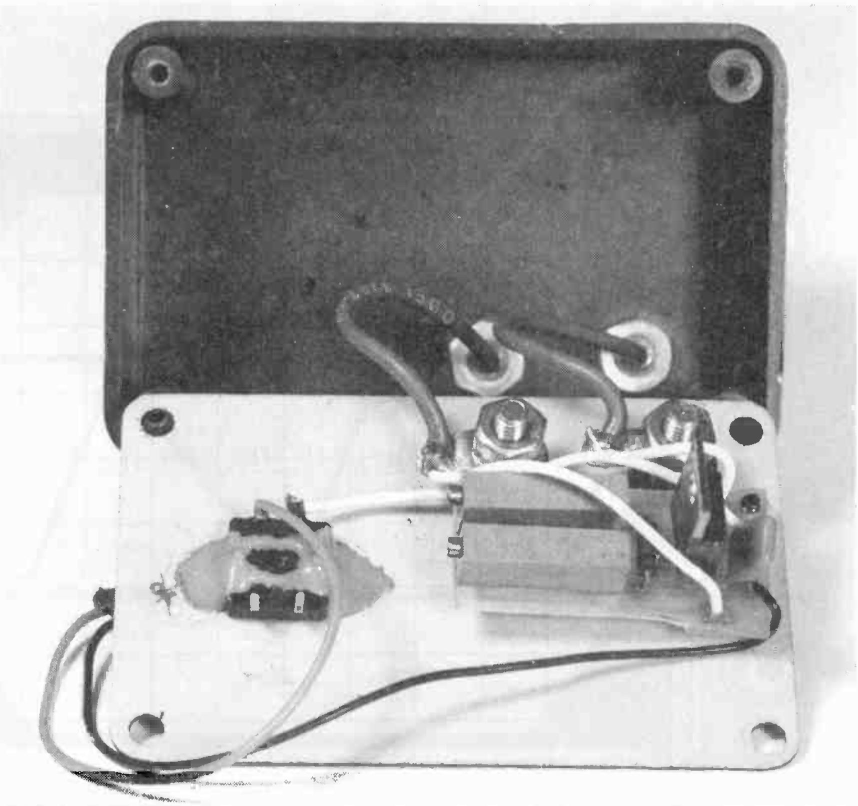

FIG. 3—THE INSIDE of the completed unit. The case is large enough to contain the 9-volt battery.

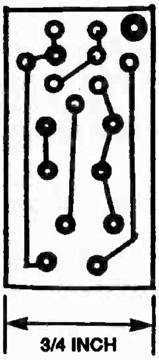

MILLIOHM ADAPTER FOIL PATTERN.

banana jacks in the case cover. Mount the PC board on the cover as close to the jacks as possible. You can secure it with No. 4-40 hardware on one corner with a plastic spacer.

Any momentary switch will work in the circuit, but make sure there's enough room in the case for the 9-volt battery. The battery connector can be made from an old 9-volt battery. Remove the old battery case and cut the leads from the cells to the connector at the top. Solder on new wires, and keep in mind that the small connector is now negative and the large one is positive.

Attach the negative side of the battery to the negative PC board input and the positive side to one side of the momentary switch. Solder another wire from the remaining side of the momentary switch to the positive PC board input. Solder one output from the PC board and one banana plug wire to a spade lug on the back of one banana jack. Do the same for the other jack. Figure 3 shows the inside of the completed unit.

Calibration

With the cover still off, plug the adapter into the DMM and set the range of the meter to 20 volts. Press S1 and adjust R2 for 6 volts DC. Next, place an ammeter across the banana jacks and adjust R3 for a reading of 100 milliamperes. To calibrate the circuit when used in conjunction with test leads, short circuit the leads together and write down the reading; then subtract that reading from any readings you take with the test leads. Now get out that junkbox and start testing those components with unknown values. Ω

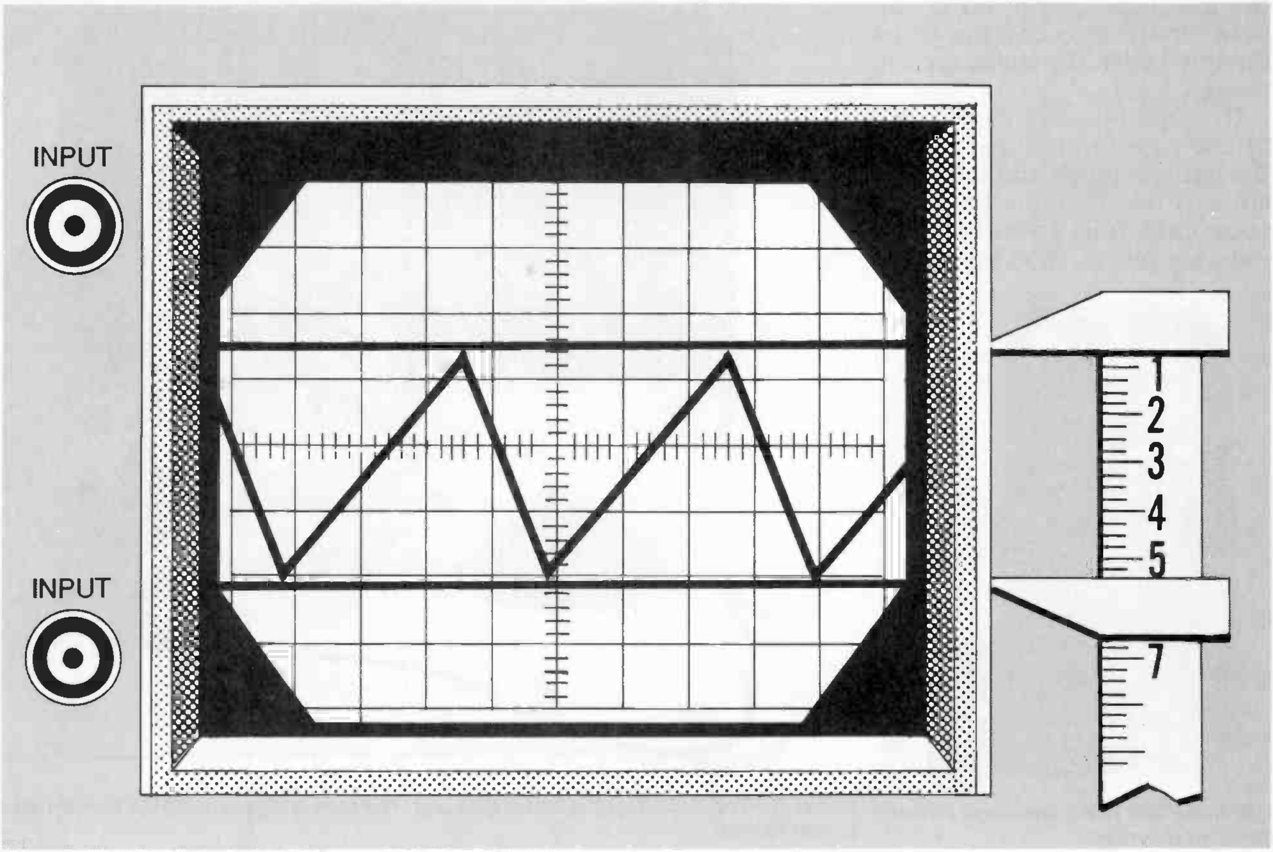

VOLTAGE CURSOR ADAPTER

Measure the voltage values of waveforms accurately and easily on your oscilloscope with the Voltage Cursor Adapter.

SKIP CAMPISI

HAVE YOU EVER HAD TROUBLE MEASURING the voltage value of a waveform on an oscilloscope screen? Are you tired of counting graticule squares and "guesstimating" those values? If so, you need the Voltage Cursor Adapter. It superimposes horizontal cursor lines on the top and bottom of the waveform—a kind of electronic calipers—to permit direct readout of the voltage value. The cursor lines extend across the entire screen.

The lower cursor (think of it as the floor) is the *zero-volt cursor* and the upper cursor (think of it as the ceiling) is the precision DC *reference voltage*. The cursors can be placed on any parts of the waveform that you want to measure, and the voltage can be read directly from a turns counter coupled to a precision potentiometer.

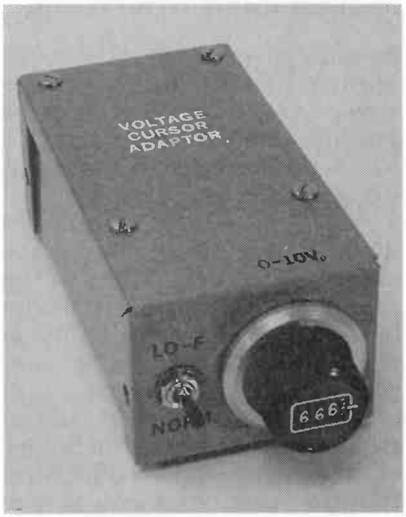

THE CURSOR ADAPTER superimposes horizontal cursor lines on the top and bottom of a waveform.

Figure 1 is the schematic diagram for the voltage cursor adaptor. The required 15 to 25-volt power to the circuit can be supplied either by batteries or a wall-mounted AC to 15-volt DC adapter. The author's prototype has a jack that will accept a plug from either the adapter or a battery pack consisting of two 9-volt batteries taped together.

The MC78L12 voltage regulator (IC1) supplies regulated 12-volts DC to the rest of the circuit. The ICM7555 timer (IC2), a CMOS version of the industry-standard 555, drives the CD4066B, a CMOS bilateral switch (IC3). This drive frequency can either be a normal frequency (NORM) of 100 Hz or a low-frequency (LO-F) of 10 kHz, depending on the setting of switch S1. Set S1 to LO-F for inputs below 500 Hz.

The DC reference voltage supplied to pin 3 of IC3 is set by R3, a 10-turn, 5000-ohm precision potentiometer. The voltage can be read directly from a turns counter dial coupled directly to the potentiometer's wiper. The accuracy of this reading can be 1% or better. Trimmer potentiometer R1 permits the voltage to R3 to be calibrated to precisely 10 volts.

The circuit is calibrated by setting the digital reading on the turns counter of R3 to the full clockwise position and adjusting R1 for a reading of 10 volts at the wiper of R3 with a digital voltmeter.

Bilateral switch IC3 converts the DC reference to a squarewave with exactly the same wiper amplitude. The squarewave output appears on the common pins 4, 9, and 10 of IC3 and coaxial plug PL1.

Building the adapter

The circuitry is simple enough to be built on an approximately 2 × 3-inch stock predrilled perforated board by point-to-point wiring methods. (The prototype circuit was built on a multipurpose board from Radio Shack.) There is nothing critical about component selection or placement. Figure 2 is a guide for positioning components on the circuit board.

The prototype was mounted in a two-part aluminum case that measured approximately 4 by 2 by 1⅝ inches. If you intend to put the circuit in similar project case, use the blank circuit board as a pattern and mark the hole locations on that part of the case with the ends folded up.

It is recommended that both

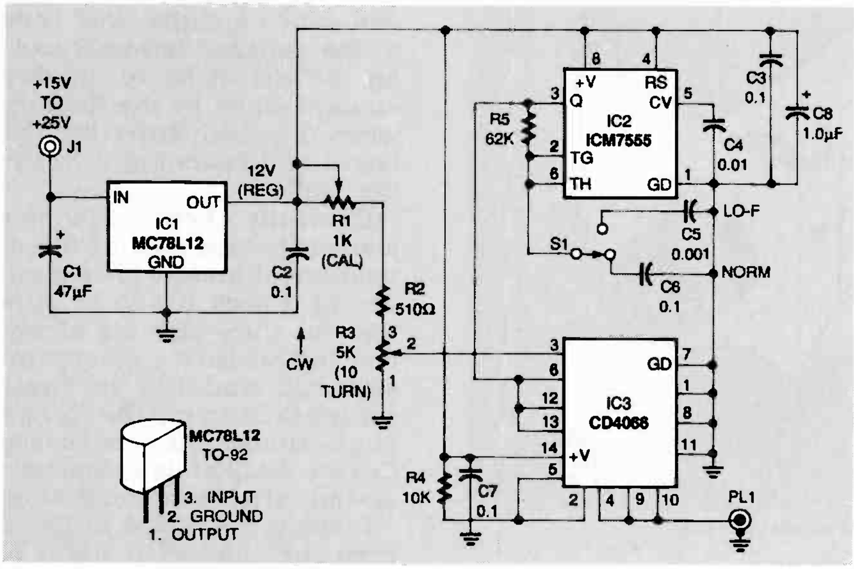

FIG. 1—SCHEMATIC DIAGRAM for the voltage cursor adapter. Power can be supplied by batteries or an AC adapter.

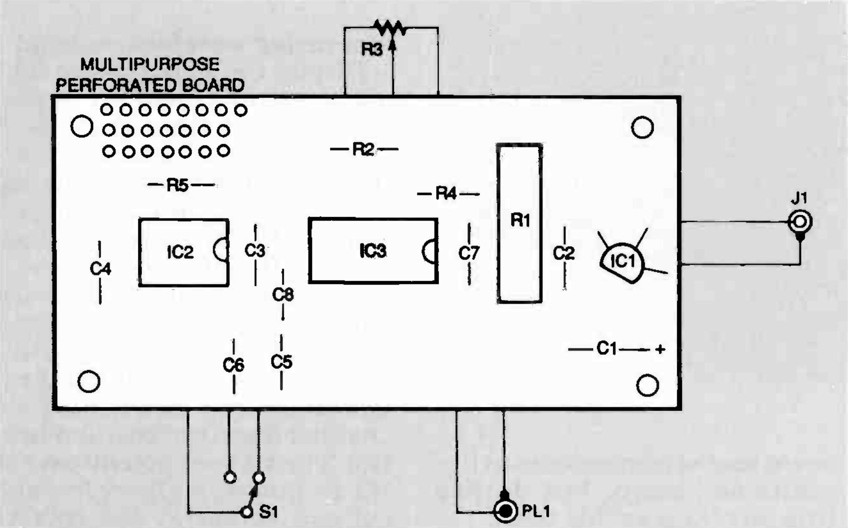

FIG. 2—USE THIS AS A GUIDE for positioning the parts on the perforated construction board.

PARTS LIST

Resistors: all fixed are ¼-watt, 5%

R1—1000-ohm multiturn trimmer potentiometer, Bourns 3005P 101 or equiv.

R2—510 ohms

R3—5000-ohm, ⅞in. precision potentiometer, ten-turn with matching multiturn dial, Clarostat 73JB with a 15-turn dial or equiv. (See text)

R4—10,000 ohms

R5—62,000 ohms

Capacitors

C1—47µF, aluminum electrolytic, 35 VDC

C2, C3, C6, C7—0.1µF, ceramic

C6—0.1µF, polyester

C4—0.01µF, ceramic

C5—0.001µF, polyester

C8—1.0µF, solid tantalum dipped

Semiconductors

IC1—MC78L12 12-volt voltage regulator, Motorola or equiv.

IC2—ICM7555 CMOS timer, Harris or equiv.

IC3—CD4066B CMOS quad-bilateral switch, Harris or equiv.

Other components

S1—SPDT switch, miniature, panel-mounted, 3A

J1—coaxial jack, panel-mounted (to mate with power plug)

PL1—panel-mounted phone plug ⅛-inch dia.

Miscellaneous: multipurpose perforated board 1 27/32 × 2 27/32 (Radio Shack 276-150 or equiv.); project case, aluminum, 4 × 2⅛· 1⅝in. (Radio Shack 270-239 or equiv.); one 8 pin DIP socket and one 14-pin DIP socket; four standoffs with screws, 11/16 inch; two 9-volt batteries with two snap connectors and attached plug or wall-mounted line AC to 15-VDC adapter, 50 mA; insulated hookup wire; solder; screws.

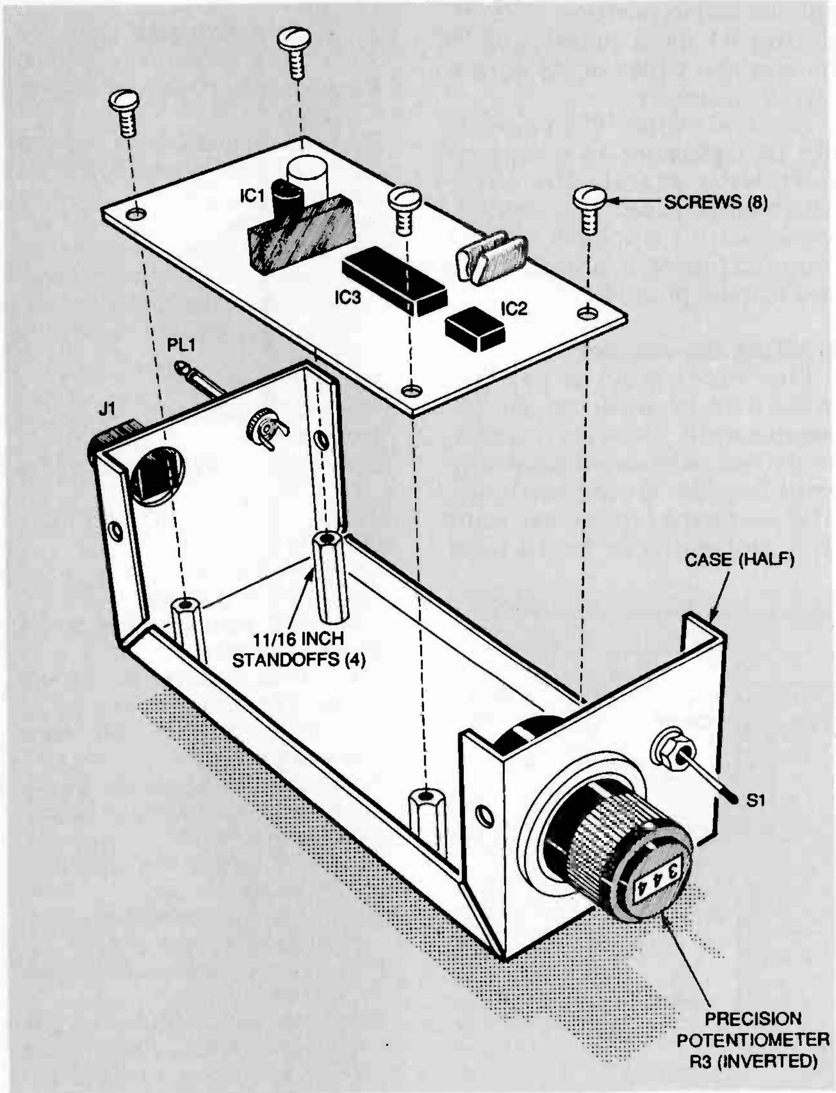

FIG. 3—HERE'S HOW EVERYTHING MOUNTS inside the case.

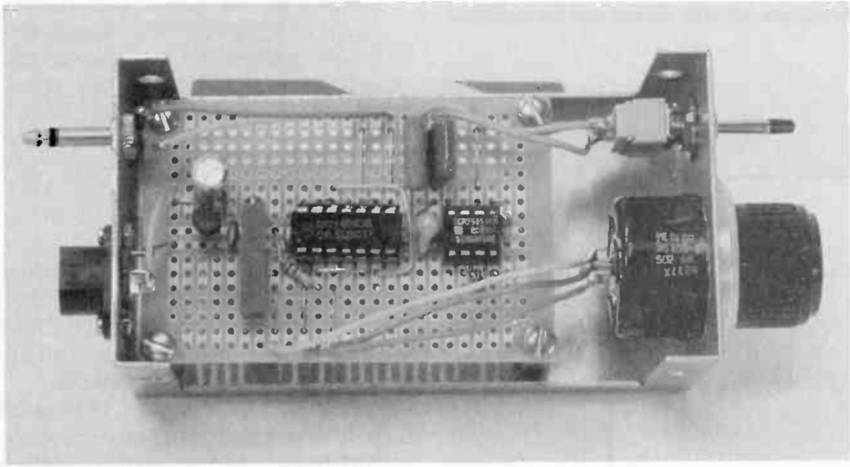

EVERYTHING FITS EASILY inside the case.

IC2 and IC3 be inserted in sockets. Insert and solder an 8-pin socket for IC2 and a 14-pin socket for IC3, as shown in Fig. 2. Insert and solder all other on-board leaded components in the positions shown, but do not trim any leads at this time.

Set the circuit board aside. Refer to Fig. 3, the mechanical assembly diagram, and drill the holes that you marked on the case half for mounting the ¾-inch standoffs. Then drill the holes for mounting jacks J1 and J2, precision potentiometer R3 with counter and switch S1 in the case end surfaces.

Mount the jacks, potentiometer and switch on the case half. Cut 3- to 4-inch lengths of No. 22 AWG stranded, insulated, hookup wires for making all connections to the board-mounted components, as shown in Fig. 2. Note: A lower cost counter can be substituted for the digital readout dial used in the author's prototype.

Solder one end of the wires from the jacks, potentiometer and switch and the other ends to the assigned terminal pads on the circuit board, leaving enough slack in the hookup wires to permit inverting the board and fastening it to the case half.

Carefully check all solder joints to be sure they are free of inadvertent bridges or cold soldering. Check IC1 to be sure that the three pins are identified and soldered correctly. Insert IC2 and IC3 in their sockets, observing the correct pin locations. After the Voltage Cursor Adapter is completely assembled it is ready for testing.

Connect the times 10 probe from the channel B input of your oscilloscope to the output plug (PL1) of the Voltage Cursor Adapter.

Measuring waveform voltage

Display the signal whose voltage you want to measure by plugging a lead from the signal source into the channel A vertical amplifier jack of your oscilloscope. Set the triggering on channel A. The output is variable from 0 to 10 volts.

Set your oscilloscope's attenuators to accommodate the 0 to 10-volt range. Using the CHOPPED/ALTERNATE mode, position in the zero cursor with the channel B vertical position control. Then adjust potentiometer R3 to position the *reference voltage cursor* at the correct level on the signal waveform you want to measure. Ω

MINI LOGIC ANALYZER

This mini analyzer is inexpensive, versatile, and easy to build.

JAMES J. BARBARELLO

AFTER YOU HAVE COMPLETED YOUR latest logic circuit design, you will probably want to build a prototype to test its operation. If the circuit is not complex, a common logic probe is all that's needed to troubleshoot it. However, if the circuit has multiple signals that must be checked for proper time phasing, or if it requires one or more complex driving signals to exercise its operation, then you need a more powerful testing tool.

Professionals use a logic analyzer and function generator to make these tests. Experimenters, hobbyists, and those on a tight budget will want to build the PC Mini Logic Analyzer. It costs only about $30 to build, but when it is linked to an IBM-compatible computer, it is sufficiently capable for most hobby applications. The PC Mini Logic Analyzer provides up to eight driving signals (outputs) and eight inputs. Each output can be programmed with up to a 64-bit pattern. In addition to its logic analyzer function, the unit can serve double duty as a digital integrated circuit (IC) tester.

The analyzer consists of a hardware interface with applications software. The interface buffers the signals that are sent from the computer's parallel port to the circuit to be tested. It also buffers the signals that are returned from the circuit to the computer, and shifts their voltage levels so that they are compatible with the PC's logic levels. The interface obtains its power from the circuit under test, so it will always recognize the proper logic levels: TTL at 5 volts, or CMOS from 3 to 15 volts. The interface can be connected directly to any parallel port on your PC.

The software displays 64 bits of the eight outputs and eight inputs simultaneously. It allows full on-screen programming of the outputs. Scan time (the time to process the 64 bits) can be adjusted from a high of about 100 bits/second on an average PC to a low of one bit every 10 seconds. The slower speeds allow you to single step through a circuit and observe how each output bit affects the prototype circuit's operation.

Best of all, the interface circuit is composed of common components, all available from most electronic component suppliers. Although a PC board layout is provided here, it is not essential for proper operation; the circuit can also be constructed on a solderless breadboard.

The hardware interface will be described first, followed by the software. Then, an actual logic analyzer application will be described, followed by an example of how to use the analyzer as a digital IC tester.

Theory of operation

The analyzer incorporates two basic functional blocks; a transistor buffer/inverter section, and an analog switch section that feeds voltage comparators. The transistor buffer/inverter section is shown in the schematic in Fig. 1.

Transistor Q1 in Fig. 1 is configured as a standard inverting switch. A signal greater than

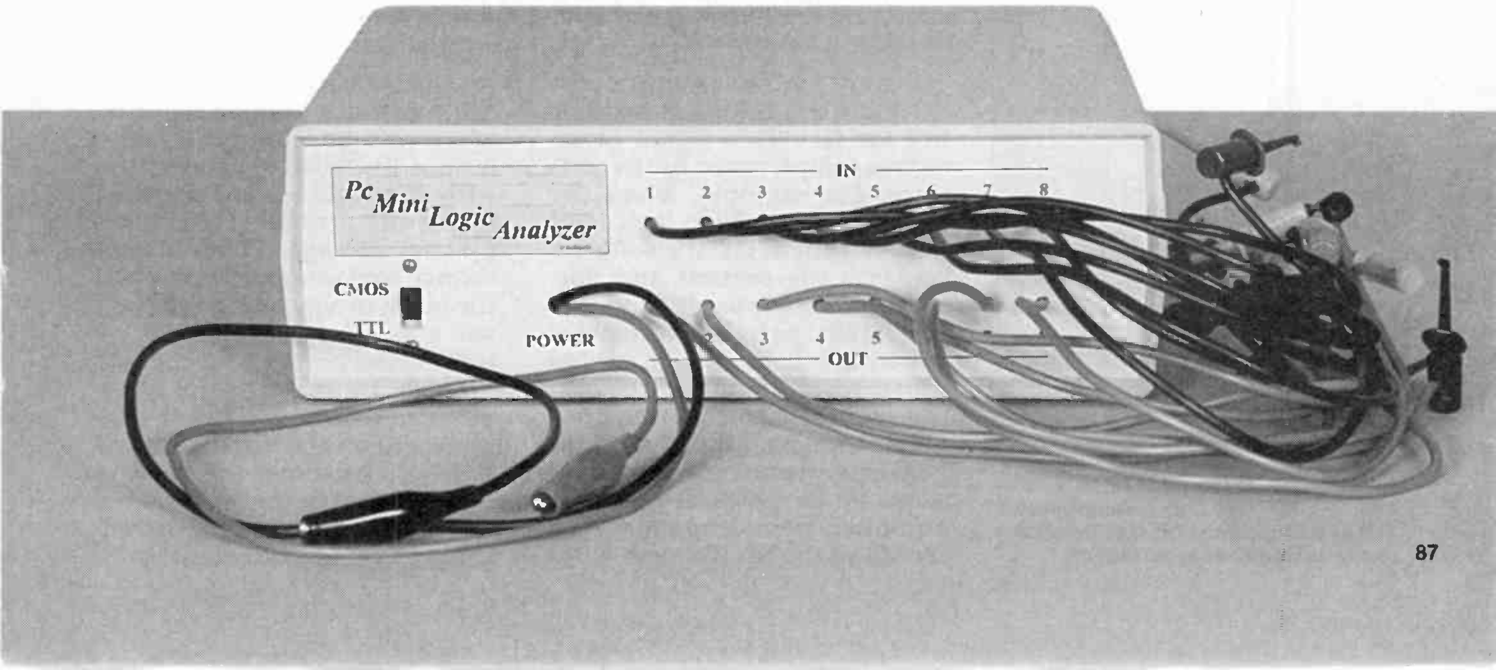

about 0.7 volt from the parallel port (pin 2 for Q1) causes the transistor to conduct, raising the output (O1 for Q1) to about 0.3 volt. When a signal less than 0.7 volt is applied, the transistor does not conduct, and current is provided to any load connected to the output through the collector resistor (R2 for Q1). This signal inversion is compensated by the software. For example, when the user requests a high logic level to be output at O1, the software converts the request and outputs a low logic level at pin 2 of the parallel port. The transistor will invert that low logic level to a logic high, producing the output requested by the user.

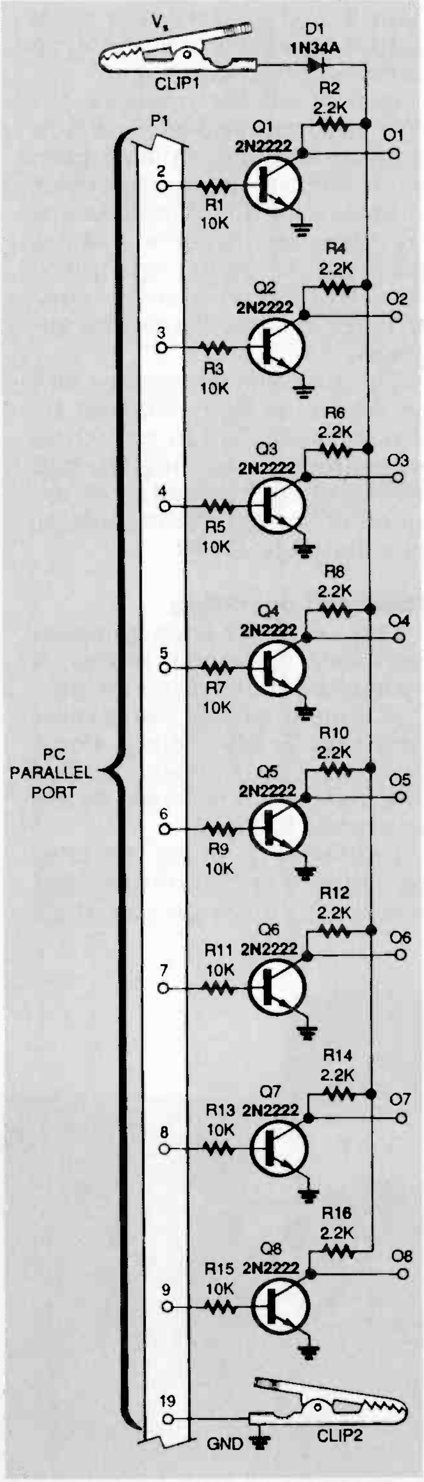
V+
D1
1N34A
CLIP1
R2
2.2K
P1
Q1
2N2222
O1
2
R1
10K
R4
2.2K
Q2
2N2222
O2
3
R3
10K
R6
2.2K
Q3
2N2222
O3
4
R5
10K
R8
2.2K
Q4
2N2222
O4
5
R7
10K
R10
2.2K
Q5
2N2222
O5
PC
PARALLEL
PORT
6
R9
10K
R12
2.2K
Q6
R11 2N2222
10K
O6
7
R14
2.2K
Q7
R13 2N2222
10K
O7
8
R16
2.2K
Q8
R15 2N2222
10K
O8
9
19
GND
CLIP2

FIG. 1—THE ANALYZER incorporates two basic functional blocks. This is the transistor buffer/inverter section.

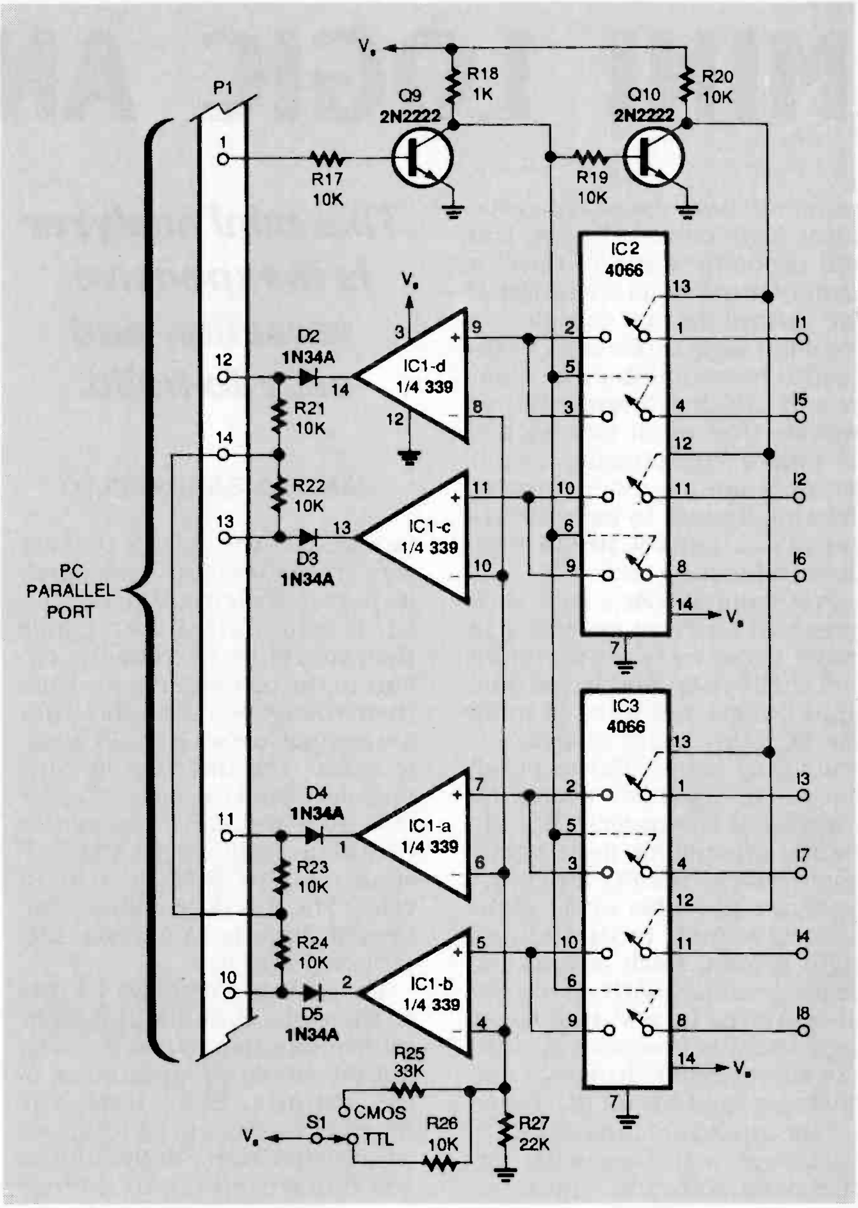
V+
R18
1K
R20
10K
P1
Q9
2N2222
Q10
2N2222
1
R17
10K
R19
10K
IC2
4066
V+
13
3
D2
1N34A
9
2
1
I1
12
IC1-d
1/4 339
14
R21
10K
12
8
5
3
4
I5
14
12
R22
10K
11
10
11
I2
13
13
IC1-c
1/4 339
6
D3
1N34A
PC
PARALLEL
PORT
10
9
8
I6
14
V+
7
IC3
4066
13
D4
1N34A
7
2
1
I3
11
IC1-a
1/4 339
1
5
R23
10K
6
3
4
I7
12
R24
10K
5
10
11
I4
10
2
IC1-b
1/4 339
6
D5
1N34A
4
9
8
I8
R25
33K
14
V+
CMOS
7
S1
V+
TTL
R26
10K
R27
22K

FIG. 2—THE ANALOG SWITCH SECTION feeds the voltage comparators. Two of the analog switches in a CD4066B quad-analog switch IC (IC2-a and IC2-b) feed their outputs to the non-inverting input of IC1-d, one of the four voltage comparators in an LM339 quad comparator IC.

The transistor inverter/buffer performs two necessary functions. First, it acts as a current amplifier, providing more current than the parallel port could provide directly. Second, since the transistor obtains its operating voltage through D1, the output levels will be consistent with the logic levels generated by the circuitry that is supplied by that voltage. (The output voltage level is not dependent on the input driving voltage, which will always be that provided by the parallel port).

Diode D1 protects the interface against reverse power supply voltage. The approximately 0.3 volt lost across the diode is insignificant in the operation of the interface. The transistor inverter/buffer is used eight times

(Q1–Q8) to provide eight outputs.

The second functional block is shown in Fig. 2. Two of the analog switches in a CD4066B CMOS quad-analog switch IC (IC2-a and IC2-b) feed their outputs to the non-inverting input of IC1-d, one of the four voltage comparators in an LM339 quad comparator IC. When the control pin of an analog switch is pulled high, the switch will pass the input signal to its output, also putting from about 50 to 200 ohms in series in the process. When the control pin is brought low, the path between input and output becomes a very high impedance, effectively disconnecting the input from the output.

The signal from pin 1 of the parallel port is inverted twice, once by transistor Q9 and a second time by Q10. Therefore, the control signal for analog switch IC2-b is 180° out of phase with the signal for IC2-a. With a high signal at pin 1 of P1, switch IC2-a will be on, and IC2-b will be off. Conversely, with a low signal at pin 1 of P1, the switch IC2-a will be off, and IC2-b will be on. This allows the two switch outputs to be connected together (sometimes called a "wired OR"), because only one switch will be on at a time.

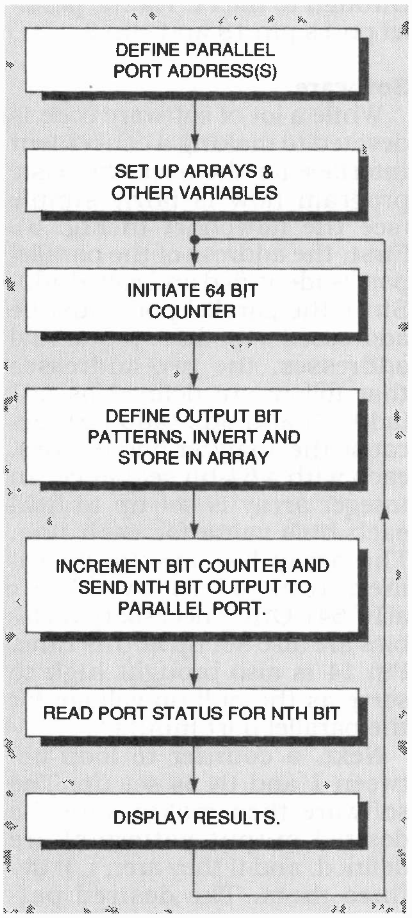
DEFINE PARALLEL PORT ADDRESS(S)
SET UP ARRAYS & OTHER VARIABLES
INITIATE 64 BIT COUNTER
DEFINE OUTPUT BIT PATTERNS. INVERT AND STORE IN ARRAY
INCREMENT BIT COUNTER AND SEND NTH BIT OUTPUT TO PARALLEL PORT.
READ PORT STATUS FOR NTH BIT
DISPLAY RESULTS.

FIG. 3—BASIC PROGRAM FLOW. First, the address of the parallel port that's going to be used is identified and called ad0.

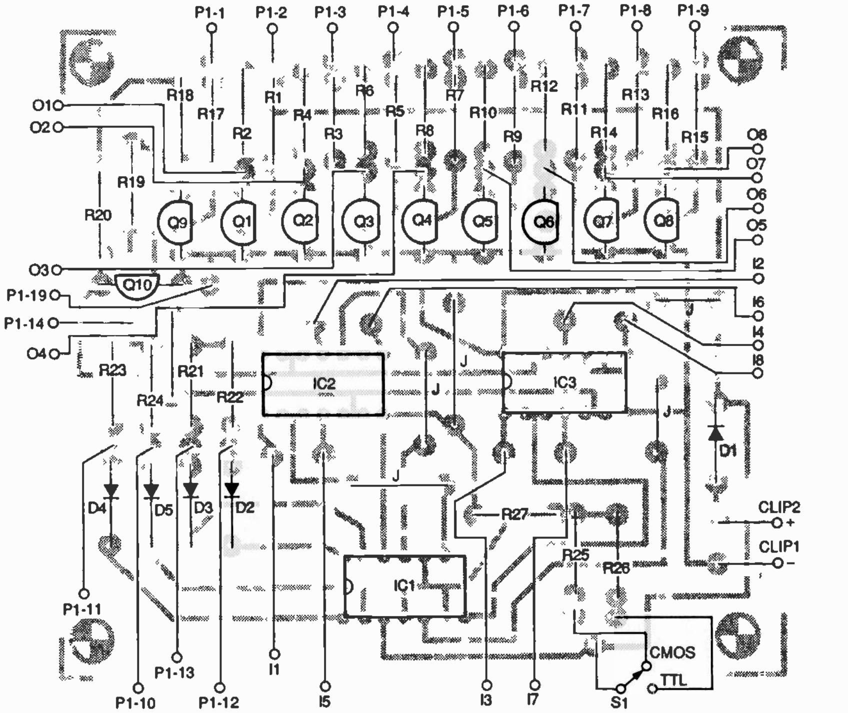
P1-1
P1-2
P1-3
P1-4
P1-5
P1-6
P1-7
P1-8
P1-9
O1
O2
O3
O4
O5
O6
O7
O8
P1-19
P1-14
P1-11
P1-13
P1-10
P1-12
I1
I2
I3
I4
I5
I6
I7
I8
R1
R2
R3
R4
R5
R6
R7
R8
R9
R10
R11
R12
R13
R14
R15
R16
R17
R18
R19
R20
R21
R22
R23
R24
R25
R26
R27
Q1
Q2
Q3
Q4
Q5
Q6
Q7
Q8
Q9
Q10
IC1
IC2
IC3
D1
D2
D3
D4
D5
J
CLIP2
+
CLIP1
−
CMOS
TTL
S1

FIG. 4—PARTS-PLACEMENT DIAGRAM. The circuit layout is not critical, so you can use either the PC board an experimenter's breadboard.

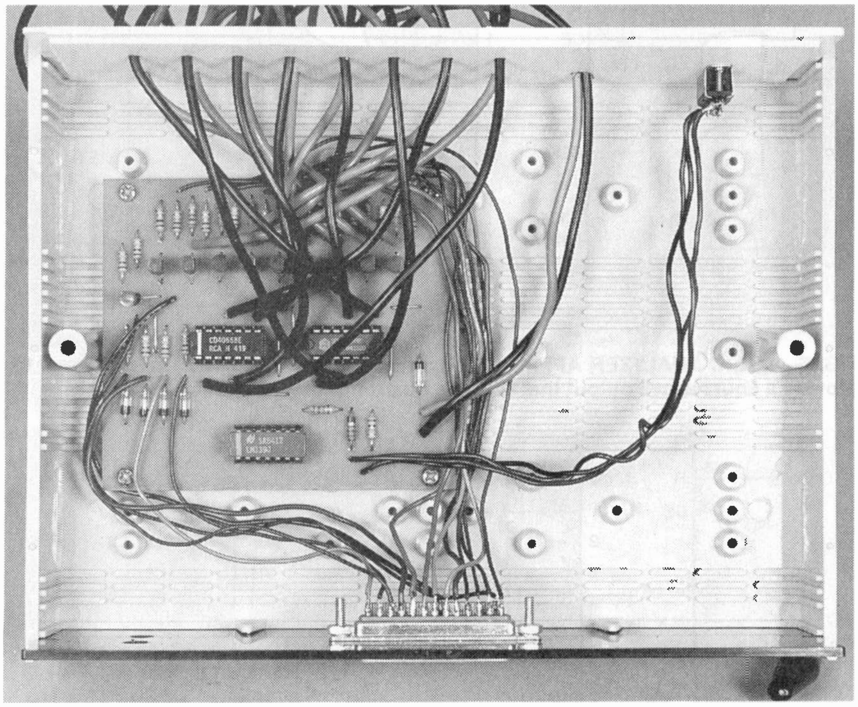

FIG. 5—AUTHOR'S PROTOTYPE. Use micro-clips for the 16 input/output lines, and insulated mini alligator clips for the power and ground leads. Route the wires through holes in the enclosure and attach the connectors.

Comparator IC1-d compares the signal at its positive input to the reference voltage at its negative input. When the signal is greater than the reference, the

TABLE 1

Connection	Output	Voltage
None	O1	Vbattery–Vdiode
None	O2	Vbattery–Vdiode
None	O3	Vbattery–Vdiode
None	O4	Vbattery–Vdiode
None	O5	Vbattery–Vdiode
None	O6	Vbattery–Vdiode
None	O7	Vbattery–Vdiode
None	O8	Vbattery–Vdiode
O1 to P3	O2	0 to 0.5V (approx)
O1 to P4	O3	0 to 0.5V (approx)
O1 to P5	O4	0 to 0.5V (approx)
O1 to P6	O5	0 to 0.5V (approx)
O1 to P7	O6	0 to 0.5V (approx)
O1 to P8	O7	0 to 0.5V (approx)
O1 to P9	O8	0 to 0.5V (approx)
O2 to P2	O1	0 to 0.5V (approx)

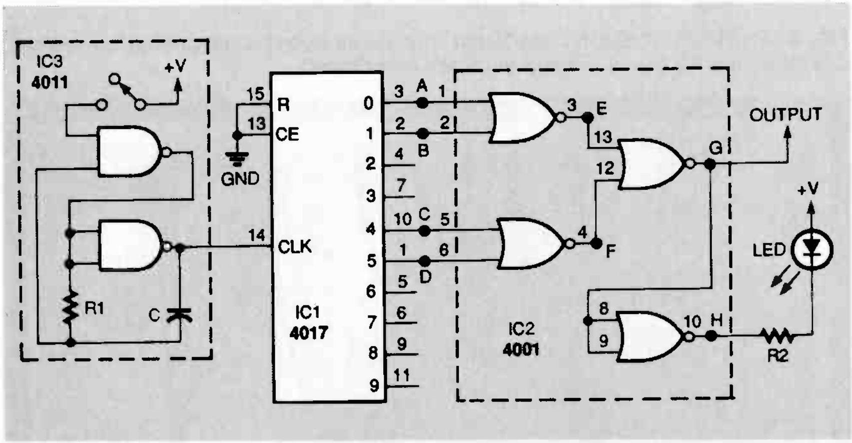

FIG. 6—LOGIC ANALYZER APPLICATION. This self-clocking circuit is supposed to provide a squarewave output that has one quarter of the input frequency.

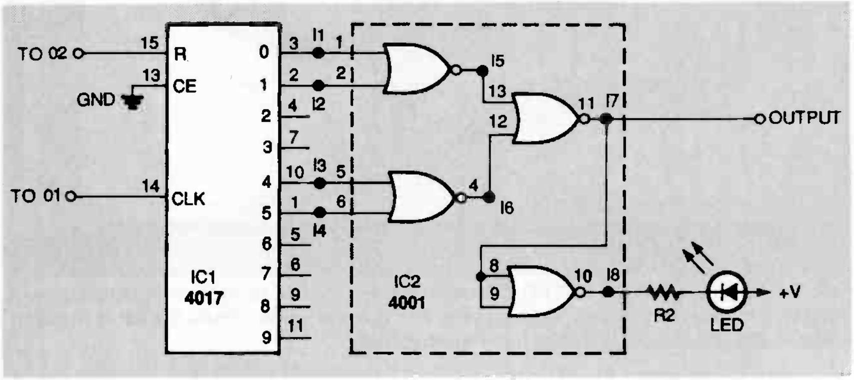

FIG. 7—MONITORING ALL POINTS in the circuit will reveal any problems quickly. Here's the test circuit with the CD4011B replaced by the analyzer.

comparator's output goes high, and blocking diode D2 is reverse-biased. The high output at pin 14 of the parallel port pulls the parallel port pin 13 input high. A low signal from the comparator output allows the parallel port pin to sink voltage from pin 14 into the comparator's output through the diode. This pulls parallel port pin 13 low when a low signal is provided to the comparator. This arrangement allows signals greater than the 5-volt level of the parallel port to be processed. The analog switch/comparator block is used four times to accept eight inputs and provide them to the four comparators for input into the PC.

The reference voltages for the comparators are created in the voltage divider consisting of R25, R26, and R27. When switch S1 is in its TTL position, the voltage across R27 (provided to the comparators) will be about 2 volts with a 5-volt power supply. With S1 in its CMOS position, the voltage across R27 will be about 70% of the power supply voltage. Those voltage levels correspond to the lower limits of valid high logic levels for those logic families. A common ground reference is provided from the power supply through to the PC via the parallel port's pin 19 and clip 2.

Software

While a lot of software code is devoted to making a convenient interface for the user, the basic program flow is fairly simple (see the flowchart in Fig. 3). First, the address of the parallel port is identified and called ad0. Since the parallel port must be addressed with three sequential addresses, the two addresses that follow are defined as ad1 (ad0 + 1) and ad2 (ad0 + 2). Because the analyzer has 16 lines, each with a 64-bit sequence, an integer array is set up to hold each bit's value for each line. The array has maximum dimensions of 16 by 64, or a(16,64). Other necessary variables are also set up at this time. Pin 14 is also brought high to serve as the pull-up voltage for the parallel port inputs.

Next, a counter to loop between 1 and 64 is set up. The software then makes sure the desired output pattern(s) are defined, and if they aren't, it defines them. The desired patterns are inverted (to compensate for the hardware inversion in the interface transistors) and stored in the integer array. Next,

LISTING 1

```
REM**** MINITEST.BAS
REM**** QUICK CHECKOUT OF MiniAnal
REM**** V940708    (c) 1994, JJ Barbarello
CLEAR : COLOR 7, 0: CLS
DEFINT A-X: DEFSTR Y-Z: DIM o(8), i(8): DEF SEG = 64
LOCATE 1, 15
PRINT "QUICK CHECKOUT OF Pc MINI LOGIC ANALYZER INTERFACE" LOCATE 2, 1: PRINT
STRING$(79, 220)
LOCATE 5, 15: INPUT "PARALLEL PORT ADDRESS (Enter for default): "; ad0
IF ad0 = 0 THEN ad0 = PEEK(8) + 256 * PEEK(9)
ad1 = ad0 + 1: ad2 = ad1 + 1:  OUT ad2, 0: PRINT : PRINT
FOR i = 0 TO 7: a = 2 ^ i
OUT ad0, a XOR 255: GOSUB status1
IF 0 <> o(i + 1) THEN
     PRINT "Problem with Output"; i + 1: "(Should be low)"
     x = 1
END IF
IF 1 <> i(i + 1) THEN
     PRINT "Problem with Input"; i + 1; "(Should be high)"
     x = 1
END IF
OUT ad0, a: GOSUB status1
IF 1 <> o(i + 1) THEN
     PRINT "Problem with Output"; i + 1; "(Should be high)"
     x = 1
END IF
IF 0 <> i(i + 1) THEN
     PRINT "Problem with Input"; i + 1; "(Should be low)"
     x = 1
END IF
NEXT i
IF x = 0 THEN LOCATE 10, 32: PRINT "INTERFACE TESTS OK"
END
REM******** GET STATUS OF PRINTER PORT PINS *******
status1:
o(1) = INP(ad0) AND 1
o(2) = (INP(ad0) AND 2) / 2
o(3) = (INP(ad0) AND 4) / 4
o(4) = (INP(ad0) AND 8) / 8
o(5) = (INP(ad0) AND 16) / 16
o(6) = (INP(ad0) AND 32) / 32
o(7) = (INP(ad0) AND 64) / 64
o(8) = (INP(ad0) AND 128) / 128
OUT ad2, 0: REM: Pin 1 High
i(1) = (INP(ad1) AND 32) / 32
i(2) = (INP(ad1) AND 16) / 16
i(3) = -(((INP(ad1) AND 128) / 128) = 0)
i(4) = (INP(ad1) AND 64) / 64
OUT ad2, 1: REM: Pin 1 low
i(5) = (INP(ad1) AND 32) / 32
i(6) = (INP(ad1) AND 16) / 16
i(7) = -(((INP(ad1) AND 128) / 128) = 0)
i(8) = (INP(ad1) AND 64) / 64
RETURN
```

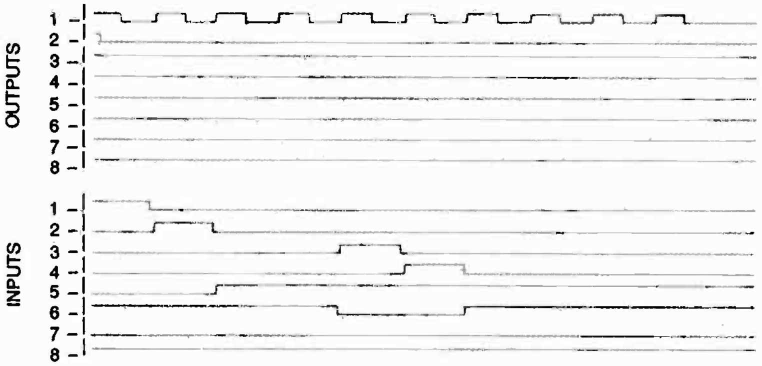

FIG. 8—RESULTING BIT PATTERN. The immediate problem is that the LED does not light, verified in the bit pattern showing that inputs I7 and I8 are not switching.

the bit counter is incremented and the next bit is sent out to the parallel port. The status of the port is then read and the results are displayed. The software then loops back to the bit counter, and the input/output process continues. There is more to the actual program code, but the basic analyzer operation is captured in just those few steps.

Construction

The circuit layout is not critical, so you can use either the PC board and parts-placement diagram shown in Fig. 4, or follow the schematic diagram and build the circuit on an experimenter's breadboard. For either assembly method, IC sockets are recommended.

Once the circuit is constructed, and before you begin the final wiring, decide on the case style (if any) for enclosing the analyzer. Because 16 wires and test clips extend from the circuit board, it is wise to choose a wire-coding arrangement that will help you easily recognize each input and output lead.

Use No. 20 stranded hookup wire for the 16 input/output leads, connecting one end to the 16 points identified as O1–O8 and I1–I8. Again using No. 20 stranded wire, connect a red lead to Clip 1 and a black lead to Clip 2, and also connect switch S1.

Decide if you want to have the parallel port cable connected to the board directly, or if you want to add a DB-25 connector. For the integral cable option, obtain 15 four-foot lengths of No. 22 or 24 stranded wire, or a four-foot length of 15-conductor cable. Then use those leads to connect the appropriate points from the circuit board to the pins of a male DB-25 plug. If you use a connector, make the appropriate connections from the circuit board to a female DB-25 socket mounted on the back of the case with wire cut to the size of the enclosure.

The last step is to attach connection devices to the 16 input/output lines and power leads. Use micro-clips for the 16 input/output lines, and insulated mini alligator clips for the power and ground leads. Route the wires through holes in the enclosure and attach the connectors. Complete the assembly

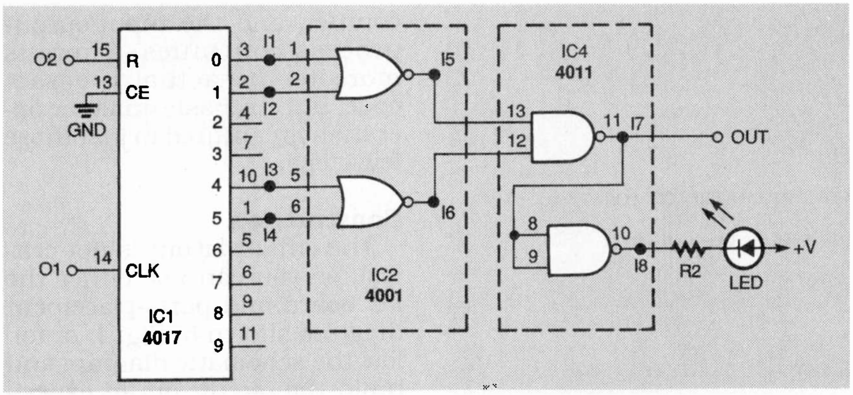

FIG. 9—A NAND FUNCTION will produce a low logic level when the two inputs are high, and produce a high logic level otherwise. Here's the revised circuit.

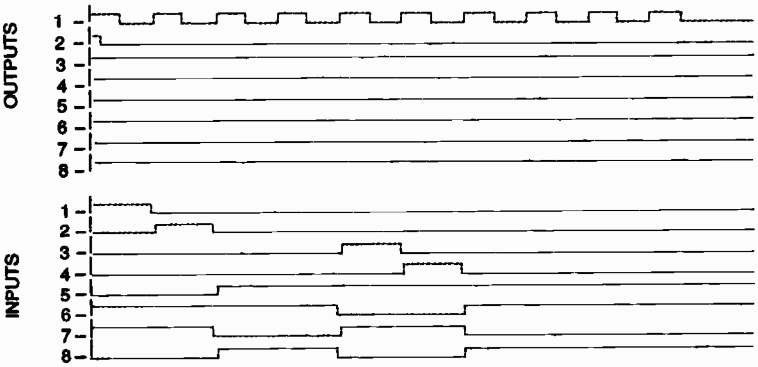

FIG. 10—RESULTANT BIT PATTERN of the circuit in Fig. 9.

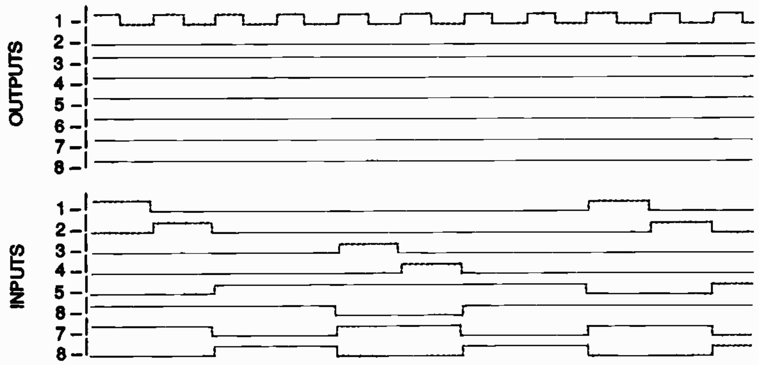

FIG. 11—THE I7 AND I8 WAVEFORMS would both be symmetrical if they ended after the eighth clock pulse. Routing the ninth output of IC1 to the reset pin will do this.

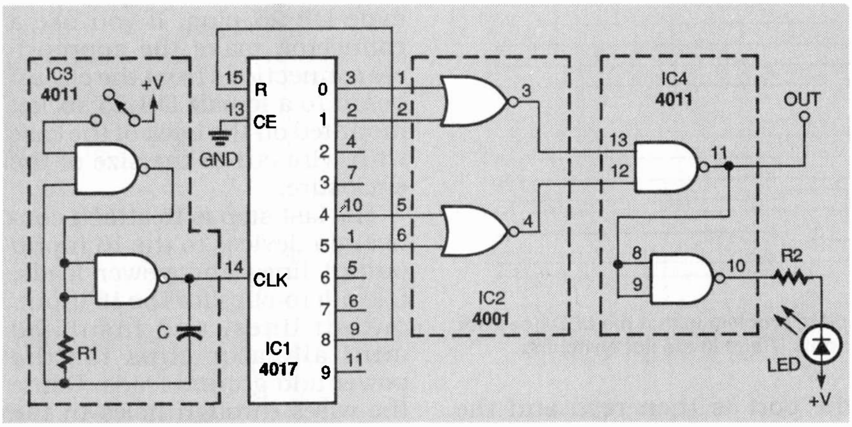

FIG. 12—THIS FINAL CIRCUIT configuration meets all of the original requirements; the output is the input divided by 4, it is symmetrical, and the LED lights.

by mounting the circuit board in the enclosure and mounting S1. Figure 5 shows the author's prototype unit.

Checkout

To test the analyzer, you'll need a 9-volt battery, a voltmeter, and a few jumpers. First make sure that there is no continuity between the board and the eight output leads O1–O8. Then connect the power terminals to the 9-volt battery (red to +, black to −). The voltage between each of the output leads and ground should be the battery voltage less the voltage drop across diode D1. For example, if your battery terminal voltage is 8.9 volts, the voltage at O1 should be about 8.6 volts. The actual voltage is not critical, because the drop across D1 and the individual transistors will vary slightly depending on the actual devices installed.

Next, connect O1 to P1 pin 3; O2 should now read between 0 and about 0.5 volt. Connect O1 to P1 pins 4 through 9 in turn, reading the outputs (O3 through O8 respectively). Finally, connect O2 to P1 pin 2 and read the output at O1. This entire procedure is summarized in Table 1.

To check the interface input circuits and parallel port connections, use the program MATEST.BAS shown in Listing 1. (All of the software for this analyzer project will be made available on the Gernsback BBS, 516-293-2283, v.32, v.42bis, as a file called MINI-ANAL.ZIP.) Extract the program MATEST.EXE from the zipped file or type it in by hand and run it from QBasic. Connect each numbered output to its corresponding input (O1 to I1, O2 to I2, etc.). Connect the power leads to a 9-volt battery, and connect the analyzer output to your computer's parallel port. Place S1 in the CMOS position.

The program will use the outputs to create a logic 1 and then logic 0 for each of the eight inputs. It will read the inputs to determine if they were processed correctly. When you execute the program, it will ask you for the parallel port address.

Press ENTER if you're using LPT1 at the standard address of 888. Otherwise, type in the decimal address of the parallel port and press ENTER. The program will then test all inputs and outputs, and give you the results. You will either get a message "Interface Tests OK," or one or more messages indicating which inputs or outputs did not perform as they should.

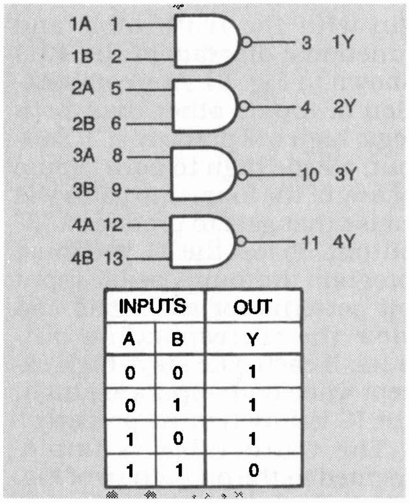

INPUTS		OUT
A	B	Y
0	0	1
0	1	1
1	0	1
1	1	0

FIG. 13—TO USE THE ANALYZER to check a CD4011B NAND gate, begin with the truth table and functional diagram.

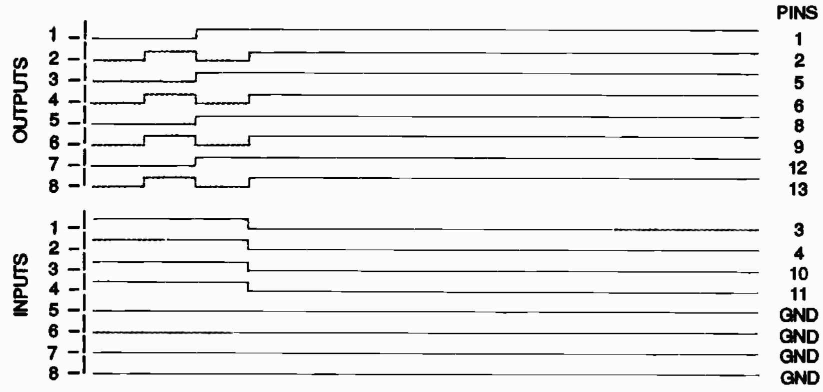

FIG. 14—THE TRUTH TABLE of the CD4011B is implemented in this bit pattern.

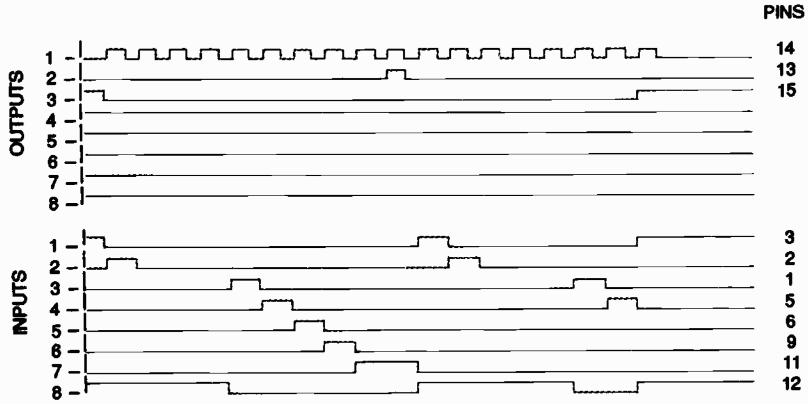

FIG. 15—BIT SCAN for the CD4017B CMOS decade counter/divider.

Logic analyzer

To use the unit as a logic analyzer, begin with the following sample problem: Assume Fig. 6 is a self-clocking circuit that you designed to provide a squarewave output that is one quarter of the input frequency. The clock can be disabled with an external switch, and the LED should turn on with the rising edge of the clock. If wired as shown, the circuit does not operate properly. Most notably, the LED doesn't flash.

Before testing the circuit, IC3, a simple squarewave generator, must be replaced with one of the analyzer's outputs. If IC3 were left in the circuit, there would be no way to synchronize it with the analyzer, and the display would drift, resembling an oscilloscope trace without proper sync. The IC3 oscillator is therefore simulated by programming an analyzer output to produce a squarewave.

Although IC1 is a divide-by-ten counter, the analyzer has a 64-bit pattern that is not evenly divisible by ten. Again to avoid pattern drifting, a second output must be programmed to reset IC1 after ten clock pulses.

Finally, the correct monitoring points must be determined. For the best understanding of circuit operation, all possible points (A through H) should be monitored. The resulting test circuit is shown in Fig. 7.

The bit pattern obtained is shown in Fig. 8. The fault with the circuit was that the LED would not light. The problem is verified in the bit pattern showing that inputs I7 and I8 are not switching. However, the outputs of IC1 are performing as expected, and the outputs of the first two NOR gates (I5 and I6) are also valid. Re-examining the circuit with the help of the analyzer reveals the fault: At least one of the inputs (I5 and I6) to the third NOR gate (I7) is always logic high. That makes the output always logic low. The circuit really needs a NAND function to produce a low logic level when the two inputs are high, and to produce a high logic level at other times. The solution is to change the circuit to incorporate a NAND gate. The revised circuit is shown in Fig. 9, and the resultant bit pattern is shown in Fig. 10.

Replacing the NOR gate with a NAND gate has provided the desired signals at I7 and I8. (The fourth NOR gate was also replaced with a NAND gate arbitrarily—either gate can serve as an effective inverter/current sink for the LED). By looking at the new bit pattern, you might notice another problem. The circuit is dividing by four, but it is not producing a symmetrical squarewave. The pattern is symmetrical through the eighth positive clock pulse (O1), but then stays low for the remaining two clock pulses. In viewing the bit pattern this will be immediately obvious, and it is much better than attempting to watch the LED while determining what's wrong.

PARTS LIST

All resistors are ¼-watt, 5%.

R1, R3, R5, R7, R9, R11, R13, R15, R17, R19–R24, R26—10,000 ohms
R2, R4, R6, R8, R10, R12, R14, R16—2200 ohms
R18—1000 ohms
R25—33,000 ohms
R27—22,000 ohms

Semiconductors

D1–D5—1N34A Germanium diode
Q1–Q10—2N2222A NPN transistor
IC1, IC2—CD4066 CMOS Quad Analog Switch
IC3—LM339A quad comparator

Other components

S1—SPDT switch
P1—Female DB-25 connector and male-to-male DB-25 cable (or use integral cable with male DB-25 on one end, see text)
Clip 1—Red miniature alligator clip
Clip 2—Black miniature alligator clip

Miscellaneous: Sixteen micro clips, three 14-pin IC sockets, 4-foot length of 15-conductor cable or 60 feet of No. 22 stranded wire (see text), No. 20 stranded wire, PC board, suitable case.

Note: The following items are available from JJ Barbarello, 817 Tennent Road, Manalapan, NJ 07726:
- **PC Board (MA-PC)—$15.00**
- **Disk containing both source code and executable file of the mini analyzer program, which includes additional capability for saving and retrieving data, on-line help, on-line directory of files, and more (MA-S)—$12.00**

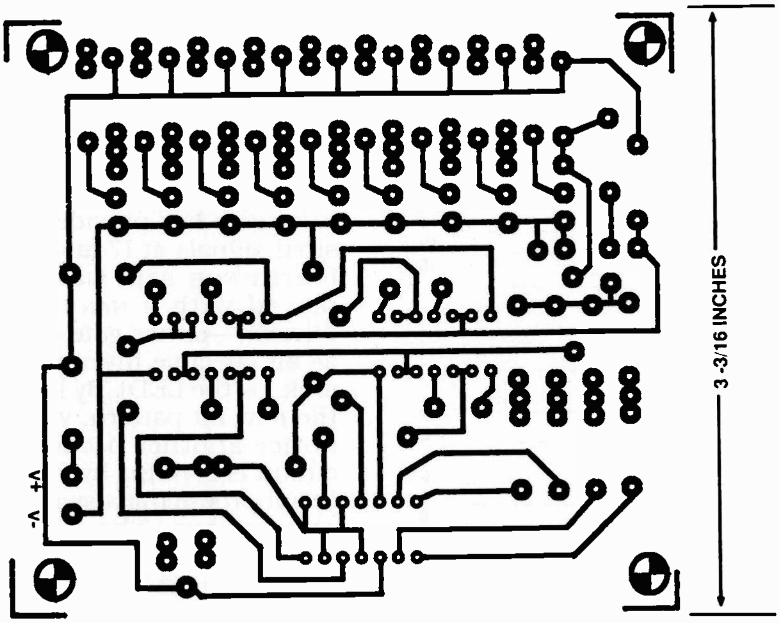

FOIL PATTERN FOR THE MINI ANALYZER.

A look at the I7 and I8 waveforms shows how to solve the problem. They would both be symmetrical if they ended after the eighth clock pulse. This can be accomplished by routing the ninth output of IC1 (called output 8 because the first one is called output 0) to the reset pin. This will create a divide-by-8 counter instead of a divide-by-10, and it will generate the bit pattern in Fig. 11. Notice that the O2 reset pulse has been removed—it is no longer needed because a hardware reset has been incorporated. Because the circuit divides by 8, and the 64-bit pattern is divisible by 8, the bit pattern is stable, or synchronized. The final circuit configuration is shown in Fig. 12. It meets all of the original requirements in that the output is the input divided by 4, it is symmetrical, and the LED lights on the rising edge of the clock.

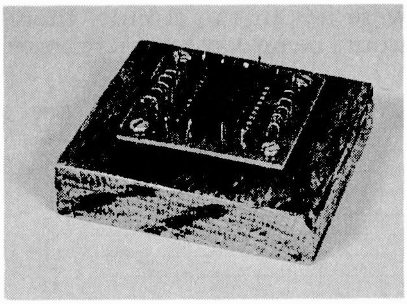

FIG. 16—THIS SIMPLE JIG makes connections to the IC pins easy.

IC checker

This example will expand on the last example. The analyzer will check for proper operation of the CD4011B NAND gate. Begin with the truth table and functional diagram of the 4011 shown in Fig. 13. Any combination of inputs other than both logic high will produce a "1" output; a logic high to both inputs of any of the four NAND gates will cause that gate to produce a "0" output. To test the IC, you must program the four possible input bit patterns for each gate and view the corresponding outputs. If each gate stays high except when both inputs are high, the IC is functioning properly.

The truth table is implemented in the bit pattern of Fig. 14. Note that the pin connections are indicated to the right of the traces. Inputs 5 through 8 are grounded to avoid stray pickup. Ground any unused inputs to avoid stray signal pickup through the high off-impedance inputs of the CD4066 switches. Alternatley, you can attach a pull-down (terminating) resistor to each input. Any resistor value of 10,000 ohms or more will work fine and not load the source circuitry.

Instead of a single bit, a group of bits (five in this case) will be used for each state. Each set of inputs to the 4011 (O1 and O2, O3 and O4, O5 and O6, and O7 and O8) have the same pattern. The first five bits of each set are low. Then the second five bits are low on the first input and high on the second. Then the pattern flips, with the next five bits high on the first input and low on the second. Finally, the bit pattern shifts to both inputs high. In each instance, the expected IC output pattern (I1, I2, I3 and I4) show a high input until the inputs shift to both high. Then the IC output goes low.

A bit scan for the CD4017B CMOS decade counter/divider is shown in Fig. 15. The input bit pattern from a reference book was used directly, except that IC outputs 2, 3, and 4 were not sensed (because the IC has 11 outputs, and the analyzer has

(Continued on page 109)

MAGNETIC FIELD METER

Determine your exposure to line-frequency magnetic-fields with our easy-to-build portable ELF gaussmeter.

REINHARD METZ

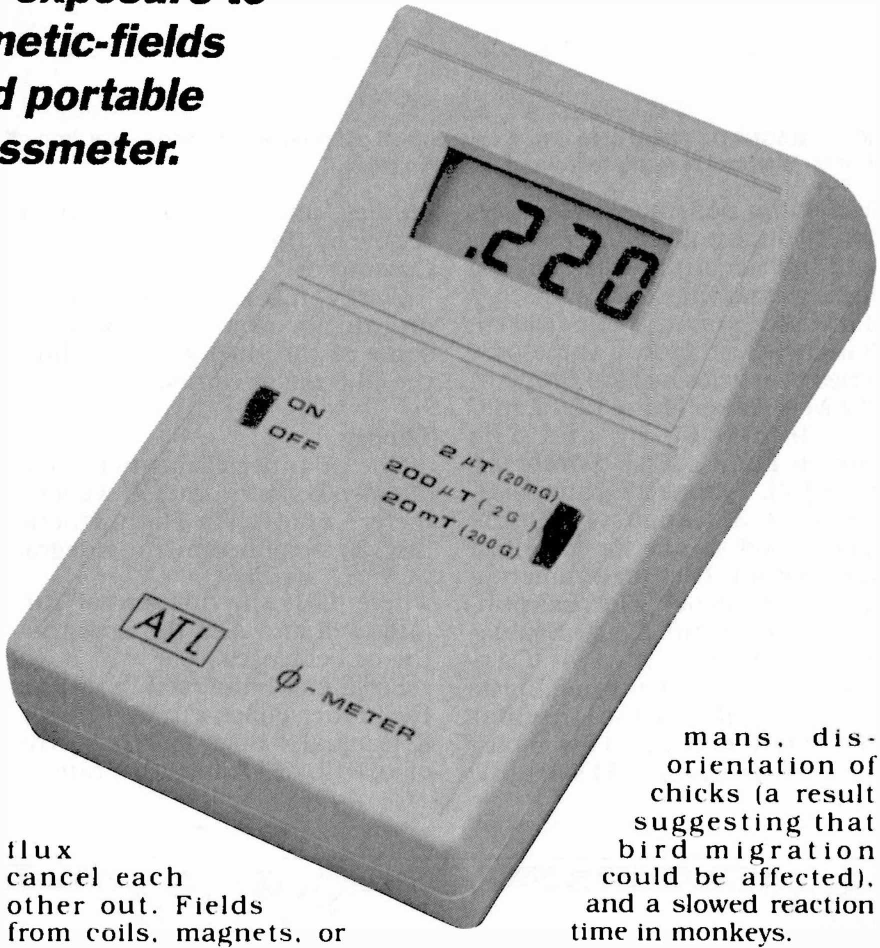

IF YOU ARE ONE IN A GROWING number of people who are concerned about the potentially harmful effects of exposure to magnetic fields, you will be interested in this important construction project. Now you can build your own gaussmeter, and determine the magnitude of magnetic flux densities in and around your home. Our hand-held, battery-operated magnetic-field meter is sensitive from 0.1 microtesla (µT) to 20 milliteslas (mT), and has a frequency range from 50 Hz to 20 kHz.

Why all the worry?

Magnetic fields are all around us. They occurr from the generation, distribution, and use of 50 and 60-Hz electricity, electronic equipment, and even from Earth's magnetic field, which has always been present throughout Man's evolution. Man has been "tuned" into Earth's steady magnetic field of about 30 µT (at sea level) for millions of years. Some sources of excessive magnetic fields that have caused the greatest public concern include power-distribution substations, power lines, CRT terminals, and use of appliances.

Magnetic field intensities can vary greatly, depending on the exposure source and the distance from that source. The rate at which the field intensity falls off with distance can vary from one source to another, depending on how well the current-carrying lines are balanced, or how well the opposing lines of magnetic flux cancel each other out. Fields from coils, magnets, or transformers drop off rapidly with distance by a factor of $1/r^3$. In power lines, if currents flow in opposite directions, the drop-off is $1/r^2$ because of partial field canceling. When unbalanced current exists, the field intensity falls off less rapidly as $1/r$.

Figure 1-*a*, -*b*, and -*c* show drop-off rates of $1/r$, $1/r^2$, and $1/r^3$, respectively. Figure 2 lists some of the many sources of magnetic field exposure, with their range of intensities and drop-off rates.

Although a great deal of controversy still prevails, many people in the scientific community believe that exposure to magnetic fields of extremely-low frequency (ELF fields of 1–100 Hz) may pose a risk to human health. Some disturbing findings of exposure to ELF fields include a significant increase in serum triglycerides (a possible stress indicator) in humans, disorientation of chicks (a result suggesting that bird migration could be affected), and a slowed reaction time in monkeys.

A study conducted by epidemiologist Nancy Wertheimer and physicist Ed Leeper, found that exposures to magnetic fields as small as 0.25 µT correlated with a rise in cancer rates. In the study, the researchers examined wiring and transformers in the neighborhood of birth homes of children who had died of leukemia between 1950 and 1975, along with those of a control group of children who did not have the disease. The results of their studies were published in *The American Journal of Epidemiology* (March, 1979). Some experts argue that other factors, such as pollution and exposure to chemical carcinogens, make interpretation of those findings very difficult.

Standards for acceptable exposure to ELF fields are emerging, as are results of studies

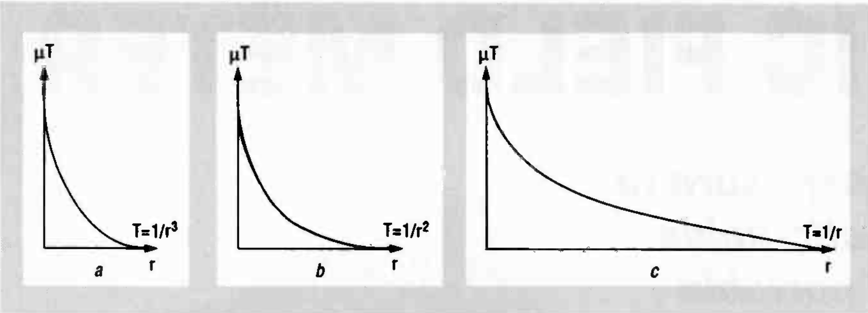

FIG. 1—MAGNETIC FIELD drop-offs. A fast drop-off of $1/r^3$ (*a*), $1/r^2$ (*b*), and a slow drop-off of $1/r$ (*c*) is typical of many sources of magnetic fields.

describing possible hazard levels. If you are more interested a detailed account of scientific findings and the political history of the effects of magnetic-field radiation, we suggest a three-part series of articles by Paul Brodeur, *The New Yorker* (June 12, 19, and 26, 1989). "60-Hz and The Human Body", *IEEE Spectrum*, Parts 1–3, Volume 27, Number 9, pages 22–35 (August, 1990) is also a good source for technical information. The Environmental Protection Agency (EPA) has published a report titled "The Evaluation of the Potential Carcinogenicity of Electromagnetic Fields", publication number EPA/600/6-90/005B. This report contains analyses of 64 scientific studies, and is currently under review by the Scientific Advisory Committee.

Well, that's enough background for now. Let's examine some of the theory behind how the ELF meter works.

Theory

The quantity of magnetic flux density, **B**, is in units of webers/meter², or tesla (T). The magnetic flux, ϕ, is defined by the integral

$$\phi = \int \mathbf{B} \cdot d\mathbf{s} = B \times A$$

where **ds** is the differential surface area and A is the area that the coil encloses.

For a coil immersed in a field, the induced open-circuit voltage, E, is equal to the number of turns of a coil, N, times the rate of change of flux through it.

$$E = |N \times d\phi/dt|$$

Note that the value of $N \times d\phi/dt$ is actually negative with respect to the induced voltage value, but for our purposes we will just consider the magnitude of the product. The direction of the induced current is such that its own magnetic field opposes the changes in flux responsible for producing it.

If we substitute for ϕ we get

$$E = N \times A(d\mathbf{B}/dt)$$

If the magnetic field of a sine wave is $\mathbf{B} = a(\sin \omega t)$, a is the amplitude in teslas and ω is the angular velocity ($2\pi f$), then

$$d\mathbf{B} = a\omega(\cos \omega t)dt, \text{ and}$$

$$E = N \times Aa\omega(\cos \omega t)$$

Since $\cos \omega t$ varies from +1 to −1, the peak magnetic field is defined as

$$E = NAa\,\omega$$

For a frequency of 60 Hz, ω equals

$$2\pi \times 60 = 377$$

For a coil size of 3½" × 3", the area is .0068 m², and therefore

$$E = 2.56\ N \times a$$

For the 12-turn pickup coil that we'll use, the sensitivity is 30 µV per µT.

Circuit description

The meter's 12-turn field pickup is integrated into the unit's

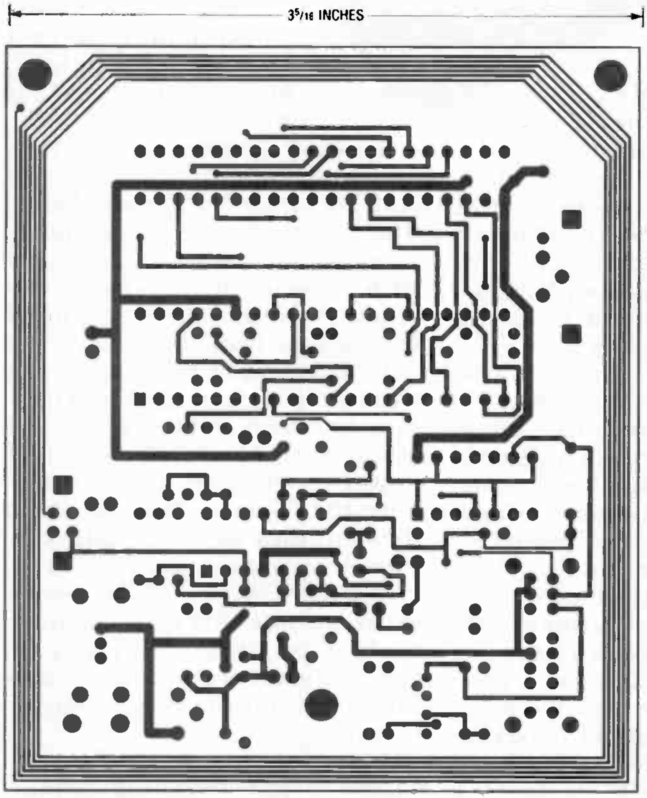

COMPONENT SIDE OF THE PC BOARD.

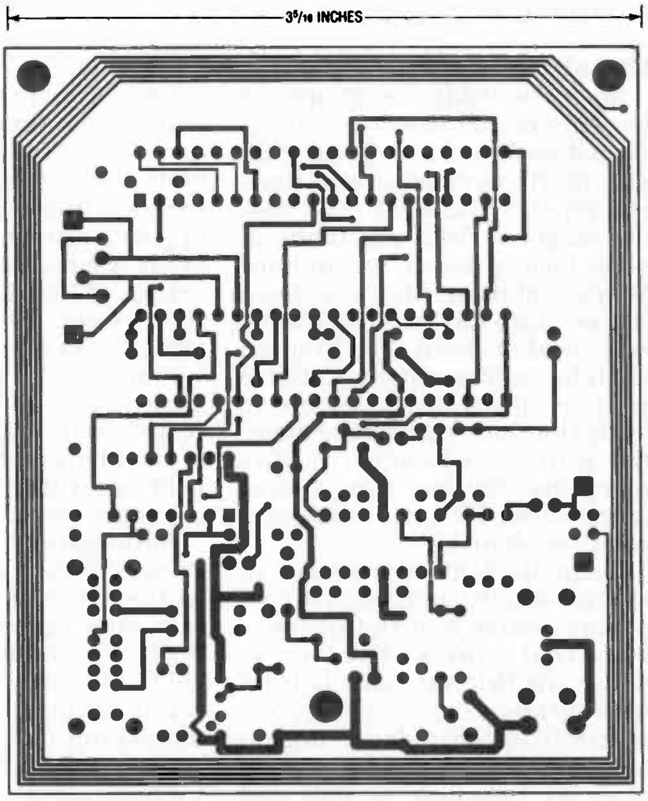

SOLDER SIDE OF THE PC BOARD.

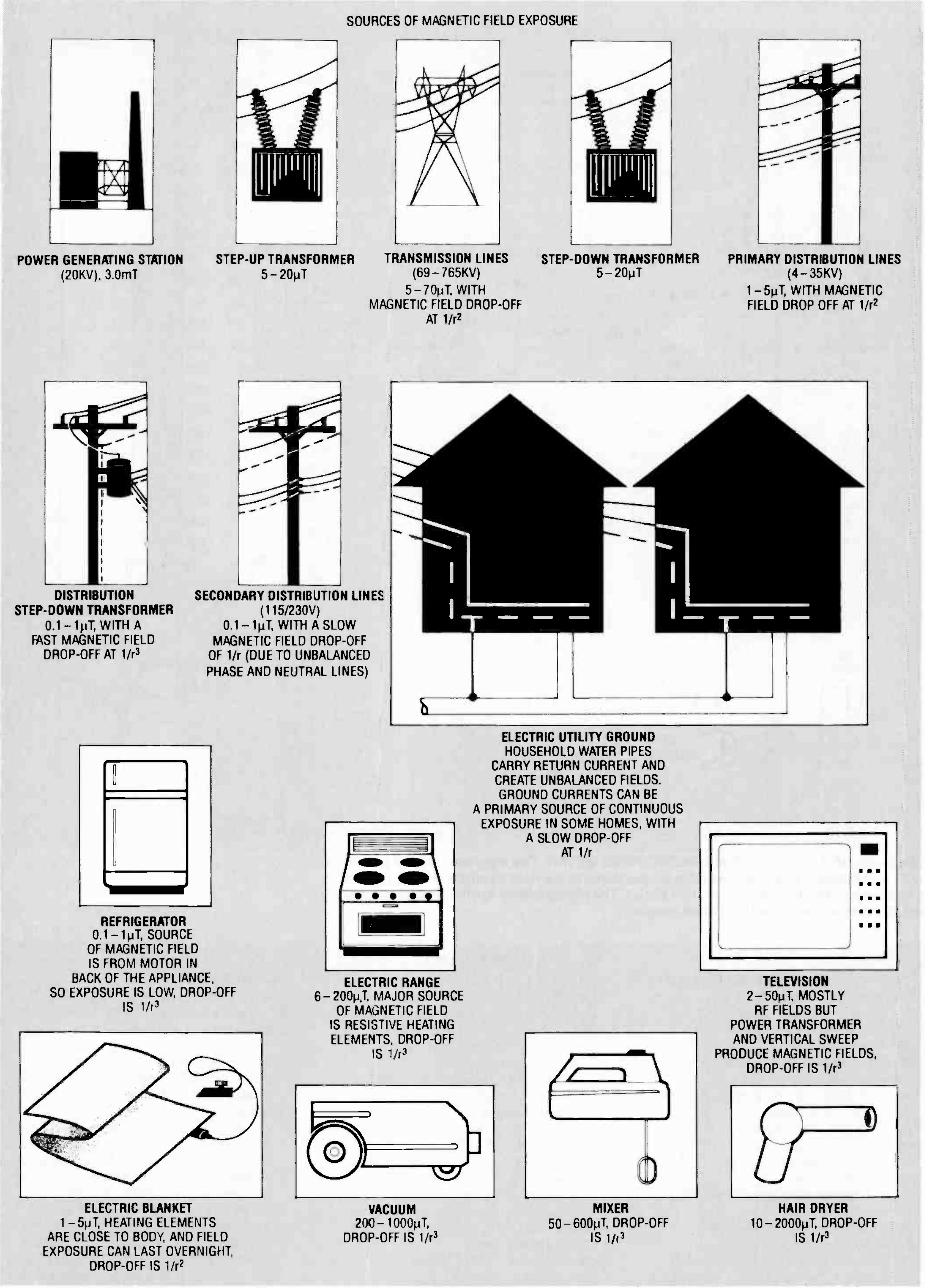
SOURCES OF MAGNETIC FIELD EXPOSURE
POWER GENERATING STATION
(20KV), 3.0mT
STEP-UP TRANSFORMER
5 – 20μT
TRANSMISSION LINES
(69 – 765KV)
5 – 70μT, WITH MAGNETIC FIELD DROP-OFF AT 1/r²
STEP-DOWN TRANSFORMER
5 – 20μT
PRIMARY DISTRIBUTION LINES
(4 – 35KV)
1 – 5μT, WITH MAGNETIC FIELD DROP OFF AT 1/r²
DISTRIBUTION STEP-DOWN TRANSFORMER
0.1 – 1μT, WITH A FAST MAGNETIC FIELD DROP-OFF AT 1/r³
SECONDARY DISTRIBUTION LINES
(115/230V)
0.1 – 1μT, WITH A SLOW MAGNETIC FIELD DROP-OFF OF 1/r (DUE TO UNBALANCED PHASE AND NEUTRAL LINES)
ELECTRIC UTILITY GROUND
HOUSEHOLD WATER PIPES CARRY RETURN CURRENT AND CREATE UNBALANCED FIELDS. GROUND CURRENTS CAN BE A PRIMARY SOURCE OF CONTINUOUS EXPOSURE IN SOME HOMES, WITH A SLOW DROP-OFF AT 1/r
REFRIGERATOR
0.1 – 1μT, SOURCE OF MAGNETIC FIELD IS FROM MOTOR IN BACK OF THE APPLIANCE, SO EXPOSURE IS LOW, DROP-OFF IS 1/r³
ELECTRIC RANGE
6 – 200μT, MAJOR SOURCE OF MAGNETIC FIELD IS RESISTIVE HEATING ELEMENTS, DROP-OFF IS 1/r³
TELEVISION
2 – 50μT, MOSTLY RF FIELDS BUT POWER TRANSFORMER AND VERTICAL SWEEP PRODUCE MAGNETIC FIELDS, DROP-OFF IS 1/r³
ELECTRIC BLANKET
1 – 5μT, HEATING ELEMENTS ARE CLOSE TO BODY, AND FIELD EXPOSURE CAN LAST OVERNIGHT, DROP-OFF IS 1/r²
VACUUM
200 – 1000μT, DROP-OFF IS 1/r³
MIXER
50 – 600μT, DROP-OFF IS 1/r³
HAIR DRYER
10 – 2000μT, DROP-OFF IS 1/r³

FIG. 2—HERE ARE SOME PRIMARY SOURCES of magnetic field exposure with the range of field intensity in teslas, and drop-off rates.

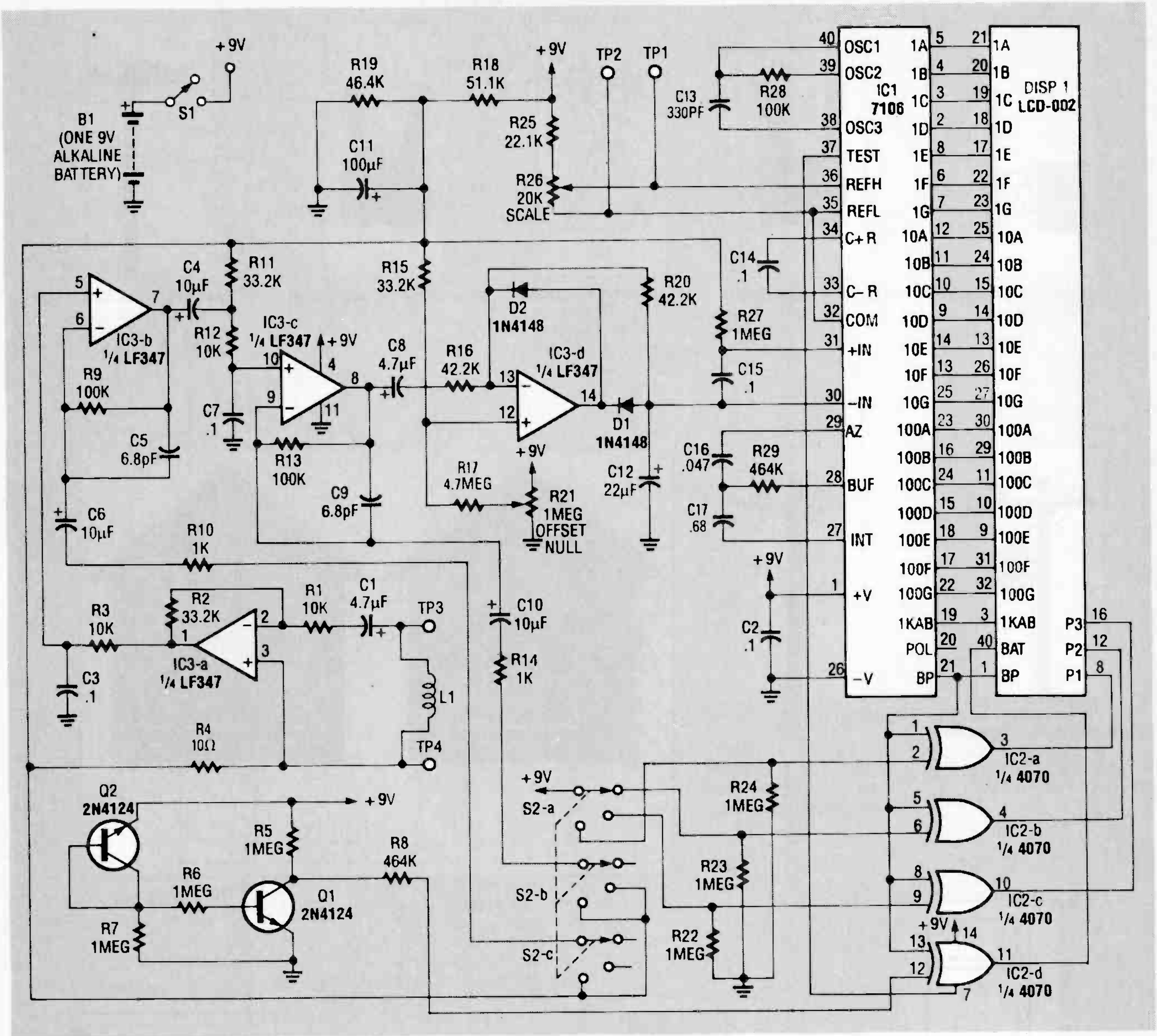
+9V
S1
B1
(ONE 9V
ALKALINE
BATTERY)
R19
46.4K
R18
51.1K
+9V
TP2
TP1
C13
330PF
R28
100K
C11
100µF
R25
22.1K
R26
20K
SCALE
C4
10µF
R11
33.2K
R15
33.2K
C14
.1
R20
42.2K
R27
1MEG
IC3-b
1/4 LF347
R12
10K
IC3-c
1/4 LF347
D2
1N4148
IC3-d
1/4 LF347
C8
4.7µF
R16
42.2K
C15
.1
R9
100K
C7
.1
D1
1N4148
C5
6.8pF
R13
100K
C16
.047
R29
464K
C12
22µF
R17
4.7MEG
R21
1MEG
OFFSET
NULL
C17
.68
C6
10µF
R10
1K
C9
6.8pF
R2
33.2K
R1
10K
C1
4.7µF
TP3
C10
10µF
R3
10K
IC3-a
1/4 LF347
R14
1K
C2
.1
C3
.1
L1
R4
10Ω
TP4
Q2
2N4124
+9V
R5
1MEG
R8
464K
Q1
2N4124
R6
1MEG
R7
1MEG
S2-a
S2-b
S2-c
R24
1MEG
R23
1MEG
R22
1MEG
IC1
7106
OSC1
OSC2
OSC3
TEST
REFH
REFL
C+R
C–R
COM
+IN
–IN
AZ
BUF
INT
+V
–V
1A
1B
1C
1D
1E
1F
1G
10A
10B
10C
10D
10E
10F
10G
100A
100B
100C
100D
100E
100F
100G
1KAB
POL
BP
DISP 1
LCD-002
BAT
P3
P2
P1
IC2-a
1/4 4070
IC2-b
1/4 4070
IC2-c
1/4 4070
IC2-d
1/4 4070

FIG. 3—SCHEMATIC OF THE MAGNETIC FIELD METER. The magnetic field is picked up by L1 and appears as a voltage that is proportional to the field strength at the input of IC3-a, which amplifies the signal to 100 µV per µT. The signal is then further amplified by IC3-b and IC3-c to achieve the three tesla ranges.

PARTS LIST

All resistors are ¼-watt, 1%, unless otherwise indicated.
R1, R3, R12—10,000 ohms
R2, R11, R15—33,200 ohms
R4—10 ohms
R5-R7, R22-R24, R27—1 megohm
R8, R29—464,000 ohms
R9, R13, R28—100,000 ohms
R10, R14—1000 ohms
R16, R20—42,200 ohms
R17—4.7 megohms
R18—51,100 ohms
R19—46,400 ohms
R21—1-megohm potentiometer, 5%
R25—22,100 ohms
R26—20,000-ohm potentiometer, 5%

Capacitors
C1, C8—4.7 µF, 10 volts, electrolytic
C2, C14—0.1 µF, electrolylic or polyester
C3, C7, C15—0.1 µF, polyester
C4, C6, C10—10 µF, electrolytic
C5, C9—6.5 pF, ceramic disc or mica
C11—100 µF, 10 volts, electrolytic
C12—22 µF, 10 volts, electrolytic
C13—330 pF, polyester
C16—0.047 µF, polyester or ceramic disc
C17—0.68 µF, polyester

Semiconductors
D1, D2—1N4148 switching diode
Q1, Q2—2N4124 NPN transistor
IC1—ICL 7106 A/D converter
IC2—4070 or 4030 quad 2-input exclusive-OR gate
IC3—LF347 quad JFET input op-amp
DISP1—LCD-002 liquid crystal display

Other components
S1—MSS1200, SPST (Alco)
S2—MSS4300, SPDT (Alco)
L1—18 turns, 3" diameter remote-sensing coil (optional, see text)
B1—9-volt alkaline battery, with connector
Case—Pac-Tec, HPS-9VB

NOTE: The following items are available from A & T Labs, P.O. Box 4884, Wheaton, IL 60187: A kit of all parts including PC board and case, without battery, $79.00; an etched, drilled and plated through PC board with solder mask and silk-screened parts placement, $15.00; a fully assembled and tested unit, $109.00. Add 6.75% sales tax for Illinois residents, 5% shipping and handling in U.S., 12% shipping and handling in Canada. Check or Money order (UPS COD in contiguous U.S. only) is accepted.

circuit board. For remote sensing, an external field coil probe can be used. Figure 3 shows the complete schematic of the circuit. The magnetic field picked up by the coil appears as a voltage, which is proportional to field strength and frequency at the input of a cascaded amplifier IC3-a, -b, and -c. With a first stage amplifier gain of 3.3 set by R12-R10, the overall sensitivity is 100 μV per μT, or 100 mV per mT. The meter sensitivity is nominally 2 volts full scale, leading to the lowest level sensitivity of 20 mT full scale.

Op-amp IC3-a amplifies the signal to a normalized level of 100 μV per 1 μT. That voltage is further amplified by 1, 100, or 10,000 by IC3-b and -c. The three amplifier stages provide the three magnetic field ranges of 2 mT, 200 μT, and 2 uT (full scale). Components R3-C3 and R12-C7 establish a frequency roll-off characteristic that compensates for the frequency-proportional sensitivity of the pickup coil, and set the 20-kHz cut-off point.

Finally, IC3-d is a precision rectifier and peak detector. Its output drives IC1, a combination analog-to-digital (A/D) converter and LCD driver. Components R25–R29 and C13–C17 are used by IC1 to set display-update times, clock generation, and reference voltages. The decimal points are driven by IC2, as determined by the range-select switch S2. Transistors Q1 and Q2 serve as a low-battery detector, and turn on the battery annunciator in the LCD when the battery voltage drops below 7 volts.

Assembly and checkout

The finished unit shown in Fig. 4 uses a double-sided PC board, which is available from the source mentioned in the parts list. We also show the component side and solder side of the PC board if you choose to make it yourself. You can, however, build the circuit on a perforated construction board if you like, but remember to include the 18-turn remote sensing coil, L1, as indicated in the Parts List. Mount all parts below the LCD display first. It's easier to fix assembly problems if a socket is used with the LCD. Install all parts as shown in Fig. 5 paying attention to component valves and capacitor polarities. If you are using the internal sensing coil, install jumpers between L1-TP3 and L1-TP4.

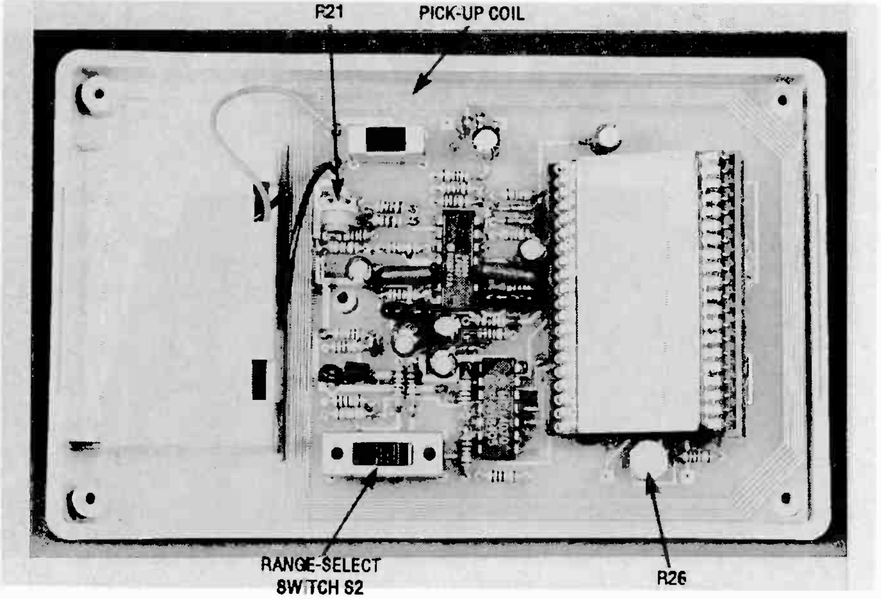

FIG. 4—THIS IS AN INTERNAL VIEW of the magnetic field meter. Assembly is easy, just install all components below the LCD first.

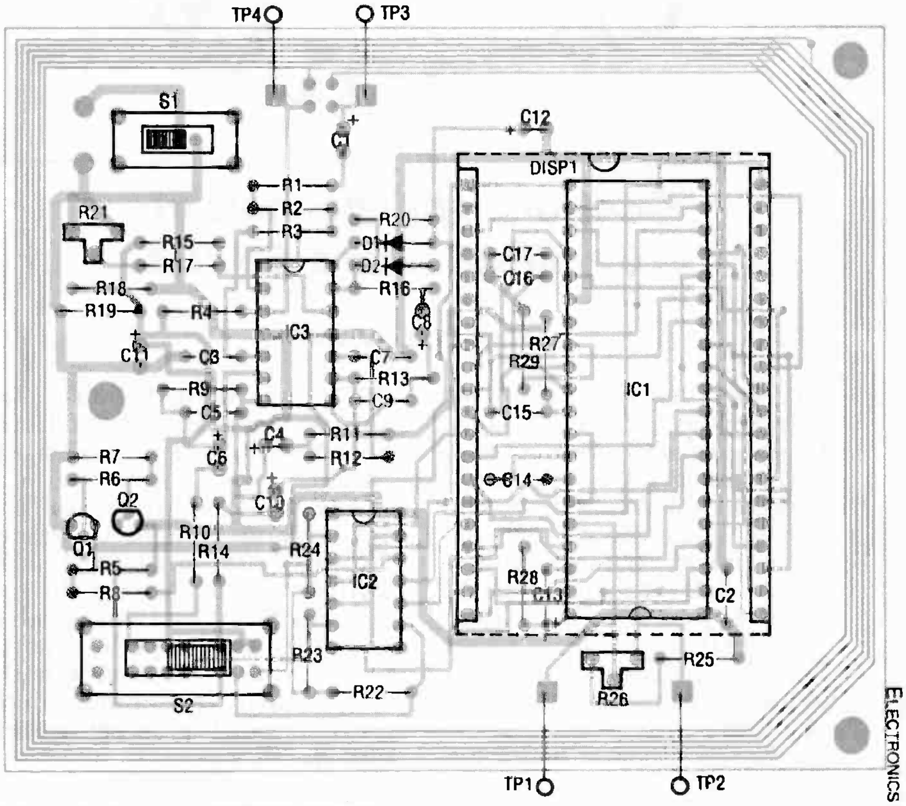

FIG. 5—PARTS PLACEMENT DIAGRAM.

If you are using the case specified in the parts list, raise and angle the display as necessary with wire-wrap IC sockets. Make holes in the front panel for S1 and S2. Mount the finished PC board in the case using a spacer for the single screw holding the center bottom of the board, and attach the battery connector.

With power on, adjust R26 for 1.000 volt between TP1 and TP2. Then, select the 20 mT range and short the pickup coil with a very

(Continued on page 108)

POWER CONTROLLER

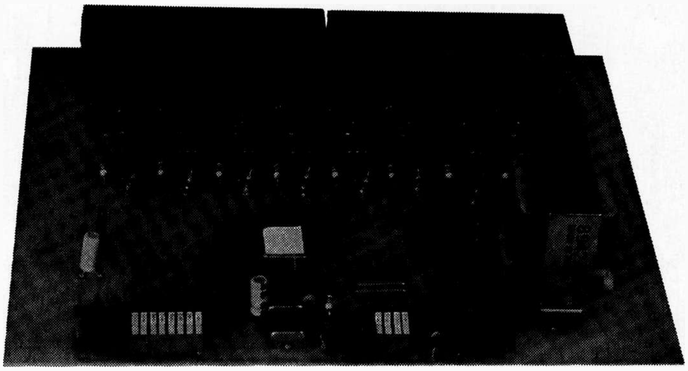

Control eight channels of 120-volt AC with a 68705.

RICHARD L. ROANE

HAVE YOU EVER WANTED TO VARY the power delivered to a 120-volt AC load under software control? In this project a 68705 microcontroller will do just that. Specifically, it can control eight discrete power Triacs, each of which delivers power in 32 smoothly graduated steps ranging from 0 to 97% of full power. The value delivered to one channel is independent of the value delivered to any other channel. Loads can include light displays, universal motors, heaters, and other appliances.

In this circuit, the power level is set by software, not a potentiometer. The software includes a basic set of routines for processing interrupts and setting the power level. The software also includes five test and demonstration routines for putting the circuit through its paces. Moreover, there's plenty of room to add your own routines to the 68705's built-in EPROM. An easy-to-build single-sided circuit board makes assembly easy.

The basic circuit is simple, yet versatile enough to accept inputs from on-board DIP switches; alternatively, the inputs can be driven from a microcomputer bus or parallel port, or a stand-alone device with TTL-compatible outputs. There are 12 input bits to set modes and specify values.

Circuit design

The circuit, shown in Fig. 1, consists of a power supply, AC line synchronizer, Triac output control, input circuitry, and support devices.

The power supply is a conventional 5-volt circuit built around a 7805 regulator (IC3) and several other components (transformer T1, bridge BR1, and C1–C4). The AC line synchronizer consists of an optocoupler (IC2), a current-limiting resistor (R1), and a bias resistor (R2). The power controller generates variable power levels with phase-control techniques. When using a Triac to control a load, the power delivered can be varied by triggering the Triac after the start of an AC half-cycle.

For example, by triggering the Triac shortly after the start of a half-cycle, the load will receive almost full power. Triggering in the middle of a half-cycle will supply the load with half power and triggering close to the end of a half-cycle will result in the load receiving little power.

In this project the controller is synchronized to AC power-line zero crossings, which occur at a rate of 120 times per second. The microprocessor's internal timer then divides the interval between crossings into 32 steps, thus delivering gradually increasing amounts of power to the load.

The output of the optocoupler is a negative-going pulse that drives the microcontroller's interrupt input directly at a rate of 120 hertz. Each time IC1 receives an interrupt, it recalculates the time to generate Triac trigger pulses on Port B of IC1.

Each bit of Port B is connected to a Triac control circuit; since they're all identical, only Port 7 will be described in detail. Port B7 of IC1 drives transistor Q1 via resistor R4. (Q1 is required because the 68705 can't sink enough current to trigger Triac-driver IC4 directly.)

When it's time to turn on Triac TR1, the microcontroller delivers a 10-microsecond negative-going pulse to port PB7. That pulse briefly enables the transistor, which causes current flow through the LED inside IC4. Although a Motorola MCP3011 Triac driver was used in this design, any similar device could be used, as long as it doesn't have an internal zero-crossing detector. That type of driver allows only two modes of operation: full on and full off.

The Triac driver's LED then triggers the photon-coupled Triac within the IC, which in turn triggers TR1 and delivers power to the external load. Resistor R5 provides current limiting for IC4, and R6 and C7 form a snubber network to filter line transients that could cause the Triac to trigger falsely.

The hot side of the power line connects to the MT1 terminal of all the power Triacs (TR1–TR8). The specified Triacs are rated for 4 amperes at 200 volts. They are isolated-tab versions, and were selected because heatsinking is mandatory. With the specified Triacs, the maximum load is 100 watts per output. Larger loads can be handled by substituting Triacs with higher current ratings. Be sure to use proper heatsink and isolation techniques.

Input circuit

The input circuit consists of Ports A and C of the microcontroller, along with two banks each of DIP switches (S1 and S2), SIP resistors (RN1 and RN2), four sets of 0.1 inch header jumpers (P1–P4) for mode selection, a 14-pin header (P5) connected to Port A, and a four-pin header (P6) for input selection.

The P1–P4 jumpers allow both the common resistor legs and the common switch poles to be tied to either ground or +5-volts DC, or to be disconnected from the circuit entirely. If programmed as inputs, the ports should not be left floating because this might cause erratic operation. Connector P6 allows an external device (e.g., a computer) to drive the controller's inputs.

The 68705P3 requires a few support components. Crystal XTAL1 runs at a frequency of 3.58 megahertz, which should not be altered. Capacitor C5 loads the oscillator to keep it at the correct frequency, and capacitor C6 keeps the reset pin low for an instant, thereby giving the 68705 time to stabilize after power up. Momentary switch S3 provides for resetting the circuit. Although unused in the prototype, an external connection is provided to the microcontroller's RESET input, which could allow an external device to reset the circuit when desired.

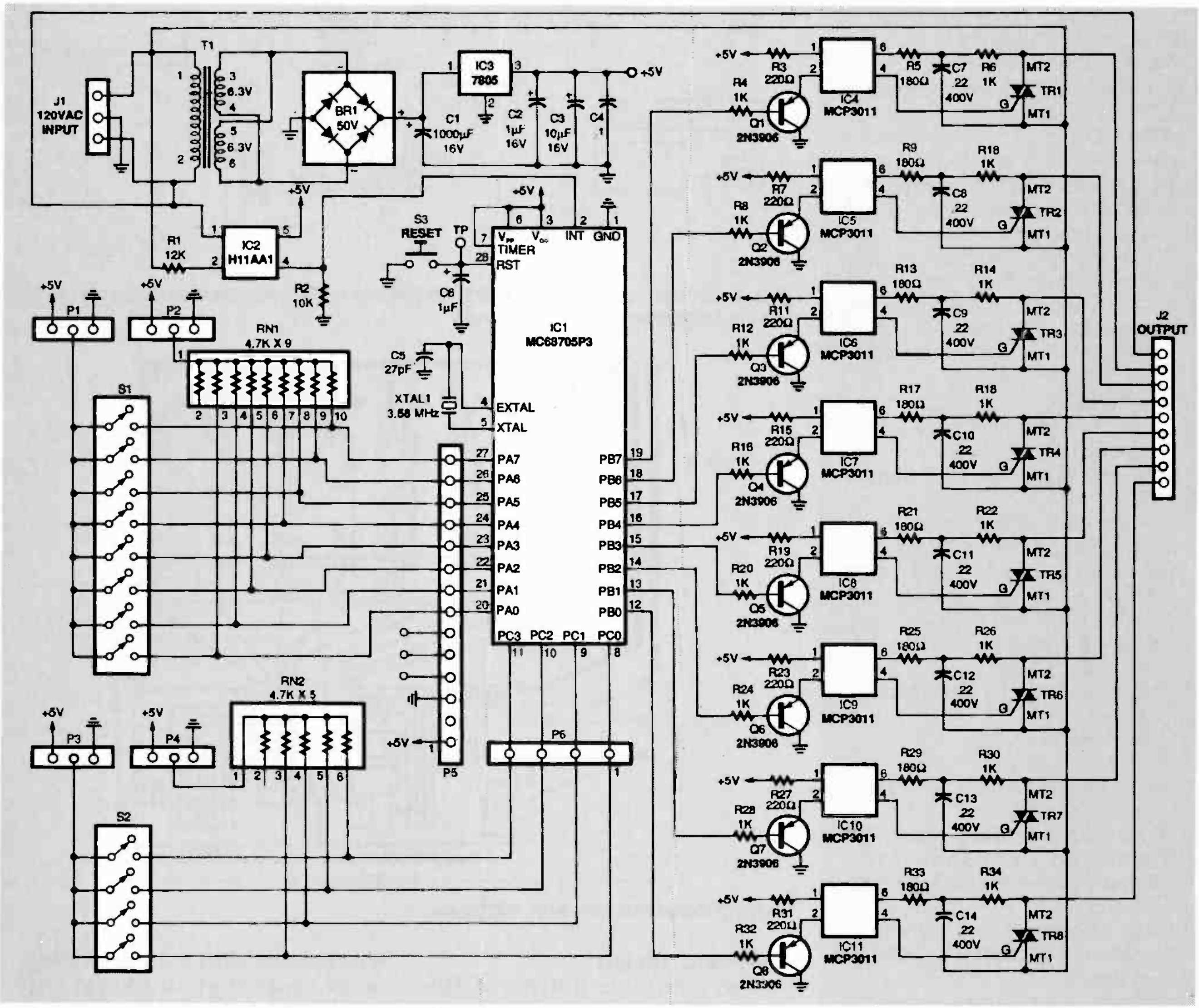
J1
120VAC
INPUT
T1
6.3V
6.3V
BR1
50V
IC3
7805
C1
1000µF
16V
C2
1µF
16V
C3
10µF
16V
C4
.1
+5V
R1
12K
IC2
H11AA1
R2
10K
P1
P2
P3
P4
RN1
4.7K X 9
RN2
4.7K X 5
S1
S2
S3
RESET
TP
C6
1µF
C5
27pF
XTAL1
3.58 MHz
IC1
MC68705P3
VPP
VCC
INT
GND
TIMER
RST
EXTAL
XTAL
PA7
PA6
PA5
PA4
PA3
PA2
PA1
PA0
PB7
PB6
PB5
PB4
PB3
PB2
PB1
PB0
PC3
PC2
PC1
PC0
P5
P6
R3
220Ω
R4
1K
Q1
2N3906
IC4
MCP3011
R5
180Ω
C7
.22
400V
R6
1K
MT2
TR1
MT1
G
R7
220Ω
R8
1K
Q2
2N3906
IC5
MCP3011
R9
180Ω
C8
.22
400V
R10
1K
TR2
R11
220Ω
R12
1K
Q3
2N3906
IC6
MCP3011
R13
180Ω
C9
.22
400V
R14
1K
TR3
R15
220Ω
R16
1K
Q4
2N3906
IC7
MCP3011
R17
180Ω
C10
.22
400V
R18
1K
TR4
R19
220Ω
R20
1K
Q5
2N3906
IC8
MCP3011
R21
180Ω
C11
.22
400V
R22
1K
TR5
R23
220Ω
R24
1K
Q6
2N3906
IC9
MCP3011
R25
180Ω
C12
.22
400V
R26
1K
TR6
R27
220Ω
R28
1K
Q7
2N3906
IC10
MCP3011
R29
180Ω
C13
.22
400V
R30
1K
TR7
R31
220Ω
R32
1K
Q8
2N3906
IC11
MCP3011
R33
180Ω
C14
.22
400V
R34
1K
TR8
J2
OUTPUT

FIG. 1—COMPLETE SCHEMATIC OF THE 68705-based power controller. The controller reads switch values from Ports A and C, and delivers software-controlled 120-volt AC power to Port B.

PARTS LIST

All resistors are ¼-watt, 5%, unless otherwise noted.

R1—12,000 ohms, 1 watt
R2—10,000 ohms
R3, R7, R11, R15, R19, R23, R27, R31—220 ohms
R4, R6, R8, R10, R12, R14, R16, R18, R20, R22, R24, R26, R28, R30, R32, R34—1000 ohms
R5, R9, R13, R17, R21, R25, R29, R33—180 ohms
RN1—4700 ohms × 9 resistor network, common terminal
RN2—4700 ohms × 5 resistor network, common terminal

Capacitors

C1—1000 µF, 16V, radial electrolytic
C2—1 µF, 16V, tantalum
C3—10 µF, 16V, radial electrolytic
C4, C15—0.1 µF, monolithic
C5—27 pF, ceramic
C6—1 µF, 16V, radial electrolytic
C7–C14—0.22 µF, 400V, poly film

Semiconductors

IC1—MC68705P3 microcontroller
IC2—H11AA1 AC optocoupler
IC3—MC7805CT 5-volt regulator
IC4–IC11—MCP3011 Triac driver
Q1–Q8—2N3906 NPN transistor
TR1–TR8—Q2004L3 Triac, 200V, 4A, isolated tab
BR1—Bridge rectifier, 50V, 1A, DIP

Other components

J1—3-terminal AC power connector
J2—9-terminal AC power connector
P1–P4—3-pin single-row 0.1 inch male header
P5—14-pin single-row 0.1 inch male header
P6—4-pin single-row 0.1 inch male header
S1—8-position DIP switch
S2—4-position DIP switch
S3—SPST momentary pushbutton switch
T1—Transformer, 120V primary, dual 6.3V secondaries (Digi-key T101-ND or equiv.)
XTAL1—3.579545 MHz crystal, HC-18

Miscellaneous

Shorting jumpers for P1—P4, heatsink for TR1—TR8, PC board, solder, wire, etc.

Note: The following items are available from Richard L. Roane, Route 601, Cologne, VA 23037:

- **Source code listing and programmed MC68705P3—$24.95 postpaid**

VA residents must add appropriate sales tax.

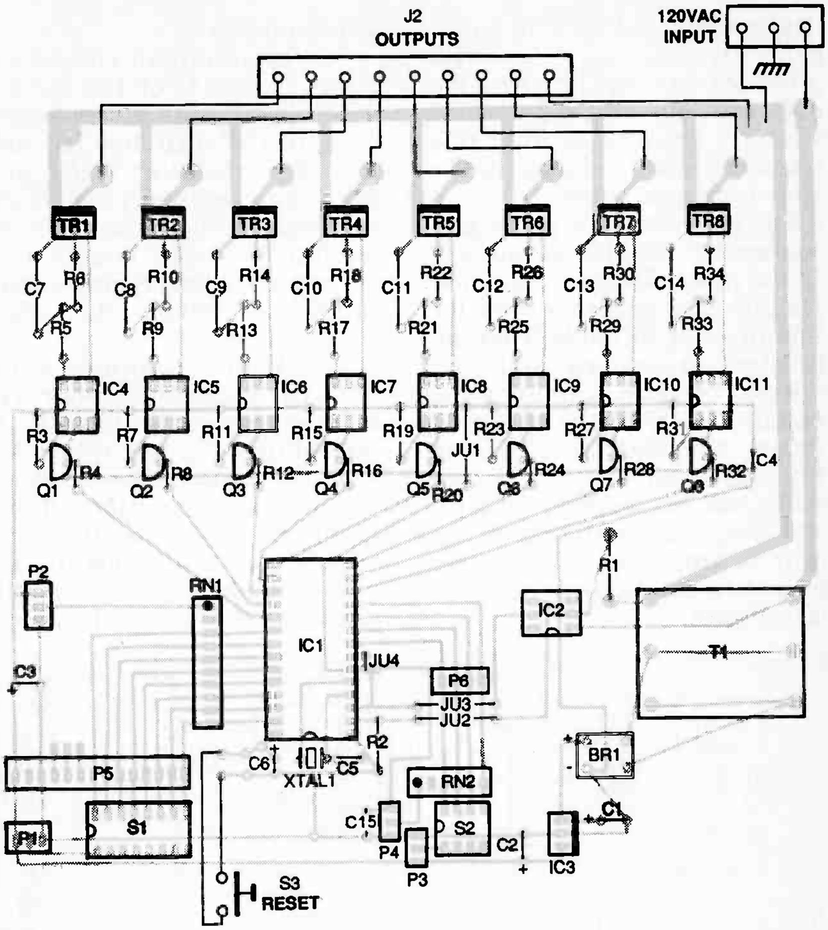
J2
OUTPUTS
120VAC
INPUT
TR1
TR2
TR3
TR4
TR5
TR6
TR7
TR8
C7
R6
R5
C8
R10
R9
C9
R14
R13
C10
R18
R17
C11
R22
R21
C12
R26
R25
C13
R30
R29
C14
R34
R33
IC4
IC5
IC6
IC7
IC8
IC9
IC10
IC11
R3
R4
Q1
R7
R8
Q2
R11
R12
Q3
R15
R16
Q4
R19
JU1
Q5
R20
R23
R24
Q6
R27
R28
Q7
R31
R32
Q8
C4
R1
P2
RN1
IC1
IC2
T1
C3
JU4
P6
JU3
JU2
R2
BR1
C6
XTAL1
C5
RN2
P5
P1
S1
C15
P4
S2
C2
IC3
C1
P3
S3
RESET

FIG. 2—MOUNT ALL COMPONENTS AS SHOWN HERE. Note that four jumpers mount on the component side of the board.

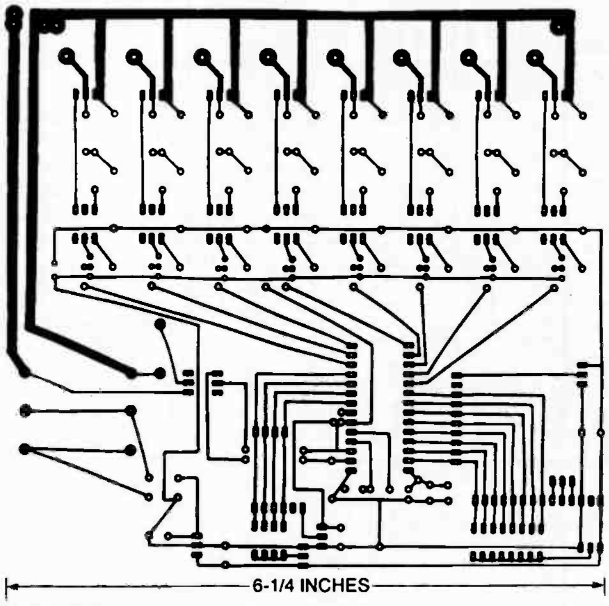

POWER CONTROLLER FOIL PATTERN.

Software design

The complete listing of the software is available from the author, as detailed in the Parts List. It is also available on the Gernsback BBS (516-293-2283, v.32, v.42bis); look for file POWCON.ASM. An overview of the major routines follows:

The interrupt routine sets up

TABLE 1—POWER LEVELS

Power Level	Voltage (VAC)	Percent On (%)
0	1.2	1.0
1	1.9	1.6
2	2.6	2.2
3	6.6	5.5
4	9.2	7.7
5	12.2	10.2
6	15.6	13.0
7	19.5	16.3
8	23.6	19.7
9	27.9	23.3
10	32.8	27.3
11	37.7	31.4
12	42.9	35.8
13	48.2	40.2
14	53.6	44.7
15	59.0	49.2
16	64.4	53.7
17	70.8	59.0
18	75.2	62.7
19	80.2	66.8
20	85.2	71.0
21	90.0	75.0
22	94.5	78.8
23	97.8	81.5
24	101.6	84.7
25	105.3	87.8
26	107.3	89.4
27	110.2	91.8
28	112.7	93.9
29	114.8	95.7
30	116.4	97.1
31	0.0	0.0

TABLE 2
APPLICATION AND JUMPER SELECTION

Value	Application	JU1	JU2	JU3	JU4
0	Manual	GND	+5	+5	GND
1	External	-	-	+5	GND
2	Sequencer	-	-	+	GND
3	Run and Flash	-	-	+	GND
4	Test	-	-	+5	GND
5-15	Not used				

TABLE 3
MANUAL MODE POWER LEVELS

Port C Data Decimal	Power Level Decimal
0	31
1	0
2	2
3	4
4	6
5	8
6	10
7	12
8	14
9	16
10	18
11	20
12	22
13	24
14	26
15	30

the microcontroller's internal timer to produce 31 timer interrupts per half-cycle. Together with the full-off value, those interrupts provide the 32 power levels. Table 1 lists power level, corresponding voltage, and percentage of full on. A power value of 0 represents the lowest on voltage, 30 represents the highest, and anything greater than 30 represents a voltage of 0.0.

Eight bytes of RAM (PWRC0–PWRC7) function as control bytes, each corresponding to a Port B output bit. Each byte stores one item of power level data (as shown in column 1 of Table 1). To see how a value in a control byte is translated into an output voltage, follow this example.

Suppose that an application calls for a power level of 16 at Output 3, and a level of 30 at Output 7. In that case simply store values of 16 and 30 in control bytes PWRC3 and PWRC7, respectively. The outputs will immediately and simultaneously reflect these levels. Note that although the desired voltages were specified in decimal form, you must store the corresponding hexadecimal value in the output-control byte.

A "hole" was left in the microcontroller's EPROM (addresses $80–$5FF) for application programs. The five demonstration routines discussed earlier occupy those addresses. You are free to create your own applications, but leave the addresses outside that range alone. Addresses below $80 are reserved for constant and variable storage; addresses above $5FF contain the reset and interrupt routines. All 68705 registers, except for the timer, are available for use by application routines.

ASSEMBLER VARIATIONS

In developing the software for this article, the intention was to provide a single source file that would assemble correctly under three assemblers: the TECI 6805 Assembler, Motorola's 6805 Assembler, and Motorola's Freeware Assembler. In the course of testing, a bug was discovered in the Freeware assembler. After encountering an instruction that uses direct addressing on page zero of memory, the assembler starts generating garbage. Both of the other assemblers force extended addressing, thus avoiding the problem. The only point in the code where this is a problem is in the jump table, where there is the instruction:

```
JMAN: JMP MAN
```

By changing that instruction as follows, the problem can be avoided:

```
JMAN: BRA MAN
```

Branching, rather than jumping, works because the target routine is within 128 bytes of the jump table.

Construction

Assembly is straightforward. Figure 2 shows the parts layout on the single-sided circuit board. Load the board by inserting lowest- to highest-profile components. Make sure that you observe polarities of all semiconductors, SIP resistors, and electrolytic capacitors.

For the prototype, four Triacs were mounted each to two 90° aluminum heatsinks. The AC power input was brought to one heavy-duty, three-terminal connector, and the eight Triac outputs plus AC common were brought to a separate nine-terminal connector.

The circuit board has 120-volt AC signals on it, so check and double-check your work before plugging in and testing the project. Figure 3 is a photo of the author's prototype.

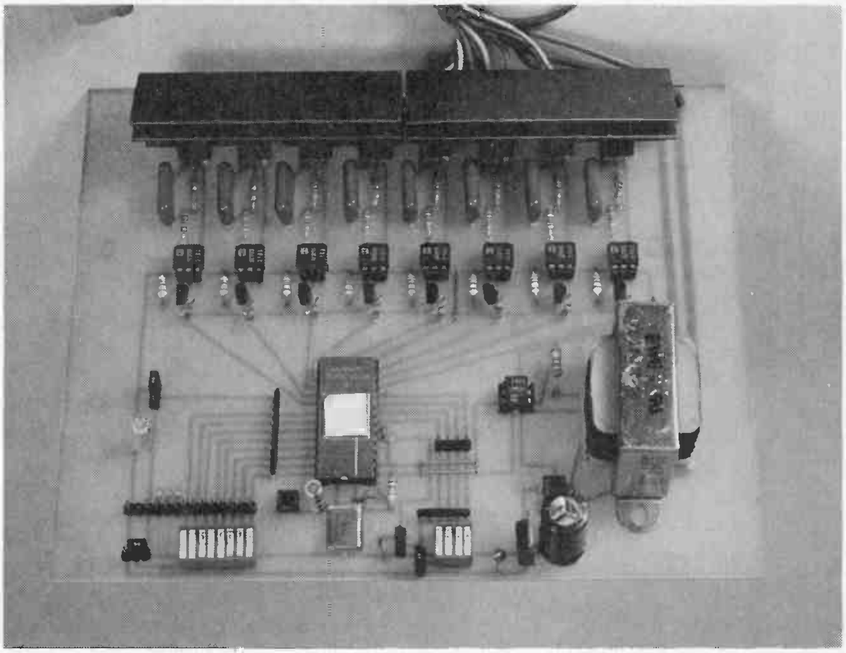

FIG. 3—COMPLETED UNIT. **The circuit board has 120-volt AC signals on it, so check and double-check your work before plugging in and testing.**

Built-in applications

On power up, a selection routine reads Port C and selects the corresponding application. Note that following power-up, Port C can be used for other purposes. Table 2 lists Port C values, along with the corresponding routines and PC-board jumper settings. In general, if no host is connected and Port A will input data from S1, set the jumpers as shown for application 0 (manual). Also note that in the design, once a routine has been selected, the only way to change to a new routine is to reset the system. The following is description of each in detail.

SOUND INPUT

The accompanying schematic shows a simple circuit for letting an audio sound source control the power-controller output. The circuit accepts input from a standard crystal microphone, and amplifies the signal using two op-amp stages. The signal then passes through an emitter-follower circuit that drives a TLC548 A/D converter. Potentiometer R10 functions as a sensitivity control.

The software, presented below, simply clocks the A/D converter eight times, accumulates the result, then stores it in all eight output registers, from where they are applied to the Triacs in the usual manner. To use the routine, you'll have to declare the variable SDDATA, add the routine itself to the jump table, and add a new S1 switch setting to the application-selection power-up routine.

SOUND MODE ROUTINE

```
SOUND:     LDA #$05
           STA DDRA
           BSET 2,PORTA
STARTAD:   CLRA
           LDX #$08
           BCLR 2,PORTA
LOOPAD:    BSET 0,PORTA
           BRSET 1,PORTA,COLLECT
COLLECT:   ROLA
           BCLR 0,PORTA
           DECX
           BNE LOOPAD
           BSET 2,PORTA
           CMP #30
           BHI ADADJUST
STORESD:   STA SDDATA
           BSR APPSD
           BRA STARTAD
ADADJUST:  LDA #30
           BRA STORESD
APPSD:     LDA SDDATA
           STA PWRC0
           STA PWRC1
           STA PWRC2
           STA PWRC3
           STA PWRC4
           STA PWRC5
           STA PWRC6
           STA PWRC7
           RTS
```

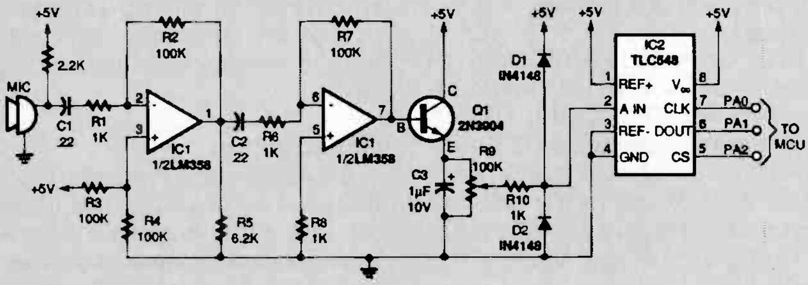

Manual

The software scans Port A waiting for a bit to go low. When that occurs, the microcontroller loads power-level data from Port C and stores it in the power-control byte specified by the value in Port A. For example, placing 05 on Port C and bringing PA2 low will set output 2 to power level 8. Because Port C has only four input bits, it can only specify 16 power values ($2^4 = 16$). Thus a look-up table (shown in Table 3) spreads the 16 specifiable values across the entire 32-value range.

Sequence

This routine demonstrates the controller's ability to power light displays. It tusn on the first output, waits, then turns it

(Continued on page 106)

5 BOOKS FOR ONLY $4.95

plus 1 book FREE upon prepayment

WHEN YOU JOIN THE **Electronics Book Club®**

VALUES TO $165.70

The leading source of information for electronics hobbyists for over 30 years!

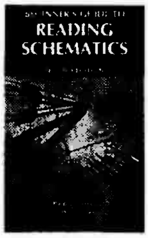
READING
SCHEMATICS

1576398 $10.95

BUILD YOUR OWN
LASER, PHASER, ION RAY GUN
& OTHER WORKING SPACE-AGE PROJECTS

1560696 $17.95

ELECTRONIC
COMPONENTS
A Complete
Reference for
Project Builders

1576606 $19.95

Troubleshooting
& Repairing
Electronic
Circuits

1573119 $19.95

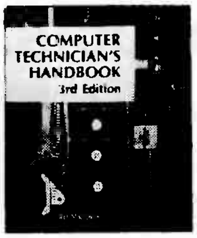
COMPUTER
TECHNICIAN'S
HANDBOOK
3rd Edition

1575472 $27.95

1577262 $21.95

COMPACT
DISC PLAYER
MAINTENANCE AND
REPAIR

1570519 $17.95

TROUBLESHOOTING & REPAIRING
AUDIO &
VIDEO
CASSETTE
PLAYERS &
RECORDERS
HOMER L. DAVIDSON
Schaum's
MATHEMATICAL
HANDBOOK
OF FORMULAS AND TABLES
Murray R. Spiegel
McGRAW-HILL
ELECTRONIC
TESTING
HANDBOOK
JOHN D. LENK
BASIC
ELECTRONICS
COURSE
2nd Edition
ELECTRONICS
POCKET
REFERENCE
SECOND EDITION
EDWARD PASAHOW
HOME
VCR
REPAIR
ILLUSTRATED

1577564 $21.95

03760‥6-XX $45.00
Counts as 2/Hardcover

0602247 $13.95

1574875 $19.95

0707613 $19.95

0487375 $24.95

ELECTRONIC
DISPLAYS

0696217-XX $55.00
Count as 2/Hardcover

MECHANICAL
DEVICES FOR THE ELECTRONICS EXPERIMENTER
HYDRAULICS
PNEUMATICS
CAMS
SOLENOIDS
GEAR TRAINS
VACUUM ACTUATORS

0535469 $29.95
Hardcover

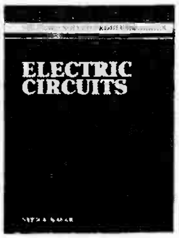
ELECTRIC
CIRCUITS

0535469 $29.95
Hardcover

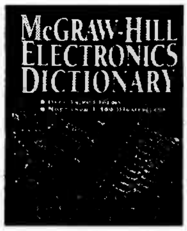
McGRAW-HILL
ELECTRONICS
DICTIONARY

0404348-XX $49.50
Counts as 2/Hardcover

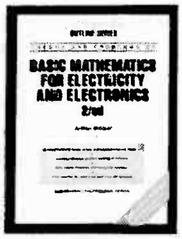
BASIC MATHEMATICS
FOR ELECTRICITY
AND ELECTRONICS

0044392 $12.95

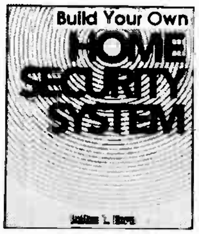
Build Your Own
HOME
SECURITY
SYSTEM

0303932 $16.95

BUILD YOUR OWN
TEST EQUIPMENT

0155593 $22.95

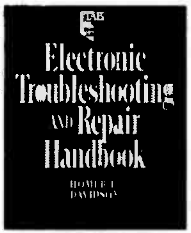
Electronic
Troubleshooting
and Repair
Handbook
HOMER L. DAVIDSON

015676X-XX 69.00
Counts as 2/Hardcover

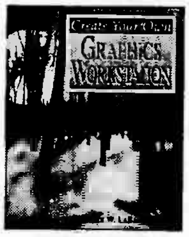
Create Your Own
GRAPHICS
WORKSTATION

0359555 $24.95

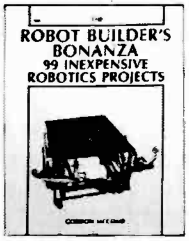
ROBOT BUILDER'S
BONANZA
99 INEXPENSIVE
ROBOTICS PROJECTS

1571469 $17.95

BUILD YOUR OWN
486/486DX

0501815-XX $32.95
Counts as 2/Hardcover

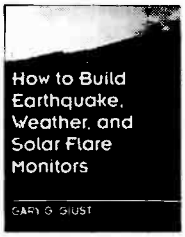
How to Build
Earthquake,
Weather, and
Solar Flare
Monitors
GARY G. GIUST

0252092 $19.95

The Complete Guide to
TROUBLESHOOTING
AND REPAIRING
CONSUMER
ELECTRONICS
WITHOUT
A SCHEMATIC

0156492 $34.95
Hardcover

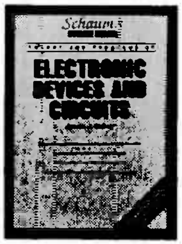
Schaum's
ELECTRONIC
DEVICES AND
CIRCUITS

0102740 $12.95

How
to test
electronic

0304068 $15.95

As a member of the ELECTRONICS BOOK CLUB®... you'll enjoy receiving Club bulletins every 3-4 weeks containing exciting offers on the latest books in the field at savings of up to 50% off the regular publishers' prices. If you want the Main selection, do nothing and it will be shipped automatically. If you want another book, or no book at all, simply return the reply form to us by the date specified. You'll have at least 10 days to decide. If you ever receive a book you don't want due to late delivery of the bulletin, you can return it at our expense. Your only obligation is to purchase 3 more books during the next 12 months, after which you may cancel your membership at any time. And you'll be eligible for FREE BOOKS through our Bonus Book Program.

A shipping/handling charge and sales tax will be added to all orders. All books are softcover unless otherwise noted. If you select a book that counts as 2 choices, write the book number in one box and XX in the next. (Publishers' Prices Shown) ©1996 EBC

REX496

If card is missing, write to: Electronics Book Club®, *A Division of The McGraw-Hill Companies*, P.O. Box 549, Blacklick, OH 43004-9918

It started in America!

The creators are the masters in manufacturing the finest video products...

You probably don't associate VCR's with American technology. Fact is, video recording has its origins in America and it was 3M that brought video recording out of the lab and into your living room. Today, 3M video tape is the choice of all the major networks. No other tape company has ever won an Oscar or an Emmy. 3M Black Watch tape follows in this tradition—service and quality go hand in hand. Here are three Black Watch products you should be using at home!

Clean up! With constant playing and using of degrading dry or wet cleaners, the output of your video tapes has slowly diminished to an unacceptable level and the VCR plays as if it has a head cold! The culprit is most likely clogged and dirty video and/or audio heads. The 3M Black Watch™ Head Cleaner Videocassette uses a patented magnetic tape-based cleaning formation to remove head clogging debris. No foreign substances such as cloth, plastics or messy liquids and no harsh abrasive materials are present. The cleaner's usable life is 400 cleanings or more!

It's easy to use. Place the 3M Black Watch™Head Cleaner Videocassette in the VCR and press the Play button. A pre-recorded message will appear clearly on your screen and an audible tone is heard, telling you that the cleaning process is now completed. No guess work; you never over clean! Priced at $19.95.

For the VCR! Once your VCR's record and playback heads are cured, and the unit plays like new, consider using the finest videocassette you can buy—the 3M Black Watch™ T120 Hi Pro VHS 4410 Videocassette. The 4410 is the highest performing videocassette available today for use with all standard format VHS recording hardware!

Here's what you hear and see....A sharp, clear picture—brightest ever colors—freedom from streaks, flashes and snow—outstanding high-fidelity audio reproduction—optimum camcorder performance—maintains recording integrity. 3M Black Watch™ video tape is 100% laser inspected to guarantee surface smoothness and drop-out free performance. Priced at $8.00

You saw it here first! 3M Black Watch™ 0900 8mm video tape cassette loaded into your Hi Band camcorder delivers the finest picture and sound possible in the 8mm format. Extremely fine particles of pure iron alloyed with nickel and cobalt deliver a video performance exceeding 400 lines of horizontal resolution. You get the advantage of an exceptional video image with superior audio reproduction. This means your Hi 8 format camcorder will produce the best video and audio definition possible. With the 3M Black Watch™ 8mm cassette, the recording capability and performance of your camcorder will be significantly enhanced. Priced at $14.95.

BLACK WATCH
P6 120
Hi8
0900
3M
For the clearest pictures you ever record!

Claggk Inc. — 3M VHS Special Offer
P.O. Box 4099,
Farmingdale,
New York 11735

Yes, I like your offer and here is my order for 3M Black Watch™ products!

___ 3M Black Watch™ 0900 Hi Band-120 8mm Cassette **($14.95 each)** $______

___ 3M Black Watch™ T120 Hi Pro VHS 4410 Videocassette **($8.00 each)** $______

___ 3M Black Watch™ Head Cleaner Videocassette **($19.95 each)** $______

Shipping and handling per order **$4.00**

Total Amount in U.S. Funds only $______

New York residents add local sales tax. Canadians **add $6.00** per order. No foreign orders. Do not send cash.

Bill my ☐ VISA ☐ MasterCard
Card No.______________________
Expire Date ___/___/___
Signature ______________________
Name (Please print)______________
Address ______________________
City____________State_____ZIP ____

POWER CONTROLLER

Continued from page 104

off. It then repeats that action for output 2, and so on. After output 7 toggles, the software then repeats the cycle.

Run and flash—This is a variation of the sequencer application. It sequences the outputs for five cycles, then flashes all eight three times and repeats the cycle.

Test—The test application tests the controller. The routine steps the outputs through all 32 power levels, pausing one second between each.

External—Here's where things get interesting. This routine allows an external source, such as an intelligent I/O module or a computer, to control power levels. To get into this mode, set switch S2 to a value of 2, install the correct jumpers, then reset the circuit. From then on, input Port C0 functions as a strobe that causes the microcontroller to read the value on Port A and perform the proper function.

When PC0 goes low, the microcontroller reads Port A. It then splits the value into a 5-bit power specification (PA0–PA4), and a 3-bit output port (PA5–PA7). The 5-bit power specification allows 1 of 32 values; the 3-bit output port allows 1 of 8 ports.

For example, assume you place the value $57 on Port A and strobe PC0 low. In binary, $57 = 0101 0111. Taking the upper three bits yields 010, or 2. Taking the lower 5 bits yields 10111, or 23. Thus the controller will set output 2 to level 23.

Summing up

As you can see, both hardware and software are simple and suitable for being customized. For example, you could connect the output of an A/D converter to Port A, then vary power levels based on some analog quantity.

When creating your own routines, be sure to include it in both the power-on Select routine and the Jump Table. That way the controller will know where to find it. Ω

THE PROJECT ANTHOLOGY

Continued from page 80

the circuit to be matched to the direct signal from the guitar.

A transconductance amplifier has an extra input, and it is this extra input that makes these components so useful in voltage-controlled filters and amplifiers. The output current is governed by the differential input current and the current fed to the control input. In effect, the gain of the amplifier is controlled by the current fed to the control input. In this case R13 is connected in series with the control input, so that it is voltage rather than current control that is obtained. R12 provides a small bias current to the control input under quiescent conditions. Without R12 there is a tendency for the output signal to decay a little too fast. Also, when the VCA provides very high attenuation levels the output signal seems to become heavily distorted, and produced a rather unmusical "buzzing" sound. R12 completely eliminates both problems.

The transconductance amplifier in the LM13700N has a fourth input, and this can be used to supply a bias current to linearizing diodes at the input of the amplifier. This enables higher signal levels to be handled. In this circuit R10 provides the bias current to the linearizing diodes. The LM13700N used for IC2 is actually a dual transconductance amplifier and buffer amplifier, but in this circuit only one amplifier and buffer stage are utilized. No connections are made to the other amplifier and buffer stage. Note that the LM13600N is virtually identical to the LM13700N, and will work just as well in this circuit.

Chip IC3 is used in the amplifier which drives the smoothing and rectifier circuit. This is an inverting amplifier which has its closed-loop voltage gain controlled by VR3. The maximum voltage gain is about 120 times, and this is obtained with VR3 at maximum resistance. The wide range of gains available should enable good results to be obtained using most guitar pick-up.

The rectifier and smoothing circuit is a simple half-wave type based on D7 and D8. Germanium diodes are used for D7 and D8 because they have lower forward voltage drops than silicon types, and they therefore give a DC output voltage which is a more accurate reflection of the input signal's amplitude. The attack and decay times of the smoothing circuit are kept quite short so that the output envelope accurately follows the envelope of the input signal. On the other hand, the decay time is made sufficiently long to avoid unwanted distortion products on the output signal.

The current consumption of the circuit is about 8 milliamps. This can be supplied by a 9-volt transistor radio battery, but it would be more economic to use a higher capacity battery (six AA cells or larger in a holder) if the unit is likely to receive a lot of use.

Adjustment. Adjust VR1 to give the best distortion effect. It is advisable to use the lowest resistance that gives good results, as higher resistances will give greater voltage gain, which will in turn encourage problems with feedback and general electrical noise. VR3 should be adjusted before giving VR2 its final setting. Adjusting VR3 is again just a matter of finding the setting that gives what is judged to be the best effect. Setting VR3 too low in value will result in rather a low-output level, and generally unimpressive results. Using a value that is too high will give plenty of output signal, but the initial part of the output envelope might be compressed slightly. This will give an output that is slightly lacking in terms of "twangyness." In between these two extremes there should be a fairly broad range of settings that give good results. Once VR3 has been given a suitable setting, VR2 is adjusted to balance the volume levels obtained with the effect switched in and out. Ω

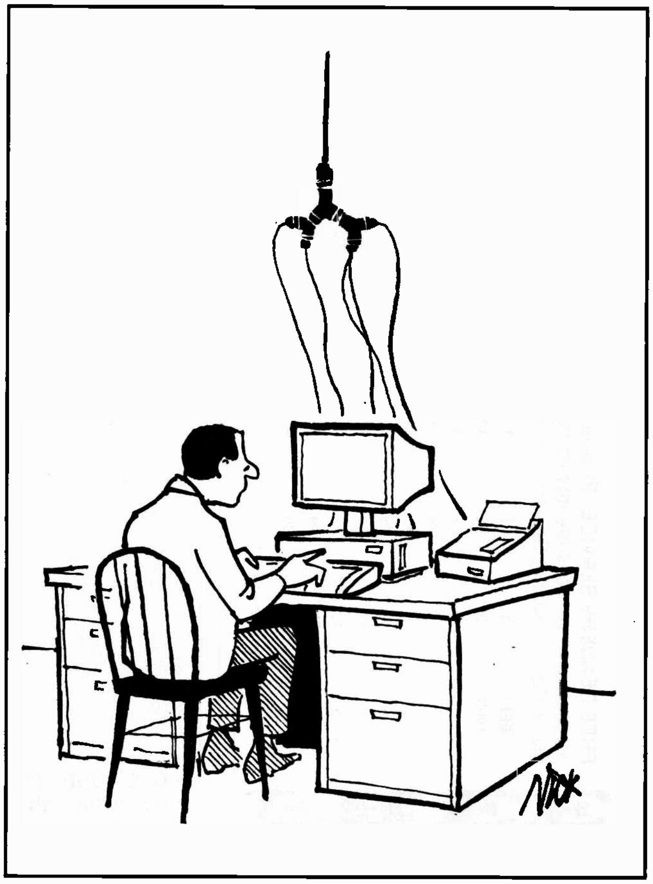

FREE
VOID AFTER JULY 1, 1996

FREE READERS SERVICE: To receive more information about products and services mentioned in this issue, please circle the applicable Reader Service Number, and mail today!

EEH 1996	11	21	31	41	51
	12	22	32	42	52
	13	23	33	43	53
	14	24	34	44	54
5	15	15	35	45	55
6	16	26	36	46	56
7	17	27	37	47	57
8	18	28	38	48	58
9	19	29	39	49	59
10	20	30	40	50	60

Additional Coupons on pages 15 and 111.

1 ☐ Please send me 12 issues of POPULAR ELECTRONICS® for $18.95 and bill me. (Canada $25.63 U.S. FUNDS — includes G.S.T.)
2 ☐ Please send me 12 issues of ELECTRONICS NOW® for $17.97 and bill me. (Canada $25.65 U.S. FUNDS — includes G.S.T.)

Mail to: ELECTRONICS EXPERIMENTER'S HANDBOOK, Reader Service Management Dept.,
PO Box 5192, Pittsfield, MA 01203-9989

Name ____________
Please Print
Company Name (if applicable) ____________
Address ____________
City ____________ State ____________ Zip +4 ____________

Allow 6-8 weeks for delivery of first issue. OFFER VALID U.S. AND CANADA ONLY.

FREE

MAGNETIC FIELD METER

Continued from page 99

short lead between TP3 and TP4. Adjust offset-null potentiometer R7 for 0.00. Remove the jumper, and the meter is complete.

Calibration

Calibration of the meter is basically determined by the pickup-coil characteristics, amplifier gains, and meter reference-voltage setting. The amplifier gains, as we previously discussed, are chosen to match the coil characteristics as closely as possible.

If you desire to calibrate your meter more exactly, you will need to generate a known magnetic field intensity. One way to do that is to pass a known current through a coil configuration whose field pattern characteristics are known. Figure 6 shows such a calibration setup. A good controllable signal source is a sine-wave generator and an audio amplifier, whose output is coupled to a coil through an 8-ohm resistor. Measuring the voltage across the resistor gives the current. Then, calculate the magnetic field according to Fig. 6. (Note that while all references to field strength here are made in teslas, gauss are also commonly used. The conversion is easy: 1 tesla = 10,000 gauss.)

Place the meter inside the coil and turn it on. Use the highest sensitivity scale that does not overrange the display. An overrange is indicated by a display of 1 followed by three blanks. In most cases, the 2 μT range is satisfactory.

Measurement interpretation

A great deal of controversy exists in the emerging understanding of potential health hazards of low-frequency magnetic fields. The International Radiation Protection Association (IRPA) has set some interim standards based on 1984 World Health Organization guidelines. Those IRPA standards specify a continuous maximum magnetic field exposure for the general public of 100 μT, and 500 μT as the maximum occupational exposure allowed over the entire working day. Ω

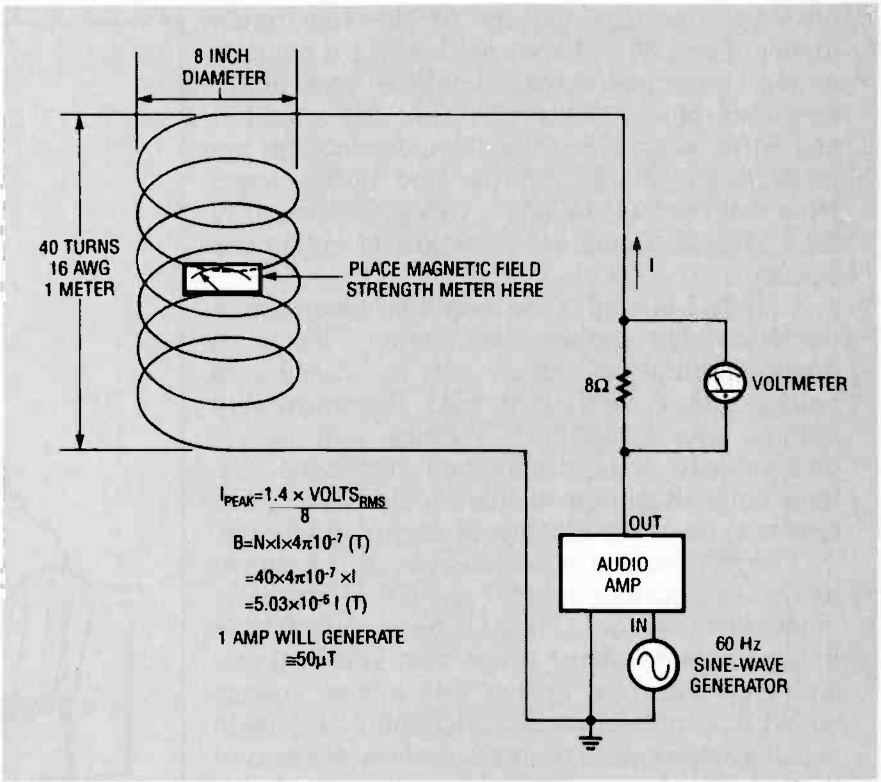
8 INCH DIAMETER
40 TURNS
16 AWG
1 METER
PLACE MAGNETIC FIELD STRENGTH METER HERE
I
8Ω
VOLTMETER
OUT
AUDIO AMP
IN
60 Hz SINE-WAVE GENERATOR
$I_{PEAK} = \frac{1.4 \times VOLTS_{RMS}}{8}$
$B = N \times I \times 4\pi 10^{-7}$ (T)
$= 40 \times 4\pi 10^{-7} \times I$
$= 5.03 \times 10^{-5}$ I (T)
1 AMP WILL GENERATE ≅50μT

FIG. 6—USE THIS TEST SETUP TO accurately calibrate your meter. A known current is passed through a coil whose field intensity is known. A sine-wave generator provides the 60 Hz frequency, and an audio amplifier is coupled to the coil by an 8-ohm resistor. Measure the voltage across the resistor, and use the calculations shown.

NEW PRODUCTS

Continued from page 12

rate tester (BERT).

The Model 904 can be connected to any personal computer through any available serial port. The analyzer uses the serial port in a way that provides a high-speed data channel along with a serial control channel. The analyzer's embedded coprocessor permits the monitoring and emulation of data lines rated to 64 kilobytes per second on a full-duplex basis.

The Model 904 supports most WAN protocols, including asynchronous and synchronous protocols such as BiSync and IPARS, and bit-oriented protocols such as HDLC, SNA, and X.25. The protocol analyzer captures all parameters of the communications link: data, control leads, time stamps, and error conditions.

That information permits extensive post-capture analysis of data, even at a different location. Any combination of data line events can be searched or trapped. This allows the review of live data or search through captured data. Captured data can be saved as a DOS file on disk for later viewing, analysis, or printing. The included protocol emulator program interface allows the use of Telebyte off-the-shelf test analysis programs or the user can write his own custom programs.

In addition to monitoring, the Model 904 has emulation capabilities. A menu-driven program on the PC allows the selection of many combinations of communications options to emulate X.25, HDLC, Async, and BiSync. A built-in BERT tester offers a choice of test patterns and modes of operation. The analyzer is supplied with standard RS-232C monitoring software and two additional programs for X.25 and HDLC.

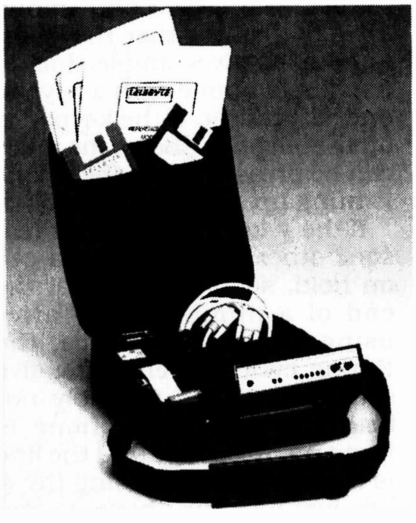

CIRCLE 45 ON FREE INFORMATION CARD

For portable field operation the Model 904 will run for 2 hours on a standard 9-volt battery. Since problems don't always happen in the same place, being able to hook up to the nearest PC without having to open the PC's case is convenient.

The Model 904, supplied in a carrying case with cables, software, and manuals, is priced at $499.00

TELEBYTE TECHNOLOGY, INC.
270 Pulaski Road
Greenlawn, NY 11740
Phone: 1-800-835-3298
Fax: 516-385-8184 or 516-385-7060
for complete data sheet
E-mail: Sales@telebyteUSA.com

NEW LITERATURE

Continued from page 15

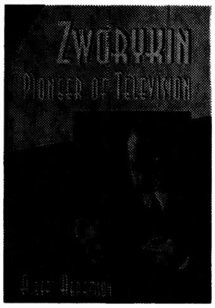

CIRCLE 39 ON FREE INFORMATION CARD

of the iconoscope, the first practical television camera tube. To write this book, Abramson researched Zworykin's original patents and both his published and unpublished notebooks and papers, and he interviewed many television pioneers, including Zworykin himself.

This biography reviews the life of this television pioneer and scientist, including its often dramatic periods. It covers Zworykon's early years in Russia as well as his career at the RCA laboratories. He was invited there by RCA founder David Sarnoff. More than 50 photographs in the book highlight Zworykin's work. Ω

THERMOMETER

Continued from page 42

perature of ice water (32°F) and 80°F water be the test points.

The temperature of the hot water can be determined accurately with a standard laboratory mercury thermometer, but a high-quality, liquid-filled confectioner's thermometer will also give satisfactory results.

Measuring temperature

Observe common sense precautions when measuring temperature with your digital thermometer. If the transistor/sensor is held near an open flame or heating coil it will be damaged or destroyed.

If you put the digital thermometer in a protective case to protect it from dust, rain, and salt spray, it will work reliably out of doors, in boats, or even on camping trips.

Keep the digital thermometer's temperature measurement limits in mind. The Harris ICL7106CPL has a rated temperature range of 0 and 70° Celsius. Ω

MINI ANALYZER

Continued from page 94

only eight). This approach checks all the functions of the 4017, including reset, clocking, clock enabling, and carry out. A second scan could be performed to sense the previously unchecked outputs instead of IC outputs 6, 7, and 8.

Most digital ICs can be tested in a similar process. That is, you must determine the necessary inputs to exercise all functions of the IC, program those inputs, determine where to observe the outputs and what those outputs should be. Then you can compare the results.

To make a simple jig for this purpose, you can use a general-purpose IC PC Board (Radio Shack Part No. 276-159) and a 20-pin IC socket. Solder the socket in place, and solder a solid wire loop on each of the input/output pads. The jig is shown in Fig. 16. Ω

Plant Trees for America™

Trees increase property values, and make our homes and neighborhoods more livable.

Trees provide food, shelter, and nesting sites for songbirds.

Trees help conserve energy. They cool our homes and entire cities in the summer, and slow cold winter winds. Shade trees and windbreaks can cut home utility bills 15-35%.

America needs more trees

The United States has lost a third of its forest cover in the last 200 years.

Our towns should have twice as many street trees as they have today.

We need more trees around our homes and throughout our communities. We need more trees to protect our farm fields and our rivers and streams. To provide wood for our homes and a thousand products we use every day.

10 Free Trees

Ten Colorado blue spruces, or other conifers selected to grow in your area will be given to each person who joins the Arbor Day Foundation.

Your trees will be shipped postpaid at the right time for planting in your area, February through May in the spring or October through mid December in the fall. The six to twelve inch trees are guaranteed to grow, or they will be replaced free.

You will also receive The Tree Book with information about tree planting and care.

To become a member and to receive your free trees, send a $10 membership contribution to Ten Blue Spruces, National Arbor Day Foundation, 100 Arbor Avenue, Nebraska City, NE 68410.

Join today, and plant your Trees for America!

The National Arbor Day Foundation

ISOLATION TRANSFORMER

Continued from page 45

agram (Fig. 1) and begin the wiring by connecting the hot side of the line cord (black wire) to one side of the fuse holder. Next solder a wire from the other side of the fuse holder to one terminal of the main power switch, S1.

Connect the other terminal of S1 to the other side of the fan and to one terminal of S2. Connect power indicator NE1 across the fan terminals so it will indicate when power is applied to the fan and to S2.

Standby switch S2 allows power to the transformer to be shut off while still allowing the fan to work. Wire indicator NE2 across S2 so that it illuminates when S2 is open, or when the unit is in the standby mode. Next, connect S2 to the other side of the transformer. Connect the neutral (white) wire from the line cord to one side of the fan and also connect it to one terminal on the input side of the transformer.

Connect both output leads of the transformer to the AC receptacle SO1. Next connect the output-power indicator NE3 across the receptacle. As a last important step, run a ground connection from the line cord (the green wire) to the base of the transformer and to the fan. If the transformer has not been grounded to the chassis, be sure to do so now with a separate wire.

Before closing up the case, plug an appliance such as a coffee maker, a lamp, or another appliance that draws about 500 watts into the isolation transformer outlet and turn on the power. Run the transformer with the load turned on for approximately 30 minutes while checking it every five minutes or so for excess heat.

The prototype was tested with an 800-watt coffee maker, and the transformer ran warm—but not hot—to the touch. Do not use the isolation transformer to power anything with that high a power rating for extended periods of time. Ω

PHONE ON HOLD

continued from page 38

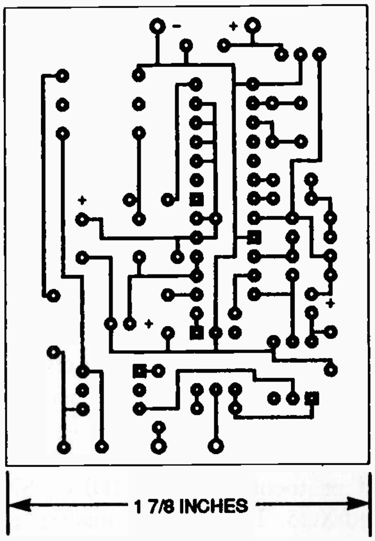

FOIL PATTERN for the universal hold circuit's single-sided board.

the case.

Before closing up the case, connect the hold circuit to the telephone line and plug the wall transformer into an AC outlet. Pick a phone up, press the # key, and hang it up within five seconds; the LED should light. Pick a phone up and the LED should go out. Press the # key, wait about eight seconds, and hang the phone up; the LED should not light.

Operation

The power supply, DTMF decoder, and timing circuits are isolated from the telephone line by the optoisolators and transformer. The five-second timing cycle that starts after the # key is released gives ample time to hang up a phone with a keypad on the handset. If the keypad is on the telephone base, the # key can be pressed while the phone is hung up.

If the # key is pressed for reasons other than to place a call on hold, such as to signal the end of a number entry after using a fax-back service, the timing cycle will end after five seconds and will probably not be active when the phone is hung up. To be sure that the line is not on hold after using the # key, just pick the phone up and hang up again. Ω

FOR SALE

THE Case Against the Patents. Thoroughly tested and proven alternatives that work in the real world. $28.50. **SYNERGETICS PRESS,** Box 809-C, Thatcher, AZ 85552. (602) 428-4073. Visa/MC.

PLANS AND KITS

60 SOLDERLESS Breadboard Projects in two easy-to-read pocket books. Complete with circuit descriptions, schematics, parts layouts, component listings, etc. Both books (BP107 & BP113) only $11.90 plus $3.50 for shipping. USA and Canada only. US funds. **ETT, INC.,** PO Box 240, Massapequa Park, NY 11762-0240.

CRYSTAL set Handbook — Visit antiquity by building the radios your grandfather built. Assemble a "Quaker Oats" rig, wind coils that work and make it look like the 1920's! Only $10.95 plus $3.50 for shipping and handling. **CLAGGK INC.,** PO Box 4099, Farmingdale, NY 11735. US funds only! USA and Canada — no foreign orders.

ALL-in-one catalog. AM/FM/ham/spy, transmitters, amplifiers, receivers. Voice scramblers/disguisers, audio, TV, tesla coils, plans, "secret" books, kits, imports, exports and more. Start your own licensed or unlicensed radio station, 60 full pages for $1.00. **PAN-COM INTERNATIONAL,** PO Box 130-FH, Paradise, CA 95967.

ADVERTISING INDEX

ELECTRONICS NOW magazine does not assume any responsibility for errors that may appear in the index below.

Free Information No.		Page
13	Amaze Electronics	11
—	CLAGGK, Inc.	4, 90, 106
—	Cleveland Institute of Electronics	17
—	C&S Sales	CV2
—	Elecrtronics Technology Today	CV3
—	Electronics Technology Today	81, 12
—	Fotronic	5
—	Information Unlimited	7
11	Interactive Image Technology	CV4
—	Mouser	35
—	NRI Schools	9
—	School of PC Repair	13
—	School of VCR Repair	13
12	Sescom	11
—	Software Science	31
—	Tab	105

Make
U.S. Savings Bonds
part of your
retirement savings
program.

For a recorded message of current rate information, call
1-800-4US BOND
1-800-487-2663

Take Stock in America U.S. SAVINGS BONDS
A public service of this magazine

BUSINESS OPPORTUNITIES

START your own technical venture! Don Lancaster's newly updated **Incredible Secret Money Machine II** tells how. We now have autographed copies of the Guru's underground classic for $18.50. **SYNERGETICS PRESS,** Box 809-C, Thatcher, AZ 85552. (602) 428-4073. Visa/MC.

LET the government finance your small business. Grants/loans to $500,000.00. Free recorded message: (707) 488-0270. (FO4).

EDUCATION/INSTRUCTION

TODAY'S electronics. Don't get left behind! Top home-study courses developed by U.S. Military. Free catalog. **FEDERAL TECHNICAL PUBLISHERS,** Box 1218, E. Glen Lake, MN 55345.

WANTED

INVENTIONS, ideas, new products! Presentation to industry/exhibition at national innovation exposition. Patent services. **1 (800) 288-IDEA.**

CABLE TV

UNIVERSAL descrambler, $129.95 + $5.00 S&H. **"Bullet" Buster,** $19.95 + $3.00 S&H. Booklets, **"Cable Unscrambled"** and **"Descrambler Schematics Revealed",** $10.00 each... and much more! Visa/MC. **ELECTROMAN,** Box 24474, New Orleans, LA 71084. (504) 482-3017.

BEST BY MAIL

Rates: Write National, Box 5, Sarasota, FL 34230

OF INTEREST TO ALL

PINBALL MACHINES BOOKS, Parts, Supplies. Catalog $1.00. MARCO, 2401-(EEH) Sunset, West Columbia, SC 29169-4717.

PERSONAL COMPUTER OWNERS CAN EARN $1000 to $5000 monthly offering simple services part time. Free list of 100 best services. C.B.C.A.Z, P.O. Box 926168, Houston, TX 77292.

Prototype it....... *FAST!*

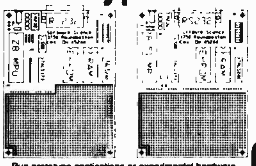

with ProtoQuick 8051 or Z8
- Complete single board computers
- Up to 32K EPROM and 8K RAM
- 12 sq. in. plated-through proto area
- RS232 C serial port w/ DB25 conn.
- Op sys in EPROM w/ source code
- Assembled, ready to run - 5v only
- MS-DOS cross-assembler included

ProtoQuick Z8 and 8051

$99.00 each

Software Science
3750 Roundbottom Road
Cincinnati, OH 45244 USA
(513) 561-2060

Run prototype applications or experimental hardware from the serial port - **WITHOUT PROGRAMMING!**

THE COLLECTED WORKS OF MOHAMMED ULLYSES FIPS

#166—By Hugo Gernsback. Here is a collection of 21 April Fools Articles, reprinted from the pages of the magazines they appeared in, as a 74-page, 8½ × 11-inch book. The stories were written between 1933 and 1964. Some of the devices actually exist today. Others are just around the corner. All are fun and almost possible. Stories include the Cordless Radio Iron, The Visi-Talkie, Electronic Razor, 30-Day LP Record, Teleyeglasses and even Electronic Brain Servicing. Get your copy today. Ask for book **#166** and include **$16.00 (includes shipping and handling)** in the US and Canada, and order from **CLAGGK Inc., P.O. Box 4099, Farmingdale, NY 11735-0793.** Payment in US funds by US bank check or International Money Order. Allow 6-8 weeks for delivery.

MA05

BUY BONDS

FREE

VOID AFTER JULY 1, 1996

1 ☐ Please send me 12 issues of POPULAR ELECTRONICS® for $18.95 and bill me. (Canada $25.63 U.S. FUNDS — includes G.S.T.)

2 ☐ Please send me 12 issues of ELECTRONICS NOW® for $17.97 and bill me. (Canada $25.65 U.S. FUNDS — includes G.S.T.)

Mail to: ELECTRONICS EXPERIMENTER'S HANDBOOK, Reader Service Management Dept.,
PO Box 5192, Pittsfield, MA 01203-9989

Name ______ Please Print

Company Name (if applicable) ______

Address ______

City ______ State ______ Zip +4 ______

Allow 6-8 weeks for delivery of first issue. OFFER VALID U.S. AND CANADA ONLY.

FREE READERS SERVICE: To receive more information about products and services mentioned in this issue, please circle the applicable Reader Service Number, and mail today!

EEH	11	21	31	41	51
1996	12	22	32	42	52
	13	23	33	43	53
	14	24	34	44	54
5	15	15	35	45	55
6	16	26	36	46	56
7	17	27	37	47	57
8	18	28	38	48	58
9	19	29	39	49	59
10	20	30	40	50	60

Additional Coupons on pages 15 and 108.

FREE

PROTOTYPING STATION

Continued from page 26

amplifier, a speaker, and a terminal strip. The signal generator circuit, shown in Fig. 12, consists of an astable 555 oscillator circuit, in which R1, in conjunction with C1 or C2, sets the output frequency. Switch S1 selects either the 0.01-µF capacitor that generates frequencies from 20 hertz to about 5 kilohertz, or the 1-µF capacitor that generates frequencies from 2 to 10 kilohertz. Potentiometer R3 controls gain or volume.

The audio amplifier circuit (Fig. 13) uses an LM 386 in a basic configuration. With just three components, a gain of about 20 is achieved; that can amplify the signal from a microphone, a Walkman-type radio, or any similar audio gear. The amplifier outputs about 500 milliwatts; if more power is needed, install an LM 380 amplifier IC that produces about 2 watts. Potentiometer R1 acts as a volume control by limiting the input voltage. The input and output to the amplifier are accessed from the terminal strip, as are the speaker and signal generator. This circuit is so simple that point-to-point wiring is used, as shown in Fig. 14.

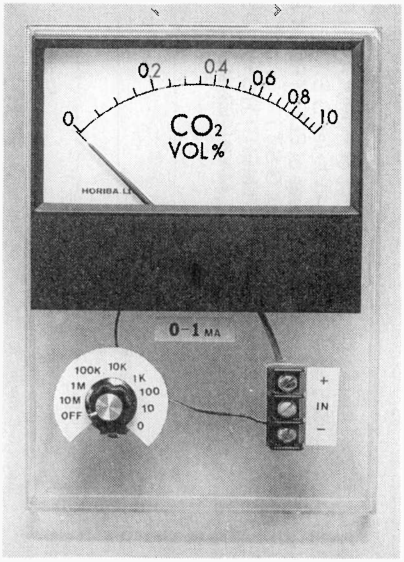

FIG. 18—THE METER MODULE contains a factory-made 0- to 1-milliampere movement.

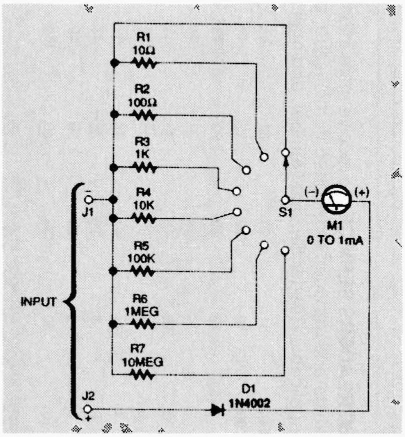

FIG. 19—AN 8-POSITION ROTARY SWITCH allows the input to be connected either directly to the negative meter pole, or through one of seven dropping resistors.

The three-digit decoder-driver shown in Fig. 15 adds another practical function block. The circuit, shown in Fig. 16, will interface with any standard BCD output to produce a digital display. The circuit contains a CD4553 three-digit BCD counter and a 74LS48 seven-segment decoder. The decoded data is routed to a three-digit, common-cathode display.

A logic probe is also included in the BCD decoder module. This simple circuit will indicate logic status, either high or low, at any given test point. Figure 17 is the schematic of the logic probe, which lights the red LED (LED2) for high logic levels, and the green LED (LED1) for low. If desired, common-anode, seven-segment displays can be substituted for the single LEDs. The cathodes of the b, c, e, f, and g segments of one display can be wired together to form an "H" to indicate a high, and the f, e, and d segments of the second display can be wired together to form an "L" to indicate a low.

The meter module shown in Fig. 18 makes use of a factory-marked 4½-inch, 0- to 1-milliampere movement, bought at surplus. As shown in Fig. 19, an 8-position rotary switch (S1) allows the input to be connected either directly to the negative meter pole, or through one of seven dropping resistors ranging in value from 10 ohms to 10 megohms. When using this meter, it is best to start with the highest setting, and then work down to prevent pinning the pointer. Ω

ELECTRONICS EXPERIMENTER'S handbook™

Gernsback Publications, Inc.
500 Bi-County Blvd.
Farmingdale, NY 11735-3931
1-(516) 293-3000
Larry Steckler, EHF/CET
President

Christina Estrada
assistant to the President

For Advertising ONLY
1-516-293-3000
Fax 1-516-293-3115

Larry Steckler
publisher

Arline Fishman
advertising director

Michele Torrillo
advertising assistant

Adria Coren
credit manager

Subscriber Customer Service
1-800-288-0652

Order Entry for New Subscribers
1-800-999-7139
7:00 AM - 6:00 PM M-F MST

ADVERTISING SALES OFFICES

EAST/SOUTHEAST
Stanley Levitan
Eastern Advertising
1 Overlook Ave.
Great Neck, NY 11021-3750
1-516-487-9357
Fax 1-516-487-8402

MIDWEST/Texas/Arkansas/Okla.
Ralph Bergen
Midwest Advertising
One Northfield Plaza, Suite 300
Northfield, IL 60093-1214
1-708-446-1444
Fax 1-708-559-0562

PACIFIC COAST
Blake Murphy
Pacific Advertising
Hutch Looney & Associates, Inc.
6310 San Vicente Blvd.
Suite 360
Los Angeles, CA 90048-5426
1-213-931-3444, Ext. 224
Fax 1-213-931-7309

Hobbyist's Paperback Budget Books

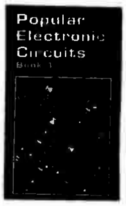

☐ **BP80/98—Popular Electronic Circuits-Books 1 and 2...$11.90.** Contains a wide range of circuits which are accompanied by text giving a brief introduction, circuit description and special notes on construction and setting-up that may be necessary.

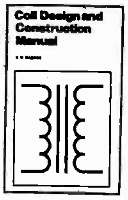

☐ **#160—Coil Design and Construction Manual...$5.95.** A complete book for the home builder on how to make RF, IF, audio and power chokes and transformers. Practically every possible type is discussed and calculations are explained in detail.

☐ **BP271—How to Expand, Modernize and Repair PC's and Compatibles...$7.75.** All the practical information you are likely to need to upgrade your PC and compatible. Also contains useful information and illustrations to help you with repairs.

☐ **BP276—Shortwave Superhet Receiver Construction...$6.95.** Provides construction details, including coil winding, of a number of advanced-design receivers which should have performance levels at least equal to commercial sets of similar complexity.

50 Projects Using IC CA3130

☐ **#223—50 Projects Using IC CA3130...$5.00.** One of the more practical and useful operational amplifiers (opamp), the CA3130 integrated circuit chip is the heart of several easy-to-assemble projects covered in the book. The projects are divided into five categories: audio projects, RF projects, test equipment, household projects and a catch-all miscellaneous group. Ideal for all skill levels.

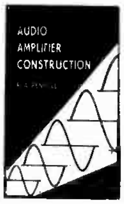

☐ **BP122—Audio Amplifier Construction...$5.75.** Practical designs are featured and include circuit diagram and description, Veroboard or printed-circuit board layout and construction notes. The text is divided into two parts. The first deals with many types of preamplifiers. The second covers power amplifiers from a simple low-power battery type to a 100-watt DC-coupled amplifier using four MOSFETs in the output stage.

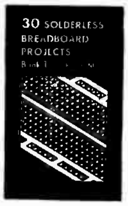

☐ **BP107—30 Solderless Breadboard Projects-Book 1...$5.95.** Each project is designed to be built on a "Verobloc" breadboard and is presented with a brief circuit description, circuit diagram, component layout diagram and components list. Notes on construction and applications are provided. Wherever possible, the components are common to several projects to keep project costs down.

☐ **BP266—Electronic Modules and Systems for Beginners...$7.25.** Shows the reader how to build a number of simple analog and digital circuit modules, all suitable for battery operation, and all based on only 1 or 2 transistors or ICs.

Electronic Music Learning Projects

☐ **BP329—Electronic Music Learning Projects...$6.25.** A beginner's book that will assist you in assembling such projects as the Elexylophone, Gildaphone, Melody Ranger, Cordmaker, Appealing Handbells, Electronic Sol-Fa, Tune-up Box and more.

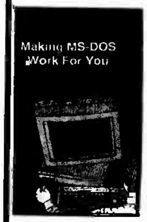

☐ **BP319—Making MS-DOS Work for You—$6.50.** What you need to know first, comes first, however the text is written in a circular style so that the reader can start anywhere in the text. This book is relevant to all versions of both MS-DOS and PC-DOS.

UNDERSTANDING PC SOFTWARE
SOFTWARE

☐ **BP303—Understanding PC Software...$6.95.** Covers main types of PC software, giving details of typical uses, the basics of how they are used and their limitations. Each chapter covers popular software and programs of a similar type.

Aerial Projects

☐ **BP105 Aerial (Antenna) Projects...$5.50.** In this book the author has considered practical antenna designs, including active, loop and ferrite antennas which perform well and are relatively simple and inexpensive to build. The complex theory and mathematics of antenna design have been avoided. Also included are construction details of a number of antenna accessories including a pre-selector, attenuator and filters.

Electronic Calculator Users Handbook

☐ **BP33—Electronic Calculator Users Handbook...$5.75.** The book eliminates the mysteries of the calculator and offers unique ideas for the simplest to most complex calculators. Covers the basic functions plus trigonometric, hyperbolic, logarithms, square roots and powers. Included are formulas and data for VAT, discounts, and mark up, currency conversion, interest, binary and octal numbers, and much more!

☐ **BP125—25 Simple Amateur Band Antennas...$5.50.** Plans to build antennas that are simple and inexpensive to construct and perform well. From the simple dipole to beam, triangle and even mini-rhombic types made from four TV masts and about 400 feet of wire. After the antenna discussion you will find a complete set of dimension tables that will help you "tune" an antenna on a particular frequency.

A Concise Introduction to dBASE

☐ **BP263—A Concise Introduction to dBASE...$6.95.** The most popular data-base program, dBASE, is difficult to master without help. This book helps you spend your time using dBase, not setting it up. A user's guide that covers all the basics up to dBase IV.

Electronics

☐ **PCP104—Electronics—Build and Learn...$9.95.** Construction details are given to build a circuit demonstrator that is used throughout the book to introduce common electronic components and how these components are built up to useful circuits.

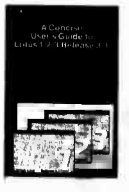

☐ **BP302—A Concise Users Guide to Lotus 1-2-3 Release 3.1...$6.25.** Written to help existing spreadsheet users upgrade to 1-2-3 Release 3.1, which has the ability to work 3-dimensionally with both multiple worksheets and files.

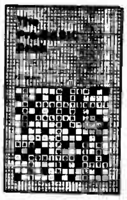

☐ **BP146—The Pre-BASIC Book...$6.95.** Another book on BASIC but with a difference. It concentrates on introducing technique by looking in depth at the most frequently used and more easily understood computer instructions.

Just What the Project Builder Is Looking For!

☐ **#160—Coil Design and Construction Manual**	**$5.95**
☐ **#219—Solid State Novelty Projects**	**$4.95**
☐ **#222—Solids State Shortwave Receivers for Beginners**	**$5.50**
☐ **#223—50 Projects Using IC CA3130**	**$5.00**
☐ **#225—Practical Introduction to Digital ICs**	**$5.25**
☐ **PCP111—Electronic Test Equipment Handbook**	**$13.95**
☐ **1596T—24 Silicon-Controlled Rectifier Projects**	**$9.95**
☐ **BP58—50 Circuits Using 7400 Series ICs**	**$5.50**
☐ **BP69—Electronic Games**	**$5.50**
☐ **BP71—Electronic House Projects**	**$5.00**
☐ **BP84—Digital IC Projects**	**$5.50**
☐ **BP93—Electronic Timer Projects**	**$5.50**

ELECTRONIC TECHNOLOGY TODAY INC.
P.O. BOX 240, Massapequa, NY 11762-0240

Name ____________________
Address ____________________
City__________ State ________ Zip__________
PE5 Allow 6-8 weeks for delivery

SHIPPING CHARGES IN USA AND CANADA

$0.01 to $5.00	$2.00
$5.01 to $10.00	$3.00
$10.01 to 20.00	$4.00
$20.01 to 30.00	$5.00
$30.01 to 40.00	$6.00
$40.01 to 50.00	$7.00
$50.01 and above	$8.50

SORRY No orders accepted outside of USA & Canada

Number of books ordered ☐

Total price of merchandise .... $________
Shipping (see chart) .... $________
Subtotal .... $________
Sales Tax (NYS only) .... $________
Total Enclosed .... $________

All payments must be in U.S. funds!

Electronics Workbench

Version 4.1

NEW WINDOWS 95/NT 32 BIT VERSION

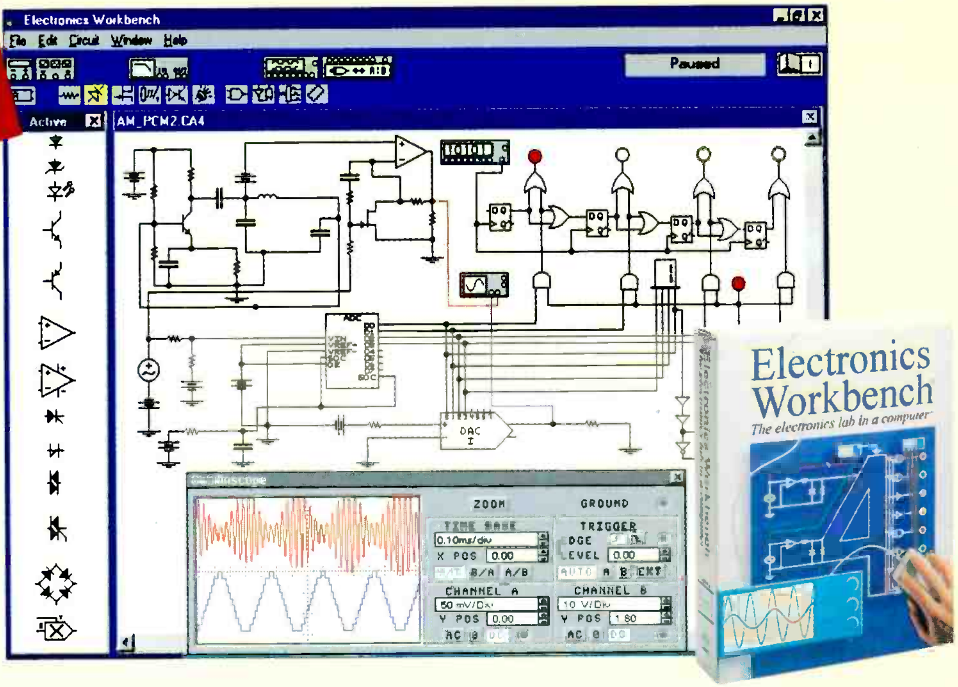

ZOOM GROUND
TIME BASE 0.10ms/div
X POS 0.00
TRIGGER EDGE
LEVEL 0.00
CHANNEL A 50 mV/Div
Y POS 0.00
CHANNEL B 10 V/Div
Y POS 1.80

Design and Verify Circuits Faster

Join over 20,000 customers using the affordable mixed-signal simulator

Design faster with Electronics Workbench. Mix the analog and digital components and ICs in any combination. And with a click of the mouse, try 'what if' scenarios and fine tune your designs. The built-in SPICE simulator gives you real-world waveforms.

All without programming or netlist syntax errors.

And in minutes. Not hours or days.

You'll be up and running sooner. And create better designs faster with Electronics Workbench. We guarantee it!

Electronics Workbench: $299

- Click & drag schematic capture
- Mixed analog/digital SPICE simulator
- Instant Bode plots and scrollable waveforms
- 50 analog components with 350 models
- 140 digital components and ICs in TTL and CMOS
- Windows 95/NT/3.1, DOS and Macintosh versions
- Free unlimited technical support
- 30-day money-back guarantee

Engineer's Pack: $599

- Electronics Workbench
- 2,450 models
- Import/Export SPICE netlists
- Export to PCB packages

To discover more about *the* affordable mixed-signal simulator, call us today at:

800-263-5552

Fax: 416-977-1818
Internet: http://www.interactiv.com

INTERACTIVE IMAGE TECHNOLOGIES LTD.
908 Niagara Falls Blvd. #068, North Tonawanda, NY 14120-2060
Telephone: 416-977-5550
BBS: 416-977-3540
CompuServe: 71333,3435
E-mail: ewb@interactiv.com

Australia: 2-519-3933 • Brazil: 11-453-5588 • Cyprus: 2-62-1068 • Czech Replubic: 42-48-23454 • Denmark: 33-25-0109 • Finland: 0-297-5033 • France: 14-908-9000 • Germany: 711-62-7740 • Greece: 1-524-9981 • Hungary: 1-215-0082 • India: 11-578-7055 Indonesia: 21-470-0815 • Israel: 3-647-5613 • Italy: 11-437-5549 • Japan: 3-3382-3136 • Malaysia: 3-774-2189 • Mexico: 5-593-5260 • Netherlands: 18031-7666 • New Zealand: 9-267-1756 • Norway: 22-16-70-45 • Philippines: 973-27-0118 • Portugal: 1-81-46609 Singapore: 777-2303 • Slovenia: 61-317-830 • South Africa: 331-6-8309 • South Korea: 2-2-222-3431 • Spain: 1-383-8335 • Sri Lanka: 1-86-5970 • Sweden: 8-740-5500 • Taiwan: 886-2-366-0080 • Thailand: 66-2-398-6952 • UAE: 4-453505 • UK: 1-203-23-3216

INTERACTIVE

9EA3FP9

Trademarks are property of their respective holders. Offer is in U.S. dollars and valid only in the United States and Canada. All orders subject to $15 shipping and handling charge.

CIRCLE 11 ON FREE INFORMATION CARD

Newnes

INCLUDES

FREE
NEWNES ONLINE
MEMBERSHIP

ELECTRONICS EXPLAINED

Lou Frenzel

Electronics Explained

Electronics Explained

The New Systems Approach to Learning Electronics

Louis E. Frenzel, Jr.

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Newnes is an imprint of Elsevier

Newnes

Newnes is an imprint of Elsevier
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

Copyright © 2010 Elsevier Inc. All rights reserved

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
Frenzel, Louis E.
Electronics explained : the new systems approach to learning electronics / Louis E. Frenzel, Jr.
p. cm.
ISBN 978-1-85617-700-9
1. Electronics—Textbooks. I. Title.
TK7816.F683 2010
621.381—dc22 2010006565

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

For information on all Newnes publications
visit our web site at www.elsevierdirect.com

Typeset by MPS Limited, a Macmillan Company, Chennai,
India www.macmillansolutions.com

Printed in the United States of America

10 11 12 13 14 10 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER BOOK AID International Sabre Foundation

With love to my one and only, Joan Ree.

List of Figures

FIGURE 1.1 Major sectors of the electronics industry and common applications. 4
FIGURE 1.2 General block diagram of how the electronics industry works from raw materials to end users. 6
FIGURE 1.3 Model of how all electronic circuits and equipment work. Input signals are processed by circuits and equipment into new output signals. 8
FIGURE 1.4 Model showing how a cell phone communicates with a cell site to make a call. 9
FIGURE 1.5 Block diagram representing any digital computer, be it a mainframe, PC, or embedded controller. 10
FIGURE 1.6 The most common form of robot emulates a human arm and hand to perform work automatically under the control of a computer. 12

FIGURE 2.1 Physics model of what a copper atom looks like. 17
FIGURE 2.2 Basic electrical circuit used to describe every electronic circuit. 19
FIGURE 2.3 Permanent bar magnet illustrating the magnetic lines of flux that surround it. 21
FIGURE 2.4 (A) Current in a wire produces a magnetic field around it. (B) Coiling the wire increases the strength of the magnetic field. 22
FIGURE 2.5 Relative motion between a magnetic field and a conductor such as a wire causes a voltage to be induced into the wire making it a voltage source. 23
FIGURE 2.6 DC voltages. (A) DC voltage source such as a battery or power supply. One side is usually grounded to serve as a reference. (B) Fixed, continuous positive DC voltage shown over time. (C) Negative DC voltage. (D) Varying positive DC voltage. (E) DC voltage pulses. 25
FIGURE 2.7 AC voltages. (A) AC voltage source drawn as a circle with a wave inside. The output polarity changes periodically. (B) Sine wave, the most common AC voltage shape. (C) Rectangular wave or AC pulses. (D) Triangular wave. (E) Random AC wave that may represent voice or video. 27

FIGURE 2.8 Basic concept of any battery. 29
FIGURE 2.9 Battery and cell symbols used in circuit diagrams. (A) Cell. (B) Four cells in series making a battery. Cell voltage add-up. (C) Symbol for battery of any number of cells. (D) Cells in parallel to increase current capability. 30
FIGURE 2.10 Panel comprised of solar cells in series and parallel make up DC source that charges the battery in a spacecraft. 32
FIGURE 2.11 Frequency-domain views. (A) 1-MHz sine wave. (B) 1-MHz square wave. 35
FIGURE 2.12 A low cost digital multimeter that measures DC and AC voltages, current, and resistance. 36

FIGURE 3.1 Hierarchy of electronics from systems to components. 41
FIGURE 3.2 System block diagram of typical personal computer. 42
FIGURE 3.3 System block diagram of iPod or MP3 music player. 43
FIGURE 3.4 Switch and how it is used. 44
FIGURE 3.5 Types of resistors and schematic symbol. 45
FIGURE 3.6 Standard resistor color code. 46
FIGURE 3.7 Voltage divider, the most common resistor circuit. 46
FIGURE 3.8 Capacitor construction and schematic symbol. 47
FIGURE 3.9 (A) Charging capacitor. (B) Discharging capacitor. 47
FIGURE 3.10 Low-pass filter, a common capacitor use. 48
FIGURE 3.11 How capacitor passes AC sine wave but blocks DC voltage. 49
FIGURE 3.12 Schematic symbol for inductor. 50
FIGURE 3.13 Schematic symbol for transformer. 50
FIGURE 3.14 Diode, its schematic symbol, and how to bias it for conduction or cut-off. 51
FIGURE 3.15 How a diode rectifies AC into DC. 51
FIGURE 3.16 Schematic symbol for bipolar junction transistor (BJT) and its inputs and outputs. 52
FIGURE 3.17 Schematic symbol for MOSFET and its input and outputs. 52
FIGURE 3.18 Simple BJT amplifier and how it works. 53
FIGURE 3.19 Simple MOSFET switch and how it works. 54
FIGURE 3.20 Typical integrated circuit today. This one is a dual-satellite TV tuner for set-top boxes. The size is only 7 × 7 mm. The pins are soldered to a printed circuit board. (*Courtesy NXP Semiconductor.*) 54
FIGURE 3.21 Common breadboarding socket for building prototypes or just experimenting. (*Courtesy Global Specialties.*) 56
FIGURE 3.22 Measuring battery voltage with multimeter. 56
FIGURE 3.23 This is the circuit used to light an LED and how to wire it on the breadboard. 57

FIGURE 3.24 Circuits for charging and discharging capacitor and respective wiring. 59

FIGURE 3.25 Circuit that turns an LED off and on with MOSFET and wiring. 60

FIGURE 4.1 Linear circuit has a straight-line response. As input varies, output varies by factor of 10 larger, producing straight-line output. 62

FIGURE 4.2 Block symbol for amplifier. Gain, A, is often given. DC power inputs are not usually shown. 62

FIGURE 4.3 Single-ended (A) and differential input (B) amplifiers. Inputs and outputs are referenced to ground. 64

FIGURE 4.4 Class A amplifier conducts continuously. 64

FIGURE 4.5 Class B amplifier conducts half the time, making it more efficient. 65

FIGURE 4.6 Class C amplifier only conducts for part of a half cycle, but the current pulse stimulates resonant tank circuit that oscillates at desired frequency. 66

FIGURE 4.7 Class D amplifier uses PWM at higher frequency to switch higher voltages for greater output. Class D amplifiers are more efficient than any other type, and thus do not produce as much heat. 67

FIGURE 4.8 Most commonly used op amp circuit configurations. (A) Inverting amplifier. (B) Noninverting amplifier. (C) Follower. (D) Differential amplifier. 68

FIGURE 4.9 Instrumentation amplifier widely used to amplify small sensor signals in industrial applications. 69

FIGURE 4.10 Amplifier input impedance (A) and output impedance (B). 70

FIGURE 4.11 When one amplifier feeds another, input and output impedances form a voltage divider that attenuates the signal and offsets some of the gain. 71

FIGURE 4.12 Cascading amplifier provides more gain. Total gain is product of individual gains or sum of dB gains. 72

FIGURE 4.13 Frequency-response curves show output versus input frequency. Curve in (A) is response of op amp that extends from DC to 5-MHz cut-off frequency. Curve in (B) is response of common stereo audio amplifier. 73

FIGURE 4.14 Four common filter types and their response curves and symbols. (A) Low pass. (B) High pass. (C) Band pass. (D) Band reject or notch. 74

FIGURE 4.15 Types of oscillators and their symbols. (A) Sine wave. (B) Rectangular pulse. (C) VCO. (D) Crystal clock oscillator. 76

FIGURE 4.16 Linear mixing with resistors (A) or an op amp summer (B). 77
FIGURE 4.17 Non-linear mixer used to generate higher and low frequencies. 77
FIGURE 4.18 Phase detector converts phase shift into proportional DC average. 78
FIGURE 4.19 Phase-locked loop. 79
FIGURE 4.20 Frequency synthesizer using phase-locked loop. Changing the divider ratio changes the frequency. 81
FIGURE 4.21 Typical power supply configuration using bus architecture, regulators, and DC–DC converters to get desired number of outputs. reg, regulator. 83
FIGURE 4.22 Concept of duty cycle and how varying pulse widths can be filtered into a proportional DC average voltage. avg, average. 85
FIGURE 4.23 Common DC-to-AC solar power system. 87

FIGURE 5.1 Binary signal that represents 0 and 1 as voltage levels. 89
FIGURE 5.2 Off/on switches are used to enter binary data into digital circuits. 91
FIGURE 5.3 How lights or LEDs are used to represent binary data. 92
FIGURE 5.4 Parallel binary data transfer over bus. 98
FIGURE 5.5 Serial data transfer where 1 bit occurs after another in time. 99
FIGURE 5.6 Logic symbols for inverter and truth table. 101
FIGURE 5.7 Logic symbols for AND gate and truth table. 102
FIGURE 5.8 How AND gate is used to turn signal off or on. (A) The AND gate showing the input and output names. (B) The waveforms of the inputs and output. 103
FIGURE 5.9 Logic symbols for OR gate and truth table. 103
FIGURE 5.10 Logic symbols for NAND gate and truth table. 104
FIGURE 5.11 Logic symbols for NOR gate and truth table. 105
FIGURE 5.12 Logic symbols for XOR and XNOR gates and truth tables. 105
FIGURE 5.13 Logic symbols for flip-flops. (A) RS. (B) D type. (C) JK. 106
FIGURE 5.14 Toggling T input on JK flip-flop causes it to change states, producing frequency division by 2. 106
FIGURE 5.15 Cascading JK flip-flops produce higher-frequency division ratios, which is also a binary counter that counts from 0000 to 1111. 107
FIGURE 5.16 Storage register from one word consists of FFs. It also has clock/load and clear/reset inputs. 107
FIGURE 5.17 (A) Shifting serial data into shift register. (B) Shifting serial data out of shift register. 108
FIGURE 5.18 Shift registers used for (A) serial-to-parallel data conversion and (B) parallel-to-serial data conversion. 109

FIGURE 5.19 Counter storing binary value equal to number of input pulses that occurred. 110
FIGURE 5.20 Digital multiplexer (mux) that selects one of four inputs to appear at the single output. 111
FIGURE 5.21 Decoders. (A) 1-AND gate decoder that identifies input 0110. (B) 4-to-16 decoder that identifies 1 of 16 different 4-bit input values. 112
FIGURE 5.22 Basic concept for random access memory (RAM). Address selects one of eight storage locations. Then data may be read from that location or data stored in it. 113
FIGURE 5.23 RAM storage cells. (A) RS FF for SRAM. (B) Capacitor cell for DRAM. 114
FIGURE 5.24 Storage cell used in EEPROM and flash memories is MOSFET with a special floating gate that keeps transistor off or on. 115
FIGURE 5.25 PLD concept. (A) Block concept of programmable AND and OR gate arrays. (B) Simple circuit of PLD showing programming by dots on AND and OR gate arrays. 117
FIGURE 5.26 Analog-to-digital conversion is performed by sampling analog signal at equal intervals and generating proportional binary value. 119
FIGURE 5.27 Symbol for ADC with both parallel and serial output examples. 120
FIGURE 5.28 Converting binary data back into analog signal with DAC. 120
FIGURE 5.29 DAC output is stepped approximation to original analog signal. 121

FIGURE 6.1 How a micro controls external devices. 125
FIGURE 6.2 How a micro monitors external devices. 125
FIGURE 6.3 Controlling liquid level in a tank automatically with a micro. 126
FIGURE 6.4 General block diagram of any digital computer or microcontroller. 127
FIGURE 6.5 How data is manipulated in a register. Examples of shifting data, incrementing and decrementing a register, and clearing (resetting) a register to 0. 129
FIGURE 6.6 How ALU adds or otherwise processes two data values. 130
FIGURE 6.7 Program counter identifies the address in memory to be accessed. 132
FIGURE 6.8 Three common instruction word formats with addresses or data. 133

FIGURE 6.9 Example of program stored in RAM for execution using various CPU registers. 134
FIGURE 6.10 Only one source may transmit at a time on a bus. Others are disconnected. All destinations can receive the transmitted data. Dest, destination. 136
FIGURE 6.11 General block diagram of 8051 microcontroller. It is very similar to most other controllers in terms of major circuits and their interconnections. 138
FIGURE 6.12 ASCII letter "j" being sent over an RS-232 interface using a UART. 140
FIGURE 6.13 Basic make-up of digital signal processor and its ADC and DAC I/O. 144

FIGURE 7.1 Model of electronic communications system. 148
FIGURE 7.2 Ways of communicating. 149
FIGURE 7.3 How an antenna creates a radio wave. 152
FIGURE 7.4 Electromagnetic frequency spectrum. 154
FIGURE 7.5 Radio propagation of shortwaves. 157
FIGURE 7.6 Line-of-sight (LOS) communications at VHF, UHF, and microwave. 158
FIGURE 7.7 Amplitude modulation. 160
FIGURE 7.8 Sidebands and bandwidth. 161
FIGURE 7.9 Frequency modulation. 162
FIGURE 7.10 Types of digital modulation. (A) ASK. (B) OOK. (C) FSK. (D) BPSK. 163
FIGURE 7.11 Radio transmitter. 165
FIGURE 7.12 Superheterodyne receiver. LO, local oscillator. 167
FIGURE 7.13 Single-chip UHF transceiver for low-speed, short-range digital data transmissions. 169
FIGURE 7.14 Quarter-wave ground-plane antenna. 171
FIGURE 7.15 Dipole antenna. 172
FIGURE 7.16 Yagi antenna. 173
FIGURE 7.17 Printed circuit board (PCB) patch antenna. 174
FIGURE 7.18 Coaxial cable transmission line. 174
FIGURE 7.19 Microwave horn antenna fed by waveguide transmission line. 174
FIGURE 7.20 How a parabolic dish works. 175
FIGURE 7.21 Concept of communications satellite. 178
FIGURE 7.22 Satellite transponder. LO, local oscillator. 180

FIGURE 8.1 (A) Cell coverage of a given area. (B) Cell area covered depends on antenna pattern and environment. 185
FIGURE 8.2 How small and large cells cover an area. 186
FIGURE 8.3 Cell phone basic operations. 187

FIGURE 8.4 Block diagram of basic cell phone. 188
FIGURE 8.5 How QPSK is generated using mixers and a local oscillator shifted 90 degrees to produce in-phase (I) and quadrature (Q) signals that are summed. 192
FIGURE 8.6 QPSK diagrams. (A) Phasor. (B) Constellation. 192
FIGURE 8.7 How 16-QAM is generated and resulting constellation diagram. 193
FIGURE 8.8 OFDM divides a channel into many subchannels or subcarriers, each of which is modulated by part of digital data to be transmitted. 194
FIGURE 8.9 Generating OFDM is done by DSP and I/Q modulation. 195
FIGURE 8.10 MIMO in a 4 × 4 configuration quadruples the data rate in an existing channel, while reducing the ill effects of multipath fading. 196
FIGURE 8.11 Frequency-division multiple access and frequency duplexing use two separate bands of frequencies divided into individual channels. 198
FIGURE 8.12 TDMA multiplexing divides a channel into sequential time slots for the digital data. 199
FIGURE 8.13 How CDMA is generated. 200
FIGURE 8.14 A block diagram of 3G cell phone with multiple radios and functions. 201
FIGURE 8.15 Concept of femtocell. 203

FIGURE 9.1 How LANs, MANs, and WANs connect to form a huge worldwide network of computers and other devices. 209
FIGURE 9.2 Three main networking topologies—star, ring, and bus. Bus is the most widely used. 210
FIGURE 9.3 Coax cable. It is now mainly used in cable TV networks. The most common is RG-6/U and uses F-type connectors. 211
FIGURE 9.4 Twisted-pair cable. Initially used for telephones and is still used. Computer network cables use four twisted pairs. A shielded version is available where noise is a problem. 212
FIGURE 9.5 Fiber optic cable. It is nonconductive because it is made of plastic and glass, and is used to transmit data as infrared (IR) light pulses. 212
FIGURE 9.6 How binary data is transmitted over fiber optic cable. 214
FIGURE 9.7 Ethernet hub or switch implements a bus that is accessed by RJ-45 connectors and twisted-pair cable. 215
FIGURE 9.8 Ethernet frame or packet. 215
FIGURE 9.9 A T-1 data frame commonly used in digital telephone systems as well as data interconnections in networks. 217

FIGURE 9.10 Internet illustrated as cloud of fiber backbone connections through network access points (NAP) that attach to WANs, MANs, and LANs. 218
FIGURE 9.11 Protocols used in Internet. (A) Transmission control protocol (TCP). (B) Internet protocol (IP). The data in each frame is transmitted from left to right and top to bottom. 219
FIGURE 9.12 Summary of short-range wireless networking technologies shown as data rate versus range. 221
FIGURE 9.13 How a Wi-Fi access point or hot spot works. 221
FIGURE 9.14 A mesh network can extend the overall range of a node and provide greater reliability through multiple possible paths. Most wireless mesh networks are ZigBee, but other wireless technologies may be used (e.g., Wi-Fi). 224
FIGURE 9.15 How RFID works. The tag has no power of its own but gets power from the transmitter in the reader. 225

FIGURE 10.1 Cross-section of speaker showing major components and how it works. 230
FIGURE 10.2 Cross-over networks keep signals of different frequencies separate for appropriate speakers. 231

FIGURE 10.3 How audio signal is digitized. In this example, each sample is assigned a 4-bit code. The stepped approximation shown is what a DAC would deliver to the speakers. 234
FIGURE 10.4 Voice signals up to 4 kHz for the telephone are sampled at an 8-kHz rate, producing an 8-bit sample, each 125 microseconds. Stringing the samples together in a serial data stream produces a digital signal at a 64-kHz rate. 236
FIGURE 10.5 General block diagram of a digital music player like Zune MP3 or Apple iPod. 237
FIGURE 10.6 Digital music on a CD is encoded as pits and lands on the CD surface. These are burned in by a laser. A laser light shined on the pits and lands produces a reflection that a photodetector picks up to create the original digital data. 238
FIGURE 10.7 Block diagram of CD player. 240
FIGURE 10.8 Block diagram of typical AV receiver, the heart of the audio reproduction part of a consumer home entertainment center. 241
FIGURE 10.9 How 5.1 surround sound system is set up. 243

FIGURE 11.1 How scanning converts a scene of light and color into an electronic signal. 249
FIGURE 11.2 TV screen size and aspect ratio. 249
FIGURE 11.3 Interlaced scanning helps reduce flicker. Sync pulses help keep the receiver and transmitter in step with one another. 251
FIGURE 11.4 Digital TV is created by digitizing the analog video signals with an ADC, and then converting it into a serial bitstream of video data words. 252
FIGURE 11.5 TV screens can be thought of as thousands of rows and columns of pixels, each represented by data. 253
FIGURE 11.6 Any color of light can be reproduced by mixing red, green, and blue light in the right proportion. 256
FIGURE 11.7 Red, green, and blue digital color signals are generated by filtering the scene into its red, green, and blue components, and then digitizing the resulting analog signals. 257
FIGURE 11.8 HDTV packet for ATSC U.S. digital color TV system. 258
FIGURE 11.9 General block diagram of TV transmitter. 258
FIGURE 11.10 U.S. ATSC system translates the digital bitstream into an eight-level signal to transmit multiple bits per level, making higher data rates in a given bandwidth possible. 259
FIGURE 11.11 Simplified block diagram of digital TV set. 261
FIGURE 11.12 Inputs and outputs of modern digital TV set. 262
FIGURE 11.13 How color pixels are formed on a TV screen. (A) Pixel format for most CRTs. (B) Pixel format typical of LCD, plasma, or LED screens. 263
FIGURE 11.14 Coax cable bandwidth up to 1 GHz can contain about 130 6-MHz channels. 266
FIGURE 11.15 General block diagram of modern cable TV system. 267
FIGURE 11.16 Simplified block diagram of modern cable box with DVR. 270
FIGURE 11.17 TV distribution satellites in geosynchronous orbits act as wireless relay stations. 271
FIGURE 11.18 Video from all sources is transmitted up-link to satellite that rebroadcasts it back to Earth over down-link to home receivers. 271
FIGURE 11.19 Satellite TV home system consists of antenna with low-noise block converter, receiver, and TV set. 272
FIGURE 11.20 Simplified block diagram of satellite TV receiver. 273
FIGURE 11.21 Block diagram of standard DVD player. 275

FIGURE 12.1 General block diagram of industrial control process. 278

FIGURE 12.2 Open-loop control system for filling vat with liquid via a pump. 278
FIGURE 12.3 Generic diagram of any closed-loop control system with feedback. 279
FIGURE 12.4 Filling vat with liquid automatically with closed-loop control system. 280
FIGURE 12.5 Closed-loop control system for motor speed control and regulation. 281
FIGURE 12.6 Most resistive sensors like an RTD or thermistor are placed in a bridge circuit to condition them to provide amplified output voltage. 282
FIGURE 12.7 Characteristic curve of thermistor and simple voltage divider circuit producing output voltage. 283
FIGURE 12.8 Strain gauge is a resistive element whose value changes with stress or pressure. 284
FIGURE 12.9 Schematic symbols for photo cells. (A) Photoconductive or photoresistive. (B) Photovoltaic or solar cell. (C) Photo transistor. 285
FIGURE 12.10 Photodiode and its simple output circuit. 286
FIGURE 12.11 Simple off/on photo control circuit. Light turns on automatically at dark and off when light is present. 287
FIGURE 12.12 Simple tachometer circuit using phototransistor to create pulses that are averaged into a DC voltage proportional to speed. 288
FIGURE 12.13 Cross-sectional view of typical relay showing major components and schematic diagram. 290
FIGURE 12.14 Common relay contacts. (A) Form A. (B) Form B. (C) Form C. 291
FIGURE 12.15 Reed relay. 291
FIGURE 12.16 Controlling a large current with a smaller current with a relay. 292
FIGURE 12.17 Latching a relay with its own contacts. 294
FIGURE 12.18 Operating a relay with a transistor is a common application. The diode protects the transistor when power is removed from the relay coil. 295
FIGURE 12.19 A solenoid produces linear motion using magnetic principles. 296
FIGURE 12.20 Using an H-bridge circuit to reverse the current flow and direction of rotation in a DC motor. 297
FIGURE 12.21 Schematic symbol of an SCR. 299
FIGURE 12.22 Circuit that shows how an SCR is turned on and off. 300
FIGURE 12.23 Using an SCR to control power to a load using AC. The SCR rectifies the AC. 301
FIGURE 12.24 Schematic symbol for triac. 302

FIGURE 12.25 Equivalent circuit of triac shown with SCRs. 302
FIGURE 12.26 Lamp dimmer circuit using RC phase shifter and diac to control current in the load. 303
FIGURE 12.27 AC waveforms in a lamp load controlled by a triac. (A) Waveforms with triggering at 60 degrees. (B) Waveforms with triggering at 120 degrees. The longer pulses produce higher average current in the load. 304
FIGURE 12.28 General block diagram of PLC. 305
FIGURE 12.29 PLC operational scan cycle. 306
FIGURE 12.30 Sample ladder diagram and PLC application program. 307

Preface

This book is for those of you who want to learn about electronics and computers but frankly are put off by all the gory details. If you have a short attention span, a nonexistent background in science, math, and technology, and are looking for near-instant gratification of electronics knowledge, this book is for you. Electronics is a complex subject, especially if you want to be an electronics engineer. However, if you are merely curious or just need a working knowledge of electronics for your job, this book will give you a basic education on the fundamentals without boring you to death or loading you down with the excruciating details that other books lay on you. What I have attempted to do is write a book that tells you the most important concepts and fundamentals but leaves out the minutiae that only an engineer will need or appreciate. This book contains exactly what you need to know today in the 21st century.

There are lots of basic electronics books around and you may already have read some. This one is really different. First, it does not go into detailed circuit analysis. Why? Simply, all electronic equipment today is made up of integrated circuits or chips, and we can't see the circuits, access them, or even repair them. If you are an electronic engineer designing integrated circuits, yes, you do need to understand their operation. If you just want to learn how electronic equipment works, then you can skip this detail. This book treats circuits as simple functional building blocks whose purpose and application are easily understood.

Second, this book presents what I call the NEW electronics. Most basic electronic books dwell on the older components and circuits and ignore how electronics really is today. These books do a good job teaching you the history of electronics but little else. What you will see in this book is a fresh new way to look at and understand electronics, especially in the context of the familiar applications you use every day made with the most advanced integrated circuits. If you complete this book and still want more, one of the more traditional basic electronics books would be your next step. See the appendix for my recommendations.

The first three chapters introduce you to the electrical principles that form the basis of electronics. You will learn the core concept of electronics, which boils down to how to generate current flow and then control it to do something useful. Magnetic fields are also covered, since magnetism is just as important in electronics as current flow. But don't worry, even if you never took a physics course you can easily get through this material.

Next, you learn about the basic components of electronics such as resistors, capacitors, inductors, transformers, diodes, transistors, and especially integrated circuits. Then you will discover all the different types of electronic circuits used to make up any electronic device. The functional block diagram approach is used here so you won't have to do the nitty-gritty circuit analysis that has you chasing electrons through components and solving simultaneous equations.

An important part of this book is the coverage of computers, and specifically of embedded controllers, that is, those tiny single-chip microcomputers or microcontrollers that are built into virtually every electronic device. Digital signal processors (DSPs) are also covered since they also appear in most modern electronic equipment. You can't name an electronic product today that does not have one of these embedded computers in it. You really need to know what these are and how they work, as they are the very heart of electronic devices today.

The rest of the book consists of descriptions of how all the major electronic products and systems work. The entire range of modern electronics is covered, including major things like cell phones, TV (cable, HDTV, DVD, VCR, etc.), audio (stereo, CDs, MP3, etc.), satellites (GPS, TV, radio, etc.), radio (broadcast, marine-aircraft-shortwave, ham, CB, family radio, etc.), and everything else you can think of. And it is all explained using basic building blocks rather than detailed circuits.

Finally, if you want involvement as you read the book, get involved with the projects that I list in the Project sidebars in each chapter. These can lead to a deeper understanding of the concepts. Furthermore, these projects could encourage you to become a true electronic hobbyist … or is that a geek or nerd?

When you finish this book you will not be an electronic engineer nor will you be able to design or analyze circuits. But you will know about all the different kinds of circuits, how they work, and how they all fit together to create electronic equipment. The benefits to you are the ability to apply, use, select, operate, and to some extent troubleshoot most of the common electronics products with confidence, while appreciating their complexity and the value they deliver. Hope you like it.

Lou Frenzel
Austin, TX

Chapter 1

Learning about Electronics for Your Own Good

A Little Bit of Perspective before You Get to the Details

In this Chapter:

- Why it is helpful to know electronics.
- The 10,000-foot view of the electronics industry and how it works.
- The systems view of electronics.

INTRODUCTION

You don't know how lucky you are. You live in high-tech heaven surrounded by lots of useful, entertaining, interesting, necessary, and even addicting electronic products, products you cling to and would not want to give up.

But do you have any idea how many electronic products you own? Do you really know how much electronics influences your life every day? Some of you do, but most of you probably do not because your electronic gadgets are pretty much taken for granted. Yet we spend the better part of our lives working with these devices. Most of us sit in front of a computer all day at work, slouch on the couch in front of the TV at night, listen to our iPods going to and from work, and spend a weekend downloading videos off of the Internet or taking photos with a digital camera. And let's be clear here, how many text messages did you send today or cell calls did you make? The impact on our lives of all this electronic equipment is almost overwhelming. In this first chapter, I encourage you to think about how electronics impacts your life.

And that brings us to the subject of this book, learning how electronics works. If you want a first book in electronics that gives you the big picture about how electronic equipment works, this is it. I am not going to beat you to death with a lot of complex circuits and theory or math, only what you really need to know. Then I will go on to show you how each and every one of the products you use every day actually works.

© 2010 Elsevier Inc. All rights reserved.

doi: 10.1016/B978-1-85617-700-9.00001-1

IT ISN'T LIKE IT USED TO BE

It used to be that you learned electronics by playing around with resistors, capacitors, transistors, and other devices. You wired up simple circuits, built a one- or two-transistor radio, or even built a kit. But today electronics is so different that it is hard to do that anymore. Oh yes, you can still build little circuits and radios and other gadgets, but that is not how electronic products are made today. Electronic equipment today is made with integrated circuits or ICs that we also call chips. These chips are tightly packed on printed circuit boards with resistors and capacitors that you can barely see. It is more difficult than ever to experiment with these devices, much less troubleshoot and repair them. We are in that age of electronics where unless it is a big pricey piece of equipment, you throw a defective electronic product away and simply buy a newer and better one at a lower price.

But despite the complexity of today's electronic products, there is something just fascinating about them. Many of you actually want to know how they work. And despite their complexity, these products are actually understandable. It is possible to learn how these devices work without a whole lot of trouble. And you don't have to learn all the math and physics required by engineers to do so. This book takes the basic ideas of electronics and reduces them down to the lowest common denominator and shows you how to use those fundamentals and apply them to all of the fabulous electronic gadgets you have today.

WHAT'S IN IT FOR YOU?

There are some big benefits to being electronic literate. Take a look at some of them.

Satisfy Your Curiosity—Surrounded by so much electronics, you often probably wondered just how all this stuff works. For some of us, not knowing how something works nags at us until we satisfy that curiosity. If you have an interest in electronics, you tend to want to know more about it. This book will get you started in the right direction.

Enhance Your Hobby—If you dabble in a hobby that uses electronics, you obviously want and need to know more about it. Lots of hobbies involve electronics such as radio-controlled models, electric trains, amateur radio, shortwave listening, citizens or family radio, audio systems, or video. Or maybe you just like playing around with your home security system, garage door opener, or your high-speed Internet connection. A major part of any hobby is learning more about it and that is very satisfying. And don't forget—electronics itself can be your hobby. You can spend your time learning more about it and then digging deeper into those things that you want to know more about. Amateur radio is certainly one of those electronic hobbies that will take you from one extreme to the next in the radio spectrum. Anything having to do with computers is also a great hobby.

With everything being controlled by an embedded microcomputer, spending more time learning about them is a great activity with nearly endless possibilities.

Job-Related Needs—You may actually be working in a job surrounded by electronic equipment of some sort, yet you do not know how it works. Often, just understanding some of the fundamentals will give you a better grasp of how to use that equipment or even troubleshoot and repair it. Understanding some of the concepts may actually help you to better select and purchase equipment that you may use.

Boost Your School Performance—For those of you going to a college or university to learn electronic engineering or technology, you regularly have your curiosity stimulated by the classes you take. Yet most of the classes are heavily involved with theory, physics, math, and circuit analysis. What you really don't get in school is the perspective or big picture as to how the electronic products actually work. This book will give you the big picture and show you how it all comes together so that all those detailed classes in circuits and theory will make more sense.

Just so You Won't Appear to Be So Dumb—With electronics everywhere, people often talk about it without really knowing what they are talking about. It is amazing the ignorance of people who use the electronics but don't have a clue as to what it does or how it does it. For instance, do you really understand what an MP3 is, or why BluRay DVDs are better, or what a 3G cell phone really is? The answers are in this book. If you want to appear a little smarter to your co-workers, friends, and relatives, this book can help out. You will be able to deal with common questions and misconceptions, such as can you get shocked by a car battery, can you really get brain cancer or pop popcorn with your cell phone, or will sitting too close to your TV set make you sterile? While this book certainly won't make you a genius, it will indeed give you the basics so that you won't appear to be an idiot when technical questions about electronics come up.

ELECTRONICS: THE BIG PICTURE

Before we get into the nitty-gritty of how electronic equipment works, you really need to know the entire scope of electronics. It is a huge industry and recent figures seem to show that it is one of the biggest, if not the biggest, markets in the entire world. To get a grasp of how big the industry is, we normally break it down into segments of specializations. There are five major fields of electronics: communications, computers, control, instrumentation and measurement, and components. Any electronic product or device will easily fall into one of those categories. In some cases, the product may overlap two or more or all of those categories. Take a look at Figure 1.1, which shows each of these segments of electronics and the products and technologies that are part of them.

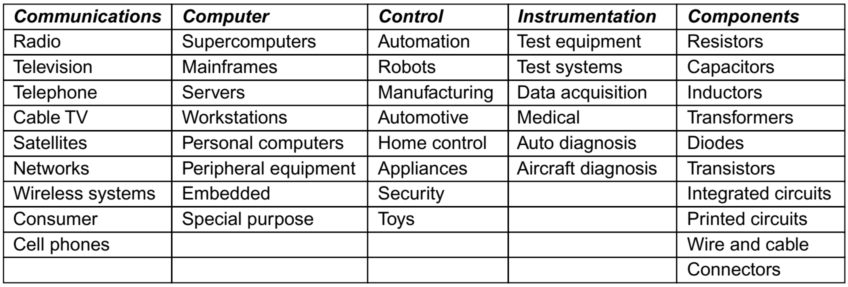

Communications	**Computer**	**Control**	**Instrumentation**	**Components**
Radio	Supercomputers	Automation	Test equipment	Resistors
Television	Mainframes	Robots	Test systems	Capacitors
Telephone	Servers	Manufacturing	Data acquisition	Inductors
Cable TV	Workstations	Automotive	Medical	Transformers
Satellites	Personal computers	Home control	Auto diagnosis	Diodes
Networks	Peripheral equipment	Appliances	Aircraft diagnosis	Transistors
Wireless systems	Embedded	Security		Integrated circuits
Consumer	Special purpose	Toys		Printed circuits
Cell phones				Wire and cable
				Connectors

FIGURE 1.1 Major sectors of the electronics industry and common applications.

Communications

Communications is the oldest segment of electronics and still the largest. Electronics actually started with communications–specifically telegraph, telephone, and radio–and grew from there. A lot of this book is about communications because it is such a huge segment of the industry and one that we use the most. And don't forget, communications comes in both wired and wireless forms. Wired systems include the vast telephone network, cable TV systems, and computer networks, including the Internet. Wireless systems comprise all of the radio, cell phone, television, satellite, radar, and other wireless systems.

Computers

Computers have only been around since the late 1940s and early 1950s but what an impact they have had on the way we work. Today virtually everyone uses and even owns a personal computer. And then there are all the larger and more powerful computers such as the servers that manage all of our networks, the mainframes still used by government and big business to manage huge databases, as well as the high-speed supercomputers that still do an amazing amount of scientific, engineering, and mathematical design and research. But the computer we are going to talk about most in this book is the embedded controller. These are small single-chip digital computers called *microcomputers* or just *micros* that are literally part of every electronic product. These are miniature digital computers dedicated to a specific function inside the products in which they exist.

Control

Control is a broad general term for monitoring and control. Monitoring, of course, means sensing various physical characteristics such as temperature, humidity, physical position, motor shaft speed, or light level. Electronic devices called *sensors* or *transducers* convert these physical characteristics into electrical signals which allow us to monitor them. Control refers to managing and exercising some degree of influence over items such as motors, lights, relays, heating elements, and other devices. Most monitoring and control functions take place in factories, chemical plants, refineries, and food processing operations. Control also occurs in the home, an example being your electronic thermostat. It also shows up in electromechanical products like a DVD player. Electronic controls appear in appliances, home control systems, security systems, automobiles, and even toys. Robots are another major segment of the control industry.

Instrumentation

Instrumentation refers to the field involved with testing and measuring electronic equipment and other mechanical or electronic items. It refers to the test instruments like volt meters, oscilloscopes, and spectrum analyzers as well as the large automated test systems used for mass testing and measuring of other devices. Instrumentation refers to data acquisition systems, medical tests and measurements, and a wide range of other products.

Components

All of the electronic products that make up each of the four major segments of the electronic industry are made up of various electronic components. In the past, electronic products were made of the individual discrete components such as resistors, capacitors, diodes, and transistors. Today most electronic equipment is made with one or more integrated circuits surrounded by a sprinkling of those other so-called *discrete* components. No longer can we access most individual components or circuits since they are sealed within integrated circuits or chips. Just keep in mind that there is a whole industry centered on making these electronic components that are in turn made available to the manufacturers of the end electronic equipment in each of the four major categories just described.

HOW THE ELECTRONICS INDUSTRY WORKS

As part of the big picture of learning electronics, you really need to understand how the industry itself works. This is summed up by the simplified block diagram in Figure 1.2. It all starts with the raw materials that the component manufacturers use to create the individual components. For example, sand is refined into silicon, which, in turn, is converted into wafers that are processed into the integrated circuits we use.

The electronic equipment manufacturers design the electronic equipment for the marketplace. It may be a TV set, a cell phone, or a military radar. These companies buy the electronic components and put them together to create the

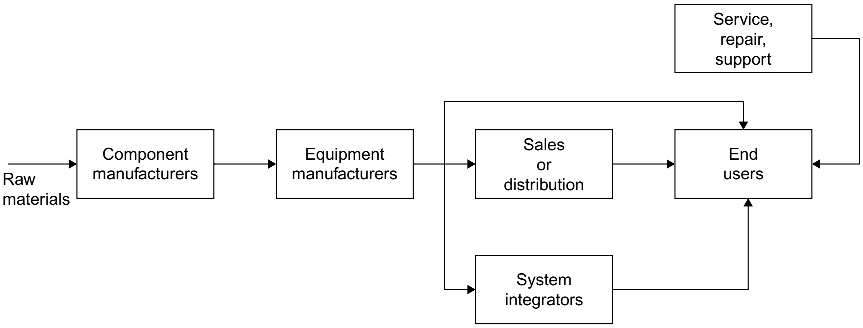

FIGURE 1.2 General block diagram of how the electronics industry works from raw materials to end users.

end product. Manufacturers may then sell the end product directly to customers and end users. In some cases, the end product will go through a sales organization or distributor of some sort. Your LCD large-screen TV set doesn't come directly to you directly from Samsung in Korea or Sony in Japan. It is sold to an organization such as Best Buy or Circuit City, which is where you buy it. Sometimes the electronic equipment is sold to so-called systems integrators. These are companies that piece together larger systems for an end user (like the military) made up of a wide range of different types of equipment.

Finally, there are the customers or end users of the equipment. These come from a wide range of backgrounds. The end users or customers may be businesses with offices, industry with factories and process control plants, or the government, military, hospitals and other medical organizations, schools, colleges and universities, and you the consumer. There is also a network of service companies designed to support customers. They may do installation or repair work that the customer is not capable of doing.

Electricity versus Electronics

One of the questions that always comes up in the early stages of learning about electronics is what the difference is between electricity and electronics. Let's get this straight up front. Electricity refers to the basic energy we use for power. It generally involves the generation and distribution of electrical energy to homes, businesses, and factories. Electricity is energy that is produced by the coal-fired or nuclear power plant and then sent to you over long transmission lines until it comes into your house where you access it through your AC wall outlets. You use that energy for lighting and appliances, and also to power most of your electronic equipment.

Electrical power may also be generated for use in other systems. The electrical system in a car is a great example. It uses a battery plus a generator driven by the automobile engine to recharge the battery and to operate all of the electrical and electronic devices in a car. Ships, airplanes, and off-shore drilling rigs also have their own electrical-power generating systems. The new electric wind generators and solar power systems are also electricity-producing systems.

All of the principles of electricity also apply to electronics. Basic electrical components such as resistors, capacitors, inductors, transformers, motors, and other similar devices are also used in electronics.

Electronics is the field of science that uses electrical principles to perform other useful functions. Electronics is the control of electricity to produce radio, television, computers, robots, and MRI machines. To control the electricity, you need more than just basic electrical components such as resistors and capacitors. What makes electronics possible are the unique components called *transistors*. Electronics began, of course, with vacuum tubes and then switched over to transistors as the main control devices. Transistors are now miniaturized and packaged in devices called *integrated circuits* that perform very specific electronic functions, such as generating signals, amplifying signals, or performing some control function. Just keep in mind that it is the control devices such as transistors and the circuits that they comprise that make electronics what it is. The transistors are used to form circuits that in turn perform all of the various functions required to produce electronic devices.

THE PAINFUL TRUTH

Okay, here comes the hard part. Electricity is moving electrons. An electron is a tiny particle that is inside the atoms that make up everything we can feel, see, and even not see. If we can get some electrons moving together in a wire, electrical or electronic component, or circuit, we can control them to do many useful things, such as amplification, for example. Moving electrons make up what we call *current flow*. And current flow is produced by electrical energy called *voltage*. Voltage comes from batteries, the AC wall outlet, and lots of other sources including power supplies, solar cells, microphones, and the like. If we can control the electrons we can produce magic results. That is what electronics is all about.

We control the electrons with components like resistors, capacitors, inductors, transistors, and all sorts of integrated circuits. We create circuits that actually process the electron flow. We can call the voltages that create current flow and the resulting current *signals*. The circuit processes the signals to produce the end effect that we want, be it a radio, a computer, or DVD player.

One of the basic models I will use in this book is shown in Figure 1.3. Think of electronic signals as the inputs to a circuit or piece of equipment that processes the signals producing new signals that represent the output we want. Think of the small voltage that comes from a microphone when you talk into it. It is basically too small to be useful. So we apply it to a circuit called an *amplifier* that makes the signal bigger and more useful. That amplifier processes the signal to make it large enough to drive a speaker as our desired

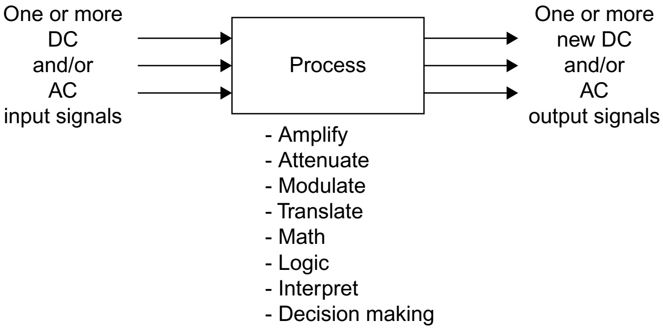

FIGURE 1.3 Model of how all electronic circuits and equipment work. Input signals are processed by circuits and equipment into new output signals.

output. The result is a public address system. Anyway, you get the picture. Just remember that *all* electronic circuits and equipment are based on this model: INPUT-PROCESS-OUTPUT. Think about it.

Here are a couple of other examples to show how most electronic things work.

Cell Phone

Take a look at Figure 1.4. On the left is a cell phone handset. It is made up of two major pieces–a transmitter (TX) and a receiver (RX). You talk into a microphone and generate a small electrical voltage or signal representing your voice. The transmitter takes that and processes it in such a way that the voice controls or modulates a high-frequency radio signal. That signal carrying your voice is then radiated by the antenna to a cell site or base station. The base station receiver picks it up and processes it back into the original voice signal where it then undergoes other processing before being sent over the telephone network to its destination. The dialing process or text messaging from the keyboard also generates input signals that are transmitted to the base station.

On the return side, the voice from your caller passes through the cell phone network to the base station and it is then used to control the transmitter to send that voice to you by radio. The transmitter sends the signal to the antenna that sends the signal to the receiver in your cell phone. The receiver processes the signal by amplifying, demodulating, and otherwise processing the voice that eventually is sent to the speaker that is glued to your ear. Another output may be a text message on the liquid crystal display (LCD).

Computers

The main function of a computer is to store and retrieve data, process data, and to provide automatic control. Data consists of information, facts, and figures. It may be numbers, letters, symbols, or text of any kind. It might represent

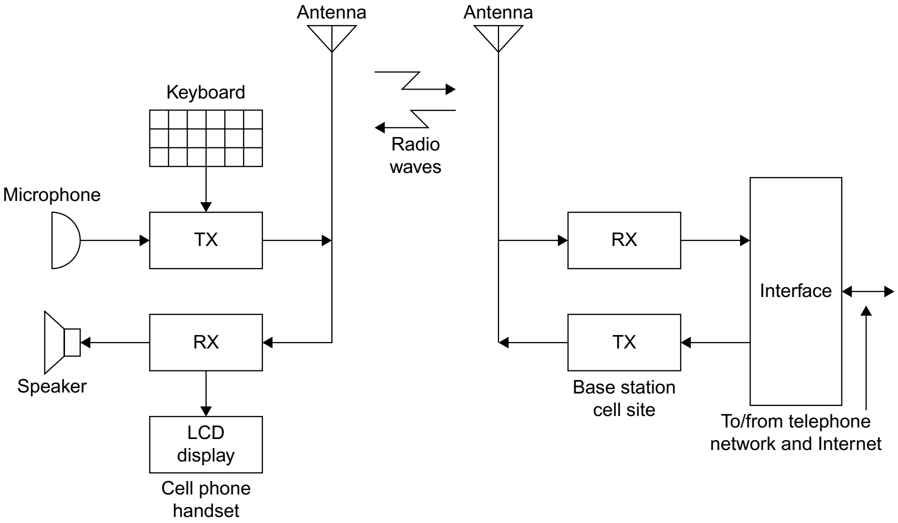

FIGURE 1.4 Model showing how a cell phone communicates with a cell site to make a call.

names and addresses, parts inventory, accounting or banking figures, control signals, or a variety of other items. Data may also be graphics and illustrations, or voice and video that are converted to data. Such data can be stored electronically in computers and rapidly accessed.

Computers process or manipulate data in many different ways. They can perform mathematical operations on it, sort it, edit it, translate it, or otherwise operate upon it to produce some useful output. Computers also make decisions based on the data.

Computers have so many applications that it is difficult to enumerate them. In general, computers are tools that help people organize and perform their work in a more efficient manner. Computers also allow people to do work that was difficult or impossible to do without computers. Computers perform accounting and banking operations; store and process mailing lists and employee records; solve complex mathematical problems in business, science, and engineering; and create graphics. The thing that makes them so valuable is that all of their operations are automatic. Humans need not be a part of the actual computation, as is necessary with a calculator. Computers carry out preprogrammed assignments to solve various problems or perform operations. Most operations are extremely fast. It is this automatic feature that makes computers so useful.

All computers, regardless of their size and features, are made up of the four basic sections shown in Figure 1.5. The memory electronically stores the data to be processed, as well as the program that tells how the data is to be manipulated. The program is a sequence of codes or instructions that cause the computer to carry out its preprogrammed assignment one step at a time.

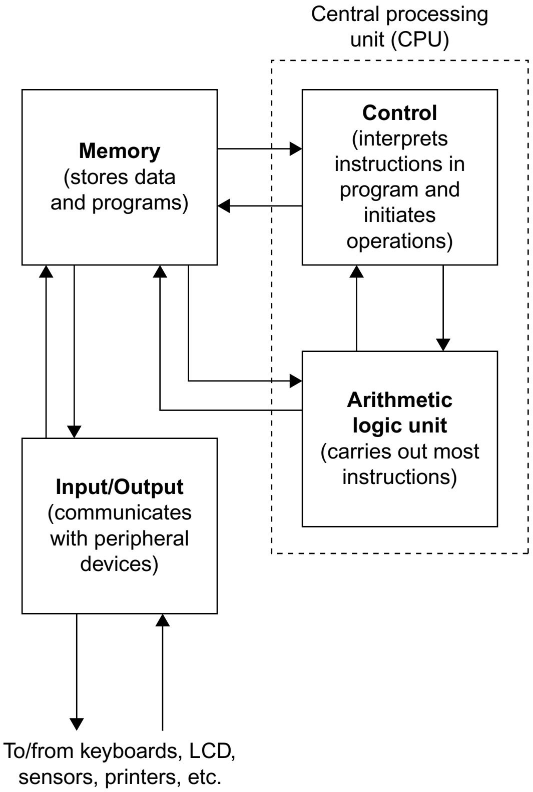
Central processing unit (CPU)
Memory (stores data and programs)
Control (interprets instructions in program and initiates operations)
Arithmetic logic unit (carries out most instructions)
Input/Output (communicates with peripheral devices)
To/from keyboards, LCD, sensors, printers, etc.

FIGURE 1.5 Block diagram representing any digital computer, be it a mainframe, PC, or embedded controller.

Refer again to Figure 1.5. The control section examines the program instructions one at a time in sequence, interprets them, and then sends control signals to all other parts of the computer. The desired operations, such as store, add, print, and so on, are carried out.

The arithmetic logic unit (ALU) is the section that actually carries out most of the processing. It performs data transfers, arithmetic operations, and logical decision making. In most computers, the control and ALU sections are integral units that work closely together. They are usually referred to as the central processing unit (CPU). A microprocessor in a personal computer is a single chip or central processing unit.

The input/output (I/O) section in Figure 1.5 is the interface between the computer and the outside world. It communicates with peripheral devices that humans use to input programs and data, and receive the output results of computation. Data and instructions are usually entered via a keyboard, while outputs are displayed on a video screen or printed out. Inputs may come from sensors and output may go to a printer or LCD readout.

There are many sizes and types of computers. Supercomputers are the largest and fastest computers. They are used primarily for difficult, complex scientific and math problems. Mainframe computers are also very large, and they are used by business and government for financial applications, data storage, and other applications. Personal computers are the smallest practical computers in use. Millions of them are used for word processing, financial spreadsheets, database management, and email and Internet access.

There are also many dedicated computers, built into other equipment to perform single-purpose operations unlike the general-purpose data processors described above. Thanks to semiconductor technology, it is now possible to put an entire computer on a single chip you can hold in your hand. These are called *embedded computers*, or *microcontrollers*. These microcomputers are used like components and are built into most other kinds of equipment such as TV sets, microwave ovens, gasoline pumps, washing machines, and many others. Virtually every other electronic product contains at least one or more embedded controllers. And even very large ICs have controllers on the same chip called *cores*.

Robots

A great example of a control device is a robot. Most robots consist of an electromechanical manipulator arm used in manufacturing operations. Robots are machines designed to mimic human capabilities. Most robots simulate a human arm and hand. As a result, a robot may be able to substitute for a human in certain manufacturing jobs. A typical robot manipulator arm is shown in Figure 1.6. Motors operate shoulder, elbow, and wrist joints, while grippers and other tools form the hand that does the work.

Robots help automate the manufacturing process to increase productivity. They lower costs, produce more goods in less time, and improve quality. Robots also replace humans in many jobs requiring repetitive, tedious, and hazardous work.

Robots perform a variety of tasks in industry. The most common are welding, spraying or painting, and “pick-and-place” operations. Various tools on the end of the manipulator arm, called *end effectors*, let the robot perform a remarkable variety of tasks. In addition, a programmable controller or computer is used to control the robot. A robot may thus be programmed to perform a wide variety of jobs and is easily reconfigured when the work changes.

In subsequent chapters you will see more of these kinds of descriptions. But in Chapters 2, 3, and 4, you will learn a few more basics that make the rest of the book more understandable.

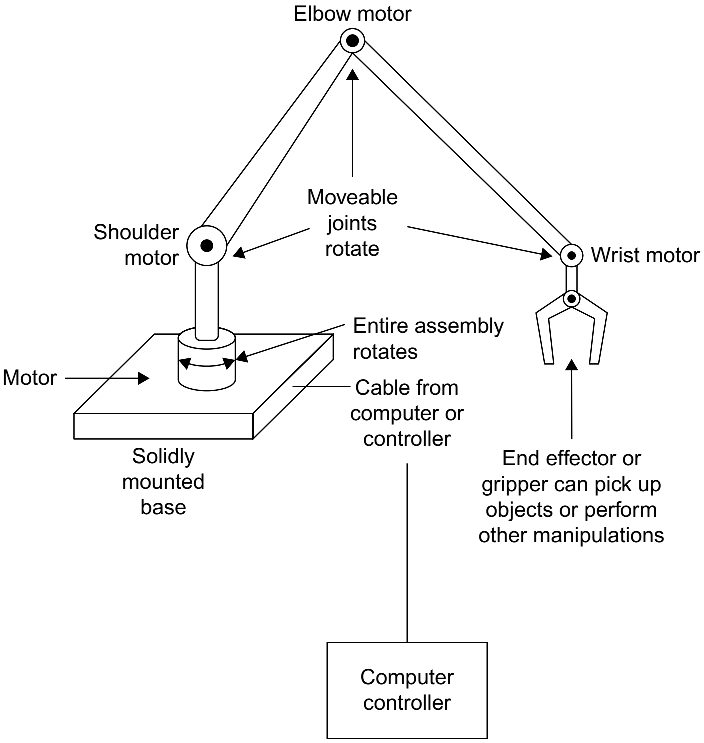
Elbow motor
Moveable joints rotate
Shoulder motor
Wrist motor
Entire assembly rotates
Motor
Cable from computer or controller
Solidly mounted base
End effector or gripper can pick up objects or perform other manipulations
Computer controller

FIGURE 1.6 The most common form of robot emulates a human arm and hand to perform work automatically under the control of a computer.

Project 1.1

How Much Electronics Do You Own?

Just to be sure that you really appreciate the extent to which electronics impacts your life, I strongly suggest that you take a few minutes and make an inventory of all the electronic devices and equipment that you use at home, at work, and away from home. Get a yellow pad and a ballpoint pen and move around your house and just write down all the electronic products that you own and use on a regular basis. TV sets, computers, and cell phones are the most obvious, but don't forget things like telephones, your wrist watch, appliances, and other entertainment products. Make a separate listing of the electrical and electronic products you use inside your car and then another one for those electronic products that you use at work. Make the list comprehensive. Study it. Be amazed.

Project 1.2

How Much Electronics Do You Use?

Now take that same pad of paper and pen and make a quick summary of how you spend your time during the day. How much time do you spend using each of the electronic products that you listed above? Some of them are only momentarily used, but others you use over longer periods of time, particularly computers and TV sets. Just take a quick estimate of the time you use each one during a typical day. Again, add up the hours. Be amazed.

Project 1.3

Decisions, Decisions

Make the following decision. If you had to give up all of your electronic products but one, which one would you keep? TV, cell phone, remote control, iPod? Only one. This will say a lot about who you are.

Chapter 2

Electronic Concepts: More Interesting Than You Think

Some Basic Stuff You Really Need to Know

In this Chapter:

- Electrons and current flow.
- DC and AC voltage.
- Voltage sources, batteries, the power line, and power supplies.
- Signal sources.
- Magnetism and induction.

INTRODUCTION

You don't have to be an engineer, physicist, or mathematical genius to learn electronics. But there are a few basic things that you need to know so you can understand what is going on in the circuits and equipment. For example, you need to know that electrons are what makes up current flow, and voltage is what makes the current flow. You also need to know about magnetic fields and electric fields, where they come from, and what they do. And how magnetism and current flow are related. With that background you are ready to learn about the various components and circuits. This chapter takes care of those basics.

ELECTRICITY AND ELECTRONICS

Electricity is a type of energy produced by a charge that is either fixed (static) or moving (dynamic). The main source of electricity is the electron, a subatomic particle that has a negative charge. When electrons are stored or moved, electricity is produced.

Electricity is one of our prime sources of energy. It is used for lighting, heating, and operating appliances of all sorts. It powers motors to produce mechanical energy, and it powers our electronic equipment. The uses of electricity are virtually infinite.

© 2010 Elsevier Inc. All rights reserved.
doi: 10.1016/B978-1-85617-700-9.00002-3

Electronics is the field of applied science that uses components such as resistors, capacitors, diodes, transistors, and integrated circuits to control and process electricity. These components are used to create circuits that convert, modify, vary, translate, or otherwise manipulate electrical charges to perform useful functions. That is the essence of it. Now for the details.

Atoms and Electrons

This section really ought to be called Physics 101 but I don't want to scare you away. Just to make you more comfortable, I have to say that it is not all that complicated or difficult to understand. So even if you avoided physics in high school, just be aware that it is pretty easy. Trust me.

All matter, whether it is a solid, a liquid, or gas, is made up of tiny entities called atoms. An atom is the smallest possible particle of substances called elements. An element is a chemical substance that cannot be subdivided into smaller, simpler substances. Typical elements are hydrogen, oxygen, gold, silver, copper, carbon, helium, silicon, and sodium. All matter is either an element or composed of several elements. Atoms of two or more elements often combine to form new substances called compounds. For example, water is a compound of hydrogen (H) and oxygen (O) or H_2O. Salt is a compound of sodium (Na) and chlorine (Cl) or NaCl. The smallest possible particle of a compound that exhibits all the characteristics of the substance is called a molecule.

According to the theories of physics, an atom is composed of a nucleus consisting of a core of tightly bound subatomic particles, called protons and neutrons. Protons have a positive charge. Neutrons are neutral, having neither a positive nor a negative charge.

Rotating around the nucleus are electrons. The electrons arrange themselves in multiple orbits in much the same way that the planets orbit around the sun. The orbits are sometimes referred to as rings or shells. Electrons have a negative charge. Figure 2.1 illustrates a copper atom. The number of electrons in orbit equals the number of protons in the nucleus. The number of positive charges equals the number of negative charges. Therefore, the atom is balanced electrically.

As far as electronics is concerned, the most important part of an atom is its electrons. Since electrons can be manipulated (i.e., stored or moved), they can be used to produce electricity. This electricity can then, in turn, be processed and controlled to perform a wide range of valuable functions. By applying an external force, the electrons in the outer shells can be stripped off to produce current flow.

Charge, Voltage, and Current

There are two basic types of electricity, static and dynamic. Static electricity is the buildup of a charge between two objects. Dynamic electricity supplies a continuous charge.

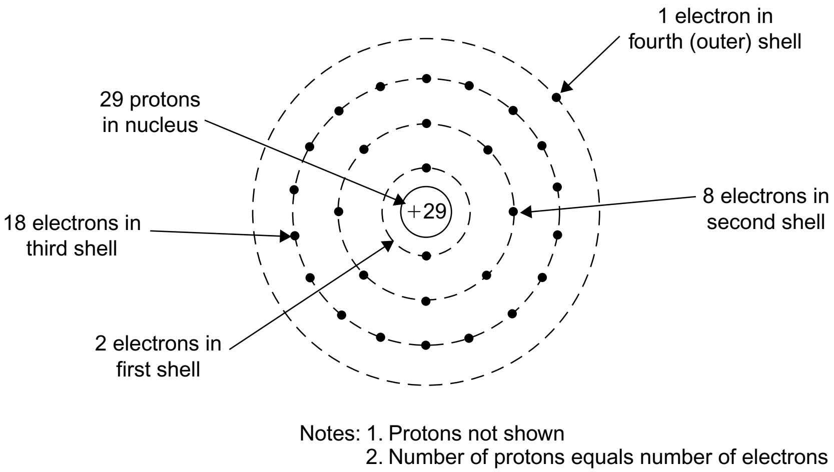
1 electron in fourth (outer) shell
29 protons in nucleus
+29
8 electrons in second shell
18 electrons in third shell
2 electrons in first shell
Notes: 1. Protons not shown
2. Number of protons equals number of electrons

FIGURE 2.1 Physics model of what a copper atom looks like.

Static Electricity

With static electricity, one object has an excess of electrons and the other has a shortage of electrons. The object with an excess of electrons is therefore negative while the object with a shortage of electrons is positive. A basic law of electricity states that opposite charges attract and like charges repel. The two oppositely charged bodies, one negative and one positive, will have a high physical attraction for one another. An invisible force field called an *electric field* will exist between the two charged objects. The object with a shortage of electrons attracts the object with excess electrons.

The two oppositely charged objects will naturally attract one another. The closer they are, the greater the attraction. This attraction of the charges sets up an electrical force field between the two bodies. This electric field, like a magnetic field, contains energy that can be released to perform useful work.

Whenever there are opposite charges on two objects, we say that an electrical potential or difference of potential exists between them. This difference of potential is called *voltage*. Voltage is an electrical pressure that can cause work to be accomplished.

The charge stored on the two objects is usually called static electricity because it is stationary. However, this charge is usually dissipated in some way. If the two charged objects get too close, or the charge builds up too high, the attractive force becomes so great that the electrons jump the gap between them and create a spark. That is how lightning occurs. The clouds and the earth become oppositely charged, the earth positive, the clouds negative. The charge becomes so great that the electrons leap across the free space gap, ionizing the air and creating a path through which the electrons travel to neutralize the positive charge on the earth. The result is the massive spark or flash that we call lightning and the roar we call thunder.

In general, static electricity is not very useful. It represents energy stored as a charge, but it quickly dissipates as a spark or a brief flow of electrons between the negatively and positively charged objects when the two charged objects get close enough or when the charge gets too high. We call this an *electrostatic discharge* (ESD). Static electricity, or ESD, is regarded as more of a disadvantage or nuisance than anything else. In fact, it is one of the main causes of damage to electronic components such as transistors and integrated circuits.

Neutralizing a charge between two objects cancels or eliminates the potential. However, if we had a constant charge that could constantly be replenished, we would have a continual source of potential (voltage) and could therefore sustain a flow of electrons. With such a source of electrical energy, useful work can be accomplished.

Dynamic Electricity and Current Flow

Electricity as we know it, whether in electrical power applications or in electronics, is actually the flow of electrons. Electrons flowing from one place to another is known as current flow. And it is voltage that causes the current to flow. When electrons flow, we have energy that can be put to practical use. Electrons flowing in a filament create light. Electrons flowing in a heating coil produce heat. Electrons flowing in a motor produce mechanical energy. The object is to convert the energy from one form to another. In electronics, we precisely control the electrons with special components and circuits to produce a variety of effects.

Figure 2.2 provides a general block diagram model that shows how all electrical or electronic circuits work. A voltage source produces the energy that causes electrons to flow. This current flow is controlled or processed by a single electronic component, or many components that make up an electronic circuit. The processed electrons flow through a load that consumes or uses the electrical energy to produce a useful output. In an electrical application, the load may be a light, a heating element, or motor, as described above. In an electronic application, the load may be an antenna that radiates the radio signal produced by the electronic circuits. Or the load may be the magnetic head that develops and records digital data on a hard disk or a laser that records a video signal on a DVD.

A couple of key facts about the circuit in Figure 2.2 are required here. First, there must be a voltage source to cause the current to flow. The electrons flow from the negative terminal of the voltage source around the circuit as they are attracted by the positive terminal. Second, the circuit must always form a complete loop from plus to minus.

Current is so important that we need some way to measure it. We do this by standardizing the number of electrons that move past one point in a conductor

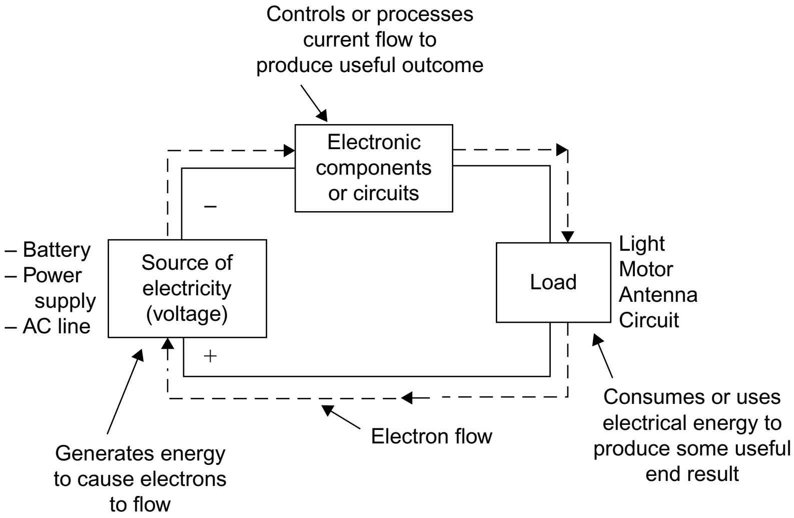
Controls or processes current flow to produce useful outcome
Electronic components or circuits
–
– Battery
– Power supply
– AC line
Source of electricity (voltage)
+
Load
Light
Motor
Antenna
Circuit
Consumes or uses electrical energy to produce some useful end result
Electron flow
Generates energy to cause electrons to flow

FIGURE 2.2 Basic electrical circuit used to describe every electronic circuit.

during a specific period of time. The charge or quantity of electrons (Q) is measured in coulombs (C). One coulomb of charge is equal to 6,242,000, 000,000,000,000 or 6.242×10^{18} electrons. If 1 coulomb moves past a point in 1 second, we say that the current is 1 ampere (A). One coulomb per second (1 C/s) equals 1 A.

Direction of Current Flow

The current flowing in electronic components and circuits consists of electrons that are influenced by a charge or voltage source. The early pioneers of electricity did not know that electricity was the movement of electrons. They knew that there was some kind of movement, of a charge or current, but they did not know what it was, or in which direction it flowed. In order to explain electrical phenomena and perform analyses of components and circuits, they simply assumed a direction of current flow that everyone agreed to and used. Their original assumption was that current flowed from positive to negative, the opposite of the actual direction. This has become known as conventional current flow.

Even after the discovery that real current was electrons moving from negative to positive, scientists and engineers continued to assume a positive-to-negative current flow. And science and engineering professors have continued to teach and support the conventional current flow concept. Most science and engineering textbooks illustrate and teach conventional current flow. Old traditions die hard, or maybe never die.

As it turns out, it does not really matter what direction current flows. Any analysis, design, or explanation can be successfully carried out using either electron flow or conventional current. One way is not any better than the other in

(*Continued*)

Direction of Current Flow (Continued)

describing or understanding electrical or electronic phenomena. Why academia has perpetuated the myth of conventional current flow rather than accepting and teaching the truth is not known. It no doubt stems from long-time heritage, custom, and convention.

Conductors, Insulators, and Semiconductors

A conductor is a material that has many electrons easily freed up by an external voltage. Current flows through conductors easily. They have what we call low resistance. Most good conductors are metals like copper, silver, or aluminum, with loosely bound electrons that can be freed by an external voltage to create current flow. Copper wire is the most commonly used conductor because of its low cost and its ability to be formed into many different shapes and sizes. Most metals are conductors but so is salt water.

Insulators are just the opposite of conductors. These are usually compounds in which the electrons are tightly bound together with the nuclei of the atoms. Even with lots of voltage applied, the electrons are hard to strip away to make current. Insulators keep current from flowing. Some common insulator materials are glass, ceramic, and plastics.

A third category of material is called *semiconductor*. It is a material that can be changed to make it a good conductor or a good insulator or anything in between. Semiconductors are used to make the transistors, diodes, and integrated circuits. The most common semiconductor is silicon. Others are germanium and carbon. Compounds such as gallium arsenide (GaAs), indium phosphide (InP), and silicon germanium (SiGe) are examples. Most integrated circuits (ICs) or chips are made of silicon.

MAGNETISM

You are probably asking, what does magnetism have to do with anything? As it turns out, it has lots to do with electricity and electronics. As you will see, the two are tied together significantly, meaning you need to know about both to get the big picture. Magnetism is essential to all electrical and electronic phenomena. Without magnetism, there would be no electricity or electronics.

Magnetic Fields

Magnetism is an invisible force that has the power to attract pieces of iron, or alloys of iron such as steel. Iron, steel, and the metals nickel and cobalt are called magnetic materials because they can be magnetized to generate a magnetic field, or they support the development and passage of a magnetic field. Magnets are pieces of such materials that exhibit this force. Magnetism has no effect on nonmagnetic materials such as other metals (aluminum, brass, etc.), or on objects made of wood, plastic, or glass.

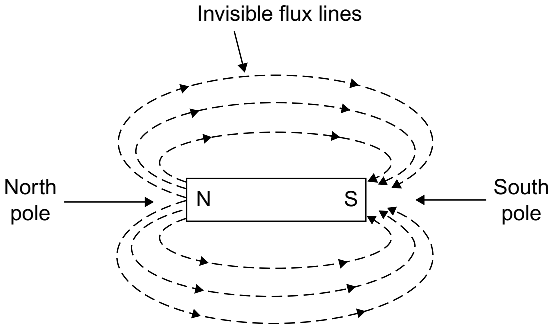

FIGURE 2.3 Permanent bar magnet illustrating the magnetic lines of flux that surround it.

You have no doubt experienced the effects of magnetism yourself. You may have played with a horseshoe magnet as a kid, or used bar magnets like the one in Figure 2.3 in a school lab. And you most likely have used flat magnets to hold paper notes on your refrigerator or file cabinet. These are called permanent magnets because they retain their magnetic force indefinitely.

A magnet is surrounded by an invisible magnetic force field made up of magnetic lines of force or flux lines. These flux lines flow out of one end of the magnet, the north (N) pole, and into the other end, the south (S) pole, as shown in Figure 2.3. The ends of the magnet are the points of the heaviest concentration of the magnetic field.

The basic law of magnetism follows: *opposite poles attract, and like poles repel.* When you get two permanent magnets together, the north and south poles will attract one another. On the other hand, two north poles will repel one another. Two south poles will also physically push away from one another. Permanent magnets are used in a variety of electrical and electronic equipment. Their most common use is in motors, where the attraction and repulsion of magnets create rotary motion.

Electromagnetism

Magnetism is also produced by electricity. Whenever electrons flow in a conductor, they produce a magnetic field. This effect is called *electromagnetism*. Figure 2.4A shows how the magnetic lines of force encircle a wire through which current is flowing. Note that the direction of the lines of force depends on the direction of the current flow.

The strength of the magnetic field around a wire depends on the magnitude of the current flowing. High current (composed of many electrons) produces a strong magnetic field. Despite the current amplitude, however, the magnetic field weakens because it spreads out quickly, even at a short distance from the wire.

If you make a coil out of the wire, as shown in Figure 2.4B, the lines of force around each turn are added together. The result is that a more powerful,

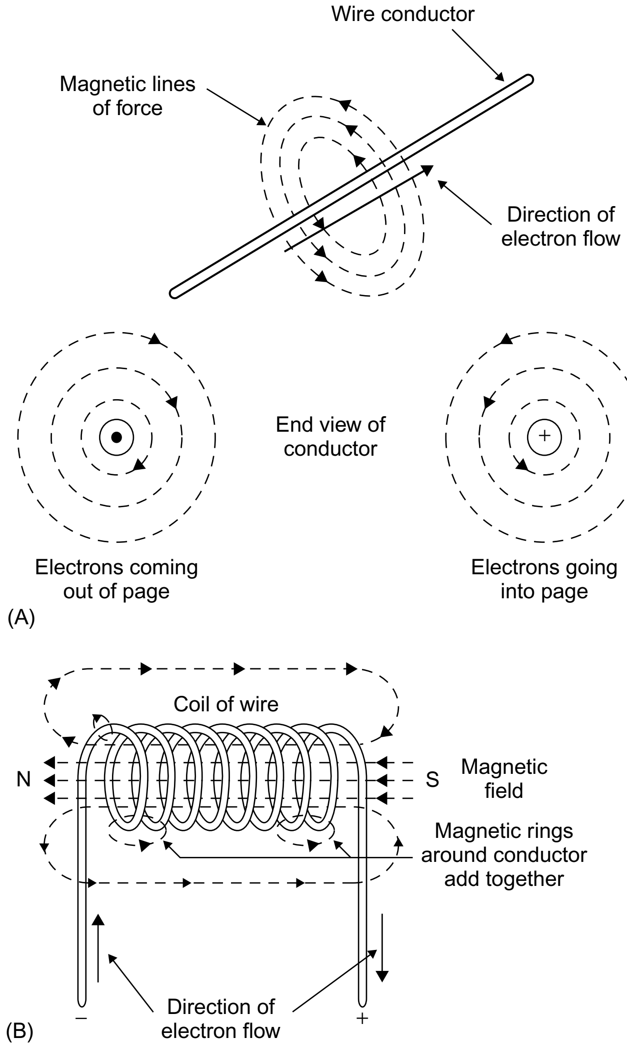
Wire conductor
Magnetic lines of force
Direction of electron flow
End view of conductor
Electrons coming out of page
Electrons going into page
(A)
Coil of wire
N
S
Magnetic field
Magnetic rings around conductor add together
Direction of electron flow
−
+
(B)

FIGURE 2.4 **(A)** Current in a wire produces a magnetic field around it. **(B)** Coiling the wire increases the strength of the magnetic field.

highly concentrated magnetic field is produced. In fact, the coil simulates a bar magnet with north and south poles.

As indicated earlier, the strength of the magnetic field depends on the current amplitude in the wire. High current produces a strong field. The magnetic field is also increased by coiling the wire, which helps concentrate the lines of force. The closer the turns of the wire, the stronger the field. You can also increase field strength for a given current level by simply adding more turns

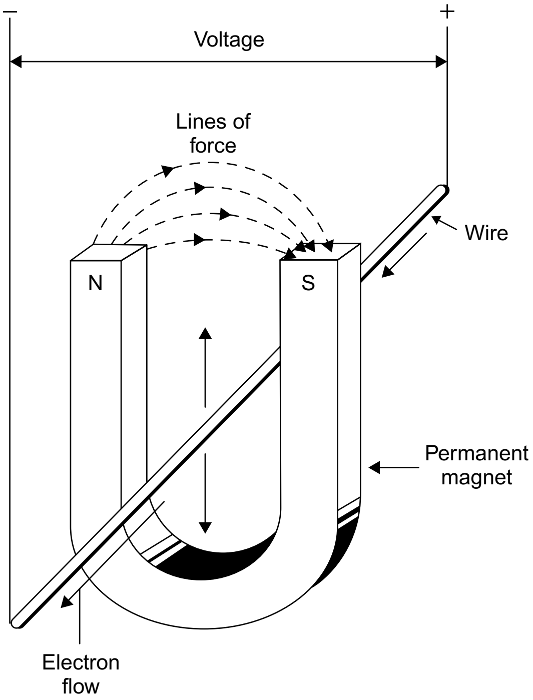

FIGURE 2.5 Relative motion between a magnetic field and a conductor such as a wire causes a voltage to be induced into the wire making it a voltage source.

to the coil. Strong electromagnets are made by tightly coiling together many turns of fine wire in several layers.

Another way to increase the strength of the magnetic field is to insert a bar or core of magnetic material inside the coil. Because flux lines flow easier in iron or steel than in air, an iron or steel bar inside the coil helps concentrate the lines of force. Most of the flux lines flow through the core because it represents a lower path of resistance than the surrounding air for the lines of force. The result is a very strong field.

Electromagnetic Induction

Another important principle of magnetism is electromagnetic induction. This principle states that whenever there is relative motion between a conductor and a magnetic field, a voltage will be induced into the conductor.

Refer to Figure 2.5. If the wire is moved up or down so that it cuts through the magnetic lines of force between the magnet poles, voltage will be induced in the wire and cause current to flow. Voltage will also be induced into the wire if the wire is held in a fixed location and the magnetic field is moved so that the flux lines cut across the wire. The wire actually becomes a voltage source. That voltage will cause current to flow. The direction of current flow in the wire depends on both the direction of the magnetic field and the direction of the relative motion between the magnetic field and the wire.

The amount of the induced voltage depends on the number of flux lines cut and the speed of cutting. The greater the number of lines of force, the higher the induced voltage and the greater the current flow. The highest induced voltage occurs when the conductor cuts across the flux lines at a right angle. If the wire moves parallel with the lines of force, no voltage is induced.

The induced voltage can also be increased by coiling the wire. More turns passing through the magnetic field will produce greater induced voltage and current flow. The motion in the magnetic field causes a small voltage to be induced into each turn of wire. All of the voltages will add together to produce a higher voltage.

In the examples above, we have assumed that the magnetic field comes from a permanent magnet. However, it doesn't really matter where the field comes from. An electromagnetic field works just as well. Electromagnetic induction is the principal method of generating electricity in the world. Most electrical power is produced by giant generators containing large coils of wire that rotate in a magnetic field. The generator or alternator in your car works the same way.

VOLTAGE SOURCES

To make current flow, we must have voltage. That voltage will produce either direct current (DC) or alternating current (AC). DC is electron flow in only one direction. AC is current that flows in one direction and then reverses and flows in the opposite direction.

DC Voltages

DC may be a fixed, steady current, one that varies, or one that switches off and on periodically. A DC voltage source is one that causes direct current to flow one way through the circuit, component, or conductor. It has positively and negatively charged terminals of a fixed polarity. It always causes electrons to flow from negative to positive.

Figure 2.6A shows a fixed, positive DC voltage source, which is usually either a battery or an electronic power supply. We call it a positive source because the negative terminal of the DC source is connected to a common point called the *ground*. The other terminal is positive with respect to ground. Voltages are always looked at or measured between two points with one point (ground) being the reference. The polarity of the voltage is the polarity of the terminal not connected to ground.

The remainder of Figure 2.6 shows graphs of DC voltage over time. In electronics we normally view DC or AC voltages with respect to time. The positive DC voltage is shown in Figure 2.6B as a horizontal dashed line.

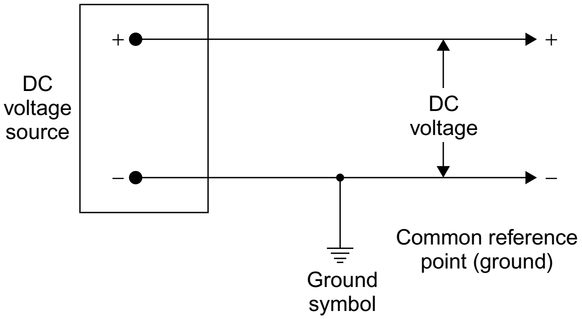
DC
voltage
source
+
−
+
−
DC
voltage
Common reference
point (ground)
Ground
symbol

(A)

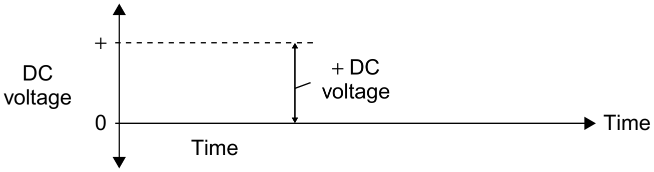

(B) *Fixed, positive DC voltage*

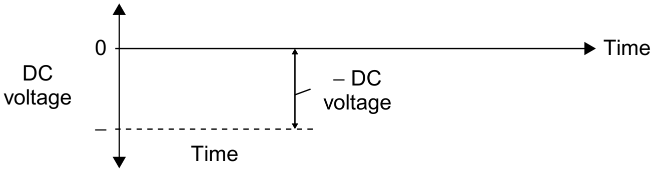

(C) *Fixed, negative DC voltage*

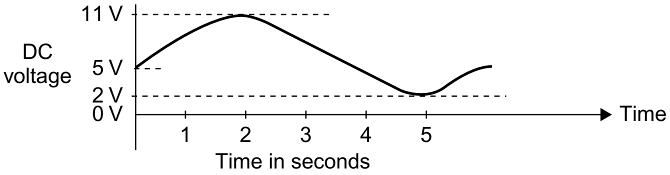

(D) *Varying, positive DC voltage*

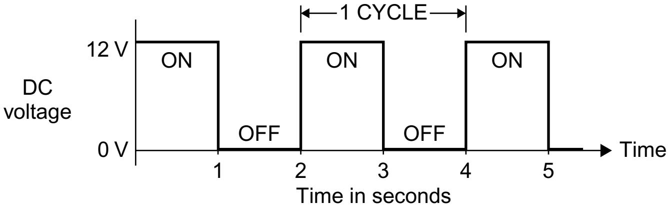

(E) *Pulsating DC voltage*

FIGURE 2.6 DC voltages. **(A)** DC voltage source such as a battery or power supply. One side is usually grounded to serve as a reference. **(B)** Fixed, continuous positive DC voltage shown over time. **(C)** Negative DC voltage. **(D)** Varying positive DC voltage. **(E)** DC voltage pulses.

A negative DC voltage is shown in Figure 2.6C. The dashed line representing the voltage below the zero line indicates that it is negative. The voltage is negative with respect to the reference point known as ground.

Figure 2.6D shows a varying positive DC voltage. It begins at +5 volts, rises upward to +11 volts, and then drops to +2 volts over a period of 5 seconds. The left-hand scale indicates the voltage level.

Figure 2.6E shows a pulsating DC, that is, one that turns off and on. This voltage switches on to +12 volts, stays on for 1 second, and then switches to zero for 1 second. Then this on–off cycle repeats. One cycle is the time of one ON pulse and one OFF pulse or in this case 2 seconds. Such pulse signals are usually called *binary* or *digital signals*.

AC Voltages

As its name implies, AC current flows in two directions in a conductor, component, or circuit, but not at the same time. The electrons move in one direction, then reverse and flow in the other direction, both for a short period of time. The direction of flow reverses repeatedly.

The polarity of an AC voltage reverses periodically. During one period, the terminals of the voltage source are + and −, but later reverse to − and +. This reversal continues to repeat as the voltage amplitude varies. Since it is the voltage that causes current to flow, if the polarity of the voltage changes, then naturally the direction of the electron flow will also change periodically. The electrons just flow back and forth in the circuit.

An AC voltage source is usually represented by just a circle with a varying wave inside as illustrated in Figure 2.7A. It may be the AC outlet in your home, a signal generator, or a type of electronic circuit called an *oscillator*. AC voltages and currents can assume many shapes. The most common electrical and electronic AC voltage is a sine wave, as shown in Figure 2.7B. The voltage rises gradually from zero to a positive peak, and then drops to zero. Then it reverses polarity, rises to a negative peak, and returns to zero. Each rise and fall is called an alternation. One positive and one negative alternation form a cycle. The cycles repeat over and over. The number of cycles that occur per second is called the frequency.

The AC power-line voltage that comes into your house and appears at the outlets is a sine wave. Its frequency is 60 cycles per second (cps). The term *hertz* (abbreviated Hz) is normally used instead of cps to express frequency. The power-line frequency, therefore, is 60 Hz. All radio signals are sine waves, but have very high frequencies.

AC signals can have a variety of other shapes as well. Figure 2.7C shows a square or rectangular wave. It is made up of pulses of alternating polarity voltages that switch from +6 to −6 volts. Figure 2.7D shows a triangular wave, and Figure 2.7E, a complex waveform that might represent voice or video. Regardless of their shape, AC signals have one thing in common: they reverse the polarity and direction of current flow periodically. Their frequencies can range from a fraction of a cycle per second to billions of cycles per second.

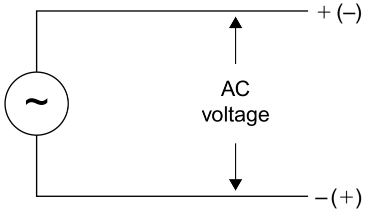

(A)

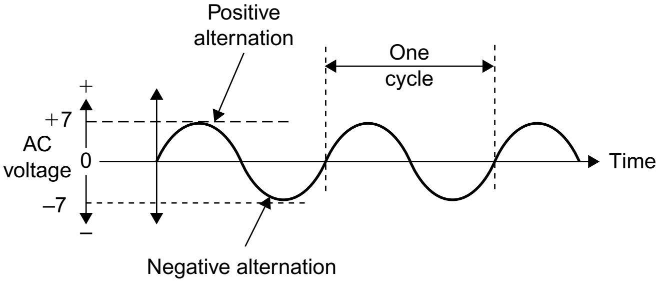
Positive
alternation
One
cycle
+
+7
AC
voltage
0
–7
–
Time
Negative alternation

(B) *Sine wave*

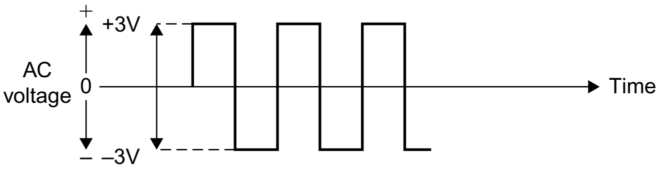
+
+3V
AC
voltage
0
–
–3V
Time

(C) *Square wave or rectangular pulses*

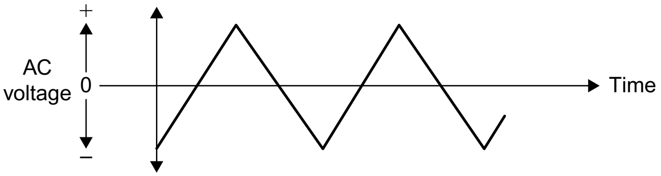
+
AC
voltage
0
–
Time

(D) *Triangular wave*

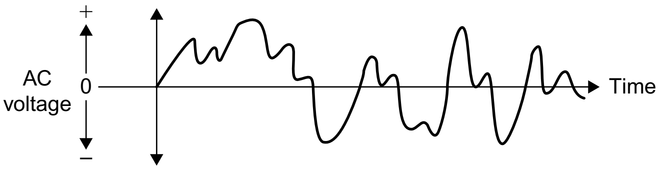
+
AC
voltage
0
–
Time

(E) *Random complex wave*

FIGURE 2.7 AC voltages. **(A)** AC voltage source drawn as a circle with a wave inside. The output polarity changes periodically. **(B)** Sine wave, the most common AC voltage shape. **(C)** Rectangular wave or AC pulses. **(D)** Triangular wave. **(E)** Random AC wave that may represent voice or video.

AC Frequency

Frequency is measured and expressed in units called cycles per second (cps) or Hz. Hertz is the more common unit. To express higher frequencies, the units kilohertz (kHz), megahertz (MHz), and gigahertz (GHz) are used.

1 kHz = 1000 Hz
1 MHz = 1 million Hz or 1000 kHz
1 GHz = 1 million Hz, 1 million kHz, or 1000 MHz

Table 2.1 presents a summary of various frequency ranges that you will encounter and the types of signals in those ranges.

AC Voltage Measurement

To measure or represent the amount of AC in a circuit, we usually state the peak, peak-to-peak, and root mean square (rms) values. The peak value of a wave is simply the voltage measured from the zero line to the maximum positive (or negative) value. In Figure 2.7B, the peak value of the sine wave (V_p) is 7 volts. The peak-to-peak value, normally abbreviated V_{pp}, is the positive peak added to the negative peak. Or, since the positive and negative peaks are identical in a sine wave, the peak-to-peak value is two times the peak value, or in this case 14 volts.

$$V_{pp} = 2V_p$$

TABLE 2.1 The Frequency Ranges of the Electromagnetic Spectrum and Their Applications

Frequency Range	Types of Signals
20 Hz to 20 kHz	Audio, sound, voice, and music.
300 kHz to 3 MHz (LF)	Low radio frequencies. AM broadcast.
3 MHz to 30 MHz (HF)	High-frequency radio, shortwaves. Broadcasting, amateur radio, CB.
30 MHz to 300 MHz (VHF)	Very-high radio frequencies. 2-way radio, mobile, marine, aircraft; TV and FM broadcast.
300 MHz to 3 GHz (UHF) Anything above 1 GHz is called microwave.	Ultra-high radio frequency. 2-way radio, TV broadcast, cell phones, GPS satellites, wireless broadband, wireless networks.
3 GHz to 30 GHz (EHF)	Extremely high radio frequencies. Radar, satellites, network radios.
30 GHz to 300 GHz	Super-high radio frequencies. Radar, satellites, network radios.

Another way to measure AC voltage is to use an average value known as the root mean square (rms), or effective value. The rms value is computed by multiplying the peak value by 0.707, or:

$$V_{rms} = 0.707\ V_p$$

Thus, the rms value of the sine wave is $0.707 \times 7 = 4.949$ volts. Most sine wave voltages are measured and stated in terms of the rms value. The amount of AC voltage available at a standard AC outlet, for example, is 120 volts rms. The rms value as discussed above is applied only to sine waves. The significance of the rms value is discussed later when we address power.

PRACTICAL DC AND AC VOLTAGE SOURCES

Just what is the source of DC and AC? Such voltages originate from a variety of sources.

DC Voltage Sources

Batteries

The most common source of DC voltage is a battery. A battery is a collection of cells. A cell is the basic unit of voltage generation that creates electricity by chemical action. A battery is made up of two or more cells connected together. Figure 2.8 shows the configuration of a cell. It consists of two metal electrodes immersed in a solution called an electrolyte. The chemical interaction between the metal electrodes and the electrolyte produces a separation of charges. This separation of charges results in an excess of electrons on one electrode (the − terminal of the cell) and a shortage of electrons on the other electrode (the + terminal of the cell). If a conductor is connected externally between the terminals, electrons will flow from − to +.

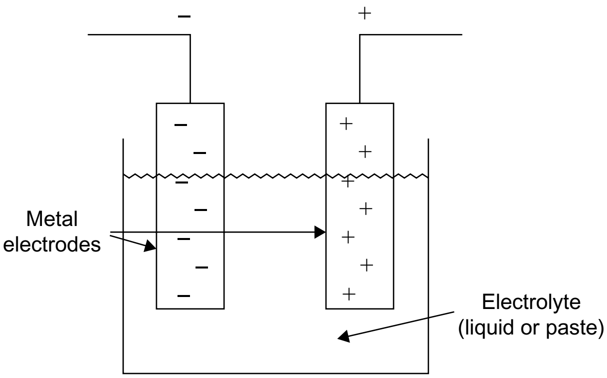

FIGURE 2.8 Basic concept of any battery.

TABLE 2.2 Most Popular Types of Cells/Batteries and Typical Applications

Materials	Voltage	Type	Applications
Alkaline	1.5	Primary	AAA, AA, C, D cells; flashlights; toys; and much more
Mercuric oxide	1.35	Primary	Watches; calculators
Silver oxide	1.6	Primary	Hearing aids; watches
Nickel-cadmium	1.2	Secondary	Tools
Nickel metal hydride	1.2	Secondary	Laptops; cell phones
Lead acid	2.1	Secondary	Cars; golf carts
Lithium ion	4.1	Secondary	Cell phones; iPods; laptops

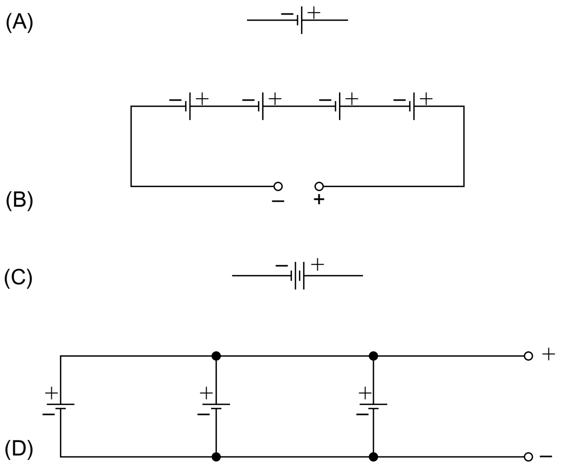

FIGURE 2.9 Battery and cell symbols used in circuit diagrams. **(A)** Cell. **(B)** Four cells in series making a battery. Cell voltage add-up. **(C)** Symbol for battery of any number of cells. **(D)** Cells in parallel to increase current capability.

The amount of voltage produced by the cell depends on the type of electrodes and the electrolyte. Various chemical combinations generate different voltages. Table 2.2 lists the most commonly used cells and the voltage they generate.

The schematic symbol of a cell is illustrated in Figure 2.9A. To create a battery, two or more cells are connected together. The cells are usually connected in series form, that is, end to end and − to +, as shown in Figure 2.9B. The cell voltages add up to produce the total battery voltage. For example, six small alkaline cells are connected in series to create a $1.5 \times 6 = 9$-volt transistor-radio battery. A car battery consists of six lead-acid cells connected in series; each produces 2.1 volts for a total battery voltage of $6 \times 2.1 = 12.6$ volts. If four nickel-metal-hydride cells are connected in series, as in Figure 2.9B, the total battery voltage is 4.8 volts.

The simplified schematic symbol for a battery is shown in Figure 2.9C. Instead of drawing all of the individual cells, the simplified battery symbol is used.

The cells listed in Table 2.2 can be separated into two categories, primary and secondary. *Primary cells* cannot be recharged. As the chemical action produces voltage, the electrodes and electrolyte are actually used up. At some point, no further chemical action is produced, and no voltage is generated. The cell must be discarded and replaced.

Secondary cells, on the other hand, can be recharged. In these cells, the chemical action can be reversed by connecting an external DC voltage to the battery. This is called *charging*. Charging forces electrons into the battery, producing a reverse chemical action that rejuvenates the electrodes and the electrolyte. The cell can be discharged and charged repeatedly, giving it a very long life.

The amount of current that can be drawn from a cell depends on its size and the quantity of the materials used. Large electrodes and electrolytes can produce more current, but the voltage stays the same; the type of material, not its volume, determines voltage.

For example, a large D-size flashlight cell and a tiny AAA penlight cell are both made of alkaline material so their output voltages are both 1.5 volts. The D-cell, however, can produce more current and will have a longer life because it is larger.

A greater current capacity can be obtained by simply connecting two or more cells in parallel. In a parallel connection, the cells are wired across one another, all + leads connected together, and all − leads connected together, as shown in Figure 2.9D. The output voltage remains the same, but the current output can be higher.

Solar Cells

A special type of voltage-producing cell is the so-called *solar cell*. Also called a *photovoltaic* (PV) cell, a solar cell converts light into electrical energy. When the sun or artificial light shines on the cell surface, a voltage is generated. The most commonly used material is silicon, and a typical cell generates about 0.5 volt. Other materials, such as cadmium sulfide and gallium arsenide, are also used.

You have probably seen solar cells used in hand-held calculators. Big arrays of solar cells are now being adopted as an energy source for homes and businesses. Another important application is spacecraft. Virtually every satellite, deep-space probe, or other space vehicle, uses solar cells for power. Huge panels of solar cells connected in a complex series–parallel array produce a powerful battery that can operate all electrical and electronic equipment on board. Typically, however, such a solar battery is primarily used to recharge the main spacecraft power source, which may be a nickel-cadmium secondary battery, as shown in Figure 2.10. The voltage produced by this array is 2 volts.

Fuel Cell

A fuel cell is a special type of battery that combines oxygen and hydrogen gas to produce voltage and electron flow. That means you have to have a supply of hydrogen gas, which is usually stored in a bottle or other container. Typically, the oxygen can be extracted from the air, but if not you will need a bottle of oxygen as well. When you combine the two, charges separate and plus and minus terminals are formed. Most fuel cells generate between 0.5 volt and 0.9 volt, so you have to put many of them in series to be useful. Usually they are bulky and inconvenient because of the need for the gases. But as long as oxygen and hydrogen are supplied, the fuel cell produces the voltage. Heat and water are by-products of the chemical process, making them essentially impractical for many applications.

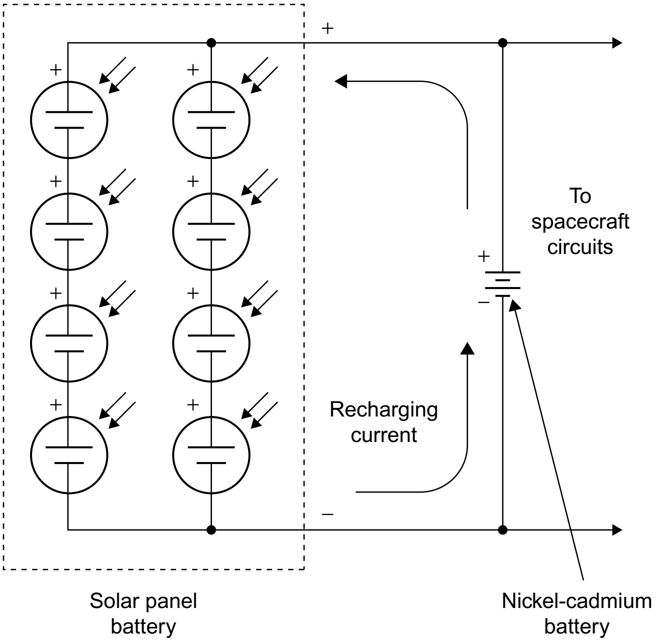

FIGURE 2.10 Panel comprised of solar cells in series and parallel make up DC source that charges the battery in a spacecraft.

Power Supplies

Electronic circuits require a source of DC for proper operation. A great many electronic products are still powered by batteries because they are small and portable, such as cell phones, iPods, and the like. Other electronic equipment is designed to operate from the AC power line. Such equipment contains an electronic power supply that converts AC into the required DC voltages. Computers, TV sets, and other nonportable electronic equipment have an AC power supply.

AC Voltage Sources

Generators

The AC that appears at the outlets of our homes and offices is produced by huge electromechanical generators owned and operated by an electrical utility company. A source of mechanical power—such as water, or a steam turbine powered by oil, coal, or nuclear fuel—operates the generator. The turbine converts steam into mechanical energy to turn the generator. The generator rotates coils in a strong magnetic field to produce a voltage. The speed of rotation of the coils determines the frequency. That voltage is a 60-Hz sine wave.

The big wind generators popping up across the landscape also produce AC sine waves, but because the rotational speed depends on the wind speed, the AC frequency will vary, which is unacceptable. Therefore, the AC voltage they produce is converted to DC, and then back into AC of the correct frequency (60 Hz) to be used or put on the AC power grid.

Another example is the alternator in your car. "Alternator" is just another word for an AC generator. The alternator, driven by the car engine through a belt, produces AC output that is converted to DC by diodes making up a rectifier. The rectifier output is DC to charge the battery.

Inverters

Electronic circuits called inverters are used to produce AC from a DC source. This is common in solar power systems. The solar cells produce DC, which keeps batteries charged. The DC is then converted to AC by the inverter.

In electronic equipment, AC signals are generated by circuits called *oscillators* or *function generators*. The AC signals are sine waves, square (rectangular) waves, or triangular waves, among others.

Analog and Digital Signals

You have no doubt heard the terms "analog" and "digital." These terms distinguish between the two main types of voltages or signals in electronic equipment. Analog signals are voltages that vary smoothly and continuously over time. You saw some in previous figures. Analog signals may be either DC or

AC. A continuous DC voltage is analog, as well as a varying DC signal. A sine wave and the random signal in Figure 2.7E are also analog signals.

Analog signals come from microphones, video cameras, and sensors that detect light level, temperature, and other physical conditions. The most common types of analog signals are voice or music from a microphone and video from a camera.

Digital signals are pulses or on–off signals. The DC pulses in Figure 2.6E and 2.7C are digital. Most digital signals are binary in nature, meaning they only have two voltage values. In Figure 2.6E the two values of voltage are +12 and zero volts. In Figure 2.7C the two voltage levels are +3 and −3 volts.

You will see both analog and digital signals in all sectors of electronics. Today, most signals are digital. Special circuits called analog-to-digital converters (ADC) and digital-to-analog converters (DAC) are used to translate between the two types of signals.

Fourier Theory and the Frequency Domain

We normally think of electronic signals as voltages that vary over time. Waveforms showing signals are said to be in the time domain. Time is the horizontal axis in a graph. But there is another way to look at signals in the frequency domain where frequency is on the horizontal axis. For example, a sine wave is shown as a single vertical line at its frequency as Figure 2.11A shows. The height of the line is representative of the voltage level, either peak or rms.

The big question is, how do you show a complex signal like a square wave in the frequency domain? The only way to do this is to use the Fourier theory that says that any complex nonsinusoidal wave is made up of a group of sine waves added together. There is a fundamental sine wave that occurs at the signal frequency plus harmonic sine waves. A harmonic is a sine wave whose frequency is an integer (whole number) multiple of the fundamental sine wave. The second harmonic of a 50-MHz sine wave is a 100-MHz sine wave. The third harmonic is a 150-MHz sine wave and so on. If you add the fundamental sine wave to multiple harmonics of differing amplitudes and phases, you can create any shape of wave.

For instance, a square wave is made up of a fundamental sine wave and all the odd harmonics. A 1-MHz square wave is made up of a fundamental 1-MHz sine wave and harmonics of 3 MHz, 5 MHz, 7 MHz, and so on. According to Fourier theory, it takes an infinite number of harmonics but in reality only a few (like up to the fifth or seventh) are needed to approximate the square wave. The frequency domain plot of a square wave therefore would look like the one in Figure 2.11B.

When looking at a time domain view of a signal, try to keep a frequency-domain picture in mind at the same time. You cannot actually see the third or fifth harmonics in the time domain plot but they are buried in there just the same.

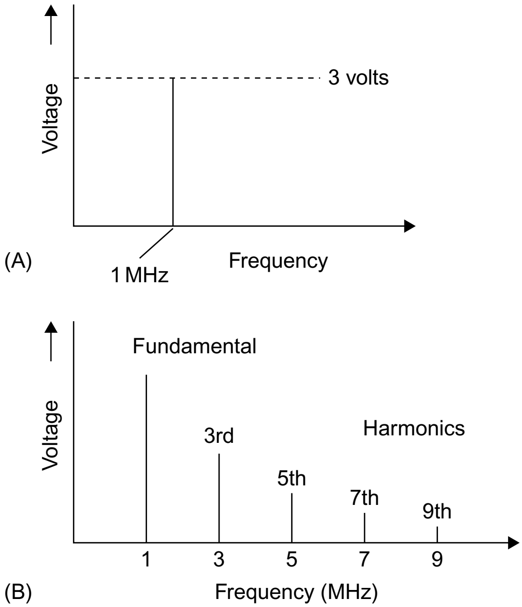

FIGURE 2.11 Frequency-domain views. **(A)** 1-MHz sine wave. **(B)** 1-MHz square wave.

Project 2.1

Getting Familiar with a Multimeter

One of the things you are going to need as you work with electronic parts, circuits, and equipment is a multimeter. This is a test instrument that measures voltage, current, and resistance. Some also measure capacitance and other things. These multimeters are usually hand held and battery operated. There are two basic types, analog and digital. Analog meters use a dial to show the measured value. Digital multimeters (DMMs) have a liquid crystal display (LCD) to show the measurement. See Figure 2.12. The two types work equally well; which one you use is typically a matter of preference for meters versus digital displays. I recommend a digital unit.

The good news is that multimeters are not expensive. You can get one for as little as $10, but a better unit may cost you several times that. The cheaper units are kits that you will need to build yourself. I do not recommend a kit if you are just starting out. Later in this book I will discuss kits in more detail. In any case, you need to buy a multimeter and learn to use it. Plan to buy one as soon as you can. Check your local Radio Shack store or buy online.

Three good sources of multimeters by mail order or web order follow:

Elenco Electronics: www.elenco.com
Jameco: www.jameco.com
Kelvin Electronics: www.kelvin.com

Ask for a printed catalog when you inquire.

(*Continued*)

Project 2.1 (Continued)

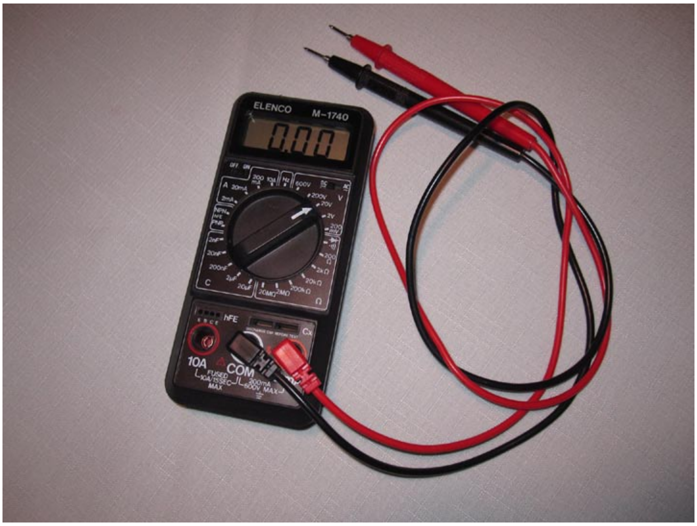

FIGURE 2.12 A low cost digital multimeter that measures DC and AC voltages, current, and resistance.

Project 2.2

Hardware Sources

As an electronic experimenter, you should be familiar with buying parts and equipment. Radio Shack stores still stock some parts but today most hobbyists and experimenters buy via the mail or online. The most popular sources of parts and equipment are All Electronics, Digi-Key, and Jameco. And you will discover others. You can go to their websites, but you really should get on their mailing list and get a catalog. Do that as soon as you can. Links to these sources follow. You can get your multimeter from one of them as well.

All Electronics: www.allelectronics.com
Digi-Key: http://www.digikey.com
Jameco: http://www.jameco.com
Ramsey Electronics: http://www.ramseyelectronics.com

Project 2.3

Magazines of Interest

As you are learning electronics, you need to be continuously aware of what is going on in the industry and with the technology. Magazines are a good source of that information. One that you should read on a regular basis is *Nuts & Volts*. It is

(*Continued*)

Project 2.3 (Continued)

available at some newsstands and bookstores, but if you cannot locate it, go to the website and subscribe. This magazine has lots of articles and columns, including construction projects. It is a great way to stay up to date and learn more. *Nuts & Volts* also publishes *Servo* magazine, which is aimed at people who like to build and play around with robots.

Another recommended magazine that has an embedded controller emphasis is *Circuit Cellar*. This publication contains great educational articles and projects. Get a subscription if you are serious about learning more.

Another interesting magazine is called *Make*—it is not strictly electronics but it usually contains some good electronic projects. It also features other mechanical projects such as robots, among other things. This one is definitely worth a look the next time you are at the newsstand.

Nuts & Volts and *Servo*: www.nutsvolts.com
Circuit Cellar: www.circuitcellar.com
Make: www.makezine.com

Project 2.4

Measuring Voltages

A good first step in using a multimeter is to measure voltages. Batteries are a good place to start. Look around the house and locate some batteries that you commonly use. Take them out of the equipment before you measure them. You will most likely find AA or C cells. Set the multimeter to measure DC according to the meter user's manual and connect the test leads to the battery. The red lead is +, the black lead is −. The LCD on the meter should show the voltage. It should be 1.5 volts. You can also check batteries in some cordless phones, cell phones, or other devices. You may be able to remove the battery from a laptop. In each case, check to determine the battery technology—lithium ion, alkaline, and so on.

Two key points to remember. First, if you get the test leads reversed, the red to the − terminal of the battery and the black to the + terminal, the meter will read a negative value. That just means that the leads are reversed, nothing more. Second, you cannot typically judge the condition of the battery by measuring its voltage. With no load on the battery, no current is drawn from it so you are just measuring the characteristic voltage. If you connect the battery to something, current will flow and its voltage will decrease slightly. If it decreases a lot, the battery is going bad. If the cell or battery voltage is several tenths of a volt lower than the characteristic value given in Table 2.2, the battery is probably bad.

You could also go out to your car and measure its battery voltage. Be careful. You cannot get shocked but you could short something. When the engine is off, the battery will probably register about 12 volts. With the engine running, the battery voltage will be a bit higher, as much as 15 volts or so. The generator is charging the battery, making its voltage a bit higher.

(*Continued*)

Project 2.4 (Continued)

One last thing: you can use the multimeter to measure the AC line voltage at an outlet. Only do this if you have probes that will fit into the slots on the outlet.

And be super careful as you can kill yourself. Be paranoid, even.

Set the multimeter for AC voltage at an appropriate level of at least 120 volts, according to the multimeter manual. Place the probe tips into the AC outlet and read the voltage.

Do NOT touch any metal with the probe tips!!

Your value will depend on things like how far you are from the power station, what appliances and other items are on, and time of day. In any case, the voltage range is about 95 to 130 volts. It is rarely lower than 110 or higher than 125 volts. The typical value is 120 volts, or what is called *nominal*. I measured mine at 117 volts.

Chapter 3

The Systems versus Components View of Electronics

A Fresh New Way to Learn about Electronics

In this Chapter:

- Systems defined.
- Examples of systems.
- Introduction to basic electronic components.

INTRODUCTION

When learning about electronics in the past, you started with basics of physics and math, and then went on to learn electrical theories. Next came instruction in how electronic components work, and then you combined them into circuits. Then you put the circuits together into specific pieces of equipment. Finally, at last, knowing the equipment, you could understand the application, such as TV or computers or audio or home system control or whatever. This approach to learning is called the bottom-up approach. You learned the details and added them all together to get to what you really wanted to know about—the application and its technology.

That is still the way you learn electronics in college today, and it is the approach taken by most other electronics books you may have seen. This book doesn't take that old legacy path. Instead it skips most of the early background and theory, goes right for details on the circuits, and then builds to the applications. It is a much more interesting way to learn electronics. And it is also far more relevant to how electronic equipment is implemented today. We call this way of learning the "systems view" of electronics.

WHY SYSTEMS AND NOT CIRCUITS?

What is the advantage of this approach? First, thanks to modern semiconductor technology, most electronic components and circuits are inside chips or integrated circuits (ICs). You can't see them, and you can't access, test, or replace

© 2010 Elsevier Inc. All rights reserved.
doi: 10.1016/B978-1-85617-700-9.00003-5

them. About all you can do is learn what they do, and then work with the various input and output signals. Since you do not have access to the components and circuits, there is some argument as to why you need to learn all that circuit theory detail. If you are an engineer designing ICs, you definitely need to know this. Otherwise, it is just a waste of time unless you want to know just for the heck of it. Some of it is nice to know, of course, but in most cases it is not necessary. What you really need to know is what the circuit is, what it does, what its characteristics are, and what the inputs and outputs should look like.

A good example is the power supply in a PC. If it goes bad, you just replace it with a $30 unit that you can buy online. Troubleshooting the power supply would take many hours and require some expensive test equipment and access to spare parts. The whole repair could take hours and at a typical rate of, say, $50/hour, or more, it could cost hundreds of dollars and take weeks to get the new parts. A quick replacement with the $30 unit takes only about a half hour at most. So for less time and money, you get your PC working faster. Not a bad deal at all.

A cell phone is the same. It takes hours of troubleshooting and very expensive test equipment to troubleshoot and repair one. A typical repair would cost hundreds of dollars. Consequently, it is rarely done. It is cheaper to get a new cell phone, and it will be a better one at that. And it may even be free if you renew your service contract. That’s electronics today.

The ability to analyze circuit operation by tracing electron flow through the various components used to be necessary. That made it necessary to know about all the components in detail and how they all come together to make a circuit. But with the systems view, you rarely need this knowledge unless you are an engineer designing the circuits or the equipment. In most cases you are not, so why invest the time?

The systems view is a better approach, as you do learn about the circuits and the equipment at a higher level. Later, if you want or need to, you can delve deeper into the circuit and component details, just for fun or to satisfy your curiosity. We won’t do that in this book. But you will get the system view and big picture that will serve you well in learning electronics.

The systems view is more important today because ICs have made it possible to largely ignore the internal details. The ICs and systems on a chip (SoC) have resulted in the throw-away electronics society in which we live. It is no longer possible or economical to repair most electronic equipment thanks to ICs. The cost of labor, the cost of test instruments to troubleshoot equipment, and the difficulty of replacing ICs has made it more practical to just discard a bad piece of electronic equipment rather than repair it. It costs less and you get a newer and better replacement product as a result. Not a bad deal.

Yes, it is still possible to repair some equipment. For example, big expensive equipment such as a magnetic resonance imaging (MRI) machine is usually repairable, but it is done by replacing not bad components, but by replacing defective modules, subassemblies, or printed circuit boards.

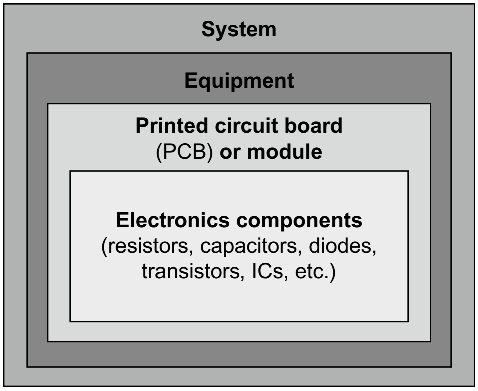

FIGURE 3.1 Hierarchy of electronics from systems to components.

You don't have to know how to analyze circuits to do that. You just troubleshoot by observing signal flow, measuring inputs and outputs, and replacing any module or subassembly that does not work. You get the equipment working sooner and at lower cost.

A good way to picture the systems view is shown in Figure 3.1. It all begins with the components that make up the circuits that in turn make up the modules and subassemblies, usually on printed circuit boards (PCBs). Those are put together to form pieces of equipment. Finally, multiple pieces of equipment may end up being interconnected to form some kind of system. A few system examples will show you the idea.

A personal computer (PC) is a great example of a system (see Figure 3.2). It is made up of modules including a mother board with its processor, memory ICs, and input/output (I/O) interfaces. There is a power supply module, a hard disk drive, a CD/DVD drive, and perhaps a video card. Then the system is completed when it is connected to the video monitor, keyboard, printer, and Internet modem.

A consumer home entertainment system is made up of a big screen HDTV set, a cable or satellite box, a DVD/Blu-ray player, and an audio amplifier/receiver with speakers. Other equipment may be involved. The TV itself or any of the other components are systems in their own right.

An iPod or MP3 player is a simple system. Shown in Figure 3.3, it consists of a memory, either a tiny hard disk drive or a solid-state flash memory that stores the music, an audio codec IC that does the digital music decompression, digital-to-analog conversion and analog-to-digital conversion for storing music, a microcontroller that controls the whole thing including the selection controls, and the LCD readout. A pair of audio amplifiers and a headset complete the system. A power management system charges the battery and doles out DC to power the various ICs.

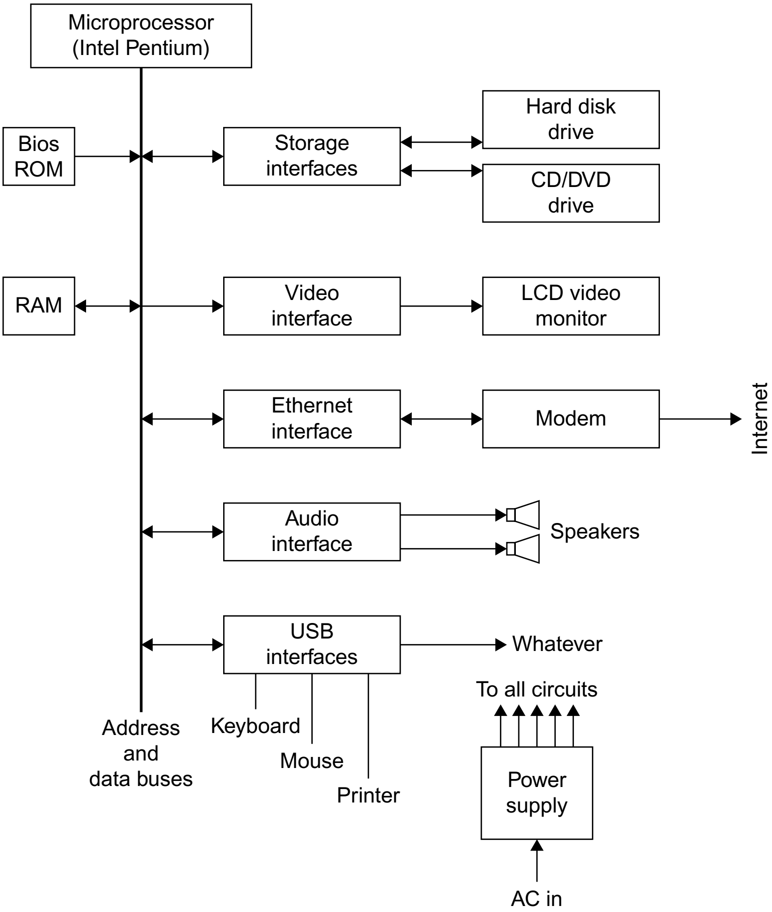
Microprocessor (Intel Pentium)
Bios ROM
Storage interfaces
Hard disk drive
CD/DVD drive
RAM
Video interface
LCD video monitor
Ethernet interface
Modem
Internet
Audio interface
Speakers
USB interfaces
Whatever
To all circuits
Address and data buses
Keyboard
Mouse
Printer
Power supply
AC in

FIGURE 3.2 System block diagram of typical personal computer.

An MRI machine is a whole system. It is made up of the central bed surrounded by a superconducting magnet and radio frequency–sensing coils, high-power radio frequency (RF) amplifiers and detectors, the video imaging section, and a computer that runs the whole thing.

Anyway, you get the picture. There are big systems, little systems, and lots of in-between systems. Their main characteristic today is that you cannot get at the individual components and circuits, but there are exceptions.

In this book we look mainly at the circuits, equipment, and systems, not the components. We will give you a quickie introduction to the main components but move on to the circuits and systems.

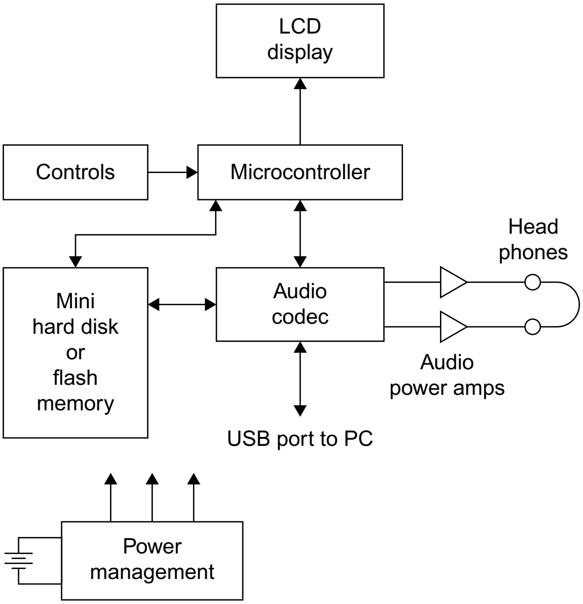

FIGURE 3.3 System block diagram of iPod or MP3 music player.

ELECTRONIC COMPONENTS

Most systems today are collections of ICs wired together on printed circuit boards and packaged in a housing of some sort. The MP3 player/iPod in Figure 3.3 is a great example. Each block in that diagram is an IC. In the past, systems were made with PCBs and other assemblies, but the circuits were made with discrete components, individual resistors, capacitors, diodes, transistors, and even some ICs. You will still see these here and there in modern equipment; just remember that most circuits and equipment are just a batch of ICs.

Just so you won't be ignorant of components, here is a nutshell introduction that should serve you well when learning electronics the system way. If you want to dig deeper, you can find many other books that will make your eyes glaze over with detail.

Switches

There is no simpler electrical component than a switch. It is just two (sometimes more) metal contacts and a mechanical device to open and close them. In one position the switch contacts are open and not touching so that no current will flow through them; the switch is "off." Moving the switch lever, the contacts are made to touch making a path for current to flow; the switch is said to be "on."

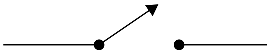

Switch open or off

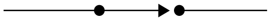

Switch closed or on

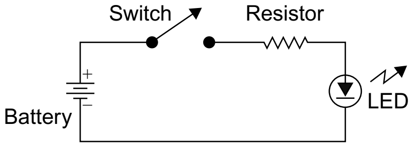

FIGURE 3.4 Switch and how it is used.

Figure 3.4 shows a switch in a circuit diagram called a *schematic*. This shows the switch connected to a battery and a light-emitting diode (LED) in a simple flashlight. The other component is a resistor to limit the current.

Resistors

Resistors are one of the most common components in electronics. These devices offer opposition to current flow; they limit or set the level of current in a circuit. The opposition to current flow is measured in ohms. The higher the ohms value, the more resistance it offers to electron flow.

Ohm's Law

There is one key electrical relationship you should know, called Ohm's law. It is simple and easy to understand. It says: The current in a circuit (I) is directly proportional to the voltage (V) and inversely to the resistance (R). It is usually expressed as a simple formula:

$$I = V/R$$

If you increase the voltage in a given resistance, the current increases. That makes sense since the voltage charge simply attracts and repels more electrons. Increasing the resistance makes the current decrease. Since resistance opposes current flow, more resistance with the same voltage will make the current decrease.

Keep this simple relationship in mind, as it is the heart of every electrical and electronic application.

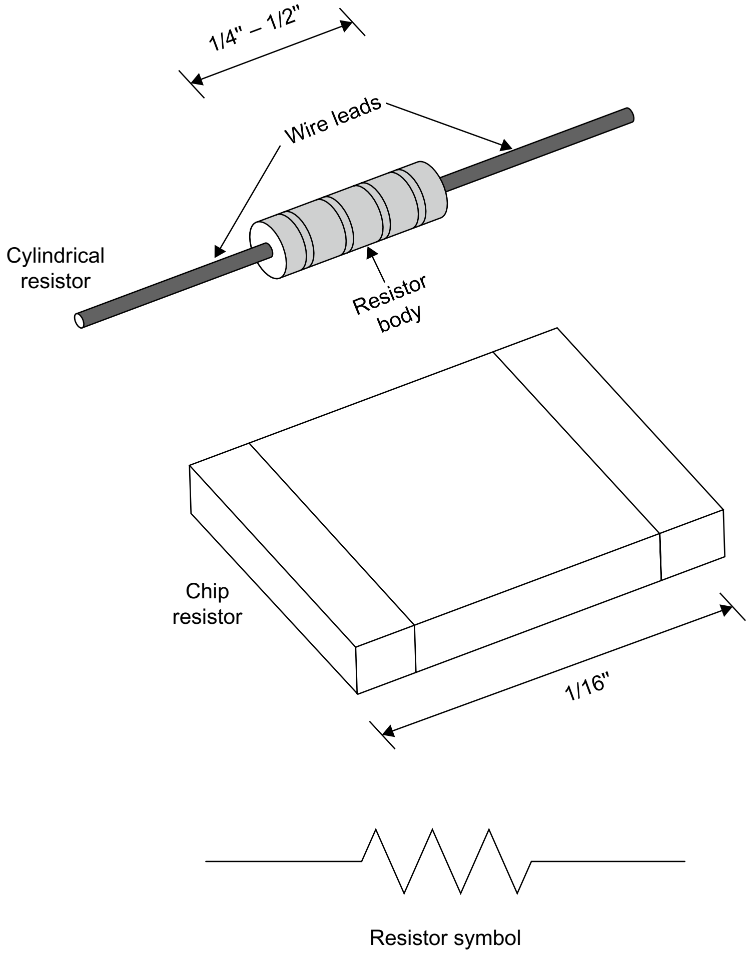

FIGURE 3.5 Types of resistors and schematic symbol.

Today most resistors are made along with all the other circuitry in an IC. But discrete component resistors are also still fairly widely used. The most common ones are called chip resistors and look something like the one shown in Figure 3.5. They are small and you can just make them out when looking at a printed circuit board. Older but still used resistors are cylindrical like the one shown in Figure 3.5. The schematic symbol used to represent a resistor in a schematic diagram is shown in Figure 3.5. Color-coded resistors use a set of color bands with a special code to indicate their value (see Figure 3.6).

An example of a common resistor circuit is the voltage divider shown in Figure 3.7. It is used to attenuate a voltage or signal or make it lower. The two resistors of the voltage divider are connected in series. When current flows, a

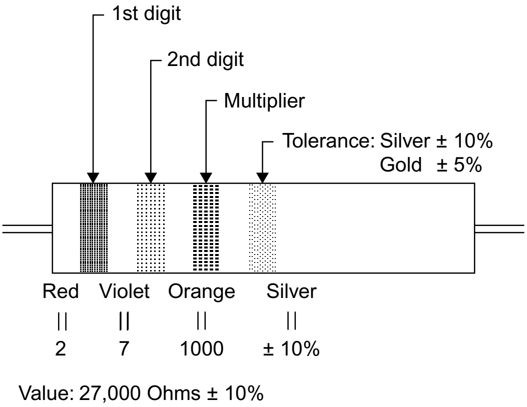
1st digit
2nd digit
Multiplier
Tolerance: Silver ± 10%
Gold ± 5%
Red = 2
Violet = 7
Orange = 1000
Silver = ± 10%
Value: 27,000 Ohms ± 10%

Resistor color code

Color of band	Value	Multiplier
Black	0	1
Brown	1	10
Red	2	100
Orange	3	1000
Yellow	4	10,000
Green	5	100,000
Blue	6	1,000,000
Violet	7	10,000,000
Gray	8	100,000,000
White	9	1,000,000,000

FIGURE 3.6 Standard resistor color code.

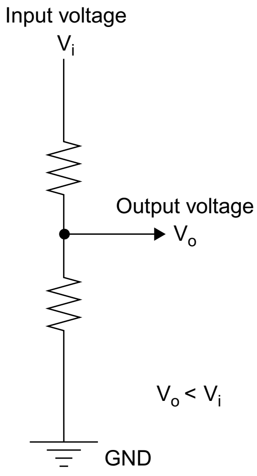
Input voltage
V_i
Output voltage
V_o
$V_o < V_i$
GND

FIGURE 3.7 Voltage divider, the most common resistor circuit.

voltage develops across the two resistors. The input voltage is divided between the two resistors according to their value. In any case, the output is taken from one of the resistors and it is lower in value than the input.

Capacitors

A capacitor is a device that stores an electrical charge. It is made up of two metallic plates separated by an insulator such as plastic or ceramic or even air (see Figure 3.8). The symbol for a capacitor in schematics is also shown.

The basic function of a capacitor is to be charged or discharged. You charge the capacitor by applying a voltage to it as shown in Figure 3.9A. The electrons pile

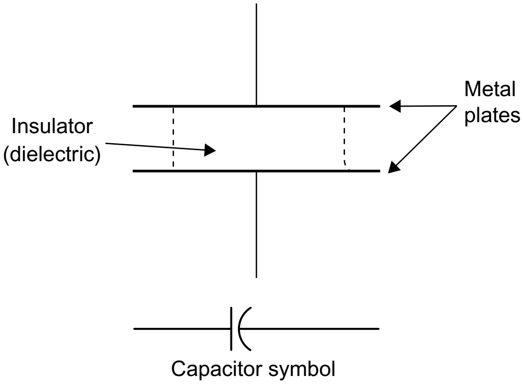
Metal
plates
Insulator
(dielectric)
Capacitor symbol

FIGURE 3.8 Capacitor construction and schematic symbol.

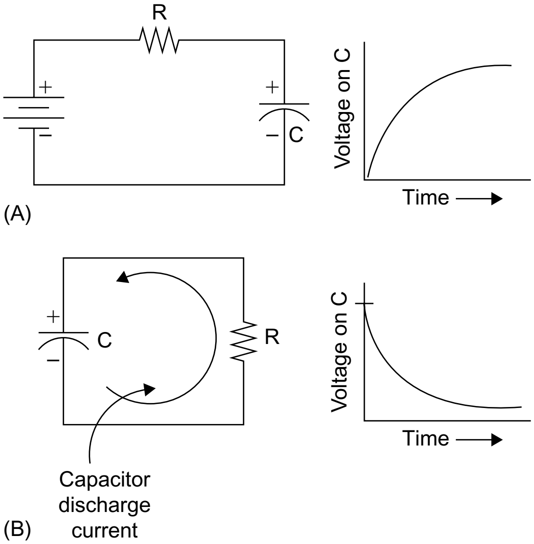
R
+
−
+
−
C
Voltage on C
Time
(A)
+
−
C
R
Voltage on C
Time
Capacitor
discharge
current
(B)

FIGURE 3.9 **(A)** Charging capacitor. **(B)** Discharging capacitor.

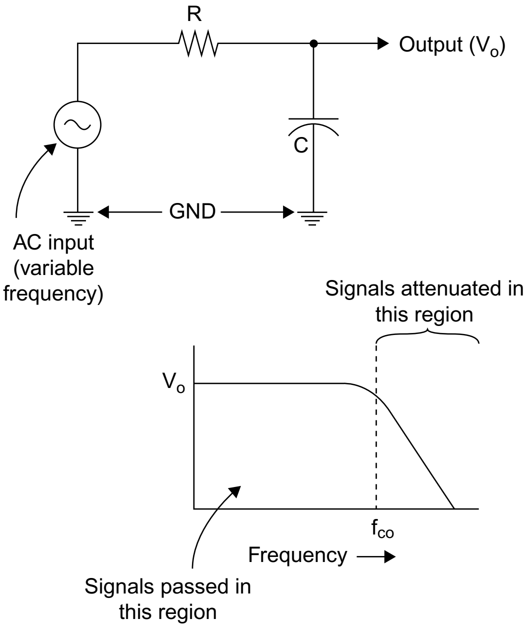
R
Output (V_o)
C
GND
AC input (variable frequency)
Signals attenuated in this region
V_o
f_{co}
Frequency
Signals passed in this region

FIGURE 3.10 Low-pass filter, a common capacitor use.

up on one plate making it negative and are pulled away from the other plate making it positive. It takes a finite amount of time for the capacitor to charge as determined by the values of capacitance (C), stated in farads (F), and resistance (R). That time is called the time constant (T). In the charged state, the capacitor stores the voltage. The capacitor in its charged state is almost like a battery, as it can supply DC voltage to a circuit or load.

The energy stored in the capacitor is used by discharging it into a load such as a resistor (see Figure 3.9B). The electrons then flow from the negative plate through the resistor to the positive plate. Current flows until the charge is dissipated. Again, it takes time for the discharge to occur.

When capacitors are used with AC signals, they offer opposition to current flow called *capacitive reactance*. It is like resistance in that it determines current flow, but it varies with the frequency of the AC and capacitor value.

Capacitors perform all sorts of functions in electronic circuits. They are used for filtering and they are used to pass AC but block DC. Figure 3.10 shows a low-pass filter comprised of a resistor and a capacitor. This is a frequency-sensitive circuit that passes AC signals below a specific frequency called the *cut-off frequency* (f_{co}) and attenuates the frequencies above the cut-off. Figure 3.11 shows how a capacitor passes AC but blocks DC.

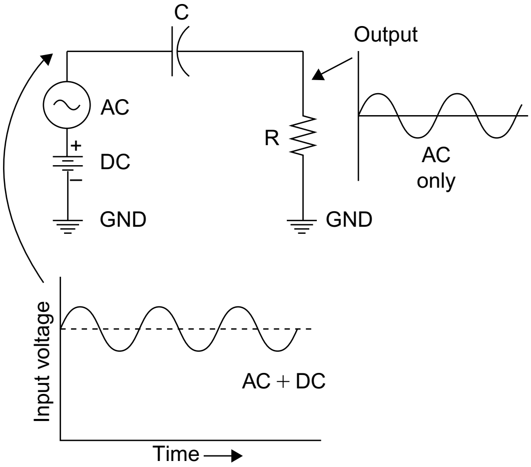

FIGURE 3.11 How capacitor passes AC sine wave but blocks DC voltage.

A Word about Ground

In schematic diagrams like Figure 3.10, you will see the triangular-shaped symbol labeled GND. This means ground. Ground is an electronic term that refers to a common connecting point. It may be a common wire, a metal chassis, a copper area on a printed circuit board, or some other conductive part or area. In this figure, it means that the bottom of the AC generator and the bottom of the capacitor are connected together.

Ground is also the common reference point for measuring all voltages.

Inductors

Inductors are mainly just coils of wire. When current flows through them, they produce a magnetic field and induce a voltage in themselves. This is called *self-induction*. The induced voltage has an opposite polarity of the applied voltage. The result is that the induced voltage causes opposition to the current flow. We often use that characteristic to control current flow in AC circuits. Like the capacitor, the inductor is an opposition to AC called *inductive reactance*. Figure 3.12 shows the symbol of an inductor as used in schematic diagrams.

Transformers

A transformer is usually comprised of two coils or windings on a common core. The schematic symbol is shown in Figure 3.13. A voltage applied to one winding called the *primary* causes current to flow and a magnetic field to develop. This magnetic field spreads out and induces voltage in the other

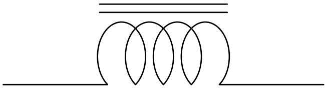

FIGURE 3.12 Schematic symbol for inductor.

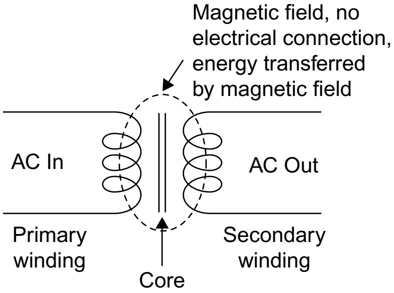
Magnetic field, no electrical connection, energy transferred by magnetic field
AC In
AC Out
Primary winding
Secondary winding
Core

FIGURE 3.13 Schematic symbol for transformer.

winding called the *secondary*, thereby transferring electrical energy for one coil to another by way of the magnetic field. Transformers are used to step up or step down voltage levels of AC signals. They are also used for signal isolation as well as impedance matching.

Diodes

A diode is a semiconductor component that lets current flow through it in one direction but blocks current in the other direction depending on the polarity of the voltage applied to it. It acts like a polarity-sensitive switch. Figure 3.14 shows a typical diode, the diode schematic symbol, and how it is biased for current or no current.

Diodes are used mainly for rectification, the process of converting AC into DC. Figure 3.15 shows how a diode converts an AC sine wave into DC pulses. If you put a capacitor across the load resistor, it will charge up to the peak sine voltage and store it. The result is that the output is a near-constant DC value. Most electronic power supplies work like this.

Special diodes are made to emit light (LED), regulate voltage (zener diode), act as a variable capacitor (varactor), or perform as a switch (PIN diode).

Transistors

A transistor is a three-terminal semiconductor device that uses a small input signal to control a much larger output signal. One type of transistor is the bipolar junction transistor (BJT) shown in Figure 3.16. A small current applied to the base element is used to control a larger current flowing from emitter to collector.

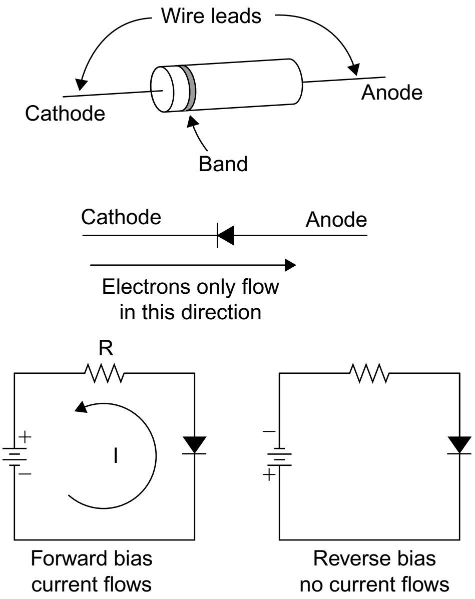
Wire leads
Cathode
Anode
Band
Cathode
Anode
Electrons only flow in this direction
R
+
−
I
Forward bias current flows
−
+
Reverse bias no current flows

FIGURE 3.14 Diode, its schematic symbol, and how to bias it for conduction or cut-off.

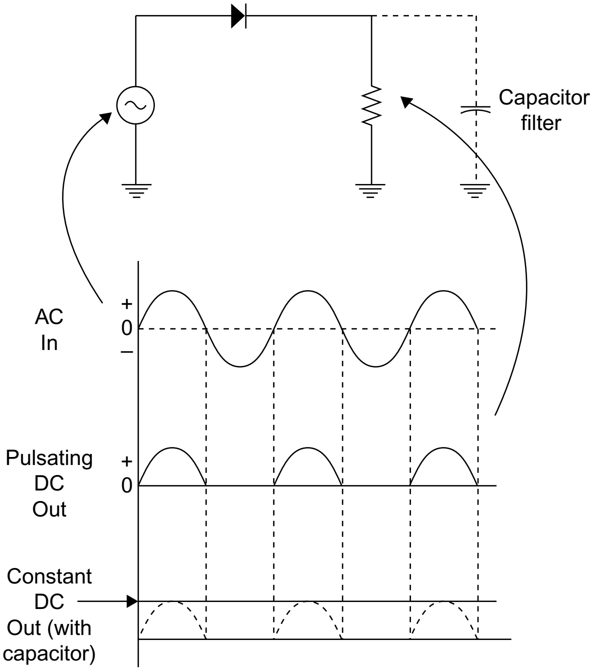
Capacitor filter
AC In
+
0
−
Pulsating DC Out
+
0
Constant DC Out (with capacitor)

FIGURE 3.15 How a diode rectifies AC into DC.

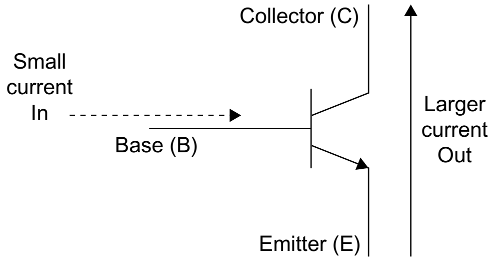

FIGURE 3.16 Schematic symbol for bipolar junction transistor (BJT) and its inputs and outputs.

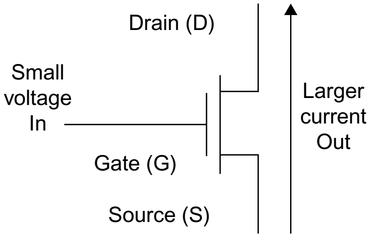

FIGURE 3.17 Schematic symbol for MOSFET and its input and outputs.

Another type of transistor is the metal oxide semiconductor-field-effect transistor (MOSFET). It too has three elements, as shown in Figure 3.17. A small voltage on the gate controls a larger current flowing from source to drain.

Transistors have two basic functions: amplification and switching. In amplification, the small input variation causes larger output current variation. The small input signal isn't really made bigger, but the larger output variation has the same shape and characteristics, resulting in the same effect. The transistor simply varies the larger DC current from the supply and shapes it like the input.

Amplifiers

Figure 3.18 shows a simple BJT amplifier. The base resistor sets the base current turning the transistor on so that current flows from emitter to collector. Its output voltage is a DC value at about half the supply voltage. Now, if a small AC signal is applied through the capacitor as shown, the base current will vary. This will vary the collector current and the voltage across the collector resistor. A small variation in the base current produces a larger variation in the collector current. The output voltage is an enlarged version of the input, except that it is reversed in phase. We say that the output is 180 degrees out of phase with the input, called *phase inversion*. The output is a DC signal, but to convert it to AC we pass it through a capacitor to the load. All amplifiers generally work like this.

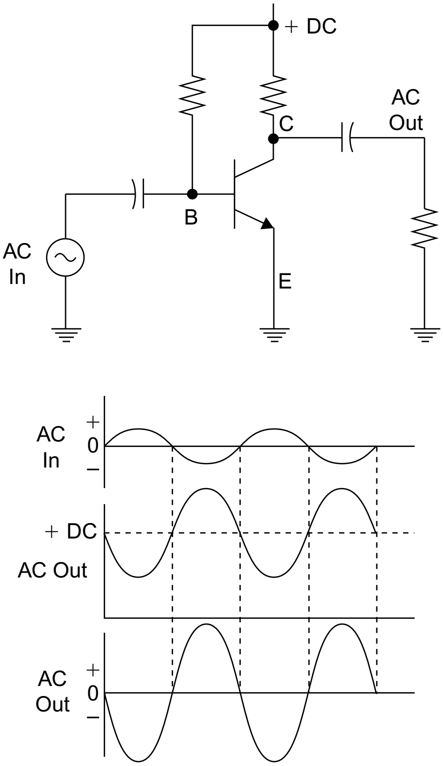

FIGURE 3.18 Simple BJT amplifier and how it works.

Switches

As a switch, the transistor output current is turned off or on by the smaller input current or voltage. A small input signal can switch a larger output current off and on. Switching transistors can switch very fast (in less than nanoseconds). Figure 3.19 shows a MOSFET switch. If the input voltage on the gate is zero, the MOSFET does not conduct. It acts like an open or off switch. The output then is just the supply voltage (+V) as seen through the resistor.

Now, if we apply voltage to the gate above a certain threshold level, the MOSFET will conduct. It acts like a closed or on switch so current flows in the resistor. The output is a very small value near zero. Figure 3.19 also shows what the inputs and outputs look like with a square wave applied.

Integrated Circuits

An IC is also called a chip. It is a complete circuit made on a single chip of silicon. All the components are made at the same time and interconnected to form an amplifier, microcomputer, memory, oscillator, radio transceiver, or

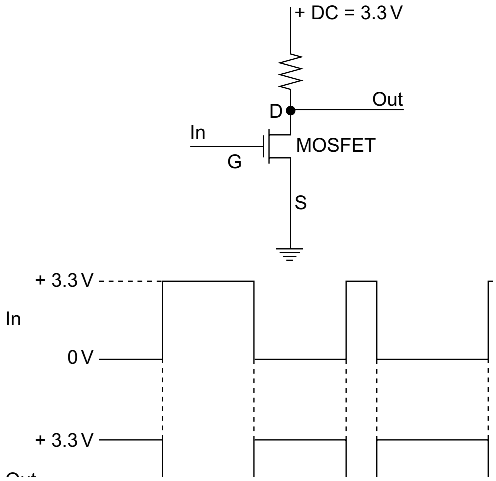

FIGURE 3.19 Simple MOSFET switch and how it works.

MDM9600
QUALCOMM

FIGURE 3.20 Typical integrated circuit today. This one is a dual-satellite TV tuner for set-top boxes. The size is only 7 × 7 mm. The pins are soldered to a printed circuit board. *(Courtesy NXP Semiconductor.)*

whatever the function. The IC is then connected to DC power and it works. All you do is apply the inputs, and new outputs are produced according to what the IC function is. Most circuits today are ICs.

There are all sizes and types of ICs. There are linear circuits for analog signals, digital circuits for binary signals, and mixed signal IC for both analog and digital signals. There are small ICs that are just an amplifier or logic circuit. Then there are even larger circuits like a computer memory or a microprocessor that is the heart of every computer. Finally, there are complete systems on a chip (SoC). Each year, it is possible to make the components on a silicon chip smaller so you can put more on a chip economically. That is why we have iPods, cell phones, and other really small devices that in the past would have been huge pieces of equipment. Figure 3.20 is representative of ICs today.

There are lots of other electronic components but these are the main ones. A few others are introduced in later chapters.

While this book attempts to get you thinking about the big picture—the systems rather than the individual components—we cannot ignore discrete components completely. Well, actually we can, but you may be curious and want to have a bit more involvement with them. If so, the projects to follow will show you how individual parts are used. The circuits presented here are trivial compared with the sophisticated technology you use every day, such as big-screen LCD HDTV sets, laptops, cell phones, and your wireless Internet broadband router. Feel free to skip these if you wish.

Project 3.1

Setting Up a Simple Lab

If you want to play around with some components just to get the feel for how it used to be, you need to set up a simple lab. If you performed the previous projects, you already have a multimeter. Now you will need a way to interconnect parts in circuits and a power source.

The power source can be four AA cells in a holder. This will provide DC voltage to power the circuits. Or you can use a common 9-volt battery. A battery clip with leads will allow you to connect the battery to the circuit.

Next you will need a breadboard. This is a connector socket with multiple holes in it like that shown in Figure 3.21. You plug the wire leads of the components into the holes to connect them to one another without soldering them.

Then you will need some components such as resistors, capacitors, and transistors. In this project, the goal is to get the battery and the breadboarding socket. The remaining projects for this chapter are circuit experiments.

You can get the breadboard from a local Radio Shack or from a mail order house like Jameco. There are usually two sizes, large and small. The small one is adequate. Figure 3.21 shows a larger one.

Get a battery holder that will take four AA batteries. Again, Radio Shack or Jameco are good sources. Some battery holders also have an on/off switch but it is not necessary. They will have two leads for connecting the battery to the circuit. The 9-volt battery clip is shown in the figure.

Once you get the parts, put the AA cells into the holder so that you connect the negative end (−) of one battery to the positive end (+) of the next so that the cells will be in series. That way their voltages will add together. The circuit should look like that in Figure 3.22.

Next, test the battery. Using your multimeter on the DC ranges, and measure the voltage across the battery holder wires. What do you get? You should read 6 volts or a bit more. The 9-volt battery should, of course, measure 9 volts.

Now you are ready for a few additional component projects.

(*Continued*)

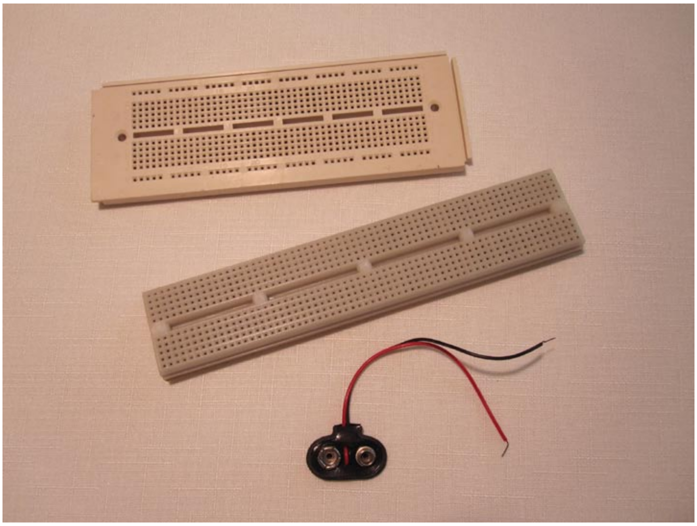

FIGURE 3.21 Common breadboarding socket for building prototypes or just experimenting. *(Courtesy Global Specialties.)*

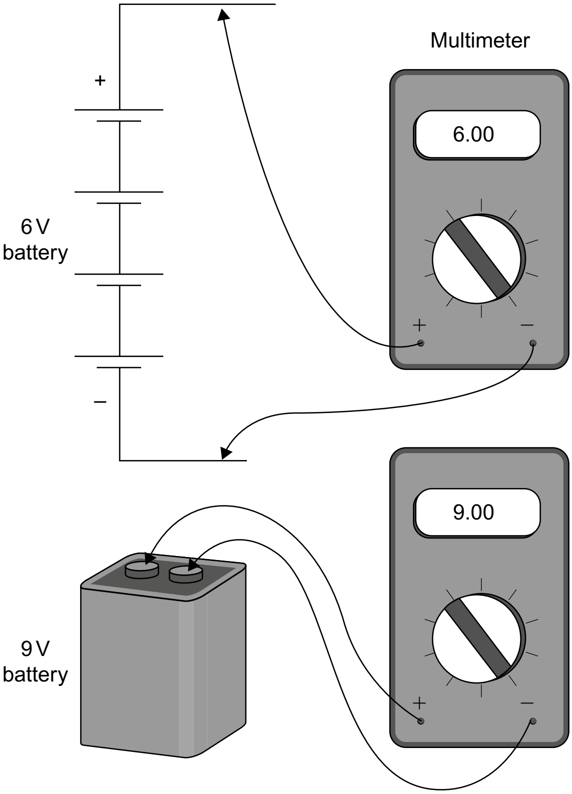
Multimeter
+
6 V
battery
–
6.00
+
–
9.00
9 V
battery
+
–

FIGURE 3.22 Measuring battery voltage with multimeter.

Project 3.2

Light Emitting Diode

When you are buying your parts, get an LED and a set of resistors. Any old common red or green LED will work. The resistor set will be a package of all the most common values of resistors.

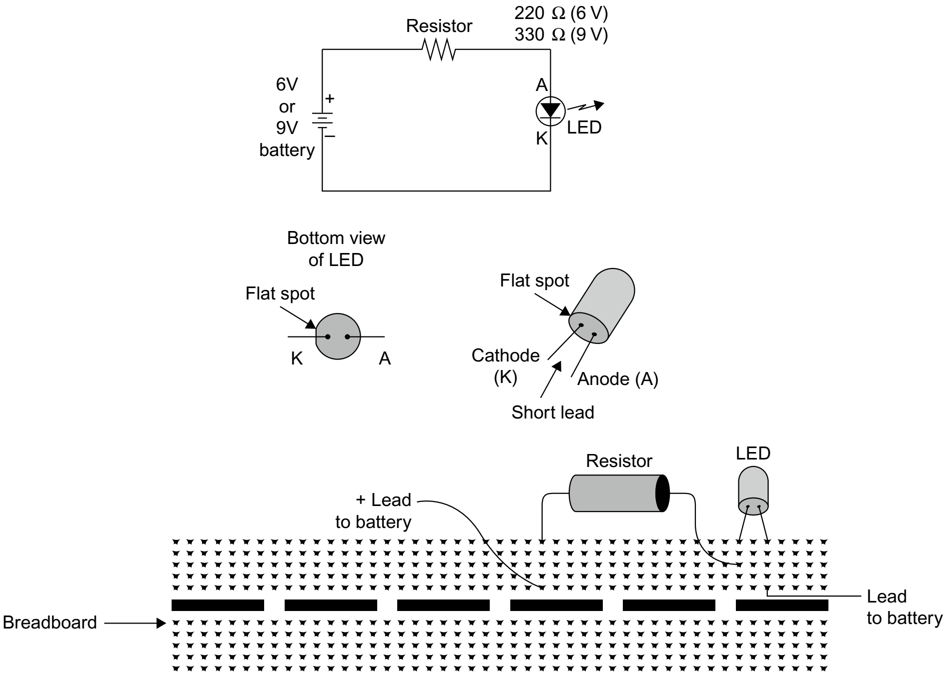
Resistor
220 Ω (6 V)
330 Ω (9 V)
6V
or
9V
battery
+
–
A
K
LED
Bottom view
of LED
Flat spot
K
A
Flat spot
Cathode
(K)
Anode (A)
Short lead
Resistor
LED
+ Lead
to battery
Lead
to battery
Breadboard

FIGURE 3.23 This is the circuit used to light an LED and how to wire it on the breadboard.

(*Continued*)

Project 3.2 (Continued)

Now, wire up the circuit in Figure 3.23. A pictorial is shown in the figure as well. Be sure to get the LED leads connected correctly, for if not, you will zap it. See the wiring details in the figure.

The resistor is needed to control or limit the current so that you do not blow out the LED. A value of 220 ohms is a good one if you use the 6-volt battery. The color code for that resistor is red-red-brown-gold. If you are using the 9-volt battery, use a 330-ohm resistor (orange-orange-brown-gold).

Wire the circuit and connect the battery. The LED should light. Now you are an electronic circuit genius.

Project 3.3

Capacitors

Capacitors store a charge and then release it. The charge is stored energy. You charge the capacitor by applying a voltage to it. Then you discharge the capacitor by connecting it to a load like a resistor. This brief project will illustrate that idea.

You will need the following parts from Radio Shack or another source. Get a 100-K ohm resistor (color code brown-black-yellow-gold) and a 100-μF (microfarad) electrolytic capacitor. This capacitor is polarized, meaning that you have to observe the polarity of the voltage applied to it. The wire leads are usually marked with either a + sign, or sometimes with a − sign.

Wire the circuit shown in Figure 3.24 on the breadboarding socket. Do not initially connect the positive lead of the battery to the resistor. Connect your multimeter across the capacitor. Now connect the 6 volts (or 9 volts) from the battery supply as indicated when you are ready to start charging the capacitor. You should see the voltage begin to rise slowly. It should take about 50 seconds for the capacitor to fully charge to 6 (or 9) volts.

Now disconnect the battery's positive lead. The capacitor retains the charge, as you should see on the meter. It will probably begin to drop slowly as it discharges into the meter.

Now use the capacitor as a voltage source. Connect the capacitor to the resistor and LED you used earlier as shown in Figure 3.24. The capacitor will discharge quickly and you should only see the LED blink briefly. To see it again, recharge the capacitor and repeat.

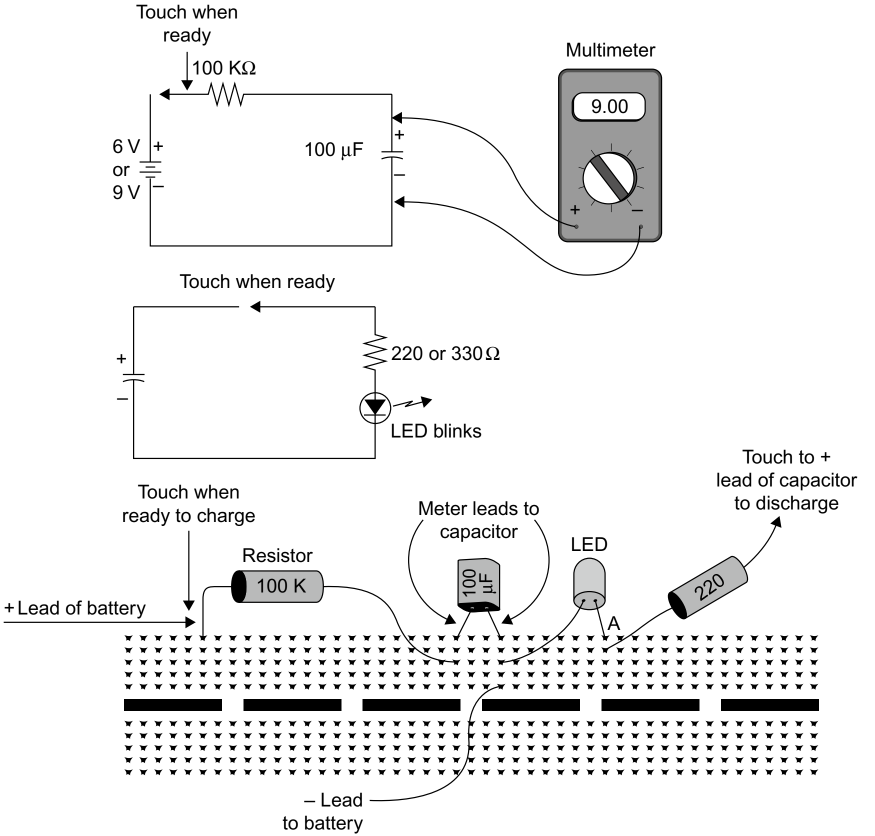
Touch when ready
100 KΩ
6 V or 9 V
+
–
100 µF
Multimeter
9.00
Touch when ready
220 or 330 Ω
LED blinks
Touch when ready to charge
+Lead of battery
Resistor
100 K
Meter leads to capacitor
100 µF
LED
A
220
Touch to + lead of capacitor to discharge
– Lead to battery

FIGURE 3.24 Circuits for charging and discharging capacitor and respective wiring.

Project 3.4

Transistor Switch

A transistor makes a good switch, especially a MOSFET. And it is easy to use. This project demonstrates how the MOSFET can be used to turn the LED off and on.

Acquire a MOSFET, such as the IRF510 or its equivalent. Then build the circuit shown in Figure 3.25. Be sure to follow the guidelines for wiring the transistor. Use the LED and 220- (or 330-) ohm resistor that you used before.

Now, connect the battery. Touch the wire from the gate on the MOSFET to the +6 volt or +9 connection like on the resistor. This will turn the transistor on and it will act as a switch to turn on the LED. Disconnect the gate from +6 volts and the transistor switch goes off, turning off the LED.

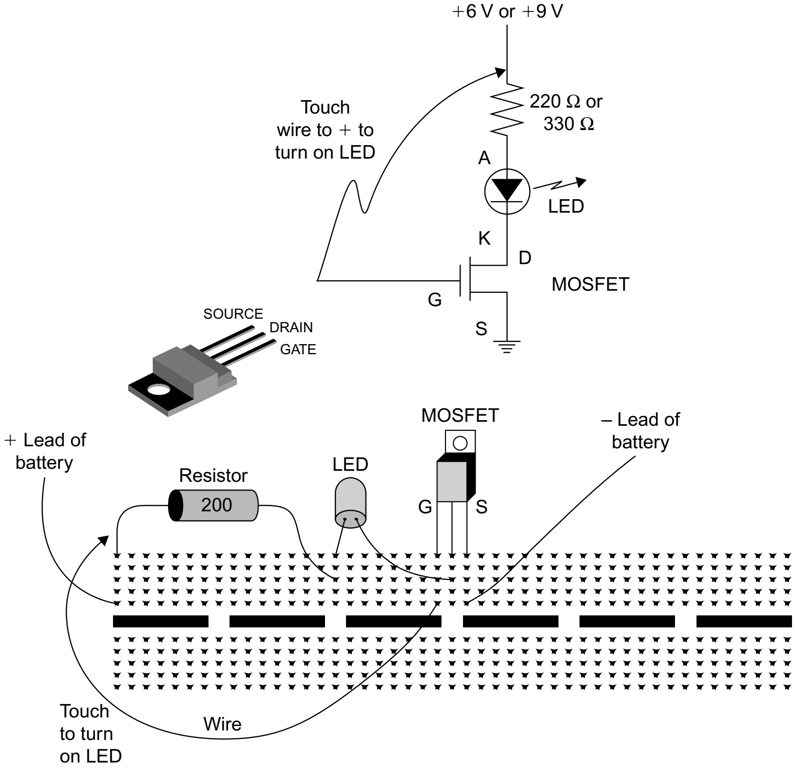
+6 V or +9 V
Touch wire to + to turn on LED
220 Ω or 330 Ω
A
LED
K
D
MOSFET
G
S
SOURCE
DRAIN
GATE
MOSFET
– Lead of battery
+ Lead of battery
Resistor
200
LED
G
S
Touch to turn on LED
Wire

FIGURE 3.25 Circuit that turns an LED off and on with MOSFET and wiring.

Chapter 4

Electronic Circuits: Linear/Analog

The Building Blocks of Electronic Equipment

In this Chapter:

- Linear circuits defined.
- Types of linear circuits.
- Characteristics and specifications of the most common linear circuits.

INTRODUCTION

The basic building blocks of electronic equipment are individual circuits made up of transistors, resistors, capacitors, and sometimes transformers, inductors or other components. It used to be that circuits were made up of individual discrete components that were wired together on a printed circuit board. But today, most electronic circuits are in integrated circuit (IC) form or more likely multiple circuits are already combined on a single silicon chip to create a larger circuit, sub-system or system on a chip (SoC).

While most circuits are ICs there are still some cases where discrete components are necessary or desirable. The most common examples are high voltage or high power circuits that would burn up a tiny silicon chip. Discrete components have not gone away entirely and you will find some of them external to an IC such as larger capacitors or inductors that just cannot be easily integrated onto a chip or transistors that can handle higher voltage or power. As mentioned before, we are not going to talk about the make-up or operation of individual circuits, Instead, this book treats circuits as just building blocks with an input and an output that process the input in some way to create a new, different output.

LINEAR CIRCUITS

There are two basic types of circuits, linear and digital. Digital circuits are covered in the next chapter. Linear circuits are those circuits that process analog signals. Recall that analog signals are those that vary smoothly and continuously

© 2010 Elsevier Inc. All rights reserved.
doi: 10.1016/B978-1-85617-700-9.00004-7

over time in contrast to pulse and digital signals that are voltages that switch off and on very fast. Analog signals are like sine waves, voice signals, video signals, radio waves, and voltages from sensors like a temperature transducer. DC, either a constant value of voltage or one that is varying, is also an analog signal.

A linear circuit is one whose output is proportional to its input. See Figure 4.1. Its so-called response is a straight line or what a mathematician would call a linear function. The best example of this is the amplifier where its output is proportional to its input so that it accurately reproduces the input faithfully but at a higher voltage or power level. In the figure, with 0.4 volt input you get 4 volt output. That means the amplifier has a gain of 10. Linear circuits do not distort the signal. This chapter covers linear circuits with most of the emphasis on amplifiers since that is the largest category of linear circuits. And as it turns out, most other types of linear circuits are just amplifiers connected in different ways to give different processing results.

Amplifiers

The basic block diagram symbol for an amplifier is a triangle. See Figure 4.2. It has an input and an output. Amplifiers also have one or more DC inputs as well. This DC from a battery or other power supply is what powers the circuit

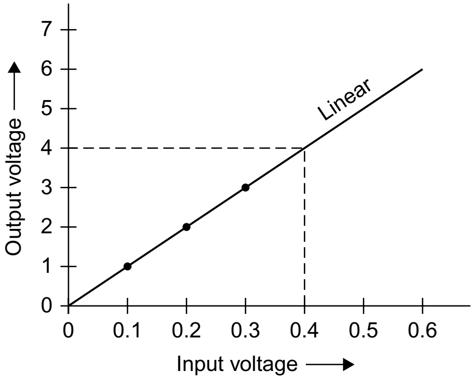

FIGURE 4.1 Linear circuit has a straight-line response. As input varies, output varies by factor of 10 larger, producing straight-line output.

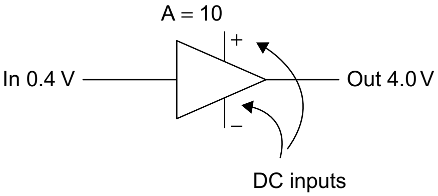

FIGURE 4.2 Block symbol for amplifier. Gain, A, is often given. DC power inputs are not usually shown.

and it is this DC that is ultimately converted into the new, larger output signal. In most cases the DC inputs are not shown in a block diagram. The amplifier processes the input to create an output of the same shape but at a larger amplitude.

The primary characteristic of any amplifier is its *gain*. The gain (usually represented by A) is simply a number equal to the ratio of the output to the input.

$$A = \frac{\text{Output}}{\text{Input}}$$

If the output is 4 volts and the input is 0.4 volt, the gain is simply:

$$A = \frac{4}{0.4} = 10$$

The amplifier multiplies the input by 10 to get the output.

$$\text{Output} = 10 \times \text{Input}$$

Classifying Amplifiers

The amplifier is the most common linear circuit. There are so many different types of amplifiers that we divide them up into different types to distinguish one from another. And there are multiple ways of doing this. Here are the most common.

Small Signal versus Large Signal

Small signal amplifiers are those that, as the name implies, only amplify small signals which are those roughly below about 1 volt. Small signal amplifiers amplify millivolt, microvolt, or nanovolt signals. Large signal amplifiers amplify larger signals and typically amplify the power of a signal rather than the voltage. These are called power amplifiers (PAs).

Frequency Classification

Amplifiers are also categorized by the frequency of the signals they amplify. Some examples are DC amplifiers, audio frequency (AF) amplifiers, intermediate frequency (IF) amplifiers, video amplifiers, radio frequency (RF) amplifiers, and microwave amplifiers.

Configuration Classification

The two main configurations are single-ended and differential. A single-ended amplifier is one in which the input and output are referenced to a common ground. These amplifiers have two inputs, one is the signal and the other is ground. See Figure 4.3A. Normally only the input is shown as the ground is just assumed to be there.

A differential amplifier has two inputs as Figure 4.3B shows. These are usually called the inverting (−) and the non-inverting (+) inputs. Both inputs are referenced to ground. The amplifier is called a differential amplifier because it actually takes the mathematical difference between the two inputs and amplifies the result. One input is subtracted from another, then the difference is amplified by the gain A. The formula for this operation is:

$$\text{Output} = (\text{Input } 2 - \text{Input } 1) \times A$$

Differential amplifiers have one neat feature in that any signal common to both inputs like noise is automatically canceled out. This feature is widely taken advantage of in very small signal amplifiers used in memory chips and high gain amplifiers that amplify small signals from sensors. Noise that is picked up simultaneously on both inputs simply subtracts itself out without bothering the desired small difference signal.

Classification by Class of Operation

One way to categorize amplifiers is to designate the conduction of the transistors in the amplifiers. The primary classes are A, B, AB, C, D, E, and F. There are a few others but these are the most common. Class A amplifiers conduct continuously. We say that the amplification occurs over the full 360 degrees of a sine wave cycle. See Figure 4.4. Note: The small t on the waveforms means time as the voltage is varying over time.

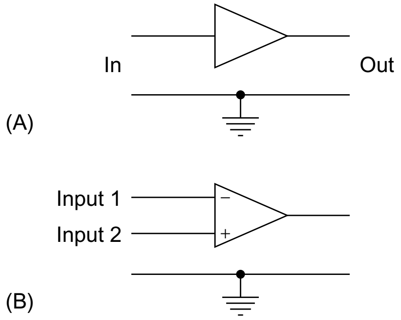

FIGURE 4.3 Single-ended (**A**) and differential input (**B**) amplifiers. Inputs and outputs are referenced to ground.

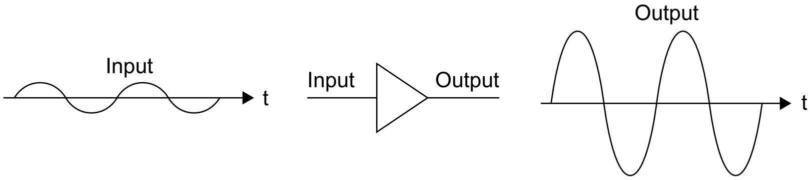

FIGURE 4.4 Class A amplifier conducts continuously.

Class B amplifier transistors only conduct for half a cycle or 180 degrees of a sine wave. See Figure 4.5. Such an amplifier seems useless since it distorts the signal. But if you combine two transistors, one to amplify positive cycles and the other to amplify negative cycles, then combine the two, the complete wave is amplified. Such an amplifier is called a push pull amplifier. The benefit of this arrangement is that the circuit is more efficient.

Class AB amplifiers are a variation of the class B. Class B amps do add a little distortion to the signal around the point where the sine wave crosses zero. This is caused by the transistors not turning on and off precisely at the zero point. This glitch shown in Figure 4.5 is called crossover distortion. To get rid of that, class AB amplifiers allow some small current to flow continuously. The result is less distortion but a little less efficiency.

Class C amplifiers are for RF signals and are usually power amplifiers. They only allow the transistor to conduct for less than 180 degrees of a sine wave input, usually 90 to 150 degrees. It produces massive distortion. However, the amplifier output is made up of an inductor and capacitor (LC) tuned circuit that oscillates at the operating frequency and removes the distortion. These amplifiers are very efficient.

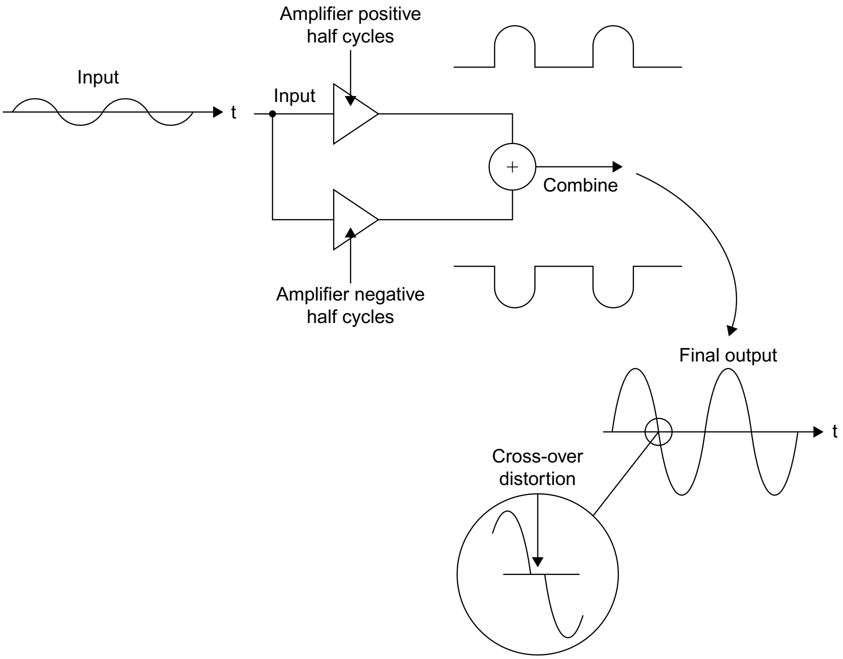
Amplifier positive half cycles
Input
t
Input
+
Combine
Amplifier negative half cycles
Final output
t
Cross-over distortion

FIGURE 4.5 Class B amplifier conducts half the time, making it more efficient.

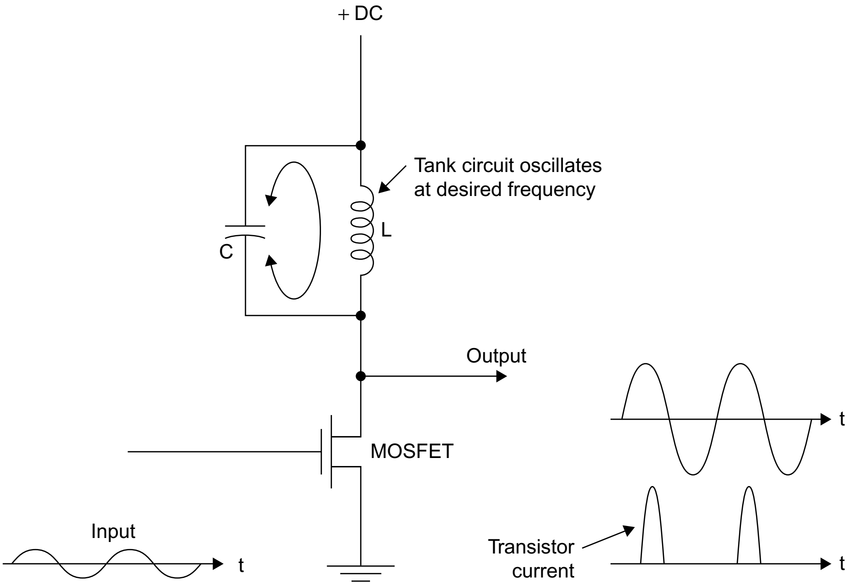

FIGURE 4.6 Class C amplifier only conducts for part of a half cycle, but the current pulse stimulates resonant tank circuit that oscillates at desired frequency.

In a class C amplifier the MOSFET acts like a switch that is turned off and on by the input signal, usually a sine wave. See Figure 4.6. When the MOSFET is on the capacitor charges up to the DC supply voltage and current flows in the inductor creating a surrounding magnetic field. When the transistor turns off, the inductor and capacitor begin to exchange energy and set up an oscillation at the resonant frequency of the LC circuit. This is called a tank circuit. The result is that the energy stored in the tank circuit produces the amplified sine wave output.

Another way of looking at this is that the transistor switch distorts the input creating a pulse waveform that contains lots of harmonics. The tank circuit acts like a selective band-pass filter that lets the fundamental sine wave pass while the harmonics are filtered out.

Don't forget the Fourier theory and harmonics. Remember that a harmonic is an integer multiple of a fundamental sine wave. If the fundamental sine wave is 20MHz, the 2nd harmonic is 40MHz, the 3rd harmonic is 60MHz, the 4th harmonic is 80MHz, and so on. Any non-sine wave signal, like a half sine wave, square wave, or pulse is said to be made up of the fundamental sine wave added to multiple harmonics. A distorted sine wave has the fundamental sine wave in it but also harmonics. This concept is called the Fourier theory after the French mathematician.

Class D amplifiers are a special class called a switching amplifier. It is made up of transistor switches rather than a real linear amplifier. This amplifier chops the input analog signal into high-frequency pulses with a varying width. See Figure 4.7. This process is called pulse width modulation (PWM).

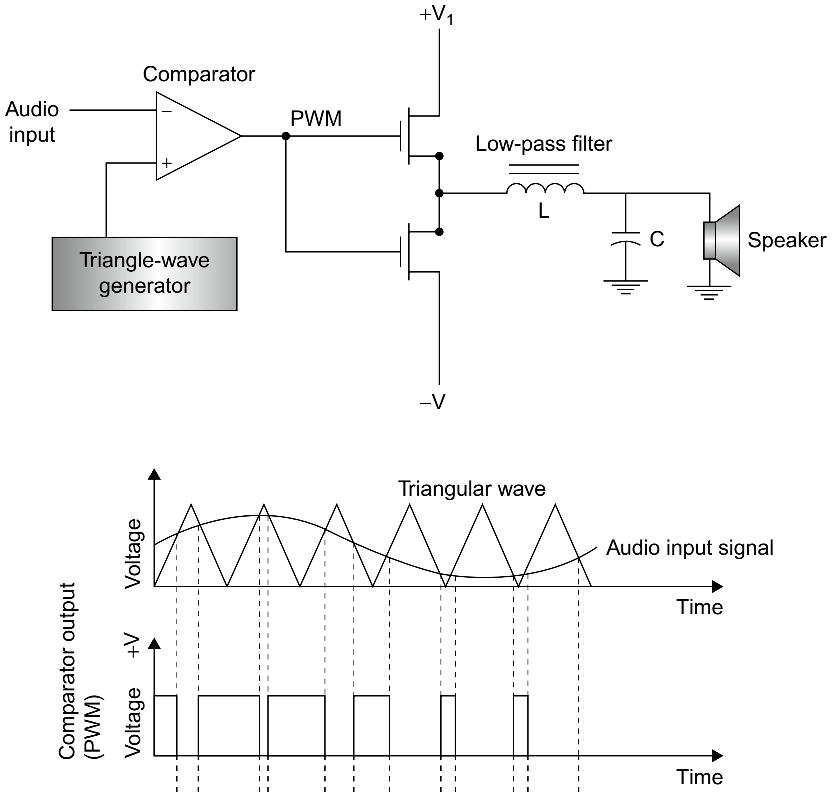
+V$_1$
Comparator
Audio input
−
+
PWM
Low-pass filter
L
C
Speaker
Triangle-wave generator
−V
Triangular wave
Audio input signal
Voltage
Time
Comparator output (PWM)
+V
Voltage
Time

FIGURE 4.7 Class D amplifier uses PWM at higher frequency to switch higher voltages for greater output. Class D amplifiers are more efficient than any other type, and thus do not produce as much heat.

The sine wave audio signal to be amplified is sent to one input of a special amplifier called a comparator along with a higher-frequency triangle wave. The comparator output switches when the triangle wave and sine wave values are equal. The resulting PWM signal is then fed to MOSFET switches to make the signal bigger. The higher amplitude output it then filtered back into an analog signal by a low-pass filter made up of a capacitor and inductor. Most class D amplifiers are audio power amplifiers whose load is a speaker or headphone. For low power less than a few watts, all the circuitry is in a single IC. For higher power, the MOSFETs are larger external devices.

The big benefit of switching amplifiers is that for a given output power they are very efficient. Whereas a class AB amplifier may only be 20 to 30% efficient, the class D is over 90% efficient. This means that the amplifier can be smaller and use less power and dissipates less heat. Class D amplifiers are great for portable battery-operated devices like MP3 players and cell phones.

Classes E and F are special switching amplifier variations used at RF frequencies only. Like the class C amplifier, they use LC circuits to filter out the harmonics they produce. They are very efficient.

Operational Amplifiers

There are many special types of amplifiers. The most widely used is the operational amplifier or op amp. This is a versatile high-gain DC differential amplifier that can be configured with external resistors, capacitors, diodes, etc. to perform a wide variety of amplification functions. The op amp may be the most widely used type of amplifier because of its versatility.

Figure 4.8 shows the most common op amp configurations. In most cases, the resistors are external to the IC amplifier. Note that the gain is determined by the values of the external resistors. Note formulas for gain (A).

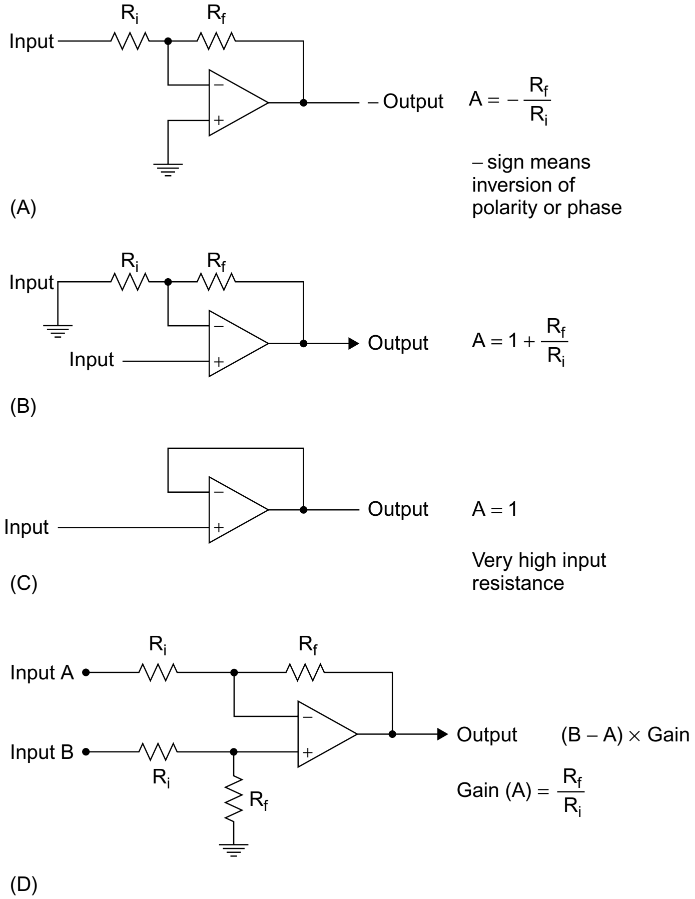

R_i
R_f
Input
–
+
– Output
$A = -\frac{R_f}{R_i}$
– sign means
inversion of
polarity or phase
(A)
R_i
R_f
Input
Input
–
+
Output
$A = 1 + \frac{R_f}{R_i}$
(B)
Input
–
+
Output
A = 1
Very high input
resistance
(C)
R_i
R_f
Input A
Input B
R_i
R_f
–
+
Output
(B – A) × Gain
Gain (A) = $\frac{R_f}{R_i}$
(D)

FIGURE 4.8 Most commonly used op amp circuit configurations. **(A)** Inverting amplifier. **(B)** Noninverting amplifier. **(C)** Follower. **(D)** Differential amplifier.

A variation of the op amp is the instrumentation amp. See Figure 4.9. It is a combination of multiple op amps that make it very useful in industrial applications such as amplifying small signals from sensors. Again the gain is determined by the resistor values. You can make this amplifier from individual op amps but usually you just buy it as a single integrated circuit. The external resistor sets the gain.

A variable or programmable gain amplifier is one whose gain can be controlled from an external DC input.

Amplifier Specifications

While you really do not need to know much about what the circuitry in an amplifier IC is, you do need to understand the specifications. Here are the most common specifications to look for.

As mentioned before, gain is the primary amplifier characteristic. It is usually a voltage or power ratio, depending on the type of amplifier.

$$\text{Voltage gain is output voltage over input voltage: } Av = \frac{Vo}{Vi}$$

$$\text{Power gain is output power over input power: } Ap = \frac{Po}{Pi}$$

Many amplifiers have their gain expressed in decibels (dB). A dB is a gain measurement based on a logarithmic scale rather than a linear scale. This has the advantage of being able to express large and small gains in smaller numbers.

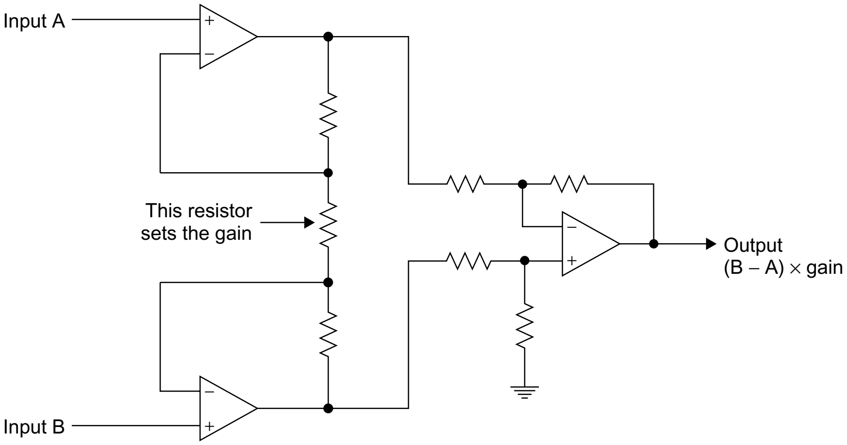

FIGURE 4.9 Instrumentation amplifier widely used to amplify small sensor signals in industrial applications.

And gains of multiple stages can be added rather than multiplied. To calculate dB gains you used these formulas:

$$\text{Voltage gain: dB} = 20 \log (Av)$$

$$\text{Power gain: dB} = 10 \log (Ap)$$

In these formulas, log means the common (base 10) logarithm of the voltage or power gain. The logarithm (log) is easy to find using a scientific calculator.

For example, a voltage gain of 50 expressed in dB is:

$$\text{dB} = 20 \log 50 = 20\,(1.7) = 34\ \text{dB}$$

A power gain 300 expressed in dB is:

$$\text{dB} = 10 \log 300 = 10(2.5) = 25\ \text{dB}$$

Input Impedance

Input impedance is the load seen by the signal source to the amplifier. It is usually resistive but typically also has some capacitance across it. Figure 4.10A shows an equivalent circuit of an amplifier. Note the input resistance. That is the input impedance. It may be as high as several megohms or could be as low

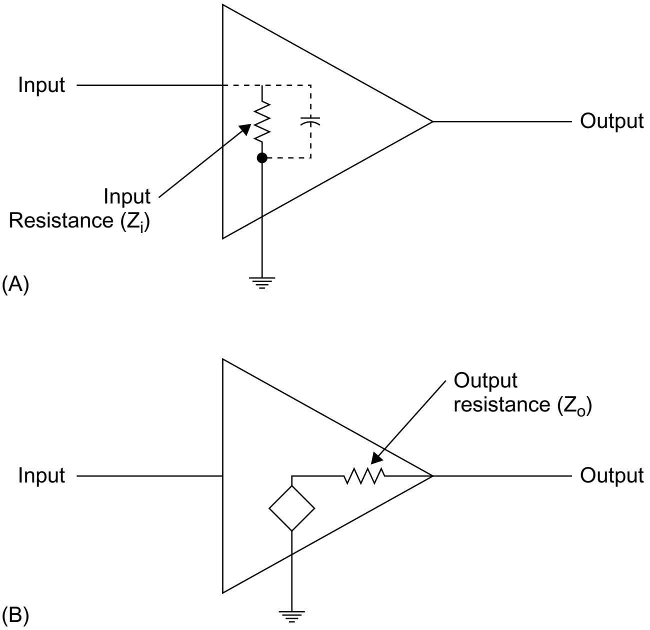

FIGURE 4.10 Amplifier input impedance **(A)** and output impedance **(B)**.

as, say, 50 ohms. Usually the higher the better for voltage amplifiers but not for power amplifiers.

Impedance

The term *impedance* means opposition to current flow. It is actually the combined opposition offered by resistance plus any capacitive or inductive reactance. If there is no reactance, the impedance is just the resistance so the two terms can be used interchangeably. Impedance is represented by the letter Z.

Output Impedance

Amplifiers act as signal generating sources because they essentially just create a new higher voltage or higher power signal with the same shape as the input. They appear as a signal generator to the next circuit or load. Figure 4.10B shows the amplifier as a generator shaped as a diamond. It has an internal impedance or series resistance called the output impedance. It is not a real resistor but it is there and it appears in series with whatever load is connected which may be another amplifier or something else.

Figure 4.11 shows the amplifier connected to another amplifier. Note that the output impedance for amplifier A_2 and the load form a voltage divider so some of the amplification is lost to the voltage divider effect. The input impedance of A_2 is the load for the output impedance of A_1.

Cascading Amplifiers

If an amplifier does not have enough gain, then multiple amplifiers can be cascaded one after the other to give the needed gain. One amplifier amplifies the output of the next and so on. See Figure 4.12. To calculate the overall gain of an amplifier you just multiply the gains as shown in the figure. This procedure assumes that one amplifier does not load the next. That is, the output impedance of each amplifier is assumed to be zero and the input impedance of the next amplifier is assumed to be infinite. That way, there is no voltage divider effect that lowers the gain.

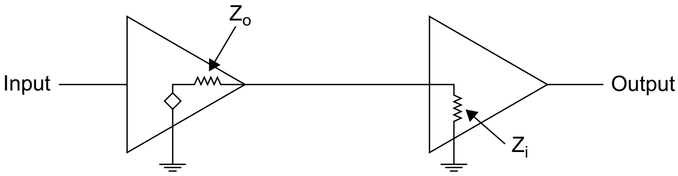

Z_o and Z_i form a voltage divider that lowers or cancels gain

FIGURE 4.11 When one amplifier feeds another, input and output impedances form a voltage divider that attenuates the signal and offsets some of the gain.

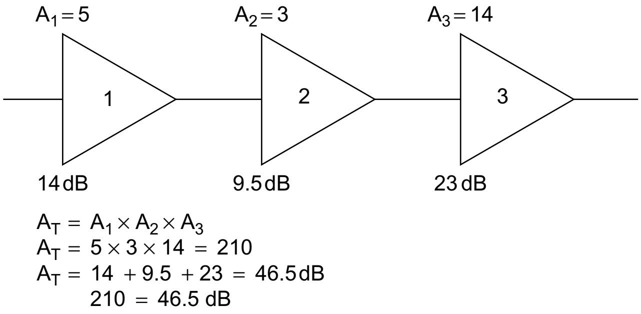

FIGURE 4.12 Cascading amplifier provides more gain. Total gain is product of individual gains or sum of dB gains.

If the amplifier gains are given in dB, you just add the dB figures to get the total gain.

Efficiency

Efficiency is the ratio of the output power to the input power expressed as a percentage.

$$\%\ \text{Eff} = \frac{\text{Po}}{\text{Pi}} \times 100$$

An amplifier takes the DC supply voltage and converts it into the AC output. The efficiency expresses how well the amplifier does this. Any lost power usually gets converted to heat which is not good. Nevertheless, some heat will always be generated just because electronic circuits are not perfect. High efficiency is a good thing of course, but often efficiency is traded off for better linearity, fidelity, and lower distortion of the signal.

Class A amplifiers have the poorest efficiency commonly in the 10 to 30% range. Classes B and AB are more efficient in the range of 50 to 70%. Class C amps have an efficiency of 70 to 85%. Class D amplifiers are the best with efficiencies in the 80 to 95% range.

Frequency Response

This is a measure of the range of frequencies that amplifiers cover. DC amplifiers may only cover DC up to about a few hundred Hz. Audio amplifiers may cover voice frequencies from 300 to 3000 Hz or stereo amps that cover 20 Hz to 20 kHz. IF amplifiers may only cover a specific band, say, 455 kHz, 9 MHz, 45 MHz, 70 MHz, or some other popular IF value. Op amps amplify signals from DC up to some very high frequency as high as several hundred MHz in some cases.

RF amplifiers cover high radio frequencies from a few MHz up to microwaves from 1 to 100 GHz. Their frequency and range of coverage is usually set by a filter or LC tuned circuit.

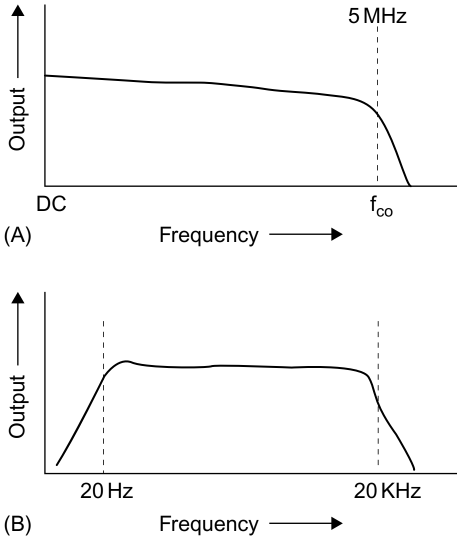

FIGURE 4.13 Frequency-response curves show output versus input frequency. Curve in **(A)** is response of op amp that extends from DC to 5-MHz cut-off frequency. Curve in **(B)** is response of common stereo audio amplifier.

Figure 4.13A shows the response curve for an op amp with an output down to DC and up to the cut-off frequency (f_{co}) or 5 MHz. Figure 4.13B is the response curve of a stereo amplifier with a range of 20 Hz to 20 kHz.

Output Capability

This specification indicates the range of output signals the amplifier is capable of. It is usually stated as a maximum voltage swing or as a maximum amount of output power. In both cases, the maximum output is usually given for a given value of load resistance. For example, an op amp may be capable of producing an output of ±12 volts across a 2 K load. A power amplifier may produce a maximum output of 10 watts across an 8 ohm speaker load or, say, 50 watts across a 50 ohm antenna load.

Filters

A filter is a frequency sensitive circuit that is designed to pass some frequencies, but reject others. There are various ways of making filter circuits that will accept some frequencies, but greatly attenuate others. There are literally thousands of applications for such circuits. But the main use of filters is to retain a desired signal, but eliminate interference and noise.

There are four basic types of filters: low-pass (LPF), high-pass (HPF), bandpass (BPF), and band-reject (BRF). The names, of course, tell their function.

A low-pass filter (LPF) is a circuit that passes all frequencies between DC and some upper cut-off frequency (f_{co}). All frequencies above the cut-off frequency are rejected. Figure 4.14A shows the response curve of an LPF. This is a plot of the filter output amplitude versus frequency. Note the gradual slope

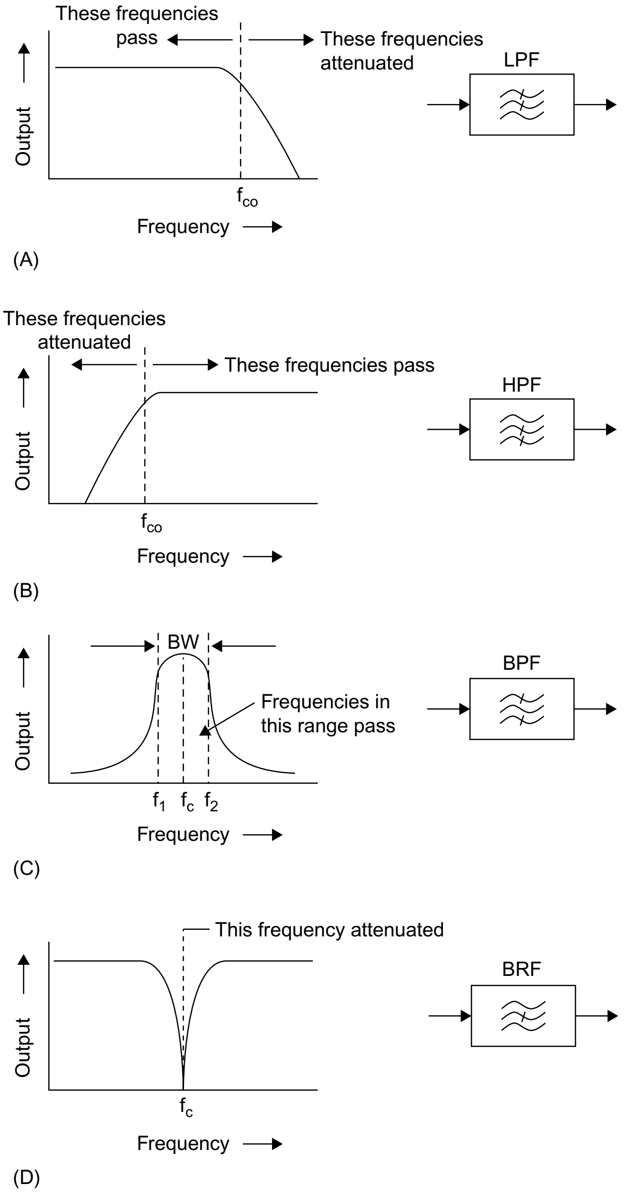
These frequencies pass
These frequencies attenuated
LPF
Output
f_co
Frequency
(A)
These frequencies attenuated
These frequencies pass
HPF
Output
f_co
Frequency
(B)
BW
BPF
Frequencies in this range pass
Output
f_1 f_c f_2
Frequency
(C)
This frequency attenuated
BRF
Output
f_c
Frequency
(D)

FIGURE 4.14 Four common filter types and their response curves and symbols. **(A)** Low-pass. **(B)** High-pass. **(C)** Band-pass. **(D)** Band-reject or notch.

as signals above the cut-off frequency (f_c) are reduced in amplitude. The block symbol is how LPF is represented in some diagrams.

A high-pass filter passes frequencies above the cut-off and greatly attenuates those below the cut-off. Figure 4.14B shows the ideal response curve of an HPF. Note the HPF symbol.

Band-pass filters (BPF) pass a certain range of frequencies, but reject all those above and below that range. Figure 4.14C shows the practical response curve of a BPF. Note that there are two cut-off frequencies, one above and one below the center frequency f_c. The difference between the upper and lower cut-offs is called the filter bandwidth (BW):

$$BW = f_2 - f_1$$

This is the range of frequencies passed. Again note the schematic symbol for a BPF.

Figure 4.14D illustrates the curve for a band-reject or notch filter. It is used to eliminate one single frequency and those close around it.

Note in the block symbols in Figure 4.14 that each has three waves representing high, middle, and low frequencies. The wave with the slash through it means that these frequencies are cut out or greatly attenuated.

Filters can be implemented in many ways. Resistor and capacitor (RC) filters are common as are inductor and capacitor (LC) filters. You can make filters with op amps called active filters. Other filters are made with mechanical vibrating elements, a ceramic or quartz crystal, or a special surface acoustic wave (SAW) filter. Many filters today are made with digital signal processing (DSP). These will be discussed in Chapter 6.

Oscillators

An oscillator is a circuit that generates a signal. It is usually a sine wave or a rectangular wave. The electronic symbols used to represent an oscillator are given in Figure 4.15.

The main specification of an oscillator is its output frequency. Most oscillators have a fixed frequency but there are oscillators whose frequency can be varied. A tunable frequency oscillator is called a variable frequency oscillator (VFO). You can often tune it by turning a knob that varies a capacitor or inductor to change the frequency. Some oscillators are tuned by an external voltage. These are called voltage-controlled oscillators (VCOs). By varying the value of a DC input voltage, the frequency is changed.

The frequency of an oscillator is set by some frequency determining element or circuit. That circuit may be made up of resistors and capacitors (RC), inductors and capacitors (LC), or a crystal. A crystal is a thin piece of quartz that can be made to vibrate at a very precise frequency. It maintains its frequency very closely unlike RC or LC networks that can drift with temperature or change from vibration or other conditions. Crystals come in a small metal package and are used to set the frequency in most oscillators today.

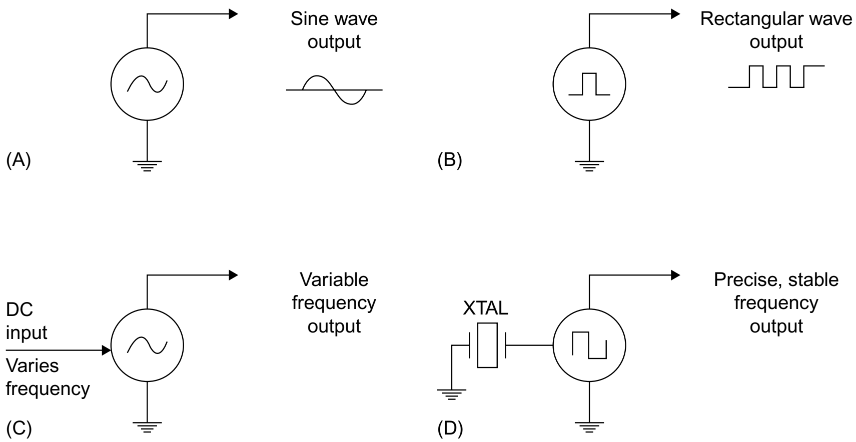

FIGURE 4.15 Types of oscillators and their symbols. **(A)** Sine wave. **(B)** Rectangular pulse. **(C)** VCO. **(D)** Crystal clock oscillator.

A special type of oscillator is called a clock oscillator. It is used as an accurate time and frequency source. The frequency determining element is usually a quartz crystal that vibrates at a precise frequency and is very stable. The clock signal is used to operate all the circuits in synchronism. An example is the clock oscillator in a PC that sets the speed of calculations. Sometimes you will see a quartz crystal (XTAL) symbol drawn with the oscillator symbol to illustrate this. See Figure 4.14D.

Mixers

A mixer is a circuit that takes two input signals and mixes them in a special way to produce new output signals. There are two types of mixers, linear and non-linear.

A linear mixer is one that takes the input signals and just adds them together. You can do this with some resistors as shown in Figure 4.16A. The sum of both input signals appears across the output resistor.

You can also make a linear mixer with an op amp as shown in Figure 4.16B. This op amp circuit is called a summer. Setting the resistor values allows you to mix and add gain at the same time. A good example of a linear mixer is those used by musicians so they can mix the outputs from microphones and musical instruments so they can all be amplified as one by the power amplifiers that drive the speakers. Note the variable resistors called pots that are used as amplitude or volume controls.

A non-linear mixer takes two input signals and effectively multiplies them together. The result is that the mixer produces four output signals. Figure 4.17 shows the symbol for a mixer. The inputs are frequencies 1 and 2 or f_1 and f_2.

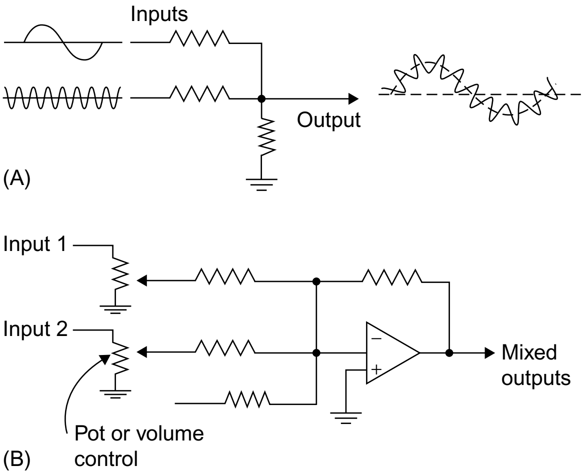
Inputs
Output
(A)
Input 1
Input 2
Mixed outputs
Pot or volume control
(B)

FIGURE 4.16 Linear mixing with resistors **(A)** or an op amp summer **(B)**.

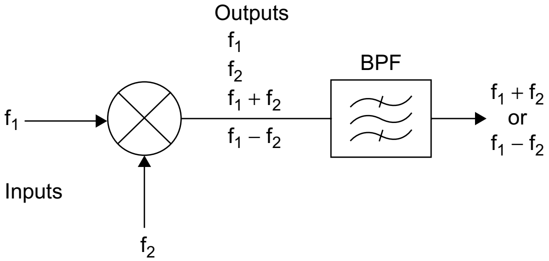
Outputs
f1
f2
f1 + f2
f1 − f2
BPF
f1 + f2
or
f1 − f2
f1
Inputs
f2

FIGURE 4.17 Non-linear mixer used to generate higher and low frequencies.

The mixer outputs are f_1, f_2, and ($f_1 + f_2$) or the sum, and ($f_1 - f_2$) the difference. The sum and difference frequencies are the ones of most interest. Usually a filter is used at the output of the mixer to select the desired frequency and eliminate the others. If the lower difference frequency is selected, the mixer is often called a down converter. If the sum frequency is selected, the mixer is called an up converter. The mixer is a way to translate a frequency to another higher or lower frequency that is better for processing. Any signal amplitude or frequency variations on the inputs are accurately carried through to the new output frequency.

Phase Detector

A phase detector is a mixer-like circuit that puts out a signal that is proportional to the phase difference between two input signals of the same frequency. See Figure 4.18. A phase shift is a time difference between two signals of the

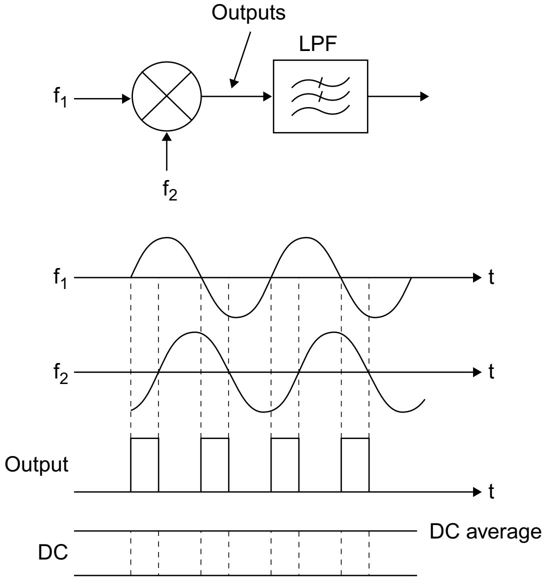

FIGURE 4.18 Phase detector converts phase shift into proportional DC average.

same frequency. We sometimes need to know how much phase shift is present. The phase detector produces a series of output pulses whose width is proportional to the phase difference. Passing the pulses through a low-pass filter smoothes them into a proportional DC voltage.

Filtering Pulses into DC

A common occurrence in electronic circuits is the filtering of pulses into a DC voltage. You saw this back in Figure 3.15 where positive pulses from a sine wave rectifier are smoothed into a DC voltage. It is the capacitor that charges up and stores the pulse voltage. Then between pulses the capacitor discharges into the load maintaining a near constant DC voltage. The low-pass filter usually includes a capacitor so performs the same function in the phase detector of Figure 4.18. The output is a proportional average DC output.

Phase-Locked Loops

The phase-locked loop (PLL) is a frequency- and phase-sensitive feedback control circuit. It consists of three major parts: a mixer or phase detector, a low-pass filter (LPF), and a voltage-controlled oscillator (VCO), as shown in Figure 4.19. The signal from the VCO is compared to the input signal. If there is a frequency (or phase) difference between the two, an error signal is generated. This error signal is filtered by the LPF into a varying DC level and is used to control the VCO frequency. This is the feedback signal.

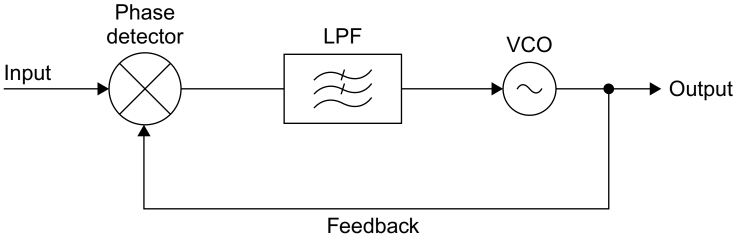

FIGURE 4.19 Phase-locked loop.

If the VCO frequency, f_o, differs from the input reference signal frequency, f_o the f_r the phase detector sees this as a phase shift. This causes the phase detector to produce an error signal. This error signal is filtered into a varying DC signal that is used to control the VCO.

The VCO can be either a sine wave oscillator or a rectangular wave oscillator depending upon the desired waveshape of the output. In either case, the VCO output frequency is made directly proportional to the DC control voltage. There is a linear voltage-to-frequency relationship in the VCO. This means that changing the control voltage produces a proportional change in output frequency. If the voltage goes up, so does the frequency.

Now if an input reference signal whose frequency is near that of the VCO is applied to the PLL, the phase detector will produce an output voltage proportional to the frequency difference. This signal is filtered and the resulting DC control voltage is applied to the VCO. The control voltage is such that it forces the VCO frequency to move in a direction that reduces the error signal. This means that the VCO frequency will change until it is equal to the input reference signal frequency. When this happens, the two signals are synchronized or “locked.” The phase difference causes the phase detector to produce the DC voltage at the VCO input to keep the PLL locked to the input signal.

If the input reference signal changes, then the phase detector will recognize a frequency (or phase) difference between the input and the VCO output. As a result, the LPF will produce a different DC control signal that will force the VCO to change such that it is equal to the new input frequency. As you can see then, the PLL will “track” an input signal frequency as it changes.

The range of frequencies over which the PLL will track an input signal and remain locked is known as the lock range. This is a range of frequencies above and below the VCO free running frequency. The PLL can track and “lock” to any input frequency in this range. If an input signal out of the lock range is applied, the PLL will not synchronize.

If the input signal is initially outside of the lock range, the PLL will not lock. But, the PLL will jump into a locked condition as soon as the input frequency gets close to the VCO frequency. In other words, the PLL will “capture” the incoming signal if it is close enough to the VCO frequency. Once the input signal is captured, the PLL is locked and will track further changes in the input signal frequency.

The range of frequencies over which a PLL can capture a signal is known as the capture range. Like the lock range, it too is centered on the free running frequency. But the capture range is narrower than the lock range. The PLL acts as a frequency sensitive circuit over a narrow range of frequencies.

Since the PLL will only capture and lock on to input signals within a certain narrow band, the PLL acts like a band-pass filter. For that reason, the PLL is an excellent signal conditioner. You can take a noisy input signal or one with undesirable interference on it and filter it with a PLL. The PLL will lock on to only the desired frequency component of the input. The VCO reproduces the input signal at the same frequency but with the noise and interference removed. The PLL not only cleans up a signal but also can track it if its frequency changes.

The PLL is widely used for a variety of purposes. It is used to recover the clock signal in some wireless applications. It is used to recover the original signal in frequency modulation (FM) radio. It is used to multiply a frequency by a fixed factor. It can be used for motor speed control. Almost all electronic products of some kind contain a PLL.

Frequency Synthesizers

One of the most popular uses of the PLL is in frequency synthesizers. A frequency synthesizer is a very stable signal source that can be varied over a specific range of frequencies in finite increments or steps. Frequency synthesizers are used as precision signal generators for test and measurement purposes and replace more conventional oscillator circuits as the primary signal sources in radio transmitters or receivers.

The big advantage of a PLL synthesizer is its ability to generate a wide range of frequencies with great accuracy and stability by using only a single stable signal source. For example, a PLL synthesizer with a single crystal controlled reference input can readily generate hundreds or even thousands of discrete frequencies that are as stable as the crystal reference. By using a synthesizer as the initial signal source in a transmitter, a wide range of channels can be obtained with a single crystal. When many channels must be used, a synthesizer greatly reduces the cost since crystals are expensive.

The synthesizer circuit in Figure 4.20 shows that it is a PLL with a frequency divider connected between the VCO output and the phase detector. The frequency divider is a circuit that divides its input frequency by some whole number. For example, the output of a divider with a division factor of 12 and an input of 48 MHz is 48/12 = 4 MHz.

The frequency synthesizer PLL also uses a stable, crystal oscillator input to the phase detector. This is usually derived from a crystal oscillator. When the PLL is locked, the stability and accuracy of the VCO output will be the same

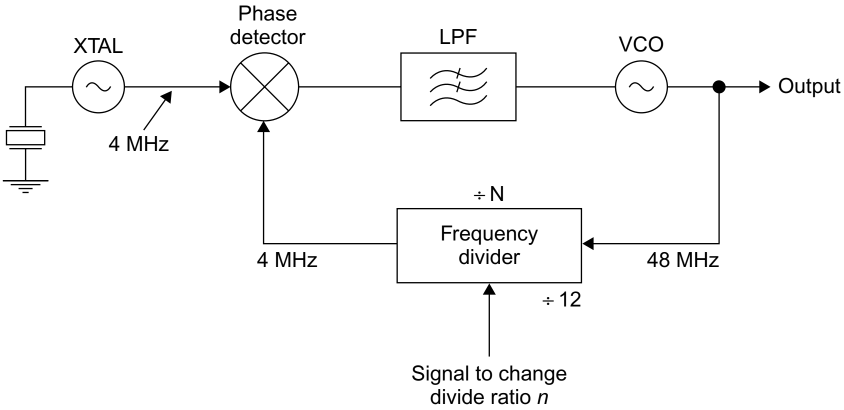

FIGURE 4.20 Frequency synthesizer using phase-locked loop. Changing the divider ratio changes the frequency.

as that of the reference input. The frequency divider is usually made so that its divide ratio can be changed by setting a switch or changing the control inputs. Changing the division ratio changes the output frequency at which the loop will lock.

The reason for this is that the two inputs to the phase detector must always be equal in frequency if a locked condition is to be achieved. The reference frequency is fixed so the only thing that can change is the VCO output if the divider ratio is altered. If the loop is locked, the output of the divider will be 24 MHz no matter what the divide ratio is. Therefore, with an output of 4 MHz, a divide by 6 circuit must have a 24 MHz input. Since the VCO output is the divider input, the VCO and synthesizer output will be 24 MHz too.

As you can see, changing the divide ratio changes the output frequency. In this circuit, if the divide ratio is changed in integer values between 6 and 12, the output frequencies will be as shown in Table 4.1. With this setup, the output is switched in 4-MHz increments.

If the input reference is made 1 MHz, then when the divide ratio is changed, the output frequency will step in 1-MHz increments. By carefully selecting the reference frequency and the divide ratio, it is possible to synthesize practically any range of frequencies in many increments from subaudio to microwave.

Note one final fact about the PLL circuit in Figure 4.20. If the reference is considered the input, and the VCO is considered the output, then the PLL is acting as a frequency multiplier. Here the output is 12 times the input. You will occasionally see the PLL used as a frequency multiplier in some applications.

The PLL synthesizer is a very versatile circuit. It is commonly available in integrated circuit form. PLL frequency synthesizers are used in TV channel selectors, stereo receivers, CB radios, cellular telephones, and other two-way radios. In such equipment the divider ratio is supplied by an input keyboard or

TABLE 4.1 PLL Divide Ratio and VCO Output

Divide Ratio	VCO Output
6	24 MHz
7	28 MHz
8	32 MHz
9	36 MHz
10	40 MHz
11	44 MHz
12	48 MHz

a built-in microcomputer programmed to supply the desired divide factor for a selected output frequency.

Power Supplies

A power supply is, as its name implies, a source of power to an electronic circuit. Most electronic circuits operate from DC as they process AC or DC signals. The most common power supply is the one that gets its input from the standard AC wall outlet that supplies 120 volts at 60 Hz. This common voltage is then converted by the power supply into one or more DC voltages that go to operate the TV set, computer, or other equipment. This section discusses the main components and circuits of a power supply.

Batteries

One of the most widely used power supplies is a battery. A battery is a great source of DC by itself and no AC source is needed. Batteries were the very first form of voltage sources for electrical circuits and they are still widely used today. What would we do without the batteries for our cell phones, iPods, cordless telephones, laptop computers, and remote controls? Just remember that a special form of power supply is the battery charger that normally uses AC to recharge lead-acid, nickel cadmium, nickel metal hydride, and lithium-ion batteries.

Standard Power Supply

A very common AC to DC power supply is shown in Figure 4.21. It normally consists of a transformer that translates the 120 volt AC line voltage into a higher or lower voltage as needed by its step up or step down characteristics. Today, since most equipment uses ICs, a lower AC voltage is needed. The lower

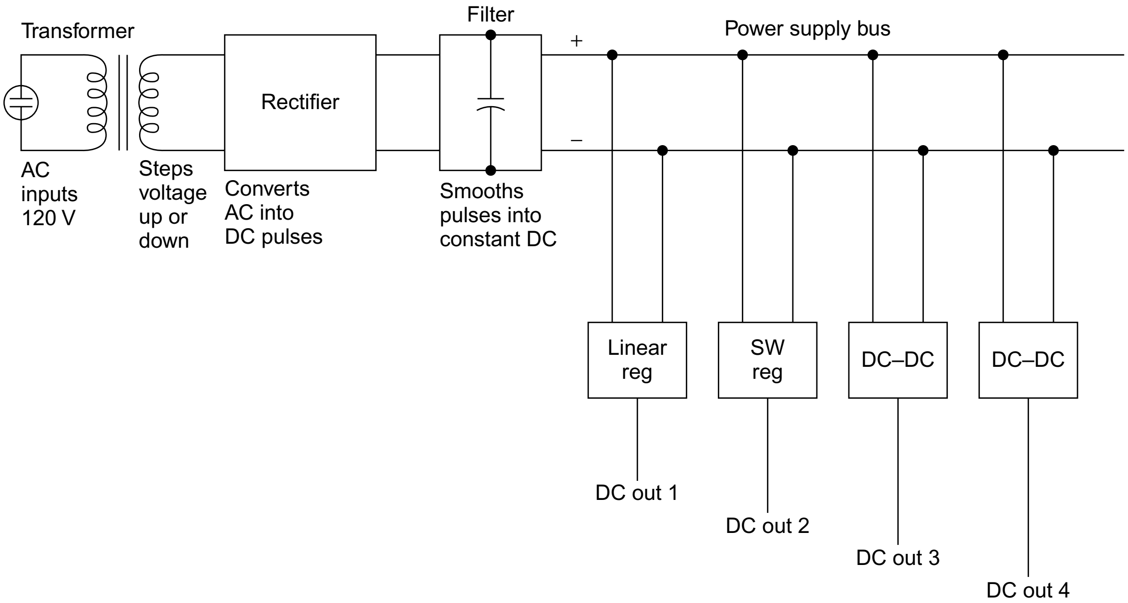
Transformer
AC inputs 120 V
Steps voltage up or down
Rectifier
Converts AC into DC pulses
Filter
Smooths pulses into constant DC
+
–
Power supply bus
Linear reg
SW reg
DC–DC
DC–DC
DC out 1
DC out 2
DC out 3
DC out 4

FIGURE 4.21 Typical power supply configuration using bus architecture, regulators, and DC–DC converters to get desired number of outputs. reg, regulator.

AC voltage is then converted into a pulsating DC voltage by a rectifier. A rectifier is one or more diodes that act like polarity sensitive switches to change the positive and/or negative sine half cycles into pulses of DC. A filter is then used to smooth the pulses into a more constant DC voltage. The filter is usually a large capacitor that charges up to the peak of the DC pulses then only discharges very slowly into the load.

Regulators

A regulator is a circuit that maintains a fixed DC output voltage of a desired value. The DC produced by the power supply at the filter output will vary over a wide range as the AC input line voltage changes. Any changes in the load meaning variations in current taken from the supply will also cause the output voltage to vary. A regulator absorbs those variations and creates a constant DC output suitable for powering ICs and other electronic circuits. Common DC output voltages are 3.3, 5, 12, 15, and 24 volts. Other typical DC supply voltages are 1.2, 1.8, 6, and 48 volts.

There are two types of regulators, linear and switching. A linear regulator inserts a transistor in series between the power supply output and the load. If the load varies, circuitry in the regulator adjusts the conduction of the transistor so that a constant voltage is maintained. It is as if a resistor in series with the load forms an adjustable voltage divider. If the output voltage drops, the series transistor conducts more to keep the output voltage the same. If the output increases, the transistor conducts less to decrease the voltage and maintain the same value.

The other type of regulator is the switching regulator. It puts a transistor in series with the load as well but the internal circuitry turns the transistor off and on at a high rate of speed. By varying the duty cycle (ratio of on to time for one period) of the on–off pulses the average DC output may be varied. Then if any output voltage changes are sensed, the duty cycle is adjusted to ensure that the average output stays the same.

Duty Cycle and PWM

The duty cycle is the ratio of the pulse width to the period of a rectangular wave as shown in Figure 4.22. The period is the reciprocal of the frequency of the signal or $T = 1/f$. It is constant. The on-time is the duration of the pulse voltage (t). The pulse duration can vary depending upon what the application is. If the period is 50 milliseconds and the pulse on-time is 10 milliseconds, the duty cycle (D.C.) is:

$$\text{D.C.} = 10/50 = 0.2 \times 100 = 20\%$$

Circuits can be made to change the on-time so that the duty cycle can be varied from a few percent to nearly 100%. This is called pulse width modulation (PWM). It was introduced earlier in the class D amplifier discussion. PWM is widely used in switching power supplies like regulators and DC–DC converters.

(*Continued*)

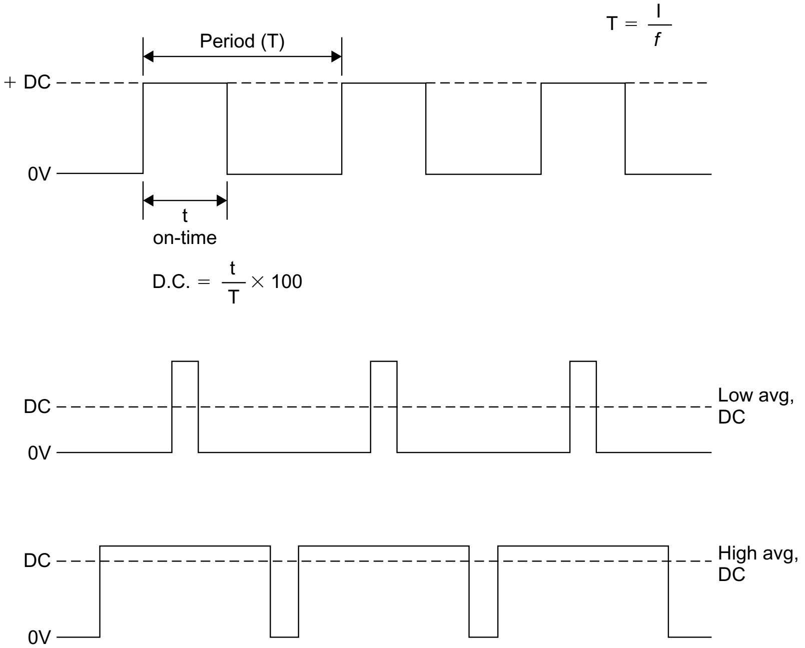

FIGURE 4.22 Concept of duty cycle and how varying pulse widths can be filtered into a proportional DC average voltage. avg, average.

Keep in mind that if you filter the pulses into an average proportional DC with a capacitor or LPF, you can vary the DC from some low value with a low duty cycle or some higher value with a higher duty cycle.

Linear regulators work very well but are not very efficient as they waste power in the series transistor. Typical efficiency is 10 to 40%. A switching regulator is far more efficient as it is off for a large part of the time. Efficiencies in the 70 to 95% range are common.

DC—DC Converters

A DC–DC converter is a circuit that does what its name implies. It translates one DC voltage to another. It can take the voltage from a 12 volt battery and generate several hundred or thousand volts. Or it can take a 4.2 volt battery and produce voltages of 3.3 or 2.5 volts and 15 volts. By building one DC power supply, several other output voltages can be obtained by using different DC–DC converters.

Bus-Oriented Architecture

Many power supplies have what we call a bus architecture. See Figure 4.21. The DC power supply produces one main output voltage. That voltage is applied to a common DC bus. A bus is just the name for the two + and − DC lines that carry the voltage. Then, regulators and DC–DC converters are used to generate the voltages needed by the equipment. Again see Figure 4.21.

Power Management

Many power supplies today include a circuit that manages the power supply voltages. It is usually a single IC that does things like turn on the various DC voltages in a power supply in a specific sequence. Some ICs require multiple DC voltages and they must be turned on one at a time in a specific sequence. In addition, the power management chip does things like monitor current usage. If one power supply is not being used, meaning current not being drawn from it, that part of the power supply is shut down to conserve power. A good example of this is the power management in a laptop computer or cell phone that monitors and controls all DC power to preserve the battery charge for longer life. Some other functions performed by power management chips are voltage monitoring with voltage alarm if a voltage fails, temperature monitoring, fault detection, over voltage or under voltage detection, and power control. Some power management chips also have the battery charging circuits built in. Every cell phone and laptop has one of these.

Inverters

An inverter is a power supply that converts DC into AC. A common unit converts 12 volts from a car battery into standard 120 volt 60Hz AC. With an inverter you can operate common household lights and appliances from a battery.

A special variation of an inverter is the uninterruptible power supply (UPS). These are widely used with computers and networks to provide power even during loss of AC input from the power line. A battery operates an inverter that is used to power the computers and other equipment. If the AC power fails, the equipment continues to operate. The battery is continuously charged by a built-in battery charger that operates from the AC power line.

Solar Power System

Any home or business solar power system is similar to a large UPS (see Figure 4.23). It uses the solar panels to charge a bank of batteries that store the energy for use later when the sun goes down. The inverters convert the DC into standard 120/240 volt AC to power the house. In some cases, unused power is sent back to the power grid.

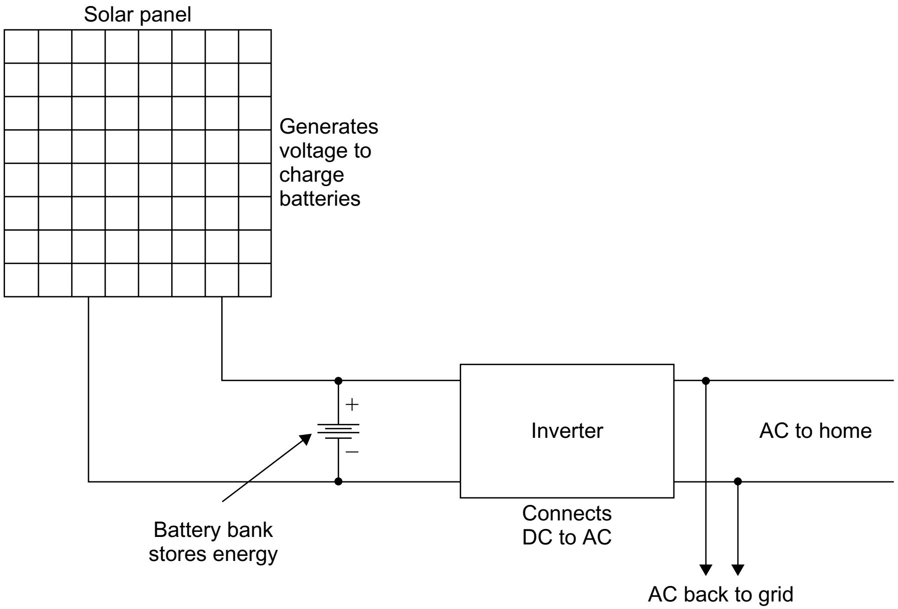
Solar panel
Generates voltage to charge batteries
+
–
Inverter
AC to home
Battery bank stores energy
Connects DC to AC
AC back to grid

FIGURE 4.23 Common DC-to-AC solar power system.

Project 4.1

Build a Kit

If you are really serious about electronics, then one of the first things you should do is build a kit. A kit, in case you really do not know, is a complete package of electronic parts to make a specific product. You build it yourself by placing components on a printed circuit board (PCB) and soldering them in place. There is no better way to see how electronic products get put together than to handle the wide variety of parts. It is a great learning experience.

There are two main questions to answer about this project. First, what kind of kit do you want? Second, can you learn to solder? You will have to answer the first question, but as for the second, yes, you can learn to solder. It's not that hard.

One way to get an overview of the kits available is to obtain catalogs from the sources mentioned earlier in this book. You can order a print copy online and see the whole range of kits available. Go to:

All Electronics: www.allelectronics.com
Elenco Electronics: www.elenco.com
Jameco: www.jameco.com
Kelvin: www.kelvin.com
Ramsey: www.ramseyelectronics.com

What I recommend is to get a kit related to one of the topics in this chapter. A power supply is a great choice, and it will give you a variable DC voltage source that you can use in future projects. The same is true for a basic audio amplifier. A solar panel kit is a good one also. Most of the companies have lots of radio- and wireless-related kits, but save these for the later chapters on wireless products.

Don't forget to get a small soldering iron (25 watts or so) with a small tip and some solder. You will also need a few hand tools such as a wire stripper, needle-nose pliers, side cutters, and a small screwdriver, if you do not already own these.

Chapter 5

Electronic Circuits: Digital

Practically Everything Is Digital These Days

In this Chapter:

- Binary number system.
- Logic gates and flip-flops.
- Common combinational and sequential circuits.
- Semiconductor memory.
- Programmable logic devices.
- Analog-to-digital and digital-to-analog conversion.

INTRODUCTION

Recall that there are two basic types of electronic signals—analog and digital. A digital signal is one that varies in discrete steps. Unlike an analog signal, which varies continuously, a digital signal has two levels or states. The signal switches or changes abruptly from one state to the other.

Figure 5.1 shows a DC digital signal that switches between two known levels such as zero volts or close to it (<0.1 volt) or 0 V and +3.3 V. The positive voltage can be anything between about 1 volt and 12 volts with 3.3 and 5 being the most common.

Digital signals with two discrete levels are also referred to as *binary signals. Binary* means two—two states or two discrete levels of voltage.

Humans use the decimal number system that represents quantities with digits 0 through 9. However, digital equipment and computers do not. Internally, digital equipment processes binary data.

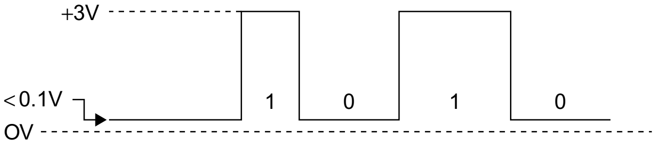

FIGURE 5.1 Binary signal that represents 0 and 1 as voltage levels.

© 2010 Elsevier Inc. All rights reserved.
doi: 10.1016/B978-1-85617-700-9.00005-9

BINARY NUMBERS

The binary number system is a set of rules and procedures for representing and processing numerical quantities in digital equipment. Because the base of the binary number system is 2, only two symbols (0 and 1) are used to represent any quantity. The symbols 0 and 1 are called *binary digits* or *bits*. For example, the 6-bit number 101101 represents the decimal number 45. Your understanding of how digital circuits, microprocessors, and related equipment process data is tied directly to an understanding of the binary number system.

The reason for using binary numbers in digital equipment is the ease with which they can be implemented. The electronic components and circuits used to represent and process binary data must be capable of assuming two discrete states to represent 0 and 1. Examples of two-state components are switches and transistors. When a switch is closed or on, it can represent a binary 1. When the switch is open or off, it can represent a binary 0. A conducting transistor may represent a 1, whereas a cut-off transistor may represent a 0. The representation may also be voltage levels. For example, a binary 1 may be represented by +3.3 V and a binary 0 by 0 V as previously shown in Figure 5.1.

Binary-to-Decimal Conversions

The binary system is similar to the decimal system in that the position of a digit in a number determines its weight. Recall that in the decimal system the weights are powers of 10. The right-most digit is units or 1's, and then 10's, 100's, 1000's, and so on, moving from right to left from one digit to the next. The position weights of a binary number are powers of 2:

$2^0 = 1$
$2^1 = 2$
$2^2 = 4$
$2^3 = 8$
$2^4 = 16$
$2^5 = 32$
$2^6 = 64$
$2^7 = 128$
$2^8 = 256$

The position weights of an 8-bit binary number follow:

Bit position	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
Weight	128 (MSB)	64	32	16	8	4	2	1 (LSB)

The most significant digit or bit (MSB) and the least significant bit (LSB) are identified.

Now let's evaluate the decimal quantity associated with a given binary number, 101101. You simply multiply each bit by its position weight and add the values to get the decimal equivalent, 45.

Position weight	32	16	8	4	2	1
Binary number	1	0	1	1	0	1

$1 \times 1 = 1$
$0 \times 2 = 0$
$1 \times 4 = 4$
$1 \times 8 = 8$
$0 \times 16 = 0$
$1 \times 32 = 32$
Decimal equivalent 45

You can see that positions with a 0 bit have no effect on the value. Therefore, they can be ignored. To quickly determine the decimal equivalent of a binary number, simply sum the weights of the positions containing a 1 bit. For example, in the number 11101, the weights of those positions with a 1 bit from right to left are 1 + 4 + 8 + 16 = 29.

Using Hardware to Represent Binary Quantities

Switches are widely used to enter binary data into computers and digital equipment. If the switch is set to the up position, a binary 1 is represented. If the switch is down, a binary 0 is represented. The switch contacts can be open or closed depending on the circuit in which the switch is connected. Figure 5.2 shows a group of slide switches set to represent a binary number, 11000101 or 197.

Indicator lights such as light-emitting diodes (LEDs) are sometimes used to read out binary data in digital equipment. An "on" light is a binary 1 and an "off" light is a binary 0. See Figure 5.3. The decimal value being represented by the display is 178.

Decimal-to-Binary Conversions

You may find it necessary to convert decimal numbers into their binary equivalents. This can be done by dividing the decimal number by 2, dividing the quotient by 2, dividing that quotient by 2, and so on, recording the remainders until the quotient is zero. The remainders form the binary equivalent.

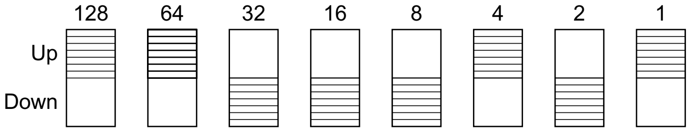

FIGURE 5.2 Off/on switches are used to enter binary data into digital circuits.

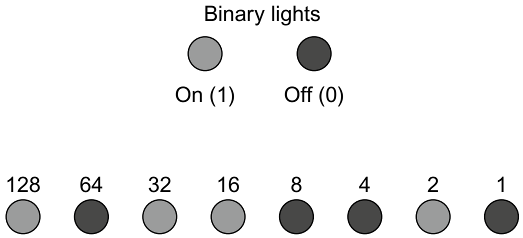

FIGURE 5.3 How lights or LEDs are used to represent binary data.

In the following example, we convert the decimal number 57 to its binary equivalent.

Quotient	Remainder
$57 \div 2 = 28$	1 (LSB)
$28 \div 2 = 14$	0
$14 \div 2 = 7$	0
$7 \div 2 = 3$	1
$3 \div 2 = 1$	1
$1 \div 2 = 0$	1 (MSB)

You can always check your work by reconverting the binary number back to decimal using the procedure described earlier. Also, keep in mind that some scientific calculators can do binary-to-decimal and decimal-to-binary conversions.

Another name for a binary number is *binary word*. The term *word* is more general. It refers to a fixed group of bits that can mean numbers, letters, or special characters and codes. We say that digital equipment processes *binary data words*.

All digital circuits and microcomputers work with a fixed-length binary word. A common binary word length in microcomputers and other digital equipment is 8 bits. All data storage, processing, manipulation, and transmission are carried out in 8-bit groups. Word lengths of 4, 8, 12, 16, 32, and 64 bits are common in digital equipment.

An 8-bit word is usually called a *byte*. In data communication applications, an 8-bit word is sometimes called an *octet*. A 4-bit word is sometimes referred to as a *nibble*.

Maximum Decimal Value for N Bits

The number of bits in a binary word determines the maximum decimal value that can be represented by that word. This maximum value is determined with the formula:

$$M = 2^N - 1$$

where M is the maximum decimal value and N is the number of bits in the word.

For example, what is the largest decimal number that can be represented by 4 bits?

$$M = 2^N - 1 = 16 - 1 = 15$$

With 4 bits, the maximum possible number is binary 1111, or decimal 15.

The maximum decimal number that can be represented with 1 byte is 255 or 11111111. An 8-bit word greatly restricts the range of numbers that can be accommodated. But this is usually overcome by using larger words.

There is one important point to know before you leave this subject. The formula $M = 2^N - 1$ determines the maximum decimal quantity (M) that can be represented with a binary word of N bits. This value is 1 less than the maximum number of values that can be represented. The maximum number of values that can be represented (Q) is determined with the formula:

$$Q = 2^N$$

With 8 bits, the maximum number of values is 256 or 0 through 255.

Table 5.1 gives the number of bits in a binary number and the maximum number of states that can be represented.

BCD and ASCII

The binary numbers we have been discussing are usually referred to as pure binary codes. But there are other types of binary codes. For example, the binary-coded decimal (BCD) system is popular. BCD is a hybrid code between the binary and decimal systems. It was developed in an attempt to simplify the conversion processes between the two systems and to improve human–machine communication.

TABLE 5.1 Number of Bits in Binary Number and Maximum Number of States

Number of Bits	Maximum States
8	256
12	4096 (4 K)
16	65,536 (64 K)
20	1,048,576 (1 M)
24	16,777,216 (16 M)
32	4,294,967,296 (4 G)

Note: K = kilobits = 1024; M = megabits = 1,048,576; G = gigabits = 1,0737,41,824.

To represent a decimal number in the BCD system, each digit is replaced by its 4-bit binary equivalent. Thus, the number 729 in BCD is:

7 2 9

0111 0010 1001

The BCD code is given in Table 5.2.

It is important to note that the 4-bit binary numbers 1010 through 1111 that represent the decimal values 10 through 15 are invalid in BCD.

To convert a BCD number to decimal, simply substitute the decimal equivalent of each 4-bit group. The BCD number 1001 0100 0110 in decimal is 946.

A special form of BCD code is used in computers and data communications systems. It is a 7- or 8-bit code that is used to represent not only numbers but also letters (both uppercase and lowercase), special symbols, and control functions. This code is called the American Standard Code for Information Interchange, or ASCII (pronounced "ask-key").

The following are examples of ASCII:

Number, letter, or symbol	ASCII
F	01000110
8	00111000
J	01101010
+	00101011

TABLE 5.2 Binary-Coded Decimal

Decimal Value	BCD Code
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001

The complete ASCII is given in Table 5.3. It provides the decimal and hexadecimal (hex) values of the codes. The code contains both uppercase and

TABLE 5.3 The ASCII Character Code

Char	Dec	Oct	Hex	Char	Dec	Oct	Hex	Char	Dec	Oct	Hex	Char	Dec	Oct	Hex
(nul)	0	0000	0×00	(sp)	32	0040	0×20	@	64	0100	0×40	`	96	0140	0×60
(soh)	1	0001	0×01	!	33	0041	0×21	A	65	0101	0×41	a	97	0141	0×61
(stx)	2	0002	0×02	“	34	0042	0×22	B	66	0102	0×42	b	98	0142	0×62
(etx)	3	0003	0×03	#	35	0043	0×23	C	67	0103	0×43	c	99	0143	0×63
(eot)	4	0004	0×04	$	36	0044	0×24	D	68	0104	0×44	d	100	0144	0×64
(enq)	5	0005	0×05	%	37	0045	0×25	E	69	0105	0×45	e	101	0145	0×65
(ack)	6	0006	0×06	&	38	0046	0×26	F	70	0106	0×46	f	102	0146	0×66
(bel)	7	0007	0×07	‘	39	0047	0×27	G	71	0107	0×47	g	103	0147	0×67
(bs)	8	0010	0×08	(	40	0050	0×28	H	72	0110	0×48	h	104	0150	0×68
(ht)	9	0011	0×09	)	41	0051	0×29	I	73	0111	0×49	i	105	0151	0×69
(nl)	10	0012	0×0a	*	42	0052	0×2a	J	74	0112	0×4a	j	106	0152	0×6a
(vt)	11	0013	0×0b	+	43	0053	0×2b	K	75	0113	0×4b	k	107	0153	0×6b
(np)	12	0014	0×0c	,	44	0054	0×2c	L	76	0114	0×4c	l	108	0154	0×6c
(cr)	13	0015	0×0d	−	45	0055	0×2d	M	77	0115	0×4d	m	109	0155	0×6d
(so)	14	0016	0×0e	.	46	0056	0×2e	N	78	0116	0×4e	n	110	0156	0×6e
(si)	15	0017	0×0f	/	47	0057	0×2f	O	79	0117	0×4f	o	111	0157	0×6f
(dle)	16	0020	0×10	0	48	0060	0×30	P	80	0120	0×50	p	112	0160	0×70
(dc1)	17	0021	0×11	1	49	0061	0×31	Q	81	0121	0×51	q	113	0161	0×71
(dc2)	18	0022	0×12	2	50	0062	0×32	R	82	0122	0×52	r	114	0162	0×72
(dc3)	19	0023	0×13	3	51	0063	0×33	S	83	0123	0×53	s	115	0163	0×73

(Continued)

TABLE 5.3 (Continued)

Char	Dec	Oct	Hex	Char	Dec	Oct	Hex	Char	Dec	Oct	Hex	Char	Dec	Oct	Hex
(dc4)	20	0024	0×14	4	52	0064	0×34	T	84	0124	0×54	t	116	0164	0×74
(nak)	21	0025	0×15	5	53	0065	0×35	U	85	0125	0×55	u	117	0165	0×75
(syn)	22	0026	0×16	6	54	0066	0×36	V	86	0126	0×56	v	118	0166	0×76
(etb)	23	0027	0×17	7	55	0067	0×37	W	87	0127	0×57	w	119	0167	0×77
(can)	24	0030	0×18	8	56	0070	0×38	X	88	0130	0×58	x	120	0170	0×78
(em)	25	0031	0×19	9	57	0071	0×39	Y	89	0131	0×59	y	121	0171	0×79
(sub)	26	0032	0×1a	:	58	0072	0×3a	Z	90	0132	0×5a	z	122	0172	0×7a
(esc)	27	0033	0×1b	;	59	0073	0×3b	[	91	0133	0×5b	{	123	0173	0×7b
(fs)	28	0034	0×1c	<	60	0074	0×3c	\	92	0134	0×5c	|	124	0174	0×7c
(gs)	29	0035	0×1d	=	61	0075	0×3d	]	93	0135	0×5d	}	125	0175	0×7d
(rs)	30	0036	0×1e	>	62	0076	0×3e	^	94	0136	0×5e	~	126	0176	0×7e
(us)	31	0037	0×1f	?	63	0077	0×3f	_	95	0137	0×5f	(del)	127	0177	0×7f

Char = ASCII character
Oct = Octal (3-bit groups)
Hex = Hexadecimal (4-bit groups)
Dec = Decimal
Characters in parentheses are control codes for specific operations.
X = "Don't care" character
Example: ASCII capital letter A
Decimal 65
Binary 0000000001000001
Octal 0/000/000/001/000/001
0 0 1 0 1
Hex 0000/0000/0100/0001
0 x 4 1

lowercase letters, the digits 0 through 9, punctuation marks, common mathematical symbols, and many special two- and three-letter codes. The special codes are known as control codes. They are used to indicate the start and end of a string of characters or to initiate special operations. Some examples follow:

STX—Start of text sequence
EOT—End of transmission of a string of characters
BEL—Ring a bell
DEL—Delete
LF—Line feed
CR—Carriage return

These codes not only represent alphanumeric data but can also make things happen in a printer or computer.

The main use of ASCII is in data communication. Computers use ASCII to "talk" to their peripheral units, such as printers, or to one another. A common example is a personal computer that receives data via a modem from the Internet. Another example is the data sent from a computer to a laser printer. ASCII is the communication code for text.

Hexadecimal Notation

Hexadecimal notation is a way to represent binary values with numbers and letters in an effort to make data easier to remember, and to reduce the errors normally associated with writing or reading long strings of 1's and 0's. The hex code is just an extension of the BCD code, with the decimal values from 11 to 15 represented by the letters A through F. See the following table.

Decimal Value	Hex Code
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001
A (10)	1010
B (11)	1011
C (12)	1100
D (13)	1101
E (14)	1110
F (15)	1111

(*Continued*)

Hexadecimal Notation (Continued)

As an example to convert the binary number 101000110101100 to hex, you divide the word into 4-bit groups starting on the right and then replacing each with the hex codes. In this case, the hex code is A35C.

Parallel and Serial Data

It is frequently necessary for digital devices to communicate. Someone using a computer may have to send data to a printer. Or two people using personal computers may have to communicate over a local area network (LAN). To accomplish either of these objectives, one must transmit binary data from one circuit to another. This is done in one of two ways: parallel transmission or serial transmission.

Parallel Data

When the parallel method is used, all bits of a binary word or number are transmitted simultaneously. Figure 5.4 shows an 8-bit word being transferred from logic circuit 1 to logic circuit 2.

A *bus* is a set of parallel data lines over which digital data is transferred. A data bus can have any number of data lines as required by the application. However, the main characteristic of a bus is parallel data transfer. In physical form, a bus is usually implemented on a connector, with lines of copper on a printed circuit board or with a ribbon cable with one wire per bit.

Because all bits are transmitted at the same time over a parallel bus, data movement is extremely fast. The main benefit of parallel data transmission is its high speed. With high-speed logic circuits, a binary word can be transferred from one point to another in as little as several nanoseconds. Parallel data transmission over buses in computers takes place at rates up to several hundred megahertz. It is not uncommon to see a rate of 100 million bits per second (Mbps)

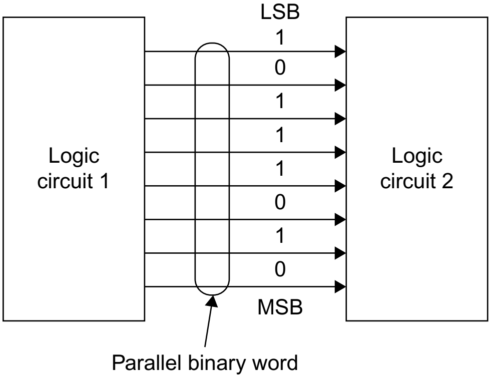

FIGURE 5.4 Parallel binary data transfer over bus.

on a personal computer data bus. On long buses, data rates are slower because the stray inductance and capacitance of the cable typically distorts the data and slows the transmission speed.

Although parallel transmission is fast, it is expensive, because there must be circuitry for each bit on both the sending and receiving ends. There must be one wire per bit in a cable plus a separate ground wire. This increases the complexity and thereby the cost of the circuitry.

Serial Data

The other method of moving binary data from one point to another is serial transmission. In the serial method, the bits of a word are transmitted one at a time, or sequentially. Figure 5.5 shows the waveform of an 8-bit serial binary word. This is the voltage that you would monitor at the single output of the circuit generating the word. It is the waveform that an oscilloscope would display. Each bit occurs for a fixed time interval such as 1.5 milliseconds. Thus, it takes $1.5 \times 8 = 12$ milliseconds (ms) to transmit an 8-bit word.

The speed of data transmission is measured in bits per second (bps). That figure can be calculated from the expression:

$$\text{bps} = \frac{1}{t}$$

Here, t is the time for 1 bit. In Figure 5.5, 1 bit takes 1.5 μs. The speed of the data, then, is

$$\text{bps} = 1 \div (1.5 \times 10^{-6}) = 666{,}666.7\,\text{bps or } 666.7\,\text{kbps}$$

If you know the speed in bps, you can calculate the bit time (t) by rearranging the formula:

$$T = \frac{1}{\text{bps}}$$

Assume a data rate of 14,400 bps. The bit time is 69.4 μs.

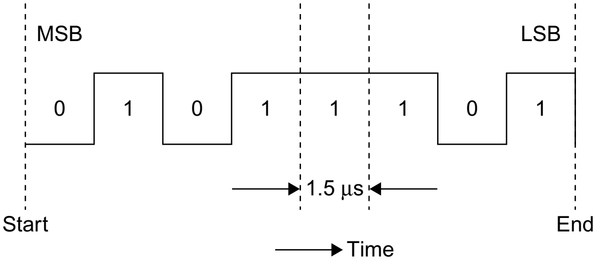

FIGURE 5.5 Serial data transfer where 1 bit occurs after another in time.

As seen in Figure 5.5, the MSB is transmitted first. Depending on the system used, however, the LSB can be sent first. In any case, the speed of transmission depends on the number of bits in the word and the duration of each bit. It is this longer transmission time that is the primary disadvantage of the serial method. However, although serial transmission is slower than parallel transmission, it is perfectly suitable for many applications. Common data rates are in excess of millions of bits per second. Data rates of more than 10 Gbps are common.

The main benefits of the serial method are its simplicity and low cost. Only a single line rather than multiple lines is needed to interconnect the equipment. In addition, only one set of sending and receiving circuits is needed. All data transmissions by copper cable or fiber optical cable exceeding about 100 Mbps are made by the serial method. All wireless data transmissions are serial.

DIGITAL LOGIC ELEMENTS

Digital logic elements are the basic circuits that are used to process the binary data. The logic element has one or more binary data inputs to be processed. The logic element processes or manipulates the binary input signals in a fixed way and generates an appropriate output signal. The output is a function of the binary states of the inputs and the unique processing capability of the logic element. The logic element "looks" at the binary input signals, and then makes a decision and generates an appropriate output. There are five elementary logic circuits: inverter or NOT, AND, OR, NAND, and NOR. These functions are performed by logic elements known as inverters and gates.

You need not be concerned with the inner workings of these logic elements. Instead, your primary concern is with the basic logic functions. Needless to say, the gates and inverters are inside every microcomputer and digital integrated circuit.

Inverter

The inverter has a single input and a single output. The logic function performed by an inverter is *inversion*. The output of the inverter is simply the inverse, or opposite, of the input. Because binary signals can assume only one of two different states, either 0 or 1, the inverter generates a 0 when the input is a 1, and a 1 when the input is a 0.

Two common symbols used to represent an inverter are shown in Figure 5.6. The input and output are given names, which are usually letters of the alphabet.

There are a variety of ways to express the operation of various logic functions. One way is to use Boolean algebraic expressions. A Boolean expression is a simple mathematical or algebraic formula that expresses the output in terms of the input. The Boolean expression for the inverter is:

$$B = A^*$$

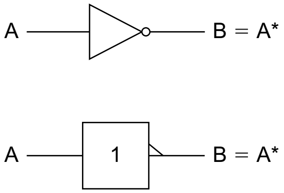

A	B
0	1
1	0

FIGURE 5.6 Logic symbols for inverter and truth table.

This equation is read, “B equals not A.” The asterisk means NOT or inversion. What this expression is telling you is that if the input is A, the output, B, is not A.

Another method of expressing the function of a logic element is a truth table. This is simply a table listing all possible combinations of the inputs and outputs of a logic element. Figure 5.6 also shows the truth table for an inverter. There are only two possible inputs, 0 and 1, and the corresponding outputs. A truth table completely defines the operation of a logic element.

AND Gate

A gate is a circuit that has two or more inputs and a single output. The gate generates a binary output that is a function of the states of the inputs and the nature of the gate itself.

The type of gate determines how the binary inputs are processed. The two basic logic gates are the AND gate and the OR gate. An AND gate is a logic circuit that generates a binary 1 output if all of its inputs are binary 1’s. Otherwise, the circuit generates a binary 0. All of the inputs have to be binary 1 for the AND gated to produce a binary 1 output. If any one of the inputs is a binary 0, the output is binary 0.

Figure 5.7 shows the logic symbols used to represent an AND gate. Letters of the alphabet are used to represent inputs and outputs. There may be more than two inputs, but there is only one output.

As with the inverter, the output of the AND gate can be expressed in terms of the inputs with a Boolean expression. The Boolean expression for an AND gate is:

$$C = AB$$

This equation is read, “C equals A and B.” Parentheses also can be used to separate the inputs; for example, $C = (A)(B)$.

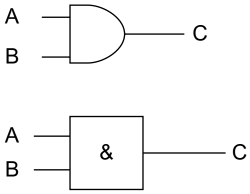

A	B	C
0	0	0
0	1	0
1	0	0
1	1	1

FIGURE 5.7 Logic symbols for AND gate and truth table.

The truth table for a simple two-input AND gate is also shown in Figure 5.7. Note that the output C is a binary 1 when both inputs are binary 1's. At all other times, the output is binary 0. With two inputs, there are a total of four different input combinations ($2^2 = 4$). With three inputs, there are $2^3 = 8$ possible input combinations for an AND gate. These inputs are the binary numbers 000 through 111.

One of the most common applications of an AND gate in digital circuits is gating. Gating simply refers to the use of one binary signal to control another. A two-input AND gate is most often used as a control gate. One input signal is the control that either keeps the other input signal from passing through to the output or allows it to pass.

Figure 5.8 shows the inputs and output of a typical control gate. Note that multiletter names called *mnemonics* are used instead of single letters to designate the input and output signals. The ENB or enable input represents the control signal. As long as the ENB input is a binary 0, the output is a binary 0. The gate is said to be inhibited because nothing more than a binary 0 output occurs.

When the ENB control input is a binary 1, the gate is enabled. At this time, the main input signal, CLK for clock, is allowed to pass through to the output BST. CLK is a periodic clock signal that switches repetitively between the binary 0 and the binary 1 levels at a fixed frequency. The output of the AND gate follows the CLK input as long as the ENB control input is a binary 1.

OR Gate

Another commonly used logic gate is the OR circuit. Like the AND gate, the OR gate has two or more inputs and a single output. The OR gate generates a binary 1 if any one or more of its inputs are binary 1. The only time the output of an OR gate is binary 0 is when all its inputs are binary 0.

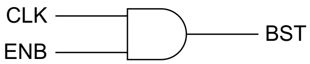

(A) Gate symbol

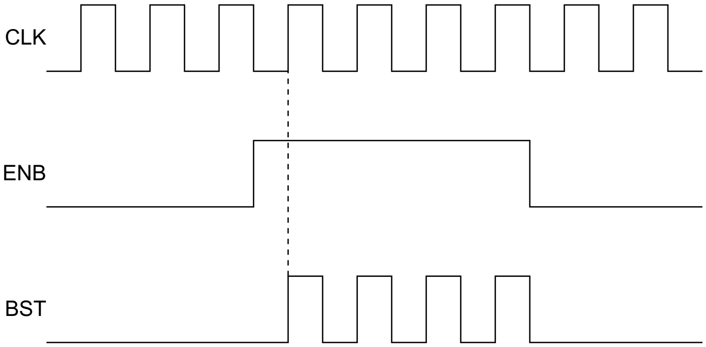

(B) Waveforms

FIGURE 5.8 How AND gate is used to turn signal off or on. **(A)** The AND gate showing the input and output names. **(B)** The waveforms of the inputs and output.

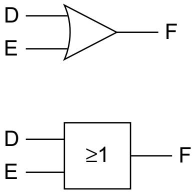

D	E	F
0	0	0
0	1	1
1	0	1
1	1	1

FIGURE 5.9 Logic symbols for OR gate and truth table.

The symbols used to represent an OR gate are shown in Figure 5.9. Letters of the alphabet are used to designate the inputs and outputs. Using these designations, the Boolean output expression for the OR gate is:

$$D = E + F$$

This equation is read, "D equals E or F." In this Boolean expression, the plus sign designates the OR function. In a Boolean logic expression, the plus sign does not mean addition. In Figure 5.9, the mathematical symbol ≥ (greater than or equal to) is used to designate the OR logic function. The truth table for an OR gate further defines its operation. The output D is a binary 1 when input E or input F or both are a binary 1. Like an AND gate, an OR gate can have any number of inputs.

NAND and NOR Gates

The AND and OR logical functions are basic to all digital systems. Although these functions are implemented with AND and OR gates, a variation of these gates is even more widely used. These are NAND and NOR gates, which are a combination of either an AND gate or an OR gate and an inverter. These gates are more flexible in their application when they are used together.

A NAND gate can be made with an AND gate followed by an inverter. Figure 5.10 shows this circuit configuration. Note the two special symbols in Figure 5.10 used to represent the logical NAND function. The truth table is also given.

A NAND gate can perform all of the gating and detection functions mentioned earlier for an AND gate. The only difference is that the output is inverted.

The other widely used digital logic element is a NOR gate. A NOR gate is a combination of an OR gate followed by an inverter (see Figure 5.11). The symbols used to represent the NOR logic function are also shown. The output of a NOR gate is simply the complement of the output of an OR gate, as shown by the truth table. An XOR actually is a 1-bit binary adder.

Exclusive OR Gate

A variation of the OR gate is the exclusive OR, usually designated XOR. Its symbols are given in Figure 5.12. The truth table is also shown. Note that the

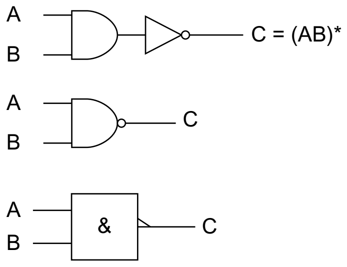

A	B	C
0	0	1
0	1	1
1	0	1
1	1	0

FIGURE 5.10 Logic symbols for NAND gate and truth table.

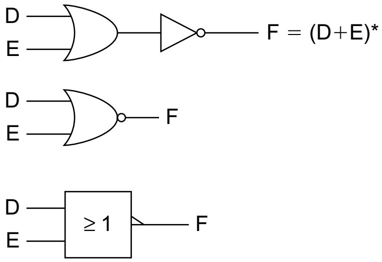

D	E	F
0	0	1
0	1	0
1	0	0
1	1	0

FIGURE 5.11 Logic symbols for NOR gate and truth table.

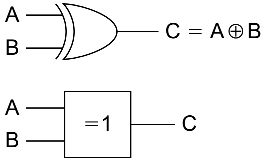

A	B	C
0	0	0
0	1	1
1	0	1
1	1	0

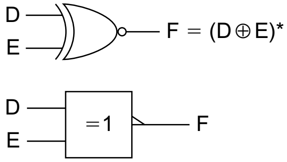

D	E	F
0	0	1
0	1	0
1	0	0
1	1	1

FIGURE 5.12 Logic symbols for XOR and XNOR gates and truth tables.

circuit only produces a 1 output if the two inputs are opposite. The output is 0 if the two inputs are the same.

An inverted XOR or XNOR has the symbols and truth table shown in Figure 5.12. It is useful as a 1-bit comparator since the output is 1 if both inputs are the same.

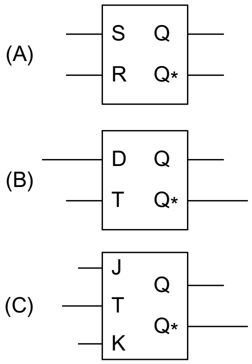

FIGURE 5.13 Logic symbols for flip-flops. **(A)** RS. **(B)** D type. **(C)** JK.

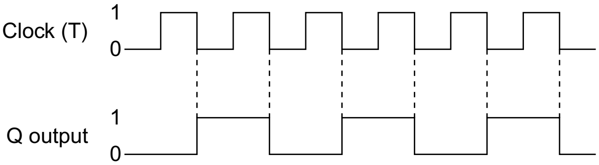

FIGURE 5.14 Toggling T input on JK flip-flop causes it to change states, producing frequency division by 2.

Flip-Flops

A flip-flop (FF) is another basic building block of electronics. It is essentially a circuit that stores 1 bit of data. The circuit has two states—set and reset. If the FF is storing a binary 1, it is set. If it is storing a binary 0, it is reset. Figure 5.13A shows the basic block diagram of a reset-set (R-S) FF. The inputs are set (S) and reset (R). The two outputs Q and Q* are complementary. You apply an input to the appropriate input to set or reset the FF.

A D-type flip-flop is shown in Figure 5.13B. It has a D or data input and a clock (T) input. When the clock occurs, the FF is either set or reset depending on the value of the D input. The D FF is used to store 1 bit of data.

Another type of FF is the JK, shown in Figure 5.13C. It has J and K inputs that correspond to S and R inputs, respectively. It also has a clock or T input that performs the set or reset operations. To store a binary 1, you make J = 1 and K = 0. When a clock input occurs, the FF is then set or reset.

The T input also causes the state of the FF to toggle or change state when a clock transition occurs (see Figure 5.14). This feature is used to produce frequency division by 2 as shown. The output is a square-wave half the frequency of the T input. Additional JK FFs can be cascaded to produce frequency division by the power of 2 or 2, 4, 8, 16, 32, 64, and so on, as shown in Figure 5.15.

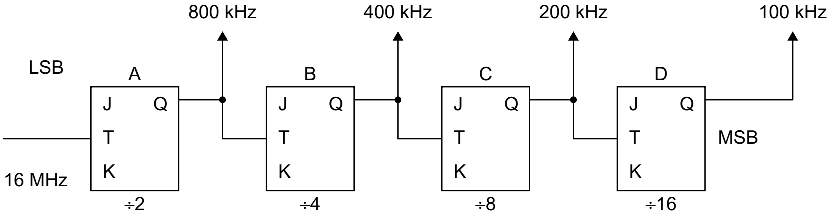
800 kHz
400 kHz
200 kHz
100 kHz
LSB
A
B
C
D
J
Q
T
K
16 MHz
MSB
÷2
÷4
÷8
÷16

FIGURE 5.15 Cascading JK flip-flops produce higher-frequency division ratios, which is also a binary counter that counts from 0000 to 1111.

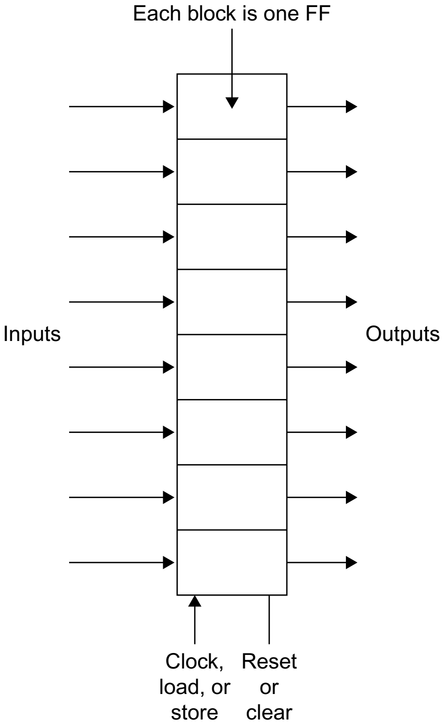
Each block is one FF
Inputs
Outputs
Clock, load, or store
Reset or clear

FIGURE 5.16 Storage register from one word consists of FFs. It also has clock/load and clear/reset inputs.

Storage Registers

A storage register is a place to store a binary word or number. It is made up of 1 FF per bit. An 8-bit register stores bytes and has 8 FFs. Figure 5.16 shows a register with eight inputs and eight outputs. Most registers also have a clock input also called LOAD or STORE that determines when the input data is stored. Some also have a CLEAR or RESET input that puts all the FFs in the binary 0 state.

Shift Registers

A shift register (SR) is like a storage register, but the bits stored there may be shifted from one FF to the next as each clock pulse occurs. Figure 5.17A shows how serial binary data is shifted into the SR for storage. As new data

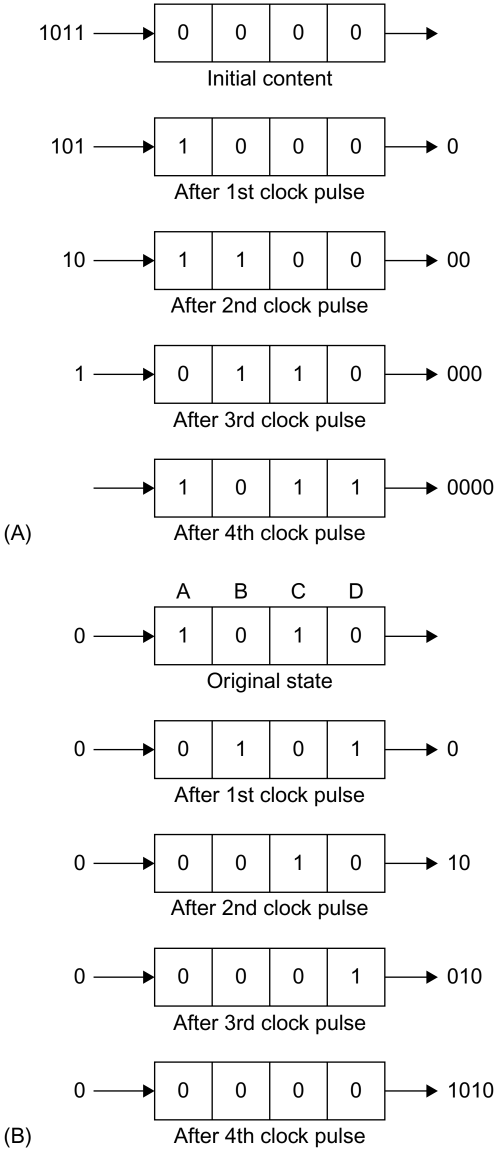
1011
0 0 0 0
Initial content
101
1 0 0 0
0
After 1st clock pulse
10
1 1 0 0
00
After 2nd clock pulse
1
0 1 1 0
000
After 3rd clock pulse
1 0 1 1
0000
(A)
After 4th clock pulse
A B C D
0
1 0 1 0
Original state
0
0 1 0 1
0
After 1st clock pulse
0
0 0 1 0
10
After 2nd clock pulse
0
0 0 0 1
010
After 3rd clock pulse
0
0 0 0 0
1010
(B)
After 4th clock pulse

FIGURE 5.17 **(A)** Shifting serial data into shift register. **(B)** Shifting serial data out of shift register.

is shifted in, the existing data will be shifted out and sent to another circuit. Figure 5.17B shows data being shifted out as binary 0's are shifted in.

Shift registers are often used for parallel-to-serial and serial-to-parallel conversions. A serial-to-parallel conversion is shown in Figure 5.18A. The SR initially contains all binary 0's. Then serial data is shifted in and the output is taken in parallel from the FF outputs. For parallel-to-serial conversion, a binary word is initially stored in the register. Then the word is shifted out as the clock pulses occur (see Figure 5.18B).

Counters

As its name implies, a counter is a circuit that counts input pulses. Counters are made of multiple cascaded FFs, and they are sometimes accompanied by logic gates. As the input pulses occur, the binary number stored in the counter FFs tells how many pulses have occurred. The frequency divider shown in Figure 5.15 is also a 4-bit counter. It can count from 0 (0000) to 15 (1111).

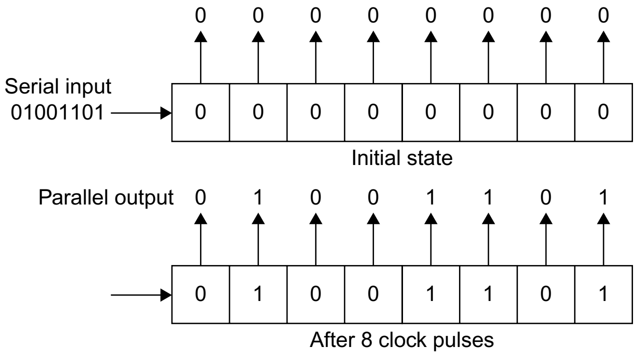
0 0 0 0 0 0 0 0
Serial input
01001101
0 0 0 0 0 0 0 0
Initial state
Parallel output 0 1 0 0 1 1 0 1
0 1 0 0 1 1 0 1
After 8 clock pulses

(A) Serial-to-parallel

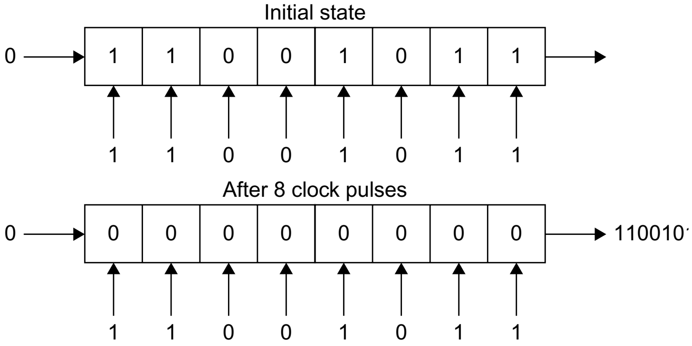
Initial state
0
1 1 0 0 1 0 1 1
1 1 0 0 1 0 1 1
After 8 clock pulses
0
0 0 0 0 0 0 0 0
110010
1 1 0 0 1 0 1 1

(B) Parallel-to-serial

FIGURE 5.18 Shift registers used for **(A)** serial-to-parallel data conversion and **(B)** parallel-to-serial data conversion.

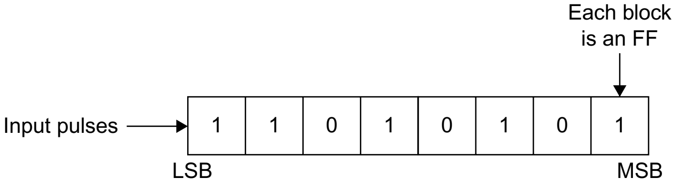

FIGURE 5.19 Counter storing binary value equal to number of input pulses that occurred.

If all FFs are initially reset to 0, then 10 pulses occur, and the FFs from A to D would be storing 0101 from left to right. The right-most FF (D) is the MSB, so you read the counter as 1010 or the binary equivalent of decimal 10.

Binary counters can store numbers up to $2^N - 1$, where N is the number of FFs. For example, an 8-bit counter would count from 00000000 to 11111111 or $2^8 - 1 = 256 - 1 = 255$.

Figure 5.19 shows a counter where 178 pulses have occurred, so the counter stores the number 10101011 that is the binary equivalent of 179.

Another popular form of counter is the BCD version that counts from 0000 to 1001 or 0 through 9. BCD counters are used to keep track of decimal quantities in binary form. BCD counters can serve as frequency dividers as well. Each divides by 10.

COMMON LOGIC CIRCUITS

There are several types of logic circuits that occur again and again in digital equipment. The counters and registers just described are good examples. A few others that you may encounter are described in this section. Remember that these circuits can be made from individual gates and FFs in IC form. You can also buy them in IC form already wired for their specific function. But mostly these circuits appear as a collection of circuits inside some larger IC.

Multiplexer

A multiplexer or mux is a circuit with two or more inputs and one output (see Figure 5.20). Here there are four inputs and one output. Two other inputs are used to select one of the inputs to appear as the output. The four selected inputs are 00, 01, 10, and 11. In the figure, note which input is routed to the output with each input select code. Multiplexers can have as many inputs as needed.

Demultiplexer

A demultiplexer or demux is the opposite of a mux. It has one input and multiple outputs. Like the mux, select inputs specify which of the output lines the input is connected to.

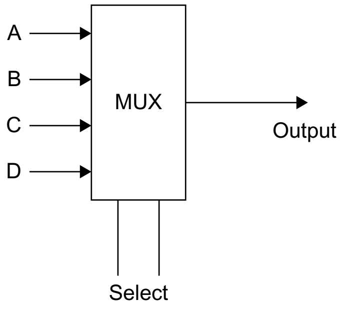

Select	Input at output
00	A
01	B
10	C
11	D

FIGURE 5.20 Digital multiplexer (mux) that selects one of four inputs to appear at the single output.

Decoder

A decoder is a circuit that can detect a specific binary code. In its simplest form, it is just an AND gate, as shown in Figure 5.21A. If the input 0110 appears at the inputs, the AND gate will have all its inputs at binary 1, so the output will be binary 1 indicating that it is detecting the desired input. For any other 4-bit input, the output will be binary 0.

Decoders can also be made to detect all states of a given binary input. For example, a 4-bit decoder like that in Figure 5.21B looks at the 4-bit inputs and decodes all states from 0000 to 1111. If the input is 1000, the 8 output is binary 1 and all other outputs are 0.

Comparators

A comparator takes two binary words as inputs and compares them. If the two inputs are equal, a binary 1 output occurs. Otherwise, if the two inputs are different, the output is 0. Some comparators have outputs that can tell if one input is larger than or smaller than the other.

Arithmetic Circuits

Digital logic does a good job of basic arithmetic functions like add and subtract. The basic computing element is called an arithmetic logic unit (ALU). It does binary addition and subtraction. It can also perform basic logic operations such as AND, OR, or XOR on two binary input words. Circuits for multiplication and division are more complex, but are common in digital computers.

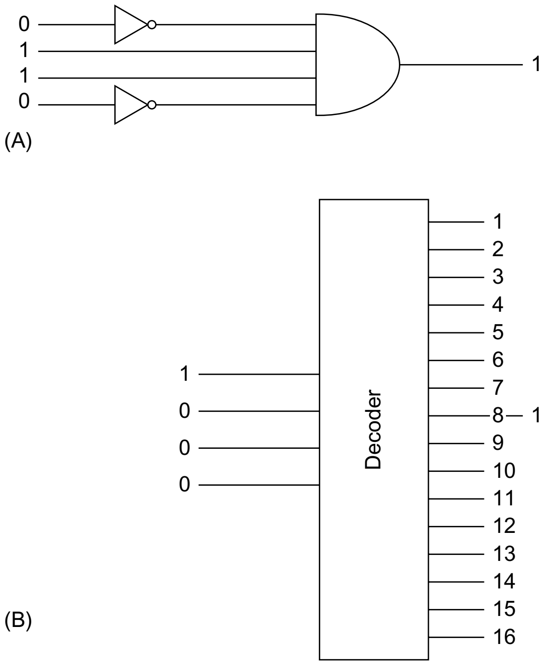

FIGURE 5.21 Decoders. **(A)** 1-AND gate decoder that identifies input 0110. **(B)** 4-to-16 decoder that identifies 1 of 16 different 4-bit input values.

DIGITAL MEMORIES

Digital memories are semiconductor circuits that store binary data. You have probably heard of the terms RAM and ROM as they apply to personal computers. These are the two main types of memory, random access memory (RAM), and read-only memory (ROM).

RAM

The random access memory is one in which you can access hundreds, thousands, or millions of binary storage locations. You can store one binary number or word in each location or read the data out of the location. Each location is referred to by a unique binary number called the address. You send the memory an address and activate the selected location. Then you can read data from the location or write data into that location. RAMs therefore are what we call read-write memories.

Figure 5.22 shows the concept of a RAM. It has eight locations for 4-bit words. To address eight locations you need a 3-bit address ($2^3 = 8$). When you apply the address to the address decoder, the decoder activates the desired location. In the figure you see that the binary number 1001 stored in location 011 is being accessed and read out.

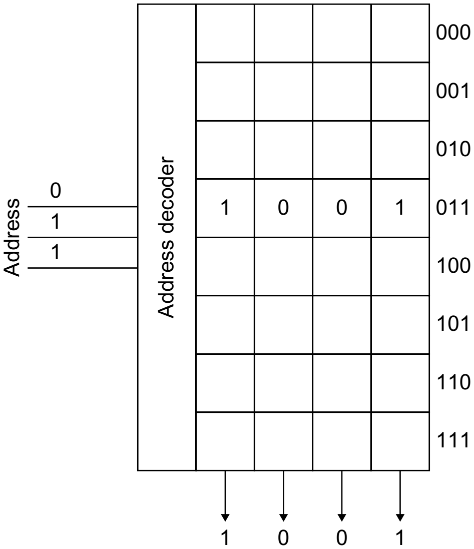

FIGURE 5.22 Basic concept for random access memory (RAM). Address selects one of eight storage locations. Then data may be read from that location or data stored in it.

Most memories are large today. We talk about megabytes (MB) or millions of bytes of data or gigabytes (GB) or billions of bytes of data in a modern PC. Embedded controllers will have fewer locations, usually in the thousands of bytes (KB) or perhaps more in the larger systems. Storage word sizes are usually 8 (bytes), 16, 32, or 64 bits.

There are two ways to make a RAM. The first way is just to build one FF for each bit to be stored. The basic set-reset FF described earlier is the basic storage circuit (see Figure 5.23A). Memories using this type of circuit are called static RAM or SRAM. They are called static because once you store the bit in the FF, it stays there as long as power is applied to the circuit.

The other form of RAM is called *dynamic RAM* or DRAM. The basic storage circuit is a capacitor as shown in Figure 5.23B. If you charge the capacitor, it is storing a binary 1. If the capacitor is discharged, a binary 0 is being stored. A MOSFET is used as a switch to access the capacitor. If the MOSFET is turned on by an input address on the gate, it connects the capacitor to the data line, where it may be read out or where a new bit may be stored. When the MOSFET is turned off, the capacitor is isolated and just stores the bit.

The reason why the term *dynamic* is used for this type of RAM is that the charge on the capacitor will leak off very quickly. The capacitor is very small and millions of them are made on a tiny chip of silicon. For this reason, the capacitor has to be recharged every so often so that the data is retained. This is called a refresh operation and it usually occurs every 2 milliseconds or so.

Comparing SRAM and DRAM, you should know that SRAM is the faster of the two in that you can store data or read data out faster than you can with a

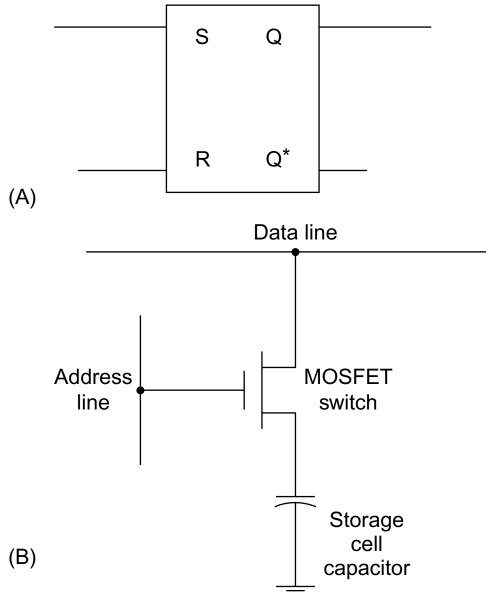

FIGURE 5.23 RAM storage cells. **(A)** RS FF for SRAM. **(B)** Capacitor cell for DRAM.

DRAM. The downside is that the SRAM is more complex and takes up more space on the chip, and thus costs more. The DRAM is slower but not much. Its advantage is that you can put more storage cells on a chip, making them cheaper. The need to refresh is a downside but easy to live with, as all the refresh circuitry is on the chip and the process is transparent to the user.

Just keep in mind a key fact. RAM is fast and reliable, but it loses all data if the DC power to the memory circuit is removed. All data is erased. RAMs are therefore said to be volatile.

ROM

Read-only memories are nonvolatile memories. Once you store data in them, they retain the data when power is removed. In some ROMs, you can change the data stored there. Incidentally, all ROMs are random access in that you can go to any desired location to read or maybe write data.

There are several categories of ROM. The older types are called masked ROM, where the mask refers to the actual wiring of the ROM to permanently fix the data in the circuitry when it is made. The storage cells are essentially a MOSFET matrix where each MOSFET is permanently turned on or off to store 1 or 0. No changes are possible. Later a programmable ROM or PROM was developed. It used a special MOSFET that could be programmed on or off. You could then erase the programming by shining ultraviolet (UV) light

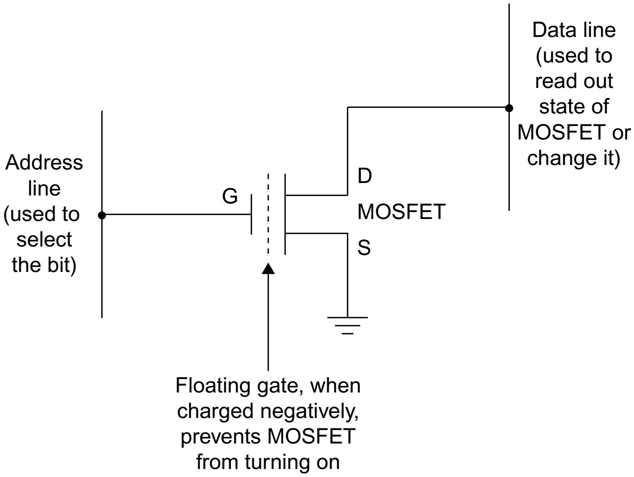
Data line (used to read out state of MOSFET or change it)
Address line (used to select the bit)
G
D
MOSFET
S
Floating gate, when charged negatively, prevents MOSFET from turning on

FIGURE 5.24 Storage cell used in EEPROM and flash memories is MOSFET with a special floating gate that keeps transistor off or on.

on the chip for a short time. Then the PROM could be reprogrammed. These are no longer used.

Most ROMs today are what we call electrically erasable programmable ROMs or EEPROMs or E^2PROMs. The storage element is a MOSFET that you turn off or on with a programmable voltage (see Figure 5.24). The MOSFET retains that state until it is electrically changed.

The main problem with EEPROMs is that they are slow to erase and reprogram. But today we have a special version of EEPROMs called flash memories. These are also EEPROMs, but they actually are almost as fast as RAMs for reading and writing data. Yet they are nonvolatile.

You will hear of two types of flash memory—NOR and NAND. NOR flash is used to store programs for microcontrollers. It can only be erased on large block sizes and not individual cells. NAND flash is faster and you can erase individual memory locations. The NAND flash is gradually replacing the NOR flash in many applications.

Flash memory has become so popular it is used in almost every electronic product today. MP3 players and iPods store music in flash memory. Cell phones store phone numbers, digital cameras store photos, and in embedded controllers, the flash usually stores the program that tells the processor what to do. And, let’s not forget the flash drive or “thumb” drive as it is called. These devices with a USB connector can store many gigabytes of data and are very inexpensive. Flash memory is also used to make solid-state drives (SSDs) that are the equivalent of the hard disk drives used in PCs and laptops. They are still expensive, but as prices decline, you can look for more SSDs to replace conventional mechanical/magnetic hard drives.

PROGRAMMABLE LOGIC DEVICES

There are two ways that digital devices are built today. The most common is to use an embedded controller or microcomputer. This is a single-chip computer that you can program to do whatever digital function you want. Most electronic products have at least one embedded controller at the heart of their design.

In some cases, the embedded controller is not fast enough to do the job. One way to make the circuitry is to use individual logic circuits such as gates, flip-flops (FFs), counters, and so on. That is the way it used to be done. Now, if you need fast custom digital logic, you build the circuit with a programmable logic device (PLD). A PLD is an IC that contains a mix of gates, FFs, and other circuits that can be programmed to implement any digital function. The programming is usually done on a computer and a file is created. The file is then downloaded to the PLD, where it is stored in a RAM or ROM. The file in the memory determines how the unique circuit is formed.

There are three basic types of PLDs: simple PLDs, complex PLDs, and field programmable gate arrays (FPGAs). The SPLDs are small devices for simpler digital circuits. The most common types are called programmable array logic (PAL), programmable logic arrays (PLAs), and generic array logic (GAL). PROMs are also used as SPLDs. For example, the address input to a ROM is some binary number or a collection of binary inputs. These form an address to locate a specific binary word stored in that location. When that particular selection of input bits occurs, it forms the address that identifies the memory location where the desired set of output responses is stored. This form of SPLD is also called a look-up table or LUT.

The other SPLDs are essentially a set of AND and OR gates that can be programmed. They all take the form shown in Figure 5.25. You can select which inputs go to the AND gates, and/or select which AND outputs go to which OR gate inputs. A PROM has the AND array fixed and you can program the OR gates. A PAL is an SPLD in which you can program the AND inputs but the OR connections are fixed. A PLA is a PLD that lets you program both the AND inputs and the OR connections. In both PLAs and PALs, the programming is usually permanent.

A GAL is a variation of the PAL with its AND inputs programmable. The difference is that the programming can be changed. The program you develop is actually stored in flash memory. The flash memory cells send the programming bits to the various AND and OR gates telling them what logic to implement. You can then erase it and reprogram if you want or need to. GALs also have one or more FFs associated with the logic gates, so you can store bits or actually build counters and registers by programming.

CPLDs are complex PLDs similar to SPLDs, only larger. They contain more gates and flip-flops so that very large logic projects can be built on a single chip. A typical CPLD is really just a large collection of PALs or GALs that can be individually programmed and then connected together with a large

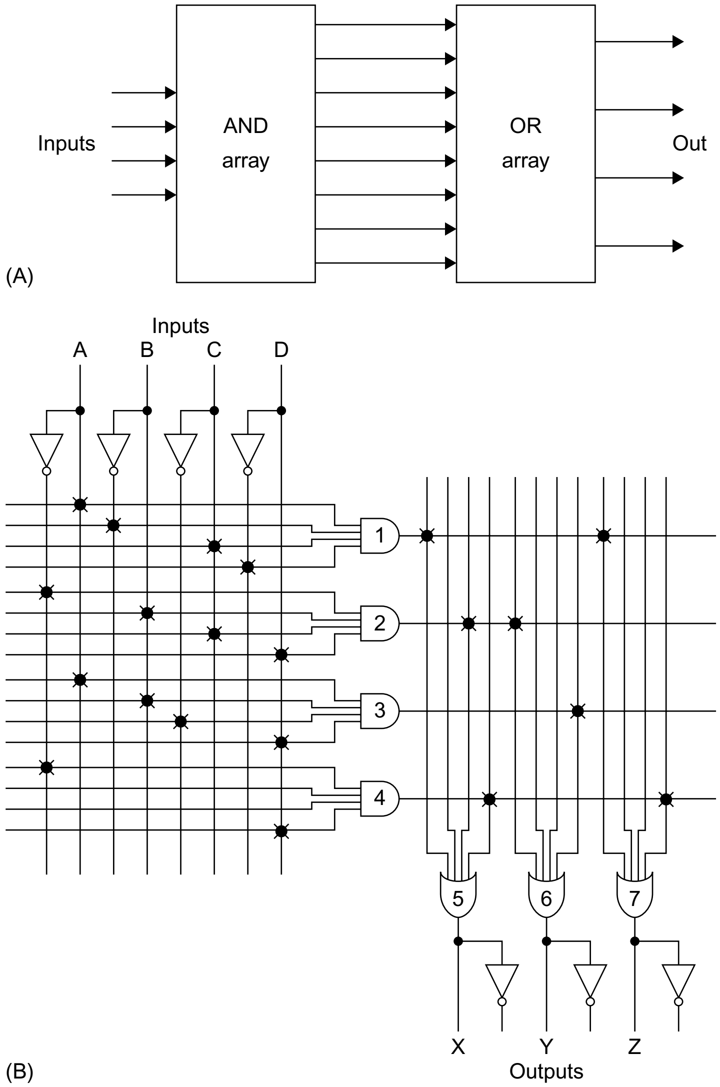

FIGURE 5.25 PLD concept. **(A)** Block concept of programmable AND and OR gate arrays. **(B)** Simple circuit of PLD showing programming by dots on AND and OR gate arrays.

interconnect matrix. Input/output (I/O) circuits also provide inversion if needed or extra power to drive external circuits. Again, flash memory is used to store the programming.

The largest and most flexible PLD is the field programmable logic array or FPGA. It has become the most popular and common way to implement digital equipment other than a microcontroller. FPGAs are large chips that contain thousands of logic circuits. The largest devices have millions of gates and other circuits. Most FPGAs are programmed by storing a bit pattern in the SRAM cells inside. These cells tell the logic circuits what configuration to take and what functions to perform. If you turn the power off, the programming goes away and all you have is a dumb FPGA with no logic interconnections or functions. When you power up an FPGA, the desired program, usually from a flash memory, or a computer loads the RAM cells with the program, thereby enabling the FPGA to perform its desired function.

The logic inside an FPGA is essentially thousands of look-up tables (LUTs) that can be individually programmed. Each LUT has its own program memory block as well as one or more FFs and I/O circuits. The LUTs may also be interconnected to one another by programming. With this arrangement virtually any digital logic operation can be programmed. You can even make your own custom computer with an FPGA, for example.

There are about a half-dozen companies that make PLDs. The two oldest and largest are Xilinx and Altera. Both have extensive lines of CPLDs and FPGAs.

DATA CONVERSION

The term *data conversion* refers to the translation of analog signals into digital form and vice versa. While it used to be an all-analog electronics world, today it is virtually all digital. We still have analog signals such as voice, video, radio waves, and sensor voltages, but mostly we process, store, and transmit data by digital means. For that reason, circuits that perform these conversions are critical to virtually all electronic applications. All cell phone voice is digital, all modern TV is digital, and sensor outputs are captured and processed in digital form. It is difficult to name a product today that does not use some form of data conversion.

Analog-to-Digital Conversion

The process of converting an analog signal into a digital one is called analog-to-digital conversion, and it is performed by an analog-to-digital converter (ADC). The process, also referred to as sampling, is illustrated in Figure 5.26. The ADC looks at the analog input and periodically takes a sample of the voltage at that instant, captures it, and then converts it into a proportional binary number. We say that we are digitizing the signal. The sample points are shown by the dots on the analog curve. The binary value of the sample is shown to the right of the curve. The conversion process actually results in a sequence of binary numbers that represent the analog waveform. These values are usually stored in a RAM or transmitted to other circuits as shown in Figure 5.27. Note the symbol for an ADC.

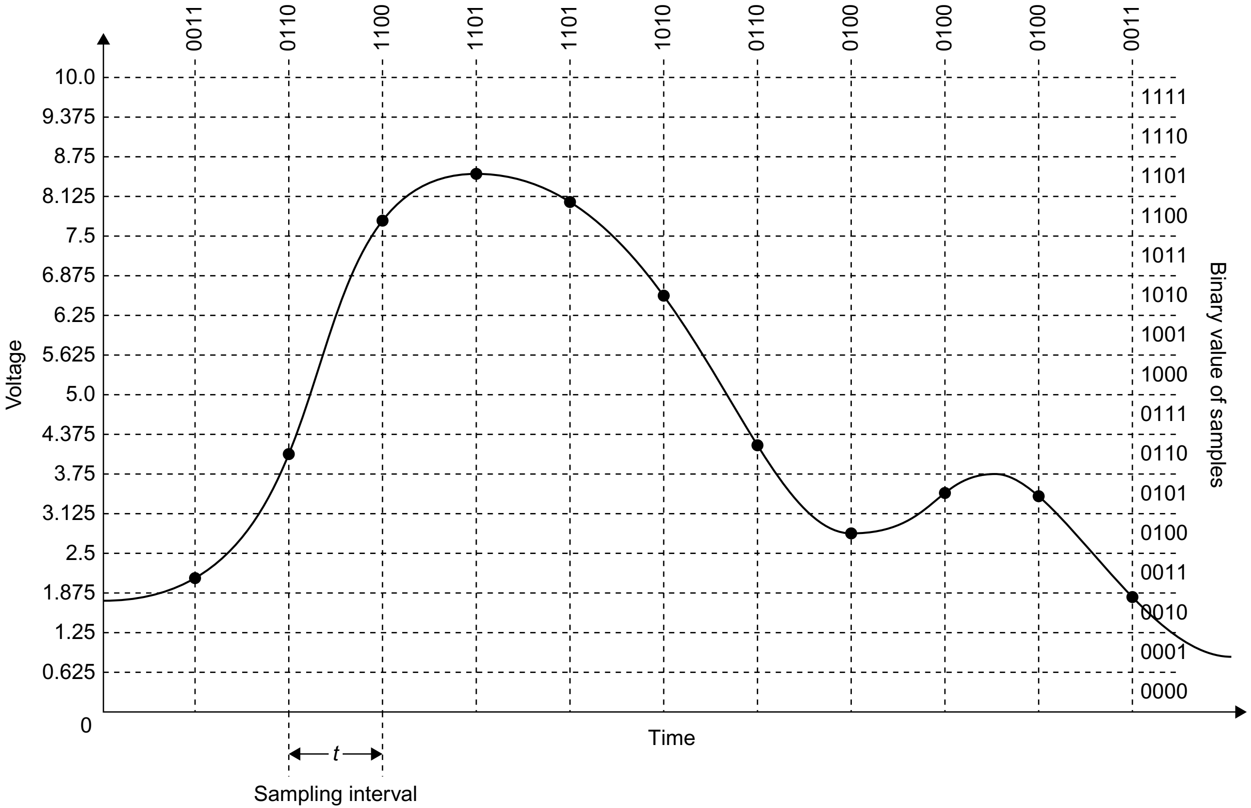

FIGURE 5.26 Analog-to-digital conversion is performed by sampling analog signal at equal intervals and generating proportional binary value.

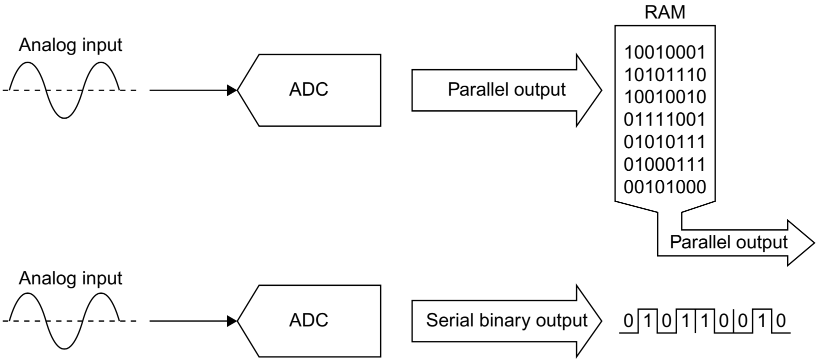
RAM
Analog input
ADC
Parallel output
10010001
10101110
10010010
01111001
01010111
01000111
00101000
Parallel output
Analog input
ADC
Serial binary output
0 1 0 1 1 0 0 1 0

FIGURE 5.27 Symbol for ADC with both parallel and serial output examples.

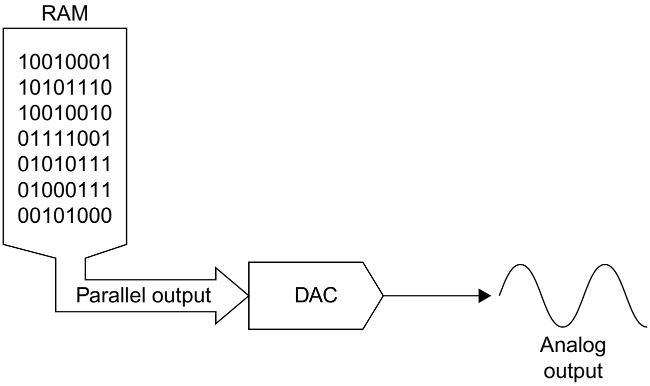
RAM
10010001
10101110
10010010
01111001
01010111
01000111
00101000
Parallel output
DAC
Analog
output

FIGURE 5.28 Converting binary data back into analog signal with DAC.

Digital-to-Analog Conversion

To recover the original signal, we put the data sequence previously captured by the ADC into a digital-to-analog-converter (DAC) (see Figure 5.28). The output is a version of the analog signal. The DAC output is not a perfect reproduction, but just an approximation. This is shown in more detail in Figure 5.29. Each binary input results in a constant voltage output from the DAC during the sample period. The result is a stepped approximation to the original signal. The rate at which the binary data is sent to the DAC must be the same as the sampling interval to recover the original frequency information in the signal.

Resolution and Sampling Interval

The key to good data conversion is to use greater resolution and faster sampling rates. Resolution refers to the number of bits used in the data conversion.

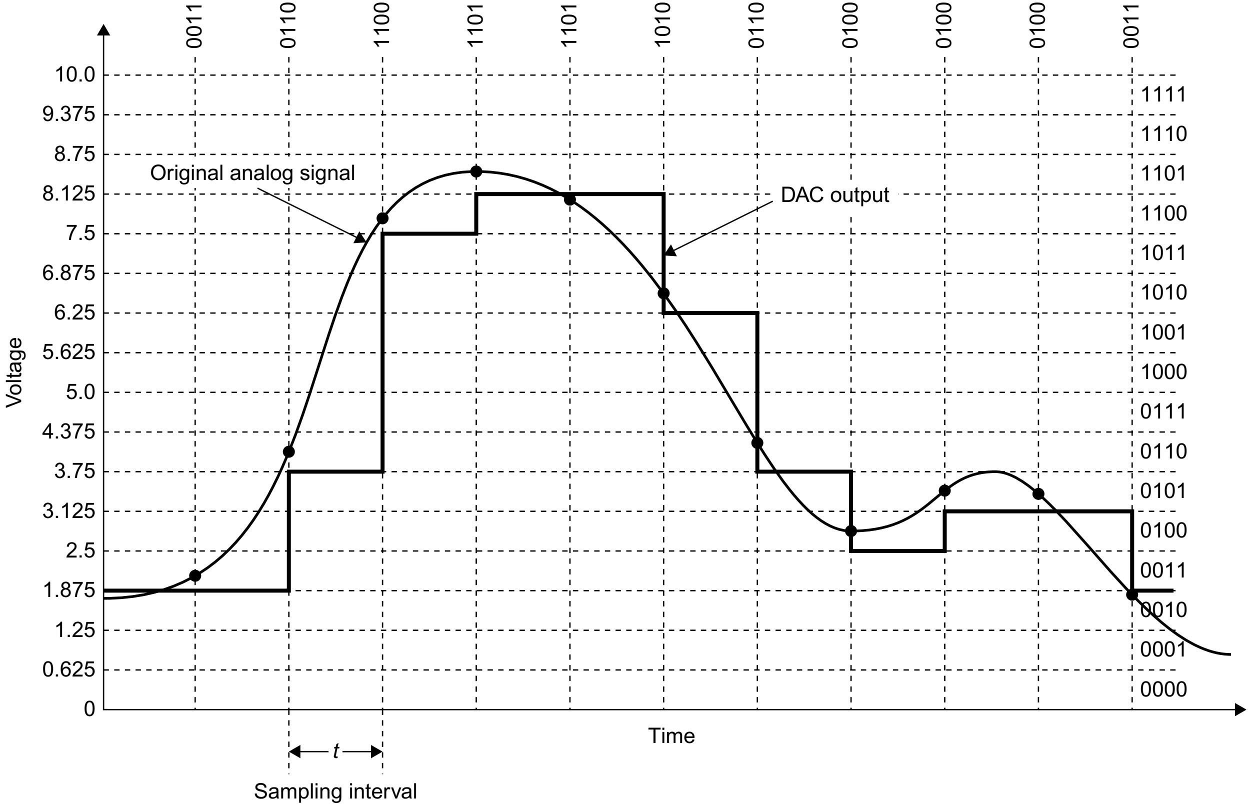

FIGURE 5.29 DAC output is stepped approximation to original analog signal.

In Figure 5.27, only 4 bits are used, so the resolution is poor. The voltage range is only divided into 16 intervals, meaning that amplitude variations at less than 0.625 volt are missed. This problem can be corrected by using more bits. ADCs are available in many bit sizes. The most common are 8, 10, and 12 bits, but 14 and 16 bits are available. Some methods of ADC produce resolutions of 20 to 26 bits. The result is a finer conversion of amplitude detail. As an example, if the 0- to 10-volt range in Figure 5.27 was sampled with a 12-bit ADC, the individual smallest voltage increment that can be detected is $10/2^{12} = 10/4096 = 2.44\,\text{mV}$ instead of the 0.625 volt in the figure.

Another critical specification is sampling rate. To retain all the frequency detail in a signal, the sampling rate must be at least twice the highest frequency in the signal. This is called the Nyquist criterion. For example, when digitizing music with a frequency range of 20 Hz to 20 kHz, the sampling rate must be at least double the 20-kHz frequency. In most systems, a rate of 44.1 kHz or 48 kHz is used.

ADCs with sampling rates to several hundred megahertz are commonly available, meaning that even radio signals can be digitized.

Project 5.1

Digging Deeper into Digital

If you want to learn more about digital circuits without investing in one or more new books, you can find some good material online. A few alternatives follow.

Do your own search on digital logic using the terms *tutorial*, *introduction to*, *fundamentals*, or *basics*. You will get lots of hits that you can sort through to find what you want.

Go to www.allaboutcircuits.com. Look at their digital section called Volume IV.

Go to openbookproject.net/electricCircuits. Look at the digital chapter.

On both of the websites, there is also a projects or experiments section that will guide you to further hands-on exploration.

Chapter 6

How Microcomputers Work

The Brains of Every Electronic Product Today

In this Chapter:

- Microprocessors, microcomputers, and embedded controllers defined.
- The stored program concept.
- Central processors and how they work.
- Types of microcomputers.
- Modern examples of embedded controllers.
- Digital signal processing.

INTRODUCTION

Whenever you hear or see the word *microcomputer*, you probably think of a personal computer. A personal computer is, of course, one type of microcomputer. But you may not be aware of other forms of microcomputers. For example, do you know that almost every piece of electronic equipment you own or use has a microcomputer in it? This type of microcomputer is called an embedded controller. It is usually a single integrated circuit that performs all of the basic functions of a computer but is dedicated to a specific task. Such micros (*micro* is short for microcomputer) are in TV sets, stereo receivers, microwave ovens, CD players, DVD players, cell phones, blenders, copiers, and most other types of electronic devices. You will find them in your car (several of them), your iPod, your bathroom scale, gasoline pumps, and dozens of other things you use every day. Microcomputers do indeed make your day—you cannot live without them.

CONCEPTS AND DEFINITIONS

A microcomputer is a small digital computer that can take several different forms. It can be a single integrated circuit, or it can be a module made up of several integrated circuits on a printed-circuit board.

© 2010 Elsevier Inc. All rights reserved.
doi: 10.1016/B978-1-85617-700-9.00006-0

The two most common types of microcomputers are the embedded microcontroller and the personal computer. One special type of microcomputer is the programmable logic controller (PLC) used for industrial control (Chapter 12).

The larger microcomputers, such as personal computers, are made with a microprocessor. A microprocessor is a large-scale integrated circuit that contains most of the digital logic circuitry usually associated with a digital computer. This logic circuitry is referred to as the central processing unit (CPU). A microprocessor is a single-chip CPU.

Another common micro is called a *core*. A core is a microprocessor or microcomputer integrated with some other circuits. The core is made on the same silicon chip as a cell phone or a security device. Today multicore micros are common. Two or more cores are used to get more processing power.

Embedded microcontrollers are complete micros on a single chip, including the CPU, memory, and input/output circuits. But in all cases, a microcomputer is an assembly of digital logic circuits, such as gates and flip-flops, that is used to process data. It is sometimes referred to as a *data processor*, or simply *processor*.

Data, of course, refers to the binary numbers and words the processor works with. Processing refers to the way the data is manipulated or handled. Types of processing include arithmetic, logic, sorting, translating, editing, counting, and searching. Any action taken on the data is called processing. Processing normally implies that the data is changed in some way or is used to create new data. Data that is not processed as suggested above is dealt with in other ways. Four common ways are storing, retrieving, input, or output. The data is not changed by any of these techniques.

Storing data means putting it in a safe place, like a semiconductor memory or disk drive, where it can be accessed later. *Retrieving*, of course, is the opposite of storing, or going to get the data for reuse.

Input means taking data into the computer to be stored or processed. *Output* means sending the data from the computer to some external device. Input and output (I/O) are ways to transmit data from one place to another.

Another feature of a microcomputer is its *decision-making* capability. During processing, the computer can make decisions and alter its sequence of operation. In other words, the computer can “make up its mind” based on the state of the data or outside conditions.

For example, the computer can tell if a number is less than, greater than, or equal to another. It can choose among alternative courses of actions or say yes or no, true or false, if given enough input facts.

Another major application of computers is *control*. Computers can be used to actuate relays and solenoids or turn lights and motors off and on (see Figure 6.1). In control applications, the computer actually determines when external devices are turned on or off. The computer serves as an electronic clock to time various operations. One example is the embedded single-chip microcontroller inside a microwave oven or a washing machine. Another example is a programmable logic controller (PLC), a special type of microcomputer used

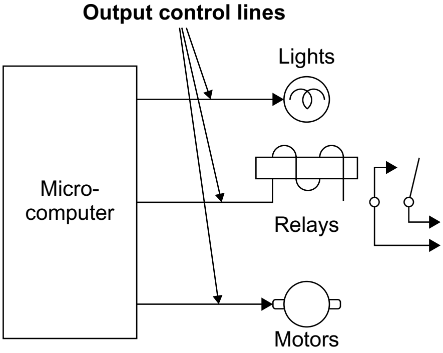
Output control lines
Lights
Micro-
computer
Relays
Motors

FIGURE 6.1 How a micro controls external devices.

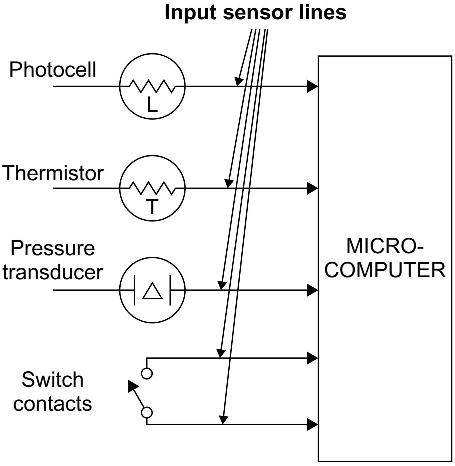
Input sensor lines
Photocell
L
Thermistor
T
Pressure
transducer
Switch
contacts
MICRO-
COMPUTER

FIGURE 6.2 How a micro monitors external devices.

in industrial applications, that is commonly used to sense, sequence, and time operations in a factory.

The most important thing to remember is that the control is automatic. The micro knows when to perform the various operations that it has been assigned because it has been programmed to do so.

In order to implement certain control applications, the micro must also perform a monitoring function. The micro "looks at" the process or devices being controlled to see what is happening. For example, the computer can monitor switch closures to determine physical state or position, pressure, and many other parameters. Transducers (sensors) convert the monitored physical characteristics, such as temperature or light level, into electrical signals that the computer can understand and respond to (see Figure 6.2). In most microcontrollers or PLCs, the controlled output change takes place only if a specific

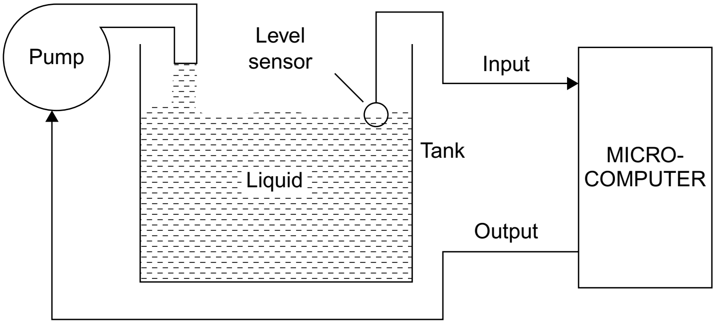

FIGURE 6.3 Controlling liquid level in a tank automatically with a micro.

input condition is sensed or not sensed. For example, if a temperature sensor indicates that the temperature has risen to a specific level, the computer will turn on a cooling fan.

Micros sometimes change the control function in response to one or more of the inputs they are monitoring. For example, if the micro senses that the liquid in a tank exceeds a given level, it can automatically turn off the pump that is filling the tank. If the liquid level goes below a predetermined level, the micro detects this and automatically starts the pump (see Figure 6.3). The key point here is that the computer makes its own decisions based on the input data that it receives. The result is full automation of some process or device.

The micro's ability to monitor and control operation may be used to automate simple processes in a toy or to operate entire factories.

COMPUTER ORGANIZATION AND OPERATION

All microcomputers are made up of four basic sections: memory, control unit, arithmetic-logic unit, and I/O unit. A general block diagram of a microcomputer showing these four sections is given in Figure 6.4. These four sections communicate with one another over multiple parallel electrical conductor data paths called a bus, as shown, to process the data or perform a control function.

Note that the control and arithmetic-logic unit (ALU) are shown together in a common structure. This is the central processing unit (CPU). The CPU is, of course, usually a microprocessor.

The memory is that part of the computer where data and programs are stored. The memory in any computer may contain thousands, millions, or even billions of locations used for storing numbers, words, or other forms of information.

Two primary types of information are stored in computer memory. The first type is the data to be processed. These are the codes, numbers, letters, and other forms of data to be manipulated.

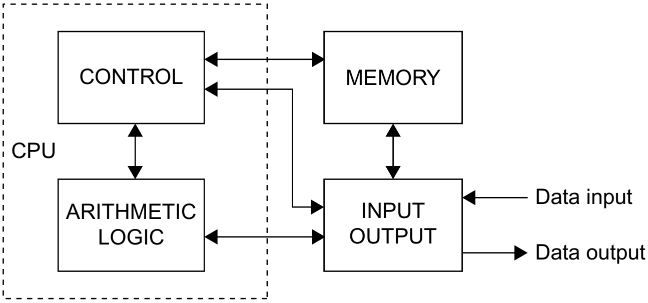

FIGURE 6.4 General block diagram of any digital computer or microcontroller.

The other type of data stored in memory is an instruction. Instructions are special binary numbers or codes that tell the computer what to do. Instructions specify the ways in which the data is to be processed. For instance, there are instructions that cause arithmetic operations to take place or data to be transferred from one place to another.

The instructions listed in a special sequence form a program. A program is a list of instructions that causes data to be processed in a unique way. A program is a step-by-step procedure that solves a problem, performs a control operation, or otherwise manipulates data according to some recipe. Programs are stored in memory along with the data that they process. In some micros there are separate memories for data and instructions. The programs that a computer uses are called software.

Data and instructions can be stored in or retrieved from memory during processing. When data is stored, we say that it is written into memory. When data is retrieved, we say that it is read from memory. A typical write operation transfers data from an analog-to-digital converter and stores it in memory. A common read operation accesses data in memory to be transferred to an LCD screen.

Now let's look at the operation of the CPU. The control unit is that portion of the digital computer responsible for the automatic operation. The control unit sequentially examines the instructions in a program and issues signals to the other sections of the computer that carry out designated operations.

Each instruction is fetched (read) from memory by the control unit, interpreted, and then executed one at a time until the program is completed. This is called the fetch-execute cycle, which is repeated on each instruction until the program runs to completion. The execution of each instruction may call for accessing one or more data words in the memory or storing a data word in memory.

The ALU is the section of the computer that carries out many of the functions that are specified by the instructions. In other words, the ALU actually processes the data. Specifically, the ALU carries out two main types of processing: arithmetic operations (such as addition, subtraction, multiplication, and division) or logic operations (such as AND, OR, complement, or exclusive OR). For example, if an add instruction is stored in memory, the control section will fetch it, interpret it, and send signals to the ALU that cause two numbers to be added.

The ALU also performs data movement operations. It can move data or instruction words from one place to another inside the CPU or it can carry out memory read/write or input/output operations. These are called load and store instructions.

A key part of the CPU associated with the ALU is the registers. Most ALUs have two or more registers. High-powered CPUs may have a group of 16 or more registers, called *general-purpose registers* (GPRs). The registers are used to temporarily store the data being processed by the ALU and the results of the computations. Other registers in the CPU store a number called an *address* where the data or instructions are stored, store the instruction being executed, or act as a stop-off place for data into or out of the CPU.

The control section, ALU, and the registers in all micros are very closely related. They operate together and are always considered as a single unit. As indicated earlier, the combination of the control and the ALU sections is called the CPU (see Figure 6.4). In addition, microprocessors are single-chip CPUs. Besides being called CPUs, microprocessors are sometimes called MPUs or *microprocessing units*. You will also see the expression μP used to refer to a microprocessor. The μ is the Greek letter mu, which means micro; the *P* means processor.

A microcomputer consists of a microprocessor (the CPU) plus external memory and I/O circuits. Or it could be an embedded controller with everything on one chip. The memory is either SRAM or DRAM for data and programs and EEPROM or flash for the program.

The I/O unit of the computer is the set of logic circuits that permits the CPU and memory to communicate with the outside world. The I/O circuits are referred to as *interfaces*. All data transfers into and out of the computer pass through the I/O interfaces.

The external peripheral devices, or *peripherals*, connected to the I/O unit are electronic or electromechanical units that are used for data entry or data display. Data is most commonly entered into the computer through an input keyboard or a disk drive.

The output data is usually displayed on a video monitor or small LCD screen. “Hard copy” on paper is created by a printer. There are a wide variety of other external input/output devices such as scanners, voice synthesizers, and barcode readers.

In an embedded microcontroller, the peripheral devices include keyboard, ADC, sensors, magnetic stripe readers, barcode scanner, and so on. Outputs go to relays, motors, lights, and liquid crystal displays (LCDs).

OPERATIONAL DETAILS

Registers and the ALU

The main circuit element in a CPU is the register. Usually the data to be processed by the CPU is taken from memory and stored temporarily in a register. Like a memory location, a register usually has a fixed length such as 8, 16, or 32 bits. But unlike a memory location, registers are often used to manipulate data as well as store it. That is, the register can alter or process the data in some way.

Examples of how a register can process data are shown in Figure 6.5. A binary word in a register can be shifted one or more bit positions to the right or left, or the register may be connected as an up/down counter so that it can be incremented (add one to the content) or decremented (subtract one from the content). A register can also be reset or cleared, thereby erasing any data in it and leaving the content zero.

Data transfers and manipulations performed on a register are initiated by individual computer instructions. For example, one instruction may cause the register to be loaded from a memory location. Another instruction may cause the word in the register to be transferred to a memory location. All data transfers are parallel, meaning that all bits are moved simultaneously from the source to the destination.

The main working register in most digital computers is called the *accumulator*. It may also be called the *A register*, *W register*, or something else.

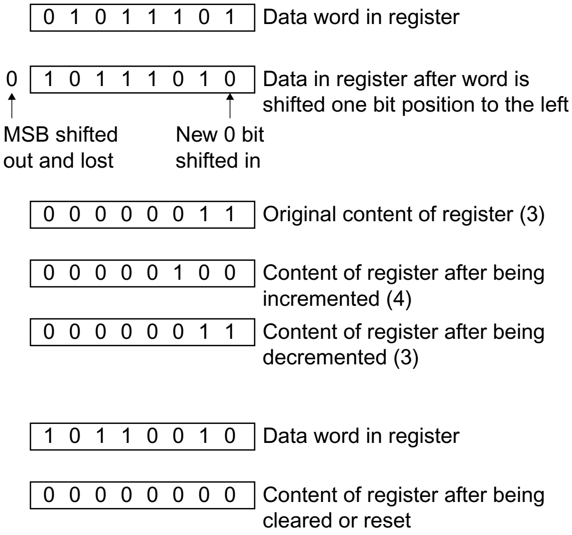
0 1 0 1 1 1 0 1
Data word in register
0
1 0 1 1 1 0 1 0
Data in register after word is shifted one bit position to the left
MSB shifted out and lost
New 0 bit shifted in
0 0 0 0 0 0 1 1
Original content of register (3)
0 0 0 0 0 1 0 0
Content of register after being incremented (4)
0 0 0 0 0 0 1 1
Content of register after being decremented (3)
1 0 1 1 0 0 1 0
Data word in register
0 0 0 0 0 0 0 0
Content of register after being cleared or reset

FIGURE 6.5 How data is manipulated in a register. Examples of shifting data, incrementing and decrementing a register, and clearing (resetting) a register to 0.

The accumulator can hold one word whose bit length is equal to that of a memory location. An 8-bit microprocessor has an 8-bit accumulator. The accumulator is part of the ALU.

The accumulator can be loaded from memory or the accumulator content can be stored in any memory location. I/O operations with peripheral units sometimes also take place through the accumulator.

Data to be processed by the arithmetic/logic section is usually held in the accumulator. This data is fed to the ALU. The ALU is a digital logic circuit that can add, subtract, and perform a wide variety of logic operations.

The ALU is capable of processing two inputs, one from the accumulator and one from the memory of another register (see Figure 6.6). Let's assume that you wish to add two binary numbers. To do this, you first load one of the numbers (the *augend*) into the accumulator. This is done with a load-accumulator instruction that takes the number from a memory location and puts it into the accumulator. Next, an ADD instruction is executed. This causes another word (the *addend*) to be taken from memory and placed in register B. It is then added to the content of the accumulator. The sum is usually stored back in the accumulator, replacing the augend originally there.

Some microprocessors have two or more accumulator registers that share a single ALU. Two accumulators provide greater flexibility in the manipulation of data than a single accumulator. Such multiple registers simplify, speed up, and shorten a program to perform a given operation. Some sophisticated CPUs have 4, 8, 16, or even more accumulators. Usually they are referred to as GPRs or a register file. These registers can each use the ALU and act as temporary storage locations for data and the intermediate results of calculations. Instructions are provided to move data from one register to another.

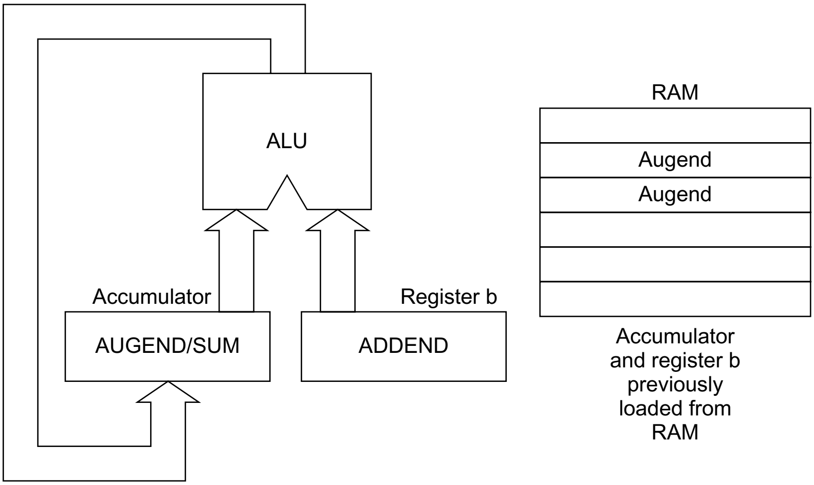

FIGURE 6.6 How ALU adds or otherwise processes two data values.

In most micros, the GPRs share a single ALU. The ALU accepts only two input words and generates a single output word. The two words to be processed by the ALU usually come from any two GPRs or any GPR and a designated memory location. The destination of the ALU output can also be one of the GPRs in some CPUs.

Control Unit

CPUs also have several other registers, including the instruction register; the program counter, also called the instruction counter; and the memory address register, also called the address buffer.

The *instruction register* (IR) is used to store the instruction word. When the CPU fetches an instruction from memory, it is temporarily stored in the IR. The instruction is a binary word or code that defines a specific operation to be performed. The instruction word is also called the *op code* or *operation code*. The CPU decodes the instruction, and then executes it.

The *program counter* (PC) is really a counter and a register. It stores a binary word that is used as the address for accessing the instructions in a program. If a program begins with an instruction stored in memory location 43, the PC is first loaded with the address 43. The address in the PC is applied to the memory, causing the instruction in location 43 to be fetched and executed. After the instruction is executed, the PC is incremented (add 1) to the next address in sequence, or 44. The instructions in a program are stored in sequential memory locations.

The *memory address register* (MAR) or *address buffer* also stores the address that references memory. This register directly drives the address bus and the memory address decoder in RAM or ROM. The MAR gets input from the PC when an instruction is to be accessed (see Figure 6.7). The MAR can also be loaded with an address that is used to access data words stored in memory. To retrieve a data word used in an arithmetic operation, the MAR is loaded with the binary word that points to the location of that word in RAM. This address is often a part of the instruction.

It is important to note that the PC and the MAR (address buffer) have a fixed length of so many bits. And that limits the amount of memory that can be accessed. For example, with a 16-bit address register, the address bus has 16 bits to address RAM and ROM. With 16 bits, a maximum of $2^{16} = 65{,}536$ words can be addressed.

There are usually two other registers, the *flag* and *stack pointer registers*. The flag or *F register* is an 8-bit register whose individual flip-flops are set and reset by the ALU as the various arithmetic and logic operations are carried out. Each flip-flop is called a flag. As an example, there are *zero* (Z) and *carry* (C) *flags*. If the accumulator content is zero after an operation is performed, the Z flag is set indicating this condition. If an arithmetic operation (addition) results in a carry from the MSB of the accumulator, the C flag is set indicating this

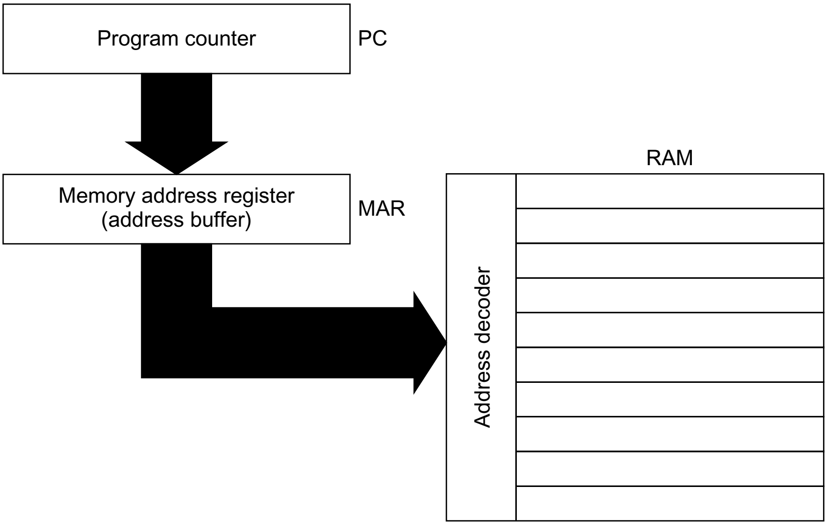

FIGURE 6.7 Program counter identifies the address in memory to be accessed.

condition. These flags can be monitored or tested by the control circuitry to change the sequence of processing.

The stack register is a 16-bit or larger register used to address a selected area of RAM known as the *stack*. This memory is used to store register contents and status information when subroutines and interrupts are used.

Instruction-Word Formats

There are three types of instruction formats used in typical 8-bit microprocessors, illustrated in Figure 6.8. In the single-word format, the instruction is a single 8-bit word. This word is called the *op code*. The op code tells the ALU, the registers, and other elements of the system what to do. In this format, no address is used. The focus of the instruction is implied in the instruction. That is, the data to be processed is already in a location designated by the instruction. Usually the data is in a register. Typical instructions using this format are register-to-register transfer, shift data left (or right), or clear to zero.

The two-word instruction format in Figure 6.8 requires two 8-bit words to define the operation. These two words are stored in sequential memory locations. The first word is the op code. The second word is usually an address that specifies a memory location where the data word to be processed is stored. For example, if the op code calls for an add operation, the address word designates the location in RAM of the number to be added to the contents of the accumulator. The 256 bytes of RAM can be addressed with 1 byte of address.

In some 2-byte instructions, the second byte is not the address. Instead, it is the data itself. This is called an immediate instruction since it is not necessary to address the data that is available immediately within the instruction itself.

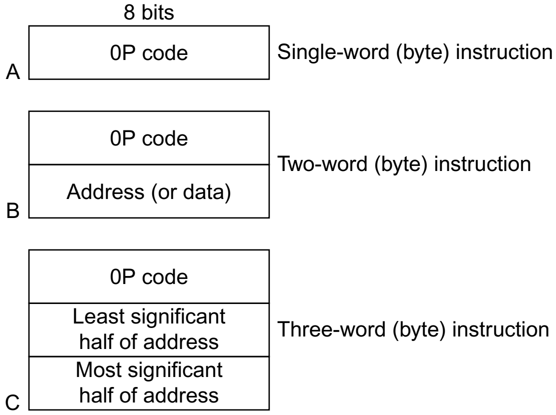
8 bits
0P code
Single-word (byte) instruction
A
0P code
Address (or data)
Two-word (byte) instruction
B
0P code
Least significant half of address
Most significant half of address
Three-word (byte) instruction
C

FIGURE 6.8 Three common instruction-word formats with addresses or data.

The three-word instruction format in Figure 6.8 is comprised of an 8-bit op code and two 8-bit address words stored in sequential memory locations. The second and third bytes together form a 16-bit address word that designates the location in RAM of the data to be processed. In the 3-byte instruction format, the first byte is the op code, the second byte is the least significant half of the address, while the third byte is the most significant part of the address. Different formats are used by other micros.

To access a word in RAM, the instruction address word must be stored in the MAR. This happens during the instruction fetch operation. When an instruction is fetched from memory, the op code is stored in the instruction register while the address is stored in the MAR. The instruction is then executed. The MAR usually gets its input from the PC. Once an instruction is fetched and executed, the PC is incremented.

The PC may be incremented once, twice, or three times, depending on the length of the instruction just executed. If a 2-byte instruction is executed, the PC is incremented twice so that the PC points to the address of the next instruction op code.

Program Execution Example

Now let's illustrate how a typical CPU executes a simple program. Here are a few common instructions similar to those found in any micro.

Move A to B (MOV B,A): This is a 1-byte instruction that causes the content of the accumulator (A) to be moved to register B. The content of the accumulator is not erased.

Load Accumulator (LDA): This is a 3-byte instruction that takes the content of the address specified by the second and third bytes and loads it into the accumulator.

Add Immediate (ADI): This is a 2-byte instruction that takes the content of the second byte of the instruction and adds it to the content of the accumulator. The sum is stored in the accumulator, replacing the number that was there previously.

Halt (HLT): Halt is a 1-byte instruction that stops processing.

Note that we refer to each instruction by a three-letter abbreviation called a *mnemonic*.

Figure 6.9 shows a complete CPU and the RAM containing a simple program. The first instruction in the program is the LDA stored in location 0000.

To begin the execution, the address of the first instruction is loaded into the PC that points to the location of the instructions in RAM.

Next, the content of the PC is transferred to the MAR that addresses RAM. The LDA instruction op code is fetched and stored in the instruction register. The address bytes of the LDA instruction in locations 0001 and 0002 are then transferred to the memory address register. The address of the data to be loaded into the accumulator in bytes 2 and 3 is 0008. The desired data word in location 0008 is retrieved and loaded into the accumulator. This is the augend (23).

After the LDA instruction is executed, the number 23 is stored in the accumulator. Next, the PC is incremented three times. The ADI instruction is then fetched and stored in the instruction register. The number to be added is stored in location 0004.

This is an ADD immediate instruction where the second byte contains the data, in this case the addend (56). The number in location 0004H is 56, which is added to 23 in the accumulator, creating the sum 79. The mnemonic of the next instruction to be executed is MOV B,A in location 0005.

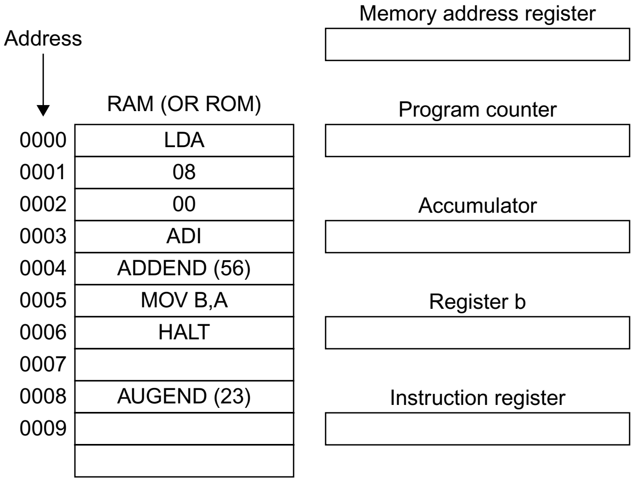

FIGURE 6.9 Example of program stored in RAM for execution using various CPU registers.

The sum in the accumulator is then moved to the B register by the MOV B,A instruction. The content of the accumulator is not changed when data is transferred to the B register.

The PC is incremented once since the HLT is a 1-byte instruction.

As you can see by this step-by-step analysis of the execution of a computer program, there is nothing mysterious about its operation. The CPU simply fetches and then executes the sequentially stored instructions at very high speed until the operation is complete. This process is sometimes referred to as the *stored program concept*.

Microcomputer Buses

Data transfers inside a microprocessor take place in parallel. This means that all bits in a word are transferred simultaneously from one place to another. It takes only a few nanoseconds (one billionth of a second) for all data bits in one register to be moved to another register.

The parallel data transfers take place over a data bus. A bus is simply multiple parallel electrical connections from a source to a destination. A number of these buses are contained within the CPU and are known as internal buses.

Microprocessors usually have three major buses—a *data bus*, an *address bus*, and a *control bus*. These are made available to external circuits in some micros. A typical 8-bit CPU has an 8-bit data bus and a 16-bit address bus. The data bus sends data to and from the CPU, RAM, ROM, and I/O sections. All data transfers between the CPU and memory or I/O sections take place over the data bus. The address bus drives all of the memory and I/O devices.

When an instruction is fetched from RAM or ROM, it is transferred over the data bus from the memory into the instruction register. Any data word retrieved from memory or a peripheral device via the input section also passes over the data bus into the accumulator or GPR.

When a store instruction is executed, the word in the accumulator is transmitted over the data bus into RAM. Data can also be transferred from the accumulator over the data bus to an external device such as an LCD. Or, data from an input device such as a keyboard passes over the data bus and is placed into the accumulator. The important point here is that data can move in either direction over the data bus. We say that it is a *bidirectional bus*.

Another key point is that the data bus can be connected to only one data source at a time. Data can originate only at a single source, but it can be sent to one or more destinations. To accomplish this, circuits called *bus multiplexers*, or *three-state line-driver circuits*, are used to connect or disconnect the various data sources to or from the bus. Figure 6.10 illustrates this concept.

The address bus is a unidirectional bus. It transfers an address from the CPU to all external circuits (memory and I/O). Address words are produced in the CPU. The PC generates the address that points to the instruction to be fetched. The content of the PC is transferred over a parallel 16-bit internal

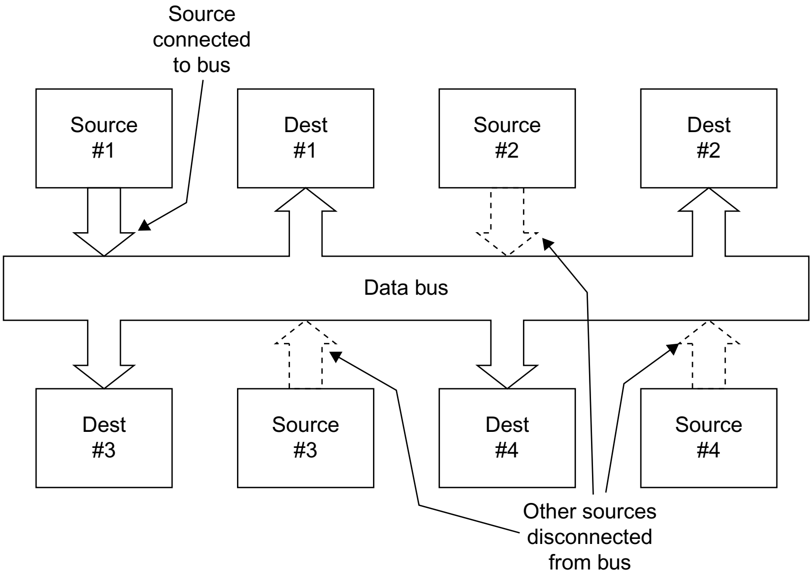

FIGURE 6.10 Only one source may transmit at a time on a bus. Others are disconnected. All destinations can receive the transmitted data. Dest=destination.

address bus to the memory address register or address buffer. The output of the MAR or address buffer is the address bus.

The control bus consists of numerous signals generated by the microprocessor and used to initiate various memory or I/O operations. The control bus also contains input lines from external circuits that tell the CPU what to do and when. The number of control signals varies from CPU to CPU.

POPULAR MICROS

The first practical microprocessors and microcomputers became available in the mid-1970s. Over the years lots of different models have been developed. As semiconductor processing technology has improved over the years, it has become easier and cheaper to put more circuits on a chip plus the memory that earlier was always external. And processing speeds have multiplied many times over. Today's embedded controllers are indeed complete computers on a chip and they run at very high clock rates. And their cost is low. In high-volume manufacturing, an embedded controller can cost less than $1.

Most of the initial processors were 8-bit devices. These are still popular today. Larger, faster processors are available for not much additional cost. Sixteen- and 32-bit processors are commonplace. Even 64-bit devices are available if needed. Many are individual chips, while others are available only as cores on a larger systems chip. Following is a summary of a few of the most widely used processors.

8- and 16-Bit Microcontrollers

8051

The 8051 is one of the oldest but still widely used. It was originally developed by Intel in the late 1970s and 1980s. It is still around and available from a half dozen or more companies. The more recent versions are faster and have more memory and I/O features than the original. Original clock speeds were only 8 to 12 MHz, but today you can get models with a clock rate up to 100 MHz.

A general block diagram of the 8051 is given in Figure 6.11. The 8051 features an internal RAM of 128 or 256 bytes but can handle 64 or 128 KB of RAM externally. The internal ROM can be had from 512 bytes to 128 KB. Some models also feature flash memory. As for I/O ports, the figure shows four 8-bit parallel I/O ports. You can get a standard serial port called a UART, and more specialized I/O ports such as SPI and I^2C. See the Interfaces sidebar in this chapter. Many models also come with an internal 10- or 12-bit ADC.

A few other items in this diagram that were not previously discussed are the clock, interrupts, and timers. The *clock* is an oscillator that controls all operations in the micro. Its frequency is usually set by an external crystal that maintains a precise and stable frequency. The higher the frequency, the faster the operations. Fast is good.

An *interrupt* is, as its name implies, an external signal that can interrupt the CPU when an important event occurs. It can signal the CPU of the event and even cause the CPU to stop executing the current program and jump off to another program to deal with the interrupt.

Timers are usually just one or more up/down counters than can be programmed with various count values. These are used to implement count functions in programs and to provide timing operations for I/O operations. Most controllers have one or more timers.

68HC11

Another widely used 8-bit controller is the Freescale 68HC11 and its various versions. There are dozens of offshoots and variations. It has two 8-bit accumulators and one 16-bit accumulator. Total memory space is 64 KB with 128, 256, or 512 bytes of RAM, and up to 48 KB of ROM program memory. Different versions have all the interfaces, both parallel and serial, and some models have a built-in 8-bit ADC. A 16-bit version called the 68HC12 is also part of the product mix.

PIC Processors

Microchip Technology makes one of the most widely used 8-bit controllers. Called PIC processors, they are extremely simple and cheap. There are hundreds of versions with different memory sizes and configurations and various mixes of interfaces. One of the simpler versions is the 16C54. It has a single

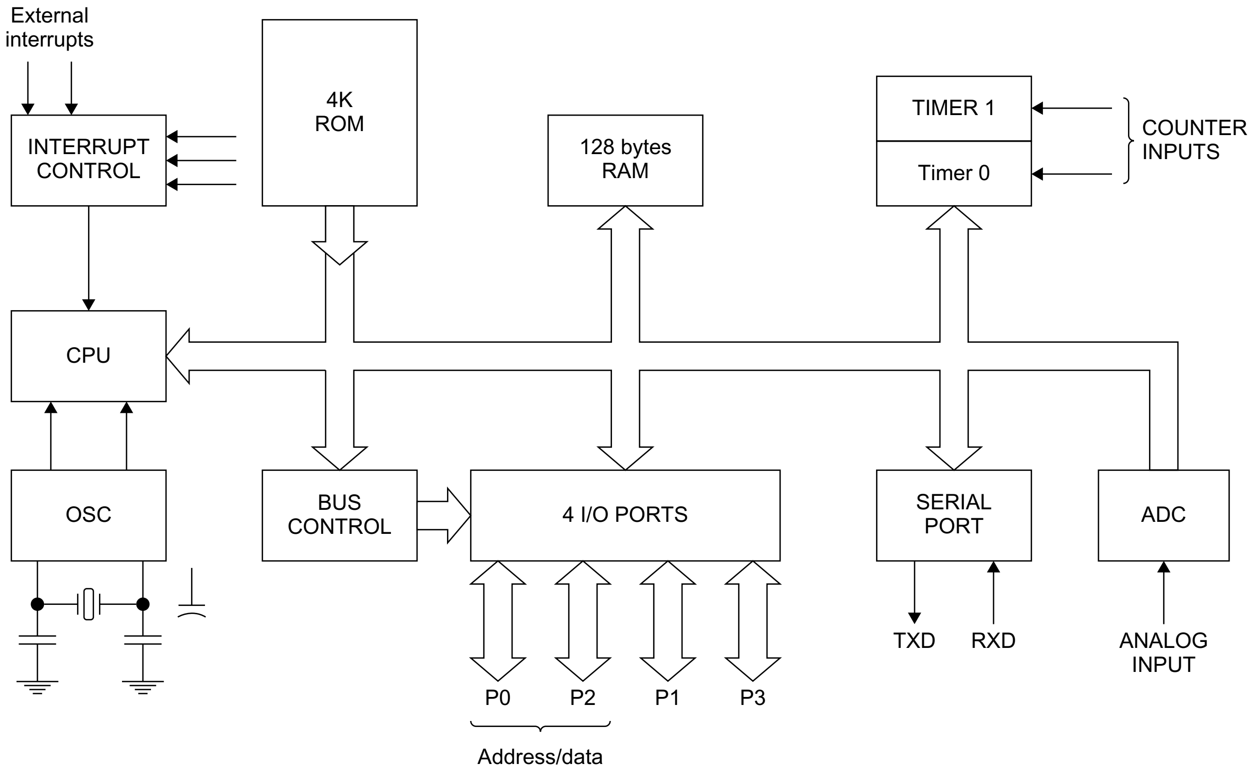
External interrupts
INTERRUPT CONTROL
4K ROM
128 bytes RAM
TIMER 1
Timer 0
COUNTER INPUTS
CPU
OSC
BUS CONTROL
4 I/O PORTS
SERIAL PORT
ADC
TXD
RXD
ANALOG INPUT
P0
P2
P1
P3
Address/data

FIGURE 6.11 General block diagram of 8051 microcontroller. It is very similar to most other controllers in terms of major circuits and their interconnections.

working register, W, that is the accumulator. It uses a bank of 24 to 73 registers for RAM. ROM for program storage ranges from 512 bytes to 2 KB. I/O occurs via 4- or 8-bit parallel ports. Larger versions have more RAM and more ROM, higher clock speeds, a built-in ADC, and special network interfaces like CAN. Several 16-bit versions are also available.

MPS430

A typical 16-bit microcontroller is the Texas Instruments MSP430. It has 16 GPRs and a clock speed of 16 MHz. On-chip RAM is SRAM, and can be had in versions with 128 bytes to 10 KB. ROM for the program is flash from 1 KB to 120 KB. It has an ADC on board, a DAC, timers, and all the common serial interfaces. It is a popular chip because its low power consumption makes it ideal for portable and battery-powered devices.

32- and 64-Bit Processors

As it turns out, it is almost as cheap and easy to make a 32-bit processor as it is to make an 8-bit device. For that reason, for a few cents more, designers can make their products faster and more powerful by selecting a 32-bit device. With this many bits being processed at the same time, it is possible to move more data faster, compute with increased resolution, and to perform much more complex operations faster and easier. Thirty-two-bit processors can also address much more memory. A few years ago, 32-bit processors were reserved for only the largest or most critical of needs. Today, the 32-bit processor is almost as widely used as the 8-bit processor.

Three devices account for most applications. They are made by ARM Holdings, Freescale, and MIPS. ARM is the name of a British company that invented a super-simple, fast, and flexible 32-bit processor that does everything well at very low cost. It is referred to as a reduced instruction set computer (RISC) processor. It has a few simple instructions but is extremely fast. ARM does not actually make chips, but instead licenses the design to others. There are several vendors of ARM processors but most of them are actually cores that are integrated on-chip with other circuits to form a system on a chip (SoC). With over 75% of the market share, you are sure to encounter an ARM device sooner or later. Chances are you own at least one inside a product such as a cell phone, which is where most are used.

Freescale's PowerPC is another popular 32-bit processor. It was originally used in Apple Macintosh computers but no more. IBM used it widely and still does in some products. Many embedded versions have since been made. Its primary usage is in communications, networking, automotive, and industrial applications.

For really super processing power, the MIPS processors are a popular choice for certain embedded applications. These devices come in 32-bit as well as 64-bit models, in case you really have a hot application for which nothing else will do. The MIPS processors are great at advanced math and have excellent resolution.

Microcontroller Interfaces

An *interface* is the circuitry that connects an embedded controller to the outside world. Initially, all I/O in early micros was just the parallel data and address buses extended to accommodate additional I/O circuitry. Today, much of the I/O interface circuits are put right on the same chip with the rest of the micro. Following are the most widely used interfaces you are likely to encounter. All of these are serial interfaces. Parallel I/O is still used, but today the preference is serial I/O because of fewer interconnecting lines.

RS-232/UART—This universal asynchronous receiver-transmitter is actually the physical circuitry implementing a popular serial interface standard referred to as RS-232. It was originally developed for connecting computers to teletype machines and early video terminals. It is still used today, especially for connecting to industrial equipment for monitoring or control.

The RS-232 interface sends and receives data in bytes (8 bits) plus two extra bits, called Start and Stop bits, that tell the receiving device when the data begins and ends (see Figure 6.12). The logic levels are −3 to −25 volts for a binary 1 and +3 to +25 volts for a binary 0. Typical voltage levels are in the 5- to 12-volt range. The data rate is slow, typically from 300 bps to 115.2 kbps. Standard data rates are used, so no clock signal is needed. Common speeds are 2400, 9600, and 19,200 bps. The physical communications medium is a two-wire (signal plus ground) cable that can be up to 50 feet long. Special 9- and 24-pin connectors are part of the standard. Many PCs and laptops still have a 9-pin RS-232 port.

I^{2}C—Shorthand for inter-integrated circuit or I squared C, this is a low-speed serial bus that is used for sending data from one chip to another on the same PC board or over short cables between two pieces of equipment. The data rate is typically in the 10 kbps to 400 kbps range, but a faster version (3.4 Mbps) is

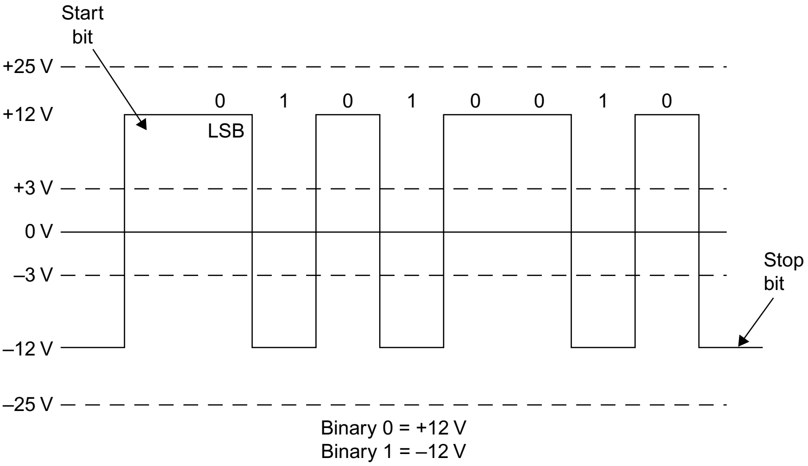

FIGURE 6.12 ASCII letter "j" being sent over an RS-232 interface using a UART.

(*Continued*)

Microcontroller Interfaces (Continued)

available. Speeds rarely exceed 100 kbps. Only three wires are needed, data, a clock signal that controls the data transfer, and a ground.

SPI—This is the serial peripheral interface. It uses four wires, data in, data out, clock, and a select signal that determines whether a device is a master or slave and either sending or receiving data. A fifth connection is ground. It is much faster than the I^2C interface, with speeds up to 20 Mbps.

CAN—This is controller area network serial bus in an interface used to link micros together in small networks. It is used in automotive and industrial applications. The usual cable is twisted pair and a ground. Speeds range from 10 kbps to 1 Mbps with 20 kbps being typical. The data is sent in specifically formatted frames as determined by a standard protocol. It uses start and stop bits like the RS-232 and some similar control codes defined in ASCII. It also includes error detection and correction capability, so it is very reliable.

USB—The universal serial bus is the most common PC interface for peripherals including keyboard, mouse, printer, flash drive, external disk, and so on. It is also used in some embedded micros. The interface uses two wires but also includes +5 volts DC and ground. Version 2.0 provides speeds to 480 Mbps over cables no longer than 16 feet. A new higher-speed version 3.0 is also available with speeds up to 5 Gbps. Data is sent in bytes in packets according to a unique protocol.

RS-485—This is an interface widely used in industrial applications. It uses twisted pair cable that can be up to 4000 feet long. The cable is actually a bus to which many devices can be attached. Data rates can be up to 10 Mbps for shorter lengths. Various protocols may be used.

A WORD ABOUT SOFTWARE

The microcontroller can perform almost any digital function from the simplest logic operation to a complex mathematical calculation. The only problem is that you have to tell it to do what you want. And you do that by programming. A program is the sequence of instructions that detail the order in which you do the various processing steps that lead to your application. The controller chip itself is a dumb piece of hardware waiting for its program. What this means is that today, designing an electronic product is as much about writing programs as it is building electronic hardware.

Software is the name given to all the programs that a processor uses for developing and executing a program. The program that implements your application is the applications program. Other types of software used with the processor follow:

Languages—A programming language is a set of rules or syntax that defines how you tell the processor what to do. A language is the tool that you actually use to create your application program. It is a kind of shorthand code that lets you express the operations in simple terms.

Editor—An editor is like a word processing program that lets you type in your program. You then save it as a file that the compiler or assembler uses to create machine code.

Compiler—A compiler is the program that translates your program code into binary code that can be stored in the memory and used by the processor. Processors only understand binary code but humans are not good with such code. So, you write your program in a language you can understand and the compiler converts it to the code that the processor understands.

Operating system—For large complex programs or for applications that require the processor to execute multiple programs at different times, you need a piece of software that manages the multiple programs and operations. The operating system (OS) oversees things like memory usage and access, timing, and I/O operations. Not all embedded controllers need an OS, just the more powerful ones with multiple functions. For example, most cell phones have an OS.

Development environment—This is a complete package of software tools that facilitates your software development. It consists of languages and compilers, as well as programs for testing and debugging your code.

Subroutines—These are short sections of code that perform a specific operation. It may be a math operation, some type of I/O process, or other function that is performed multiple times. You only write the subroutine once, and then access it when you need to do that operation. Many development systems have libraries of subroutines that you can tap so you do not have to program common functions yourself. Subroutines shorten your development time and speed execution.

Programming Languages

There are dozens of languages used from programming embedded controllers. The more popular ones are assembly language, C, and BASIC.

Assembly Language

Assembly language uses shorthand mnemonics such as three- or four-letter designations for the instructions that the processor can execute. Instead of programming with binary code, you use the mnemonics. In addition, you use names that you make up to identify memory locations, I/O ports, and other things normally referenced by a binary code. Following is what a simple assembly program looks like:

```
ORG $0000
LDAA FIRST
MUL SECOND
STA PRODUCT
OUT PORT 1
END
```

This programs loads one number from a memory location given the name FIRST, multiplies it by another number in memory location named SECOND. It then stores the result in another memory location called PRODUCT. It next outputs the value to an I/O port called PORT 1. The ORG $0000 simply says to start the program whose first instruction is in memory location 0000, a hex address.

The software that converts the assembly language program to binary code is the *assembler*. The assembler gives actual binary assignments to the memory locations named as well as the instruction op codes.

C Language

The C language is referred to as a higher-level language in that the various syntax and terminology of the language usually translates into multiple binary machine instructions. The C language lets you write the program in a more conversational or casual format using English instead of cryptic mnemonics and other terms. The language is a bit of a challenge to learn, but after that programming goes much faster. A compiler converts the C-language program code into binary code.

A simple C program may appear like this:

```
int x, y, z
x = 38
y = 57
z = 57 − 38
printf z
```

The first line identifies three integer values x, y, and z; 38 is assigned to x, and 57 to y. Then the difference between the two is designated z. Finally, the printf line accesses a subroutine that causes z to be printed.

BASIC Language

Some embedded controllers use a simplified version of BASIC. BASIC was invented before personal computers and initially became the main language of personal computer programming. Microsoft still sells its Visual BASIC software for program development. BASIC is very simple higher-level language that is easy to learn and use. The program created is then translated into binary code by a compiler. Alternately, the controller runs a program called an *interpreter* that resides in memory and executes the program a command at a time.

The following program is written in PBASIC, a simple version of BASIC used with an embedded controller called the BASIC Stamp made by Parallax. It uses a PIC microprocessor and stores a program in flash memory. The program causes an LED to flash off and on.

```
Flash
Output 7
High 7
Pause 500
Low 7
Pause 500
Goto Flash
```

The name of this program is Flash. The first line tells the controller to make pin 7 the output port. The next line tells pin 7 to go high or to binary 1. The next command says pause for 500 milliseconds. Next, pin 7 is made to go low or to binary 0 via the Low command. Pause makes it stay there for another 500 milliseconds. Finally, the Goto command causes the program to look back and repeat itself again and again. The result is an LED that turns off and on at a 1000-millisecond (1-second) rate.

DIGITAL SIGNAL PROCESSING

Digital signal processing (DSP) is the technique of programming a microcomputer to perform operations normally carried out by analog or linear circuits. The method is illustrated in Figure 6.13. Assume you want to process an analog signal by filtering. You could use a hardware filter made up of inductors and capacitors or resistors and capacitors. Another way to do it is with DSP. The analog signal to be filtered is first digitized with an ADC and the resulting data stored in a RAM. This data is then acted upon by a processor programmed to perform a filter function. Special mathematical algorithms have been developed to do this. The resulting calculations generate new data that is also stored in the RAM. This new data is then sent to a DAC where it is converted back to an analog signal. The result is that the signal is filtered as if were processed by an electronic circuit of capacitors and inductors. The processing was mathematical.

There are many other operations that can be performed by DSP. Filtering is the main application, but you can also do things like modulation and demodulation as needed in wireless applications. It will perform other things such as data compression, equalization, tone control, and even frequency analysis.

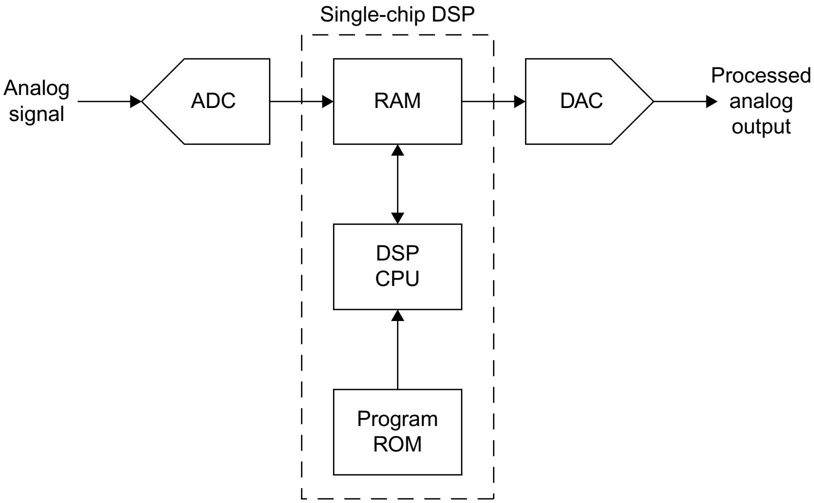

FIGURE 6.13 Basic make-up of digital signal processor and its ADC and DAC I/O.

Using a program called the fast Fourier transform (FFT), the DSP can analyze a signal and tell you exactly what frequency components are in it. It identifies the harmonic frequencies and their various amplitudes and phases just as in a frequency-domain plot.

The big question is why use a complicated technique such as DSP when you can do it with simple analog/linear hardware? Probably the best answer is "just because we can." With cheap single-chip processors, DSP is often no more expensive than an equivalent analog circuit. Since everything else is digital these days, it makes some sense to do as many analog functions digitally as well. But the best reason is that DSP provides superior results over analog methods. Filtering is substantially improved with better selectivity that was not possible with analog filters.

While DSP can be programmed on any processor or controller, it is best done on special DSP chips designed for the purpose. They use special architectures and processing schemes to speed up and improve the processing over standard conventional processors. A good example is that DSP chips implement a multiply and add or accumulate function, called MAC, that is common to most DSP processes. The MAC has to be programmed in a standard processor, but in a DSP chip it is implemented with special hardware that speeds up the process and simplifies programming. Some more conventional processors also incorporate DSP instructions to facilitate DSP operations.

Now that DSP chips are cheap despite their complexity and speed, DSP is widely used. It is found in most electronic products today, such as TV sets, cell phones, MP3 players, cable modems, and many other common products.

Project 6.1

Use a Microcontroller

One of the fastest ways to get familiar with a microcontroller is to use the BASIC Stamp made by Parallax. It comes as a complete unit with a microchip technology PIC processor and I/O circuits. One is a USB port to a personal computer. The software provided lets you program the controller in PBASIC. You write the program on a personal computer and download it to the EEPROM in the PIC. Then you run the program. It is ideal as a way to monitor and control lots of simple devices. And it is a fast and easy way to see firsthand what a controller does.

Go to the Parallax website at www.parallax.com and find the BASIC Stamp info and just order one. There are various models, from the ultra simple to the complex, and lots of accessories to help you implement various applications.

Project 6.2

Learn More about DSP

Go to the Texas Instruments website at www.ti.com. TI is the leader in DSP chips and has lots of useful information for your learning purposes.

Chapter 7

Radio/Wireless

The Invisible Cables of Modern Electronics

In this Chapter:

- The oldest electronic application.
- What are radio waves?
- The electromagnetic frequency spectrum.
- The hardware of radio: transmitters, receivers, and antennas.
- Two-way radio applications.
- Satellites.
- Radio telescopes.

INTRODUCTION

First they called it wireless. Then they called it radio. Today, the term *wireless* is trendy again. Whatever you call it, radio is communication over a distance without any physical connection. Radio is cool because it is completely silent and invisible. It can occur over a distance of only a few feet, all the way around the world, or between planets and spacecraft. Radio, despite the fact that it has been around for over a century, and the fact that we take it for granted on a grand scale, is truly black magic. When you finish reading this chapter, you will know how this magic works.

RADIO COMMUNICATION SYSTEMS

Radio is an electronic communications system. All electronic communications systems have the basic form shown in Figure 7.1, whether it is a wired or wireless system. The information to be communicated, usually voice, is first converted into an electrical signal by a microphone. That signal is then processed by the circuits in a transmitter (TX), and then sent via a communications medium to a receiver (RX) that processes the signal and converts it back into voice by a speaker or headphones. In the telephone system, the communications medium is a cable made of copper wire. In a cable TV system, the medium is a coaxial cable, twisted-pair cable, or fiber optic cable. But in radio, the communications channel is free space. The radio signal travels from one place to another almost instantaneously and it passes right through

© 2010 Elsevier Inc. All rights reserved.
doi: 10.1016/B978-1-85617-700-9.00007-2

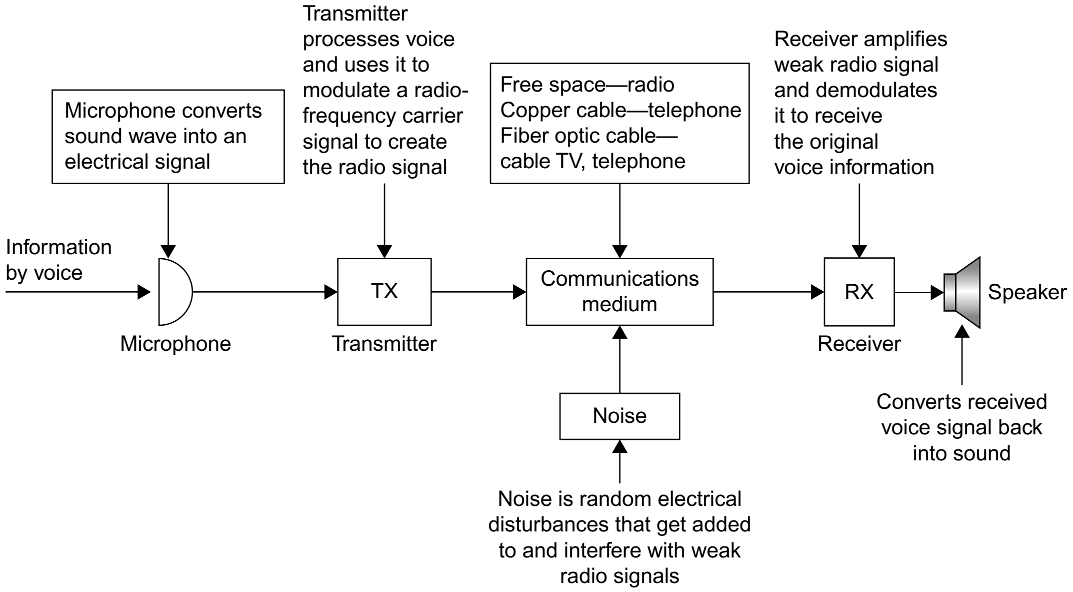
Microphone converts sound wave into an electrical signal
Transmitter processes voice and uses it to modulate a radio-frequency carrier signal to create the radio signal
Free space—radio
Copper cable—telephone
Fiber optic cable—cable TV, telephone
Receiver amplifies weak radio signal and demodulates it to receive the original voice information
Information by voice
Microphone
TX
Transmitter
Communications medium
RX
Receiver
Speaker
Noise
Converts received voice signal back into sound
Noise is random electrical disturbances that get added to and interfere with weak radio signals

FIGURE 7.1 Model of electronic communications system.

most obstructions like walls, buildings, smog, trees, and even people. Note that noise is added to the signal as it passes through the channel or medium.

Noise

Noise is any random electrical voltage that interferes with the weak radio signal. Noise comes from the atmosphere, space, human-made objects, and electronic components. Atmospheric noise is lightning. Space noise comprises random signals from the Sun and remote stars. Human-made noise comes from electrical power generation, distribution, and switching. It also comes from fluorescent lights, motors, auto ignitions, and many other sources. The worst noise comes from the electronic components that actually process the weak radio signal. Resistors, transistors, and integrated circuits generate random noise because of thermal agitation. Heat produces random movement of the electrons in the components, and that produces a noise voltage which is added to the radio signal. Even though noise voltages are very small (microvolts or nanovolts), they are still large enough to interfere with and in some cases obliterate the very small received radio signal. In most radio systems, it is noise that limits intelligibility, transmitted data speed, and transmitted distances.

THREE BASIC WAYS OF COMMUNICATIONS

Radio communications systems work in one of three basic ways, simplex, half duplex, and full duplex (see Figure 7.2).

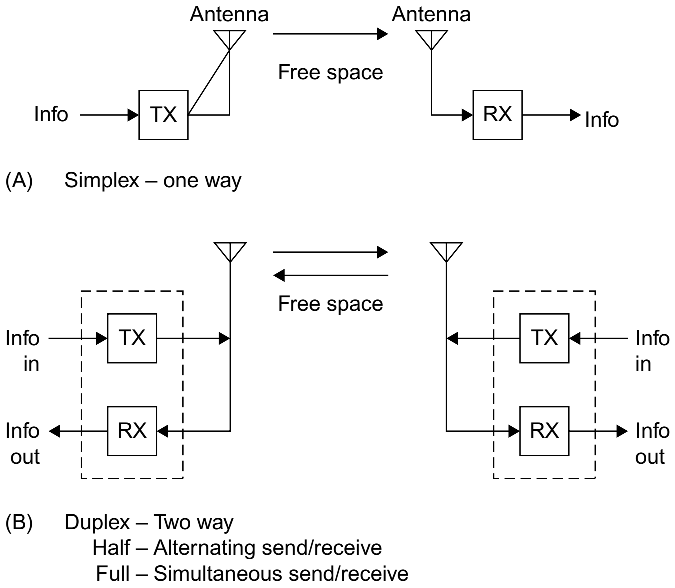
Antenna
Antenna
Free space
Info
TX
RX
Info
(A) Simplex – one way
Free space
Info in
TX
Info out
RX
TX
Info in
RX
Info out
(B) Duplex – Two way
Half – Alternating send/receive
Full – Simultaneous send/receive

FIGURE 7.2 Ways of communicating.

Simplex

Simplex is one-way communications. Examples are radio broadcasting, paging, and remote control. A radio telescope is also one-way: there is a transmitter at one end and a receiver at the other.

Half Duplex

Duplex means two-way, and most radio communications are really two-way affairs. There is a transmitter and a receiver at both ends of the communications system. Both parties can send and receive. And there are two types of duplex systems, half duplex and full duplex.

In half-duplex systems, only one party can transmit at a time. Those communicating with one another take turns talking. This is the kind of radio where you have to say "over" or "come back" after you finish speaking to let the other party know that it is time for her or him to talk. CB radio is like this. Most two-way communications services (police, fire, taxi, aircraft, marine, etc.) use this method because it is cheap and simple. Fax machines are half duplex. Most computer modems are half duplex.

Full Duplex

Full duplex is more complex and expensive. This is two-way radio where both parties can both send and receive simultaneously. Probably the best example of this is a cordless or cellular telephone. It is great to be able to talk and listen at the same time.

WHAT IS A RADIO WAVE ANYWAY?

Any discussion of radio always gets around to this basic question. How do you describe something that you cannot see, feel, or hear? You imagine it.

Electric and Magnetic Fields

A radio wave is an invisible force field that is made up of both a magnetic field and an electric field. In combination, we call them an *electromagnetic field*. We can't see or feel them, but we can experience their effects. In Chapter 2, electric and magnetic fields are described briefly. To review, a magnetic field is a force generated by a magnet, either a permanent magnet or electromagnet. Current flow in a wire or other conductor produces the magnetic field. Invisible magnetic flux lines emanate from the magnet and may influence objects in or near them.

An electric field derives from a voltage, or what we call a *potential*, across two conductors. The attraction of the positive and negative charges across an open space introduces another kind of invisible field that can also influence external objects.

Now if you combine these two types of fields into a self-supporting entity, you would have an electromagnetic field or *radio wave*. We create a radio signal with the circuits in a transmitter, but it is the magic of a device called an *antenna* that produces the combined electric and magnetic field that we refer to as a radio wave.

In the 19th century, a British scientist named James Clerk Maxwell figured out that if you generate a changing electric field, it would, in turn, generate a magnetic field. If the magnetic field is changing and moving forward, it will generate an electric field. The two fields support or regenerate one another as they change and move outward from the antenna that produces them. Maxwell expressed his theories in his famous mathematical equations that form the foundation of radio and light transmission.

Trying to draw a picture of a radio wave is somewhat like trying to draw a picture of the wind. Nevertheless, Figure 7.3A is an attempt. Try to think in three dimensions. Current flow in the antenna wire creates the magnetic field lines. The voltage across the two segments of the antenna produces the electric field lines. The picture is somewhat simplified if we just show the field lines as in Figure 7.3B. The electric and magnetic field lines are always at right angles (90 degrees) to one another. The radio wave moves in a direction at a right angle to both field lines, in this case either into or out of the page. As it gets some distance from the antenna, it becomes self-supporting. That is, the magnetic field sustains the generation of the electric field and vice versa.

Polarization

Polarization of a radio wave means how it is oriented with respect to the surface of the Earth. If the electric field is vertical to the surface of the Earth, the wave is vertically polarized. If the electric field is horizontal to the Earth, the wave is horizontally polarized. It is the orientation of the antenna generating the radio signal that determines the polarization. A vertical antenna produces vertical polarization, and a horizontal antenna produces horizontal polarization. For the best reception, the transmitting and receiving antennas should both be either vertical or horizontal.

There is also circular polarization. Some antennas make the signal rotate clockwise or counterclockwise as the signal travels from transmitter to receiver. Clockwise rotation, as viewed from the transmitter, is called right-hand circular polarization (RHCP). Counterclockwise rotation is called left-hand circular polarization (LHCP).

Signal Speed

Radio waves travel at the speed of light in space. That is 300,000,000 meters per second or 186,410 miles per second. So while radio transmission is not

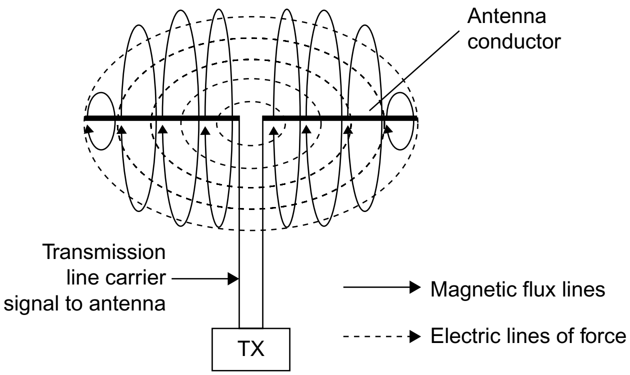
Antenna
conductor
Transmission
line carrier
signal to antenna
Magnetic flux lines
Electric lines of force
TX

(A) Antenna produces both electric and magnetic fields

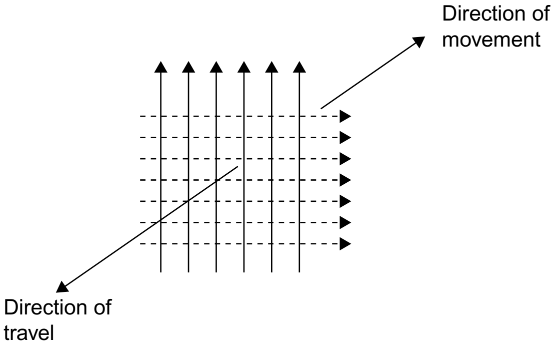
Direction of
movement
Direction of
travel

(B) The electric and magnetic field lines are at a right angle to one another and to the direction of radiation

FIGURE 7.3 How an antenna creates a radio wave.

instantaneous, it is pretty fast. Just as an example, to transmit a signal to the moon that is 250,000 miles away takes 250,000/186,400 = 1.34 seconds.

Signal Strength

As the radio wave gets farther and farther from the antenna, it gets weaker and weaker. The power in a radio wave, in fact, decreases by the square of the distance from the antenna. For example, doubling the distance decreases the received power by two squared (2^2) or four. Even when the transmitter puts many hundreds or thousands of watts of power into the antenna, the signal received at a distance is many times smaller. Typical radio signals are usually expressed in terms of microwatts (1-millionth of a watt) or nanowatts (1-billionth of a watt).

As the radio wave encounters a receiving antenna, the magnetic and electric fields induce a voltage into the antenna. This tiny voltage is then amplified and translated back into the original information signal by the receiver.

ELECTROMAGNETIC FREQUENCY SPECTRUM

The radio signal applied to the antenna by the transmitter is a sine wave. Remember that all sine waves have both an amplitude and a frequency. Frequency is stated in terms of cycles per second or hertz (Hz). Radio transmissions can take place over a very wide frequency range. We call this range the *electromagnetic frequency spectrum*. It extends from approximately 10 kHz to 300 MHz. Figure 7.4 shows a diagram designating the major divisions of the electromagnetic spectrum.

The electromagnetic spectrum is in free space, not only here on Earth but throughout the universe. The spectrum space may be used by anyone and, in fact, on Earth we all share it. The spectrum is divided up and used for many different radio services. It is also partitioned such that various frequency ranges are assigned to or used by different countries for their various radio services. In many cases, portions of the frequency spectrum are shared concurrently among several countries. This allocation of the frequency spectrum is necessary to prevent radio stations from interfering with one another.

In the United States, the frequency spectrum is regulated by the Federal Communications Commission (FCC). Congress established the Communications Act of 1934, which has subsequently been amended to permit the FCC to regulate all forms of radio and wired (telephone, cable TV, etc.) communications services. The National Telecommunications Information Administration (NTIA) manages the spectrum for the government and the military. Throughout the world, the frequency spectrum is allocated and managed by a worldwide organization known as the International Telecommunications Union (ITU).

Major Segments of the Frequency Spectrum

30 kHz to 300 kHz—Low frequencies (LF)
300 kHz to 3 MHz—Medium frequencies (MF)
3 MHz to 30 MHz—High frequencies (HF) or shortwaves
30 MHz to 300 MHz—Very-high frequencies (VHF)
300 MHz to 3 GHz—Ultra-high frequencies (UHF)
3 GHz to 30 GHz—Super-high frequencies (SHF)
30 GHz to 300 GHz—Extremely high frequencies (EHF)

The various areas of the frequency spectrum are assigned to different types of communications services. For example, in the low-frequency (LF) range, most communications services are navigational systems used by the Coast Guard and military. The AM broadcast band (530 kHz to 1710 kHz) is in the

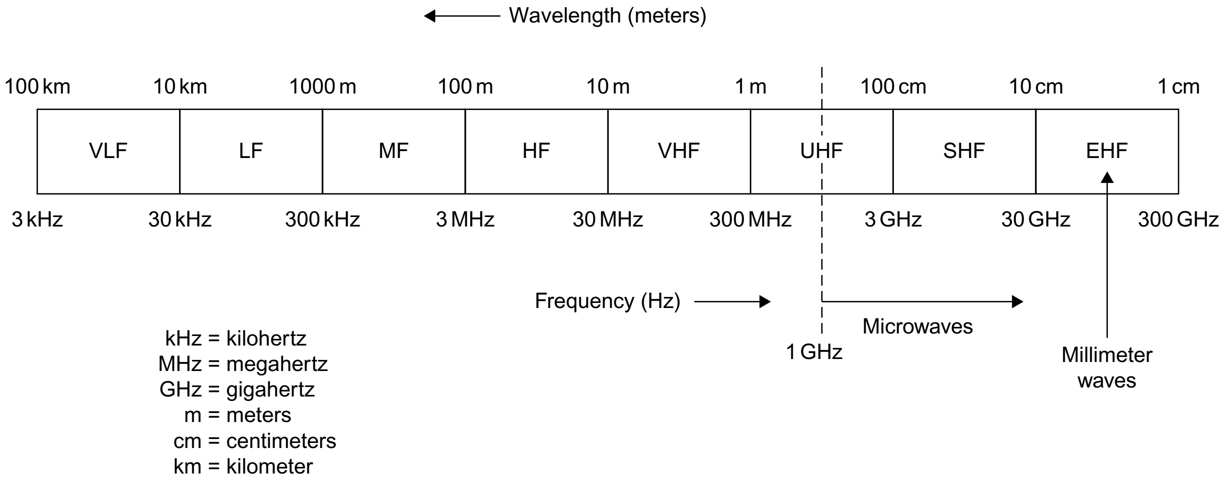
Wavelength (meters)
100 km
10 km
1000 m
100 m
10 m
1 m
100 cm
10 cm
1 cm
VLF
LF
MF
HF
VHF
UHF
SHF
EHF
3 kHz
30 kHz
300 kHz
3 MHz
30 MHz
300 MHz
3 GHz
30 GHz
300 GHz
Frequency (Hz)
Microwaves
1 GHz
Millimeter waves
kHz = kilohertz
MHz = megahertz
GHz = gigahertz
m = meters
cm = centimeters
km = kilometer

FIGURE 7.4 Electromagnetic frequency spectrum.

medium-frequency (MF) range. Shortwave (SW) radio broadcasts, CB radio, amateur radio, and a variety of other communications services are in the 3 to 30 MHz or high-frequency (HF) range.

One of the most widely used areas of the spectrum is the VHF range from 30 to 300 MHz. Virtually all two-way radio communications for marine, aircraft, mobile, and other services are in this range. FM broadcasting (88 to 108 MHz) and TV channels 2 through 7 are also in this region.

The UHF frequency range, from 300 MHz to 3000 MHz (3 GHz), also contains a wide variety of communications services. It includes two-way mobile radio, TV, and cellular telephone.

Frequencies above 1 GHz are designated as microwaves. In the microwave region between 1 and 3 GHz are satellite and navigational systems. For example, the widely used global positioning system (GPS) operates in this range. This area is also occupied by the newer cellular telephone service and wireless networks such as Wi-Fi and Bluetooth.

The frequencies from 30 GHz to 300 GHz are called millimeter waves because their wavelength is in millimeters. These frequencies are typically used only by satellite and radar systems. Other services are point-to-point wireless data services and cellular backhaul. Some short-range wireless devices have also been defined for the 60-GHz range. These will be used in consumer applications as well as in faster wireless networks.

Critical Issue: Spectrum Space

There is only so much radio spectrum space and we must all share it. It is a true natural resource. That's why spectrum space is regulated by governments. But we have just about used it all up. Virtually every hertz of space is spoken for. Diverse radio services compete for space. In some cases, an old service must die before a new one can begin. Sometimes space is stolen from one service to give to another, such as in the mid-1980s when the upper UHF TV channels were sacrificed to make room for cellular telephones. Lately, the U.S. government has been selling space to the highest bidders for new cell phone and other wireless services. During the recent switchover to digital TV in the United States, a huge chunk of the spectrum was freed up (108 MHz, from 698 to 806 MHz) as TV stations abandoned this region. Known as the 700-MHz band, it was auctioned off by the government for $19.6 billion for more cell phone spectrum as well as broadband wireless services and cell phone TV.

The trend has been to push higher and higher in the frequency range to find more space. During the past several decades, the microwave region (above 1 GHz) has been developed. This region, in the UHF, SHF, and EHF ranges, was previously used only for radar and satellites. But now it is being used and developed for digital cell phones, navigation, high-speed networking, broadband (Internet) wireless, and digital TV.

Beyond the millimeter wave range above 300 MHz is the optical spectrum. That is, above 400 GHz, the electromagnetic radiation becomes light, namely infrared, visible light, and ultraviolet light. But that's another story altogether.

Wavelength

While we normally use frequency to refer to location of a radio signal in the spectrum, sometimes you will hear about frequencies or ranges of frequencies designated by wavelength.

Wavelength is the actual physical length of a radio wave. It is the amount of space the wave occupies. Specifically, wavelength is the length of one cycle of the sinusoidal variation of the electromagnetic fields. If we could see the electric and magnetic fields, we could measure the distance between the maximum intensity points in meters (1 meter = 3.28 feet or 39.37 inches). This is the wavelength. Wavelength is also the distance that the radio wave travels in the time of one cycle.

You can determine the wavelength of a radio signal with the following simple formula:

$$\lambda = 300{,}000{,}000/f$$

Wavelength is represented by the Greek letter lambda (λ) and frequency in hertz is designated by the letter "f." The value 300,000,000 is the speed of light of a radio wave in meters per second in free space. (This translates to a speed of 186,410 miles per second.)

A radio signal with a frequency of 150 MHz is expressed in hertz as 150,000,000 Hz. Therefore,

$$\lambda = 300{,}000{,}000/150{,}000{,}000 = 2 \text{ meters}$$

If we give the frequency in megahertz, the formula can be shortened to:

$$\lambda = 300/f$$

For example, we can calculate the wavelength of the shortwave or high-frequency part of the spectrum (3 to 30 MHz) as follows:

$$\lambda = 300/3 = 100 \text{ meters}$$
$$\lambda = 300/30 = 10 \text{ meters}$$

If you rearrange the formula, you can also find the frequency if you know the wavelength:

$$f = 300/\lambda$$

What is the frequency of the 25-meter band?

$$f = 300/25 = 12 \text{ MHz}$$

RADIO WAVE PROPAGATION

Propagation refers to the way in which a radio signal travels from the transmitting antenna to the receiving antenna. The way the signal behaves is directly related to its frequency.

Low and Medium Frequencies

For example, in the low-frequency and medium-frequency ranges, the radio signals tend to hug the Earth as they move from transmitter to receiver. These ground wave signals can only travel relatively short distances because they are absorbed and attenuated by the Earth, trees, buildings, and other obstructions. Depending on the frequency of operation, transmitter power, and the nature of the terrain, distances are limited to several hundred miles. Distances can be extended with greater power or if the signal travels over water.

High Frequencies

In the HF or shortwave frequency range, the ground waves are extremely weak and travel only short distances, usually several miles. Instead, propagation is by way of refraction of the ionosphere. The ionosphere is an electrified area above the Earth located from approximately 30 miles to 250 miles above the Earth. The Sun ionizes the gases in the air, thereby giving them electrical characteristics. The closer the ionosphere to the Sun, the greater the degree the ionization. The ionosphere is actually several different layers or ionization levels as shown in Figure 7.5. The upper layers have greater ionization than the lower layers. Obviously, the degree of ionization is determined by whether it is day or night. During the daylight hours, the ionosphere is highly charged electrically. At night, the ion layers essentially disappear.

A radio wave in the 3- to 30-MHz range is transmitted up to the ionosphere where it is refracted or bent by ionization. The degree of bending depends on the frequency of operation and how high in the ionosphere the radio wave penetrates. In most cases, the radio wave is bent so much that it actually returns to Earth. The effect appears to be similar to that of a light beam being reflected from a mirror.

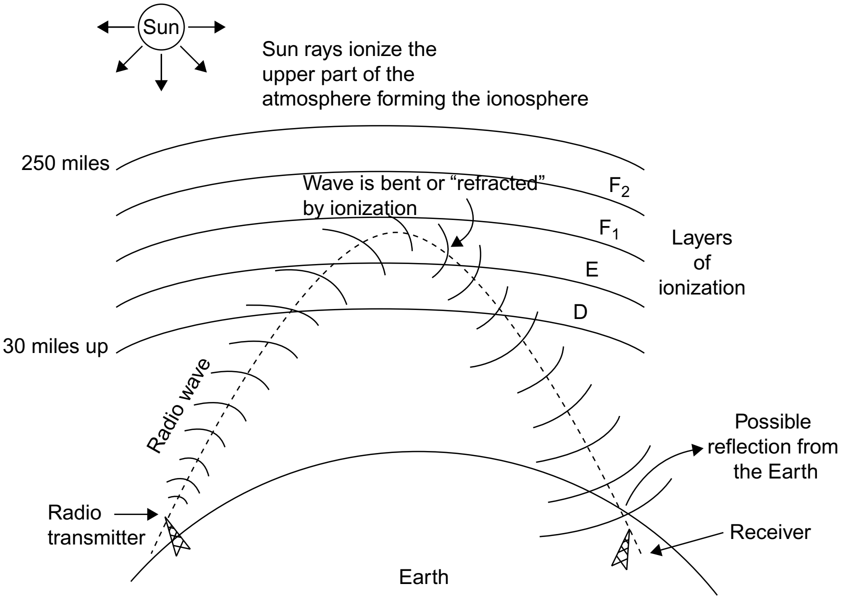
Sun
Sun rays ionize the upper part of the atmosphere forming the ionosphere
250 miles
Wave is bent or "refracted" by ionization
F_2
F_1
E
D
Layers of ionization
30 miles up
Radio wave
Possible reflection from the Earth
Radio transmitter
Receiver
Earth

FIGURE 7.5 Radio propagation of shortwaves.

As you can see, the radio wave actually returns to the Earth at some distance from the transmitter. Receiving stations in the area of the refracted signal can receive it. In some cases, if the signal is strong enough and the characteristics of the Earth are right, the radio signal may actually be reflected by the Earth back up to the ionosphere where it is again bent and returned to Earth. Under the right conditions, the radio signals can make several hops, thereby greatly extending the distance of communications. With the optimum frequency, power level, and time of day, communications halfway around the world can easily be achieved. That is why it is easy for you to listen to worldwide radio broadcasts that occur in the shortwave frequency bands.

VHF, UHF, and Microwaves

Radio wave propagation in the frequency range above 30 MHz occurs essentially by direct antenna-to-antenna space waves. This is usually called *line-of-sight (LOS) communications*. Further, the higher the frequency of the radio wave, the less it is affected by the ionosphere. UHF and microwave signals penetrate the ionosphere, making satellite and long-distance spacecraft communication possible.

LOS communications means that the transmitting antenna must be able to "see" the receiving antenna. If the distance of the receiving antenna is beyond the curvature of the Earth, the transmitted signal will simply pass over the antenna. To receive the signal, the receiving antenna or transmitting antenna can be increased in height, thereby extending the total range of communications. Figure 7.6 shows this idea. The simple formula along with it allows you to calculate the maximum LOS communications if you know the height of the receiving and transmitting antennas. That's why at VHF, UHF, and microwave

$$D = \sqrt{2H_t} + \sqrt{2H_r}$$

$D =$ Line-of-sight distance (miles)
$H_t =$ Height of transmitting antenna
$H_r =$ Height of receiving antenna

If $H_t = 200$ ft and $H_r = 40$ ft,

$$D = \sqrt{2(200)} + \sqrt{2(40)} = \sqrt{400} + \sqrt{80}$$

$$D = 20 + 8.94 = 28.94 \text{ miles}$$

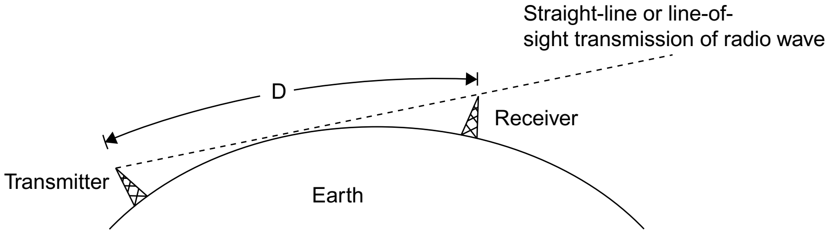

FIGURE 7.6 Line-of-sight (LOS) communications at VHF, UHF, and microwave.

frequencies, the height of the antenna is the most critical determining factor in overall communications distance.

MODULATION

A radio wave is a sine wave of voltage that has been converted into electric and magnetic fields by the antenna. The transmitter generates the sine wave and amplifies it to a high power level. The question is, how does the information to be transmitted fit into this?

The information or intelligence to be conveyed is usually a voice signal produced by a microphone. Other forms of intelligence signals include pictures or video and digital (binary) data from a computer. The purpose of the high-frequency sine wave radio signal is to give the intelligence signal a ride. That's why the radio signal generated by the transmitter is called the carrier. Modulation is the process by which the information is impressed on or embedded in the carrier. Modulation causes the carrier to be modified by the information signal.

There are three basic forms of modulation used to put information on a carrier: *amplitude modulation* (AM), *frequency modulation* (FM), and *phase modulation* (PM).

Amplitude Modulation

In amplitude modulation, the voltage or power level of the information signal changes the amplitude of the carrier in proportion (see Figure 7.7). With no modulation, the AM carrier is transmitted by itself. When the modulating information signal (a sine wave) is applied, the carrier amplitude rises and falls in accordance. The carrier frequency remains constant during amplitude modulation.

Amplitude modulation is widely used in radio. AM broadcast stations are, of course, amplitude modulated. So are citizens band radios, aircraft radios, and the video modulation of a TV broadcast transmitter. A special form of amplitude modulation, known as quadrature modulation (QAM), is also widely used in modems to transmit digital data over cable or wireless.

Sidebands

The modulation process causes new signals to be generated. These new sine wave signals are called sidebands. Their frequencies are the sum and difference of the carrier and modulating signal frequencies. For example, in an AM radio station, audio frequencies as high as 5 kHz can be transmitted. If a 5-kHz sine wave tone is to be transmitted, the modulation process causes sidebands 5 kHz below and 5 kHz above the carrier to be produced. This is illustrated in Figure 7.8. For an AM radio station with a carrier frequency of 860 kHz, the

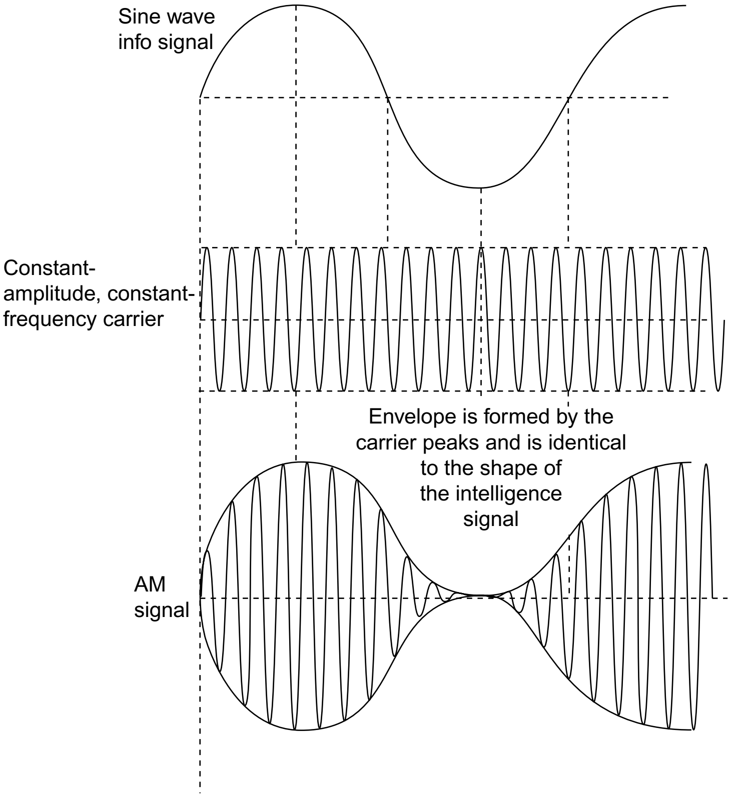
Sine wave
info signal
Constant-
amplitude, constant-
frequency carrier
Envelope is formed by the
carrier peaks and is identical
to the shape of
the intelligence
signal
AM
signal

FIGURE 7.7 Amplitude modulation.

lower sideband (LSB) would occur at 860 − 5 = 855 kHz, while the upper sideband (USB) occurs at 860 + 5 = 865 kHz. The carrier and the sidebands combined produce the composite waveform shown at the bottom of Figure 7.7.

Bandwidth

This brings up the very important concept of bandwidth. Bandwidth refers to a range of frequencies over which a radio signal operates. This is also referred to as a radio channel. As you can see from Figure 7.8, the AM signal consists of the carrier and the sidebands. Together these signals occupy a bandwidth of 10 kHz. You can figure out the bandwidth of any signal by simply subtracting the lower sideband frequency from the upper sideband frequency.

$$\text{BW} = \text{USB} - \text{LSB} = 865 - 855 = 10\ \text{kHz}$$

All of the receiver circuits must be set to pass signals in this 10-kHz range in order to avoid distortion of the signal or lost information.

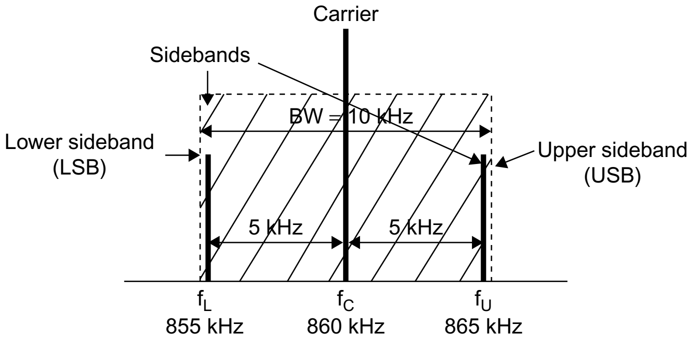

f_M = Frequency of modulating signal (5 kHz)
f_C = Carrier frequency (860 kHz)
f_L = Lower sideband frequency
f_U = Upper sideband frequency

$$f_L = f_C - f_M = 860 - 5 = 855 \text{ kHz}$$
$$f_U = f_C + f_M = 860 + 5 = 865 \text{ kHz}$$

$$BW = f_U - f_L = 865 - 855 = 10 \text{ kHz}$$

FIGURE 7.8 Sidebands and bandwidth.

Frequency Modulation

In frequency modulation, the carrier amplitude remains constant but its frequency is changed in accordance with the modulating signal. Specifically, the higher the amplitude of the information signal, the greater the frequency change. The actual carrier frequency deviates above and below the center carrier frequency as the information signal amplitude varies. Figure 7.9 shows frequency modulation with a sine wave information signal. Note that the carrier frequency gets higher on the positive peaks and lower on the negative peaks of the information signal.

Like AM, FM also produces sidebands. But unlike AM, which produces a single pair of sidebands for each frequency in the modulating signal, the frequency-modulation process produces an infinite number of pairs of sidebands for each frequency in the information signal. As a result, the bandwidth occupied by an FM signal is enormous. Luckily, the number of sidebands produced can be controlled by properly selecting the amount of deviation permitted in the carrier. Small deviations result in fewer sidebands. Further, some of the higher-order sidebands are extremely low in amplitude and, therefore, contribute little to the FM signal. But while the bandwidth of an FM signal can be controlled and established to fit a desired frequency range, it does nevertheless usually take more room in the spectrum than an AM signal.

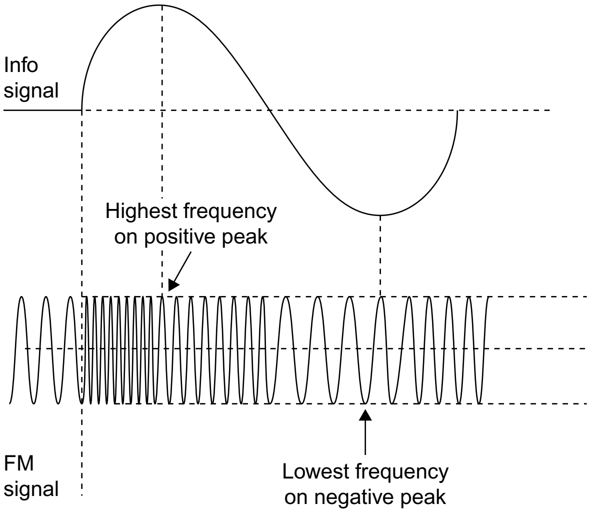

FIGURE 7.9 Frequency modulation.

The primary benefit of FM is that it is less sensitive to noise, which consists of undesirable amplitude variations that get involuntarily added to a signal. Noise is easily eliminated in an FM system where a constant carrier amplitude is used. Some of the most common applications of FM include FM radio broadcasting, and two-way mobile and marine radios.

Phase Modulation

The third type of modulation is phase modulation. Remember that phase shift is a time shift between two sine waves of the same frequency. We can use the information signal to shift the phase of the carrier with respect to the carrier reference. The result of this, however, is a signal that looks virtually the same as an FM signal (Figure 7.9). Phase modulation is usually a little easier to implement electronically than FM, so most so-called FM systems typically use phase modulation instead. Both FM and PM are often referred as types of *angle modulation*.

Digital Modulation

All of the three basic types of modulation, AM, FM, and PM, can also be used to transmit digital or binary data. However, they are usually given different names. Amplitude modulation of a carrier by a binary signal is usually referred to as amplitude shift keying (ASK). This is illustrated in Figure 7.10A. The binary signal simply shifts the carrier amplitude between two specific levels. A special form of ASK is called on–off keying (OOK). This is illustrated in Figure 7.10B. Here the binary signal simply turns the carrier on for a binary 1 and off for a binary 0.

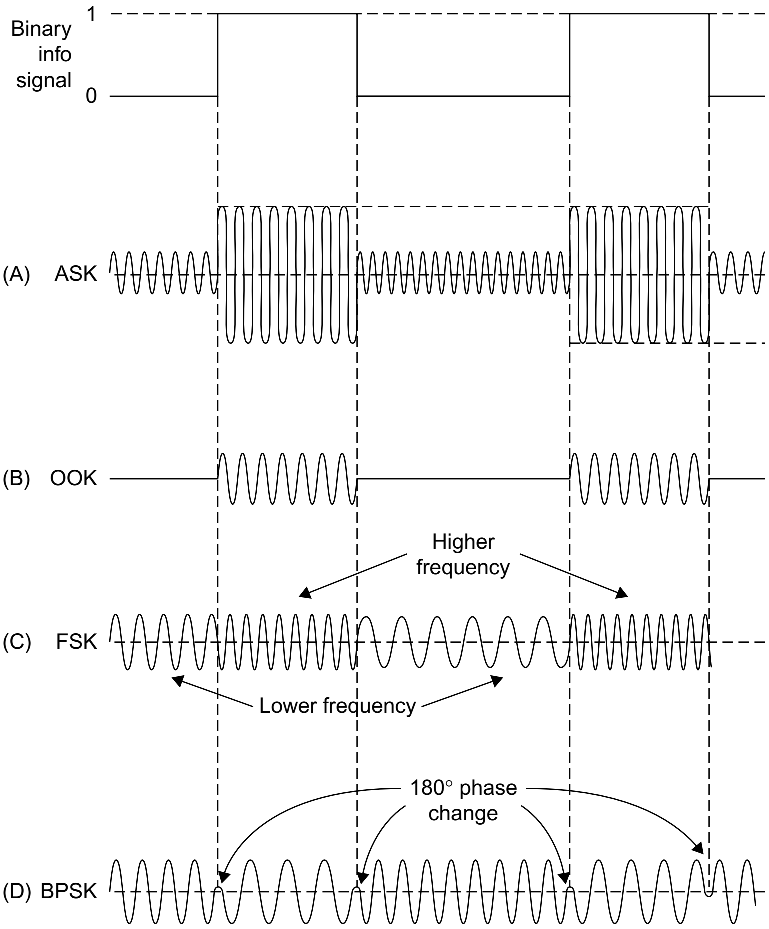

FIGURE 7.10 Types of digital modulation. **(A)** ASK. **(B)** OOK. **(C)** FSK. **(D)** BPSK.

Frequency modulation of a carrier by a binary signal is called frequency shift keying (FSK). Here the binary signal shifts the carrier between two discrete frequencies (see Figure 7.10C).

Phase modulation of a carrier by a binary signal is referred to as phase shift keying (PSK). The term binary PSK (BPSK) is also used. The phase of the carrier is changed as the binary signal switches from 0 to 1 or 1 to 0 (see Figure 7.10D). Phase shift is 180 degrees, which is easily detected at the receiver.

One of the most widely used forms of digital modulation is QAM, which is really a composite of both AM and PSK. The binary signal is converted into a multilevel digital signal that simultaneously modifies both the amplitude and the phase of the signal. QAM is used in digital data transmission in satellites and in modems sending digital data in audio form over telephone lines or cable

TV systems. It is also widely used in the newer digital wireless networks. See Chapter 8 for more detail.

Because binary signals contain many higher-frequency harmonics, the resulting signals produced by modulation have an enormous bandwidth. Many schemes have been developed to limit the spectrum of the modulating signal and, as a result, to reduce the amount of bandwidth occupied by a binary modulated radio signal.

Spread Spectrum

Spread spectrum (SS) is unusual in that it is not only a form of modulation but also a system for encrypting the information signal for security purposes and for multiplexing multiple signals over a range of frequencies.

Spread spectrum was originally developed during World War II as a way to prevent the enemy from interpreting radio transmissions. For years the government and military protected the concepts of SS modulation. However, today it is widely used in radio and telephone communications. Most of the newer cellular telephones are based on this concept, specifically code division multiple access (CDMA) cell phones.

When using SS, the analog information signal is converted into a digital signal before transmission. The digital voice signal is then used to modify the carrier in a special way. The two most common ways of modifying the carrier are referred to as *direct-sequence* (DS) and *frequency-hopping* (FH) *spread spectrum*. In direct-sequence SS, the serial digital signal representing the voice is chopped up into random sequences and then used to PSK the carrier. In frequency-hopping SS, the carrier is rapidly switched from one frequency to another for a short dwell time. The result of SS modulation is that the signal itself is broken up at random, and spectral energy is spread over a very wide frequency range. Because of the random sequences of the coding and frequency hopping, such a radio signal is nearly impossible to put back together at a receiver without knowing the code. SS receivers designed to pick up the SS signal know the code, and can unscramble the signal and reassemble it into the original modulating information signal.

RADIO HARDWARE

All radio communications systems consist of three basic pieces of hardware: *transmitter*, *receiver*, and *antenna*. In some systems, the transmitter and receiver are packaged and used separately, while in others they are combined into a single package and called a *transceiver*.

Transmitters

The radio signal is produced by the transmitter. A general block diagram that applies to virtually every transmitter is shown in Figure 7.11. The radio signal

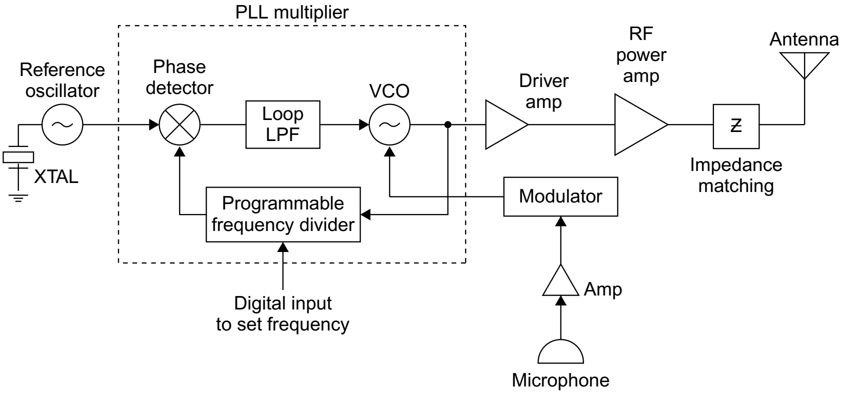

FIGURE 7.11 Radio transmitter.

is generated at a low power level by a frequency synthesizer. This uses a phase-locked loop (PLL) with a crystal (XTAL) oscillator reference. A variable-frequency divider is used to set the frequency of operation. The FCC is really fussy about the frequency at which transmitters work and, therefore, the frequency must be precisely set and extremely stable. A quartz crystal reference oscillator is normally used to achieve this objective.

The low-level signal produced by the voltage-controlled oscillator (VCO) is then amplified by one or more driver amplifiers to boost the power level. Next, the microphone signal is fed to the modulator. The modulator changes the frequency or phase of the RF signal in accordance with the voice, digital data, or other information to be transmitted. That signal is applied to the PLL VCO.

The modulated signal is sent to additional RF power amplifiers, usually referred to as *drivers*. Each succeeding stage boosts the signal power by a specific amount. A final RF power amplifier produces the desired amount of output power. From there, impedance (Z) matching circuits are used to match the high-power amplifier to the antenna for maximum power radiation.

The physical realization of the transmitter can vary all the way from a single integrated circuit chip the size of a postage stamp to something that can fill up a moderately large building. The size depends on the operating frequency and the amount of output power desired. Very-low-power, high-frequency radio transmitters for garage door openers and remote keyless entry systems on cars are fully self-contained within a single integrated circuit. High-power radio broadcast transmitters at low frequencies may fill a small room. The higher the power and the lower the frequency, the larger the transmitter.

Receivers

The job of the receiver is to translate the weak radio signal picked up by the antenna into an output that fully recovers the originally transmitted information.

By the time the radio signal reaches the receiving antenna, it is extremely weak and, therefore, the primary job of the receiver is to boost the signal level. Therefore, the receiver consists mainly of a series of amplifiers that increase the level of the small signal to a point where the original information can be recovered.

The amount of amplification provided by a receiver determines its *sensitivity*. Another job of the receiver is to provide selectivity so that the desired signal can be picked out of the thousands of signals being transmitted simultaneously by other transmitters. The receiver's job is to select just the desired one and to reject all of the others. This is what we refer to as the *selectivity* of the receiver.

A third function of the receiver is to recover the original information. A circuit called a *demodulator* removes the voice, video, or digital data from the carrier. Finally, output amplifiers reproduce the signal as desired, such as a voice signal in a speaker.

Figure 7.12 is a block diagram of a superheterodyne receiver. The incoming signal is usually amplified by an *RF input amplifier* often called a *low-noise amplifier* (LNA) and then fed to a circuit called a *mixer*. The mixer is a frequency-translation circuit that converts the incoming signal, regardless of its frequency, along with its modulation to a lower intermediate frequency (IF). This allows the receiver to use a fixed-frequency amplifier system where most of the gain and selectivity of the receiver is achieved. The mixer takes the input signal and mixes it with a high-frequency sine wave signal generated internally in the receiver by what is called a local oscillator (LO). The LO is usually a frequency synthesizer. The result of the mixing is output signals that are the sum and difference frequencies of the input and oscillator signals. The IF amplifier is normally tuned to select only the difference frequency. Tuning is accomplished by changing the local oscillator frequency. The local oscillator in most channelized receivers is some type of PLL frequency synthesizer.

The IF amplifier boosts the signal level considerably. Its selectivity ensures that signals whose frequencies are above or below the desired signal are effectively eliminated. The IF amplifier also has a built-in feedback control circuit that automatically adjusts its gain to the level of the incoming signal. This feature is called *automatic gain control* (AGC). AGC permits the receiver to pick up both strong local signals and very weak distant signals. When very strong signals are received, the AGC automatically reduces the gain of the IF amplifier to minimize circuit overload and signal distortion. When very weak signals are received, the AGC adjusts the gain of the IF amplifier for higher amplification.

The *demodulator circuit*, sometimes called the *detector*, follows the IF amplifier. It removes the carrier and recovers the desired information signal. From there, additional processing takes place to either boost the signal level or to shape it before it appears at the output.

Most receivers that we are familiar with have a speaker as the primary output. But there are exceptions, such as the picture tube in a TV set or the motor in a garage door opener. The receiver may also put out digital data that goes straight to a computer.

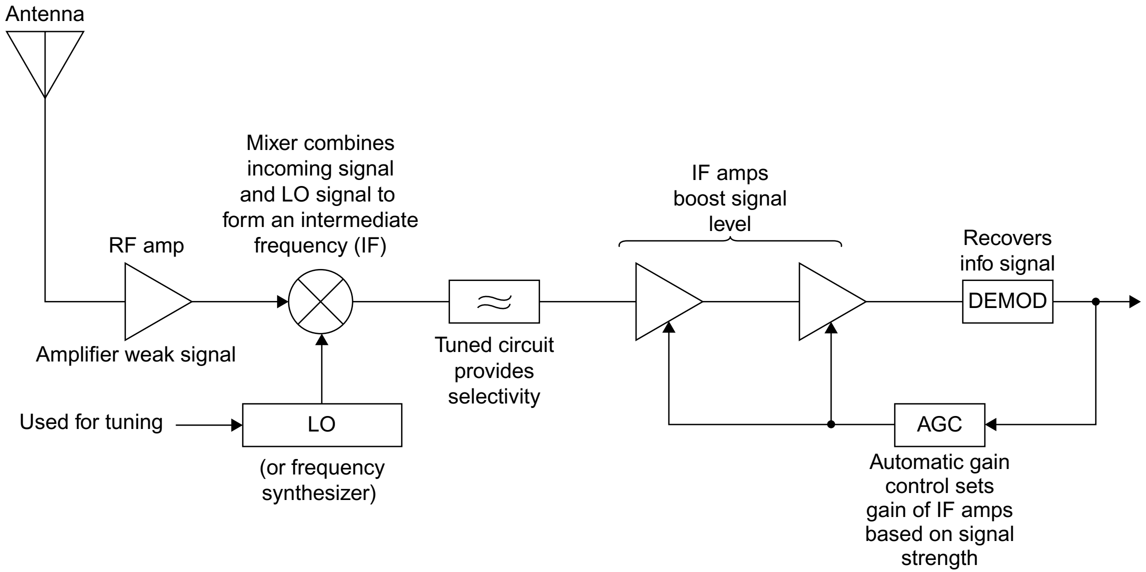
Antenna
RF amp
Mixer combines incoming signal and LO signal to form an intermediate frequency (IF)
IF amps boost signal level
Recovers info signal
DEMOD
Amplifier weak signal
Tuned circuit provides selectivity
Used for tuning
LO
(or frequency synthesizer)
AGC
Automatic gain control sets gain of IF amps based on signal strength

FIGURE 7.12 Superheterodyne receiver. LO, local oscillator.

Transceivers

Most communications systems use both a transmitter and a receiver. In two-way radio, both send and receive capability is needed at both ends of the communications channel. Therefore, the receiver and transmitter are packaged together in the same housing to create a transceiver. Small portable transceivers are called hand-helds, walkie-talkies, or cellular telephones.

The transmitters and receivers are typically the same as those described earlier. However, in some cases, the transmitter and receiver may share circuits. They virtually always share the antenna and a common power supply like a battery. In some cases, a single-frequency synthesizer may generate both the receiver LO signal as well as the transmitter carrier.

ISM Band Transceiver

The ISM band refers to specific frequencies set aside for wireless industrial, scientific, and medical (ISM) applications. The most common frequencies in the United States are 315, 433, and 915 MHz. The frequency 868 MHz is widely used in Europe. These unlicensed frequencies are used for simple monitor and control applications using low-speed digital data over short ranges. Some common examples are garage door openers, remote keyless entry on cars, remote thermometers, tire pressure gauges, security and fire alarms, home automation, and any telemetry or remote control application. The radios used for this are typically single ICs to which an antenna, a battery, a housing, and the input data are added. The devices are available as a receiver, a transmitter, or a complete transceiver. Figure 7.13 is a block diagram of a typical transceiver.

The transmitter section uses an external crystal to serve as the reference to an internal PLL frequency synthesizer to set the frequency. The PLL actually multiplies the crystal frequency by some N value to get the desired output. If the crystal frequency is 6.77 MHz and the divide ratio is 64, the output would be 433.28 MHz. The serial digital data input coming from a sensor, an analog-to-digital converter (ADC), or a computer frequency modulates the crystal oscillator to create the FSK signal. A power amplifier (PA) after the transmit VCO boosts the power up into the 1- to 100-milliwatt range. The output goes to the antenna. A transmit/receive switch allows the antenna to be shared in a half-duplex system.

The receiver is a special version of a superheterodyne. The input signal from the antenna is boosted in amplitude by the LNA and applied to the mixers. The local oscillator is set to the incoming signal frequency (also the transmit frequency), so the mixers generate a difference of zero. This is called *direct conversion* or *zero IF*. Two mixers are used and the local oscillator is shifted 90 degrees in phase between the two. This gives two zero IF outputs that are filtered. Zero IF does not mean a nonexistent output signal. It just means that the FSK signal is recovered directly at its data rate. The I and Q signals are 90 degrees out

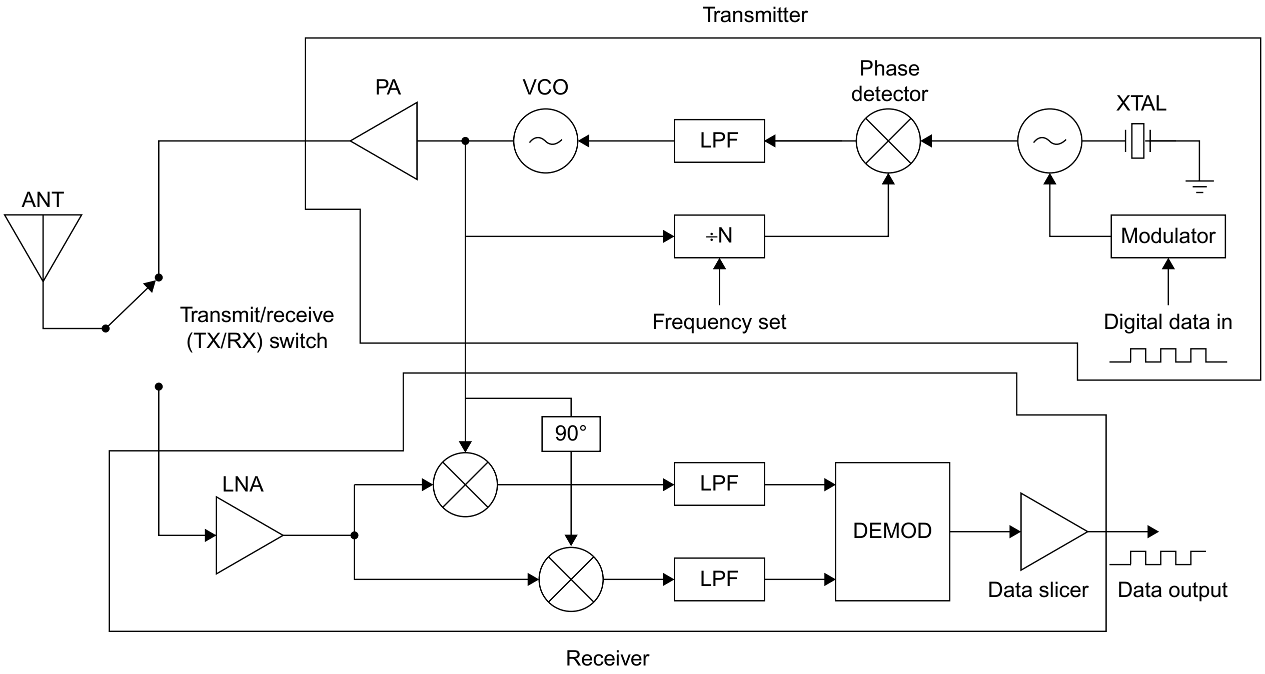
Transmitter
PA
VCO
Phase detector
XTAL
LPF
÷N
Modulator
ANT
Transmit/receive (TX/RX) switch
Frequency set
Digital data in
90°
LNA
LPF
LPF
DEMOD
Data slicer
Data output
Receiver

FIGURE 7.13 Single-chip UHF transceiver for low-speed, short-range digital data transmissions.

of phase and this allows the digital data to be more easily demodulated. A data slicer, a circuit like a comparator, is used to shape the signal into a clean binary bit stream. With this kind of radio, the range is limited. It may be only a few feet or up to a mile or so depending on the antenna, environment, or application.

ANTENNAS

The antenna takes the voltage and current produced by the transmitter and converts it into the electromagnetic wave we know as a radio wave. At the receiver, the antenna converts the electromagnetic field back into a voltage we know as the radio signal. The radio waves pass over the receiving antenna, inducing a voltage that causes current to flow in the receiver.

Antennas can have an amazing variety of shapes. Essentially an antenna or aerial is some arrangement of electrical conductors whose size and length are critical. For optimum operation, the antenna length must be related to the wavelength (λ) of the signal. Common antenna sizes are one-quarter ($\lambda/4$) and one-half ($\lambda/2$) wavelength at the operating frequency. The higher the frequency, the shorter the wavelength and thus the smaller the antenna. For example, one wavelength at 150 MHz, a popular two-way radio frequency range measures 2 meters. A half wavelength is 1 meter and a quarter wavelength is 0.5 meter. Since 1 meter is 39.37 inches, a half wavelength is 19.68 inches. At 900 MHz, the frequency range of cell phones, one wavelength is 0.3333 meter. A quarter wavelength antenna would be $0.3333/4 = 0.08333$ meter or 3.28 inches.

Ground-Plane Antenna

Antennas can be anywhere from a long horizontal copper wire to a vertical steel or aluminum tower several hundred feet tall. Or it may be an array of short conductors like those in a TV antenna. One of the most common antennas is a short quarter wavelength vertical that is widely used in mobile radios, cell phones, and other hand-held units (see Figure 7.14). The quarter-wave or ground-plane antenna as it is sometimes called, radiates equally well in all directions, making it omnidirectional. We say that its radiation or reception pattern is a circle.

Dipole

Another common antenna type is the dipole, which is one-half wavelength long and usually mounted horizontally. Figure 7.15 shows an example. Its radiation or reception pattern is a figure 8. The greatest radiation or strongest reception occurs at a right angle to the dipole element itself.

Yagi

A variation of the dipole is the Yagi, a multielement antenna based on the dipole shown in Figure 7.16. To the dipole is added a reflector element that is

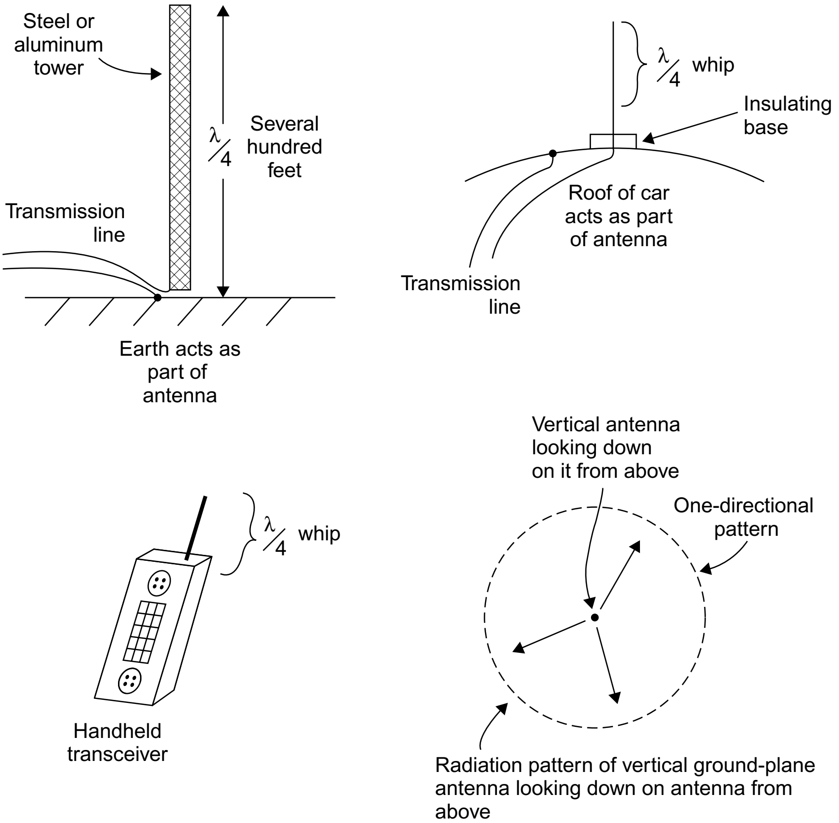
Steel or aluminum tower
λ/4
Several hundred feet
Transmission line
Earth acts as part of antenna
λ/4 whip
Insulating base
Roof of car acts as part of antenna
Transmission line
λ/4 whip
Handheld transceiver
Vertical antenna looking down on it from above
One-directional pattern
Radiation pattern of vertical ground-plane antenna looking down on antenna from above

FIGURE 7.14 Quarter-wave ground-plane antenna.

slightly longer than one-half wavelength, and one or more director elements that are progressively shorter than the dipole. The directors and reflector make the Yagi highly directional. The extra elements, called *parasitic elements*, concentrate and aim the beam of radiation over a narrow range. At the transmitter, this has the effect of increasing the transmitted power, making the signal stronger at the receiver as if the transmitter power had been increased. A similar gain effect occurs at the receiver.

Other Antennas

There are many other antenna variations. Antennas formed with copper patterns on a printed circuit board (PCB) are widely used. One example is the *patch antenna* (see Figure 7.17). It is simply a square area of copper about one-half wavelength on a side made on a PCB. Another is the *loop antenna* that is just a loop of copper on the PCB. A popular cell phone antenna is the

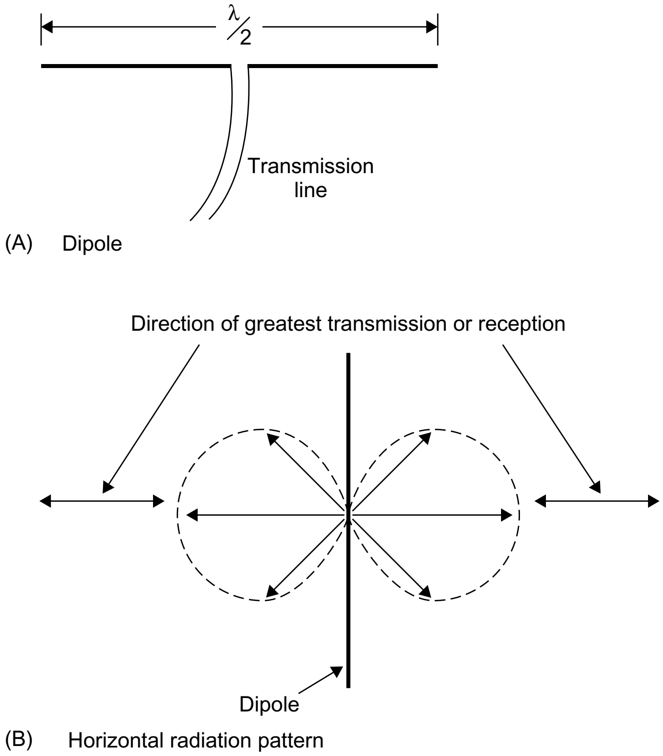
λ/2
Transmission line
(A) Dipole
Direction of greatest transmission or reception
Dipole
(B) Horizontal radiation pattern

FIGURE 7.15 Dipole antenna.

inverted F, which is an F-shaped pattern on the PCB. Antennas also come as components. For microwave frequencies, there are special antennas made on a ceramic base that are small and can be soldered to a PCB.

Transmission Lines

The antenna is usually connected to the transmitter or receiver by way of a coaxial cable (see Figure 7.18). It consists of a copper center conductor surrounded by an insulator, such as Teflon. Around that is a shielded copper braid or foil. The whole thing is covered with a plastic insulator. Coaxial cable is the most commonly used transmission line to connect the equipment to the antenna. It provides an efficient method of transferring the voltages and currents to and from the antenna while providing a certain amount of shielding from noise. For PCB antennas, the transmission line is just a short pair of copper lines from the transceiver IC to the antenna.

Microwave Antennas and Waveguides

Antennas used at microwave frequencies are quite different from those used at the lower frequencies. Perhaps the most common is the *horn antenna*. It looks

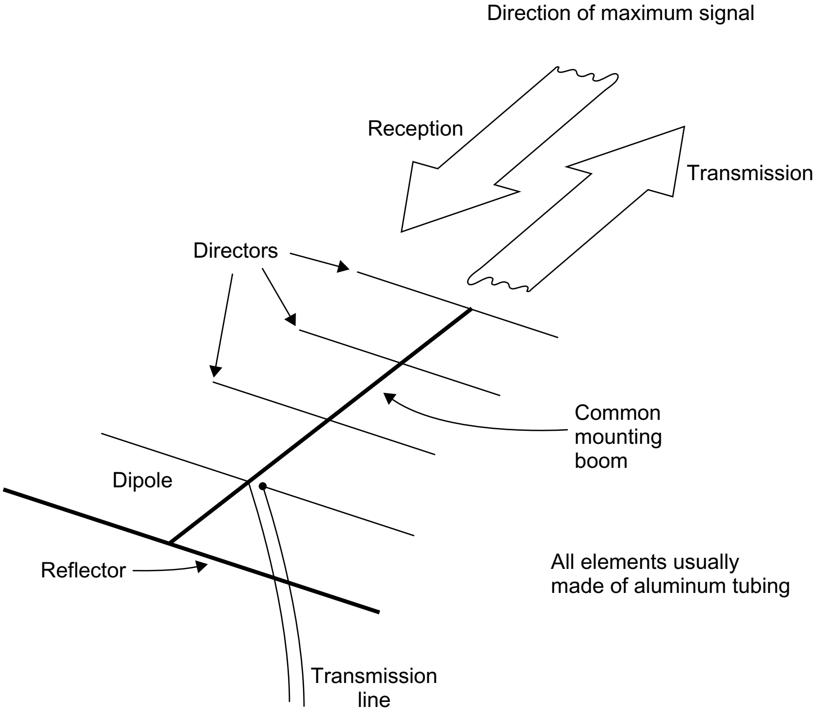
Direction of maximum signal
Reception
Transmission
Directors
Common mounting boom
Dipole
Reflector
All elements usually made of aluminum tubing
Transmission line

(A) Yagi

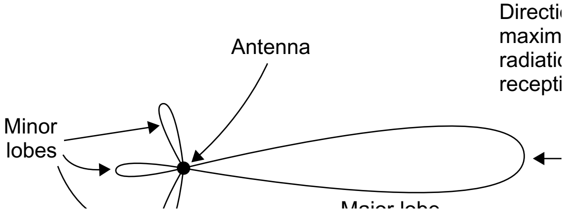
Directi
maxim
radiati
recepti
Antenna
Minor lobes
Major lobe

FIGURE 7.16 Yagi antenna.

like the flared end of a musical horn, only most of them have an opening or aperture that is rectangular rather than round.

As seen in Figure 7.19, the horn is directional and has gain. The microwave energy is fed to the horn by a transmission line. Coax can be used for the lower microwave frequencies, but at the higher microwave frequencies, coaxial cable loss is too high for long runs. A high percentage of the signal power is lost in the coax cable itself. This problem is overcome by using a waveguide. A waveguide is nothing more than a hollow copper pipe, usually with a rectangular cross-section through which the microwave energy passes from transmitter to antenna or from antenna to receiver.

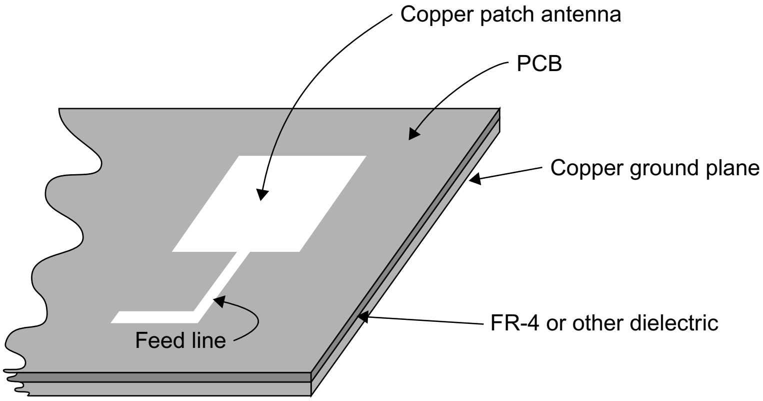
Copper patch antenna
PCB
Copper ground plane
FR-4 or other dielectric
Feed line

FIGURE 7.17 Printed circuit board (PCB) patch antenna.

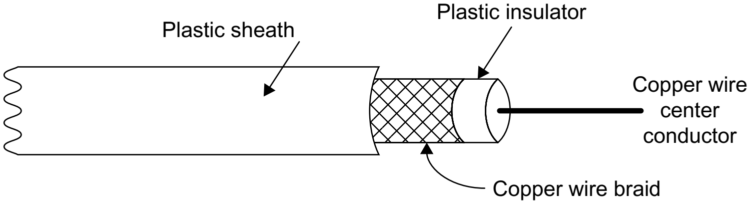
Plastic insulator
Plastic sheath
Copper wire center conductor
Copper wire braid

FIGURE 7.18 Coaxial cable transmission line.

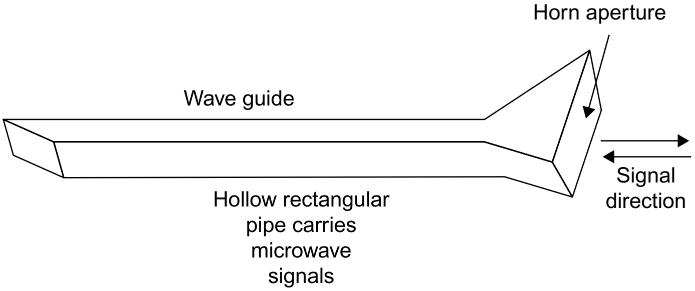
Horn aperture
Wave guide
Signal direction
Hollow rectangular pipe carries microwave signals

FIGURE 7.19 Microwave horn antenna fed by waveguide transmission line.

The most effective and widely used microwave antenna is the parabolic dish (see Figure 7.20). This is usually a horn antenna combined with a reflector shaped like a dish. Its shape is a mathematical parabola. Signals transmitted by the horn are reflected forward by the parabolic dish shaping the radio waves into a very narrow beam. Received energy is reflected by the dish into the horn antenna. The highly directional nature of the dish gives it a very

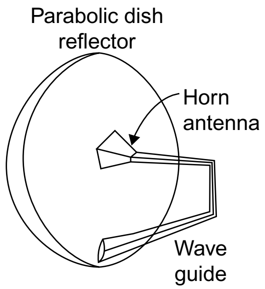
Parabolic dish reflector
Horn antenna
Wave guide

(A) Parabolic dish with horn

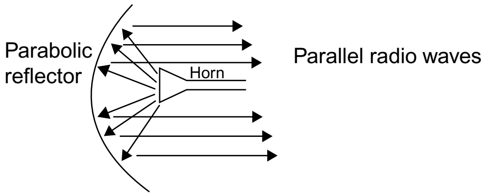
Parabolic reflector
Horn
Parallel radio waves

(B) Transmission

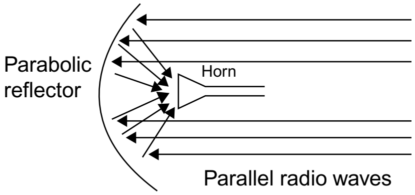
Parabolic reflector
Horn
Parallel radio waves

(C) Reception

FIGURE 7.20 How a parabolic dish works.

narrow beamwidth and enormous gain. These are widely used in satellite TV and radar.

TWO-WAY RADIO

Two-way radio communications between individuals are by far the most widespread use of radio. And there are many forms of it. You have seen such radio communications as they take place and you have no doubt engaged in some form of it yourself. In any case, this section takes a look at some of the diverse uses of such two-way communications.

Most two-way radios operate in the VHF and UHF bands. The simplest two-way radios operate on multiple frequencies usually selected by a frequency synthesizer. Virtually all two-way radio communications are half duplex, meaning that one party transmits while the other receives.

Aircraft

All airplanes from the smallest single-prop planes to the largest jumbo super jets use two-way radio to communicate with airport towers and the major air traffic management centers across the country. Two-way radio is absolutely essential to the safety of any flight.

Most aircraft radios are transmitters installed in front of the pilot or co-pilot. They usually operate in the 108- to 135-MHz band and use amplitude modulation (AM). Specific channels are assigned for different regions and for different communications functions. The pilot or co-pilot uses front panel switches to select the desired communications frequency. Hand-held microphones with a push-to-talk switch are the most commonly used input devices, but other forms of microphones such as throat mikes or microphones attached to headsets are also widely used. Most transceivers operate at very low power and transmission distances are considerable. The height of the airplane gives the radio transceiver an enormous transmission range even with low power at VHF frequencies.

Citizens Band

You have no doubt heard of citizens band (CB) radio and perhaps even used one yourself. Anyone can buy and use a citizens band radio for two-way communications. The FCC has allocated 40 frequencies or channels in the 27-MHz frequency range for citizens band operation. The maximum transmission power is restricted to 5 watts, although some hand-held units use less. AM is used.

Citizens band radio has enjoyed wide popularity with individuals and sportsmen as well as truckers. While the advent of the cellular telephone has reduced the number of individuals using citizens band radio, it is still widely used. The primary limitation of most CB radios is the antenna. Since CB radios operate in the 27-MHz range, the required minimum antenna length is approximately 9 feet. This can be shrunk somewhat by the use of coils attached to the antenna, but effectiveness is reduced. A CB radio antenna of 5 or 6 feet is common and certainly manageable on a car, truck, or boat. In hand-held transceivers, any antenna longer than about 18 inches is a nuisance.

A 27-MHz frequency allocation of CB radio also gives it some peculiar transmission characteristics. While communications is reliable over short distances of up to several miles, at certain times of the day or evening, worldwide communication is also possible. This is probably more a disadvantage than an advantage, primarily because interference can be received from stations literally anywhere in the world.

Amateur Radio

Amateur or ham radio is a hobby. Technically knowledgeable individuals are licensed by the FCC to use selected frequency bands across the spectrum for personal communications. Hams use voice as well as CW (continuous wave) or Morse code transmissions. Hams talk to one another about their equipment, the weather, and other common matters. Communication is worldwide and helps to promote good international relationships. Hams frequently design, build, and experiment with their own equipment, although many hams prefer commercially manufactured radios. Hams experiment with digital radios, satellites, and TV.

Family Radio

Family radio is another radio service recently instituted by the FCC for the purpose of short-range personal communications. It supplements CB radio in a wide range of personal communications needs. Because of a different frequency allocation, family radios are smaller, more portable, and have more reliable communications over distances of several miles.

All family radios are hand-helds using FM and operating in the 450-MHz range. Only several frequencies in the 450 MHz band are allocated to the Family Radio Service. While many users will share these frequencies, the short communication distances minimize interference between stations using the same frequency. The transceivers are restricted to 2 watts in power. At a frequency of 450 MHz, a quarter-wave whip antenna is only about 6 inches long. Communications are extremely reliable up to 1 mile and even farther if transmission takes place outdoors.

You can use family radios in a variety of ways. They are particularly handy when you are camping, hiking, or doing other outdoor activities. Many families use them in shopping malls and even around the house to keep in touch. They are great on farms and ranches and for communications between two nearby cars.

Marine

The very first application of two-way radio in the world was on ships at sea. Ships were always in danger without any form of communication. If a ship was disabled or began sinking, no one would know for a long period of time or maybe never. With a two-way radio or even a one-way radio on a ship, the operator could communicate with a shore-based radio station to signal danger with the familiar SOS.

Today, a radio is a necessity on virtually every boat and ship. While small boats are not required to carry radios, most larger boats are. And what seaman would be without a radio for safety reasons?

Marine radios have operated on a wide range of frequencies over the years. Ocean-going liners have operated on shortwave frequencies so that long-distance communications are possible. Ships also operate on lower frequencies for exchanging messages and at the higher frequency for reliable communications near land and in port.

The typical marine radio for communicating near ports is a transceiver of 25 to 40 watts of power operating with frequency modulation in the 150- to 170-MHz range. Most of these units transmit on multiple frequencies, making it always possible to find a clear channel. Channel 16, operating on a frequency of 156.8 MHz, is used for emergency calls. This frequency is typically monitored by other boaters and always by the Coast Guard.

Mobile Radio

Mobile radio refers to two-way radios mounted in cars and trucks. These are widely used by police, fire, and other civil services. They are also widely used by taxis, buses, and by companies using fleets of cars or trucks for convenient communications. The typical unit is mounted in the vehicle and provides relatively high power in the 25- to 50-watt range. Most communications take place in the 150- to 170-MHz range. Other mobile radio frequency ranges are 450 to 470 MHz. Most radios today are hand-held types.

SATELLITES

A satellite is a space station that is designed to orbit the Earth and provide a remote radio signal relaying station for the purpose of extending the range of communications worldwide. See Figure 7.21. Such communications satellites contain both a receiver and transmitter that operate simultaneously. The receiver picks up signals from the Earth on one frequency and retransmits them simultaneously on another frequency. In this way, the transmitter and receiver in the satellite do not interfere with one another.

Most communications satellites are geosynchronous. These are satellites positioned to orbit directly around the Earth's equator at a distance of 22,300 miles from Earth. When located in this position, the satellite rotates in exact synchronization with the Earth's rotation. Thus, the term *geosynchronous*. Because the satellite and the Earth are rotating together, it appears as though

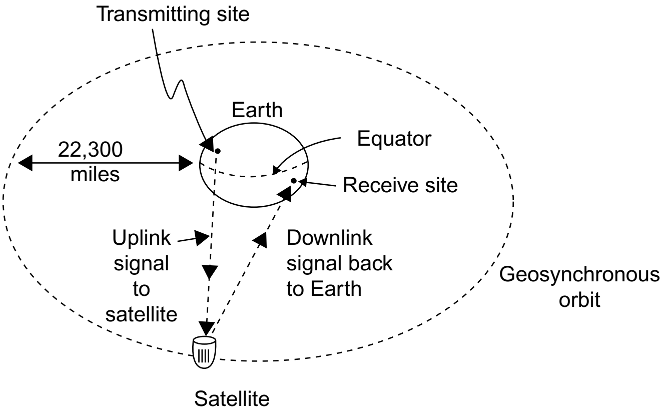

FIGURE 7.21 Concept of communications satellite.

the satellite is stationary overhead to an observer on Earth. With the satellite in a fixed position with respect to the Earth, the satellite becomes a near-perfect radio repeater station.

Figure 7.22 shows a general block diagram of a satellite. The communications part of a satellite consists of a combined receiver and transmitter called a transponder. Satellites operate in the microwave region. The C band from 4 to 6 GHz and the K_u band from 11 to 18 GHz are the most popular operating ranges. A low-noise amplifier (LNA) picks up the very small 6 GHz uplink signal from Earth. It is mixed with a 2-GHz LO signal and down-converted to a 4-GHz downlink signal, which is then amplified by a high-power amplifier (HPA) and retransmitted back to Earth. Most satellites contain many transponders with wide bandwidth capable of handling many signals simultaneously.

The satellite power source is batteries combined with a system of solar cells. The solar cells convert light from the Sun into electrical energy that is used to recharge the batteries. This system gives the satellite long life in space.

Communications satellites are used primarily in long-distance telephone communications and for distribution of TV signals. Some satellites are used for navigation, a good example being the global positioning system (GPS) widely used by the military and commercial organizations. It is also widely used by individuals as they find their way with personal navigation devices (PNDs) like those from Garmin and Tom Tom. The GPS system uses 26 satellites about 10,800 miles above the Earth orbiting in overlapping patterns so that at least three of four are “visible” by wireless at any point on the Earth.

Other satellites are used for surveillance. Spy satellites only 102 miles high pick up radio signals as well as infrared images and photographs. Weather satellites monitor cloud formations and movements. There are also two satellite telephone networks with low-Earth-orbit satellites to permit communications to/from any point on the planet.

RADIO TELESCOPES

Observation of the planets and stars is usually done with an optical telescope. However, in 1932 researchers discovered that distant stars or suns emitted radio waves. All stars, like our Sun, emit radio frequencies. Most of this energy occurs in the microwave region at 1.4218 GHz. This is the frequency of ionized hydrogen. By building a highly sensitive radio receiver and coupling it to a large highly directional radio antenna, it is possible to plot the position of distant stars just as effectively as an optical telescope. Radio telescopes are, in fact, even more effective in that they can more easily detect distant stars that cannot be seen even by the best optical telescopes.

Radio astronomy observatories consist of an enormous parabolic dish antenna usually 60 to 250 feet in diameter, which can be rotated and tilted so that it scans the skies. An array of multiple smaller dishes can also be used as in the SETI (search for extraterrestrial intelligence) project in New Mexico. The antennas are coupled to one or more very sensitive radio receivers. These

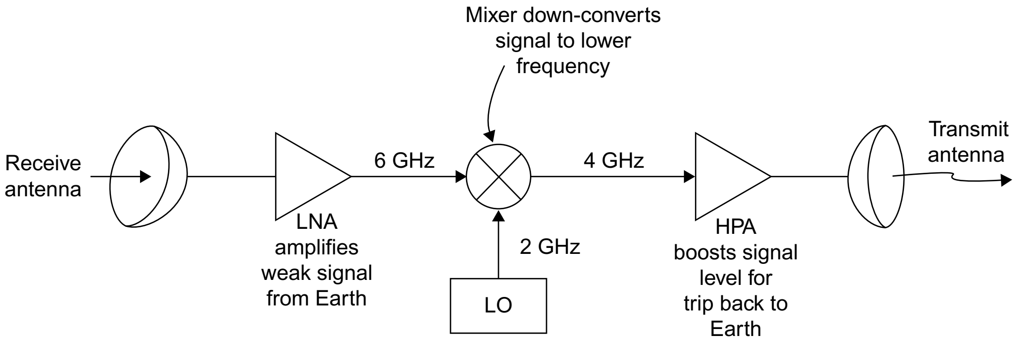
Mixer down-converts
signal to lower
frequency
Receive
antenna
6 GHz
4 GHz
Transmit
antenna
LNA
amplifies
weak signal
from Earth
2 GHz
LO
HPA
boosts signal
level for
trip back to
Earth

FIGURE 7.22 Satellite transponder. LO, local oscillator.

receivers have very high gain so that they can sufficiently amplify very small signals received from stars that are many light years away. Often special cooling systems are used to reduce the temperature of the receiver front end to minimize noise, thereby permitting even weaker signals to be received. The largest radio telescope is a parabolic dish reflector built of screen mesh 1000 feet in diameter. The dish is built into the valley between adjacent mountains near Arecibo, Puerto Rico. The horn antenna is moveable on a long cable so it can be positioned to see any part of the sky.

Wireless Everything

It seems like everything is going wireless these days. Cordless telephones, wireless local area networks, and your TV remote control. But wireless has always been around just not by that name. Most radio applications are old and well known. A few examples you may have encountered follow.

Garage door opener. A tiny transmitter in your car transmits a binary coded message to the receiver in your garage hooked up to the door motor.

Remote keyless entry. A miniature transmitter on your key ring transmits a signal to a receiver in your car that opens the door locks. Some even turn on lights and start your car from 10 to 30 feet away.

Wireless data acquisition. Remote transmitters are set up with solar batteries at remote sites to gather data from sensors and transmit it back to a collection center. A good example is electric, gas, and water utilities that use such systems to monitor critical conditions at remote sites and transmit the data back to a central monitoring facility where decisions can be made. Home wireless thermometers are very common.

Radio control of models. This is very common in toy cars, model airplanes, and boats.

RFID. Ultra miniature radios are being embedded in shipping labels, cars, and other objects for identification purposes. The units transmit their unique ID code when scanned by a nearby receiver.

HD radio. This is digital radio broadcasting in the United States. Most U.S. AM and FM stations now transmit a digital version of their signals on the same frequencies as their analog transmissions. They use a form of digital modulation called OFDM (see Chapter 8). It produces greater frequency response than the regular analog broadcasts, making AM sound more like FM and FM sound more like a CD. A special radio is needed to receive it. The digital technology makes noise and fading less of a problem. And it permits each station to add several additional channels thanks to the multiplexing capabilities of the HD technology.

Security tags. The plastic tags clipped on new clothes to prevent their being stolen contain a tiny antenna and electronic circuit that make their presence known at a nearby transmitter/receiver unit.

Surveillance and bugging, or secret monitoring or electronic spying. This is not just something you see in the movies or on TV. It is more widely used in private investigation, police work, and industrial espionage than you may think. Scary.

Face it, wireless is truly everywhere.

Project 7.1

Investigate HD Radio

HD radio is digital radio in the United States that shares the same frequencies as the regular AM and FM bands. You can buy stand-alone HD receivers, and some stereo receivers include it. HD radio is also available in some cars. To learn more about this technology, go to the website of the developer iBiquity or www.ibiquity.com. Also try a Google/Bing/Yahoo search on HD radio.

Project 7.2

Become Familiar with ISM Radio

ISM radio is the term used to refer to short-range data radios used in the FCC's unlicensed spectrum designated for industrial-scientific-medical (ISM) applications. Got to the FCC website www.fcc.gov and find the most commonly used ISM frequencies of operation in Part 15 of the rules and regulations. Do a general search of ISM to determine some common applications.

Project 7.3

Radar

Radar (radio detection and ranging) is a type of wireless. It has many uses, mainly military but also commercial and civilian. Marine and aircraft navigation are key applications. Police radar is widely used. Radar is also showing up in automobiles as detectors of distance between vehicles for automatic braking. You may wish to learn more about this technology. Do a search on radar.

Project 7.4

Shortwave Radio

Buy a low-cost shortwave radio and experience listening to international broadcasts and amateur radio operation.

Project 7.5

Experience CB or FRS

Go to your local Radio Shack and buy a pair of Family Radio Service walkie-talkies. Then test them to determine maximum range. As an alternative, buy a CB radio and listen to what is going on with the truckers or other locals.

Chapter 8

Cell Phones

It Is Now Possible to Do Anything Wirelessly: Talk, Text, Email, Web Browse, Games, Whatever

In this Chapter:

- How the cellular system works.
- Cell phone radio technologies.
- Digital modulation and access schemes.
- Cell phone data capabilities.
- Smartphones, femto cells, MIMO, and location technology.

INTRODUCTION

The cell phone is one of the coolest, most useful, and convenient electronic devices of all time. It also happens to be one of the most complex technically, not only internally but because it is part of the largest communications network in the universe, the telephone system. The reason for a separate chapter on this topic is not only because of the advanced wireless technology involved, but also its importance to all of us personally and to the national economy. Since practically everyone has a cell phone today, it plays a huge role in our lives. And for you techies out there, it is worthwhile knowing a bit of the details of how phone calls work, how text messages get sent, and how you can now access the Internet via your cell phone.

CAN YOU HEAR ME NOW?

We've all said that before. Why? For one simple reason: cell phones are two-way radios with all the usual faults. The cell phone industry usually refers to itself as the wireless industry, but it's all the same thing—radio. Your cell phone contains a *transmitter* and a *receiver* that collectively are called a *transceiver*. You will also hear them called *handsets*. People in the industry also call them *mobile terminals*. Whatever. Because cell phones have to be so small to be practical, they are limited in the amount of power they can transmit and the type of antenna that is practical. Battery drain is a major factor in cell phone design, as you want to get as much talk time as you can on a battery charge. But by limiting the transmit power you automatically minimize the distance

© 2010 Elsevier Inc. All rights reserved.
doi: 10.1016/B978-1-85617-700-9.00008-4

over which the unit will transmit. In any case, communications range is related to power as well as antenna type and orientation. And the typical range of a cell phone is a mile or so, depending on conditions. Take a look at the sidebar to review some of the facts about wireless and radio covered in Chapter 7.

Factors Affecting Range and Reliability to Radio Transmission

Range—The greater the distance between any transmitter and receiver, the smaller the signal. Signal strength varies inversely by the square of the distance between transmitter and receiver.

Power—The higher the power the greater the transmission range. More power translates into increased communications distance up to a point.

Antenna gain—Some antennas focus the signal in a specific direction. Doing so gives the antenna *gain*, meaning it multiplies the power of the transmitter and the signal strength at the receiver.

Environment—Cell phones and any other microwave wireless device always transmit farther in clear open space. The ideal is direct line of sight (LOS), meaning that the two antennas can "see" one another. While cell phones don't have to have such a clear path to work, they always work better if they do. Environment also means transmitting from inside a building. Yes, radio waves do pass through walls and other obstructions but as they do they lose considerable strength as they encounter obstacles. Walls, ceilings, floors, and any other obstructions greatly attenuate (make it smaller) the radio signal. So if you are inside a car, shopping mall, office, or airplane, your signal strength loses over half its strength and that severely restricts your range.

Height—Getting your cell phone to the highest possible point will greatly increase transmission distance. This helps the line of sight effect so important to reliable transmission. Getting higher typically isn't practical, but almost nothing helps more than height, especially outside.

Receiver sensitivity—This means how much gain the receiver has. High gain translates into greater range.

Multipath—Whenever you are transmitting in an environment with lots of obstructions, your signal will bounce off many objects and be reflected. Some reflected signals will get lost and never get to the receiver. Others will bounce off objects that will ultimately direct them to their intended destination. These reflected signals will be delayed in time. Depending on factors such as operating frequency and distance traveled, these signals could arrive in phase and actually boost reception level. Or they could arrive out of phase and literally cancel the more direct signal that arrived earlier.

Noise—Don't forget that noise also affects communications distance and reliability. See discussion in Chapter 7.

THE CELLULAR CONCEPT

The original designers of the cell phone created a system that helps overcome the range and power problem described in the sidebar. If distances are kept

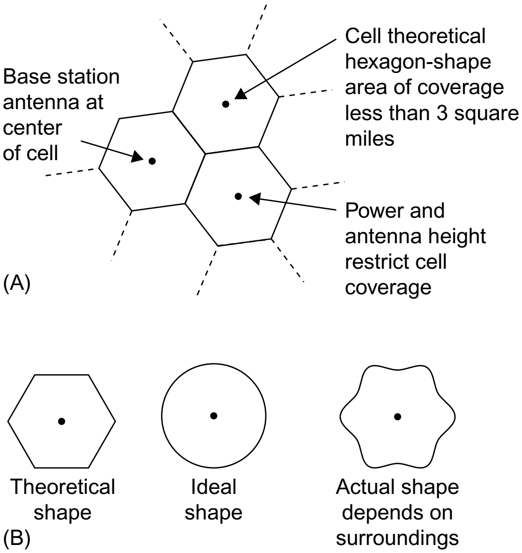
Base station antenna at center of cell
Cell theoretical hexagon-shape area of coverage less than 3 square miles
Power and antenna height restrict cell coverage
(A)
Theoretical shape
Ideal shape
Actual shape depends on surroundings
(B)

FIGURE 8.1 **(A)** Cell coverage of a given area. **(B)** Cell area covered depends on antenna pattern and environment.

short, then other limitations in power and environment are offset. The working word here is "cell." A cell or cell site is one of many transmit/receive stations set up to communicate with individual cell phones. Also known as base stations, these facilities are easily recognized by their tall antenna towers. The small building at the base of the tower houses racks of transceivers that share the big antennas at the top of the tower. Long coax cables carry the signals to and from the antenna. They actually look like pipes going up the side of the tower.

Cell phone systems have dozens to hundreds of cell sites. An ideal design is shown in Figure 8.1A. Each cell represents an area covered by the base station antennas. Cell range and overall coverage are deliberately restricted by antenna height and radiation pattern as well as transmitting power. The hexagon-shaped cells are only theoretical. In reality, the shape of the cell is more circular or rather an irregular circular shape because of antenna characteristics and environmental characteristics that affect the cell shape (nearby trees, buildings, etc.) (see Figure 8.1B).

While equal-size cells are desirable and operation more predictable, in the real world, the coverage of each cell depends a great deal on the number of cell phone subscribers as well as the terrain and environment. In large cities with lots of people and buildings, smaller cell sizes are used to increase subscriber capacity and provide the most reliable coverage possible in a given area. Many cities have what are called microcells that may only cover one block or picocells that cover one area of a building. On the other hand, in rural areas and along

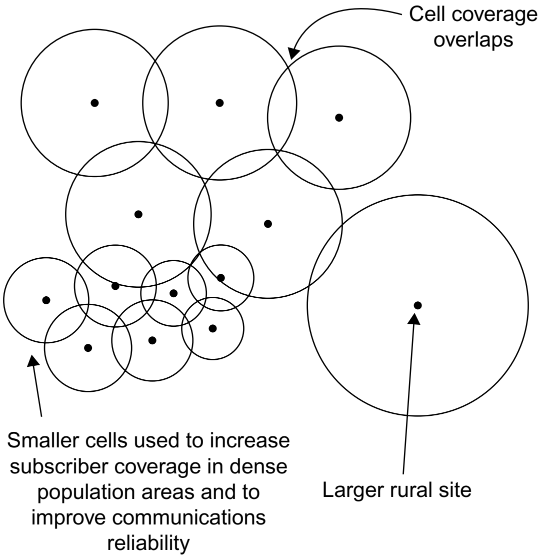
Cell coverage overlaps
Smaller cells used to increase subscriber coverage in dense population areas and to improve communications reliability
Larger rural site

FIGURE 8.2 How small and large cells cover an area.

major highways, the cell sizes are larger and cover fewer users. The cell coverage may look more like that in Figure 8.2.

A key feature of the cellular concept is that by keeping the cell sizes small, the frequency channels assigned to cellular service can be reused. One frequency may carry different calls if the cell sites are far enough apart and do not interfere with one another. This frequency reuse concept multiplies the overall number of channels available to handle calls.

One other factor to keep in mind is that in most cities, there are two or more cellular companies vying for your business. Each company has its own cellular system with base stations with coverage that overlays the cell coverage of other companies. The competing systems do not interfere with one another because they usually operate on different frequencies with different radio technology, and the cell phones they support know which base stations to talk to.

One other thing: when you are talking on your cell phone, you are talking to only one base station—typically the one closest to you since both the power and range of your handset and that of the base station are limited. If you are moving (such as in a car), you will eventually leave the coverage area of one cell and enter the coverage area of an adjacent cell. You never really know when this happens, but it is a pretty neat technical trick to make this happen. The process of having your call transferred from one cell site to another automatically is called the *hand-off*.

All of the base stations are tied to a central office known as the *mobile telephone switching office* (MTSO). Many companies simply refer to it as the *switch*, as its main job is switching between base stations as well as linking to

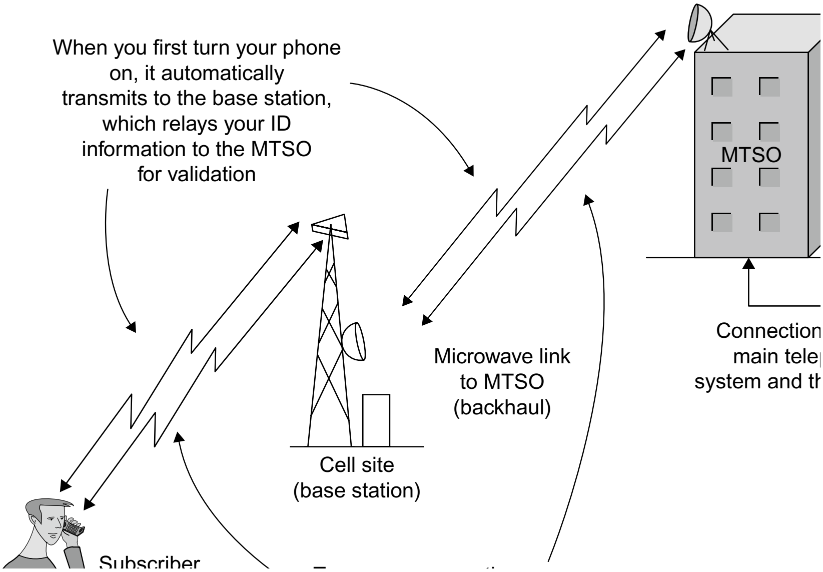
When you first turn your phone on, it automatically transmits to the base station, which relays your ID information to the MTSO for validation
MTSO
Microwave link to MTSO (backhaul)
Cell site (base station)
Connection
main tele
system and th
Subscriber

FIGURE 8.3 Cell phone basic operations.

the main telephone system. The connection between the base stations and the switching station can be by copper cable or fiber optic cable or in many cases by a separate microwave wireless link. This is known as *backhaul*. If a cell site tower has one or more small "dish" antennas on it you can bet that your signal eventually is communicated back to the switch via another wireless link.

The big job of the MTSO is to control and keep track of everything. It validates you as a subscriber when you turn on your phone and make a call. It keeps track of how many minutes of time you use. And it will get messages from remote switches when you are in the roam mode outside of your usual home subscriber area. The cell phone companies have the ability to use other systems in most major cities and in many foreign countries, so you can use the phone virtually anywhere. Figure 8.3 shows just what goes on when you turn on your phone and make or receive a call.

WHAT'S INSIDE A CELL PHONE?

A cell phone can legitimately be called the world's most complicated two-way radio. Sure, it is just a transceiver, but it uses practically every radio trick in the book to ensure that you can make your calls reliably anytime with minimal hassle. It is a two-way radio that is used just like any other phone you are familiar with. In other words, you do not have to say "over," "come back," or any other phrase when you are finished talking and want to listen, as with some two-way radios (CB, family radio, ham radio, aircraft, marine, etc.).

Simultaneous send and receive is referred to as *full duplex*. All telephones are full duplex. It is easy to do in the wired telephone system but complex and costly to do by radio. Yet, full duplex is what makes cell phones so comfortable to use. There is none of this “roger and over” stuff. What this means is that your transmitter and receiver are working at the same time. There is more about how duplexing works later.

Figure 8.4 shows a general block diagram of what’s in a cell phone. The receiver and transmitter making up the transceiver are obvious and they share the single antenna. The transceiver is usually a single chip. What is not so obvious is what we call the baseband part of the cell phone. Baseband refers to the voice and data to be transmitted or received. In most modern digital cell phones, this is a very complex integrated circuit that handles the translation of voice between analog and digital, modulation and demodulation, and voice compression. All these functions are done digitally by a digital signal processor (DSP) or special DSP circuits in combination with a powerful embedded processor. This processor or a separate processor handles all the transparent housekeeping jobs of managing the keyboard and the LCD screen as well as the basic control features of the cell system, which also include automatic operating frequency selection and automatic power control under the direction of the MTSO via the cell site. The DSP also implements the messaging functions.

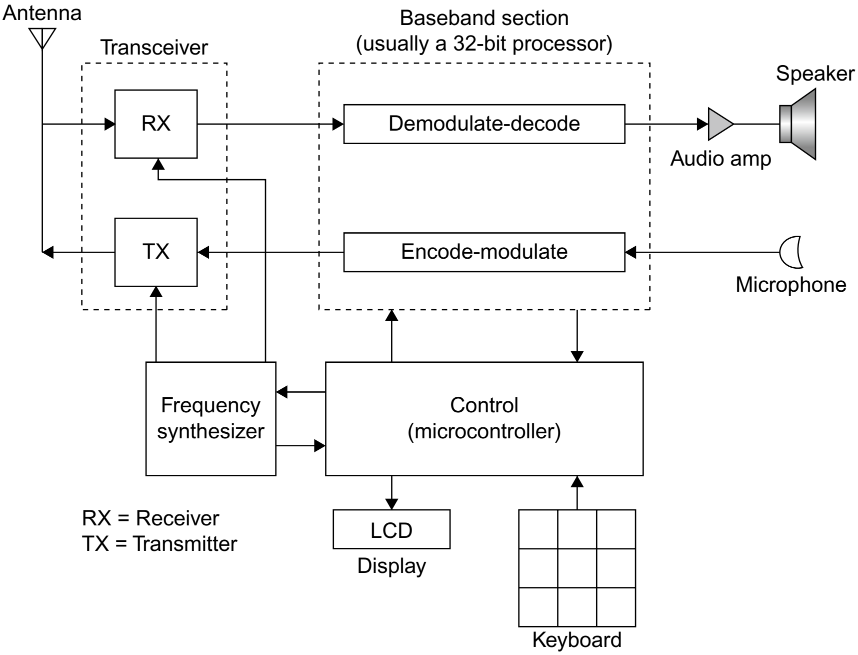
Antenna
Transceiver
Baseband section
(usually a 32-bit processor)
Speaker
RX
Demodulate-decode
Audio amp
TX
Encode-modulate
Microphone
Frequency synthesizer
Control (microcontroller)
RX = Receiver
TX = Transmitter
LCD
Display
Keyboard

FIGURE 8.4 Block diagram of basic cell phone.

What Type of Cell Phone Do You Have?

Shame on you if you don't know. Well, not really. Actually, it is tough to know just what type of cell phone you have. You can't tell by looking, even inside, since even if you open up your cell phone, God forbid, all you see is a group of integrated circuits, and you can't tell one from another. Furthermore, your cell phone carrier doesn't tell you and usually your user's manual never really says. You can probably ask by calling your carrier, but even then you may not get an answer. I guess the bottom line here is, do you really care? If it does what you want, what difference does it make what technology you are using? On the other hand, if you are an electronics geek, you want to know. I do.

So even though no one knows what type of phone you are using, it does impact the overall system and what phone features and capabilities you have. Multiple types of cell phone technologies are in use concurrently, and that is what makes the cell phone system so bloody complex and expensive. Most of the original carriers have to continue to support the older original technologies as well as implement new ones to stay competitive with other carriers offering their hot new data technologies and features such as email, instant messaging, games, Internet access, mobile TV, and the like.

The first characteristic of a cell phone is the frequencies it uses. Most cell phones operate in the 800- to 900-MHz band where there are hundreds of channels for cell phone calls. Another range of frequencies is the so-called personal communications system (PCS) band from 1850 to 1990 MHz (1.85 to 1.99 GHz), which has hundreds more channels. Some of the newer phones use the advanced wireless services (AWS), 1700- and 2100-MHz spectrum assignments. Because of the higher frequency, PCS and AWS systems have a shorter range, and thus use smaller area cells, which means many more cells sites to cover a given area. Some cell phones actually operate on two or more bands.

Next, a cell phone is known by the technology and access methods it uses. The early original analog phones used first-generation (1G) technology, or FM, but have now been phased out so that only second-generation (2G) digital phones are in use. We are now in the third generation (3G) of cell phone technology and 4G is almost with us. A summary of these technologies follows.

GSM—The Global System for Mobile Communications (originally called Groupe Spécial Mobile) is the 2G digital system developed in Europe. Now all of Europe, most of the United States, and the rest of the world use this excellent but seriously complex technology. It uses time-division multiple access (TDMA), a digital system using digital modulation and a method of multiplying the number of channels in a given amount of frequency spectrum. Your voice is converted to binary 1's and 0's before being transmitted. This allows the carrier to put eight subscribers in one 200-kHz-wide frequency channel, thereby greatly multiplying their subscriber capacity and income. GSM uses a form of FSK called Gaussian minimum shift keying (GMSK). Today, U.S. carriers ATT Wireless and T-Mobile use GSM. By going to

GSM, these carriers can take advantage of several enhancements that facilitate high-speed data transmissions such as the General Packet Radio System (GPRS) and Enhanced Data Rate for GSM Evolution (EDGE).

GPRS—This is an extension of GSM that gives users an always-on packet data transmission capability that they can use for email, Internet access, messaging, or games. GPRS steals one or more of the eight TDMA channels to transmit or receive data instead of digital voice. You can achieve a data rate of about 20 to 60 kbps. GPRS is known as a two-and-a-half generation (2.5G) technology, and is generally available in most larger cities.

EDGE—This 2.5G system is a software upgrade from GPRS/GSM systems that gives you even higher packet data rates. It uses a multilevel modulation called 8PSK that triples the speed of data transmission. Practical speeds up to 180 kbps are possible.

CDMA—Code division multiple access is another 2G digital cell phone technology developed by San Diego–based Qualcomm Inc. The basic technology is known as *spread spectrum* where many signals are transmitted simultaneously over a very wide 1.25-MHz frequency band without interfering with one another. The original standard is called IS-95, and CDMA signals coexist in the same spectrum with TDMA systems in many areas.

cdma2000—This is the 2.5G version of CDMA. It adds high-speed packet data transmission called 1xRTT with a data rate up to 144 kbps. A more recent upgrade called 1xRTT EV-DO has even higher data rates up to 2 Mbps, making Internet access a cinch. Many cdma2000 systems are in operation around the United States by Verizon and Sprint and in Korea.

WCDMA—This is wideband CDMA, which is a third-generation (3G) technology. The access technology is CDMA, but it uses wider 5-MHz bands to provide greater user capacity and very high potential data rates from 384 kbps to 2 Mbps. WCDMA is the upgrade from GSM/EDGE. Carriers using GSM upgraded their systems to WCDMA for greater data speeds and services at greater subscriber capacity. Base stations support both GSM and WCDMA.

HSDPA/HSUPA—High-speed downlink packet access (HSDPA) and high-speed uplink packet access (HSUPA) are upgrades from the WCDMA system. Both use QAM to boost data speeds in the same 5-MHz channels. Data rates as high as 14 Mbps down-link and 5 Mbps up-link are possible. Most cellular companies have now implemented this technology. More advanced versions called HSPA or HSPA+ can give even higher data rates.

cdma2000 Rev A and **Rev B**—These are more advanced CDMA versions that are used to boost data speeds in cdma2000 handsets. Rev A boosts download data rates to 3.1 Mbps using QPSK and 16-QAM in the standard 1.25-MHz channel. Rev B is not widely implemented, but can use three 1.25-MHz channels and boost download speeds to 14.7 Mbps.

Long-Term Evolution—Long-Term Evolution (LTE) is the next or fourth-generation (4G) cell phone technology. It is just now being implemented. It is appearing in Europe first and the United States later. It uses an entirely different radio technology called orthogonal frequency division multiplexing (OFDM) to give even higher data rates up to 100 Mbps. More on that later.

How can you tell which of the above phone technologies you have? You cannot tell by the manufacturer, such as RIM BlackBerry, Samsung, LG, Motorola, Nokia, Sony Ericsson, and so on, because they all make multiple types. However, a few clues follow.

Carriers AT&T and T-Mobile provide GSM/WCDMA phones. Most also have some form of HSPA. These carriers will eventually morph their systems to LTE. Carriers Sprint and Verizon use cdma2000 phones. Most phones today are also what we call multimode phones in that they incorporate two or more different technologies. The most common combination is GSM/WCDMA/HSPA.

DIGITAL MODULATION AND WIRELESS TRANSMISSION METHODS

In Chapter 7, basic digital modulation methods ASK/OOK and FSK were covered. These are not used except for an advanced FSK as used in GSM. Instead, all cellular systems used advanced digital modulation methods. This section provides a look at these techniques.

Quadrature Amplitude Modulation

Quadrature amplitude modulation (QAM) is a technique that mixes both amplitude and phase variations in a carrier at the same time. What this technique does is to allow higher data rates within the same bandwidth. See the sidebar on spectral efficiency. The technique consists of transmitting multiple bits for each time interval of the carrier *symbol*. The term "symbol" means some unique combination of phase and amplitude. The term M as in M-QAM indicates how many bits are transmitted per time interval or symbol for each unique amplitude/phase combination.

The simplest form of QAM is 2-QAM, more commonly called QPSK or quadrature phase shift keying. It is produced by the circuit shown in Figure 8.5. It takes the serial bitstream and passes it through a 2-bit shift register producing two parallel bitstreams at half the rate. One is called the in-phase (I) bitstream and the other the quadrature (Q) bitstream. The I and Q signals are each applied to a mixer along with a local oscillator (LO) signal. The mixer output becomes part of the RF carrier. Note that the LO signal is applied directly to one mixer, but is shifted by 90 degrees and applied to the other mixer. The term "quadrature" means shifted by 90 degrees. The two mixer outputs are AM

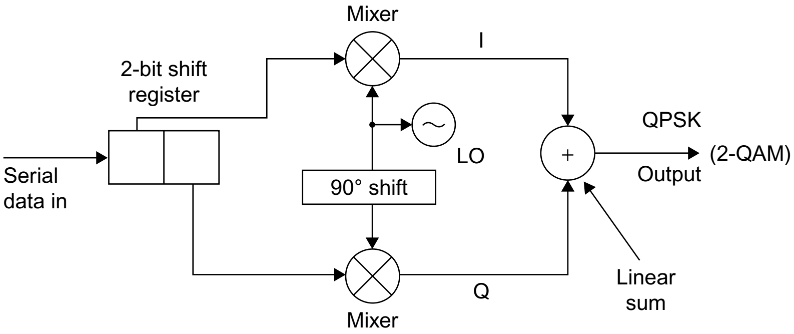
Mixer
I
2-bit shift register
Serial data in
LO
90° shift
QPSK
(2-QAM)
Output
Linear sum
Q
Mixer

FIGURE 8.5 How QPSK is generated using mixers and a local oscillator shifted 90 degrees to produce in-phase (I) and quadrature (Q) signals that are summed.

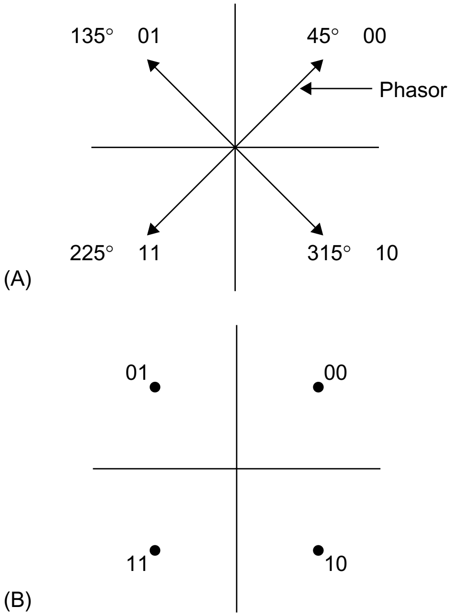
135° 01
45° 00
Phasor
225° 11
315° 10
(A)
01
00
11
10
(B)

FIGURE 8.6 QPSK diagrams. **(A)** Phasor. **(B)** Constellation.

signals but at different phases. But when they are added together, the QPSK signal is generated. With this arrangement, you can transmit data at twice the rate of BPSK in the same bandwidth.

Figure 8.6 shows how we illustrate this with what we call the phasor and constellation diagrams. Figure 8.6A shows the amplitude and phase of each possible combination of 2 bits. The length of the arrow called a *phasor* is the amplitude, and its direction or angle is the phase. If the receiver gets a signal

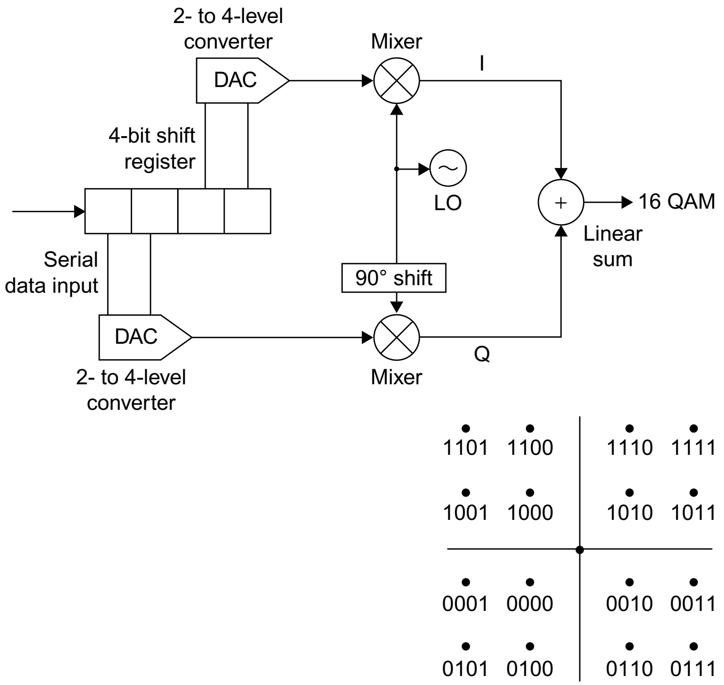
2- to 4-level converter
DAC
Mixer
I
4-bit shift register
LO
16 QAM
Linear sum
Serial data input
90° shift
DAC
Q
2- to 4-level converter
Mixer
1101
1100
1110
1111
1001
1000
1010
1011
0001
0000
0010
0011
0101
0100
0110
0111

FIGURE 8.7 How 16-QAM is generated and resulting constellation diagram.

that is at a phase angle of 135 degrees, it means that the 2-bit combination 01 is being transmitted. The constellation diagram in Figure 8.6B is usually shown without the arrows or phasors.

This technique can be modified to produce 16-QAM, which transmits 4 bits per time interval or symbol causing the data rate to quadruple in the same bandwidth. The circuit for producing 16-QAM is shown in Figure 8.7. The 4-bit shift register produces two groups of I and Q bits. These are converted into four levels by a two- to four-level converter that works like a digital-to-analog converter (DAC). The resulting constellation diagram for 16-QAM is shown in Figure 8.6 as well. Other common versions are 64-QAM and 256-QAM, which transmit 6 and 8 bits per symbol, respectively, in the same bandwidth.

QPSK and QAM modulation methods are common in cable TV for digital TV as well as high-speed Internet service. They are also used with satellites and any broadband wireless application. All modern cell phones and most other digital wireless schemes use this type of modulation because of its spectral efficiency.

Orthogonal Frequency Division Multiplexing

Orthogonal frequency division multiplexing (OFDM) is a broadband modulation method like spread spectrum/CDMA. It takes a high-speed serial binary

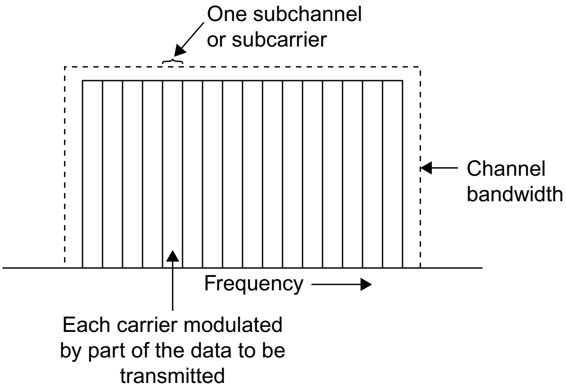

FIGURE 8.8 OFDM divides a channel into many subchannels or subcarriers, each of which is modulated by part of digital data to be transmitted.

signal and spreads it over a wide bandwidth. The serial data is passed through a circuit that maps out the constellation diagram for the modulation to be used. It then divides it into many slower-speed serial bitstreams. Each bitstream modulates a carrier on one of many adjacent carriers in the available bandwidth. This technique effectively divides a wide bandwidth into many narrower subchannels or subcarriers as shown in Figure 8.8. Sometimes dozens of channels are used, and in other cases hundreds or even thousands of carriers are used. The carrier frequencies are selected so they are orthogonal and as a result they will not interfere with one another even though they are directly adjacent. All the modulated channels are then added together and the combination transmitted in the available bandwidth. The type of modulation varies with the application but it is usually BPSK, QPSK, 16-QAM, or 64-QAM.

The basic technique is shown in Figure 8.9. The only practical way to create an OFDM signal is by using DSP. The DSP executes an algorithm called the inverse fast Fourier transform (IFFT) that creates the multiple adjacent modulated carriers. The DSP data is applied to DACs to create I and Q signals that are sent to mixers for creation of the final signal. At the receiver, the data is recovered by another DSP executing the FFT. The output is the original fast serial data.

OFDM seems impossibly complex while seemingly hogging spectrum space. Yet, it can transmit higher speeds in smaller bandwidths than most other types of digital modulation. It is very spectrally efficient. Furthermore, it is more resistant to multipath interference that can cause microwave links to lose data due to a reflected signal interfering with another or a direct signal causing cancellation and fading. Most of the new wireless technologies today use OFDM, including the forthcoming LTE cellular systems. Some examples of other OFDM uses include wireless local area networks (WLAN), wireless broadband (WiMAX), digital subscriber line (DSL) Internet access, digital TV, and AC power-line networking.

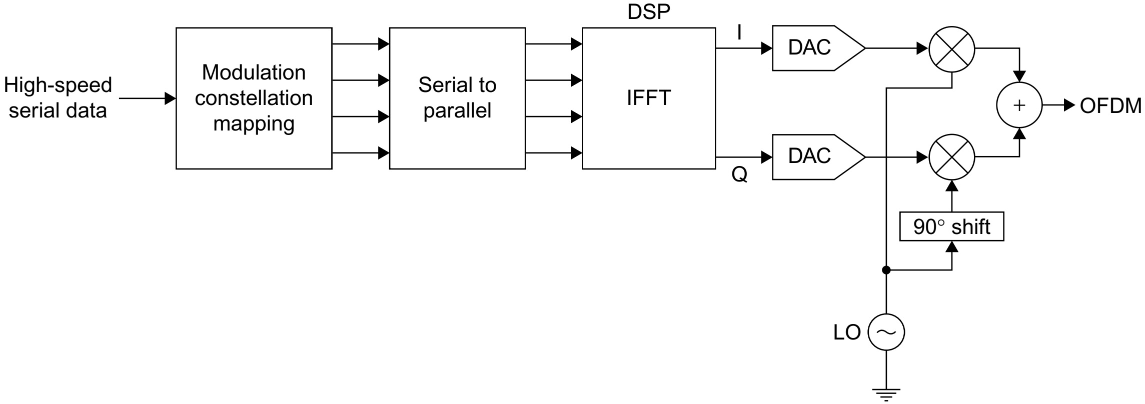
High-speed serial data
Modulation constellation mapping
Serial to parallel
DSP
IFFT
I
Q
DAC
DAC
+
OFDM
90° shift
LO

FIGURE 8.9 Generating OFDM is done by DSP and I/Q modulation.

Multiple Input Multiple Output

Another technique for boosting speed within a given bandwidth is multiple input multiple output (MIMO). Also known as spatial multiplexing, it is the use of multiple transmitters transmitting to multiple receivers, each sending different bitstreams on the same channel. The high-speed serial data is divided into two, three, four, or even more bitstreams, and each is used to modulate a different transmitter but operating on the same frequency. The signals from each transmitter reach the antennas of each receiver along with any reflected signals. See Figure 8.10, which shows four transmitters (Tx) and four receivers (Rx) or 4 × 4 MIMO. What allows the receivers to sort out each bitstream is the fact that each signal travels a different path, and will have different characteristics or spatial signature that can be sorted out by DSP in the receiver.

There are several things that make MIMO practical. First, MIMO is used at very high frequencies, especially microwave (>1 GHz). This means that antennas are short and inter-antenna space may be one wavelength or more, which is required for MIMO to work. The distance between the antennas ensures different spatial paths with different characteristics that let the receivers recover each bitstream. Second, thanks to semiconductor technology, it is possible to make the multiple transmitters and receivers very inexpensively, typically all on a single silicon chip.

MIMO has several benefits. First, it allows a higher data rate in the same channel bandwidth. Using a 4 × 4 MIMO, for example, would boost speed by a factor of 4. Configurations of MIMO can be any combination of transmitters and receivers. The 2 × 2 mode is popular but 4 × 4 is also used. Second, MIMO provides a more reliable link. It actually takes advantage of the multipath problems that usually plague microwave wireless and makes it

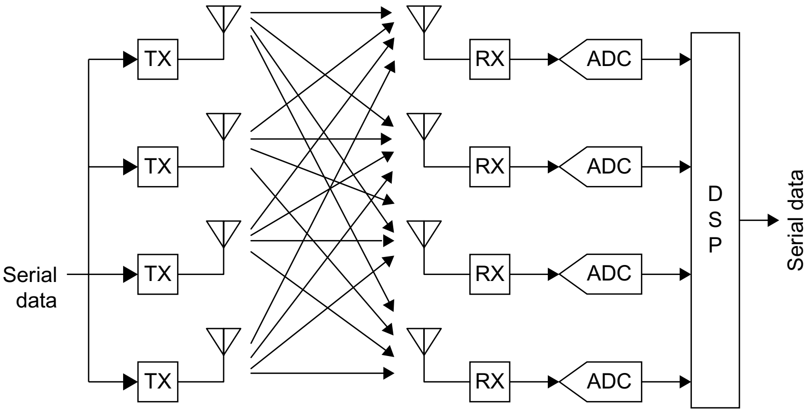

FIGURE 8.10 MIMO in a 4 × 4 configuration quadruples the data rate in an existing channel, while reducing the ill effects of multipath fading.

better. Finally, MIMO is usually combined with OFDM, which provides the ultimate data-rate boost, reliability, and spatial efficiency. MIMO is already used in wireless local area networks (WLAN) like Wi-Fi and in broadband wireless systems using WiMAX. Both of these technologies will be covered in the next chapter. MIMO will also be used with LTE, the next cell phone system. No, handsets probably will not have multiple antennas but the base stations will.

Spectral Efficiency

Spectral efficiency refers to how much data you can transmit in a given bandwidth. It is usually measured in bits per second per hertz (bps/Hz). ASK/OOK and FSK have efficiencies of <1. Special forms of FSK, such as Gaussian minimum shift keying (GMSK) used with GSM cell phones, are a bit better. BPSK has a spectral efficiency of 1. Table 8.1 gives the efficiencies of most of the popular digital wireless modulation schemes.

One last thing: the type of modulation is not the only thing that affects spectral efficiency. The level of noise in a channel also determines the number of bit errors that occur in transmission. The higher levels of QAM have more errors, so this limits their efficiency to some extent. Plain FSK, BPSK, and QPSK are very resistant to noise, so are more suitable in applications where noise is a problem. Noise and bit rates have to be traded off for good reliable transmission. And by the way, the spectral efficiency issue applies to wired as well as wireless communications channels.

TABLE 8.1 Spectral Efficiency of Selected Digital Wireless Modulation Schemes

Modulation Method	Spectral Efficiency (bps/Hz)
FSK	<1
GMSK	1.35
BPSK	1
QPSK	2
16-QAM	4
64-QAM	6
256-QAM	8
OFDM	Depends on modulation of subcarrier and number of subcarriers
MIMO	Depends on configuration

ACCESS METHODS AND DUPLEXING

Access methods are wireless techniques that allow multiple users to share a given part of the frequency spectrum. Access methods are also called *multiplexing*. Multiplexing means allowing two or more signals to be transmitted *concurrently* over the same communications medium such as a cable or within a given frequency band. Access methods allow many subscribers to share precious spectrum space. And they provide the means to implement full-duplex simultaneous transmit/receive that most phone users expect. There are four types: frequency division multiple access (FDMA), time division multiple access (TDMA), code division multiple access (CDMA), and orthogonal frequency division multiplex access (OFDMA). For full-duplex operation, there are two methods—frequency-division duplexing (FDD) and time-division duplexing (TDD).

FDMA—A given bandwidth is divided into many narrower channels as shown in Figure 8.11. A given radio will cover all of the bands but has the ability to select a specific operating frequency. GSM divides the band into 200-kHz channels. cdma2000 uses 1.25-MHz channels and WCDMA uses 5-MHz channels. The newer OFDM uses 5-, 10-, 15-, or 20-MHz channels.

FDD—Simultaneous transmit and receive is accomplished by using two different bands where half are used for transmit and the other half for receive. The transmitter and receiver operate at the same time but on different frequencies so they do not interfere with one another (see Figure 8.11).

TDMA—With this method, a single-frequency channel is used, but the signal is divided into fixed-duration time slots into which digitized segments of voice or data are transmitted. In digital voice transmission, the analog voice

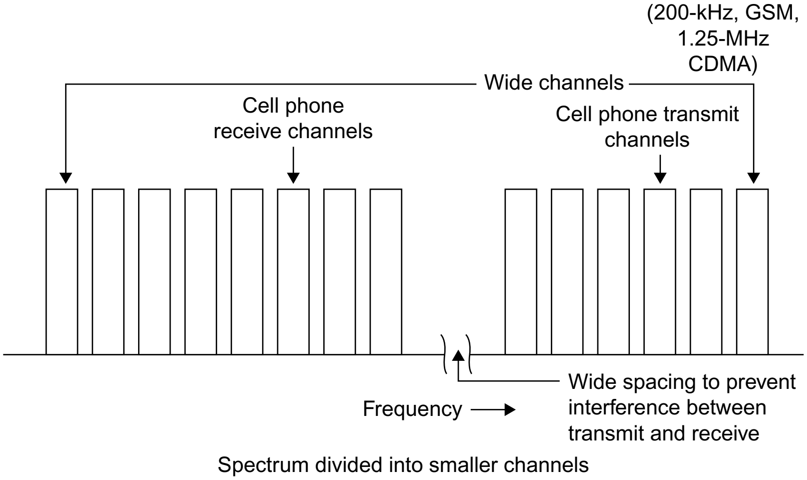
(200-kHz, GSM, 1.25-MHz CDMA)
Wide channels
Cell phone receive channels
Cell phone transmit channels
Wide spacing to prevent interference between transmit and receive
Frequency
Spectrum divided into smaller channels

FIGURE 8.11 Frequency-division multiple access and frequency duplexing use two separate bands of frequencies divided into individual channels.

signal is periodically sampled and digitized into a stream of binary numbers. These binary values are then compressed into a smaller number of bits by a vocoder circuit. The resulting bits are then transmitted to the receiver. With TDMA, binary values from two or more voice sources can be interleaved and sequenced, and then sent over the single channel. For example, in GSM systems, there are eight time slots, as shown in Figure 8.12. The data is transmitted at a 270-kbps rate in a 200-kHz wide channel. Each time slot contains 1 byte of the voice data for one phone call. The frame of eight signals is then repeated.

TDD—Time-division duplexing facilitates concurrent send and receive by assigning transmitted signals in one time slot and received signals in another time slot. They share the same frequency channel. Because all this is happening at a high rate of speed, you cannot determine that transmit and receive are not simultaneous, although in reality they are not.

CDMA—This method uses spread-spectrum technology. It takes the digitized voice and further processes it to give it special coding, and then spreads it over a wide bandwidth. In cdma2000 phones, the channel bandwidth is 1.25 MHz wide, and all cell phone signals are coded and spread over this bandwidth. The process goes like this: the analog voice signal produced by the microphone is digitized in an ADC, and then compressed by a special digital circuit called a *vocoder* or *voice coder*. This reduces the total number of bits and the speed of the digital signal. This vocoded signal is then mixed with a high-speed serial code signal, making it unique so that the receiver will recognize it. This is done with an XOR gate as shown in Figure 8.13. The higher-frequency signal, at 1.23 Mbps, is called the *chipping data*. The resulting signal fills the 1.25-MHz channel. In WCDMA systems, the process is the same, but the chipping signal is 3.84 Mbps and the channel width is 5 MHz.

Many cell phone calls are then transmitted on the same channel. They actually all mix together, but because they occur randomly with one another, they do not interfere with one another. Typically, up to about 64 users can share the same bandwidth. One signal simply appears as a low-level noise to another. The desired signal is picked up at the receiver by recognizing the unique code given to the transmitted signal. The two most common types of spread spectrum are direct

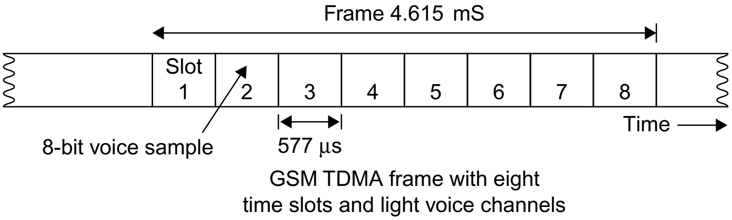
Frame 4.615 mS
Slot 1
2
3
4
5
6
7
8
Time
8-bit voice sample
577 μs
GSM TDMA frame with eight
time slots and light voice channels

FIGURE 8.12 TDMA multiplexing divides a channel into sequential time slots for the digital data.

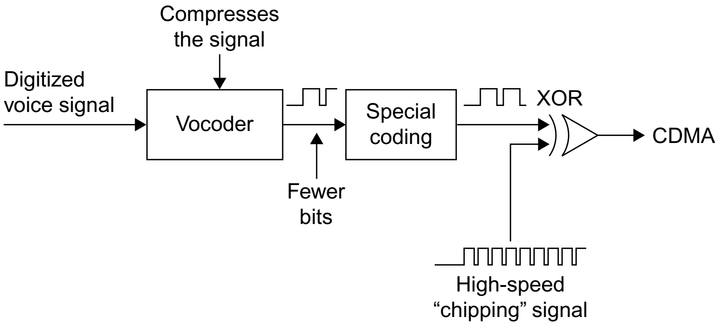

FIGURE 8.13 How CDMA is generated.

sequence and frequency hopping. Direct-sequence spread spectrum (DSSS) is used in cell phones as illustrated in Figure 8.13. Frequency hopping is used in Bluetooth wireless systems (see Chapter 9). Most duplexing is by FDD, but some of the newer versions of WCDMA also use TDD.

OFDMA

Access with OFDM is done by just assigning one or more of the subchannels to each phone call. With each subchannel 15 kHz wide, each is capable of carrying a compressed (vocoded) voice signal so that many calls can be handled. In practice, more than one channel may be used, depending on the specific frequency assignments, bandwidths, and other features of the LTE system being used.

A LOOK INSIDE A 3G CELL PHONE

Figure 8.14 shows the insides of a modern 3G cell phone. This one uses the WCDMA/HSPA 3G technology, but is also backward compatible with GSM/EDGE if an area does not have 3G yet. There is a separate radio transceiver (XCVR) for each. All of the baseband functions like modulation and demodulation and handset management are handled by one or more processors. One processor may be DSP. This handset also has a Bluetooth radio for the headset (not shown) and a Wi-Fi transceiver for Internet access via a wireless LAN access point or hot spot. These technologies will be covered in the next chapter. A GPS radio provides navigation as well as fulfills the E911 requirement. Note that all radios have separate antennas. A camera in included. The CCD is the charge-coupled device that converts the light and color variations in the scene into electrical signals. The stereo headphones are for the MP3 music player.

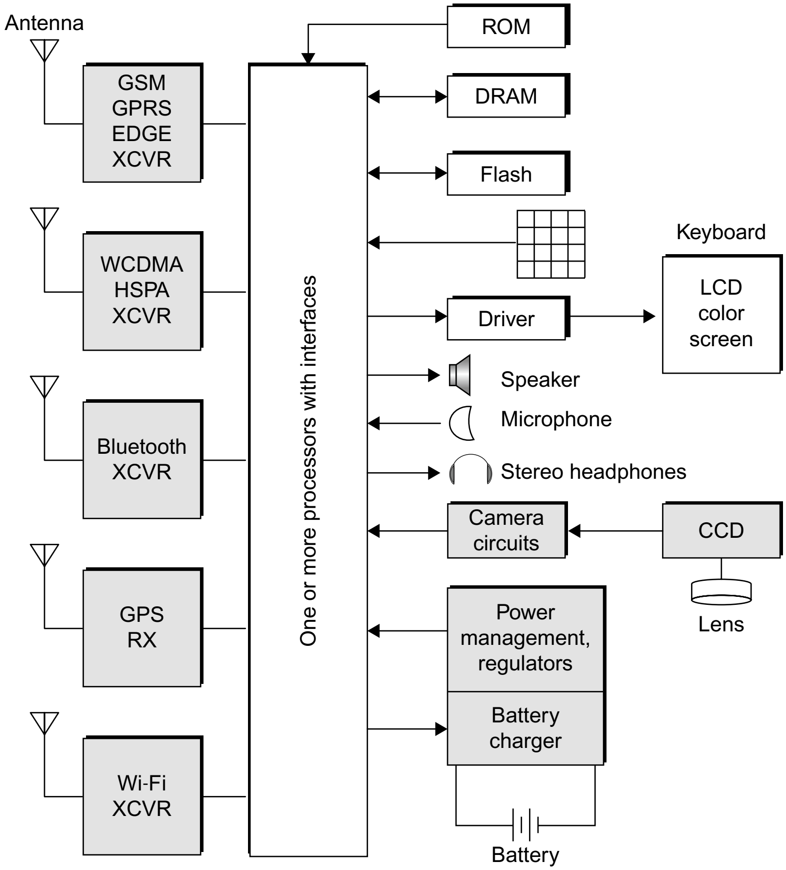
Antenna
GSM
GPRS
EDGE
XCVR
WCDMA
HSPA
XCVR
Bluetooth
XCVR
GPS
RX
Wi-Fi
XCVR
One or more processors with interfaces
ROM
DRAM
Flash
Keyboard
Driver
LCD
color
screen
Speaker
Microphone
Stereo headphones
Camera
circuits
CCD
Lens
Power
management,
regulators
Battery
charger
Battery

FIGURE 8.14 A block diagram of 3G cell phone with multiple radios and functions.

THE LATEST CELL PHONE TECHNOLOGY

The cell phone industry is one of the fastest changing in electronics. New products come out almost daily. And new features and capabilities are invented regularly to keep the technology fresh and the sales continuous. Here is a snapshot in time of the latest cell phone developments.

Smartphones

A smartphone is a basic cell phone but with advanced features that expand its use. The basic cell phone does phone calls or at most handles text messaging as well. But a smartphone has multiple functions. Some examples of smartphones are the BlackBerry models from Research in Motion (RIM), Apple's iPhone, some Nokia models, the Palm Pre, and various models from Samsung, LG, and Motorola. They are too numerous to mention and they go out of date monthly or yearly.

High-speed data capability. Most smartphones use 3G technology, so they can deliver almost as fast a data rate as your home Internet connections, usually 1 to 2 Mbps or better.

Internet access. A built-in browser lets you access the Internet just as you would from your home laptop. The main limitation is the small screen size, which prevents you from doing as much as you would like, but you can still do it.

Email. Email is one of the key features of a smartphone. It lets you stay in touch without a PC or laptop.

Messaging. It goes without saying that text messaging is a main data feature.

Digital camera. Most smartphones have an excellent high-resolution camera built in. It is not as great as a stand-alone digital camera, but it does have 1 to 3 megapixels of resolution.

Camcorder. Some of the digital cameras have video capability as well. The main limitation is the amount of memory available to store videos.

MP3/iPod. A music player is a standard feature of many smart phones. The key here is to have sufficient flash memory to store the songs you want.

Wireless local area network (WLAN) access. You will learn more about this in the next chapter, but a smartphone has a built-in separate transceiver to connect to a Wi-Fi hot spot or other access point in airports, hotels, coffee shops, and other locations just like a laptop. You can access the Internet or email as an alternative to doing it through the cellular network.

Bluetooth. Bluetooth is another short-range wireless technology covered in Chapter 9. It is used in smartphones for wireless headsets. A headset contains a headphone and a microphone that fits in the ear. It talks to the cell phone wirelessly via a Bluetooth link.

GPS. The global positioning system is the worldwide satellite navigation system. Many phones have a built-in GPS receiver so that you can find your way. Some even offer full navigation software and display as in those personal navigation devices (PNDs) like those from Garmin and Tom Tom.

Femtocells

One of the most frustrating problems cell phone users have is poor coverage at home. Many homes are not located near cell sites or have poor indoor performance. This is maddening to many, especially those who want to give up their regular wired home phone and go only wireless. One solution to this problem is the femtocell. This is a small base station designed to be installed at home. The concept is illustrated in Figure 8.15. It can usually accommodate calls from up to four cell phones. The femtocell connects back to the carrier via your high-speed Internet connection, a DSL line, or a cable TV link.

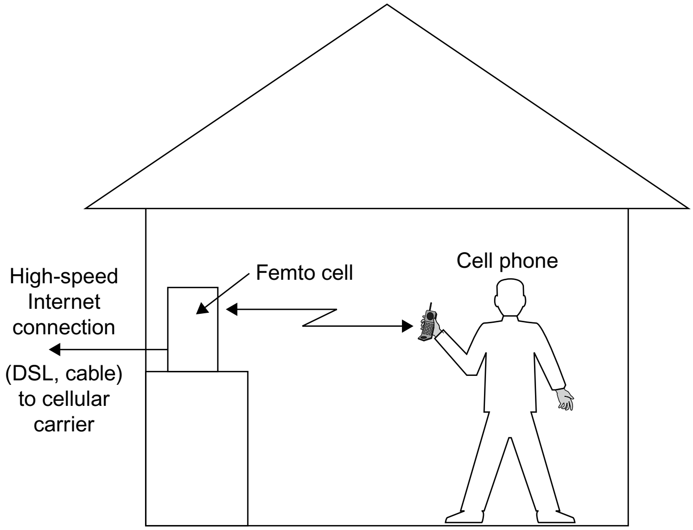
High-speed Internet connection
(DSL, cable) to cellular carrier
Femto cell
Cell phone

FIGURE 8.15 Concept of femto cell.

It provides great in-home coverage and has the added benefit that it offloads to the cell site and the carrier, providing more capacity to others.

Femtocells are just now becoming available but will soon be rolled out nationally. It is expected that the new LTE systems will make extensive use of femto cells.

Mobile TV

Video is now available on some phones. This video comes to you over the cell phone network, so it uses the high-speed data connection. Video is very data intensive, so it really loads down the network, limits subscriber capacity, and slows speed in some cases and raises your monthly bill. Yet, you can get various TV programs plus news, weather, sports, and comedy this way.

It has been decided that the best way to offer TV on a cell phone is to broadcast it directly to a small receiver in the cell phone. This keeps the video data overload off the cellular network and provides a better picture and more programming choices. This over the air (OTA) cell phone video is now available in many U.S. cities with more to come. A company and service called MediaFLO is transmitting TV to those who subscribe. Up to 15 or 20 channels provide a mix of programming. Some TV is also available in Europe using an OFDM system called Digital Video Broadcast-Handset (DVB-H). Variations are also available in Japan and Korea.

It is expected that some time in the near future, the U.S. digital TV system called the Advanced Television Systems Committee (ATSC) standard will be available in a mobile version that can be received via free OTA if the phone has a receiver built in. In all systems, the video is highly compressed using a standard referred to as H.264 and reformatted for a 360 × 240 pixel screen.

Location Technology

Location technology refers to systems and circuits that are designed to pinpoint the location of a cell phone in use. This can be done now to a certain extent, as a carrier can always determine which cell site a phone is currently using. That only locates the phone within an area of a few miles. The FCC has mandated that all cell phones can be located within 50 to 100 feet in case of an emergency. This is part of the E911 system for all cell phones. E911 means that your cell phone company has a system that will send your physical location to emergency services if you call 911 from your cell phone. Some systems use GPS, but others have another type of location technology. The whole thing is transparent to a user, and most users actually do not know that their phone has this capability.

The location capability has many concerned for privacy reasons. But if you don't want to be located, turn off your phone. Otherwise, learn to live with it, as it many save your life one day. Location technology is also supposed to lead to services that are based on location, such as advertisements for restaurants and other businesses near you.

When GPS is used, the GPS receiver computes your coordinates from the satellite signals and sends them to the carrier for use if called for. In another system, your location is determined by actually triangulating your phone with three nearby cell sites. No additional electronics are needed inside the phone, but special antennas, receivers, and other gear are needed at the cell sites. Both systems can usually pinpoint you within 50 feet or so depending on terrain and environment.

Project 8.1

Examine a Base Station

Go look at a cell site. The wireless phone company won't let you inside, but you can drive close to many and examine what you can see. Especially note the antenna arrangement at the top of the tower. Most have a triangular arrangement of three antenna arrays that divide the coverage into three 120-degree wide areas. You will see multiple coax cables going up the tower to carry the received and transmitted signals. Check for a dish that connects the site to the main switching station.

Project 8.2

Check Out the Latest Handsets, Features, and Services

Go to a cell phone store and ask about what digital services are available in your area. Try to find out what the technology is, such as GPRS, EDGE, cdma2000, 1xRTT, WCDMA, HSPA, and so on. Do you have email, instant messaging, games, Internet access, transmission of color digital photos, or what?

Project 8.3

Get an Area Coverage Map

Call your cellular carrier and ask for a coverage map. Some carriers actually give you this, but many times you have to ask for it. It may also be on your carrier website. The map will show you the actual range and coverage of the system. If you leave the area, you may have no coverage or you may switch over to another system in a roam mode.

Project 8.4

ID the Carriers

Identify the cell carriers (ATT Wireless, Cingular, Verizon, Sprint, Alltel, MetroPCS, etc.) in your area. If they advertise locally in the newspapers or in cell phone stores, they cover your area.

Project 8.5

Relevant Websites

There are several great websites where you can get additional information on how cell phones work. Here are a couple that I like: www.howstuffworks.com and www.sss-mag.com.

Chapter 9

Networking: Wired and Wireless

All Devices Talking to One Another

In this Chapter:

- Types of networks defined: WAN, MAN, LAN, and PAN.
- Ethernet, Sonet, and other wired networking technologies.
- Internet transmission fundamentals.
- Wireless networking technologies, including Wi-Fi, Bluetooth, WiMAX, ZigBee, and others.

INTRODUCTION

You already know from reading this book that virtually every electronic product has a digital computer built into it in the form of an embedded controller or microcomputer. But one other fact about electronic equipment that you may not know is that a large percentage of these devices are now being interconnected to one another to form networks. This is particularly true of computers. After all, what computer today is not connected to some form of network? And while we may be nearing the point where all computers are networked, the trend continues toward networking many other non-computer devices. This chapter takes a look at networking concepts and the most popular forms of networks being used today, both wired and wireless.

IS EVERYTHING NETWORKED?

A network is just a system of people or things that are connected to one another. If someone was to ask you to give an example of a network, you probably would use the computer network definition. But there are all sorts of other networks in which we all participate in some form or another. One of the biggest networks is our interconnecting road system. Our highways and byways interconnect cities with one another and provide transportation routes for people and goods. The railroad is a massive network. You could also call the airline system a massive network of cities connected by airplane flights. Our whole electrical grid is a massive network. And one of the largest networks in

© 2010 Elsevier Inc. All rights reserved.

doi: 10.1016/B978-1-85617-700-9.00009-6

the world is the telephone system, both wired and wireless. And then again, let's not forget the Internet itself which connects us all to everything else. For our purposes here, our network definition is a system designed for communications via computers talking to one another or individuals talking to one another over cell phones. In all cases, a network is involved.

TYPES OF NETWORKS

There are lots of different ways to interconnect individual computers or other electronic products. You can do it by electrical cable or fiber optic cable or you can do it in a wireless manner. In any case, each computer or user in the network is called a *node*. The goal of a network is to connect each node to all other nodes in some way so that they can communicate. Over the years, several different levels of networks have evolved.

Wide Area Networks

A wide area network (WAN) is the largest network of all and we could consider the entire telephone network or the Internet as a large WAN. Sometimes WANs are smaller and may be localized to a country, a state, a city, or other large geographical area. The Internet is certainly a WAN. Most WANs are made up of fiber optic cable so that they can carry lots of data at very high speeds.

Metropolitan Area Networks

A metropolitan area network (MAN) is a network that covers a smaller geographical area such as a city or a large college campus system. Good examples of MANs are the local telephone company and your local cable TV company. MANs are also widely used in large companies and in governments to interconnect their computers. Most MANs are also implemented with fiber optic cable to maintain speed and data capacity.

Local Area Networks

A local area network (LAN) covers a small group of computers, typically a thousand computers or less. LANs are implemented in companies, small businesses, and even in homes. A LAN may only have a couple of users but could have several hundred, depending on the circumstances. Even two connected PCs at home comprise a LAN. In most cases, LANs are connected externally to a MAN or a WAN for external communications.

Personal Area Networks

Most personal area networks (PANs) are short-range wireless networks. A PAN is formed when two or more computers or cell phones interconnect to one another wirelessly over a short range, typically less than about 30 feet. Most PANs are what we call ad hoc networks that are only set up temporarily for a specific purpose. The Bluetooth headset option on a cell phone is an example of a PAN.

Home Area Networks

A home area network (HAN) is a network inside the home used to provide monitoring and control over energy usage. It connects to the utility's electric meter and monitors energy usage so the home owner can see what energy is being used and where. It is also used to give the utility a way to control the heating and air conditioning to save energy. The HAN also provides a way to conveniently turn appliances off and on. The HAN may use wireless or communications over the AC power line. The HAN is the home portion of the national Smart Grid effort to conserve energy.

Storage Area Networks

A storage area network (SAN), as its name implies, is a network that interconnects large data storage devices to computers and to one another. Large arrays of hard disks are usually set up to store massive amounts of data needed by business, the government, and other organizations. The SAN provides a way to access all this information in a fast and easy way.

Network Relationships

As it turns out, the different types of networks are connected to one another, the result being one massive network. Figure 9.1 is an example. For example, local

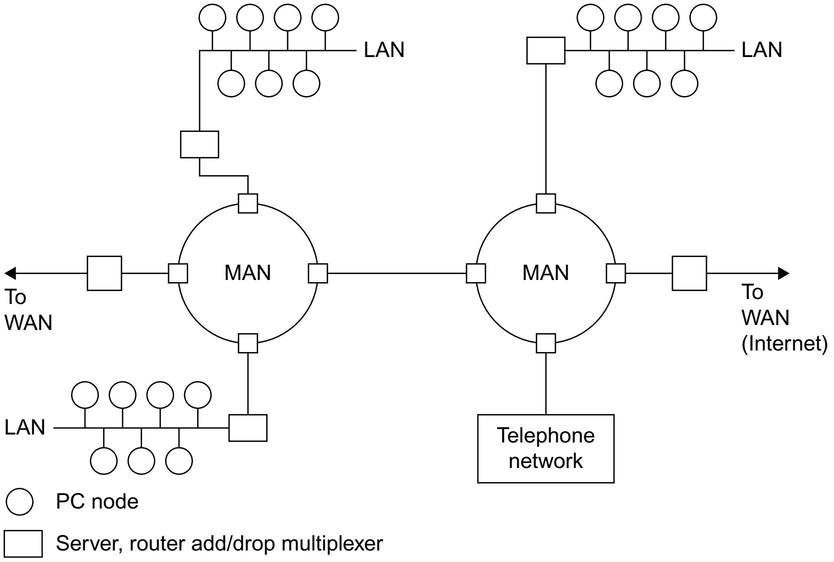

FIGURE 9.1 How LANs, MANs, and WANs connect to form a huge worldwide network of computers and other devices.

LANs are connected to MANs, while the MANs themselves are connected to WANs, and so on. In this figure, the circles represent PCs while the squares represent network connecting points. These include servers, the computers that manage LANs and MANs; routers, which determine connections in the network; and add/drop multiplexers that put data on the network and take it off. All the interconnecting links are fiber optic cable except for the copper cable in the LANs. This complex hierarchy essentially lets any computer talk to any other computer under the right conditions. The Internet implements such a hierarchy of networks.

NETWORK INTERCONNECTION METHODS

As you can imagine, there are lots of different ways to connect computers or other electronic devices to one another. For example, you could run a cable between each of the computers you want to interconnect. The result would be a large, complex, and very messy and expensive interconnection system. It is called a *mesh*. Except for a few computers, such systems are not used. Instead, over the years, several different interconnection methods called network topologies have emerged. These are the star, the ring, and the bus, all simply illustrated in Figure 9.2.

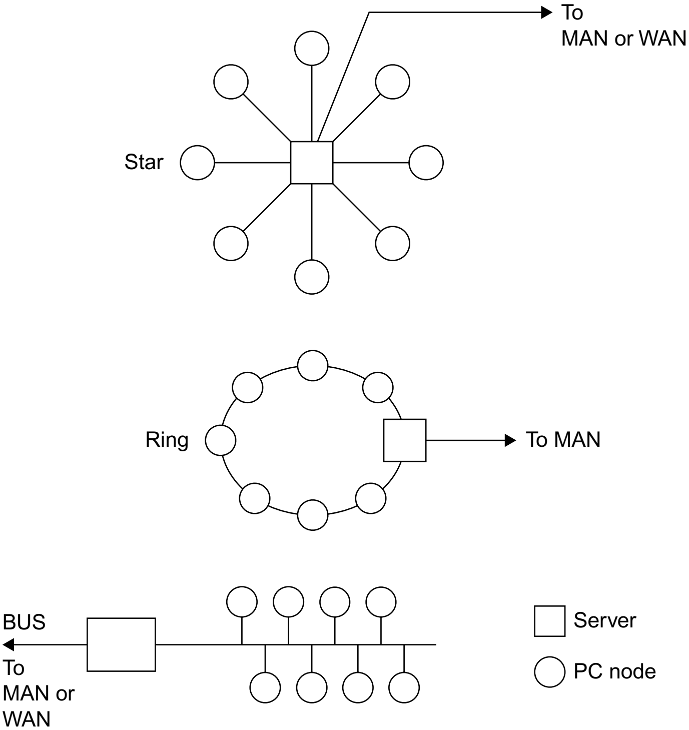

FIGURE 9.2 Three main networking topologies—star, ring, and bus. Bus is the most widely used.

In the *star* method, each node is connected to a central control computer or server that manages the interconnections from any PC to any other. In the *ring* interconnection, the computers are simply connected into a single closed loop. Then data is transferred from one computer to the next by sending it around the ring. The computer destined for the information will recognize it and grab it off the ring as it comes by. The *bus* method of interconnection is one of the most popular, as it is simply a common cable or connecting point to which all computers are attached. The problem with the bus is that there can only be one sender or transmitting party of the information on the bus at a time. This requires some method of managing who transmits to whom and when. All nodes can receive.

The mesh network mentioned earlier is a messy thing to implement with cables. However, it can be implemented wirelessly. In a mesh, all nodes can talk to all other nodes within their range. This is advantageous in that there are multiple paths for data to take if one path should be blocked for some reason. More on that later.

WIRED NETWORKING TECHNOLOGIES

Wired networks use three types of cable: twisted pair, coax, or fiber optic. The telephone system has used twisted-pair cable from the very beginning and it is still in use today. It is also the predominant way that LANs are interconnected. Coax cable was once popular but is not widely used today except in cable TV systems. Fiber optic cables were not practical until the 1970s, but since then have been developed into the best way to wire computer networks because of their high speed and low loss.

Types of Cables

Coax cable was the original networking medium (see Figure 9.3). It has a copper wire center conductor surrounded by an insulator like Teflon or another type of plastic, which in turn is usually covered with a wire-mesh braid.

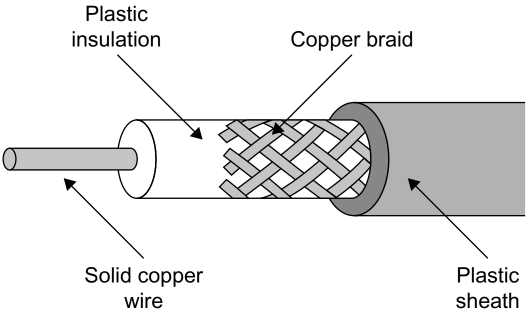

FIGURE 9.3 Coax cable. It is now mainly used in cable TV networks. The most common is RG-6/U and uses F-type connectors.

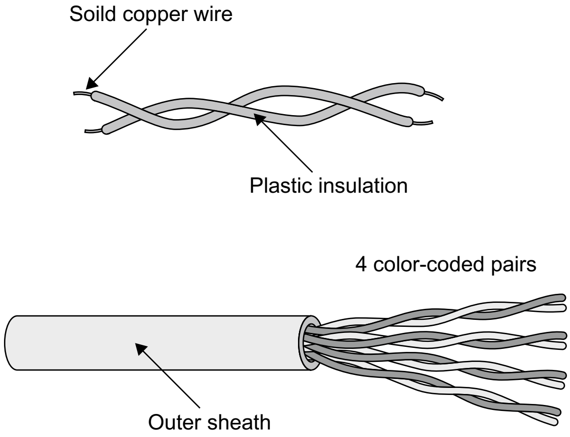

FIGURE 9.4 Twisted-pair cable. Initially used for telephones and is still used. Computer network cables use four twisted pairs. A shielded version is available where noise is a problem.

TABLE 9.1 Types of Twisted-Pair Cable

Category	Maximum Data Rate	Application/Comment
CAT5	100 Mbps	Ethernet LAN, telephones
CAT5e	1 Gbps	Ethernet LAN
CAT6	10 Gbps	Ethernet LAN, video
CAT7	10 Gbps	LAN, point to point

A plastic outer jacket covers the whole thing. The center wire and the mesh are the two connection points. Coax can handle frequencies up to many gigahertz. But it has high losses over long distances. Coax is mainly used in cable TV networks.

Twisted pair is the widely used cable for telephones and LANs. It consists of two solid insulated wires twisted together as shown in Figure 9.4. Most telephone or LAN cables contain multiple pairs. Typical LAN cables have four pairs in a single cable. These cables are standardized and categorized as shown in Table 9.1. CAT5 and CAT5e are the most common. The eight wires are usually terminated in a connector called an RJ-45.

Fiber optic cable consists of a thin glass or plastic center piece surrounded by a plastic outer cover. Then that is further wrapped by an outer plastic covering (see Figure 9.5). Light waves are passed down the center glass or plastic cable. Most light comes from an infrared (IR) light laser. You cannot see IR but it is light nonetheless. IR light is what your TV remote control uses. The IR light is pulsed off and on to create binary signals. These are picked up

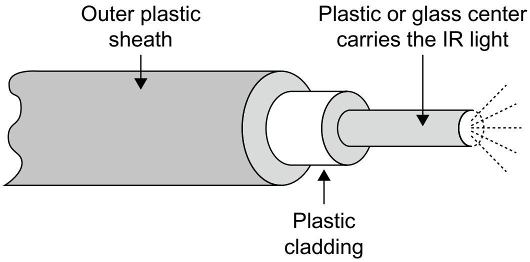

FIGURE 9.5 Fiber optic cable. It is nonconductive because it is made of plastic and glass, and is used to transmit data as infrared (IR) light pulses.

at the receiving end by a light detector and converted back into binary 0 and 1 voltages (see Figure 9.6).

Over the years, dozens of different wired networks have been developed. But in the past two decades, only two types of wired networks are dominant: Ethernet and Sonet.

Ethernet

Ethernet was originally developed as a LAN in the 1970s. It used coax cable in a bus topology to interconnect multiple computers. Ethernet is also known by its Institute of Electrical and Electronic Engineers (IEEE) standard designation 802.3. There are many variations of this technology, but the most common one uses unshielded twisted-pair cable. The bus topology is generally implemented inside a piece of equipment called a *hub* or *switch* (see Figure 9.7). The cable from each computer or other device is connected to one of the jacks on the hub or switch. The hub or switch connects to a master computer called a *server* that manages the network. Generally speaking, connections to the hub or switch are limited in length to about 100 meters or 300 feet.

The initial Ethernet LAN had a maximum data rate of 2.93 Mbps over coax. When Ethernet was standardized, the basic data rate was 10 Mbps. Over the years, Ethernet has been improved and developed to create versions with data speeds over twisted-pair cable of 100 Mbps, 1 Gbps, and 10 Gbps. While 10 Mbps can be implemented over a single twisted pair, to achieve speeds of 100 Mbps, 1 Gbps, and 10 Gbps, all four twisted pairs in the CAT5 cable are used. The data is divided between and transmitted in parallel at the same time using special multilevel encoding that permits such higher data rates to be achieved.

There are also versions of Ethernet that use fiber optic cable. Fiber optic cable is normally used on the 1-Gbps and 10-Gbps versions. The latest versions of Ethernet provide 40-Gbps and 100-Gbps data rates.

How Is Ethernet Used?

Ethernet uses a bus topology connection where only one node can transmit at a time. This requires the use of an access method that permits the nodes on

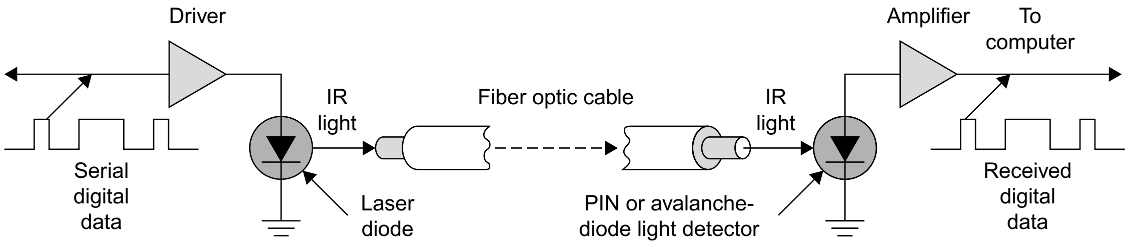
Driver
Serial digital data
IR light
Laser diode
Fiber optic cable
IR light
PIN or avalanche-diode light detector
Amplifier
To computer
Received digital data

FIGURE 9.6 How binary data is transmitted over fiber optic cable.

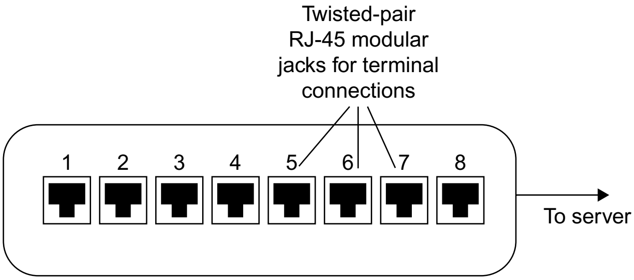

FIGURE 9.7 Ethernet hub or switch implements a bus that is accessed by RJ-45 connectors and twisted-pair cable.

Preamble	Start of frame delimiter	Destination address	Source address	Length	Data	Frame check sequence
7 bytes	1 byte	6 bytes	6 bytes	2 bytes	46–1500 bytes	4 bytes

FIGURE 9.8 Ethernet frame or packet.

the bus to share it. Ethernet uses a technique known as *carrier sense multiple access with collision detection* (CSMA/CD). This is part of the IEEE 802.3 standard. Essentially what happens is that if a computer wants to transmit, it monitors the bus, and if no one is transmitting, it will begin transmitting its signal. If someone else tries to transmit at the same time, a collision occurs and both computers stop transmitting. Each will wait a random period of time and then try again. Whoever captures the bus first will be allowed to transmit. The data is transmitted over the bus and any computer connected to it can receive it. However, typically the transmission is only intended for one other computer. To identify that computer, each is given an Ethernet address that is transmitted along with the data to be sent.

The data to be transmitted are packaged together in a group of bits called a packet. Figure 9.8 shows what a typical Ethernet packet looks like. The preamble consists of 7 bytes of alternating 1's and 0's that help the receiving computer establish clock synchronization with the transmitting computer. The start frame delimiter presents a unique 1-byte code indicating that a frame of Ethernet data is to be transmitted. Next, a 6-byte destination address identifies the desired receiving node. The next 6 bytes identify the sending node. This is followed by another 2 bytes of data that indicate how long the data message is. The block of data is then transmitted. It can be any length from about 46 to 1500 bytes. The packet ends with a frame check sequence that is essentially a unique code designed for error detection and correction. If bit errors occur during transmission, the receiving computer can perform specific logic functions that allow the error to be detected and then corrected.

TABLE 9.2 Sonet/SDH Designations and Data Rates

Sonet Level	Data Rate
OC-1	51.84 Mbps
OC-3	155.52 Mbps
OC-12	622.08 Mbps
OC-48	2.488 Gbps
OC-192	9.953 Gbps
OC-768	39.812 Gbps

While Ethernet started out as a LAN, it has morphed into other larger forms of networks. Because of its high-speed capability with fiber optic cable, it is also used to form MANs and even WANs. A special version for MANs and WANs is called *Carrier Ethernet*. There is also a version of Ethernet used in storage area networks, called *Internet Small Systems Computer Interface* or *iSCSI*. There is also a wireless version of Ethernet that will be discussed in another section in this chapter.

Sonet

Sonet means *synchronous optical network*. Also known by its international standard name, *synchronous digital hierarchy*, this fiber optic network system was designed specifically for high-speed data transmission in telephone networks and computer systems. Basically, Sonet defines how data is transmitted over the fiber optic cables. This includes the data format as well as data speeds. Table 9.2 shows the various standardized transmission speeds from 51.84 to 39.812 Gbps. The bulk of Sonet installations, which are used primarily in the telephone networks and the Internet, use 2.488 Gbps (also generally called 2.5 Gbps) and 9.953 Gbps (generally referred to as 10-Gbps Sonet). More recently, new 39.812 (40) Gbps networks are being built for the Internet. Data is transmitted in fixed-length frames made up of 810 bytes, most of which is data. The network topology is either point to point or i n a ring.

Most big WANs and MANs use the Sonet system. However, Sonet is often used to transmit other data, including the telephone company T1 and T3 line data, asynchronous transfer mode (ATM), frame relay, and other telephone company formats. It is also widely used to carry Ethernet packets as well as the unique packets that are part of the Internet protocol.

Telephone Data Systems

When the wired telephone system converted to digital many years ago, several key data transmission standards were developed. The most common are the T carrier systems, as well as frame relay and asynchronous transfer mode (ATM). The most widely used digital telephone networking standard is the T-1 system. It was developed to transmit 24 digitized telephone lines over a single cable. The voice signals were digitized into 8-bit chunks and then transmitted sequentially over a T-1 line from one central office to another. The basic T-1 frame format is shown in Figure 9.9. It consists of 192 data bits made up of 24 bytes representing individual chunks of telephone calls. An additional synchronizing bit is added. The result is a T-1 signal that transmits at a rate of 1.544 Mbps. And while T-1 is still widely used for digital telephone calls, it is also used for broadband computer data connections and cell phone backhaul.

There are other T carrier systems that are used to multiplex T-1 lines onto larger cables. The most common is the T-3, which allows dozens of T-1 lines to be multiplexed on a single cable at a 44.736-Mbps data rate.

Frame relay transmits data mostly in telephone systems. Both the T carrier and frame relay are synchronous systems, meaning that they are synchronized to a master clock signal.

Asynchronous transfer mode (ATM), as its name implies, is an asynchronous system that transmits data in packets. Each packet is made up of 53 bytes—5 bytes for a header that designates the destination and a 48-byte payload segment. Packets may be digital voice or computer data.

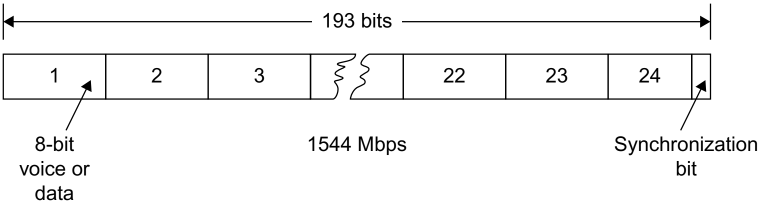

FIGURE 9.9 A T-1 data frame commonly used in digital telephone systems as well as data interconnections in networks.

Optical Transport Network

The optical transport network (OTN) is a technology used to implement the Internet backbone network. This is the core long haul fiber optical network that connects the world together. OTN uses special polarization and modulation techniques to transmit data rates of 40-Gbps and 100-Gbps.

HOW THE INTERNET WORKS

The Internet is an enormous WAN covering the entire world. It consists of many super-fast, fiber optic cable links that are generally referred to as the backbone of

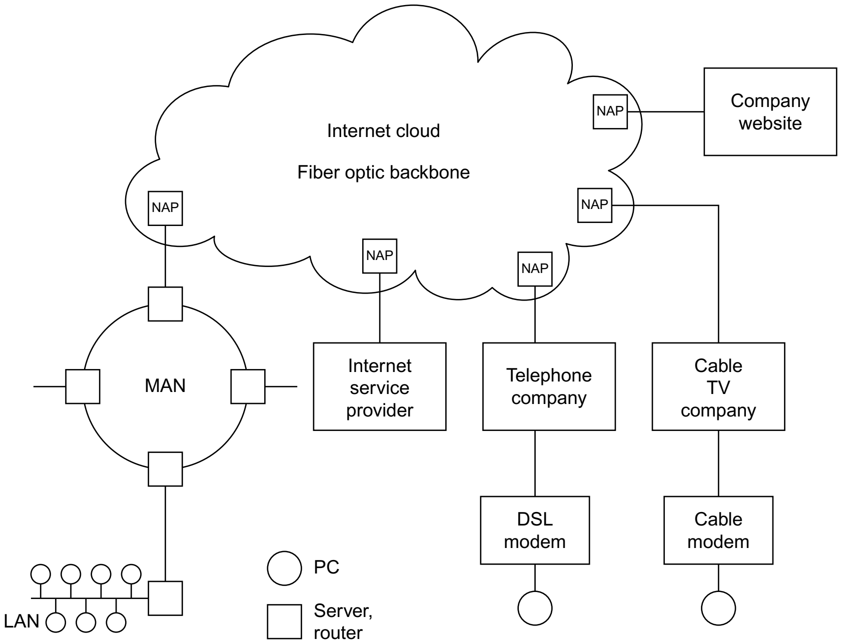

FIGURE 9.10 Internet illustrated as cloud of fiber backbone connections through network access points (NAP) that attach to WANs, MANs, and LANs.

the network. These high-speed links are owned by companies and governments and provide connection points for companies, governments, WANs, Internet service providers, and hosts for websites. These connection points are known as network access points (NAPs). See Figure 9.10. There are dozens of NAPs in the United States and they provide the link to the backbone connections. These backbone connections are worldwide. Because they are complex and interlinked, they are often referred to as the Internet cloud. Since data being transmitted from one point to the other can go by multiple paths in the backbone, there is no way to know the exact path, so the cloud designation is appropriate. You put the data transmission into one NAP entry point to the cloud and it comes out in another. From there it goes to other MANs and LANs for connections to other networks and to individual users.

Data is transmitted over the Internet using a standardized method referred to as TCP/IP. This means transmission control protocol/Internet protocol. Both are protocols for building packets that are transmitted over the Internet.

Assume that a LAN and MAN using Ethernet want to transmit information over the Internet. The Ethernet packets are then sent to a NAP where they are assembled into the packets that actually get transmitted over the fiber optic cable. The TCP assembles the data to be transmitted into unique packets and attaches a header to it, giving the source and destination addresses, a sequence

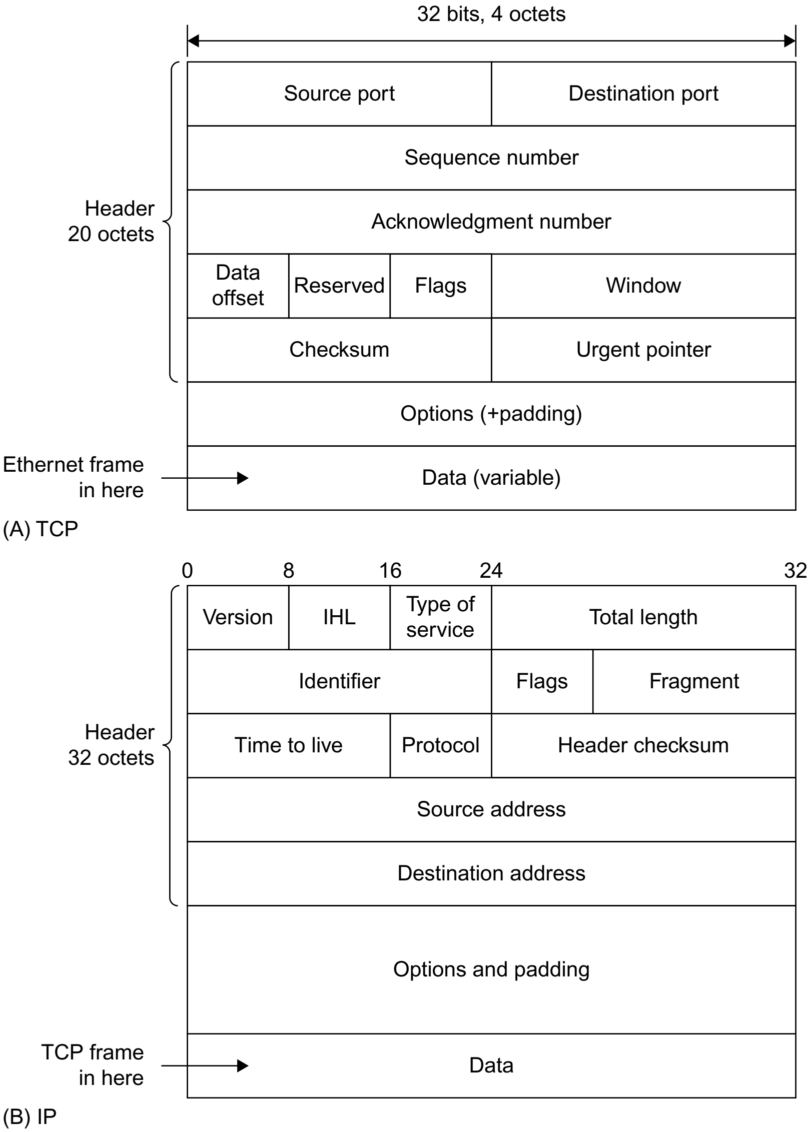
32 bits, 4 octets
Header
20 octets
Source port
Destination port
Sequence number
Acknowledgment number
Data offset
Reserved
Flags
Window
Checksum
Urgent pointer
Options (+padding)
Ethernet frame in here
Data (variable)
(A) TCP
0
8
16
24
32
Header
32 octets
Version
IHL
Type of service
Total length
Identifier
Flags
Fragment
Time to live
Protocol
Header checksum
Source address
Destination address
Options and padding
TCP frame in here
Data
(B) IP

FIGURE 9.11 Protocols used in Internet. **(A)** Transmission control protocol (TCP). **(B)** Internet protocol (IP). The data in each frame is transmitted from left to right and top to bottom.

number, and some error detection and correction information. The general format is shown in Figure 9.11A.

The TCP packets then become the data that is assembled further into Internet packets (see Figure 9.11B). The Internet protocol attaches a header giving source and destination addresses as well as other information about where and how the data is to be sent.

The IP header contains a 32-bit destination address identifying the location of the computer receiving the data. This 32-bit address is divided into four 8-bit segments. Each 8-bit segment is usually referred to by its binary equivalent. The result is the IP address. Every computer and most other devices have an IP address assigned by an organization set up for assigning and managing Internet addresses. You may have seen the address referred to in the format 51.78.23.189. This is called the *dotted decimal format*.

With a 32-bit address, over 4 billion individual computers and other devices can be assigned a unique address. However, we are already beginning to run out of addresses and a new updated form of Internet protocol called IPv6 (Internet Protocol version 6) is gradually replacing it. It has a 128-bit address that provides billions and trillions of additional addresses so that even the smallest inconsequential device can have its own Internet address.

As the data is transmitted over the Internet, it passes through many devices referred to as *routers*. The router is the basic transmission control device of the Internet. What the router does is examine all of the packets transmitted to it and specifically looks at their IP destination addresses. The router temporarily will store data being transmitted and then decide which part of the network it should be connected to next to reach its desired destination. Routers store sets of information called *routing tables* that help the router decide the best transmission path through the Internet.

The data being transmitted ultimately reaches the NAP connected to the network where the destination exists. This may be a company, MAN, or LAN, or it could be an Internet service provider, such as the telephone company or cable TV company that will ultimately connect to the source computer. The source computer then strips away the TCP/IP headers to recover the original Sonet or Ethernet data.

WIRELESS NETWORKS

Wireless networks are usually LANs or PANs, but may also be MANs. Most wireless networks are short-range wireless connections that bring greater mobility as well as freedom from interconnecting cables to networking. Figure 9.12 shows a graph designating the various types of wireless networking schemes in use today. In this graph, each networking technology is defined in terms of its data rate as well as its range. In this section, we take a look at all of these technologies, most of which you have heard of or already use on a regular basis.

Wi-Fi

The most widely used wireless networking technology is referred to by its trade name Wi-Fi, meaning *wireless fidelity*. This the name given to the IEEE's

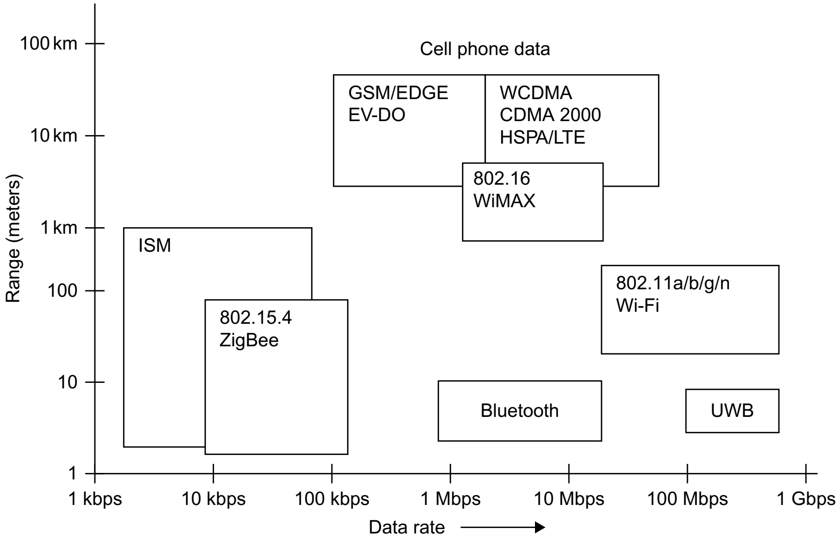

FIGURE 9.12 Summary of short-range wireless networking technologies shown as data rate versus range.

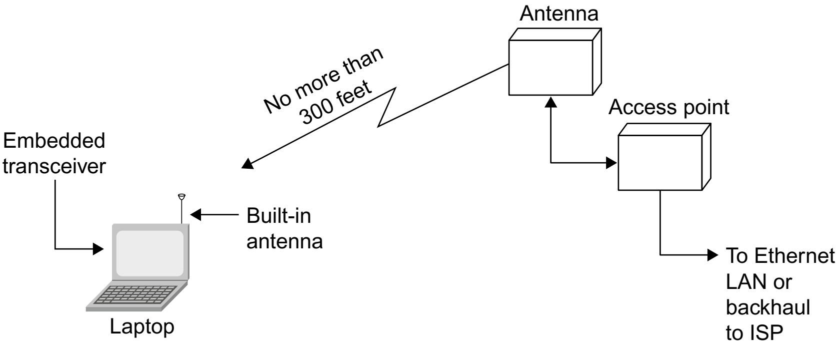
Antenna
No more than 300 feet
Access point
Embedded transceiver
Built-in antenna
Laptop
To Ethernet LAN or backhaul to ISP

FIGURE 9.13 How a Wi-Fi access point or hot spot works.

wireless Ethernet standard designated as 802.11. Wi-Fi was originally developed as a wireless extension to normal Ethernet LANs. It allowed laptops and personal computers and even cell phones to connect to a company or organizational LAN wirelessly instead of the normal CAT5 connection. This not only simplified and in many case eliminated new wiring but also gave employees greater freedom and mobility, particularly since most employees use laptop computers rather than fixed PCs.

Figure 9.13 shows how Wi-Fi works. Each PC, laptop, or cell phone has a built-in data transceiver and antenna that communicate with a wireless access point (AP). This wireless access point is a transceiver that connects directly to an Ethernet switch, which, in turn, connects to the LAN wiring and its server. Access points are placed at strategic locations within buildings to allow as many people as possible to connect wirelessly.

There are also public access points known as *hot spots*. These are present in coffee shops, airports, hotels, and in many other locations, allowing anyone with a laptop or cell phone to connect. These hot spots talk back to a central point by a backhaul method, such as a T1 telephone line or microwave relay link that connects to an Internet service provider that, in turn, is connected to the Internet.

The 802.11 standard exists in several basic forms. The most popular version of this standard is 802.11b. It operates in the 2.4- to 2.483-GHz unlicensed band. It uses the direct sequence spread spectrum (DSSS) and is capable of data rates up to 11 Mbps in a range of up to 100 meters. This frequency band is divided into 11 channels, each 22 MHz wide. The computer and access point decide on which channel to use to minimize interference with other transmissions. If you are close enough to the access point, you can actually get the 11-Mbps data rate, but if you are not, the system automatically ratchets down to 5.5, 2, or even 1 bps depending on the range, interference, and noise or other environmental conditions.

Another popular version of the standard is 802.11g. It is a faster version of Wi-Fi that also operates in the 2.4-GHz band. It can achieve data rates as high as 54 Mbps up to 100 meters. It uses OFDM with 52 subcarriers. The most recent standard is 802.11n, which uses OFDM and MIMO to achieve even higher data rates up to approximately 600 Mbps. Another less popular version is 802.11a, which uses the 5-GHz unlicensed band. It too uses OFDM and can achieve data rates to 54 Mbps. The IEEE is also currently working on an advanced wireless standard that should ultimately be able to achieve a 1-Gbps data rate in the near future.

Wi-Fi is also built into most of the new smartphones. While a smartphone has data capability because of its 3G digital technology, Wi-Fi gives it one more way to access the Internet and email.

Wi-Fi is also the most popular home networking technology. With no wires to install, it is easy to set up and use with a broadband DSL connection or cable TV access.

Bluetooth

Bluetooth was developed as a personal area network (PAN) for connecting computers, cell phones, and other devices up to a range of about 30 feet. It is a low-speed data transmission method. Bluetooth operates in the same 2.4- to 2.483-GHz unlicensed spectrum as Wi-Fi. It uses a technique known as frequency hopping spread spectrum (FHSS), where the data is divided into chunks and transmitted via a carrier that hops from one random frequency to another. Data is transmitted at a 1-Mbps rate using FSK. An enhanced data rate (EDR) version of Bluetooth is also available to transmit at higher speeds up to 3 Mbps. Even newer versions can give data rates as high as 24 Mbps and even higher rates are possible in future versions of the standard. The standard is managed by the Bluetooth Special Interest Group (SIG).

One of the basic features of Bluetooth is that it is capable of forming small networks called piconets. It does this by linking two Bluetooth devices together. One serves as a master controller and it can connect to seven other Bluetooth slave devices. Once the PAN has been set up, the various connected devices can exchange information with one another through the master.

By far the most common application for Bluetooth is cordless headsets for cell phones. But you will also encounter it in some wireless connections between laptops and cell phones. It is also used in some computer peripheral devices such as PC-to-printer connections.

ZigBee

ZigBee is another short-range PAN network technology with the IEEE designation 802.15.4. It uses low power, so the range is typically 100 meters or less, basically depending on the antennas and the physical environment where it is used. ZigBee also operates in that unlicensed spectrum from 2.4 to 2.483 GHz.

The IEEE 802.15.4 standard defines the basic radio technology. It uses direct-frequency spread spectrum (DSSS) and a version of QPSK that gives a data rate of 250 Kbps. The frequency spectrum is divided into 16 1-MHz channels.

The basic function for ZigBee is monitoring and control. Monitoring refers to telemetry that allows sensors to collect information, which is then transmitted by radio to some central collection point. It also allows for remote control of the devices such as lights, motors, and other items. The ZigBee Alliance, the organization that promotes and enhances this wireless technology, has developed a series of applications for industrial monitoring and control, home monitoring and control, energy monitoring (automatic meter reading) and control, and several others.

A major feature of the ZigBee technology is its ability to perform mesh networks. While a typical ZigBee radio node may only have a range of 30 meters or less, the range can be extended by simply transmitting data from one node to another in a mesh network that may cover hundreds of meters, or in some cases even miles. A mesh is shown in Figure 9.14. Each circuit is a ZigBee transceiver node and each line is a wireless path. Mesh networks have many different paths for data and, therefore, are extremely reliable. If one path is blocked or disabled, data can usually find its way to the desired destination by way of another path. For example, in Figure 9.14, a direct path may be from A-D-F-I-K, but if the D-to-F link is blocked or the I node is disabled, the mesh would automatically reroute the data through another path such as A-D-G-J-K.

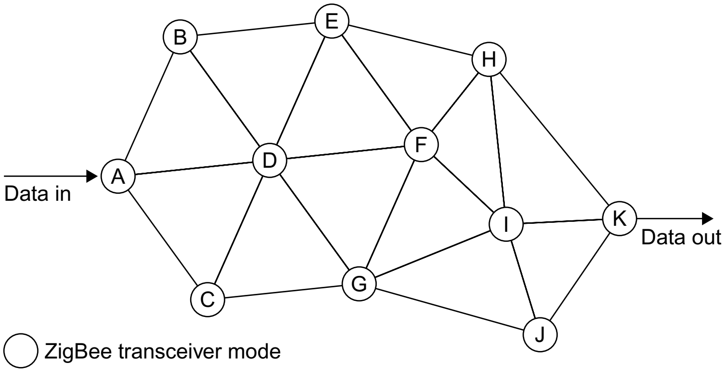

FIGURE 9.14 A mesh network can extend the overall range of a node and provide greater reliability through multiple possible paths. Most wireless mesh networks are ZigBee, but other wireless technologies may be used (e.g., Wi-Fi).

A forthcoming application of the 802.15.4 standard is the replacement of the current infrared (IR) TV remote controls. An RF version called RF4CE is gradually expected to replace IR remotes in all consumer electronic equipment providing longer range and no line of sight limitations. Two-way communications are expected to be implemented.

Ultra Wideband

Ultra wideband (UWB) technology is a very broadband, high-speed PAN designed for transmitting audio, video, and high-speed data. It uses OFDM in the frequency spectrum from 3.1 GHz to roughly 5 GHz. Higher frequencies up to 10.6 GHz are available but not widely used. Ultra wideband has a maximum data rate of 480 Mbps, but the actual rate depends on the range between nodes.

WiMAX

WiMAX is actually a wireless MAN. It was initially designed to offer a wireless alternative to Internet connections to homes, usually by way of cable TV systems or DSL lines of a telephone company. WiMAX is also standardized by the IEEE as 802.16. It comes in two variations, fixed and mobile. The fixed version is used for Internet connectivity in a fixed mode where the nodes in a star-like network are essentially in one position. WiMAX services are now available through an organization such as Clearwire in the United States with base stations with a range of several miles. Laptop computers containing WiMAX transceivers can connect to the network. Data rates will vary depending on the service offered by the company but can be from 1 Mbps to 10 Mbps or more. The technology uses OFDM with modulation schemes including BPSK, QPSK, 16-QAM, and 64-QAM. MIMO may or may not be used.

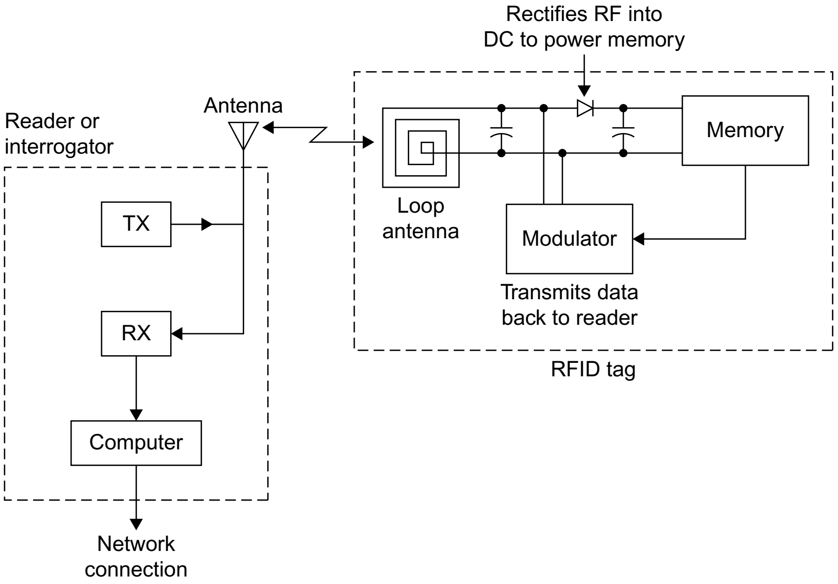
Rectifies RF into
DC to power memory
Antenna
Reader or
interrogator
TX
RX
Computer
Network
connection
Loop
antenna
Memory
Modulator
Transmits data
back to reader
RFID tag

FIGURE 9.15 How RFID works. The tag has no power of its own but gets power from the transmitter in the reader.

A mobile version of WiMAX is also available. It is designed for connectivity where laptops and other mobile devices can connect to access points or base stations even while in motion.

Most WiMAX activity takes place in the United States in the 2.5-GHz frequency range and the 3.5-GHz range in Europe, Canada, and Asia. It is expected that WiMAX will be used in the new 700-MHz band for high-speed Internet access.

Radio Frequency Identification and Near-Field Communications

Radio frequency identification (RFID) is a very short-range technology that is intended to substitute for standard bar codes. Practically everything has a printed bar code on it today so that with optical scanners the device can be recognized by a particular part number or other information that identifies it. The bar code readers are normally attached to a computer and, in turn, a network.

RFID devices are small thin plastic tags containing a radio chip that can be read by a wireless reader. The RFID chip contains an EEPROM memory in which is stored the device's unique identification number and other related information. A key characteristic of the RFID tag is that it contains no power source such as a battery. Yet it is designed to transmit its digital code to some external reading device whenever it is interrogated. Figure 9.15 shows how it works.

The interrogation takes place when an external reader made up of a transmitter (TX) sends a radio frequency (RF) signal to the nearby tag. The radio frequency energy received by the tag's antenna is then rectified and filtered into a DC voltage that is used to power up the memory and the modulator circuit that sends data back to the reader. The reader also contains a receiver (RX) that picks up the small signal from the tag and sends it on to a computer where identification takes place.

RFID tags are generally used for what is known as asset tracking. RFID tags are attached to any relatively expensive item whose position is to be monitored, tracked, or otherwise controlled. It is good for keeping track of capital equipment, baggage, manufacturing tools, and other items. It is also used for personnel security on badges that employees wear and is good for animal tracking. One of the most widely used applications is automatic toll collection on highways.

RFID tags come in a variety of sizes, shapes, and technologies. Some of the older tags work at a frequency of 125 kHz or 13.56 MHz. The newer tags work at UHF frequencies in the 902- to 928-MHz range. All of these tags are passive devices with no power other than that received from the reader. Some longer-range RFID tags with small batteries have also been developed.

The unique code stored in the EEPROM is stored during the tag's manufacture. It is then clocked out at a relatively slow rate of speed around 70 kbps. The serial binary data is used in a form of amplitude modulation called backscatter modulation that produces minor variations in the reader's signal that can be converted back into a binary code.

Another similar RFID-like technology is known as near-field communications (NFC). It has an even shorter range of only about 1 foot. It operates on 13.56 MHz. It is mainly designed to be used in credit cards and in cell phones to produce automatic transactions such as credit card purchases or payment for transportation (e.g., train, subway, bus) or access to some facility.

BROADBAND TECHNOLOGY

Whenever you hear the term "broadband," it generally refers to a high-speed interconnection or Internet access. Most broadband systems provide data rates from roughly 1 Mbps or less to well over 10 Mbps, depending on the level of service purchased. Such broadband services are usually provided by cable TV companies or telephone carriers. Other smaller organizations also offer various other types of Internet connectivity.

Cable TV Connectivity

Most cable TV companies offer high-speed Internet connectivity services. What they do is allocate one or more of the 6-MHz wide channels on the cable to this service. (See Chapter 11 on TV.) By using QPSK, 16-QAM, or 64-QAM, a wide range of customers can be offered multiple levels of Internet connectivity service. The customer gets a cable TV modem that connects to the cable along with the TV set. This modem provides the connectivity to a computer via Ethernet. A Wi-Fi access point is often connected to the cable modem to provide wireless connections to the Internet for PCs and laptops.

Digital Subscriber Lines

Digital subscriber lines (DSLs) are high-speed Internet connections that use standard telephone lines. DSL uses the standard twisted-pair telephone lines that come into every home for normal telephone service. These lines, because they were developed only for voice signals, are restricted in bandwidth and data rate. However, special techniques have been developed to allow very high-speed data transmissions on them. DSLs, also referred to as *asymmetric digital subscriber lines* (ADSL), use a variation of OFDM called *discrete multitone* (DMT). What it does is divide the restricted bandwidth of the twisted-pair cable into multiple OFDM channels, each 4 kHz wide. Then the data to be transmitted is divided into parallel paths and modulated using some form of QAM. All of this is handled by a DSL modem connected to subscribers' computers and the home telephone lines.

The speed potential of an ADSL line depends on how far away the subscriber is from the central office. The greater the distance, the lower the data rate. For even the longest runs from 12,000 to 18,000 feet, data rates of up to about 2 Mbps are possible.

Newer versions of ADSL have also been developed to permit data rates of up to about 12 Mbps at a range of 8000 feet and 20 Mbps at a range of about 4000 feet. A newer version referred to as video DSL or VDSL extends the bandwidth further and uses higher-level versions of QAM to get data rates of up to 52 Mbps.

Wireless Broadband

Wireless broadband services are only beginning to become available. A relatively new WiMAX wireless network known as XOHM is offered by Clearwire, a spinoff of Sprint Nextel. It provides fixed wireless data services with data rates in the 1- to 3-Mbps range. WiMAX modems are connected to computers that access nearby base stations within several miles of one another.

The beauty of wireless broadband is that no cables or other hard connectivity is required. These are easy to set up, although at microwave frequencies, range is a problem as are multipath problems with buildings, trees, and other obstructions. However, with high enough antennas and sufficient power, wireless broadband has proved to be practical. It is growing rapidly, particularly in those areas where DSL and cable TV services are not currently available. This is particularly true in rural areas and in some small towns and communities.

Don't forget that wireless Internet connectivity can also be achieved by using the regular cell phone networks. You can purchase a data modem that plugs into your laptop USB connector on a PC. This modem connects you just as it would a cell phone to your cellular network. From there you can use the 3G data service to connect to the Internet, access email, or perform other networking operations.

ISM-BAND RADIOS

The industrial-scientific-medical (ISM) band has been mentioned before. This comprises a group of frequencies or bands set aside for unlicensed data transmission over short ranges. The main application is telemetry monitoring and remote control. The U.S. frequencies are 13.56 MHz, 315 MHz, 433 MHz, and 902 to 928 MHz, with 915 MHz the most common. In Europe, 868 MHz is an unlicensed frequency. Transmitters and receivers are tiny single ICs with power levels from about 1 mW to 1 watt. Modulation is usually ASK/OOK or FSK. Data rates are low from a few kbps to usually no more than 100 kbps.

Some typical applications are garage door openers, remote keyless entry for vehicles, remote thermometers, medical telemetry, remote-controlled toy cars and boats, and sensor reading in industrial applications. Range is no more than a hundred feet or so, but longer ranges to several miles can be achieved with higher antennas.

Project 9.1

Become More Familiar with Wireless Data Protocols

You may want to find more information on the most popular short-range wireless data methods. The following websites are recommended. And you may want to do a general Internet search with Bing, Google, or Yahoo!.

Wi-Fi: www.wi-fi.org
Bluetooth: www.bluetooth.com
ZigBee: www.zigbee.org
General sites: www.palowireless.com, www.howstuffworks.com

The Wikipedia entries are also quite good.

Chapter 10

Audio Electronics

Digital Voice and Music Dominate

In this Chapter:

- Sound and the electronics that deliver it.
- Digital audio.
- The compact disc (CD).
- Voice and music electronics.

INTRODUCTION

While the telephone, telegraph, and radio are generally considered the first real electronic applications, audio was certainly next. By "audio," I mean sound: voice and music, things we can hear. Audio signals have frequencies within the range of the human ear. Over the years many electronic devices have been developed to amplify, capture, and reproduce sounds of all types. Today, like everything else in electronics, audio electronics is predominantly digital, although many analog devices and systems are still in use. This chapter is a summary of audio devices you know and use every day from stereo systems to iPods.

THE NATURE OF SOUND

Sound is just pressure waves in the air that your ear can hear. Anything that produces a noise or other disturbance produces pressure waves that travel to your ears where the sound waves are converted into signals that your brain recognizes as sound. The frequency range of these air vibrations is 20 Hz to 20 kHz. Our ears do not respond to lower or higher frequencies even though they may be present. Human hearing also varies widely due to age and health status. Age brings on hearing loss, especially of the higher frequencies.

There are four key things needed in audio electronics:

- The ability to convert sound waves into electrical signals.
- The means to convert the electrical signals into sound.

© 2010 Elsevier Inc. All rights reserved.
doi: 10.1016/B978-1-85617-700-9.00010-2

- A way to amplify the signals.
- Media to store or record the signals.

This is what audio electronics is all about.

Microphones

The microphone is the transducer that converts sound waves into an analog signal. The sound waves impinge on the microphone and it causes vibrations that are translated into a signal by either capacitive element or an inductive coil. A technical discussion about microphones is way beyond the scope of this book. In any case, the resulting signal is usually very small, in the millivolt range or even microvolt range. Therefore, it must usually be amplified before it is useful. This is the function of a small-signal amplifier usually called a *preamplifier*. What you do with the signal after that is up to you. Typically it will be amplified further for use in a public address (PA) or other sound system. Or alternately, the voice may be recorded. Many voice signals will be used in a telephone, cordless or cellular.

Speakers

A speaker is a transducer that converts the audio signal into sound waves that you can hear. You have obviously seen a speaker, so little explanation is needed. Its main element is a paper or plastic cone that vibrates at the audio frequencies applied to it. Then it produces the sound pressure waves that our ears will hear.

Figure 10.1 shows how a speaker works. The cone is attached to a form around which a coil of wire is wound. This is the so-called *voice coil*. That voice coil is positioned inside a strong permanent magnet with north and

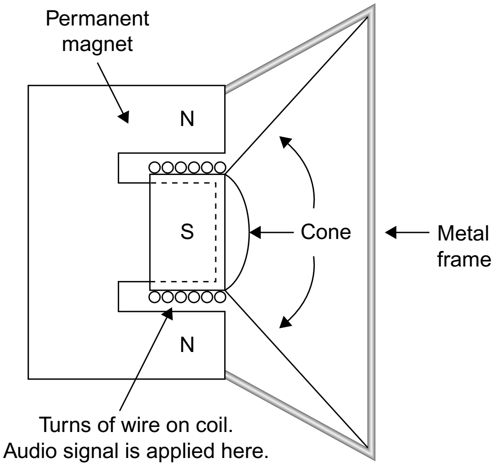

FIGURE 10.1 Cross-section of speaker showing major components and how it works.

south poles. Now when you apply the audio signal to the coil, current flows and causes the coil to become an electromagnet. The electromagnet generates alternating north and south poles as the voice or music signal occurs. These magnetic fields interact with the permanent magnet, causing attraction and repulsion of the voice coil, and the cone moves. The motion of the cone accurately reproduces the original sound waves.

Speakers come in a wide range of sizes and types. Cone speakers are usually called mid-range speakers in that they can reproduce most of the frequency range except for the very low and very high frequencies. Special speakers are used for those frequencies. A woofer or subwoofer is a larger speaker designed for frequencies of about 600 Hz and below. A tweeter is a special speaker optimized for the higher frequencies above 10 kHz. Special filters called *crossover networks* separate the signal into low, mid-, and high frequencies as shown in Figure 10.2.

The remainder of this chapter addresses the amplification and recording or storage of audio signals. Along the way, most analog audio signals are digitized. And many are also compressed to save storage space or reduce the binary data rate for transmission.

DIGITAL AUDIO

The sounds we hear and the sounds we generate are analog signals. A microphone picks up the voice and music and produces an analog signal to be amplified or stored. We have amplifiers and speakers that will accurately reproduce the

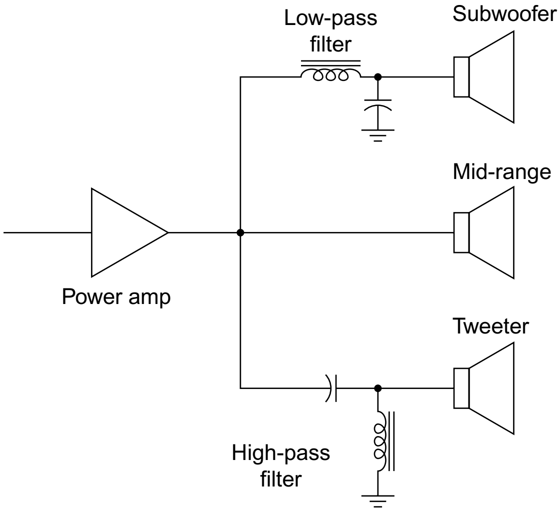

FIGURE 10.2 Cross-over networks keep signals of different frequencies separate for appropriate speakers.

sounds. But the real challenge with audio is capturing, storing, or recording the sounds for later reproduction. With today's digital techniques, we can store not only voice and music more accurately than before, but we can also transmit it more easily by radio, TV, or the Internet.

Past Recording Media

Edison invented the phonograph in the late 1800s. This was the first device ever conceived to record voice. A microphone converted the sound into vibrations that moved a sharp needle in direct accordance with the sound. The needle cut a varying groove in the wax coating on a rotating cylinder. Later, a needle placed in the groove moved the diaphragm in a horn speaker that translated the movement into sound. The quality was poor but it worked, and the concept was soon translated into what we know as a *phonograph record*. The record disc is rotated while a needle cuts a spiral groove into the plastic that accurately captures the sound. To reproduce the sound the disc is rotated at the same speed and a needle is placed in the groove to reproduce the sound in a speaker. A transducer converts the movement of the needle in the groove into an analog signal that is amplified.

Early records spun at a speed of 78 rpm and could only hold a few minutes of sound. Later a smaller 45-rpm record was invented to play minutes of music at a lower cost. Then a larger long-playing record spinning at only 33⅓ rpm was produced that could hold an hour or so of sound. Stereo or two channel sound was also effectively recorded into a V-shaped groove in the record.

Records were soon upstaged by magnetic tape. A plastic such as Mylar was coated with a powdered iron or ferrite magnetic material. The music or sound was then applied to a magnetic coil that put a magnetic pattern on the tape representing the sound. A coil passed near the tape would convert the magnetic pattern variations into a small voltage to recover the sound. Reel-to-reel tape was popular for a while, but magnetic tape became the recording media of choice for years when the Philips cassette tape was invented. Other formats like the ever-popular 8-track unit were also used.

All of these formats worked well but had their limitations. Record grooves wore out with usage causing the high-frequency response to be eroded. Scratches and dirt on the record also produced noise. As for tape, it had good frequency response but had its own noise that could not be reduced below a certain level. The dynamic range of both was poor.

Digital technology has solved all these problems. Now the most popular medium is the compact disc (CD), which records sound in full digital format. It has very low, practically undetectable noise, wide frequency response, and a very long life. The cost is also low.

Digital sound is also routinely stored in computer memories like flash EEPROMs and transmitted over the Internet. With digital techniques, sound can also be processed with DSP such as compression, filtering, and equalization.

Dynamic Range

Dynamic range refers to the difference between the highest and lowest volumes of sound that a system can handle. It is a measure of how wide an amplitude range a particular storage media can record or an audio system can accommodate. It is also a way to state the maximum sound pressures that our ears can withstand and the lowest levels that we can hear.

The upper level of amplitude is set by our ears and the level at which distortion occurs. The lowest levels of the dynamic range are determined by the noise in the system. Noise is the random voltage variation caused by thermal agitation of electronics in electronic components and random variations in the magnetism on a magnetic tape or disc. As the volume of a sound gets lower at some point it will be smaller than any noise present and not distinguishable from the noise background.

In general, the wider the dynamic range the better. It usually means lower noise and greater volume (loudness) to better match the range of human hearing.

Dynamic range can be stated in terms of power or voltage. It is usually a very wide range, so we resort to a unique mathematical method to express it. The term decibel (dB) refers to how dynamic range is presented. In math form you take the ratio of the maximum and minimum voltage, take the logarithm of that, and multiply by 20:

$$\mathrm{dB} = 20\log\left(V_{max}/V_{min}\right)$$

For power,

$$\mathrm{dB} = 10\log\left(P_{max}/P_{min}\right)$$

The dynamic range of a typical phonograph record is probably about 50 dB, and magnetic tape, about 55 dB. But a compact disc (CD) is approximately 90 dB, the best we have.

Dynamic range can be improved by lowering the noise in a system. This can be done with special signal-processing techniques. The most notable are methods produced by Dolby.

Digitizing Sound

Most digital sound and music is digitized with a sampling rate of 44.1 kHz or 48 kHz. Recall the Nyquist requirement that to capture and retain the frequency content accurately, the analog signal has to be sampled at least twice the highest frequency of the sound. Since most music and the human ear cut off at about 20 kHz, then any sampling rate greater than $2 \times 20\,\mathrm{kHz} = 40\,\mathrm{kHz}$ will satisfy that requirement. The rate of 44.1 kHz was standardized for CDs and 48 kHz for professional recordings. At 44.1 kHz, a sample is taken every $1/44{,}100 = 22.676$ microseconds.

Figure 10.3 shows the sampling of an analog signal. The smooth curve is the analog music or voice. At each vertical interval a sample is taken. In this

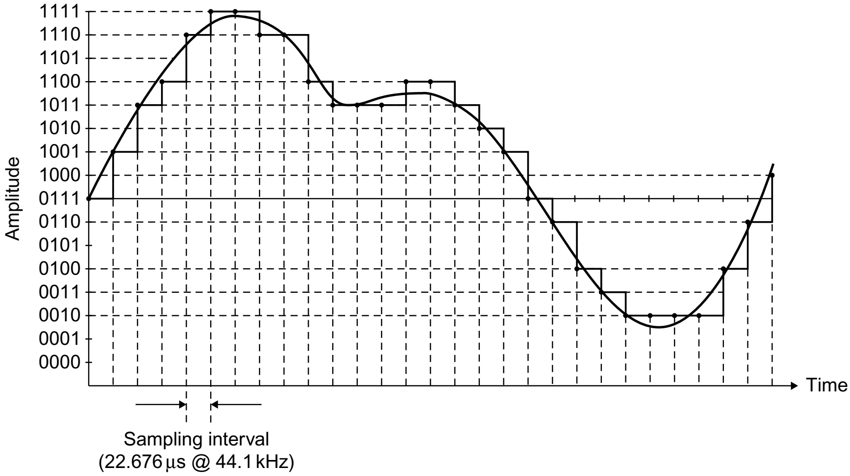

FIGURE 10.3 How audio signal is digitized. In this example, each sample is assigned a 4-bit code. The stepped approximation shown is what a DAC would deliver to the speakers.

example, a 4-bit ADC is used so there are $2^4 = 16$ possible levels. Note the 4-bit code associated with each sample amplitude. The binary samples are stored for later playback. When sent to a DAC, the samples will reproduce the audio in a stepped approximation shown overlaid with the analog signal. Some low-pass filtering of that signal will produce a sound which almost duplicates the original.

Digitizing audio usually results in each sample producing a 16-bit word. With 16 bits you can represent $2^{16} = 65{,}536$ levels. That produces a dynamic range from the highest to lowest amplitude levels of 65,536 to 1. On the decibel (dB) scale, that is a range of 96 dB—far greater than any dynamic range of previous analog recording methods. And noise is virtually nonexistent. It is so low no one can hear it. The overall result is a highly accurate way to capture sound and music. The 16-bit words can then be conveniently stored in a memory, or captured on a compact disc.

Digital Compression

Digital compression is a mathematical technique that greatly reduces the size of a digital word or bitstream so that it may be transmitted faster or stored in a smaller memory. Digitizing sound creates a huge number of bits. Assume stereo music that sampled at a rate of 44.1 kHz to create 16-bit words for each sample. One second of stereo music, then, produces $41{,}000 \times 16 \times 2 = 1{,}411{,}200$ bits. A 3-minute song is $60 \times 3 = 180$ seconds long. The result is $1{,}411{,}200 \times 180 = 254{,}016{,}000$ bits. Since there are 8 bits per byte, the result is 31,742,000 bytes or nearly 32 MB or megabytes. That is an enormous amount of memory for just one song. With a recording medium like the CD with a storage capacity of about 700 MB, that is okay. But for computers or portable music devices, it is impractical, not to mention expensive. And to transmit that over the Internet would take about 4 minutes at a 1 Mb/s rate. Pretty slow by today's standards.

The solution to this storage and transmission problem is to compress the bitstream into fewer bits. This is done by a variety of mathematical algorithms that greatly reduce the number of bits without materially affecting the quality of the sound. The process is called *digital compression*. The music is compressed before it is stored or transmitted. Then it has to be decompressed to hear the original sound.

The two most commonly used music compression algorithms are MP3 and AAC. MP3 is short for MPEG-1 Audio Layer 3, the algorithm developed by the Motion Picture Experts Group as part of a system that compressed video as well as audio. AAC means advanced audio coding. MP3 is by far the most widely used for storing music in MP3 music players and sending music over the Internet. AAC is used in the Apple iPod and iPhone and used on the iTunes site to send music. It is also part of later MPEG2 and MPEG4 video compression formats. Both methods significantly reduce the number of bits to roughly a tenth of their original size, greatly speeding up transmission and easing storage requirements. There are many more compression standards out there, but these are by far the most used and the ones you will most likely encounter.

To perform the compression process you actually need a special CPU or processor. It is typically a special DSP device programmed with the algorithm for either compressing or decompressing the audio.

There are also a number of compression methods used just for voice. Voice compression was created to produce signals for telephony. Most phone systems assume a maximum voice frequency of 4 kHz. The most common digitizing rate is twice that or 8 kHz. Eight-bit samples are typical. If you digitize voice creating a stream of samples in serial format, the signal would look like that in Figure 10.4. Each sample produces an 8-bit word where each bit is $125/8 = 15.625\,\mu s$ long. That translates to a serial data rate of $1/15.625\,\mu s = 64$ kbps. This takes up too much bandwidth in a telephone system, so compression is used. The International Telecommunications Union (ITU), an international standards organization, has created a whole family of compression standards. These are designated as G.711, G.723, G.729, and others. These mathematical algorithms reduce the bit rate for transmission to about 8 kbps. You will see them used in VoIP (voice over Internet phone) digital phones, which are gradually replacing regular old-style analog phones.

There are many other forms of audio compression. Another common one is Dolby Digital or AC-3 that is used in digital movie theater presentations and some DVD players.

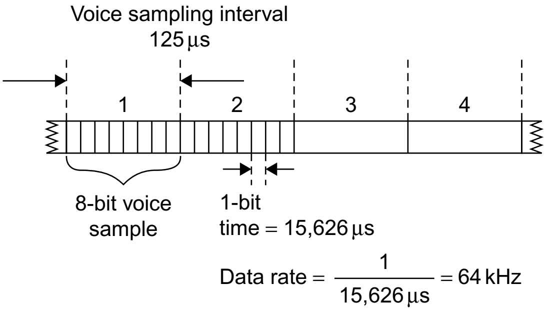

FIGURE 10.4 Voice signals up to 4 kHz for the telephone are sampled at an 8-kHz rate, producing an 8-bit sample, each 125 microseconds. Stringing the samples together in a serial data stream produces a digital signal at a 64-kHz rate.

How an MP3 Player Works

MP3 players such as the Microsoft Zune or Apple iPod are amazing electronic systems in a hand-held package. They all take the general form shown in Figure 10.5. At the heart of the player is a processor. It may be a general processor or a DSP. In some cases, two processors are used, one to control the player buttons, controls, and display, and other general housekeeping functions; a separate DSP handles the compression/decompression. A ROM holds the general control program for the device. A large flash EEPROM is used to store the music in compressed form. Some devices also provide for an external flash memory plug-in device to hold more music. A USB port is the main I/O interface to connect to a PC for the music downloads.

The codec in the device is the coder-decoder. This chip holds the ADC to digitize music from an external source such as a microphone (coder) and the DACs that decode or translate the decompressed audio back into the original analog music for amplification. Class-D switching power amplifiers are used for the headphones to reduce power consumption. One or more internal speakers may also be used. Other circuits include the LCD display and its drivers and LED back-lights if used. A power management chip is normally used to provide multiple DC voltages to the different circuits and to manage power to conserve it. This section also contains the battery charger.

The Compact Disc

The compact disc (CD) is by far the most popular digital audio medium in use today. It has been around since the 1980s and is still going strong. It has virtually totally replaced the Philips cassette magnetic tape cartridges and vinyl phonograph discs. About the only competition the CD has is from flash EEPROM devices such as those used to store MP3/AAC music for portable music players.

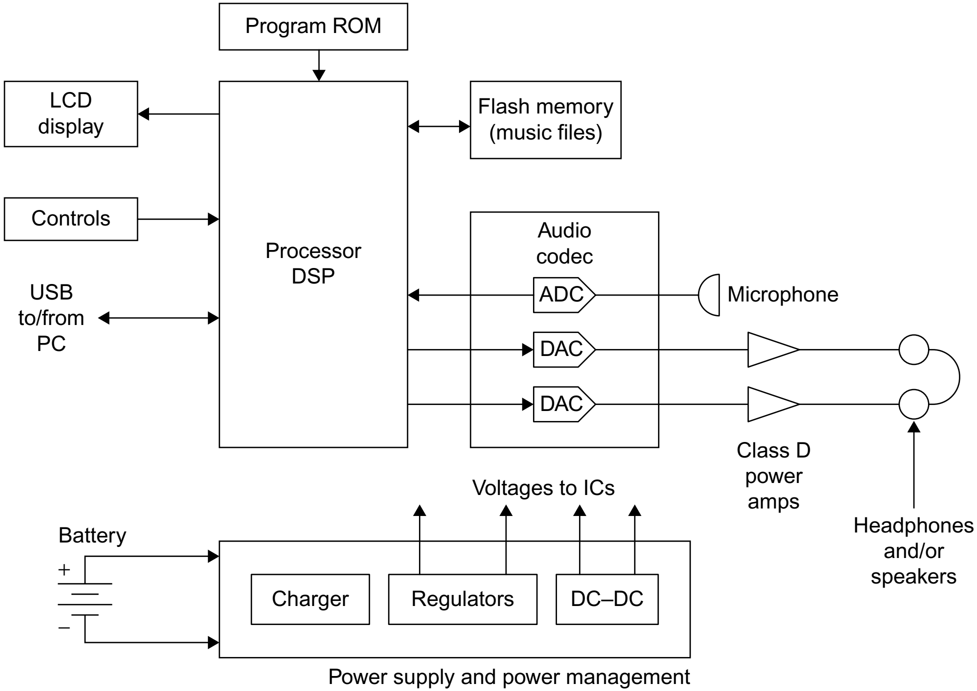
Program ROM
LCD display
Controls
USB to/from PC
Processor DSP
Flash memory (music files)
Audio codec
ADC
DAC
DAC
Microphone
Class D power amps
Headphones and/or speakers
Voltages to ICs
Battery
+
−
Charger
Regulators
DC–DC
Power supply and power management

FIGURE 10.5 General block diagram of a digital music player like Zune MP3 or Apple iPod.

You have probably seen a CD. Let's add a little background so you will appreciate the CD. It is a 4.72-inch (120-mm) diameter disc made of plastic. It is about 0.05 inch (1.2 mm) thick and has a hole in the center for mounting on the motor that spins the disc. It is actually a sandwich of clear plastic and a thin metal layer for reflecting light.

The music is recorded on the disc in digital format. A laser actually burns *pits* into the plastic representing the 1's and 0's of the digital data (see Figure 10.6). These pits are extremely small, about 0.5 μm wide (a micron designated μm is 1-millionth of a meter). The flat areas between the pits are called *lands*. A binary 1 occurs at a transition from a pit to a land or vice versa while the land is a string of binary 0s. The binary data is recorded as a continuous spiral track about 1.6 μm wide, starting at the center and working outward. That is a very efficient way to store the data, but it means that the motor

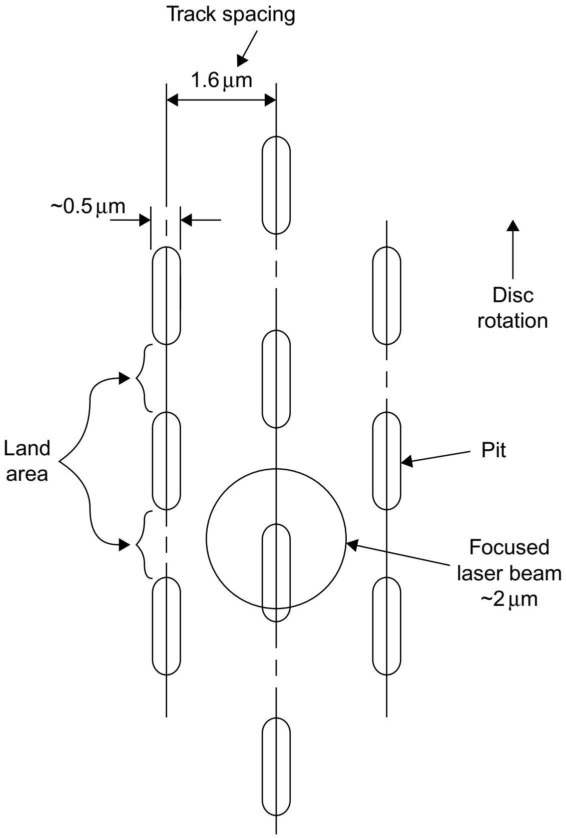

FIGURE 10.6 Digital music on a CD is encoded as pits and lands on the CD surface. These are burned in by a laser. A laser light shined on the pits and lands produces a reflection that a photodetector picks up to create the original digital data.

speed must vary to ensure a steady bit rate when the music is played back. The pickup mechanism will experience higher speeds at the center of the disc and lower speeds at the outer edges of the disc, so that a speed control mechanism keeps the recovered data rate constant.

The music on a CD is derived from two stereo channels of music with a frequency response from 20 Hz to 20 kHz. Digitization occurs at the standard 44.1-kHz rate. Each sample is 16 bits long. The 16-bit words from the left and right channels are alternated and formatted into a serial data stream that occurs at a rate of 44.1 kHz $\times$ 16 $\times$ 2 $=$ 1.4112 MHz. The 16-bit words are then encoded in a special way. First, they undergo an error detection and correction encoding scheme using what is called *cross-interleaved Reed-Solomon code* (CIRC). This coding helps detect errors in reading the disc caused by dirt, scratches, or other distortion. The CIRC adds extra bits that are used to find the errors and fix them prior to playback.

Next, the serial data string is then processed using what is called eight-to-fourteen modulation (EFM). Each 8-bit piece of data is translated into a 14-bit word by a look-up table. EFM formats the data for the pit-and-land encoding scheme. Finally, the completed data is formatted into frames 588 bits long and occurring at a rate of 4.32 Mbps. That is the speed of the serial data coming from the pickup assembly in a CD player before processing.

To recover the audio from the CD, the disc is put into the player and a motor rotates it. Refer to Figure 10.7. A laser beam is shined on the disc as indicated in Figure 10.6. The reflections from the pits and lands produce an optical light pattern that is picked up by a photodetector that converts the light variations into the 4.32 Mbps bitstream. A motor control system varies the speed of the motor to keep the data rate steady. The data stream goes to a batch of processing circuits. First, the EFM is removed, and then a CIRC decoder identifies and repairs any errors and recovers the original 1.41-Mbps data stream. A demultiplexer separates the left- and right-channel 16-bit words and sends them to the DACs, where the original analog music is recovered and sent to the power amplifiers and speakers.

The CD can store lots of data. Its capacity is in the 650- to 700-MB range. That translates into a maximum of about 74 minutes of audio. And this is not compressed.

AV RECEIVER

An HDTV receiver is usually at the heart of most consumer electronics home entertainment systems. A secondary piece is the AV (audio–video) receiver, which provides the audio component of the entertainment. The systems usually include the CD player, several radio options, and all the audio power amplifiers that operate the multiple speakers. The AV receiver also accepts inputs from the TV set, DVD player, and other external devices to provide higher-quality sound. This section provides a look at this piece of equipment and how it works.

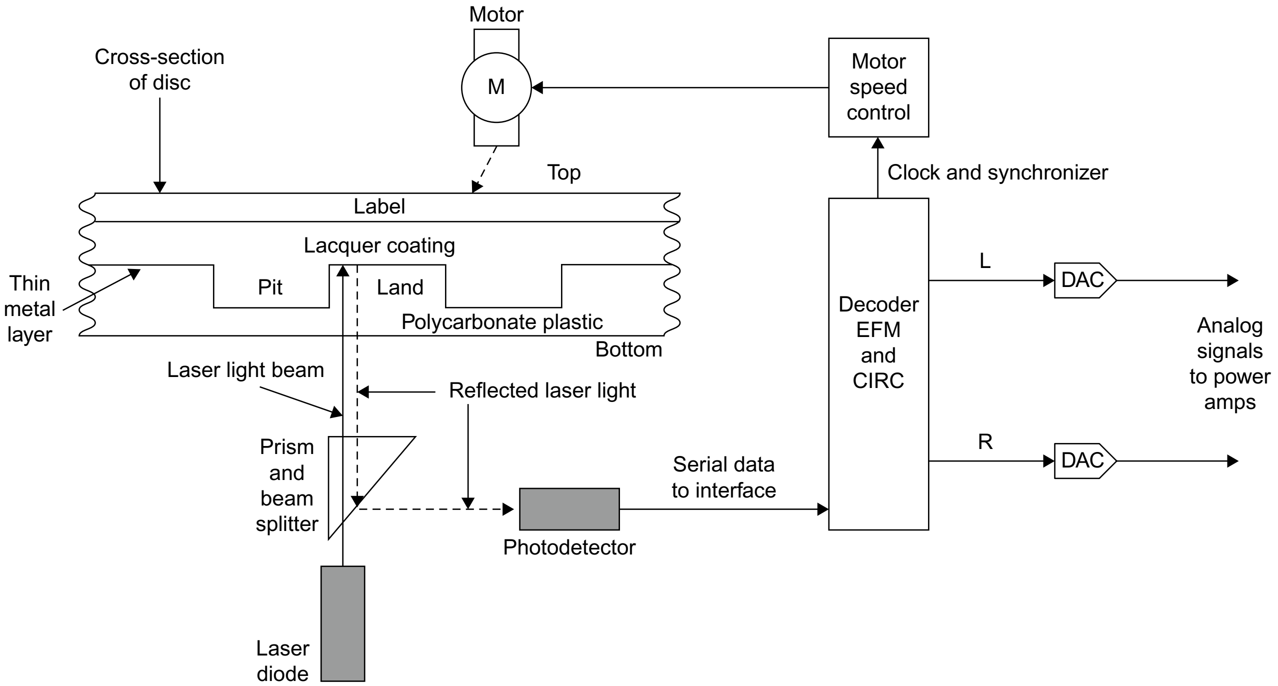
Motor
M
Motor speed control
Cross-section of disc
Top
Clock and synchronizer
Label
Lacquer coating
Thin metal layer
Pit
Land
Polycarbonate plastic
Bottom
L
DAC
Decoder EFM and CIRC
Analog signals to power amps
R
DAC
Laser light beam
Reflected laser light
Prism and beam splitter
Serial data to interface
Photodetector
Laser diode

FIGURE 10.7 Block diagram of CD player.

Figure 10.8 shows a general block diagram of the AV receiver. Note the multiple power amplifiers on the right. These drive the speakers that are part of a surround sound system. Most of these are class AB linear amplifiers that deliver power levels from about 20 watts per channel to over 100 watts in the larger units. The frequency response is 20 Hz to 20 kHz or more with a distortion level of 0.5 to 0.9% or less.

The power amplifiers get their inputs from multiple sources. These sources include audio from the internal radio receivers and the CD player as well as many other external sources. These sources include the TV receiver, DVD player, VCR, and a cable or satellite box, as well as audio from older devices like tape players and phonographs. Many receivers also accept inputs from an iPod or other MP3 music player. A large switching matrix is used to select the

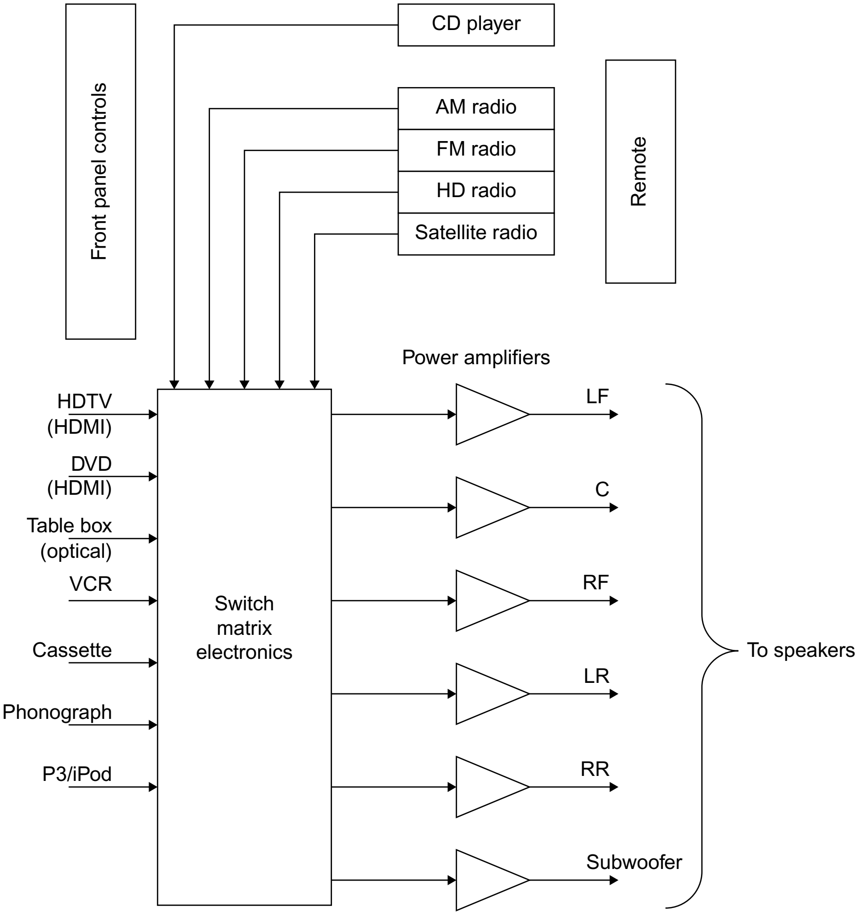
Front panel controls
CD player
AM radio
FM radio
HD radio
Satellite radio
Remote
Power amplifiers
HDTV (HDMI)
DVD (HDMI)
Table box (optical)
VCR
Cassette
Phonograph
P3/iPod
Switch matrix electronics
LF
C
RF
LR
RR
Subwoofer
To speakers

FIGURE 10.8 Block diagram of typical AV receiver, the heart of the audio reproduction part of a consumer home entertainment center.

desired input. That selection can be made from the front panel controls or via the remote control.

Radio Choices

All AV receivers contain the traditional AM and FM analog radios. These are still popular sources of music. But more and more the most recent receivers contain one or more digital radio sources such as HD radio or satellite radio.

HD radio is becoming more popular each year. This is a digital radio broadcast option available at most U.S. radio stations today. The AM and FM sources are digitized and broadcast on the same frequencies but with digital modulation techniques (OFDMs). The result is an improvement in the signal. Digital signals offer slightly better frequency response on both AM and FM. AM signals will sound more like FM and FM signals will be nearly CD quality. Furthermore, the digital signals are more noise- and fade-free than the standard AM and FM signals. While HD radio was first made available in automobiles, it is now widely available for home use. Separate HD radios can be purchased at most consumer electronics stores. And more HD radios are being built into AV receivers.

Satellite radio was also first made popular in automobiles. Yet separate home receivers are available. Some AV receivers now include satellite radio capability. Satellite radio has been available from two sources, Sirius Radio and XM Radio. Both are subscription services with hundreds of radio, music, news, sports, and other channels. The signals are digital, delivering near CD-quality sound. Both systems use the 2.3-GHz microwave band. The main difficulty when using satellite radio at home is that an outside antenna is needed. Some radios supply a small antenna for window mounting so that the antenna can "see" the satellites in the sky. Otherwise, poor or no reception may occur.

Surround Sound

Virtually all audio sound today is stereo. That is, the music comes from at least two speakers, one on the left and one on the right. The sound is recorded as two separate channels with different microphones to provide a stereophonic experience to the listener. A minimal system uses two speakers but some add a subwoofer for good bass reproduction. This is sometimes referred to as a 2.1 stereo system, that is, two speakers and one subwoofer.

New audio systems offer surround sound. These systems feature six, seven, or eight speakers. The idea is to totally envelop the listener with sound as it might be heard in a concert hall or theater. There are speakers in front and at the rear. The first surround-sound implementations were analog but today most surround sound is fully digital. The most common surround-sound method is Dolby Digital 5.1, which uses five speakers and a subwoofer as shown in Figure 10.9. These are left front (LF), right front (RF), center (C), left rear (LR), and right rear (RR). Then there is one subwoofer. The woofer can be positioned

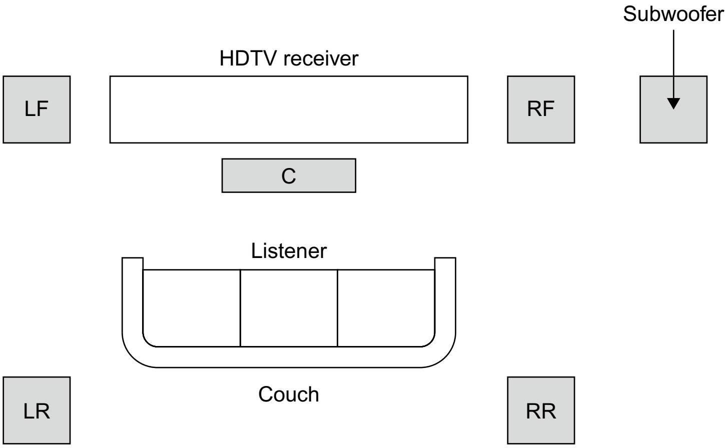

FIGURE 10.9 How 5.1 surround sound system is set up.

anywhere but is usually at the front. Each speaker is driven by its own amplifier in the AV receiver.

The big question that you may have is where do the six channels of sound come from? Not from the radio or a CD for sure, both of which are stereo only. In most cases, the sound signals will come from a DVD player. Right now it is about the only common source of 5.1 surround sound, although some HDTV programs support it. During recording of the original material, one microphone is used for each channel. The digitized audio for each of the channels is stored on the DVD.

Incidentally, there are more elaborate versions of surround sound. For example, 7.1 adds two more speakers on the left and right. As of now there are few, if any, 7.1 music sources available.

When you play the DVD, the digital signals from each channel are sent to a special chip that decodes them and otherwise performs necessary noise reduction, equalization, and other operations. The recovered signals are then sent to DACs for conversion to their original analog format for amplification.

SPECIAL SOUND APPLICATIONS

When you think of audio or sound, you think mainly of voice or music reproduction. Yet, electronic sound is used in lots of additional places. Some interesting examples are in musical instruments, sonar, ultrasound, and hearing aids.

Musical Instruments

Musical instruments produce their own sounds, of course. Each has its own unique signature made up of the basic notes or tones to which are added the mix of harmonics or overtones unique to the physical structure of the instrument. Some instruments like horns and drums rarely need amplification. Other instruments like guitars and violins commonly need some amplification, especially in large venues such as on stage addressing hundreds or thousands of people. Special microphone-like devices called pickups are used with string instruments. The pickup is usually one that converts the string vibrations into a small signal that is then amplified.

One of the most popular electronic instruments is the *synthesizer*. It has a keyboard like a piano. The keys are equivalent to those on a piano. Each key then triggers the electronics to generate a tone of the desired frequency. That tone is especially modified by DSP circuits to sound like that note played on a piano, an organ, or any other musical instrument. The DSP synthesizes or constructs the tones digitally and sends them to a DAC that converts them into the analog sound to be amplified.

Are Vacuum-Tube Amplifiers the Best Audio Amplifiers?

Vacuum tubes? I'm kidding, right? Actually, no. There are many audio experts, especially guitar players and some recording studio engineers, who still believe that vacuum-tube amplifiers are superior to solid-state amplifiers. And as it turns out, the largest market for general vacuum tubes today is guitar, high-end stereo, and recording studio amplifiers.

Contrary to popular belief, vacuum tubes did not really go away when transistors and integrated circuits came along. They are still widely used in China, Russia, and a few other countries. And that is where most tubes are made these days. Vacuum tubes are still widely used. Cathode ray tubes (CRTs) or picture tubes are still at the heart of many TV sets and computer video monitors. Magnetrons are vacuum tubes that are in every microwave oven. And microwave vacuum tubes, such as traveling-wave tubes (TWTs) and klystrons, are still used in high-power microwave transmitters. And vacuum-tube audio amplifiers.

Why are vacuum-tube amplifiers so popular still? It is the unique sound they produce. It is a sound that musicians appreciate and want in their sound mix. And the sound is difficult and expensive to reproduce in a solid-state amplifier. The sound is smoother and less harsh. The sound is "warmer," and the compression that such amplifiers produce when overloaded is exactly what makes rock and country music sound like it does. It looks like vacuum tubes will be around much longer, but prices are sky high and some tube models are hard to find.

Noise-Canceling Earphones

You have probably heard of these or maybe seen or used them. They make clever use of analog signal processing to greatly minimize surrounding noise when you are trying to listen to music with headphones.

First, most of these headphones try to isolate the actual headphone from the surroundings with a tight-fitting ear bud or soft foam pads around the earphone that limit the amount of external noise getting to your eardrum. This is called *passive noise control*.

Second, each earphone contains a built-in microphone that picks up the surrounding noise. It amplifies it then inverts it in phase. The original noise from the microphone is then added to the inverted noise. Being the same but exactly out of phase with one another, the two noise signals cancel one another. The cancellation is not perfect, but the noise reduction is significant. It can greatly minimize noise on a plane or in a car or other types of noisy environments.

Sonar

Sonar means sound navigation and ranging. It is basically the underwater equivalent of radar. It is used on ships and submarines for making depth measurements, detecting underwater objects, and for warfare. The two types of sonar are passive and active.

Passive sonar is simply the idea of putting underwater microphones called *hydrophones* on a long wire or arrays of such microphones in the water to listen for any sounds that occur. Amplifiers boost the signal levels and various filters help sift through all the complex signals that can be heard. In military applications, the sounds are digitized and various DSP programs help filter and identify specific sounds. Many ships create unique sounds called signatures that can be digitized and stored for comparison to any sounds picked up for positive identification. The sounds include fish and other sea life, boat and ship engines, motion through the water by a ship or boat, or any noise that occurs inside a submarine. Sound travels very easily through water so is easy to pick up. With practice and experience, a sonar operator can identify almost everything occurring over a wide range around the microphone array.

The other form of sonar is *active*. This is like radar where a high-frequency pulse of sound is applied to a transducer to convert the pulse into sound waves for transmission through the water. The frequency of the pulse is usually ultrasonic, that is, above most human hearing. Frequencies from about 15 kHz to 1 MHz are used, depending on the application. The pulse of sound travels out from the transducer and is then reflected off distant objects. The reflections go back to the source where they are picked up by a microphone. Then, knowing the speed of sound transmission in water, the distance to the object can be computed. Sonar depth sounders point the transducer downward to get a measure of the distance to the bottom. Sonars can easily pick up reflections from ships miles away for navigation safety purposes or for military actions.

Hearing Aids

Electronic hearing aids have been around a long time. The older analog types were large and cumbersome but worked fine. Integrated circuits made them much smaller and more acceptable to the hearing impaired. Today, the newer hearing aids use digital technology. The microphone picks up the sound and amplifies it then digitizes the sound in an ADC. A DAC can then translate the sound back into audio for amplification and application to the earphone. Most digital hearing aids now incorporate DSP filters. These filters may be customized to the frequency response of the ear, correcting for different frequencies. Mostly the high-frequency response is lost due to aging. Using an external computer, each hearing aid can be adjusted from volume and frequency response to fit the exact deficiency.

Project 10.1

Experience Digital Radio

Digital radio in the United States is available from satellites and via the existing AM and FM analog stations through HD radio. Both offer commercial-free music and other content in multiple channels. HD radio is free but satellite radio is a pay-for-listen service.

In the United States, two satellite radio services were established in the early 2000s, XM Radio and Sirius Radio. Recently the two services merged to form Sirius XM Satellite Radio. Both offer subscription services to music, news, sports, and other information. Both services operate in the 2.3-GHz band, and get their signals from satellites. Special receivers and antennas are required.

Satellite radio is a digital service and both systems used compressed audio for signal transmission. HD radio is simultaneously broadcast with existing AM and FM analog signals. A special receiver is needed to receive it. The digital techniques greatly improve sound quality.

A good project if you are a music lover is to try one or both of these services. You can get an HD or satellite radio for your vehicle or you can buy a home unit. Some high-end AV receivers have one or the other in addition to the traditional AM and FM radios. You can purchase home desktop units from Best Buy and other electronics dealers for a reasonable price.

HD radio is free and you will find that some stations transmit additional digital channels not available in analog form. These give you more music genres. No special antenna is needed.

To learn more about HD radio, do a Bing, Google, or Yahoo! search. And go directly to the company that invented HD radio, iBiquity (www.ibiquity.com), for more details.

You can also buy a satellite radio from several local sources. But you will need to subscribe. Go to www.sirius.com and www.xmradio.com.

Project 10.2

Test Your Hearing Response

To check the frequency response of your ears, go to www.audiocheck.net/audiotests_frequencycheckhigh.php. There are a few online audio tests for your hearing. This website has a whole batch of other audio response tests that you may be interested in.

Chapter 11

Video Technology

TV and Video Is All Digital Now

In this Chapter:

- How scanning works to create video.
- Basic color TV principles.
- U.S. and other standards for digital and high-definition TV.
- How a digital TV set works.
- TV screen technologies.
- Cable TV.
- Satellite TV.
- Cell phone TV.
- DVD players.

INTRODUCTION

Nothing in electronics has affected us more than the development of television. Okay, I haven't forgotten the PC and the Internet, which have also seriously impacted our lives. But TV came first and has been working its magic on us for far longer. The impact of television or video on our lives is significant in that we spend more than half our waking hours in front of a TV screen, either watching TV or working with a video monitor on a computer. The social and political impact has been enormous. And, of course, you also interact with the liquid crystal display (LCD) screens on cell phones, laptops, and personal navigation devices. For this reason it helps to understand a little bit about how video works. Video is perhaps one of the most complex segments of electronics, but this chapter makes those fundamentals easy to understand.

VIDEO FUNDAMENTALS FOR THE IMPATIENT

The main feature of any video device is the screen. The first video or TV screens were cathode ray tubes (CRT) and those are still used in older TV sets and computer monitors. The most commonly used TV and video screen today is the LCD. LCD screens are used in the newer TV sets and in most personal computers.

© 2010 Elsevier Inc. All rights reserved.
doi: 10.1016/B978-1-85617-700-9.00011-4

They are also used in laptops, netbooks, cell phones, personal navigation devices, and anything else that has a screen on it. TV sets also use plasma, digital light-processing (DLP) projection technology, and other methods to display very large pictures. There is more detail on TV screens later in this chapter, but for now, the most important part of understanding video is to learn the methodology used to create and present video regardless of the screen type.

A video system is a method for converting a picture or scene into an electronic signal that can be transmitted by radio or cable or stored electronically. Some type of device is needed to take in the light and color information from a picture or scene and convert it into an electrical signal. Once in electrical form, the video can be modulated onto a radio carrier for transmission or electronically stored. The process of converting the scene or picture into an electrical signal is known as scanning.

Scanning Fundamentals

A video signal is usually generated by a video camera. The camera takes the scene and scans it electronically, dividing it into hundreds or even thousands of very thin horizontal lines (see Figure 11.1).

The horizontal scan lines comprise a sequence of light and color variations. The camera produces an analog signal whose frequency and amplitude vary with the brightness and color changes along the scan line. The whole idea in video is to transmit the video signal one scan line at a time at very high speeds. If you are not transmitting the signal by radio or cable, you can store it, such as recording it on magnetic tape. In any case, the received signals are then converted back into scan lines and applied to the screen by scanning the lines across the receiving screen very fast from left to right and top to bottom.

The secret to producing a high-quality video picture is to use enough scan lines with enough detail in each so that the color, brightness, and fine detail in the picture are retained. Note in Figure 11.1 that the analog signal for one scanned line is such that the higher-amplitude voltages are closer to black while the lower-amplitude signals are closer to the white level. This is what a scan line from a monochrome or black and white (B&W) video system produces. It is called the *brightness* or *Y signal*. Color will be covered later.

Aspect Ratio

When talking about TV screens, the key factors are size and aspect ratio. The *size* is usually the diagonal measure of the screen in inches. The *aspect ratio* is the ratio of the picture width to the picture height. Standard TV sets and video monitors use an aspect ratio of 4:3 as shown in Figure 11.2. Most of the newer TV screens and video monitors have an aspect ratio of 16:9. This wider format was chosen because it is a better fit with the wide-screen movie formats used by Hollywood.

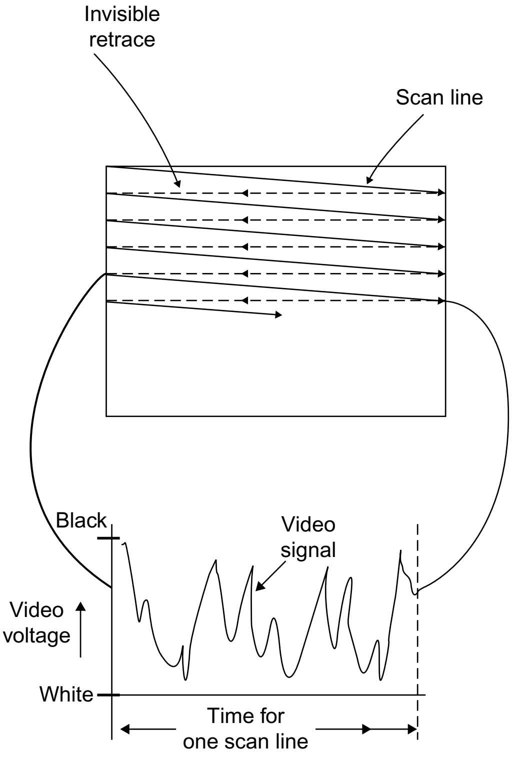
Invisible
retrace
Scan line
Black
Video
signal
Video
voltage
White
Time for
one scan line

FIGURE 11.1 How scanning converts a scene of light and color into an electronic signal.

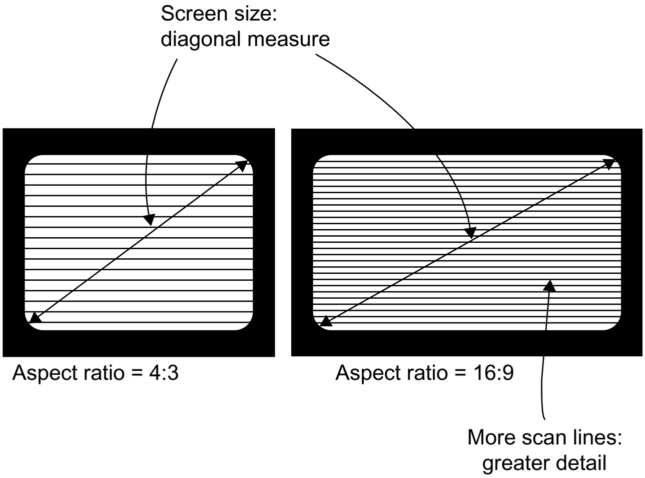
Screen size:
diagonal measure
Aspect ratio = 4:3
Aspect ratio = 16:9
More scan lines:
greater detail

FIGURE 11.2 TV screen size and aspect ratio.

Persistence of Vision

When you think about it, it just doesn't seem possible that a complete picture or scene can be reproduced in electronic format by scanning. The scanning process just breaks the picture or scene down into hundreds or perhaps thousands of scan-line signals, and the only way that a picture can be created is for those scan lines to occur at a very high rate of speed. Luckily, this technique works because we all have a visual characteristic known as persistence of vision. Persistence is that characteristic of the eye which prevents it from following very high-speed changes in light variations. In other words, if the scan lines are traced across the screen at a very high rate of speed, your eye simply cannot follow them one at a time but instead puts the whole thing together and sees a single steady picture. If you scan at too low a rate, you will actually begin to see the scan lines and your eye will interpret this as a kind of flicker.

The concept of persistence of vision has been known for decades. Motion picture films take advantage of this characteristic of our eyes as well. Remember that movies are just a high-speed sequence of individual still pictures recorded on film. If you can get the frame rate up to a level of approximately 24 frames per second or faster, your eye will translate the changes from one still frame to the next as motion.

Analog TV

As you probably know, the Federal Communications Commission eliminated analog TV broadcasts in the United States as of June 12, 2009. But while analog TV has been replaced by digital TV, that doesn't mean that it isn't still around. You will still find analog TV in many other countries of the world, and it is also still widely used in closed-circuit television. In any case, it is such a widely used standard that understanding it is useful.

Back in the late 1940s and early 1950s, the United States created what is called the National Television Standards Committee (NTSC) TV standard. It basically specifies that a picture is made up of 525 horizontal scan lines, known as a video *frame*. As it turns out, one frame is actually made up of two fields of 262.5 scan lines. The scene or picture is first scanned 262.5 times from top to bottom, creating one field. This field is then transmitted at a rate of 60 fields per second. Then the scene is scanned again with another 262.5 lines and these are transmitted next. The second field is interlaced with the first field to form the complete 525 line picture. See Figure 11.3. The result is that you will see 30 complete 525-line frames per second. This is fast enough for persistence of vision to work nicely. The interlacing helps significantly in reducing any perceived flicker.

One thing that you may be wondering is how do you keep the TV receiver screen in step with the video camera or other video source? This is solved by creating synchronizing pulses that are transmitted before and after each scan line (see Figure 11.3). After one field is scanned, a larger group of synchronizing pulses occurs between fields to cause the receiver to start the scanning back at

Other World TV Systems

The U.S. NTSC analog system is used in Canada, Mexico, and Japan. But other analog systems are used throughout the world. In Europe, specifically the UK, Germany, Italy, and Spain, the PAL (phase alternation line) standard is used. It uses 625 scan lines with a frame rate of 25 frames per second. In France and in some other countries, a system called SECAM for Sequenteil Couleur a Memorie (or sequential chrominance and memory) is used. It too uses 625 scan lines and a 25-frame-per-second rate. PAL and SECAM systems have a better vertical resolution for picture detail because of the greater number of scan lines, giving a perceptibly improved picture over the NTSC system. Most of the other countries of the world have adopted one of these three systems.

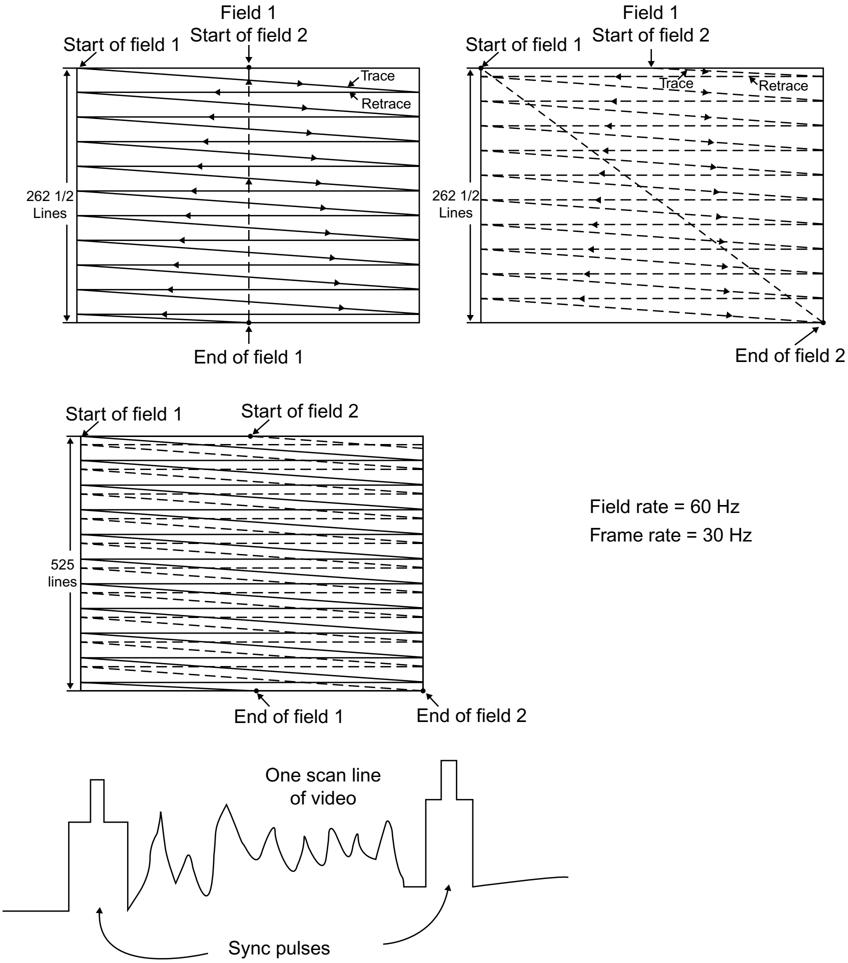
Field 1
Start of field 2
Start of field 1
Trace
Retrace
262 1/2 Lines
End of field 1
Field 1
Start of field 2
Start of field 1
Trace
Retrace
262 1/2 Lines
End of field 2
Start of field 1
Start of field 2
525 lines
End of field 1
End of field 2
Field rate = 60 Hz
Frame rate = 30 Hz
One scan line of video
Sync pulses

FIGURE 11.3 Interlaced scanning helps reduce flicker. Sync pulses help keep the receiver and transmitter in step with one another.

the top again. The synchronizing pulses are transmitted or stored along with the analog video signal for each line. These synchronizing pulses are then used to trigger the receiver's circuitry to make sure that the scene is sequenced properly on the screen.

Digital TV

Most TV and video now is transmitted and stored in digital or binary format. Scanning is still the process used to convert the scene into an electrical signal, in this case a stream of binary numbers representing the brightness and color detail. Each scan line is still essentially an analog signal but that is converted into a sequence of binary numbers by an analog-to-digital converter (ADC). The basic process is illustrated in Figure 11.4. The ADC samples are measures of the analog signal at a constant rate, as shown by the dots along the curve

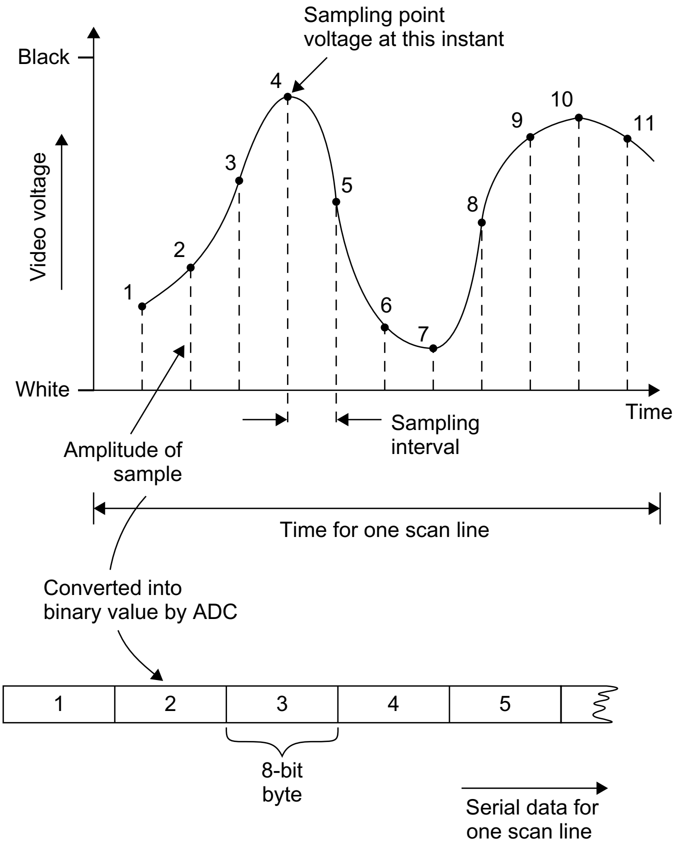
Sampling point
voltage at this instant
Black
White
Video voltage
1
2
3
4
5
6
7
8
9
10
11
Time
Sampling
interval
Amplitude of
sample
Time for one scan line
Converted into
binary value by ADC
1
2
3
4
5
8-bit
byte
Serial data for
one scan line

FIGURE 11.4 Digital TV is created by digitizing the analog video signals with an ADC, and then converting it into a serial bitstream of video data words.

in Figure 11.4. Remember that in order to retain the information in an analog signal when it is converted into digital, the signal must be sampled at a rate at least twice as fast as the highest-frequency signal contained in the analog signal. Since video signals contain extremely high frequencies, the sampling rate is very fast. Most video contains scene details with frequencies beyond 4 MHz. That means a sampling rate of at least 8 MHz to capture the detail. In any case, each sample results in an 8-bit binary number. These bytes, one representing each of the sequential samples, are sent one after another in a serial binary data stream as shown. The video data can then be transmitted or stored.

Each of the binary numbers produced by sampling produces what we call a *pixel* or picture element. This particular binary number will ultimately translate to a specific analog voltage value and produce a specific brightness level along the scan line.

One way to visualize a digital video display is to think of it as an array, or horizontal scan lines, each made up of a sequence of pixels (see Figure 11.5). Each pixel is a dot of light or, in this case, a square. That pixel is derived from a specific binary number developed during the scanning and sampling process. Each of those pixels represents a specific brightness level from white to black. A digital-to-analog converter (DAC) converts the binary brightness value back into a specific shade of gray.

Digital TV (DTV) uses the interlaced scanning principles described earlier to help minimize flicker. But it also uses a faster progressive scan. This is where each scan line making up the screen is scanned sequentially one after the other from top to bottom. To minimize flicker, *progressive scanning* must be done at a higher rate of speed than interlace scanning. Progressive scanning

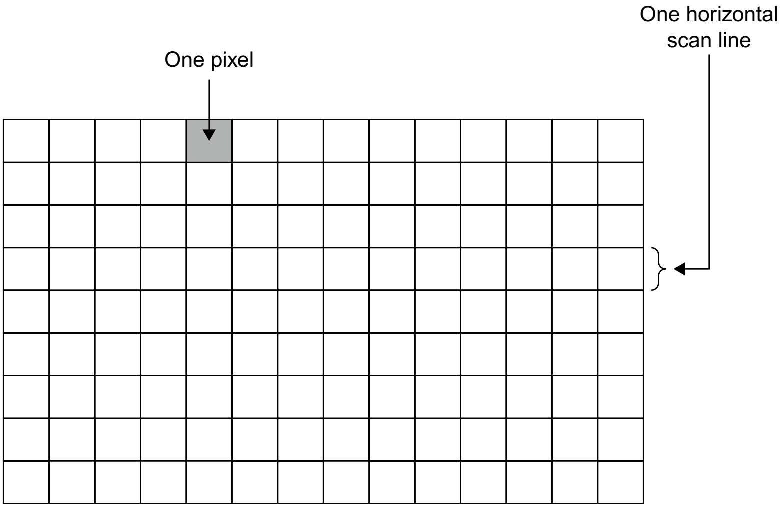

FIGURE 11.5 TV screens can be thought of as thousands of rows and columns of pixels, each represented by data.

is used simply because it creates a sequence of digital numbers that is more conveniently stored and processed one after the other than in a two-step interlaced process.

DTV uses a variety of different formats. In the basic format, it produces 480 scan lines with 640 pixels per line. This is roughly the equivalent of what you would see with an analog TV signal. This is what we call the DTV standard-resolution format. Both interlaced and progressive scanning formats can be used with rates of 24, 30, and 60 frames per second.

Picture Resolution

One thing that you should understand about video is the concept of resolution. This refers to how much fine detail you can see on the screen. In analog TV, brightness and color detail results in very-high-frequency analog signals. The older original analog TV was limited to frequencies of approximately 4.2 MHz. Modern HDTV uses much higher-frequency responses up to about 10 MHz. To see that picture detail, you must transmit the analog signal over a channel that has sufficient bandwidth. Most radio channels and cable systems have bandwidth limitations, and essentially act as a low-pass filter that will remove some of the higher frequencies. While you will still see a picture, it will lack some of the fine detail that your eye may or may not notice.

Resolution as it relates to DTV is a combination of how many scan lines plus how many pixels per line are used. Obviously, the greater the number of scan lines and the greater the number of pixels, the more detail you are going to be able to transmit and ultimately see.

A higher-resolution format to produce what we call high-definition (HD) TV uses 720 scan lines with 1280 pixels per line. The ultimate high-definition format uses 1080 horizontal lines with 1920 pixels per line. Both of those high-definition formats can use either interlace or progressive scanning at rates of 30, 60, 120, and 240 frames per second. Just remember that the frame rate simply refers to the number of times per second that each frame is presented on the screen. The earlier slow frame rates of 30 and 60 frames per second often produces a kind of picture smearing when the scenes are changing at a high rate of speed. This is noticeable in action movies or in sports events. The newer 120- and 240-frames-per-second scan rates essentially eliminate this problem.

Color Video Principles

Thus far, the signals we have described are what we call monochrome or black-and-white video. In *monochrome video*, all we capture and transmit and reproduce is the brightness variations along a scan line that range from black to white and an infinite number of shades of gray in between. But, of course, today most video is color. Capturing, transmitting, processing, and displaying color is a far more complex process than monochrome.

One physics principle that makes color video possible is the concept that any color of light can be reproduced by simply mixing the right combination of red, green, and blue light. Take a look at Figure 11.6. Mixing red and green creates yellow. Mixing red and blue produces violet (magenta), while mixing green and blue produces the color cyan. In any case, as it turns out virtually any color can be created using just the right mix and intensities of red, green, and blue. Strangely enough, there is one unique combination of the three colors that creates white. Black is produced, of course, by just turning off all the light sources. The color TV problem is how do you capture the individual colors and then how do you reproduce them?

Most video cameras use a light-sensitive imaging chip called a charge-coupled device (CCD). An optical lens focuses the scene on the chip. The chip itself is made up of millions of tiny light-sensitive elements that are charged like a capacitor with a voltage depending on the brightness level reaching them. The analog video signal for a scan line is then created by sampling those light-sensitive elements one at a time horizontally. The result is an analog signal whose amplitude is proportional to the light level. The result is a black-and-white brightness or Y signal.

Figure 11.7 shows how the color signals are produced. The scene is actually passed through a lens and a beam splitter that divides the scene into three identical segments. Each is passed through individual red, green, and blue filters. Three light-sensitive imaging chips (CCDs) then produce three separate analog signals for the red, green, and blue elements in a scene. The red, green, and blue analog signals are then individually amplified and sampled by an analog-to-digital converter (ADC) to translate them into a binary sequence. The three binary streams are then transmitted serially as a sequence of red, green, and blue for each picture element on the imaging chip. Each of these RGB colors will then later form a pixel on the viewing screen.

Video Compression

Generating a color video signal produces a serial bitstream that varies at a rate of hundreds of millions of bits per second. Consider these calculations:

- For each pixel, there are three color signals of 8 bits each making a total of $8 \times 3 = 24$ bits per pixel.
- For a standard DTV screen, there are 480 lines of 640 pixels for a total of 307,200 pixels per frame or screen.
- For one full screen or frame, there are $307{,}200 \times 24 = 7{,}372{,}800$ bits.
- If we want to display 60 frames per second, then the transmission rate is $7{,}372{,}800 \times 60 = 442{,}368{,}000$ bits per second, or 442.368 Mbps. That is extremely high frequency.

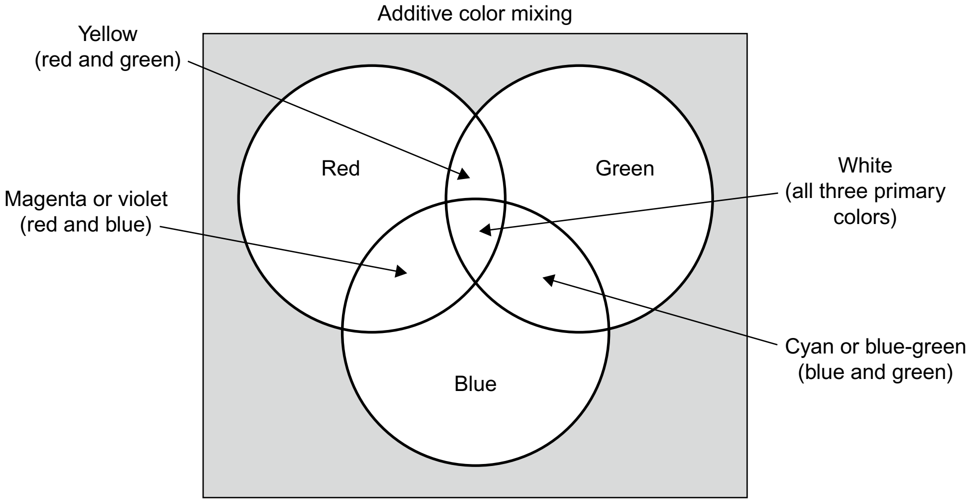
Additive color mixing
Yellow
(red and green)
Red
Green
White
(all three primary colors)
Magenta or violet
(red and blue)
Cyan or blue-green
(blue and green)
Blue

FIGURE 11.6 Any color of light can be reproduced by mixing red, green, and blue light in the right proportion.

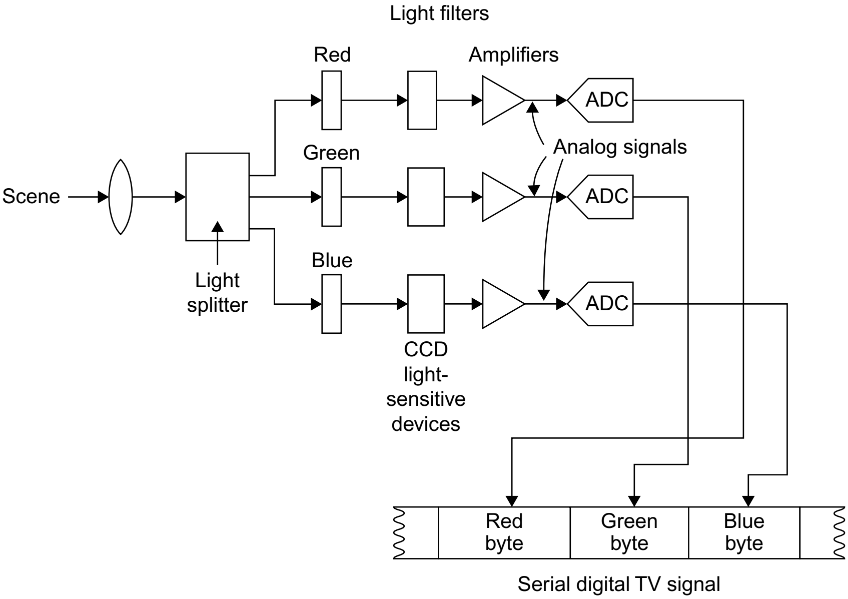

FIGURE 11.7 Red, green, and blue digital color signals are generated by filtering the scene into its red, green, and blue components, and then digitizing the resulting analog signals.

- To store one frame of video of this format, you would need a memory with 55,296,000 bytes or 55.296 MB. And that is just one frame.
- To store 1 minute of video you would need a memory of 3,317,760,000 bytes or just over 3.3 gigabytes (GB). And that is just 1 minute. Multiply by 60 to get the amount of memory for an hour.

Imagine the speeds and memory requirements for 1080p or a 120 frames-per-second format. And we have not even factored in the stereo audio digital signals that go along with this. Anyway, you are probably getting the picture here. First, data rates of 442.368 Mbps are not impossible, but they are impractical as they require too much bandwidth over the air and on a cable. And they are expensive. Furthermore, while memory devices like CDs and DVDs are available to store huge amounts of data, they are still not capable of those figures. Therefore, when it comes to transmitting and storing digital video information, some technique must be used to reduce the speed of that digital signal and the amount of data that it produces.

This problem is handled by a digital technique known as *compression*. Digital compression is essentially a mathematical algorithm that takes the individual color pixel binary numbers and processes them in such a way as to reduce the total number of bits representing the color information. The whole compression process is way beyond the scope of this book, but suffice it to say that it is a technique that works well and produces a bitstream at a much lower rate. And, the compressed video will take up less storage in a computer memory chip.

The digital compression technique used in digital TV is known as MPEG-2. MPEG refers to the Motion Picture Experts Group, an organization that develops video compression and other video standards.

What you have to think about when you consider the transmission of digital video from one place to another is that it is typically accompanied by audio. The audio is also in serial digital format. Those digital words will be transmitted along with the video digital words to create a complete TV signal.

TV Transmission

The compressed audio and video of a TV signal are assembled into packets or frames that look like that in Figure 11.8. The 4 bytes up front provide synchronizing bits for the receiver. The ID header tells the packet number, sequence, and video format used. The remaining bits are the serial compressed video and audio.

Figure 11.9 shows a simplified TV transmitter. The serial digital video is put through the MPEG-2 compression process. This is followed by additional

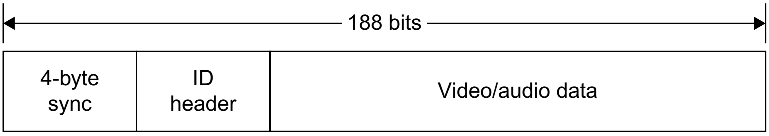
188 bits
4-byte sync
ID header
Video/audio data

FIGURE 11.8 HDTV packet for ATSC U.S. digital color TV system.

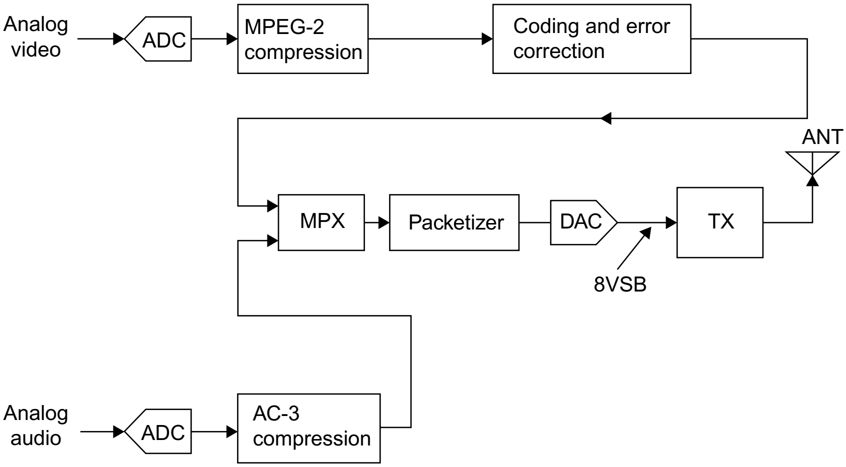
Analog video
ADC
MPEG-2 compression
Coding and error correction
ANT
MPX
Packetizer
DAC
TX
8VSB
Analog audio
ADC
AC-3 compression

FIGURE 11.9 General block diagram of TV transmitter.

bit manipulation and coding for error correction and improved signal recovery by the receiver. This serial data is then multiplexed with the serial compressed audio data. Then the packets are formed. This is what will modulate the transmitter.

The modulation begins by translating the fast serial bitstream into an eight-level signal with a DAC. Each 3-bit segment of the serial data is assigned a specific voltage level as shown in Figure 11.10. Since each level represents 3 bits, you can transmit more bits per second in the same time period. It is this eight-level signal that modulates the transmitter carrier. The modulation is AM, where all of the upper sidebands are transmitted but only part of the lower sidebands. This saves spectrum. The modulation is referred to as 8VSB for *vestigial sideband*, where vestigial refers to part of the lower sideband. The signal is then up-converted to the final frequency by a mixer. A high-power amplifier in the transmitter sends the signal via coax cable to the antenna. Typical TV station range is about 70 miles maximum, and that is determined by tower height and terrain.

In the United States, each TV channel is 6 MHz wide. There are 50 channels starting at channel 2 from 54 to 60 MHz. Channels 3 through 6 range from 60 to 88 MHz. Channels 7 through 13 run from 174 to 216 MHz. Channels 14 through 51 run from 470 MHz and ending with channel 51 from 692 to 698 MHz.

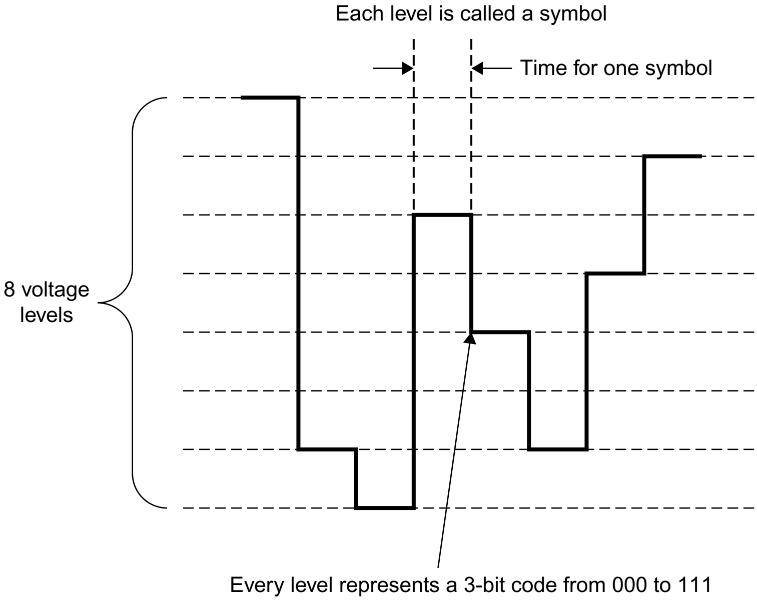
Each level is called a symbol
Time for one symbol
8 voltage levels
Every level represents a 3-bit code from 000 to 111

FIGURE 11.10 U.S. ATSC system translates the digital bitstream into an eight-level signal to transmit multiple bits per level, making higher data rates in a given bandwidth possible.

TV Receiver

Figure 11.11 shows a general block diagram of a modern digital TV receiver. The signal from the antenna or cable TV box is sent to the tuner where the signal is amplified and sent to a mixer along with a local oscillator signal to down-convert the signal to a lower frequency called the *intermediate frequency*. This is commonly 44 MHz. The local oscillator is a frequency synthesizer that selects the desired channel, usually with a remote control. The IF signal is filtered to select only the desired 6-MHz channel.

The IF is then sent to an ADC for digitizing. The 8VSB signal is then recovered in the demodulator. It is then translated into the original bitstream, and any coding or error correction is removed. The signal is then demultiplexed into the video and audio streams. The video goes to an MPEG-2 decompression circuit for recovery of the original video. The individual RGB components are sent to DACs to recover the picture detail. The RGB signals are then amplified and sent to the screen for display. The audio is also decompressed and sent to the DACs and amplifiers and speakers.

The whole TV set looks something like Figure 11.12. Note all the possible inputs. The set gets its RF inputs from the antenna or cable box. The cable box may also have a digital signal output, meaning the cable box does all the decoding, demultiplexing, and decompression. The output comprises a serial data stream of the video and audio. A special high-speed digital interface called the *high-definition media interface* (HDMI) is used. It uses a special cable no more than 15 feet long and with a special 19-pin connector. The HDMI carries video and audio from the TV to cable box or DVD player or vice versa.

Most TV sets also accept analog video inputs from other sources like VCRs, video cameras, and digital cameras. A special fiber optic cable input called TOSLINK serves as the interface for TV sets, cable boxes, and DVD players.

Finally, some TV sets include standard computer interfaces like USB ports or the IEEE 1394 interface, a fast digital serial interface for video and audio developed by Apple.

TV set outputs, of course, are sent to the screen and the speakers. A remote control sends signals to an embedded controller that tells the set what to do.

TV Screen Technology

As indicated earlier, there are several ways that the video signal is translated back into a viewable picture. The cathode ray tube is still present in older sets, but LCD screens are by far the most popular and most widely used today. Plasma screens are still available but slowly fading away, and DLP projection screens are popular because of their low cost. There are other emerging technologies that we may see in the future. In any case, what we need is a light source to reproduce the varying intensities of red, green, and blue light.

Color video screens are made up of a very large number of tiny red, green, and blue light sources. These are used to make up the pixels or dots of light

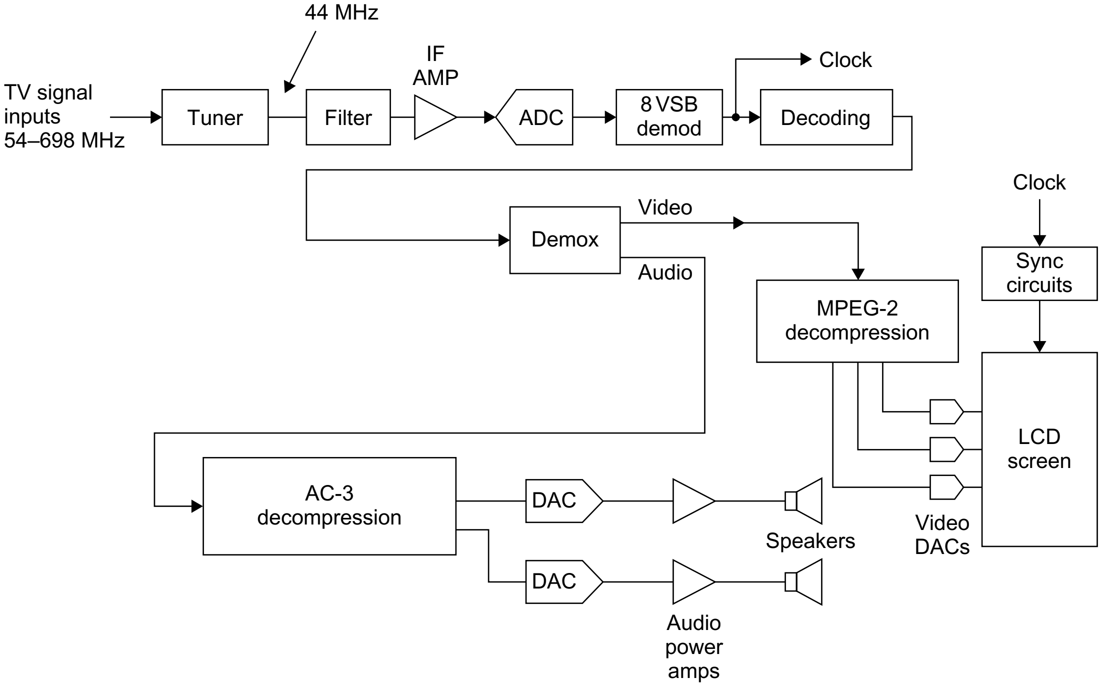
44 MHz
IF AMP
Clock
TV signal inputs 54–698 MHz
Tuner
Filter
ADC
8 VSB demod
Decoding
Clock
Demox
Video
Audio
Sync circuits
MPEG-2 decompression
LCD screen
AC-3 decompression
DAC
DAC
Speakers
Video DACs
Audio power amps

FIGURE 11.11 Simplified block diagram of digital TV set.

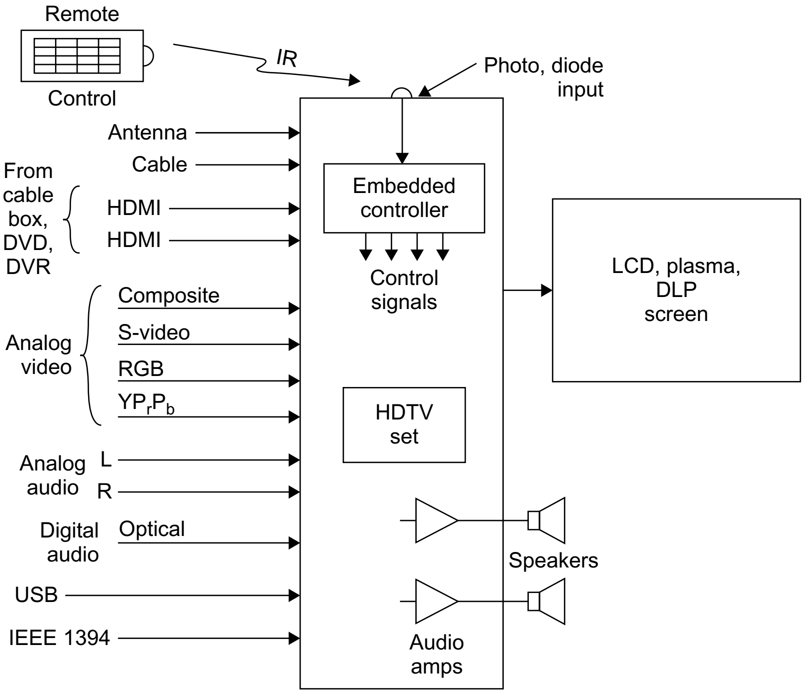
Remote
Control
IR
Photo, diode input
Antenna
Cable
From cable box, DVD, DVR
HDMI
HDMI
Analog video
Composite
S-video
RGB
YP_rP_b
Analog audio
L
R
Digital audio
Optical
USB
IEEE 1394
Embedded controller
Control signals
HDTV set
Audio amps
Speakers
LCD, plasma, DLP screen

FIGURE 11.12 Inputs and outputs of modern digital TV set.

that re-create the picture. These pixels are very small, so that your eye cannot actually see them even when you are relatively close to the screen.

Figure 11.13 shows some of the screen pixel formats in use today. Cathode ray tubes were made up of a group of three color phosphor as shown in Figure 11.13A. These color triads were grouped tightly together to form the complete screen. When an electron beam strikes each of the red (R), blue (B), and green (G) pixels, they cause the phosphor dots to glow. The eye automatically mixes the individual red, blue, and green colors to reproduce the desired color.

LCD and plasma screens use a pixel format somewhat like that shown in Figure 11.13B. These are usually rectangular red, green, and blue areas that are backlighted with the appropriate degree of intensity.

One pixel is not necessarily equal to the three color group shown in Figure 11.13A. Instead, 1 pixel as displayed on the screen might actually be made up of a collection of these individual groups. The actual pixel size may be something like Figure 11.13B, made up of multiple groupings of red, blue, and green elements as shown.

The actual format displayed on the screen is determined by the TV display standards. As indicated earlier, a standard DTV screen uses 480 horizontal lines with 640 pixels per line. That creates a screen with a total of 307,200 pixels. The 720 × 1280 screen puts 1280 pixels on 720 lines for a total of

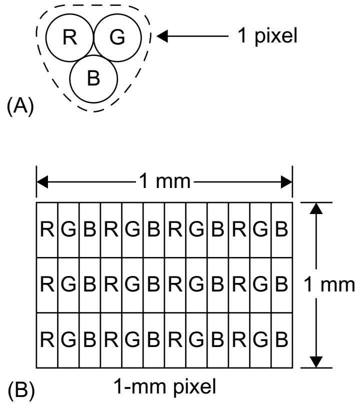

FIGURE 11.13 How color pixels are formed on a TV screen. (**A**) Pixel format for most CRTs. (**B**) Pixel format typical of LCD, plasma, or LED screens.

921,600 pixels. The big daddy of all screens uses 1080 lines with 1920 pixels per line for a total of 2,073,600 pixels. Just remember that the pixels are turned on one at a time from left to right and top to bottom on the screen at a very high rate of speed.

Summary of Screen Technologies

The actual technical details of each of the different types of screens are way beyond the scope of this book, but a summary might be useful. A quickie review of each of these screen types follows.

CRT—A cathode ray tube contains three electron-beam guns, each one driven by the analog color signals from the DACs. The intensity of the electron beam depends on the intensity of the light and color signal being fed to it. These electron beams are aligned so that they strike the red, green, and blue phosphor dots covering the screen. Each dot emits light with an intensity proportional to the strength of the electron beam striking it. The electron beams are scanned simultaneously across the screen at very high speed, stimulating the pixels in the right proportion to reproduce the desired color.

LCD—Liquid crystals are, as their name implies, a liquid made up of molecules whose physical orientation can be changed by the application of an electric signal. Different voltages change the polarization of the light through the liquid. By applying one level of voltage to the liquid crystal, it can be made almost totally transparent where maximum light passes through it. Another level of voltage will twist the molecules and practically block all light from being passed. In this way, the liquid crystal acts almost like a camera shutter but in this case with varying degrees of transparency being possible. The liquid crystal material is contained within glass plates. A bright light source such as a fluorescent lamp or high-brightness LEDs are placed behind the liquid crystal.

In front of the liquid crystal is another transparent sheet containing the red, blue, and green pixels described earlier. The whole thing is sandwiched together in a very thin screen. The electrical signals applied to a matrix of wires cause the liquid crystals to be properly oriented for each pixel and to allow the correct proportion of light through to each of the red, blue, and green light filters.

Plasma—A plasma screen is also made up of many tiny light sources. Each light source or subpixel is actually a small chamber containing xenon and neon gases. When a voltage is applied to each of these chambers, the gases ionize and emit ultraviolet light. The ultraviolet light then actually strikes a phosphor coated inside the chamber. The chamber then emits red, green, or blue light. The degree of ionization determines the brightness of each subpixel. In any case, electrical signals are then applied in sequence to the subpixels to turn them on with the correct degree of brightness to produce the desired color.

DLP—Digital light processing is a technique invented by Texas Instruments. The heart of it is an integrated circuit chip made up of hundreds of thousands or millions of tiny mirrors. The mirrors can be moved to one of two positions. This is an example of a *microelectromechanical system* (MEMS). MEMS is a technique for making mechanical moving objects using semiconductor manufacturing techniques.

Each of the tiny mirrors can be moved so that in one position it will reflect maximum light (representing white) and the other position where no light is reflected (representing black). You can then produce any shade of gray simply by switching the individual mirrors off and on at a high rate of speed. The light source is shined on the mirror and the resulting monochrome scene produced on the DLP chip is then focused on a screen from the rear by a lens system.

To produce color, the light source is passed through a rotating disc with alternating red, green, and blue segments. The speed of rotation and the frequency of each color on the disc allow each mirror to then reflect a specific color of light. In the newer DLP screens, high-intensity red, green, and blue LEDs are used to send light to the mirrors, which are then mixed and projected onto the screen.

LED—The newest TV screens are made of tiny red, green, and blue light emitting diodes arranged in a pixel matrix as described earlier. The LED screens are the brightest available and respond fast to action scenes. However, they are the most expensive screens.

3D TV

Yes, you actually can buy a 3D TV set that gives you that same 3D “coming-at-you” feeling of theater 3D. As this is written, these sets are very expensive and there is very little 3D material to see. Right now it is only available on DVD. Eventually it may be available by cable or satellite or over the Internet. And, yes, you still have to wear the special glasses. These glasses are not the red and green lens type but a special electronic version. The lenses use electrical polarization like a shutter that turns the left and right lenses off and on

Which Screen Type Is Best?

It is very difficult to find a new TV set with a CRT. You can still buy them, but they are typically only available in the smaller screen sizes such as 9, 13, and 19 inches.

LCD screens are available in a wide range of sizes, but the minimum today for a TV set is about 19 inches. Remember that these dimensions are the diagonal dimension from an upper corner of the screen to the lower diagonal corner. The most popular LCD screen size is 32 inches, but they are available in a wide range of sizes from about 42 to 65 inches. Even bigger ones are available. Plasma screens come in sizes from about 42 inches to about 70 inches. A DLP screen can be had in sizes from 50 to 100 inches.

As for prices, LCD and plasma screens are approximately comparable. DLP projection screens are significantly less expensive for the same size.

There are a couple of other factors to consider in evaluating screen types. As far as brightness goes, they are all very good, with the DLP probably slightly less bright than the other two types. Contrast, meaning the ratio of the brightest white to the darkest black, is also excellent on all types of screens, a little bit less so on the LCD.

Viewing angle means whether you can see the picture from any angle. Most screens are viewed directly straight on, but many viewers are off to the side or at an angle. DLP screens generally have poor viewing angles: when you get too far off to the side, you will not be able to see the picture or it will fade away. LCDs have this problem but to a lesser extent. Plasma screens have the widest viewing angles.

Finally, the ability of a screen to produce high-speed motion is generally excellent on plasma and DLP screens. LCDs did produce some smearing, but recently the higher scan rates of 120 and 240 frames per second have essentially eliminated this problem.

LCD screens dominate TV today and are gradually replacing all other types as their price continues to drop with better manufacturing methods. Plasmas and DLP are still available but are gradually fading. LED screens are growing in popularity and will become more widely used as prices come down in the future.

in synchronization with the two alternating sets of frames on the screen. A wireless signal (IR or radio) signals the glasses when to switch. The switching occurs at very high speed so the effect is to see the 3D scene. Only time will tell if this is just a gimmick or niche TV product.

CABLE TELEVISION

It is estimated that about 15% of all U.S. citizens get their TV over the air with an antenna. But over 70% get TV via cable. The remaining 15% get satellite TV. Why? Mainly because cable offers not only better, more reliable signals than by antenna, but also offers hundreds of channels of programming. But that's not all. Most of those who have cable also use the cable connection for

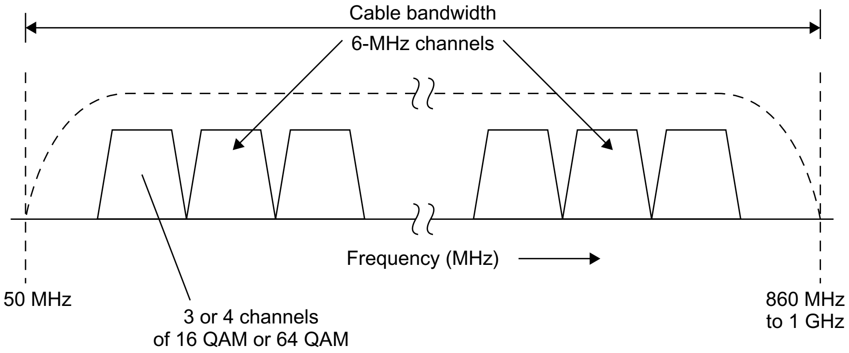

FIGURE 11.14 Coax cable bandwidth up to 1 GHz can contain about 130 6-MHz channels.

high-speed Internet access service. Of course, you pay for all that, and it has become like a monthly utility bill. With the cable industry being so big, it pays to know a little more about it.

It all starts with a facility called the *cable head end*. Another name you might hear is cable modem termination system (CMTS), which specifically is the equipment at the head end. This is a place where the company collects video and programming from many different sources. It snatches signals off the air, it receives premium programming like HBO, ESPN, Showtime, and Fox via satellites, stores movies on its large computers called video servers, and may even do some of its own local programming. The result is a hundred channels or more.

Now each of those TV signals—video plus audio—is modulated onto carriers that are spaced 6 MHz apart just like the regular over-the-air TV channels. Then all of those modulated carriers are added together to form a very complex signal. That signal is then amplified and put on the outgoing cable. Figure 11.14 shows what the spectrum looks like. Most cables have a bandwidth from about 50 MHz up to about 860 MHz to 1 GHz, which will accommodate up to about 130 6-MHz channels. We call this process frequency-division multiplexing, where a single cable can carry many separate channels of data, each on a different frequency and bandwidth. And since the cable is shielded, it does not radiate, so it, in effect, duplicates the free-space electromagnetic spectrum.

Since most TV today is digital, digital modulation methods are used to put multiple TV signals within each 6-MHz channel. Using 16-QAM or 64-QAM, a cable company can squeeze several channels of video/audio into each 6-MHz chunk of spectrum, providing hundreds of additional slots for programming. That is why cable companies can advertise up to 500 channels of entertainment, news, and so on.

All of the digital video/audio signals are then remodulated onto different channel assignments. All of these 6-MHz channels are added together or linearly mixed in a multiplexer (MUX) to create a massively complex composite signal. This is what goes out on the cable (see Figure 11.15).

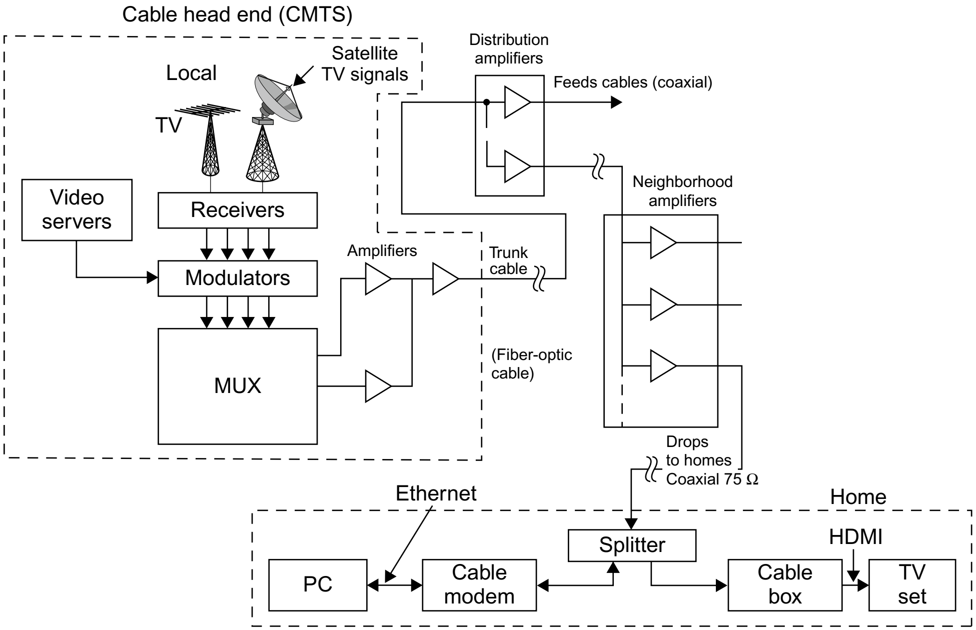
Cable head end (CMTS)
Local
Satellite TV signals
TV
Video servers
Receivers
Modulators
MUX
Amplifiers
Trunk cable
(Fiber-optic cable)
Distribution amplifiers
Feeds cables (coaxial)
Neighborhood amplifiers
Drops to homes Coaxial 75 Ω
Ethernet
Home
HDMI
PC
Cable modem
Splitter
Cable box
TV set

FIGURE 11.15 General block diagram of modern cable TV system.

The signal distribution network is called a *hybrid fiber cable* (HFC) system. This multiplexed signal is applied to a fiber optic cable, called a *trunk cable*. It is routed around the city and it terminates at selected points where the signal is divided into multiple paths of coax cable. These so-called *feeder cables* are connected to amplifiers along the way to overcome the large attenuation of the cables. At some point, the feeder cables come into your neighborhood and are further amplified and split into multiple signals. These signals are then sent to your house by a coax drop cable. The most common type is 75-ohm RG-6/U coax with F-connectors. The cable comes into your home and is usually split two or more times and sent to wall connectors in the various rooms. These connectors connect the signal to a cable box and/or a cable modem for the PC.

The cable box is like a tuner in that its main job is to select one of the many channels and send it to your TV set. Figure 11.16 shows a simplified block diagram of a typical cable box, usually called a set-top box (STB). The input is a TV tuner with channel selection via a frequency-synthesizer local oscillator by way of your remote control. The selected signal is then demodulated and decompressed by a digital media processor chip, and then sent to the TV set via one of the many possible interfaces, usually the HDMI or *high-definition multimedia interface*. This is a very fast digital interface with special cable and connector. Most TV sets have this as the most common input. Other video format interfaces such as RGB, PrPbPy, or S-video, are usually provided to accommodate older analog TV sets. An additional feature of many of the newer STBs is a built-in digital video recorder (DVR). This is essentially a computer hard disc drive that records the program you want to save for viewing later.

Other Wired TV Distribution Systems

There are a couple of other ways that digital video is distributed to homes. One is a fully fiber optic system. The fiber cables go all the way from the head end to the homes. This was once too expensive, but today, thanks to improved fiber and other technologies, it is cost effective. Such systems are generally referred to as *passive optical networks* (PONs). "Passive" simply means that the signals are strong enough and the cable runs short enough so that no amplification is needed along the way. The digital video and audio are modulated onto infrared (IR) light beams and sent down the cable. Because fiber has such wide bandwidth, data rates are extremely high, over 1 GB/second in the newer systems. This allows not only digital video to be transmitted but also very-high-speed Internet service. Such systems use standards like Ethernet PON or GPON for gigabit PON. A typical U.S. system is Verizon's FiOS.

Another new system is called IPTV for *Internet protocol TV*. This is video distribution via putting the compressed video on a fiber optic cable using the familiar TCP/IP protocols. The signals travel via fiber to the neighborhoods. From there they are split off and the signals travel via in-place twisted-pair telephone lines.

(*Continued*)

Other Wired TV Distribution Systems (Continued)

New forms of modems using a form of OFDM called *discrete multitone* (DMT) transmit the signal over the twisted pair at high speeds. These lines are now used just for high-speed Internet service and are referred to as *asynchronous digital subscriber lines* (ADSL or just DSL). Newer versions called ADSL2 and ADSL2+ or VDSL for video DSL provide data rates from 26 to 52 Mbps on standard phone lines. An example of a system in the United States is AT&T's U-verse system.

For Internet access, you connect a cable modem to the coax cable and it automatically selects the channel where the data services are assigned. The modem goes to your computer via an Ethernet cable. Access to the Internet is by way of your computer's browser.

SATELLITE TV

Satellite TV technically is also over-the-air TV or radio, but instead of the signals coming from one or more local TV stations, broadcast towers, the signals emanate from a geosynchronous satellite 22,300 miles away orbiting around the Equator (see Figure 11.17). The satellite transponders operate in the Ku band with frequencies of 10.95 to 14.5 GHz. This system is referred to as *direct broadcast satellite* (DBS).

The satellite TV company gets its TV like the cable companies, and brings it all together at a single Earth station head end where it is all collected and transmitted on an up-link to the satellite in the 14- to 14.5-GHz range (see Figure 11.18). The signals are then amplified in the satellite and rebroadcast on the down-link frequencies at 10.95 to 12.75 GHz. A high-gain antenna focuses the signals on the United States or other country of interest.

Down on Earth, satellite TV subscribers have a special receiver and antenna. These are illustrated in Figure 11.19. The antenna is usually a small dish or parabolic reflector antenna about 18 inches in diameter. The antenna is usually a small horn. Also inside the antenna assembly is a low-noise amplifier (LNA) to boost the small signal from the satellite. This is then sent to a mixer along with a local oscillator signal to down-convert the signals into the 950- to 2100-MHz range. This signal is then routed by coax to the satellite receiver. The LNA/mixer assembly at the antenna is called a low-noise block (LNB) converter. This receiver signal processing at the antenna is necessary because the Ku band frequencies of 10.95 to 12.75 GHz are so high they would be far too greatly attenuated by the antenna coax cable to the receiver. So down-converting to the lower frequency allows coax to be used to the receiver.

The satellite receiver is like the cable box in that it is used to select the desired channel and demodulate it for the TV set. The satellite signals are all digital and transmitted in packets 147 bytes long. A satellite receiver is shown in Figure 11.20. MPEG-2 video decompression is used. The recovered signals

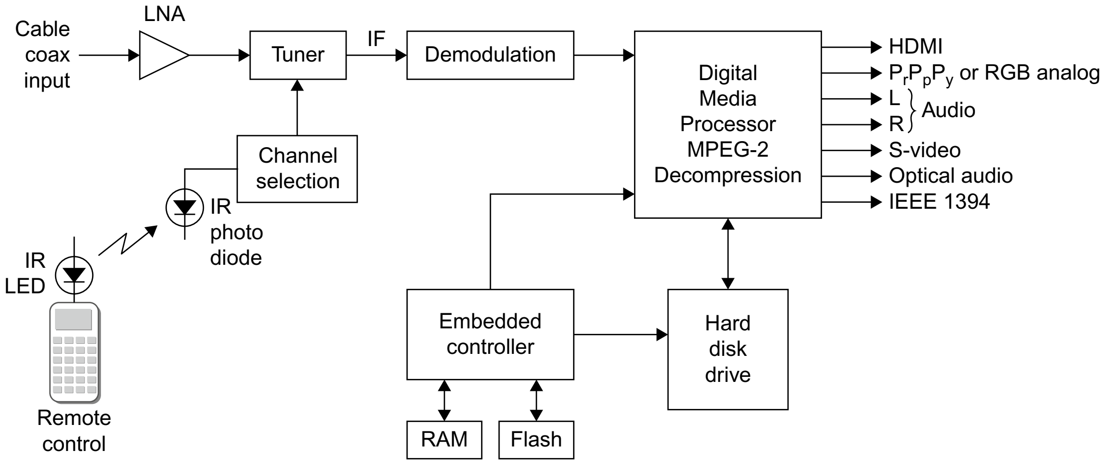
Cable coax input
LNA
Tuner
IF
Demodulation
Digital Media Processor MPEG-2 Decompression
HDMI
$P_rP_pP_y$ or RGB analog
L
R
Audio
S-video
Optical audio
IEEE 1394
Channel selection
IR photo diode
IR LED
Remote control
Embedded controller
Hard disk drive
RAM
Flash

FIGURE 11.16 Simplified block diagram of modern cable box with DVR.

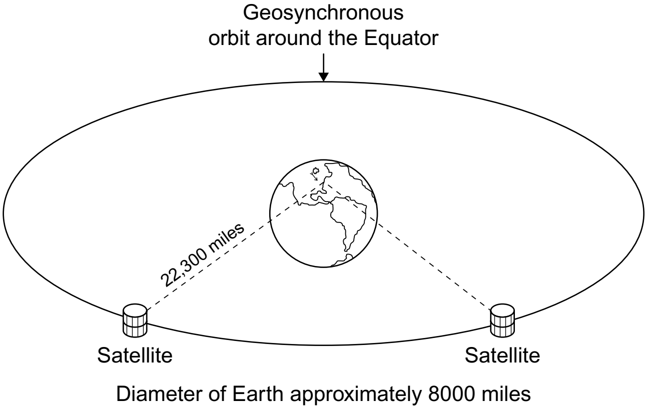
Geosynchronous
orbit around the Equator
22,300 miles
Satellite
Satellite
Diameter of Earth approximately 8000 miles

FIGURE 11.17 TV distribution satellites in geosynchronous orbits act as wireless relay stations.

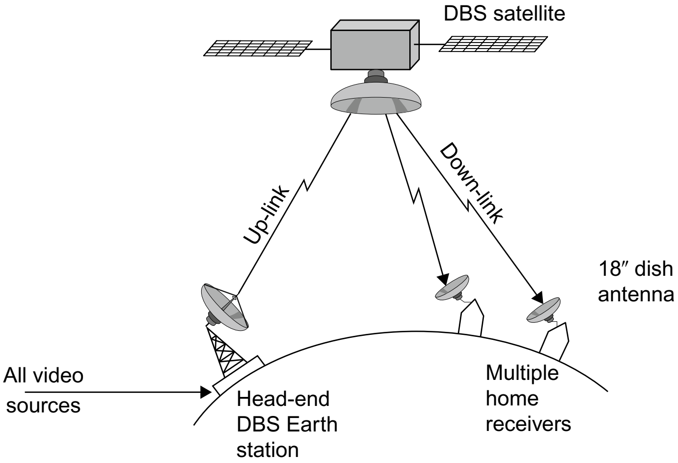
DBS satellite
Up-link
Down-link
18″ dish
antenna
All video
sources
Head-end
DBS Earth
station
Multiple
home
receivers

FIGURE 11.18 Video from all sources is transmitted up-link to satellite that rebroadcasts it back to Earth over down-link to home receivers.

are then sent to a digital TV receiver, typically via an HDMI port. Or in older systems, the video and audio are remodulated on a channel 3 or 4 TV carrier and sent to an analog TV set.

CELL PHONE TV

You knew it would come. Actual TV on your cell phone. You can easily get video via the Internet connection on a 3G phone over the network or via a Wi-Fi hot spot. Typical video is like that from YouTube or just video clips that users attach.

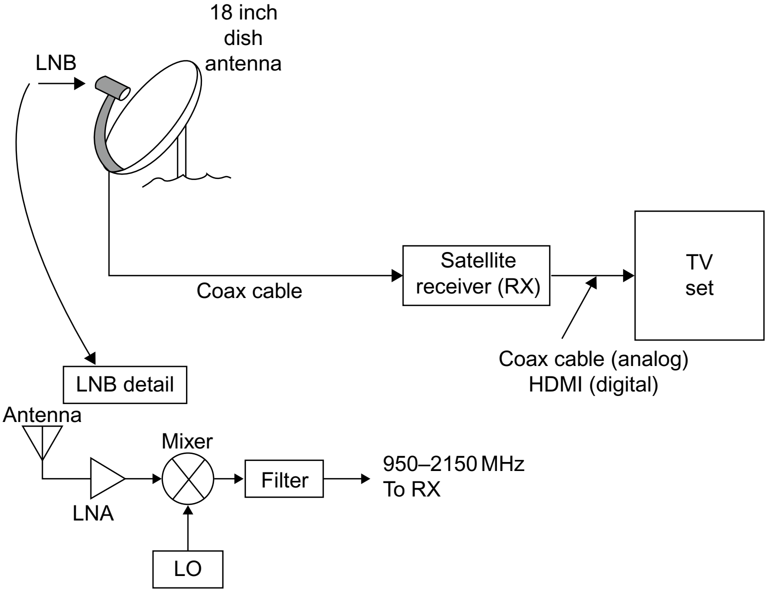

FIGURE 11.19 Satellite TV home system consists of antenna with low-noise block converter, receiver, and TV set.

Some cellular carriers also offer several video channels over the network for an extra fee each month. Such video is available from AT&T and Verizon, and consists mainly of news, weather, sports, and a few comedy channels.

Since video is so data intensive and requires lots of extra bandwidth and high speeds, too much video service can overload most cellular networks. It has been decided that the best way to deliver TV service to a cell phone is to broadcast it. Several methods are being implemented. For example, in Europe the DVB-H system is being used to broadcast programming to handsets. Similar systems in Japan and Korea are also used. Both are digital. In the United States, MediaFLO broadcasts video to handsets on the old TV channel 55 (716 to 722 MHz). It uses OFDM and about 15 to 20 channels are available. The service is provided for a monthly fee.

In the near future, standard ASTC digital TV will be available in the United States. A special version for mobile handsets has been developed and will likely be adopted by existing TV stations for broadcast to cell phones. This is expected to be free TV, but no doubt ads will be used to pay the bills. If making a phone call or texting by cell phone is distracting to drivers now, wait until TV is available. (Now that is something to look forward to.)

CLOSED-CIRCUIT TV

Closed-circuit TV (CCTV) is used primarily for surveillance. You have seen the small wall- or ceiling-mounted TV cameras in stores, airports, companies,

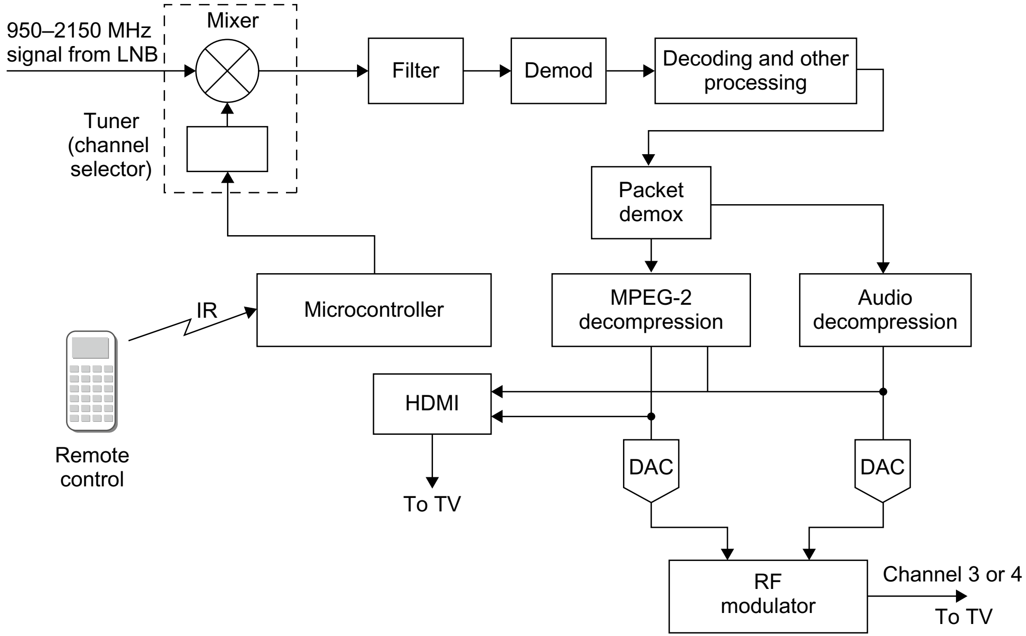
950–2150 MHz signal from LNB
Mixer
Tuner (channel selector)
Filter
Demod
Decoding and other processing
Packet demox
Microcontroller
IR
MPEG-2 decompression
Audio decompression
HDMI
To TV
DAC
DAC
Remote control
RF modulator
Channel 3 or 4
To TV

FIGURE 11.20 Simplified block diagram of satellite TV receiver.

and today on the streets. The video signals are routed to a central point where multiple TV sets or monitors are set up for observation.

In many installations, the video is transmitted directly over coax cable to the monitors. There is no modulation involved. However, in some cases, when remote cameras are used, a wireless link via 2.4 GHz is typical. Of course, modulation is then used in the transmission so that a receiver is needed at the viewing end to recover the video.

Many installations also have video recorders (VHS tape or DVD) to capture several minutes or hours of video. In the past, most cameras were black and white, but today more and more use color.

DIGITAL VIDEO DISCS

The formal meaning of DVD is *digital versatile disc*. Its main use is video but it can also store computer data or audio. It is the storage medium of choice for movies, music videos, and other video programming. It has virtually replaced the VCR.

The basic DVD storage medium is a disc that is almost identical to the CD discussed in Chapter 10. The disc is the same size (120-mm diameter), but it is formatted differently to store many more bits and bytes of data. While the CD can store up to about 700 MB, the DVD can store about 4.7 GB, that is, gigabytes or billions of bytes. And that is just the basic one-sided disc. There are also single-side/double-layer discs, double-sided/single-layer, and double-sided/double-layer versions that can store even more, up to 15.9 GB.

The storage method is the same. A laser beam burns pits into a spiral track. The data is represented as pits and lands as described earlier in the CD discussion (see Figure 10.6). The big difference is that the size of the pits is significantly smaller and the track spacing is less than half the CD spacing, that is, 1.6 microns (1600 nanometers) versus 740 nanometers (nm). A nanometer is 1-billionth of a meter. With smaller pits and tighter spacing, the disc can hold significantly more data.

But that is not the whole story. The video is also compressed. Unlike the audio on a CD, which is not compressed, digital video is compressed with the MPEG-2 standard described earlier. This gives roughly a 40-to-1 compression. In that way, a full 2 hours of video can be stored.

As for audio, the format is as described in Chapter 10 for CDs, but DVD audio is Digital Dolby Surround Sound 5.1 using six channels of 16-bit, 44.1-kHz digital sound. Audio DVDs are also available with 192-kHz sampling of the audio with 24-bit ADCs, giving even greater fidelity and low-noise performance.

How a DVD Player Works

Figure 11.21 depicts the block diagram of a typical DVD player. First, note the motor that drives the disc. It is a variable-speed device that changes speeds as the pickup mechanism moves from inside to outside as it reads the spiral track.

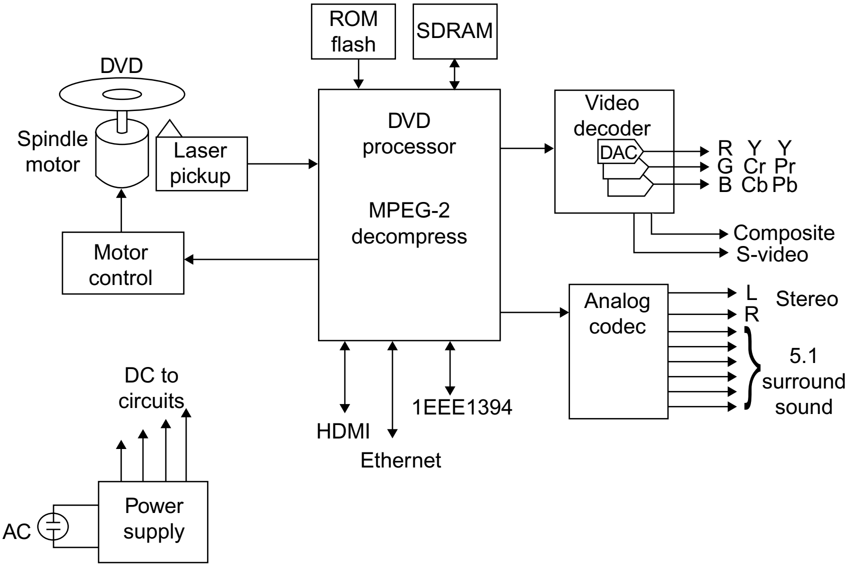
ROM flash
SDRAM
DVD
Spindle motor
Laser pickup
DVD processor
MPEG-2 decompress
Video decoder
DAC
R Y Y
G Cr Pr
B Cb Pb
Composite
S-video
Motor control
Analog codec
L
R
Stereo
5.1 surround sound
DC to circuits
HDMI
Ethernet
1EEE1394
AC
Power supply

FIGURE 11.21 Block diagram of standard DVD player.

The pickup head uses a laser that shines on the disc to read the pits and lands. Because of the closer spacing, a higher-frequency (shorter wavelength) laser is used. CD players use a 780-nm laser, while DVD players use a 640-nm laser that has a smaller focused beam.

The photodetector in the head picks up the laser reflections and sends the 10 Mb/second data stream to the processing chips. In some players, a single-chip application-specific device handles all of the digital processing from demultiplexing, to error detection and correction, decoding, and MPEG-2 decompression. All this may be in separate chips in some players.

In any case, the MPEG-2 data goes to a video decoder and to multiple DACs that re-create the original analog RGB signals. Various video formats are used, such as RGB, YCrCb, and YPrPb, each of which needs three DACs each. Most DVD players also decode regular composite video and S-video.

The separated audio signals go to an audio codec that contains the DACs which recover the audio content either in basic stereo or 5.1 surround sound. As for interfaces, an HDMI output is provided. Most players also have an Ethernet port, IEEE 1394, and USB connections.

Don't forget that some of the later models are also DVD recorders. They play DVDs but also burn new ones with video from the TV set, cable box, satellite box, or a built-in tuner.

A Word about Blu-ray

Blu-ray is the next generation of DVD medium and players. The discs are the same size but thanks to even smaller pits, lands, and track spacing, even greater storage capacity is available. With 320-nm track spacing and a 405-nm (blue light wavelength) laser, the single-sided disc can hold up to 25 GB of data. That is enough to store two full-length high-definition movies or 13 hours of standard-definition video. A double-layer version is also available that doubles those storage figures. Blu-ray uses MPEG-2 compression, but also supports other video compression standards such as MPEG-4/H.264 and VC-1.

For a while, there were several other video disc standards competing for dominance. The most notable was HD-DVD, which did bring players to market. Most movie companies and other media sources selected Blu-ray as the distribution medium, making Blu-ray players the de facto standard and eliminating HD-DVD.

Project 11.1

Look at a Pixel

Get a magnifying glass and hold it up to your TV screen to see the individual color pixels or subpixels. Are they triads or rectangular areas?

Project 11.2

Diagram Your Own TV System

Draw a block diagram of your own home TV setup. Just draw a box for each and indicate how each of them is connected. Identify the type of cable, connector, and interface between each box.

Project 11.3

Try Over-the-Air TV

If you get your TV by cable or satellite, try receiving local channels on an antenna. Use an indoor antenna for simplicity. You can buy a set of "rabbit ears" from a local Radio Shack. Or try an outdoor antenna from Radio Shack. You can get some channels on a simple loop of wire 1 yard long.

If you still have an older analog TV, you probably also have a cable box and went through this process before.

Chapter 12

Industrial Control

How to Automate: Monitor, Process, and Control

In this Chapter:

- Control defined.
- Open and closed control systems.
- Sensors and transducers.
- Actuators like relays, solenoids, and motors.
- Thyristors.
- Programmable logic controllers.

INTRODUCTION

Industrial control is one of the four major applications of electronics. In industrial control, electronic components, circuits, and equipment are used to operate various types of machines in manufacturing plants. Most industrial machines like robots are mechanical in nature but electrically operated. Most are powered by electric motors or other actuators. To operate the machine, electrical power must be properly controlled. Industrial control electronics is used to turn machines off and on at the appropriate time, control their speed of operation, and otherwise produce the desired manipulations.

Electronic control is not confined to industry. There is lots going on in the home and car as well. The principles given here apply to those applications as well.

OPEN- AND CLOSED-LOOP CONTROL

The basic industrial control process is illustrated in Figure 12.1. Electrical inputs, such as switches, sensors, or other devices, are used to initiate an electronic control process. The signal-processing circuits generate control output signals that are used to operate the industrial machines. This basic process is known as open-loop control. The typical device being controlled by the output signals is an electric motor that operates a machine. Other outputs could operate lights, relays, solenoids, or a variety of other devices.

© 2010 Elsevier Inc. All rights reserved.
doi: 10.1016/B978-1-85617-700-9.00012-6

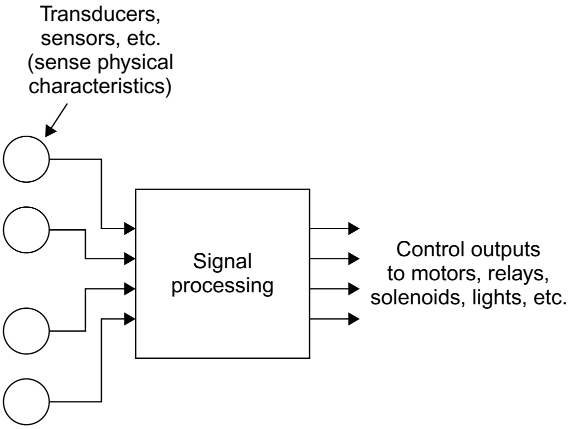
Transducers,
sensors, etc.
(sense physical
characteristics)
Signal
processing
Control outputs
to motors, relays,
solenoids, lights, etc.

FIGURE 12.1 General block diagram of industrial control process.

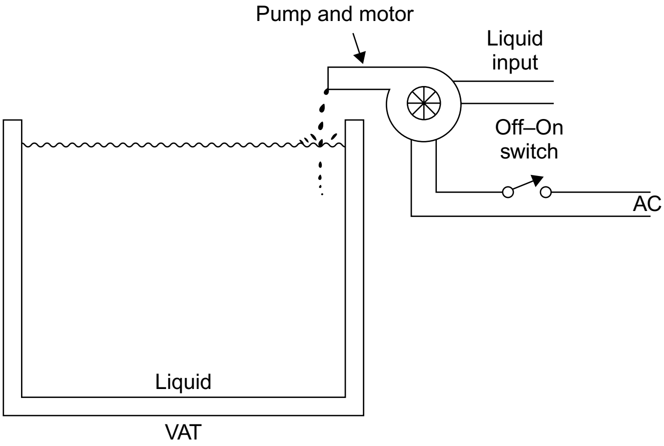
Pump and motor
Liquid
input
Off–On
switch
AC
Liquid
VAT

FIGURE 12.2 Open-loop control system for filling vat with liquid via a pump.

Open-Loop Control

A simple example of an open-loop control system is shown in Figure 12.2. The input is a switch that applies electrical power to a motor. The motor, in turn, operates a pump that causes liquid to be put into a vat or tank. As soon as the vat is full, the operator turns the switch off. The basic control process in this system is simply turning power off and on to control the pump. The operator visually monitors the level in the vat until the correct level is reached.

Closed-Loop Control

An improved form of industrial control system is illustrated in Figure 12.3. It is known as a closed-loop control system. The system is given an initial input

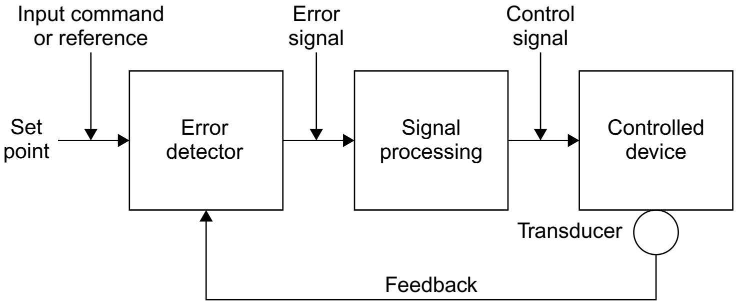

FIGURE 12.3 Generic diagram of any closed-loop control system with feedback.

reference called the set point. The set point is usually a voltage that represents some physical value that is to be achieved. It is applied to an error detector that compares this input to a signal from the controlled devices. If the two signals are different, an error output is produced. This error signal is processed (usually amplified), and an output control signal is derived from it. It is this output signal that operates the controlled device. A transducer monitors the controlled device to see that the desired outcome is obtained. The signal from this transducer is called *feedback*. The feedback signal is compared to the input to produce the error signal. The control signal operates the controlled device until the error is reduced to zero. At that time, the desired output condition, where the feedback equals the set point, is obtained.

The key to closed-loop control is the feedback. Feedback signals tell the system whether the machine is performing correctly. The feedback signal allows the system to adjust itself in such a way that the desired output is continuously accomplished. Control is automatic.

A closed-loop version of our system to fill a vat with liquid is shown in Figure 12.4. An initial input signal is applied to start the process. If the vat is not full, the transducer will sense it and send a signal to the error detector that says the liquid level is low. The error detector generates an output signal that is processed and used to start the pump motor. The motor rotates the pump, which puts liquid in the tank. When the vat is full, the sensor will generate a signal that is sent to the error detector. The error is now zero, so the pump motor is turned off.

The primary benefit of a closed-loop control system is that its operation is automatic. In the open-loop system, an operator has to turn the system off and on manually to fill the tank. In the closed-loop control system, the operator gives an initial input, called the set point, to which the error detector will compare the feedback signal from the sensor. He or she turns the system on. From that point on, operation is automatic. The pump turns itself off when the vat is full. If the liquid level drops below the desired level due to usage, the sensor indicates a low level and automatically turns on the motor so that the pump will again fill the tank. When the tank is full, the unit shuts off by itself.

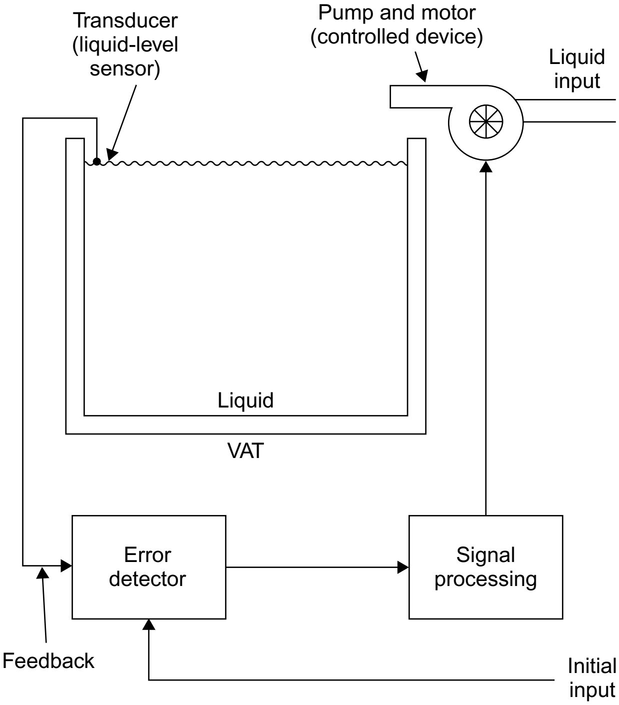
Transducer (liquid-level sensor)
Pump and motor (controlled device)
Liquid input
Liquid
VAT
Error detector
Signal processing
Feedback
Initial input

FIGURE 12.4 Filling vat with liquid automatically with closed-loop control system.

Most industrial control systems are closed loop to make them automatic. And your home has some as well. Your toilet flusher is a closed loop as is your heating and air conditioning (HVAC) thermostat. Appliances like your washer, dryer, and dishwasher are other examples of closed-loop control. All of them contain a microcontroller with sensors, timers, and feedback to make them totally automatic.

Controllers

While many industrial control applications are implemented with simple control circuits, most industrial control is carried out by specialized instruments called controllers. These units accept inputs from sensors, condition them, process them, and then generate output control signals.

A good example of an analog controller is one used to control the temperature in a system. A resistive temperature sensor is used to monitor the temperature of the liquid in a pipe. If the temperature goes up, the sensor output goes up. This is sensed by the controller circuitry. Inside the controller, a comparator is set to a desired temperature threshold. Should the temperature exceed a specified set point, the controller will turn off the device heating the

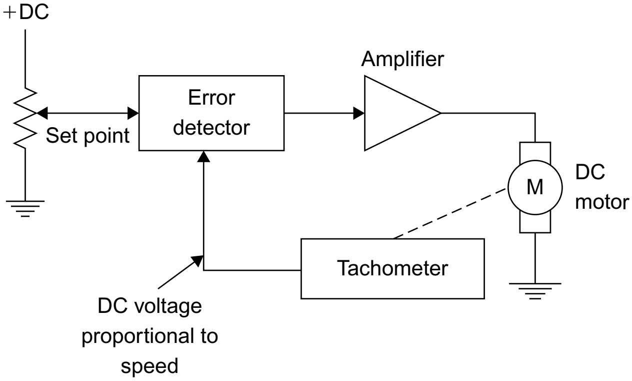

FIGURE 12.5 Closed-loop control system for motor speed control and regulation.

pipe and liquid. Should the temperature drop below another set point temperature, the controller will turn on the heating device in an effort to maintain the set point temperature. While the heating element will be turning off and on, it will produce the desired average overall temperature.

A continuous analog controller will provide proportional control rather than an off/off response. For instance, suppose that you want to control the speed of a motor. The system depicted in Figure 12.5 does this. You want the motor to stay set to the speed you want, say 1200 rpm. A pot connected to a voltage represents that speed value. The voltage is applied to an error detector along with the feedback input from a tachometer. A tachometer is a device that senses rotational speed and develops a proportional DC voltage. That voltage is sent to the error detector and compared to the set point. If they are different, an error voltage is developed. This voltage is then amplified and applied to the motor. The speed of a DC motor is proportional to the voltage applied to it. So if the motor slows down, the error voltage is increased and that is amplified and applied to the motor speeding it up to compensate for the speed drop. A speed up in the motor produces an increased tachometer output. The resulting error signal causes the motor speed to drop until the error is zero. With such a circuit, the motor speed remains constant.

An analog controller is a product made up of the set point pot, error detector, and a selection of amplifiers whose gain can be varied and an output amplifier that can drive the motor or other device. Many different types of commercial electronic controllers have been developed to monitor and control physical characteristics such as pressure, physical strain, weight, liquid flow rate, physical position, or liquid level in a tank. Digital controllers that contain a microcomputer are also available.

SENSORS

An essential element of most control systems is the inputs from sensors. Sensors are the components that detect physical changes or events and convert them into

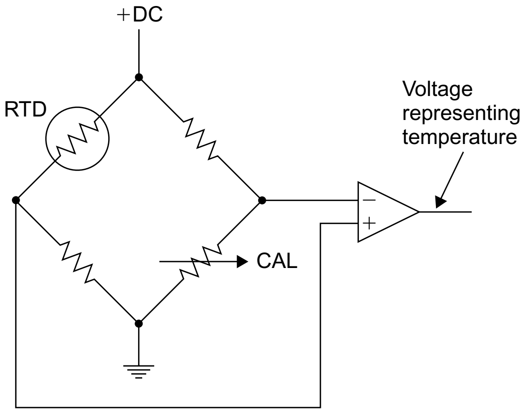

FIGURE 12.6 Most resistive sensors like an RTD or thermistor are placed in a bridge circuit to condition them to provide amplified output voltage.

electrical signals to be processed. Sometimes the term transducer is used to refer to the sensor. A transducer is a device that converts one type of energy into another such as mechanical energy to electrical. There are literally hundreds of different types of sensors. A few examples of the most common types follow.

Temperature Sensors

Temperature is probably the most often sensed physical characteristic. Temperature sensors convert temperature to a resistance change or voltage change. The most commonly used temperature sensors are RTDs, thermistors, and thermocouples.

RTD—An RTD is a resistive temperature device. It is essentially just a platinum wire whose resistance varies with temperature. Most of them have a resistance of 100 ohms at zero degrees Celsius (0°C). As the temperature changes, the resistance changes to reflect the change. RTDs have what we call a linear positive temperature coefficient. That is, as the temperature goes up the resistance goes up and vice versa.

To convert the resistance variation into a voltage, the RTD is usually put into a bridge circuit as shown in Figure 12.6. The bridge is initially balanced with the variable calibrate control (CAL) to produce zero output at a given temperature. Then if the temperature varies, the bridge becomes unbalanced and a voltage is produced. This is then amplified into a larger voltage that is a measure of the temperature. The output can be measured with a meter or it can be digitized for use as in input to a digital system for display or control.

Thermistor—A thermistor is a resistor whose resistance has a negative temperature coefficient. As temperature goes up the resistance goes down and vice versa (see Figure 12.7A). The resistance change is much greater than that

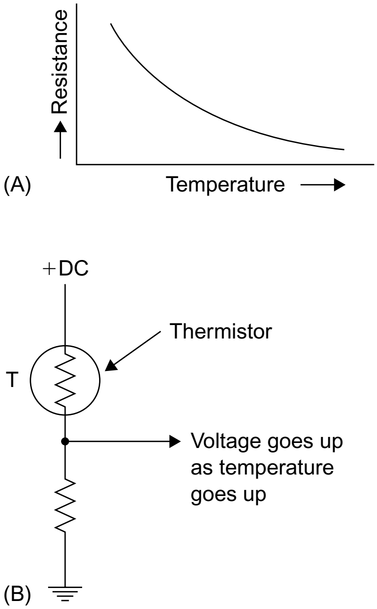

FIGURE 12.7 Characteristic curve of thermistor and simple voltage divider circuit producing output voltage.

of an RTD. The thermistor can be used in a simple voltage divider to develop an output voltage or in a bridge circuit like the RTD (see Figure 12.7B).

Thermocouple—A thermocouple is a unique type of temperature sensor, as it develops a voltage rather than a resistance change. It is formed with two dissimilar metals. If the junction of the metals is heated, a voltage is developed. The voltage is usually in the millivolt range so that it usually has to be amplified before it becomes useful. The thermocouple's main advantage is its accuracy at very high temperatures.

Solid State—Diodes and transistors make good temperature sensors. A silicon diode has a very linear voltage drop variation that decreases with an increase in temperature. This is true of the emitter-base junction of a bipolar transistor as well. Zener regulator diodes exhibit a similar response. The sensor is usually packaged into a complete integrated circuit with an amplifier that produces a voltage output proportional to the temperature variation.

Pressure Sensors

A pressure sensor responds to force or pressure. There are many different types. A widely used kind is called a *strain gauge*. A strain gauge is essentially a thin pattern of metal deposited on a plastic base like that shown in Figure 12.8. It has a specific value of resistance. A value of 120 ohms is common. The strain

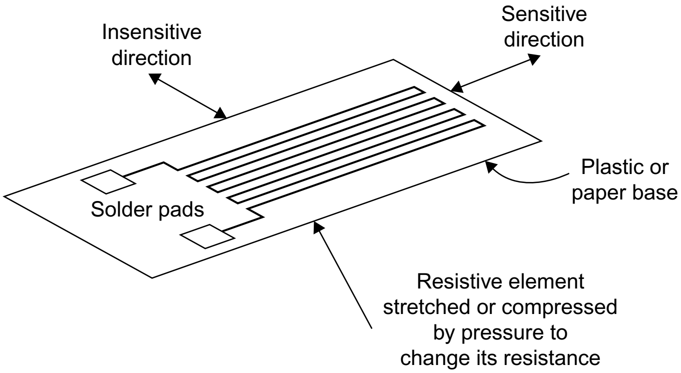
Insensitive direction
Sensitive direction
Plastic or paper base
Solder pads
Resistive element stretched or compressed by pressure to change its resistance

FIGURE 12.8 Strain gauge is a resistive element whose value changes with stress or pressure.

gauge is then glued or cemented with an adhesive to the object to which the pressure is to be applied. An example is a steel beam. When pressure is applied to the beam, it will bend. As it bends, it stretches or compresses the strain gauge resistance element. If the resistive element is stretched, its resistance increases. If it is compressed the resistance decreases. Knowing the specific characteristics of the resistance change per pound of pressure, an exact measure of the pressure can be determined. Mostly strain gauges are used in a bridge circuit like that shown in Figure 12.6. Since the resistance variation is small and the resulting bridge circuit output is small, a special amplifier called an *instrumentation amplifier* is used. It has differential inputs and very high gain.

A special pressure sensor called a *load cell* is typically made up of strain gauges. It is fully calibrated to provide a specific voltage output with a specific pressure. Load cells are widely used for weight measurement. An example application is the use of load cells in truck weighing scales.

There are other forms of pressure transducers. Many are solid-state devices with an output voltage proportional to pressure. *Piezoelectric pressure transducers* use the voltage produced by certain types of crystals or ceramics when pressure is applied to them.

Switches

Switches are widely used as sensors. Switches are mechanical devices with electrical contacts that open or close with a particular physical movement. Switches are used to detect mechanical movement and specifically to determine when the limit of some physical device has been reached. Called *limit switches*, these devices have various arms and mechanisms that permit ease of use in a variety of situations. Examples of such switches are those that can detect when a cabinet door is open or closed or a switch that senses when a mechanical device reaches a specific point where motion must be stopped. And

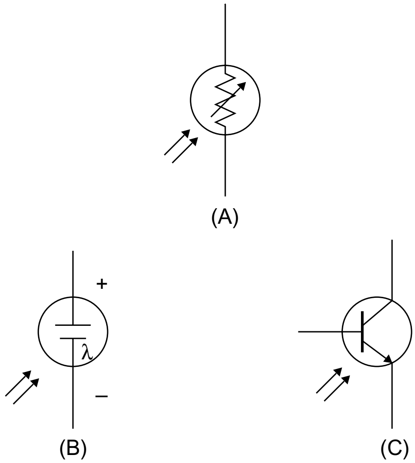

FIGURE 12.9 Schematic symbols for photo cells. **(A)** Photoconductive or photoresistive. **(B)** Photovoltaic or solar cell. **(C)** Photo transistor.

of course, a human-operated switch may be the most common. It is usually a button that is pushed to initiate or stop some action.

Potentiometers

A potentiometer is a variable resistor or variable voltage divider. While most electronic pots are rotary devices, they also come in a linear variety. The variable arm of the pot is controlled by some external mechanical device. Then by using the resistance value of the pot or the voltage it produces as a variable voltage divider, it can be used to sense mechanical position.

Photoelectric Sensors

One of the most widely used industrial transducers is a photoelectric sensor. A photoelectric sensor is a component that responds to light and produces an electrical signal that can initiate some operation.

There are three basic types of photoelectric sensors. These are the *photoresistive* or *photoconductive cell*, the *photovoltaic cell*, and the *phototransistor*. The photoconductive sensor is a light-sensitive resistor. Its schematic symbol is shown in Figure 12.9A. The resistance of the device varies with the amount of light falling on it. With no light falling on the sensor, its resistance will be very high, 100,000 ohms or more. When bright light shines on the sensor, however, its resistance will drop to a very low value, usually several hundred ohms or less. Of course, at a light level between bright and dark, the resistance of the device will be somewhere between 100 ohms and several hundred thousand ohms. The resistance is inversely proportional to the light level.

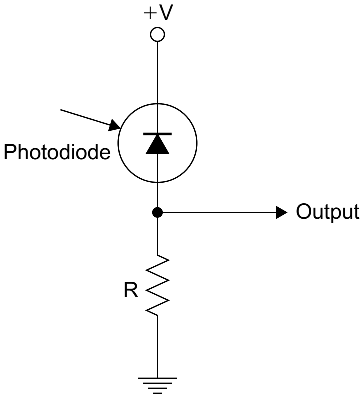

FIGURE 12.10 Photodiode and its simple output circuit.

Another type of photoelectric device is the photovoltaic cell, whose schematic symbol is illustrated in Figure 12.9B. This is the solar cell described in an earlier chapter. Whenever bright light falls on the cell, it will generate a small DC voltage. The maximum output of a typical photovoltaic cell is approximately 0.45 to 0.5 volt. At lower light levels, the output voltage will be less. Photovoltaic cells are used primarily in power generation systems for producing charging voltage for batteries in satellites and other remote systems. Multiple cells are connected in various series and parallel combinations to get the desired voltage level and current capacity. In some applications, the photovoltaic cell can be used for industrial control. These are also found in portable calculators. Two AA cells at 1.5 volts each produce 3 volts in series. To generate 3 volts with 0.5-volt solar cells, you would need to connect six in series ($6 \times 0.5 = 3$ volts).

A silicon diode can also be used as a light detector. Special diodes in clear housings are often used to detect light. These are called photodiodes. Photodiodes are operated in the reverse-biased direction, as shown in Figure 12.10. With no light falling on the diode, it is cut off, so there is no voltage across R_1. When light strikes the diode, its reverse leakage current increases dramatically, causing a voltage drop to occur across R_1.

A phototransistor is a light-sensitive transistor. Its symbol is given in Figure 12.9C. It only has two terminals, the emitter and the collector. The device has a base, but typically it has no lead. Instead, the device is built so that when the base is exposed to light, it will cause the emitter-base junction to conduct. When light strikes the base, it causes ionization, which simulates base-current flow. Therefore, the transistor conducts between emitter and collector. The primary advantage of a phototransistor is its very high sensitivity over a photodiode. It will conduct even with a small amount of light applied to the base. The transistor essentially provides amplification so that small light levels can control a large current.

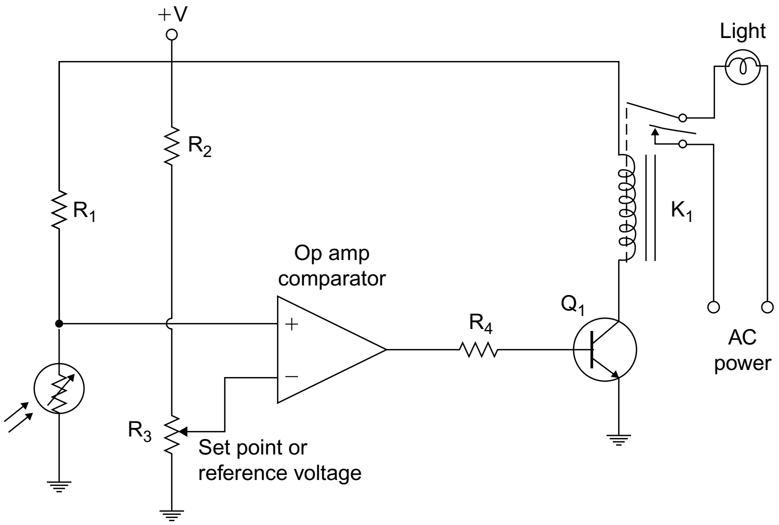

FIGURE 12.11 Simple off/on photo control circuit. Light turns on automatically at dark and off when light is present.

A typical closed-loop control circuit using a photoelectric cell is illustrated in Figure 12.11. This circuit will automatically turn a light on when it gets dark and turn the light off when it is light again. Note that a photoresistive cell is connected as part of a voltage divider with resistor R_1. The DC voltage from this voltage divider is applied to one input of an op amp comparator. The DC reference input to the comparator comes from a voltage divider made up of R_2 and R_3. The potentiometer can be used to adjust the DC reference voltage or set point. The output of the comparator drives a switching transistor Q_1, which will turn the relay K_1 off or on. This, in turn, operates the light.

When the photoresistive cell is exposed to light, its resistance will be very low. Therefore, the voltage from the voltage divider will be low. It will be less than the DC set point voltage and the comparator will not be triggered. The output of the comparator will be a negative voltage, which will reverse bias the emitter-base junction of the transistor and keep it turned off. As soon as it gets dark outside, the resistance of the photocell will increase. This will cause the voltage-divider output to increase. At some point, the voltage will be higher than the set point voltage, causing the comparator to switch. This applies a positive voltage to the transistor base, turning it on, and then turning on the relay and the light. The light will remain on as long as the photoconductive cell is in the dark. But as soon as light shines on the photocell, the voltage-divider output voltage will drop, thereby triggering the comparator and turning off the transistor, the relay, and the light bulb. By adjusting the reference or set point voltage with potentiometer R_3, the triggering point can be varied, so that the light will be turned off or on at the desired level of ambient lighting.

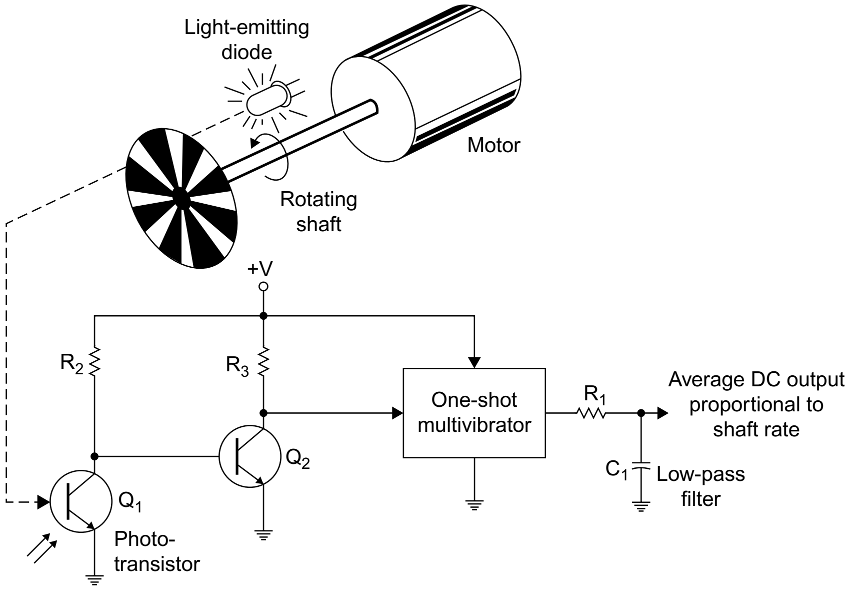

FIGURE 12.12 Simple tachometer circuit using phototransistor to create pulses that are averaged into a DC voltage proportional to speed.

This same basic concept can be used to produce a variety of industrial control operations. For example, you could use the circuit as an entry or intrusion detector. Shine a light on the photocell. If someone or something comes between the light and photocell, the circuit will trigger and the relay could then be used to operate a bell or buzzer or light.

Another example of a photoelectric controller is a tachometer circuit used for measuring the speed of a rotating shaft. Figure 12.12 shows a small, plastic wheel connected to the shaft of a motor. There are alternate clear and opaque areas on the disk. The disk is positioned so that it is between the light-emitting diode (LED) and the phototransistor. As the disk rotates, the dark opaque areas will block the light, while the clear areas pass the light from the LED to the phototransistor. When a clear area on the disk passes, the LED light will shine through it, hitting the phototransistor, and turning it on. When a dark area passes, the light will be blocked and the phototransistor will turn off. The rapidly rotating disk will cause the transistors to turn off and on very rapidly. This rapid conduction and nonconduction of the phototransistor causes a series of pulses to occur at the output of Q_2. These pulses are used to trigger the one-shot multivibrator.

The one-shot multivibrator produces a series of fixed, amplitude-width pulses. These are applied to an RC low-pass filter made up of C_1 and R_1. This filter will average the pulses into a DC voltage. This output voltage is proportional to the motor speed. If the speed of the motor should increase, the pulses occur at a more rapid rate. This means that the pulses occur closer together, and thus the average voltage across C_1 goes up. A DC voltmeter connected to the output can be used to indicate the speed. At low speeds, the meter indication will be low, while at high speeds it will be high. The meter dial itself can be calibrated in terms of revolutions per minute (rpm), thereby giving an accurate speed indication.

Flow Sensors

These devices have liquid or gas pass through them and measure the amount of flow in gallons per minute or other measure. They develop an output voltage proportional to flow rate.

There too many other sensors to catalog here but in most cases there is a sensor to convert any physical variable into a voltage to be measured.

OUTPUT DEVICES

An output device or an *actuator* is the thing that is most often controlled in a control system. The two most common types of actuators are relays and motors. There are many types of both.

Relays

A relay is an electromagnetically operated switch. A typical relay, and its schematic symbol, is illustrated in Figure 12.13. Its main components are a magnetic coil and a set of switching contacts. With no current applied to the coil, the spring keeps the armature pulled down, so that the contact arm makes connection with the upper, normally closed (NC) contact. When current is applied to the coil, a magnetic field is produced. This magnetic field attracts the steel armature, opening the upper contact, and closing the lower, normally open (NO) contact. The contacts form a single-pole, double-throw (SPDT) switch. When power is disconnected from the coil, the magnetic field ceases, and the contacts return to their initial state because of the spring.

Relay coils are designed to be operated from either AC or DC. The contacts are small circular disks of silver or tungsten. These materials are chosen because they will handle a large amount of current with minimum burning, arcing, and pitting.

The various kinds of switching contacts used in relays are illustrated in Figure 12.14. Normally open (NO) contacts are referred to as *form A*. Applying power to the coil closes the contacts. Normally closed (NC) contacts are referred to as *form B*. When power is applied to the relay coil, the contacts open. Single-pole, double-throw switching contacts are referred to as *form C*. When power is applied to the relay coil, the contact arm moves from contact A to contact B. Form C contacts are usually of the break-before-make variety. That is, contact A opens before the arm touches or “makes” contact B.

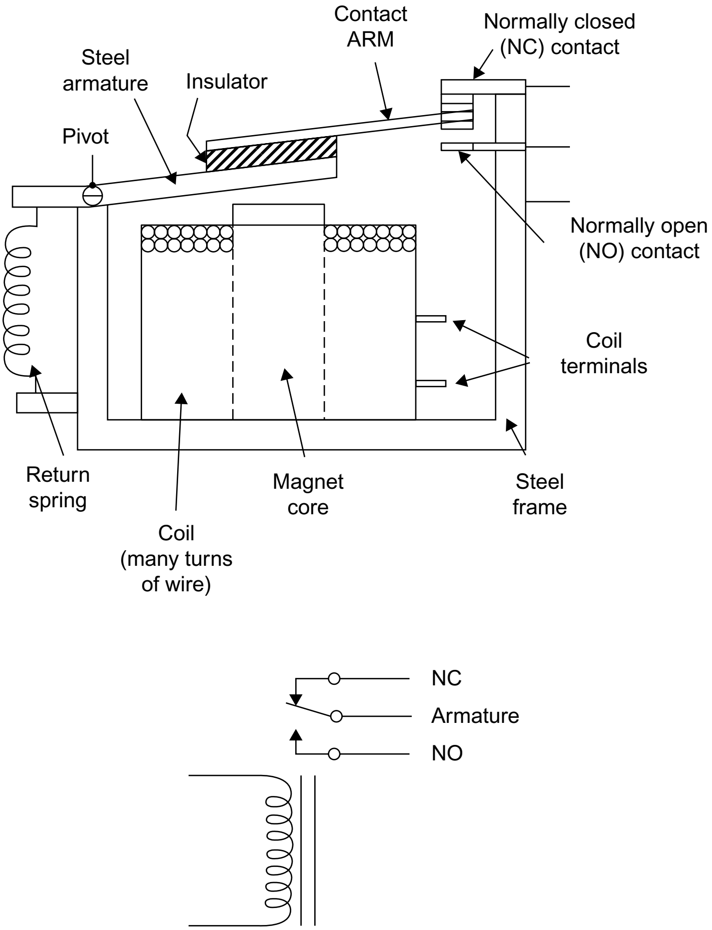
Contact
ARM
Normally closed
(NC) contact
Steel
armature
Insulator
Pivot
Normally open
(NO) contact
Coil
terminals
Return
spring
Magnet
core
Steel
frame
Coil
(many turns
of wire)
NC
Armature
NO

FIGURE 12.13 Cross-sectional view of typical relay showing major components and schematic diagram.

Sometimes you will hear the expression that a relay has been "picked" or "picked up." This means that power has been applied to the relay coil, and its contacts have been moved according to their function, and switching of a circuit has occurred.

Most relays usually have more than one set of contacts, so that multiple switching operations can be performed simultaneously. The various contact arms are ganged together so that they operate together when power is applied to the coil. A common configuration is two form-A contacts ganged to form a double-pole, single-throw (DPST) relay. Putting two form-C contacts together on the same relay produces a double-pole, double-throw (DPDT) relay.

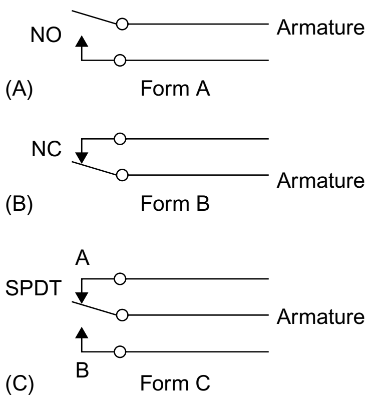

FIGURE 12.14 Common relay contacts. (**A**) Form A. (**B**) Form B. (**C**) Form C.

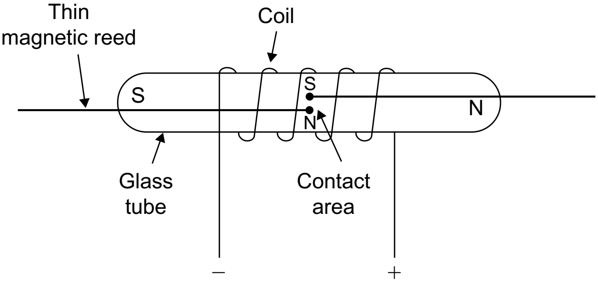

FIGURE 12.15 Reed relay.

Another widely used type of relay is the *reed relay*, illustrated in Figure 12.15. The heart of the reed relay is a reed switch that consists of two thin, metallic, reed contacts sealed in a glass tube. The reed contacts are made of a magnetic material, so that they will be magnetized when a magnetic field is applied to them. A coil consisting of many turns of fine wire is wound on a form and placed over the reed switch. When power is applied to the coil, a magnetic field is produced. This magnetic field magnetizes the two reeds as if they were bar magnets. One end of each reed will have a north pole and a south pole. The reeds are positioned so that the north pole of one will attract the south pole of the other. Remember, magnetic theory says that unlike poles attract, while like poles repel. The result is that the contacts will move toward one another and touch, making a good electrical connection. When power is removed from the coil, the contacts become demagnetized and spring apart. The reed relay will also operate by moving a permanent magnet near it. When the magnet is near, the contacts will close. This is a common way to sense open windows or doors in a home security system.

The contacts on reed relays cannot handle as much power as the contacts on conventional relays because they are smaller. As a result, they are used only in lower current power applications. However, the contacts on a reed relay open and close faster than those of a larger standard relay. Reed relay contacts can open and close in a few hundred microseconds, while standard relays require many milliseconds to open or close because of the larger contacts and the related structures that must be moved.

Relays are used in two basic ways in industrial control:

1. To control a larger current with a smaller current.
2. For remote control.

Let's examine the use of a relay for controlling a larger current with a smaller current. A good example of this is the starting system in your car. A large, heavy-duty DC motor is used to spin the flywheel on the car's engine to start it. This motor draws a current of hundreds of amperes. To turn the engine and start the car, the battery must be connected to the starter motor. This could be done directly by having the ignition switch connect the battery to the motor. However, this switch would have to have very large contacts to handle the high current. Such a switch would be large and expensive to build. The large heavy wires from the battery to the motor would have to be lengthened and brought up into the car.

A better solution is to use a smaller ignition switch to control the current applied to a relay coil (see Figure 12.16). The relay coil draws considerably less current than the motor; therefore, the switch contacts need not be large

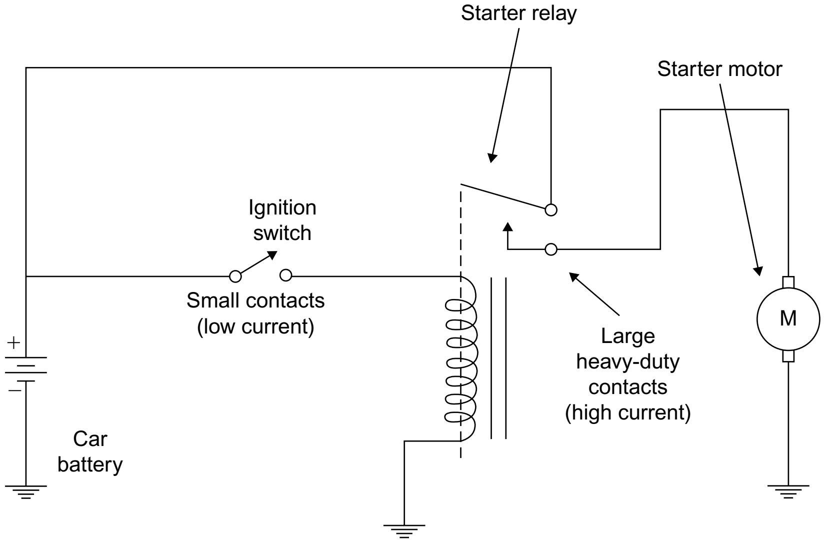
Starter relay
Starter motor
Ignition switch
Small contacts (low current)
+
−
Car battery
Large heavy-duty contacts (high current)
M

FIGURE 12.16 Controlling a large current with a smaller current with a relay.

or expensive. The switching contacts on the relay are much larger and will carry the high motor current. When the ignition switch is turned, its contacts close, applying current to the relay coil. The form-A contacts close, connecting the battery to the starting motor. When the car starts, the ignition switch is released, the relay drops out, and the starter motor shuts off.

Another application of a relay is remote control. Again, the starting system of a car is a good example. The load to be controlled, in this case the starter motor, is located remotely from the ignition switch. If the ignition switch were used to apply current to the motor, long wires would be required to connect the battery, the motor, and the switch. Heavy wire conductors would have to be used to carry the high current. The longer the interconnecting wires, the higher the resistance of the circuit. While large conductors have minimum resistance, when high current flows through them, voltage drops still occur. The voltage dropped across the interconnecting lines subtract from the battery voltage available for the motor. The voltage drops might be excessive, unless very large, expensive, and inconvenient conductors are used.

To eliminate this problem, a remote relay is used. The relay itself is located near the starter motor and the battery. In this way, the interconnecting leads can be kept short, thereby minimizing voltage drops. Since the relay operates from very low current, longer interconnecting wires can be used to attach it to the ignition switch without noticeable voltage drops.

In some industrial control applications, the load to be operated is located at extreme distances from the switch used to control it. The current controlling the device does not have to be passed through very long conductors if a relay is used. Small conductors running to the switch remotely from the load control the relay while the relay turns the load off and on.

Relay contacts stay closed (or open) as long as power is applied to the coil. This means that the switch operating the relay coil must be kept closed if the relay is to remain actuated. There are applications, however, where it is desirable to use a momentary contact switch to operate the relay. An example is a pushbutton that, when depressed, causes contacts to be closed and power to be applied to the relay. When the button is released, a spring causes the contacts to open. With such an arrangement, the relay will close momentarily while the button is depressed, but will drop out when the button is released. The objective is to cause the relay to latch into its actuated position when the coil is momentarily pulsed. Such a relay is called a *latching relay*. Some latching relays use a mechanical arrangement on the armature to cause the contacts to latch closed when the coil is momentarily pulsed.

A relay can be made to latch electrically if it has an extra set of contacts. A typical latching circuit is shown in Figure 12.17. This relay has two sets of form-C contacts that operate simultaneously. The lower set of contacts is used to operate the load. The upper set of contacts is used to automatically latch the relay once it is pulsed. Switch S_1 is a normally open (NO) momentary contact pushbutton that is used to apply power to the relay coil. Depressing the button

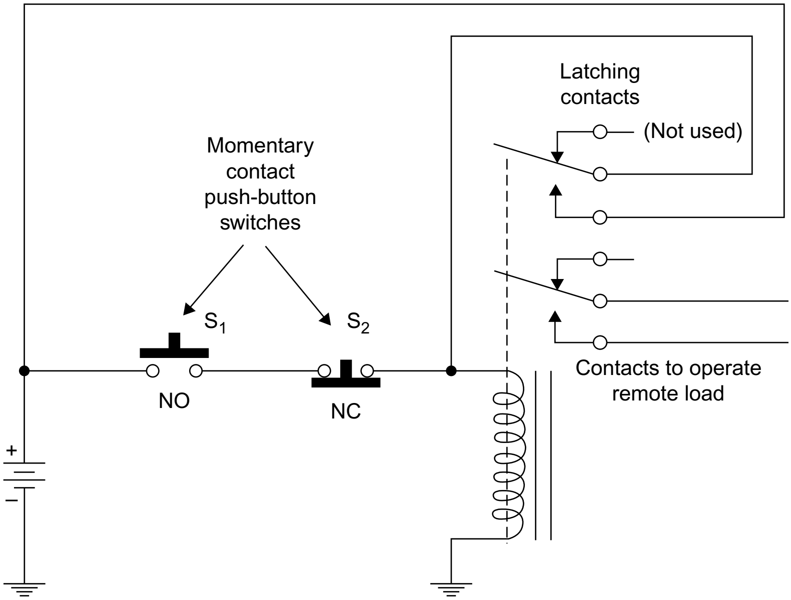

FIGURE 12.17 Latching a relay with its own contacts.

causes the relay coil to operate, closing the contacts. The lower set of contacts operates the load as desired. The upper set of contacts applies the power to the coil. Thus, when pushbutton S_1 is released, the circuit to the relay coil remains closed through the upper-latching contacts. To turn off the circuit, normally closed (NC) pushbutton S_2 must be depressed. This breaks the relay coil circuit, removing power. The relay drops out, opening the load contacts and the latching contacts.

While many relay coils are operated by mechanical switches, others are operated by a transistor. A typical transistor-controlled relay circuit is shown in Figure 12.18. The transistor is operated as a switch. When the base is grounded through contact B on S_1, the transistor is cut off and acts as an open switch. Therefore, no current flows in the relay coil. When a voltage is applied to the base resistor, through contact A on S_1, the transistor saturates, acting like a very-low-resistance or "on" switch. Current flows through the relay coil, operating the contacts. Removing the base current from the transistor turns the transistor off, terminating current in the relay. With this arrangement, an even smaller current can be used to control the already small relay current. A base current, of only microamperes, is sufficient to cause the transistor to operate.

In the circuit of Figure 12.18, when base current is removed from the transistor, the transistor will cut off. When current through the relay coil ceases, the magnetic field around it collapses. In doing so, the magnetic field induces a very high voltage in the relay coil. This very high voltage will be applied to the collector of the transistor. This voltage could be as high as several hundred, or

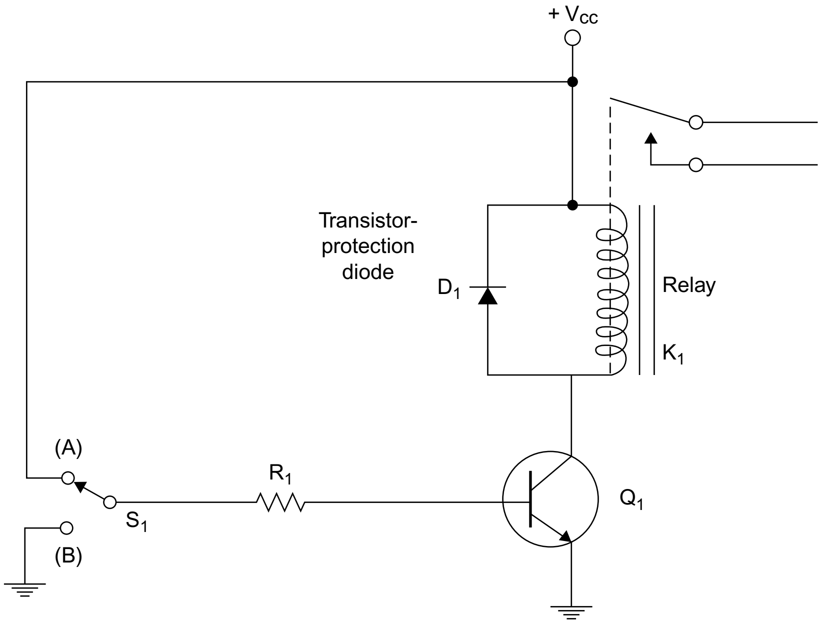

FIGURE 12.18 Operating a relay with a transistor is a common application. The diode protects the transistor when power is removed from the relay coil.

even thousands, of volts, depending on the relay coil. Such a high voltage will damage the transistor. To protect the transistor against such occurrences, some kind of voltage-spike eliminator or surge suppressor must be used across the relay coil. A commonly used protection device is a silicon diode D_1, as shown in Figure 12.18. When current is applied to the relay coil through the transistor, the diode is reverse biased and has absolutely no effect on the circuit. When the transistor turns off, the high voltage will be inducted into the coil with the polarity reversed from the supply voltage V_{cc}. This causes the diode to become forward biased and conduct. Thus, the diode effectively shorts out the high voltage spike, protecting the transistor.

Solenoids

A solenoid is a device that produces linear motion with an electromagnetic coil (see Figure 12.19). A magnetized steel rod called a plunger is inserted into a coil of wire. When a voltage is applied to the coil, a magnetic field is produced. It interacts with the magnetic rod producing attraction and/or repulsion that causes the rod to move into or out of the coil. The rod is usually physically constrained in one or both directions, and is often spring loaded so that it always returns to the same place when power is removed.

Solenoids produce linear motion. They are usually connected to a lever or ratchet mechanism that produces other forms of motion including rotation.

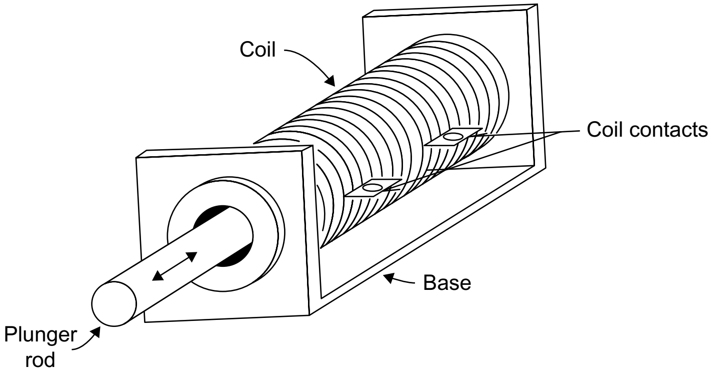

FIGURE 12.19 A solenoid produces linear motion using magnetic principles.

Solenoids are also used in valves to turn off the flow of a liquid or gas. By continuously controlling the current in the coil, the rod may also be precisely set to some position other than off or on. This effect is used in valves to continuously control the opening.

Motors

A motor is a device that converts electrical energy to mechanical energy. The mechanical output is typically the rotation of a shaft. That rotation can then be used directly or converted into other mechanical forms with levers, gears, ratchets, or cams. There are dozens of different motor types but most are categorized as either DC or AC.

Motors: Billions and Billions Sold and Used

Motors are one of those devices we take for granted. We use them every day but don't realize it. Most motors are hidden but do an enormous amount of physical manipulation. In industry, motors drive machine tools, conveyer belts, pumps, and other mechanical systems. Elevators use big motors.

At home motors are everywhere. Just think a minute and try to identify all the motors you have in your home and in your car. All of your appliances have motors, such as the washer, dryer, dishwasher, and HVAC units. There are motors in your mixers, blenders, and can openers. Shavers and hair dryers have motors. An analog clock is a motor. Then there are the motors in your electronics. In a PC or laptop, there are the fan motor and the ones in your hard disk and CD drives. Your printer, scanner, and copier and fax machine have motors. Your DVD player, CD player, and VCR all have multiple motors.

Then in your car is the starter motor, as well as the motors for windows, seats, mirrors, and windshield wipers.

Can you name some others?

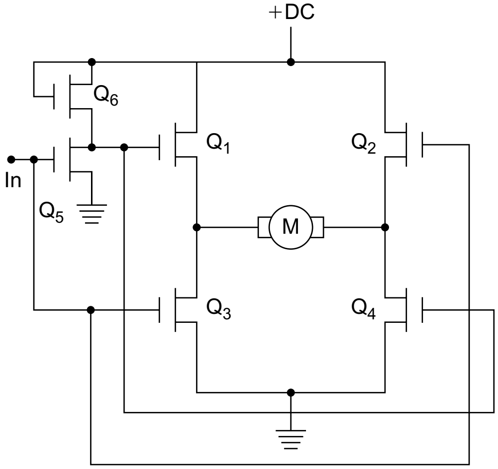

FIGURE 12.20 Using an H-bridge circuit to reverse the current flow and direction of rotation in a DC motor.

DC Motors

A DC motor operates from a DC voltage source like a battery or power supply. It is made up of a magnet and a rotating coil. When a current is applied to the coil, it forms an electromagnet. The magnetism of the coil interacts with the magnet field, producing rotary motion. The magnet is commonly a permanent magnet in smaller DC motors but an electromagnet in larger motors. In the larger motors, a special field coil is energized with DC voltage as well to create the magnetic field.

The main specifications of a motor are its horsepower rating, speed, and torque. Speed and torque are controlled by the amount of current flowing in the coil and/or the strength of the magnetic field. Increasing the voltage applied to the motor usually causes speed and torque to increase. Some kind of electronic control circuit is usually involved to set the speed to the desired value. A variable resistor in series with the motor will control its speed, but usually more complex electronic circuits are involved.

Incidentally, the direction of rotation is determined by the polarity of the DC voltage. Reversing the polarity reverses the magnetic field, thus changing the direction of rotation. Figure 12.20 shows a neat circuit used to change the direction of rotation of a DC motor. This is called an H-bridge and is made up of MOSFET switches. Q_2 and Q_3 are turned on by a positive input voltage IN. Q_5 inverts this voltage to zero, which keeps Q_1 and Q_4 off. Therefore, current (electrons) flows from ground through Q_3 through the motor (M) from left to right through Q_2. Setting the input to zero turns off Q_2 and Q_3 and turns on Q_1 and Q_4. Now current flows from ground through Q_4 and the motor from right to left and then through Q_1. This changes the direction of the motor with just a simple off/off input control voltage.

Most DC motors rotate too fast for many applications. A typical speed range is 1000 to 10,000 rpm. To achieve lower speeds, the motor is used in conjunction with a gear box or pulleys.

A special type of DC motor is the *stepper motor*. This is a DC motor driven by two DC pulse signals. The pulses operate coils or windings that produce a rotating magnetic field. A permanent magnet rotor interacts with the rotating magnetic field to produce motion. The interesting characteristic is that the rotation occurs in increments or steps instead of a continuous rotation as with a standard DC motor. Depending on the number of windings and coils and elements on the rotor, the increments usually vary in increments of 22.5, 18, 15, 7.5, or 3 degrees or smaller. The stepper motor therefore permits very precise positioning. Yet it can be rotated at several thousand rpm if needed.

AC Motors

AC motors usually operate from a 50- or 60-Hz power line. Again the rotation is produced by interacting magnetic fields. The speed of rotation is fixed by the number of coils and magnet poles inside the motor. That speed is also often higher than needed, so gears and/or pulleys are used to drop the speed to the desired value.

To change the speed of an AC motor, you must vary the frequency of the AC voltage applied to it. This means that you cannot operate the motor directly from the AC power line if the speed must be changed. Instead, you must use a so-called AC drive. This is a box of electronics that generates a variable frequency AC voltage to apply to the motor. An AC drive takes the AC line input and uses semiconductor switches to produce an AC signal at a different frequency.

Other Actuators

Besides motors, solenoids, and relays, there are a variety of different controlled devices. Valves are common device. Valves are used to turn the flow of a liquid or gas on or off. Proportional valves very carefully regulate the amount of flow.

Other devices are hydraulic or pneumatic devices. Hydraulic devices use oil in cylinders to provide mechanical force where needed or position something mechanically. Pneumatic cylinders work with compressed air to apply force or control the position of a device.

THYRISTORS

Many industrial switching circuits are implemented with MOSFETs designed for high voltage and high power. You will sometimes hear these devices called *insulated gate field-effect transistors* (IGFETs).

Other power-switching devices are also widely used, especially thyristors. A thyristor is a semiconductor device used for switching purposes in industrial control. Like the relay, it is used to apply and control electrical power to motors, heating elements, lights, and other loads in industrial applications. In many applications, thyristors replace relays. A thyristor acts as a switch, but because of its solid-state nature, is far faster in switching than a relay. There are three basic types of thyristors: silicon-controlled rectifiers (SCRs), triacs, and diacs.

Silicon-Controlled Rectifier

The SCR is a three-terminal thyristor that acts like a silicon rectifier diode whose conductor is controlled by an input current. The schematic symbol for an SCR is shown in Figure 12.21. The symbol is similar to that of a diode with a cathode and an anode. Note that the third element of the SCR is known as the gate. The SCR will conduct current between cathode and anode, but only if the proper control current is applied to the gate. The gate must be made positive with respect to the cathode in order for the SCR to conduct. When conducting, the SCR acts like a closed switch. The voltage drop across the cathode and anode will be approximately 0.7 to 1.8 volts, depending on the size of the SCR and how much current is flowing through it. When the cathode and anode are reverse biased, current will not flow through the device.

Thyristors, like relays, are generally used to control a larger current with a smaller current. Figure 12.22 shows how an SCR is used as a switch to apply DC power to a light bulb. The load could also be a motor or heating element. The DC voltage is connected so that the cathode and anode of the SCR are forward biased, but no current will flow through the device until a current is applied to the gate. This is done with the switch S_1. With the switch open, no current flows in the gate. Closing the switch, however, applies a positive voltage to the gate through resistor R_1. This causes the SCR to turn on. When it conducts, it acts as a low-resistance "on" switch and the light bulb goes on. At this point, switch S_1 may be opened. It is not necessary to maintain current in the gate in order for the device to continue to conduct. The gate current only

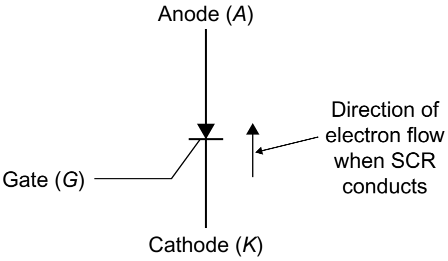

FIGURE 12.21 Schematic symbol of an SCR.

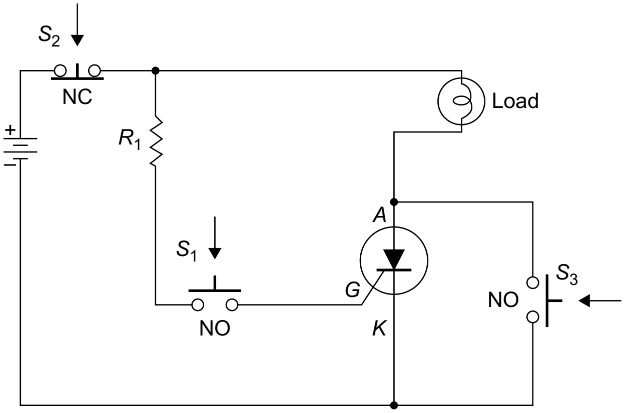

FIGURE 12.22 Circuit that shows how an SCR is turned on and off.

needs to be momentary, for it is only required to turn the device on. The device remains on like a latching relay that has been pulsed.

The gate element is only used to turn the SCR on. Removing gate current will not turn the device off. To cause the SCR to stop conducting, the circuit must be broken. This can be done with a momentary-contact, normally closed pushbutton, in series with the circuit, like S_2 in Figure 12.22. Depressing this pushbutton will break the circuit, stopping the current in the SCR, and the light bulb will go off. To turn the bulb back on, S_1 must again be depressed to apply gate current momentarily.

Another way to stop conduction is to momentarily short around the SCR, as shown in Figure 12.15. When switch S_3 is closed momentarily, current will flow through it and the bulb, bypassing the SCR. The current in the SCR will drop to zero. When switch S_3 is opened, the circuit will be off.

While SCRs are sometimes used to control DC power, in most applications they are used to control AC. Figure 12.23A shows the SCR used to apply AC to a light bulb. If switch S_1 is closed, gate current will be applied to the device. The device, however, will only conduct when the anode is made positive with respect to the cathode. This, of course, occurs when the applied AC voltage has the correct polarity. Since the SCR operates as a rectifier diode, current will only flow through the device on the positive half-cycles of the sine wave. The current through the light bulb will be a pulsating DC, as illustrated in Figure 12.23B. The brightness of the lamp will depend on the average amount of current flowing.

The average amount of current through the light bulb, or other load, can be controlled by using the gate element. Various electronic circuits are used to adjust the time when the SCR turns on. By turning on the SCR at the appropriate time in the positive half-cycle, the duration of the current flow can be controlled. The longer the current is allowed to flow, the higher the average

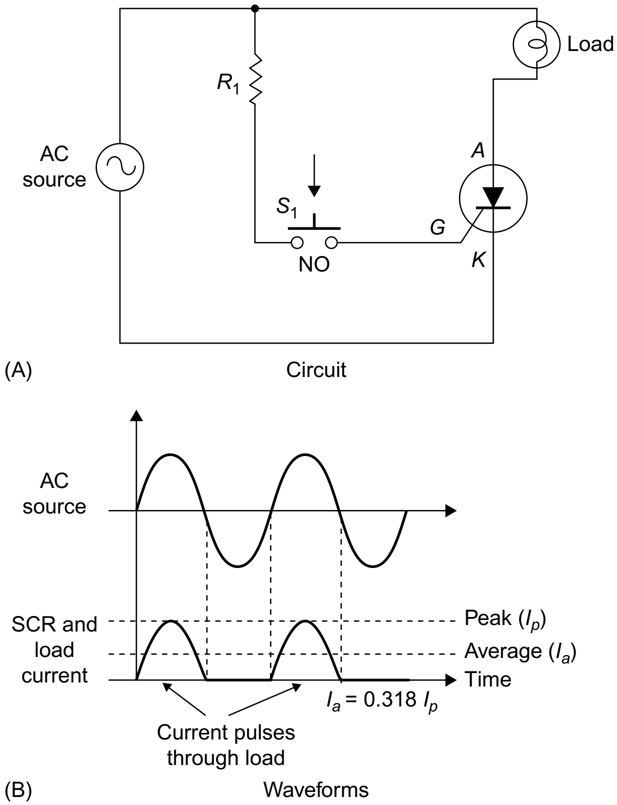
Load
R1
AC source
A
S1
G
K
NO
(A)
Circuit
AC source
SCR and load current
Peak (I_p)
Average (I_a)
Time
$I_a = 0.318\ I_p$
Current pulses through load
(B)
Waveforms

FIGURE 12.23 Using an SCR to control power to a load using AC. The SCR rectifies the AC.

current in the load. By changing the control point, the brightness of the bulb can be varied.

Diac

A diac is a two-terminal semiconductor device that conducts in either direction when a voltage of a specific level is exceeded. This voltage is called the *triggering voltage*. Typical diac triggering voltages are in the 20- to 45-volt range. The diac is used as a triggering device for SCR and triacs as it determines when the device will conduct or cut off.

Triac

A triac is another three-terminal thyristor. Its operation is similar to that of an SCR, but the triac will conduct in either direction. The schematic symbol for the triac is shown in Figure 12.24. Current will flow from main terminal 1 (MT1)

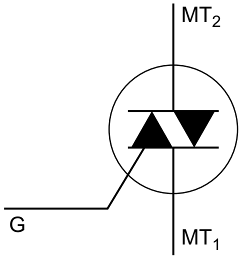

FIGURE 12.24 Schematic symbol for triac.

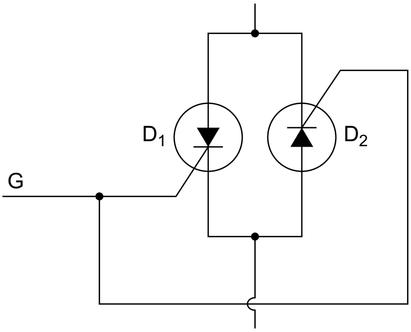

FIGURE 12.25 Equivalent circuit of triac shown with SCRs.

to main terminal 2 (MT2), or from main terminal 2 (MT2) to main terminal 1 (MT1), depending on the polarity of the voltage applied, and if the gate triggers the device on.

An approximate, equivalent circuit of the triac is shown in Figure 12.25. Here, two SCRs are connected back to back with their gates connected together. Current can flow in either direction, depending on whether D_1 or D_2 conducts. Of course, SCR conduction depends on whether a gate-triggering current is applied. Triacs, like SCRs, are used primarily to switch AC voltage to a load. If gate current is applied continuously, the triac will conduct in both forward and backward directions, causing both positive and negative half-cycles of the AC to be applied to a load.

Figure 12.26 shows a circuit used for light-dimming purposes. The load is a light bulb. An RC phase shifter and diac-triggering device are used on the gate of the triac. Assume that the diac has a triggering voltage of 30 volts. If the voltage across the diac is less than 30 volts, it will not conduct. As soon as the 30-volt point is reached or exceeded, the diac rapidly conducts and fires the SCR. Keep in mind that the diac is a bidirectional device, and that current will flow through it in either direction if the bias voltage in either direction exceeds the triggering voltage level.

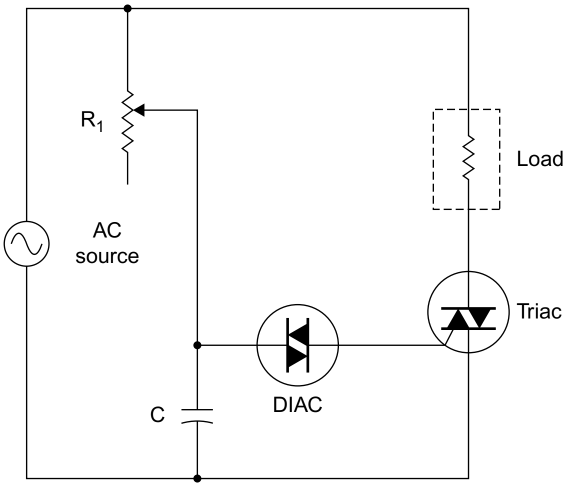

FIGURE 12.26 Lamp dimmer circuit using RC phase shifter and diac to control current in the load.

The triac will conduct on both positive and negative half-cycles of the applied AC. The phase shifter made up of R_1 and C will set the triggering delay. It will determine at what point in the AC cycle the diac will fire and apply gate current to the triac. Figure 12.27 shows the waveforms at two different phase-shift points in the circuit. Note that current flows through the load in both the positive and negative half-cycles, but the duration of current flow depends on the point where the gate triggers. In Figure 12.27, the average current through the load will be relatively high at a 60-degree phase shift, and the bulb will be bright. At 120 degrees, the triggering point is late in each half-cycle, allowing only a small amount of current to flow through the bulb. The bulb, therefore, will glow dimly. Adjusting R_1 will allow the brightness of the bulb to be varied from nearly full on to full off. This is the circuit normally used in household and industrial light dimmers.

This same circuit can be used to control the speed of AC motors. In connecting a hand drill as the load, for example, R_1 will vary the speed of rotation.

Triacs are also widely used as solid-state relays.

PROGRAMMABLE LOGIC CONTROLLERS

A very popular type of controller used in industry and process control is a specialized computer called a programmable logic controller (PLC), or programmable controller. It uses sensors to monitor physical variables and generates output control signals to operate heating elements, motors, pumps, solenoids, and other devices. The PLC, however, contains a microprocessor that can be

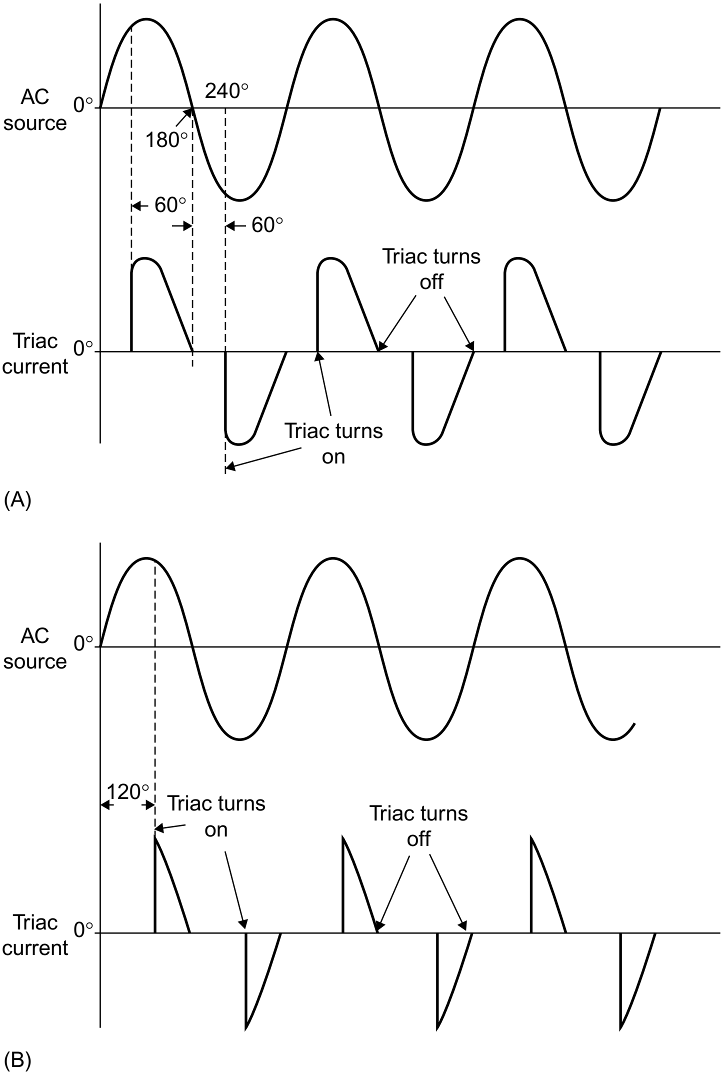
AC source
0°
240°
180°
60°
60°
Triac turns off
Triac current
0°
Triac turns on
(A)
AC source
0°
120°
Triac turns on
Triac turns off
Triac current
0°
(B)

FIGURE 12.27 AC waveforms in a lamp load controlled by a triac. **(A)** Waveforms with triggering at 60 degrees. **(B)** Waveforms with triggering at 120 degrees. The longer pulses produce higher average current in the load.

programmed in a special language that permits the PLC to be customized to the particular control function. In this way, a standard PLC can be programmed and configured for virtually any application. The PLC is by far one of the most popular and versatile control devices in use in industry.

The main function of a PLC is factory automation. It is a versatile special-purpose computer that can be quickly and easily reconfigured to meet changes

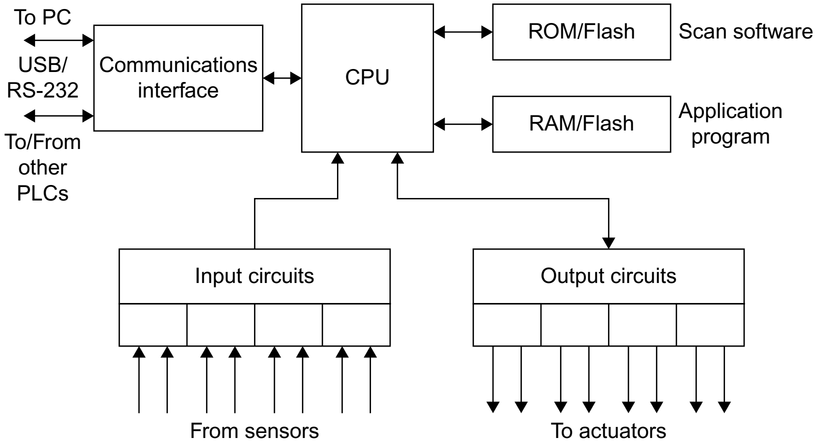

FIGURE 12.28 General block diagram of PLC.

and additions in factories, plants, and other operations that routinely need to control and sequence various machines and processes. While an embedded controller or personal computer could be used in some operations, they are harder to program and apply and are less flexible when rapid changes are needed. Furthermore, these common microcomputers are not designed for the factory environment. But PLCs are rugged and hardened against the harsh environment of most factories and plants. They can withstand the temperature ranges, vibration, and dirty conditions usually existing in a manufacturing or process control setting.

Organization

A general block diagram of a PLC is shown in Figure 12.28. It is like most other microcomputer diagrams but there are some differences. A microprocessor is at the heart of the design. This CPU is coupled to a ROM where a special control program is stored. This program implements the basic scan operational mode of the PLC. RAM is also available to store the application program. Flash memory may also be used in place of RAM.

The key to the PLC's versatility is its extensive I/O section. It has a variety of input and output modules. These are fixed in small, single-function PLCs, but in the larger more flexible units, the PLC is designed to accept a wide variety of separate I/O modules. These usually plug into a rack containing the main PLC unit and power supply. The different I/O modules let the user customize the PLC to the specific application.

Some examples of I/O modules are digital inputs and outputs with common logic levels or conditioned levels, analog-input modules that accept analog signals from sensors, and analog output modules that generate analog outputs for controlling actuators of almost any kind. A huge variety is available.

A communications interface is also available so that the PLC can talk to external devices like a PC or other PLCs that are connected to form a larger control system. RS-232 and RS-485 serial interfaces are common as is USB. The program describing the operations to be implemented is developed on an external PC, and then downloaded to the PLC RAM or flash memory via the USB or RS-232 interfaces.

Operation

The general operation of a PLC is shown in Figure 12.29. When the PLC is started, the internal operational program looks at the application program and interprets what is to be done. The sequence of operations is to first scan the inputs. The PLC looks at what the various sensors are doing. It gets a status update from the application itself. The program then interprets what to do next based on the application program. It processes the input to determine what the outputs should be. Then it updates the outputs to carry out the application. Then the cycle repeats itself at a high rate of speed. The inputs are scanned and outputs updated every 1 millisecond or so, slower or faster as the application requires.

A simple example is using a switch to turn on a motor. The switch contacts are connected to an input module by wires. Output wires connect to a relay that in turn is used to operate the motor. As the PLC is scanning, it sees that the switch is off, so it does nothing. But as soon as the switch is turned

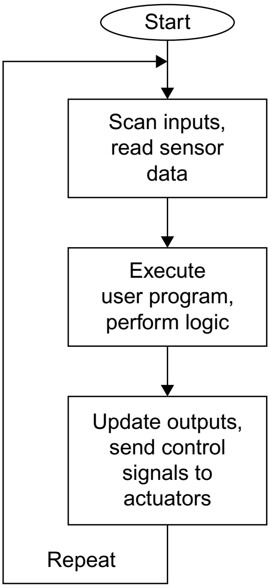
Start
Scan inputs, read sensor data
Execute user program, perform logic
Update outputs, send control signals to actuators
Repeat

FIGURE 12.29 PLC operational scan cycle.

on, the next scan detects this new condition. The program then tells the CPU that if the switch is on, turn on the relay. An output signal is sent to the output module, which then actuates an external relay. The relay contacts then turn the motor on. If the switch is then turned off, the next scan detects this condition and causes the output to turn off the relay and motor.

Programming

As for programming, PLCs can be programmed in various ways. Most PLC vendors have special languages to do the programming. In some cases, standard BASIC or C languages may be used. One widely used approach is a graphical language called *ladder logic*. Ladder logic is the technology that the PLC replaced decades ago. Control systems were made with relays, and they were wired so the diagram looked like a ladder (see Figure 12.30). The two vertical lines represent a DC or AC voltage source. Circuits are then connected between these two "rails," forming rungs on the ladder. The circuits are connections of switches, relays, and other components. The software allows the user to build the ladder logic program on the screen, while a translation program in the PC converts it into the binary code that the PLC can execute.

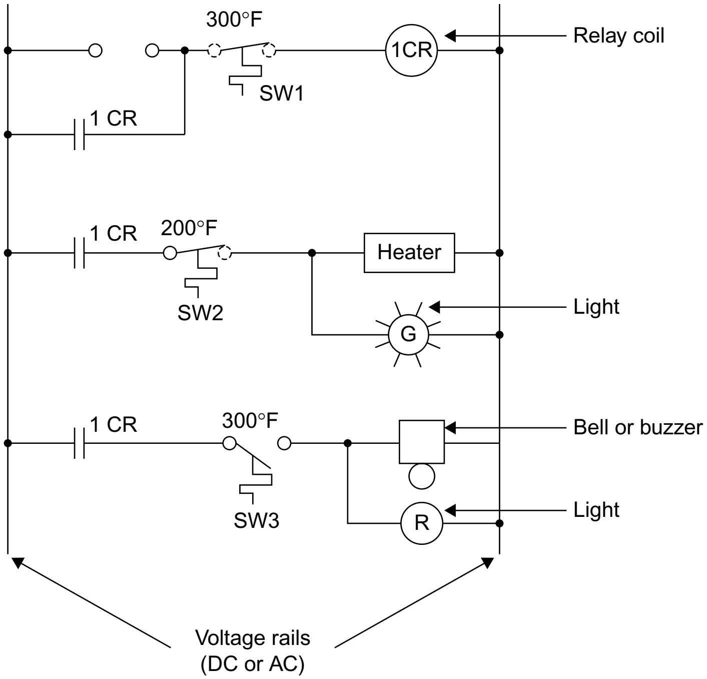

FIGURE 12.30 Sample ladder diagram and PLC application program.

This particular circuit in Figure 12.30 does the following: when the ON pushbutton is pressed, current is applied to relay coil 1CR through normally closed contacts of SW1. SW1 is a thermal switch whose contacts stay closed as long as the temperature is less than 300 degrees Fahrenheit (300°F). If the temperature rises above that level, the contacts open, thereby breaking the circuits. Relay contacts 1CR around the ON pushbutton latch the circuit on when the ON pushbutton is released.

In the next "rung" of the ladder contacts 1CR are on the relay. They close when the relay turns on. This applies voltage to a heating element. This current flows through the normally closed contacts of thermal switch SW2. A green light in parallel with the heating elements turns on to indicate the heater is on.

When the heater temperature reaches 200°F, SW2 opens, turning off the heater and the green light. As the heater cools, SW2 will then close again, applying voltage to the heater. The heater will then cycle off and on keeping its temperature near the 200°F value.

If the temperature should rise above 300°F due to a failure of SW2 °F then SW1 opens. Relay 1CR is turned off, and its contacts open, releasing the latch on rung 1 and opening the circuit on rung 2. Thermal switch SW3 closes at 300°F, turning on a red light and ringing an alarm bell.

This example could be programmed with a graphical ladder logic program or with special commands and syntax of a traditional programming language.

In addition to being able do simple off/off tasks common to industrial control, special hardware in the PLC can also implement counters and timers so that external events can be counted and the sequencing of events can be programmed.

Project 12.1

Build a Robot

Robots are great examples of basic industrial control. While most industrial robots are usually manipulator arms, you can actually buy a mobile robot with wheels that can be programmed for motion and other functions. A wide range of kits are available. It is a great way to get familiar with sensors, a microcomputer controller, and various actuators. Typical inputs are switches, photo cells, temperature sensors, and common outputs are relays, lights, and motors.

Some recommended kits are those made by Lego and Parallax. Suggested websites to learn more follow:

Lego Mindstorms NXT: www.mindstorms.lego.com

Parallax: www.parallax.com

Two magazines regularly cover robots from a hobbyist/experimenter viewpoint. These are *Nuts & Volts* and *Servo*. You may be able to find these at a newsstand; otherwise, go to their websites:

Nuts & Volts Magazine: www.nutsvolts.com

Servo Magazine: www.servomagazine.com

BOOK REFERENCES

Here is a list of books for those who wish to dig deeper. You will find lots of basic electronics books out there, but these are my favorites as well as some of my own.

Ashby D. *Electrical engineering 101*. Newnes/Elsevier; 2008. A good basic book that delves more into the circuits than this book. A good next step from here.

Frenzel LE. *Crash course in electronics technology*. 2nd ed. Newnes/Elsevier; 1996. Out of print now but still available in the used market. A programmed introduction to electronics.

Frenzel LE. *Crash course in digital technology*. 2nd ed. Newnes/Elsevier; 1998. Out of print now but still available in the used market. A programmed introduction to digital circuits. Getting a bit dated now but still valid.

Frenzel LE. *Crash course in PC and microcontroller technology*. Newnes/Elsevier; 1999. Out of print now but still available in the used market. A programmed introduction to micros. Getting dated but the fundamentals are still good.

Frenzel LE. *Principles of electronic communications systems*. 3rd ed. McGraw-Hill; 2008. Recently updated and it takes the systems approach to the subject. A college text but very readable.

Horowitz P, Hill W. *The art of electronics*. 2nd ed. Cambridge University Press; 1989. A bit dated now but a good example of how to present electronics. A great reference. A lab manual is also available. I wish the authors would update this book, but knowing how much work is involved, I might be reluctant too.

Kuphaldt, TR. *Lessons in electric circuits*. Self-published; 2000–2009. A complete electronics book online and free for download. It covers DC and AC circuits, semiconductors and basic circuits, and digital. A lab manual is available. Go to www.openbookproject.new/electricCircuits.

Maxfield C. *Bebop to the Boolean Boogie*. 3rd ed. Newnes/Elsevier; 2009. First published in 1995, it was a hit. A great introduction to digital practice and circuits. It was recently updated.

Michael NA. *Introduction to telecommunication electronics*. 2nd ed. Artech House; 1995. A great little book. Interesting format with each topic presented in two pages. Lots of basic electronic info as well as communications.

© 2010 Elsevier Inc. All rights reserved.
doi: 10.1016/B978-1-85617-700-9.00017-5

Index

#

3D TV, 264-265

A

AC. *See* Alternating current
Accumulator
  in ALU, 130
  register, 129–130
Actuator, 289–298. *See also* Motor; Relay; Solenoid
ADC. *See* Analog-to-digital converter
Address buffer, 131, 132f
Address bus, 135
Address register, memory, 131, 132f
Addressing, in RAM, 113f
Aircraft, radio in, 176
Alkaline battery, 30t
Alternating current (AC), 26–29, 27f
  capacitors' passing of, 49f
  frequency of, 28
  generators for, 33
  inverters for, 33
  motor, 298
  oscillator in, 26, 33
  polarity of, 26
  signal shapes of, 26, 27f
  in solar power system, 87f
  sources for, 33
  voltage measurement for, 28–29
Alternator, 33
ALU. *See* Arithmetic logic unit
AM. *See* Amplitude modulation
American Standard Code for Information Interchange (ASCII), 93–98
  character code of, 95t–96t
  control codes of, 94–97
Ampere, 18–19
Amplifier, 7–8, 52, 62–63, 69–73
  audio frequency, 63
  BJT, 53f
  cascading, 71–72, 71f, 72f
  class A, 64, 64f
  class AB, 65
  class B, 65, 65f
  class C, 65, 66, 66f
  class D, 66–67, 67f
  class E, 68
  class F, 68
  classification of, 63–69, 64–68
  configuration of, 63–64
  DC, 63
  differential, 64, 64f, 68f
  efficiency of, 72
  follower, 68f
  frequency response of, 72–73, 73f
  gain of, 63, 72f
  impedance in, 70–72, 70f
  instrumentation, 69, 69f, 283–284
  intermediate frequency, 63
  inverting and noninverting, 68f
  as linear circuit, 63
  microwave, 63
  operational, 68–69, 68f
  output capability of, 73
  power, 63
  pre-, 230
  radio frequency, 63
  in radio receiver, 166
  signal size of, 63
  single-ended, 63, 64f
  switching, 66–67, 67f
  symbol for, 62f
  vacuum-tube, 244

© 2010 Elsevier Inc. All rights reserved.
doi: 10.1016/B978-1-85617-700-9.00002-3

Amplitude modulation (AM), 159, 160f.
    See also Quadrature amplitude modulation
Amplitude shift key (ASK), 162, 163f
Analog-to-digital converter (ADC)
  for data conversion, 34, 118
  for digital TV, 252f
  parallel output of, 120f
  serial output of, 120f
Antenna, 151, 170–175
  dipole, 170, 172f
  electromagnetic fields from, 152f
  ground-plane, 170, 171f
  horn, 172–173
  loop, 171–172
  microwave, 172–175, 174f
  patch, 171–172, 174f
  transmission lines to, 172, 174f
  Yagi, 170–171, 173f
Arithmetic circuit, 111
Arithmetic logic unit (ALU), 10, 10f
  accumulator in, 130
  in computer, 10, 10f, 126
  data processing by, 130f
  as logic circuit, 111
  operation of, 129–131
ASCII. *See* American Standard Code for Information Interchange
ASK. *See* Amplitude shift key
Aspect ratio, 248, 249f
Assembly (programming language), 142–143
Astronomy, radio, 179–181
Asynchronous transfer mode (ATM), 216, 217
ATM. *See* Asynchronous transfer mode
Atom, 16, 17f
Audio
  amplifier, 63
  digital, 231–239
  dynamic range of, 233
  early recording media for, 232
  electronics, 229
Audio-video receiver (AV receiver), 239–243
  diagram of, 241f
  radio in, 242
Automatic gain control, 166
AV receiver. *See* Audio-video receiver

B

Band-pass filter (BPF)
  bandwidth of, 75
  response curve of, 74f
  symbol for, 74f
Band-reject filter (BRF), 74f
Bandwidth
  of BPF, 75
  modulation and, 161f
  of radio, 160, 161f
BASIC (programming language), 143–144
Battery, 29f
  cell voltages in, 30t
  charging of, 31
  as DC source, 29–31
  diagram of, 30f
  multimeter measurement of, 56f
  as power supply, 82
  primary cells in, 31
  secondary cells in, 31
  types of, 30t
BCD. *See* Binary-coded decimal
Binary data word, 92
Binary digits, 90
Binary numbers, 90–100
  decimal conversion from and to, 90–92
  hardware representation of, 91
  maximum number of states in, 93t
Binary phase shift keying (BPSK), 163, 163f
  spectral efficiency of, 197t
Binary-coded decimal (BCD), 93–98

- code of, 94t
- hexadecimal notation in, 97–98

Bipolar junction transistor (BJT), 50, 52f, 53f

Bits. *See also* Microcontroller; Microprocessor
- binary numbers as, 90–100
- maximum decimal value calculation for, 92–93
- significant, most and least, 90

BJT. *See* Bipolar junction transistor

Bluetooth, 222–223

Blu-ray, 275–276

Boat, radio on, 177–178

Boolean algebraic expressions, 100

BPF. *See* Band-pass filter

BPSK. *See* Binary phase shift keying

Breadboard
- in home lab, 55
- LED wired on, 57–58
- socket, 56f

BRF. *See* Band-reject filter

Broadband technology
- network with, 226–228
- wireless, 227–228

Bus, 98. *See also* Universal serial bus
- address, 135
- bidirectional, 135
- computer, 135–136
- control, 135
- data, 135, 136f
- multiplexer, 135
- network topology, 210f, 211
- parallel data transfer over, 98f
- power supply and, 86
- sources connected to, 136f

Bus-oriented architecture, 86

Byte, 92

C

C (programming language), 143

Cable, 211–213
- coax, 211–212, 211f, 266f
- fiber optic, 212–213, 213f, 214f, 268
- twisted pair, 212, 212f

Cable TV, 265–269
- diagram of, 267f, 270f
- Internet connectivity with, 226–227

CAN. *See* Controller area network

Capacitive resistance, 48

Capacitors, 47–49
- AC *vs.* DC with, 49f
- charging of, 47f, 58
- construction of, 47f
- discharging from, 47f
- in LPF, 48f
- symbol for, 47f

Carbon, as semiconductor, 20

Cascading amplifier, 71–72, 71f, 72f

Cathode ray tube (CRT), 98

CB. *See* Citizens band

CCD. *See* Charge-coupled device

CD. *See* Compact disc

CDMA. *See* Code division multiple access

Cell. *See also* Battery; Solar cells
- fuel, 32
- load of, 284
- photo, 285–289, 285f
- symbols for, 285f

Cell phone, 183
- 3G, 200, 201f
- CDMA, 190, 199
- cellular concept of, 184–187, 185f, 186f
- communication model of, 8, 9f
- components of, 187–191
- diagram of, 188f
- digital modulation in, 191–197
- duplex capability of, 188
- EDGE, 190
- femtocell as, 202–203, 203f
- GPRS, 190
- GPS in, 204

Cell phone (*Continued*)
  GSM, 189–190
  hand-off of, 186
  HSPA, 190
  location technology in, 204
  LTE, 191
  mobile television on, 203
  operation of, 187f
  as radio, 183–184
  range and reliability of, 184
  smartphone as, 201–202
  spread spectrum in, 164
  transmission frequency of, 28t
  TV on, 271–272
  types of, 189–191
Central processing unit (CPU)
  components of, 126
  control unit of, 126, 131–132
  instruction-word formats in, 132–133
  program execution in, 133–135
  registers in, 128, 129
Charge, 16–18
Charge-coupled device (CCD), 255
Chip. *See* Integrated circuit
Circuit, 61. *See also* Digital circuit; Linear circuit; Logic circuit
  arithmetic, 111
  filter as, 73–75
  mixer as, 76
  model of, 7–8, 8f, 19f
  photo control, 287, 287f
  theory of, 39–40
  three-state line-driver, 135
Citizens band (CB), 176, 182
Clock
  crystal, 76, 76f
  in microprocessor, 137
Coax cable, 211–212, 211f, 266f
Code. *See also* American Standard Code for Information Interchange; Binary-coded decimal; Decoder
  operation, 131, 132–133
  for resistors, 46f
Code division multiple access (CDMA)
  cell phone use of, 190, 199
  generation of, 200f
Codec, 236
Communication. *See also* Duplex communication
  cell phone model of, 8, 9f
  as industry sector, 4, 4f
  LOS, 158–159, 158f
  radio system for, 147–149, 148f
  with satellite, 178–179, 178f
  simplex, 149f, 150
Compact disc (CD), 236–239
  dynamic range of, 233
  encoding on, 238f
  past recording media *vs.*, 232
  player, diagram of, 240f
Comparator, 111
Compiler, 142
Components
  capacitors as, 47–49
  of cell phone, 187–191
  of CPU, 126
  diodes as, 50
  in electronics hierarchy, 41f
  IC as, 53–55
  inductors as, 49
  as industry sector, 4f, 5
  of microprocessor, 126
  resistors as, 44–47
switches as, 43–44, 53, 90
  transformers as, 49–50
  transistors as, 7, 50–53, 90
  types of, 43–55
Compounds, atomic elements in, 16
Compression
  of sound, 234–235
  of video, 255–258
Computer, 123–136, 127f. *See also* Central processing unit; Memory; Microprocessor
  ALU in, 10, 10f, 126

- applications of, 9
- bus, 135–136
- control with, 124–125
- diagram of, 42f
- embedded, 11
- external devices with, 125f
- FFs in, 124
- function of, 8–11
- gates in, 124
- as industry sector, 4, 4f
- input-output system of, 10–11, 10f, 124, 126
- memory of, 126
- as microcontroller, 124
- monitoring by, 125f
- personal, 124
- repairing, 40
- RISC, 139
- sections of, 9, 10f
- as system, 41
- types of, 124

Conductor, 20. *See also* Semiconductor

Control. *See also* Industrial control; Microcontroller; Programmable logic controller; Relay
- of actuator, 289–298
- bus, 135
- closed-loop, 277–281, 279f, 280f
- computers for, 124–125
- gain, 166
- as industry sector, 4–5, 4f
- open-loop, 277–281, 278f
- of output device, 289–298
- photo, 287, 287f
- robots as device for, 11
- with thyristor, 298–303

Controller, 280–281. *See also* Microcontroller; Programmable logic controller

Controller area network (CAN), 141

Copper, atomic composition of, 17f

Core, 124

Coulomb, 18–19

Counter, 109–110
- data conversion with, 109f
- FFs in, 109–110
- program, 131, 132f

CPU. *See* Central processing unit

Crossover, 231, 231f

CRT. *See* Cathode ray tube

Current flow, 7, 16–19. *See also* Alternating current; Direct current
- conventional, 19–20
- direction of, 19–20
- electrons in, 15
- in magnetic fields, 21–22, 22f
- measurement of, 18–19
- voltage in, 15

Cut-off frequency, 48–49

D

DAC. *See* Digital-to-analog converter

Data
- ALU’s processing of, 130f
- bus, 135, 136f
- on fiber optic cable, 214f
- Internet’s protocols for, 219f
- LED representation of, 92f
- parallel transmission of, 98–100, 98f
- register’s manipulation of, 129f
- retrieval of, 124
- serial transmission of, 98–100
- storage of, 124
- word, binary, 92

Data conversion, 118–122
- ADC for, 34, 118
- with counter, 109f
- DAC for, 34, 120, 121f
- resolution in, 120–122
- sampling in, 118, 119f, 120–122, 121f

Data processor. *See* Processor

dB. *See* Decibels

DC. *See* Direct current

Decibels (dB), 69–70
Decimal number system. *See also* Binary-coded decimal; Hexadecimal notation
  binary conversion from and to, 90–92
  bits' maximum value calculation in, 92–93
Decoder, 111, 112f
Demodulator, 166
Demultiplexer, 110
Demux. *See* Demultiplexer
Detector. *See* Demodulator
Development environment, 142
Diac, 301
Differential amplifier, 64, 64f, 68f
Digital audio, 231–239
Digital circuit, 89. *See also* Binary numbers; Data conversion; Gate; Logic circuit; Memory; Programmable logic device
  logic elements in, 100–110
Digital compression, 234–235
Digital light processing (DLP)
  LCD and plasma screens *vs.*, 265
  operation of, 99
  TVs made with, 97
Digital modulation
  advanced, 191
  basic, 162–164, 163f
  in cell phone, 191–197
  MIMO in, 196–197, 196f
  OFDM in, 193–194, 194f, 195f, 200
  QAM, 191–193, 192f, 193f
  spectral efficiency of, 197, 197t
Digital radio, 182, 246
Digital signal processing (DSP), 144–145, 144f
Digital subscriber line (DSL), 227
Digital video disc (DVD), 274–276
  Blu-ray as next generation of, 275–276
  player, 274–275, 275f
Digital video recorder (DVR), 270f
Digital-to-analog converter (DAC), 34, 120, 121f
Diode. *See also* Light emitting diode
  bias of, 51f
  as component, 50
  photo, 286f
  rectification by, 51f
  silicon, 286
  symbol for, 51f
  temperature sensing with, 283
  zener, 50
Dipole antenna, 170, 172f
Direct current (DC), 24–26, 25f, 29–33
  amplifier, 63
  from batteries, 29–31
  capacitors' blocking of, 49f
  conversion of, 85, 87f
  ground in, 24
  motor, 297–298
  negative, 26
  positive, 24
  power supplies for, 33
  pulsating, 26
  pulse filtering to, 78
  from solar cells, 31–32
  in solar power system, 87f
Dish, parabolic, 174–175, 175f
DLP. *See* Digital light processing
Driver
  radio amplification with, 165
  three-state line, 135
DSL. *See* Digital subscriber line
DSP. *See* Digital signal processing
Duplex communication, 149f, 150
  access methods and, 198–200
  of cell phone, 188
  FDD, 198
  TDD, 199
Duplexing, 198, 199
Duty cycle, 84–85, 85f

DVD. *See* Digital video disc
DVR. *See* Digital video recorder
Dynamic range, audio's, 233

E

Earphones. *See* Headphones, noise-cancelling
EDGE, 190
Editor, 141
Effector, end, 11
Efficiency, 72
  of amplifiers, 72
  spectral, 197, 197t
8051 microcontroller (Intel), 137, 138f
Electricity
  dynamic, 18–19
  electronics *vs.*, 6–7
  electrons in, 7, 15–20
  as energy source, 15
  production of, 15
  static, 16, 17–18
Electromagnetic induction, 23–24, 23f
Electromagnetism, 21–23. *See also* Magnetic field; Magnetism
  from antenna, 152f
  frequency spectrum of, 28t, 153–155, 154f
  in radio, 150–151
  spectrum space in, 155
  wavelength in, 156
Electronics
  as applied science, 16
  audio, 229
  concepts in, 15
  in daily life, 1
  electricity *vs.*, 6–7
  hierarchy of, 41f
  input-process-output model of, 7–8, 8f
  learning about, 2–3
Electronics industry
  communications as sector of, 4, 4f
  components as sector of, 4f, 5
  computers as sector of, 4, 4f
  control as sector of, 4–5, 4f
  instrumentation as sector of, 4f, 5
  manufacturing in, 5–6, 6f
  raw materials in, 5, 6f
  workings of, 5–7, 6f
Electrons
  in atoms, 16
  in current flow, 15
  in electricity, 7, 15–20
Electrostatic discharge (ESD), 18
Elements, atomic, 16
End effector, 11
ESD. *See* Electrostatic discharge
Ethernet
  frame of, 215f
  hub, 215f
  networking with, 213
  operation of, 213–216

F

Fast Fourier transform (FFT), 144–145
FCC. *See* Federal Communications Commission
FDD. *See* Frequency-division duplexing
FDMA. *See* Frequency division multiple access
Federal Communications Commission (FCC), 153
Feedback, in closed-loop control, 279f
Femtocell, 202–203, 203f
FF. *See* Flip-flop
FFT. *See* Fast Fourier transform
Fiber optic cable, 212–213, 213f
  data transmission over, 214f
  TV over, 268

Filter. *See also* Band-pass filter; Band-reject filter; High-pass filter; Low-pass filter
  as circuit, 73–75
  pulse to DC, 78
  SAW, 75
Flag, 131–132
Flag register, 131–132
Flip-flop (FF), 106
  cascading, 107f
  in computer, 124
  in counters, 109–110
  in storage register, 107f
  symbols for, 106f
FM. *See* Frequency modulation
Fourier theory, 34. *See also* Fast Fourier transform
  harmonics and, 66
Frequency. *See also* Amplifier
  of AC, 28
  of cell phone transmission, 28t
  cut-off, 48–49
  domain, 34, 35f
  of electromagnetic spectrum, 28t, 153–155, 154f
  Fourier theory and, 34
  intermediate, 260
  of oscillators, 75–76
  of radar transmission, 28t
  of satellite transmission, 28t
  synthesizers, 80–82, 81f
  of TV transmission, 28t
Frequency division multiple access (FDMA), 198
Frequency modulation (FM), 161–162, 162f
  sidebands on, 161
Frequency shift keying (FSK), 163, 163f
  spectral efficiency of, 197t
Frequency-division duplexing (FDD), 198
FSK. *See* Frequency shift keying
Fuel cell, 32
Function generators, 33

G

Gain
  of amplifier, 63, 72f
  control, 166
  dB for measurement of, 69–70
  power, 69–70
  voltage, 69–70
GAL. *See* Generic array logic
Gate
  AND, 101–102, 102f, 103f
  OR, 102–104, 103f
  in computer, 124
  NAND, 104, 104f
  NOR, 104, 105f
  XNOR, 105f
  XOR, 104–105, 105f
Gauge, strain, 283–284, 284f
Gaussian minimum shift keying, 189–190
General Packet Radio System (GPRS), 190
General-purpose register (GPR), 128
Generator
  AC, 33
  function, 33
Generic array logic (GAL), 116
Germanium, 20
Global positioning system (GPS), 179
  in cell phones, 204
Global System for Mobile Communications (GSM), 189–190
GPR. *See* General-purpose register
GPRS. *See* General Packet Radio System
GPS. *See* Global positioning system

Ground, 49
  in DC, 24
  symbol for, 49f
Ground-plane antenna, 170, 171f
Groupe Spécial Mobile. *See* Global System for Mobile Communications
GSM. *See* Global System for Mobile Communications

H

Ham radio, 176–177
HAN. *See* Home area network
Hard copy, 128
Hardware
  binary numbers represented by, 91
  purchasing, 36
  of radio, 164–170
Harmonics, 66
HDMI. *See* High-definition media interface
Headphones, noise-cancelling, 244–245
Hearing
  aids, 245–246
  testing of, 246
Hertz (Hz), 26
Hexadecimal notation, 97–98
High frequency radio, 157–158
High-definition media interface (HDMI), 260
High-definition TV (HDTV)
  packet in, 258f
  resolution of, 254
High-pass filter (HPF), 74f
High-speed packet access (HSPA), 190
Home area network, 209
Home entertainment system
  satellite TV in, 272f
  as system, 41
Horn antenna, 172–173
HPF. *See* High-pass filter
HSPA. *See* High-speed packet access
Hydraulic device, 298
Hydrophone, 245
Hz. *See* Hertz

I

IC. *See* Integrated circuit
Impedance, 71
  in amplifier, 70–72, 70f
  input, 70–71
  output, 71
Induction
  electromagnetic, 23–24, 23f
  self-, 49
Inductive reactance, 49
Inductors, 49, 50f
Industrial, scientific, and medical band (ISM band)
  applications of, 228
  project on, 182
  transceiver, 168–170
Industrial control, 277
  actuator in, 289–298
  closed-loop, 277–281, 279f, 280f
  controllers in, 280–281
  diagram of, 278f
  motor in, 296–297
  open-loop, 277–281, 278f
  output device in, 289–298
  relay in, 292
  sensors in, 281–289
  solenoid in, 295–296, 296f
  with thyristor, 298–303
Instruction register (IR), 131
Instruction-word formats, 132–133, 133f
Instrumentation, as industry sector, 4f, 5
Instrumentation amplifier, 69, 69f, 283–284

Instruments, musical, 243–244
Insulators, 20
Integrated circuit (IC), 2, 53–55
  transistors in, 7
Interface, 128
  high-definition media, 260
  to microcontroller, 140–141
  serial peripheral, 141
International Telecommunications Union (ITU), 153
Internet, 217–220
  on cable TV, 226–227
  diagram of, 218f
  protocols used on, 219f
Internet protocol TV (IPTV), 268–269
Interrupt, 137
Inversion, 100
  in amplifiers, 68f
  phase, 52
Inverters
  in AC, 33
  logic of, 100–101
  as power supply, 86
  symbols for, 101f
IPTV. *See* Internet protocol TV
IR. *See* Instruction register
ISM band. *See* Industrial, scientific, and medical band
ITU. *See* International Telecommunications Union

K

Kit building, 88

L

Lab, set up of, 55
LAN. *See* Local area network
Language. *See* Programming language
LCD. *See* Liquid crystal display
Lead acid battery, 30t
LED. *See* Light emitting diode
Light
  photo control circuit for, 287, 287f
  speed of, 151–152
Light emitting diode (LED), 50
  binary data represented by, 92f
  on breadboard, 57–58
  MOSFET switching of, 60
  as screen technology, 264, 265
Lightning, 17–18
Linear circuit, 61–86
  amplifier as, 63
  filter as, 73–75
  mixer as, 76
  phase detector as, 77–78
  straight-line response of, 62f
Line-of-sight communication (LOS communication), 158–159, 158f
Liquid crystal display (LCD)
  DLP and plasma screens *vs.*, 265
  operation of, 98–99
  TVs made with, 97
Lithium ion battery, 30t
Local area network (LAN), 208
Logic. *See also* Gate; Programmable logic controller
  Boolean, 100
  digital, 100–110
  generic array, 116
  of inverter, 100–101
  ladder, 307, 307f
  in PAL, 116
  in PLA, 116
  in PLD, 116–118, 117f
Logic circuit, 110–111
  ALU as, 111
  comparator as, 111
  decoder as, 111, 112f
  demultiplexer as, 110
  multiplexer as, 110, 111f
Logic gate. *See* Gate
Long-Term Evolution (LTE), 191

Loop antenna, 171–172
LOS communication. *See* Line-of-sight communication
Low frequency radio, 156–157
Low-pass filter (LPF)
  capacitors in, 48f
  function of, 74–75
  in PLL, 78
  response curve of, 74f
  symbol for, 74f
LPF. *See* Low-pass filter
LTE. *See* Long-Term Evolution

M

Magazines, 36–37
Magnetic field, 20–21, 21f
  current flow in, 21–22, 22f
  voltage induction with, 23f
Magnetic tape, 232
Magnetism, 20–24
  attraction and repulsion in, 21
  flux lines in, 21f
  magnetic fields in, 20–21, 21f
  materials exhibiting, 20
MAN. *See* Metropolitan area network
Manipulator arm, robot's, 11, 12f
Manufacturing, electronics industry, 5–6, 6f
Marine radio, 177–178
Memory. *See also* Random access memory; Read-only memory
  in computer, 126
Memory address register, 131, 132f
Mercuric oxide battery, 30t
Mesh network topology, 210, 224f
Metal oxide semiconductor-field-effect transistor (MOSFET), 52, 52f
  in class C amplifier, 66
  LED switching by, 60
  switch, 54f
Metropolitan area network (MAN), 208
Microcomputer. *See* Computer
Microcontroller, 11
  8-bit, 137–139
  16-bit, 137–139
  32-bit, 139
  64-bit, 139
  computer as, 124
  8051 (Intel), 137, 138f
  interfaces to, 140–141
  MPS430 (Texas Instruments), 139
  68HC11 (Freescale), 137
Microphone, 230
Microprocessing unit. *See* Central processing unit; Microprocessor
Microprocessor, 124. *See also* Central processing unit
  8-bit, 137–139
  16-bit, 137–139
  32-bit, 139
  64-bit, 139
  clock in, 137
  components of, 126
  instruction-word formats in, 132–133
  popular, 136–139
  RISC, 139
Microwave
  amplifier, 63
  antenna, 172–175, 174f
  propagation of, 158–159
  radio, 158–159
MIMO. *See* Multiple input multiple output
Mixer, 76–77
  as circuit, 76
  linear, 76, 77f
  nonlinear, 76, 77f
  outputs, 76–77
  in radio receiver, 166
Mobile telephone switching office, 186–187

Modulation, 159–164. *See also* Amplitude modulation; Digital modulation; Frequency modulation; Quadrature amplitude modulation
  angle, 162
  ASK in, 162
  bandwidth and, 161f
  demodulator and, 166
  OOK in, 162
  phase, 162
  pulse width, 66–67, 84
  sidebands from, 159–160
MOSFET. *See* Metal oxide semiconductor-field-effect transistor
Motor
  AC, 298
  applications of, 296
  DC, 297–298
  industrial control with, 296–297
  speed regulation of, 281f
  stepper, 298
MP3 player
  diagram of, 42f, 237f
  operation of, 236
  as system, 41
MPS430 microcontroller (Texas Instruments), 139
Multimeter, 35
  battery measurement with, 56f
Multiple input multiple output (MIMO)
  in digital modulation, 196–197, 196f
  spectral efficiency of, 197t
Multiplexer. *See also* Demultiplexer
  bus, 135
  as logic circuit, 110, 111f
Multiplexing, 198–200. *See also* Orthogonal frequency division multiplexing
Musical instruments, 243–244
Mux. *See* Multiplexer

N

Network, 207–208
  Bluetooth, 222–223
  with broadband technology, 226–228
  CAN as, 141
  DSL for, 227
  with Ethernet, 213
  HAN as, 209
  interconnection methods for, 210–211, 210f
  LAN as, 208
  MAN as, 208
  near-field, 225–226
  PAN as, 208–209
  relationships of, 209–210
  RFID as, 225–226, 225f
  SAN as, 209
  with Sonet, 216, 216t
  topologies of, 210, 210f, 211, 224f
  types of, 208–210
  ultra wideband, 224
  WAN as, 208
  Wi-Fi, 220–222
  WiMAX, 224–225, 227
  wired, 211–216
  wireless, 220–226, 221f
  ZigBee, 223–224
Nibble, 92
Nickel metal hydride battery, 30t
Nickel-cadmium battery, 30t
Noise, in radio, 149
Noise-cancelling headphones, 244–245

O

Octet, 92
OFDM. *See* Orthogonal frequency division multiplexing
Ohm's law, 44
On-off keying (OOK), 162, 163f

OOK. *See* On-off keying
Operating system (OS), 142
Operation code, 131–133
Orthogonal frequency division multiplexing (OFDM)
  digital modulation with, 193–194, 194f, 195f, 200
  spectral efficiency of, 197t
OS. *See* Operating system
Oscillator, 75–76. *See also* Clock; Voltage-controlled oscillator
  in AC, 26, 33
  frequency of, 75–76
  types of, 76f
OTN. *See* Optical transport network
Optical transport network, 217

P

PAL. *See* Programmable array logic
PAN. *See* Personal area network
Parabolic dish, 174–175, 175f
Patch antenna, 171–172, 174f
PCB. *See* Printed circuit board
Peripherals, 128. *See also* Serial peripheral interface
Personal area network (PAN), 208–209
Phase detector, 77–78, 78f
  as circuit, 77–78
Phase inversion, 52
Phase modulation, 162
Phase shift keying (PSK), 163, 163f
Phase-locked loop (PLL), 78–80
  divide ratio of, 82t
  in frequency synthesizers, 80–82, 81f
  LPF in, 78
  VCO in, 78–80
Phonograph record, 232, 233
Photo cell, 285–289, 285f
Photodiode, 286f
Photoelectric sensor, 285–289
Phototransistor, 285f, 286, 288f
Photovoltaic (PV), 31, 286
PIC processor (Microchip Technology), 137–139
Piezoelectric pressure transducer, 284
Pixel, 253, 253f, 263f, 276
PLA. *See* Programmable logic array
Plasma screen
  DLP and LCD *vs.*, 265
  operation of, 99
  TVs made with, 97
PLC. *See* Programmable logic controller
PLD. *See* Programmable logic device
PLL. *See* Phase-locked loop
Pneumatic device, 298
Polarization, radio, 151
Potential, 150–151
Potentiometer, 285
Power amplifier, 63
Power gain, 69–70
Power supply, 82–86
  battery as, 82
  bus-oriented architecture in, 86
  for DC, 33
  DC-DC converters in, 85
  inverters as, 86
  power management in, 86
  regulators in, 84–85
  solar, DC-to-AC, 87f
  standard, 82–84, 83f
PowerPC (Freescale), 139
Preamplifier, 230
Printed circuit board (PCB)
  in kit, 88
  patch antenna on, 174f
Processor, 124. *See also* Central processing unit; Microprocessor
Program
  counter, 131, 132f
  CPU's execution of, 133–135
  in RAM, 134f
  storage concept for, 135

Programmable array logic (PAL), 116
Programmable logic array (PLA), 116
Programmable logic controller (PLC), 303–308
  diagram of, 305f
  operation of, 306–307
  organization of, 305–306
  programming of, 307–308
  scan cycle of, 306f
Programmable logic device (PLD), 116–118, 117f
Programming language, 141, 142–144
  assembly, 142–143
  BASIC, 143–144
  C, 143
  with PLC, 307–308
PSK. *See* Phase shift keying
Pulse filter into DC, 78
Pulse width modulation (PWM), 66–67
  duty cycle and, 84
PV. *See* Photovoltaic
PWM. *See* Pulse width modulation

Q

QAM. *See* Quadrature amplitude modulation
Quadrature amplitude modulation (QAM)
  digital modulation with, 191–193, 192f, 193f
  spectral efficiency of, 197t
Quadrature phase shift keying. *See* Quadrature amplitude modulation

R

Radar
  project, 182
  transmission frequency of, 28t
Radio, 151. *See also* Duplex communication; Modulation; Sidebands
  in aircraft, 176
  AM, 159, 160f
  AV receiver with, 242
  bandwidth of, 160, 161f
  CB, 176, 182
  cell phone as, 183–184
  communication system, 147–149, 148f
  digital, 182, 246
  driver for, 165
  electromagnetic field as, 150–151
  family, 177
  frequency amplifier, 63
  frequency types, 28t, 156–159
  GPRS, 190
  ham, 176–177
  hardware of, 164–170
  LOS communication with, 158–159, 158f
  marine, 177–178
  microwave, 158–159
  mobile, 178
  modes of, 149–150
  multipath transmission of, 184
  noise in, 149
  polarization of, 151
  propagation of, 156–158
  range and reliability of, 184
  receiver, 165–166, 167f
  shortwave, 157f
  signal speed and strength of, 151–153
  simplex, 149f, 150
  spread spectrum in, 164
  transceiver, 168
  transmission lines in, 172, 174f
  transmitter, 164–165, 165f
  two-way, 175–178
Radio astronomy, 179–181
Radio frequency identification (RFID), 225–226, 225f

Radio telescope, 179–181

RAM. *See* Random access memory

Random access memory (RAM), 112–114
- addressing in, 113f
- dynamic, 113, 114f
- program stored in, 134f
- static, 113, 114f

Reactance, inductive, 49

Read-only memory (ROM), 114–115

Receiver. *See also* Audio-video receiver
- radio, 165–166, 167f
- satellite TV, 273f
- superheterodyne, 167f
- TV, 260

Recommended Standard 232 (RS-232), 140, 140f

Recommended Standard 485 (RS-485), 141

Record, phonograph, 232, 233

Reduced instruction set computer (RISC), 139

Reed relay, 291, 291f

Register
- accumulator, 129–130
- in CPU, 128, 129
- data manipulation in, 129f
- flag, 131–132
- general-purpose, 128
- instruction, 131
- memory address, 131, 132f
- operation of, 129–131
- shift, 108–109, 108f
- stack pointer, 131–132
- storage, 107, 107f

Regulator
- linear, 84
- in power supplies, 84–85
- switching, 84

Relay, 289–295
- common contacts of, 291f
- current control with, 292
- diagram of, 290f
- industrial control with, 292
- latching of, 293–294, 294f
- reed, 291, 291f
- remote control with, 293
- as switch, 289
- transistor with, 295f

Resistance, capacitive, 48

Resistive sensor, 282f

Resistive temperature device (RTD), 282, 282f

Resistors, 44–47
- color codes for, 46f
- symbol for, 45f
- types of, 45f
- in voltage divider, 46f

Resolution
- in data conversion, 120–122
- of HDTV, 254
- of video, 254

RFID. *See* Radio frequency identification

Ring network topology, 210f, 211

RISC. *See* Reduced instruction set computer

Robots, 11–13
- building of, 308b
- as control device, 11
- end effectors of, 11
- manipulator arm of, 11, 12f

ROM. *See* Read-only memory

Routers, 220

RS-232. *See* Recommended Standard 232

RS-485. *See* Recommended Standard 485

RTD. *See* Resistive temperature device

S

Sampling
- in data conversion, 118, 119f, 120–122, 121f
- sound, 236f

SAN. *See* Storage area network
Satellite
  communication, 178–179, 178f
  geosynchronous orbit of, 178–179
  transmission frequency of, 28t
  transponder of, 180f
Satellite TV, 269–271
  distribution system for, 271f
  in home entertainment system, 272f
  receiver diagram with, 273f
SAW. *See* Surface acoustic wave
Scanning, 248, 249f
  interlaced, 251f
  progressive, 253–254
SCR. *See* Silicon-controlled rectifier
Self-induction, 49
Semiconductor, 20. *See also* Metal oxide semiconductor-field-effect transistor; Thyristor
  carbon as, 20
  germanium as, 20
  silicon as, 20
Sensor
  of flow, 289
  in industrial control, 281–289
  photoelectric, 285–289
  potentiometer as, 285
  pressure, 283–284
  resistive, 282f
  switch as, 284–285
  of temperature, 282–283
Serial peripheral interface (SPI), 141
Shift key, 162, 163f. *See also* Quadrature amplitude modulation
  with ASK, 162, 163f
  with BPSK, 163, 163f, 197t
  with FSK, 163, 163f, 197t
  with Gaussian minimum, 189–190
  with OOK, 162, 163f
  with PSK, 163, 163f
Shift register, 108–109, 108f
Sidebands
  on FM, 161
  from modulation, 159–160
  vestigial, 259
Signal, 7. *See also* Digital signal processing
  in AC, 26, 27f
  from amplifier, 63
  analog, 33–34, 61–62
  binary, 26, 89, 89f
  digital, 26, 33–34
  distortion of, 62
  radio, 151–153
Silicon
  diode made of, 286
  as semiconductor, 20
Silicon-controlled rectifier (SCR), 299–301, 299f, 300f, 301f
Silver oxide battery, 30t
Simplex communication, 149f, 150
68HC11 microcontroller (Freescale), 137
Smartphone, 201–202
SoC. *See* System on a chip
Software, 141–144
Solar cells
  as DC source, 31–32
  in panel, 32f
  symbol for, 285f
Solar power system, DC-to-AC, 87f
Solenoid, 295–296, 296f
Sonar, 245
Sonet. *See* Synchronous optical network
Sound, 229–231
  compression of, 234–235
  digitization of, 233–234, 234f
  frequency of, 28t
  sampling of, 236f
  in special applications, 243–246
  surround, 242–243, 243f
Speakers, 230–231
  diagram of, 230f
  surround sound with, 242–243, 243f

  as transducer, 230
  types of, 231
Spectral efficiency, 197, 197t
Spectrum
  electromagnetic frequency, 28t, 153–155, 154f
  space, 155
  spread, 164
SPI. *See* Serial peripheral interface
Spread spectrum, 164
Stack, 132
Stack pointer register, 131–132
Star network topology, 210f, 211
Static electricity, 16, 17–18
Stepper motor, 298
Storage area network (SAN), 209
Storage register, 107
  FFs in, 107f
Stored program concept, 135
Strain gauge, 283–284, 284f
Subroutines, 142
Surface acoustic wave (SAW), 75
Surround sound, 242–243, 243f
Switch, 43–44
  as component, 43–44, 53, 90
  limit, 284–285
  MOSFET, 54f
  relay as, 289
  schematic of, 44f
  as sensor, 284–285
Switching
  amplifier, 66–67, 67f
  of LED, 60
  mobile telephone, 186–187
  regulator, 84
  transistor, 53
Synchronous digital hierarchy. *See* Synchronous optical network
Synchronous optical network (Sonet), 216, 216t
Synthesizer, frequency, 80–82, 81f
System on a chip (SoC), 61
Systems approach, 1, 39–42

T

Tachometer, phototransistor in, 288f
TDD. *See* Time-division duplexing
TDMA. *See* Time division multiple access
Telephone, mobile switching of, 186–187
Telephone data system, 217
  ATM in, 217
Television (TV). *See also* Cable TV; High-definition TV; Internet protocol TV; Satellite TV
  ADC for, 252f
  analog, 250–252
  aspect ratio of, 248, 249f
  on cell phone, 271–272
  diagram of, set, 105f
  digital, 252–254
  with DLP, 97
  frequency of, 28t
  inputs and outputs of, set, 105f
  international systems for, 251
  with LCD, 97
  mobile, 203
  over fiber optic cable, 268
  with plasma screen, 97
  receiver, 260
  screen technology of, 100–102, 106f
  transmission of, 258–259, 258f
Temperature sensor, 282–283
Thermistor, 282–283
Thermocouple, 283
Thyristor
  control with, 298–303
  diac, 301
  SCR, 299–301, 299f, 300f, 301f
  triac, 301–303, 304f
Time division multiple access (TDMA), 198–199, 199f
Time-division duplexing (TDD), 199
Timer, 137

Transceiver
  ISM band, 168–170
  radio, 168
  single-chip, 169f
Transducer
  piezoelectric pressure, 284
  speaker as, 230
Transformer, 49–50, 50f
Transistor, 50–53. *See also* Metal oxide semiconductor-field-effect transistor
  bipolar junction, 50, 52f, 53f
  as component, 7, 50–53, 90
  IC in, 7
  photo-, 285f, 286, 288f
  relay with, 295f
  switching, 53
  symbols for, 52f
  temperature sensing with, 283
  types of, 50
Transmitter
  radio, 164–165, 165f
  TV, 258–259, 258f
  UART as, 140, 140f
Transponder, satellite, 180f
Triac, 301–303, 304f
Truth table, 101
  for AND gate, 102f
  for OR gate, 103f
  inverter for, 101f
  for NAND gate, 104f
  for NOR gate, 105f
  for XNOR gate, 105f
  for XOR gate, 105f
TV. *See* Television
Twisted pair cable, 212, 212f

U

UART. *See* Universal asynchronous receiver-transmitter
Ultra high frequency radio, 158–159
Ultra wideband network, 224
Universal asynchronous receiver-transmitter (UART), 140, 140f
Universal serial bus (USB), 141
USB. *See* Universal serial bus

V

Vacuum tubes, 244
Valve, 298
Varactor, 50
VCO. *See* Voltage-controlled oscillator
Very high frequency radio, 158–159
Video, 247–264
  aspect ratio of, 248, 249f
  CCD for, 255
  color, 254–255, 256f, 257f
  compression of, 255–258
  monochrome, 254–255
  persistence of vision in, 250
  resolution of, 254
  scanning in, 248, 249f, 251f, 253–254
Vision, persistence of, 250
Voice
  coil, 230–231
  sampling of, 236f
Voltage, 7, 16–18
  of AC, 28–29
  of battery cell, 30t
  in current flow, 15
  divider, 46f
  as electrical potential difference, 17
  gain, 69–70
  induction of, 23f
  measurement of, 37–38
  sources of, 24–29
Voltage-controlled oscillator (VCO), 75, 76f
  in PLL, 78–80

W

WAN. *See* Wide area networks
Wave. *See also* Microwave
  short-, 157f
  surface acoustic, 75
Waveguide, 172–175
Wavelength, in electromagnetism, 156
Wide area networks (WAN), 208
Wi-Fi, 220–222
WiMAX, 224–225, 227
Wireless, 147. *See also* Radio
  broadband, 227–228
  networks, 220–226, 221f
  trend in, 181
Word
  binary data, 92
  instruction, 132–133, 133f

Y

Yagi antenna, 170–171, 173f

Z

ZigBee, 223–224

Atom

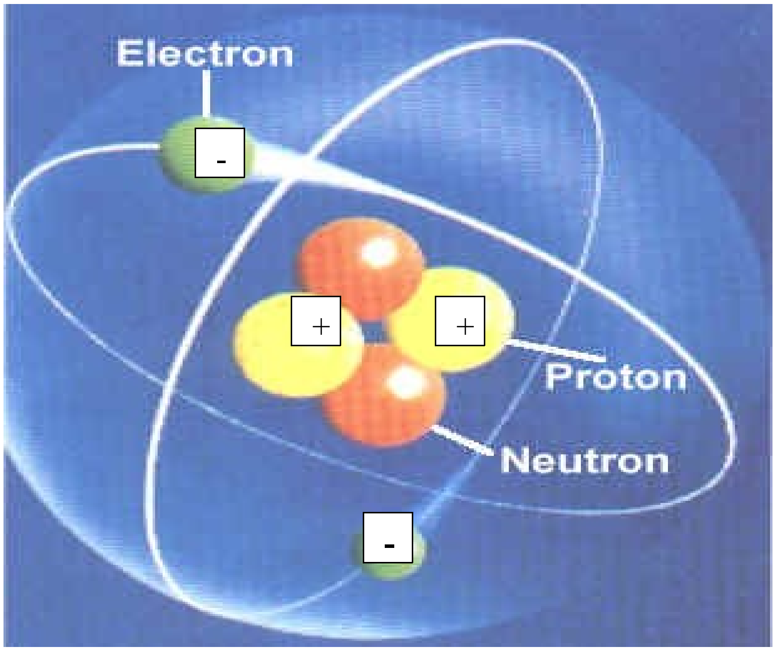

The **atom** is a basic unit of matter consisting of a dense, central nucleus surrounded by a cloud of negatively charged electrons. The atomic nucleus contains a mix of positively charged protons and electrically neutral neutrons (except in the case of Hydrogen-1, which is the only stable nuclide with no neutron). The electrons of an atom are bound to the nucleus by the electromagnetic force. Likewise, a group of atoms can remain bound to each other, forming a molecule. An atom containing an equal number of protons and electrons is electrically neutral, otherwise it has a positive or negative charge and is an ion. An atom is classified according to the number of protons and neutrons in its nucleus: the number of protons determines the chemical element, and the number of neutrons determine the isotope of the element.

Electron

The **electron** is a subatomic particle that carries a negative electric charge. It has no known substructure and is believed to be a point particle.[2] Electrons participate in gravitational, electromagnetic and weak interactions. Like its rest mass and elementary charge, the intrinsic angular momentum (or spin) of an electron has a constant value. In the collision of an electron and a positron, the electron's antiparticle, both are annihilated. An electron–positron pair can be produced from gamma ray photons with sufficient energy.

Proton

The **proton** is a subatomic particle with an electric charge of +1 elementary charge. It is found in the nucleus of each atom but is also stable by itself and has a second identity as the hydrogen ion, H^+. It is composed of 3 even more fundamental particles comprising two up quarks and one down quark.

Neutron

The **neutron** is a subatomic particle with no net electric charge and a mass slightly larger than that of a proton.

Neutrons are usually found in atomic nuclei. The nuclei of most atoms consist of protons and neutrons, which are therefore collectively referred to as nucleons. The number of protons in a nucleus is the atomic number and defines the type of element the atom forms. The number of neutrons determines the isotope of an element. For example, the carbon-12 isotope has 6 protons and 6 neutrons, while the carbon-14 isotope has 6 protons and 8 neutrons.

Electrical conductor

In science and engineering, an electrical **conductor** is a material which contains movable electric charges. In metallic conductors, such as copper or aluminum, the movable charged particles are electrons (See electrical conduction). Positive charges may also be mobile in the form of atoms in a lattice missing electrons (called "holes") or ions, such as in the electrolyte of a battery.

Note: The following applies to direct current only. When the direction of voltage/current alternates, other effects (inductive and capacitive reactance) come into play also.

All conductors contain electric charges which will move when an electric potential difference (measured in volts) is applied across separate points on the material. This flow of charge (measured in amperes) is what is meant by *electric current*. In most materials, the rate of current is proportional to the voltage (as determined by Ohm's law) provided the temperature remains constant and the material remains in the same shape and state.

Most familiar conductors are metallic. Copper is the most common material for electrical wiring (silver is the best but expensive), and gold for high-quality surface-to-surface contacts. However, there are also many non-metallic conductors, including graphite, solutions of salts, and all plasmas. See electrical conduction for more information on the physical mechanism for charge flow in materials.

Non-conducting materials lack mobile charges, and so resist the flow of electric current, generating heat. In fact, all non-superconducting materials offer some resistance and warm up when a current flows. Thus, proper design of an electrical conductor takes into account the temperature that the conductor needs to be able to endure without damage, as well as the quantity of electrical current. The motion of charges also creates an electromagnetic field around the conductor that exerts a mechanical radial squeezing force on the conductor. A conductor of a given material and volume (length x cross-sectional area) has no real limit to the current it can carry without being destroyed as long as the heat generated by the resistive loss is removed and the conductor can withstand the radial forces. This effect is especially critical in printed circuits, where conductors are relatively small and close together, and inside an enclosure: the heat produced, if not properly removed, can cause fusing (melting) of the tracks.

Since all non-superconducting conductors have some resistance, and all insulators will carry some current, there is no theoretical dividing line between conductors and insulators. However, there is a large gap between the conductance of materials that will carry a *useful current* at working voltages and those that will carry a negligible current for the purpose in hand, so the categories of *insulator* and *conductor* do have practical utility.

Thermal and electrical conductivity often go together (for instance, most metals are both electrical and thermal conductors). However, some materials are practical electrical conductors without being a good thermal conductor.

(Electrical) Non Conductor (Insulator)

An **insulator**, also called a *dielectric*, is a material that resists the flow of electric current. An insulating material has atoms with tightly bonded valence electrons. These materials are used in parts of electrical equipment, also called *insulators* or *insulation*, intended to support or separate electrical conductors without passing current through themselves. The term is also used more specifically to refer to insulating supports that attach electric power transmission wires to utility poles or pylons.

Some materials such as glass or Teflon are very good electrical insulators. A much larger class of materials, for example rubber-like polymers and most plastics are still "good enough" to insulate electrical wiring and cables even though they may have lower bulk resistivity. These materials can serve as practical and safe insulators for low to moderate voltages (hundreds, or even thousands, of volts).

Semiconductor

A **semiconductor** is a material that has electrical conductivity between those of a conductor and an insulator. The conductivity of a semiconductor material can be varied under an external electrical field. Devices made from semiconductor materials are the foundation of modern electronics, including radio, computers, telephones, and many other devices. Semiconductor devices include the transistor, many kinds of diodes including the light-emitting diode, the silicon controlled rectifier, and digital and analog integrated circuits. Solar photovoltaic panels are large semiconductor devices that directly convert light energy into electrical energy.

In a metallic conductor, current is carried by the flow of electrons. In semiconductors, current can be carried either by the flow of electrons or by the flow of positively-charged "holes" in the electron structure of the material.

Silicon is used to create most semiconductors commercially. Dozens of other materials are used, including germanium, gallium arsenide, and silicon carbide. A pure semiconductor is often called an “intrinsic” semiconductor. The conductivity, or ability to conduct, of semiconductor material can be drastically changed by adding other elements, called “impurities” to the melted intrinsic material and then allowing the melt to solidify into a new and different crystal. This process is called "doping".

CURRENT:

The flow of Electron through a conductor is called current. The no. of electronics moving through a conductor per second called Ampere.

The unit of current is **“Ampere”.**

Sub Unit:

nA = Nano Ampere = 10^{-9} = 1/1000000000
µA = Micro Ampere = 10^{-6} = 1-1000000
mA = Milli Ampere = 10^{-3} = 1/1000
A = Ampere = 1
KA = Kilo Ampere = 10^{3} = 1000
MA = Mega Ampere = 10^{6} = 1000000
GA = Giga Ampere = 10^{9} = 1000000000

current Conversion:

1000 nA = 1 µA

1000 µ A = 1 mA

1000 mA = 1 A

1000 A = 1 KA

1000 KA = 1 MA

1000 MA = 1 GA

Types of Current:
There are two types 1. AC Current
2. DC Current

AC – Current & DC - Current

Alternating Current (AC)

Alternating Current (AC) flows one way, then the other way, continually reversing direction.

An AC voltage is continually changing between positive (+) and negative (-).

The rate of changing direction is called the **frequency** of the AC and it is measured in **hertz (Hz)** which is the number of forwards-backwards **cycles per second**.

Mains electricity in the UK has a frequency of 50Hz.

See below for more details of signal properties.

An AC supply is suitable for powering some devices such as lamps and heaters but almost all electronic circuits require a steady DC supply (see below).

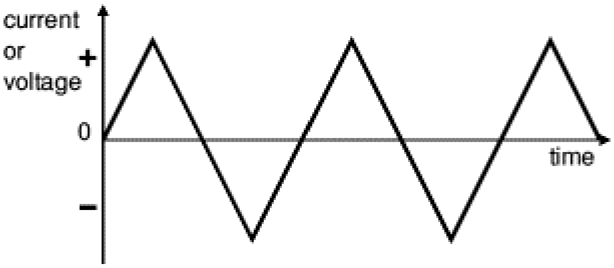

This triangular signal is AC because it changes between positive (+) and negative (-).

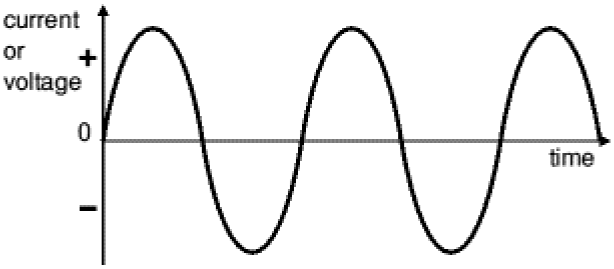

AC from a power supply
This shape is called a **sine wave**.

Direct Current (DC)

Direct Current (DC) always flows in the same direction, but it may increase and decrease.

A DC voltage is always positive (or always negative), but it may increase and decrease.

Electronic circuits normally require a **steady DC** supply which is constant at one value or a **smooth DC** supply which has a small variation called **ripple**.

Cells, batteries and regulated power supplies provide **steady DC** which is ideal for electronic circuits.

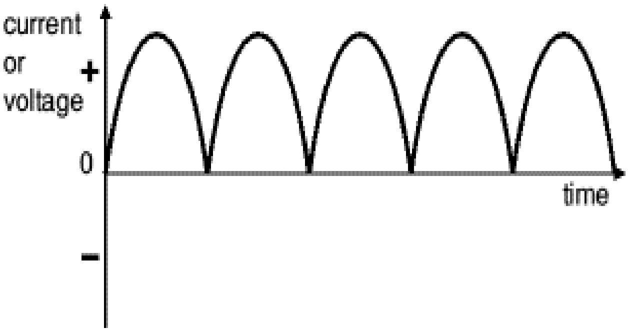

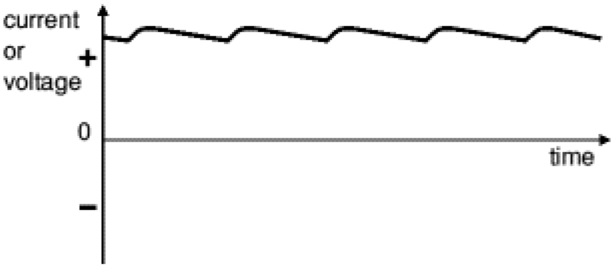

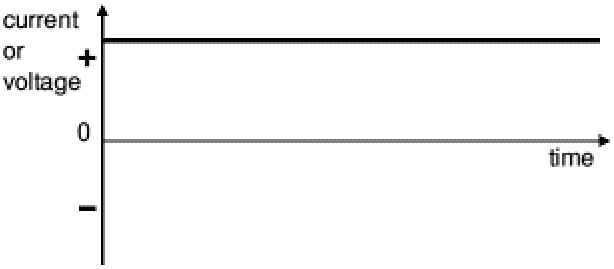

Power supplies contain a transformer which converts the mains AC supply to a safe low voltage AC. Then the AC is converted to DC by a bridge rectifier but the output is **varying DC** which is unsuitable for electronic circuits.

Some power supplies include a capacitor to provide **smooth DC** which is suitable for less-sensitive electronic circuits, including most of the projects on this website.

Lamps, heaters and motors will work with any DC supply.

Voltage

What is voltage?

Voltage is a measure of the electric force available to cause the movement or flow of electrons. Thus, voltage in itself implies no movement of electrons, but the potential to cause electrons to move.

Voltage measurements of direct current

When an electric force is available to cause the movement of electrons, a voltmeter is used to measure the potential. When that potential is unchanging, it is said to be a direct current or DC potential. DC electricity typically comes from a battery, but may come from a filtered, rectified power supply. More on rectification and filtering later.

Voltage measurements of alternating current

When an electrical force is available to cause the movement of electrons, it can sometimes not be measured accurately because the value is changing instant-by-instant. In a typical generator, for example, it can be changing in value between -110Volts and +110 volts in a sinusoidal (sine wave) fashion. Voltage that changes instant-by-instant, such as your household power, is called AC or Alternating Current.

In these cases, a rectified value is extracted and filtered in order that an average, 'positive' voltage value can be measured. More on the mechanism of rectification later but know that this is a method of converting AC into DC electricity.

Voltage:

The difference of Electric potential which exists between two points of conduction carrying a constant of one Ampere. When the power dissipate between these points in one watt.

The unit of Voltage **"Volt" = V**

nV = Nano Volt = 10^{-9} = 1/1000000000
µV = Micro Volt = 10^{-6} = 1-1000000
mV = Milli Volt = 10^{-3} = 1/1000
V = Volt = 1
KV = Kilo Volt = 10^{3} = 1000
MV = Mega Volt = 10^{6} = 1000000
GV = Giga Volt = 10^{9} = 1000000000

Current Conversion:

1000 nV = 1 µV

1000 µ V = 1 mV

1000mV = 1 V

1000 V = 1 KV

1000 KV = 1 MV

1000 MV = 1 GV

There are two types of Voltage

AC Voltage & DC Voltage.

Electric power

Electric power is defined as the rate at which electrical energy is transferred by an electric circuit. The SI unit of power is the watt.

When electric current flows in a circuit, it can transfer energy to do mechanical or thermodynamic work. Devices convert electrical energy into many useful forms, such as heat (electric heaters), light (light bulbs), motion (electric motors), sound (loudspeaker) or chemical changes. Electricity can be produced mechanically by generation, or chemically, or by direct conversion from light in photovoltaic cells, also it can be stored chemically in batteries.

The electric power in watts associated with a complete electric circuit or a circuit component represents the rate at which energy is converted from the electrical energy of the moving charges to some other form, e.g., heat, mechanical energy, or energy stored in electric fields or magnetic fields. For a resistor in a D C Circuit the power is given by the product of applied voltage and the electric current:

P = VI

Power = Voltage x Current

☐ Watts = ☐ volts x ☐ amperes

Enter data in any two of the boxes, then click on the active text for the quantity you wish to calculate.

Ohm's law

$$\frac{V}{I * R}$$

V = I*R
I = V/R
R =V/I

Ohm's law applies to electrical circuits; it states that the current through a conductor between two points is directly proportional to the potential difference or voltage across the two points, and inversely proportional to the resistance between them.

The mathematical equation that describes this relationship is:

Where I is the current in amperes, V is the potential difference in volts, and R is a circuit parameter called the resistance (measured in ohms, also equivalent to volts per ampere). The potential difference is also known as the voltage drop, and is sometimes denoted by U, E or emf (electromotive force) instead of V.

The law was named after the German physicist Georg Ohm, who, in a treatise published in 1827, described measurements of applied voltage and current through simple electrical circuits containing various lengths of wire. He presented a slightly more complex equation than the one above to explain his experimental results. The above equation is the modern form of Ohm's law.

The resistance of most resistive devices (resistors) is constant over a large range of values of current and voltage. When a resistor is used under these conditions, the resistor is referred to as an *ohmic device* (or an *ohmic resistor*) because a single value for the resistance suffices to describe the resistive behavior of the device over the range. When sufficiently high voltages are applied to a resistor, forcing a high current through it, the device is no longer ohmic because its resistance, when measured under such electrically stressed conditions, is different (typically greater) from the value measured under standard conditions (see temperature effects, below).

Ohm's law, in the form above, is an extremely useful equation in the field of electrical/electronic engineering because it describes how voltage, current and resistance are interrelated on a "macroscopic" level, that is, commonly, as circuit elements in an electrical circuit. Physicists who study the electrical properties of matter at the microscopic level use a closely related and more general vector equation, sometimes also referred to as Ohm's law, having variables that are closely related to the I, V and R scalar variables of Ohm's law, but are each functions of position within the conductor. See the Physics and Relation to heat conduction sections below.

Electrical circuits consist of electrical devices connected by wires (or other suitable conductors). (See the article electrical circuits for some basic combinations.) The above diagram shows one of the simplest electrical circuits that can be constructed. One electrical device is shown as a circle with + and - terminals, which represents a voltage source such as a battery. The other device is illustrated by a zig-zag symbol and has an R beside it. This symbol represents a resistor, and the R designates its resistance. The + or positive terminal of the voltage source is connected to one of the terminals of the resistor using a wire of negligible resistance, and through this wire a current I is shown, in a specified direction illustrated by the arrow. The other terminal of the resistor is connected to the - or negative terminal of the voltage source by a second wire. This configuration forms a complete circuit because all the current that leaves one terminal of the voltage source must return to the other terminal of the voltage source. (While not shown, because electrical engineers assume that it exists, there is an implied current I, and an arrow pointing to the left, associated with the second wire.)

Voltage is the electrical driver that moves (negatively charged) electrons through wires and electrical devices, current is the rate of electron flow, and resistance is the property of a resistor (or other device that obeys Ohm's law) that limits current to an amount proportional to the applied voltage. So, for a given resistance R (ohms),

and a given voltage V (volts) established across the resistance, Ohm's law provides the equation (I=V/R) for calculating the current through the resistor (or device).

The "conductor" mentioned by Ohm's law is a circuit element across which the voltage is measured. Resistors are conductors that slow down the passage of electric charge. A resistor with a high value of resistance, say greater than 10 megohms, is a poor conductor, while a resistor with a low value, say less than 0.1 ohm, is a good conductor. (Insulators are materials that, for most practical purposes, do not allow a current when a voltage is applied.)

In a circuit diagram, like the one above, the various components may be joined by connectors, contacts, welds or solder joints of various kinds, but for simplicity these connections are usually not shown.

Electronic Symbols

Wires and connections

Component	Circuit Symbol	Function of Component
Wire	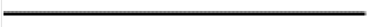	To pass current very easily from one part of a circuit to another.
Wires joined	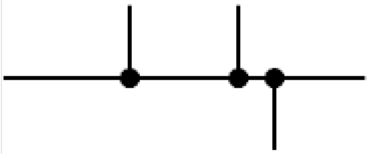	A 'blob' should be drawn where wires are connected (joined), but it is sometimes omitted. Wires connected at 'crossroads' should be staggered slightly to form two T-junctions, as shown on the right.
Wires not joined	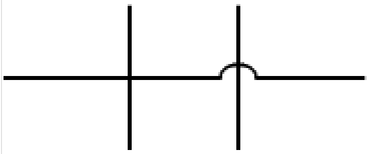	In complex diagrams it is often necessary to draw wires crossing even though they are not connected. I prefer the 'bridge' symbol shown on the right because the simple crossing on the left may be misread as a join where you have forgotten to add a 'blob'!

Power Supplies

Component	Circuit Symbol	Function of Component
Cell	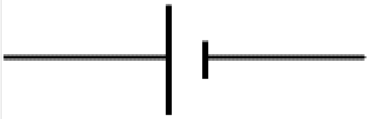	Supplies electrical energy. The larger terminal (on the left) is positive (+). A single cell is often called a battery, but strictly a battery is two or more cells joined together.
Battery	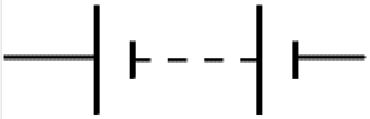	Supplies electrical energy. A battery is more than one cell. The larger terminal (on the left) is positive (+).
DC supply	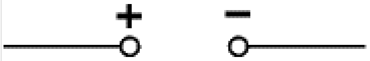	Supplies electrical energy. DC = Direct Current, always flowing in one direction.
AC supply	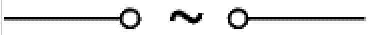	Supplies electrical energy. AC = Alternating Current, continually changing direction.
Fuse	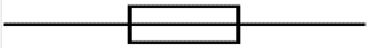	A safety device which will 'blow' (melt) if the current flowing through it exceeds a specified value.
Transformer	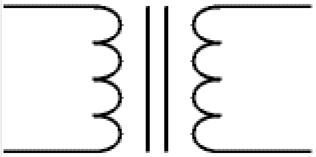	Two coils of wire linked by an iron core. Transformers are used to step up (increase) and step down (decrease) AC voltages. Energy is transferred between the coils by the magnetic field in the core. There is no electrical connection between the coils.

Earth (Ground)	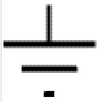	A connection to earth. For many electronic circuits this is the 0V (zero volts) of the power supply, but for mains electricity and some radio circuits it really means the earth. It is also known as ground.

Output Devices: Lamps, Heater, Motor, etc.

Component	**Circuit Symbol**	**Function of Component**
Lamp (lighting)	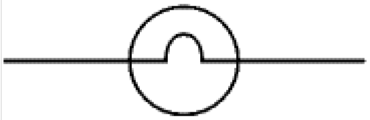	A transducer which converts electrical energy to light. This symbol is used for a lamp providing illumination, for example a car headlamp or torch bulb.
Lamp (indicator)	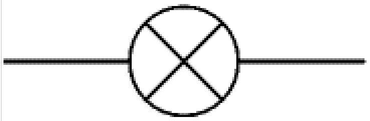	A transducer which converts electrical energy to light. This symbol is used for a lamp which is an indicator, for example a warning light on a car dashboard.
Heater	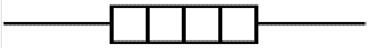	A transducer which converts electrical energy to heat.
Motor	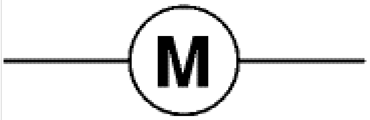	A transducer which converts electrical energy to kinetic energy (motion).
Bell	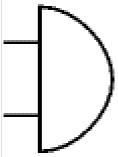	A transducer which converts electrical energy to sound.
Buzzer	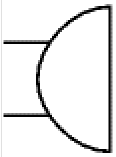	A transducer which converts electrical energy to sound.
Inductor (Coil, Solenoid)	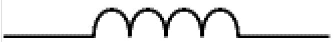	A coil of wire which creates a magnetic field when current passes through it. It may have an iron core inside the coil. It can be used as a transducer converting electrical energy to mechanical energy by pulling on something.

Switches

Component	**Circuit Symbol**	**Function of Component**
Push Switch (push-to-make)	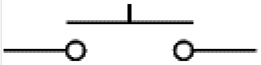	A push switch allows current to flow only when the button is pressed. This is the switch used to operate a doorbell.
Push-to-Break Switch	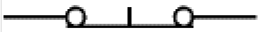	This type of push switch is normally closed (on), it is open (off) only when the button is pressed.
On-Off Switch (SPST)	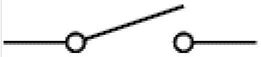	SPST = Single Pole, Single Throw. An on-off switch allows current to flow only when it is in the closed (on) position.
2-way Switch (SPDT)	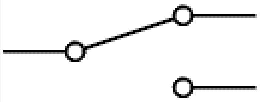	SPDT = Single Pole, Double Throw. A 2-way changeover switch directs the flow of current to one of two routes according to its position. Some SPDT switches have a central off position and are described as 'on-off-on'.

Dual On-Off Switch (DPST)	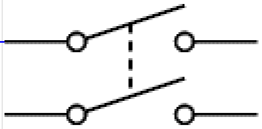	DPST = Double Pole, Single Throw. A dual on-off switch which is often used to switch mains electricity because it can isolate both the live and neutral connections.
Reversing Switch (DPDT)	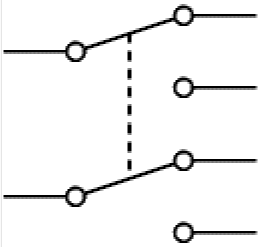	DPDT = Double Pole, Double Throw. This switch can be wired up as a reversing switch for a motor. Some DPDT switches have a central off position.
Relay	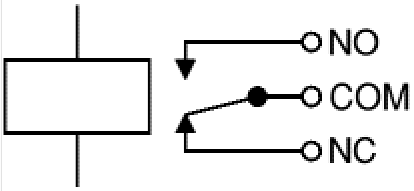	An electrically operated switch, for example a 9V battery circuit connected to the coil can switch a 230V AC mains circuit. NO = Normally Open, COM = Common, NC = Normally Closed.

Resistors

Component	**Circuit Symbol**	**Function of Component**
Resistor	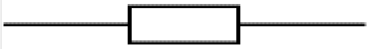	A resistor restricts the flow of current, for example to limit the current passing through an LED. A resistor is used with a capacitor in a timing circuit. Some publications still use the old resistor symbol: 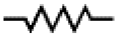
Variable Resistor (Rheostat)	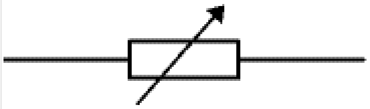	This type of variable resistor with 2 contacts (a rheostat) is usually used to control current. Examples include: adjusting lamp brightness, adjusting motor speed, and adjusting the rate of flow of charge into a capacitor in a timing circuit.
Variable Resistor (Potentiometer)	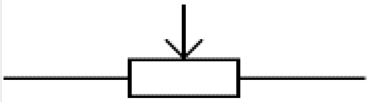	This type of variable resistor with 3 contacts (a potentiometer) is usually used to control voltage. It can be used like this as a transducer converting position (angle of the control spindle) to an electrical signal.
Variable Resistor (Preset)	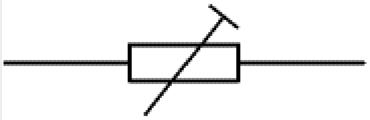	This type of variable resistor (a preset) is operated with a small screwdriver or similar tool. It is designed to be set when the circuit is made and then left without further adjustment. Presets are cheaper

		than normal variable resistors so they are often used in projects to reduce the cost.

Capacitors

Component	Circuit Symbol	Function of Component
Capacitor	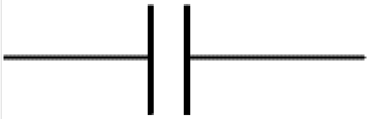	A capacitor stores electric charge. A capacitor is used with a resistor in a timing circuit. It can also be used as a filter, to block DC signals but pass AC signals.
Capacitor, polarised	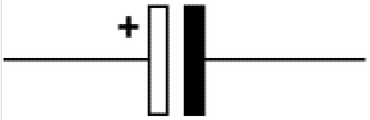	A capacitor stores electric charge. This type must be connected the correct way round. A capacitor is used with a resistor in a timing circuit. It can also be used as a filter, to block DC signals but pass AC signals.
Variable Capacitor	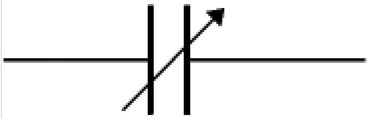	A variable capacitor is used in a radio tuner.
Trimmer Capacitor	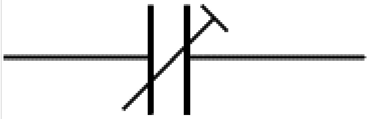	This type of variable capacitor (a trimmer) is operated with a small screwdriver or similar tool. It is designed to be set when the circuit is made and then left without further adjustment.

Diodes

Component	Circuit Symbol	Function of Component
Diode	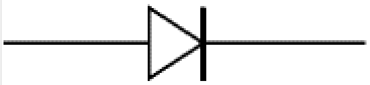	A device which only allows current to flow in one direction.
LED Light Emitting Diode	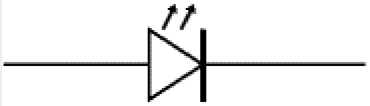	A transducer which converts electrical energy to light.
Zener Diode	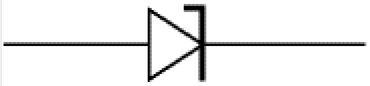	A special diode which is used to maintain a fixed voltage across its terminals.
Photodiode	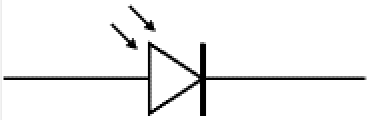	A light-sensitive diode.

Transistors

Component	Circuit Symbol	Function of Component
Transistor NPN	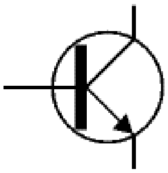	A transistor amplifies current. It can be used with other components to make an amplifier or switching circuit.
Transistor PNP	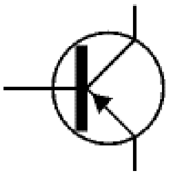	A transistor amplifies current. It can be used with other components to make an amplifier or switching circuit.
Phototransistor	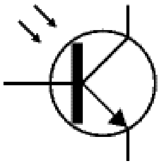	A light-sensitive transistor.

Audio and Radio Devices

Component	Circuit Symbol	Function of Component
Microphone	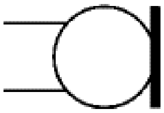	A transducer which converts sound to electrical energy.
Earphone	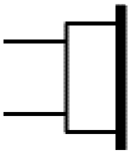	A transducer which converts electrical energy to sound.
Loudspeaker	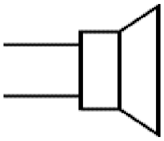	A transducer which converts electrical energy to sound.
Piezo Transducer	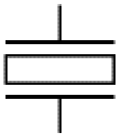	A transducer which converts electrical energy to sound.
Amplifier (general symbol)	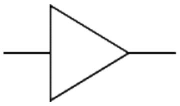	An amplifier circuit with one input. Really it is a block diagram symbol because it represents a circuit rather than just one component.

Aerial (Antenna)	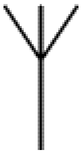	A device which is designed to receive or transmit radio signals. It is also known as an antenna.

Meters and Oscilloscope

Component	Circuit Symbol	Function of Component
Voltmeter	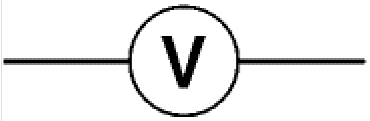	A voltmeter is used to measure voltage. The proper name for voltage is 'potential difference', but most people prefer to say voltage!
Ammeter	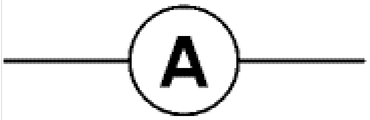	An ammeter is used to measure current.
Galvanometer	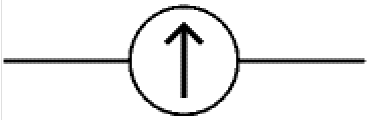	A galvanometer is a very sensitive meter which is used to measure tiny currents, usually 1mA or less.
Ohmmeter	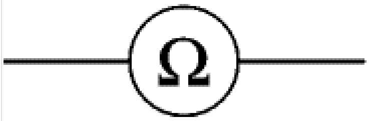	An ohmmeter is used to measure resistance. Most multimeters have an ohmmeter setting.
Oscilloscope	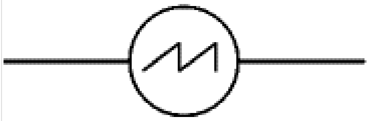	An oscilloscope is used to display the shape of electrical signals and it can be used to measure their voltage and time period.

Sensors (input devices)

Component	Circuit Symbol	Function of Component
LDR	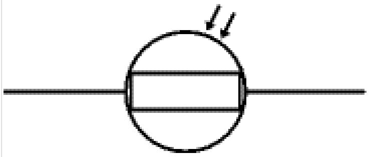	A transducer which converts brightness (light) to resistance (an electrical property). LDR = Light Dependent Resistor
Thermistor	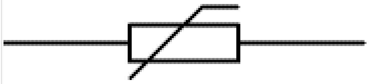	A transducer which converts temperature (heat) to resistance (an electrical property).

Logic Gates

Logic gates process signals which represent **true** (1, high, +Vs, on) or **false** (0, low, 0V, off). For more information please see the Logic Gates page. There are two sets of symbols: traditional and IEC (International Electrotechnical Commission).

Gate Type	Traditional Symbol	IEC Symbol	Function of Gate
NOT	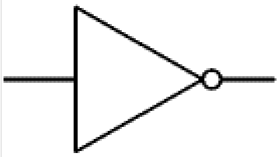	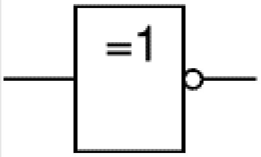	A NOT gate can only have one input. The 'o' on the output means 'not'. The output of a NOT gate is the inverse (opposite) of its input, so the output is true when the input is false. A NOT gate is also called an inverter.
AND	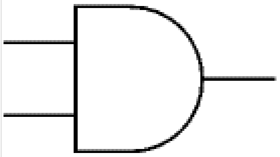	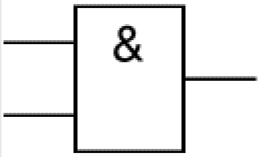	An AND gate can have two or more inputs. The output of an AND gate is true when all its inputs are true.
NAND	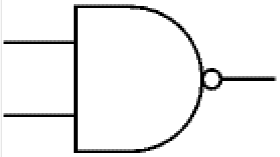	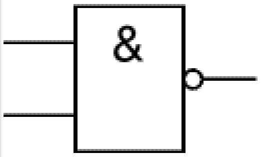	A NAND gate can have two or more inputs. The 'o' on the output means 'not' showing that it is a Not AND gate. The output of a NAND gate is true unless all its inputs are true.
OR	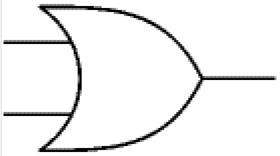	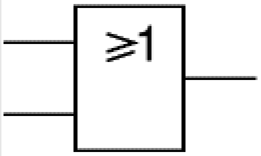	An OR gate can have two or more inputs. The output of an OR gate is true when at least one of its inputs is true.
NOR	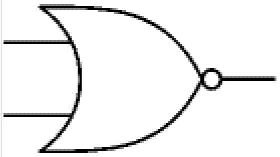	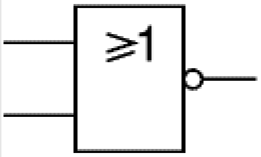	A NOR gate can have two or more inputs. The 'o' on the output means 'not' showing that it is a Not OR gate. The output of a NOR gate is true when none of its inputs are true.
EX-OR	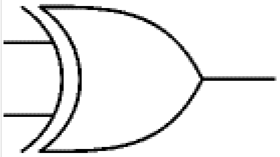	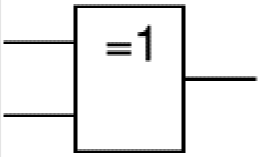	An EX-OR gate can only have two inputs. The output of an EX-OR gate is true when its inputs are different (one true, one false).
EX-NOR	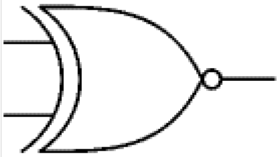	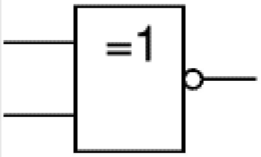	An EX-NOR gate can only have two inputs. The 'o' on the output means 'not' showing that it is a Not EX-OR gate. The output of an EX-NOR gate is true when its inputs are the same (both true or both false).

Multimeters

Multimeters are very useful test instruments. By operating a multi-position switch on the meter they can be quickly and easily set to be a voltmeter, an ammeter or an ohmmeter. They have several settings (called 'ranges') for each type of meter and the choice of AC or DC. Some multimeters have additional features such as transistor testing and ranges for measuring capacitance and frequency.

Choosing a multimeter

The photographs below show modestly priced multimeters which are suitable for general electronics use, you should be able to buy meters like these for less than £15. A digital multimeter is the best choice for your first multimeter; even the cheapest will be suitable for testing simple projects.

If you are buying an analogue multimeter make sure it has a high sensitivity of 20k/V or greater on DC voltage ranges, anything less is not suitable for electronics. The sensitivity is normally marked in a corner of the scale, ignore the lower AC value (sensitivity on AC ranges is less important), the higher DC value is the critical one. Beware of cheap analogue multimeters sold for electrical work on cars because their sensitivity is likely to be too low.

Digital multimeters

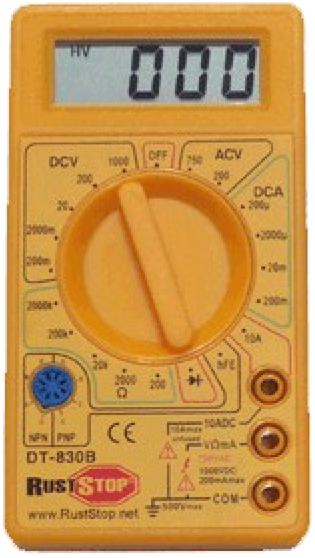

All digital meters contain a battery to power the display so they use virtually no power from the circuit under test. This means that on their DC voltage ranges they have a very high resistance (usually called input impedance) of 1M or more, usually 10M, and they are very unlikely to affect the circuit under test.

Typical ranges for digital multimeters like the one illustrated:

(The values given are the maximum reading on each range)

DC Voltage: 200mV, 2000mV, 20V, 200V, 600V.

AC Voltage: 200V, 600V.

DC Current: 200µA, 2000µA, 20mA, 200mA, 10A*.

*The 10A range is usually unfused and connected via a special socket.

AC Current: None. (You are unlikely to need to measure this).

Resistance: 200 ohms, 2000 ohms, 20k ohms, 200k ohms, 2000k ohms, Diode Test.

Digital meters have a special diode test setting because their resistance ranges cannot be used to test diodes and other semiconductors.

Analogue multimeters

Analogue meters take a little power from the circuit under test to operate their pointer. They must have a high sensitivity of at least 20k/V or they may upset the circuit under test and give an incorrect reading. See the section below on sensitivity for more details.

Batteries inside the meter provide power for the resistance ranges, they will last several years but you should avoid leaving the meter set to a resistance range in case the leads touch accidentally and run the battery flat.

Typical ranges for analogue multimeters like the one illustrated:

(The voltage and current values given are the maximum reading on each range)

DC Voltage: 0.5V, 2.5V, 10V, 50V, 250V, 1000V.

AC Voltage: 10V, 50V, 250V, 1000V.

DC Current: 50µA, 2.5mA, 25mA, 250mA.

PHOTOGRAPHERS
DIRECT.COM

A high current range is often missing from this type of meter.

AC Current: None. (You are unlikely to need to measure this).

Resistance: 20 ohms, 200 ohms, 2k ohms, 20k ohms, 200k ohms.

These resistance values are in the middle of the scale for each range.

It is a good idea to leave an analogue multimeter set to a DC voltage range such as 10V when not in use. It is less likely to be damaged by careless use on this range, and there is a good chance that it will be the range you need to use next anyway!

Sensitivity of an analogue multimeter

Multimeters must have a high sensitivity of at least 20k/V otherwise their resistance on DC voltage ranges may be too low to avoid upsetting the circuit under test and giving an incorrect reading. To obtain valid readings the meter resistance should be at least 10 times the circuit resistance (take this to be the highest resistor value near where the meter is connected). You can increase the meter resistance by selecting a higher voltage range, but this may give a reading which is too small to read accurately!

On any DC voltage range:

Analogue Meter Resistance = Sensitivity × Max. reading of range

e.g. a meter with 20k/V sensitivity on its 10V range has a resistance of 20k/V × 10V = 200k.

By contrast, digital multimeters have a constant resistance of at least 1M ohms (often 10M ohms) on all their DC voltage ranges. This is more than enough for almost all circuits.

RESISTORS

A **Resistor** is a two-terminal electronic component designed to oppose an electric current by producing a voltage drop between its terminals in proportion to the current, that is, in accordance with Ohm's law:

$$V = IR$$

Resistors are used as part of electrical networks and electronic circuits. They are extremely commonplace in most electronic equipment. Practical resistors can be made of various compounds and films, as well as resistance wire (wire made of a high-resistivity alloy, such as nickel/chrome).

The primary characteristics of resistors are their resistance and the power they can dissipate. Other characteristics include temperature coefficient, noise, and inductance. Less well-known is critical resistance, the value below which power dissipation limits the maximum permitted current flow, and above which the limit is applied voltage. Critical resistance depends upon the materials constituting the resistor as well as its physical dimensions; it's determined by design.

Resistors can be integrated into hybrid and printed circuits, as well as integrated circuits. Size, and position of leads (or terminals) are relevant to equipment designers; resistors must be physically large enough not to overheat when dissipating their power.

TYPICAL RESISTOR	TYPE	SYMBOL
A 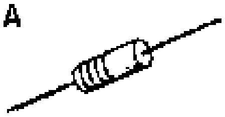	FIXED CARBON	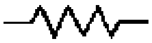
B 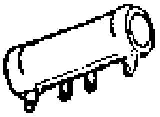	FIXED WIREWOUND (TAPPED)	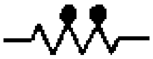
C 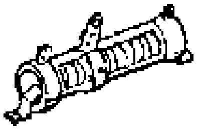	ADJUSTABLE WIREWOUND	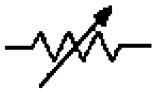
D 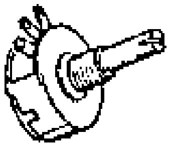	POTENTIOMETER	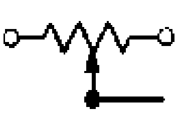
E 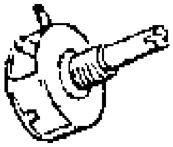	RHEOSTAT	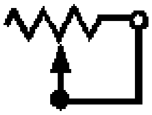

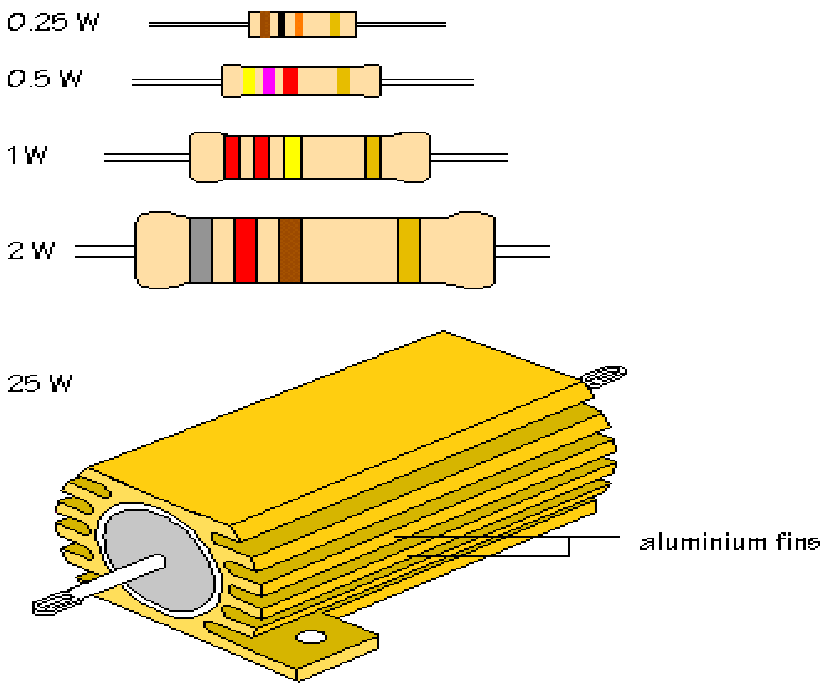

Units

The **ohm** (symbol: Ω) is a SI-driven unit of electrical resistance, named after Georg Simon Ohm. Commonly used multiples and submultiples in electrical and electronic usage are the milliohm, kilohm, and megohm.

Examples	
R47	0.47 ohms
4R7	4.7 ohms
470R	470 ohms
4K7	4.7K ohms
47K	47K ohms
47K3	47.3K ohms

470K	470K ohms
4M7	4.7M ohms

The unit of Resistor is **“OHMS”.**

Sub Unit:

nΩ = Nano OHMS = 10^{-9} = 1/1000000000
µΩ = Micro OHMS = 10^{-6} = 1-1000000
mΩ = Milli OHMS = 10^{-3} = 1/1000
A = OHMS = 1
KΩ = Kilo OHMS = 10^{3} = 1000
MΩ = Mega OHMS = 10^{6} = 1000000
GΩ = Giga OHMS = 10^{9} = 1000000000

Resistor Conversion:

1000 nΩ = 1 µΩ

1000 µ Ω = 1 mΩ

1000 mΩ = 1 Ω

1000 Ω = 1 KΩ

1000 KΩ = 1 MΩ

1000 MΩ = 1 GΩ

Carbon film

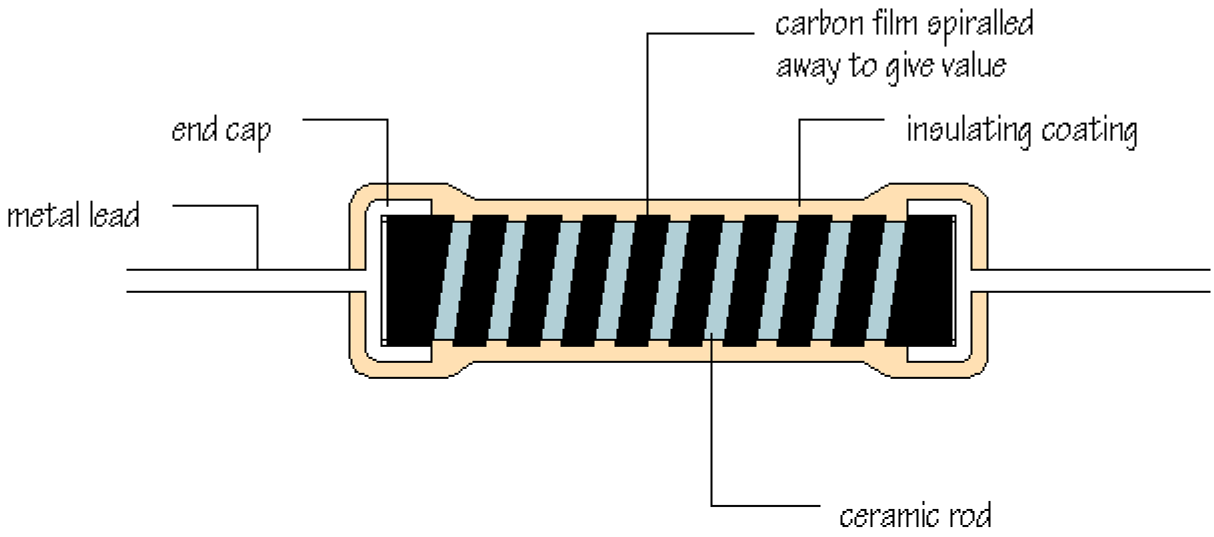
carbon film spiralled away to give value
end cap
insulating coating
metal lead
ceramic rod

A carbon film is deposited on an insulating substrate, and a helix cut in it to create a long, narrow resistive path. Varying shapes, coupled with the resistivity of carbon, (ranging from 9 to 40 µΩ-cm) can provide a variety of resistances.[1] Carbon film resistors feature a power rating range of 1/6 W to 5 W at 70°C. Resistances available range from 1 ohm to 10 megohm. The carbon film resistor can operate between temperatures of -55°C to 155°C. It has 200 to 600 volts maximum working voltage range

Metal film

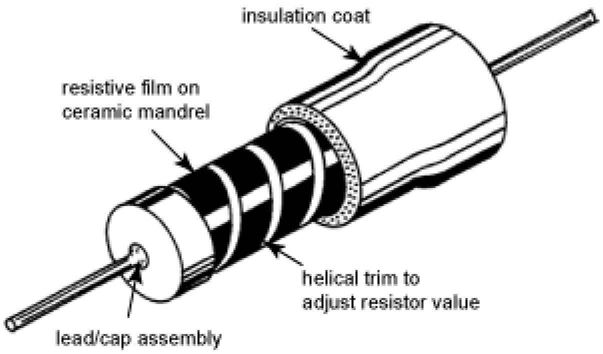
insulation coat
resistive film on ceramic mandrel
helical trim to adjust resistor value
lead/cap assembly

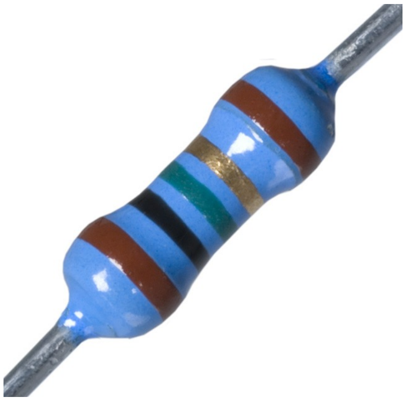

A common type of axial resistor today is referred to as a metal-film resistor. MELF (Metal Electrode Leadless Face) resistors often use the same technology, but are a cylindrically shaped resistor designed for surface mounting. [Note that other types of resistors, e.g. carbon composition, are also available in "MELF" packages].

Metal film resistors are usually coated with nickel chromium (NiCr), but might be coated with any of the cermet materials listed above for thin film resistors. Unlike thin film resistors, the material may be applied using different techniques than sputtering (though that is one such technique). Also, unlike thin-film resistors, the resistance value is determined by cutting a helix through the coating rather than by etching. [This is similar to the way carbon resistors are made.] The result is a reasonable tolerance (0.5, 1, or 2%) and a temperature coefficient of (usually) 25 or 50 ppm/K.

Wire wound

Wire wound resistors are commonly made by winding a metal wire around a ceramic, plastic, or fiberglass core. The ends of the wire are soldered or welded to two caps, attached to the ends of the core. The assembly is protected with a layer of paint, molded plastic, or an enamel coating baked at high temperature. The wire leads are usually between 0.6 and 0.8 mm in diameter and tinned for ease of soldering. For higher power wire wound resistors, either a ceramic outer case or an aluminum outer case on top of an insulating layer is used. The aluminum-cased types are designed to be attached to a heat sink to dissipate the heat; the rated power is dependent on being used with a suitable heat sink, e.g., a 50 W power rated resistor will overheat at around one fifth of the power dissipation if not used with a heat sink.

Actual lengths will vary

Because wire wound resistors are coils they have more undesirable inductance than other types of resistor, although winding the wire in sections with alternately reversed direction can minimize inductance.

Cermet

A **cer met** is a composite material composed of ceramic (cer) and metallic (met) materials. A cer met is ideally designed to have the optimal properties of both a ceramic, such as high temperature resistance and hardness, and those of a metal, such as the ability to undergo plastic deformation. The metal is used as a binder for an oxide, boride, carbide, or alumina. Generally, the metallic elements used are nickel, molybdenum, and cobalt. Depending on the physical structure of the material, cermets can also be metal matrix composites, but cermets are usually less than 20% metal by volume.

Cermets are used in the manufacture of resistors (especially potentiometers), capacitors, and other electronic components which may experience high temperatures.

Cermets are being used instead of tungsten carbide in saws and other brazed tools due to their superior wear and corrosion properties. TiCN, TiC, TiN and similar can be brazed like tungsten carbide if properly prepared however they require special handling during grinding.

More complex materials, known as Cermet 2 or Cermet II, are being utilized since they give considerably longer life in cutting tools while both brazing and grinding like tungsten carbide.

Some types of cermets are also being considered for use as spacecraft shielding as they resist the high velocity impacts of micrometeoroids and orbital debris much more effectively than more traditional spacecraft materials such as aluminum and other metals.

Measurement

Nearly always, the resistance value is of interest. The value of a resistor can be measured with an ohmmeter, which may be one function of a multimeter. Usually, probes on the ends of test leads connect to the resistor.

Measuring low-value resistors, such as fractional-ohm resistors, with acceptable accuracy requires four-terminal connections. One pair of terminals applies a known, calibrated current to the resistor, while the other pair senses the voltage drop across the resistor. Some laboratory test instruments have spring-loaded pairs of contacts, with neighboring contacts electrically isolated from each other. Better digital multimeters have four terminals on their panels, generally used with special test leads. These comprise four wires in all, and have special test clips with jaws insulated from each other. One jaw provides the measuring current, while the other senses the voltage drop.

Series and parallel resistors

Main article: Series and parallel circuits

Resistors in a parallel configuration each have the same potential difference (voltage). To find their total equivalent resistance (R_{eq}):

$$1/R = 1/R1 + 1/R2 + \ldots.. + 1/Rn$$

The parallel property can be represented in equations by two vertical lines "||" (as in geometry) to simplify equations. For two resistors,

$R = R1 // R2 = R1,R2 / R1 + R2$

The current through resistors in series stays the same, but the voltage across each resistor can be different. The sum of the potential differences (voltage) is equal to the total voltage. To find their total resistance:

$$R = R1 + R2 \ldots..+ Rn$$

A resistor network that is a combination of parallel and series can sometimes be broken up into smaller parts that are either one or the other. For instance,

However, many resistor networks cannot be split up in this way. Consider a cube, each edge of which has been replaced by a resistor. For example, determining the resistance between two opposite vertices requires matrix methods for the general case. However, if all twelve resistors are equal, the corner-to-corner resistance is ⅚ of any one of them.

The practical application to resistors is that a resistance of any non-standard value can be obtained by connecting standard values in series or in parallel.

Fixed and Variable Resistors

There are two kinds of resistors, FIXED and VARIABLE. The fixed resistor will have one value and will never change (other than through temperature, age, etc.). The resistors shown in A and B of figure 1-29are classed as fixed resistors. The tapped resistor illustrated in B has several fixed taps and makes more than one resistance value available. The sliding contact resistor shown in C has an adjustable collar that can be moved to tap off any resistance within the ohmic value range of the resistor.

There are two types of variable resistors, one called a POTENTIOMETER and the other a RHEOSTAT (see views D and E of fig. 1-29.)An example of the potentiometer is the volume control on your radio, and an example of the rheostat is the dimmer control for the dash lights in an automobile. There is a slight difference between them. Rheostats usually have two connections, one fixed and the other moveable. Any variable resistor can properly be called a rheostat. The potentiometer always has three connections, two fixed and one moveable. Generally, the rheostat has a limited range of values and a high current-handling capability. The potentiometer has a wide range of values, but it usually has a limited current-handling capability. Potentiometers are always connected as voltage dividers.

Construction

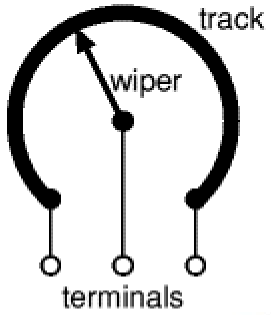

Standard Variable Resistor

Variable resistors consist of a resistance **track** with connections at both ends and a **wiper** which moves along the track as you turn the spindle. The track may be made from carbon, cer met (ceramic and metal mixture) or a coil of wire (for low resistances). The track is usually rotary but straight track versions, usually called sliders, are also available.

Variable resistors may be used as a rheostat with **two** connections (the wiper and just one end of the track) or as a potentiometer with all **three** connections in use. Miniature versions called presets are made for setting up circuits which will not require further adjustment.

Variable resistors are often called **potentiometers** in books and catalogues. They are specified by their maximum resistance, linear or logarithmic track, and their physical size. The standard spindle diameter is 6mm.

The resistance and type of track are marked on the body:

4K7 LIN means 4.7 kΩ linear track.

1M LOG means 1 MΩ logarithmic track.

Some variable resistors are designed to be mounted directly on the circuit board, but most are for mounting through a hole drilled in the case containing the circuit with stranded wire connecting their terminals to the circuit board.

Linear (LIN) and Logarithmic (LOG) tracks

Linear (LIN) track means that the resistance changes at a constant rate as you move the wiper. This is the standard arrangement and you should assume this type is required if a project does not specify the type of track. Presets always have linear tracks.

Logarithmic (LOG) track means that the resistance changes slowly at one end of the track and rapidly at the other end, so halfway along the track is **not** half the total resistance! This arrangement is used for volume (loudness) controls because the human ear has a logarithmic response to loudness so fine control (slow change) is required at low volumes and coarser control (rapid change) at high volumes. It is important to connect the ends of the track the correct way round, if you find that turning the spindle increases the volume rapidly followed by little further change you should swap the connections to the ends of the track.

Rheostat

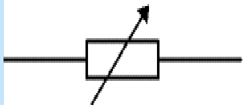

This is the simplest way of using a variable resistor. **Two terminals** are used: one connected to an end of the track, the other to the moveable wiper. Turning the spindle changes the resistance between the two terminals from zero up to the maximum resistance.

Rheostats are often used to **vary current**, for example to control the brightness of a lamp or the rate at which a capacitor charges.

If the rheostat is mounted on a printed circuit board you may find that all three terminals are connected! However, one of them will be linked to the wiper terminal. This improves the mechanical strength of the mounting but it serves no function electrically.

Potentiometer

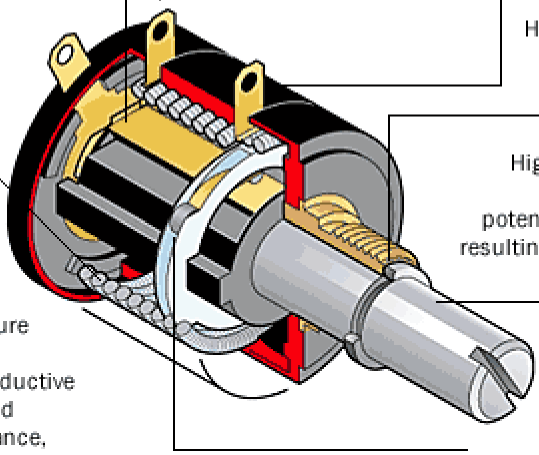

GOLD PLATED TERMINALS AND CONTINUITY BAR:
Do not corrode or tarnish.

HOUSING:
High temp. plastic
Durable in harsh environments.

BRASS BUSHINGS:
Hign quality brass bushings. Provde better support for potentiometer shaft side loads, resulting in long life expectancies

ELEMENT:
Wirewound (shown): Most commonly used in multi-turns. Offers better stability and linearity. Low temperature coefficient.
Hybrid: Made with conductive plastic over a wirewound element. Lower inductance, better resolution, and longer life

STAINLESS STEEL SHAFT:
Non-corrosive. Many modifications available for ease of linking to your system

PRECIOUS METAL WIPER:
Platinum alloy ensures long life and low noise.

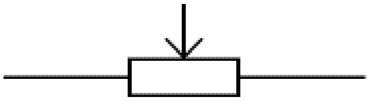

Potentiometer Symbol

Variable resistors used as potentiometers have all **three terminals** connected.

This arrangement is normally used to **vary voltage**, for example to set the switching point of a circuit with a sensor, or control the volume (loudness) in an amplifier circuit. If the terminals at the ends of the track are connected across the power supply

then the wiper terminal will provide a voltage which can be varied from zero up to the maximum of the supply.

Presets

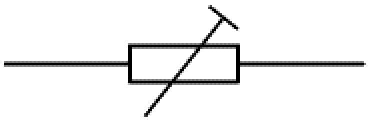

Preset Symbol

These are miniature versions of the standard variable resistor. They are designed to be mounted directly onto the circuit board and adjusted only when the circuit is built. For example to set the frequency of an alarm tone or the sensitivity of a light-sensitive circuit. A small screwdriver or similar tool is required to adjust presets.

Presets are much cheaper than standard variable resistors so they are sometimes used in projects where a standard variable resistor would normally be used.

Multiturn presets are used where very precise adjustments must be made. The screw must be turned many times (10+) to move the slider from one end of the track to the other, giving very fine control.

Preset
(open style)

Presets
(closed style)

Multiturn preset

Four-band resistors

Main article: Electronic color code

Four-band identification is the most commonly used color-coding scheme on all resistors. It consists of four colored bands that are painted around the body of the resistor. The first two bands encode the first two significant digits of the resistance value, the third is a power-of-ten multiplier or number-of-zeroes, and the fourth is the tolerance accuracy, or acceptable error, of the value. Sometimes a fifth band identifies the thermal coefficient, but this must be distinguished from the true 5-color system, with 3 significant digits.

For example, green-blue-yellow-red is $56\times10^4\ \Omega = 560\ \mathrm{k}\Omega \pm 2\%$. An easier description can be as followed: the first band, green, has a value of 5 and the second band, blue, has a value of 6, and is counted as 56. The third band, yellow, has a value of 10^4, which adds four 0's to the end, creating 560,000Ω at ±2% tolerance accuracy. 560,000Ω changes to 560 kΩ ±2% (as a kilo- is 10^3).

Each color corresponds to a certain digit, progressing from darker to lighter colors, as shown in the chart below.

5 band axial resistors

5-band identification is used for higher precision (lower tolerance) resistors (1%, 0.5%, 0.25%, 0.1%), to specify a third significant digit. The first three bands represent the significant digits, the fourth is the multiplier, and the fifth is the tolerance. Five-band resistors with a gold or silver 4th band are sometimes encountered, generally on older or specialized resistors. The 4th band is the tolerance and the 5th the temperature coefficient.

IV Band Colour Coding of Resistor

V Band Colour Coding of Resistor

Color	1st band	2nd band	3rd band	4th band (multiplier)	5th band (tolerance)	Temp. Coefficient
Black	_	0	0	_	_	
Brown	1	1	1	0	±1% (F)	100 ppm
Red	2	2	2	00	±2% (G)	50 ppm
Orange	3	3	3	000	_	15 ppm
Yellow	4	4	4	0000	_	25 ppm
Green	5	5	5	00000	_	
Blue	6	6	6	000000	_	
Violet	7	7	7	0000000	_	
Gray	8	8	8	00000000	_	
White	9	9	9	000000000	_	
Gold	_	_	_	×1/10	±5% (J)	
Silver	_	_	_	×1/100	±10% (K)	
No Band	_	_	_		±20% (M)	

Thermistor

A **thermistor** is a type of resistor whose resistance varies with temperature. The word is a portmanteau of *thermal* and *resistor*. Thermistors are widely used as inrush current limiters, temperature sensors, self-resetting over current protectors, and self-regulating heating elements.

Thermistors differ from resistance temperature detectors (RTD) in that the material used in a thermistor is generally a ceramic or polymer, while RTDs use pure metals. The temperature response is also different; RTDs are useful over larger temperature ranges, while thermistors typically achieve a higher precision within a limited temperature range.

PTC thermistors

MZ11A 75B
MZ11A 75A
MZ11A 75S
MZ11A 75BHV
shiheng.en.alibaba.com

PTC thermistors for over current protection

Ceramic PTC thermistors are used instead of conventional fuses to protect loads such as motors, transformers, etc. or electronic circuits against over current. They not only respond to inadmissibly high currents but also if a preset temperature limit is exceeded. Thermistor fuses limit the power dissipation of the overall circuit by increasing their resistance and thus reducing the current to a harmless residual value. In contrast to conventional fuses, they do not have to be replaced after elimination of the fault but resume their protective function immediately after a short cooling-downtime.

As opposed to PTC thermistors made of plastic materials, ceramic PTC thermistors always return to their initial resistance value, even after frequent heating/cooling cycles.

PTC thermistors as limit temperature sensors

With PTC thermistors as temperature sensors only the steep region of the *R/T* characteristic is Used. The resistance of the PTC thermistor is to be regarded as a function of the ambient temperature [*R*PTC = f (*T*A)].

The precondition for this relationship between resistance and ambient temperature is that self-heating and/or the varistor effect are excluded. This means that these PTC thermistors must be operated in the lowest possible field strengths. To enable a fast response, thermistor sensors have especially small dimensions. High control accuracy is achieved by using materials with an extra steep resistance/temperature characteristic.

Limit temperature measurement for industrial applications

A PTC thermistor can be used effectively to detect whether a temperature limit in industrial equipment, liquids, etc. is exceeded. The PTC thermistor is mounted in thermal contact with the medium that has to be protected from overheating. When the rated temperature limit of the medium is reached, the resistance of the PTC thermistor increases abruptly. In connection with a control unit, this signal can be used to automatically switch off the power supply of a load. EPCOS offers leaded disks, metal tags, screw-type metal cases as well as surface-mounted devices.

NTC thermistors

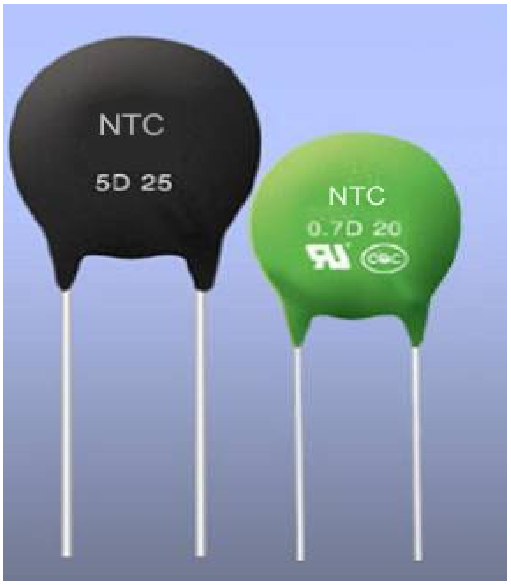

1. Introduction

An NTC thermistor has to be connected in series to the power source circuit to avoid the surge current at the instant when the electronic circuits are turned on. The device can effectively suppress the surge current, and its resistance and power consumption can be greatly reduced after that through the continuous effect of the current so as not to affect the normal work current. Therefore the Power NTC thermistor is the most convenient and efficient instrument to curb the surge current and protect the electronic devices from being damaged.

2. Applications

Applicable to the protection of the power circuits of conversion power supply, switching power supply, UPS power supply, electric heaters, electronic energy-saving lamps, electronic ballasts and other electronic devices, and the filament protection of color picture tubes, incandescent lamps and other lights.

3. Characteristics:

¡¤Small size, Strong power and strong capability of surge current protection.
¡¤Characteristics Fast response to the rapidly surge.
¡¤Big material constant (B value), Small remain resistance.
¡¤Longevity of service, High reliability.
¡¤Integral series, Extensive operating range.

Varistor

Avaristor is an electronic component with a significant non-ohmic current–voltage characteristic. The name is a portmanteau of *variable resistor*. Varistors are often used to protect circuits against excessive transient voltages by incorporating them into the circuit in such a way that, when triggered, they will shunt the current created by the high voltage away from the sensitive components. A varistor is also known as *Voltage Dependent Resistor* or **VDR**. A varistor’s function is to conduct significantly increased current when voltage is excessive.

*Note: only non-ohmic variable resistors are usually called varistors. Other, ohmic types of variable resistor include the potentiometer and the rheostat.

Light Dependent Resistor

A **Light Dependent Resistor** (aka LDR, photoconductor, or photocell) is a device which has a resistance which varies according to the amount of light falling on its surface.

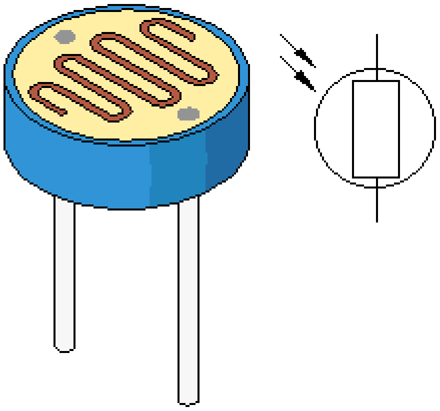

A typical light dependent resistor is pictured above together with (on the right hand side) its circuit diagram symbol. Different LDR's have different specifications, however the **LDR**'s we sell (1) in the REUK Shop (2) are fairly standard and have a resistance in total darkness of 1 MOhm, and a resistance of a couple of kOhm in bright light (*10-20kOhm @ 10 lux, 2-4kOhm @ 100 lux*).

Uses for Light Dependent Resistors

Light dependent resistors are a vital component in any electric circuit (3) which is to be turned on and off automatically according to the level of ambient light - for example, solar powered garden lights, and night security lighting.
An LDR can even be used in a simple remote control circuit using the backlight of a mobile phone to turn on a device - call the mobile from anywhere in the world, it lights up the LDR, and lighting (or a garden sprinkler (4)) can be turned on remotely!

Light Dependent Resistor Circuits

There are two basic circuits using **light dependent resistors** - the first is activated by darkness, the second is activated by light. The two circuits are very similar and just require an **LDR**, some standard resistors (5), a variable resistor (6) (aka potentiometer), and any small signal transistor (7)

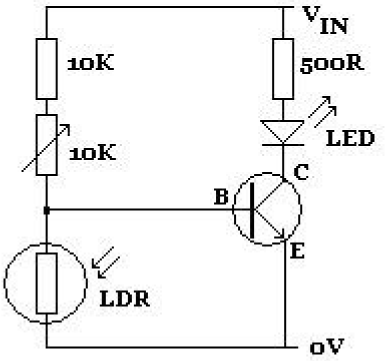

In the circuit diagram above, the **LED (8) lights up** whenever the LDR is in **darkness**. The 10K variable resistor is used to fine-tune the level of darkness required before the LED lights up. The 10K standard resistor can be changed as required to achieve the desired effect, although any replacement must be at least **1K** to protect the transistor from being damaged by excessive current.

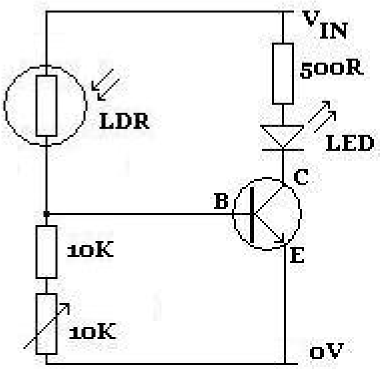

By swapping the LDR over with the 10K and 10K variable resistors (as shown above), the circuit will be activated instead by light. Whenever sufficient light falls on the LDR (manually fine-tuned using the 10K variable resistor), the LED will light up.

Using an LDR in the Real World

The circuits shown above are not practically useful. In a real world circuit, the LED (and resistor) between the positive voltage input (Vin) and the collector (C) of the transistor would be replaced with the device to be powered.

Typically a **relay** is used - particularly when the low voltage light detecting circuit is used to switch on (or off) a 240V mains powered device. A diagram of that part of the circuit is shown above. When darkness falls (if the LDR circuit is configured that way around), the **relay is triggered** and the 240V device - for example a security light –switches on.

NEW For details of a more advanced light/dark sensor circuit click here to read our new article **LM741 Light/Dark Sensor Circuit** We have these circuits available for sale.

Inductor

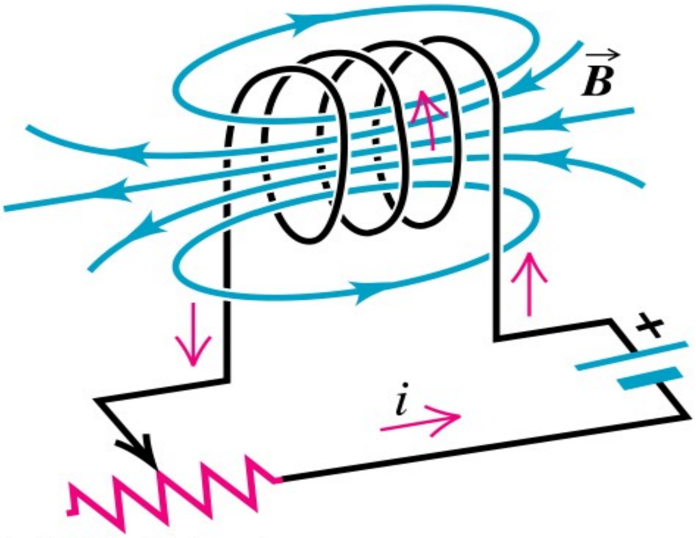

Copyright © Addison Wesley Longman, Inc.

An **inductor** or **reactor** is a passive electrical component that can store energy in a magnetic field created by the electric current passing through it. An inductor's ability to store magnetic energy is measured by its inductance, in units of henries. Typically an inductor is a conducting wire shaped as a coil, the loops help create a strong magnetic field inside the coil due to Faraday's law of induction. Inductors are one of the basic electronic components used in electronics where current and voltage change with time, due to the ability of inductors to delay and reshape alternating currents.

Overview

Inductance (*L*) (measured in henries) is an effect resulting from the magnetic field that forms around a current-carrying conductor that tends to resist changes in the current. Electric current through the conductor creates a magnetic flux proportional to the current. A change in this current creates a change in magnetic flux that, in turn, by Faraday's law generates an electromotive force (EMF) that acts to oppose this change in current. Inductance is a measure of the amount of EMF generated for a unit change in current. For example, an inductor with an inductance of 1 henry produces an EMF of 1 volt when the current through the inductor changes at the rate of 1 ampere per second. The number of loops, the size of each loop, and the material it is wrapped around all affect the inductance. For example, the magnetic flux linking these turns can be increased by coiling the conductor around a material with a high permeability such as iron. This can increase the inductance by 2000 times, although less so at high frequencies.

The unit of Inductor is **“Henry”.**

Sub Unit:

nH = Nano **Henry** = 10^{-9} = 1/1000000000
µH = Micro **Henry** = 10^{-6} = 1-1000000
mH = Milli **Henry** = 10^{-3} = 1/1000
H = **Henry** = 1
KH = Kilo **Henry** = 10^{3} = 1000

Inductor Conversion:

1000 nH = 1 µH

1000 µ H = 1 mH

1000 mH = 1 H

1000 H = 1 KH

1000 KH = 1 MH

Applications

An inductor with two 47mH windings, as may be found in a power supply.

Inductors are used extensively in analog circuits and signal processing. Inductors in conjunction with capacitors and other components form tuned circuits which can emphasize or filter out specific signal frequencies. Applications range from the use of large inductors in power supplies, which in conjunction with filter capacitors remove residual hums known as the Mains hum or other fluctuations from the direct current output, to the small inductance of the ferrite bead or torus installed around a cable to

Prevent radio frequency interference from being transmitted down the wire. Smaller inductor/capacitor combinations provide tuned circuits used in radio reception and broadcasting, for instance.

Two (or more) inductors which have coupled magnetic flux form a transformer, which is a fundamental component of every electric utility power grid. The efficiency of a transformer may decrease as the frequency increases due to eddy currents in the core material and skin effect on the windings. Size of the core can be decreased at higher frequencies and, for this reason, aircraft use 400 hertz alternating current rather than the usual 50 or 60 hertz, allowing a great saving in weight from the use of smaller transformers.

An inductor is used as the energy storage device in some switched-mode power supplies. The inductor is energized for a specific fraction of the regulator's switching frequency, and de-energized for the remainder of the cycle. This energy transfer ratio determines the input-voltage to output-voltage ratio. This X_L is used in complement with an active semiconductor device to maintain very accurate voltage control.

Inductors are also employed in electrical transmission systems, where they are used to depress voltages from lightning strikes and to limit switching currents and fault current. In this field, they are more commonly referred to as reactors.

Larger value inductors may be simulated by use of gyrator circuits.

Inductor construction

An inductor is usually constructed as a coil of conducting material, typically copper wire, wrapped around a core either of air or of ferromagnetic material. Core materials with a higher permeability than air increase the magnetic field and confine it closely to the inductor, thereby increasing the inductance. Low frequency inductors are constructed like transformers, with cores of electrical steel laminated to prevent eddy currents. 'Soft' ferrites are widely used for cores above audio frequencies, since they don't cause the large energy losses at high frequencies that ordinary iron alloys do. This is because of their narrow hysteresis curves, and their high resistivity prevents eddy currents. Inductors come in many shapes. Most are constructed as enamel coated wire wrapped around a ferrite bobbin with wire exposed on the outside, while some enclose the wire completely in ferrite and are called "shielded". Some inductors have an adjustable core, which enables changing of the inductance. Inductors used to block very high frequencies are sometimes made by stringing a ferrite cylinder or bead on a wire.

Small inductors can be etched directly onto a printed circuit board by laying out the trace in a spiral pattern. Some such planar inductors use a planar core.

Small value inductors can also be built on integrated circuits using the same processes that are used to make transistors. Aluminum interconnect is typically used, laid out in a spiral coil pattern. However, the small dimensions limit the inductance, and it is far more common to use a circuit called a "gyrator" which uses a capacitor and active components to behave similarly to an inductor.

Transformer

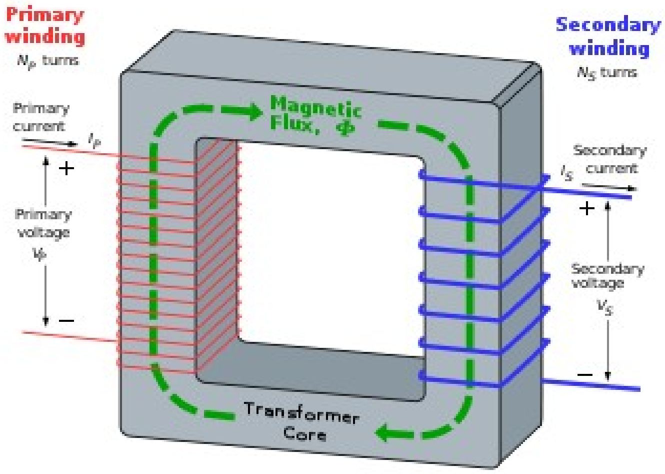
Primary winding
N_p turns
Primary current
I_p
+
Primary voltage
V_p
−
Magnetic Flux, Φ
Transformer Core
Secondary winding
N_S turns
Secondary current
I_S
+
Secondary voltage
V_S
−

0
120V
0
120V
Rapid Electronics
88-0270
6VA
6V 6VA
6V

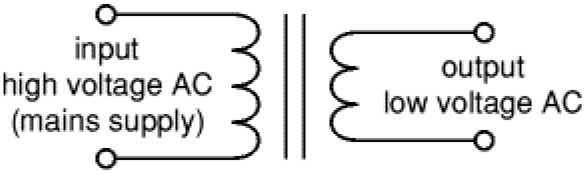
input
high voltage AC
(mains supply)
output
low voltage AC

Transformer

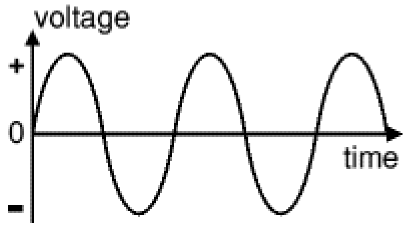

Output: low voltage AC

Transformers convert AC electricity from one voltage to another with little loss of power. Transformers work only with AC and this is one of the reasons why mains electricity is AC.

Step-up transformers increase voltage, step-down transformers reduce voltage. Most power supplies use a step-down transformer to reduce the dangerously high mains voltage (230V in UK) to a safer low voltage.

The input coil is called the **primary** and the output coil is called the **secondary**. There is no electrical connection between the two coils, instead they are linked by an alternating magnetic field created in the soft-iron core of the transformer. The two lines in the middle of the circuit symbol represent the core.

Transformers waste very little power so the power out is (almost) equal to the power in. Note that as voltage is stepped down current is stepped up.

The ratio of the number of turns on each coil, called the **turn's ratio**, determines the ratio of the voltages. A step-down transformer has a large number of turns on its primary (input) coil which is connected to the high voltage mains supply, and a small number of turns on its secondary (output) coil to give a low output voltage.

Transformer types

From Wikipedia, the free encyclopedia

Circuit symbols

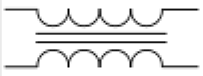	Transformer with two windings and iron core.
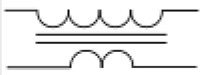	Step-down or step-up transformer. The symbol shows which winding has more turns, but not usually the exact ratio.
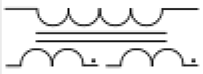	Transformer with three windings. The dots show the relative configuration of the windings.

A variety of types of electrical transformer are made for different purposes. Despite their design differences, the various types employ the same basic principle as discovered in 1831 by Michael Faraday, and share several key functional parts.

Power transformers

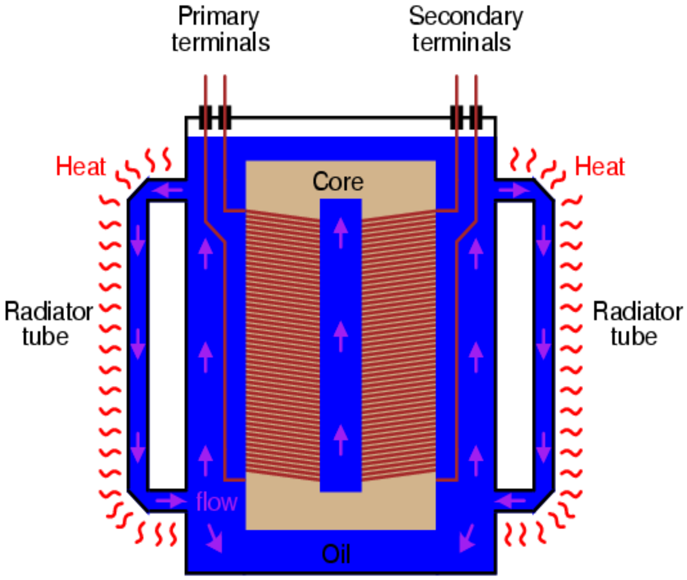

This is the most common type of transformer, widely used in appliances to convert mains voltage to low voltage to power electronics

- Widely available in power ratings ranging from mW to MW
- Insulated laminations minimize eddy current losses
- Small appliance and electronic transformers may use a split bobbin, giving a high level of insulation between the windings
- Rectangular core
- Core laminate stampings are usually in EI shape pairs. Other shape pairs are sometimes used.
- Mumetal shields can be fitted to reduce EMI (electromagnetic interference)
- A screen winding is occasionally used between the 2 power windings
- Small appliance and electronics transformers may have a thermal cut out built in
- Occasionally seen in low profile format for use in restricted spaces
- laminated core made with silicon steel with high permeability

Variac / Autotransformer

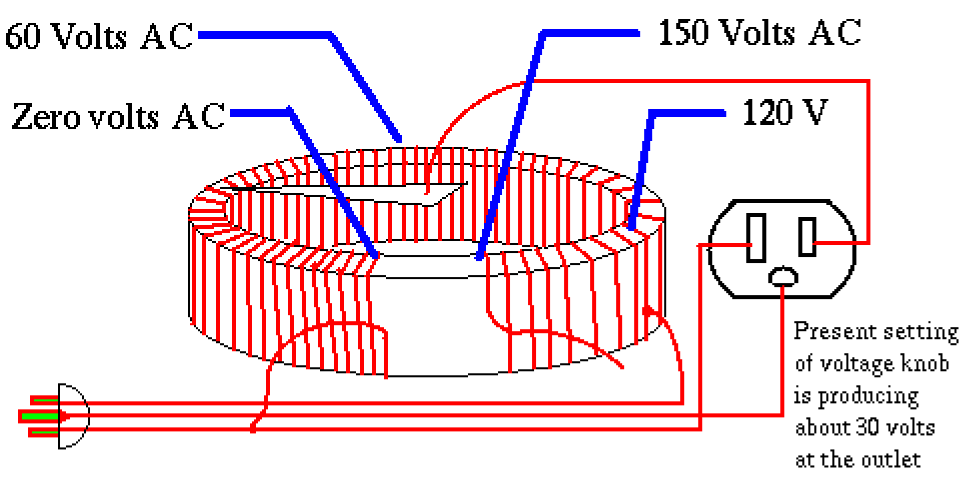
60 Volts AC
Zero volts AC
150 Volts AC
120 V
Present setting of voltage knob is producing about 30 volts at the outlet

An autotransformer has only a single winding, which is tapped at some point along the winding. AC or pulsed voltage is applied across a portion of the winding, and a higher (or lower) voltage is produced across another portion of the same winding. While theoretically separate parts of the winding can be used for input and output, in practice the higher voltage will be connected to the ends of the winding, and the lower voltage from one end to a tap. For example, a transformer with a tap at the center of the winding can be used with 230 volts across the entire winding, and 115 volts between one end and the tap. It can be connected to a 230-volt supply to drive 115-volt equipment, or reversed to drive 230-volt equipment from 115 volts. Since the current in the

Windings is lower, the transformer is smaller, lighter cheaper and more efficient. For voltage ratios not exceeding about 3:1, an autotransformer is cheaper, lighter, smaller and more efficient than an isolating (two-winding) transformer of the same rating.

In practice, transformer losses mean that autotransformers are not perfectly reversible; one designed for stepping down a voltage will deliver slightly less voltage than required if used to step up. The difference is usually slight enough to allow reversal where the actual voltage level is not critical. This is true of isolated winding transformers too.

By exposing part of the winding coils of an autotransformer, and making the secondary connection through a sliding carbon brush, an autotransformer with a near-continuously variable turns ratio can be obtained, allowing for wide voltage adjustment in very small increments.

Constant voltage transformer (Ferro-resonance)

By arranging particular magnetic properties of a transformer core, and installing a resonant tank circuit (a capacitor and an additional winding), a transformer can be arranged to automatically keep the secondary winding voltage constant regardless (within some limits) of any variance in the primary supply without additional circuitry or manual adjustment. CVA transformers run hotter than standard power transformers, for the regulating action is dependent on core saturation, which reduces efficiency somewhat. The output waveform features heavy distortion

Ferrite Core

Ferrite core power transformers are widely used in switched mode power supplies (SMPSUs). The powder core enables high frequency operation, and hence much smaller size to power ratio than laminated iron transformers.

Ferrite transformers are not usable as power transformers at mains frequency

Current transformers

ANNAI 安耐
HTTP://WWW.ANNAICHINA.COM

Current transformers used in metering equipment for three-phase 400 ampere electricity supply

A current transformer (CT) is a measurement device designed to provide a current in its secondary coil proportional to the current flowing in its primary. Current transformers are commonly used in metering and protective relaying in the electrical power industry where they facilitate the safe measurement of large currents, often in the presence of high voltages. The current transformer safely isolates measurement and control circuitry from the high voltages typically present on the circuit being measured.

Current transformers are often constructed by passing a single primary turn (either an insulated cable or an uninsulated bus bar) through a well-insulated toroidal core wrapped with many turns of wire. The CT is typically described by its current ratio from primary to secondary. For example, a 4000:5 CT would provide an output current of 5 amperes when the primary was passing 4000 amperes. The secondary winding can be single ratio or have several tap points to provide a range of ratios. Care must be taken that the secondary winding is not disconnected from its load while current flows in the primary, as this will produce a dangerously high voltage across the open secondary and may permanently affect the accuracy of the transformer.

Specially constructed wideband CTs are also used, usually with an oscilloscope, to measure high frequency waveforms or pulsed currents within pulsed power systems. One type provides a voltage output that is proportional to the measured current; another, called a Rogowski coil, requires an external integrator in order to provide a proportional output.

Voltage transformers

Voltage transformers (VTs) or potential transformers (PTs) are another type of instrument transformer, used for metering and protection in high-voltage circuits. They are designed to present negligible load to the supply being measured and to have a precise voltage ratio to accurately step down high voltages so that metering and protective relay equipment can be operated at a lower potential. Typically the secondary of a voltage transformer is rated for 69 or 120 Volts at rated primary voltage, to match the input ratings of protection relays.

The transformer winding high-voltage connection points are typically labelled as H1, H2 (sometimes H0 if it is internally grounded) and X1, X2, and sometimes an X3 tap may be present. Sometimes a second isolated winding (Y1, Y2, Y3) may also be available on the same voltage transformer. The high side (primary) may be connected phase to ground or phase to phase. The low side (secondary) is usually phase to ground.

The terminal identifications (H1, X1, Y1, etc.) are often referred to as polarity. This applies to current transformers as well. At any instant terminals with the same suffix numeral have the same polarity and phase. Correct identification of terminals and wiring is essential for proper operation of metering and protection relays.

While VTs were formerly used for all voltages greater than 240V primary, modern meters eliminate the need VTs for most secondary service voltages. VTs are typically used in circuits where the system voltage level is above 600 V. Modern meters eliminate the need of VT's since the voltage remains constant and it is measured in the incoming supply.

Pulse transformers

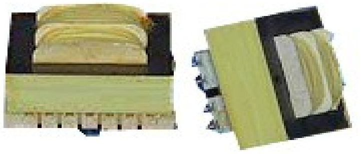

A **pulse transformer** is a transformer that is optimised for transmitting rectangular electrical pulses (that is, pulses with fast rise and fall times and a relatively constant amplitude). Small versions called *signal* types are used in digital logic and telecommunications circuits, often for matching logic drivers to transmission lines. Medium-sized *power* versions are used in power-control circuits such as camera flash controllers. Larger *power* versions are used in the electrical power distribution industry to interface low-voltage control circuitry to the high-voltage gates of power semiconductors. Special high voltage pulse transformers are also used to generate high power pulses for radar, particle accelerators, or other high energy pulsed power applications.

To minimise distortion of the pulse shape, a pulse transformer needs to have low values of leakage inductance and distributed capacitance, and a high open-circuit inductance. In power-type pulse transformers, a low coupling capacitance (between the primary and secondary) is important to protect the circuitry on the primary side from high-powered transients created by the load. For the same reason, high insulation resistance and high breakdown voltage are required. A good transient response is necessary to maintain the rectangular pulse shape at the secondary, because a pulse with slow edges would create switching losses in the power semiconductors.

The product of the peak pulse voltage and the duration of the pulse (or more accurately, the voltage-time integral) is often used to characterise pulse transformers. Generally speaking, the larger this product, the larger and more expensive the transformer.

Pulse transformers by definition have a duty cycle of less than 1, whatever energy stored in the coil during the pulse must be "dumped" out before the pulse is fired again.

Air-core transformers

These are used for high frequency work. The lack of a core means very low inductance. Such transformers may be nothing more than a few turns of wire soldered onto a printed circuit board.

Ferrite-core transformers

Widely used in intermediate frequency (IF) stages in super heterodyne radio receivers. These are mostly tuned transformers, containing a threaded ferrite slug that is screwed in or out to adjust IF tuning. The transformers are usually canned for stability and to reduce interference

Audio transformers

Transformers in a tube amplifier. Output transformers are on the left. The power supply toroidal transformer is on right.

Audio transformers are usually the factor which limits sound quality when used; electronic circuits with wide frequency response and low distortion are relatively simple to design.

Transformers are also used in DI boxes to convert high-impedance instrument signals (e.g. bass guitar) to low impedance signals to enable them to be connected to a microphone input on the mixing console.

A particularly critical component is the output transformer of an audio power amplifier. Valve circuits for quality reproduction have long been produced with no other (inter-stage) audio transformers, but an output transformer is needed to couple the relatively high impedance (up to a few hundred ohms depending upon configuration) of the output valve(s) to the low impedance of a loudspeaker. (The valves can deliver a low current at a high voltage; the speakers require high current at low voltage.) Most solid-state power amplifiers need no output transformer at all.

For good low-frequency response a relatively large iron core is required; high power handling increases the required core size. Good high-frequency response requires carefully designed and implemented windings without excessive leakage inductance or stray capacitance. All this makes for an expensive component.

Early transistor audio power amplifiers often had output transformers, but they were eliminated as designers discovered how to design amplifiers without them.

Capacitors

Function

Capacitors store electric charge. They are used with resistors in timing circuits because it takes time for a capacitor to fill with charge. They are used to smooth varying DC supplies by acting as a reservoir of charge. They are also used in filter circuits because capacitors easily pass AC (changing) signals but they block DC (constant) signals.

Capacitance

This is a measure of a capacitor's ability to store charge. A large capacitance means that more charge can be stored. Capacitance is measured in farads, symbol F. However 1F is very large, so prefixes are used to show the smaller values.

Three prefixes (multipliers) are used, µ (micro), n (nano) and p (pico):

- µ means 10^{-6} (millionth), so 1000000µF = 1F
- n means 10^{-9} (thousand-millionth), so 1000nF = 1µF
- p means 10^{-12} (million-millionth), so 1000pF = 1nF

Capacitor values can be very difficult to find because there are many types of capacitor with different labeling systems!

There are many types of capacitor but they can be split into two groups, **polarised** and **unpolarised**. Each group has its own circuit symbol.

Polarised capacitors (large values, 1µF +)

Examples:

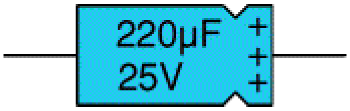

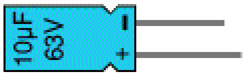

Circuit symbol:

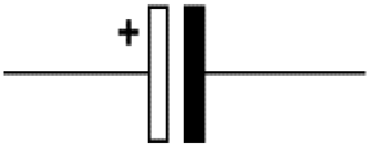

Electrolytic Capacitors

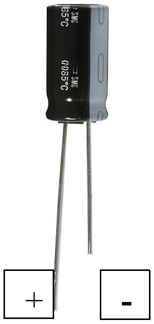

Electrolytic capacitors are polarised and **they must be connected the correct way round**, at least one of their leads will be marked + or -. They are not damaged by heat when soldering.

There are two designs of electrolytic capacitors; **axial** where the leads are attached to each end (220µF in picture) and **radial** where both leads are at the same end (10µF in picture). Radial capacitors tend to be a little smaller and they stand upright on the circuit board.

It is easy to find the value of electrolytic capacitors because they are clearly printed with their capacitance and voltage rating. The voltage rating can be quite low (6V for example) and it should always be checked when selecting an electrolytic capacitor. If the project parts list does not specify a voltage, choose a capacitor with a rating which is greater than the project's power supply voltage. 25V is a sensible minimum most battery circuits.

Tantalum Bead Capacitors

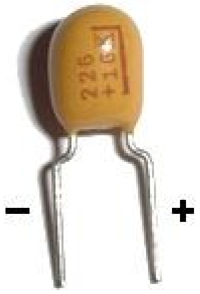

Tantalum bead capacitors are polarised and have low voltage ratings like electrolytic capacitors. They are expensive but very small, so they are used where a large capacitance is needed in a small size.

Modern tantalum bead capacitors are printed with their capacitance, voltage and polarity in full. However older ones use a colour-code system which has two stripes (for the two digits) and a spot of colour for the number of zeros to give the value in µF. The standard colour code is used, but for the spot, **grey** is used to mean × 0.01 and **white** means × 0.1 so that values of less than 10µF can be shown. A third colour stripe near the leads shows the voltage (yellow 6.3V, black 10V, green 16V, blue 20V, grey 25V, white 30V, pink 35V). The positive (+) lead is to the right when the spot is facing you: '**when the spot is in sight, the positive is to the right**'.

For	example:	**blue,**	**grey,**	**black**	**spot**	means	68µF
For	example:	**blue,**	**grey,**	**white**	**spot**	means	6.8µF
For	example:	**blue,**	**grey,**	**grey**	**spot**	means	0.68µF

Unpolarised capacitors (small values, up to 1µF)

Examples:

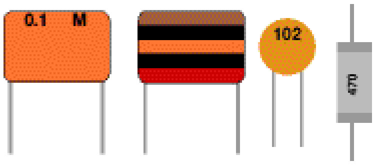

Circuit symbol:

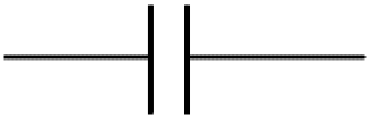

Small value capacitors are unpolarised and may be connected either way round. They are not damaged by heat when soldering, except for one unusual type (polystyrene). They have high voltage ratings of at least 50V, usually 250V or so. It can be difficult to find the values of these small capacitors because there are many types of them and several different labelling systems!

Many small value capacitors have their value printed but without a multiplier, so you need to use experience to work out what the multiplier should be!

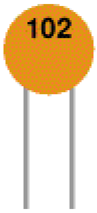

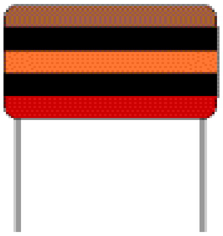

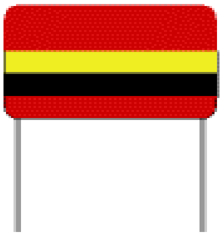

For example **0.1** means 0.1µF = 100nF.

Sometimes the multiplier is used in place of the decimal point:
For example: **4n7** means 4.7nF.

Capacitor Number Code

A number code is often used on small capacitors where printing is difficult:

- the 1st number is the 1st digit,
- the 2nd number is the 2nd digit,
- the 3rd number is the number of zeros to give the capacitance in pF.
- Ignore any letters - they just indicate tolerance and voltage rating.

For example: **102** means 1000pF = 1nF *(not 102pF!)*

For example: **472J** means 4700pF = 4.7nF (J means 5% tolerance).

Capacitor Colour Code

Colour Code	
Colour	Number
Black	0
Brown	1
Red	2
Orange	3
Yellow	4
Green	5
Blue	6
Violet	7
Grey	8
White	9

A colour code was used on polyester capacitors for many years. It is now obsolete, but of course there are many still around. The colours should be read like the resistor code, the top three colour bands giving the value in pF. Ignore the 4th band (tolerance) and 5th band (voltage rating).

For example:

Brown, black, orange means 10000pF = 10nF = 0.01µF.

Note that there are no gaps between the colour bands, so 2 identical bands actually appear as a wide band.

For example:

Wide red, yellow means 220nF = 0.22µF.

Polystyrene Capacitors

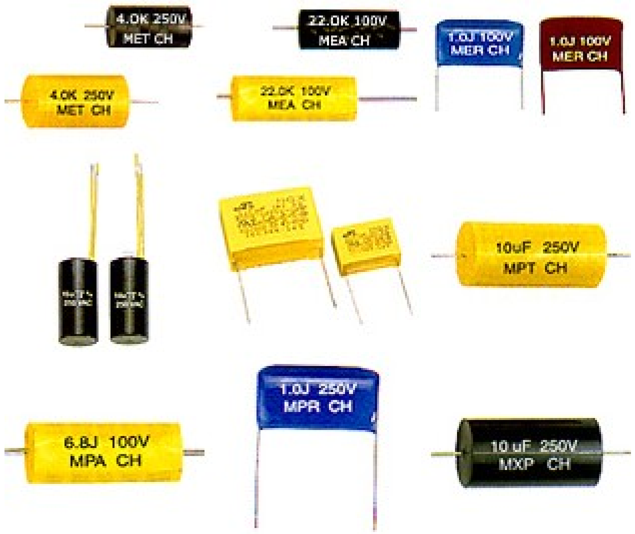
4.0K 250V
MET CH
22.0K 100V
MEA CH
1.0J 100V
MER CH
1.0J 100V
MER CH
4.0K 250V
MET CH
22.0K 100V
MEA CH
10uF 250V
MPT CH
1.0J 250V
MPR CH
6.8J 100V
MPA CH
10 uF 250V
MXP CH

This type is rarely used now. Their value (in pF) is normally printed without units. Polystyrene capacitors can be damaged by heat when soldering (it melts the polystyrene!) so you should use a heat sink (such as a crocodile clip). Clip the heat sink to the lead between the capacitor and the joint.

Real capacitor values (the E3 and E6 series)

You may have noticed that capacitors are not available with every possible value, for example 22µF and 47µF are readily available, but 25µF and 50µF are not!

Why is this? Imagine that you decided to make capacitors every 10µF giving 10, 20, 30, 40, 50 and so on. That seems fine, but what happens when you reach 1000? It would be pointless to make 1000, 1010, 1020, 1030 and so on because for these values 10 is a very small difference, too small to be noticeable in most circuits and capacitors cannot be made with that accuracy.

To produce a sensible range of capacitor values you need to increase the size of the 'step' as the value increases. The standard capacitor values are based on this idea and they form a series which follows the same pattern for every multiple of ten.

The E3 series (3 values for each multiple of ten)
10, 22, 47, ... then it continues 100, 220, 470, 1000, 2200, 4700, 10000 etc.
Notice how the step size increases as the value increases (values roughly double each time).

The E6 series (6 values for each multiple of ten)
10, 15, 22, 33, 47, 68, ... then it continues 100, 150, 220, 330, 470, 680, 1000 etc.
Notice how this is the E3 series with an extra value in the gaps.

The E3 series is the one most frequently used for capacitors because many types cannot be made with very accurate values.

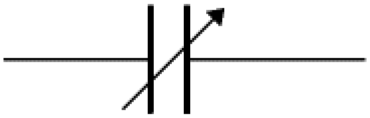
Variable Capacitor Symbol

Variable capacitors

Variable capacitors are mostly used in radio tuning circuits and they are sometimes called 'tuning capacitors'. They have very small capacitance values, typically between 100pF and 500pF (100pF = 0.0001µF). The type illustrated usually has trimmers built in (for making small adjustments - see below) as well as the main variable capacitor.

Variable Capacitor

Many variable capacitors have very short spindles which are not suitable for the standard knobs used for variable resistors and rotary switches. It would be wise to check that a suitable knob is available before ordering a variable capacitor.

Variable capacitors are **not** normally used in timing circuits because their capacitance is too small to be practical and the range of values available is very limited. Instead timing circuits use a fixed capacitor and a variable resistor if it is necessary to vary the time period.

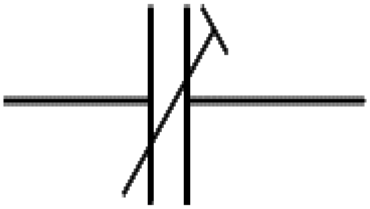
Trimmer Capacitor Symbol

Trimmer capacitors

Trimmer capacitors (trimmers) are miniature variable capacitors. They are designed to be mounted directly onto the circuit board and adjusted only when the circuit is built.

Trimmer Capacitor

A small screwdriver or similar tool is required to adjust trimmers. The process of adjusting them requires patience because the presence of your hand and the tool will slightly change the capacitance of the circuit in the region of the trimmer!

Trimmer capacitors are only available with very small capacitances, normally less than 100pF. It is impossible to reduce their capacitance to zero, so they are usually specified by their minimum and maximum values, for example 2-10pF.

Trimmers are the capacitor equivalent of presets which are miniature variable resistors.

Standard Switches

Type of Switch	Circuit Symbol	Example
ON-OFF **Single Pole, Single Throw = SPST** A simple on-off switch. This type can be used to switch the power supply to a circuit. When used with mains electricity this type of switch *must* be in the live wire, but it is better to use a DPST switch to isolate both live and neutral.	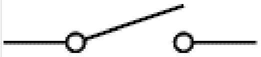	 SPST toggle switch
(ON)-OFF **Push-to-make = SPST Momentary** A push-to-make switch returns to its normally open (off) position when you release the button, this is shown by the brackets around ON. This is the standard doorbell switch.	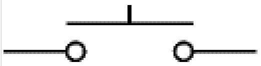	 Push-to-make switch
ON-(OFF) **Push-to-break = SPST Momentary** A push-to-break switch returns to its normally closed (on) position when you release the button.	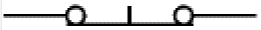	 Push-to-break switch

ON-ON **Single Pole, Double Throw = SPDT** This switch can be on in both positions, switching on a separate device in each case. It is often called a **changeover switch.** For example, a SPDT switch can be used to switch on a red lamp in one position and a green lamp in the other position. A SPDT toggle switch may be used as a simple on-off switch by connecting to COM and one of the A or B terminals shown in the diagram. A and B are interchangeable so switches are usually not labelled. **ON-OFF-ON** **SPDT Centre Off** A special version of the standard SPDT switch. It has a third switching position in the centre which is off. Momentary (ON)-OFF-(ON) versions are also available where the switch returns to the central off position when released.	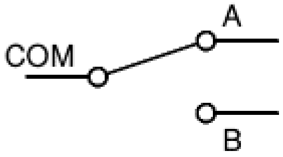 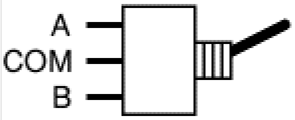	 SPDT toggle switch SPDT slide switch (PCB mounting) SPDT rocker switch
Dual ON-OFF **Double Pole, Single Throw = DPST** A pair of on-off switches which operate together (shown by the dotted line in the circuit symbol). A DPST switch is often used to switch mains electricity because it can isolate both the live and neutral connections.	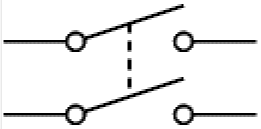	 DPST rocker switch

Special Switches

Type of Switch	Example

Push-Push Switch (e.g. SPST = ON-OFF)

This looks like a momentary action push switch but it is a standard on-off switch: push once to switch on, push again to switch off. This is called a **latching action**.

Micro switch (usually SPDT = ON-ON)

Micro switches are designed to switch fully open or closed in response to small movements. They are available with levers and rollers attached.

Key switch

A key operated switch. The example shown is SPST.

Tilt Switch (SPST)

Tilt switches contain a conductive liquid and when tilted this bridges the contacts inside, closing the switch. They can be used as a sensor to detect the position of an object. Some tilt switches contain mercury which is poisonous.

Reed Switch (usually SPST)

The contacts of a reed switch are closed by bringing a small magnet near the switch. They are used in security circuits, for example to check that doors are closed. Standard reed switches are SPST (simple on-off) but SPDT (changeover) versions are also available.

DIP Switch (DIP = Dual In-line Parallel)

This is a set of miniature SPST on-off switches, the example shown has 8 switches. The package is the same size as a standard DIL (Dual In-Line) integrated circuit.

ON
DIP
1
2
3
4
5
6
7
8

This type of switch is used to set up circuits, e.g. setting the code of a remote control.

Multi-pole Switch

The picture shows a 6-pole double throw switch, also known as a 6-pole changeover switch. It can be set to have momentary or latching action. Latching action means it behaves as a push-push switch, push once for the first position, push again for the second position etc.

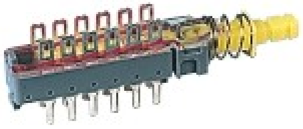

Multi-way Switch

Multi-way switches have 3 or more conducting positions. They may have several poles (contact sets). A popular type has a rotary action and it is available with a range of contact arrangements from 1-pole 12-way to 4-pole 3 way.

The number of ways (switch positions) may be reduced by adjusting a stop under the fixing nut. For example if you need a 2-pole 5-way switch you can buy the 2-pole 6-way version and adjust the stop.

Contrast this multi-way switch (many switch positions) with the multi-pole switch (many contact sets) described above.

Multi-way rotary switch

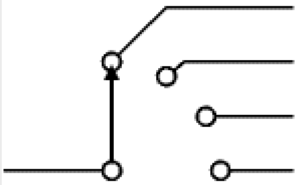

1-pole 4-way switch symbol

Relays

A relay is an electrically operated switch. Current flowing through the coil of the relay creates a magnetic field which attracts a lever and changes the switch contacts. The coil current can be on or off so relays have two switch positions and they are double throw (changeover) switches.

Relays allow one circuit to switch a second circuit which can be completely separate from the first. For example a low voltage battery circuit can use a relay to switch a 230V AC mains circuit. There is no electrical connection inside the relay between the two circuits, the link is magnetic and mechanical.

The coil of a relay passes a relatively large current, typically 30mA for a 12V relay, but it can be as much as 100mA for relays designed to operate from lower voltages. Most ICs (chips) cannot provide this current and a transistor is usually used to amplify the small IC current to the larger value required for the relay coil. The maximum output current for the popular 555 timer IC is 200mA so these devices can supply relay coils directly without amplification.

Relays are usuallly SPDT or DPDT but they can have many more sets of switch contacts, for example relays with 4 sets of changeover contacts are readily available. For further information about switch contacts and the terms used to describe them please see the page on switches.

Most relays are designed for PCB mounting but you can solder wires directly to the pins providing you take care to avoid melting the plastic case of the relay.

The supplier's catalogue should show you the relay's connections. The coil will be obvious and it may be connected either way round. Relay coils produce brief high voltage 'spikes' when they are switched off and this can destroy transistors and ICs in the circuit. To prevent damage you must connect a protection diode across the relay coil.

The animated picture shows a working relay with its coil and switch contacts. You can see a lever on the left being attracted by magnetism when the coil is switched on. This lever moves the switch contacts. There is one set of contacts (SPDT) in the foreground and another behind them, making the relay DPDT.

The relay's switch connections are usually labelled COM, NC and NO:

COM = Common, always connect to this, it is the moving part of the switch.

NC = Normally Closed, COM is connected to this when the relay coil is off.

NO = Normally Open, COM is connected to this when the relay coil is on.

Connect to COM and NO if you want the switched circuit to be on when the relay coil is on.

Connect to COM and NC if you want the switched circuit to be on when the relay coil is off.

Choosing a relay

You need to consider several features when choosing a relay:

Physical size and pin arrangement

If you are choosing a relay for an existing PCB you will need to ensure that its dimensions and pin arrangement are suitable. You should find this information in the supplier's catalogue.

Coil voltage

The relay's coil voltage rating and resistance must suit the circuit powering the relay coil. Many relays have a coil rated for a 12V supply but 5V and 24V relays are also readily available. Some relays operate perfectly well with a supply voltage which is a little lower than their rated value.

Coil resistance

The circuit must be able to supply the current required by the relay coil. You can use Ohm's law to calculate the current: Relay coil current = supply voltage

Coil resistance

For example: A 12V supply relay with a coil resistance of 400 passes a current of 30mA. This is OK for a 555 timer IC (maximum output current 200mA), but it is too much for most ICs and they will require a transistor to amplify the current.

Switch ratings (voltage and current)

The relay's switch contacts must be suitable for the circuit they are to control. You will need to check the voltage and current ratings. Note that the voltage rating is usually higher for AC, for example: "5A at 24V DC or 125V AC".

Switch contact arrangement (SPDT, DPDT etc)

Most relays are SPDT or DPDT which are often described as "single pole changeover" (SPCO) or "double pole changeover" (DPCO). For further information please see the page on switches.

Protection diodes for relays

Transistors and ICs must be protected from the brief high voltage produced when a relay coil is switched off. The diagram shows how a signal diode (eg 1N4148) is connected 'backwards' across the relay coil to provide this protection.

Current flowing through a relay coil creates a magnetic field which collapses suddenly when the current is switched off. The sudden collapse of the magnetic field induces a brief high voltage across the relay coil which is very likely to damage transistors and ICs. The protection diode allows the induced voltage to drive a brief current through the coil (and diode) so the magnetic field dies away quickly rather than instantly. This prevents the induced voltage becoming high enough to cause damage to transistors and ICs.

Reed relays

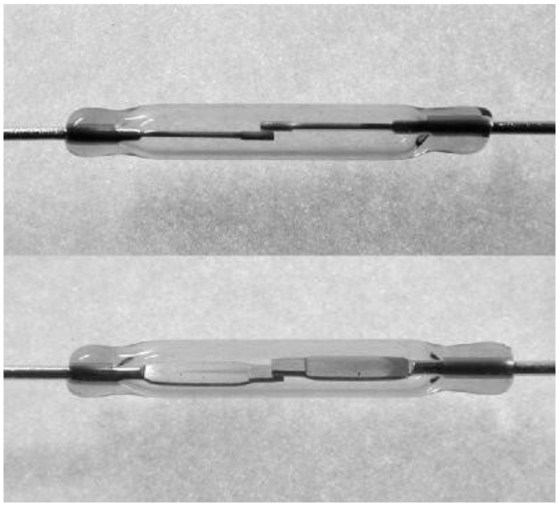

_

Reed relays consist of a coil surrounding a reed switch. Reed switches are normally operated with a magnet, but in a reed relay current flows through the coil to create a magnetic field and close the reed switch.

Reed relays generally have higher coil resistances than standard relays (1000 for example) and a wide range of supply voltages (9-20V for example). They are capable of switching much more rapidly than standard relays, up to several hundred times per second; but they can only switch low currents (500mA maximum for example).

The reed relay shown in the photograph will plug into a standard 14-pin DIL socket ('IC holder').

For further information about reed switches please see the page on switche

---

_

Diode

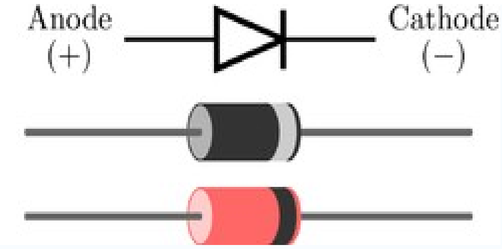

Figure 1: Closeup of a diode, showing the square shaped semiconductor crystal

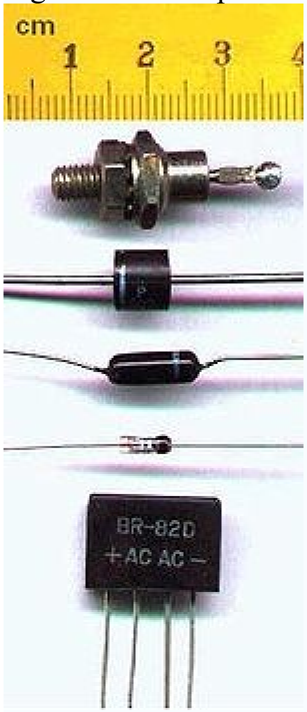

In electronics, a **diode** is a two-terminal device (thermionic diodes may also have one or two ancillary terminals for a heater).

Diodes have two active electrodes between which the signal of interest may flow, and most are used for their unidirectional electric current property. The varicap diode is used as an electrically adjustable capacitor.

The directionality of current flow most diodes exhibit is sometimes generically called the *rectifying* property. The most common function of a diode is to allow an electric current to pass in one direction (called the *forward biased* condition) and to block the current in the opposite direction (the reverse biased condition). Thus, the diode can be thought of as an electronic version of a check valve.

Real diodes do not display such a perfect on-off directionality but have a more complex non-linear electrical characteristic, which depends on the particular type of diode technology. Diodes also have many other functions in which they are not designed to operate in this on-off manner.

Early diodes included “cat’s whisker” crystals and vacuum tube devices (also called **thermionic valves**). Today the most common diodes are made from semiconductor materials such as silicon or germanium

Semiconductor diodes

Most modern diodes are based on semiconductor p-n junctions. In a p-n diode, conventional current can flow from the p-type side (the anode) to the n-type side (the cathode), but cannot flow in the opposite direction. Another type of semiconductor diode, the Schottky diode, is formed from the contact between a metal and a semiconductor rather than by a p-n junction

Types of semiconductor diode

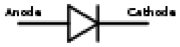

Diode

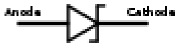

Zener diode

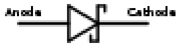

Schottky diode

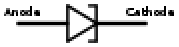

Tunnel diode

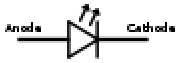

Light-emitting diode

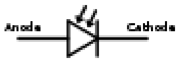

Photodiode

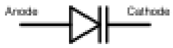

Varector

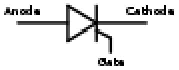

Silicon controlled rectifier

Figure: Typical diode packages in same alignment as diode symbol. Thin bar depicts the cathode.

There are several types of junction diodes, which either emphasize a different physical aspect of a diode often by geometric scaling, doping level, choosing the right electrodes, are just an application of a diode in a special circuit, or are really different devices like the Gunn and laser diode and the MOSFET:

Normal (p-n) diodes, which operate as described above, are usually made of doped silicon or, more rarely, germanium. Before the development of modern silicon power rectifier diodes, cuprous oxide and later selenium was used; its low efficiency gave it a much higher forward voltage drop (typically 1.4–1.7 V per “cell”, with multiple cells stacked to increase the peak inverse voltage rating in high voltage rectifiers), and required a large heat sink (often an extension of the diode’s metal substrate), much larger than a silicon diode of the same current ratings would require. The vast majority of all diodes are the p-n diodes found in CMOS integrated circuits, which include two diodes per pin and many other internal diodes.

Zener diode

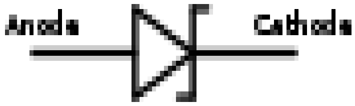

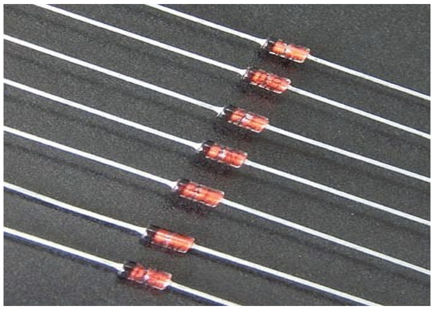

Zener diode schematic symbol

Current-voltage characteristic of a Zener diode with a breakdown voltage of 17 volt. Notice the change of voltage scale between the forward biased (positive) direction and the reverse biased (negative) direction.

A **Zener diode** is a type of diode that permits current in the forward direction like a normal diode, but also in the reverse direction if the voltage is larger than the breakdown voltage known as "Zener knee voltage" or "Zener voltage". The device was named after Clarence Zener, who discovered this electrical property.

A conventional solid-state diode will not allow significant current if it is reverse-biased below its reverse breakdown voltage. When the reverse bias breakdown voltage is exceeded, a conventional diode is subject to high current due to avalanche breakdown. Unless this current is limited by external circuitry, the diode will be permanently damaged. In case of large forward bias (current in the direction of the arrow), the diode exhibits a voltage drop due to its junction built-in voltage and internal resistance. The amount of the voltage drop depends on the semiconductor material and the doping concentrations.

A **Zener diode** exhibits almost the same properties, except the device is specially designed so as to have a greatly reduced breakdown voltage, the so-called **Zener voltage**. A Zener diode contains a heavily doped p-n junction allowing electrons to tunnel from the valence band of the p-type material to the conduction band of the n-type material. In the atomic scale, this tunneling corresponds to the transport of valence band electrons into the empty conduction band states; as a result of the reduced barrier between these bands and high electric fields that are induced due to the relatively high levels of dopings on both sides. A reverse-biased Zener diode will exhibit a controlled breakdown and allow the current to keep the voltage across the Zener diode at the Zener voltage. For example, a diode with a Zener breakdown voltage of 3.2 V will exhibit a voltage drop of 3.2 V if reverse bias voltage applied across it is more than its Zener voltage. However, the current is not unlimited, so the Zener diode is typically used to generate a reference voltage for an amplifier stage, or as a voltage stabilizer for low-current applications.

The breakdown voltage can be controlled quite accurately in the doping process. While tolerances within 0.05% are available, the most widely used tolerances are 5% and 10%.

Another mechanism that produces a similar effect is the avalanche effect as in the avalanche diode. The two types of diode are in fact constructed the same way and both effects are present in diodes of this type. In silicon diodes up to about 5.6 volts, the Zener effect is the predominant effect and shows a marked negative temperature coefficient. Above 5.6 volts, the avalanche effect becomes predominant and exhibits a positive temperature coefficient.

In a 5.6 V diode, the two effects occur together and their temperature coefficients neatly cancel each other out, thus the 5.6 V diode is the component of choice in temperature-critical applications.

Modern manufacturing techniques have produced devices with voltages lower than 5.6 V with negligible temperature coefficients, but as higher voltage devices are encountered, the temperature coefficient rises dramatically. A 75 V diode has 10 times the coefficient of a 12 V diode.

All such diodes, regardless of breakdown voltage, are usually marketed under the umbrella term of "Zener diode".

Uses

Zener diode shown with typical packages. *Reverse* current − i_Z is shown.

Zener diodes are widely used to regulate the voltage across a circuit. When connected in parallel with a variable voltage source so that it is reverse biased, a Zener diode conducts when the voltage reaches the diode's reverse breakdown voltage. From that point it keeps the voltage at that value.

Tunnel diode

Tunnel diode schematic symbol

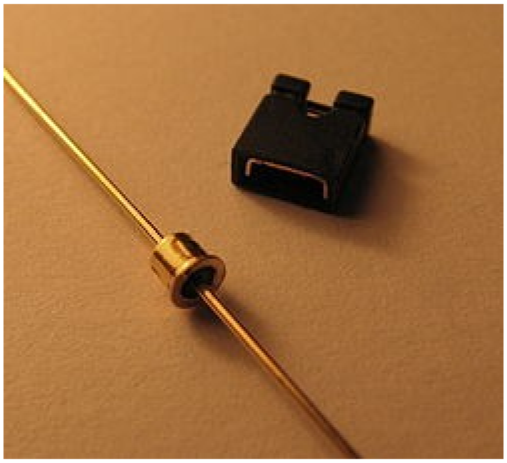

1N3716 tunnel diode (with jumper for scale)

A **tunnel diode** or **Esaki diode** is a type of semiconductor diode which is capable of very fast operation, well into the microwave frequency region, by using quantum mechanical effects.

It was invented in August 1957 by Leo Esaki when he was with Tokyo Tsushin Kogyo (now known as Sony), who in 1973 received the Nobel Prize in Physics for discovering the electron tunneling effect used in these diodes.

These diodes have a heavily doped p-n junction only some 10 nm (100 Å) wide. The heavy doping results in a broken bandgap, where conduction band electron states on the n-side are more or less aligned with valence band hole states on the p-side.

Tunnel diodes were manufactured by SONY for the first time in 1957 [1] followed by General Electric and other companies from about 1960, and are still made in low volume today. [2] Tunnel diodes are usually made from germanium, but can also be made in gallium arsenide and silicon materials. They can be used as oscillators, amplifiers, frequency converters and detectors. [3]

Light-emitting diode

.

Blue, green, and red LEDs; these can be combined to produce most perceptible colors, including white. Infrared and ultraviolet (UVA) LEDs are also available.

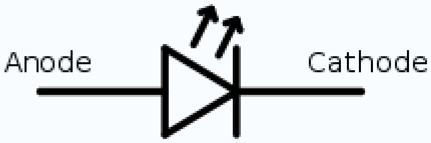

LED schematic symbol

.

A **light-emitting-diode** (**LED**) is a semiconductor diode that emits light when an electric current is applied in the forward direction of the device, as in the simple LED circuit. The effect is a form of electroluminescence where incoherent and narrow-spectrum light is emitted from the p-n junction in a solid state material.

LEDs are widely used as indicator lights on electronic devices and increasingly in higher power applications such as flashlights and area lighting. An LED is usually a small area (less than 1 mm^2) light source, often with optics added directly on top of the chip to shape its radiation pattern and assist in reflection.[2][3] The color of the emitted light depends on the composition and condition of the semi conducting material used, and can be infrared, visible, or ultraviolet. Besides lighting, interesting applications include using UV-LEDs for sterilization of water and disinfection of devices,[4] and as a grow light to enhance photosynthesis in plants.[5]

Photodiode

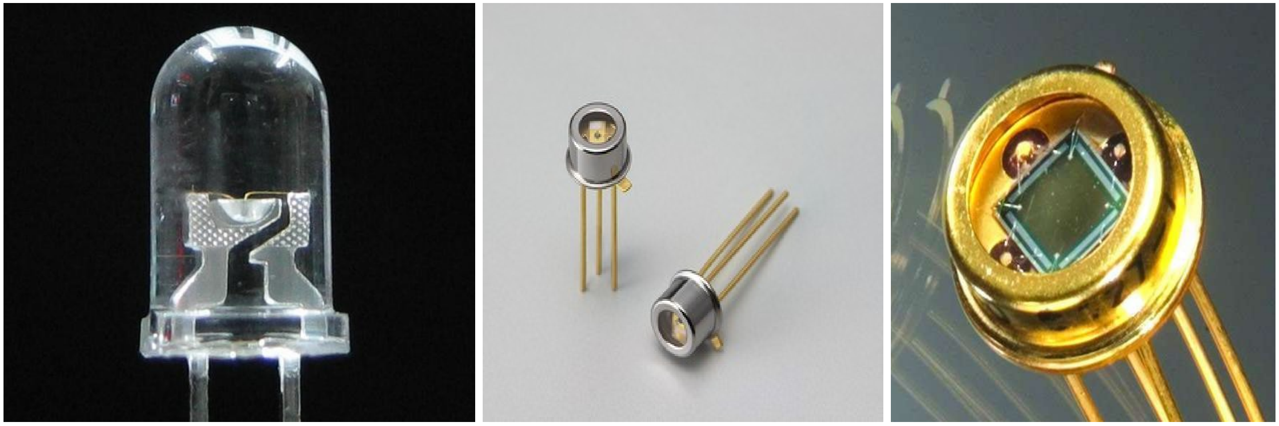

A photodiode

A **photodiode** is a type of photodetector capable of converting light into either current or voltage, depending upon the mode of operation.[1]

Photodiodes are similar to regular semiconductor diodes except that they may be either exposed (to detect vacuum UV or X-rays) or packaged with a window or optical fibre connection to allow light to reach the sensitive part of the device. Many diodes designed for use specifically as a photodiode will also use a PIN junction rather than the typical PN junction.

Varactor Diode

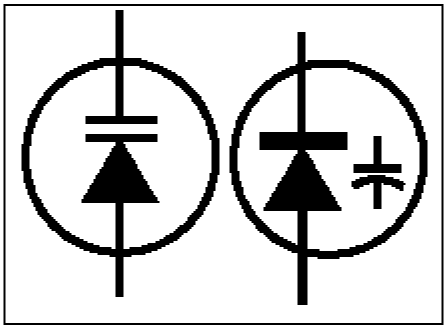

From Wikipedia, the free encyclopedia
(Redirected from Varactor diode)
Jump to: navigation, search

Varicap schematic symbol

In electronics, a **varicap diode**, **varactor diode**, **variable capacitance diode** or **tuning diode** is a type of diode which has a variable capacitance that is a function of the voltage impressed on its terminals._

Applications

Varactors are principally used as a voltage-controlled capacitor, rather than as rectifiers. They are commonly used in parametric amplifiers, parametric oscillators and voltage-controlled oscillators as part of phase-locked loops and frequency synthesizers.

Transformer + Rectifier

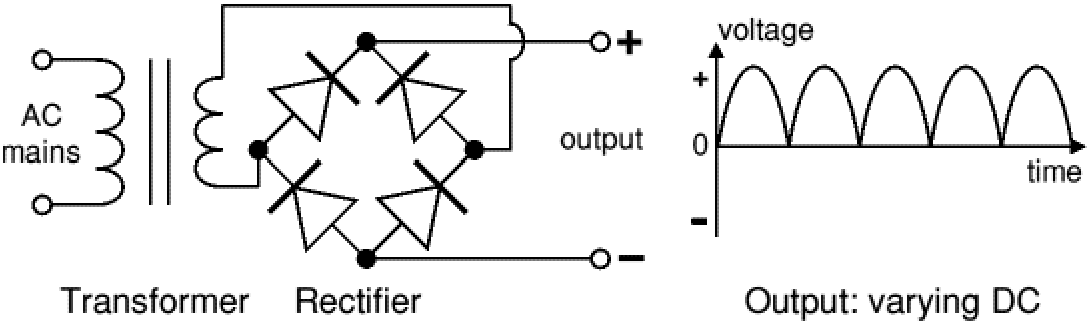
AC
mains
output
+
−
voltage
+
0
-
time
Transformer
Rectifier
Output: varying DC

The **varying DC** output is suitable for lamps, heaters and standard motors. It is **not** suitable for electronic circuits unless they include a smoothing capacitor.

Transformer + Rectifier + Smoothing

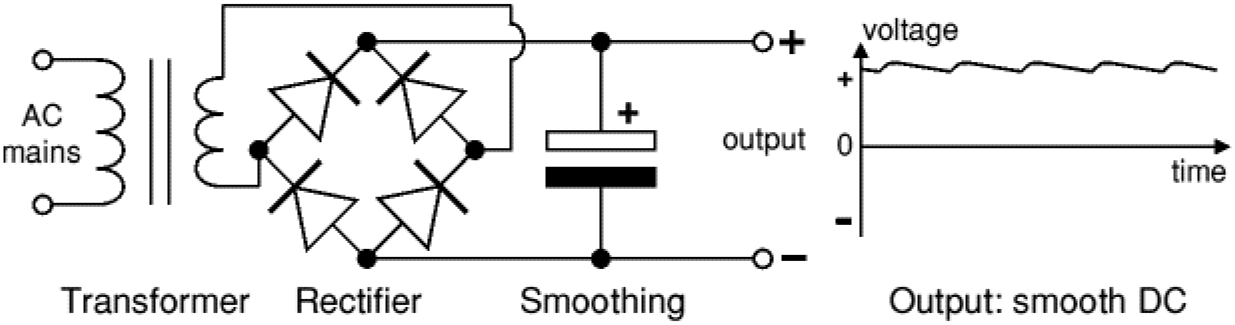
AC
mains
+
output
+
−
voltage
+
0
-
time
Transformer
Rectifier
Smoothing
Output: smooth DC

The **smooth DC** output has a small ripple. It is suitable for most electronic circuits.

Transformer + Rectifier + Smoothing + Regulator

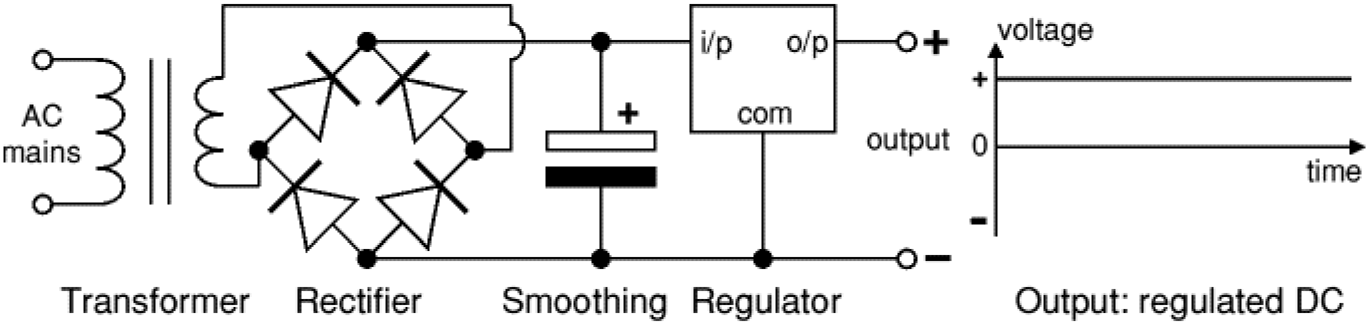
AC
mains
i/p
o/p
com
+
+
-
voltage
+
0
-
output
time
Transformer
Rectifier
Smoothing
Regulator
Output: regulated DC

The **regulated DC** output is very smooth with no ripple. It is suitable for all electronic circuits.

Dual Supplies

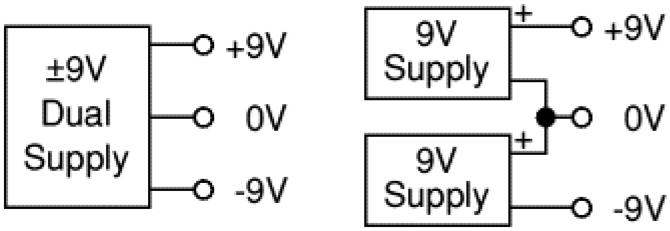
±9V
Dual
Supply
+9V
0V
-9V
9V
Supply
9V
Supply
+
+
+9V
0V
-9V

Some electronic circuits require a power supply with positive and negative outputs as well as zero volts (0V). This is called a 'dual supply' because it is like two ordinary supplies connected together as shown in the diagram.

Dual supplies have three outputs, for example a ±9V supply has +9V, 0V and -9V outputs.

Transistors

Function

Transistors **amplify current**, for example they can be used to amplify the small output current from a logic IC so that it can operate a lamp, relay or other high current device. In many circuits a resistor is used to convert the changing current to a changing voltage, so the transistor is being used to **amplify voltage**.

A transistor may be used as a **switch** (either fully on with maximum current, or fully off with no current) and as an **amplifier** (always partly on).

The amount of current amplification is called the **current gain**, symbol h_{FE}. For further information please see the Transistor Circuits page.

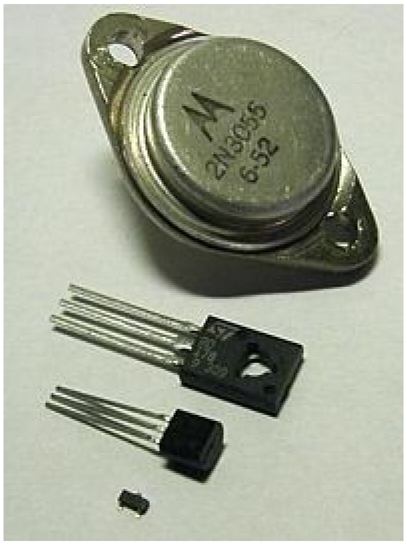

Types of transistor

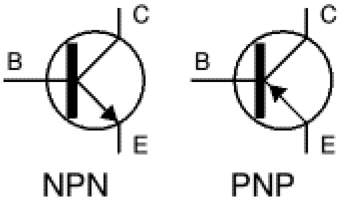

Transistor circuit symbols

There are two types of standard transistors, **NPN** and **PNP**, with different circuit symbols. The letters refer to the layers of semiconductor material used to make the transistor. Most transistors used today are NPN because this is the easiest type to make from silicon. If you are new to electronics it is best to start by learning how to use NPN transistors.

The leads are labelled **base** (B), **collector** (C) and **emitter** (E). These terms refer to the internal operation of a transistor but they are not much help in understanding how a transistor is used, so just treat them as labels!

NPN

The symbol of an NPN Bipolar Junction Transistor.

NPN is one of the two types of bipolar transistors, in which the letters "N" and "P" refer to the majority charge carriers inside the different regions of the transistor. Most bipolar transistors used today are NPN, because electron mobility is higher than hole mobility in semiconductors, allowing greater currents and faster operation.

NPN transistors consist of a layer of P-doped semiconductor (the "base") between two N-doped layers. A small current entering the base in common-emitter mode is amplified in the collector output. In other terms, an NPN transistor is "on" when its base is pulled **high** relative to the emitter.

The arrow in the NPN transistor symbol is on the emitter leg and points in the direction of the conventional current flow when the device is in forward active mode.

One mnemonic device for identifying the symbol for the NPN transistor is "**n**ot **p**ointing i**n**"[5] or "**n**ever **p**oints i**n**".

PNP

The other type of BJT is the PNP with the letters "P" and "N" referring to the majority charge carriers inside the different regions of the transistor.

The symbol of a PNP Bipolar Junction Transistor.

PNP transistors consist of a layer of N-doped semiconductor between two layers of P-doped material. A small current leaving the base in common-emitter mode is amplified in the collector output. In other terms, a PNP transistor is "on" when its base is pulled **low** relative to the emitter.

The arrow in the PNP transistor symbol is on the emitter leg and points in the direction of the conventional current flow when the device is in forward active mode.

One mnemonic device for identifying the symbol for the PNP transistor is "**p**oints i**n** **p**roudly"[5] or "**p**oints i**n** **p**ermanently".

A Darlington pair is two transistors connected together to give a very high current gain.

In addition to standard (bipolar junction) transistors, there are **field-effect transistors** which are usually referred to as **FET**s. They have different circuit symbols and properties and they are not (yet) covered by this page.

Connecting

Transistors have three leads which must be connected the correct way round. Please take care with this because a wrongly connected transistor may be damaged instantly when you switch on.

If you are lucky the orientation of the transistor will be clear from the PCB or stripboard layout diagram, otherwise you will need to refer to a supplier's catalogue to identify the leads.

The drawings on the right show the leads for some of the most common case styles.

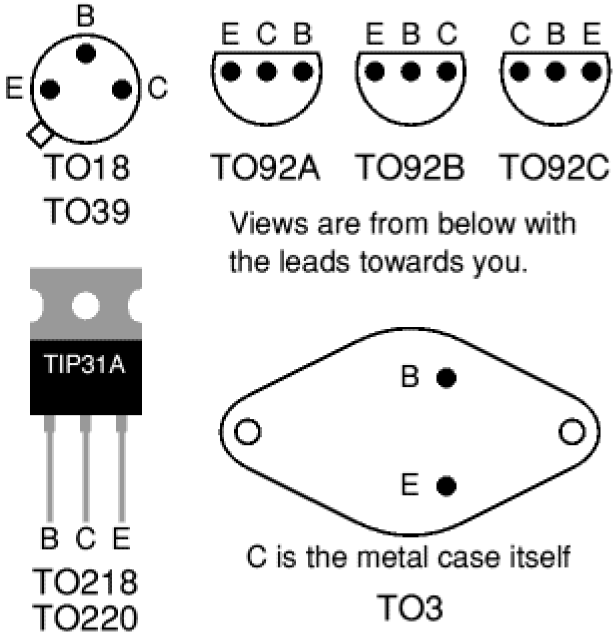
B
E
C
TO18
TO39
E C B
E B C
C B E
TO92A
TO92B
TO92C
Views are from below with the leads towards you.
TIP31A
B C E
TO218
TO220
B
E
C is the metal case itself
TO3

Transistor leads for some common case styles.

Please note that transistor lead diagrams show the view from **below** with the leads towards you. This is the opposite of IC (chip) pin diagrams which show the view from above.

Testing a transistor

Transistors can be damaged by heat when soldering or by misuse in a circuit. If you suspect that a transistor may be damaged there are two easy ways to test it:

1. Testing with a multimeter

Use a multimeter or a simple tester (battery, resistor and LED) to check each pair of leads for conduction. Set a digital multimeter to diode test and an analogue multimeter to a low resistance range.

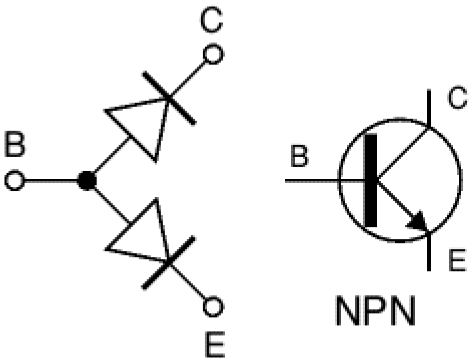

Testing an NPN transistor

Test each pair of leads both ways (six tests in total):

- The **base-emitter (BE)** junction should behave like a diode and conduct one way only.
- The **base-collector (BC)** junction should behave like a diode and conduct one way only.
- The **collector-emitter (CE)** should not conduct either way.

The diagram shows how the junctions behave in an NPN transistor. The diodes are reversed in a PNP transistor but the same test procedure can be used.

Testing in a simple switching circuit

Connect the transistor into the circuit shown on the right which uses the transistor as a switch. The supply voltage is not critical, anything between 5 and 12V is suitable. This circuit can be quickly built on breadboard for example. Take care to include the 10kΩ resistor in the base connection or you will destroy the transistor as you test it!

If the transistor is OK the LED should light when the switch is pressed and not light when the switch is released.

To test a PNP transistor use the same circuit but reverse the LED and the supply voltage.

Some multimeters have a 'transistor test' function which provides a known base current and measures the collector current so as to display the transistor's DC current gain h_{FE}.

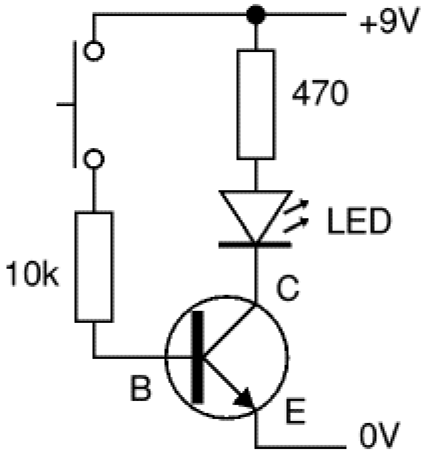

How a transistor works

Amplifier circuit, standard common-emitter configuration.

Simple circuit using a transistor.

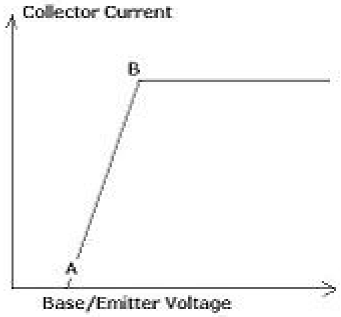

Operation graph of a transistor.

The essential usefulness of a transistor comes from its ability to use a small signal applied between one pair of its terminals to control a much larger signal at another pair of terminals. This property is called "gain". A transistor can control its output in proportion to the input signal; this is called an "amplifier".

Or, the transistor can be used to turn current on or off in a circuit like an electrically controlled "switch", where the amount of current is determined by other circuit elements.

The two types of transistors have slight differences in how they are used in a circuit. A bipolar transistor has terminals labelled base, collector and emitter. A small current at base terminal can control or switch a much larger current between collector and emitter terminals. For a field-effect transistor, the terminals are labelled gate, source, and drain, and a voltage at the gate can control a current between source and drain.

The image to the right represents a typical bipolar transistor in a circuit. Charge will flow between emitter and collector terminals depending on the current in the base. Since internally the base and emitter connections behave like a semiconductor diode, a voltage drop develops between base and emitter while the base current exists. The size of this voltage depends on the material the transistor is made from, and is referred to as V_{be}.

Transistor as a switch

Transistors are commonly used as electronic switches, for both high power applications including switched-mode power supplies and low power applications such as logic gates.

It can be seen from the graph that once the base voltage reaches a certain level, shown at B, no more current will exist and the output will be held at a fixed voltage. The transistor is then said to be saturated. Hence, values of input voltage can be chosen such that the output is either completely off,[7] or completely on. The transistor is acting as a switch, and this type of operation is common in digital circuits where only "on" and "off" values are relevant.

Transistor as an amplifier

The above common emitter amplifier is designed so that a small change in voltage in (V_{in}) changes the small current through the base of the transistor and the transistor's current amplification combined with the properties of the circuit mean that small swings in V_{in} produce large changes in V_{out}.

It is important that the operating parameters of the transistor are chosen and the circuit designed such that as far as possible the transistor operates within a linear portion of the graph, such as that shown between A and B, otherwise the output signal will suffer distortion.

Various configurations of single transistor amplifier are possible, with some providing current gain, some voltage gain, and some both.

From mobile phones to televisions, vast numbers of products include amplifiers for sound reproduction, radio transmission, and signal processing. The first discrete transistor audio amplifiers barely supplied a few hundred milliwatts, but power and audio fidelity gradually increased as better transistors became available and amplifier architecture evolved.

Modern transistor audio amplifiers of up to a few hundred watts are common and relatively inexpensive.

Some musical instrument amplifier manufacturers mix transistors and vacuum tubes in the same circuit, as some believe tubes have a distinctive sound.

Field-effect transistor

MF6
NIKOS
P45N02LD
L22
BC24
R78
R82
L27
C85
R104

High-power N-channel field-effect transistor

The **field-effect transistor** (FET) is a type of transistor that relies on an electric field to control the shape and hence the conductivity of a channel of one type of charge carrier in a semiconductor material. FETs are sometimes called *unipolar transistors* to contrast their single-carrier-type operation with the dual-carrier-type operation of bipolar (junction) transistors (BJT). The *concept* of the FET predates the BJT, though it was not physically implemented until *after* BJTs due to the limitations of semiconductor materials and relative ease of manufacturing BJTs compared to FETs at the time.

Terminals

All FETs have a *gate*, *drain*, and *source* terminal that are roughly similar to the *base*, *collector*, and *emitter* of BJTs. Aside from the JFET, all FETs also have a fourth terminal called the *body*, *base*, *bulk*, or *substrate.* This fourth terminal serves the technical purpose of biasing the transistor into operation; it is rare to make non-trivial use of the body terminal in circuit designs, but its presence is important when setting up the physical layout of an integrated circuit.

Cross Section of an n-type MOSFET

The names of the terminals refer to their functions. The gate terminal may be thought of as controlling the opening and closing of a physical gate. This gate permits electrons to flow through or blocks their passage by creating or eliminating a channel between the source and drain. Electrons flow from the source terminal towards the drain terminal if influenced by an applied voltage. The body simply refers to the bulk of the semiconductor in which the gate, source and drain lie. Usually the body terminal is connected to the highest or lowest voltage within the circuit, depending on type. The body terminal and the source terminal are sometimes connected together since the source is also sometimes connected to the highest or lowest voltage within the circuit, however there are several uses of FETs which do not have such a configuration, such as transmission gates and cascode circuits

FET operation

The FET controls the flow of electrons (or electron holes) from the source to drain by affecting the size and shape of a "conductive channel" created and influenced by voltage (or lack of voltage) applied across the gate and source terminals. (For ease of discussion, this assumes body and source are connected). This conductive channel is the "stream" through which electrons flow from source to drain.

Consider an **n-channel "depletion-mode" device**. A negative gate-to-source voltage causes a *depletion region* to expand in width and encroach on the channel from the sides, narrowing the channel. If the depletion region expands to completely close the channel, the resistance of the channel from source to drain becomes large, and the FET is effectively turned off like a switch. Likewise a positive gate-to-source voltage increases the channel size and allows electrons to flow easily.

Now consider an **n-channel "enhancement-mode" device**. A positive gate-to-source voltage is necessary to create a conductive channel, since one does not exist naturally within the transistor. The positive voltage attracts free-floating electrons within the body towards the gate, forming a conductive channel. But first, enough electrons must be attracted near the gate to counter the dopant ions added to the body of the FET; this forms a region free of mobile carriers called a depletion region, and the phenomenon is referred to as the *threshold voltage* of the FET. Further gate-to-source voltage increase will attract even more electrons towards the gate which are able to create a conductive channel from source to drain; this process is called *inversion*.

For either enhancement- or depletion-mode devices, at drain-to-source voltages much less than gate-to-source voltages, changing the gate voltage will alter the channel resistance, and drain current will be proportional to drain voltage (referenced to source voltage). In this mode the FET operates like a variable resistor and the FET is said to be operating in a *linear mode* or *ohmic mode*.[1][2]

If drain-to-source voltage is increased, this creates a significant asymmetrical change in the shape of the channel due to a gradient of voltage potential from source to drain. The shape of the inversion region becomes "pinched-off" near the drain end of the channel. If drain-to-source voltage is increased further, the pinch-off point of the channel begins to move away from the drain towards the source. The FET is said to be in *saturation mode*;[3] some authors refer to it as *active mode*, for a better analogy with bipolar transistor operating regions.[4][5] The saturation mode, or the region between ohmic and saturation, is used when amplification is needed. The in-between region is sometimes considered to be part of the ohmic or linear region, even where drain current is not approximately linear with drain voltage.

Even though the conductive channel formed by gate-to-source voltage no longer connects source to drain during saturation mode, carriers are not blocked from flowing. Considering again an n-channel device, a depletion region exists in the p-type body, surrounding the conductive channel and drain and source regions. The electrons which comprise the channel are free to move out of the channel through the depletion region if attracted to the drain by drain-to-source voltage. The depletion region is free of carriers and has a resistance similar to silicon. Any increase of the drain-to-source voltage will increase the distance from drain to the pinch-off point, increasing resistance due to the depletion region proportionally to the applied drain-to-source voltage. This proportional change causes the drain-to-source current to remain relatively fixed independent of changes to the drain-to-source voltage and quite unlike the linear mode operation. Thus in saturation mode, the FET behaves as a constant-current source rather than as a resistor and can be used most effectively as a voltage amplifier. In this case, the gate-to-source voltage determines the level of constant current through the channel.

Uses

The most commonly used FET is the MOSFET. The CMOS (complementary-symmetry metal oxide semiconductor) process technology is the basis for modern digital integrated circuits. This process technology uses an arrangement where the (usually "enhancement-mode") p-channel MOSFET and n-channel MOSFET are connected in series such that when one is on, the other is off.

The fragile insulating layer of the MOSFET between the gate and channel makes it vulnerable to electrostatic damage during handling. This is not usually a problem after the device has been installed.

In FETs electrons can flow in either direction through the channel when operated in the linear mode, and the naming convention of drain terminal and source terminal is somewhat arbitrary, as the devices are typically (but not always) built symmetrically from source to drain. This makes FETs suitable for switching analog signals between paths (multiplexing). With this concept, one can construct a solid-state mixing board, for example.

Silicon-controlled rectifier (SCR)

A **silicon-controlled rectifier** (or **semiconductor-controlled rectifier**) is a four-layer solid state device that controls current. The name "silicon controlled rectifier" or **SCR** is General Electric's trade name for a type of thyristor. The SCR was developed by a team of power engineers led by Gordon Hall and commercialised by Frank W. "Bill" Gutzwiller in 1957.

Theory of operation

An SCR is a type of rectifier, controlled by a logic gate signal. It is a four-layer, three-terminal device. A p-type layer acts as an anode and an n-type layer as a cathode; the p-type layer closer to the n-type(cathode) acts as a gate.

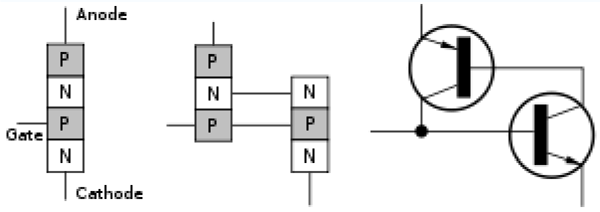

An SCR (left) can be thought of as two BJT's working together (right).

Modes of operation

In the normal "off" state, the device restricts current to the leakage current. When the gate to cathode voltage exceeds a certain threshold, the device turns "on" and conducts current. The device will remain in the "on" state even after gate current is removed so long as current through the device remains above the holding current. Once current falls below the holding current for an appropriate period of time, the device will switch "off".

If the applied voltage increases rapidly enough, capacitive coupling may induce enough charge into the gate to trigger the device into the "on" state; this is referred to as "dv/dt triggering." This is usually prevented by limiting the rate of voltage rise across the device, perhaps by using a snubber. "dv/dt triggering" may not switch the SCR into full conduction rapidly and the partially-triggered SCR may dissipate more power than is usual, possibly harming the device.

SCRs can also be triggered by increasing the forward voltage beyond their rated breakdown voltage (also called as breakover voltage), but again, this does not rapidly switch the entire device into conduction and so may be harmful so this mode of operation is also usually avoided. Also, the actual breakdown voltage may be substantially higher than the rated breakdown voltage, so the exact trigger point will vary from device to device.

SCRs are made with voltage ratings of up to 7500 volts, and with current ratings up to 3000 RMS amperes per device. Some of the larger ones can take over 50 kA in single-pulse operation. SCRs are used in power switching, phase control, chopper, battery chargers, and inverter circuits. Industrially they are applied to produce variable DC voltages for motors (from a few to several thousand HP) from AC line voltage. They control the bulk of the dimmers used in stage lighting, and can also be used in some electric vehicles to modulate the working voltage in a Jacobson circuit. Another common application is phase control circuits used with inductive loads. SCRs can also be found in welding power supplies where they are used to maintain a constant output current or voltage. Large silicon-controlled rectifer assemblies with many individual devices connected in series are used in high-voltage DC converter stations.

Two SCRs in "inverse parallel" are often used in place of a TRIAC for switching inductive loads on AC circuits. Because each SCR only conducts for half of the power cycle and is reverse-biased for the other half-cycle, turn-off of the SCRs is assured. By comparison, the TRIAC is capable of conducting current in both directions and assuring that it switches "off" during the brief zero-crossing of current can be difficult.

Typical electrostatic discharge (ESD) protection structures in integrated circuits produce a parasitic SCR. This SCR is undesired; if it is triggered by accident, the IC can go into latchup and potentially be destroyed

---

TRIAC

A **TRIAC**, or **TRIode for Alternating Current** is an electronic component approximately equivalent to two silicon-controlled rectifiers (SCRs/thyristors) joined in inverse parallel (paralleled but with the polarity reversed) and with their gates connected together. Formal name for a TRIAC is **bidirectional triode thyristor**. This results in a bidirectional electronic switch which can conduct current in either direction when it is triggered (turned on). And thus doesn't have any polarity. It can be triggered by either a positive or a negative voltage being applied to its *gate* electrode (with respect to A1, otherwise known as MT1). Once triggered, the device continues to conduct until the current through it drops below a certain threshold value, such as at the end of a half-cycle of alternating current (AC) mains power. This makes the TRIAC a very convenient switch for AC circuits, allowing the control of very large power flows with milliampere-scale control currents. In addition, applying a trigger pulse at a controllable point in an AC cycle allows one to control the percentage of current that flows through the TRIAC to the load (so-called *phase control*).

Low power TRIACs are used in many applications such as light dimmers, speed controls for electric fans and other electric motors, and in the modern computerized control circuits of many household small

and major appliances. However, when used with inductive loads such as electric fans, care must be taken to assure that the TRIAC will turn off correctly at the end of each half-cycle of the ac power.

A snubber circuit is often used to assist this turn off. Snubber circuits are also used to prevent premature triggering. For higher-powered, more-demanding loads, two SCRs in inverse parallel may be used instead of one TRIAC. Because each SCR will have an entire half-cycle of reverse polarity voltage applied to it, turn-off of the SCRs is assured, no matter what the character of the load.

Triac Schematic Symbol

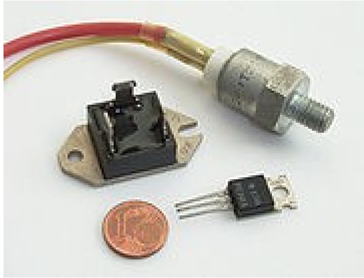

Triacs

DIAC

The **DIAC**, or **diode for alternating current**, is a bidirectional trigger diode that conducts current only after its breakdown voltage has been exceeded momentarily. When this occurs, the resistance of the diode abruptly decreases, leading to a sharp decrease in the voltage drop across the diode and, usually, a sharp increase in current flow through the diode. The diode remains "in conduction" until the current flow through it drops below a value characteristic for the device, called the holding current. Below this value, the diode switches back to its high-resistance (non-conducting) state. When used in AC applications this automatically happens when the current reverses polarity.

Typical Diac voltage and current relationships. Once the voltage exceeds the turn-on threshold, the device turns on and the voltage rapidly falls while the current increases.

The behavior is typically the same for both directions of current flow. Most DIACs have a breakdown voltage around 30 V. In this way, their behavior is somewhat similar to (but much more precisely controlled and taking place at lower voltages than) a neon lamp.

DIACs are a form of thyristor but without a gate electrode. They are typically used for triggering both thyristors and TRIACs - a bidirectional member of the thyristor family. Because of this common usage, many TRIACs contain a built-in DIAC in series with the TRIAC's "gate" terminal.

DIACs are also called *symmetrical trigger diodes* due to the symmetry of their characteristic curve. Because DIACs are bidirectional devices, their terminals are not labeled as *anode* or *cathode* but as A1 and A2 or MT1 ("Main Terminal") and MT2.

ICs

Analog

The MOSFET's advantages in most digital circuits do not translate into supremacy in all analog circuits. The two types of circuit draw upon different features of transistor behavior. Digital circuits switch, spending most of their time outside the switching region, while analog circuits depend on MOSFET behavior held precisely in the switching region of operation. The bipolar junction transistor (BJT) has traditionally been the analog designer's transistor of choice, due largely to its higher transconductance and its higher output impedance (drain-voltage independence) in the switching region.

Nevertheless, MOSFETs are widely used in many types of analog circuits because of certain advantages. The characteristics and performance of many analog circuits can be designed by changing the sizes (length and width) of the MOSFETs used. By comparison, in most bipolar transistors the size of the device does not significantly affect the performance. MOSFETs' ideal characteristics regarding gate current (zero) and drain-source offset voltage (zero) also make them nearly ideal switch elements, and also make switched capacitor analog circuits practical. In their linear region, MOSFETs can be used as precision resistors, which can have a much higher controlled resistance than BJTs. In high power circuits, MOSFETs sometimes have the advantage of not suffering from thermal runaway as BJTs do. Also, they can be formed into capacitors and gyrator circuits which allow op-amps made from them to appear as inductors, thereby allowing all of the normal analog devices, except for diodes (which can be made smaller than a MOSFET anyway), to be built entirely out of MOSFETs. This allows for complete analog circuits to be made on a silicon chip in a much smaller space.

Some ICs combine analog and digital MOSFET circuitry on a single mixed-signal integrated circuit, making the needed board space even smaller. This creates a need to isolate the analog circuits from the digital circuits on a chip level, leading to the use of isolation rings and Silicon-On-Insulator (SOI). The main advantage of BJTs versus MOSFETs in the analog design process is the ability of BJTs to handle a larger current in a smaller space. Fabrication processes exist that incorporate BJTs and MOSFETs into a single device. Mixed-transistor devices are called Bi-FETs (Bipolar-FETs) if they contain just one BJT-FET and BiCMOS (bipolar-CMOS) if they contain complementary BJT-FETs. Such devices have the advantages of both insulated gates and higher current density.

BJTs have some advantages over MOSFETs for at least two digital applications. Firstly, in high speed switching, they do not have the "larger" capacitance from the gate, which when multiplied by the resistance of the channel gives the intrinsic time constant of the process. The intrinsic time constant places a limit on the speed a MOSFET can operate at because higher frequency signals are filtered out. Widening the channel reduces the resistance of the channel, but increases the capacitance by the exact same amount. Reducing the width of the channel increases the resistance, but reduces the capacitance by the same amount. R*C=Tc1, 0.5R*2C=Tc1, 2R*0.5C=Tc1. There is no way to minimize the intrinsic time constant for a certain process. Different processes using different channel lengths, channel heights, gate thicknesses and materials will have different intrinsic time constants. This problem is mostly avoided with a BJT because it does not have a gate.

The second application where BJTs have an advantage over MOSFETs stems from the first. When driving many other gates, called fanout, the resistance of the MOSFET is in series with the gate capacitances of the other FETs, creating a secondary time constant. Delay circuits use this fact to create a fixed signal delay by using a small CMOS device to send a signal to many other, many times larger

CMOS devices. The secondary time constant can be minimized by increasing the driving FET's channel width to decrease its resistance and decreasing the channel widths of the FETs being driven, decreasing their capacitance. The drawback is that it increases the capacitance of the driving FET and increases the resistance of the FETs being driven, but usually these drawbacks are a minimal problem when compared to the timing problem. BJTs are better able to drive the other gates because they can output more current than MOSFETs, allowing for the FETs being driven to charge faster. Many chips use MOSFET inputs and BiCMOS (see above) outputs.

Digital

The growth of digital technologies like the microprocessor has provided the motivation to advance MOSFET technology faster than any other type of silicon-based transistor. A timeline can be found at computerhistory.org.[21] A big advantage of MOSFETs for digital switching is that the oxide layer between the gate and the channel prevents DC current from flowing through the gate, further reducing power consumption and giving a very large input impedance. The insulating oxide between the gate and channel effectively isolates a MOSFET in one logic stage from earlier and later stages, which allows a single MOSFET output to drive a considerable number of MOSFET inputs. Bipolar transistor-based logic (such as TTL) does not have such a high fanout capacity. This isolation also makes it easier for the designers to ignore to some extent loading effects between logic stages independently. That extent is defined by the operating frequency: as frequencies increase, the input impedance of the MOSFETs decreases.

CMOS

The principal reason for the success of the MOSFET was the development of digital CMOS logic, which uses p- and n-channel MOSFETs as building blocks. Overheating is a major concern in integrated circuits since ever more transistors are packed into ever smaller chips. CMOS logic reduces power consumption because no current flows (ideally), and thus no power is consumed, except when the inputs to logic gates are being switched. CMOS accomplishes this current reduction by complementing every nMOSFET with a pMOSFET and connecting both gates and both drains together. A high voltage on the gates will cause the nMOSFET to conduct and the pMOSFET not to conduct and a low voltage on the gates causes the reverse. During the switching time as the voltage goes from one state to another, both MOSFETs will conduct briefly. This arrangement greatly reduces power consumption and heat generation. Digital and analog CMOS applications are described below.

Introduction – IC 555 & 556

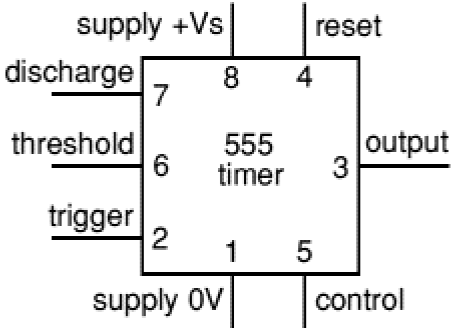
supply +Vs
reset
discharge
7
8
4
threshold
6
555
timer
3
output
trigger
2
1
5
supply 0V
control

Example circuit symbol (above)

Actual pin arrangements (below)

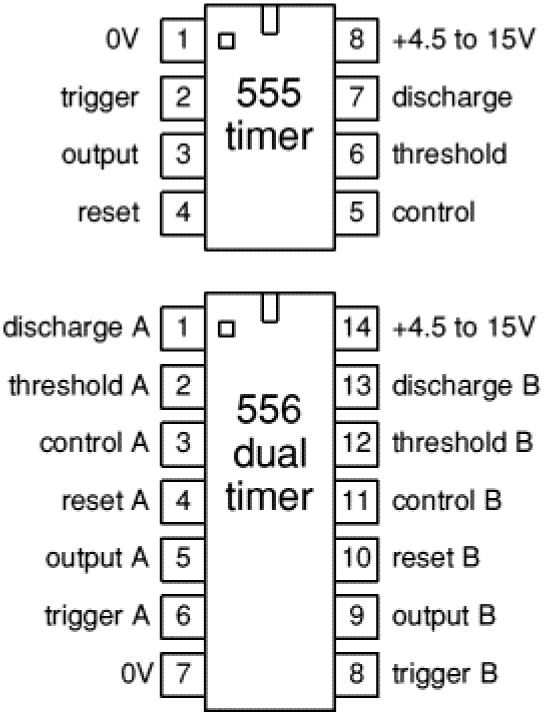
0V
1
8
+4.5 to 15V
trigger
2
555
timer
7
discharge
output
3
6
threshold
reset
4
5
control
discharge A
1
14
+4.5 to 15V
threshold A
2
13
discharge B
control A
3
556
dual
timer
12
threshold B
reset A
4
11
control B
output A
5
10
reset B
trigger A
6
9
output B
0V
7
8
trigger B

The 8-pin 555 timer must be one of the most useful ICs ever made and it is used in many projects. With just a few external components it can be used to build many circuits, not all of them involve timing!

A popular version is the NE555 and this is suitable in most cases where a '555 timer' is specified. The 556 is a dual version of the 555 housed in a 14-pin package, the two timers (A and B) share the same power supply pins. The circuit diagrams on this page show a 555, but they could all be adapted to use one half of a 556.

Low power versions of the 555 are made, such as the ICM7555, but these should only be used when specified (to increase battery life) because their maximum output current of about 20mA (with a 9V supply) is too low for many standard 555 circuits. The ICM7555 has the same pin arrangement as a standard 555.

The circuit symbol for a 555 (and 556) is a box with the pins arranged to suit the circuit diagram: for example 555 pin 8 at the top for the +Vs supply, 555 pin 3 output on the right. Usually just the pin numbers are used and they are not labelled with their function.

The 555 and 556 can be used with a supply voltage (Vs) in the range 4.5 to 15V (18V absolute maximum).

Standard 555 and 556 ICs create a significant 'glitch' on the supply when their output changes state. This is rarely a problem in simple circuits with no other ICs, but in more complex circuits a **smoothing capacitor** (eg 100µF) should be connected across the +Vs and 0V supply near the 555 or 556.

The input and output pin functions are described briefly below and there are fuller explanations covering the various circuits:

BIT

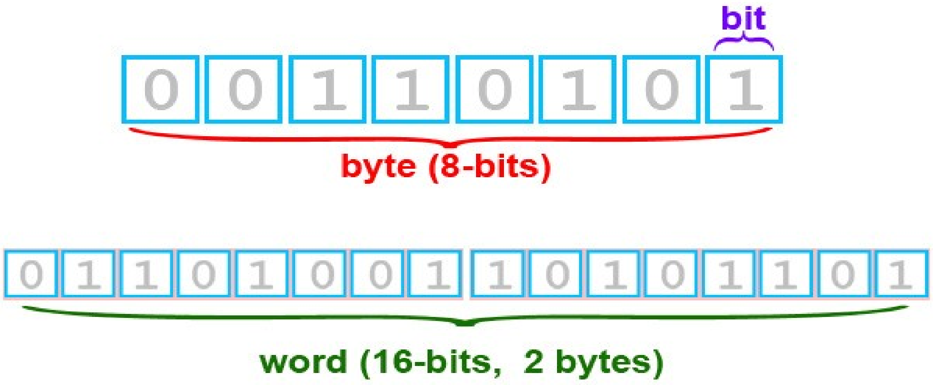

A **bit** or **binary digit** is the basic unit of information in computing and telecommunications. The two values can also be interpreted as logical values (*true*/*false*, *yes*/*no*), algebraic signs (+/−), activation states (*on*/*off*), or any other two-valued attribute. In several popular programming languages, numeric 0 is equivalent (or convertible) to logical *false*, and 1 to *true*.

There are several units of information which are defined as multiples of bits, such as byte (8 bits), kilobit (either 1000 or 2^{10} = 1024 bits), megabyte (either 8000000 or 8×2^{20} = 8388608bits), etc.

v • d • e

Multiples of bits

SI decimal prefixes			IEC binary prefixes	
Name (Symbol)	**Standard SI**	**Binary usage**	**Name (Symbol)**	**Value**
kilobit (kbit)	10^3	2^{10}	kibibit (Kibit)	2^{10}
megabit (Mbit)	10^6	2^{20}	mebibit (Mibit)	2^{20}
gigabit (Gbit)	10^9	2^{30}	gibibit (Gibit)	2^{30}
terabit (Tbit)	10^{12}	2^{40}	tebibit (Tibit)	2^{40}
petabit (Pbit)	10^{15}	2^{50}	pebibit (Pibit)	2^{50}
exabit (Ebit)	10^{18}	2^{60}	exbibit (Eibit)	2^{60}
zettabit (Zbit)	10^{21}	2^{70}	zebibit (Zibit)	2^{70}
yottabit (Ybit)	10^{24}	2^{80}	yobibit (Yibit)	2^{80}

Byte

Binary Arithmetic

For some important aspects of Internet engineering, most notably IP Addressing, an understanding of binary arithmetic is critical. Many strange-looking decimal numbers can only be understood by converting them (at least mentally) to binary.

All digital computers represent data as a collection of *bits*. A bit is the smallest possible unit of information. It can be in one of two states - off or on, 0 or 1. The meaning of the bit, which can represent almost anything, is unimportant at this point. The thing to remember is that *all* computer data - a text file on disk, a program in memory, a packet on a network - is ultimately a collection of bits.

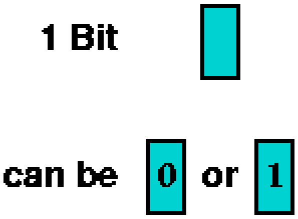

If one bit has two different states, how many states do two bits have? The answer is four. Likewise, three bits have eight states. For example, if a computer display had eight colors available, and you wished to select one of these to draw a diagram in, three bits would be sufficient to represent this information. Each of the eight colors would be assigned to one of the three-bit combinations. Then, you could pick one of the colors by picking the right three-bit combination.

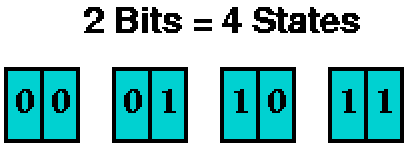

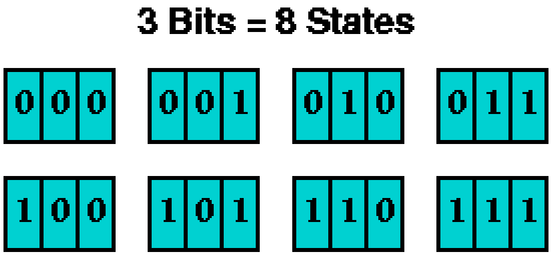

A common and convenient grouping of bits is the *byte* or *octet*, composed of eight bits. If two bits have four combinations, and three bits have eight combinations, how many combinations do eight bits have? If you don't want to write out all the possible byte patterns, just multiply eight twos together - one two for each bit. Two times two is four, so the number of combinations of two bits is four. Two times two times two is eight, so the number of combinations of three bits is eight. Do this eight times - or just compute two to the eighth power - and you discover that a byte has 256 possible states.

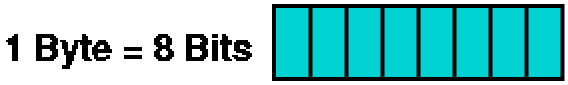

Obviously, if a byte has 256 possible states, its exact state can be represented by a number from 1 to 256. However, since zero is a very important number, a byte is more typically represented by a number from 0 to 255. This is very common, and with bit pattern 00000000 representing zero, and bit pattern 11111111 representing 255. The numbers matching these two patterns, and everything in between, can be computed by assigning a weight to each bit, multiplying each bit's value (0 or 1) by its weight, and then adding the totals. For example, here's how 217 is represented as 11011001 in binary:

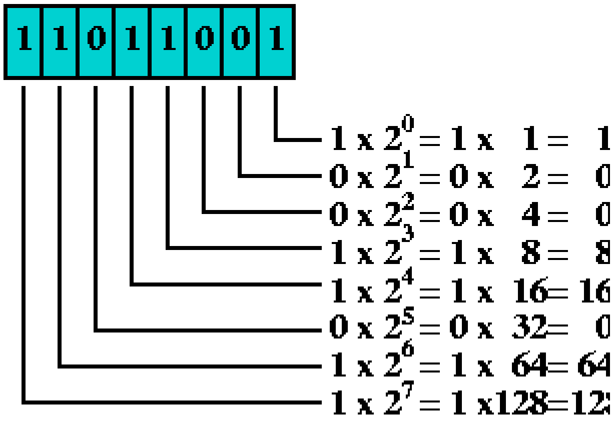

$$1 + 8 + 16 + 64 + 128 = 217$$

To convert a number from decimal to binary, begin at leftmost bit position (128). If the number is larger than or equal to the bit's weight, write a 1 in the bit position, subtract the bit's weight from the number, and continue with the difference. If the number is less than the bit's weight, write a 0 in the bit position and continue without any subtraction. Here's an illustration of converting 141 to binary:

There is a simpler way to convert bytes back and forth between binary and decimal; akin to memorizing multiplication tables. The byte can split into two four-bit halves, each half called a *nibble*. Memorize the decimal values for the high nibble (they're just the multiples of 16). The low nibble is trivial. Every number between 0 and 255 is the sum of one of the high nibble values and one of the low nibble values. Write the high nibble next to the low nibble, and you have the byte value in binary. Conversely, an eight-bit binary byte can be split in half, each nibble converted to decimal and two decimal numbers added together.

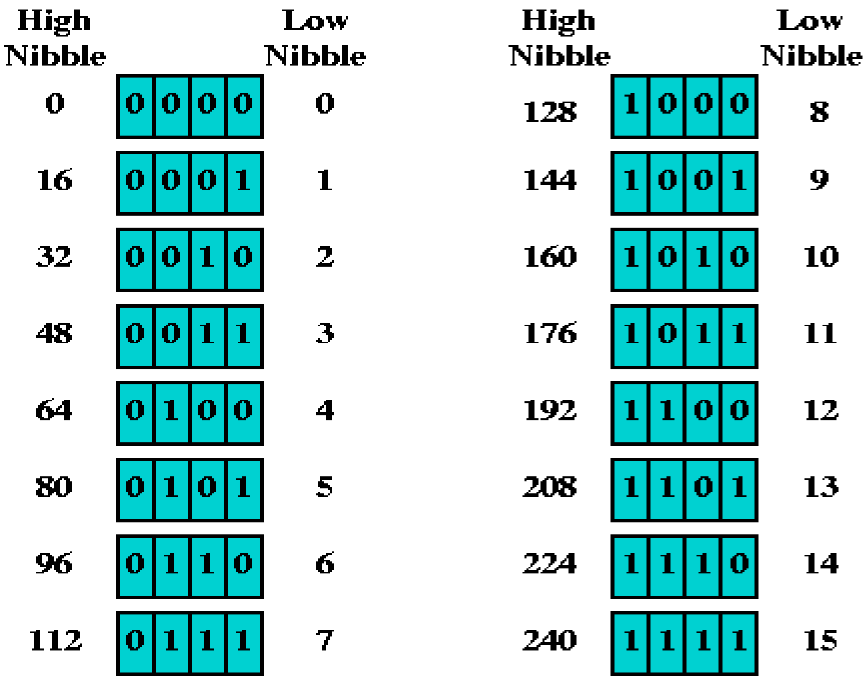

High Nibble		Low Nibble	High Nibble		Low Nibble
0	0 0 0 0	0	128	1 0 0 0	8
16	0 0 0 1	1	144	1 0 0 1	9
32	0 0 1 0	2	160	1 0 1 0	10
48	0 0 1 1	3	176	1 0 1 1	11
64	0 1 0 0	4	192	1 1 0 0	12
80	0 1 0 1	5	208	1 1 0 1	13
96	0 1 1 0	6	224	1 1 1 0	14
112	0 1 1 1	7	240	1 1 1 1	15

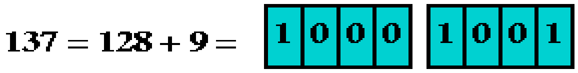
137 = 128 + 9 =
1 0 0 0
1 0 0 1

The most common bit patterns in Internet engineering are those with a string of one bits, followed by a string of zero bits. Here are all such bytes, along with their decimal representation, computed just like the example using 217.

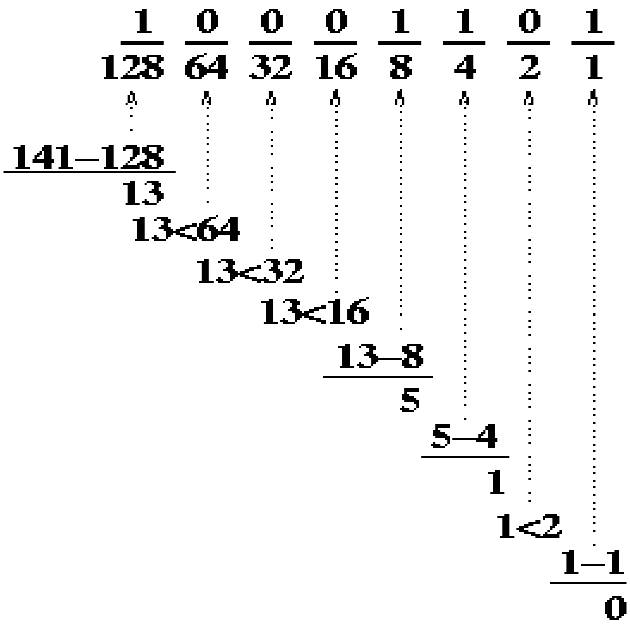
1 0 0 0 1 1 0 1
128 64 32 16 8 4 2 1
141–128
13
13<64
13<32
13<16
13–8
5
5–4
1
1<2
1–1
0

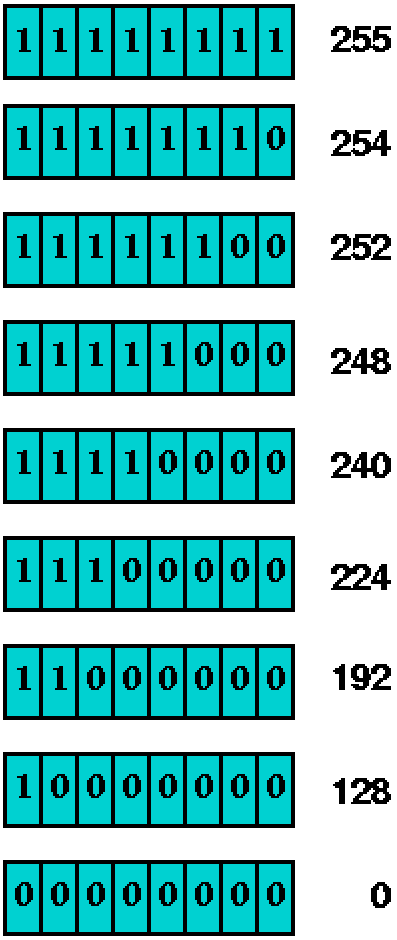

1	1	1	1	1	1	1	1	255
1	1	1	1	1	1	1	0	254
1	1	1	1	1	1	0	0	252
1	1	1	1	1	0	0	0	248
1	1	1	1	0	0	0	0	240
1	1	1	0	0	0	0	0	224
1	1	0	0	0	0	0	0	192
1	0	0	0	0	0	0	0	128
0	0	0	0	0	0	0	0	0

Unit symbol or abbreviation

Prefixes for bit and byte multiples							
Decimal			**Binary**				
Value	**SI**		**Value**	**IEC**		**JEDEC**	
1000	k	kilo	1024	Ki	kibi	K	kilo
1000^2	M	mega	1024^2	Mi	mebi	M	mega
1000^3	G	giga	1024^3	Gi	gibi	G	giga
1000^4	T	tera	1024^4	Ti	tebi		
1000^5	P	peta	1024^5	Pi	pebi		
1000^6	E	exa	1024^6	Ei	exbi		
1000^7	Z	zetta	1024^7	Zi	zebi		
1000^8	Y	yotta	1024^8	Yi	yobi		

Nibble

In computing, a **nibble** (often, **nybble**, or even **nyble** to simulate the spelling of **byte**) is a four-bit aggregation,[1] or half an octet. As a nibble contains 4 bits, there are sixteen (2^4) possible values, so a nibble corresponds to a single hexadecimal digit (thus, it is often referred to as a "hex digit" or "hexit")A full byte (octet) is represented by two hexadecimal digits; therefore, it is common to display a byte of information as two nibbles. The nibble is often called a "semioctet" or a "quartet" in a networking or telecommunication context. Sometimes the set of all 256 byte values is represented as a table 16×16, which gives easily readable hexadecimal codes for each value.

The sixteen nibbles and their equivalents in other numeral systems:

$\mathbf{0}_{hex}$	=	0_{dec}	=	0_{oct}	0	0	0	0
$\mathbf{1}_{hex}$	=	1_{dec}	=	1_{oct}	0	0	0	1
$\mathbf{2}_{hex}$	=	2_{dec}	=	2_{oct}	0	0	1	0
$\mathbf{3}_{hex}$	=	3_{dec}	=	3_{oct}	0	0	1	1
$\mathbf{4}_{hex}$	=	4_{dec}	=	4_{oct}	0	1	0	0
$\mathbf{5}_{hex}$	=	5_{dec}	=	5_{oct}	0	1	0	1
$\mathbf{6}_{hex}$	=	6_{dec}	=	6_{oct}	0	1	1	0
$\mathbf{7}_{hex}$	=	7_{dec}	=	7_{oct}	0	1	1	1
$\mathbf{8}_{hex}$	=	8_{dec}	=	10_{oct}	1	0	0	0
$\mathbf{9}_{hex}$	=	9_{dec}	=	11_{oct}	1	0	0	1
$\mathbf{A}_{hex}$	=	10_{dec}	=	12_{oct}	1	0	1	0
$\mathbf{B}_{hex}$	=	11_{dec}	=	13_{oct}	1	0	1	1
$\mathbf{C}_{hex}$	=	12_{dec}	=	14_{oct}	1	1	0	0
$\mathbf{D}_{hex}$	=	13_{dec}	=	15_{oct}	1	1	0	1
$\mathbf{E}_{hex}$	=	14_{dec}	=	16_{oct}	1	1	1	0
$\mathbf{F}_{hex}$	=	15_{dec}	=	17_{oct}	1	1	1	1

Logic gate

From Wikipedia, the free encyclopedia

Jump to: navigation, search

A **logic gate** performs a logical operation on one or more logic inputs and produces a single logic output. The logic normally performed is Boolean logic and is most commonly found in digital circuits.

Logic gates are primarily implemented electronically using diodes or transistors, but can also be constructed using electromagnetic relays, fluidics, optics, molecules, or even mechanical elements.

In electronic logic, a logic level is represented by a voltage or current, (which depends on the type of electronic logic in use). Each logic gate requires power so that it can source and sink currents to achieve the correct output voltage. In logic circuit diagrams the power is not shown, but in a full electronic schematic, power connections are required

Type	**Distinctive shape**	**Rectangular shape**	**Boolean algebra between A& B**	**Truth table**
AND	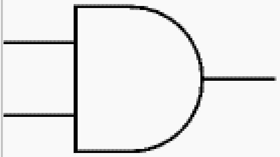	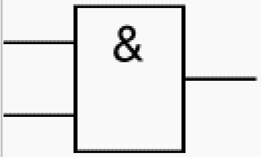	A * B	**INPUT** **OUTPUT** A B A AND B 0 0 0 0 1 0 1 0 0 1 1 1
OR	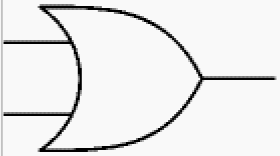	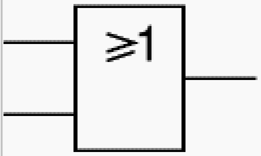	*A* + *B*	**INPUT** **OUTPUT** A B A OR B 0 0 0 0 1 1 1 0 1 1 1 1
NOT	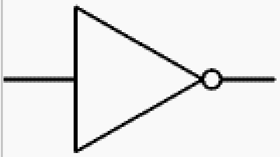	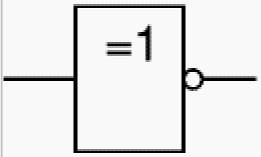	As a Inverter	**INPUT** **OUTPUT** A NOT A 0 1 1 0
In electronics a NOT gate is more commonly called an inverter. The circle on the symbol is called a *bubble*, and is generally used in circuit diagrams to indicate an inverted (active-low) input or output.				

<table>
<tr><td colspan="5"></td></tr>
<tr>
<td>NAND</td>
<td>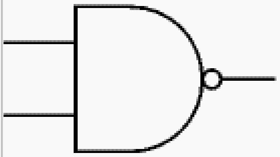</td>
<td>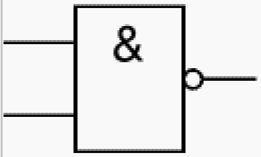 </td>
<td>A * B = A* B</td>
<td><table>
<tr><th colspan="2">INPUT</th><th>OUTPUT</th></tr>
<tr><td>A</td><td>B</td><td>A NAND B</td></tr>
<tr><td>0</td><td>0</td><td>1</td></tr>
<tr><td>0</td><td>1</td><td>1</td></tr>
<tr><td>1</td><td>0</td><td>1</td></tr>
<tr><td>1</td><td>1</td><td>0</td></tr>
</table></td>
</tr>
<tr>
<td>NOR</td>
<td>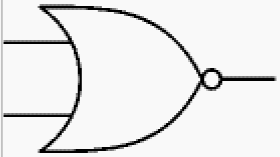</td>
<td>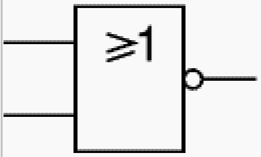 </td>
<td>–</td>
<td><table>
<tr><th colspan="2">INPUT</th><th>OUTPUT</th></tr>
<tr><td>A</td><td>B</td><td>A NOR B</td></tr>
<tr><td>0</td><td>0</td><td>1</td></tr>
<tr><td>0</td><td>1</td><td>0</td></tr>
<tr><td>1</td><td>0</td><td>0</td></tr>
<tr><td>1</td><td>1</td><td>0</td></tr>
</table></td>
</tr>
<tr>
<td>XOR</td>
<td>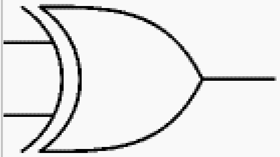</td>
<td>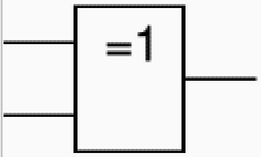 </td>
<td>–</td>
<td><table>
<tr><th colspan="2">INPUT</th><th>OUTPUT</th></tr>
<tr><td>A</td><td>B</td><td>A XOR B</td></tr>
<tr><td>0</td><td>0</td><td>0</td></tr>
<tr><td>0</td><td>1</td><td>1</td></tr>
<tr><td>1</td><td>0</td><td>1</td></tr>
<tr><td>1</td><td>1</td><td>0</td></tr>
</table></td>
</tr>
<tr>
<td>XNOR</td>
<td>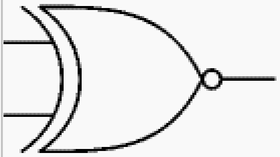</td>
<td>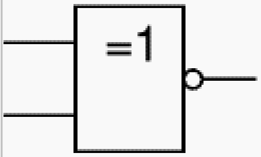 </td>
<td>–</td>
<td><table>
<tr><th colspan="2">INPUT</th><th>OUTPUT</th></tr>
<tr><td>A</td><td>B</td><td>A XNOR B</td></tr>
<tr><td>0</td><td>0</td><td>1</td></tr>
</table></td>
</tr>
</table>

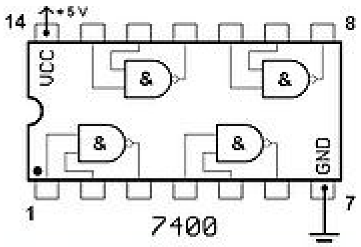

SN 7400N
7645

Flip-flop

In digital circuits, a flip-flop is a term referring to an electronic circuit (a bistable multivibrator) that has two stable states and thereby is capable of serving as one bit of memory. Today, the term flip-flop has come to mostly denote non-transparent (clocked or edge-triggered) devices, while the simpler transparent ones are often referred to as latches; however, as this distinction is quite new, the two words are sometimes used interchangeably (see history).

A flip-flop is usually controlled by one or two control signals and/or a gate or clock signal. The output often includes the complement as well as the normal output. As flip-flops are implemented electronically, they require power and ground connections.

Set-Reset flip-flops (SR flip-flops)

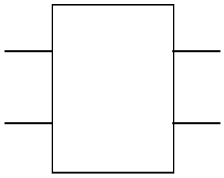

The symbol for an SR latch

The fundamental latch is the simple SR flip-flop , where S and R stand for set and reset respectively. It can be constructed from a pair of cross-coupled NOR logic gates. The stored bit is present on the output marked Q.

Normally, in storage mode, the S and R inputs are both low, and feedback maintains the Q and Q outputs in a constant state, with Q the complement of Q. If S (Set) is pulsed high while R is held low, then the Q output is forced high, and stays high even after S returns low; similarly, if R (Reset) is pulsed high while S is held low, then the Q output is forced low, and stays low even after R returns low.

SR Flip-Flop operation

Characteristic table

S	R	Action
0	0	Keep state
0	1	Q = 0
1	0	Q = 1
1	1	Unstable combination, see race condition

Excitation table

Q (t)	Q (t+1)	S	R	Action
0	0	0	X	No change
0	1	1	0	set
1	0	0	1	reset
1	1	X	0	No change

('X' denotes a don’t care condition; meaning the signal is irrelevant)

JK flip-flop

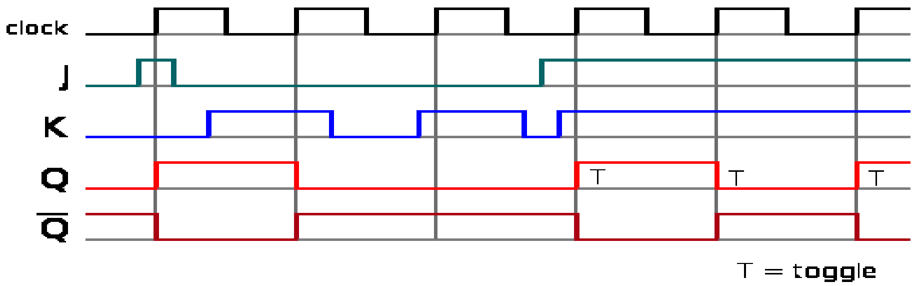

JK flip-flop timing diagram

The JK flip-flop augments the behavior of the SR flip-flop (J=Set, K=Reset) by interpreting the S = R = 1 condition as a "flip" or toggle command. Specifically, the combination J = 1, K = 0 is a command to set the flip-flop; the combination J = 0, K = 1 is a command to reset the flip-flop; and the combination J = K = 1 is a command to toggle the flip-flop, i.e., change its output to the logical complement of its current value. Setting J = K = 0 does NOT result in a D flip-flop, but rather, will hold the current state. To synthesize a D flip-flop, simply set K equal to the complement of J. The JK flip-flop is therefore a universal flip-flop, because it can be configured to work as an SR flip-flop, a D flip-flop, or a T flip-flop. NOTE: The flip flop is positive edge triggered (Clock Pulse) as seen in the timing diagram.

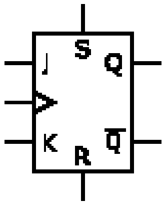

A circuit symbol for a JK flip-flop, where > is the clock input, J and K are data inputs, Q is the stored data output, and Q' is the inverse of Q.

The characteristic equation of the JK flip-flop is:

$$Q_{next} = J\overline{Q} + \overline{K}Q$$

and the corresponding truth table is:

JK Flip Flop operation

Characteristic table

J	K	Qnext	Comment
0	0		hold state
0	1		reset 0
1	0		set 1
1	1		toggle 1

Excitation table

Q	Qnext	J	K	Comment
0	0	0	X	No change
1	1	X	Set	
0	X	1	Reset	
1	X	0	No change	

D flip-flop

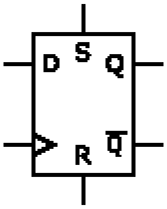

D flip-flop symbol

The Q output always takes on the state of the D input at the moment of a rising clock edge (or falling edge if the clock input is active low). It is called the D flip-flop for this reason, since the output takes the value of the D input or Data input, and Delays it by one clock count. The D flip-flop can be interpreted as a primitive memory cell, zero-order hold, or delay line.

Truth table:

Clock	D	Q	Qprev
Rising edge	0	0	X
Rising edge	1	1	X
Non-Rising	X	Qprev	

('X' denotes a Don't care condition, meaning the signal is irrelevant)

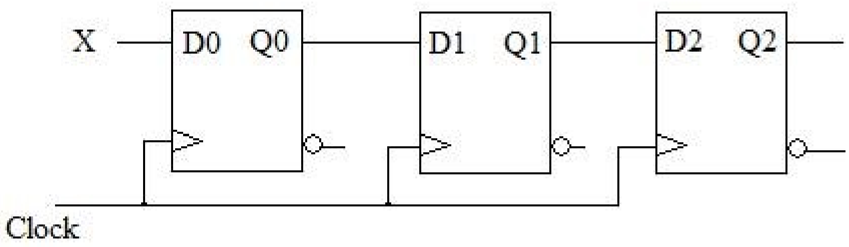

3-bit shift register

These flip flops are very useful, as they form the basis for shift registers, which are an essential part of many electronic devices. The advantage of the D flip-flop over the D-type latch is that it "captures" the signal at the moment the clock goes high, and subsequent changes of the data line do not influence Q until the next rising clock edge. An exception is that some flip-flops have a 'reset' signal input, which will reset Q (to zero), and may be either asynchronous or synchronous with the clock.

The above circuit shifts the contents of the register to the right, one bit position on each active transition of the clock. The input X is shifted into the leftmost bit position.

Master-slave (pulse-triggered) D flip-flop

A master-slave D flip-flop is created by connecting two gated D latches in series, and inverting the enable input to one of them. It is called master-slave because the second latch in the series only changes in response to a change in the first (master) latch. The term pulse triggered means that data are entered on the rising edge of the clock pulse, but the output doesn't reflect the change until the falling edge of the clock pulse.

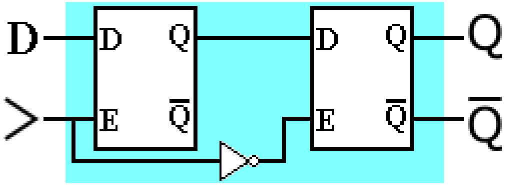

A master slave D flip flop. It responds on the negative edge of the enable input (usually a clock).

For a positive-edge triggered master-slave D flip-flop, when the clock signal is low (logical 0) the “enable” seen by the first or “master” D latch (the inverted clock signal) is high (logical 1). This allows the “master” latch to store the input value when the clock signal transitions from low to high. As the clock signal goes high (0 to 1) the inverted “enable” of the first latch goes low (1 to 0) and the value seen at the input to the master latch is “locked”. Nearly simultaneously, the twice inverted “enable” of the second or “slave” D latch transitions from low to high (0 to 1) with the clock signal. This allows the signal captured at the rising edge of the clock by the now “locked” master latch to pass through the “slave” latch. When the clock signal returns to low (1 to 0), the output of the "slave" latch is "locked", and the value seen at the last rising edge of the clock is held while the “master” latch begins to accept new values in preparation for the next rising clock edge.

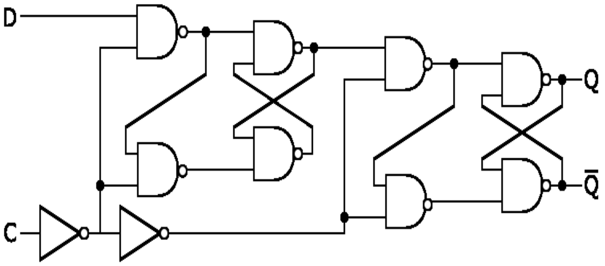

An implementation of a master-slave D flip-flop that is triggered on the positive edge of the clock.

By removing the left-most inverter in the above circuit, a D-type flip flop that strobes on the falling edge of a clock signal can be obtained. This has a truth table like this:

D	Q	>	Qnext
0	X	Falling	0
1	X	Falling	1

Most D-type flip-flops in ICs have the capability to be set and reset, much like an SR flip-flop. Usually, the illegal S = R = 1 condition is resolved in D-type flip-flops.

Inputs				Outputs	
S	R	D	>	Q	Q'
0	1	X	X	0	1
1	0	X	X	1	0
1	1	X	X	1	1

By setting S = R = 0, the flip-flop can be used as described above.

Edge-triggered D flip-flop

A more efficient way to make a D flip-flop is not so easy to understand, but it works the same way. While the master-slave D flip flop is also triggered on the edge of a clock, its components are each triggered by clock levels. The "edge-triggered D flip flop" does not have the master slave properties.

How to Solder

Soldering Advice for Components

It is very tempting to start soldering components onto the circuit board straight away, but please take time to identify all the parts first. You are much less likely to make a mistake if you do this!

1. **Stick all the components onto a sheet of paper using sticky tape.**
2. **Identify each component** and write its name or value beside it.
3. **Add the code (R1, R2, C1 etc.) if necessary.** Many projects from books and magazines label the components with codes (R1, R2, C1, D1 etc.) and you should use the project's parts list to find these codes if they are given.
4. **Resistor values** can be found using the resistor colour code which is explained on our Resistors page. You can print out and make your own Resistor Colour Code Calculator to help you.
5. **Capacitor values** can be difficult to find because there are many types with different labelling systems! The various systems are explained on our Capacitors page.

Some components require special care when soldering. Many must be placed the correct way round and a few are easily damaged by the heat from soldering. Appropriate warnings are given in the table below, together with other advice which may be useful when soldering.

For most projects it is best to put the components onto the board in the order given below:

	Components	Pictures	Reminders and Warnings
1	**IC Holders** (DIL sockets)	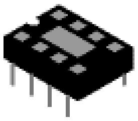	**Connect the correct way round** by making sure the notch is at the correct end. Do NOT put the ICs (chips) in yet.
2	**Resistors**	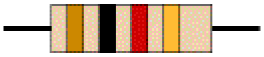	No special precautions are needed with resistors.

3	**Small value capacitors** (usually less than 1µF)	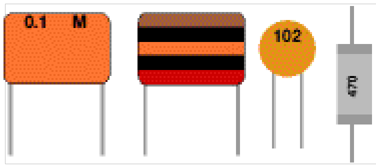	These may be connected either way round. Take care with polystyrene capacitors because they are easily damaged by heat.
4	**Electrolytic capacitors** (1µF and greater)	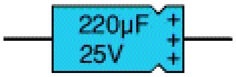	**Connect the correct way round.** They will be marked with a + or - near one lead.
5	**Diodes**	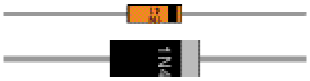	**Connect the correct way round.** Take care with germanium diodes (e.g. OA91) because they are easily damaged by heat.
6	**LEDs**	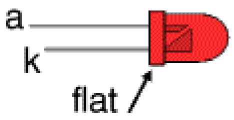	**Connect the correct way round.** The diagram may be labelled **a** or **+** for anode and **k** or **-** for cathode; yes, it really is k, not c, for cathode! The cathode is the short lead and there may be a slight flat on the body of round LEDs.
7	**Transistors**	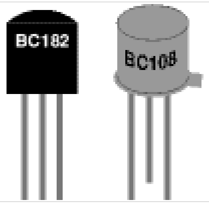	**Connect the correct way round.** Transistors have 3 'legs' (leads) so extra care is needed to ensure the connections are correct. Easily damaged by heat.
8	**Wire Links** between points on the circuit board.	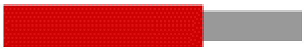 single core wire	Use single core wire, this is one solid wire which is plastic-coated. If there is no danger of touching other parts you can use tinned copper wire, this has no plastic coating and looks just like solder but it is stiffer.
9	**Battery clips**, **buzzers** and other parts with their own wires		**Connect the correct way round.**
10	**Wires** to parts off the circuit board, including **switches**, **relays**, **variable resistors** and **loudspeakers**.	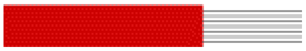 stranded wire	You should use stranded wire which is flexible and plastic-coated. Do not use single core wire because this will break when it is repeatedly flexed.
11	**ICs (chips)**	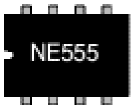	**Connect the correct way round.** **Many ICs are static sensitive.** Leave ICs in their antistatic packaging until you need them, then earth your hands by touching a metal water pipe or window frame before touching the ICs. **Carefully insert ICs in their holders**: make sure all the pins are lined up with the socket then push down firmly with your thumb.

CLASSIFICATION OF FLUXES

1. INORGANIX FLUX
2. ORGANIC FLUX
3. ROSIN BASED FLUX
4. NO CLEAN FLUX

1. INORGANIC FLUX:

Consist of mixture of salts/acids such as HCL, H3PO4, $Zncl_2$, Nh_4Cl, Na_2Bo, having picking action through generated HCI acid in contact with moisture or added water. These fluxes are very corrosive in nature and generally not used in electronic assemblies.

2. NON ROSIN BASED FLUXES

These fluxes do not contain Rosin but contain only activators based on GLIMTIC, LACTIC ACIDS/HYDROCHLORIDES, water soluble and meditatively corrosive, hence must be cleaned after soldering process thoroughly.

3. ROSIN BASED FLUXES

Rosin which is naturally available from pine a tree by tapping is made into crystalline structure with pale yellow to Dark brown colour is used as base.

Rosin is a mixture of several organic compounds like ABETIC ACID, PIMARIC ACID, etc., pure Rosin is a weak pickling agent. For heavier passivation layer, surface activators like Bromides/ chlorides are added. The acid contents for various R type of fluxes are given below.

R – ROSIN NON ACTIVATED.	ACID CONTENT: < 0.2%
RAM – ROSIN MILDLY ACTIVATED.	ACID CONTENT: 0.2% TO 0.4%
RA – ROSIN ACTIVATED.	ACID CONTENT: 0.4% to 4%
RSA – ROSIN SUPER ACTIVATED.	ACID CONTENT: > 4%

4. NO CLEAN FLUX

Does not contain natural Rosin but obtained by synthetic resins solvents, foaming agents and activators. Generally does not leave any residues.

1. SOLDER ALLOY

WHY TIN & LEAD

1. Tin is not appreciably affected by air or water.
2. The corrosion resistance of tin is an important factor for using as a coating on Copper wire to protect against corrosion.
3. Tin reacts and alloys with metal easily.
4. Lead is soft and dense, but surface gets quickly corroded.
5. Lead reduces the melting temperature and brittleness thus adding to the overall strength of alloy.

METALLURGY OF SOLDER JOINTS

When two metals “A” and “B” are soldered with a filler material “S”. The following regions occur in the final joint structure.

1. PURE “A”
2. ALLOY OF “A” AND “S”
3. PURE “S”
4. ALLOY OF “S” AND “B”
5. PURE “B”

The alloy region consists of irreversible compounds created out of the metals and the filler. The filler, the thickness of this alloy being 0.1 to 1.0 microns. This zone is usually called as intermetallic layer in case of Cu-Solder (sn-pb) combination. They are typically cu3sn, cu6sn5. This intermetallic layer is mainly responsible for the strength of the joint. This layer thickness increases with time and temperature of the soldering. This intermatellic layer is brittle in nature. Hence the thickness of this layer has to be controlled.

CHEMICAL COMPOSITION OF Sn63, Sn60

	Sn63	**Sn60**
TIN Composition	: 62.5-62.5%	59.5% - 61.5%
Solid Melting Range	: 183 Deg.C	183 Deg.C
Liquid Melting range	: 183 Deg.C	188 Deg. C
Plastic Region	: Nil	Slight

What is solder?

Reels of solder

Solder is an alloy (mixture) of tin and lead, typically 60% tin and 40% lead. It melts at a temperature of about 200°C. Coating a surface with solder is called 'tinning' because of the tin content of solder. Lead is poisonous and you should always wash your hands after using solder.

Solder for electronics use contains tiny cores of flux, like the wires inside a mains flex. The flux is corrosive, like an acid, and it cleans the metal surfaces as the solder melts. This is why you must melt the solder actually on the joint, not on the iron tip. Without flux most joints would fail because metals quickly oxidise and the solder itself will not flow properly onto a dirty, oxidised, metal surface.

The best size of solder for electronics is 22swg (SWG= standard wire gauge).

Desoldering

At some stage you will probably need to desolder a joint to remove or re-position a wire or component. There are two ways to remove the solder:

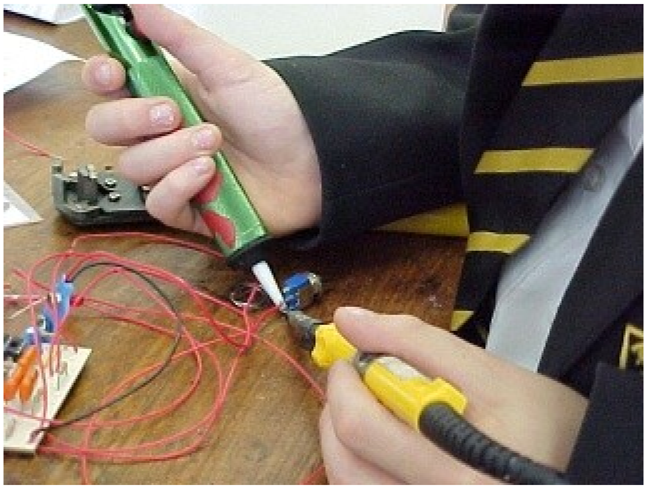

Using a desoldering pump (solder sucker)

1. **With a desoldering pump (solder sucker)**

- Set the pump by pushing the spring-loaded plunger down until it locks.
- Apply both the pump nozzle and the tip of your soldering iron to the joint.
- Wait a second or two for the solder to melt.
- Then press the button on the pump to release the plunger and suck the molten solder into the tool.
- Repeat if necessary to remove as much solder as possible.
- The pump will need emptying occasionally by unscrewing the nozzle.

XYTRO WICK
SOLDER REMOVER
SW205

Solder remover wick

2. **With solder remover wick (copper braid)**

- Apply both the end of the wick and the tip of your soldering iron to the joint.
- As the solder melts most of it will flow onto the wick, away from the joint.
- Remove the wick first, then the soldering iron.
- Cut off and discard the end of the wick coated with solder.

After removing most of the solder from the joint(s) you may be able to remove the wire or component lead straight away (allow a few seconds for it to cool). If the joint will not come apart easily apply your soldering iron to melt the remaining traces of solder at the same time as pulling the joint apart, taking care to avoid burning yourself.

Soldering iron

For electronics work the best type is one powered by mains electricity (230V in the UK), it should have a heatproof cable for safety. The iron's power rating should be 15 to 25W and it should be fitted with a small bit of 2 to 3mm diameter.

Other types of soldering iron

Low voltage soldering irons are available, but their extra safety is undermined if you have a mains lead to their power supply! **Temperature controlled** irons are excellent for frequent use, but not worth the extra expense if you are a beginner. **Gas-powered** irons are designed for use where no mains supply is available and are not suitable for everyday use. Pistol shaped **solder guns** are far too powerful and cumbersome for normal electronics use.

Soldering iron stand

You must have a safe place to put the iron when you are not holding it. The stand should include a sponge which can be dampened for cleaning the tip of the iron.

Desoldering pump (solder sucker)

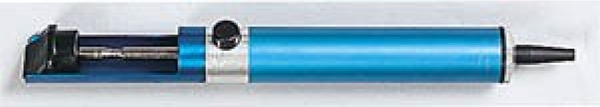

A tool for removing solder when desoldering a joint to correct a mistake or replace a component.

Solder remover wick (copper braid)

This is an alternative to the desoldering pump shown above.

XYTRO WICK
SOLDER REMOVER
SW205

Reel of solder

The best size for electronics is 22swg
(SWG standardwire gauge).

Side cutters

For trimming component leads close to the circuit board.

Wire strippers

Most designs include a cutter as suitable for trimming component leads.

Small pliers

Usually called 'snipe nose' pliers, these are for bending component leads etc. If you put a strong rubber band across the handles the pliers make a convenient holder for parts such as switches while you solder the contacts.

Small flat-blade screwdriver

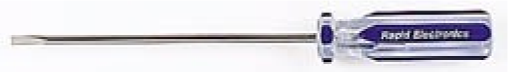

For scraping away excess flux and dirt between tracks, as well as driving screws!

Heat sink

You can buy a special tool, but a standard crocodile clip works just as well and is cheaper.

The following tool is only required if you are using stripboard:

Track cutter

A 3mm drill bit can be used instead, in fact the tool is usually just a 3mm drill bit with a proper handle fitted.

The following tools are only required if you make your own PCBs:

PCB rubber

This is an abrasive rubber for cleaning PCBs. It can also be used to clean strip board where the copper tracks have become dull and tarnished.

Breadboard

A breadboard (solder less breadboard, protoboard, plug board) is reusable sometimes [1] solder less device used to build a (generally temporary) prototype of an electronic circuit and for experimenting with circuit designs. This is in contrast to strip board (Vero board) and similar prototyping printed circuit boards, which are used to build more permanent soldered prototypes or one-offs, and cannot easily be reused. A variety of electronic systems may be prototyped by using breadboards, from small analog and digital circuits to complete central processing units (CPUs).

The term breadboard is derived from an early form of point-to-point construction: in particular, the practice of constructing simple circuits (usually using valves/tubes) on a convenient wooden base, similar to a cutting board like the kind used for slicing bread with a knife. It can also be viewed as bread with a large number of pores (holes for connection); like the bread most commonly used in America and Europe, a modern prototyping board is typically white or off-white.

Jump wires

The jump wires for breadboarding can be obtained in ready-to-use jump wire sets or can be manually manufactured. The latter can become tedious work for larger circuits. Ready-to-use jump wires come in different qualities, some even with tiny plugs attached to the wire ends. Jump wire material for ready-made or home-made wires should usually be 22 AWG (0.33 mm²) solid copper, tin-plated wire - assuming no tiny plugs are to be attached to the wire ends. The wire ends should be stripped 3/16" to 5/16" (approx. 5 mm to 8 mm). Shorter stripped wires might result in bad contact with the board's spring clips (insulation being caught in the springs). Longer stripped wires increase the likelihood of short-circuits on the board. Needle-nose pliers and tweezers are helpful when inserting or removing wires, particularly on crowded boards.

Differently colored wires and color coding discipline are often adhered to for consistency. However, the number of available colors is typically far less than the number of signal types or paths. So typically a few wire colors get reserved for the supply voltages and ground (e.g. red, blue, black), some more for main signals, while the rest often get random colors. There are ready-to-use jump wire sets on the market where the color indicates the length of the wires; however, these sets do not allow applying a meaningful color coding schema.

Diagram

A "full size" terminal breadboard strip typically consists of around 56 to 65 rows of connectors, each row containing the above mentioned two sets of connected clips (A to E and F to J). "Small size" strips typically come with around 30 rows.

Terminal Strip: **Breadboard**

```
  A B C D E   F G H I J

 1 o-o-o-o-o v o-o-o-o-o
 2 o-o-o-o-o   o-o-o-o-o
 3 o-o-o-o-o   o-o-o-o-o
 ~
 ~
61 o-o-o-o-o   o-o-o-o-o
62 o-o-o-o-o   o-o-o-o-o
63 o-o-o-o-o ^ o-o-o-o-o
```

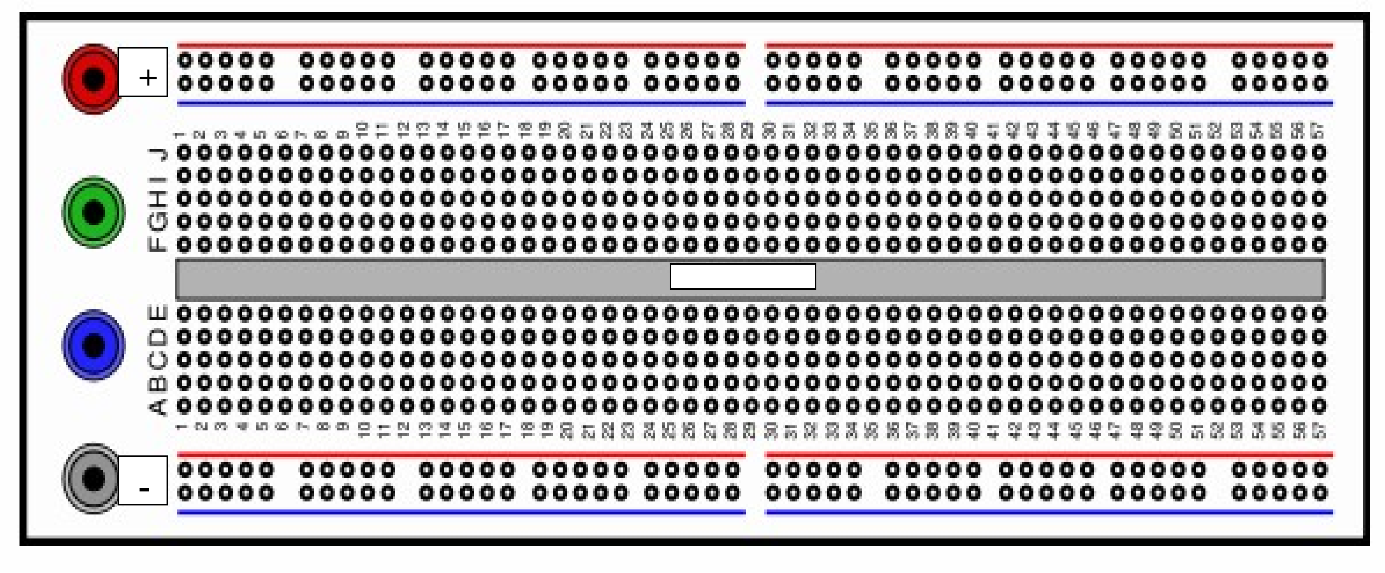
+
F G H I J
A B C D E
-
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57

Printed circuit board (PCB)

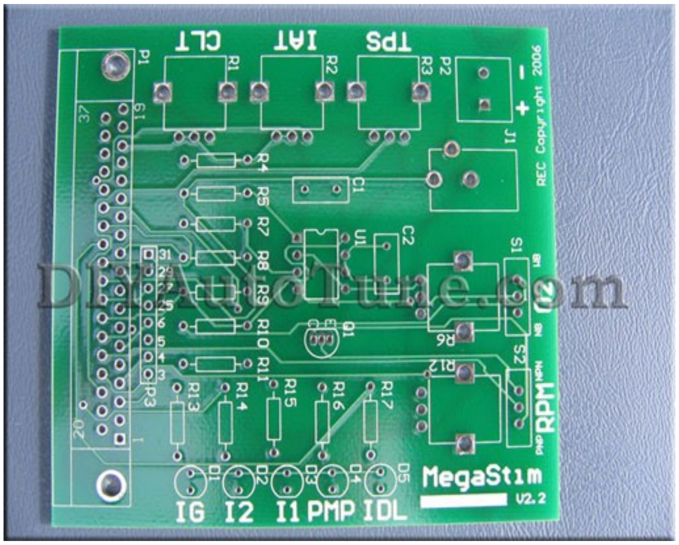
CLT
IAT
TPS
P1
P2
J1
REC Copyright 2006
S1
S2
RPM
Q1
U1
C1
C2
P3
MegaStim
V2.2
IG
I2
I1
PMP
IDL
DIYAutoTune.com

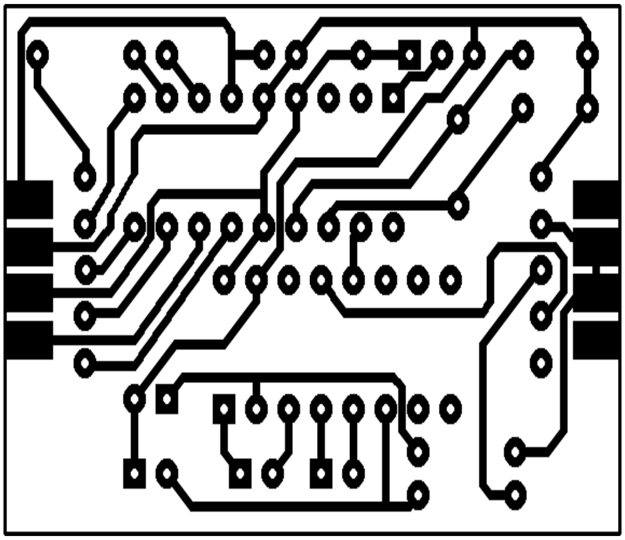

PCBs are boards whereupon electronic circuits have been etched. PCBs are rugged, inexpensive, and can be highly reliable. They require much more layout effort and higher initial cost than either wire-wrapped or point-to-point constructed circuits, but are much cheaper and faster for high-volume production. Much of the electronics industry's PCB design, assembly, and quality control needs are set by standards that are published by the IPC organization.

ELECTRONIC PRODUCTS

Consisting of electronic assemblies, a combination of electronic components and devices and PCBs. Components need to be interconnected for their functioning, to get the desired results.

Generally, these components are mounted on Printed wiring board which provides support and required interconnection.

ABOUT THE PCB

Printed wiring Board, an essential part of total electronic packaging system, which provides interconnection between components, and physical support for the whole assembly. Various types and configuration of components are mounted on Printed wiring board to make the board functional as an electronic device.

The printed wiring boards are classified as follows:

i. Single sided
ii. Double sided
iii. Multi layer
iv. Flexible

The single sided PCB is the simplest, and has the conductor pattern printed only on one side. The components are inserted and mounted on the other side of the PCB. The double sided PCB has conductor patterns on both sides and components are mounted on one side, while in a multilayer PCB has several layers of conductors which are sand witched together.

The PCB consists of conductor tracks of Solder plated copper and required size of Holes (or plated through holes) for mounting various components as per design.

The basic substrate for the PCBs is made up of one of the following insulating materials:

i. Paper phenol
ii. Glass epoxy
iii. Glass polyamide
iv. Ceramic

The flexible boards are generally made up of polyamide substrate.

The Ceramic substrate is generally used in Defense applications where the ceramic components are used.

SOLDER MASK: Solder mask is applied on the Bare printed Circuit board for avoiding solder coating on unwanted conductor tracks during wave soldering and preventing conductor tracks from exposing to atmosphere. Normal conformal coating materials are

1. Acrylic
2. Polyurethane
3. Silicone
4. Paraxylylene
5. Epoxy

These coatings are applied by means of dipping, spraying, brushing or vacuum deposition depending upon material used.

Surface-mount technology

Surface-mount technology (**SMT**) is a method for constructing electronic circuits in which the components (SMC or Surface Mounted Components) are mounted directly onto the surface of printed circuit boards (PCBs). Electronic devices so made are called *surface-mount devices* or **SMD**s. In the industry it has largely replaced the through-hole technology construction method of fitting components with wire leads into holes in the circuit board.

An SMT component is usually smaller than its through-hole counterpart because it has either smaller leads or no leads at all. It may have short pins or leads of various styles, flat contacts, a matrix of solder balls (BGAs), or terminations on the body of the component.

GETTING STARTED IN ELECTRONICS

BY

FORREST M. MIMS, III

FIRST PRINTING - 1983
SECOND PRINTING - 1984
THIRD PRINTING - 1985
FOURTH PRINTING - 1986
FIFTH PRINTING - 1987
SIXTH PRINTING - 1987
SEVENTH PRINTING - 1988
EIGHTH PRINTING - 1991
NINTH PRINTING - 1991
TENTH PRINTING - 1992
ELEVENTH PRINTING - 1993
TWELFTH PRINTING - 1994

COPYRIGHT © 1983 BY FORREST M. MIMS, III
ALL RIGHTS RESERVED
PRINTED IN THE UNITED STATES OF AMERICA

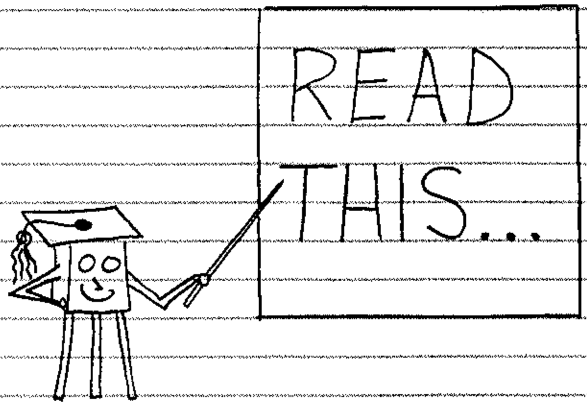

THIS BOOK IS FOR THE ENTERTAINMENT AND EDIFICATION OF ITS READERS. WHILE REASONABLE CARE HAS BEEN EXERCISED WITH RESPECT TO ITS ACCURACY, THE AUTHOR AND RADIO SHACK ASSUME NO RESPONSIBILITY FOR ERRORS, OMISSIONS OR SUITABILITY OF ITS CONTENTS FOR ANY APPLICATION. NEITHER DO WE ASSUME ANY LIABILITY FOR ANY DAMAGES RESULTING FROM USE OF INFORMATION IN THIS BOOK. IT IS YOUR RESPONSIBILITY TO DETERMINE IF USE, MANUFACTURE OR SALE OF ANY DEVICE THAT INCORPORATES INFORMATION IN THIS BOOK INFRINGES ANY PATENTS, COPYRIGHTS OR OTHER RIGHTS.

CAUTION: THIS BOOK INCLUDES SEVERAL REFERENCES TO ELECTRICAL SAFETY WHICH MUST BE HEEDED. IT IS ESSENTIAL THAT CAREFUL SUPERVISION BE GIVEN CHILDREN WORKING WITH LINE-POWERED ELECTRONIC CIRCUITS AND SOLDERING IRONS.

DUE TO THE MANY CUSTOMER INQUIRIES RECEIVED BY RADIO SHACK AND THE AUTHOR, IT IS IMPOSSIBLE TO ANSWER REQUESTS FOR ADDITIONAL INFORMATION (CUSTOM CIRCUIT DESIGNS, TECHNICAL ADVICE, TROUBLESHOOTING ASSISTANCE, ETC.). BUT THOUGH WE CANNOT ACKNOWLEDGE INDIVIDUAL INQUIRIES, WE WILL BE HAPPY TO RECEIVE ANY COMMENTS, IMPRESSIONS OR SUGGESTIONS.

THANKS IN ADVANCE TO THOSE OF YOU WHO WRITE! BUT PLEASE REMEMBER WE ARE UNABLE TO GIVE YOU A PERSONAL REPLY.

CONTENTS

GETTING STARTED IN ELECTRONICS 5
1. ELECTRICITY 6
  PUTTING ELECTRICITY TO WORK 7
  BACK TO BASICS 8-9
  STATIC ELECTRICITY 10-12
  ELECTRICAL CURRENT 13
  DIRECT CURRENT ELECTRICITY 14-17
  ALTERNATING CURRENT ELECTRICITY 18
  MEASURING AC AND DC 19
  ELECTRICAL CIRCUITS 20-21
  PULSES, WAVES, SIGNALS AND NOISE 22-23
2. ELECTRONIC COMPONENTS 24
  WIRE AND CABLE 24
  SWITCHES 25
  RELAYS 26
  MOVING COIL METERS 26
  MICROPHONES AND SPEAKERS 27
  RESISTORS 28-31
  CAPACITORS 32-35
  RESISTOR AND CAPACITOR APPLICATIONS 36-37
  COILS 38-39
  TRANSFORMERS 40-41
3. SEMICONDUCTORS 42
  SILICON 42-43
  THE DIODE 44-47
  THE TRANSISTOR 48
    BIPOLAR TRANSISTORS 48-51
    FIELD-EFFECT TRANSISTORS 52-56
  THE UNIJUNCTION TRANSISTOR 57
  THE THYRISTOR 58
    SILICON-CONTROLLED RECTIFIERS 58-59
    TRIACS 60-61
4. PHOTONIC SEMICONDUCTORS 62
  LIGHT 62-63
  OPTICAL COMPONENTS 64
  HOW CONVEX LENSES ARE USED 65
  SEMICONDUCTOR LIGHT SOURCES 66
    LIGHT EMITTING DIODES 66-69
  SEMICONDUCTOR LIGHT DETECTORS 70
    PHOTORESISTIVE LIGHT DETECTORS 70-71
    PN JUNCTION LIGHT DETECTORS 72
      PHOTODIODES 72-73
      PHOTOTRANSISTORS 74-75
      PHOTOTHYRISTORS 76
      SOLAR CELLS 77

5. INTEGRATED CIRCUITS 78-79

6. DIGITAL INTEGRATED CIRCUITS 80
- MECHANICAL SWITCH GATES 80
- THE BINARY CONNECTION 81
- DIODE GATES 82
- TRANSISTOR GATES 83
- GATE SYMBOLS 84-85
- DATA HIGHWAYS 85
- HOW GATES ARE USED 86
- COMBINATIONAL LOGIC CIRCUITS 86-87
- SEQUENTIAL LOGIC CIRCUITS 88-89
- A COMBINATIONAL-SEQUENTIAL LOGIC SYSTEM 90
- DIGITAL IC FAMILIES 91

7. LINEAR INTEGRATED CIRCUITS 92
- THE BASIC LINEAR CIRCUIT 92
- OPERATIONAL AMPLIFIERS 93
- TIMERS 94
- FUNCTION GENERATORS 94
- VOLTAGE REGULATORS 95
- OTHER LINEAR IC'S 95

8. CIRCUIT ASSEMBLY TIPS 96
- TEMPORARY CIRCUITS 96
- PERMANENT CIRCUITS 96-97
- HOW TO SOLDER 98
- POWERING ELECTRONIC CIRCUITS 99

9. 100 ELECTRONIC CIRCUITS 100
- DIODE CIRCUITS 101
  - SMALL SIGNAL DIODES AND RECTIFIERS 101-102
  - ZENER DIODE CIRCUITS 103
- TRANSISTOR CIRCUITS 104
  - BIPOLAR TRANSISTOR CIRCUITS 104-105
  - JUNCTION FET AND POWER MOSFET CIRCUITS 106-107
- UNIJUNCTION TRANSISTOR CIRCUITS 108-109
- THYRISTOR CIRCUITS 110
  - SCR CIRCUITS 110
  - TRIAC CIRCUITS 111
- PHOTONIC CIRCUITS 112
  - LIGHT EMITTING DIODE (LED) CIRCUITS 112-113
  - SEMICONDUCTOR LIGHT DETECTOR CIRCUITS 114-115
- DIGITAL IC CIRCUITS 116
  - TTL CIRCUITS 116-117
  - CMOS CIRCUITS 118-121
- LINEAR IC CIRCUITS 122
  - OPERATIONAL AMPLIFIER (OP-AMP) CIRCUITS 122-123
  - COMPARATOR CIRCUITS 124
  - VOLTAGE REGULATOR CIRCUITS 125
  - TIMER CIRCUITS 126-127

INDEX 128

GETTING STARTED IN ELECTRONICS

WELCOME TO THE WORLD OF ELECTRONICS, ONE OF THE FASTEST GROWING OF TODAY'S "HIGH-TECH" FIELDS *AND* AN EDUCATIONAL AND ENTERTAINING HOBBY. THIS BOOK WILL TAKE YOU FROM STATIC ELECTRICITY TO SOLID-STATE ELECTRONICS. ALONG THE WAY WE'LL COVER ELECTRICITY, ELECTRONIC COMPONENTS AND INTEGRATED CIRCUITS (IC'S). CHAPTERS 3-7 SHOW HOW COMPONENTS AND IC'S ARE USED TO FORM ELECTRONIC CIRCUITS. CHAPTER 9 GIVES PLANS FOR 100 CIRCUITS, EACH OF WHICH I'VE BUILT AND TESTED. "PAGE ARROWS" (P.xx ⇨) THROUGHOUT THE BOOK REFER YOU TO RELATED TOPICS IN FUTURE CHAPTERS (LIKE WORKING VERSIONS OF MANY EXAMPLE CIRCUITS IN CHAPTERS 3-7). I HOPE YOU FIND THIS BOOK USEFUL, EDUCATIONAL AND, ESPECIALLY, FUN! Forrest M. Mims, III

GOING FURTHER IN ELECTRONICS

I HOPE THIS BOOK ENCOURAGES YOU TO GO FURTHER IN ELECTRONICS. BEGIN BY GETTING RADIO SHACK'S "SEMICONDUCTOR REFERENCE HANDBOOK" AND "ENGINEER'S MINI-NOTEBOOK" SERIES. READ ELECTRONICS MAGAZINES LIKE *COMPUTERCRAFT*, *RADIO-ELECTRONICS*, *BYTE* AND *QST*. YOU MAY ALSO WISH TO READ "ELECTRONICS NOTEBOOK," A COLUMN I WRITE EACH MONTH FOR *COMPUTERCRAFT*. MANY OF THESE COLUMNS HAVE BEEN COMPILED IN A SERIES OF BOOKS, INCLUDING "THE FORREST MIMS CIRCUIT SCRAPBOOK" (McGRAW-HILL, 1983) AND "FORREST MIMS' CIRCUIT SCRAPBOOK II (HOWARD W. SAMS, 1986). THESE BOOKS DESCRIBE IN DETAIL SEVERAL OF THE CIRCUITS IN *THIS* BOOK AND MANY OTHER CIRCUITS. QUESTIONS? THIS BOOK WILL RAISE MANY! IF YOU CAN'T FIND ANSWERS IN THE REFERENCES ABOVE, TRY A *GOOD* LIBRARY. FINDING ANSWERS MAY TAKE TIME, BUT YOU'LL LEARN MUCH IN THE PROCESS. PERHAPS YOU'LL EVEN CONSIDER FORMAL TRAINING FOR A CAREER IN ELECTRONICS.

A SPECIAL NOTE TO EDUCATORS

THIS BOOK CAN GIVE YOUR STUDENTS A BASIC KNOWLEDGE OF ELECTRONICS. YOU CAN ASSIGN THE BOOK FOR OUTSIDE READING AND TEST STUDENTS AS THEY PROGRESS. OR YOU CAN DEVELOP A TOTAL COURSE, COMPLETE WITH DEMONSTRATIONS, EXPERIMENTS AND LECTURES. THANKS TO RADIO SHACK'S SOLDERLESS MODULAR SOCKETS, YOU AND YOUR STUDENTS SHOULD BE ABLE TO ASSEMBLE TEST VERSIONS OF VIRTUALLY EVERY CIRCUIT IN CHAPTER 9 ("100 ELECTRONIC CIRCUITS"). INCIDENTALLY, VOLUME BUYERS CAN RECEIVE A PRICE DISCOUNT FROM RADIO SHACK ON THIS BOOK AND ELECTRONIC COMPONENTS. SEE THE LATEST RADIO SHACK CATALOG FOR DETAILS. (PRICE DISCOUNTS ARE OPTIONAL AT RADIO SHACK DEALERS AND FRANCHISE STORES.)

1. ELECTRICITY

THE ONLY DIFFERENCE BETWEEN A BOLT OF LIGHTNING AND THE SPARK BETWEEN YOUR FINGER AND A DOORKNOB ON A DRY DAY IS QUANTITY. BOTH ARE ELECTRICITY. BENJAMIN FRANKLIN FIRST CONFIRMED THIS WITH HIS FAMOUS KITE EXPERIMENT.

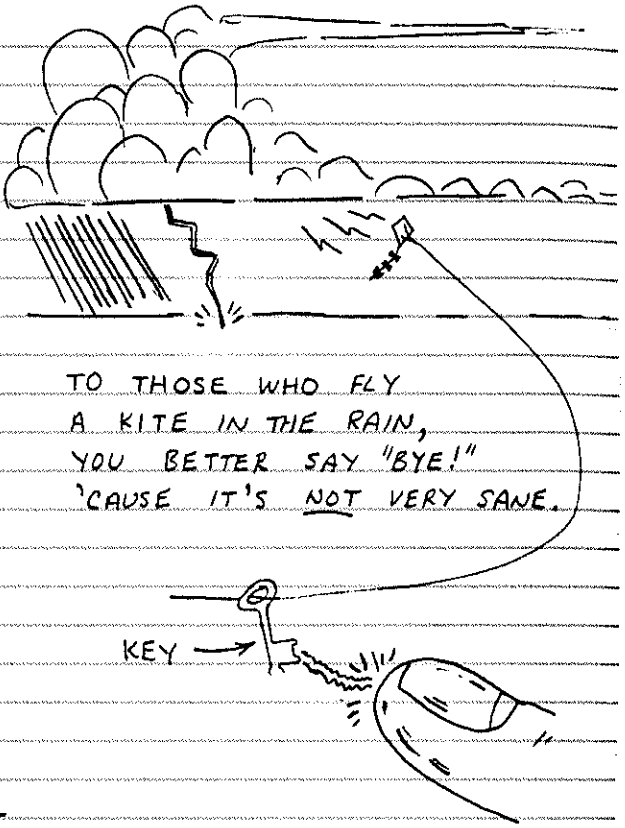

TO THOSE WHO FLY
A KITE IN THE RAIN,
YOU BETTER SAY "BYE!"
'CAUSE IT'S NOT VERY SANE.

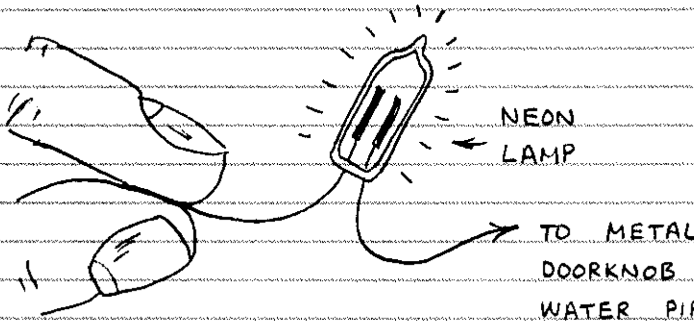

HERE'S A NEAT WAY TO "SEE" ELECTRICITY WITHOUT BEING ZAPPED: GRASP ONE LEAD FROM A NEON LAMP, WALK ACROSS A CARPET WHILE WEARING HARD SOLED SHOES AND TOUCH THE SECOND LEAD FROM THE LAMP TO A METAL OBJECT. THE LAMP WILL FLASH (UNLESS THE RELATIVE HUMIDITY IS HIGH).

OF COURSE, YOU CANNOT "SEE" ELECTRICITY! YOU SEE ITS EFFECT UPON AIR AND THE NEON IN THE LAMP. THE EFFECTS OF ELECTRICITY WHICH CAN BE SEEN ARE MANY. HERE ARE SOME MORE:

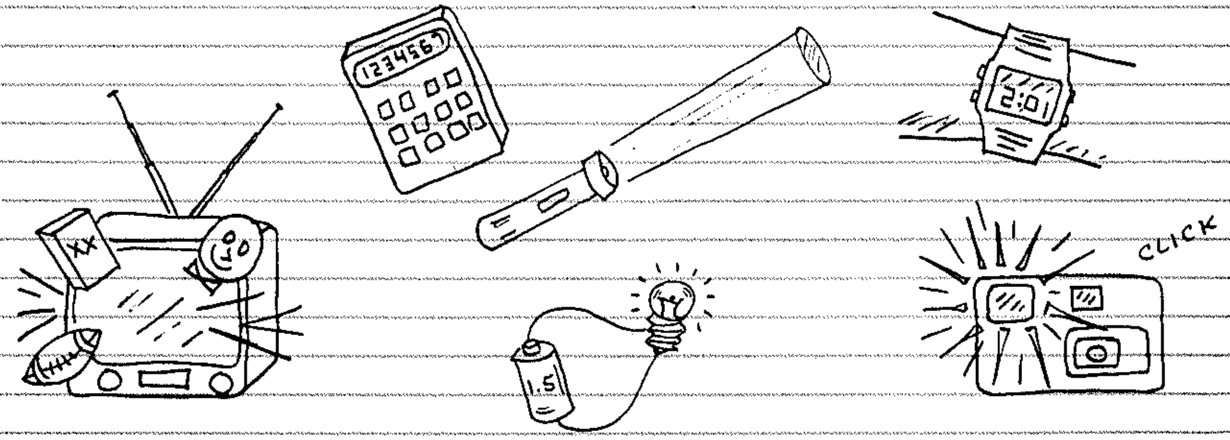

PUTTING ELECTRICITY TO WORK

ALL MATTER HAS ELECTRICAL PROPERTIES. THAT'S WHY SCIENTISTS OVER THE PAST FEW CENTURIES HAVE BEEN ABLE TO INVENT HUNDREDS OF GADGETS THAT GENERATE, STORE, CONTROL AND SWITCH ELECTRICITY. THESE DEVICES HAVE COMBINED TO CARRY US INTO...

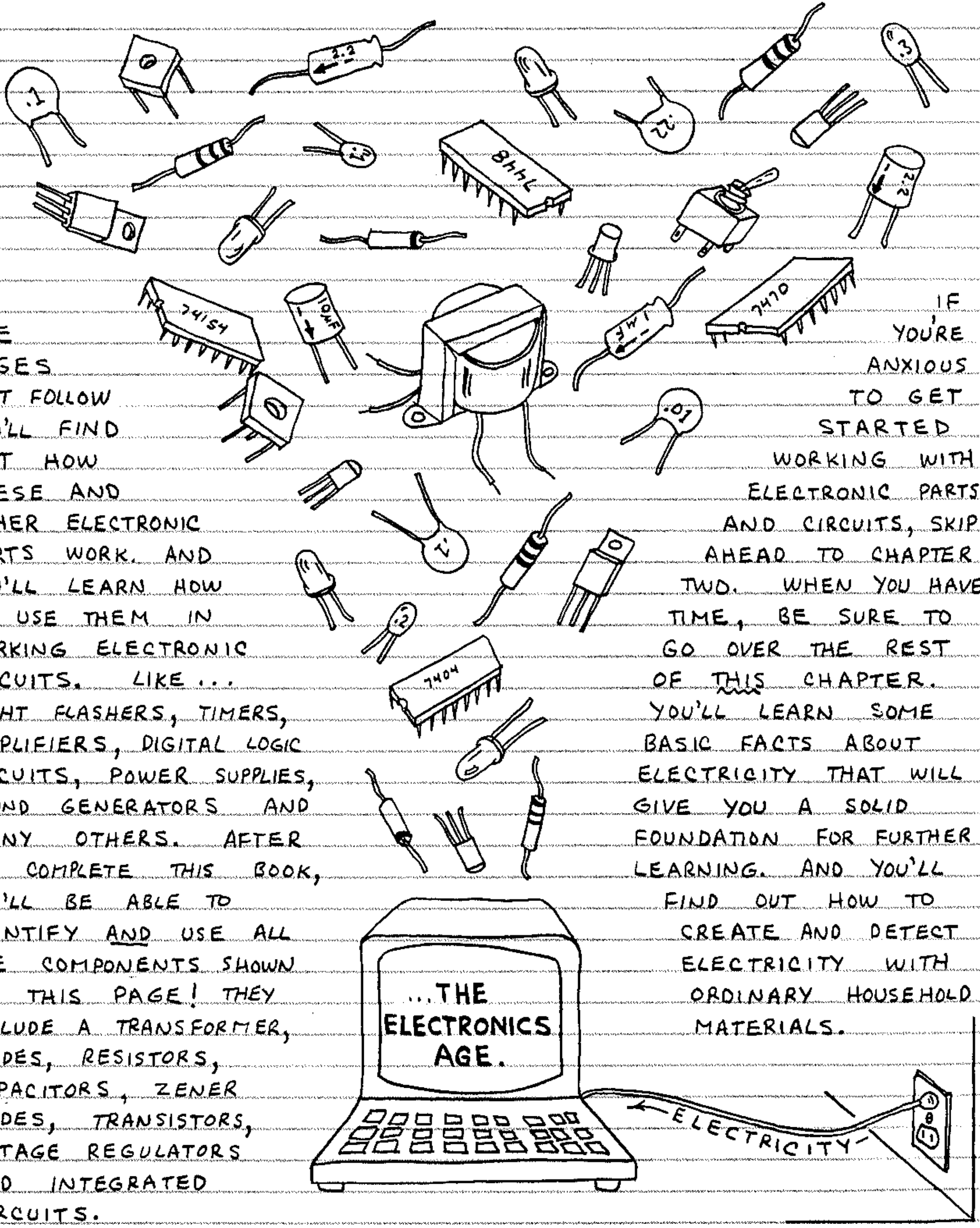

IN THE PAGES THAT FOLLOW YOU'LL FIND OUT HOW THESE AND OTHER ELECTRONIC PARTS WORK. AND YOU'LL LEARN HOW TO USE THEM IN WORKING ELECTRONIC CIRCUITS. LIKE... LIGHT FLASHERS, TIMERS, AMPLIFIERS, DIGITAL LOGIC CIRCUITS, POWER SUPPLIES, SOUND GENERATORS AND MANY OTHERS. AFTER YOU COMPLETE THIS BOOK, YOU'LL BE ABLE TO IDENTIFY AND USE ALL THE COMPONENTS SHOWN ON THIS PAGE! THEY INCLUDE A TRANSFORMER, DIODES, RESISTORS, CAPACITORS, ZENER DIODES, TRANSISTORS, VOLTAGE REGULATORS AND INTEGRATED CIRCUITS.

IF YOU'RE ANXIOUS TO GET STARTED WORKING WITH ELECTRONIC PARTS AND CIRCUITS, SKIP AHEAD TO CHAPTER TWO. WHEN YOU HAVE TIME, BE SURE TO GO OVER THE REST OF THIS CHAPTER. YOU'LL LEARN SOME BASIC FACTS ABOUT ELECTRICITY THAT WILL GIVE YOU A SOLID FOUNDATION FOR FURTHER LEARNING. AND YOU'LL FIND OUT HOW TO CREATE AND DETECT ELECTRICITY WITH ORDINARY HOUSEHOLD MATERIALS.

BACK TO BASICS

ELECTRICITY IS AN ESSENTIAL INGREDIENT OF MATTER. THE BEST WAY TO UNDERSTAND THE NATURE OF ELECTRICITY IS TO EXAMINE THE SMALLEST COMPONENT OF EVERY ELEMENT, THE ATOM.

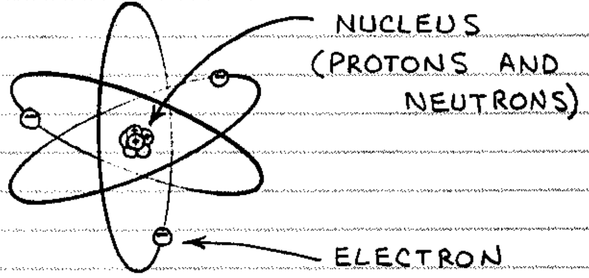

LITHIUM ATOM

THIS IS A LITHIUM ATOM. THE THIRD SIMPLEST ATOM AFTER HYDROGEN AND HELIUM, LITHIUM ATOMS HAVE 3 ELECTRONS THAT ENCIRCLE A NUCLEUS OF 3 PROTONS AND 4 NEUTRONS.

⊖ ELECTRONS HAVE A NEGATIVE ELECTRICAL CHARGE.
⊕ PROTONS HAVE A POSITIVE ELECTRICAL CHARGE.
○ NEUTRONS HAVE NO ELECTRICAL CHARGE.

☐ IONS — NORMALLY AN ATOM HAS AN EQUAL NUMBER OF ELECTRONS AND PROTONS. THE CHARGES CANCEL TO GIVE THE ATOM NO NET ELECTRICAL CHARGE. IT'S POSSIBLE TO DISLODGE ONE OR MORE ELECTRONS FROM MOST ATOMS. THIS CAUSES THE ATOM TO HAVE A NET POSITIVE CHARGE. IT'S THEN CALLED A POSITIVE ION. IF A STRAY ELECTRON COMBINES WITH A NORMAL ATOM, THE ATOM HAS A NET NEGATIVE CHARGE AND IS CALLED A NEGATIVE ION.

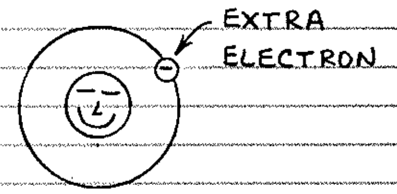

NEGATIVE ION

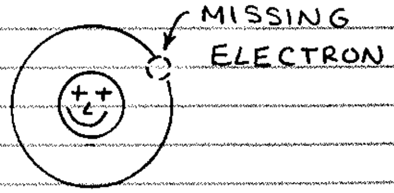

POSITIVE ION

☐ ELECTRONS — FREE ELECTRONS CAN MOVE AT HIGH SPEED THROUGH METALS, GASES AND A VACUUM. OR THEY CAN REST ON A SURFACE.

☐ MORE ABOUT FREE ELECTRONS — MANY TRILLIONS OF ELECTRONS CAN REST ON A SURFACE OR TRAVEL THROUGH SPACE OR MATTER AT NEAR THE SPEED OF LIGHT (186,000 MILES PER SECOND)!

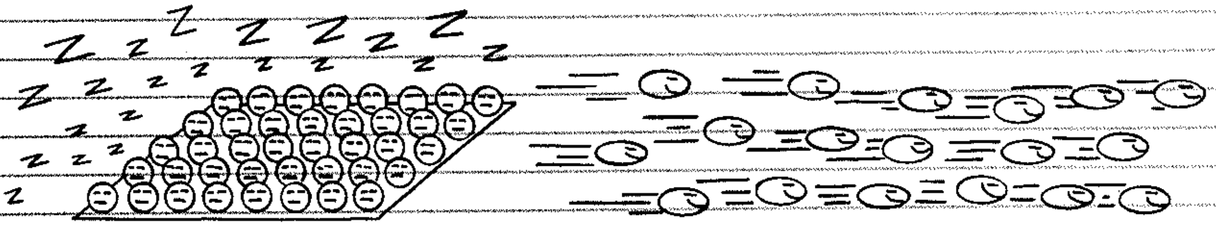

RESTING ELECTRONS MOVING ELECTRONS

☐ RESTING ELECTRONS — A GROUP OF NEGATIVE ELECTRONS ON A SURFACE CAUSES THE SURFACE TO BE NEGATIVELY CHARGED. SINCE THE ELECTRONS ARE NOT MOVING, THE SURFACE CAN BE SAID TO HAVE A NEGATIVE STATIC ELECTRICAL CHARGE.

☐ MOVING ELECTRONS — A STREAM OF MOVING ELECTRONS IS CALLED AN ELECTRICAL CURRENT. RESTING ELECTRONS CAN QUICKLY FORM AN ELECTRICAL CURRENT IF PLACED NEAR A CLUSTER OF POSITIVE IONS. THE POSITIVELY CHARGED IONS WILL ATTRACT THE ELECTRONS WHICH WILL RUSH IN TO FILL THE "HOLES" OR VOIDS LEFT BY THE MISSING ELECTRONS.

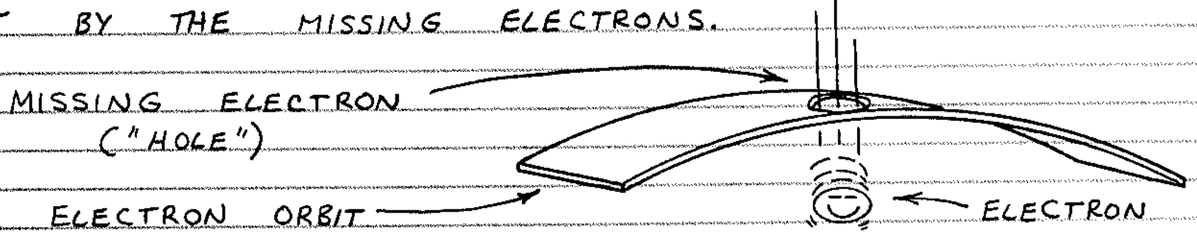
MISSING ELECTRON ("HOLE")
ELECTRON ORBIT
ELECTRON

☐ MISSING ELECTRONS — MECHANICAL FRICTION, LIGHT, HEAT OR A CHEMICAL REACTION MAY REMOVE ELECTRONS FROM A SURFACE. THIS CAUSES THE SURFACE TO BE POSITIVELY CHARGED. SINCE THE POSITIVELY CHARGED ATOMS ARE AT REST, THE SURFACE CAN BE SAID TO HAVE A POSITIVE STATIC ELECTRICAL CHARGE.

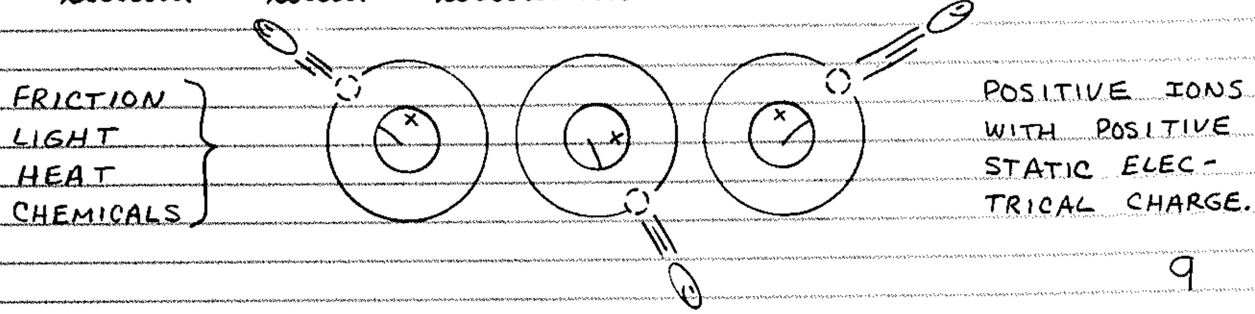
FRICTION
LIGHT
HEAT
CHEMICALS
POSITIVE IONS WITH POSITIVE STATIC ELECTRICAL CHARGE.

STATIC ELECTRICITY

YOU GENERATE STATIC ELECTRICITY EVERY TIME YOU WALK ACROSS A CARPET, PULL TAPE FROM A ROLL, REMOVE YOUR CLOTHING OR DRY CLOTHES IN A DRIER. MUCH OF THE TIME YOU DON'T EVEN REALIZE IT UNLESS THE AIR IS DRY AND THE STATIC CHARGE SUDDENLY CRACKLES, POPS AND FLASHES ITS WAY TO A NEW HOME. THESE STATIC CHARGES ARE CAUSED BY MECHANICAL FRICTION. BACK IN 600 B.C., THALES OF GREECE EXPERIMENTED WITH THE STATIC ELECTRICITY PRODUCED WHEN AMBER IS RUBBED WITH WOOL.

☐ AMBER — ONCE UPON A TIME SAP FLOWING FROM TREES HARDENED INTO CLEAR GOLDEN NODULES WHICH WERE EVENTUALLY BURIED IN THE EARTH. SOMETIMES, BEFORE IT HARDENED INTO AMBER, THE STICKY SAP ENTOMBED BITS OF PLANT MATTER, INSECTS AND EVEN DROPLETS OF WATER! A KIND OF NATURAL CASTING PLASTIC, AMBER IS EASILY ELECTRIFIED BY FRICTION. IT THEN ATTRACTS BITS OF PAPER.

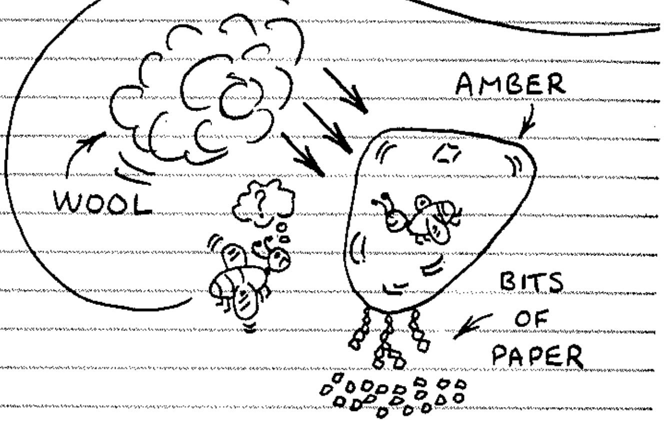

FAMOUS FACT: THE ELECTRON IS NAMED AFTER THE GREEK WORD FOR AMBER!

☐ ELECTRIFIED PLASTIC AND GLASS — RUN A PLASTIC COMB THROUGH YOUR HAIR ON A DRY DAY AND YOU'LL TRANSFER ELECTRONS FROM YOUR HAIR TO THE COMB. RUB A GLASS ROD WITH SILK OR THE SYNTHETIC FIBERS OF A PAINT BRUSH AND YOU'LL REMOVE ELECTRONS FROM THE GLASS. BOTH THE NEGATIVELY CHARGED COMB AND THE POSITIVELY CHARGED GLASS ROD WILL, LIKE AMBER, ATTRACT BITS OF PAPER. YOU CAN ELECTRIFY OR CHARGE MANY MATERIALS BY RUBBING THEM WITH FUR, WOOL, ETC. METAL? NO, THE CHARGE LEAKS AWAY.

COMB (AFTER STROKING HAIR) GLASS ROD (RUBBED WITH SILK)

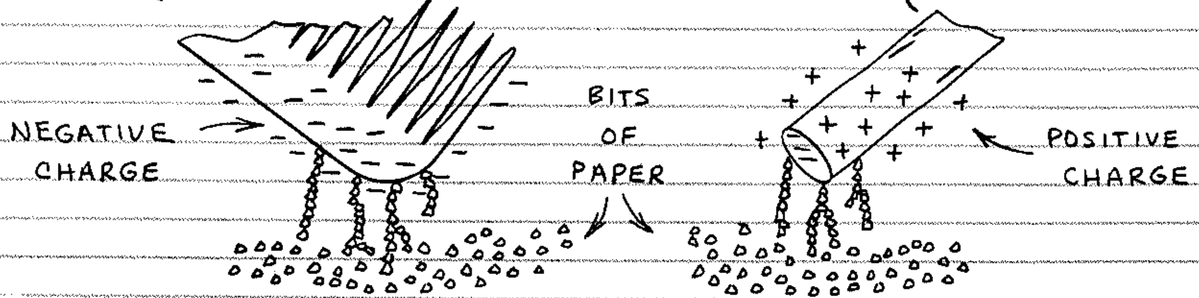

☐ OPPOSITE AND LIKE CHARGES — HOW DO WE KNOW THE COMB AND GLASS ROD HAVE OPPOSITE CHARGES? A FUNDAMENTAL RULE OF ELECTRICITY IS LIKE CHARGES REPEL AND UNLIKE CHARGES ATTRACT. HERE'S AN EXPERIMENT THAT PROVES THE RULE AND ANSWERS THE QUESTION:

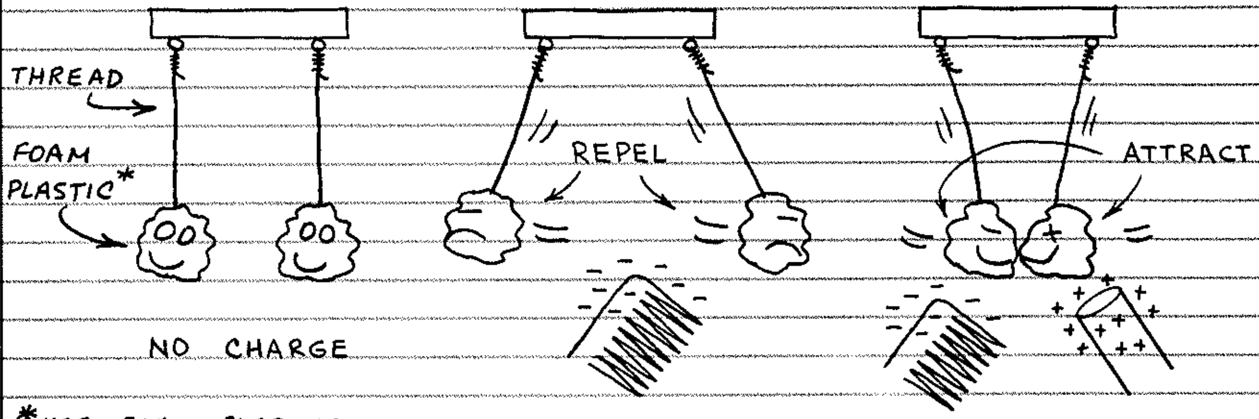
THREAD
FOAM PLASTIC*
NO CHARGE
REPEL
ATTRACT

*USE FOAM PLASTIC PACKING MATERIAL.

REMEMBER: UNLIKE CHARGES ATTRACT — 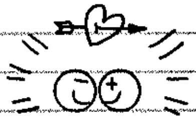

LIKE CHARGES REPEL —

☐ THE ELECTROSCOPE — THE FIRST INSTRUMENT DESIGNED TO DETECT AND MEASURE STATIC ELECTRICITY WAS THE ELECTROSCOPE. YOU CAN EASILY MAKE ONE.

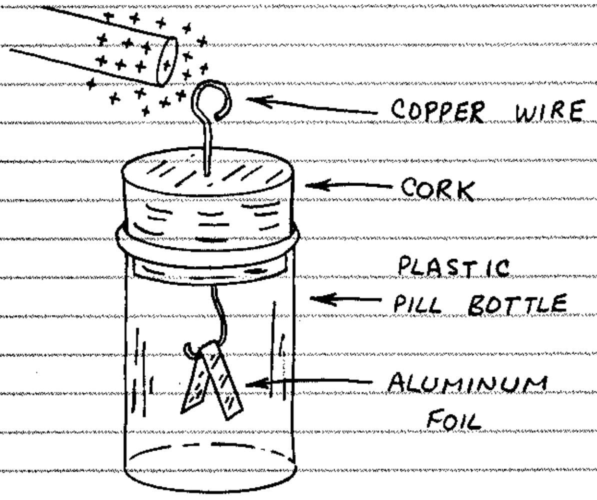
COPPER WIRE
CORK
PLASTIC PILL BOTTLE
ALUMINUM FOIL

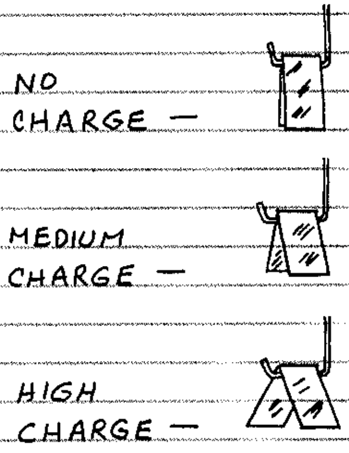
NO CHARGE —
MEDIUM CHARGE —
HIGH CHARGE —

BE SURE THE FOLDED FOIL STRIP IS CLEAN AND DRY. WHEN YOU TOUCH A CHARGED OBJECT TO THE WIRE, THE TWO HALVES OF THE FOIL STRIP WILL BE GIVEN THE SAME CHARGE AND WILL THEREFORE FLY APART.

☐ CONDUCTORS AND INSULATORS— YOU CAN USE YOUR ELECTROSCOPE TO PROVE THAT ELECTRONS TRAVEL THROUGH SOME MATERIALS BUT NOT OTHERS. HINT: TRY THIS ON A DRY DAY! ELECTRONS CAN TRAVEL THROUGH MOIST AIR SO THE CHARGE ON YOUR ELECTROSCOPE WILL QUICKLY LEAK AWAY ON HUMID DAYS.

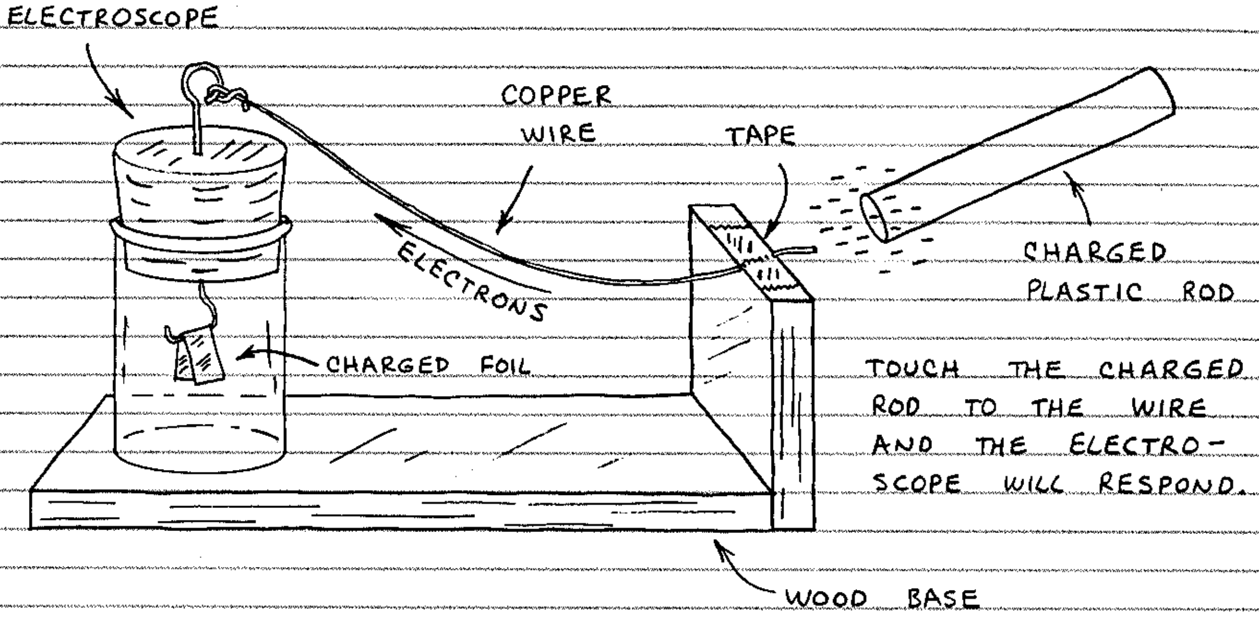
ELECTROSCOPE
COPPER WIRE
TAPE
ELECTRONS
CHARGED PLASTIC ROD
CHARGED FOIL
WOOD BASE

TOUCH THE CHARGED ROD TO THE WIRE AND THE ELECTROSCOPE WILL RESPOND.

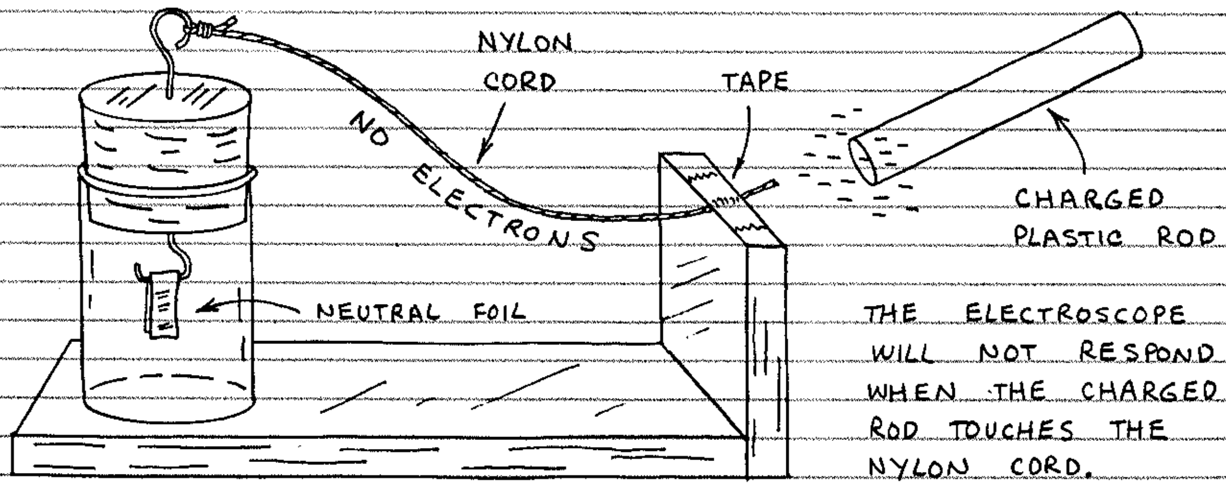
NYLON CORD
TAPE
NO ELECTRONS
CHARGED PLASTIC ROD
NEUTRAL FOIL

THE ELECTROSCOPE WILL NOT RESPOND WHEN THE CHARGED ROD TOUCHES THE NYLON CORD.

THIS DEMONSTRATION SHOWS THAT ELECTRONS CAN TRAVEL THROUGH SOME MATERIALS BUT NOT OTHERS. MATERIALS THROUGH WHICH ELECTRONS TRAVEL ARE CONDUCTORS. MATERIALS THROUGH WHICH ELECTRONS TRAVEL POORLY OR NOT AT ALL ARE INSULATORS.

CONDUCTORS INCLUDE SILVER, GOLD, IRON, COPPER, ETC.

INSULATORS INCLUDE GLASS, PLASTIC, RUBBER, WOOD, ETC.

ELECTRICAL CURRENT

THE CONDUCTOR — INSULATOR DEMONSTATION ILLUSTRATES TWO OTHER IMPORTANT POINTS :

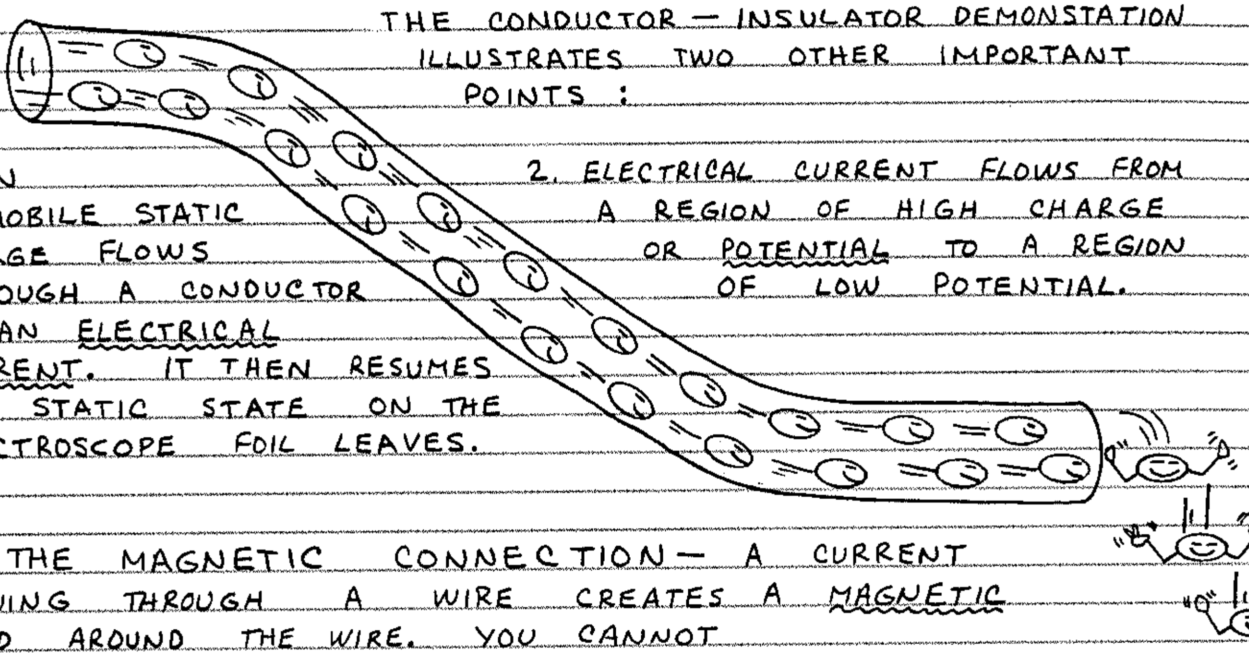

1. AN IMMOBILE STATIC CHARGE FLOWS THROUGH A CONDUCTOR AS AN ELECTRICAL CURRENT. IT THEN RESUMES ITS STATIC STATE ON THE ELECTROSCOPE FOIL LEAVES.

2. ELECTRICAL CURRENT FLOWS FROM A REGION OF HIGH CHARGE OR POTENTIAL TO A REGION OF LOW POTENTIAL.

☐ THE MAGNETIC CONNECTION — A CURRENT FLOWING THROUGH A WIRE CREATES A MAGNETIC FIELD AROUND THE WIRE. YOU CANNOT SEE THE FIELD, BUT YOU CAN OBSERVE ITS EFFECT. ORIENT A COMPASS SO ITS NEEDLE POINTS TO THE NORTH (N) MARK. PLACE A COPPER WIRE OVER AND PARALLEL TO THE NEEDLE. THEN CONNECT A FLASHLIGHT CELL ACROSS THE WIRE AND THE NEEDLE WILL MOVE AWAY FROM ITS NORTH-SOUTH ORIENTATION. (LEAVE THE WIRE CONNECTED FOR ONLY AN INSTANT TO PREVENT THE CELL FROM OVERHEATING!)

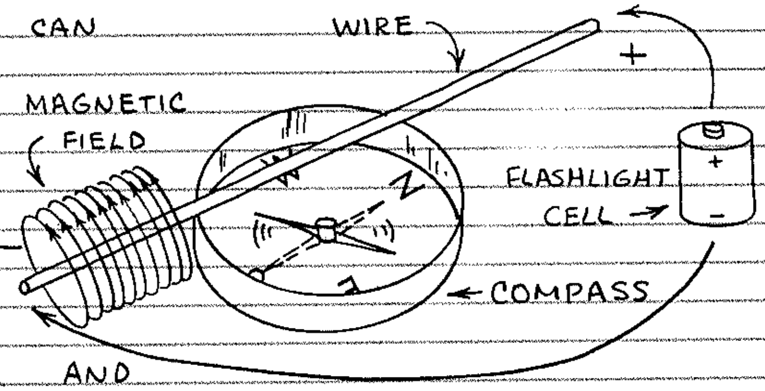

☐ MEASURING CURRENT ELECTRICITY —

THE PHYSICAL (OR MECHANICAL) MOTION OF A MAGNETIC COMPASS NEEDLE IN A MAGNETIC FIELD PROVIDES A CONVENIENT WAY TO MEASURE THE QUANTITY OF CURRENT FLOWING IN A WIRE. THIS IS THE BASIS OF THE MOVING COIL CURRENT METER USED IN THE ANALOG MULTIMETER. TO PROVIDE HIGH SENSITIVITY, THE WIRE IS WRAPPED AS A COIL.

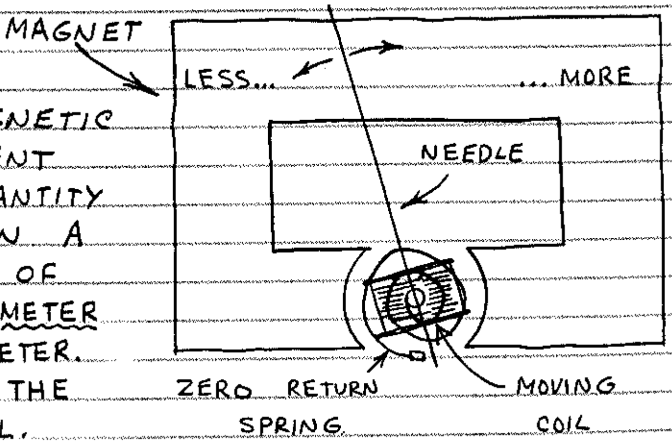

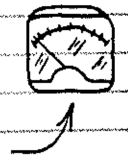

P. 26

DIRECT CURRENT ELECTRICITY

AN ELECTRICAL CURRENT CAN FLOW IN EITHER OF TWO DIRECTIONS THROUGH A CONDUCTOR. IF IT FLOWS IN ONLY ONE DIRECTION, WHETHER STEADILY OR IN PULSES, IT'S CALLED *DIRECT CURRENT* (DC). IT'S IMPORTANT TO BE ABLE TO SPECIFY THE QUANTITY AND POWER OF A DIRECT CURRENT. HERE ARE THE KEY TERMS:

☐ CURRENT (I) — CURRENT IS THE QUANTITY OF ELECTRONS PASSING A GIVEN POINT. THE UNIT OF CURRENT IS THE *AMPERE*. ONE AMPERE IS 6,280,000,000,000,000,000 (6.28×10^{18}) ELECTRONS PASSING A POINT IN ONE SECOND.

☐ VOLTAGE (V OR E) — VOLTAGE IS ELECTRICAL PRESSURE OR FORCE. VOLTAGE IS SOMETIMES REFERRED TO AS *POTENTIAL*. *VOLTAGE DROP* IS THE DIFFERENCE IN VOLTAGE BETWEEN THE TWO ENDS OF A CONDUCTOR THROUGH WHICH CURRENT IS FLOWING. IF WE COMPARE CURRENT TO WATER FLOWING THROUGH A PIPE, THEN VOLTAGE IS THE WATER PRESSURE.

☐ POWER (P) — THE WORK PERFORMED BY AN ELECTRICAL CURRENT IS CALLED POWER. THE UNIT OF POWER IS THE *WATT*. THE POWER OF A DIRECT CURRENT IS ITS VOLTAGE TIMES ITS CURRENT.

☐ RESISTANCE (R) — CONDUCTORS ARE NOT PERFECT. THEY RESIST TO SOME DEGREE THE FLOW OF CURRENT. THE UNIT OF RESISTANCE IS THE *OHM* (Ω). A POTENTIAL DIFFERENCE OF ONE VOLT WILL FORCE A CURRENT OF ONE AMPERE THROUGH A RESISTANCE OF ONE OHM. THE RESISTANCE OF A CONDUCTOR IS ITS VOLTAGE DROP DIVIDED BY THE CURRENT FLOWING THROUGH THE CONDUCTOR.

☐ MR. OHM'S LAW — GIVEN ANY TWO OF THE ABOVE, YOU CAN FIND THE OTHER TWO USING THESE FORMULAS KNOWN AS *OHM'S LAW*:

$$V = I \times R$$
$$I = V / R$$
$$R = V / I$$
$$P = V \times I \text{ (OR) } I^2 \times R$$

WE'LL REFER TO OHM'S LAW LATER IN THIS BOOK...

☐ SUMMING UP — THIS IS THE "WATER ANALOGY":

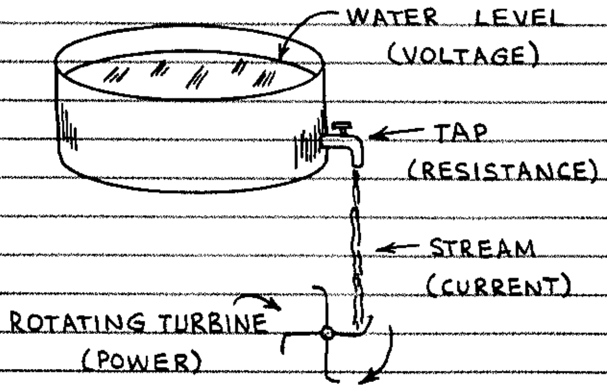

USING DIRECT CURRENT

THERE ARE SO MANY USES FOR DIRECT CURRENT ELECTRICITY NO SINGLE BOOK CAN DESCRIBE THEM ALL. HERE'S A PAGE OF SEVERAL DESIGNED AROUND A SINGLE WIRE COIL YOU CAN EASILY MAKE FROM A 1-½ TO 3-INCH SECTION OF A SODA STRAW AND AT LEAST 30-FEET OF 30 GAUGE, LACQUER COATED WIRE. SECURE THE COIL IN PLACE WITH TAPE. REMOVE INSULATION FROM ENDS OF COIL WITH FINE SAND PAPER.

(TWICE ACTUAL SIZE)

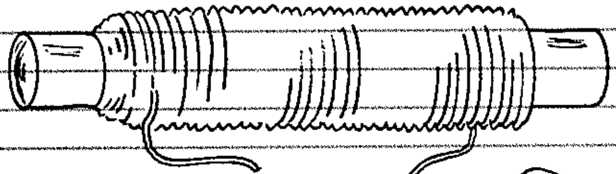

☐ ELECTROMAGNET — INSERT A STEEL NAIL IN THE COIL, CONNECT THE LEADS TO A 9-VOLT BATTERY, AND THE NAIL WILL BECOME A MAGNET UNTIL THE POWER IS DISCONNECTED. (IT MAY RETAIN SOME MAGNETISM.)

☐ SOLENOID — THIS IS A "SUCKING MAGNET." APPLY POWER TO COIL AND NAIL WILL BE PULLED RAPIDLY INSIDE.

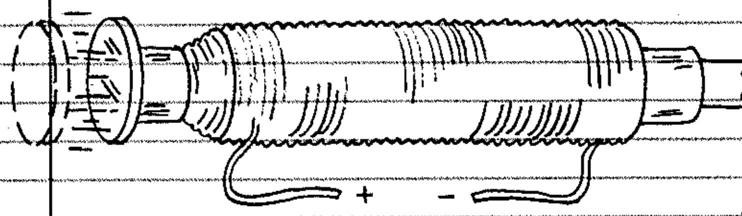

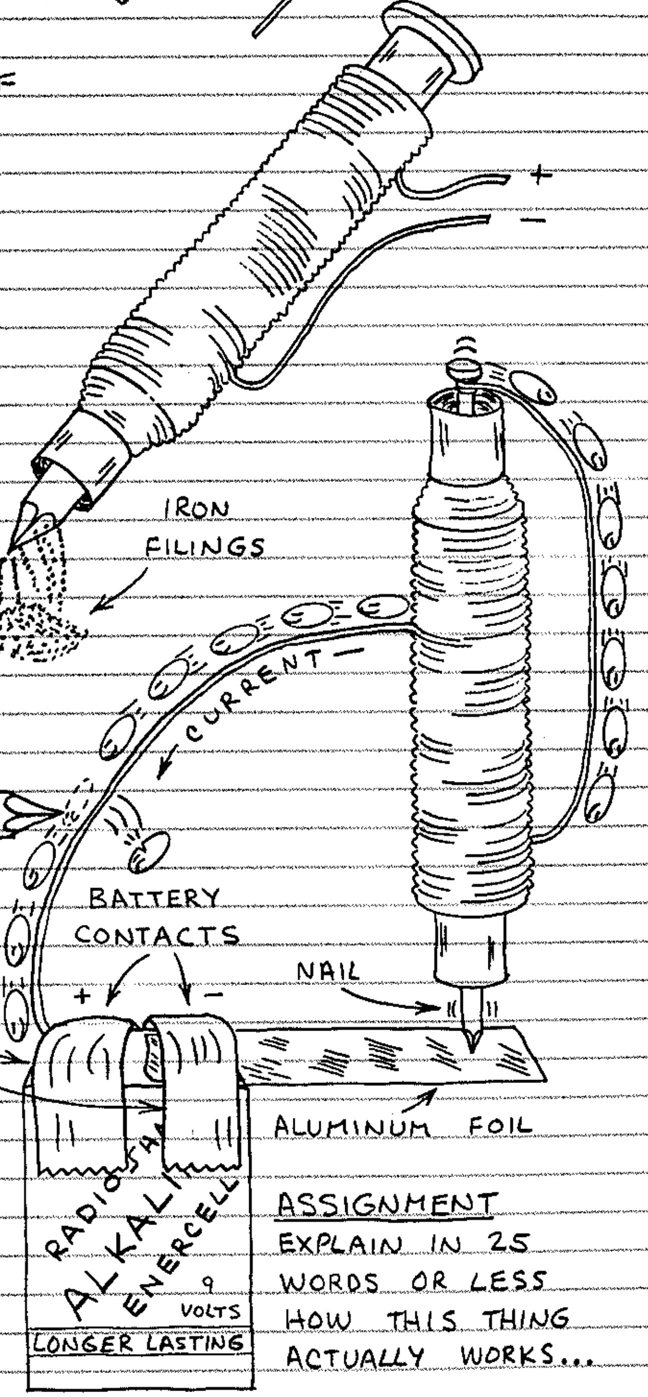
+
−
IRON FILINGS
CURRENT
BATTERY CONTACTS
+
−
TAPE
NAIL
ALUMINUM FOIL
RADIOSHA
ALKALI
ENERCELL
9 VOLTS
LONGER LASTING

☐ MOTOR — MAYBE NOT YOUR IDEA OF A MOTOR, BUT THIS ELEGANT APPARATUS QUALIFIES UNDER THE DICTIONARY DEFINITION. USE A LIGHT WEIGHT NAIL. SECURE ONE COIL LEAD TO NAIL. ADJUST HEIGHT OF COIL UNTIL NAIL JUMPS UP AND DOWN.

ASSIGNMENT

EXPLAIN IN 25 WORDS OR LESS HOW THIS THING ACTUALLY WORKS...

MAKING DIRECT CURRENT ELECTRICITY

A SURPRISING NUMBER OF WAYS EXIST FOR PRODUCING DIRECT CURRENT. HERE ARE THE BIGGIES:

☐ CHEMICAL GENERATORS — ELECTROLYTES ARE CHEMICAL SOLUTIONS THAT CONTAIN MANY IONS. FOR EXAMPLE, DISSOLVE TABLE SALT IN WATER AND THE SALT WILL BREAK DOWN INTO POSITIVE SODIUM IONS AND NEGATIVE CHLORINE IONS. IF TWO DISSIMILAR METAL PLATES ARE IMMERSED IN THE SALT SOLUTION, THE POSITIVE IONS WILL MIGRATE TOWARD ONE PLATE AND THE NEGATIVE IONS WILL MIGRATE TOWARD THE OTHER. IF THE TWO PLATES ARE CONNECTED TOGETHER BY A CONDUCTOR, A CURRENT WILL FLOW THROUGH THE SOLUTION (AS IONS) AND THE CONDUCTOR (AS ELECTRONS). THIS KIND OF GENERATOR IS CALLED A WET CELL. CELLS IN WHICH THE ELECTROLYTE IS ABSORBED BY PAPER OR FORMED INTO A PASTE ARE CALLED DRY CELLS. HERE ARE SOME CHEMICAL GENERATORS YOU CAN MAKE. HAVE FUN!

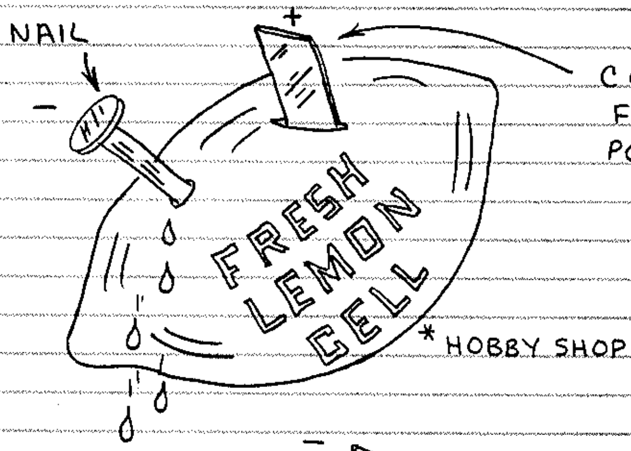
NAIL
+
−
FRESH LEMON CELL
* HOBBY SHOP

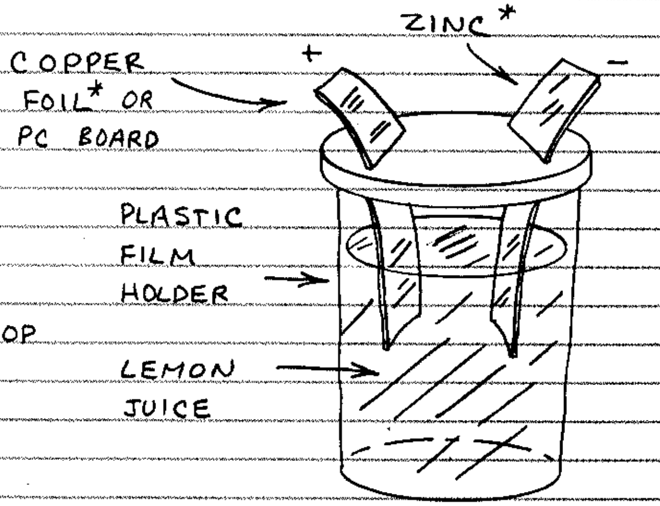
COPPER FOIL* OR PC BOARD
ZINC*
+
−
PLASTIC FILM HOLDER
LEMON JUICE

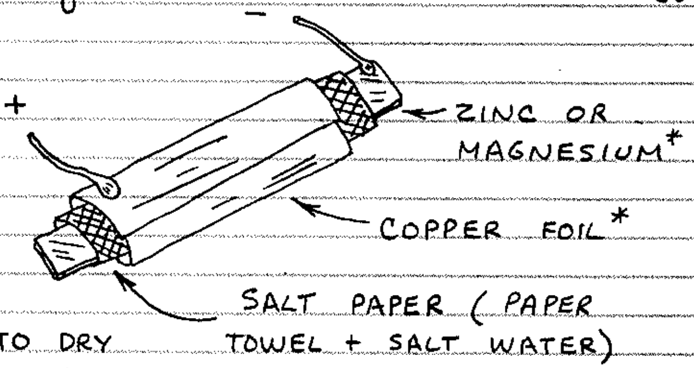
−
+
ZINC OR MAGNESIUM*
COPPER FOIL*
SALT PAPER (PAPER TOWEL + SALT WATER)

OK TO DRY PAPER AND ACTIVATE WITH WATER...

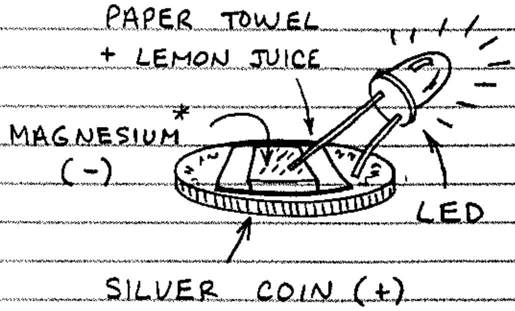
PAPER TOWEL + LEMON JUICE
MAGNESIUM* (−)
LED
SILVER COIN (+)

CONNECT TWO OR MORE CELLS IN SERIES TO FORM A BATTERY WITH TOTAL VOLTAGE EQUAL TO SUM OF CELL VOLTAGES.

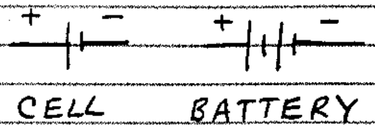
+
−
+
−
CELL
BATTERY

☐ ELECTROMAGNETIC GENERATORS— A CURRENT FLOWING THROUGH A CONDUCTOR ESTABLISHES A MAGNETIC FIELD AROUND THE CONDUCTOR. THIS EFFECT WORKS BOTH WAYS SO THAT A CURRENT WILL FLOW IN A CONDUCTOR WHICH IS MOVED THROUGH A MAGNETIC FIELD. YOU CAN EASILY DEMONSTRATE ELECTROMAGNETIC CURRENT GENERATION WITH A COIL OF WIRE AND A SMALL MAGNET. (THE COIL SHOWN ON PAGE 15 WORKS FINE.) CONNECT THE LEADS OF THE COIL TO A METER DESIGNED TO SENSE MICROAMPERES. INSERT A STEEL NAIL THROUGH THE COIL AND STROKE THE MAGNET BACK AND FORTH ACROSS THE COIL. THE METER WILL INDICATE A FEW MICROAMPERES EACH STROKE. THE POLARITY (DIRECTION) OF THE CURRENT WILL REVERSE ON THE BACK STROKES. WANT A READY-MADE GENERATOR? JUST ROTATE THE SHAFT OF A SMALL DC MOTOR. MOST SUCH MOTORS WILL PRODUCE A POTENTIAL DIFFERENCE OF UP TO SEVERAL VOLTS! YOU CAN ADD A PROPELLER TO MAKE A WIND POWERED GENERATOR.

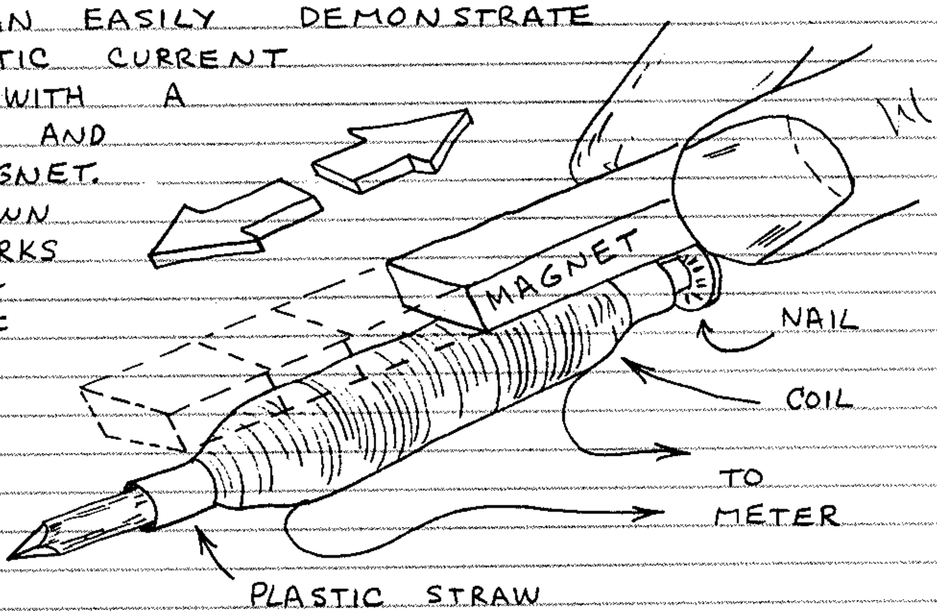

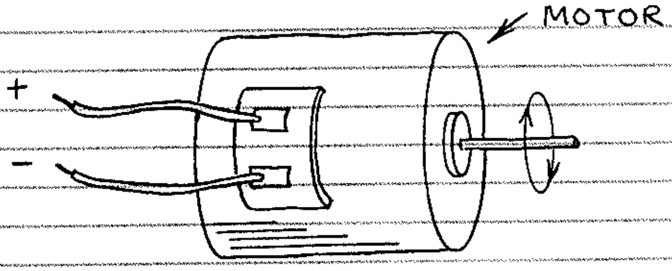

☐ THERMOELECTRIC GENERATORS— IF A JUNCTION OF TWO DISSIMILAR METALS IS HEATED, A CURRENT WILL BE PRODUCED. COPPER WIRE WRAPPED AROUND THE END OF A STEEL NAIL WILL GENERATE A FEW THOUSANDTHS OF A VOLT WHEN HEATED BY THE FLAME OF A MATCH. JUNCTIONS LIKE IRON AND CONSTANTAN PRODUCE MUCH HIGHER VOLTAGES. (THIS IS THE SEEBECK EFFECT.)

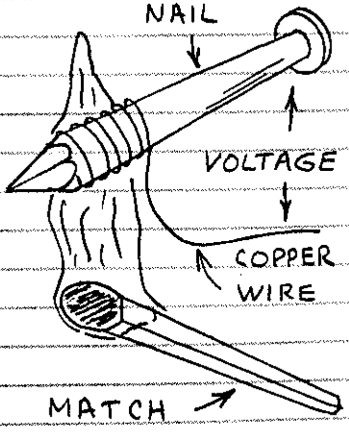

ALTERNATING CURRENT ELECTRICITY

LOOK BACK AT THE HOMEMADE COIL AND MAGNET "GENERATOR" ON THE PRECEEDING PAGE. WHEN THE MAGNET IS STROKED IN ONE DIRECTION ALONG THE COIL, ELECTRONS IN THE WIRE ARE MOVED IN ONE DIRECTION AND A DIRECT CURRENT IS PRODUCED. ON THE BACK STROKE, UNLESS THE MAGNET IS MOVED AWAY FROM THE COIL, THE DIRECTION OF CURRENT FLOW IS REVERSED. THEREFORE, IF THE MAGNET IS STROKED BACK AND FORTH ALONG THE COIL, A CURRENT WHICH ALTERNATES IN DIRECTION OR POLARITY IS PRODUCED. IT'S CALLED AN ALTERNATING CURRENT. ALTERNATING CURRENT (AC) IS USUALLY PRODUCED BY ROTATING A COIL IN A MAGNETIC FIELD.

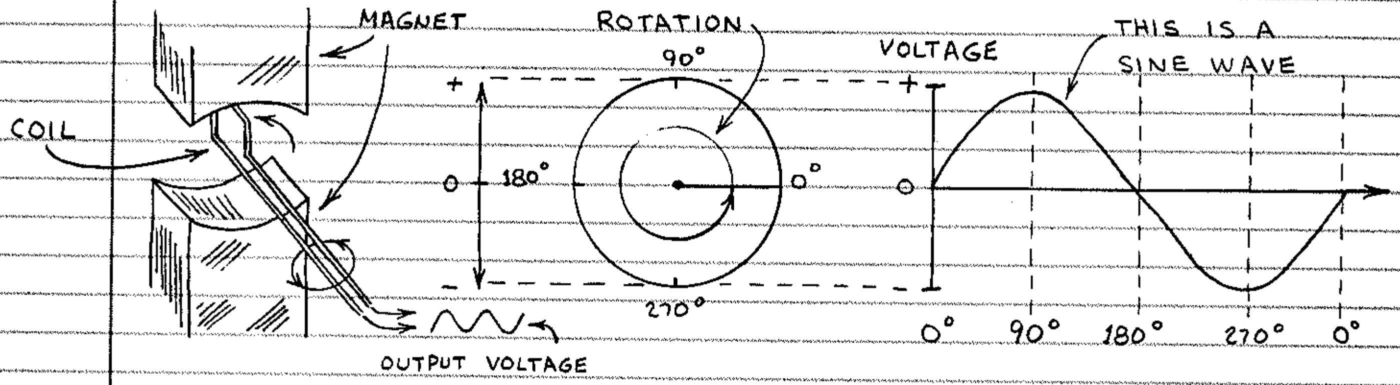

ROTATING COIL VOLTAGE OUTPUT AC SINE WAVE

☐ SINE WAVE MEASUREMENT — AC VOLTAGE IS USUALLY SPECIFIED AT A VALUE EQUAL TO THE DC VOLTAGE CAPABLE OF DOING THE SAME WORK. FOR A SINE WAVE THIS VALUE IS 0.707 TIMES THE PEAK VOLTAGE. IT'S CALLED THE RMS (ROOT-MEAN-SQUARE) VOLTAGE. THE PEAK VOLTAGE (OR CURRENT) IS 1.41 TIMES THE RMS VALUE. HOUSEHOLD LINE VOLTAGE IS SPECIFIED ACCORDING TO ITS RMS VALUE. THEREFORE, A HOUSEHOLD VOLTAGE OF 120-VOLTS CORRESPONDS TO A PEAK VOLTAGE OF 120 × 1.41 OR 169.2-VOLTS.

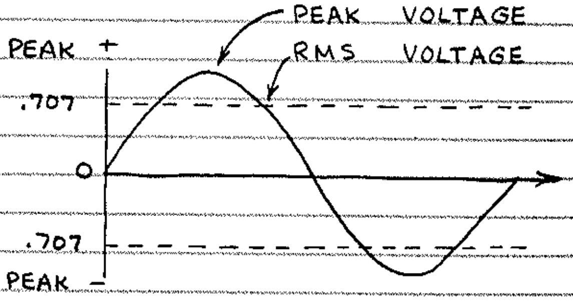

☐ WHY AC IS USED — AC IS BETTER SUITED THAN DC FOR TRANSMISSION THROUGH LONG DISTANCE POWER LINES. A WIRE CARRYING AC WILL INDUCE A CURRENT IN A NEARBY WIRE. THIS IS THE PRINCIPLE BEHIND THE TRANSFORMER.

MEASURING AC AND DC

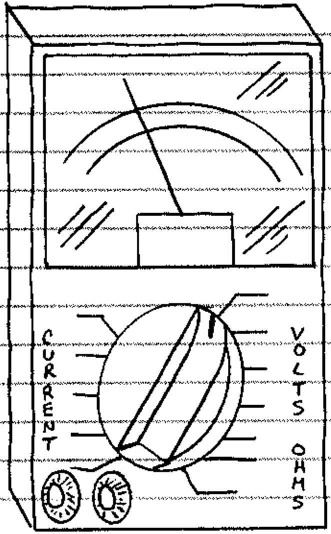

YOU CAN EASILY MEASURE AC AND DC VOLTAGE AND CURRENT WITH AN INSTRUMENT CALLED THE MULTIMETER. ANALOG MULTIMETERS USE A MOVING COIL METER. DIGITAL MULTIMETERS HAVE A DIGITAL READOUT. THE MULTIMETER IS THE SINGLE MOST IMPORTANT ELECTRONIC TEST INSTRUMENT.

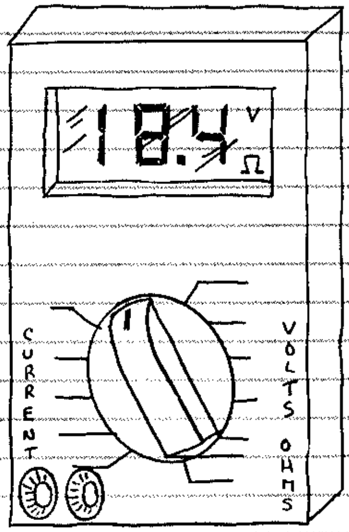

☐ ANALOG MULTIMETER — LESS EXPENSIVE, SOMEWHAT LESS PRECISE THAN DIGITAL TYPES. BEST BY FAR FOR OBSERVING THE TREND OF A SLOWLY CHANGING VOLTAGE, CURRENT OR RESISTANCE.

☐ DIGITAL MULTIMETER — HIGHLY ACCURATE AND EASIER TO READ THAN ANALOG TYPES. BEST FOR FINDING THE PRECISE VALUE OF A VOLTAGE, CURRENT OR RESISTANCE.

☐ SUMMING UP MULTIMETERS — THEY'RE INDISPENSABLE! EVEN IF YOU HAVE ONLY A PASSING INTEREST YOU SHOULD CONSIDER BUYING ONE BECAUSE IT HAS MANY USES IN THE HOME, ON THE JOB AND WHEN WORKING WITH APPLIANCES AND MOTOR VEHICLES. IF YOU'RE SERIOUS ABOUT ELECTRONICS, CONSIDER BUYING A QUALITY HIGH-IMPEDANCE MULTIMETER THAT WILL HAVE LITTLE OR NO EFFECT ON THE DEVICE OR CIRCUIT YOU'RE MEASURING. IDEALLY, YOU SHOULD HAVE BOTH THE ANALOG AND DIGITAL TYPES.

ELECTRICAL SAFETY

ELECTRICITY CAN KILL! IF YOU WANT TO BE AROUND LONG ENOUGH TO ENJOY EXPERIMENTING WITH ELECTRONICS, ALWAYS TREAT ELECTRICITY WITH THE RESPECT IT DESERVES. WE'LL LOOK AT SAFETY AGAIN LATER.

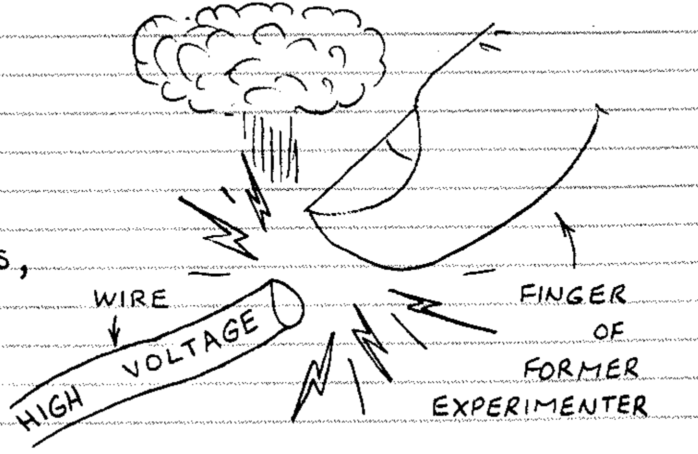

ELECTRICAL CIRCUITS

AN ELECTRICAL CIRCUIT IS ANY ARRANGEMENT THAT PERMITS AN ELECTRICAL CURRENT TO FLOW. A CIRCUIT CAN BE AS SIMPLE AS A BATTERY CONNECTED TO A LAMP OR AS COMPLICATED AS A DIGITAL COMPUTER.

☐ A BASIC CIRCUIT — THIS BASIC CIRCUIT CONSISTS OF A SOURCE OF ELECTRICAL CURRENT (A BATTERY), A LAMP AND TWO CONNECTION WIRES. THE PART OF A CIRCUIT WHICH PERFORMS WORK IS CALLED THE LOAD. HERE THE LOAD IS THE LAMP. IN OTHER CIRCUITS THE LOAD CAN BE A MOTOR, A HEATING ELEMENT, AN ELECTROMAGNET, ETC.

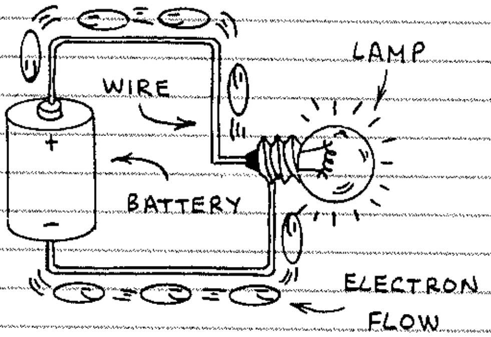

☐ A SERIES CIRCUIT — A CIRCUIT MAY INCLUDE MORE THAN ONE COMPONENT (SWITCH, LAMP, MOTOR, ETC.). A SERIES CIRCUIT IS FORMED WHEN CURRENT FLOWING THROUGH ONE COMPONENT FIRST FLOWS THROUGH ANOTHER. (ARROWS SHOW DIRECTION OF ELECTRON FLOW.)

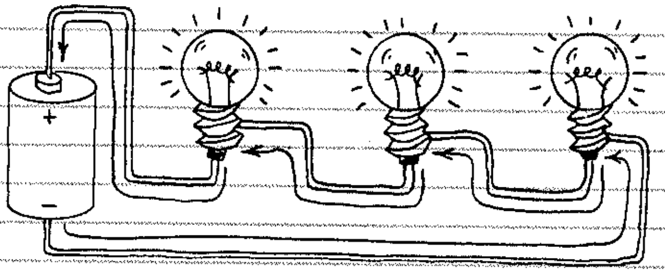

☐ A PARALLEL CIRCUIT — A PARALLEL CIRCUIT IS FORMED WHEN TWO OR MORE COMPONENTS ARE CONNECTED SO CURRENT CAN FLOW THROUGH ONE COMPONENT WITHOUT HAVING FIRST TO FLOW THROUGH ANOTHER.

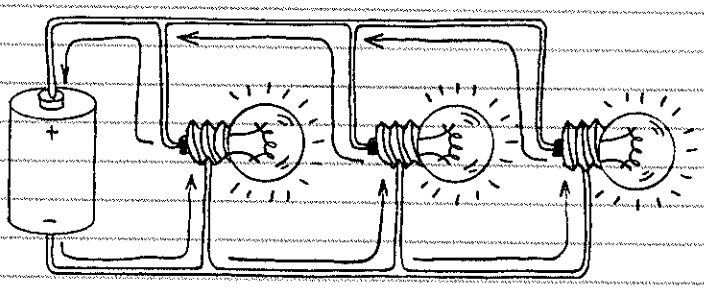

☐ A SERIES-PARALLEL CIRCUIT — MANY ELECTRICAL CIRCUITS ARE BOTH SERIES AND PARALLEL. ALL PROVIDE A COMPLETE PATH BETWEEN THE CIRCUIT AND ITS POWER SUPPLY.

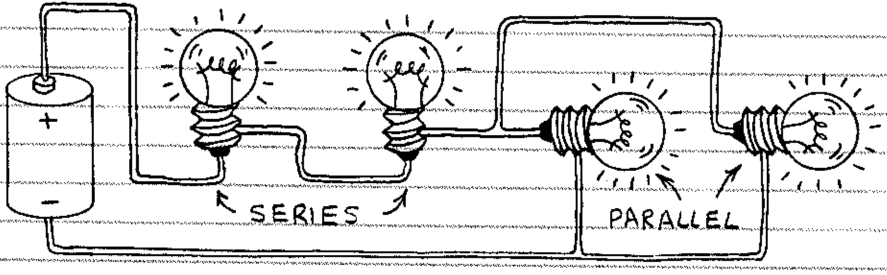

☐ CIRCUIT DIAGRAMS— THUS FAR THE ELECTRICAL CIRCUITS SHOWN IN THIS BOOK HAVE BEEN ILLUSTRATED IN PICTORIAL FORM. PICTORIAL VERSIONS OF CIRCUITS WILL BE USED IN THE NEXT SEVERAL CHAPTERS AS WELL. LATER IN THE BOOK THE PICTORIALS WILL BE REPLACED BY CIRCUIT DIAGRAMS. IN A CIRCUIT DIAGRAM PICTORIAL VIEWS OF COMPONENTS ARE REPLACED BY COMPONENT SYMBOLS.

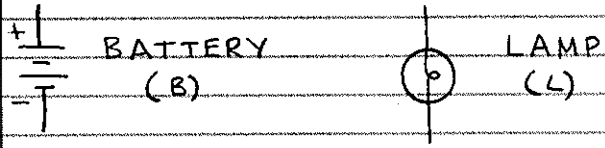

COMPONENT SYMBOLS

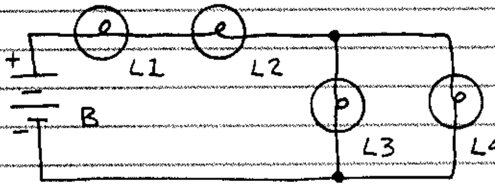

SERIES—PARALLEL CIRCUIT

☐ ELECTRICAL "SHORT" CIRCUIT— WHEN A WIRE OR OTHER CONDUCTOR IS PLACED ACROSS THE CONNECTIONS OF A COMPONENT, SOME OR ALL OF ANY CURRENT IN THE CIRCUIT MAY TAKE A SHORTCUT THROUGH THE CONDUCTOR. "SHORT" CIRCUITS SUCH AS THIS ARE USUALLY UNDESIRABLE AT BEST. THEY CAN CAUSE BATTERIES TO RAPIDLY LOSE THEIR CAPACITY. AND THEY CAN CAUSE DAMAGE TO WIRING AND COMPONENTS. "SHORT" CIRCUITS CAN EVEN CAUSE ENOUGH HEAT TO IGNITE THE INSULATION ON A WIRE! CAUTION: THE HUMAN BODY CONDUCTS ELECTRICITY. THEREFORE CARELESSLY TOUCHING AN ELECTRICAL CIRCUIT MAY CAUSE A "SHORT" CIRCUIT. IF THE VOLTAGE AND CURRENT ARE HIGH ENOUGH, YOU MAY RECEIVE A DANGEROUS OR EVEN LETHAL SHOCK.

☐ ELECTRICAL "GROUND"— ONE OF THE WIRES OF THE AC LINE IS CONNECTED TO EARTH BY A METAL ROD. METAL ENCLOSURES OF ELECTRICALLY POWERED DEVICES ARE CONNECTED TO THIS GROUND WIRE. THIS PREVENTS A SHOCK HAZARD SHOULD A NON-GROUNDED WIRE MAKE CONTACT WITH THE METAL ENCLOSURE. WITHOUT THE GROUND CONNECTION, A PERSON TOUCHING THE DEVICE WHILE STANDING ON THE GROUND OR A WET FLOOR MIGHT RECEIVE A DANGEROUS SHOCK.

GROUND ALSO REFERS TO THE POINT IN A CIRCUIT AT ZERO VOLTAGE, WHETHER OR NOT IT'S CONNECTED TO GROUND. FOR INSTANCE, THE MINUS (-) SIDE OF THE BATTERY IN THE CIRCUITS ABOVE AND ON THE PRECEEDING PAGE CAN BE CONSIDERED GROUND.

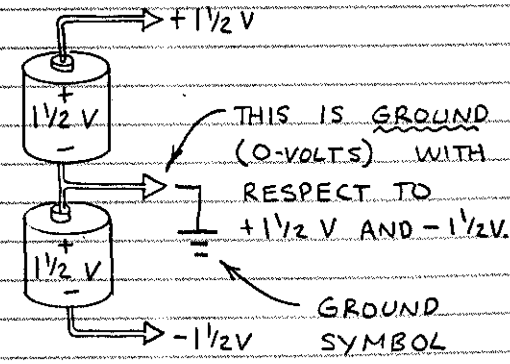

PULSES, WAVES, SIGNALS AND NOISE

ELECTRONICS IS THE STUDY AND APPLICATION OF ELECTRONS, THEIR BEHAVIOUR AND THEIR EFFECTS. THE SIMPLEST APPLICATIONS FOR ELECTRONS ARE STRAIGHTFORWARD AC AND DC CIRCUITS IN WHICH A CURRENT IS USED TO POWER LAMPS, ELECTROMAGNETS, MOTORS, SOLENOIDS AND SIMILAR DEVICES. WHAT TAKES ELECTRONICS FAR BEYOND THESE BASIC APPLICATIONS IS THE EASE WITH WHICH STREAMS OF ELECTRONS CAN BE CONTROLLED AND MANIPULATED.

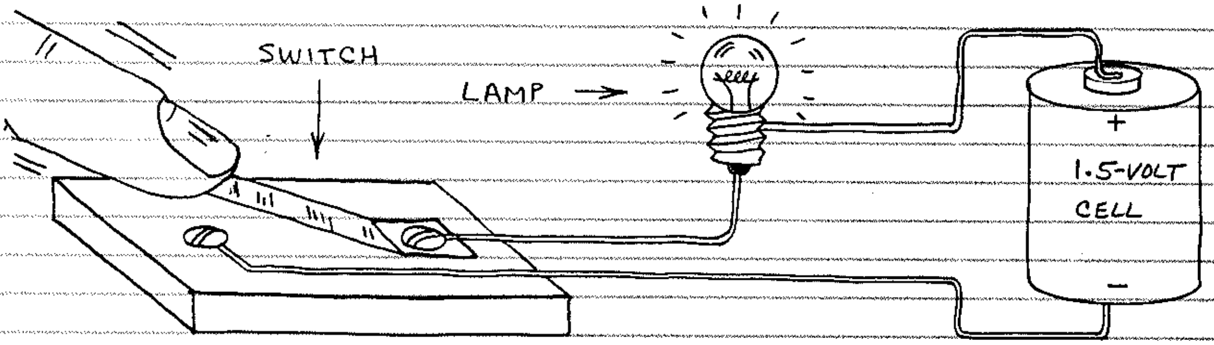

THIS SIMPLE CIRCUIT IS REALLY MORE USEFUL THAN IT FIRST APPEARS BECAUSE IT CAN SEND INFORMATION BY CONVERTING A PLANNED SEQUENCE OF SWITCH CLOSURES INTO FLASHES OF LIGHT.

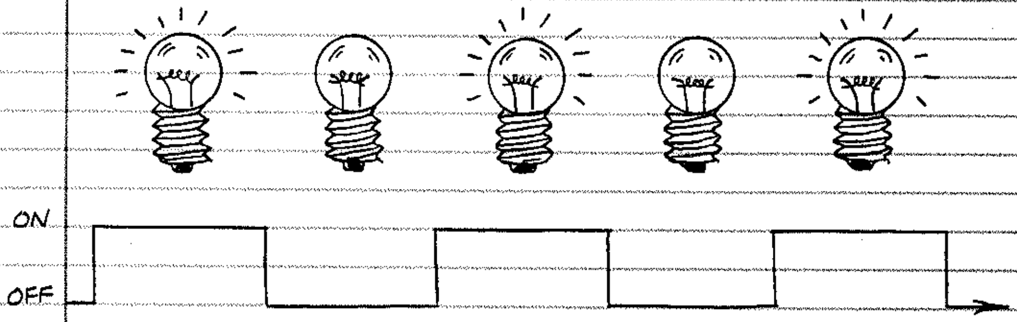

THE FLASHES OF THE LAMP CAN BE REPRESENTED BY A DIAGRAM LIKE THIS...

PATTERNS OF FLASHES OR PULSES LIKE THESE CAN REPRESENT COMPLEX INFORMATION LIKE SPEECH. OR SPEECH CAN BE TRANSFORMED INTO PROPORTIONAL VARIATIONS IN THE BRIGHTNESS OF A LAMP. HERE'S A SIMPLE WAY TO SEND VOICE OVER A BEAM OF REFLECTED LIGHT:

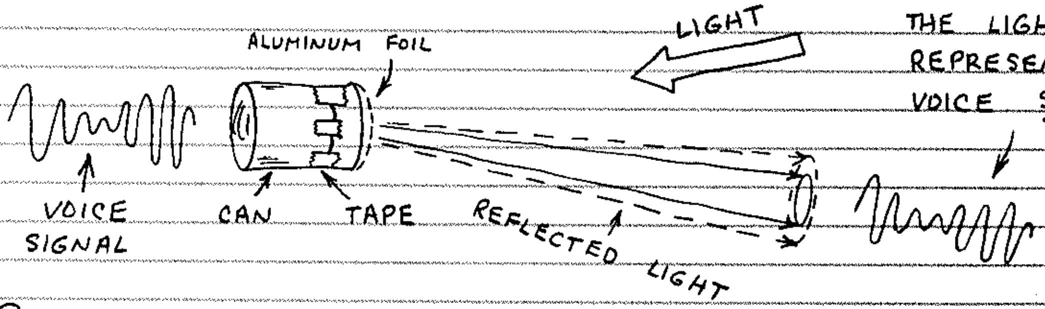

THE LIGHT INTENSITY REPRESENTS THE VOICE SIGNAL.

P. 123

☐ PULSES — A PULSE IS A SUDDEN, BRIEF INCREASE OR DECREASE IN A CURRENT FLOW. THE IDEAL PULSE WOULD HAVE AN INSTANTANEOUS RISE AND FALL, BUT REAL PULSES ARE NOT SO IDEAL.

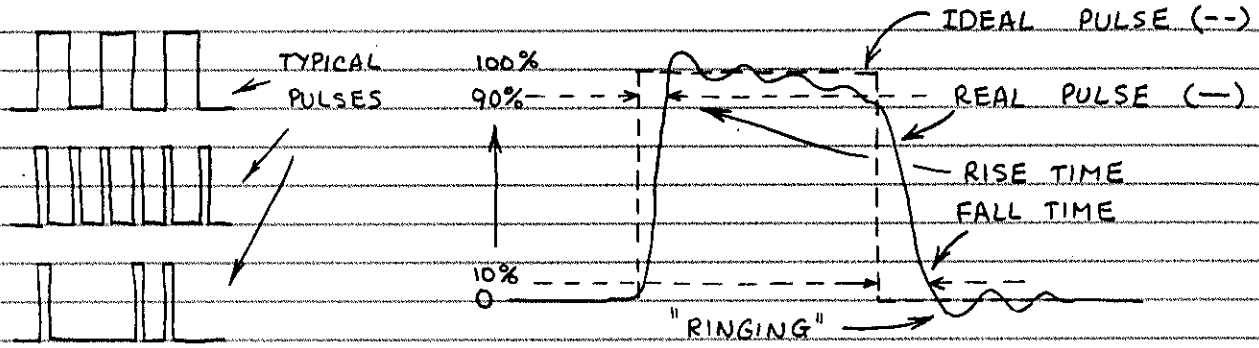

☐ WAVES — A WAVE IS A PERIODIC FLUCTUATION IN A CURRENT OR VOLTAGE. WAVES MAY HAVE A SINGLE POLARITY (DC) OR BOTH POSITIVE AND NEGATIVE COMPONENTS (AC). THERE ARE MANY KINDS OF WAVES. HERE ARE A FEW:

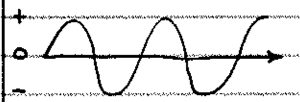

AC SINE WAVE

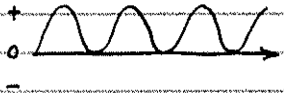

DC SINE WAVE

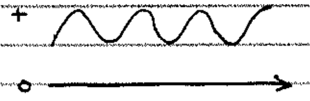

DC+ SINE WAVE

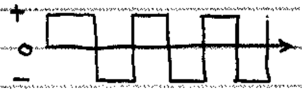

AC SQUARE WAVE

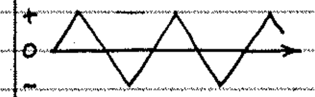

AC TRIANGLE WAVE

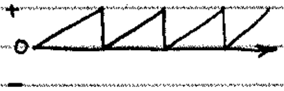

DC RAMP

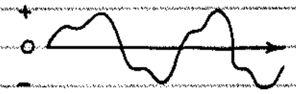

COMPLEX SINE WAVE

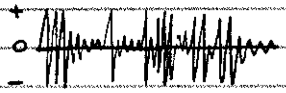

VOICE

☐ SIGNALS — A SIGNAL IS A PERIODIC WAVEFORM THAT CONVEYS INFORMATION. THE PROCESS THAT GENERATES THE WAVEFORM IS CALLED MODULATION. SIGNALS CAN BE AC, DC OR AC RIDING ON A DC LEVEL. THEIR ENEMY IS...

☐ NOISE — ALL ELECTRONIC DEVICES AND CIRCUITS GENERATE SMALL, RANDOM ELECTRICAL CURRENTS. WHEN THESE CURRENTS ARE UNWANTED, THEY'RE CALLED NOISE. NOISE CAN ALSO ENTER ELECTRONIC CIRCUITS BY MEANS OF THE ELECTROMAGNETIC WAVES GENERATED BY LIGHTNING, AUTOMOBILE IGNITION SYSTEMS, ELECTRIC MOTORS AND POWER LINES. WHILE NOISE MAY HAVE A LEVEL OF ONLY A FEW MILLIONTHS OF A VOLT OR AMPERE, IT MAY EASILY OBSCURE AN EQUALLY LOW LEVEL SIGNAL.

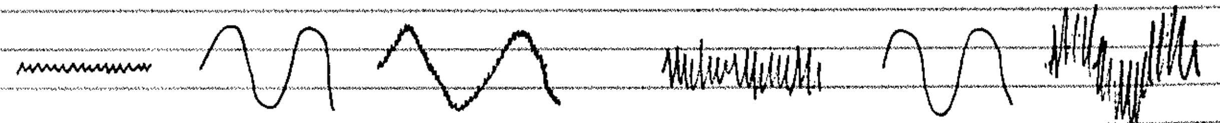

NOISE + SIGNAL = NOISY SIGNAL   NOISE + SIGNAL = ???

2. ELECTRONIC COMPONENTS

DOZENS OF DIFFERENT FAMILIES OF PARTS AND COMPONENTS BLOCK, CARRY, CONTROL, SELECT, STEER, SWITCH, STORE, MANIPULATE, REPLICATE, MODULATE AND EXPLOIT AN ELECTRICAL CURRENT. THOSE THAT USE SEMICONDUCTING CRYSTALS ARE SO IMPORTANT WE'LL DEVOTE AN ENTIRE CHAPTER TO THEM. YOU'LL FIND JUST ABOUT ALL THE REMAINING PARTS YOU SHOULD KNOW ABOUT IN THIS CHAPTER.

WIRE AND CABLE

USED TO CARRY AN ELECTRICAL CURRENT. MOST WIRE IS MADE FROM A LOW RESISTANCE METAL LIKE COPPER. SOLID WIRE IS A SINGLE CONDUCTOR. STRANDED WIRE IS TWO OR MORE TWISTED OR BRAIDED BARE CONDUCTORS. MOST WIRE IS PROTECTED BY AN INSULATING COVERING OF PLASTIC, RUBBER OR LACQUER. WIRE WHICH HAS BEEN TINNED IS EASIER TO SOLDER.

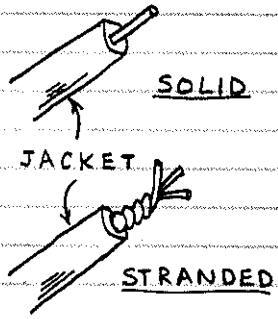

SPECIFICATIONS FOR BARE COPPER WIRE

GAUGE	DIAMETER (INCHES)	FEET/POUND	FEET/OHM
16	.05082	127.9	249.00
18	.04030	203.4	156.50
20	.03196	323.4	98.50
22	.02535	514.2	61.96
24	.02010	817.7	38.96
26	.01594	1300.0	24.50
28	.01264	2067.0	15.41
30	.01003	3287.0	9.69

CABLES HAVE ONE OR MORE CONDUCTORS AND MORE INSULATION THAN ORDINARY WIRE. COAXIAL CABLE CAN CARRY HIGH FREQUENCY SIGNALS (LIKE TELEVISION).

☐ CAUTION! ALWAYS USE WIRE RATED FOR THE CURRENT IT IS TO CARRY. IF A WIRE IS HOT TO THE TOUCH, IT'S CARRYING TOO MUCH CURRENT. USE A HEAVIER GAUGE WIRE OR REDUCE THE CURRENT. OTHERWISE...

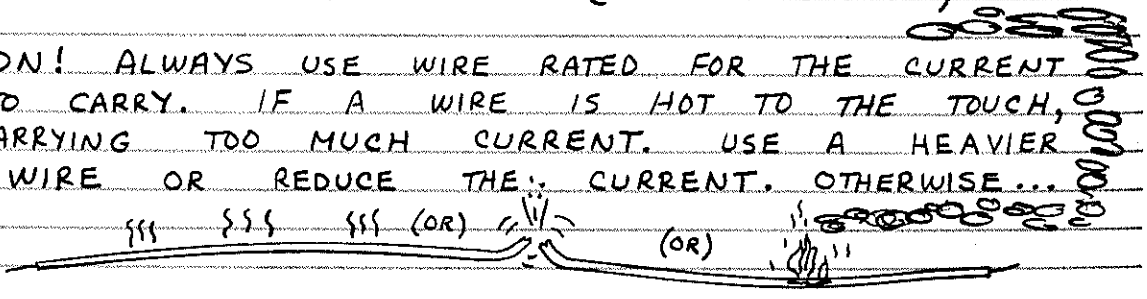

SWITCHES

MECHANICAL SWITCHES PERMIT OR INTERRUPT THE FLOW OF CURRENT. THEY ARE ALSO USED TO DIRECT CURRENT TO VARIOUS POINTS.

☐ THE BASIC KNIFE SWITCH — THE SIMPLEST SWITCH ...

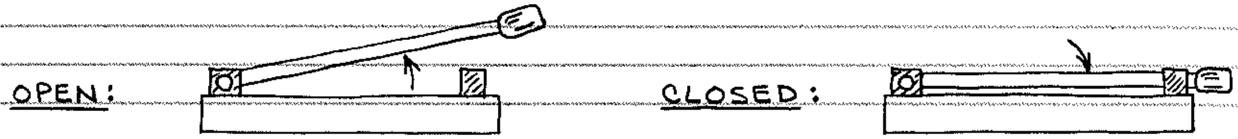

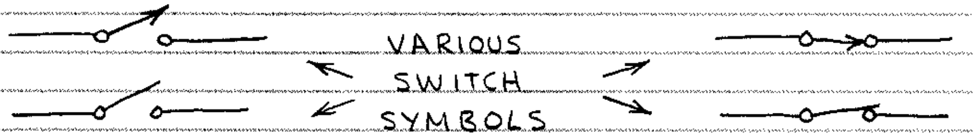

THIS IS CALLED AN SPST (SINGLE-POLE, SINGLE-THROW) SWITCH.

☐ MULTIPLE CONTACT SWITCHES — HERE ARE SYMBOLS FOR THE MAJOR KINDS:

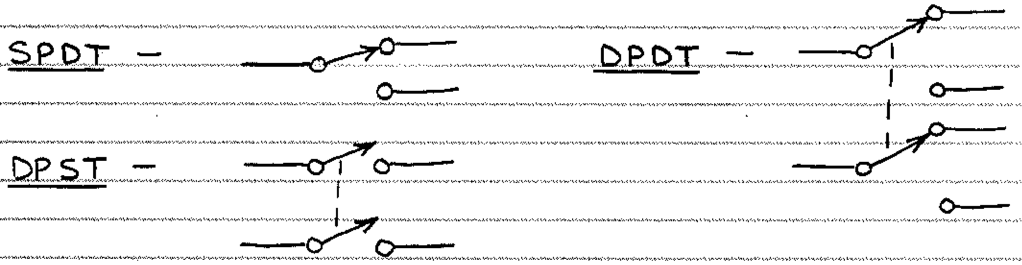

SPDT - SINGLE-POLE, DOUBLE-THROW
DPST - DOUBLE-POLE, SINGLE-THROW
DPDT - DOUBLE-POLE, DOUBLE-THROW

(THE DASHED LINE MEANS BOTH SIDES MOVE TOGETHER.)

☐ OTHER SWITCHES —

PUSHBUTTON. USUALLY SPST, NORMALLY OPEN (NO) OR NORMALLY CLOSED (NC).

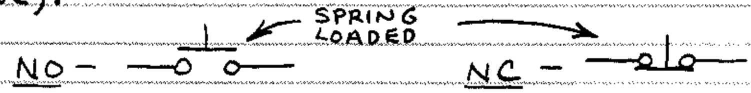

ROTARY. WAFER-LIKE WITH ONE POLE AND 2 OR MORE CONTACTS. WAFERS CAN BE STACKED TO PROVIDE MORE POLES. MANY VARIATIONS ARE POSSIBLE.

MERCURY. MERCURY BLOB CLOSES SWITCH. POSITION SENSITIVE.

OTHER. MANY KINDS OF TOGGLE, ROCKER, LEVER, SLIDE, PUSH-ON/PUSH-OFF, ILLUMINATED AND OTHER SWITCHES ARE AVAILABLE.

RELAYS

A RELAY IS AN ELECTROMAGNETIC SWITCH. A SMALL CURRENT FLOWING THROUGH A COIL IN THE RELAY CREATES A MAGNETIC FIELD THAT PULLS ONE SWITCH CONTACT AGAINST OR AWAY FROM ANOTHER.

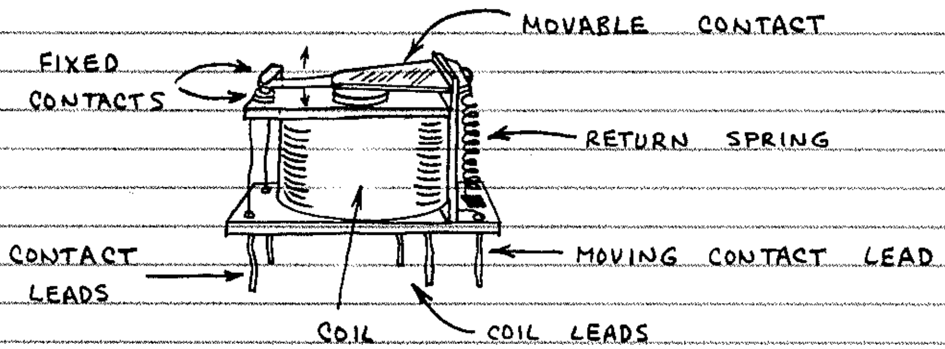
MOVABLE CONTACT
FIXED CONTACTS
RETURN SPRING
CONTACT LEADS
MOVING CONTACT LEAD
COIL
COIL LEADS

☐ RELAY SYMBOL — THE ARRANGEMENT OF CONTACTS CAN PROVIDE SPST, SPDT, DPST, DPDT AND OTHER SWITCH OPERATIONS.

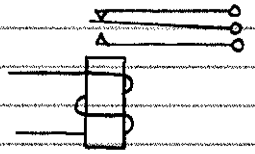

THIS IS THE SYMBOL FOR A RELAY WITH SPDT CONTACTS.

☐ REED SWITCH RELAYS — AN ENCLOSED GLASS TUBE HOUSING A PAIR OF CLOSELY SPACED SWITCH CONTACTS IS A REED SWITCH. A MAGNETIC FIELD WILL CLOSE THE CONTACTS. THIS MAKES POSSIBLE A VERY SIMPLE SPST RELAY.

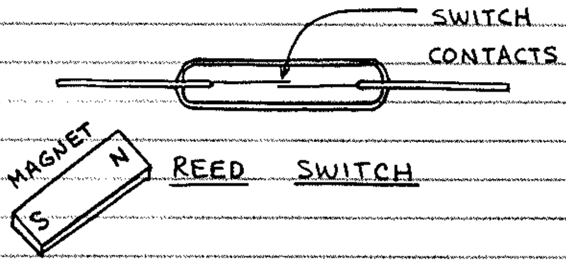
SWITCH CONTACTS
MAGNET
N
S

REED SWITCH

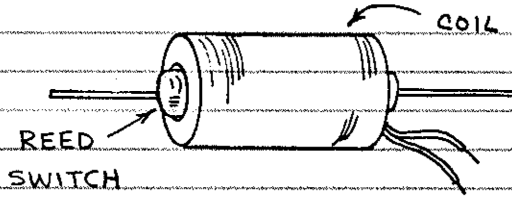
COIL
REED SWITCH

REED RELAY

MOVING COIL METER

A COIL ON A PIVOT BETWEEN THE POLES OF A U-SHAPED MAGNET WILL ROTATE WHEN A CURRENT IS PASSED THROUGH THE COIL. THIS IS THE PRINCIPLE OF THE MOVING COIL METER.

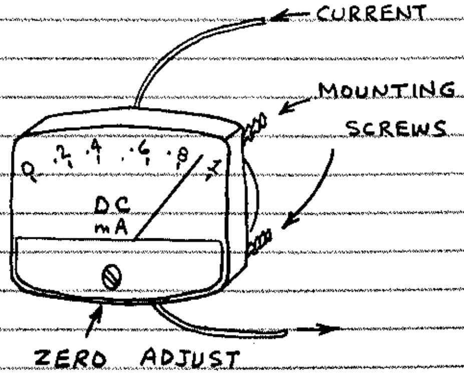
CURRENT
MOUNTING SCREWS
0 .2 .4 .6 .8 1
DC mA
ZERO ADJUST

MICROPHONES AND SPEAKERS

A MICROPHONE CONVERTS SOUND WAVE VARIATIONS INTO CORRESPONDING VARIATIONS IN AN ELECTRICAL CURRENT. THE SOUND WAVE VARIATIONS ARE FIRST CONVERTED TO BACK-AND-FORTH MOVEMENTS OF A FLEXIBLE FILM OR FOIL CALLED A DIAPHRAGM. THESE MOVEMENTS THEN CAUSE VARIATIONS IN AN ELECTRICAL CURRENT BY ANY OF THE FOLLOWING MEANS:

☐ CARBON— MOVEMENT OF THE DIAPHRAGM CHANGES THE PRESSURE APPLIED TO A CAPSULE OF CARBON PARTICLES. THIS CAUSES PROPORTIONAL CHANGES IN THE RESISTANCE OF THE CAPSULE.

☐ DYNAMIC— A SMALL COIL IS MOVED THROUGH A MAGNETIC FIELD AS THE DIAPHRAGM MOVES. THIS CAUSES A PROPORTIONAL OUTPUT CURRENT TO BE GENERATED.

☐ CONDENSER— THE MOVING DIAPHRAGM ALTERS THE DISTANCE BETWEEN TWO METAL PLATES. THE RESULT IS A PROPORTIONAL CHANGE IN THE CAPACITANCE OF THE PLATES.

☐ CRYSTAL— A WAFER OF PIEZOELECTRIC MATERIAL (WHICH PRODUCES A VOLTAGE WHEN BENT BY THE PRESSURE OF SOUND WAVES) FORMS THE DIAPHRAGM OR IS MECHANICALLY LINKED TO THE DIAPHRAGM.

A SPEAKER CONVERTS VARIATIONS IN A CURRENT OR VOLTAGE INTO SOUND WAVES. THE TWO MOST COMMON SPEAKERS ARE:

☐ MAGNETIC— SIMILAR IN PRINCIPLE TO A DYNAMIC MICROPHONE. IN FACT, A MAGNETIC SPEAKER CAN BE USED AS A MICROPHONE.

☐ CRYSTAL— SIMILAR IN PRINCIPLE TO A CRYSTAL MICROPHONE. A CRYSTAL SPEAKER CAN DOUBLE AS A MICROPHONE.

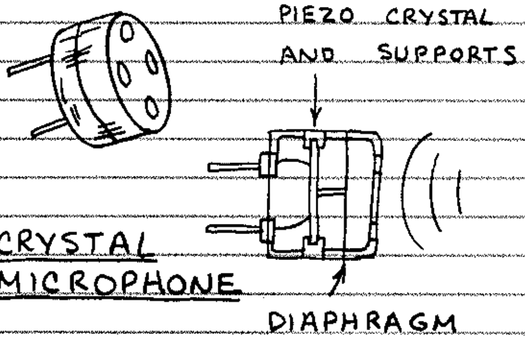

CRYSTAL MICROPHONE

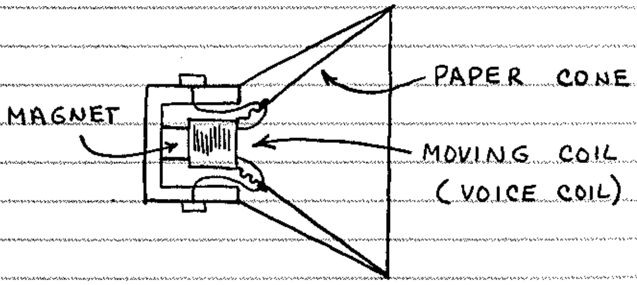

MAGNETIC SPEAKER

RESISTORS

RESISTORS COME IN DOZENS OF SIZES AND SHAPES BUT THEY ALL DO THE SAME THING: LIMIT* CURRENT. MORE ABOUT THAT LATER. FIRST, LET'S SEE HOW A TYPICAL RESISTOR IS MADE:

*OR RESIST

TYPICAL CARBON COMPOSITION RESISTOR

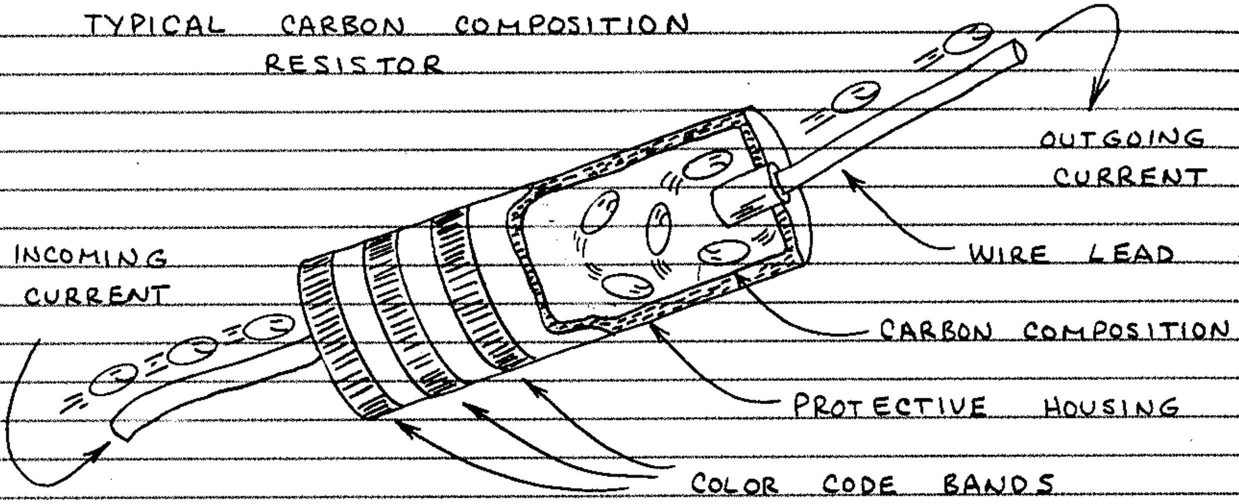
OUTGOING CURRENT
WIRE LEAD
INCOMING CURRENT
CARBON COMPOSITION
PROTECTIVE HOUSING
COLOR CODE BANDS

"CARBON COMPOSITION" IS JUST A FANCY WAY OF DESCRIBING POWDERED CARBON MIXED WITH A GLUE-LIKE BINDER. THIS KIND OF RESISTOR IS EASY TO MAKE. AND ITS RESISTANCE CAN BE CHANGED FROM ONE RESISTOR TO THE NEXT SIMPLY BY CHANGING THE RATIO OF CARBON PARTICLES TO BINDER. MORE CARBON GIVES LESS RESISTANCE.

☐ DO-IT-YOURSELF RESISTORS — YOU CAN MAKE A RESISTOR BY DRAWING A LINE WITH A SOFT LEAD PENCIL ON A SHEET OF PAPER. MEASURE THE RESISTANCE OF THE LINE OR POINTS ALONG IT BY TOUCHING THE PROBES OF A MULTIMETER TO THE LINE. BE SURE TO SET THE MULTIMETER TO ITS HIGHEST RESISTANCE SCALE. THE RESISTANCE OF A SINGLE LINE MAY BE TOO HIGH TO MEASURE. IF SO, DRAW OVER THE LINE A DOZEN OR SO TIMES. HERE'S WHAT I MEASURED:

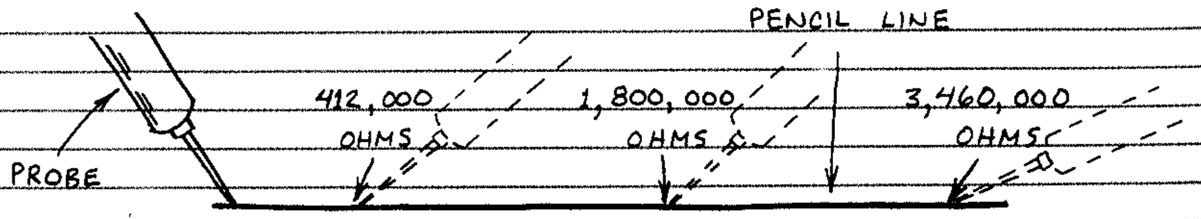
PENCIL LINE
412,000 OHMS
1,800,000 OHMS
3,460,000 OHMS
PROBE

☐ RESISTOR COLOR CODE — SEE THOSE COLOR CODE BANDS ON THE RESISTOR PICTORIAL? IN REAL LIFE THEY'RE KIND OF PRETTY. BUT THEY HAVE A FAR MORE IMPORTANT PURPOSE: THEY INDICATE THE RESISTANCE OF THE RESISTOR THEY DECORATE. HERE'S HOW:

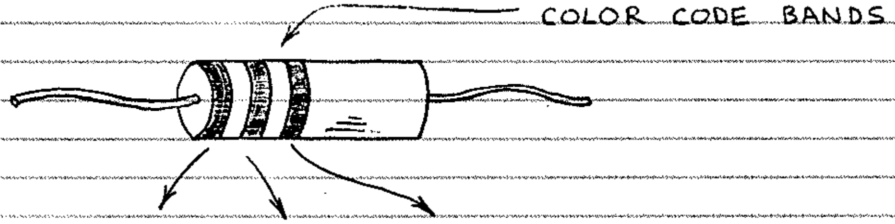

COLOR	1	2	3 (MULTIPLIER)
BLACK	0	0	1
BROWN	1	1	10
RED	2	2	100
ORANGE	3	3	1,000
YELLOW	4	4	10,000
GREEN	5	5	100,000
BLUE	6	6	1,000,000
VIOLET	7	7	10,000,000
GRAY	8	8	100,000,000
WHITE	9	9	(NONE)

NOTE: SOMETIMES THERE'S A FOURTH BAND. IT INDICATES THE TOLERANCE* OF THE RESISTOR:

GOLD = ±5%
SILVER = ±10%
NONE = ±20%

* OR ACCURACY

LOOKS COMPLICATED THE FIRST TIME... BUT YOU'LL QUICKLY LEARN HOW TO USE IT. FOR EXAMPLE, WHAT'S THE RESISTANCE OF A RESISTOR COLOR CODED YELLOW, VIOLET AND RED? YELLOW IS THE FIRST COLOR SO THE FIRST NUMBER IS 4. VIOLET IS THE SECOND COLOR SO THE SECOND NUMBER IS 7. SINCE THE THIRD COLOR IS RED, THE MULTIPLIER IS 100. THEREFORE, THE RESISTANCE IS 47 × 100 OR 4700 OHMS. NO FOURTH COLOR BAND MEANS THE ACTUAL RESISTANCE IS 4700 ± 20%. 20% OF 4700 IS 940. THEREFORE, THE ACTUAL VALUE IS BETWEEN 3760 AND 5640 OHMS.

☐ SUBSTITUTING RESISTORS — WHAT IF YOU NEED A 6700-OHM RESISTOR BUT CAN ONLY FIND A 6800-OHM UNIT? YOU CAN ALMOST ALWAYS USE ANY VALUE WITHIN 10 OR 20% OF THE REQUIRED VALUE SO GO AHEAD AND USE IT. IF A PARTICULAR CIRCUIT REQUIRES MORE ACCURACY IT WILL TELL YOU. OF COURSE YOU CAN BUILD UP CUSTOM RESISTANCES BY CONNECTING TWO OR MORE RESISTORS IN SERIES OR IN PARALLEL. MORE ABOUT THAT LATER.

☐ RESISTOR SUBSTITUTION PRECAUTIONS — RESISTORS THAT CONDUCT LOTS OF CURRENT CAN BECOME VERY HOT! THEREFORE, ALWAYS USE RESISTORS HAVING THE PROPER POWER RATING. IF A PROJECT YOU'RE BUILDING DOESN'T SPECIFY THE POWER RATING FOR ITS RESISTORS, IT'S USUALLY OK TO USE 1/4 OR 1/2 WATT UNITS.

☐ SOME RESISTOR SHORTHAND — OFTEN YOU'LL SEE RESISTORS DESIGNATED WITH A K OR M SUFFIX, LIKE 47K OR 10M. K MEANS KILO, AFTER THE GREEK WORD FOR 1,000. THEREFORE, 47K MEANS 47 × 1,000 OR 47,000. M IS SHORT FOR MEGOHM OR 1,000,000 OHMS. THEREFORE A 1M RESISTOR HAS A RESISTANCE OF 1 × 1,000,000 OR 1,000,000 OHMS. SUMMING UP...

K = × 1,000 (47K = 47 × 1,000 = 47,000 OHMS)

M = × 1,000,000 (2.2M = 2.2 × 1,000,000 = 2,200,000 OHMS)

☐ OTHER KINDS OF RESISTORS — THE CARBON COMPOSITION RESISTOR IS ONLY ONE OF SEVERAL MAJOR KINDS OF RESISTORS. HERE ARE OTHERS:

METAL FILM RESISTORS. VARIOUS KINDS OF RESISTORS THAT USE A THIN FILM OF METAL OR A METAL PARTICLE MIXTURE TO ACHIEVE VARIOUS RESISTANCES.

CARBON FILM RESISTORS. THESE ARE MADE BY DEPOSITING A CARBON FILM ON A SMALL CERAMIC CYLINDER. A SPIRAL GROOVE CUT INTO THE FILM CONTROLS THE LENGTH OF CARBON BETWEEN THE LEADS, HENCE THE RESISTANCE.

WIRE-WOUND RESISTORS. THESE CONSIST OF A TUBULAR FORM WRAPPED WITH COILS OF RESISTANCE WIRE. THEY ARE VERY ACCURATE AND CAN TAKE LOTS OF HEAT.

PHOTORESISTORS. ALSO CALLED PHOTOCELLS. MADE FROM A LIGHT SENSITIVE MATERIAL LIKE CADMIUM SULFIDE. INCREASING THE LIGHT LEVEL DECREASES THE RESISTANCE. MORE ABOUT THIS LATER.

THERMISTORS. THIS IS A TEMPERATURE SENSITIVE RESISTOR. INCREASING THE TEMPERATURE DECREASES THE RESISTANCE (IN MOST CASES).

☐ VARIABLE RESISTORS — OFTEN IT'S NECESSARY TO CHANGE THE RESISTANCE OF A RESISTOR. VARIABLE RESISTORS ARE CALLED *POTENTIOMETERS*. THEY ARE USED TO ALTER THE VOLUME OF A RADIO, CHANGE THE BRIGHTNESS OF A LAMP, ADJUST THE CALIBRATION OF A METER, ETC. *TRIMMERS* ARE POTENTIOMETERS EQUIPPED WITH A PLASTIC THUMBWHEEL OR A SLOT FOR A SCREWDRIVER BLADE. THEY ARE DESIGNED FOR OCCASIONAL ADJUSTMENT.

☐ RESISTOR SYMBOLS:

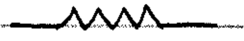
FIXED RESISTOR

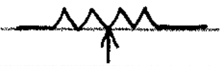
POTENTIOMETER

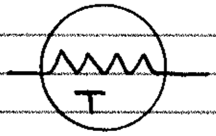

THERMISTOR

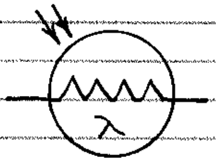

PHOTORESISTOR

HOW RESISTORS ARE USED

☐ SERIES CIRCUIT — OFTEN RESISTORS ARE CONNECTED IN SERIES LIKE THIS:

THE *TOTAL* RESISTANCE IS SIMPLY THE *SUM* OF THE INDIVIDUAL RESISTANCES.

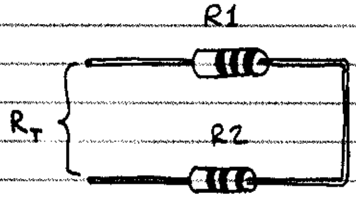

$$R_T = R1 + R2$$

☐ PARALLEL CIRCUIT — RESISTORS CAN ALSO BE CONNECTED IN PARALLEL LIKE THIS:

THE *TOTAL* RESISTANCE IS THE PRODUCT OF THE TWO RESISTANCES DIVIDED BY THEIR SUM.

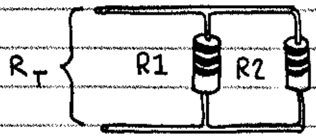

$$R_T = \frac{R1 \times R2}{R1 + R2}$$

FOR THREE OR MORE IN PARALLEL, GO FIND YOUR CALCULATOR BECAUSE...

$$R_T = \frac{1}{\frac{1}{R1} + \frac{1}{R2} + \frac{1}{R3} \ldots \text{ETC.}}$$

☐ VOLTAGE DIVISION — SUPER IMPORTANT! V_{OUT} IS DETERMINED BY RATIO OF R1 AND R2. HERE'S THE FORMULA:

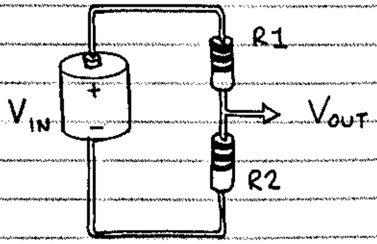

$$V_{OUT} = V_{IN} \left(\frac{R2}{R1 + R2} \right)$$

CAPACITORS

THERE ARE MANY KINDS OF CAPACITORS, BUT THEY ALL DO THE SAME THING: STORE ELECTRONS. THE SIMPLEST CAPACITOR IS TWO CONDUCTORS SEPARATED BY AN INSULATING MATERIAL CALLED THE DIELECTRIC. LIKE THIS:

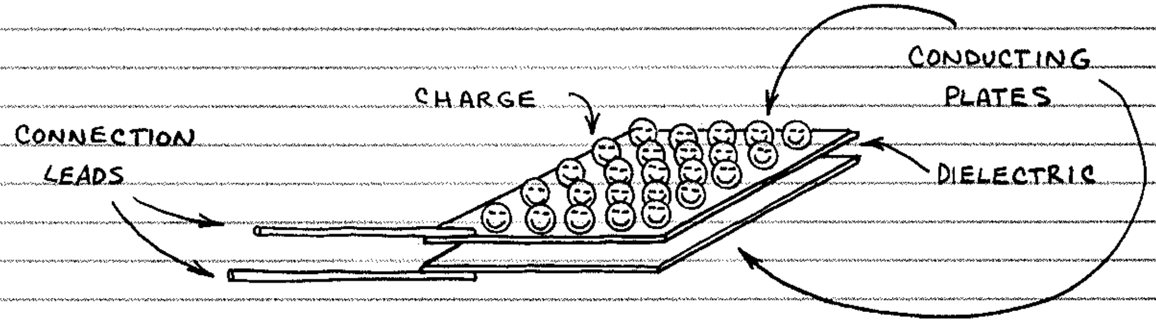

THE DIELECTRIC CAN BE PAPER, PLASTIC FILM, MICA, GLASS, CERAMIC, AIR OR A VACUUM. THE PLATES CAN BE ALUMINUM DISCS, ALUMINUM FOIL OR A THIN FILM OF METAL APPLIED TO OPPOSITE SIDES OF A SOLID DIELECTRIC. THE CONDUCTOR – DIELECTRIC – CONDUCTOR SANDWICH CAN BE ROLLED INTO A CYLINDER OR LEFT FLAT. MORE ABOUT TYPES OF CAPACITORS LATER.

HOW TO MAKE A CAPACITOR

YOU CAN MAKE A CAPACITOR FROM TWO SHEETS OF ALUMINUM FOIL AND ONE SHEET OF WAXED PAPER. FOLD THE PAPER AROUND ONE FOIL SHEET AND STACK THE SHEETS LIKE THIS:

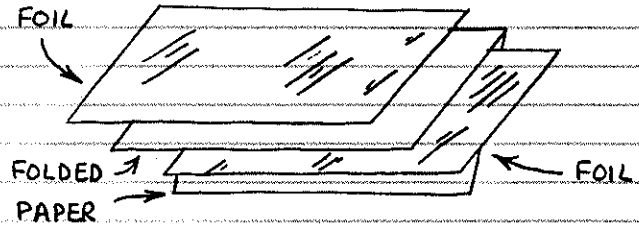

THEN FOLD THE SHEETS LIKE THIS:

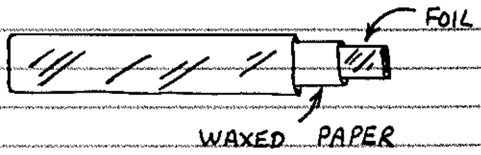

BE SURE THE FOIL SHEETS DON'T TOUCH! PRESS THE CONTACTS OF A 9-VOLT BATTERY BRIEFLY TO THE EXPOSED ENDS OF THE FOIL SHEETS. THEN TOUCH THE PROBES OF A HIGH-IMPEDANCE MULTIMETER TO THE FOIL SHEETS. THE METER WILL INDICATE A SMALL VOLTAGE FOR A FEW SECONDS. THE VOLTAGE WILL THEN FALL TO ZERO.

☐ CHARGING A CAPACITOR — THE MINUS SIDE OF OUR HOMEMADE CAPACITOR IS CHARGED WITH ELECTRONS ALMOST IMMEDIATELY. SINCE RESISTORS LIMIT CURRENT YOU CAN SLOW DOWN THE CHARGING TIME BY PLACING A RESISTOR BETWEEN THE CAPACITOR AND THE 9-VOLT BATTERY:

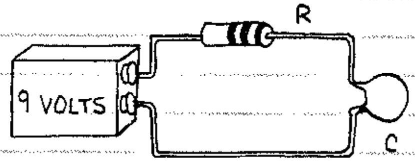

HERE'S A GRAPH OF THE CHARGING TIME:

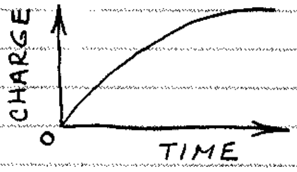

☐ DISCHARGING A CAPACITOR — THE ELECTRONS IN A CHARGED CAPACITOR WILL GRADUALLY LEAK THROUGH THE DIELECTRIC UNTIL BOTH PLATES HAVE AN EQUAL CHARGE. THE CAPACITOR IS THEN DISCHARGED. THE CAPACITOR CAN BE DISCHARGED VERY QUICKLY BY CONNECTING ITS PLATES TOGETHER. OR IT CAN BE DISCHARGED MORE SLOWLY BY CONNECTING A RESISTOR ACROSS IT:

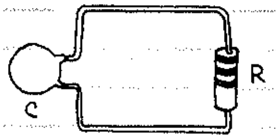

HERE'S A GRAPH OF THE DISCHARGE TIME:

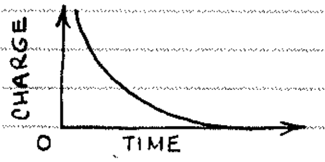

☐ SPECIFYING CAPACITORS — THE ABILITY TO STORE ELECTRONS IS KNOWN AS CAPACITANCE. CAPACITANCE IS SPECIFIED IN FARADS. A 1-FARAD CAPACITOR CONNECTED TO A 1-VOLT SUPPLY WILL STORE 6,280,000,000,000,000,000 (6.28×10^{18}) ELECTRONS! MOST CAPACITORS HAVE MUCH SMALLER VALUES. SMALL CAPACITORS ARE SPECIFIED IN PICOFARADS (TRILLIONTHS OF A FARAD) AND LARGER CAPACITORS ARE SPECIFIED IN MICROFARADS (MILLIONTHS OF A FARAD). SUMMING UP:

1-FARAD = 1 F
1-MICROFARAD = $1\,\mu F = 10^{-6}\,F = 0.000\,001\,F$
1-PICOFARAD = $1\,pF = 10^{-12}\,F = 0.000\,000\,000\,001\,F$

☐ SUBSTITUTING CAPACITORS — THE CAPACITANCE SPECIFIED FOR MOST CAPACITORS MAY BE FROM 5 TO 100 % AWAY FROM THE ACTUAL VALUE. THEREFORE YOU CAN OFTEN SUBSTITUTE CLOSE VALUES FOR A SPECIFIED VALUE. BE SURE, HOWEVER, TO USE A CAPACITOR RATED AT THE EXPECTED MAXIMUM VOLTAGE LEVEL!

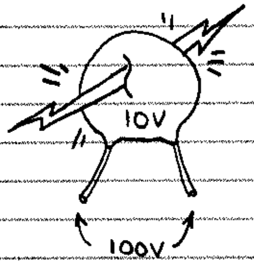

☐ CAPACITOR SUBSTITUTION PRECAUTIONS — YOU MUST MAKE SURE THE CAPACITOR YOU PLAN TO USE MEETS OR EXCEEDS THE REQUIRED VOLTAGE RATING. OTHERWISE ITS DIELECTRIC MAY BE ZAPPED BY THE STORED CHARGE. THE VOLTAGE RATING IS USUALLY PRINTED ON THE CAPACITOR. V MEANS VOLTS. WV IS WORKING VOLTS (SAME THING).

☐ KINDS OF CAPACITORS — CAPACITORS ARE OFTEN LABELED ACCORDING TO THEIR DIELECTRIC. THUS YOU'LL SEE REFERENCES TO CERAMIC, MICA, POLYSTYRENE AND MANY OTHERS. ALL THESE ARE FIXED VALUE CAPACITORS. SOME CAPACITORS HAVE A VARIABLE CAPACITY AND A SPECIAL CLASS OF FIXED CAPACITORS HAS MUCH MORE CAPACITY THAN OTHER CAPACITORS. HERE'S MORE:

VARIABLE CAPACITORS. THESE USUALLY HAVE ONE OR MORE NON-MOVING PLATES AND ONE OR MORE MOVING PLATES. THE CAPACITANCE IS CHANGED BY ROTATING A ROD AFFIXED TO ONE SIDE OF THE MOVABLE PLATES.

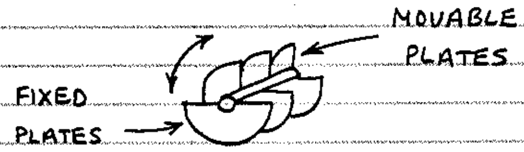

THIS KIND IS USED TO TUNE RADIO RECEIVERS AND TRANSMITTERS. THE DIELECTRIC IS USUALLY AIR.

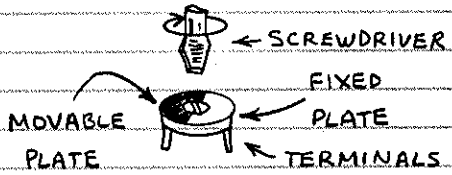

THIS KIND IS USED TO TUNE OSCILLATORS LIKE THOSE USED IN DIGITAL WATCHES. THEY'RE SMALL.

ELECTROLYTIC CAPACITORS. UNIQUE IN THAT A THIN OXIDE LAYER FORMED ON ALUMINUM OR TANTALUM FOIL IS THE DIELECTRIC. MUCH HIGHER CAPACITANCE THAN NON-ELECTROLYTIC TYPES. TANTALUM UNITS HAVE MORE CAPACITANCE PER VOLUME AND A LONGER LIFE THAN ALUMINUM ELECTROLYTICS. BUT THEY COST MORE. MOST ELECTROLYTICS ARE POLARIZED. THEY MUST BE CONNECTED INTO A CIRCUIT IN THE PROPER DIRECTION:

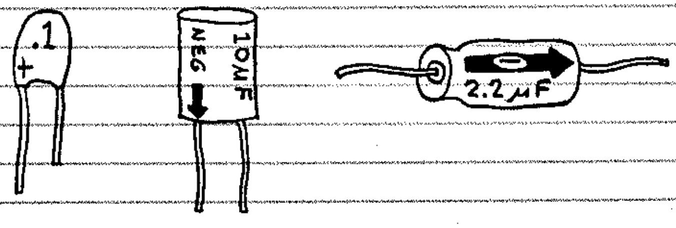

POSITIVE LEAD MUST GO TO MOST POSITIVE CONNECTION POINT!

☐ CAPACITOR SYMBOLS:

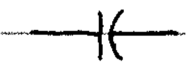
FIXED

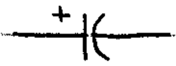
FIXED

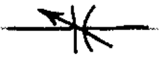
VARIABLE

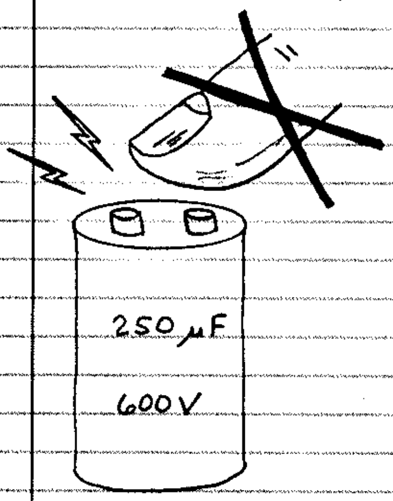

☐ WARNING! CAPACITORS CAN STORE A CHARGE FOR A CONSIDERABLE TIME AFTER THE POWER TO THEM HAS BEEN SWITCHED OFF. THIS CHARGE CAN BE DANGEROUS! A LARGE ELECTROLYTIC CHARGED TO ONLY 5 OR 10 VOLTS CAN MELT THE TIP OF A SCREWDRIVER PLACED ACROSS ITS TERMINALS! HIGH VOLTAGE CAPACITORS LIKE THOSE USED IN TELEVISION SETS AND PHOTOFLASH UNITS CAN STORE A LETHAL CHARGE! NEVER TOUCH THE LEADS OF SUCH A CAPACITOR. AT THE VERY LEAST THE JOLT CAN THROW YOU ACROSS A ROOM!

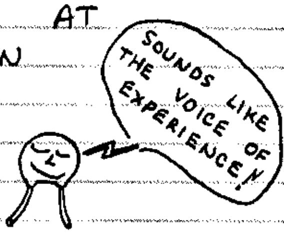
SOUNDS LIKE THE VOICE OF EXPERIENCE!

HOW CAPACITORS ARE USED

☐ PARALLEL CIRCUIT — OFTEN CAPACITORS ARE CONNECTED IN PARALLEL LIKE THIS:

THE TOTAL CAPACITANCE IS THE SUM OF THE INDIVIDUAL CAPACITANCES.

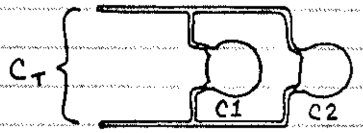

$$C_T = C1 + C2$$

☐ SERIES CIRCUIT — SOMETIMES CAPACITORS ARE CONNECTED IN SERIES LIKE THIS:

THE TOTAL CAPACITANCE IS THE PRODUCT OF THE TWO CAPACITANCES DIVIDED BY THEIR SUM.

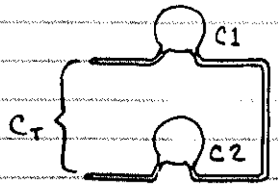

$$C_T = \frac{C1 \times C2}{C1 + C2}$$

THREE OR MORE CAPACITORS IN SERIES? HERE'S THE FORMULA:

$$C_T = \frac{1}{\frac{1}{C1} + \frac{1}{C2} + \frac{1}{C3} \ldots \text{ETC.}}$$

☐ AND MORE — THERE ARE MANY OTHER WAYS TO USE CAPACITORS, SOME OF WHICH ARE SHOWN NEXT...

RESISTOR AND CAPACITOR APPLICATIONS

RESISTORS AND CAPACITORS ARE THE KEY INGREDIENTS OF MANY ELECTRONIC CIRCUITS. HERE ARE SOME REASONS WHY:

☐ POWER SUPPLY FILTER — A CAPACITOR WILL SMOOTH (FILTER) THE PULSATING VOLTAGE FROM A POWER SUPPLY INTO A STEADY DIRECT CURRENT (DC).

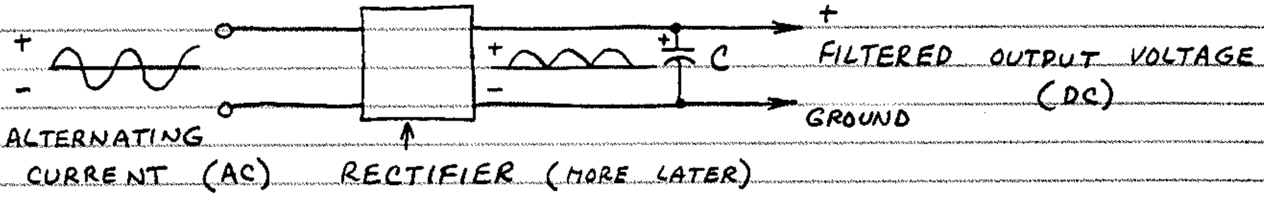
+
-
+
-
C
+
FILTERED OUTPUT VOLTAGE (DC)
GROUND
ALTERNATING CURRENT (AC)
RECTIFIER (MORE LATER)

☐ SPIKE REMOVER — DIGITAL LOGIC CIRCUITS, WHICH WE'LL FIND OUT MORE ABOUT LATER, CAN USE LOTS OF CURRENT MOMENTARILY WHEN THEY SWITCH FROM OFF TO ON OR VICE VERSA. THIS CAN CAUSE VERY BRIEF BUT SUBSTANTIAL REDUCTIONS IN POWER APPLIED TO NEARBY CIRCUITS. THESE POWER SPIKES (OR GLITCHES, AS THEY ARE SOMETIMES CALLED) CAN BE ELIMINATED BY PLACING A SMALL (0.1 μF) CAPACITOR ACROSS THE POWER LEADS OF THE LOGIC CIRCUIT:

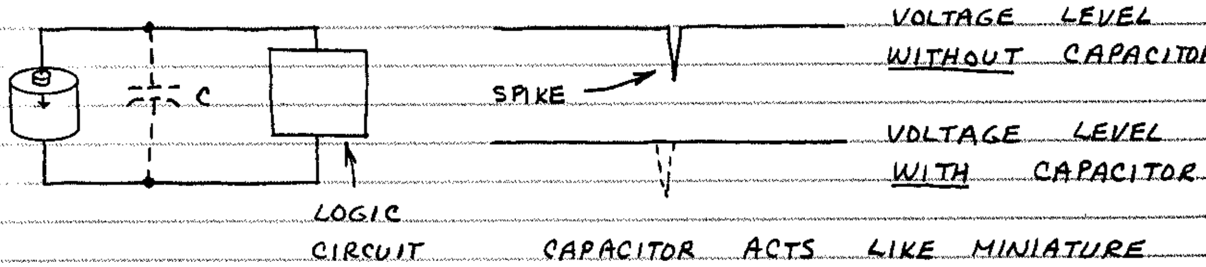
C
LOGIC CIRCUIT
SPIKE
VOLTAGE LEVEL WITHOUT CAPACITOR
VOLTAGE LEVEL WITH CAPACITOR

CAPACITOR ACTS LIKE MINIATURE BATTERY THAT SUPPLIES POWER DURING THE SPIKE.

☐ AC-DC SELECTIVE FILTER — OFTEN AN ELECTRICAL SIGNAL WILL BE RIDING ATOP A STEADY DC SIGNAL. FOR EXAMPLE, THE SIGNAL FROM A LIGHTWAVE COMMUNICATION SYSTEM MAY LOOK LIKE THIS WHEN IT'S DARK: BUT SUNLIGHT CAUSES THIS:

P. 123

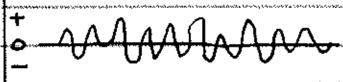

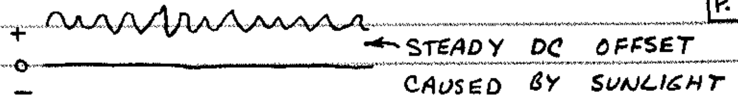
+
0
-
STEADY DC OFFSET CAUSED BY SUNLIGHT

A CAPACITOR WILL PASS THE FLUCTUATING SIGNAL AND COMPLETELY BLOCK THE STEADY DC LEVEL.

☐ R-C CIRCUITS — TWO CIRCUITS THAT COMBINE A RESISTOR (R) AND CAPACITOR (C) ARE VERY IMPORTANT. THEY ARE THE INTEGRATOR AND DIFFERENTIATOR. BOTH THESE CIRCUITS ARE USED TO RESHAPE AN INCOMING STREAM OF WAVES OR PULSES.

THE PRODUCT OF R AND C IN THESE CIRCUITS IS CALLED THE RC TIME CONSTANT. FOR THE CIRCUITS SHOWN BELOW, THE RC TIME CONSTANT (IN SECONDS) IS AT LEAST TEN TIMES THE INTERVAL BETWEEN INCOMING CYCLES OR PULSES.

1. INTEGRATOR. HERE'S A BASIC RC INTEGRATOR:

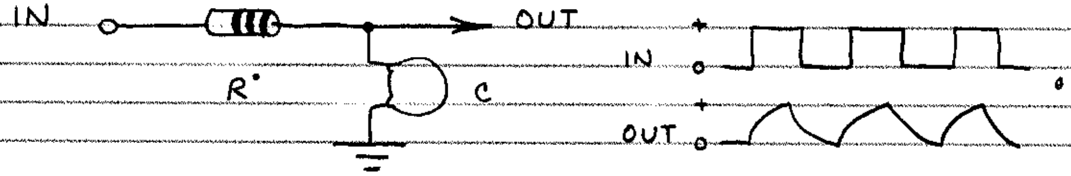

IF THE INPUT PULSES ARE SPEEDED UP, THE OUTPUT WAVEFORMS (OFTEN CALLED A SAWTOOTH) WILL NOT REACH THEIR FULL HEIGHT (AMPLITUDE). IT'S EASY TO DESIGN AN AMPLIFIER THAT IGNORES WAVES WITH LESS THAN A DESIRED AMPLITUDE. THEREFORE, THE INTEGRATOR CAN FUNCTION AS A FILTER WHICH PASSES ONLY SIGNALS BELOW A CERTAIN FREQUENCY.

2. DIFFERENTIATOR. HERE'S A BASIC RC DIFFERENTIATOR:

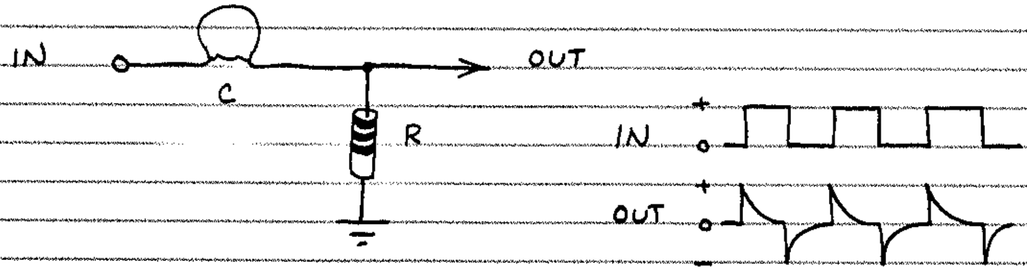

THIS CIRCUIT PRODUCES SYMMETRICAL OUTPUT WAVES WITH SHARP POSITIVE AND NEGATIVE PEAKS. IT'S USED TO MAKE NARROW PULSE GENERATORS FOR TELEVISION RECEIVERS AND TO TRIGGER DIGITAL LOGIC CIRCUITS.

☐ MORE ABOUT R-C — YOU WILL OFTEN SEE REFERENCES TO THE RC TIME CONSTANT OF A CIRCUIT. IT'S THE TIME IN SECONDS FOR A CHARGING OR DISCHARGING CAPACITOR TO GO THROUGH 63.3% OF THE CHANGE IN CHARGE.

COILS

ELECTRONS MOVING THROUGH A WIRE CAUSE AN *ELECTROMAGNETIC FIELD* TO ENCIRCLE THE WIRE. AS YOU KNOW FROM CHAPTER 1, PASSING A CURRENT THROUGH A WIRE THAT'S BEEN WRAPPED AS A COIL (P. 15) CREATES AN EVEN STRONGER FIELD. THIS FIELD MAKES POSSIBLE SOLENOIDS, MOTORS AND ELECTROMAGNETS. COILS HAVE OTHER IMPORTANT ROLES, TOO:

1. COILS RESIST *RAPID* CHANGES IN THE CURRENT FLOWING THROUGH THEM WHILE FREELY PASSING STEADY (DC) CURRENT. HERE ARE SOME EXAMPLES:

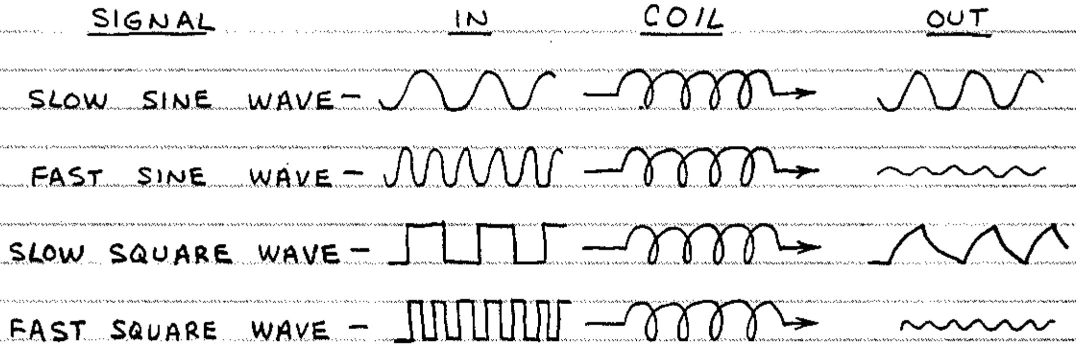
SIGNAL
IN
COIL
OUT
SLOW SINE WAVE -
FAST SINE WAVE -
SLOW SQUARE WAVE -
FAST SQUARE WAVE -

SOMETIMES A COIL WILL ADD *RINGING* TO A SQUARE WAVE PASSING THROUGH IT. THIS CAN HAPPEN WHEN THE RESISTANCE OF THE EXTERNAL CURRENT PATH THAT CONNECTS THE ENDS OF THE COIL IS HIGH.

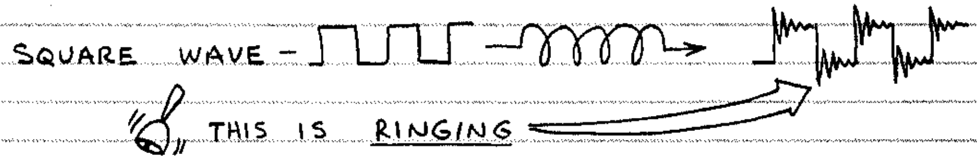
SQUARE WAVE -
THIS IS RINGING

2. SOME OF THE ENERGY IN THE FIELD AROUND A COIL CAN BE *INDUCED* (TRANSFERRED) INTO A SECOND, NEARBY COIL. THIS IS THE PRINCIPLE OF THE *TRANSFORMER*:

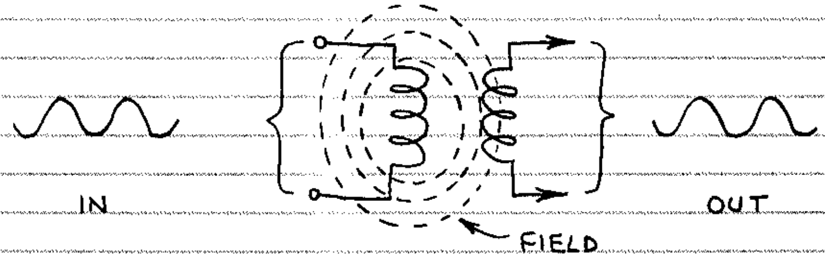
IN
OUT
FIELD

THE INPUT SIDE OF THE TRANSFORMER IS CALLED THE *PRIMARY*. THE OUTPUT SIDE IS CALLED THE *SECONDARY*.

☐ TYPES OF COILS — THERE ARE MANY DIFFERENT TYPES OF COILS. HERE ARE SOME OF THEM:

TUNING COIL. RADIOS USE VARIOUS COILS TO HELP SELECT A DESIRED SIGNAL. TUNING COILS HAVE A SERIES OF TAPS OR A MOVABLE CORE SO THEIR INDUCTANCE*, HENCE RESONANT FREQUENCY, CAN BE CHANGED.

TYPICAL TUNING COIL

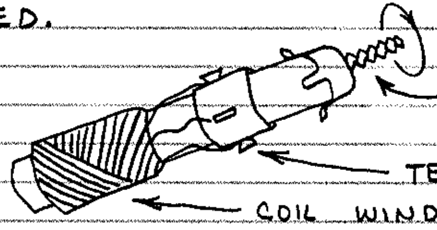
CORE ADJUSTMENT SCREW
TERMINALS
COIL WINDING

*OPPOSITION TO CURRENT CHANGES.

ANTENNA COIL. RADIOS OFTEN USE A BROADLY TUNED COIL TO PICK UP RADIO SIGNALS.

TYPICAL ANTENNA COIL

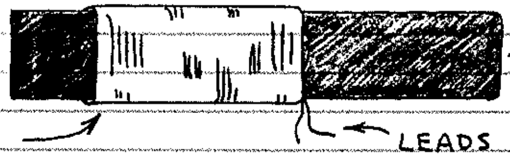
FERRITE CORE (BAR-SHAPED)
COIL
LEADS

CHOKE. USED IN MANY CIRCUITS TO LIMIT OR SUPPRESS FLUCTUATING SIGNALS WHILE PASSING A STEADY CURRENT. CHOKES ARE AVAILABLE IN MANY SHAPES AND SIZES:

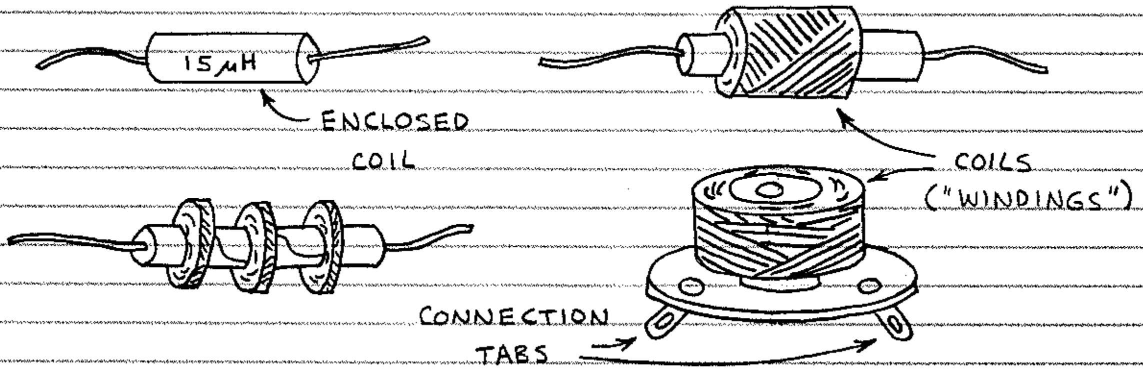
15 μH
ENCLOSED COIL
COILS ("WINDINGS")
CONNECTION TABS

TRANSFORMER. SO IMPORTANT WE'LL DEVOTE A COMPLETE SECTION TO IT.

☐ APPLICATIONS FOR COILS — IN ADDITION TO THOSE ALREADY DESCRIBED, COILS ARE USED IN FILTERS THAT SELECTIVELY PASS SPECIFIC FREQUENCY BANDS.

☐ CAUTION! A HIGH VOLTAGE PULSE CAN BE PRODUCED IN A CHOKE WHEN THE CURRENT FLOWING THROUGH IT IS INTERRUPTED. BE CAREFUL.

TRANSFORMERS

TRANSFORMERS ARE A MAJOR CLASS OF COILS HAVING TWO OR MORE WINDINGS USUALLY WRAPPED AROUND A COMMON CORE MADE FROM LAMINATED IRON SHEETS. HERE'S A SIMPLE TRANSFORMER:

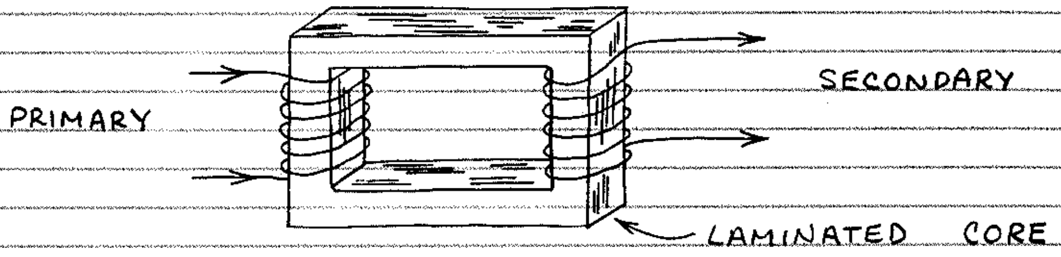

IF THE CURRENT FLOWING THROUGH THE PRIMARY COIL IS FLUCTUATING, THEN A CURRENT WILL BE INDUCED INTO THE SECONDARY WINDING. A STEADY (DC) CURRENT WILL NOT BE TRANSFERRED FROM ONE COIL TO THE OTHER.

☐ HOW THEY WORK — TRANSFORMERS HAVE THE ABILITY TO TRANSFORM VOLTAGE AND CURRENT TO HIGHER OR LOWER LEVELS. THEY DO NOT, OF COURSE, CREATE POWER FROM NOTHING. THEREFORE, IF A TRANSFORMER BOOSTS THE VOLTAGE OF A SIGNAL, IT REDUCES ITS CURRENT. AND IF IT CUTS THE VOLTAGE OF A SIGNAL, IT RAISES ITS CURRENT. IN OTHER WORDS... THE POWER FLOWING FROM A TRANSFORMER CANNOT EXCEED THE INCOMING POWER!

☐ TURNS RATIO — THE RATIO OF PRIMARY TO SECONDARY TURNS DETERMINES A TRANSFORMER'S VOLTAGE RATIO...

1:1 RATIO.

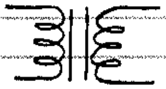

THE PRIMARY VOLTAGE AND CURRENT ARE TRANSFERRED UNALTERED TO THE SECONDARY. OFTEN CALLED AN ISOLATION TRANSFORMER.

STEP-UP.

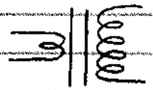

THE VOLTAGE IS INCREASED BY THE TURNS RATIO. THUS A 1:5 TURNS RATIO WILL BOOST 5-VOLTS AT THE PRIMARY INTO 25-VOLTS AT THE SECONDAR

STEP-DOWN.

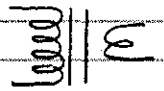

THE VOLTAGE IS REDUCED BY THE TURNS RATIO. THUS A 5:1 TURNS RATIO WILL DROP 25-VOLTS AT THE PRIMARY TO 5-VOLTS AT THE SECONDARY.

☐ TRANSFORMER TYPES AND APPLICATIONS — HERE ARE SOME OF THE MAJOR TRANSFORMER TYPES:

ISOLATION.

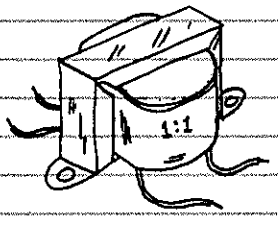

USED TO ISOLATE DIFFERENT PARTS OF A CIRCUIT AND TO PROVIDE PROTECTION FROM ELECTRICAL SHOCK.

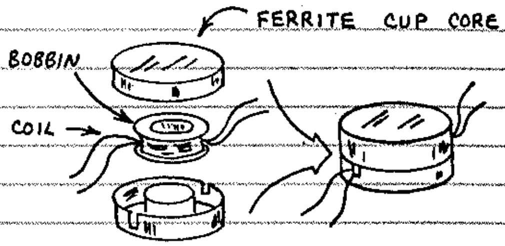

STANDARD 1:1
ISOLATION

MINIATURE 1:1
ISOLATION

POWER CONVERSION.

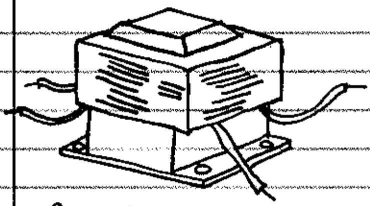

OFTEN USED TO REDUCE POWER LINE VOLTAGE TO USABLE LEVEL.

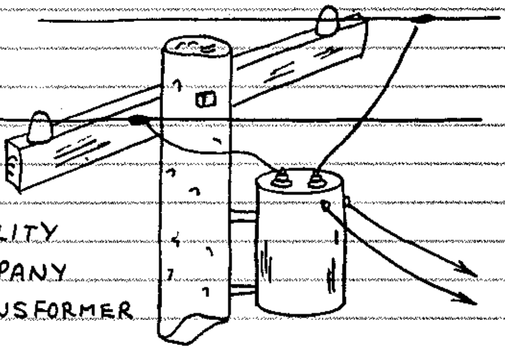

UTILITY
COMPANY
TRANSFORMER

POWER
TRANSFORMER

HIGH - VOLTAGE.

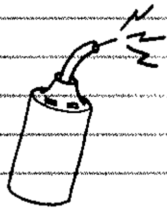

AUTOMOTIVE
IGNITION
COIL

USED TO PRODUCE IGNITION SPARKS IN GASOLINE ENGINES. ALSO USED TO POWER TV PICTURE TUBES, SOME LASERS, NEON LIGHTS, ETC.

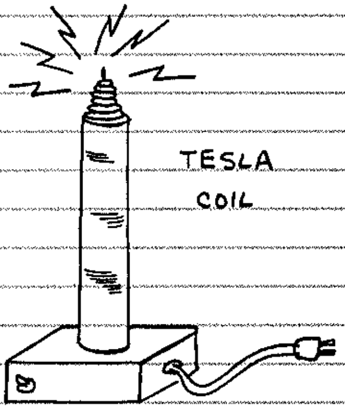

TESLA
COIL

AUDIO.

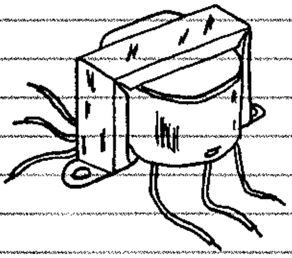

USED TO MATCH THE IMPEDANCE* OF AN AMPLIFIER TO THAT OF A MICROPHONE, SPEAKER OR OTHER DEVICE.

*OPPOSITION TO THE FLOW OF ALTERNATING CURRENT.

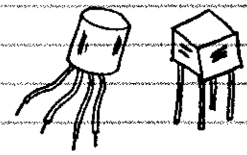

MINIATURE

TAPPED PRIMARY
AND SECONDARY
WINDINGS

NOTE: LEADS OF TRANSFORMERS ARE COLOR CODED.

3. SEMICONDUCTORS

THE MOST EXCITING AND IMPORTANT ELECTRONIC COMPONENTS ARE MADE FROM CRYSTALS CALLED SEMICONDUCTORS. DEPENDING ON CERTAIN CONDITIONS, A SEMICONDUCTOR CAN ACT LIKE A CONDUCTOR OR AN INSULATOR.

SILICON

THERE ARE MANY DIFFERENT SEMICONDUCTING MATERIALS, BUT SILICON, THE MAIN INGREDIENT OF SAND, IS THE MOST POPULAR.

SILICON ATOM

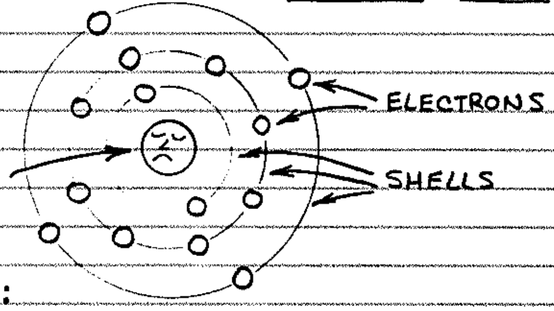

A SILICON ATOM HAS BUT FOUR ELECTRONS IN ITS OUTERMOST SHELL, BUT IT WOULD LIKE TO HAVE EIGHT. THEREFORE, A SILICON ATOM WILL LINK UP WITH FOUR OF ITS NEIGHBORS TO SHARE ELECTRONS:

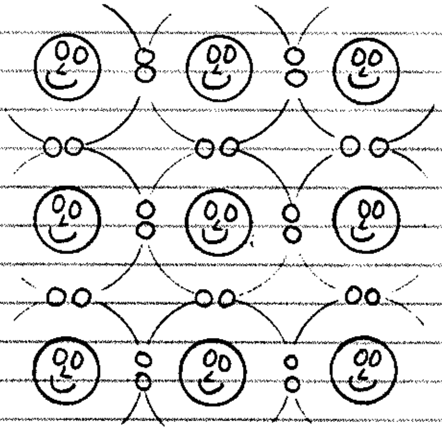

A CLUSTER OF SILICON ATOMS SHARING OUTER ELECTRONS FORMS A REGULAR ARRANGEMENT CALLED A CRYSTAL.

← THIS IS A MAGNIFIED VIEW OF A SILICON CRYSTAL. TO KEEP THINGS SIMPLE, ONLY THE OUTER ELECTRONS OF EACH ATOM ARE SHOWN.

SILICON FORMS 27.7 % OF THE EARTH'S CRUST! ONLY OXYGEN IS MORE COMMON. IT'S NEVER FOUND IN THE PURE STATE. WHEN PURIFIED, IT'S DARK GRAY IN COLOR.

SILICON AND DIAMOND SHARE THE SAME CRYSTAL STRUCTURE AND OTHER PROPERTIES. BUT SILICON IS NOT TRANSPARENT.

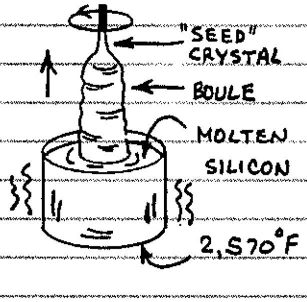

SILICON CAN BE GROWN INTO BIG CRYSTALS. IT'S CUT INTO WAFERS FOR MAKING ELECTRONIC PARTS.

☐ SILICON RECIPES— PURE SILICON ISN'T VERY USEFUL. THAT'S WHY SILICON MAKERS SPICE UP THEIR SILICON RECIPES WITH A DASH OF PHOSPHORUS, BORON OR OTHER GOODIES. THIS IS CALLED DOPING THE SILICON. WHEN GROWN INTO CRYSTALS, DOPED SILICON HAS VERY USEFUL ELECTRONIC PROPERTIES!

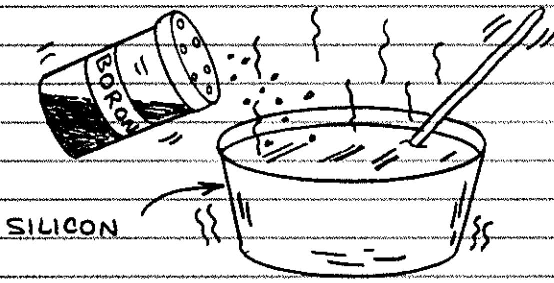

☐ P & N SPICED SILICON LOAF— BORON, PHOSPHORUS AND CERTAIN OTHER ATOMS CAN JOIN WITH SILICON ATOMS TO FORM CRYSTALS. HERS'S THE CATCH: A BORON ATOM HAS ONLY THREE ELECTRON IN ITS OUTER SHELL. AND A PHOSPHORUS ATOM HAS FIVE ELECTRONS IN ITS OUTER SHELL. SILICON WITH EXTRA PHOSPHORUS ELECTRONS IS CALLED N-TYPE SILICON (N = NEGATIVE). SILICON WITH ELECTRON DEFICIENT BORON ATOMS IS CALLED P-TYPE SILICON (P = POSITIVE).

☐ P-TYPE SILICON—
A BORON ATOM IN A CLUSTER OF SILICON ATOMS LEAVES A VACANT ELECTRON OPENING CALLED A HOLE. IT'S POSSIBLE FOR AN ELECTRON FROM A NEARBY ATOM TO "FALL" INTO THE HOLE. THEREFORE, THE HOLE HAS MOVED TO A NEW LOCATION. REMEMBER, HOLES CAN MOVE THROUGH SILICON (JUST AS BUBBLES MOVE THROUGH WATER).

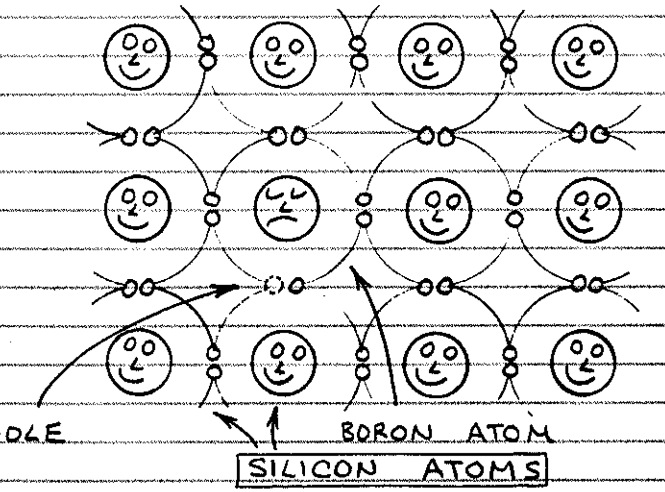

☐ N-TYPE SILICON—
A PHOSPHORUS ATOM IN A CLUSTER OF SILICON ATOMS DONATES AN EXTRA ELECTRON. THIS EXTRA ELECTRON CAN MOVE THROUGH THE CRYSTAL WITH COMPARATIVE EASE. IN OTHER WORDS, N-TYPE SILICON CAN CARRY AN ELECTRICAL CURRENT. BUT SO CAN P-TYPE SILICON! HOLES "CARRY" THE CURRENT.

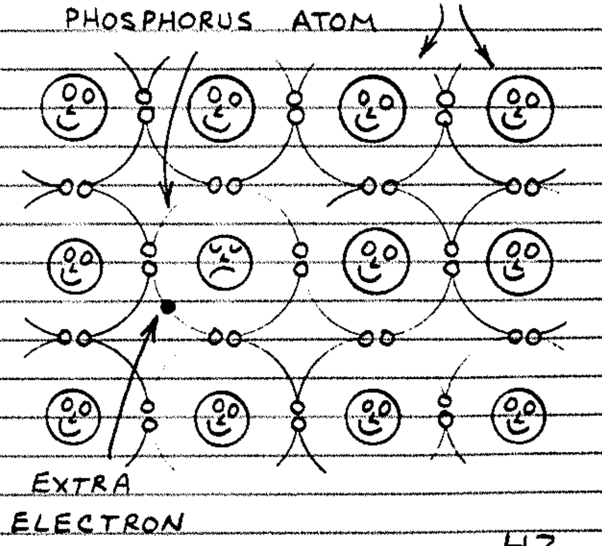

THE DIODE

BOTH P-TYPE AND N-TYPE SILICON CONDUCT ELECTRICITY. THE RESISTANCE OF BOTH TYPES IS DETERMINED BY THE PROPORTION OF HOLES OR SURPLUS ELECTRONS. THEREFORE BOTH TYPES CAN FUNCTION AS RESISTORS. AND THEY WILL CONDUCT ELECTRICITY IN ANY DIRECTION.

BY FORMING SOME P-TYPE SILICON IN A CHIP OF N-TYPE SILICON, ELECTRONS WILL FLOW THROUGH THE SILICON IN ONLY ONE DIRECTION. THIS IS THE PRINCIPLE OF THE DIODE. THE P-N INTERFACE IS CALLED THE PN JUNCTION.

☐ HOW THE DIODE WORKS — HERE'S A SIMPLIFIED EXPLANATION OF HOW A DIODE CONDUCTS ELECTRICITY IN ONE DIRECTION (FORWARD) WHILE BLOCKING THE FLOW OF CURRENT IN THE OPPOSITE DIRECTION (REVERSE).

FORWARD BIAS

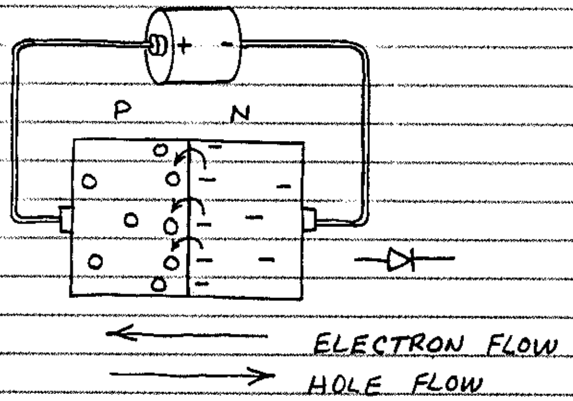

REVERSE BIAS

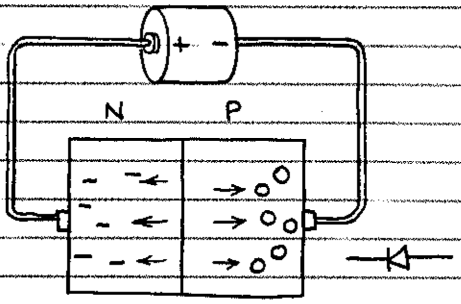

NO CURRENT FLOW

HERE THE CHARGE FROM THE BATTERY REPELS HOLES AND ELECTRONS TOWARD THE JUNCTION. IF THE VOLTAGE EXCEEDS 0.6-VOLT (SILICON), THEN ELECTRONS WILL CROSS THE JUNCTION AND COMBINE WITH HOLES. A CURRENT THEN FLOWS.

HERE THE CHARGE FROM THE BATTERY ATTRACTS HOLES AND ELECTRONS AWAY FROM THE JUNCTION. THEREFORE, NO CURRENT CAN FLOW.

☐ A TYPICAL DIODE — DIODES ARE COMMONLY ENCLOSED IN SMALL GLASS CYLINDERS. A DARK BAND MARKS THE CATHODE TERMINAL. THE OPPOSITE TERMINAL IS THE ANODE

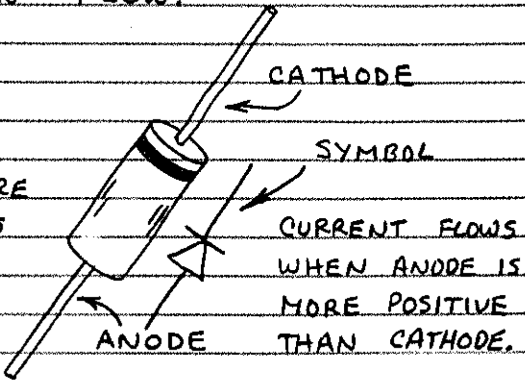

CURRENT FLOWS WHEN ANODE IS MORE POSITIVE THAN CATHODE.

☐ DIODE OPERATION — YOU ALREADY KNOW A DIODE IS LIKE AN ELECTRONIC ONE-WAY VALVE. IT'S IMPORTANT TO UNDERSTAND SOME ADDITIONAL ASPECTS OF DIODE OPERATION. HERE ARE SOME KEY ONES:

1. A DIODE WILL NOT CONDUCT UNTIL THE FORWARD VOLTAGE REACHES A CERTAIN THRESHOLD POINT. FOR SILICON DIODES THIS VOLTAGE IS ABOUT 0.6-VOLT.

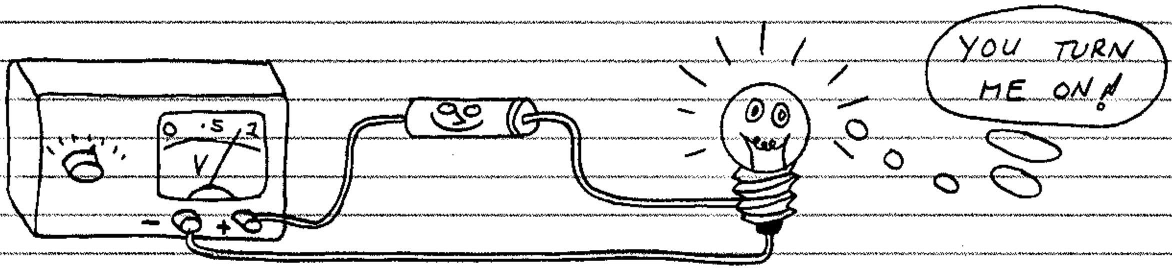

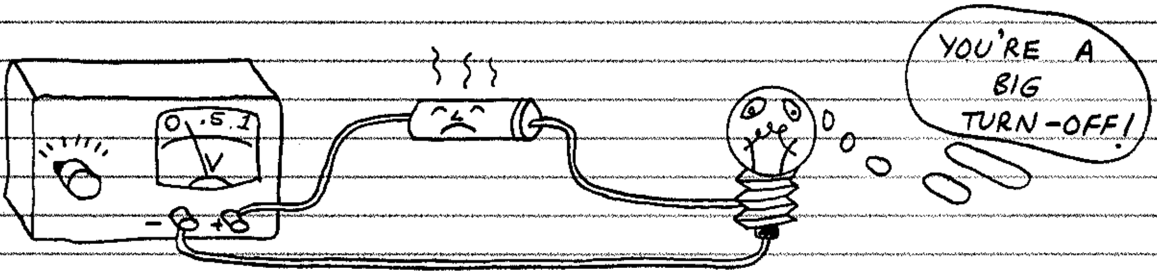

2. IF THE FORWARD CURRENT BECOMES EXCESSIVE, THE SEMICONDUCTOR CHIP MAY CRACK OR MELT! AND THE CONTACTS MAY SEPARATE. IF THE CHIP MELTS, THE DIODE MAY SUDDENLY CONDUCT IN BOTH DIRECTIONS. THE RESULTING HEAT MAY VAPORIZE THE CHIP!

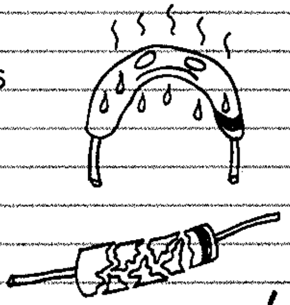

3. TOO MUCH REVERSE VOLTAGE WILL CAUSE A DIODE TO CONDUCT IN THE WRONG DIRECTION. SINCE THIS VOLTAGE IS FAIRLY HIGH, THE SUDDEN CURRENT SURGE MAY ZAP THE DIODE.

☐ SUMMING UP DIODE OPERATION — THIS GRAPH SUMS UP DIODE OPERATION. (IT'S APPROXIMATE.)

V_F = FORWARD VOLTAGE
V_R = REVERSE VOLTAGE
I_F = FORWARD CURRENT
I_R = REVERSE CURRENT

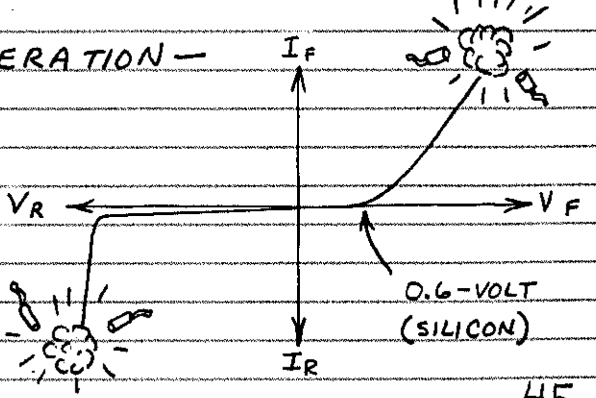

☐ TYPES OF DIODES — MANY DIFFERENT KINDS OF DIODES ARE AVAILABLE. HERE ARE SOME OF THE MAJOR TYPES:

SMALL SIGNAL.

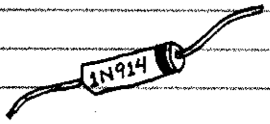

SMALL SIGNAL DIODES ARE USED TO TRANSFORM LOW CURRENT AC TO DC, DETECT (DEMODULATE) RADIO SIGNALS, MULTIPLY VOLTAGE, PERFORM LOGIC, ABSORB VOLTAGE SPIKES, ETC.

POWER RECTIFIER.

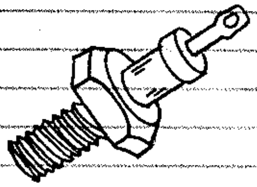

FUNCTIONALLY IDENTICAL TO SMALL SIGNAL DIODES, POWER RECTIFIERS CAN HANDLE *MUCH* MORE CURRENT. THEY ARE INSTALLED IN LARGE METAL PACKAGES THAT SOAK UP EXCESS HEAT AND TRANSFER IT TO A METAL HEAT SINK. USED MAINLY IN POWER SUPPLIES.

ZENER.

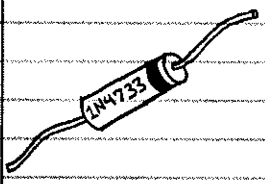

THE ZENER DIODE IS DESIGNED TO HAVE A SPECIFIC *REVERSE BREAKDOWN* (CONDUCTION) *VOLTAGE*. THIS MEANS ZENER DIODES CAN FUNCTION LIKE A VOLTAGE SENSITIVE SWITCH. ZENER DIODES HAVING BREAKDOWN VOLTAGES (V_z) OF FROM ABOUT 2-VOLTS TO 200-VOLTS ARE AVAILABLE.

LIGHT-EMITTING.

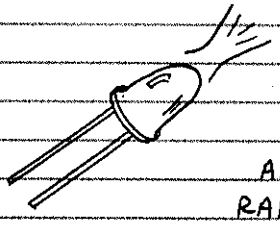

ALL DIODES EMIT SOME ELECTROMAGNETIC RADIATION WHEN FORWARD BIASED. DIODES MADE FROM CERTAIN SEMICONDUCTORS (LIKE GALLIUM ARSENIDE PHOSPHIDE) EMIT *CONSIDERABLY* MORE RADIATION THAN SILICON DIODES. THEY'RE CALLED LIGHT-EMITTING DIODES (LEDs).

PHOTODIODE.

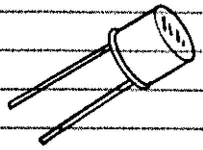

ALL DIODES RESPOND TO SOME DEGREE WHEN ILLUMINATED BY LIGHT. DIODES DESIGNED SPECIFICALLY TO DETECT LIGHT ARE CALLED PHOTODIODES. THEY INCLUDE A GLASS OR PLASTIC WINDOW THROUGH WHICH THE LIGHT ENTERS. OFTEN THEY HAVE A LARGE, EXPOSED JUNCTION REGION. SILICON MAKES GOOD PHOTODIODES.

HOW DIODES ARE USED

IN CHAPTER 9 YOU'LL SEE HOW VARIOUS TYPES OF DIODES ARE USED IN MANY APPLICATIONS. FOR NOW HERE ARE TWO OF THE MOST IMPORTANT ROLES FOR SMALL SIGNAL DIODES AND RECTIFIERS:

☐ HALF-WAVE RECTIFIER

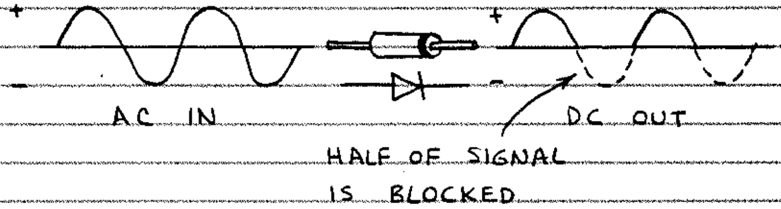

AN UNDULATING (AC) SIGNAL (OR VOLTAGE) IS RECTIFIED INTO A SINGLE POLARITY (DC) SIGNAL (OR VOLTAGE).

☐ FULL-WAVE RECTIFIER

P. 101

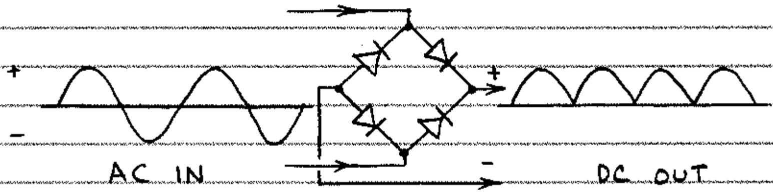

THIS 4-DIODE "NETWORK" (OR BRIDGE RECTIFIER) RECTIFIES BOTH HALVES OF AN AC SIGNAL.

MORE ABOUT THE DIRECTION OF CURRENT FLOW

AN ELECTRICAL CURRENT IS THE MOVEMENT OF ELECTRONS THROUGH A CONDUCTOR OR SEMICONDUCTOR. SINCE ELECTRONS MOVE FROM A NEGATIVELY CHARGED TO A POSITIVELY CHARGED REGION, WHY DOES THE ARROWHEAD IN A DIODE SYMBOL POINT IN THE OPPOSITE DIRECTION? THERE ARE TWO REASONS:

1. BEGINNING WITH BENJAMIN FRANKLIN, IT WAS TRADITIONALLY ASSUMED ELECTRICITY FLOWS FROM A POSITIVELY CHARGED TO A NEGATIVELY CHARGED REGION. THE DISCOVERY OF THE ELECTRON CORRECTED THAT. (BUT MOST ELECTRICAL CIRCUIT DIAGRAMS TODAY STILL FOLLOW THE OLD TRADITION IN WHICH THE POSITIVE POWER SUPPLY CONNECTION IS PLACED ABOVE THE NEGATIVE CONNECTION AS IF GRAVITY SOMEHOW INFLUENCES THE FLOW OF A CURRENT.)

2. IN A SEMICONDUCTOR, AS SHOWN ON PAGE 44, HOLES FLOW IN THE DIRECTION OPPOSITE THAT OF ELECTRON FLOW. IT'S THEREFORE COMMON TO REFER TO POSITIVE CURRENT FLOW IN SEMICONDUCTORS.

FOR ACCURACY, IN THIS BOOK "CURRENT FLOW" REFERS TO ELECTRON FLOW. BUT WE'RE STUCK WITH SYMBOLS THAT INDICATE HOLE FLOW.

THE TRANSISTOR

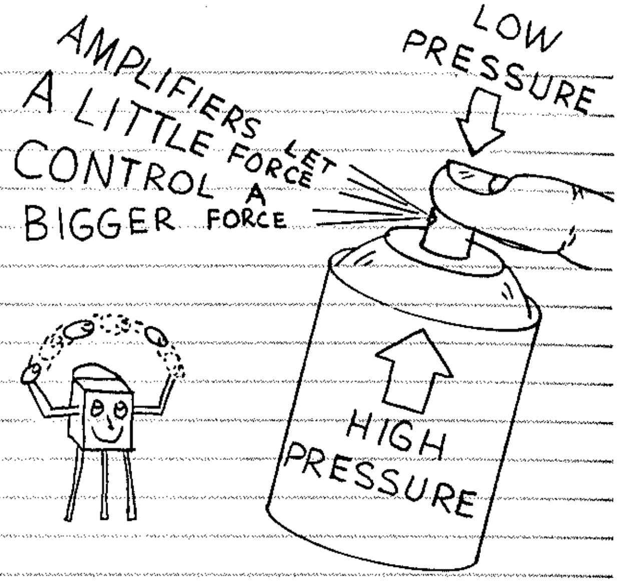
AMPLIFIERS LET A LITTLE FORCE CONTROL A BIGGER FORCE
LOW PRESSURE
HIGH PRESSURE

TRANSISTORS ARE SEMICONDUCTOR DEVICES WITH THREE LEADS. A VERY SMALL CURRENT OR VOLTAGE AT ONE LEAD CAN CONTROL A MUCH LARGER CURRENT FLOWING THROUGH THE OTHER TWO LEADS. THIS MEANS TRANSISTORS CAN BE USED AS AMPLIFIERS AND SWITCHES. THERE ARE TWO MAIN FAMILIES OF TRANSISTORS: BIPOLAR AND FIELD-EFFECT.

BIPOLAR TRANSISTORS

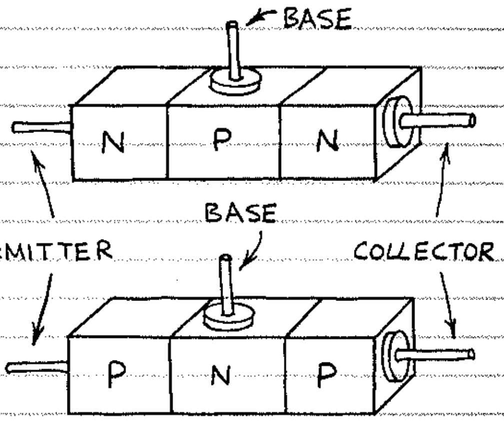
BASE
N
P
N
BASE
EMITTER
COLLECTOR
P
N
P

ADD A SECOND JUNCTION TO A PN JUNCTION DIODE AND YOU GET A 3-LAYER SILICON SANDWICH. THE SANDWICH CAN BE EITHER NPN OR PNP. EITHER WAY, THE MIDDLE LAYER ACTS LIKE A FAUCET OR GATE THAT CONTROLS THE CURRENT MOVING THROUGH THE THREE LAYERS.

☐ BIPOLAR TRANSISTOR OPERATION — THE THREE LAYERS OF A BIPOLAR TRANSISTOR ARE THE EMITTER, BASE AND COLLECTOR. THE BASE IS VERY THIN AND HAS FEWER DOPING ATOMS THAN THE EMITTER AND COLLECTOR. THEREFORE A VERY SMALL EMITTER-BASE CURRENT WILL CAUSE A MUCH LARGER EMITTER-COLLECTOR CURRENT TO FLOW.

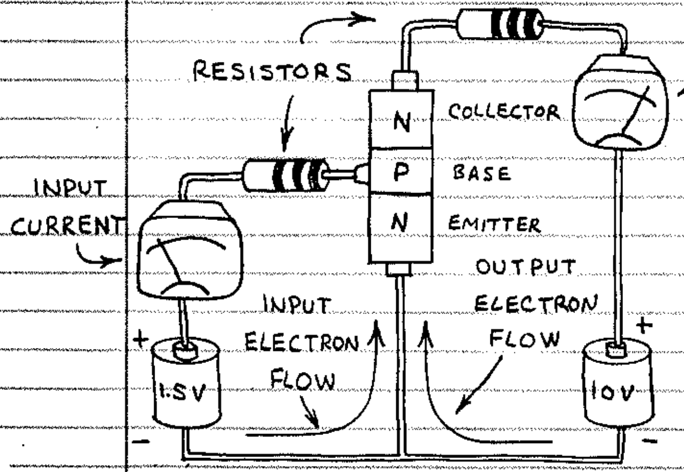
RESISTORS
N
P
N
COLLECTOR
BASE
EMITTER
AMPLIFIED OUTPUT CURRENT
INPUT CURRENT
INPUT ELECTRON FLOW
OUTPUT ELECTRON FLOW
+
-
1.5V
+
-
10V

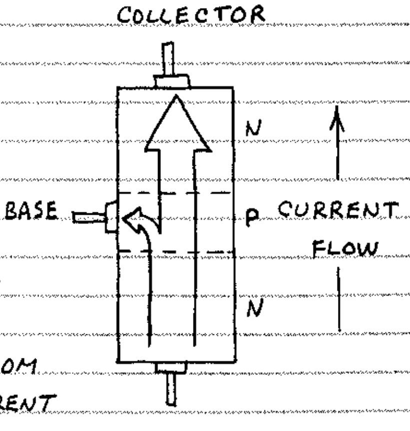
COLLECTOR
N
P
N
BASE
CURRENT FLOW
EMITTER

THE RESISTORS PROTECT THE TRANSISTOR FROM TOO MUCH CURRENT (WHICH CAN CAUSE EXCESSIVE HEAT).

☐ MORE ABOUT BIPOLAR TRANSISTOR OPERATION — DIODES AND TRANSISTORS SHARE SEVERAL KEY FEATURES:

1. THE BASE-EMITTER JUNCTION (OR DIODE) WILL NOT CONDUCT UNTIL THE FORWARD VOLTAGE EXCEEDS 0.6-VOLT.

2. TOO MUCH CURRENT WILL CAUSE A TRANSISTOR TO BECOME HOT AND OPERATE IMPROPERLY. IF A TRANSISTOR IS HOT WHEN TOUCHED, DISCONNECT THE POWER TO IT!

3. TOO MUCH CURRENT OR VOLTAGE MAY DAMAGE OR PERMANENTLY DESTROY THE SEMICONDUCTOR CHIP THAT FORMS A TRANSISTOR. IF THE CHIP ISN'T HARMED, ITS TINY CONNECTION WIRES MAY MELT OR SEPARATE FROM THE CHIP. NEVER CONNECT A TRANSISTOR BACKWARDS!

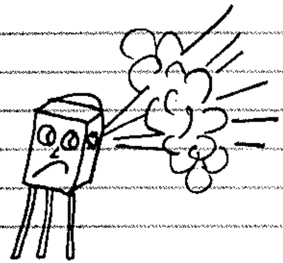

☐ KINDS OF TRANSISTORS — MANY DIFFERENT KINDS OF TRANSISTORS ARE AVAILABLE. HERE ARE EXAMPLES OF THE MOST IMPORTANT:

SMALL SIGNAL AND SWITCHING.

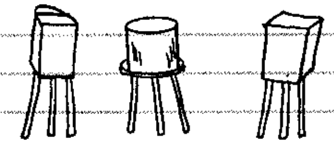

SMALL SIGNAL TRANSISTORS ARE USED TO AMPLIFY LOW LEVEL SIGNALS. SWITCHING TRANSISTORS ARE DESIGNED TO BE OPERATED FULLY ON OR OFF. SOME TRANSISTORS CAN BOTH AMPLIFY AND SWITCH EQUALLY WELL.

POWER.

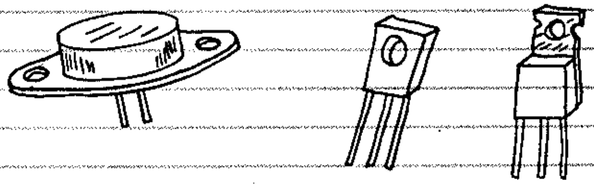

POWER TRANSISTORS ARE USED IN HIGH POWER AMPLIFIERS AND POWER SUPPLIES. LARGE SIZE AND EXPOSED METAL SURFACES KEEP THEM COOL.

HIGH-FREQUENCY.

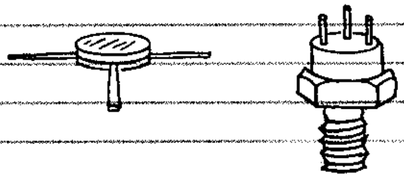

HIGH-FREQUENCY TRANSISTORS OPERATE AT RADIO, TELEVISION AND MICROWAVE FREQUENCIES. THE BASE REGION IS VERY THIN AND THE ACTUAL CHIP IS VERY SMALL.

☐ BIPOLAR TRANSISTOR SYMBOLS — ARROWS POINT IN DIRECTION OF HOLE FLOW.

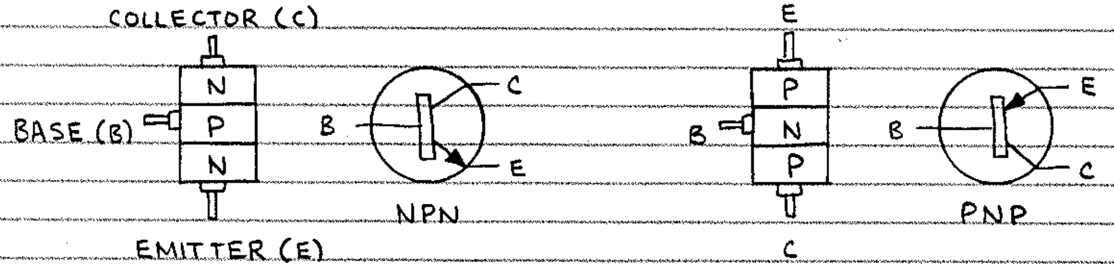
COLLECTOR (C)
BASE (B)
EMITTER (E)
N
P
N
B
C
E
NPN
E
B
C
P
N
P
B
E
C
PNP

HOW BIPOLAR TRANSISTORS ARE USED

WHEN THE BASE OF AN NPN TRANSISTOR IS GROUNDED (0 VOLTS), NO CURRENT FLOWS FROM THE EMITTER TO THE COLLECTOR (THE TRANSISTOR IS "OFF"). IF THE BASE IS FORWARD-BIASED BY AT LEAST 0.6 VOLT, A CURRENT WILL FLOW FROM THE EMITTER TO THE COLLECTOR (THE TRANSISTOR IS "ON"). WHEN OPERATED IN ONLY THESE TWO MODES, THE TRANSISTOR FUNCTIONS AS A SWITCH. IF THE BASE IS FORWARD-BIASED, THE EMITTER-COLLECTOR CURRENT WILL FOLLOW VARIATIONS IN A MUCH SMALLER BASE CURRENT. THE TRANSISTOR THEN FUNCTIONS AS AN AMPLIFIER. THIS DISCUSSION APPLIES TO A TRANSISTOR IN WHICH THE EMITTER IS THE GROUND CONNECTION FOR BOTH THE INPUT AND OUTPUT AND IS CALLED THE COMMON-EMITTER CIRCUIT. SOME SIMPLIFIED COMMON-EMITTER CIRCUITS ARE SHOWN BELOW. SO YOU CAN SEE HOW THEY ARE USED IN REAL CIRCUITS, EACH EXAMPLE REFERS TO A TYPICAL WORKING APPLICATION IN CHAPTER 9.

P. 92

☐ A BIPOLAR TRANSISTOR SWITCH

P. 107

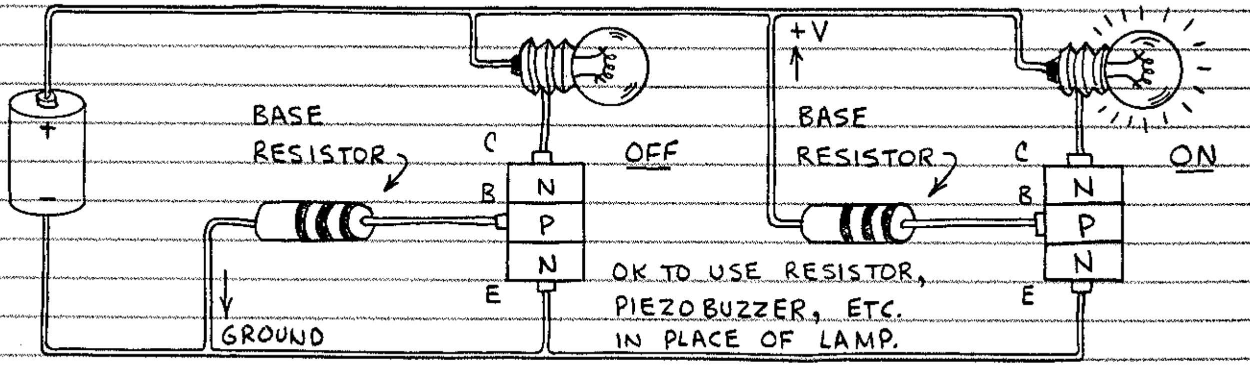
+
−
BASE RESISTOR
C
B
E
N
P
N
OFF
GROUND
+V
BASE RESISTOR
C
B
E
N
P
N
ON
OK TO USE RESISTOR, PIEZOBUZZER, ETC. IN PLACE OF LAMP.

ONLY TWO INPUTS ARE POSSIBLE: GROUND (0 VOLTS) AND THE POSITIVE BATTERY VOLTAGE (+V). THEREFORE THE TRANSISTOR IS OFF OR ON. A TYPICAL BASE RESISTANCE IS 5,000 TO 10,000 OHMS. (IF THE RESISTOR IS REPLACED BY A WIRE, THE LAMP CAN BE SWITCHED ON OR OFF FROM A CONSIDERABLE DISTANCE.)

☐ A BIPOLAR TRANSISTOR DC AMPLIFIER — THE VARIABLE RESISTOR FORWARD BIASES THE TRANSISTOR AND CONTROLS THE INPUT (BASE-EMITTER) CURRENT. THE METER INDICATES THE OUTPUT (COLLECTOR – EMITTER) CURRENT. THE SERIES RESISTOR PROTECTS THE METER FROM EXCESSIVE CURRENT.

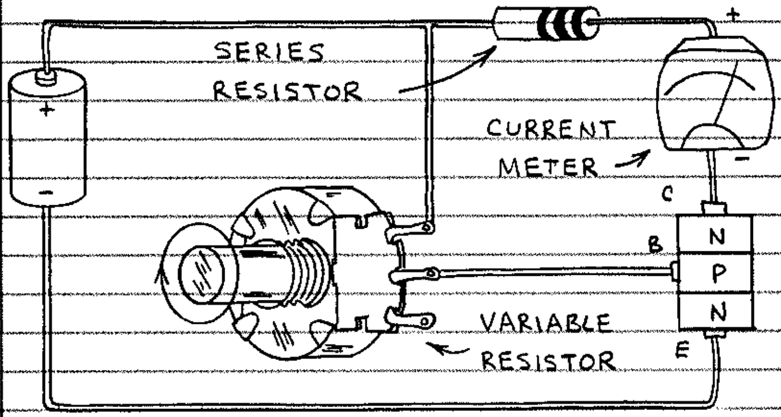
SERIES RESISTOR
CURRENT METER
VARIABLE RESISTOR
+
-
C
B
E
N
P
N

P. 104

IN A WORKING CIRCUIT, THE VARIABLE RESISTOR MAY BE IN SERIES WITH A SECOND COMPONENT HAVING A RESISTANCE THAT VARIES WITH TEMPERATURE, LIGHT, MOISTURE, ETC. (WATER IS THE VARIABLE RESISTANCE IN THE MOISTURE METER ON P. 104.) WHEN THE INPUT SIGNAL CHANGES RAPIDLY, AN AC AMPLIFIER SUCH AS THE ONE BELOW IS USED.

☐ A BIPOLAR TRANSISTOR AC AMPLIFIER — THIS IS THE SIMPLEST OF SEVERAL BASIC AC AMPLIFIERS. THE INPUT CAPACITOR BLOCKS ANY DC IN THE INPUT SIGNAL.

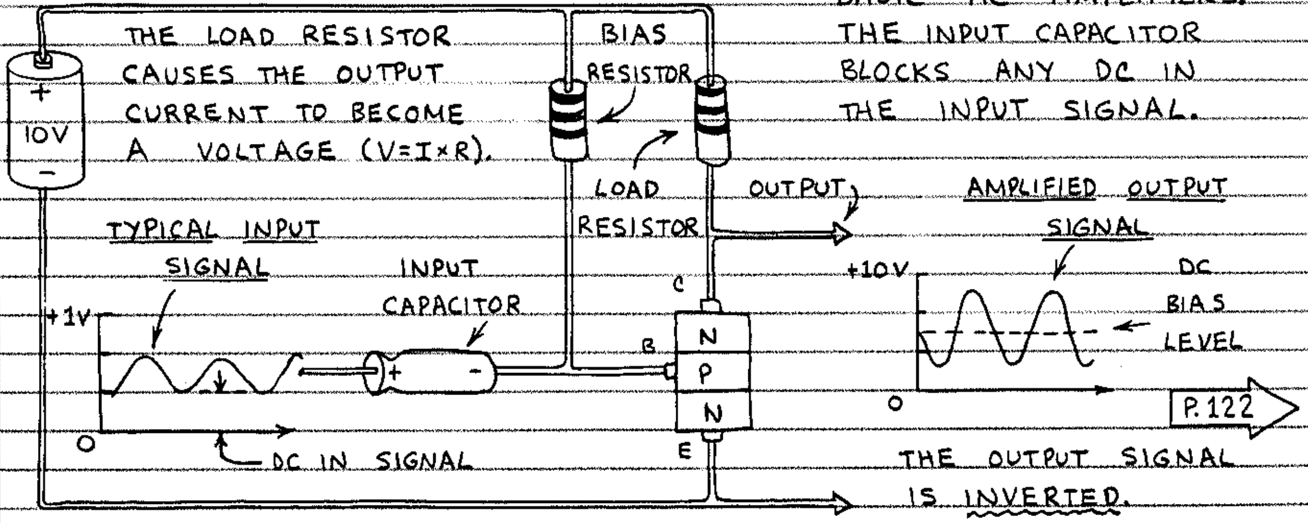
THE LOAD RESISTOR CAUSES THE OUTPUT CURRENT TO BECOME A VOLTAGE (V=I×R).
+
10V
-
BIAS RESISTOR
LOAD RESISTOR
OUTPUT
AMPLIFIED OUTPUT SIGNAL
+10V
DC BIAS LEVEL
0
TYPICAL INPUT SIGNAL
+1V
0
DC IN SIGNAL
INPUT CAPACITOR
+
-
C
B
E
N
P
N

P. 122

THE OUTPUT SIGNAL IS INVERTED.

THE BIAS RESISTOR IS SELECTED TO GIVE AN OUTPUT VOLTAGE OF ABOUT HALF THE BATTERY VOLTAGE. THE AMPLIFIED SIGNAL "RIDES" ON THIS STEADY OUTPUT VOLTAGE AND VARIES ABOVE AND BELOW IT. (WITHOUT THE BIAS RESISTOR, ONLY THE POSITIVE HALF OF THE INPUT SIGNAL ABOVE 0.6 VOLT (SEE P. 45) WILL BE AMPLIFIED. THIS WILL CAUSE SEVERE DISTORTION.) TO SEE ONE WAY A WORKING VERSION OF THIS AMPLIFIER IS USED, TURN NOW TO P. 122 AND LOOK AT THE OUTPUT SECTION OF THE LIGHTWAVE TRANSMITTER.

FIELD-EFFECT TRANSISTORS

FIELD-EFFECT TRANSISTORS (OR FETs) HAVE BECOME MORE IMPORTANT THAN BIPOLAR TRANSISTORS. THEY ARE EASY TO MAKE AND REQUIRE LESS SILICON. THERE ARE TWO MAJOR FET FAMILIES, JUNCTION AND METAL-OXIDE-SEMICONDUCTOR. IN BOTH KINDS AN OUTPUT CURRENT IS CONTROLLED BY A SMALL INPUT VOLTAGE AND PRACTICALLY NO INPUT CURRENT!

JUNCTION FETs

THE TWO MAIN KINDS OF FETs ARE N-CHANNEL AND P-CHANNEL. THE CHANNEL IS LIKE A SILICON RESISTOR THAT CONDUCTS CURRENT MOVING FROM THE SOURCE TO THE DRAIN. A VOLTAGE AT THE GATE INCREASES THE CHANNEL RESISTANCE AND REDUCES THE DRAIN-SOURCE CURRENT. THEREFORE THE FET CAN BE USED AS AN AMPLIFIER OR A SWITCH.

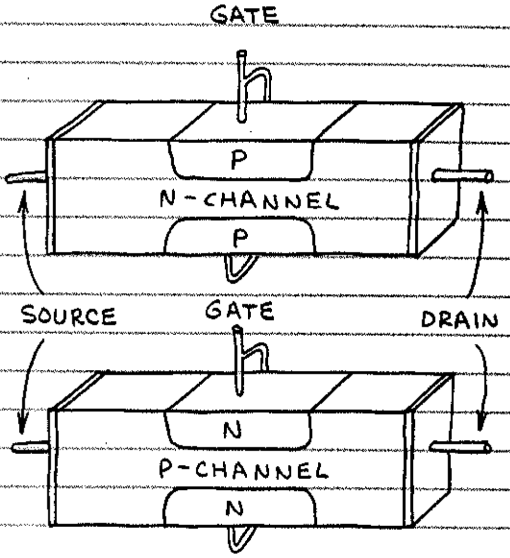
GATE
P
N-CHANNEL
P
SOURCE
GATE
DRAIN
N
P-CHANNEL
N

☐ JUNCTION FET OPERATION— THE ARRANGEMENT BELOW SHOWS HOW AN N-CHANNEL FET WORKS. A NEGATIVE GATE VOLTAGE CREATES TWO HIGH RESISTANCE REGIONS (THE FIELD) IN THE CHANNEL ADJACENT TO THE P-TYPE SILICON. MORE GATE VOLTAGE WILL CAUSE THE FIELDS TO MERGE TOGETHER AND COMPLETELY BLOCK THE CURRENT. THE GATE-CHANNEL RESISTANCE IS VERY HIGH.

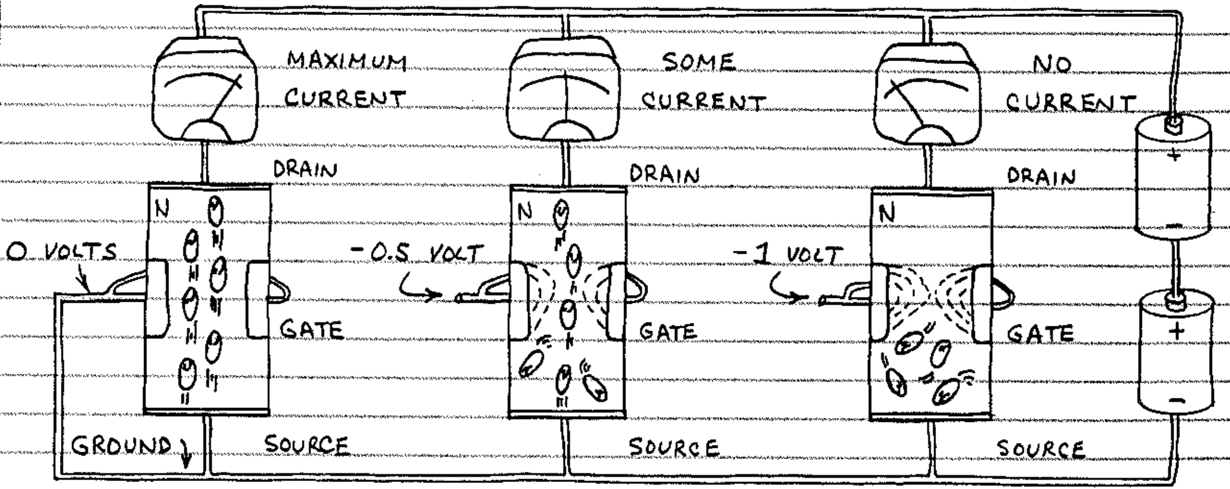
MAXIMUM CURRENT
SOME CURRENT
NO CURRENT
DRAIN
DRAIN
DRAIN
N
N
N
0 VOLTS
-0.5 VOLT
-1 VOLT
GATE
GATE
GATE
+
-
+
-
GROUND
SOURCE
SOURCE
SOURCE

☐ MORE ABOUT JUNCTION FETs — SINCE THEY ARE VOLTAGE CONTROLLED, JUNCTION FETs (OR JFETs) HAVE IMPORTANT ADVANTAGES OVER CURRENT-CONTROLLED BIPOLAR TRANSISTORS:

1. THE GATE-CHANNEL RESISTANCE OF A JFET IS VERY HIGH (MILLIONS OF OHMS). THEREFORE THE JFET HAS LITTLE OR NO EFFECT ON EXTERNAL COMPONENTS OR CIRCUITS CONNECTED TO ITS GATE.

2. THE VERY HIGH GATE-CHANNEL RESISTANCE MEANS PRACTICALLY NO CURRENT FLOWS IN THE GATE CIRCUIT. (WHY IS THE RESISTANCE SO HIGH? THE GATE AND CHANNEL FORM A DIODE. SO LONG AS THE INPUT SIGNAL REVERSE BIASES THIS DIODE, THE GATE HAS VERY HIGH INPUT RESISTANCE.)

3. LIKE BIPOLAR TRANSISTORS, JFETs CAN BE DAMAGED OR DESTROYED BY EXCESSIVE CURRENT OR VOLTAGE.

☐ KINDS OF JUNCTION FETs — JFETs ARE USED IN MANY DIFFERENT APPLICATIONS. SINCE THEY CANNOT BE USED FOR HIGH POWER ROLES, MOST ARE INSTALLED IN SMALL PLASTIC OR METAL PACKAGES. HERE ARE THE MAIN CATEGORIES:

SMALL SIGNAL AND SWITCHING.

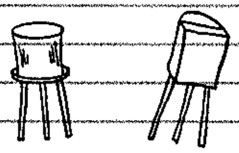

SMALL SIGNAL JFETs ARE USED AT THE INPUT STAGE OF AMPLIFIERS TO PROVIDE A HIGH RESISTANCE INPUT. THEY ARE ALSO USED AS SWITCHES.

HIGH FREQUENCY.

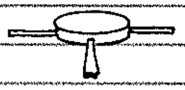

HIGH FREQUENCY JFETS ARE USED TO AMPLIFY OR PRODUCE HIGH FREQUENCY SIGNALS.

☐ JUNCTION FET SYMBOLS — GATES INTERNALLY CONNECTED.

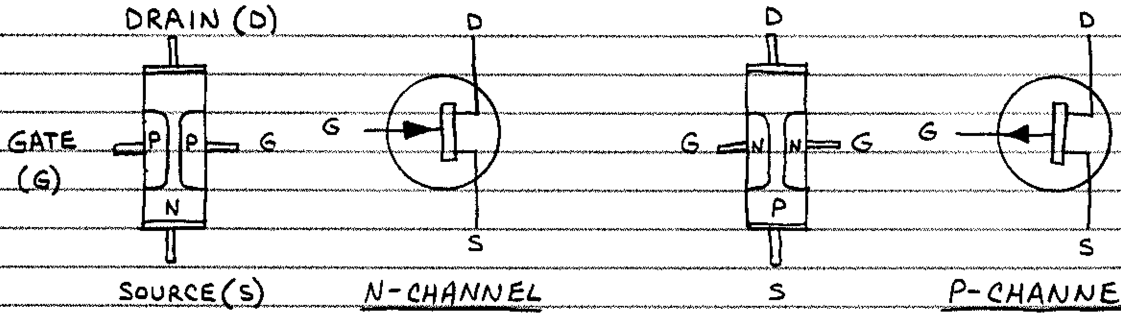

N-CHANNEL

P-CHANNEL

METAL-OXIDE-SEMICONDUCTOR FETs

THE METAL-OXIDE-SEMICONDUCTOR FET (OR MOSFET) HAS BECOME THE MOST IMPORTANT TRANSISTOR. MOST MICROCOMPUTER AND MEMORY INTEGRATED CIRCUITS ARE ARRAYS OF THOUSANDS OF MOSFETs ON A SMALL SLIVER OF SILICON. WHY? MOSFETs ARE EASY TO MAKE, THEY CAN BE VERY SMALL, AND SOME MOSFET CIRCUITS CONSUME NEGLIGIBLE POWER. NEW KINDS OF POWER MOSFETs ARE ALSO VERY USEFUL.

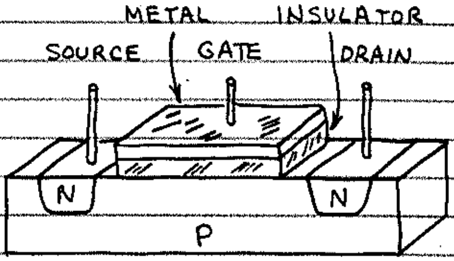
METAL
INSULATOR
SOURCE
GATE
DRAIN
N
N
P

N-MOSFET

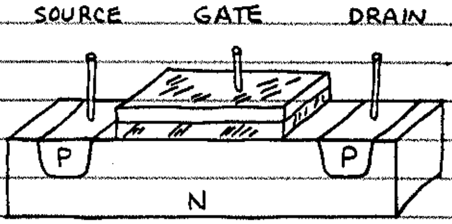
SOURCE
GATE
DRAIN
P
P
N

P- MOSFET

☐ MOSFET OPERATION— ALL MOSFETs ARE N-TYPE OR P-TYPE. UNLIKE THE JUNCTION FET, THE GATE OF A MOSFET HAS NO ELECTRICAL CONTACT WITH THE SOURCE AND DRAIN. A GLASS-LIKE LAYER OF SILICON-DIOXIDE (AN INSULATOR) SEPARATES THE GATE'S METAL CONTACT FROM THE REST OF THE TRANSISTOR.

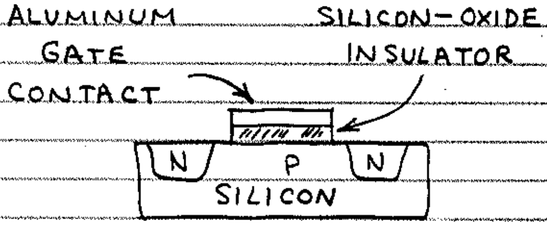
ALUMINUM GATE CONTACT
SILICON-OXIDE INSULATOR
N
P
N
SILICON

A POSITIVE GATE VOLTAGE ATTRACTS ELECTRONS TO THE REGION BELOW THE GATE. THIS CREATES A THIN N-TYPE CHANNEL IN THE P-TYPE SILICON BETWEEN THE SOURCE AND DRAIN. CURRENT CAN THEN FLOW THROUGH THE CHANNEL. THE GATE VOLTAGE DETERMINES THE RESISTANCE OF THE CHANNEL.

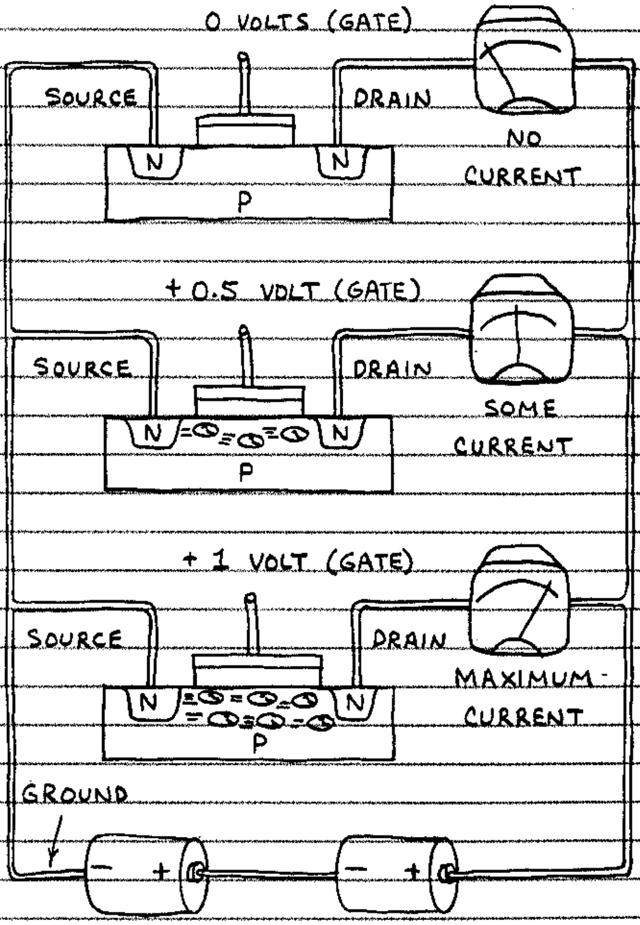
0 VOLTS (GATE)
SOURCE
DRAIN
N
N
P
NO CURRENT
+0.5 VOLT (GATE)
SOURCE
DRAIN
N
N
P
SOME CURRENT
+1 VOLT (GATE)
SOURCE
DRAIN
N
N
P
MAXIMUM CURRENT
GROUND
− +
− +

☐ MORE ABOUT MOSFETs — THE INPUT RESISTANCE OF THE MOSFET IS THE HIGHEST OF ANY TRANSISTOR. THIS AND OTHER FACTORS GIVE MOSFETs IMPORTANT ADVANTAGES:

1. THE GATE-CHANNEL RESISTANCE IS ALMOST INFINITE (TYPICALLY 1,000,000,000,000,000-OHMS). THIS MEANS THE GATE PULLS NO CURRENT FROM EXTERNAL CIRCUITS. (WELL, IT MAY BORROW A FEW TRILLIONTHS OF AN AMPERE.)

2. MOSFETs CAN FUNCTION AS VOLTAGE-CONTROLLED VARIABLE RESISTORS. THE GATE VOLTAGE CONTROLS CHANNEL RESISTANCE.

3. NEW KINDS OF MOSFETs CAN SWITCH VERY HIGH CURRENTS IN A FEW BILLIONTHS OF A SECOND.

☐ CAUTION — BECAUSE THE GLASS-LIKE SILICON OXIDE LAYER BELOW THE GATE IS SO THIN, IT CAN BE PIERCED BY TOO MUCH VOLTAGE OR EVEN STATIC ELECTRICITY. EVEN THE STATIC CHARGE GENERATED BY CLOTHING OR A CELLOPHANE WRAPPER CAN ZAP THE GATE OF A MOSFET!

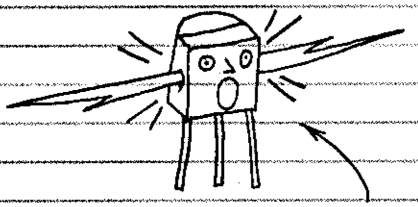
ZAPPED MOSFET

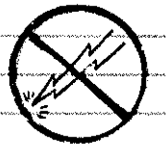
CAUTION SYMBOL

☐ KINDS OF MOSFETs — LIKE JFETs, MOSFETs INSTALLED IN SMALL METAL OR PLASTIC PACKAGES ARE USED TO GIVE AMPLIFIERS AN ULTRA-HIGH INPUT RESISTANCE. THEY ARE ALSO USED AS VOLTAGE CONTROLLED RESISTORS AND SWITCHES. THE MOST IMPORTANT CATEGORY HAS BECOME:

POWER.

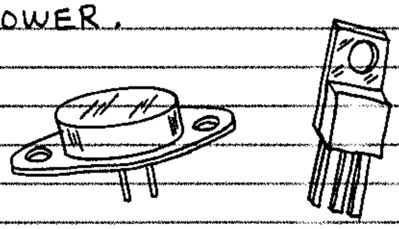

POWER MOSFETs ALLOW A FEW VOLTS TO SWITCH OR AMPLIFY MANY AMPERES AT VERY FAST SPEEDS.

☐ MOSFET SYMBOLS — THESE ARE THE MOST COMMON.

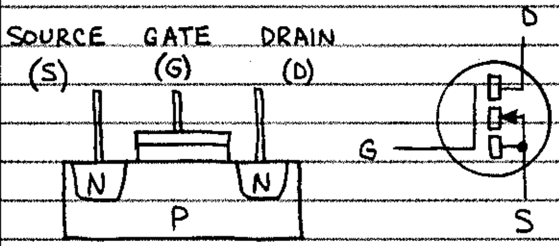

N-MOSFET

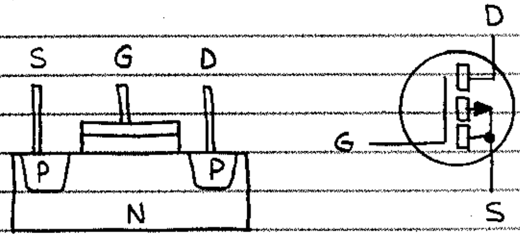

P-MOSFET

HOW FETs ARE USED

FIELD-EFFECT TRANSISTORS ARE USED AS AMPLIFIERS, SWITCHES AND VOLTAGE-CONTROLLED RESISTORS. HERE ARE SOME TYPICAL CIRCUIT ARRANGEMENTS.

☐ A JFET ELECTROMETER — THIS ULTRA-SIMPLE CIRCUIT IS THE ELECTRONIC VERSION OF THE ELECTROSCOPE. THE GATE LEAD OF AN N-CHANNEL JFET IS LEFT DISCONNECTED. NORMALLY A CURRENT FLOWS FROM SOURCE TO DRAIN. WHEN A NEGATIVELY CHARGED OBJECT (LIKE A PLASTIC COMB THAT'S BEEN STROKED THROUGH YOUR HAIR) IS PLACED NEAR THE GATE, THE CURRENT FLOW IS REDUCED OR STOPPED.

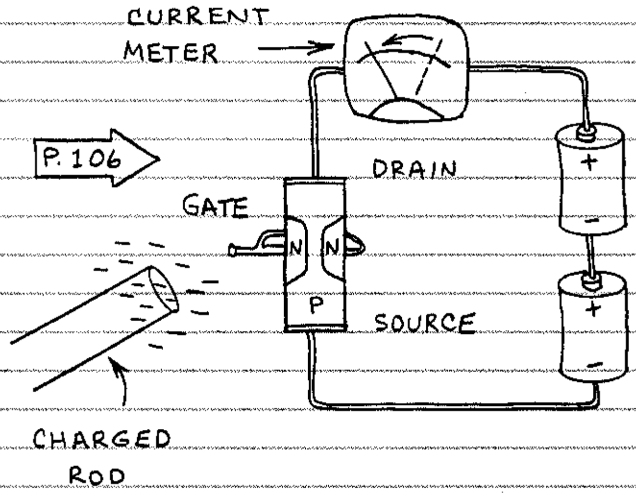
CURRENT METER
P. 106
DRAIN
GATE
N
N
P
SOURCE
+
-
+
-
CHARGED ROD

☐ A MOSFET LAMP DRIVER — THIS CIRCUIT SHOWS HOW A POWER MOSFET CAN BE USED TO SWITCH ON A LAMP OR OTHER DC POWERED DEVICE. SINCE THE POWER MOSFET HAS AN ALMOST INFINITE INPUT RESISTANCE, THE SWITCH CAN BE REPLACED BY A TINY INPUT SIGNAL.

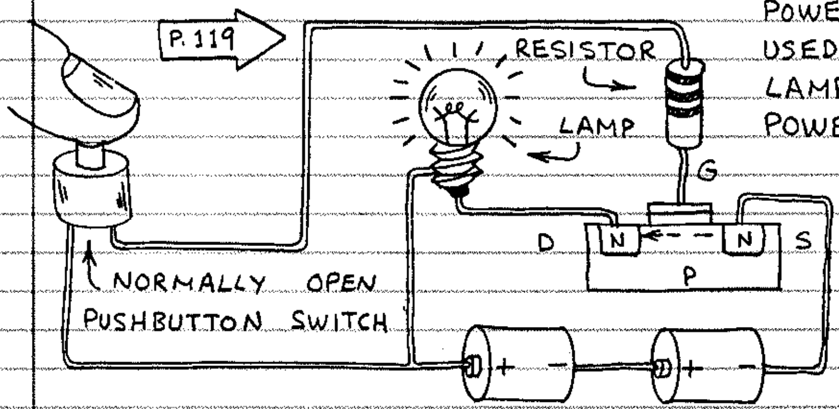
P. 119
RESISTOR
LAMP
G
D
N
N
S
P
NORMALLY OPEN PUSHBUTTON SWITCH
+
-
+
-

☐ A MOSFET LAMP DIMMER — THIS CIRCUIT USES A POWER MOSFET AS A VOLTAGE CONTROLLED RESISTOR.

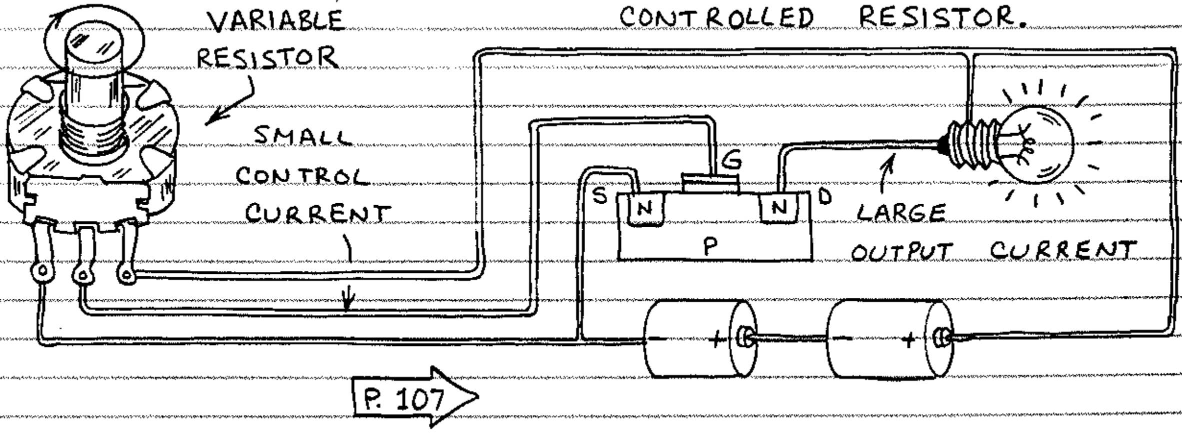
VARIABLE RESISTOR
SMALL CONTROL CURRENT
G
S
N
N
D
P
LARGE OUTPUT CURRENT
-
+
-
+
P. 107

THE UNIJUNCTION TRANSISTOR

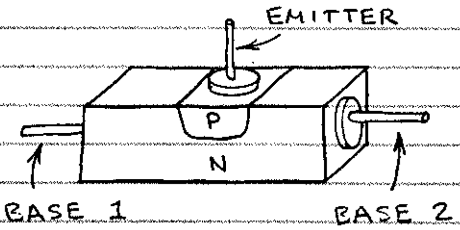
EMITTER
P
N
BASE 1
BASE 2

THE UNIJUNCTION TRANSISTOR (UJT) IS NOT A TRUE TRANSISTOR. IT'S MORE LIKE A DIODE WITH TWO CATHODE CONNECTIONS. IT WORKS LIKE A VOLTAGE-CONTROLLED SWITCH AND DOES NOT AMPLIFY.

☐ UJT OPERATION —

NORMALLY A SMALL CURRENT WILL FLOW FROM BASE 1 TO BASE 2. WHEN THE VOLTAGE APPLIED TO THE EMITTER REACHES A CERTAIN THRESHOLD (SEVERAL VOLTS), THE UJT SWITCHES ON AND A HIGH CURRENT FLOWS FROM BASE 1 TO THE EMITTER. BELOW THE THRESHOLD VOLTAGE, NO CURRENT FLOWS FROM BASE 1 TO THE EMITTER.

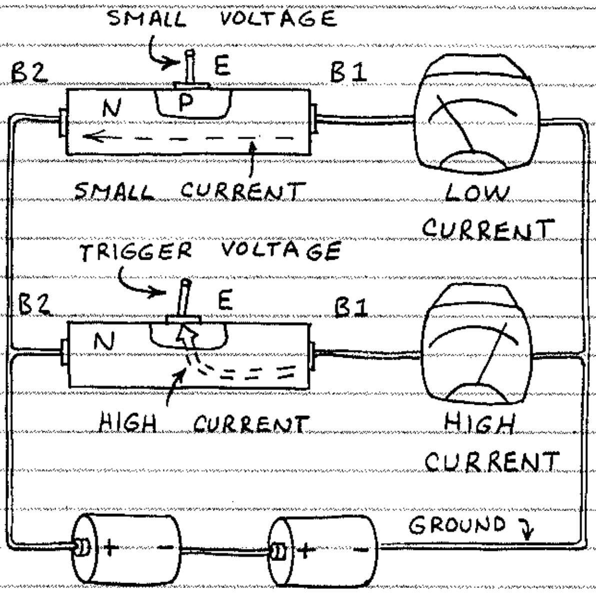
SMALL VOLTAGE
B2
E
B1
N
P
SMALL CURRENT
LOW CURRENT
TRIGGER VOLTAGE
B2
E
B1
N
HIGH CURRENT
HIGH CURRENT
GROUND
+
-
+
-

☐ UJT SYMBOL —

THE SYMBOL FOR THE UJT RESEMBLES THAT OF A JFET.

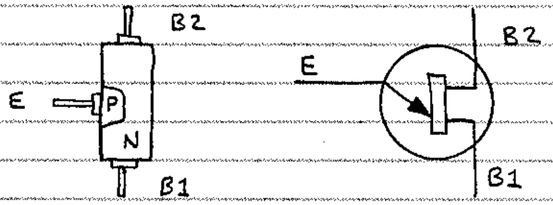
B2
E
P
N
B1
B2
E
B1

HOW UNIJUNCTION TRANSISTORS ARE USED

THIS ARRANGEMENT ALLOWS A UJT TO FLASH A LIGHT-EMITTING DIODE (LED). CURRENT FLOWS INTO THE CAPACITOR UNTIL THE UJT'S TRIGGER VOLTAGE IS REACHED. THE CURRENT IN THE CAPACITOR IS THEN "DUMPED" THROUGH THE LED. THE LED GLOWS UNTIL THE CAPACITOR IS DISCHARGED. THE CHARGE-DISCHARGE CYCLE THEN REPEATS.

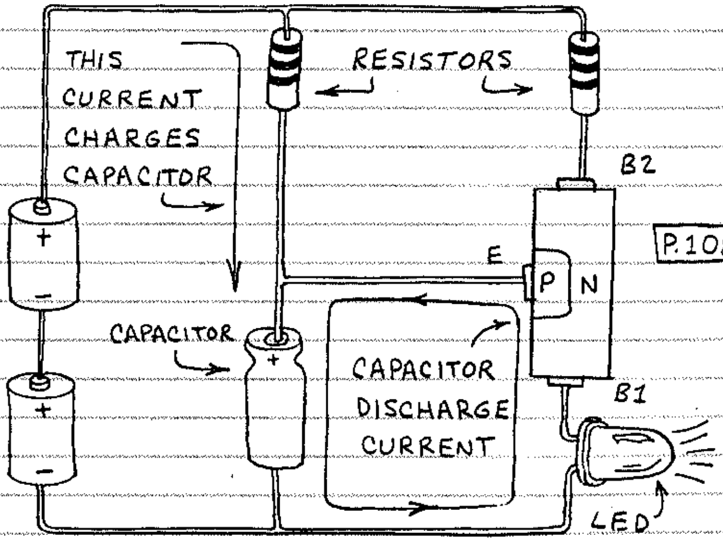
THIS CURRENT CHARGES CAPACITOR
RESISTORS
B2
E
P
N
+
-
+
-
CAPACITOR
+
CAPACITOR DISCHARGE CURRENT
B1
LED

P.108

THE THYRISTOR

THYRISTORS ARE SEMICONDUCTOR DEVICES WITH THREE LEADS. A SMALL CURRENT AT ONE LEAD WILL ALLOW A MUCH LARGER CURRENT TO FLOW THROUGH THE OTHER TWO LEADS. THE CONTROLLED CURRENT IS EITHER ON OR OFF. THEREFORE THYRISTORS DO NOT AMPLIFY FLUCTUATING SIGNALS LIKE TRANSISTORS DO. INSTEAD THEY ARE SOLID-STATE SWITCHES. THERE ARE TWO FAMILIES OF THYRISTORS, SILICON-CONTROLLED RECTIFIERS (SCRs) AND TRIACs. SCRs SWITCH DIRECT CURRENT AND TRIACs SWITCH ALTERNATING CURRENT.

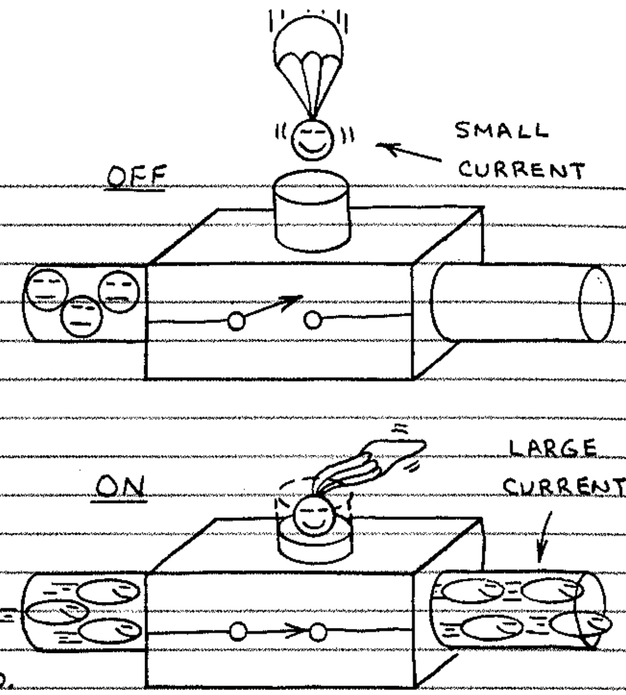
OFF
SMALL CURRENT
ON
LARGE CURRENT

SILICON-CONTROLLED RECTIFIERS (SCRs)

THE SCR IS SIMILAR TO A BIPOLAR TRANSISTOR WITH A FOURTH LAYER AND THEREFORE THREE PN JUNCTIONS. IT IS SOMETIMES CALLED A 4-LAYER PNPN DIODE SINCE IT PASSES A CURRENT IN ONLY ONE DIRECTION.

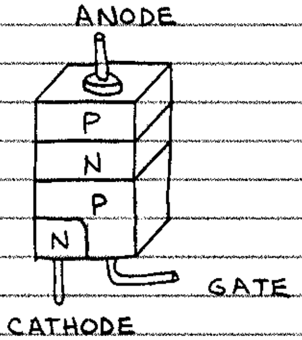
ANODE
P
N
P
N
GATE
CATHODE

☐ SCR OPERATION — IF THE ANODE OF AN SCR IS MADE MORE POSITIVE THAN THE CATHODE THE TWO OUTERMOST PN JUNCTIONS ARE FORWARD BIASED. THE MIDDLE PN JUNCTION, HOWEVER, IS REVERSE BIASED AND CURRENT CANNOT FLOW. A SMALL GATE CURRENT FORWARD BIASES THE MIDDLE PN JUNCTION AND ALLOWS A MUCH LARGER CURRENT TO FLOW THROUGH THE DEVICE. THE SCR STAYS ON EVEN IF THE GATE CURRENT IS REMOVED! (UNTIL POWER IS DISCONNECTED.)

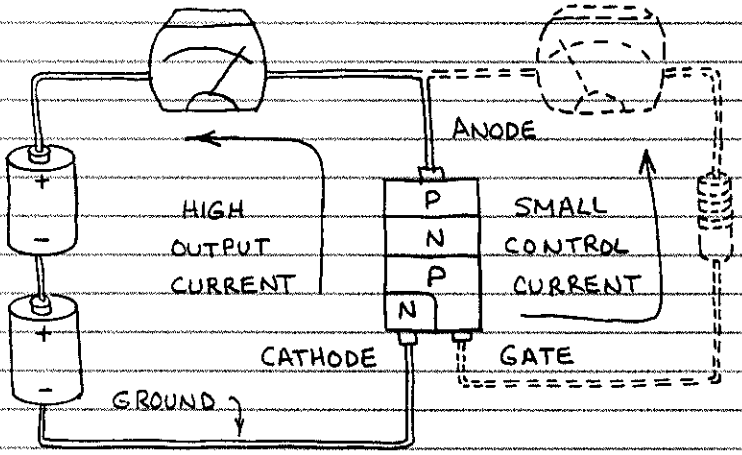
ANODE
+
-
+
-
HIGH OUTPUT CURRENT
P
N
P
N
SMALL CONTROL CURRENT
CATHODE
GATE
GROUND

☐ KINDS OF SCRs — SCRs ARE CATEGORIZED ACCORDING TO THE CURRENT THEY CAN SWITCH. HERE ARE THREE GENERAL CATEGORIES (MANY OTHER CASE STYLES ARE AVAILABLE):

LOW CURRENT.

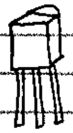

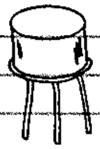

LOW CURRENT SCRs INCLUDE THOSE THAT SWITCH UP TO 1-AMPERE AT UP TO 100-VOLTS.

MEDIUM CURRENT.

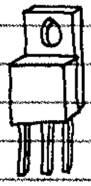

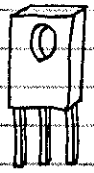

THESE SCRs SWITCH UP TO 10-AMPERES AT UP TO SEVERAL HUNDRED VOLTS. ONE COMMON USE IS SOLID-STATE SWITCHING FOR AUTO ENGINES.

HIGH CURRENT.

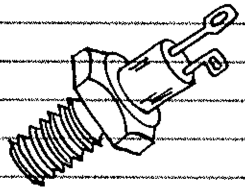

THESE SCRs CAN SWITCH UP TO 2,500-AMPERES AT UP TO SEVERAL THOUSAND VOLTS! THEY CONTROL MOTORS, LIGHTS, APPLIANCES, ETC.

☐ SCR SYMBOL.

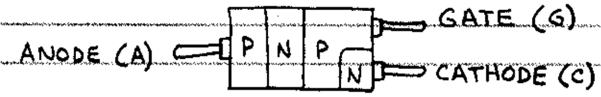
GATE (G)
ANODE (A)
P
N
P
N
CATHODE (C)

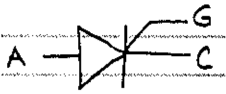

HOW SCRs ARE USED

THIS ARRANGEMENT SHOWS HOW AN SCR IS USED TO SWITCH ON AN INCANDESCENT LAMP. OTHER DEVICES CAN ALSO BE CONTROLLED.

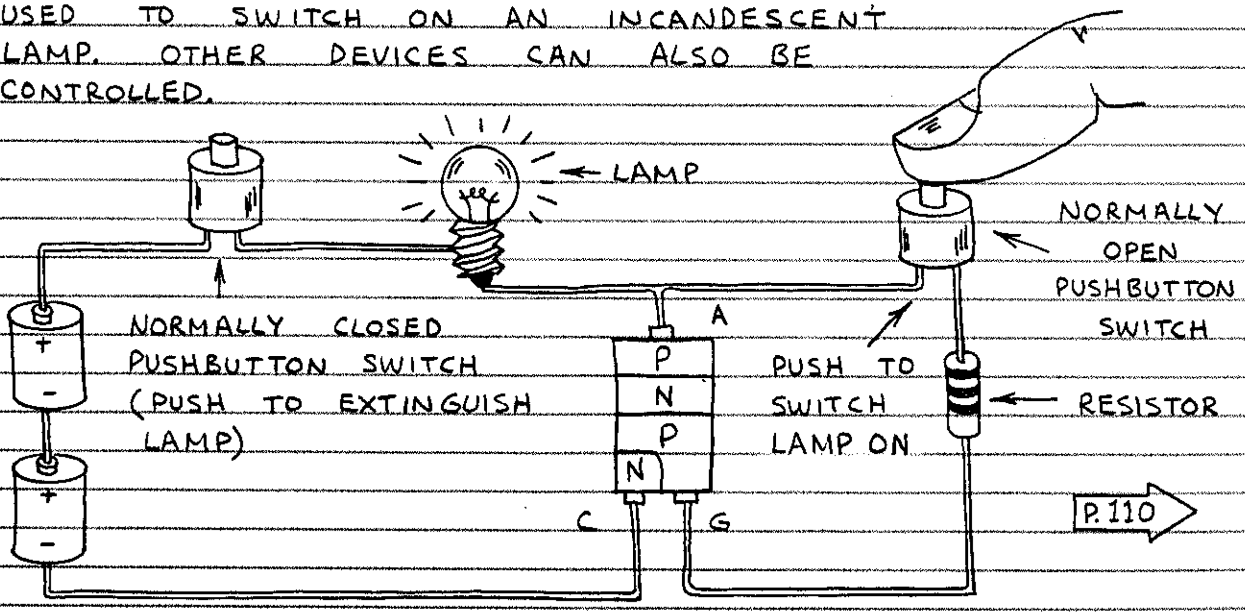
LAMP
NORMALLY OPEN PUSHBUTTON SWITCH
A
+
-
+
-
NORMALLY CLOSED PUSHBUTTON SWITCH (PUSH TO EXTINGUISH LAMP)
P
N
P
N
PUSH TO SWITCH LAMP ON
RESISTOR
C
G

P. 110

TRIACS

THE TRIAC IS EQUIVALENT TO TWO SCRs CONNECTED IN PARALLEL. THIS MEANS TRIACS CAN SWITCH BOTH DIRECT AND ALTERNATING CURRENT. NOTICE THAT THE TRIAC HAS FIVE LAYERS PLUS AN EXTRA N-TYPE REGION. ALSO NOTE HOW ALL THREE LEADS MAKE CONTACT WITH TWO LAYERS.

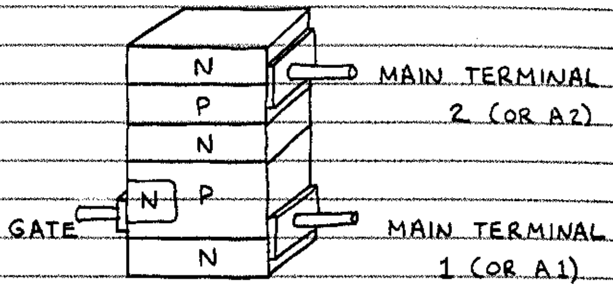
N
P
N
N
P
N
MAIN TERMINAL 2 (OR A2)
GATE
MAIN TERMINAL 1 (OR A1)

☐ TRIAC OPERATION — THE TWO PARALLEL SCRs THAT FORM A TRIAC FACE IN OPPOSITE DIRECTIONS (REVERSE-PARALLEL). WHEN USED TO SWITCH ALTERNATING CURRENT, THE TRIAC STAYS ON ONLY WHEN THE GATE RECEIVES CURRENT. REMOVE THE GATE CURRENT AND IT SWITCHES OFF WHEN THE AC PASSES THROUGH 0 VOLTS.

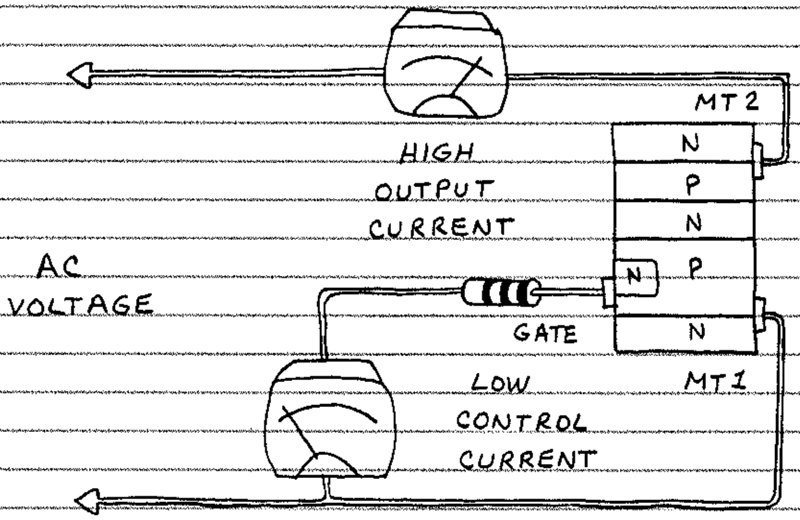
MT2
HIGH
OUTPUT
CURRENT
N
P
N
N
P
N
AC
VOLTAGE
GATE
LOW
CONTROL
CURRENT
MT1

☐ KINDS OF TRIACS — TRIACS, LIKE SCRs, ARE CATEGORIZED ACCORDING TO THE CURRENT THEY CAN SWITCH. TRIACS DON'T HAVE THE VERY HIGH POWER CAPABILITY OF HIGH CURRENT SCRs. HERE ARE TWO CATEGORIES:

LOW CURRENT.

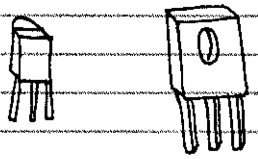

LOW CURRENT TRIACS SWITCH UP TO 1-AMPERE AT UP TO SEVERAL HUNDRED VOLTS. OTHER CASE STYLES ALSO USED.

MEDIUM CURRENT.

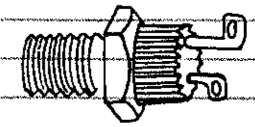

THESE TRIACS SWITCH UP TO 40-AMPERES AT UP TO 1,000-VOLTS. MANY CASE STYLES ARE AVAILABLE.

☐ TRIAC SYMBOL — REMEMBER, THE TRIAC IS THE SAME AS TWO REVERSE-PARALLEL SCRs:

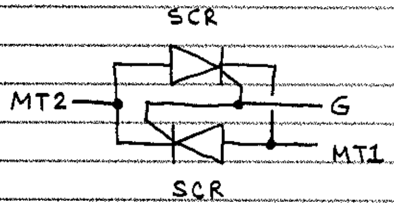
SCR
MT2
G
MT1
SCR

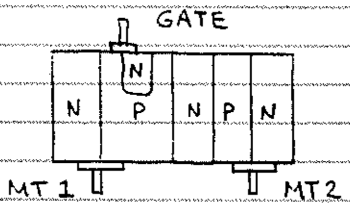
GATE
N
N
P
N
P
N
MT1
MT2

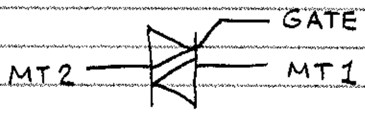
GATE
MT2
MT1

HOW TRIACS ARE USED

THIS ARRANGEMENT SHOWS HOW A TRIAC CAN SWITCH ON A LAMP POWERED BY HOUSEHOLD LINE CURRENT. MOTORS AND OTHER DEVICES CAN ALSO BE CONTROLLED.

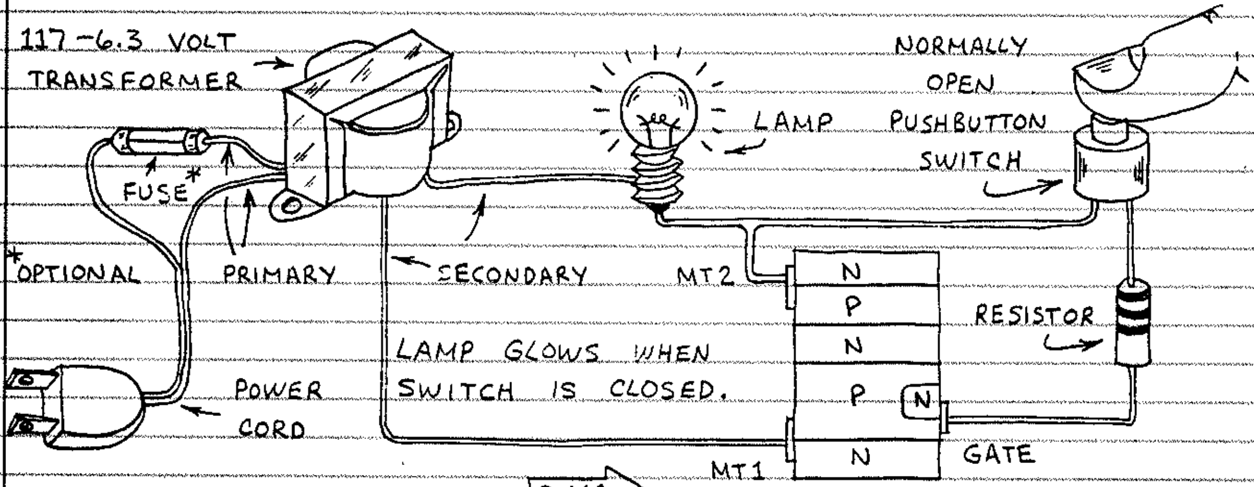
117-6.3 VOLT TRANSFORMER
FUSE*
*OPTIONAL
PRIMARY
SECONDARY
POWER CORD
LAMP
NORMALLY OPEN PUSHBUTTON SWITCH
MT2
N
P
N
P
N
N
RESISTOR
GATE
MT1

LAMP GLOWS WHEN SWITCH IS CLOSED.

CAUTION: DO NOT ASSEMBLE.

P. 111

TWO-LEAD THYRISTORS

AN SCR OR TRIAC WILL SWITCH ON WITHOUT A GATE SIGNAL IF THE VOLTAGE ACROSS ITS OTHER TWO LEADS REACHES A CERTAIN LEVEL (THE BREAKDOWN VOLTAGE). THIS SELF-SWITCHING ABILITY MAKES POSSIBLE TWO-LEAD THYRISTORS.

FOUR-LAYER DIODE.

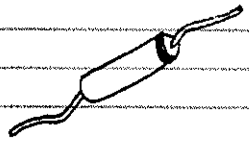

A FOUR-LAYER DIODE IS AN SCR WITHOUT A GATE. IT SWITCHES DC VOLTAGE.

DIAC.

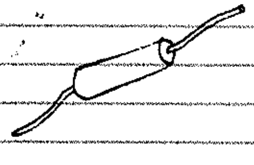

A DIAC IS A THREE-LAYER DEVICE SIMILAR TO A PNP JUNCTION TRANSISTOR WITHOUT A BASE LEAD. IT SWITCHES AC VOLTAGE.

4. PHOTONIC SEMICONDUCTORS

PHOTONICS IS THE FAST GROWING FIELD OF ELECTRONICS INVOLVING SEMICONDUCTOR DEVICES THAT EMIT AND DETECT LIGHT. BEFORE LOOKING AT SOME PHOTONIC COMPONENTS, LET'S TAKE A QUICK LOOK AT SOME FACTS ABOUT LIGHT.

LIGHT

"LET THERE BE LIGHT..."

LIGHT IS COMPOSED OF PARTICLES CALLED PHOTONS THAT BEHAVE LIKE WAVES OF ENERGY. PHOTONS ARE NOT NECESSARILY VISIBLE AND ONLY THOSE YOU CAN SEE ARE COLLECTIVELY CALLED LIGHT. PHOTONS ARE PRODUCED WHEN AN ELECTRON THAT'S BEEN EXCITED TO A HIGHER THAN NORMAL ENERGY LEVEL FALLS BACK TO ITS NORMAL LEVEL.

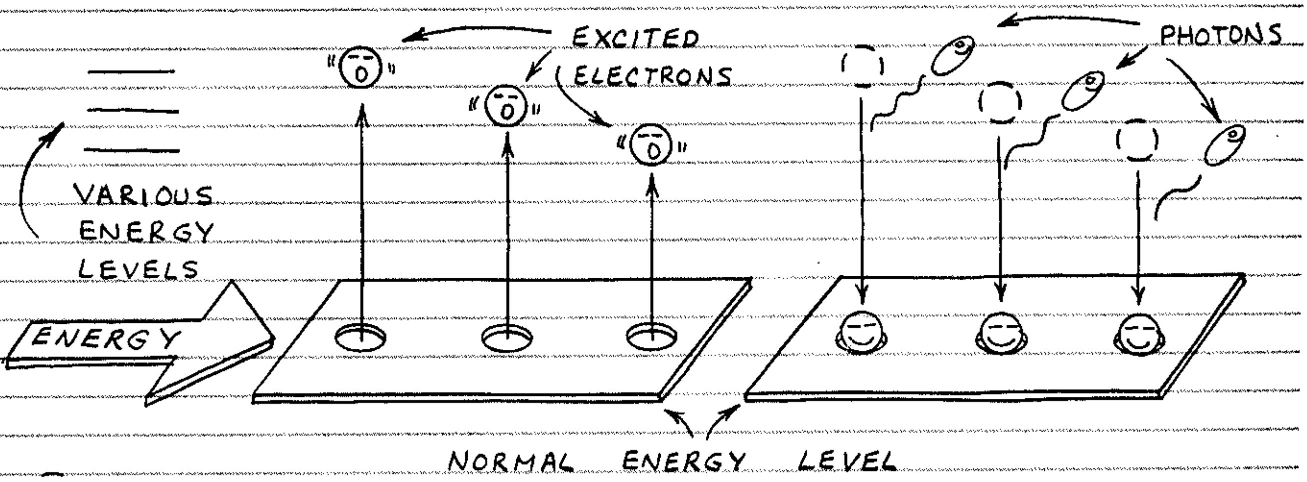
EXCITED ELECTRONS
PHOTONS
VARIOUS ENERGY LEVELS
ENERGY
NORMAL ENERGY LEVEL

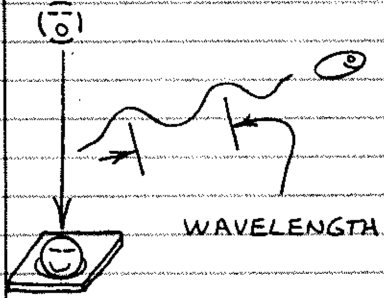
WAVELENGTH

REMEMBER, PHOTONS ACT LIKE WAVES. THE DISTANCE BETWEEN CRESTS IS THE WAVELENGTH. ELECTRONS EXCITED TO HIGHER ENERGY LEVELS EMIT PHOTONS WITH SHORTER WAVELENGTHS THAN ELECTRONS EXCITED TO LOWER LEVELS.

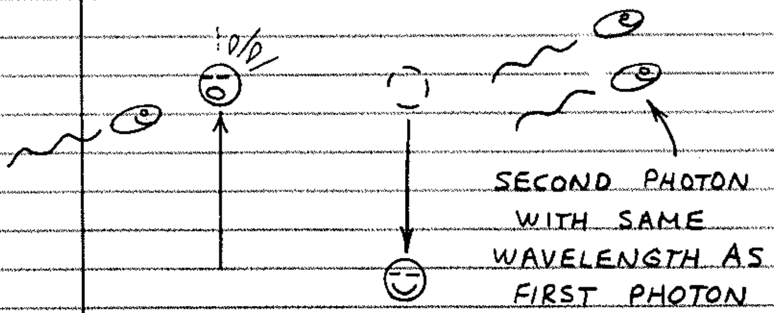
SECOND PHOTON WITH SAME WAVELENGTH AS FIRST PHOTON

EXCITED ELECTRONS CAN RESUME THEIR NORMAL LEVEL SPONTANEOUSLY. OR A PHOTON WITH THE PROPER WAVELENGTH CAN STIMULATE AN EXCITED ELECTRON TO RETURN TO ITS NORMAL LEVEL.

☐ THE ELECTROMAGNETIC SPECTRUM — VISIBLE LIGHT IS A FORM OF ELECTROMAGNETIC RADIATION. THE WAVELENGTH OF LIGHT IS SPECIFIED IN NANOMETERS (1-NANOMETER IS A BILLIONTH OF A METER). THE DIAGRAM BELOW SHOWS THE RELATIONSHIP OF LIGHT TO OTHER FORMS OF ELECTROMAGNETIC RADIATION.

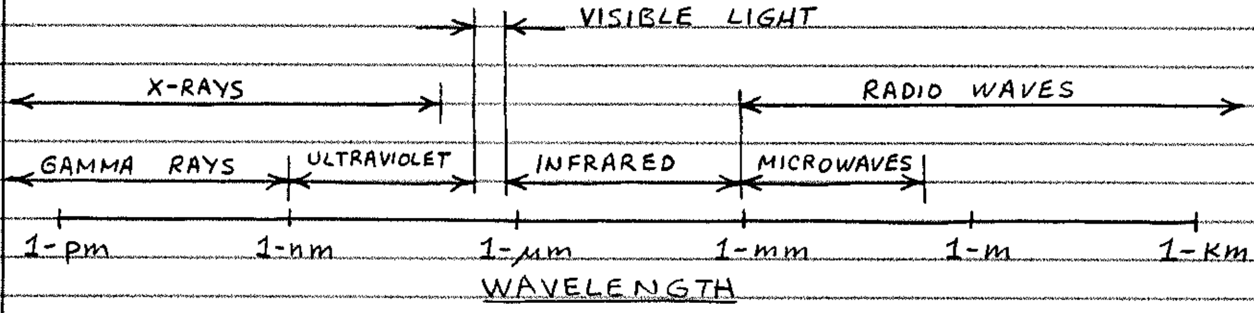
VISIBLE LIGHT
X-RAYS
RADIO WAVES
GAMMA RAYS
ULTRAVIOLET
INFRARED
MICROWAVES
1-pm
1-nm
1-μm
1-mm
1-m
1-Km
WAVELENGTH

1-pm = 1 PICOMETER (0.000 000 000 001 METER)
1-nm = 1 NANOMETER (0.000 000 001 METER)
1-μm = 1 MICROMETER (0.000 001 METER)
1-mm = 1 MILLIMETER (0.001 METER)
1-m = 1 METER (39.37 INCHES)
1-Km = 1 KILOMETER (1000 METERS)

→ ||
THESE LINES ARE ONE MILLIMETER (1-mm) APART.

☐ THE OPTICAL SPECTRUM — ULTRAVIOLET, VISIBLE AND INFRARED RADIATION ARE TOGETHER CALLED THE *OPTICAL SPECTRUM*. HERE'S AN EXPANDED DIAGRAM OF THE OPTICAL SPECTRUM:

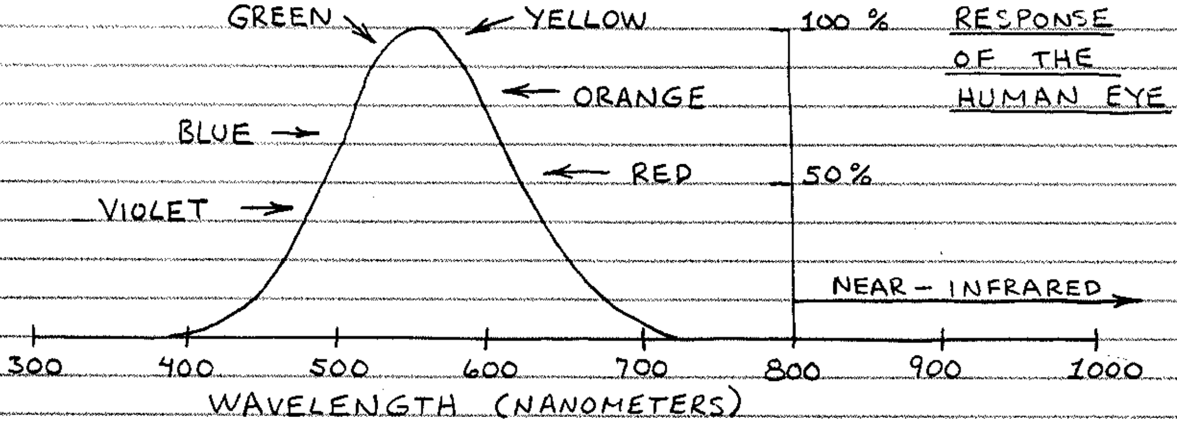

MANY PHOTONIC SEMICONDUCTORS EMIT OR DETECT *NEAR-INFRARED* RADIATION. SILICON, FOR EXAMPLE, CAN DETECT VISIBLE LIGHT. BUT IT IS MOST SENSITIVE TO NEAR-INFRARED AT ABOUT 880 nm. BECAUSE SO MANY PHOTONIC COMPONENTS CAN OPERATE IN BOTH THE VISIBLE AND NEAR-INFRARED, IT'S COMMON TO REFER TO NEAR-INFRARED AS LIGHT.

OPTICAL COMPONENTS

OPTICAL COMPONENTS CONDUCT, BEND OR CHANGE THE CHARACTERISTICS OF LIGHT. SEVERAL ARE VERY IMPORTANT IN MANY APPLICATIONS OF PHOTONIC SEMICONDUCTORS:

1. FILTERS TRANSMIT ONLY A NARROW BAND OF OPTICAL WAVELENGTHS.

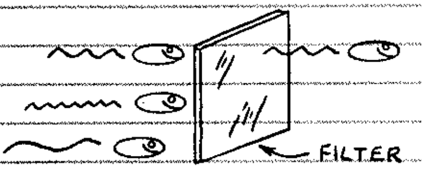

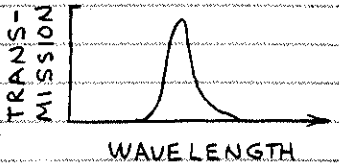

2. REFLECTORS REFLECT SOME OR MOST OF AN ONCOMING LIGHT BEAM. SOME LIGHT MAY OR MAY NOT BE TRANSMITTED.

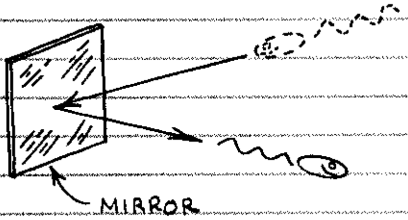

THOSE WITH A VERY SMOOTH SURFACE (LIKE MIRRORS) ARE CALLED SPECULAR REFLECTORS.

3. BEAMSPLITTERS REFLECT PART OF AN ONCOMING LIGHT BEAM AND TRANSMIT THE REMAINDER.

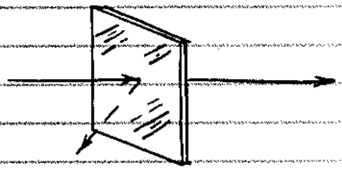

A GLASS MICROSCOPE SLIDE MAKES A GOOD BEAM SPLITTER. (EACH SURFACE REFLECTS 4%.)

4. LENSES BEND LIGHT. THE MOST IMPORTANT ARE:

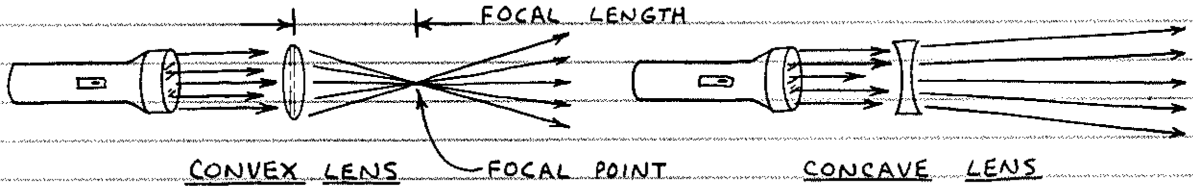

CONVEX LENS

CONCAVE LENS

CONVEX LENSES ARE OFTEN USED IN CONJUNCTION WITH SEMICONDUCTOR LIGHT SOURCES AND DETECTORS. FOR EXAMPLE, THEY CAN COLLECT AND FOCUS LIGHT ONTO A MINIATURE DETECTOR.

5. OPTICAL FIBERS ARE THIN, FLEXIBLE STRANDS OF HIGHLY TRANSPARENT GLASS OR PLASTIC THAT CONDUCT LIGHT. THE LIGHT TRAVELS THROUGH A CORE SURROUNDED BY A THIN CLADDING. PLASTIC FIBERS ARE INEXPENSIVE. GLASS FIBERS ARE MUCH MORE TRANSPARENT. BOTH KINDS TRANSMIT SOME WAVELENGTHS MUCH BETTER THAN OTHERS. HIGH QUALITY FIBERS ARE USED TO SEND TELEPHONE CALLS AND COMPUTER DATA VIA PULSES OF LIGHT.

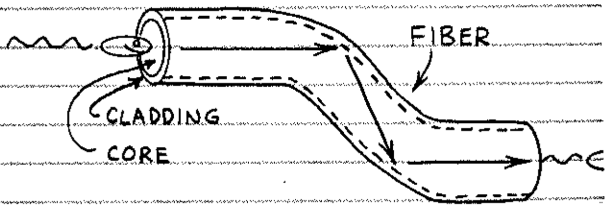

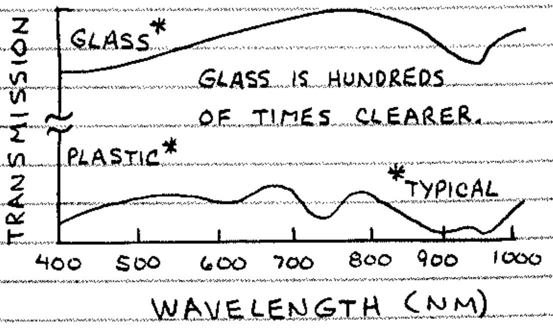

HOW CONVEX LENSES ARE USED

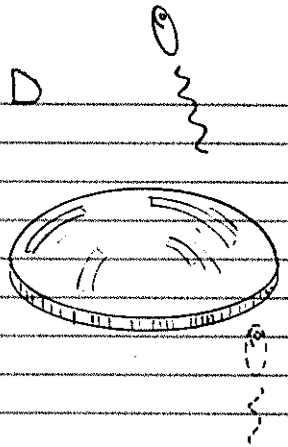

MANY SEMICONDUCTOR LIGHT SOURCES AND DETECTORS ARE EQUIPPED WITH A BUILT-IN CONVEX LENS. THIS PAGE EXPLAINS WHY AND SHOWS HOW EXTERNAL LENS ARE USED WITH SOURCES AND DETECTORS.

☐ THE INVERSE SQUARE LAW.

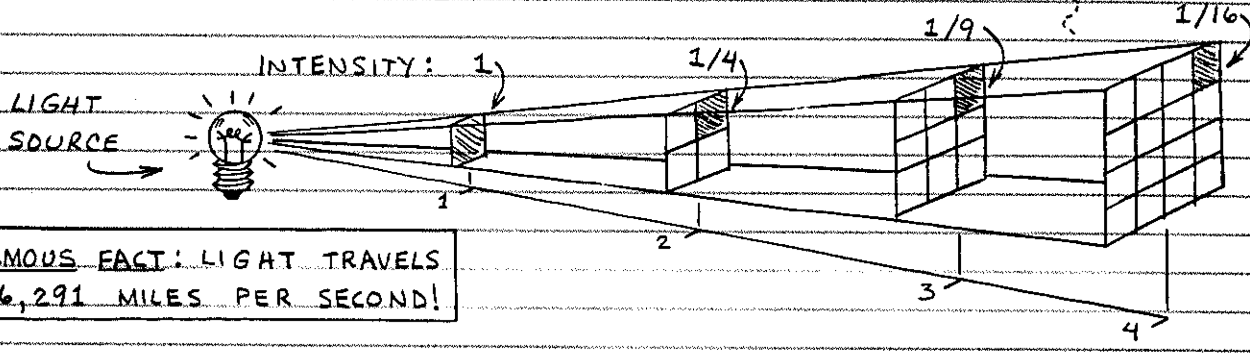
INTENSITY: 1
1/4
1/9
1/16
LIGHT SOURCE
1
2
3
4

FAMOUS FACT: LIGHT TRAVELS 186,291 MILES PER SECOND!

AS LIGHT FROM A VERY SMALL SOURCE SPREADS OUTWARD, ITS INTENSITY IS INVERSELY PROPORTIONAL TO THE SQUARE OF THE DISTANCE. IN OTHER WORDS, IF THE DISTANCE IS 3, THEN THE INTENSITY IS 1/9 THE INTENSITY WHEN THE DISTANCE IS 1. A CONVEX LENS CAN CANCEL THIS INTENSITY REDUCTION.

☐ THE CONVEX LENS.

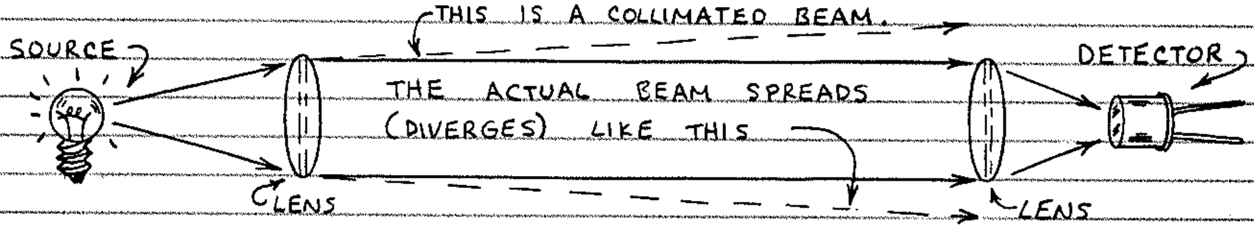
THIS IS A COLLIMATED BEAM.
SOURCE
DETECTOR
THE ACTUAL BEAM SPREADS (DIVERGES) LIKE THIS
LENS
LENS

THE BEAM SPREAD ANGLE (THE DIVERGENCE) IN RADIANS* IS THE DIAMETER OF THE SOURCE DIVIDED BY THE FOCAL LENGTH OF THE LENS. THIS MEANS LENSES WITH LONGER FOCAL LENGTHS GIVE NARROWER BEAMS. (BUT LENSES WITH LONG FOCAL LENGTHS COLLECT LESS LIGHT THAN THOSE WITH SHORT FOCAL LENGTHS...)

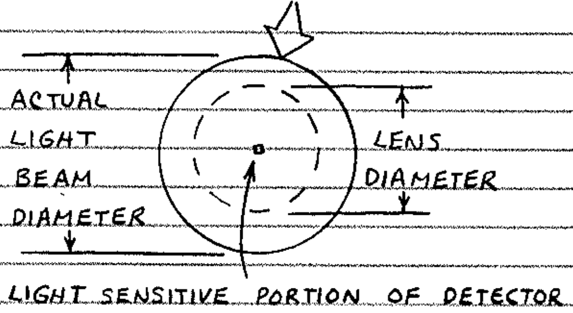
ACTUAL LIGHT BEAM DIAMETER
LENS DIAMETER
LIGHT SENSITIVE PORTION OF DETECTOR

WITH CAREFUL LENS PLACEMENT ALL THE LIGHT WITHIN THE DASHED CIRCLE CAN BE FOCUSED ONTO THE LIGHT SENSITIVE PORTION OF THE DETECTOR. (SURE BEATS THE INVERSE SQUARE LAW!)

*ONE RADIAN IS 57.3 DEGREES. (THERE ARE 360° IN A CIRCLE.)

SEMICONDUCTOR LIGHT SOURCES

WHEN BOMBARDED BY LIGHT, HEAT, ELECTRONS AND OTHER FORMS OF ENERGY, MOST SEMICONDUCTOR CRYSTALS WILL EMIT VISIBLE OR INFRARED LIGHT. THE BEST SEMICONDUCTOR LIGHT SOURCES, HOWEVER, ARE PN JUNCTION DIODES.

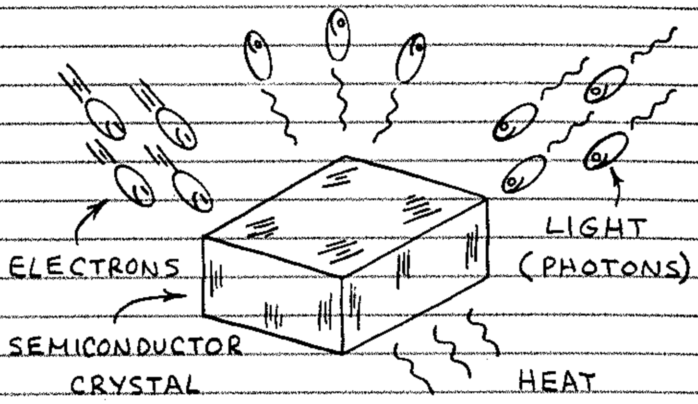
ELECTRONS
LIGHT (PHOTONS)
SEMICONDUCTOR CRYSTAL
HEAT

LIGHT EMITTING DIODES

THE LIGHT EMITTING DIODE CONVERTS AN ELECTRICAL CURRENT DIRECTLY INTO LIGHT. THEREFORE THE LIGHT EMITTING DIODE (LED) IS MORE EFFICIENT THAN MANY OTHER LIGHT SOURCES.

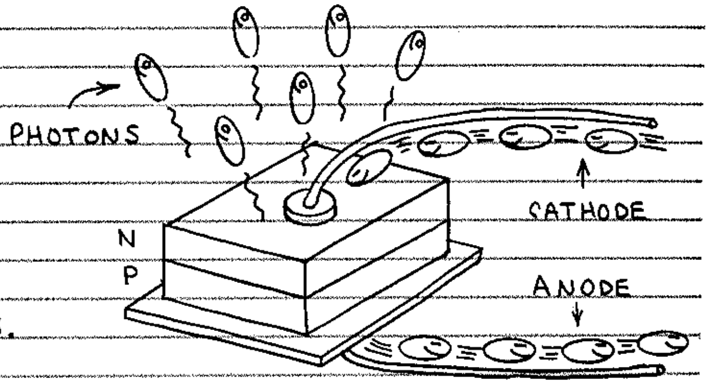
PHOTONS
CATHODE
N
P
ANODE

☐ LED OPERATION— THE FORWARD VOLTAGE ACROSS A DIODE MUST EXCEED A THRESHOLD LEVEL BEFORE A CURRENT CAN CROSS THE JUNCTION. FOR SILICON, WHICH EMITS A TINY AMOUNT OF NEAR-INFRARED, THE THRESHOLD IS 0.6-VOLT. FOR GALLIUM ARSENIDE, WHICH EMITS CONSIDERABLE NEAR-INFRARED, THE THRESHOLD IS 1.3-VOLTS. THIS VOLTAGE EXCITES THE ELECTRONS. WHEN THE ELECTRONS CROSS THE JUNCTION AND COMBINE WITH HOLES, THEY EMIT PHOTONS.

FORWARD BIAS

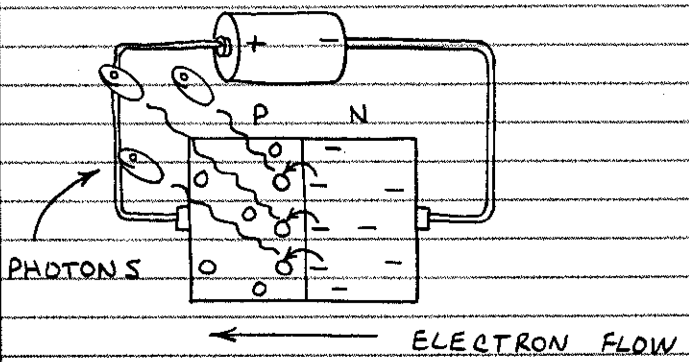
+
−
P
N
PHOTONS
ELECTRON FLOW

REVERSE BIAS

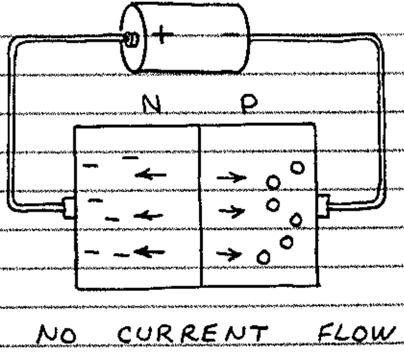
+
−
N
P
NO CURRENT FLOW

☐ MORE ABOUT LED OPERATION — HERE ARE SOME KEY ASPECTS OF LED OPERATION YOU SHOULD KNOW ABOUT:

1. THE LIGHT EMITTED BY AN INCANDESCENT LAMP CONTAINS MANY WAVELENGTHS. THE LIGHT EMITTED BY AN LED HAS A NARROW WAVELENGTH RANGE. (THIS IS BECAUSE THE ELECTRONS IN THE LED ARE ALL EXCITED TO THE SAME LEVEL.)

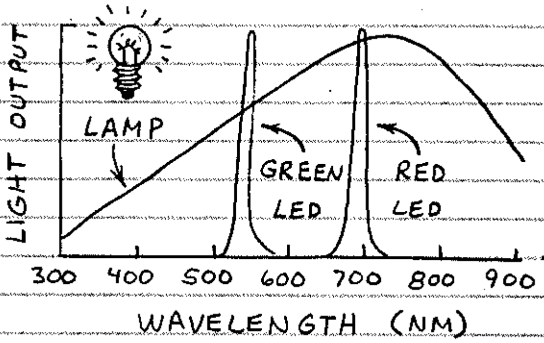

2. WHEN AN LED BEGINS TO CONDUCT, THE VOLTAGE INCREASES GRADUALLY WHILE THE CURRENT INCREASES RAPIDLY. TOO MUCH CURRENT WILL OVERHEAT THE LED AND POSSIBLY SEPARATE THE LEADS OR MELT THE SEMICONDUCTOR CHIP.

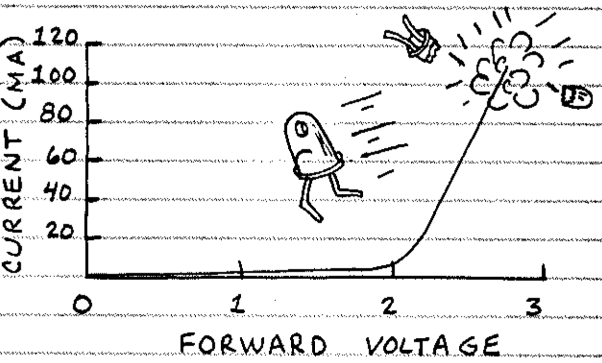

3. THE LIGHT EMITTED BY AN LED IS DIRECTLY PROPORTIONAL TO CURRENT THROUGH THE LED. THIS MEANS LEDS ARE IDEAL FOR TRANSMITTING INFORMATION. THE LIGHT OUTPUT FROM AN OVERHEATED LED WILL SOON DECREASE. THE LED MAY EVEN BE DAMAGED.

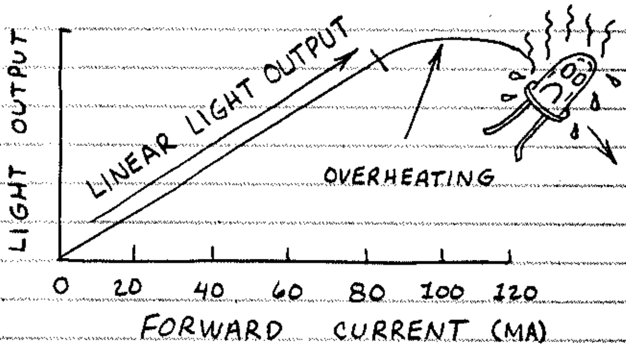

4. THE FORWARD VOLTAGE AND WAVELENGTH OF AN LED ARE DIRECTLY RELATED. THEREFORE IT'S NOT ALWAYS POSSIBLE TO SUBSTITUTE DIFFERENT LEDS WITHOUT CHANGING THE VOLTAGE AND CURRENT. MANY DIFFERENT SEMICONDUCTORS ARE USED TO MAKE VARIOUS LEDS. VISIBLE LIGHT EMITTING LEDs EMIT UP TO A MILLIWATT OR SO OF POWER. SOME INFRARED LEDs (LIKE 880 NM UNITS) EMIT 15 OR MORE MILLIWATTS! (A FLASHLIGHT EMITS 10 OR MORE MILLIWATTS.)

WAVELENGTH (NM)	VOLTAGE
565 (GREEN)	2.2 — 3.0
590 (YELLOW)	2.2 - 3.0
615 (ORANGE)	1.8 - 2.7
640 (RED)	1.6 - 2.0
690 (RED)	2.2 - 3.0
880 (INFRARED)	2.0 - 2.5
900 (INFRARED)	1.2 - 1.6
940 (INFRARED)	1.3 - 1.7

☐ KINDS OF LEDs — SINCE THE LED IS A LIGHT SOURCE, IT'S HELPFUL TO KNOW WHAT'S INSIDE THE PLASTIC OR METAL LED CASE. SHOWN HERE IS A TYPICAL LED. THE HEAVY LEADS HELP CONDUCT HEAT AWAY FROM THE CHIP. THE REFLECTOR COLLECTS LIGHT EMITTED FROM THE EDGES OF THE CHIP. THE EPOXY IS USUALLY COLORED WHEN THE LED IS A VISIBLE LIGHT EMITTER. LIGHT SCATTERING PARTICLES ARE OFTEN ADDED TO THE EPOXY. THIS DIFFUSES THE LIGHT AND CAUSES THE END OF THE LED TO APPEAR BRIGHTER.

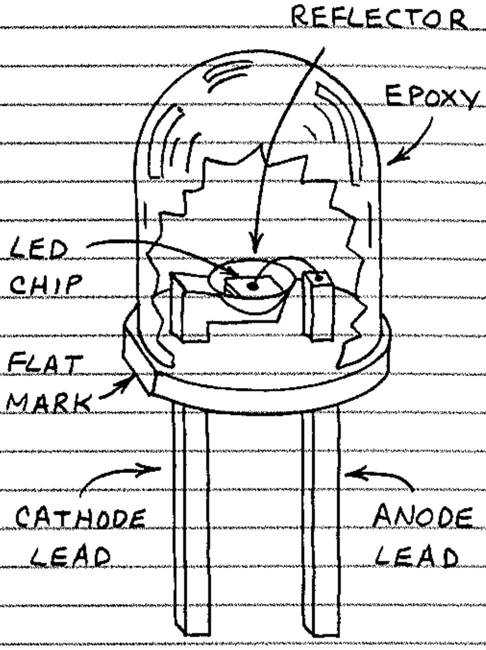

VISIBLE LIGHT LEDs.

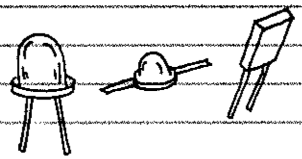

THESE INEXPENSIVE LEDs ARE USED AS INDICATOR LIGHTS. CERTAIN RED LEDs ARE USED TO TRANSMIT INFORMATION. MOST ARE ENCAPSULATED IN EPOXY.

LED DISPLAYS.

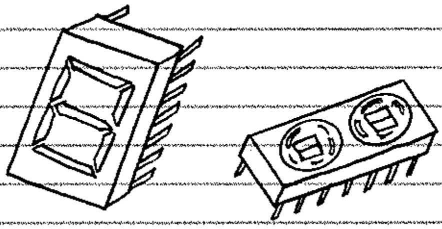

MANY KINDS OF LED READOUTS CAPABLE OF DISPLAYING DIGITS AND CHARACTERS ARE AVAILABLE. THEY ARE MORE RUGGED THAN LIQUID CRYSTAL DISPLAYS, BUT THEY USE MUCH MORE CURRENT.

INFRARED LEDs.

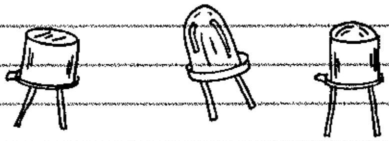

INFRARED LEDs SHOULD BE CALLED INFRARED EMITTING DIODES. THEY ARE USED TO TRANSMIT INFORMATION. THEY ARE ALSO USED IN INTRUSION ALARMS, REMOTE CONTROL DEVICES, ETC. A SPECIAL KIND OF INFRARED LED IS THE DIODE LASER. SOME EMIT SEVERAL WATTS!

☐ LED SYMBOL.

BOTH SYMBOLS SHOWN HERE ARE USED.

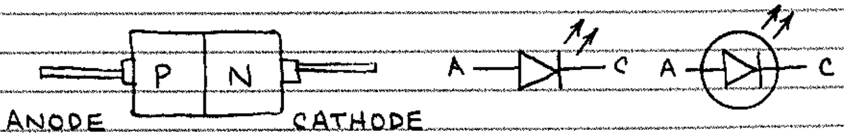

HOW LEDs ARE USED

LEDs CAN BE POWERED BY CONTINUOUS CURRENT OR BY BRIEF PULSES OF CURRENT. WHEN OPERATED CONTINUOUSLY, THE CURRENT CAN BE VARIED TO CHANGE THE LIGHT OUTPUT.

☐ LED DRIVE CIRCUIT — BECAUSE LEDs ARE CURRENT DEPENDENT, IT'S USUALLY NECESSARY TO PROTECT THEM FROM EXCESSIVE CURRENT WITH A SERIES RESISTOR. SOME LEDs INCLUDE A BUILT-IN SERIES RESISTOR. MOST DO NOT. IT'S IMPORTANT TO KNOW HOW TO DETERMINE THE REQUIRED SERIES RESISTANCE (R_S). THE FORMULA IS:

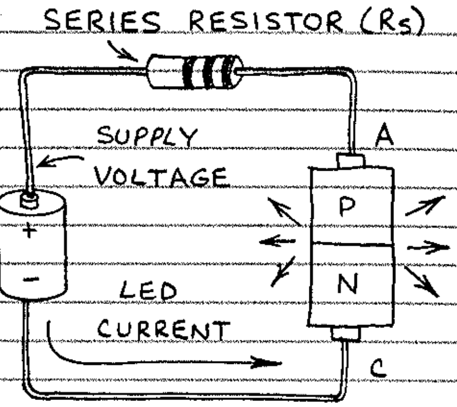

$$R_S = \frac{\text{SUPPLY VOLTAGE} - \text{LED VOLTAGE}}{\text{LED CURRENT}} \quad \text{OR} \quad R_S = \frac{V - V_{LED}}{I_{LED}}$$

P.112

EXAMPLE: SUPPOSE YOU WANT TO OPERATE A RED LED AT A FORWARD CURRENT (I_{LED}) OF 10-MILLIAMPERES FROM A 5-VOLT SUPPLY (V). V_{LED} IS 1.7-VOLTS (FROM DATA SHEET). THEREFORE R_S IS $(5-1.7)/0.01$ OR 330-OHMS.

☐ LED POLARITY INDICATOR — TWO REVERSE-PARALLEL LEDs FORM A POLARITY INDICATOR. BOTH LEDs GLOW IF THE TESTED VOLTAGE IS AC. THE SERIES RESISTOR MUST BE USED!

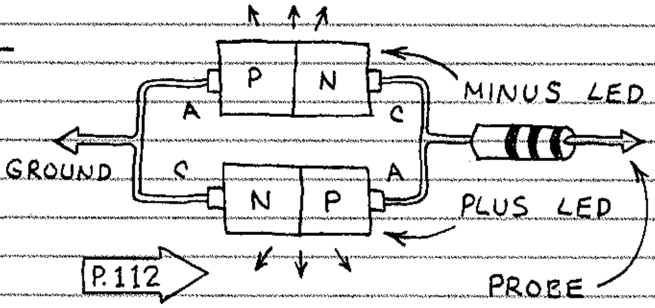

P.112

☐ PULSED LED — WHEN OPERATED CONTINUOUSLY, AN INFRARED LED MIGHT HAVE A MAXIMUM CURRENT OF 100-MILLIAMPERES. WHEN DRIVEN BY BRIEF PULSES OF CURRENT, THE SAME LED MAY SAFELY ACCEPT HUGE 10-AMPERE PULSES!

NOTE:
A SERIES RESISTOR MAY NOT BE REQUIRED IF THE PULSES DO NOT EXCEED THE MAXIMUM LEVELS SPECIFIED FOR THE LED.

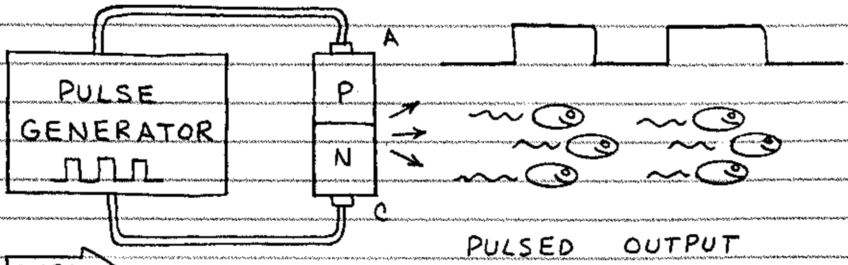

P.126

SEMICONDUCTOR LIGHT DETECTORS

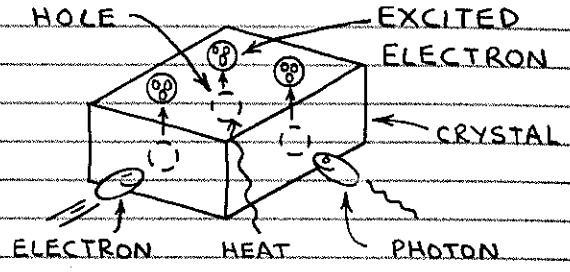
HOLE
EXCITED ELECTRON
CRYSTAL
ELECTRON
HEAT
PHOTON

ENERGY ENTERING A SEMICONDUCTOR CRYSTAL EXCITES ELECTRONS TO HIGHER LEVELS, LEAVING BEHIND HOLES. THESE ELECTRONS AND HOLES CAN RECOMBINE AND EMIT PHOTONS, OR THEY CAN MOVE AWAY FROM ONE ANOTHER AND FORM A CURRENT. THIS IS THE BASIS OF SEMICONDUCTOR LIGHT DETECTORS. THERE ARE TWO MAJOR CLASSES OF SEMICONDUCTOR LIGHT DETECTORS, THOSE WITH AND THOSE WITHOUT PN JUNCTIONS.

PHOTORESISTIVE LIGHT DETECTORS

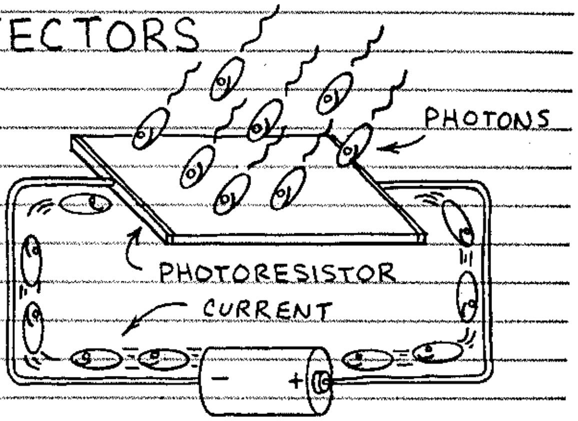
PHOTONS
PHOTORESISTOR
CURRENT
−
+

PHOTORESISTORS ARE SEMICONDUCTOR LIGHT DETECTORS WITHOUT A PN JUNCTION. THEIR RESISTANCE IS VERY HIGH (UP TO MILLIONS OF OHMS) WHEN NO LIGHT IS PRESENT. WHEN ILLUMINATED, THEIR RESISTANCE IS VERY LOW (HUNDREDS OF OHMS).

☐ PHOTORESISTOR OPERATION. THIS PANEL SHOWS HOW A PHOTON CREATES A HOLE-ELECTRON PAIR. AN EXTERNAL VOLTAGE WILL FORCE THE HOLE AND ELECTRON TO MOVE.

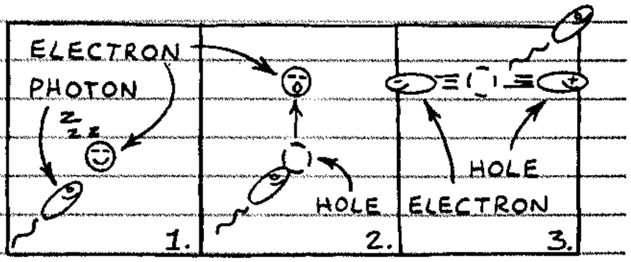
ELECTRON
PHOTON
1.
HOLE
2.
HOLE
ELECTRON
3.

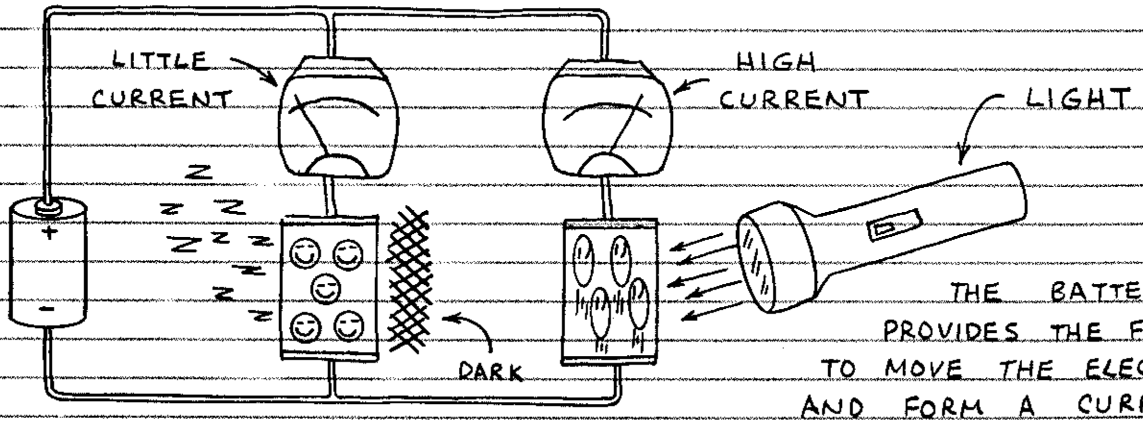
LITTLE CURRENT
HIGH CURRENT
LIGHT
+
−
DARK

THE BATTERY PROVIDES THE FORCE TO MOVE THE ELECTRONS AND FORM A CURRENT.

☐ MORE ABOUT PHOTORESISTOR OPERATION — HERE ARE SOME IMPORTANT ASPECTS OF PHOTORESISTOR OPERATION:

1. PHOTORESISTORS MAY REQUIRE A FEW MILLISECONDS OR MORE TO FULLY RESPOND TO A CHANGE IN LIGHT INTENSITY (THAT'S PRETTY SLOW). THEY MAY REQUIRE MANY MINUTES TO RETURN TO THEIR NORMAL DARK RESISTANCE WHEN LIGHT IS REMOVED (THE MEMORY EFFECT).

2. THE SEMICONDUCTOR MOST OFTEN USED IN PHOTORESISTORS IS CADMIUM SULFIDE. ITS SENSITIVITY TO LIGHT IS VERY SIMILAR TO THAT OF THE HUMAN EYE! LEAD SULFIDE IS USED TO DETECT INFRARED (OUT TO 3-MICROMETERS).

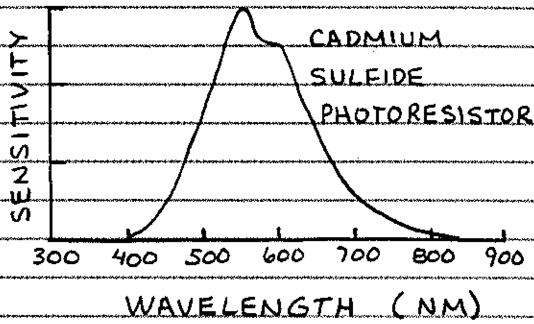

☐ KINDS OF PHOTORESISTORS — MANY DIFFERENT KINDS ARE AVAILABLE. IN MOST THE LIGHT SENSITIVE SEMICONDUCTOR IS COATED BETWEEN INTERLEAVED ELECTRODES TO INCREASE THE EXPOSED SURFACE. A PLASTIC OR GLASS WINDOW MAY OR MAY NOT BE USED.

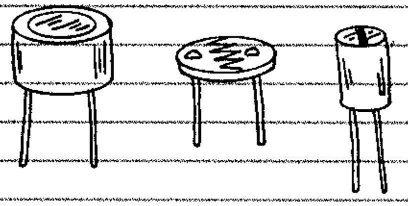

☐ PHOTORESISTOR SYMBOL. BOTH SYMBOLS SHOWN HERE ARE USED.

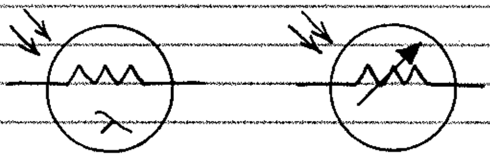

HOW PHOTORESISTORS ARE USED

P. 114

PHOTORESISTORS ARE USED IN LIGHT CONTROLLED RELAYS AND LIGHT METERS.

☐ LIGHT METER. THE ARRANGEMENT SHOWN HERE INDICATES ON A CURRENT METER THE INTENSITY OF LIGHT ILLUMINATING A CADMIUM SULFIDE PHOTORESISTOR.

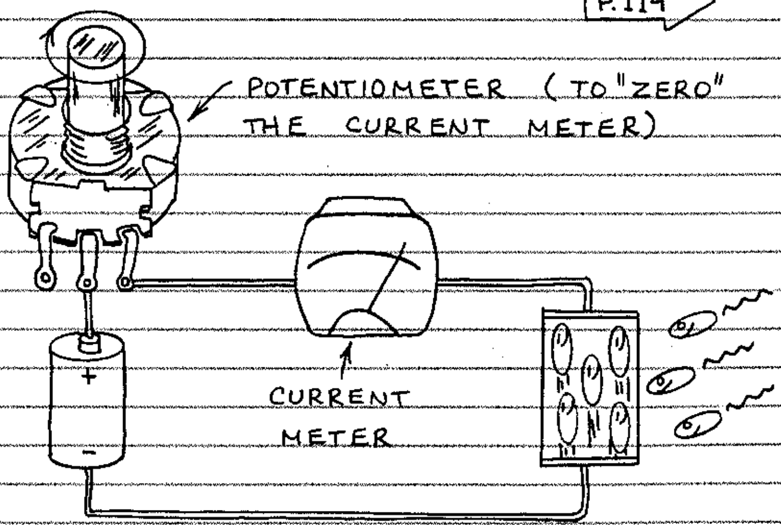

PN JUNCTION LIGHT DETECTORS

PN JUNCTION LIGHT DETECTORS FORM THE LARGEST FAMILY OF PHOTONIC SEMICONDUCTORS. MOST ARE MADE FROM SILICON AND CAN DETECT BOTH VISIBLE LIGHT AND NEAR-INFRARED.

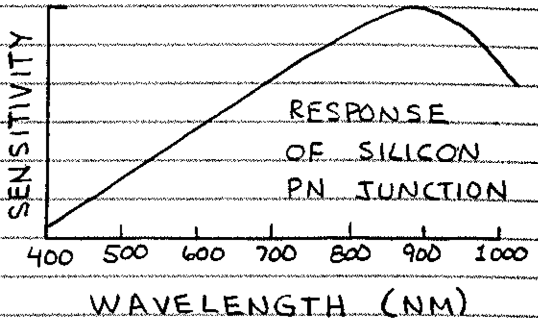

PHOTODIODES

ALL PN JUNCTIONS ARE LIGHT SENSITIVE. PHOTODIODES ARE PN JUNCTIONS SPECIFICALLY DESIGNED FOR LIGHT DETECTION. THEY ARE USED IN CAMERAS, INTRUSION ALARMS, LIGHTWAVE COMMUNICATORS, ETC.

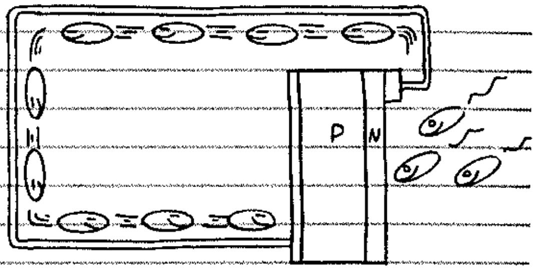

☐ PHOTODIODE OPERATION.
A PHOTON WILL CREATE A HOLE-ELECTRON PAIR AT A PN JUNCTION. A CURRENT WILL FLOW IF THE TWO SIDES OF THE JUNCTION ARE CONNECTED. TWO OPERATING MODES ARE POSSIBLE:

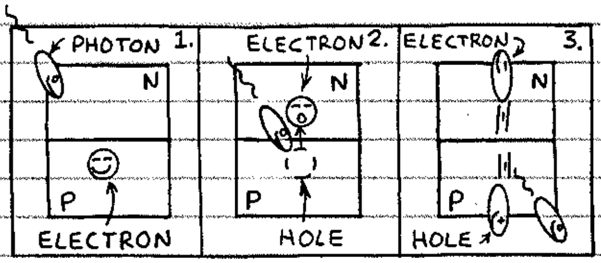
PHOTON 1.
N
P
ELECTRON
ELECTRON 2.
N
P
HOLE
ELECTRON 3.
N
P
HOLE

1. PHOTOVOLTAIC OPERATION — HERE THE PHOTODIODE BECOMES A CURRENT SOURCE WHEN IT IS ILLUMINATED.

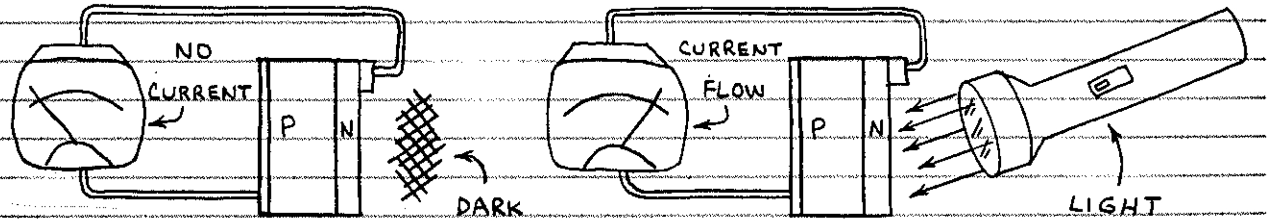
NO CURRENT
P
N
DARK
CURRENT FLOW
P
N
LIGHT

2. PHOTOCONDUCTIVE OPERATION — HERE THE PHOTODIODE IS REVERSE-BIASED. A CURRENT FLOWS WHEN THE PN JUNCTION IS ILLUMINATED. (WHEN DARK, A TINY CURRENT CALLED THE DARK CURRENT WILL FLOW.)

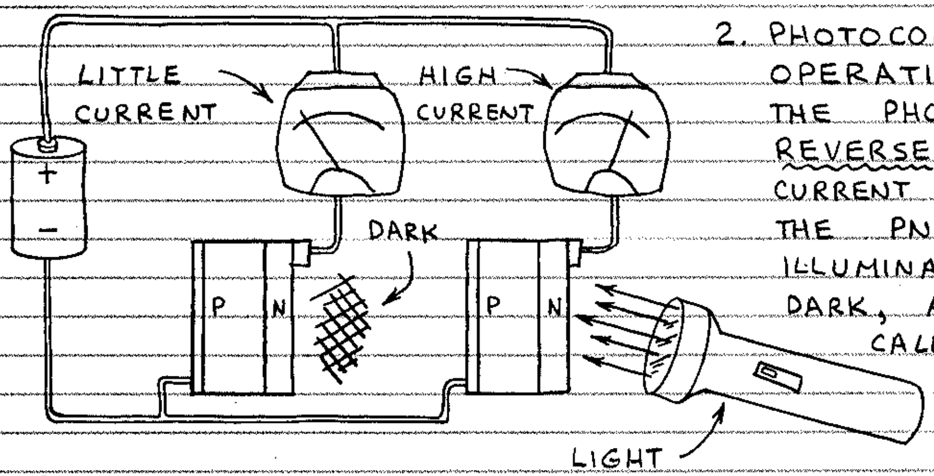
LITTLE CURRENT
HIGH CURRENT
+
−
DARK
P
N
P
N
LIGHT

☐ KINDS OF PHOTODIODES. SHOWN HERE IS A TYPICAL PHOTODIODE. MANY OTHER CASE STYLES ARE ALSO USED (PLASTIC HOUSING, BUILT-IN LENSES AND FILTERS, ETC.). THE MOST IMPORTANT DISTINCTION IS THE SIZE OF THE SEMICONDUCTOR CHIP. SPECIALIZED CHIP DESIGNS MAY BE USED TO GIVE BETTER RESPONSE TO CERTAIN WAVELENGTHS OF LIGHT.

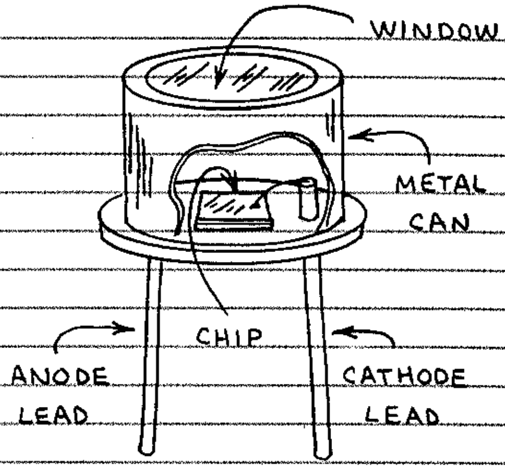
WINDOW
METAL CAN
CHIP
ANODE LEAD
CATHODE LEAD

FAMOUS FACT: LEDs CAN BOTH EMIT AND DETECT LIGHT!

SMALL AREA PHOTODIODES.

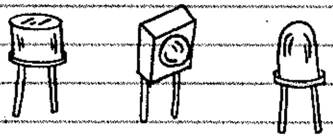

THESE PHOTODIODES HAVE VERY FAST RESPONSE TIMES WHEN USED IN THE REVERSE-BIASED PHOTOCONDUCTIVE MODE.

LARGE AREA PHOTODIODES.

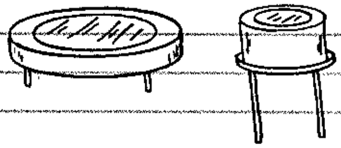

THOUGH SLOWER RESPONDING THAN SMALL AREA PHOTODIODES, THEIR LARGE AREA PROVIDES HIGH SENSITIVITY.

☐ PHOTODIODE SYMBOL. BOTH SYMBOLS SHOWN HERE ARE USED.

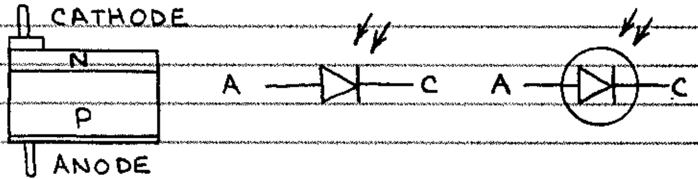
CATHODE
N
P
ANODE
A
C
A
C

HOW PHOTODIODES ARE USED

PHOTODIODES ARE COMMONLY USED TO DETECT FAST PULSES OF NEAR INFRARED (AS IN LIGHTWAVE COMMUNICATIONS).

☐ LIGHT METER. THIS ARRANGEMENT PROVIDES A BASIC PHOTOCONDUCTIVE MODE LIGHT METER. ITS RESPONSE IS VERY LINEAR.

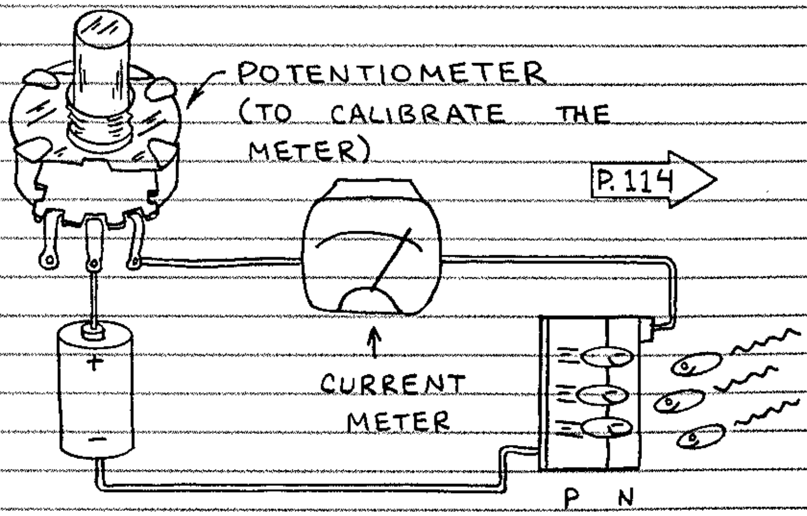
POTENTIOMETER (TO CALIBRATE THE METER)
+
−
CURRENT METER
P
N

P. 114

PHOTOTRANSISTORS

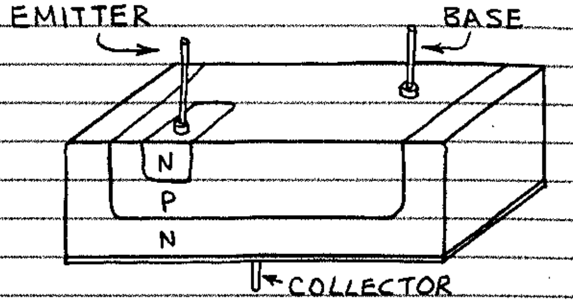
EMITTER
BASE
N
P
N
COLLECTOR

ALL TRANSISTORS ARE LIGHT SENSITIVE. PHOTOTRANSISTORS ARE SPECIFICALLY DESIGNED TO TAKE ADVANTAGE OF THIS IMPORTANT PROPERTY. LIGHT SENSITIVE FETs ARE AVAILABLE, BUT THE MOST COMMON PHOTO-TRANSISTOR IS AN NPN JUNCTION TRANSISTOR WITH A LARGE, EXPOSED BASE REGION. PHOTONS ENTERING THE BASE REPLACE THE BASE-EMITTER CURRENT OF ORDINARY NPN TRANSISTORS. THEREFORE A PHOTOTRANSISTOR DIRECTLY AMPLIFIES VARIATIONS IN THE NUMBER OF PHOTONS.

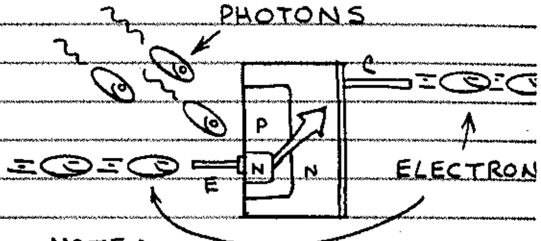
PHOTONS
C
P
N
N
E
ELECTRON

NOTE:
BASE LEAD IS OPTIONAL.

☐ NPN PHOTOTRANSISTOR OPERATION— TWO TYPES OF NPN PHOTOTRANSISTORS ARE AVAILABLE. ONE IS AN NPN TRANSISTOR AS SHOWN ABOVE. THE OTHER INCLUDES A SECOND NPN TRANSISTOR TO PROVIDE MORE AMPLIFICATION.

1. NPN PHOTOTRANSISTOR.

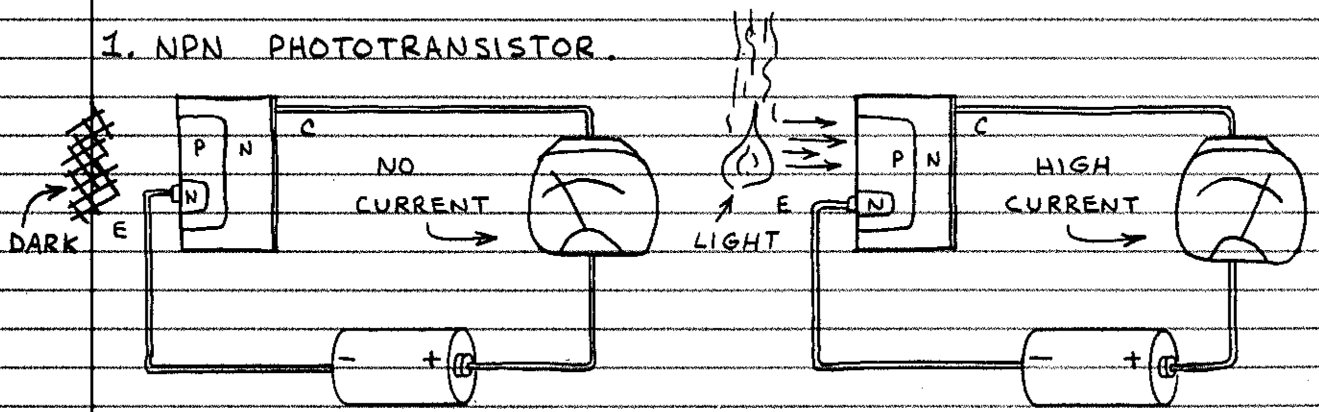
DARK
P
N
N
E
C
NO CURRENT
−
+
LIGHT
P
N
N
E
C
HIGH CURRENT
−
+

2. PHOTODARLINGTON TRANSISTOR.

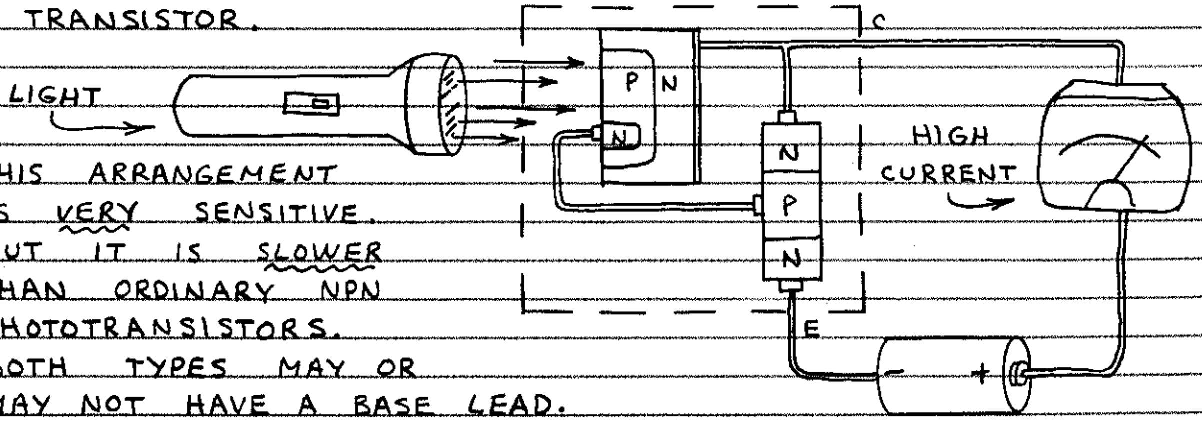
LIGHT
P
N
N
C
N
P
N
E
HIGH CURRENT
−
+

THIS ARRANGEMENT IS VERY SENSITIVE. BUT IT IS SLOWER THAN ORDINARY NPN PHOTOTRANSISTORS. BOTH TYPES MAY OR MAY NOT HAVE A BASE LEAD.

☐ KINDS OF PHOTOTRANSISTORS. SHOWN HERE IS A TYPICAL LOW COST NPN PHOTOTRANSISTOR. MANY OTHER CASE STYLES ARE ALSO USED (METAL CANS, GLASS LENSES, FLAT WINDOWS, ETC.). IMPORTANT: THE BASE LEAD MAY OR MAY NOT BE PRESENT. MANY PHOTOTRANSISTOR CIRCUITS DO NOT USE THE BASE CONNECTION.

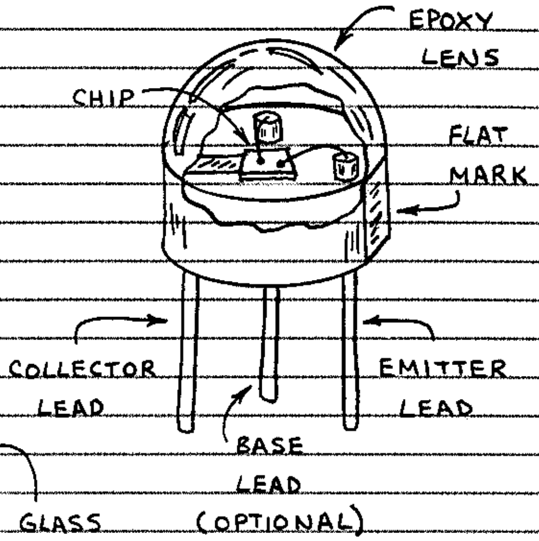
EPOXY LENS
CHIP
FLAT MARK
COLLECTOR LEAD
EMITTER LEAD
BASE LEAD (OPTIONAL)

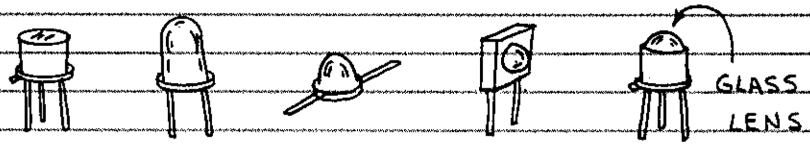

THESE ARE TYPICAL PHOTOTRANSISTORS.

☐ PHOTOTRANSISTOR SYMBOLS.

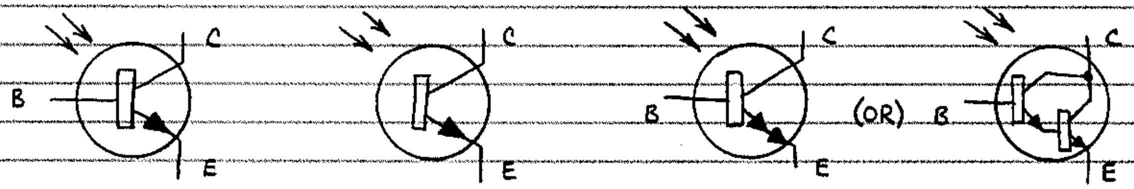
B
C
E
C
E
B
C
E
(OR)
B
C
E

NPN NPN (NO BASE LEAD) PHOTODARLINGTONS

HOW PHOTOTRANSISTORS ARE USED

PHOTOTRANSISTORS ARE OFTEN USED TO DETECT FLUCTUATING (AC) LIGHT SIGNALS. THIS ARRANGEMENT USES A STEADY (DC) LIGHT TO ENERGIZE A RELAY.

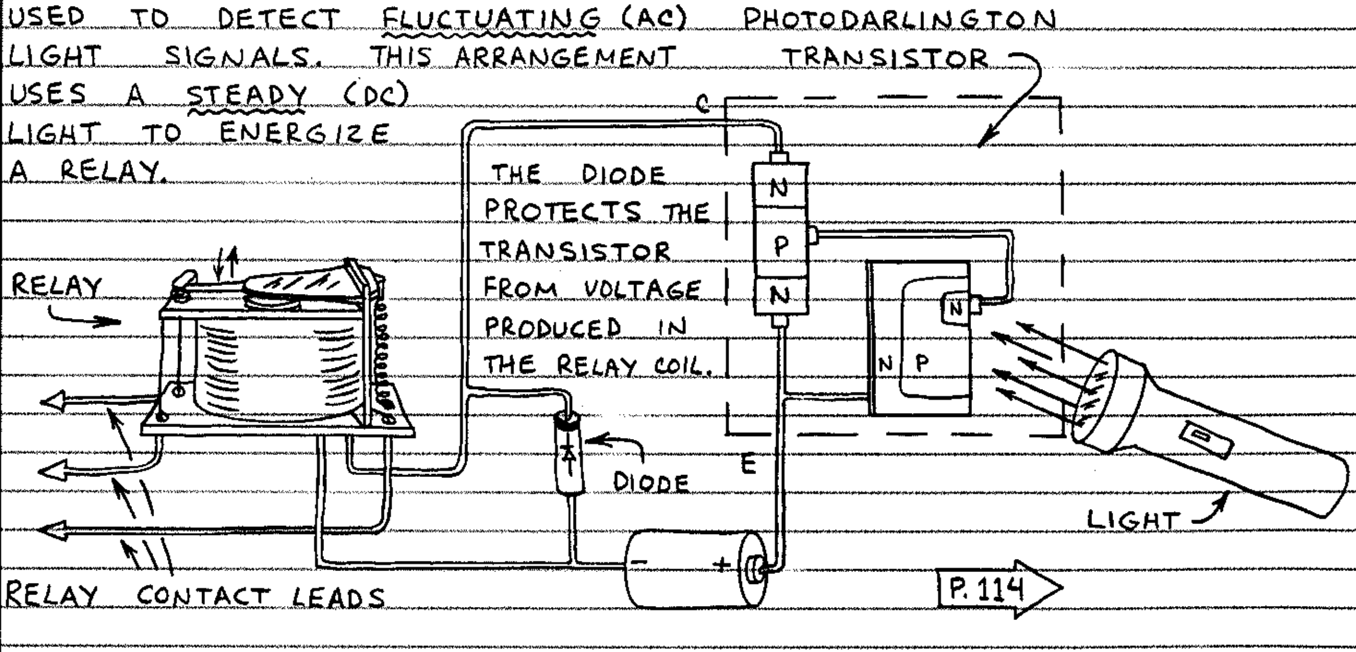
PHOTODARLINGTON TRANSISTOR
C
N
P
N
N
N
P
E
THE DIODE PROTECTS THE TRANSISTOR FROM VOLTAGE PRODUCED IN THE RELAY COIL.
RELAY
DIODE
–
+
LIGHT
RELAY CONTACT LEADS

P. 114

PHOTOTHYRISTORS

PHOTOTHYRISTORS ARE VARIOUS KINDS OF LIGHT-ACTIVATED THYRISTORS. YOU CAN THINK OF THEM AS LIGHT-ACTIVATED SWITCHES. THE MOST IMPORTANT MEMBER OF THE FAMILY IS THE LIGHT-ACTIVATED SILICON CONTROLLED RECTIFIER (LASCR). LIGHT ACTIVATED TRIACS ARE ALSO MADE. NEITHER CAN SWITCH AS MUCH CURRENT AS CONVENTIONAL THYRISTORS.

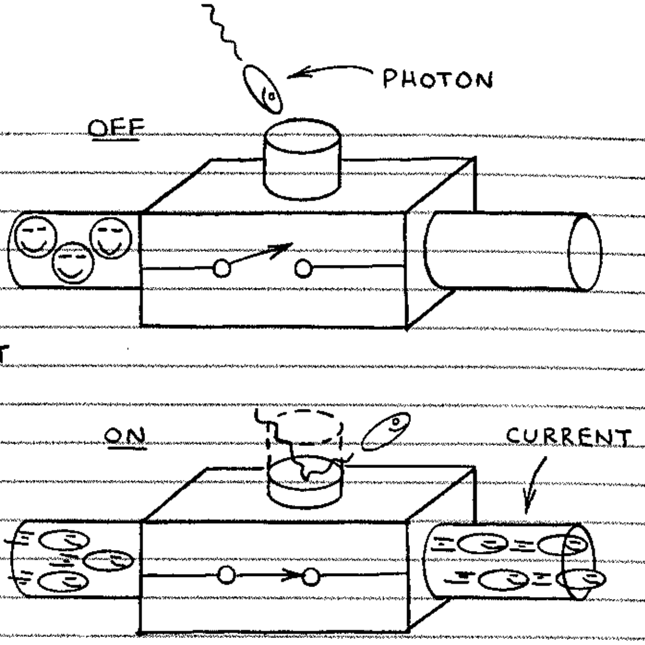
PHOTON
OFF
ON
CURRENT

LIGHT ACTIVATED SCRs (LASCRs)

TO IMPROVE THEIR SENSITIVITY TO LIGHT, THE LASCR IS MADE THINNER THAN STANDARD SCRs. THIS LIMITS THE AMOUNT OF CURRENT THEY CAN SWITCH. FOR HIGH CURRENT APPLICATIONS A LASCR CAN BE USED TO TRIGGER A CONVENTIONAL SCR.

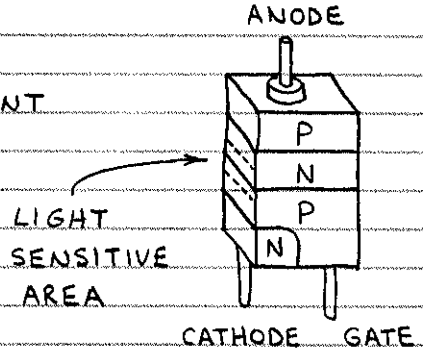
ANODE
P
N
P
N
LIGHT SENSITIVE AREA
CATHODE
GATE

☐ KINDS OF LASCRs.

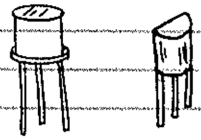

MOST LASCRs CAN SWITCH UP TO A FEW HUNDRED VOLTS. MAXIMUM CURRENT IS ONLY A FEW TENTHS OF AN AMPERE.

HOW LASCRs ARE USED

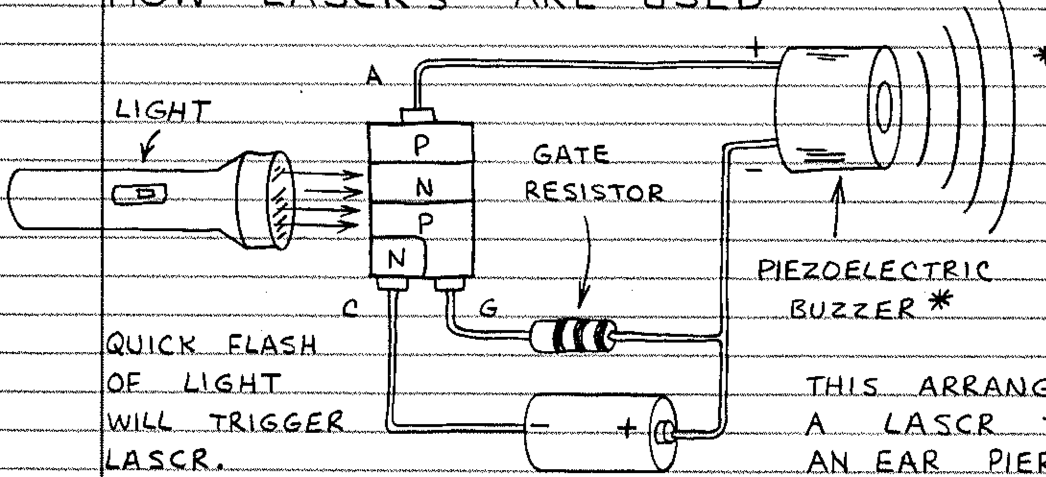
A
LIGHT
P
N
P
N
C
G
GATE RESISTOR
+
-
PIEZOELECTRIC BUZZER *
-
+

*ONCE TRIGGERED, BUZZER REMAINS ON UNTIL POWER IS INTERRUPTED.

QUICK FLASH OF LIGHT WILL TRIGGER LASCR.

P. 115

THIS ARRANGEMENT ALLOWS A LASCR TO TURN ON AN EAR PIERCING BUZZER.

SOLAR CELLS

SOLAR CELLS ARE PN JUNCTION PHOTODIODES WITH AN EXCEPTIONALLY LARGE LIGHT SENSITIVE AREA. A SINGLE SILICON SOLAR CELL GENERATES 0.5 VOLT IN BRIGHT SUNLIGHT.

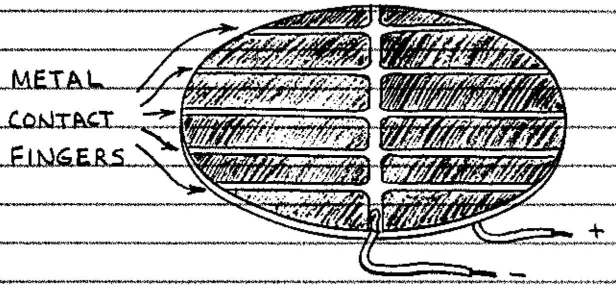
METAL
CONTACT
FINGERS
+
−

☐ SOLAR CELL OPERATION

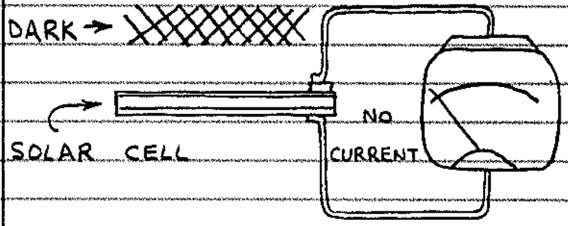
DARK
SOLAR CELL
NO CURRENT

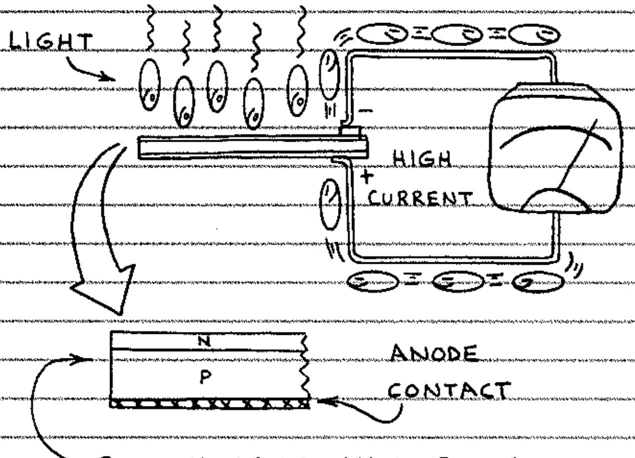
LIGHT
−
+
HIGH CURRENT
N
P
ANODE CONTACT

SILICON CELLS MAY BE P ON N.

A CELL THIS SIZE GENERATES 0.1 AMPERE.*

* IN BRIGHT SUNLIGHT.

☐ KINDS OF SOLAR CELLS.

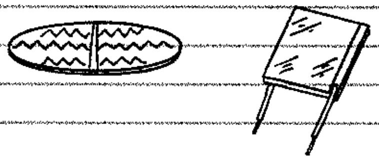

MANY DIFFERENT KINDS OF SILICON SOLAR CELLS ARE MADE. OFTEN INDIVIDUAL CELLS ARE CONNECTED IN SERIES OR PARALLEL.

SERIES: OUTPUT VOLTAGE IS SUM OF CELL VOLTAGES.

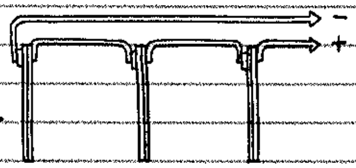

PARALLEL: OUTPUT CURRENT IS SUM OF CELL CURRENTS.

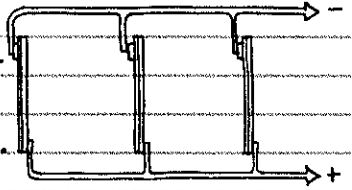

☐ SOLAR CELL SYMBOL.

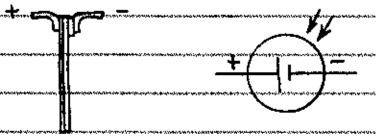

REMEMBER, CELLS MAY BE P ON N.

HOW SOLAR CELLS ARE USED

P. 115

ARRAYS OF SOLAR CELLS CAN CHARGE RECHARGEABLE CELLS AND BATTERIES.

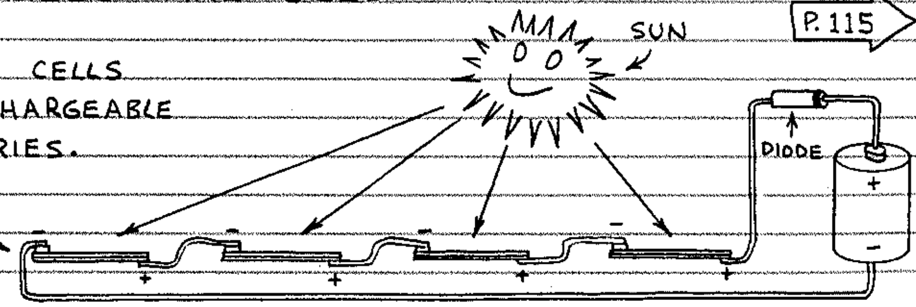

SERIES CONNECTED SOLAR CELL ARRAY

5. INTEGRATED CIRCUITS

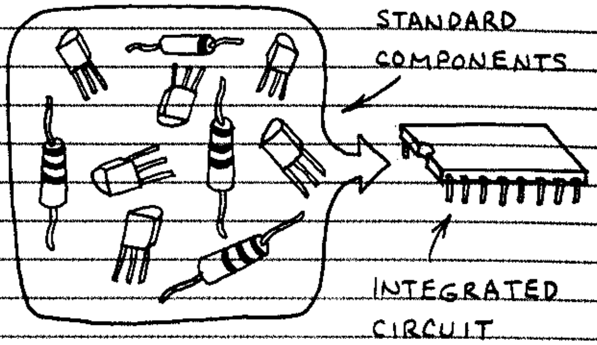

ELECTRONIC CIRCUITS CAN BE MADE BY SIMULTANEOUSLY FORMING INDIVIDUAL TRANSISTORS, DIODES AND RESISTORS ON A SMALL CHIP OF SILICON. THE COMPONENTS ARE CONNECTED TO ONE ANOTHER WITH ALUMINUM "WIRES" DEPOSITED ON THE SURFACE OF THE CHIP. THE RESULT IS AN INTEGRATED CIRCUIT. INTEGRATED CIRCUITS (OR IC'S) CAN CONTAIN AS FEW AS SEVERAL TO AS MANY AS HUNDREDS OF THOUSANDS OF TRANSISTORS. THEY HAVE MADE POSSIBLE VIDEO GAMES, DIGITAL WATCHES, AFFORDABLE COMPUTERS AND MANY OTHER VERY SOPHISTICATED PRODUCTS. HERE'S A SIMPLIFIED AND HIGHLY MAGNIFIED VIEW OF A SECTION OF A *BIPOLAR* INTEGRATED CIRCUIT:

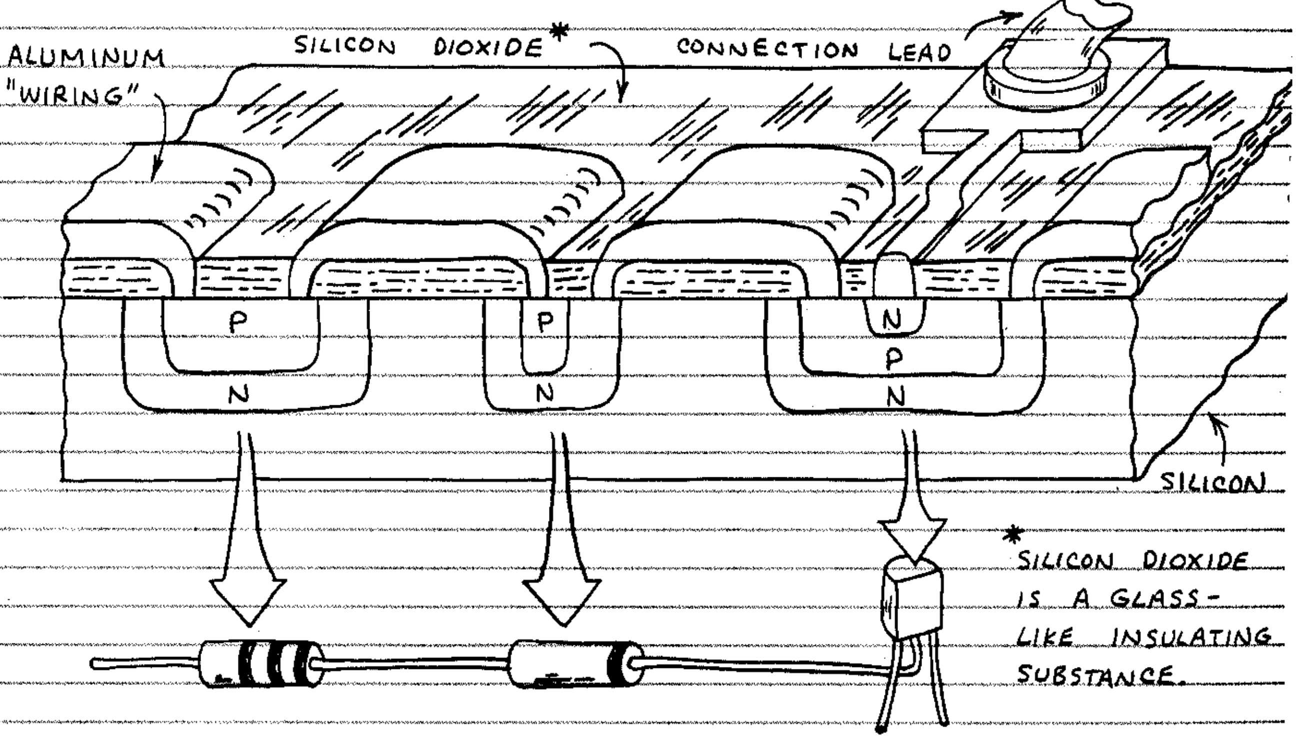

* SILICON DIOXIDE IS A GLASS-LIKE INSULATING SUBSTANCE.

RESISTOR — A SMALL SECTION OF P-TYPE SILICON FORMS A RESISTOR.

DIODE — A PN JUNCTION FORMS A DIODE.

TRANSISTOR — A PAIR OF PN JUNCTIONS FORMS AN NPN TRANSISTOR

OF COURSE THE CONVENTIONAL COMPONENTS SHOWN BELOW THE HIGHLY MAGNIFIED SECTION OF THE IC ARE NOT DRAWN TO THE SAME SCALE. FOR EXAMPLE, ONE KIND OF IC INCLUDES 262,144 TRANSISTORS ON A SILICON CHIP ONLY ABOUT 1/4 INCH SQUARE!

☐ KINDS OF INTEGRATED CIRCUITS — INTEGRATED CIRCUITS ARE GROUPED INTO TWO MAJOR CATEGORIES:

1. ANALOG (OR LINEAR) IC'S PRODUCE, AMPLIFY OR RESPOND TO VARIABLE VOLTAGES. ANALOG IC'S INCLUDE MANY KINDS OF AMPLIFIERS, TIMERS, OSCILLATORS AND VOLTAGE REGULATORS.

2. DIGITAL (OR LOGIC) IC'S RESPOND TO OR PRODUCE SIGNALS HAVING ONLY TWO VOLTAGE LEVELS. DIGITAL IC'S INCLUDE MICROPROCESSORS, MEMORIES, MICROCOMPUTERS AND MANY KINDS OF SIMPLER CHIPS.

SOME IC'S COMBINE ANALOG AND DIGITAL FUNCTIONS ON A SINGLE CHIP. FOR EXAMPLE, A DIGITAL CHIP MAY INCLUDE A BUILT-IN ANALOG VOLTAGE REGULATOR SECTION. AND AN ANALOG TIMER CHIP MAY INCLUDE AN ON-CHIP DIGITAL COUNTER TO GIVE MUCH LONGER TIME DELAYS THAN POSSIBLE WITH THE TIMER ALONE.

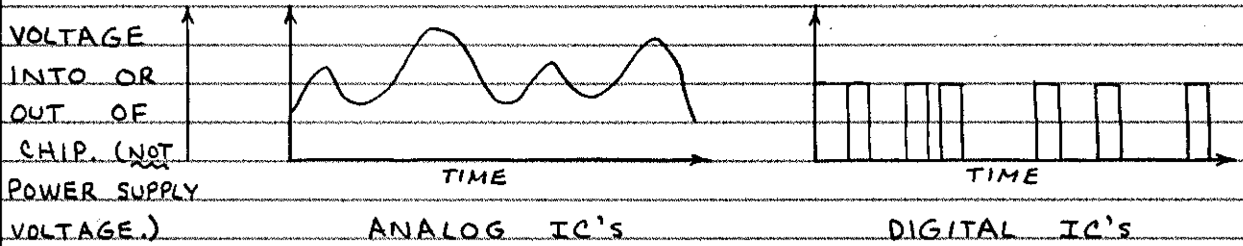

☐ KINDS OF INTEGRATED CIRCUIT PACKAGES — IC CHIPS ARE SUPPLIED IN MANY DIFFERENT PACKAGES. BY FAR THE MOST COMMON ARE VARIATIONS OF THE DUAL IN-LINE PACKAGE (OR DIP). THE DIP IS MADE FROM PLASTIC (CHEAP) OR CERAMIC (MORE ROBUST). MOST DIPS HAVE 14 OR 16 PINS, BUT THE PIN COUNT CAN RANGE FROM 4 TO 64. HERE'S A TYPICAL DIP:

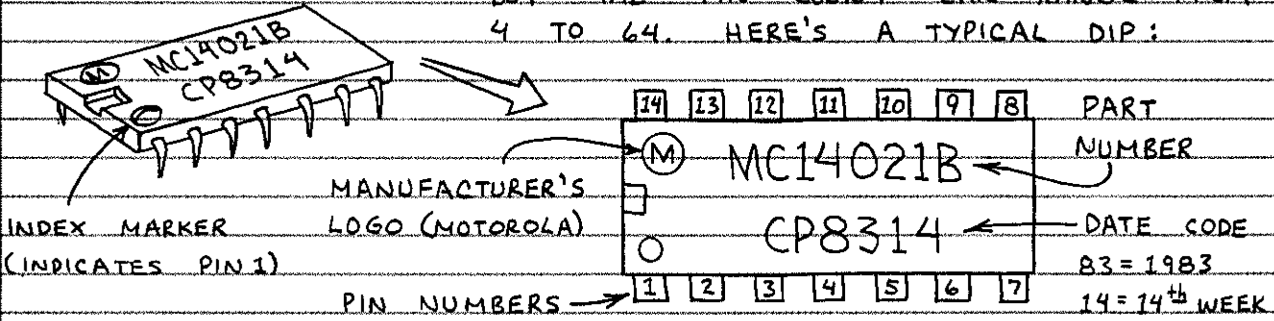
MC14021B
CP8314
14
13
12
11
10
9
8
PART NUMBER
MANUFACTURER'S LOGO (MOTOROLA)
INDEX MARKER (INDICATES PIN 1)
DATE CODE
83 = 1983
14 = 14th WEEK
PIN NUMBERS
1
2
3
4
5
6
7

ANOTHER IC PACKAGE IS THE TO-5 METAL CAN. THOUGH VERY STURDY, IT'S BEING REPLACED IN MANY CASES BY CHEAPER PLASTIC DIPS.

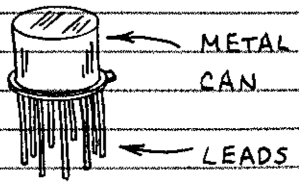
METAL CAN
LEADS

6. DIGITAL INTEGRATED CIRCUITS

NO MATTER HOW COMPLICATED, *ALL* DIGITAL INTEGRATED CIRCUITS ARE MADE FROM SIMPLE BUILDING BLOCKS CALLED *GATES*. GATES ARE LIKE ELECTRONICALLY CONTROLLED SWITCHES. THEY ARE EITHER *ON* OR *OFF*. HOW DO GATES WORK? LET'S START WITH THE BASICS...

MECHANICAL SWITCH GATES

THE THREE SIMPLEST GATES CAN BE DEMONSTRATED WITH SOME PUSHBUTTON SWITCHES, A BATTERY AND A LAMP.

☐ SWITCH "AND" GATE.

THE LAMP GLOWS ONLY WHEN SWITCHES A *AND* B ARE CLOSED. THE TABLE SUMMARIZES THE GATE'S OPERATION. IT'S CALLED A *TRUTH TABLE*.

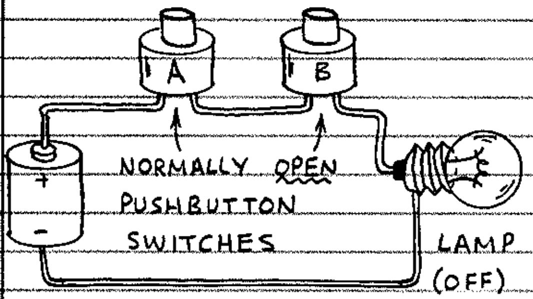
A
B
+
-
NORMALLY OPEN
PUSHBUTTON
SWITCHES
LAMP
(OFF)

OPEN SWITCH = OFF
CLOSED SWITCH = ON

A	B	OUT
OFF	OFF	OFF
OFF	ON	OFF
ON	OFF	OFF
ON	ON	ON

ALL POSSIBLE ON-OFF COMBINATIONS

☐ SWITCH "OR" GATE.

THE LAMP GLOWS ONLY WHEN SWITCH A *OR* SWITCH B *OR* BOTH SWITCHES A AND B ARE CLOSED. HERE'S THE TRUTH TABLE:

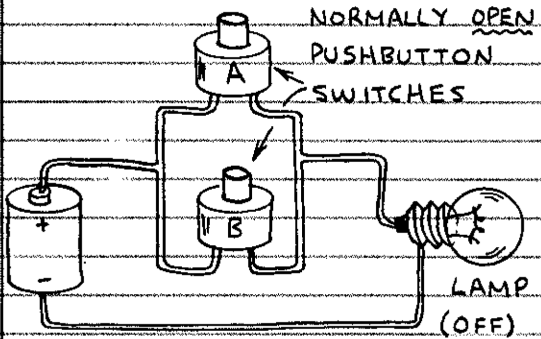
NORMALLY OPEN
PUSHBUTTON
SWITCHES
A
B
+
-
LAMP
(OFF)

THE SWITCHES ARE THE GATE'S *INPUTS*. THE LEAD WITHOUT SWITCHES IS THE *COMMON* OR *GROUND* LEAD.

A	B	OUT
OFF	OFF	OFF
OFF	ON	ON
ON	OFF	ON
ON	ON	ON

☐ SWITCH "NOT" GATE.

THE LAMP NORMALLY GLOWS. ONLY WHEN THE SWITCH IS OPENED IS THE LAMP OFF. IN OTHER WORDS, THE "NOT" GATE REVERSES (INVERTS) THE USUAL ACTION OF A SWITCH. HERE'S THE TRUTH TABLE:

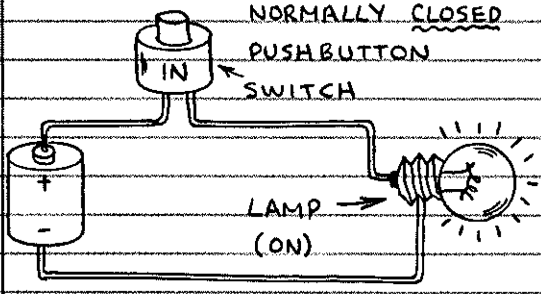
NORMALLY CLOSED
PUSHBUTTON
SWITCH
IN
+
-
LAMP
(ON)

THE "NOT" GATE IS USUALLY CALLED THE *INVERTER*.

IN	OUT
OFF	ON
ON	OFF

THE BINARY CONNECTION

IT'S POSSIBLE TO SUBSTITUTE THE DIGITS 0 AND 1 FOR THE OFF AND ON STATES OF A SWITCH. THE TRUTH TABLES FOR THE GATES ON THE PREVIOUS PAGE THEN BECOME:

"AND" GATE

A	B	OUT
0	0	0
0	1	0
1	0	0
1	1	1

"OR" GATE

A	B	OUT
0	0	0
0	1	1
1	0	1
1	1	1

"NOT" GATE

IN	OUT
0	1
1	0

THE 0 AND 1 INPUT (A & B) COMBINATIONS FORM NUMBERS IN THE TWO DIGIT (OR BIT) BINARY NUMBER SYSTEM. IN DIGITAL ELECTRONICS, BINARY NUMBERS SERVE AS CODES THAT REPRESENT DECIMAL NUMBERS, LETTERS OF THE ALPHABET, VOLTAGES AND MANY OTHER KINDS OF INFORMATION.

DECIMAL	BINARY	BINARY-CODED DECIMAL (BCD)
0	0	0000 0000
1	1	0000 0001
2	10	0000 0010
3	11	0000 0011
4	100	0000 0100
5	101	0000 0101
6	110	0000 0110
7	111	0000 0111
8	1000	0000 1000
9	1001	0000 1001
10	1010	0001 0000
11	1011	0001 0001
12	1100	0001 0010
13	1101	0001 0011
14	1110	0001 0100
15	1111	0001 0101

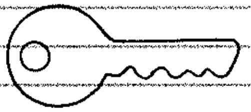

BINARY FACTS

A BINARY 0 OR 1 IS A BIT.
A PATTERN OF 4 BITS IS A NIBBLE.
A PATTERN OF 8 BITS IS A BYTE.

BCD— EACH DECIMAL DIGIT IS ASSIGNED ITS BINARY EQUIVALENT. NOTE THAT LEADING ZEROS ARE SHOWN. IN DIGITAL ELECTRONICS ALL BIT LOCATIONS ARE OCCUPIED.

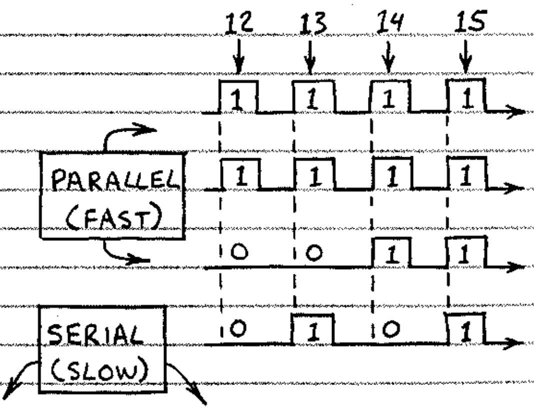

BINARY NUMBERS CAN BE SENT THROUGH WIRES (BUSES) ALL AT ONCE (PARALLEL) OR A BIT AT A TIME (SERIAL). SHOWN HERE ARE SERIAL AND PARALLEL TRANSMISSION OF 15... 14... 13... 12.

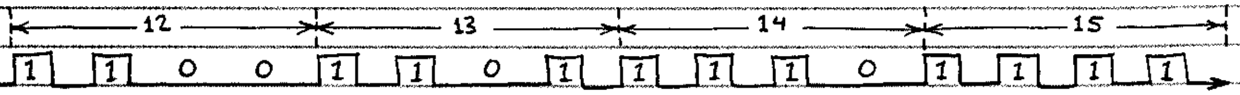

DIODE GATES

OFTEN IT'S DESIRABLE TO CONTROL A GATE ELECTRICALLY RATHER THAN MECHANICALLY. THE SIMPLEST ELECTRICALLY CONTROLLED GATE USES PN JUNCTION DIODES THAT ARE SWITCHED ON (FORWARD BIAS) OR OFF (REVERSE BIAS) BY AN INPUT SIGNAL OF SEVERAL VOLTS (BINARY 1 OR HIGH) OR AN INPUT NEAR OR AT GROUND (BINARY 0 OR LOW).

☐ DIODE "OR" GATE

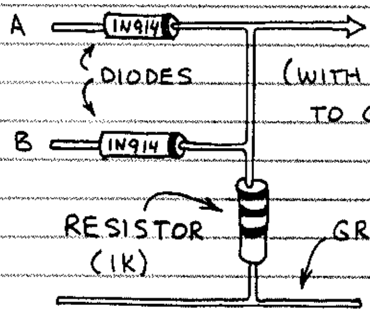

WHEN THE INPUT VOLTAGE AT A OR B IS MORE POSITIVE THAN GROUND, IT PASSES THROUGH THE FORWARD BIASED DIODE(S) AND APPEARS AT THE OUTPUT. OTHERWISE THE OUTPUT IS AT OR NEAR GROUND. THE TRUTH TABLE IS VALID FOR INPUTS OF 0 VOLT (0 OR LOW) AND +6 VOLTS (1 OR HIGH).

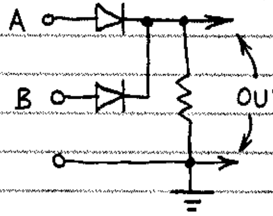

A	B	OUT
0V	0V	0V
0V	6V	5.4V
6V	0V	5.4V
6V	6V	5.4V

THE OUTPUT DOES NOT REACH A FULL 6 VOLTS WHEN HIGH BECAUSE THE DIODES REQUIRE A FORWARD VOLTAGE OF 0.6 VOLT. THIS VOLTAGE IS SUBTRACTED FROM THE OUTPUT VOLTAGE. (IN ELECTRONICS JARGON A SILICON DIODE CAUSES A "VOLTAGE DROP" OF 0.6 VOLT.)

☐ DIODE "AND" GATE

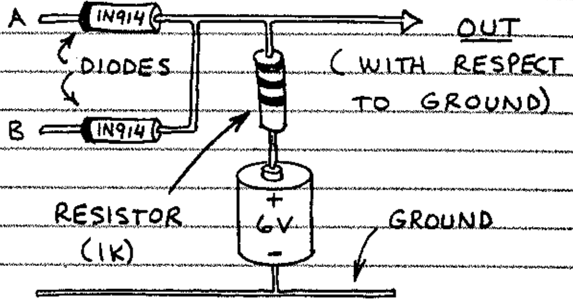

WHEN THE INPUT VOLTAGE AT A AND B IS MORE POSITIVE THAN GROUND, CURRENT FLOWS FROM THE BATTERY THROUGH THE RESISTOR TO THE OUTPUT. IF EITHER A OR B IS AT OR NEAR GROUND, ONE OR BOTH DIODES BECOME FORWARD BIASED AND CURRENT FLOWS AWAY FROM THE OUTPUT.

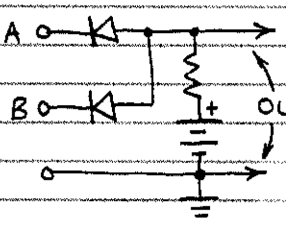

A	B	OUT
0V	0V	0V
0V	6V	.5V
6V	0V	.5V
6V	6V	5.4V

AS CIRCUITS BECOME MORE COMPLICATED, PICTORIAL VIEWS ARE NOT PRACTICAL. THAT'S WHY THIS PAGE INTRODUCES CIRCUIT DIAGRAMS FOR EACH OF THE TWO PICTORIALS SHOWN ABOVE. WE'LL FIND OUT MORE ABOUT CIRCUIT DIAGRAMS LATER. IN THE MEANTIME, THE NEXT PAGE SHOWS MORE OF THEM...

TRANSISTOR GATES

THE VOLTAGE DROP OF DIODE GATES MEANS AMPLIFICATION IS REQUIRED IN ORDER TO CONNECT TOGETHER A SERIES OF GATES. WHILE TRANSISTORS CAN PROVIDE THE NECESSARY AMPLIFICATION, TRANSISTORS CAN FUNCTION AS GATES! BOTH BIPOLAR AND FIELD-EFFECT TRANSISTORS CAN BE USED. ON THIS PAGE ARE SHOWN CIRCUIT DIAGRAMS FOR SOME OF THE SIMPLEST BIPOLAR TRANSISTOR GATES. TOGETHER THEY FORM THE RESISTOR-TRANSISTOR DIGITAL LOGIC FAMILY. YOU CAN ACTUALLY MAKE THESE GATES. BUT THE MAIN REASON THEY'RE HERE IS TO GIVE YOU AN APPRECIATION FOR THE INTEGRATED CIRCUIT GATES WE'LL BE LOOKING AT SHORTLY....

☐ "NOT" GATE (INVERTER)

WHEN IN IS AT +V (BINARY 1 OR HIGH), TRANSISTOR Q1 SWITCHES ON AND CONNECTS OUT DIRECTLY TO GROUND (BINARY 0 OR LOW). WHEN IN IS LOW, Q1 SWITCHES OFF AND OUT BECOMES (THROUGH R1) +V. "NOT" GATES LIKE THIS MAKE POSSIBLE IMPORTANT NEW LOGIC GATES.

IN	OUT
L	H
H	L

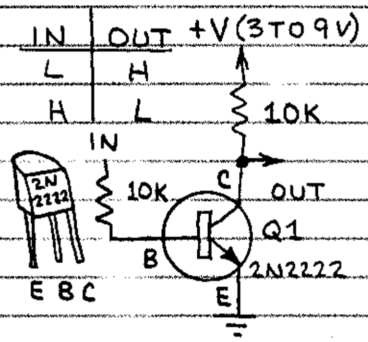

☐ "AND" GATE

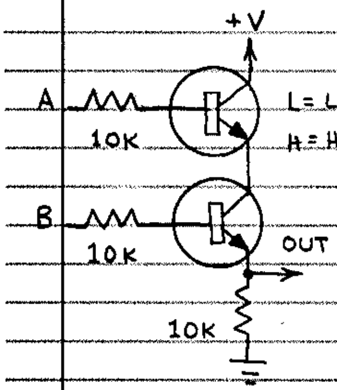

L = LOW
H = HIGH

A	B	OUT
L	L	L
L	H	L
H	L	L
H	H	H

USE 2N2222 OR ANY COMMON NPN TRANSISTOR FOR ALL THESE GATES.

☐ "NAND" (NOT-AND) GATE

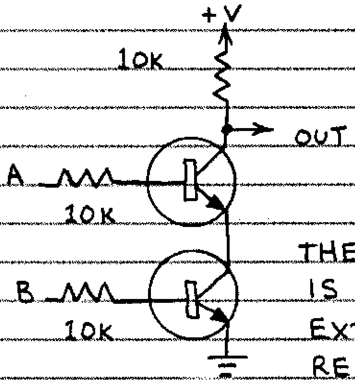

A	B	OUT
L	L	H
L	H	H
H	L	H
H	H	L

THE "NOT" FUNCTION IS "BUILT-IN" (NO EXTRA TRANSISTOR REQUIRED).

☐ "OR" GATE

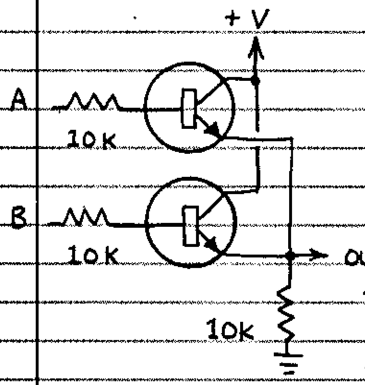

A	B	OUT
L	L	L
L	H	H
H	L	H
H	H	H

+V FOR ALL THESE GATES CAN BE +3 TO +9 VOLTS.

☐ "NOR" (NOT-OR) GATE

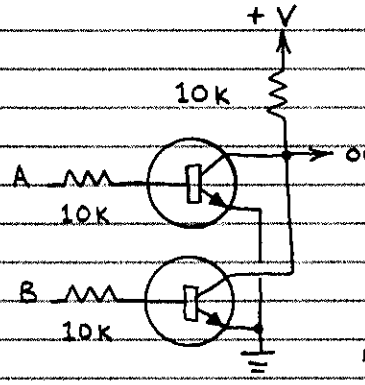

A	B	OUT
L	L	H
L	H	L
H	L	L
H	H	L

LIKE THE "NAND" GATE, THE "NOT" FUNCTION IS "BUILT-IN."

GATE SYMBOLS

BEFORE MOVING ON TO DIGITAL INTEGRATED CIRCUITS, LET'S LOOK AT THE SYMBOLS FOR THE VARIOUS KINDS OF GATES. THIS IS ALSO A GOOD PLACE TO INTRODUCE SEVERAL GATES WE'VE NOT YET ENCOUNTERED.

☐ "AND" GATE

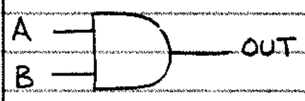

A	B	OUT
L	L	L
L	H	L
H	L	L
H	H	H

☐ "NAND" GATE

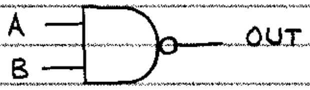

A	B	OUT
L	L	H
L	H	H
H	L	H
H	H	L

☐ "OR" GATE

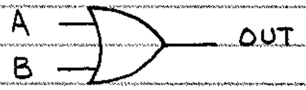

A	B	OUT
L	L	L
L	H	H
H	L	H
H	H	H

☐ "NOR" GATE

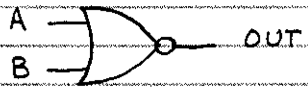

A	B	OUT
L	L	H
L	H	L
H	L	L
H	H	L

☐ "EXCLUSIVE-OR" GATE

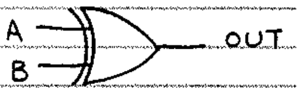

A	B	OUT
L	L	L
L	H	H
H	L	H
H	H	L

☐ "EXCLUSIVE - NOR" GATE

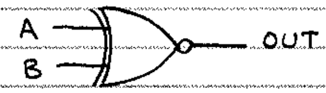

A	B	OUT
L	L	H
L	H	L
H	L	L
H	H	H

☐ LOGIC GATES WITH MORE THAN TWO INPUTS — THE GATES SHOWN ABOVE ARE CALLED LOGIC CIRCUITS BECAUSE THEY MAKE LOGICAL DECISIONS. LOGIC GATES OFTEN HAVE MORE THAN TWO INPUTS. ADDITIONAL INPUTS INCREASE THE DECISION MAKING POWER OF A GATE. THEY ALSO INCREASE THE NUMBER OF WAYS GATES CAN BE CONNECTED TO ONE ANOTHER TO FORM ADVANCED DIGITAL LOGIC CIRCUITS. HERE ARE TWO EXAMPLES:

3-INPUT "AND" GATE

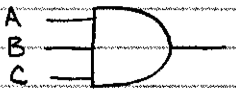

A	B	C	OUT
L	L	L	L
L	L	H	L
L	H	L	L
L	H	H	L
H	L	L	L
H	L	H	L
H	H	L	L
H	H	H	H

3-INPUT "NAND" GATE

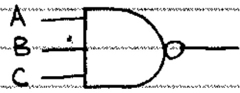

A	B	C	OUT
L	L	L	H
L	L	H	H
L	H	L	H
L	H	H	H
H	L	L	H
H	L	H	H
H	H	L	H
H	H	H	L

☐ SINGLE-INPUT GATES — THE "NOT" GATE OR INVERTER IS VERY IMPORTANT SINCE IT CAN *INVERT* (REVERSE) THE OUTPUT FROM ANOTHER GATE. STRICTLY SPEAKING, HOWEVER, THE INVERTER IS *NOT* A DECISION MAKING CIRCUIT (LIKE GATES WITH *TWO* OR MORE INPUTS). A CLOSE RELATIVE OF THE INVERTER IS THE *BUFFER*, A NON-INVERTING CIRCUIT THAT ISOLATES GATES FROM OTHER CIRCUITS OR ALLOWS THEM TO *DRIVE* HIGHER THAN NORMAL LOADS. THREE-STATE INVERTERS AND BUFFERS HAVE AN OUTPUT THAT CAN BE ELECTRONICALLY DISCONNECTED FROM THE REMAINDER OF THE CIRCUIT. THE OUTPUT IS THEN NEITHER HIGH NOR LOW. INSTEAD IT "FLOATS" AND APPEARS AS A *VERY* HIGH RESISTANCE.

☐ BUFFER

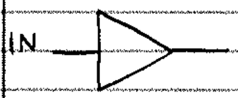

IN	OUT
L	L
H	H

☐ INVERTER ("NOT" GATE)

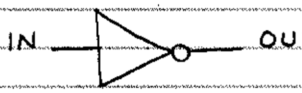

IN	OUT
L	H
H	L

☐ 3-STATE BUFFER

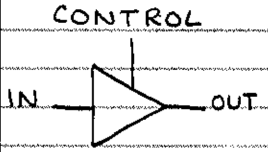

CONTROL	IN	OUT
L	L	L
L	H	H
H	X	HI-Z

☐ 3-STATE INVERTER

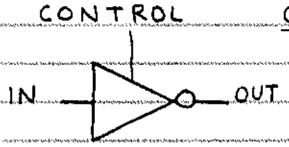

CONTROL	IN	OUT
L	L	H
L	H	L
H	X	HI-Z

"X" MEANS "DOESN'T MATTER." HI-Z MEANS HIGH OUTPUT RESISTANCE.

DATA "HIGHWAYS"

OFTEN CIRCUITS MADE FROM GATES EXCHANGE INFORMATION (BINARY 0'S AND 1'S ENCODED AS LOW AND HIGH VOLTAGE LEVELS). THE INFORMATION IS USUALLY SENT OVER WIRES CALLED *BUSES*. A BUS IS LIKE A *DATA HIGHWAY*. IT MAY BE ONE WIRE THROUGH WHICH INFORMATION IS SENT SERIALLY (BIT BY BIT). OR IT MAY BE UP TO EIGHT (OR MORE) WIRES THROUGH WHICH INFORMATION IS SENT IN PARALLEL (A BYTE OR MORE AT A TIME). IN BOTH CASES, OF COURSE, A GROUND IS REQUIRED TO COMPLETE THE CIRCUIT.

☐ 3-STATE TRAFFIC COPS — 3-STATE GATES CAN STOP "TRAFFIC JAMS" ON BUSES. FOR INSTANCE:

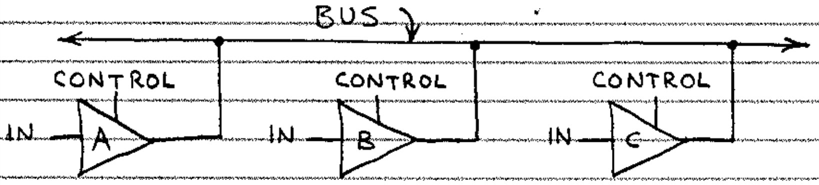

ONLY DATA ENTERING THE SELECTED BUFFER (CONTROL = L) GETS ON THE BUS.

HOW GATES ARE USED

GATES CAN BE USED INDIVIDUALLY OR CONNECTED TOGETHER TO FORM A "NETWORK" OF GATES CALLED A LOGIC CIRCUIT. ALMOST ALL LOGIC CIRCUITS CAN BE PLACED IN ONE OF TWO CATEGORIES: COMBINATIONAL OR SEQUENTIAL.

COMBINATIONAL LOGIC CIRCUITS

COMBINATIONAL LOGIC CIRCUITS RESPOND TO INCOMING DATA (0'S AND 1'S) ALMOST IMMEDIATELY AND WITHOUT REGARD TO EARLIER EVENTS. (THIS WILL MAKE MORE SENSE WHEN YOU READ ABOUT SEQUENTIAL CIRCUITS...) COMBINATIONAL LOGIC CIRCUITS CAN BE VERY SIMPLE OR IMMENSELY COMPLICATED. VIRTUALLY ANY COMBINATIONAL CIRCUIT CAN BE IMPLEMENTED WITH ONLY "NAND" OR "NOR" GATES. LIKE THESE "NAND" GATE CIRCUITS...

4-INPUT "NAND" GATE

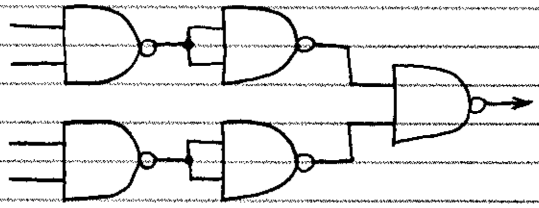

"AND" GATE

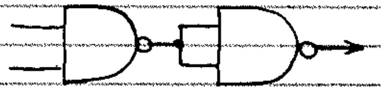

"OR" GATE

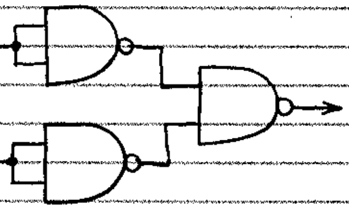

"NOR" GATE

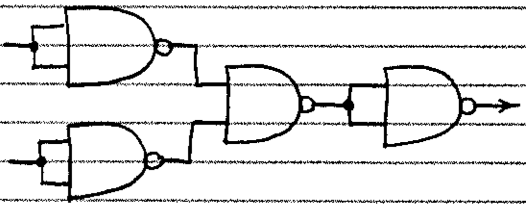

NOTE:
THESE CIRCUITS DO NOT SHOW THE GROUND CONNECTION THAT MUST BE PRESENT. USUALLY THE GROUND IS COMMON TO THE INPUT AND OUTPUT.

INVERTERS

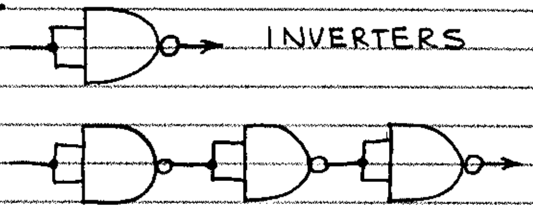

BUFFER

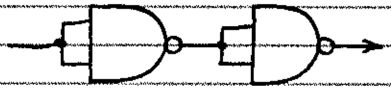

"EXCLUSIVE-OR" GATE

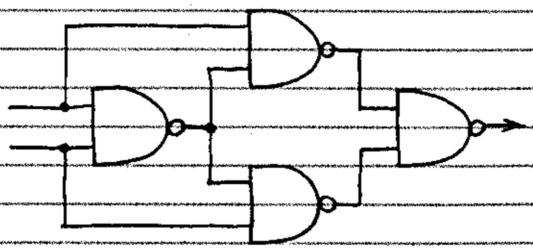

"EXCLUSIVE-NOR" GATE

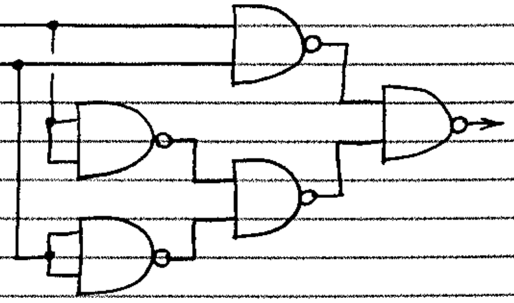

☐ COMBINING DIFFERENT GATES — HERE ARE TWO EXAMPLES OF COMBINATIONAL NETWORKS THAT USE MORE THAN ONE KIND OF GATE. (REMEMBER, BOTH THESE CIRCUITS CAN ALSO BE MADE ENTIRELY FROM "NAND" GATES!)

DATA SELECTOR

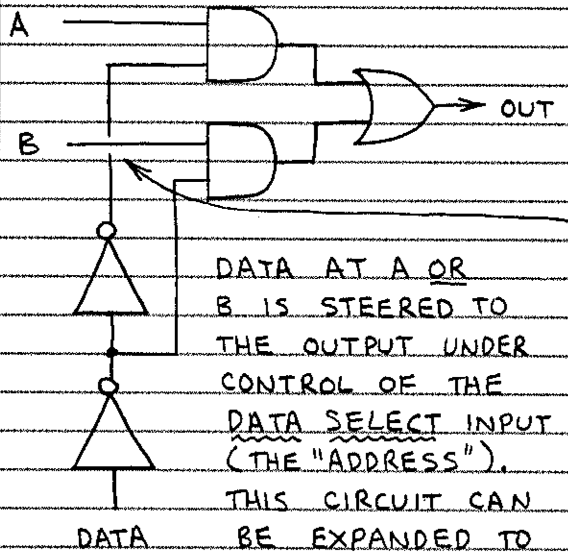
A
B
OUT
DATA SELECT:
L SELECTS A
H SELECTS B

DATA AT A OR B IS STEERED TO THE OUTPUT UNDER CONTROL OF THE DATA SELECT INPUT (THE "ADDRESS").

THIS CIRCUIT CAN BE EXPANDED TO INCLUDE MANY MORE INPUTS AND ADDRESSES.

BINARY TO DECIMAL DECODER

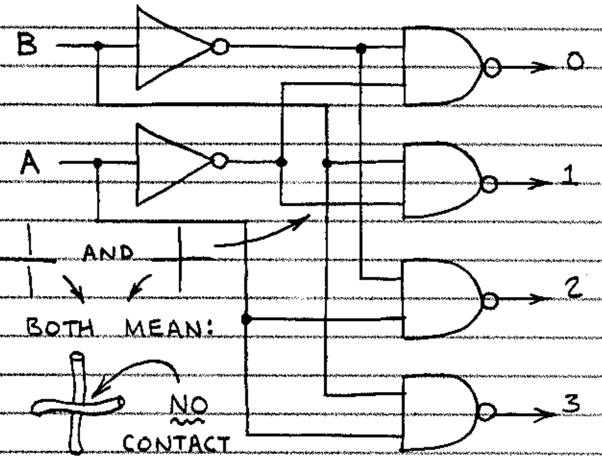
B
A
0
1
2
3
AND
BOTH MEAN:
NO CONTACT

THIS CONVERTS A TWO-BIT BINARY NUMBER TO ITS DECIMAL EQUIVALENT.

A	B	0	1	2	3
L	L	L	H	H	H
L	H	H	L	H	H
H	L	H	H	L	H
H	H	H	H	H	L

☐ ADVANCED COMBINATIONAL NETWORKS — HERE ARE SOME SIMPLE EXAMPLES FROM FOUR MAJOR FAMILIES OF COMBINATIONAL NETWORKS. THESE AND MANY OTHER NETWORK FAMILIES ARE AVAILABLE AS INTEGRATED CIRCUITS. BOXES LIKE THOSE SHOWN ARE LOGIC CIRCUIT SYMBOLS THAT REPRESENT COMPLICATED NETWORKS OF GATES.

MULTIPLEXER (DATA SELECTOR)

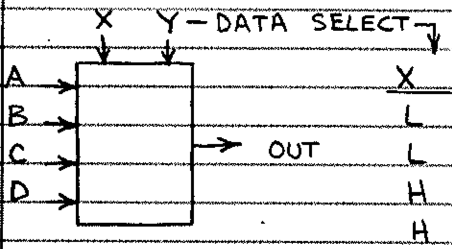
X
Y — DATA SELECT
A
B
C
D
OUT

X	Y	OUT
L	L	A
L	H	B
H	L	C
H	H	D

DEMULTIPLEXER

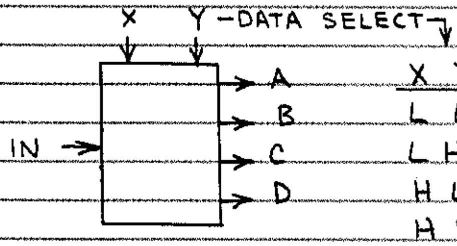
X
Y — DATA SELECT
IN
A
B
C
D

X	Y	IN TO...
L	L	A
L	H	B
H	L	C
H	H	D

ENCODER

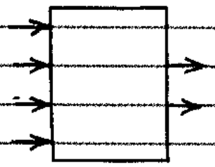

CONVERTS DECIMAL AND OTHER DATA TO BINARY. USES "OR" GATES.

DECODER

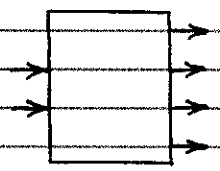

CONVERTS BINARY TO DECIMAL AND BCD TO DECIMAL DIGITS ON DIGITAL READOUTS.

SEQUENTIAL LOGIC CIRCUITS

THE OUTPUT STATE OF A SEQUENTIAL LOGIC CIRCUIT IS DETERMINED BY THE PREVIOUS STATE OF THE INPUT. IN OTHER WORDS, BITS OF DATA MOVE THROUGH SEQUENTIAL CIRCUITS STEP-BY-STEP. OFTEN THE DATA ADVANCES ONE STEP WHEN A PULSE IS RECEIVED FROM A "CLOCK" (A CIRCUIT THAT EMITS A STEADY STREAM OF PULSES). THE SEQUENTIAL LOGIC BUILDING BLOCK IS THE FLIP-FLOP. HERE'S A QUICK FLIP-FLOP REVIEW:

☐ THE BASIC "RS" (RESET-SET) FLIP-FLOP

ALSO CALLED A LATCH. THE OUTPUTS (Q AND $\bar{Q}$) ARE ALWAYS IN OPPOSITE STATES. ($\bar{Q}$ MEANS "NOT Q".)

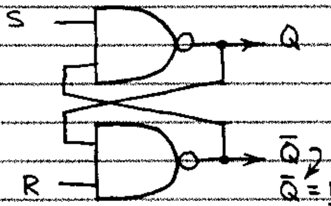

$\bar{Q}$ = NOT Q (IF Q = 0, $\bar{Q}$ = 1).

S	R	Q	$\bar{Q}$
L	L	(DISALLOWED)	
L	H	H	L
H	L	L	H
H	H	NO CHANGE	

☐ CLOCKED "RS" FLIP-FLOP

THIS LATCH IGNORES DATA AT S AND R UNTIL A "CLOCK" (OR ENABLE) PULSE ARRIVES. THEN IT CHANGES STATES.

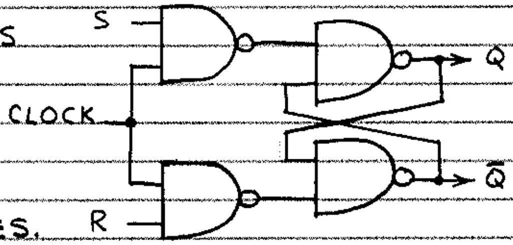

S	R	Q	$\bar{Q}$
L	L	NO CHANGE	
L	H	L	H
H	L	H	L
H	H	(DISALLOWED)	

VALID AFTER CLOCK PULSE ARRIVES.

☐ "D" (DATA OR DELAY) FLIP-FLOP

D FLIP-FLOPS STORE THE CURRENT OUTPUTS BETWEEN CLOCK PULSES.

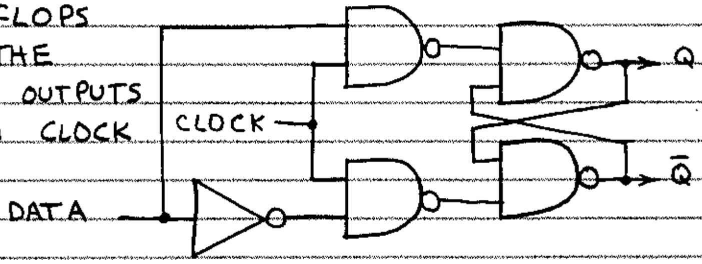

D	Q	$\bar{Q}$
L	L	H
H	H	L
	(OR)	
0	0	1
1	1	0

VALID AFTER CLOCK PULSE ARRIVES.

☐ "JK" FLIP-FLOP

THE "JK" FLIP-FLOP ALLOWS BOTH INPUTS TO BE H. (IN WHICH CASE ITS OUTPUTS "TOGGLE" OR SWITCH STATES AT EACH CLOCK PULSE.)

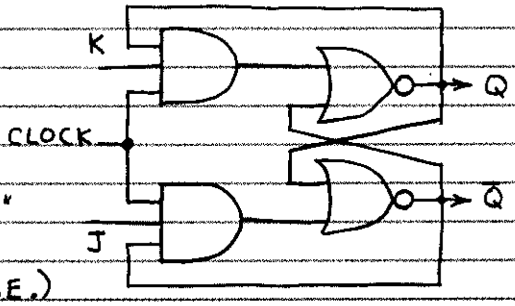

J	K	Q	$\bar{Q}$
L	L	NO CHANGE	
L	H	L	H
H	L	H	L
H	H	"TOGGLE"	

VALID AFTER CLOCK PULSE ARRIVES.

☐ "T" (TOGGLE) FLIP-FLOP

THE Q (OR $\bar{Q}$) OUTPUT IS L (OR H) FOR EVERY OTHER INPUT PULSE. THEREFORE THE INPUT PULSES ARE DIVIDED BY TWO. HERE ARE SEVERAL WAYS TO MAKE A "T" FLIP-FLOP:

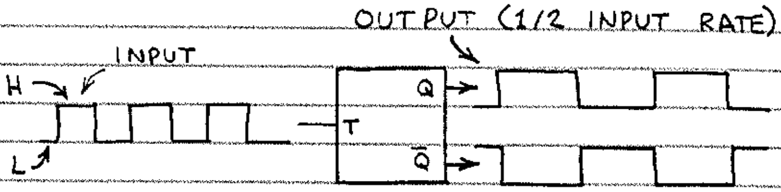
OUTPUT (1/2 INPUT RATE)
INPUT
H
L
T
Q
$\bar{Q}$

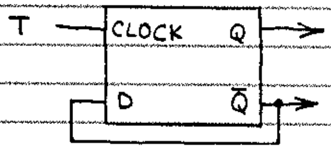
T
CLOCK
Q
D
$\bar{Q}$

"D" FLIP-FLOP

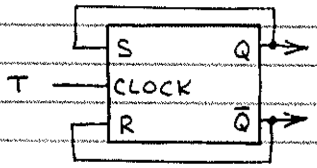
T
S
CLOCK
R
Q
$\bar{Q}$

CLOCKED "RS" FLIP-FLOP

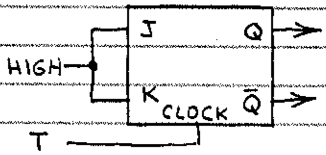
HIGH
J
K
CLOCK
Q
$\bar{Q}$
T

JK FLIP-FLOP

☐ "D" FLIP-FLOP DATA STORAGE REGISTER

HERE'S HOW FOUR "D" FLIP-FLOPS FORM A STORAGE REGISTER OR MEMORY THAT "LOADS" (SAVES) THE 4-BIT BINARY NIBBLE AT INPUTS A–D WHEN THE "CLOCK" INPUT IS "STROBED" (PULSED). MANY TYPES OF IC REGISTERS ARE AVAILABLE.

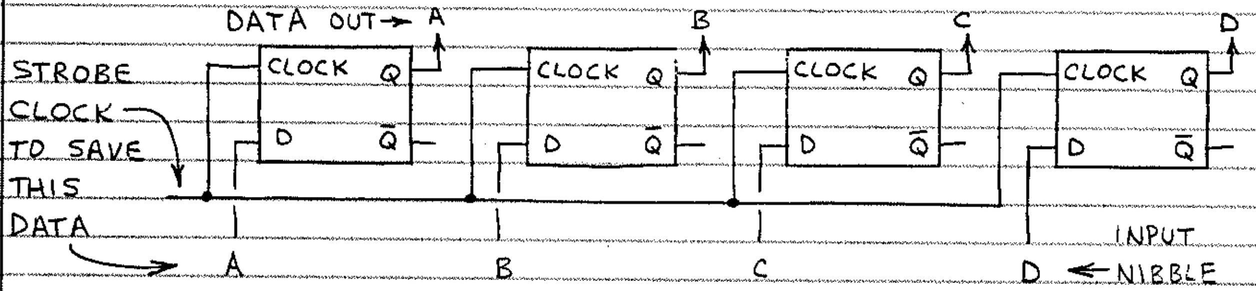
DATA OUT → A
B
C
D
STROBE CLOCK TO SAVE THIS DATA
CLOCK
Q
D
$\bar{Q}$
A
B
C
D
INPUT NIBBLE

☐ "T" FLIP-FLOP COUNTER

HERE'S HOW FOUR "T" FLIP-FLOPS FORM A 4-BIT BINARY COUNTER:

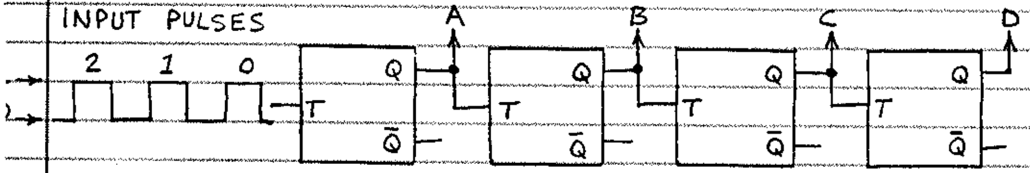
INPUT PULSES
2
1
0
T
Q
$\bar{Q}$
A
B
C
D

COUNT	D C B A
0	0 0 0 0
1	0 0 0 1
2	0 0 1 0
3	0 0 1 1
4	0 1 0 0
5	0 1 0 1
6	0 1 1 0
7	0 1 1 1
8	1 0 0 0
9	1 0 0 1
10	1 0 1 0
11	1 0 1 1
12	1 1 0 0
13	1 1 0 1
14	1 1 1 0
15	1 1 1 1

EACH "T" FLIP-FLOP DIVIDES INCOMING PULSES BY TWO. AS THE BINARY TRUTH TABLE REVEALS, THE RESULT IS A 0000 → 1111 BINARY COUNT. (THE COUNT RECYCLES TO 0000 AFTER THE 16TH INCOMING PULSE.) THERE ARE MANY TYPES OF IC COUNTERS, MOST OF WHICH INCLUDE SPECIAL FEATURES (COUNT UP OR DOWN, RESET, ETC.).

A COMBINATIONAL – SEQUENTIAL LOGIC SYSTEM

HERE'S HOW COMBINATIONAL AND SEQUENTIAL LOGIC IC's CAN FORM A DECIMAL COUNTING CIRCUIT, A VERY SIMPLE DIGITAL LOGIC SYSTEM.

1. THE BLOCK DIAGRAM

P. 117

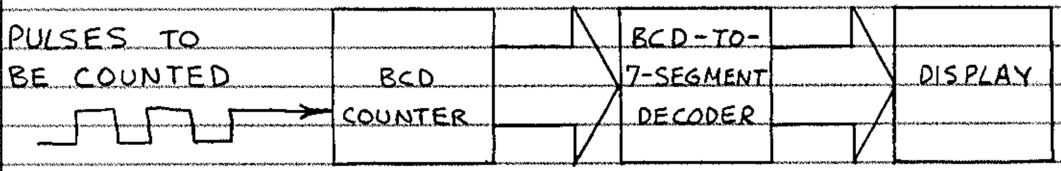
PULSES TO
BE COUNTED
BCD
COUNTER
BCD-TO-
7-SEGMENT
DECODER
DISPLAY

THE BCD COUNTER ADVANCES ONE COUNT FOR EACH INCOMING PULSE. WHEN THE COUNT REACHES 1001 (DECIMAL 9), THE COUNTER RECYCLES TO 0000. THE DECODER ACTIVATES THE APPROPRIATE SEGMENTS OF AN LED DISPLAY.

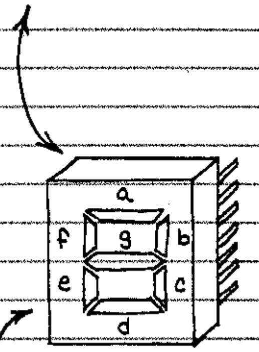

2. THE CIRCUIT DIAGRAM

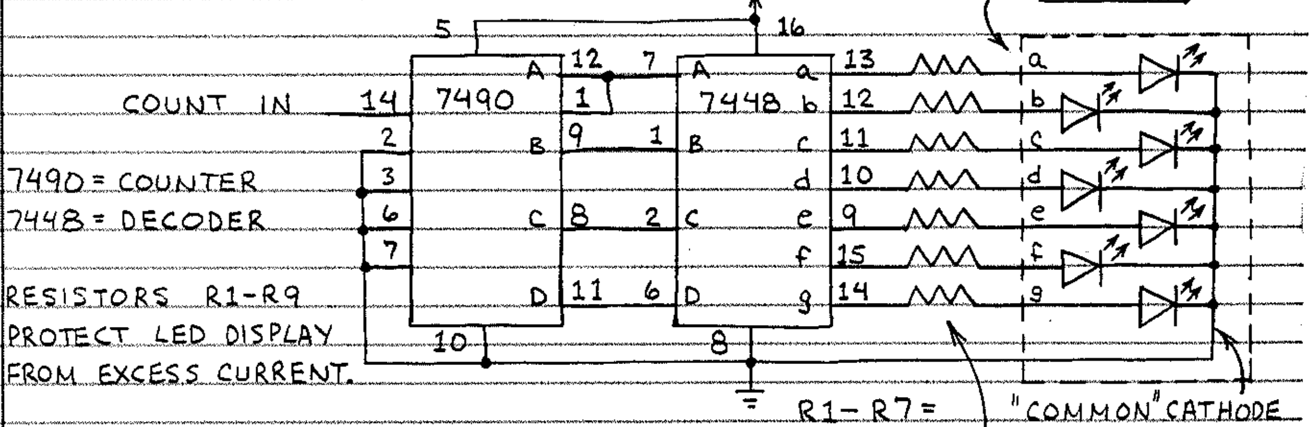
+5 V
5
16
A
12
7
A
a
13
a
COUNT IN
14
7490
1
7448
b
12
b
2
B
9
1
B
c
11
c
7490 = COUNTER
3
d
10
d
7448 = DECODER
6
C
8
2
C
e
9
e
7
f
15
f
RESISTORS R1-R9
PROTECT LED DISPLAY
FROM EXCESS CURRENT.
D
11
6
D
g
14
g
10
8

R1 – R7 = 270-OHMS

"COMMON" CATHODE

3. THE ACTUAL CIRCUIT

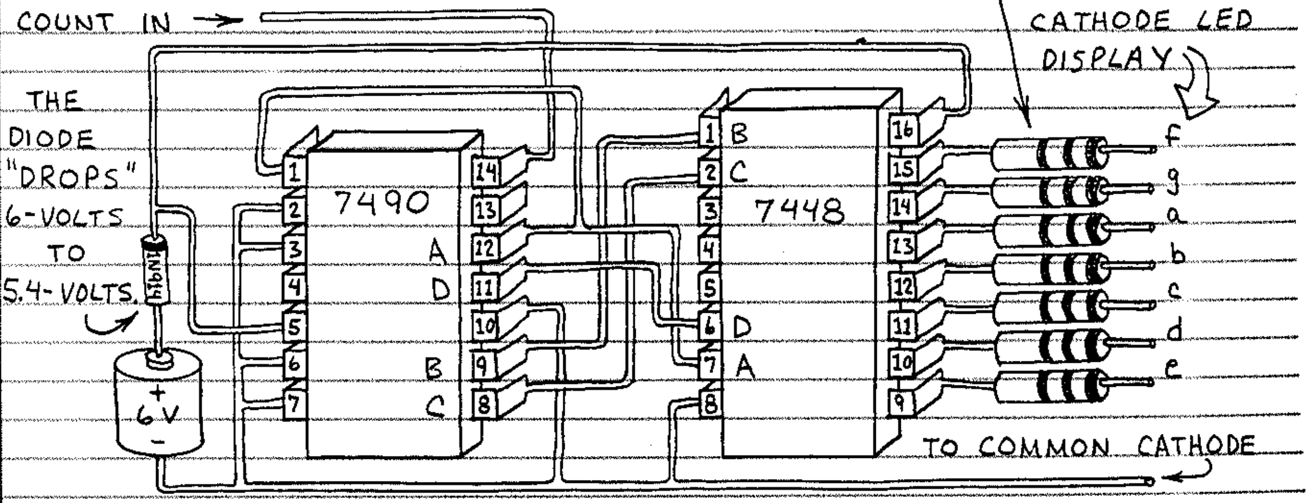
COUNT IN
1
2
3
4
5
6
7
7490
14
13
12
11
10
9
8
A
D
B
C
1N914
+
6V
–
1 B
2 C
3
4
5
6 D
7 A
8
7448
16
15
14
13
12
11
10
9
f
g
a
b
c
d
e

THE DIODE "DROPS" 6-VOLTS TO 5.4-VOLTS.

TO COMMON CATHODE LED DISPLAY

TO COMMON CATHODE

DIGITAL IC FAMILIES

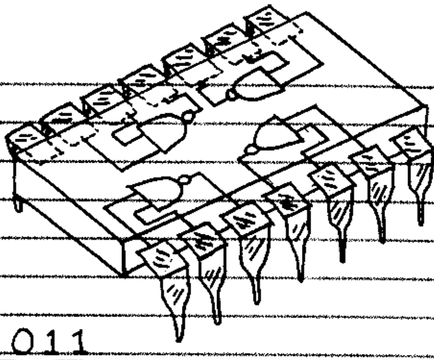

4011

A "QUAD" (FOUR) OF 2-INPUT CMOS "NAND" GATES

THERE ARE MORE THAN A DOZEN MAJOR FAMILIES OF BIPOLAR AND MOS INTEGRATED CIRCUITS. EACH IC (OR "CHIP") CONTAINS A SPECIFIC LOGIC NETWORK OR ASSORTMENT OF VARIOUS LOGIC FUNCTIONS. HERE ARE SOME OF THE MAJOR DIGITAL IC FAMILIES:

☐ BIPOLAR DIGITAL IC'S

1. TRANSISTOR-TRANSISTOR LOGIC (TTL OR T^2L). THE LARGEST AND FORMERLY MOST POPULAR DIGITAL IC FAMILY. CAN CHANGE STATES MORE THAN 20,000,000 TIMES PER SECOND. VERY INEXPENSIVE. DRAWBACKS: MUST BE POWERED BY 5-VOLT SUPPLY. USES LOTS OF POWER. (INDIVIDUAL GATES REQUIRE 3 OR 4 MILLIAMPERES.) MOST WIDELY USED IS THE 7400 SERIES. THE 7404, FOR EXAMPLE, CONTAINS FOUR INVERTERS.

2. LOW-POWER SCHOTTKY TTL (LS). A NEWER KIND OF TTL THAT CONSUMES ONLY 20% AS MUCH POWER. DRAWBACK: MORE EXPENSIVE THAN STANDARD TTL. MOST WIDELY USED IS THE 74LS00 SERIES.

☐ MOSFET DIGITAL IC'S

1. P- AND N-CHANNEL MOS (PMOS AND NMOS). CONTAIN MORE GATES PER CHIP THAN TTL. MANY SPECIAL PURPOSE CHIPS (MICROPROCESSORS, MEMORIES, ETC.). DRAWBACKS: FEW COUNTERPARTS TO POPULAR TTL CHIPS. SLOWER THAN TTL. MAY REQUIRE TWO OR MORE SUPPLY VOLTAGES. MAY BE DAMAGED BY STATIC ELECTRICAL DISCHARGE.

2. COMPLEMENTARY MOS (CMOS). FASTEST GROWING AND MOST VERSATILE DIGITAL IC FAMILY. THERE ARE CMOS VERSIONS OF MOST POPULAR TTL CHIPS. ONE SERIES USES THE SAME DESIGNATION NUMBERS. THE 74C04, FOR EXAMPLE, IS THE CMOS EQUIVALENT OF THE TTL 7404. NEW HIGH-SPEED CMOS JUST AS FAST AS TTL. MOST CMOS HAS A WIDE SUPPLY VOLTAGE RANGE (TYPICALLY +3 TO +18 VOLTS). USES LESS POWER THAN ANY OTHER DIGITAL IC FAMILY. (INDIVIDUAL GATES REQUIRE 0.1 MILLIAMPERE.) DRAWBACK: MAY BE DAMAGED BY STATIC ELECTRICAL DISCHARGE. MOST WIDELY USED ARE 74C00 AND 4000 SERIES.

7. LINEAR INTEGRATED CIRCUITS

THE INPUT AND OUTPUT VOLTAGE LEVELS OF LINEAR INTEGRATED CIRCUITS CAN VARY OVER A WIDE RANGE. OFTEN THE OUTPUT VOLTAGE IS PROPORTIONAL TO THE INPUT VOLTAGE. THEREFORE, A GRAPH OF INPUT VERSUS OUTPUT IS A STRAIGHT (LINEAR) LINE. THERE ARE MANY TYPES OF LINEAR IC'S. ONLY THE MAJOR TYPES ARE COVERED HERE. FIRST LET'S COMPARE THE BASIC DIGITAL AND LINEAR CIRCUITS:

THE BASIC LINEAR CIRCUIT

A SINGLE BIPOLAR OR FIELD-EFFECT TRANSISTOR CAN FUNCTION AS A DIGITAL OR LINEAR CIRCUIT. IN BOTH CASES, THE TRANSISTOR CAN INVERT THE SIGNAL AT ITS INPUT. HERE'S HOW AN NPN BIPOLAR TRANSISTOR CAN PERFORM ALL FOUR FUNCTIONS:

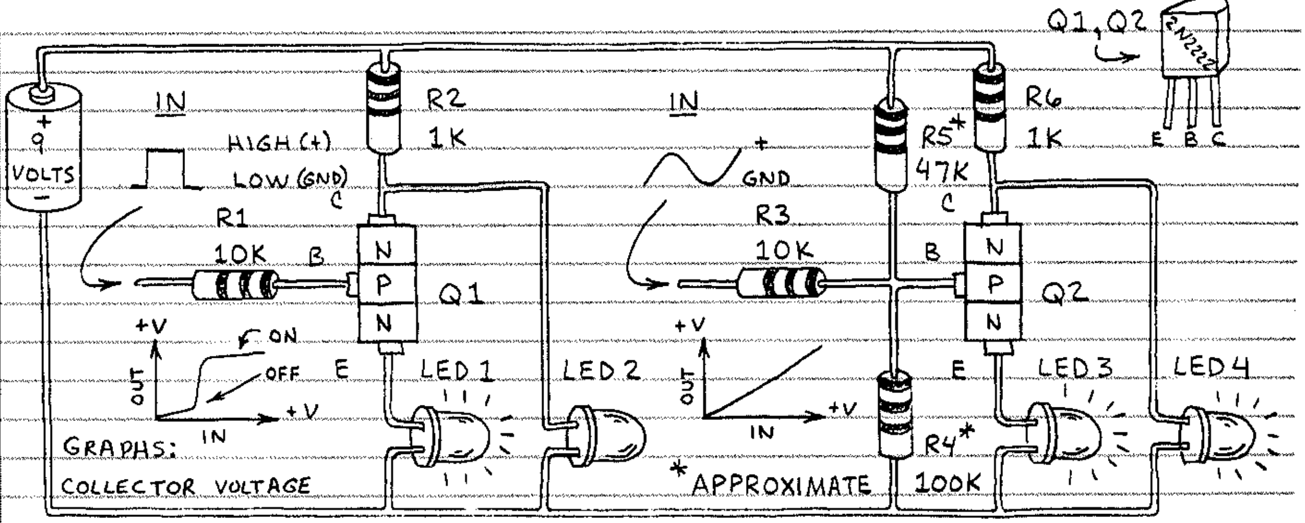
Q1, Q2
2N2222
E B C
+
9 VOLTS
−
IN
HIGH (+)
LOW (GND)
R1
10K
B
C
R2
1K
N
P
N
Q1
E
LED 1
LED 2
+V
ON
OFF
OUT
IN
+V
GRAPHS:
COLLECTOR VOLTAGE
IN
+
GND
R3
10K
R5*
47K
B
C
R6
1K
N
P
N
Q2
E
LED 3
LED 4
+V
OUT
IN
+V
R4*
100K
*APPROXIMATE

DIGITAL:

HERE TRANSISTOR Q1 IS USED AS A SWITCH. WHEN THE INPUT IS NEAR +V (OR HIGH), Q1 TURNS ON AND LED 1 IS ILLUMINATED. WHEN THE INPUT IS NEAR GROUND (OR LOW), Q1 TURNS OFF. THIS TURNS LED 1 OFF AND ALLOWS LED 2 TO GLOW. (R2 CONTROLS THE CURRENT THROUGH BOTH LEDs.) THIS CIRCUIT IS THEN A COMBINED DIGITAL BUFFER AND INVERTER.

LINEAR:

HERE Q2 IS AN AMPLIFIER THAT OPERATES OVER THE ENTIRE RANGE FROM FULL OFF TO FULL ON. R4 AND R5 FORM A VOLTAGE DIVIDER THAT APPLIES A SMALL VOLTAGE TO Q2'S BASE TO KEEP Q2 SLIGHTLY ON EVEN WHEN NO INPUT IS PRESENT. THIS ALLOWS Q2 TO OPERATE IN A LINEAR MODE. AS THE INPUT VOLTAGE RISES, LED 3 BRIGHTENS AND LED 4 DIMS.

OPERATIONAL AMPLIFIERS

OPERATIONAL AMPLIFIERS (OR "OP-AMPS") ARE BY FAR THE MOST VERSATILE OF LINEAR IC'S. THEY'RE CALLED "OPERATIONAL" AMPLIFIERS SINCE THEY WERE ORIGINALLY DESIGNED TO DO MATHEMATICAL OPERATIONS. OP-AMPS AMPLIFY THE DIFFERENCE BETWEEN VOLTAGES OR SIGNALS (AC OR DC) APPLIED TO THEIR TWO INPUTS. THE VOLTAGE APPLIED TO ONLY ONE INPUT WILL BE AMPLIFIED IF THE SECOND INPUT IS GROUNDED OR MAINTAINED AT SOME VOLTAGE LEVEL.

☐ OP-AMP OPERATION — THE OP-AMP HAS AN INVERTING AND NON-INVERTING INPUT. THE POLARITY OF A VOLTAGE APPLIED TO THE INVERTING INPUT IS REVERSED AT THE OUTPUT. (INVERTING INPUT IS − ; NON-INVERTING INPUT IS +.)

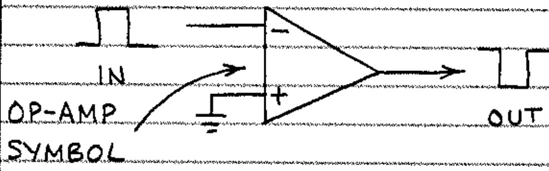

INVERTING MODE

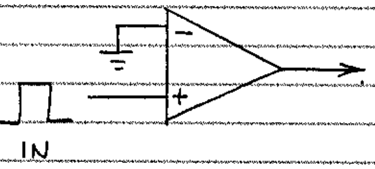

NON-INVERTING MODE

P.122

☐ OP-AMP "FEEDBACK" — THE CIRCUITS SHOWN ABOVE ALLOW THE OP-AMP TO OPERATE AT ITS MAXIMUM AMPLIFICATION LEVEL (OR GAIN). USUALLY THE GAIN IS REDUCED TO A MORE PRACTICAL LEVEL BY FEEDING SOME OF THE OUTPUT BACK TO THE INVERTING (−) INPUT. FOR EXAMPLE:

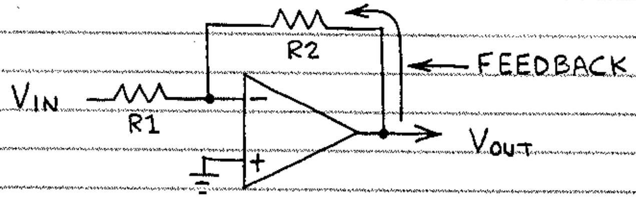

INVERTING AMPLIFIER

GAIN = R2/R1

$V_{OUT} = -V_{IN}(R2/R1)$

P.122

☐ OP-AMP COMPARATOR — WHEN OPERATED WITHOUT A FEEDBACK RESISTOR (R2 ABOVE), THE OUTPUT VOLTAGE WILL SWING FROM FULL ON TO FULL OFF (OR VICE VERSA) WHEN THE VOLTAGES APPLIED TO THE INPUTS DIFFER BY ONLY ABOUT 0.001 VOLT! THIS DIGITAL-LIKE MODE MAKES POSSIBLE MANY USEFUL APPLICATIONS.

P.124

☐ TYPES OF OP-AMPS — BOTH BIPOLAR AND MOSFET IC OP-AMPS ARE AVAILABLE. SOME BIPOLAR OP-AMPS HAVE FET OR MOSFET INPUTS TO PROVIDE VERY HIGH INPUT RESISTANCE. MANY DIFFERENT OP-AMPS ARE MADE. A SINGLE IC MAY INCLUDE UP TO FOUR INDIVIDUAL OP-AMPS.

TIMERS

WHEN OPERATED AS A COMPARATOR, THE OP-AMP CAN BE USED AS A TIMER. ALL THAT'S REQUIRED IS AN RC (RESISTOR-CAPACITOR) CIRCUIT LIKE THIS:

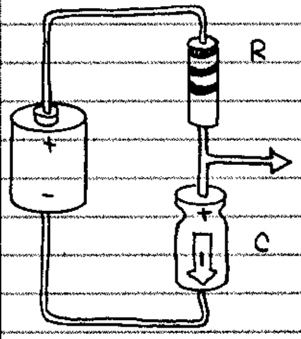

RC CIRCUIT

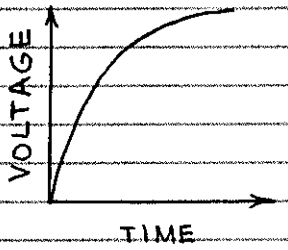

RC CHARGING GRAPH

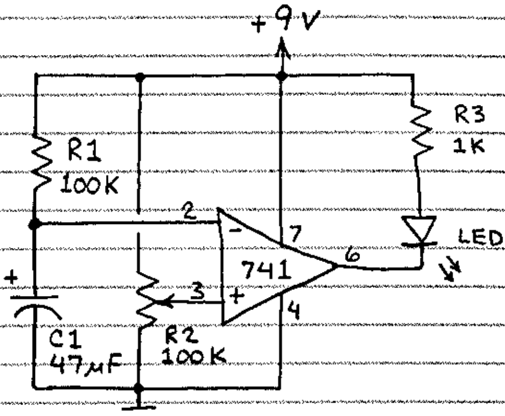

TIMER

IN THE CIRCUIT DIAGRAM (ABOVE RIGHT), R1 AND C1 FORM AN RC CIRCUIT. C1 GRADUALLY CHARGES TO +9 VOLTS THROUGH R1. WHEN THE VOLTAGE ON C1 EXCEEDS THE REFERENCE VOLTAGE SUPPLIED TO THE NON-INVERTING INPUT OF THE OP-AMP, ITS OUTPUT SWINGS FROM HIGH TO LOW AND THE LED GLOWS. THE TIME DELAY CAN BE CHANGED BY ALTERING THE VALUES OF R1 AND C1 OR THE SETTING OF R2. DISCHARGE C1 (USE PUSHBUTTON SWITCH) FOR NEW CYCLE.

☐ IC TIMERS — THE SIMPLE CIRCUIT ABOVE IS THE KEY INGREDIENT OF MOST IC TIMERS. MOST INCLUDE AN OUTPUT FLIP-FLOP TO GIVE DEFINITE HIGH OR LOW OUTPUT. SOME INCLUDE A BINARY COUNTER THAT ADVANCES ONE COUNT PER DELAY PERIOD (OR CYCLE). THE TIMER IS RECYCLED EACH TIME THE COUNT ADVANCES. A DECODER AT THE COUNTER OUTPUT ALLOWS TOTAL DELAYS OF FROM DAYS TO A YEAR OR MORE TO BE SELECTED. BOTH BIPOLAR AND CMOS TIMERS ARE AVAILABLE.

P.126

FAMOUS FACT: ANALOG COMPUTERS USE OP-AMPS TO SOLVE COMPLEX EQUATIONS!

FUNCTION GENERATORS

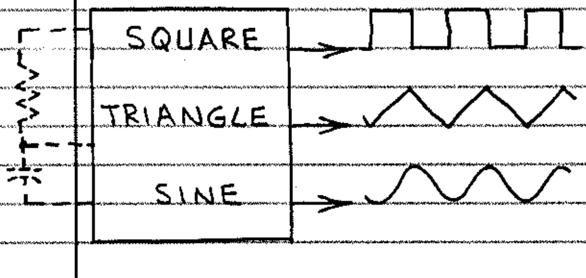

THESE IC'S GENERATE VARIOUS KINDS OF OUTPUT WAVES SUCH AS THOSE SHOWN HERE. THE FREQUENCY OF THE WAVES CAN BE CONTROLLED BY AN EXTERNAL RC CIRCUIT.

VOLTAGE REGULATORS

VOLTAGE REGULATORS CONVERT A VOLTAGE APPLIED TO THEIR INPUT INTO A FIXED OR VARIABLE (BUT USUALLY LOWER) VOLTAGE. IN MOST A SMALL, FIXED REFERENCE VOLTAGE (USUALLY A VOLT OR SO) IS APPLIED TO THE NON-INVERTING INPUT OF AN OP-AMP. THE RERERENCE VOLTAGE (OR V_{REF}) IS THEN AMPLIFIED BY THE RATIO OF THE FEEDBACK AND INPUT RESISTORS (THE GAIN). IF ONE OF THE RESISTORS IS A POTENTIOMETER, THE OUTPUT VOLTAGE (V_{OUT}) CAN BE VARIED FROM V_{REF} TO +V (THE CHIP SUPPLY VOLTAGE). ACTUAL IC REGULATORS INCLUDE EXTRA TRANSISTORS TO PROVIDE V_{REF} AND TO ALLOW THE CHIP TO DRIVE LOADS THAT REQUIRE MORE POWER THAN AN OP-AMP ALONE CAN DELIVER.

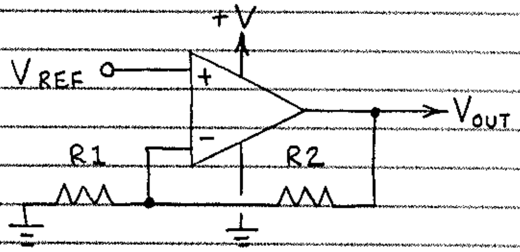

BASIC VOLTAGE REGULATOR

☐ IC REGULATORS — MANY TYPES OF FIXED AND VARIABLE OUTPUT IC REGULATORS ARE AVAILABLE. MOST ARE INSTALLED IN PACKAGES MADE OF METAL OR HAVING METAL TABS TO HELP RADIATE EXCESSIVE HEAT INTO THE SURROUNDING AIR. CAUTION: MANUFACTURER'S OPERATING INSTRUCTIONS AND STANDARD SAFETY PRECAUTIONS MUST BE FOLLOWED FOR BEST RESULTS.

P.125

OTHER LINEAR IC's

THERE ARE NUMEROUS SPECIAL FUNCTION LINEAR IC's, MANY OF WHICH INCORPORATE OP-AMPS. FOR EXAMPLE:

☐ AUDIO AMPLIFIERS — MANY KINDS AVAILABLE. SOME INCLUDE TWO AMPLIFIERS ON ONE CHIP (FOR STEREO).

☐ PHASE-LOCKED LOOPS — BASED ON AN OLD BUT CLEVER IDEA IN WHICH AN ON-CHIP OSCILLATOR DUPLICATES (OR TRACKS) THE FREQUENCY OF AN INCOMING SIGNAL. USED TO DETECT THE PRESENCE OF CERTAIN FREQUENCIES (LIKE TOUCH-TONE® TONES) AND TO DEMODULATE FM RADIO SIGNALS.

☐ OTHER LINEAR IC's — INCLUDED ARE MANY KINDS OF CHIPS FOR TELEPHONE, RADIO, TELEVISION AND COMPUTER COMMUNICATIONS. ALSO, MANY KINDS OF IC's THAT DETECT TEMPERATURE, LIGHT AND PRESSURE.

8. CIRCUIT ASSEMBLY TIPS

THERE ARE SEVERAL WAYS TO MAKE EITHER TEMPORARY OR PERMANENT VERSIONS OF ELECTRONIC CIRCUITS. IN THIS CHAPTER WE'LL LOOK AT SOME CIRCUIT ASSEMBLY TIPS YOU MAY FIND HELPFUL.

TEMPORARY CIRCUITS

IT'S ALWAYS WISE TO BUILD A TEMPORARY VERSION OF A CIRCUIT BEFORE ASSEMBLING IT IN PERMANENT FORM. YOU CAN THEN MAKE CHANGES AND FIND OUT HOW WELL THE CIRCUIT WORKS. BY FAR THE MOST IMPORTANT TOOL FOR TEMPORARY CIRCUIT ASSEMBLY IS THE PLASTIC SOLDERLESS MODULAR BREADBOARD SOCKET. IT'S A GOOD IDEA TO KEEP SEVERAL ON YOUR WORKBENCH. THEY WILL LET YOU BUILD ENTIRE CIRCUITS IN MINUTES. USE "JUMPER" WIRES TO INTERCONNECT PARTS WHOSE LEADS ARE NOT INSERTED IN THE SAME ROW OF TERMINALS. TO AVOID BENDING THEIR PINS (AND PRICKING YOUR FINGERS), INSTALL AND REMOVE IC'S CAREFULLY.

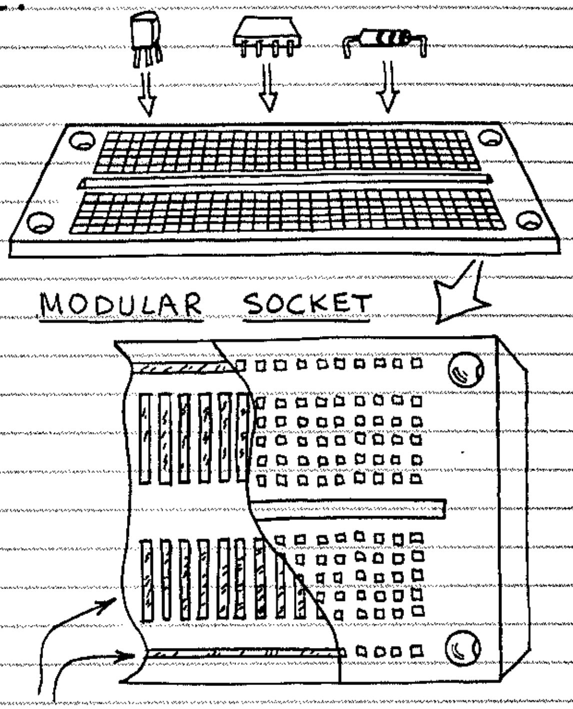

CUTAWAY SHOWING COMMON TERMINAL CONNECTIONS.

HINT: INSTALL SOCKET ON BASE AND ADD POTENTIOMETERS, BATTERY, LEDs, SWITCHES, ETC.

PERMANENT CIRCUITS

WITH THE EXCEPTION OF SOME VERY SIMPLE CIRCUITS, MOST PERMANENT CIRCUITS ARE ASSEMBLED ON SOME FORM OF CIRCUIT BOARD.

☐ PERFORATED BOARD CONSTRUCTION — COMPONENT LEADS ARE INSERTED THROUGH PERFORATIONS IN A PHENOLIC OR SIMILAR BOARD AND SOLDERED TOGETHER ON THE BACK SIDE OF THE BOARD. OFTEN INSULATED CONNECTION WIRES MUST BE USED. ONCE ASSEMBLED, "PERFBOARD" CIRCUITS ARE DIFFICULT TO REPAIR SINCE COMPONENT LEADS ARE OFTEN TWISTED AND SOLDERED.

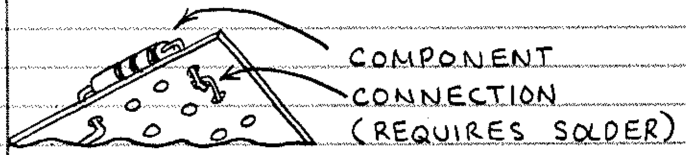

☐ WIRE-WRAP — FASTEST WAY TO ASSEMBLE CIRCUITS THAT USE MORE THAN A FEW IC'S. USE WIRE-WRAP IC SOCKETS (WITH SQUARE CONNECTION PINS). BOTH HAND AND MOTOR POWERED WRAPPING TOOLS ARE AVAILABLE. IF YOU USE THE KIND THAT REQUIRES SOME OF THE WIRE'S INSULATION BE REMOVED, WRAP A FEW TURNS OF INSULATED WIRE AROUND THE CONNECTION PIN TO STRENGTHEN THE CONNECTION.

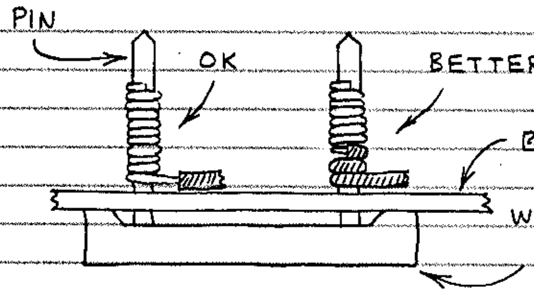
PIN
OK
BETTER
BOARD
WIRE-WRAP
SOCKET

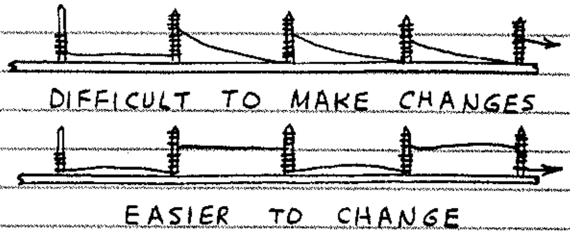
DIFFICULT TO MAKE CHANGES
EASIER TO CHANGE

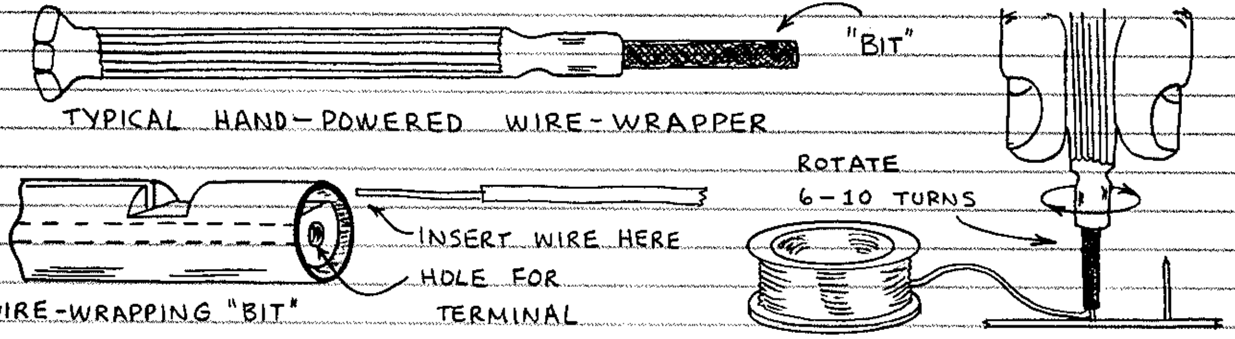
"BIT"
TYPICAL HAND-POWERED WIRE-WRAPPER
ROTATE
6-10 TURNS
INSERT WIRE HERE
HOLE FOR
TERMINAL
WIRE-WRAPPING "BIT"

☐ PRINTED CIRCUIT (PC) — PROVIDES THE NEATEST AND MOST PROFESSIONAL APPEARING COMPLETED CIRCUIT. SOCKETS NOT REQUIRED, BUT COMPONENT LEADS MUST BE SOLDERED TO THE COPPER PATTERNS ON THE BOARD. THERE ARE MANY TYPES OF PC BOARDS. TWO TYPES USED BY EXPERIMENTERS ARE:

1. PRE-ETCHED PERFORATED GRID BOARDS HAVE A ROUND, COPPER FOIL SOLDER PAD AT EACH HOLE. ON MANY BOARDS ROWS OF HOLES ARE CONNECTED BY COMMON COPPER FOIL STRIPS (LIKE A SOLDERLESS BREADBOARD). IT'S USUALLY NECESSARY TO JOIN SOME OF THE CONTACTS ON THE BOARD WITH "JUMPERS" (SHORT LENGTHS OF INSULATED HOOKUP OR WRAPPING WIRE).

2. CUSTOM PC BOARDS ARE MADE BY APPLYING A TAPE OR CHEMICAL COATING (THE "RESIST") TO THE CLEAN COPPER FOIL OF A PC BOARD. THE UNCOATED COPPER IS THEN CHEMICALLY ETCHED AWAY, LEAVING BEHIND A FOIL WIRING PATTERN. HOLES ARE DRILLED FOR COMPONENT LEADS. TAKES LOTS OF TIME, BUT PRODUCES NEAT CIRCUITS.

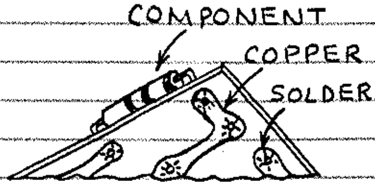
COMPONENT
COPPER
SOLDER

HOW TO SOLDER

GOOD SOLDERING PRACTICES ARE ESSENTIAL FOR RELIABLE OPERATION OF A CIRCUIT WITH SOLDERED CONNECTIONS. HERE ARE SIX STEPS FOR SUCCESSFUL SOLDERING:

1. ALWAYS USE A LOW-WATTAGE SOLDERING IRON (25 TO 40 WATTS). BE SURE TO TIN THE TIP ACCORDING TO THE MANUFACTURER'S INSTRUCTIONS.

2. ALWAYS USE ROSIN CORE SOLDER WHEN SOLDERING ELECTRONIC COMPONENTS. NEVER USE ACID CORE SOLDER SINCE IT WILL CORRODE THE SOLDERED LEAD.

3. SOLDER DOES NOT ADHERE TO PAINT, GREASE, OIL, WAX OR MELTED INSULATION. REMOVE ALL SUCH FOREIGN MATTER WITH A SOLVENT, STEEL WOOL OR FINE SANDPAPER. ALWAYS BUFF THE COPPER FOIL OF A PC BOARD WITH STEEL WOOL BEFORE SOLDERING. (THE COPPER SHOULD BE SHINY.)

4. TO SOLDER, FIRST HEAT THE CONNECTION (NOT THE SOLDER!) FOR A FEW SECONDS WITH THE HOT TIP OF THE IRON. THEN LEAVE THE IRON IN PLACE AND APPLY SOLDER.

5. ALLOW THE SOLDER TO FLOW THROUGH AND AROUND THE CONNECTION BEFORE REMOVING THE IRON. DON'T APPLY TOO MUCH SOLDER OR MOVE THE CONNECTION BEFORE IT COOLS.

6. KEEP THE IRON'S TIP CLEAN AND SHINY. WIPE OFF DEBRIS WITH A DAMP SPONGE OR CLOTH.

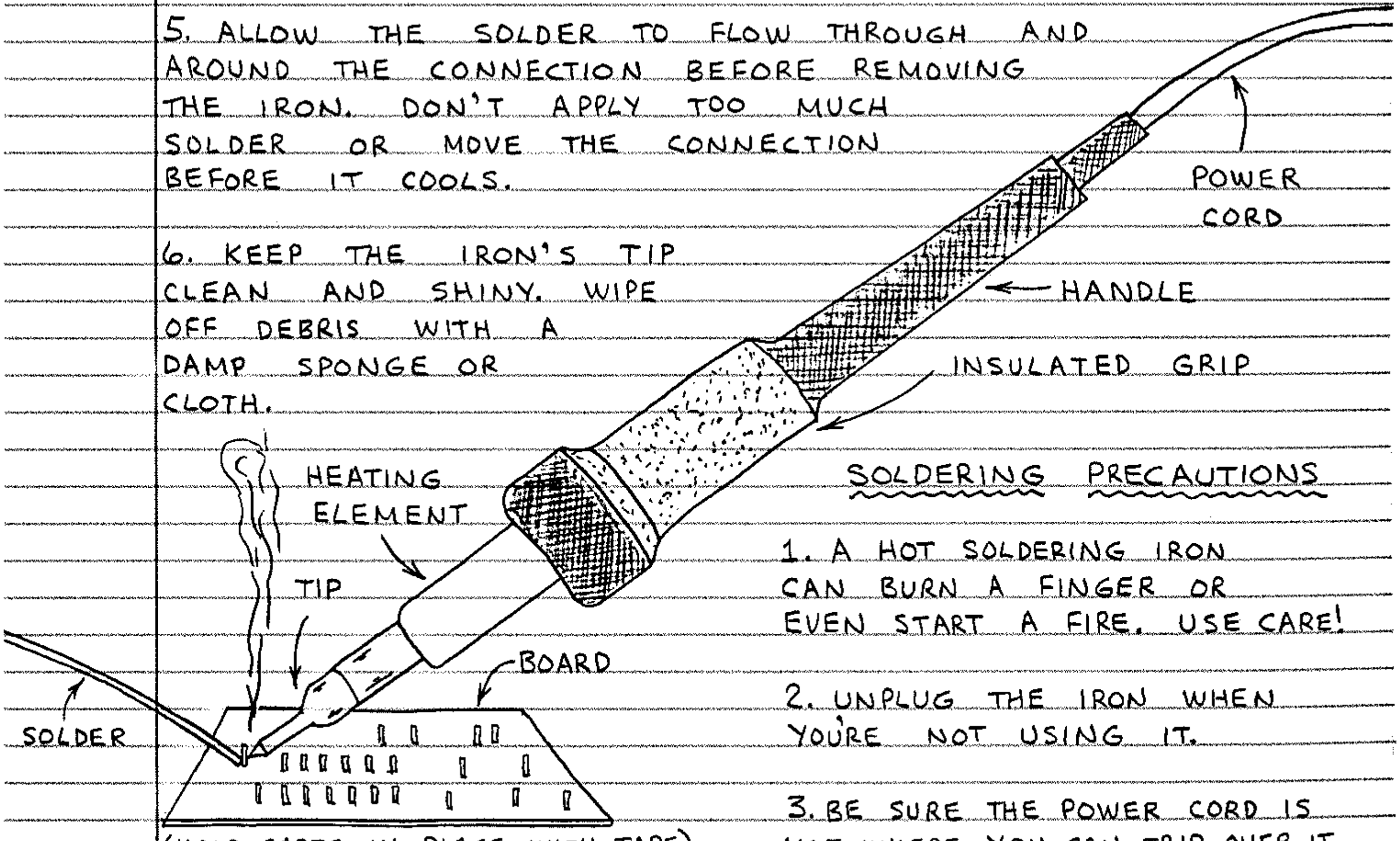

(HOLD PARTS IN PLACE WITH TAPE)

SOLDERING PRECAUTIONS

1. A HOT SOLDERING IRON CAN BURN A FINGER OR EVEN START A FIRE. USE CARE!

2. UNPLUG THE IRON WHEN YOU'RE NOT USING IT.

3. BE SURE THE POWER CORD IS NOT WHERE YOU CAN TRIP OVER IT.

POWERING ELECTRONIC CIRCUITS

☐ BATTERY POWER — MANY CIRCUITS USE SO LITTLE POWER THEY CAN BE POWERED BY BATTERIES. THIS KEEPS THE COMPLETED CIRCUIT COMPACT AND ALLOWS IT TO BE OPERATED ANYWHERE.

☐ SOLAR POWER — SOLAR CELLS CAN POWER YOUR CIRCUITS DIRECTLY. OR YOU CAN USE AN ARRAY OF SOLAR CELLS TO CHARGE A RECHARGEABLE BATTERY.

☐ LINE POWER - THE SIMPLEST LINE POWERED SUPPLY IS THE SO-CALLED AC ADAPTER. THESE MODULAR UNITS ARE COMPACT AND EASY TO USE. UNITS HAVING VARIOUS OUTPUT VOLTAGES AND CURRENTS ARE AVAILABLE. YOU CAN MAKE YOUR OWN LINE POWERED SUPPLY USING AN IC VOLTAGE REGULATOR.

☐ CAUTION — SAFETY SHOULD BE YOUR FIRST CONCERN WHEN BUILDING YOUR OWN LINE POWERED SUPPLY. THE POWER CORD MUST BE PROTECTED FROM THE SHARP EDGES OF A HOLE DRILLED IN A METAL CABINET. (USE A PLASTIC STRAIN RELIEF.) ALL CONNECTIONS TO THE AC LINE MUST BE INSIDE A FULLY ENCLOSED HOUSING! LEAVING SUCH CONNECTIONS EXPOSED IS A POTENTIAL SHOCK HAZARD. MAKE SURE ALL COMPONENTS THAT ARE CONNECTED TO THE AC LINE (SWITCHES, FUSES, TRANSFORMERS, ETC.) MEET OR EXCEED THE POWER REQUIREMENT OF YOUR CIRCUIT.

SUMMING UP CIRCUIT ASSEMBLY

THE REMAINDER OF THIS BOOK INCLUDES MANY CIRCUITS YOU CAN QUICKLY ASSEMBLE ON A SOLDERLESS BREADBOARD. CHANCES ARE YOU'LL WANT TO MAKE PERMANENT VERSIONS OF SOME. FOR BEST RESULTS, PLAN THE PROJECT CAREFULLY. A NEATLY ASSEMBLED PROJECT WILL BE MORE RELIABLE THAN ONE HASTILY ASSEMBLED.

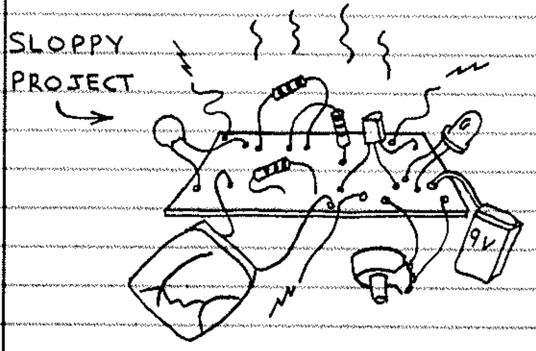

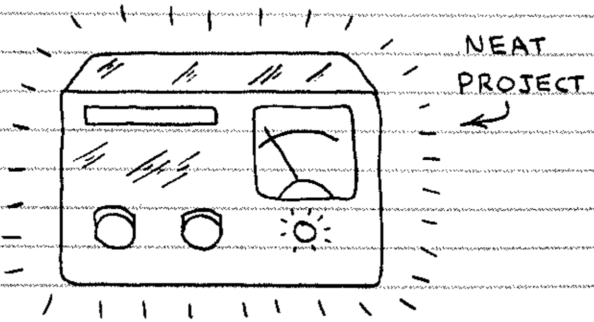

9. 100 ELECTRONIC CIRCUITS

HERE'S A COLLECTION OF 100 ELECTRONIC CIRCUITS. I'VE ASSEMBLED EACH CIRCUIT TO MAKE SURE ALL OF THEM WORK.

☐ SELECTING AND SUBSTITUTING COMPONENTS — YOU CAN FIND MOST OF THE COMPONENTS AT RADIO SHACK STORES. SAVE TIME AND MAKE A LIST OF WHAT YOU NEED BEFORE YOU VISIT RADIO SHACK. (YOU CAN FIND CURRENT CATALOG NUMBERS IN THE LATEST RADIO SHACK CATALOG.) IF A COMPONENT IS UNAVAILABLE, TRY ELSEWHERE. SOMETIMES YOU CAN SUBSTITUTE COMPONENTS. FOR EXAMPLE, IT'S OFTEN OK TO SUBSTITUTE NPN SWITCHING TRANSISTORS FOR ONE ANOTHER (2N3904 FOR 2N2222, ETC.). NEARBY VALUES OF RESISTORS AND CAPACITORS CAN OFTEN BE USED (1.2K FOR 1K RESISTOR, 0.33 μF FOR 0.47 μF CAPACITOR, ETC.). ALWAYS FOLLOW APPROPRIATE VOLTAGE AND POWER RATINGS!

☐ WHEN A CIRCUIT DOESN'T WORK — MAKE SURE THE CIRCUIT IS RECEIVING ADEQUATE POWER. IF IT IS OR IF YOU SMELL OR FEEL A HOT COMPONENT, IMMEDIATELY DISCONNECT THE POWER AND FOLLOW THESE STEPS: ① RECHECK ALL CONNECTIONS. (IS A WIRE MISSING? IS AN IC PIN BENT? IS A SOLDER CONNECTION BAD? IS A WIRE "SHORTED"? IS A DIODE BACKWARDS?) ② IS A COMPONENT DEFECTIVE? ③ SOMETIMES, ESPECIALLY WHEN POWER SUPPLY LEADS ARE MORE THAN SIX INCHES LONG, IC CIRCUITS WILL WORK IMPROPERLY OR NOT AT ALL UNLESS YOU CONNECT A 0.1 μF CAPACITOR ACROSS THE POWER SUPPLY PINS OF EACH CHIP. IT MAY ALSO BE NECESSARY TO CONNECT A 1 TO 10 μF CAPACITOR ACROSS THE POWER LEADS WHERE THEY ENTER THE BOARD. ④ DOES THE PUBLISHED CIRCUIT CONTAIN AN ERROR?

☐ SAFETY FIRST — BE SURE TO FOLLOW APPROPRIATE PRECAUTIONS WHEN WORKING WITH AC LINE POWERED CIRCUITS. BE CAREFUL WHEN SOLDERING. CIRCUITS WITH SPEAKERS CAN PRODUCE VERY LOUD SOUNDS. KEEP YOUR DISTANCE, AND DON'T USE HEADPHONES.

☐ GOING FURTHER — TRY EXPERIMENTING WITH THE VALUES OF COMPONENTS IN RC CIRCUITS (P. 37). TRY SUBSTITUTING OTHER OUTPUT DEVICES IN CIRCUITS THAT DRIVE A RELAY, PIEZO BUZZER, ETC. (BE SURE TO FOLLOW VOLTAGE AND CURRENT RATINGS. USE OHM'S LAW AND, IF NECESSARY, ADD A SERIES RESISTOR TO REDUCE CURRENT.) BEFORE BUILDING A PERMANENT VERSION OF A CIRCUIT, ALWAYS ASSEMBLE AND TEST A BREADBOARD VERSION. FINALLY, BE SURE TO BUY RADIO SHACK'S CURRENT "SEMICONDUCTOR REFERENCE GUIDE" AND "ENGINEER'S NOTEBOOK." FOR MORE ADVANCED CIRCUITS AND INFORMATION ABOUT NEW DEVELOPMENTS, READ MY COLUMN ("THE ELECTRONICS SCIENTIST") IN COMPUTERS & ELECTRONICS.

DIODE CIRCUITS

THE VARIOUS KINDS OF DIODES HAVE MANY APPLICATIONS. HERE ARE SOME TYPICAL CIRCUITS:

SMALL SIGNAL DIODES AND RECTIFIERS

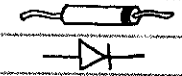

☐ VOLTAGE REGULATOR

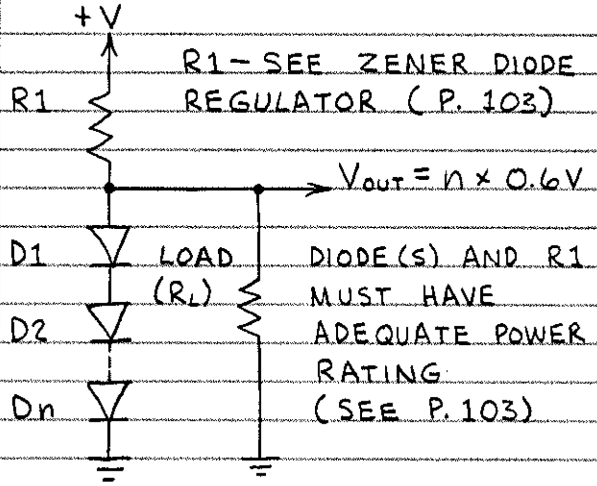
+V
R1
D1
D2
Dn
LOAD
(R$_L$)
$V_{OUT} = n \times 0.6V$

R1 — SEE ZENER DIODE REGULATOR (P. 103)

DIODE(S) AND R1 MUST HAVE ADEQUATE POWER RATING (SEE P. 103)

ONE OR MORE SILICON DIODES CAN REGULATE A VOLTAGE IN 0.6 VOLT STEPS.

☐ VOLTAGE DROPPER

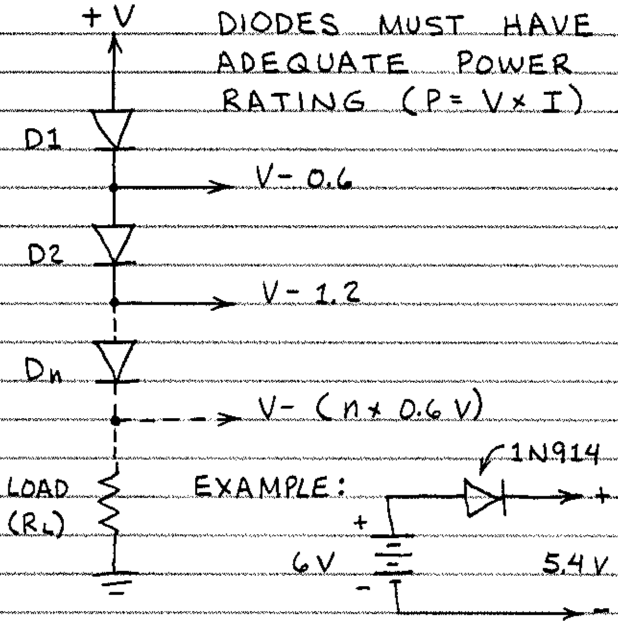
+V
D1
$V - 0.6$
D2
$V - 1.2$
Dn
$V - (n \times 0.6\,V)$
LOAD
(R$_L$)
EXAMPLE:
1N914
+
6V
-
+
5.4V
-

DIODES MUST HAVE ADEQUATE POWER RATING ($P = V \times I$)

☐ 9 VOLT POWER SUPPLY

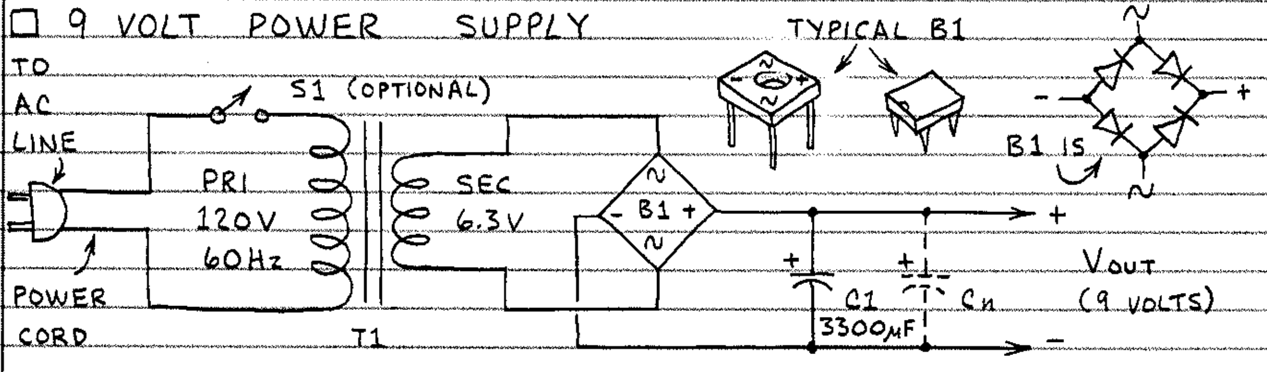
TO AC LINE
S1 (OPTIONAL)
PRI
120V
60Hz
SEC
6.3V
POWER CORD
T1
TYPICAL B1
B1 IS
~
- B1 +
~
+
C1
3300µF
Cn
+
VOUT
(9 VOLTS)
-

THIS IS A BASIC AC LINE OPERATED 9 VOLT POWER SUPPLY. FOR LOW RIPPLE (SUPERIMPOSED AC AT V_{OUT}), USE LARGE VALUE FOR C1. (OK TO ADD ONE OR MORE CAPACITORS (C_n) IN PARALLEL WITH C1 FOR MORE CAPACITANCE.) CAPACITORS MUST HAVE A DC WORKING VOLTAGE (WVDC) OF AT LEAST 12 VOLTS. RECTIFIER BRIDGE B1 MUST HAVE PEAK INVERSE VOLTAGE (PIV) OF AT LEAST 12 VOLTS. T1 AND B1 MUST HAVE ADEQUATE POWER AND CURRENT RATINGS. (USE OHMS LAW...) CAUTION: YOU MUST INSULATE OR ENCLOSE ALL EXPOSED AC LINE CONNECTIONS! THE POWER CORD MUST BE UNPLUGGED WHEN YOU ASSEMBLE OR SERVICE THE CIRCUIT!

CAUTION: THESE CIRCU
CAN PRODUC
HIGH VOLTAG

☐ VOLTAGE DOUBLERS

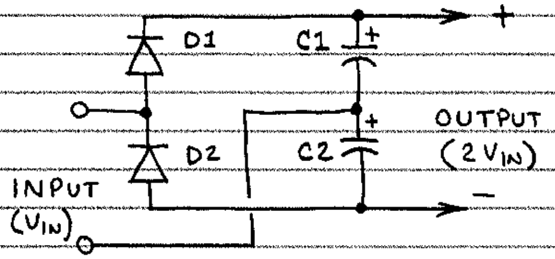
D1
C1
+
D2
C2
+
OUTPUT
(2 V_{IN})
+
−
INPUT
(V_{IN})

CONVENTIONAL

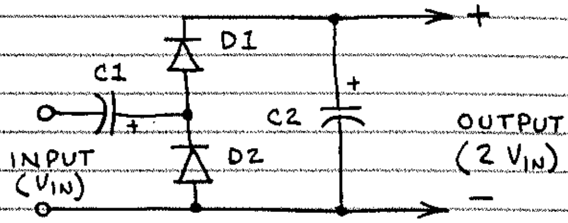
D1
C1
+
D2
C2
+
INPUT
(V_{IN})
OUTPUT
(2 V_{IN})
+
−

CASCADE

THESE CIRCUITS APPROXIMATELY DOUBLE AN INCOMING AC VOLTAGE. THE OUTPUT IS DC. USE CAPACITORS AND DIODES RATED FOR TWICE THE INPUT VOLTAGE. OUTPUT RIPPLE (∼∼∼) CAN BE REDUCED BY USING LARGE VALUES FOR C1 AND C2.

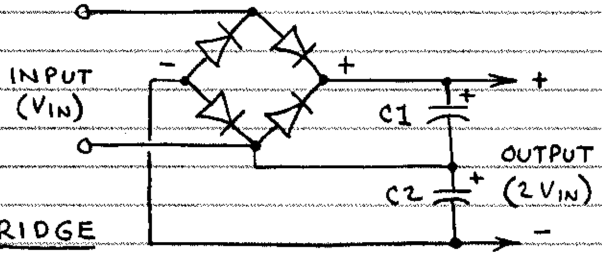
INPUT
(V_{IN})
−
+
C1
+
C2
+
OUTPUT
(2 V_{IN})
+
−

BRIDGE

THE BRIDGE DOUBLER IS MORE EFFICIENT THAN THE CONVENTIONAL AND CASCADE DOUBLERS. SINCE 4-DIODE BRIDGE RECTIFIERS ARE AVAILABLE, IT'S EASY TO MAKE.

☐ VOLTAGE TRIPLER

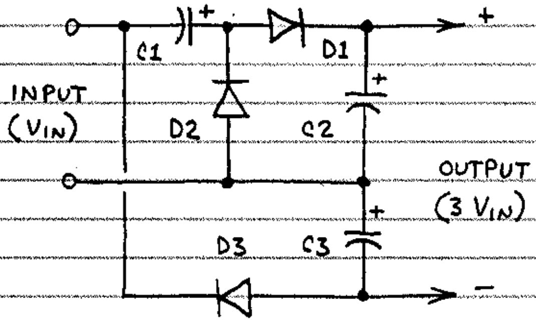
INPUT
(V_{IN})
C1
+
D1
D2
C2
+
OUTPUT
(3 V_{IN})
D3
C3
+
+
−

TRIPLES AND CONVERTS TO DC AN INCOMING AC VOLTAGE. C2, D1, D2 AND D3 RATED AT >2 V_{IN}.

☐ VOLTAGE QUADRUPLER

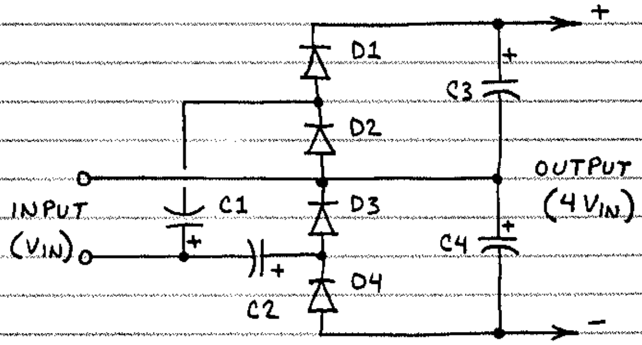
D1
D2
D3
D4
C1
C2
C3
C4
+
+
+
+
INPUT
(V_{IN})
OUTPUT
(4 V_{IN})
+
−

QUADRUPLES AND CONVERTS TO DC AN INCOMING AC VOLTAGE. ALL COMPONENTS RATED AT >2 V_{IN}.

☐ CASCADE MULTIPLIER

ADD MORE STAGES FOR MORE MULTIPLICATION. ALL COMPONENTS RATED AT >2 V_{IN}.

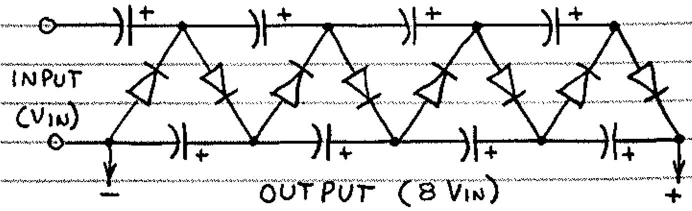
INPUT
(V_{IN})
−
+
OUTPUT (8 V_{IN})

ZENER DIODE CIRCUITS

☐ VOLTAGE REGULATOR

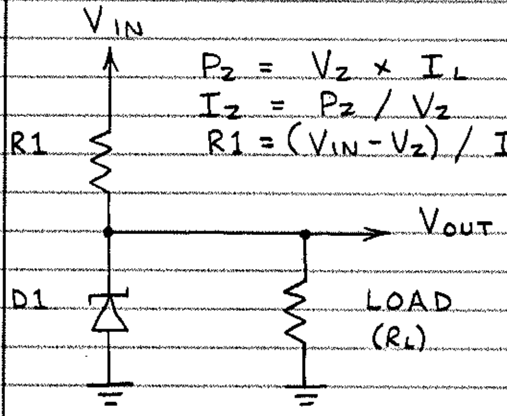

$P_Z = V_Z \times I_L$

$I_Z = P_Z / V_Z$

$R1 = (V_{IN} - V_Z) / I$

I_L = MAXIMUM LOAD CURRENT
I_Z = MAXIMUM ZENER CURRENT
I = CURRENT THROUGH R1
V_Z = ZENER DIODE VOLTAGE
P_Z = ZENER DIODE POWER

THIS CIRCUIT SUPPLIES A STEADY VOLTAGE (V_{OUT}) TO A LOAD FROM AN UNREGULATED SUPPLY (LIKE A BATTERY). V_{IN} CAN VARY BUT MUST BE AT LEAST 1 VOLT ABOVE DESIRED V_{OUT}. I_L CAN VARY FROM 0 MA TO PLANNED MAXIMUM VALUE. I DOES NOT CHANGE IF I_L FALLS TO 0 MA. SINCE $I = I_L + I_Z$, I_Z RISES AS I_L FALLS. IN OTHER WORDS, THE REGULATOR ALWAYS USES THE SAME CURRENT, EVEN WHEN THE LOAD IS REMOVED. CAUTION: D1 AND R1 MUST HAVE PROPER POWER RATING. USE OHM'S LAW (P. 14).

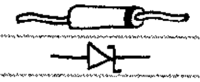

EXAMPLE: A RADIO DRAWS FROM 20 TO 50 MA FROM A 9 VOLT BATTERY. TO POWER IT FROM A 12 VOLT BATTERY, USE A 9 VOLT, 1/2 WATT ZENER DIODE. R1 SHOULD BE CLOSE TO 60 Ω AND RATED FOR AT LEAST 0.15 W.

EXAMPLE CIRCUIT:

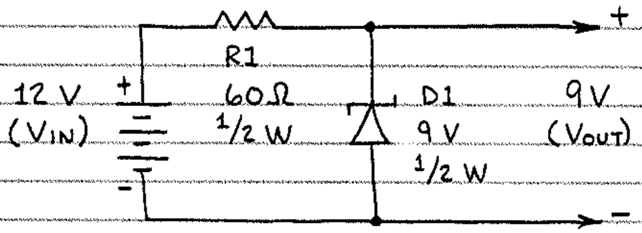

☐ WAVEFORM CLIPPER

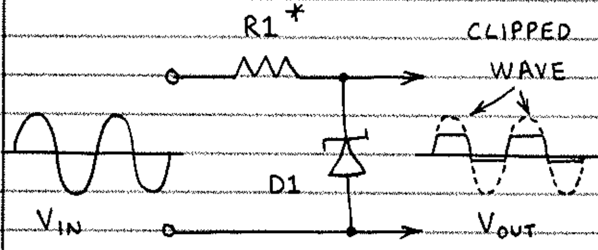

THIS CIRCUIT IS HANDY FOR REDUCING AN INCOMING SIGNAL VOLTAGE TO A LOWER, MORE MANAGEABLE LEVEL. IT CAN ALSO CONVERT A SINE OR TRIANGLE WAVE TO AN APPROXIMATION OF A SQUARE WAVE. *R1: SEE ABOVE (LET I = 2 MA MINIMUM).

☐ DUAL WAVEFORM CLIPPER

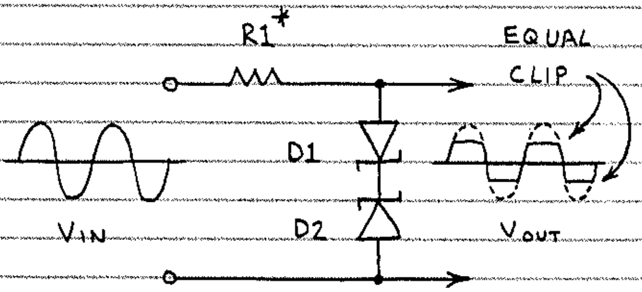

THIS IS A SYMMETRICAL FORM OF THE ADJACENT CIRCUIT. IT CLIPS BOTH HALVES OF AN INCOMING WAVE EQUALLY (IF V_Z = D1 = D2). USE TO PROTECT SPEAKERS AND PHONES FROM EXCESSIVE SIGNAL LEVELS OR TO MAKE SQUARE WAVES.

TRANSISTOR CIRCUITS

INTEGRATED CIRCUITS RECEIVE MORE ATTENTION, BUT BOTH BIPOLAR AND FIELD-EFFECT TRANSISTORS HAVE MANY APPLICATIONS.

BIPOLAR TRANSISTOR CIRCUITS

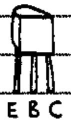

2N2222
2N2907,
ETC.

☐ MOISTURE METER

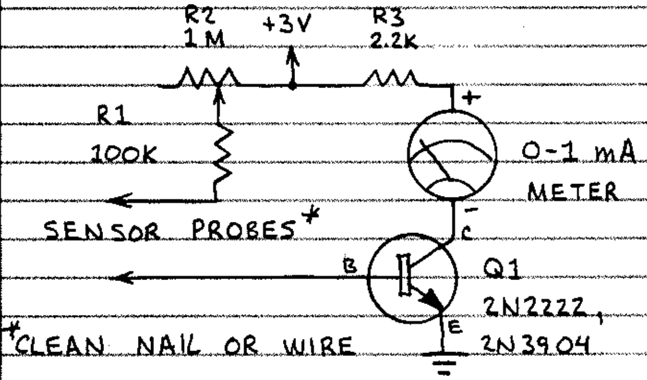
R2
1M
+3V
R3
2.2K
R1
100K
+
0-1 mA
METER
−
SENSOR PROBES*
C
B
Q1
2N2222,
2N3904
E

*CLEAN NAIL OR WIRE

THIS CIRCUIT WILL MEASURE THE MOISTURE LEVEL OF SOIL IN YOUR GARDEN. ADJUST R2 FOR METER READING OF 1 mA WHEN SOIL MOISTURE IS AT DESIRED LEVEL. METER WILL THEN INDICATE LOWER MOISTURE LEVELS.

☐ MOISTURE ACTUATED RELAY

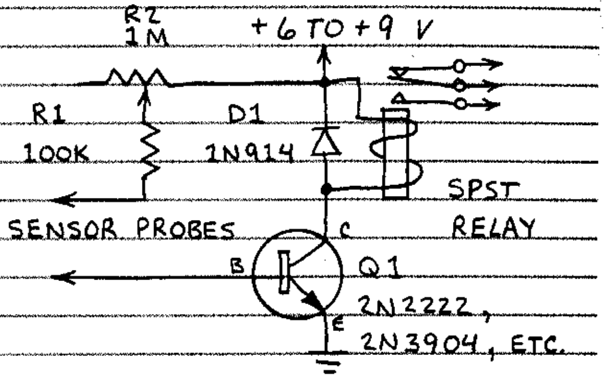
R2
1M
+6 TO +9 V
R1
100K
D1
1N914
SPST
RELAY
SENSOR PROBES
C
B
Q1
2N2222,
2N3904, ETC.
E

THIS CIRCUIT CLOSES A RELAY (6-9 V, 500 OHM COIL) WHEN MOISTURE LEVEL EXCEEDS A LEVEL SET BY R2. RELAY CONTACTS THEN SWITCH ON A LIGHT OR OTHER DEVICE. CAN DETECT RAIN.

☐ METRONOME

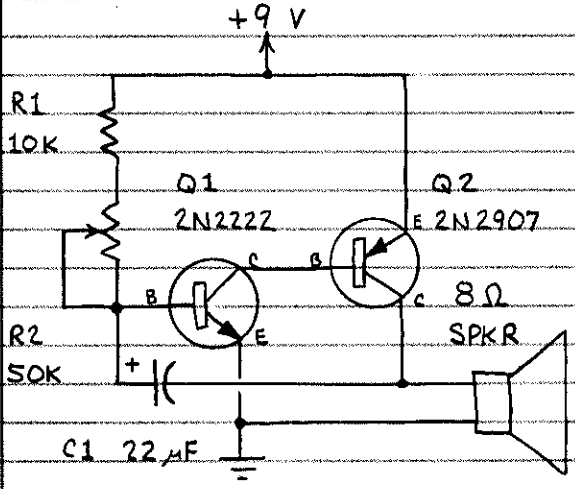
+9 V
R1
10K
Q1
2N2222
Q2
E 2N2907
C
B
B
E
C
8Ω
SPKR
R2
50K
+
C1 22 µF

A METRONOME MARKS TIME BY PRODUCING A REGULAR SEQUENCE OF "CLICKS" OR "POCKS." ADJUST THE CLICK RATE BY ADJUSTING R2 OR CHANGING C1'S VALUE.

☐ LIGHT FLASHER

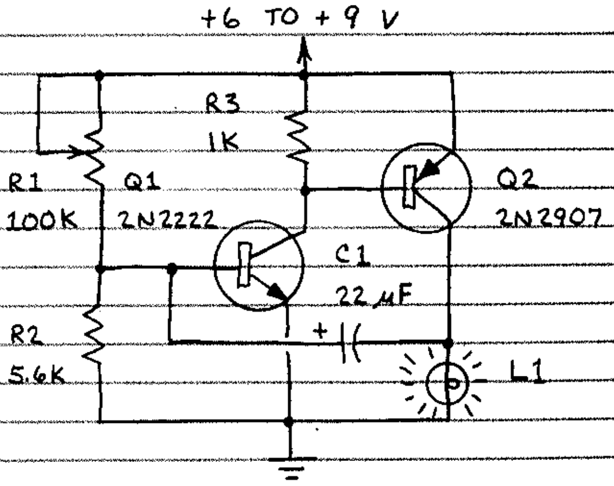
+6 TO +9 V
R3
1K
R1
100K
Q1
2N2222
Q2
2N2907
C1
22 µF
+
R2
5.6K
L1

THIS CIRCUIT PRODUCES BRIGHT FLASHES EVERY SECOND OR SO. R1 CONTROLS FLASH RATE. USE NO. 122 OR 222 MINIATURE LAMP FOR L1.

☐ SIREN

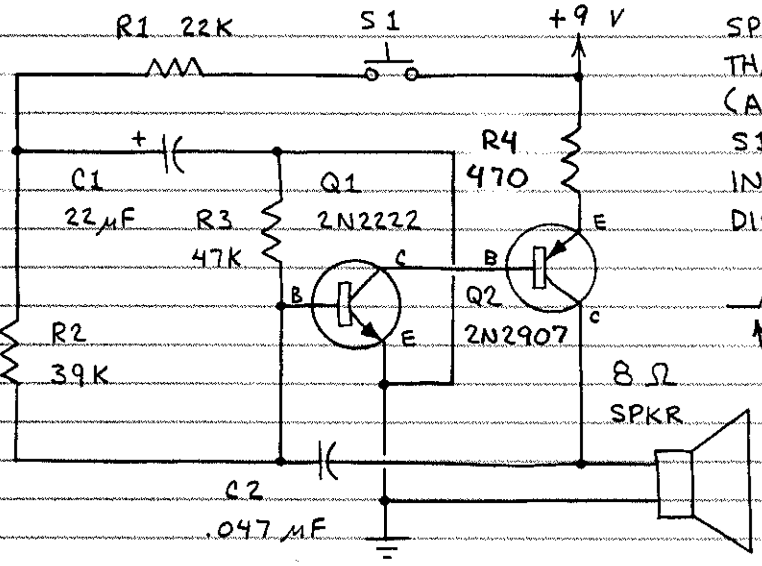
R1 22K
S1
+9 V
+
C1
22 µF
R3
47K
Q1
2N2222
R4
470
E
C
B
Q2
2N2907
R2
39K
E
C
8 Ω
SPKR
C2
.047 µF

CLOSE S1 AND THE SPEAKER EMITS A TONE THAT RISES IN FREQUENCY (AS C1 CHARGES). OPEN S1 AND THE TONE FALLS IN FREQUENCY (AS C1 DISCHARGES). LIKE THIS:

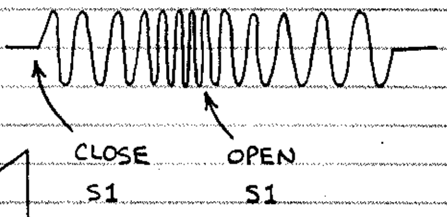
CLOSE
S1
OPEN
S1

HINT: CHANGE R1 TO CHANGE UP-DOWN TIME.

☐ HIGH VOLTAGE POWER SUPPLY

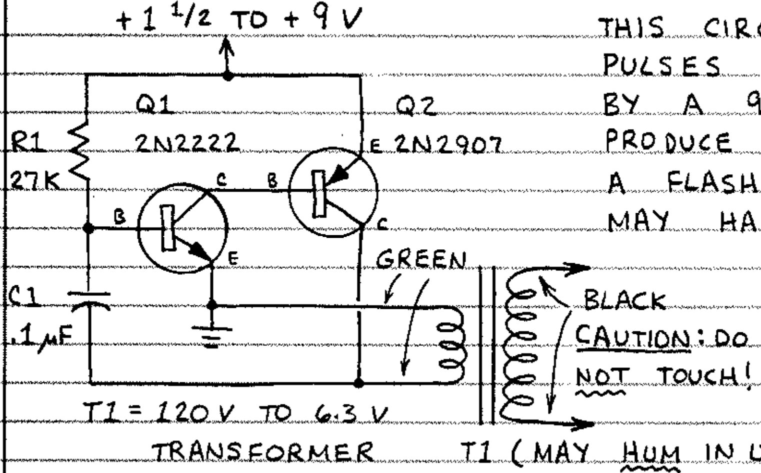
+1 1/2 TO +9 V
Q1
2N2222
Q2
E 2N2907
R1
27K
C
B
B
E
C
C1
.1 µF
GREEN
BLACK
CAUTION: DO NOT TOUCH!
T1 = 120 V TO 6.3 V TRANSFORMER
T1 (MAY HUM IN USE)

THIS CIRCUIT PRODUCES 220 VOLT PULSES OF DC WHEN POWERED BY A 9-VOLT BATTERY. IT WILL PRODUCE UP TO 170 VOLTS FROM A FLASHLIGHT CELL (BUT YOU MAY HAVE TO EXPERIMENT WITH C1'S VALUE. THE CIRCUIT WILL POWER ONE OR MORE NEON LAMPS THROUGH A 1M SERIES RESISTOR.

☐ BURGLAR ALARM

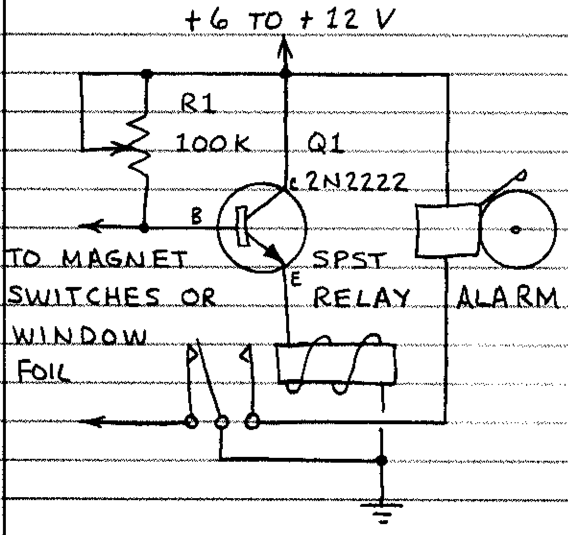
+6 TO +12 V
R1
100K
Q1
C 2N2222
B
E
TO MAGNET SWITCHES OR WINDOW FOIL
SPST
RELAY
ALARM

ALARM SOUNDS AND STAYS ON (UNTIL POWER IS DISCONNECTED) WHEN MAGNET SWITCH IS OPENED OR WINDOW FOIL IS BROKEN. R1: SET TO MAXIMUM VALUE. DISCONNECT ONE LEAD TO SWITCHES / FOIL. THEN REDUCE VALUE OF R1 TO JUST PAST POINT WHERE ALARM SOUNDS. CIRCUIT USES ONLY 0.3 MILLIAMPERES AT 6 VOLTS.

RELAY: USE 6 V, 500 Ω UNIT WHEN SUPPLY IS 6-9 V. USE 12 V, 1200 Ω UNIT WHEN SUPPLY IS 12 V.

NOTE: ASSEMBLE, INSTALL AND CONCEAL THIS SYSTEM WITH CARE.

JUNCTION FET CIRCUITS

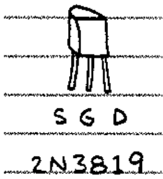

☐ ELECTROMETER

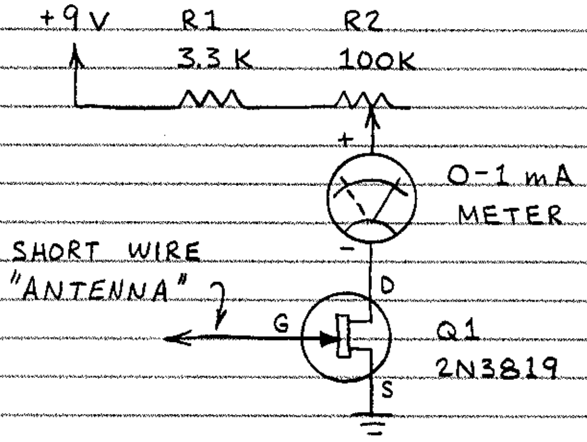
+9V
R1
3.3K
R2
100K
+
0-1 mA
METER
−
SHORT WIRE
"ANTENNA"
D
G
Q1
2N3819
S

THIS CIRCUIT DETECTS STATIC ELECTRICITY FROM A CHARGED OBJECT (PLASTIC COMB, ETC.) OVER A FOOT AWAY! ADJUST R1 SO METER INDICATES 1 MA. CHARGED OBJECT NEAR "ANTENNA" WILL DECREASE METER READING.

☐ TOUCH SWITCH

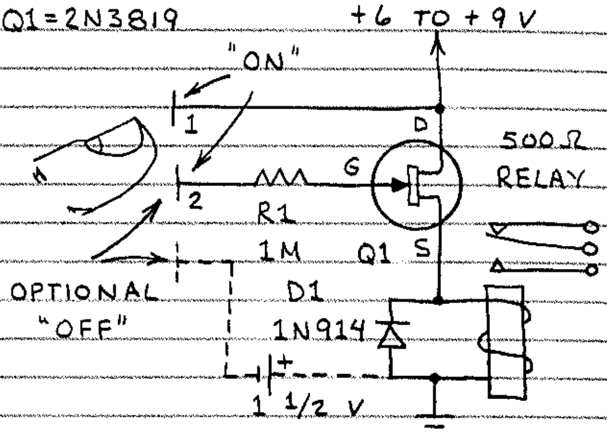
Q1=2N3819
+6 TO +9V
"ON"
1
2
D
500Ω
RELAY
G
R1
1M
Q1
S
OPTIONAL
"OFF"
D1
1N914
+
1 1/2 V

OUTDOORS:
AWAY FROM POWER LINES, BRIEFLY TOUCH "ON" CONTACTS TO ACTUATE RELAY. TOUCH CONTACT 2 ONLY TO DEACTUATE RELAY.
INDOORS:
MAY BE NECESSARY TO INCLUDE OPTIONAL "OFF" CIRCUIT.

☐ TIMER

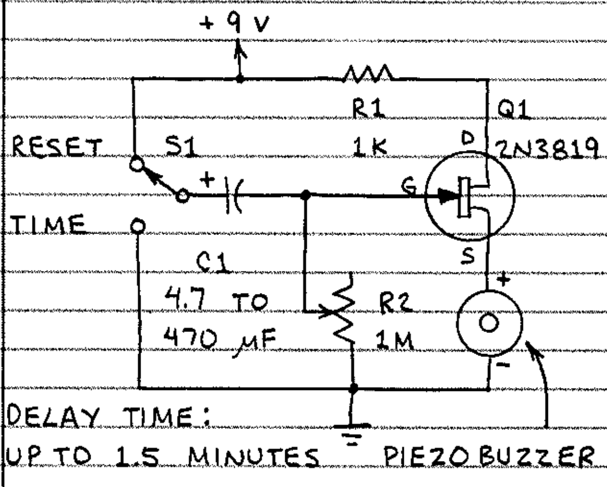
+9V
R1
1K
Q1
2N3819
RESET
S1
D
G
+
TIME
S
C1
4.7 TO
470 μF
R2
1M
+
−
PIEZO BUZZER

DELAY TIME:
UP TO 1.5 MINUTES

SET S1 TO "RESET" (BUZZER WILL SOUND). THEN SET S1 TO "TIME." BUZZER WILL BE SILENT UNTIL DELAY IS COMPLETE. BUZZER WILL THEN SOUND. INCREASE C1 OR R1 FOR LONGER DELAYS. REDUCE R2'S RESISTANCE DURING RESET MODE (SPEEDS UP RESET).

☐ AUDIO MIXER

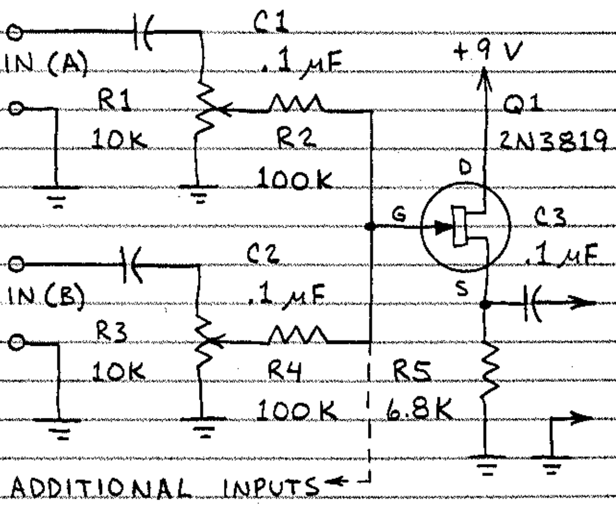
IN (A)
C1
.1 μF
+9V
R1
10K
R2
100K
Q1
2N3819
D
G
C3
.1 μF
S
IN (B)
C2
.1 μF
R3
10K
R4
100K
R5
6.8K
ADDITIONAL INPUTS

THIS CIRCUIT ALLOWS TWO (OR MORE) MICROPHONES OR OTHER DEVICES TO BE CONNECTED TO THE SAME AMPLIFIER. R1 AND R3 CONTROL ATTENUATION OF EACH INPUT. THEREFORE R1 AND R3 ARE BALANCE CONTROLS.

POWER MOSFET (DMOS, VMOS, ETC) CIRCUITS

☐ LINEAR LIGHT DIMMER

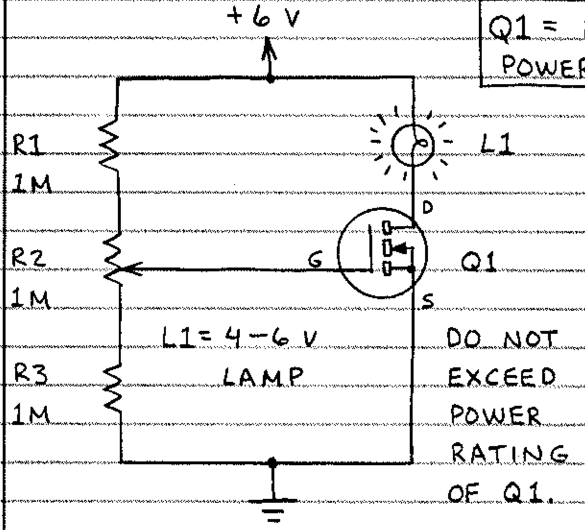
+6 V
R1
1M
R2
1M
R3
1M
L1
D
G
S
Q1
L1 = 4-6 V
LAMP

DO NOT EXCEED POWER RATING OF Q1.

CHANGE SETTING OF R2 TO CHANGE INTENSITY OF LAMP.

THIS CIRCUIT SHOWS HOW POWER MOSFET CAN BE USED AS A VARIABLE RESISTOR.

☐ AUDIO AMPLIFIER

Q1 = N-CHANNEL POWER MOSFET

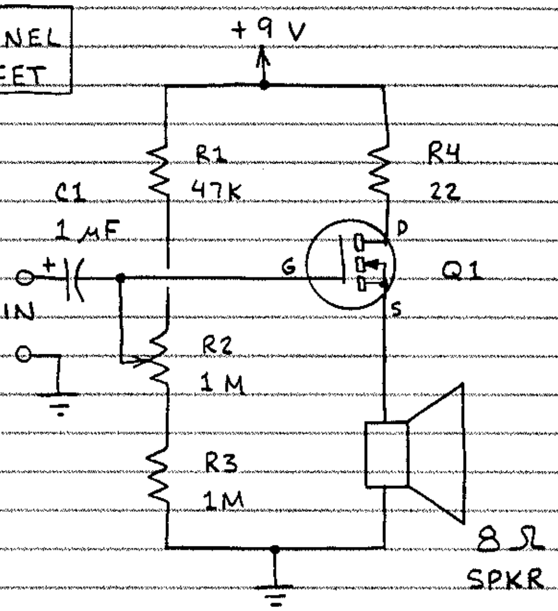
+9 V
C1
1 μF
+
IN
R1
47K
R2
1 M
R3
1M
R4
22
D
G
S
Q1
8 Ω
SPKR

USE TO AMPLIFY SIGNALS AND TONES FROM OTHER CIRCUITS. R2 CONTROLS GAIN (VOLUME).

☐ LONG DURATION TIME DELAY CIRCUITS

1. *OFF* AFTER DELAY

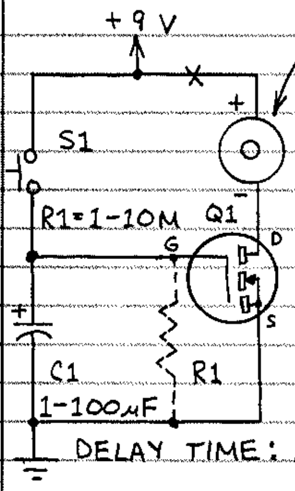
+9 V
X
+
S1
R1=1-10M
Q1
G
D
S
+
C1
1-100 μF
R1

PIEZOBUZZER (OR PORTABLE RADIO, LIGHT, ETC.). DO *NOT* EXCEED POWER RATING OF Q1. USE SERIES RESISTOR AT "X" TO REDUCE CURRENT THROUGH Q1 AND THE LOAD.

DELAY TIME: MINUTES TO HOURS!

CLOSE AND THEN OPEN S1 TO ACTIVATE BUZZER. AFTER C1 DISCHARGES INTERNALLY OR THROUGH R1 (OPTIONAL), Q1 TURNS OFF AND SILENCES THE BUZZER. *LONG* DELAYS POSSIBLE.

2. *ON* AFTER DELAY

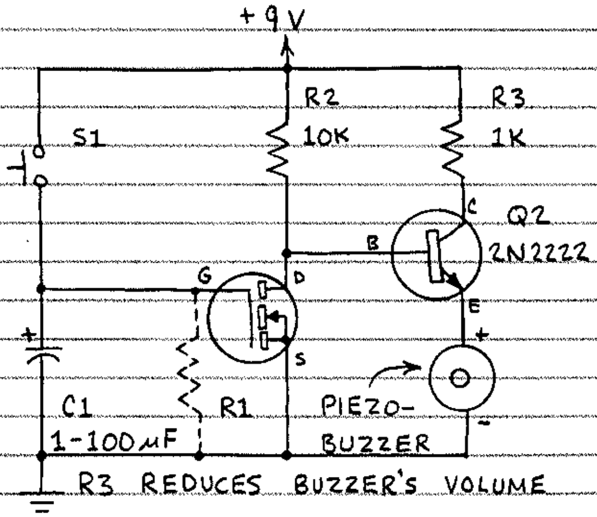
+9 V
S1
R2
10K
R3
1K
C
B
E
Q2
2N2222
G
D
S
+
C1
1-100 μF
R1
+
-
PIEZO-
BUZZER

R3 REDUCES BUZZER'S VOLUME

Q2 INVERTS THE STATUS OF Q1. THEREFORE THE BUZZER SOUNDS *AFTER* THE TIME DELAY IS COMPLETE. INCREASE C1'S VALUE TO INCREASE DELAY.

UNIJUNCTION TRANSISTOR CIRCUITS

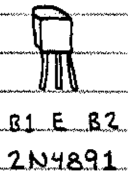
B1 E B2
2N4891

☐ TIME BASE

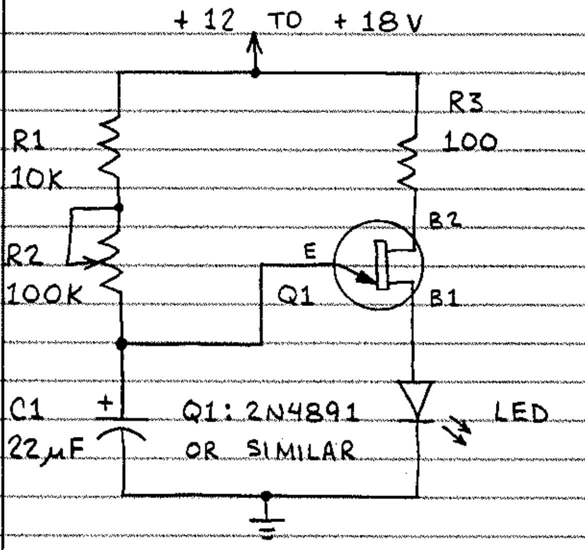
+12 TO +18V
R1
10K
R2
100K
R3
100
E
Q1
B2
B1
C1
22μF
+
Q1: 2N4891
OR SIMILAR
LED

EACH TIME THE UJT (Q1) TURNS ON, THE CHARGE ON C1 IS "DUMPED" THROUGH THE LED. THE LED FLASHES IN RESPONSE. BETWEEN FLASHES, THE LED GLOWS DIMLY. R2 CONTROLS FLASH RATE AND CAN BE SET FOR ONE FLASH PER SECOND (TIME BASE ROLE).

☐ TONE GENERATOR

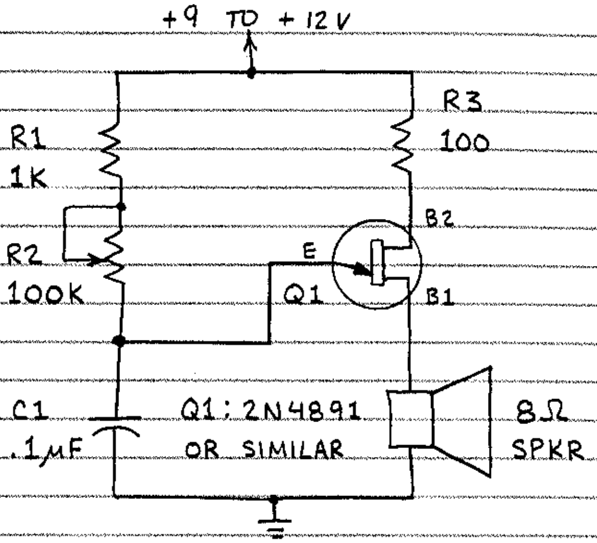
+9 TO +12V
R1
1K
R2
100K
R3
100
E
Q1
B2
B1
C1
.1μF
Q1: 2N4891
OR SIMILAR
8Ω
SPKR

THIS CIRCUIT IS IDENTICAL IN PRINCIPLE TO THE ADJACENT CIRCUIT. C1 IS MUCH SMALLER TO SPEED UP THE CHARGE-DISCHARGE CYCLE. THE RESULT IS AN AUDIO FREQUENCY TONE WHICH IS EMITTED BY THE SPEAKER. THIS CIRCUIT CAN BE EXPANDED (SEE BELOW).

☐ ORGAN

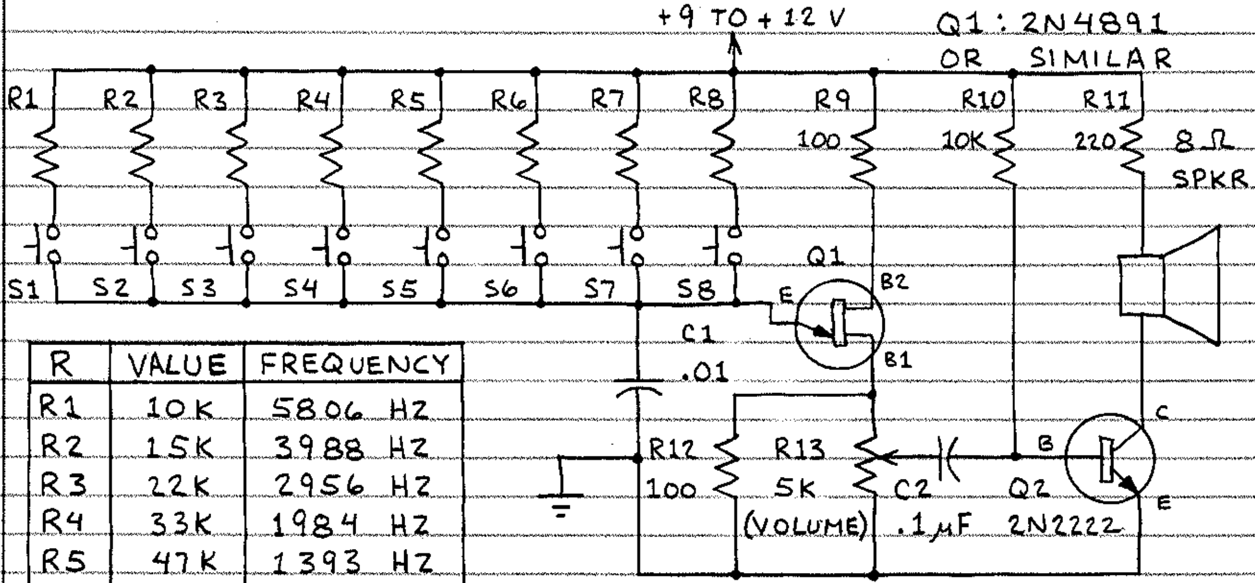
+9 TO +12 V
Q1: 2N4891
OR SIMILAR
R1
R2
R3
R4
R5
R6
R7
R8
R9
100
R10
10K
R11
220
8Ω
SPKR
S1
S2
S3
S4
S5
S6
S7
S8
Q1
E
B2
B1
C1
.01
R12
100
R13
5K
(VOLUME)
C2
.1μF
B
C
E
Q2
2N2222

R	VALUE	FREQUENCY
R1	10K	5806 HZ
R2	15K	3988 HZ
R3	22K	2956 HZ
R4	33K	1984 HZ
R5	47K	1393 HZ
R6	68K	941 HZ
R7	100K	583 HZ
R8	150K	430 HZ

TABLE SHOWS TYPICAL FREQUENCIES. CHANGE C1 TO ALTER THE OVERALL FREQUENCY RANGE.

☐ RAMP GENERATOR

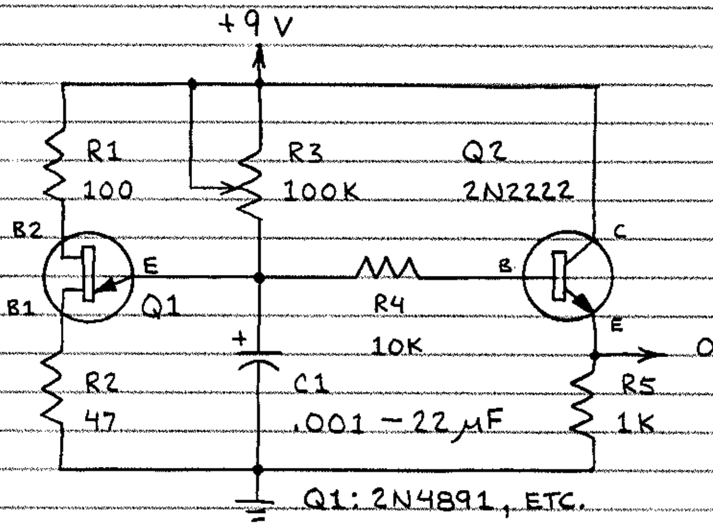
+9V
R1
100
R3
100K
Q2
2N2222
B2
E
Q1
B1
R4
10K
B
C
E
+
C1
.001 – 22 µF
R2
47
R5
1K
Q1: 2N4891, ETC.

Q2 SAMPLES THE VOLTAGE ON C1 AND OUTPUTS IT AS A RAMP (OR "SAWTOOTH" WAVE). R3 CONTROLS THE RATE AT WHICH RAMPS ARE PRODUCED.

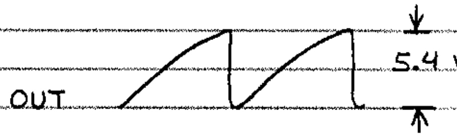

RAMPS SUPPLY GRADUALLY INCREASING VOLTAGE TO MANY KINDS OF CIRCUITS.

☐ CHIRP GENERATOR

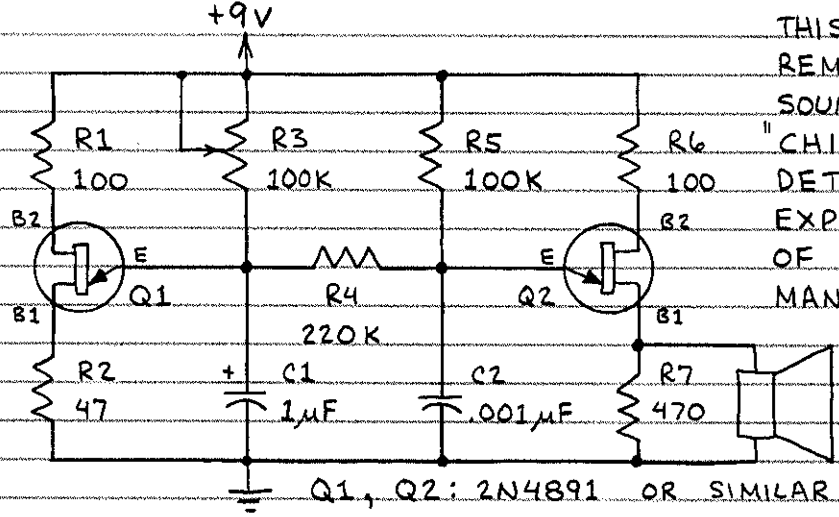
+9V
R1
100
R3
100K
R5
100K
R6
100
B2
E
Q1
B1
R4
220K
E
Q2
B2
B1
R2
47
+
C1
1µF
C2
.001µF
R7
470
PIEZO TWEETER
SPEAKER
Q1, Q2: 2N4891 OR SIMILAR

THIS CIRCUIT PRODUCES A REMARKABLE VARIETY OF SOUNDS. AS SHOWN, IT "CHIRPS" AT A RATE DETERMINED BY R3. EXPERIMENT WITH VALUES OF C1, R5 AND C2 FOR MANY OTHER EFFECTS.

☐ VOLTAGE SENSITIVE OSCILLATOR

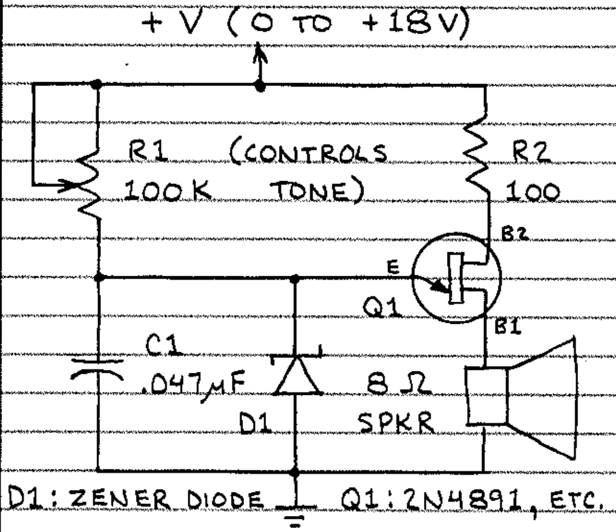
+V (0 TO +18V)
R1 (CONTROLS
100K TONE)
R2
100
E
Q1
B2
B1
C1
.047µF
D1
8 Ω
SPKR
D1: ZENER DIODE
Q1: 2N4891, ETC.

THIS CIRCUIT SOUNDS A TONE WHEN V_{IN} IS BELOW V_Z OF D1. SELECT V_Z OF D1 FOR DESIRED TURN-OFF LEVEL. THIS CIRCUIT CAN BE USED TO INDICATE WHEN VOLTAGE OF A BATTERY (WHICH CAN BE POWERING ANOTHER CIRCUIT) FALLS BELOW A CERTAIN LEVEL. THIS IS AN EXCELLENT EXAMPLE OF A SIMPLE, YET SOPHISTICATED, CIRCUIT.

THYRISTOR CIRCUITS

SILICON CONTROLLED RECTIFIERS AND TRIACS HAVE MANY APPLICATIONS AS SOLID-STATE SWITCHES.

SCR CIRCUITS

☐ LATCHING SWITCH

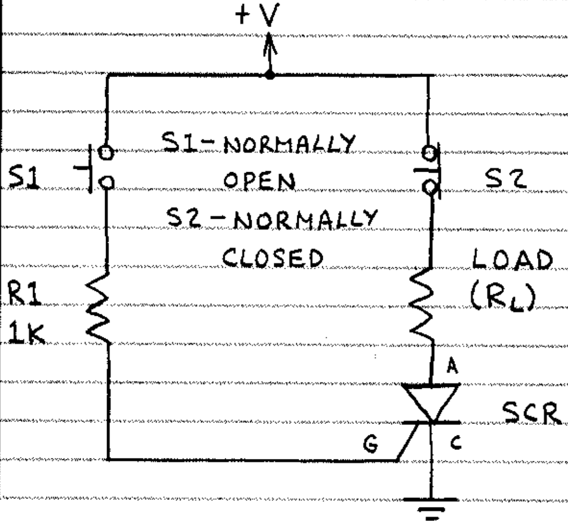
+V
S1
S1-NORMALLY OPEN
S2-NORMALLY CLOSED
S2
R1
1K
LOAD
(R_L)
A
SCR
G
C

CLOSE S1 TO TURN ON THE SCR AND SUPPLY CURRENT TO THE LOAD. THE SCR WILL REMAIN ON AFTER S1 IS OPENED UNLESS THE LOAD IS A DC MOTOR OR UNTIL S2 IS BRIEFLY OPENED.

☐ TEST CIRCUIT

S1 IS "SPRING RETURN TO CENTER OFF" SWITCH.

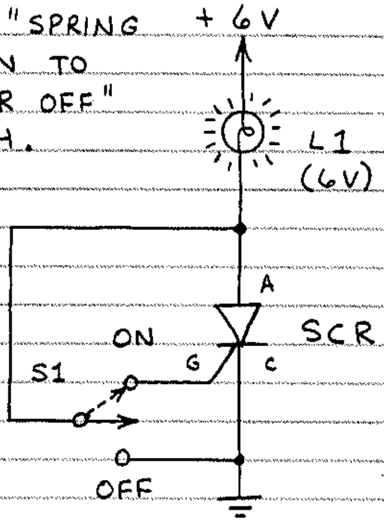
+6V
L1
(6V)
A
SCR
ON
G
C
S1
OFF

S1 IS AN SPDT SWITCH. IN THE "ON" POSITION, THE SCR IS TURNED ON AND THE LAMP GLOWS. IN THE "OFF" POSITION, CURRENT IS SHUNTED AWAY FROM SCR, THUS TURNING IT OFF.

☐ CAPACITOR DISCHARGE LED FLASHER

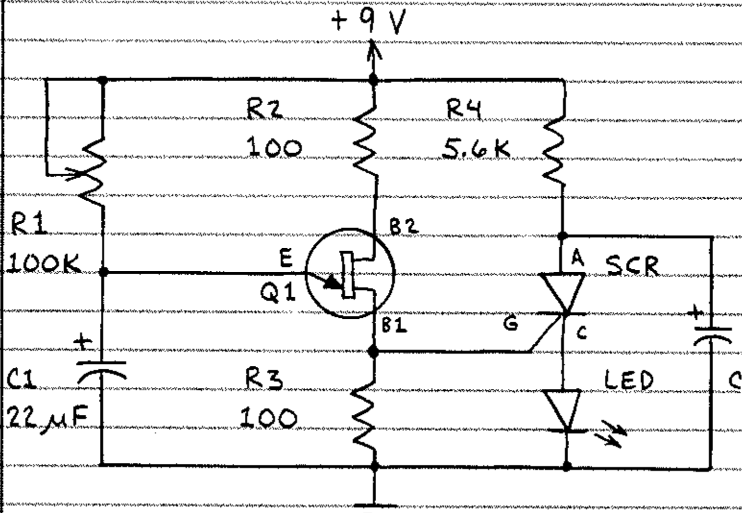
+9V
R2
100
R4
5.6K
R1
100K
E
Q1
B2
B1
A
SCR
G
C
+
C1
22µF
R3
100
LED
+
C2

Q1: 2N4891 UJT

R1 CONTROLS RATE.

WHEN THE SCR IS OFF, C2 CHARGES THROUGH R4. WHEN THE SCR IS TURNED ON BY A PULSE FROM UJT Q1, THE CHARGE IN C1 IS RAPIDLY "DUMPED" THROUGH THE LED. THE SCR (AND LED) THEN TURNS OFF SINCE THERE IS NO LONGER SUFFICIENT HOLDING CURRENT. THE CYCLE THEN REPEATS.

TRIAC CIRCUITS

☐ TEST CIRCUIT

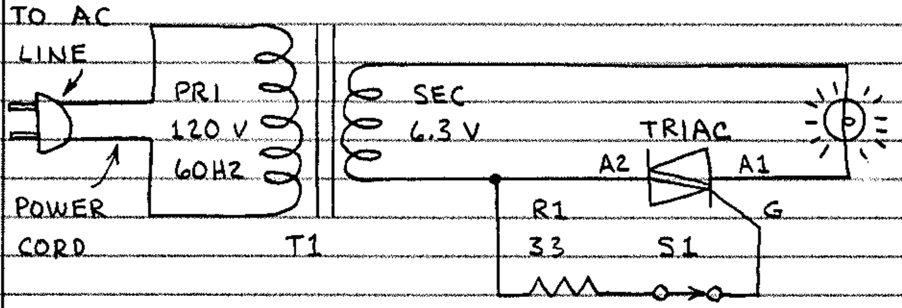
TO AC LINE
POWER CORD
PRI
120 V
60HZ
SEC
6.3 V
T1
TRIAC
A2
A1
G
R1
33
S1
L1
(NO. 47 LAMP)

L1 GLOWS WHEN S1 IS CLOSED.
L1 DOES NOT GLOW WHEN S1 IS OPEN.

CAUTION:
TRIACS ARE DESIGNED SPECIFICALLY FOR AC OPERATION. BE SURE TO OBSERVE COMMON SENSE SAFETY WHEN WORKING WITH HOUSEHOLD LINE CURRENT! MAKE SURE ALL CONNECTIONS TO AC LINE ARE INSULATED OR ENCLOSED.

☐ LIGHT DIMMER CIRCUITS

1. 6.3 VOLT DIMMER

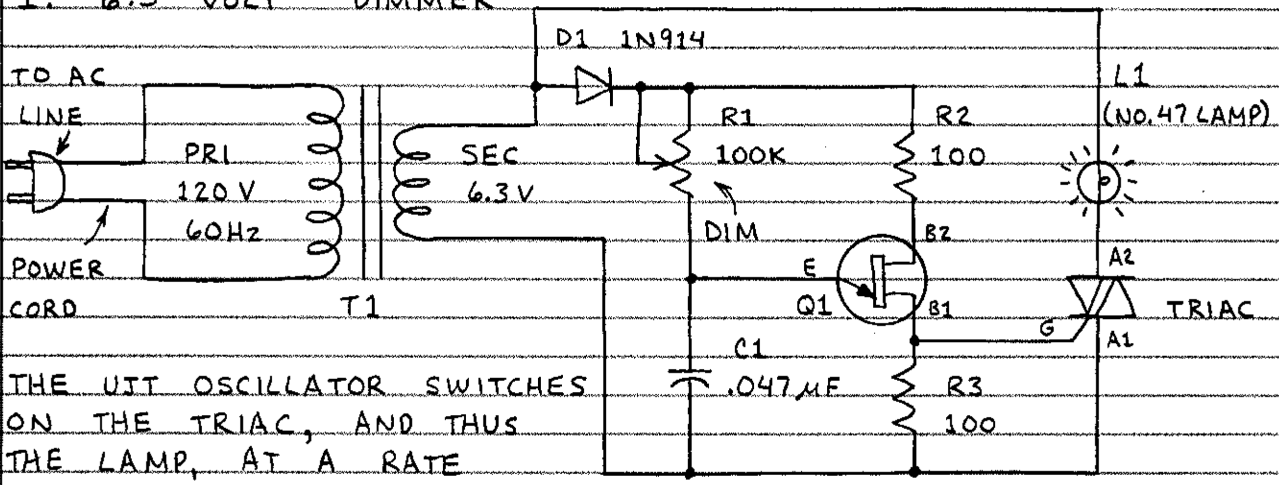
D1 1N914
TO AC LINE
POWER CORD
PRI
120 V
60HZ
SEC
6.3 V
T1
R1
100K
DIM
R2
100
L1
(NO. 47 LAMP)
E
Q1
B2
B1
A2
G
A1
TRIAC
C1
.047 µF
R3
100

THE UJT OSCILLATOR SWITCHES ON THE TRIAC, AND THUS THE LAMP, AT A RATE CONTROLLED BY R1.

Q1: 2N4891 OR SIMILAR UJT.

2. 120 VOLT DIMMER

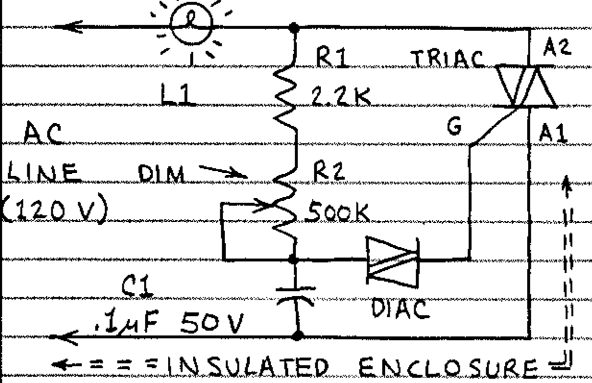
L1
R1
2.2K
TRIAC
A2
G
A1
AC LINE (120 V)
DIM
R2
500K
C1
.1µF 50V
DIAC
INSULATED ENCLOSURE

THIS IS HOW MANY HOUSEHOLD DIMMER SWITCHES ARE MADE. L1 CAN BE UP TO 100 WATT (120 VOLT) LAMP. USE HEAT SINK ON TRIAC IF IT BECOMES HOT. THE DIAC IS A BIDIRECTIONAL TRIGGER DIODE. CAUTION: THIS CIRCUIT MUST BE FULLY ENCLOSED AT ALL TIMES WHEN POWER IS APPLIED!

PHOTONIC CIRCUITS

CIRCUITS USING PHOTONIC COMPONENTS ARE AMONG THE MOST VERSATILE AND INTERESTING OF ALL CIRCUITS.

LIGHT EMITTING DIODE (LED) CIRCUITS

☐ LED DRIVE CIRCUIT

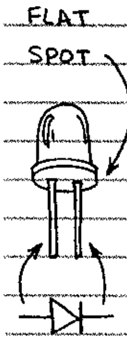

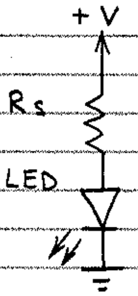

A SERIES RESISTOR MUST BE USED TO LIMIT CURRENT THROUGH AN LED. (EXCEPTIONS: CERTAIN PULSE CIRCUITS AND IC LED DRIVERS.)

$$R_S = \frac{+V - (V_{LED})}{I_{LED}}$$

EXAMPLE: ASSUME YOU WANT TO OPERATE A RED LED AT A FORWARD CURRENT (I_{LED}) OF 10 MA (OR 0.01 AMPERE) FROM A 5 VOLT SUPPLY. THE DATA SHEET FOR THE LED LISTS AN LED VOLTAGE (V_{LED}) OF 1.7 VOLTS. THEREFORE R_S IS (5-1.7)/.01 OR 330 OHMS.

☐ VARIABLE BRIGHTNESS LED

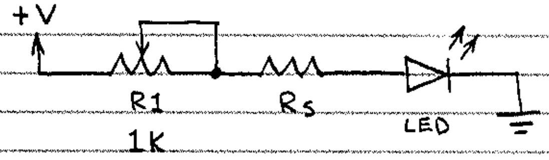
+V
R1
1K
R_S
LED

ADJUST R1 TO ALTER CURRENT THROUGH THE LED, THUS VARYING ITS BRIGHTNESS. R_S MUST BE USED (SEE ABOVE).

☐ POLARITY INDICATOR

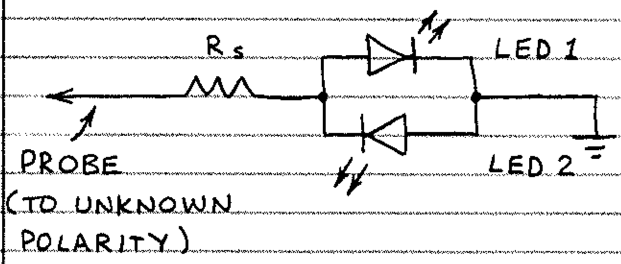
R_S
LED 1
LED 2
PROBE
(TO UNKNOWN
POLARITY)

THIS CIRCUIT INDICATES THE POLARITY OF A VOLTAGE. R_S MUST BE USED (SEE ABOVE).

LED	+	−	AC (±)
1	ON	OFF	ON
2	OFF	ON	ON

☐ TRI-STATE POLARITY INDICATOR

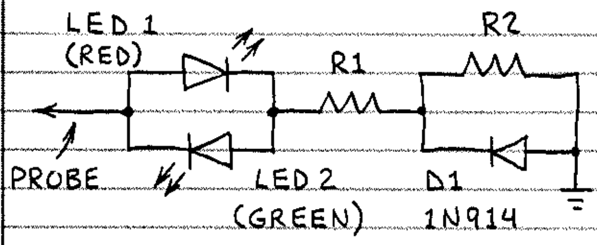
LED 1
(RED)
R1
R2
PROBE
LED 2
(GREEN)
D1
1N914

THIS IS A MORE COLORFUL VERSION OF THE CIRCUIT ABOVE.

V_{IN}	COLOR
+	RED
−	GREEN
AC(±)	YELLOW*

$$R1 = \frac{V_{IN} - (V_{LED2} + 0.6)}{I_{LED2}} \qquad R1 + R2 = \frac{V_{IN} - V_{LED1}}{I_{LED1}}$$

* WHEN TWO-CHIP LED USED.

☐ DUAL LED FLASHER

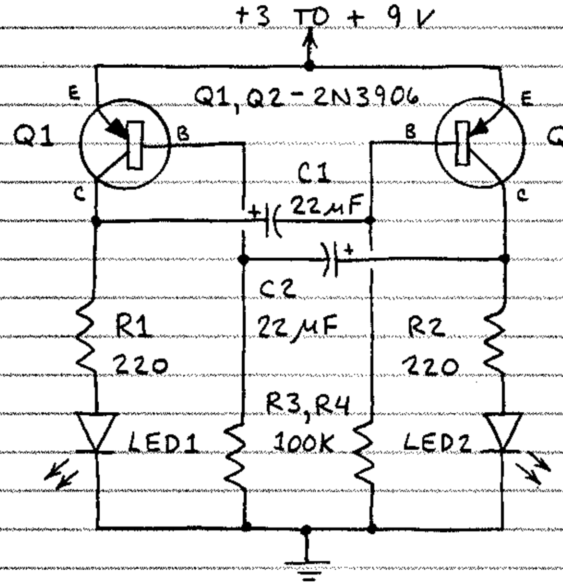
+3 TO + 9V
Q1, Q2 - 2N3906
E
Q1
B
C
B
E
Q2
C
C1
22µF
C2
22µF
R1
220
R2
220
R3, R4
100K
LED1
LED2

THIS CIRCUIT IS CALLED A FREE-RUNNING MULTIVIBRATOR. IT'S IDENTICAL TO A FLIP-FLOP THAT TRIGGERS ITSELF REPEATEDLY. Q1 AND Q2 ARE GENERAL PURPOSE PNP TRANSISTORS (2N3906, 2N2907, ETC.). R1 AND R2 LIMIT THE CURRENT TO THE LEDs (WHICH FLASH ALTERNATELY). INCREASING THE VALUES OF C1 AND C2 WILL SLOW THE FLASH RATE.

☐ VOLTAGE LEVEL INDICATOR

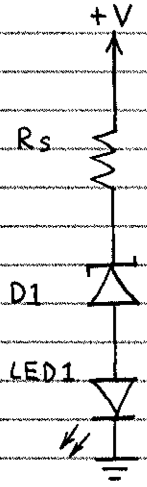

IN EACH CIRCUIT THE LED(s) GLOWS WHEN V+ REACHES THE BREAKDOWN VOLTAGE (V_Z) OF ITS ZENER + V_{LED}. BE SURE TO USE INDIVIDUAL R_S FOR EACH LED. (SEE FACING PAGE.) THE CIRCUIT AT RIGHT PROVIDES A BAR-GRAPH READOUT WHEN ZENERS WITH PROGRESSIVELY HIGHER V_Z ARE USED. CONNECT ZENERS IN SERIES FOR HIGHER TOTAL V_Z.

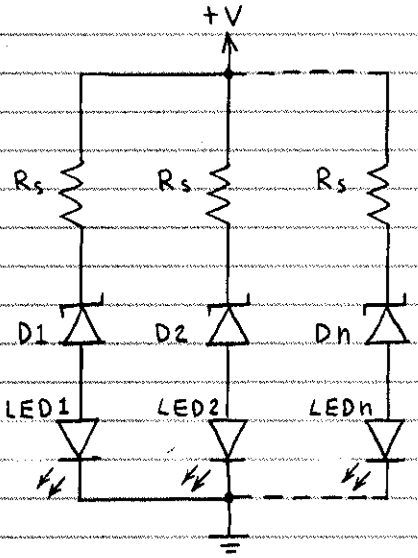
+V
Rs
Rs
Rs
D1
D2
Dn
LED1
LED2
LEDn

☐ FLASHER LED + RELAY

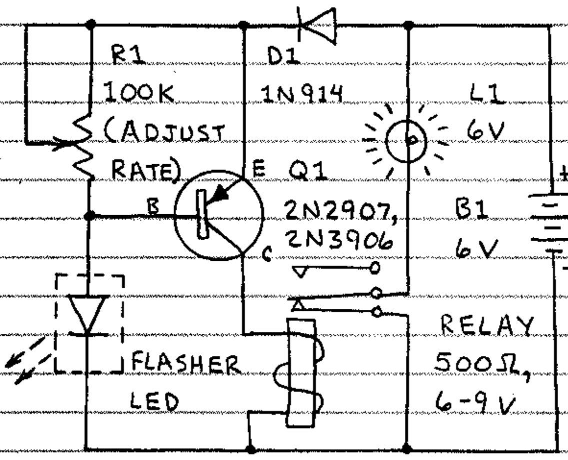
R1
100K
(ADJUST
RATE)
D1
1N914
L1
6V
E
Q1
B
C
2N2907,
2N3906
B1
6V
+
−
RELAY
500Ω,
6-9V
FLASHER
LED

THE FLASHER LED INCLUDES A BUILT-IN IC THAT FLASHES THE LED SEVERAL TIMES EACH SECOND. THIS CIRCUIT SHOWS HOW TO "TAP" THIS FLASH RATE (VIA Q1) TO FORM AN ULTRA-SIMPLE PULSE GENERATOR THAT ACTUATES A RELAY AND, IN TURN, A LAMP. D1 IS NEEDED TO KEEP THE VOLTAGE TO THE FLASHER LED NEAR 5 VOLTS.

SEMICONDUCTOR LIGHT DETECTOR CIRCUITS

☐ LIGHT METER CIRCUITS

1. PHOTORESISTOR

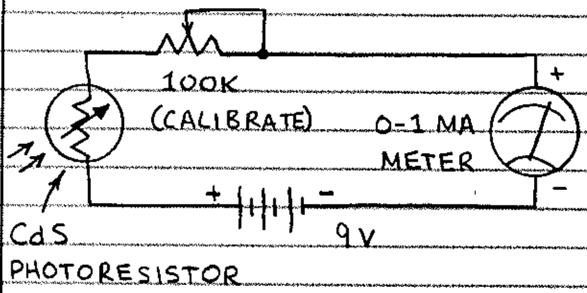
100K
(CALIBRATE)
0-1 MA
METER
+
-
+
-
9V
CdS
PHOTORESISTOR

2. SOLAR CELL

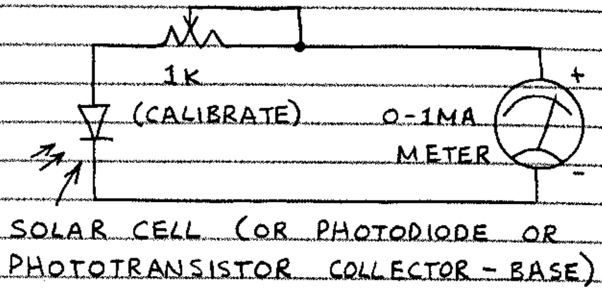
1K
(CALIBRATE)
0-1MA
METER
+
-
SOLAR CELL (OR PHOTODIODE OR
PHOTOTRANSISTOR COLLECTOR - BASE)

☐ LIGHT ACTUATED RELAY CIRCUITS

1. PHOTORESISTOR

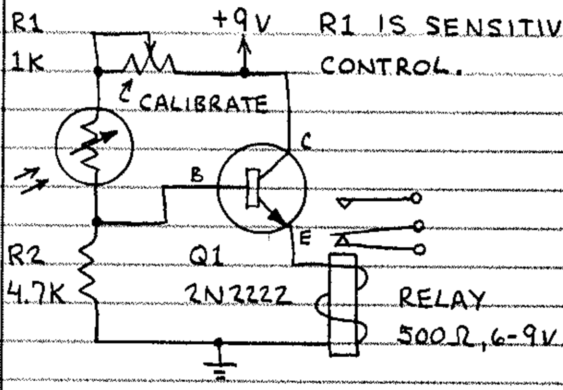
R1
1K
+9V
CALIBRATE
C
B
E
R2
4.7K
Q1
2N2222
RELAY
500Ω, 6-9V

R1 IS SENSITIVITY CONTROL.

RELAY REMAINS ACTUATED FOR BRIEF TIME AFTER LIGHT REMOVED.

2. PHOTOTRANSISTOR

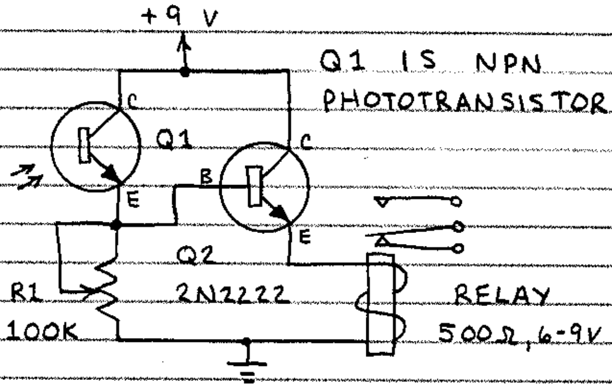
+9 V
C
Q1
E
C
B
E
Q2
2N2222
R1
100K
RELAY
500Ω, 6-9V

Q1 IS NPN PHOTOTRANSISTOR.

RESPONDS FASTER THAN PHOTORESISTOR CIRCUIT. NO DELAY WHEN LIGHT IS REMOVED.

☐ LIGHT DEACTUATED RELAY CIRCUITS

1. PHOTORESISTOR

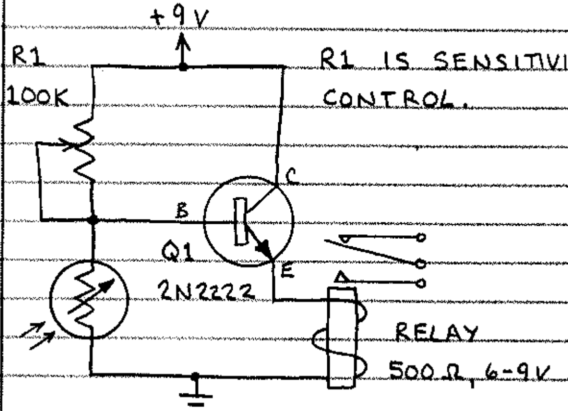
+9V
R1
100K
C
B
E
Q1
2N2222
RELAY
500Ω, 6-9V

R1 IS SENSITIVITY CONTROL.

RELAY IS ACTUATED ONLY WHEN THE PHOTORESISTOR IS DARK.

2. PHOTO TRANSISTOR

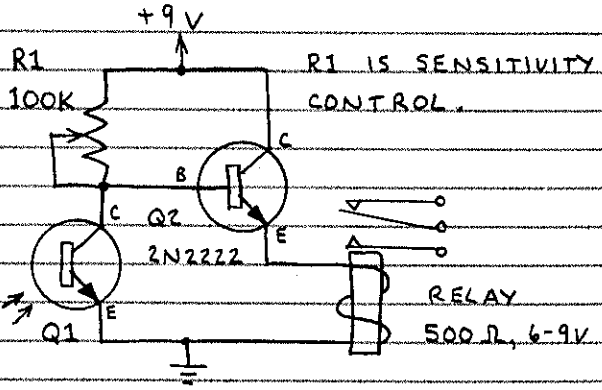
+9V
R1
100K
C
B
E
C
Q2
2N2222
E
Q1
RELAY
500Ω, 6-9V

R1 IS SENSITIVITY CONTROL.

RELAY IS ACTUATED WHEN Q1 IS DARK. LIGHT ON Q1 DEACTUATES RELAY. CALIBRATE WITH R1.

☐ AUDIBLE LIGHT PROBE

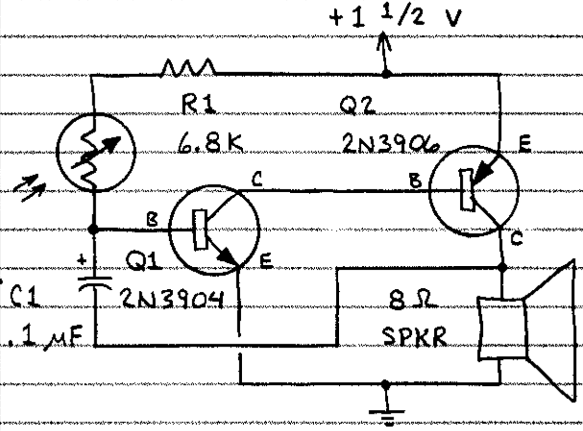
+1 1/2 V
R1
6.8K
Q2
2N3906
E
C
B
B
C
Q1
E
C1
.1 μF
+
2N3904
8 Ω
SPKR

THIS IS AMONG THE MOST ENTERTAINING CIRCUITS IN THIS BOOK. VARIOUS PNP AND NPN TRANSISTORS CAN BE USED FOR Q1 AND Q2. THE TONE FROM THE SPEAKER INCREASES IN FREQUENCY AS THE INTENSITY OF LIGHT ON THE PHOTORESISTOR INCREASES. VERY SENSITIVE! TRY THIS: OPERATE CIRCUIT IN A DARK ROOM UNTIL TONE SLOWS TO A SERIES OF CLICKS. THEN SHINE BEAM FROM FLASHLIGHT ON THE PHOTORESISTOR...

☐ SOLAR CELL BATTERY CHARGER

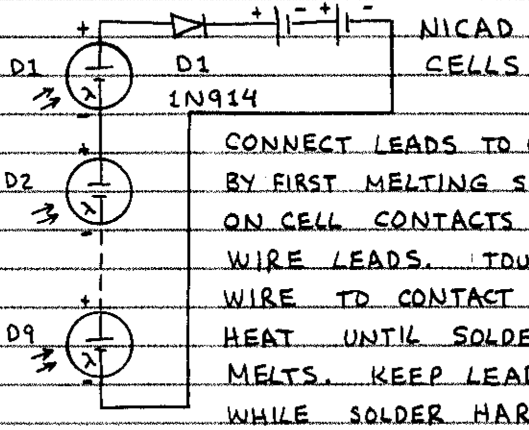
NICAD
CELLS
D1
1N914
D1
D2
D9
+
-

CONNECT LEADS TO CELLS BY FIRST MELTING SOLDER ON CELL CONTACTS AND WIRE LEADS. TOUCH WIRE TO CONTACT AND HEAT UNTIL SOLDER MELTS. KEEP LEAD STILL WHILE SOLDER HARDENS.

USE NINE CELLS TO CHARGE TWO NICADS. CURRENT FROM SOLAR CELLS MUST NOT EXCEED MAXIMUM CHARGING RATE FOR NICADS! YOU CAN MONITOR CURRENT BY CONNECTING MULTIMETER BETWEEN NICADS AND D1. INSERT SERIES RESISTOR OR REMOVE SOLAR CELL TO REDUCE CURRENT. D1 KEEPS NICADS FROM DISCHARGING THROUGH SOLAR CELLS (WHEN DARK). SOLAR CELLS ARE FRAGILE. SOLDER AND MOUNT WITH CARE.

☐ LIGHT ACTUATED LATCHING CIRCUITS

1. RELAY

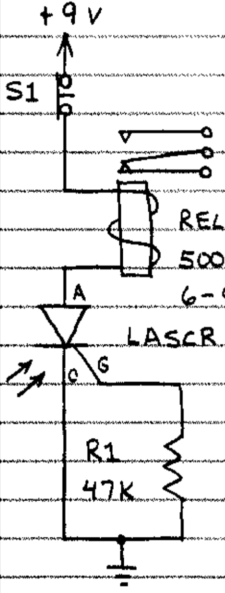
+9V
S1
RELAY
500 Ω,
6-9V
A
LASCR
C
G
R1
47K

2. LED (OR LAMP)

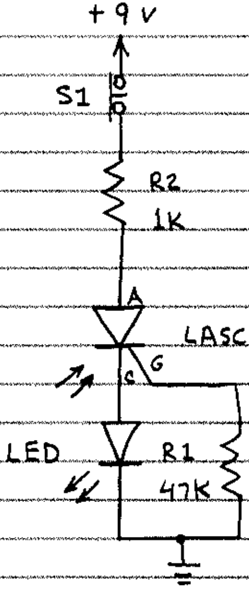
+9V
S1
R2
1K
A
LASCR
C
G
LED
R1
47K

3. BUZZER

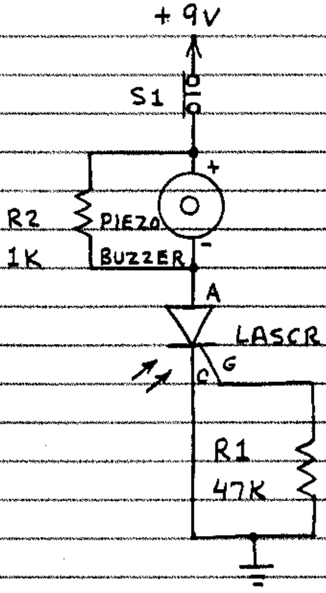
+9V
S1
+
R2
1K
PIEZO
BUZZER
-
A
LASCR
C
G
R1
47K

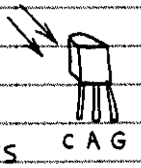
C A G
LASCR
(MAY
VARY)

THESE CIRCUITS ARE ACTUATED BY LIGHT. OPEN S1 TO SWITCH LASCR OFF. SOME LASCRs ARE MORE LIGHT SENSITIVE THAN OTHERS. MOST WILL TRIGGER IN RESPONSE TO LIGHT FROM CAMERA STROBE (XENON FLASH UNIT).

DIGITAL IC CIRCUITS

DIGITAL ICs ARE VERY EASY TO USE. HERE'S A SELECTION OF TTL AND CMOS CIRCUITS:

TTL CIRCUITS

☐ OPERATING REQUIREMENTS

1. POWER SUPPLY MUST NOT EXCEED 5.25 VOLTS. SEE PAGE 125 OR USE:

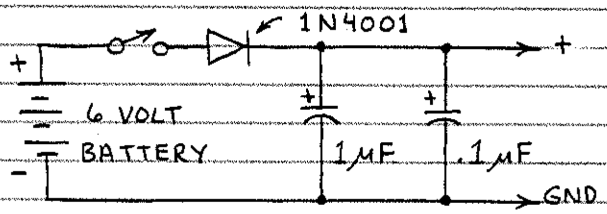

1N4001
+
6 VOLT
BATTERY
−
1 μF
.1 μF
+
GND

2. INPUTS MUST NOT EXCEED +5.25 V.
3. INPUTS SHOULD ALWAYS GO SOMEWHERE (NOT LEFT "FLOATING").
4. FORCE OUTPUTS OF UNUSED GATES "H" TO SAVE POWER. (EXAMPLE: UNUSED "NAND"—MAKE ONE INPUT H.)
5. AVOID LONG WIRES IN CIRCUITS.
6. CONNECT 1 TO 10 μF CAPACITOR ACROSS POWER LEADS WHERE THEY ENTER CIRCUIT.
7. CONNECT 0.1 μF CAPACITOR ACROSS POWER PINS OF EACH TTL CHIP IN MULTI-CHIP CIRCUITS.
8. REMEMBER, TTL USES MUCH MORE CURRENT THAN LS OR CMOS.

☐ "D" FLIP-FLOP

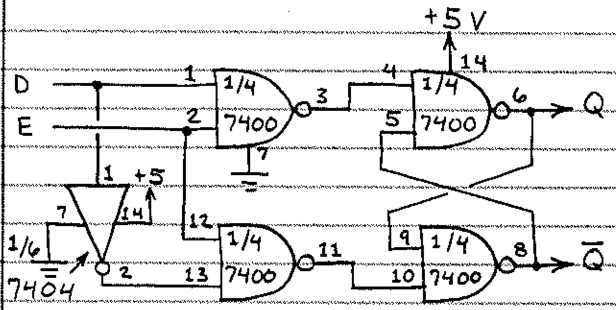

+5 V
D
E
1/4 7400
1/4 7400
1/4 7400
1/4 7400
1/6 7404
Q
Q̄
1
2
3
4
5
6
7
8
9
10
11
12
13
14
+5

Q = D WHEN ENABLE (E) IS HIGH. NO CHANGE WHEN E IS LOW.

☐ CLOCKED "RS" FLIP-FLOP

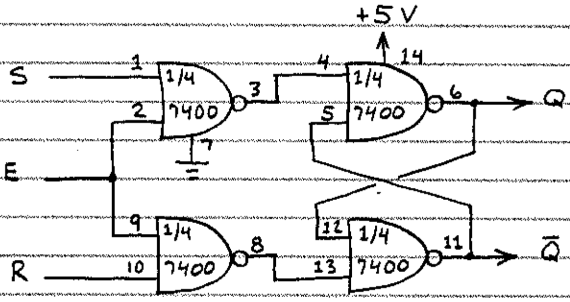

+5 V
S
E
R
1/4 7400
1/4 7400
1/4 7400
1/4 7400
Q
Q̄
1
2
3
4
5
6
7
8
9
10
11
12
13
14

FUNCTIONS AS "RS" FLIP-FLOP WHEN ENABLE (E) IS HIGH.

☐ DUAL LED FLASHER

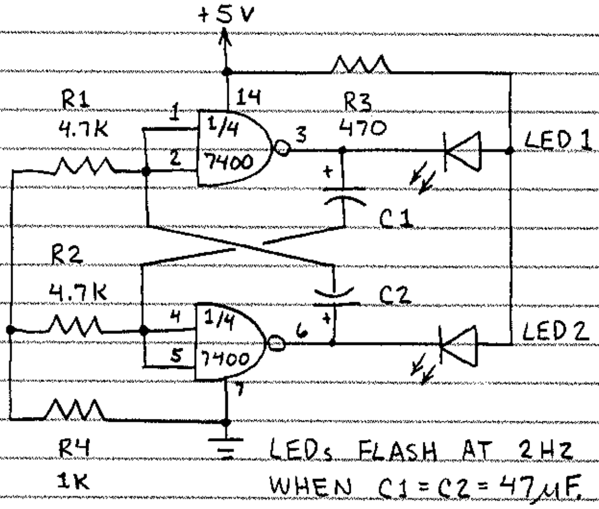

+5 V
R1
4.7K
R2
4.7K
R3
470
R4
1K
1/4 7400
1/4 7400
LED 1
LED 2
C1
C2
1
2
3
4
5
6
7
14

LEDs FLASH AT 2HZ WHEN C1 = C2 = 47 μF.

☐ TONE GENERATOR

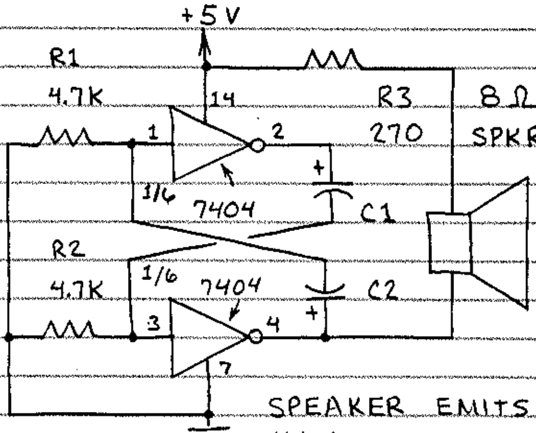

+5 V
R1
4.7K
R2
4.7K
R3
270
8 Ω
SPKR
1/6 7404
1/6 7404
C1
C2
1
2
3
4
7
14

SPEAKER EMITS 4 KHZ TONE WHEN C1 = C2 = 0.1 μF.

☐ 0 TO 9 SECOND (OR MINUTE) TIMER

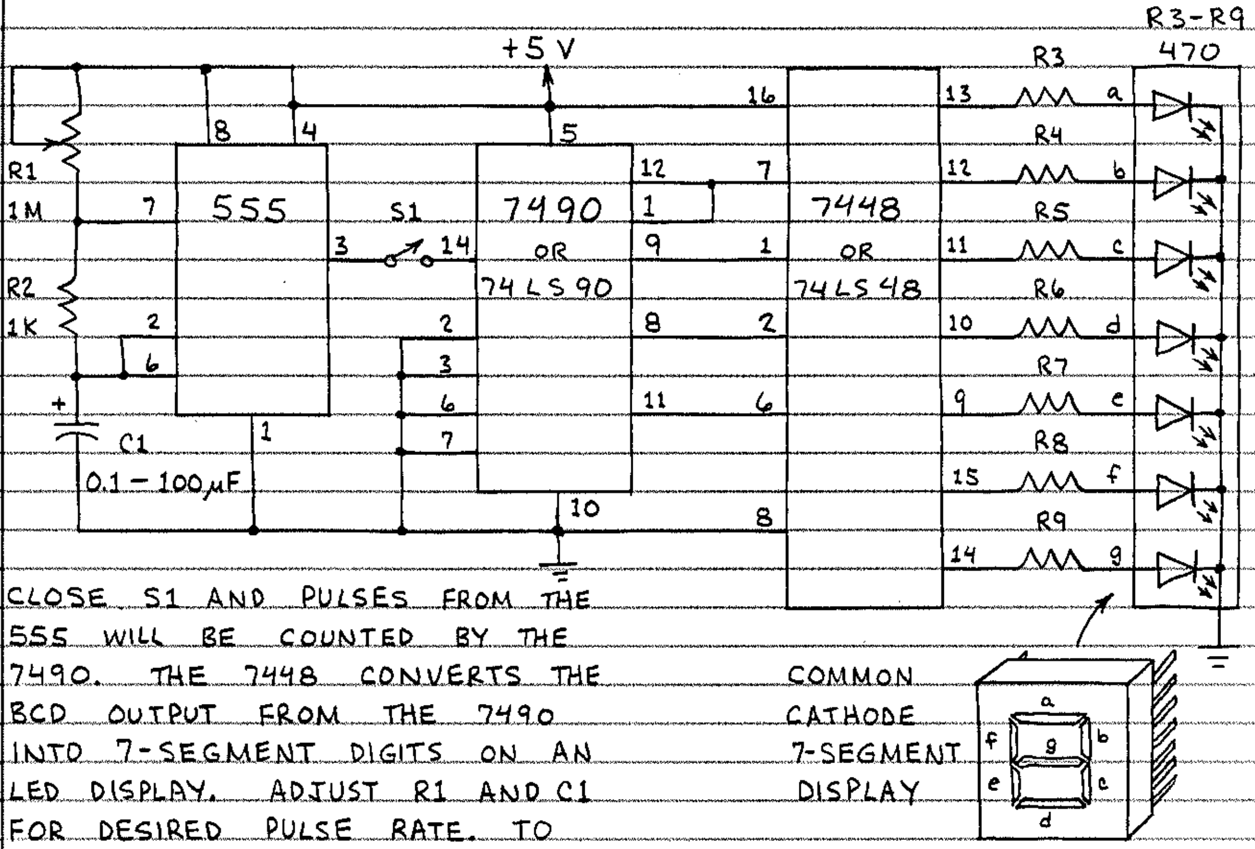
+5 V
R3-R9
470
R1
1M
R2
1K
C1
0.1 – 100 µF
555
S1
7490
OR
74LS90
7448
OR
74LS48
R3
R4
R5
R6
R7
R8
R9
a
b
c
d
e
f
g
8
4
7
2
6
1
3
14
5
12
1
9
8
11
10
16
7
1
2
6
8
13
12
11
10
9
15
14
COMMON
CATHODE
7-SEGMENT
DISPLAY

CLOSE S1 AND PULSES FROM THE 555 WILL BE COUNTED BY THE 7490. THE 7448 CONVERTS THE BCD OUTPUT FROM THE 7490 INTO 7-SEGMENT DIGITS ON AN LED DISPLAY. ADJUST R1 AND C1 FOR DESIRED PULSE RATE. TO ADD ADDITIONAL DIGIT:

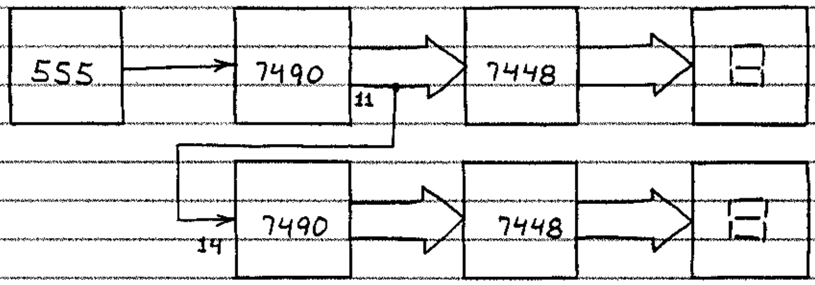
555
7490
7448
11
7490
7448
14

HINT: THIS BASIC CIRCUIT WILL COUNT PULSES FROM OTHER CIRCUITS. ELIMINATE 555 AND APPLY PULSES (5 VOLTS MAXIMUM) TO 7490.

☐ DIVIDE-BY-5 COUNTER

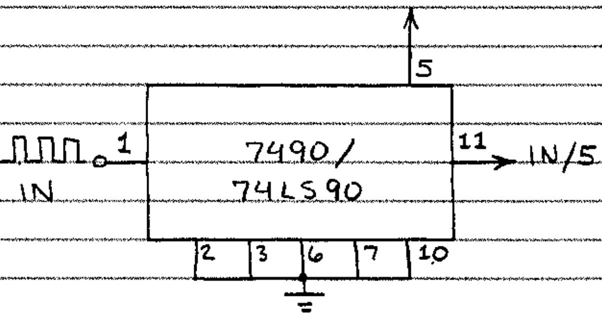
5
1
IN
7490/
74LS90
11
IN/5
2
3
6
7
10

DIVIDES INCOMING "TRAIN" OF PULSES BY 5. OK TO USE AT INPUT OF CIRCUIT ABOVE.

☐ DIVIDE-BY-10 COUNTER

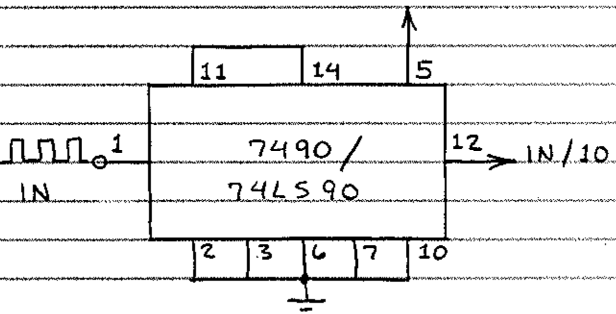
11
14
5
1
IN
7490/
74LS90
12
IN/10
2
3
6
7
10

DIVIDES INCOMING "TRAIN" OF PULSES BY 10. OK TO USE AT INPUT OF CIRCUIT ABOVE.

CMOS CIRCUITS

☐ OPERATING REQUIREMENTS

1. THE POSITIVE POWER TO A CMOS CHIP (V_{DD}) CAN RANGE FROM +3 TO +15 (OR +18) VOLTS. USE POWER SUPPLIES ON PAGE 125 OR A BATTERY (BEST):

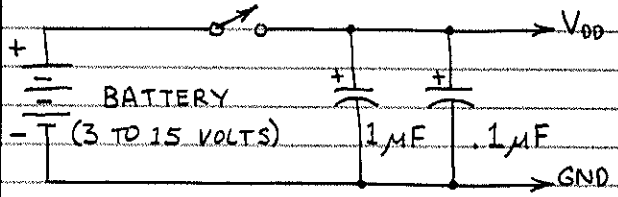
V_{DD}
+
BATTERY
− (3 TO 15 VOLTS)
+
+
1μF
.1μF
GND

2. INPUTS MUST NOT EXCEED V_{DD}.
3. ALL UNUSED INPUTS MUST GO TO V_{DD} OR GND (⏚).
4. NEVER APPLY AN INPUT SIGNAL TO AN UNPOWERED CMOS CIRCUIT.

☐ HANDLING PRECAUTIONS

1. PLACE CMOS ICs PINS DOWN ON AN ALUMINUM FOIL SHEET OR TRAY WHEN THEY ARE NOT IN A CIRCUIT OR PROPERLY STORED.
2. NEVER STORE CMOS ICs IN NONCONDUCTIVE PLASTIC "SNOW," TRAYS, BAGS OR FOAM. PLUG THEM IN CONDUCTIVE FOAM OR FOAMED PLASTIC WRAPPED IN ALUMINUM FOIL.
3. AVOID USING AN AC LINE POWERED IRON TO SOLDER PINS OF CMOS ICs. USE IC SOCKETS, WIRE-WRAP OR BATTERY POWERED IRON.
4. DRAIN STATIC CHARGE ON YOUR BODY BY TOUCHING GROUNDED OBJECT.

☐ BOUNCELESS SWITCH

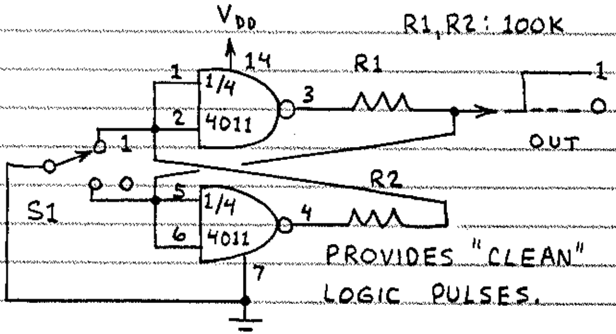
V_{DD}
R1, R2: 100K
14
R1
1
2
1/4
4011
3
1
OUT
1
S1
R2
5
6
1/4
4011
4
7

PROVIDES "CLEAN" LOGIC PULSES.

☐ SINGLE LED "GATED" FLASHER

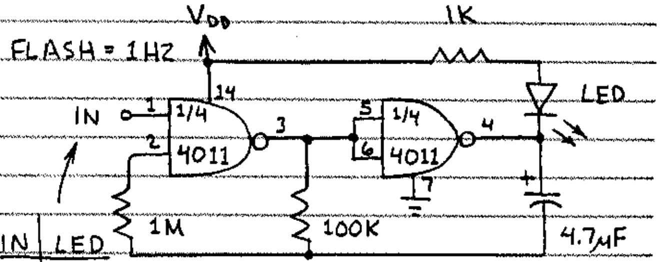
V_{DD}
1K
FLASH = 1HZ
14
LED
IN
1
2
1/4
4011
3
5
6
1/4
4011
4
7
1M
100K
+
4.7μF

IN	LED
L	ON
H	FLASH

(ADD INVERTER BETWEEN PIN 4 AND LED FOR L = OFF.)

☐ "ONE-SHOT" TOUCH SWITCH

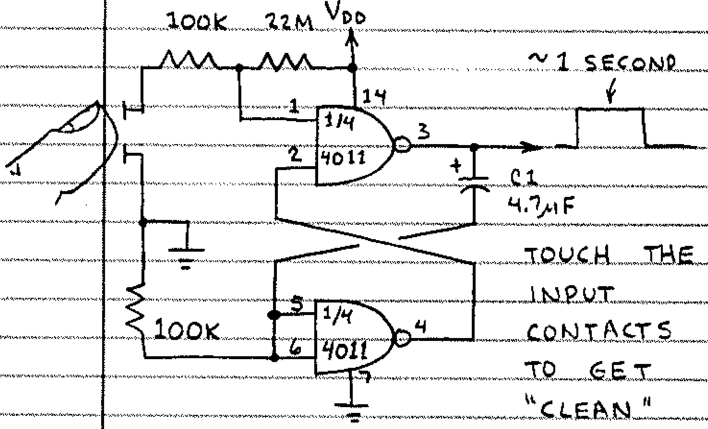
100K
22M
V_{DD}
14
1
2
1/4
4011
3
~1 SECOND
+
C1
4.7μF
100K
5
6
1/4
4011
4
7

TOUCH THE INPUT CONTACTS TO GET "CLEAN" OUTPUT PULSE.

C1 CONTROLS LENGTH OF OUTPUT PULSE.

☐ DUAL LED FLASHER

LEDs FLASH ALTERNATELY AT ~1 HZ.

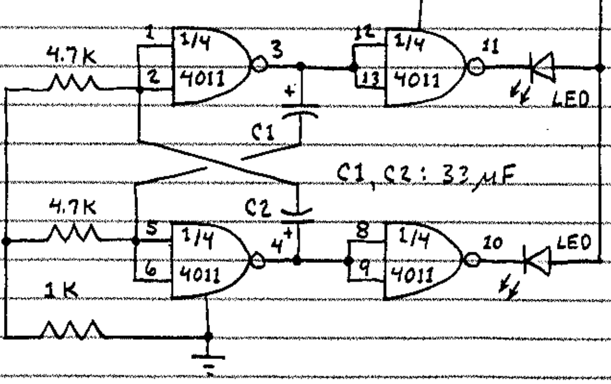
V_{DD}
1K
1
2
1/4
4011
3
12
13
1/4
4011
11
LED
4.7K
+
C1
C1, C2: 33μF
C2
4.7K
5
6
1/4
4011
4
+
8
9
1/4
4011
10
LED
1K

CHANGE C1 AND C2 TO CHANGE RATE.

☐ MINI RADIO TRANSMITTER

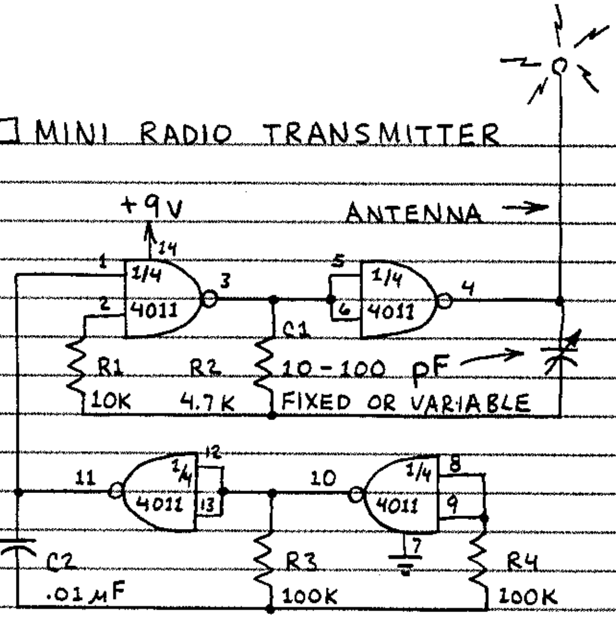

+9V
ANTENNA →
14
1
2
1/4 4011
3
5
6
1/4 4011
4
C1
R1 10K
R2 4.7K
10-100 pF
FIXED OR VARIABLE
11
1/4 4011
12
13
10
1/4 4011
8
9
7
C2 .01μF
R3 100K
R4 100K

SENDS TONE TO NEARBY RADIO. TUNE RADIO AND/OR C1 TO HEAR TONE. USE INSULATED TOOL TO TUNE C1. C2 CONTROLS TONE FREQUENCY. USE 1-2 FEET OF WIRE FOR ANTENNA.

☐ "GATED" TONE GENERATOR

*USE AT LEAST 180Ω WHEN V_{DD} = 9V.

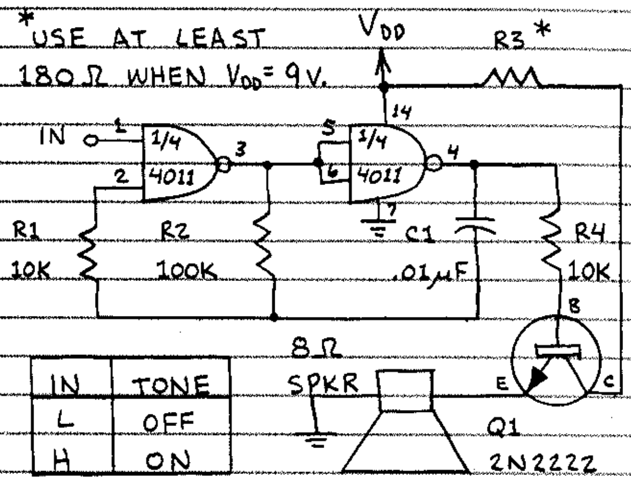

V_{DD}
R3*
IN
1
2
1/4 4011
3
5
6
1/4 4011
4
14
7
R1 10K
R2 100K
C1 .01μF
R4 10K
8Ω SPKR
B
E
C
Q1 2N2222

IN	TONE
L	OFF
H	ON

VALUES SHOWN GIVE 1.3 KHZ TONE. C1 AND R2 CONTROL TONE FREQUENCY. R3 CONTROLS VOLUME. INPUT CAN BE SWITCH (V_{DD} / GND) OR LOGIC SIGNAL.

☐ STANDARD TOUCH SWITCH

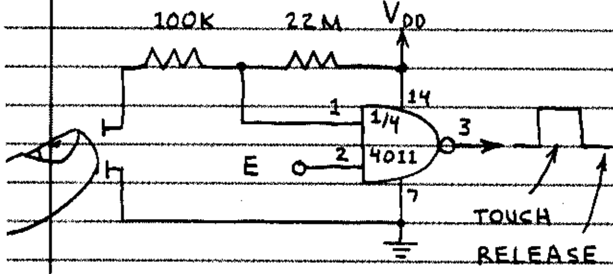

100K
22M
V_{DD}
14
1
2
1/4 4011
3
E
7
TOUCH
RELEASE

WHEN THE ENABLE INPUT (E) IS HIGH, THE CIRCUIT RESPONDS.

☐ LINEAR ×10 AMPLIFIER

R1: 1M

R2: 10M

GAIN = R2/R1

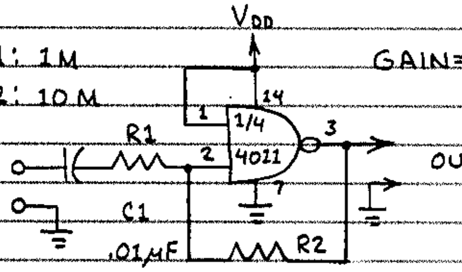

V_{DD}
14
1
2
1/4 4011
3
7
R1
C1 .01μF
R2
OUT

THIS IS A NON-DIGITAL ROLE FOR A DIGITAL GATE.

☐ FREQUENCY GENERATOR

R1 CONTROLS FREQUENCY.

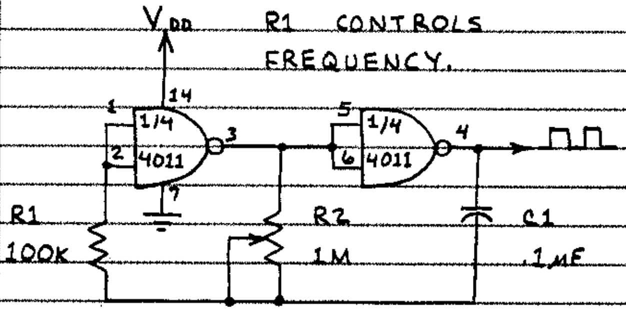

V_{DD}
14
1
2
1/4 4011
3
7
5
6
1/4 4011
4
R1 100K
R2 1M
C1 .1μF

USE TO SUPPLY "CLOCK" PULSES TO OTHER CMOS CIRCUITS.

☐ LAMP FLASHER

RATE IS ~1 HZ AS SHOWN.

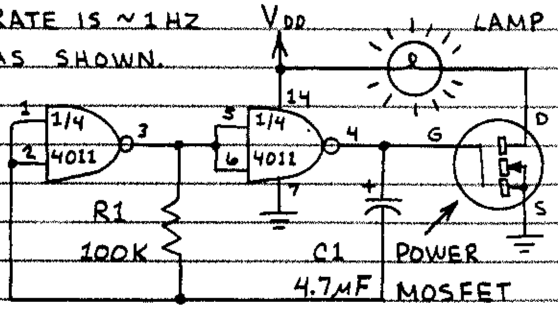

V_{DD}
LAMP
14
1
2
1/4 4011
3
5
6
1/4 4011
4
7
R1 100K
C1 4.7μF
G
D
S
POWER MOSFET

C1 AND R1 CONTROL RATE. OK TO USE SEPARATE SUPPLY FOR LAMP.

☐ ELECTRONIC COIN TOSSER

PRESS S1 MOMENTARILY.
TYPICAL TEST RESULTS:

TRIAL	HEADS	TAILS
1	23	27
2	25	25

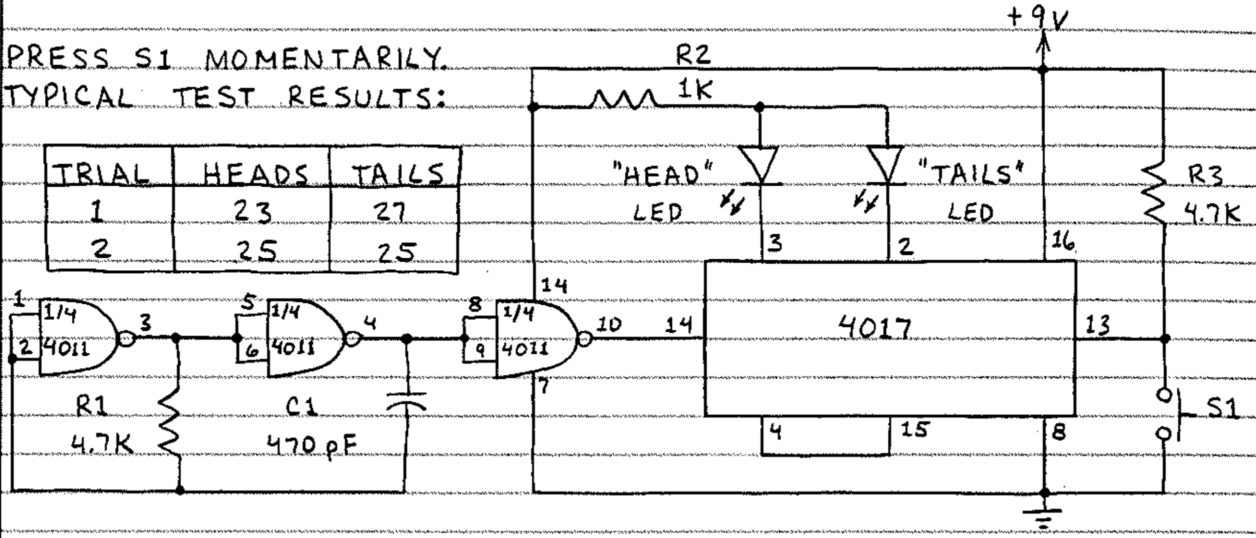
+9V
R2
1K
"HEAD" LED
"TAILS" LED
R3
4.7K
3
2
16
14
1
2
1/4 4011
3
5
6
1/4 4011
4
8
9
1/4 4011
10
7
14
4017
13
S1
4
15
8
R1
4.7K
C1
470 pF

☐ RANDOM NUMBER GENERATOR

HINT:
INCREASE C1 TO 4.7 µF AND LEAVE S1 CLOSED TO MAKE SEQUENTIAL FLASHER.

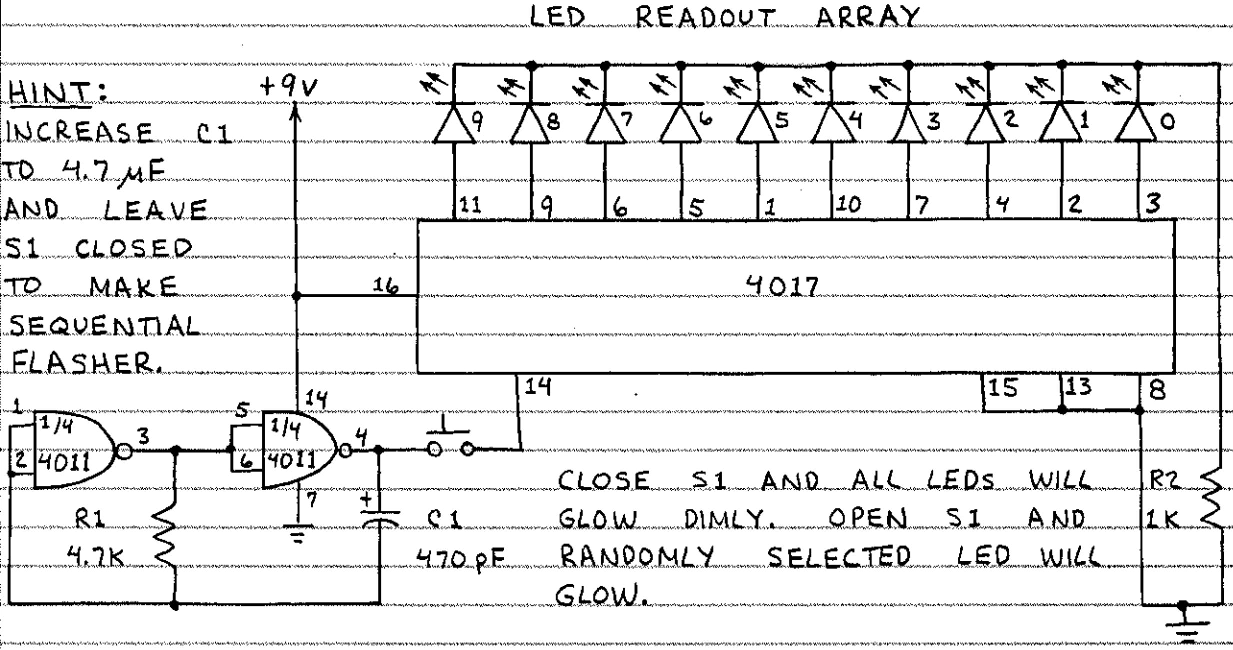
LED READOUT ARRAY
9
8
7
6
5
4
3
2
1
0
11
9
6
5
1
10
7
4
2
3
+9V
16
4017
14
15
13
8
1
2
1/4 4011
3
5
6
1/4 4011
14
4
7
R1
4.7K
C1
470 pF
R2
1K

CLOSE S1 AND ALL LEDs WILL GLOW DIMLY. OPEN S1 AND RANDOMLY SELECTED LED WILL GLOW.

☐ 1-OF-4 SEQUENCER

USE TO ACTUATE RELAYS, ETC.

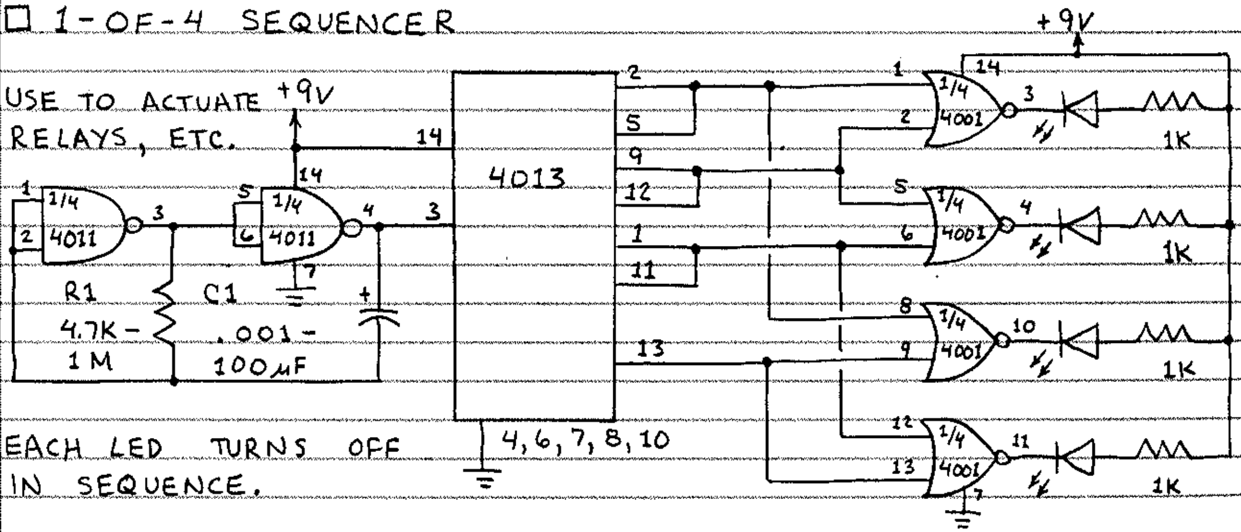
+9V
+9V
14
4013
3
2
5
9
12
1
11
13
4, 6, 7, 8, 10
1
2
1/4 4011
3
5
6
1/4 4011
14
4
7
R1
4.7K -
1 M
C1
.001 -
100 µF
14
1
2
1/4 4001
3
5
6
1/4 4001
4
8
9
1/4 4001
10
12
13
1/4 4001
11
7
1K
1K
1K
1K

EACH LED TURNS OFF IN SEQUENCE.

☐ MODEL RAILROAD CROSSING FLASHER LIGHTS

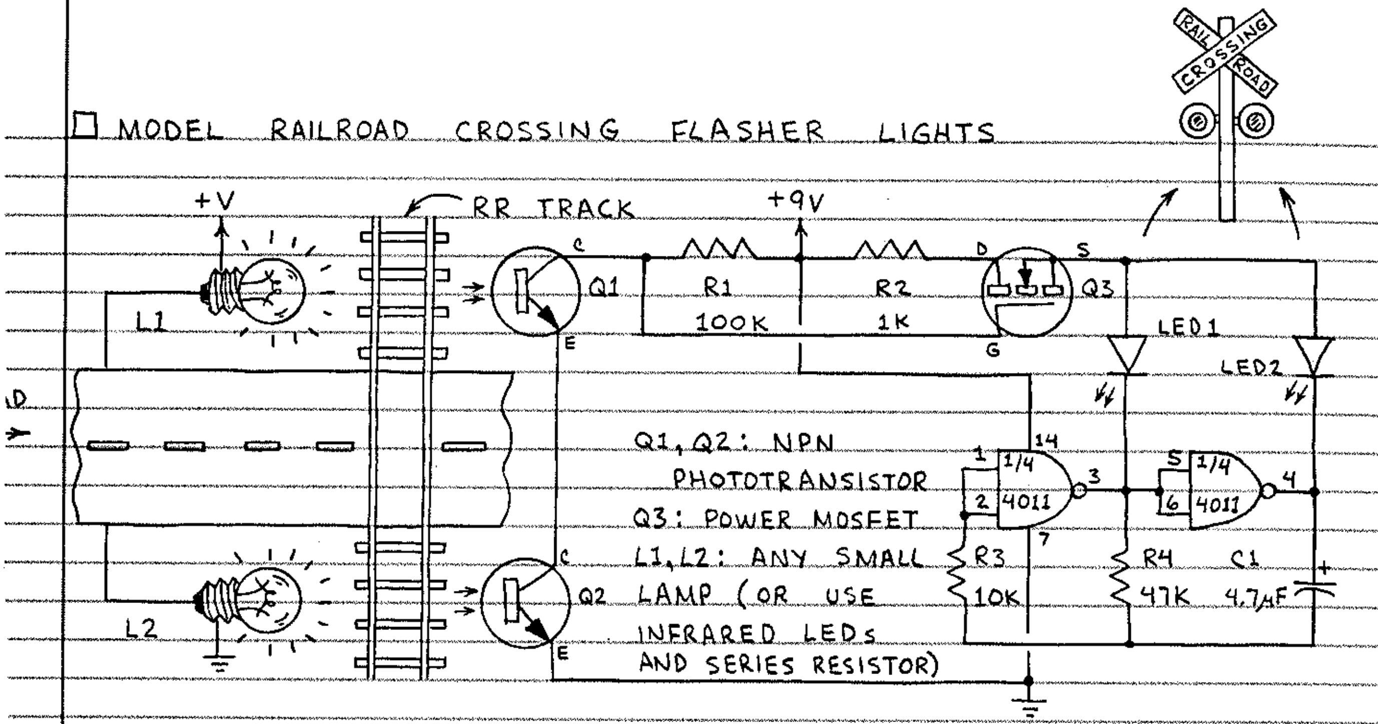
RAIL ROAD CROSSING
+V
RR TRACK
+9V
L1
L2
Q1
Q2
C
E
R1
100K
R2
1K
D
S
G
Q3
LED1
LED2
1/4
4011
1
2
3
14
7
5
6
4
R3
10K
R4
47K
C1
4.7µF
+

Q1, Q2: NPN PHOTOTRANSISTOR
Q3: POWER MOSFET
L1, L2: ANY SMALL LAMP (OR USE INFRARED LEDs AND SERIES RESISTOR)

LEDS FLASH ALTERNATELY WHEN TRAIN BREAKS LIGHT BEAM TO EITHER Q1 OR Q2. LEDS CONTINUE FLASHING UNTIL TRAIN PASSES. SHIELD Q1 AND Q2 FROM ROOM LIGHTS WITH 1" HEAT SHRINK TUBING.

☐ PROGRAMMABLE GAIN OPERATIONAL AMPLIFIER

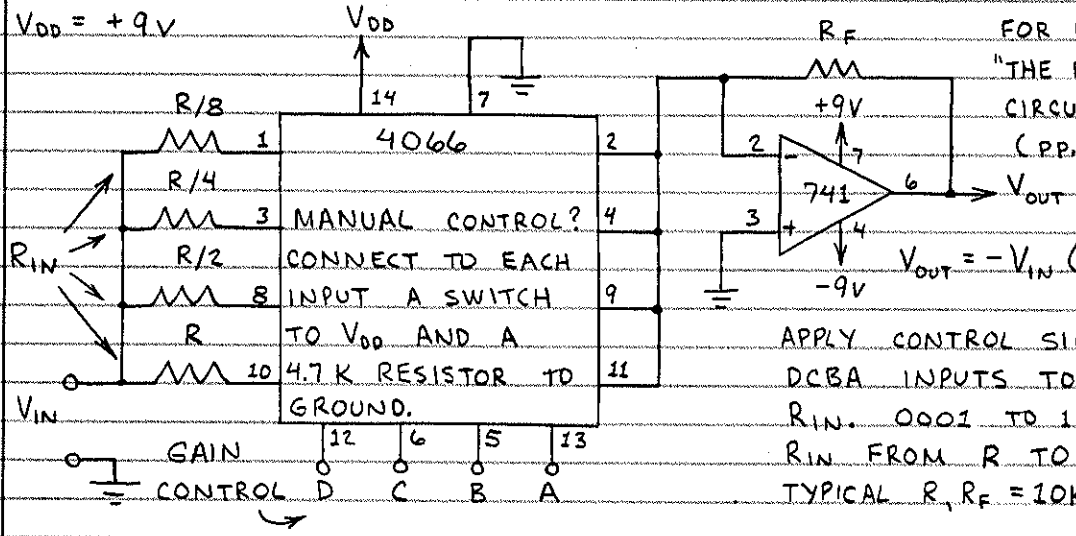
V_{DD} = +9V
V_{DD}
14
7
4066
R/8
R/4
R/2
R
1
3
8
10
2
4
9
11
12
6
5
13
R_{IN}
V_{IN}
GAIN CONTROL
D
C
B
A
R_F
+9V
-9V
741
2
3
7
4
6
V_{OUT}

MANUAL CONTROL? CONNECT TO EACH INPUT A SWITCH TO V_{DD} AND A 4.7K RESISTOR TO GROUND.

$V_{OUT} = -V_{IN}(R_F/R_{IN})$

FOR DETAILS SEE "THE FORREST MIMS CIRCUIT SCRAPBOOK" (PP. 21-23).

APPLY CONTROL SIGNALS AT DCBA INPUTS TO ALTER R_{IN}. 0001 TO 1111 VARIES R_{IN} FROM R TO R/15. TYPICAL R, R_F = 10K.

☐ ALL ON — ALL OFF SEQUENCER

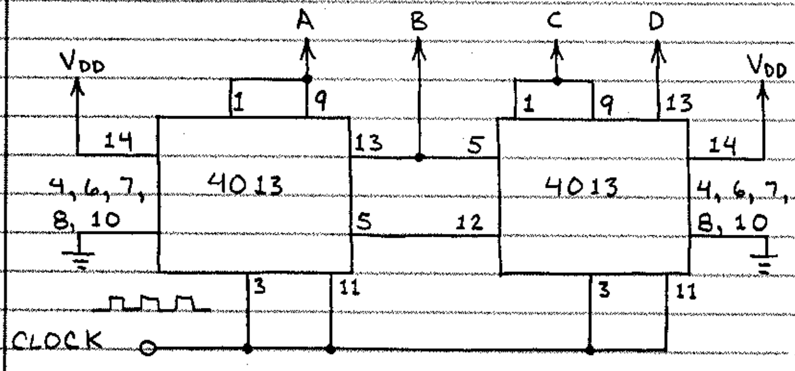
A
B
C
D
V_{DD}
V_{DD}
4013
4013
1
9
13
5
14
4, 6, 7, 8, 10
5
12
3
11
1
9
13
14
4, 6, 7, 8, 10
3
11
CLOCK

ALL OUTPUTS GO LOW, THEN HIGH, IN SEQUENCE (A...B...C...D...A...B...ETC.). USE WITH LEDs FOR EYE-CATCHING DISPLAY. FOR "BUCKET-BRIGADE" OPERATION, CONNECT PIN 5 OF FIRST 4013 TO PIN 13 (NOT 12) OF SECOND 4013.

LINEAR IC CIRCUITS

YOU CAN MAKE AN AMAZING VARIETY OF CIRCUITS WITH LINEAR ICs. HERE ARE A FEW OF THE MANY POSSIBILITIES:

OPERATIONAL AMPLIFIER (OP-AMP) CIRCUITS

☐ AUDIO AMPLIFIER

741 IS PREAMPLIFIER (R2 SETS GAIN)
386 IS POWER AMPLIFIER

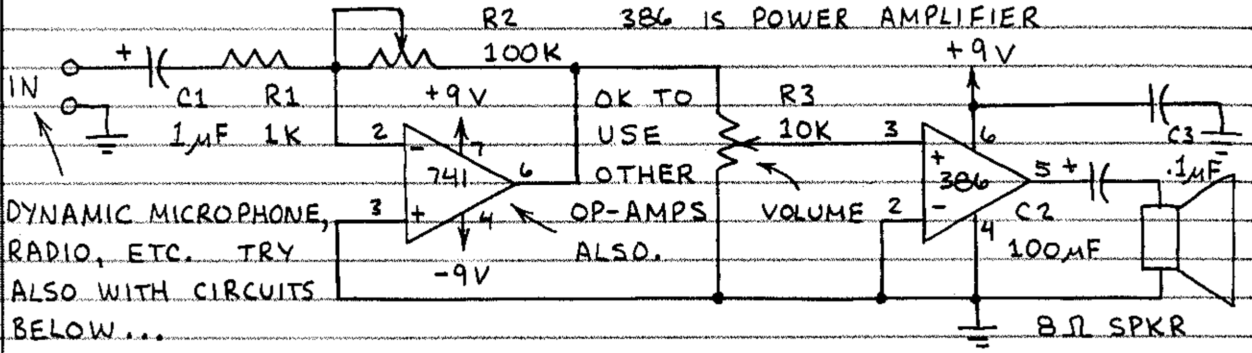
IN
+
C1
1µF
R1
1K
R2
100K
+9V
2
7
741
6
3
4
-9V
OK TO USE OTHER OP-AMPS ALSO.
R3
10K
VOLUME
+9V
3
6
386
5
2
4
C2
100µF
C3
.1µF
8Ω SPKR

DYNAMIC MICROPHONE, RADIO, ETC. TRY ALSO WITH CIRCUITS BELOW...

☐ MIXER

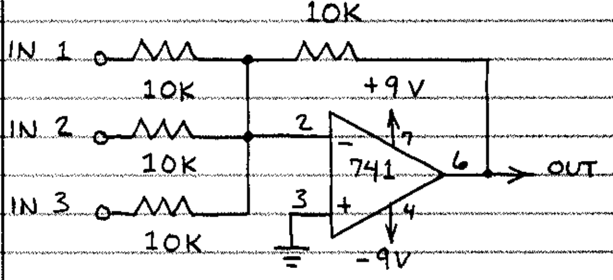
IN 1
IN 2
IN 3
10K
10K
10K
10K
+9V
2
7
741
6
3
4
-9V
OUT

SEVERAL SIMULTANEOUS INPUTS.
USE WITH AUDIO AMPLIFIER ABOVE.

☐ DIFFERENCE AMPLIFIER

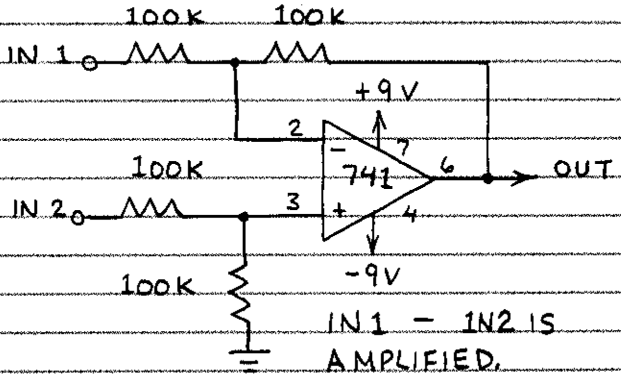
IN 1
100K
100K
+9V
2
7
741
6
IN 2
100K
3
4
-9V
100K
OUT

IN1 − IN2 IS AMPLIFIED.

☐ LIGHTWAVE VOICE TRANSMITTER*

MIC IS CRYSTAL OR ELECTRET* MICROPHONE. LED IS INFRARED TYPE. USE LENS TO FOCUS LIGHT FROM LED INTO NARROW BEAM. TO TEST, PLACE RADIO EARPHONE NEAR MICROPHONE. ADJUST R1 AND R6 FOR BEST RECEIVER SOUND.

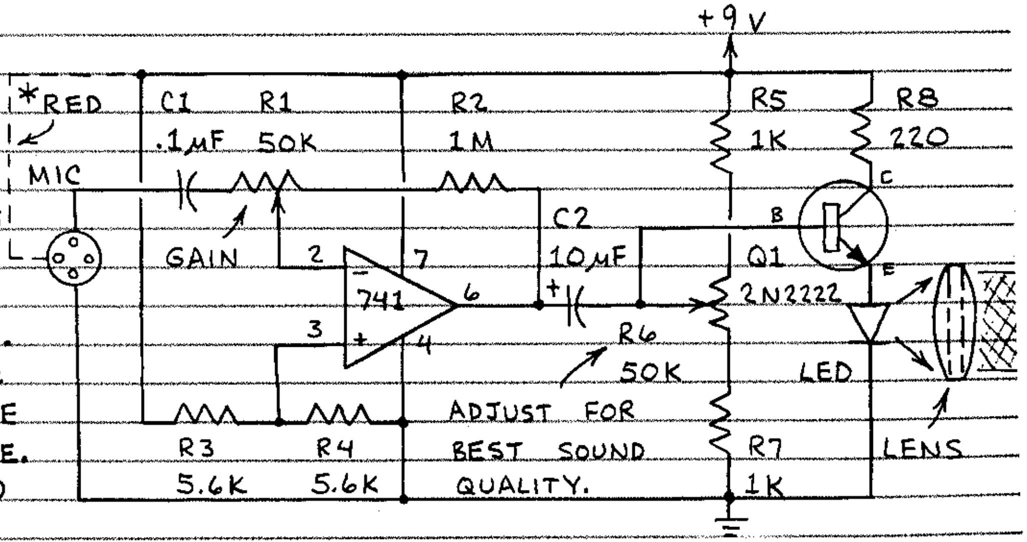
+9V
*RED
MIC
C1
.1µF
R1
50K
GAIN
R2
1M
2
7
741
6
3
4
R3
5.6K
R4
5.6K
C2
10µF
+
R6
50K
ADJUST FOR BEST SOUND QUALITY.
R5
1K
R7
1K
R8
220
B
C
E
Q1
2N2222
LED
LENS

*SEE "THE FORREST MIMS CIRCUIT SCRAPBOOK" (PP. 35-42).

☐ PERCUSSION SYNTHESIZER (BELL, DRUM, ETC.)

SET R2 AND R3 AT CENTER POSITIONS. ADJUST R4 TO JUST PAST POINT WHERE OSCILLATION STOPS. CLOSE S1. ADJUST R2, R3 AND R4 AND TRY AGAIN. IN TIME YOU'LL HAVE AN ELECTRONIC BELL, DRUM, BONGO, ETC.

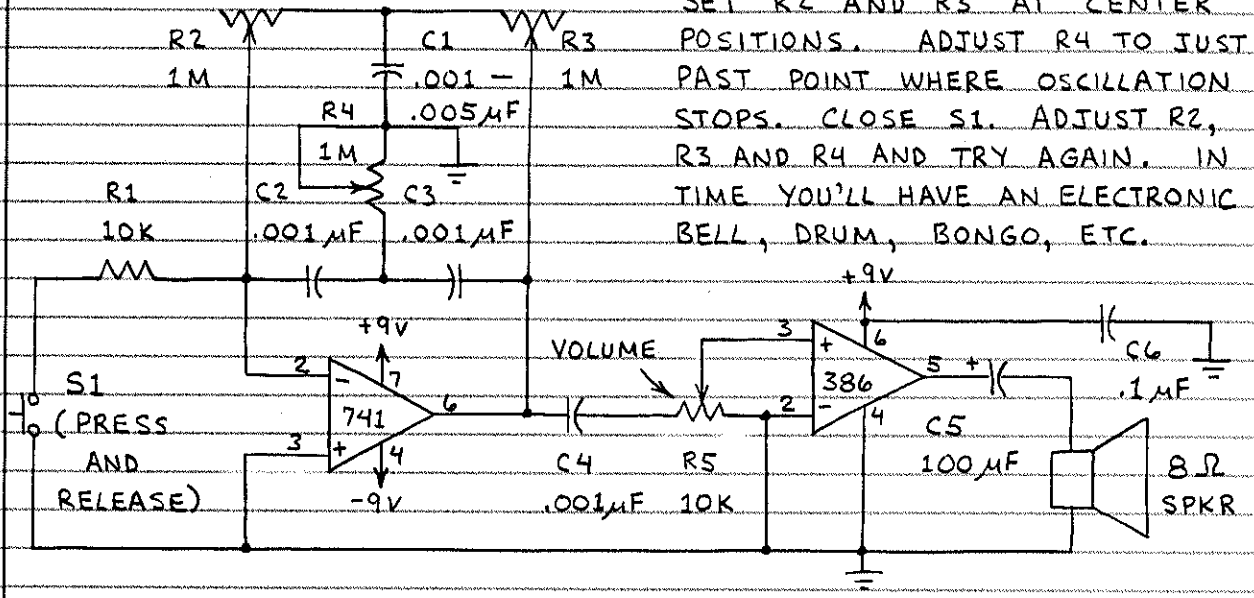
R2
1M
C1
.001 – .005μF
R3
1M
R4
1M
R1
10K
C2
.001μF
C3
.001μF
+9V
S1
(PRESS AND RELEASE)
2
3
7
4
6
741
-9V
VOLUME
C4
.001μF
R5
10K
+9V
3
2
6
4
5
386
+
C6
.1μF
C5
100μF
8Ω
SPKR

☐ LIGHT SENSITIVE UP-DOWN TONE GENERATOR

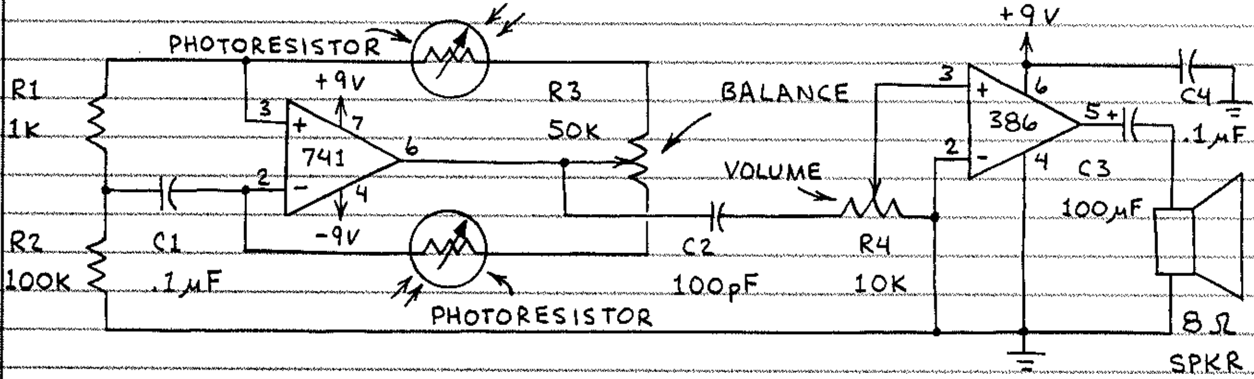
PHOTORESISTOR
R1
1K
+9V
3
7
741
6
2
4
-9V
R3
50K
BALANCE
R2
100K
C1
.1μF
PHOTORESISTOR
C2
100pF
VOLUME
R4
10K
+9V
3
2
6
4
5
386
+
C4
.1μF
C3
100μF
8Ω
SPKR

TAKE THIS CIRCUIT AND A FLASHLIGHT INTO A DARK ROOM...

☐ LIGHTWAVE COMMUNICATIONS RECEIVER

WILL DETECT VOICE OR TONE MODULATED LIGHT BEAM. USE LENS FOR MORE RANGE. KEEP DETECTOR AWAY FROM SUNLIGHT.

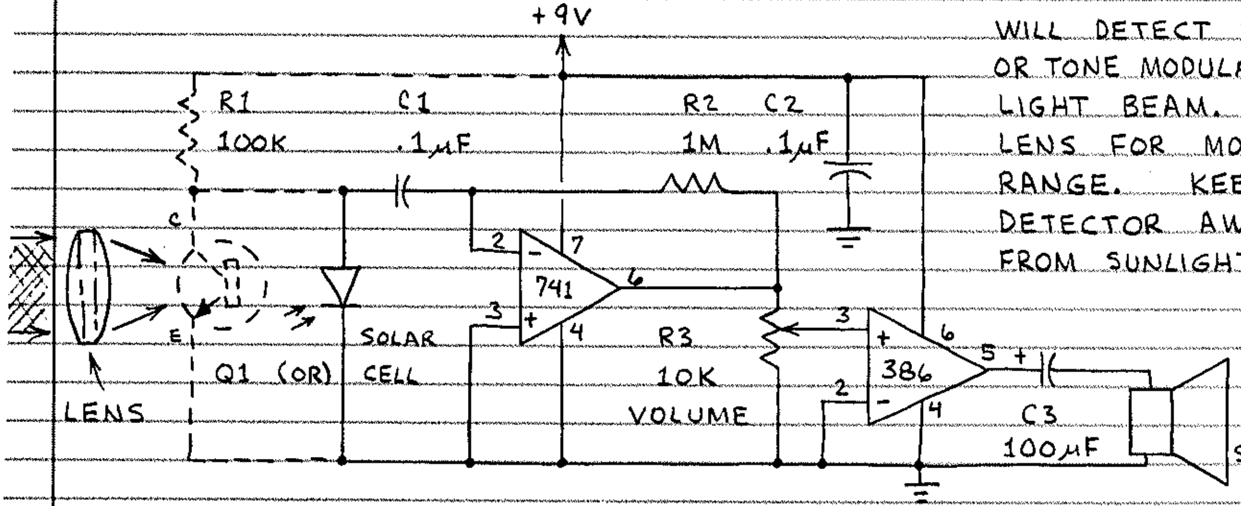
+9V
R1
100K
C1
.1μF
R2
1M
C2
.1μF
C
E
2
3
7
4
741
6
LENS
Q1 (OR)
SOLAR CELL
R3
10K
VOLUME
3
2
6
4
5
386
+
C3
100μF
8Ω
SPKR

DETECTOR: USE R1 AND Q1 OR SOLAR CELL. (Q1: PHOTOTRANSISTOR)

COMPARATOR CIRCUITS

☐ VOLTAGE MONITOR

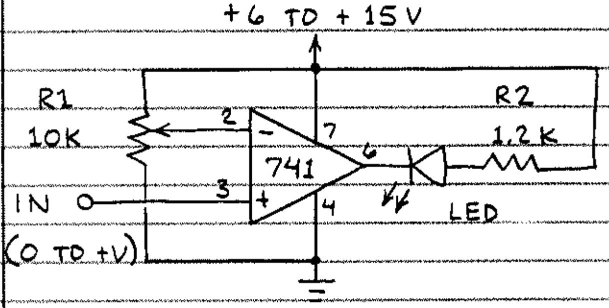
+6 TO +15V
R1
10K
2
7
R2
1.2K
741
6
3
IN
(0 TO +V)
4
LED

WHEN THE INPUT VOLTAGE IS 0, THE LED GLOWS. THE LED SWITCHES OFF WHEN THE INPUT VOLTAGE RISES TO A LEVEL DETERMINED BY R1. EXCHANGE CONNECTIONS TO PINS 2 AND 3 TO REVERSE OPERATING MODE.

☐ LIGHT LEVEL INDICATOR

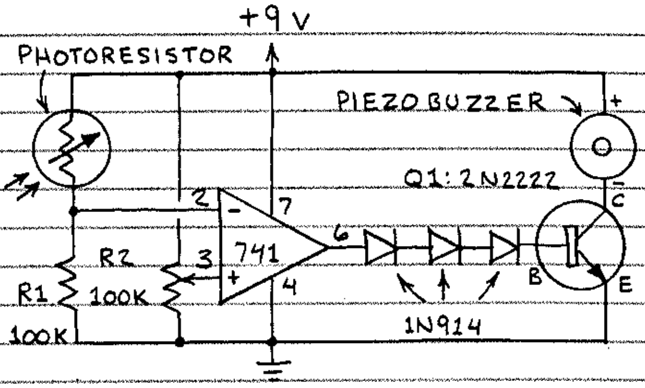
+9V
PHOTORESISTOR
PIEZOBUZZER
+
−
Q1: 2N2222
C
2
7
741
6
3
4
R2
100K
R1
100K
B
E
1N914

BUZZER SOUNDS WHEN LIGHT FALLS TO LEVEL DETERMINED BY R2. REVERSE CONNECTIONS TO PINS 2 AND 3 TO SOUND TONE WHEN LIGHT INCREASES.

☐ BARGRAPH VOLTAGE INDICATOR

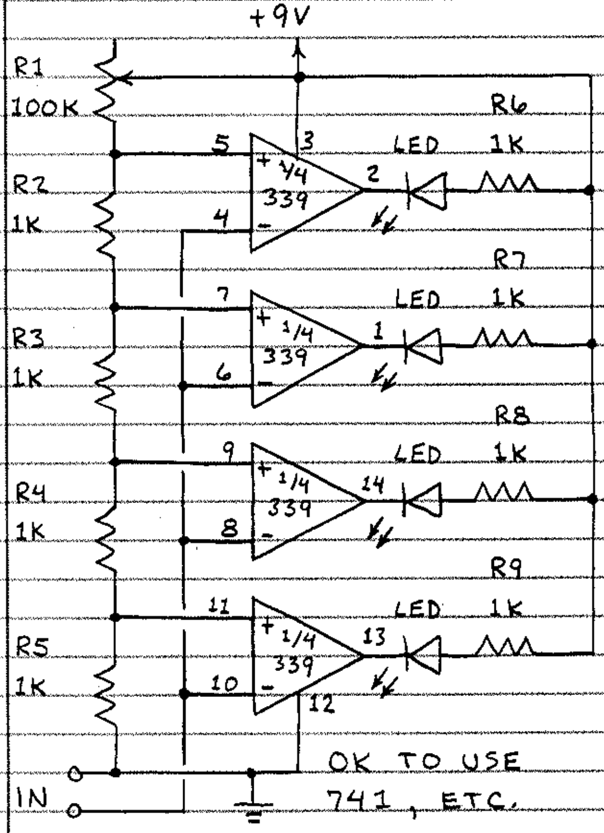
+9V
R1
100K
R2
1K
R3
1K
R4
1K
R5
1K
5
3
1/4
339
2
4
LED
R6
1K
7
6
1
R7
1K
9
8
14
R8
1K
11
10
13
12
R9
1K
IN
OK TO USE 741, ETC.

LEDs GLOW IN SEQUENCE AS INPUT VOLTAGE RISES. R1 CONTROLS SENSITIVITY.

☐ "WINDOW" COMPARATOR

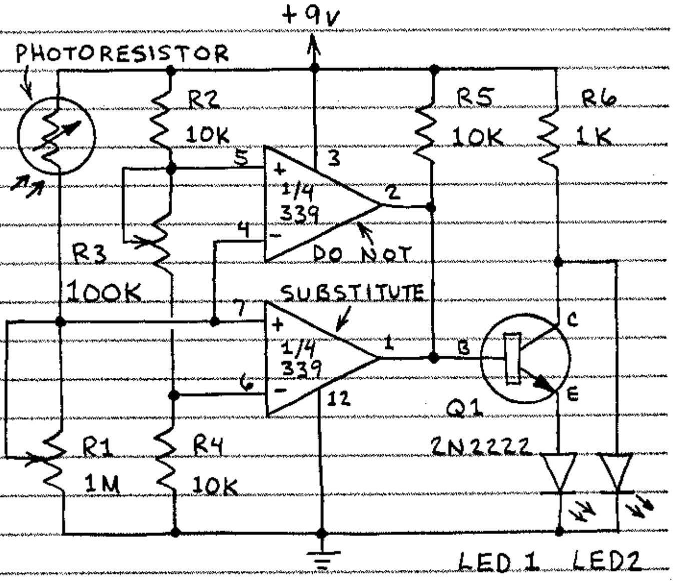
+9V
PHOTORESISTOR
R2
10K
R5
10K
R6
1K
5
3
1/4
339
2
4
DO NOT
SUBSTITUTE
R3
100K
7
1
6
12
B
C
E
Q1
2N2222
R1
1M
R4
10K
LED 1
LED2

SET R1 TO CENTER POSITION. TURN OFF LIGHTS AND ROTATE R3 JUST PAST POINT WHERE LED 2 GLOWS. CIRCUIT WILL THEN RESPOND LIKE THIS: (R1 & R3 CONTROL RESPONSE)

DARK ⟶ LIGHT
LED 1 = OFF ←–ON–→ OFF
LED 2 = ON ←–OFF–→ ON

VOLTAGE REGULATOR CIRCUITS

☐ FIXED OUTPUT LINE POWERED SUPPLY

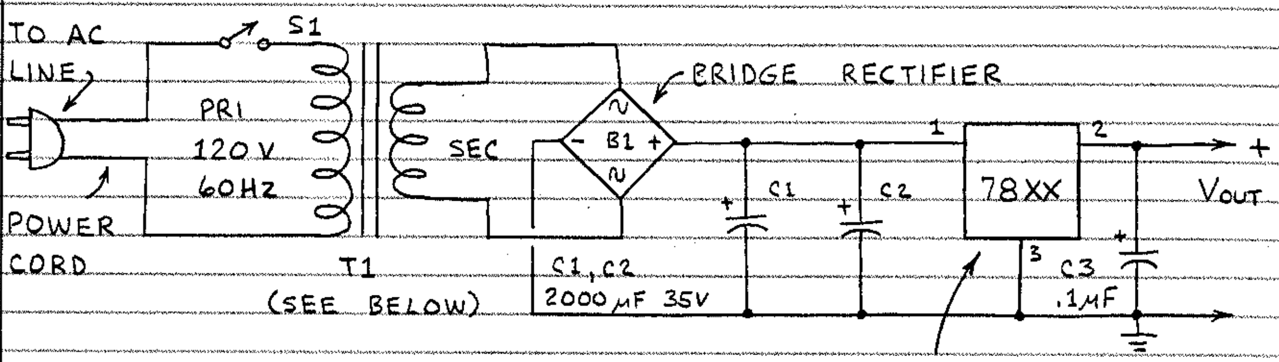
TO AC LINE
S1
PRI
120 V
60 HZ
POWER CORD
T1
(SEE BELOW)
SEC
BRIDGE RECTIFIER
~
- B1 +
~
1
2
78XX
+
VOUT
C1
C2
+
+
C1, C2
2000 µF 35V
3
C3
.1µF

THIS BASIC SUPPLY WILL DELIVER UP TO 1.5 AMPERES AT THE RATED OUTPUT IF PROPERLY HEAT SUNK. YOU MUST USE A TRANSFORMER RATED AT THE PROPER VOLTAGE AND CURRENT. THE REGULATOR WILL "SHUT DOWN" IF THE CHIP BECOMES OVERHEATED. FOR BEST RESULTS APPLY SILICONE COMPOUND BETWEEN TAB AND HEAT SINK. ALL CONNECTIONS TO THE AC LINE MUST BE INSULATED OR ENCLOSED!

REGULATOR IC

7805 = 5 VOLTS

7812 = 12 VOLTS

7815 = 15 VOLTS

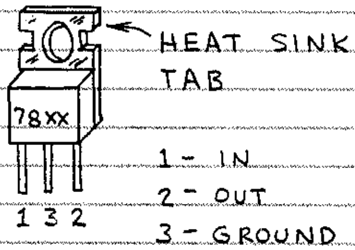
HEAT SINK TAB
78XX
1 3 2

1 - IN

2 - OUT

3 - GROUND

CAUTION: AC LINE OPERATED CIRCUITS.

☐ VARIABLE OUTPUT POWER SUPPLY

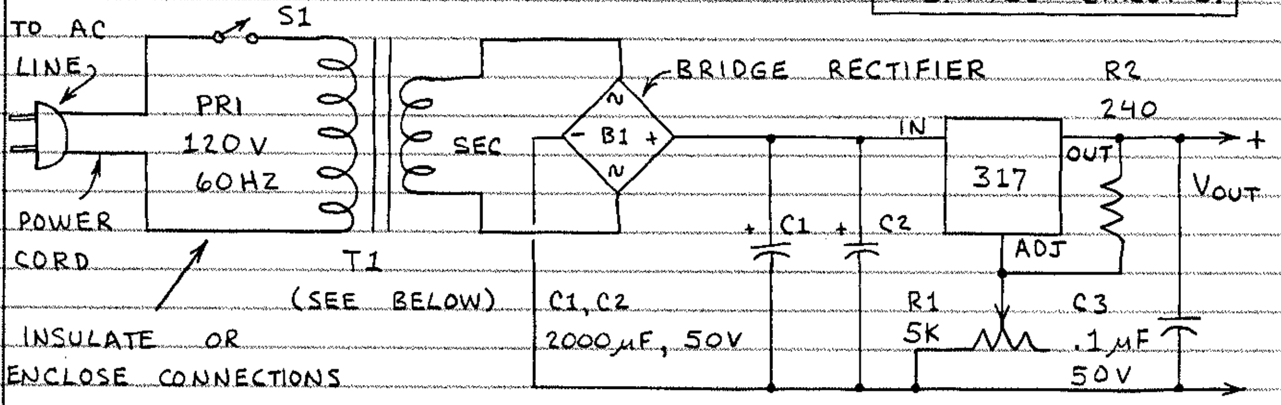
TO AC LINE
S1
PRI
120 V
60 HZ
POWER CORD
T1
(SEE BELOW)
INSULATE OR ENCLOSE CONNECTIONS
SEC
BRIDGE RECTIFIER
~
- B1 +
~
IN
317
OUT
ADJ
R2
240
+
VOUT
+ C1
+ C2
C1, C2
2000 µF, 50V
R1
5K
C3
.1 µF
50V

THIS ADJUSTABLE SUPPLY WILL DELIVER FROM 1.2 TO 37 VOLTS AT UP TO 1.5 AMPERES. R1 CONTROLS VOUT. (IF VOUT DOES NOT GO TO 1.2 VOLT MINIMUM, R1 MAY NOT BE ABLE TO ACHIEVE SUFFICIENTLY LOW RESISTANCE.) T1 SHOULD HAVE 25 VOLT (OR HIGHER) SECONDARY AND BE RATED FOR 2 AMPERES OR MORE.

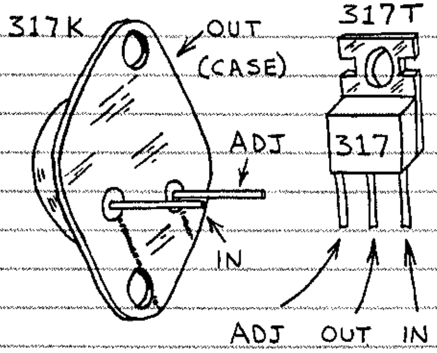
317K
OUT (CASE)
ADJ
IN
317T
317
ADJ
OUT
IN

TIMER CIRCUITS

☐ BASIC TIMER

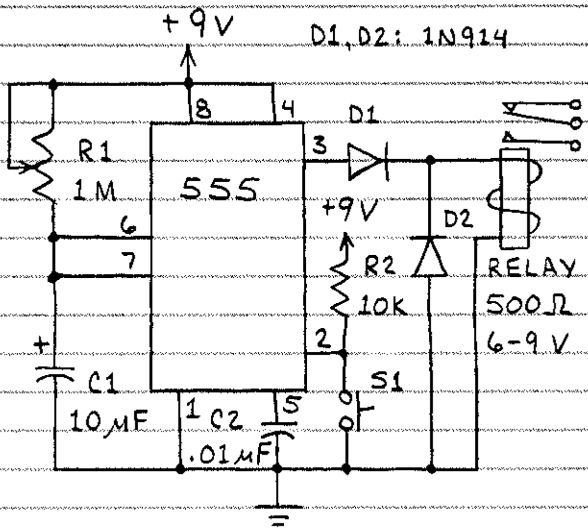
+9V
D1, D2: 1N914
8
4
D1
3
R1
1M
555
6
7
+9V
D2
R2
10K
RELAY
500Ω
6-9V
2
S1
+
C1
10µF
1
C2
.01µF
5

PRESS S1 MOMENTARILY TO START TIMING CYCLE. RELAY WILL BE ACTUATED (PULLED IN) UNTIL CYCLE IS COMPLETE. R1 AND C1 CONTROL LENGTH OF TIME DELAY. USE VERY LARGE VALUE FOR C1 TO GET LONG DELAYS. CIRCUIT WILL RESPOND TO LOGIC PULSES, TOO.

☐ TONE BURST GENERATOR

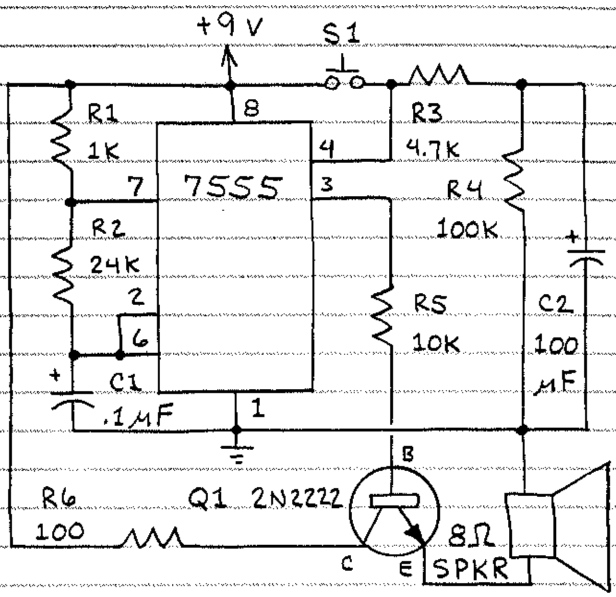
+9V
S1
R1
1K
8
4
R3
4.7K
7
7555
3
R4
100K
R2
24K
+
2
6
R5
10K
C2
100
µF
+
C1
.1µF
1
R6
100
Q1 2N2222
B
C
E
8Ω
SPKR

PRESS S1 AND THE SPEAKER EMITS A TONE. RELEASE S1 AND THE TONE CONTINUES FOR SEVERAL SECONDS. C2 AND R4 CONTROL DELAY. C1 CONTROLS FREQUENCY. (USE 7555 ONLY. 555 USES TOO MUCH CURRENT.)

☐ PULSE GENERATOR

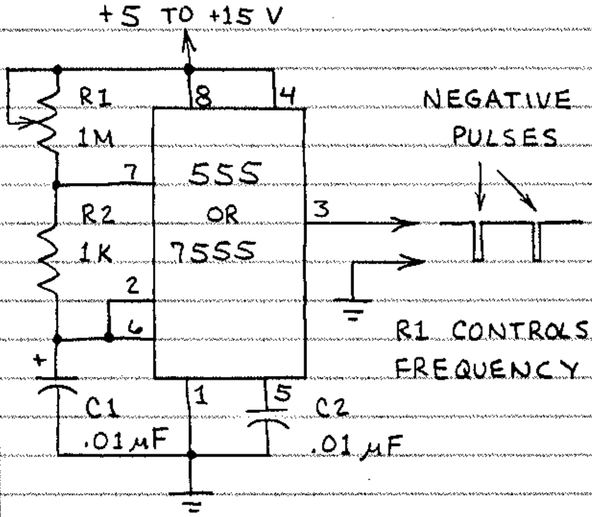
+5 TO +15 V
R1
1M
8
4
NEGATIVE
PULSES
7
555
OR
7555
3
R2
1K
2
6
+
C1
.01µF
1
5
C2
.01µF

R1 CONTROLS FREQUENCY

USE TO SUPPLY PULSES TO DIGITAL LOGIC CIRCUITS, ETC.

☐ LED TONE TRANSMITTER

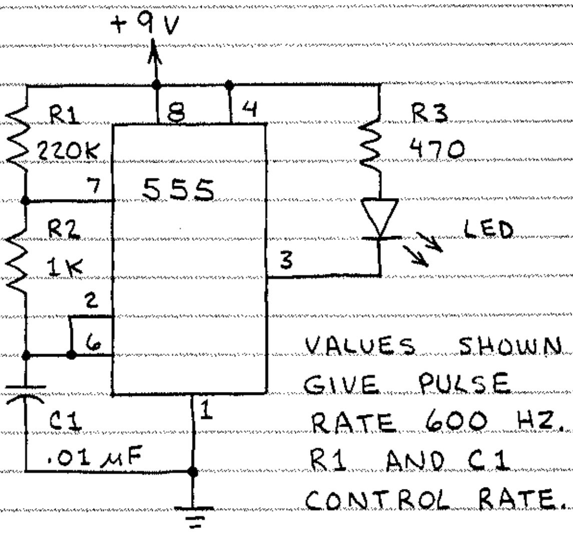
+9V
R1
220K
8
4
R3
470
7
555
LED
R2
1K
3
2
6
C1
.01µF
1

VALUES SHOWN GIVE PULSE RATE 600 HZ. R1 AND C1 CONTROL RATE.

USE TO TEST LIGHTWAVE RECEIVERS.

☐ LIGHT / DARK DETECTOR

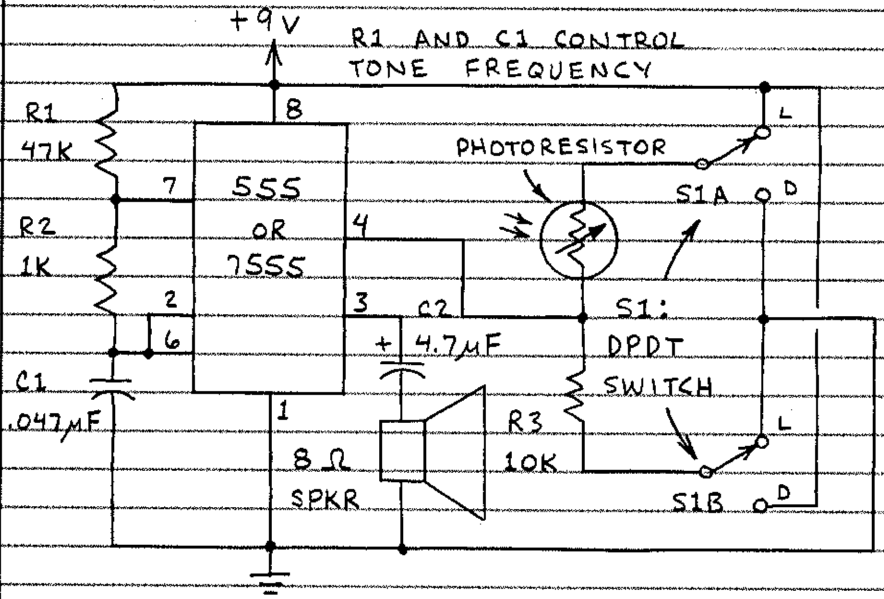
+9V
R1 AND C1 CONTROL TONE FREQUENCY
R1 47K
R2 1K
C1 .047μF
555 OR 7555
8
7
2
6
4
3
1
PHOTORESISTOR
S1A
L
D
C2 + 4.7μF
S1: DPDT SWITCH
R3 10K
8 Ω SPKR
S1B
L
D

WHEN S1 IS IN POSITION "L," THE SPEAKER EMITS A TONE WHEN LIGHT ILLUMINATES THE PHOTORESISTOR. WHEN S1 IS IN POSITION "D," THE SPEAKER EMITS A TONE WHEN THE PHOTORESISTOR IS NOT ILLUMINATED.

☐ 3-STATE TONE GENERATOR

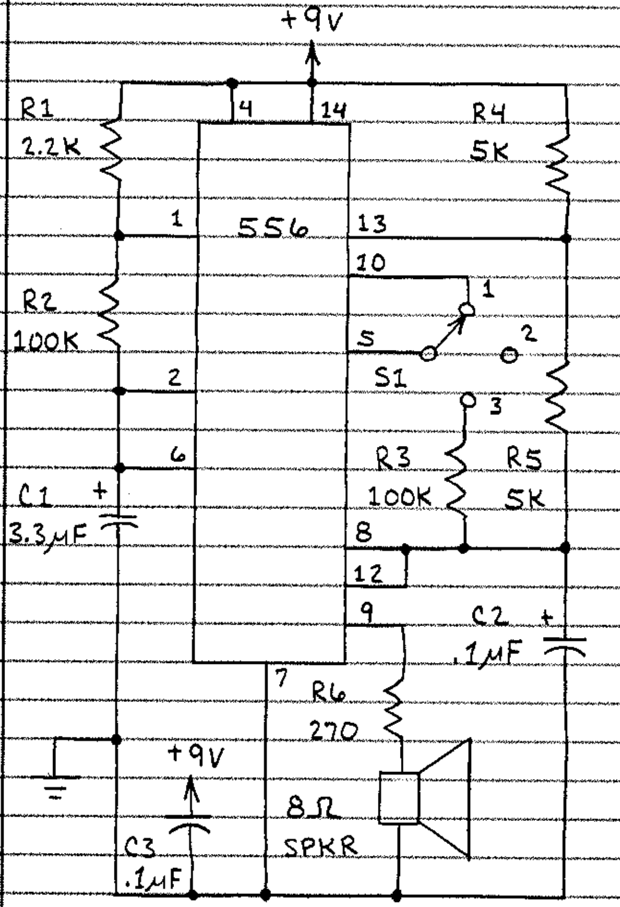
+9V
R1 2.2K
R2 100K
C1 + 3.3μF
556
4
14
1
2
6
13
10
5
8
12
9
7
R4 5K
S1
1
2
3
R3 100K
R5 5K
C2 + .1μF
R6 270
8Ω SPKR
+9V
C3 .1μF

S1:

1- TONE BURST

2- STEADY TONE

3- TWO TONE

☐ EVENT FAILURE ALARM

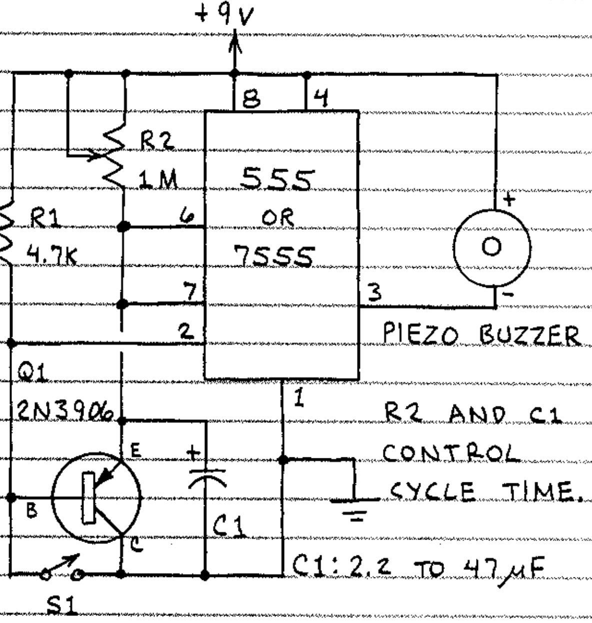
+9V
8
4
R2 1M
R1 4.7K
555 OR 7555
6
7
2
3
1
+
-
PIEZO BUZZER
Q1 2N3906
E
B
C
+
C1
S1
R2 AND C1 CONTROL CYCLE TIME.
C1: 2.2 TO 47μF

WHEN POWER IS APPLIED, THE 555 ENTERS A TIMING CYCLE. UNLESS S1 IS CLOSED BEFORE THE CYCLE ENDS, THE PIEZO BUZZER SOUNDS. THE CYCLE CAN BE RESET ANY TIME BY CLOSING S1. NOTE: S1 CAN BE REPLACED BY A SIGNAL FROM AN EXTERNAL CIRCUIT.

INDEX

A
ALTERNATING CURRENT, 18, 36
AMPLIFIER, 92-93, 95, 107, 119, 122
ATOM, 8
B
BATTERY, 17, 44
BINARY, 81, 89
BIPOLAR, 48, 51, 104-105
C
CAPACITOR, 32-35, 36-37
CIRCUIT, 20-21, 82, 90
CMOS, 91, 118-121
COIL, 38-39
COMBINATIONAL, 86-87, 90
COMPARATOR, 93, 124
CONDUCTOR, 12, 43
CURRENT, 9, 13, 14, 45, 59, 67
D
DETECTOR, 70-77
DIFFERENTIATOR, 37
DIGITAL, 79, 80, 83, 91, 116-121
DIODE, 44-47, 78, 82, 101-103
DIRECT CURRENT, 14, 36
E
ELECTROMAGNET, 15
ELECTROMAGNETIC SPECTRUM, 63
ELECTRON, 8-9, 42-43, 62, 66, 70
ELECTROSCOPE, 11-12
F
FIELD-EFFECT, 52-56, 92
FLIP-FLOP, 88-89, 94, 116
G
GATE, 80, 82-87
GROUND, 21
I
INSULATOR, 12, 43
INTEGRATED CIRCUIT, 78-95
INTEGRATOR, 37
INVERSE SQUARE LAW, 65
ION, 8
J
JUNCTION FET, 52-53, 56, 106
L
LASCR, 76
LENS, 65, 122, 123
LIGHT, 62, 63, 66-68, 77, 111, 114, 122
LIGHT EMITTING DIODE, 46, 66-69, 112
LINEAR (IC), 79, 92, 95
M
MAGNETIC FIELD, 13
METER, 13, 25, 114
MICROPHONE, 27
MOSFET, 54-55, 56, 107
MULTIMETER, 19
N
NOISE, 22-23
NUCLEUS, 8
O
OHM'S LAW, 14, 103
OPERATIONAL AMPLIFIER, 93, 122-123
OPTICAL COMPONENTS, 64
OPTICAL SPECTRUM, 63
P
PARALLEL, 20-21, 31, 35, 69, 77
PHOTODIODE, 72-73
PHOTON, 62, 66, 70, 72, 74, 76
PHOTORESISTOR, 70-71, 114-115, 124
PHOTOTHYRISTOR, 76, 115
PHOTOTRANSISTOR, 74-75, 114, 121, 123
PN JUNCTION, 44, 66, 70, 72
POWER, 14, 116, 125
PROTON, 8
PULSE, 22-23, 119, 126
R
RC CIRCUITS, 37, 94, 100
RELAY, 25, 75, 114, 115
RESISTANCE, 14
RESISTOR, 28-31, 69, 78
S
SAFETY, 19, 35, 98, 100, 111, 125
SCR, 58-59, 110
SERIES, 20-21, 31, 35, 69, 77
SEMICONDUCTOR, 42-43, 62-63, 66, 70
SEQUENTIAL, 88-90
SIGNAL, 22-23
SILICON, 42-43, 72
SINE WAVE, 18
SOLAR CELL, 77-115
SOLDER, 98
SOLENOID, 15
SPEAKER, 27
STATIC ELECTRICITY, 10, 106, 118
SWITCH(ES), 25, 26, 58, 80
T
THERMOELECTRIC, 17
THYRISTOR, 58-61, 110-111
TIMER, 93, 106-107, 117, 126
TRANSFORMER, 40-41
TRANSISTOR, 48-57, 74-75, 78, 83, 104
TRIAC, 60-61, 111
TTL, 91, 116-117
U
UNIJUNCTION, 57, 108-109
V
VOLTAGE, 14, 45, 68
VOLTAGE REGULATOR, 95, 103, 125
W
WAVE(S), 22-23
WIRE, 24
WIRE-WRAP, 97

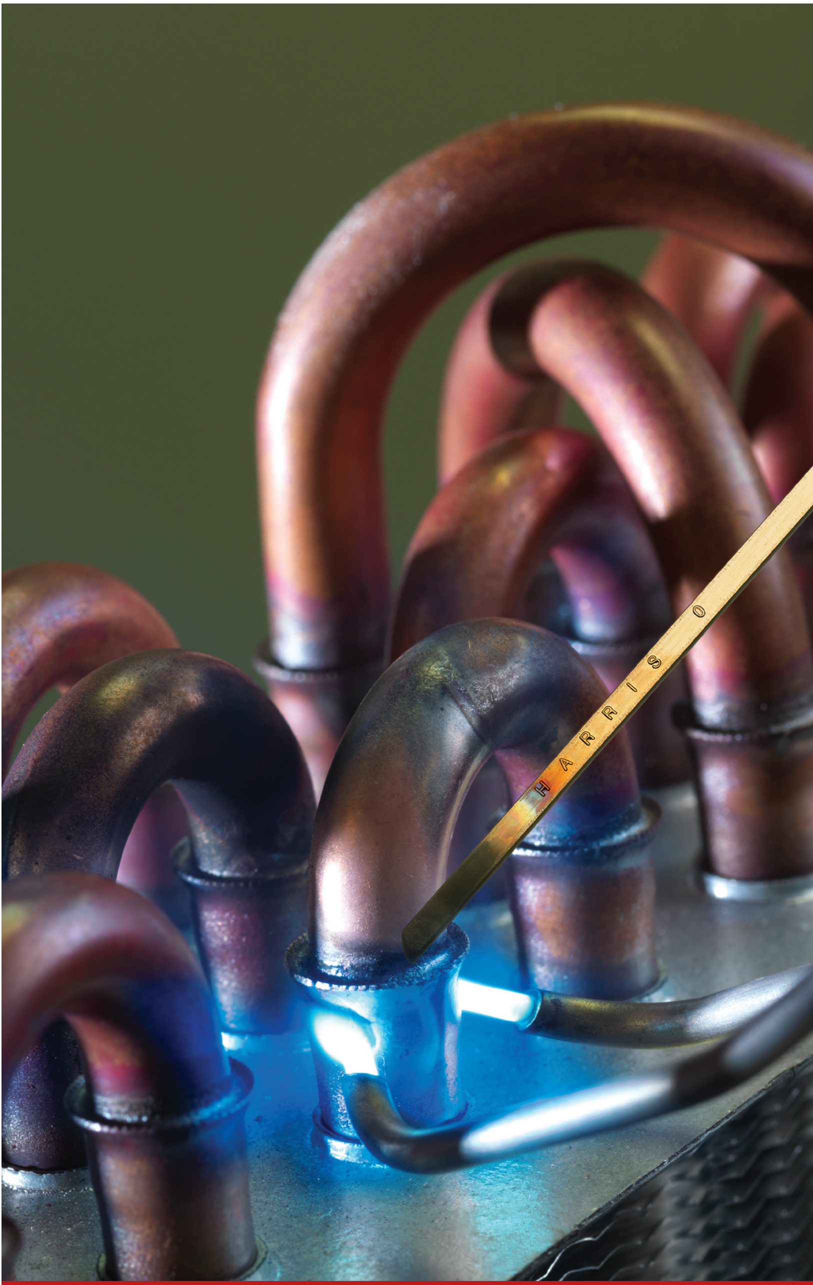

BRAZING AND SOLDERING

HARRIS
A LINCOLN ELECTRIC COMPANY

Turn to the Pros

SERVICEMAN'S GUIDE

BRAZING WORKSHOP

Individuals who desire further training in correct soldering, brazing, and welding procedures should consider attending The Harris Products Group's brazing workshop. This two-day, comprehensive course, conducted by The Harris Products Group's Technical Staff, covers fundamentals of torch brazing. The course combines classroom discussion with hands-on metal joining experience including basic metallurgy, base metals, filler metals, equipment, technique, and safety.

WHO SHOULD ATTEND?

Supervisors, foremen, quality control, service technicians, and others involved with supervision, instruction, or production of brazed assemblies should attend. Participants are encouraged to share their application problems for class discussion. For more information and course outline, contact The Harris Products Group Technical Services Department at 513.754.2000.

As a NATE®, recognized training provider, we offer several courses designed to improve your brazing skills and achieve required continuing education hours. Contact Technical Services for information.

Harris celebrates 100 years in the brazing industry. Harris is the largest manufacturer of welding, brazing and soldering consumables in the world. Our products are used in over 90 countries.

CONTENTS

PROCEDURES AND TECHNICAL INFORMATION

Procedures for Brazing Pipe and Tubing 6-11
Procedures for Troubleshooting Brazing Pipe and Tubing 12-13

PRODUCTS AND SPECIFICATIONS

Ring of Fire™ 16-17
Stay-Brite® Silver-bearing Solders 18-19
Common Lead-Free Solders 20
Bridgit® Lead-Free Solders 21
Safety-Silv® High Silver Brazing Alloys 22
Estimating Brazing Alloy Amounts 23
Stay-Silv® Phos/Copper Brazing Alloys 24-25
Aluminum Alloys 27
Safety-Silv® High Silver Brazing Alloys 28
Solders 29
Fluxes 30
Heating Equipment 36-37

CHARTS

Ring of Fire™ Ordering Information 17
Stay-Brite® Joint vs. Brazing Joint 19
Tin/Lead Solders 20
Estimating Brazing Alloy Amounts 23
Aluminum Alloys 27
Safety-Silv® 56 28
Safety-Silv® 45 28
Safety-Silv® 40 28
Stay-Brite® Solders 29
50/50 Solders 29
60/40 Solders 29
Fluxes 30
Phos-Copper Brazing Alloys 32-33
Lead-Free Solders 32-33
Filler Metals Selection Chart 34-35
Acetylene Welding/Brazing Tips 37

FOREWORD

Brazing is a metal joining process utilizing a filler metal that melts above 840ºF and below the melting point of the base metals. The filler metal is drawn into the joint by capillary attraction producing a sound, leak-proof connection. The above sentences briefly describe a process that is an integral part of manufacturing, installing, and repairing refrigeration and air conditioning systems. This handbook is designed to assist the serviceman or contractor in making sound brazed joints and selecting the correct filler metal for each application.

WARNING:
Protect yourself and others. Read and understand this information.
Brazing and soldering alloys and fluxes may produce FUMES AND GASES hazardous to your health.

- Before use, read and understand the manufacturer's instructions, Material Safety Data Sheets (MSDS) and your employer's safety practices.
- Keep your head out of the fumes.
- Use enough ventilation or exhaust at the flame to keep fumes and gases from your breathing zone and the general area.
- For maximum safety, be certified for and wear a respirator at all times when welding or brazing.
- Wear correct eye, ear, and body protection.
- See American National Standard Z49.1, Safety in Welding, Cutting and Allied Processes, published by the American Welding Society, 8669 NW 36 Street, #130, Miami, Florida 33166-6672;OSHA Safety and Health Standards, 29 CFR 1910, available from the U.S. Government Printing Office, Superintendent of Documents, P.O. Box 371954, Pittsburgh, PA 15250-7954.
- Material Safety Data Sheets are available for all Harris products. Each MSDS contains detailed safety and health information about possible hazards associated with use of each particular product. MSDS are available from your employer or by contacting The Harris Products Group, 4501 Quality Place, Mason, OH 45040-1971

STATEMENT OF LIABILITY :
This information and recommendations contained in this publication have been compiled from sources believed to be reliable and to represent the best information on the subject at the time of issue. No warranty, guarantee, or representation is made by The Harris Products Group as to the absolute correctness or sufficiency of any representation contained in this and other publications. The Harris Products Group assumes no responsibility in connection herewith, nor can it be assumed that all acceptable safety measures are contained in this (and other) publications, or that other or additional measures may not be required under particular or exceptional conditions or circumstances.

PROCEDURES AND TECHNICAL INFORMATION

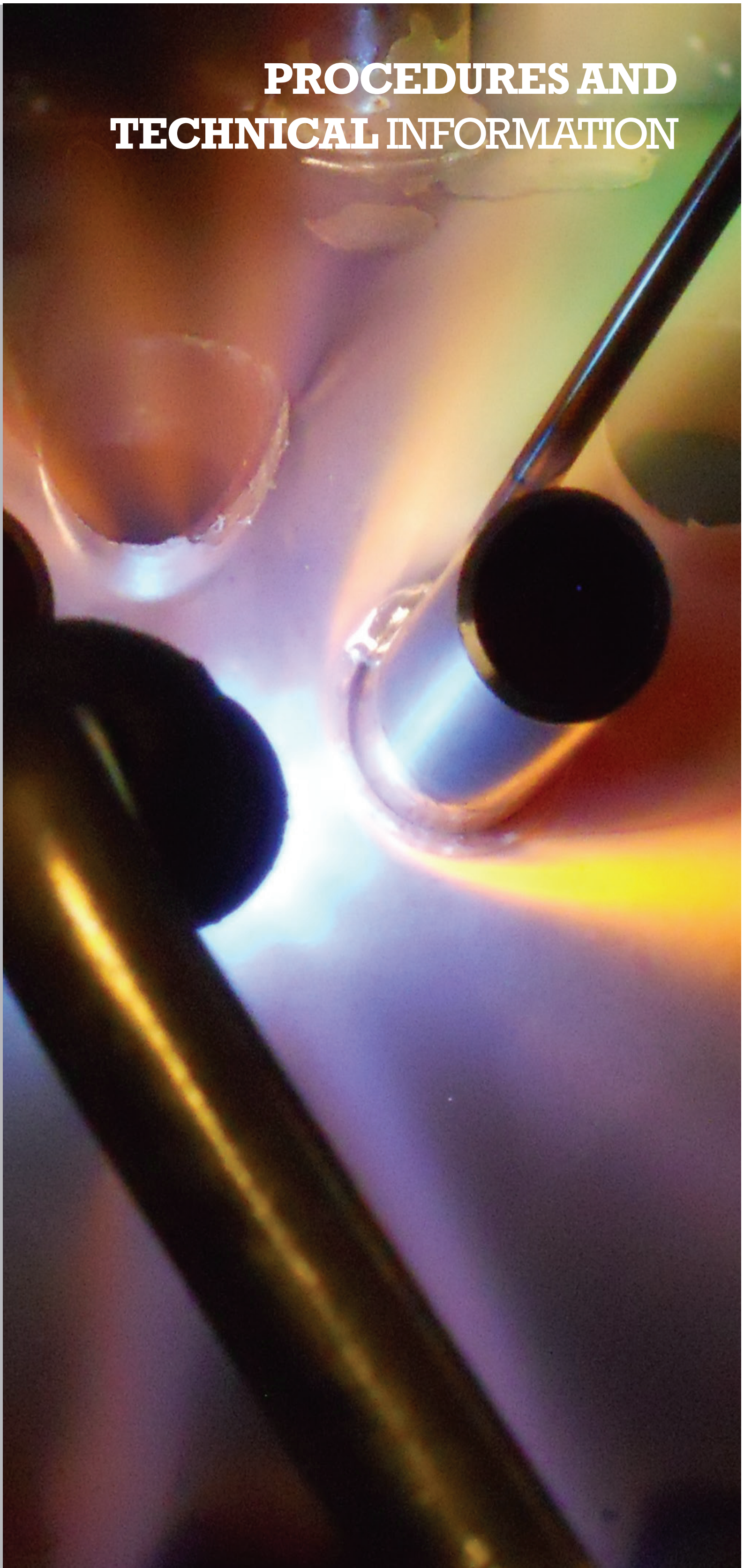

CUT TUBE SQUARE.

Cut to the exact length required using a tube cutter or hacksaw. If a hacksaw is used, a sawing fixture should also be used to ensure square cuts. Remove all inside and outside burrs with a reamer, file, or other sharp edge scraping tool. If tube is out of round, it should be brought to true dimension and roundness with a sizing tool.

CLEAN TUBE END AND INSIDE SURFACE OF FITTING.

The joint surface areas should be clean and free from oil, grease, or oxide contamination. Surfaces may be properly cleaned for brazing by brushing with a stainless steel wire brush or by a stiff rubbing with emery cloth or Scotch-Brite®*. If oil or grease is present, clean with a commercial solvent. Remember to remove small foreign particles such as emery dust, by wiping with a clean dry cloth. The joint surface MUST be clean.

SELECT BRAZING ALLOY.

Refer to the Filler Metal Selection Chart on page 32 for recommended brazing filler metal selection. When brazing copper to copper, alloys such as Dynaflow®, Stay-Silv® 5, or Stay-Silv® 15 are recommended. These alloys contain phosphorus and are self-fluxing on copper. When brazing brass or bronze fittings, Stay-Silv® white flux is required with these alloys. When brazing iron, steel or other ferrous metals, select one of the Safety-Silv® brazing alloys such as Safety-Silv® 45 or Safety-Silv® 56 with Stay-Silv® white flux. Do not use phosphorus bearing alloys as the joint may be brittle.

*Scotch-Brite is a trademark of 3M

PERFORM PROPER FLUXING

Proper fluxing is important because the flux absorbs oxides formed during heating and promotes the flow of the filler metal. When using Stay-Silv® white flux, apply it only with a brush. To prevent excess flux residue inside refrigeration lines, apply a thin layer of flux to only the male tubing. Insert the tube into the fitting and, if possible, rotate the fitting once or twice on the tube to ensure uniform coverage.

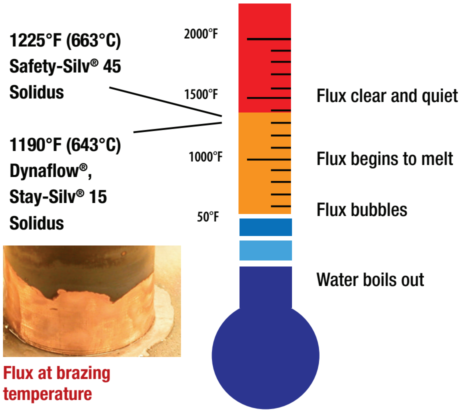

Flux at brazing temperature

FLUX APPLICATION

White flux is used for most applications. Black flux is helpful for long heating cycles or localized heating with induction. It is also used when brazing stainless steel.

Flux goes through physical changes during heating and turns clear at about 1100°F / 593°C. This is an indication that parts are close to brazing temperature. Stir flux before use. If flux is dried out add a small amount of water until flux reaches a paste consistency.

STAY-SILV® WHITE BRAZING FLUX IS AVAILABLE IN THESE CONTAINERS:

- 7 oz.
- 1/4 lb.
- 1/2 lb.
- 1 lb.
- 5 lb.
- 25 lb.
- 60 lb.

ADJUST TORCH FLAME

OXYGEN/ACETYLENE

For most brazing jobs using oxy-acetylene gases, a slightly carburizing or neutral flame should be used. The neutral flame has a well defined inner cone. Avoid an oxidizing flame.

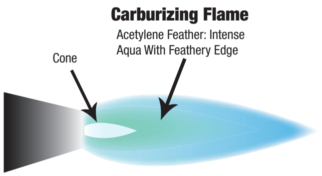

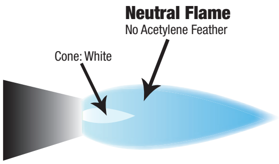

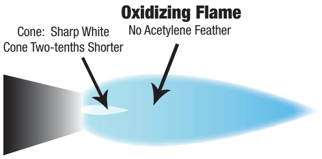

AIR/ACETYLENE USING INFERNO® SWIRL COMBUSTION TIPS

Brazing with air/acetylene torches is a popular alternative to oxygen mixed fuel gas. The fuel gas flow aspirates air into a mixer that contains an internal vane that spins the gas to improve combustion and increase flame temperature.

If the tank has a delivery pressure gauge, set the delivery pressure to 14-15 PSI. If the tank has only a contents gauge delivery pressure is preset at the factory. Open the regulator adjusting screw all the way by turning it clockwise until it bottoms.

OPEN THE TORCH VALVE

Opening the torch valve about 3/4 of a turn will provide sufficient fuel gas delivery. Do not try to meter pressure (reducing the flame) by using the torch handle valve. If a higher or lower flame is required, change to a different tip size.

OTHER FUEL GASES

Alternate fuel gases such as propane, propylene, and natural gas can be mixed with oxygen for brazing. Refer to the Harris equipment catalog or website for equipment and setting information.

HEATING THE JOINT AREA.

Always keep the torch in short motion.

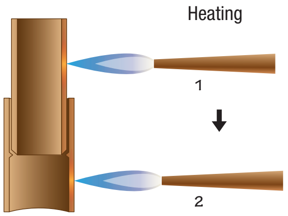

1. Start heating the tube, by first applying the flame to a point just adjacent to the fitting. Work the flame alternately around the tube and fitting until both reach brazing temperature, before applying the brazing filler metal.

2. When a flux is used, it will be a good temperature guide. Continue heating the tube until the flux passes the “bubbling” temperature range and becomes quiet, completely fluid and transparent. Watch for this on both sides of the joint to ensure even heating.

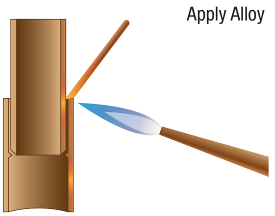

3. Direct the flame from the tube to the fitting. When alloy is applied it should quickly melt and flow into the joint.

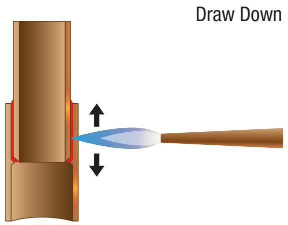

4. Sweep the flame back and forth along the axis of the assembled joint, tube, and fitting to reach and then maintain uniform heat in both parts.

APPLY THE BRAZING ALLOY.

Feed the alloy into the joint between the tube and the fitting. Only after the base metals have been heated to brazing temperatures should the filler metal be added. At that time, the flame may be deflected momentarily to the tip of the filler metal to begin the melting process. Always keep both the fitting and the tube heated by playing the flame over the tube and the fitting as the brazing alloy is drawn into the joint. The brazing alloy will diffuse into and completely fill all joint areas. Do not continue feeding brazing alloy after the joint area is filled. Excess fillets do not improve the quality or the dependability of the braze and are a waste of material.

WHEN MAKING VERTICAL ALLOY-UP JOINTS

Heat the tube first, then apply heat to the fitting. It is important to bring both pieces up to temperature evenly. Keep the flame directed toward the fitting. If the tube is overheated, the brazing alloy may run down the tube rather than into the joint.

CLEAN AFTER BRAZING

All flux residue must be removed for inspection and pressure testing. Immediately after the brazing alloy has set, quench or apply a wet brush or swab to crack and remove the flux residue. Use emery cloth or a wire brush, if necessary.

TO SEPARATE A BRAZING JOINT

First clean the joint thoroughly, then flux the visible alloy and all adjacent areas of the tube and fitting. Next, heat the joint (tube and fitting) evenly, especially the flange of the fitting. When brazing alloy becomes fluid throughout the joint area, the tube can be easily removed. To re-braze the joint, clean the tube end and the inside of the fitting and proceed as directed to make a new brazed joint.

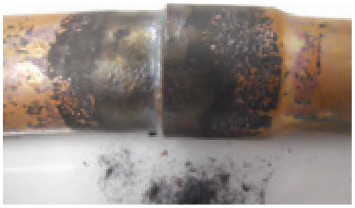

NITROGEN PURGE

During braze heating, oxide scale forms on the inside of the copper tube. These dark scales flake off and are carried by refrigerant and can potentially clog small orifices.

For HVAC/R and medical gas installations flow nitrogen through the tube during brazing to prevent internal scale formation. Use a low flow rate to avoid excess pressure inside the tube. A small hole at the line end will allow the nitrogen to escape.

BRAZING PIPE AND TUBING

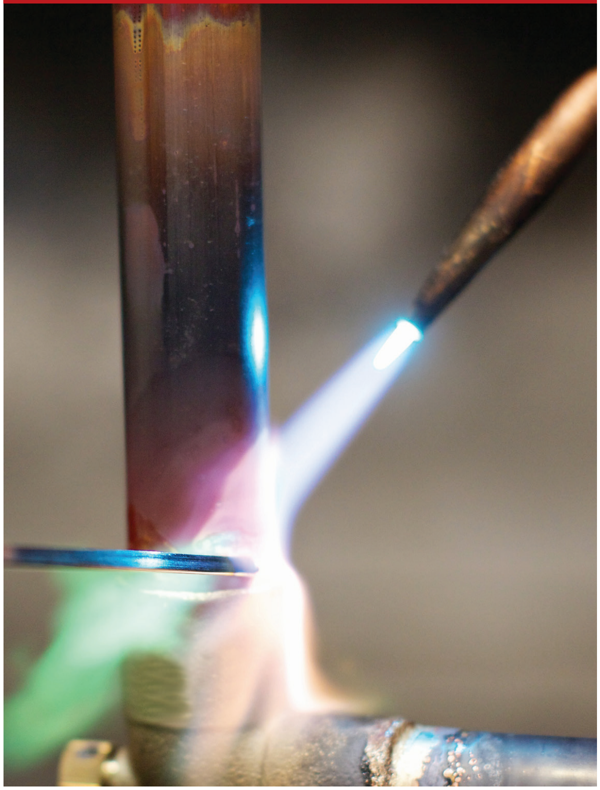

The art of brazing is relatively simple, and the rules of common sense apply. Occasionally, however, things do go wrong, and the brazing process fails to do its job satisfactorily. The check lists below have been prepared to assist in such instances. They are intended to provide practical tips on what to what to look for and steps to correct them.

IF BRAZING ALLOY DOES NOT FLOW INTO THE JOINT, EVEN THOUGH IT MELTS AND FORMS A FILLET

1. The outside of the joint is hot, but the inside is not up to brazing temperature.
   Review correct heating procedure on page 9. Remember to heat the tube first to conduct heat inside the fitting.

2. There is a flux breakdown due to excessive heat. If overheated, the flux can become saturated with oxides and the brazing alloy won't flow. Try using a softer flame and/or applying a heavier coating of flux. On thick sections where heating is prolonged, or on stainless steel, Harris Stay-Silv® black flux is recommended.

IF BRAZING ALLOY DOES NOT WET SURFACES BUT BALLS UP INSTEAD OF RUNNING INTO THE JOINT

1. Review heating techniques:
   (a) The base metals are not up to brazing temperature, and the alloy has been melted by the torch flame.
   (b) The joint has been overheated and the flux is no longer active.

2. Base metals have not been properly cleaned.

IF BRAZING ALLOY FLOWS AWAY FROM INSTEAD OF INTO THE JOINT

1. Make sure fitting is up to temperature and the flame is directed towards the fitting.

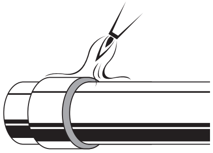

IF THE FILLER METAL CRACKS AFTER IT SOLIDIFIES

1. When brazing dissimilar metals, the different coefficient of expansion may put the filler metal in tension just below the liquidus temperature during cooling. This sometimes occurs in a copper-to-steel joint. The copper expands and contracts at a greater rate than the steel. Brazing alloys are stronger in compression, so a steel-to-copper assembly would help alleviate the problem.

2. Brazing steel (or other ferrous metals) with an alloy containing phosphorus can lead to formation of a brittle phosphide, that is prone to cracking. Braze ferrous metals with non-phosphorus content alloys.

3. Excessive joint clearance can lead to filler metals cracking under stress or vibration. Make sure clearances are held to .002" - .006" at brazing temperature (depending on alloy).

4. Too rapid quenching can sometimes cause cracking. Let joint cool more before washing off flux residue.

IF JOINT LEAKS IN SERVICE

90% of "leakers" in service are due to incorrect brazing techniques. The most common causes are:

1. Improper or uneven heating of joint. The effect of this is inadequate or incomplete penetration by the filler metal. Review proper techniques on page 9.

2. Overheating, causing volatilization of elements (phosphorus, zinc, etc.).

3. Incorrect torch flame adjustment, leading to deposition of carbon or causing excessive oxidation.

REPAIR OF LEAKS

Pinhole leaks in copper-to-copper joints brazed with phosphorus/copper or phosphorus/copper/silver filler metals can often be repaired using Blockade®. If care is taken, you can re-braze the joint with Blockade® without re-melting the original braze. Clean thoroughly before brazing.

We **DO NOT** recommend brazing over joints previously soldered with tin/lead solders. The low melting elements in the solder may prevent proper filler metal / base metal alloying.

Pinhole leaks in joints brazed with either the phosphorus or high silver alloys can usually be repaired with Stay-Brite® solder. Clean the joint thoroughly before soldering and use Stay-Clean® liquid flux.

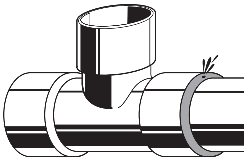

PRODUCTS &
SPECIFICATIONS

HARRIS
Stay-Silv®
White Brazing Flux
Fundente de soldadura blanco
Flux de brasage fort blanc
DANGER: BEFORE USE, READ MSDS
PELIGRO: ANTES DEL USO LEÍDO MSDS
DANGER: AVANT L'EMPLOI LU MSDS
NET WT. 6.5 oz. (184 g)

RING OF FIRE™

STAY-SILV® 15% BRAZING RINGS

CONVENIENT | CONSISTENT | QUALITY

Ring of Fire™ is the perfect alternative to any brazing project. It's easy to use, provides a consistent braze every time and meets the same high standards expected from Harris.

THE HARRIS ADVANTAGE:

- Made in the USA.
- One of *Industry Week's* Best Plants
- Provides consistent joints
- Available in a variety of sizes
- Inventory is better managed with smaller packages
- Purchase and carry only what is needed
- Visual verification of a completed joint
- Can never use too much or too little alloy

When working in the field, reliable supplies and tools are a necessity to make the job easier. Ring of Fire™ is a great alternative to rod for any brazing project. It's easy to use and provides a consistent braze every time. It's the ring form of Harris' Stay-Silv® 15 (15% silver alloy) using the same formulation, testing and standards. Work with confidence knowing that joints are tight and leak-proof when Ring of Fire™ brazing rings are used on projects.

Ask for Harris' Ring of Fire™ brazing rings when purchasing other brazing alloys and HVAC/R supplies.

Scan the QR code for more information.

ORDERING INFORMATION

PART #	DESCRIPTION	SIZE
RF15250	SS15 - 1/4" Joint Ring x 25 pkg.	1/4"
RF15375	SS15 - 3/8" Joint Ring x 25 pkg.	3/8"
RF15500	SS15 - 1/2" Joint Ring x 25 pkg.	1/2"
RF15625	SS15 - 5/8" Joint Ring x 25 pkg.	5/8"
RF15750	SS15 - 3/4" Joint Ring x 25 pkg.	3/4"
RF15875	SS15 - 7/8" Joint Ring x 25 pkg.	7/8"
RF15125	SS15 - 1 1/8" Joint Ring x 10 pkg.	1 1/8"
RF15VAR	SS15 - Tubing Joint Ring Variety Pack	

3 EASY STEPS FOR A TIGHT, LEAK- PROOF JOINT

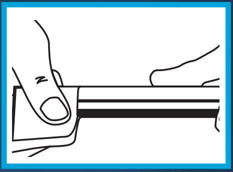

1 Clean the tube, pipe and fittings to be brazed.

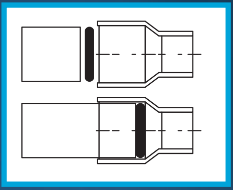

2 Place the brazing rings inside the coupling and insert the tube. Be sure the ring is at the bottom and touches the end of the tube.

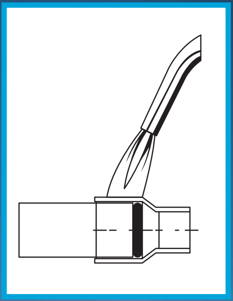

3 Heat the tube and coupling, applying heat to all the parts. Heat will transfer throughout the part and draw the alloy through the capillary space. The molten braze will appear around the top of the coupling for easy visual verification of the completed braze. Read the complete instructions inside the package and refer to the Material Safety Data Sheet (MSDS) before using.

Extremely versatile, Stay-Brite® silver-bearing solders are widely used throughout the industry as a better-than-brazing method in many situations.

The important advantage of Stay-Brite® solders is the greater strength of the overall component. After joining, its lower working temperature eliminates the weakening of the base metals caused by annealment from high brazing heat. The result is a stronger, more economical joint.

Stay-Brite® silver bearing solders were introduced to the air conditioning industry more than 40 years ago.

Today, Stay-Brite® is the most used of all solders.

Stay-Brite® silver bearing solders have the same excellent affinity as Safety-Silv® alloys for bonding with most of the ferrous and nonferrous alloys (including stainless steel, nickel, copper, brass, etc.) It provides a considerably higher than necessary elongation for sound dissimilar metal joints and vibration applications. Stay-Brite® alloys range in temperature from 430°F to 535°F.

Stay-Brite® solder eliminates oxide scale formed from brazing heat. No nitrogen purge is required for HVAC/R tubes. They are cadmium free and nontoxic.

Stay-Brite® is a eutectic alloy with a single melting point of 430°F (221°C).

Stay-Brite® 8 has a solidus of 430°F (221°C) and a liquidus of 535°F (279°C). This melting range provides the ability to fill wider clearance parts.

Stay-Brite
Stay Brite
SILVER BEARING SOLDER
NSF
RoHS COMPLIANT

Stay-Brite® the industry's most popular solder.

STAY-BRITE® JOINT vs. BRAZING JOINT

Stay-Brite® offers these important advantages over silver brazing:

- Lower material costs up to 66%
- Lowers temperature up to 60%
- Speeds production up to 400%
- Faster post cleaning with little metal discoloration
- Elimination of base metal distortion
- Elimination of base metal annealment
- Elimination of oxide scale formed by heat
- Cadmium-free and nontoxic
- No internal oxide scale, so nitrogen purge is unnecessary

STAY-BRITE®

Certified to ANSI/NSF 61 and NSF/ANSI 372 for drinking water systems. Conforms to lead-free content requirements of the U.S. Safe Drinking Water Act.

STAY-BRITE® 8

Certified to ANSI/NSF 51- Food Equipment Materials

STAY-BRITE® JOINT vs. BRAZING JOINT

PROPERTIES	STAY-BRITE®	BRAZED JOINT
Tube burst strength (before heating; type M-1)	3,800 PSI	3,800 PSI
Tube burst strength (after joining)	3,800 PSI	1,800 PSI
Oxide scale	None	Heavy
Annealment	None	Heavy

COMMON WIRE SOLDERS FOR TIN-LEAD

TIN-LEAD 50/50, 40/60, 60/40

Meets ASTM B32. With some exceptions, the tin-lead solders can be used to solder copper to most copper, lead, high-nickel and steel alloys.

Tin-lead solders are not recommended for use in high stress or vibration joints in the cooling industry due to lack of sufficient elongation properties. Heat sources for use with solder include soldering guns, irons, and torch applications.

TIN-ANTIMONY 95/5

Meets ASTM B32. The 95/5 tin-antimony solder is useful for applications where moderately elevated temperature is a factor. It has a higher electrical conductivity and is sometimes used where lead contamination must be avoided. The tin-antimony solders are not recommended for use on brass.

Harris 95/5 is certified to ANSI/NSF 61 and NSF/ANSI 372 for drinking water systems. Conforms to lead-free content requirements of the U.S. Safe Drinking Water Act.

FLUX SELECTION FOR TIN-LEAD, AND TIN-ANTIMONY SOLDERS

Both Stay-Clean® paste and liquid soldering fluxes are recommended, except in electrical or electronic applications that require the use of a non-corrosive flux.

TIN/LEAD SOLDERS

COMPOSITION/ PROPERTIES

Solder	Tin	Lead	Antimony	Solidus	Liquidus
*40/60	40	60	-	360°F (182°C)	460°F (238°C)
*60/40	60	40	-	360°F (182°C)	375°F (191°C)
*50/50	50	50	-	360°F (182°C)	420°F (216°C)
95/5	95	-	5	452°F (233°C)	464°F (240°C)

* 40/60, 60/40, 50/50 solders are available in solid, rosin or acid core.

BRIDGIT®

A patented, high-performance, lead-free solder developed by The Harris Products Group in response to the Federal ban on the use of lead solders in drinking water systems. This lead-free solder is specially formulated to fill both tight and loose-fitting connections. The nickel content in Bridgit® solder creates joints that are substantially stronger than those joints soldered with 50/50, 95/5 or common lead-free solders. Bridgit® has the ability to fill large gaps and "cap" joints with ease. Meets ASTM B32.

Extensive laboratory and field tests show that Bridgit® eliminates the problems associated with 95/5. Bridgit® begins to melt at 460°F, only 40° higher than 50/50. Bridgit® flux in a one pound spool provides 20% more wire than a roll of 1/8" diameter 50/50, resulting in 75 more soldered joints per roll.

Bridgit® paste flux and Bridgit® water soluble flux are certified to ANSI/NSF 61 and NSF/ANSI 372 for drinking water systems. Conforms to lead-free content requirements of the U.S. Safe Drinking Water Act. Certified to ANSI/NSF 51-Food Equipment Materials.

BRIDGIT® PASTE FLUX

Bridgit® flux is burn resistant, reducing black carbon formations that can cause leaks. This flux is unmatched for use in soldering copper, brass, bronze, galvanized, and other plumbing fittings. Works equally well with other solders.

BRIDGIT® WATER SOLUBLE FLUX

This flux is formulated so water flushing will remove flux residue from copper tube runs. Water soluble flux conforms to ASTM B813-93, and meets state and local regulations for use in potable water systems.

LEAD-FREE
BRIDGIT
NICKEL-BEARING SOLDER
Superior Strength!
FOR JOINING TIGHT & LOOSE FITTING WATER PIPE
HARRIS
SILVER BEARING SOFT SOLDER
NET WT. 1 lb (454 g)

SAFETY-SILV®

The Harris Products Group manufactures its complete line of cadmium-free, high silver brazing alloys with the same attention to quality found in their phosphorus / copper products. Only the purest metals are used. Precision production procedures ensure consistency in product quality and performance. Safety-Silv® alloys are strongly recommended as replacements for all cadmium-bearing brazing filler metals.

SAFETY-SILV® 56

This high silver (56%) content alloy makes first-quality brazes. It is free-flowing with unequaled capillary attraction and deep penetration. Ductility is high and corrosion resistance is suitable for all but strong chemical applications. NSF 51 listed for use with food equipment materials. The silver color is an excellent match to stainless steel and silverware applications.

SAFETY-SILV®45T

Performs like a 45% silver, cadmium-bearing alloy. Lower melting temperature than Safety-Silv® 45. Excellent fillet forming qualities. Produces high-strength, ductile joints. Listed with NSF 51 listed for use with food equipment materials.

SAFETY-SILV®45

Excellent general purpose nontoxic brazing alloy. Good ductility and capillary flow. Color is silver to light yellow matches polished brass.

HARRIS
SAFETY-SILV
45
3/32" (2.4 mm)
Made in USA
(31 g)

SAFETY-SILV®40

Ductile, free-flowing alloy offers economy, good penetration into tight connections, and medium temperature. Silver to light yellow color matches polished brass.

ESTIMATING AMOUNTS OF BRAZING ALLOYS REQUIRED

1. Locate the tube diameter to be joined and the wire size to be used. Where the row and the column intersect is the approximate length in inches of alloy required per joint.
2. Multiply the length of the alloy needed per joint by the total number of joints.
3. To convert the total length to pounds or troy ounces, divide by the inches of alloy/lb. in row A or the inches of alloy/troy oz. in row B.

ESTIMATING BRAZING ALLOY AMOUNTS

TUBE DIAMETER	3/64" WIRE	1/16" WIRE	3/32" WIRE	.050"x 1/8" ROD	TIP SIZE	ESTIMATED ACETYLENE USE (C.F.H.)
1/4"	1 1/4"	3/4"			4	6-14
3/8"	1 1/2"	1"			4	6-14
1/2"	2"	1 1/2"	3/4"	7/8"	5	8-18
3/4"	3"	2"	1"	1 1/8"	5	8-18
1"		3"	1 1/2"	1 5/8"	6	10-20
1 1/4"		4"	2"	2 1/2"	6	10-20
1 1/2"			2 1/2"	2 3/4"	7	13-25
2"			3 3/4"	4 1/2"	8	16-32
2 1/2"			6"	7 1/2"	8	16-32
3"			10"	11 1/2"	9	20-37
3 1/2"			12"	13 3/4"	9	20-37
4"			14"	16"	10	24-42
6"			21"	23 3/4"	10	24-42
A	1900"	1068"	475"	513"	in. of alloy/lb.	
B	118"	67"	29"		in. of alloy/troy oz.	

A- Phos/copper/silver alloys. Dynaflow®, Harris® 15, etc.
B- Silver Brazing alloys, Safety-Silv® 40, 45, 45T, 56
The above figures are approximate and will vary depending on joint clearance, depth, and operator technique.

HARRIS® 0

Low-cost alloy for many copper-to-copper applications where good fit-up can be maintained and brazing temperature is not critical.

STAY-SILV® 5 AND STAY-SILV® 6

Medium-range alloys, Stay-Silv® 5 is useful primarily where fit-up cannot be tightly controlled. Stay-Silv® 6 is slightly more fluid and can be used where closer tolerances are available. Both alloys are somewhat more ductile than Harris® 0.

DYNAFLOW®

Premium, medium-range silver alloy, formulated to even tighter specifications than the Stay-Silv® alloys to mirror the performance characteristics of the 15% silver brazing filler metals. Excellent for brazing both tight and poorly-fitted connections. Dynaflow's® proven reliability and acceptance by field service engineers has made it the leading choice of brazing operators.

STAY-SILV® 15

The industry standard for air conditioning and refrigeration applications. Still widely used, but now often replaced by Dynaflow® in many HVAC/R applications.

BLOCKADE®

Blockade's® unique silicon addition offers significant advantages over phos/copper and silver/phos/copper, (BCuP), brazing alloys. Among the benefits are: bright alloy color during cooling that allows visual verification of a good braze joint, improved ductility over BCuP-2 alloy, reduced brazing temperature, and reduced material costs. It is often used with Stay-Silv® white brazing flux as a replacement for high silver alloys when brazing brass.

HARRIS 0
PHOS-COPPER BRAZING ALLOY
Stay-Silv 5
PHOS-COPPER SILVER BRAZING ALLOY
Stay-Silv 6
Dynaflow
Stay-Silv 15
BLOCKADE
BRAZING ALLOY WITH SILICON FOR COPPER & BRASS

Phos/copper and silver/phos/copper alloys are used to braze copper-to-copper and copper-to-brass. The phosphorus content in these alloys makes them self-fluxing on copper. When brazing brass or copper to brass, use Stay-Silv® white brazing flux. These alloys are not recommended for brazing steel or other ferrous metals.

The amount of phosphorus in the phosphorus/copper filler metals (AWS BCuP series) is the critical factor that determines the melting range of the alloys and thus, how they perform. The Harris Products Group is the brazing industry's forerunner in developing the technology to control phosphorus content. This controls alloy melting temperatures so precisely that brazing operators no longer need to make temperature adjustments from one batch of filler metals to the next. Operators know that with Harris alloys, the result will be the same with every batch, every time.

Over the decades many things have changed in the industry. However, Harris' dedication to manufacturing the world's purest and most consistent brazing alloys has not changed. Harris is committed to providing the best products every time, all the time, so all jobs are completed successfully.

THE HARRIS PRODUCTS GROUP *guarantees users a liquidus temperature variation of no more than plus or minus 6˚F, a much tighter standard than the industry requires.*

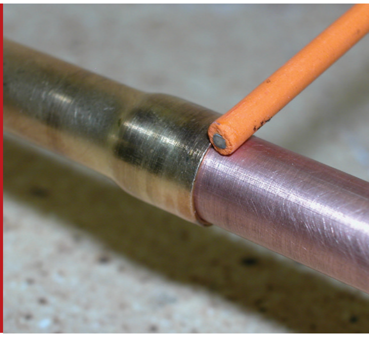

THE HARRIS PRODUCTS GROUP, MASON, OH 45040
HARRIS
ALCoR
FLUX-CORED
ALUMINUM BRAZING ALLOY
PT #AL200RC
2 mm DIA. x 36
READ WARNING ON REVERSE SIDE!
Al-Braze 1070
Al-Solder 500
Kit Que Suelda De Aluminio • Kit De Soudure
READ MSDS
Contents:
Hecho en EE.UU.-Fabriqué aux É.-U.

BRAZING AND SOLDERING ALUMINUM

Aluminum can be soldered or brazed if the correct procedure is followed. Pre cleaning is essential to break up the tough aluminum oxide film. Thorough brushing with a stainless steel wire brush is recommended. Most common aluminum alloys such as 1100 and 3003 can be readily soldered or brazed. It should be noted that some alloys are difficult to join. Dissimilar metal connections may be subject to galvanic corrosion in certain service conditions.

ALCOR®

Aluminum alloy with non-corrosive flux inside the wire, thus no external flux required. Very good fluidity with good capillary attraction. Post-braze cleaning unnecessary. Its low operating temperature often makes it a better choice than aluminum silicon alloys for aluminum coil repair.

AL-SOLDER®

Forms excellent corrosion-resistant joints on the tough-to-solder aluminum alloys. Joints all solderable aluminum alloys to each other and to dissimilar metals, both ferrous and non-ferrous. Also beneficial as a high-temperature solder on most other metals.

AL-BRAZE®

Superior brazing alloy for the joining of aluminum-to-aluminum. Not recommended for brazing aluminum directly to non-aluminum alloys, as the joint may become brittle. Al-Braze® is free flowing with unequaled capillary attraction, ductility and penetration. Excellent corrosion resistance.

ALUMINUM ALLOYS

PART #	DESCRIPTION	CHEMICAL COMP.	SOLIDUS	LIQUIDUS	TYPICAL APPLICATION
Al200RC	Alcor® Flux-Cored Aluminum Alloy	Zn Al	824°F 440°C	824°F 440°C	A new approach to joining aluminum. A low temperature, free-flowing, flux-cored solder for aluminum joining or repair.
500K	Al-Solder® 500 Aluminum Solder Kit	15% Zn 85% Sn	391°F 199°C	482°F 250°C	Forms excellent corrosion-resistant joints on the tough-to-solder aluminum alloys. Use with copper-to-aluminum connections.
1070K	Al-Braze® 1070 Aluminum Brazing Kit	88% Al 12% Si	1070°F 577°C	1080°F 582°C	Superior brazing alloy for joining aluminum-to-aluminum. Excellent capillary attraction.

Safety-Silv®56

Safety-Silv®45 (Flux-Coated)

Safety-Silv®40

SAFETY-SILV® 56

CHEMICAL COMPOSITION	SPECIFICATIONS	SOLIDUS	LIQUIDUS	TYPICAL APPLICATION
56 % Ag 22 % Cu 17 % Zn 5% Sn	AWS A5.8 BAg-7 NSF 51	1145°F 618°C	1205°F 652°C	For ferrous and nonferrous alloys. Often used to braze stainless steel for use with food equipment materials.

SAFETY-SILV® 45

CHEMICAL COMPOSITION	SPECIFICATIONS	SOLIDUS	LIQUIDUS	TYPICAL APPLICATION
45 % Ag 30 % Cu 25 % Zn	AWS A5.8 BAg-5	1225°F 663°C	1370°F 743°C	General purpose filler for steel and copper alloys. Melting range useful for wide clearances.

SAFETY-SILV® 40

CHEMICAL COMPOSITION	SPECIFICATIONS	SOLIDUS	LIQUIDUS	TYPICAL APPLICATION
40% Ag 30.5 % Cu 29.5 % Zn	Harris Internal	1250°F 677°C	1350°F 732°C	Ductile, free-flowing alloy that offers economy, good penetration into tight connections and medium temperature.

Stay-Brite®

50/50

60/40

STAY-BRITE® SOLDERS

CHEMICAL COMPOSITION	SPECIFICATIONS	SOLIDUS	LIQUIDUS	TYPICAL APPLICATION
4 % Ag 96 % Sn	ASTM B32 Sn96 NSF 51 J-STD-006 Sn96 Ag 04A	430°F 221°C	1205°F 652°C	Use with all metals except aluminum. Low temperature solder used in HVAC/R industry.

50/50 SOLDERS

CHEMICAL COMPOSITION	SPECIFICATIONS	SOLIDUS	LIQUIDUS	TYPICAL APPLICATION
50 % Sn 50 % Pb	ASTM B32 Sn50 J-STD-006 Sn50pB50a	360°F 182°C	420°F 216°C	Used to join copper to most copper, lead, nickel, and steel, with some exceptions. Note: Illegal for potable water systems

60/40 SOLDERS

CHEMICAL COMPOSITION	SPECIFICATIONS	SOLIDUS	LIQUIDUS	TYPICAL APPLICATION
60 % Sn 40 % Pb	ASTM B32 Sn60	360°F 182°C	375°F 191°C	Similar to 50/50 but flows faster due to narrow melting range. Note: Illegal for potable water systems

Other solders and production packs available, please contact customer service.

FLUXES

DESCRIPTION	SPECIFICATION	ACTIVE TEMP	TYPICAL APPLICATION
STAY-SILV® WHITE BRAZING FLUX 7 oz., 1/4 lb., 1/2 lb.,1 lb., 5 lbs., 25 lbs., 60 lbs.	Meets Federal Spec. (OF499,Type B AWS A5.31) Class FB3-A AMS 3410	1050-1600°F (565 - 870° C)	All-purpose low temperature brazing flux used to braze most ferrous and non-ferrous metal, (exceptions are aluminum bronze, titanium, magnesium, and aluminum).
STAY-SILV® BLACK BRAZING FLUX 1 lb., .5 lbs., 5 lbs., 30 lbs., 60 lbs.	Meets Federal Spec. (OF499,Type B AWS A5.31) Class FB3-C AMS 3411	1050 -1700°F (565 - 926°C)	A boron modified formula designed for use where heating cycles are prolonged or where there is concentrated local heat as with induction brazing. Also effective in brazing stainless steel and tungsten carbide.
STAY-CLEAN® LIQUID SOLDERING FLUX 1 gl., 4 oz., 16 oz., 32 oz., 55 gl.	Meets Commercial Spec. A-A-51145 Type II Form B	300 - 700°F (148 - 371°C)	An active flux for soldering copper, brass, steel, nickel, and stainless steel. It can be used effectively with Stay-Brite® and most other solders. Remove flux residue after soldering.
STAY-CLEAN® PASTE SOLDERING FLUX 4 oz.	Meets Commercial Spec. A-A-51145 Type I Form A	300 - 600° (148 - 315°C)	Excellent flux for joining copper-to-copper and copper-to-brass. Not recommended for electrical or electronic applications.
BRIDGIT® HIGH TEMPERATURE PASTE SOLDERING FLUX 4 oz.	NSF/ANSI 61	300 – 700°F (148 - 371°C)	Designed for lead-free solders. Well suited for use in larger connections where prolonged heating will cause other fluxes to burn.
BRIDGIT® WATER SOLUBLE SOLDERING FLUX 4 oz.	ASTM B813 NSF/ANSI 61	350 - 600°F (176 - 316°C)	This flux is formulated so water flushing will remove flux residue from copper tube runs. The flux meets state and local regulations for use in potable water systems.

HARRIS
THE HARRIS PRODUCTS GROUP
Stay-Silv®
White brazing flux
Fundente de soldadura blanco
stay-clean®
LIQUID • LIQUIDO • LIQUIDE
Soldering & Tinning Flux
Fundente para Soldar y Estañar
Flux décapant et de mouillage
HARRIS
CONTENTS: 16 fl. oz. (.473 L)
BRIDGIT
NET WT. 4 oz. (113.4 g)
Stay-Silv®
NET WT. 1 lb (454 g)

PHOS-COPPER BRAZING ALLOYS

DESCRIPTION	CHEMICAL COMPOSITION	SPECIFICATIONS	SOLIDUS	LIQUIDUS	TYPICAL APPLICATION
STAY-SILV® 15 	15% Ag 5% P 80% Cu	AWS A5.8 BCuP-5	1190°F 643°C	1475°F 802°C	Designed primarily for copper to copper brazing application, it may also be used in brazing brass with the use of Stay-Silv® brazing flux.
STAY-SILV® 5 	5% Ag 6% P 89% Cu	AWS A5.8 BCuP-3	1190°F 643°C	1500°F 816°C	Well suited where close fit-up cannot necessarily be maintained.
DYNAFLOW® 	6% Ag 6% P 88% Cu	--------	1190°F 643°C	1465°F 796°C	Designed for copper-to-copper applications, it may also be used on brass. Provides the ability to braze poorly fitted as well as tight connections.
HARRIS® 0 	7.2% P 92.8% Cu	AWS A5.8 BCuP-5	1310°F 710°C	1460°F 793°C	Designed for copper-to-copper. Suitable where tight tolerances can be maintained. Harris®0 has a higher solidus temperature so more heat is required, but alloy is fluid at brazing temperature.

LEAD-FREE SOLDER

DESCRIPTION	CHEMICAL COMPOSITION	SPECIFICATIONS	SOLIDUS	LIQUIDUS	TYPICAL APPLICATION
BRIDGIT® SOLDER 	-------	ASTM B32 Alloy HB NSF 61	460°F 238°C	630°F 332°C	Widely used in plumbing applications where lead-bearing solders are prohibited. Contains nickel, making joints tremendously strong. Wide plastic range makes Bridgit® an excellent alloy for large diameter fittings and ill-fitted or non-concentric pipes. Fills gaps and caps off easily and effectively.
STAY-BRITE® SOLDER 	4% Ag 96% Sn	ASTM B32 SN96 NSF 51• J-STD-006 SN96AG04A	430°F 221°C	430°F 221°C	Use for all metals with the exception of aluminum. Low temperature solder often used in residential HVAC. Flows quickly to penetrate tight clearances. Use Stay-Brite® 8 where wider gaps must be filled.
95/5 	95% Sn 5% Sb	ASTM B32 • Sb5 J-STD-006 Sn95Sb05A NSF 61	452°F 233°C	464°F 240°C	Well-suited for applications where moderately elevated temperature is a factor. Requires close part tolerance due to it's narrow melting range.

FILLER METALS SELECTION CHART

METALS TO BE JOINED	SOLDERS	BRAZING FILLER METALS	SOLIDUS	LIQUIDUS	FLUIDITY RATING*	FLUX	TORCHES & FLAMES
COPPER OR BRASS TO COPPER OR BRASS	Stay-Brite® Stay-Brite® 8 Bridgit®		430°F / 221°C 430°F / 221°C 460°F / 238°C	430°F / 221°C 535°F / 279°C 630°F / 332°C	10 8 6	Stay-Clean® soldering fluxes, liquid or paste. Bridgit® paste or water soluble paste flux.	Harris Inferno® Air-fuel equipment
		Blockade® Harris® 0 Stay-Silv® 5 Dynaflow® Stay-Silv® 6 Stay-Silv® 15	1178°F / 637°C 1310°F / 710°C 1190°F / 643°C 1190°F / 643°C 1190°F / 643°C 1190°F / 643°C	1247°F / 674°C 1475°F / 802°C 1500°F / 816°C 1465°F / 796°C 1425°F / 774°C 1480°F / 804°C	7 5 3 3 5 3	No flux required for copper to copper joints with the phosphorus-bearing filler metals. For brass and other copper alloys, use Stay-Silv® white brazing flux.	Harris Inferno® Air-fuel equipment or Harris oxy-acetylene equipment (Neutral flame)
COPPER OR BRASS TO STEEL OR STAINLESS STEEL	Stay-Brite® Stay-Brite® 8		430°F / 221°C 430°F / 221°C	430°F / 221°C 535°F / 279°C	10 8	Stay-Clean® liquid soldering flux.	Harris Inferno® Air-fuel equipment
		Safety-Silv® 56 Safety-Silv® 38T Safety-Silv® 45 Safety-Silv® 45T	1145°F / 618°C 1120°F / 660°C 1125°F / 663°C 1195°F / 646°C	1205°F / 652°C 1325°F / 718°C 1370°F / 743°C 1265°F / 685°C	8 7 6.5 7	Stay-Silv® white brazing flux. Stay-Silv® black flux for stainless steel.	Harris Inferno® Air-fuel equipment or Harris oxy-acetylene equipment (Slightly reducing flame)
STEEL OR STAINLESS STEEL TO STEEL OR STAINLESS STEEL	Stay-Brite® Stay-Brite®		430°F / 221°C 430°F / 221°C	430°F / 221°C 535°F / 279°C	10 8	Stay-Clean® liquid soldering flux.	Harris Inferno® Air-fuel equipment
		Safety-Silv® 56 Safety-Silv® 38T Safety-Silv® 40Ni2 Safety-Silv® 45 Safety-Silv® 45T Safety-Silv® 50N	1145°F / 618°C 1220°F / 660°C 1220°F / 660°C 1225°F / 663°C 1195°F / 646°C 1220°F / 660°C	1205°F / 652°C 1325°F / 718°C 1435°F / 779°C 1370°F / 743°C 1265°F / 685°C 1305°F / 707°C	8 7 4.5 6.5 7 7	Stay-Silv® white brazing flux. Stay-Silv® black flux for stainless steel.	Harris Inferno® Air-fuel equipment or Harris oxy-acetylene equipment (Slightly reducing flame)
STEEL OR STAINLESS STEEL TO CARBIDES	Not Recommended	Safety-Silv® 40Ni2 Safety-Silv® 50N	1220°F / 660°C 1220°F / 660°C	1435°F / 779°C 1305°F / 707°C	4.5 7	Stay-Silv® black flux for stainless steel.	Harris oxy-acetylene equipment (Reducing flame)
ALUMINUM-TO-ALUMINUM (1) OR ALUMINUM- TO-COPPER OR BRASS (2) OR ALUMINUM-TO-STEEL OR STAINLESS STEEL (2)	Al-Solder® 500 Alcor®		391°F / 199°C ---	482°F / 250°C 824°F / 440°C	---- ----	Stay-Clean® aluminum soldering flux. Flux is contained inside the wire.	Harris Inferno® Air-fuel equipment
		Al-Braze® 4043 Al-Braze® 1070	1065°F / 574°C 1070°F / 577°C	1170°F / 632°C 1080°F / 582°C	7 9	Al-Braze® 1070 flux	Harris Inferno® Air-fuel equipment or Harris Oxy-acetylene equipment (Reducing flame)

NOTE: ALUMINUM TO DISSIMILAR METAL JOINTS MAY BE SUBJECT TO GALVANIC CORROSION

*The higher the fluidity rating the faster the alloy flows within the melting range

HARRIS INFERNO®

Harris Inferno® air-fuel equipment allows for soldering and brazing of a variety of tube sizes. Equipment attaches to a B (40 cu. ft.) or MC (10 cu. ft.) acetylene cylinder.

The Inferno® tip provides a special swirl combustion approach to increase flame velocity and homogenize the gas mixture.

For air-acetylene, fuel gas delivery pressure should be set at 14-15 PSI. When soldering or brazing different size connections, change tip size rather than reducing flow. The air fuel flame provides a wider heating pattern than oxy-acetylene. This is helpful in soldering or brazing larger tube diameters as heat tends to wrap around the pipe.

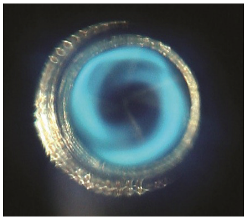

SWIRL COMBUSTION TECHNOLOGY

The revolutionary Inferno® tip has a specially designed insert that delivers reliable swirl combustion performance every time. The swirl combustion of the Inferno® tip is unlike anything offered on the market. Contractors get a consistent, hotter flame that will engulf and wrap around the work piece for maximum efficiency.

AIR-FUEL FLAME PATTERN

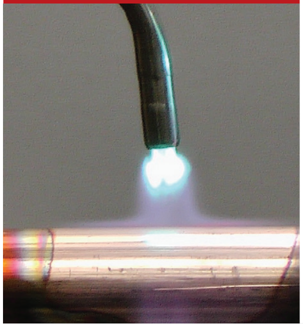

OXY-FUEL FLAME PATTERN

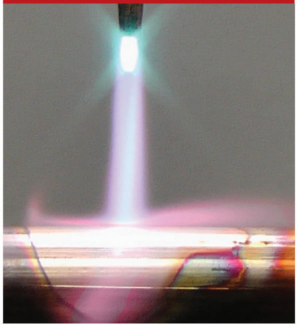

Scan the QR code for more information.

HARRIS PORT-A-TORCH®

The Harris Port-A-Torch® contains all the quality equipment needed for cutting, welding and brazing packaged in a rugged, molded plastic carrying case. The outfit is designed to carry one MC acetylene cylinder and one 20 cu. ft. oxygen cylinder. As supplied, the outfit is capable of cutting up to a 1” plate and welding up to a 1⁄16”plate . The outfit can cut up to a 4” plate and weld up to a 1⁄2” plate with larger tips and acetylene cylinder.

An oxy-acetylene flame produces the highest flame temperature with a more focused heat zone. Optimum delivery pressure settings for this gas mixture depends on the tip size. Refer to the tip chart for guidance.

Acetylene Welding/Brazing Tips

METAL THICKNESS	SIZE	23A90 PART #	OXYGEN PSIG	ACETYLENE PSIG
1/64”	0	1600840	1	1
1/32”	1	1600850	1	1
3/64”	2	1600860	2	2
1/16”	3	1600840	3	3
3/32”	4	1600840	4	4
1/8”	5	1600840	5	5
3/16”	6	1600840	6	6
1/4”	7	1600840	7	7
5/16”	8	1600840	8	8
3/8”	9	1600840	9	9
1/2”	10	1600840	10	10

port-a-torch
HVAC

NOTES:

A LINCOLN ELECTRIC COMPANY

The Harris Products Group
4501 Quality Place, Mason, OH 45040-1971
513-754-2000
www.harrisproductsgroup.com

ISO 9001 Registered
PN: 9700016 REV B
092014

UNO
ARDUINO
DIGITAL (PWM~)

HACKING ELECTRONICS

An illustrated DIY guide for makers and hobbyists

TAB
Make Great Stuff!

Simon Monk

UNO
ARDUINO
DIGITAL (PWM~)

HACKING ELECTRONICS

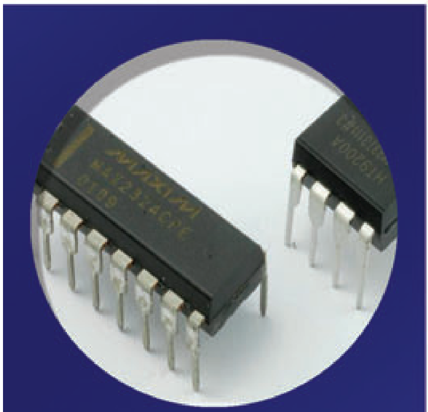

About the Author

Dr. Simon Monk (Preston, UK) has a degree in Cybernetics and Computer Science and a PhD in Software Engineering. Monk spent several years as an academic before he returned to industry, co-founding the mobile software company Momote Ltd. He has been an active electronics hobbyist since his early teens and is a full-time writer on hobby electronics and open-source hardware. Dr. Monk is the author of numerous electronics books, specializing in open-source hardware platforms, especially Arduino and Raspberry Pi. He is also co-author with Paul Scherz of *Practical Electronics for Inventors,* 3rd edition. You can follow Simon on Twitter, where he is @simonmonk2.

Hacking Electronics

An Illustrated DIY Guide for Makers and Hobbyists

Simon Monk

Mc Graw Hill Education

New York Chicago San Francisco Lisbon
London Madrid Mexico City Milan New Delhi
San Juan Seoul Singapore Sydney Toronto

Copyright © 2013 by The McGraw-Hill Companies. All rights reserved. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

ISBN: 978-0-07-180237-6

MHID: 0-07-180237-1

The material in this e-book also appears in the print version of this title: ISBN: 978-0-07-180236-9,
MHID: 0-07-180236-3

McGraw-Hill e-books are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. To contact a representative please e-mail us at bulksales@mcgraw-hill.com.

All trademarks are trademarks of their respective owners. Rather than put a trademark symbol after every occurrence of a trademarked name, we use names in an editorial fashion only, and to the benefit of the trademark owner, with no intention of infringement of the trademark. Where such designations appear in this book, they have been printed with initial caps.

Information has been obtained by McGraw-Hill from sources believed to be reliable. However, because of the possibility of human or mechanical error by our sources, McGraw-Hill, or others, McGraw-Hill does not guarantee the accuracy, adequacy, or completeness of any information and is not responsible for any errors or omissions or the results obtained from the use of such information.

TERMS OF USE

This is a copyrighted work and McGraw-Hill and its licensors reserve all rights in and to the work. Use of this work is subject to these terms. Except as permitted under the Copyright Act of 1976 and the right to store and retrieve one copy of the work, you may not decompile, disassemble, reverse engineer, reproduce, modify, create derivative works based upon, transmit, distribute, disseminate, sell, publish or sublicense the work or any part of it without McGraw-Hill's prior consent. You may use the work for your own noncommercial and personal use; any other use of the work is strictly prohibited. Your right to use the work may be terminated if you fail to comply with these terms.

THE WORK IS PROVIDED "AS IS." THE McGRAW-HILL COMPANIES AND ITS LICENSORS MAKE NO GUARANTEES OR WARRANTIES AS TO THE ACCURACY, ADEQUACY OR COMPLETENESS OF OR RESULTS TO BE OBTAINED FROM USING THE WORK, INCLUDING ANY INFORMATION THAT CAN BE ACCESSED THROUGH THE WORK VIA HYPERLINK OR OTHERWISE, AND EXPRESSLY DISCLAIM ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. McGraw-Hill and its licensors do not warrant or guarantee that the functions contained in the work will meet your requirements or that its operation will be uninterrupted or error free. Neither McGraw-Hill nor its licensors shall be liable to you or anyone else for any inaccuracy, error or omission, regardless of cause, in the work or for any damages resulting therefrom. McGraw-Hill has no responsibility for the content of any information accessed through the work. Under no circumstances shall McGraw-Hill and/or its licensors be liable for any indirect, incidental, special, punitive, consequential or similar damages that result from the use of or inability to use the work, even if any of them has been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause arises in contract, tort or otherwise.

To Roger, for making it possible for me to turn a hobby into an occupation.

This page intentionally left blank

Contents at a Glance

1 Getting Started 1
2 Theory and Practice 19
3 Basic Hacks 33
4 LEDs 55
5 Batteries and Power 83
6 Hacking Arduino 105
7 Hacking with Modules 149
8 Hacking with Sensors 193
9 Audio Hacks 213
10 Mending and Breaking Electronics 235
11 Tools 245
A Parts 257

Index 263

This page intentionally left blank

Contents

Acknowledgments xix
Introduction xxi

CHAPTER 1 **Getting Started** **1**

Getting Stuff 1
    Buying Components 1
    Where to Buy Things to Hack 2
    A Basic Toolkit 3
How to Strip a Wire 5
    You Will Need 5
How to Join Wires Together by Twisting 7
    You Will Need 7
How to Join Wires by Soldering 8
    Safety 8
    You Will Need 9
    Soldering 10
    Joining Wires 11
How to Test a Connection 12
    You Will Need 12
How to Hack a Computer Fan to Keep Soldering Fumes Away 14
    You Will Need 14
    Construction 14
Summary 18

CHAPTER 2 **Theory and Practice** **19**

How to Assemble a Starter Kit of Components 19
    You Will Need 19
How to Identify Electronic Components 20
    Resistors 20
    Capacitors 22
    Diodes 23
    LEDs 23

Transistors 24
Integrated Circuits 24
Other Stuff 25
Surface Mount Components 25
What Are Current, Resistance, and Voltage? 25
Current 26
Resistance 26
Voltage 26
Ohm's Law 27
What Is Power? 28
How to Read a Schematic Diagram 29
The First Rule of Schematics: Positive Voltages Are Uppermost 29
Second Rule of Schematics: Things Happen Left to Right 30
Names and Values 30
Component Symbols 30
Summary 31

CHAPTER 3 Basic Hacks 33

How to Make a Resistor Get Hot 33
You Will Need 33
The Experiment 33
How to Use Resistors to Divide a Voltage 34
You Will Need 34
How to Convert a Resistance to a Voltage (and Make a Light Meter) 37
You Will Need 37
Hack a Push Light to Make It Light Sensing 39
You Will Need 39
Breadboard 40
Construction 41
How to Choose a Bipolar Transistor 45
Datasheets 45
MOSFET Transistors 46
PNP and N-Channel Transistors 46
Common Transistors 47
How to Use a Power MOSFET to Control a Motor 48
You Will Need 48
Breadboard 48
How to Select the Right Switch 50
Push-Button Switches 50
Microswitches 51
Toggle Switches 51
Summary 53

CHAPTER 4 **LEDs** **55**
How to Stop an LED from Burning Out 55
You Will Need 55
Diodes 56
LEDs 56
Trying It Out 58
How to Select the Right LED for the Job 59
You Will Need 59
Brightness and Angle 59
Multicolor 60
IR and UV 60
LEDs for Illumination 61
How to Use a LM317 to Make a Constant Current Driver 62
You Will Need 62
Design 63
Breadboard 64
Construction 65
How to Measure the Forward Voltage of an LED 66
You Will Need 67
How to Power Large Numbers of LEDs 68
How to Make LEDs Flash 69
You Will Need 69
Breadboard 69
How to Use Stripboard (LED Flasher) 71
Designing the Stripboard Layout 71
You Will Need 73
Construction 74
Troubleshooting 77
How to Use a Laser Diode Module 78
Hacking a Slot Car Racer 78
You Will Need 78
Storing Charge in a Capacitor 79
Design 80
Construction 81
Testing 81
Summary 82

CHAPTER 5 **Batteries and Power** **83**
Selecting the Right Battery 83
Battery Capacity 83
Maximum Discharge Rate 84
Single-Use Batteries 84
Rechargeable Batteries 86

Charging Batteries (in General) 88
  C 88
  Over-Charging 89
  Over-Discharging 89
  Battery Life 89
How to Charge a NiMH Battery 89
  Simple Charging 90
  Fast Charging 91
How to Charge a Sealed Lead–Acid Battery 91
  Charging with a Variable Power Supply 91
How to Charge a LiPo Battery 92
Hacking a Cell Phone Battery 93
Controlling the Voltage from a Battery 95
  You Will Need 96
  Breadboard 97
Boosting Voltage 97
Calculating How Long a Battery Will Last 98
How to Design for Battery Backup 99
  Diodes 99
  Trickle Charging 100
How to Use Solar Cells 101
  Testing a Solar Panel 102
  Trickle Charging with a Solar Panel 103
  Minimizing Power Consumption 104
Summary 104

CHAPTER 6 **Hacking Arduino 105**
How to Set up Arduino (and Blink an LED) 106
  You Will Need 106
  Setting Up Arduino 106
  Modifying the Blink Sketch 109
How to Make an Arduino Control a Relay 112
  Relays 112
  Arduino Outputs 112
  You Will Need 114
  Construction 114
  Software 115
How to Hack a Toy for Arduino Control 116
  You Will Need 116
  Construction 116
  The Serial Monitor 118
  Software 118

How to Measure Voltage with an Arduino 119
- You Will Need 120
- Construction 120
- Software 120

How to Use an Arduino to Control an LED 122
- You Will Need 122
- Construction 122
- Software (Flashing) 123
- Software (Brightness) 124

How to Play a Sound with an Arduino 125
- You Will Need 125
- Construction 126
- Software 126

How to Use Arduino Shields 127

How to Control a Relay from a Web Page 128
- You Will Need 129
- Construction 130
- Network Configuration 131
- Testing 131
- Software 132

How to Use a Alphanumeric LCD Shield with Arduino 136
- You Will Need 137
- Construction 137
- Software 137

How to Drive a Servo Motor with an Arduino 139
- You Will Need 140
- Construction 140
- Software 140

How to Charlieplex LEDs 142
- You Will Need 143
- Construction 143
- Software 143

How to Type Passwords Automatically 145
- You Will Need 146
- Construction 146
- Software 146

Summary 147

CHAPTER 7 **Hacking with Modules** **149**

How to Use a PIR Motion Sensor Module 149
- You Will Need (PIR and LED) 150
- Breadboard 150
- You Will Need (PIR and Arduino) 151

Construction 151
Software 152
How to Use Ultrasonic Rangefinder Modules 153
You Will Need 154
The HC-SR04 Rangefinder 155
The MaxBotix LV-EZ1 Rangefinder 157
How to Use a Wireless Remote Module 159
You Will Need 159
Breadboard 159
How to Use a Wireless Remote Module with Arduino 161
You Will Need 161
Software 161
How to Control Motor Speed with a Power MOSFET 163
You Will Need 163
Breadboard 164
Software 165
How to Control DC Motors with an H-Bridge Module 166
You Will Need 170
Breadboard 170
Using the Control Pins 171
How to Control a Stepper Motor with an H-Bridge Module 172
You Will Need 174
Construction 175
Software 175
How to Make a Simple Robot Rover 177
You Will Need 178
Construction 179
Testing 181
Software 182
How to Use a Seven-Segment LED Display Module 183
You Will Need 185
Construction 186
Software 187
How to Use a Real-Time Clock Module 188
You Will Need 189
Construction 190
Software 191
Summary 192

CHAPTER 8 **Hacking with Sensors** **193**
How to Detect Noxious Gas 193
You Will Need 193
The LM311 Comparator 194

    Breadboard . . . 195
    Using a Gas Sensor with an Arduino . . . 196
How to Measure Something's Color . . . 197
    You Will Need . . . 198
    Construction . . . 199
    Software . . . 199
How to Detect Vibration . . . 202
    You Will Need . . . 202
    Construction . . . 202
    Software . . . 203
How to Measure Temperature . . . 204
    You Will Need . . . 205
    Construction . . . 205
    Software . . . 205
How to Use an Accelerometer . . . 206
    You Will Need . . . 207
    Construction . . . 207
    Software . . . 208
How to Sense Magnetic Fields . . . 210
    You Will Need . . . 211
    Construction . . . 211
    Software . . . 212
Summary . . . 212

CHAPTER 9 **Audio Hacks . . . 213**
Hacking Audio Leads . . . 213
    General Principals . . . 213
    Soldering Audio Connectors . . . 215
    Converting a Stereo Signal to Mono . . . 217
How to Use a Microphone Module . . . 218
How to Make an FM Bug . . . 220
    You Will Need . . . 221
    Construction . . . 221
    Testing . . . 223
Selecting Loudspeakers . . . 223
How to Make a 1-Watt Audio Amplifier . . . 224
    You Will Need . . . 225
    Construction . . . 226
    Testing . . . 227
How to Generate Tones with a 555 Timer . . . 227
    You Will Need . . . 229
    Construction . . . 229

How to Make a USB Music Controller 229
    You Will Need 230
    Construction 230
    Software 230
How to Make a Software VU Meter 232
    You Will Need 233
    Construction 233
    Software 233
Summary 234

CHAPTER 10 **Mending and Breaking Electronics** **235**
How to Avoid Electrocution 235
How to Take Something Apart AND Put It Back Together Again 236
How to Check a Fuse 237
How to Test a Battery 239
How to Test a Heating Element 240
Finding and Replacing Failed Components 240
    Testing Components 240
    Desoldering 241
    Replacement 242
How to Scavenge Useful Components 242
How to Reuse a Cell Phone Power Adapter 243
Summary 244

CHAPTER 11 **Tools** **245**
How to Use a Multimeter (General) 245
    Continuity and Diode Test 245
    Resistance 246
    Capacitance 247
    Temperature 247
    AC Voltage 248
    DC Voltage 249
    DC Current 249
    AC Current 250
    Frequency 250
How to Use a Multimeter to Test a Transistor 250
How to Use a Lab Power Supply 251
Introducing: The Oscilloscope 252
Software Tools 253
    Simulation 253
    Fritzing 253
    EAGLE PCB 254
    Online Calculators 256
Summary 256

Appendix **Parts** **257**
Tools 257
Components 258
  Component Starter Kits 258
  Resistors 258
  Capacitors 259
  Semiconductors 259
  Hardware and Miscellaneous 260
  Modules 261

Index **263**

This page intentionally left blank

Acknowledgments

Many thanks to all those at McGraw-Hill Education who have done such a great job in producing this book. In particular, thanks to my editor Roger Stewart and to Vastavikta Sharma, Jody McKenzie, Mike McGee, and Claire Splan.

Special thanks are due to Duncan Amos, John Heath, and John Hutchinson for their technical review of the material and encouragement.

And last but not least, thanks once again to Linda, for her patience and generosity in giving me space to do this.

This page intentionally left blank

Introduction

This is a book about "hacking" electronics. It is not a formal, theory-based book about electronics. Its sole aim is to equip the reader with the skills he or she needs to use electronics to make something, whether it's starting from scratch, connecting together modules, or adapting existing electronic devices for some new use.

You will learn how to experiment and get your ideas into some kind of order, so that what you make will work. Along the way, you'll gain an appreciation for why things work and the limits of what they can do, and learn how to make prototypes on solderless breadboard, how to solder components directly to each other, and how to use stripboard.

You will also learn how to use the popular Arduino microcontroller board, which has become one of the most important tools available to the electronics hacker. There are over 20 examples of how to use an Arduino with electronics in this book.

Electronics has changed. This is a modern book that avoids theory you will likely never use and instead concentrates on how you can build things using readymade modules when they are available. There is, after all, no point in reinventing the wheel.

Some of the things explained and described in the book include

- Using LEDs, including high-power Lumileds
- Using LiPo battery packs and buck-boost power supply modules
- Using sensors to measure light, temperature, vibration, acceleration, sound level, and color
- Interfacing with Arduino microcontroller boards, including using Arduino shields such as the Ethernet and LCD display shields
- Using servo and stepper motors

Some of the things described in the book that you can make along the way include

- A noxious gas detector
- An Internet-controlled hacked electric toy
- A device for measuring color
- An ultrasonic rangefinder
- A remote control robotic rover
- An accelerometer-based version of the “egg and spoon” race
- A one-watt audio amplifier
- A bug made from a hacked MP3 FM transmitter
- Working brakes and head lights that can be added to a slot car

You Will Need

This is a very practical, hands-on type of book. You will therefore need some tools and components to get the most out of it.

As far as tools go, you will need little more than a multimeter and soldering equipment.

When it comes to areas of electronics where a microcontroller would be useful, an Arduino Uno board is best. So you may wish to buy one of these microcontroller boards before attempting some of the projects.

Every component used in this book is listed in the Appendix, along with sources where it can be obtained. The majority of the components can be found in a starter kit from SparkFun, but most electronic starter kits will provide a lot of what you will need.

In many of the “how-tos,” there will be a You Will Need section. This will refer to a code in the Appendix that explains where to get the component.

How to Use This Book

The book is organized into chapters, each of which has a theme. Within each chapter, most of the numbered sections contain a "how-to" on some topic of electronics.

The book contains the following chapters:

Chapter	Title	Description
Chapter 1	Getting Started	The book starts off by telling you where you can buy equipment and components, as well as things to hack. This chapter also deals with the basics of soldering and focuses on a project to hack an old computer fan to make a fume extractor for use while soldering.
Chapter 2	Theory and Practice	This chapter introduces electronic components—or at least the ones you are likely to use—and explains how to identify them and describes what they do. It also introduces a small amount of essential theory, which you will use over and over again.
Chapter 3	Basic Hacks	This chapter contains a set of fairly basic "hacking" how-tos, introducing concepts like using transistors with example projects. It includes hacking a "push light" to make it automatically turn on when it gets dark and how to control a motor using power MOSFETs.
Chapter 4	LEDs	In addition to discussing regular LEDs and how to use them and make them flash and so on, this chapter also looks at using constant current drivers for LEDs and how to power large numbers of LEDs and laser diode modules.
Chapter 5	Batteries and Power	This chapter discusses the various types of battery, both single use and rechargeable. It also covers how to charge batteries including LiPos. Automatic battery backup, voltage regulation, and solar charging are also explained.
Chapter 6	Hacking Arduino	The Arduino has become the microcontroller board of choice for electronics hackers. Its open-source hardware design makes using a complex device like a microcontroller very straightforward. The chapter gets you started with the Arduino and includes a few simple how-tos, like controlling a relay, playing sounds, and controlling servo motors from an Arduino. It also covers the use of Arduino expansion shields.

Chapter	Title	Description
Chapter 7	Hacking with Modules	When you want to make something, you can often use readymade modules at least for part of the project. Modules exist for all sorts of things, from wireless remotes to motor drivers.
Chapter 8	Hacking with Sensors	Sensor ICs and modules are available for sensing everything from gas to acceleration. In this chapter, we explore a good range of them and explain how to use them and connect some of them to an Arduino.
Chapter 9	Audio Hacks	This chapter has a number of useful how-tos relating to electronics and sound. It includes making and adapting audio leads, as well as audio amplifiers, and discusses the use of microphones.
Chapter 10	Mending and Breaking Electronics	Mending electronics and scavenging useful parts from dead electronics are a worthy activity for the electronics hacker. This chapter explains how to take things apart and sometimes put them back together again.
Chapter 11	Tools	The final chapter of the book is intended as a reference to explain more about how to get the most out of tools such as multimeters and lab power supplies.

This page intentionally left blank

HACKING ELECTRONICS

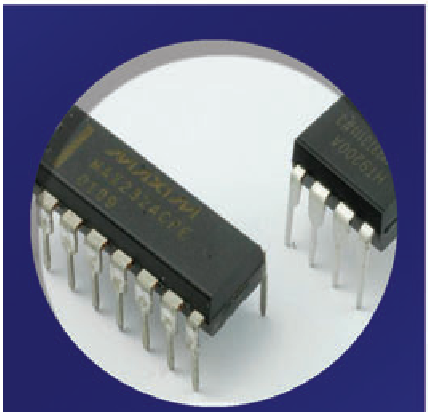

1

Getting Started

In this first chapter, we will investigate some of the tools and techniques needed to hack electronics. We will start with a little soldering, and wire up an old computer fan to help keep the solder fumes out of our lungs.

As it says in the title, this book is all about "hacking electronics." The word "hacking" has come to mean many things. But in this book, "hacking" means "just do it!" You don't need a degree in electronic engineering to create or modify something electronic. The best way to learn is by having a go at it. You will learn as much from your mistakes as from your successes.

As you start to make things and experiment, you will likely want to understand more of the theory behind it all. Traditional electronics textbooks are pretty terrifying unless you have a good grasp of complex mathematics. This book strives to, above all else, enable you to do things first and worry about the theory later.

To get started, you will need some tools, and also find out where to get components and parts to use in your projects.

Getting Stuff

In addition to buying components and tools, there are lots of low-cost and interesting electronic consumer items that can be hacked and used for new purposes, or that can act as donors of interesting components.

Buying Components

Most component purchases happen on the Internet, although there are local electronic stores like RadioShack (in the U.S.) and Maplin (in the UK) where you can buy components. At traditional brick-and-mortar stores like those, the product range is often limited and the prices can be on the high side. They do, after all, have a shop to pay for. These stores are invaluable, however, on the odd occasion where you need something in a hurry. Perhaps you need an LED because you accidentally destroyed one, or maybe you want to look at the enclosures they sell for projects.

Sometimes it's just nice to hold a box or look at tools for real, rather than trying to size them up from pictures on a web site.

As you get into electronics, you will likely gradually accumulate a set of components and tools that you can draw from when you start a new project. Components are relatively cheap, so when I need one of something, I generally order two or three or even five if they are cheap, enough that I have extras on hand that can be used another time. This way, you will often find that when you start to work on something, you actually have pretty much everything you need already.

Component buying really depends on where you are in the world. In the U.S., Mouser and DigiKey are the largest suppliers of electronic components to the hobby electronics market. In fact, both of these suppliers sell worldwide. Farnell also supplies pretty much anything you could want, anywhere in the world.

When it comes to buying ready-made electronics modules for your projects, the SparkFun, Seeed Studio, Adafruit, and ITead Studio web sites can help. All have a wide range of modules, and much enjoyment can be had simply from browsing their online catalogs.

Nearly all the components used in this book have part codes for one or more of the suppliers I just mentioned. The only exceptions are for a few unusual modules that are better to buy from eBay.

There is also no end to the electronic components available on online auction sites, many coming direct from countries in the far east and often at extremely low prices. This is frequently the place to go for unusual components and things like laser modules and high-power LEDs that can be expensive in regular component suppliers. They are also very good for buying components in bulk. Sometimes these components are not grade A, however, so read the descriptions carefully and don't be disappointed if some of the items in the batch are dead-on-arrival.

Where to Buy Things to Hack

The first thing to consider, now that you are into hacking electronics, is an effect that your household and friends will have on you. You will become the recipient of dead electronics. But keep an eye open in your new role as refuse collector. Sometimes these "dead" items may actually be candidates for straightforward resurrection.

Another major source of useful bits is the dollar/pound/euro (delete as appropriate) store. Find the aisle with the electronic stuff: flashlights, fans, solar toys, illuminated cooling laptop bases, and so on. It's amazing what can be bought for a single unit of currency. Often you will find motors and arrays of LEDs for a lower price than you would the raw components from a conventional supplier.

Supermarkets are another source of cheap electronics that can be hacked. Good examples of useful gadgets are cheap powered computer speakers, mice, power supplies, radio receivers, LED flashlights, and computer keyboards.

A Basic Toolkit

Don't think you are going to get through this chapter without doing some soldering. Given this, you will need some basic tools. These do not have to be expensive. In fact, when you are starting out on something new, it's a good idea to learn to use things that are inexpensive, so it doesn't matter if you spoil them. After all, you wouldn't learn the violin on a Stradivarius. Plus, what will you have to look forward to if you buy all your high-end tools now!

Many starter toolkits are available. For our purposes, you will need a basic soldering iron, solder, a soldering iron stand, some pliers, snips, and a screwdriver or two. SparkFun sells just such a kit (SKU TOL-09465), so buy that one or look for something similar.

FIGURE 1-1 A digital multimeter

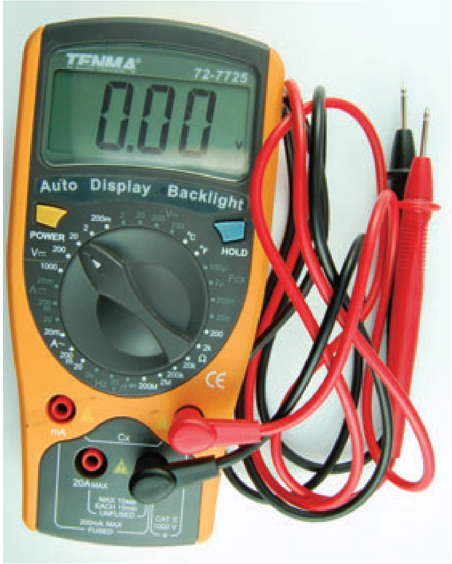

You will also need a multimeter (Figure 1-1). I would suggest a low-cost digital multimeter (don't even think of going above USD 20). Even if you end up buying a better one, you will still end up using the other one since it's often useful to measure more than one thing at a time. The key things you need are DC Volts, DC current, resistance, and a continuity test. Everything else is fluff that you will only need once in a blue moon. Again, look for something similar to this model from SparkFun (SKU TOL-09141) or the slightly higher specification meter shown in Figure 1-1.

Solderless breadboards (Figure 1-2) are very useful for quickly trying out designs before you commit them to solder. You poke the leads of components into the sockets, and metal clips behind the holes connect all the holes on a row together. They are not expensive (see T5 in the Appendix).

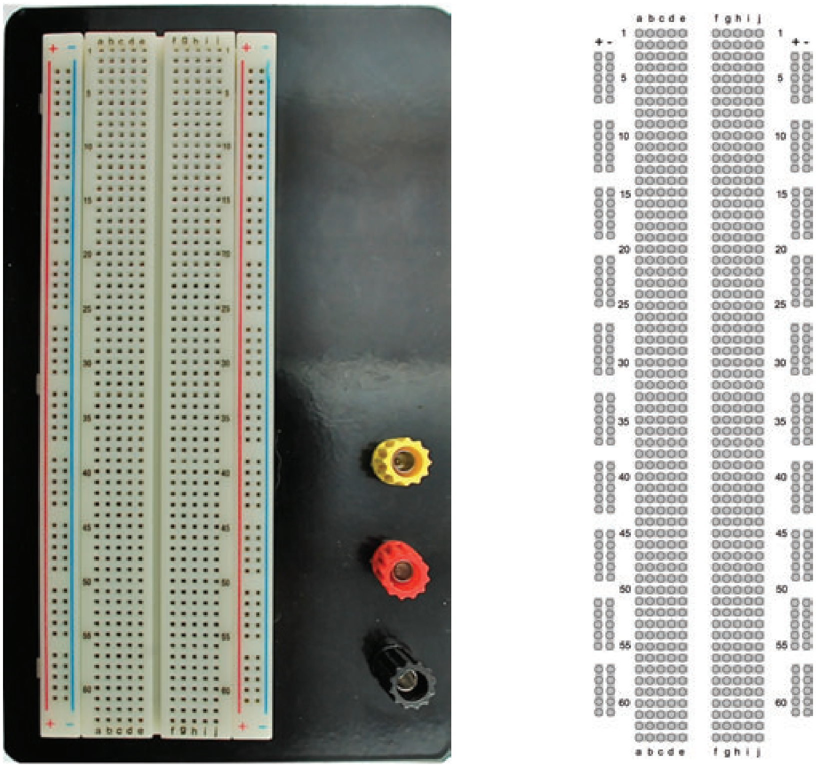

(a)

(b)

Figure 1-2 Solderless breadboard

You will also need some solid core wire in different colors (T6) to make bridging connections on the breadboard. Another good idea is to buy special-purpose jumper wires with little plugs on the end—although these are useful, they are by no means essential.

Breadboard comes in all shapes and sizes, but a big one is probably most useful. Where I use solderless breadboard in the book, I use the one specified in T5 in the Appendix. This has 63 rows by 2 columns with two supply strips down each side (Figure 1-2a). It is also mounted on an aluminum base with rubber feet to stop it moving about on the table. This is a very common size of breadboard and most suppliers will have something similar.

Figure 1-2b shows how the conductive strips are arranged underneath the plastic top surface of the board. All the holes that share a common gray area beneath are connected together in rows of five connectors. The long strips down each side are used for the power supply to the components. One positive and one negative. They are color-coded red and green.

How to Strip a Wire

Let's start with some basic techniques you need to know when hacking electronics. Perhaps the most basic of these is stripping wire.

You Will Need

Quantity	Item	Appendix Code
	Wire to be stripped	T9 or scrap
1	Pliers	T1
1	Snips	T1

Whenever you hack electronics, there is likely to be some wire involved, so you need to know how to use it. Figure 1-3 shows a selection of commonly used types of wire, set beside a matchstick to give them perspective.

On the left, next to the matchstick, are three lengths of solid-core wire, sometimes called hookup wire. This is mostly used with solderless breadboard, because being made of a single core of wire inside plastic insulation, it will eventually break if it is bent. Being made of a single strand of wire does mean it is much easier to push into sockets when prototyping since it doesn't bunch up like multi-core wire.

When using it with breadboard, you can either buy already-stripped lengths of wire in various colors as a kit (see Appendix, T6) or reels of wire that you can cut to the lengths you want yourself (see Appendix, T7, T8, T9). It is useful to have at least

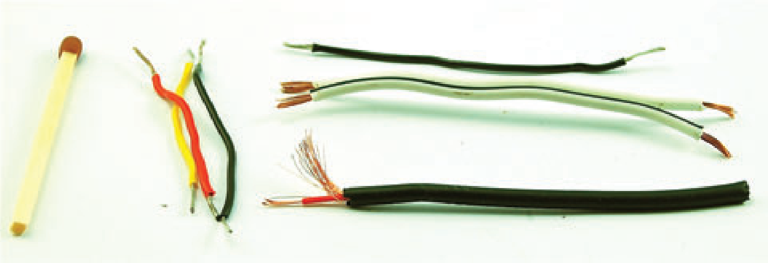

FIGURE 1-3 Common types of wire

three colors: red, yellow, and black are a good choice. It makes it easier to see how a project is connected up if you use red for the positive power supply, black for negative, and yellow for any other wires needed.

The top right of Figure 1-3 shows a length of multi-core wire, as well as some twin-strand multi-core wire. Multi-core wire is used when connecting up modules of a project. For instance, the wires to a loudspeaker from an amplifier module might use some twin, multi-core wire. It's useful to have some of this wire around. It is easily reclaimed from broken electronic devices, and relatively cheap to buy new (see Appendix, T10 and T11).

The wire at the bottom right of Figure 1-3 is screened wire. This is the type of wire you find in audio and headphone leads. It has an inner core of multi-core insulated wire surrounded by a screened wire on the outside. This type of wire is used where you don't want electrical noise from the environment such as mains hum (60 Hz electrical noise from 110V equipment) to influence the signal running through the central wire. The outer wire screens the inner wire from any stray signals and noise. There are variations of this where there is more than one core surrounded by the screening—for example, in a stereo audio lead.

Insulated wire is of no use to us unless we have a way of taking some of the insulation off it at the end, as this is where we will connect it to something. This is called "stripping" the wire. You can buy special-purpose wire strippers for this, which you can adjust to the diameter of the wire you want to strip. This implies that you know the width of the wire, however. If you are using some wire that you scavenged from a dead electronic appliance, you won't know the width. Having said that, with a bit of practice you will find you can strip wire just as well using a pair of pliers and some wire snips.

Both of these are essential tools for the electronics hacker. Neither tool needs to be expensive. In fact, snips tend to get notches in them that make them annoying to use, so a cheap pair (I usually pay about USD 2) that can be replaced regularly is a good idea.

Figures 1-4a and 1-4b show how to strip a wire with pliers and snips. The pliers are used to hold things still with a firm grip, while the snips do the actual stripping.

Grip the wire in the pliers, about an inch away from the end (Figure 1-4a). Use the snips to grip the insulation where you want to take it off. Sometimes it helps to just nip the insulation all the way around before gripping it tightly with the snips, and then pull the insulation off (Figure 1-4b).

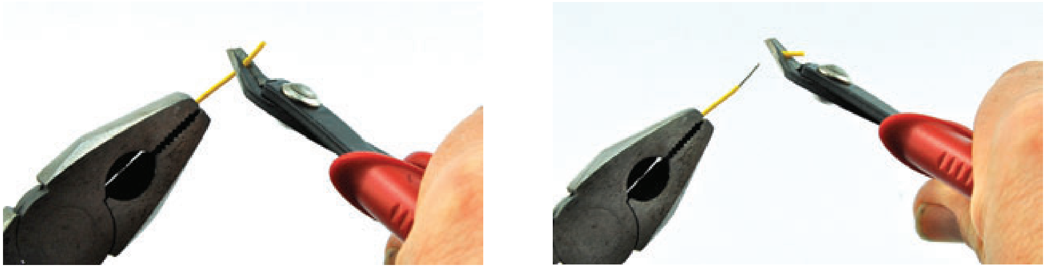

(a) (b)

FIGURE 1-4 Stripping wire

For longer lengths of wire, you can just wrap the wire around your finger a few times instead of using pliers.

This takes a bit of practice. Sometimes you will have the snips grip it too tightly and accidentally cut the wire all the way through, while other times you won't grip it hard enough with the snips and the insulation will stay in place or stretch. Before attempting anything important, practice with an old length of wire.

How to Join Wires Together by Twisting

It is possible to join wires without soldering. Soldering is more permanent, but sometimes this technique is good enough.

One of the simplest ways of joining wires is to simply twist the bare ends together. This works much better for multi-core wire than the single-core variety, but if done properly with the single-core, it will still make a reliable connection.

You Will Need

To try out joining two wires by twisting (there is slightly more to it than you might expect), you will need the following.

Quantity	Item	Appendix Code
2	Wires to be joined	T10
1	Roll of PVC insulating tape	T3

If you need to strip the wires first to get at the copper, refer back to the section "How to Strip a Wire."

Figures 1-5a thru 1-5d show the sequence of events in joining two wires by twisting them.

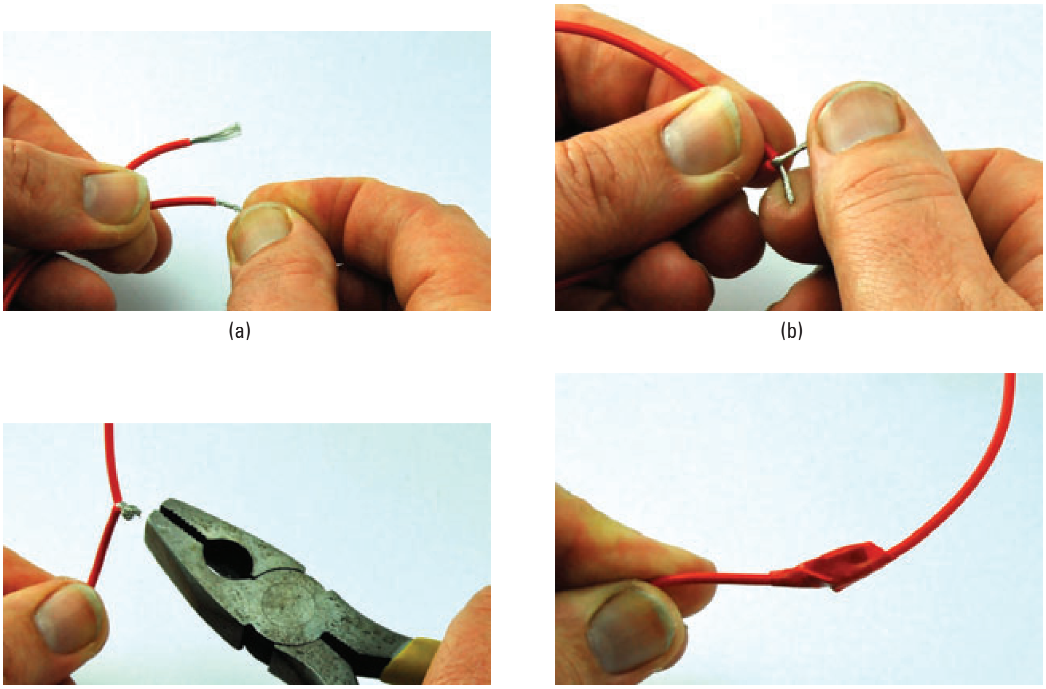

(a) (b)

(c) (d)

Figure 1-5 Joining wires by twisting

First, twist the strands of each wire up clockwise (Figure 1-5a). This just tidies up any straggling strands of the multi-core wire. Then, twist together the two pre-twisted wires (Figure 1-5b) so they are both twisting around each other. Try to avoid the situation where one of the wires twists around the second, while the second remains straight. If it does this, it is very easy for the first wire to just slip off the second. Next, twist the joined wires up into a neat little knot (Figure 1-5c). Note that a pair of pliers may be easier to use when making the knot, especially if the wire is on the thick side. Lastly, cover the joint with four or five turns of PVC insulating tape (Figure 1-5d).

How to Join Wires by Soldering

Soldering is the main skill necessary for hacking electronics.

Safety

I don't want to put you off, but … be aware that soldering involves melting metal at very high temperatures. Not only that, but melting metal that's coupled with noxious fumes. It is a law of nature that anyone who has a motorbike eventually falls off it, and anyone who solders will burn their fingers. So be careful and follow these safety tips:

- Always put the iron back in its stand when you are not actually soldering something. If you leave it resting on the bench, sooner or later it will roll off. Or you could catch the wires with your elbow and if it falls to the floor, your natural reflex will be to try and catch it—and chances are you will catch the hot end. If you try and juggle it in one hand, while looking for something or arranging some components ready to solder, sooner or later you will either solder your fingers or burn something precious.
- Wear safety glasses. Blobs of molten solder will sometimes flick up, especially when soldering a wire or component that is under tension. You do not want a blob of molten solder in your eye. If you are long-sighted, magnifying goggles may not look cool, but they will serve the dual purpose of protecting your eyes and letting you see properly.
- If you do burn yourself, run cold water over the burned skin for at least a minute. If the burn is bad, seek medical attention.
- Solder in a ventilated room, and ideally set up a little fan to draw the fumes away from you and the soldering iron. Preferably have it blowing out of a window. A fun little project to practice your wire joining skills on is making a fan using an old computer (see the section "How to Hack a Computer Fan to Keep Soldering Fumes Away").

You Will Need

To practice joining some wires with solder, you will need the following items.

Quantity	Item	Appendix Code
2	Wires to be joined	T10
1	Roll of PVC insulating tape	T3
1	Soldering kit	T1
1	Magic hands (optional)	T4
1	Coffee mug (essential)	

Magic hands are a great help during soldering because they solve the problem that, when soldering, you really need three hands: one to hold the iron, one to hold the solder, and one to hold the thing or things you are trying to solder. You generally use the magic hands to hold the thing or things you are trying to solder. Magic hands are comprised of a small weighted bracket with crocodile clips that can be used to hold things in place and off the work surface.

An alternative that works well for wires is to bend them a little so that the end you are soldering will stick up from the workbench. It usually helps to place something heavy like a coffee mug on the wire to keep it from moving.

Soldering

Before we get onto the business of joining these two wires, let's have a look at soldering. If you haven't soldered before, Figures 1-6a thru 1-6c show you how it's done.

1. Make sure your soldering iron has fully heated up.
2. Clean the tip by wiping it on the damp (not sopping wet) sponge on the soldering iron stand.

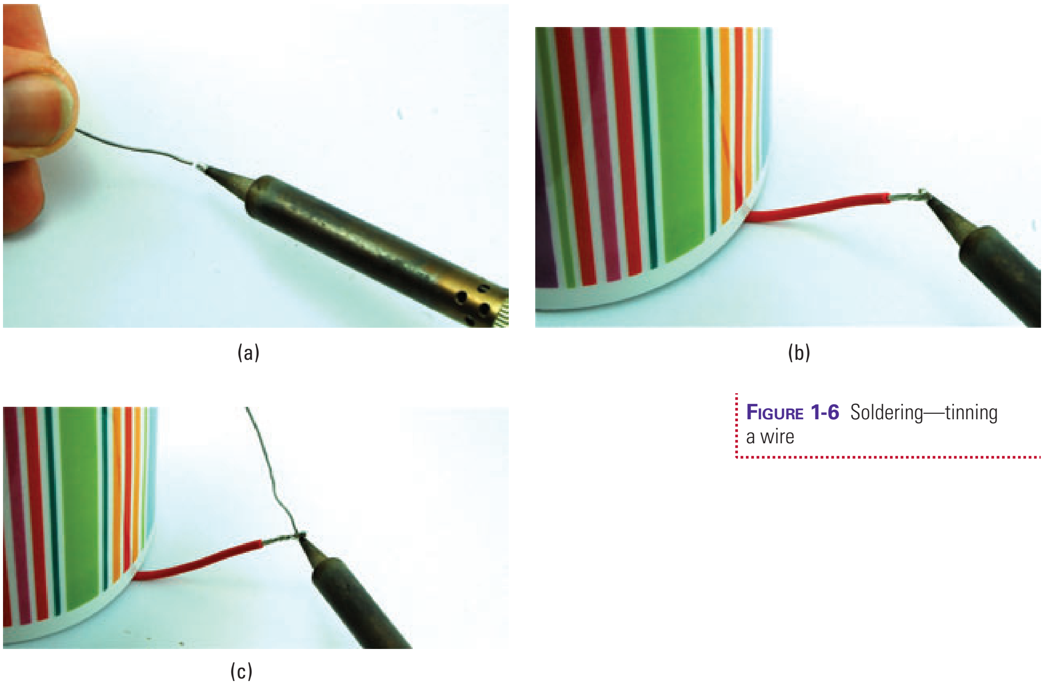
(a)

(b)

(c)

FIGURE 1-6 Soldering—tinning a wire

3. Touch a bit of solder onto the tip of the iron to "tin" it (see Figure 1-6a). After you have done this, the tip should be bright and shiny. If the solder doesn't melt, then your iron probably isn't hot enough yet. If the solder forms into a ball and doesn't coat the tip of the iron, the tip of it may be dirty, so wipe it on the sponge and try again.
4. Hold the soldering iron to the wire and leave it there for a second or two (Figure 1-6b).
5. Touch the solder to the wire near the soldering iron. It should flow into the wire (Figure 1-6c).

Soldering is something of an art. Some people are naturally very neat at soldering. So do not worry if your results are a bit blobby at first. You will get better. The main thing to remember is that you heat up the item you want to solder and only apply the solder when that thing is hot enough for the solder to melt onto it. If you are struggling, it sometimes helps to apply the solder to the spot where the soldering iron meets the thing being soldered.

The following section offers a bit more soldering practice for you—in this case, by soldering wires together.

Joining Wires

To join two wires with solder, you can use the same approach described in the section "How to Join Wires Together by Twisting" and then flow solder into the little knot. An alternative way—that makes for a less lumpy joined wire—is illustrated in Figures 1-7a thru 1-7d.

1. The first step is to twist each end. If it is multi-core wire (a), tin it with solder as shown in Figure 1-7a.
2. Hold the wires side by side and heat them with the iron (see Figure 1-7b). Note the chopstick technique of holding both the second wire and the solder in one hand.
3. Introduce the solder to the wires so they join together into one wire and look something like that shown Figure 1-7c.
4. Wrap the joint in three or four turns of insulating tape—half an inch is probably enough (see Figure 1-7d).

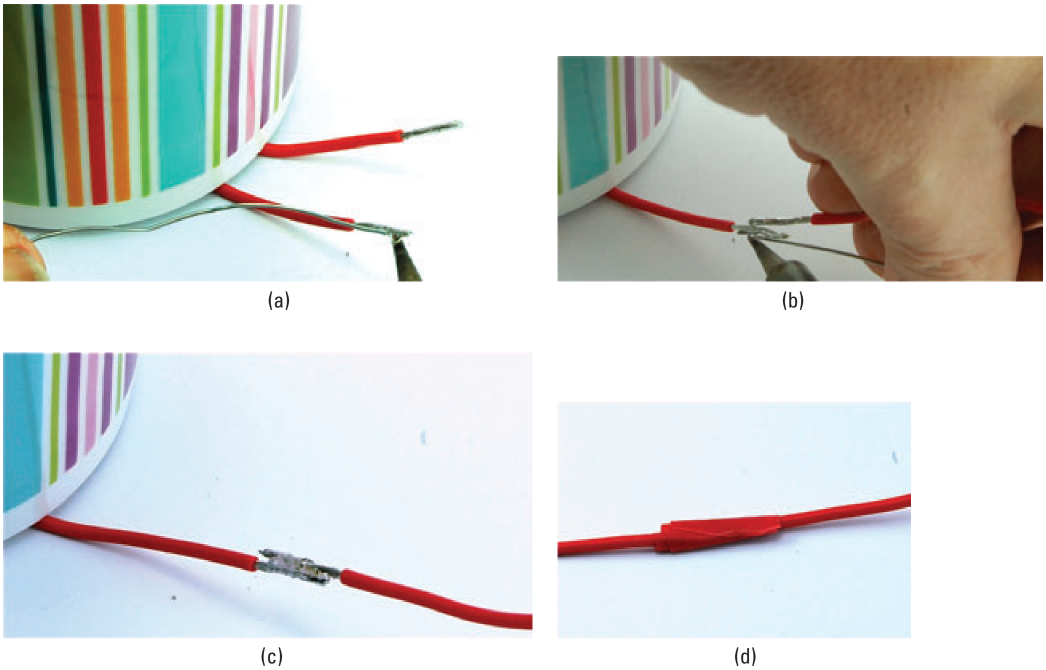

(a) (b) (c) (d)

FIGURE 1-7 Joining wires by soldering

How to Test a Connection

For the joints that we have made in the section "How to Join Wires by Soldering," it is fairly obvious that they are connected. However, especially with solid-core wire, it is not uncommon for the wire core to break somewhere under the insulation. If you own an electric guitar, you will probably be familiar with the problem of a broken guitar lead.

You Will Need

Quantity	Item	Appendix Code
1	Multimeter	T2
1	Connections to be tested	

Nearly all multimeters have a "Continuity" mode. When set in this useful mode, the multimeter will beep when the leads are connected to each other.

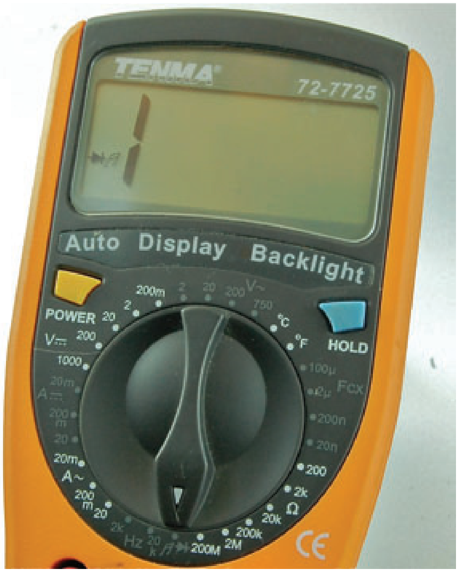

Figure 1-8 A multimeter in Continuity mode

Set your multimeter to "Continuity mode," and then try touching the leads together. Now take a length of wire and try touching the multimeter leads to each end of the wire (Figure 1-8). The buzzer should sound if the wire is okay.

You can use this technique on circuit boards. If you have an old bit of circuit board from something, try testing between the soldered connections on the same track (Figure 1-9).

If there is no connection where you would expect there to be a connection, then there may be a "dry joint," where the solder hasn't flowed properly or there is a crack in the track on the circuit board (this sometimes happens if the board gets flexed).

A dry joint is easily fixed by just applying a bit of solder and making sure it flows properly. Cracks on a circuit board can be fixed by scraping away some of the protective lacquer over the track and then soldering up the split in the track.

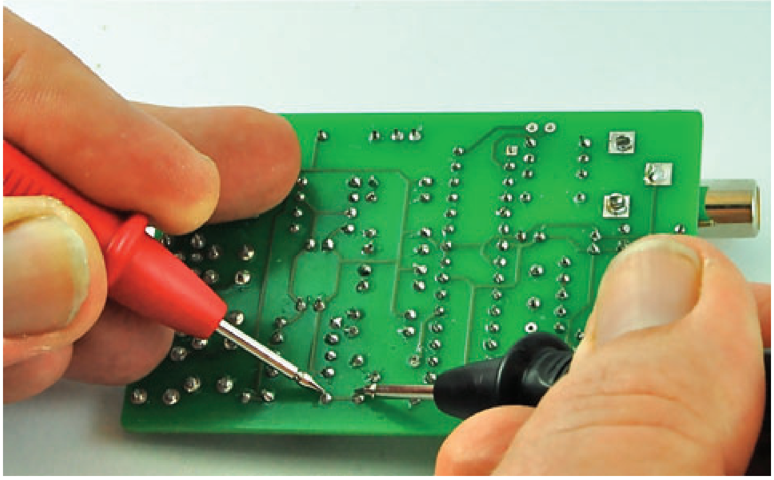

Figure 1-9 Testing a circuit board

How to Hack a Computer Fan to Keep Soldering Fumes Away

Solder fumes are unpleasant and bad for you. If you can sit by an open window while you solder, then great. If not, then this is a good little construction project to enhance your electronics hacking skills (Figure 1-10).

Okay, so it's not going to win any awards for style, but attached to my work light (which is always close to whatever I am soldering), the fumes will at least be directed away from my face.

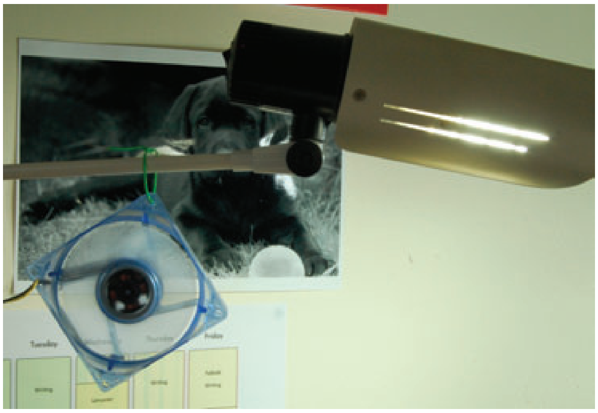

FIGURE 1-10 A homemade fume extractor

You Will Need

Quantity	Item	Appendix Code
1	Soldering equipment	T1
1	An old computer fan (two-lead)	
1	12V power supply	M1
1	SPST switch	K1

Construction

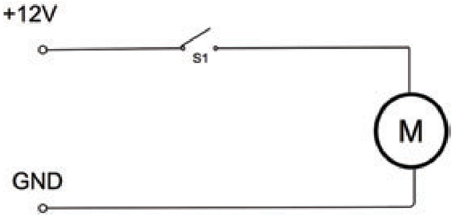

FIGURE 1-11 The schematic diagram for the fume extractor

Figure 1-11 shows the schematic diagram for this mini-project.

Newcomers to electronics often view schematic diagrams like this with suspicion, thinking it better just to show the components as they actually are, with wires where wires need to be—just like in Figure 1-12. It is worth learning how to read a schematic diagram. It really isn't that hard and in the long term it will pay dividends. Not least because of the vast number of useful circuit diagrams published on the Internet. It's a bit like being able to read music. You can get so far playing by ear, but there are more options if you can read and write musical notation.

So, let's examine our schematic diagram. Over on the left we have two labels that say "+12V" and "GND." The first is the

FIGURE 1-12 The wiring diagram for the fume extractor

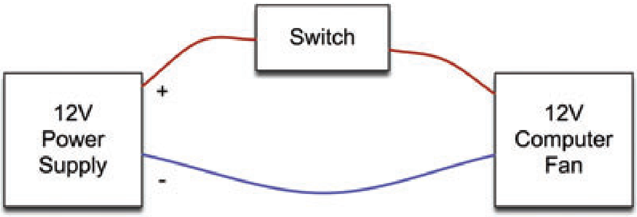

12V positive supply from the 12V power supply. GND actually refers to the negative connection of the power supply. GND is short for "ground" and just means zero volts. Voltage is relative, so the 12V connection of the power supply is 12V above the other connection (the GND connection). We will learn more about voltage in the next chapter.

Moving toward the right, we have a switch. This is labeled "S1," and if we had more than one switch in a schematic, they would be labeled "S2," "S3," and so on. The symbol for a switch shows how it operates. When the switch is turned to the on position, its two connections are connected together, and when it is in the off position, they aren't. It's as simple as that.

The switch is just controlling the supply of electricity to the motor of the fan (M) as if it were a faucet.

Step 1. Strip the Power Supply Leads

We have a power supply and we are going to cut the plug off the end of it and strip the wires (see the section "How to Strip a Wire"). Before you cut off the plug, make sure the power supply is NOT plugged in. Otherwise, if you snip both wires at the same time, the cutters will probably short the two connections together, which may damage the power supply.

Step 2. Identify the Power Supply Lead Polarity

Having cut the wires, we need to know which one is the positive one. To do this, let's use a multimeter. Set the multimeter to its 20V DC range. Your multimeter will probably have two voltage ranges, one for AC and one for DC. You need to use the DC range. This is often marked by a solid line above a dotted line. The AC range will either be marked as AC or have a picture of a little sine wave next to it. If you select AC instead of DC, it will not damage the meter, but you will not get a meaningful reading. (See Chapter 11 if you need more information on multimeters.)

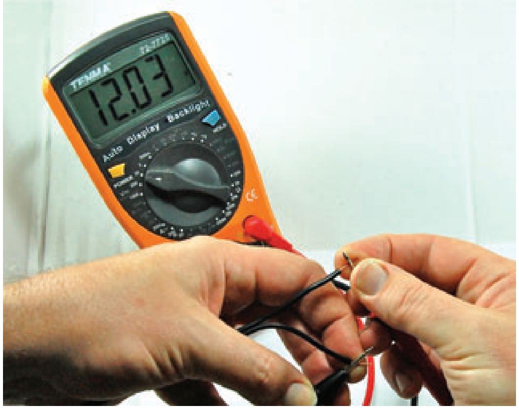

FIGURE 1-13 Using a multimeter to find the power supply polarity

First making sure that the stripped leads from the power supply are not touching, plug the power supply in and turn it on.

Now touch the two test leads from the multimeter to the leads from the power supply (Figure 1-13). If the number on the multimeter is not negative, then the red test lead of the multimeter is connected to the positive lead. Mark the lead in some way (I tied a knot in it). If the multimeter shows a negative voltage, then the leads are swapped over, so tie a knot in the power supply lead connected to the black test lead of the multimeter—in this case, this is the positive lead from the power supply.

Step 3. Connect the Negative Leads Together

Unplug your power supply. You should never solder anything that is powered up.

Cut any plug off the end of your computer fan and strip the two wires. Mine had one black (negative) and one yellow (positive) lead. Three lead fans are more complex and should be avoided. If you get the leads the wrong way around, no harm will befall you. The fan will just rotate in the opposite direction.

We are now going to join the negative lead of the fan to the negative lead (no knot) of the power supply (Figure 1-14).

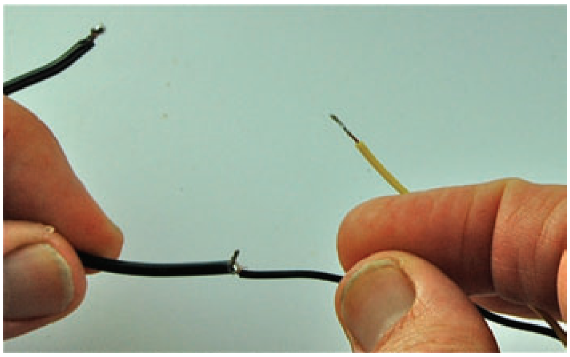

FIGURE 1-14 Connecting the negative leads together

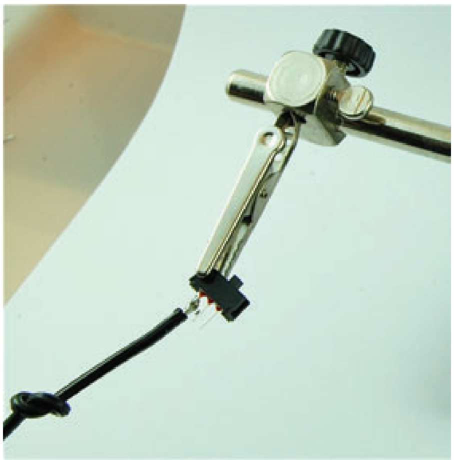

FIGURE 1-15 Connecting the positive lead to the switch

Step 4. Connect the Positive Lead to the Switch

Solder the positive lead from the power supply to one of the outer connections on the switch (it doesn't matter which). (See Figure 1-15.) It will help to tin the switch connection with a little solder before you start.

Finally, connect the remaining lead from the fan to the center connection of the switch (see Figure 1-16).

Step 5. Try It Out

Wrap the bare connections with insulating tape, plug it in, turn it on, and presto! When you flick the switch, the fan should come on.

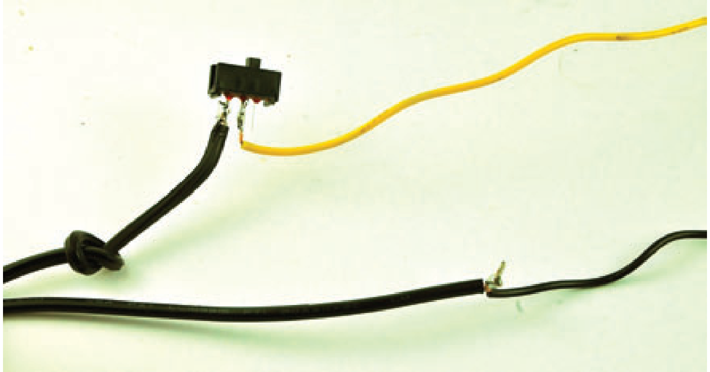

FIGURE 1-16 Connecting the fan to the switch

Summary

Now that we have the basics and are confident about a bit of soldering and dealing with wires and switches, we can now move on to Chapter 2. There, we will start looking at a few electronic components, as well as some of the basic ideas you will need to understand to successfully hack electronics.

2

Theory and Practice

There are a few fundamentals that will help us get the most out of our electronics. I have no intention of overloading you with theory, so you may find you come back to this chapter as and when you need to. But before we start on any theory, let's look at getting together some of the components we will use.

How to Assemble a Starter Kit of Components

In Chapter 1, we assembled a few tools and did some soldering. The only thing we made used a scavenged computer fan, an off-the-shelf power supply, and a switch.

Certain components you will find that you use over and over again. To get yourself a basic stock of components, I recommend you buy a starter kit. SparkFun sells such a kit (see the Appendix, K1), but it does not contain any resistors, so you will need to buy a resistor set, too (K2). Once you have these, you will have a useful collection of components that should cover 80 percent of what you need.

Other suppliers sell starter kits, and although none of them will contain everything you need for this book, most will give you a very good starting point.

You Will Need

The SparkFun Starter Kit contains the following items, and the items used directly in this book are marked with a *, so if buying an alternative kit, look for one that has the majority of these components. Also see the Appendix for a list of other components used in the book.

Quantity	Item	Quantity	Item
10	0.1uF capacitor *	3	20-pin male header *
5	100uF capacitor *	3	Mini power switch *
5	10uF capacitor *	2	Push buttons *
5	1uF capacitor	1	10k trimpot *
5	10nF capacitor	2	LM358 OpAmp
5	1nF capacitor	2	3.3V regulator
5	100pF capacitor	2	5V regulator *
5	10pF capacitor	1	555 timer *
5	1N4148 diode	1	Green LED *
5	1N4001 diode *	1	Yellow LED *
5	2N3906 PNP transistor	1	Red LED *
5	2N3904 NPN transistor *	1	7-segment red LED
3	20-pin female header	1	Mini photocell *

The separate SparkFun resistor kit (K2) contains resistors of the following values:

0Ω, 1.5Ω, 4.7Ω, 10Ω, 47Ω
110Ω, 220Ω, 330Ω, 470Ω, 680Ω
1kΩ, 2.2kΩ, 3.3kΩ, 4.7kΩ, 10kΩ
22kΩ, 47kΩ, 100kΩ, 330kΩ, 1MΩ

How to Identify Electronic Components

So, what have we just bought here? Let's go through the components in the SparkFun starter kits and explain what they do, starting with the resistors.

Resistors

Figure 2-1 shows an assortment of resistors. Resistors come in different sizes to be able to cope with different amounts of power. High-power resistors are physically big to cope with the heat they produce. Since "parts getting hot" is generally a bad thing in electronics, we will mostly avoid that. Nearly all of the time we can use the 0.25-watt resistors as provided in the SparkFun kit, which are perfect for general use.

As well as having a maximum power rating, resistors also have a "resistance." As the word suggests, resistance is actually

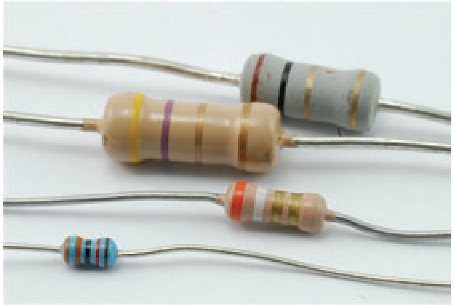

FIGURE 2-1 Assorted resistors

resistance to the flow of current. So a high-resistance resistor will not allow much current to flow, while a low-value resistor will allow lots of current to flow.

Resistors are the most commonly used component you can find. Since we will be using them a lot, we will go into greater detail on the subject in the section "What Are Current, Resistance, and Voltage?" later in this chapter.

Resistors have little stripes on them that tell you their value. You can learn to read the stripes (more in a moment on that) or you can avoid all of this by storing them in a bag or in the drawer of a component box with the value written on the box or bag. If in doubt, check the value with the resistance measurement feature of your multimeter.

However, an essential piece of geekiness is to know your resistor color-codes. Each color has a value, as shown next:

Color	Value
Black	0
Brown	1
Red	2
Orange	3
Yellow	4
Green	5
Blue	6
Violet	7
Gray	8
White	9
Gold	1/10
Silver	1/100

Gold and silver, as well as representing the fractions 1/10 and 1/100, are also used to indicate how accurate the resistor is. So gold is ±5% and silver is ±10%.

There will generally be three of these bands grouped together at one end of the resistor. This is followed by a gap, and then a single band at the other end of the resistor. The single band indicates the accuracy of the resistor value. Since none of the projects in this book require very accurate resistors, there is no need to select your resistors on the basis of accuracy.

Figure 2-2 shows the arrangement of the colored bands. The resistor value uses just the three bands. The first band is the first digit, the second the second digit, and the third "multiplier" band is how many zeros to put after the first two digits.

So, a 270Ω (*ohm*) resistor will have first digit 2 (red), second digit 7 (violet), and a multiplier of 1 (brown). Similarly, a 10KΩ resistor will have bands of brown, black, and orange (1, 0, 000).

In addition to fixed resistors, there are also variable resistors (a.k.a., potentiometers or pots). This comes in handy with volume controls, where turning a knob changes the resistance and alters the level of sound.

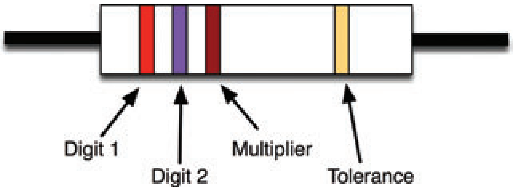

FIGURE 2-2 Resistor stripes

Capacitors

When hacking electronics, you will occasionally need to use a capacitor. Luckily, you do not need to know much about what they do. They are often used to head-off problems like the instability of a circuit or unwanted noise. Their use is often given a name like "decoupling capacitor" or "smoothing capacitor." There are simple rules you can follow about where you need a capacitor. These will be highlighted as we encounter them in later sections.

For the curious, capacitors store charge, a bit like a battery, but not much charge, and they can store the charge and release it very quickly.

Figure 2-3 shows a selection of capacitors.

If you look closely at the second capacitor from the left, you will see the number 103. This is actually the value of the capacitor in picofarads. The unit of capacitance is farad, but a 1F capacitor would be considered a huge capacitor, storing a great deal of charge. So, while such beasts do exist, everyday capacitors are either measured in nanofarads (nF = 1/1,000,000,000F) or microfarads (μF = 1/1,000,000F). You will also find capacitors in the picofarad range (pF = 1/1,000,000,000,000F).

Returning to 103. … Rather like resistors, this means 10 and then 3 zeros, in units of pF. So in this case that's 10,000pF or 10nF.

Larger capacitors, like those on the right of Figure 2-3, are called electrolytic capacitors. They are usually in the μF

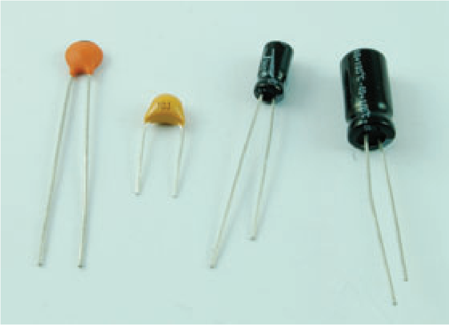

FIGURE 2-3 Assorted capacitors

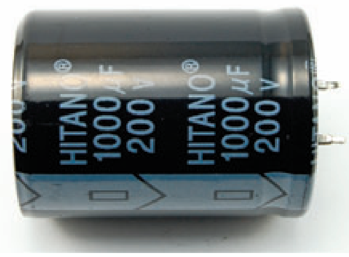

Figure 2-4 An electrolytic capacitor

range and have their value written on their side. They also have a + and a – side, and unlike most other capacitors must be connected the right way around.

Figure 2-4 shows a large electrolytic, with value (1000μF) and its negative lead clearly indicated at the bottom of the figure. If the capacitor has one lead longer than the other, the longer one will normally be the positive lead.

The capacitor in Figure 2-4 also has a voltage written on it (200V). This is the capacitor's maximum voltage. So if you put more than 200V across its leads, it will fail. Big electrolytic capacitors like this have a reputation for failing spectacularly and may burst, spewing forth goo.

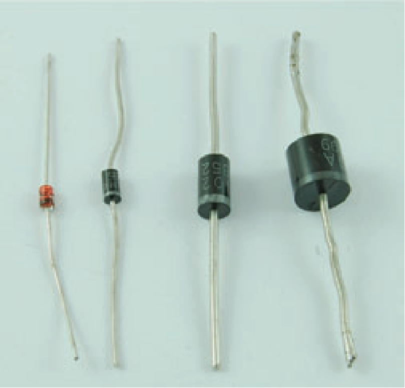

Figure 2-5 A selection of diodes

Diodes

You will occasionally need to use diodes. They are kind of a one-way valve, only allowing current to flow in one direction. They are therefore often used to protect sensitive components from accidental reverse voltage that could damage them.

Diodes (Figure 2-5) have a stripe at one end. That end is called the cathode, while the other end is called the anode. We will hear more about diodes later.

As with resistors, the bigger the diode physically, the more power it can cope with before it gets too hot and expires. Ninety percent of the time, you will just be using one of the two diodes on the left-hand side of the figure.

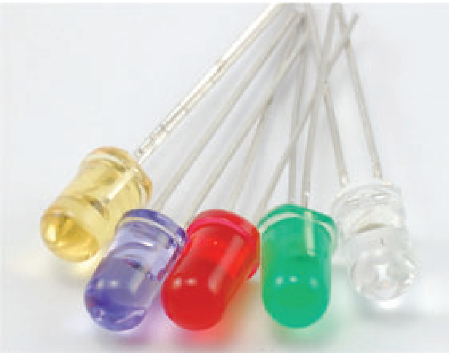

Figure 2-6 Assorted LEDs

LEDs

LEDs light up, and generally look pretty. Figure 2-6 shows a selection of LEDs.

LEDs are a little sensitive, so you should not connect them directly to a battery. Instead you have to use a resistor to reduce the current flowing into the LED. If you do not do this, the LED will probably die almost instantly.

Later on, we will see how to select the right resistor for the job.

Just like regular diodes, LEDs have a positive and a negative lead (anode and cathode). The anode is the longer of the two leads. There is also usually a flat side to the LED case on the cathode side.

As well as single LEDs, you also get LEDs in more complicated arrangements within a single package. Figure 2-7 shows some interesting-looking LEDs.

From left to right, these LEDs are an ultraviolet LED, an LED with both red and green LEDs in the same package, a high-power RGB (red, green, blue) LED that can be controlled to produce any color of light, a seven-segment LED display, and an LED bar graph display.

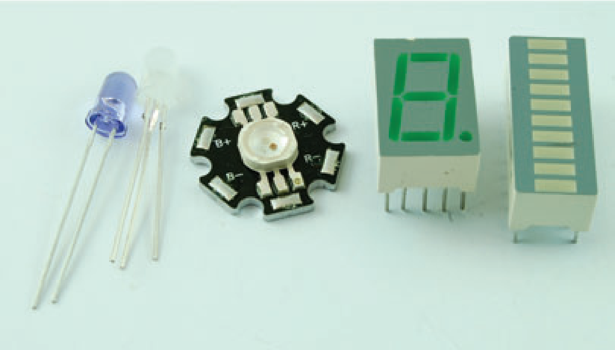

FIGURE 2-7 More LEDs

This is just a small selection of LED types. There are many others to choose from. In later sections, we will explore some of these more exotic LEDs.

Transistors

While transistors can be used in audio amplifiers and in many circumstances, for the casual electronics hacker, the transistor can be thought of as a switch. But rather than a switch controlled by a lever, it is a switch that switches a big current, yet is controlled by a small current.

Generally speaking, the physical size of the transistor (Figure 2-8) determines how big the current that it switches can be before it starts producing smoke.

Of the transistors in Figure 2-8, the right-hand two are quite specialized and employed for high power use.

Generally, the rule for a component is that if it's ugly and has three legs, it's probably some kind of transistor.

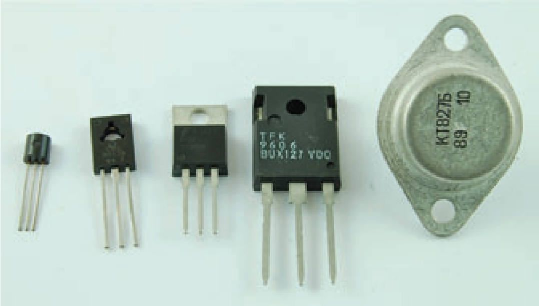

FIGURE 2-8 Transistors

Integrated Circuits

An integrated circuit (IC), or just "chip," is a load of transistors and other components printed onto silicon. The purpose of the IC varies wildly. It can be a microcontroller (mini-computer), or an entire audio amplifier, or a computer memory, or any one of thousands of other possibilities.

ICs make life easy, because as they say, often "there's a chip for that." Indeed, if there is something you want to make, there may well be a chip for it already, and if there isn't, then there will probably be a general-purpose chip that takes you halfway there.

ICs look like bugs (Figure 2-9).

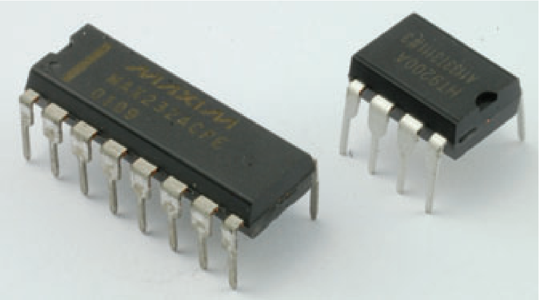

Figure 2-9 Integrated circuits

Other Stuff

There are so many other components out there, some of which are very familiar, such as batteries and switches. Others are less familiar and include potentiometers (variable resistors found in volume controls), phototransistors, rotary encoders, light dependent resistors, and so on. We will explore these as they arise later in the book.

Surface Mount Components

Let's touch a little on the subject of surface mount devices (SMDs). These components are just resistors, transistors, capacitors, ICs, and so on, but in tiny packages designed to be soldered onto the top surface of circuit boards by machines.

Figure 2-10 shows a selection of SMDs.

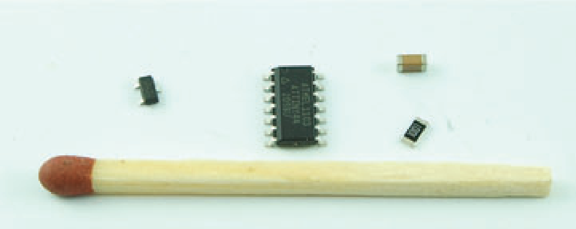

Figure 2-10 Surface mount components

The matchstick shows you just how small these devices are. It is perfectly possible to do surface mount soldering by hand, but you need a steady hand and a high-quality soldering iron. Not to mention a lot of patience. You are also likely to need a means of making circuit boards, as they are not easy to use with breadboard or other prototyping tools.

In this book, we mostly look at using the conventional "through-hole" components rather than SMDs. However, as your experience grows and you feel you might like working with SMDs, do not be afraid to try.

What Are Current, Resistance, and Voltage?

Voltage, current, and resistance are three properties that are fundamental to almost everything you will do in electronics. They are intimately related, and if you can master the relationship between them, you will be a wise hacker indeed.

Please take the time to read and understand this little bit of theory. Once you understand it, many other things should automatically fall into place.

Current

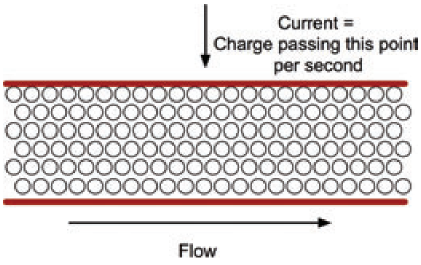

Figure 2-11 Current

The problem with electrons is that you cannot see them, so you just have to imagine how they do things. I like to think of electrons as little balls flowing through pipes. Any physicists reading this will probably be clutching their heads or hurling this book to the floor in disgust now. But it works for me.

Each electron has a charge and it's always the same—lots of electrons, lots of charge, few electrons, and a little bit of charge.

Current, rather like the current in a river, is measured by counting how much charge passes you per second (Figure 2-11).

Resistance

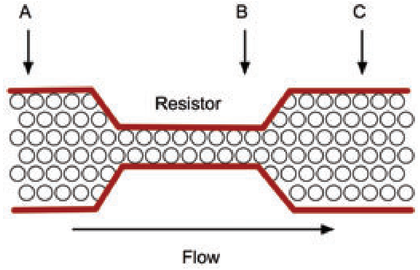

Figure 2-12 A resistor

A resistor's job is to provide resistance to the flow of current. So, if we keep thinking about our river, it is like a constriction in a river (Figure 2-12).

The resistor has reduced the amount of charge that can pass by a point. And it doesn't matter which point you measure at (A, B, or C) because, if you look upstream of the resistor, the charge is hanging around waiting to move through the resistor. Therefore, less is moving past A per second. In the resistor (B), it's restricted.

The "speed" analogy does not really hold true for electrons, but one important point is that the current will be the same wherever you measure it.

Imagine what happens when a resistor stops too much current from flowing through an LED.

Voltage

Voltage is the final part of the equation (that we will come to in a minute). If we persist with the water-in-a-river analogy, then voltage is like the height that the river drops over a given distance (Figure 2-13).

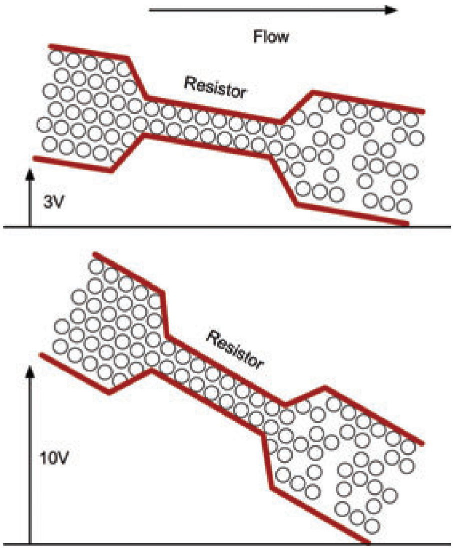

FIGURE 2-13 Voltage

As everyone knows, a river that loses height quickly flows fast and furious, whereas a relatively gently sloped river will have a correspondingly gentle current.

This analogy helps with the concept of voltage being relative. That is, it does not matter if the river is falling from 10,000 ft to 5,000 ft or from 5,000 ft to 0 ft. The drop is the same and so will be the rate of flow.

Ohm's Law

Before we get into the math of this, let's think for a moment about current, voltage, and resistance and how they relate to each other.

Try this little quiz. Think in terms of the river if you find it helps.

1. If the voltage increases, will the current (a) increase or (b) decrease?
2. If the resistance increases, will the current (a) increase or (b) decrease?

Did you get the answers (a) and (b) correct, respectively?

If you write this down as an equation, it is called Ohm's law and can be written as:

$$I = V / R$$

I for current (I guess "C" was already taken), V for Volts, and R for resistance.

So, the current flowing through a resistor, or any wire connecting to it, will be the voltage across the resistor divided by the resistance of the resistor.

The units of resistance are in Ω (the abbreviation for ohms), while units of current are in A (short for amps, which is short for amperes) and in voltage V (the easy one).

So, if we have a voltage of 10V across a resistor of 100Ω the current flowing will be:

$$10V / 100\Omega = 0.1A$$

For convenience, we often use mA (1/1000 of an amp). So 0.1A is also 100mA.

That's enough about Ohm's law for now, we will meet it again later. It is the single most useful thing you can know about electronics. In the next section, we will look at the only other truly essential math you will need—power.

What Is Power?

Power is all about energy and time. So, in a way, it's a bit like current. But, instead of being the amount of charge passing a point, it is the amount of energy transformed into heat per second when a current passes through something that resists the flow (like a resistor). Forget the river, it doesn't really help much here.

Restricting the flow of a current generates heat, and the amount of heat generated can be calculated as the voltage across a resistor times the current flowing through it. The units of power are the watt (W). You would write this in math as:

$$P = I \times V$$

So, in our earlier example, we had 10V across a 100Ω resistor, so the current through the resistor was 100mA and will generate 0.1A × 10V, or 1 W of power. Given that the resistors that we have from the SparkFun kit are 250 mW (0.25 W). Our resistor will get hot and may eventually break.

If you don't know the current, but you do know the resistance, another useful formula for calculating the power is:

$$P = V^2 / R$$

Or, power is voltage squared (times itself) divided by the resistance. So, for the example earlier:

$$P = 10 \times 10 / 100 = 1\ W$$

That's reassuringly the same answer as we got before.

Most components have a maximum power rating like this, so when selecting a resistor, transistor, diode, and so on, it is worth doing a quick check and multiplying the voltage across the component by the current that you expect to flow through it. Then, choose a component with a maximum power rating comfortably greater than the expected power.

Power is the best measure of how much electricity is being used. It is the electrical energy being used per second, and unlike current it can be compared for devices operating from

Device	Power
Battery-powered FM radio (volume down)	20 mW
Battery-powered FM radio (volume up)	500 mW
Arduino Uno microcontroller board (9V supply)	200 mW
Home WiFi router	10 W
Compact fluorescent (low-power) light bulb	15 W
Filament light bulb	60 W
LCD TV 40-inch	200 W
Electric kettle	3000 W (3 kW)

TABLE 2-1 Power Usage

both 110 volt outlets and low voltage. It is good to have a basic understanding of just how much—or how little—electricity devices use. Table 2-1 shows some devices you might find around the home and lists how much power they use.

So, now you know why you don't get battery-powered kettles!

How to Read a Schematic Diagram

Hacking electronics often involves trawling the Internet, looking for people who have made something like the thing you want to make or adapt. You will often find schematic diagrams that tell you how to make and do things. So you need to be able to understand these schematics in order to turn them into real electronics.

These may at first sight seem a little baffling, but schematics obey a few simple rules and tend to use the same patterns over and over again. So there is a lot less to learn than you might think.

Ponder Figure 2-14 while we consider some of these rules—or more accurately conventions—because sometimes they are broken.

Figure 2-14 goes a long way to explaining why we sometimes talk of electronic circuits. It's kind of a loop. The current is flowing out of the battery, through the switch (when it's closed), through the resistor and LED (D1), and then back to the battery. The lines on the schematic can be thought of as perfect wires without any resistance.

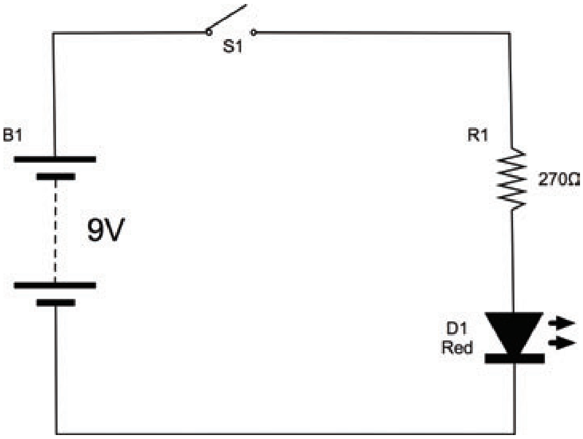

FIGURE 2-14 A simple schematic

The First Rule of Schematics: Positive Voltages Are Uppermost

A convention that most people follow when drawing a schematic is to put the higher voltages near the top, so on the left-hand side of the diagram, we have a 9V battery. The bottom of the battery is at 0V or GND (Ground), while the top of the battery will by 9V higher than that.

Notice that we draw the resistor R1 above the LED (D1). This way, we can think of some of the voltage as being lost across the resistor, before the remainder is lost through the diode and flows back to the negative connection of the battery.

Second Rule of Schematics: Things Happen Left to Right

Western civilization invented electronics and writes from left to right. You read from left to right and, culturally, more things happen from left to right. Electronics is no different, so it is common to start with the source of the electricity—the battery or power supply on the left—and then work our way from left to right across the diagram.

So, next we have our switch, which controls the flow of the electricity, and then the resistor and LED.

Names and Values

It is normal to give every component in a schematic a name. So, in this case the battery pack is called B1, the switch S1, the resistor R1, and the LED D1. This means that when you go from a schematic to a breadboard layout and eventually a circuit board, you can see which components on the schematic correspond to which components on the breadboard or circuit board.

It is also normal to specify the value of each of the components where appropriate. So, for example, the resistors' value of 270Ω is marked on the diagram. The rest of the components don't need much else said about them.

Component Symbols

Table 2-2 lists the most common circuit symbols you will encounter. This is nothing like a complete list, but we will discuss other symbols later in the book.

There are two main styles of circuit symbol: American and European. Fortunately, they are similar enough to avoid difficulties in recognizing them.

In this book, we will use the U.S. circuit symbols.

Symbol (U.S.)	Symbol (European)	Photo	Component	Use
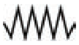	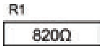		Resistor	Resisting
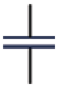	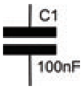		Capacitor	Temporary charge storage
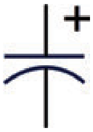	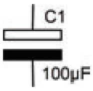		Capacitor (polarized)	
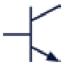	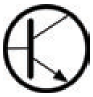		Transistor (bipolar NPN)	Using a small current to control a larger current
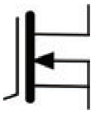	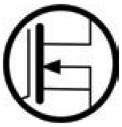		Transistor (MOSFET N-channel)	Using a very small current to control a larger current
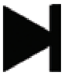	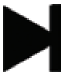		Diode	Prevents current from flowing in the wrong direction
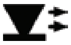	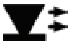		LED	Indication and illumination
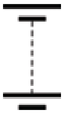	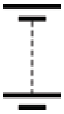		Battery	Power supply
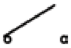	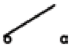		Switch	Turning things on and off; control

TABLE 2-2 Common Schematic Symbols

Summary

In the next chapter, we get a much more practical look at some basic hacks and hone our electronic construction skills. This includes using prototyping boards and taking our soldering beyond simply connecting wires to other wires.

We will also learn how to use solderless breadboard so we can build electronics quickly and get underway.

UNO
ARDUINO
DIGITAL (PWM~)

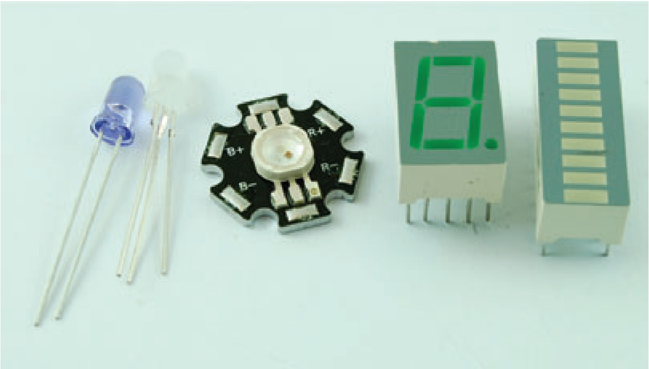
B+
R+
B-
R-

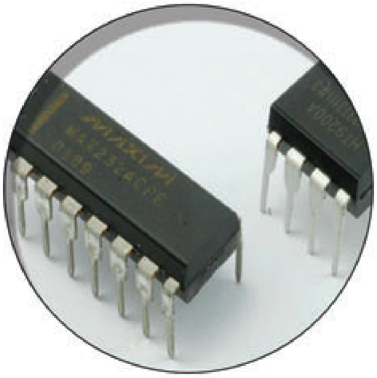
MAXIM
MAX232CPE

3

Basic Hacks

This chapter contains a set of fairly basic “hacking” how-tos. These build and use various electronic construction techniques. So this is probably a good chapter to at least skim through so that when you attempt more advanced how-tos, you can refer back to it if needed.

How to Make a Resistor Get Hot

Sometimes things will get hot when you are hacking electronics. It’s always better when this is expected rather than when it’s a surprise, so it’s worth doing a little experimenting in this area.

You Will Need

Quantity	Item	Appendix Code
1	100Ω 0.25-watt resistor	K2
1	4 × AA battery holder	H1
1	4 × AA batteries (the rechargeable type is a good idea)	

Figure 3-1 shows the schematic diagram.

The Experiment

All we will do is connect the 100Ω resistor across the battery terminals and see how hot it gets.

Caution Be careful when doing this because the resistor’s temperature will rise to about 50°C/122°F. The resistor’s leads, however, will not get very hot.

We are using a battery holder that takes four AA cells, each providing about 1.5V. They are each connected, one after the other, providing us with 6V total. Figure 3-2 shows how the batteries are actually connected within the battery box as a schematic diagram. In this kind of arrangement, the batteries are said to be in series.

Figure 3-3 shows the resistor heater in action.

Simply touch a finger to the resistor to confirm it's hot.

Is this bad/good? Will the resistor eventually break because it's warm? No, it won't. Resistors are designed to cope with a bit of heat. If we do the math, the power that the resistor is burning is the voltage squared divided by the resistance, which is:

(6 × 6) / 100 = 0.36W

If it is a 0.25-W resistor, then we are exceeding its maximum power. This would be a foolish thing to do if we were designing a product for mass production. However, that's *not* what we are doing, and the chances are the resistor would continue to work like that indefinitely.

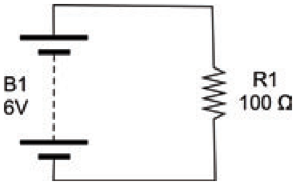

FIGURE 3-1 The schematic for heating a resistor

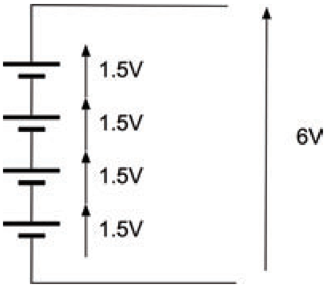

FIGURE 3-2 The schematic diagram for a battery holder

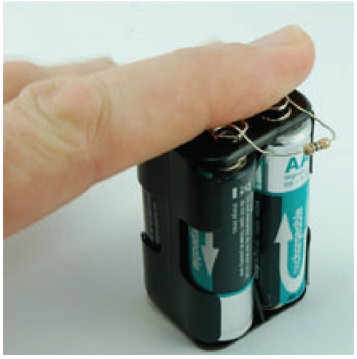

FIGURE 3-3 Making a resistor get hot

How to Use Resistors to Divide a Voltage

Sometimes voltages are too big. For example, in an FM radio, the signal going from the radio part to the audio amplifier part will be deliberately too large so it can be reduced using the volume knob.

Another example might be when you have a sensor that produces a voltage between 0 and 10V but you want to connect it to an Arduino microcontroller that expects it to be between 0 and 5V.

A very common technique in electronics is to use a pair of resistors (or a single variable resistor) as a "voltage divider."

You Will Need

Quantity	Item	Appendix Code
1	10kΩ trimpot (tiny variable resistor)	K1, R1
1	Solderless breadboard	T5
	Solid-core jumper wire	T6
1	4 × AA battery holder	H1
1	4 × AA batteries	
1	Battery clip	H2
1	Multimeter	T2

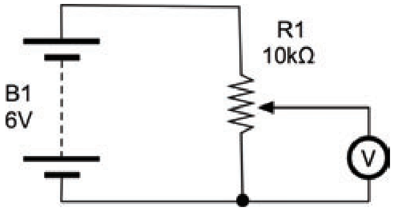

FIGURE 3-4 A voltage divider schematic diagram

Figure 3-4 shows the schematic diagram for our experiment. There are a couple of new schematic symbols here. The first is the variable resistor (or pot). This looks like a regular resistor symbol, but has a line with an arrow connecting to the resistor. This is the moving slider connection of the variable resistor.

The second new symbol is the circle with a "V" in it. This is a voltmeter, which in our case is the multimeter set to its DC voltage range.

The variable resistor we will use has three leads. One lead is fixed at each end of a conductive track, while a third connection to the central slider moves from one end of the track to the other. The overall resistance of the whole track is 10kΩ.

Our voltage in is going to be supplied from the battery pack and will be roughly 6V. We are going to use a multimeter to measure the output voltage and see how much it is being reduced by our voltage divider.

If you remember, the grey bars indicate where the connections underneath the holes are connected together. Take some time to follow the lines on the stripboard and reassure yourself that everything is connected in the same way as the schematic (Figure 3-4).

Plug the trimpot into the breadboard as shown, and then wire up the battery by carefully pushing the leads into the + and – power supply lines on the breadboard: red to +, black to –. If you struggle to get the multi-core wires of the battery clip into the holes, solder a bit of solid-core wire to the end of the leads.

Attach wires between the positive supply and the top connection of the trimpot, and the negative supply and the bottom connection of the trimpot. Finally, attach the multimeter. If your multimeter has alligator clips, use these in preference to the normal probes, clipping short jumper wires into the alligator clips and then pushing the other ends into the positions shown in Figure 3-5. When you have done all this, your breadboard should look something like Figures 3-6a and 3-6b.

Turn the trimpot to its fully clockwise position. The multimeter should read 0V (Figure 3-6a). Now turn it fully anti-clockwise and it should read something around 6V (Figure 3-6b)—in other words, the full battery voltage. Finally, turn it to roughly the middle position and you should see that the meter indicates about 3V (Figure 3-6c).

Think of the variable resistor as behaving a bit like two resistors, R1 and R2, as shown in Figure 3-7.

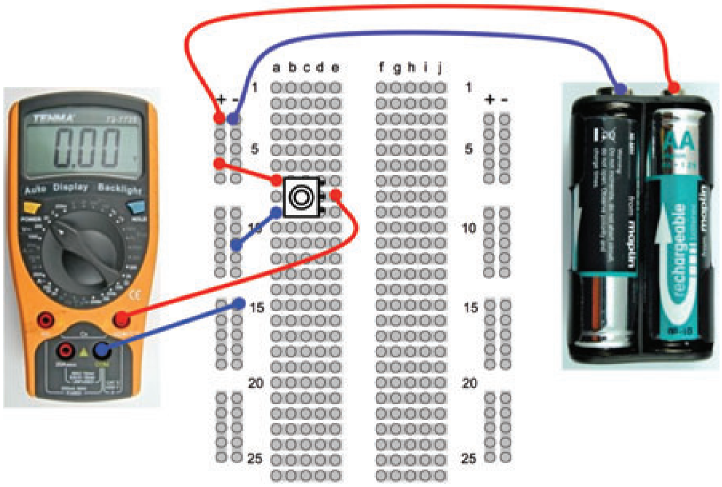

FIGURE 3-5 A voltage divider breadboard layout

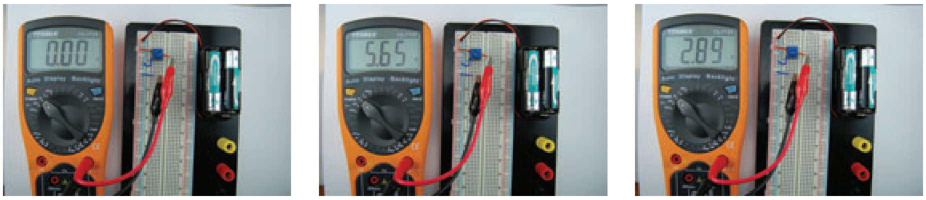

(a) (b) (c)

FIGURE 3-6 A voltage divider breadboard

The formula to calculate Vout if we know Vin, R1 and R2 is as follows:

$$Vout = Vin * R2 / (R1 + R2)$$

So, if R1 and R2 are both 5 kΩ and Vin is 6V, then:

$$Vout = 6V * 5k\Omega / (5k\Omega + 5k\Omega) = 30 / 10 = 3V$$

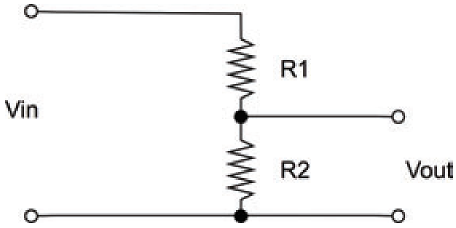

FIGURE 3-7 A voltage divider with fixed resistors

This ties in with what we found when we put the trimpot to its middle position. It is exactly the same as having two fixed resistors of 5 kΩ each.

As with many of the calculations you make in electronics, people have made handy calculating tools. If you type "voltage divider calculator" into a search engine, you will find them. One such example can be found here: www.electronics2000.co.uk/calc/potential-divider-calculator.php.

These calculators will also usually match to the nearest available fixed resistor value.

How to Convert a Resistance to a Voltage (and Make a Light Meter)

An LDR (light-dependent resistor; a.k.a., photoresistor) is a resistor whose resistance changes depending on the amount of light falling on its transparent window. We will use one of these devices to demonstrate the idea of converting a resistance to a voltage by using it as one-half of a potential divider.

You Will Need

Quantity	Item	Appendix Code
1	Light-dependent resistor	K1, R2
1	Solderless breadboard	T5
	Solid-core jumper wire	T6
1	4 × AA battery holder	H1
1	4 × AA batteries	
1	Battery clip	H2
1	Multimeter	T2

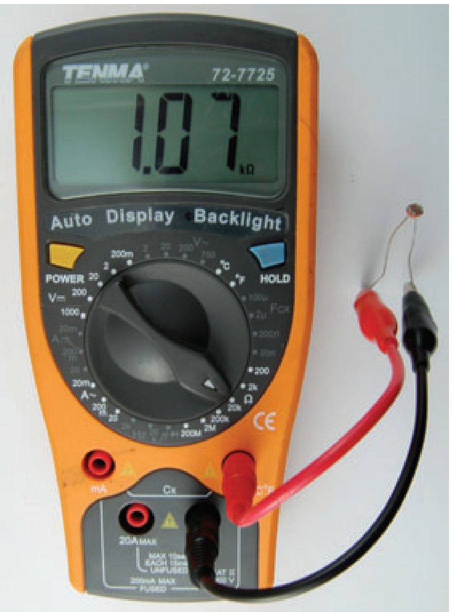

FIGURE 3-8 Measuring the LDR resistance

Before we get the breadboard out, let's just experiment directly with the LDR. Figure 3-8 shows the LDR connected directly to the multimeter on its 20kΩ resistance setting. As you can see, the resistance of my LDR was 1.07kΩ. Putting my hand over the LDR to screen out some of the light increased that resistance to a few tens of kΩ. So, the way the LDR works, the more light that reaches it, the lower the resistance.

Microcontrollers such as the Arduino can measure voltages and do things with them, but not directly measure resistance. So to convert our LDR's resistance into a more easily used voltage, we can put it in a voltage divider as one of the resistors (Figure 3-9).

Note that the symbol for the LDR is like a resistor but with little arrows pointing to it to indicate its sensitivity to light.

We can make up this schematic on our breadboard, this time setting our multimeter to the 20V DC range and watching how the voltage changes as we cover the LDR to reduce the light getting to it (Figures 3-10 and 3-11).

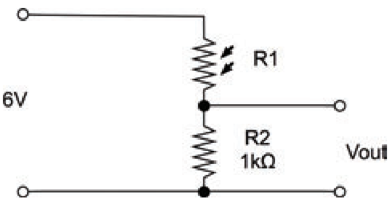
R1
6V
R2
1kΩ
Vout

FIGURE 3-9 Measuring light level with an LDR and voltage divider

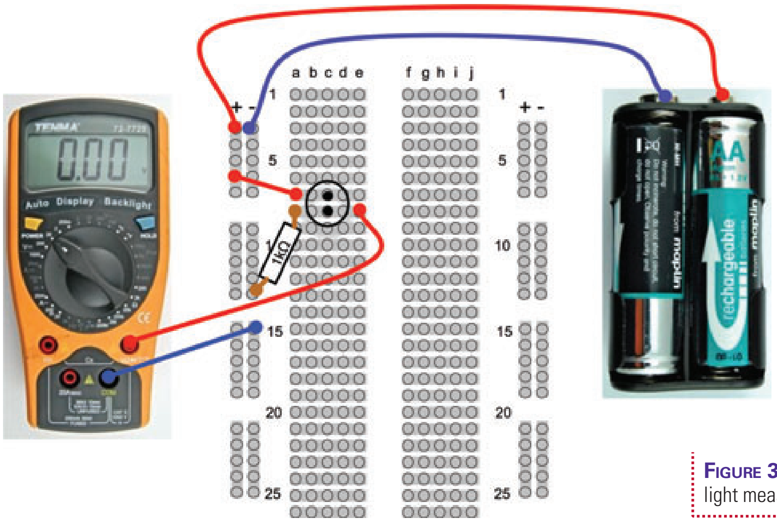
a b c d e
f g h i j
+ -
1
5
10
15
20
25
1kΩ
TENMA
0.00
Auto Display Backlight
AA
rechargeable
maplin

FIGURE 3-10 A breadboard layout for light measurement

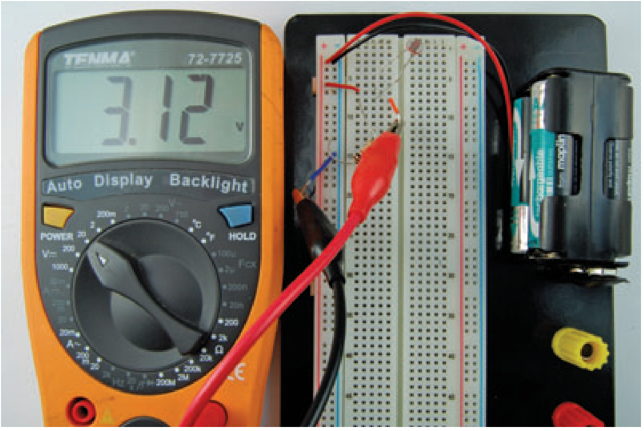
TENMA
72-7725
3.12
Auto Display Backlight
POWER
HOLD

FIGURE 3-11 Light measurement

Hack a Push Light to Make It Light Sensing

Battery-powered push lights are one of the many glorious bargains you are likely to find in a dollar/euro/pound store. These are intended for use in cupboards and other dark locations where a bit of extra light would be useful. Push them once and they light, push them again and they turn off.

It will not surprise you to hear that we are going to use our LDR to turn the light on and off. But we are also going to use a transistor.

Our approach will be to get it working on breadboard first and then solder up the design onto the push light. In fact, we will use a single LED in place of the push lamp until we know that it will work.

You Will Need

Quantity	Name	Item	Appendix Code
1	R1	Light-dependent resistor	K1, R2
1	T1	Transistor 2N3904	K1, S1
1	R2	Resistor 10kΩ	K2
1*	R3	Resistor 220Ω	K2
1*	D1	Red LED or high-brightness LED	K1 or S2
*		Solid-core jumper wire	T6
1		Push light	

* These components are only needed for the breadboard experiment.

We want the LDR to control an LED, so a first thought at a circuit might be as shown in Figure 3-12.

There are two fatal flaws in this design. First, as more light falls on the LDR its resistance decreases, allowing more current to flow so the LED will get brighter. This is the opposite of what we want. We want the LED to come on when it's dark.

We need to use a transistor.

The basic operation of a transistor is shown in Figure 3-13. There are many different types of transistors, and probably the most common (and the type we will use) is called an NPN bipolar transistor.

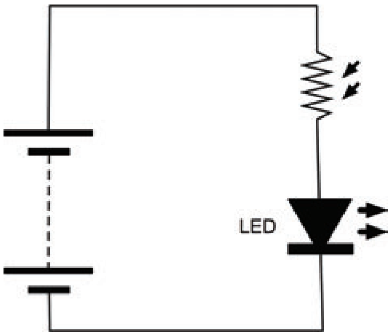

FIGURE 3-12 An LED and LDR

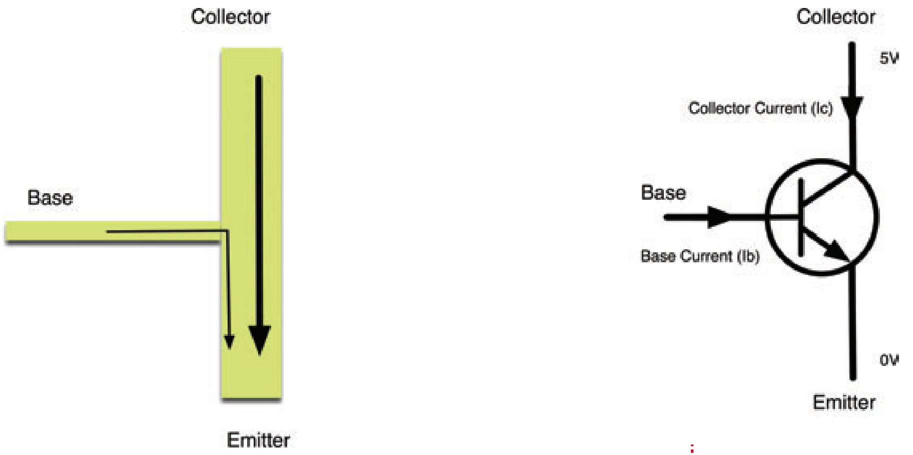

FIGURE 3-13 A bipolar transistor

This transistor has three leads: the emitter, the collector, and the base. The basic principal is that a small current flowing through the base will allow a much bigger current to flow between the collector and the emitter.

Just how much bigger the current is depends on the transistor, but it's typically a factor of 100.

Breadboard

Figure 3-14 shows the schematic diagram we will build on the breadboard. To understand this circuit, let's consider two cases.

Case 1: When It's Dark

In this case, the LDR R1 will have a very high resistance, so you could almost imagine that it isn't there at all. In that case, current will flow through R2, through the base and emitter of the transistor, allowing as much current as it needs to flow through R3, the LED, and T1 into its collector and out through the emitter. When enough current flows into the base of a transistor to allow current to flow from the collector to the emitter, this is called "turning on" the transistor.

We can calculate the base current using Ohm's law. In this situation, the base of the transistor will be at only about half a volt, so we can assume there is more or less the full 6V across the 10kΩ resistor R2. Since I = V / R, the current will be 6 / 10,000 A or 0.6mA.

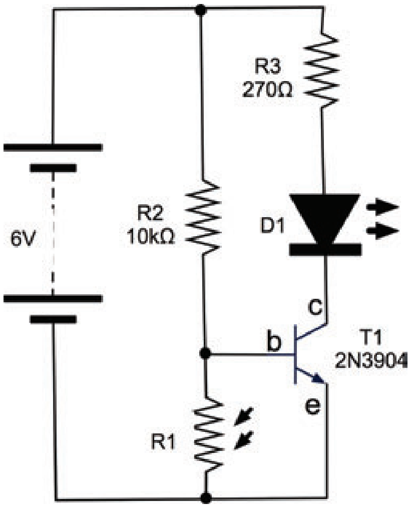

FIGURE 3-14 Using an LDR and transistor to switch an LED

Case 2: When It's Light

When it is light, we have to consider the resistance of the LDR R1. The lighter it is, the lower the resistance of R1 and the more of the current otherwise destined for the base of the transistor will be diverted through R1, preventing the transistor from turning on.

I think the time has come to build the project on breadboard. Figure 3-15 shows the breadboard layout, and Figures 3-16a and 3-16b the finished breadboard.

When placing the LED on the breadboard, make sure you get it the right way around. The longer lead is the positive lead, and it should be on row 10 connected to R3. (See Figure 3-16a.)

If everything is fine, you should find that when you cover the LDR, the LED should light (Figure 3-16b).

Construction

Now that we have proved our circuit works, we can get on with modding the push light. Figure 3-17 shows the push light the author used. Unless you are very lucky, yours is likely to be

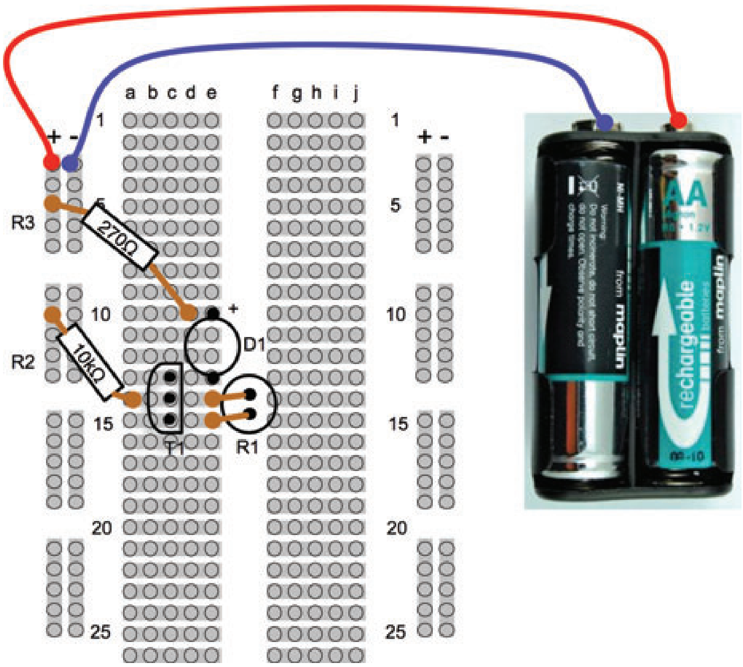

FIGURE 3-15 The light switch breadboard layout

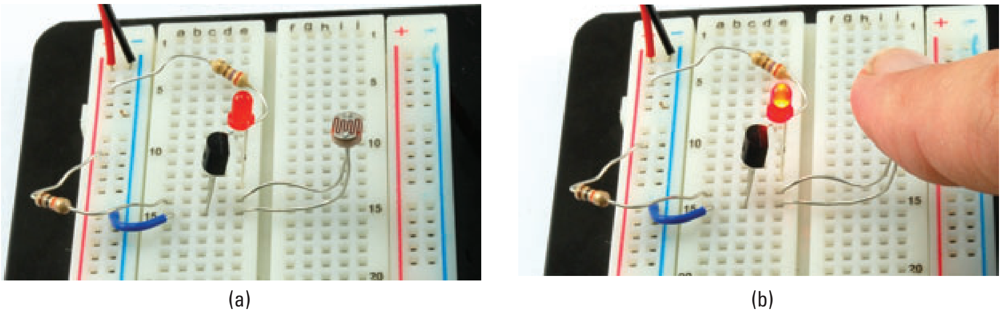

FIGURE 3-16 The light switch breadboard

different, so read through the following sections carefully and you should be able to work out how to change your light. To make life easy for yourself, try and find a push light that operates from 6V (4 AA or AAA cells).

You will probably find screws on the back of the push light. Remove these and put them somewhere safe. The inside of the push light is shown in Figure 3-18. The various connections on the light are marked. You can find the corresponding connections on your light using a multimeter.

Setting the multimeter to its 20V DC range will let you determine which battery lead is positive and which is negative. Looking at the wiring, we can draw a schematic diagram for the light as it stands, before we start altering it (Figure 3-19).

FIGURE 3-17 A push light

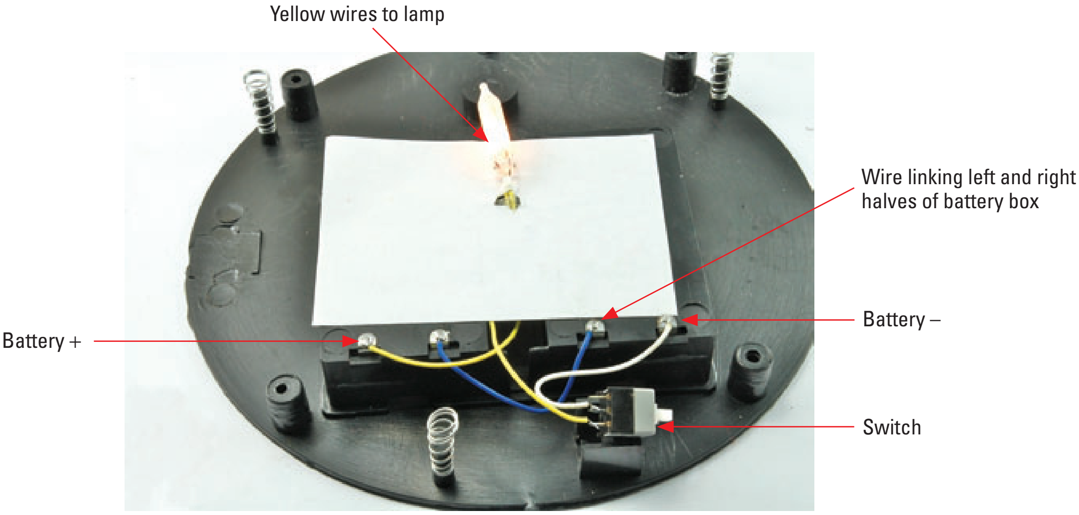

FIGURE 3-18 Inside the push light

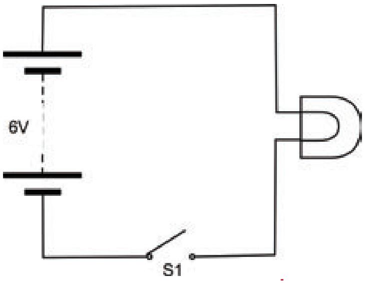

FIGURE 3-19 The schematic diagram for the original push light

This light uses an old-fashioned incandescent bulb. We will replace that with a high-brightness LED. If you don't have one of these, a regular LED of the color of your choice will work, but not be very bright.

Figure 3-20 shows how we replaced the bulb with the LED and the 220Ω resistor. Make sure the longer positive lead of the LED is connected to the resistor and the far side of that resistor is connected to the positive terminal of the battery.

Try pressing the switch to make sure the LED is working.

We can now draw a schematic that combines what we have in the existing lamp and our LDR circuit (Figure 3-21).

In fact, all this really amounts to is adding in the switch to the original LED schematic. We have already installed R3 and D1 when we replaced the bulb with an LED. The switch is already there, so all we need to add is the transistor, LDR, and R2. Figure 3-22 shows how we will rewire the push light.

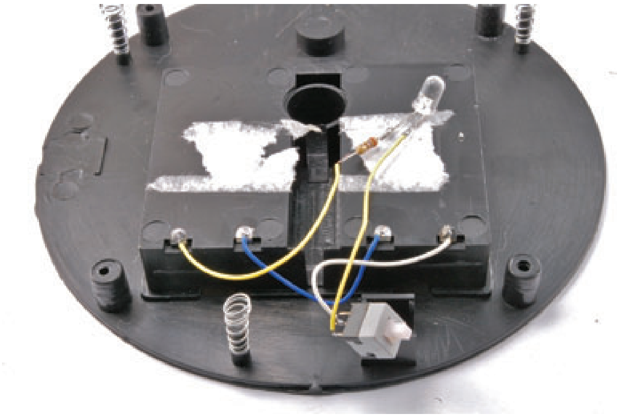

FIGURE 3-20 Replacing the bulb with an LED and a resistor

Figure 3-23 shows the sequence of steps in soldering the extra components onto the light.

1. Start by desoldering the lead from the switch that isn't connected to the negative battery terminal (Figure 3-23a).
2. Solder the 10kΩ resistor R2 between the middle lead of the transistor (the base) and the positive terminal on the battery box.
3. With the flat of the transistor facing upward, as shown in the diagram, connect the left-hand lead of the transistor to the wire you just disconnected from the switch (Figure 3-23b).
4. Solder the LDR between the left and middle pins of the transistor, and connect the combined left-hand transistor lead and LDR lead to the connection on the switch that the wire used to be attached to. (See Figure 3-23c.)
5. Tuck the components away neatly, bending the leads to make sure there is no way the bare leads can touch each other. (See Figure 3-23d.)

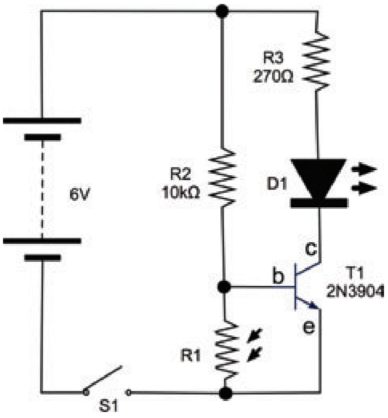

FIGURE 3-21 The final schematic

There you are! You have hacked some electronics.

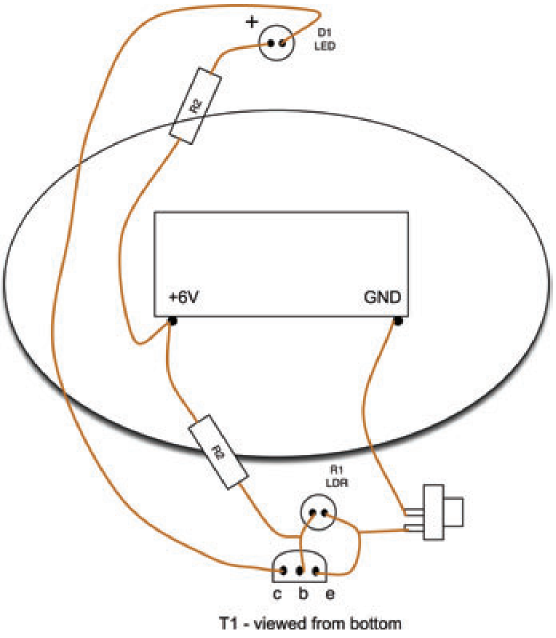
+
D1
LED
R2
+6V
GND
R2
R1
LDR
c b e
T1 - viewed from bottom

FIGURE 3-22 The push light wiring diagram

FIGURE 3-23 Soldering the project

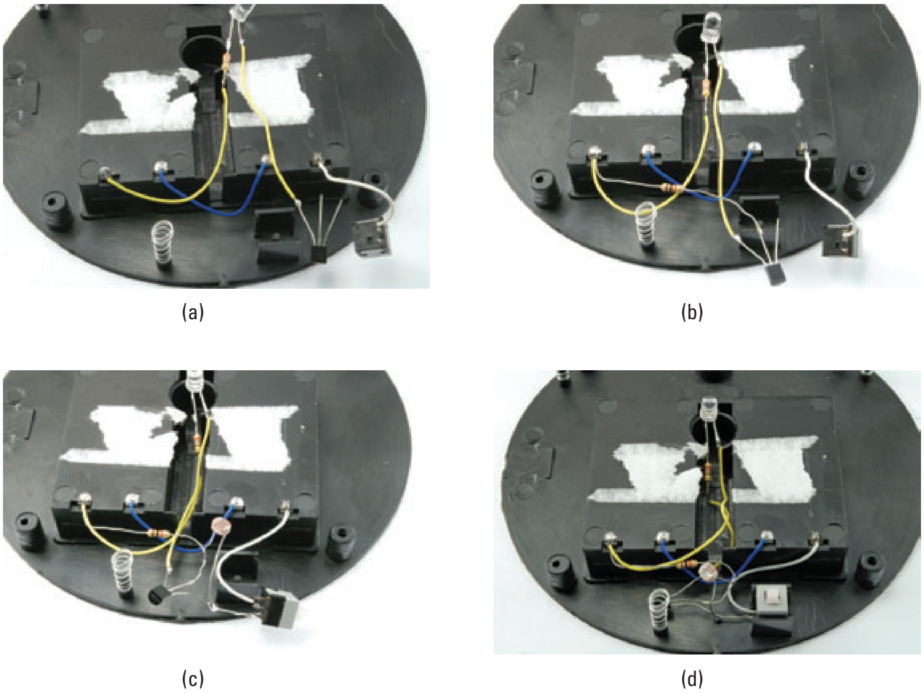
(a)
(b)
(c)
(d)

How to Choose a Bipolar Transistor

The transistor we used in the previous section, "Hack a Push Light to Make It Light Sensing," is a useful general-purpose transistor. But there are many other types of transistors that we can use for different purposes. This section is to help you find the right transistor and try to use it in such a way that it doesn't die in a puff of smoke.

Datasheets

Transistors have a number of parameters we need to know about. All transistors have an associated datasheet. This is produced by the manufacturer and specifies everything you could possibly want to know about the device, from the dimensions of its leads to its electrical characteristics.

Most of the time you will use one of three or four transistors and will not need to look into the exact details of how the transistor behaves, but if you do need to, it's there on the datasheet. So, you may just want to skip to the next subsection where we look at a few different types of transistors—just the useful ones, nothing exotic.

Table 3-1 shows some of the data you will find on the 2N3904's datasheet under maximum ratings.

Maximum Collector-Emitter and Collector-Base voltages of 40V and 60V mean we do not have to worry about exceeding them in battery-powered devices. We need to be careful that we do not exceed the emitter-base voltage though.

The maximum collector current of 200mA is quite healthy though. It means we could in theory control ten LEDs, all taking 20mA at the same time. If we do exceed this value, then the transistor will get hot and eventually fail.

The one electrical characteristic we are most interested in is DC current gain or h_{FE} as it will be called on the datasheet. This is listed in the electrical characteristics section of the datasheet.

You may remember that the DC gain is the multiplier that determines how much more current can flow in through the base than the collector. Looking at Table 3-2, this means that at a collector current of 10mA and a collector emitter voltage of 1.0V (it's

Absolute Maximum Ratings

Symbol	Parameter	Value	Units
V_{CEO}	Collector-Emitter Voltage	40	V
V_{CBO}	Collector-Base Voltage	60	V
V_{EBO}	Emitter-Base Voltage	6.0	V
I_C	Collector Current – Continuous	200	mA

TABLE 3-1 2N3904 Maximum Ratings

Symbol	Parameter	Test Condition	Min	Max	Units
ON CHARACTERISTICS					
h_{FE}	DC Current Gain	I_C = 0.1 mA, V_{CE} = 1.0V	40		
		I_C = 1.0 mA, V_{CE} = 1.0V	70		
		I_C = 10 mA, V_{CE} = 1.0V	100	300	
		I_C = 50 mA, V_{CE} = 1.0V	60		
		I_C = 100 mA, V_{CE} = 1.0V	30		

TABLE 3-2 2N3904 Electrical Characteristics

nearly always about that), the typical gain will be 100, meaning that only 10mA / 100 = 100nA needs to be flowing into the base for this amount of current to flow through the collector.

MOSFET Transistors

The 2N3904 is what is called a bipolar transistor. It's basically a device that amplifies current. A small current into the base controls a much bigger current flowing through the collector. Sometimes, the current gain of just 100 or so is not nearly enough.

There is another type of transistor that does not suffer from this limitation called the MOSFET (Metal Oxide Semiconductor Field Effect Transistor). You can see why it gets shortened to MOSFET. These transistors are controlled by voltage rather than current and make very good switches.

MOSFETs do not have emitters, bases, and collectors, they have "sources," "gates," and "drains." They turn on when the gate voltage passes a threshold, usually about 2V. Once on, quite large currents can flow through the "drain" to the "source" rather like a bipolar transistor. But since the gate is isolated from the rest of the transistor by a layer of insulating glass, hardly any current flows into the gate. It is the voltage at the gate that determines what current will flow.

We will meet MOSFETs again later in the section "How to Use a Power MOSFET to Control a Motor," and in Chapter 7 in the section "How to Control Motor Speed with a Power MOSFET."

PNP and N-Channel Transistors

The automated light switch of the previous section switched on the "negative side of the load." That is, if you go back to Figure 3-21, you can see that the resistor and LED that make up the light are not connected to GND except through the

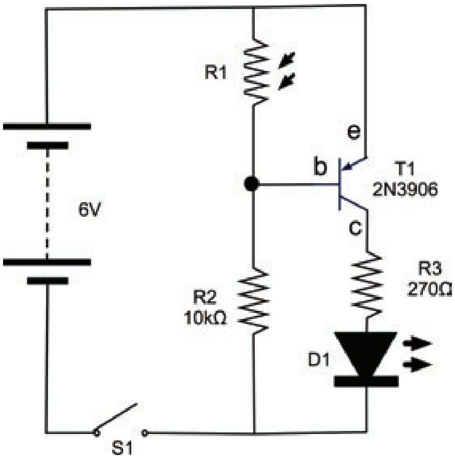

Figure 3-24 Using a PNP bipolar transistor

transistor. If for some reason (and this does happen) we wanted to switch the positive side, then we would need to use a PNP equivalent of the NPN 2N3904, such as the 2N3906. NPN stands for Negative-Positive-Negative, and yes, you can guess what PNP stands for. That is because transistors are kind of semiconductor sandwiches, with material of either N or P type as the bread. If the bread is N type (the most common), then the base voltage needs to be higher than the emitter voltage (by about 0.5V) before the transistor starts to turn on. On the other hand, a PNP transistor turns on when the base voltage is more than 0.5V lower than the emitter voltage.

If we wanted to switch the positive side, we could use a PNP transistor (as shown in the PNP alternative to Figure 3-21) displayed in Figure 3-24.

MOSFETs also have their own equivalent of PNP transistors called P-channel, their version of the more common NPN being called N-channel.

Common Transistors

The transistors in Table 3-3 will cover a wide range of transistor applications. There are thousands and thousands of other transistors, but in this book we only really use them for switching, so these will cover most "bases"!

Name	Appendix Code	Type	Max Switching Current	Notes
Low/medium-current switching				
2N3904	S1	NPN bipolar	200mA	Current gain about 100
2N3906	S4	PNP bipolar	200mA	Current gain about 100
2N7000	S3	N-channel MOSFET	200mA	2.1V gate-source threshold voltage; turns on when gate is 2.1V higher than source
High-current switching				
FQP30N06	S6	N-channel MOSFET	30A	2.0V gate-source threshold voltage; turns on when gate is 2.0V higher than source

TABLE 3-3 Really Useful Transistors

How to Use a Power MOSFET to Control a Motor

Figure 3-25 shows the schematic symbol and the pinout for the FQP30N06 N-channel MOSFET.

This MOSFET is capable of controlling loads of up to 30A. We are not going to push it any way near that far, we are just going to use it to control the power to a small electric motor that might have a peak load of 1 or 2A. While this would be too much for the bipolar transistors that we have been using so far, this MOSFET will hardly notice!

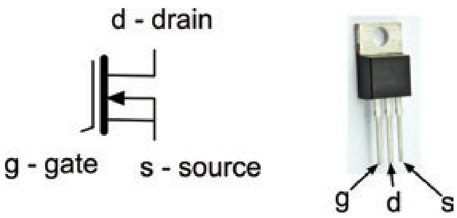

Figure 3-25 The FQP30N06 N-channel MOSFET

You Will Need

To try out this high-power MOSFET, you will need the following items.

Quantity	Item	Appendix Code
1	Solderless breadboard	T5
	Solid-core jumper wire	T6
1	4 × AA battery holder	H1
1	4 × AA batteries	
1	Battery clip	H2
1	Multimeter	T2
1	10kΩ trimpot	K1
1	FQP30N06 MOSFET	S6
1	6V DC motor or gear motor	H6

The DC motor can be any small motor you can find that is around 6V. A motor rated at 12V should still turn at 6V. To test it, just connect its terminals directly to the 6V battery.

Breadboard

The schematic diagram for what we will make is shown in Figure 3-26.

The variable resistor will control the voltage at the gate of the MOSFET. When that gate voltage exceeds the gate threshold, the transistor will turn on and the motor will start.

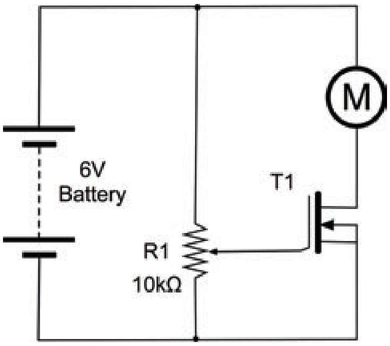

Figure 3-26 A schematic for the MOSFET experiment

Figure 3-27 The breadboard layout for the MOSFET experiment

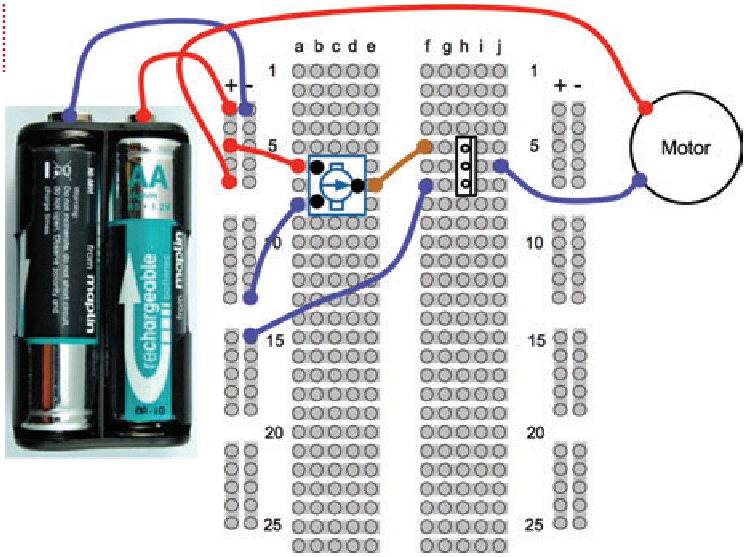

The breadboard layout for the project and a photograph of the experiment in action are shown in Figure 3-27 and Figure 3-28.

To connect the motor to the breadboard, you will probably need to solder a pair of leads to it. It does not matter which way around you connect the motor. The polarity just determines which direction the motor turns. So if you swap the motor leads over, it will turn in the opposite direction.

Try turning the knob on the variable resistor. You will see that you do not have a great deal of control over the speed of

Figure 3-28 The MOSFET experiment

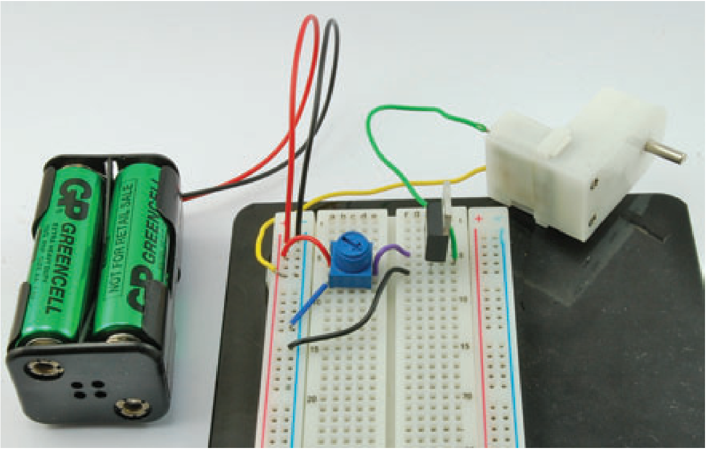

the motor. If you hover around the threshold voltage, you can control the motor speed, but you can probably see why the MOSFET is most commonly used as a switch that is either on or off.

This kind of MOSFET is called a logic level MOSFET, because its gate voltage is low enough to be controlled directly by digital output pins on a microcontroller. This is not true of all MOSFETs. Some have gate threshold voltages of 6V or more.

In Chapter 7, you will use a MOSFET to finely control the motor speed.

How to Select the Right Switch

On the face of it, a switch is a very simple thing. It closes two contacts, making a connection. Often, that is all you need, but other times you will require something more complicated. For example, let's say you want to switch two things at the same time.

There are also switches that only make the contact while you are pressing them, or ones that latch in one position. Switches may be push button, toggle, or rotary. There are many options to choose from and in this section we will attempt to explain the options.

Figure 3-29 shows a selection of switches.

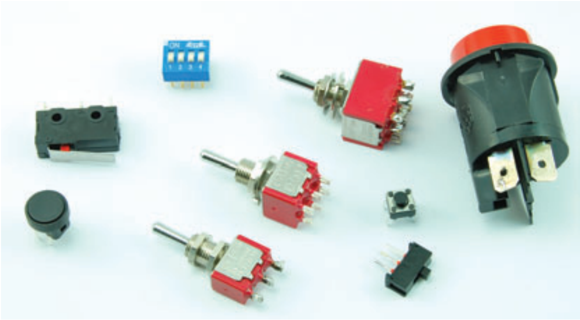

FIGURE 3-29 Switches

Push-Button Switches

Where so many things use a microcontroller, a simple push switch is probably the most common type of switch (Figure 3-30).

This kind of switch is designed to be soldered directly onto a circuit board. It will also fit onto our breadboard, which makes it quite handy.

The confusing thing about this switch is that it has four connections where you would only expect there to be two. Looking at Figure 3-30, you can see that connections B and C are always connected together, as are A and D. However, when the button is pressed, all four pins are all connected together.

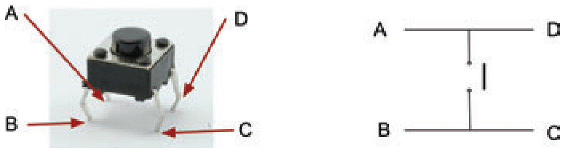

FIGURE 3-30 A push switch

This does mean that you need to be careful to find the right pins or your switch will be connected all the time.

If there is any doubt about how the switch works, use your multimeter set to Continuity mode to work out what is connected to what—first without the switch pressed and then with the switch pressed.

Microswitches

A microswitch is another type of handy switch. They are not designed to be pressed directly, but are often used in things like microwave ovens to detect that the door is closed, or as anti-tamper switches that detect when the cover is removed from an intruder alarm box.

Figure 3-31 shows a microswitch—with three pins!

The reason a microswitch has three pins rather than just two is that it is what is known as a "double throw" or "change-over" switch. In other words, there is one common connection C and two other connections. The common connection will always be connected to one of those contacts, but never both at the same time. The normally open (n.o.) connection is only closed when the button is pressed; however, the normally closed (n.c.) connection is normally closed, and only opens when the button is released.

If you have one of these switches, you might like to connect your multimeter to it. Attach one lead to the common connection and use it to find the n.c. connection, then press the button and the beep should stop.

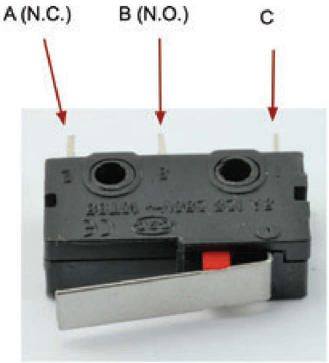

FIGURE 3-31 A microswitch

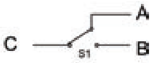

Toggle Switches

If you look through a component catalog (which every good electronics hacker should), you will find a bewildering array of toggle switches. Some will be described as DPDT, SPDT, SPST, or SPST, momentary on, and so forth.

Let's untangle some of this jargon, with a key for these cryptic letters:

- D = Double
- S = Single
- P = Pole
- T = Throw

So, a DPDT switch is double pole, double throw. The word "pole" refers to the number of separate switch contacts that are controlled from the one mechanical lever. So, a double pole switch can switch two things on and off independently. A single throw switch can only open or close a contact (or two contacts if it is double pole). However, a double throw switch can make the common contact be connected to one of two other contacts. So, a microswitch is a double throw switch because it has both normally closed and normally open contacts.

Figure 3-32 summarizes this.

Notice in Figure 3-32 that when drawing a schematic with a double pole switch, it is normal to draw the switch as two switches (S1a and S1b) and connect them with a dotted line to show they are linked mechanically.

The matter is further complicated because you can have three poles or even more on a switch, and double throw switches are sometimes sprung, so they do not stay in one or both of these positions. They may also have a center-off position where the common contact is not connected to anything.

FIGURE 3-32 Toggle switches—poles and throws

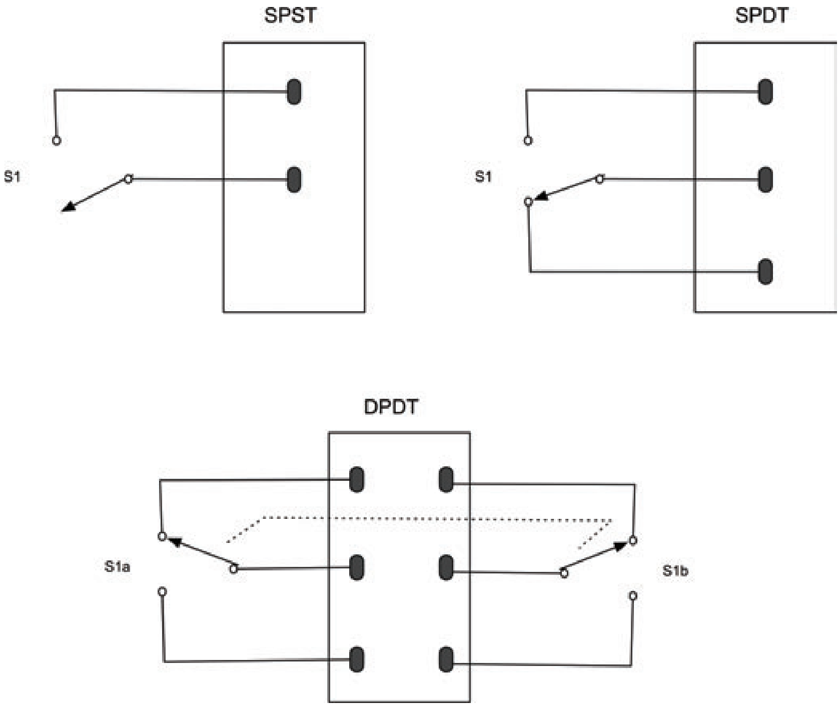

You might see a switch described as "DPDT, On-Off-Mom." Well, we know what the DPDT bit means. It will have six legs for a start. The "On-Off-Mom" part means that it also has a center position, where the common connection is not made to anything. Switch it one way and it will be on to one set of contacts and stay in that position. Switch it the other way and it will be sprung to return to the central position, allowing you to make a "momentary" connection.

A lot of this terminology applies to other kinds of switches in addition to toggle switches.

Summary

We now know a bit about voltage, current resistance, and power. In the next chapter, we will use these ideas in looking at how to use LEDs.

UNO
ARDUINO
DIGITAL (PWM~)

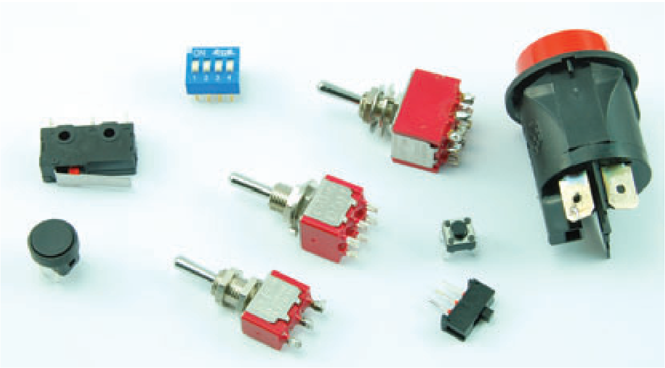

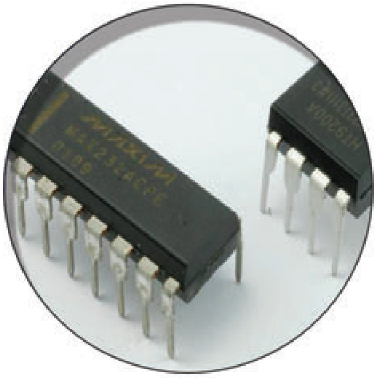
MAXIM

4

LEDs

LEDs (light-emitting diodes) are diodes that emit light when a current passes through them. Well on the way to completely replacing filament light bulbs in almost all applications, they can be used as indicators and, with the very high brightness types of LED, can provide illumination.

They are much more efficient than conventional light bulbs, producing far more light per watt of power, and are a lot less delicate.

LEDs do, however, require a little more thought when used. They have to be powered with the correct polarity and require circuitry to limit the current flowing through them.

How to Stop an LED from Burning Out

LEDs are delicate little things and quite easy to destroy accidentally. One of the quickest ways of destroying an LED is to attach it to a battery without using a resistor to limit the current.

To get to grips with LEDs, we will put three different color LEDs on our breadboard (Figure 4-1).

You Will Need

Quantity	Names	Item	Appendix Code
1		Solderless breadboard	T5
1	D1	Red LED	K1
1	D2	Yellow LED	K1
1	D3	Green LED	K1
1	R1	330Ω resistor	K2
2	R2, R3	220Ω resistor	K2
		Jumper wires	T6
1		4 × AA battery holder	H1
1		Battery clip	H2
4		AA batteries	

Diodes

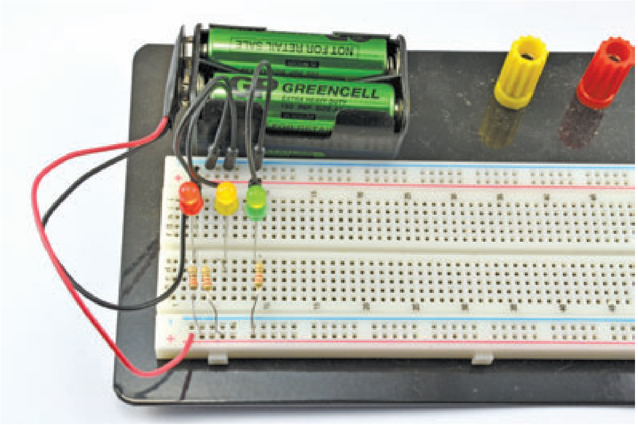

Figure 4-1 LEDs on a breadboard

We need to understand LEDs a little better if we are to successfully use them. LED stands for light-emitting diode, so let's start by looking at what a diode is (Figure 4-2).

A diode is a component that only lets current flow in one direction. It has two leads, one called the anode, and the other called the cathode. If the anode is at a higher voltage than the cathode (it has to be greater by about half a volt), then it will conduct electricity and is said to be "forward-biased." If on the other hand the anode isn't at least half a volt higher than the cathode, it is said to be "reverse-biased" and no current flows.

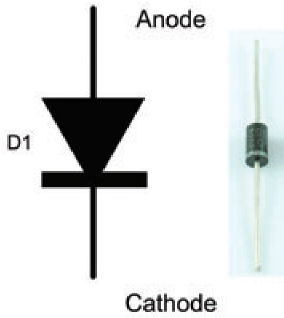

Figure 4-2 A diode

LEDs

An LED is just like a regular diode except that when it is forward-biased, it conducts and generates light. It also differs from a regular diode in that the anode usually needs to be at least 2V higher than the cathode for it to be forward-biased.

Figure 4-3 shows the schematic diagram for driving an LED.

The key to this circuit is to use a resistor to limit the current flowing through the LED. A normal red LED will typically just be lit at about 5mA and is designed to be used at around 10 to 20mA (this is called the "forward current" or I_F). We will aim for 15mA for our LED. We can also assume that when it is conducting, there will be about 2V across it. This is called the "forward voltage" or V_F. That means there will be 6 – 2 = 4V across the resistor.

So, we have a resistor that has a current flowing through it (and the LED) of 15mA and a voltage across it of 4V. We can use Ohm's law to calculate the value of resistance we need to achieve this:

$$R = V / I = 4V / 0.015A = 267\Omega$$

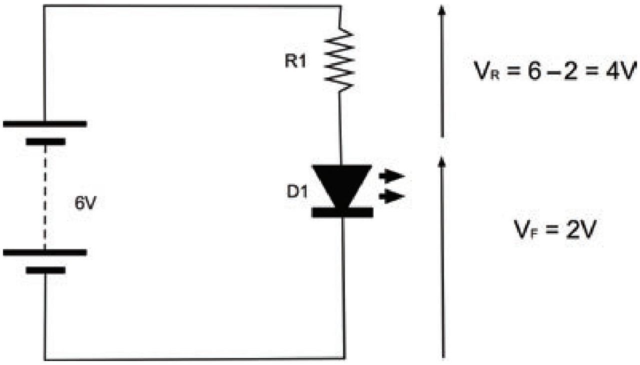

Figure 4-3 Limiting current to an LED

Parameter	Red	Green	Yellow	Orange	Blue	Units
Maximum forward current (I_F)	25	25	25	25	30	mA
Typical forward voltage (V_F)	1.7	2.1	2.1	2.1	3.6	V
Maximum forward voltage	2	3	3	3	4	V
Maximum reverse voltage	3	5	5	5	5	V

TABLE 4-1 An LED Datasheet

Resistors come in standard values, and the nearest higher value in our resistor starter kit is 330Ω.

As I mentioned earlier, a red LED will almost always light quite brightly with something like 10–20mA. The exact current is not critical. It needs to be high enough to make the LED light, but not exceed the maximum forward current of the LED (for a small red LED, typically 25mA).

Table 4-1 shows a section of the datasheet for a typical range of LEDs of different colors. Note how V_F changes for different color LEDs. This will mean you may need to use a different resistor, but usually if the supply voltage is above say 6V, then small variations in V_F for color will not require a different resistor value.

The other parameter you should be aware of is the "maximum reverse voltage." If you exceed this by, say, wiring your LED the wrong way around, it is likely to break the LED.

Many online series resistor calculators are available that—given the supply voltage V_F and current I_F for your LED—will calculate the series resistor for you. For example:

www.electronics2000.co.uk/calc/led-series-resistor-calculator.php

Table 4-2 is a useful rough guide, assuming a forward current of around 15mA.

Supply Voltage (V)	Red	Green, Yellow, Orange	Blue
3	91Ω	60Ω	none
5	220Ω	180Ω	91Ω
6	270Ω / 330Ω	220Ω	180Ω
9	470Ω	470Ω	360Ω
12	680Ω	660Ω	560Ω

TABLE 4-2 Series Resistors for LEDs

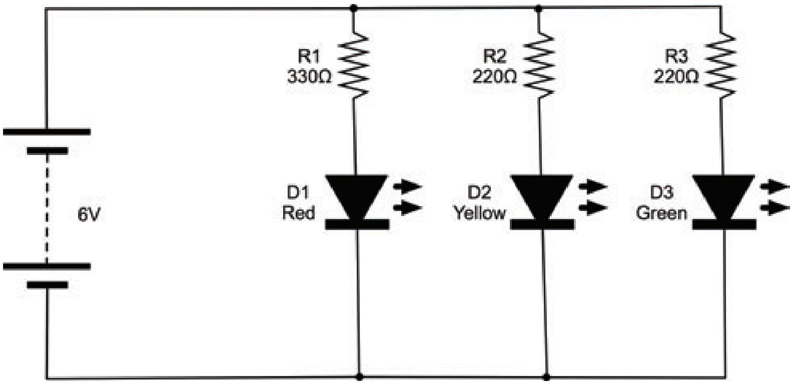
R1
330Ω
R2
220Ω
R3
220Ω
6V
D1
Red
D2
Yellow
D3
Green

Figure 4-4 An LED's schematic

Trying It Out

You might like to try out your LEDs and get them lit up on the breadboard. So, using Figures 4-4 and 4-5 as a guide, wire up your breadboard. Remember that the longer lead of the LED is normally the anode (positive) and thus should be to the left of the breadboard.

An important point to notice here is that each LED has its own series resistor. It is tempting to use one lower value current limiting resistor and put the LEDs in parallel, but don't do this. If you do, the LED with the lowest V_F will hog all the current and probably burn out, at which point the LED with the next lowest V_F will do the same, until all the LEDs are dead.

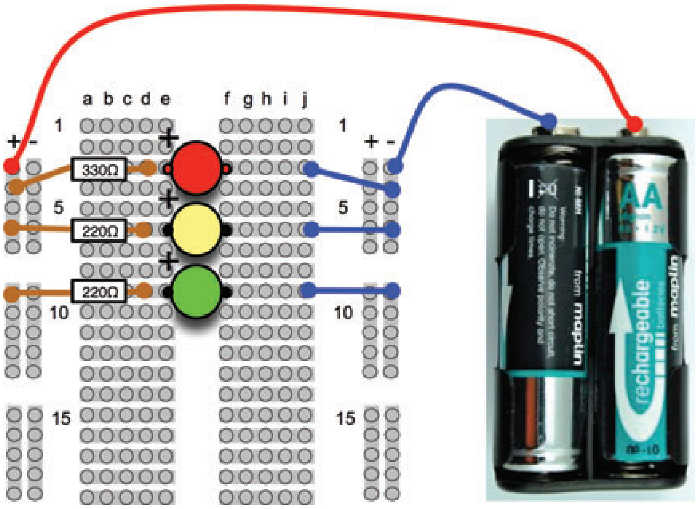
a b c d e
f g h i j
1
5
10
15
+ -
330Ω
220Ω
220Ω
+
AA
rechargeable
from maplin

Figure 4-5 An LED breadboard layout

How to Select the Right LED for the Job

LEDs come in all colors, shapes, and sizes. Many times, you just want a little indicator light, in which case a standard red LED is usually fine. However, there are many other options, including LEDs bright enough to be used as lamps.

You Will Need

Quantity	Names	Item	Appendix Code
1		Solderless breadboard	T5
1	D1	RGB common cathode LED	S4
3	R1–R3	500Ω trimpot	R3
1	R1	330Ω resistor	K2
2	R2, R3	220Ω resistor	K2
		Jumper wires	T6
1		4 × AA battery holder	H1
1		Battery clip	H2
4		AA batteries	

Brightness and Angle

When selecting an LED, they may simply be described as “standard” or “high brightness” or “ultra-bright.” These terms are subjective and open to abuse by unscrupulous vendors. What you really want to know is the LED’s luminous intensity, which is how much light the LED produces. You also want to know the angle over which the LED spreads the light.

So, for a flashlight, you would use LEDs with a high luminous intensity and a narrow angle. Whereas for an indicator light to show that your gadget is turned on, you would probably use an LED with a lower luminous intensity but a wider angle.

Luminous intensity is measured in millicandela or mcd, and a standard indicator type LED will typically be around 10 to 100 mcd, with a fairly wide viewing angle being 50 degrees. A “high brightness” LED might be up to 2000 or 3000 mcd, and an ultra-bright anything up to 20,000 mcd. A narrow beam LED is about 20 degrees.

Multicolor

We have already explored the more common LED colors, but you can also get LED packages that actually contain two or three LEDs of different colors in the same package. Common varieties are red/green as well as full-color RGB (Red Green Blue). By varying the proportion of each color, you can change the color of the light produced by the LED package.

Figure 4-6 shows the schematic we can use to try out a little experiment with an RGB LED.

We are going to use a variable resistor with each of the red, green, and blue LEDs. The fixed resistors (R4, R5, and R6) are to prevent too much current flowing when the slider of the variable resistor is right at 6V.

Figure 4-7 shows the breadboard layout for this. The common lead of the LED is the longest lead, while the other three are the three-color anodes.

Once all the components are in the board and you have attached the battery, you should be able to mix various colors by changing the position of the three sliders. Figure 4-8 shows the circuit in operation.

IR and UV

As well as visible LEDs, you can also buy LEDs whose light is invisible. This is not as pointless as you might think. Infrared LEDs are used in TV remote controls, and ultraviolet LEDs are

Figure 4-6 An RGB LED test schematic

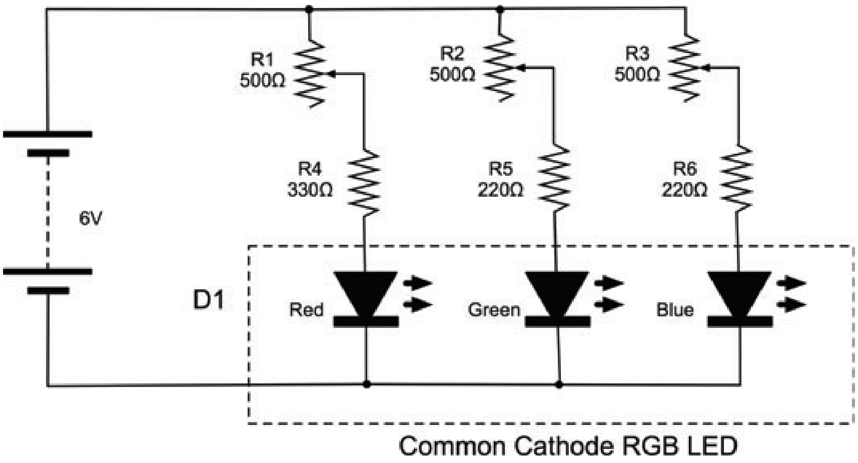

FIGURE 4-7 An RGB LED test breadboard layout

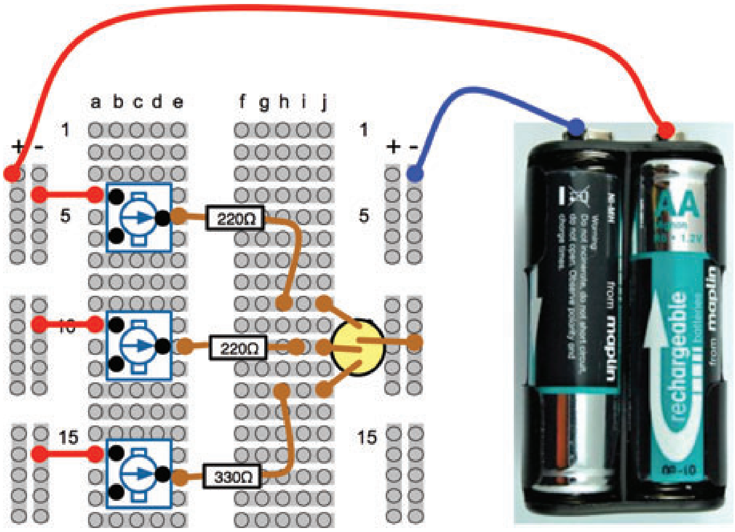

used in specialist applications such as checking the authenticity of bank notes and making people's white clothes light up in clubs.

FIGURE 4-8 An RGB test

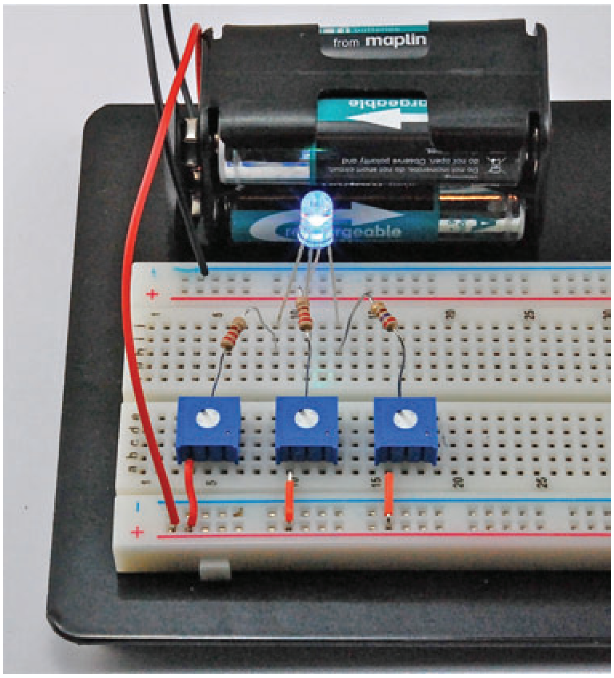

Use these LEDs just like any other LED. They will have a recommended forward current and voltage and will need a series resistor. Of course, checking that they are working is trickier. Digital cameras are often a little sensitive to infrared and you may see a red glow on the screen.

LEDs for Illumination

LEDs are also finding their way into general household lighting. This has come about because of improvements in LED technology that have produced LEDs with a brightness comparable to incandescent light bulbs—well, they are getting there anyway. Figure 4-9 shows one such high-brightness LED. In this case, it's a 1-W LED, although 3-W and 5-W LED modules are also available.

The cool-looking star shape is thanks to the aluminum heat sink that the LED is attached to. At full power, these LEDs produce enough heat to warrant such a heat sink to disperse the heat into the air.

These LEDs can use a resistor to limit the current, but a quick calculation will show you that you will need quite a high-power resistor. A better approach to using these LEDs is to use a constant current driver, which is the subject of the next section.

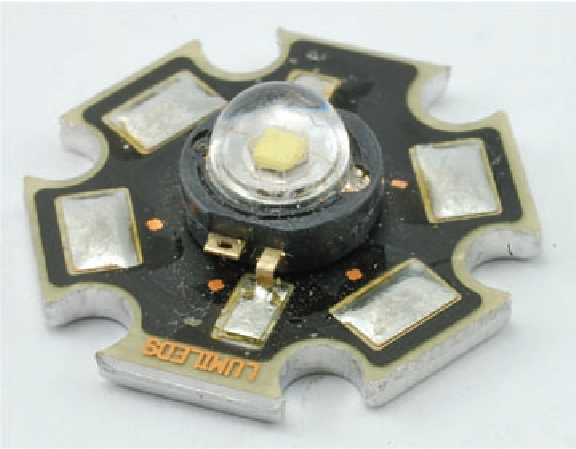

Figure 4-9 A high-power LED

How to Use a LM317 to Make a Constant Current Driver

Using a resistor to limit the current is all right for small LEDs. However, it is a bit hit or miss since it is very dependent on the LED being used and the power being provided. So for low-power LEDs, where the supply current is not critical it works okay. For high-power LEDs, you can use a series resistor (it will need to be quite high power), but a better way is to use a constant current driver.

As the name suggests, the constant current driver will supply the same current whatever voltage it is supplied with and whatever the forward voltage of the LED. You just set the current and that is how much current will flow through the high-power LED.

A very useful IC that is often used for this purpose is the LM317. This IC is primarily intended as an adjustable voltage regulator, but can easily be adapted for use in regulating current.

This project will start off on breadboard and then we will cut the top off a battery clip and solder the LM317 and resistor to it to make an emergency 1-W LED light.

You Will Need

Quantity	Names	Item	Appendix Code
1		Solderless breadboard	T5
1	D1	1-W white Lumileds LED	S3

Quantity	Names	Item	Appendix Code
3	R1	4.7Ω resistor	K2
1		Battery clip (to destroy)	H2
1		PP3 9V battery	
		Jumper wires	T6

Design

Figure 4-10 shows the schematic diagram for regulating the current to a high-power LED like the one shown in Figure 4-9.

The LM317 is very easy to use in a constant current mode. It will always strive to keep its output voltage at exactly 1.25V above whatever voltage the Adj (adjust) pin is at.

The LED we are going to use is a 1W white light LED. It has an I_f (forward current) of 300mA and a V_f (forward voltage) of 3.4V.

The formula for calculating the right value for R1 for use with the LM317 is:

$$R = 1.25V / I$$

so in this case, R = 1.25 / 0.3 = 4.2Ω

If we used a standard resistor value of 4.7Ω, then this would reduce the current to:

$$I = 1.25V / 4.7\Omega = 266 \text{ mA}$$

Checking the power rating for the resistor, the LM317 will always have 1.25V between Out and Adj. So:

$$P = V \times I = 1.25V \times 266mA = 0.33W$$

Figure 4-10 An LM317 constant current LED driver schematic

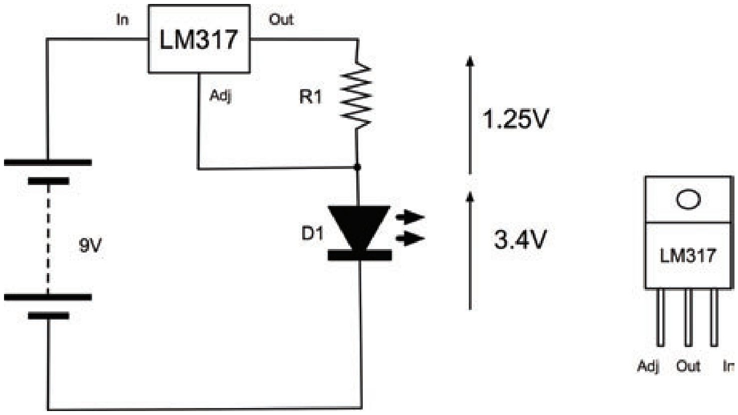

A half-watt resistor will therefore be fine.

The LM317 also needs its input to be about 3V higher than its output to guarantee 1.25V between Adj and the output. This means that a 6V battery would not be quite high enough because the forward voltage is 3.4V. However, we could drive the circuit using a 9V battery or even a 12V power supply without modification, since whatever the input voltage, the current will always be limited to about 260mA.

A quick calculation of the power consumed by the LM317 will reassure us that we are not going to come near exceeding its maximum power rating.

For a 9V battery, the voltage between In and Out will be 9 – (1.25 + 3.4) = 4.35V. The current is 260mA, so the power is: 4.35 × 0.26 = 1.13W.

According to its data sheet, the maximum power handling capability of the LM317 is 20W, and it can cope with a current of up to 2.2A for a supply voltage of less than 15V. So we are fine.

Breadboard

Figure 4-11 shows the breadboard layout for this, and Figure 4-12 displays the actual breadboard. These LEDs are almost painfully bright, so avoid staring at them. When working with them, I cover them with a sheet of paper so I can see when they are on without being temporarily blinded!

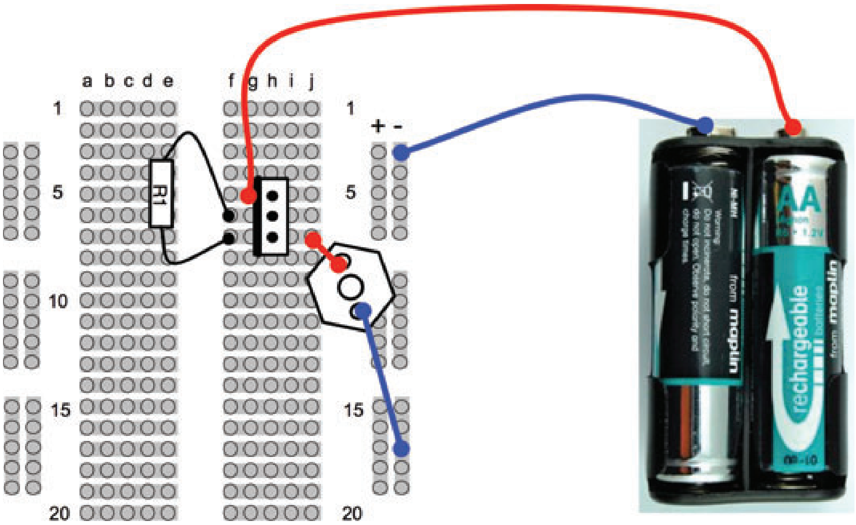

FIGURE 4-11 The LED constant current driver breadboard layout

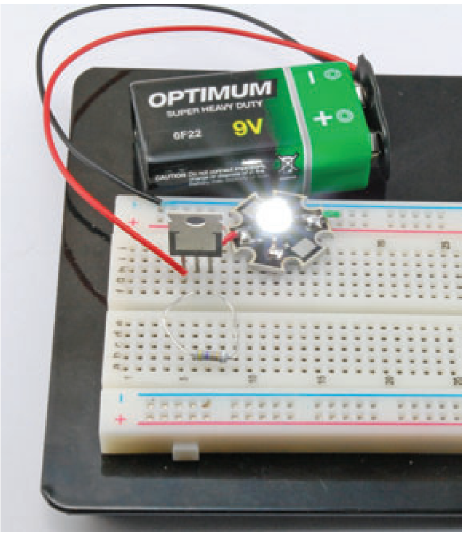

FIGURE 4-12 The LED constant current driver

You will need to solder lengths of solid-core wire to the LED's terminals so it can plug into the breadboard. It is a good idea to leave the insulation on so there is no chance of the bare wires touching the heat sink and shorting.

Construction

We will use this to make a little emergency lantern by hacking a battery clip to build the electronics on top of it so it can be clipped on top of a PP3 battery in the event of a power failure (Figure 4-13).

Figures 4-14a thru 4-14d shows the stages involved in soldering this up.

First, remove the plastic from the back of the battery clip using a craft knife. Then, unsolder the exposed leads (Figure 4-14a).

The next step (Figure 4-14b) is to solder the Input lead of the LM317 to the positive terminal of the battery clip. Remember that the positive connector on the clip will be the opposite of the connector on the battery itself, so the positive connector on the clip is the socket-shaped connector. Gently bend the leads of the LM317 apart a little to make this easier.

Now solder the LED in place making sure the cathode of the LED goes to the negative connection on the battery clip (Figure 4-14c).

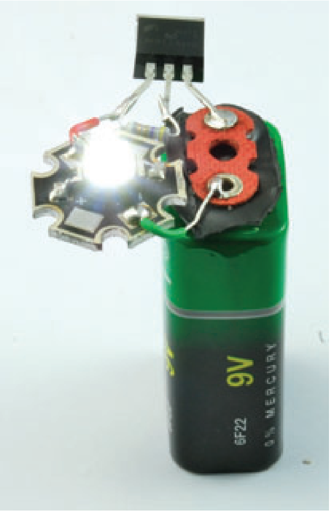

FIGURE 4-13 Emergency LED lighting

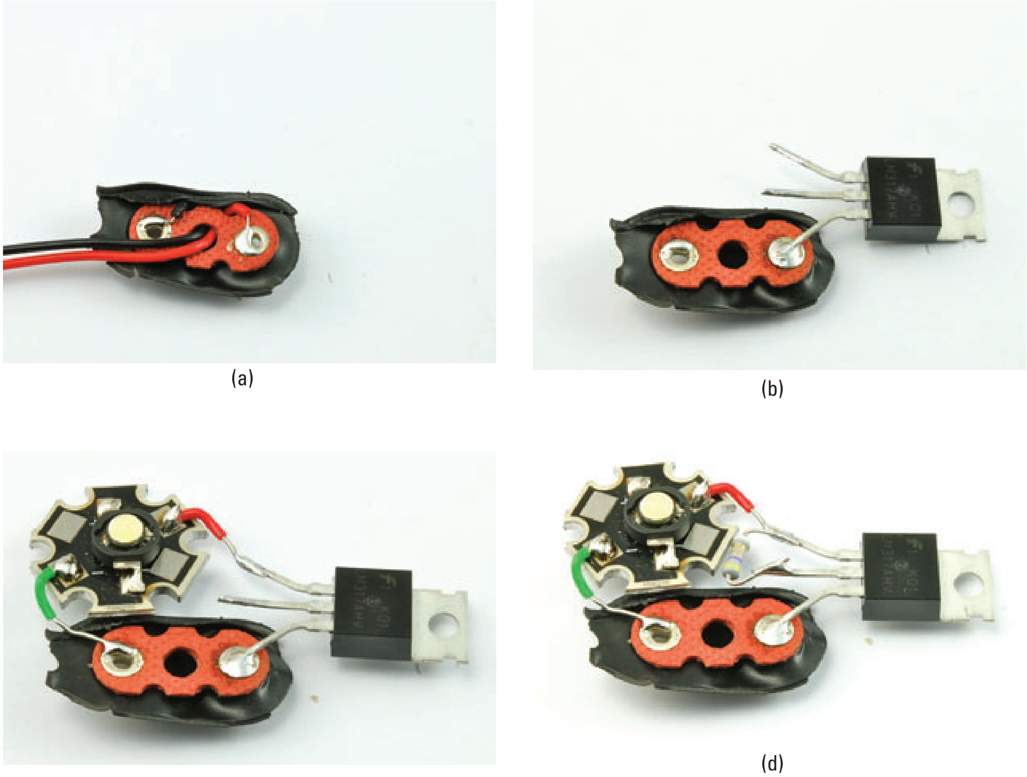

(a)

(b)

(c)

(d)

Figure 4-14 Making an emergency 1-W LED light

Finally, solder the resistor across the two topmost leads of the LM317 (Figure 4-14d).

How to Measure the Forward Voltage of an LED

If you want to power lots of LEDs at the same time, it is a good idea to test a few of the LEDs you actually intend to use and measure the forward voltage at the current you intend to use. Figure 4-15 shows how you would do this.

Figure 4-15a shows the schematic diagram. A variable resistor is used to vary the current through the LED and when the desired forward current is set, the voltage can be read.

The current and voltage do not have to be read at the same time, but if you do have two meters, this does make life easier.

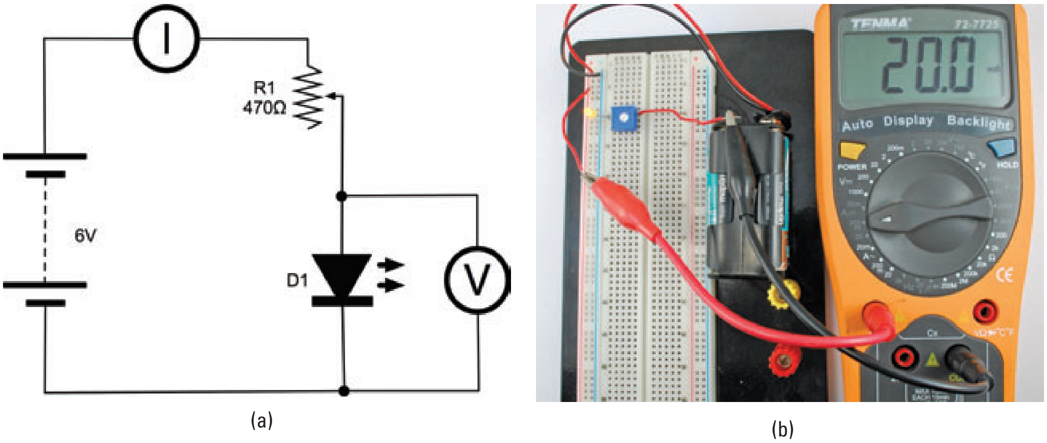
I
R1
470Ω
6V
D1
V
TENMA
72-7725
20.0
Auto Display Backlight
POWER
HOLD

(a)

(b)

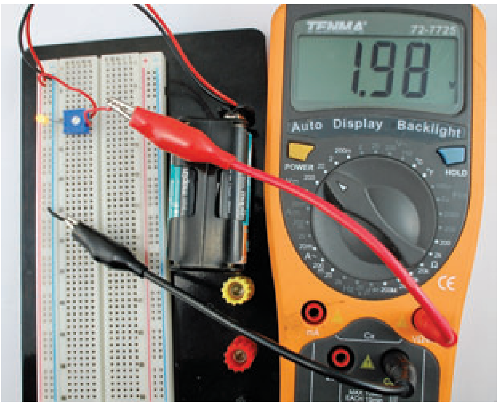
TENMA
72-7725
1.98
Auto Display Backlight
POWER
HOLD

(c)

Figure 4-15 Measuring the forward voltage of an LED

Set the variable resistor to its middle point and wire up the breadboard as shown in Figure 4-15b. You will probably have to move the positive lead of your multimeter to a different socket when measuring current. Select a current range of 200mA DC. Now adjust the variable resistor until the current reads 20mA.

We can now measure the voltage across the LED. To do this, first disconnect the multimeter, put the positive multimeter back in the right socket for measuring voltage, and then change the range to 20V DC. Wire it up as shown in Figure 4-15c and measure the voltage. In this case, it was 1.98V.

You Will Need

Quantity	Names	Item	Appendix Code
1		Solderless breadboard	T5
1	D1	LED	K1
3	R1	500Ω trimpot	R3
		Jumper wires	T6
1		4 × AA battery holder	H1
1		Battery clip	H2
4		AA batteries	

How to Power Large Numbers of LEDs

If you use something like a 12V power supply, you can put a number of LEDs in series, with just one LED. In fact, if you know the forward voltages fairly accurately, and the power supply is well regulated, you can get away without any series resistor at all.

So, if you have fairly standard LEDs that have a forward voltage of 2V, then you could just put six of them in series. However, it will not be terribly easy to predict how much current the LEDs will take.

A safer approach is to arrange the LEDs in parallel strings, each string having its own current-limiting resistor (Figure 4-16).

Although the math for this isn't too hard, there can be a fair bit of it, so you can save yourself a lot of time by using an online calculator like http://led.linear1.org/led.wiz (Figure 4-17).

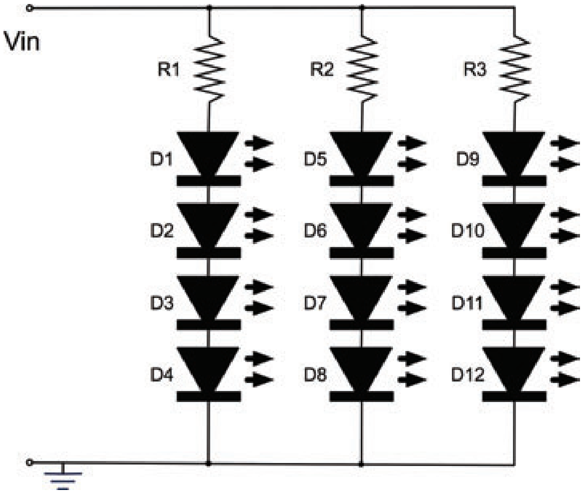

Figure 4-16 Powering multiple LEDs

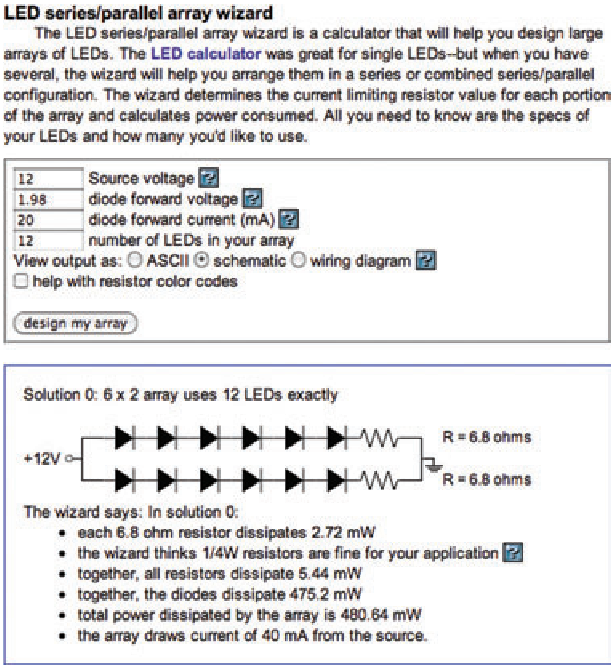

LED series/parallel array wizard

The LED series/parallel array wizard is a calculator that will help you design large arrays of LEDs. The LED calculator was great for single LEDs--but when you have several, the wizard will help you arrange them in a series or combined series/parallel configuration. The wizard determines the current limiting resistor value for each portion of the array and calculates power consumed. All you need to know are the specs of your LEDs and how many you'd like to use.

12	Source voltage
1.98	diode forward voltage
20	diode forward current (mA)
12	number of LEDs in your array

View output as: ○ ASCII ● schematic ○ wiring diagram

☐ help with resistor color codes

Solution 0: 6 x 2 array uses 12 LEDs exactly

The wizard says: In solution 0:

- each 6.8 ohm resistor dissipates 2.72 mW
- the wizard thinks 1/4W resistors are fine for your application
- together, all resistors dissipate 5.44 mW
- together, the diodes dissipate 475.2 mW
- total power dissipated by the array is 480.64 mW
- the array draws current of 40 mA from the source.

Figure 4-17 The LED wizard

In this particular designer, you enter the source voltage for the overall supply, the LED forward voltage, the desired current for each LED, and the number of LEDs you want to light. The wizard then does the math and works out a few different layouts.

One consideration is that where you have a string of LEDs in series, if any of the LEDs fail, then all the LEDs will be off.

How to Make LEDs Flash

The 555 timer IC is a useful little IC that can be used for many different purposes, but is particularly convenient for making LEDs flash or generating higher frequency oscillations suitable for making audible tones (see Chapter 9).

We are going to make this design on breadboard and then transfer it to a more permanent home on a bit of stripboard.

You Will Need

Quantity	Names	Item	Appendix Code
1		Solderless breadboard	T5
1	D1	LED red	K1
1	D1	LED green	K1
1	R1	1kΩ resistor	K2
1	R2	470kΩ resistor	K2
2	R3, R4	220Ω resistor	K2
1	C1	1µF capacitor	K2
1	IC1	555 timer	K2
		Jumper wires	T6
1		4 × AA battery holder	H1
1		Battery clip	H2
4		AA batteries	

Breadboard

The schematic for the LED flasher is shown in Figure 4-18.

The breadboard layout is shown in Figure 4-19. Make sure you have the IC the right way up. There will be a notch in the IC body next to the top (pins 1 and 8). The capacitor and LEDs must both be the correct way around, too.

Figure 4-20 shows the finished breadboard. You should find that the LEDs alternate, each staying on for about a second.

Now that we know the design is right and everything works, try swapping out R2 with a 100kΩ resistor and notice the effect on the flashing.

The 555 timer is a very versatile device, and in this configuration it oscillates at a frequency determined by the formula:

frequency = 1.44 / ([R1 + 2 * R2] * C)

where the units of R1, R2, and C1 are in Ω and F. Plugging in the values for this design, we get:

frequency = 1.44 / ([1000 + 2 * 470000] * 0.000001) = 1.53 Hz

One hertz (or Hz) means one oscillation per second. When we use the 555 timer in a later chapter to generate an audible tone, we will be using the same circuit to generate a frequency in the hundreds of hertz.

As with so many electronic calculations, there are also online calculators for the 555 timer.

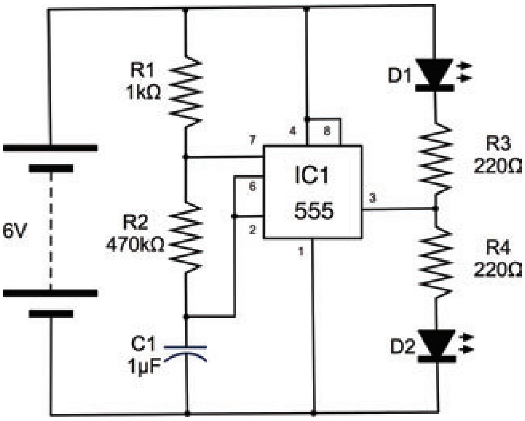
R1
1kΩ
D1
7
4
8
IC1
555
R3
220Ω
6
3
2
1
6V
R2
470kΩ
R4
220Ω
C1
1µF
D2

FIGURE 4-18 The LED flasher schematic

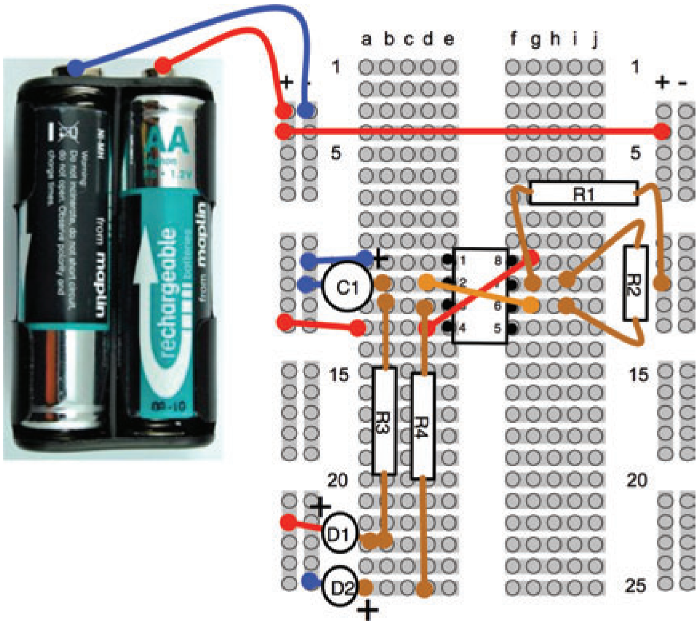
a b c d e
f g h i j
1
5
15
20
25
+ -
+
R1
R2
R3
R4
C1
D1
D2
1
8
2
6
4
5
AA
rechargeable
maplin

FIGURE 4-19 The LED flasher breadboard layout

FIGURE 4-20 The LED flasher on breadboard

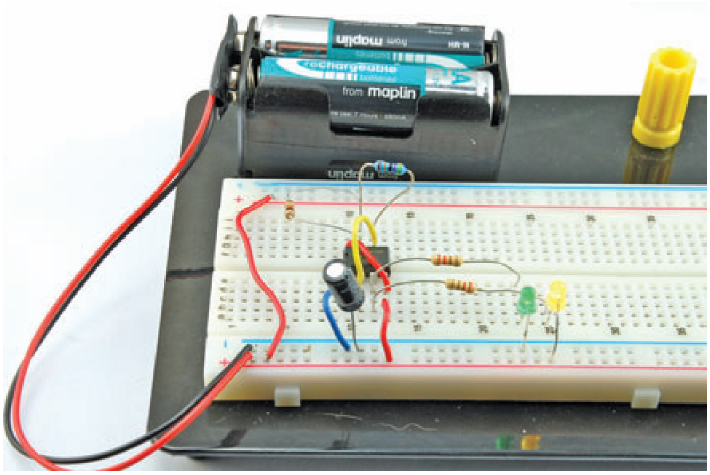

How to Use Stripboard (LED Flasher)

Breadboard is very useful for trying things out, but not so useful as a permanent home for your electronics. The problem is that the wires tend to fall out, and it's all a bit big and bulky.

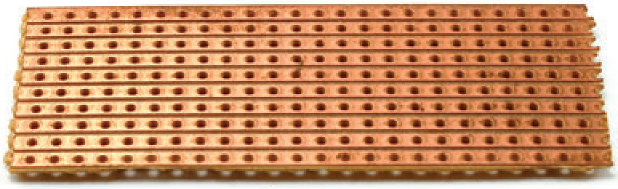

FIGURE 4-21 Stripboard

Stripboard (Figure 4-21) is a bit like general-purpose printed circuit board. It is a perforated board with conductive strips running underneath, rather like breadboard. The board can be cut to the size you need and components and wires soldered onto it.

Designing the Stripboard Layout

Figure 4-22 shows the final stripboard layout for the LED flasher that we made in the previous section. It is not easy to explain how we got to this from the schematic and breadboard layout. There is a certain amount of trial and error, but there are a few principals you can follow to try and make it easier.

The first is to use a drawing tool with a stripboard template. For Mac users, with OmniGraffle, a template is available for download from the book's web site (www.hackingelectronics.com). There is also an image file that can be printed out and used as a template to sketch out the design.

The Xs underneath the IC are breaks in the track, which we will make with a drill bit. One of the goals of a good stripboard

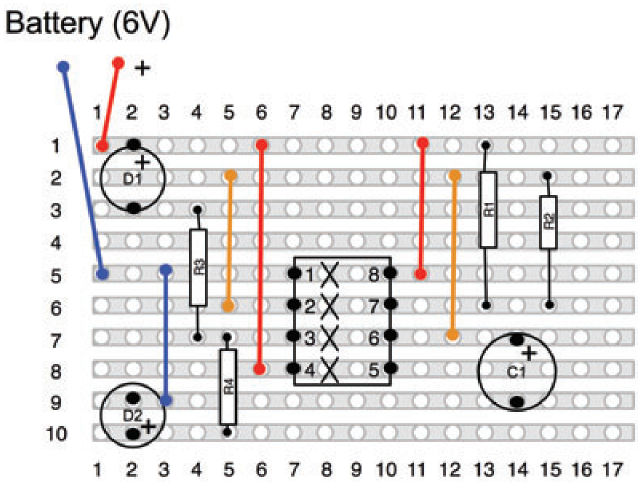

FIGURE 3-22 An LED flasher stripboard layout

layout is to try and avoid making too many breaks in the track. Breaks are unavoidable for an IC like this. If we did not make them, pin 1 would be connected to pin 8, pin 2 to pin 7, and so on, and nothing would work.

The colored lines on the board are linking wires. So, for instance, from the schematic diagram of Figure 4-18, we can see that pins 4 and 8 of the IC should be connected together and both go to the positive supply. This is accomplished by the two red linking wires. Similarly, pins 2 and 6 need to be connected together. This is accomplished by using the orange leads.

Although logically the stripboard layout is the same as the schematic, the components are in rather different places. The LEDs are on the left in the stripboard layout and on the right on the schematic. It is not always like this, and it's easier if they are similar, but in this case the left-hand pins of the IC include the output pin 3 that the LEDs need, and the pins connected to R1, R3, and C1 are all on the right-hand side of the IC.

Try making a stripboard layout from the schematic, you may well come up with a different and better layout than the one I produced.

The steps I went through in designing this layout are as follows:

1. Place the IC fairly centrally, with a bit more room above than below and with pin 1 uppermost (convention).
2. Find a good place for R3 and R4 to be put so the strips are at least three holes apart for one resistor lead, when the other lead of each resistor goes to pin 3.

3. Pick the top track of the stripboard to be +V so it can be close to the positive end of one of the LEDs
4. Pick row 5 to be the ground connection. This way it can run straight on to pin 1 of the IC.
5. Add a link wire from row 5 to row 9 to provide the negative connection for the LED D2.
6. Put a jumper wire from pin 4 of the IC to row 1 (+V).

Turning now to the right-hand side of the board:

1. Put a jumper in connecting pin 8 of the IC to row 1 (+V).
2. R1 and R2 both have one end connected to pin 7, so put them side by side with the far end of R1 going to row 1 (+V).
3. R2 needs to connect to pin 6, but pin 6 and pin 7 of the IC are too close together for the resistor to lie flat, so take that lead up to the unused row 2 instead, then put jumpers from row 2 down to both pin 6 and pin 2 of the IC.
4. Finally, C1 needs to go between pin 6 (or pin 2, but 6 is easier) and GND (row 9).

A good way of checking that you have made all the connections you need is to print off the schematic and then go through each connection on the stripboard and check off its counterpart on the schematic.

This may all seem a little like magic, but try it. It's not as hard to do as it is to describe.

You Will Need

You will need all the components listed in the section "How to Make LEDs Flash," plus the following items.

Quantity	Item	Appendix Code
1	Stripboard 10 strips by 17 holes	H3
1	Soldering kit	T1
1	Drill bit (1/8 inch)	

Before we start soldering, it is worth considering what kind of LEDs you want to use for this project. You may decide to use higher-brightness LEDs or power the project from a lower voltage. If you do decide on this, recalculate the values for R3 and R4 and try it out on the breadboard layout. The 555 timer IC needs a supply voltage between 4.5V and 16V, and the output can supply up to 200mA.

Construction

Step 1. Cut the Stripboard to Size

There is no point in having a large bit of stripboard with just a few components on it, so the first thing we need to do is cut the stripboard to the right size. In this case, it's ten strips each of 17 holes. Stripboard doesn't actually cut very well. You can use a rotary tool, but wear a mask because the dust from the stripboard is nasty and you really do not want it in your lungs. I find the easiest way to cut stripboard is actually to score it with a craft knife and metal ruler on both sides and then break it over the edge of your work surface.

Score it across the holes, not between them. When the board is cut, the copper underside will look like Figure 4-23.

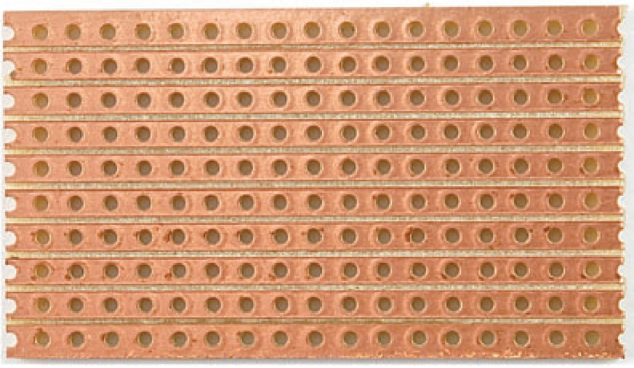

Figure 4-23 A stripboard cut to size

Step 2. Make the Breaks in the Tracks

A good tip is to use a permanent marker and put a dot in the top left corner. Otherwise, it is very easy to get the board turned around, resulting in breaks and links being put in the wrong place.

To make the breaks, count the position in rows and columns of the break from the top of the board layout and then push a bit of wire so you can identify the right hole on the copper side of the stripboard (Figure 4-24a). Using a drill bit, just "twizzle" it between thumb and forefinger to cut through the copper track. It usually only takes a couple of twists (Figure 4-24b and c).

When you have cut all four breaks, the bottom side of the breadboard should look like Figure 4-25. Check very carefully that none of the burrs from the copper have lodged between the tracks and that the breaks are complete. Photographing the board and then zooming in is a great way of actually checking the board.

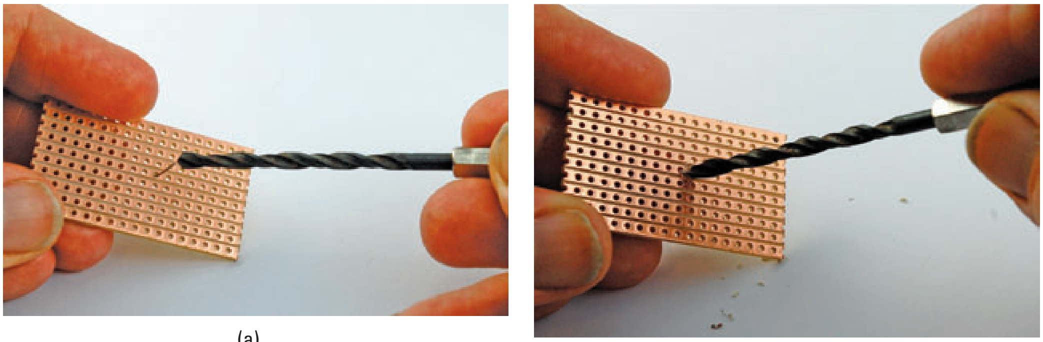

(a)

(b)

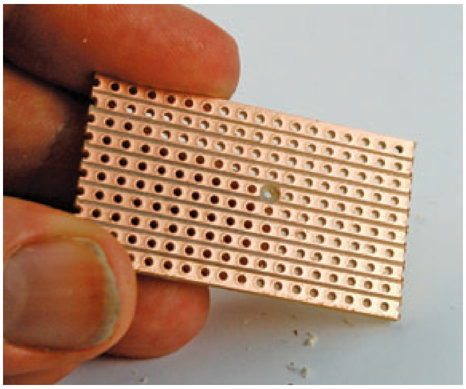

(c)

FIGURE 4-24 Cutting a track on stripboard

Step 3. Make the Wire Links

A golden rule of any type of circuit board construction, including when using stripboard, is to start with the lowest-lying components. This is so that when you turn the board on its back to solder it, the thing being soldered will stay in place through the weight of the board.

In this case, the first thing to solder is the links.

Strip and cut solid-core wire to slightly longer than the length of the link. Bend it into a U-shape and push it through the holes at the top, counting the rows and columns to get the right holes (Figure 4-26a). Some people get very skilled at bending the wires with pliers to just the right length. I find it easier to bend the wires with a bit of a curve so they will kind of squash into the right holes. I find this easier than trying to get the length just right from the start.

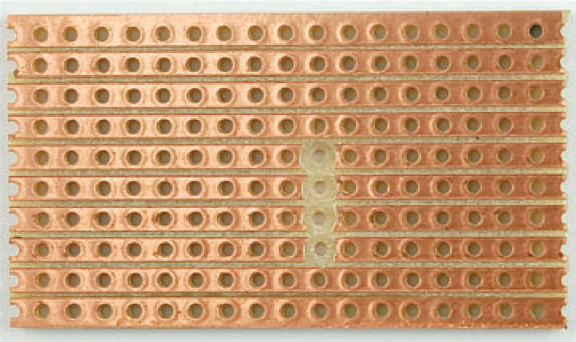

FIGURE 4-25 The stripboard with breaks cut

Turn the board over (see how the wire is held in place) and solder the wire by holding the iron to the point where the wire emerges from the hole. Heat it for a second or two and then apply the solder until it slows along the track, covering the hole and wire (Figure 4-26b and c).

Repeat this process for the other end of the lead and then snip off the excess wire (Figure 4-26d and e).

When you have soldered all the links, your board should look like the one in Figure 4-27.

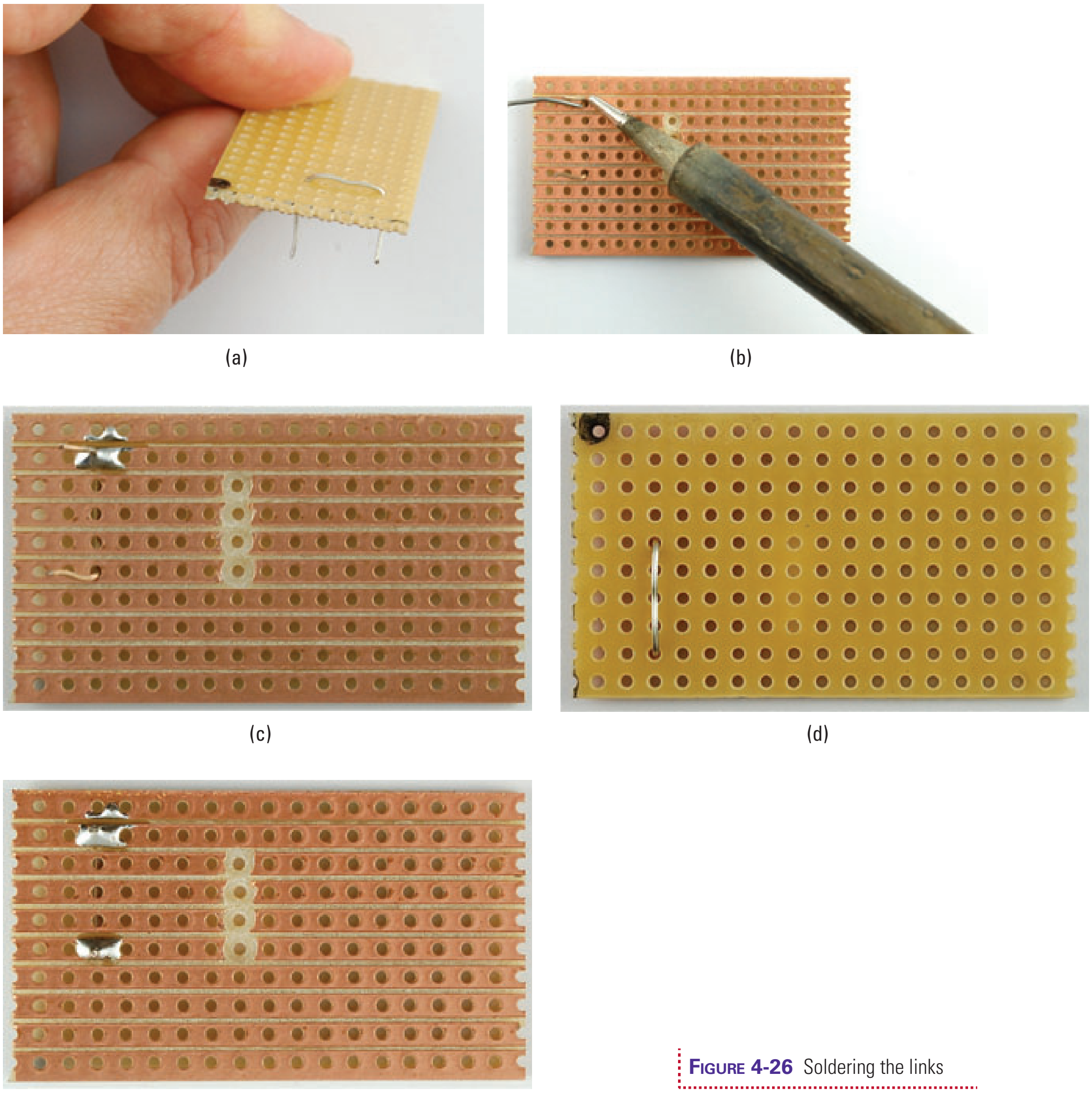

(a)

(b)

(c)

(d)

(e)

Figure 4-26 Soldering the links

Step 4. Resistors

The resistors are the next lowest components to the board, so solder these next, in the same way as you did the links. When they are all soldered, the stripboard should look like Figure 4-28.

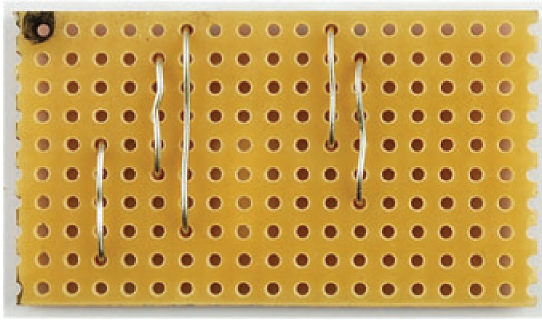

FIGURE 4-27 The stripboard with all its links

Step 5. Solder the Remaining Components

Next, solder the LED, capacitor (which can be laid on its side as shown in Figure 4-29), and finally the LEDs and connectors to the battery clip.

That's it. Now it's time for the moment of truth. Before you plug it in, do a very careful inspection for any shorts on the underside of the board.

If everything seems in order, connect the battery clip to the battery.

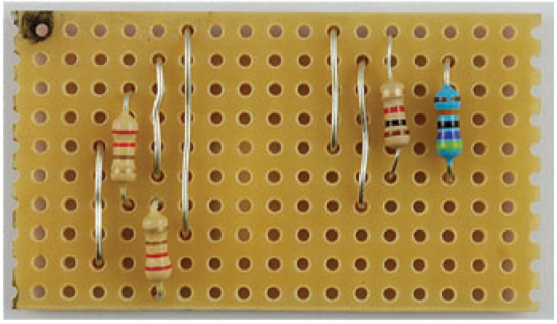

FIGURE 4-28 The stripboard with resistors in place

Troubleshooting

If it does not work, immediately unplug it and go back though and check everything, especially that the LEDs, IC, and capacitor are the correct way around. Also check that the batteries are okay.

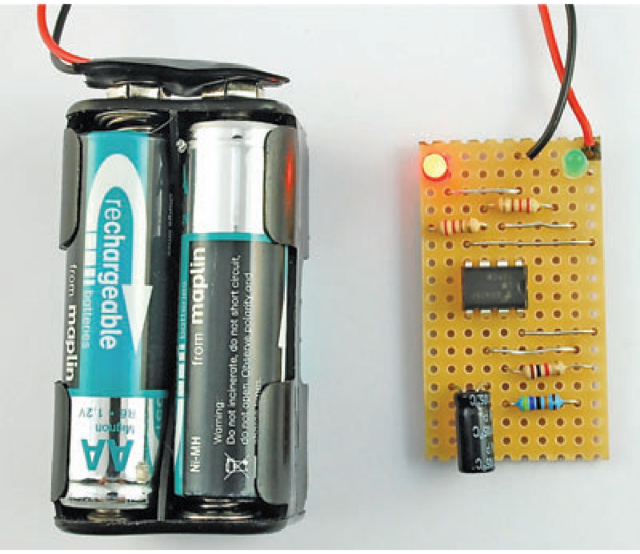

FIGURE 4-29 The LED flasher on the stripboard

How to Use a Laser Diode Module

Lasers are best bought as laser modules. The difference between a laser module and a laser diode is that the module includes a laser diode as well as a lens to focus the beam of laser light and a drive circuit to control the current to the laser diode.

If you buy a laser diode, you will have to do all this yourself.

A laser diode module, such as the 1-mW module shown in Figure 4-30, comes with a datasheet that says it needs to be supplied with 3V. So, all you need to do is find it a 3V battery and connect it up.

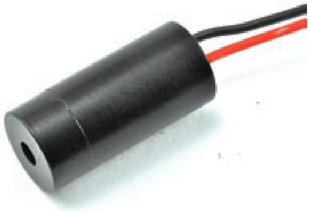

FIGURE 4-30 A laser diode module

Hacking a Slot Car Racer

Slot cars are a lot of fun, but could be improved by adding headlights and working brake lights to the car (Figure 4-31).

LEDs are just the right size to be fitted front and back into a slot car.

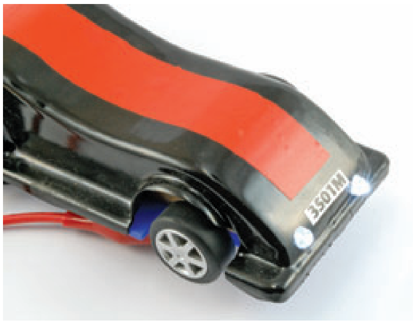

FIGURE 4-31 A modified slot car racer

You Will Need

You will need the following items to add lights to your slot car.

Quantity	Name	Item	Appendix Code
1		Slot car racer for modification	
1	D1	1N4001	S5, K1
2	D2, D3	High-brightness white LED 5rmm	S2
2	D4, D5	Red LED 5mm	S11
4	R1–4	1kΩ resistor	K2
1	C1	1000µF 16V-capacitor	C1
		Red, yellow, and black hookup wire	T7, T8, T9
1		* Two-way header plug and socket	

* I used a scavenged two-way header socket and plug to make it easier to work on the two halves of the car. This is not essential.

The slot car used here was part of a build-your-own slot car that has plenty of room inside for the electronics. Plan ahead and make sure you can fit everything in.

Storing Charge in a Capacitor

To make the brake lights stay on for a few moments after the car has stopped, you will need a capacitor to store charge.

If you think back to the idea of electricity as water flowing in a river, then a capacitor is a bit like a storage tank. Figure 4-32 shows how a capacitor can be used to store charge.

Figure 4-32a shows a tank (c1) being filled with water through A. Throughout this, water will also flow along the top and drive a water wheel, turning electrical energy into motion, a bit like how a light bulb or LED turns electrical energy into light. The water falls out of the bottom, returning to ground. Imagine a pump (like a battery) pulling the water back up for another circuit. If the water stops flowing into C1 through A, then C1 will still be full of water that will keep the water wheel turning until the water level in C1 drops below that of the outlet of the water well.

Figure 4-32b shows the electronic equivalent of this circuit. While the voltage at A is higher than GND, C1 will fill with charge and the light will be lit.

When the voltage at A is disconnected, the capacitor will discharge through the light bulb, lighting it. As the level of voltage drops in the capacitor, the bulb will gradually dim until it goes out as it reaches GND.

Figure 4-32 A capacitor as a tank

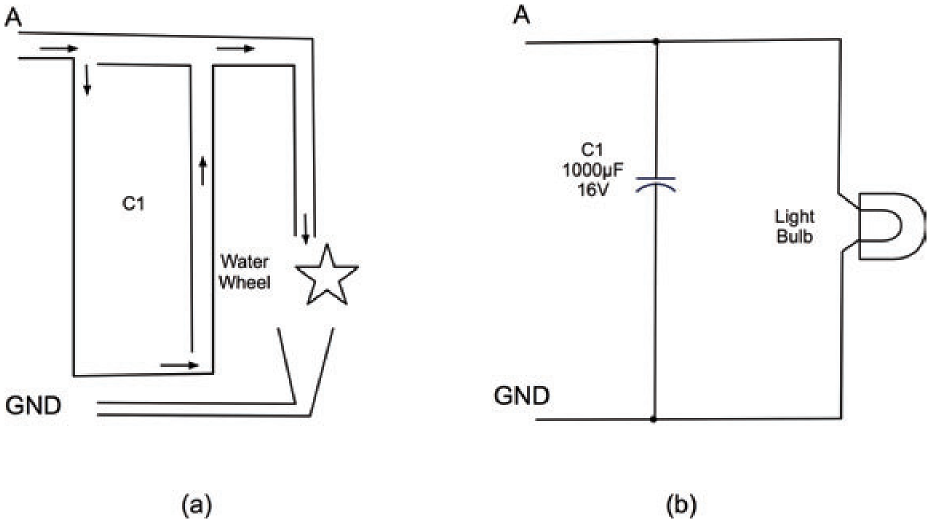

On the face of it, you can think of capacitors as being a bit like batteries. Both store charge. However, there are some very important differences.

- Capacitors only store a tiny fraction of the charge that a battery of the same size can store.
- Batteries use a chemical reaction to store electrical energy. This means their voltage remains relatively constant until they are spent, at which time it falls off rapidly. Capacitors, however, drop evenly in voltage as they discharge, just like the level of water decreasing in a tank.

Design

Figure 4-33 shows the schematic diagram for this modification.

The headlights (D2 and D3) are powered all the time from the slot car's connection to the track, so whenever the motor is running the LEDs will light.

The brake lights are more interesting. These will automatically come on when the car stops, and then go off after a few seconds. To do this, we make use of a capacitor C1.

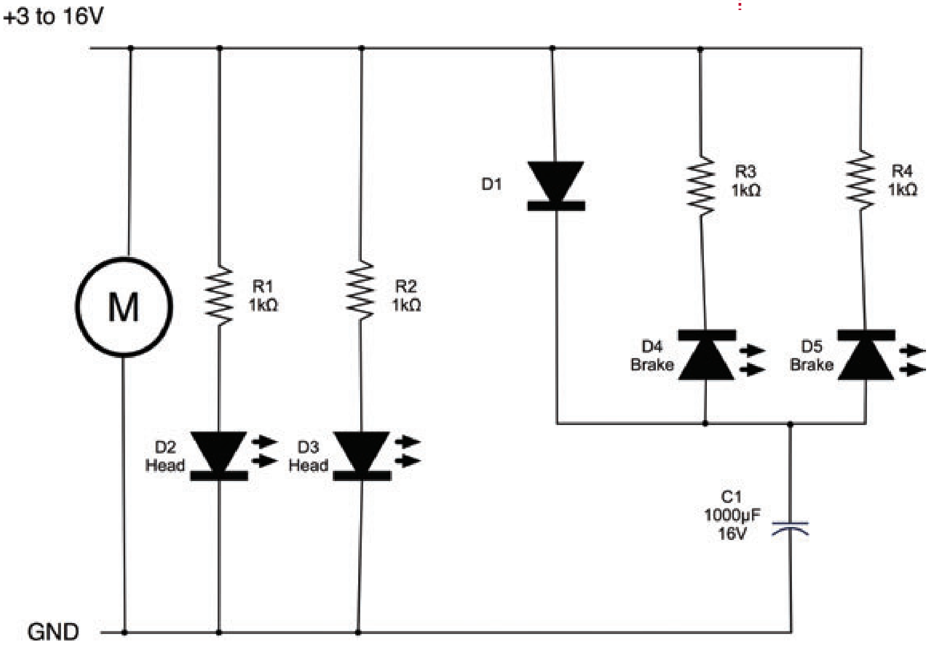

FIGURE 4-33 A schematic diagram for the slot car modification

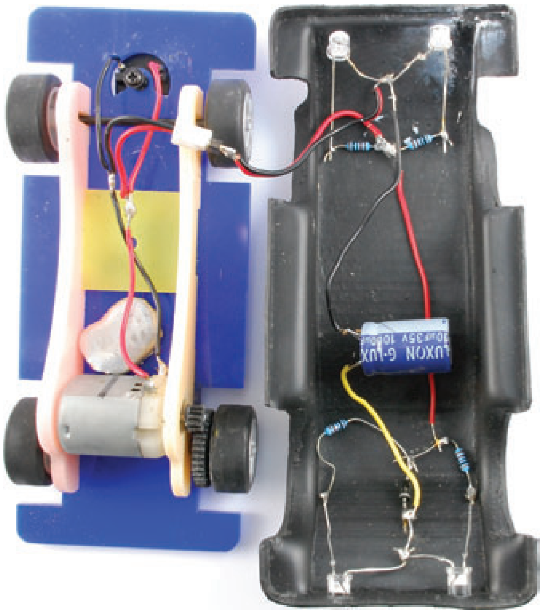

FIGURE 4-34 The components inside the car

When the car is powered, C1 will be charged through D1. At this point, the brake lights D4 and D5 will not be lit because they will be reverse-biased—that is, the voltage going in from the car tracks will be higher than the voltage at the top of the capacitor.

When you release the trigger on the controller, there will be no voltage coming in. Now the voltage at the top of the capacitor will be higher than the voltage coming in, so the capacitor will discharge through D4 and D5, making them light.

Construction

Figure 4-34 shows how the components are laid out in the two halves of the car.

How you lay these out in your vehicle may vary depending on how much space you have.

Holes were drilled in the case to take the 5mm LEDs. The LEDs are a snug fit in the holes and stay in place without any glue.

Figure 4-35 shows a wiring diagram for the arrangement that makes it easier to see what is going on.

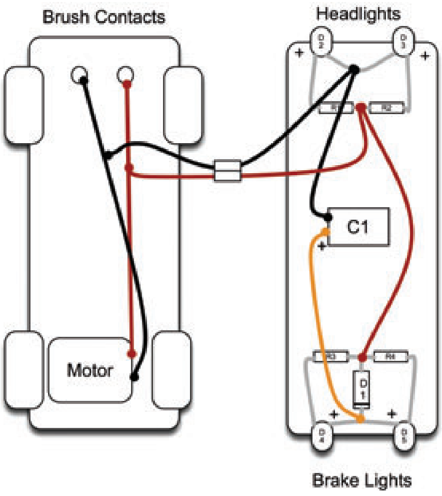

FIGURE 4-35 The wiring diagram for the modified car

Use your multimeter of the 20V range to identify which is the positive power connector on the contacts at the front of the car. This contact is connected to the red lead.

The longer leads of the LEDs are the positive connections, and the capacitor's negative lead should be marked with a "-".

The optional connector makes it easier to work on the two halves of the car separately.

Testing

Testing really just involves trying out the car on the track. If the headlight LEDs are not on as soon as you squeeze the trigger on the controller, check the wiring, paying special attention to the polarity of the LEDs.

Summary

We have learned how to use LEDs in this chapter, as well as picked up some good building skills so we can make our creations a bit more permanent using stripboard.

In the next chapter, we will examine sources of power, including batteries, power supplies, and solar panels. We will also look at how to select the right kind of battery, repurpose old rechargeable batteries, and use them in our projects.

5

Batteries and Power

Everything that you make or adapt is going to need to get its power from somewhere. This might be from a household electricity adapter, solar panels, rechargeable batteries of various sorts, or just standard AA batteries.

In this chapter, you will find out all about batteries and power, starting with batteries.

Note I use the word "battery" to describe both batteries and cells. Strictly speaking, a battery is a collection of cells wired one after the other to give the desired voltage.

Selecting the Right Battery

There are many types of battery on the market. So, to simplify things, in this chapter we will just look at the most common types of battery, those that are readily available and that will be used in most of the devices in this book.

Battery Capacity

Whether single-use or rechargeable, batteries have a capacity—that is, they hold a certain amount of electricity. Manufacturers of single-use batteries often don't specify this capacity in the batteries you buy from a supermarket. They just label them heavy duty / light duty, and so on. This is a little like having to buy milk and being given the choice of "big bottle" or "small bottle" without being able to see how big the bottle is or be told how many pints or liters it contains. One can speculate as to the reasons for this. One reason might be that battery producers think the public isn't intelligent enough to understand a stated battery capacity. Another might be that the longer a battery is on the shelf, the more its capacity shrinks. Still another is that the capacity actually varies a lot with the current drawn from the battery.

Anyway, if a battery manufacturer is kind enough to tell you what you are buying, the capacity figure will be stated in Ah or mAh. So a battery that claims to have a capacity of 3000mAh (typical of a single-use alkaline AA cell) can supply 3000mA for one hour. Or, alternatively, 3A for an hour. But it doesn't have to draw 3A. If your project only uses 30mA, you can expect the battery to last 100 hours (3000/30). In truth, the relationship is not quite that simple, because as you draw more current, the capacity decreases. Nevertheless, this will do as a rule of thumb.

Maximum Discharge Rate

You cannot take a tiny battery like a CR2032 with a capacity of 200mAh and expect to power a big electric motor at 20A for 1/100 of an hour (six minutes). There are two reasons for this. First, all batteries actually have an internal resistance. So, it is as if there is a resistor connected to one of the terminals. This varies depending on the current being drawn from the battery, but may be as high as a few tens of ohms. This will naturally limit the current.

Second, when a battery is discharged too quickly, by too high a current, it gets hot—sometimes very hot, sometimes "on fire" hot. This will damage the battery.

Batteries therefore also have a safe discharge rate, which is the maximum current you can safely draw from it.

Single-Use Batteries

Although somewhat wasteful, sometimes it makes sense to use single-use batteries that cannot be recharged. You should consider single-use batteries if:

- The project uses very little power, so they will last a long time anyway.
- The project will never be close to someplace where it can be charged up.

Table 5-1 shows some common single-use batteries. These figures are typical values and will vary a lot between actual devices.

Especially when it comes to the maximum discharge rate, you may get away with a lot more, or the battery may fail or get very hot with considerably less. It will also depend on how well ventilated a box they are in, as heating under high currents is a big problem.

So in the spirit of hacking electronics, spend less time planning and more time trying. See how hot it gets and how long it lasts. After all, we are having fun here, not designing a product.

Type	Typical Capacity	Voltage	Max. Discharge Current	Features	Common Uses
Lithium button cell (e.g., CR2032)	200mAh	3V	4mA with pulses up to 12mA	High temperature range (–30 to 80ºC); small	Low-power devices; RF remote controls; LED key ring lights, etc.
Alkaline PP3 battery	500mAh	9V	800mA	Low cost; readily available	Small portable electronic devices; smoke alarms; guitar pedals
Lithium PP3	1200mAh	9V	400mA pulses up to 800mA	Expensive; light; high-capacity	Radio receivers
AAA cell	800mAh	1.5V	1.5A continuous	Low-cost; readily available	Small motorized toys; remote controls
AA cell	3000mAh	1.5V	2A continuous	Low cost; readily available	Motorized toys
C cell	6000mAh	1.5V	Probably get away with 4A	High-capacity	Motorized toys; high-powered flashlights
D cell	15,000mAh	1.5V	Probably get away with 6A	High-capacity	Motorized toys; high-powered flashlights

TABLE 5-1 Single-Use Battery Types

Note Some of the photographs in the table are of branded batteries. The figures shown are for batteries of that type, not specifically the batteries listed.

Roll Your Own Battery

A single-cell battery with a voltage of just 1.5V is probably not going to be of any use. You will normally need to put a number of these cells in series (end to end) to produce a battery of the desired voltage.

When you do this, you do not increase the capacity. If each cell was 2000mAh, then if you put four 1.5V batteries in series, the capacity would still be 2000mAh, but at 6V rather than 1.5V.

Battery holders such as the one shown in Figure 5-1 are a great way of doing this. Look closely at how the battery holder is constructed and you'll notice how the positive of one battery is connected to the negative of the next, and so on.

This holder is designed to take six AA batteries so as to produce an overall voltage of 9V. Battery holders like this are available to take two, four, six, eight, or ten cells, both in AA and AAA.

Another advantage of using a battery holder is that you can use rechargeable batteries instead of single-use batteries. However, rechargeable cells normally have a lower voltage, so you have to take this into account when calculating the overall voltage of your battery pack.

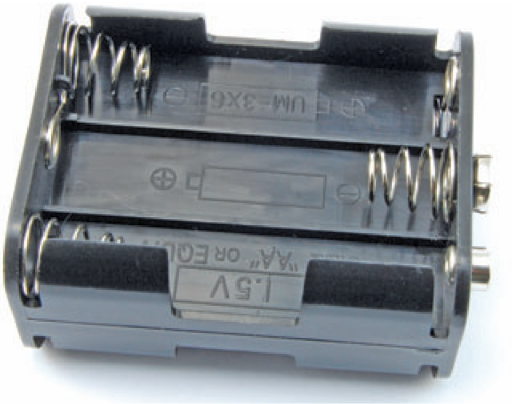

FIGURE 5-1 A battery holder

Selecting a Battery

Table 5-2 should help you decide on a suitable battery for your project. There is not always a best answer to the question "Which battery should I use?" and this table is definitely in the territory of rules of thumb.

You should also do the math and include how frequently the battery will need replacing.

	Voltage			
Power	**3V**	**6V**	**9V**	**12V**
Less than 4mA (short bursts) or 12mA continuous	Lithium button cell (e.g., CR2032)	2 × Lithium button cell (e.g., CR2032)	PP3	Unlikely
Less than 3A (short bursts) or 1.5A continuous	2 × AAA battery pack	4 × AAA battery pack	6 × AAA battery pack	8 × AAA battery pack
Less than 5A (short bursts) or 2A continuous	2 × AAA battery pack	4 × AAA battery pack	6 × AAA battery pack	8 × AAA battery pack
Even more	2 × C or D battery pack	4 × C or D battery pack	6 × C or D battery pack	8 × C or D battery pack

TABLE 5-2 Selecting a Single-Use Battery

Rechargeable Batteries

Rechargeable batteries can provide both cost and green benefits over single-use batteries. They are available in different types and in different capacities. Some, such as rechargeable AA or

AAA batteries, are designed as replacements for single-use batteries, and you remove them to charge in a separate charger. Other batteries are intended to be built into your project so all you have to do is plug a power adapter into your project to charge the batteries without removing them. The advent of cheap, high-capacity, low-weight lithium polymer (LiPo) batteries has made this a common approach for many items of consumer electronics.

Table 5-3 shows some commonly used types of rechargeable batteries.

Although there are many more types than this, these are the most commonly used batteries. Each type of battery has its own needs when it comes to charging, and we will look at each in later sections.

Type		Typical Capacity	Voltage	Features	Common Uses
NiMH button cell pack		80mAh	2.4 or 3.6V	Small	Battery backup
NiMH AAA cell		750mAh	1.25V	Low cost	Replacement for single-use AAA cell
NiMH AA cell		2000mAh	1.25V	Low cost	Replacement for single-use AA cell
NiMH C cell		4000mAh	1.25V	High capacity	Replacement for single-use C cell
Small LiPo cell		50mAh	3.7V	Low cost; high capacity for weight and size	Micro-helicopters
LC18650 LiPo cell		2200mAh	3.7V	Low cost; high capacity for weight and size; slightly bigger than AA	High-power flashlights; Tesla Roadster (yes, really—about 6800 of them)
LiPo pack		900mAh	7.4V	Low cost; high capacity for weight and size	Cell phones, iPods, etc.
Sealed lead–acid battery		1200mAh	6/12V	Easy to charge and use; heavy	Intruder alarms; small electric vehicles / wheelchairs

TABLE 5-3 Rechargeable Batteries

	NiMH	LiPo	Lead–Acid
Cost per mAh	Medium	Medium	Low
Weight per mAh	Medium	Low	High
Self-discharge	High (flat in 2–3 months)	Low (6% per month)	Low (4% per month)
Handling of full charge/ discharge cycles	Good	Good	Good
Handling of shallow discharge/charge	Medium (regular full discharge prolongs battery life)	Medium (not well-suited to trickle charging)	Good

TABLE 5-4 Characteristics of Different Battery Technologies

Table 5-4 summarizes the features of NiMH, LiPo, and lead–acid battery technologies.

If you want your project to charge a battery in place, then a LiPo or sealed lead–acid battery is probably the best choice. However, if you want the option to remove the battery and/or use single-use batteries, then a AA battery pack is a good compromise between capacity and size.

For ultra-high-power projects, lead–acid batteries, despite being an ancient technology, still perform pretty well, just as long as you don't have to carry them around! They are also easy to charge and are the most robust of the technologies, offering the least chance of fire or explosion.

Charging Batteries (in General)

Certain principals apply no matter what kind of battery you are charging. So read this section before reading those that follow it concerning specific battery types.

C

The letter C is used to denote the capacity of a battery in Ah or mAh. So, when people talk about charging a battery, they often talk about charging at 0.1C or C/10. Charging a battery at 0.1C means charging it at 1/10 of its capacity per hour. For example, if a battery has a capacity of 2000mAh, then charging it at 0.1C means charging it with a constant current of 200mA.

Over-Charging

Most batteries do not respond well to being over-charged. If you keep supplying them with a high charging current, you will damage them. They often also get hot. In the case of LiPo batteries, this can be "hot" in a fiery sort of way.

For this reason, chargers often charge at a low rate (called trickle charging), so that the low current will not damage the battery. Clearly this makes charging slow. Or, they will use a timer or other circuitry to detect when a battery is full and either stop charging altogether, or switch to trickle charging, which keeps the battery topped up until you are ready to use it.

With some kinds of battery, notably LiPos and the lead–acid variety, if you charge the battery with a constant voltage, then as the battery becomes charged, its voltage rises to match the charging voltage and the current naturally levels off.

Many LiPo batteries now come with a little built-in chip that prevents over-charging automatically. Always look for batteries with such protection.

Over-Discharging

You are probably starting to get the impression that rechargeable batteries are fussy. If so, you're right. Most types of battery are equally unhappy if you over-discharge them and let them go completely flat.

Battery Life

Anyone with a laptop more than a few years old will notice that the capacity of the battery gradually decreases until the laptop only works when plugged in, since the battery has become completely useless. Rechargeable batteries (whatever the technology) can only be recharged a few hundred (perhaps 500) times before needing to be replaced.

Many manufacturers of consumer electronics now build the battery into the device in such a way that it is not "user serviceable," with the rationale that the life of the battery is probably longer than the attention span of the consumer.

How to Charge a NiMH Battery

If you are going to remove your batteries to charge them, this section is pretty trivial. You take them out and put them in a commercial NiMH battery charger that will charge them until they are full and then stop. You can then put them back into your project and you are done.

If, on the other hand, you want to leave the batteries in place while you charge them, then you need to understand a little more about the best way to charge your NiMH batteries.

Simple Charging

The easiest way to charge a NiMH battery pack is to trickle charge it, limiting the current with a resistor. Figure 5-2 shows the schematic for charging a battery pack of four NiMH batteries using a 12V DC adaptor like the one we used back in Chapter 1 to make our fume extractor.

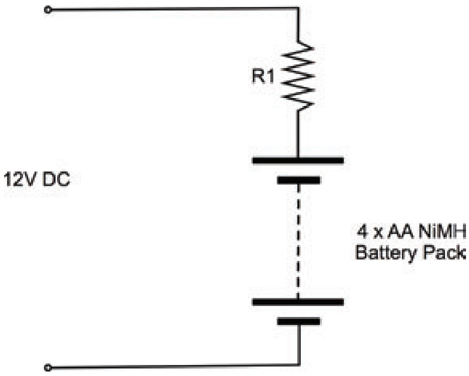

Figure 5-2 Schematic for trickle charging a NiMH battery pack

To calculate the value of R1, we first have to decide what current we want to charge our battery with. Generally, a NiMH battery can be safely trickle charged with less than 0.1C indefinitely. If the AA batteries we have each hold a C of 2000mAh, then we can charge them at up to 200mA. To be on the safe side, and if we planned to allow the batteries to "trickle" charge most of the time—for, say, a battery backup project—I would probably use a lower current of 0.05C or more conveniently C/20, which is 100mA.

Typically, the charge time for NiMH batteries is about 3C times the charging current, so at 100mA, we could expect our batteries to take 3 × 2000mAh / 100mA = 60 hours.

Back to calculating R1. When the batteries are discharged, each will be at a voltage of about 1.0V, so the voltage across the resistor will be 12V – 4V = 8V.

Using Ohm's law, R = V / I = 8V / 0.1A = 80Ω.

Let's be conservative and choose the convenient resistor value of 100Ω. Feeding this back in, the actual current will be I = V / R = 8V / 100Ω = 80mA.

When the batteries are fully charged, their voltage will rise to about 1.3V so the current will reduce to: I = V / R = (12V – 1.3V × 4) / 100Ω = 68mA.

That all sounds just fine, our 100Ω will be great. Now we just need to find out what maximum power rating we need for R1.

$$P = I\,V = 0.08A \times 8 = 0.64W = 640\text{ mW}$$

So, we should probably use a 1-W resistor.

Fast Charging

If you want to charge the batteries faster than that, then it is probably best to use a commercial charger, which will monitor the batteries and turn itself off or reduce the charge to a trickle when the batteries are full.

How to Charge a Sealed Lead–Acid Battery

These batteries are the least delicate of the battery types and could easily be trickle charged using the same approach as for NiMH batteries.

Charging with a Variable Power Supply

However, if you want to charge them faster, then it is best to charge them with a fixed voltage, with some current limiting (a resistor again). For a 12V battery (halve this for a 6V battery) until a discharged battery gets to around 14.4V, you can charge it with almost as much current as your power supply can take. It's only when it gets to this voltage that you need to slow down the charging to a trickle to prevent the battery from getting hot.

The reason we need to limit the current when the battery first starts to charge is that even if the battery doesn't get hot, the wires to it might get hot and whatever is supplying the voltage will only be able to supply a certain amount of current.

Figure 5-3 shows an adjustable power supply. Once you get into electronics, this is one of the first pieces of test equipment you should buy. You can use it in place of batteries while you are working on a project, and also use it to charge up pretty much any type of rechargeable battery.

A variable power supply lets you set both an output voltage and a maximum current. So the power supply will try and supply the specified voltage until the current limit is reached, at which point, the voltage will drop until the current falls back below the set current.

Figure 5-3a shows the power supply set to 14.4V and we have attached the power leads to an empty 12V 1.3Ah sealed lead–acid battery. We will start by adjusting the current setting of the power

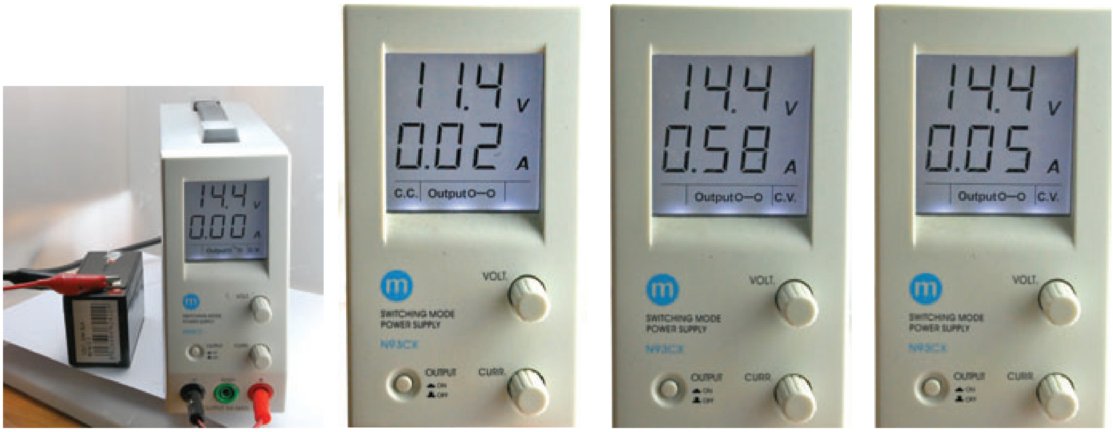
(a) (b) (c) (d)

Figure 5-3 Using a variable power supply to charge a lead–acid battery

supply to minimum, so as to prevent any nasty surprises. The voltage immediately drops to 11.4V (Figure 5-3b), so we can gradually increase the maximum current. In actual fact, even with no current limiting (turning the current knob to maximum), the current only rose to 580mA and the voltage increases to 14.4V (Figure 5-3c). After about two hours, the current has dropped to just 200mA, indicating that our battery is getting full. Finally, after four hours, the current is just 50mA and the battery is now fully charged (Figure 5-3d).

How to Charge a LiPo Battery

The technique we have just used on a lead–acid battery using a variable power supply will work just as well on a LiPo battery if we adjust the voltage and current accordingly.

For a LiPo cell, the voltage should be set to 4.2V and the current limited (usually to 0.5A) for a smallish cell, but currents up to C are sometimes used in radio-controlled vehicles.

However, unlike lead–acid and NiMH batteries, you cannot put a number of cells in series and charge the whole lot as one battery. Instead, you have to charge them separately, or use a "balanced charger" that monitors the voltage at each cell separately and controls the power to each.

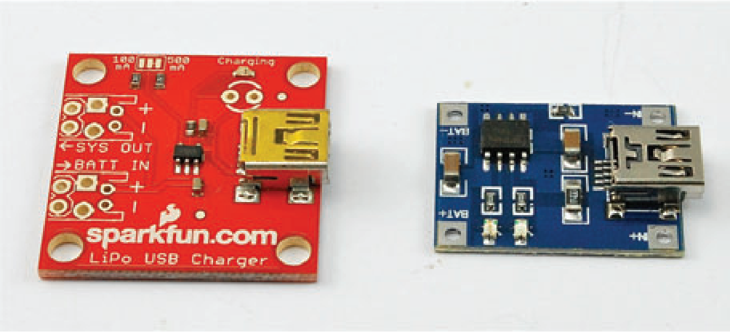

FIGURE 5-4 SparkFun and generic LiPo chargers

The safest and most reliable way to charge a LiPo is to use one of the chips that exist just for that purpose. These chips are cheap, but generally only available as surface-mounted components. However, there are plenty of ready-made modules available, many of which use the MCP73831 IC. Figure 5-4 shows two of these—one from SparkFun (see the Appendix, M16) and one for just a few dollars from eBay.

Both are used in the same manner. They will charge a single LiPo cell (3.7V) from a USB input of 5V. The SparkFun board has space on the PCB for two other connectors, one to which the battery is connected and the other for a second connection to the battery—the intention is that you connect the electronics that will use the battery to the second socket. The sockets can be either JST connectors as are often found on the end of the leads of a LiPo batter, or just screw terminals. The SparkFun module allows you to select the charging current, using a connection pad.

The generic module on the right has a fixed charge rate of 500mA and just a single pair of connections for the battery.

It is not a good idea to trickle charge a LiPo. If you want to keep them topped up, for say a battery backup solution, then leave them attached to the charger.

Hacking a Cell Phone Battery

Most of us have a cell phone or two languishing in a drawer somewhere, and one of the useful components that can usually be scavenged (assuming it's not the reason the phone is in the drawer) is the battery. The power supply is another useful item.

Figure 5-5a shows a fairly typical vintage cell phone battery. The battery is 3.7V (a single cell) and is 1600mAh (pretty good).

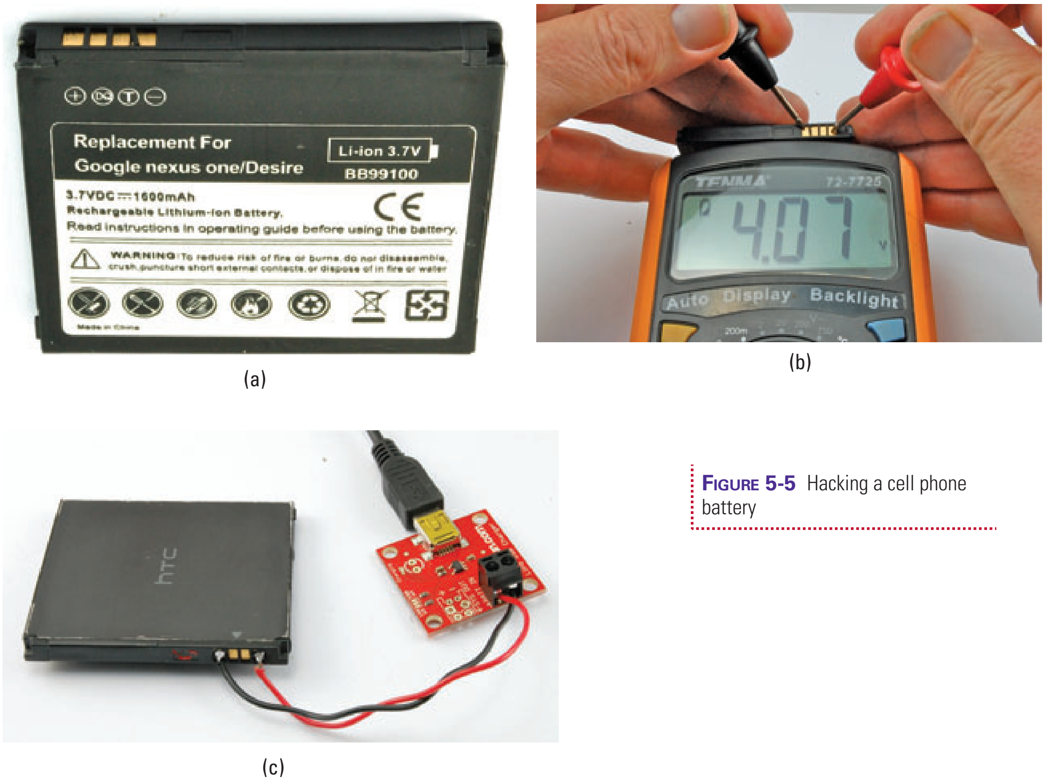
Replacement For
Google nexus one/Desire
Li-ion 3.7V
BB99100
3.7VDC 1600mAh
Rechargeable Lithium-ion Battery.
Read instructions in operating guide before using the battery.
CE
WARNING
Made in China
TENMA
72-7725
4.07
V
Auto Display Backlight
htc

(a)

(b)

(c)

FIGURE 5-5 Hacking a cell phone battery

Cell phone batteries normally have more than just the usual two connections for positive and negative. So the first task must be to identify the connections on the battery.

To identify the positive and negative connections to the cell, just put your multimeter into the 20V DC range and test each combination of pairs until you get the meter to read something over 3.5V, depending on how well charged the battery is (Figure 5-5b).

The batteries often have gold-plated contacts that make them very easy to solder leads to. Once they have leads attached, you can use a charger like the one described in the previous section. Figure 5-5c shows the SparkFun charger module being used for just that purpose.

Caution When using a LiPo battery, remember that if you discharge them too far (below about 3V per cell), you can permanently damage them. Most new LiPo batteries will include an automatic cut-off circuit, built into the battery package, to prevent over-discharging, but this may not be the case for a scavenged battery.

Controlling the Voltage from a Battery

The thing with batteries is that even though they may say 1.5V, 3.7V, or 9V on the package, their voltage will drop as they discharge—often by quite a high percentage.

For example, a 1.5V alkaline AA battery when brand new will be about 1.5V and will quickly fall to about 1.3V under load but still deliver useful amounts of power down to about 1V. This means that in a pack of four AA batteries, the voltage could be anything between 6V and 4V. Most types of battery, whether single-use or rechargeable, exhibit a similar voltage drop.

This may not matter much; it just depends on what the battery is powering. If it is powering a motor or an LED, then the motor will just go a bit more slowly, or the LED will be a little dimmer as the battery discharges. However, some ICs have a very narrow voltage tolerance. There are ICs designed to work at 3.3V that specify a maximum working voltage of 3.6V. Similarly, if the voltage drops too low, the device will also stop working.

In fact, many digital chips such as microcontrollers are designed to work at a standard voltage of 3.3V or 5V.

To ensure a steady voltage, we need to use something called a voltage regulator. Fortunately for us, voltage regulators come as convenient three-pin, low-cost chips that are very easy to use. In fact, the packages just look like transistors, and the bigger the package, the more current they can control.

Figure 5-6 shows how you would use the most common of voltage regulators, called the 7805.

Using just a voltage regulator IC and two capacitors, any input voltage between 7V and 25V can be regulated to a constant 5V. The capacitors provide little reservoirs of charge that keep the regulator IC operating in a stable manner.

Figure 5-6 A voltage regulator schematic

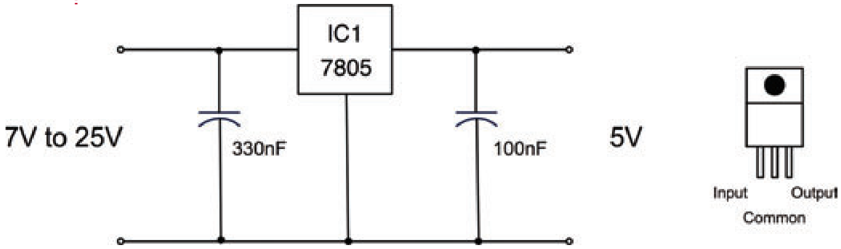

In the following experiment with a 7805, we will omit the capacitors, as the supply voltage is a steady 9V battery and the load on the output is just a resistor (Figure 5-7).

The capacitors become much more necessary when the load varies (in other words, in the amount of current that it draws), something that is true of most circuits.

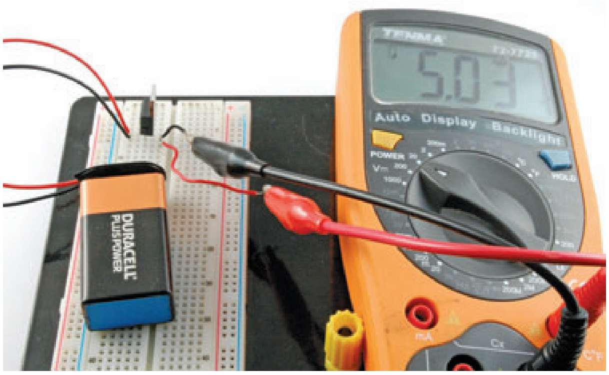

Figure 5-7 Experimenting with the 7805

You Will Need

Quantity	Names	Item	Appendix Code
1		Solderless breadboard	T5
1	IC1	7805 voltage regulator	K1, S4
1		Battery clip	H2
1		9V PP3 battery	

Wire up the breadboard as shown in Figure 5-8.

Figure 5-8 The 7805 breadboard layout

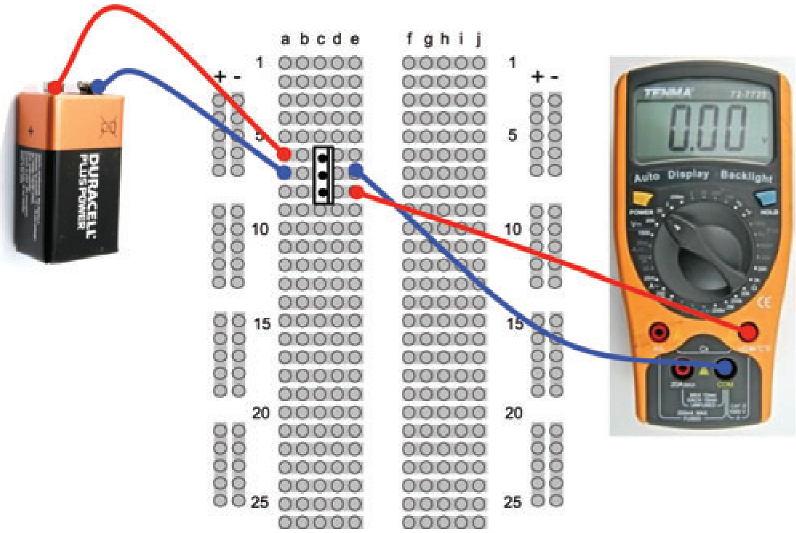

Breadboard

With the battery connected, the multimeter should display a voltage of close to 5V.

Although 5V is a very common voltage, there are voltage regulators for most common voltages, as well as the LM317 voltage regulator that we discussed in Chapter 4, that as well as providing constant current, can also be configured as a voltage regulator.

Table 5-5 lists some common voltage regulators that provide different output voltages and different current handling capabilities.

Output Voltage	100mA	1–2A
3.3V	78L33	LF33CV
5V	78L05	7805 (App A S4) (7–25V in)
9V	78L09	7809
12V	78L12	7812

TABLE 5-5 Voltage Regulators

Boosting Voltage

The voltage regulator ICs in the section titled “Controlling the Voltage from a Battery” only work if the input voltage is greater than the output voltage. In fact, it normally has to be a couple of volts higher, but some more expensive voltage regulators called LDO (low drop out) regulators are available that only require about half a volt more on the input than the output.

Sometimes, however—and cellular phones are a good example of this—it is very convenient to use a single-cell LiPo battery of 3.7V when we require a higher voltage (often 5V) for the circuit.

In these situations, you can employ a very useful circuit called a buck-boost converter. These use an IC and a small inductor (coil of wire) and by applying pulses to the inductor, produce a higher voltage. Actually, it’s more complex than that, but you get the idea.

Buck-boost converters are readily available as modules on well-known auction sites. You can find 1A adjustable modules that will provide an adjustable output of 5V to 25V from 3.7V for a few dollars. Try searching for “Boost Step-Up 3.7V.” The main module suppliers also provide such boards for around USD 5.

SparkFun sells an interesting module (see the Appendix, M17) that combines a LiPo battery charger with a buck-boost, so you can both charge your LiPo from an external USB 5V input and use the 3.7V LiPo cell to provide an output of 5V using the buck-boost (Figure 5-9).

This actually takes all the difficulties away from using a LiPo in a situation where you want to charge the LiPo battery in situ. Your 5V microcontroller circuit or whatever you are using is just attached to the VCC and GND connections, the battery is clipped into the socket, and to charge the device, you just plug in a USB cable.

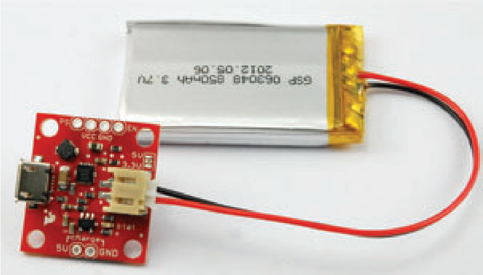

Figure 5-9 Combined LiPo charger and booster

Calculating How Long a Battery Will Last

We have already touched on the capacity of a battery—that is, the number of mAh it can supply. However, other factors come into play that we should think about when deciding if the batteries we are considering for a project are going to last long enough.

It's really just a matter of common sense, but nevertheless it's easy to make false assumptions about what you need.

As an example, I recently built an automated door for my chicken house. It opens at dawn and closes when it gets dark. It uses an electric motor, and electric motors use a lot of current, so I needed to decide what kind of batteries to use. My first thought was to use big D cells or a lead–acid battery. But when it came to do the math, I found this wasn't really necessary.

Although the motor uses 1A each time it is in operation, it is only in operation twice a day, and each time only for about three seconds. I measured the control circuit as using 1mA all the time. So, let's work out how many mAh the control circuit and motors each use in a day, and then see how many days various types of battery will last.

Let's start with the motors:

1A × 3 seconds × 2 = 6As = 6/3600Ah = 0.0016 Ah = 1.6mAh per day

On the other hand, the controller, which I had assumed was the low power part of the project would consume:

1mA × 24 hours = 24mAh per day

This means we can pretty much ignore the power consumed by the motor since it is less than a tenth of the juice required by the controller. Let's say the total requirement is 25mAh/day.

AA batteries are typically 3000mAh, so if we powered the project from AA batteries, we could expect them to last 3000mAh / 25mAh per day = 120 days.

So we do not really need to look much further, AAs will be fine. In the end, I used solar power for this project, which we will visit again in the section titled “How to Use Solar Cells” later in this chapter.

How to Design for Battery Backup

Replacing batteries is a nuisance, and expensive, so it is often cheaper and more convenient to power things from a wall-wart power supply. However, this brings its own disadvantages:

- The device is now tethered to a wire.
- If the household electricity fails, the device will stop working.

The best of both worlds can be achieved by arranging for automatic battery backup of a device that is powered by your household electricity. So, both batteries and a power supply are used, but the batteries are only used if the power supply is not available.

Diodes

What we do not want to happen is for both the batteries and the voltage from the power supply to conflict with each other when both are available. For instance, if the power supply is at a higher voltage than the batteries, it would charge them. But without anything to limit the current, this could be disastrous, even if the batteries were of the rechargeable sort.

FIGURE 5-10 Battery backup schematic

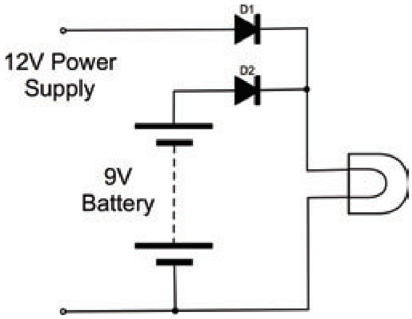

Figure 5-10 shows the basic schematic for this. The power supply always needs to be a higher voltage than the battery, so in this case it is 12V and the battery 9V. The schematic also assumes that the battery backup is being used to drive a light bulb.

Recall that diodes act rather like one-way valves. They only allow current to flow in the direction of the arrow. So, let’s look at the three possible cases of how power could be supplied here. This is simply the power supply: just batteries and both the battery and power supply (Figure 5-11).

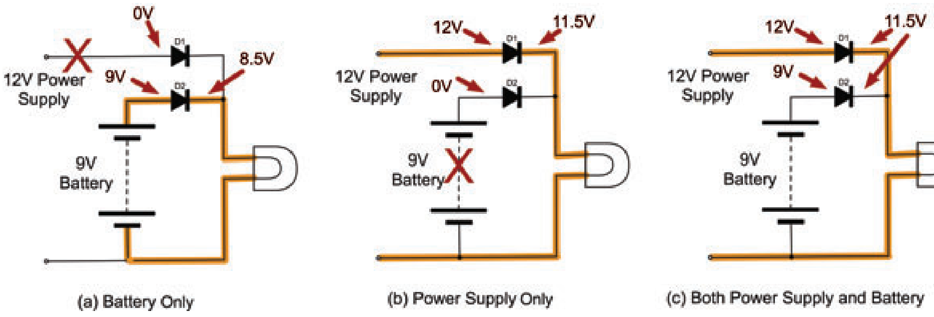
0V
D1
8.5V
12V Power Supply
9V
D2
9V Battery
(a) Battery Only
12V
11.5V
D1
12V Power Supply
0V
D2
9V Battery
(b) Power Supply Only
12V
11.5V
D1
12V Power Supply
9V
D2
9V Battery
(c) Both Power Supply and Battery

FIGURE 5-11 Diodes for battery backup

Just the Battery

If only the battery has a voltage greater than zero (in other words, the power supply is not plugged in), then the situation is as shown in Figure 5-11a. The 9V from the battery will be at the anode of D2, and the cathode of D2 will be pulled toward ground by the load of the light bulb. This will cause D2 to be forward-biased and conduct the current through the light bulb. A forward-biased diode will have an almost constant voltage of 0.5V across it, which is why we can say that the voltage after the diode is 8.5V.

On the other hand, D1 will have a higher voltage (8.5V) on its cathode (right-hand side in the diagram) than its anode (0V), so no current will flow through D1.

Just the Power Supply

If just the power supply is connected (Figure 5-11b), then the role of the diodes is reversed and now the current flows through D1 to the light bulb.

Both the Power Supply and the Battery

Figure 5-11c shows the situation where both the power supply and the battery are connected. The 12V of the power supply will ensure that the cathode of D2 is at 11.5V. Since the anode of D2 is at 9V from the battery, the diode will remain reverse-biased and no current will flow through it.

Trickle Charging

As we already have a battery and a power supply, we have most of the ingredients we need to charge the battery. We could for example use six AA rechargeable batteries in a battery box and arrange to charge them at C/20 (assuming C = 2000mAh) or 100mA from the power supply.

That way, the batteries would always be charged, and provide light whenever the power failed. Figure 5-12 shows the schematic for this.

You may not have been expecting the extra diode D3. This is really just to account for the fact that we do not know exactly how the power supply is designed, so we do not know what would happen if the battery was connected to its output (via R1) when it was turned off. This may discharge the battery or damage the power supply. The diode D3 just protects it and makes sure no current can flow back into it.

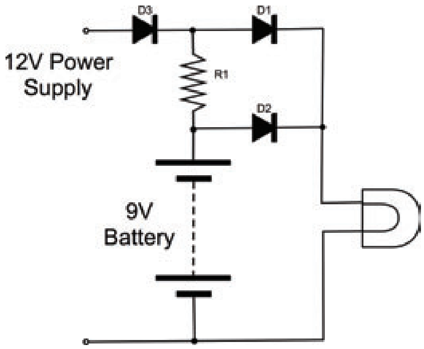

FIGURE 5-12 Battery backup and charging

We want a charging current of 100mA to flow through R1, and we know that when both the power supply and battery are connected, there will be a voltage across R1 of 12V – 0.5V – 9V or 2.5V. So, using Ohm's law, the value of the resistor should be:

$$R = V / I = 2.5 / 0.1A = 25\Omega$$

The nearest standard value to this is probably 27Ω.

Its power requirement: $P = V^2 / R = 2.5^2 / 27 = 0.23W$

This means a standard half- or quarter-watt resistor will be fine.

How to Use Solar Cells

On the face of it, solar cells seem like the perfect power source. They convert light into electricity, and so in theory you need never change a battery or be plugged into a wall outlet again!

However, as always, the reality is not quite so simple. Solar cells, unless they are very large, produce fairly small amounts of electricity and so are most suited to low-power devices and projects that are outdoors away from household electricity.

If you are thinking of trying a solar project that will be installed indoors, unless it will be sited against a south-facing window, I really wouldn't try it. Solar cells do not require direct sunlight, but to produce any useful amounts of electricity, they really need a good unobstructed view of the sky.

Two solar projects I have developed are a solar-powered radio (the solar panel is as big as the radio and, yes, it needs to be next to the window), and a solar-operated chicken house door. If you are lucky enough to live somewhere sunny, then solar power is obviously a lot easier.

Figure 5-13 shows a typical solar panel. This one was scavenged from a security light installation. It is about six inches by four inches and has a swivel mount that allows it to be angled toward the sun. It is the panel I used for the chicken house door.

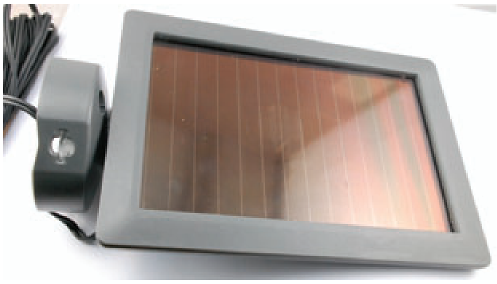

FIGURE 5-13 A solar panel

Projects that use a solar panel to provide power nearly always also use a rechargeable battery. So the panel charges the battery and the project draws its power from the battery.

Small solar cells generally only produce around half a volt, so they are usually combined into panels of many cells that increase the voltage to a level that is high enough to charge a battery.

The voltage you find on a solar panel normally refers to the voltage of battery that the solar panel is capable of charging. So, it is quite common to find 6V or 12V solar panels. When you measure the voltage from these in bright sunlight, the reading will be much higher, possibly 20V for a 12V panel. But, under the load of charging a battery, this drops rapidly.

Testing a Solar Panel

A solar panel will have a certain number of watts and a nominal voltage specified for it. These tend to be for ideal conditions, so when I get a solar panel that I want to use in a project, I like to test it to find what it is really capable of. Without knowing how much power it can provide in a real situation where it's installed, it is hard to make safe assumptions about battery capacities and how low you need to keep the current consumption.

When testing out a solar panel, you should use a resistor as a "dummy load," and then try out the solar panel in various locations and levels of brightness, measuring the voltage across the resistor. From this, you can calculate the current being provided by the panel.

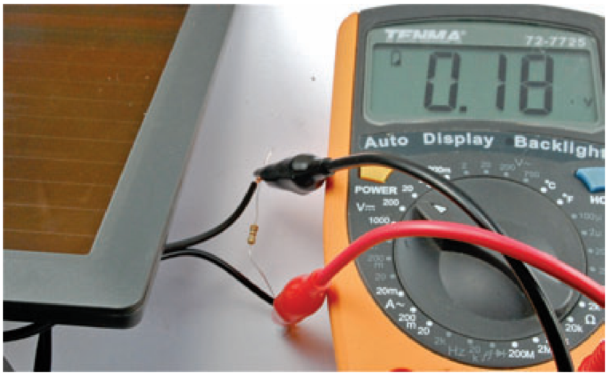

FIGURE 5-14 Testing a solar panel

Figure 5-14 shows such an arrangement for my "chicken house" solar panel. The meter is showing just 0.18V with a 100Ω load resistor inside the light box that I use for my photography. That equates to just 1.8 mA.

I find a spreadsheet a useful way of recording how the solar panel performs. Figure 5-15 shows an excerpt from the spreadsheet, complete with graph. You can then file this away until the next time you wish to use a solar panel in a project.

The spreadsheet can be downloaded from www.hackingelectronics.com, but there is really nothing complex about the math.

As you can see, the solar panel produces only 1 or 2mA indoors even under bright artificial lighting. The results outdoors with a clear view of the sky are better, but it really only produces quite high power in direct sunlight.

Trickle Charging with a Solar Panel

Since the solar panels produce a reasonable voltage, even in relatively low light conditions, they can easily be used to trickle charge a battery. However, you should always use a diode to protect the solar panel from the situation where the battery is at a higher voltage than the panel (say at night), since such a reverse flow will damage the solar panel.

A typical simple trickle charge schematic is shown in Figure 5-16.

FIGURE 5-15 Solar panel data

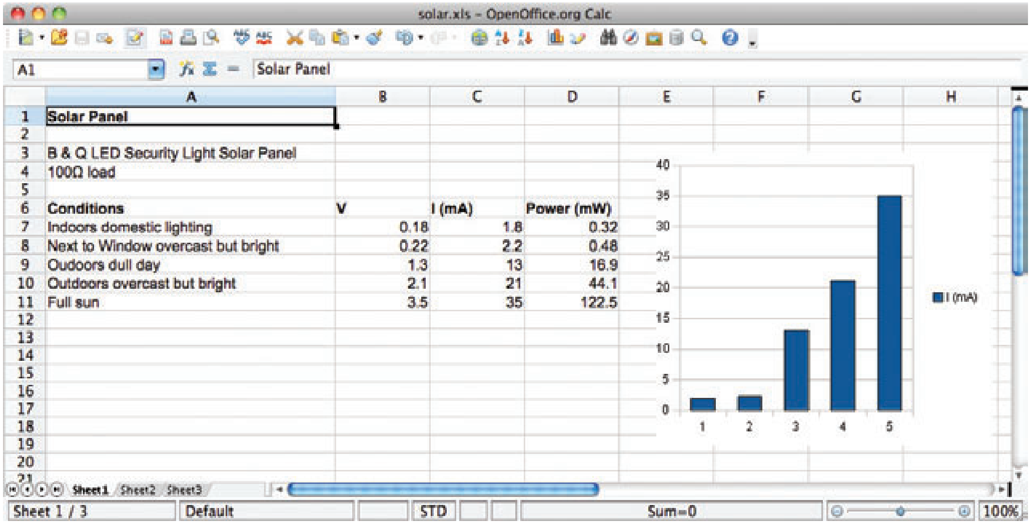

solar.xls - OpenOffice.org Calc

A1 Solar Panel

Solar Panel

B & Q LED Security Light Solar Panel

100Ω load

Conditions	V	I (mA)	Power (mW)
Indoors domestic lighting	0.18	1.8	0.32
Next to Window overcast but bright	0.22	2.2	0.48
Oudoors dull day	1.3	13	16.9
Outdoors overcast but bright	2.1	21	44.1
Full sun	3.5	35	122.5

Lead–acid batteries are still a very popular choice for trickle charging from solar. This is mainly because they are very forgiving of gentle over-charging and have a lower self-discharge rate than, say, NiMH batteries.

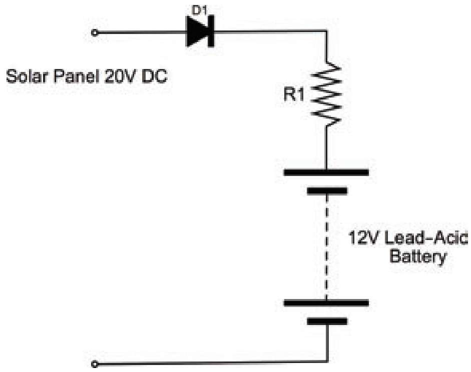

Figure 5-16 Schematic for solar trickle charging

Minimizing Power Consumption

When planning solar power for some small outdoor project, you need to make sure the solar panel charging the battery can keep up with demand.

If you live in southern California, the design for using solar panels is pretty easy. You can count on quite a lot of sun all year long. However, if you live a long way from the equator, say, in a maritime climate where it's often quite dull during the day, then you will have short winter days. You may get weeks of dull weather with short days. If your system is to work all year long, you either need to have a large battery that will last for a few weeks of dull weather, or use a larger solar panel.

The sums are pretty easy. There are mAh going into the battery from the solar panel, and mAh coming out for the device it is powering. The device might be running all the time, but the solar panel is only active half the time (daylight). So, you need to work out what you think the worst case for solar input might be for a week or two, and then design it accordingly.

It will probably be easier and cheaper to put your efforts into minimizing the current consumed by the system rather than increasing the size of the solar panel and battery.

Summary

In this chapter, we have learned about how to power our projects. In the next chapter, you will learn how to use the very popular Arduino microcontroller board.

6

Hacking Arduino

Microcontrollers are essentially low-powered computers on a chip. They have input/output pins to which you can attach electronics so the microcontroller can, well, control things. Utilizing a microcontroller used to be quite a complex process, largely because the microcontroller needed to be programmed. This was often done in assembler or complex C. But there was a lot to learn before you could do anything useful. Because of this, it discouraged their use in casual projects where you just wanted to hack something together.

Enter the Arduino (Figure 6-1). The Arduino is a simple-to-use, low-cost, readymade board that lets you use a microcontroller in your projects with a minimum of fuss.

The Arduino sells in vast quantities and has become the platform of choice for makers and hackers in need of microcontrollers.

The popularity of Arduino is due to many factors, including its:

- Low cost
- Open-source hardware design
- Easy-to-use integrated development environment (IDE) to program it with
- Plug-in shields that add features like displays and motor drivers that clip onto the top of the Arduino

All the programs for the Arduino used in this and later chapters are available for download from the book's accompanying web site (www.hackingelectronics.com).

The examples in this book are designed to work with both the Arduino Uno and the Arduino Leonardo. However, two of the projects, in the sections "How to Type Passwords Automatically" and "How to Make a USB Music Controller" (see Chapter 9), only work with Arduino Leonardo.

The Leonardo is the newer board. You may have some compatibility problems with this board and some Arduino shields. This includes any Ethernet Shield prior to the R3 Ethernet Shield. So, if you have an older Ethernet Shield, then the section "How to Control a Relay from a Web Page" will work on an Arduino Uno, but not a Leonardo, unless you have an R3 shield.

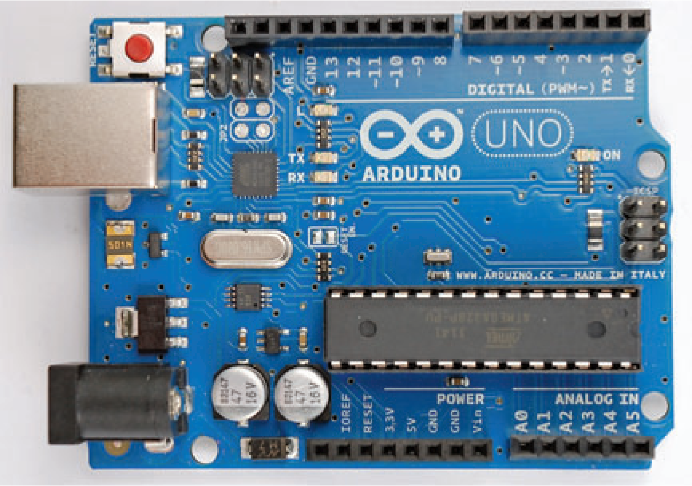

Figure 6-1 An Arduino Uno board

How to Set up Arduino (and Blink an LED)

To be able to program an Arduino, we first have to install the Arduino integrated development environment (IDE) on our computer. Arduino is available for Windows, Mac, and Linux.

You Will Need

Quantity	Item	Appendix Code
1	Arduino Uno/Leonardo	M2/M21
1	USB lead; Type B for Uno, Micro USB for Leonardo	

Setting Up Arduino

The first step is to download the software for your computer type from the official Arduino web site here: http://arduino.cc/en/Main/Software.

Once this is downloaded, you can find detailed instructions for the installation of each platform here: http://arduino.cc/en/Guide/HomePage.

One of the nice things about the Arduino is that to get started with it, all you need is an Arduino, a computer, and a USB lead to connect the two together. The Arduino can even be powered

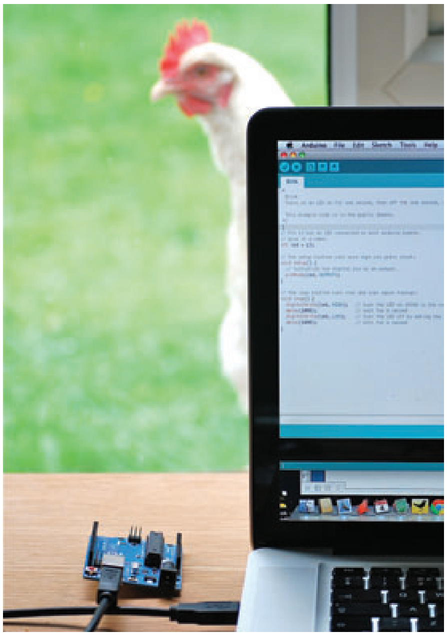

FIGURE 6-2 The Arduino, laptop, and chicken

over the USB connection to the computer. Figure 6-2 shows an Arduino Uno (the most common type of Arduino) connected to a laptop running the Arduino IDE.

To prove that the Arduino is working, we will program it to flash an LED that is on the Arduino board labeled "L" and hence known as the "L" LED.

Start by launching the Arduino IDE on your computer. Then, from the File menu (Figure 6-3), select Examples | 01.Basics | Blink.

In an attempt to make programming the Arduino sound less daunting to non-programmers, programs on the Arduino are referred to as "sketches." Before we can send the Blink sketch to your Arduino, we need to tell the Arduino IDE what type of Arduino we are using. The most common type is the Arduino Uno, and in this chapter we will assume that is what you have. So from the Tools | Board menu, select Arduino Uno (Figure 6-4).

As well as selecting the board type, we also need to select the port it is connected to. In Windows, this is easy because it is always COM3 or COM4 (see Figure 6-5). However, on a Mac or Linux, there will generally be more serial

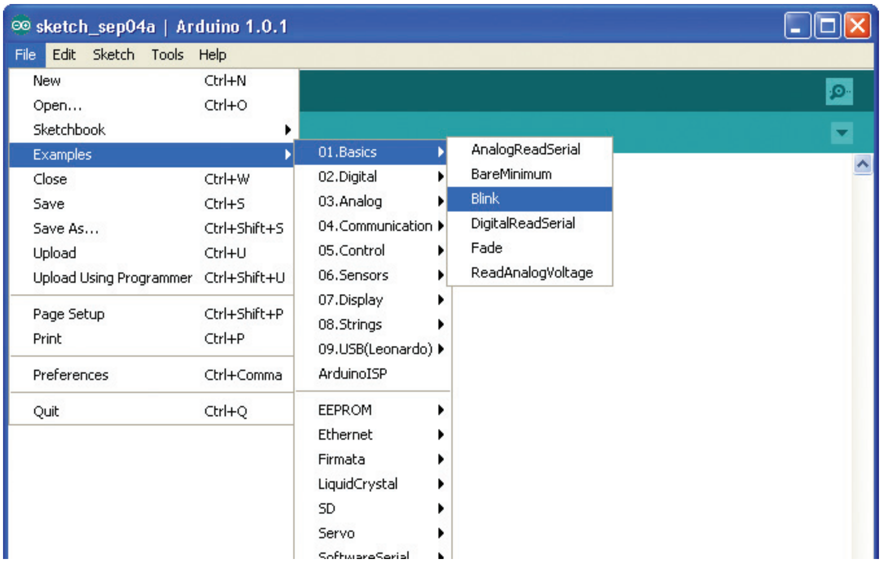
sketch_sep04a | Arduino 1.0.1
File
Edit
Sketch
Tools
Help
New Ctrl+N
Open... Ctrl+O
Sketchbook
Examples
Close Ctrl+W
Save Ctrl+S
Save As... Ctrl+Shift+S
Upload Ctrl+U
Upload Using Programmer Ctrl+Shift+U
Page Setup Ctrl+Shift+P
Print Ctrl+P
Preferences Ctrl+Comma
Quit Ctrl+Q
01.Basics
02.Digital
03.Analog
04.Communication
05.Control
06.Sensors
07.Display
08.Strings
09.USB(Leonardo)
ArduinoISP
EEPROM
Ethernet
Firmata
LiquidCrystal
SD
Servo
SoftwareSerial
AnalogReadSerial
BareMinimum
Blink
DigitalReadSerial
Fade
ReadAnalogVoltage

FIGURE 6-3 Loading the "Blink" sketch

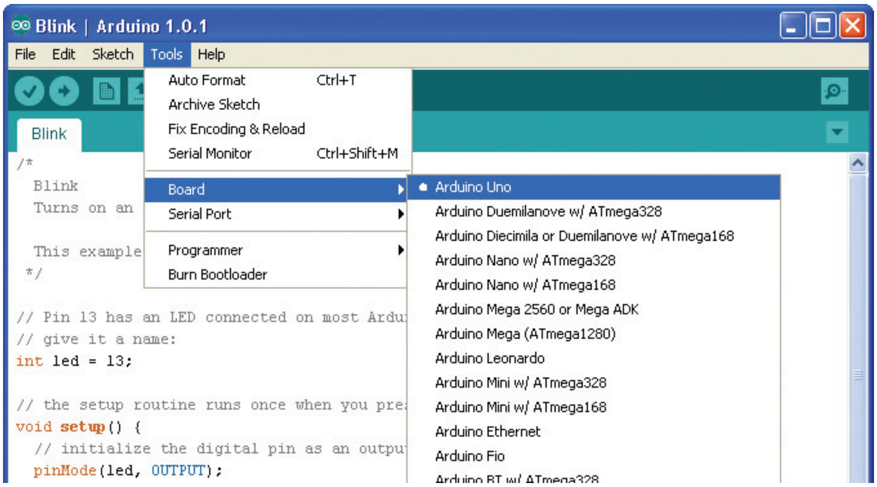
Blink | Arduino 1.0.1
File
Edit
Sketch
Tools
Help
Auto Format Ctrl+T
Archive Sketch
Fix Encoding & Reload
Serial Monitor Ctrl+Shift+M
Board
Serial Port
Programmer
Burn Bootloader
Blink
● Arduino Uno
Arduino Duemilanove w/ ATmega328
Arduino Diecimila or Duemilanove w/ ATmega168
Arduino Nano w/ ATmega328
Arduino Nano w/ ATmega168
Arduino Mega 2560 or Mega ADK
Arduino Mega (ATmega1280)
Arduino Leonardo
Arduino Mini w/ ATmega328
Arduino Mini w/ ATmega168
Arduino Ethernet
Arduino Fio
Arduino BT w/ ATmega328

```
/*
  Blink
  Turns on an

  This example
 */

// Pin 13 has an LED connected on most Ardu
// give it a name:
int led = 13;

// the setup routine runs once when you pre
void setup() {
  // initialize the digital pin as an outpu
  pinMode(led, OUTPUT);
```

FIGURE 6-4 Selecting the board type

devices. However, the Arduino IDE shows them with the most recently connected first, so your Arduino board should be at the top of the list.

To actually upload the sketch onto the Arduino board, click the Upload button on the tool bar. This is the second button on the toolbar, highlighted in Figure 6-6.

Once, you press the Upload button, a few things should happen. First, a progress bar will appear as the Arduino IDE first compiles the sketch (converts it into a suitable form for uploading). Then, the LEDs on the Arduino—labeled Rx and Tx—should flicker for a while.

Finally, the LED labeled L should start to blink. The Arduino IDE will also show a message that looks something like "Binary sketch size: 1,084 bytes (of a 32,256 byte maximum)." This means

FIGURE 6-5 Selecting the serial port

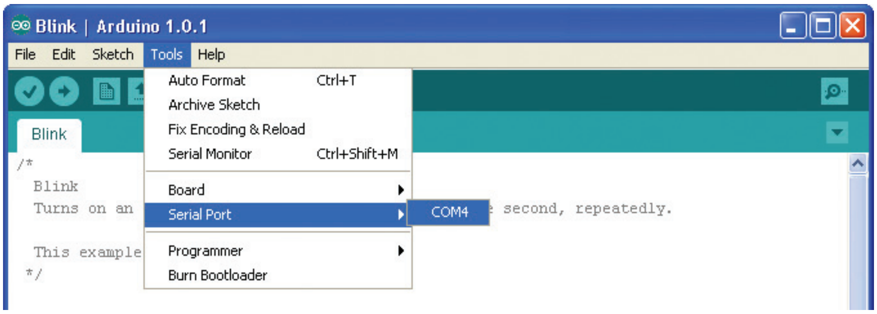
Blink | Arduino 1.0.1
File
Edit
Sketch
Tools
Help
Auto Format Ctrl+T
Archive Sketch
Fix Encoding & Reload
Serial Monitor Ctrl+Shift+M
Board
Serial Port
Programmer
Burn Bootloader
Blink
COM4

```
/*
  Blink
  Turns on an                               second, repeatedly.

  This example
 */
```

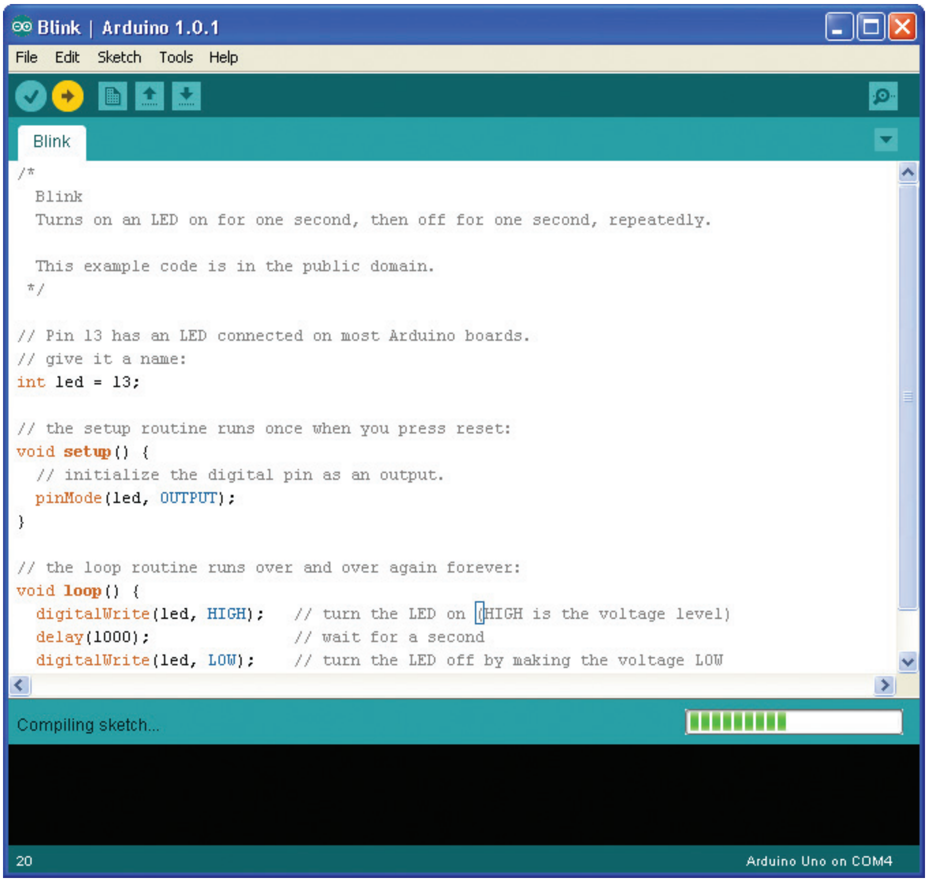

```
/*
  Blink
  Turns on an LED on for one second, then off for one second, repeatedly.

  This example code is in the public domain.
 */

// Pin 13 has an LED connected on most Arduino boards.
// give it a name:
int led = 13;

// the setup routine runs once when you press reset:
void setup() {
  // initialize the digital pin as an output.
  pinMode(led, OUTPUT);
}

// the loop routine runs over and over again forever:
void loop() {
  digitalWrite(led, HIGH);   // turn the LED on (HIGH is the voltage level)
  delay(1000);               // wait for a second
  digitalWrite(led, LOW);    // turn the LED off by making the voltage LOW
```

Compiling sketch...

20 Arduino Uno on COM4

Figure 6-6 Uploading the Blink sketch

the sketch has used about 1kb of the 32k of flash memory available for programs on the Arduino.

Note that if you are using a Leonardo, you may have to keep the Reset button depressed until you see the message “Uploading…” in the Arduino software.

Modifying the Blink Sketch

It may be that your Arduino was already blinking when you first plugged it in. That is because the Arduino is often shipped with the Blink sketch already installed.

If this is the case, you might like to prove to yourself that you have actually done something by changing the blink rate. We will now examine the Blink sketch and see how we could change it to make it blink faster.

The first part of the sketch is just a comment, to tell someone looking at the sketch what it is supposed to do. This is not actual program code, and part of the preparation for the code being uploaded is for all such "comments" to be stripped out. Anything between "/*" and "*/" is ignored.

```
/*
  Blink
  Turns on an LED on for one second, then off for one
second, repeatedly.

  This example code is in the public domain.
 */
```

We then have a couple of individual line comments that start with //. Just like the block comments earlier, these are simply to inform us of what is happening. In this case, it helpfully tells us to use pin 13 as the pin we will flash. We have chosen that pin because on an Arduino Uno that pin is connected to the built-in "L" LED.

```
// Pin 13 has an LED connected on most Arduino boards.
// give it a name:
int led = 13;
```

The next part of the sketch is the "setup" function. Every Arduino sketch must have a "setup" function, and this function runs every time the Arduino is reset, either because (as the comment says) the reset button is pressed, or the Arduino is powered up.

```
// the setup routine runs once when you press reset:
void setup() {
  // initialize the digital pin as an output.
  pinMode(led, OUTPUT);
}
```

The structure of this text is a little confusing if you are new to programming. A function is a section of code that has been given a name (in this case, the name is "setup"). For now, simply use the text just cited as a template and understand that it must start with the first line "void setup() {". Afterward, you place each of the commands you want to issue on a line ending with ";", and then mark the end of the function with a "}" symbol.

In this case, the only command we expect the Arduino to carry out is to issue the "pinMode(led, OUTPUT)" command that not unsurprisingly sets that pin to be an output.

Next comes the juicy part of the sketch: the "loop" function.

Like the "setup" function, every Arduino sketch must have a "loop" function. Unlike "setup" which only runs once after a reset, the "loop" function runs continuously. That is, as soon as all its instructions have been done, it starts again.

In the "loop" function, we first turn on the LED by issuing the "digitalWrite (led, HIGH)" instruction. We then pause for a second by using the command "delay(1000)". The value is 1000 for 1000 milliseconds, or one second. We then turn the LED back off again, and delay for another second before the whole process starts over.

```
// the loop routine runs over and over again forever:

void loop() {
  digitalWrite(led, HIGH);   // turn the LED on (HIGH is the voltage level)
  delay(1000);               // wait for a second
  digitalWrite(led, LOW);    // turn the LED off by making the voltage LOW
  delay(1000);               // wait for a second
}
```

To modify this sketch to make the LED blink faster, change both occurrences of 1000 to 200. These changes are both in the "loop" function, so your function will now look like this:

```
void loop() {
  digitalWrite(led, HIGH);   // turn the LED on (HIGH is the voltage level)
  delay(200);               // wait for a second
  digitalWrite(led, LOW);    // turn the LED off by making the voltage LOW
  delay(200);               // wait for a second
}
```

If you try to save the sketch before uploading it, you will be reminded that it is a "read-only" example sketch, but the Arduino IDE will offer you the option to save it as a copy, which you can then modify to your heart's content.

You do not have to do this, of course. You can just upload the sketch unsaved, but if you do decide to save this or any other sketch, you will find that it then appears in the File | Sketchbook menu on the Arduino IDE.

So, either way, click the Upload button again. When the uploading is complete, the Arduino will reset itself and the LED should start to blink much faster.

How to Make an Arduino Control a Relay

The USB connection of an Arduino can be used for more than just programming the Arduino. You can also use it to send data between the Arduino and your computer. If we attach a relay to the Arduino, we could send a command from our computer to turn the relay on and off.

Relays

A relay (Figure 6-7) is an electromechanical switch. It's very old technology, but relays are cheap and very easy to use.

A relay is basically an electromagnet that closes switch contacts. The fact that the coil and the contacts are electrically isolated from one another makes relays great for things like switching home-powered devices on and off from something like an Arduino.

Whereas the coil of a relay is often energized by between 5V and 12V, the switch contacts can control high-power, high-voltage loads. For example, the relay photographed in Figure 6-7 claims a maximum current of 10A at 120V AC (household power) as well as 10A at 24V DC.

Arduino Outputs

Arduino outputs, and for that matter inputs, are referred to as "pins," even though if you look at the connectors along the sides of the Arduino, they are most definitely sockets rather than pins. The name harkens back to the pins on the microcontroller IC at the heart of the Arduino that were connected to the sockets.

Each of these "pins" can be configured to act as either an input or an output. When they are acting as an output, each pin can provide up to 40mA. This is more than enough to light

FIGURE 6-7 A relay

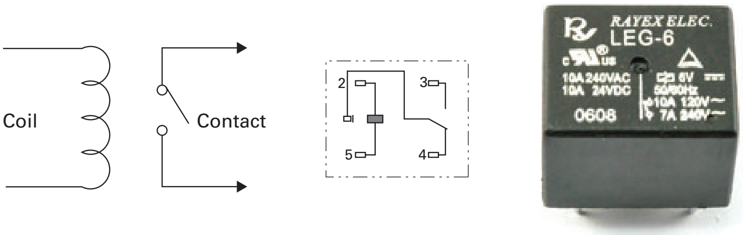

Relay Schemtic

Relay Package

A Relay

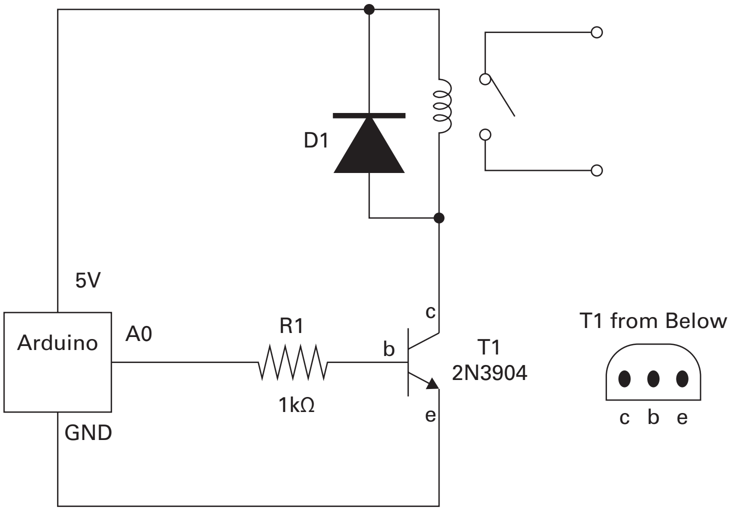

FIGURE 6-8 Schematic diagram of an Arduino-controlled relay

an LED, but not enough to energize a relay coil, which typically requires more like 100mA.

This is a problem we have already discussed. Since we want to use a small current to control a larger one, a good way to do this is by using a transistor.

Figure 6-8 shows a schematic diagram of what we are going to build.

We are using a transistor just like we did when we were controlling an LED. One difference in the schematic is that there is a diode across the relay coil. This is required because when you turn the relay off and the magnetic field in the coil collapses, you get a spike of voltage. The diode prevents this from damaging anything.

We are going to solder the components to the relay and then attach the necessary leads to a header strip that will plug into the Arduino (Figure 6-9). The header strip has 15 pins and spans both of the connector sockets on the side of the Arduino closer to the microcontroller chip. There is a gap between the two connector strips, so one of the header pins will not actually be fitted into a socket.

FIGURE 6-9 The Arduino relay interface

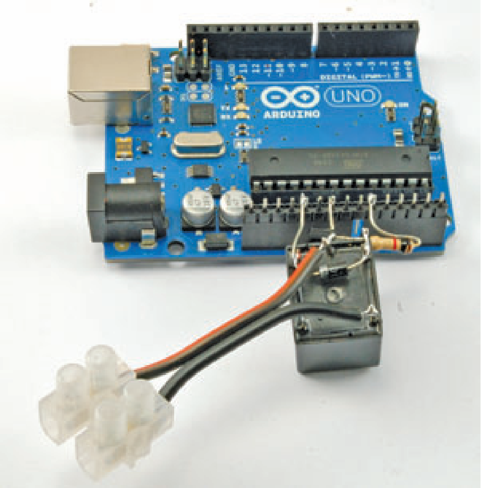

You Will Need

Quantity	Names	Item	Appendix Code
1		Arduino Uno/Leonardo	M2/M21
1		USB lead; Type B for Uno, Micro USB for Leonardo	
1		Transistor 2N3904	K1, S1
1	R1	1kΩ 0.25-W resistor	K2
1	D1	1N4001 diode	K1, S5
1	Relay	5V Relay	H16
1		* Pin header 15-way	K1, H4
1		Two-way screw terminal	H5

* Pin headers are usually supplied in long lengths designed to be snapped into whatever number of connections you need.

Construction

Figure 6-10 shows how the components are attached. First, solder the diode to the relay coil contacts. These are the two pins on the far side of the relay that have three pins more or less in a row. The stripe on the diode should be to the right, as shown in Figure 6-10.

After soldering the diode across the relay coils, bend out the leads of the transistor and position it as shown in Figure 6-10 with the flat side against the relay. Shorten the base (middle) lead of the transistor, shorten the leads on the resistor and attach it to the base lead.

Finally, solder the three leads to the connector strip. The resistor lead should go to the 6th lead from the left, the emitter of the transistor to the 9th from the left, and the diode lead to the 11th from the left.

Before we attach leads to the relay contacts, we can test our work using a multimeter in Continuity mode, so attach the header pins to the Arduino as shown in Figure 6-9 and clip one lead of the multimeter (on Continuity mode) to the middle contact of the relay (in between the diode leads). Attach the other lead of the multimeter to each of the two unconnected contacts on the relay. One will buzz and the other will not. Attach the lead to

Figure 6-10 Wiring the relay interface

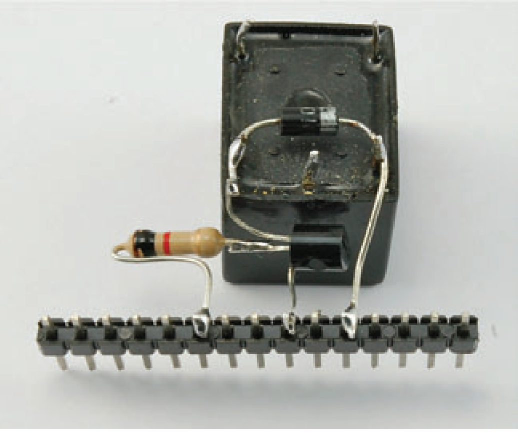

the one that does not cause the multimeter to beep—this is the n.o. (normally open contact).

Load the sketch "relay_test" into Arduino and upload it to the Arduino board. When the Arduino restarts, you should find that every two seconds the relay will flip from being open to being closed.

Software

The sketch for this is much the same as the Blink sketch.

```
// relay_test

int relayPin = A0;

void setup()
{
  pinMode(relayPin, OUTPUT);
}

void loop()
{
  digitalWrite(relayPin, HIGH);
  delay(2000);
  digitalWrite(relayPin, LOW);
  delay(2000);
}
```

The only real difference is that we are using pin A0 rather than pin 13. Arduino has a feature that allows you to use the analog input pins A0 to A5 as digital inputs or outputs as well as analog inputs, but you have to put the letter A in front of them when using them as digital pins.

If all is well, then to make it easier to attach things to the relay contacts, we can solder some wires to them and use a two-way screw terminal block (Figure 6-11).

The relay module can be used to control all sorts of things. It can be used to control 110 or 240V home-powered devices, but be very sure you know what you are doing before you attempt that. If you plan to do this, everything must

FIGURE 6-11 Attaching leads to the relay contacts

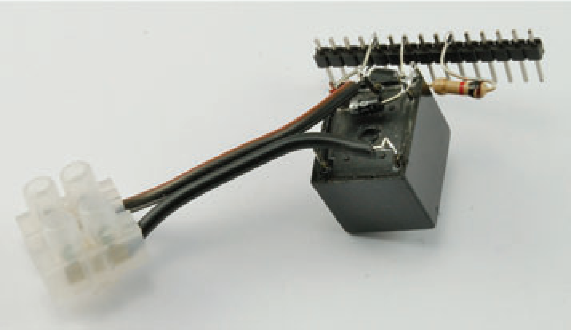

be insulated and the whole project enclosed in a plastic box, especially since touching a "hot" wire can and does kill many people every year.

In the next section, you will hack an electrical toy so it can be turned on and off using the Arduino and relay module you have just built.

How to Hack a Toy for Arduino Control

The great thing about a relay is that it behaves just like a switch. This means that if you have something you want to turn on and off from your Arduino and that item has a switch, then all you need to do is solder some wires to the switch and attach them to the relay. This would allow both the relay and the switch to turn the device on and off. But if you do not want to keep the original switch, it can be removed, as it will be in this case.

The toy that the author chose to hack is a little electric bug (Figure 6-12).

Figure 6-12 The hapless electric bug awaiting dissection

You Will Need

As well as the relay module built in the section "How to Make an Arduino Control a Relay," you will also need the following items.

Quantity	Item	Appendix Code
1	Arduino Uno/Leonardo	M2/21
1	USB lead; Type B for Uno, Micro USB for Leonardo	
1	An electric toy (battery-powered) with an on–off switch	
1	Twin multi-core wire	

Construction

Taking the toy apart, you can see the connections to the switch (Figure 6-13a). De-solder the switch and attach wires to the leads that used to go to the switch (Figure 6-13b). You should always put insulating tape around the bare wires to prevent accidental shorts (Figure 6-13c).

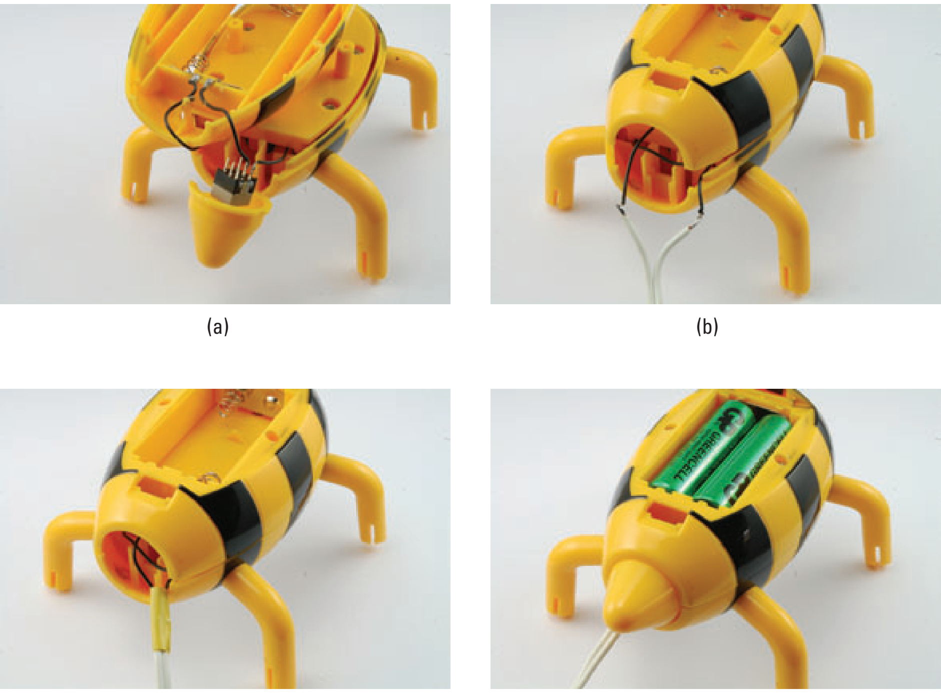

(a) (b)

(c) (d)

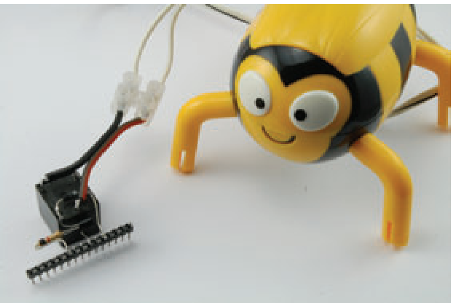

(e)

Figure 6-13 Hacking the toy

The toy can then be assembled with the wire leaving through a gap in the case (Figure 6-13d). If there is no suitable gap, you will probably need to drill a hole.

Finally, the toy is ready to use, so plug the relay interface into the Arduino and connect the wires to its screw terminals (Figure 6-13e). If the test sketch is still installed, you should find that the toy repeatedly turns on for a couple of seconds, and then turns back off again.

This is okay, but not terribly useful. We will use another sketch that will allow us to send commands to the Arduino from your computer. The sketch is called “relay_remote”.

Upload this sketch to the Arduino. Then, open the Serial Monitor by clicking the button on the right-hand side of the Arduino IDE (circled in Figure 6-14).

The Serial Monitor

The Serial Monitor is part of the Arduino IDE that allows you to send and receive data between your computer and the Arduino board (Figure 6-15).

At the top of the Serial Monitor is an area where we can type commands. When we click the Send button, these are sent to the Arduino. We can see any messages that the Arduino has sent in the area below this.

Try this out by typing in the number 1 and clicking Send. This should start your toy. Entering "0" should turn it off again.

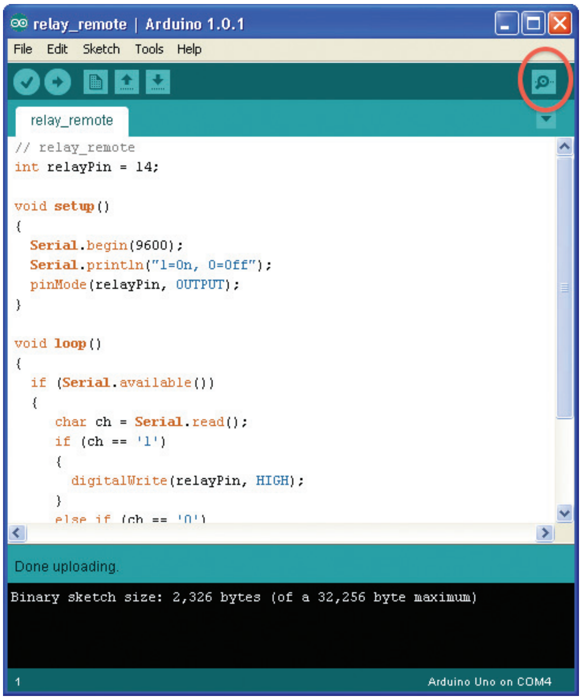

```
// relay_remote
int relayPin = 14;

void setup()
{
  Serial.begin(9600);
  Serial.println("1=On, 0=Off");
  pinMode(relayPin, OUTPUT);
}

void loop()
{
  if (Serial.available())
  {
     char ch = Serial.read();
     if (ch == '1')
     {
       digitalWrite(relayPin, HIGH);
     }
     else if (ch == '0')
```

```
Done uploading.
Binary sketch size: 2,326 bytes (of a 32,256 byte maximum)
```

FIGURE 6-14 Opening the Serial Monitor

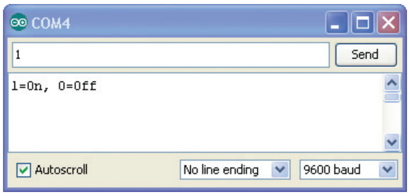

1

```
1=On, 0=Off
```

☑ Autoscroll

No line ending

9600 baud

FIGURE 6-15 The Serial Monitor

Software

Let's now look at the sketch.

```
// relay_remote
int relayPin = A0;

void setup()
{
  Serial.begin(9600);
  Serial.println("1=On, 0=Off");
  pinMode(relayPin, OUTPUT);
}

void loop()
{
  if (Serial.available())
  {
     char ch = Serial.read();
     if (ch == '1')
     {
       digitalWrite(relayPin, HIGH);
     }
     else if (ch == '0')
     {
       digitalWrite(relayPin, LOW);
     }
  }
}
```

Notice that the "setup" function now has two new commands in it.

```
Serial.begin(9600);
Serial.println("1=On, 0=Off");
```

The first of these starts serial communications over the serial port at 9600 baud. The second sends the prompt message, so that we know what to do when the Serial Monitor opens.

The "loop" function first uses the function "Serial.available()" to check if there is any communication from the computer waiting to be processed. If there is, then this is read into a character variable.

We then have two if statements. The first checks to see if the character is a "1", and if it is, it turns on the toy. If, on the other hand, the character read is a "0", it turns it off.

We have made a little bit of a leap from our first flashing sketch, and if you need more help understanding how the sketch works, you might consider buying the book *Programming Arduino: Getting Started with Sketches* by this author.

How to Measure Voltage with an Arduino

The pins labeled A0 to A5 on an Arduino are analog inputs. That is, you can use them to measure voltage. To demonstrate this, you will use the variable resistor (trimpot) as a voltage divider connected to A3 (Figure 6-16).

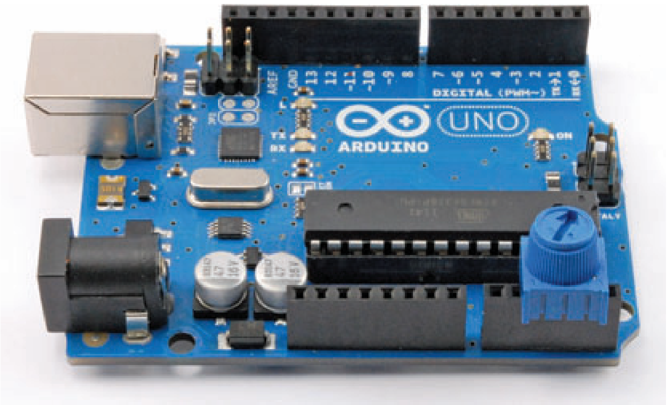

FIGURE 6-16 A variable resistor and Arduino

If you skipped the section on voltage dividers in Chapter 3 titled “How to Use Resistors to Divide a Voltage,” you probably should go back and have a quick look now.

You Will Need

To try this example, you will need the following items.

Quantity	Names	Item	Appendix Code
1		Arduino Uno/Leonardo	M2/M21
1		USB lead; Type B for Uno, Micro USB for Leonardo	
1	R1	10kΩ variable resistor	K1, R1

Construction

The construction of this project is very simple. There is no actual soldering involved, we will just push the three pins of the variable resistor into the Arduino sockets A2, A3, and A4. Figure 6-17 shows the schematic diagram for this.

You may be wondering how this will work, since you would normally expect the top of the variable resistor to be connected to 5V and the bottom to GND. Well, since a 10kΩ resistor at 5V will only be allowing 0.5mA of current to flow, we can use pins A2 as digital outputs and set A2 to 0V and A4 to 5V.

Plug the variable resistor into the Arduino so that the middle slider connection is to pin A3 and the pins on either side are connected to A2 and A4.

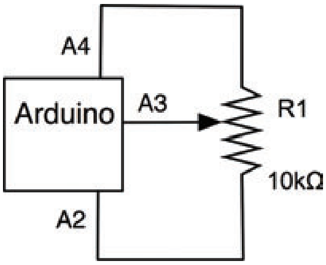

FIGURE 6-17 Schematic diagram—measuring voltage with an Arduino

Software

Load the sketch “voltmeter” into the Arduino IDE and then program your Arduino board with it. Open the Serial Monitor and you should see something like Figure 6-18.

Try twiddling the knob from one end of the range to the other. You should find that you can set the voltage to anything between 0 and 5V.

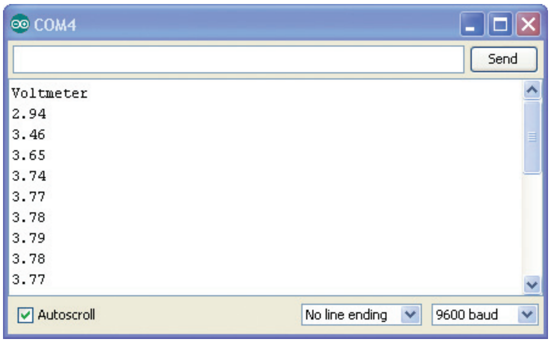

```
Voltmeter
2.94
3.46
3.65
3.74
3.77
3.78
3.79
3.78
3.77
```

☑ Autoscroll

FIGURE 6-18 The Serial Monitor showing voltage at A3

```
// voltmeter

int voltsInPin = 3;
int gndPin = A2;
int plusPin = A4;
```

```
void setup()
{
  pinMode(gndPin, OUTPUT);
  digitalWrite(gndPin, LOW);
  pinMode(plusPin, OUTPUT);
  digitalWrite(plusPin, HIGH);
  Serial.begin(9600);
  Serial.println("Voltmeter");
}

void loop()
{
  int rawReading = analogRead(voltsInPin);
  float volts = rawReading / 204.8;
  Serial.println(volts);
  delay(200);
}
```

The sketch defines the pins as usual. Note that when referring to the analog input pins for use as analog inputs (as with “voltsInPin”), we just use the pin number. So for A3 we specify just 3. However, because we are using A2 and A4 as digital outputs, we have to add the letter A to the front.

The “setup” function sets the pin modes, but also sets “gndPin” and “plusPin” to LOW and HIGH, respectively, before starting serial communication and sending a welcome message.

Inside the loop, we use “analogRead” to give us a raw value of between 0 and 1023, where 0 means 0V and 1023 means 5V. To convert this into an actual voltage, we need to divide it by 204.8 (1023/5). In dividing a raw reading that is an integer (whole number) by a decimal number of 204.8 (called floats in Arduino), the result will be a float, and so we specify the type of the “volts” variable to be “float”.

Finally, we print out the voltage and then wait 200 milliseconds before we take the next reading. We don’t have to wait before taking the next reading, it just stops the readings from flying up the screen too fast to read.

In the next section, we will use the same hardware with the addition of an external LED and a slightly different sketch to change the rate at which an external LED flashes.

How to Use an Arduino to Control an LED

There are three useful things to be learned here. The first is how to make an Arduino drive an LED. The second is how to control the rate of flashing using a reading from a variable resistor, and finally we will show how to use the Arduino to control the power going to the LED and thus determine its brightness (Figure 6-19).

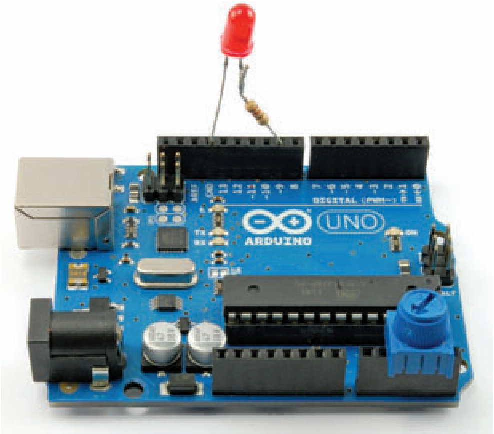

FIGURE 6-19 An Arduino, variable resistor, and LED

You Will Need

To try this example, you will need the following items.

Quantity	Names	Item	Appendix Code
1		Arduino Uno/Leonardo	M2/M21
1		USB lead; Type B for Uno, Micro USB for Leonardo	
1	R1	10kΩ variable resistor	K1, R1
1	R2	220Ω Resistor	K2
1	D1	LED	K1

Construction

As we discussed in Chapter 4, LEDs need a resistor to stop them drawing too much current. This means we cannot just push one straight into an output pin of an Arduino. So, we are going to start by taking an LED and resistor, shortening the leads, soldering them together and making an LED resistor combo that we can just plug into our Arduino. Figure 6-20 shows the steps involved in making this.

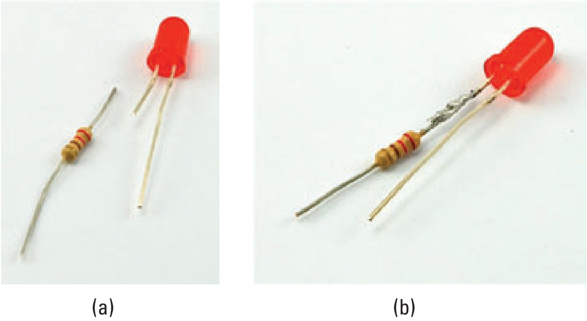
(a) (b)

FIGURE 6-20 Making an LED resistor combo

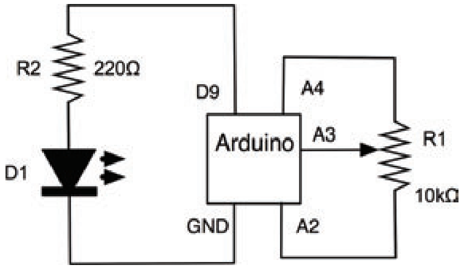

FIGURE 6-21 Schematic for an LED, Arduino, and a variable resistor

To avoid confusion, put the resistor in the anode (positive and longer) lead and keep the combined lead the longer lead, so you know it is the positive end of the combo.

The schematic diagram for the arrangement is shown in Figure 6-21.

We will use pin 9 as a digital output for the LED. The other end of the LED combo being connected to a convenient GND connection.

Keep this Arduino-friendly LED combo. You will use it again in several later sections.

Software (Flashing)

You will use two different sketches with this arrangement of hardware. The first uses the variable resistor to control the speed of the flashing, while the second will control the brightness of the LED.

Attach the LED resistor combo, as shown in Figure 6-19, and load the sketch "variable_led_flash" onto your Arduino board. You should find that turning the knob controls the rate at which the LED flashes.

```
// variable_led_flash

int voltsInPin = 3;
int gndPin = A2;
int plusPin = A4;
int ledPin = 9;

void setup()
{
  pinMode(gndPin, OUTPUT);
  digitalWrite(gndPin, LOW);
  pinMode(plusPin, OUTPUT);
  digitalWrite(plusPin, HIGH);
  pinMode(ledPin, OUTPUT);
}

void loop()
{
  int rawReading = analogRead(voltsInPin);
  int period = map(rawReading, 0, 1023, 100, 500);
  digitalWrite(ledPin, HIGH);
  delay(period);
  digitalWrite(ledPin, LOW);
  delay(period);
}
```

The sketch is quite similar to that in the previous section; however, we no longer use the Serial Monitor, so all that code is gone. We do need to define a new pin "ledPin" to use for the LED.

The "loop" function still reads the raw value from the analog pin A3, but it then uses the "map" function to convert the "rawReading" value of between 0 to 1023 to a range of 100 to 500.

The "map" function is a standard Arduino command that adjusts the range of the value passed in as the first parameter. The second and third parameters are the range of the raw value, while the fourth and fifth are the desired range you want to compress or expand the value into.

We then flash the LED using this number (100 to 500) as the delay between turning the LED on and off. The end result of this is that the LED will flash faster the closer A3 is to 0V.

Software (Brightness)

We can use exactly the same hardware, but with different software to control the brightness of the LED instead of its rate of flashing. This will use the Arduino "analogWrite" function to vary the power going to the pin. This feature is only available for those pins marked with a "~" on the Arduino board. Fortunately, we thought ahead and chose such a pin to connect the LED to.

These pins can use a technique called pulse-width modulation (PWM) to control how much power goes to the output. This works by sending out a series of pulses, around 500 times per second. These pulses may be high for only a short time, in which case little power is delivered, or high until it's nearly time for the next pulse, in which case lots of power is delivered.

In the case of the LED, this means that in each cycle, the LED is either off, on for some of the time, or on the whole time. Our eyes cannot keep up with such a fast-changing event, so it just appears that the brightness of the LED varies.

Load the sketch "variable_led_brightness" onto your Arduino. You should find that, now, the variable resistor controls the brightness of the LED rather than its rate of flashing.

Most of the sketch is the same as the previous one, the difference lies in the "loop" function.

```
void loop()
{
  int rawReading = analogRead(voltsInPin);
  int brightness = rawReading / 4;
  analogWrite(ledPin, brightness);
}
```

The function “analogWrite” expects a value between 0 and 255, so we can take our raw analog reading of between 0 and 1023 and divide it by 4 to put it into roughly the right range.

How to Play a Sound with an Arduino

The first Arduino sketch we tried at the start of this chapter flashed an LED on and off. If we turn a digital output pin on and off much faster than this, we can drive a sounder to create a sound. Figure 6-22 shows a simple sound generator that plays one of two notes when a button is pressed.

You Will Need

To have your Arduino make sounds, you will need the following items.

Quantity	Names	Item	Appendix Code
1		Arduino Uno/Leonardo	M2/M21
1		USB lead; Type B for Uno, Micro USB for Leonardo	
2	S1, S2	Miniature push switches	K1
1	Sounder	Small piezo sounder	M3
1		Breadboard	T5
		Jumper wires or solid-core wire	T6

Figure 6-22 A simple Arduino tone generator

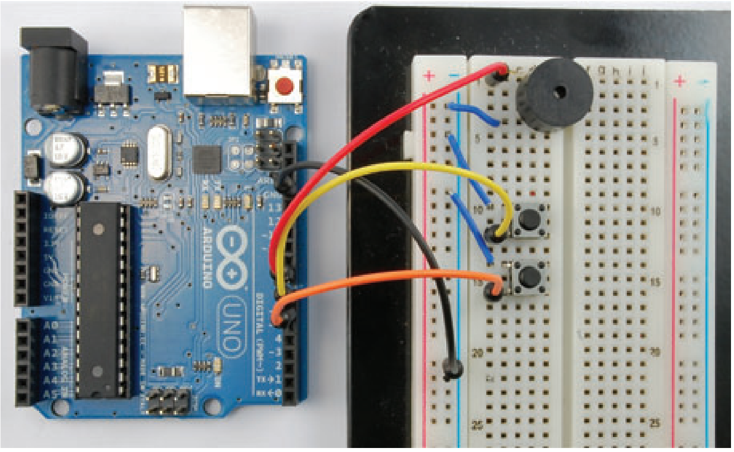

Construction

Figure 6-23 shows the schematic for the tone generator, while Figure 6-24 displays the breadboard layout.

Make sure the push switches are the right way around. They should be positioned so the pins extend out of the sides rather than the top and bottom. The piezo sounder may have one pin indicated as positive. Position this at the top of the breadboard.

Attach the components as shown and connect the jumper leads to the Arduino.

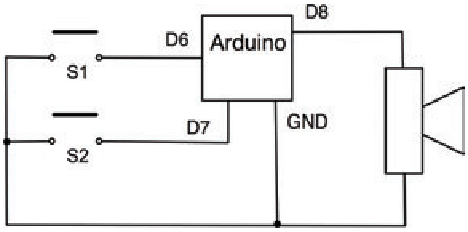

FIGURE 6-23 Schematic diagram for the tone generator

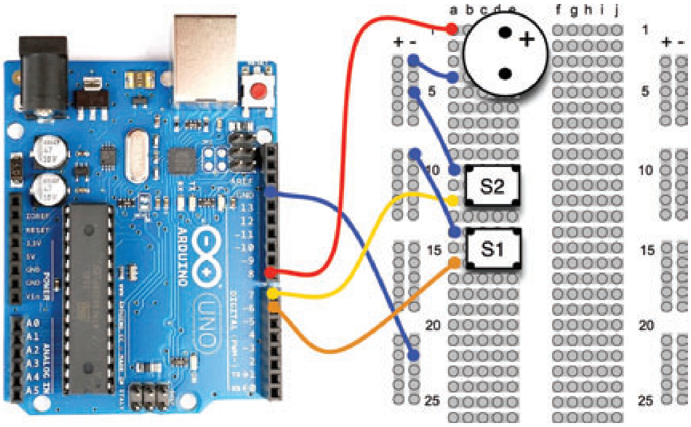

FIGURE 6-24 Breadboard layout for the tone generator

Software

The sketch is quite straightforward and follows what should be a familiar pattern by now.

```
// arduino_sounds

int sw1pin = 6;
int sw2pin = 7;
int soundPin = 8;

void setup()
{
  pinMode(sw1pin, INPUT_PULLUP);
  pinMode(sw2pin, INPUT_PULLUP);
  pinMode(soundPin, OUTPUT);
}

void loop()
{
  if (! digitalRead(sw1pin))
  {
    tone(soundPin, 220);
  }
  else if (! digitalRead(sw2pin))
  {
    tone(soundPin, 300);
  }
  else
  {
    noTone(soundPin);
  }
}
```

First, we define the variables for the pins. The switches will be connected to "sw1pin" and "sw2pin". These will be digital inputs, while the "soundPin" will be a digital output.

Note that in the setup function for the switch pins, we use the command "pinMode" with the parameter INPUT_PULLUP. This sets the pin to be an input, but also enables a "pull-up" resistor built into the Arduino, which keeps the input pin HIGH, unless we pull it LOW by pressing the button.

It is because the input pins are normally high that in the "loop" function, when we are checking to see if a button is pressed, we have to use the "!" (logical not). In other words, the following will only sound the tone if the digital input pin "sw1pin" is LOW.

```
if (! digitalRead(sw1pin))
{
tone(soundPin, 220);
}
```

The "tone" function is a useful built-in Arduino function that plays a tone on a particular pin. The second parameter is the frequency of the tone in Hertz (cycles per second).

If no key is pressed, then the function "noTone" is called and stops any tone that is playing.

How to Use Arduino Shields

The success of Arduino had been in no small part due to the wide range of plug-in shields that add useful features to a basic Arduino board. A shield is designed to fit into the header sockets of the main Arduino board. Most shields will then pass through these connections in another row of header sockets, making it possible to construct stacks of shields with an Arduino at the bottom. Shields that have, say, a display on them, will not normally pass through in this way. You also need to be aware that if you stack shields like this, you need to make sure there are no incompatibilities, such as two of the shields using the same pin. Some shields get around this problem by providing jumpers to add some flexibility to pin assignments.

Shield	Description	URL
Motor	Ardumoto shield. Dual H-bridge bidirectional motor control at up to 2A per channel.	www.sparkfun.com/products/9815
Ethernet	Ethernet and SD card shield.	http://arduino.cc/en/Main/ArduinoEthernetShield
Relay	Controls four relays. Screw terminals for relay contacts.	www.robotshop.com/seeedstudio-arduino-relay-shield.html
LCD	16 × 2 character alphanumeric LCD shield with joystick.	www.freetronics.com/products/lcd-keypad-shield

TABLE 6-1 Some Commonly Used Shields

The web site http://shieldlist.org/ will tell you which pins are used by any particular shield.

There are shields available for almost anything you could want an Arduino to do. They range from relay control to LED displays and audio file players.

Most of these are designed with the Arduino Uno in mind, but are also usually compatible with the bigger Arduino Mega and the new Arduino Leonardo.

For an encyclopedic list, that includes useful technical details about the pin usage of these shields, it can be found at http://shieldlist.org/.

Some of the author's favorite shields are listed in Table 6-1.

How to Control a Relay from a Web Page

By using an Ethernet Shield and connecting it to your home hub, you can turn your Arduino into a tiny web server. Since it is still an Arduino, you can still attach electronics to it. So, by using the hacked toy we made in the section "How to Hack a Toy for Arduino Control" and a web interface on the Arduino, we can control our toy over our local network, or if we open up our firewall from the Internet!

Figure 6-25 shows the toy attached to the shield and Arduino along with the browser interface that we will use to control it—first on our computer (Figure 6-25b) and then from a smartphone (Figure 6-25c).

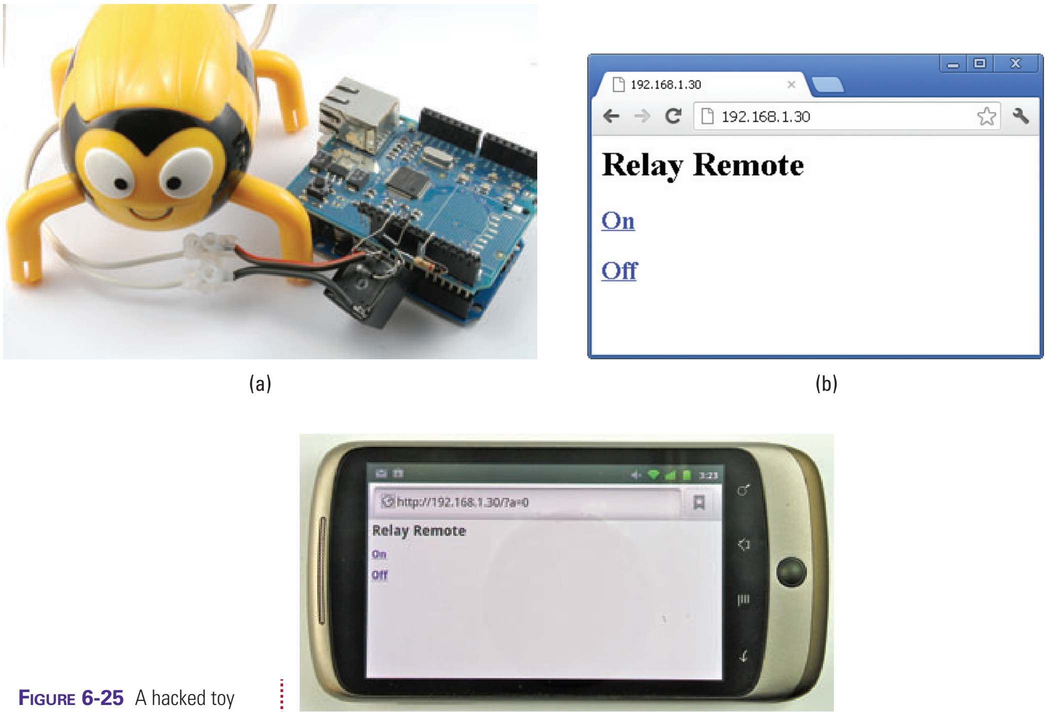
192.168.1.30
192.168.1.30
Relay Remote
On
Off
http://192.168.1.30/?a=0
Relay Remote
On
Off

(a) (b)

(c)

FIGURE 6-25 A hacked toy controllable over the network

You Will Need

To make your web-controlled toy, you will first need to complete the section titled "How to Hack a Toy for Arduino Control." In addition, you will need the following items:

Quantity	Item	Appendix Code
1	Arduino Ethernet Shield	M4
1	Ethernet patch cable	T6
1	9V or 12V 500mA Power Supply	M1

Note that this project will only work with the Arduino Leonardo if the Ethernet Shield is the latest R3 Ethernet Shield. If you have an older Ethernet Shield, you will either need to buy a new Ethernet Shield or use an Arduino Uno.

Construction

In this project, the Arduino is powered from an external power supply rather than the USB connection to the computer. There are two reasons for this. One is that the Ethernet Shield will not operate just from USB power, and the other is that once the Arduino has been programmed—there is no need for it to be connected to your computer, so it may as well be powered from a separate adapter.

The wiring diagram for the project is shown in Figure 6-26.

Connect everything up as shown in Figure 6-26 and load the sketch "web_relay" into the Arduino IDE. Don't upload it onto the Arduino itself just yet. There are some configuration changes we need to make first.

FIGURE 6-26 Wiring diagram for remote relay control

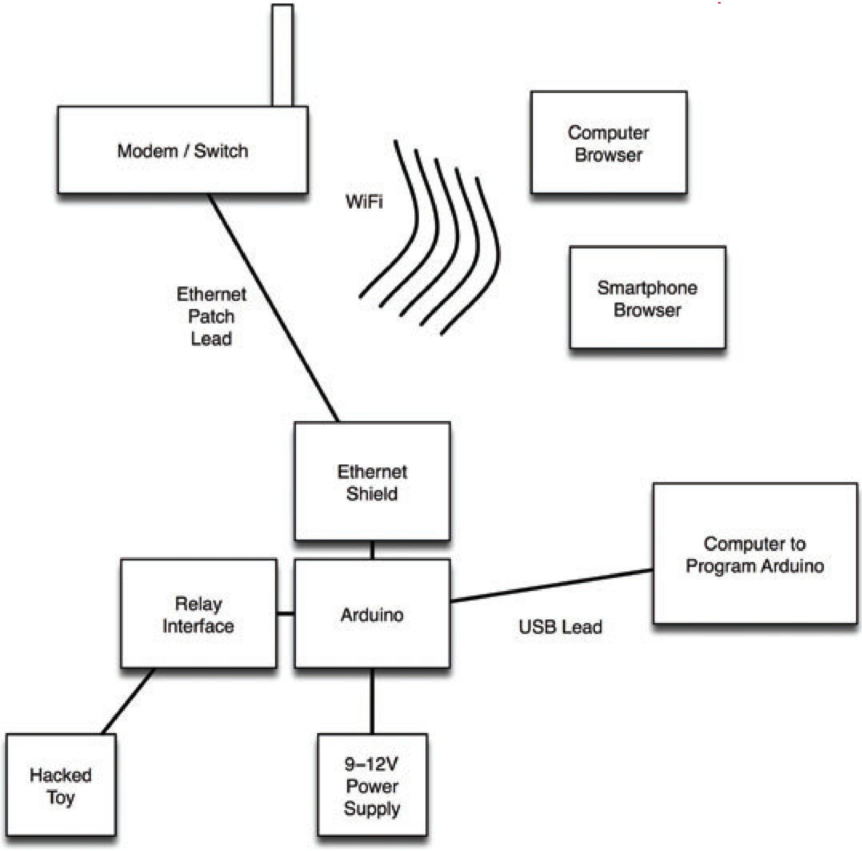
Modem / Switch
WiFi
Computer Browser
Smartphone Browser
Ethernet Patch Lead
Ethernet Shield
Computer to Program Arduino
Relay Interface
Arduino
USB Lead
Hacked Toy
9–12V Power Supply

Network Configuration

If you look at the top of the sketch, you will see the lines:

```
byte mac[] = { 0xDE, 0xAD, 0xBE, 0xEF, 0xFE, 0xED };
byte ip[] = { 192, 168, 1, 30 };
```

The first of these, the "mac address" just has to be unique amongst all the devices connected to your network. Note that some of the newer Ethernet Shields have a Mac address printed on them. If you have one of these boards, enter that here. The second one is the IP address. Most devices you connect to your home network will have IP addresses assigned to them automatically by a process called DHCP. That's just fine if you don't really need to know the IP address of your shield and don't mind it changing (say, when using the shield to act like a browser rather than a web server). But in this project, the Arduino and Ethernet Shield are going to act like a web server, so we need to know what the IP address is so we can type it into the address bar of a browser.

You will manually define an IP address. This cannot be any four numbers; they must be numbers that qualify as being internal IP addresses and fit in the range of IP addresses expected by your home router. Typically, the first three numbers will be something like 10.0.1.*x* or 192.168.1.*x*, where *x* is some number between 0 and 255. Some of these IP addresses will be in use by other devices on your network. To find an unused but valid IP address, connect to the administration page for your home router and look for an option that says DHCP. You should somewhere find a list of devices and their IP addresses similar to that in Figure 6-27. Select a final number to use in your IP address. In this case, 192.168.1.30 looked like a good bet and indeed it worked fine.

Set the IP address in the sketch and upload it to your Arduino board.

Testing

Open up a browser on your computer, tablet, or smartphone and navigate to the IP address. If you used the same IP address as the author, that would be http://192.168.1.30. You should then see the web page of Figure 6-25b and 6-25c displayed.

Click the On button and the relay should click on, turning the hacked toy on. The page will then reload in the browser. Clicking "Off" will turn the relay off again.

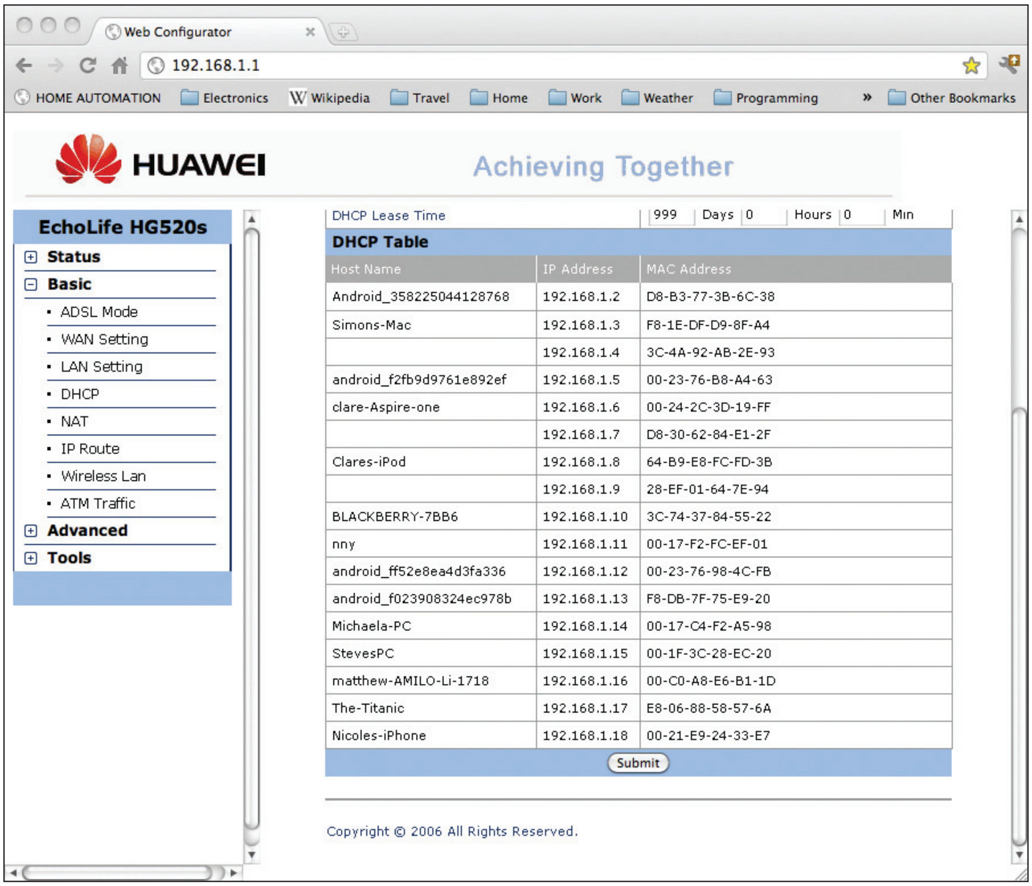
Web Configurator
192.168.1.1
HOME AUTOMATION
Electronics
Wikipedia
Travel
Home
Work
Weather
Programming
Other Bookmarks
HUAWEI
Achieving Together
EchoLife HG520s
Status
Basic
ADSL Mode
WAN Setting
LAN Setting
DHCP
NAT
IP Route
Wireless Lan
ATM Traffic
Advanced
Tools
Submit
Copyright © 2006 All Rights Reserved.

DHCP Lease Time 999 Days 0 Hours 0 Min

DHCP Table

Host Name	IP Address	MAC Address
Android_358225044128768	192.168.1.2	D8-B3-77-3B-6C-38
Simons-Mac	192.168.1.3	F8-1E-DF-D9-8F-A4
	192.168.1.4	3C-4A-92-AB-2E-93
android_f2fb9d9761e892ef	192.168.1.5	00-23-76-B8-A4-63
clare-Aspire-one	192.168.1.6	00-24-2C-3D-19-FF
	192.168.1.7	D8-30-62-84-E1-2F
Clares-iPod	192.168.1.8	64-B9-E8-FC-FD-3B
	192.168.1.9	28-EF-01-64-7E-94
BLACKBERRY-7BB6	192.168.1.10	3C-74-37-84-55-22
nny	192.168.1.11	00-17-F2-FC-EF-01
android_ff52e8ea4d3fa336	192.168.1.12	00-23-76-98-4C-FB
android_f023908324ec978b	192.168.1.13	F8-DB-7F-75-E9-20
Michaela-PC	192.168.1.14	00-17-C4-F2-A5-98
StevesPC	192.168.1.15	00-1F-3C-28-EC-20
matthew-AMILO-Li-1718	192.168.1.16	00-C0-A8-E6-B1-1D
The-Titanic	192.168.1.17	E8-06-88-58-57-6A
Nicoles-iPhone	192.168.1.18	00-21-E9-24-33-E7

Figure 6-27 Selecting an IP address to use

Software

This sketch is one of the most complicated in the book. However, it is one that can easily be used as a template for other hacks using an Arduino as a web server.

Let's break this sketch down and look at it a section at a time.

```
// web_relay

#include <SPI.h>
#include <Ethernet.h>

// MAC address just has to be unique. This should work
byte mac[] = { 0xDE, 0xAD, 0xBE, 0xEF, 0xFE, 0xED };
// The IP address will be dependent on your local
network:
```

```
byte ip[] = { 192, 168, 1, 30 };
EthernetServer server(80);
int relayPin = A0;
char line1[100];
```

Two libraries have to be included for use with the Ethernet Shield: “SPI” and “Ethernet.” Libraries contain a useful collection of functions for, say, a shield. They greatly simplify the writing of sketches as they can just make use of the library functions.

The “SPI” library is for a kind of serial communication that the Arduino uses to send instructions to the Ethernet Shield, and the “Ethernet” library defines some useful functions for us to use with the Ethernet Shield.

After the two variables that define the Mac address and IP address, the next command creates a new “EthernetServer” object that is used every time we want to do something with the Ethernet. We then define the “relayPin” to use and also create a line buffer of 100 characters that is used later in the code when reading the header that arrives from a browser when you navigate to the page being served by the Arduino.

```
void setup()
{
  pinMode(relayPin, OUTPUT);
  Ethernet.begin(mac, ip);
  server.begin();
}
```

The setup function initializes the Ethernet library using the Mac and IP addresses that we set earlier. It also sets the pin mode of the “relayPin” to be an OUTPUT.

```
void loop()
{
  EthernetClient client = server.available();
  if (client)
  {
    while (client.connected())
    {
      readHeader(client);
      if (! pageNameIs("/"))
      {
        client.stop();
        return;
      }
      digitalWrite(relayPin, valueOfParam('a'));
```

```
        client.println("HTTP/1.1 200 OK");
        client.println("Content-Type: text/html");
        client.println();

        // send the body
        client.println("<html><body>");
        client.println("<h1>Relay Remote</h1>");

        client.println("<h2><a href='?a=1'/>On</a></h2>");
        client.println("<h2><a href='?a=0'/>Off</a></h2>");
        client.println("</body></html>");

        client.stop();
      }
    }
  }
```

The loop function is responsible for servicing any requests that come to the web server from a browser. If a request is waiting for a response, then calling "server.available" will return us a "client". If client exists (tested by the first "if" statement), we can then determine if it is connected to the web server by calling "client.connected".

We will come to the "readHeader" function later. This function and "pageNameIs" are used to determine that the browser is actually contacting the page for setting the relay. This is because browsers will often send two requests to a server page, one to try and find an icon for the web site, and a second to the page itself. This code allows us to ignore the icon request.

The next line sets the relay pin using "digitalWrite". The value it sets the output to is whatever value the request parameter "a" is set to. This will be either "1" or "0".

The next three lines of code print out a return header. This just tells the browser what type of content to display. In this case, just HTML.

Once the header has been written, it simply remains to write the remaining HTML back to the browser. This must include the usual "<html>" and "<body>" tags, and also includes a "<h1>" header tag and two "<h2>" tags that are also hyperlinks to this same page, but with the request parameter "a" set to either "0" or "1".

Finally, "client.stop" tells the browser that the message is complete and the browser will display the page.

```
void readHeader(EthernetClient client)
{
  // read first line of header
  char ch;
  int i = 0;
  while (ch != '\n')
  {
    if (client.available())
    {
      ch = client.read();
      line1[i] = ch;
      i ++;
    }
  }
  line1[i] = '\0';
  Serial.println(line1);
}
```

The final three functions in the sketch are general-purpose functions that I tend to use over and over again when making an Arduino web server like this.

The first, “readHeader”, reads the header of the request coming from the browser into the buffer “line”. We can then use this in the next two functions.

```
boolean pageNameIs(char* name)
{
   // page name starts at char pos 4
   // ends with space
   int i = 4;
   char ch = line1[i];
   while (ch != ' ' && ch != '\n' && ch != '?')
   {
     if (name[i-4] != line1[i])
     {
       return false;
     }
     i++;
     ch = line1[i];
   }
   return true;
}
```

The function “pageNameIs” returns true if the page name part of the header matches the argument supplied. This is what we use in the "loop" function to ignore the icon request from the browser.

```
int valueOfParam(char param)
{
  for (int i = 0; i < strlen(line1); i++)
  {
    if (line1[i] == param && line1[i+1] == '=')
    {
      return (line1[i+2] - '0');
    }
  }
  return 0;
}
```

The "valueOfParam" lets you read the value of the request parameter supplied as an argument. This is much more restricted than the kind of request parameter you will be used to if you have done any web programming. First, the request parameter name must be a single character, and second, its value must be a single digit between 0 and 9. The function will return the value or 0 if there is no parameter of that name.

This is one of those projects that can be adapted for all sorts of purposes.

How to Use an Alphanumeric LCD Shield with Arduino

Another commonly used Arduino shield is the LCD shield (Figure 6-28).

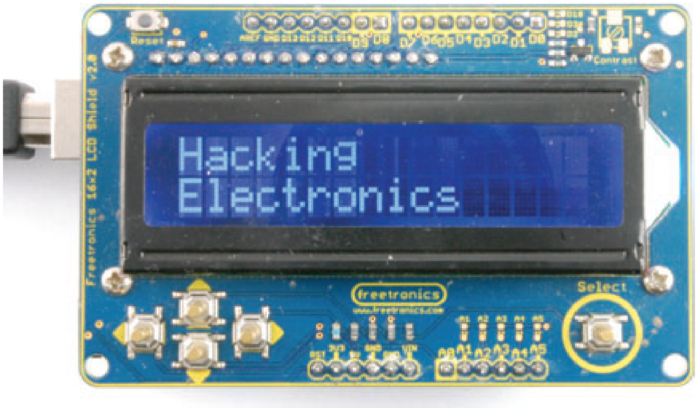

FIGURE 6-28 An LCD shield

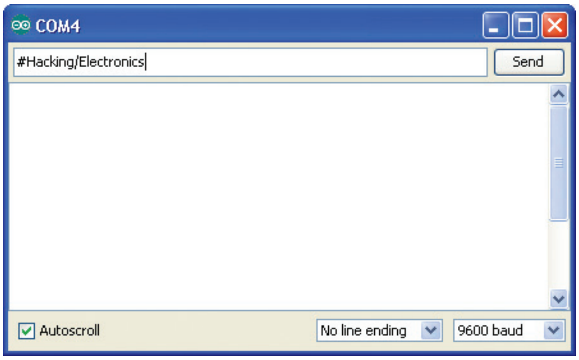

#Hacking/Electronics

☑ Autoscroll

No line ending

9600 baud

FIGURE 6-29 Sending a message with the Serial Monitor

There are many different shields available and most use an LCD module based on the HD44780 LCD driver chip. The model used here is the Freetronics LCD and Keypad Shield (www.freetronics.com). Most other LCD projects will work with this example code, but you may have to change the pin allocations (discussed later).

This project lets you send a short message (the display is only two lines of 16 characters) using the Serial Monitor (Figure 6-29).

You Will Need

To experiment with an LCD display, you will need the following items.

Quantity	Item	Appendix Code
1	Arduino Uno	M2
1	USB Type A to Type B (as commonly used for USB printers)	
1	LCD Shield	M18

Construction

There is really not very much to construct here. Just plug the LCD shield onto the Arduino and plug in the Arduino to your computer via a USB port.

Software

The software is pretty straightforward too. Again, most of the work is done in the library.

```
// LCD_messageboard

#include <LiquidCrystal.h>

// LiquidCrystal display with:
// rs on pin 8
// rw on pin 11
// enable on pin 9
// d4-7 on pins 4-7
LiquidCrystal lcd(8, 11, 9, 4, 5, 6, 7);

void setup()
{
  Serial.begin(9600);
  lcd.begin(2, 16);
  lcd.clear();
  lcd.setCursor(0,0);
  lcd.print("Hacking");
  lcd.setCursor(0,1);
  lcd.print("Electronics");
}

void loop()
{
  if (Serial.available())
  {
    char ch = Serial.read();
    if (ch == '#')
    {
      lcd.clear();
    }
    else if (ch == '/')
    {
      lcd.setCursor(0,1);
    }
    else
    {
      lcd.write(ch);
    }
  }
}
```

If you are using a different LCD shield, then check the specification to see which pins it uses for what. You may need to modify the line:

```
LiquidCrystal lcd(8, 11, 9, 4, 5, 6, 7);
```

The parameters to this are the pins that the shield uses for (rs, rw, e, d4, d5, d6, d7). Note that not all shields use the rw pin. If this is the case, just pick the number of a pin not being used for anything else.

The loop reads any input, and if it is a # character, it clears the display. If it is a "/" character, it moves to the second row; otherwise, it just displays the character that was sent.

For example, to send the text displayed in Figure 6-28, you would enter the following into the Serial Monitor:

```
#Hacking/Electronics
```

Notice that the LCD library provides you with the "lcd .setCursor" function to set the position for the next text to be written. The text is then written using the "lcd.write" function.

How to Drive a Servo Motor with an Arduino

Servo motors are a combination of motor, gearbox, and sensor that are often found in remote-controlled vehicles to control steering or the angles of surfaces on remote-controlled airplanes and helicopters.

Unless they are special-purpose servo motors, servo motors do not rotate continuously. They usually only rotate through about 180 degrees, but can be accurately set to any position by sending a stream of pulses.

Figure 6-30 displays a servo motor and shows how the length of the pulses determines the position of the servo.

A servo will have three connections: GND, a positive power supply (5 to 6V), and a control connection. The GND connection is usually connected to a brown or black lead, the

FIGURE 6-30 Controlling a servo motor with pulses

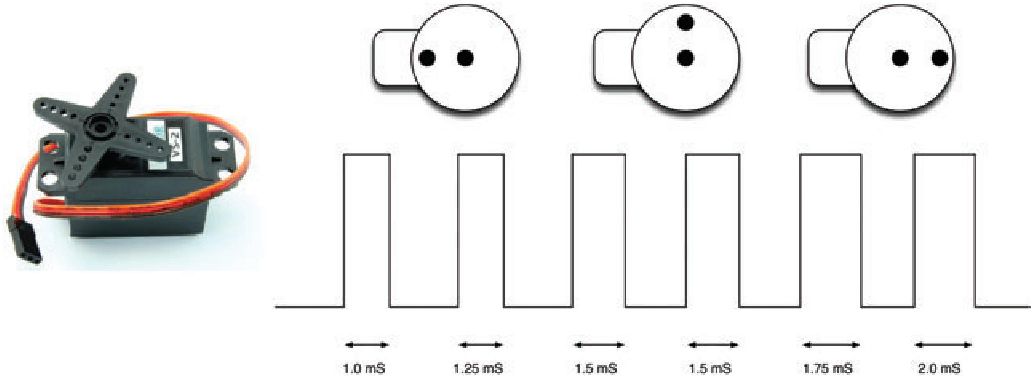

positive connection to a red lead, and the control connection to an orange or yellow lead.

The control connection draws very little current. The server expects to receive a pulse every 20 ms or so. If the pulse is 1.5 ms in duration, then the servo will sit at its middle position. If the pulse is shorter, it will settle in a position to one side, and if the pulse is longer, it will move to a position on the other side of the center position.

You Will Need

To experiment with a servo and Arduino, you will need the following items.

Quantity	Item	Appendix Code
1	Arduino Uno/Leonardo	M2/M21
1	USB lead; Type B for Uno, Micro USB for Leonardo	
1	9g servo motor	H10
1	10kΩ variable resistor	K1, R1
	Jumper wires or solid-core wire	T6

Construction

Figure 6-31 shows a servo connected to an Arduino using a jumper.

Before you power the servo motor from the 5V supply of an Arduino, first check that the Arduino can supply the current requirement. Most small servos will be just fine, such as the tiny 9g servo shown in Figure 6-31.

In Figure 6-31, you can also see the little blue trimpot used to set the position of the servo. This is connected to A1, but uses A0 and A2 to provide GND and +5V to the track ends of the variable resistor.

Figure 6-31 Connecting a servo to an Arduino

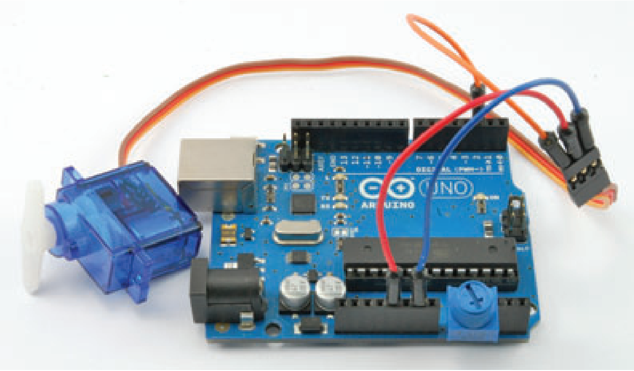

Software

The Arduino has a library specifically designed for generating the pulses that the servo needs. The following example sketch (called "servo") will use this library to set the position of the servo arm to follow the position of the knob on a variable resistor.

```
// servo

#include <Servo.h>

int gndPin = A0;
int plusPin = A2;
int potPin = 1;
int servoControlPin = 2;
```

After defining the pins to be used, the servo library requires the following line of code to set up the servo.

```
Servo servo;
```

The "setup" function sets up the pins and associates the "servo" with the "servoControlPin".

```
void setup()
{
  pinMode(gndPin, OUTPUT);
  digitalWrite(gndPin, LOW);
  pinMode(plusPin, OUTPUT);
  digitalWrite(plusPin, HIGH);
  servo.attach(servoControlPin);
}
```

The "loop" function continuously reads A1 to determine the position of the variable resistor (a number between 0 and 1023) and divides this number by 6 to convert it to an angle between 0 and 170. This is the angle in degrees to which the servo is then set.

```
void loop()
{
  int potPosition = analogRead(potPin);  // 0 to 1023
  int angle = potPosition / 6;           // 0 to 170
  servo.write(angle);
}
```

How to Charlieplex LEDs

An Arduino only has so many IO pins, so when looking to minimize the number of pins used to display a matrix of LEDs, an interesting technique called Charlieplexing can be used. The name comes from the inventor Charlie Allen at the company Maxim, and the technique takes advantage of the feature of Arduino and other microcontroller IO pins that allows them to be changed from outputs to inputs while a sketch is running.

Figure 6-32 shows the arrangement for controlling six LEDs with three pins.

Table 6-2 shows how the pins should be set to light a particular LED.

The number of LEDs that can be controlled per microcontroller pin is given by the following formula:

$$\text{LEDs} = n^2 - n$$

So, if we use four pins, we can have 16 – 4 or 12 LEDs, and 10 pins would give us a massive 90 LEDs, but an awful lot of wiring to do.

There are, however, problems with scaling Charlieplexing up, not the least of which is that the refresh rate needs to be fast enough to fool the eye, and a large number of pins will need a lot of sequence steps to energize all the LEDs that need energizing in a refresh cycle. This will also result in the LEDs becoming dim because their duty cycle will be low. You can compensate for this to some extent by increasing the current through the LEDs, which will cope with fairly large peak currents for a small duration. This does lead to the problem that

LED	Pin 1	Pin 2	Pin 3
A	High	Low	Input
B	Low	High	Input
C	Input	High	Low
D	Input	Low	High
E	High	Input	Low
F	Low	Input	High

TABLE 6-2 Charlieplexing

FIGURE 6-32 Charlieplexing LEDs

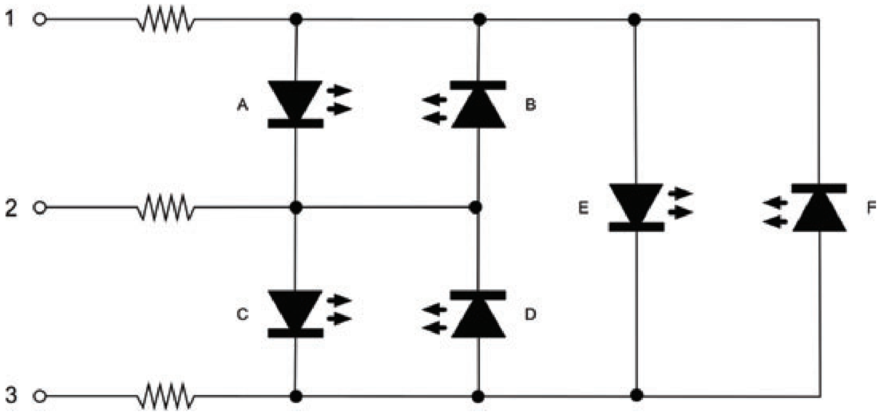

if the microcontroller freezes for some reason, then LEDs could burn out.

You Will Need

To Charlieplex six LEDs, you will need the following items.

Quantity	Item	Appendix Code
1	Arduino Uno/Leonardo	M2/M21
1	USB lead; Type B for Uno, Micro USB for Leonardo	
6	LEDs	S11
3	220Ω resistors	K2
	Jumper wires or solid-core wire	T6

Construction

You will use breadboard to Charlieplex these LEDs (Figure 6-33).

The breadboard layout for this is shown in Figure 6-34. When constructing it, take special care to get the polarity of each LED correct.

The resistors are used to connect Arduino pins D12, D11, and D10 to the breadboard. The resistor leads will stay in the Arduino sockets better if you put a little zigzag kink in the leads using pliers.

The LEDs are very close together, so it is easier to make this with 3mm LEDs. The positive leads of the LEDs (anodes) are shown in red in Figure 6-34.

FIGURE 6-33 Six Charlieplexed LEDs on breadboard

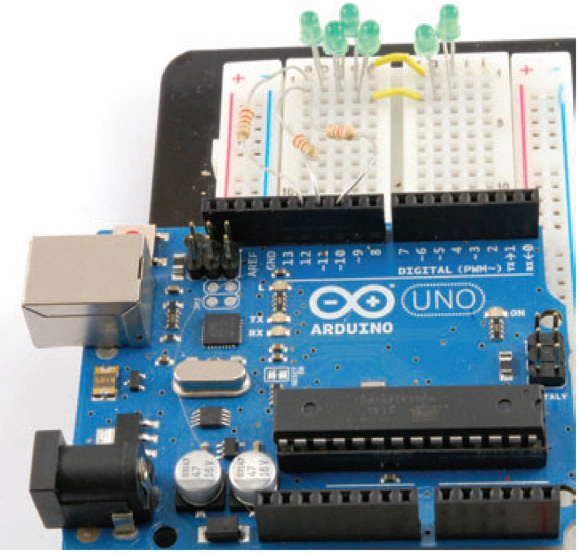

Software

Load the sketch for this ("charlieplexing") onto your Arduino board, and you should see it cycle through the LEDs in the order A to F, as per Figure 6-32.

The sketch first defines three pins to be used as the control pins.

```
// charlieplexing

int pin1 = 12;
int pin2 = 11;
int pin3 = 10;
```

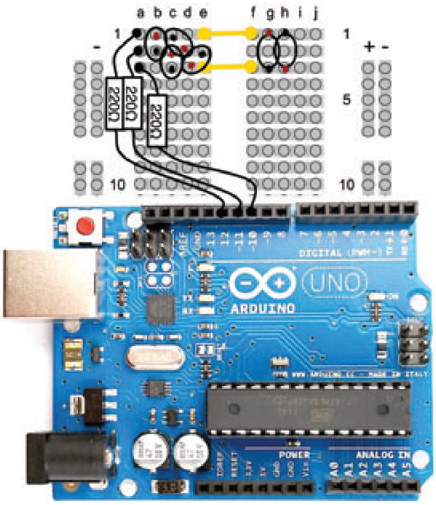

FIGURE 6-34 Breadboard layout for Charlieplexing

The pin states that we need to control each LED, as shown in Table 6-2, are contained in an array called "pinStates". Each element of this array is itself an array of three elements, one for each control pin. A value of 1 indicates that the control pin should be HIGH for that LED, 0 for a LOW, and –1 for an INPUT.

```
int pinStates[][3] = {
  {1, 0, -1}, // A
  {0, 1, -1}, // B
  {-1, 1, 0}, // C
  {-1, 0, 1}, // D
  {1, -1, 0}, // E
  {0, -1, 1}  // F
};
```

As we will be changing the pin mode of the control pins as we go, there is nothing to put in the "setup" function. You do, however, have to have the function there, even if it has nothing in it.

```
void setup()
{
}
```

The loop function steps through each of the LEDs and sets the control pin states according to the values in the appropriate row of the array, using the function "setPins".

```
void loop()
{
  for (int i = 0; i < 6; i++)
  {
    setPins(pinStates[i][0], pinStates[i][1], pinStates[i][2]);
    delay(1000);
  }
}
```

The function “setPins” does not do much except conveniently put the commands to set the state for each control pin into a single convenient line. Most of the logic lives in the “setPin” function that it calls.

```
void setPins(int p1, int p2, int p3)
{
  setPin(pin1, p1);
  setPin(pin2, p2);
  setPin(pin3, p3);
}
```

The “setPin” function sets the state of the pin supplied as its first argument. If the value of the state is –1, then the pin mode for the pin is set to INPUT. Otherwise, it is assumed to be a 1 or a 0 and the pin is set to be an OUTPUT and set to the value supplied with a call to “digitalWrite”.

```
void setPin(int pin, int value)
{
  if (value == -1)
  {
    pinMode(pin, INPUT);
  }
  else
  {
    pinMode(pin, OUTPUT);
    digitalWrite(pin, value);
  }
}
```

How to Type Passwords Automatically

The Arduino Leonardo can be used to impersonate a USB keyboard. Unfortunately, this is not true of the Arduino Uno, so in this section you will need an Arduino Leonardo.

Figure 6-35 shows the device we are going to construct.

All that happens when you press the button is that the Arduino Leonardo pretends to be a keyboard and types the password set in the sketch, wherever the cursor happens to be.

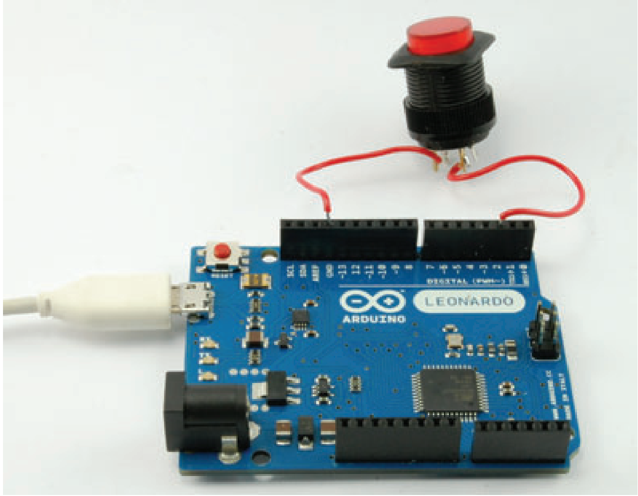

Figure 6-35 Entering passwords automatically with Arduino Leonardo

You Will Need

To build this, you will need the following items.

Quantity	Item	Appendix Code
1	Arduino Leonardo	M21
1	Micro USB lead for the Leonardo	
1	Impressive switch	H15
	Hookup wire	T7

Construction

Solder leads to the switch, and tin the ends so they can be pushed directly into the sockets on the Arduino. One lead from the switch should go to digital pin 2 and the other to GND.

Program the Arduino Leonardo with the sketch "password". Note that when programming the Leonardo, you may have to hold down the reset button until the message "uploading…" appears in the Arduino software.

Software

To use the project, just position your mouse over a password field and press the button. Please note that this project is really just to illustrate what you can do with an Arduino Leonardo. To find your password, all someone would have to do is press the button while in a word processor. So, in terms of security, it is about as secure as writing your password on a sticky note and attaching it to your computer monitor!

The sketch is very simple. The first step is to define a variable to contain your password. You will need to change this to your password. We then define the pin to use for the switch.

```
// password
// Arduino Leonardo Only

char* password = "mysecretpassword";

const int buttonPin = 2;
```

The Leonardo has access to special keyboard and mouse features not available to other types of Arduino. So, in the "setup" function, the Keyboard feature is started with the line "Keyboard.begin()".

```
void setup()
{
  pinMode(buttonPin, INPUT_PULLUP);
  Keyboard.begin();
}
```

In the main loop, the button is checked with a digital read. If the button is pressed, then the Leonardo uses "Keyboard.print" to send the password. It then waits two seconds to prevent the password being sent multiple times.

```
void loop()
{
  if (! digitalRead(buttonPin))
  {
    Keyboard.print(password);
    delay(2000);
  }
}
```

Summary

This chapter should have got you started using the Arduino and given you some food for thought for clever hacks using it. It has, however, only scratched the surface of what is possible with this versatile board.

For more information on programming the Arduino, you may wish to look at some of the author's other books on this topic. *Programming Arduino: Getting Started with Sketches* assumes no prior programming experience and will show you how to program the Arduino from first principals. *30 Arduino Projects for the Evil Genius* is a project-based book that explains both the hardware and programming side of Arduino, and is illustrated with example projects, nearly all of which are built on breadboard.

The official Arduino web site, www.arduino.cc, has a wealth of information on using the Arduino, as well as the official documentation for the Arduino commands and libraries.

7

Hacking with Modules

There are many modules available that provide a great shortcut when hacking together a project. These modules are usually a tiny PCB with a few components on them and some convenient connection points. They make it very easy to use some surface-mounted ICs that would otherwise be very difficult to solder connections to. Many of these modules are designed to be used with microcontrollers like the Arduino.

In this chapter, you will explore some of the more fun and useful modules available from suppliers like SparkFun and Adafruit, most of whose modules are also open-source hardware. So you'll get to see the schematics for them and even make your own modules using the design if you wish.

Access to the schematics and data sheets is very useful when trying to use a module. There are a few important things you need to know about any module before you use it:

- What is the range of supply voltage?
- How much current does it consume?
- How much current can any outputs supply?

How to Use a PIR Motion Sensor Module

PIR motion sensors are used in intruder alarms and for automatic security alarms. They detect movement using infrared light. They are also cheap and easy to use.

In this example, you will first experiment with a PIR module using it to light an LED, and then look at how it could be hooked up to an Arduino to send a warning message to the Serial Console.

You Will Need (PIR and LED)

Quantity	Names	Item	Appendix Code
1		PIR module (5–9V)	M5
1	D1	LED	K1
1	R1	470Ω resistor	K2
1		Solderless breadboard	T5
		Solid-core jumper wire	T6
1		4 × AA battery holder	H1
1		4 × AA batteries	
1		Battery clip	H2

Breadboard

Figure 7-1 shows the schematic diagram for this experiment.

Looking at the datasheet for this particular module, the supply voltage range is 5V to 7V, so it will work just fine with our four AA batteries.

The module is very easy to use. You just supply it with power and its output goes high (to supply voltage) when movement is detected and then back low again after a second or two.

The datasheet also says that the output can supply up to 10mA. That isn't a great deal, but is enough to light an LED. By choosing a 470Ω resistor, we will be limiting the current to:

I = V / R = (6V – 2V) / 470Ω = 4 / 470 = 8.5mA

Figure 7-2 shows the breadboard layout, while Figure 7-3 offers a photograph of the actual breadboard.

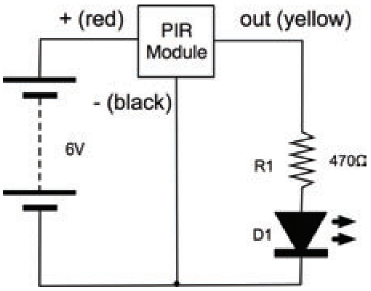

Figure 7-1 Schematic diagram—using a PIR module with an LED

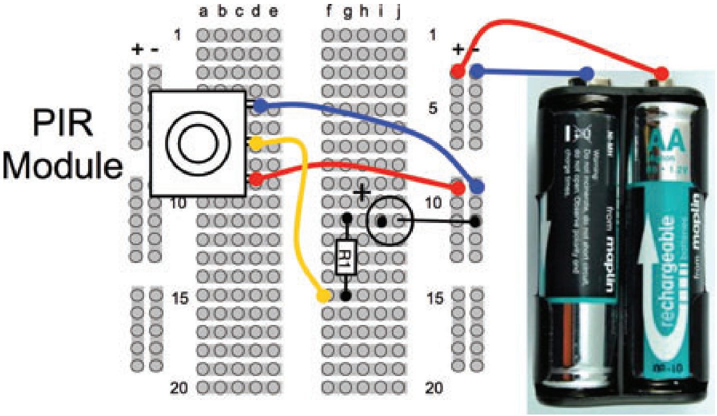

Figure 7-2 Breadboard layout—using a PIR module with an LED

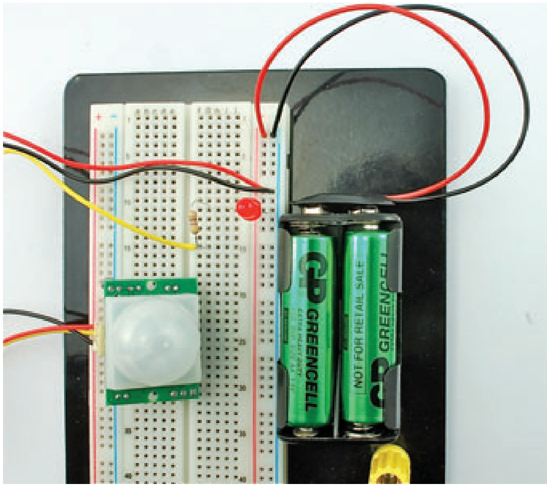

FIGURE 7-3 Using a PIR module with an LED

The PIR module has three pins labeled +5V, GND, and OUT. The supplied connector lead has red, black, and yellow leads. Hook it up so the red lead connects to the connection labeled +5V.

When it's powered up, the LED will light every time movement is detected.

Having already discussed the PIR sensor so we know what to expect of it, it's time to interface it with an Arduino.

You Will Need (PIR and Arduino)

To interface the PIR sensor with an Arduino, you really only need the PIR sensor and an Arduino.

Quantity	Item	Appendix Code
1	PIR module (5–9V)	M5
1	Arduino Uno/Leonardo	M2/M21
1	USB lead; Type B for Uno, Micro USB for Leonardo	

Construction

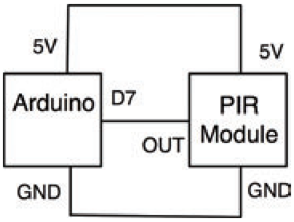

FIGURE 7-4 Schematic diagram for the Arduino and PIR sensor

Figure 7-4 shows the schematic diagram for this, while Figure 7-5 shows how the PIR module is wired to the module. To get the wires to stay in the Arduino sockets, it helps to put a little zigzag bend in the tinned end of the wire lead.

Before you move onto the next stage of programming the Arduino, temporarily remove the OUT lead from its Arduino socket. The reason for this is that you do not know what sketch was last running on the Arduino. It might have been something where pin 7 was an output, and if it was, this could easily damage the output electronics of the PIR sensor.

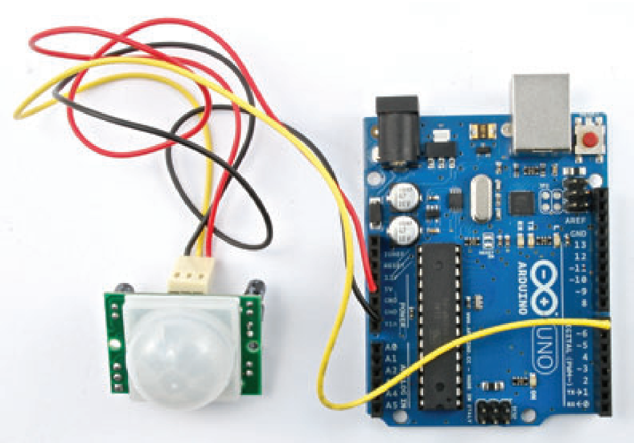

FIGURE 7-5 The Arduino and the PIR sensor

Software

Load the sketch "pir_warning" into the Arduino IDE and onto the Arduino board, and then plug the yellow "OUT" lead back into pin 7 on the Arduino.

When you launch the Serial Monitor (Figure 7-6), you will see an event appear every time movement is detected. Imagine leaving this running while away from your computer—to detect snoopers!

The sketch is very straightforward.

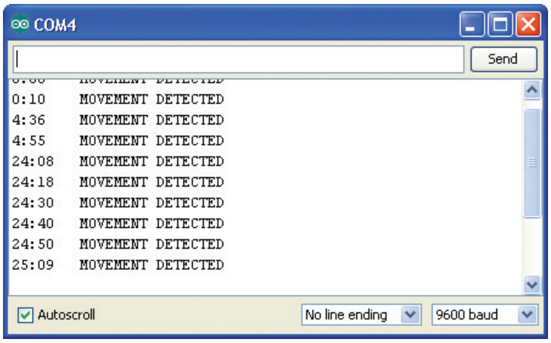

```
0:10     MOVEMENT DETECTED
4:36     MOVEMENT DETECTED
4:55     MOVEMENT DETECTED
24:08    MOVEMENT DETECTED
24:18    MOVEMENT DETECTED
24:30    MOVEMENT DETECTED
24:40    MOVEMENT DETECTED
24:50    MOVEMENT DETECTED
25:09    MOVEMENT DETECTED
```

FIGURE 7-6 The Serial Monitor showing intruder alerts

```
// pir_warning

int pirPin = 7;

void setup()
{
  pinMode(pirPin, INPUT);
  Serial.begin(9600);
}

void loop()
{
  if (digitalRead(pirPin))
  {
    int totalSeconds = millis() / 1000;
```

```
      int seconds = totalSeconds % 60;
      int mins = totalSeconds / 60;
      Serial.print(mins);
      Serial.print(":");
      if (seconds < 10) Serial.print("0");
      Serial.print(seconds);
      Serial.println("\tMOVEMENT DETECTED");
      delay(10000);
    }
}
```

The only part of the code that is a bit different than the other sketches we have seen deals with displaying an elapsed time in minutes and seconds next to each event.

This code uses the Arduino "millis" function, which returns the number of milliseconds since the Arduino was last reset. This is then separated into its minute and second components and the various parts printed out as a message. The last part to be displayed uses the "println" command that adds a line feed to the end of the text so the next text starts on a new line.

The special character "\t" in this "println" is a tab character, to line the output up neatly.

How to Use Ultrasonic Rangefinder Modules

Ultrasonic rangefinders use ultrasound (higher frequency than the human ear can hear) to measure the distance to a sound-reflective object. They measure the time it takes for a pulse of sound to travel to the object and back. Figure 7-7 shows two different types of sonar. On the left is a low-cost sonar module (less than USD 5) with separate ultrasonic transducers for

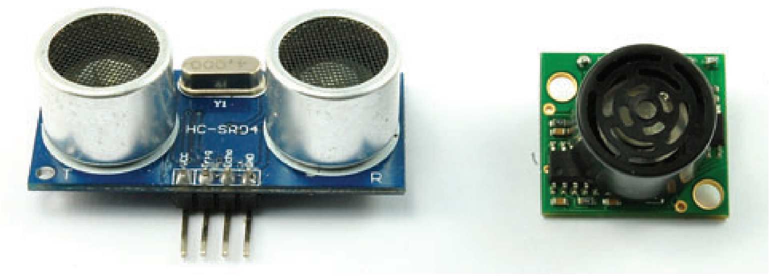

FIGURE 7-7 Ultrasonic rangefinders

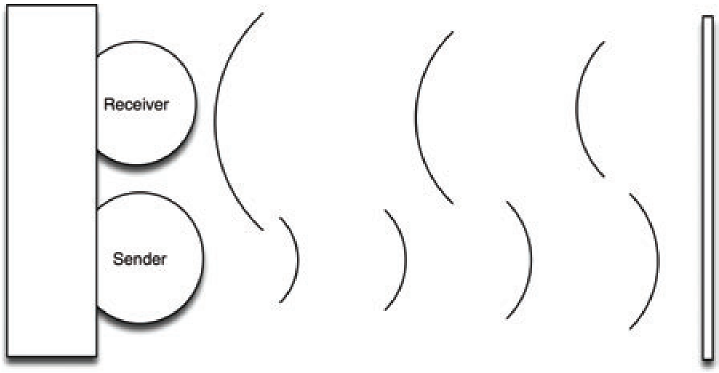

Figure 7-8 Ultrasonic range finding

sending the pulse and receiving the echo, and a much more expensive (around USD 25) but highly specified module made by MaxBotix Inc.

To see how to use these modules with an Arduino, we will try out each in turn.

Ultrasonic range finding works the same as sonar used by ships and submarines. A sound wave is sent out from a sender, hits an object, and bounces back. Since we know the speed of sound, the distance to the sound-reflecting object can be calculated from the time it takes for the sound to come back to the receiver (Figure 7-8).

The sound used is at a high frequency—hence, it is called ultrasonic. Most units operate at a frequency of about 40 kHz. Not many people can hear sounds above 20 kHz.

You Will Need

To try out both rangefinders, you will need the following items.

Quantity	Item	Appendix Code
1	Arduino Uno/Leonardo	M2/M21
1	USB Lead; Type B for Uno, Micro USB for Leonardo	
1	MaxBotix LV-EZ1 rangefinder	M6
1	HC-SR04 rangefinder	M7
1	Solderless breadboard	T5
	Solid-core jumper wire	T6

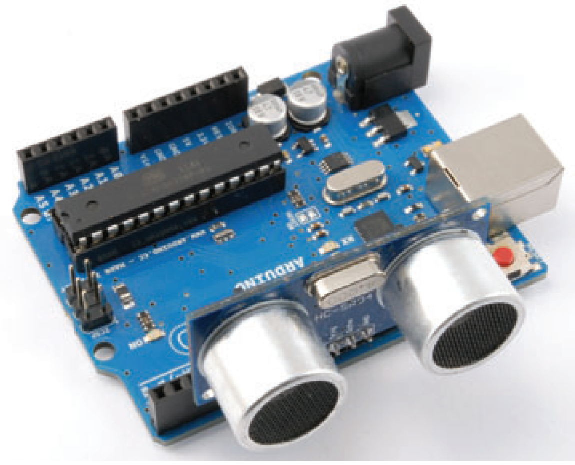

Figure 7-9 An HC-SR04 rangefinder on an Arduino

The HC-SR04 Rangefinder

These modules make the Arduino do a lot of the work, which is one reason why they are so much cheaper than the MaxBotix modules. They do, however, have the advantage that they can just fit into the side connector of an Arduino if you can spare the pins to supply them with current using two output pins (Figure 7-9).

Load the sketch "range_finder_budget" onto the Arduino and then plug the rangefinder module into the Arduino, as shown in Figure 7-9.

When you open the Serial Monitor, you will see a stream of distances in inches appear (Figure 7-10). Try pointing the rangefinder in different directions—say, a wall a few feet away—and confirm that the reading is reasonably accurate with a tape measure.

The Arduino code for measuring the range is all contained within the "takeSounding_cm" function. This sends a single 10-microsecond pulse to the "trigger" pin of the ultrasonic module, which then uses the built-in Arduino function "pulseIn" to measure the time period before the echo pin goes high.

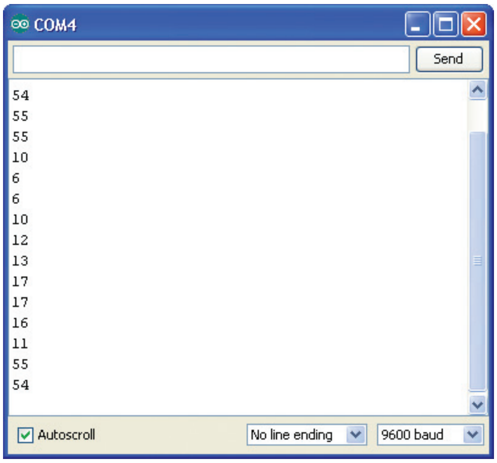

```
54
55
55
10
6
6
10
12
13
17
17
16
11
55
54
```

☑ Autoscroll

No line ending

9600 baud

Figure 7-10 Distance readings in the Serial Monitor

```
// range_finder_budget

int trigPin = 9;
int echoPin = 10;
int gndPin = 11;
int plusPin = 8;

int lastDistance = 0;

void setup()
{
  Serial.begin(9600);
  pinMode(trigPin, OUTPUT);
  pinMode(echoPin, INPUT);
  pinMode(gndPin, OUTPUT);
  digitalWrite(gndPin, LOW);
  pinMode(plusPin, OUTPUT);
  digitalWrite(plusPin, HIGH);
}
```

```
void loop()
{
  Serial.println(takeSounding_in());
  delay(500);
}

int takeSounding_cm()
{
  digitalWrite(trigPin, LOW);
  delayMicroseconds(2);
  digitalWrite(trigPin, HIGH);
  delayMicroseconds(10);
  digitalWrite(trigPin, LOW);
  delayMicroseconds(2);
  int duration = pulseIn(echoPin, HIGH);
  int distance = duration / 29 / 2;
  if (distance > 500)
  {
    return lastDistance;
  }
  else
  {
    lastDistance = distance;
    return distance;
  }
}

int takeSounding_in()
{
  return takeSounding_cm() * 2 / 5;
}
```

We then need to convert that time in milliseconds into a distance in centimeters. If there is no reflection because there is no object that is close enough, or the object is reflecting the sound wave away rather than letting it bounce back to the receiver, then the time of the pulse will be very large and so the distance will also be recorded as very large.

To filter out these long readings, we disregard any measurement that is greater than 5m, returning that last sensible reading we got.

The speed of sound is roughly 343 m/s in dry air at 20 degrees C, or 34,300 cm/s.

Or, put another way, 34,300 / 1,000,000 cm / microsecond.

That is, 0.0343 cm/microsecond.

Put another way, 1/0.0343 microseconds/cm.

Or, 29.15 microseconds/cm.

Thus, a time of 291.5 microseconds would indicate a distance of 10 cm.

The "takeSounding_cm" function approximates 29.15 to 29 and then also divides the answer by 2, as we don't want the distance of the whole return journey, just the distance to the subject.

In actual fact, many factors affect the speed of sound, so this approach will only ever give an approximate answer. The temperature and the humidity of the air will both affect the measurement.

The MaxBotix LV-EZ1 Rangefinder

The HC-SR04 rangefinder only has a single type of interface, and we have to tell it to generate the sonar pulse, and then time how long it takes to come back ourselves.

In contrast, the MaxBotix device does all of this for us, and what's more it provides us with no less than three ways to get the distance readings:

- Serial data readings
- Analog (Vcc / 512) / inch
- Pulse width (147 µS/inch)

FIGURE 7-11 The MaxBotix rangefinder and Arduino

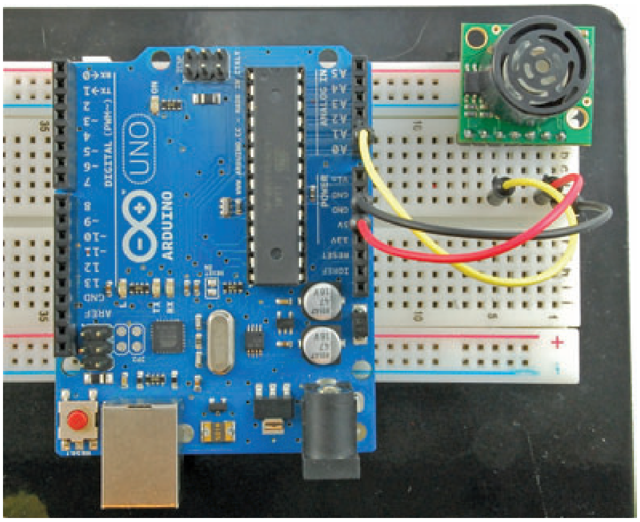

We will use the analog method to test out this device. The figure of Vcc / 512 per inch means that the analog output will be the supply voltage divided by 512 per inch. So if the object was 10 inches away, then the analog output voltage would be:

10 inches × 5V/ 512 = 0.098V

The MaxBotix module has too many pins to easily fit directly onto the Arduino connectors, so you will need to use breadboard.

Figure 7-11 shows the unit and the Arduino on the breadboard, while Figure 7-12 shows the breadboard layout.

Load up the sketch "range_finder_maxsonar" and then connect up the module as shown in Figure 7-11.

The sketch is much simpler than for the other module, and the distance in inches is just the raw analog reading (between 0 and 1023) divided by two.

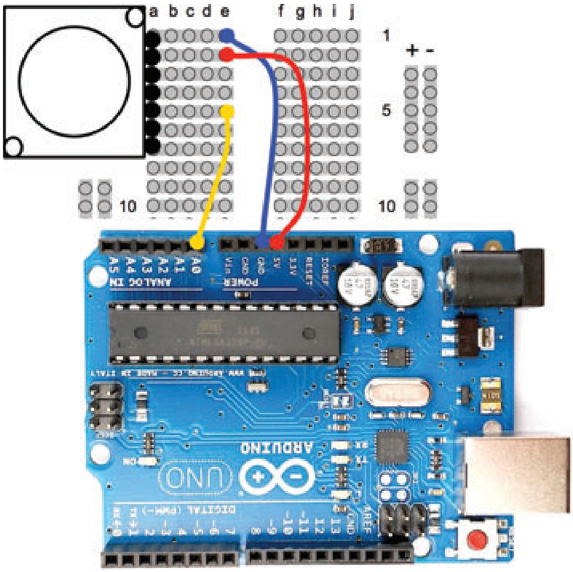

FIGURE 7-12 The MaxBotix rangefinder and Arduino breadboard layout

```
// range_finder_maxsonar

int readingPin = 0;

int lastDistance = 0;

void setup()
{
  Serial.begin(9600);
}

void loop()
{
  Serial.println(takeSounding_in());
  delay(500);
}

int takeSounding_in()
{
  int rawReading = analogRead(readingPin);
  return rawReading / 2;
}

int takeSounding_cm()
{
  return takeSounding_cm() * 5 / 2;
}
```

Opening the Serial Monitor will produce the same stream of distance measurements as the other module.

Note that both sketches have a metric and imperial distance measurement flavor for your convenience.

How to Use a Wireless Remote Module

Radio frequency circuits usually are not worth making yourself when extremely useful modules like the one shown in Figure 7-13 are readily available for just a few dollars.

The module shown can be easily found on eBay and has a handy little key-fob sized remote with four buttons on it. These buttons can toggle four digital pins on and off on the corresponding receiver module.

It is worth noting that modules like this are also available with relays instead of digital outputs, making it very easy to hack your own remote control projects.

You will first experiment with the module on breadboard, just turning on an LED, and then in the following section, you can try connecting it to an Arduino.

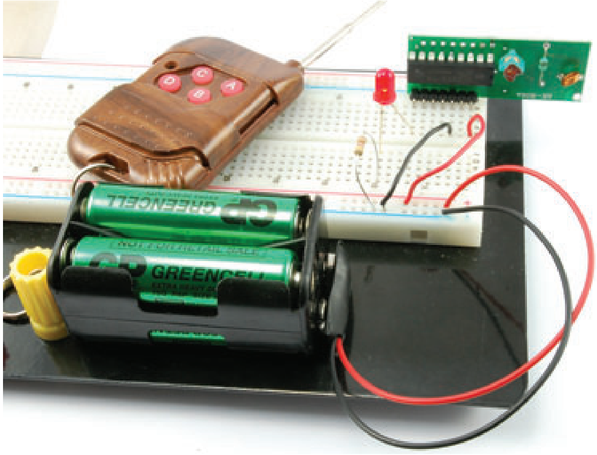

FIGURE 7-13 An RF module on breadboard

You Will Need

To try out the wireless remote on breadboard, you will need the following items.

Quantity	Names	Item	Appendix Code
1		Solderless breadboard	T5
		Solid-core jumper wire	T6
1		Wireless remote kit	M8
1	D1	LED	K1
1	R1	470Ω resistor	K2
1		4 ×AA battery holder	H1
1		Battery clip	H2
4		AA batteries	

Breadboard

Figure 7-14 shows the breadboard layout used to test the remote. You could, if you wished, add three more LEDs so there was one for each channel.

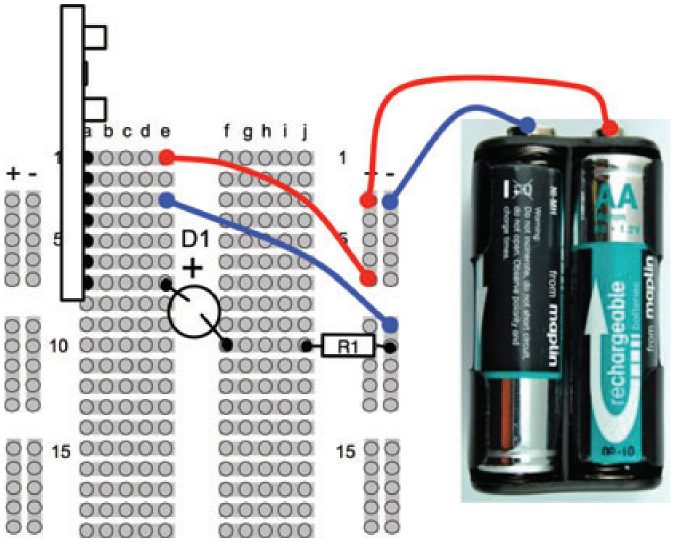

Figure 7-14 Breadboard layout for testing the RF module

The datasheet for this module shows that the pins are as shown in Table 7-1.

Put the module on the breadboard with pin 1 at the top of the breadboard, and wire it up as shown in Figure 7-14.

That really is all there is to it. Pressing button A should toggle the LED on and off. If you wanted to, you could add more LEDs so there was one for each channel, or try moving the LED to a different output to check that they all work.

Pin Number	Pin Name	Purpose
1	Vcc	Positive supply 4.5 to 7V
2	VT	Switch voltage—no connection needed
3	GND	Ground
4	D3	Digital output 3
5	D2	Digital output 2
6	D1	Digital output 1
7	D0	Digital output 0

TABLE 7-1 RF Receiver Pinout

How to Use a Wireless Remote Module with Arduino

If we are prepared to lose one of the four channels of the remote from the section "How to Use a Wireless Remote Module," then we can plug the receiver straight into the Arduino socket A0 to A5 (see Figure 7-15).

You Will Need

To try out the wireless remote with an Arduino, you will need the following items.

Quantity	Item	Appendix Code
1	Arduino Uno/Leonardo	M2/M21
1	USB lead; Type B for Uno, Micro USB for Leonardo	
1	Wireless Remote Kit	M8

Before plugging the remote receiver into the Arduino, upload the sketch "rf_remote".

Software

With the software uploaded and the RF receiver attached, when you open the Serial Monitor you should see something like Figure 7-16.

Figure 7-15 Using a RF remote with an Arduino

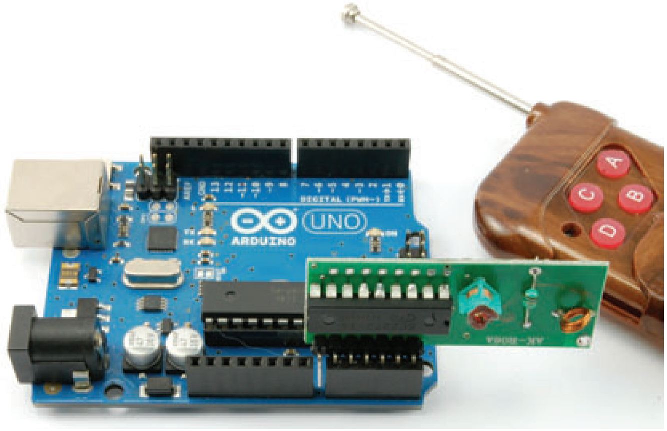

The sketch displays as a 1 or 0 the current state of the remote control channels. So button A will not do anything (that is the button we sacrificed), but pressing the other buttons should toggle the appropriate column between 0 and 1.

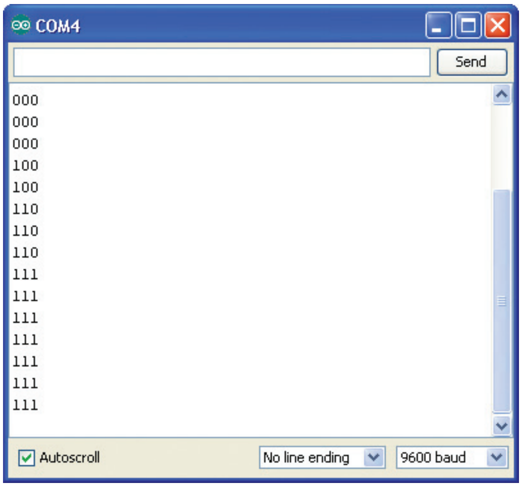

```
000
000
000
100
100
110
110
110
111
111
111
111
111
111
111
```

☑ Autoscroll

No line ending

9600 baud

FIGURE 7-16 Remote control messages to your computer

```
// rf_remote

int gndPin = A3;
int plusPin = A5;
int bPin = A2;
int cPin = A1;
int dPin = A0;

void setup()
{
  pinMode(gndPin, OUTPUT);
  digitalWrite(gndPin, LOW);
  pinMode(plusPin, OUTPUT);
  digitalWrite(plusPin, HIGH);
  pinMode(bPin, INPUT);
  pinMode(cPin, INPUT);
  pinMode(dPin, INPUT);
  Serial.begin(9600);
}

void loop()
{
  Serial.print(digitalRead(bPin));
  Serial.print(digitalRead(cPin));
  Serial.println(digitalRead(dPin));
  delay(500);
}
```

The RF receiver uses very little current, so there is no problem powering it from a digital output. In fact, doing so has the added benefit that we can actually turn it off to save power simply by setting the “plusPin” low.

How to Control Motor Speed with a Power MOSFET

This section is a little out of place because MOSFET transistors are not modules. However, this section does lead in nicely to the next section on using motor controller modules.

We first used power MOSFETs in Chapter 3. They are a kind of transistor that is particularly suited to switching high-current loads efficiently. By efficiently, I mean that they run pretty cool, working very well as electronic switches. They have a very low "on" resistance and a very high "off resistance."

Back in Chapter 6, you used a technique called PWM (pulse-width modulation) to control the brightness of an LED by varying the length of pulses. You can use exactly the same trick on a DC motor. However, unlike an LED, motors use too much current to be driven directly from an Arduino output, so you will use a MOSFET controlled by the Arduino.

You Will Need

To build this, you will need the following items.

Quantity	Name	Item	Appendix Code
1		Solderless breadboard	T5
		Solid-core jumper wire	T6
1		4 × AA battery holder	H1
1		4 × AA batteries	
1		Battery clip	H2
1	R1	10kΩ trimpot	K1
1	R2	1kΩ resistor	K2
1	T1	FQP30N06 MOSFET	S6
1		6V DC motor or gear motor	H6
1		Arduino Uno/Leonardo	M2/M21
1		USB lead; Type B for Uno, Micro USB for Leonardo	

The DC motor can be any small motor you can find that is around 6V.

Breadboard

Figure 7-17 shows the schematic diagram.

Notice that we actually have two sources of power here. We have the Arduino, which will be getting its power from your computer's USB port and a separate battery that supplied the power to the MOSFET. This is quite a common arrangement, because the Arduino's 5V output is not really suitable for high-current loads like a motor. Indeed, motors can create all sorts of problems for delicate electronics, so it is best not to power them from the Arduino.

There is less of a problem if the Arduino and motor share a power supply. For example, a 9V battery provides power to the Arduino through its power jack, and at the same time provides the positive supply to the motor.

I have included a resistor R2 between the Arduino output pin and the MOSFET. The circuit would work fine just connecting D5 directly to the gate; however, the gate acts like a capacitor, which means that when switched at very high speed it can actually cause quite a lot of current to flow from the digital output. This will not be a problem at the relatively slow PWM speeds used by the Arduino, but it is considered "good practice" to use a resistor here.

Figures 7-18 and 7-19 show the actual circuit and the breadboard layouts, respectively.

Figure 7-17 Schematic diagram for the MOSFET motor control

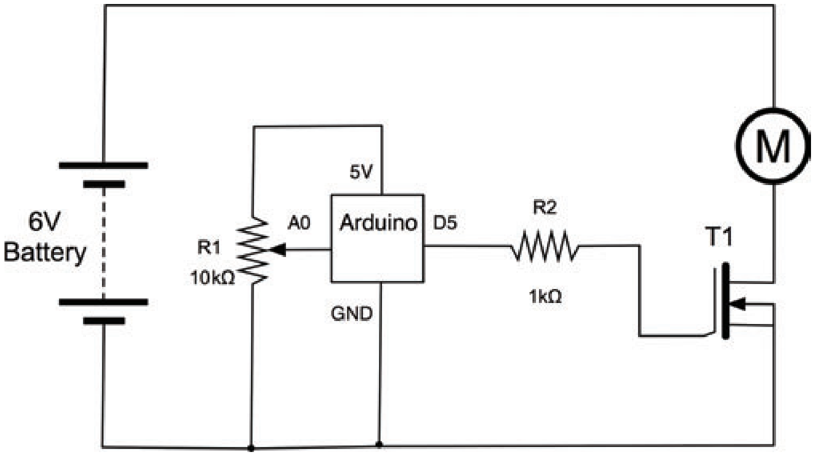

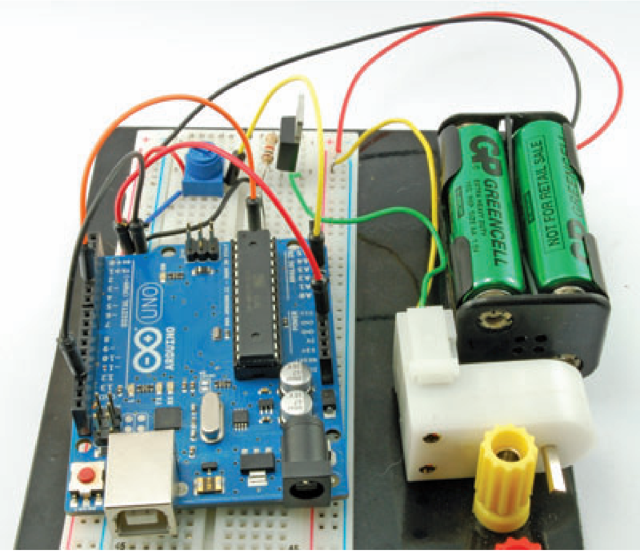

Figure 7-18 The MOSFET motor control

Software

Load the sketch "mosfet_motor_speed" onto the Arduino and connect the battery. You should find that by turning the variable resistor you have much finer control of the motor's speed than you did way back in Chapter 3 when you were just controlling the gate voltage of the MOSFET.

The sketch is very similar to the sketch we used to control the brightness of an LED from an Arduino in Chapter 6.

```
// mosfet_motor_speed
int voltsInPin = 0;
int motorPin = 5;

void setup()
{
  pinMode(motorPin, OUTPUT);
}

void loop()
{
  int rawReading = analogRead(voltsInPin);
  int power = rawReading / 4;
  analogWrite(motorPin, power);
}
```

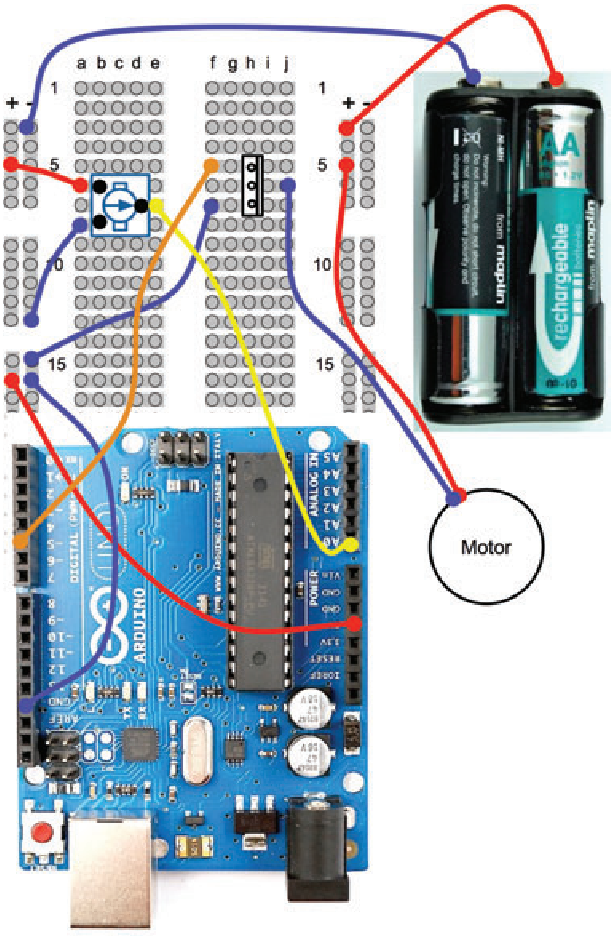

Figure 7-19 Breadboard layout for the MOSFET motor control

In the “loop” function, the raw reading of between 0 and 1023 from the analog input is divided by 4 to give us a number between 0 and 255 that is suitable for use with “analogWrite”.

How to Control DC Motors with an H-Bridge Module

In the earlier section of this chapter, “How to Control Motor Speed with a Power MOSFET,” we saw how you can use a MOSFET to control the speed of a motor. This is fine as long as you always want the motor to turn in the same direction. If you want to be able to reverse the direction of the motor, you need to use something called an H-Bridge.

To change the direction in which a motor turns, you have to reverse the direction in which the current flows. To do this requires four switches or transistors. Figure 7-20 shows how this works, using switches in an arrangement. You can now see why it is called an H-Bridge.

In Figure 7-20, S1 and S4 are closed, while S2 and S3 are open. This allows current to flow through the motor with terminal "A" positive and terminal "B" negative. If we were to reverse the switches so that S2 and S3 are closed and S1 and S4 are open, then "B" will be positive and "A" will be negative, and the motor will turn in the opposite direction.

You may, however, have spotted a danger with this circuit. That is, if by some chance S1 and S2 are both closed, then the positive supply will be directly connected to the negative supply and we will have a short circuit. The same is true if S3 and S4 are both closed at the same time.

You can build an H-Bridge yourself using transistors, and Figure 7-21 shows a typical H-Bridge schematic.

This schematic requires some six transistors and a good few other components. If you wanted to control two motors, you would need some 12 transistors, which causes everything to become quite complicated.

Fortunately, help is on hand as there are several H-Bridge ICs available that usually have two H-Bridges on a single chip and make controlling motors very easy. One such chip is

FIGURE 7-20 An H-Bridge using switches

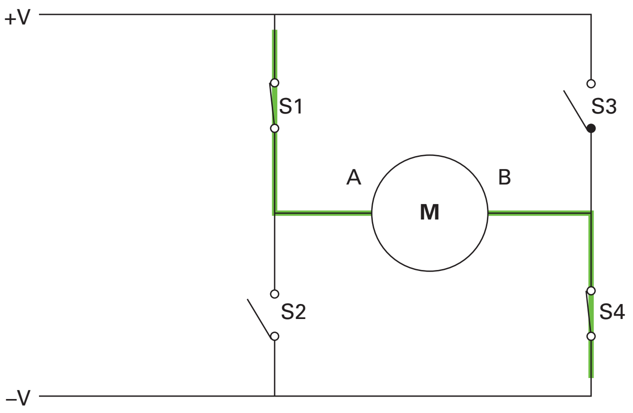

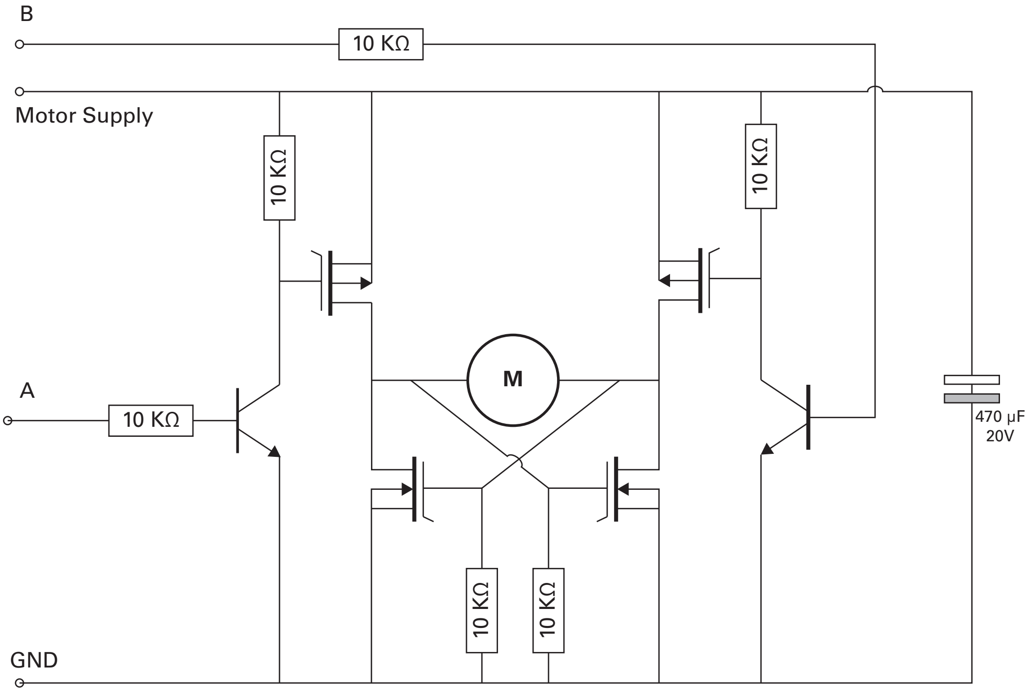
B
10 KΩ
Motor Supply
10 KΩ
10 KΩ
M
A
10 KΩ
470 μF
20V
10 KΩ
10 KΩ
GND

Figure 7-21 An example schematic for an H-Bridge

available as a module from SparkFun (Figure 7-22). You will find similar modules available from other module suppliers.

Figure 7-22 actually shows two of these modules so you can see both sides. The modules are supplied without connectors and the module on the left has pin headers soldered to it. This makes it very easy to use with breadboard.

Table 7-2 shows the pins of this module and explains the purpose of each. The module has two motor channels called A

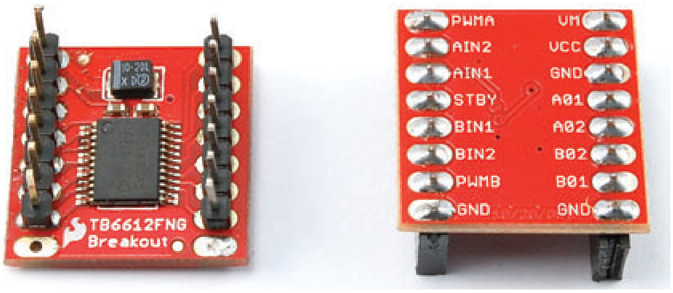
TB6612FNG
Breakout
PWMA
AIN2
AIN1
STBY
BIN1
BIN2
PWMB
GND
VM
VCC
GND
A01
A02
B02
B01
GND

Figure 7-22 A SparkFun H-Bridge module

Pin Name	Purpose	Purpose	Pin Name
PWMA	PWM input for Channel A	Motor supply voltage (VCC to 15V)	VM
AIN2	Control input 2 for A; high for counter-clockwise	Logic supply (2.7 to 5.5V); only requires 2mA	VCC
AIN1	Control input 1 for A; high for clockwise		GND
STBY	To connect to GND to put the device into “standby” mode.	Motor A connection 1	A01
BIN1	Control input 1 for B; high for clockwise	Motor A connection 2	A02
BIN2	Control input 2 for B; high for counterclockwise	Motor B connection 2	B02
PWMB	PWM input for Channel A	Motor B connection 1	B01
GND			GND

TABLE 7-2 The SparkFun TB6612FNG Breakout Board Pinout

and B and can drive motors with a current of 1.2A per channel with peak currents of over twice that.

We will experiment with this module using just one of its two H-Bridge channels (Figure 7-23).

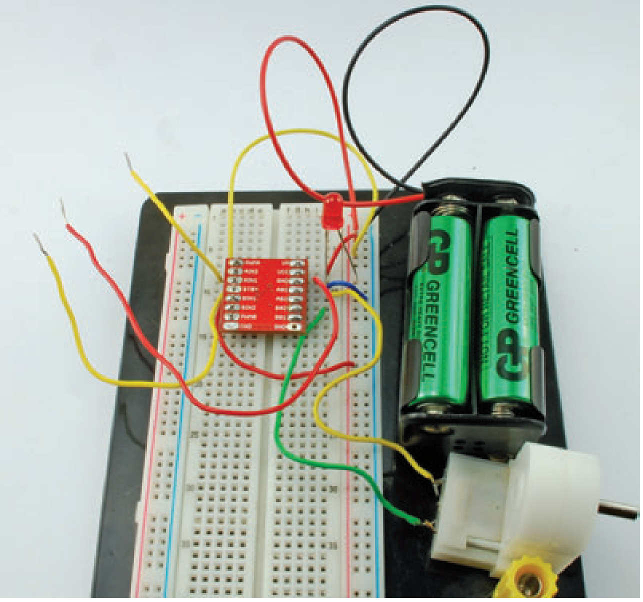

FIGURE 7-23 Experimenting with the SparkFun TB6612FNG breakout board

You Will Need

To build this, you will need the following items.

Quantity	Name	Item	Appendix Code
1		Solderless breadboard	T5
		Solid-core jumper wire	T6
1		4 × AA battery holder	H1
1		4 × AA batteries	
1		Battery clip	H2
1		LED	K1
1		SparkFun TB6612FNG Breakout Board	M9
1		6V DC motor or gear motor	H6
1		Header pins	K1, H4

The DC motor can be any small motor around 6V.

Breadboard

Before fitting the module onto the breadboard, you need to solder the header pins into place as shown in Figure 7-22. We won't use the bottom two GND connections, so you can just solder the top seven pins on each side.

Figure 7-24 shows the schematic diagram for the experiment, while Figure 7-25 displays the breadboard layout.

FIGURE 7-24 Schematic diagram for H-Bridge experiment

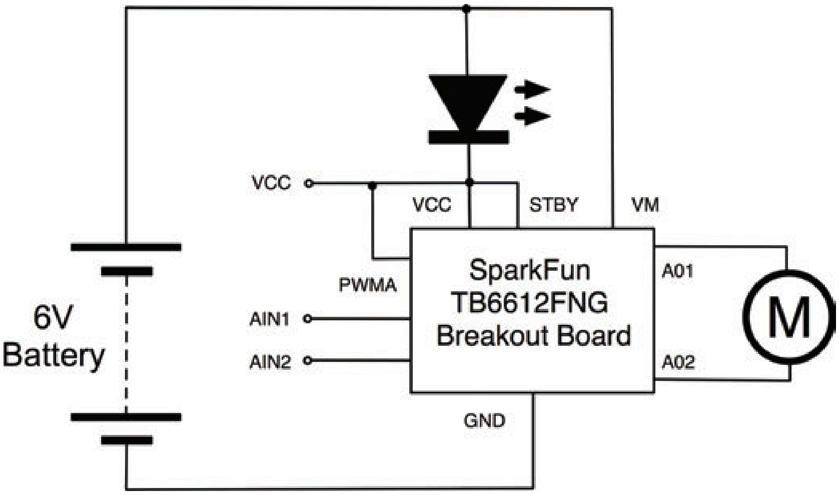

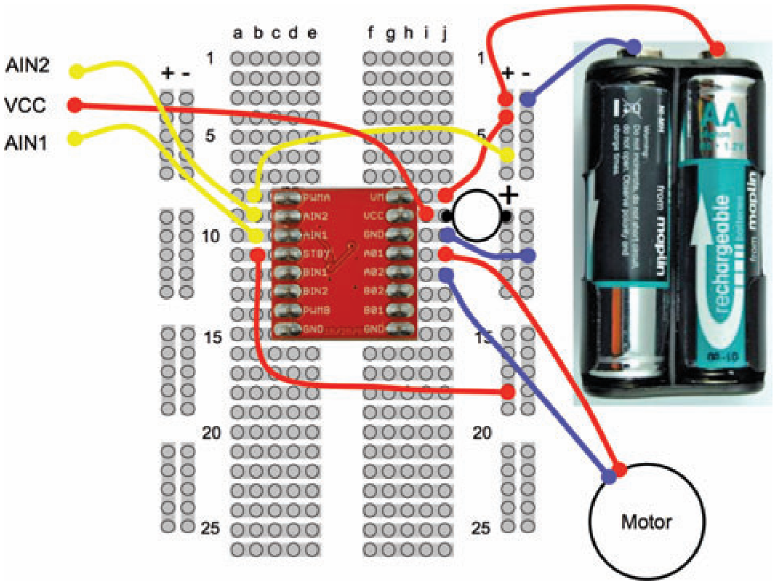
AIN2
VCC
AIN1
a b c d e
f g h i j
+ -
+ -
1
5
10
15
20
25
PWMA
AIN2
AIN1
STBY
BIN1
BIN2
PWMB
GND
VM
VCC
GND
A01
A02
B02
B01
GND
+
rechargeable
AA
from maplin
Motor

Figure 7-25 Breadboard layout for the H-Bridge experiment

The 6V battery pack is actually a slightly higher voltage than is (strictly speaking) allowed for VCC on the module. You would probably get away with the extra half volt above the nominal maximum voltage of 5.5V, but to play it safe, we can use an LED to drop 2V, so that VCC will be around 4V, which is well within its range.

This is a useful trick, but only use it when the current flowing is less than the maximum forward current of the LED. In fact, in this experiment, the current required for VCC is not even enough to make the LED glow.

The PWMA pin is connected to VCC, which simulates the PWN control signal being on all the time—in other words, there is full power to the motor.

Next, put everything on the breadboard as shown in Figure 7-25.

Using the Control Pins

Three of the leads from the breadboard do not actually go anywhere. You will control the motor, touching the red lead going to VCC to AIN1, and then to AIN2, in turn. Note how the motor turns first in one direction and then the other.

You might be wondering why there are two control pins, as well as the PWM pin for each motor channel. In theory, you could have one direction pin and one PWM pin, and if the PWM power was zero, then the motor would not turn at all.

The reason we have three pins to control each motor (PWM, IN1, and IN2) rather than just two is that if both IN1 and IN2 are high (connected to VCC), then the H-Bridge operates in a "braking" mode, which provides electrical braking of the motor, slowing it down. This feature is not often used, but can be useful if you want to stop the motor quickly.

How to Control a Stepper Motor with an H-Bridge Module

Normal DC motors are nice and easy to use. There are just two connections to make and if the voltage is applied one way it turns clockwise; reverse the polarity and it turns counterclockwise. The down side to normal DC motors is that if you want to know what position it has turned to, you have to use some kind of sensor.

Stepper motors are entirely different kinds of motors. They commonly have four connections. Figure 7-26 shows how a stepper motor works. Or more specifically, a bipolar stepper motor, which is the one we will try out.

The motor contains a toothed rotor where each of the teeth of the rotor are magnets, of alternating north and south poles.

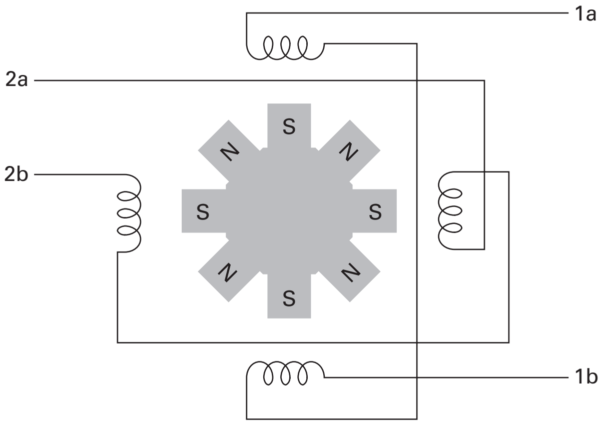

FIGURE 7-26 How a bipolar stepper motor works

Four coils, acting as electromagnets will, when energized in the right order, move the rotor around one step. The coils are arranged in pairs, wired so that as one pushes, its opposite number pulls.

Most stepper motors will have far more steps than the eight shown in Figure 7-26, sometimes 200 or more. This makes the motors very flexible, because they can run freely just like any other motor, by sending the stepping pulses quickly, or they can be controlled very precisely by just moving them forward one step at a time. For this reason, you will find stepper motors in inkjet printers and also 3D printers.

Because the stepper motor will only turn, if we generate a series of pulses in the right order, and need to be able to reverse the direction of current flow in the coils, we can use an Arduino to generate the control signals and an H-Bridge module to supply the power to the coils (Figure 7-27).

Figure 7-28 shows the schematic diagram for this arrangement.

Identifying which lead is which on a stepper motor sometimes requires a bit of trial and error. Using a multimeter, you can measure the resistance between pairs of leads and therefore work out which leads are connected to the same coil.

Another way of finding the leads that belong to the same coil is to hold two of the leads together and see if it makes the shaft of the motor more difficult to turn. Strange, but true!

FIGURE 7-27 Controlling a stepper motor with an Arduino and an H-Bridge

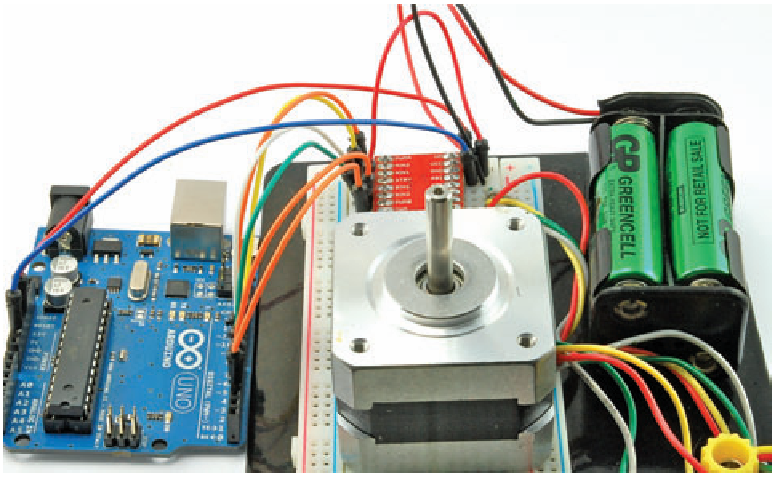

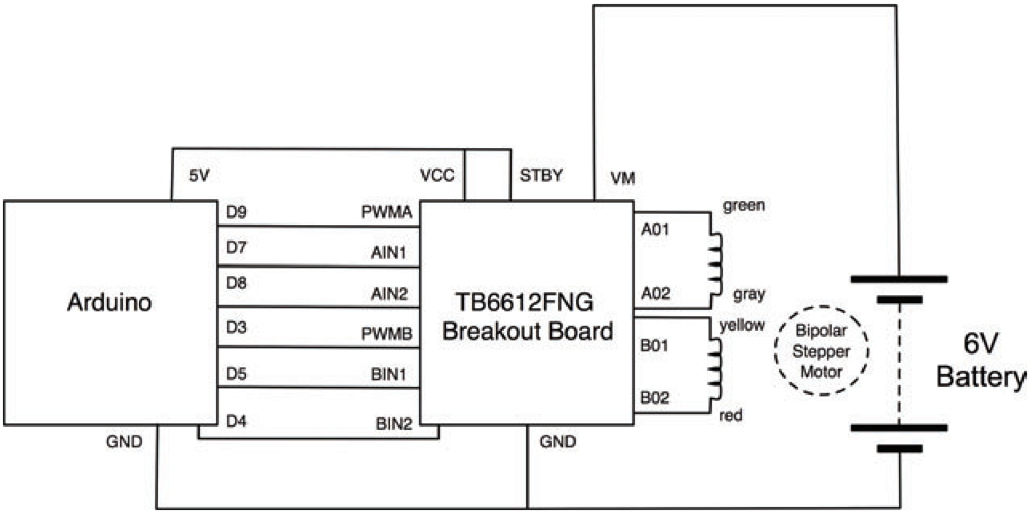
5V
VCC
STBY
VM
Arduino
D9
PWMA
D7
AIN1
D8
AIN2
D3
PWMB
D5
BIN1
D4
BIN2
TB6612FNG
Breakout Board
A01
A02
B01
B02
green
gray
yellow
red
Bipolar
Stepper
Motor
6V
Battery
GND
GND

FIGURE 7-28 Schematic diagram for a stepper motor control

If when you turn it on, the motor doesn't turn, you will only need to swap over one of the coil's leads. The colors indicated in Figure 7-28 match the lead colors for the Adafruit motor.

Although the motors suggested are 12V, they will still work using the 6V supplied by the battery pack. However, do not try and power them from the 5V supply of the Arduino. They draw too much current.

You Will Need

To build this, you will need the following items.

Quantity	Item	Appendix Code
1	Solderless breadboard	T5
	Solid-core jumper wire	T6
1	6 × AA battery holder	H8
1	6 × AA batteries	
1	TB6612FNG breakout board	M9
1	Bipolar stepper motor	H13
1	Arduino Uno/Leonardo	M2/M21
1	USB lead; Type B for Uno, Micro USB for Leonardo	

Construction

Figure 7-29 shows the breadboard layout.

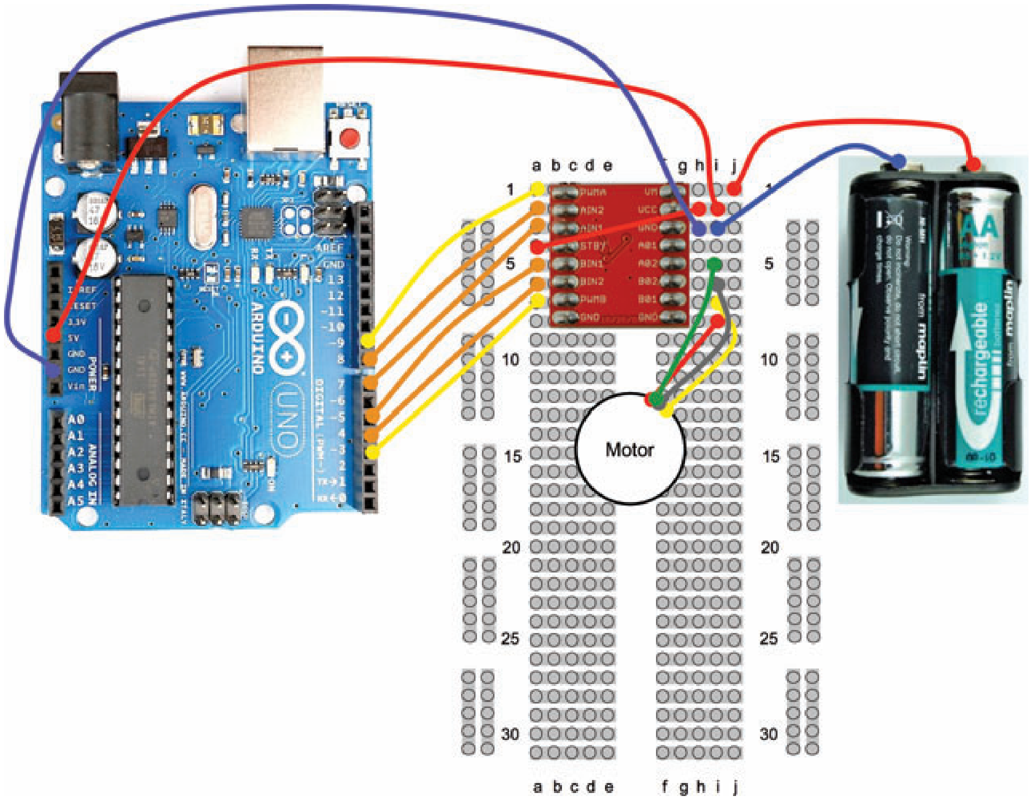
a b c d e
f g h i j
1
5
10
15
20
25
30
PWMA
AIN2
AIN1
STBY
BIN1
BIN2
PWMB
GND
VM
VCC
GND
A01
A02
B02
B01
GND
Motor
ARDUINO
UNO
AREF
GND
13
12
~11
~10
~9
8
7
~6
~5
4
~3
2
TX→1
RX←0
DIGITAL (PWM~)
IOREF
RESET
3.3V
5V
GND
GND
Vin
POWER
A0
A1
A2
A3
A4
A5
ANALOG IN
ON
TX
RX
L
AA
rechargeable
maplin

FIGURE 7-29 Breadboard layout for the stepper motor

Software

The example sketch ("stepper") will first of all turn the motor in one direction through 200 steps and then pause for a second before turning it in the opposite direction through 200 steps. For a 200-step motor, each turn will be a full 360 degrees.

First, the pin variables are defined and the "setup" function sets them all to be outputs.

```
// stepper

int PWMApin = 9;
int AIN1pin = 7;
int AIN2pin = 8;
int PWMBpin = 3;
int BIN1pin = 5;
int BIN2pin = 4;

void setup()
{
  pinMode(PWMApin, OUTPUT);
  pinMode(AIN1pin, OUTPUT);
  pinMode(AIN2pin, OUTPUT);
  pinMode(PWMBpin, OUTPUT);
  pinMode(BIN1pin, OUTPUT);
  pinMode(BIN2pin, OUTPUT);
}
```

The "loop" function then directs the motor in one direction (forward) for 200 steps, pauses a second, and then directs the motor back the same number of steps, but with half the delay between steps before pausing another second. It will continue this indefinitely.

```
void loop()
{
  forward(10, 200);
  delay(1000);
  back(5, 200);
  delay(1000);
}
```

The functions "forward" and "back" both take two parameters. The first is the delay between each step in milliseconds, and the second is the number of steps to take.

The "forward" and "back" functions use the function "setStep" to set the right polarities of the two coils, in the pattern 1010, 0110, 0101, 1001.

```
void forward(int d, int steps)
{
  for (int i = 0; i < steps / 4; i++)
  {
    setStep(1, 0, 1, 0);
    delay(d);
    setStep(0, 1, 1, 0);
    delay(d);
    setStep(0, 1, 0, 1);
    delay(d);
    setStep(1, 0, 0, 1);
    delay(d);
  }
}
```

To make the motor turn in the opposite direction, the pattern is just reversed.

```
void back(int d, int steps)
{
  for (int i = 0; i < steps / 4; i++)
  {
    setStep(1, 0, 0, 1);
    delay(d);
    setStep(0, 1, 0, 1);
    delay(d);
    setStep(0, 1, 1, 0);
    delay(d);
    setStep(1, 0, 1, 0);
    delay(d);
  }
}
```

The “setStep” function actually sets the appropriate outputs of the motor controller.

```
void setStep(int w1, int w2, int w3, int w4)
{
  digitalWrite(AIN1pin, w1);
  digitalWrite(AIN2pin, w2);
  digitalWrite(PWMApin, 1);
  digitalWrite(BIN1pin, w3);
  digitalWrite(BIN2pin, w4);
  digitalWrite(PWMBpin, 1);
}
```

How to Make a Simple Robot Rover

In this project, we will create a little roving robot. To do this, we will use the RF remote control we used in the section “How to Use a Wireless Remote Module,” with the H-Bridge module we

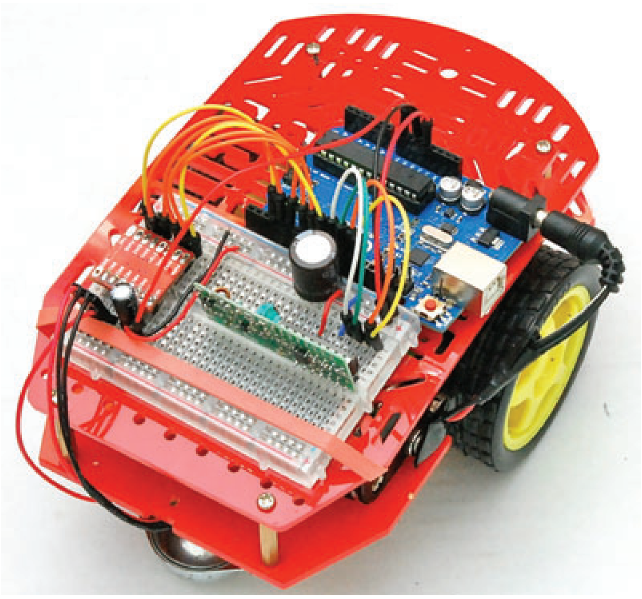

Figure 7-30 The robot rover

just discussed in the section "How to Control DC Motors with an H-Bridge Module," along with an Arduino.

The project will demonstrate how to use an Arduino to control a motor module.

The robot (Figure 7-30) will be built using a low-cost robot chassis kit that includes two gear motors.

The robot is built using a small breadboard that holds both the motor module and the RF receiver module. So apart from putting pin headers on the motor controller, there is no soldering to be done in this project.

You Will Need

To build this, you will need the following items.

Quantity	Name	Item	Appendix Code
1		Small solderless breadboard	H12
		Solid-core jumper wire	T6
1		6 × AA battery holder	H8
1		6 × AA batteries	
*1		Battery clip to 2.1mm DC jack adapter	H9

Quantity	Name	Item	Appendix Code
1		LED	K1
1		SparkFun TB6612FNG breakout board	M9
1		Magician chassis	H7
1		Header pins	K1, H4
1	C1	1000µF 16V capacitor	C1
1	C2	100µF 16V capacitor	C2
1		Arduino Uno/Leonardo	M2/M21
1		USB lead; Type B for Uno, Micro USB for Leonardo	

* If you use the Adafruit battery box that is already terminated in a 2.1mm plug, then you do not need this.

Construction

Figure 7-31 shows the schematic diagram for the rover.

The use of modules simplifies this greatly. The only additional components that have been added are the two capacitors: C1 and C2. These are necessary to prevent sudden drops in battery voltage (as the motors start) from causing the Arduino to reset.

Figure 7-31 Schematic diagram for the rover

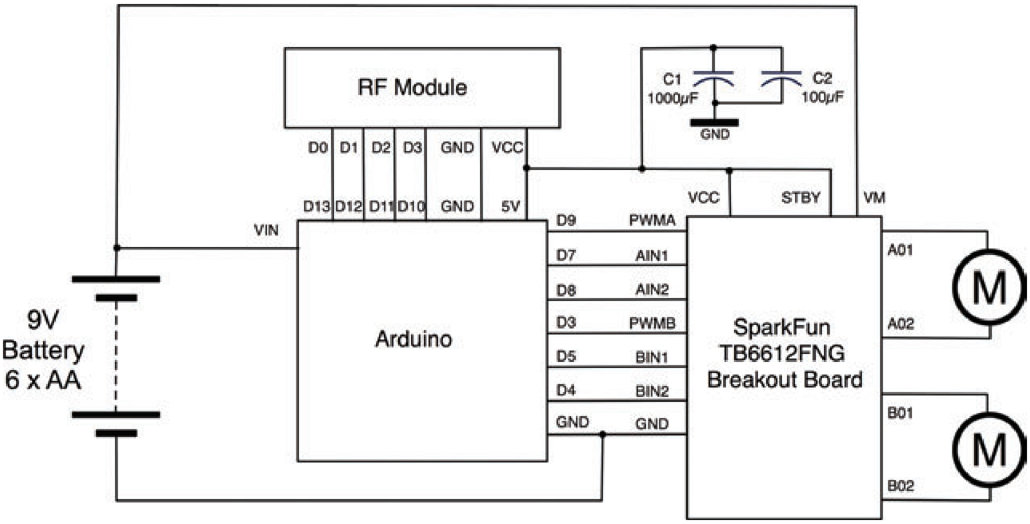
RF Module
D0
D1
D2
D3
GND
VCC
C1
1000µF
C2
100µF
GND
D13
D12
D11
D10
GND
5V
VIN
9V
Battery
6 x AA
Arduino
D9
D7
D8
D3
D5
D4
GND
PWMA
AIN1
AIN2
PWMB
BIN1
BIN2
GND
VCC
STBY
VM
SparkFun
TB6612FNG
Breakout Board
A01
A02
B01
B02
M
M

Step 1. Construct the Magician Chassis

The project is built around the Magician Chassis (Figure 7-32). This comes as a kit that is fixed together with nuts and bolts. Follow the instructions that come with the Chassis, but do not attach the battery box that comes with the kit or the support pillar that is right in the middle of the board. This is because, to power the Arduino, you need a bit more than the 5V to 6V that four AAs can supply. So, you are going to replace the battery box with one that takes six AA batteries rather than just four.

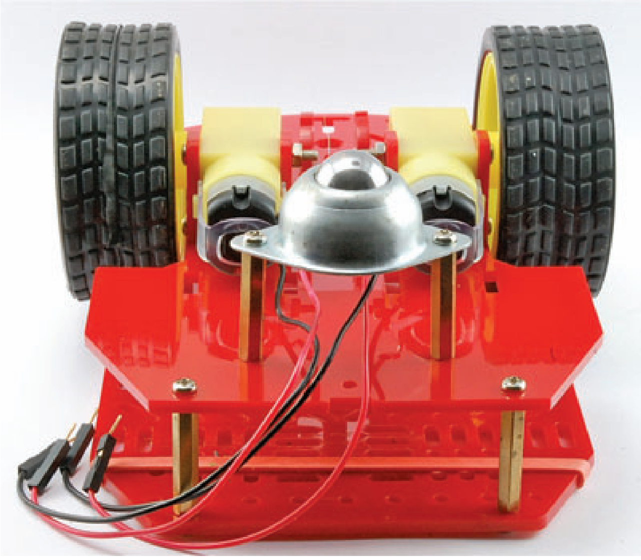

Figure 7-32 The Magician Chassis

Step 2. Program the Arduino

It is a good idea to program the Arduino with the sketch before you start attaching the electronics. Load the sketch "rover" onto the Arduino.

Step 3. Attach the Arduino and Breadboard

Find suitable mounting holes on the chassis and use small nuts and bolts to attach the Arduino. You can also use an elastic band for this. Some breadboards come with a self-adhesive backing, and you can use this to attach it to the chassis. For a less permanent way of attaching it, a rubber band will work just fine.

Step 4. Build the Breadboard

Figure 7-33 shows the breadboard layout for the project and how it is wired up to the Arduino.

There are quite a lot of wires on this project, so check all the connections once you think you have finished. Photocopying the page of the book and checking off with a pencil each connection as it's made is a good way of ensuring they are all there.

Also, not unlike our normal large breadboard, when wiring up this breadboard we have used the outer supply rail as GND and the inner rail as 5V.

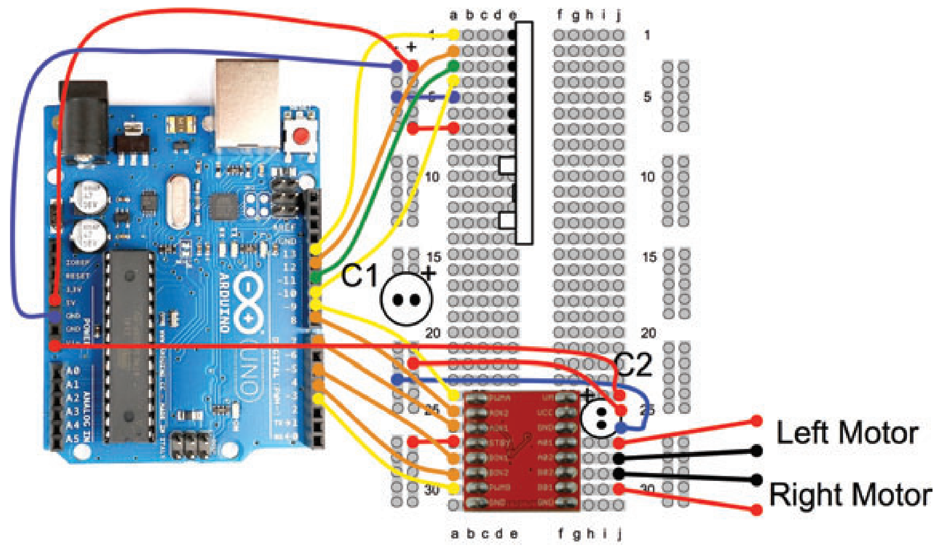

FIGURE 7-33 Breadboard layout for the rover

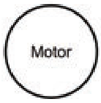

Step 5. Wire up the Motors

Each motor has a red and a black lead. So find the leads going to the left motor and attach them to the breadboard rows connected to the A01 and A02 connections of the motor module. Then, do the same for the right motor, connecting them to B01 and B02.

Step 6. Attach the Battery

If the battery box is made up of two rows of batteries, it will be quite a snug fit and the top surface of the chassis will need to bend out a little to accommodate it. If it is all in one row, like the Adafruit box, then you can attach it to the bottom layer of the chassis using small nuts and bolts.

Testing

When everything is assembled and ready to go, attach the battery and try the project out by pressing the buttons on the remote. The C button will set the robot running forwards, the B button will make it rotate to the right on the spot, and D will make it turn to the left. The A button will bring the robot to a halt.

Software

The sketch for this project is too long to list in full here, so we will just look at some of the main points.

The RF receiver toggles the output for the button you press. So press it once and it turns on, press it again and it turns off. However, this is not really the way we want it to work. We just want to know when a button has been pressed.

To do that, we keep track of the last state of each of the outputs, and only when the output has changed do we report the change. This uses the following array to store the output states and another array "remotePins":

```
int remotePins[] = {10, 11, 12, 13};
int lastPinStates[] = {0, 0, 0, 0};
```

The function that detects that a change has occurred is as follows:

```
int getKeyPress()
{
  // the outputs on the RF module toggle
  // so see what's changed and that's the
  // key that was pressed
  int result = -1;
  for (int i = 0; i < 4; i++)
  {
    int remoteInput = digitalRead(remotePins[i]);
    //Serial.print(remoteInput);
    if (remoteInput != lastPinStates[i])
    {
      result = i;
    }
    lastPinStates[i] = remoteInput;
  }
  return result;
}
```

The main loop calls this function to detect any key presses and then calls the appropriate function for the button.

```
void loop()
{
   int keyPressed = getKeyPress();
   Serial.println(keyPressed);
   if (keyPressed == 3)
   {
     stopMotors();
   }
   else if (keyPressed == 0)
   {
     turnLeft();
   }
   else if (keyPressed == 2)
   {
     turnRight();
   }
   else if (keyPressed == 1)
   {
     forward();
   }
   delay(20);
}
```

The functions that control the movement are all very similar. The function for turning left is shown next.

```
void turnLeft()
{
  digitalWrite(AIN1pin, HIGH);
  digitalWrite(AIN2pin, LOW);
  analogWrite(PWMApin, slowPower);
  digitalWrite(BIN1pin, LOW);
  digitalWrite(BIN2pin, HIGH);
  analogWrite(PWMBpin, slowPower);
}
```

This sets the AIN and BIN pins—in this case, to set the motors turning in opposite directions. The PWM power is controlled by a call to "analogWrite" using one of two values held in variables ("fullPower" and "slowPower").

How to Use a Seven-Segment LED Display Module

Seven-segment LED displays have a nice retro feel to them.

LED displays made up of a number of LEDs contained in a single package can be a challenge to control. Such displays will normally be controlled using a microcontroller; however, it is not necessary to use a microcontroller output pin to each individual LED. But rather, multi-LED displays are organized as "common anode" or "common cathode," with all the LED terminals of the anode or cathode connected together and then brought out through one pin. Figure 7-34 shows how a common cathode seven-segment display might be wired internally.

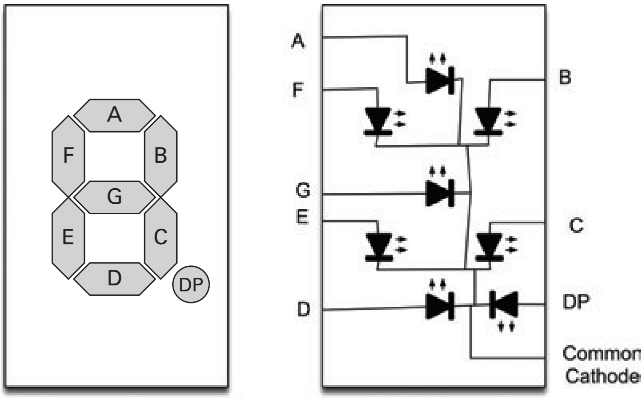

Figure 7-34 A common cathode LED display

In a common cathode display like this, the common cathode would be connected to ground and each segment anode driven by a microcontroller pin through a separate resistor. Do not be tempted to use one resistor on the common pin, and don't use any resistors on the non-common connections, since the current will be limited no matter how many LEDs are lit. Because of this, the display will get dimmer the more LEDs are illuminated.

It is quite common for multiple displays to be contained in the same case—for example, the three-digit, seven-segment common cathode LED display shown in Figure 7-35.

In this kind of display, each digit of the display is like the single-digit display of Figure 7-35, and has its own common cathode. In addition, all the A segment anodes are connected together, as are each segment.

Figure 7-35 A three-digit, seven-segment LED display

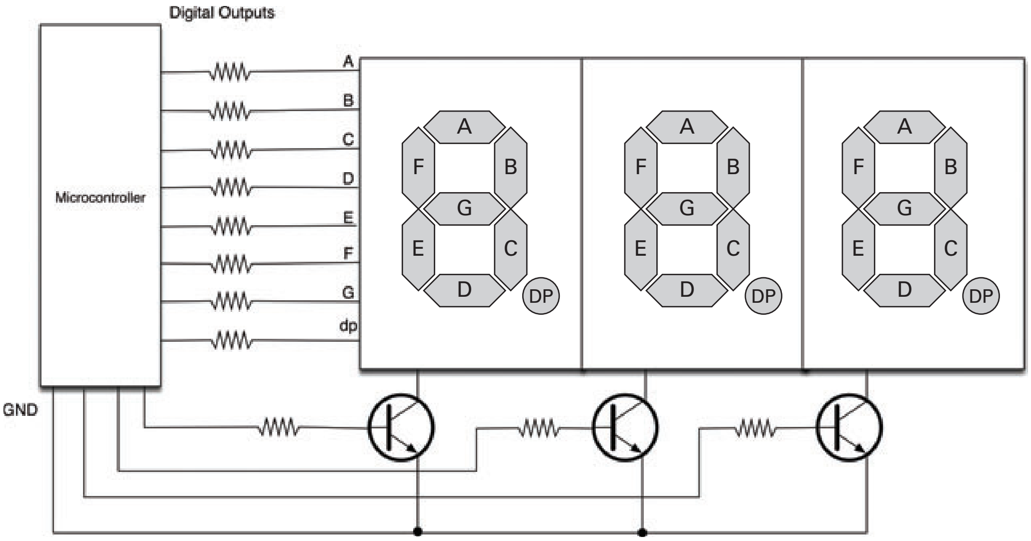

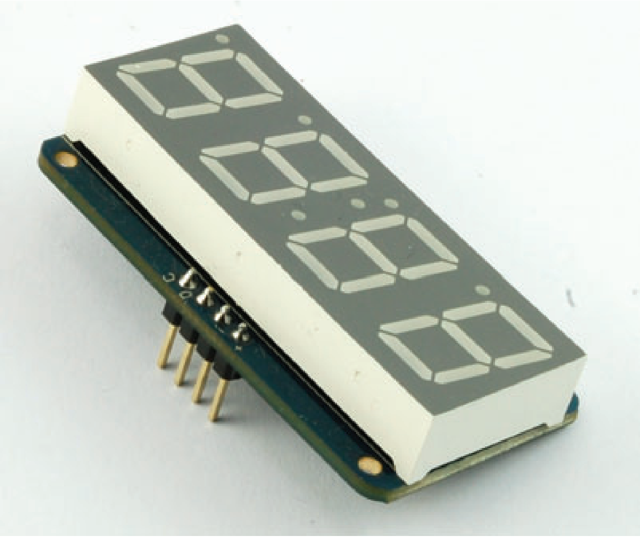

FIGURE 7-36 A four-digit, seven-segment I2C display

The Arduino using the display will then activate each common cathode in turn and then turn on the appropriate segments for that digit, and then move onto the next digit, and so on. This refresh happens very quickly so that the display appears to display different numbers on each digit. This is called multiplexing.

Note the use of transistors to control the common cathodes. This is simply to handle the current of potentially eight LEDs at once, which would be too much for most microcontrollers.

Fortunately for us, there is a much simpler way to use multi-digit, seven-segment LED displays. Modules ride to the rescue once again!

Figure 7-36 shows a four-digit, seven-segment LED display that has just four pins on its connector, and two of them are for power.

You Will Need

To build this, you will need the following items.

Quantity	Item	Appendix Code
1	Solderless breadboard	T5
	Solid-core jumper wire	T6
1	Arduino Uno/Leonardo	M2/M21
1	USB lead; Type B for Uno, Micro USB for Leonardo	
1	Adafruit seven-segment display w/I2C backpack	M19

Construction

The module comes as a kit, so start by following the instructions that accompany the module to assemble it.

The LED module uses a type of serial interface on the Arduino called I2C (pronounced "I squared C"). This requires just two pins, but they have to be the two pins above "AREF" on the Arduino Uno. These pins are named SDA and SCL.

This means that, frustratingly, the module will not just plug straight into the Arduino, we will need to use breadboard.

Figure 7-37 shows the breadboard layout and Figure 7-38 the breadboard itself, with the seven-segment display in action.

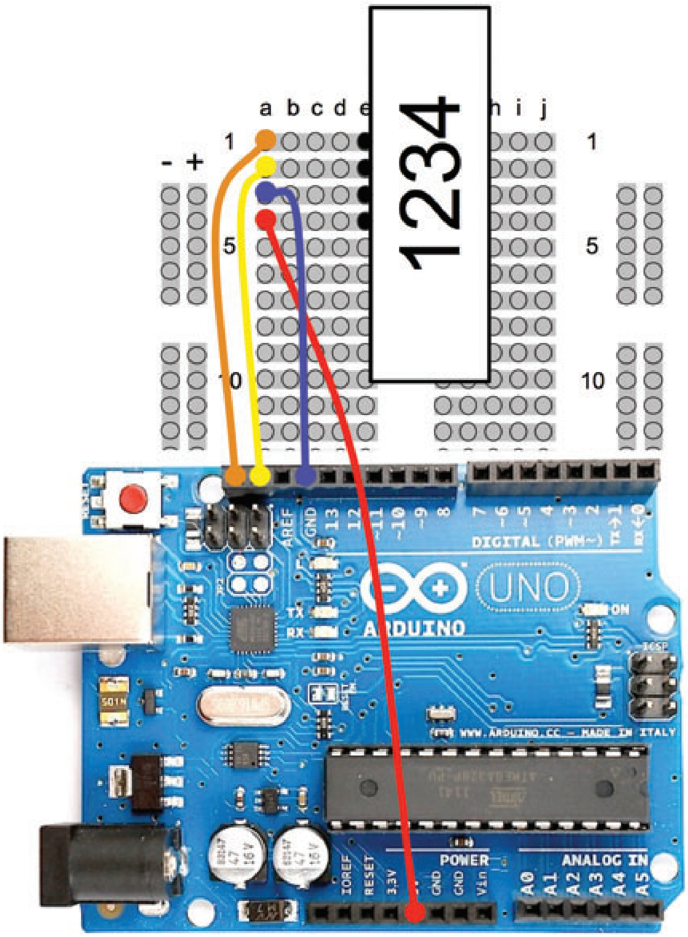
a b c d e
h i j
1
5
10
- +
1234
AREF
GND
13
12
DIGITAL (PWM~)
UNO
ARDUINO
TX
RX
ON
ICSP
WWW.ARDUINO.CC – MADE IN ITALY
POWER
ANALOG IN
IOREF
RESET
3.3V
5V
GND
GND
Vin
A0
A1
A2
A3
A4
A5

FIGURE 7-37 Breadboard layout for using the seven-segment display

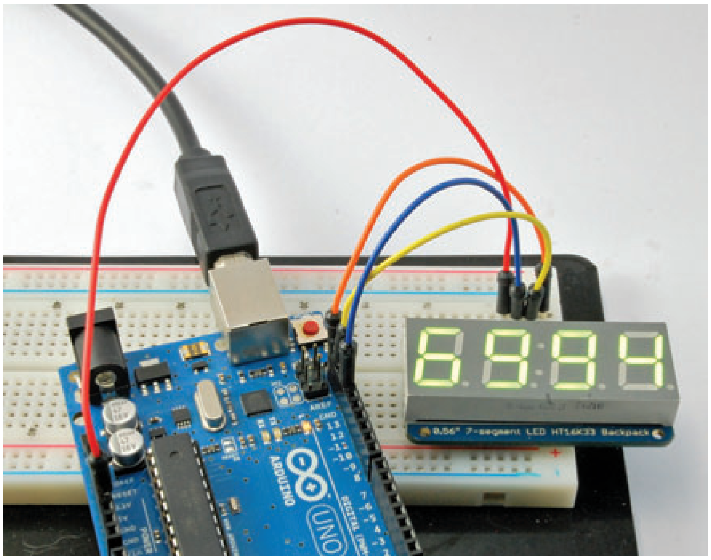

Figure 7-38 The seven-segment display in action

Software

Adafruit provides a library to simplify the use of the module. You need to download this and copy the library folder into the "libraries" folder in your Arduino documents folder. See the instructions on Adafruit's web site at www.adafruit.com/products/880.

The three libraries that the module requires are loaded using the #includes statements.

```
// seven_seg_display

#include <Wire.h>
#include "Adafruit_LEDBackpack.h"
#include "Adafruit_GFX.h"
```

The following line assigns a variable to the display object so we can tell it what to display.

```
Adafruit_7segment disp = Adafruit_7segment();
```

The "setup" function begins serial communication on the I2C pins and then initializes the display. The value 0x70 is the I2C address of the display module. This is the default value for its address, but there are solder connections on the module you can short together to change the address. You might want to do this if you need to use more than one display, since each display must have a different address.

```
void setup()
{
  Wire.begin();
  disp.begin(0x70);
}
```

The “loop” function simply displays the current number of milliseconds since the board was reset, divided by 10. The display will therefore count up in 1/100ths of a second.

```
void loop()
{
  disp.print(millis() / 10);
  disp.writeDisplay();
  delay(10);
}
```

How to Use a Real-Time Clock Module

You could write an Arduino sketch to keep track of the time, but as soon as you unplugged it, it would forget the time. The way around this problem is to use an RTC (real-time clock) like the one shown in Figure 7-39.

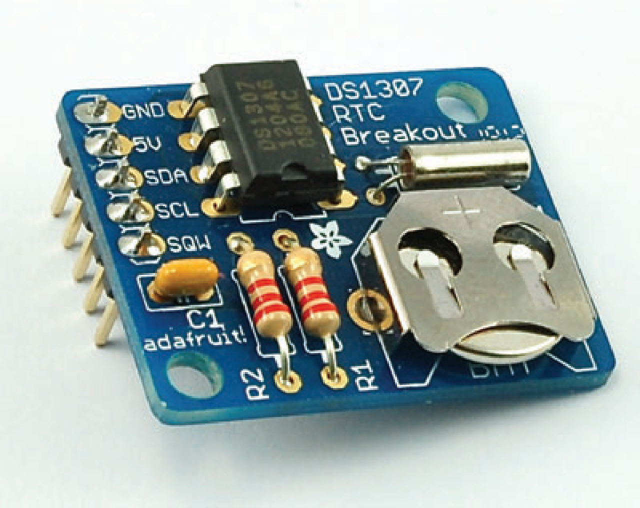

Figure 7-39 An RTC module

FIGURE 7-40 A "simple" digital clock

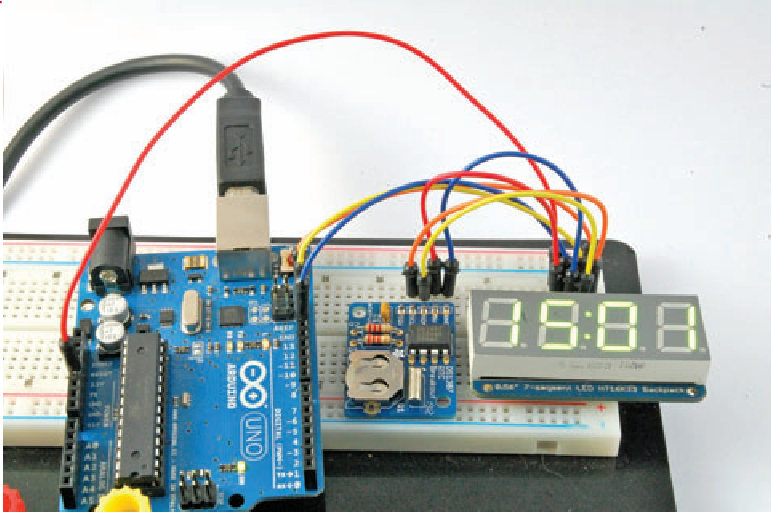

This particular module is also an Adafruit product. There are lots of similar modules out there, but their pin allocations may be different.

The RTC includes a lithium battery that will last for years, and provides enough power to keep the correct time when the module is not powered.

We can combine the RTC module with the seven-segment display module we used previously and make ourselves a simple digital clock (Figure 7-40).

You Will Need

To build this, you will need the following items.

Quantity	Item	Appendix Code
1	Solderless breadboard	T5
	Solid-core jumper wire	T6
1	Arduino Uno/Leonardo	M2/M21
1	USB lead; Type B for Uno, Micro USB for Leonardo	
1	Adafruit seven-segment display w/I2C backpack	M19
1	RTC module	

Construction

The RTC module also comes as a kit, so start by following the instructions that accompany the module to assemble it.

The RTC module also uses I2C and has a different address to the display, so we do not need to change anything.

Figure 7-41 shows the breadboard layout for the clock.

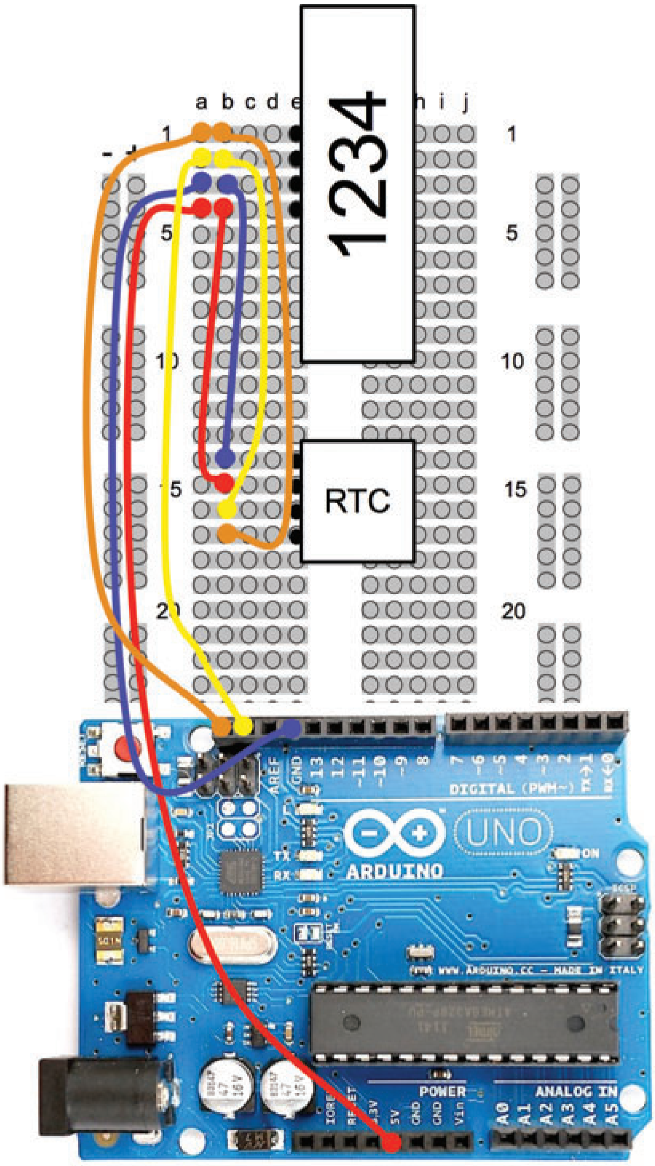
a b c d e
h i j
1
5
10
15
20
1234
RTC
ARDUINO
UNO
DIGITAL (PWM~)
POWER
ANALOG IN

FIGURE 7-41 Breadboard layout for the clock

Software

Load up the sketch "clock" onto your Arduino. The display should immediately start showing the time your computer is set to.

Much of this sketch is the same as that in the section "How to Use a Seven-Segment LED Display Module." But there is one additional library for the RTC module that we need to import. Instructions for downloading this are linked from the product page for the RTC module (www.adafruit.com/products/264).

```
// clock

#include <Wire.h>
#include "Adafruit_LEDBackpack.h"
#include "Adafruit_GFX.h"
#include "RTClib.h"
```

In addition to creating a display to use, we now have to give the RTC a name. Let's call it "RTC".

```
RTC_DS1307 RTC;
Adafruit_7segment disp = Adafruit_7segment();
```

The "setup" function now has an additional command to start the RTC so it is ready to receive commands. The "if" statement checks to see if the clock part of the RTC is active. If this is the first time it has been used, it will not be, so if this is the case, it initializes it to the programming computer's time.

```
void setup()
{
  Wire.begin();
  RTC.begin();
  if (! RTC.isrunning())
  {
    RTC.adjust(DateTime(__DATE__, __TIME__));
  }
  disp.begin(0x70);
}
```

The main loop now reads the time from the RTC and displays it. It also uses the display libraries' "drawColon" function to make the colon flash by turning it on and off with a half-second delay in between.

```
void loop()
{
  disp.print(getDecimalTime());
  disp.drawColon(true);
  disp.writeDisplay();
  delay(500);
  disp.drawColon(false);
  disp.writeDisplay();
  delay(500);
}
```

The “getDecimalTime” function reads the hours and minutes from the RTC and turns them into a decimal number that can be written to the display. The first two digits will contain the hour, and the left two digits the minute.

```
int getDecimalTime()
{
  DateTime now = RTC.now();
  int decimalTime = now.hour() * 100 + now.minute();
  return decimalTime;
}
```

Summary

In addition to the modules here, you will find lots of other useful modules on the web sites of companies like Adafruit and SparkFun. The web sites also include some information on how to use the modules and their specifications. If you find a module you would like to make use of, the first step is to research how you could use it. As well as the datasheets and tutorial information on the supplier’s web site, you will often find instructions on building the projects if you search for the module on the Internet.

8

Hacking with Sensors

Chapters 6, 7, and 8 all overlap somewhat, as many sensors are also modules and both can often be used with an Arduino.

In this chapter, we will look at how to use a range of sensors, whether with a little supporting electronics or as an input to an Arduino, or sometimes both.

How to Detect Noxious Gas

In this section, you will use a methane sensor (Figure 8-1).

While they look like they should be expensive, these sensors are really quite low cost. They include a small heater (connected between the two H connections) and a catalytic sensing element whose resistance changes depending on the concentration of methane. Although the project will run on batteries, it will burn through them pretty quickly because these sensors have a heating element that will consume 150 to 200 mA.

The sensing of methane does have lots of sensible scientific and industrial uses. However, we will use this technological know-how for the puerile activity of … detecting farts.

You Will Need

To experiment with this gas sensor, you will need the following items.

Quantity	Names	Item	Appendix Code
1	D1	LED	K1
1	R1	10kΩ trimpot	K1
1	R2	10kΩ resistor	K2
1	R3	470Ω resistor	K2
1	IC1	LM311 comparator	S7

Quantity	Item	Appendix Code
1	Methane sensor MQ-4	M11
1	Piezo buzzer (with own oscillator)	M10
1	Solderless breadboard	T5
	Solid-core jumper wire	T6
1	4 × AA battery holder	H1
1	4 × AA batteries	
1	Battery clip	H2
1	* Arduino Uno/Leonardo	M2/M21
1	* USB lead; Type B for Uno, Micro USB for Leonardo	

* Only required if you want to connect the detector to an Arduino.

The piezo sounder must be of the type that includes its own oscillator circuit and will work at 6V.

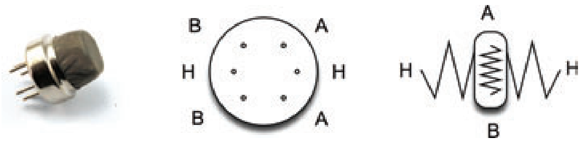

Figure 8-1 A methane sensor

The LM311 Comparator

Figure 8-2 shows the schematic diagram for the gas detector.

The key to this circuit is the comparator IC (LM311). Comparators, as the name suggests, compare voltages. If the voltage at its "+" connection is greater than the voltage at its "–"

Figure 8-2 Schematic diagram for the gas detector

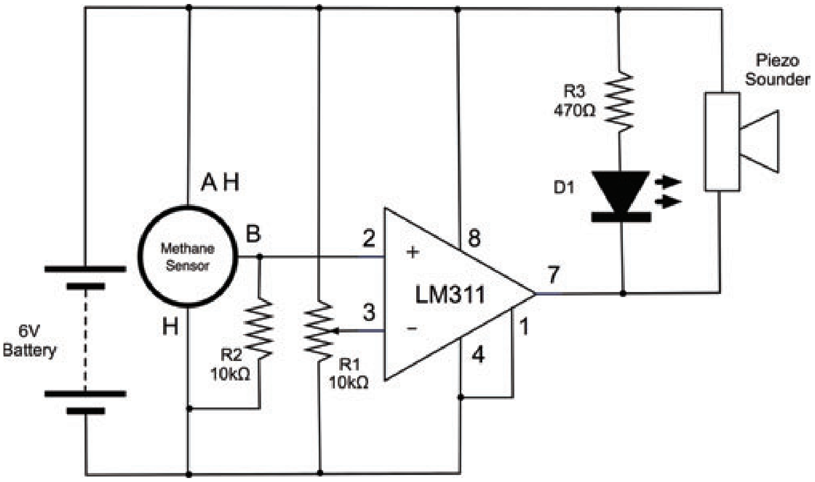

connection, then its output turns on. In this case, that will light the LED and sound a buzzer.

The trimpot supplies a threshold voltage to the negative input of the comparator. To use the gas detector, the trimpot is turned until the LED just goes out. It will come back on if the output from the sensor increases enough to exceed the value at the comparator's negative input.

The sensor has rather unusual connections. It has six connections, but some of them are doubled up and connected behind the scenes (see Figure 8-1). The H connectors supply a heating element that warms the catalyst layer between A and B. When methane is detected, the resistance between A and B falls. R2 forms a voltage divider with the sensing element. One benefit of the sensor basically being two resistors—one acting as a heater and the other as a sensor—is that the pin connections are reversible.

The sensor leads are thick and at a strange spacing, so they will not fit in breadboard. For this reason, we solder some leads to them (Figure 8-3).

FIGURE 8-3 Attaching leads to the sensor

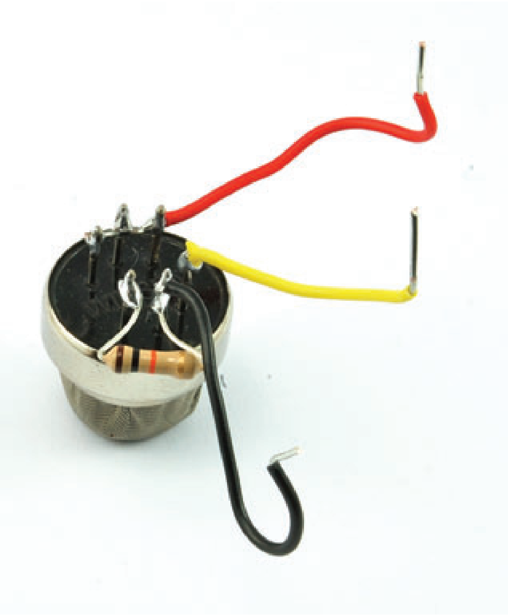

Rather than solder wires to all the leads, we can just solder the following connections:

- A red positive supply lead to all the pins on one side of the sensor (the two A pins and one H connection)
- The resistor R2 between B and the GND side of the heater
- A GND lead to the GND side of the heater (black)
- An output lead to B (yellow)

Breadboard

Figure 8-4 shows the breadboard layout for the gas detector, while Figure 8-5 displays the project itself.

The breadboard layout is very straightforward, but do make sure that the IC is the correct way around. When it is all assembled, I will leave you to find your own way of testing it. Just a note that breathing on the sensor will also set it off.

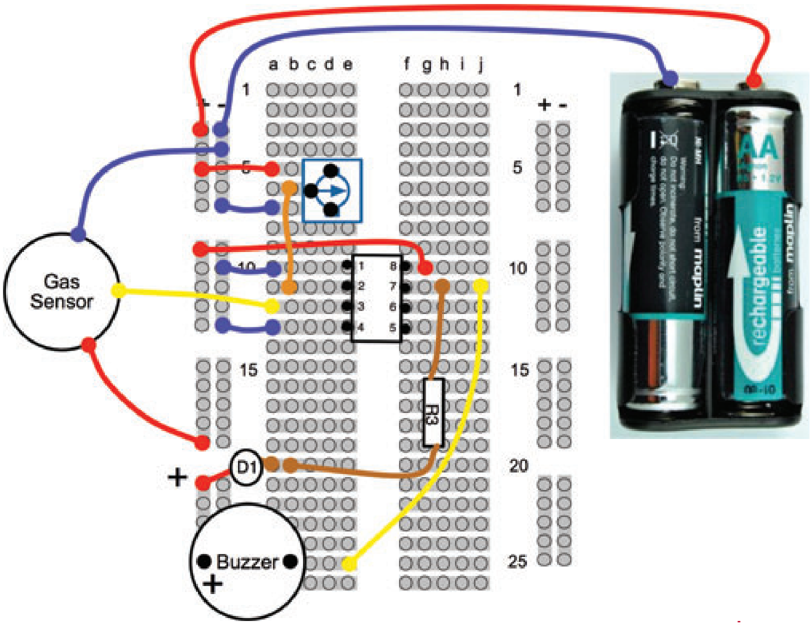
a b c d e
f g h i j
1
5
10
15
20
25
+ -
Gas Sensor
D1
+
Buzzer
+
R3
1
2
3
4
5
6
7
8
AA
rechargeable
maplin

Figure 8-4 Breadboard layout for the gas detector

Using a Gas Sensor with an Arduino

In soldering the three leads onto the methane sensor, we have also made something that we can easily attach directly to an Arduino (Figure 8-6).

Connect the positive supply connection from the sensor to 5V on the Arduino, GND on the sensor to GND on the Arduino, and the output of the sensor to A3.

Since this sensor can use up to 200 mA, you must power it from the real 5V and GND connections on the Arduino and not use the trick of powering it from a digital output.

The following sketch ("methane") prints the readings from the sensor into the Serial Monitor. Again, note that if you breathe on the sensor, the reading will increase.

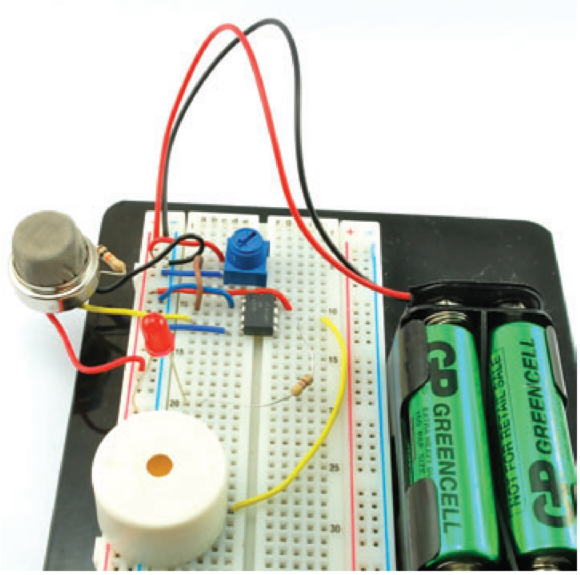
GP GREENCELL
NOT FOR RETAIL SALE
GP GREENCELL

Figure 8-5 The gas detector

FIGURE 8-6 Using the gas sensor with an Arduino

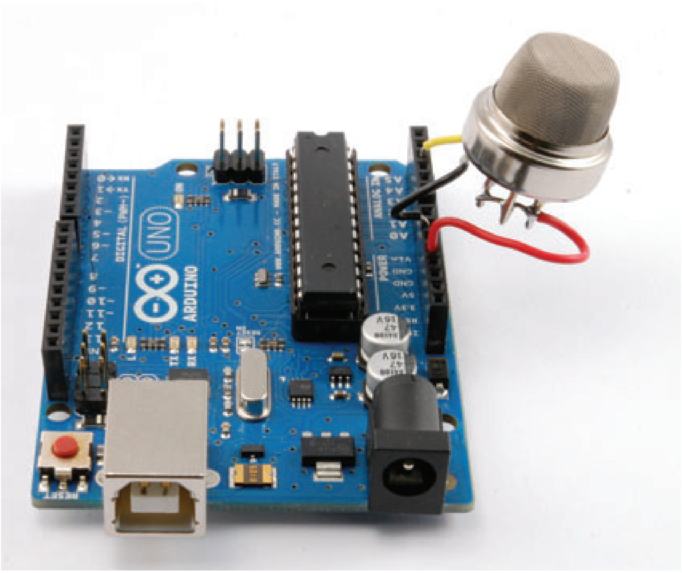

```
// methane

int analogPin = 3;

void setup()
{
  Serial.begin(9600);
  Serial.println("Methane Detector");
}

void loop()
{
  Serial.println(analogRead(analogPin));
  delay(500);
}
```

How to Measure Something's Color

The TCS3200 is a handy little IC for measuring the color of something. There are several different variations on this chip, but they all work the same way. The chip has a transparent case, and dotting its surface are photodiodes with different color filters over them (red, green, and blue). You can read the relative amounts of each primary color.

The easiest way to use the chip is to buy a module like the one shown in Figure 8-7.

This module, which cost less than USD 10, also has four white LEDs that illuminate the object whose color you want to measure, as well as convenient header pins.

Table 8-1 shows the connections on the module and their purpose. With the exception of the power to the LEDs, these connections are taken straight from the IC, so any module you find that uses the TCS3200 is likely to have the same connections, even if they are not quite in the same place.

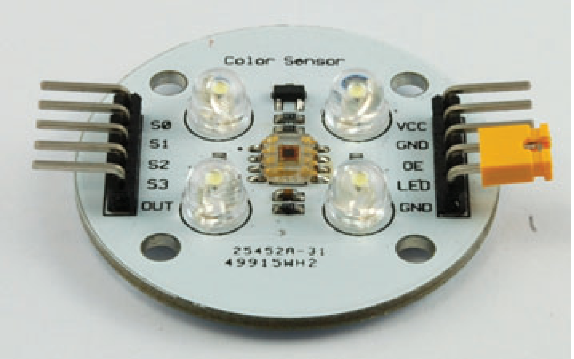

FIGURE 8-7 A light-sensing module

The IC does not produce an analog output, but instead varies the frequency of a train of pulses. You choose which color the pulse frequency corresponds to by changing the values on the digital inputs S2 and S3.

You Will Need

Quantity	Item	Appendix Code
1	Arduino Uno/Leonardo	M2/M21
1	USB lead; Type B for Uno, Micro USB for Leonardo	
1	Color-sensing module	M12
1	Male-to-female jumper set	T12

Pin	Description	Description	Pin
S0	S0 and S1 select the frequency range. Both should be set HIGH.	2.5V to 5.5V	VCC
S1		Ground	GND
S2	Red—S2 and S3 LOW Green—S2 and S3 HIGH Blue—S2 LOW, S3 HIGH White—S2 HIGH, S3 LOW	Output Enable—set to LOW to effectively turn the chip on.	OE
S3		Tie to ground with the attached jumper to turn the LEDs on.	LED
OUT	The output pulses.		GND

TABLE 8-1 Color-Sensing Module Pinout

Construction

Construction is perhaps too strong a word for it. The module will fit directly into the Arduino (Figure 8-8), facing outward. It will make the following connections:

- S0 module to D3 Arduino
- S1 module to D4 Arduino
- S2 module to D5 Arduino
- S3 module to D6 Arduino
- OUT module to D7 Arduino

You will also need three male-to-female jumper leads to connect:

- VCC module to 5V Arduino
- GND module to GND Arduino
- OE module to GND Arduino

Figure 8-9 shows the module sensing colors on a Rubik's cube.

Software

The sketch "color_sensing" demonstrates the use of this module.

```
// color_sensing
int pulsePin = 7;
int prescale0Pin = 3;
int prescale1Pin = 4;
int colorSelect0pin = 5;
int colorSelect1pin = 6;
```

The pins are named according to their function rather than using the module pin names.

The "setup" function sets the appropriate pin modes and then sets both the "prescale" pins that control the

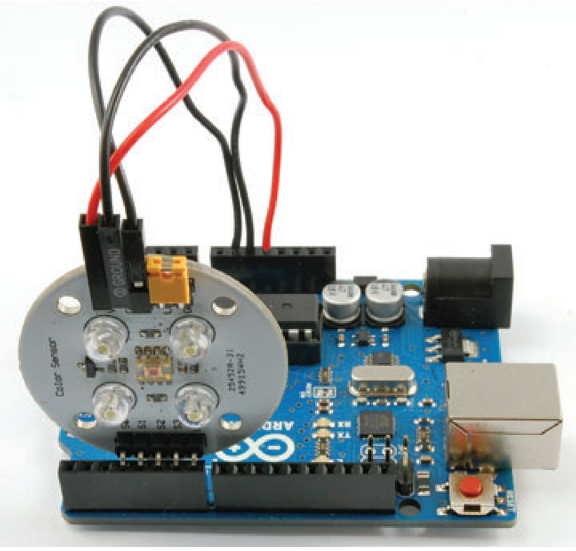

FIGURE 8-8 A light sensor attached to an Arduino

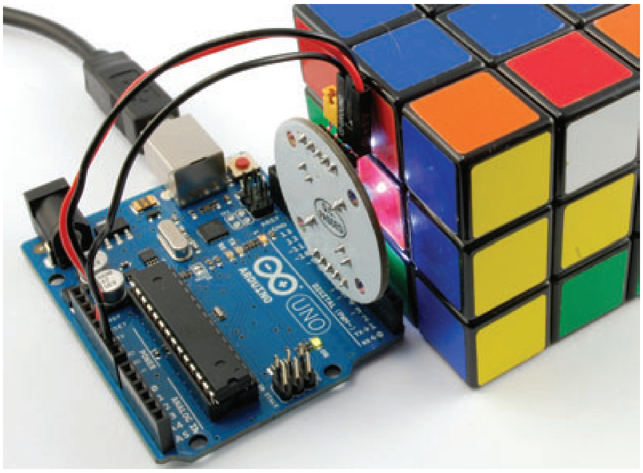

FIGURE 8-9 Sensing colors on a Rubik's cube

output frequency range to HIGH, starts serial communication, and then displays a welcome message.

```
void setup()
{
  pinMode(prescale0Pin, OUTPUT);
  pinMode(prescale1Pin, OUTPUT);
  // set maximum prescale
  digitalWrite(prescale0Pin, HIGH);
  digitalWrite(prescale1Pin, HIGH);
  pinMode(colorSelect0pin, OUTPUT);
  pinMode(colorSelect1pin, OUTPUT);
  pinMode(pulsePin, INPUT);
  Serial.begin(9600);
  Serial.println("Color Reader");
}
```

The “loop” function reads the three different colors (more on that later) and displays a message depending on the dominant color. Note that the lower the value, the brighter that particular color.

```
void loop()
{
  long red = readRed();
  long green = readGreen();
  long blue = readBlue();
  if (red < green && red < blue)
  {
    Serial.println("RED");
  }
  if (green < red && green < blue)
  {
    Serial.println("GREEN");
  }
  if (blue < green && blue < red)
  {
    Serial.println("BLUE");
  }
  delay(500);
}
```

Each of the functions—“readRed”, “readGreen”, “readBlue”, and “readWhite”—just call a function “readColor” with the appropriate values for S2 and S3.

```
long readRed()
{
  return (readColor(LOW, LOW));
}
```

The function “readColor” first sets the appropriate pins for the color and records a start time in the variable “start”. It then waits for 1000 pulses to happen. Afterward, it returns the difference between the current time and the start time.

```
long readColor(int bit0, int bit1)
{
  digitalWrite(colorSelect0pin, bit0);
  digitalWrite(colorSelect1pin, bit1);
  long start = millis();
  for (int i=0; i< 1000; i++)
  {
    pulseIn(pulsePin, HIGH);
  }
  return (millis() - start);
}
```

Although not actually used, there is also a function in the sketch that writes the color values to the Serial Monitor.

```
void printRGB()
{
  Serial.print(readRed()); Serial.print("\t");
  Serial.print(readGreen()); Serial.print("\t");
  Serial.print(readBlue()); Serial.print("\t");
  Serial.println(readWhite());
}
```

How to Detect Vibration

Piezo vibration sensors, like the one from SparkFun shown in Figure 8-10, are very easy to use with an Arduino.

The sensors are a thin strip of piezoelectric material with a rivet in the end acting as a weight. When there is a vibration, the weight moves, stressing the piezo material that produces a spike in voltage. Measured with the right test equipment, this spike can be as high as 80V. However, because we are going to connect it to an analog input on an Arduino, the resistance of that input will be sufficient to damp the voltage to a level that will not harm our Arduino.

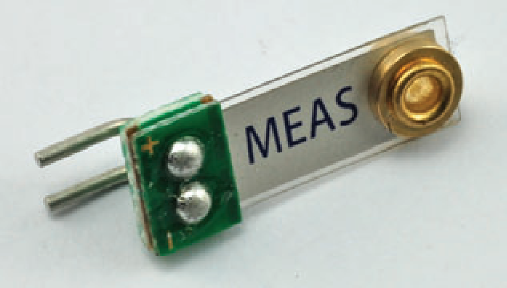

Figure 8-10 A piezo vibration sensor

You Will Need

To detect vibration with your piezo sensor, you will need the following items.

Quantity	Item	Appendix Code
1	Arduino Uno/Leonardo	M2/M21
1	USB lead; Type B for Uno, Micro USB for Leonardo	
1	Piezo vibration sensor	M13
1	LED	K1
1	220Ω resistor	K2

Construction

The piezo sensor is another very Arduino-friendly sensor. It can be just plugged into the Arduino sockets. In this case, it is plugged into pins A0 and A1. A0 will be set to an output LOW and used to provide the ground connection to the sensor (Figure 8-11). Note that the module is marked with a "+" on one side. Connect this side to "A1".

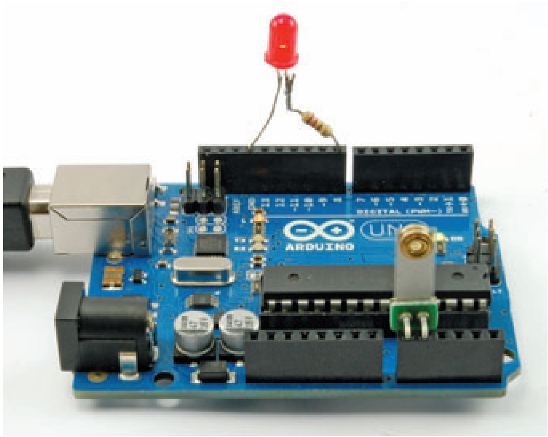

FIGURE 8-11 Sensing vibration with an Arduino

The LED is joined to a resistor as described back in Chapter 6. This can then be plugged into sockets 8 and GND on the Arduino, with the positive connection of the LED connected to 8.

Software

The software that follows uses the technique of calibrating itself as it starts, to get the "no vibration" reading from the sensor. It then waits until the sensor reading exceeds the threshold set, at which point it lights the LED. Pressing the Arduino "reset" button will cause the sensor to detect movement again.

```
// vibration_sensor

int gndPin = A0;
int sensePin = 1;
int ledPin = 8;
```

After defining the pins to use, we then define two variables. The variable "normalReading" is used during calibration (more on that in a minute), and the variable "threshold" should be set to the amount that the analog reading is allowed to exceed "normalReading" by before the LED is turned on.

```
int normalReading = 0;
int threshold = 10;
```

The "setup" function sets the appropriate pin modes and then calls the "calibrate" function to find the reading for the sensor when there is no vibration.

```
void setup()
{
  pinMode(gndPin, OUTPUT);
  digitalWrite(gndPin, LOW);
  pinMode(ledPin, OUTPUT);
  normalReading = calibrate();
}
```

The “loop” function simply takes a reading and checks to see if it has exceeded the threshold. If it has, it turns the LED on.

```
void loop()
{
  int reading = analogRead(sensePin);
  if (reading > normalReading + threshold)
  {
    digitalWrite(ledPin, HIGH);
  }
}
```

To calibrate the sensor, 100 readings are made with a one-millisecond delay between each reading, and the average is returned. A variable of type “long” is used to hold the total, as this number may be too big to fit in the usual “int” type.

```
int calibrate()
{
  int n = 100;
  long total = 0;
  for (int i = 0; i < n; i++)
  {
    total = total + analogRead(sensePin);
    delay(1);
  }
  return total / n;
}
```

How to Measure Temperature

A number of different sensor ICs are designed for measuring temperature. Perhaps the simplest to use is the TMP36 (Figure 8-12).

You can experiment with the sensor, just printing the temperature to the Serial Monitor, or you can combine the sensor with the relay module we made in Chapter 6.

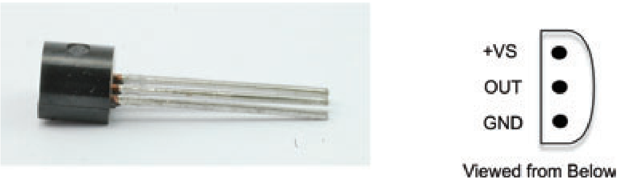

Figure 8-12 The TMP36

You Will Need

To use this temperature measurement IC, you will need the following items.

Quantity	Item	Appendix Code
1	Arduino Uno/Leonardo	M2/M21
1	USB lead; Type B for Uno, Micro USB for Leonardo	
1	TMP36 temperature sensor IC	S8

Construction

The TMP36 has just three pins, two for the power supply and one analog output. The power supply needs to be between 2.7V and 5.5V, making it ideal for use with the 5V of an Arduino. In fact, we can supply the power to it through digital outputs and just plug the whole chip into three pins on the analog connector of the Arduino (Figure 8-13).

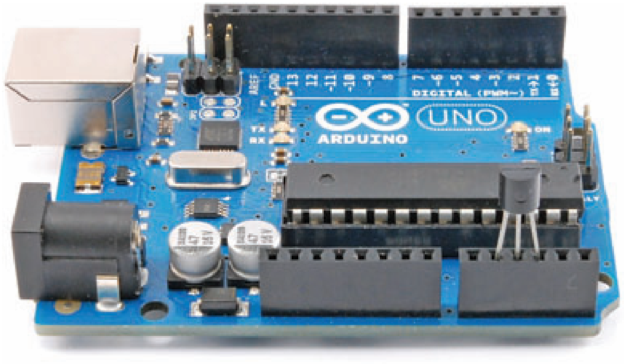

Figure 8-13 The TMP36 attached to an Arduino

Software

The sketch ("temperature_sensor") follows what should now be a fairly familiar pattern. The pins are defined, and then in the "setup" function the output pins that provide power to the sensor are set to LOW for GND and HIGH for the positive supply.

```
// temperature_sensor

int gndPin = A1;
int sensePin = 2;
int plusPin = A3;

void setup()
{
  pinMode(gndPin, OUTPUT);
  digitalWrite(gndPin, LOW);
  pinMode(plusPin, OUTPUT);
  digitalWrite(plusPin, HIGH);
  Serial.begin(9600);
}
```

The main loop reads the value from the analog input and then does a bit of arithmetic to calculate the actual temperature.

First, the voltage at the analog input is calculated. This will be the raw value (between 0 and 1023) divided by 205. It is divided by 205 because a span of 1024 values occupies 5V, or 1024 / 5 = 205 per volt.

The TMP36 outputs a voltage from which the temperature in degrees C can be calculated from the equation:

tempC = 100.0 * volts – 50

For good measure, the sketch also converts this into degrees F and prints both out to the Serial Monitor.

```
void loop()
{
  int raw = analogRead(sensePin);
  float volts = raw / 205.0;
  float tempC = 100.0 * volts - 50;
  float tempF = tempC * 9.0 / 5.0 + 32.0;
  Serial.print(tempC);
  Serial.print(" C ");
  Serial.print(tempF);
  Serial.println(" F");
  delay(1000);
}
```

How to Use an Accelerometer

Tiny accelerometer modules (Figure 8-14) are now available at low cost. The two models shown are very similar, both being 5V compatible and providing analog outputs for each axis. The one on the left is from Freetronics (www.freetronics.com/am3x) and the one on the right is from Adafruit (www.adafruit.com/products/163).

These modules are three axis accelerometers that measure the force applied to a tiny weight inside the chip. Two of the dimensions, X and Y are parallel to the modules PCB. The third dimension (Z) is at 90 degrees to the module’s surface. There will normally be a constant force acting on this dimension due to gravity. So if you tip the module, the effect of gravity starts to increase on the dimension in which you tip it (see Figure 8-15).

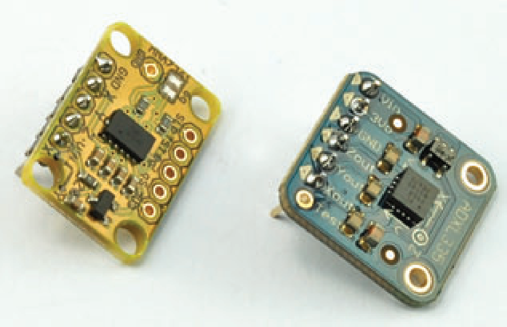

FIGURE 8-14 Accelerometer modules

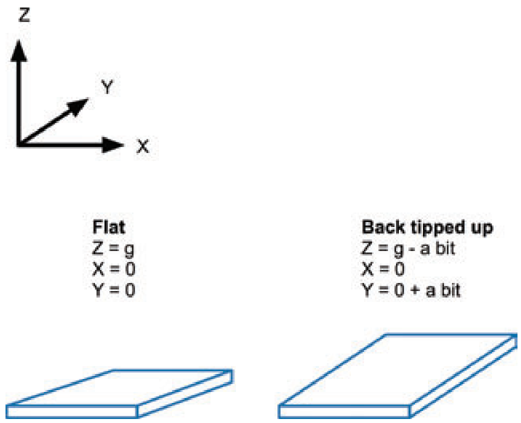

FIGURE 8-15 The effect of gravity on the accelerometer

As a vehicle to test one of these accelerometers, we are going to build an electronic version of the children's game of egg and spoon. The idea behind this is to use the accelerometer to detect the level of tilt of the "spoon" and flash an LED when it starts to be in danger of losing the egg. A buzzer sounds when the level of tilt is extreme enough for the egg to have fallen off (Figure 8-16).

You Will Need

To participate in an Arduino and spoon race, you will need the following items.

Quantity	Item	Appendix Code
1	Arduino Uno/Leonardo	M2/M21
1	USB lead; Type B for Uno, Micro USB for Leonardo	
1	Accelerometer	M15
1	Piezo buzzer	M3
1	LED	K1
1	220Ω resistor	K2
1	Battery clip to 2.1mm jack adapter	H9
1	Wooden spoon	
1	PP3 9V battery	

Construction

With a bit of thought, both of the accelerometer modules are capable of being plugged directly into the Arduino, as are the buzzer and LED. You should program the Arduino with the right sketch for the accelerometer module you are using before you attach the module, just in case some of the pins on the A0 to A5 connector are set to be outputs from a previous sketch.

Figure 8-17 shows the schematic diagram for the Arduino Egg and Spoon.

As you can see from Figure 8-18, all the components fit into the sockets on

FIGURE 8-16 An Arduino and spoon race

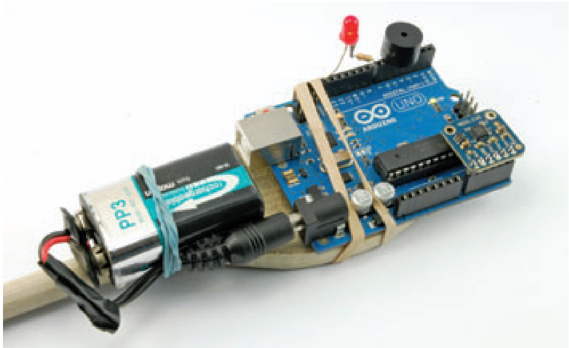

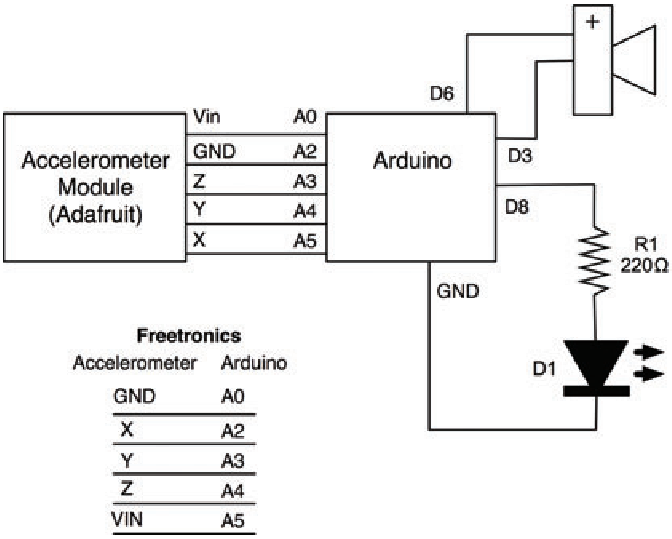

Freetronics

Accelerometer	Arduino
GND	A0
X	A2
Y	A3
Z	A4
VIN	A5

FIGURE 8-17 The schematic diagram for the Arduino Egg and Spoon

the Arduino. The LED/resistor combo is the same as we used in Chapter 6. The positive end goes to digital pin 8 on the Arduino and the negative end to GND. The buzzer fits between pins D3 and D6—D6 being connected to the positive end of the buzzer. If the pins on your buzzer are at a different spacing, then you can pick other pins, but remember to change the variables "gndPin2" and "buzzerPin" to whatever pins you end up using.

Both of the accelerometer modules will fit in the Arduino sockets A0 to A5, as shown in Figure 8-18. However, their pin allocations are quite different.

The project is powered from a 9V battery using an adapter, and the Arduino and battery are attached to the spoon with rubber bands.

FIGURE 8-18 The components attached to the Arduino

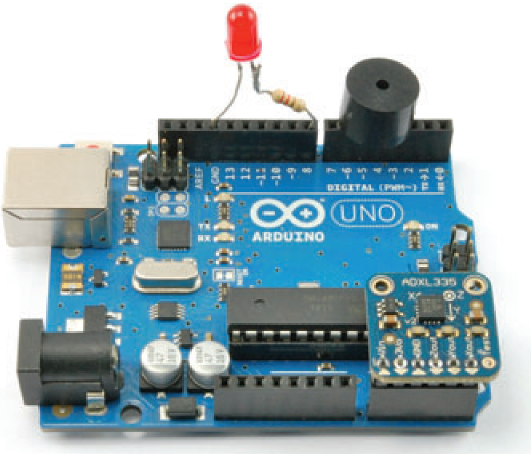

Software

There are two versions of the sketch provided: "egg_and_spoon_adafruit" and "egg_and_spoon_freetronics". Make sure you get the right one, and then program the Arduino with it BEFORE you attach the accelerometer.

The only difference between the two sketches is the pin allocations.

This is the sketch for the Adafruit version.

We start by defining the pins used.

```
// egg_and_spoon_adafruit
int gndPin1 = A2;
int gndPin2 = 3;
int xPin = 5;
int yPin = 4;
int zPin = 3;
int plusPin = A0;
int ledPin = 8;
int buzzerPin = 6;
```

The two variables “levelX” and “levelY” are used to measure the resting values of acceleration for X and Y if the spoon is level.

```
int levelX = 0;
int levelY = 0;
```

The “ledThreshold” and “buzzerThreshold” can be adjusted to set the degree of wobble before the LED lights and the buzzer sounds to indicate a “dropped egg.”

```
int ledThreshold = 10;
int buzzerThreshold = 40;
```

The “setup” function initializes the pins and then calls the function “calibrate” that sets the values of “levelX” and “levelY”.

```
void setup()
{
  pinMode(gndPin1, OUTPUT);
  digitalWrite(gndPin1, LOW);
  pinMode(gndPin2, OUTPUT);
  digitalWrite(gndPin2, LOW);
  pinMode(plusPin, OUTPUT);
  pinMode(ledPin, OUTPUT);
  pinMode(buzzerPin, OUTPUT);
  digitalWrite(plusPin, HIGH);
  calibrate();
}
```

In the main loop, we read the X and Y accelerations and see how much they have strayed from the values of “levelX” and “levelY”. The “abs” function returns the absolute value of a number, so if the difference is negative, it is turned into a positive value, and it is this that is compared with the thresholds that have been set.

```
void loop()
{
  int x = analogRead(xPin);
  int y = analogRead(yPin);
  boolean shakey = (abs(x - levelX) > ledThreshold || abs(y - levelY) >
ledThreshold);
  digitalWrite(ledPin, shakey);
  boolean lost = (x > levelX + buzzerThreshold || y > levelY + buzzerThreshold);
  if (lost)
  {
    tone(buzzerPin, 400);
  }
}
```

The only complication in the “calibrate” function is that we must wait for 200 milliseconds before we can take the readings. This gives the accelerometer time to turn on properly.

```
void calibrate()
{
  delay(200); // give accelerometer time to turn on
  levelX = analogRead(xPin);
  levelY = analogRead(yPin);
}
```

How to Sense Magnetic Fields

Sensing a magnetic field is made easy using a three-pin sensor IC like the A1302 linear hall effect sensor. You can use this chip in very much the same way as we did the TMP36 temperature sensor in the section “How to Measure Temperature” earlier in this chapter.

You Will Need

To use this temperature measurement IC, you will need the following items.

Quantity	Item	Appendix Code
1	Arduino Uno/Leonardo	M2/M21
1	USB lead; Type B for Uno, Micro USB for Leonardo	
1	A1302 linear hall effect sensor	S12

Construction

Just like the TMP36, the A1302 has just three pins, two for the power supply and one analog output. The power supply needs to be between 4.5V and 6V, making it ideal for use with the 5V of an Arduino.

In fact, we can supply the power to it through digital outputs and just plug the whole chip into three pins on the analog connector of the Arduino (Figure 8-19). The chip should be oriented with the dot facing outward.

Program the Arduino with the sketch before you plug in the sensor, in case A1 is set to be an output.

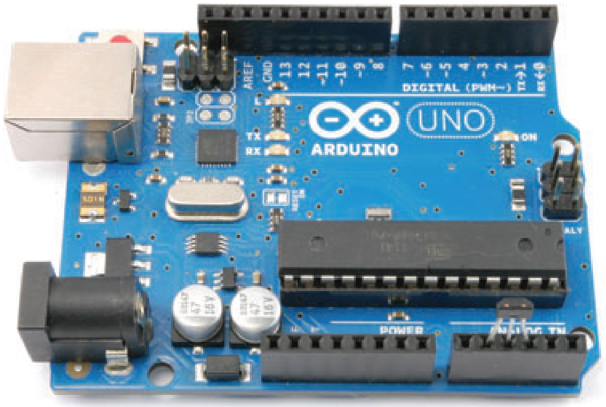

FIGURE 8-19 The A1302 magnetic sensor attached to an Arduino

Software

The sketch for the magnetic sensor is very similar to that of the temperature sensor.

First, the three pins are set up: digital pins 15 and 17 (A0 and A2), and A1 is set as the sensor pin.

```
// magnetic_sensor

int gndPin = A1;
int sensePin = 2;
int plusPin = A3;

void setup()
{
  pinMode(gndPin, OUTPUT);
  digitalWrite(gndPin, LOW);
  pinMode(plusPin, OUTPUT);
  digitalWrite(plusPin, HIGH);
  Serial.begin(9600);
}
```

The main loop just takes the raw reading and sends it to the Serial Monitor.

The device is not terribly sensitive, but if you hold a magnet next to it you should see a change in the reading coming from the Serial Monitor.

```
void loop()
{
  int raw = analogRead(sensePin);
  Serial.println(raw);
  delay(1000);
}
```

Summary

There are many other sensors out there, and many will interface to an Arduino quite easily using an analog input, or employing pulse length, letting you adapt the sketches used for other sensors to different sensors.

In the next chapter, we will change tack and look at sound and audio electronics.

9

Audio Hacks

In this chapter, you will look at audio electronics and find out how to make and amplify sounds so you can drive a loudspeaker.

You will also discover how to hack an FM transmitter intended for use with MP3 players in the car, so that it works as a surveillance bug.

First though, we will look at the more mundane topic of audio leads, how to use them, mend them, and make your own.

Hacking Audio Leads

Ready-to-use audio leads are pretty cheap to buy unless you go for the high-end connectors. Sometimes though, if you need a lead in a hurry, or an unusual lead, it helps to know how to wire one up from parts in your junk box or from connectors you have bought.

Many items of consumer electronics are supplied with a range of leads, and you do not always need them for use with the item you bought. Keep them in your junk box since you never know when you might need to make some kind of lead.

Figure 9-1 shows a selection of audio plugs, some designed to have leads soldered to them, and others that have plastic moldings around the lead and cable, which cannot be soldered to. Plugs with plastic moldings around them are still useful, however. It just means you will have to cut and strip the wire that leads to the plug rather than solder it to the plug itself.

General Principals

Audio leads carry audio signals, often on their way to an amplifier, and the last thing you want is for them to pick up electrical noise that will affect the quality of the sound. For this reason, audio leads are normally screened (see Figure 9-2).

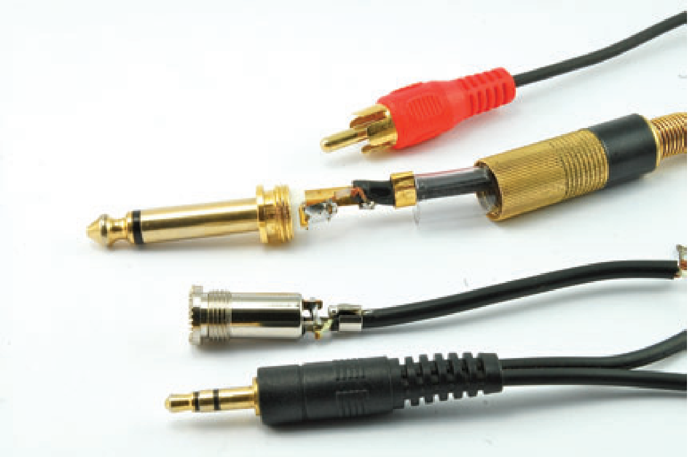

Figure 9-1 A variety of audio plugs

The audio signal itself (or two audio signals for stereo) is carried on insulated multi-core wires that are then enclosed in an outer conductive sheath of screening wire that carries the ground connection.

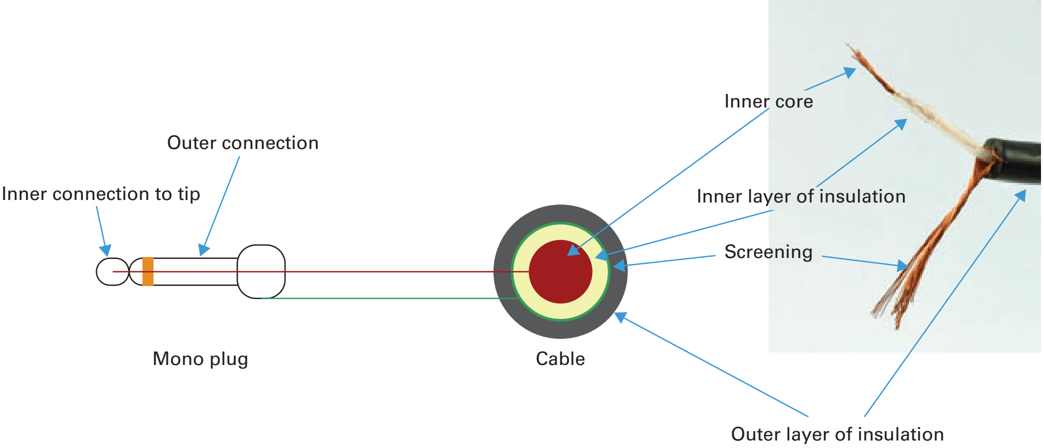

Figure 9-2 A screened audio lead

Inner core
Outer connection
Inner connection to tip
Inner layer of insulation
Screening
Mono plug
Cable
Outer layer of insulation

The exception to this is for leads to loudspeakers. These are not screened because the signal has been amplified to such a degree that any noise the speaker cables might pick up would be undetectable.

Soldering Audio Connectors

Stripping audio connectors is made more difficult by the fact that there is more than one layer of insulation. It is very easy to accidentally cut through the shielding. Nicking the outer insulation all around before stripping it will usually help with this problem.

Figure 9-3 shows the sequence involved in soldering a screened lead to a 6.3mm jack plug of the sort often used to connect an electric guitar to its amplifier.

The first step is to strip off the outer insulation about 20mm (a bit less than an inch) from the end of the lead and tease the shielding wires around to one side of the lead and twist them together. Strip about 5mm of insulation off the inner core insulation (Figure 9-3a). Then, tin both bare ends (Figure 9-3b).

The jack plug has two solder tags: one for the outer part of the plug and one connected to the tip. Both will usually have holes in them. Figure 9-3c shows the screening trimmed to a shorter length and pushed through the hole ready to solder. Once the screening is soldered into place, solder the inner core to the solder tag for the tip (Figure 9-3d).

These wires are quite delicate, so make sure the inner core wire has some extra length (as shown in Figure 9-3e) so that if the plug flexes, it will not break the connection. Notice that the strain relief tabs at the end of the plug have been pinched around the outer insulation. Finally, the plug will often have a plastic sleeve that protects the connections. Slide this over the connections and then screw in the plug casing.

Tip

If there is a plug on the other end of the lead, remember to push the new plug enclosure and plastic sleeve onto the lead BEFORE you solder the second plug on, otherwise you will end up having to unsolder everything to put it on. The author has made this mistake more times than he cares to admit.

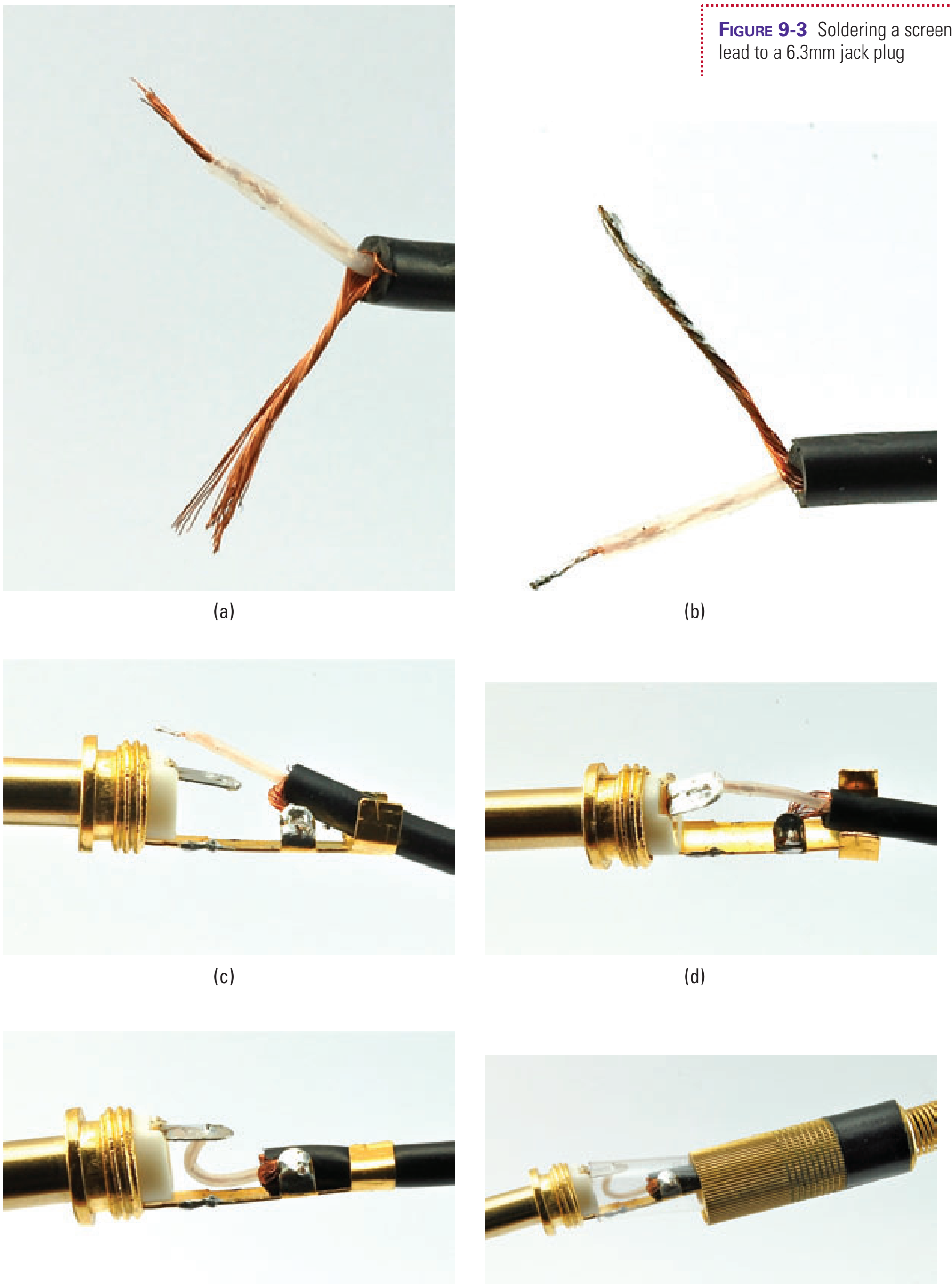

(a) (b)

(c) (d)

(e) (f)

Figure 9-3 Soldering a screened lead to a 6.3mm jack plug

Converting a Stereo Signal to Mono

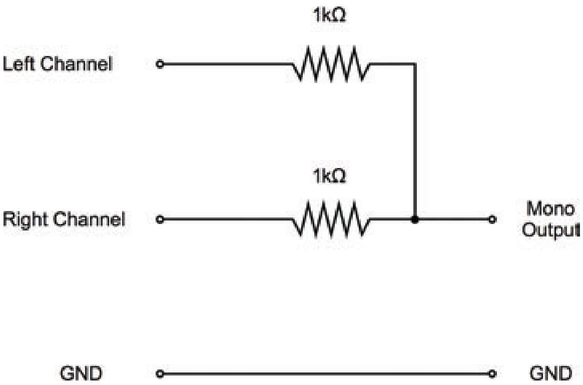

FIGURE 9-4 Mixing stereo to mono

Stereo audio is made up of two slightly different audio signals that give the stereo effect when played through two separate speakers. Sometimes, you have a stereo output that you want to input to a single channel (mono) amplifier.

You could use just one of the channels of the stereo signal (say, the left channel), but then you will lose whatever is on the right channel. So a better way of converting stereo to mono is to use a pair of resistors to mix the two channels into one (Figure 9-4).

Looking at the schematic of Figure 9-4, you could be forgiven for thinking all you need to do is connect the left and right channels to each other directly. This is not a good idea, because if the signals are very different, there is the potential for a damaging current to flow from one to the other.

As an example, we could use the mono 6.3mm jack we just soldered leads to, and combine it with a pair of resistors and a stereo 3.5mm jack plug so we could, for example, plug an MP3 player into a guitar practice amplifier.

Figure 9-5 shows the steps involved in this. To make it easier to photograph, the author's lead is made very short. You will

FIGURE 9-5 Making a lead

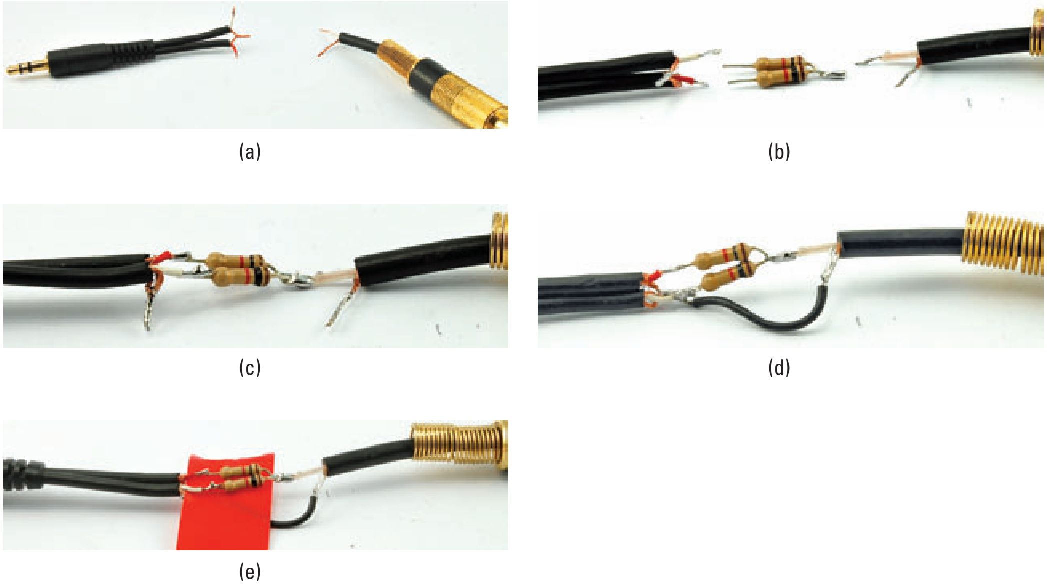
(a) (b) (c) (d) (e)

probably want to make yours longer. This is not a problem, unless you plan to make it longer than a few yards or meters.

The 3.5mm plug is of the plastic molded variety, reclaimed from some unwanted lead. The first step is to strip both leads (Figure 9-5a). Note that the stereo plug has two screened connections in one twin cable. The screened ground connections of both channels of the stereo plug can be twisted together.

Tin the ends of all the leads, and then solder the resistors together, as shown in Figure 9-5b.

Next, solder the stereo and mono leads to the resistors, as shown in Figure 9-5c, and cut and tin a short length of wire to bridge the ground connections. Solder it into place (Figure 9-5d) and then wrap everything in insulating tape, taking care to put tape in between any places where wires could short together (Figure 9-5e).

Figure 9-6 A microphone module

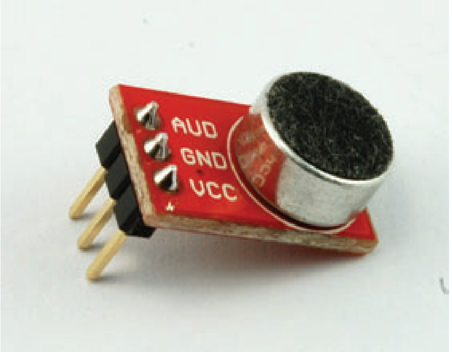

How to Use a Microphone Module

Microphones (mics) respond to sound waves, but sound waves are just small changes in air pressure, so it is not surprising that the signal you get from a mic is usually very faint. It requires amplification to bring it up to a useable level.

While it is perfectly possible to make a little amplifier to boost the signal from your mic, you can also buy a mic module that has an amplifier built in. Figure 9-6 shows such a module.

The mic module just requires a supply voltage between 2.7V and 5.5V. This makes it ideal for interfacing with an Arduino.

In Chapter 11, you will find out a bit more about oscilloscopes. But for now, here is a sneak preview of what an oscilloscope will display (see Figure 9-7) when connected to the mic module while a constant tone is being generated and the module is supplied with 5V.

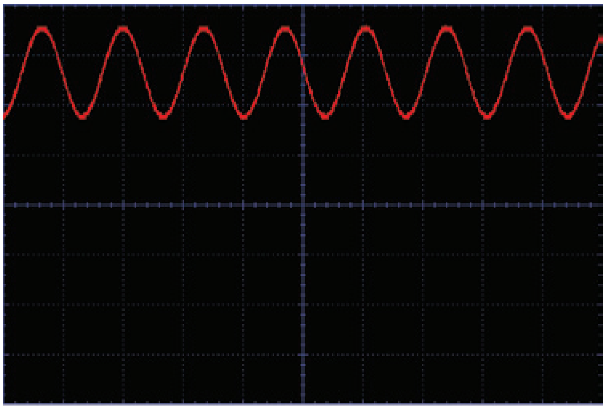

Figure 9-7 The output of the microphone module

The oscilloscope is displaying the sound. In this case, a constant and rather irritating tone of 7.4 kHz. The horizontal axis is time, and each blue square represents 100 microseconds. The vertical axis is the voltage and each square is 1V. The output of the mic module is a voltage that varies very quickly between about 1.8V and 3.5V. If you draw a horizontal line straight down the middle of the waveform, it would be at about 2.5V. This is halfway between 0V and 5V. So if there is no sound at all, there will just be a flat line at 2.5V, and as the sound gets louder, the waveform will swing further and further either way. It will not, however, go higher than 5V or lower than 0V. Instead, the signal will clip and become distorted.

Figure 9-8 The schematic diagram for a mic module

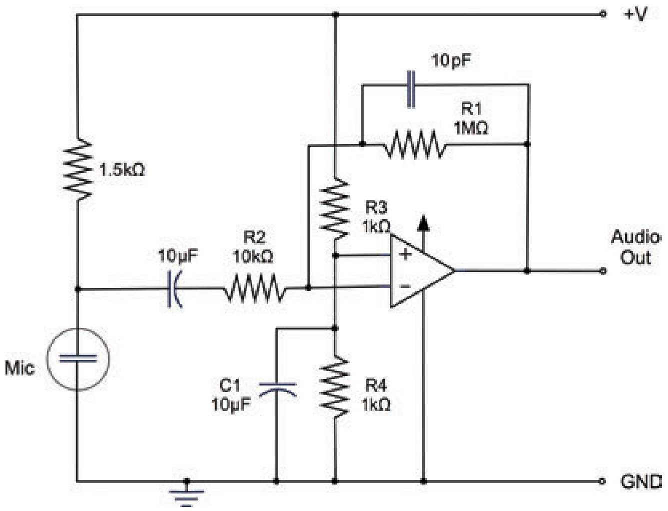

The mic module shown is sold by SparkFun (BOB-09964). The schematic for this, along with all its design files have been made public. Figure 9-8 shows a schematic for a typical microphone pre-amp.

The chip at the center of this design has a similar circuit symbol to the comparator you used in the "How to Detect Noxious Gas" section at the beginning of Chapter 8. However, it is not a comparator; it is an amplifier IC of a type known as an "operational amplifier" (or "op amp" for short).

Whereas a comparator turns its output on when the "+" input is higher than the "–" input, an op amp amplifies the difference between the "+" and "–" inputs. Left to its own devices, it amplifies this by a factor of millions. This means that the tiniest signal or noise on the input would be turned into meaningless thrashing of the output from 0 to 5V. To tame the op amp and reduce its amplification factor (called "gain"), something called "feedback" is used.

The trick is to take a portion of the output and feed it into the negative input of the op amp. This reduces the gain to an amount determined by the ratio of R1 to R2, shown in Figure 9-8. In this case, R1 is 1MΩ and R2 is 10kΩ, so the gain is 1,000,000 / 10,000 or 100.

The signal from the microphone is being amplified by a factor of 100. This shows just how weak the signal is in the first place.

The “+” input to the op amp is held halfway between GND and 5V (2.5V) by using R3 and R4 as a voltage divider. C1 helps to keep this constant.

From the schematic, you can see how you could build the module yourself on, say, stripboard. Op amps like the one used (which is a surface-mounted device) are also available in the eight-pin DIP form. However, a module like this will save you a lot of effort and may even turn out cheaper than buying and building a module from scratch.

I realize this is a rather cursory introduction to op amps. These are very useful devices, but unfortunately require more space to explain fully than this book can accommodate. You will find good information on op amps at the Wikipedia site, as well as in books with a more theoretical bent like *Practical Electronics for Inventors*, Third Edition, by Paul Sherz and Simon Monk, which has a chapter devoted exclusively to op amps.

In the next section, we will combine this module with a hacked FM transmitter of the sort used to let you play your MP3 player through your car radio, thus creating an audio “bug.”

How to Make an FM Bug

To make an FM transmitter that will broadcast sound picked up from a microphone to a nearby FM radio receiver would require a lot of effort. We are hackers, so we are going to cheat and take apart an FM transmitter and wire it up to a mic module. Figure 9-9 shows the end result of this hack.

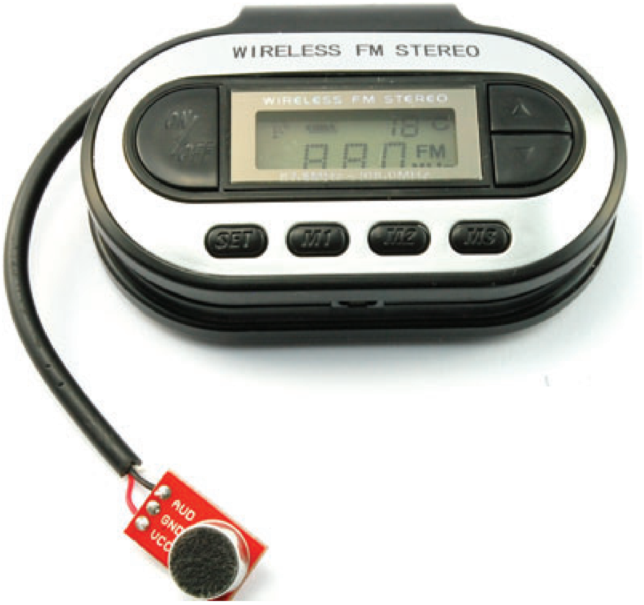

FIGURE 9-9 An FM radio bug

You Will Need

To build the bug, you will need the following.

Quantity	Item	Appendix Code
1	Microphone module	M5
1	* FM transmitter for MP3 players	
1	FM radio receiver	

* For suitable FM transmitters, try searching on eBay using the search terms "fm transmitter mp3 car." Expect to pay about USD 5 and look for the most basic of models. You do not need remote control or an SD card interface. You just want something that has an audio input lead, and for simplicity purposes runs on two AA or AAA batteries (3V).

Construction

This is a very easy project to make. Figure 9-10 shows the schematic diagram for the bug.

The 3V battery of the FM transmitter is used to provide power to the mic module, and the single output of the mic module is connected to both the left and right inputs of the stereo FM transmitter.

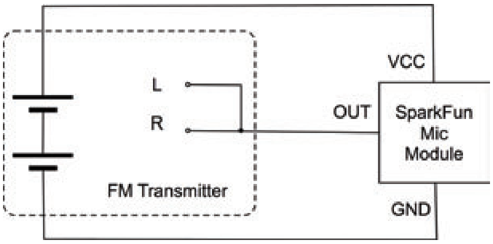

Figure 9-10 Schematic diagram for the radio bug

Figure 9-11 shows how the FM transmitter is modified to connect the mic module to it.

The first step is to unscrew any screws that hold the case together and pull it apart. Then, chop off the plug, leaving most of the lead in place since the lead often doubles as an antenna in these devices. Strip and tin the three wires inside the lead (Figure 9-11a).

Looking at the three wires, in Figure 9-11a, the red wire is the right signal, the white the left, and the black ground. This is a common convention, but if you are not sure it applies to your transmitter, you can check by stripping the wires on the plug end of the lead you cut off and using the continuity setting on your multimeter to see which lead is connected to what on the plug. The farthest tip and next ring should be the left and right signals, and the metal nearest the plastic should be the ground connection.

We are going to leave the ground and left connections as they are, but disconnect the wire and connect it to the 3V connection of the battery (Figure 9-11b). In this transmitter, the positive terminal of the battery box underneath the PCB is soldered to the top surface of the PCB.

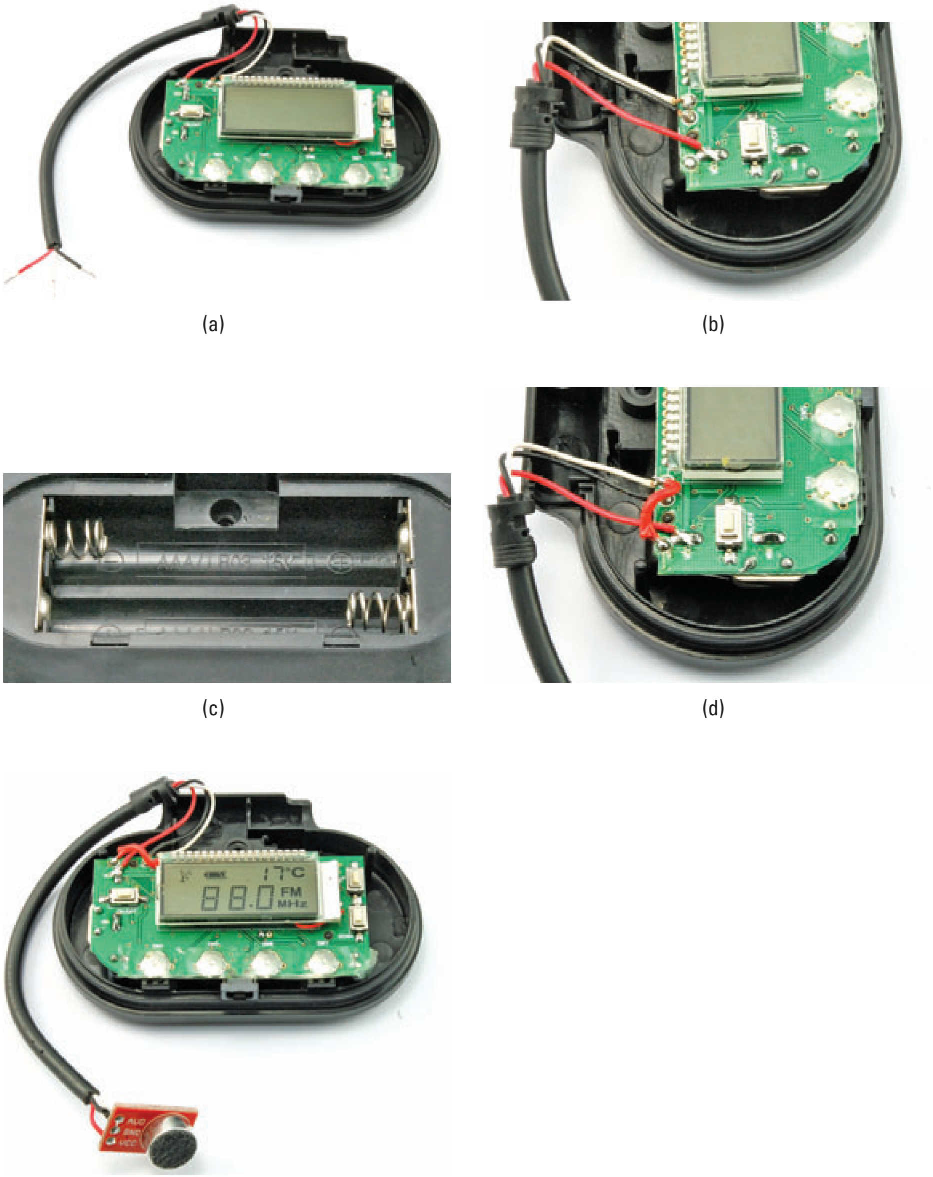
17°C
88.0 FM MHz
AUD
GND
VCC

(a) (b) (c) (d) (e)

Figure 9-11 Modding the FM transmitter

To find the positive connection, look carefully at the battery box. In Figure 9-11c, you can see that the metal piece on the left of the figure links the negative of the top cell to the positive of the bottom cell. The 3V connection will therefore be the top right connection of the battery box, so trace where this comes out on the top of the PCB. If it is attached by wires, then find an appropriate place for the red wire of the audio jack lead to be joined to it.

Referring back to the schematic diagram of Figure 9-10, we need to make a little wire just to link the left and right channels (Figure 9-11d). When all the changes are complete, it should look like Figure 9-11e.

Testing

Note that the on/off button of the transmitter will have no effect on the power going to the mic module. So to fully turn off the bug, remove the batteries.

To test the module out, set the frequency of the FM transmitter to one not occupied by a radio station and then set the radio receiver to the same frequency. You may well hear the howl of feedback through the radio. To prevent that, take the radio receiver to a different room. You should find that you can hear what is happening in the room with the bug in it pretty clearly.

Selecting Loudspeakers

Loudspeakers have remained largely unchanged in design since the early days of radio. Figure 9-12 shows how a loudspeaker works.

The cone (often still made of paper) has a light coil around the end that sits within a fixed magnet attached to the frame of the loudspeaker. When the coil is driven by an amplified audio signal, it moves toward and away from the magnet in time with the audio. This creates pressure waves in the air, producing a sound.

Electronically speaking, a loudspeaker just looks like a coil. When you buy a speaker like this, it will have a number of ohms associated with it. Most speakers are 8Ω, but you can also commonly find 4Ω and 60Ω speakers. If you measure the resistance of the coil of an 8Ω speaker, you should find that it is indeed about 8Ω.

FIGURE 9-12 How a loudspeaker works

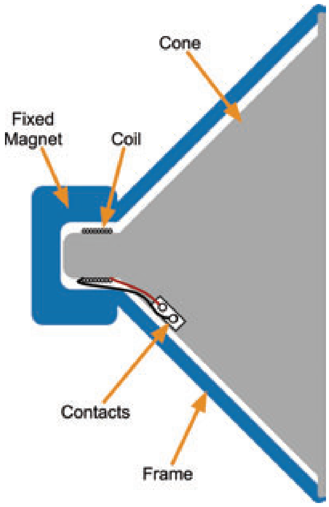

Another figure that is normally stated with the speaker is the power. This specifies how hard the loudspeaker can be driven before the coil will get too hot and burn out. For a small loudspeaker such as one you might put in a small radio receiver, values of 250 mW and up are not untypical. As you progress toward the kind of speakers you would use with a hi-fi set, you will see figures in the tens of watts, or even hundreds of watts.

It is very hard to build speakers that can cover the whole range of audio frequencies, which is generally standardized as 20 Hz up to 20 kHz. So you will often find hi-fi speakers that group a number of speakers into a single box. This might be a "woofer" (for low frequencies) and a "tweeter" (for high frequencies). Because woofers cannot keep up with the high frequencies, a module called a "crossover network" is used to separate the low and high frequencies and drive the two types of speakers separately. Sometimes this is taken a step further and three drive units are used: one for bass, one for mid-range tones, and a tweeter for high frequencies.

The human ear can pick out the direction of a high-frequency sound very easily. If you hear a bird tweeting in a tree, you will probably be able to look straight at it without having to think about where it is. The same is not true of low frequencies. For this reason, surround-sound systems often have a single low-frequency "woofer" and a number of other speakers that handle midrange and higher frequencies. This makes life easier, because bass speakers have to be much larger than higher-frequency units in order to push large amounts of air about relatively slowly to produce bass sounds.

How to Make a 1-Watt Audio Amplifier

Building a small amplifier to drive a loudspeaker is made easier by an IC like the TDA7052, which contains pretty much all the components you need, on a chip costing less than $1. In this section, you will make a little amplifier module on stripboard (Figure 9-13).

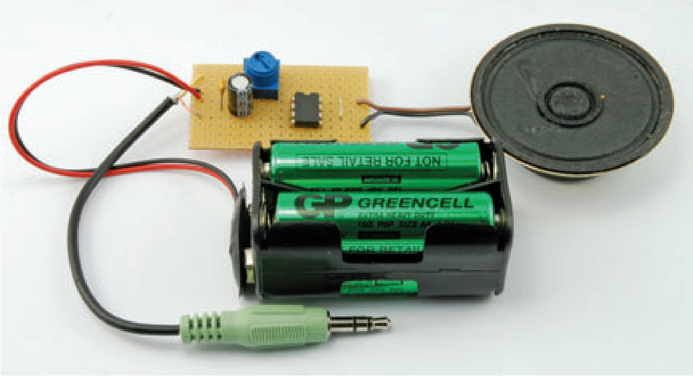

FIGURE 9-13 A 1-watt amplifier module

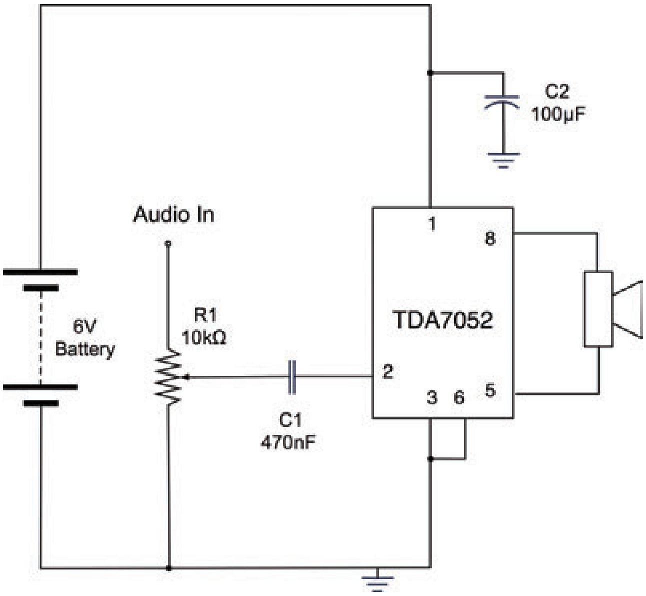

FIGURE 9-14 A typical TDA7052 amplifier schematic

An alternative to making your own amplifier is to buy a ready-made module. You will find these available for a wide range of different powers and in mono and stereo configurations. eBay is a good source for such modules, as are SparkFun (BOB-11044) and Adafruit (product ID 987). These modules often use an advanced type of design called "class-D," which is far more efficient in its use of energy than the module we are going to build.

Figure 9-14 shows the typical schematic for a TDA7052 amplifier.

R1 acts as a volume control, reducing the signal before amplification.

C1 is used to pass the audio signal on to the input to the amplifier IC without passing on any bias voltage that the signal may have from the audio device producing the signal. For this reason, when you use a capacitor like this, it is called a coupling capacitor.

C2 is used to provide a reservoir of charge that can be drawn on quickly by the amplifier when it needs it for very rapid changes in the power supplied to the speaker. This capacitor should be positioned close to the IC.

You Will Need

To build the amplifier module, you will need the following.

Quantity	Name	Item	Appendix Code
1	IC1	TDA7052	S9
1	R1	10kΩ variable resistor	K1, R1
1	C1	470nF capacitor	C3
1	C2	100µF capacitor	K1, C2
1		8Ω speaker	H14
1		Stripboard	H3

Construction

Figure 9-15 shows the stripboard layout for the amplifier module. If you have not used stripboard before, read through the section titled “How to Use Stripboard (LED Flasher)” in Chapter 4.

To build the module, follow the steps shown in Figure 9-16.

First, cut the stripboard to size and make the three cuts in the tracks using a drill bit (Figure 9-16a).

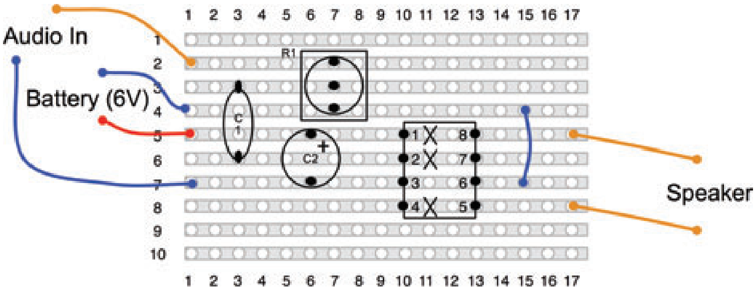
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
1 2 3 4 5 6 7 8 9 10
Audio In
Battery (6V)
R1
C1
C2
+
1 2 3 4 5 6 7 8
Speaker

Figure 9-15 The stripboard layout for an amplifier module

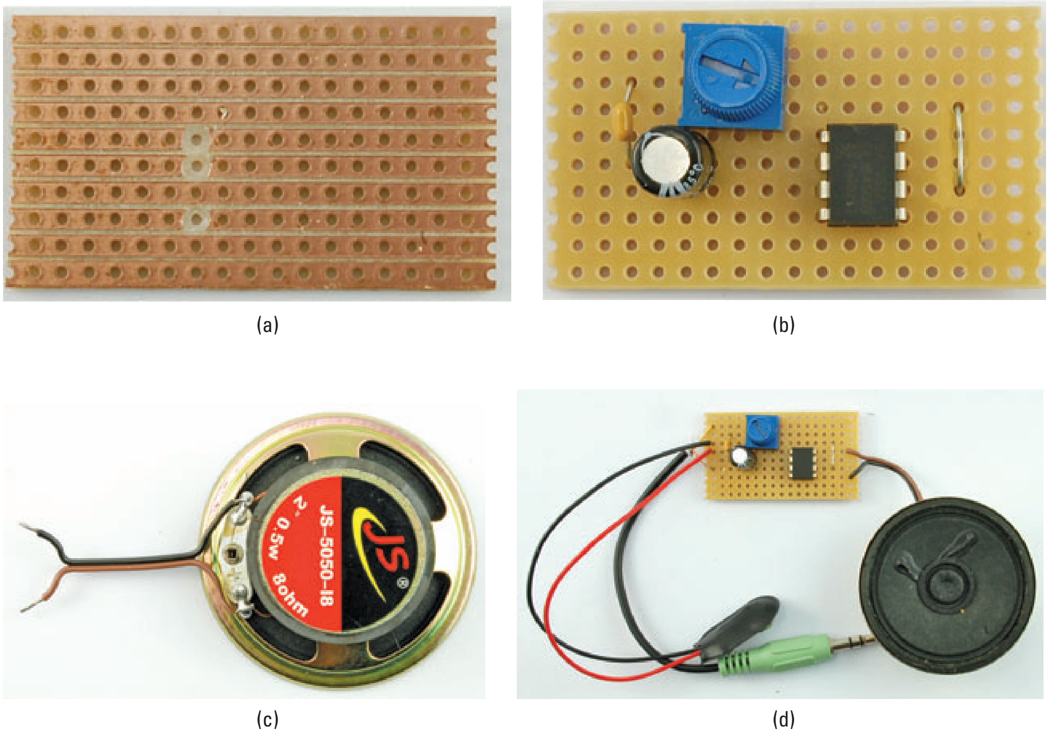
JS
JS-5050-I8
2" 0.5w 8ohm

(a) (b)

(c) (d)

Figure 9-16 Building the audio amplifier module

The next step is to solder the link into place, and then the IC, C1, C2, and R1 in that order (Figure 9-16b). It is easiest to solder the components that are lowest to the board first.

Attach leads to the speaker (Figure 9-16c) and finally attach the battery clip and a lead ending in a 3.5mm stereo jack plug (Figure 9-16d). Note that only one channel of the audio lead is used. If you want to use both left and right channels, you should use a pair of resistors (see the section "Hacking Audio Leads" at the beginning of this chapter).

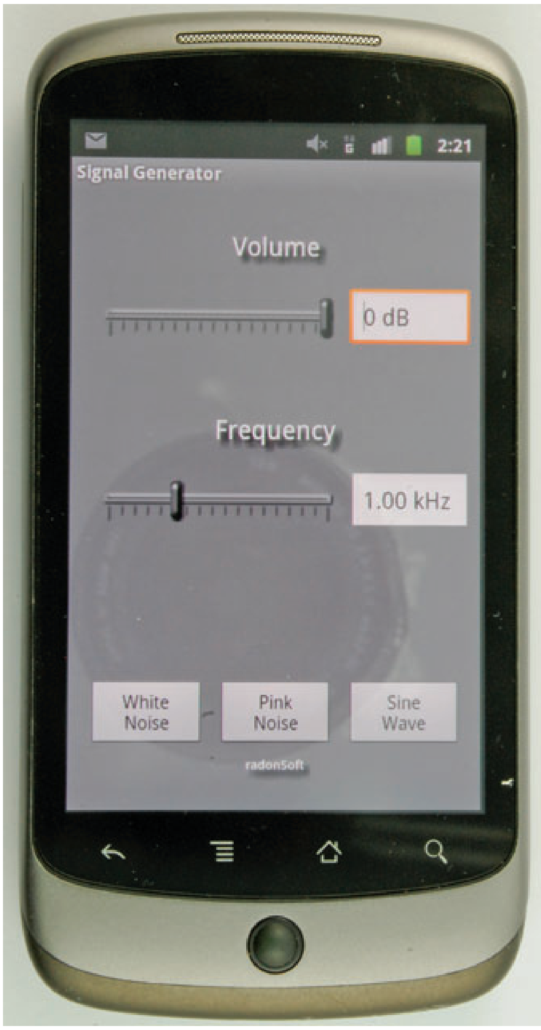

Signal Generator

Volume: 0 dB

Frequency: 1.00 kHz

FIGURE 9-17 A signal generator app

Testing

You can try the amplifier out by plugging it into an MP3 player, or, if you have an Android phone or an iPhone, download a signal generator app like the one shown in Figure 9-17. There are a number of such apps, many of them free, including this one for Android from RadonSoft.

With this, you can play a tone at a frequency you select. By noting when the volume of the speaker starts to drop off, you can work out the useful frequency range of your amplifier module.

How to Generate Tones with a 555 Timer

Back in Chapter 4, you used a 555 timer to blink a pair of LEDs. In this section, we will see how to use a 555 timer IC oscillating at much higher frequencies to generate audio tones.

The pitch will be controlled using a light-dependent resistor (LDR) so that as you wave your hand over the light sensor, the pitch will change in a theremin-like manner.

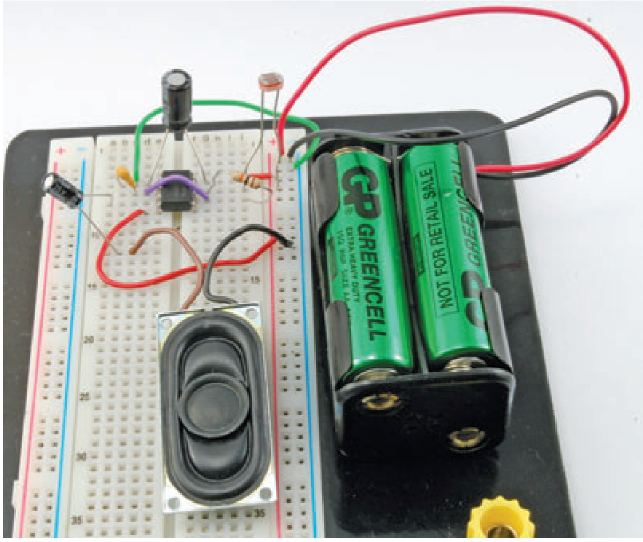
GP GREENCELL
EXTRA HEAVY DUTY
NOT FOR RETAIL SALE
GP GREENCELL

FIGURE 9-18 Generating tones with a 555 timer IC

Figure 9-18 shows the tone generator built onto breadboard.

Figure 9-19 shows the schematic diagram for the tone generator.

This is similar to the design of the LED flasher in Chapter 4. In this case, instead of two fixed resistors and a capacitor setting the frequency, R1 is the LDR, whose resistance will vary between about 1kΩ and 4kΩ depending on the light falling on it. We need a much higher frequency than our LED flashing circuit—in fact, if we aim for a maximum frequency of around 1 kHz, we need a frequency of about 1000 times what we had before.

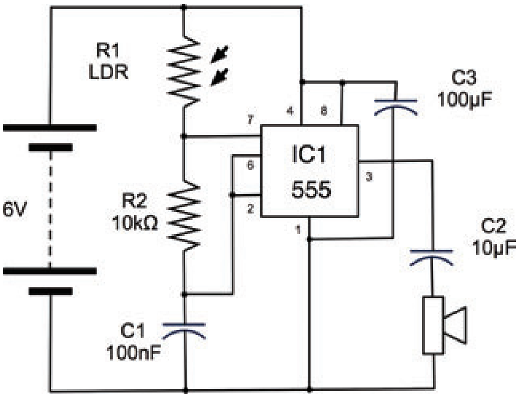
R1
LDR
6V
R2
10kΩ
C1
100nF
7
4
8
6
2
1
3
IC1
555
C3
100μF
C2
10μF

FIGURE 9-19 Schematic diagram for a 555 tone generator

The 555 timer oscillates at a frequency determined by the formula:

frequency = 1.44 / ((R1 + 2 * R2) * C)

where the units of R1, R2, and C1 are in Ω and F.

So, if we use a 100nF capacitor for C1, and R2 is 10kΩ, and R1 (the LDR) has a minimum frequency of 1kΩ, then we can expect a frequency of:

1.44 / ((1000 + 20000) * 0.0000001) = 686 Hz

If the LDR's resistance increases to 4kΩ, then the frequency will drop to:

1.44 / ((4000 + 20000) * 0.0000001) = 320 Hz

To calculate the frequency and when deciding what values of R1, R2, and C1 to use, there are online calculators like this one at www.bowdenshobbycircuits.info/555.htm that will calculate the frequency for you.

You Will Need

To build the amplifier module, you will need the following.

Quantity	Name	Item	Appendix Code
1	IC1	555 timer IC	K1, S10
1	R1	LDR	K1, R2
1	R2	10kΩ resistor	K2
1	C1	100nF capacitor	K1, C4
1	C2	10μF capacitor	K1, C5
1		8Ω speaker	H14

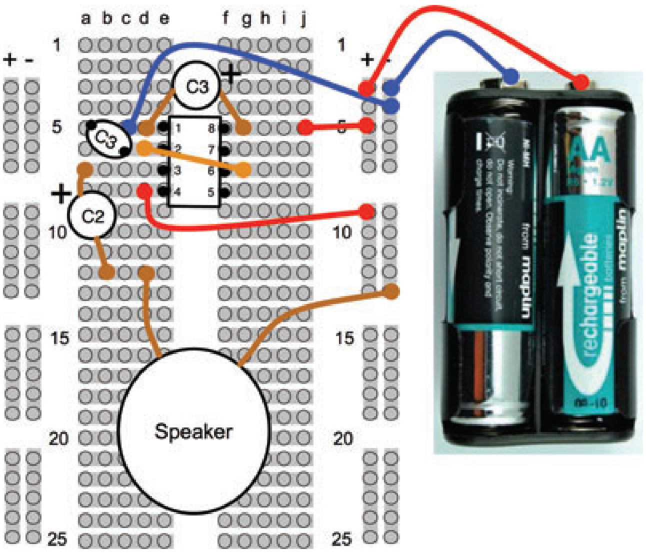

FIGURE 9-20 A signal generator app

Construction

Figure 9-20 shows the breadboard layout for the tone generator.

It would be quite straightforward to build this design onto stripboard. The stripboard layout in the section "How to Use Stripboard (LED Flasher)" in Chapter 4 would be a good starting point.

How to Make a USB Music Controller

Music software like Ableton Live™ is designed to allow USB controllers that emulate a keyboard to control virtual musical instruments and do all kinds of exciting things.

You can use the USB keyboard emulation features of the Arduino Leonardo with an accelerometer so that tilting the board produces a key press of a number between 0 and 8, with 4 being pressed if the board is level, 0 if tilted almost vertically to the right, and 8 being pressed when it is tilted the other way.

The only hardware on the Arduino is the accelerometer (Figure 9-21).

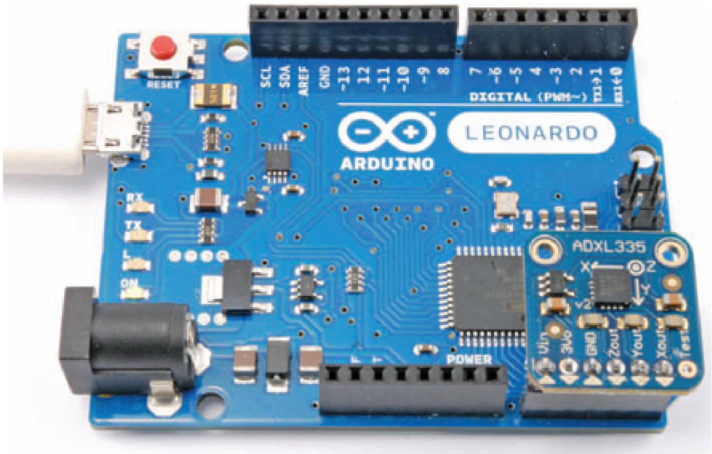

Figure 9-21 A USB music controller

You Will Need

To build this controller, you will need the following items.

Quantity	Item	Appendix Code
1	Arduino Leonardo	M21
1	Micro USB for Leonardo	
1	Accelerometer	M15 (Adafruit version)

Construction

There is actually very little to construct in this project. The schematic is actually the same as in the section titled "How to Use an Accelerometer" in Chapter 8. The Freetronics accelerator will also work, but you will need to change the pin assignments before attaching the accelerometer.

Software

The software for the music controller combines code for sensing the angle of tilt on the X-axis with emulating a keyboard press.

The first step is to assign the pins to be used. As in the section "How to Use an Accelerometer" in Chapter 8, the accelerometer module is powered from output pins.

```
// music_controller

int gndPin = A2;
int xPin = 5;
int yPin = 4;
int zPin = 3;
int plusPin = A0;
```

The variable “levelX” is used during calibration and holds the analog value when the accelerometer is flat.

The “oldTilt” variable contains the old value of the tilt of the board, which is a value between 0 and 8, where 4 means level. The old value is remembered, so that a key press is only sent if the tilt angle changes.

```
int levelX = 0;
int oldTilt = 4;
```

The “setup” function sets the output pins to power the accelerometer, calls “calibrate”, and starts the Leonardo keyboard emulation mode.

```
void setup()
{
  pinMode(gndPin, OUTPUT);
  digitalWrite(gndPin, LOW);

  pinMode(plusPin, OUTPUT);
  digitalWrite(plusPin, HIGH);
  calibrate();
  Keyboard.begin();
}
```

In the main loop, the accelerometer reading is converted to a number between 0 and 8, and if it has changed since the last reading, a key press is generated.

```
void loop()
{
  int x = analogRead(xPin);
  // levelX-70 levelX levelX + 70
  int tilt = (x - levelX) / 14 + 4;
  if (tilt < 0) tilt = 0;
  if (tilt > 8) tilt = 8;
  // 0 left, 4 is level, 8 is right
  if (tilt != oldTilt)
  {
      Keyboard.print(tilt);
      oldTilt = tilt;
  }
}
```

The “calibrate” function takes an initial reading of the acceleration on the X-axis, after waiting 200 milliseconds for the accelerometer to turn on properly.

```
void calibrate()
{
  delay(200); // give accelerometer time to turn on
  levelX = analogRead(xPin);
}
```

How to Make a Software VU Meter

The mic module you used in the section “How to Make an FM Bug” is also perfectly suited for use with microcontrollers like the Arduino. Figure 9-22 shows the module with pins attached to it and pushed into the analog connector strip of the Arduino.

The mic module can be used to measure the sound level and write a number of “*”s to the Serial Monitor to indicate the loudness of the sound (Figure 9-23).

Figure 9-22 Attaching a mic module to an Arduino

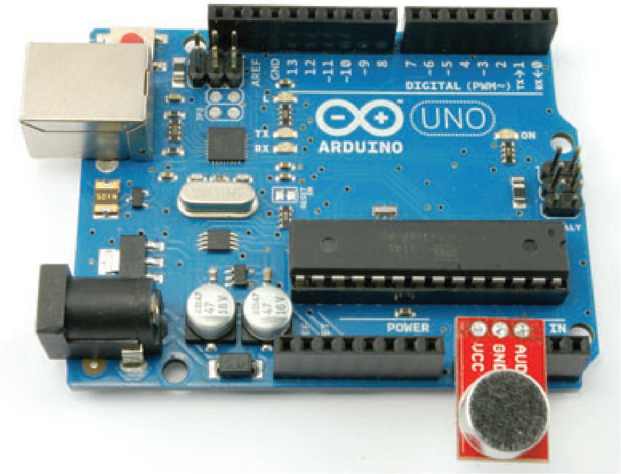

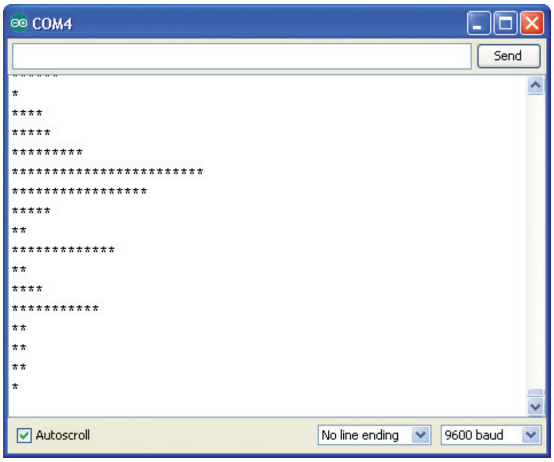

```
*
****
*****
*********
************************
*****************
*****
**
*************
**
****
***********
**
**
**
*
```

☑ Autoscroll No line ending 9600 baud

Figure 9-23 The Serial Monitor as a VU meter

You Will Need

To build this VU meter, you will need the following items.

Quantity	Item	Appendix Code
1	Arduino Uno/Leonardo	M2, M21
1	USB lead; Type B for Uno, Micro USB for Leonardo	
1	Mic module	M14
1	Header pins (three-way)	H4

Construction

Upload the sketch "vu_meter" before attaching the mic module.

Solder the header pins to the module so they will fit into the Arduino sockets A0 to A2 with the microphone facing outward, as shown in Figure 9-22.

Software

The mic module uses very little current, so for convenience we can use A0 and A1 to provide the power to it.

The sketch begins by defining the pins to use and setting them up in the "setup" function. Serial communication is also started here.

```
// vu_meter

int gndPin = A1;
int plusPin = A0;
int soundPin = 2;

void setup()
{
  pinMode(gndPin, OUTPUT);
  digitalWrite(gndPin, LOW);
  pinMode(plusPin, OUTPUT);
  digitalWrite(plusPin, HIGH);
  Serial.begin(9600);
}
```

The loop function reads the raw value for the analog input A2. The mic module produces an output of 2.5V when there is no signal, and swings above and blows that with the sound's waveform. So to find the "loudness" we need to first subtract 511 from the raw value—511 being equivalent to the 2.5V offset in the raw analog reading that spans from 0 to 1023.

The "abs" function makes any negative number positive and then divides the whole result by 10 to give us a number between 0 and 51 and assigns it to the variable "topLED". We are not actually using LEDs, but you could think of each "*" as being an LED illuminated on a bar graph display.

The "for" loop then prints a number of "*"s equal to the value held in "topLED". Finally, a new line is printed and we delay for 1/10th of a second.

```
void loop()
{
  int value = analogRead(soundPin);
  int topLED = 1 + abs(value - 511) / 10;

  for (int i = 0; i < topLED; i++)
  {
      Serial.print("*");
  }
  Serial.println();
  delay(100);
}
```

Summary

In addition to the how-tos just covered, there are lots of audio modules you can make use of. Low-cost stereo power amplifiers are available from eBay and suppliers like SparkFun and Adafruit.

You can also buy ultra-low-cost amplified speakers intended for computers and reuse them in your projects.

10

Mending and Breaking Electronics

In this chapter, we will look at taking things apart and putting them back together again, or just taking them apart to salvage components.

In today's throw-away society, many consumer electronics items that stop working go directly into the garbage. Economically, they are simply not worth paying someone to repair. However, that does not mean it is not worth *trying* to repair them. Even if the attempt fails, some serviceable components may be scavenged for use in your projects.

How to Avoid Electrocution

When working on something that is powered by household electricity, NEVER work on it when it is plugged into the outlet. I actually like to have the electrical plug for the appliance right in front of me, so that I know it is not plugged in. Household electricity kills many people every year. Take it seriously!

Some devices, such as switch mode power supplies, contain high-value capacitors that will hold their charge for hours after the device has been unplugged. These capacitors are simply biding their time, waiting for some unsuspecting fingers to complete the circuit.

Unless it is a very small capacitor, it should not be discharged by shorting the leads with a screwdriver. A large capacitor at high voltage can supply huge amounts of charge in a fraction of a second, melting the end of the screwdriver and flinging molten metal around. People have been blinded by capacitors exploding in this manner, so don't do it.

Figure 10-1 shows the safe way to discharge a capacitor.

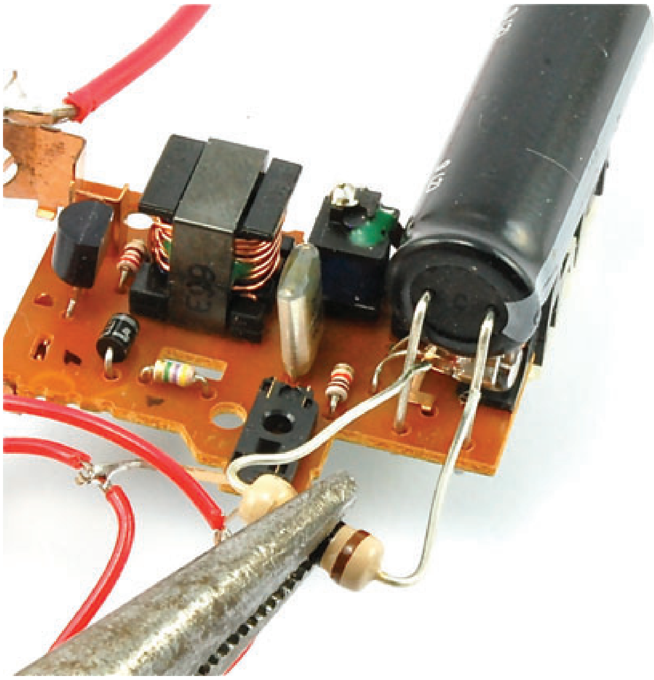

Figure 10-1 Safely discharging a capacitor

The legs of a 100Ω resistor are bent to the right fit for the capacitor contacts and held in the teeth of a pair of pliers for a few seconds. You can use the highest setting of your voltmeter to check that the capacitor has discharged to a safe level (say, 50V). If you have a high-wattage resistor, all the better. If it is not high enough power, it will break, but not in as spectacular a way as a capacitor being discharged dangerously.

Some devices that can pack a painful and sometimes lethal punch are:

- Old glass CRT TVs
- Switch-mode power supplies
- Camera flash guns and disposable cameras with a flash

How to Take Something Apart AND Put It Back Together Again

It is often said that “any fool can take something apart, but putting it back together is a totally different matter.”

Just remember that taking things apart usually voids their warranty.

By following a few simple rules, you shouldn't have any problems.

- Have a clear working area with lots of room.
- As you take out the screws, place them in the same pattern on your worktop as they were in the case they came out of. Sometimes the screws can be different sizes. If they are likely to be knocked or roll about on the surface, then push them into a piece of expanded polystyrene or something similar.
- After undoing the screws, when you come to take the case apart, watch out for any little plastic bits like switch buttons that might fall out. Try and keep them in place until you are ready to remove them.
- If something looks tricky, draw a sketch or take a photograph. (I tend to take a lot of photographs when repairing things, like with a hair dryer or straighteners, that have a large mechanical design component.)
- Try not to force things apart. Look to see where the clips are.
- If all else fails, try cutting the case apart with a handsaw (something your author has resorted to in the past), and then later glue the case back together.

How to Check a Fuse

The most convenient problem to fix in an appliance is the fuse. It's convenient because it is easy to test and easy to fix. Fuses are basically just wires designed to burn out when the current flowing through them gets too high. This prevents further damage to more expensive components, or can stop a fire from starting.

Sometimes fuses are clear, so you can see that the wire inside them has broken and that they have "blown." Fuses are rated in amps and will generally be labeled to show the maximum current in A or mA they can take. Fuses also come as "fast blow" and "slow blow." As you would expect, this determines how fast the fuses react to over-current.

Some household electrical plugs contain a fuse holder, and you can also find fuses on PCBs. Figures 10-2a-c show the

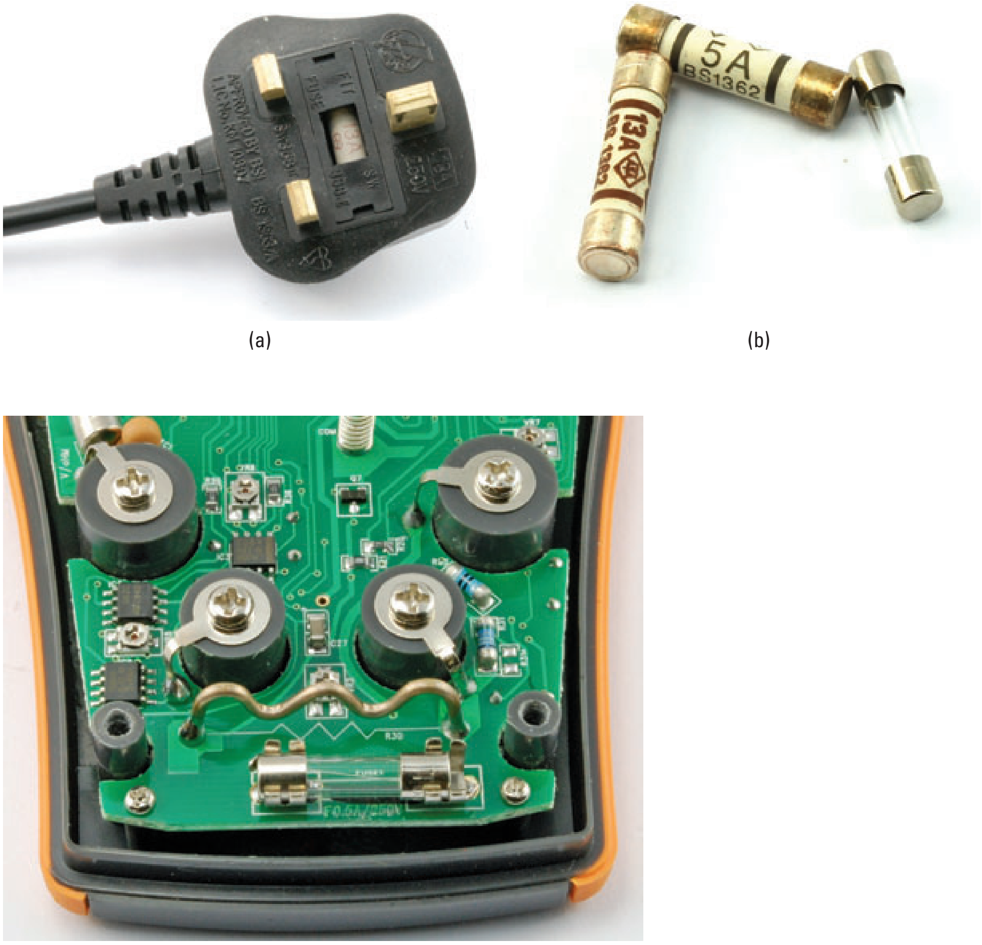

(a) (b)

(c)

Figure 10-2 Fuses

inside of a UK fused plug and also a fuse holder on the PCB for the author's multimeter.

You have used your multimeter in Continuity mode enough times now that you can probably guess how to test a fuse (Figure 10-3).

If a fuse has blown, there may be a good reason for this. Occasionally, however, they blow for other reasons, such as a momentary spike in the electric power lines or when turning on a heating element on a particularly cold day. So, generally, if there is no obvious sign of a problem with the device

Figure 10-3 Testing a fuse with a multimeter

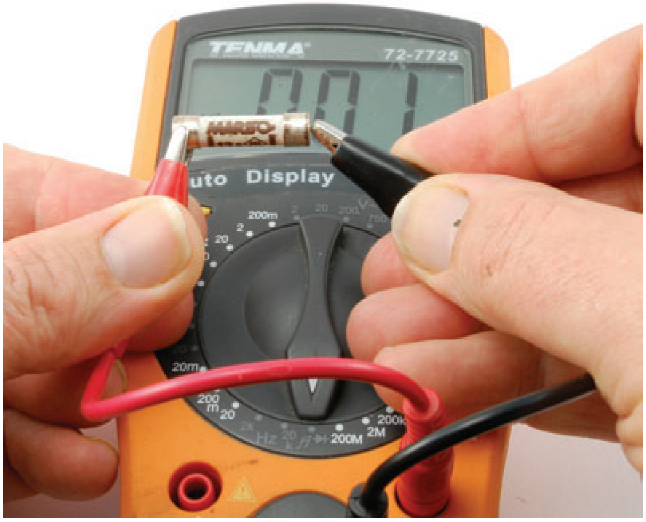

(look for wires that have come lose, or any sign of charring), then try replacing the fuse.

If the fuse immediately blows again, don't try another. You should instead find the source of the problem.

How to Test a Battery

Spent batteries are, of course, another common reason for something not working. Simply measuring the voltage will tell you very quickly if the battery is empty.

Figure 10-4 Testing a battery using a resistor and multimeter

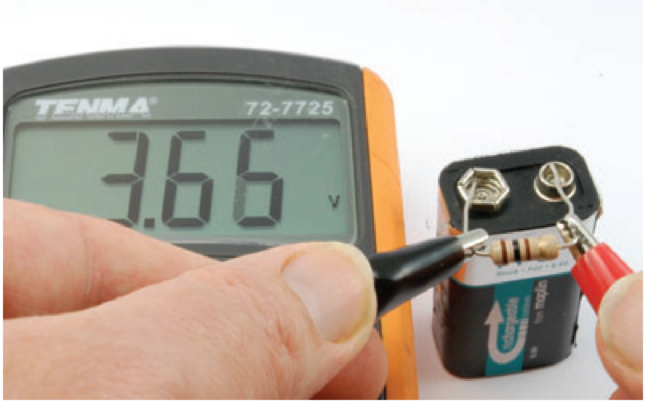

During testing, if a 1.5V battery like an AA or AAA is showing less than 1.2V, or a 9V battery is showing less than 8V, it is probably time to throw it away. However, the voltage of a battery shown when it is not powering anything can be a little misleading. For a more accurate picture, use a 100Ω resistor as a "dummy" load. Figure 10-4 shows a resistor and multimeter being used to assess the state of the battery.

How to Test a Heating Element

If you have a suspect heating element from an oven, hair dryer, or so on, you can check it by measuring its resistance. As with anything using household electricity, only do this when the appliance is completely disconnected.

It's a good idea to roughly work out what you think the resistance should be before you measure it. So, for example, if you have a 2-kW 220V heating element, then rearrange:

$P = V^2 / R$

to

$R = V^2 / P = 220 \times 220 / 2000 = 24\Omega$

Calculating what you expect before you measure it is always a good idea, because if you measure it first, it is all too easy to convince yourself that it was what you were expecting. For instance, one time your humble author convinced himself that a suspect element was fine because it was showing a resistance of a few hundred ohms. Eventually, it transpired that there was a light bulb in parallel with the heating element and that the element itself was instead broken.

Finding and Replacing Failed Components

When something stops working on a PCB, it is often the result of something burning out. This sometimes leads to charring around the component. Resistors and transistors are common culprits.

Testing Components

Resistors are easy to test with a multimeter set to its resistance range. Although the results can be misleading, you can test them without removing them. Most of the time, you are looking for an open circuit, very high resistance, or sometimes a short (0Ω).

If your multimeter has a capacitance range, these too can easily be tested.

Other components are less easily identified. It is usually possible to make out some kind of device name on the case.

A magnifying glass is sometimes useful, as is taking a digital photograph and then zooming in to a high magnification. Having found some kind of identifying mark, type it into your favorite search engine.

Bipolar transistors can also be tested (see the section “How to Use a Multimeter to Test a Transistor” in Chapter 11). However, if you have a spare, it is often easier just to replace it.

Desoldering

There is definitely a knack to desoldering. You often have to add more solder to get the solder to flow. I find it quite effective to draw the solder off onto the tip of the soldering iron, which I keep cleaning using the sponge.

Desoldering braid (Appendix – code T13) is also quite effective. Figure 10-5 shows the steps involved in using desoldering braid to remove the solder from around a component lead so it can be removed.

FIGURE 10-5 Using desoldering braid

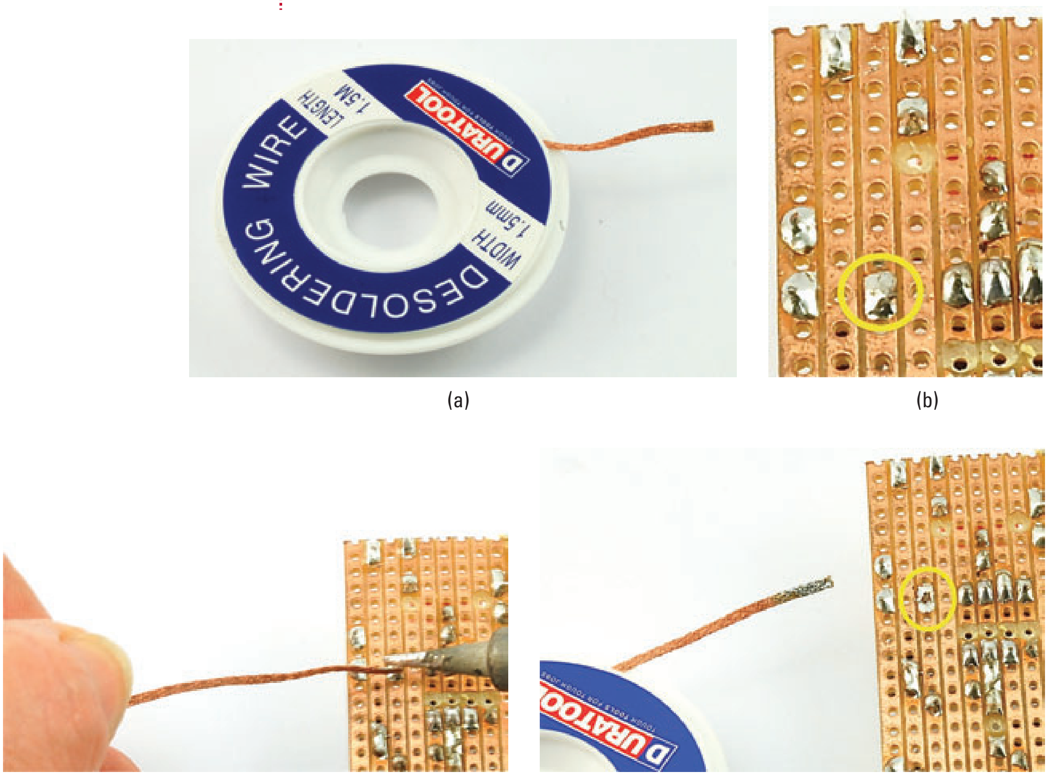

(a) (b)

(c) (d)

Desoldering braid (Figure 10-5a) is supplied in small lengths on a small reel. You do not need much. It is braided wire impregnated with flux that encourages the solder to flow into it and off the PCB or stripboard copper.

Figure 10-5b shows the joint (circled in yellow) that we are going to remove the solder from. Press the braid onto the joint with the soldering iron (Figure 10-5c) and you should feel the blob of solder on the joint start to melt into the braid. Remove the braid while everything is hot and you should see a nice clean joint with the solder transferred to the braid (Figure 10-5d).

Cut off the section of the braid with solder on it and throw it away.

You may have to do this a couple of times to remove enough solder to release the component.

Replacement

Soldering in the replacement component is straightforward, you just have to make sure you get it the right way around. This is where photographing the board before making the replacement can be a good idea.

How to Scavenge Useful Components

Dead consumer electronics are a good source of components. But be selective, because some components are really not worth saving. Resistors are so cheap that it is really not worth the effort of removing them.

Here is what I look for when scavenging:

- Any kind of motors
- Connectors
- Hookup wire
- Seven-segment LED displays
- Loudspeakers
- Switches
- Large transistors and diodes
- Large or unusual capacitors
- Screws nuts and bolts

FIGURE 10-6 Scavenging from a VCR

Figure 10-6 shows the insides of a dead video cassette recorder, with some of the more interesting parts for scavenging labeled.

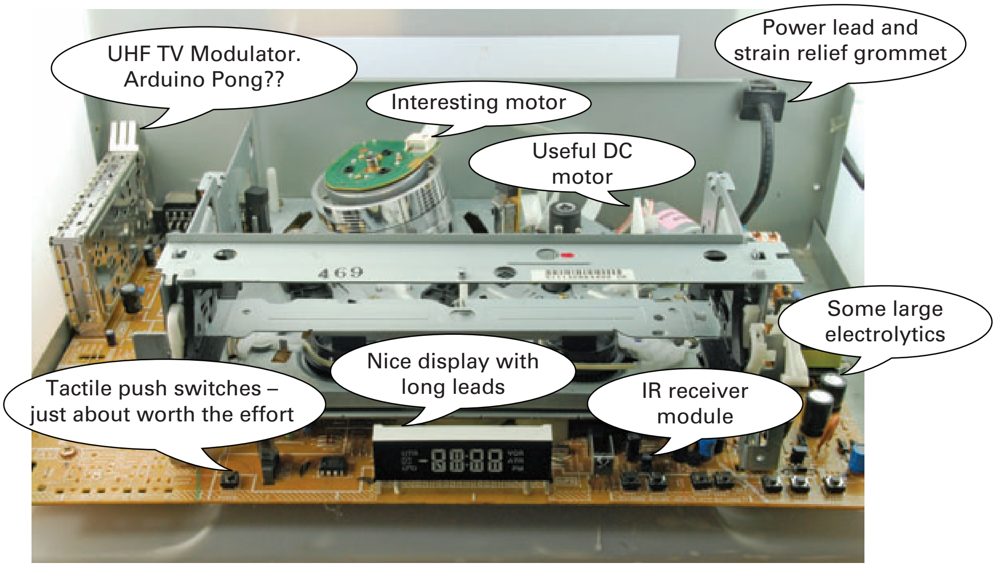
UHF TV Modulator.
Arduino Pong??
Power lead and strain relief grommet
Interesting motor
Useful DC motor
Some large electrolytics
Nice display with long leads
Tactile push switches – just about worth the effort
IR receiver module

The easiest way to remove a lot of components, and things like hookup wire, is simply to snip them with wire cutters. The same applies to large electrolytic capacitors and other items, as long as they still leave leads long enough to use. Alternatively, you can desolder the items.

How to Reuse a Cell Phone Power Adapter

Everything you make in electronics requires a power source of some sort. Sometimes this will be batteries, but often it is more convenient to power the device from your household electricity.

Given that most of us have drawers stuffed with obsolete mobile phones and their charges, it makes sense to be able to reuse an old mobile phone charger. If they are newish phones, they may well have some kind of standard connector on them like a mini-USB or micro-USB, but many older phone models had a proprietary plug, used only by that phone manufacturer.

There is nothing to stop us from taking such an adapter and putting a more standard plug on the end of it, or even connecting the bare wires to screw terminals.

Figure 10-7 shows the steps involved in putting a different type of connector, such as a 2.1mm barrel jack, on the end of an old cell phone charger.

The charger is of the "wall-wart" type that plugs directly into an electrical outlet. The connector is of a type long since discontinued (Figure 10-7a). The charger has a label saying that it can supply 5V at 700mA, so the first step is (making sure the charger is unplugged) to chop off the existing connector and strip the bare wires. There should be two wires, and if one is black and one is red, then the red one is usually positive and the black negative. In this case, the wires are red and yellow. Whatever colors the wires are, it is always a good idea to use a multimeter to check the polarity (Figure 10-7b).

Remember to put the lead through the plastic body of the barrel jack before you start soldering!

You can then solder on a barrel jack plug (Appendix–code H11). This is much the same procedure we used for an audio lead in the section "Hacking Audio Leads" in Chapter 9. Figure 10-7c shows the plug ready to solder, while Figure 10-7d displays the final lead ready to use.

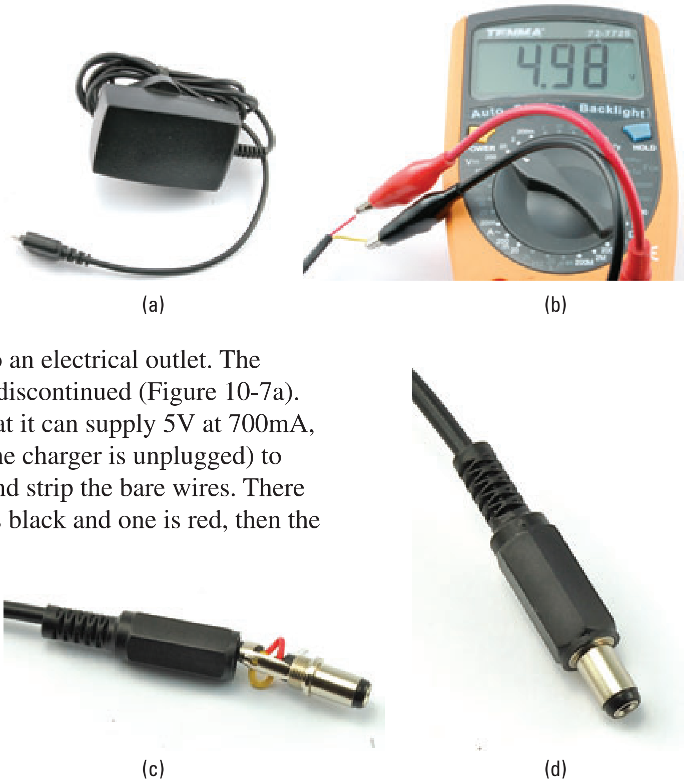

Figure 10-7 Attaching a barrel jack to a cell phone charger

Summary

In this chapter, we have discovered some of the treasures that can be rescued from dead electronic equipment and also briefly looked at testing and mending.

If you want to learn more about mending things, I recommend the book *How to Diagnose and Fix Everything Electronic* by Michael Geier (McGraw-Hill/TAB, 2011).

11

Tools

This chapter is mainly for reference. You have already met some of the techniques described here while working your way through the book.

How to Use a Multimeter (General)

Figure 11-1 shows a close-up of the range selector of my multimeter.

This is typical of a medium-range multimeter costing around USD 20. We have probably only used four or five of the settings during the course of this book, so it is worth pointing out some of the other features of a multimeter like this.

Continuity and Diode Test

Starting at 6 o'clock, we have the Continuity mode, represented by a little music symbol and also a diode symbol. We have used the Continuity mode many times. It just beeps when there is very low resistance between the leads.

The reason a diode symbol appears here is because this mode also doubles for testing diodes. With some multimeters, this feature will also work on LEDs, allowing you to measure the forward voltage.

Connect the anode of the diode (the end without a stripe in a normal diode, and with a longer lead on an LED) to the red test lead of the multimeter, and then the other end of the diode to the black lead. The meter will then tell you the forward voltage of the diode. So, expect to see about 0.5V for a normal diode and 1.7V to 2.5V for an LED. You will probably also find that the LED glows a little.

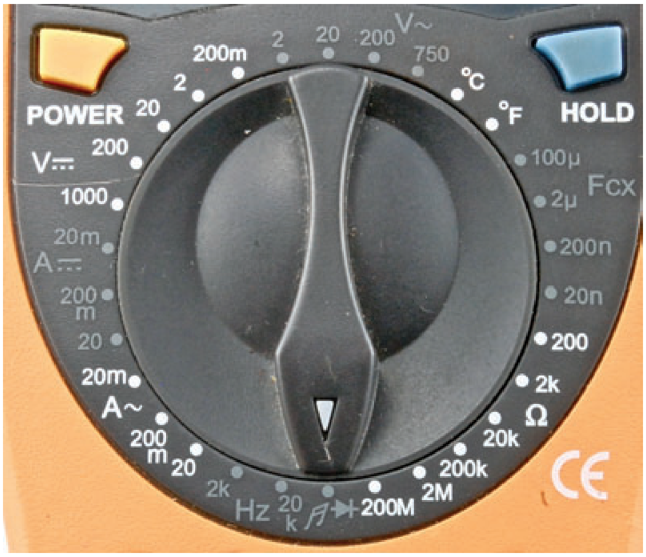

Figure 11-1 Multimeter range selection

Resistance

The multimeter in Figure 11-1 has six resistance ranges, from 200MΩ down to 200Ω. If you pick a range that has a maximum resistance lower than the resistor you are measuring, then the meter will indicate this. Mine does so by displaying a "1" on its own without any further digits. This tells me I need to switch to a higher resistance range. Even better, start at the maximum range and work your way down until you get a precise reading. For the most precise reading, you need the meter to be on the range above the one that tells you it's out of range.

When measuring high-value resistors of 100kΩ and up, remember that you yourself are also a big resistor, so if you hold the test lead to the resistor at both ends (see Figure 11-2), you are measuring both the resistor in question and your own resistance.

Use test leads with crocodile clips, or pin the resistor to your work surface with the flat of the test leads.

Figure 11-2 How not to measure high-value resistors

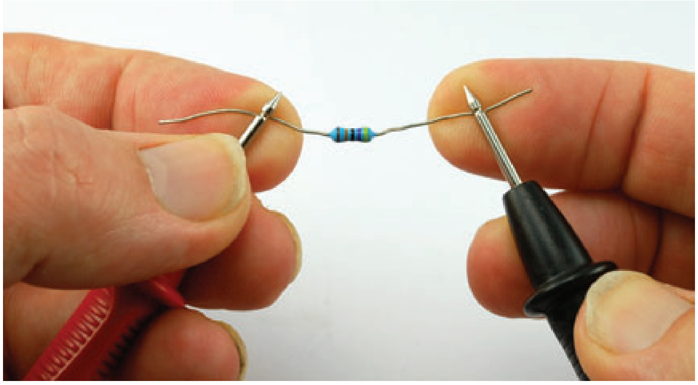

Capacitance

Some multimeters include a capacitance range. While not particularly useful for finding the value of unknown capacitors (capacitors have their value written on them), being able to test a capacitor and make sure it still has a capacitance something close to its stated value is useful.

The capacitance range on most meters is quite inaccurate, but then the values of actual capacitors—especially electrolytics—often have quite a wide tolerance.

In other words, if your meter tells you that your 100μF capacitor is actually 120μF, then that is to be expected.

Temperature

If your multimeter has a temperature range, it probably also comes with a special set of leads for measuring it, such as those shown in Figure 11-3.

The leads are actually a thermocouple that can measure the temperature of the tiny metal bead on the end of the leads. This thermometer is a lot more useful than your average digital thermometer. Check the manual for your meter, but the range of temperature is likely to be something like –40°C to 1000°C (–40°F to 1832°F).

So, you can use it to check how hot your soldering iron is getting, or if you have a component in a project that seems to be getting a bit toasty, you can use this to check just how hot it is getting.

FIGURE 11-3 Thermocouple leads for temperature measurement

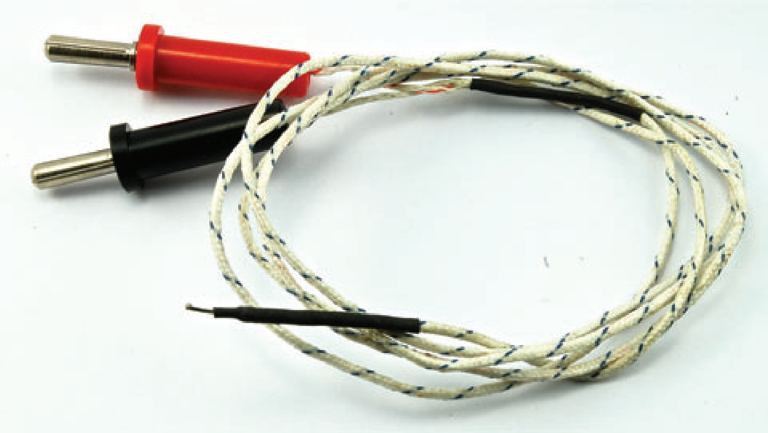

AC Voltage

We have not talked about AC very much in this book. AC stands for alternating current and refers to the type of electricity you get in a home wall socket's 110V or 220V supply. Figure 11-4 shows how 110V AC household electricity voltage varies over time.

From Figure 11-4, it is apparent that the voltage actually reaches a peak of 155V and swings all the way negative to –155V. So you might be wondering why it is referred to as 110V at all.

The answer is that since a lot of the time, the voltage is quite low, at those times, it delivers very little power. So the 110V is a kind of average. It's not the normal average voltage, because that would be (110 – 110) / 2 = 0V, and because half the time it is negative.

110V is the RMS voltage (root mean squared). This is the peak positive voltage divided by the square root of 2 (1.4). You can think of this as the DC equivalent voltage. So a light bulb running on 110V AC would appear to be the same brightness as if it were running on 110V DC.

You are unlikely to need to measure AC unless you are doing something exotic and dangerous, and you should not do that unless you are very sure about what you are doing and therefore probably already knew what I just told you.

Figure 11-4 Alternating current

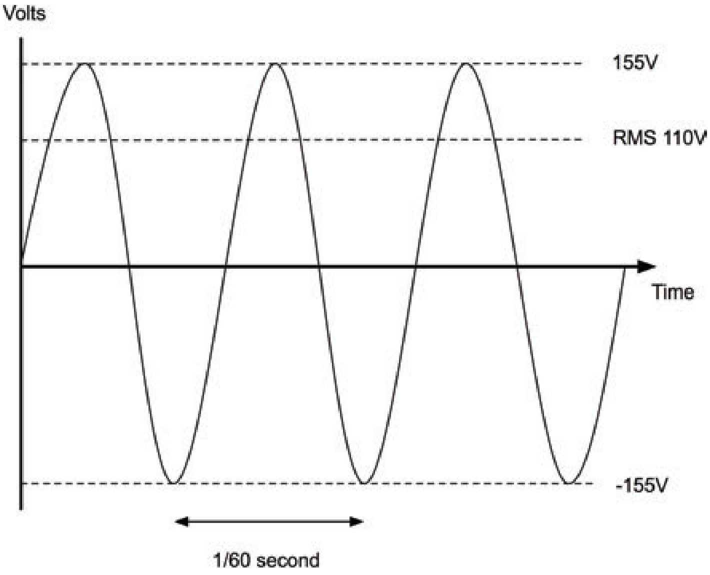

DC Voltage

We have already measured DC voltage quite a lot—mostly at the 0 to 20V range.

There is nothing much more to say about this, except to always start with the highest voltage range you believe you are about to measure and then work your way down.

DC Current

When measuring current, you will probably find that for all current ranges you will need to use different sockets on the multimeter for the positive probe lead. There is usually one connection for low currents and a separate one for the high-current ranges (20A on the author's multimeter; see Figure 11-5).

There are two important points to consider here. First, if you exceed the current range, your meter will not just give you a warning, it may well blow a fuse within the meter.

The second point is that when the probe leads are in the sockets for current measurement, there is a very low resistance between them. After all, they need to allow as much of the original current as possible to flow through them. So, if you forget that the leads are in these sockets and go to measure a voltage elsewhere in the circuit, you will effectively short-out your circuit and probably blow the fuse on your multimeter at the same time.

So, just to reiterate, if you have been using your multimeter to measure current, ALWAYS put the probe leads back to their

Figure 11-5 High-current measurement

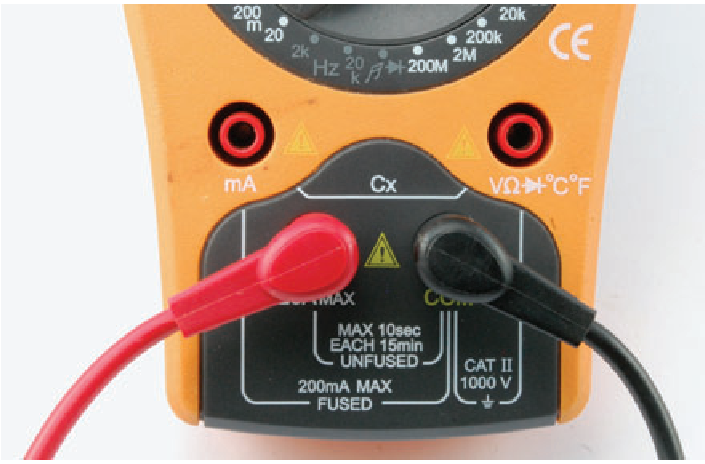

voltage sockets, as these are more likely to be used next. If you try and measure current with the leads in the voltage sockets, then all that will happen is that you get a reading of zero.

AC Current

The same argument that we gave for measuring AC voltage probably applies here. Exercise extreme caution.

Frequency

If your multimeter has a frequency setting, this can be useful. For example, in the section "How to Generate Tones with a 555 Timer" in Chapter 9 where we create an audio tone using a 555 timer, you could use this feature to measure the frequency of the tone being produced. This can be handy if you do not have access to an oscilloscope.

How to Use a Multimeter to Test a Transistor

Some multimeters actually have a transistor test socket where you can plug in a transistor. The multimeter will not only tell you if the transistor is alive or dead, but also what its gain (Hfe) is.

If your multimeter does not have such a feature, you can use the diode test feature to at least tell you if the transistor is undamaged.

Figure 11-6 shows the steps involved in testing an NPN bipolar transistor like the 2N3906.

Put the multimeter into diode test mode and attach the negative lead of the meter to the center base connection of the transistor, and the positive lead to one of the other leads of the transistor. It does not matter if it is the emitter or collector (check the transistor's pinout to find the base). You should get a reading, somewhere between 500 and 900. This is the forward voltage in mA between the base and whichever other connection you chose (Figure 11-6a). Then, move the positive lead to the other lead of the transistor (Figure 11-6b) and you should see a similar figure. If either reading is zero, either your transistor is dead, or it is a PNP type of transistor, in which case you need to carry out the same procedure but with the positive and negative leads to the multimeter reversed.

FIGURE 11-6 Testing a transistor

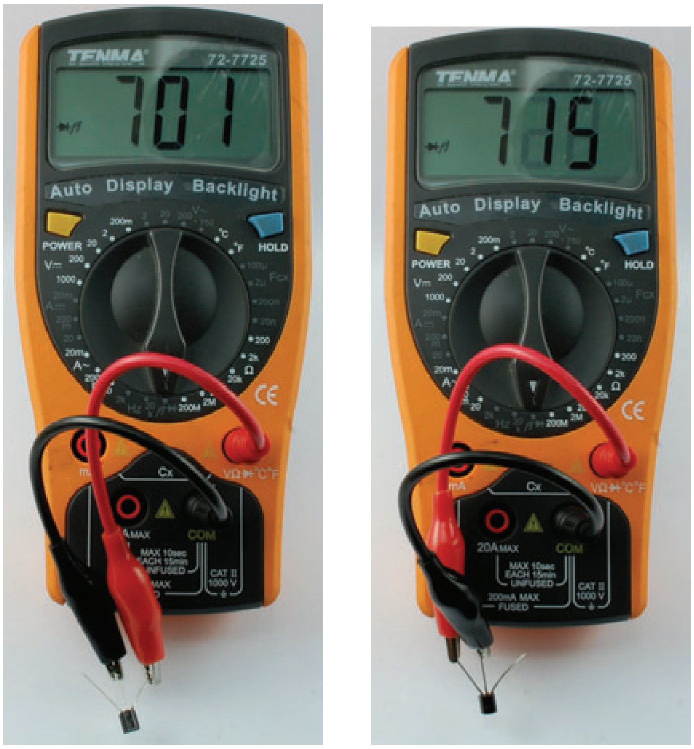

(a) (b)

How to Use a Lab Power Supply

We came across a lab power supply back in Chapter 5. If you have your soldering equipment and a multimeter, then a lab power supply (Figure 11-7) is probably the next item to invest in. It will get a lot of use.

The power supply shown in Figure 11-7 is a simple-to-use basic design. In the figure, it is being used to charge a lead–acid battery. You will find that you use it to power your projects while developing them. You should be able to get something similar for under USD 100.

It plugs into your home electrical socket and can deliver up to 20V at 4A, which is more than enough for most purposes. The screen displays the voltage at the top, and the current being consumed at the bottom.

The reasons why it is more convenient than using batteries or a fixed power supply are:

- It displays how much current is being consumed.
- You can limit the current consumption.
- You can use it in constant current mode when testing LEDs.
- You can adjust the voltage easily.

The control panel has an Output switch that turns the output voltage on and off, and two knobs that control the voltage and current.

If I am powering up some project for the first time, I will often follow this procedure:

1. Set the current to its minimum setting.
2. Set the desired voltage.
3. Turn on the output (the voltage will probably drop).
4. Increase the current and watch the voltage rise, making sure that the current isn't rising to an unexpected level.

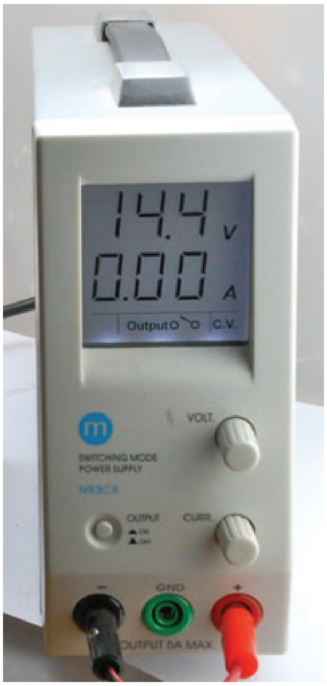
14.4 V
0.00 A
Output
C.V.
VOLT
SWITCHING MODE POWER SUPPLY
OUTPUT
CURR.
GND
OUTPUT 5A MAX.

FIGURE 11-7 A lab power supply

Introducing: The Oscilloscope

Oscilloscopes (Figure 11-8) are an indispensable tool for any kind of electronics design or test where you are looking at a signal that changes over time. They are a relatively expensive

FIGURE 11-8 A low-cost digital oscilloscope

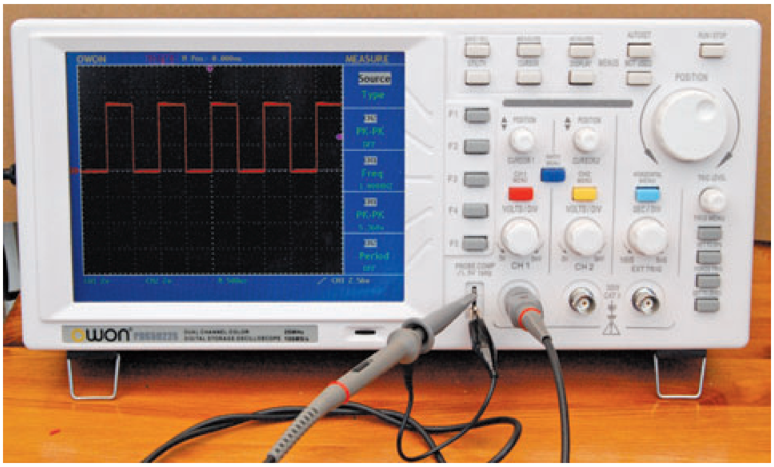
OWON
MEASURE
Source
POSITION
CH 1
CH 2

bit of equipment (from USD 200 on up) and there are various kinds. One of the most cost-effective types does not have any display at all, but connects to your computer over USB. If you don't want to risk blobs of solder on your laptop, or wait for it to boot up, then a dedicated oscilloscope is probably best.

Entire books have been written about using an oscilloscope effectively, and every oscilloscope is different, so we will just cover the basics here.

As you can see from Figure 11-8, the waveform is displayed over the top of a grid. The vertical grid is in units of some fraction of volts, which on this screen is 2V per division. So the voltage of the square wave in total is 2.5 × 2 or 5V.

The horizontal axis is the time axis, and this is calibrated in seconds. In this case, 500 microseconds (µS) per division. So the length of one complete cycle of the wave is 1000 µS—or 1 millisecond—indicating a frequency of 1 KHz.

The other advantage of an oscilloscope is that the test leads are very high impedance, which means that they have very little effect on the thing you are trying to measure.

Software Tools

As well as hardware tools for hacking electronics, there are lots of useful software tools that can help us out.

Simulation

If you like the idea of trying out electronic designs in a virtual world, you should try one of the online simulators like CircuitLab (www.circuitlab.com). This online tool (Figure 11-9) allows you to draw your circuits online and simulate how they will behave.

You will have to pick up a bit more theory than this book covers, but a tool like this can save you a lot of effort.

Fritzing

Fritzing (www.fritzing.org) is a really interesting open-source software project that lets you design projects. It is intended

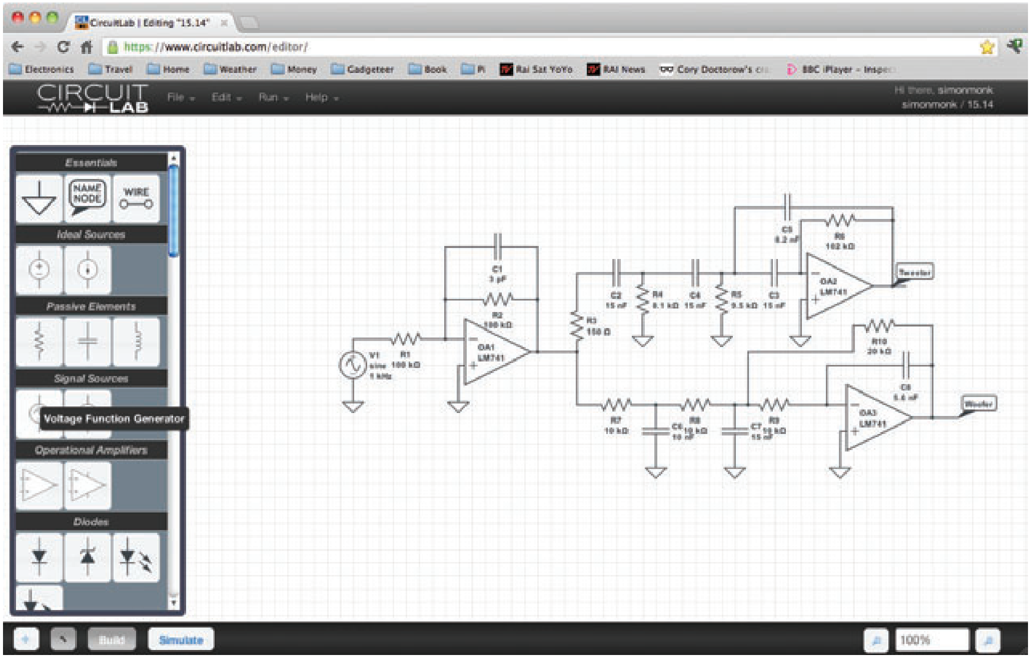
CircuitLab | Editing "15.14"
https://www.circuitlab.com/editor/
Electronics
Travel
Home
Weather
Money
Gadgeteer
Book
Pi
Rai Sat YoYo
RAI News
Cory Doctorow's
BBC iPlayer - Inspec
CIRCUIT LAB
File
Edit
Run
Help
Hi there, simonmonk
simonmonk / 15.14
Essentials
NAME NODE
WIRE
Ideal Sources
Passive Elements
Signal Sources
Voltage Function Generator
Operational Amplifiers
Diodes
V1 sine 1 kHz
R1 100 kΩ
C1 3 pF
R2 100 kΩ
OA1 LM741
R3 150 Ω
C2 15 nF
R4 8.1 kΩ
C4 15 nF
R5 9.5 kΩ
C3 15 nF
C5 8.2 nF
R6 102 kΩ
OA2 LM741
Tweeter
R10 20 kΩ
C8 5.6 nF
OA3 LM741
Woofer
R7 10 kΩ
C6 10 nF
R8 10 kΩ
C7 15 nF
R9 10 kΩ
Build
Simulate
100%

Figure 11-9 The CircuitLab simulator

primarily for breadboard design and includes libraries of components and modules, such as an Arduino, that can all be wired up (Figure 11-10).

EAGLE PCB

If you want to start creating your own PCBs for your electronics designs, then look for the most popular tool for this, which is called EAGLE PCB (Figure 11-11). It allows you to draw a schematic diagram and then switch to a PCB view where you can route the connections between components before creating the CAM (computer-aided manufacturing) files, which you can then send off to a PCB fabrication shop.

Creating PCBs is a subject in its own right. For more information on this, take a look at the book *Make Your Own PCBs with EAGLE: From Schematic Designs to Finished Boards* by Simon Monk (TAB, 2013).

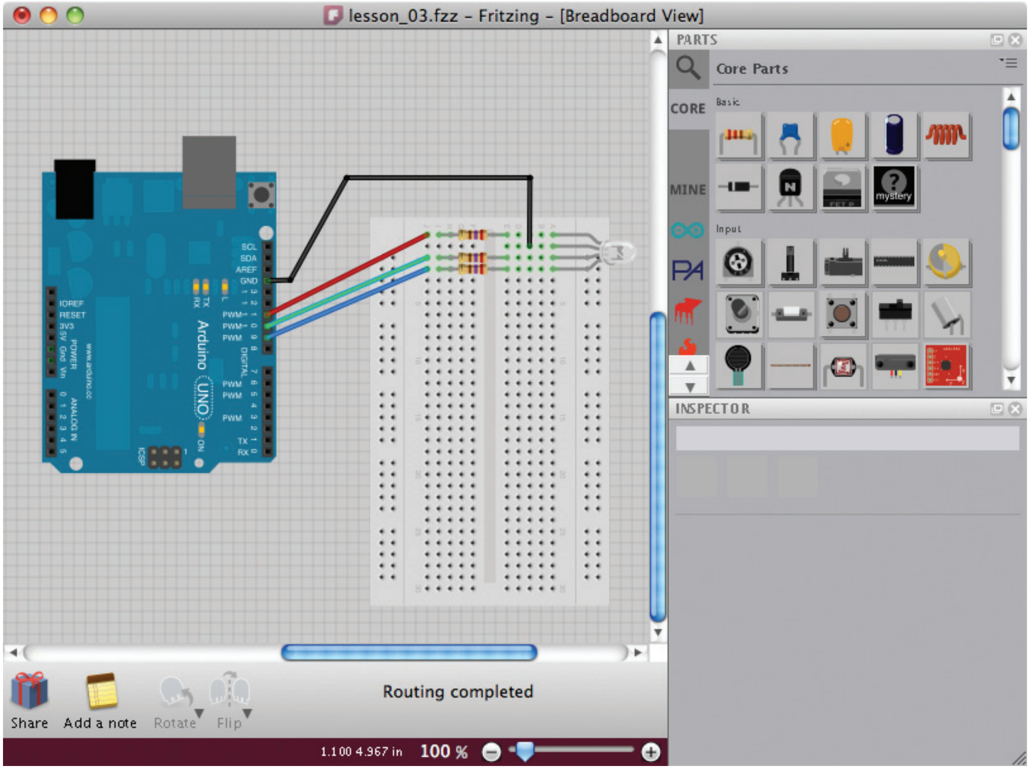
lesson_03.fzz – Fritzing – [Breadboard View]
PARTS
Core Parts
CORE
MINE
Basic
Input
INSPECTOR
Routing completed
Share
Add a note
Rotate
Flip
1.100 4.967 in
100 %

FIGURE 11-10 Fritzing

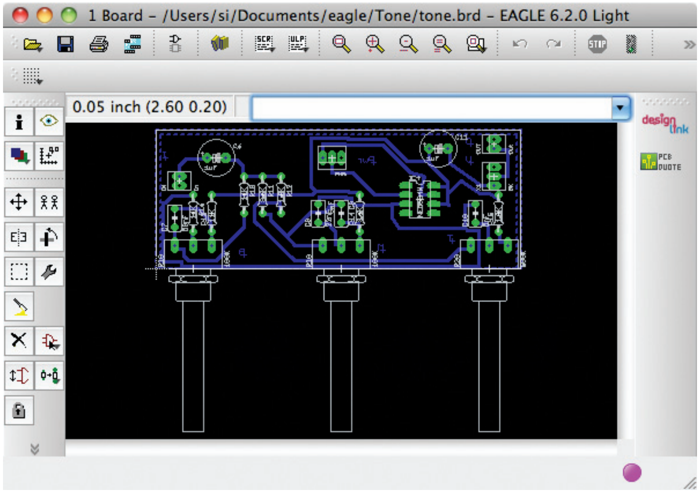
1 Board – /Users/si/Documents/eagle/Tone/tone.brd – EAGLE 6.2.0 Light
0.05 inch (2.60 0.20)

FIGURE 11-11 EAGLE PCB

Online Calculators

Online calculators can make your electronics math a whole lot easier. Some of the more useful ones are:

- **http://led.linear1.org/1led.wiz** A series resistor calculator for LEDs
- **http://led.linear1.org/led.wiz** Designed for driving large numbers of LEDs
- **www.bowdenshobbycircuits.info/555.htm** A 555 timer IC component calculator

Summary

This is the last chapter in this book and I hope it will help you get started "hacking electronics." There is much satisfaction in making something physical, or modifying a device so it does just what you want.

The line between producer and consumer is blurring more and more today as people start designing and building their own electronic devices.

The Internet offers many useful resources. The following web sites are worth a special mention:

- www.hacknmod.com
- www.instructables.com
- www.arduino.cc (for Arduino)
- www.sparkfun.com (modules and interesting components)
- www.adafruit.com (more cool stuff)
- www.dealextreme.com (bargains; search for LEDs, etc.)
- www.ebay.com (search for the same items as that in the other URLs in this list)

See also the components suppliers mentioned in the Appendix.

Appendix

Parts

Prices of components vary enormously, so please treat the following lists as a guide and shop around.

I know some people who buy almost everything on eBay. But beware. Though things are often very cheap there, occasionally they are much more expensive than at other suppliers.

I have listed part codes for the tools, modules, and so on from SparkFun and Adafruit, as these suppliers are very accessible to hobbyists and also provide good accompanying documentation. They also have distributors throughout the world, so you do not have to buy direct from either company if you live outside the U.S.

For other components, I have tried to list product codes for Mouser and DigiKey since these predominate as suppliers to hobbyists in the U.S., and also Farnell, who are UK-based but will ship to anywhere.

Please also see the book's web site (www.hackingelectronics.com) as updates for component availability will appear here.

Tools

Book Code	Description	SparkFun	Adafruit
T1	Beginner toolkit (soldering kit, pliers, snips)	TOL-09465	
T2	Multimeter	TOL-09141	
T3	PVC insulating tape	PRT-10688	
T4	Helping hands	TOL-09317	ID: 291
T5	Solderless breadboard	PRT-00112	ID: 239
T6	Solid-core jumper wire set	PRT-00124	ID: 758
T7	Red hookup wire (22 AWG)	PRT-08023	ID: 288
T8	Black hookup wire (22 AWG)	PRT-08022	ID: 290

Book Code	Description	SparkFun	Adafruit
T9	Yellow hookup wire (22 AWG)	PRT-08024	ID: 289
T10	Red multi-core wire (22 AWG)	PRT-08865	
T11	Black multi-core wire (22 AWG)	PRT-08867	
T12	Male-to-female jumper set	PRT-09385	ID: 825
T13	Desoldering braid / wick	TOL-09327	ID: 149

Components

To get yourselves a basic stock of components, you are strongly recommended to buy a starter kit of components. SparkFun sells such a set, but it doesn't include resistors, so you will need to buy a resistor set, too. Once you have these, you will own a useful collection of components that should cover the majority of what you need.

Component Starter Kits

The SparkFun Beginner Parts Kit and Resistor Kit will give you a good initial stock of parts.

Book Code	Description	SparkFun
K1	SparkFun Beginner Parts Kit (KIT-10003)	KIT-10003
K2	SparkFun Resistor Kit	COM-10969

Resistors

Book Code	Description	SparkFun	Adafruit	Other
R1	10kΩ trimpot, 0.1-inch pitch (also in K1)	COM-09806	ID: 356	DigiKey: 3362P-103LF-ND Mouser: 652-3362P-1-103LF Farnell: 9354301
R2 also in K1	LDR (also in K1)	SEN-09088	ID: 161	DigiKey: PDV-P8001-ND Farnell: 1652637
R3	500Ω trimpot			DigiKey: CT6EP501-ND Mouser: 652-3386P-1-501LF Farnell: 9355103

Capacitors

Book Code	Description	SparkFun	Other
C1	1000μF 16V electrolytic		DigiKey: P10373TB-ND Mouser: 667-ECA-1CM102 Farnell: 2113031
C2	100μF 16V electrolytic (also in K1)	COM-00096	DigiKey: P5529-ND Mouser: 647-UST1C101MDD Farnell: 8126240
C3	470nF capacitor		DigiKey: 445-8413-ND Mouser: 810-FK28X5R1E474K Farnell: 1179637
C4	100nF capacitor (also in K1)	COM-08375	DigiKey: 445-5258-ND Mouser: 810-FK18X7R1E104K Farnell: 1216438 Adafruit: 753
C5	10μF capacitor (also in K1)	COM-00523	DigiKey: P14482-ND Mouser: 667-EEA-GA1C100 Farnell: 8766894

Semiconductors

Book Code	Description	SparkFun	Adafruit	Other
S1	2N3904 (also in K1)	COM-00521	756	DigiKey: 2N3904-APTB-ND Mouser: 610-2N3904 Farnell: 9846743
S2	High-brightness white LED (5mm)	COM-00531	754	DigiKey: C513A-WSN-CV0Y0151-ND Mouser: 941-C503CWASCBADB152 Farnell: 1716696
S3	1-W Lumiled LED on heatsink	BOB-09656	518	DigiKey: 160-1751-ND Mouser: 859-LOPL-E011WA Farnell: 1106587
S4	7805 voltage regulator (also in K1)	COM-00107		DigiKey: 296-13996-5-ND Mouser: 512-KA7805ETU Farnell: 2142988
S5	1N4001 diode (also in K1)	COM-08589	755	DigiKey: 1N4001-E3/54GITR-ND Mouser: 512-1N4001 Farnell: 1651089
S6	FQP30N06	COM-10213	355	DigiKey: FQP30N06L-ND Mouser: 512-FQP30N06 Farnell: 1695498

Book Code	Description	SparkFun	Adafruit	Other
S7	LM311 comparator			DigiKey: 497-1570-5-ND Mouser: 511-LM311N Farnell: 9755942
S8	TMP36 Temperature IC	SEN-10988	165	DigiKey: TMP36GT9Z-ND Farnell: 1438760
S9	TDA7052			DigiKey: 568-1138-5-ND Mouser: 771-TDA7052AN Farnell: 526198
S10	NE555 timer IC (also in K1)	COM-09273		DigiKey: 497-1963-5-ND Mouser: 595-NE555P Farnell: 1467742
S11	Red LED 5mm	COM-09590	297	DigiKey: 751-1118-ND Mouser: 941-C503BRANCY0B0AA1 Farnell: 1249928
S12	Linear hall effect sensor			DigiKey: 620-1022-ND Mouser: 785-SS496B Farnell: 1791388

Hardware and Miscellaneous

Book Code	Description	SparkFun	Adafruit	Other
H1	4 × AA battery holder	PRT-00550	830	DigiKey: 2476K-ND Mouser: 534-2476 Farnell: 4529923
H2	Battery clip			DigiKey: BS61KIT-ND Mouser: 563-HH-3449 Farnell: 1183124
H3	Stripboard			eBay—search for "stripboard" Farnell: 1201473
H4	Pin header strip	PRT-00116	392	
H5	2A two-way screw terminal			eBay—search for "terminal block" Mouser: 538-39100-1002
H6	6V gear motor			Part of H7 eBay—search for "gear motor" or "gearmotor"
H7	Magician chassis	ROB-10825		
H8	6 × AA battery holder		248	DigiKey: BH26AASF-ND Farnell: 3829571

Book Code	Description	SparkFun	Adafruit	Other
H9	Battery clip to 2.1mm jack adapter		80	
H10	9g servo motor	ROB-09065	169	
H11	2.1mm barrel jack plug			DigiKey: CP3-1000-ND Farnell: 1737256
H12	Small solderless breadboard	PRT-09567	64	
H13	12V bipolar stepper motor	ROB-09238	324	
H14	8Ω speaker	COM-09151		
H15	Large pushbutton switch	COM-09336	559	
H16	5V relay	COM-00100		Digikey: T7CV1D-05-ND

Modules

Book Code	Description	SparkFun	Adafruit	Other
M1	12V 500mA power supply	TOL-09442	798	Note: U.S. model listed here.
M2	Arduino Uno R3	DEV-11021	50	
M3	Piezo sounder	COM-07950	160	
M4	Arduino Ethernet Shield	DEV-09026	201	
M5	PIR module	SEN-08630	189	
M6	MaxBotix LV-EZ1 rangefinder	SEN-00639	172	
M7	HC-SR04 rangefinder			eBay—search for “HC-SR04”
M8	AK-R06A RF Kit			eBay—search for “433MHZ 4 Channel RF Radio”
M9	SparkFun TB6612FNG breakout board	ROB-09457		
M10	Piezo sounder (built-in oscillator)			eBay—search for “Active Buzzer 5V”
M11	Methane sensor MQ-4	SEN-09404		
M12	Color sensing module			eBay—search for “TCS3200D Arduino”
M13	Piezo vibration sensor	SEN-09199		

Book Code	Description	SparkFun	Adafruit	Other
M14	SparkFun mic module	BOB-09964		
M15	Accelerometer module		163	Freetronics: AM3X
M16	USB LiPo charger	PRT-10161	259	
M17	Combined LiPo charger, Buck-booster	PRT-11231		
M18	Arduino LCD shield			Freetronics: LCD Keypad Shield
M19	4-digit, 7-segment display w/I2C backpack		880	
M20	RTC module		264	
M21	Arduino Leonardo	DEV-11286	849	

Index

1-watt audio amplifier, 224–227
30 Arduino Projects for the Evil Genius (Monk), 148
555 timer IC, 69, 70, 227–229
7805 voltage regulator, 95, 96

A

A1302 magnetic sensor, 210, 211
Ableton Live software, 229
AC current, 250
AC voltage, 248
accelerometer, 206–210
  construction of, 207–208
  items needed for, 207
  music controller with, 230–232
  schematic diagram, 208
  software sketch, 208–210
Adafruit
  accelerometer module, 206
  amplifier module, 225
  module library, 187
  part codes, 257–262
  RTC module, 189
  web site, 187, 256
alkaline batteries, 85
Allen, Charlie, 142
amperes, 27
amplifier module, 224–227
  construction of, 226–227
  items needed for, 225
  schematic diagram, 225
  testing process, 227
amplifiers
  building 1-watt, 224–227
  mic module, 218–220
analogWrite function, 124–125
angle/brightness of LEDs, 59
anodes and cathodes, 56
Arduino, 105–148
  accelerometer module and, 207–208
  Blink sketch changes and, 109–111
  Charlieplexing LEDs with, 142–145
  color-sensing module and, 199
  electric toy project, 116–119
  Ethernet shield for, 128–136
  explanation and overview, 105
  gas sensor used with, 196–197
  input/output pins, 112
  keyboard emulation, 145–147, 229–232
  LCD shield for, 128, 136–139
  LED control project, 122–125
  magnetic field sensor with, 211
  mic module attached to, 232
  MOSFET motor control project, 163–166
  passwords typed using, 145–147
  PIR module used with, 151
  playing sounds with, 125–127
  relay control projects, 112–116, 128–136
  schematic diagrams, 113

Arduino (*Cont.*)
  Serial Monitor, 118, 232
  servo motor project, 139–141
  setting up, 106–109
  shields used by, 127–128
  temperature sensor with, 205
  tone generator project, 125–127
  ultrasonic rangefinders and, 155–158
  vibration sensor with, 202–204
  voltage measurement, 119–121
  web-controlled relays, 128–136
  web resources for, 106, 128, 148
  wireless remote modules and, 161–162
Arduino Leonardo, 105, 128, 129, 145–147, 230
Arduino Uno, 107, 128, 129
auction sites, 2
audio frequencies, 224
audio hacks, 213–234
  amplifier module, 224–227
  audio leads, 213–218
  microphone module, 218–220
  software VU meter, 232–234
  tone generator, 227–229
  USB music controller, 229–232
audio leads, 213–218
  general principles of, 213–215
  soldering audio connectors, 215, 216
  stereo-to-mono conversion, 217–218
automatic battery backup, 99–101

B

barrel jack, 244
batteries, 83–101
  automatic backup with, 99–101
  boosting voltage from, 97–98
  calculating requirements for, 98–99
  capacitors compared to, 80
  capacity of, 83–84, 88, 98
  charging guidelines for, 88–93
  controlling voltage from, 95–97
  hacking cell phone, 93–94
  holders for, 85–86
  life of, 89
  maximum discharge rate of, 84
  rechargeable, 86–93
  single-use, 84–86
  testing, 239
  trickle charging, 89, 90, 100–101
  types/characteristics of, 85, 86, 87, 88
  See also power; solar panels
battery backup, 99–101
  diodes for, 99–100
  schematic diagram, 99
  trickle charging, 100–101
bipolar transistors, 39–40
  commonly used, 47
  datasheets, 45–46
  MOSFET, 46
  N-channel, 47
  NPN, 39–40, 47
  operational diagram, 40
  PNP, 46–47
Blink sketch, 107, 109–111
boosting voltage, 97–98
braid, desoldering, 241–242
breadboard, solderless, 3–5
breadboard layouts
  Charlieplexed LEDs, 143
  constant current driver, 64–65
  flashing LEDs, 69–71
  gas detector, 195–196
  H-Bridge experiment, 171
  LED lighting, 58, 61
  light switch, 41–42
  MaxBotix rangefinder and Arduino, 158
  MOSFET motor control, 49, 166
  PIR motion sensor module, 150–151
  real-time clock, 190
  robot rover, 181
  seven-segment LED display, 186

- stepper motor control, 175
- tone generator, 126, 229
- voltage divider, 36
- voltage regulator, 96

brightness of LEDs, 59, 124–125
buck-boost converters, 97–98
buying
- components, 1–2
- electronics to hack, 2–3
- solderless breadboard, 3–5
- starter toolkit, 3

C

calculators
- frequency, 229
- LED wizard, 68
- online, 256
- resistor, 57
- voltage divider, 36

calibrate function, 232
capacitance range, 247
capacitors, 22–23
- batteries vs., 80
- part codes for, 259
- safely discharging, 235–236
- storing charge in, 79
- testing with multimeters, 247

capacity of batteries, 83–84, 88
cathodes and anodes, 56
cell phones
- hacking batteries of, 93–94
- reusing power adapters from, 243–244

charging batteries, 88–89
- lead-acid batteries, 91–92
- LiPo batteries, 92–93
- NiMh batteries, 89–91

Charlieplexed LEDs project, 142–145
- construction of, 143
- explanation of, 142–143
- items needed for, 143
- software sketch, 143–145

chips. *See* integrated circuits
circuit boards
- stripboards and, 71, 75
- testing connections on, 13

CircuitLab simulator, 253, 254
color codes for resistors, 21
colors of LEDs, 60
color-sensing module, 197–202
- attaching to an Arduino, 199
- pin purposes table, 198
- software sketch, 199–202
- TCS3200 IC and, 197–198

common anode LED display, 184
common cathode LED display, 184
comparators, 194–195, 219
components, 258–262
- buying, 1–2
- capacitors, 22–23, 259
- desoldering, 241–242
- diodes, 23
- hardware, 260–261
- identifying, 20–25
- integrated circuits, 24–25
- LEDs, 23–24
- modules, 261–262
- part codes for, 258–262
- replacing, 242
- resistors, 20–23, 258
- scavenging, 242–243
- semiconductors, 259–260
- starter kits, 19–20, 258
- surface mount, 25
- symbols used for, 30–31
- testing, 240–241
- transistors, 24

computer fan fume extractor, 14–17
connections, testing, 12–13
constant current driver, 62–66
- breadboard layout, 64–65
- construction process, 65–66
- design overview, 63–64
- LM317 IC used for, 62, 63
- schematic diagram, 63
- *See also* voltage regulators

Continuity mode, 12–13, 245
crossover network, 224
current, 26, 249–250

D

datasheets
    LED, 57
    transistor, 45–46
DC current, 249–250
DC motor control, 166–172
    breadboard layout, 171
    H-Bridge module, 167–169
    items needed for, 170
    schematic diagram, 170
DC voltage, 249
desoldering process, 241–242
digitalWrite function, 134
diodes, 23
    battery backup, 99–100
    forward- vs. reverse biased, 56
    laser diode modules, 78–81
    testing with multimeters, 245
discharge rate, 84
discharging capacitors, 235–236
double pole switches, 52
double throw switches, 52
DPDT switches, 52, 53
drawColon function, 191–192
dry joints, 13

E

EAGLE PCB, 254, 255
eBay, 2, 221, 225, 256, 257
electric toy project, 116–119
    construction of, 116–117
    items needed for, 116
    software sketch, 118–119
    See also web-controlled toy project
electrocution avoidance, 235–236
electrolytic capacitors, 23
electronics
    buying to hack, 2–3
    rules for taking apart, 237
    toolkit for hacking, 3–5
emergency lantern project, 65–66
Ethernet library, 133
Ethernet shield, 128–136

F

fast charging, 91
fixed resistors, 22
flashing LEDs, 69–77
    breadboard project, 69–71
    stripboard project, 71–77
FM radio bug, 220–223
    construction of, 221–223
    items needed for, 221
    schematic diagram, 221
    testing process, 223
FM transmitters, 220–223
forward current, 56
forward voltage, 56, 66–67
forward-biased diode, 56
Freetronics
    accelerometer module, 206
    LCD and Keypad Shield, 137
frequency
    calculating, 229
    measuring, 250
fritzing, 253–254, 255
fume extractor project, 14–17
    hacking process, 15–17
    items needed for, 14
    schematic diagram, 14
    wiring diagram, 15
fuses, checking, 237–239

G

gas detector, 193–197
    Arduino used for, 196–197
    breadboard layout, 195–196

- items needed for, 193–194
- LM311 comparator for, 194–195
- schematic diagram, 194

Geier, Michael, 244
getDecimalTime function, 192

H

hacking
- basic toolkit for, 3–5
- buying electronics for, 2–3

hackingelectronics.com web site, 71, 105, 257
hardware part codes, 260–261
H-Bridge module, 166–177
- control pin usage, 171–172
- DC motor control, 166–172
- pin purposes table, 169
- schematic diagrams, 168, 170
- stepper motor control, 172–177

HC-SR04 rangefinder, 155–157
heating element test, 240
hertz (Hz), 70
high-value resistors, 246
How to Diagnose and Fix Everything Electronic (Geier), 244

I

I2C serial interface, 186
infrared LEDs, 60–61
insulated wire, 6
integrated circuits (ICs), 24–25
- 555 timer, 69, 70, 227–229
- A1302 magnetic sensor, 210, 211
- LM311 comparator, 194–195
- LM317 voltage regulator, 62–66, 97
- MCP73831 chip, 93
- TCS3200 chip, 197–198
- TDA7052 chip, 224, 225

integrated development environment (IDE), 105, 106
Internet resources. *See* web resources
IP addresses, 131, 132

J

joining wires, 7–12
- by soldering, 11–12
- by twisting, 7–8

jumper wires, 4

K

keyboard emulation
- automatic password entry, 145–147
- USB music controller, 229–232

L

lab power supply, 251–252
laser diode modules, 78–81
LCD display project, 136–139
- construction of, 137
- items needed for, 137
- software sketch, 137–139

LCD shield, 128, 136–139
LDO voltage regulators, 97
LDRs (light dependent resistors), 37
- LEDs controlled using, 39–44
- multimeter measurement of, 37
- tones controlled using, 227–229

lead-acid batteries
- characteristics of, 87, 88
- how to charge, 91–92

LEDs (light-emitting diodes), 23–24, 55–82
- Arduino for controlling, 122–125
- brightness and angle of, 59
- burnout prevention for, 55–58
- calculator for using, 68
- Charlieplexing with Arduino, 142–145
- constant current driver for, 62–66
- datasheet example, 57
- flashing projects, 69–77, 123–124
- forward voltage measurement, 66–67
- high-brightness, 61–62
- infrared and ultraviolet, 60–61
- laser diode modules and, 78–81
- LDRs for controlling, 39–44

LEDs (*Cont.*)
    mixing colored lights, 60
    PIR module used with, 150–151
    powering large numbers of, 68–69
    resistors used with, 56–57, 122–123
    schematic diagrams, 56, 58
    selecting for jobs, 59–62
    seven-segment display, 183–188
    slot car racer project, 78–81
libraries, 133, 187
light dependent resistors. *See* LDRs
light meter project example, 37–38
light sensing push light, 39–44
    construction process, 41–44
    items needed for, 39
    schematic diagrams, 40, 43
    situational considerations, 40–41
    wiring diagram, 44
light-emitting diodes. *See* LEDs
light-sensing module, 198
LiPo batteries
    characteristics of, 87, 88
    how to charge, 92–93
    over-charging, 89
    over-discharging, 94
lithium batteries, 85
LM311 comparator, 194–195
LM317 voltage regulator, 62–66, 97
loop function, 111, 124, 134, 141, 144
loudspeakers, 223–224
luminous intensity, 59

M

magic hands, 10
Magician Chassis, 180
magnetic field sensor, 210–212
    construction of, 211
    software sketch, 212
Make Your Own PCBs with EAGLE (Monk), 254
map function, 124
MaxBotix LV-EZ1 rangefinder, 157–158
maximum discharge rate, 84
MCP73831 IC, 93
measurements
    AC current, 250
    AC voltage, 248
    Arduino voltage, 119–121
    capacitance, 247
    DC current, 249–250
    DC voltage, 249
    frequency, 250
    LED forward voltage, 67
    resistance, 37, 246
    temperature, 247
methane detector. *See* gas detector
microcontrollers, 105
microphone module, 218–220
    attached to Arduino, 232
    FM radio bug hack, 220–223
    oscilloscope display, 218–219
    schematic diagram, 219
    VU meter creation, 232–234
microswitches, 51
modules, 149–192, 261–262
    accelerometer, 206–210
    amplifier, 224–227
    H-Bridge, 166–177
    light-sensing, 198
    microphone, 218–220
    part codes for, 261–262
    PIR motion sensor, 149–153
    real-time clock, 188–192
    seven-segment LED display, 183–188
    ultrasonic rangefinder, 153–158
    wireless remote, 159–162
    See also sensors
mono-from-stereo conversion, 217–218
mosfet_motor_speed sketch, 165–166
MOSFET transistors, 46, 163–166
    Arduino software sketch, 165–166
    breadboard layouts using, 49, 166

- motor control experiments, 48–50, 163–166
- schematic diagrams using, 48, 164

motion sensor. *See* PIR motion sensor module
motor controllers
- Arduino shield, 128
- H-Bridge module, 166–177
- MOSFET transistor, 48–50, 163–166

multi-core wire, 6
multimeters, 245–251
- battery test using, 239
- capacitance range on, 247
- Continuity mode, 12–13, 245
- current measurements, 249–250
- diode test using, 245
- frequency setting on, 250
- fuse test using, 238, 239
- LDR resistance measurement, 37
- LED forward voltage measurement, 67
- polarity test using, 15–16
- resistance ranges on, 246
- suggestions for buying, 3
- temperature range on, 247
- testing connections with, 12–13
- transistor test using, 250–251
- voltage measurements, 248–249

N

N-channel transistors, 47
network configuration, 131
NiMh batteries
- characteristics of, 87, 88
- how to charge, 89–91

NPN transistors, 39–40, 47

O

Ohm's law, 27–28
OmniGraffle software, 71
online tools
- calculators, 256
- fritzing, 253–254, 255
- simulators, 253, 254
- *See also* software tools; web resources

operational amplifier (op amp), 219–220
oscilloscope, 218–219, 252–253
over-charging batteries, 89
over-discharging batteries, 89, 94

P

pageNameIs function, 134, 135
part codes
- for capacitors, 259
- for components, 258–262
- for hardware/miscellaneous, 260–261
- for modules, 261–262
- for resistors, 258
- for semiconductors, 259–260
- for starter kits, 258
- for tools, 257–258
- *See also* components

password typing project, 145–147
- construction of, 146
- items needed for, 146
- software sketch, 146–147

PCBs, creating, 254
photoresistor, 37
piezo buzzer/sounder, 194
piezo vibration sensor, 202
pins, Arduino, 112
PIR motion sensor module, 149–153
- breadboard layout, 150–151
- construction process, 151–152
- interfacing with an Arduino, 151
- LED experiment, 150–151
- schematic diagrams, 150, 151
- software sketch, 152–153

pliers, 6–7
PNP transistors, 46–47
potentiometers (pots), 22
power, 28–29
- batteries used for, 83–101
- formula for calculating, 28

power (*Cont.*)
  household device usage of, 29
  solar panels used for, 101–104
power adapter hack, 243–244
power rating, 28
power supplies
  battery backup and, 100
  lab power supply, 251–252
  polarity test for, 15–16
  variable power supply, 91–92
Practical Electronics for Inventors (Sherz and Monk), 220
Programming Arduino: Getting Started with Sketches (Monk), 119, 148
pulse-width modulation (PWM), 124
push light hacking project. *See* light sensing push light
push-button switches, 50–51

R

radio bug. *See* FM radio bug
range_finder_budget sketch, 155
readColor function, 201
readHeader function, 134, 135
real-time clock (RTC) module, 188–192
  construction of, 190
  items needed for, 189
  software sketch, 191–192
rechargeable batteries, 86–88
  battery life, 89
  characteristics of, 88
  commonly used types of, 87
  guidelines for charging, 88–93
  lead-acid, 87, 88, 91–92
  LiPo, 87, 88, 89, 92–93
  NiMh, 87, 88, 89–91
  See also single-use batteries
regulating voltage, 95–97
Relay shield, 128
relays
  Arduino-controlled, 112–116, 128–136
  controlling from a web page, 128–136
  description/illustration, 112
relay_test sketch, 115
resistance, 26
  converting to voltage, 37–38
  ranges on multimeters, 246
resistors, 20–22
  calculators for, 57
  color codes for, 21
  dividing voltage with, 34–36
  experiment on heating, 33–34
  LEDs and, 56–57, 122–123
  measuring with multimeter, 246
  part codes for, 258
  reading stripes on, 22
reverse-biased diode, 56
RF modules. *See* wireless remote modules
RGB LEDs, 60
  breadboard layout, 61
  test schematic, 60
robot rover project, 177–183
  construction of, 179–181
  items needed for, 178–179
  schematic diagram, 179
  software sketch, 182–183
  testing process, 181

S

safety glasses, 9
scavenging components, 242–243
schematic diagrams
  basic rules of, 29–30
  component symbols on, 30–31
  names and values on, 30
  reading, 14, 29–31
schematic diagrams (specific)
  accelerometer, 208
  amplifier module, 225
  Arduino-controlled relay, 113
  battery backup, 99

  - constant current driver, 63
  - FM radio bug, 221
  - fume extractor, 14
  - gas detector, 194
  - H-bridge module, 168, 170
  - LED light generation, 56, 58
  - light sensing push light, 40, 43
  - microphone module, 219
  - modified slot car, 80
  - MOSFET motor control, 48, 166
  - PIR motion sensor module, 150, 151
  - resistor heater, 34
  - robot rover, 179
  - seven-segment LED display, 184
  - solar trickle charging, 104
  - stepper motor control, 174
  - tone generator, 126, 228
  - voltage divider, 35
  - voltage regulator, 95
  - *See also* wiring diagrams
- screened wire, 6
- semiconductors, 259–260
- sensors, 193–212
  - color, 197–202
  - gas, 193–197
  - magnetic field, 210–212
  - motion, 149–153
  - temperature, 204–206
  - vibration, 202–204
  - *See also* modules
- Serial Monitor, 118, 232
- servo motor project, 139–141
  - construction of, 140
  - items needed for, 140
  - software sketch, 140–141
- setPins function, 145–146
- setup function, 110, 121, 141, 144
- seven-segment LED display module, 183–188
  - construction of, 186–187
  - description of, 183–185
  - items needed for, 185
  - schematic diagram, 184
  - software sketch, 187–188
- shields (Arduino), 127–128
  - commonly used, 128
  - Ethernet, 128–136
  - LCD, 128, 136–139
- signal generator app, 227
- simulation tools, 253, 254
- single throw switches, 52
- single-use batteries, 84–86
  - battery holders for, 85–86
  - types/characteristics of, 85, 86
  - *See also* rechargeable batteries
- sketches (Arduino), 107
  - accelerometer, 208–210
  - Blink sketch, 107, 109–111
  - Charlieplexed LEDs project, 143–145
  - color-sensing module, 199–202
  - electric toy project, 118–119
  - LCD display, 137–139, 187–188
  - LED control project, 123–125
  - magnetic field sensor, 212
  - MOSFET motor speed, 165–166
  - password typing project, 146–147
  - PIR motion sensor module, 152–153
  - real-time clock module, 191–192
  - robot rover project, 182–183
  - servo motor project, 140–141
  - stepper motor control, 175–177
  - temperature sensor, 205–206
  - tone generator project, 126–127
  - ultrasonic rangefinder module, 155–156, 158
  - USB music controller, 230–232
  - voltmeter sketch, 120–121
  - web-controlled toy, 132–136
  - wireless remote module, 161–162
- slot car racer project, 78–81
  - capacitor used in, 79–80
  - construction of, 81
  - design overview, 80–81
  - items needed for, 78
  - schematic diagram, 80
  - testing, 81
  - wiring diagram, 81

snips, wire, 6–7
software sketches. *See* sketches
software tools, 253–256
  EAGLE PCB, 254, 255
  fritzing, 253–254, 255
  online calculators, 256
  simulation, 253
software VU meter, 232–234
  construction of, 233
  software sketch, 233–234
solar panels, 101–104
  overview of using, 101–102
  power consumption issues, 104
  spreadsheet for monitoring, 103
  testing the performance of, 102–103
  trickle charging with, 103–104
  See also batteries; power
soldering, 8–12
  audio connectors, 215, 216
  dangers related to, 8–9
  desoldering and, 241–242
  hacking a fan for, 14–17
  items needed for, 9–10
  joining wires by, 11–12
  process overview, 10–11
  safety tips for, 9
solderless breadboard, 3–5
solid-core wire, 4, 5–6
sounds
  audio frequencies and, 224
  playing with Arduino, 125–127
  See also audio hacks
SparkFun
  amplifier module, 225
  H-Bridge module, 168–169
  LiPo battery charger, 93
  mic module, 219
  part codes, 257–262
  Starter Kits, 3, 19–20, 258
SPDT/SPST switches, 52
SPI library, 133
starter kits
  component, 19–20, 258
  toolkit, 3, 257
stepper motor control, 172–177
  Arduino used in, 173
  construction of, 175
  description of, 172–174
  H-bridge module, 173
  items needed for, 174
  schematic diagram, 174
  software sketch, 175–177
stereo-to-mono conversion, 217–218
stripboard, 71
  amplifier module layout, 226
  component soldering, 77
  cutting to size, 74
  LED flasher layout, 71–77
  making track breaks, 74–75
  resistor placement, 76, 77
  wire link soldering, 75–76
stripping wire, 5, 6–7
suppliers, 1–2, 3, 257
surface mount devices (SMDs), 25
switches, 50–53
  H-bridge using, 167
  labeled on diagrams, 15
  microswitches, 51
  push-button, 50–51
  toggle, 51–53
symbols, component, 30–31

T

TCS3200 IC, 197–198
TDA7052 IC, 224, 225
temperature
  measuring with multimeters, 247
  sensor for measuring, 204–206
temperature sensor, 204–206
  construction of, 205
  software sketch, 205–206
templates, stripboard, 71

testing
  amplifier module, 227
  batteries, 239
  components, 240–241
  connections, 12–13
  FM radio bug, 223
  fuses, 237–239
  heating elements, 240
  power supply polarity, 15–16
  robot rover, 181
  slot car racer, 81
  solar panels, 102–103
  transistors, 250–251
  web-controlled relays, 131
TMP36 sensor, 204–206
toggle switches, 51–53
tone generator, 125–127, 227–229
  555 timer-based, 227–229
  Arduino-based, 125–127
  construction of, 126, 229
  items needed for, 125, 229
  schematic diagrams, 126, 228
  software sketch, 126–127
toolkits, 3–5
tools, 3–5, 245–256
  EAGLE PCB, 254, 255
  fritzing, 253–254, 255
  lab power supply, 251–252
  multimeter, 245–251
  online calculators, 256
  oscilloscope, 252–253
  part codes for, 257–258
  simulation, 253, 254
  software, 253–256
toy hacking projects
  Arduino-controlled toy, 116–119
  web-controlled toy, 128–136
transistors, 24
  bipolar, 39–40, 45–47
  commonly used, 47
  datasheets for, 45–46
  MOSFET, 46, 163–166
  NPN, 39–40, 47
  PNP, 46–47
  testing, 250–251
trickle charging, 89, 90
  battery backup, 100–101
  solar panels for, 103–104
tweeters and woofers, 224
twisting wires together, 7–8

U

ultrasonic range finding, 154
ultrasonic rangefinder modules, 153–158
  general description of, 153–154
  HC-SR04 rangefinder, 155–157
  MaxBotix LV-EZ1 rangefinder, 157–158
ultraviolet LEDs, 60–61
USB music controller, 229–232
  construction of, 230
  software sketch, 230–232

V

valueOfParam function, 136
variable power supply, 91–92
variable resistors, 22
variable_led_brightness sketch, 124–125
variable_led_flash sketch, 123–124
VCR scavenging, 243
vibration detector, 202–204
  construction of, 202–203
  items needed for, 202
  software sketch, 203–204
voltage, 26–27
  Arduino for measuring, 119–121
  boosting from batteries, 97–98
  controlling from batteries, 95–97
  converting a resistance to, 37–38
  determining for LEDs, 66–67
  dividing with resistors, 34–36
  measuring with multimeters, 248–249
voltage dividers, 36, 119–121

voltage regulators, 95–97
  breadboard layout, 96
  common types of, 97
  sample schematic, 95
  See also constant current driver
voltmeter sketch, 120–121
VU meter, 232–234
  construction of, 233
  software sketch, 233–234

W

watts, 28
web resources, 256
  Arduino-related, 106, 128, 148, 256
  Hacking Electronics book, 71, 105, 257
  software tools, 253–256
web-controlled toy project, 128–136
  construction of, 130
  items needed for, 129
  network configuration, 131, 132
  software sketch, 132–136
  testing process, 131
  wiring diagram, 130
  See also electric toy project
wire
  desoldering, 241–242
  joining, 7–8, 11–12
  multi-core, 6
  part codes for, 257–258
  purchasing, 4
  screened, 6
  soldering, 11–12
  solid-core, 4, 5–6
  stripping, 5, 6–7
  twisting, 7–8
  types of, 5–6
wire snips, 6–7
wire strippers, 6
wireless remote modules, 159–162
  breadboard layout, 159–160
  items needed for, 159, 161
  software sketch, 161–162
  using with Arduino, 161–162
wiring diagrams
  fume extractor, 15
  light sensing push light, 44
  modified slot car, 81
  remote relay control, 130
  See also schematic diagrams
woofers and tweeters, 224

This page intentionally left blank

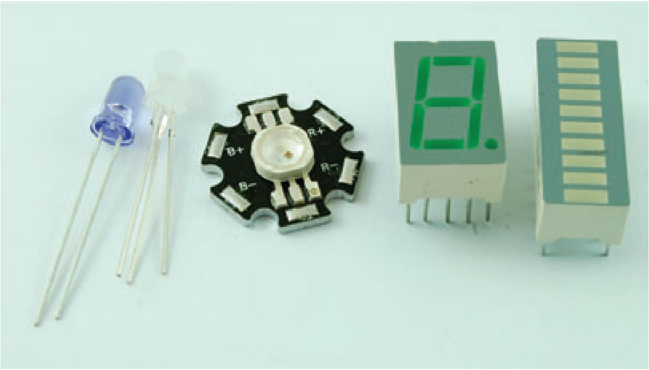

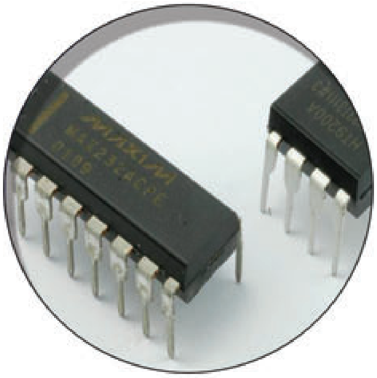

This page intentionally left blank

UNO
ARDUINO
DIGITAL (PWM~)

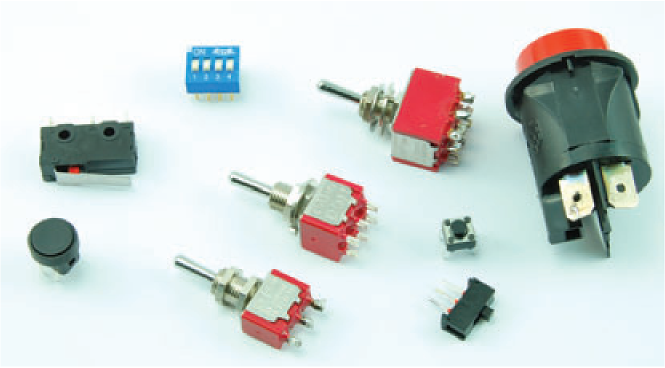

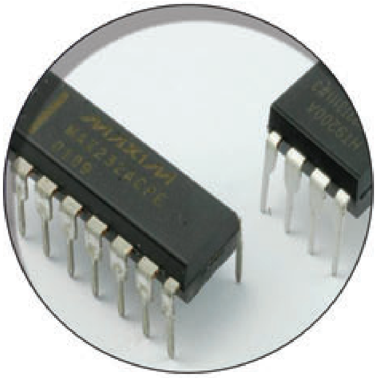
MAXIM

Edited by

Andrei D. Polyanin

Alexei I. Chernoutsan

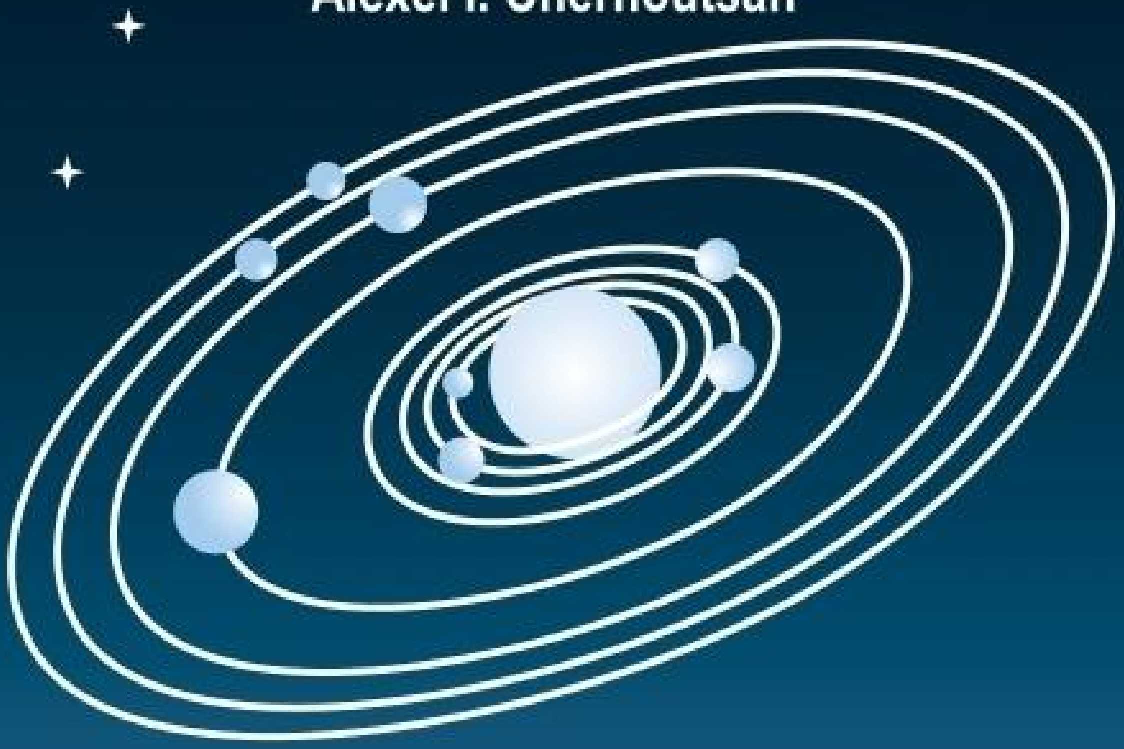

A CONCISE HANDBOOK OF MATHEMATICS, PHYSICS, AND ENGINEERING SCIENCES

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

A CONCISE HANDBOOK OF MATHEMATICS,

9/10/10 9:28 AM

A CONCISE HANDBOOK OF MATHEMATICS, PHYSICS, AND ENGINEERING SCIENCES

A CONCISE HANDBOOK OF MATHEMATICS, PHYSICS, AND ENGINEERING SCIENCES

Edited by

Andrei D. Polyanin

Alexei I. Chernoutsan

Authors

A.I. Chernoutsan, A.V. Egorov, A.V. Manzhirov, A.D. Polyanin, V.D. Polyanin, V.A. Popov, B.V. Putyatin, Yu.V. Repina, V.M. Safrai, A.I. Zhurov

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2011 by Taylor and Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-1-4398-0640-1 (Ebook-PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

CONTENTS

Preface . . . . **xxv**

Editors . . . . **xxvii**

Part I. Mathematics **1**

M1. Arithmetic and Elementary Algebra **3**

M1.1. Real Numbers 3
  M1.1.1. Integer Numbers 3
  M1.1.2. Real, Rational, and Irrational Numbers 4

M1.2. Equalities and Inequalities. Arithmetic Operations. Absolute Value 5
  M1.2.1. Equalities and Inequalities 5
  M1.2.2. Addition and Multiplication of Numbers 5
  M1.2.3. Ratios and Proportions 6
  M1.2.4. Percentage 6
  M1.2.5. Absolute Value of a Number (Modulus of a Number) 7

M1.3. Powers and Logarithms 7
  M1.3.1. Powers and Roots 7
  M1.3.2. Logarithms 8

M1.4. Binomial Theorem and Related Formulas 9
  M1.4.1. Factorials. Binomial Coefficients. Binomial Theorem 9
  M1.4.2. Related Formulas 9

M1.5. Progressions 10
  M1.5.1. Arithmetic Progression 10
  M1.5.2. Geometric Progression 10

M1.6. Mean Values and Some Inequalities 11
  M1.6.1. Arithmetic Mean, Geometric Mean, and Other Mean Values 11
  M1.6.2. Inequalities for Mean Values 11
  M1.6.3. Some Inequalities of General Form 11

M1.7. Some Mathematical Methods 12
  M1.7.1. Proof by Contradiction 12
  M1.7.2. Mathematical Induction 12
  M1.7.3. Proof by Counterexample 13

Bibliography for Chapter M1 13

M2. Elementary Functions **15**

M2.1. Power, Exponential, and Logarithmic Functions 15
  M2.1.1. Power Function: $y = x^\alpha$ 15
  M2.1.2. Exponential Function: $y = a^x$ 17
  M2.1.3. Logarithmic Function: $y = \log_a x$ 18

M2.2. Trigonometric Functions 19
  M2.2.1. Trigonometric Circle. Definition of Trigonometric Functions 19
  M2.2.2. Graphs of Trigonometric Functions 21
  M2.2.3. Properties of Trigonometric Functions 22

M2.3. Inverse Trigonometric Functions 26
    M2.3.1. Definitions. Graphs of Inverse Trigonometric Functions 26
    M2.3.2. Properties of Inverse Trigonometric Functions 27

M2.4. Hyperbolic Functions 29
    M2.4.1. Definitions. Graphs of Hyperbolic Functions 29
    M2.4.2. Properties of Hyperbolic Functions 30

M2.5. Inverse Hyperbolic Functions 33
    M2.5.1. Definitions. Graphs of Inverse Hyperbolic Functions 33
    M2.5.2. Properties of Inverse Hyperbolic Functions 34

Bibliography for Chapter M2 35

M3. Elementary Geometry 37

M3.1. Plane Geometry 37
    M3.1.1. Triangles 37
    M3.1.2. Polygons 43
    M3.1.3. Circle 46

M3.2. Solid Geometry 49
    M3.2.1. Straight Lines, Planes, and Angles in Space 49
    M3.2.2. Polyhedra 51
    M3.2.3. Solids Formed by Revolution of Lines 55

Bibliography for Chapter M3 60

M4. Analytic Geometry 61

M4.1. Points, Segments, and Coordinate Plane 61
    M4.1.1. Cartesian and Polar Coordinates on Plane 61
    M4.1.2. Distance Between Points. Division of Segment in Given Ratio. Area of a Polygon 62

M4.2. Straight Lines and Points on Plane 64
    M4.2.1. Equations of Straight Lines on Plane 64
    M4.2.2. Mutual Arrangement of Points and Straight Lines 66

M4.3. Quadratic Curves 68
    M4.3.1. Circle 68
    M4.3.2. Ellipse 69
    M4.3.3. Hyperbola 71
    M4.3.4. Parabola 74
    M4.3.5. Transformation of Quadratic Curves to Canonical Form 75

M4.4. Coordinates, Vectors, Curves, and Surfaces in Space 78
    M4.4.1. Vectors and Their Properties 78
    M4.4.2. Coordinate Systems 79
    M4.4.3. Scalar, Cross, and Scalar Triple Products of Vectors 82

M4.5. Line and Plane in Space 84
    M4.5.1. Plane in Space 84
    M4.5.2. Line in Space 89
    M4.5.3. Mutual Arrangement of Points, Lines, and Planes 91

M4.6. Quadric Surfaces (Quadrics) 96
    M4.6.1. Quadrics and Their Canonical Equations 96
    M4.6.2. Quadrics (General Theory) 99

Bibliography for Chapter M4 102

M5. Algebra **103**

M5.1. Polynomials and Algebraic Equations 103
  M5.1.1. Polynomials and Their Properties 103
  M5.1.2. Linear and Quadratic Equations 105
  M5.1.3. Cubic Equations 105
  M5.1.4. Fourth-Degree Equation 107
  M5.1.5. Algebraic Equations of Arbitrary Degree and Their Properties 108

M5.2. Determinants and Matrices 112
  M5.2.1. Determinants 112
  M5.2.2. Matrices. Types of Matrices. Operations with Matrices 114
  M5.2.3. Inverse Matrix. Functions of Matrices 119
  M5.2.4. Eigenvalues and Characteristic Equation of a Matrix. The Cayley–Hamilton Theorem 120

M5.3. Systems of Linear Algebraic Equations 123
  M5.3.1. Consistency Condition for a Linear System 123
  M5.3.2. Finding Solutions of a System of Linear Equations 124

M5.4. Quadratic Forms 127
  M5.4.1. Quadratic Forms and Their Transformations 127
  M5.4.2. Canonical and Normal Representations of a Quadratic Form 129

M5.5. Linear Spaces 131
  M5.5.1. Concept of a Linear Space. Its Basis and Dimension 131
  M5.5.2. Subspaces of Linear Spaces 134
  M5.5.3. Coordinate Transformations Corresponding to Basis Transformations in a Linear Space 135
  M5.5.4. Euclidean Space 136

Bibliography for Chapter M5 138

M6. Limits and Derivatives **139**

M6.1. Basic Concepts of Mathematical Analysis 139
  M6.1.1. Number Sets. Functions of Real Variable 139
  M6.1.2. Limit of a Sequence 141
  M6.1.3. Limit of a Function. Asymptotes 144
  M6.1.4. Infinitely Small and Infinitely Large Functions 146
  M6.1.5. Continuous Functions. Discontinuities of the First and the Second Kind 147
  M6.1.6. Convex and Concave Functions 148
  M6.1.7. Convergence of Functions 149

M6.2. Differential Calculus for Functions of a Single Variable 150
  M6.2.1. Derivative and Differential: Their Geometrical and Physical Meaning 150
  M6.2.2. Table of Derivatives and Differentiation Rules 151
  M6.2.3. Theorems about Differentiable Functions. L’Hospital Rule 153
  M6.2.4. Higher-Order Derivatives and Differentials. Taylor’s Formula 154
  M6.2.5. Extremal Points. Points of Inflection 156
  M6.2.6. Qualitative Analysis of Functions and Construction of Graphs 158
  M6.2.7. Approximate Solution of Equations (Root-Finding Algorithms for Continuous Functions) 160

M6.3. Functions of Several Variables. Partial Derivatives 162
  M6.3.1. Point Sets. Functions. Limits and Continuity 162
  M6.3.2. Differentiation of Functions of Several Variables 163
  M6.3.3. Directional Derivative. Gradient. Geometrical Applications 166
  M6.3.4. Extremal Points of Functions of Several Variables 167
  M6.3.5. Differential Operators of the Field Theory 170

Bibliography for Chapter M6 170

M7. Integrals **171**

M7.1. Indefinite Integral 171
  M7.1.1. Antiderivative. Indefinite Integral and Its Properties 171
  M7.1.2. Table of Basic Integrals. Properties of the Indefinite Integral. Examples of Integration 172
  M7.1.3. Integration of Rational Functions 174
  M7.1.4. Integration of Irrational Functions 177
  M7.1.5. Integration of Exponential and Trigonometric Functions 179
  M7.1.6. Integration of Polynomials Multiplied by Elementary Functions . . . 181

M7.2. Definite Integral 183
  M7.2.1. Basic Definitions. Classes of Integrable Functions. Geometrical Meaning of the Definite Integral 183
  M7.2.2. Properties of Definite Integrals and Useful Formulas 184
  M7.2.3. Asymptotic Formulas for the Calculation of Integrals 185
  M7.2.4. Mean Value Theorems. Properties of Integrals in Terms of Inequalities 187
  M7.2.5. Geometric and Physical Applications of the Definite Integral 188
  M7.2.6. Improper Integrals with Infinite Integration Limits 190
  M7.2.7. Improper Integrals of Unbounded Functions 193
  M7.2.8. Approximate (Numerical) Methods for Computation of Definite Integrals 194

M7.3. Double and Triple Integrals 195
  M7.3.1. Definition and Properties of the Double Integral 195
  M7.3.2. Computation of the Double Integral 197
  M7.3.3. Geometric and Physical Applications of the Double Integral 199
  M7.3.4. Definition and Properties of the Triple Integral 200
  M7.3.5. Computation of the Triple Integral. Some Applications. Iterated Integrals and Asymptotic Formulas 201

M7.4. Line and Surface Integrals 203
  M7.4.1. Line Integral of the First Kind 203
  M7.4.2. Line Integral of the Second Kind 205
  M7.4.3. Surface Integral of the First Kind 206
  M7.4.4. Surface Integral of the Second Kind 207
  M7.4.5. Integral Formulas of Vector Calculus 208

Bibliography for Chapter M7 209

M8. Series **211**

M8.1. Numerical Series and Infinite Products 211
  M8.1.1. Convergent Numerical Series and Their Properties. Cauchy's Criterion 211
  M8.1.2. Convergence Criteria for Series with Positive (Nonnegative) Terms 212
  M8.1.3. Convergence Criteria for Arbitrary Numerical Series. Absolute and Conditional Convergence 213

M8.1.4. Multiplication of Series. Some Inequalities 215

M8.2. Function Series 215
M8.2.1. Pointwise and Uniform Convergence of Function Series 215
M8.2.2. Basic Criteria of Uniform Convergence. Properties of Uniformly Convergent Series 216

M8.3. Power Series 218
M8.3.1. Radius of Convergence of Power Series. Properties of Power Series 218
M8.3.2. Taylor and Maclaurin Power Series 220
M8.3.3. Operations with Power Series. Summation Formulas for Power Series 221

M8.4. Fourier Series 223
M8.4.1. Representation of 2π-Periodic Functions by Fourier Series. Main Results 223
M8.4.2. Fourier Expansions of Periodic, Nonperiodic, Even, and Odd Functions 225
M8.4.3. Criteria of Uniform and Mean-Square Convergence of Fourier Series 226

Bibliography for Chapter M8 227

M9. Functions of Complex Variable **229**

M9.1. Complex Numbers 229
M9.1.1. Definition of a Complex Number. Arithmetic Operations with Complex Numbers 229
M9.1.2. Trigonometric Form of Complex Numbers. Powers and Radicals .. 230

M9.2. Functions of Complex Variables 231
M9.2.1. Basic Concepts. Differentiation of a Function of a Complex Variable 231
M9.2.2. Integration of Functions of Complex Variables 238
M9.2.3. Taylor and Laurent Series 241
M9.2.4. Zeros and Isolated Singularities of Analytic Functions 243
M9.2.5. Residues. Calculation of Definite Integrals 245

Bibliography for Chapter M9 249

M10. Integral Transforms **251**

M10.1. General Form of Integral Transforms. Inversion Formulas 251

M10.2. Laplace Transform 251
M10.2.1. Laplace Transform and the Inverse Laplace Transform 251
M10.2.2. Main Properties of the Laplace Transform. Inversion Formulas for Some Functions 252
M10.2.3. Limit Theorems. Representation of Inverse Transforms as Convergent Series 255

M10.3. Various Forms of the Fourier Transform 255
M10.3.1. Fourier Transform and the Inverse Fourier Transform 255
M10.3.2. Fourier Cosine and Sine Transforms 257

M10.4. Mellin Transform and Other Transforms 258
M10.4.1. Mellin Transform and the Inversion Formula 258
M10.4.2. Main Properties of the Mellin Transform. Relation Among the Mellin, Laplace, and Fourier Transforms 259
M10.4.3. Summary Table of Integral Transforms 259

Bibliography for Chapter M10 260

M11. Ordinary Differential Equations **261**

M11.1. First-Order Differential Equations 261
  M11.1.1. General Concepts. The Cauchy Problem. Uniqueness and Existence Theorems 261
  M11.1.2. Equations Solved for the Derivative. Simplest Techniques of Integration 263
  M11.1.3. Exact Differential Equations. Integrating Factor 265
  M11.1.4. Riccati Equation 267
  M11.1.5. Equations Not Solved for the Derivative 268
  M11.1.6. Approximate Analytic Methods for Solution of Equations 269
  M11.1.7. Numerical Integration of Differential Equations 270

M11.2. Second-Order Linear Differential Equations 271
  M11.2.1. Formulas for the General Solution. Some Transformations 271
  M11.2.2. Representation of Solutions as a Series in the Independent Variable 274
  M11.2.3. Boundary Value Problems 276
  M11.2.4. Eigenvalue Problems 277

M11.3. Second-Order Nonlinear Differential Equations 281
  M11.3.1. Form of the General Solution. Cauchy Problem 281
  M11.3.2. Equations Admitting Reduction of Order 281
  M11.3.3. Methods of Regular Series Expansions with Respect to the Independent Variable 283
  M11.3.4. Perturbation Methods in Problems with a Small Parameter 284
  M11.3.5. Galerkin Method and Its Modifications (Projection Methods) 287
  M11.3.6. Iteration and Numerical Methods 291

M11.4. Linear Equations of Arbitrary Order and Linear Systems of Equations 292
  M11.4.1. Linear Equations with Constant Coefficients 292
  M11.4.2. Linear Equations with Variable Coefficients 296
  M11.4.3. Systems of Linear Equations with Constant Coefficients 298

M11.5. Nonlinear Equations of Arbitrary Order 301
  M11.5.1. Structure of the General Solution. Cauchy Problem 301
  M11.5.2. Equations Admitting Reduction of Order 302

Bibliography for Chapter M11 303

M12. Partial Differential Equations **305**

M12.1. First-Order Quasilinear Partial Differential Equations 305
  M12.1.1. Characteristic System. General Solution 305
  M12.1.2. Cauchy Problem 306

M12.2. Classification of Second-Order Linear Partial Differential Equations 307
  M12.2.1. Equations with Two Independent Variables 307
  M12.2.2. Equations with Many Independent Variables 310

M12.3. Basic Problems for Linear Equations of Mathematical Physics 311
  M12.3.1. Initial and Boundary Conditions. Cauchy Problem. Boundary Value Problems 311
  M12.3.2. First, Second, Third, and Mixed Boundary Value Problems 313

M12.4. Properties and Exact Solutions of Linear Equations 314
  M12.4.1. Homogeneous Linear Equations and Their Particular Solutions . . 314
  M12.4.2. Nonhomogeneous Linear Equations and Their Particular Solutions 319

M12.5. Method of Separation of Variables (Fourier Method) 320
  M12.5.1. Solution of Problems for Parabolic Equations 320
  M12.5.2. Solution of Problems for Hyperbolic Equations 324
  M12.5.3. Solution of Problems for Elliptic Equations 326

M12.6. Integral Transforms Method 328
  M12.6.1. Laplace Transform and Its Application to Linear Mathematical Physics Equations 328
  M12.6.2. Fourier Transform and Its Application for Linear Equations of Mathematical Physics 331

M12.7. Boundary Value Problems for Parabolic Equations with One Space Variable. Green's Function 332
  M12.7.1. Representation of Solutions via the Green's Function 332
  M12.7.2. Problems for Parabolic Equation $\rho(x)\frac{\partial w}{\partial t} = \frac{\partial}{\partial x}\left[p(x)\frac{\partial w}{\partial x}\right] - q(x)w + \Phi(x,t)$ 334

M12.8. Boundary Value Problems for Hyperbolic Equations with One Space Variable. Green's Function. Goursat Problem 336
  M12.8.1. Representation of Solutions in terms of the Green's Function 336
  M12.8.2. Problems for Hyperbolic Equation $\rho(x)\frac{\partial^2 w}{\partial t^2} = \frac{\partial}{\partial x}\left[p(x)\frac{\partial w}{\partial x}\right] - q(x)w + \Phi(x,t)$ 338
  M12.8.3. Goursat Problem (a Problem with Initial Data of Characteristics) 340

M12.9. Boundary Value Problems for Elliptic Equations with Two Space Variables 342
  M12.9.1. Problems and the Green's Functions for Equation $a(x)\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + b(x)\frac{\partial w}{\partial x} + c(x)w = -\Phi(x,y)$ 342
  M12.9.2. Representation of Solutions to Boundary Value Problems via the Green's Functions 343
  M12.9.3. Solutions of Problems for the Laplace Equation in the Polar Coordinate System 344

M12.10. Problems with Many Space Variables. Representation of Solutions via Green's Functions 345
  M12.10.1. Boundary Value Problems for Parabolic Equations 345
  M12.10.2. Boundary Value Problems for Hyperbolic Equations 347
  M12.10.3. Boundary Value Problems for Elliptic Equations 348
  M12.10.4. Construction of the Green's Functions. General Formulas and Relations 348

M12.11. Duhamel's Principles in Nonstationary Problems 349
  M12.11.1. Problems for Homogeneous Linear Equations 349
  M12.11.2. Problems for Nonhomogeneous Linear Equations 351

Bibliography for Chapter M12 352

M13. Special Functions and Their Properties 355

M13.1. Some Coefficients, Symbols, and Numbers 355
  M13.1.1. Factorials. Binomial Coefficients. Pochhammer Symbol 355
  M13.1.2. Bernoulli Numbers 356

M13.2. Error Functions. Exponential and Logarithmic Integrals 356
  M13.2.1. Error Function and Complementary Error Function 356
  M13.2.2. Exponential Integral 357
  M13.2.3. Logarithmic Integral 357

M13.3. Sine Integral and Cosine Integral. Fresnel Integrals 358
  M13.3.1. Sine Integral 358
  M13.3.2. Cosine Integral 358
  M13.3.3. Fresnel Integrals 359

M13.4. Gamma Function, Psi Function, and Beta Function 359
  M13.4.1. Gamma Function 359
  M13.4.2. Psi Function (Digamma Function) 360
  M13.4.3. Beta Function 361

M13.5. Incomplete Gamma Function 361
  M13.5.1. Integral Representations. Recurrence Formulas 361
  M13.5.2. Expansions as $x \to 0$ and $x \to \infty$. Relation to Other Functions 362

M13.6. Bessel Functions (Cylinder Functions) 362
  M13.6.1. Definitions and Basic Formulas 362
  M13.6.2. Integral Representations and Asymptotic Expansions 364

M13.7. Modified Bessel Functions 365
  M13.7.1. Definitions. Basic Formulas 365
  M13.7.2. Integral Representations and Asymptotic Expansions 366

M13.8. Degenerate Hypergeometric Functions (Kummer Functions) 367
  M13.8.1. Definitions and Basic Formulas 367
  M13.8.2. Integral Representations and Asymptotic Expansions 370

M13.9. Elliptic Integrals 370
  M13.9.1. Complete Elliptic Integrals 370
  M13.9.2. Incomplete Elliptic Integrals 372

M13.10. Orthogonal Polynomials 373
  M13.10.1. Legendre Polynomials and Legendre Functions 373
  M13.10.2. Laguerre Polynomials and Generalized Laguerre Polynomials . 375
  M13.10.3. Chebyshev Polynomials 376
  M13.10.4. Hermite Polynomials 377

Bibliography for Chapter M13 378

M14. Probability Theory 379

M14.1. Basic Concepts and Simplest Probabilistic Models 379
  M14.1.1. Rules and Formulas of Combinatorics 379
  M14.1.2. Probabilities of Random Events 380
  M14.1.3. Conditional Probability. Total Probability Formula 382
  M14.1.4. Sequence of Trials 384

M14.2. Random Variables and Their Characteristics 386
  M14.2.1. One-Dimensional Random Variables 386
  M14.2.2. Expectation, Variance and Moments of a Random Variable 388
  M14.2.3. Main Discrete Distributions and Their Numerical Characteristics 390
  M14.2.4. Main Continuous Distributions and Their Numerical Characteristics 393
  M14.2.5. Two-dimensional and Multivariate Random Variables 396

M14.3. Limit Theorems 399
  M14.3.1. Convergence of Random Variables 399
  M14.3.2. Limit Theorems 400

Bibliography for Chapter M14 402

Part II. Physics **403**

P1. Physical Foundations of Mechanics **405**

P1.1. Kinematics of a Point 405
  P1.1.1. Basic Definitions. Velocity and Acceleration 405
  P1.1.2. The Direct and Inverse Problems of Kinematics 406
  P1.1.3. Circular Motion 407
  P1.1.4. Relativity of Motion. Addition of Velocities 407

P1.2. Kinematics of a Rigid Body 408
  P1.2.1. Translational and Rotational Motion. Plane Motion 408
  P1.2.2. Instantaneous Axis of Rotation. Addition of Angular Velocities. Motion Relative to a Rotational Frame 408

P1.3. Dynamics 410
  P1.3.1. Newton's First Law. Mass. Momentum. Force 410
  P1.3.2. Newton's Second and Third Laws 411
  P1.3.3. Action at a Distance. Causality. Natural Forces 411
  P1.3.4. Galilean Relativity Principle 412

P1.4. Law of Conservation of Momentum 413
  P1.4.1. Center of Mass. Momentum of a System of Particles 413
  P1.4.2. Law of Conservation of Momentum. Motion of Center of Mass 414

P1.5. Law of Conservation of Energy 415
  P1.5.1. Work and Power. Kinetic Energy 415
  P1.5.2. Conservative Forces. Potential Energy 416
  P1.5.3. Mechanical Energy of a System of Particles. Law of Conservation of Mechanical Energy 418
  P1.5.4. Potential Curves. Stability 419

P1.6. Law of Conservation of Angular Momentum 420
  P1.6.1. Moment of Force. Net Force 420
  P1.6.2. Angular Momentum of a Particle. Motion in Central Field 421
  P1.6.3. Conservation of Angular Momentum for a System of Particles 423
  P1.6.4. Two-Body Problem. Reduced Mass 424

P1.7. Gravitational Field 425
  P1.7.1. Law of Universal Gravitation. Superposition Principle 425
  P1.7.2. Strength and Potential of Gravitational Field 426
  P1.7.3. Motion in a Central Field of Gravity. Kepler's Laws. Orbital and Escape Velocities 427

P1.8. Non-Inertial Reference Frames 427
  P1.8.1. Definition of Fictitious Forces 427
  P1.8.2. Translationally Moving Non-Inertial Frames 428
  P1.8.3. Uniformly Rotating Reference Frame 428

P1.9. Dynamics of Rigid Body 429
  P1.9.1. Rotation About a Fixed Axis 429
  P1.9.2. Properties of Moment of Inertia 430
  P1.9.3. Plane Motion of a Rigid Body 431
  P1.9.4. Motion with a Fixed Point 432

P1.10. Special Theory of Relativity 434
  P1.10.1. Basics of the Special Theory of Relativity 434
  P1.10.2. Consequences of the Lorentz Transformations. STR Kinematics . . 435
  P1.10.3. STR Dynamics 437

Bibliography for Chapter P1 439

P2. Molecular Physics and Thermodynamics **441**

P2.1. Basic Principles and Definitions 441
  P2.1.1. Two Approaches to Studying Matter 441
  P2.1.2. Equation of State 441

P2.2. First Law of Thermodynamics 443
  P2.2.1. Quantity of Heat. Internal Energy 443
  P2.2.2. First Law of Thermodynamics 443
  P2.2.3. First Law of Thermodynamics for an Ideal Gas 445

P2.3. Second Law of Thermodynamics 446
  P2.3.1. Formulations 446
  P2.3.2. Carnot's Theorem 446
  P2.3.3. Calculation of the Internal Energy 447

P2.4. Entropy 448
  P2.4.1. Clausius Inequality 448
  P2.4.2. Definition of Entropy 449
  P2.4.3. Direction of Non-Equilibrium Processes in a Thermally Insulated System 449
  P2.4.4. Statistical Meaning of Entropy 450
  P2.4.5. Third Law of Thermodynamics 451
  P2.4.6. Thermodynamic Potentials 451

P2.5. Kinetic Theory of Ideal Gases 453
  P2.5.1. The Basic Equation of the Kinetic Theory 453
  P2.5.2. Temperature and Energy of an Ideal Gas 454
  P2.5.3. Maxwell's Distribution 456
  P2.5.4. Boltzmann's Distribution 457

P2.6. Real Gases. Van Der Waals Equation 458
  P2.6.1. Isotherms of a Real Gas 458
  P2.6.2. Van Der Waals Equation 459

P2.7. Phase Equilibrium. Phase Transitions 461
  P2.7.1. Conditions for Phase Equilibrium 461
  P2.7.2. First-Order Phase Transitions 462

P2.8. Surface Tension 462
  P2.8.1. Surface Energy 462
  P2.8.2. Additional Pressure Under a Curved Surface 463
  P2.8.3. Wetting. Capillary Effects 464

P2.9. Transfer Phenomena in Gases 464
  P2.9.1. Collision of Molecules in a Gas 464
  P2.9.2. Qualitative Model of Transfer Phenomena in Gases 465

Bibliography for Chapter P2 467

P3. Electrodynamics **469**

P3.1. Electric Charge. Coulomb's Law 469
  P3.1.1. Electric Charge and Its Properties 469
  P3.1.2. Coulomb's Law 470

P3.2. Electric Field. Strength and Potential 470
  P3.2.1. Strength of Electric Field 470
  P3.2.2. Potential of an Electrostatic Field 472

P3.3. Gauss's Theorem 474
  P3.3.1. Statement of the Theorem 474
  P3.3.2. Application of Gauss's Theorem 475

P3.4. Conductors and Capacitors 477
  P3.4.1. Electrostatics of Conductors 477
  P3.4.2. Capacitors 479

P3.5. Electrostatics of Dielectrics 481
  P3.5.1. Polarization of Dielectrics 481
  P3.5.2. Types of Dielectric and Polarization Mechanisms 484

P3.6. Energy of Electrostatic Field 485
  P3.6.1. Energy of a System of Charges 485
  P3.6.2. Energy of an Electrostatic Field 486

P3.7. Direct Current 488
  P3.7.1. Electric Current in a Circuit 488
  P3.7.2. Current in a Closed Circuit 490
  P3.7.3. Classical Free-Electron Theory of Metals 491

P3.8. Constant Magnetic Field 492
  P3.8.1. Lorentz Force and Ampère Force 492
  P3.8.2. Calculation of Magnetic Induction 494
  P3.8.3. Ampère's Circuital Law and Flux of Magnetic Induction 496

P3.9. Magnetic Field in Matter 498
  P3.9.1. Magnetization. Molecular Currents 498
  P3.9.2. Types of Magnetic Materials and Magnetization Mechanisms 500

P3.10. Electromagnetic Induction 502
  P3.10.1. Faraday's Law of Electromagnetic Induction 502
  P3.10.2. Self-Induction and Mutual Induction 504
  P3.10.3. Energy of Magnetic Field 505

P3.11. Maxwell's Equations 506
  P3.11.1. Maxwell's Equations in Vacuum and Matter 506
  P3.11.2. Law of Conservation of Energy for Electromagnetic Fields 508
  P3.11.3. Relativistic Properties of Fields 509

Bibliography for Chapter P3 510

P4. Oscillations and Waves 511

P4.1. Oscillations 511
  P4.1.1. Harmonic Oscillations. Composition of Oscillations 511
  P4.1.2. Free Undamped Oscillations 513
  P4.1.3. Damped and Forced Oscillations 517

P4.2. Wave Processes 521
  P4.2.1. Elastic and Acoustic Waves 521
  P4.2.2. Electromagnetic Waves 528

Bibliography for Chapter P4 533

P5. Optics 535

P5.1. Geometric Optics. Radiometry 535
  P5.1.1. Geometric Optics 535
  P5.1.2. Radiometry and Photometry 537

P5.2. Wave Properties of Light 539
  P5.2.1. Interference of Light 539
  P5.2.2. Diffraction 545

P5.3. Interaction Between Light and Matter 551
  P5.3.1. Light Polarization. Fresnel Formulas 551
  P5.3.2. Light Dispersion and Absorption 555

P5.4. Foundations of Quantum Optics 558
  P5.4.1. Thermal Radiation 558
  P5.4.2. Light Quanta 562
  P5.4.3. Principle of Laser Operation 565

Bibliography for Chapter P5 568

P6. Quantum Mechanics. Atomic Physics 569

P6.1. Atomic Structure. Bohr Model 569
  P6.1.1. Nuclear Atomic Model 569
  P6.1.2. Bohr Model 570

P6.2. Elements of Quantum Mechanics 573
  P6.2.1. Wave Function. Uncertainty Relation 573
  P6.2.2. The Schrödinger Equation 574
  P6.2.3. Operator Approach in Quantum Mechanics 578
  P6.2.4. Motion in a Central Field. Hydrogen Atom 580
  P6.2.5. Electron Spin and Fine Structure of the Spectrum 585

P6.3. Structure and Spectra of Complex Atoms 586
  P6.3.1. Pauli Exclusion Principle. Structure of Complex Atoms 586
  P6.3.2. Spectra of Complex Atoms 588

P6.4. Elements of Physics of Molecules 591
  P6.4.1. Classification of Levels of a Molecule 592
  P6.4.2. Electron Energy. Theory of Chemical Bond 593
  P6.4.3. Vibrational and Rotational Spectra 594

Bibliography for Chapter P6 596

P7. Quantum Theory of Crystals 597

P7.1. Quantum Statistical Distributions and Their Application 597
  P7.1.1. Elements of Quantum Statistics 597
  P7.1.2. Electron Gas in Metal 601
  P7.1.3. Heat Capacity of Crystal Lattice. Phonons 604

P7.2. Electrons in a Periodic Field. Energy Bands 608
  P7.2.1. Qualitative Theory of Energy Bands 609
  P7.2.2. Quantum Mechanical Description of Energy Regions 611
  P7.2.3. Dynamics of Electron in Crystal 612
  P7.2.4. Band Theory of Semiconductors 614

Bibliography for Chapter P7 616

P8. Elements of Nuclear Physics 617

P8.1. Basic Properties of Nuclei 617
  P8.1.1. Characteristics of Nuclei 617
  P8.1.2. Nuclear Forces 621
  P8.1.3. Nuclear Models 623

P8.2. Nuclear Transformations . . . 625
P8.2.1. Radioactivity . . . 625
P8.2.2. Nuclear Reactions . . . 632
P8.3. Nuclear Power Engineering . . . 637
P8.3.1. Nuclear Fission . . . 637
P8.3.2. Fusion Reactions . . . 641
P8.4. Subatomic Particles . . . 642
P8.4.1. Classes of Subatomic Particles . . . 643
P8.4.2. Characteristics of Particles . . . 647
P8.4.3. Interaction Unification . . . 650
Bibliography for Chapter P8 . . . 651

Part III. Elements of Applied and Engineering Sciences 653

E1. Dimensions and Similarity . . . 655
E1.1. Dimensions . . . 655
E1.1.1. Basic and Derived Units of Measurement. Systems of Units of Measurement . . . 655
E1.1.2. Dimensional and Dimensionless Quantities. Dimensional Formula . . 655
E1.2. The π-Theorem and Its Practical Use . . . 656
E1.2.1. The π-Theorem . . . 656
E1.2.2. Simplest Examples of the Use of the π-Theorem . . . 659
E1.3. Similarity and Simulation. General Remarks . . . 664
E1.3.1. Similarity and Simulation . . . 664
E1.3.2. General Concluding Remarks . . . 665
Bibliography for Chapter E1 . . . 665

E2. Mechanics of Point Particles and Rigid Bodies . . . 667
E2.1. Kinematics . . . 667
E2.1.1. Kinematics of a Point . . . 667
E2.1.2. Kinematics of a Rigid Body . . . 671
E2.1.3. Plane-Parallel (Plane) Motion of a Rigid Body . . . 675
E2.1.4. Arbitrary Motion of a Rigid Body . . . 678
E2.1.5. Compound Motion of a Point . . . 681
E2.1.6. Compound Motion of a Body . . . 684
E2.2. Basic Notions and Laws of Mechanics . . . 687
E2.2.1. Basic Notions of Mechanics . . . 687
E2.2.2. Basic Laws of Mechanics . . . 689
E2.3. Statics . . . 690
E2.3.1. Basic Laws and Theorems of Statics . . . 690
E2.3.2. Balance Conditions for a System of Forces Applied to a Rigid Body 692
E2.3.3. Solution of the Problems of Statics . . . 693
E2.3.4. Center of Parallel Forces. Center of Gravity of a Rigid Body . . . 698
E2.3.5. Distributed Forces . . . 699
E2.3.6. Friction Laws (Coulomb Laws) . . . 700
E2.4. Dynamics of a Point Particle . . . 701
E2.4.1. Equations of Motion of a Point Particle . . . 701
E2.4.2. First and Second Problems of Dynamics . . . 702

E2.5. General Theorems of Dynamics of a Mechanical System 703
E2.5.1. Basic Notions and Definitions 703
E2.5.2. Theorem on the Motion of the Center of Mass 707
E2.5.3. Theorem on the Momentum 708
E2.5.4. Theorem on the Angular Momentum 710
E2.5.5. Theorem on the Kinetic Energy 712

E2.6. Elements of Analytical Mechanics 716
E2.6.1. D'Alembert's Principle 716
E2.6.2. Classification of Mechanical Constraints. Generalized Coordinates . 718
E2.6.3. Principle of Virtual Displacements 720
E2.6.4. General Equation of Dynamics (d'Alembert–Lagrange Principle) ... 722
E2.6.5. Lagrange Equations of the Second Kind 723

E2.7. Small Oscillations of Mechanical Systems 724
E2.7.1. Preliminaries 724
E2.7.2. Small Oscillations of a Conservative System 725
E2.7.3. Normal Coordinates 727
E2.7.4. Influence of Small Dissipative Forces on System Oscillations 728
E2.7.5. Forced Oscillations 729

E2.8. Dynamics of an Absolutely Rigid Body 730
E2.8.1. Rotation of a Rigid Body about a Fixed Axis 730
E2.8.2. Plane-Parallel (Plane) Motion of a Rigid Body 731
E2.8.3. Motion of a Rigid Body about a Fixed Point 732
E2.8.4. Elementary Theory of Gyroscope 737

E2.9. Elementary Theory of Impact 738

Bibliography for Chapter E2 743

E3. Elements of Strength of Materials 745

E3.1. Basic Notions 745
E3.1.1. Introduction. External and Internal Forces 745
E3.1.2. Stresses and Strains at a Point 746
E3.1.3. Basic Notions and Hypotheses 747

E3.2. Stress–Strain State at a Point 748
E3.2.1. Types of Stress State 748
E3.2.2. Uniaxial Tension and Compression 748
E3.2.3. Simple Shear 749

E3.3. Central Tension and Compression 750
E3.3.1. Longitudinal Force 750
E3.3.2. Stresses and Strains under Tension or Compression 751
E3.3.3. Strength Analysis under Tension and Compression 752

E3.4. Torsion 753
E3.4.1. Twisting Moment 753
E3.4.2. Stresses and Strains in Torsion 753
E3.4.3. Strength Analysis in Torsion 755

E3.5. Symmetrical Bending 755
E3.5.1. Bending Moment and Transverse Force 755
E3.5.2. Stresses and Strains under Symmetric Pure Bending 758
E3.5.3. Stresses and Strains in Symmetric Lateral Bending 760
E3.5.4. Strength Analysis in the Case of Symmetric Bending 761

E3.6. Combined Stress 763
  E3.6.1. Oblique Bending 763
  E3.6.2. Eccentric Extension or Compression 765
  E3.6.3. Bending with Torsion 766

E3.7. Stability of Compressed Rods 768
  E3.7.1. Critical Force 768
  E3.7.2. Euler Formula 768
  E3.7.3. Influence of Methods for Fixation of the Rod Ends on the Value of the Critical Force 770
  E3.7.4. Scope of the Euler Formula 770
  E3.7.5. Stability Analysis of Compressed Rods 771

Bibliography for Chapter E3 771

E4. Hydrodynamics **773**

E4.1. Hydrostatics 773
  E4.1.1. Properties of Incompressible Fluid at Rest. Basic Law of Hydrostatics 773
  E4.1.2. Force of Pressure on a Plane and on a Curved Wall 774
  E4.1.3. Archimedes Principle. Stability of Floating Bodies 777

E4.2. Hydrodynamic Equations and Boundary Conditions 778
  E4.2.1. Laminar Flows. Navier–Stokes Equations. Euler Equations 779
  E4.2.2. Initial Conditions and the Simplest Boundary Conditions 781
  E4.2.3. Turbulent Flows. Reynolds Equations 782

E4.3. Hydrodynamics of Thin Films 784
  E4.3.1. Preliminary Remarks. Different Regimes of Film Type Flow 784
  E4.3.2. Films on an Inclined Plane and on a Cylindrical Surface 785

E4.4. Fluid Flows in Tubes 786
  E4.4.1. Laminar Flows in Tubes of Various Cross-Sections 786
  E4.4.2. Turbulent Flows in Tubes 791

E4.5. Hydrodynamic Boundary Layers on a Flat Plate 796
  E4.5.1. Laminar Boundary Layer 796
  E4.5.2. Turbulent Boundary Layer 798

E4.6. Spherical Particles and Circular Cylinders in Translational Flow 800
  E4.6.1. Stokes Equations and Their Solution for the Axisymmetric Case 800
  E4.6.2. Spherical Particles in Translational Stokes Flow (at $\mathsf{Re} \to 0$) 802
  E4.6.3. Spherical Particles in Translational Flow in a Wide Range of Re 803
  E4.6.4. Translational Flow Past a Cylinder (the Plane Problem) 804

E4.7. Spherical Drops and Bubbles in Translational Flow 806
  E4.7.1. Spherical Drops and Bubbles in Translational Stokes Flow (at $\mathsf{Re} \to 0$) 806
  E4.7.2. Spherical Drops and Bubbles in Translational Flow at Various Reynolds Numbers 808

E4.8. Flow Past Nonspherical Particles 809
  E4.8.1. Translational Stokes Flow Past Ellipsoidal Particles 809
  E4.8.2. Translational Stokes Flow Past Bodies of Revolution 811
  E4.8.3. Translational Stokes Flow Past Particles of Arbitrary Shape 812
  E4.8.4. Sedimentation of Isotropic and Nonisotropic Particles 814
  E4.8.5. Flow Past Nonspherical Particles at Various Reynolds Numbers 815

E4.9. Motion of Several Particles. Effective Viscosity of Suspensions 816
E4.9.1. Motion of Two Spheres. Gravitational Sedimentation of Several Spheres 816
E4.9.2. Rate of Suspension Precipitation. Effective Viscosity of Suspensions 818

Bibliography for Chapter E4 819

E5. Mass and Heat Transfer 821

E5.1. Convective Mass and Heat Transfer. Equations and Boundary Conditions ... 821
E5.1.1. Mass and Heat Transfer in Laminar Flows 821
E5.1.2. Basic Dimensionless Parameters. Diffusion Fluxes and the Sherwood Number 824
E5.1.3. Mass and Heat Transfer in Turbulent Flows 825

E5.2. Mass Transfer in Liquid Films 826
E5.2.1. Mass Exchange Between Gases and Liquid Films 826
E5.2.2. Dissolution of a Plate by a Laminar Liquid Film 829

E5.3. Heat Transfer to a Flat Plate 831
E5.3.1. Heat Transfer in Laminar Flow 831
E5.3.2. Heat Transfer in Turbulent Flow 832

E5.4. Convective Heat and Mass Transfer in Tubes and Channels 834
E5.4.1. Heat and Mass Transfer in a Circular Tube with Constant Temperature of the Wall for Laminar Flows 834
E5.4.2. Heat and Mass Transfer in a Circular Tube with Constant Heat Flux at the Wall for Laminar Flows 838
E5.4.3. Heat and Mass Transfer in Plane Channels for Laminar Flows 839
E5.4.4. Turbulent Heat Transfer in Circular Tube and Plane Channel 841
E5.4.5. Limit Nusselt Numbers for Tubes of Various Cross-Sections 843

E5.5. Interior Heat Problems for Solid Bodies of Various Shapes 845
E5.5.1. Statement of the Problem 845
E5.5.2. Bulk Temperature for Bodies of Various Shapes 845

E5.6. Mass and Heat Exchange Between Particles of Various Shapes and a Quiescent Medium 847
E5.6.1. Stationary Mass and Heat Exchange 847
E5.6.2. Transient Mass and Heat Exchange 849

E5.7. Mass and Heat Transfer to a Spherical Particle in a Translational Flow 850
E5.7.1. Statement of the Problem 850
E5.7.2. Mass Transfer at Low Reynolds Numbers and Various Peclet Numbers 851
E5.7.3. Mass Transfer at Moderate and High Reynolds Numbers 852

E5.8. Mass and Heat Transfer to Spherical Drops and Bubbles in Translational Flow 853
E5.8.1. Statement of the Problem. Some Remarks 853
E5.8.2. Mass Transfer at Low Reynolds Numbers and Various Peclet Numbers 853
E5.8.3. Mass Transfer at Moderate and High Reynolds Numbers 855
E5.8.4. General Correlations for the Sherwood Number 855

E5.9. Mass Transfer to Nonspherical Particles in Translational Flow 856
E5.9.1. Mass and Heat Transfer to a Particle of Arbitrary Shape at Low Peclet Numbers 856
E5.9.2. Mass Transfer to a Ellipsoidal Particle in Translational Flow 857
E5.9.3. Mass Transfer to Particles of Arbitrary Shape 858
E5.9.4. Mass and Heat Transfer to a Circular Cylinder in Translational Flow 859

E5.10. Transient Mass Transfer to Particles, Drops, and Bubbles in Steady-State Flows 860
E5.10.1. Statement of the Problem 860
E5.10.2. General Correlations for the Sherwood Number 861

Bibliography for Chapter E5 862

E6. Electrical Engineering 865

E6.1. Generalities 865
E6.1.1. Introduction. Basic Notions, Definitions, and Notation 865
E6.1.2. Perfect Elements. Active and Passive Elements 866
E6.1.3. Electric Circuits and Networks. Classification of Electric Circuits and Their Operation Modes 867
E6.1.4. Structure Elements of Network. Notion of the Graph of a Network . . 868
E6.1.5. Physical Laws for Structure Elements 868
E6.1.6. Problems of Analysis and Synthesis of a Network. An Analysis of a Network by Physical Laws. Potential Diagram 869

E6.2. Linear Circuits of Direct Current 871
E6.2.1. Power of Direct Current 871
E6.2.2. Notion of Nonideal Source. Equivalent Circuits and External Characteristics 872
E6.2.3. Problem of Maximal Power Transfer. The Source Efficiency Factor . 874
E6.2.4. Theorem on Active Impedor (Two-Terminal Network) and Its Application 875
E6.2.5. Two Compensation Theorems. Properties of Linear Circuits: Linearity Theorem and Reciprocity Theorem 876
E6.2.6. Equivalent Transformations in Electrical Networks 877
E6.2.7. Application of Equivalent Transformations in Network Analysis. Convolution Method 878
E6.2.8. Superposition Principle. The Notion of Input and Transfer Conductances (Admittances) 880
E6.2.9. Concise Methods for Analyzing Electric Circuits 881

E6.3. Nonlinear Electric Circuits of Direct Current 884
E6.3.1. Notion of a Nonlinear Element. Several Typical Characteristics of Nonlinear Elements 884
E6.3.2. Equivalent Transformations of Characteristics of Nonlinear Elements 886
E6.3.3. Graphical Method for Analyzing Circuits with Nonlinear Elements . 887

E6.4. Magnetic Circuits 890
E6.4.1. Magnetic Field and Its Parameters 890
E6.4.2. Ampère's Circuital Law 891
E6.4.3. Ohm's Law for Magnetic Circuit 892
E6.4.4. Ferromagnetic Materials and Their Properties 893
E6.4.5. Kirchhoff's Laws for Magnetic Circuits 894
E6.4.6. Analysis of an Unbranched Magnetic Circuit 896
E6.4.7. Analysis of a Branched Magnetic Circuit 898
E6.4.8. Analysis of a Branched Homogeneous Magnetic Circuit with Permanent Magnet 899

E6.5. Single-Phase Harmonic Current Circuits 900
E6.5.1. Effective and Average Values of Sinusoidal Quantities 900
E6.5.2. Representation of Sinusoidal Quantities by Vectors on Complex Plane 901
E6.5.3. Active Resistance in a Sinusoidal Current Circuit 904

E6.5.4. Inductance in a Sinusoidal Current Circuit 905
E6.5.5. Capacitance in a Sinusoidal Current Circuit 906
E6.5.6. Unbranched Electric Circuit 907
E6.5.7. Branched Electric Circuit 910
E6.5.8. Topographic Diagram 911
E6.5.9. Resonance Phenomena 912
E6.5.10. Improvement of the Power Factor 917

Bibliography for Chapter E6 920

E7. Empirical and Engineering Formulas and Criteria for Their Applicability 921

E7.1. Empirical, Engineering and Interpolation Formulas. Least-Squares Method . 921
E7.1.1. Empirical and Engineering Formulas 921
E7.1.2. Least-Squares Method 922
E7.1.3. Refinement of Empirical Formulas 924
E7.1.4. Interpolation Formulas 925

E7.2. Criteria for Applicability of Empirical Formulas 926
E7.2.1. Direct Method for Verifying Two-Parameter Formulas 926
E7.2.2. Three-Parameter Empirical Formulas 931
E7.2.3. Combined Method for Verifying Empirical Formulas 932

E7.3. Construction of Engineering Formulas by the Method of Asymptotic Analogies 933
E7.3.1. Description of the Method of Asymptotic Analogies 933
E7.3.2. Illustrating Examples: Interior Heat Exchange Problems for Bodies of Various Shape 936
E7.3.3. Illustrating Examples: Mass Exchange Between Bubbles/Particles and Flow 939

Bibliography for Chapter E7 942

Part IV. Supplements 943

S1. Integrals **945**

S1.1. Indefinite Integrals 945
S1.1.1. Integrals Involving Rational Functions 945
S1.1.2. Integrals Involving Irrational Functions 949
S1.1.3. Integrals Involving Exponential Functions 952
S1.1.4. Integrals Involving Hyperbolic Functions 953
S1.1.5. Integrals Involving Logarithmic Functions 956
S1.1.6. Integrals Involving Trigonometric Functions 957
S1.1.7. Integrals Involving Inverse Trigonometric Functions 962

S1.2. Tables of Definite Integrals 963
S1.2.1. Integrals Involving Power-Law Functions 963
S1.2.2. Integrals Involving Exponential Functions 966
S1.2.3. Integrals Involving Hyperbolic Functions 967
S1.2.4. Integrals Involving Logarithmic Functions 967
S1.2.5. Integrals Involving Trigonometric Functions 968

S1.3. General Reduction Formulas for the Evaluation of Definite Integrals 970
S1.3.1. Integrals Involving Functions of a Linear or Rational Argument 970
S1.3.2. Integrals Involving Functions with Trigonometric Argument 971
S1.3.3. Integrals Involving Logarithmic Functions 971
S1.3.4. General Reduction Formulas for the Calculation of Improper Integrals 972

Bibliography for Chapter S1 973

S2. Integral Transforms **975**

S2.1. Tables of Laplace Transforms 975
  S2.1.1. General Formulas 975
  S2.1.2. Expressions Involving Power Functions 977
  S2.1.3. Expressions Involving Exponential Functions 977
  S2.1.4. Expressions Involving Hyperbolic Functions 978
  S2.1.5. Expressions Involving Logarithmic Functions 979
  S2.1.6. Expressions Involving Trigonometric Functions 979
  S2.1.7. Expressions Involving Special Functions 981

S2.2. Tables of Inverse Laplace Transforms 982
  S2.2.1. General Formulas 982
  S2.2.2. Expressions Involving Rational Functions 984
  S2.2.3. Expressions Involving Square Roots 988
  S2.2.4. Expressions Involving Arbitrary Powers 990
  S2.2.5. Expressions Involving Exponential Functions 990
  S2.2.6. Expressions Involving Hyperbolic Functions 992
  S2.2.7. Expressions Involving Logarithmic Functions 992
  S2.2.8. Expressions Involving Trigonometric Functions 993
  S2.2.9. Expressions Involving Special Functions 994

S2.3. Tables of Fourier Cosine Transforms 995
  S2.3.1. General Formulas 995
  S2.3.2. Expressions Involving Power Functions 995
  S2.3.3. Expressions Involving Exponential Functions 996
  S2.3.4. Expressions Involving Hyperbolic Functions 997
  S2.3.5. Expressions Involving Logarithmic Functions 997
  S2.3.6. Expressions Involving Trigonometric Functions 998
  S2.3.7. Expressions Involving Special Functions 999

S2.4. Tables of Fourier Sine Transforms 1000
  S2.4.1. General Formulas 1000
  S2.4.2. Expressions Involving Power Functions 1000
  S2.4.3. Expressions Involving Exponential Functions 1001
  S2.4.4. Expressions Involving Hyperbolic Functions 1002
  S2.4.5. Expressions Involving Logarithmic Functions 1002
  S2.4.6. Expressions Involving Trigonometric Functions 1003
  S2.4.7. Expressions Involving Special Functions 1004

Bibliography for Chapter S2 1004

S3. Orthogonal Curvilinear Systems of Coordinates **1005**

S3.1. Arbitrary Curvilinear Coordinate Systems 1005
  S3.1.1. General Nonorthogonal Curvilinear Coordinates 1005
  S3.1.2. General Orthogonal Curvilinear Coordinates 1006

S3.2. Cylindrical and Spherical Coordinate Systems 1008
  S3.2.1. Cylindrical Coordinates 1008
  S3.2.2. Spherical Coordinates 1009

Bibliography for Chapter S3 1010

S4. Ordinary Differential Equations **1011**

S4.1. First-Order Equations 1011
  S4.1.1. Simplest Equations Integrable in Closed Form 1011
  S4.1.2. Riccati Equation 1012
  S4.1.3. Other Equations 1014

S4.2. Second-Order Linear Equations 1016
  S4.2.1. Equations Involving Power Functions 1017
  S4.2.2. Equations Involving Exponential and Other Functions 1024
  S4.2.3. Equations Involving Arbitrary Functions 1027

S4.3. Second-Order Nonlinear Equations 1028
  S4.3.1. Equations of the Form $y''_{xx} = f(x,y)$ 1028
  S4.3.2. Equations of the Form $f(x,y)y''_{xx} = g(x,y,y'_x)$ 1030

Bibliography for Chapter S4 1034

S5. Some Useful Electronic Mathematical Resources **1035**

S6. Physical Tables **1037**

S6.1. Symbols and Units 1037
  S6.1.1. Greek Letters 1037
  S6.1.2. SI Prefixes 1037
  S6.1.3. SI Base Units 1038
  S6.1.4. SI Derived Units with Special Names 1038
  S6.1.5. Non-SI Units Accepted for Use with SI 1039
  S6.1.6. Other Common Non-SI Units 1039
  S6.1.7. Base Planck Units 1040

S6.2. Physical Constants 1040
  S6.2.1. Universal Constants 1040
  S6.2.2. Astronomical Constants 1040
  S6.2.3. Planetary, Lunar, and Solar Parameters 1041
  S6.2.4. Electromagnetic Constants 1041
  S6.2.5. Atomic and Nuclear Constants 1041
  S6.2.6. Other Constants 1042

S6.3. Nuclear Physics 1043
  S6.3.1. Masses of Light Nuclides 1043
  S6.3.2. Leptons 1043
  S6.3.3. Quarks 1044
  S6.3.4. Elementary Bosons 1044
  S6.3.5. Some Baryons 1044
  S6.3.6. Some Mesons 1045

Bibliography for Chapter S6 1046

S7. Periodic Table **1047**

Bibliography for Chapter S7 1049

Index **1051**

PREFACE

This is a concise multi-subject handbook, which consists of three major parts: mathematics, physics, and applied and engineering sciences. It presents basic notions, formulas, equations, problems, theorems, methods, and laws on each of the subjects in brief form. The absence of proofs and a concise presentation has permitted combining a substantial amount of reference material in a single volume. The handbook is intended for a wide audience of engineers and researchers (not specialized in mathematics or theoretical physics) as well as graduate and postgraduate students.

- The first part of the book contains chapters on arithmetics, elementary and analytic geometry, algebra, differential and integral calculus, functions of complex variable, integral transforms, ordinary and partial differential equations, special functions, probability theory, etc.
- The second part of the book contains chapters on molecular physics and thermodynamics, electricity and magnetism, oscillations and waves, optics, special relativity, quantum mechanics, atomic physics, etc.
- The third part of the book contains chapters on dimensional analysis and similarity, mechanics of point masses and rigid bodies, strength of materials, hydrodynamics, mass and heat transfer, electrical engineering, and methods for constructing empirical and engineering formulas.

A compact and clear presentation of the material allows the reader to get quick help on (or revise) the desired topic. Special attention is paid to issues that many engineers and students may find difficult to understand.

When selecting the material, the authors have given a pronounced preference to practical aspects; namely, to formulas, problems, methods, and laws that most frequently occur in sciences and engineering applications and university education. Many results are represented in tabular form.

For the convenience of a wider audience with different mathematical backgrounds, the authors tried to avoid special terminology whenever possible. Therefore, some of the topics and methods are outlined in a schematic and somewhat simplified manner, which is sufficient for them to be used successfully in most cases. Many sections were written so that they could be read independently. The material within subsections is arranged in increasing order of complexity. This allows the reader to get to the heart of the matter quickly.

The material of the reference book can be roughly categorized into the following three groups according to meaning:

1. The main text containing a concise, coherent survey of the most important definitions, formulas, equations, methods, theorems, and laws.
2. For the reader's better understanding of the topics and methods under study, numerous examples are given throughout the book.
3. Discussion of additional issues of interest, given in the form of remarks in small print.

For the reader's convenience, several long mathematical tables—indefinite and definite integrals, direct and inverse integral transforms (Laplace, Mellin, and Fourier transforms), and exact solutions of differential equations—which contain a large amount of information, are presented in the supplement of the book. Also included are some physical tables and the periodic table of the chemical elements.

This handbook consists of parts, chapters, sections, and subsections. Figures and tables are numbered separately in each section, while formulas (equations) and examples are numbered separately in each subsection. When citing a formula, we use notation like (M3.1.2.5), which means formula 5 in Subsection M3.1.2. For the reader's convenience, each citation number is preceded by a letter to indicate one of the major parts: mathematics (M), physics (P), engineering sciences (E), or supplements (S). At the end of each chapter, we present a list of main and additional literature sources containing more detailed information about topics of interest to the reader.

Special font highlighting in the text, cross-references, an extensive table of contents, and a detailed index help the reader to find the desired information.

Chapters M1, M2, and M6–M9 were written by V. M. Safrai and A. I. Zhurov, Chapters M3–M5, M10, and M14 by A. V. Manzhirov and V. A. Popov, Chapters M11–M13, E1, E4, E5, E7, and S1–S5 by A. D. Polyanin, Chapters P1–P8 by A. I. Chernoutsan, Chapter E2 by V. D. Polyanin, Chapter E3 by B. V. Putyatin, Chapter E6 by A. V. Egorov and Yu. V. Repina, and Chapters S6 and S7 by A. I. Chernoutsan and A. I. Zhurov. Part M was edited by A. D. Polyanin and parts E and S were edited by A. D. Polyanin and A. I. Chernoutsan.

We would like to express our deep gratitude to Vladimir Nazaikinskii for translating several chapters of this handbook.

The authors hope that this book will be helpful for a wide range of engineers, scientists, university teachers, and students engaged in the fields of physics, mechanics, engineering sciences, chemistry, biology, ecology, medicine as well as social and economical sciences.

Andrei D. Polyanin
Alexei I. Chernoutsan

EDITORS

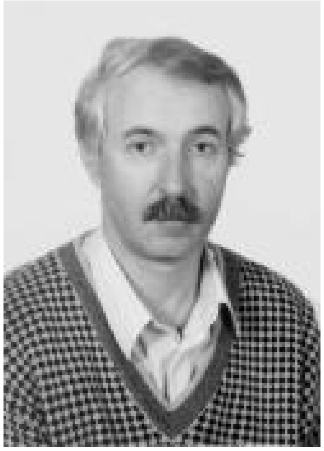

Andrei D. Polyanin, D.Sc., Ph.D., is a well-known scientist of broad interests and is active in various areas of mathematics, mechanics, and chemical engineering sciences. He is one of the most prominent authors in the field of reference literature on mathematics.

Professor Polyanin graduated with honors from the Department of Mechanics and Mathematics of Moscow State University in 1974. He received his Ph.D. degree in 1981 and D.Sc. degree in 1986 at the Institute for Problems in Mechanics of the Russian (former USSR) Academy of Sciences. Since 1975, Professor Polyanin has been working at the Institute for Problems in Mechanics of the Russian Academy of Sciences; he is also Professor of Mathematics at Bauman Moscow State Technical University. He is a member of the Russian National Committee on Theoretical and Applied Mechanics and of the Mathematics and Mechanics Expert Council of the Higher Certification Committee of the Russian Federation.

Professor Polyanin is an author of more than 30 books in English, Russian, German, and Bulgarian as well as over 140 research papers and three patents. He has written a number of fundamental handbooks, including A. D. Polyanin and V. F. Zaitsev, *Handbook of Exact Solutions for Ordinary Differential Equations*, CRC Press, 1995 and 2003; A. D. Polyanin and A. V. Manzhirov, *Handbook of Integral Equations*, CRC Press, 1998 and 2008; A. D. Polyanin, *Handbook of Linear Partial Differential Equations for Engineers and Scientists*, Chapman & Hall/CRC Press, 2002; A. D. Polyanin, V. F. Zaitsev, and A. Moussiaux, *Handbook of First Order Partial Differential Equations*, Taylor & Francis, 2002; A. D. Polyanin and V. F. Zaitsev, *Handbook of Nonlinear Partial Differential Equations*, Chapman & Hall/CRC Press, 2004, and A. D. Polyanin and A. V. Manzhirov, *Handbook of Mathematics for Engineers and Scientists*, Chapman & Hall/CRC Press, 2007.

Professor Polyanin is Editor-in-Chief of the international scientific-educational Website *EqWorld — The World of Mathematical Equations* (http://eqworld.ipmnet.ru), which is visited by several thousands of users a day worldwide. He is also Editor of the book series *Differential and Integral Equations and Their Applications*, Chapman & Hall/CRC Press, London/Boca Raton. Professor Polyanin is a member of the Editorial Board of the journal *Theoretical Foundations of Chemical Engineering.*

In 1991, Professor Polyanin was awarded a Chaplygin Prize of the Russian Academy of Sciences for his research in mechanics. In 2001, he received an award from the Ministry of Education of the Russian Federation.

Address: Institute for Problems in Mechanics, 101 Vernadsky Ave., Bldg. 1, 119526 Moscow, Russia.

Home page: http://eqworld.ipmnet.ru/polyanin-ew.htm

Alexei I. Chernoutsan, Ph.D., is a prominent scientist in the fields of statistical physics, fluctuation kinetics, phase transitions, and percolation theory. He graduated from the Faculty of Physics of the Moscow State University in 1975. He received his Ph.D. degree in 1980 at the Landau Institute for Theoretical Physics of the Russian (former USSR) Academy of Sciences. Professor Chernoutsan has been a member of staff of the Gubkin Russian State University of Oil and Gas since 1980, Dean of the Faculty of Natural Sciences Education since 2002, and Head of the Department of Physics since 2009.

An experienced and multi-skilled lecturer in general physics, Professor Chernoutsan has taught students of the Gubkin Russian State University of Oil and Gas for many years. He is an active author in the field of popular science and since 1991 he has been Deputy Editor-in-Chief of *Kvant*, a famous Russian educational magazine for students. He is an author of 3 books and more than 80 scientific and educational publications.

Address: Gubkin Russian State University of Oil and Gas, 65 Leninskii Ave., 119991 Moscow, Russia.
E-mail: acher@gubkin.ru

Part I

Mathematics

Chapter M1

Arithmetic and Elementary Algebra

M1.1. Real Numbers

M1.1.1. Integer Numbers

► **Natural, integer, even, and odd numbers.** *Natural numbers*: 1, 2, 3, ... (all positive whole numbers).

Integer numbers (or simply *integers*): 0, ±1, ±2, ±3, ...

Even numbers: 0, 2, 4, ... (all nonnegative integers that can be divided evenly by 2). An even number can generally be represented as $n = 2k$, where $k = 0, 1, 2, \ldots$

Remark 1. Sometimes all integers that are multiples of 2, such as 0, ±2, ±4, ..., are considered to be even numbers.

Odd numbers: 1, 3, 5, ... (all natural numbers that cannot be divided evenly by 2). An odd number can generally be represented as $n = 2k + 1$, where $k = 0, 1, 2, \ldots$

Remark 2. Sometimes all integers that are not multiples of 2, such as ±1, ±3, ±5, ..., are considered to be odd numbers.

All integers as well as even numbers and odd numbers form *infinite countable sets*, which means that the elements of these sets can be enumerated using the natural numbers 1, 2, 3, ...

► **Prime and composite numbers.** A *prime number* is a positive integer that is greater than 1 and has no positive integer divisors other than 1 and itself. The prime numbers form an infinite countable set. The first ten prime numbers are: 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, ...

A *composite number* is a positive integer that is greater than 1 and is not prime, i.e., has factors other than 1 and itself. Any composite number can be uniquely factored into a product of prime numbers. The following numbers are composite: $4 = 2 \times 2$, $6 = 2 \times 3$, $8 = 2^3$, $9 = 3^2$, $10 = 2 \times 5$, $12 = 2^2 \times 3$, ...

The number 1 is a special case that is considered to be neither composite nor prime.

► **Divisibility tests.** Below are some simple rules helping to determine if an integer is divisible by another integer.

All integers are divisible by 1.

Divisibility by 2: last digit is divisible by 2.

Divisibility by 3: sum of digits is divisible by 3.

Divisibility by 4: two last digits form a number divisible by 4.

Divisibility by 5: last digit is either 0 or 5.

Divisibility by 6: divisible by both 2 and 3.

Divisibility by 9: sum of digits is divisible by 9.

Divisibility by 10: last digit is 0.

Divisibility by 11: the difference between the sum of the odd-numbered digits (1st, 3rd, 5th, etc.) and the sum of the even-numbered digits (2nd, 4th, etc.) is divisible by 11.

Example 1. Let us show that the number 80729 is divisible by 11.

The sum of the odd-numbered digits is $\Sigma_1 = 8 + 7 + 9 = 24$. The sum of the even-numbered digits is $\Sigma_2 = 0 + 2 = 2$. The difference between them is $\Sigma_1 - \Sigma_2 = 22$ and is divisible by 11. Consequently, the original number is also divisible by 11.

► **Greatest common divisor and least common multiple.**

1°. The *greatest common divisor* of natural numbers $a_1, a_2, \ldots, a_n$ is the largest natural number, b, which is a common divisor to $a_1, \ldots, a_n$.

Suppose some positive numbers $a_1, a_2, \ldots, a_n$ are factored into products of primes so that

$$a_1 = p_1^{k_{11}} p_2^{k_{12}} \ldots p_m^{k_{1m}}, \quad a_2 = p_1^{k_{21}} p_2^{k_{22}} \ldots p_m^{k_{2m}}, \quad \ldots, \quad a_n = p_1^{k_{n1}} p_2^{k_{n2}} \ldots p_m^{k_{nm}},$$

where $p_1, p_2, \ldots, p_m$ are different prime numbers and the k_{ij} are nonnegative integers ($i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m$). Then the greatest common divisor b of $a_1, a_2, \ldots, a_n$ is calculated as

$$b = p_1^{\sigma_1} p_2^{\sigma_2} \ldots p_m^{\sigma_m}, \qquad \sigma_j = \min_{1 \le i \le n} k_{ij}.$$

Example 2. The greatest common divisor of 180 and 280 is $2^2 \times 5 = 20$ due to the following factorization:

$$180 = 2^2 \times 3^2 \times 5 = 2^2 \times 3^2 \times 5^1 \times 7^0,$$
$$280 = 2^3 \times 5 \times 7 = 2^3 \times 3^0 \times 5^1 \times 7^1.$$

2°. The *least common multiple* of n natural numbers $a_1, a_2, \ldots, a_n$ is the smallest natural number, A, that is a multiple of all the a_k.

Suppose some natural numbers $a_1, \ldots, a_n$ are factored into products of primes just as in Item 1°. Then the least common multiple of all the a_k is calculated as

$$A = p_1^{\nu_1} p_2^{\nu_2} \ldots p_m^{\nu_m}, \qquad \nu_j = \max_{1 \le i \le n} k_{ij}.$$

Example 3. The least common multiple of 180 and 280 is equal to $2^3 \times 3^2 \times 5^1 \times 7^1 = 2520$ due to the factorization given in Example 2.

M1.1.2. Real, Rational, and Irrational Numbers

► **Real numbers.** The *real numbers* are all the positive numbers, negative numbers, and zero. Any real number can be represented by a *decimal fraction* (or simply *decimal*), finite or infinite. The set of all real numbers is denoted by $\mathbb{R}$.

All real numbers are categorized into two classes: the *rational* numbers and *irrational* numbers.

► **Rational numbers.** A *rational number* is a real number that can be written as a fraction (ratio) p/q with integer p and q ($q \neq 0$). It is only the rational numbers that can be written in the form of finite (terminating) or periodic (recurring) decimals (e.g., $1/8 = 0.125$ and $1/6 = 0.16666\ldots$). Any integer is a rational number.

The rational numbers form an infinite countable set. The set of all rational numbers is everywhere dense. This means that, for any two distinct rational numbers a and b such that $a < b$, there exists at least one more rational number c such that $a < c < b$, and hence there are infinitely many rational numbers between a and b. (Between any two rational numbers, there always exist irrational numbers.)

► **Irrational numbers.** An *irrational number* is a real number that is not rational; no irrational number can be written as a fraction p/q with integer p and q ($q \neq 0$). To the irrational numbers there correspond nonperiodic (nonrepeating) decimals. Here are examples of irrational numbers: $\sqrt{3} = 1.73205\ldots$, $\pi = 3.14159\ldots$

The set of irrational numbers is everywhere dense, which means that between any two distinct irrational numbers, there are both rational and irrational numbers. The set of irrational numbers is uncountable.

M1.2. Equalities and Inequalities. Arithmetic Operations. Absolute Value

M1.2.1. Equalities and Inequalities

Throughout Subsection 1.2.1, it is assumed that a, b, c, d are real numbers.

► **Basic properties of equalities.**

1. If $a = b$, then $b = a$.
2. If $a = b$, then $a + c = b + c$, where c is any real number; furthermore, if $a + c = b + c$, then $a = b$.
3. If $a = b$, then $ac = bc$, where c is any real number; furthermore, if $ac = bc$ and $c \neq 0$, then $a = b$.
4. If $a = b$ and $b = c$, then $a = c$.
5. If $ab = 0$, then either $a = 0$ or $b = 0$; furthermore, if $ab \neq 0$, then $a \neq 0$ and $b \neq 0$.

► **Basic properties of inequalities.**

1. If $a < b$, then $b > a$.
2. If $a \leq b$ and $b \leq a$, then $a = b$.
3. If $a \leq b$ and $b \leq c$, then $a \leq c$.
4. If $a < b$ and $b \leq c$ (or $a \leq b$ and $b < c$), then $a < c$.
5. If $a < b$ and $c < d$ (or $c = d$), then $a + c < b + d$.
6. If $a \leq b$ and $c > 0$, then $ac \leq bc$.
7. If $a \leq b$ and $c < 0$, then $ac \geq bc$.
8. If $0 < a \leq b$ (or $a \leq b < 0$), then $1/a \geq 1/b$.

M1.2.2. Addition and Multiplication of Numbers

► **Addition of real numbers.** The sum of real numbers is a real number.
Properties of addition:

$$\begin{aligned} &a + 0 = a && \text{(property of zero)},\\ &a + b = b + a && \text{(addition is commutative)},\\ &a + (b + c) = (a + b) + c = a + b + c && \text{(addition is associative)}, \end{aligned}$$

where a, b, c are arbitrary real numbers.

For any real number a, there exists its unique *additive inverse*, or its *opposite*, denoted by $-a$, such that

$$a + (-a) = a - a = 0.$$

► **Multiplication of real numbers.** The product of real numbers is a real number.
Properties of multiplication:

$$\begin{aligned} &a \times 0 = 0 && \text{(property of zero)},\\ &ab = ba && \text{(multiplication is commutative)},\\ &a(bc) = (ab)c = abc && \text{(multiplication is associative)},\\ &a \times 1 = 1 \times a = a && \text{(multiplication by unity)},\\ &a(b + c) = ab + ac && \text{(multiplication is distributive)}, \end{aligned}$$

where a, b, c are arbitrary real numbers.

For any nonzero real number a, there exists its unique *multiplicative inverse*, or its *reciprocal*, denoted by a^{-1} or $1/a$, such that

$$aa^{-1} = 1 \quad (a \neq 0).$$

M1.2.3. Ratios and Proportions

► **Operations with fractions and properties of fractions.** Ratios are written as fractions: $a : b = a/b$. The number a is called the *numerator* and the number b ($b \neq 0$) is called the *denominator* of a fraction.

Properties of fractions and operations with fractions:

$$\frac{a}{1} = a, \qquad \frac{a}{b} = \frac{ab}{bc} = \frac{a:c}{b:c} \qquad \text{(simplest properties of fractions);}$$

$$\frac{a}{b} \pm \frac{c}{b} = \frac{a \pm c}{b}, \qquad \frac{a}{b} \pm \frac{c}{d} = \frac{ad \pm bc}{bd} \qquad \text{(addition and subtraction of fractions);}$$

$$\frac{a}{b} \times c = \frac{ac}{b}, \qquad \frac{a}{b} \times \frac{c}{d} = \frac{ac}{bc} \qquad \text{(multiplication by a number and by a fraction);}$$

$$\frac{a}{b} : c = \frac{a}{bc}, \qquad \frac{a}{b} : \frac{c}{d} = \frac{ad}{bc} \qquad \text{(division by a number and by a fraction).}$$

► **Proportions. Simplest relations. Derivative proportions.** A proportion is an equation with a ratio on each side. A proportion is denoted by $a/b = c/d$ or $a : b = c : d$.

1°. The following simplest relations follow from $a/b = c/d$:

$$ad = bc, \quad \frac{a}{c} = \frac{b}{d}, \quad a = \frac{bc}{d}, \quad b = \frac{ad}{c}.$$

2°. The following derivative proportions follow from $a/b = c/d$:

$$\frac{ma + nb}{pa + qb} = \frac{mc + nd}{pc + qd},$$
$$\frac{ma + nc}{pa + qc} = \frac{mb + nd}{pb + qd},$$

where m, n, p, q are arbitrary real numbers.

Some special cases of the above formulas:

$$\frac{a \pm b}{b} = \frac{c \pm d}{d}, \qquad \frac{a - b}{a + b} = \frac{c - d}{c + d}.$$

M1.2.4. Percentage

► **Definition. Main percentage problems.** A *percentage* is a way of expressing a ratio or a fraction as a whole number, by using 100 as the denominator. One *percent* is one per one hundred, or one hundredth of a whole number; notation: 1%.

Below are the statements of main percentage problems and their solutions.

1°. Find the number b that makes up $p\%$ of a number a. Answer: $b = \frac{ap}{100}$.

2°. Find the number a whose $p\%$ is equal to a number b. Answer: $a = \frac{100\,b}{p}$.

3°. What percentage does a number b make up of a number a? Answer: $p = \frac{100\,b}{a}\%$.

► **Simple and compound percentage.**

1°. *Simple percentage*. Suppose a cash deposit is increased yearly by the same amount defined as a percentage, $p\%$, of the initial deposit, a. Then the amount accumulated after t years is calculated by the simple percentage formula

$$x = a\left(1 + \frac{pt}{100}\right).$$

2°. *Compound percentage*. Suppose a cash deposit is increased yearly by an amount defined as a percentage, $p\%$, of the deposit in the previous year. If a is the initial deposit, then the amount accumulated after t years is calculated by the compound percentage formula

$$x = a\left(1 + \frac{p}{100}\right)^t.$$

M1.2.5. Absolute Value of a Number (Modulus of a Number)

► **Definition.** The absolute value of a real number a, denoted by $|a|$, is defined by the formula

$$|a| = \begin{cases} a & \text{if } a \ge 0, \\ -a & \text{if } a < 0. \end{cases}$$

An important property: $|a| \ge 0$.

► **Some formulas and inequalities.**

1°. The following relations hold true:

$$\begin{aligned} &|a| = |-a| = \sqrt{a^2}, \quad a \le |a|, \\ &\big||a| - |b|\big| \le |a+b| \le |a| + |b|, \\ &\big||a| - |b|\big| \le |a-b| \le |a| + |b|, \\ &|ab| = |a|\,|b|, \quad |a/b| = |a|/|b|. \end{aligned}$$

2°. From the inequalities $|a| \le A$ and $|b| \le B$ it follows that $|a+b| \le A+B$ and $|ab| \le AB$.

M1.3. Powers and Logarithms

M1.3.1. Powers and Roots

► **Powers and roots: the main definitions.** Given a positive real number a and a positive integer n, the nth power of a, written as a^n, is defined as the multiplication of a by itself repeated n times:

$$a^n = \underbrace{a \times a \times a \times \cdots \times a}_{n \text{ multipliers}}.$$

The number a is called the *base* and n is called the *exponent*.

Obvious properties: $0^n = 0$, $1^n = 1$, $a^1 = a$.

Raising to the zeroth power: $a^0 = 1$, where $a \neq 0$.

Raising to a negative power: $a^{-n} = \dfrac{1}{a^n}$, where n is a positive integer.

If a is a positive real number and n is a positive integer, then the *nth arithmetic root* or *radical* of a, written as $\sqrt[n]{a}$, is the unique positive real number b such that $b^n = a$. In the case of $n = 2$, the brief notation $\sqrt{a}$ is used to denote $\sqrt[2]{a}$.

The following relations hold:

$$\sqrt[n]{0} = 0, \qquad \sqrt[n]{1} = 1, \qquad \left(\sqrt[n]{a}\right)^n = a.$$

Raising to a fractional power $p = m/n$, where m and n are natural numbers:

$$a^p = a^{m/n} = \sqrt[n]{a^m}, \qquad a \ge 0.$$

► **Operations with powers and roots.** The properties given below are valid for any real exponents p and q ($a > 0$, $b > 0$):

$$a^{-p} = \frac{1}{a^p}, \quad a^p a^q = a^{p+q}, \quad \frac{a^p}{a^q} = a^{p-q},$$

$$(ab)^p = a^p b^p, \quad \left(\frac{a}{b}\right)^q = \frac{a^q}{b^q}, \quad (a^p)^q = a^{pq}.$$

In operations with roots (radicals) the following properties are used:

$$\sqrt[n]{ab} = \sqrt[n]{a}\,\sqrt[n]{b}, \quad \sqrt[n]{\frac{a}{b}} = \frac{\sqrt[n]{a}}{\sqrt[n]{b}}, \quad \sqrt[n]{a^m} = \left(\sqrt[n]{a}\right)^m, \quad \sqrt[n]{\sqrt[m]{a}} = \sqrt[mn]{a}.$$

Remark. It often pays to represent roots as powers with rational exponents and apply the properties of operations with powers.

M1.3.2. Logarithms

► **Definition. The main logarithmic identity.** The *logarithm of a positive number b to a given base a* is the exponent of the power c to which the base a must be raised to produce b. It is written as $\log_a b = c$.

Equivalent representations:

$$\log_a b = c \quad \Longleftrightarrow \quad a^c = b,$$

where $a > 0$, $a \ne 1$, and $b > 0$.

Main logarithmic identity:

$$a^{\log_a b} = b.$$

Simple properties:

$$\log_a 1 = 0, \qquad \log_a a = 1.$$

► **Properties of logarithms. The common and natural logarithms.** Properties of logarithms:

$$\log_a(bc) = \log_a b + \log_a c, \qquad \log_a\left(\frac{b}{c}\right) = \log_a b - \log_a c,$$

$$\log_a(b^k) = k \log_a b, \qquad \log_{a^k} b = \frac{1}{k}\log_a b \quad (k \ne 0),$$

$$\log_a b = \frac{1}{\log_b a} \quad (b \ne 1), \qquad \log_a b = \frac{\log_c b}{\log_c a} \quad (c \ne 1),$$

where $a > 0$, $a \ne 1$, $b > 0$, $c > 0$, and k is any number.

The logarithm to the base 10 is called the *common* or *decadic logarithm* and written as

$$\log_{10} b = \log b \quad \text{or sometimes} \quad \log_{10} b = \lg b.$$

The logarithm to the base e (the *base of natural logarithms*) is called the *natural logarithm* and written as

$$\log_e b = \ln b,$$

where $e = \lim\limits_{n\to\infty}\left(1 + \frac{1}{n}\right)^n = 2.718281\ldots$

The following relations hold:

$$\ln b \approx 2.30259 \lg b, \quad \lg b \approx 0.43429 \ln b$$

M1.4. Binomial Theorem and Related Formulas

M1.4.1. Factorials. Binomial Coefficients. Binomial Theorem

► **Factorials. Binomial coefficients.**
Factorial:

$$0! = 1! = 1,$$
$$n! = 1\times 2\times 3\times\cdots\times(n-1)\times n, \quad n = 2,\ 3,\ 4,\ \ldots$$

Double factorial:

$$\begin{aligned}0!! &= 1!! = 1,\\ n!! &= \begin{cases}(2k)!! & \text{if } n=2k,\\ (2k+1)!! & \text{if } n=2k+1,\end{cases}\\ (2k)!! &= 2\times 4\times 6\times\cdots\times(2k-2)\times(2k) = 2^k k!,\\ (2k+1)!! &= 1\times 3\times 5\times\cdots\times(2k-1)\times(2k+1),\end{aligned}$$

where n and k are natural numbers.
Binomial coefficients:

$$C_n^k = \binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{n(n-1)\ldots(n-k+1)}{k!}, \quad k = 1,\ 2,\ 3,\ \ldots,\ n;$$
$$C_a^k = \frac{a(a-1)\ldots(a-k+1)}{k!}, \quad \text{where} \quad k = 1,\ 2,\ 3,\ \ldots,$$

where n is a natural number and a is any number.

► **Binomial theorem.** Let a, b, and c be real (or complex) numbers. The following formulas hold true:

$$\begin{aligned}(a\pm b)^2 &= a^2 \pm 2ab + b^2,\\ (a\pm b)^3 &= a^3 \pm 3a^2b + 3ab^2 \pm b^3,\\ (a\pm b)^4 &= a^4 \pm 4a^3b + 6a^2b^2 \pm 4ab^3 + b^4,\\ &\cdots\cdots\cdots\cdots\cdots\cdots\\ (a+b)^n &= \sum_{k=0}^{n} C_n^k a^{n-k} b^k, \quad n = 1,\ 2,\ \ldots\end{aligned}$$

The last formula is known as the *binomial theorem*, where the C_n^k are binomial coefficients.

M1.4.2. Related Formulas

► **Formulas involving powers ≤ 4.**

$$\begin{aligned}a^2 - b^2 &= (a-b)(a+b),\\ a^3 + b^3 &= (a+b)(a^2 - ab + b^2),\\ a^3 - b^3 &= (a-b)(a^2 + ab + b^2),\\ a^4 - b^4 &= (a-b)(a+b)(a^2 + b^2),\\ (a+b+c)^2 &= a^2 + b^2 + c^2 + 2ab + 2ac + 2bc,\\ a^4 + a^2b^2 + b^4 &= (a^2 + ab + b^2)(a^2 - ab + b^2).\end{aligned}$$

► **Formulas involving arbitrary powers.** Let n be any positive integer. Then

$$a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + \cdots + ab^{n-2} + b^{n-1}).$$

If n is a positive even number, then

$$\begin{aligned} a^n - b^n &= (a+b)(a^{n-1} - a^{n-2}b + \cdots + ab^{n-2} - b^{n-1}) \\ &= (a-b)(a+b)(a^{n-2} + a^{n-4}b^2 + \cdots + a^2b^{n-4} + b^{n-2}). \end{aligned}$$

If n is a positive odd number, then

$$a^n + b^n = (a+b)(a^{n-1} - a^{n-2}b + \cdots - ab^{n-2} + b^{n-1}).$$

M1.5. Progressions

M1.5.1. Arithmetic Progression

1°. An *arithmetic progression*, or *arithmetic sequence*, is a sequence of real numbers for which each term, starting from the second, is the previous term plus a constant d, called the *common difference*, so that $a_{n+1} = a_n + d$, $n = 1, 2, 3, \ldots$ In general, the terms of an arithmetic progression are expressed as

$$a_n = a_1 + (n-1)d, \qquad n = 1,\ 2,\ 3,\ \ldots,$$

where a_1 is the first term of the progression. An arithmetic progression is called *increasing* if $d > 0$ and *decreasing* if $d < 0$.

2°. An arithmetic progression has the property

$$a_n = \tfrac{1}{2}(a_{n-1} + a_{n+1}).$$

3°. The sum of n first terms of an arithmetic progression is calculated as

$$S_n = a_1 + \cdots + a_n = \tfrac{1}{2}(a_1 + a_n)n = \tfrac{1}{2}[2a_1 + (n-1)d]n.$$

M1.5.2. Geometric Progression

1°. A *geometric progression*, or *geometric sequence*, is a sequence of real numbers for which each term, starting from the second, is the previous term multiplied by a constant q, called the *common ratio*, so that $a_{n+1} = a_n q$, $n = 1, 2, 3, \ldots$ In general, the terms of a geometric progression are expressed as

$$a_n = a_1 q^{n-1}, \qquad n = 1,\ 2,\ 3,\ \ldots,$$

where a_1 is the first term of the progression.

2°. A geometric progression with positive terms has the property

$$a_n = \sqrt{a_{n-1}a_{n+1}}.$$

3°. The sum of n first terms of a geometric progression is calculated as $(q \neq 1)$

$$S_n = a_1 + \cdots + a_n = a_1 \frac{1-q^n}{1-q}.$$

M1.6. Mean Values and Some Inequalities

M1.6.1. Arithmetic Mean, Geometric Mean, and Other Mean Values

The *arithmetic mean* of a set of n real numbers $a_1, a_2, \ldots, a_n$ is defined as

$$m_{\rm a} = \frac{a_1 + a_2 + \cdots + a_n}{n}. \qquad (1.6.1.1)$$

Geometric mean of n positive numbers $a_1, a_2, \ldots, a_n$:

$$m_{\rm g} = (a_1 a_2 \ldots a_n)^{1/n}. \qquad (1.6.1.2)$$

Harmonic mean of n real numbers $a_1, a_2, \ldots, a_n$:

$$m_{\rm h} = \frac{n}{(1/a_1) + (1/a_2) + \cdots + (1/a_n)}, \qquad a_k \neq 0. \qquad (1.6.1.3)$$

Quadratic mean (or *root mean square*) of n real numbers $a_1, a_2, \ldots, a_n$:

$$m_{\rm q} = \sqrt{\frac{a_1^2 + a_2^2 + \cdots + a_n^2}{n}}. \qquad (1.6.1.4)$$

M1.6.2. Inequalities for Mean Values

Given n positive numbers $a_1, a_2, \ldots, a_n$, the following inequalities hold true:

$$m_{\rm h} \leq m_{\rm g} \leq m_{\rm a} \leq m_{\rm q}, \qquad (1.6.1.5)$$

where the mean values are defined above by (1.6.1.1)–(1.6.1.4). The equalities in (1.6.1.5) are attained only if $a_1 = a_2 = \cdots = a_n$.

To make it easier to remember, let us rewrite inequalities (1.6.1.5) in words as

$$\boxed{\text{harmonic mean}} \leq \boxed{\text{geometric mean}} \leq \boxed{\text{arithmetic mean}} \leq \boxed{\text{quadratic mean}}.$$

M1.6.3. Some Inequalities of General Form

Let a_k and b_k be real numbers with $k = 1, 2, \ldots, n$.

Generalized triangle inequality:

$$\left|\sum_{k=1}^{n} a_k\right| \leq \sum_{k=1}^{n} |a_k|.$$

Cauchy's inequality (also known as the *Cauchy–Bunyakovsky inequality* or *Cauchy–Schwarz–Bunyakovsky inequality*):

$$\left(\sum_{k=1}^{n} a_k b_k\right)^2 \leq \left(\sum_{k=1}^{n} a_k^2\right)\left(\sum_{k=1}^{n} b_k^2\right).$$

Minkowski's inequality:

$$\left(\sum_{k=1}^{n} |a_k + b_k|^p\right)^{\frac{1}{p}} \leq \left(\sum_{k=1}^{n} |a_k|^p\right)^{\frac{1}{p}} + \left(\sum_{k=1}^{n} |b_k|^p\right)^{\frac{1}{p}}, \qquad p \geq 1.$$

M1.7. Some Mathematical Methods

M1.7.1. Proof by Contradiction

Proof by contradiction (also known as *reductio ad absurdum*) is an indirect method of mathematical proof. It is based on the law of non-contradiction (a statement cannot be true and false at the same time) and includes the following reasoning:

1. Suppose one has to prove some statement S.
2. One assumes that the opposite of S is true.
3. Based on known axioms, definitions, theorems, formulas, and the assumption of Item 2, one arrives at a contradiction (deduces some obviously false statement).
4. One concludes that the assumption of Item 2 is false and hence the original statement S is true, which was to be proved.

Example. (Euclid's proof of the irrationality of the square root of 2 by contradiction.)

1. It is required to prove that $\sqrt{2}$ is an irrational number, that is, a real number that cannot be represented as a fraction p/q, where p and q are both integers.

2. Assume the opposite: $\sqrt{2}$ is a rational number. This means that $\sqrt{2}$ can be represented as a fraction

$$\sqrt{2} = p/q. \tag{1.7.1.1}$$

Without loss of generality the fraction p/q is assumed to be irreducible, implying that p and q are mutually prime (have no common factor other than 1).

3. Square both sides of (1.7.1.1) and then multiply by q^2 to obtain

$$2q^2 = p^2. \tag{1.7.1.2}$$

The left-hand side is divisible by 2. Then the right-hand side, p^2, and hence p is also divisible by 2. Consequently, p is an even number so that

$$p = 2n, \tag{1.7.1.3}$$

where n is an integer. Substituting (1.7.1.3) into (1.7.1.2) and then dividing by 2 yields

$$q^2 = 2p^2. \tag{1.7.1.4}$$

Now it can be concluded, just as above, that q^2 and hence q must be divisible by 2. Consequently, q is an even number so that

$$q = 2m, \tag{1.7.1.5}$$

where m is an integer.

It is now apparent from (1.7.1.3) and (1.7.1.5) that the fraction p/q is not simple, since p and q have a common factor 2. This contradicts the assumption made in Item 2.

4. It follows from the results of Item 3 that the representation of $\sqrt{2}$ in the form of a fraction (1.7.1.1) is false, which means that $\sqrt{2}$ is irrational.

M1.7.2. Mathematical Induction

The method of proof by (complete) *mathematical induction* is based on the following reasoning:

1. Let $A(n)$ be a statement dependent on n with $n = 1, 2, \ldots$ (A is a hypothesis at this stage).
2. Base case. Suppose the initial statement $A(1)$ is true. This is usually established by direct substitution $n = 1$.
3. Induction step. Assume that $A(n)$ is true for any n and then, based on this assumption, prove that $A(n+1)$ is also true.
4. Principle of mathematical induction. From the results of Items 2–3 it is concluded that the statement $A(n)$ is true for any n.

Example.

1. Prove the formula for the sum of odd numbers

$$1 + 3 + 5 + \cdots + (2n - 1) = n^2 \tag{1.7.2.1}$$

for any natural n.

2. For $n = 1$, we have an obvious identity: $1 = 1$.

3. Let us assume that formula (1.7.2.1) holds for any n. To consider the case of $n + 1$, let us add the next term, $(2n + 1)$, to both sides of (1.7.2.1) to obtain

$$1 + 3 + 5 + \cdots + (2n - 1) + (2n + 1) = n^2 + (2n + 1) = (n + 1)^2.$$

Thus, from the assumption of the validity of formula (1.7.2.1) for any n it follows that (1.7.2.1) is also valid for $n + 1$.

4. According to the principle of mathematical induction, this proves formula (1.7.2.1).

Remark. The first step, the formulation of an original hypothesis, is the most difficult part of the method of mathematical induction. This step is often omitted from the method.

M1.7.3. Proof by Counterexample

A *counterexample* is an example which is used to prove that a statement (proposition) is false. Counterexamples play an important role in mathematics. Whereas a complicated proof may be the only way to demonstrate the validity of a particular theorem, a single counterexample is all that is needed to refute the validity of a proposed theorem.

In general, the scheme of a proof by counterexample is as follows:

1. Given a proposition: all elements a that belong to a set A also belong to a set (possess a property) B.

2. Refutation of the proposition: one specifies an element a_* (counterexample) that belongs to A but does not belong to B.

Example. *Proposition:* Numbers in the form $2^{2^n} + 1$, where n is a positive integer, were once thought to be prime.

These numbers are prime for $n = 1, 2, 3, 4$. But for $n = 5$, we have a counterexample, since

$$2^{2^5} + 1 = 4294967297 = 641 \times 6700417;$$

it is a composite number.

Conclusion: When faced with a number in the form $2^{2^n} + 1$, we are not allowed to assume it is either prime or composite, unless we know for sure for some other reason.

Bibliography for Chapter M1

Bronshtein, I. N. and Semendyayev, K. A., *Handbook of Mathematics, 4th Edition*, Springer-Verlag, Berlin, 2004.

Courant, R. and Robbins, H., *What Is Mathematics?: An Elementary Approach to Ideas and Methods, 2nd Edition*, Oxford University Press, Oxford, 1996.

Franklin, J. and Daoud, A., *Introduction to Proofs in Mathematics*, Prentice Hall, New York, 1988.

Garnier, R. and Taylor, J., *100% Mathematical Proof*, John Wiley & Sons, New York, 1996.

Gelbaum, B. R. and Olmsted, J. M. H., *Theorems and Counterexamples in Mathematics*, Springer-Verlag, New York, 1990.

Jordan, B. E. and Palow, W. P., *Integrated Arithmetic & Algebra*, Addison-Wesley, Boston, 1999.

Krantz, S. G., *Dictionary of Algebra, Arithmetic, and Trigonometry*, CRC Press, Boca Raton, Florida, 2001.

Pólya, G., *Mathematics and Plausible Reasoning, Vol. 1: Induction and Analogy in Mathematics*, Princeton University Press, Princeton, 1990.

Rossi, R. J., *Theorems, Corollaries, Lemmas, and Methods of Proof*, Wiley-Interscience, Hoboken, N.J., 2006.

Thompson, J. E., *Arithmetic for the Practical Man*, Van Nostrand Reinhold, New York, 1973.

Weisstein, E. W., *CRC Concise Encyclopedia of Mathematics, 2nd Edition*, CRC Press, Boca Raton, Florida, 2003.

Zwillinger, D., *CRC Standard Mathematical Tables and Formulae, 31st Edition*, CRC Press, Boca Raton, Florida, 2002.

Chapter M2

Elementary Functions

Basic elementary functions: power, exponential, logarithmic, trigonometric, and inverse trigonometric (arc-trigonometric or antitrigonometric) functions. All other elementary functions are obtained from the basic elementary functions and constants by means of the four arithmetic operations (addition, subtraction, multiplication, and division) and the operation of composition (composite functions).

The graphs and the main properties of the basic as well as some other frequently occurring elementary functions of the real variable are described below.

M2.1. Power, Exponential, and Logarithmic Functions

M2.1.1. Power Function: $y = x^\alpha$ (α is an Arbitrary Real Number)

► **Graphs of the power function.** General properties of the graphs: the point $(1, 1)$ belongs to all the graphs, and $y > 0$ for $x > 0$. For $\alpha > 0$, the graphs pass through the origin $(0, 0)$; for $\alpha < 0$, the graphs have the vertical asymptote $x = 0$ ($y \to +\infty$ as $x \to 0$). For $\alpha = 0$, the graph is a straight line parallel to the x-axis.

Consider more closely the following cases.

Case 1: $y = x^{2n}$, where n is a positive integer ($n = 1, 2, \ldots$). This function is defined for all real x and its range consists of all $y \geq 0$. This function is even, nonperiodic, and unbounded. It crosses the axis Oy and is tangential to the axis Ox at the origin $x = 0$, $y = 0$. On the interval $(-\infty, 0)$ this function decreases, and it increases on the interval $(0, +\infty)$. It attains its minimum value $y = 0$ at $x = 0$. The graph of the function $y = x^2$ (parabola) is given in Fig. M2.1 *a*.

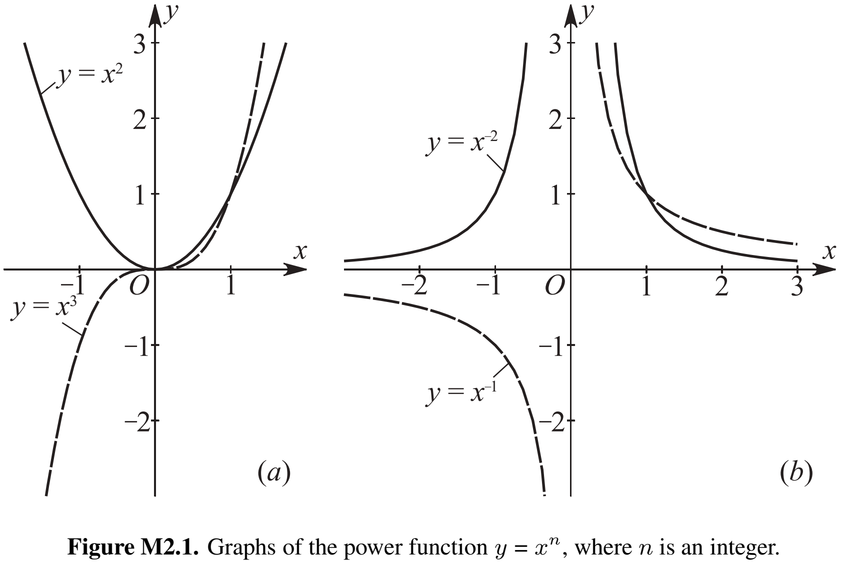

Figure M2.1. Graphs of the power function $y = x^n$, where n is an integer.

Case 2: $y = x^{2n+1}$, where n is a positive integer. This function is defined on the entire x-axis and its range coincides with the y-axis. This function is odd, nonperiodic, and unbounded. It crosses the x-axis and the y-axis at the origin $x = 0$, $y = 0$. It is an increasing function on the entire axis with no points of extremum, the origin being its inflection point. The graph of the function $y = x^3$ (cubic parabola) is shown in Fig. M2.1 *a*.

Case 3: $y = x^{-2n}$, where n is a positive integer. This function is defined for all $x \neq 0$, and its range is the semiaxis $y > 0$. It is an even, nonperiodic, unbounded function having no intersection with the coordinate axes. It increases on the interval $(-\infty, 0)$, decreases on the interval $(0, +\infty)$, and has no points of extremum. The graph of the function has a vertical asymptote $x = 0$. The graph of the function $y = x^{-2}$ is given in Fig. M2.1 *b*.

Case 4: $y = x^{-2n+1}$, where n is a positive integer. This function is defined for all $x \neq 0$, and its range is the entire y-axis. It is an odd, nonperiodic, unbounded function with no intersections with the coordinate axes. This is a decreasing function on the entire axis with no points of extremum. It has a vertical asymptote $x = 0$. The graph of the function $y = x^{-1}$ is given in Fig. M2.1 *b*.

Case 5: $y = x^\alpha$ with a noninteger $\alpha > 0$. This function is defined for all* $x \geq 0$ and its range is the semiaxis $y \geq 0$. This function is neither odd nor even and it is nonperiodic and unbounded. It crosses the axes Ox and Oy at the origin $x = 0$, $y = 0$ and increases everywhere in its domain, taking its smallest value at the point $x = 0$, $y = 0$. The graph of the function $y = x^{1/2}$ is given in Fig. M2.2.

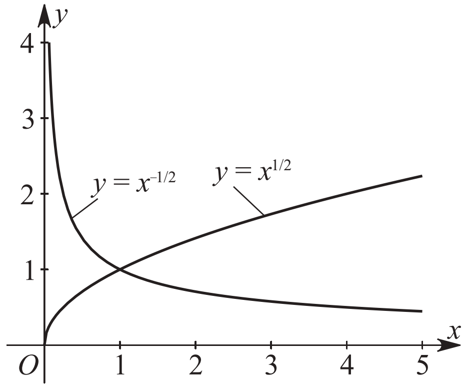

Figure M2.2. Graphs of the power function $y = x^\alpha$, where α is a noninteger.

Case 6: $y = x^\alpha$ with a noninteger $\alpha < 0$. This function is defined for all $x > 0$ and its range is the semiaxis $y > 0$. This function is neither odd nor even, it is nonperiodic and unbounded, and it has no intersections with the coordinate axes, which coincide with its horizontal and vertical asymptotes. This function is decreasing on its entire domain and has no points of extremum. The graph of the function $y = x^{-1/2}$ is given in Fig. M2.2.

► **Properties of the power function.** Basic properties of the power function:

$$x^\alpha x^\beta = x^{\alpha+\beta}, \quad (x_1 x_2)^\alpha = x_1^\alpha x_2^\alpha, \quad (x^\alpha)^\beta = x^{\alpha\beta},$$

for any α and β, where $x > 0$, $x_1 > 0$, $x_2 > 0$.

---

* In fact, the power function $y = x^{1/n}$ with an odd integer n is also defined for all $x < 0$. Here, however, it is always assumed that $x \geq 0$. A similar assumption is made with regard to the functions of the form $y = x^{m/n}$, where m is a positive integer and m/n is an irreducible fraction.

Differentiation and integration formulas:

$$(x^\alpha)' = \alpha x^{\alpha-1}, \qquad \int x^\alpha\,dx = \begin{cases} \dfrac{x^{\alpha+1}}{\alpha+1} + C & \text{if } \alpha \neq -1, \\ \ln|x| + C & \text{if } \alpha = -1. \end{cases}$$

The Taylor series expansion in a neighborhood of an arbitrary point x_0:

$$x^\alpha = \sum_{n=0}^{\infty} C_\alpha^n x_0^{\alpha-n}(x - x_0)^n \quad \text{for} \quad |x - x_0| < |x_0|,$$

where $C_\alpha^n = \dfrac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!}$ are binomial coefficients.

M2.1.2. Exponential Function: $y = a^x$ $(a > 0,\ a \neq 1)$

► **Graphs of the exponential function.** This function is defined for all x and its range is the semiaxis $y > 0$. This function is neither odd nor even, it is nonperiodic and unbounded, and it crosses the axis Oy at $y = 1$ and does not cross the axis Ox. For $a > 1$, it is an increasing function on the entire x-axis; for $0 < a < 1$, it is a decreasing function. This function has no extremal points; the axis Ox is its horizontal asymptote. The graphs of these functions have the following common property: they pass through the point $(0, 1)$. The graph of $y = a^x$ is symmetrical to the graph of $y = (1/a)^x$ with respect to the y-axis. For $a > 1$, the function a^x grows faster than any power of x as $x \to +\infty$, and it decays faster than any power of $1/x$ as $x \to -\infty$. The graphs of the functions $y = 2^x$ and $y = (1/2)^x$ are given in Fig. M2.3.

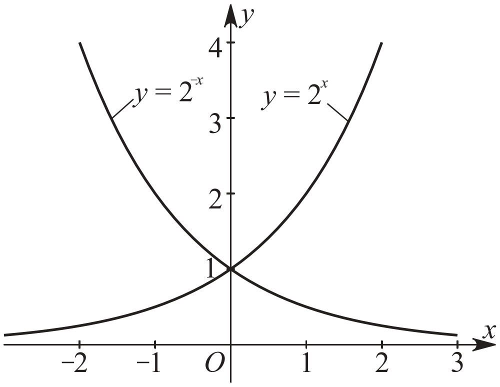

Figure M2.3. Graphs of the exponential function.

► **Properties of the exponential function.** Basic properties of the exponential function:

$$a^{x_1}a^{x_2} = a^{x_1+x_2}, \qquad a^x b^x = (ab)^x, \qquad (a^{x_1})^{x_2} = a^{x_1x_2}.$$

Number e, *base of natural (Napierian) logarithms*, and the function e^x:

$$e = \lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n = 2.718281\dots, \qquad e^x = \lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n.$$

The formula for passing from an arbitrary base a to the base e of natural logarithms:

$$a^x = e^{x\ln a}.$$

The inequality

$$a^{x_1} > a^{x_2} \iff \begin{cases} x_1 > x_2 & \text{if } a > 1, \\ x_1 < x_2 & \text{if } 0 < a < 1. \end{cases}$$

The limit relations for any $a > 1$ and $b > 0$:

$$\lim_{x\to+\infty} \frac{a^x}{|x|^b} = \infty, \qquad \lim_{x\to-\infty} a^x|x|^b = 0.$$

Differentiation and integration formulas:

$$(e^x)' = e^x, \qquad \int e^x\,dx = e^x + C;$$

$$(a^x)' = a^x \ln a, \qquad \int a^x\,dx = \frac{a^x}{\ln a} + C.$$

Power series expansion:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

M2.1.3. Logarithmic Function: $y = \log_a x$ $(a > 0,\ a \neq 1)$

► **Graphs of the logarithmic function.** This function is defined for all $x > 0$ and its range is the entire y-axis. The function is neither odd nor even; it is nonperiodic and unbounded; it crosses the axis Ox at $x = 1$ and does not cross the axis Oy. For $a > 1$, this function is increasing, and for $0 < a < 1$, it is a decreasing function; it has no extremal points, and the axis Oy is its vertical asymptote. The common property of the graphs of such functions is that they all pass through the point $(1, 0)$. The graph of the function $y = \log_a x$ is symmetric to that of $y = \log_{1/a} x$ with respect to the x-axis. The modulus of the logarithmic function tends to infinity slower than any power of x as $x \to +\infty$ and slower than any power of $1/x$ as $x \to +0$. The graphs of the functions $y = \log_2 x$ and $y = \log_{1/2} x$ are shown in Fig. M2.4.

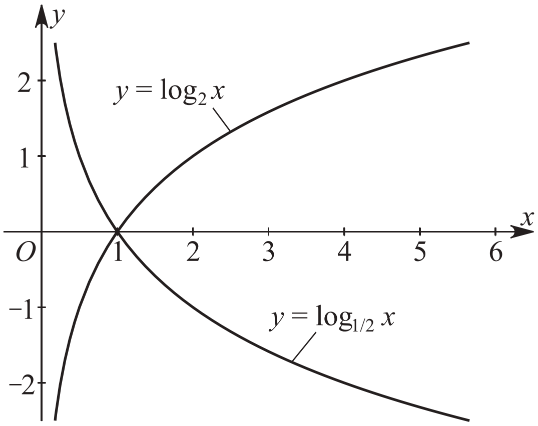

Figure M2.4. Graphs of the logarithmic function.

► **Properties of the logarithmic function.** By definition, the logarithmic function is the inverse of the exponential function. The following equivalence relation holds:

$$y = \log_a x \quad\Longleftrightarrow\quad x = a^y,$$

where $a > 0$, $a \neq 1$.

Basic properties of the logarithmic function:

$$a^{\log_a x} = x, \qquad \log_a(x_1 x_2) = \log_a x_1 + \log_a x_2,$$
$$\log_a(x^k) = k\log_a x, \quad \log_a x = \frac{\log_b x}{\log_b a},$$

where $x > 0$, $x_1 > 0$, $x_2 > 0$, $a > 0$, $a \neq 1$, $b > 0$, $b \neq 1$.

The simplest inequality:

$$\log_a x_1 > \log_a x_2 \quad\Longleftrightarrow\quad \begin{cases} x_1 > x_2 & \text{if } a > 1, \\ x_1 < x_2 & \text{if } 0 < a < 1. \end{cases}$$

For any $b > 0$, the following limit relations hold:

$$\lim_{x\to+\infty} \frac{\log_a x}{x^b} = 0, \qquad \lim_{x\to+0} x^b \log_a x = 0.$$

The logarithmic function with the base e (*base of natural logarithms, Napierian base*) is denoted by

$$\log_e x = \ln x,$$

where $e = \lim\limits_{n\to\infty}\left(1 + \frac{1}{n}\right)^n = 2.718281\ldots$

Formulas for passing from an arbitrary base a to the Napierian base e:

$$\log_a x = \frac{\ln x}{\ln a}.$$

Differentiation and integration formulas:

$$(\ln x)' = \frac{1}{x}, \qquad \int \ln x\,dx = x\ln x - x + C.$$

Power series expansion:

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n-1}\frac{x^n}{n} + \cdots = \sum_{k=1}^{\infty}(-1)^{k-1}\frac{x^k}{k}, \qquad -1 < x \leq 1.$$

M2.2. Trigonometric Functions

M2.2.1. Trigonometric Circle. Definition of Trigonometric Functions

► **Trigonometric circle. Degrees and radians.** *Trigonometric circle* is the circle of unit radius with center at the origin of an orthogonal coordinate system Oxy. The coordinate axes divide the circle into four quarters (*quadrants*); see Fig. M2.5. Consider rotation of the polar radius issuing from the origin O and ending at a point M of the trigonometric circle.

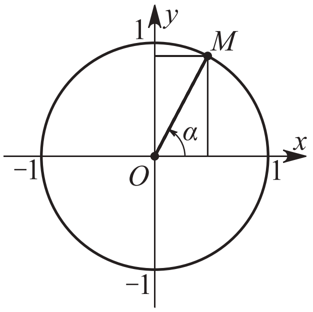

Figure M2.5. Trigonometric circle.

Let α be the angle between the x-axis and the polar radius OM measured from the positive direction of the x-axis. This angle is assumed positive in the case of counterclockwise rotation and negative in the case of clockwise rotation.

Angles are measured either in radians or in degrees. One radian is the angle at the vertex of the sector of the trigonometric circle supported by its arc of unit length. One degree is the angle at the vertex of the sector of the trigonometric circle supported by its arc of length $\pi/180$. The radians are related to the degrees by the formulas

$$1 \text{ radian} = \frac{180^\circ}{\pi}; \qquad 1^\circ = \frac{\pi}{180}.$$

► **Definition of trigonometric functions.** The *sine* of α is the ordinate (the projection to the axis Oy) of the point on the trigonometric circle corresponding to the angle of α radians. The *cosine* of α is the abscissa (projection to the axis Ox) of that point (see Fig. M2.5). The sine and the cosine are basic trigonometric functions and are denoted, respectively, by $\sin\alpha$ and $\cos\alpha$.

Other trigonometric functions are *tangent*, *cotangent*, *secant*, and *cosecant*. These are derived from the basic trigonometric functions, sine and cosine, as follows:

$$\tan\alpha = \frac{\sin\alpha}{\cos\alpha}, \qquad \cot\alpha = \frac{\cos\alpha}{\sin\alpha}, \qquad \sec\,\alpha = \frac{1}{\cos\alpha}, \qquad \operatorname{cosec}\alpha = \frac{1}{\sin\alpha}.$$

Table M2.1 gives the signs of the trigonometric functions in different quadrants. The signs and the values of $\sin\alpha$ and $\cos\alpha$ do not change if the argument α is incremented by $\pm 2\pi n$, where $n = 1,\ 2,\ \ldots$ The signs and the values of $\tan\alpha$ and $\cot\alpha$ do not change if the argument α is incremented by $\pm\pi n$, where $n = 1,\ 2,\ \ldots$

TABLE M2.1
Signs of trigonometric functions in different quarters.

Quarter	Angle in radians	$\sin\alpha$	$\cos\alpha$	$\tan\alpha$	$\cot\alpha$	$\sec\alpha$	$\operatorname{cosec}\alpha$
I	$0<\alpha<\frac{\pi}{2}$	+	+	+	+	+	+
II	$\frac{\pi}{2}<\alpha<\pi$	+	−	−	−	−	+
III	$\pi<\alpha<\frac{3\pi}{2}$	−	−	+	+	−	−
IV	$\frac{3\pi}{2}<\alpha<2\pi$	−	+	−	−	+	−

TABLE M2.2
Numerical values of trigonometric functions for some angles α (in radians).

Angle α	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{3\pi}{4}$	$\frac{5\pi}{6}$	π
$\sin\alpha$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0
$\cos\alpha$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$	$-\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{3}}{2}$	-1
$\tan\alpha$	0	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$	∞	$-\sqrt{3}$	-1	$-\frac{\sqrt{3}}{3}$	0
$\cot\alpha$	∞	$\sqrt{3}$	1	$\frac{\sqrt{3}}{3}$	0	$-\frac{\sqrt{3}}{3}$	-1	$-\sqrt{3}$	∞

Table M2.2 gives the values of trigonometric functions for some values of their argument (the symbol ∞ means that the function is undefined for the corresponding value of its argument).

M2.2.2. Graphs of Trigonometric Functions

► **Sine: $y = \sin x$.** This function is defined for all x and its range is $y \in [-1, 1]$. The sine is an odd, bounded, periodic function (with period 2π). It crosses the axis Oy at the point $y = 0$ and crosses the axis Ox at the points $x = \pi n$, $n = 0, \pm 1, \pm 2, \ldots$ The sine is an increasing function on every segment $[-\frac{\pi}{2} + 2\pi n, \frac{\pi}{2} + 2\pi n]$ and is a decreasing function on every segment $[\frac{\pi}{2} + 2\pi n, \frac{3}{2}\pi + 2\pi n]$. For $x = \frac{\pi}{2} + 2\pi n$, it attains its maximal value ($y = 1$), and for $x = -\frac{\pi}{2} + 2\pi n$ it attains its minimal value ($y = -1$). The graph of the function $y = \sin x$ is called the *sinusoid* or *sine curve* and is shown in Fig. M2.6.

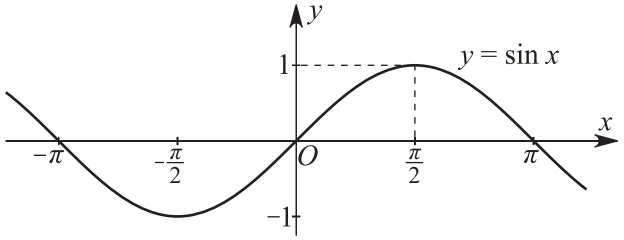

Figure M2.6. Graph of the function $y = \sin x$.

► **Cosine: $y = \cos x$.** This function is defined for all x and its range is $y \in [-1, 1]$. The cosine is a bounded, even, periodic function (with period 2π). It crosses the axis Oy at the point $y = 1$, and crosses the axis Ox at the points $x = \frac{\pi}{2} + \pi n$. The cosine is an increasing function on every segment $[-\pi + 2\pi n, 2\pi n]$ and is a decreasing function on every segment $[2\pi n, \pi + 2\pi n]$, $n = 0, \pm 1, \pm 2, \ldots$ For $x = 2\pi n$ it attains its maximal value ($y = 1$), and for $x = \pi + 2\pi n$ it attains its minimal value ($y = -1$). The graph of the function $y = \cos x$ is a sinusoid obtained by shifting the graph of the function $y = \sin x$ by $\frac{\pi}{2}$ to the left along the axis Ox (see Fig. M2.7).

► **Tangent: $y = \tan x$.** This function is defined for all $x \neq \frac{\pi}{2} + \pi n$, $n = 0, \pm 1, \pm 2, \ldots$, and its range is the entire y-axis. The tangent is an unbounded, odd, periodic function (with period π). It crosses the axis Oy at the point $y = 0$ and crosses the axis Ox at the points $x = \pi n$. This is an increasing function on every interval $(-\frac{\pi}{2} + \pi n, \frac{\pi}{2} + \pi n)$. This function has no points of extremum and has vertical asymptotes at $x = \frac{\pi}{2} + \pi n$, $n = 0, \pm 1, \pm 2, \ldots$ The graph of the function $y = \tan x$ is given in Fig. M2.8.

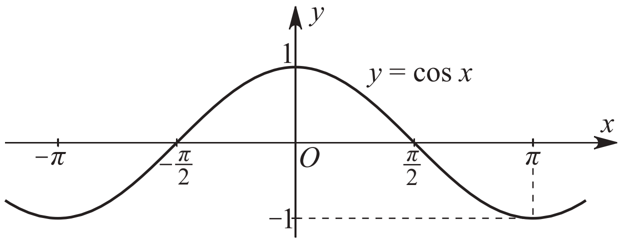

Figure M2.7. Graph of the function $y = \cos x$.

► **Cotangent: $y = \cot x$.** This function is defined for all $x \neq \pi n$, $n = 0, \pm 1, \pm 2, \ldots$, and its range is the entire y-axis. The cotangent is an unbounded, odd, periodic function (with period π). It crosses the axis Ox at the points $x = \frac{\pi}{2} + \pi n$, and does not cross the axis Oy. This is a decreasing function on every interval $(\pi n, \pi + \pi n)$. This function has no extremal points and has vertical asymptotes at $x = \pi n$, $n = 0, \pm 1, \pm 2, \ldots$ The graph of the function $y = \cot x$ is given in Fig. M2.9.

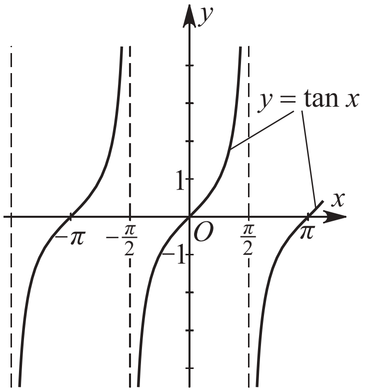

Figure M2.8. Graph of the function $y = \tan x$.

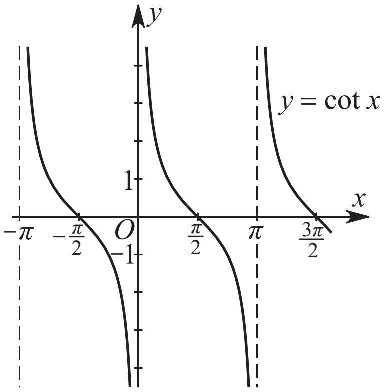

Figure M2.9. Graph of the function $y = \cot x$.

M2.2.3. Properties of Trigonometric Functions

► **Simplest relations.**

$$
\begin{aligned}
&\sin^2 x + \cos^2 x = 1, && \tan x \cot x = 1,\\
&\sin(-x) = -\sin x, && \cos(-x) = \cos x,\\
&\tan x = \frac{\sin x}{\cos x}, && \cot x = \frac{\cos x}{\sin x},\\
&\tan(-x) = -\tan x, && \cot(-x) = -\cot x,\\
&1 + \tan^2 x = \frac{1}{\cos^2 x}, && 1 + \cot^2 x = \frac{1}{\sin^2 x}.
\end{aligned}
$$

► **Reduction formulas.**

$$
\begin{aligned}
&\sin(x \pm 2n\pi) = \sin x, && \cos(x \pm 2n\pi) = \cos x,\\
&\sin(x \pm n\pi) = (-1)^n \sin x, && \cos(x \pm n\pi) = (-1)^n \cos x,\\
&\sin\Bigl(x \pm \frac{2n+1}{2}\pi\Bigr) = \pm(-1)^n \cos x, && \cos\Bigl(x \pm \frac{2n+1}{2}\pi\Bigr) = \mp(-1)^n \sin x,
\end{aligned}
$$

$$\sin\Bigl(x\pm\frac{\pi}{4}\Bigr)=\frac{\sqrt{2}}{2}(\sin x\pm\cos x),\qquad \cos\Bigl(x\pm\frac{\pi}{4}\Bigr)=\frac{\sqrt{2}}{2}(\cos x\mp\sin x),$$
$$\tan(x\pm n\pi)=\tan x,\qquad \cot(x\pm n\pi)=\cot x,$$
$$\tan\Bigl(x\pm\frac{2n+1}{2}\pi\Bigr)=-\cot x,\qquad \cot\Bigl(x\pm\frac{2n+1}{2}\pi\Bigr)=-\tan x,$$
$$\tan\Bigl(x\pm\frac{\pi}{4}\Bigr)=\frac{\tan x\pm 1}{1\mp\tan x},\qquad \cot\Bigl(x\pm\frac{\pi}{4}\Bigr)=\frac{\cot x\mp 1}{1\pm\cot x},$$

where $n = 1, 2, \ldots$

► **Relations between trigonometric functions of single argument.**

$$\begin{aligned}
\sin x &= \pm\sqrt{1-\cos^2 x}=\pm\frac{\tan x}{\sqrt{1+\tan^2 x}}=\pm\frac{1}{\sqrt{1+\cot^2 x}},\\
\cos x &= \pm\sqrt{1-\sin^2 x}=\pm\frac{1}{\sqrt{1+\tan^2 x}}=\pm\frac{\cot x}{\sqrt{1+\cot^2 x}},\\
\tan x &= \pm\frac{\sin x}{\sqrt{1-\sin^2 x}}=\pm\frac{\sqrt{1-\cos^2 x}}{\cos x}=\frac{1}{\cot x},\\
\cot x &= \pm\frac{\sqrt{1-\sin^2 x}}{\sin x}=\pm\frac{\cos x}{\sqrt{1-\cos^2 x}}=\frac{1}{\tan x}.
\end{aligned}$$

The sign before the radical is determined by the quarter in which the argument takes its values.

► **Addition and subtraction of trigonometric functions.**

$$\begin{aligned}
&\sin x+\sin y=2\sin\Bigl(\frac{x+y}{2}\Bigr)\cos\Bigl(\frac{x-y}{2}\Bigr),\\
&\sin x-\sin y=2\sin\Bigl(\frac{x-y}{2}\Bigr)\cos\Bigl(\frac{x+y}{2}\Bigr),\\
&\cos x+\cos y=2\cos\Bigl(\frac{x+y}{2}\Bigr)\cos\Bigl(\frac{x-y}{2}\Bigr),\\
&\cos x-\cos y=-2\sin\Bigl(\frac{x+y}{2}\Bigr)\sin\Bigl(\frac{x-y}{2}\Bigr),\\
&\sin^2 x-\sin^2 y=\cos^2 y-\cos^2 x=\sin(x+y)\sin(x-y),\\
&\sin^2 x-\cos^2 y=-\cos(x+y)\cos(x-y),\\
&\tan x\pm\tan y=\frac{\sin(x\pm y)}{\cos x\cos y},\quad \cot x\pm\cot y=\frac{\sin(y\pm x)}{\sin x\sin y},\\
&a\cos x+b\sin x=r\sin(x+\varphi)=r\cos(x-\psi).
\end{aligned}$$

Here, $r=\sqrt{a^2+b^2}$, $\sin\varphi=a/r$, $\cos\varphi=b/r$, $\sin\psi=b/r$, and $\cos\psi=a/r$.

► **Products of trigonometric functions.**

$$\begin{aligned}
\sin x\sin y&=\tfrac{1}{2}[\cos(x-y)-\cos(x+y)],\\
\cos x\cos y&=\tfrac{1}{2}[\cos(x-y)+\cos(x+y)],\\
\sin x\cos y&=\tfrac{1}{2}[\sin(x-y)+\sin(x+y)].
\end{aligned}$$

▶ Powers of trigonometric functions.

$$\cos^2 x = \tfrac{1}{2}\cos 2x + \tfrac{1}{2}, \qquad \sin^2 x = -\tfrac{1}{2}\cos 2x + \tfrac{1}{2},$$

$$\cos^3 x = \tfrac{1}{4}\cos 3x + \tfrac{3}{4}\cos x, \qquad \sin^3 x = -\tfrac{1}{4}\sin 3x + \tfrac{3}{4}\sin x,$$

$$\cos^4 x = \tfrac{1}{8}\cos 4x + \tfrac{1}{2}\cos 2x + \tfrac{3}{8}, \qquad \sin^4 x = \tfrac{1}{8}\cos 4x - \tfrac{1}{2}\cos 2x + \tfrac{3}{8},$$

$$\cos^5 x = \tfrac{1}{16}\cos 5x + \tfrac{5}{16}\cos 3x + \tfrac{5}{8}\cos x, \qquad \sin^5 x = \tfrac{1}{16}\sin 5x - \tfrac{5}{16}\sin 3x + \tfrac{5}{8}\sin x,$$

$$\cos^{2n} x = \frac{1}{2^{2n-1}}\sum_{k=0}^{n-1} C_{2n}^{k}\cos[2(n-k)x] + \frac{1}{2^{2n}}C_{2n}^{n},$$

$$\cos^{2n+1} x = \frac{1}{2^{2n}}\sum_{k=0}^{n} C_{2n+1}^{k}\cos[(2n-2k+1)x],$$

$$\sin^{2n} x = \frac{1}{2^{2n-1}}\sum_{k=0}^{n-1}(-1)^{n-k} C_{2n}^{k}\cos[2(n-k)x] + \frac{1}{2^{2n}}C_{2n}^{n},$$

$$\sin^{2n+1} x = \frac{1}{2^{2n}}\sum_{k=0}^{n}(-1)^{n-k} C_{2n+1}^{k}\sin[(2n-2k+1)x].$$

Here, $n = 1,\ 2,\ \ldots$ and $C_m^k = \dfrac{m!}{k!\,(m-k)!}$ are binomial coefficients $(0! = 1)$.

▶ Addition formulas.

$$\sin(x \pm y) = \sin x\cos y \pm \cos x\sin y, \qquad \cos(x \pm y) = \cos x\cos y \mp \sin x\sin y,$$

$$\tan(x \pm y) = \frac{\tan x \pm \tan y}{1 \mp \tan x\tan y}, \qquad \cot(x \pm y) = \frac{1 \mp \tan x\tan y}{\tan x \pm \tan y}.$$

▶ Trigonometric functions of multiple arguments.

$$\cos 2x = 2\cos^2 x - 1 = 1 - 2\sin^2 x, \qquad \sin 2x = 2\sin x\cos x,$$

$$\cos 3x = -3\cos x + 4\cos^3 x, \qquad \sin 3x = 3\sin x - 4\sin^3 x,$$

$$\cos 4x = 1 - 8\cos^2 x + 8\cos^4 x, \qquad \sin 4x = 4\cos x\,(\sin x - 2\sin^3 x),$$

$$\cos 5x = 5\cos x - 20\cos^3 x + 16\cos^5 x, \qquad \sin 5x = 5\sin x - 20\sin^3 x + 16\sin^5 x,$$

$$\cos(2nx) = 1 + \sum_{k=1}^{n}(-1)^k \frac{n^2(n^2-1)\ldots[n^2-(k-1)^2]}{(2k)!}4^k\sin^{2k} x,$$

$$\cos[(2n+1)x] = \cos x\left\{1+\sum_{k=1}^{n}(-1)^k \frac{[(2n+1)^2-1][(2n+1)^2-3^2]\ldots[(2n+1)^2-(2k-1)^2]}{(2k)!}\sin^{2k} x\right\},$$

$$\sin(2nx) = 2n\cos x\left[\sin x + \sum_{k=1}^{n}(-4)^k \frac{(n^2-1)(n^2-2^2)\ldots(n^2-k^2)}{(2k-1)!}\sin^{2k-1} x\right],$$

$$\sin[(2n+1)x] = (2n+1)\left\{\sin x + \sum_{k=1}^{n}(-1)^k \frac{[(2n+1)^2-1][(2n+1)^2-3^2]\ldots[(2n+1)^2-(2k-1)^2]}{(2k+1)!}\sin^{2k+1} x\right\},$$

$$\tan 2x = \frac{2\tan x}{1-\tan^2 x}, \qquad \tan 3x = \frac{3\tan x - \tan^3 x}{1 - 3\tan^2 x}, \qquad \tan 4x = \frac{4\tan x - 4\tan^3 x}{1 - 6\tan^2 x + \tan^4 x},$$

where $n = 1,\ 2,\ \ldots$

▶ Trigonometric functions of half argument.

$$\sin^2\frac{x}{2}=\frac{1-\cos x}{2},\qquad \cos^2\frac{x}{2}=\frac{1+\cos x}{2},$$
$$\tan\frac{x}{2}=\frac{\sin x}{1+\cos x}=\frac{1-\cos x}{\sin x},\quad \cot\frac{x}{2}=\frac{\sin x}{1-\cos x}=\frac{1+\cos x}{\sin x},$$
$$\sin x=\frac{2\tan\frac{x}{2}}{1+\tan^2\frac{x}{2}},\quad \cos x=\frac{1-\tan^2\frac{x}{2}}{1+\tan^2\frac{x}{2}},\quad \tan x=\frac{2\tan\frac{x}{2}}{1-\tan^2\frac{x}{2}}.$$

▶ Differentiation formulas.

$$\frac{d\sin x}{dx}=\cos x,\quad \frac{d\cos x}{dx}=-\sin x,\quad \frac{d\tan x}{dx}=\frac{1}{\cos^2 x},\quad \frac{d\cot x}{dx}=-\frac{1}{\sin^2 x}.$$

▶ Integration formulas.

$$\int\sin x\,dx=-\cos x+C,\qquad \int\cos x\,dx=\sin x+C,$$
$$\int\tan x\,dx=-\ln|\cos x|+C,\quad \int\cot x\,dx=\ln|\sin x|+C,$$

where C is an arbitrary constant.

▶ Power series expansions.

$$\cos x=1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\cdots+(-1)^n\frac{x^{2n}}{(2n)!}+\cdots\qquad (|x|<\infty),$$
$$\sin x=x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\cdots+(-1)^n\frac{x^{2n+1}}{(2n+1)!}+\cdots\qquad (|x|<\infty),$$
$$\tan x=x+\frac{x^3}{3}+\frac{2x^5}{15}+\frac{17x^7}{315}+\cdots+\frac{2^{2n}(2^{2n}-1)|B_{2n}|}{(2n)!}x^{2n-1}+\cdots\quad (|x|<\pi/2),$$
$$\cot x=\frac{1}{x}-\left(\frac{x}{3}+\frac{x^3}{45}+\frac{2x^5}{945}+\cdots+\frac{2^{2n}|B_{2n}|}{(2n)!}x^{2n-1}+\cdots\right)\qquad (0<|x|<\pi),$$

where B_n are Bernoulli numbers (see Subsection M13.1.2).

▶ Representation in the form of infinite products.

$$\sin x=x\left(1-\frac{x^2}{\pi^2}\right)\left(1-\frac{x^2}{4\pi^2}\right)\left(1-\frac{x^2}{9\pi^2}\right)\ldots\left(1-\frac{x^2}{n^2\pi^2}\right)\ldots$$
$$\cos x=\left(1-\frac{4x^2}{\pi^2}\right)\left(1-\frac{4x^2}{9\pi^2}\right)\left(1-\frac{4x^2}{25\pi^2}\right)\ldots\left(1-\frac{4x^2}{(2n+1)^2\pi^2}\right)\ldots$$

▶ Euler and de Moivre formulas. Relation to hyperbolic functions.

$$e^{y+ix}=e^y(\cos x+i\sin x),\quad (\cos x+i\sin x)^n=\cos(nx)+i\sin(nx),\quad i^2=-1,$$
$$\sin(ix)=i\sinh x,\quad \cos(ix)=\cosh x,\quad \tan(ix)=i\tanh x,\quad \cot(ix)=-i\coth x.$$

M2.3. Inverse Trigonometric Functions

M2.3.1. Definitions. Graphs of Inverse Trigonometric Functions

► **Definitions of inverse trigonometric functions.** *Inverse trigonometric functions* (*arc functions*) are the functions that are inverse to the trigonometric functions. Since the trigonometric functions $\sin x$, $\cos x$, $\tan x$, $\cot x$ are periodic, the corresponding inverse functions, denoted by $\operatorname{Arcsin} x$, $\operatorname{Arccos} x$, $\operatorname{Arctan} x$, $\operatorname{Arccot} x$, are multi-valued. The following relations define the multi-valued inverse trigonometric functions:

$$\begin{aligned} &\sin(\operatorname{Arcsin} x) = x, \quad &&\cos(\operatorname{Arccos} x) = x,\\ &\tan(\operatorname{Arctan} x) = x, \quad &&\cot(\operatorname{Arccot} x) = x. \end{aligned}$$

These functions admit the following verbal definitions: $\operatorname{Arcsin} x$ is the angle whose sine is equal to x; $\operatorname{Arccos} x$ is the angle whose cosine is equal to x; $\operatorname{Arctan} x$ is the angle whose tangent is equal to x; $\operatorname{Arccot} x$ is the angle whose cotangent is equal to x.

The principal (single-valued) branches of the inverse trigonometric functions are denoted by

$$\begin{aligned} &\arcsin x \equiv \sin^{-1} x &&\text{(arcsine is the inverse of sine)},\\ &\arccos x \equiv \cos^{-1} x &&\text{(arccosine is the inverse of cosine)},\\ &\arctan x \equiv \tan^{-1} x &&\text{(arctangent is the inverse of tangent)},\\ &\operatorname{arccot} x \equiv \cot^{-1} x &&\text{(arccotangent is the inverse of cotangent)} \end{aligned}$$

and are determined by the inequalities

$$\begin{aligned} &-\tfrac{\pi}{2} \le \arcsin x \le \tfrac{\pi}{2}, \quad 0 \le \arccos x \le \pi \qquad (-1 \le x \le 1);\\ &-\tfrac{\pi}{2} < \arctan x < \tfrac{\pi}{2}, \quad 0 < \operatorname{arccot} x < \pi \qquad (-\infty < x < \infty). \end{aligned}$$

The following equivalent relations can be taken as definitions of single-valued inverse trigonometric functions:

$$\begin{aligned} &y = \arcsin x, \quad -1 \le x \le 1 &&\Longleftrightarrow \quad x = \sin y, \quad -\frac{\pi}{2} \le y \le \frac{\pi}{2};\\ &y = \arccos x, \quad -1 \le x \le 1 &&\Longleftrightarrow \quad x = \cos y, \quad 0 \le y \le \pi;\\ &y = \arctan x, \quad -\infty < x < +\infty &&\Longleftrightarrow \quad x = \tan y, \quad -\frac{\pi}{2} < y < \frac{\pi}{2};\\ &y = \operatorname{arccot} x, \quad -\infty < x < +\infty &&\Longleftrightarrow \quad x = \cot y, \quad 0 < y < \pi. \end{aligned}$$

The multi-valued and the single-valued inverse trigonometric functions are related by the formulas

$$\begin{aligned} \operatorname{Arcsin} x &= (-1)^n \arcsin x + \pi n,\\ \operatorname{Arccos} x &= \pm \arccos x + 2\pi n,\\ \operatorname{Arctan} x &= \arctan x + \pi n,\\ \operatorname{Arccot} x &= \operatorname{arccot} x + \pi n, \end{aligned}$$

where $n = 0, \pm 1, \pm 2, \ldots$

The graphs of inverse trigonometric functions are obtained from the graphs of the corresponding trigonometric functions by mirror reflection with respect to the straight line $y = x$ (with the domain of each function being taken into account).

▶ **Arcsine: $y = \arcsin x$.** This function is defined for all $x \in [-1, 1]$ and its range is $y \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. The arcsine is an odd, nonperiodic, bounded function that crosses the axes Ox and Oy at the origin $x = 0$, $y = 0$. This is an increasing function in its domain, and it takes its smallest value $y = -\frac{\pi}{2}$ at the point $x = -1$; it takes its largest value $y = \frac{\pi}{2}$ at the point $x = 1$. The graph of the function $y = \arcsin x$ is given in Fig. M2.10.

▶ **Arccosine: $y = \arccos x$.** This function is defined for all $x \in [-1, 1]$ and its range is $y \in [0, \pi]$. It is neither odd nor even. It is a nonperiodic, bounded function that crosses the axis Oy at the point $y = \frac{\pi}{2}$ and crosses the axis Ox at the point $x = 1$. This is a decreasing function in its domain, and at the point $x = -1$ it takes its largest value $y = \pi$; at the point $x = 1$ it takes its smallest value $y = 0$. For all x in its domain, the following relation holds: $\arccos x = \frac{\pi}{2} - \arcsin x$. The graph of the function $y = \arccos x$ is given in Fig. M2.11.

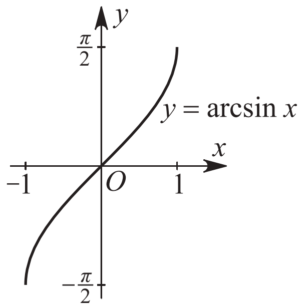

Figure M2.10. Graph of the function $y = \arcsin x$.

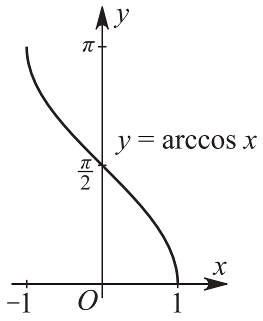

Figure M2.11. Graph of the function $y = \arccos x$.

▶ **Arctangent: $y = \arctan x$.** This function is defined for all x, and its range is $y \in (-\frac{\pi}{2}, \frac{\pi}{2})$. The arctangent is an odd, nonperiodic, bounded function that crosses the coordinate axes at the origin $x = 0$, $y = 0$. This is an increasing function with no points of extremum. It has two horizontal asymptotes: $y = -\frac{\pi}{2}$ (as $x \to -\infty$) and $y = \frac{\pi}{2}$ (as $x \to +\infty$). The graph of the function $y = \arctan x$ is given in Fig. M2.12.

▶ **Arccotangent: $y = \operatorname{arccot} x$.** This function is defined for all x, and its range is $y \in (0, \pi)$. The arccotangent is neither odd nor even. It is a nonperiodic, bounded function that crosses the axis Oy at the point $y = \frac{\pi}{2}$ and does not cross the axis Ox. This is a decreasing function on the entire x-axis with no points of extremum. It has two horizontal asymptotes $y = 0$ (as $x \to +\infty$) and $y = \pi$ (as $x \to -\infty$). For all x, the following relation holds: $\operatorname{arccot} x = \frac{\pi}{2} - \arctan x$. The graph of the function $y = \operatorname{arccot} x$ is given in Fig. M2.13.

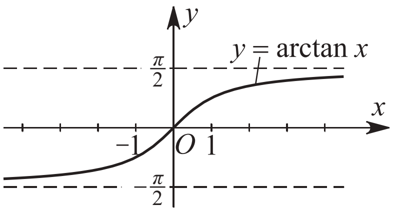

Figure M2.12. Graph of the function $y = \arctan x$.

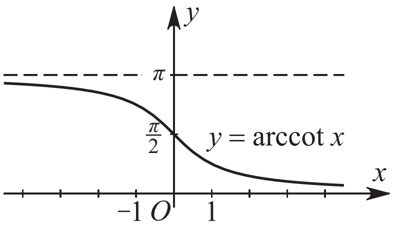

Figure M2.13. Graph of the function $y = \operatorname{arccot} x$.

M2.3.2. Properties of Inverse Trigonometric Functions

▶ **Simplest formulas.**

$$\sin(\arcsin x) = x, \quad \cos(\arccos x) = x,$$
$$\tan(\arctan x) = x, \quad \cot(\operatorname{arccot} x) = x.$$

► Some properties.

$$\arcsin(-x) = -\arcsin x, \qquad \arccos(-x) = \pi - \arccos x,$$
$$\arctan(-x) = -\arctan x, \qquad \operatorname{arccot}(-x) = \pi - \operatorname{arccot} x,$$

$$\arcsin(\sin x) = \begin{cases} x - 2n\pi & \text{if } 2n\pi - \frac{\pi}{2} \le x \le 2n\pi + \frac{\pi}{2}, \\ -x + 2(n+1)\pi & \text{if } (2n+1)\pi - \frac{\pi}{2} \le x \le 2(n+1)\pi + \frac{\pi}{2}, \end{cases}$$

$$\arccos(\cos x) = \begin{cases} x - 2n\pi & \text{if } 2n\pi \le x \le (2n+1)\pi, \\ -x + 2(n+1)\pi & \text{if } (2n+1)\pi \le x \le 2(n+1)\pi, \end{cases}$$

$$\arctan(\tan x) = x - n\pi \quad \text{if} \quad n\pi - \tfrac{\pi}{2} < x < n\pi + \tfrac{\pi}{2},$$
$$\operatorname{arccot}(\cot x) = x - n\pi \quad \text{if} \quad n\pi < x < (n+1)\pi.$$

► Relations between inverse trigonometric functions.

$$\arcsin x + \arccos x = \tfrac{\pi}{2}, \qquad \arctan x + \operatorname{arccot} x = \tfrac{\pi}{2};$$

$$\arcsin x = \begin{cases} \arccos\sqrt{1-x^2} & \text{if } 0 \le x \le 1, \\ -\arccos\sqrt{1-x^2} & \text{if } -1 \le x \le 0, \\ \arctan\dfrac{x}{\sqrt{1-x^2}} & \text{if } -1 < x < 1, \\ \operatorname{arccot}\dfrac{\sqrt{1-x^2}}{x} - \pi & \text{if } -1 \le x < 0; \end{cases} \qquad \arccos x = \begin{cases} \arcsin\sqrt{1-x^2} & \text{if } 0 \le x \le 1, \\ \pi - \arcsin\sqrt{1-x^2} & \text{if } -1 \le x \le 0, \\ \arctan\dfrac{\sqrt{1-x^2}}{x} & \text{if } 0 < x \le 1, \\ \operatorname{arccot}\dfrac{x}{\sqrt{1-x^2}} & \text{if } -1 < x < 1; \end{cases}$$

$$\arctan x = \begin{cases} \arcsin\dfrac{x}{\sqrt{1+x^2}} & \text{for any } x, \\ \arccos\dfrac{1}{\sqrt{1+x^2}} & \text{if } x \ge 0, \\ -\arccos\dfrac{1}{\sqrt{1+x^2}} & \text{if } x \le 0, \\ \operatorname{arccot}\dfrac{1}{x} & \text{if } x > 0; \end{cases} \qquad \operatorname{arccot} x = \begin{cases} \arcsin\dfrac{1}{\sqrt{1+x^2}} & \text{if } x \ge 0, \\ \pi - \arcsin\dfrac{1}{\sqrt{1+x^2}} & \text{if } x \le 0, \\ \arctan\dfrac{1}{x} & \text{if } x > 0, \\ \pi + \arctan\dfrac{1}{x} & \text{if } x < 0. \end{cases}$$

► Addition and subtraction of inverse trigonometric functions.

$$\arcsin x + \arcsin y = \arcsin\bigl(x\sqrt{1-y^2} + y\sqrt{1-x^2}\,\bigr) \qquad \text{for} \quad x^2 + y^2 \le 1,$$
$$\arccos x \pm \arccos y = \pm\arccos\bigl[xy \mp \sqrt{(1-x^2)(1-y^2)}\,\bigr] \qquad \text{for} \quad x \pm y \ge 0,$$
$$\arctan x + \arctan y = \arctan\frac{x+y}{1-xy} \quad \text{for} \quad xy < 1,$$
$$\arctan x - \arctan y = \arctan\frac{x-y}{1+xy} \quad \text{for} \quad xy > -1.$$

► Differentiation formulas.

$$\frac{d}{dx}\arcsin x = \frac{1}{\sqrt{1-x^2}}, \qquad \frac{d}{dx}\arccos x = -\frac{1}{\sqrt{1-x^2}},$$
$$\frac{d}{dx}\arctan x = \frac{1}{1+x^2}, \qquad \frac{d}{dx}\operatorname{arccot} x = -\frac{1}{1+x^2}.$$

► **Integration formulas.**

$$\int \arcsin x\,dx = x\arcsin x + \sqrt{1-x^2} + C, \qquad \int \arccos x\,dx = x\arccos x - \sqrt{1-x^2} + C,$$

$$\int \arctan x\,dx = x\arctan x - \frac{1}{2}\ln(1+x^2) + C, \qquad \int \operatorname{arccot} x\,dx = x\operatorname{arccot} x + \frac{1}{2}\ln(1+x^2) + C,$$

where C is an arbitrary constant.

► **Power series expansions.**

$$\arcsin x = x + \frac{1}{2}\frac{x^3}{3} + \frac{1\times 3}{2\times 4}\frac{x^5}{5} + \frac{1\times 3\times 5}{2\times 4\times 6}\frac{x^7}{7} + \cdots + \frac{1\times 3\times\cdots\times(2n-1)}{2\times 4\times\cdots\times(2n)}\frac{x^{2n+1}}{2n+1} + \cdots \quad (|x|<1),$$

$$\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots + (-1)^{n-1}\frac{x^{2n-1}}{2n-1} + \cdots \quad (|x|\le 1),$$

$$\arctan x = \frac{\pi}{2} - \frac{1}{x} + \frac{1}{3x^3} - \frac{1}{5x^5} + \cdots + (-1)^n\frac{1}{(2n-1)x^{2n-1}} + \cdots \quad (|x|>1).$$

The expansions for $\arccos x$ and $\operatorname{arccot} x$ can be obtained from the relations $\arccos x = \frac{\pi}{2} - \arcsin x$ and $\operatorname{arccot} x = \frac{\pi}{2} - \arctan x$.

M2.4. Hyperbolic Functions

M2.4.1. Definitions. Graphs of Hyperbolic Functions

► **Definitions of hyperbolic functions.** Hyperbolic functions are defined in terms of the exponential functions as follows:

$$\sinh x = \frac{e^x - e^{-x}}{2}, \quad \cosh x = \frac{e^x + e^{-x}}{2}, \quad \tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad \coth x = \frac{e^x + e^{-x}}{e^x - e^{-x}}.$$

The graphs of hyperbolic functions are given below.

► **Hyperbolic sine: $y = \sinh x$.** This function is defined for all x and its range is the entire y-axis. The hyperbolic sine is an odd, nonperiodic, unbounded function that crosses the axes Ox and Oy at the origin $x = 0$, $y = 0$. This is an increasing function in its domain with no points of extremum. The graph of the function $y = \sinh x$ is given in Fig. M2.14.

► **Hyperbolic cosine: $y = \cosh x$.** This function is defined for all x, and its range is $y \in [1, +\infty)$. The hyperbolic cosine is an even, nonperiodic, unbounded function that crosses the axis Oy at $y = 1$ and does not cross the axis Ox. This function is decreasing on the interval $(-\infty, 0)$ and increasing on the interval $(0, +\infty)$; it takes its smallest value $y = 1$ at $x = 0$. The graph of the function $y = \cosh x$ is given in Fig. M2.15.

► **Hyperbolic tangent: $y = \tanh x$.** This function is defined for all x, and its range is $y \in (-1, 1)$. The hyperbolic tangent is an odd, nonperiodic, bounded function that crosses the coordinate axes at the origin $x = 0$, $y = 0$. This is an increasing function on the entire x-axis and has two horizontal asymptotes: $y = -1$ (as $x \to -\infty$) and $y = 1$ (as $x \to +\infty$). The graph of the function $y = \tanh x$ is given in Fig. M2.16.

► **Hyperbolic cotangent: $y = \coth x$.** This function is defined for all $x \ne 0$, and its range consists of all $y \in (-\infty, -1)$ and $y \in (1, +\infty)$. The hyperbolic cotangent is an odd, nonperiodic, unbounded function that does not cross the coordinate axes. This is a decreasing function on each of the semiaxes of its domain; it has no points of extremum. It has two horizontal asymptotes: $y = -1$ (as $x \to -\infty$) and $y = 1$ (as $x \to +\infty$). The graph of the function $y = \coth x$ is given in Fig. M2.17.

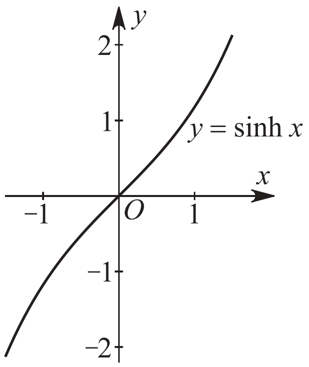

Figure M2.14. Graph of the function $y = \sinh x$.

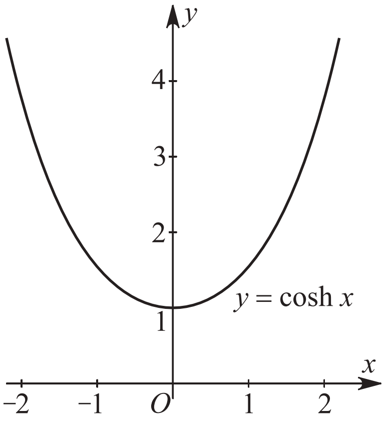

Figure M2.15. Graph of the function $y = \cosh x$.

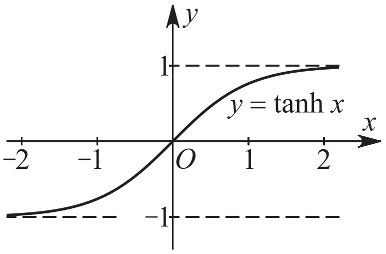

Figure M2.16. Graph of the function $y = \tanh x$.

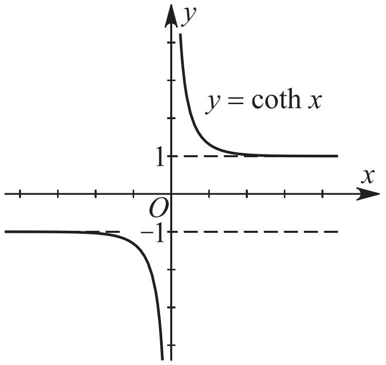

Figure M2.17. Graph of the function $y = \coth x$.

M2.4.2. Properties of Hyperbolic Functions

► **Simplest relations.**

$$\cosh^2 x - \sinh^2 x = 1, \qquad \tanh x \coth x = 1,$$
$$\sinh(-x) = -\sinh x, \qquad \cosh(-x) = \cosh x,$$
$$\tanh x = \frac{\sinh x}{\cosh x}, \qquad \coth x = \frac{\cosh x}{\sinh x},$$
$$\tanh(-x) = -\tanh x, \qquad \coth(-x) = -\coth x,$$
$$1 - \tanh^2 x = \frac{1}{\cosh^2 x}, \qquad \coth^2 x - 1 = \frac{1}{\sinh^2 x}.$$

► **Relations between hyperbolic functions of single argument ($x \geq 0$).**

$$\sinh x = \sqrt{\cosh^2 x - 1} = \frac{\tanh x}{\sqrt{1 - \tanh^2 x}} = \frac{1}{\sqrt{\coth^2 x - 1}},$$
$$\cosh x = \sqrt{\sinh^2 x + 1} = \frac{1}{\sqrt{1 - \tanh^2 x}} = \frac{\coth x}{\sqrt{\coth^2 x - 1}},$$
$$\tanh x = \frac{\sinh x}{\sqrt{\sinh^2 x + 1}} = \frac{\sqrt{\cosh^2 x - 1}}{\cosh x} = \frac{1}{\coth x},$$

$$\coth x = \frac{\sqrt{\sinh^2 x + 1}}{\sinh x} = \frac{\cosh x}{\sqrt{\cosh^2 x - 1}} = \frac{1}{\tanh x}.$$

► **Addition formulas.**

$$\sinh(x \pm y) = \sinh x \cosh y \pm \sinh y \cosh x, \qquad \cosh(x \pm y) = \cosh x \cosh y \pm \sinh x \sinh y,$$

$$\tanh(x \pm y) = \frac{\tanh x \pm \tanh y}{1 \pm \tanh x \tanh y}, \qquad \coth(x \pm y) = \frac{\coth x \coth y \pm 1}{\coth y \pm \coth x}.$$

► **Addition and subtraction of hyperbolic functions.**

$$\sinh x \pm \sinh y = 2\sinh\Big(\frac{x \pm y}{2}\Big)\cosh\Big(\frac{x \mp y}{2}\Big),$$

$$\cosh x + \cosh y = 2\cosh\Big(\frac{x + y}{2}\Big)\cosh\Big(\frac{x - y}{2}\Big),$$

$$\cosh x - \cosh y = 2\sinh\Big(\frac{x + y}{2}\Big)\sinh\Big(\frac{x - y}{2}\Big),$$

$$\sinh^2 x - \sinh^2 y = \cosh^2 x - \cosh^2 y = \sinh(x+y)\sinh(x-y),$$

$$\sinh^2 x + \cosh^2 y = \cosh(x+y)\cosh(x-y),$$

$$(\cosh x \pm \sinh x)^n = \cosh(nx) \pm \sinh(nx),$$

$$\tanh x \pm \tanh y = \frac{\sinh(x \pm y)}{\cosh x \cosh y}, \qquad \coth x \pm \coth y = \pm\frac{\sinh(x \pm y)}{\sinh x \sinh y},$$

where $n = 0, \pm1, \pm2, \ldots$

► **Products of hyperbolic functions.**

$$\sinh x \sinh y = \tfrac{1}{2}[\cosh(x+y) - \cosh(x-y)],$$

$$\cosh x \cosh y = \tfrac{1}{2}[\cosh(x+y) + \cosh(x-y)],$$

$$\sinh x \cosh y = \tfrac{1}{2}[\sinh(x+y) + \sinh(x-y)].$$

► **Powers of hyperbolic functions.**

$$\cosh^2 x = \tfrac{1}{2}\cosh 2x + \tfrac{1}{2}, \qquad \sinh^2 x = \tfrac{1}{2}\cosh 2x - \tfrac{1}{2},$$

$$\cosh^3 x = \tfrac{1}{4}\cosh 3x + \tfrac{3}{4}\cosh x, \qquad \sinh^3 x = \tfrac{1}{4}\sinh 3x - \tfrac{3}{4}\sinh x,$$

$$\cosh^4 x = \tfrac{1}{8}\cosh 4x + \tfrac{1}{2}\cosh 2x + \tfrac{3}{8}, \qquad \sinh^4 x = \tfrac{1}{8}\cosh 4x - \tfrac{1}{2}\cosh 2x + \tfrac{3}{8},$$

$$\cosh^5 x = \tfrac{1}{16}\cosh 5x + \tfrac{5}{16}\cosh 3x + \tfrac{5}{8}\cosh x, \qquad \sinh^5 x = \tfrac{1}{16}\sinh 5x - \tfrac{5}{16}\sinh 3x + \tfrac{5}{8}\sinh x,$$

$$\cosh^{2n} x = \frac{1}{2^{2n-1}}\sum_{k=0}^{n-1} C_{2n}^k \cosh[2(n-k)x] + \frac{1}{2^{2n}}C_{2n}^n,$$

$$\cosh^{2n+1} x = \frac{1}{2^{2n}}\sum_{k=0}^{n} C_{2n+1}^k \cosh[(2n-2k+1)x],$$

$$\sinh^{2n} x = \frac{1}{2^{2n-1}}\sum_{k=0}^{n-1} (-1)^k C_{2n}^k \cosh[2(n-k)x] + \frac{(-1)^n}{2^{2n}}C_{2n}^n,$$

$$\sinh^{2n+1} x = \frac{1}{2^{2n}}\sum_{k=0}^{n} (-1)^k C_{2n+1}^k \sinh[(2n-2k+1)x].$$

Here, $n = 1, 2, \ldots$ and C_m^k are binomial coefficients.

► **Hyperbolic functions of multiple argument.**

$$\cosh 2x = 2\cosh^2 x - 1, \qquad \sinh 2x = 2\sinh x\cosh x,$$

$$\cosh 3x = -3\cosh x + 4\cosh^3 x, \qquad \sinh 3x = 3\sinh x + 4\sinh^3 x,$$

$$\cosh 4x = 1 - 8\cosh^2 x + 8\cosh^4 x, \qquad \sinh 4x = 4\cosh x(\sinh x + 2\sinh^3 x),$$

$$\cosh 5x = 5\cosh x - 20\cosh^3 x + 16\cosh^5 x, \qquad \sinh 5x = 5\sinh x + 20\sinh^3 x + 16\sinh^5 x.$$

$$\cosh(nx) = 2^{n-1}\cosh^n x + \frac{n}{2}\sum_{k=0}^{[n/2]} \frac{(-1)^{k+1}}{k+1} C_{n-k-2}^{k-2} 2^{n-2k-2} (\cosh x)^{n-2k-2},$$

$$\sinh(nx) = \sinh x \sum_{k=0}^{[(n-1)/2]} 2^{n-k-1} C_{n-k-1}^{k} (\cosh x)^{n-2k-1}.$$

Here, C_m^k are binomial coefficients and $[A]$ stands for the integer part of the number A.

► **Hyperbolic functions of half argument.**

$$\sinh\frac{x}{2} = \operatorname{sign} x\sqrt{\frac{\cosh x - 1}{2}}, \qquad \cosh\frac{x}{2} = \sqrt{\frac{\cosh x + 1}{2}},$$

$$\tanh\frac{x}{2} = \frac{\sinh x}{\cosh x + 1} = \frac{\cosh x - 1}{\sinh x}, \qquad \coth\frac{x}{2} = \frac{\sinh x}{\cosh x - 1} = \frac{\cosh x + 1}{\sinh x}.$$

► **Differentiation formulas.**

$$\frac{d\sinh x}{dx} = \cosh x, \qquad \frac{d\cosh x}{dx} = \sinh x,$$

$$\frac{d\tanh x}{dx} = \frac{1}{\cosh^2 x}, \qquad \frac{d\coth x}{dx} = -\frac{1}{\sinh^2 x}.$$

► **Integration formulas.**

$$\int \sinh x\,dx = \cosh x + C, \qquad \int \cosh x\,dx = \sinh x + C,$$

$$\int \tanh x\,dx = \ln\cosh x + C, \qquad \int \coth x\,dx = \ln|\sinh x| + C,$$

where C is an arbitrary constant.

► **Power series expansions.**

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \cdots + \frac{x^{2n}}{(2n)!} + \cdots \qquad (|x| < \infty),$$

$$\sinh x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} + \cdots + \frac{x^{2n+1}}{(2n+1)!} + \cdots \qquad (|x| < \infty),$$

$$\tanh x = x - \frac{x^3}{3} + \frac{2x^5}{15} - \frac{17x^7}{315} + \cdots + (-1)^{n-1}\frac{2^{2n}(2^{2n}-1)|B_{2n}|x^{2n-1}}{(2n)!} + \cdots \qquad (|x| < \pi/2),$$

$$\coth x = \frac{1}{x} + \frac{x}{3} - \frac{x^3}{45} + \frac{2x^5}{945} - \cdots + (-1)^{n-1}\frac{2^{2n}|B_{2n}|x^{2n-1}}{(2n)!} + \cdots \qquad (|x| < \pi),$$

where B_n are Bernoulli numbers (see Subsection M13.1.2).

► **Relation to trigonometric functions.**

$$\sinh(ix)=i\sin x,\quad \cosh(ix)=\cos x,\quad \tanh(ix)=i\tan x,\quad \coth(ix)=-i\cot x,\qquad i^2=-1.$$

M2.5. Inverse Hyperbolic Functions

M2.5.1. Definitions. Graphs of Inverse Hyperbolic Functions

► **Definitions of inverse hyperbolic functions.** The *inverse hyperbolic functions* (also known as the *area hyperbolic functions*) are the inverses of the respective hyperbolic functions. The following notation is used for inverse hyperbolic functions:

$$\begin{aligned}
&\operatorname{arcsinh} x \equiv \operatorname{arsinh} x \equiv \sinh^{-1} x && \text{(inverse of hyperbolic sine)},\\
&\operatorname{arccosh} x \equiv \operatorname{arcosh} x \equiv \cosh^{-1} x && \text{(inverse of hyperbolic cosine)},\\
&\operatorname{arctanh} x \equiv \operatorname{artanh} x \equiv \tanh^{-1} x && \text{(inverse of hyperbolic tangent)},\\
&\operatorname{arccoth} x \equiv \operatorname{arcoth} x \equiv \coth^{-1} x && \text{(inverse of hyperbolic cotangent)}.
\end{aligned}$$

Inverse hyperbolic functions can be expressed in terms of logarithmic functions:

$$\begin{aligned}
&\operatorname{arcsinh} x = \ln\bigl(x+\sqrt{x^2+1}\,\bigr) \quad (x \text{ is any}); && \operatorname{arccosh} x = \ln\bigl(x+\sqrt{x^2-1}\,\bigr) \quad (x \ge 1);\\
&\operatorname{arctanh} x = \frac{1}{2}\ln\frac{1+x}{1-x} \quad (|x|<1); && \operatorname{arccoth} x = \frac{1}{2}\ln\frac{x+1}{x-1} \quad (|x|>1).
\end{aligned}$$

Here, only one (principal) branch of the function arccosh x is listed, the function itself being double-valued. In order to write out both branches of arccosh x, the symbol ± should be placed before the logarithm on the right-hand side of the formula.

The graphs of the inverse hyperbolic functions are given below. These are obtained from the graphs of the corresponding hyperbolic functions by mirror reflection with respect to the straight line $y = x$ (with the domain of each function taken into account).

► **Inverse hyperbolic sine: $y = \operatorname{arcsinh} x$.** This function is defined for all x, and its range coincides with the y-axis. The arcsinh x is an odd, nonperiodic, unbounded function that crosses the axes Ox and Oy at the origin $x = 0$, $y = 0$. This is an increasing function on the entire x-axis with no points of extremum. The graph of the function $y = \operatorname{arcsinh} x$ is given in Fig. M2.18.

► **Inverse hyperbolic cosine: $y = \operatorname{arccosh} x$.** This function is defined for all $x \in [1, +\infty)$, and its range consists of $y \in [0, +\infty)$. The arccosh x is neither odd nor even; it is nonperiodic and unbounded. It does not cross the axis Oy and crosses the axis Ox at the point $x = 1$. It is an increasing function in its domain with the minimal value $y = 0$ at $x = 1$. The graph of the function $y = \operatorname{arccosh} x$ is given in Fig. M2.19.

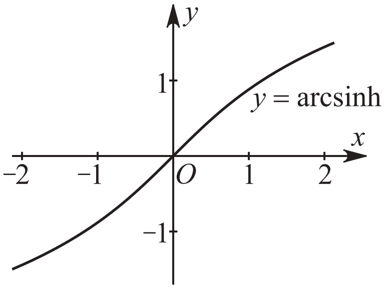

Figure M2.18. Graph of the function $y = \operatorname{arcsinh} x$.

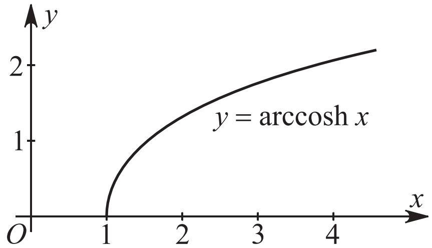

Figure M2.19. Graph of the function $y = \operatorname{arccosh} x$.

► **Inverse hyperbolic tangent: $y = \operatorname{arctanh} x$.** This function is defined for all $x \in (-1, 1)$, and its range consists of all y. The $\operatorname{arctanh} x$ is an odd, nonperiodic, unbounded function that crosses the coordinate axes at the origin $x = 0$, $y = 0$. This is an increasing function in its domain with no points of extremum and an inflection point at the origin. It has two vertical asymptotes: $x = \pm 1$. The graph of the function $y = \operatorname{arctanh} x$ is given in Fig. M2.20.

► **Inverse hyperbolic cotangent: $y = \operatorname{arccoth} x$.** This function is defined for $x \in (-\infty, -1)$ and $x \in (1, +\infty)$. Its range consists of all $y \neq 0$. The $\operatorname{arccoth} x$ is an odd, nonperiodic, unbounded function that does not cross the coordinate axes. It is a decreasing function on each of the semiaxes of its domain. This function has no points of extremum and has one horizontal asymptote $y = 0$ and two vertical asymptotes $x = \pm 1$. The graph of the function $y = \operatorname{arccoth} x$ is given in Fig. M2.21.

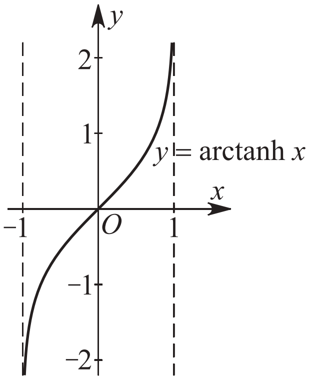

Figure M2.20. Graph of the function $y = \operatorname{arctanh} x$.

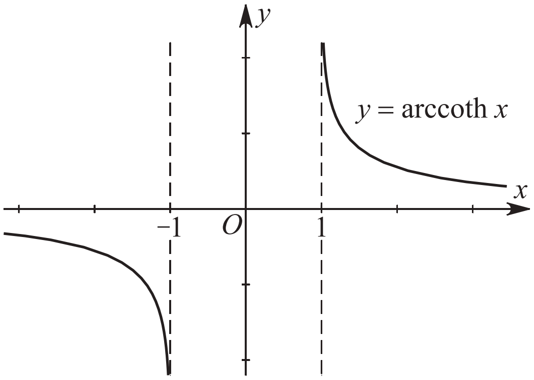

Figure M2.21. Graph of the function $y = \operatorname{arccoth} x$.

M2.5.2. Properties of Inverse Hyperbolic Functions

► **Simplest relations.**

$$\operatorname{arcsinh}(-x) = -\operatorname{arcsinh} x, \quad \operatorname{arctanh}(-x) = -\operatorname{arctanh} x, \quad \operatorname{arccoth}(-x) = -\operatorname{arccoth} x.$$

► **Relations between inverse hyperbolic functions.**

$$\begin{aligned}
\operatorname{arcsinh} x &= \operatorname{arccosh} \sqrt{x^2+1} = \operatorname{arctanh} \frac{x}{\sqrt{x^2+1}},\\
\operatorname{arccosh} x &= \operatorname{arcsinh} \sqrt{x^2-1} = \operatorname{arctanh} \frac{\sqrt{x^2-1}}{x},\\
\operatorname{arctanh} x &= \operatorname{arcsinh} \frac{x}{\sqrt{1-x^2}} = \operatorname{arccosh} \frac{1}{\sqrt{1-x^2}} = \operatorname{arccoth} \frac{1}{x}.
\end{aligned}$$

► **Addition and subtraction of inverse hyperbolic functions.**

$$\begin{gathered}
\operatorname{arcsinh} x \pm \operatorname{arcsinh} y = \operatorname{arcsinh}\left(x\sqrt{1+y^2} \pm y\sqrt{1+x^2}\right),\\
\operatorname{arccosh} x \pm \operatorname{arccosh} y = \operatorname{arccosh}\left[xy \pm \sqrt{(x^2-1)(y^2-1)}\right],\\
\operatorname{arcsinh} x \pm \operatorname{arccosh} y = \operatorname{arcsinh}\left[xy \pm \sqrt{(x^2+1)(y^2-1)}\right],\\
\operatorname{arctanh} x \pm \operatorname{arctanh} y = \operatorname{arctanh} \frac{x \pm y}{1 \pm xy}, \quad \operatorname{arctanh} x \pm \operatorname{arccoth} y = \operatorname{arctanh} \frac{xy \pm 1}{y \pm x}.
\end{gathered}$$

► **Differentiation formulas.**

$$\frac{d}{dx}\operatorname{arcsinh} x = \frac{1}{\sqrt{x^2+1}}, \qquad \frac{d}{dx}\operatorname{arccosh} x = \frac{1}{\sqrt{x^2-1}},$$
$$\frac{d}{dx}\operatorname{arctanh} x = \frac{1}{1-x^2} \quad (x^2<1), \qquad \frac{d}{dx}\operatorname{arccoth} x = \frac{1}{1-x^2} \quad (x^2>1).$$

► **Integration formulas.**

$$\int \operatorname{arcsinh} x\,dx = x\operatorname{arcsinh} x - \sqrt{1+x^2} + C,$$
$$\int \operatorname{arccosh} x\,dx = x\operatorname{arccosh} x - \sqrt{x^2-1} + C,$$
$$\int \operatorname{arctanh} x\,dx = x\operatorname{arctanh} x + \frac{1}{2}\ln(1-x^2) + C,$$
$$\int \operatorname{arccoth} x\,dx = x\operatorname{arccoth} x + \frac{1}{2}\ln(x^2-1) + C,$$

where C is an arbitrary constant.

► **Power series expansions.**

$$\operatorname{arcsinh} x = x - \frac{1}{2}\frac{x^3}{3} + \frac{1\times 3}{2\times 4}\frac{x^5}{5} - \cdots + (-1)^n \frac{1\times 3\times\cdots\times(2n-1)}{2\times 4\times\cdots\times(2n)}\frac{x^{2n+1}}{2n+1} + \cdots \quad (|x|<1),$$
$$\operatorname{arcsinh} x = \ln(2x) + \frac{1}{2}\frac{1}{2x^2} + \frac{1\times 3}{2\times 4}\frac{1}{4x^4} + \cdots + \frac{1\times 3\times\cdots\times(2n-1)}{2\times 4\times\cdots\times(2n)}\frac{1}{2nx^{2n}} + \cdots \quad (|x|>1),$$
$$\operatorname{arccosh} x = \ln(2x) - \frac{1}{2}\frac{1}{2x^2} - \frac{1\times 3}{2\times 4}\frac{1}{4x^4} - \cdots - \frac{1\times 3\times\cdots\times(2n-1)}{2\times 4\times\cdots\times(2n)}\frac{1}{2nx^{2n}} - \cdots \quad (|x|>1),$$
$$\operatorname{arctanh} x = x + \frac{x^3}{3} + \frac{x^5}{5} + \frac{x^7}{7} + \cdots + \frac{x^{2n+1}}{2n+1} + \cdots \quad (|x|<1),$$
$$\operatorname{arccoth} x = \frac{1}{x} + \frac{1}{3x^3} + \frac{1}{5x^5} + \frac{1}{7x^7} + \cdots + \frac{1}{(2n+1)x^{2n+1}} + \cdots \quad (|x|>1).$$

Bibliography for Chapter M2

Abramowitz, M. and Stegun, I. A. (Editors), *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, National Bureau of Standards Applied Mathematics, Washington, D.C., 1964.

Adams, R., *Calculus: A Complete Course, 6th Edition*, Pearson Education, Toronto, 2006.

Anton, H., Bivens, I., and Davis, S., *Calculus: Early Transcendentals Single and Multivariable, 8th Edition*, John Wiley & Sons, New York, 2005.

Bronshtein, I. N. and Semendyayev, K. A., *Handbook of Mathematics, 4th Edition*, Springer-Verlag, Berlin, 2004.

Courant, R. and John, F., *Introduction to Calculus and Analysis, Vol. 1*, Springer-Verlag, New York, 1999.

Edwards, C. H., and Penney, D., *Calculus, 6th Edition*, Pearson Education, Toronto, 2002.

Gradshteyn, I. S. and Ryzhik, I. M., *Tables of Integrals, Series, and Products, 6th Edition*, Academic Press, New York, 2000.

Kline, M., *Calculus: An Intuitive and Physical Approach, 2nd Edition*, Dover Publications, New York, 1998.

Korn, G. A. and Korn, T. M., *Mathematical Handbook for Scientists and Engineers, 2nd Edition*, Dover Publications, New York, 2000.

Prudnikov, A. P., Brychkov, Yu. A., and Marichev, O. I., *Integrals and Series, Vol. 1, Elementary Functions*, Gordon & Breach, New York, 1986.

Sullivan, M., *Trigonometry, 7th Edition*, Prentice Hall, Englewood Cliffs, New Jersey, 2004.

Thomas, G. B. and Finney, R. L., *Calculus and Analytic Geometry, 9th Edition*, Addison Wesley, Reading, Massachusetts, 1996.

Weisstein, E. W., *CRC Concise Encyclopedia of Mathematics, 2nd Edition*, CRC Press, Boca Raton, Florida, 2003.

Zill, D. G. and Dewar, J. M., *Trigonometry, 2nd Edition*, McGraw-Hill, New York, 1990.

Zwillinger, D., *CRC Standard Mathematical Tables and Formulae, 31st Edition*, CRC Press, Boca Raton, Florida, 2002.

Chapter M3

Elementary Geometry

M3.1. Plane Geometry

M3.1.1. Triangles

► **Plane triangle and its properties.**

1°. A *plane triangle*, or simply a *triangle*, is a plane figure bounded by three straight line segments (*sides*) connecting three noncollinear points (*vertices*) (Fig. M3.1*a*). The smaller angle between the two rays issuing from a vertex and passing through the other two vertices is called an (*interior*) *angle* of the triangle. The angle adjacent to an interior angle is called an *external angle* of the triangle. An external angle is equal to the sum of the two interior angles to which it is not adjacent.

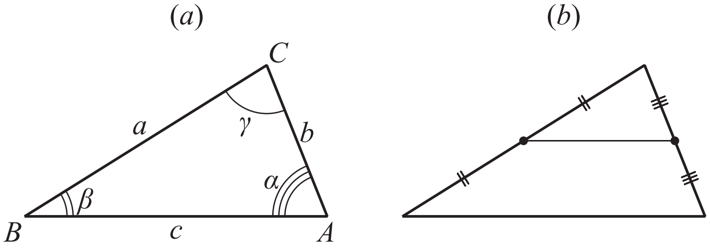

Figure M3.1. Plane triangle (*a*). Midline of a triangle (*b*).

A triangle is uniquely determined by any of the following sets of its parts:

1. Two angles and their included side.
2. Two sides and their included angle.
3. Three sides.

Depending on the angles, a triangle is said to be:

1. *Acute* if all three angles are acute.
2. *Right* (or *right-angled*) if one of the angles is right.
3. *Obtuse* if one of the angles is obtuse.

Depending on the relation between the side lengths, a triangle is said to be:

1. *Regular* (or *equilateral*) if all sides have the same length.
2. *Isosceles* if two of the sides are of equal length.
3. *Scalene* if all sides have different lengths.

2°. *Congruence tests for triangles:*

1. If two sides of a triangle and their included angle are congruent to the corresponding parts of another triangle, then the triangles are congruent.
2. If two angles of a triangle and their included side are congruent to the corresponding parts of another triangle, then the triangles are congruent.

3. If three sides of a triangle are congruent to the corresponding sides of another triangle, then the triangles are congruent.

3°. Triangles are said to be *similar* if their corresponding angles are equal and their corresponding sides are proportional.

Similarity tests for triangles:

1. If all three pairs of corresponding sides in a pair of triangles are in proportion, then the triangles are similar.
2. If two pairs of corresponding angles in a pair of triangles are congruent, then the triangles are similar.
3. If two pairs of corresponding sides in a pair of triangles are in proportion and the included angles are congruent, then the triangles are similar.

The areas of similar triangles are proportional to the squares of the corresponding linear parts (such as sides, altitudes, medians, etc.).

4°. The line connecting the midpoints of two sides of a triangle is called a *midline* of the triangle. The midline is parallel to and half as long as the third side (Fig. M3.1b).

Let a, b, and c be the lengths of the sides of a triangle; let α, β, and γ be the respective opposite angles (Fig. M3.1a); let R and r be the circumradius and the inradius, respectively; and let $p = \frac{1}{2}(a+b+c)$ be the semiperimeter.

Table M3.1 represents the basic properties and relations characterizing triangles.

TABLE M3.1
Basic properties and relations characterizing plane triangles.

No.	The name of property	Properties and relations
1	Triangle inequality	The length of any side of a triangle does not exceed the sum of lengths of the other two sides
2	Sum of angles of a triangle	$\alpha+\beta+\gamma=180^\circ$
3	Law of sines	$\dfrac{a}{\sin\alpha}=\dfrac{b}{\sin\beta}=\dfrac{c}{\sin\gamma}=2R$
4	Law of cosines	$c^2=a^2+b^2-2ab\cos\gamma$
5	Law of tangents	$\dfrac{a+b}{a-b}=\dfrac{\tan\left[\frac{1}{2}(\alpha+\beta)\right]}{\tan\left[\frac{1}{2}(\alpha-\beta)\right]}=\dfrac{\cot\left(\frac{1}{2}\gamma\right)}{\tan\left[\frac{1}{2}(\alpha-\beta)\right]}$
6	Theorem on projections (law of cosines)	$c=a\cos\beta+b\cos\alpha$
7	Trigonometric angle formulas	$\sin\dfrac{\gamma}{2}=\sqrt{\dfrac{(p-a)(p-b)}{ab}},\quad \cos\dfrac{\gamma}{2}=\sqrt{\dfrac{p(p-c)}{ab}},$ $\tan\dfrac{\gamma}{2}=\sqrt{\dfrac{(p-a)(p-b)}{p(p-c)}},\quad \sin\gamma=\dfrac{2}{ab}\sqrt{p(p-a)(p-b)(p-c)}$
8	Law of tangents	$\tan\gamma=\dfrac{c\sin\alpha}{b-c\cos\alpha}=\dfrac{c\sin\beta}{a-c\cos\beta}$
9	Mollweide's formulas	$\dfrac{a+b}{c}=\dfrac{\cos\left[\frac{1}{2}(\alpha-\beta)\right]}{\sin\left(\frac{1}{2}\gamma\right)}=\dfrac{\cos\left[\frac{1}{2}(\alpha-\beta)\right]}{\cos\left[\frac{1}{2}(\alpha+\beta)\right]},$ $\dfrac{a-b}{c}=\dfrac{\sin\left[\frac{1}{2}(\alpha-\beta)\right]}{\cos\left(\frac{1}{2}\gamma\right)}=\dfrac{\sin\left[\frac{1}{2}(\alpha-\beta)\right]}{\sin\left[\frac{1}{2}(\alpha+\beta)\right]}$

Table M3.2 permits one to find the sides and angles of an arbitrary triangle if three appropriately chosen sides and/or angles are given. From the relations given in Tables M3.1 and M3.2, one can derive all missing relations by cyclic permutations of the sides a, b, and c and the angles α, β, and γ.

TABLE M3.2
Solution of plane triangles.

No.	Three parts specified	Formulas for the remaining parts
1	Three sides a, b, c	*First method.* One of the angles is determined by the law of cosines, $\cos\alpha = \dfrac{b^2+c^2-a^2}{2bc}$. Then either the law of sines or the law of cosines is applied. *Second method.* One of the angles is determined by trigonometric angle formulas. Further proceed in a similar way. Remark. The sum of lengths of any two sides must be greater than the length of the third side.
2	Two sides a, b and the included angle γ	*First method.* The side c is determined by the law of cosines, $c = \sqrt{a^2+b^2-2ab\cos\gamma}$. The angle α is determined by either the law of cosines or the law of sines. The angle β is determined from the sum of angles in triangle, $\beta = 180^\circ - \alpha - \gamma$. *Second method.* $\alpha+\beta$ is found from the sum of angles in triangle, $\alpha+\beta = 180^\circ - \gamma$; $\alpha-\beta$ is found from the law of tangents, $\tan\dfrac{\alpha-\beta}{2} = \dfrac{a-b}{a+b}\cot\dfrac{\gamma}{2}$. Then α and β can be found. The third side c is determined by either the law of cosines or the law of sines.
3	A side c and the two angles α, β adjacent to it	The third angle γ is found from the sum of angles in triangle, $\gamma = 180^\circ - \alpha - \beta$. Sides a and b are determined by the law of sines.
4	Two sides a, b and the angle α opposite one of them	The second angle is determined by the law of sines, $\sin\beta = \dfrac{b}{a}\sin\alpha$. The third angle is $\gamma = 180^\circ - \alpha - \beta$. The third side is determined by the law of sines, $c = a\dfrac{\sin\gamma}{\sin\alpha}$. Remark. Five cases are possible: 1. $a > b$; i.e., the angle is opposite the greater side. Then $\alpha > \beta$, $\beta < 90^\circ$ (the larger angle is opposite the larger side), and the triangle is determined uniquely. 2. $a = b$; i.e., the triangle is isosceles and is determined uniquely. 3. $a < b$ and $b\sin\alpha < a$. Then there are two solutions, $\beta_1 + \beta_2 = 180^\circ$. 4. $a < b$ and $b\sin\alpha = a$. Then the solution is unique, $\beta = 90^\circ$. 5. $a < b$ and $b\sin\alpha > a$. Then there are no solutions.

► **Medians, angle bisectors, and altitudes of a triangle.** A straight line through a vertex of a triangle and the midpoint of the opposite side is called a *median* of the triangle (Fig. M3.2*a*). The three medians of a triangle intersect in a single point lying strictly inside the triangle, which is called the *centroid* or *center of gravity* of the triangle. This point cuts the medians in the ratio 2 : 1 (counting from the corresponding vertices).

The length of the median m_a to the side a is equal to

$$m_a = \frac{1}{2}\sqrt{2(b^2+c^2)-a^2} = \frac{1}{2}\sqrt{a^2+4b^2-4ab\cos\gamma}\,. \qquad \text{(M3.1.1.1)}$$

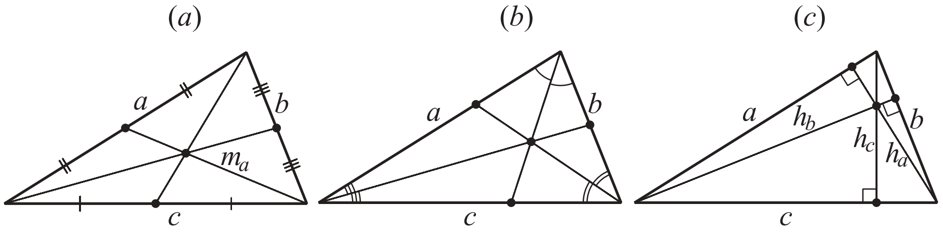

Figure M3.2. Medians (*a*), angle bisectors (*b*), and altitudes (*c*) of a triangle.

An *angle bisector* of a triangle is a line segment between a vertex and a point of the opposite side and dividing the angle at that vertex into two equal parts (Fig. M3.2*b*). The three angle bisectors intersect in a single point lying strictly inside the triangle. This point is equidistant from all sides and is called the *incenter* (the center of the *incircle* of the triangle). The angle bisector through a vertex cuts the opposite side in ratio proportional to the adjacent sides of the triangle.

The length of the angle bisector l_a drawn to the side a is given by the formulas

$$\begin{aligned} l_a &= \frac{\sqrt{bc[(b+c)^2-a^2]}}{b+c} = \frac{\sqrt{4p(p-a)bc}}{b+c}, \\ l_a &= \frac{2cb\cos\left(\frac{1}{2}\alpha\right)}{b+c} = 2R\frac{\sin\beta\sin\gamma}{\cos\left[\frac{1}{2}(\beta-\gamma)\right]} = 2p\frac{\sin\left(\frac{1}{2}\beta\right)\sin\left(\frac{1}{2}\gamma\right)}{\sin\beta+\sin\gamma}, \end{aligned} \qquad \text{(M3.1.1.2)}$$

where R is the circumradius (see below).

An *altitude* of a triangle is a straight line passing through a vertex and perpendicular to the straight line containing the opposite side (Fig. M3.2*c*). The three altitudes of a triangle intersect in a single point, called the *orthocenter* of the triangle.

The length of the altitude h_a to the side a is given by the formulas

$$\begin{aligned} h_a &= b\sin\gamma = c\sin\beta = \frac{bc}{2R}, \\ h_a &= 2(p-a)\cos\frac{\alpha}{2}\cos\frac{\beta}{2}\cos\frac{\gamma}{2} = 2(p-b)\sin\frac{\alpha}{2}\sin\frac{\beta}{2}\cos\frac{\gamma}{2}. \end{aligned} \qquad \text{(M3.1.1.3)}$$

The lengths of the altitude, the angle bisector, and the median through the same vertex satisfy the inequality $h_a \le l_a \le m_a$. If $h_a = l_a = m_a$, then the triangle is isosceles; moreover, the first equality implies the second, and vice versa.

► **Circumcircle and incircle.** A straight line passing through the midpoint of a segment and perpendicular to it is called the *perpendicular bisector* of the segment. The circle passing through the vertices of a triangle is called the *circumcircle* of the triangle. The center O_1 of the circumcircle, called the *circumcenter*, is the point where the perpendicular bisectors of the sides of the triangle meet (Fig. M3.3*a*). The feet of the perpendiculars drawn from a point Q on the circumcircle to the three sides of the triangle lie on the same straight line called the *Simpson line* of Q with respect to the triangle (Fig. M3.3*b*). The circumcenter, the orthocenter, and the centroid lie on a single line, called the *Euler line* (Fig. M3.3*c*).

The circle tangent to the three sides of a triangle and lying inside the triangle is called the *incircle* of the triangle. The center O_2 of the incircle (the *incenter*) is the point where the angle bisectors meet (Fig. M3.4*a*). The straight lines connecting the vertices of a triangle with the points at which the incircle is tangent to the respective opposite sides intersect in a single point G called the *Gergonne point* (Fig. M3.4*b*).

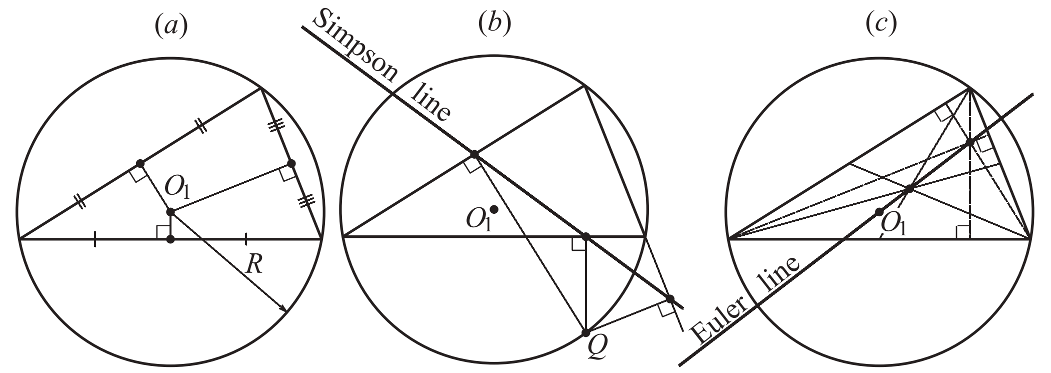

Figure M3.3. The circumcircle of a triangle. The circumcenter (*a*), the Simpson line (*b*), and the Euler line (*c*).

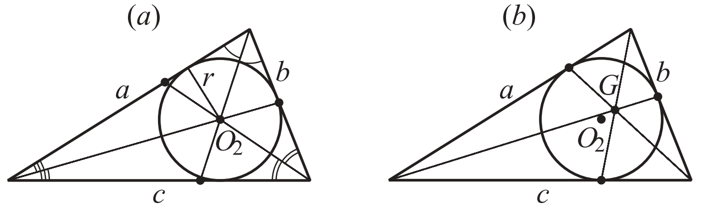

Figure M3.4. The incircle of a triangle (*a*). The incenter and the Gergonne point (*b*).

The inradius r and the circumradius R satisfy the relations

$$r = \sqrt{\frac{(p-a)(p-b)(p-c)}{p}} = p\tan\frac{\alpha}{2}\tan\frac{\beta}{2}\tan\frac{\gamma}{2} = (p-c)\tan\frac{\gamma}{2}, \quad \text{(M3.1.1.4)}$$

$$R = \frac{a}{2\sin\alpha} = \frac{b}{2\sin\beta} = \frac{c}{2\sin\gamma} = \frac{p}{4\cos\left(\frac{1}{2}\alpha\right)\cos\left(\frac{1}{2}\beta\right)\cos\left(\frac{1}{2}\gamma\right)}. \quad \text{(M3.1.1.5)}$$

The distance d between the circumcenter and the incenter is given by the expression

$$d = \sqrt{R^2 - 2Rr}. \quad \text{(M3.1.1.6)}$$

► **Area of a triangle.** The area S of a triangle is given by the formulas

$$S = \tfrac{1}{2}ah_a = \tfrac{1}{2}ab\sin\gamma = rp, \quad \text{(M3.1.1.7)}$$

$$S = \sqrt{p(p-a)(p-b)(p-c)} \quad (\textit{Heron's formula}), \quad \text{(M3.1.1.8)}$$

$$S = \frac{abc}{4R} = 2R^2\sin\alpha\sin\beta\sin\gamma, \quad \text{(M3.1.1.9)}$$

$$S = c^2\frac{\sin\alpha\sin\beta}{2\sin\gamma} = c^2\frac{\sin\alpha\sin\beta}{2\sin(\alpha+\beta)}. \quad \text{(M3.1.1.10)}$$

► **Right (right-angled) triangles.** A *right* triangle is a triangle with a right angle. The side opposite the right angle is called the *hypotenuse*, and the other two sides are called the *legs* (Fig. M3.5).

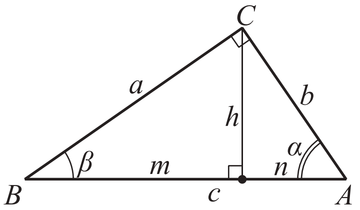

Figure M3.5. A right triangle.

The hypotenuse c, the legs a and b, and the angles α and β opposite the legs satisfy the following relations:

$$\begin{aligned}&\alpha+\beta=90^\circ;\\ &\sin\alpha=\cos\beta=\frac{a}{c},\quad \sin\beta=\cos\alpha=\frac{b}{c},\\ &\tan\alpha=\cot\beta=\frac{a}{b},\quad \tan\beta=\cot\alpha=\frac{b}{a}.\end{aligned}\tag{M3.1.1.11}$$

One also has

$$a^2+b^2=c^2\quad \text{(PYTHAGOREAN THEOREM)},\tag{M3.1.1.12}$$

$$h^2=mn,\quad a^2=mc,\quad b^2=nc,\tag{M3.1.1.13}$$

where h is the length of the altitude drawn to the hypotenuse; moreover, the altitude cuts the hypotenuse into segments of lengths m and n.

In a right triangle, the length of the median m_c drawn from the vertex of the right angle coincides with the circumradius R and is equal to half the length of the hypotenuse c, $m_c=R=\frac{1}{2}c$. The inradius is given by the formula $r=\frac{1}{2}(a+b-c)$. The area of the right triangle is $S=\frac{1}{2}ah_a=\frac{1}{2}ab$ (see also formulas (M3.1.1.4), (M3.1.1.5), and (M3.1.1.9)).

► Isosceles and equilateral triangles.

1°. An *isosceles* triangle is a triangle with two equal sides. These sides are called the *legs*, and the third side is called the *base* (Fig. M3.6*a*).

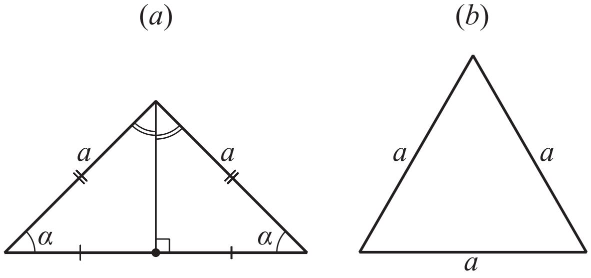

Figure M3.6. An isosceles triangle (*a*). An equilateral triangle (*b*).

Properties of isosceles triangles:

1. In an isosceles triangle, the angles adjacent to the base are equal.
2. In an isosceles triangle, the median drawn to the base is the angle bisector and the altitude.
3. In an isosceles triangle, the sum of distances from a point of the base to the legs is constant.

Criteria for a triangle to be isosceles:

1. If two angles in a triangle are equal, then the triangle is isosceles.
2. If a median in a triangle is also an altitude, then the triangle is isosceles.
3. If a bisector in a triangle is also an altitude, then the triangle is isosceles.

2°. An *equilateral* (or *regular*) triangle is a triangle with all three sides equal (Fig. M3.6*b*). All angles of an equilateral triangle are equal to 60°. In an equilateral triangle, the circumradius R and the inradius r satisfy the relation $R = 2r$.

For an equilateral triangle with side length a, the circumradius and the inradius are given by the formulas $R = \frac{\sqrt{3}}{3}a$ and $r = \frac{\sqrt{3}}{6}a$, and the area is equal to $S = \frac{\sqrt{3}}{4}a^2$.

M3.1.2. Polygons

► **Polygons. Basic information.** A *polygon* is a plane figure bounded by a closed broken line. The straight line segments forming a polygon are called its *sides* (or *edges*). The points at which two sides meet are called the *vertices* (or *corners*) of the polygon. Two sides sharing a vertex, as well as two successive vertices (the endpoints of the same edge), are said to be adjacent. A polygon is said to be *convex* if it lies on one side of any straight line passing through two neighboring vertices. In what follows, we consider only simple convex polygons.

An *(interior) angle* of a convex polygon is the angle between two sides meeting in a vertex. A convex polygon is said to be *inscribed* in a circle if all of its vertices lie on the circle. A polygon is said to be *circumscribed* about a circle if all of its sides are tangent to the circle.

For a convex polygon with n sides, the sum of interior angles is equal to $180°(n-2)$.

One can find the area of an arbitrary polygon by dividing it into triangles.

► **Properties of quadrilaterals.**

1. The diagonals of a convex quadrilateral meet.
2. The sum of interior angles of a convex quadrilateral equals 360° (Figs. M3.7*a* and *b*).
3. The lengths of the sides a, b, c, and d, the diagonals d_1 and d_2, and the segment m connecting the midpoints of the diagonals satisfy the relation $a^2 + b^2 + c^2 + d^2 = d_1^2 + d_2^2 + 4m^2$.
4. A convex quadrilateral is circumscribed if and only if $a + c = b + d$.
5. A convex quadrilateral is inscribed if and only if $\alpha + \gamma = \beta + \delta$.
6. The relation $ac + bd = d_1 d_2$ holds for inscribed quadrilaterals (PTOLEMY'S THEOREM).

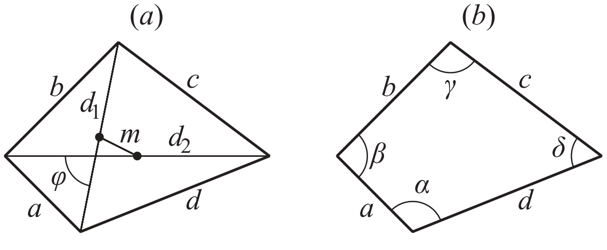

Figure M3.7. Quadrilaterals.

► **Areas of quadrilaterals.** The area of a convex quadrilateral is equal to

$$S = \frac{1}{2} d_1 d_2 \sin\varphi = \sqrt{p(p-a)(p-b)(p-c)(p-d) - abcd\cos^2\frac{\beta+\delta}{2}}, \qquad \text{(M3.1.2.1)}$$

where φ is the angle between the diagonals d_1 and d_2 and $p = \frac{1}{2}(a + b + c + d)$.

The area of an inscribed quadrilateral is

$$S = \sqrt{p(p-a)(p-b)(p-c)(p-d)}. \qquad \text{(M3.1.2.2)}$$

The area of a circumscribed quadrilateral is

$$S = \sqrt{abcd\sin^2\frac{\beta+\delta}{2}}. \qquad \text{(M3.1.2.3)}$$

► **Basic quadrilaterals.**

1°. A *parallelogram* is a quadrilateral such that both pairs of opposite sides are parallel (Fig. M3.8*a*).

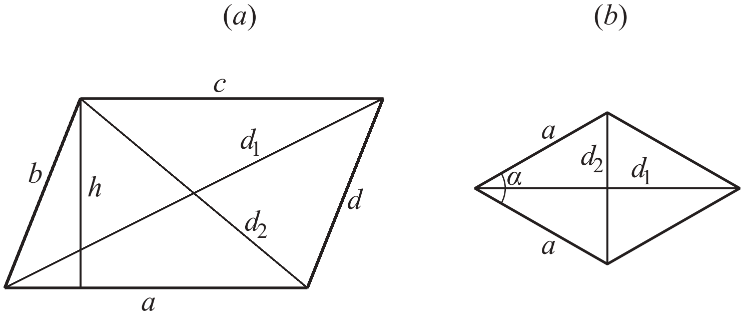

Figure M3.8. A parallelogram (*a*) and a rhombus (*b*).

Attributes of parallelograms (a quadrilateral is a parallelogram if):

1. Both pairs of opposite sides have equal length.
2. Both pairs of opposite angles are equal.
3. Two opposite sides are parallel and have equal length.

Properties of parallelograms:

1. The diagonals meet and bisect each other.
2. Opposite sides have equal length, and opposite angles are equal.
3. The diagonals and the sides satisfy the relation $d_1^2 + d_2^2 = 2(a^2 + b^2)$.
4. The area of a parallelogram is $S = ah$, where h is the altitude.

2°. A *rhombus* is a parallelogram in which all sides are of equal length (Fig. M3.8*b*).

Properties of rhombi:

1. The diagonals are perpendicular.
2. The diagonals are angle bisectors.
3. The area of a rhombus is $S = ah = a^2 \sin\alpha = \frac{1}{2}d_1 d_2$.

3°. A *rectangle* is a parallelogram in which all angles are right angles (Fig. M3.9*a*).

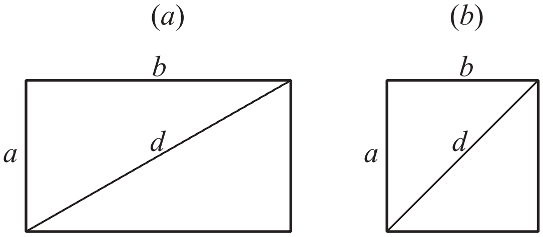

Figure M3.9. A rectangle (*a*) and a square (*b*).

Properties of rectangles:

1. The diagonals have equal lengths.
2. The area of a rectangle is $S = ab$.

4°. A *square* is a rectangle in which all sides have equal lengths (Fig. M3.9*b*). A square is also a rhombus with right angles.

Properties of squares:

1. All angles are right angles.
2. The diagonals are equal to $d = a\sqrt{2}$.
3. The diagonals meet at a right angle and are angle bisectors.
4. The area of a square is equal to $S = a^2 = \frac{1}{2}d^2$.

5°. A *trapezoid* is a quadrilateral in which two sides are parallel and the other two sides are nonparallel (Fig. M3.10). The parallel sides a and b are called the *bases* of the trapezoid, and the other two sides are called the *legs*. In an *isosceles* trapezoid, the legs are of equal length. The line segment connecting the midpoints of the legs is called the *median* of the trapezoid. The length of the median is equal to half the sum of the lengths of the bases, $m = \frac{1}{2}(a + b)$.

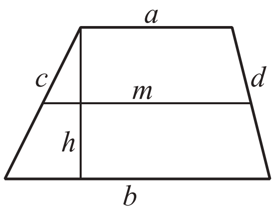

Figure M3.10. A trapezoid.

The perpendicular distance between the bases is called the *altitude* of a trapezoid.

Properties of trapezoids:

1. A trapezoid is circumscribed if and only if $a + b = c + d$.
2. A trapezoid is inscribed if and only if it is isosceles.
3. The area of a trapezoid is $S = \frac{1}{2}(a + b)h = mh = \frac{1}{2}d_1 d_2 \sin\varphi$, where φ is the angle between the diagonals d_1 and d_2.
4. The segment connecting the midpoints of the diagonals is parallel to the bases and has the length $\frac{1}{2}(b - a)$.

► **Regular polygons.** A convex polygon is said to be *regular* if all of its sides have the same length and all of its interior angles are equal. A convex n-gon is regular if and only if it is taken to itself by the rotation by an angle of $2\pi/n$ about some point O. The point O is called the *center* of the regular polygon. The angle between two rays issuing from the center and passing through two neighboring vertices is called the *central angle* (Fig. M3.11).

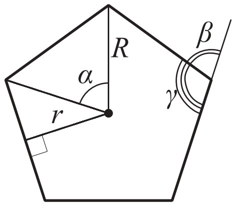

Figure M3.11. A regular polygon.

Properties of regular polygons:

1. The center is equidistant from all vertices as well as from all sides of a regular polygon.
2. A regular polygon is simultaneously inscribed and circumscribed; the centers of the circumcircle and the incircle coincide with the center of the polygon itself.
3. In a regular polygon, the central angle is $\alpha = 360°/n$, the external angle is $\beta = 360°/n$, and the interior angle is $\gamma = 180° - \beta$.
4. The circumradius R, the inradius r, and the side length a of a regular polygon satisfy the relations

$$a = 2\sqrt{R^2 - r^2} = 2R \sin\frac{\alpha}{2} = 2r \tan\frac{\alpha}{2}. \qquad \text{(M3.1.2.4)}$$

5. The area S of a regular n-gon is given by the formula

$$S = \frac{arn}{2} = nr^2 \tan\frac{\alpha}{2} = nR^2 \sin\frac{\alpha}{2} = \frac{1}{4}na^2 \cot\frac{\alpha}{2}. \qquad \text{(M3.1.2.5)}$$

Table M3.3 presents several useful formulas for regular polygons.

TABLE M3.3
Regular polygons (a is the side length).

No.	Name	Inradius r	Circumradius R	Area S
1	Regular polygon	$\dfrac{a}{2\tan\frac{\pi}{n}}$	$\dfrac{a}{2\sin\frac{\pi}{n}}$	$\dfrac{1}{2}arn$
2	Triangle	$\dfrac{\sqrt{3}}{6}a$	$\dfrac{\sqrt{3}}{3}a$	$\dfrac{\sqrt{3}}{4}a^2$
3	Square	$\dfrac{1}{2}a$	$\dfrac{1}{\sqrt{2}}a$	a^2
4	Pentagon	$\sqrt{\dfrac{5+2\sqrt{5}}{20}}\,a$	$\sqrt{\dfrac{5+\sqrt{5}}{10}}\,a$	$\dfrac{\sqrt{25+10\sqrt{5}}}{4}a^2$
5	Hexagon	$\dfrac{\sqrt{3}}{2}a$	a	$\dfrac{3\sqrt{3}}{2}a^2$
6	Octagon	$\dfrac{1+\sqrt{2}}{2}a$	$\dfrac{\sqrt{2+\sqrt{2}}}{2}a$	$2(1+\sqrt{2})a^2$
7	Enneagon	$\dfrac{5+2\sqrt{5}}{2}a$	$\dfrac{1+\sqrt{5}}{2}a$	$\dfrac{\sqrt{5+2\sqrt{5}}}{2}a^2$
8	Dodecagon	$\dfrac{2+\sqrt{3}}{2}a$	$\dfrac{3+\sqrt{3}}{\sqrt{6}}a$	$3(2+\sqrt{3})a^2$

M3.1.3. Circle

► **Some definitions and formulas.** A *circle* is the set of all points in the plane that are the same fixed distance R from a fixed point O (Fig. M3.12*a*). The distance R is called the *radius* of the circle and the point O is called its *center*. A plane figure bounded by a circle, including its interior, is called a *disk*. A segment connecting two points on a circle is called a *chord*. A chord passing through the center of a circle is called a *diameter* of the circle (Fig. M3.12*b*). The length of a diameter is $d = 2R$. A straight line that touches a circle at a single point is called a *tangent*, and the common point is called the *point of tangency* (Fig. M3.12*c*). A straight line that cuts a circle at two points, an extended chord, is called a

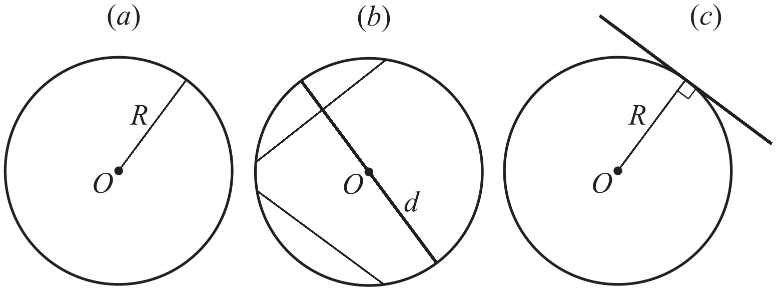

Figure M3.12. A circle (*a*), a diameter (*b*) and a tangent (*c*) of a circle.

secant. The angle formed by two radii is called a *central angle*. The angle formed by two chords with a common endpoint is called an *inscribed angle*.

Properties of circles and disks:

1. The circumference is $L = 2\pi R = \pi d = 2\sqrt{\pi S}$.
2. The area of a disk is $S = \pi R^2 = \frac{1}{4}\pi d^2 = \frac{1}{4}Ld$.
3. The diameter of a circle is a longest chord.
4. The diameter passing through the midpoint of a chord is perpendicular to the chord.
5. The radius drawn to the point of tangency is perpendicular to the tangent.
6. An inscribed angle is half the central angle subtended by the same chord, $\alpha = \frac{1}{2}\angle BOC$ (Fig. M3.13*a*).

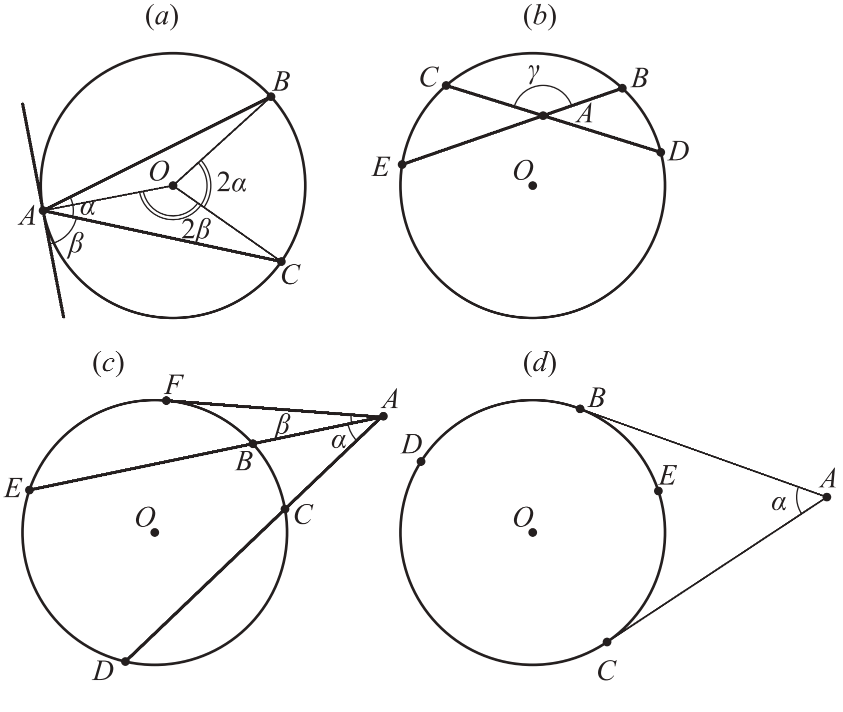

Figure M3.13. Properties of circles and disks.

7. The angle between a chord, AC, and the tangent to the circle at an endpoint, A, of the chord is equal to $\beta = \frac{1}{2}\angle AOC$ (Fig. M3.13*a*).
8. The angle between two chords, BE and CD, is $\gamma = \frac{1}{2}(\overset{\smile}{BC} + \overset{\smile}{ED})$ (Fig. M3.13*b*).

9. The angle between two secants, AD and AE, is $\alpha = \frac{1}{2}(\smile{DE} - \smile{BC})$ (Fig. M3.13*c*).
10. The angle between a secant, AE, and the tangent, AF, to the circle at a point, F, is equal to $\beta = \frac{1}{2}(\smile{FE} - \smile{BF})$ (Fig. M3.13*c*).
11. The angle between two tangents, AB and AC, is $\alpha = \frac{1}{2}(\smile{BDC} - \smile{BEC})$ (Fig. M3.13*d*).
12. If two chords, BE and CD, meet, then $AC \cdot AD = AB \cdot AE = R^2 - m^2$, $m = OA$ (Fig. M3.13*b*).
13. For secants, AE and AD, the relations $AC\cdot AD = AB\cdot AE = m^2 - R^2$ hold (Fig. M3.13*c*).
14. For a tangent, AF, and a secant, AD, the relation $AF^2 = AC \cdot AD$ holds (Fig. M3.13*c*).

► **Segment and sector.** A plane figure bounded by two radii and one of the subtended arcs is called a (*circular*) *sector*. A plane figure bounded by an arc and the corresponding chord is called a *segment* (Fig. M3.14*a*). If R is the radius of the circle, l is the arc length, a is the chord length, α is the central angle (in degrees), and h is the height of the segment, then the following relations hold:

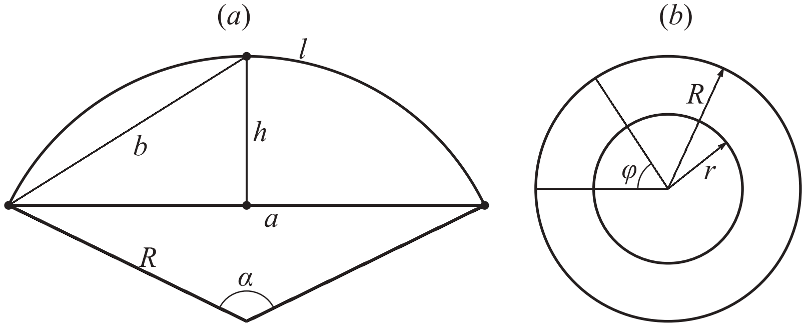

Figure M3.14. A segment (*a*) and an annulus (*b*).

$$
\begin{aligned}
a &= 2\sqrt{2hR - h^2} = 2R\sin\frac{\alpha}{2},\\
h &= R - \sqrt{R^2 - \frac{a^2}{4}} = R\left(1 - \cos\frac{\alpha}{2}\right) = \frac{a}{2}\tan\frac{\alpha}{4},\\
l &= \frac{2\pi R\alpha}{360} \approx 0.01745\, R\alpha.
\end{aligned}
\qquad \text{(M3.1.3.1)}
$$

The area of a circular sector is given by the formula

$$S = \frac{lR}{2} = \frac{\pi R^2\alpha}{360} \approx 0.00873\, R^2\alpha, \qquad \text{(M3.1.3.2)}$$

and the area of a segment not equal to a half-disk is given by

$$S_1 = \frac{\pi R^2\alpha}{360} \pm S_\Delta, \qquad \text{(M3.1.3.3)}$$

where S_Δ is the area of the triangle with vertices at the center of the disk and at the endpoints of the radii bounding the corresponding sector. One takes the minus sign for $\alpha < 180$ and the plus sign for $\alpha > 180$.

The arc length and the area of a segment can be found by the approximate formulas

$$\begin{aligned} &l \approx \frac{8b-a}{3}, \quad l \approx \sqrt{a^2 + \frac{16h^2}{3}},\\ &S_1 \approx \frac{h(6a+8b)}{15}, \end{aligned} \tag{M3.1.3.4}$$

where b is the chord of the half-segment (see Fig. M3.14a).

► **Annulus.** An *annulus* is a plane figure bounded by two concentric circles of distinct radii (Fig. M3.14b). Let R be the outer radius of an annulus (the radius of the outer bounding circle), and let r be the inner radius (the radius of the inner bounding circle). Then the area of the annulus is given by the formula

$$S = \pi(R^2 - r^2) = \frac{\pi}{4}(D^2 - d^2) = 2\pi\rho\delta, \tag{M3.1.3.5}$$

where $D = 2R$ and $d = 2r$ are the outer and inner diameters, $\rho = \frac{1}{2}(R+r)$ is the midradius, and $\delta = R - r$ is the width of the annulus.

The area of the part of the annulus contained in a sector of central angle φ, given in degrees (see Fig. M3.14b), is given by the formula

$$S = \frac{\pi\varphi}{360}(R^2 - r^2) = \frac{\pi\varphi}{1440}(D^2 - d^2) = \frac{\pi\varphi}{180}\rho\delta. \tag{M3.1.3.6}$$

M3.2. Solid Geometry

M3.2.1. Straight Lines, Planes, and Angles in Space

► **Mutual arrangement of straight lines and planes.**

1°. Two distinct straight lines lying in a single plane either have exactly one point of intersection or do not meet at all. In the latter case, they are said to be *parallel*. If two straight lines do not lie in a single plane, then they are called *skew lines*.

The angle between skew lines is determined as the angle between lines parallel to them and lying in a single plane (Fig. M3.15a). The distance between skew lines is the length of the straight line segment that meets both lines and is perpendicular to them.

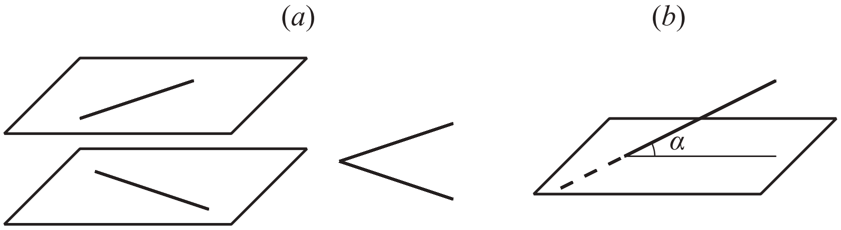

Figure M3.15. The angle between skew lines (a). The angle between a line and a plane (b).

2°. Two distinct planes either intersect in a straight line or do not have common points. In the latter case, they are said to be *parallel*. Coinciding planes are also assumed to be parallel. If two planes are perpendicular to a single straight line or each of them contains a pair of intersecting straight lines parallel to the corresponding lines in the other pair, then the planes are parallel.

3°. A straight line either entirely lies in the plane, meets the plane at a single point, or has no common points with the plane. In the last case, the line is said to be *parallel* to the plane.

The angle between a straight line and a plane is equal to the angle between the line and its projection onto the plane (Fig. M3.15b). If a straight line is perpendicular to two intersecting straight lines on a plane, then it is perpendicular to each line on the plane, i.e., *perpendicular to the plane*.

► Polyhedral angles.

1°. A *dihedral angle* is a figure in space formed by two half-planes issuing from a single straight line as well as the part of space bounded by these half-planes. The half-planes are called the *faces* of the dihedral angle, and their common straight line is called the *edge*. A dihedral angle is measured by its linear angle ABC (Fig. M3.16a), i.e., by the angle between the perpendiculars raised to the edge DE of the dihedral angle in both planes (*faces*) at the same point.

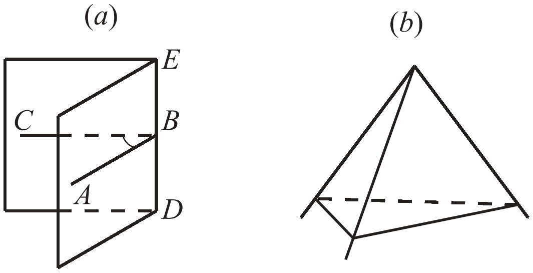

Figure M3.16. A dihedral (a) and a trihedral (b) angle.

2°. A part of space bounded by an infinite triangular pyramid is called a *trihedral angle* (Fig. M3.16b). The faces of this pyramid are called the *faces* of the trihedral angle, and the vertex of the pyramid is called the *vertex of a trihedral angle*. The rays in which the faces intersect are called the *edges* of a trihedral angle. The edges form *face angles*, and the faces form the dihedral angles of the trihedral angle. As a rule, one considers trihedral angles with dihedral angles less than π (or 180°), i.e., convex trihedral angles. Each face angle of a convex trihedral angle is less than the sum of the other two face angles and greater than their difference.

Two trihedral angles are equal if one of the following conditions is satisfied:

1. Two face angles, together with the included dihedral angle, of the first trihedral angle are equal to the respective parts (arranged in the same order) of the second trihedral angle.
2. Two dihedral angles, together with the included face angle, of the first trihedral angle are equal to the respective parts (arranged in the same order) of the second trihedral angle.
3. The three face angles of the first trihedral angle are equal to the respective face angles (arranged in the same order) of the second trihedral angle.
4. The three dihedral angles of the first trihedral angle are equal to the respective dihedral angles (arranged in the same order) of the second trihedral angle.

3°. A *polyhedral angle* $OABCDE$ (Fig. M3.17a) is formed by several planes (*faces*) having a single common point (the *vertex*) and successively intersecting along straight lines OA, OB, $\ldots$, OE (the *edges*). Two edges belonging to the same face form a *face angle* of the polyhedral angle, and two neighboring faces form a *dihedral angle*.

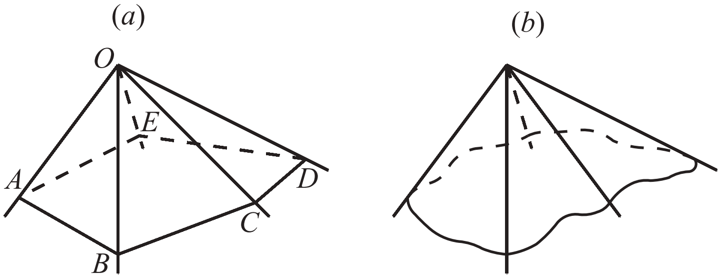

Figure M3.17. A polyhedral (*a*) and a solid (*b*) angle.

Polyhedral angles are equal (*congruent*) if one can be transformed into the other by translations and rotations. For polyhedral angles to be congruent, the corresponding parts (face and dihedral angles) must be equal.

A convex polyhedral angle lies entirely on one side of each of its faces. The sum $\angle AOB + \angle BOC + \cdots + \angle EOA$ of face angles (Fig. M3.17*a*) of any convex polyhedral angle is less that 2π (or 360°).

4°. A *solid angle* is a part of space bounded by straight lines issuing from a single point (vertex) to all points of some closed curve (Fig. M3.17*b*). Trihedral and polyhedral angles are special cases of solid angles. A solid angle is measured by the area cut by the solid angle on the sphere of unit radius centered at the vertex. Solid angles are measured in steradians. The entire sphere forms a solid angle of 4π steradians.

M3.2.2. Polyhedra

► **General concepts.** A *polyhedron* is a closed object formed by intersecting planes. In other words, a polyhedron is a set of finitely many plane polygons satisfying the following conditions:

1. Each side of each polygon is simultaneously a side of a unique other polygon, which is said to be adjacent to the first polygon (via this side).
2. From each of the polygons forming a polyhedron, one can reach any other polygon by successively passing to adjacent polygons.

These polygons are called the *faces*, their sides are called the *edges*, and their vertices are called the *vertices* of a polyhedron.

A polyhedron is said to be *convex* if it lies entirely on one side of the plane of any of its faces; if a polyhedron is convex, then so are its faces.

EULER'S THEOREM. *If the number of vertices in a convex polyhedron is v, the number of edges is e, and the number of faces is f, then $v + f - e = 2$.*

► **Prism. Parallelepiped.**

1°. An n-sided *prism* is a polyhedron in which two faces are equal n-gons (the *base faces*) that lie on parallel planes and have respectively parallel sides, and the remaining n faces (*joining* or *lateral faces*) are parallelograms; see Fig. M3.18*a*. A *right* prism is a prism in which the lateral faces are perpendicular to the base faces; otherwise it is an *oblique* prism. A right prism is said to be *regular* if its base faces are regular polygons.

If l is the lateral edge length, S is the area of the base face, H is the height of the prism (perpendicular distance between the planes of the bases), P_{sec} is the perimeter of a normal

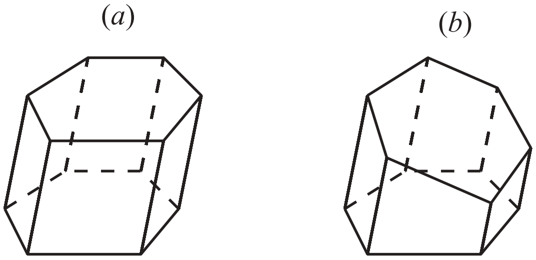

Figure M3.18. A prism (*a*) and a truncated prism (*b*).

section, one perpendicular to a lateral edge, and $S_{\rm sec}$ is the area of a normal section, then the area of the lateral surface $S_{\rm lat}$ and the volume V of the prism are given by the formulas

$$\begin{aligned} S_{\rm lat} &= P_{\rm sec}l \\ V &= SH = S_{\rm sec}l. \end{aligned} \tag{M3.2.2.1}$$

The portion of a prism between one of the bases and a plane nonparallel to it is called a *truncated prism* (Fig. M3.18*b*). The volume of a truncated prism is

$$V = LS_1, \tag{M3.2.2.2}$$

where L is the length of the segment connecting the centroids of the base faces and S_1 is the area of the section of the prism by a plane perpendicular to this segment.

The volume of a truncated regular prism (its base being a regular n-gon) is expressed as

$$V = S_{\rm sec}\frac{l_1 + l_2 + \cdots + l_n}{n},$$

where $S_{\rm sec}$ is the area of a normal section and l_1, l_2, …, l_n are the lengths of the lateral edges.

2°. A prism whose bases are parallelograms is called a *parallelepiped*. All four diagonals in a parallelepiped intersect at a single point and bisect each other (Fig. M3.19*a*). A parallelepiped is said to be *rectangular* if it is a right prism and its base faces are rectangles. In a rectangular parallelepiped, all diagonals are equal (Fig. M3.19*b*).

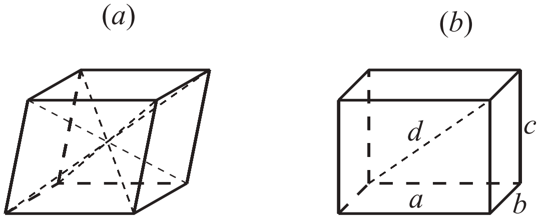

Figure M3.19. A parallelepiped (*a*) and a rectangular parallelepiped (*b*).

If a, b, and c are the lengths of the edges of a rectangular parallelepiped, then the diagonal d can be determined by the formula $d^2 = a^2 + b^2 + c^2$. The volume of a rectangular parallelepiped is given by the formula $V = abc$, and the lateral surface area is $S_{\rm lat} = PH$, where P is the perimeter of the base face.

3°. A rectangular parallelepiped all of whose edges are equal ($a = b = c$) is called a *cube*. The diagonal of a cube is given by the formula $d^2 = 3a^2$. The volume of the cube is $V = a^3$, and the lateral surface area is $S_{\rm lat} = 4a^2$.

▶ Pyramid, obelisk, and wedge.

1°. A *pyramid* is a polyhedron in which one face (the *base* of the pyramid) is an arbitrary polygon and the other (*lateral*) faces are triangles with a common vertex, called the *apex* of the pyramid (Fig. M3.20*a*). The base of an n*-sided* pyramid is an n-gon. The perpendicular through the apex to the base of a pyramid is called the *altitude* (*height*) of the pyramid.

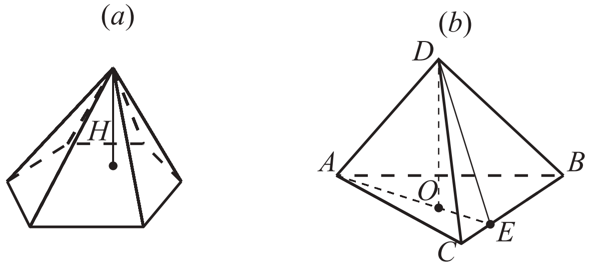

Figure M3.20. A pyramid (*a*). The altitude DO, the plane DAE, and the side BC in a triangular pyramid (*b*).

The volume of a pyramid is given by the formula

$$V = \frac{1}{3}SH, \qquad \text{(M3.2.2.3)}$$

where S is the area of the base and H is the altitude of the pyramid.

If DO is the altitude of the pyramid $ABCD$ and $DA \perp BC$, then the plane DAE is perpendicular to BC (Fig. M3.20*b*).

If the pyramid is cut by a plane (Fig. M3.21*a*) parallel to the base, then

$$\begin{gathered}\frac{SA_1}{A_1A} = \frac{SB_1}{B_1B} = \cdots = \frac{SO_1}{O_1O},\\ \frac{S_{ABCDEF}}{S_{A_1B_1C_1D_1E_1F_1}} = \left(\frac{SO}{SO_1}\right)^2,\end{gathered} \qquad \text{(M3.2.2.4)}$$

where SO is the altitude of the pyramid.

The altitude of a *triangular pyramid* passes through the orthocenter of its base if and only if all pairs of opposite edges of the pyramid are perpendicular.

Given the length of the edges, $DA = a$, $DB = b$, $DC = c$, $BC = p$, $AC = q$, and $AB = r$, of a triangular pyramid (Fig. M3.21*b*), its volume can be found from the relation

$$V^2 = \frac{1}{288}\begin{vmatrix} 0 & r^2 & q^2 & a^2 & 1\\ r^2 & 0 & p^2 & b^2 & 1\\ q^2 & p^2 & 0 & c^2 & 1\\ a^2 & b^2 & c^2 & 0 & 1\\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}, \qquad \text{(M3.2.2.5)}$$

where the right-hand side contains a determinant.

A pyramid is said to be *regular* if its base is a regular n-gon and the altitude passes through the center of the base. The altitude (dropped from the apex) of a lateral face is called the *apothem* of a regular pyramid. For a regular pyramid, the lateral surface area is

$$S_{\text{lat}} = \frac{1}{2}Pl, \qquad \text{(M3.2.2.6)}$$

where P is the perimeter of the base and l is the apothem.

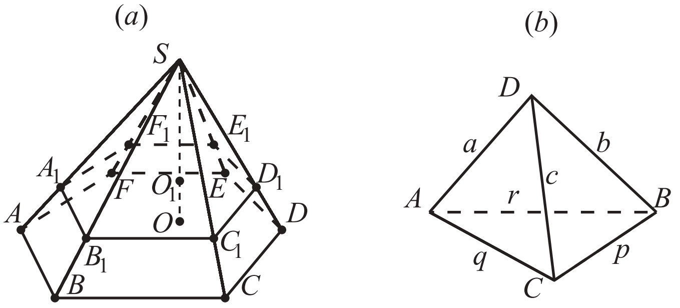

Figure M3.21. The original pyramid and a pyramid cut off by a plane (*a*). A triangular pyramid (*b*).

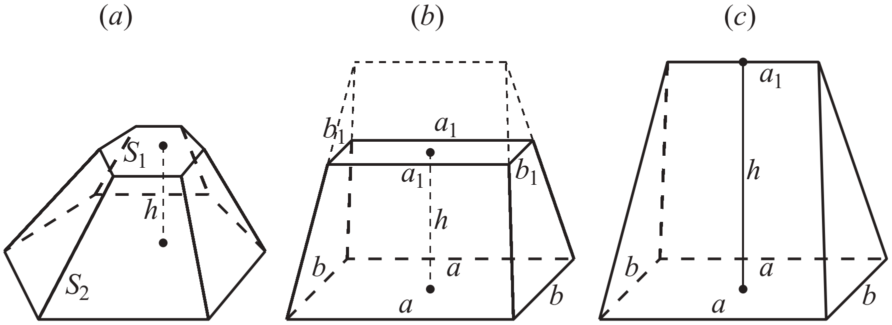

Figure M3.22. A frustum of a pyramid (*a*), an obelisk (*b*), and a wedge (*c*).

2°. If a pyramid is cut by a plane parallel to the base, then it splits into two parts: a pyramid similar to the original pyramid and the *frustum* (Fig. M3.22*a*). The volume of the frustum is

$$V = \frac{1}{3}h(S_1 + S_2 + \sqrt{S_1S_2}\,) = \frac{1}{3}hS_2\left(1 + \frac{a}{A} + \frac{a^2}{A^2}\right), \qquad \text{(M3.2.2.7)}$$

where S_1 and S_2 are the areas of the bases, a and A are two respective sides of the bases, and h is the altitude (the perpendicular distance between the bases).

For a regular frustum, the lateral surface area is

$$S_{\text{lat}} = \frac{1}{2}(P_1 + P_2)l, \qquad \text{(M3.2.2.8)}$$

where P_1 and P_2 are the perimeters of the bases and l is the altitude of the lateral face.

3°. A hexahedron whose bases are rectangles lying in parallel planes and whose lateral faces form equal angles with the base, but do not meet at a single point, is called an *obelisk* (Fig. M3.22*b*). If a, b and a_1, b_1 are the sides of the bases and h is the altitude, then the volume of the hexahedron is

$$V = \frac{h}{6}[(2a + a_1)b + (2a_1 + a)b_1]. \qquad \text{(M3.2.2.9)}$$

4°. A pentahedron whose base is a rectangle and whose lateral faces are isosceles triangles and isosceles trapezoids is called a *wedge* (Fig. M3.22*c*). The volume of the wedge is

$$V = \frac{h}{6}(2a + a_1)b. \qquad \text{(M3.2.2.10)}$$

► **Regular polyhedra.** A polyhedron is said to be *regular* if all of its faces are equal regular polygons and all polyhedral angles are equal to each other. There exist five regular polyhedra (Fig. M3.23), whose properties are given in Table M3.4.

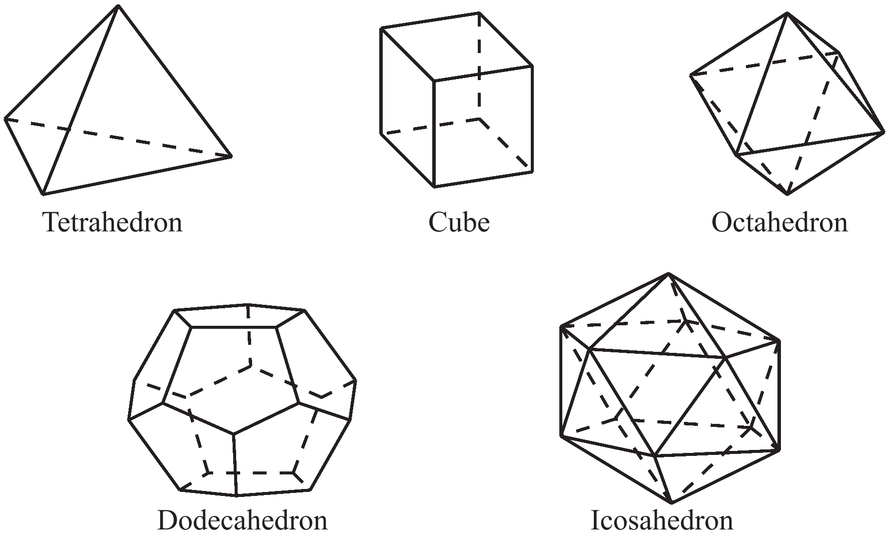

Figure M3.23. Five regular polyhedra.

TABLE M3.4
Regular polyhedra (a is the edge length).

No.	Name	Number of faces and their shapes	Number of vertices	Number of edges	Total surface area	Volume
1	Tetrahedron	4 triangles	4	6	$a^2\sqrt{3}$	$\frac{a^3\sqrt{2}}{12}$
2	Cube	6 squares	8	12	$6a^2$	a^3
3	Octahedron	8 triangles	6	12	$2a^2\sqrt{3}$	$\frac{a^3\sqrt{2}}{3}$
4	Dodecahedron	12 pentagons	20	30	$3a^2\sqrt{25+10\sqrt{5}}$	$\frac{a^3}{4}(15+7\sqrt{5})$
5	Icosahedron	20 triangles	12	30	$3a^2\sqrt{3}$	$\frac{5a^3}{12}(3+\sqrt{5})$

M3.2.3. Solids Formed by Revolution of Lines

► **Cylinder.** A *cylindrical surface* is a surface in space swept by a straight line (the *generator*) moving parallel to a given direction along some curve (the *directrix*) (Fig. M3.24*a*).

1°. A solid bounded by a closed cylindrical surface and two planes is called a *cylinder*; the planes are called the *bases* of the cylinder (Fig. M3.24*b*).

If P is the perimeter of the base, P_{sec} is the perimeter of the section perpendicular to the generator, S_{sec} is the area of this section, S_{bas} is the area of the base, and l is the length of the generator, then the lateral surface area S_{lat} and the volume V of the cylinder are given by the formulas

$$\begin{aligned} S_{\text{lat}} &= PH = P_{\text{sec}}l, \\ V &= S_{\text{bas}}H = S_{\text{sec}}l. \end{aligned} \qquad \text{(M3.2.3.1)}$$

In a *right* cylinder, the bases are perpendicular to the generator. In particular, if the bases are disks, then one speaks of a *right circular cylinder*. The volume, the lateral surface

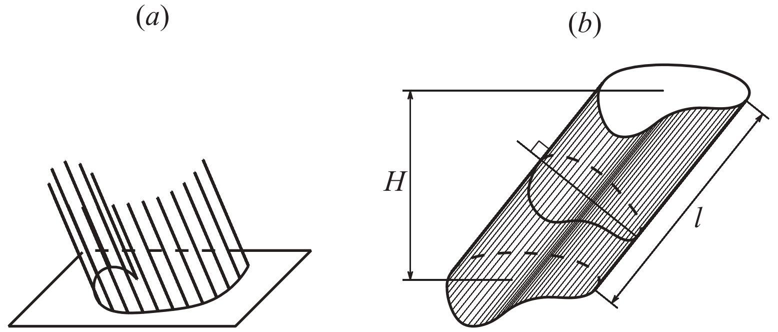

Figure M3.24. A cylindrical surface (*a*). A cylinder (*b*).

area, and the total surface area of a right circular cylinder are given by the formulas

$$\begin{aligned} V &= \pi R^2 H, \\ S_{\text{lat}} &= 2\pi R H, \\ S &= 2\pi R(R+H), \end{aligned} \tag{M3.2.3.2}$$

where R is the radius of the base.

A right circular cylinder is also called a *round cylinder*, or simply a *cylinder*.

2°. The part of a cylinder cut by a plane nonparallel to the base is called a *truncated cylinder* (Fig. M3.25*a*).

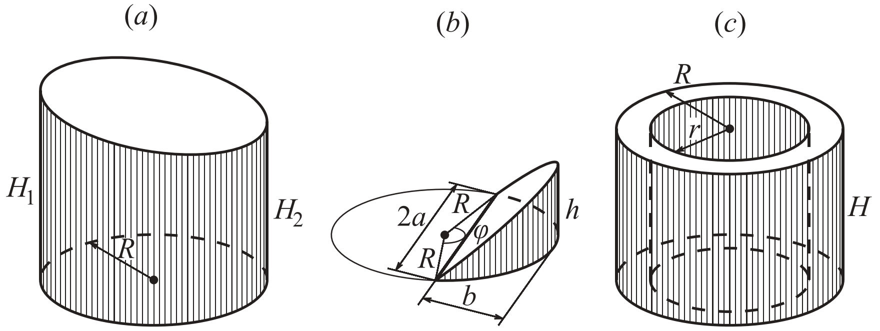

Figure M3.25. A truncated cylinder (*a*), a “hoof” (*b*), and a cylindrical tube (*c*).

The volume, the lateral surface area, and the total surface area of a truncated cylinder are given by the formulas

$$\begin{aligned} V &= \pi R^2 \frac{H_1+H_2}{2}, \\ S_{\text{lat}} &= \pi R(H_1+H_2), \\ S &= \pi R\left[H_1+H_2+R+\sqrt{R^2+\left(\frac{H_2-H_1}{2}\right)^2}\,\right], \end{aligned} \tag{M3.2.3.3}$$

where H_1 and H_2 are the maximal and minimal generators.

3°. A *segment of a round cylinder (a "hoof")* is a portion of the cylinder cut by a plane that is nonparallel to the base and intersects it. If R is the radius of the cylindrical segment, h is the height of the "hoof," and b is its width (for the other notation, see Fig. M3.25*b*), then the volume V and the lateral surface area S_{lat} of the "hoof" can be determined by the formulas

$$\begin{aligned} V &= \frac{h}{3b}\big[a(3R^2-a^2)+3R^2(b-R)\alpha\big] = \frac{hR^3}{b}\left(\sin\alpha-\frac{\sin^3\alpha}{3}-\alpha\cos\alpha\right), \\ S_{\text{lat}} &= \frac{2\pi R}{b}[(b-R)\alpha+a], \end{aligned} \tag{M3.2.3.4}$$

where $\alpha=\frac{1}{2}\varphi$ is measured in radians.

4°. A solid bounded by two closed cylindrical surfaces and two planes is called a *cylindrical tube*; the planes are called the bases of the tube. The volume of a round cylindrical tube (Fig. M3.25*c*) is

$$V=\pi H(R^2-r^2)=\pi H\delta(2R-r)=\pi H\delta(2r+\delta)=2\pi H\delta\rho, \tag{M3.2.3.5}$$

where R and r are the outer and inner radii, $\delta=R-r$ is the thickness, $\rho=\frac{1}{2}(R+r)$ is the midradius, and H is the height of the pipe.

► **Conical surface. Cone. Frustum of cone.** A *conical surface* is the union of straight lines (*generators*) passing through a fixed point (the *apex*) in space and any point of some space curve (the *directrix*) (Fig. M3.26*a*).

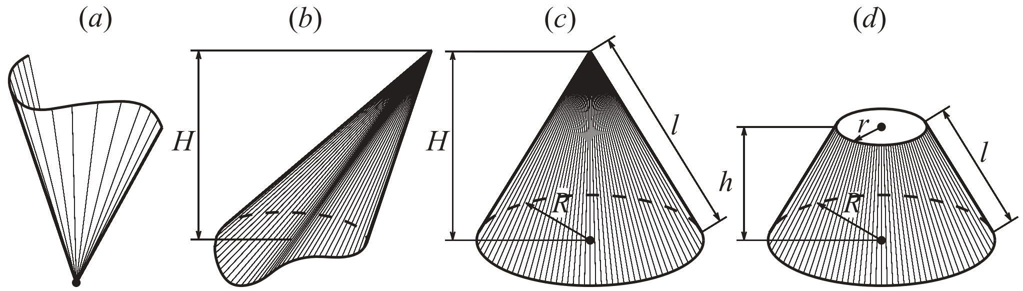

Figure M3.26. Conical surface (*a*). A cone (*b*), a right circular cone (*c*), and a frustum of a cone (*d*).

1°. A solid bounded by a conical surface with closed directrix and a plane is called a *cone*; the plane is the base of the cone (Fig. M3.26*b*). The volume of an arbitrary cone is given by the formula

$$V=\frac{1}{3}HS_{\text{bas}}, \tag{M3.2.3.6}$$

where H is the altitude of the cone and S_{bas} is the area of the base.

A *right circular cone* (Fig. M3.26*c*) has a disk as the base, and its vertex is projected onto the center of the disk. If l is the length of the generator and R is the radius of the base, then the volume, the lateral surface area, and the total surface area of the right circular cone are given by the formulas

$$\begin{aligned} V &= \frac{1}{3}\pi R^2 H, \\ S_{\text{lat}} &= \pi R l = \pi R\sqrt{R^2+H^2}, \\ S &= \pi R(R+l). \end{aligned} \tag{M3.2.3.7}$$

2°. If a cone is cut by a plane parallel to the base, then we obtain a *frustum of a cone* (Fig. M3.26*d*). The length l of the generator, the volume V, the lateral surface area S_{lat}, and the total surface area S of the frustum of a right circular cone are given by the formulas

$$\begin{aligned} l &= \sqrt{h^2 + (R-r)^2}, \\ V &= \frac{\pi h}{3}(R^2 + r^2 + Rr), \\ S_{\text{lat}} &= \pi l(R+r), \\ S &= \pi[l(R+r) + R^2 + r^2], \end{aligned} \qquad \text{(M3.2.3.8)}$$

where r is the radius of the upper base and h is the altitude of the frustum of a cone.

► Sphere. Spherical parts. The torus.

1°. A *sphere* is the set of all points in space that are the same distance R from a fixed point O (Fig. M3.27*a*). The distance R is called the *radius* of the sphere and the point O is called its *center*. A straight line segment that passes through the center of a sphere and whose endpoints are on the sphere is called a *diameter* of the sphere. A solid formed by a sphere together with its interior is called a *ball*. Any section of the sphere by a plane is a circle. The section of the sphere by a plane passing through its center is called a *great circle* of radius R. There exists exactly one great circle passing through two arbitrary points on the sphere that are not antipodal (i.e., are not the opposite endpoints of a diameter); the smaller arc of this great circle is the shortest distance on the sphere between these points. The surface area S of a sphere and the volume V of the ball bounded by the sphere are given by

$$\begin{aligned} S &= 4\pi R^2 = \pi D^2 = \sqrt[3]{36\pi V^2}, \\ V &= \frac{4\pi R^3}{3} = \frac{\pi D^3}{6} = \frac{1}{6}\sqrt{\frac{S^3}{\pi}}, \end{aligned} \qquad \text{(M3.2.3.9)}$$

where $D = 2R$ is the diameter of the sphere.

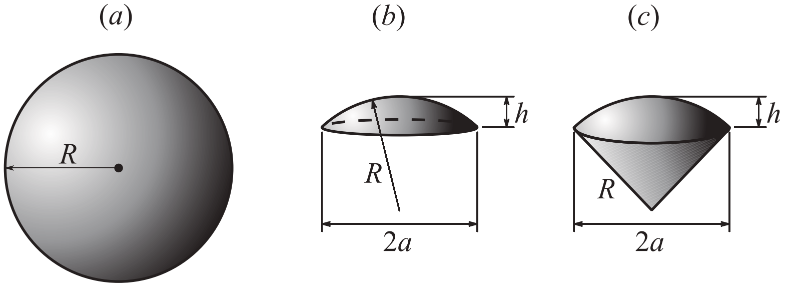

Figure M3.27. A sphere (*a*), a spherical cap (*b*), and a spherical sector (*c*).

2°. A portion of a ball cut from it by a plane is called a *spherical cap* (Fig. M3.27*b*). The width a (base radius), the area S_{lat} of the curved surface, the total surface area S, and the volume V of a spherical cap can be found from the formulas

$$\begin{aligned} a^2 &= h(2R-h), \\ S_{\text{lat}} &= 2\pi Rh = \pi(a^2 + h^2), \\ S &= S_{\text{lat}} + \pi a^2 = \pi(2Rh + a^2) = \pi(h^2 + 2a^2), \\ V &= \frac{\pi h}{6}(3a^2 + h^2) = \frac{\pi h^2}{3}(3R-h), \end{aligned} \qquad \text{(M3.2.3.10)}$$

where R and h are the radius and the height of the spherical cap.

3°. A portion of a ball bounded by the curved surface of a spherical cap and the conical surface whose base is the base of the cap and whose vertex is the center of the ball is called a *spherical sector* (Fig. M3.27*c*). The total surface area S and the volume V of a spherical sector are given by the formulas

$$\begin{aligned} S &= \pi R(2h + a), \\ V &= \frac{2}{3}\pi R^2 h, \end{aligned} \tag{M3.2.3.11}$$

where a is the width of the spherical cap, h is its height, and R is the radius of the sector.

4°. A portion of a ball contained between two parallel plane secants is called a *spherical segment* (Fig. M3.28*a*). The curved surface of a spherical segment is called a *spherical zone*, and the plane circular surfaces are the *bases* of a spherical segment. The radius R of the ball, the radii a and b of the bases, and the height h of a spherical segment satisfy the relation

$$R^2 = a^2 + \left(\frac{a^2 - b^2 - h^2}{2h}\right)^2. \tag{M3.2.3.12}$$

The curved surface area S_{lat}, the total surface area S, and the volume V of a spherical segment are given by the formulas

$$\begin{aligned} S_{\text{lat}} &= 2\pi Rh, \\ S &= S_{\text{lat}} + \pi(a^2 + b^2) = \pi(2Rh + a^2 + b^2), \\ V &= \frac{\pi h}{6}(3a^2 + 3b^2 + h^2). \end{aligned} \tag{M3.2.3.13}$$

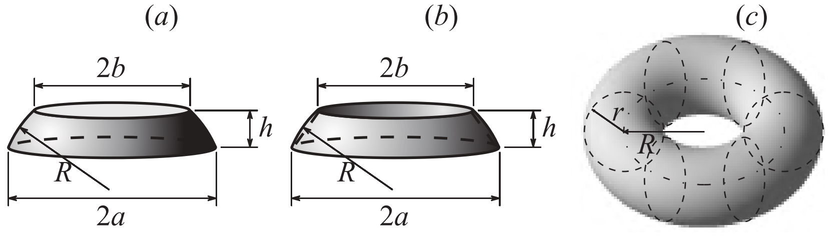

Figure M3.28. A spherical segment (*a*) and a spherical segment without the truncated cone inscribed in it (*b*). A torus (*c*).

If V_1 is the volume of the truncated cone inscribed in a spherical segment (Fig. M3.28*b*) and l is the length of its generator, then $V - V_1 = \frac{\pi h l^2}{6}$.

5°. A *torus* is a surface generated by revolving a circle about an axis coplanar with the circle but not intersecting it (Fig. M3.28*c*). If r is the radius of the circle being rotated and R is the distance from its center to the axis of revolution ($R > r$), then the surface area and the volume of the torus are given by

$$\begin{aligned} S &= 4\pi^2 Rr = \pi^2 Dd, \\ V &= 2\pi^2 Rr^2 = \frac{\pi^2 Dd^2}{4}, \end{aligned}$$

where $d = 2r$ and $D = 2R$ are the diameters of the generating circle and the circle of revolution.

Bibliography for Chapter M3

Alexander, D. C. and Koeberlein, G. M., *Elementary Geometry for College Students, 3rd Edition*, Houghton Mifflin Company, Boston, Massachusetts, 2002.

Gustafson, R. D. and Frisk, P. D., *Elementary Geometry, 3rd Edition*, Wiley, New York, 1991.

Moise, E., *Elementary Geometry from an Advanced Standpoint, 3rd Edition*, Addison Wesley, Boston, Massachusetts, 1990.

Musser, G. L. and Trimpe, L. E., *College Geometry: A Problem Solving Approach with Applications*, Prentice Hall, Englewood Cliffs, New Jersey, 1994.

Schultze, A. and Sevenoak, F. L., *Plane and Solid Geometry*, Adamant Media Corporation, Boston, Massachusetts, 2004.

Tussy, A. S. and Gustafson, R. D., *Basic Geometry for College Students: An Overview of the Fundamental Concepts of Geometry*, Thomson-Brooks/Cole, Pacific Grove, California, 2002.

Chapter M4

Analytic Geometry

M4.1. Points, Segments, and Coordinate Plane

M4.1.1. Cartesian and Polar Coordinates on Plane

► **Rectangular Cartesian coordinates in the plane.** A *rectangular Cartesian coordinate system* consists of two mutually perpendicular directed lines, called *coordinate axes*, each treated as a number line (see Subsection M6.1.1). The point of intersection of the axes is called the *origin* and usually labeled with the letter O, while the axes themselves are called the *coordinate axes*. As a rule, one of the coordinate axes is horizontal, directed from left to right, and called the *abscissa axis*. The other axis is vertical, directed upwards, and called the *ordinate axis*. The two axes are usually denoted by X or OX and Y or OY, respectively, and the coordinate system itself is denoted by XY or OXY. The two coordinate axes divide the plane into four parts, which are called *quadrants* and numbered I, II, III, and IV counterclockwise as shown in Fig. M4.1.

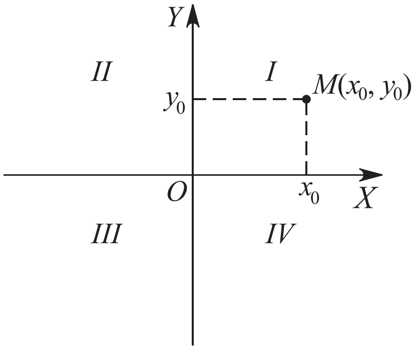

Figure M4.1. A rectangular Cartesian coordinate system.

Each point M in the plane is uniquely defined by a pair of real numbers $(x_0,\ y_0)$, called its *coordinates*, which specify its projections onto the X- and Y-axes. The numbers x_0 and y_0 are called, respectively, the *abscissa* and the *ordinate* of the point M.

Remark. Strictly speaking, the coordinate system introduced above is a *right rectangular Cartesian coordinate system*. A *left rectangular Cartesian coordinate system* can, for example, be obtained by changing the direction of one of the axes. A right rectangular Cartesian coordinate system is usually called simply a Cartesian coordinate system.

If A and B are two points in the plane, then the length of the segment AB will be denoted $|AB|$.

► **Polar coordinates.** A *polar coordinate system* is determined by a point O called the *pole*, a ray OA issuing from this point, which is called the *polar axis*, a scale segment for measuring lengths, and the positive sense of rotation around the pole. Usually, the counterclockwise sense is assumed to be positive (see Fig. M4.2*a*).

The position of each point M on the plane is determined by two *polar coordinates*, the *polar radius* $\rho = |OM|$ and the *polar angle* $\theta = \angle AOM$ (the values of the angle θ are defined up to an additive term $2\pi n$, where n is an integer). To be definite, one usually assumes that $0 \le \theta \le 2\pi$ or $-\pi \le \theta \le \pi$. The polar radius of the pole is zero, and its polar angle does not have any definite value.

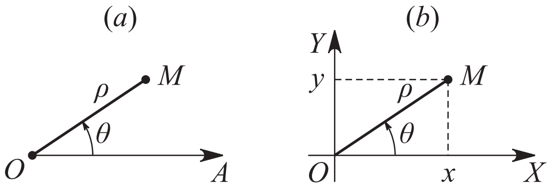

Figure M4.2. A polar coordinate system (*a*). Relationship between Cartesian and polar coordinates (*b*).

► **Relationship between Cartesian and polar coordinates.** Suppose that M is an arbitrary point in the plane, (x, y) are its rectangular Cartesian coordinates, and (ρ, θ) are its polar coordinates (see Fig. M4.2*b*). The transformation from one coordinate system to the other is expressed by the formulas

$$\begin{aligned} x &= \rho\cos\theta, \\ y &= \rho\sin\theta \end{aligned} \qquad \text{or} \qquad \begin{aligned} \rho &= \sqrt{x^2+y^2}, \\ \tan\theta &= y/x, \end{aligned} \tag{M4.1.1.1}$$

where the polar angle θ is determined with regard to the quadrant where the point M lies.

Example. Let us find the polar coordinates ρ, θ $(0 \le \theta \le 2\pi)$ of the point M whose Cartesian coordinates are $x = -3$, $y = -3$.

From formulas (M4.1.1.1), we obtain $\rho = \sqrt{(-3)^2 + (-3)^2} = 3\sqrt{2}$ and $\tan\theta = \frac{-3}{-3} = 1$. Since the point M lies in the third quadrant, we have $\theta = \arctan 1 + \pi = \frac{5}{4}\pi$.

M4.1.2. Distance Between Points. Division of Segment in Given Ratio. Area of a Polygon

► **Distance between points on plane.** The distance d between two arbitrary points $A_1(x_1, y_1)$ and $A_2(x_2, y_2)$ on the plane is given by the formula

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$

where x and y with the respective subscripts are the Cartesian coordinates of these points, and by the formula

$$d = \sqrt{\rho_1^2 + \rho_2^2 - 2\rho_1\rho_2\cos(\theta_2 - \theta_1)},$$

where ρ and θ with the respective subscripts are the polar coordinates of these points.

► **Angles between segments.** The angle β between arbitrary segments A_1A_2 and A_3A_4 joining the points $A_1(x_1, y_1)$, $A_2(x_2, y_2)$ and $A_3(x_3, y_3)$, $A_4(x_4, y_4)$, respectively, can be found from the relation

$$\cos\beta = \frac{(x_2 - x_1)(x_4 - x_3) + (y_2 - y_1)(y_4 - y_3)}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}\sqrt{(x_4 - x_3)^2 + (y_4 - y_3)^2}}.$$

► **Division of a segment in a given ratio.** Given two points, $A(x_1, y_1)$ and $B(x_2, y_2)$, and a number λ, the coordinates of the point $M(x, y)$ dividing the segment AB in the ratio $\lambda = |AM| : |MB|$ are expressed as

$$x = \frac{x_1 + \lambda x_2}{1 + \lambda}, \qquad y = \frac{y_1 + \lambda y_2}{1 + \lambda}. \tag{M4.1.2.1}$$

Example. Find the coordinates of the midpoint of a segment AB.

The midpoint of the segment corresponds to $\lambda = 1$. Substituting this value into (M4.1.2.1) gives $x = \dfrac{x_1 + x_2}{2}$, $y = \dfrac{y_1 + y_2}{2}$.

► **Area of a triangle.** The area S_3 of the triangle with vertices A_1, A_2, and A_3 is given by the formula

$$\begin{aligned} \pm S_3 &= \frac{1}{2}[(x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1)] \\ &= \frac{1}{2}\begin{vmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_1 & y_3 - y_1 \end{vmatrix} = \frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}, \end{aligned} \tag{M4.1.2.2}$$

where x and y with respective subscripts are the Cartesian coordinates of the vertices, and by the formula

$$\pm S_3 = \frac{1}{2}[\rho_1\rho_2 \sin(\theta_2 - \theta_1) + \rho_2\rho_3 \sin(\theta_3 - \theta_2) + \rho_3\rho_1 \sin(\theta_1 - \theta_3)], \tag{M4.1.2.3}$$

where ρ and θ with respective subscripts are the polar coordinates of the vertices. In formulas (M4.1.2.2) and (M4.1.2.3), one takes the plus sign if the vertices are numbered counterclockwise (see Fig. M4.3*a*) and the minus sign otherwise.

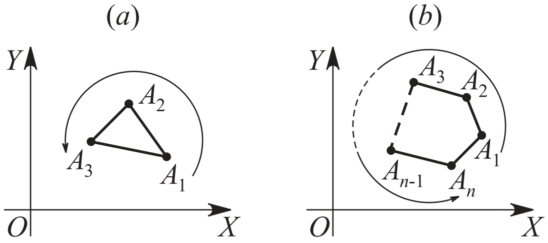

Figure M4.3. Area of a triangle (*a*) and of a polygon (*b*).

► **Area of a polygon.** The area S_n of the polygon with vertices $A_1, \ldots, A_n$ is given by the formula

$$\pm S_n = \frac{1}{2}[(x_1 - x_2)(y_1 + y_2) + (x_2 - x_3)(y_2 + y_3) + \cdots + (x_n - x_1)(y_n + y_1)], \tag{M4.1.2.4}$$

where x and y with respective subscripts are the Cartesian coordinates of the vertices, and by the formula

$$\pm S_n = \frac{1}{2}[\rho_1\rho_2 \sin(\theta_2 - \theta_1) + \rho_2\rho_3 \sin(\theta_3 - \theta_2) + \cdots + \rho_n\rho_1 \sin(\theta_1 - \theta_n)], \tag{M4.1.2.5}$$

where ρ and θ with respective subscripts are the polar coordinates of the vertices. In formulas (M4.1.2.4) and (M4.1.2.5), one takes the plus sign if the vertices are numbered counterclockwise (see Fig. M4.3*b*) and the minus sign otherwise.

Remark. One often says that formulas (M4.1.2.2)–(M4.1.2.5) express the *oriented area* of the corresponding figures.

M4.2. Straight Lines and Points on Plane

M4.2.1. Equations of Straight Lines on Plane

► **Equation of a curve.** Given a coordinate system, the set of all points in the plane can be treated as the set of various pairs of numbers x, y. Relations imposed upon x and y define subsets of the plane.

A line in the plane is usually defined using an equation relating the Cartesian coordinates x and y. An equation

$$F(x,y)=0$$

is an *equation of a curve in the plane* if the coordinates of all the points lying on the curve satisfy the equation and the coordinates of all those points that do not lie on the curve do not satisfy it.

Example 1. Derive an equation of the line all of whose points are equidistant from the points $A(0,2)$ and $B(4,-2)$.

Let $M(x,y)$ be a point that belongs to the line. For the distances to A and B, we have

$$\rho(A,M)=\sqrt{x^2+(y-2)^2},\qquad \rho(B,M)=\sqrt{(x-4)^2+(y+2)^2}.$$

It follows from the relation $\rho(A,M)=\rho(B,M)$ that $x^2+(y-2)^2=(x-4)^2+(y+2)^2$. Expanding and collecting similar terms yields $y=x-2$, which is an equation that determines a straight line.

Parametric equations of a curve on the plane have the form

$$x=\varphi(t),\qquad y=\psi(t),$$

where x and y are treated as the coordinates of some point M for each value of the *variable parameter* t.

► **Slope-intercept equation of a straight line.** The *slope-intercept equation of a straight line* in the rectangular Cartesian coordinate system OXY has the form

$$y=kx+b,\tag{M4.2.1.1}$$

where $k=\tan\varphi$ is the slope of the line and b is the y-intercept of the line, i.e., the signed distance from the point of intersection of the line with the ordinate axis to the origin. Equation (M4.2.1.1) is meaningful for any straight line that is not perpendicular to the abscissa axis (see Fig. M4.4*a*).

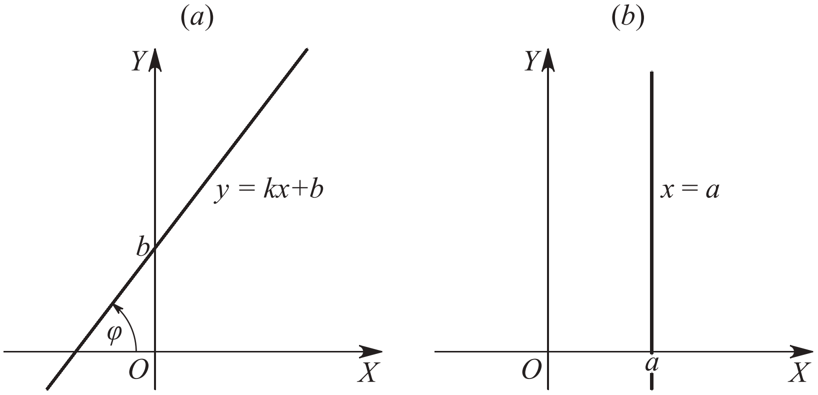

Figure M4.4. Straight lines on plane.

If a straight line is not perpendicular to the OX-axis, then its equation can be written as (M4.2.1.1), but if a straight line is perpendicular to the OX-axis, then its equation can be written as

$$x = a, \tag{M4.2.1.2}$$

where a is the abscissa of the point of intersection of this line with the OX-axis (see Fig. M4.4*b*).

For the slope of a straight line, we also have the formula

$$k = \frac{y_2 - y_1}{x_2 - x_1}, \tag{M4.2.1.3}$$

where $A_1(x_1, y_1)$ and $A_2(x_2, y_2)$ are two arbitrary points of the line.

► **Point-slope equation of a straight line.** In the rectangular Cartesian coordinate system OXY, the equation of a straight line with slope k passing through a point $A(x_1, y_1)$ has the form

$$y - y_1 = k(x - x_1). \tag{M4.2.1.4}$$

If we set $x_1 = 0$ and $y_1 = b$ in equation (M4.2.1.4), then we obtain equation (M4.2.1.1).

► **Equation of a straight line passing through two given points.** The equation of a straight line passing through two distinct points $A_1(x_1, y_1)$ and $A_2(x_2, y_2)$ has the form

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} \qquad (x_1 \neq x_2,\ y_1 \neq y_2). \tag{M4.2.1.5}$$

If $x_1 = x_2$, this equation degenerates into $x = x_1$. If $y_1 = y_2$, the equation becomes $y = y_1$.

Example 2. Let us derive the equation of the straight line passing through the points $A_1(5, 1)$ and $A_2(7, 3)$. Substituting the coordinates of these points into formula (M4.2.1.5), we obtain

$$\frac{x - 5}{2} = \frac{y - 1}{2} \qquad \text{or} \qquad y = x - 4.$$

► **General equation of a straight line.** A linear equation of the form

$$Ax + By + C = 0 \qquad (A^2 + B^2 \neq 0) \tag{M4.2.1.6}$$

is called the *general equation of a straight line* in the rectangular Cartesian coordinate system OXY. In rectangular Cartesian coordinates, each straight line is determined by an equation of degree 1, and, conversely, each equation of degree 1 determines a straight line.

If $B \neq 0$, then equation (M4.2.1.6) can be written as (M4.2.1.1), where $k = -A/B$ and $b = -C/B$.

Special cases of equation (M4.2.1.6):

1. If $A = 0$ and $B \neq 0$, then the equation becomes $y = -C/B$ and determines a straight line parallel to the axis OX.
2. If $B = 0$ and $A \neq 0$, then the equation becomes $x = -C/A$ and determines a straight line parallel to the axis OY.
3. If $C = 0$, then the equation becomes $Ax + By = 0$ and determines a straight line passing through the origin.

► **General equation of a straight line passing through a given point.** In the rectangular Cartesian coordinate system OXY, the general equation of a straight line passing through a point $M(x_1, y_1)$ in the plane has the form

$$A(x - x_1) + B(y - y_1) = 0. \tag{M4.2.1.7}$$

► **Parametric equations of a straight line.** In the rectangular Cartesian coordinate system OXY, a straight line passing through a point $M(x_1, y_1)$ in the plane can be represented by the parametric equations

$$x = x_1 + At, \quad y = y_1 + Bt, \tag{M4.2.1.8}$$

where A and B are constants and t is a variable parameter.

A straight line passing through two points, $M_1(x_1, y_1)$ and $M_2(x_2, y_2)$, can be represented by the parametric equations

$$\begin{aligned} x &= x_1(1-t) + x_2 t, \\ y &= y_1(1-t) + y_2 t. \end{aligned} \tag{M4.2.1.9}$$

► **Intercept-intercept equation of a straight line.** The general equation of a straight line can be rewritten in the form

$$\frac{x}{a} + \frac{y}{b} = 1,$$

which is called the *intercept-intercept equation of a straight line*. The numbers a and b are the x- and y-intercepts of the straight line, i.e., the signed distances from the origin to the points at which the straight line crosses the coordinate axes (see Fig. M4.5).

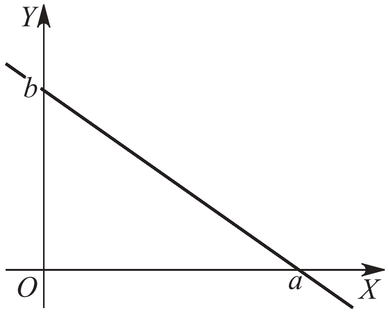

Figure M4.5. A straight line with intercept-intercept equation.

► **Equation of a pencil of straight lines.** The set of all straight lines passing through a fixed point M in the plane is called a *pencil of straight lines*, and the point M itself is called the *center of the pencil*. The equation determining all straight lines in the pencil is called the *equation of the pencil*.

1°. Given the Cartesian coordinates of the pencil center $M(x_1, y_1)$, then the equation of any straight line in the pencil has the form (M4.2.1.7), where A and B are arbitrary constants.

2°. If the equations of two straight lines in the pencil are known, $A_1x + B_1y + C_1 = 0$ and $A_2x + B_2y + C_2 = 0$, then the equation of the pencil can be written as

$$\alpha(A_1x + B_1y + C_1) + \beta(A_2x + B_2y + C_2) = 0,$$

where α and β are any numbers that are not simultaneously zero.

M4.2.2. Mutual Arrangement of Points and Straight Lines

► **Condition for three points to be collinear.** Suppose there are three distinct points, $M_1(x_1, y_1)$, $M_2(x_2, y_2)$, and $M_3(x_3, y_3)$, given in the Cartesian coordinate system OXY on the plane. They are collinear (lie on the same straight line) if and only if

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = x_1y_2 + x_2y_3 + x_3y_1 - x_1y_3 - x_2y_1 - x_3y_2 = 0.$$

This condition reflects the fact that the area of the triangle with vertices at the above points is zero; cf. Eq. (M4.1.2.2).

► **Distance from a point to a straight line.** The distance d from a point $M(x_0, y_0)$ to a straight line given by the general equation $Ax + By + C = 0$ can be calculated as

$$d = \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}}. \tag{M4.2.2.1}$$

Example 1. Let us find the distance from the point $M(2, 1)$ to the straight line $3x - 4y + 8 = 0$. We use formula (M4.2.2.1) to obtain

$$d = \frac{|3 \cdot 2 - 4 \cdot 1 + 8|}{\sqrt{3^2 + 4^2}} = \frac{10}{5} = 2.$$

► **Angle between two straight lines.**

1°. If two straight lines are given by the equations

$$y = k_1 x + b_1,$$
$$y = k_2 x + b_2,$$

where $k_1 = \tan\varphi_1$ and $k_2 = \tan\varphi_2$ are the slopes of the respective lines (see Fig. M4.6), the angle α between these straight lines is determined by the formula

$$\tan\alpha = \frac{k_2 - k_1}{1 + k_1 k_2} \qquad (k_1 k_2 \neq -1).$$

If $k_1 = k_2$, the straight lines are *parallel* ($\alpha = 0$).

If $k_1 k_2 = -1$, the straight lines are *perpendicular* ($\alpha = \frac{1}{2}\pi$).

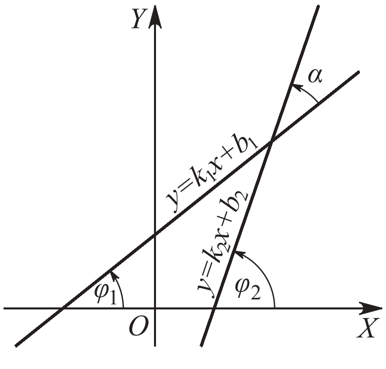

Figure M4.6. Angle between two straight lines.

Example 2. Given a triangle with vertices $A(-2, 0)$, $B(2, 4)$, and $C(4, 0)$, derive the equations of the side BC and the altitude AH.

Using (M4.2.1.5), one finds an equation for the side BC: $\frac{x-2}{4-2} = \frac{y-4}{0-4}$, or $y = -2x + 8$. It follows that the slope of this straight line is $k_{BC} = -2$. The above condition for two straight lines to be perpendicular gives $k_{AH} = -\frac{1}{k_{BC}} = \frac{1}{2}$. Using equation (M4.2.1.4), one obtains the equation for the altitude AH: $y - 0 = \frac{1}{2}(x + 2)$ or $y = \frac{1}{2}x + 1$.

2°. If two straight lines are defined by the general equations

$$\begin{aligned} A_1x + B_1y + C_1 &= 0,\\ A_2x + B_2y + C_2 &= 0, \end{aligned} \tag{M4.2.2.2}$$

the angle α between them can be calculated from

$$\tan\alpha = \frac{A_1B_2 - A_2B_1}{A_1A_2 + B_1B_2} \qquad (A_1A_2 + B_1B_2 \neq 0).$$

If $A_1B_2 - A_2B_1 = 0$ (or $A_1/A_2 = B_1/B_2$), the straight lines are *parallel*.
If $A_1A_2 + B_1B_2 = 0$, the straight lines are *perpendicular*.

► **Point of intersection of straight lines.** Suppose that two straight lines are defined by general equations in the form (M4.2.2.2). Each common solution of equations (M4.2.2.2) determines a common point of the two lines.

If the determinant of system (M4.2.2.2) is not zero, i.e.,

$$\begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix} = A_1B_2 - A_2B_1 \neq 0,$$

then the system is consistent and has a unique solution; hence, these straight lines are distinct and nonparallel and meet at the point $A(x_0, y_0)$, where

$$x_0 = \frac{B_1C_2 - B_2C_1}{A_1B_2 - A_2B_1}, \quad y_0 = \frac{C_1A_2 - C_2A_1}{A_1B_2 - A_2B_1}.$$

► **Distance between parallel lines.** The distance between the parallel lines given by equations

$$A_1x + B_1y + C_1 = 0 \quad \text{and} \quad A_1x + B_1y + C_2 = 0$$

can be found using the formula

$$d = \frac{|C_1 - C_2|}{\sqrt{A_1^2 + B_1^2}}.$$

M4.3. Quadratic Curves

M4.3.1. Circle

► **Equations of a circle in the Cartesian coordinate system.** The *canonical equation of a circle* in a rectangular Cartesian coordinate system OXY has the form

$$x^2 + y^2 = a^2, \tag{M4.3.1.1}$$

where the point $O(0, 0)$ is the center of the circle and $a > 0$ is its radius (see Fig. M4.7*a*). The circle defined by equation (M4.3.1.1) is the locus of points equidistant (lying at the distance a) from its center.

The circle with radius a and center $A(x_0, y_0)$ is defined by the equation

$$(x - x_0)^2 + (y - y_0)^2 = a^2. \tag{M4.3.1.2}$$

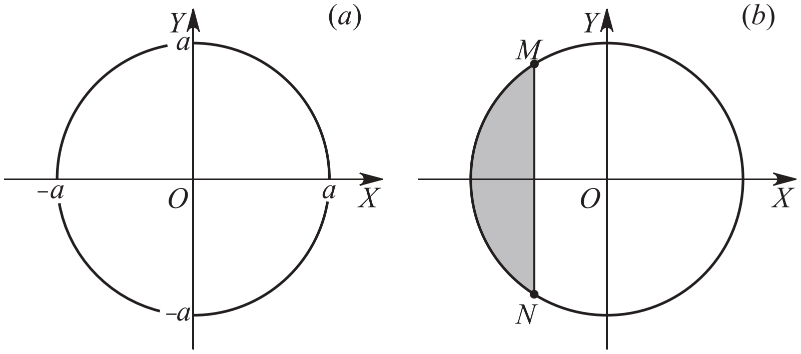

Figure M4.7. Circle.

The circle that passes through three noncollinear points $A_1(x_1,y_1)$, $A_2(x_2,y_2)$, and $A_3(x_3,y_3)$ can be described by the determinant equation

$$\begin{vmatrix} x^2+y^2 & x & y & 1 \\ x_1^2+y_1^2 & x_1 & y_1 & 1 \\ x_2^2+y_2^2 & x_2 & y_2 & 1 \\ x_3^2+y_3^2 & x_3 & y_3 & 1 \end{vmatrix} = 0.$$

The area of the *disk* bounded by a circle of radius a is given by the formula $S=\pi a^2$. The circumference of this circle is $L=2\pi a$. The area of the figure bounded by the circle and a chord with endpoints $M(x_0,y_0)$ and $N(x_0,-y_0)$, shaded in Fig. 4.7*b*, is expressed as

$$S=\frac{\pi a^2}{2}+x_0\sqrt{a^2-x_0^2}+a^2\arcsin\frac{x_0}{a}.$$

See also Subsection M3.1.3.

► **Other equations of a circle.** The equation of the circle (M4.3.1.1) can be represented in parametric form as

$$x=a\cos\theta,\quad y=a\sin\theta,$$

where the polar angle θ plays the role of the variable parameter.

In the polar coordinate system, the equation of the circle (M4.3.1.1) becomes

$$\rho=a.$$

Note that it does not contain the polar angle θ.

M4.3.2. Ellipse

► **Definition and the canonical equation of an ellipse.** An *ellipse* is the locus of points in the plane the sum of whose distances to two points, F_1 and F_2, is a constant quantity, denoted $2a$; see Fig. M4.8 *a*. Either of the points F_1 and F_2 is called a *focus* of the ellipse and the distance between them, $\rho(F_1,F_2)=2c$, is called the *focal distance*.

In the rectangular Cartesian coordinate system where the X-axis is the straight line passing through the foci, the origin O coincides with the midpoint of the segment F_1F_2, and the Y-axis passes through O and is perpendicular to the X-axis, as shown in Fig. M4.8 *a*, the equation of the ellipse has the simplest form

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1. \qquad \text{(M4.3.2.1)}$$

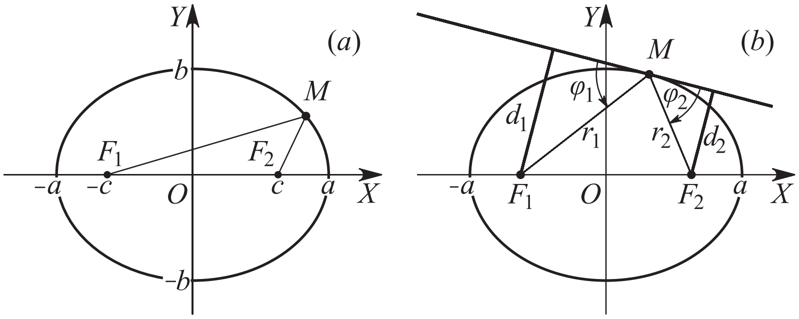

Figure M4.8. Ellipse (*a*). Tangent to an ellipse and the optical property of an ellipse (*b*).

Equation (M4.3.2.1) is called the *canonical equation of the ellipse*. Here, it is assumed that $a \ge b > 0$.

The positive numbers a and b are called, respectively, the *semimajor axis* and *semiminor axis* of the ellipse, with $b = \sqrt{a^2 - c^2}$. The number $c = \sqrt{a^2 - b^2}$ is called the *linear eccentricity* of the ellipse. The number $e = c/a = \sqrt{1 - b^2/a^2}$ is called the *eccentricity* or the *numerical eccentricity* of the ellipse, with $0 \le e < 1$. The number $p = b^2/a$ is called the *focal parameter* (or simply the *parameter*) of the ellipse.

The point $O(0,0)$ is called the *center* of the ellipse. The points of intersection of the ellipse with the axes of symmetry, $A_1(-a,0)$, $A_2(a,0)$ and $B_1(0,-b)$, $B_2(0,b)$, are called its *vertices*. The straight line passing through the foci of an ellipse is known as its major axis and is sometimes called its *focal axis*. Either of the straight lines $x = \pm a/e$ $(e \ne 0)$ is called a *directrix* of the ellipse. The focus $F_2(c,0)$ and the directrix $x = a/e$ are said to be *right*, and the focus $F_1(-c,0)$ and the directrix $x = -a/e$ are said to be *left*.

Remark. For $a = b$ $(c = 0)$, equation (M4.3.2.1) becomes $x^2 + y^2 = a^2$ and determines a circle.

The area of the figure bounded by the ellipse is given by the formula $S = \pi ab$. The length of the ellipse can be calculated approximately by the formula $L \approx \pi\bigl[1.5\,(a+b) - \sqrt{ab}\,\bigr]$.

► **Focal and focus-directrix properties of an ellipse.** The segments joining a point $M(x,y)$ of an ellipse with the foci $F_1(-c,0)$ and $F_2(c,0)$ are called the *left* and *right focal radii* of this point. We denote the lengths of the left and right focal radii by $r_1 = |F_1M|$ and $r_2 = |F_2M|$, respectively (see Fig. M4.8 *b*). By the definition of an ellipse,

$$r_1 + r_2 = 2a,$$

where r_1 and r_2 satisfy the relations

$$r_1 = \sqrt{(x+c)^2 + y^2} = a + ex, \qquad r_2 = \sqrt{(x-c)^2 + y^2} = a - ex.$$

The ellipse determined by equation (M4.3.2.1) on the plane is the locus of points for which the ratio of distances to a focus and the like directrix is equal to e:

$$r_1\left|x + \frac{a}{e}\right|^{-1} = e, \qquad r_2\left|x - \frac{a}{e}\right|^{-1} = e.$$

(A focus and a directrix are said to be *like* if both of them are right or left simultaneously.)

► **Equation of a tangent and the optical property of an ellipse.** The tangent to the ellipse (M4.3.2.1) at an arbitrary point $M_0(x_0, y_0)$ is given by the equation

$$\frac{x_0 x}{a^2} + \frac{y_0 y}{b^2} = 1.$$

The distances d_1 and d_2 from the foci $F_1(-c,0)$ and $F_2(c,0)$ to the tangent to the ellipse at the point $M_0(x_0,y_0)$ are expressed as

$$d_1 = \frac{1}{aN}|x_0e + a| = \frac{r_1}{aN}, \qquad N = \sqrt{\left(\frac{x_0}{a^2}\right)^2 + \left(\frac{y_0}{b^2}\right)^2},$$
$$d_2 = \frac{1}{aN}|x_0e - a| = \frac{r_2}{aN},$$

where $r_1 = r_1(M_0)$ and $r_2 = r_2(M_0)$ are the lengths of the focal radii of M_0 (see Fig. M4.8 *b*, where $M = M_0$).

The tangent at an arbitrary point $M_0(x_0,y_0)$ of an ellipse forms acute angles φ_1 and φ_2 with the focal radii of the point of tangency, and

$$\sin\varphi_1 = \frac{d_1}{r_1} = \frac{1}{aN}, \qquad \sin\varphi_2 = \frac{d_2}{r_2} = \frac{1}{aN}.$$

This fact, written as

$$\varphi_1 = \varphi_2,$$

is known as the *optical property of an ellipse*. It means that a light ray issued from one focus of the ellipse will reflect to the other focus (see Fig. M4.8 *b*, where $M = M_0$).

► **Equations of an ellipse in polar coordinates and parametric equations.** In polar coordinates (ρ,φ), with the pole coinciding with the right focus and the polar axis directed along the X-axis, the equation of an ellipse has the form

$$\rho = \frac{p}{1 + e\cos\varphi},$$

where $0 \le \varphi \le 2\pi$, $p = b^2/a$, and $e = \sqrt{1 - b^2/a^2}$. If the pole is taken at the left focus, this equation becomes

$$\rho = \frac{p}{1 - e\cos\varphi}.$$

The equation of an ellipse can also be represented in the parametric form

$$x = a\cos t, \quad y = b\sin t,$$

with the parameter t assuming any values from 0 to 2π.

M4.3.3. Hyperbola

► **Definition and the canonical equation of a hyperbola.** A *hyperbola* is the locus of points in the plane the absolute difference of whose distances to two points, F_1 and F_2, is a constant quantity, denoted $2a$; see Fig. M4.9 *a*. Either of the points F_1 and F_2 is called a *focus* of the hyperbola and the distance between them, $\rho(F_1,F_2) = 2c$, is called the *focal distance*.

In the rectangular Cartesian coordinate system where the X-axis is the straight line passing through the foci, the origin O coincides with the midpoint of the segment F_1F_2, and the Y-axis passes through O and is perpendicular to the X-axis, as shown in Fig. M4.8 *a*, the equation of the hyperbola has the simplest form

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \qquad \text{(M4.3.3.1)}$$

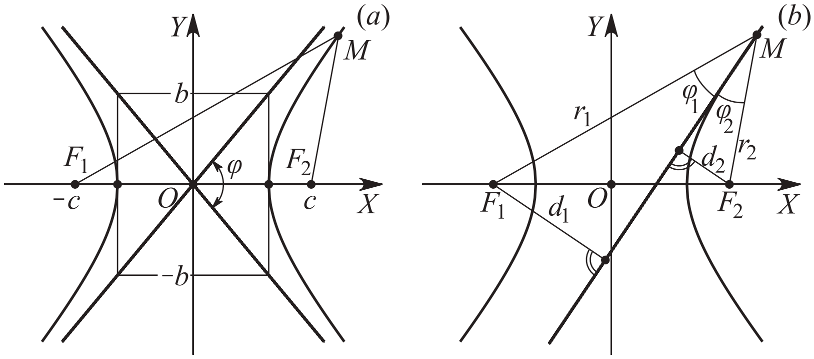

Figure M4.9. Hyperbola (*a*). The tangent to the hyperbola and optical property of a hyperbola (*b*).

and is called the *canonical equation of the hyperbola*.

The number a is called the *real semiaxis* of the hyperbola and the number b is called its *imaginary semiaxis*, with $b = \sqrt{c^2 - a^2}$. The number $c = \sqrt{a^2 + b^2}$ is known as the *linear eccentricity* of the hyperbola and the number $e = c/a = \sqrt{1 + b^2/a^2}$ is its *eccentricity*, $e > 1$. The number $p = b^2/a$ is called the *focal parameter* (or simply the *parameter*) of the hyperbola.

The point $O(0,0)$ is called the *center* of the hyperbola. The points $(-a,0)$ and $(a,0)$ at which the hyperbola crosses the X-axis are known as the *vertices* of the hyperbola. Either of the straight lines $x = \pm a/e$ is called a *directrix* of the hyperbola. The focus $F_2(c,0)$ and the directrix $x = a/e$ are said to be *right*, while the focus $F_1(-c,0)$ and the directrix $x = -a/e$ are said to be *left*.

A hyperbola consists of two parts, called its *branches*, lying in the domains $x \geq a$ and $x \leq -a$. It has two *asymptotes*, straight lines the hyperbola approaches at large distances from its center, which are given by

$$y = \frac{b}{a}x \quad \text{and} \quad y = -\frac{b}{a}x.$$

The branches of a hyperbola lie within two vertical angles formed by the asymptotes and are called its *left* and *right* branches. The angle φ between the asymptotes of a hyperbola is determined by the equation

$$\tan\frac{\varphi}{2} = \frac{b}{a}.$$

Remark. If $a = b$ ($e = \sqrt{2}$), then $\varphi = \frac{1}{2}\pi$. In this case, the hyperbola is said to be *equilateral* or *rectangular* and its asymptotes are mutually perpendicular. The equation of an equilateral hyperbola has the form $x^2 - y^2 = a^2$. If the asymptotes are taken to be the coordinate axes, then the equation of the hyperbola becomes $xy = a^2/2$; i.e., an equilateral hyperbola represents an inverse proportionality dependence.

► **Focal and focus-directrix properties of a hyperbola.** The segments joining a point $M(x,y)$ of the hyperbola with the foci $F_1(-c,0)$ and $F_2(c,0)$ are called the *left* and *right focal radii* of this point. We denote the lengths of the left and right focal radii by $r_1 = |F_1M|$ and $r_2 = |F_2M|$, respectively (see Fig. M4.9 *b*). By the definition of a hyperbola,

$$|r_1 - r_2| = 2a,$$

where r_1 and r_2 satisfy the relations

$$r_1 = \sqrt{(x+c)^2 + y^2} = \begin{cases} a + ex & \text{for } x > 0, \\ -a - ex & \text{for } x < 0, \end{cases}$$

$$r_2 = \sqrt{(x-c)^2 + y^2} = \begin{cases} -a + ex & \text{for } x > 0, \\ a - ex & \text{for } x < 0. \end{cases}$$

The hyperbola defined by equation (M4.3.3.1) on the plane is the locus of points for which the ratio of distances to a focus and the like directrix is equal to e:

$$r_1\left|x+\frac{a}{e}\right|^{-1}=e,\qquad r_2\left|x-\frac{a}{e}\right|^{-1}=e.$$

(A focus and a directrix are said to be *like* if both of them are right or left simultaneously.)

▶ **Equation of a tangent and the optical property of a hyperbola.** The tangent to the hyperbola (M4.3.3.1) at an arbitrary point $M_0(x_0,y_0)$ is given by the equation

$$\frac{x_0x}{a^2}-\frac{y_0y}{b^2}=1.$$

The distances d_1 and d_2 from the foci $F_1(-c,0)$ and $F_2(0,c)$ to the tangent to the hyperbola at the point $M_0(x_0,y_0)$ are expressed as

$$d_1=\frac{|x_0e+a|}{aN}=\frac{r_1}{aN},\qquad N=\sqrt{\left(\frac{x_0}{a^2}\right)^2+\left(\frac{y_0}{b^2}\right)^2},$$
$$d_2=\frac{|x_0e-a|}{aN}=\frac{r_2}{aN},$$

where r_1 and r_2 are the lengths of the focal radii of the point M_0 (see Fig. M4.9 b, where $M=M_0$).

The tangent at any point $M_0(x_0,y_0)$ of the hyperbola forms acute angles φ_1 and φ_2 with the focal radii of the point of tangency, and

$$\sin\varphi_1=\frac{d_1}{r_1}=\frac{1}{aN},\quad \sin\varphi_2=\frac{d_2}{r_2}=\frac{1}{aN}.$$

This fact, written as

$$\varphi_1=\varphi_2,$$

is known as the *optical property of a hyperbola*. It means that a light ray issued from a focus of the hyperbola will reflect so as to appear as though issued from the other focus (see Fig. M4.9 b, where $M=M_0$).

The tangent to a hyperbola at any point bisects the angles between the straight lines joining this point with the foci. The tangent to a hyperbola at either of its vertices intersects the asymptotes at two points such that the distance between them is equal to $2b$.

▶ **Equations of a hyperbola in polar coordinates and parametric equations.** In polar coordinates (ρ,φ), with the pole coinciding with the right focus and the polar axis directed along the X-axis, the equation of the hyperbola has the form

$$\rho=\frac{p}{1-e\cos\varphi},$$

where $0\le\varphi\le2\pi$, $p=b^2/a$, and $e=\sqrt{1+b^2/a^2}$. If the pole is taken at the left focus, the equation of the hyperbola becomes

$$\rho=\frac{p}{1+e\cos\varphi}.$$

A parametric representation for the right branch of a hyperbola is given by the equations

$$x=a\cosh t,\quad y=b\sinh t,$$

with the parameter t assuming any real values.

A parametric representation that covers both branches of a hyperbola is given by the equations

$$x=a\sec t,\quad y=b\tan t,$$

with $-\pi\le t\le\pi$ and $t\ne\pm\frac{1}{2}\pi$.

M4.3.4. Parabola

► **Definition and the canonical equation of a parabola.** A *parabola* is the locus of all points in the plane equidistant from a given point F and a given straight line l, with $F \notin l$; see Fig. M4.10 a. The point F is called the *focus* of the parabola and the straight line l is its *directrix*.

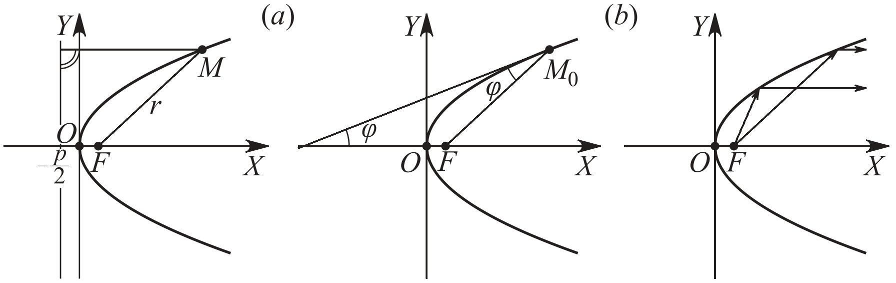

Figure M4.10. Parabola (a). Tangent to a parabola (b). Optical property of a parabola (c).

Let us draw a straight line through the focus F and perpendicularly to the directrix and denote the point at which this line crosses the directrix by C. Introduce the following Cartesian coordinate system: take the above line to be the X-axis (directed from C to F), the midpoint of the segment CF to be the origin O, and the perpendicular line through O to be the Y-axis. In this coordinate system, the parabola is determined by the equation

$$y^2 = 2px, \tag{M4.3.4.1}$$

where $p = |FC| > 0$. The number p is called *focal parameter* and equation (M4.3.4.1) is called the *canonical equation of the parabola*.

A parabola consists of an infinite branch symmetric about the X-axis. The point $O(0,0)$ is called the *vertex* of the parabola. The directrix of the parabola is given by the equation $x = -p/2$. The number $p/2$ is known as the *focal distance*. The segment joining a point $M(x,y)$ on the parabola with the focus $F(p/2,0)$ is called the *focal radius* of the point.

► **Focal properties of a parabola.** If r denotes the length of the focal radius FM, then by the definition of a parabola,

$$r = x + \frac{p}{2}.$$

As is apparent from Fig. M4.10a, the number r also satisfies the relation

$$r = \sqrt{\left(x - \frac{p}{2}\right)^2 + y^2}.$$

► **Equation of a tangent and the optical property of a parabola.** The tangent to the parabola (M4.3.4.1) at an arbitrary point $M_0(x_0, y_0)$ is given by the equation

$$yy_0 = p(x + x_0). \tag{M4.3.4.2}$$

The angle φ between the tangent to the parabola at a point $M_0(x_0, y_0)$ and the focal radius FM_0 is determined by

$$\cos\varphi = \frac{y_0}{\sqrt{y_0^2 + p^2}}.$$

The same relation holds for the angle between the tangent (M4.3.4.2) and the X-axis. This property of a parabola is called its *optical property*: a light ray issued from the focus reflects off the parabola in the direction parallel to the parabola axis (see Fig. M4.10 *c*).

► **Equation of a parabola in polar coordinates and parametric equations.** In polar coordinates (ρ, φ), with the pole at the focus of the parabola and the polar axis directed along the parabola axis, the equation of the parabola has the form

$$\rho = \frac{p}{1 - \cos\varphi},$$

where $0 < \varphi < 2\pi$.

Parametric equations of a parabola are

$$x = \tfrac{1}{2}pt^2, \quad y = pt,$$

with the parameter t assuming any real values.

M4.3.5. Transformation of Quadratic Curves to Canonical Form

► **General equation of a quadratic curve. Translation and rotation.** A set of points in the plane whose coordinates in the rectangular Cartesian coordinate system satisfy the general second-order algebraic equation

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0 \qquad \text{(M4.3.5.1)}$$

is called a (bivariate) *quadratic curve* (or just *quadratic*); it is also known as a *second-order curve*. If equation (M4.3.5.1) does not determine a real geometric object, this equation is said to determine an *imaginary quadratic curve*.

Equation (M4.3.5.1) may be simplified using the following transformations of the Cartesian coordinate system:

1. Translation:

$$x = \bar{x} + x_0, \qquad y = \bar{y} + y_0. \qquad \text{(M4.3.5.2)}$$

It means that the origin $O(0,0)$ is transferred to the point $\bar{O}(x_0, y_0)$ and the coordinate axes are moved parallel to the original ones; $\bar{x}$ and $\bar{y}$ denote the new coordinates.

2. Rotation:

$$x = \hat{x}\cos\varphi - \hat{y}\sin\varphi, \qquad y = \hat{x}\sin\varphi + \hat{y}\cos\varphi. \qquad \text{(M4.3.5.3)}$$

The coordinate axes are rotated about the origin, which does not move, by the angle φ counterclockwise; $\hat{x}$ and $\hat{y}$ denote the new coordinates.

► **Canonical equations of quadratic curves. The classification table.** With transformations (M4.3.5.2)–(M4.3.5.3), equation (M4.3.5.1) can be reduced to one of the *nine canonical forms* classified in Table M4.1. The first five curves, with $\delta \neq 0$, are *nondegenerate* (their canonical equations contain two quadratic terms proportional to x^2 and y^2). The last four curves, with $\delta = 0$, are *degenerate* (their canonical equations contain only one quadratic term, x^2 or y^2). Curves 3, 5, 7, 8, and 9, with $\Delta = 0$, split into straight lines; their equations can be represented as the product of two factors linear in the coordinates, each having the form $(\alpha_n x + \beta_n y + \gamma_n)$, on the left-hand side and zero on the right-hand side.

TABLE M4.1
Classification of quadratic curves.

No.	Curve name	Canonical equation	Conditions for invariants
1	Ellipse	$\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=1$	$\delta>0,\ I\Delta<0$
2	Imaginary ellipse (no real points)	$\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=-1$	$\delta>0,\ I\Delta>0$
3	Pair of imaginary straight lines intersecting at a real point	$\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=0$	$\delta>0,\ \Delta=0$
4	Hyperbola	$\dfrac{x^2}{a^2}-\dfrac{y^2}{b^2}=1$	$\delta<0,\ \Delta\neq0$
5	Pair of intersecting straight lines (degenerate hyperbola)	$\dfrac{x^2}{a^2}-\dfrac{y^2}{b^2}=0$	$\delta<0,\ \Delta=0$
6	Parabola	$y^2=2px$	$\delta=0,\ \Delta\neq0$
7	Pair of parallel straight lines	$x^2-a^2=0$	$\delta=\Delta=0,\ \sigma<0$
8	Pair of imaginary parallel straight lines (no real points)	$x^2+a^2=0$	$\delta=\Delta=0,\ \sigma>0$
9	Pair of coinciding straight lines	$x^2=0$	$\delta=\Delta=\sigma=0$

► **Invariants of quadratic curves.** Quadratic curves can be studied using the three *invariants*

$$I=a_{11}+a_{22},\quad \delta=\begin{vmatrix}a_{11}&a_{12}\\a_{12}&a_{22}\end{vmatrix},\quad \Delta=\begin{vmatrix}a_{11}&a_{12}&a_{13}\\a_{12}&a_{22}&a_{23}\\a_{13}&a_{23}&a_{33}\end{vmatrix},\tag{M4.3.5.4}$$

whose values do not change under parallel translations and rotations of the coordinate axes, and the sign of the quantity

$$\sigma=\begin{vmatrix}a_{11}&a_{13}\\a_{13}&a_{33}\end{vmatrix}+\begin{vmatrix}a_{22}&a_{23}\\a_{23}&a_{33}\end{vmatrix}.\tag{M4.3.5.5}$$

The invariant Δ is called the *large discriminant* of equation (M4.3.5.1). The invariant δ is called the *small discriminant*.

The quadratic curves can be classified based on the values of the invariants, specified in the last column in Table M4.1.

► **Characteristic equation of quadratic curves.** The properties of quadratic curves can be studied using the *characteristic equation*

$$\begin{vmatrix}a_{11}-\lambda&a_{12}\\a_{12}&a_{22}-\lambda\end{vmatrix}=0\quad\text{or}\quad\lambda^2-I\lambda+\delta=0.\tag{M4.3.5.6}$$

The roots λ_1 and λ_2 of the characteristic equation (M4.3.5.6) are eigenvalues of a real symmetric matrix, $[a_{ij}]$, and hence are real.

The invariants I and δ are expressed in terms of the roots λ_1 and λ_2 as follows:

$$I=\lambda_1+\lambda_2,\quad \delta=\lambda_1\lambda_2.\tag{M4.3.5.7}$$

► **Nondegenerate case $\delta\neq0$. Reduction of quadratic curves to canonical form.** First, by applying the translation transformation (M4.3.5.2) with

$$x_0=-\frac{1}{\delta}\begin{vmatrix}a_{13}&a_{12}\\a_{23}&a_{22}\end{vmatrix},\quad y_0=-\frac{1}{\delta}\begin{vmatrix}a_{11}&a_{13}\\a_{12}&a_{23}\end{vmatrix},$$

one reduces equation (M4.3.5.1) to the form

$$a_{11}\bar{x}^2 + 2a_{12}\bar{x}\bar{y} + a_{22}\bar{y}^2 + \frac{\Delta}{\delta} = 0. \qquad \text{(M4.3.5.8)}$$

Then, with the rotation transformation (M4.3.5.3) where $\bar{x}$ and $\bar{y}$ are substituted for x and y and the angle φ determined by

$$\tan 2\varphi = \frac{2a_{12}}{a_{11} - a_{22}} \qquad \Bigl(\text{if } a_{11} = a_{22}, \text{ then } \varphi = \frac{\pi}{4}\Bigr),$$

equation (M4.3.5.8) is transformed into

$$\lambda_1 \hat{x}^2 + \lambda_2 \hat{y}^2 + \frac{\Delta}{\delta} = 0,$$

where λ_1 and λ_2 are the roots of the characteristic equation (M4.3.5.6).

Note the following formulas for an ellipse:

$$a^2 = -\frac{1}{\lambda_2}\frac{\Delta}{\delta} = -\frac{\Delta}{\lambda_1\lambda_2^2}, \quad b^2 = -\frac{1}{\lambda_1}\frac{\Delta}{\delta} = -\frac{\Delta}{\lambda_1^2\lambda_2} \qquad (\lambda_1 \ge \lambda_2),$$

where a and b are the semimajor and semiminor axes of the ellipse.

Similar formulas for a hyperbola have the form

$$a^2 = -\frac{1}{\lambda_1}\frac{\Delta}{\delta} = -\frac{\Delta}{\lambda_1^2\lambda_2}, \quad b^2 = \frac{1}{\lambda_1}\frac{\Delta}{\delta} = \frac{\Delta}{\lambda_1^2\lambda_2} \qquad (\lambda_1 \ge \lambda_2).$$

► **Degenerate case $\delta = 0$. Reduction of quadratic curves to canonical form.** If $\delta = 0$, equation (M4.3.5.1) can be rewritten as

$$(\alpha x + \beta y)^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0. \qquad \text{(M4.3.5.9)}$$

If the coefficients a_{13} and a_{23} are respectively proportional to α and β, i.e., $a_{13} = k\alpha$ and $a_{23} = k\beta$, then equation (M4.3.5.9) becomes $(\alpha x + \beta y)^2 + 2k(\alpha x + \beta y) + a_{33} = 0$, and hence

$$\alpha x + \beta y = -k \pm \sqrt{k^2 - a_{33}},$$

which determines a pair of real (or imaginary) parallel straight lines.

If a_{13} and a_{23} are not proportional to α and β, then equation (M4.3.5.9) can be rewritten as

$$(\alpha x + \beta y + \gamma)^2 + 2k(\beta x - \alpha x + q) = 0. \qquad \text{(M4.3.5.10)}$$

The parameters k, γ, and q can be determined by comparing the coefficients in equations (M4.3.5.9) and (M4.3.5.10). If the line $\alpha x + \beta y + \gamma = 0$ is treated as the axis $\bar{O}\bar{X}$ and the line $\beta x - \alpha x + q = 0$ as the axis $\bar{O}\bar{Y}$ and the new coordinates are expressed as

$$\hat{x} = \frac{\beta x - \alpha x + q}{\pm\sqrt{\alpha^2 + \beta^2}}, \quad \hat{y} = \frac{\alpha x + \beta y + \gamma}{\pm\sqrt{\alpha^2 + \beta^2}},$$

then equation (M4.3.5.10) acquires the form

$$\hat{y}^2 = 2p\hat{x},$$

where $p = |k|/\sqrt{\alpha^2 + \beta^2}$. The axis $\bar{O}\bar{X}$ points to the half-plane where the sign of $\beta x - \alpha x + q$ is opposite to that of k.

The focal parameter p of a parabola is expressed in terms of the invariants I, δ, and Δ and the roots λ_1 and λ_2 ($\lambda_1 \ge \lambda_2$) of the characteristic equation (M4.3.5.6) as follows:

$$p = \frac{1}{I}\sqrt{-\frac{\Delta}{I}} = \frac{1}{\lambda_1}\sqrt{-\frac{\Delta}{\lambda_1}} > 0, \quad \lambda_2 = 0.$$

M4.4. Coordinates, Vectors, Curves, and Surfaces in Space

M4.4.1. Vectors and Their Properties

► **Notion of a vector.** A directed line segment connecting an initial point A and a terminal point B (see Fig. M4.11) is called a *vector* and denoted $\overrightarrow{AB}$. The nonnegative number equal to the length of the segment AB is called the *length* (or *magnitude*) of the *vector* $\overrightarrow{AB}$ and denoted $|\overrightarrow{AB}|$. The vector $\overrightarrow{BA}$ is said to be *opposite* to the vector $\overrightarrow{AB}$; it has the same magnitude but opposite direction. Vectors are usually denoted by a single lowercase letter, either with an arrow above (e.g., $\vec{a}$) or without (e.g., $\mathbf{a}$); the latter is the most common notation for vectors, in which case a boldface lowercase letter is used.

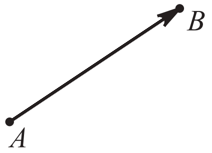

Figure M4.11. Vector $\overrightarrow{AB}$.

Two vectors are said to be *collinear* (*parallel*) if they lie on the same straight line or on parallel lines. Three vectors are said to be *coplanar* if they lie in the same plane or in parallel planes. A vector 0 whose initial and terminal points coincide is called the *zero vector* (or *null vector*); its length is zero ($|0| = 0$) and its direction is assumed to be arbitrary. A vector $\mathbf{e}$ of length one is called a *unit vector*.

Two vectors are called *equal* is they are collinear and have the same magnitude and direction. It follows that, for any vector $\mathbf{a}$ and any point A, there exists a unique vector $\overrightarrow{AB}$ with its start point at A that is equal to $\mathbf{a}$. For this reason, vectors in analytical geometry are defined up to their position, so that all vectors obtained from each other by parallel transport are considered to be the same.

► **Sum and difference of vectors.** The *sum* $\mathbf{a}+\mathbf{b}$ *of vectors* $\mathbf{a}$ *and* $\mathbf{b}$ is defined as the vector directed from the initial point of $\mathbf{a}$ to the terminal point of $\mathbf{b}$ where the start of $\mathbf{b}$ is placed at the tip of $\mathbf{a}$. This method of the addition of vectors is called the *triangle rule* (see Fig. M4.12 *a*). The sum $\mathbf{a}+\mathbf{b}$ can also be found using the *parallelogram rule* as shown in Fig. M4.12 *b*. The *difference* $\mathbf{a}-\mathbf{b}$ *of vectors* $\mathbf{a}$ *and* $\mathbf{b}$ is defined as the vector that must be added to $\mathbf{b}$ to get $\mathbf{a}$: $\mathbf{b}+(\mathbf{a}-\mathbf{b}) = \mathbf{a}$ (see Fig. M4.12 *c*).

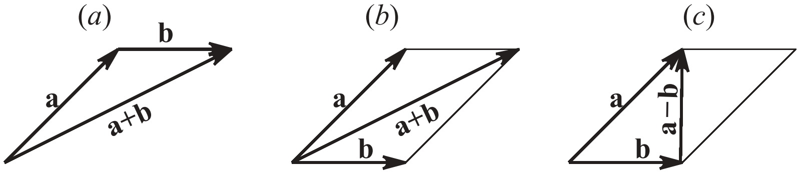

Figure M4.12. The sum of vectors: triangle rule (*a*) and parallelogram rule (*b*). The difference of vectors (*c*).

The *product* $\lambda\mathbf{a}$ *of a vector* $\mathbf{a}$ *by a number* λ is defined as the vector whose magnitude is equal to $|\lambda\mathbf{a}| = |\lambda||\mathbf{a}|$ and direction coincides with that of $\mathbf{a}$ if $\lambda > 0$ or is opposite to it if $\lambda < 0$.

Remark. If $\mathbf{a} = 0$ or $\lambda = 0$, then the resulting product is the zero vector. In this case, the direction of the product $\lambda\mathbf{a}$ is undetermined.

► **Main properties of operations with vectors.**

1. $\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}$ (commutativity).
2. $\mathbf{a}+(\mathbf{b}+\mathbf{c})=(\mathbf{a}+\mathbf{b})+\mathbf{c}$ (associativity of addition).
3. $\mathbf{a}+0=\mathbf{a}$ (existence of zero vector).
4. $\mathbf{a}+(-\mathbf{a})=0$ (existence of opposite vector).
5. $\lambda(\mathbf{a}+\mathbf{b})=\lambda\mathbf{a}+\lambda\mathbf{b}$ (distributivity with respect to addition of vectors).
6. $(\lambda+\mu)\mathbf{a}=\lambda\mathbf{a}+\mu\mathbf{a}$ (distributivity with respect to addition of constants).
7. $\lambda(\mu\mathbf{a})=(\lambda\mu)\mathbf{a}$ (associativity of product).
8. $1\mathbf{a}=\mathbf{a}$ (multiplication by unity).

M4.4.2. Coordinate Systems

► **Cartesian coordinate system. Some useful formulas.** A *rectangular Cartesian coordinate system* (also called just *rectangular coordinate system* or *Cartesian coordinate system*) is defined by three pairwise perpendicular directed straight lines OX, OY, and OZ (the *coordinate axes*) concurrent at a single point O (the *origin*).

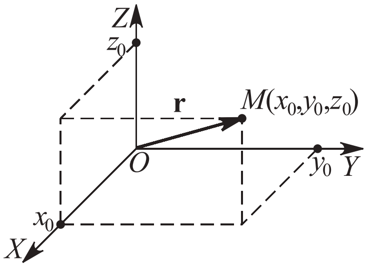

Figure M4.13. A point in a rectangular Cartesian coordinate system.

For an arbitrary point M in space, let us draw through it three planes parallel to the planes OYZ, OXZ, and OXY. These planes will intersect the coordinate axes OX, OY, and OZ at three points. Denote by x_0, y_0, and z_0 the distances from these points to the origin O (see Fig. M4.13). The numbers x_0, y_0, and z_0 are, respectively, called the x-coordinate (or *abscissa*), the y-coordinate (or *ordinate*), and the z-coordinate of the point M. One usually uses the notation $M(x_0,y_0,z_0)$ to specify that the point M has the coordinates (x_0,y_0,z_0).

Planes parallel to the coordinate planes are coordinate surfaces on which one of the coordinates is constant. Straight lines parallel to the coordinate axes are coordinate lines along which only one coordinate varies and the other two remain constant. Coordinate surfaces meet at coordinate lines.

Each point M in three-dimensional space uniquely defines a vector $\overrightarrow{OM}$, which is called the *position vector* of the point M. The coordinates of the position vector coincide with those of M and one usually writes $\mathbf{r}=\overrightarrow{OM}=(x_0,y_0,z_0)$.

The distance between two points, $M_1(x_1,y_1,z_1)$ and $M_2(x_2,y_2,z_2)$, is given by the formula

$$d=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2}=|\mathbf{r}_2-\mathbf{r}_1|, \qquad \text{(M4.4.2.1)}$$

where $\mathbf{r}_2=\overrightarrow{OM}_2$ and $\mathbf{r}_1=\overrightarrow{OM}_1$ are the position vectors of the points M_1 and M_2, respectively (see Fig. M4.14).

Any triple of numbers (x,y,z) can be identified with a point P and a position vector $\overrightarrow{OP}$, whose coordinates are these numbers. An arbitrary vector (x,y,z) can be represented as

$$(x,y,z)=x\mathbf{i}+y\mathbf{j}+z\mathbf{k},$$

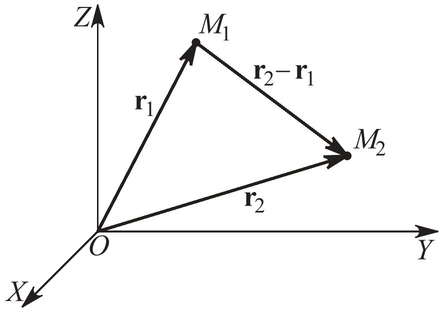

Figure M4.14. Distance between points.

where $\mathbf{i} = (1,0,0)$, $\mathbf{j} = (0,1,0)$, and $\mathbf{k} = (0,0,1)$ are the unit vectors with the same directions as the coordinate axes OX, OY, and OZ (*basis vectors*).

Two vectors $\mathbf{r}_1 = (x_1, y_1, z_1)$ and $\mathbf{r}_2 = (x_2, y_2, z_2)$ are equal to each other if and only if the relations

$$x_1 = x_2, \quad y_1 = y_2, \quad z_1 = z_2$$

hold simultaneously. The coordinates of the sum or difference of vectors and the product of a vector by a scalar are calculated as

$$(x_1, y_1, z_1) \pm (x_2, y_2, z_2) = (x_1 \pm x_2, y_1 \pm y_2, z_1 \pm z_2),$$
$$\alpha\,(x, y, z) = (\alpha x, \alpha y, \alpha z).$$

If a point M divides a directed segment $\overrightarrow{M_1M_2}$ in a ratio λ, then the coordinates of this point are given by

$$x = \frac{x_1 + \lambda x_2}{1+\lambda}, \quad y = \frac{y_1 + \lambda y_2}{1+\lambda}, \quad z = \frac{z_1 + \lambda z_2}{1+\lambda} \qquad \text{or} \qquad \mathbf{r} = \frac{\mathbf{r}_1 + \lambda \mathbf{r}_2}{1+\lambda}, \qquad \text{(M4.4.2.2)}$$

where $\lambda = |\overrightarrow{M_1M}|/|\overrightarrow{MM_2}|$. The special case where M is the midpoint of $\overrightarrow{M_1M_2}$ corresponds to $\lambda = 1$.

The angles α, β, and γ between $\overrightarrow{M_1M_2}$ and the coordinate axes OX, OY, and OZ are determined by

$$\cos\alpha = \frac{x_2 - x_1}{|\mathbf{r}_2 - \mathbf{r}_1|}, \quad \cos\beta = \frac{y_2 - y_1}{|\mathbf{r}_2 - \mathbf{r}_1|}, \quad \cos\gamma = \frac{z_2 - z_1}{|\mathbf{r}_2 - \mathbf{r}_1|},$$

with

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1.$$

The numbers $\cos\alpha$, $\cos\beta$, and $\cos\gamma$ are called the *direction cosines* of the vector $\overrightarrow{M_1M_2}$.

The angle φ between two vectors $\overrightarrow{M_1M_2}$ and $\overrightarrow{M_3M_4}$ defined by the points $M_1(x_1, y_1, z_1)$, $M_2(x_2, y_2, z_2)$, $M_3(x_3, y_3, z_3)$, and $M_4(x_4, y_4, z_4)$ can be found from

$$\cos\varphi = \frac{(x_2 - x_1)(x_4 - x_3) + (y_2 - y_1)(y_4 - y_3) + (z_2 - z_1)(z_4 - z_3)}{|\mathbf{r}_2 - \mathbf{r}_1|\,|\mathbf{r}_4 - \mathbf{r}_3|}.$$

The area of the triangle with vertices M_1, M_2, and M_3 is given by the formula

$$S = \frac{1}{4}\sqrt{\begin{vmatrix} y_1 & z_1 & 1 \\ y_2 & z_2 & 1 \\ y_3 & z_3 & 1 \end{vmatrix}^2 + \begin{vmatrix} z_1 & x_1 & 1 \\ z_2 & x_2 & 1 \\ z_3 & x_3 & 1 \end{vmatrix}^2 + \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}^2}.$$

The volume of the pyramid with vertices M_1, M_2, M_3, and M_4 is equal to

$$V = \frac{1}{6}|D|, \qquad D = \begin{vmatrix} x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ x_3 - x_1 & y_3 - y_1 & z_3 - z_1 \\ x_4 - x_1 & y_4 - y_1 & z_4 - z_1 \end{vmatrix} = \begin{vmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \\ 1 & x_4 & y_4 & z_4 \end{vmatrix},$$

and the volume of the parallelepiped spanned by vectors $\overrightarrow{M_1M_2}$, $\overrightarrow{M_1M_3}$, and $\overrightarrow{M_1M_4}$ is equal to

$$V = |D|.$$

► **Cylindrical coordinates.** *Cylindrical coordinates* are a generalization of polar coordinates (see Subsection M4.1.1) that adds a third dimension. If a point M is specified by its cylindrical coordinates, they are the polar coordinates ρ and φ of the projection of M onto a base plane (usually OXY) and the distance (usually z) of M from this base plane (see Fig. M4.15 *a*). It is usually assumed that $0 \le \varphi \le 2\pi$ (or $-\pi \le \varphi \le \pi$). For cylindrical coordinates, the coordinate surfaces are planes $z = \text{const}$ perpendicular to the axis OZ, half-planes $\varphi = \text{const}$ bounded by the axis OZ, and cylindrical surfaces $\rho = \text{const}$ with axis OZ.

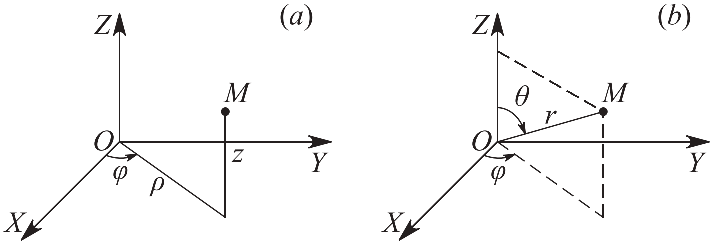

Figure M4.15. Point in cylindrical (*a*) and spherical (*b*) coordinates.

Let M be an arbitrary point in space with its Cartesian coordinates (x, y, z) and cylindrical coordinates (ρ, φ, z). The conversion formulas from Cartesian to cylindrical and from cylindrical to Cartesian coordinates are as follows:

$$\begin{aligned} x &= \rho\cos\varphi, & \rho &= \sqrt{x^2 + y^2}, \\ y &= \rho\sin\varphi, & \tan\varphi &= y/x, \\ z &= z, & z &= z, \end{aligned}$$

where the polar angle φ is taken with regard to the quadrant in which the projection of the point M onto the base plane lies.

► **Spherical coordinates.** The *spherical coordinates* of a point M are defined as the length $r = |\overrightarrow{OM}|$ of its position vector, the *azimuthal angle* φ from the positive direction of the axis OX to the projection of M onto the plane OXY, and the *zenithal angle* θ from the positive direction of the axis OZ to M (see Fig. M4.15 *b*). It is usually assumed that $0 \le \varphi \le 2\pi$ and $0 \le \theta \le \pi$ (or $-\pi \le \varphi \le \pi$ and $0 \le \theta \le \pi$). For spherical coordinates, the coordinate surfaces are spheres $r = \text{const}$ centered at the origin, half-planes $\varphi = \text{const}$ bounded by the axis OZ, and cones $\theta = \text{const}$ with vertex O and axis OZ.

The conversion formulas from the Cartesian coordinates (x, y, z) to the spherical coordinates (r, φ, θ) and back are as follows:

$$\begin{aligned} x &= r\sin\theta\cos\varphi, & r &= \sqrt{x^2 + y^2 + z^2}, \\ y &= r\sin\theta\sin\varphi, & \tan\varphi &= y/x, \\ z &= r\cos\theta, & \tan\theta &= \sqrt{x^2 + y^2}/z, \end{aligned}$$

where the angle φ is determined from the same considerations as in the case of cylindrical coordinates.

M4.4.3. Scalar, Cross, and Scalar Triple Products of Vectors

► **Scalar product of two vectors.** The *scalar product* (also known as the *dot product*) of two vectors **a** and **b**, is defined as the product of their magnitudes by the cosine of the angle between the vectors (see Fig. M4.16),

$$\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}||\mathbf{b}|\cos\varphi.$$

It follows that $\mathbf{a}\cdot\mathbf{b} > 0$ if the angle between **a** and **b** is acute, $\mathbf{a}\cdot\mathbf{b} < 0$ if it is obtuse, and $\mathbf{a}\cdot\mathbf{b} = 0$ if it is right.

Remark. The scalar product of a vector **a** by a vector **b** is also denoted by $(\mathbf{a}\cdot\mathbf{b})$, $(\mathbf{a},\mathbf{b})$, and $\mathbf{ab}$.

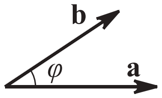

Figure M4.16. Scalar product of two vectors.

Properties of the scalar product:

1. $\mathbf{a}\cdot\mathbf{b} = \mathbf{b}\cdot\mathbf{a}$ (commutativity).
2. $\mathbf{a}\cdot(\mathbf{b}+\mathbf{c}) = \mathbf{a}\cdot\mathbf{b}+\mathbf{a}\cdot\mathbf{c}$ (distributivity with respect to addition of vectors). This property holds for any number of summands.
3. If **a** and **b** are collinear, then $\mathbf{a}\cdot\mathbf{b} = \pm|\mathbf{a}||\mathbf{b}|$. (The plus sign is taken if **a** and **b** have the same direction, and the minus sign is taken if they have opposite directions.)
4. $(\lambda\mathbf{a})\cdot\mathbf{b} = \lambda(\mathbf{a}\cdot\mathbf{b})$ (associativity with respect to a scalar factor).
5. $\mathbf{a}\cdot\mathbf{a} = |\mathbf{a}|^2$. The scalar product $\mathbf{a}\cdot\mathbf{a}$ is denoted by $\mathbf{a}^2$ (the *scalar square* of the vector **a**).
6. The magnitude of a vector is expressed via the scalar product as

$$|\mathbf{a}| = \sqrt{\mathbf{a}\cdot\mathbf{a}} = \sqrt{\mathbf{a}^2}.$$

7. Two nonzero vectors **a** and **b** are perpendicular if and only if $\mathbf{a}\cdot\mathbf{b} = 0$.
8. The scalar products of the basis vectors are

$$\mathbf{i}\cdot\mathbf{j} = \mathbf{i}\cdot\mathbf{k} = \mathbf{j}\cdot\mathbf{k} = 0, \quad \mathbf{i}\cdot\mathbf{i} = \mathbf{j}\cdot\mathbf{j} = \mathbf{k}\cdot\mathbf{k} = 1.$$

9. If vectors are given by their coordinates, $\mathbf{a} = (a_x, a_y, a_z)$ and $\mathbf{b} = (b_x, b_y, b_z)$, then

$$\mathbf{a}\cdot\mathbf{b} = (a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k})(b_x\mathbf{i} + b_y\mathbf{j} + b_z\mathbf{k}) = a_xb_x + a_yb_y + a_zb_z.$$

10. The *Cauchy–Schwarz inequality*

$$|\mathbf{a}\cdot\mathbf{b}| \le |\mathbf{a}||\mathbf{b}|.$$

11. The *Minkowski inequality*

$$|\mathbf{a}+\mathbf{b}| \le |\mathbf{a}| + |\mathbf{b}|.$$

12. The *angle φ between vectors* **a** and **b** is determined by the formula

$$\cos\varphi = \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}||\mathbf{b}|} = \frac{a_xb_x + a_yb_y + a_zb_z}{\sqrt{a_x^2 + a_y^2 + a_z^2}\sqrt{b_x^2 + b_y^2 + b_z^2}}.$$

► **Cross product of two vectors.** The *cross product* of a vector **a** by a vector **b** is defined as the vector **c** denoted by $\mathbf{a}\times\mathbf{b}$ (see Fig. M4.17), satisfying the following three conditions:

1. Its absolute value (magnitude) is equal to the area of the parallelogram spanned by the vectors **a** and **b**; i.e.,
$$|\mathbf{c}| = |\mathbf{a}\times\mathbf{b}| = |\mathbf{a}||\mathbf{b}|\sin\varphi.$$
2. It is perpendicular to the plane of the parallelogram; i.e., $\mathbf{c}\perp\mathbf{a}$ and $\mathbf{c}\perp\mathbf{b}$.
3. The vectors **a**, **b**, and **c** form a *right-handed trihedral*; i.e., the vector **c** points to the side from which the sense of the shortest rotation from **a** to **b** is counterclockwise.

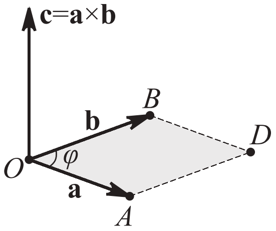

Figure M4.17. Cross product of two vectors.

Remark. The cross product of a vector **a** by a vector **b** is also denoted by $\mathbf{c} = [\mathbf{a},\mathbf{b}]$.

Properties of cross product:

1. $\mathbf{a}\times\mathbf{b} = -\mathbf{b}\times\mathbf{a}$ (anticommutativity).
2. $\mathbf{a}\times(\mathbf{b}+\mathbf{c}) = \mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c}$ (distributivity with respect to the addition of vectors). This property holds for any number of summands.
3. Vectors **a** and **b** are collinear if and only if $\mathbf{a}\times\mathbf{b} = 0$. In particular, $\mathbf{a}\times\mathbf{a} = 0$.
4. $(\lambda\mathbf{a})\times\mathbf{b} = \mathbf{a}\times(\lambda\mathbf{b}) = \lambda(\mathbf{a}\times\mathbf{b})$ (associativity with respect to a scalar factor).
5. The cross products of basis vectors are
$$\mathbf{i}\times\mathbf{i} = \mathbf{j}\times\mathbf{j} = \mathbf{k}\times\mathbf{k} = 0, \quad \mathbf{i}\times\mathbf{j} = \mathbf{k}, \quad \mathbf{j}\times\mathbf{k} = \mathbf{i}, \quad \mathbf{k}\times\mathbf{i} = \mathbf{j}.$$
6. If the vectors are given by their coordinates $\mathbf{a} = (a_x, a_y, a_z)$ and $\mathbf{b} = (b_x, b_y, b_z)$, then
$$\mathbf{a}\times\mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix} = (a_y b_z - a_z b_y)\mathbf{i} + (a_z b_x - a_x b_z)\mathbf{j} + (a_x b_y - a_y b_x)\mathbf{k}.$$
7. The area of the parallelogram spanned by vectors **a** and **b** is equal to
$$S = |\mathbf{a}\times\mathbf{b}| = \sqrt{\begin{vmatrix} a_y & a_z \\ b_y & b_z \end{vmatrix}^2 + \begin{vmatrix} a_x & a_z \\ b_x & b_z \end{vmatrix}^2 + \begin{vmatrix} a_x & a_y \\ b_x & b_y \end{vmatrix}^2}.$$
8. The area of the triangle spanned by vectors **a** and **b** is equal to
$$S = \frac{1}{2}|\mathbf{a}\times\mathbf{b}| = \frac{1}{2}\sqrt{\begin{vmatrix} a_y & a_z \\ b_y & b_z \end{vmatrix}^2 + \begin{vmatrix} a_x & a_z \\ b_x & b_z \end{vmatrix}^2 + \begin{vmatrix} a_x & a_y \\ b_x & b_y \end{vmatrix}^2}.$$

► **Conditions for vectors to be parallel or perpendicular.**
A vector **a** is collinear to a vector **b** if
$$\mathbf{b} = \lambda\mathbf{a} \quad\text{or}\quad \mathbf{a}\times\mathbf{b} = 0.$$

A vector **a** is perpendicular to a vector **b** if

$$\mathbf{a}\cdot\mathbf{b} = 0.$$

Remark. In general, the condition $\mathbf{a}\cdot\mathbf{b} = 0$ implies that the vectors **a** and **b** are perpendicular or one of them is the zero vector. The zero vector can be viewed to be perpendicular to any other vector.

► **Scalar triple product of three vectors.** The *scalar triple product* of vectors **a**, **b**, and **c** is defined as the scalar product of **a** by the cross product of **b** and **c**:

$$[\mathbf{abc}] = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}).$$

Remark. The scalar triple product of three vectors **a**, **b**, and **c** is also denoted by **abc**.

Properties of scalar triple product:

1. $[\mathbf{abc}] = [\mathbf{bca}] = [\mathbf{cab}] = -[\mathbf{bac}] = -[\mathbf{cba}] = -[\mathbf{acb}]$.
2. $[\mathbf{aab}] = [\mathbf{bab}] = 0$ or $\mathbf{a}\cdot(\mathbf{a}\times\mathbf{b}) = \mathbf{b}\cdot(\mathbf{a}\times\mathbf{b}) = 0$.
3. $[(\mathbf{a}+\mathbf{b})\mathbf{cd}] = [\mathbf{acd}] + [\mathbf{bcd}]$ (distributivity with respect to addition of vectors). This property holds for any number of summands.
4. $[\lambda\mathbf{abc}] = \lambda[\mathbf{abc}]$ (associativity with respect to a scalar factor).
5. If the vectors are given by their coordinates $\mathbf{a} = (a_x, a_y, a_z)$, $\mathbf{b} = (b_x, b_y, b_z)$, and $\mathbf{c} = (c_x, c_y, c_z)$, then
$$[\mathbf{abc}] = \begin{vmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{vmatrix}.$$
6. The scalar triple product [**abc**] is equal to the volume V of the parallelepiped spanned by the vectors **a**, **b**, and **c** taken with the sign + if the vectors **a**, **b**, and **c** form a right-handed trihedral and the sign – if the vectors form a left-handed trihedral,
$$[\mathbf{abc}] = \pm V.$$
7. Three nonzero vectors **a**, **b**, and **c** are coplanar if and only if $[\mathbf{abc}] = 0$. In this case, the vectors **a**, **b**, and **c** are linearly dependent; they satisfy a relation of the form $\alpha\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c} = 0$.

M4.5. Line and Plane in Space

M4.5.1. Plane in Space

► **General equation of a plane.** In a Cartesian coordinate system, a plane is given by a first-order algebraic equation.

The *general* (*complete*) *equation of a plane* has the form

$$Ax + By + Cz + D = 0, \tag{M4.5.1.1}$$

where $A^2 + B^2 + C^2 \neq 0$.

1. For $D = 0$, the equation defines a plane passing through the origin.
2. For $A = 0$ (respectively, $B = 0$ or $C = 0$), the equation defines a plane parallel to the axis OX (respectively, OY or OZ).
3. For $A = D = 0$ (respectively, $B = D = 0$ or $C = D = 0$), the equation defines a plane passing through the axis OX (respectively, OY or OZ).
4. For $A = B = 0$ (respectively, $A = C = 0$ or $B = C = 0$), the equation defines a plane parallel to the plane OXY (respectively, OXZ or OYZ).

► **Intercept equation of a plane.** A plane $Ax+By+Cz+D=0$ that is not parallel to the axis OX (i.e., $A\neq 0$) meets this axis at a (signed) distance $a=-D/A$ from the origin (see Fig. M4.18). The number a is called the x-intercept of the plane. Similarly, one defines the y-intercepts $b=-D/B$ (for $B\neq 0$) and the z-intercept $c=-D/C$ (for $C\neq 0$). Then such a plane can be defined by the equation

$$\frac{x}{a}+\frac{y}{b}+\frac{z}{c}=1,$$

which is called the *intercept equation of the plane*.

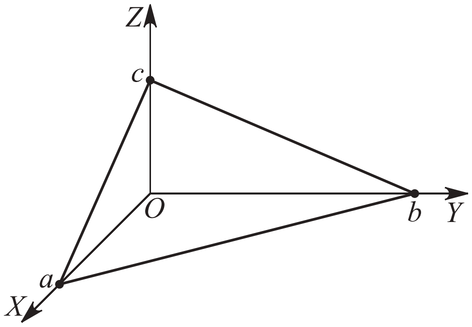

Figure M4.18. A plane with intercept equation.

Remark. A plane parallel to the axis OX but nonparallel to the other two axes is defined by the equation $y/b+z/c=1$, where b and c are the y- and z-intercepts of the plane. A plane simultaneously parallel to the axes OX and OY can be represented in the form $z/c=1$.

► **Equation of the plane passing through a point M_0 and perpendicular to a vector N.** The equation of the plane passing through a point $M_0(x_0,y_0,z_0)$ and perpendicular to a vector $\mathbf{N}=(A,B,C)$ has the form

$$A(x-x_0)+B(y-y_0)+C(z-z_0)=0, \qquad \text{or} \qquad (\mathbf{r}-\mathbf{r}_0)\cdot\mathbf{N}=0, \tag{M4.5.1.2}$$

where $\mathbf{r}$ and $\mathbf{r}_0$ are the position vectors of the points $M(x,y,z)$ and $M_0(x_0,y_0,z_0)$, respectively (see Fig. M4.19). The vector $\mathbf{N}$ is called a *normal vector*. Its direction cosines are

$$\cos\alpha=\frac{A}{\sqrt{A^2+B^2+C^2}},\quad \cos\beta=\frac{B}{\sqrt{A^2+B^2+C^2}},\quad \cos\gamma=\frac{C}{\sqrt{A^2+B^2+C^2}}.$$

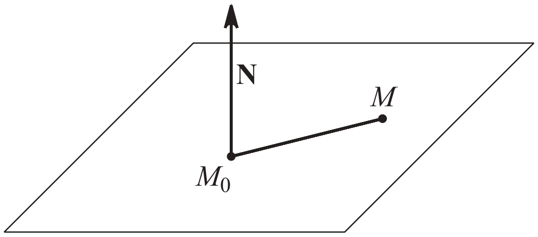

Figure M4.19. Plane passing through a point M_0 and perpendicular to a vector $\mathbf{N}$.

► **Equation of the plane passing through a point and parallel to another plane.** The plane that passes through a point $M_0(x_0, y_0, z_0)$ and is parallel to a plane $Ax+By+Cz+D=0$ is given by equation (M4.5.1.2).

► **Equation of the plane passing through three points.** The plane passing through three points $M_1(x_1, y_1, z_1)$, $M_2(x_2, y_2, z_2)$, and $M_3(x_3, y_3, z_3)$ (see Fig. M4.20) is described by the equation

$$\begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ x_2-x_1 & y_2-y_1 & z_2-z_1 \\ x_3-x_1 & y_3-y_1 & z_3-z_1 \end{vmatrix} = 0, \qquad \text{or} \qquad \big[(\mathbf{r}-\mathbf{r}_1)(\mathbf{r}_2-\mathbf{r}_1)(\mathbf{r}_3-\mathbf{r}_1)\big] = 0, \quad \text{(M4.5.1.3)}$$

where $\mathbf{r}$, $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{r}_3$ are the position vectors of the points $M(x, y, z)$, $M_1(x_1, y_1, z_1)$, $M_2(x_2, y_2, z_2)$, and $M_3(x_3, y_3, z_3)$, respectively.

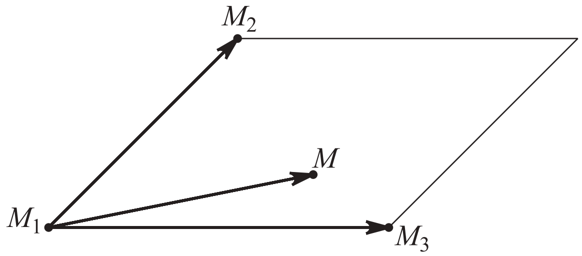

Figure M4.20. Plane passing through three points.

Remark 1. Equation (M4.5.1.3) means that the vectors $\overrightarrow{M_1M}$, $\overrightarrow{M_1M_2}$, and $\overrightarrow{M_1M_3}$ are coplanar.

Remark 2. If the three points $M_1(x_1, y_1, z_1)$, $M_2(x_2, y_2, z_2)$, and $M_3(x_3, y_3, z_3)$ are collinear, then equation (M4.5.1.3) is satisfied identically.

Example 1. Let us construct an equation of the plane passing through the three points $M_1(1, 1, 1)$, $M_2(2, 2, 1)$, and $M_3(1, 2, 2)$.

Obviously, the points M_1, M_2, and M_3 are not collinear, since the vectors $\overrightarrow{M_1M_2} = (1, 1, 0)$ and $\overrightarrow{M_1M_3} = (0, 1, 1)$ are not collinear. According to (M4.5.1.3), we have

$$\begin{vmatrix} x-1 & y-1 & z-1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{vmatrix} = 0,$$

whence the desired equation is $x - y + z - 1 = 0$.

► **Equation of the plane passing through two points and parallel to a straight line.** The plane passing through two points $M_1(x_1, y_1, z_1)$ and $M_2(x_2, y_2, z_2)$ and parallel to a straight line with direction vector $\mathbf{R} = (l, m, n)$ (see Fig. M4.21) is given by the equation

$$\begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ x_2-x_1 & y_2-y_1 & z_2-z_1 \\ l & m & n \end{vmatrix} = 0, \qquad \text{or} \qquad \big[(\mathbf{r}-\mathbf{r}_1)(\mathbf{r}_2-\mathbf{r}_1)\mathbf{R}\big] = 0, \qquad \text{(M4.5.1.4)}$$

where $\mathbf{r}$, $\mathbf{r}_1$, and $\mathbf{r}_2$ are the position vectors of the points $M(x, y, z)$, $M_1(x_1, y_1, z_1)$, and $M_2(x_2, y_2, z_2)$, respectively.

Remark. If the vectors $\overrightarrow{M_1M_2}$ and $\mathbf{R}$ are collinear, then equations (M4.5.1.4) become identities.

Example 2. Find an equation of the plane passing through the points $M_1(0, 1, 0)$ and $M_2(1, 1, 1)$ and parallel to the straight line with direction vector $\mathbf{R} = (0, 1, 1)$.

According to (M4.5.1.4), we have

$$\begin{vmatrix} x-0 & y-1 & z-0 \\ 1-0 & 1-1 & 1-0 \\ 0 & 1 & 1 \end{vmatrix} = 0,$$

whence the desired equation is $-x - y + z + 1 = 0$.

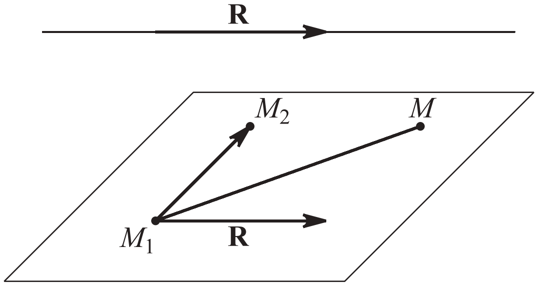

Figure M4.21. Plane passing through two points and parallel to a line.

► **Equation of the plane passing through a point and parallel to two straight lines.** The plane passing through a point $M_0(x_0, y_0, z_0)$ and parallel to two straight lines with direction vectors $\mathbf{R}_1 = (l_1, m_1, n_1)$ and $\mathbf{R}_2 = (l_2, m_2, n_2)$ is given by the equation

$$\begin{vmatrix} x-x_0 & y-y_0 & z-z_0 \\ l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \end{vmatrix} = 0, \qquad \text{or} \qquad \bigl[(\mathbf{r}-\mathbf{r}_0)\mathbf{R}_1\mathbf{R}_2\bigr] = 0,$$

where $\mathbf{r}$ and $\mathbf{r}_0$ are the position vectors of the points $M(x, y, z)$ and $M_0(x_0, y_0, z_0)$, respectively.

The equation of the plane passing through a point $M_0(x_0, y_0, z_0)$ and parallel to two noncollinear vectors $\mathbf{R}_1 = (l_1, m_1, n_1)$ and $\mathbf{R}_2 = (l_2, m_2, n_2)$ can be represented in the form (M4.5.1.2) with A, B, and C being the coordinates of the vector $\mathbf{R} = \mathbf{R}_1 \times \mathbf{R}_2$.

Example 3. Find an equation of the plane P that passes through the point $M_0(2, -1, 1)$ and is perpendicular to the planes P_1 and P_2 defined by $3x + 2y - z + 4 = 0$ and $x + y + z - 3 = 0$.

The vectors $\mathbf{N}_1 = (3, 2, -1)$ and $\mathbf{N}_2 = (1, 1, 1)$ are normal to P_1 and P_2 and parallel to P. Their cross product is

$$\mathbf{N} = \mathbf{N}_1 \times \mathbf{N}_2 = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 3 & 2 & -1 \\ 1 & 1 & 1 \end{vmatrix} = 3\mathbf{i} - 4\mathbf{j} + 1\mathbf{k}.$$

The vector $\mathbf{N}$ is perpendicular to the desired plane P, which therefore satisfies the equation

$$3(x-2) - 4(y+1) + (z-1) = 0 \qquad \text{or} \qquad 3x - 4y + z - 11 = 0.$$

► **Equation of the plane passing through two points and perpendicular to a given plane.** The plane passing through two points $M_1(x_1, y_1, z_1)$ and $M_2(x_2, y_2, z_2)$ and perpendicular to the plane $Ax + By + Cz + D = 0$ (see Fig. M4.22) is determined by the equation

$$\begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ x_2-x_1 & y_2-y_1 & z_2-z_1 \\ A & B & C \end{vmatrix} = 0, \qquad \text{or} \qquad \bigl[(\mathbf{r}-\mathbf{r}_1)(\mathbf{r}_2-\mathbf{r}_1)\mathbf{N}\bigr] = 0, \qquad \text{(M4.5.1.5)}$$

where $\mathbf{r}$, $\mathbf{r}_1$, and $\mathbf{r}_2$ are the position vectors of the points $M(x, y, z)$, $M_1(x_1, y_1, z_1)$, and $M_2(x_2, y_2, z_2)$, respectively.

Remark. If the straight line passing through the points $M_1(x_1, y_1, z_1)$ and $M_2(x_2, y_2, z_2)$ is perpendicular to the original plane, then the desired plane is undetermined and equations (M4.5.1.5) become identities.

► **Equation of the plane passing through a point and perpendicular to two planes.** The plane passing through a point $M_1(x_1, y_1, z_1)$ and perpendicular to two (nonparallel)

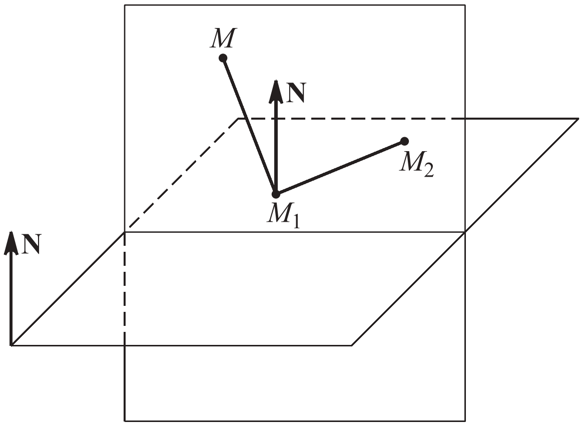

Figure M4.22. Plane passing through two points and perpendicular to a given plane.

planes $A_1x+B_1y+C_1z+D_1=0$ and $A_2x+B_2y+C_2z+D_2=0$ (see Fig. M4.23) is given by the equation

$$\begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{vmatrix}=0, \qquad \text{or} \qquad \big[(\mathbf{r}-\mathbf{r}_1)\mathbf{N}_1\mathbf{N}_2\big]=0, \qquad \text{(M4.5.1.6)}$$

where $\mathbf{N}_1=(A_1,B_1,C_1)$ and $\mathbf{N}_2=(A_2,B_2,C_2)$ are normals to the given planes and $\mathbf{r}$ and $\mathbf{r}_1$ are the position vectors of the points $M(x,y,z)$ and $M_1(x_1,y_1,z_1)$, respectively.

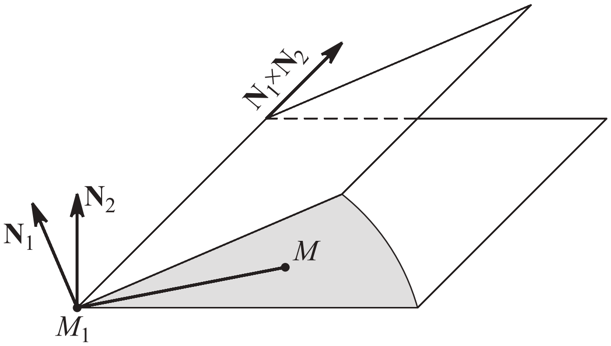

Figure M4.23. Plane passing through a point and perpendicular to two planes.

Remark 1. Equations (M4.5.1.6) mean that the vectors $\overrightarrow{M_1M}$, $\mathbf{N}_1$, and $\mathbf{N}_2$ are coplanar.

Remark 2. If the original planes are parallel, then the desired plane is undetermined. In this case, equations (M4.5.1.6) become identities.

Example 4. Let us find an equation of the plane passing through the point $M_1(0,1,2)$ and perpendicular to the planes $x-y+z-3=0$ and $-x+y+z+4=0$.

According to (M4.5.1.6), we have

$$\begin{vmatrix} x-0 & y-1 & z-2 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \end{vmatrix}=0,$$

whence the desired equation is $x+y-1=0$.

► **Equation of planes passing through the line of intersection of planes.** The planes passing through the line of intersection of the planes $A_1x + B_1y + C_1z + D_1 = 0$ and $A_2x + B_2y + C_2z + D_2 = 0$ are given by the equation

$$\alpha(A_1x + B_1y + C_1z + D_1) + \beta(A_2x + B_2y + C_2z + D_2) = 0,$$

which is called the *equation of a pencil of planes*. Here α and β are arbitrary parameters ($\alpha^2 + \beta^2 \neq 0$).

M4.5.2. Line in Space

► **Parametric equations of a straight line.** The *parametric equations of the line* that passes through a point $M_1(x_1, y_1, z_1)$ and is parallel to a direction vector $\mathbf{R} = (l, m, n)$ (see Fig. M4.24) are

$$x = x_1 + lt, \quad y = y_1 + mt, \quad z = z_1 + nt, \qquad \text{or} \qquad \mathbf{r} = \mathbf{r}_1 + t\mathbf{R}, \qquad \text{(M4.5.2.1)}$$

where $\mathbf{r} = \overrightarrow{OM}$ and $\mathbf{r}_1 = \overrightarrow{OM}_1$. As the parameter t varies from $-\infty$ to $+\infty$, the point M with position vector $\mathbf{r} = (x, y, z)$ determined by formula (M4.5.2.1) runs over the entire straight line in question. It is convenient to use parametric equations (M4.5.2.1) if one needs to find the point of intersection of a straight line with a plane.

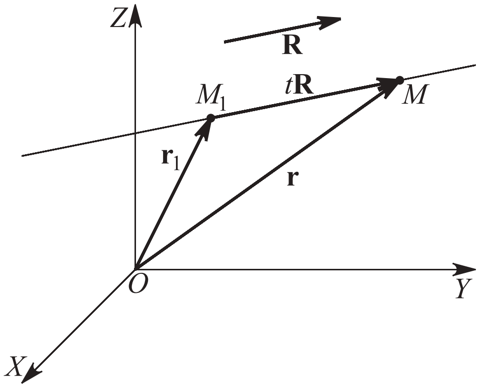

Figure M4.24. Straight line passing through a point and parallel to a direction vector.

The numbers l, m, and n characterize the direction of the straight line in space; they are called the *direction coefficients* of the straight line. For a unit vector $\mathbf{R} = \mathbf{R}^0$, the coefficients l, m, n are the cosines of the angles α, β, and γ formed by this straight line (the direction vector $\mathbf{R}^0$) with the coordinate axes OX, OY, and OZ. These cosines can be expressed via the coordinates of the direction vector $\mathbf{R}$ as

$$\cos\alpha = \frac{l}{\sqrt{l^2 + m^2 + n^2}}, \quad \cos\beta = \frac{m}{\sqrt{l^2 + m^2 + n^2}}, \quad \cos\gamma = \frac{n}{\sqrt{l^2 + m^2 + n^2}}.$$

► **Canonical equations of a straight line.** The equations

$$\frac{x - x_1}{l} = \frac{y - y_1}{m} = \frac{z - z_1}{n}, \qquad \text{or} \qquad (\mathbf{r} - \mathbf{r}_1) \times \mathbf{R} = 0, \qquad \text{(M4.5.2.2)}$$

are called the *canonical equations of the straight line* through the point $M_1(x_1, y_1, z_1)$ with the position vector $\mathbf{r}_1 = (x_1, y_1, z_1)$ and parallel to the direction vector $\mathbf{R} = (l, m, n)$.

Remark 1. One can obtain canonical equations (M4.5.2.2) from parametric equations (M4.5.2.1) by eliminating the parameter t.

Remark 2. In the canonical equations, all coefficients l, m, and n cannot be zero simultaneously, since $|\mathbf{R}| \neq 0$. But some of them may be zero. If one of the denominators in equations (M4.5.2.2) is zero, this means that the corresponding numerator is also zero.

► **General equation of a straight line.** The *general equation of a straight line in space* defines it as the line of intersection of two planes (see Fig. M4.25) and is given analytically by a system of two linear equations

$$\begin{aligned} A_1x + B_1y + C_1z + D_1 &= 0,\\ A_2x + B_2y + C_2z + D_2 &= 0. \end{aligned} \qquad \text{(M4.5.2.3)}$$

The normals to the planes are $\mathbf{N}_1 = (A_1, B_1, C_1)$ and $\mathbf{N}_2 = (A_2, B_2, C_2)$. The direction vector $\mathbf{R}$ is equal to the cross product of the normals $\mathbf{N}_1$ and $\mathbf{N}_2$; i.e.,

$$\mathbf{R} = \mathbf{N}_1 \times \mathbf{N}_2, \qquad \text{(M4.5.2.4)}$$

and its coordinates l, m, and n can be obtained by the formulas

$$l = \begin{vmatrix} B_1 & C_1 \\ B_2 & C_2 \end{vmatrix}, \quad m = \begin{vmatrix} C_1 & A_1 \\ C_2 & A_2 \end{vmatrix}, \quad n = \begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix}.$$

Remark 1. Simultaneous equations of the form (M4.5.2.3) define a straight line if and only if the coefficients A_1, B_1, and C_1 in one of them are not proportional to the respective coefficients A_2, B_2, and C_2 in the other.

Remark 2. For $D_1 = D_2 = 0$ (and only in this case), the line passes through the origin.

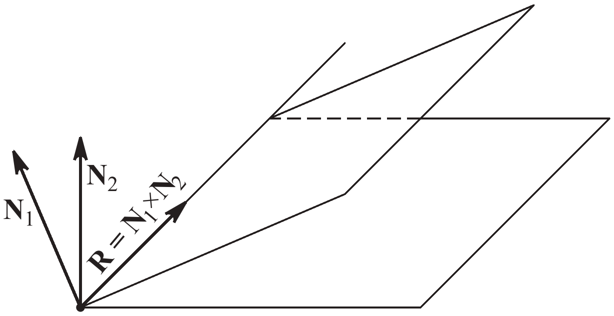

Figure M4.25. Straight line as intersection of two planes.

Example. Let us reduce the equation of the straight line

$$x + 2y - z + 1 = 0, \quad x - y + z + 3 = 0$$

to canonical form.

We choose one of the coordinates arbitrarily; say, $x = 0$. Then

$$2y - z + 1 = 0, \quad -y + z + 3 = 0,$$

and hence $y = -4$, $z = -7$. Thus the desired line contains the point $M(0, -4, -7)$. We find the cross product of the vectors $\mathbf{N}_1 = (1, 2, -1)$ and $\mathbf{N}_2 = (1, -1, 1)$ and, according to (M4.5.2.4), obtain the direction vector $\mathbf{R} = (1, -2, -3)$ of the desired line. Therefore, with (M4.5.2.2) taken into account, the equations of the line become

$$\frac{x}{1} = \frac{y+4}{-2} = \frac{z+7}{-3}.$$

► **Equations of a straight line passing through two points.** The canonical equations of the straight line (see Fig. M4.26) passing through two points $M_1(x_1, y_1, z_1)$ and $M_2(x_2, y_2, z_2)$ are

$$\frac{x-x_1}{x_2-x_1} = \frac{y-y_1}{y_2-y_1} = \frac{z-z_1}{z_2-z_1}, \qquad \text{or} \qquad (\mathbf{r}-\mathbf{r}_1)\times(\mathbf{r}_2-\mathbf{r}_1) = 0, \tag{M4.5.2.5}$$

where $\mathbf{r}$, $\mathbf{r}_1$, and $\mathbf{r}_2$ are the position vectors of the points $M(x, y, z)$, $M_1(x_1, y_1, z_1)$, and $M_2(x_2, y_2, z_2)$, respectively.

The parametric equations of this line are

$$\begin{aligned} x &= x_1(1-t) + x_2 t,\\ y &= y_1(1-t) + y_2 t,\\ z &= z_1(1-t) + z_2 t, \end{aligned} \qquad \text{or} \qquad \mathbf{r} = (1-t)\mathbf{r}_1 + t\mathbf{r}_2. \tag{M4.5.2.6}$$

Remark. Eliminating the parameter t from equations (M4.5.2.6), we obtain equations (M4.5.2.5).

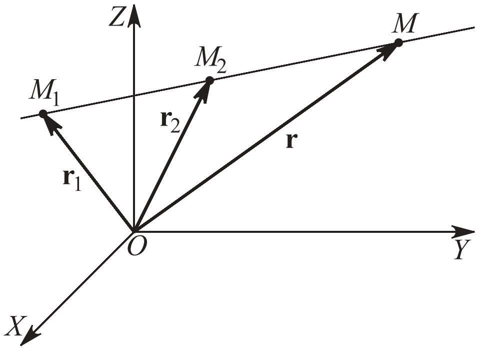

Figure M4.26. Straight line passing through two points.

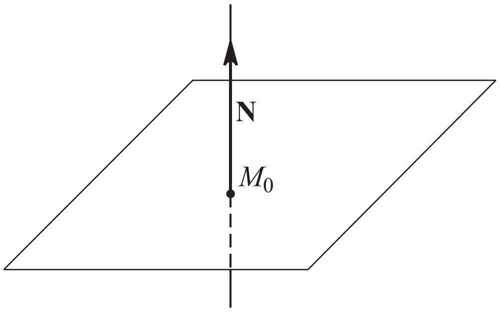

Figure M4.27. Straight line passing through a point and perpendicular to a plane.

► **Equations of a straight line passing through a point and perpendicular to a plane.** The equations of the straight line passing through a point $M_0(x_0, y_0, z_0)$ and perpendicular to the plane given by the equation $Ax + By + Cz + D = 0$ (see Fig. M4.27) are

$$\frac{x-x_0}{A} = \frac{y-y_0}{B} = \frac{z-z_0}{C}.$$

M4.5.3. Mutual Arrangement of Points, Lines, and Planes

► **Angles between lines in space.** Consider two straight lines determined by vector parametric equations $\mathbf{r} = \mathbf{r}_1 + t\mathbf{R}_1$ and $\mathbf{r} = \mathbf{r}_2 + t\mathbf{R}_2$. The angle φ between these lines (see Fig. M4.28) can be obtained from the formulas

$$\cos\varphi = \frac{\mathbf{R}_1\cdot\mathbf{R}_2}{|\mathbf{R}_1|\,|\mathbf{R}_2|}, \qquad \sin\varphi = \frac{|\mathbf{R}_1\times\mathbf{R}_2|}{|\mathbf{R}_1|\,|\mathbf{R}_2|}.$$

If the lines are given by the canonical equations

$$\frac{x-x_1}{l_1} = \frac{y-y_1}{m_1} = \frac{z-z_1}{n_1} \quad \text{and} \quad \frac{x-x_2}{l_2} = \frac{y-y_2}{m_2} = \frac{z-z_2}{n_2}, \tag{M4.5.3.1}$$

then the angle φ between the lines can be found from the formulas

$$\cos\varphi=\frac{l_1l_2+m_1m_2+n_1n_2}{\sqrt{l_1^2+m_1^2+n_1^2}\sqrt{l_2^2+m_2^2+n_2^2}},$$
$$\sin\varphi=\frac{\sqrt{\begin{vmatrix} m_1 & n_1\\ m_2 & n_2\end{vmatrix}^2+\begin{vmatrix} n_1 & l_1\\ n_2 & l_2\end{vmatrix}^2+\begin{vmatrix} l_1 & m_1\\ l_2 & m_2\end{vmatrix}^2}}{\sqrt{l_1^2+m_1^2+n_1^2}\sqrt{l_2^2+m_2^2+n_2^2}}. \qquad \text{(M4.5.3.2)}$$

Example 1. Let us find the angle between the lines

$$\frac{x}{1}=\frac{y-2}{2}=\frac{z+1}{2} \quad\text{and}\quad \frac{x}{0}=\frac{y-2}{3}=\frac{z+1}{4}.$$

Using the first formula in (M4.5.3.2), we obtain

$$\cos\varphi=\frac{1\cdot0+2\cdot3+2\cdot4}{\sqrt{1^2+2^2+2^2}\,\sqrt{0^2+3^2+4^2}}=\frac{14}{15},$$

and hence $\varphi\approx0.3672$ rad.

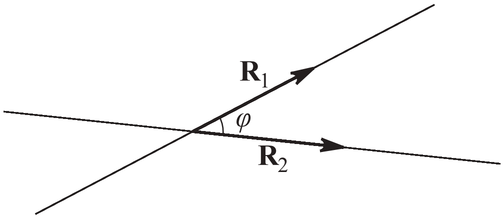

Figure M4.28. Angle between two lines in space.

► **Conditions for two lines to be parallel.** Two straight lines given by vector parametric equations $\mathbf{r}=\mathbf{r}_1+t\mathbf{R}_1$ and $\mathbf{r}=\mathbf{r}_2+t\mathbf{R}_2$ are parallel if

$$\mathbf{R}_2=\lambda\mathbf{R}_1 \qquad\text{or}\qquad \mathbf{R}_2\times\mathbf{R}_1=0,$$

i.e., if their direction vectors $\mathbf{R}_1$ and $\mathbf{R}_2$ are collinear. This can be written as

$$\frac{l_1}{l_2}=\frac{m_1}{m_2}=\frac{n_1}{n_2}.$$

Remark. If parallel lines have a common point (i.e., $\mathbf{r}_1=\mathbf{r}_2$ in parametric equations), then they coincide.

► **Conditions for two lines to be perpendicular.** Two straight lines given by vector parametric equations $\mathbf{r}=\mathbf{r}_1+t\mathbf{R}_1$ and $\mathbf{r}=\mathbf{r}_2+t\mathbf{R}_2$ are perpendicular if

$$\mathbf{R}_1\cdot\mathbf{R}_2=0. \qquad \text{(M4.5.3.3)}$$

This condition can be written as

$$l_1l_2+m_1m_2+n_1n_2=0. \qquad \text{(M4.5.3.4)}$$

Example 2. Let us show that the lines

$$\frac{x-1}{2}=\frac{y-3}{1}=\frac{z}{2} \quad\text{and}\quad \frac{x-2}{1}=\frac{y+1}{2}=\frac{z}{-2}$$

are perpendicular.

Indeed, condition (M4.5.3.4) is satisfied,

$$2\cdot1+1\cdot2+2\cdot(-2)=0,$$

and hence the lines are perpendicular.

► **Theorem on the arrangement of two lines in space.** *Two straight lines in space can:*
a) *be skew;*
b) *lie in the same plane and not meet each other, i.e., be parallel;*
c) *meet at a point;*
d) *coincide.*

A general characteristic of all four cases is the determinant of the matrix

$$\begin{pmatrix} x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \end{pmatrix}, \qquad \text{(M4.5.3.5)}$$

whose entries are taken from the canonical equations of the lines (M4.5.3.1).

In cases *a*–*d* of the theorem, for the matrix (M4.5.3.5) we have, respectively:
a) the determinant is nonzero;
b) the last two rows are proportional to each other but are not proportional to the first row;
c) the last two rows are not proportional, and the first row is their linear combination;
d) all rows are proportional.

In cases *b*–*d* the determinant is zero.

► **Angle between planes.** Consider two planes given by the general equations

$$\begin{aligned} A_1x + B_1y + C_1z + D_1 &= 0, \\ A_2x + B_2y + C_2z + D_2 &= 0. \end{aligned} \qquad \text{(M4.5.3.6)}$$

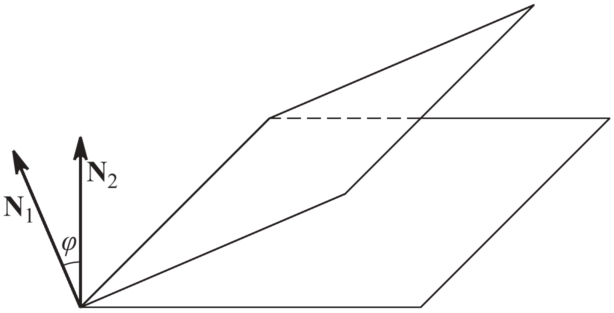

Figure M4.29. Angle between two planes.

The angle between two planes (see Fig. M4.29) is defined as any of the two adjacent dihedral angles formed by the planes (if the planes are parallel, then the angle between them is by definition equal to 0 or π). One of these dihedral angles is equal to the angle φ between the normal vectors $\mathbf{N}_1 = (A_1, B_1, C_1)$ and $\mathbf{N}_2 = (A_2, B_2, C_2)$ to the planes, which can be determined by the formula

$$\cos\varphi = \frac{A_1A_2 + B_1B_2 + C_1C_2}{\sqrt{A_1^2 + B_1^2 + C_1^2}\sqrt{A_2^2 + B_2^2 + C_2^2}} = \frac{\mathbf{N}_1 \cdot \mathbf{N}_2}{|\mathbf{N}_1|\,|\mathbf{N}_2|}.$$

► **Conditions for two planes to be parallel.** Two planes given by the general equations (M4.5.3.6) are parallel if and only if the following *condition for the planes to be parallel* is satisfied:

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2} \neq \frac{D_1}{D_2};$$

in this case, the planes do not coincide.

Two planes coincide if they are parallel and have a common point. Two planes given by the general equations (M4.5.3.6) coincide if and only if the following *condition for the planes to coincide* is satisfied:

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2} = \frac{D_1}{D_2}.$$

► **Conditions for two planes to be perpendicular.** Planes are perpendicular if their normals are perpendicular. Two planes determined by the general equations (M4.5.3.6) are perpendicular if and only if the following *condition for the planes to be perpendicular* is satisfied:

$$A_1A_2 + B_1B_2 + C_1C_2 = 0 \qquad \text{or} \qquad \mathbf{N}_1\cdot\mathbf{N}_2 = 0, \tag{M4.5.3.7}$$

where $\mathbf{N}_1 = (A_1, B_1, C_1)$ and $\mathbf{N}_2 = (A_2, B_2, C_2)$ are the normals to the planes.

Example 3. Let us show that the planes $x - y + z = 0$ and $x - y - 2z + 5 = 0$ are perpendicular.
Since condition (M4.5.3.7) is satisfied,

$$1\cdot 1 + (-1)\cdot(-1) + 1\cdot(-2) = 0,$$

we see that the planes are perpendicular.

► **Angle between a straight line and a plane.** Consider a plane given by the general equation

$$Ax + By + Cz + D = 0 \tag{M4.5.3.8}$$

and a line given by the canonical equations

$$\frac{x - x_1}{l} = \frac{y - y_1}{m} = \frac{z - z_1}{n}. \tag{M4.5.3.9}$$

The angle between the line and the plane (see Fig. M4.30) is defined as the complementary angle θ of the angle φ between the direction vector $\mathbf{R} = (l, m, n)$ of the line and the normal $\mathbf{N} = (A, B, C)$ to the plane. For this angle, one has the formula

$$\sin\theta = |\cos\varphi| = \frac{|Al + Bm + Cn|}{\sqrt{A^2 + B^2 + C^2}\sqrt{l^2 + m^2 + n^2}} = \frac{|\mathbf{N}\cdot\mathbf{R}|}{|\mathbf{N}|\,|\mathbf{R}|}.$$

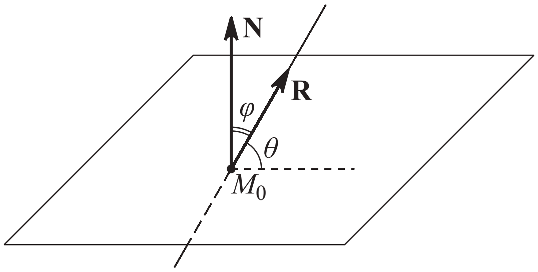

Figure M4.30. Angle between a straight line and a plane.

► **Conditions for a straight line and a plane to be parallel.** A plane given by the general equation (M4.5.3.8) and a line given by canonical equations (M4.5.3.9) are parallel if the following two conditions hold:

$$Al + Bm + Cn = 0,$$
$$Ax_1 + By_1 + Cz_1 + D \neq 0.$$

The first condition means that the direction vector of the straight line is perpendicular to the normal to the plane and the second condition means that the line is not contained in the plane.

▶ **Condition for a straight line and a plane to be perpendicular.** A line given by canonical equations (M4.5.3.9) and a plane given by the general equation (M4.5.3.8) are perpendicular if the line is collinear to the normal to the plane (is a normal itself), i.e., if

$$\frac{A}{l}=\frac{B}{m}=\frac{C}{n},\qquad \text{or}\qquad \mathbf{N}=\lambda\mathbf{R}.$$

▶ **Intersection of a straight line and a plane.** Consider a plane given by the general equation (M4.5.3.8) and a straight line given by parametric equations

$$x=x_1+lt,\quad y=y_1+mt,\quad z=z_1+nt.$$

The coordinates of the point $M_0(x_0,y_0,z_0)$ of intersection of the line with the plane (see Fig. M4.30), if the point exists at all, are determined by the formulas

$$x_0=x_1+lt_0,\quad y_0=y_1+mt_0,\quad z_0=z_1+nt_0,$$

where

$$t_0=-\frac{Ax_1+By_1+Cz_1+D}{Al+Bm+Cn}.$$

▶ **Distance from a point to a plane.** The distance from a point $M_0(x_0,y_0,z_0)$ to a plane given by the general equation (M4.5.1.1) is determined by the formula

$$d=\frac{|Ax_0+By_0+Cz_0+D|}{\sqrt{A^2+B^2+C^2}}.$$

▶ **Distance between two parallel planes.** We consider two parallel planes given by the general equations $Ax+By+Cz+D_1=0$ and $Ax+By+Cz+D_2=0$. The distance between them is

$$d=\frac{|D_1-D_2|}{\sqrt{A^2+B^2+C^2}}.$$

▶ **Distance from a point to a straight line.** The distance from a point $M_0(x_0,y_0,z_0)$ to a line given by canonical equations (M4.5.2.2) is determined by the formula

$$d=\frac{\sqrt{\begin{vmatrix} m & n\\ y_1-y_0 & z_1-z_0\end{vmatrix}^2+\begin{vmatrix} n & l\\ z_1-z_0 & x_1-x_0\end{vmatrix}^2+\begin{vmatrix} l & m\\ x_1-x_0 & y_1-y_0\end{vmatrix}^2}}{\sqrt{l^2+m^2+n^2}}.$$

▶ **Distance between straight lines.** Consider two nonparallel lines given in the canonical form

$$\frac{x-x_1}{l_1}=\frac{y-y_1}{m_1}=\frac{z-z_1}{n_1},$$
$$\frac{x-x_2}{l_2}=\frac{y-y_2}{m_2}=\frac{z-z_2}{n_2}.$$

The distance between them can be calculated by the formula

$$d=\frac{\pm\begin{vmatrix} x_1-x_2 & y_1-y_2 & z_1-z_2\\ l_1 & m_1 & n_1\\ l_2 & m_2 & n_2\end{vmatrix}}{\sqrt{\begin{vmatrix} l_1 & m_1\\ l_2 & m_2\end{vmatrix}^2+\begin{vmatrix} m_1 & n_1\\ m_2 & n_2\end{vmatrix}^2+\begin{vmatrix} n_1 & l_1\\ n_2 & l_2\end{vmatrix}^2}}\tag{M4.5.3.10}$$

(minus sign should be taken if the determinant is negative). The condition that the determinant in the numerator in (M4.5.3.10) is zero is the *condition for the two lines in space to meet*.

M4.6. Quadric Surfaces (Quadrics)

M4.6.1. Quadrics and Their Canonical Equations

► **Central surfaces.** A segment joining two points of a surface is called a *chord*. If there exists a point in space, not necessarily lying on the surface, that bisects all chords passing through it, then the surface is said to be *central* and the point is called the *center* of the surface.

The equations listed below for central surfaces are given in *canonical form*; i.e., the center of a surface is at the origin, and the surface symmetry axes are the coordinate axes. Moreover, the coordinate planes are symmetry planes.

► **Ellipsoid.** An *ellipsoid* is a central surface defined by the equation

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1, \tag{M4.6.1.1}$$

where the numbers a, b, and c are the lengths of the segments called the *semiaxes* of the ellipsoid (see Fig. M4.31 *a*). The coordinates of all points of the ellipsoid satisfy the inequalities $-a \le x \le a$, $-b \le y \le b$, and $-c \le z \le c$.

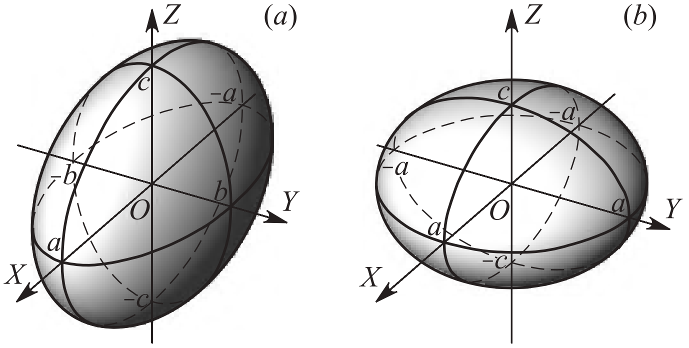

Figure M4.31. Triaxial ellipsoid (*a*) and spheroid (*b*).

If $a \ne b \ne c$, then the ellipsoid is said to be *triaxial*, or *scalene*. If $a = b \ne c$, then the ellipsoid is called a *spheroid*; it can be obtained by rotating the ellipse $x^2/a^2 + z^2/c^2 = 1$, $y = 0$ lying in the plane OXZ about the axis OZ (see Fig. M4.31 *b*). If $a = b > c$, then the ellipsoid is an *oblate spheroid*, and if $a = b < c$, then the ellipsoid is a *prolate spheroid*. If $a = b = c$, then the ellipsoid is the sphere of radius a given by the equation $x^2 + y^2 + z^2 = a^2$.

An arbitrary plane section of an ellipsoid is an ellipse (or, in a special case, a circle).

The volume of an ellipsoid is equal to $V = \frac{4}{3}\pi abc$.

Remark. About the sphere, see also Subsection M3.2.3.

► **Hyperboloids.** A *one-sheeted hyperboloid* is a central surface defined by the equation

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1, \tag{M4.6.1.2}$$

where a and b are the *real semiaxes* and c is the *imaginary semiaxis* (see Fig. M4.32 *a*).

A *two-sheeted hyperboloid* is a central surface defined by the equation

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1, \tag{M4.6.1.3}$$

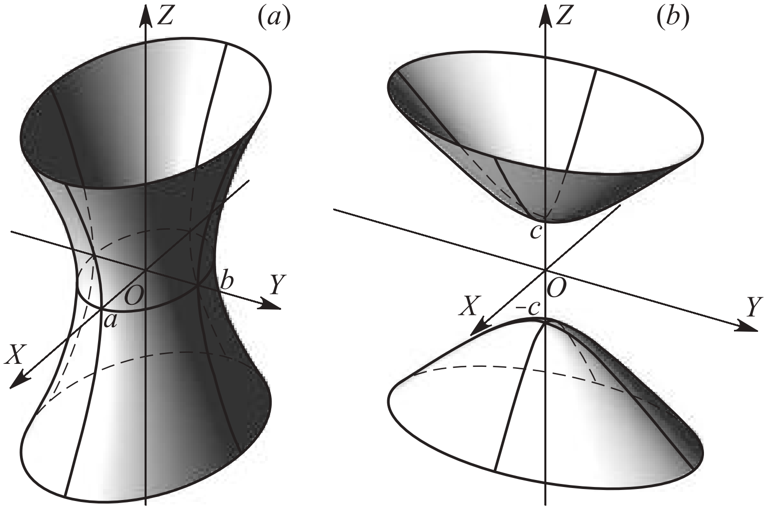

Figure M4.32. One-sheeted (*a*) and two-sheeted (*b*) hyperboloids.

where c is the *real semiaxis* and a and b are the *imaginary semiaxes* (see Fig. M4.32 *b*). A two-sheeted hyperboloid consists of two parts whose points lie at $z \le -c$ and $z \ge c$.

A hyperboloid approaches the surface

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0,$$

which is called an *asymptotic cone*, infinitely closely.

A plane passing through the axis OZ intersects each of the hyperboloids (M4.6.1.2) and (M4.6.1.3) in two hyperbolas and the asymptotic cone in two straight lines, which are the asymptotes of these hyperbolas. The section of a hyperboloid by a plane parallel to OXY is an ellipse. The section of a one-sheeted hyperboloid by the plane $z = 0$ is an ellipse, which is called the *gorge* or *throat* ellipse.

For $a = b$, we deal with the *hyperboloid of revolution* obtained by rotating a hyperbola with semiaxes a and c about its focal axis $2c$ (which is an imaginary axis for a one-sheeted hyperboloid and a real axis for a two-sheeted hyperboloid). If $a = b = c$, then the hyperboloid of revolution is said to be *right*, and its sections by the planes OXZ and OYZ are equilateral hyperbolas.

► **Cone.** A *cone* is a central surface defined by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0. \qquad \text{(M4.6.1.4)}$$

The cone (see Fig. M4.33) defined by (M4.6.1.4) has vertex at the origin, and for its base we can take the ellipse with semiaxes a and b in the plane perpendicular to the axis OZ at the distance c from the origin. This cone is the asymptotic cone for the hyperboloids (M4.6.1.2) and (M4.6.1.3). For $a = b$, we obtain a *right circular cone*.

Remark. About the cone, see also Subsection M3.2.3.

► **Paraboloids.** In contrast to the surfaces considered above, paraboloids are not central surfaces. For the equations listed below, the vertex of a paraboloid lies at the origin, the axis OZ is the symmetry axis, and the planes OXZ and OYZ are symmetry planes.

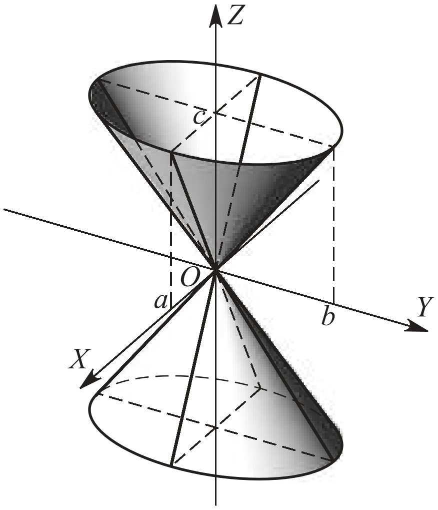

Figure M4.33. A cone.

An *elliptic paraboloid* (see Fig. M4.34 *a*) is a noncentral surface defined by the equation

$$\frac{x^2}{p} + \frac{y^2}{q} = 2z, \qquad \text{(M4.6.1.5)}$$

where $p > 0$ and $q > 0$ are parameters. All points of an elliptic paraboloid lie in the domain $z \geq 0$.

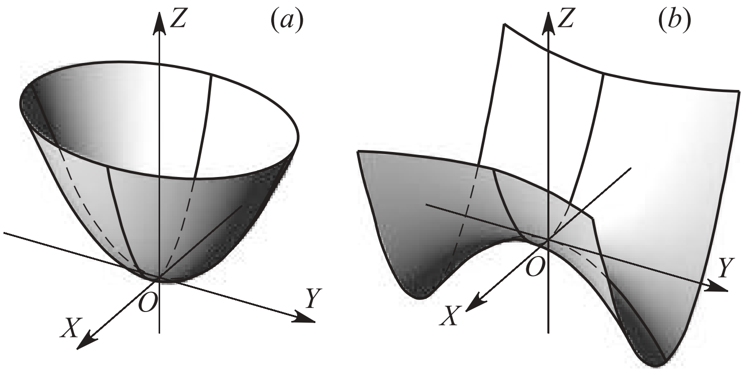

Figure M4.34. Elliptic (*a*) and hyperbolic (*b*) paraboloids.

The sections of an elliptic paraboloid by planes parallel to the axis OZ are parabolas, and the sections by planes parallel to the plane OXY are ellipses. If $p = q$, then we have a *paraboloid of revolution*, which is obtained by rotating the parabola $2pz = x^2$ lying in the plane OXZ about its axis.

The volume of the part of an elliptic paraboloid cut by the plane perpendicular to its axis at a height h is equal to $V = \frac{1}{2}\pi abh$, i.e., half the volume of the elliptic cylinder with the same base and altitude.

A *hyperbolic paraboloid* (see Fig. M4.34 *b*) is a noncentral surface defined by the equation

$$\frac{x^2}{p} - \frac{y^2}{q} = 2z, \tag{M4.6.1.6}$$

where $p > 0$ and $q > 0$ are parameters.

The sections of a hyperbolic paraboloid by planes parallel to the axis OZ are parabolas, and the sections by planes parallel to the plane OXY are hyperbolas.

M4.6.2. Quadrics (General Theory)

► **General equation of a quadric. Translation and rotation.** A *quadric* is a set of points in three-dimensional space whose coordinates in the rectangular Cartesian coordinate system satisfy a second-order algebraic equation

$$a_{11}x^2+a_{22}y^2+a_{33}z^2+2a_{12}xy+2a_{13}xz+2a_{23}yz+2a_{14}x+2a_{24}y+2a_{34}z+a_{44}=0, \tag{M4.6.2.1}$$

or

$$\begin{aligned}&(a_{11}x + a_{12}y + a_{13}z + a_{14})x + (a_{21}x + a_{22}y + a_{23}z + a_{24})y\\ &\qquad + (a_{31}x + a_{32}y + a_{33}z + a_{34})z + a_{41}x + a_{42}y + a_{43}z + a_{44} = 0,\end{aligned}$$

with symmetric coefficients, $a_{ij} = a_{ji}$ $(i, j = 1, 2, 3, 4)$; the factors 2 appearing in some terms are introduced for further convenience. If equation (M4.6.2.1) does not define a real geometric object, then one says that this equation defines an *imaginary quadric*.

Equation (M4.6.2.1) can be simplified using the transformations of translation and rotation.

1. Translation:

$$x = \bar{x} + x_0, \quad y = \bar{y} + y_0, \quad z = \bar{z} + z_0. \tag{M4.6.2.2}$$

This transformation means that the origin $O(0, 0, 0)$ is translated to the point $\bar{O}(x_0, y_0, z_0)$ with the new axes of coordinates remaining parallel to the original ones; $\bar{x}$, $\bar{y}$, and $\bar{z}$ are the new coordinates.

2. Rotation:

$$x = e_{11}\hat{x} + e_{12}\hat{y} + e_{13}\hat{z}, \quad y = e_{21}\hat{x} + e_{22}\hat{y} + e_{23}\hat{z}, \quad z = e_{31}\hat{x} + e_{32}\hat{y} + e_{33}\hat{z}. \tag{M4.6.2.3}$$

This transformation means that all points are rotated about the origin O, with e_{11}, e_{21}, e_{31} being the direction cosines of the axis $O\widehat{X}$, e_{12}, e_{22}, e_{32} those of the axis $O\widehat{Y}$, and e_{13}, e_{23}, e_{33} those of the axis $O\widehat{Z}$ in the initial coordinate system $OXYZ$.

► **Classification of quadrics.** With successive application of transformations (M4.6.2.2)–(M4.6.2.3), equation (M4.6.2.1) can be reduced to one of the following 17 *canonical forms*, each of which is associated with a certain class of quadrics (see Table M4.2). The first six surfaces, with $\delta \neq 0$, are *nondegenerate*; their canonical equations contain three quadratic terms proportional to x^2, y^2, and z^2. The other surfaces, 7–17, with $\delta = 0$, are *degenerate*; their canonical equations contain only two (proportional to x^2 and y^2) or even one (x^2) quadratic term. The last five surfaces, 13–17, disintegrate into planes (real or imaginary); their equations can be represented as the product of two factors linear in coordinates on the left-hand side and zero on the right-hand side.

TABLE M4.2
Classification of quadrics.

No.	Surface	Canonical equation	Conditions for invariants
1	Ellipsoid	$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1$	$\delta\neq 0,\ \Delta<0,\ S\delta>0,\ T>0$
2	Imaginary ellipsoid	$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=-1$	$\delta\neq 0,\ \Delta>0,\ S\delta>0,\ T>0$
3	Imaginary cone with real vertex	$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=0$	$\delta\neq 0,\ \Delta=0,\ S\delta>0,\ T>0$
4	One-sheeted hyperboloid	$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1$	$\delta\neq 0,\ \Delta>0,$ $S\delta>0$ or $T>0$ (not both >0)
5	Two-sheeted hyperboloid	$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1$	$\delta\neq 0,\ \Delta<0,$ $S\delta>0$ or $T>0$ (not both >0)
6	Real cone	$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=0$	$\delta\neq 0,\ \Delta=0,$ $S\delta>0$ or $T>0$ (not both >0)
7	Elliptic paraboloid	$\frac{x^2}{p}+\frac{y^2}{q}=2z\quad(p,q>0)$	$\delta=0,\ \Delta<0,\ T>0$
8	Hyperbolic paraboloid	$\frac{x^2}{p}-\frac{y^2}{q}=2z\quad(p,q>0)$	$\delta=0,\ \Delta>0,\ T<0$
9	Elliptic cylinder	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$	$\delta=\Delta=0,\ T>0,\ S\sigma<0$
10	Imaginary elliptic cylinder	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=-1$	$\delta=\Delta=0,\ T>0,\ S\sigma>0$
11	Hyperbolic cylinder	$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$	$\delta=\Delta=0,\ T<0,\ \sigma\neq 0$
12	Parabolic cylinder	$y^2=2px$	$\delta=\Delta=0,\ T=0$
13	Pair of real intersecting planes	$\frac{x^2}{a^2}-\frac{y^2}{b^2}=0$	$\delta=\Delta=0,\ T<0,\ \sigma=0$
14	Pair of imaginary planes intersecting in a real straight line	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=0$	$\delta=\Delta=0,\ T>0,\ \sigma=0$
15	Pair of real parallel planes	$x^2=a^2$	$\delta=\Delta=T=0,\ \Sigma<0$
16	Pair of imaginary parallel planes	$x^2=-a^2$	$\delta=\Delta=T=0,\ \Sigma>0$
17	Pair of real coinciding planes	$x^2=0$	$\delta=\Delta=T=0,\ \Sigma=0$

► **Invariants of quadrics.** The shape of a quadric can be identified using four *invariants* and two *semi-invariants* without reducing equation (M4.6.2.1) to canonical form.

The four main invariants are

$$S=a_{11}+a_{22}+a_{33}, \tag{M4.6.2.4}$$

$$T=\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}+\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}+\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}, \tag{M4.6.2.5}$$

$$\delta=\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{vmatrix}, \tag{M4.6.2.6}$$

$$\Delta=\begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12} & a_{22} & a_{23} & a_{24} \\ a_{13} & a_{23} & a_{33} & a_{34} \\ a_{14} & a_{24} & a_{34} & a_{44} \end{vmatrix}, \tag{M4.6.2.7}$$

whose values are preserved under parallel translations and rotations of the coordinate axes.

The semi-invariants are

$$\sigma = \Delta_{11} + \Delta_{22} + \Delta_{33},$$
$$\Sigma = \begin{vmatrix} a_{11} & a_{14} \\ a_{41} & a_{44} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{24} \\ a_{42} & a_{44} \end{vmatrix} + \begin{vmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{vmatrix},$$

whose values are preserved only under rotations of the coordinate axes, with Δ_{ij} being the cofactor of the entry a_{ij} in Δ.

The last column in Table M4.2 allows the classification of the quadrics in accordance with the values of the invariants S, T, δ, and Δ and semi-invariants σ and Σ.

► **Characteristic quadratic form of a quadric.** The *characteristic quadratic form*

$$F(x,y,z) = a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{13}xz + a_{23}yz$$

corresponding to equation (M4.6.2.1) and its *characteristic equation*

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & a_{13} \\ a_{12} & a_{22}-\lambda & a_{23} \\ a_{13} & a_{23} & a_{33}-\lambda \end{vmatrix} = 0, \qquad \text{or} \qquad \lambda^3 - S\lambda^2 + T\lambda - \delta = 0 \qquad \text{(M4.6.2.8)}$$

permit studying the main properties of quadrics.

The roots λ_1, λ_2, and λ_3 of the characteristic equation (M4.6.2.8) are the eigenvalues of the real symmetric matrix $[a_{ij}]$ and hence are always real. The invariants S, T, and δ can be expressed in terms of the roots λ_1, λ_2, and λ_3 as follows:

$$S = \lambda_1 + \lambda_2 + \lambda_3, \quad T = \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3, \quad \delta = \lambda_1\lambda_2\lambda_3.$$

The expressions of the parameters of the main quadrics via the invariants T, δ, and Δ and the roots λ_1, λ_2, and λ_3 of the characteristic equation are listed in Table M4.3.

TABLE M4.3

Expressions of the parameters of the main quadrics via the invariants (M4.6.2.4)–(M4.6.2.7) and the roots of the characteristic equation (M4.6.2.8).

Surface	Canonical equation	Parameters of the quadrics	Remarks
Ellipsoid	$\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$	$a^2 = -\dfrac{1}{\lambda_3}\dfrac{\Delta}{\delta}$, $b^2 = -\dfrac{1}{\lambda_2}\dfrac{\Delta}{\delta}$, $c^2 = -\dfrac{1}{\lambda_1}\dfrac{\Delta}{\delta}$, $\delta = \lambda_1\lambda_2\lambda_3$	$a \ge b \ge c$, $\lambda_1 \ge \lambda_2 \ge \lambda_3 > 0$
One-sheeted hyperboloid	$\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = 1$	$a^2 = -\dfrac{1}{\lambda_2}\dfrac{\Delta}{\delta}$, $b^2 = -\dfrac{1}{\lambda_1}\dfrac{\Delta}{\delta}$, $c^2 = \dfrac{1}{\lambda_3}\dfrac{\Delta}{\delta}$, $\delta = \lambda_1\lambda_2\lambda_3$	$a \ge b$, $\lambda_1 \ge \lambda_2 > 0 > \lambda_3$
Two-sheeted hyperboloid	$\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = -1$	$a^2 = \dfrac{1}{\lambda_3}\dfrac{\Delta}{\delta}$, $b^2 = \dfrac{1}{\lambda_2}\dfrac{\Delta}{\delta}$, $c^2 = -\dfrac{1}{\lambda_1}\dfrac{\Delta}{\delta}$, $\delta = \lambda_1\lambda_2\lambda_3$	$a \ge b$, $\lambda_1 > 0 > \lambda_2 \ge \lambda_3$
Elliptic paraboloid	$\dfrac{x^2}{p} + \dfrac{y^2}{q} = 2z$	$p = \dfrac{1}{\lambda_2}\sqrt{-\dfrac{\Delta}{T}}$, $q = \dfrac{1}{\lambda_1}\sqrt{-\dfrac{\Delta}{T}}$, $T = \lambda_1\lambda_2$	$p > 0$, $q > 0$, $\lambda_1 \ge \lambda_2 > \lambda_3 = 0$
Hyperbolic paraboloid	$\dfrac{x^2}{p} - \dfrac{y^2}{q} = 2z$	$p = \dfrac{1}{\lambda_1}\sqrt{-\dfrac{\Delta}{T}}$, $q = -\dfrac{1}{\lambda_3}\sqrt{-\dfrac{\Delta}{T}}$, $T = \lambda_1\lambda_3$	$p > 0$, $q > 0$, $\lambda_1 > \lambda_2 = 0 > \lambda_3$

Bibliography for Chapter M4

Alexander, D. C. and Koeberlein, G. M., *Elementary Geometry for College Students, 3rd Edition*, Houghton Mifflin Company, Boston, Massachusetts, 2002.

Blau, H. I., *Foundations of Plane Geometry*, Prentice Hall, Englewood Cliffs, New Jersey, 2002.

Chauvenet, W., *A Treatise on Elementary Geometry*, Adamant Media Corporation, Boston, Massachusetts, 2001.

Efimov, N. V., *Higher Geometry*, Mir Publishers, Moscow, 1980.

Fogiel, M. (Editor), *High School Geometry Tutor, 2nd Edition*, Research & Education Association, Englewood Cliffs, New Jersey, 2003.

Fuller, G. and Tarwater, D., *Analytic Geometry, 7th Edition*, Addison Wesley, Boston, Massachusetts, 1993.

Gustafson, R. D. and Frisk, P. D., *Elementary Geometry, 3rd Edition*, Wiley, New York, 1991.

Hartshorne, R., *Geometry: Euclid and Beyond*, Springer, New York, 2005.

Jacobs, H. R., *Geometry, 2nd Edition*, W. H. Freeman & Company, New York, 1987.

Jacobs, H. R., *Geometry: Seeing, Doing, Understanding, 3rd Edition*, W. H. Freeman & Company, New York, 2003.

Jurgensen, R. and Brown, R. G., *Geometry*, McDougal Littell/Houghton Mifflin, Boston, Massachusetts, 2000.

Kay, D., *College Geometry: A Discovery Approach, 2nd Edition*, Addison Wesley, Boston, Massachusetts, 2000.

Kletenik, D. V., *Problems in Analytic Geometry*, University Press of the Pacific, Honolulu, Hawaii, 2002.

Kostrikin, A. I. (Editor), *Exercises in Algebra: A Collection of Exercises in Algebra, Linear Algebra and Geometry*, Gordon & Breach, New York, 1996.

Kostrikin, A. I. and Shafarevich, I. R. (Editors), *Algebra I: Basic Notions of Algebra*, Springer-Verlag, Berlin, 1990.

Leff, L. S., *Geometry the Easy Way, 3rd Edition*, Barron's Educational Series, Hauppauge, New York, 1997.

Moise, E., *Elementary Geometry from an Advanced Standpoint, 3rd Edition*, Addison Wesley, Boston, Massachusetts, 1990.

Musser, G. L., Burger, W. F., and Peterson, B. E., *Mathematics for Elementary Teachers: A Contemporary Approach, 6th Edition*, Wiley, New York, 2002.

Musser, G. L., Burger, W. F., and Peterson, B. E., *Essentials of Mathematics for Elementary Teachers: A Contemporary Approach, 6th Edition*, Wiley, New York, 2003.

Musser, G. L. and Trimpe, L. E., *College Geometry: A Problem Solving Approach with Applications*, Prentice Hall, Englewood Cliffs, New Jersey, 1994.

Postnikov, M. M., *Lectures in Geometry, Semester I, Analytic Geometry*, Mir Publishers, Moscow, 1982.

Privalov, I. I., *Analytic Geometry, 32nd Edition [in Russian]*, Lan, Moscow, 2003.

Riddle, D. R., *Analytic Geometry, 6th Edition*, Brooks Cole, Stamford, 1995.

Roe, J., *Elementary Geometry*, Oxford University Press, Oxford, 1993.

Schultze, A. and Sevenoak, F. L., *Plane and Solid Geometry*, Adamant Media Corporation, Boston, Massachusetts, 2004.

Suetin, P. K., Kostrikin, A. I., and Manin, Yu. I., *Linear Algebra and Geometry*, Gordon & Breach, New York, 1997.

Tussy, A. S. and Gustafson, R. D., *Basic Geometry for College Students: An Overview of the Fundamental Concepts of Geometry*, Thomson-Brooks/Cole, Pacific Grove, California, 2002.

Vygodskii, M. Ya., *Mathematical Handbook: Higher Mathematics*, Mir Publishers, Moscow, 1971.

Weeks, A. W. and Adkins, J. B., *A Course in Geometry: Plane and Solid*, Bates Publishing Company, Sandwich, Massachusetts, 1982.

Woods, F. S., *Higher Geometry: An Introduction to Advanced Methods in Analytic Geometry, Phoenix Edition*, Dover Publications, New York, 2005.

Chapter M5

Algebra

M5.1. Polynomials and Algebraic Equations

M5.1.1. Polynomials and Their Properties

► **Definition of a polynomial.** A *polynomial of degree* n of a scalar variable x is an expression of the form

$$f(x) \equiv a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \quad (a_n \neq 0), \tag{M5.1.1.1}$$

where $a_0, \ldots, a_n$ are real or complex numbers ($n = 0, 1, 2, \ldots$). Polynomials of degree zero are nonzero numbers.

Two polynomials are *equal* if they have the same coefficients of like powers of the variable.

► **Main operations over polynomials.**

1°. The *sum* (*difference*) of two polynomials $f(x)$ of degree n and $g(x)$ of degree m is the polynomial of degree $l \le \max\{n, m\}$ whose coefficient of each power of x is equal to the sum (difference) of the coefficients of the same power of x in $f(x)$ and $g(x)$, i.e., if

$$g(x) \equiv b_m x^m + b_{m-1} x^{m-1} + \cdots + b_1 x + b_0, \tag{M5.1.1.2}$$

then the sum (difference) of polynomials (M5.1.1.1) and (M5.1.1.2) is

$$f(x) \pm g(x) = c_l x^l + c_{l-1} x^{l-1} + \cdots + c_1 x + c_0, \quad \text{where} \quad c_k = a_k \pm b_k \quad (k = 0, 1, \ldots, l).$$

If $n > m$ then $b_{m+1} = \cdots = b_n = 0$; if $n < m$ then $a_{n+1} = \cdots = a_m = 0$.

2°. To *multiply* a polynomial $f(x)$ of degree n by a polynomial $g(x)$ of degree m, one should multiply each term in $f(x)$ by each term in $g(x)$, add the products, and collect similar terms. The degree of the resulting polynomial is $n+m$. The product of polynomials (M5.1.1.1) and (M5.1.1.2) is

$$f(x)g(x) = c_{n+m} x^{n+m} + c_{n+m-1} x^{n+m-1} + \cdots + c_1 x + c_0, \qquad c_k = \sum_{i,j=0}^{i+j=k} a_i b_j,$$

where $k = 0, 1, \ldots, n+m$.

3°. Each polynomial $f(x)$ of degree n can be *divided* by any other polynomial $p(x)$ of degree m ($p(x) \neq 0$) *with remainder*, i.e., uniquely represented in the form $f(x) = p(x)q(x) + r(x)$, where $q(x)$ is a polynomial of degree $n - m$ (for $m \le n$) or $q(x) = 0$ (for $m > n$), referred to as the *quotient*, and $r(x)$ is a polynomial of degree $l < m$ or $r(x) = 0$, referred to as the *remainder*.

If $r(x) = 0$, then $f(x)$ is said to be *divisible* by $p(x)$ (without remainder).

If $m > n$, then $q(x) = 0$ and $r(x) \equiv f(x)$.

► Methods for finding quotient and remainder.

1°. *Horner's scheme*. To divide a polynomial $f(x)$ of degree n (see (M5.1.1.1)) by the polynomial $p(x) = x - b$, one uses Horner's scheme: the coefficients of $f(x)$ are written out in a row, starting from a_n; b is written on the left; then one writes the number a_n under a_n, the number $a_n b + a_{n-1} = b_{n-1}$ under a_{n-1}, the number $b_{n-1}b + a_{n-2} = b_{n-2}$ under $a_{n-2}, \dots,$ the number $b_1 b + a_0 = b_0$ under a_0. The number b_0 is the remainder in the division of $f(x)$ by $p(x)$, and $a_n, b_{n-1}, \dots, b_1$ are the coefficients of the quotient.

Remark. To divide $f(x)$ by $p(x) = ax + b$ ($a \neq 0$) with remainder, one first uses Horner's scheme to divide by $p_1(x) = x - (-\frac{b}{a})$; now if $q_1(x)$ and r_1 are the quotient and remainder in the division of $f(x)$ by $p_1(x)$, then $q(x) = \frac{1}{a} q_1(x)$ and $r = r_1$ are the quotient and remainder in the division of $f(x)$ by $p(x)$.

Example 1. Let us divide $f(x) = x^3 - 2x^2 - 10x + 3$ by $p(x) = 2x + 5$.
We use Horner's scheme to divide $f(x)$ by $p_1(x) = x + 5/2$:

$$\begin{array}{c|cccc} & 1 & -2 & -10 & 3 \\ \hline -\frac{5}{2} & 1 & -\frac{9}{2} & \frac{5}{4} & -\frac{1}{8} \end{array}$$

Thus $f(x) = p(x)q(x) + r(x)$, where

$$q(x) = \frac{1}{2}\Big(x^2 - \frac{9}{2}x + \frac{5}{4}\Big) = \frac{1}{2}x^2 - \frac{9}{4}x + \frac{5}{8}, \quad r = -\frac{1}{8}.$$

POLYNOMIAL REMAINDER THEOREM. *The remainder in the division of a polynomial $f(x)$ by the polynomial $p(x) = x - b$ is the number equal to the value of the polynomial $f(x)$ at $x = b$.*

2°. *Long division.* To divide a polynomial $f(x)$ of degree n by a polynomial $p(x)$ of degree $m \le n$, one can use *long division*.

Example 2. Let us divide $f(x) = x^3 + 8x^2 + 14x - 5$ by $p(x) = x^2 + 3x - 1$.
We use long division:

$$\begin{array}{rr|l} - & x^3 + 8x^2 + 14x - 5 & x^2 + 3x - 1 \\ \cline{3-3} & x^3 + 3x^2 \quad\;\; - x & x + 5 \\ \cline{2-2} - & 5x^2 + 15x - 5 & \\ & 5x^2 + 15x - 5 & \\ \cline{2-2} & 0 & \end{array}$$

Thus $f(x) = p(x)q(x) + r(x)$, where $q(x) = x + 5$ and $r(x) = 0$; i.e., $f(x)$ is divisible by $p(x)$.

Example 3. Let us divide $f(x) = x^3 - 4x^2 + x + 1$ by $p(x) = x^2 + 1$.
We use long division:

$$\begin{array}{rr|l} - & x^3 - 4x^2 + x + 1 & x^2 + 1 \\ \cline{3-3} & x^3 \qquad\;\; + x \quad\; & x - 4 \\ \cline{2-2} - & -4x^2 \qquad + 1 & \\ & -4x^2 \qquad - 4 & \\ \cline{2-2} & 5 & \end{array}$$

Thus $f(x) = p(x)q(x) + r(x)$, where $q(x) = x - 4$ and $r(x) = 5$.

► Expansion of polynomials in powers of a linear binomial. For each polynomial $f(x)$ given by equation (M5.1.1.1) and any number c, one can write out the *expansion of $f(x)$ in powers of $x - c$*:

$$f(x) = b_n(x - c)^n + b_{n-1}(x - c)^{n-1} + \cdots + b_1(x - c) + b_0.$$

To find the coefficients $b_0, \dots, b_n$ of this expansion, one first divides $f(x)$ by $x - c$ with remainder. The remainder is b_0, and the quotient is some polynomial $g_0(x)$. Then one divides

$g_0(x)$ by $x-c$ with remainder. The remainder is b_1, and the quotient is some polynomial $g_1(x)$. Then one divides $g_1(x)$ by $x-c$, obtaining the coefficient b_2 as the remainder, etc. It is convenient to perform the computations by Horner's scheme (see above).

The coefficients in the expansion of a polynomial $f(x)$ in powers of the difference $x-c$ are related to the values of the polynomial and its derivatives at $x=c$ by the formulas

$$b_0 = f(c), \quad b_1 = \frac{f'_x(c)}{1!}, \quad b_2 = \frac{f''_{xx}(c)}{2!}, \quad \ldots, \quad b_n = \frac{f^{(n)}_x(c)}{n!},$$

where the *derivative of a polynomial* $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + x_0$ with real or complex coefficients $a_0, \ldots, a_n$ is the polynomial $f'_x(x) = na_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + a_1$, $f''_{xx}(x) = [f'_x(x)]'_x$, etc. (see Subsection M6.2.1).

M5.1.2. Linear and Quadratic Equations

► **Linear equations.** The linear equation

$$ax + b = 0 \quad (a \neq 0)$$

has the solution

$$x = -\frac{b}{a}.$$

► **Quadratic equations.** The quadratic equation

$$ax^2 + bx + c = 0 \quad (a \neq 0) \tag{M5.1.2.1}$$

has the roots

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

The existence of real or complex roots is determined by the sign of the discriminant $D = b^2 - 4ac$:

Case $D > 0$. There are two distinct real roots.

Case $D < 0$. There are two distinct complex conjugate roots.

Case $D = 0$. There are two equal real roots.

VIÈTE THEOREM. *The roots of a quadratic equation (M5.1.2.1) satisfy the following relations:*

$$x_1 + x_2 = -\frac{b}{a}, \qquad x_1 x_2 = \frac{c}{a}.$$

M5.1.3. Cubic Equations

► **Incomplete cubic equation.**

1°. *Cardano's solution.* The roots of the incomplete cubic equation

$$y^3 + py + q = 0 \tag{M5.1.3.1}$$

have the form

$$y_1 = A + B, \quad y_{2,3} = -\frac{1}{2}(A+B) \pm i\frac{\sqrt{3}}{2}(A-B),$$

where

$$A=\left(-\frac{q}{2}+\sqrt{D}\right)^{1/3},\quad B=\left(-\frac{q}{2}-\sqrt{D}\right)^{1/3},\quad D=\left(\frac{p}{3}\right)^3+\left(\frac{q}{2}\right)^2,\quad i^2=-1,$$

and A, B are arbitrary values of the cubic roots such that $AB=-\frac{1}{3}p$.

The number of real roots of a cubic equation depends on the sign of the discriminant D:

Case $D>0$. There is one real and two complex conjugate roots.

Case $D<0$. There are three real roots.

Case $D=0$. There is one real root and another real root of double multiplicity.

2°. *Trigonometric solution.* If an incomplete cubic equation (M5.1.3.1) has real coefficients p and q, then its solutions can be found with the help of the trigonometric formulas given below.

(*a*) Let $p<0$ and $D<0$. Then

$$y_1=2\sqrt{-\frac{p}{3}}\cos\frac{\alpha}{3},\quad y_{2,3}=-2\sqrt{-\frac{p}{3}}\cos\left(\frac{\alpha}{3}\pm\frac{\pi}{3}\right),$$

where the values of the trigonometric functions are calculated from the relation

$$\cos\alpha=-\frac{q}{2\sqrt{-(p/3)^3}}.$$

(*b*) Let $p>0$ and $D\ge 0$. Then

$$y_1=2\sqrt{\frac{p}{3}}\cot(2\alpha),\quad y_{2,3}=\sqrt{\frac{p}{3}}\left[\cot(2\alpha)\pm i\frac{\sqrt{3}}{\sin(2\alpha)}\right],$$

where the values of the trigonometric functions are calculated from the relations

$$\tan\alpha=\left(\tan\frac{\beta}{2}\right)^{1/3},\quad \tan\beta=\frac{2}{q}\left(\frac{p}{3}\right)^{3/2},\quad |\alpha|\le\frac{\pi}{4},\quad |\beta|\le\frac{\pi}{2}.$$

(*c*) Let $p<0$ and $D\ge 0$. Then

$$y_1=-2\sqrt{-\frac{p}{3}}\frac{1}{\sin(2\alpha)},\quad y_{2,3}=\sqrt{-\frac{p}{3}}\left[\frac{1}{\sin(2\alpha)}\pm i\sqrt{3}\cot(2\alpha)\right],$$

where the values of the trigonometric functions are calculated from the relations

$$\tan\alpha=\left(\tan\frac{\beta}{2}\right)^{1/3},\quad \sin\beta=\frac{2}{q}\left(-\frac{p}{3}\right)^{3/2},\quad |\alpha|\le\frac{\pi}{4},\quad |\beta|\le\frac{\pi}{2}.$$

In the above three cases, the real value of the cubic root should be taken.

► **Complete cubic equation**. The roots of a complete cubic equation

$$ax^3+bx^2+cx+d=0\quad (a\neq 0)\qquad\text{(M5.1.3.2)}$$

are calculated by the formulas

$$x_k=y_k-\frac{b}{3a},\quad k=1,\,2,\,3,$$

where y_k are the roots of the incomplete cubic equation (M5.1.3.1) with the coefficients

$$p=-\frac{1}{3}\left(\frac{b}{a}\right)^2+\frac{c}{a},\quad q=\frac{2}{27}\left(\frac{b}{a}\right)^3-\frac{bc}{3a^2}+\frac{d}{a}.$$

VIÈTE THEOREM. *The roots of a complete cubic equation (M5.1.3.2) satisfy the following relations:*

$$x_1+x_2+x_3=-\frac{b}{a},\quad x_1x_2+x_1x_3+x_2x_3=\frac{c}{a},\quad x_1x_2x_3=-\frac{d}{a}.$$

M5.1.4. Fourth-Degree Equation

► Special cases of fourth-degree equations.

1°. The *biquadratic equation*

$$ax^4 + bx^2 + c = 0$$

can be reduced to a quadratic equation (M5.1.2.1) by the substitution $\xi = x^2$. Therefore, the roots of the biquadratic equations are given by

$$x_{1,2} = \pm\sqrt{\frac{-b + \sqrt{b^2 - 4ac}}{2a}}, \quad x_{3,4} = \pm\sqrt{\frac{-b - \sqrt{b^2 - 4ac}}{2a}}.$$

2°. The *reciprocal (algebraic) equation*

$$ax^4 + bx^3 + cx^2 + bx + a = 0$$

can be reduced to a quadratic equation by the substitution

$$y = x + \frac{1}{x}.$$

The resulting quadratic equation has the form

$$ay^2 + by + c - 2a = 0.$$

3°. The *generalized reciprocal equation*

$$ax^4 + bx^3 + cx^2 + \lambda bx + \lambda^2 a = 0$$

can be reduced to a quadratic equation by the substitution

$$y = x + \frac{\lambda}{x}.$$

The resulting quadratic equation has the form

$$ay^2 + by + c - 2a\lambda = 0.$$

► General fourth-degree equation.

1°. *Reduction of a general fourth-degree equation to an incomplete equation.* The general fourth-degree equation

$$ax^4 + bx^3 + cx^2 + dx + e = 0 \quad (a \neq 0)$$

can be reduced to an incomplete equation of the form

$$y^4 + py^2 + qy + r = 0 \qquad \text{(M5.1.4.1)}$$

by the substitution

$$x = y - \frac{b}{4a}.$$

2°. *Descartes–Euler solution.* The roots of the incomplete equation (M5.1.4.1) are given by the formulas

$$\begin{aligned} y_1 &= \tfrac{1}{2}\left(\sqrt{z_1} + \sqrt{z_2} + \sqrt{z_3}\right), & y_2 &= \tfrac{1}{2}\left(\sqrt{z_1} - \sqrt{z_2} - \sqrt{z_3}\right), \\ y_3 &= \tfrac{1}{2}\left(-\sqrt{z_1} + \sqrt{z_2} - \sqrt{z_3}\right), & y_4 &= \tfrac{1}{2}\left(-\sqrt{z_1} - \sqrt{z_2} + \sqrt{z_3}\right), \end{aligned} \qquad \text{(M5.1.4.2)}$$

where z_1, z_2, z_3 are the roots of the cubic equation (cubic resolvent of equation (M5.1.4.1))

$$z^3 + 2pz^2 + (p^2 - 4r)z - q^2 = 0. \qquad \text{(M5.1.4.3)}$$

The signs of the roots in (M5.1.4.2) are chosen from the condition

$$\sqrt{z_1}\sqrt{z_2}\sqrt{z_3} = -q.$$

The roots of the fourth-degree equation (M5.1.4.1) are determined by the roots of the cubic resolvent (M5.1.4.3); see Table M5.1.

TABLE M5.1
Relations between the roots of an incomplete equation of fourth-degree and the roots of its cubic resolvent.

Cubic resolvent (M5.1.4.3)	Fourth-degree equation (M5.1.4.1)
All roots are real and positive*	Four real roots
All roots are real: one is positive and two are negative*	Two pairs of complex conjugate roots
One real root and two complex conjugate roots	Two real roots and two complex conjugate roots

* By the Viète theorem, the product of the roots z_1, z_2, z_3 is equal to $q^2 \geq 0$.

3°. *Ferrari solution.* Let z_0 be any of the roots of the auxiliary cubic equation (M5.1.4.3). Then the four roots of the incomplete equation (M5.1.4.1) are found by solving the following two quadratic equations:

$$y^2 - \sqrt{z_0}\,y + \frac{p+z_0}{2} + \frac{q}{2\sqrt{z_0}} = 0,$$
$$y^2 + \sqrt{z_0}\,y + \frac{p+z_0}{2} - \frac{q}{2\sqrt{z_0}} = 0.$$

M5.1.5. Algebraic Equations of Arbitrary Degree and Their Properties

► **Simplest equations of degree n and their solutions.**

1°. The *binomial algebraic equation*

$$x^n - a = 0 \quad (a \neq 0)$$

has the solutions

$$x_{k+1} = \begin{cases} a^{1/n}\left(\cos\dfrac{2k\pi}{n} + i\sin\dfrac{2k\pi}{n}\right) & \text{for } a > 0, \\ |a|^{1/n}\left(\cos\dfrac{(2k+1)\pi}{n} + i\sin\dfrac{(2k+1)\pi}{n}\right) & \text{for } a < 0, \end{cases}$$

where $k = 0, 1, \ldots, n-1$ and $i^2 = -1$.

2°. Equations of the form

$$x^{2n} + ax^n + b = 0,$$
$$x^{3n} + ax^{2n} + bx^n + c = 0,$$
$$x^{4n} + ax^{3n} + bx^{2n} + cx^n + d = 0$$

are reduced by the substitution $y = x^n$ to a quadratic, cubic, and fourth-degree equation, respectively, whose solution can be expressed by radicals (see Subsections M5.1.2–M5.1.4).

Remark. In the above equations, n can be noninteger.

3°. The *generalized reciprocal (algebraic) equation*

$$a_0x^{2n}+a_1x^{2n-1}+\cdots+a_{n-1}x^{n+1}+a_nx^n + \lambda a_{n-1}x^{n-1}+\lambda^2a_{n-2}x^{n-2}+\cdots+\lambda^{n-1}a_1x+\lambda^na_0=0 \qquad (a_0\neq 0).$$

can be reduced to an equation of degree n by the substitution

$$y=x+\frac{\lambda}{x}.$$

Example 1. The equation

$$ax^6+bx^5+cx^4+dx^3+cx^2+bx+a=0,$$

which is a special case of the reciprocal equation with $n=3$ and $\lambda=1$, can be reduced to the cubic equation

$$ay^3+by^2+(c-3a)y+d-2b=0$$

by the substitution $y=x+1/x$.

► **Equations of general form and their properties.** An *algebraic equation of degree* n has the form

$$a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0=0 \quad (a_n\neq 0), \tag{M5.1.5.1}$$

where a_k are real or complex coefficients. Denote the polynomial of degree n on the left-hand side in equation (M5.1.5.1) by

$$P_n(x)\equiv a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0 \quad (a_n\neq 0). \tag{M5.1.5.2}$$

A value $x=x_1$ such that $P_n(x_1)=0$ is called a *root* of equation (M5.1.5.1) (and also a root of the polynomial $P_n(x)$). A value $x=x_1$ is called a *root of multiplicity* m if $P_n(x)=(x-x_1)^mQ_{n-m}(x)$, where m is an integer $(1\leq m\leq n)$, and $Q_{n-m}(x)$ is a polynomial of degree $n-m$ such that $Q_{n-m}(x_1)\neq 0$.

THEOREM 1 (FUNDAMENTAL THEOREM OF ALGEBRA). *Any algebraic equation of degree n has exactly n roots (real or complex), each root counted according to its multiplicity.*

Thus, the left-hand side of equation (M5.1.5.1) with roots $x_1, x_2, \ldots, x_s$ of the respective multiplicities $k_1, k_2, \ldots, k_s$ $(k_1+k_2+\cdots+k_s=n)$ can be factorized as follows:

$$P_n(x)=a_n(x-x_1)^{k_1}(x-x_2)^{k_2}\ldots(x-x_s)^{k_s}.$$

THEOREM 2. *Any algebraic equation of an odd degree with real coefficients has at least one real root.*

THEOREM 3. *Suppose that equation (M5.1.5.1) with real coefficients has a complex root $x_1=\alpha+i\beta$. Then this equation has the complex conjugate root $x_2=\alpha-i\beta$, and the roots x_1, x_2 have the same multiplicity.*

THEOREM 4. *Any rational root of equation (M5.1.5.1) with integer coefficients a_k is an irreducible fraction of the form p/q, where p is a divisor of a_0 and q is a divisor of a_n. If $a_n=1$, then all rational roots of equation (M5.1.5.1) (if they exist) are integer divisors of the free term.*

THEOREM 5 (ABEL–RUFFINI THEOREM). *Any equation (M5.1.5.1) of degree $n\leq 4$ is solvable by radicals, i.e., its roots can be expressed via its coefficients by the operations of addition, subtraction, multiplication, division, and taking roots (see Subsections M5.1.2–M5.1.4). In general, equation (M5.1.5.1) of degree $n>4$ cannot be solved by radicals.*

► Relations between roots and coefficients. Discriminant of an equation.

VIÈTE THEOREM. *The roots of equation (M5.1.5.1) (counted according to their multiplicity) and its coefficients satisfy the following relations:*

$$(-1)^k \frac{a_{n-k}}{a_n} = S_k \quad (k = 1,\,2,\,\ldots,\,n),$$

where S_k are elementary symmetric functions of $x_1, x_2, \ldots, x_n$:

$$S_1 = \sum_{i=1}^{n} x_i, \quad S_2 = \sum_{1\le i<j}^{n} x_i x_j, \quad S_3 = \sum_{1\le i<j<k}^{n} x_i x_j x_k, \quad \ldots, \quad S_n = x_1 x_2 \ldots x_n.$$

Note also the following relations:

$$(n-k)a_{n-k} + \sum_{j=1}^{k} a_{n-(k-j)} s_j = 0 \quad (k = 1,\,2,\,\ldots,\,n)$$

with symmetric functions $s_j = \sum\limits_{i=1}^{n} x_i^j$.

The *discriminant D of an algebraic equation* is the product of a_n^{2n-2} and the squared Vandermonde determinant $\Delta(x_1, x_2, \ldots, x_n)$ of its roots:

$$D = a_n^{2n-2}[\Delta(x_1, x_2, \ldots, x_n)]^2 = a_n^{2n-2} \prod_{1\le j<i\le n} (x_i - x_j)^2.$$

The discriminant D is a symmetric function of the roots $x_1, x_2, \ldots, x_n$, and is equal to zero if and only if the polynomial $P_n(x)$ has at least one multiple root.

► Bounds for the roots of algebraic equations with real coefficients.

1°. All roots of equation (M5.1.5.1) in absolute value do not exceed

$$N = 1 + \frac{A}{|a_n|}, \qquad \text{(M5.1.5.3)}$$

where A is the largest of $|a_0|, |a_1|, \ldots, |a_{n-1}|$.

The last result admits the following generalization: all roots of equation (M5.1.5.1) in absolute value do not exceed

$$N_1 = \rho + \frac{A_1}{|a_n|}, \qquad \text{(M5.1.5.4)}$$

where $\rho > 0$ is arbitrary and A_1 is the largest of

$$|a_{n-1}|, \quad \frac{|a_{n-2}|}{\rho}, \quad \frac{|a_{n-3}|}{\rho^2}, \quad \ldots, \quad \frac{|a_0|}{\rho^{n-1}}.$$

For $\rho = 1$, formula (M5.1.5.4) turns into (M5.1.5.3).

Remark. Formulas (M5.1.5.3) and (M5.1.5.4) can also be used for equations with complex coefficients.

Example 2. Consider the following equation of degree 4:

$$P_4(x) = 9x^4 - 9x^2 - 36x + 1.$$

Formula (M5.1.5.3) for $n = 4$, $|a_n| = 9$, $A = 36$ yields a fairly rough estimate $N = 5$, i.e., the roots of the equation belong to the interval $[-5, 5]$. Formula (M5.1.5.4) for $\rho = 2$, $n = 4$, $|a_n| = 9$, $A_1 = 9$ yields a better estimate for the bounds of the roots of this polynomial, $N_1 = 3$.

2°. A constant K is called an upper bound for the real roots of equation (M5.1.5.1) or the polynomial $P_n(x)$ if equation (M5.1.5.1) has no real roots greater than or equal to K; in a similar way, one defines a lower and an upper bound for positive and negative roots of an equation or the corresponding polynomial.

Let

K_1 be an upper bound for the positive roots of the polynomial $P_n(x)$,

K_2 be an upper bound for the positive roots of the polynomial $P_n(-x)$,

$K_3 > 0$ be an upper bound for the positive roots of the polynomial $x^n P_n(1/x)$,

$K_4 > 0$ be an upper bound for the positive roots of the polynomial $x^n P_n(-1/x)$.

Then all nonzero real roots of the polynomial $P_n(x)$ (if they exist) belong to the intervals $(-K_2, -1/K_4)$ and $(1/K_3, K_1)$.

Next, we describe three methods for finding upper bounds for positive roots of a polynomial.

Maclaurin method. Suppose that the first m leading coefficients of the polynomial (M5.1.5.2) are nonnegative, i.e., $a_n > 0$, $a_{n-1} \geq 0$, …, $a_{n-m+1} \geq 0$, and the next coefficient is negative, $a_{n-m} < 0$. Then

$$K = 1 + \left(\frac{B}{a_n}\right)^{1/m} \tag{M5.1.5.5}$$

is an upper bound for the positive roots of this polynomial, where B is the largest of the absolute values of negative coefficients of $P_n(x)$.

Example 3. Consider the fourth-degree equation from Example 2. In this case, $m = 2$, $B = 36$ and formula (M5.1.5.5) yields $K = K_1 = 1 + (36/9)^{1/2} = 3$. Now, consider the polynomial $P_4(-x) = 9x^4 - 9x^2 + 36x + 1$. Its positive roots has the upper bound $K_2 = 1 + (9/9)^{1/2} = 2$. For the polynomial $x^4 P_4(1/x) = x^4 - 36x^3 - 9x^2 + 9$, we have $m = 1$, $K_3 = 1 + 36 = 37$. Finally, for the polynomial $x^4 P_4(-1/x) = x^4 + 36x^3 - 9x^2 + 9$, we have $m = 2$, $k_4 = 1 + 9^{1/2} = 4$. Thus if $P_4(x)$ has real roots, they must belong to the intervals $(-2, -1/4)$ and $(1/37, 3)$.

Newton method. Suppose that the polynomial $P_n(x)$ and all its derivatives $P_n'(x)$, …, $P_n^{(n)}(x)$ take positive values for $x = c$. Then c is an upper bound for the positive roots of $P_n(x)$.

Example 4. Consider the polynomial from Example 2 and calculate the derivatives

$$P_4(x) = 9x^4 - 9x^2 - 36x + 1,\quad P_4'(x) = 36x^3 - 18x - 36,\quad P_4''(x) = 108x^2 - 18,\quad P_4'''(x) = 216x,\quad P_4''''(x) = 216.$$

It is easy to check that for $x = 2$ this polynomial and all its derivatives take positive values, and therefore $c = 2$ is an upper bound for its positive roots.

► **Theorems on the number of real roots of polynomials.** The number of all negative roots of a polynomial $P_n(x)$ is equal to the number of all positive roots of the polynomial $P_n(-x)$.

1°. The exact number of positive roots of a polynomial whose coefficients form a sequence that does not change sign or changes sign only once can be found with the help of the Descartes theorem (rule of signs).

DESCARTES THEOREM. *The number of positive roots (counted considering their multiplicity) of a polynomial $P_n(x)$ with real coefficients is either equal to the number of sign alterations between consecutive nonzero coefficients or is less than it by a multiple of 2.*

Applying the Descartes theorem to $P_n(-x)$, we obtain a similar theorem for the negative roots of the polynomial $P_n(x)$.

Example 5. Consider the cubic polynomial

$$P_3(x) = x^3 - 3x + a^2 \quad (a \neq 0).$$

Its coefficients have the signs + – +, and therefore we have two alterations of sign. Therefore, the number of positive roots of $P_3(x)$ is equal either to 2 or to 0. Now, consider the polynomial $P_3(-x) = -x^3 + 3x + a^2$. The sequence of its coefficients changes sign only once. Therefore, the original equation has one negative root.

2°. *A stronger version of the Descartes theorem.* Suppose that all roots of a polynomial $P_n(x)$ are real*; then the number of positive roots of $P_n(x)$ is equal to the number of sign alterations in the sequence of its coefficients, and the number of its negative roots is equal to the number of sign alterations in the sequence of coefficients of the polynomial $P_n(-x)$.

Example 6. Consider the characteristic polynomial of the symmetric matrix

$$P_3(x) = \begin{vmatrix} -2-x & 1 & 1 \\ 1 & 1-x & 3 \\ 1 & 3 & 1-x \end{vmatrix} = -x^3 + 14x + 20,$$

which has only real roots. The sequence of its coefficients changes sign only once, and therefore it has a single positive root. The number of its negative roots is equal to two, since this polynomial has three nonzero real roots and only one of them can be positive.

3°. If two neighboring coefficients of a polynomial $P_n(x)$ are equal to zero, then the roots of the polynomial cannot be all real (in this case, the stronger version of the Descartes theorem cannot be used).

4°. The number of real roots of a polynomial $P_n(x)$ greater than a fixed c is either equal to the number of sign alterations in the sequence $P_n(c), \dots, P_n^{(n)}(c)$ or is by an even number less. If all roots of $P_n(x)$ are real, then the number of its roots greater than c coincides with the number of sign alterations in the sequence $P_n(c), \dots, P_n^{(n)}(c)$.

Example 7. Consider the polynomial

$$P_4(x) = x^4 - 3x^3 + 2x^2 - 2a^2 x + a^2.$$

For $x = 1$, we have $P_4(1) = -a^2$, $P_4'(1) = -1 - 2a^2$, $P_4''(1) = -2$, $P_4'''(1) = 6$, $P_4''''(1) = 24$. Thus, there is a single sign alteration, and therefore the polynomial has a single real root greater than unity.

M5.2. Determinants and Matrices

M5.2.1. Determinants

► Second-order, third-order, and nth-order determinants.

1°. The *second-order determinant* is a number Δ associated with 4 scalar quantities a_{11}, a_{12}, a_{21}, a_{22}, arranged in a 2×2 square table. It is denoted and calculated as

$$\Delta = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

The numbers a_{11}, a_{12}, a_{21}, and a_{22} are called elements of the determinant Δ.

2°. The *third-order determinant* is a number Δ associated with a 3×3 square table of 9 scalar quantities; it is denoted and calculated as

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

$$= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}.$$

---

* This is the case, for instance, if we are dealing with the characteristic polynomial of a symmetric matrix.

This expression is obtained by the *triangle rule* (*Sarrus scheme*), illustrated by the following diagrams, where entries occurring in the same product with a given sign are joined by segments:

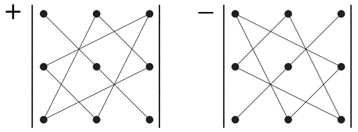

3°. The *nth-order determinant* is a number Δ associated with an $n\times n$ square table of n^2 scalar quantities; it is denoted by

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}. \qquad \text{(M5.2.1.1)}$$

The numbers a_{ij} are elements of the determinant Δ.

The determinant (M5.2.1.1) is calculated using the formulas

$$\begin{aligned} \Delta &= a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} \\ &= a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj} \end{aligned} \qquad \text{(M5.2.1.2)}$$

for any ith row and jth column. Here, A_{ij} is the *cofactor* of the element a_{ij}, which is defined as $A_{ij} = (-1)^{i+j}M_{ij}$, where M_{ij} is the *minor* corresponding to a_{ij}. The minor M_{ij} is defined as the $(n-1)$st-order determinant of size $(n-1)\times(n-1)$ obtained from the original determinant by removing the ith row and the jth column (i.e., the row and the column that intersect at a_{ij}). It follows from (M5.2.1.2) that the calculation of an nth-order determinant is reduced to the calculation of n determinants of order $n-1$.

The first formula in (M5.2.1.2) is called the *cofactor expansion of the determinant along row i* and the other one is called the *cofactor expansion of the determinant along column j*.

► Properties of determinants.

1. If a determinant contains a row (column) consisting of all zeroes, then this determinant is equal to zero.
2. If a determinant has two proportional rows (columns), then the determinant is zero.
3. If a determinant has a row (column) that is a linear combination of its other rows (columns), then the determinant is zero.
4. If two rows (columns) are interchanged, the determinant changes its sign.
5. If each element of a row (column) is divisible by a common number, this number can be factored out of the determinant.
6. The determinant does not change if a linear combination of some of its rows (columns) is added to another row (column).

Remark. The determinant is equal to zero if and only if its rows (columns) are linearly dependent.

► Calculation of determinants.

1°. Determinants can be calculated using the above properties.

Example 1. Find the determinant

$$\Delta = \begin{vmatrix} -13 & 25 & 17 \\ 26 & -34 & -26 \\ 36 & -33 & -24 \end{vmatrix}.$$

We first factor out the common divisor 2 of the elements in the second row, then add the resulting second row to the first one and then add the second row multiplied by –2 to the third one to obtain

$$\Delta = 2\begin{vmatrix} -13 & 25 & 17 \\ 13 & -17 & -13 \\ 36 & -33 & -24 \end{vmatrix} = 2\begin{vmatrix} 0 & 8 & 4 \\ 13 & -17 & -13 \\ 10 & 1 & 2 \end{vmatrix}.$$

In the last determinant, by adding the third column multiplied by –2 to the second one and by using the cofactor expansion along the first row, one obtains

$$\Delta = 2\begin{vmatrix} 0 & 0 & 4 \\ 13 & 9 & -13 \\ 10 & -3 & 2 \end{vmatrix} = 2\times 4\begin{vmatrix} 13 & 9 \\ 10 & -3 \end{vmatrix} = 8\,(-39-90) = -1032.$$

2°. Determinants are often calculated using the cofactor expansion formulas (M5.2.1.2). To this end, its is convenient to take a row or a column that contains many zero elements.

Example 2. Find the third-order determinant

$$A = \begin{vmatrix} 1 & -1 & 2 \\ 6 & 1 & 5 \\ 2 & -1 & -4 \end{vmatrix}.$$

We perform the cofactor expansion along the second column:

$$\det A = \sum_{k=1}^{3}(-1)^{k+2}a_{k2}M_{k2} = (-1)^{1+2}\times(-1)\times\begin{vmatrix} 6 & 5 \\ 2 & -4 \end{vmatrix} + (-1)^{2+2}\times 1\times\begin{vmatrix} 1 & 2 \\ 2 & -4 \end{vmatrix} + (-1)^{3+2}\times(-1)\times\begin{vmatrix} 1 & 2 \\ 6 & 5 \end{vmatrix}$$

$$= 1\times[6\times(-4)-5\times 2] + 1\times[1\times(-4)-2\times 2] + 1\times[1\times 5-2\times 6] = -49.$$

M5.2.2. Matrices. Types of Matrices. Operations with Matrices

► **Definition of a matrix. Types of matrices.** A *matrix* of *size* (or *dimension*) $m\times n$ is a rectangular table with *entries* a_{ij} ($i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$) arranged in m rows and n columns:

$$A \equiv \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}. \qquad \text{(M5.2.2.1)}$$

Note that, for each entry a_{ij}, the index i refers to the ith row and the index j to the jth column. Matrices are briefly denoted by uppercase letters (for instance, A, as here), or by the symbol $[a_{ij}]$, sometimes with more details: $A \equiv [a_{ij}]$ ($i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$). The numbers m and n are called the *dimensions* of the matrix.

The *null* or *zero matrix* is a matrix whose entries are all equal to zero: $a_{ij} = 0$ ($i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$).

A *column vector* or *column* is a matrix of size $m\times 1$. A *row vector* or *row* is a matrix of size $1\times n$. Both column and row vectors are often simply called *vectors*.

A *square matrix* is a matrix of size $n\times n$, and n is called the dimension of this square matrix. The *main diagonal* of a square matrix is its diagonal from the top left corner to the bottom right corner with the entries $a_{11}\ a_{22}\ \dots\ a_{nn}$. Table M5.2 lists the main types of square matrices.

► **Basic operations with matrices.** Two matrices are *equal* if they are of the same size and their respective entries are equal.

The *sum* of two matrices $A \equiv [a_{ij}]$ and $B \equiv [b_{ij}]$ of the same size $m\times n$ is the matrix $C \equiv [c_{ij}]$ of size $m\times n$ with the entries

$$c_{ij} = a_{ij} + b_{ij}.$$

TABLE M5.2
Some types of square matrices.

Type of square matrix $[a_{ij}]$	Entries
Unit (identity) $I = [\delta_{ij}]$	$a_{ij} = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j \end{cases}$ (δ_{ij} is the Kronecker delta)
Diagonal	$a_{ij} = \begin{cases} \text{any}, & i = j, \\ 0, & i \neq j \end{cases}$
Upper triangular (superdiagonal)	$a_{ij} = \begin{cases} \text{any}, & i \le j, \\ 0, & i > j \end{cases}$
Lower triangular (subdiagonal)	$a_{ij} = \begin{cases} \text{any}, & i \ge j, \\ 0, & i < j \end{cases}$
Symmetric	$a_{ij} = a_{ji}$
Skew-symmetric (antisymmetric)	$a_{ij} = -a_{ji}$
Hermitian (self-adjoint)	$a_{ij} = \bar{a}_{ji}$ ($\bar{a}_{ji}$ is the complex conjugate of a number a_{ji})

The sum of two matrices is denoted by $C = A + B$, and the operation is called *addition of matrices*.

Properties of addition of matrices:

$$\begin{aligned} &A + O = A && \text{(property of zero matrix)}, \\ &A + B = B + A && \text{(commutativity)}, \\ &(A + B) + C = A + (B + C) && \text{(associativity)}, \end{aligned}$$

where matrices A, B, C, and zero matrix O have the same size.

The *difference* of two matrices $A \equiv [a_{ij}]$ and $B \equiv [b_{ij}]$ of the same size $m \times n$ is the matrix $C \equiv [c_{ij}]$ of size $m \times n$ with entries

$$c_{ij} = a_{ij} - b_{ij} \quad (i = 1,\ 2,\ \ldots,\ m;\ j = 1,\ 2,\ \ldots,\ n).$$

The difference of two matrices is denoted by $C = A - B$, and the operation is called *subtraction of matrices*.

The *product* of a matrix $A \equiv [a_{ij}]$ of size $m \times n$ by a scalar λ is the matrix $C \equiv [c_{ij}]$ of size $m \times n$ with entries

$$c_{ij} = \lambda a_{ij} \quad (i = 1,\ 2,\ \ldots,\ m;\ j = 1,\ 2,\ \ldots,\ n).$$

The product of a matrix by a scalar is denoted by $C = \lambda A$, and the operation is called *multiplication of a matrix by a scalar*.

Properties of multiplication of a matrix by a scalar:

$$\begin{aligned} &0A = O && \text{(property of zero)}, \\ &(\lambda\mu)A = \lambda(\mu A) && \text{(associativity with respect to a scalar factor)}, \\ &\lambda(A + B) = \lambda A + \lambda B && \text{(distributivity with respect to addition of matrices)}, \\ &(\lambda + \mu)A = \lambda A + \mu A && \text{(distributivity with respect to addition of scalars)}, \end{aligned}$$

where λ and μ are scalars, matrices A, B, C, and zero matrix O have the same size.

The *product* of a matrix $A \equiv [a_{ij}]$ of size $m \times p$ and a matrix $B \equiv [b_{ij}]$ of size $p \times n$ is the matrix $C \equiv [c_{ij}]$ of size $m \times n$ with entries

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj} \quad (i = 1,\, 2,\, \ldots,\, m;\; j = 1,\, 2,\, \ldots,\, n);$$

i.e., the entry c_{ij} in the ith row and jth column of the matrix C is equal to the sum of products of the respective entries in the ith row of A and the jth column of B. Note that the product is defined for matrices of *compatible size*; i.e., the number of columns in the first matrix should be equal to the number of rows in the second matrix. The product of two matrices A and B is denoted by $C = AB$, and the operation is called *multiplication of matrices*.

Example 1. Consider two matrices

$$A = \begin{pmatrix} 1 & 2 \\ 6 & -3 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 10 & 1 \\ -6 & -0.5 & 20 \end{pmatrix}.$$

The product of the matrix A and the matrix B is the matrix

$$\begin{aligned} C = AB &= \begin{pmatrix} 1 & 2 \\ 6 & -3 \end{pmatrix} \begin{pmatrix} 0 & 10 & 1 \\ -6 & -0.5 & 20 \end{pmatrix} \\ &= \begin{pmatrix} 1\times 0 + 2\times(-6) & 1\times 10 + 2\times(-0.5) & 1\times 1 + 2\times 20 \\ 6\times 0 + (-3)\times(-6) & 6\times 10 + (-3)\times(-0.5) & 6\times 1 + (-3)\times 20 \end{pmatrix} = \begin{pmatrix} -12 & 9 & 41 \\ 18 & 61.5 & -54 \end{pmatrix}. \end{aligned}$$

Properties of multiplication of matrices:

$$\begin{aligned} &AO = O_1 && \text{(property of zero matrix)},\\ &(AB)C = A(BC) && \text{(associativity of the product of three matrices)},\\ &AI = A && \text{(multiplication by unit matrix)},\\ &A(B + C) = AB + AC && \text{(distributivity with respect to a sum of two matrices)},\\ &\lambda(AB) = (\lambda A)B = A(\lambda B) && \text{(associativity of the product of a scalar and two matrices)},\\ &SD = DS && \text{(commutativity for any square and any diagonal matrices)}, \end{aligned}$$

where λ is a scalar, matrices A, B, C, square matrix S, diagonal matrix D, zero matrices O and O_1, and unit matrix I have the compatible sizes.

Two square matrices A and B are said to *commute* if $AB = BA$, i.e., if their multiplication is subject to the commutative law (in general, this is not the case).

► **Transpose, orthogonal, and adjoint matrix.** The *transpose* of a matrix $A \equiv [a_{ij}]$ of size $m \times n$ is the matrix $C \equiv [c_{ij}]$ of size $n \times m$ with entries

$$c_{ij} = a_{ji} \quad (i = 1,\, 2,\, \ldots,\, n;\; j = 1,\, 2,\, \ldots,\, m).$$

The transpose is denoted by $C = A^T$.

Example 2. If $A = (a_1, a_2)$ then $A^T = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$.

Properties of transposes:

$$\begin{aligned} &(A + B)^T = A^T + B^T, \quad && (\lambda A)^T = \lambda A^T, \quad && (A^T)^T = A,\\ &(AC)^T = C^T A^T, && O^T = O_1, && I^T = I, \end{aligned}$$

where λ is a scalar; matrices A, B, and zero matrix O have size $m \times n$; matrix C has size $n \times l$; zero matrix O_1 has size $n \times m$.

THEOREM (DECOMPOSITION OF MATRICES). *For any square matrix A, the matrix $S_1 = \frac{1}{2}(A + A^T)$ is symmetric and the matrix $S_2 = \frac{1}{2}(A - A^T)$ is skew-symmetric. The representation of A as the sum of symmetric and skew-symmetric matrices is unique: $A = S_1 + S_2$.*

A square matrix A is said to be *orthogonal* if $A^T A = AA^T = I$, i.e., $A^T = A^{-1}$, where A^{-1} is the inverse of A (see Subsection M5.2.3).

Properties of orthogonal matrices:

1. If A is an orthogonal matrix, then A^T is also orthogonal.
2. The product of two orthogonal matrices is an orthogonal matrix.
3. Any symmetric orthogonal matrix is involutive, i.e., $AA = I$.

The *complex conjugate* of a matrix $A \equiv [a_{ij}]$ of size $m \times n$ is the matrix $C \equiv [c_{ij}]$ of size $m \times n$ with entries

$$c_{ij} = \bar{a}_{ij} \quad (i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n),$$

where $\bar{a}_{ij}$ is the complex conjugate of a_{ij}. The complex conjugate matrix is denoted by $C = \overline{A}$.

The *adjoint matrix* of a matrix $A \equiv [a_{ij}]$ of size $m \times n$ is the matrix $C \equiv [c_{ij}]$ of size $n \times m$ with entries

$$c_{ij} = \bar{a}_{ji} \quad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m).$$

The adjoint matrix is denoted by $C = A^*$.

Properties of adjoint matrices:

$$\begin{array}{lll} (A + B)^* = A^* + B^*, & (\lambda A)^* = \bar{\lambda} A^*, & (A^*)^* = A, \\ (AC)^* = C^* A^*, & O^* = O_1, & I^* = I, \end{array}$$

where λ is a scalar; matrices A, B, and zero matrix O have size $m \times n$; matrix C has size $n \times l$; zero matrix O_1 has size $n \times m$.

Remark. If a matrix is *real* (i.e., all its entries are real), then the corresponding transpose and the adjoint matrix coincide.

A square matrix A is said to be *normal* if $A^* A = AA^*$. A normal matrix A is said to be *unitary* if $A^* A = AA^* = I$, i.e., $A^* = A^{-1}$, where A^{-1} is the inverse of A (see Subsection M5.2.3).

► **Trace of a matrix.** The *trace* of a square matrix $A \equiv [a_{ij}]$ of size $n \times n$ is the sum of its diagonal entries,

$$\operatorname{Tr}(A) = \sum_{i=1}^{n} a_{ii}.$$

If λ is a scalar and square matrices A and B have the same size, then

$$\operatorname{Tr}(A + B) = \operatorname{Tr}(A) + \operatorname{Tr}(B), \quad \operatorname{Tr}(\lambda A) = \lambda \operatorname{Tr}(A), \quad \operatorname{Tr}(AB) = \operatorname{Tr}(BA),$$

► **Minors. Rank and defect of a matrix.** In a square or rectangular matrix, let us select k arbitrary rows and k arbitrary columns to make up a square submatrix. The kth-order determinant formed by the entries where the selected rows and columns intersect is called a kth-order *minor* of the matrix.

The *rank of a matrix* A is the maximum order of nonzero minors of A. The rank of a matrix A is denoted $\operatorname{rank}(A)$. If all entries of a matrix are zero, the rank of the matrix is taken to be zero.

Properties of the rank of a matrix:

1. The rank of a matrix does not change if: a row (column) whose entries are all zero is deleted; some rows (columns) are interchanged; a row (column) is multiplied by a nonzero number; the entries of one row (column) multiplied by any number are added to the respective entries of another row (column); and the rows are substituted by columns while the columns are substituted by the respective rows (for square matrices).

Example 3. Find the rank of the matrix $A = \begin{pmatrix} 1 & 2 & 3 & 6 \\ 2 & 3 & 1 & 6 \\ 3 & 1 & 2 & 6 \end{pmatrix}$.

Subtract the sum of the first three columns from column 4 and then delete the resulting column, whose entries are now all zero, to obtain the matrix $A_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix}$, which has the same rank as A. Since $\det(A_1) = -18 \neq 0$, we have $\operatorname{rank}(A_1) = 3$, and hence $\operatorname{rank}(A) = 3$.

2. For any matrices A and B of the same size the following inequality holds:

$$\operatorname{rank}(A + B) \le \operatorname{rank}(A) + \operatorname{rank}(B).$$

3. For a matrix A of size $m \times n$ and a matrix B of size $n \times k$, the *Sylvester inequalities* hold:

$$\operatorname{rank}(A) + \operatorname{rank}(B) - n \le \operatorname{rank}(AB) \le \min\{\operatorname{rank}(A), \operatorname{rank}(B)\}.$$

For a square matrix A of size $n \times n$, the value $d = n - \operatorname{rank}(A)$ is called the *defect* of the matrix A, and A is called a *d-fold degenerate matrix*. The rank of a nondegenerate square matrix $A \equiv [a_{ij}]$ of size $n \times n$ is equal to n.

4. Let r be the rank of a matrix; basic minor of this matrix is its nonzero minor of the order r. Basic rows (columns) of the matrix are the rows (columns) forming the basic minor.

THEOREM ON BASIC MINOR. *Basic rows (resp., basic columns) of a matrix are linearly independent. Any row (resp., any column) of a matrix is a linear combination of its basic rows (resp., columns).*

► **Linear dependence of row vectors (column vectors).** A row vector (column vector) B is a *linear combination* of row vectors (column vectors) $A_1, \dots, A_k$ if there exist scalars $\alpha_1, \dots, \alpha_k$ such that

$$B = \alpha_1 A_1 + \cdots + \alpha_k A_k.$$

Row vectors (column vectors) $A_1, \dots, A_k$ are said to be *linearly dependent* if there exist scalars $\alpha_1, \dots, \alpha_k$ ($\alpha_1^2 + \cdots + \alpha_k^2 \neq 0$) such that

$$\alpha_1 A_1 + \cdots + \alpha_k A_k = O,$$

where O is the zero row vector (column vector).

Row vectors (column vectors) $A_1, \dots, A_k$ are said to be *linearly independent* if, for any $\alpha_1, \dots, \alpha_k$ ($\alpha_1^2 + \cdots + \alpha_k^2 \neq 0$) we have

$$\alpha_1 A_1 + \cdots + \alpha_k A_k \neq O.$$

Remark. Row vectors (column vectors) $A_1, \dots, A_k$ are linearly dependent if and only if at least one of them is a linear combination of the others.

► **Determinant of a matrix.** For any square matrix A of the form (M5.2.2.1), one can calculate its determinant (M5.2.1.1), denoted $\det A$, $\det(A)$, or $|A|$.

The determinant of a matrix has the following properties:

1. The determinant of a triangular (upper or lower) and a diagonal matrices is equal to the product of its entries on the main diagonal. In particular, the determinant of the unit matrix is equal to 1.
2. The determinant of the product of two matrices A and B of the same size is equal to the product of their determinants,

$$\det(AB) = \det A \det B.$$

3. The determinant is invariant under matrix transposition:

$$\det A = \det A^T.$$

M5.2.3. Inverse Matrix. Functions of Matrices

► **Inverse matrices.** Let A be an $n\times n$ square matrix and let I be the unit matrix of the same size.

A square matrix A is called *nonsingular* or *nondegenerate* if $\det A \neq 0$.

THEOREM. *A square matrix is nondegenerate if and only if its rows (columns) are linearly independent.*

A square matrix A is called *invertible* if one can find a matrix B such that $AB = BA = I$. The matrix B is called the *inverse* of A and denoted A^{-1}. An invertible matrix A has a unique inverse.

THEOREM. *A square matrix A is invertible if and only if its determinant is nonzero (i.e., A is nonsingular).*

If the matrix A is defined by the table (M5.2.2.1), then its inverse is calculated as

$$A^{-1} = \begin{pmatrix} \frac{A_{11}}{\det A} & \frac{A_{21}}{\det A} & \cdots & \frac{A_{n1}}{\det A} \\ \frac{A_{12}}{\det A} & \frac{A_{22}}{\det A} & \cdots & \frac{A_{n2}}{\det A} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{A_{1n}}{\det A} & \frac{A_{2n}}{\det A} & \cdots & \frac{A_{nn}}{\det A} \end{pmatrix}. \qquad \text{(M5.2.3.1)}$$

where A_{ij} is the cofactor of the element a_{ij} of the determinant of A; the definition of A_{ij} can be found after formula (M5.2.1.2).

Properties of the inverse of a matrix:

$$(AB)^{-1} = B^{-1}A^{-1}, \quad (\lambda A)^{-1} = \frac{1}{\lambda}A^{-1},$$
$$(A^{-1})^{-1} = A, \quad (A^{-1})^T = (A^T)^{-1}, \quad (A^{-1})^* = (A^*)^{-1},$$

where the square matrices A and B are assumed to be nonsingular and the scalar λ to be nonzero.

► **Powers of square matrices.** The product of several identical square matrices A can be written as a *positive integer power* of the matrix A: $AA = A^2$, $AAA = A^2A = A^3$, etc. For a positive integer k, one defines $A^k = A^{k-1}A$ as the kth power of A. For a nondegenerate matrix A, one defines $A^0 = AA^{-1} = I$, $A^{-k} = (A^{-1})^k$. Powers of a matrix have the following properties:

$$A^pA^q = A^{p+q}, \quad (A^p)^q = A^{pq},$$

where p and q are arbitrary positive integers and A is an arbitrary square matrix; or p and q are arbitrary integers and A is an arbitrary nondegenerate matrix.

There exist matrices A^k whose positive integer power is equal to the zero matrix, even if $A \neq O$. If $A^k = O$ for some integer $k > 1$, then A is called a *nilpotent matrix*.

A matrix A is said to be *involutive* if it coincides with its inverse: $A = A^{-1}$ or $A^2 = I$.

► **Polynomials and functions of matrices.** A *polynomial with matrix argument* is the expression obtained from a scalar polynomial $f(x)$ by replacing the scalar argument x with a square matrix X:

$$f(X) = a_0I + a_1X + a_2X^2 + \cdots,$$

where a_i $(i = 0, 1, 2, \ldots)$ are real or complex coefficients. The polynomial $f(X)$ is a square matrix of the same size as X.

The exponential function of a square matrix X can be represented as the following convergent series:

$$e^X = 1 + X + \frac{X^2}{2!} + \frac{X^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{X^k}{k!}.$$

The inverse matrix has the form

$$(e^X)^{-1} = e^{-X} = 1 - X + \frac{X^2}{2!} - \frac{X^3}{3!} + \cdots = \sum_{k=0}^{\infty} (-1)^k \frac{X^k}{k!}.$$

Remark. Note that $e^Xe^Y \neq e^Ye^X$, in general. The relation $e^Xe^Y = e^{X+Y}$ holds only for commuting matrices X and Y.

Some other functions of matrices can be expressed in terms of the exponential function:

$$\sin X = \frac{1}{2i}(e^{iX} - e^{-iX}), \quad \cos X = \frac{1}{2}(e^{iX} + e^{-iX}),$$
$$\sinh X = \frac{1}{2}(e^X - e^{-X}), \quad \cosh X = \frac{1}{2}(e^X + e^{-X}).$$

M5.2.4. Eigenvalues and Characteristic Equation of a Matrix. The Cayley–Hamilton Theorem

► **Eigenvalues and spectra of square matrices.** An *eigenvalue* of a square matrix A is any real or complex λ for which the matrix $F(\lambda) \equiv A - \lambda I$ is degenerate. The set of all eigenvalues of a matrix A is called its *spectrum*, and $F(\lambda)$ is called its *characteristic matrix*. The inverse of an eigenvalue, $\mu = 1/\lambda$, is called a *characteristic value*.

A square matrix is nondegenerate if and only if all its eigenvalues are different from zero.

A nonzero (column) vector X satisfying the condition

$$AX = \lambda X$$

is called an *eigenvector of the matrix A* corresponding to the eigenvalue λ. Eigenvectors corresponding to distinct eigenvalues of A are linearly independent.

► **Characteristic equation of a matrix.** The algebraic equation of degree n

$$f_A(\lambda) \equiv \det(A - \lambda I) \equiv \det\,[a_{ij} - \lambda\delta_{ij}] \equiv \begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{vmatrix} = 0$$

is called the *characteristic equation* of the matrix A of size $n \times n$, and $f_A(\lambda)$ is called its *characteristic polynomial*. The spectrum of the matrix A (i.e., the set of all its eigenvalues) coincides with the set of all roots of its characteristic equation.

Example 1. The characteristic equation of the matrix

$$A = \begin{pmatrix} 4 & -8 & 1 \\ 5 & -9 & 1 \\ 4 & -6 & -1 \end{pmatrix}$$

has the form

$$f_A(\lambda) \equiv \det \begin{pmatrix} 4-\lambda & -8 & 1 \\ 5 & -9-\lambda & 1 \\ 4 & -6 & -1-\lambda \end{pmatrix} = -\lambda^3 - 6\lambda^2 - 11\lambda - 6 = -(\lambda+1)(\lambda+2)(\lambda+3) = 0.$$

Therefore the spectrum of the matrix A consists of three eigenvalues: $\lambda_1 = -1$, $\lambda_2 = -2$, and $\lambda_3 = -3$.

Let λ_j be an eigenvalue of a square matrix A. Then:

1) $\alpha\lambda_j$ is an eigenvalue of the matrix αA for any scalar α;
2) λ_j^p is an eigenvalue of the matrix A^p ($p = 0, \pm 1, \ldots, \pm N$ for a nondegenerate A; otherwise, $p = 0, 1, \ldots, N$), where N is a natural number;
3) a polynomial $f(A)$ of the matrix A has the eigenvalue $f(\lambda)$.

The matrix power series $\sum\limits_{k=0}^{\infty} \alpha_k A^k$ is convergent if and only if the power series $\sum\limits_{k=0}^{\infty} \alpha_k \lambda_j^k$ is convergent for each eigenvalue λ_j of A.

Regarding bounds for eigenvalues, see Subsection M5.1.5.

Let the positive integer s_i be the *multiplicity* of the eigenvalue λ_i of the characteristic equation of the matrix A of size $n \times n$. Note that $\sum\limits_i s_i = n$.

The determinant $\det A$ is equal to the product of all eigenvalues of A, each eigenvalue counted according to its multiplicity, i.e.,

$$\det A = \prod_i \lambda_i^{s_i}.$$

The trace $\mathrm{Tr}(A)$ is equal to the sum of all eigenvalues of A, each eigenvalue counted according to its multiplicity, i.e.,

$$\mathrm{Tr}(A) = \sum_i s_i \lambda_i.$$

► **Cayley–Hamilton theorem. Sylvester theorem.**

CAYLEY–HAMILTON THEOREM. *Each square matrix A satisfies its own characteristic equation; i.e., $f_A(A) = 0$.*

Example 2. Let us illustrate the Cayley–Hamilton theorem by the matrix in Example 1:

$$\begin{aligned} f_A(A) &= -A^3 - 6A^2 - 11A - 6I \\ &= -\begin{pmatrix} 70 & -116 & 19 \\ 71 & -117 & 19 \\ 64 & -102 & 11 \end{pmatrix} - 6\begin{pmatrix} -20 & 34 & -5 \\ -21 & 35 & -5 \\ -18 & 28 & -1 \end{pmatrix} - 11\begin{pmatrix} 4 & -8 & 1 \\ 5 & -9 & 1 \\ 4 & -6 & -1 \end{pmatrix} - 6\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = O. \end{aligned}$$

A scalar polynomial $p(\lambda)$ is called an *annihilating polynomial* of a square matrix A if $p(A) = 0$. For example, the characteristic polynomial $f_A(\lambda)$ is an annihilating polynomial of A. The unique monic annihilating polynomial of least degree is called the *minimal polynomial* of A and is denoted by $\psi(\lambda)$. The minimal polynomial is a divisor of every annihilating polynomial.

By dividing an arbitrary polynomial $f(\lambda)$ of degree n by an annihilating polynomial $p(\lambda)$ of degree m ($p(\lambda) \neq 0$), one obtains the representation

$$f(\lambda) = p(\lambda)q(\lambda) + r(\lambda),$$

where $q(\lambda)$ is a polynomial of degree $n - m$ (if $m \le n$) or $q(\lambda) = 0$ (if $m > n$) and $r(\lambda)$ is a polynomial of degree $l < m$ or $r(\lambda) = 0$. Hence

$$f(A) = p(A)q(A) + r(A),$$

where $p(A) = 0$ and $f(A) = r(A)$. The polynomial $r(\lambda)$ in this representation is called the *interpolation polynomial* of A.

Example 3. Let

$$f(A) = A^4 + 4A^3 + 2A^2 - 12A - 10I,$$

where the matrix A is defined in Example 1. Dividing $f(\lambda)$ by the characteristic polynomial $f_A(\lambda) = -\lambda^3 - 6\lambda^2 - 11\lambda - 6$, we obtain the remainder $r(\lambda) = 3\lambda^2 + 4\lambda + 2$. Consequently,

$$f(A) = r(A) = 3A^2 + 4A + 2I.$$

The Cayley–Hamilton theorem can also be used to find the powers and the inverse of a matrix A (since if $f_A(A) = 0$, then $A^k f_A(A) = 0$ for any positive integer k).

Example 4. For the matrix in Examples 1–3, one has

$$f_A(A) = -A^3 - 6A^2 - 11A - 6I = 0.$$

Hence we obtain

$$A^3 = -6A^2 - 11A - 6I.$$

By multiplying this expression by A, we obtain

$$A^4 = -6A^3 - 11A^2 - 6A.$$

Now we use the representation of the cube of A via lower powers of A and eventually arrive at the formula

$$A^4 = 25A^2 + 60A + 36I.$$

For the inverse matrix, by analogy with the preceding, we obtain

$$A^{-1} f_A(A) = A^{-1}(-A^3 - 6A^2 - 11A - 6I) = -A^2 - 6A - 11I - 6A^{-1} = 0.$$

The definitive result is

$$A^{-1} = -\frac{1}{6}(A^2 + 6A + 11I).$$

THEOREM. *Every analytic function of a square $n \times n$ matrix A can be represented as a polynomial of the same matrix,*

$$f(A) = \frac{1}{\Delta(\lambda_1, \lambda_2, \ldots, \lambda_n)} \sum_{k=1}^{n} \Delta_{n-k} A^{n-k},$$

where $\Delta(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is the Vandermonde determinant

$$\Delta(\lambda_1, \lambda_2, \ldots, \lambda_n) \equiv \begin{vmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \lambda_1^2 & \lambda_2^2 & \cdots & \lambda_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \cdots & \lambda_n^{n-1} \end{vmatrix} = \prod_{1 \le j < i \le n} (\lambda_i - \lambda_j)$$

and Δ_i *is obtained from* Δ *by replacing the* $(i+1)$*st row by* $(f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n))$.

In some cases, an analytic function of a matrix A can be computed by a formula in the following theorem.

SYLVESTER'S THEOREM. *If all eigenvalues of a matrix* A *are distinct, then*

$$f(A) = \sum_{k=1}^{n} f(\lambda_k) Z_k, \qquad Z_k = \frac{\prod_{i\neq k}(A - \lambda_i I)}{\prod_{i\neq k}(\lambda_k - \lambda_i)}.$$

M5.3. Systems of Linear Algebraic Equations

M5.3.1. Consistency Condition for a Linear System

► **Notion of a system of linear algebraic equations.** A *system of* m *linear equations with* n *unknown quantities* has the form

$$\begin{aligned} &a_{11}x_1 + a_{12}x_2 + \cdots + a_{1k}x_k + \cdots + a_{1n}x_n = b_1,\\ &a_{21}x_1 + a_{22}x_2 + \cdots + a_{2k}x_k + \cdots + a_{2n}x_n = b_2,\\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\ &a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mk}x_k + \cdots + a_{mn}x_n = b_m, \end{aligned} \tag{M5.3.1.1}$$

where $a_{11}, a_{12}, \dots, a_{mn}$ are the *coefficients of the system*; $b_1, b_2, \dots, b_m$ are its *constant terms*; and $x_1, x_2, \dots, x_n$ are the unknown quantities.

System (M5.3.1.1) is said to be *homogeneous* if all its constant terms are equal to zero. Otherwise (i.e., if there is at least one nonzero free term) the system is called *nonhomogeneous*.

If the number of equations is equal to that of the unknown quantities ($m = n$), system (M5.3.1.1) is called a *square system*.

A *solution* of system (M5.3.1.1) is a set of n numbers $x_1, x_2, \dots, x_n$ satisfying the equations of the system. A system is said to be *consistent* if it admits at least one solution. If a system has no solutions, it is said to be *inconsistent*. A consistent system of the form (M5.3.1.1) is called a *determined system* if it has a unique solution. A consistent system with more than one solution is said to be *undetermined*.

It is convenient to use matrix notation for systems of the form (M5.3.1.1),

$$AX = B, \tag{M5.3.1.2}$$

where $A \equiv [a_{ij}]$ is a matrix of size $m\times n$ called the *basic matrix* of the system; $X \equiv [x_i]$ is a column vector of size n; $B \equiv [b_i]$ is a column vector of size m.

► **Consistency condition for a general linear system.** System (M5.3.1.1) or (M5.3.1.2) is associated with two matrices: the basic matrix A of size $m\times n$ and the *augmented matrix* A_1 of size $m\times(n+1)$ formed by the matrix A supplemented with the column of the constant terms, i.e.,

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad A_1 = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1\\ a_{21} & a_{22} & \cdots & a_{2n} & b_2\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}. \tag{M5.3.1.3}$$

KRONECKER–CAPELLI THEOREM. *A linear system (M5.3.1.1) or (M5.3.1.2) is consistent if and only if its basic matrix and its augmented matrix (M5.3.1.3) have the same rank, i.e.,* $\operatorname{rank}(A_1) = \operatorname{rank}(A)$.

► **Equivalent systems of equations. The elementary transformations.** Two systems are said to be *equivalent* if their sets of solutions coincide.

Systems of linear equations can be simplified using the following three types of elementary transformations:

1. Interchange of two equations (or the corresponding rows of the augmented matrix).
2. Multiplication of both sides of one equation (or the corresponding row of the augmented matrix) by a nonzero constant.
3. Adding to both sides of one equation both sides of another equation multiplied by a constant (adding to some row of the augmented matrix its other row multiplied by a constant).

Under the above elementary transformations, a system of linear equations reduces to an equivalent system of equations.

M5.3.2. Finding Solutions of a System of Linear Equations

► **System of two equations with two unknown quantities.** A system of two equations with two unknown quantities has the form

$$\begin{aligned} a_1x+b_1y&=c_1,\\ a_2x+b_2y&=c_2. \end{aligned} \qquad \text{(M5.3.2.1)}$$

Depending on the coefficients a_k, b_k, c_k, the following three cases are possible:

1°. If $\Delta=a_1b_2-a_2b_1\neq 0$, then system (M5.3.2.1) has a unique solution,

$$x=\frac{c_1b_2-c_2b_1}{a_1b_2-a_2b_1},\qquad y=\frac{a_1c_2-a_2c_1}{a_1b_2-a_2b_1}.$$

2°. If $\Delta=a_1b_2-a_2b_1=0$ and $a_1c_2-a_2c_1=0$ (the case of proportional coefficients), then system (M5.3.2.1) has infinitely many solutions described by the formulas

$$x=t,\qquad y=\frac{c_1-a_1t}{b_1}\qquad (b_1\neq 0),$$

where t is arbitrary.

3°. If $\Delta=a_1b_2-a_2b_1=0$ and $a_1c_2-a_2c_1\neq 0$, then system (M5.3.2.1) has no solutions.

► **General square system of linear equations with $m = n$.** A square system of linear equations has the form (M5.3.1.1) with $m=n$.

1°. *Cramer's rule.* If the determinant of the matrix of system (M5.3.1.1) with $m=n$ is different from zero, i.e., $\Delta=\det A\neq 0$, then the system admits a unique solution, which is expressed by formulas

$$x_1=\frac{\Delta_1}{\Delta},\quad x_2=\frac{\Delta_2}{\Delta},\quad \dots,\quad x_n=\frac{\Delta_n}{\Delta}, \qquad \text{(M5.3.2.2)}$$

where Δ_k ($k=1, 2, \dots, n$) is the determinant of the matrix obtained from A by replacing its kth column with the column of constant terms:

$$\Delta_k=\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1\,k-1} & b_1 & a_{1\,k+1} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2\,k-1} & b_2 & a_{2\,k+1} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{n\,k-1} & b_n & a_{n\,k+1} & \cdots & a_{nn} \end{vmatrix}.$$

Example 1. Using Cramer's rule, let us find the solution of the system of linear equations

$$\begin{aligned} 2x_1 + x_2 + 4x_3 &= 16,\\ 3x_1 + 2x_2 + x_3 &= 10,\\ x_1 + 3x_2 + 3x_3 &= 16. \end{aligned}$$

The determinant of its basic matrix is different from zero,

$$\Delta = \begin{vmatrix} 2 & 1 & 4\\ 3 & 2 & 1\\ 1 & 3 & 3 \end{vmatrix} = 26 \neq 0,$$

and we have

$$\Delta_1 = \begin{vmatrix} 16 & 1 & 4\\ 10 & 2 & 1\\ 16 & 3 & 3 \end{vmatrix} = 26, \quad \Delta_2 = \begin{vmatrix} 2 & 16 & 4\\ 3 & 10 & 1\\ 1 & 16 & 3 \end{vmatrix} = 52, \quad \Delta_3 = \begin{vmatrix} 2 & 1 & 16\\ 3 & 2 & 10\\ 1 & 3 & 16 \end{vmatrix} = 78.$$

Therefore, by Cramer's rule (M5.3.2.2), the only solution of the system has the form

$$x_1 = \frac{\Delta_1}{\Delta} = \frac{26}{26} = 1, \quad x_2 = \frac{\Delta_2}{\Delta} = \frac{52}{26} = 2, \quad x_3 = \frac{\Delta_3}{\Delta} = \frac{78}{26} = 3.$$

2°. System (M5.3.1.1) with $m = n$ can be treated in the matrix form (M5.3.1.2) where A is a square matrix. If $\det A \neq 0$, the system has a unique solution

$$X = A^{-1}B,$$

expressed in terms of the inverse A^{-1}, which can be found by formula (M5.2.3.1).

3°. *Reduction of a system to a triangular form* (*Gaussian method*). Suppose that $\det A \neq 0$. The Gaussian method is based on elementary transformations (see Subsection M5.3.1) used for the reduction of a given system to an equivalent system having the triangular form

$$\begin{aligned} x_1 + \alpha_{12}x_2 + \alpha_{13}x_3 + \cdots + \alpha_{1n}x_n &= \beta_1,\\ x_2 + \alpha_{23}x_3 + \cdots + \alpha_{2n}x_n &= \beta_2,\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots &\\ x_{n-1} + \alpha_{n-1,n}x_n &= \beta_{n-1},\\ x_n &= \beta_n. \end{aligned}$$

This system can be easily solved: inserting $x_n = \beta_n$ (from the last equation) into the preceding $(n-1)$st equation, one finds x_{n-1}. Then, inserting the values obtained for x_n, x_{n-1} into the $(n-2)$nd equation, one finds x_{n-2}. Proceeding in this way, one finally finds x_1. This back substitution process is described by the formulas

$$x_k = \beta_k - \sum_{s=k+1}^{n} \alpha_{ks}x_s \quad (k = n-1, n-2, \dots, 1).$$

Example 2. Solve the system

$$\begin{aligned} x_1 + x_2 - 2x_3 &= -2,\\ 2x_1 + 3x_2 + x_3 &= 9,\\ 3x_1 + 2x_2 + 2x_3 &= 7. \end{aligned}$$

Multiply the first equation by –2 and add to the second one. Multiply the first equation by –3 and add to the third one. As a result, the first equation together with the two obtained make up the equivalent system

$$\begin{aligned} x_1 + x_2 - 2x_3 &= -2,\\ x_2 + 5x_3 &= 13,\\ -x_2 + 8x_3 &= 13. \end{aligned}$$

Adding together the last two equations and dividing the result by 13, one arrives at the triangular system

$$\begin{aligned} x_1 + x_2 - 2x_3 &= -2,\\ x_2 + 5x_3 &= 13,\\ x_3 &= 2. \end{aligned}$$

Solving this system from bottom to top, one finds that

$$x_3 = 2, \quad x_2 = 13 - 5x_3 = 3, \quad x_1 = -2 - x_2 + 2x_3 = -1.$$

4°. *The Jordan–Gauss method*. Let us introduce some definitions. An unknown x_i is called *resolved* or *basic* if it enters only in one equation of the system with coefficient 1 and is not contained in the other equations.

If each equation of the system contains a resolved unknown, this system is called *resolved*. The unknowns of the system that are not basic are called *free*.

In order to find all solutions of a consistent system of linear equations, it suffices to find an equivalent resolved system. If all the unknowns happen to be basic, the resolved system gives the values of these unknowns. Otherwise, the basic unknowns are expressed in terms of the free ones.

Description of the method. Let us write down the system of linear equations (M5.3.1.1) as the table

x_1	$\cdots$	x_k	$\cdots$	x_n	
a_{11}	$\cdots$	a_{1k}	$\cdots$	a_{1n}	b_1
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
a_{r1}	$\cdots$	a_{rk}	$\cdots$	a_{rn}	b_r
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
a_{m1}	$\cdots$	a_{mk}	$\cdots$	a_{mn}	b_m

For a resolving entry $a_{rk} \neq 0$, the following procedure is called the *Jordan transformation*:

1) multiply the rth row of the table by $1/a_{rk}$;

2) add the resulting rth row multiplied by $-a_{1k}$ to the first row;

3) add the rth row multiplied by $-a_{2k}$ to the second row; and so on for all remaining rows.

After that, the unknown x_k becomes resolved, with all entries of the kth column equal to zero except that $a_{rk} = 1$.

By choosing other resolving entries in different rows and performing the respective Jordan transformations, one arrives at a resolved system equivalent to the original one.

If, at some point, the coefficients of the unknowns in a row become all zero and the free term of that row is nonzero, then the system of equations is inconsistent. If all entries of a row, including the free term, become zero, then this row is crossed out from the table.

Example 3. Solve the system of equations

$$\begin{aligned} 2x_1 - 3x_2 + 5x_3 &= 1,\\ x_1 + 2x_2 - 3x_3 &= -7,\\ 2x_1 \phantom{{}+2x_2} + 5x_3 &= 4. \end{aligned}$$

Rewrite this system as a table and reduce it to a resolved form in six steps:

$$\begin{array}{r|r|r|r} x_1 & x_2 & x_3 & \\ \hline 2 & -3 & 5 & 1\\ \boxed{1} & 2 & -3 & -7\\ 2 & 0 & 5 & 4 \end{array} \Longrightarrow \begin{array}{r|r|r|r} x_1 & x_2 & x_3 & \\ \hline 0 & -7 & 11 & 15\\ 1 & 2 & -3 & -7\\ 0 & -4 & 11 & 18 \end{array} \Longrightarrow \begin{array}{r|r|r|r} x_1 & x_2 & x_3 & \\ \hline 0 & -7 & 11 & 15\\ 1 & 2 & -3 & -7\\ 0 & 3 & 0 & 3 \end{array}$$

$$\Longrightarrow \begin{array}{r|r|r|r} x_1 & x_2 & x_3 & \\ \hline 0 & -7 & 11 & 15\\ 1 & 2 & -3 & -7\\ 0 & \boxed{1} & 0 & 1 \end{array} \Longrightarrow \begin{array}{r|r|r|r} x_1 & x_2 & x_3 & \\ \hline 0 & 0 & 11 & 22\\ 1 & 0 & -3 & -9\\ 0 & 1 & 0 & 1 \end{array} \Longrightarrow \begin{array}{r|r|r|r} x_1 & x_2 & x_3 & \\ \hline 0 & 0 & \boxed{1} & 2\\ 1 & 0 & -3 & -9\\ 0 & 1 & 0 & 1 \end{array} \Longrightarrow \begin{array}{r|r|r|r} x_1 & x_2 & x_3 & \\ \hline 0 & 0 & 1 & 2\\ 1 & 0 & 0 & -3\\ 0 & 1 & 0 & 1 \end{array}$$

In the tables above, the resolving entries are boxed. The following sequence of actions has been performed: (1) double the second row has been subtracted from the first and third ones, (2) the first row has been subtracted from the third one, (3) the third row has been divided by 3, (4) the third row multiplied by 7 (resp., by –2) has been added to the first (resp., second) one, (5) the first row has been divided by 11, and (6) the first row multiplied by 3 has been added to the second one. Thus, the original system acquires the resolved form

$$\begin{cases} 0\cdot x_1 + 0\cdot x_2 + 1\cdot x_3 = 2, \\ 1\cdot x_1 + 0\cdot x_2 + 0\cdot x_3 = -3, \\ 0\cdot x_1 + 1\cdot x_2 + 0\cdot x_3 = 1. \end{cases}$$

The resulting solution is $x_1 = -3$, $x_2 = 1$, $x_3 = 2$.

Example 4. Solve the system of equations

$$\begin{cases} 2x_1 + 7x_2 + 3x_3 + x_4 = 6, \\ 3x_1 + 5x_2 + 2x_3 + 2x_4 = 4, \\ 9x_1 + 4x_2 + x_3 + 7x_4 = 2. \end{cases}$$

With Jordan transformations, this system is reduced to the resolved form

$$\begin{cases} - 11x_2 - 5x_3 + x_4 = -10, \\ x_1 + 9x_2 + 4x_3 \phantom{{}+x_4} = 8. \end{cases}$$

Hence, the set of all solutions to the original system is given by

$$x_1 = 8 - 9x_2 - 4x_3, \quad x_4 = -10 + 11x_2 + 5x_3,$$

with x_2 and x_3 assuming any real values.

► **General system of m linear equations with n unknown quantities.** Suppose that system (M5.3.1.1) is consistent and its basic matrix A has rank r. First, in the matrix A, one finds a submatrix of size $r\times r$ with a nonzero rth-order determinant and drops the $m-r$ equations whose coefficients do not belong to this submatrix (the dropped equations follow from the remaining ones and can, therefore, be neglected). In the remaining equations, the $n-r$ unknown quantities (free unknown quantities) that are not involved in the said submatrix should be transferred to the right-hand sides. Thus, one obtains a system of r equations with r unknown quantities, which can be solved by any of the methods described above in the current subsection.

Remark. If the rank r of the basic matrix and the rank of the augmented matrix of system (M5.3.1.1) are equal to the number of the unknown quantities n, then the system has a unique solution.

► **Existence of nontrivial solutions of a homogeneous system.** Consider the homogeneous system (M5.3.1.1), with $b_1 = b_2 = \cdots = b_m = 0$. This system is always consistent, since it always has the so-called *trivial solution* $x_1 = x_2 = \cdots = x_n = 0$.

THEOREM. *A homogeneous system has a nontrivial solution if and only if the rank of the matrix A is less than the number of the unknown quantities n.*

It follows that a square homogeneous system has a nontrivial solution if and only if the determinant of its matrix of coefficients is equal to zero, $\det A = 0$.

M5.4. Quadratic Forms

M5.4.1. Quadratic Forms and Their Transformations

► **Quadratic form with n variables.** A real quadratic form is a homogeneous polynomial of degree 2 in n variables $x_1, x_2, \ldots, x_n$ of the form

$$A_n(x_1, x_2, \ldots, x_n) \equiv \sum_{i,j=1}^{n} a_{ij}x_ix_j, \qquad \text{(M5.4.1.1)}$$

with real coefficients a_{ij} satisfying the symmetry condition $a_{ij} = a_{ji}$.

The quadratic form (M5.4.1.1) can be conveniently written in short matrix notation

$$A_n(X) = X^T A X, \qquad \text{(M5.4.1.2)}$$

where $X \equiv [x_i]$ is a column vector consisting of n elements, X^T is its transpose, and $A \equiv [a_{ij}]$ is an $n \times n$ symmetric matrix, called the *matrix of the quadratic form*.

A real-valued quadratic form $A_n(X)$ is said to be:

a) *positive definite* (resp., *negative definite*) if $A_n(X) > 0$ (resp., $A_n(X) < 0$) for any $X \neq 0$;
b) *indefinite* if there exist vectors X and Y such that $A_n(X) > 0$ and $A_n(Y) < 0$;
c) *nonnegative* (resp., *nonpositive*) if $A_n(X) \geq 0$ (resp., $A_n(X) \leq 0$) for all $X \neq 0$.

The determinant $\det A$ of the matrix A is called the *discriminant of the quadratic form* $A_n(X)$. A quadratic form is called *degenerate* if its discriminant is zero.

► **Criteria of positive and negative definiteness of a quadratic form.**

1°. A real quadratic form $A_n(X)$ is positive definite, negative definite, indefinite, nonnegative, nonpositive if the eigenvalues λ_i of its matrix $A \equiv [a_{ij}]$ are all positive, are all negative, some are positive and some negative, are all nonnegative, are all nonpositive, respectively.

2°. *Sylvester criterion.* A real quadratic form $A_n(X)$ is positive definite if and only if it satisfies the conditions

$$\Delta_1 \equiv a_{11} > 0, \quad \Delta_2 \equiv \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \quad \dots, \quad \Delta_n \equiv \det A > 0.$$

If the signs of the minor determinants alternate,

$$\Delta_1 < 0, \quad \Delta_2 > 0, \quad \Delta_3 < 0, \dots,$$

then the quadratic form is negative definite.

► **Transformations of a real quadratic form.** Let us find out how the coefficient matrix changes under a linear transformation of the variables

$$x_i = \sum_{k=1}^{n} b_{ik} y_k \qquad (i = 1,\, 2,\, \dots, n), \qquad \text{(M5.4.1.3)}$$

where b_{ik} are real numbers. In matrix notation, transformation (M5.4.1.3) becomes

$$X = BY, \qquad \text{(M5.4.1.4)}$$

where $Y \equiv [y_i]$ is a column vector of size n and $B \equiv [b_{ij}]$ is a transformation matrix of size $n \times n$.

Substituting (M5.4.1.4) into (M5.4.1.2) gives

$$A_n(X) = Y^T B^T A B Y = Y^T \widetilde{A} Y = \widetilde{A}_n(Y),$$

where

$$\widetilde{A} = B^T A B. \qquad \text{(M5.4.1.5)}$$

It follows that the discriminant of a quadratic form changes according to the rule

$$\det \widetilde{A} = \det A \, (\det B)^2.$$

In what follows, only nondegenerate transformations of variables are considered, i.e., those with $\det B \neq 0$. The rank of the coefficient matrix remains unchanged under such transformations. The rank of the coefficient matrix is usually said to be the *rank of the quadratic form*.

M5.4.2. Canonical and Normal Representations of a Quadratic Form

► **Canonical representation of a quadratic form.** Any real quadratic form (M5.4.1.1) can be reduced to the form

$$A_n(X) = \sum_{i=1}^{r} \lambda_i y_i^2 \equiv A_r(Y) \qquad \text{(M5.4.2.1)}$$

using an appropriate nondegenerate linear transformation (M5.4.1.3).

This representation is called a *canonical representation* of the quadratic form, the real coefficients $\lambda_1, \ldots, \lambda_r$ are called the *canonical coefficients*.

Reduction of the quadratic form (M5.4.1.1) to the canonical form (M5.4.2.1) is not unique and can be performed using various linear transformations of the form (M5.4.1.3).

LAW OF INERTIA OF QUADRATIC FORMS. *The number of terms with positive coefficients and the number of terms with negative coefficients in any canonical representation of a real quadratic form does not depend on the method used to obtain such a representation.*

The *index of inertia* of a real quadratic form is the integer r equal to the number of nonzero coefficients in its canonical representation (this number coincides with the rank of the quadratic form). Its *positive index of inertia* is the integer p equal to the number of positive coefficients in the canonical representation of the form, and its *negative index of inertia* is the integer q equal to the number of its negative canonical coefficients. The integer $s = p - q$ is called the *signature* of the quadratic form.

A real quadratic form $A_n(X)$ is

a) positive definite (resp., negative definite) if $p = n$ (resp., $q = n$);
b) indefinite if $p \neq 0$ and $q \neq 0$;
c) nonnegative (resp., nonpositive) if $q = 0$, $p < n$ (resp., $p = 0$, $q < n$).

THEOREM. *For any real symmetric quadratic form $A_n(X)$ there exists a real orthogonal transformation (M5.4.1.3), whose matrix B possesses the property $B^T B = BB^T = I$, that reduces the quadratic form to the canonical form (M5.4.2.1). The canonical coefficients $\lambda_1, \ldots, \lambda_n$ are eigenvalues of the quadratic form matrix A.*

► **Lagrange's method of reduction of a quadratic form to a canonical form.** For the canonical form (M5.4.1.1), consider the following two cases.

Case 1. Suppose that $a_{mm} \neq 0$ for some m $(1 \leq m \leq n)$. By letting

$$A_n(X) = \frac{1}{a_{mm}} \left(\sum_{k=1}^{n} a_{mk} x_k \right)^2 + A_{n-1}(X), \qquad \text{(M5.4.2.2)}$$

one can easily verify that the quadratic form $A_{n-1}(X)$ does not contain the variable x_m (it contains $n-1$ variables or fewer). This method of isolating a perfect square in a quadratic form can always be applied if the matrix $[a_{ij}]$ $(i, j = 1, 2, \ldots, n)$ contains nonzero diagonal elements.

Case 2. Suppose that $a_{mm} = a_{ss} = 0$, but $a_{ms} \neq 0$. In this case, the quadratic form can be represented as

$$A_n(X) = \frac{1}{2a_{ms}} \left[\sum_{k=1}^{n} (a_{mk} + a_{sk}) x_k \right]^2 - \frac{1}{2a_{ms}} \left[\sum_{k=1}^{n} (a_{mk} - a_{sk}) x_k \right]^2 + A_{n-2}(X), \qquad \text{(M5.4.2.3)}$$

where $A_{n-2}(X)$ does not contain the variables x_m and x_s (it contains $n-2$ variables), and the linear forms in square brackets are linearly independent (and therefore can be taken as new independent variables or coordinates).

By combining the above two procedures, the quadratic form $A_n(X)$ can always be represented in terms of squared linear forms; these forms are linearly independent, since each contains a variable which is absent from the other linear forms. By taking the linear forms to be new independent variables, one obtains the canonical representation of the quadratic form (M5.4.2.1).

Note that the main formulas (M5.4.2.2) and (M5.4.2.3) can be rewritten as

$$A_n(X) = \frac{1}{4a_{mm}}\left(\frac{\partial A_n}{\partial x_m}\right)^2 + A_{n-1}(X), \qquad \text{(M5.4.2.2}a\text{)}$$

$$A_n(X) = \frac{1}{8a_{ms}}\left[\left(\frac{\partial A_n}{\partial x_m} + \frac{\partial A_n}{\partial x_s}\right)^2 - \left(\frac{\partial A_n}{\partial x_m} - \frac{\partial A_n}{\partial x_s}\right)^2\right] + A_{n-2}(X). \qquad \text{(M5.4.2.3}a\text{)}$$

Example. Reduce the quadratic form

$$A_3(X) = 4x_1^2 + x_2^2 + x_3^2 - 4x_1x_2 - 4x_1x_3 + 4x_2x_3$$

to a canonical form.

Using formula (M5.4.2.2a) with $m=1$, we get

$$A_3(X) = \tfrac{1}{16}(8x_1 - 4x_2 - 4x_3)^2 + 2x_2x_3 = (2x_1 - x_2 - x_3)^2 + A_2(X).$$

Further applying formula (M5.4.2.3a) with $m=2$ and $s=3$ to $A_2(X) = 2x_2x_3$, we obtain

$$A_2(X) = 2x_2x_3 = \tfrac{1}{8}(2x_2 + 2x_3)^2 - \tfrac{1}{8}(2x_3 - 2x_2)^2 = \tfrac{1}{2}(x_2 + x_3)^2 - \tfrac{1}{2}(x_2 - x_3)^2.$$

The two formulas just obtained yield a canonical representation of the original form:

$$A_3(X) = y_1^2 + \tfrac{1}{2}y_2^2 - \tfrac{1}{2}y_3^2,$$

where

$$y_1 = 2x_1 - x_2 - x_3, \qquad y_2 = x_2 + x_3, \qquad y_3 = x_2 - x_3.$$

► **Jacobi's formula.** Introduce the following notation:

$$A\begin{pmatrix} x_1 & x_2 & \dots & x_k \\ y_1 & y_2 & \dots & y_k \end{pmatrix} = \sum_{i,j=1}^{k} a_{ij}x_ix_j$$

Let

$$D_k = A\begin{pmatrix} 1 & 2 & \dots & k \\ 1 & 2 & \dots & k \end{pmatrix} \neq 0 \qquad (k = 1,\, 2,\, \dots,\, r),$$

where r is the rank of the quadratic form (M5.4.1.1). Then the form (M5.4.1.1) admits the canonical representation

$$A_n(X) = \frac{1}{a_{11}}\,y_1^2 + \sum_{k=2}^{r} \frac{1}{D_{k-1}D_k}\,y_k^2,$$

where

$$y_k = c_{kk}x_k + c_{k,k+1}x_{k+1} + \dots + c_{kn}x_n, \quad c_{kq} = A\begin{pmatrix} 1 & 2 & \dots & k-1 & k \\ 1 & 2 & \dots & k-1 & q \end{pmatrix};$$
$$k = 1,\, 2,\, \dots,\, r; \quad q = k,\, k+1,\, \dots,\, n.$$

► **Normal representation of a real quadratic form.** Any real quadratic form (M5.4.1.1) admits the *normal representation*

$$A_n(X)=\sum_{i=1}^{n}\varepsilon_i z_i^2,$$

where $z_1, \dots, z_n$ are the new variables and $\varepsilon_1, \dots, \varepsilon_n$ are coefficients taking the values $-1, 0, 1$.

A normal representation can be obtained by the following transformations:

1. One obtains the canonical representation (M5.4.2.1), for example, by Lagrange's method.
2. With the nondegenerate coordinate transformation

$$y_i=\begin{cases}\dfrac{1}{\sqrt{\lambda_i}}\,z_i & \text{for } \lambda_i>0,\\[2mm] \dfrac{1}{\sqrt{-\lambda_i}}\,z_i & \text{for } \lambda_i<0,\\[2mm] z_i & \text{for } \lambda_i=0,\end{cases}$$

the canonical representation can be converted to a normal representation.

► **Simultaneous reduction of two quadratic forms to sums of squares.**

THEOREM. *Let $A_n(X)$ and $B_n(X)$ be real symmetric quadratic forms in n variables and let $B_n(X)$ be positive definite. Then there exists a real transformation (M5.4.1.3) that reduces the two forms to*

$$A(X)=\sum_{k=1}^{n}\lambda_k y_k^2,\qquad B(X)=\sum_{k=1}^{n}y_k^2,$$

where y_k are new variables. The set of real $\lambda_1, \dots, \lambda_n$ coincides with the spectrum of eigenvalues of the matrix $B^{-1}A$; this set consists of the roots of the algebraic equation

$$\det(A-\lambda B)=0.$$

M5.5. Linear Spaces

M5.5.1. Concept of a Linear Space. Its Basis and Dimension

► **Definition of a linear space.** A *linear space* or a *vector space* over a field of *scalars* (usually, the field of real numbers or the field of complex numbers) is a set $\mathcal{V}$ of elements $\mathbf{x}, \mathbf{y}, \mathbf{z}, \dots$ (also called *vectors*) of any nature for which the following conditions hold:

I. There is a rule that establishes correspondence between any pair of elements $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ and a third element $\mathbf{z} \in \mathcal{V}$, called the *sum* of the elements $\mathbf{x}, \mathbf{y}$ and denoted by $\mathbf{z}=\mathbf{x}+\mathbf{y}$.

II. There is a rule that establishes correspondence between any pair $\mathbf{x}$, λ, where $\mathbf{x}$ is an element of $\mathcal{V}$ and λ is a scalar, and an element $\mathbf{u} \in \mathcal{V}$, called the *product of a scalar λ and a vector* $\mathbf{x}$ and denoted by $\mathbf{u}=\lambda\mathbf{x}$.

III. The following eight axioms are assumed for the above two operations:

   1. Commutativity of the sum: $\mathbf{x}+\mathbf{y}=\mathbf{y}+\mathbf{x}$.
   2. Associativity of the sum: $(\mathbf{x}+\mathbf{y})+\mathbf{z}=\mathbf{x}+(\mathbf{y}+\mathbf{z})$.
   3. There is a zero element 0 such that $\mathbf{x}+0=\mathbf{x}$ for any $\mathbf{x}$.

4. For any element $\mathbf{x}$ there is an opposite element $\mathbf{x}'$ such that $\mathbf{x} + \mathbf{x}' = 0$.
5. A special role of the unit scalar 1: $1 \cdot \mathbf{x} = \mathbf{x}$ for any element $\mathbf{x}$.
6. Associativity of the multiplication by scalars: $\lambda(\mu\mathbf{x}) = (\lambda\mu)\mathbf{x}$.
7. Distributivity with respect to the addition of scalars: $(\lambda + \mu)\mathbf{x} = \lambda\mathbf{x} + \mu\mathbf{x}$.
8. Distributivity with respect to a sum of vectors: $\lambda(\mathbf{x} + \mathbf{y}) = \lambda\mathbf{x} + \lambda\mathbf{y}$.

This is the definition of an abstract linear space. We obtain a *specific linear space* if the nature of the elements and the operations of addition and multiplication by scalars are concretized.

Example 1. Consider the set of all free vectors in three-dimensional space. If addition of these vectors and their multiplication by scalars are defined as in analytic geometry (see Subsection M4.5.1), this set becomes a linear space denoted by B_3.

Example 2. Consider the *n-dimensional coordinate space* $\mathbb{R}^n$, whose elements are ordered sets of n arbitrary real numbers $(x_1, \dots, x_n)$. The generic element of this space is denoted by $\mathbf{x}$, i.e., $\mathbf{x} = (x_1, \dots, x_n)$, and the reals $x_1, \dots, x_n$ are called the *coordinates* of the element $\mathbf{x}$. From the algebraic standpoint, the set $\mathbb{R}^n$ may be regarded as the set of all row vectors with n real components.

The operations of addition of elements of $\mathbb{R}^n$ and their multiplication by scalars are defined by the following rules:

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n),$$
$$\lambda(x_1, \dots, x_n) = (\lambda x_1, \dots, \lambda x_n).$$

Remark. If the field of scalars λ, μ, … in the above definition is the field of all real numbers, the corresponding linear spaces are called *real linear spaces*. If the field of scalars is that of all complex numbers, the corresponding space is called a *complex linear space*. In many situations, it is clear from the context which field of scalars is meant.

The above axioms imply the following properties of an arbitrary linear space:

1. The zero vector is unique, and for any element $\mathbf{x}$ the opposite element is unique.
2. The zero vector 0 is equal to the product of any element $\mathbf{x}$ by the scalar 0.
3. For any element $\mathbf{x}$, the opposite element is equal to the product of $\mathbf{x}$ by the scalar –1.
4. The *difference* of two elements $\mathbf{x}$ and $\mathbf{y}$, i.e., the element $\mathbf{z}$ such that $\mathbf{z} + \mathbf{y} = \mathbf{x}$, is unique.

► **Basis and dimension of a linear space. Isomorphism of linear spaces.** An element $\mathbf{y}$ is called a *linear combination* of elements $\mathbf{x}_1, \dots, \mathbf{x}_k$ of a linear space $\mathcal{V}$ if there exist scalars $\alpha_1, \dots, \alpha_k$ such that

$$\mathbf{y} = \alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k.$$

Elements $\mathbf{x}_1, \dots, \mathbf{x}_k$ of the space $\mathcal{V}$ are said to be *linearly dependent* if there exist scalars $\alpha_1, \dots, \alpha_k$ such that $|\alpha_1|^2 + \cdots + |\alpha_k|^2 \neq 0$ and

$$\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k = 0,$$

where 0 is the zero element of $\mathcal{V}$.

Elements $\mathbf{x}_1, \dots, \mathbf{x}_k$ of the space $\mathcal{V}$ are said to be *linearly independent* if for any scalars $\alpha_1, \dots, \alpha_k$ such that $|\alpha_1|^2 + \cdots + |\alpha_k|^2 \neq 0$, we have

$$\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k \neq 0.$$

Remark 1. Elements $\mathbf{x}_1, \dots, \mathbf{x}_k$ of a linear space $\mathcal{V}$ are linearly dependent if and only if at least one of them is a linear combination of the others.

Remark 2. If at least one of the elements $\mathbf{x}_1, \dots, \mathbf{x}_k$ is equal to zero, then these elements are linearly dependent. If some of the elements $\mathbf{x}_1, \dots, \mathbf{x}_k$ are linearly dependent, then all these elements are linearly dependent.

Example 3. The elements $\mathbf{i}_1 = (1, 0, \ldots, 0)$, $\mathbf{i}_2 = (0, 1, \ldots, 0)$, …, $\mathbf{i}_n = (0, 0, \ldots, 1)$ of the space $\mathbb{R}^n$ (see Example 2) are linearly independent. For any $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, the vectors $\mathbf{x}$, $\mathbf{i}_1, \ldots, \mathbf{i}_n$ are linearly dependent.

A *basis* of a linear space $\mathcal{V}$ is defined as any system of linearly independent vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ such that for any element $\mathbf{x}$ of the space $\mathcal{V}$ there exist scalars $x_1, \ldots, x_n$ such that

$$\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n.$$

This relation is called the *representation of an element* $\mathbf{x}$ *in terms of the basis* $\mathbf{e}_1, \ldots, \mathbf{e}_n$, and the scalars $x_1, \ldots, x_n$ are called the *coordinates* of the element $\mathbf{x}$ in that basis.

UNIQUENESS THEOREM. *The representation of any element* $\mathbf{x} \in \mathcal{V}$ *in terms of a given basis* $\mathbf{e}_1, \ldots, \mathbf{e}_n$ *is unique.*

Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be any basis in $\mathcal{V}$ and vectors $\mathbf{x}$ and $\mathbf{y}$ have the coordinates $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ in that basis. Then the coordinates of the vector $\mathbf{x} + \mathbf{y}$ in that basis are $x_1 + y_1$, …, $x_n + y_n$, and the coordinates of the vector $\lambda\mathbf{x}$ are $\lambda x_1, \ldots, \lambda x_n$ for any scalar λ.

Example 4. Any three noncoplanar vectors form a basis in the linear space B_3 of all free vectors. The n elements $\mathbf{i}_1 = (1, 0, \ldots, 0)$, $\mathbf{i}_2 = (0, 1, \ldots, 0)$, …, $\mathbf{i}_n = (0, 0, \ldots, 1)$ form a basis in the linear space $\mathbb{R}^n$.

A linear space $\mathcal{V}$ is said to be n-*dimensional* if it contains n linearly independent elements and any $n + 1$ elements are linearly dependent. The number n is called the *dimension* of that space, $n = \dim \mathcal{V}$.

A linear space $\mathcal{V}$ is said to be *infinite-dimensional* ($\dim \mathcal{V} = \infty$) if for any positive integer N it contains N linearly independent elements.

THEOREM 1. *If* $\mathcal{V}$ *is a linear space of dimension* n*, then any* n *linearly independent elements of that space form its basis.*

THEOREM 2. *If a linear space* $\mathcal{V}$ *has a basis consisting of* n *elements, then* $\dim \mathcal{V} = n$.

Example 5. The dimension of the space B_3 of all free vectors is equal to 3. The dimension of the space $\mathbb{R}^n$ is equal to n.

Two linear spaces $\mathcal{V}$ and $\mathcal{V}'$ over the same field of scalars are said to be *isomorphic* if there is a one-to-one correspondence between the elements of these spaces such that if elements $\mathbf{x}$ and $\mathbf{y}$ from $\mathcal{V}$ correspond to elements $\mathbf{x}'$ and $\mathbf{y}'$ from $\mathcal{V}'$, then the element $\mathbf{x} + \mathbf{y}$ corresponds to $\mathbf{x}' + \mathbf{y}'$ and the element $\lambda\mathbf{x}$ corresponds to $\lambda\mathbf{x}'$ for any scalar λ.

Remark. If linear spaces $\mathcal{V}$ and $\mathcal{V}'$ are isomorphic, then the zero element of one space corresponds to the zero element of the other.

THEOREM. *Any two* n*-dimensional real (or complex) spaces* $\mathcal{V}$ *and* $\mathcal{V}'$ *are isomorphic.*

► **Affine space.** An *affine space* is a nonempty set $\mathcal{A}$ that consists of elements of any nature, called *points*, for which the following conditions hold:

I. There is a given linear (vector) space $\mathcal{V}$, called the *associated linear space*.

II. There is a rule by which any ordered pair of points $A, B \in \mathcal{A}$ is associated with an element (vector) from $\mathcal{V}$; this vector is denoted by $\overrightarrow{AB}$ and is called the *vector issuing from the point* A *with endpoint at* B.

III. The following conditions (called *axioms of affine space*) hold:
   1. For any point $A \in \mathcal{A}$ and any vector $\mathbf{a} \in \mathcal{V}$, there is a unique point $B \in \mathcal{A}$ such that $\overrightarrow{AB} = \mathbf{a}$.
   2. $\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$ for any three points $A, B, C \in \mathcal{A}$.

By definition, the *dimension of an affine space* $\mathcal{A}$ is the dimension of the associated linear space $\mathcal{V}$, $\dim \mathcal{A} = \dim \mathcal{V}$.

Any linear space may be regarded as an affine space.

In particular, the space $\mathbb{R}^n$ can be naturally considered as an affine space. Thus if $A = (a_1, \ldots, a_n)$ and $B = (b_1, \ldots, b_n)$ are points of the affine space $\mathbb{R}^n$, then the corresponding vector $\overrightarrow{AB}$ from the linear space $\mathbb{R}^n$ is defined by $\overrightarrow{AB} = (b_1 - a_1, \ldots, b_n - a_n)$.

Let $\mathcal{A}$ be an n-dimensional affine space with the associated linear space $\mathcal{V}$. A *coordinate system* in the affine space $\mathcal{A}$ is a fixed point $O \in \mathcal{A}$, together with a fixed basis $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathcal{V}$. The point O is called the *origin* of this coordinate system.

Let M be a point of an affine space $\mathcal{A}$ with a coordinate system $O\mathbf{e}_1 \ldots \mathbf{e}_n$. One says that the point M has *affine coordinates* (or simply coordinates) $x_1, \ldots, x_n$ in this coordinate system, and one writes $M = (x_1, \ldots, x_n)$ if $x_1, \ldots x_n$ are the coordinates of the radius-vector $\overrightarrow{OM}$ in the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$, i.e., $\overrightarrow{OM} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$.

M5.5.2. Subspaces of Linear Spaces

► **Concept of a linear subspace and a linear span.** A subset $\mathcal{L}$ of a linear space $\mathcal{V}$ is called a *linear subspace* of $\mathcal{V}$ if the following conditions hold:

1. If $\mathbf{x}$ and $\mathbf{y}$ belong to $\mathcal{L}$, then the sum $\mathbf{x} + \mathbf{y}$ belongs to $\mathcal{L}$.
2. If $\mathbf{x}$ belongs to $\mathcal{L}$ and λ is an arbitrary scalar, then the element $\lambda\mathbf{x}$ belongs to $\mathcal{L}$.

The *null subspace* in a linear space $\mathcal{V}$ is its subset consisting of the single element zero. The space $\mathcal{V}$ itself can be regarded as its own subspace. These two subspaces are called *improper subspaces*. All other subspaces are called *proper subspaces*.

Example 1. A subset B_2 consisting of all free vectors parallel to a given plane is a subspace in the linear space B_3 of all free vectors.

The *linear span* $L(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_m$ in a linear space $\mathcal{V}$ is, by definition, the set of all linear combinations of these vectors, i.e., the set of all vectors of the form

$$\alpha_1\mathbf{x}_1 + \cdots + \alpha_m\mathbf{x}_m,$$

where $\alpha_1, \ldots, \alpha_m$ are arbitrary scalars. The linear span $L(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ is the least subspace of $\mathcal{V}$ containing the elements $\mathbf{x}_1, \ldots, \mathbf{x}_m$.

If a subspace $\mathcal{L}$ of an n-dimensional space $\mathcal{V}$ does not coincide with $\mathcal{V}$, then $\dim\mathcal{L} < n = \dim\mathcal{V}$.

Let elements $\mathbf{e}_1, \ldots, \mathbf{e}_k$ form a basis in a k-dimensional subspace of an n-dimensional linear space $\mathcal{V}$. Then this basis can be supplemented by elements $\mathbf{e}_{k+1}, \ldots, \mathbf{e}_n$ of the space $\mathcal{V}$, so that the system $\mathbf{e}_1, \ldots, \mathbf{e}_k, \mathbf{e}_{k+1}, \ldots, \mathbf{e}_n$ forms a basis in the space $\mathcal{V}$.

THEOREM ON THE DIMENSION OF A LINEAR SPAN. *The dimension of a linear span $L(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ of elements $\mathbf{x}_1, \ldots, \mathbf{x}_m$ is equal to the maximal number of linearly independent vectors in the system $\mathbf{x}_1, \ldots, \mathbf{x}_m$.*

► **Sum and intersection of subspaces.** The *intersection* of subspaces $\mathcal{L}_1$ and $\mathcal{L}_2$ of one and the same linear space $\mathcal{V}$ is, by definition, the set of all elements $\mathbf{x}$ of $\mathcal{V}$ that belong simultaneously to both spaces $\mathcal{L}_1$ and $\mathcal{L}_2$. Such elements form a subspace of $\mathcal{V}$.

The *sum* of subspaces $\mathcal{L}_1$ and $\mathcal{L}_2$ of one and the same linear space $\mathcal{V}$ is, by definition, the set of all elements of $\mathcal{V}$ that can be represented in the form $\mathbf{y} + \mathbf{z}$, where $\mathbf{y}$ is an element of $\mathcal{V}_1$ and $\mathbf{z}$ is an element of $\mathcal{L}_2$. The sum of subspaces is also a subspace of $\mathcal{V}$.

THEOREM. *The sum of dimensions of arbitrary subspaces $\mathcal{L}_1$ and $\mathcal{L}_2$ of a finite-dimensional space $\mathcal{V}$ is equal to the sum of the dimension of their intersection and the dimension of their sum.*

Example 2. Let B_3 be the linear space of all free vectors (in three-dimensional space). Denote by $\mathcal{L}_1$ the subspace of all free vectors parallel to the plane OXY, and by $\mathcal{L}_2$ the subspace of all free vectors parallel to the plane OXZ. Then the sum of the subspaces $\mathcal{L}_1$ and $\mathcal{L}_2$ coincides with B_3, and their intersection consists of all free vectors parallel to the axis OX.

The dimension of each space $\mathcal{L}_1$ and $\mathcal{L}_2$ is equal to two, the dimension of their sum is equal to three, and the dimension of their intersection is equal to unity.

M5.5.3. Coordinate Transformations Corresponding to Basis Transformations in a Linear Space

► **Basis transformation and its inverse.** Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ and $\widetilde{\mathbf{e}}_1, \dots, \widetilde{\mathbf{e}}_n$ be two arbitrary bases of an n-dimensional linear space $\mathcal{V}$. Suppose that the elements $\widetilde{\mathbf{e}}_1, \dots, \widetilde{\mathbf{e}}_n$ are expressed via $\mathbf{e}_1, \dots, \mathbf{e}_n$ by the formulas

$$\begin{aligned}
\widetilde{\mathbf{e}}_1 &= a_{11}\mathbf{e}_1 + a_{12}\mathbf{e}_2 + \cdots + a_{1n}\mathbf{e}_n,\\
\widetilde{\mathbf{e}}_2 &= a_{21}\mathbf{e}_1 + a_{22}\mathbf{e}_2 + \cdots + a_{2n}\mathbf{e}_n,\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
\widetilde{\mathbf{e}}_n &= a_{n1}\mathbf{e}_1 + a_{n2}\mathbf{e}_2 + \cdots + a_{nn}\mathbf{e}_n.
\end{aligned}$$

Thus, the transition from the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ to the basis $\widetilde{\mathbf{e}}_1, \dots, \widetilde{\mathbf{e}}_n$ is determined by the matrix

$$A \equiv \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

Note that $\det A \neq 0$, i.e., the matrix A is nondegenerate.

The transition from the basis $\widetilde{\mathbf{e}}_1, \dots, \widetilde{\mathbf{e}}_n$ to the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ is determined by the matrix $B \equiv [b_{ij}] = A^{-1}$. Thus, we can write

$$\widetilde{\mathbf{e}}_i = \sum_{j=1}^{n} a_{ij}\mathbf{e}_j, \qquad \mathbf{e}_k = \sum_{j=1}^{n} b_{kj}\widetilde{\mathbf{e}}_j \qquad (i,k = 1,\,2,\,\dots,\,n). \tag{M5.5.3.1}$$

► **Relations between coordinate transformations and basis transformations.** Suppose that in a linear n-dimensional space $\mathcal{V}$, the transition from its basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ to another basis $\widetilde{\mathbf{e}}_1, \dots, \widetilde{\mathbf{e}}_n$ is determined by the matrix A (see above). Let $\mathbf{x}$ be any element of the space $\mathcal{V}$ with the coordinates $(x_1, \dots, x_n)$ in the basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ and the coordinates $(\widetilde{x}_1, \dots, \widetilde{x}_n)$ in the basis $\widetilde{\mathbf{e}}_1, \dots, \widetilde{\mathbf{e}}_n$, i.e.,

$$\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n = \widetilde{x}_1\widetilde{\mathbf{e}}_1 + \cdots + \widetilde{x}_n\widetilde{\mathbf{e}}_n.$$

Then using formulas (M5.5.3.1), we obtain the following relations between these coordinates:

$$x_j = \sum_{i=1}^{n} \widetilde{x}_i a_{ij}, \qquad \widetilde{x}_k = \sum_{l=1}^{n} x_l b_{lk}, \qquad j,k = 1,\dots,n.$$

In terms of matrices and row vectors, these relations can be written as follows:

$$(x_1,\dots,x_n) = (\widetilde{x}_1,\dots,\widetilde{x}_n)A, \qquad (\widetilde{x}_1,\dots,\widetilde{x}_n) = (x_1,\dots,x_n)A^{-1}$$

or, in terms of column vectors,

$$(x_1,\dots,x_n)^T = A^T(\widetilde{x}_1,\dots,\widetilde{x}_n)^T, \qquad (\widetilde{x}_1,\dots,\widetilde{x}_n)^T = (A^{-1})^T(x_1,\dots,x_n)^T,$$

where the superscript T indicates the transpose of a matrix.

M5.5.4. Euclidean Space

► **Definition and properties of a Euclidean space.** A *real Euclidean space* (or simply, *Euclidean space*) is a real linear space $\mathcal{V}$ endowed with a *scalar product* (also known as *inner product* and *dot product*), which is a real-valued function of two arguments $\mathbf{x} \in \mathcal{V}$, $\mathbf{y} \in \mathcal{V}$ denoted by $\mathbf{x}\cdot\mathbf{y}$, and satisfying the following conditions (axioms of the scalar product):

1. Symmetry: $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$.
2. Distributivity: $(\mathbf{x}_1 + \mathbf{x}_2) \cdot \mathbf{y} = \mathbf{x}_1 \cdot \mathbf{y} + \mathbf{x}_2 \cdot \mathbf{y}$.
3. Homogeneity: $(\lambda\mathbf{x}) \cdot \mathbf{y} = \lambda(\mathbf{x} \cdot \mathbf{y})$ for any real λ.
4. Positive definiteness: $\mathbf{x} \cdot \mathbf{x} \geq 0$ for any $\mathbf{x}$, and $\mathbf{x} \cdot \mathbf{x} = 0$ if and only if $\mathbf{x} = 0$.

Example 1. Consider the linear space B_3 of all free vectors in three-dimensional space. The space B_3 becomes a Euclidean space if the scalar product is introduced as in analytic geometry (see Subsection M4.5.3):

$$\mathbf{x} \cdot \mathbf{y} = |\mathbf{x}|\,|\mathbf{y}| \cos\varphi,$$

where φ is the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$.

Example 2. Consider the n-dimensional coordinate space $\mathbb{R}^n$ whose elements are ordered systems of n arbitrary real numbers, $\mathbf{x} = (x_1, \dots, x_n)$. Endowing this space with the scalar product

$$\mathbf{x} \cdot \mathbf{y} = x_1y_1 + \cdots + x_ny_n,$$

we obtain a Euclidean space.

THEOREM. *For any two elements* $\mathbf{x}$ *and* $\mathbf{y}$ *of a Euclidean space, the Cauchy–Schwarz inequality holds:*

$$(\mathbf{x} \cdot \mathbf{y})^2 \leq (\mathbf{x} \cdot \mathbf{x})(\mathbf{y} \cdot \mathbf{y}).$$

Here equality holds if and only if one of the vectors is 0 *or one vector is a multiple of the other.*

A linear space $\mathcal{V}$ is called a *normed space* if it is endowed with a *norm*, which is a real-valued function of $\mathbf{x} \in \mathcal{V}$, denoted by $\|\mathbf{x}\|$ and satisfying the following conditions:

1. Homogeneity: $\|\lambda\mathbf{x}\| = |\lambda|\,\|\mathbf{x}\|$ for any real λ.
2. Positive definiteness: $\|\mathbf{x}\| \geq 0$ and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = 0$.
3. The *triangle inequality* (also called the *Minkowski inequality*) holds for all elements $\mathbf{x}$ and $\mathbf{y}$:

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|. \tag{M5.5.4.1}$$

The value $\|\mathbf{x}\|$ is called the *norm of an element* $\mathbf{x}$ or its *length*.

THEOREM. *Any Euclidean space becomes a normed space if the norm is introduced by*

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}}. \tag{M5.5.4.2}$$

COROLLARY. *In any Euclidean space with the norm (M5.5.4.2), the triangle inequality (M5.5.4.1) holds for all its elements* $\mathbf{x}$ *and* $\mathbf{y}$.

The *distance between elements* $\mathbf{x}$ and $\mathbf{y}$ of a Euclidean space is defined by

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|.$$

One says that φ is the *angle* between two elements $\mathbf{x}$ and $\mathbf{y}$ of a Euclidean space if

$$\cos\varphi = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}.$$

Two elements $\mathbf{x}$ and $\mathbf{y}$ of a Euclidean space are said to be *orthogonal* if their scalar product is equal to zero, $\mathbf{x} \cdot \mathbf{y} = 0$.

PYTHAGOREAN THEOREM. *Let* $\mathbf{x}_1, \dots \mathbf{x}_m$ *be mutually orthogonal elements of a Euclidean space, i.e.,* $\mathbf{x}_i \cdot \mathbf{x}_j = 0$ *for* $i \neq j$. *Then*

$$\|\mathbf{x}_1 + \cdots + \mathbf{x}_m\|^2 = \|\mathbf{x}_1\|^2 + \cdots + \|\mathbf{x}_m\|^2.$$

Example 3. In the Euclidean space B_3 of free vectors with the usual scalar product (see Example 1), the following relations hold:

$$\|\mathbf{a}\| = |\mathbf{a}|, \quad (\mathbf{a}\cdot\mathbf{b})^2 \le |\mathbf{a}|^2|\mathbf{b}|^2, \quad |\mathbf{a}+\mathbf{b}| \le |\mathbf{a}| + |\mathbf{b}|.$$

In the Euclidean space $\mathbb{R}^n$ of ordered systems of n numbers with the scalar product defined in Example 2, the following relations hold:

$$\begin{aligned}
&\|\mathbf{x}\| = \sqrt{x_1^2 + \cdots + x_n^2},\\
&(x_1y_1 + \cdots + x_ny_n)^2 \le (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2),\\
&\sqrt{(x_1+y_1)^2 + \cdots + (x_n+y_n)^2} \le \sqrt{x_1^2 + \cdots + x_n^2} + \sqrt{y_1^2 + \cdots + y_n^2}.
\end{aligned}$$

► **Orthonormal basis in a finite-dimensional Euclidean space.** For elements $\mathbf{x}_1, \ldots, \mathbf{x}_m$ of a Euclidean space, the mth-order determinant $\det[\mathbf{x}_i\cdot\mathbf{x}_j]$ is called their *Gram determinant*. These elements are linearly independent if and only if their Gram determinant is different from zero.

One says that n elements $\mathbf{i}_1, \ldots, \mathbf{i}_n$ of an n-dimensional Euclidean space $\mathcal{V}$ form its *orthonormal basis* if these elements have unit norm and are mutually orthogonal, i.e.,

$$\mathbf{i}_i\cdot\mathbf{i}_j = \begin{cases} 1 & \text{for } i = j,\\ 0 & \text{for } i \ne j.\end{cases}$$

THEOREM. *In any n-dimensional Euclidean space $\mathcal{V}$, there exists an orthonormal basis.*

Orthogonalization of linearly independent elements:

Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be n linearly independent vectors of an n-dimensional Euclidean space $\mathcal{V}$. From these vectors, one can construct an orthonormal basis of $\mathcal{V}$ using the following algorithm (called *Gram–Schmidt orthogonalization*):

$$\mathbf{i}_i = \frac{\mathbf{g}_i}{\sqrt{\mathbf{g}_i\cdot\mathbf{g}_i}}, \quad \text{where} \quad \mathbf{g}_i = \mathbf{e}_i - \sum_{j=1}^{i}(\mathbf{e}_i\cdot\mathbf{i}_j)\mathbf{i}_j \quad (i = 1, 2, \ldots, n). \qquad \text{(M5.5.4.3)}$$

Remark. In any n-dimensional ($n > 1$) Euclidean space $\mathcal{V}$, there exist infinitely many orthonormal bases.

Properties of an orthonormal basis of a Euclidean space:

1. Let $\mathbf{i}_1, \ldots, \mathbf{i}_n$ be an orthonormal basis of a Euclidean space $\mathcal{V}$. Then the scalar product of two elements $\mathbf{x} = x_1\mathbf{i}_1 + \cdots + x_n\mathbf{i}_n$ and $\mathbf{y} = y_1\mathbf{i}_1 + \cdots + y_n\mathbf{i}_n$ is equal to the sum of products of their respective coordinates:
$$\mathbf{x}\cdot\mathbf{y} = x_1y_1 + \cdots + x_ny_n.$$
2. The coordinates of any vector $\mathbf{x}$ in an orthonormal basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$ are equal to the scalar product of $\mathbf{x}$ and the corresponding vector of the basis (or the projection of the element $\mathbf{x}$ on the axis in the direction of the corresponding vector of the basis):
$$x_k = \mathbf{x}\cdot\mathbf{i}_k \quad (k = 1, 2, \ldots, n).$$

Remark. In an arbitrary basis $\mathbf{e}_1, \ldots, \mathbf{e_n}$ of a Euclidean space, the scalar product of two elements $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$ and $\mathbf{y} = y_1\mathbf{e}_1 + \cdots + y_n\mathbf{e}_n$ has the form

$$\mathbf{x}\cdot\mathbf{y} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_iy_j,$$

where $a_{ij} = \mathbf{e}_i\cdot\mathbf{e}_j$ $(i, j = 1, 2, \ldots, n)$.

Two Euclidean spaces $\mathcal{V}$ and $\widetilde{\mathcal{V}}$ are said to be *isomorphic* if one can establish a one-to-one correspondence between the elements of these spaces satisfying the following conditions: if elements $\mathbf{x}$ and $\mathbf{y}$ of $\mathcal{V}$ correspond to elements $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$ of $\widetilde{\mathcal{V}}$, then the element $\mathbf{x}+\mathbf{y}$ corresponds to $\widetilde{\mathbf{x}}+\widetilde{\mathbf{y}}$; the element $\lambda\mathbf{x}$ corresponds to $\lambda\widetilde{\mathbf{x}}$ for any λ; the scalar product $(\mathbf{x}\cdot\mathbf{y})_{\mathcal{V}}$ is equal to the scalar product $(\widetilde{\mathbf{x}}\cdot\widetilde{\mathbf{y}})_{\widetilde{\mathcal{V}}}$.

THEOREM. *Any two n-dimensional Euclidean spaces $\mathcal{V}$ and $\widetilde{\mathcal{V}}$ are isomorphic.*

Bibliography for Chapter M5

Anton, H., *Elementary Linear Algebra, 8th Edition*, Wiley, New York, 2000.

Beecher, J. A., Penna, J. A., and Bittinger, M. L., *College Algebra, 2nd Edition*, Addison Wesley, Boston, Massachusetts, 2004.

Bellman, R. E., *Introduction to Matrix Analysis* (McGraw-Hill Series in Matrix Theory), McGraw-Hill, New York, 1960.

Bernstein, D. S., *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*, Princeton University Press, Princeton, 2005.

Blitzer, R. F., *College Algebra, 3rd Edition*, Prentice Hall, Englewood Cliffs, New Jersey, 2003.

Cullen, C. G., *Matrices and Linear Transformations, 2nd Edition*, Dover Publications, New York, 1990.

Dugopolski, M., *College Algebra, 3rd Edition*, Addison Wesley, Boston, 2002.

Eves, H., *Elementary Matrix Theory*, Dover Publications, New York, 1980.

Franklin, J. N., *Matrix Theory*, Dover Publications, New York, 2000.

Gantmacher, F. R., *Matrix Theory, Vol. 1, 2nd Edition*, American Mathematical Society, Providence, Rhode Island, 1990.

Gantmacher, F. R., *Applications of the Theory of Matrices*, Dover Publications, New York, 2005.

Gelfand, I. M., *Lectures on Linear Algebra*, Dover Publications, New York, 1989.

Gelfand, I. M. and Shen, A., *Algebra*, Birkhauser, Boston, Massachusetts, 2003.

Gilbert, J. and Gilbert, L., *Linear Algebra and Matrix Theory, 2nd Edition*, Brooks Cole, Stamford, Connecticut, 2004.

Hazewinkel, M. (Editor), *Handbook of Algebra, Vol. 1*, North Holland, Amsterdam, 1996.

Hazewinkel, M. (Editor), *Handbook of Algebra, Vol. 2*, North Holland, Amsterdam, 2000.

Hazewinkel, M. (Editor), *Handbook of Algebra, Vol. 3*, North Holland, Amsterdam, 2003.

Horn, R. A. and Johnson, C. R., *Matrix Analysis*, Cambridge University Press, Cambridge, England, 1990.

Korn, G. A. and Korn, T. M., *Mathematical Handbook for Scientists and Engineers, 2nd Edition*, Dover Publications, New York, 2000.

Kostrikin, A. I. and Manin, Yu. I., *Linear Algebra and Geometry*, Gordon & Breach, New York, 1997.

Kostrikin, A. I. and Shafarevich, I. R., *Algebra I: Basic Notions of Algebra*, Springer-Verlag, Berlin, 1990.

Kurosh, A. G., *Lectures on General Algebra*, Chelsea Publishing, New York, 1965.

Kurosh, A. G., *Algebraic Equations of Arbitrary Degrees*, Firebird Publishing, New York, 1977.

Lancaster, P. and Tismenetsky, M., *The Theory of Matrices, Second Edition: With Applications* (Computer Science and Scientific Computing), Academic Press, Boston, Massachusetts, 1985.

Lay, D. C., *Linear Algebra and Its Applications, 3rd Edition*, Addison Wesley, Boston, Massachusetts, 2002.

Lial, M. L., Hornsby, J., and Schneider, D. I., *College Algebra, 9th Edition*, Addison Wesley, Boston, Massachusetts, 2004.

Lipschutz, S. and Lipson, M., *Schaum's Outline of Linear Algebra, 3rd Edition*, McGraw-Hill, New York, 2000.

MacDuffee, C. C., *The Theory of Matrices*, Dover Publications, New York, 2004.

Meyer, C. D., *Matrix Analysis and Applied Linear Algebra, Package Edition*, Society for Industrial & Applied Mathematics, University City Science Center, Philadelphia, 2001.

Mikhalev, A. V. and Pilz, G., *The Concise Handbook of Algebra*, Kluwer Academic, Dordrecht, Boston, 2002.

Perlis, S., *Theory of Matrices*, Dover Publications, New York, 1991.

Schneider, H. and Barker, G. Ph., *Matrices and Linear Algebra, 2nd Edition*, Dover Publications, New York, 1989.

Shilov, G. E., *Linear Algebra*, Dover Publications, New York, 1977.

Strang, G., *Introduction to Linear Algebra, 3rd Edition*, Wellesley Cambridge Pr., Wellesley, 2003.

Strang, G., *Linear Algebra and Its Applications, 4th Edition*, Brooks Cole, Stamford, Connecticut, 2005.

Sullivan, M., *College Algebra, 7th Edition*, Prentice Hall, Englewood Cliffs, New Jersey, 2004.

Tobey, J. and Slater, J., *Intermediate Algebra, 5th Edition*, Prentice Hall, Englewood Cliffs, New Jersey, 2005.

Turnbull, H. W. and Aitken, A. C., *An Introduction to the Theory of Canonical Matrices, Phoenix Edition*, Dover Publications, New York, 2004.

Zhang, F., *Matrix Theory*, Springer, New York, 1999.

Chapter M6

Limits and Derivatives

M6.1. Basic Concepts of Mathematical Analysis

M6.1.1. Number Sets. Functions of Real Variable

► **Real axis, intervals, and segments.** The *real axis* is a straight line with a point O chosen as the origin, a positive direction, and a scale unit.

There is a one-to-one correspondence between the set of all real numbers $\mathbb{R}$ and the set of all points of the real axis, with each real x being represented by a point on the real axis separated from O by the distance $|x|$ and lying to the right of O for $x > 0$, or to the left of O for $x < 0$.

One often has to deal with the following number sets (sets of real numbers or sets on the real axis).

1. Sets of the form (a, b), $(-\infty, b)$, $(a, +\infty)$, and $(-\infty, +\infty)$ consisting, respectively, of all $x \in \mathbb{R}$ such that $a < x < b$, $x < b$, $x > a$, and x is arbitrary are called *open intervals* (sometimes simply *intervals*).

2. Sets of the form $[a, b]$ consisting of all $x \in \mathbb{R}$ such that $a \le x \le b$ are called *closed intervals* or *segments*.

3. Sets of the form $(a, b]$, $[a, b)$, $(-\infty, b]$, $[a, +\infty)$ consisting of all x such that $a < x \le b$, $a \le x < b$, $x \le b$, $x \ge a$ are called *half-open intervals*.

A *neighborhood of a point* $x_0 \in \mathbb{R}$ is defined as any open interval (a, b) containing x_0 ($a < x_0 < b$). A neighborhood of the "point" $+\infty$, $-\infty$, or ∞ is defined, respectively, as any set of the form $(b, +\infty)$, $(-\infty, c)$ or $(-\infty, -a) \cup (a, +\infty)$ (here, $a \ge 0$).

► **Lower and upper bound of a set on a straight line.** The *upper bound* of a set of real numbers is the least number that bounds the set from above. The *lower bound* of a set of real numbers is the largest number that bounds the set from below.

In more details: let a set of real numbers $X \in \mathbb{R}$ be given. A number β is called its upper bound and denoted $\sup X$ if for any $x \in X$ the inequality $x \le \beta$ holds and for any $\beta_1 < \beta$ there exists an $x_1 \in X$ such that $x_1 > \beta_1$. A number α is called the lower bound of X and denoted $\inf X$ if for any $x \in X$ the inequality $x \ge \alpha$ holds and for any $\alpha_1 > \alpha$ there exists an $x_1 \in X$ such that $x_1 < \alpha_1$.

Example 1. For a set X consisting of two numbers a and b ($a < b$), we have

$$\inf X = a, \qquad \sup X = b.$$

Example 2. For intervals (open, closed, and half-open), we have

$$\inf(a, b) = \inf[a, b] = \inf(a, b] = \inf[a, b) = a,$$
$$\sup(a, b) = \sup[a, b] = \sup(a, b] = \sup[a, b) = b.$$

One can see that the upper and lower bounds may belong to a given set (e.g., for closed intervals) and may not (e.g., for open intervals).

The symbol $+\infty$ (resp., $-\infty$) is called the upper (resp., lower) bound of a set unbounded from above (resp., from below).

► Real-valued functions of real variable. Methods of defining a function.

1°. Let D and E be two sets of real numbers. Suppose that there is a relation between the points of D and E such that to each $x \in D$ there corresponds some $y \in E$, denoted by $y = f(x)$. In this case, one speaks of a *function* f defined on the set D and taking its values in the set E. The set D is called the *domain of the function* f, and the subset of E consisting of all elements $f(x)$ is called the *range of the function* f. This functional relation is often denoted by $y = f(x)$, $f : D \to E$, $f : x \mapsto y$.

The following terms are also used: x is the *independent variable* or the *argument*; y is the *dependent variable*.

2°. The most common and convenient way to define a function is the *analytic method*: the function is defined explicitly by means of a formula (or several formulas) depending on the argument x; for instance, $y = 2\sin x + 1$.

Implicit definition of a function consists of using an equation of the form $F(x, y) = 0$, from which one calculates the value y for any fixed value of the argument x.

Parametric definition of a function consists of defining the values of the independent variable x and the dependent variable y by a pair of formulas depending on an auxiliary variable t (parameter): $x = p(t)$, $y = q(t)$.

Quite often functions are defined in terms of convergent series or by means of tables or graphs. There are some other methods of defining functions.

3°. The *graph of a function* is the representation of a function $y = f(x)$ as a line on the plane with orthogonal coordinates x, y, the points of the line having the coordinates x, $y = f(x)$, where x is an arbitrary point from the domain of the function.

► Single-valued, periodic, odd and even functions.

1°. A function is *single-valued* if each value of its argument corresponds to a unique value of the function. A function is *multi-valued* if there is at least one value of its argument corresponding to two or more values of the function. In what follows, we consider only single-valued functions, unless indicated otherwise.

2°. A function $f(x)$ is called *periodic* with period T (or T-*periodic*) if $f(x + T) = f(x)$ for any x.

3°. A function $f(x)$ is called *even* if it satisfies the condition $f(x) = f(-x)$ for any x. A function $f(x)$ is called *odd* if it satisfies the condition $f(x) = -f(-x)$ for any x.

► Decreasing, increasing, monotone, and bounded functions.

1°. A function $f(x)$ is called *increasing or strictly increasing* (resp., *nondecreasing*) on a set $D \subset \mathbb{R}$ if for any $x_1, x_2 \in D$ such that $x_1 > x_2$, we have $f(x_1) > f(x_2)$ (resp., $f(x_1) \geq f(x_2)$). A function $f(x)$ is called *decreasing or strictly decreasing* (resp., *nonincreasing*) on a set D if for all $x_1, x_2 \in D$ such that $x_1 > x_2$, we have $f(x_1) < f(x_2)$ (resp., $f(x_1) \leq f(x_2)$). All such functions are called *monotone functions*. Strictly increasing or decreasing functions are called *strictly monotone*.

2°. A function $f(x)$ is called *bounded* on a set D if $|f(x)| < M$ for all $x \in D$, where M is a finite constant. A function $f(x)$ is called *bounded from above* (*bounded from below*) on a set D if $f(x) < M$ $(M < f(x))$ for all $x \in D$, where M is a real constant.

► Composite and inverse functions.

1°. Consider a function $u = u(x)$, $x \in D$, with values $u \in E$, and let $y = f(u)$ be a function defined on E. Then the function $y = f\big(u(x)\big)$, $x \in D$, is called a *composite function* or the *superposition* of the functions f and u.

2°. Consider a function $y = f(x)$ that maps $x \in D$ into $y \in E$. The *inverse function* of $y = f(x)$ is a function $x = g(y)$ defined on E and such that $x = g(f(x))$ for all $x \in D$. The inverse function is often denoted by $g = f^{-1}$.

For strictly monotone functions $f(x)$, the inverse function always exists. In order to construct the inverse function $g(y)$, one should use the relation $y = f(x)$ to express x through y. The function $g(y)$ is monotonically increasing or decreasing together with $f(x)$.

M6.1.2. Limit of a Sequence

► **Some definitions.** Suppose that there is a correspondence between each positive integer n and some (real or complex) number denoted, for instance, by x_n. In this case, one says that a *numerical sequence* (or, simply, a *sequence*) $x_1, x_2, \dots, x_n, \dots$ is defined. Such a sequence is often denoted by $\{x_n\}$; x_n is called the *generic term* of the sequence.

Example 1. For the sequence $\{n^2 - 2\}$, we have $x_1 = -1$, $x_2 = 2$, $x_3 = 7$, $x_4 = 14$, etc.

A sequence is called *bounded* (bounded from above, bounded from below) if there is a constant M such that $|x_n| < M$ (respectively, $x_n < M$, $x_n > M$) for all $n = 1, 2, \dots$

► **Limit of a sequence.** A number b is called the *limit of a sequence* $x_1, x_2, \dots, x_n, \dots$ if for any $\varepsilon > 0$ there is $N = N(\varepsilon)$ such that $|x_n - b| < \varepsilon$ for all $n > N$.

If b is the limit of the sequence $\{x_n\}$, one writes $\lim\limits_{n\to\infty} x_n = b$ or $x_n \to b$ as $n \to \infty$.

The limit of a constant sequence $\{x_n = c\}$ exists and is equal to c, i.e., $\lim\limits_{n\to\infty} c = c$. In this case, the inequality $|x_n - c| < \varepsilon$ takes the form $0 < \varepsilon$ and holds for all n.

Example 2. Let us show that $\lim\limits_{n\to\infty} \dfrac{n}{n+1} = 1$.

Consider the difference $\left|\dfrac{n}{n+1} - 1\right| = \dfrac{1}{n+1}$. The inequality $\dfrac{1}{n+1} < \varepsilon$ holds for all $n > \dfrac{1}{\varepsilon} - 1 = N(\varepsilon)$. Therefore, for any positive ε there exists an $N = \dfrac{1}{\varepsilon} - 1$ such that for $n > N$ we have $\left|\dfrac{n}{n+1} - 1\right| < \varepsilon$.

It may happen that a sequence $\{x_n\}$ has no limit at all. For example, this is the case for the sequence $\{x_n\} = \{(-1)^n\}$. A sequence that has a finite limit is called *convergent*.

THEOREM (BOLZANO–CAUCHY). *A sequence x_n has a finite limit if and only if for any $\varepsilon > 0$, there is N such that the inequality*

$$|x_n - x_m| < \varepsilon$$

holds for all $n > N$ and $m > N$.

► **Properties of convergent sequences.**

1. Any convergent sequence can have only one limit.

2. Any convergent sequence is bounded. From any bounded sequence one can extract a convergent subsequence.*

3. If a sequence converges to b, then any of its subsequences also converges to b.

4. If $\{x_n\}$, $\{y_n\}$ are two convergent sequences, then the sequences $\{x_n \pm y_n\}$, $\{x_n \cdot y_n\}$, and $\{x_n/y_n\}$ (in this ratio, it is assumed that $y_n \neq 0$ and $\lim\limits_{n\to\infty} y_n \neq 0$) are also convergent

---

* Let $\{x_n\}$ be a given sequence and let $\{n_k\}$ be a strictly increasing sequence with k and n_k being natural numbers. The sequence $\{x_{n_k}\}$ is called a *subsequence* of the sequence $\{x_n\}$.

and

$$\begin{aligned}
&\lim_{n\to\infty}(x_n \pm y_n) = \lim_{n\to\infty} x_n \pm \lim_{n\to\infty} y_n;\\
&\lim_{n\to\infty}(cx_n) = c\lim_{n\to\infty} x_n \qquad (c = \text{const});\\
&\lim_{n\to\infty}(x_n \cdot y_n) = \lim_{n\to\infty} x_n \cdot \lim_{n\to\infty} y_n;\\
&\lim_{n\to\infty}\frac{x_n}{y_n} = \frac{\lim\limits_{n\to\infty} x_n}{\lim\limits_{n\to\infty} y_n}.
\end{aligned}$$

5. If $\{x_n\}$, $\{y_n\}$ are convergent sequences and the inequality $x_n \le y_n$ holds for all n, then $\lim\limits_{n\to\infty} x_n \le \lim\limits_{n\to\infty} y_n$.

6. If the inequalities $x_n \le y_n \le z_n$ hold for all n and $\lim\limits_{n\to\infty} x_n = \lim\limits_{n\to\infty} z_n = b$, then $\lim\limits_{n\to\infty} y_n = b$.

► **Increasing, decreasing, and monotone sequences.** A sequence $\{x_n\}$ is called *increasing or strictly increasing* (resp., *nondecreasing*) if the inequality $x_{n+1} > x_n$ (resp., $x_{n+1} \ge x_n$) holds for all n. A sequence $\{x_n\}$ is called *decreasing or strictly decreasing* (resp., *nonincreasing*) if the inequality $x_{n+1} < x_n$ (resp., $x_{n+1} \le x_n$) holds for all n. All such sequences are called *monotone sequences*. Strictly increasing or decreasing sequences are called *strictly monotone*.

THEOREM. *Any monotone bounded sequence has a finite limit.*

Example 3. It can be shown that the sequence $\left\{\left(1+\frac{1}{n}\right)^n\right\}$ is bounded and increasing. Therefore, it is convergent. Its limit is denoted by the letter e:

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \qquad (e \approx 2.71828).$$

Logarithms with the base e are called *natural* or *Napierian*, and $\log_e x$ is denoted by $\ln x$.

► **Properties of positive sequences.**

1°. If a sequence x_n ($x_n > 0$) has a limit (finite or infinite), then the sequence

$$y_n = \sqrt[n]{x_1 \cdot x_2 \ldots x_n}$$

has the same limit.

2°. From property 1° for the sequence

$$x_1,\ \frac{x_2}{x_1},\ \frac{x_3}{x_2},\ \ldots,\ \frac{x_n}{x_{n-1}},\ \frac{x_{n+1}}{x_n},\ \ldots,$$

we obtain a useful corollary

$$\lim_{n\to\infty}\sqrt[n]{x_n} = \lim_{n\to\infty}\frac{x_{n+1}}{x_n},$$

under the assumption that the second limit exists.

Example 4. Let us show that $\lim\limits_{n\to\infty} \dfrac{n}{\sqrt[n]{n!}} = e$.

Taking $x_n = \dfrac{n^n}{n!}$ and using property 2°, we get

$$\lim_{n\to\infty} \frac{n}{\sqrt[n]{n!}} = \lim_{n\to\infty} \frac{x_{n+1}}{x_n} = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = e.$$

► **Infinitely small and infinitely large sequences.** A sequence x_n converging to zero is called *infinitely small* or *infinitesimal*.

A sequence x_n whose terms infinitely grow in absolute values with the growth of n is called *infinitely large* or "*tending to infinity*." In this case, the following notation is used: $\lim\limits_{n\to\infty} x_n = \infty$. If, in addition, all terms of the sequence starting from some number are positive (negative), then one says that the sequence x_n converges to "plus (minus) infinity," and one writes $\lim\limits_{n\to\infty} x_n = +\infty$ $\bigl(\lim\limits_{n\to\infty} x_n = -\infty\bigr)$. For instance, $\lim\limits_{n\to\infty} (-1)^n n^2 = \infty$, $\lim\limits_{n\to\infty} \sqrt{n} = +\infty$, $\lim\limits_{n\to\infty} (-n) = -\infty$.

THEOREM (STOLZ). *Let x_n and y_n be two infinitely large sequences, $y_n \to +\infty$, and y_n increases with the growth of n (at least for sufficiently large n): $y_{n+1} > y_n$. Then*

$$\lim_{n\to\infty} \frac{x_n}{y_n} = \lim_{n\to\infty} \frac{x_n - x_{n-1}}{y_n - y_{n-1}},$$

provided that the right limit exists (finite or infinite).

Example 5. Let us find the limit of the sequence

$$z_n = \frac{1^k + 2^k + \cdots + n^k}{n^{k+1}}.$$

Taking $x_n = 1^k + 2^k + \cdots + n^k$ and $y_n = n^{k+1}$ in the Stolz theorem, we get

$$\lim_{n\to\infty} z_n = \lim_{n\to\infty} \frac{n^k}{n^{k+1} - (n-1)^{k+1}}.$$

Since $(n-1)^{k+1} = n^{k+1} - (k+1)n^k + \cdots$, we have $n^{k+1} - (n-1)^{k+1} = (k+1)n^k + \cdots$, and therefore

$$\lim_{n\to\infty} z_n = \lim_{n\to\infty} \frac{n^k}{(k+1)n^k + \cdots} = \frac{1}{k+1}.$$

► **Upper and lower limits of a sequence.** The limit (finite or infinite) of a subsequence of a given sequence x_n is called a partial limit of x_n. In the set of all partial limits of any sequence of real numbers, there always exists the largest and the least (finite or infinite). The largest (resp., least) partial limit of a sequence is called its *upper* (resp., *lower*) *limit*. The upper and lower limits of a sequence x_n are denoted, respectively,

$$\varlimsup_{n\to\infty} x_n, \qquad \varliminf_{n\to\infty} x_n.$$

Example 6. The upper and lower limits of the sequence $x_n = (-1)^n$ are, respectively,

$$\varlimsup_{n\to\infty} x_n = 1, \qquad \varliminf_{n\to\infty} x_n = -1.$$

A sequence x_n has a limit (finite or infinite) if and only if its upper limit coincides with its lower limit:

$$\lim_{n\to\infty} x_n = \varlimsup_{n\to\infty} x_n = \varliminf_{n\to\infty} x_n.$$

M6.1.3. Limit of a Function. Asymptotes

► Definition of the limit of a function. One-sided limits.

1°. One says that b is the *limit of a function* $f(x)$ as x tends to a if for any $\varepsilon > 0$ there is $\delta = \delta(\varepsilon) > 0$ such that $|f(x) - b| < \varepsilon$ for all x such that $0 < |x - a| < \delta$.

Notation: $\lim\limits_{x\to a} f(x) = b$ or $f(x) \to b$ as $x \to a$.

One says that b is the limit of a function $f(x)$ as x tends to $+\infty$ if for any $\varepsilon > 0$ there is $N = N(\varepsilon) > 0$ such that $|f(x) - b| < \varepsilon$ for all $x > N$.

Notation: $\lim\limits_{x\to+\infty} f(x) = b$ or $f(x) \to b$ as $x \to +\infty$.

In a similar way, one defines the limits for $x \to -\infty$ or $x \to \infty$.

THEOREM (BOLZANO–CAUCHY 1). *A function $f(x)$ has a finite limit as x tends to a (a is assumed finite) if and only if for any $\varepsilon > 0$ there is $\delta > 0$ such that the inequality*

$$|f(x_1) - f(x_2)| < \varepsilon \qquad \text{(M6.1.3.1)}$$

holds for all x_1, x_2 such that $|x_1 - a| < \delta$ and $|x_2 - a| < \delta$.

THEOREM (BOLZANO–CAUCHY 2). *A function $f(x)$ has a finite limit as x tends to $+\infty$ if and only if for any $\varepsilon > 0$ there is $\Delta > 0$ such that the inequality (M6.1.3.1) holds for all $x_1 > \Delta$ and $x_2 > \Delta$.*

2°. One says that b is the *left-hand limit* (resp., *right-hand limit*) of a function $f(x)$ as x tends to a if for any $\varepsilon > 0$ there is $\delta = \delta(\varepsilon) > 0$ such that $|f(x) - b| < \varepsilon$ for $a - \delta < x < a$ (resp., for $a < x < a + \delta$).

Notation: $\lim\limits_{x\to a-0} f(x) = b$ or $f(a - 0) = b$ (resp., $\lim\limits_{x\to a+0} f(x) = b$ or $f(a + 0) = b$).

► Properties of limits.

Let a be a number or any of the symbols ∞, $+\infty$, $-\infty$.

1. If a function has a limit at some point, this limit is unique.
2. If c is a constant function of x, then $\lim\limits_{x\to a} c = c$.
3. If there exist $\lim\limits_{x\to a} f(x)$ and $\lim\limits_{x\to a} g(x)$, then

$$\begin{aligned}
&\lim_{x\to a}\big[f(x) \pm g(x)\big] = \lim_{x\to a} f(x) \pm \lim_{x\to a} g(x);\\
&\lim_{x\to a} cf(x) = c\lim_{x\to a} f(x) \qquad (c = \text{const});\\
&\lim_{x\to a} f(x)\cdot g(x) = \lim_{x\to a} f(x)\cdot \lim_{x\to a} g(x);\\
&\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\lim\limits_{x\to a} f(x)}{\lim\limits_{x\to a} g(x)} \qquad \Big(\text{if } g(x) \neq 0,\ \lim_{x\to a} g(x) \neq 0\Big).
\end{aligned}$$

4. Let $f(x) \le g(x)$ in a neighborhood of a point a ($x \neq a$). Then $\lim\limits_{x\to a} f(x) \le \lim\limits_{x\to a} g(x)$, provided that these limits exist.

5. If $f(x) \le g(x) \le h(x)$ in a neighborhood of a point a and $\lim\limits_{x\to a} f(x) = \lim\limits_{x\to a} h(x) = b$, then $\lim\limits_{x\to a} g(x) = b$.

These properties hold also for one-sided limits.

► Limits of some functions.

$$\textit{First noteworthy limit}: \qquad \lim_{x\to 0} \frac{\sin x}{x} = 1.$$

$$\textit{Second noteworthy limit}: \qquad \lim_{x\to \infty} \Big(1 + \frac{1}{x}\Big)^x = e.$$

Some other frequently used limits:

$$\lim_{x\to 0}\frac{(1+x)^n-1}{x}=n,\quad \lim_{x\to\infty}\frac{a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0}{b_nx^n+b_{n-1}x^{n-1}+\cdots+b_1x+b_0}=\frac{a_n}{b_n},$$

$$\lim_{x\to 0}\frac{1-\cos x}{x^2}=\frac{1}{2},\quad \lim_{x\to 0}\frac{\tan x}{x}=1,\quad \lim_{x\to 0}\frac{\arcsin x}{x}=1,\quad \lim_{x\to 0}\frac{\arctan x}{x}=1,$$

$$\lim_{x\to 0}\frac{e^x-1}{x}=1,\quad \lim_{x\to 0}\frac{a^x-1}{x}=\ln a,\quad \lim_{x\to 0}\frac{\ln(1+x)}{x}=1,\quad \lim_{x\to 0}\frac{\log_a(1+x)}{x}=\log_a e,$$

$$\lim_{x\to 0}\frac{\sinh x}{x}=1,\quad \lim_{x\to 0}\frac{\tanh x}{x}=1,\quad \lim_{x\to 0}\frac{\operatorname{arcsinh} x}{x}=1,\quad \lim_{x\to 0}\frac{\operatorname{arctanh} x}{x}=1,$$

$$\lim_{x\to +0}x^a\ln x=0,\quad \lim_{x\to +\infty}x^{-a}\ln x=0,\quad \lim_{x\to +\infty}x^ae^{-x}=0,\quad \lim_{x\to +0}x^x=1,$$

where $a>0$ and $b_n\neq 0$.

► See Subsection M6.2.3, where L'Hospital rules for calculating limits with the help of derivatives are given.

► **Asymptotes of the graph of a function.** An *asymptote* of the graph of a function $y=f(x)$ is a straight line whose distance from a point (x,y) on the graph of $y=f(x)$ tends to zero if at least one of the coordinates (x,y) tends to infinity.

The line $x=a$ is a *vertical asymptote* of the graph of the function $y=f(x)$ if at least one of the one-sided limits of $f(x)$ as $x\to a\pm 0$ is equal to $+\infty$ or $-\infty$.

The line $y=kx+b$ is an *oblique asymptote* of the graph of $y=f(x)$ if at least one of the limit relations holds:

$$\lim_{x\to+\infty}[f(x)-kx-b]=0\quad\text{or}\quad\lim_{x\to-\infty}[f(x)-kx-b]=0.$$

If there exist finite limits

$$\lim_{x\to+\infty}\frac{f(x)}{x}=k,\qquad \lim_{x\to+\infty}[f(x)-kx]=b,\tag{M6.1.3.2}$$

then the line $y=kx+b$ is an oblique asymptote of the graph for $x\to+\infty$ (in a similar way, one defines an asymptote for $x\to-\infty$).

Example. Let us find the asymptotes of the graph of the function $y=\dfrac{x^2}{x-1}$.

1°. The graph has a vertical asymptote $x=1$, since $\lim\limits_{x\to 1}\dfrac{x^2}{x-1}=\infty$.

2°. Moreover, for $x\to\pm\infty$, there is an oblique asymptote $y=kx+b$ whose coefficients are determined by the formulas (M6.1.3.2):

$$k=\lim_{x\to\pm\infty}\frac{x}{x-1}=1,\qquad b=\lim_{x\to\pm\infty}\left(\frac{x^2}{x-1}-x\right)=\lim_{x\to\pm\infty}\frac{x}{x-1}=1.$$

Thus, the equation of the oblique asymptote has the form $y=x+1$. Fig. M6.1 shows the graph of the function under consideration and its asymptotes.

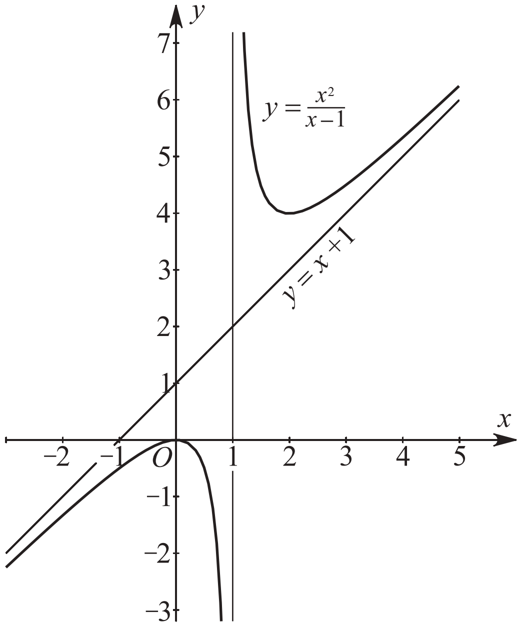

Figure M6.1. The graph of the function $y = \dfrac{x^2}{x-1}$ and its asymptotes.

M6.1.4. Infinitely Small and Infinitely Large Functions

► **Definitions.** A function $f(x)$ is called *infinitely small* for $x \to a$ if $\lim\limits_{x\to a} f(x) = 0$.

A function $f(x)$ is said to be *infinitely large* for $x \to a$ if for any $K > 0$ the inequality $|f(x)| > K$ holds for all $x \neq a$ in a small neighborhood of the point a. In this case, one writes $f(x) \to \infty$ as $x \to a$ or $\lim\limits_{x\to a} f(x) = \infty$. (In these definitions, a is a finite number or any of the symbols ∞, $+\infty$, $-\infty$.) If $f(x)$ is infinitely large for $x \to a$ and $f(x) > 0$ ($f(x) < 0$) in a neighborhood of a (for $x \neq a$), one writes $\lim\limits_{x\to a} f(x) = +\infty$ (resp., $\lim\limits_{x\to a} f(x) = -\infty$).

► **Properties of infinitely small and infinitely large functions.**

1. The sum and the product of finitely many infinitely small functions for $x \to a$ is an infinitely small function.

2. The product of an infinitely small function $f(x)$ for $x \to a$ and a function $g(x)$ which is bounded in a neighborhood U of the point a (i.e., $|g(x)| < M$ for all $x \in U$, where $M > 0$ is a constant) is an infinitely small function.

3. $\lim\limits_{x\to a} f(x) = b$ if and only if $f(x) = b + g(x)$, where $g(x)$ is infinitely small for $x \to a$.

4. A function $f(x)$ is infinitely large at some point if and only if the function $g(x) = 1/f(x)$ is infinitely small at the same point.

► **Comparison of infinitely small quantities. Symbols of the order: *O* and *o*.** Functions $f(x)$ and $g(x)$ that are infinitely small for $x \to a$ are called *equivalent* near a if $\lim\limits_{x\to a} \dfrac{f(x)}{g(x)} = 1$. In this case one writes $f(x) \sim g(x)$.

Examples of equivalent infinitely small functions:

$$(1+\varepsilon)^n - 1 \sim n\varepsilon, \qquad a^\varepsilon - 1 \sim \varepsilon \ln a, \qquad \log_a(1+\varepsilon) \sim \varepsilon \log_a e,$$
$$\sin\varepsilon \sim \varepsilon, \quad \tan\varepsilon \sim \varepsilon, \quad 1 - \cos\varepsilon \sim \tfrac{1}{2}\varepsilon^2, \quad \arcsin\varepsilon \sim \varepsilon, \quad \arctan\varepsilon \sim \varepsilon,$$

where $\varepsilon = \varepsilon(x)$ is infinitely small for $x \to a$.

Functions $f(x)$ and $g(x)$ are said to be of the *same order* for $x \to a$, and one writes $f(x) = O\bigl(g(x)\bigr)$ if $\lim\limits_{x\to a} \dfrac{f(x)}{g(x)} = K,\ 0 < |K| < \infty$.*

A function$f(x)$ is of a *higher order of smallness* compared with $g(x)$ for $x \to a$ if $\lim\limits_{x\to a} \dfrac{f(x)}{g(x)} = 0$, and in this case, one writes $f(x) = o\bigl(g(x)\bigr)$.

M6.1.5. Continuous Functions. Discontinuities of the First and the Second Kind

► **Continuous functions.** A function $f(x)$ is called *continuous* at a point $x = a$ if it is defined at that point and its neighborhood and $\lim\limits_{x\to a} f(x) = f(a)$.

For continuous functions, a small variation of their argument $\Delta x = x - a$ results in a small variation of the function $\Delta y = f(x) - f(a)$, i.e., $\Delta y \to 0$ as $\Delta x \to 0$. (This property is often used as a definition of continuity.)

A function $f(x)$ is called *right-continuous* at a point $x = a$ if it is defined at that point (and to its right) and $\lim\limits_{x\to a+0} f(x) = f(a)$. A function $f(x)$ is called *left-continuous* at a point $x = a$ if it is defined at that point (and to its left) and $\lim\limits_{x\to a-0} f(x) = f(a)$.

► **Properties of continuous functions.**

1. Suppose that functions $f(x)$ and $g(x)$ are continuous at some point a. Then the functions $f(x) \pm g(x)$, $cf(x)$, $f(x)g(x)$, $\dfrac{f(x)}{g(x)}$ $(g(a) \neq 0)$ are also continuous at a.

2. Suppose that a function $f(x)$ is continuous on the segment $[a, b]$ and takes values of different signs at its endpoints, i.e., $f(a)f(b) < 0$. Then there is a point c between a and b at which $f(x)$ vanishes:

$$f(c) = 0 \qquad (a < c < b).$$

3. If $f(x)$ is continuous at a point a and $f(a) > 0$ (resp., $f(a) < 0$), then there is $\delta > 0$ such that $f(x) > 0$ (resp., $f(x) < 0$) for all $x \in (a - \delta, a + \delta)$.

4. Any function $f(x)$ that is continuous at each point of a segment $[a, b]$ attains its largest and its smallest values, M and m, on that segment.

5. A function $f(x)$ that is continuous on a segment $[a, b]$ takes any value $c \in [m, M]$ on that segment, where m and M are, respectively, its smallest and its largest values on $[a, b]$.

6. If $f(x)$ is continuous and increasing (resp., decreasing) on a segment $[a, b]$, then on the segment $\bigl[f(a), f(b)\bigr]$ (resp., $\bigl[f(b), f(a)\bigr]$) the inverse function $x = g(y)$ exists, and is continuous and increasing (resp., decreasing).

7. If $u(x)$ is continuous at a point a and $f(u)$ is continuous at $b = u(a)$, then the composite function $f\bigl(u(x)\bigr)$ is continuous at a.

Remark. Any elementary function is continuous at each point of its domain.

► **Points of discontinuity of a function.** A point a is called a *point of discontinuity of the first kind* for a function $f(x)$ if there exist finite one-sided limits $f(a + 0)$ and $f(a - 0)$, but the relations $\lim\limits_{x\to a+0} f(x) = \lim\limits_{x\to a-0} f(x) = f(a)$ do not hold. The value $|f(a + 0) - f(a - 0)|$ is called the *jump* of the function at the point a. In particular, if $f(a + 0) = f(a - 0) \neq f(a)$, then a is called a *point of removable discontinuity*.

---

* There is another definition of the symbol O. Namely, $f(x) = O\bigl(g(x)\bigr)$ for $x \to a$ if the inequality $|f(x)| \le K|g(x)|$, $K =$ const, holds in some neighborhood of the point a (for $x \neq a$).

Examples of functions with discontinuities of the first kind.

1. The function $f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$ has a jump equal to 1 at the discontinuity point $x = 0$.

2. The function $f(x) = \begin{cases} 0 & \text{for } x \neq 0 \\ 1 & \text{for } x = 0 \end{cases}$ has a removable discontinuity at the point $x = 0$.

A point a is called a *point of discontinuity of the second kind* if at least one of the one-sided limits $f(a+0)$ or $f(a-0)$ does not exist or is equal to infinity.

Examples of functions with discontinuities of the second kind.

1. The function $f(x) = \sin\dfrac{1}{x}$ has a second-kind discontinuity at the point $x = 0$ (since this function has no one-sided limits as $x \to \pm 0$).

2. The function $f(x) = 1/x$ has an infinite limit as $x \to 0$, so it has a second-kind discontinuity at the point $x = 0$.

M6.1.6. Convex and Concave Functions

► Definition of convex and concave functions.

1°. A function $f(x)$ defined and continuous on a segment $[a, b]$ is called *convex* (or *convex downward*) if for any x_1, x_2 in $[a, b]$, the *Jensen inequality* holds:

$$f\left(\frac{x_1 + x_2}{2}\right) \leq \frac{f(x_1) + f(x_2)}{2}. \qquad \text{(M6.1.6.1)}$$

The geometrical meaning of convexity is that all points of the graph curve between two graph points lie below or on the rectilinear segment joining the two graph points (see Fig. M6.2 *a*).

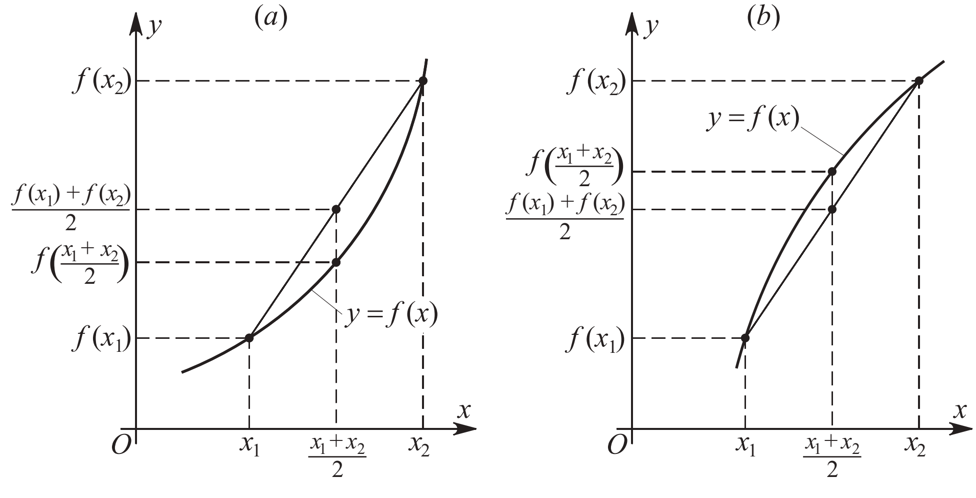

Figure M6.2. Graphs of convex (*a*) and concave (*b*) functions.

If for $x_1 \neq x_2$, condition (M6.1.6.1) holds with < instead of ≤, then the function $f(x)$ is called *strictly convex*.

2°. A function $f(x)$ defined and continuous on a segment $[a, b]$ is called *concave* (or *convex upward*) if for any x_1, x_2 in $[a, b]$ the following inequality holds:

$$f\left(\frac{x_1 + x_2}{2}\right) \geq \frac{f(x_1) + f(x_2)}{2}. \qquad \text{(M6.1.6.2)}$$

The geometrical meaning of concavity is that all points of the graph curve between two graph points lie above or on the rectilinear segment joining the two graph points (see Fig. M6.2 *b*).

If for $x_1 \neq x_2$, condition (M6.1.6.2) holds with > instead of ≥, then the function $f(x)$ is called *strictly concave*.

► **Some properties of convex and concave functions.**

1. The product of a convex (concave) function and a positive constant is a convex (concave) function.

2. The sum of two or more convex (concave) functions is a convex (concave) function.

3. A non-constant convex (resp., concave) function $f(x)$ on a segment $[a, b]$ cannot attain its largest (resp., smallest) value inside the segment.

4. A function $f(x)$ that is continuous on a segment $[a, b]$ and twice differentiable on the interval (a, b) is convex downward (resp., convex upward) if and only if $f''(x) \geq 0$ (resp., $f''(x) \leq 0$) on that interval.

M6.1.7. Convergence of Functions

► **Pointwise, uniform, and nonuniform convergence of functions.** Let $\{f_n(x)\}$ be a sequence of functions defined on a set $X \subset \mathbb{R}$. The sequence $\{f_n(x)\}$ is said to be *pointwise convergent* to $f(x)$ as $n \to \infty$ if for any fixed $x \in X$, the numerical sequence $\{f_n(x)\}$ converges to $f(x)$. The sequence $\{f_n(x)\}$ is said to be *uniformly convergent* to a function $f(x)$ on X as $n \to \infty$ if for any $\varepsilon > 0$ there is an integer $N = N(\varepsilon)$ such that for all $n > N$ and all $x \in X$, the following inequality holds:

$$|f_n(x) - f(x)| < \varepsilon. \qquad \text{(M6.1.7.1)}$$

Note that in this definition, N is independent of x. For a sequence $\{f_n(x)\}$ pointwise convergent to $f(x)$ as $n \to \infty$, by definition, for any $\varepsilon > 0$ and any $x \in X$, there is $N = N(\varepsilon, x)$ such that (M6.1.7.1) holds for all $n > N(\varepsilon, x)$.

If one cannot find such an N independent of x and depending only on ε (i.e., one cannot ensure (M6.1.7.1) uniformly; to be more precise, there is an $\varepsilon > 0$ such that for any $N > 0$ there is a $k_N > N$ and $x_N \in X$ such that $|f_{k_N}(x_N) - f(x_N)| \geq \varepsilon$), then one says that the sequence $\{f_n(x)\}$ *converges nonuniformly* to $f(x)$ on the set X.

► **Basic theorems.** Let X be an interval on the real axis.

THEOREM. *Let $f_n(x)$ be a sequence of continuous functions uniformly convergent to $f(x)$ on X. Then $f(x)$ is continuous on X.*

COROLLARY. *If the limit function $f(x)$ of a pointwise convergent sequence of continuous functions $\{f_n(x)\}$ is discontinuous, then the convergence of the sequence $\{f_n(x)\}$ is nonuniform.*

Example. The sequence $\{f_n(x)\} = \{x^n\}$ converges to $f(x) \equiv 0$ as $n \to \infty$ uniformly on each segment $[0, a]$, $0 < a < 1$. However, on the segment $[0, 1]$ this sequence converges nonuniformly to the discontinuous function $f(x) = \begin{cases} 0 & \text{for } 0 \leq x < 1, \\ 1 & \text{for } x = 1. \end{cases}$

CAUCHY CRITERION. *A sequence of functions $\{f_n(x)\}$ defined on a set $X \in \mathbb{R}$ uniformly converges to $f(x)$ as $n \to \infty$ if and only if for any $\varepsilon > 0$ there is an integer $N = N(\varepsilon) > 0$ such that for all $n > N$ and $m > N$, the inequality $|f_n(x) - f_m(x)| < \varepsilon$ holds for all $x \in X$.*

► **Geometrical meaning of uniform convergence.** Let $f_n(x)$ be continuous functions on the segment $[a, b]$ and suppose that $\{f_n(x)\}$ uniformly converges to a continuous function $f(x)$ as $n \to \infty$. Then all curves $y = f_n(x)$, for sufficiently large $n > N$, belong to the strip between the two curves $y = f(x) - \varepsilon$ and $y = f(x) + \varepsilon$ (see Fig. M6.3).

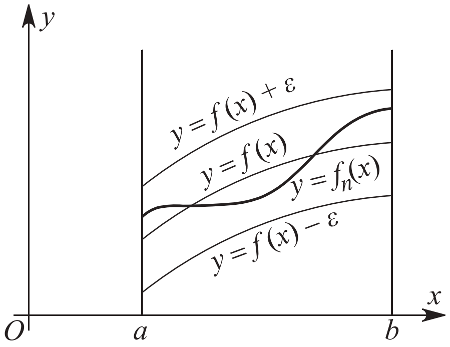

Figure M6.3. Geometrical meaning of uniform convergence of a sequence of functions $\{f_n(x)\}$ to a continuous function $f(x)$.

M6.2. Differential Calculus for Functions of a Single Variable

M6.2.1. Derivative and Differential: Their Geometrical and Physical Meaning

► **Definition of derivative and differential.** The *derivative* of a function $y = f(x)$ at a point x is the limit of the ratio

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x},$$

where $\Delta y = f(x+\Delta x) - f(x)$ is the increment of the function corresponding to the increment of the argument Δx. The derivative y' is also denoted by y'_x, $\dot{y}$, $\dfrac{dy}{dx}$, $f'(x)$, $\dfrac{df(x)}{dx}$.

Example 1. Let us calculate the derivative of the function $f(x) = x^2$.
By definition, we have

$$f'(x) = \lim_{\Delta x \to 0} \frac{(x + \Delta x)^2 - x^2}{\Delta x} = \lim_{\Delta x \to 0} (2x + \Delta x) = 2x.$$

The increment Δx is also called the differential of the independent variable x and is denoted by dx.

A function $f(x)$ that has a derivative at a point x is called *differentiable* at that point. The differentiability of $f(x)$ at a point x is equivalent to the condition that the increment of the function, $\Delta y = f(x + dx) - f(x)$, at that point can be represented in the form $\Delta y = f'(x)\,dx + o(dx)$ (the second term is an infinitely small quantity compared with dx as $dx \to 0$; see Subsection M6.1.4).

A function differentiable at some point x is continuous at that point. The converse is not true, in general; continuity does not always imply differentiability.

A function $f(x)$ is called *differentiable* on a set D (interval, segment, etc.) if for any $x \in D$ there exists the derivative $f'(x)$. A function $f(x)$ is called *continuously differentiable* on D if it has the derivative $f'(x)$ at each point $x \in D$ and $f'(x)$ is a continuous function on D.

The *differential* dy of a function $y = f(x)$ is the principal linear part of its increment Δy at the point x, so that $dy = f'(x)dx$, $\Delta y = dy + o(dx)$.

The approximate relation $\Delta y \approx dy$ or $f(x + \Delta x) \approx f(x) + f'(x)\Delta x$ (for small Δx) is often used in numerical analysis.

► Physical and geometrical meaning of the derivative. Tangent line.

1°. Let $y = f(x)$ be the function describing the path y traversed by a body by the time x. Then the derivative $f'(x)$ is the velocity of the body at the instant x.

2°. The *tangent line* or simply the *tangent* to the graph of the function $y = f(x)$ at a point $M(x_0, y_0)$, where $y_0 = f(x_0)$, is defined as the straight line determined by the limit position of the secant MN as the point N tends to M along the graph. If α is the angle between the x-axis and the tangent line, then $f'(x_0) = \tan\alpha$ is the slope ratio of the tangential line (Fig. M6.4).

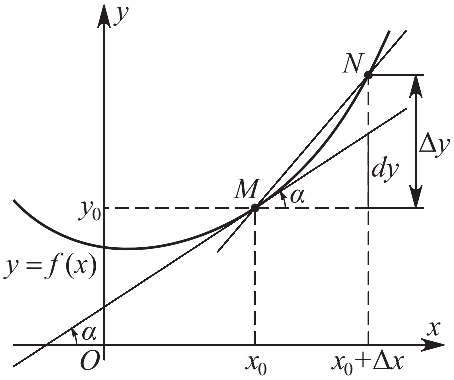

Figure M6.4. The tangent to the graph of a function $y = f(x)$ at a point (x_0, y_0).

Equation of the tangent line to the graph of a function $y = f(x)$ at a point (x_0, y_0):

$$y - y_0 = f'(x_0)(x - x_0).$$

Equation of the normal to the graph of a function $y = f(x)$ at a point (x_0, y_0):

$$y - y_0 = -\frac{1}{f'(x_0)}(x - x_0).$$

M6.2.2. Table of Derivatives and Differentiation Rules

The derivative of any elementary function can be calculated with the help of derivatives of basic elementary functions and differentiation rules.

► Table of derivatives of basic elementary functions (a = const).

$$(a)' = 0, \qquad (x^a)' = ax^{a-1},$$
$$(e^x)' = e^x, \qquad (a^x)' = a^x \ln a,$$
$$(\ln x)' = \frac{1}{x}, \qquad (\log_a x)' = \frac{1}{x\ln a},$$
$$(\sin x)' = \cos x, \qquad (\cos x)' = -\sin x,$$
$$(\tan x)' = \frac{1}{\cos^2 x}, \qquad (\cot x)' = -\frac{1}{\sin^2 x},$$
$$(\arcsin x)' = \frac{1}{\sqrt{1-x^2}}, \qquad (\arccos x)' = -\frac{1}{\sqrt{1-x^2}},$$
$$(\arctan x)' = \frac{1}{1+x^2}, \qquad (\text{arccot}\, x)' = -\frac{1}{1+x^2},$$
$$(\sinh x)' = \cosh x, \qquad (\cosh x)' = \sinh x,$$

$$(\tanh x)' = \frac{1}{\cosh^2 x}, \qquad (\coth x)' = -\frac{1}{\sinh^2 x},$$
$$(\operatorname{arcsinh} x)' = \frac{1}{\sqrt{1+x^2}}, \qquad (\operatorname{arccosh} x)' = \frac{1}{\sqrt{x^2-1}},$$
$$(\operatorname{arctanh} x)' = \frac{1}{1-x^2} \quad (x^2 < 1), \qquad (\operatorname{arccoth} x)' = \frac{1}{1-x^2} \quad (x^2 > 1).$$

► **Differentiation rules.**

1. *Derivative of the sum (difference) of functions*:

$$[u(x) \pm v(x)]' = u'(x) \pm v'(x).$$

2. *Derivative of the product of a function and a constant*:

$$[au(x)]' = au'(x) \qquad (a = \text{const}).$$

3. *Derivative of the product of functions*:

$$[u(x)v(x)]' = u'(x)v(x) + u(x)v'(x).$$

4. *Derivative of the quotient of functions*:

$$\left[\frac{u(x)}{v(x)}\right]' = \frac{u'(x)v(x) - u(x)v'(x)}{v^2(x)}.$$

5. *Derivative of a composite function*:

$$\big[f(u(x))\big]' = f'_u(u)u'(x).$$

6. *Derivative of a parametrically defined function* $x = x(t)$, $y = y(t)$:

$$y'_x = \frac{y'_t}{x'_t}.$$

7. *Derivative of an implicit function defined by the equation* $F(x, y) = 0$:

$$y'_x = -\frac{F_x}{F_y} \qquad (F_x \text{ and } F_y \text{ are partial derivatives}).$$

8. *Derivative of the inverse function* $x = x(y)$ (for details see footnote*):

$$x'_y = \frac{1}{y'_x}.$$

9. *Derivative of a composite exponential function*:

$$[u(x)^{v(x)}]' = u'vu^{v-1} + v'u^v \ln u = u^v\Big(u'\frac{v}{u} + v' \ln u\Big).$$

---

* Let $y = f(x)$ be a differentiable monotone function on the interval (a, b) and $f'(x_0) \neq 0$, where $x_0 \in (a, b)$. Then the inverse function $x = g(y)$ is differentiable at the point $y_0 = f(x_0)$ and $g'(y_0) = \dfrac{1}{f'(x_0)}$.

10. *Derivative of a composite function of two arguments*:

$$\bigl[f\bigl(u(x),v(x)\bigr)\bigr]' = f_u(u,v)u' + f_v(u,v)v' \qquad (f_u \text{ and } f_v \text{ are partial derivatives}).$$

11. *Logarithmic derivative*:

$$[\ln u(x)]' = \frac{u'(x)}{u(x)}.$$

Example 1. Let us calculate the derivative of the function $\dfrac{x^2}{2x+1}$.
Using the rule of differentiating the ratio of two functions, we obtain

$$\left(\frac{x^2}{2x+1}\right)' = \frac{(x^2)'(2x+1) - x^2(2x+1)'}{(2x+1)^2} = \frac{2x(2x+1)-2x^2}{(2x+1)^2} = \frac{2x^2+2x}{(2x+1)^2}.$$

Example 2. Let us calculate the derivative of the function $\ln\cos x$.
Using the chain rule or the logarithmic derivative formula, we get

$$(\ln\cos x)' = \frac{1}{\cos x}(\cos x)' = -\tan x.$$

Example 3. Let us calculate the derivative of the function x^x. Using the rule of differentiating the composite exponential function with $u(x) = v(x) = x$, we have

$$(x^x)' = x^x \ln x + x x^{x-1} = x^x(\ln x + 1).$$

M6.2.3. Theorems about Differentiable Functions. L'Hospital Rule

► Main theorems about differentiable functions.

ROLLE THEOREM. *If the function $y = f(x)$ is continuous on the segment $[a,b]$, differentiable on the interval (a,b), and $f(a) = f(b)$, then there is a point $c \in (a,b)$ such that $f'(c) = 0$.*

LAGRANGE THEOREM. *If the function $y = f(x)$ is continuous on the segment $[a,b]$ and differentiable on the interval (a,b), then there is a point $c \in (a,b)$ such that*

$$f(b) - f(a) = f'(c)(b-a).$$

This relation is called the *formula of finite increments*.

CAUCHY THEOREM. *Let $f(x)$ and $g(x)$ be two functions that are continuous on the segment $[a,b]$, differentiable on the interval (a,b), and $g'(x) \neq 0$ for all $x \in (a,b)$. Then there is a point $c \in (a,b)$ such that*

$$\frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(c)}{g'(c)}.$$

► L'Hospital's rules on indeterminate expressions of the form $0/0$ and ∞/∞.

THEOREM 1. *Let $f(x)$ and $g(x)$ be two functions defined in a neighborhood of a point a, vanishing at this point, $f(a) = g(a) = 0$, and having the derivatives $f'(a)$ and $g'(a)$, with $g'(a) \neq 0$. Then*

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)}.$$

Example 1. Let us calculate the limit $\displaystyle\lim_{x\to 0} \frac{\sin x}{1-e^{-2x}}$.
Here, both the numerator and the denominator vanish for $x = 0$. Let us calculate the derivatives

$$f'(x) = (\sin x)' = \cos x \qquad \Longrightarrow \qquad f'(0) = 1,$$

$$g'(x) = (1-e^{-2x})' = 2e^{-2x} \qquad \Longrightarrow \qquad g'(0) = 2 \neq 0.$$

By the L'Hospital rule, we find that

$$\lim_{x\to 0} \frac{\sin x}{1-e^{-2x}} = \frac{f'(0)}{g'(0)} = \frac{1}{2}.$$

THEOREM 2. *Let $f(x)$ and $g(x)$ be two functions defined in a neighborhood of a point a, vanishing at a, together with their derivatives up to the order $n-1$ inclusively. Suppose also that the derivatives $f^{(n)}(a)$ and $g^{(n)}(a)$ exist and are finite, $g^{(n)}(a) \neq 0$. Then*

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{f^{(n)}(a)}{g^{(n)}(a)}.$$

THEOREM 3. *Let $f(x)$ and $g(x)$ be differentiable functions and $g'(x) \neq 0$ in a neighborhood of a point a $(x \neq a)$. If $f(x)$ and $g(x)$ are infinitely small or infinitely large functions for $x \to a$, i.e., the ratio $\dfrac{f(x)}{g(x)}$ at the point a is an indeterminate expression of the form $\dfrac{0}{0}$ or $\dfrac{\infty}{\infty}$, then*

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}$$

(provided that there exists a finite or infinite limit of the ratio of the derivatives).

Remark. The L'Hospital rule 3 is applicable also in the case of a being one of the symbols ∞, $+\infty$, $-\infty$.

► **Methods for interpreting other indeterminate expressions.**

1°. Expressions of the form $0\cdot\infty$ and $\infty-\infty$ can be reduced to indeterminate expressions $\dfrac{0}{0}$ or $\dfrac{\infty}{\infty}$ by means of algebraic transformations, for instance:

$$u(x)\,v(x) = \frac{u(x)}{1/v(x)} \qquad \text{transformation rule} \quad 0\cdot\infty \Longrightarrow \frac{0}{0},$$

$$u(x)-v(x) = \left(\frac{1}{v(x)} - \frac{1}{u(x)}\right) : \frac{1}{u(x)v(x)} \qquad \text{transformation rule} \quad \infty-\infty \Longrightarrow \frac{0}{0}.$$

2°. Indeterminate expressions of the form 1^∞, ∞^0, 0^0 can be reduced to expressions of the form $\dfrac{0}{0}$ or $\dfrac{\infty}{\infty}$ by taking logarithm and using the formulas $\ln u^v = v\ln u = \dfrac{\ln u}{1/v}$.

Example 2. Let us calculate the limit $\lim\limits_{x\to 0}(\cos x)^{1/x^2}$.

We have the indeterminate expression 1^∞. We find that

$$\ln \lim_{x\to 0}(\cos x)^{1/x^2} = \lim_{x\to 0}\ln(\cos x)^{1/x^2} = \lim_{x\to 0}\frac{\ln\cos x}{x^2} = \lim_{x\to 0}\frac{(\ln\cos x)'}{(x^2)'} = \lim_{x\to 0}\frac{(-\tan x)}{2x} = -\frac{1}{2}.$$

Therefore, $\lim\limits_{x\to 0}(\cos x)^{1/x^2} = e^{-1/2} = \dfrac{1}{\sqrt{e}}$.

M6.2.4. Higher-Order Derivatives and Differentials. Taylor's Formula

► **Derivatives and differentials of higher orders.** The *second-order derivative* or the second derivative of a function $y = f(x)$ is the derivative of the derivative $f'(x)$. The second derivative is denoted by y'' and also by y''_{xx}, $\dfrac{d^2y}{dx^2}$, $f''(x)$.

The derivative of the second derivative of a function $y = f(x)$ is called the third-order derivative, $y''' = (y'')'$. The nth-order derivative of the function $y = f(x)$ is defined as the derivative of its $(n-1)$th derivative:

$$y^{(n)} = (y^{(n-1)})'.$$

The nth-order derivative is also denoted by $y_x^{(n)}$, $\dfrac{d^n y}{dx^n}$, $f^{(n)}(x)$.

The *second-order differential* is the differential of the first-order differential, $d^2y = d(dy)$. If x is the independent variable, then $d^2y = y'' \cdot (dx)^2$. In a similar way, one defines differentials of higher orders.

► **Table of higher-order derivatives of some elementary functions.**

$$(x^a)^{(n)} = a(a-1)\dots(a-n+1)x^{a-n}, \qquad (a^x)^{(n)} = (\ln a)^n a^x,$$

$$(\ln x)^{(n)} = (-1)^{n-1}(n-1)!\,\frac{1}{x^n}, \qquad (\log_a x)^{(n)} = (-1)^{n-1}\frac{(n-1)!}{\ln a}\,\frac{1}{x^n},$$

$$(\sin x)^{(n)} = \sin\left(x + \frac{\pi n}{2}\right), \qquad (\cos x)^{(n)} = \cos\left(x + \frac{\pi n}{2}\right),$$

$$(\sinh x)^{(n)} = \begin{cases} \cosh x & \text{if } n \text{ is odd,} \\ \sinh x & \text{if } n \text{ is even,} \end{cases} \qquad (\cosh x)^{(n)} = \begin{cases} \cosh x & \text{if } n \text{ is even,} \\ \sinh x & \text{if } n \text{ is odd.} \end{cases}$$

► **Rules for calculating higher-order derivatives.**

1. *Derivative of a sum (difference) of functions*:

$$[u(x) \pm v(x)]^{(n)} = u^{(n)}(x) \pm v^{(n)}(x).$$

2. *Derivatives of a function multiplied by a constant*:

$$[au(x)]^{(n)} = au^{(n)}(x) \qquad (a = \text{const}).$$

3. *Derivatives of a product*:

$$[u(x)v(x)]'' = u''(x)v(x) + 2u'(x)v'(x) + u(x)v''(x),$$
$$[u(x)v(x)]''' = u'''(x)v(x) + 3u''(x)v'(x) + 3u'(x)v''(x) + u(x)v'''(x),$$
$$[u(x)v(x)]^{(n)} = \sum_{k=0}^{n} C_n^k u^{(k)}(x)v^{(n-k)}(x) \qquad \text{(Leibniz formula),}$$

where C_n^k are binomial coefficients, $u^{(0)}(x) = u(x)$, $v^{(0)}(x) = v(x)$.

4. *Derivatives of a composite function*:

$$\big[f(u(x))\big]'' = f''_{uu}(u'_x)^2 + f'_u u''_{xx},$$
$$\big[f(u(x))\big]''' = f'''_{uuu}(u'_x)^3 + 3f''_{uu}u'_x u''_{xx} + f'_u u'''_{xxx}.$$

5. *Derivatives of a parametrically defined function* $x = x(t)$, $y = y(t)$:

$$y'' = \frac{x'_t y''_{tt} - y'_t x''_{tt}}{(x'_t)^3}, \quad y''' = \frac{(x'_t)^2 y'''_{ttt} - 3x'_t x''_{tt} y''_{tt} + 3y'_t (x''_{tt})^2 - x'_t y'_t x'''_{ttt}}{(x'_t)^5}, \quad y^{(n)} = \frac{(y^{(n-1)})'_t}{x'_t}.$$

6. *Derivatives of an implicit function defined by the equation* $F(x, y) = 0$:

$$y'' = \frac{1}{F_y^3}\big(-F_y^2 F_{xx} + 2F_x F_y F_{xy} - F_x^2 F_{yy}\big),$$
$$y''' = \frac{1}{F_y^5}\big(-F_y^4 F_{xxx} + 3F_x F_y^3 F_{xxy} - 3F_x^2 F_y^3 F_{xyy} + F_x^3 F_y F_{yyy} + 3F_y^3 F_{xx} F_{xy}$$
$$- 3F_x F_y^2 F_{xx} F_{yy} - 6F_x F_y^2 F_{xy}^2 - 3F_x^3 F_{yy}^2 + 9F_x^2 F_y F_{xy} F_{yy}\big),$$

where the subscripts denote the corresponding partial derivatives.

7. *Derivatives of the inverse function* $x = x(y)$:

$$x''_{yy} = -\frac{y''_{xx}}{(y'_x)^3}, \qquad x'''_{yyy} = -\frac{y'''_{xxx}}{(y'_x)^4} + 3\frac{(y''_{xx})^2}{(y'_x)^5}, \qquad x^{(n)}_y = \frac{1}{y'_x}[x^{(n-1)}_y]'_x.$$

► **Taylor's formula.** Suppose that in a neighborhood of a point $x = a$, the function $y = f(x)$ has derivatives up to the order $(n + 1)$ inclusively. Then for all x in that neighborhood, the following representation holds:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n(x), \quad \text{(M6.2.4.1)}$$

where $R_n(x)$ is the *remainder term* in Taylor's formula.

The remainder term can be represented in different forms:

$$R_n(x) = o[(x-a)^n] \qquad \text{(Peano)},$$

$$R_n(x) = \frac{f^{(n+1)}\big(a + k(x-a)\big)}{(n+1)!}(x-a)^{n+1} \qquad \text{(Lagrange)},$$

$$R_n(x) = \frac{f^{(n+1)}\big(a + k(x-a)\big)}{n!}(1-k)^n(x-a)^{n+1} \qquad \text{(Cauchy)},$$

$$R_n(x) = \frac{f^{(n+1)}\big(a + k(x-a)\big)}{n!\,p}(1-k)^{n+1-p}(x-a)^{n+1} \qquad \text{(Schlömilch and Roche)},$$

$$R_n(x) = \frac{1}{n!}\int_a^x f^{(n+1)}(t)(x-t)^n\,dt \qquad \text{(integral form)},$$

where $0 < k < 1$ and $p > 0$; k depends on x, n, and the structure of the remainder term. The remainders in the form of Lagrange and Cauchy can be obtained as special cases of the Schlömilch formula with $p = n + 1$ and $p = 1$, respectively.

For $a = 0$, the Taylor's formula (M6.2.4.1) turns into

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + R_n(x)$$

and is called the *Maclaurin formula*.

The Maclaurin formula for some functions:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + R_n(x),$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^n\frac{x^{2n+1}}{(2n+1)!} + R_{2n+1}(x),$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^n\frac{x^{2n}}{(2n)!} + R_{2n}(x).$$

M6.2.5. Extremal Points. Points of Inflection

► **Maximum and minimum. Points of extremum.** Let $f(x)$ be a differentiable function on the interval (a, b) and $f'(x) > 0$ (resp., $f'(x) < 0$) on (a, b). Then $f(x)$ is an *increasing* (resp., *decreasing*) function on that interval*.

Suppose that there is a neighborhood of a point x_0 such that for all $x \neq x_0$ in that neighborhood we have $f(x) > f(x_0)$ (resp., $f(x) < f(x_0)$). Then x_0 is called a point of *local minimum* (resp., *local maximum*) of the function $f(x)$.

Points of local minimum or maximum are called points of *extremum*.

---

* At some isolated points of the interval, the derivative may vanish.

► **Necessary and sufficient conditions for the existence of extremum.** Suppose that $f(x)$ is continuous in some neighborhood $(x_0-\delta,\ x_0+\delta)$ of a point x_0 and differentiable at all points of the neighborhood except, possibly, x_0.

NECESSARY CONDITION OF EXTREMUM. *A function $f(x)$ can have an extremum only at points in which its derivative either vanishes or does not exist (or is infinite).*

FIRST SUFFICIENT CONDITION OF EXTREMUM. *If $f'(x)>0$ for $x\in(x_0-\delta,\ x_0)$ and $f'(x)<0$ for $x\in(x_0,\ x_0+\delta)$, then x_0 is a point of local maximum of this function. If $f'(x)<0$ for $x\in(x_0-\delta,\ x_0)$ and $f'(x)>0$ for $x\in(x_0,\ x_0+\delta)$, then x_0 is a point of local minimum of this function.*

If $f'(x)$ is of the same sign for all $x\neq x_0$, $x\in(x_0-\delta,\ x_0+\delta)$, then x_0 cannot be a point of extremum.

SECOND SUFFICIENT CONDITION OF EXTREMUM. *Let $f(x)$ be a twice differentiable function in a neighborhood of x_0. Then the following statements hold:*

$$\begin{aligned}&\text{(i)}\quad f'(x_0)=0\ \text{ and }\ f''(x_0)<0 \quad\Longrightarrow\quad f(x)\ \text{has a local maximum at the point}\ x_0;\\&\text{(ii)}\quad f'(x_0)=0\ \text{ and }\ f''(x_0)>0 \quad\Longrightarrow\quad f(x)\ \text{has a local minimum at the point}\ x_0.\end{aligned}$$

THIRD SUFFICIENT CONDITION OF EXTREMUM. *Let $f(x)$ be a function that is n times differentiable in a neighborhood of a point x_0 and $f'(x_0)=f''(x_0)=\cdots=f^{(n-1)}(x_0)=0$, but $f^{(n)}(x_0)\neq 0$. Then the following statements hold:*

$$\begin{aligned}&\textit{(i)}\quad n\ \text{is even and}\ f^{(n)}(x_0)<0 \quad\Longrightarrow\quad f(x)\ \text{has a local maximum at the point}\ x_0;\\&\textit{(ii)}\quad n\ \text{is even and}\ f^{(n)}(x_0)>0 \quad\Longrightarrow\quad f(x)\ \text{has a local minimum at the point}\ x_0.\end{aligned}$$

If n is odd, then x_0 cannot be a point of extremum.

► **Largest and the smallest values of a function.** Let $y=f(x)$ be continuous on the segment $[a,b]$ and differentiable at all points of this segment except, possibly, finitely many points. Then the largest and the smallest values of $f(x)$ on $[a,b]$ belong to the set consisting of $f(a)$, $f(b)$, and the values $f(x_i)$, where $x_i\in(a,b)$ are the points at which $f'(x)$ is either equal to zero or does not exist (is infinite).

► **Direction of the convexity of the graph of a function.** The graph of a differentiable function $y=f(x)$ is said to be *convex upward* (resp., *convex downward*) on the interval (a,b) if for each point of this interval, the graph lies below (resp., above) the tangent line at that point.

If the function $y=f(x)$ is twice differentiable on the interval (a,b) and $f''(x)<0$ (resp., $f''(x)>0$), then its graph is convex upward (resp., downward) on that interval. (At some isolated points of the interval, the second derivative may vanish.)

Thus, in order to find the intervals on which the graph of a twice differentiable function $f(x)$ is convex upward (resp., downward), one should solve the inequality $f''(x)<0$ (resp., $f''(x)>0$).

► **Inflection points.** An *inflection point* on the graph of a function $y=f(x)$ is defined as a point $(x_0,\ f(x_0))$ at which the graph passes from one side of its tangent line to another. At an inflection point, the graph changes the direction of its convexity.

Suppose that the function $y=f(x)$ has a continuous second derivative $f''(x)$ in some neighborhood of a point x_0. If $f''(x_0)=0$ and $f''(x)$ changes sign as x passes through the point x_0, then $(x_0,\ f(x_0))$ is an inflection point.

M6.2.6. Qualitative Analysis of Functions and Construction of Graphs

► **General scheme of analysis of a function and construction of its graph.**

1. Determine the domain in which the function is defined.
2. Determine whether the function is odd or even and whether it is periodic.
3. Find the points at which the graph crosses the coordinate axes.
4. Find the asymptotes of the graph.
5. Find extremal points and intervals of monotonicity.
6. Determine the directions of convexity of the graph and its inflection points.
7. Draw the graph, using the properties 1 to 6.

Example. Let us examine the function $y = \dfrac{\ln x}{x}$ and construct its graph.

We use the above general scheme.

1. This function is defined for all x such that $0 < x < +\infty$.

2. This function is neither odd nor even, since it is defined only for $x > 0$ and the relations $f(-x) = f(x)$ or $f(-x) = -f(x)$ cannot hold. Obviously, this function is nonperiodic.

3. The graph of this function does not cross the y-axis, since for $x = 0$ the function is undefined. Further, $y = 0$ only if $x = 1$, i.e., the graph crosses the x-axis only at the point $(1, 0)$.

4. The straight line $x = 0$ is a vertical asymptote, since $\lim\limits_{x\to+0} \dfrac{\ln x}{x} = -\infty$. We find the oblique asymptotes:

$$k = \lim_{x\to+\infty} \frac{y}{x} = 0, \qquad b = \lim_{x\to+\infty} (y - kx) = 0.$$

Therefore, the line $y = 0$ is a horizontal asymptote of the graph.

5. The derivative $y' = \dfrac{1 - \ln x}{x^2}$ vanishes for $\ln x = 1$. Therefore, the function may have an extremum at $x = e$. For $x \in (0, e)$, we have $y' > 0$, i.e., the function is increasing on this interval. For $x \in (e, +\infty)$, we have $y' < 0$, and therefore the function is decreasing on this interval. At $x = e$ the function attains its maximal value $y_{\max} = \dfrac{1}{e}$.

One should also examine the points at which the derivative does not exist. There is only one such point, $x = 0$, and it corresponds to the vertical asymptote (see Item 4).

6. The second derivative $y'' = \dfrac{2\ln x - 3}{x^3}$ vanishes for $x = e^{3/2}$. On the interval $(0, e^{3/2})$, we have $y'' < 0$, and therefore the graph is convex upward on this interval. For $x \in (e^{3/2}, +\infty)$, we have $y'' > 0$, and therefore the graph is convex downward on this interval. The value $x = e^{3/2}$ corresponds to an inflection point of the graph, with the ordinate $y = \frac{3}{2}e^{-3/2}$.

7. Using the above results, we construct the graph (Fig. M6.5).

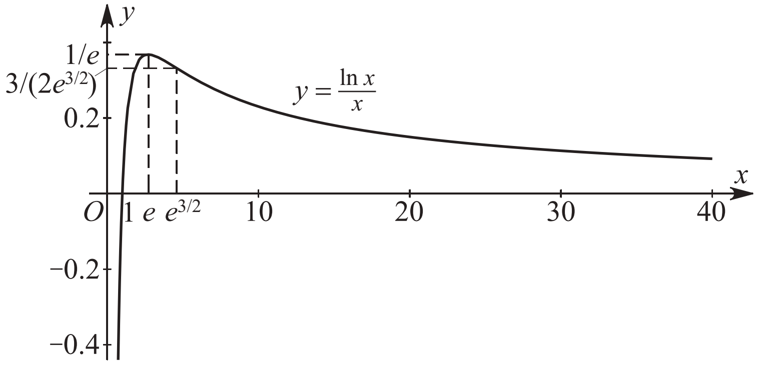

Figure M6.5. Graph of the function $y = \frac{\ln x}{x}$.

► **Transformations of graphs of functions.** Let us describe some methods which in many cases allow us to construct the graph of a function if we have the graph of a simpler function.

1. The graph of the function $y = f(x) + a$ is obtained from that of $y = f(x)$ by shifting the latter along the axis Oy by the distance $|a|$. For $a > 0$ the shift is upward, and for $a < 0$ downward (see Fig. M6.6 a).

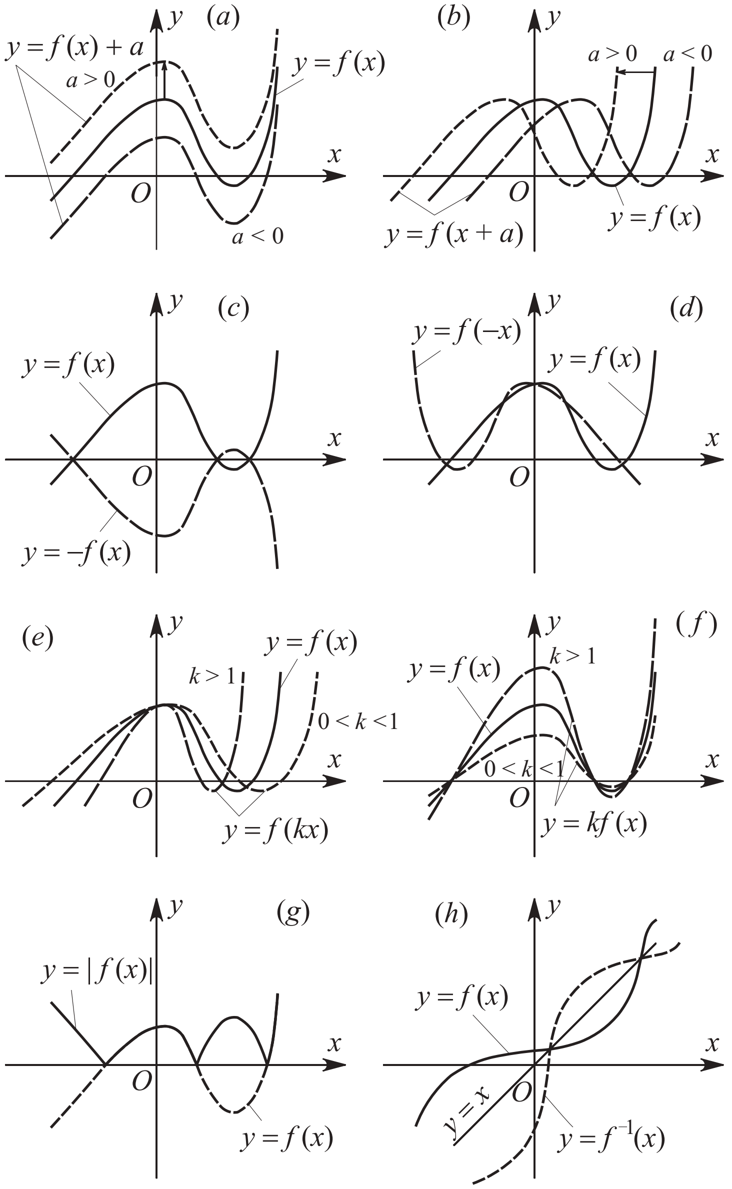
(a)
y
y = f (x) + a
a > 0
y = f (x)
x
O
a < 0
(b)
y
a > 0
a < 0
x
O
y = f (x + a)
y = f (x)
(c)
y
y = f (x)
x
O
y = −f (x)
(d)
y
y = f (−x)
y = f (x)
x
O
(e)
y
y = f (x)
k > 1
0 < k < 1
x
O
y = f (kx)
(f)
y
k > 1
y = f (x)
0 < k < 1
x
O
y = kf (x)
(g)
y
y = | f (x)|
x
O
y = f (x)
(h)
y
y = f (x)
x
y = x
O
y = f^{−1}(x)

Figure M6.6. Transformations of graphs of functions.

2. The graph of the function $y = f(x + a)$ is obtained from that of $y = f(x)$ by shifting the latter along the Ox by the distance $|a|$. For $a > 0$ the shift is to the left, and for $a < 0$ to the right (see Fig. M6.6 b).

3. The graph of the function $y = -f(x)$ is obtained from that of $y = f(x)$ by symmetric reflection with respect to the axis Ox (see Fig. M6.6 *c*).

4. The graph of the function $y = f(-x)$ is obtained from that of $y = f(x)$ by symmetric reflection with respect to the axis Oy (see Fig. M6.6 *d*).

5. The graph of the function $y = f(kx)$ for $k > 1$ is obtained from that of $y = f(x)$ by contracting the latter k times to the axis Oy, and for $0 < k < 1$ by extending the latter $1/k$ times from the axis Oy. The points at which the graph crosses the axis Oy remain unchanged (see Fig. M6.6 *e*).

6. The graph of the function $y = kf(x)$ for $k > 1$ is obtained from that of $y = f(x)$ by extending the latter k times from the axis Ox, and for $0 < k < 1$ by contracting the latter $1/k$ times to the axis Ox. The points at which the graph crosses the axis Ox remain unchanged (see Fig. M6.6*f*).

7. The graph of the function $y = |f(x)|$ is obtained from that of $y = f(x)$ by preserving the parts of the latter for which $f(x) \geq 0$ and symmetric reflection, with respect to the axis Ox, of the parts for which $f(x) < 0$ (see Fig. M6.6 *g*).

8. The graph of the inverse function $y = f^{-1}(x)$ is obtained from that of $y = f(x)$ by symmetric reflection with respect to the straight line $y = x$ (see Fig. M6.6 *h*).

M6.2.7. Approximate Solution of Equations (Root-Finding Algorithms for Continuous Functions)

► **Preliminaries.** For a vast majority of algebraic (transcendental) equations of the form

$$f(x) = 0, \qquad \text{(M6.2.7.1)}$$

where $f(x)$ is a continuous function, there are no exact formulas for the roots.

When solving the equation approximately, the first step is to bracket the roots, i.e., find sufficiently small intervals containing exactly one root each. Such an interval $[a, b]$, where the numbers a and b satisfy the condition $f(a)f(b) < 0$ (which is assumed to hold in what follows), can be found, say, graphically.

The second step is to compute successive approximations $x_n \in [a, b]$ ($n = 1, 2, \dots$) to the desired root $c = \lim\limits_{n\to\infty} x_n$, usually by one of the following methods.

► **Bisection method.** To find the root of equation (M6.2.7.1) on the interval $[a, b]$, we bisect the interval. If $f\left(\frac{a+b}{2}\right) = 0$, then $c = \frac{a+b}{2}$ is the desired root. If $f\left(\frac{a+b}{2}\right) \neq 0$, then of the two intervals $\left[a, \frac{a+b}{2}\right]$ and $\left[\frac{a+b}{2}, b\right]$ we take the one at whose endpoints the function $f(x)$ has opposite signs. Now we bisect the new, smaller interval, etc. As a result, we obtain either an exact root of equation (M6.2.7.1) at some step or an infinite sequence of nested intervals $[a_1, b_1]$, $[a_2, b_2]$, … such that $f(a_n)f(b_n) < 0$. The root is given by the formula $c = \lim\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} b_n$, and the estimate

$$0 \leq c - a_n \leq \frac{1}{2^n}(b - a)$$

is valid.

The following two methods are more efficient.

► **Regula falsi method (false position method).** Suppose that the derivatives $f'(x)$ and $f''(x)$ exist on the interval $[a, b]$ and the inequalities $f'(x) \neq 0$ and $f''(x) \neq 0$ hold for all $x \in [a, b]$.

If $f'(a)f''(a) > 0$, then we take $x_0 = a$ for the zero approximation; the subsequent approximations are given by the formulas

$$x_{n+1} = x_n - \frac{f(x_n)}{f(b) - f(x_n)}(b - x_n), \qquad n = 0,\ 1,\ \ldots$$

If $f'(a)f''(a) < 0$, then we take $x_0 = b$ for the zero approximation; the subsequent approximations are given by the formulas

$$x_{n+1} = x_n - \frac{f(x_n)}{f(a) - f(x_n)}(a - x_n), \qquad n = 0,\ 1,\ \ldots$$

The regula falsi method has the first order of local convergence as $n \to \infty$, that is,

$$|x_{n+1} - c| \le k|x_n - c|,$$

where k is a constant depending on $f(x)$ and c is the root of equation (M6.2.7.1).

The regula falsi method has a simple geometric interpretation. The straight line (secant) passing through the points $(a, f(a))$ and $(b, f(b))$ of the curve $y = f(x)$ meets the abscissa axis at the point x_1; the value x_{n+1} is the abscissa of the point where the line passing through the points $(x_0, f(x_0))$ and $(x_n, f(x_n))$ meets the x-axis (see Fig. M6.7 *a*).

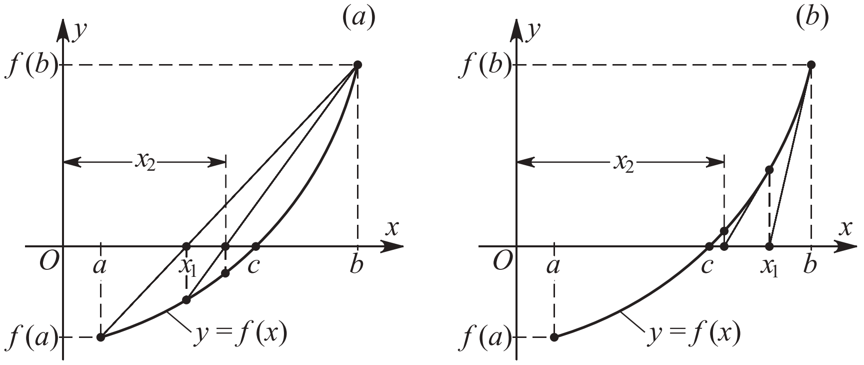

Figure M6.7. Graphical construction of successive approximations to the root of equation (M6.2.7.1) by the regula falsi method (*a*) and the Newton–Raphson method (*b*).

► **Newton–Raphson method.** Suppose that the derivatives $f'(x)$ and $f''(x)$ exist on the interval $[a, b]$ and the inequalities $f'(x) \neq 0$ and $f''(x) \neq 0$ hold for all $x \in [a, b]$.

If $f(a)f''(a) > 0$, then we take $x_0 = a$ for the zero approximation; if $f(b)f''(b) > 0$, then $x_0 = b$. The subsequent approximations are computed by the formulas

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \qquad n = 0,\ 1,\ \ldots$$

If the initial approximation x_0 is sufficiently close to the desired root c, then the Newton–Raphson method exhibits quadratic convergence:

$$|x_{n+1} - c| \le \frac{M}{2m}|x_n - c|^2,$$

where $M = \max\limits_{a \le x \le b} |f''(x)|$ and $m = \min\limits_{a \le x \le b} |f'(x)|$.

The Newton–Raphson method has a simple geometric interpretation. The tangent to the curve $y = f(x)$ through the point $(x_n, f(x_n))$ meets the abscissa axis at the point x_{n+1} (see Fig. M6.7 *b*).

The Newton–Raphson method has a higher order of convergence than the regula falsi method. Hence the former is more often used in practice.

M6.3. Functions of Several Variables. Partial Derivatives

M6.3.1. Point Sets. Functions. Limits and Continuity

► **Sets on the plane and in space.** The distance between two points A and B on the plane and in space can be defined as follows:

$$\rho(A,B)=\sqrt{(x_A-x_B)^2+(y_A-y_B)^2} \qquad \text{(on the plane)},$$

$$\rho(A,B)=\sqrt{(x_A-x_B)^2+(y_A-y_B)^2+(z_A-z_B)^2} \qquad \text{(in three-dimensional space)},$$

$$\rho(A,B)=\sqrt{(x_{1A}-x_{1B})^2+\cdots+(x_{nA}-x_{nB})^2} \qquad \text{(in n-dimensional space)},$$

where x_A, y_A and x_B, y_B, and x_A, y_A, z_A and x_B, y_B, z_B, and x_{1A}, $\ldots$, x_{nA} and x_{1B}, $\ldots$, x_{nB} are Cartesian coordinates of the respective points.

An *ε-neighborhood of a point* M_0 (on the plane or in space) is the set consisting of all points M (resp., on the plane or in space) such that $\rho(M,M_0)<\varepsilon$, where it is assumed that $\varepsilon>0$. An *ε-neighborhood of a set* K (on the plane or in space) is the set consisting of all points M (resp., on the plane or in space) such that $\inf\limits_{M_0\in K}\rho(M,M_0)<\varepsilon$, where it is assumed that $\varepsilon>0$.

An *interior point* of a set D is a point belonging to D, together with some neighborhood of that point. An *open set* is a set containing only interior points. A *boundary point* of a set D is a point such that any of its neighborhoods contains points both inside and outside D. A *closed set* is a set containing all its boundary points. A set D is called a *bounded set* if $\rho(A,B)<C$ for any points $A,B\in D$, where C is a constant independent of A,B. Otherwise (i.e., if there is no such constant), the set D is called *unbounded*.

► **Functions of two or three variables.** A (numerical) *function* on a set D is, by definition, a relation that sets up a correspondence between each point $M\in D$ and a unique numerical value. If D is a plane set, then each point $M\in D$ is determined by two coordinates x, y, and a function $z=f(M)=f(x,y)$ is called a *function of two variables*. If D belongs to a three-dimensional space, then one speaks of a *function of three variables*. The set D on which the function is defined is called the *domain* of the function. For instance, the function $z=\sqrt{1-x^2-y^2}$ is defined on the closed circle $x^2+y^2\le 1$, which is its domain.

The *graph* of a function $z=f(x,y)$ is the surface formed by the points $(x,y,f(x,y))$ in three-dimensional space. For instance, the graph of the function $z=ax+by+c$ is a plane, and the graph of the function $z=\sqrt{1-x^2-y^2}$ is a semisphere.

A *level line* of a function $z=f(x,y)$ is a line on the plane x, y with the following property: the function takes one and the same value $z=c$ at all points of that line. Thus, the equation of a level line has the form $f(x,y)=c$. A *level surface* of a function $u=f(x,y,z)$ is a surface on which the function takes a constant value, $u=c$; the equation of a level surface has the form $f(x,y,z)=c$.

A function $f(M)$ is called *bounded* on a set D if there is a constant C such that $|f(M)|\le C$ for all $M\in D$.

► **Limit of a function at a point and its continuity.** Let M be a point that comes infinitely close to some point M_0, i.e., $\rho=\rho(M_0,M)\to 0$. It is possible that the values $f(M)$ come close to some constant b.

One says that b is the *limit of the function* $f(M)$ at the point M_0 if for any (arbitrarily small) $\varepsilon>0$, there is $\delta>0$ such that for all points M belonging to the domain of the function and satisfying the inequality $0<\rho(M_0,M)<\delta$, we have $|f(M)-b|<\varepsilon$. In this case, one writes $\lim\limits_{\rho(M,M_0)\to 0}f(M)=b$.

A function $f(M)$ is called *continuous* at a point M_0 if $\lim\limits_{\rho(M,M_0)\to 0} f(M) = f(M_0)$. A function is called *continuous on a set* D if it is continuous at each point of D. Any continuous function $f(M)$ on a closed bounded set is bounded on that set and attains its smallest and its largest values on that set.

M6.3.2. Differentiation of Functions of Several Variables

For the sake of brevity, we consider the case of a function of two variables. However, all statements can be easily extended to the case of n variables.

► **Total and partial increments of a function. Partial derivatives.** A *total increment* of a function $z = f(x, y)$ at a point (x, y) is

$$\Delta z = f(x + \Delta x, y + \Delta y) - f(x, y),$$

where Δx, Δy are increments of the independent variables. *Partial increments* in x and in y are, respectively,

$$\Delta_x z = f(x + \Delta x, y) - f(x, y),$$
$$\Delta_y z = f(x, y + \Delta y) - f(x, y).$$

Partial derivatives of a function z with respect to x and to y at a point (x, y) are defined as follows:

$$\frac{\partial z}{\partial x} = \lim_{\Delta x\to 0} \frac{\Delta_x z}{\Delta x}, \qquad \frac{\partial z}{\partial y} = \lim_{\Delta y\to 0} \frac{\Delta_y z}{\Delta y}$$

(provided that these limits exist). Partial derivatives are also denoted by z_x and z_y, $\partial_x z$ and $\partial_y z$, or $f_x(x, y)$ and $f_y(x, y)$.

► **Differentiable functions. Differential.** A function $z = f(x, y)$ is called *differentiable* at a point (x, y) if its increment at that point can be represented in the form

$$\Delta z = A(x, y)\Delta x + B(x, y)\Delta y + o(\rho), \qquad \rho = \sqrt{(\Delta x)^2 + (\Delta y)^2},$$

where $o(\rho)$ is a quantity of a higher order of smallness compared with ρ as $\rho \to 0$ (i.e., $o(\rho)/\rho \to 0$ as $\rho \to 0$). In this case, there exist partial derivatives at the point (x, y), and $z_x = A(x, y)$, $z_y = B(x, y)$.

A function that has continuous partial derivatives at a point (x, y) is differentiable at that point.

The *differential* dz *of a function* $z = f(x, y)$ is defined as follows:

$$dz = f_x(x, y)\Delta x + f_y(x, y)\Delta y.$$

Taking the differentials dx and dy of the independent variables equal to Δx and Δy, respectively, one can also write $dz = f_x(x, y)\,dx + f_y(x, y)\,dy$.

The relation $\Delta z = dz + o(\rho)$ for small Δx and Δy is widely used for approximate calculations, in particular, for finding errors in numerical calculations of values of a function.

Example 1. Suppose that the values of the arguments of the function $z = x^2y^5$ are known with the error $x = 2 \pm 0.01$, $y = 1 \pm 0.01$. Let us calculate the approximate value of the function.

We find the increment of the function z at the point $x = 2$, $y = 1$ for $\Delta x = \Delta y = 0.01$, using the formula $\Delta z \approx dz = 2 \cdot 2 \cdot 1^5 \cdot 0.01 + 5 \cdot 2^2 \cdot 1^4 \cdot 0.01 = 0.24$. Therefore, we can accept the approximation $z = 4 \pm 0.24$.

If a function $z = f(x, y)$ is differentiable at a point (x_0, y_0), then

$$f(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) + o(\rho).$$

Hence, for small ρ (i.e., for $x \approx x_0$, $y \approx y_0$), we obtain the approximate formula

$$f(x, y) \approx f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

The replacement of a function by this linear expression near a given point is called *linearization*.

► **Composite function.** Consider a function $z = f(x, y)$ and let $x = x(u, v)$, $y = y(u, v)$. Suppose that for $(u, v) \in D$, the functions $x(u, v)$, $y(u, v)$ take values for which the function $z = f(x, y)$ is defined. In this way, one defines a *composite function* on the set D, namely, $z(u, v) = f\bigl(x(u, v), y(u, v)\bigr)$. In this situation, $f(x, y)$ is called the outer function and $x(u, v)$, $y(u, v)$ are called the inner functions.

Partial derivatives of a composite function are expressed by

$$\frac{\partial z}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u},$$
$$\frac{\partial z}{\partial v} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v}.$$

For $z = z(t, x, y)$, let $x = x(t)$, $y = y(t)$. Thus, z is actually a function of only one variable t. The derivative $\frac{dz}{dt}$ is calculated by

$$\frac{dz}{dt} = \frac{\partial z}{\partial t} + \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.$$

This derivative, in contrast to the partial derivative $\frac{\partial z}{\partial t}$, is called a *total derivative*.

► **Second partial derivatives and second differentials.** The *second partial derivatives* of a function $z = f(x, y)$ are defined as the derivatives of its first partial derivatives and are denoted as follows:

$$\frac{\partial^2 z}{\partial x^2} = z_{xx} \equiv (z_x)_x, \qquad \frac{\partial^2 z}{\partial x\,\partial y} = z_{xy} \equiv (z_x)_y,$$
$$\frac{\partial^2 z}{\partial y\,\partial x} = z_{yx} \equiv (z_y)_x, \qquad \frac{\partial^2 z}{\partial y^2} = z_{yy} \equiv (z_y)_y.$$

The derivatives z_{xy} and z_{yx} are called *mixed derivatives*. If the mixed derivatives are continuous at some point, then they coincide at that point, $z_{xy} = z_{yx}$.

In a similar way, one defines higher-order partial derivatives.

The *second differential* of a function $z = f(x, y)$ is the expression

$$d^2 z = d(dz) = (dz)_x \Delta x + (dz)_y \Delta y = z_{xx}(\Delta x)^2 + 2z_{xy}\Delta x \Delta y + z_{yy}(\Delta y)^2.$$

In a similar way, one defines $d^3 z$, $d^4 z$, etc.

► **Implicit functions and their differentiation.** Consider the equation $F(x, y) = 0$ with a solution (x_0, y_0). Suppose that the derivative $F_y(x, y)$ is continuous in a neighborhood of the point (x_0, y_0) and $F_y(x, y) \neq 0$ in that neighborhood. Then the equation $F(x, y) = 0$ defines a continuous function $y = y(x)$ (called an *implicit function*) of the variable x in a neighborhood of the point x_0. Moreover, if in a neighborhood of (x_0, y_0) there exists a continuous derivative F_x, then the implicit function $y = y(x)$ has a continuous derivative expressed by $\dfrac{dy}{dx} = -\dfrac{F_x}{F_y}$.

Consider the equation $F(x, y, z) = 0$ that establishes a relation between the variables x, y, z. If $F(x_0, y_0, z_0) = 0$ and in a neighborhood of the point (x_0, y_0, z_0) there exist continuous partial derivatives F_x, F_y, F_z such that $F_z(x_0, y_0, z_0) \neq 0$, then equation $F(x, y, z) = 0$, in a neighborhood of (x_0, y_0), has a unique solution $z = \varphi(x, y)$ such that $\varphi(x_0, y_0) = z_0$;

moreover, the function $z = \varphi(x, y)$ is continuous and has continuous partial derivatives expressed by

$$\frac{\partial z}{\partial x} = -\frac{F_x}{F_z}, \qquad \frac{\partial z}{\partial y} = -\frac{F_y}{F_z}.$$

Example 2. For the equation $x \sin y + z + e^z = 0$ we have $F_z = 1 + e^z \neq 0$. Therefore, this equation defines an implicit function $z = \varphi(x, y)$ on the entire plane, and its derivatives have the form $\dfrac{\partial z}{\partial x} = -\dfrac{\sin y}{1 + e^z}$, $\dfrac{\partial z}{\partial y} = -\dfrac{x \cos y}{1 + e^z}$.

► Jacobian. Dependent and independent functions. Invertible transformations.

1°. Two functions $f(x, y)$ and $g(x, y)$ are called *dependent* if there is a function $\Phi(z)$ such that $g(x, y) = \Phi(f(x, y))$; otherwise, the functions $f(x, y)$ and $g(x, y)$ are called *independent*.

The *Jacobian* is the determinant of the matrix whose elements are the first partial derivatives of the functions $f(x, y)$ and $g(x, y)$:

$$\frac{\partial(f, g)}{\partial(x, y)} \equiv \begin{vmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{vmatrix}. \tag{M6.3.2.1}$$

1) If the Jacobian (M6.3.2.1) in a domain D is identically equal to zero, then the functions $f(x, y)$ and $g(x, y)$ are dependent in D.

2) If the Jacobian (M6.3.2.1) is nonzero in D, then the functions $f(x, y)$ and $g(x, y)$ are independent in D.

2°. Functions $f_k(x_1, x_2, \dots, x_n)$, $k = 1, 2, \dots, n$, are called dependent in a domain D if there is a function $\Phi(z_1, z_2, \dots, z_n)$ such that

$$\Phi\bigl(f_1(x_1, x_2, \dots, x_n), f_2(x_1, x_2, \dots, x_n), \dots, f_n(x_1, x_2, \dots, x_n)\bigr) = 0 \qquad \text{(in } D\text{);}$$

otherwise, these functions are called independent.

The *Jacobian* is the determinant of the matrix whose elements are the first partial derivatives:

$$\frac{\partial(f_1, f_2, \dots, f_n)}{\partial(x_1, x_2, \dots, x_n)} \equiv \det\left(\frac{\partial f_i}{\partial x_j}\right). \tag{M6.3.2.2}$$

The functions $f_k(x_1, x_2, \dots, x_n)$ are dependent in a domain D if the Jacobian (M6.3.2.2) is identically equal to zero in D. The functions $f_k(x_1, x_2, \dots, x_n)$ are independent in D if the Jacobian (M6.3.2.2) does not vanish in D.

3°. Consider the transformation

$$y_k = f_k(x_1, x_2, \dots, x_n), \qquad k = 1, 2, \dots, n. \tag{M6.3.2.3}$$

Suppose that the functions f_k are continuously differentiable and the Jacobian (M6.3.2.2) differs from zero at a point $(x_1^\circ, x_2^\circ, \dots, x_n^\circ)$. Then, in a sufficiently small neighborhood of this point, equations (M6.3.2.3) specify a one-to-one correspondence between the points of that neighborhood and the set of points $(y_1, y_2, \dots, y_n)$ consisting of the values of the functions (M6.3.2.3) in the corresponding neighborhood of the point $(y_1^\circ, y_2^\circ, \dots, y_n^\circ)$. This means that the system (M6.3.2.3) is locally solvable in a neighborhood of the point $(x_1^\circ, x_2^\circ, \dots, x_n^\circ)$, i.e., the following representation holds:

$$x_k = g_k(y_1, y_2, \dots, y_n), \qquad k = 1, 2, \dots, n,$$

where g_k are continuously differentiable functions in the corresponding neighborhood of the point $(y_1^\circ, y_2^\circ, \dots, y_n^\circ)$.

M6.3.3. Directional Derivative. Gradient. Geometrical Applications

► **Directional derivative.** One says that a *scalar field* is defined in a domain D if any point $M(x,y)$ of that domain is associated with a certain value $z = f(M) = f(x,y)$. Thus, a thermal field and a pressure field are examples of scalar fields. A *level line* of a scalar field is a level line of the function that specifies the field (see Subsection M6.3.1). Thus, isothermal and isobaric curves are, respectively, level lines of thermal and pressure fields.

In order to examine the behavior of a field $z = f(x,y)$ at a point $M_0(x_0,y_0)$ in the direction of a vector $\mathbf{a} = \{a_1, a_2\}$, one should construct a straight line passing through M_0 in the direction of the vector **a** (this line can be specified by the parametric equations $x = x_0 + a_1 t$, $y = y_0 + a_2 t$) and study the function $z(t) = f(x_0 + a_1 t,\ y_0 + a_2 t)$. The derivative of the function $z(t)$ at the point M_0 (i.e., for $t = 0$) characterizes the change rate of the field at that point in the direction **a**. Dividing $z'(0)$ by $|\mathbf{a}| = \sqrt{a_1^2 + a_2^2}$, we obtain the so-called *derivative in the direction* **a** of the given field at the given point:

$$\frac{\partial f}{\partial \mathbf{a}} = \frac{1}{|\mathbf{a}|}\bigl[\,a_1 f_x(x_0,y_0) + a_2 f_y(x_0,y_0)\bigr].$$

The *gradient* of the scalar field $z = f(x,y)$ is, by definition, the vector-valued function

$$\operatorname{grad} f = f_x(x,y)\mathbf{i} + f_y(x,y)\mathbf{j},$$

where **i** and **j** are unit vectors along the coordinate axes x and y. At each point, the gradient of a scalar field is orthogonal to the level line passing through that point. The gradient indicates the direction of maximal growth of the field. In terms of the gradient, the directional derivative can be expressed as follows:

$$\frac{\partial f}{\partial \mathbf{a}} = \frac{\mathbf{a}}{|\mathbf{a}|}\operatorname{grad} f.$$

The gradient is also denoted by $\nabla f = \operatorname{grad} f$.

Remark. The above facts for a plane scalar field obviously can be extended to the case of a spatial scalar field.

► **Geometrical applications of the theory of functions of several variables.**

1. The *equation of the tangent plane* to the surface $z = f(x,y)$ at a point (x_0, y_0, z_0), where $z_0 = f(x_0,y_0)$, has the form

$$z = f(x_0,y_0) + f_x(x_0,y_0)(x - x_0) + f_y(x_0,y_0)(y - y_0).$$

The vector of the normal to the surface at that point is

$$\mathbf{n} = \bigl\{-f_x(x_0,y_0),\ -f_y(x_0,y_0),\ 1\bigr\}.$$

2. If a surface is defined by the equation $\Phi(x,y,z) = 0$, then the equation of its tangent plane at the point (x_0,y_0,z_0) has the form

$$\Phi_x(x_0,y_0,z_0)(x - x_0) + \Phi_y(x_0,y_0,z_0)(y - y_0) + \Phi_z(x_0,y_0,z_0)(z - z_0) = 0.$$

A normal vector to the surface at this point is

$$\mathbf{n} = \bigl\{\Phi_x(x_0,y_0,z_0),\ \Phi_y(x_0,y_0,z_0),\ \Phi_z(x_0,y_0,z_0)\bigr\}.$$

3. Consider a surface defined by the parametric equations

$$x = x(u,v), \quad y = y(u,v), \quad z = z(u,v)$$

or, in vector form, $\mathbf{r} = \mathbf{r}(u,v)$, where $\mathbf{r} = \{x,y,z\}$, and let $M_0\bigl(x(u_0,v_0),\ y(u_0,v_0),\ z(u_0,v_0)\bigr)$ be the point of the surface corresponding to the parameter values $u = u_0$, $v = v_0$. Then the vector of the normal to the surface at the point M_0 can be expressed by

$$\mathbf{n}(u,v) = \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{vmatrix},$$

where all partial derivatives are calculated at the point M_0.

M6.3.4. Extremal Points of Functions of Several Variables

► **Conditions of extremum of a function of two variables.**

1°. *Points of minimum, maximum, or extremum.* A point (x_0, y_0) is called a *point of local minimum* (resp., *maximum*) of a function $z = f(x, y)$ if there is a neighborhood of (x_0, y_0) in which the function is defined and satisfies the inequality $f(x, y) > f(x_0, y_0)$ (resp., $f(x, y) < f(x_0, y_0)$). Points of maximum or minimum are called *points of extremum.*

2°. A *necessary condition of extremum.* If a function has the first partial derivatives at a point of its extremum, these derivatives must vanish at that point. It follows that in order to find points of extremum of such a function $z = f(x, y)$, one should find solutions of the system of equations

$$f_x(x, y) = 0, \quad f_y(x, y) = 0.$$

The points whose coordinates satisfy this system are called *stationary points.* Any point of extremum of a differentiable function is its stationary point, but not every stationary point is a point of its extremum.

3°. *Sufficient conditions of extremum* are used for the identification of points of extremum among stationary points. Some conditions of this type are given below.

Suppose that the function $z = f(x, y)$ has continuous second derivatives at a stationary point. Let us calculate the following quantity at this point:

$$\Delta = f_{xx} f_{yy} - f_{xy}^2.$$

The following statements hold:

1) If $\Delta > 0$, $f_{xx} > 0$, then the stationary point is a point of local minimum;
2) If $\Delta > 0$, $f_{xx} < 0$, then the stationary point is a point of local maximum;
3) If $\Delta < 0$, then the stationary point cannot be a point of extremum.

In the degenerate case, $\Delta = 0$, a more delicate analysis of a stationary point is required. In this case, a stationary point may happen to be a point of extremum and may not.

Remark. In order to find points of extremum, one should check not only stationary points, but also points at which the first derivatives do not exist or are infinite.

4°. *The smallest and the largest values of a function.* Let $f(x, y)$ be a continuous function in a closed bounded domain D. Any such function takes its smallest and its largest values in D.

If the function has partial derivatives in D, except at some points, then the following method can be helpful for determining the coordinates of the points $(x_{\min}, y_{\min})$ and $(x_{\max}, y_{\max})$ at which the function attains its minimum and maximum, respectively. One should find all internal stationary points and all points at which the derivatives are infinite or do not exist. Then one should calculate the values of the function at these points and compare these with its values at the boundary points of the domain, and then choose the largest and the smallest values.

► **Extremal points of functions of three variables.** For functions of three variables, points of extremum are defined in exactly the same way as for functions of two variables. Let us briefly describe the scheme of finding extremal points of a function $u = \Phi(x, y, z)$. Finding solutions of the system of equations

$$\Phi_x(x, y, z) = 0, \quad \Phi_y(x, y, z) = 0, \quad \Phi_z(x, y, z) = 0,$$

we determine stationary points. For each stationary point, we calculate the values of

$$\Delta_1 = \Phi_{xx}, \quad \Delta_2 = \begin{vmatrix} \Phi_{xx} & \Phi_{xy} \\ \Phi_{xy} & \Phi_{yy} \end{vmatrix}, \quad \Delta_3 = \begin{vmatrix} \Phi_{xx} & \Phi_{xy} & \Phi_{xz} \\ \Phi_{xy} & \Phi_{yy} & \Phi_{yz} \\ \Phi_{xz} & \Phi_{yz} & \Phi_{zz} \end{vmatrix}.$$

The following statements hold:

1) If $\Delta_1 > 0$, $\Delta_2 > 0$, $\Delta_3 > 0$, then the stationary point is a point of local minimum;
2) If $\Delta_1 < 0$, $\Delta_2 > 0$, $\Delta_3 < 0$, then the stationary point is a point of local maximum.

► **Conditional extremum of a function of two variables. Lagrange function.** A point (x_0, y_0) is called a *point of conditional or constrained minimum* (resp., *maximum*) of a function

$$z = f(x, y) \qquad \text{(M6.3.4.1)}$$

under the additional condition*

$$\varphi(x, y) = 0 \qquad \text{(M6.3.4.2)}$$

if there is a neighborhood of the point (x_0, y_0) in which $f(x, y) > f(x_0, y_0)$ (resp., $f(x, y) < f(x_0, y_0)$) for all points (x, y) satisfying the condition (M6.3.4.2).

For the determination of points of conditional extremum, it is common to use the *Lagrange function*

$$\Phi(x, y, \lambda) = f(x, y) + \lambda\varphi(x, y),$$

where λ is the so-called *Lagrange multiplier*. Solving the system of three equations (the last equation coincides with the condition (M6.3.4.2))

$$\frac{\partial\Phi}{\partial x} = 0, \quad \frac{\partial\Phi}{\partial y} = 0, \quad \frac{\partial\Phi}{\partial\lambda} = 0,$$

one finds stationary points of the Lagrange function (and also the value of the multiplier λ). The stationary points may happen to be points of extremum. The above system yields only necessary conditions of extremum, but these conditions may be insufficient; it may happen that there is no extremum at some stationary points. However, with the help of other properties of the function under consideration, it is often possible to establish the character of a critical point.

Example 1. Let us find an extremum of the function

$$z = x^n y, \qquad \text{(M6.3.4.3)}$$

under the condition

$$x + y = a \qquad (a > 0,\ n > 0,\ x \ge 0,\ y \ge 0). \qquad \text{(M6.3.4.4)}$$

Taking $\varphi(x, y) = x + y - a$, we construct the Lagrange function

$$\Phi(x, y, \lambda) = x^n y + \lambda(x + y - a).$$

Solving the system of equations

$$\begin{aligned} \Phi_x &\equiv nx^{n-1}y + \lambda = 0, \\ \Phi_y &\equiv x^n + \lambda = 0, \\ \Phi_\lambda &\equiv x + y - a = 0, \end{aligned}$$

we find the coordinates of a unique stationary point,

$$x_\circ = \frac{an}{n+1}, \quad y_\circ = \frac{a}{n+1}, \quad \lambda_\circ = -\left(\frac{an}{n+1}\right)^n,$$

which corresponds to the conditional maximum of the given function, $z_{\max} = \dfrac{a^{n+1}n^n}{(n+1)^{n+1}}$.

---

* This condition is also called a *constraint*.

Remark. In order to find points of conditional extremum of functions of two variables, it is often convenient to express the variable y through x (or vice versa) from the additional equation (M6.3.4.2) and substitute the resulting expression into the right-hand side of (M6.3.4.1). In this way, the original problem is reduced to the problem of extremum for a function of a single variable.

Example 2. Consider again the extremum problem of Example 1 for the function of two variables (M6.3.4.3) with the constraint (M6.3.4.4). After the elimination of the variable y from (M6.3.4.3)–(M6.3.4.4), the original problem is reduced to the extremum problem for the function $z = x^n(a - x)$ of one variable.

► **Conditional extrema of functions of several variables.** Consider a function $u = f(x_1, \dots, x_n)$ of n variables under the condition that x_1, …, x_n satisfy m equations ($m < n$):

$$\begin{cases} \varphi_1(x_1, \dots, x_n) = 0, \\ \varphi_2(x_1, \dots, x_n) = 0, \\ \dots\dots\dots\dots\dots\dots, \\ \varphi_m(x_1, \dots, x_n) = 0. \end{cases}$$

In order to find the values of x_1, …, x_n for which f may have a conditional maximum or minimum, one should construct the Lagrange function

$$\Phi(x_1, \dots, x_n; \lambda_1, \dots, \lambda_m) = f + \lambda_1\varphi_1 + \lambda_2\varphi_2 + \dots + \lambda_m\varphi_m$$

and equate to zero its first partial derivatives with respect to the variables $x_1, \dots, x_n$ and the parameters $\lambda_1, \dots, \lambda_m$. From the resulting $n+m$ equations, one finds $x_1, \dots, x_n$ (and also the values of the unknown Lagrange multipliers $\lambda_1, \dots, \lambda_m$). As in the case of functions of two variables, the question whether the given function has points of conditional extremum can be answered on the basis of additional investigation.

Example 3. Consider the problem of finding the shortest distance from the point (x_0, y_0, z_0) to the plane

$$Ax + By + Cz + D = 0. \qquad \text{(M6.3.4.5)}$$

The squared distance between the points (x_0, y_0, z_0) and (x, y, z) is equal to

$$R^2 = (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2. \qquad \text{(M6.3.4.6)}$$

In our case, the coordinates (x, y, z) should satisfy equation (M6.3.4.5) (this point should belong to the plane). Thus, our problem is to find the minimum of the expression (M6.3.4.6) under the condition (M6.3.4.5). The Lagrange function has the form

$$\Phi = (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 + \lambda(Ax + By + Cz + D).$$

Equating to zero the derivatives of Φ with respect to x, y, z, and λ, we obtain the following system of algebraic equations:

$$2(x - x_0) + A\lambda = 0, \quad 2(y - y_0) + B\lambda = 0, \quad 2(z - z_0) + C\lambda = 0, \quad Ax + By + Cz + D = 0.$$

Its solution has the form

$$x = x_0 - \frac{1}{2}A\lambda, \quad y = y_0 - \frac{1}{2}B\lambda, \quad z = z_0 - \frac{1}{2}C\lambda, \quad \lambda = \frac{2(Ax_0 + By_0 + Cz_0 + D)}{A^2 + B^2 + C^2}. \qquad \text{(M6.3.4.7)}$$

Thus we have a unique answer, and since the distance between a given point and the plane can be realized at a single point (x, y, z), the values obtained should correspond to that distance. Substituting the values (M6.3.4.7) into (M6.3.4.6), we find the squared distance

$$R^2 = \frac{(Ax_0 + By_0 + Cz_0 + D)^2}{A^2 + B^2 + C^2}.$$

M6.3.5. Differential Operators of the Field Theory

► **Hamilton's operator and first-order differential operators.** Hamilton's operator, commonly known as the *nabla vector operator* or the *gradient operator*, is the symbolic vector

$$\nabla = \mathbf{i}\,\frac{\partial}{\partial x} + \mathbf{j}\,\frac{\partial}{\partial y} + \mathbf{k}\,\frac{\partial}{\partial z}.$$

This vector can be used for expressing the following differential operators:

1) gradient of a scalar function $u(x,y,z)$:

$$\operatorname{grad} u = \mathbf{i}\,\frac{\partial u}{\partial x} + \mathbf{j}\,\frac{\partial u}{\partial y} + \mathbf{k}\,\frac{\partial u}{\partial z} = \nabla u;$$

2) divergence of a vector field $\mathbf{a} = P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k}$:

$$\operatorname{div} \mathbf{a} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} = \nabla \cdot \mathbf{a}$$

(scalar product of the nabla vector and the vector **a**);

3) rotation of a vector field $\mathbf{a} = P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k}$:

$$\operatorname{curl} \mathbf{a} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{vmatrix} = \nabla \times \mathbf{a}$$

(vector product of the nabla vector and the vector **a**).

Each scalar field $u(x,y,z)$ generates a vector field $\operatorname{grad} u$. A vector field $\mathbf{a}(x,y,z)$ generates two fields: the scalar field div **a** and the vector field curl **a**.

► **Second-order differential operators.** The following differential identities hold:

$$\begin{aligned} &1)\quad \operatorname{curl}\operatorname{grad} u = \mathbf{0} \quad \text{or} \quad (\nabla \times \nabla)\,u = \mathbf{0},\\ &2)\quad \operatorname{div}\operatorname{curl} \mathbf{a} = 0 \quad \text{or} \quad \nabla \cdot (\nabla \times \mathbf{a}) = 0. \end{aligned}$$

The following differential relations hold:

$$\begin{aligned} &1)\quad \operatorname{div}\operatorname{grad} u = \Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2},\\ &2)\quad \operatorname{curl}\operatorname{curl} \mathbf{a} = \operatorname{grad}\operatorname{div} \mathbf{a} - \Delta \mathbf{a}, \end{aligned}$$

where Δ is the *Laplace operator*, $\Delta u = \nabla \cdot (\nabla u) = \nabla^2 u$.

Bibliography for Chapter M6

Adams, R., *Calculus: A Complete Course, 6th Edition*, Pearson Education, Toronto, 2006.

Boyer, C. B., *The History of the Calculus and Its Conceptual Development*, Dover Publications, New York, 1989.

Brannan, D., *A First Course in Mathematical Analysis*, Cambridge University Press, Cambridge, England, 2006.

Browder, A., *Mathematical Analysis: An Introduction*, Springer-Verlag, New York, 1996.

Courant, R. and John, F., *Introduction to Calculus and Analysis, Vol. 1*, Springer-Verlag, New York, 1999.

Edwards, C. H., and Penney, D., *Calculus, 6th Edition*, Pearson Education, Toronto, 2002.

Kline, M., *Calculus: An Intuitive and Physical Approach, 2nd Edition*, Dover Publications, New York, 1998.

Landau, E., *Differential and Integral Calculus*, American Mathematical Society, Providence, Rhode Island, 2001.

Silverman, R. A., *Essential Calculus with Applications*, Dover Publications, New York, 1989.

Zorich, V. A., *Mathematical Analysis*, Springer-Verlag, Berlin, 2004.

Chapter M7
Integrals

M7.1. Indefinite Integral

M7.1.1. Antiderivative. Indefinite Integral and Its Properties

► **Antiderivative.** An *antiderivative* (or *primitive function*) of a given function $f(x)$ on an interval (a, b) is a differentiable function $F(x)$ such that its derivative is equal to $f(x)$ for all $x \in (a, b)$:

$$F'(x) = f(x).$$

Example 1. Let $f(x) = 2x$. Then the functions $F(x) = x^2$ and $F_1(x) = x^2 - 1$ are antiderivatives of $f(x)$, since $(x^2)' = 2x$ and $(x^2 - 1)' = 2x$.

THEOREM. *Any function $f(x)$ continuous on an interval (a, b) has infinitely many continuous antiderivatives on (a, b). If $F(x)$ is one of them, then any other antiderivative has the form $F(x) + C$, where C is a constant.*

► **Indefinite integral.** The *indefinite integral* of a function $f(x)$ is the set, $F(x) + C$, of all its antiderivatives. This fact is written as

$$\int f(x)\,dx = F(x) + C.$$

Here, $f(x)$ is called the *integrand* (or the *integrand function*). The process of finding an integral is called integration. The differential dx indicates that the integration is carried out with respect to x.

Example 2. $\int 6x^2\,dx = 2x^3 + C$, since $(2x^3)' = 6x^2$.

► **Most important corollaries of the definition of the indefinite integral.** *Differentiation is the inverse of integration*:

$$\frac{d}{dx}\left(\int f(x)\,dx\right) = f(x).$$

Integration is the inverse of differentiation:*

$$\int f'(x)\,dx = f(x) + C.$$

The latter formula serves to make up tables of indefinite integrals. The procedure is often reversed here: an integral is first given in explicit form (i.e., the function $f(x)$ on the right-hand side is prescribed), and then the integrand is obtained by differentiation.

---

* Integration recovers the function from its derivative, to an additive constant.

M7.1.2. Table of Basic Integrals. Properties of the Indefinite Integral. Examples of Integration

► **Table of basic integrals.** Listed below are most common indefinite integrals, which are important for the integration of more complicated expressions:

$$\int x^a\,dx = \frac{x^{a+1}}{a+1} + C \quad (a \neq -1), \qquad \int \frac{dx}{x} = \ln|x| + C,$$

$$\int \frac{dx}{x^2+a^2} = \frac{1}{a}\arctan\frac{x}{a} + C, \qquad \int \frac{dx}{x^2-a^2} = \frac{1}{2a}\ln\left|\frac{x-a}{x+a}\right| + C,$$

$$\int \frac{dx}{\sqrt{a^2-x^2}} = \arcsin\frac{x}{a} + C, \qquad \int \frac{dx}{\sqrt{x^2+a}} = \ln\left|x+\sqrt{x^2+a}\,\right| + C,$$

$$\int e^x\,dx = e^x + C, \qquad \int a^x\,dx = \frac{a^x}{\ln a} + C,$$

$$\int \ln x\,dx = x\ln x - x + C, \qquad \int \ln ax\,dx = x\ln ax - x + C,$$

$$\int \sin x\,dx = -\cos x + C, \qquad \int \cos x\,dx = \sin x + C,$$

$$\int \tan x\,dx = -\ln|\cos x| + C, \qquad \int \cot x\,dx = \ln|\sin x| + C,$$

$$\int \frac{dx}{\sin x} = \ln\left|\tan\frac{x}{2}\right| + C, \qquad \int \frac{dx}{\cos x} = \ln\left|\tan\left(\frac{x}{2}+\frac{\pi}{4}\right)\right| + C,$$

$$\int \frac{dx}{\sin^2 x} = -\cot x + C, \qquad \int \frac{dx}{\cos^2 x} = \tan x + C,$$

$$\int \arcsin x\,dx = x\arcsin x + \sqrt{1-x^2} + C, \qquad \int \arccos x\,dx = x\arccos x - \sqrt{1-x^2} + C,$$

$$\int \arctan x\,dx = x\arctan x - \frac{1}{2}\ln(1+x^2) + C, \qquad \int \operatorname{arccot} x\,dx = x\operatorname{arccot} x + \frac{1}{2}\ln(1+x^2) + C,$$

$$\int \sinh x\,dx = \cosh x + C, \qquad \int \cosh x\,dx = \sinh x + C,$$

$$\int \tanh x\,dx = \ln\cosh x + C, \qquad \int \coth x\,dx = \ln|\sinh x| + C,$$

$$\int \frac{dx}{\sinh x} = \ln\left|\tanh\frac{x}{2}\right| + C, \qquad \int \frac{dx}{\cosh x} = 2\arctan e^x + C,$$

$$\int \frac{dx}{\sinh^2 x} = -\coth x + C, \qquad \int \frac{dx}{\cosh^2 x} = \tanh x + C,$$

$$\int \operatorname{arcsinh} x\,dx = x\operatorname{arcsinh} x - \sqrt{1+x^2} + C, \qquad \int \operatorname{arccosh} x\,dx = x\operatorname{arccosh} x - \sqrt{x^2-1} + C,$$

$$\int \operatorname{arctanh} x\,dx = x\operatorname{arctanh} x + \frac{1}{2}\ln(1-x^2) + C, \qquad \int \operatorname{arccoth} x\,dx = x\operatorname{arccoth} x + \frac{1}{2}\ln(x^2-1) + C,$$

where C is an arbitrary constant.

A more extensive table of indefinite integrals can be found in Section S1.1.

► **Properties of the indefinite integral.**

1. *A constant factor can be taken outside the integral sign*:

$$\int af(x)\,dx = a\int f(x)\,dx \qquad (a = \text{const}).$$

2. *Integral of the sum or difference of functions* (*additivity*):

$$\int [f(x) \pm g(x)]\,dx = \int f(x)\,dx \pm \int g(x)\,dx.$$

3. *Integration by parts*:

$$\int f(x)g'(x)\,dx = f(x)g(x) - \int f'(x)g(x)\,dx.$$

4. *Repeated integration by parts* (generalization of the previous formula):

$$\int f(x)g^{(n+1)}(x)\,dx = f(x)g^{(n)}(x) - f'(x)g^{(n-1)}(x) + \cdots + (-1)^n f^{(n)}(x)g(x) + (-1)^{n+1}\int f^{(n+1)}(x)g(x)\,dx, \qquad n = 0,\ 1,\ \ldots$$

5. *Change of variable* (*integration by substitution*):

$$\int f(x)\,dx = \int f(\varphi(t))\,\varphi'(t)\,dt, \qquad x = \varphi(t).$$

On computing the integral using the change of variable $x = \varphi(t)$, one should rewrite the resulting expression in terms of the original variable x using the inverse substitution $t = \varphi^{-1}(x)$.

► Examples of direct integration of elementary functions.

1°. With simple algebraic manipulation and the properties listed above, the integration may often be reduced to tabulated integrals.

Example 1. $\displaystyle\int \frac{2x-1}{\sqrt{x}}\,dx = \int\left(2\sqrt{x} - \frac{1}{\sqrt{x}}\right)dx = 2\int x^{1/2}\,dx - \int x^{-1/2}\,dx = \frac{4}{3}x^{3/2} - 2x^{1/2} + C.$

2°. Tabulated integrals can also be used where any function $\varphi(x)$ appears in place of x; for example,

$$\int e^x\,dx = e^x + C \quad\Longrightarrow\quad \int e^{\varphi(x)}\,d\varphi(x) = e^{\varphi(x)} + C;$$

$$\int \frac{dx}{x} = \ln|x| + C \quad\Longrightarrow\quad \int \frac{d\sin x}{\sin x} = \ln|\sin x| + C.$$

The reduction of an integral to a tabulated one may often be achieved by taking some function inside the differential sign.

Example 2. $\displaystyle\int \tan x\,dx = \int \frac{\sin x\,dx}{\cos x} = \int \frac{-d\cos x}{\cos x} = -\int \frac{d\cos x}{\cos x} = -\ln|\cos x| + C.$

3°. Integrals of the form $\displaystyle\int \frac{dx}{ax^2+bx+c}, \int \frac{dx}{\sqrt{ax^2+bx+c}}$ can be computed by making a perfect square:

$$ax^2 + bx + c = a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a} + c.$$

Then one should replace dx with the equal differential $d\bigl(x + \frac{b}{2a}\bigr)$ and use one of the four formulas in the second and third rows in the table of integrals given at the beginning of the current subsection.

Example 3. $\displaystyle\int \frac{dx}{\sqrt{2x-x^2}} = \int \frac{dx}{\sqrt{1-(x-1)^2}} = \int \frac{d(x-1)}{\sqrt{1-(x-1)^2}} = \arcsin(x-1) + C.$

4°. The integration of a polynomial multiplied by an exponential function can be accomplished by using the formula of integration by parts (or repeated integration by parts) given above.

Example 4. Compute the integral $\int (3x+1)\,e^{2x}\,dx$.

Taking $f(x)=3x+1$ and $g'(x)=e^{2x}$, one finds that $f'(x)=3$ and $g(x)=\frac{1}{2}e^{2x}$. On substituting these expressions into the formula of integration by parts, one obtains

$$\int (3x+1)\,e^{2x}\,dx=\frac{1}{2}(3x+1)\,e^{2x}-\frac{3}{2}\int e^{2x}\,dx=\frac{1}{2}(3x+1)\,e^{2x}-\frac{3}{4}e^{2x}+C=\Bigl(\frac{3}{2}x-\frac{1}{4}\Bigr)\,e^{2x}+C.$$

Remark 1. More complex examples of the application of integration by parts or repeated integration by parts can be found in Subsection M7.1.6.

Remark 2. Examples of using a change of variable (see Property 5 above) for the computation of integrals can be found in Subsections M7.1.4 and M7.1.5.

M7.1.3. Integration of Rational Functions

► **Partial fraction decomposition of a rational function.** A *rational function* (also known as a *rational polynomial function*) is a quotient of polynomials:

$$R(x)=\frac{P_n(x)}{Q_m(x)}, \tag{M7.1.3.1}$$

where

$$P_n(x)=a_nx^n+\cdots+a_1x+a_0,$$
$$Q_m(x)=b_mx^m+\cdots+b_1x+b_0.$$

The fraction (M7.1.3.1) is called *proper* if $m>n$ and *improper* if $m\le n$.

Every proper fraction (M7.1.3.1) can be decomposed into a sum of partial fractions. To this end, one should factorize the denominator $Q_m(x)$ into irreducible multipliers of the form

$$(x-\alpha_i)^{p_i}, \qquad i=1,2,\dots,k; \tag{M7.1.3.2a}$$
$$(x^2+\beta_jx+\gamma_j)^{q_j}, \qquad j=1,2,\dots,s, \tag{M7.1.3.2b}$$

where the p_i and q_j are positive integers satisfying the condition $p_1+\cdots+p_k+2(q_1+\cdots+q_s)=m$; $\beta_j^2-4\gamma_j<0$. The rational function (M7.1.3.1) can be represented as a sum of irreducibles and to each irreducible of the form (M7.1.3.2) there correspond as many terms as the power p_i or q_i:

$$\frac{A_{i,1}}{x-\alpha_i}+\frac{A_{i,2}}{(x-\alpha_i)^2}+\cdots+\frac{A_{i,p_i}}{(x-\alpha_i)^{p_i}}; \tag{M7.1.3.3a}$$
$$\frac{B_{j,1}x+D_{j,1}}{x^2+\beta_jx+\gamma_j}+\frac{B_{j,2}x+D_{j,2}}{(x^2+\beta_jx+\gamma_j)^2}+\cdots+\frac{B_{j,q_j}x+D_{j,q_j}}{(x^2+\beta_jx+\gamma_j)^{q_j}}. \tag{M7.1.3.3b}$$

The constants $A_{i,l}$, $B_{j,r}$, $D_{j,r}$ are found by the *method of undetermined coefficients*. To that end, one should equate the original rational fraction (M7.1.3.1) with the sum of the above partial fractions (M7.1.3.3) and reduce both sides of the resulting equation to a common denominator. Then, one collects the coefficients of like powers of x and equates them with zero, thus arriving at a system of linear algebraic equations for the $A_{i,l}$, $B_{j,r}$, and $D_{j,r}$.

Example 1. This is an illustration of how a proper fraction can be decomposed into partial fractions:

$$\frac{b_5x^5+b_4x^4+b_3x^3+b_2x^2+b_1x+b_0}{(x+a)(x+c)^3(x^2+k^2)}=\frac{A_{1,1}}{x+a}+\frac{A_{2,1}}{x+c}+\frac{A_{2,2}}{(x+c)^2}+\frac{A_{2,3}}{(x+c)^3}+\frac{Bx+D}{x^2+k^2}.$$

► Integration of a proper fraction.

1°. To integrate a proper fraction, one should first rewrite the integrand (M7.1.3.1) in the form of a sum of partial fractions. Below are the integrals of most common partial fractions (M7.1.3.3a) and (M7.1.3.3b) (with $q_j=1$):

$$\begin{aligned}&\int\frac{A}{x-\alpha}\,dx=A\ln|x-\alpha|,\qquad \int\frac{A}{(x-\alpha)^p}\,dx=-\frac{A}{(p-1)(x-\alpha)^{p-1}},\\&\int\frac{Bx+D}{x^2+\beta x+\gamma}\,dx=\frac{B}{2}\ln(x^2+\beta x+\gamma)+\frac{2D-B\beta}{\sqrt{4\gamma-\beta^2}}\arctan\frac{2x+\beta}{\sqrt{4\gamma-\beta^2}}.\end{aligned}\tag{M7.1.3.4}$$

The constant of integration C has been omitted here. More complex integrals of partial fractions (M7.1.3.3b) with $q_j>1$ can be computed using the formula

$$\int\frac{Bx+D}{(x^2+\beta x+\gamma)^q}\,dx=\frac{P(x)}{(x^2+\beta x+\gamma)^{q-1}}+\lambda\int\frac{dx}{x^2+\beta x+\gamma},\tag{M7.1.3.5}$$

where $P(x)$ is a polynomial of degree $2q-3$. The coefficients of $P(x)$ and the constant λ can be found by the method of undetermined coefficients by differentiating formula (M7.1.3.5).

Remark. The following recurrence relation may be used in order to compute the integrals on the left-hand side in (M7.1.3.5):

$$\int\frac{Bx+D}{(x^2+\beta x+\gamma)^q}\,dx=\frac{(2D-B\beta)x+D\beta-2B\gamma}{(q-1)(4\gamma-\beta^2)(x^2+\beta x+\gamma)^{q-1}}+\frac{(2q-3)(2D-B\beta)}{(q-1)(4\gamma-\beta^2)}\int\frac{dx}{(x^2+\beta x+\gamma)^{q-1}}.$$

Example 2. Compute the integral $\displaystyle\int\frac{3x^2-x-2}{x^3+8}\,dx$.

Let us factor the denominator of the integrand, $x^3+8=(x+2)(x^2-2x+4)$, and perform the partial fraction decomposition:

$$\frac{3x^2-x-2}{(x+2)(x^2-2x+4)}=\frac{A}{x+2}+\frac{Bx+D}{x^2-2x+4}.$$

Multiplying both sides by the common denominator and collecting the coefficients of like powers of x, we obtain

$$(A+B-3)x^2+(-2A+2B+D+1)x+4A+2D+2=0.$$

Now equating the coefficients of the different powers of x with zero, we arrive at a system of algebraic equations for A, B, and D:

$$A+B-3=0,\quad -2A+2B+D+1=0,\quad 4A+2D+2=0.$$

Its solution is: $A=1$, $B=2$, $D=-3$. Hence, we have

$$\begin{aligned}\int\frac{3x^2-x-2}{x^3+8}\,dx&=\int\frac{1}{x+2}\,dx+\int\frac{2x-3}{x^2-2x+4}\,dx\\&=\ln|x+2|+\ln\bigl(x^2-2x+4\bigr)-\frac{1}{\sqrt3}\arctan\frac{x-1}{\sqrt3}+C.\end{aligned}$$

Here, the last integral of (M7.1.3.4) has been used.

2°. The integrals of proper rational functions defined as the ratio of a polynomial to a power function $(x-\alpha)^m$ are given by the formulas

$$\int \frac{P_n(x)}{(x-\alpha)^m}\,dx = -\sum_{k=0}^{n} \frac{P_n^{(k)}(\alpha)}{k!\,(m-k-1)(x-\alpha)^{m-k-1}} + C, \qquad m > n+1;$$

$$\int \frac{P_n(x)}{(x-\alpha)^{n+1}}\,dx = -\sum_{k=0}^{n-1} \frac{P_n^{(k)}(\alpha)}{k!\,(n-k)(x-\alpha)^{n-k}} + \frac{P_n^{(n)}(\alpha)}{n!}\ln|x-\alpha| + C,$$

where $P_n(x)$ is a polynomial of degree n and $P_n^{(k)}(\alpha)$ is its kth derivative at $x=\alpha$.

3°. Suppose the roots in the factorization of the denominator of the fraction (M7.1.3.1) are all real and distinct:

$$Q_m(x) = b_m x^m + \cdots + b_1 x + b_0 = b_m(x-\alpha_1)(x-\alpha_2)\ldots(x-\alpha_m), \qquad \alpha_i \neq \alpha_j.$$

Then the following formula holds:

$$\int \frac{P_n(x)}{Q_m(x)}\,dx = \sum_{k=1}^{m} \frac{P_n(\alpha_k)}{Q'_m(\alpha_k)}\ln|x-\alpha_k| + C,$$

where $m > n$ and the prime denotes a derivative.

► **Integration of improper fractions.**

1°. In order to integrate an improper fraction, one should first isolate a proper fraction by division with remainder. As a result, the improper fraction is represented as the sum of a polynomial and a proper fraction,

$$\frac{a_n x^n + \cdots + a_1 x + a_0}{b_m x^m + \cdots + b_1 x + b_0} = c_m x^{n-m} + \cdots + c_1 x + c_0 + \frac{s_{m-1}x^{m-1} + \cdots + s_1 x + s_0}{b_m x^m + \cdots + b_1 x + b_0} \qquad (n \geq m),$$

which are then integrated separately.

Example 3. Evaluate the integral $I = \int \frac{x^2}{x-1}\,dx$.

Let us rewrite the integrand (improper fraction) as the sum of a polynomial and a proper fraction: $\frac{x^2}{x-1} = x + 1 + \frac{1}{x-1}$. Hence, $I = \int \left(x + 1 + \frac{1}{x-1}\right)dx = \frac{1}{2}x^2 + x + \ln|x-1| + C$.

2°. The integrals of improper rational functions defined as the ratio of a polynomial to a simple power function $(x-\alpha)^m$ are evaluated by the formula

$$\int \frac{P_n(x)}{(x-\alpha)^m}\,dx = \sum_{k=m}^{n} \frac{P_n^{(k)}(\alpha)}{k!\,(k-m+1)}(x-\alpha)^{k-m+1} + \frac{P_n^{(m-1)}(\alpha)}{(m-1)!}\ln|x-\alpha|$$
$$-\sum_{k=0}^{m-2} \frac{P_n^{(k)}(\alpha)}{k!\,(m-k-1)(x-\alpha)^{m-k-1}} + C,$$

where $n \geq m$.

Remark 1. The indefinite integrals of rational functions are always expressed in terms of elementary functions.

Remark 2. Some of the integrals reducible to integrals of rational functions are considered in Subsections M7.1.5 and M7.1.6.

M7.1.4. Integration of Irrational Functions

The integration of some irrational functions can be reduced to that of rational functions using a suitable change of variables. In what follows, the functions $R(x,y)$ and $R(x_1,\dots,x_k)$ are assumed to be rational functions in each of the arguments.

► **Integration of expressions involving radicals of linear-fractional functions.**

1°. The integrals with roots of linear functions

$$\int R\bigl(x,\ \sqrt[n]{ax+b}\bigr)\,dx$$

are reduced to integrals of rational functions by the change of variable $z=\sqrt[n]{ax+b}$.

Example 1. Evaluate the integral $I=\int x\sqrt{1-x}\,dx$.

With the change of variable $\sqrt{1-x}=z$, we have $x=1-z^2$ and $dx=-2z\,dz$. Substituting these expressions into the integral yields

$$I=-2\int(1-z^2)z^2\,dz=-\frac{2}{3}z^3+\frac{2}{5}z^5+C=-\frac{2}{3}\sqrt{(1-x)^3}+\frac{2}{5}\sqrt{(1-x)^5}+C.$$

2°. The integrals with roots of linear-fractional functions

$$\int R\left(x,\ \sqrt[n]{\frac{ax+b}{cx+d}}\right)dx$$

are reduced to integrals of rational functions by the substitution $z=\sqrt[n]{\dfrac{ax+b}{cx+d}}$.

3°. Integrals containing the product of a polynomial by a simple power function of the form $(x-a)^\beta$ are evaluated by the formula

$$\int P_n(x)(x-a)^\beta\,dx=\sum_{k=0}^{n}\frac{P_n^{(k)}(a)}{k!\,(k+\beta+1)}(x-a)^{k+\beta+1},$$

where $P_n(x)$ is a polynomial of degree n, $P_n^{(k)}(a)$ is its kth derivative at $x=a$, and β is any positive or negative proper fraction (to be more precise, $\beta\neq-1,-2,\dots,-n-1$).

► **Euler substitutions. Trigonometric substitutions.** We will be considering integrals involving the radical of a quadratic trinomial:

$$\int R\bigl(x,\sqrt{ax^2+bx+c}\bigr)\,dx,$$

where $b^2\neq 4ac$. Such integrals are expressible in terms of elementary functions.

1°. *Euler substitutions*. The given integral is reduced to the integral of a rational fraction by one of the following three Euler substitutions:

1) $\sqrt{ax^2+bx+c}=t\mp x\sqrt{a}$ if $a>0$;

2) $\sqrt{ax^2+bx+c}=xt\pm\sqrt{c}$ if $c>0$;

3) $\sqrt{ax^2+bx+c}=t(x-x_1)$ if $4ac-b^2<0$,

where x_1 is a root of the quadratic equation $ax^2+bx+c=0$. In all three cases, the variable x and the radical $\sqrt{ax^2+bx+c}$ are expressible in terms of the new variable t as (the formulas correspond to the upper signs in the substitutions):

$$\begin{aligned}
&1)\;\; x=\frac{t^2-c}{2\sqrt{a}\,t+b}, && \sqrt{ax^2+bx+c}=\frac{\sqrt{a}\,t^2+bt+c\sqrt{a}}{2\sqrt{a}\,t+b}, && dx=2\frac{\sqrt{a}\,t^2+bt+c\sqrt{a}}{(2\sqrt{a}\,t+b)^2}\,dt;\\
&2)\;\; x=\frac{2\sqrt{c}\,t-b}{a-t^2}, && \sqrt{ax^2+bx+c}=\frac{\sqrt{c}\,t^2-bt+c\sqrt{a}}{a-t^2}, && dx=2\frac{\sqrt{c}\,t^2-bt+c\sqrt{a}}{(a-t^2)^2}\,dt;\\
&3)\;\; x=\frac{(t^2+a)x_1+b}{t^2-a}, && \sqrt{ax^2+bx+c}=\frac{(2ax_1+b)t}{t^2-a}, && dx=-2\frac{(2ax_1+b)t}{(t^2-a)^2}\,dt.
\end{aligned}$$

2°. *Trigonometric substitutions*. The function $\sqrt{ax^2+bx+c}$ can be reduced, by making a perfect square in the radicand, to one of the three forms:

$$\begin{aligned}
&1)\;\; \sqrt{a}\sqrt{(x-p)^2+q^2} && \text{if}\quad a>0;\\
&2)\;\; \sqrt{a}\sqrt{(x-p)^2-q^2} && \text{if}\quad a>0;\\
&3)\;\; \sqrt{-a}\sqrt{q^2-(x-p)^2} && \text{if}\quad a<0,
\end{aligned}$$

where $p=-\frac{1}{2}b/a$. Different trigonometric substitutions are further used in each case to evaluate the integral:

$$\begin{aligned}
&1)\;\; x-p=q\tan t, && \sqrt{(x-p)^2+q^2}=\frac{q}{\cos t}, && dx=\frac{q\,dt}{\cos^2 t};\\
&2)\;\; x-p=\frac{q}{\cos t}, && \sqrt{(x-p)^2-q^2}=q\tan t, && dx=\frac{q\sin t\,dt}{\cos^2 t};\\
&3)\;\; x-p=q\sin t, && \sqrt{q^2-(x-p)^2}=q\cos t, && dx=q\cos t\,dt.
\end{aligned}$$

Example 2. Evaluate the integral $\int\sqrt{6+4x-2x^2}\,dx$.

This integral corresponds to case 3 with $a=-2$, $p=1$, and $q=2$. The integrand can be rewritten in the form:

$$\sqrt{6+4x-2x^2}=\sqrt{2}\sqrt{3+2x-x^2}=\sqrt{2}\sqrt{4-(x-1)^2}.$$

Using the trigonometric substitution $x-1=2\sin t$ and the formulas $\sqrt{3+2x-x^2}=2\cos t$ and $dx=2\cos t\,dt$, we obtain

$$\begin{aligned}
\int\sqrt{6+4x-2x^2}\,dx&=4\sqrt{2}\int\cos^2 t\,dt=2\sqrt{2}\int(1+\cos 2t)\,dt\\
&=2\sqrt{2}t+\sqrt{2}\sin 2t+C=2\sqrt{2}\arcsin\frac{x-1}{2}+\sqrt{2}\sin\Bigl(2\arcsin\frac{x-1}{2}\Bigr)+C\\
&=2\sqrt{2}\arcsin\frac{x-1}{2}+\frac{\sqrt{2}}{2}(x-1)\sqrt{4-(x-1)^2}+C.
\end{aligned}$$

► **Integral of a differential binomial.** The *integral of a differential binomial*,

$$\int x^m(a+bx^n)^p\,dx,$$

where a and b are constants, and n, m, p are rational numbers, is expressible in terms of elementary functions in the following three cases only:

1) If p is an integer. For $p \geq 0$, removing the brackets gives the sum of power functions. For $p < 0$, the substitution $x = t^r$, where r is the common denominator of the fractions m and n, leads to the integral of a rational function.

2) If $\frac{m+1}{n}$ is an integer. One uses the substitution $a+bx^n = t^k$, where k is the denominator of the fraction p.

3) If $\frac{m+1}{n} + p$ is an integer. One uses the substitution $ax^{-n} + b = t^k$, where k is the denominator of the fraction p.

Remark. In cases 2 and 3, the substitution $z = x^n$ leads to integrals of the form 3° above.

M7.1.5. Integration of Exponential and Trigonometric Functions

► Integration of exponential and hyperbolic functions.

1. Integrals of the form $\int R(e^{px}, e^{qx})\,dx$, where $R(x, y)$ is a rational function of its arguments and p, q are rational numbers, may be evaluated using the substitution $z^m = e^x$, where m is the common denominator of the fractions p and q. In the special case of integer p and q, we have $m = 1$, and the substitution becomes $z = e^x$.

Example 1. Evaluate the integral $\int \frac{e^{3x}\,dx}{e^x + 2}$.

This integral corresponds to integer p and q: $p = 1$ and $q = 3$. So we use the substitution $z = e^x$. Then $x = \ln z$ and $dx = \frac{dz}{z}$. Therefore,

$$\int \frac{e^{3x}\,dx}{e^x + 2} = \int \frac{z^2\,dz}{z + 2} = \int \Bigl(z - 2 + \frac{4}{z + 2}\Bigr) dz = \frac{1}{2}z^2 - 2z + 4\ln|z + 2| + C = \frac{1}{2}e^{2x} - 2e^x + 4\ln(e^x + 2) + C.$$

2. Integrals of the form $\int R(\sinh ax, \cosh ax)\,dx$ are evaluated by converting the hyperbolic functions to exponentials, using the formulas $\sinh ax = \frac{1}{2}(e^{ax} - e^{-ax})$ and $\cosh ax = \frac{1}{2}(e^{ax} + e^{-ax})$, and performing the substitution $z = e^{ax}$. Then

$$\int R(\sinh ax, \cosh ax)\,dx = \frac{1}{a}\int R\biggl(\frac{z^2 - 1}{2z}, \frac{z^2 + 1}{2z}\biggr)\frac{dz}{z}.$$

Alternatively, the substitution $t = \tanh\Bigl(\frac{ax}{2}\Bigr)$ can also be used to evaluate integrals of the above form. Then

$$\int R(\sinh ax, \cosh ax)\,dx = \frac{2}{a}\int R\biggl(\frac{2t}{1 - t^2}, \frac{1 + t^2}{1 - t^2}\biggr)\frac{dt}{1 - t^2}.$$

► Integration of trigonometric functions.

1. Integrals of the form $\int R(\sin ax, \cos ax)\,dx$ can be converted to integrals of rational functions using the fundamental trigonometric substitution $t = \tan\Bigl(\frac{ax}{2}\Bigr)$:

$$\int R(\sin ax, \cos ax)\,dx = \frac{2}{a}\int R\biggl(\frac{2t}{1 + t^2}, \frac{1 - t^2}{1 + t^2}\biggr)\frac{dt}{1 + t^2}.$$

Example 2. Evaluate the integral $\int \frac{dx}{2+\sin x}$.

Using the fundamental trigonometric substitution $t = \tan\frac{x}{2}$, we have

$$\int \frac{dx}{2+\sin x} = 2\int \frac{dt}{\left(2+\frac{2t}{1+t^2}\right)(1+t^2)} = \int \frac{dt}{t^2+t+1} = 2\int \frac{d(2t+1)}{(2t+1)^2+3}$$
$$= \frac{2}{\sqrt{3}}\arctan\frac{2t+1}{\sqrt{3}} + C = \frac{2}{\sqrt{3}}\arctan\left(\frac{2}{\sqrt{3}}\tan\frac{x}{2} + \frac{1}{\sqrt{3}}\right) + C.$$

2. Integrals of the form $\int R(\sin^2 ax, \cos^2 ax, \tan ax)\,dx$ are converted to integrals of rational functions with the change of variable $z = \tan ax$:

$$\int R(\sin^2 ax, \cos^2 ax, \tan ax)\,dx = \frac{1}{a}\int R\left(\frac{z^2}{1+z^2}, \frac{1}{1+z^2}, z\right)\frac{dz}{1+z^2}.$$

3. Integrals of the form

$$\int \sin ax \cos bx\,dx, \qquad \int \cos ax \cos bx\,dx, \qquad \int \sin ax \sin bx\,dx$$

are evaluated using the formulas

$$\begin{aligned}\sin\alpha\cos\beta &= \tfrac{1}{2}[\sin(\alpha+\beta)+\sin(\alpha-\beta)],\\ \cos\alpha\cos\beta &= \tfrac{1}{2}[\cos(\alpha+\beta)+\cos(\alpha-\beta)],\\ \sin\alpha\sin\beta &= \tfrac{1}{2}[\cos(\alpha-\beta)-\cos(\alpha+\beta)].\end{aligned}$$

4. Integrals of the form $\int \sin^m x\cos^n x\,dx$, where m and n are integers, are evaluated as follows:

(a) if m is odd, one uses the change of variable $\cos x = z$, with $\sin x\,dx = -dz$;
(b) if n is odd, one uses the change of variable $\sin x = z$, with $\cos x\,dx = dz$;
(c) if m and n are both even nonnegative integers, one should use the degree reduction formulas

$$\sin^2 x = \tfrac{1}{2}(1-\cos 2x), \quad \cos^2 x = \tfrac{1}{2}(1+\cos 2x), \quad \sin x\cos x = \tfrac{1}{2}\sin 2x.$$

Example 3. Evaluate the integral $\int \sin^5 x\,dx$.

This integral corresponds to odd m: $m = 5$. With simple rearrangement and the change of variable $\cos x = z$, we have

$$\int \sin^5 x\,dx = \int (\sin^2 x)^2 \sin x\,dx = -\int (1-\cos^2 x)^2\,d\cos x = -\int (1-z^2)^2\,dz$$
$$= \tfrac{2}{3}z^3 - \tfrac{1}{5}z^5 - z + C = \tfrac{2}{3}\cos^3 x - \tfrac{1}{5}\cos^5 x - \cos x + C.$$

Remark. In general, the integrals $\int \sin^p x\cos^q x\,dx$ are reduced to the integral of a differential binomial by the substitution $y = \sin x$.

M7.1.6. Integration of Polynomials Multiplied by Elementary Functions

Throughout this section, $P_n(x)$ designates a polynomial of degree n.

► **Integration of the product of a polynomial by exponential functions.** General formulas:

$$\int P_n(x)e^{ax}\,dx = e^{ax}\left[\frac{P_n(x)}{a} - \frac{P_n'(x)}{a^2} + \cdots + (-1)^n\frac{P_n^{(n)}(x)}{a^{n+1}}\right] + C,$$

$$\int P_n(x)\cosh(ax)\,dx = \sinh(ax)\left[\frac{P_n(x)}{a} + \frac{P_n''(x)}{a^3} + \cdots\right] - \cosh(ax)\left[\frac{P_n'(x)}{a^2} + \frac{P_n'''(x)}{a^4} + \cdots\right] + C,$$

$$\int P_n(x)\sinh(ax)\,dx = \cosh(ax)\left[\frac{P_n(x)}{a} + \frac{P_n''(x)}{a^3} + \cdots\right] - \sinh(ax)\left[\frac{P_n'(x)}{a^2} + \frac{P_n'''(x)}{a^4} + \cdots\right] + C.$$

These formulas are obtained by repeated integration by parts; see Property 4 from Subsection M7.1.2 with $f(x) = P_n(x)$ for $g^{(n+1)}(x) = e^{ax}$, $g^{(n+1)}(x) = \cosh(ax)$, and $g^{(n+1)}(x) = \sinh(ax)$, respectively.

In the special case $P_n(x) = x^n$, the first formula gives

$$\int x^n e^{ax}\,dx = e^{ax}\sum_{k=0}^{n}\frac{(-1)^{n-k}}{a^{n+1-k}}\frac{n!}{k!}x^k + C. \qquad \text{(M7.1.6.1)}$$

► **Integration of the product of a polynomial by a trigonometric function.**

1°. General formulas:

$$\int P_n(x)\cos(ax)\,dx = \sin(ax)\left[\frac{P_n(x)}{a} - \frac{P_n''(x)}{a^3} + \cdots\right] + \cos(ax)\left[\frac{P_n'(x)}{a^2} - \frac{P_n'''(x)}{a^4} + \cdots\right] + C,$$

$$\int P_n(x)\sin(ax)\,dx = \sin(ax)\left[\frac{P_n'(x)}{a^2} - \frac{P_n'''(x)}{a^4} + \cdots\right] - \cos(ax)\left[\frac{P_n(x)}{a} - \frac{P_n''(x)}{a^3} + \cdots\right] + C.$$

These formulas are obtained by repeated integration by parts; see Property 4 from Subsection M7.1.2 with $f(x) = P_n(x)$ for $g^{(n+1)}(x) = \cos(ax)$ and $g^{(n+1)}(x) = \sin(ax)$, respectively.

2°. To evaluate integrals of the form

$$\int P_n(x)\cos^m(ax)\,dx, \qquad \int P_n(x)\sin^m(ax)\,dx,$$

with $m = 2, 3, \ldots$, one should first use the trigonometric formulas

$$\cos^{2k}(ax) = \frac{1}{2^{2k-1}}\sum_{i=0}^{k-1}C_{2k}^{i}\cos[2(k-i)ax] + \frac{1}{2^{2k}}C_{2k}^{k} \qquad (m = 2k),$$

$$\cos^{2k+1}(ax) = \frac{1}{2^{2k}}\sum_{i=0}^{k}C_{2k+1}^{i}\cos[(2k-2i+1)ax] \qquad (m = 2k+1),$$

$$\sin^{2k}(ax) = \frac{1}{2^{2k-1}}\sum_{i=0}^{k-1}(-1)^{k-i}C_{2k}^{i}\cos[2(k-i)ax] + \frac{1}{2^{2k}}C_{2k}^{k} \qquad (m = 2k),$$

$$\sin^{2k+1}(ax) = \frac{1}{2^{2k}}\sum_{i=0}^{k}(-1)^{k-i}C_{2k+1}^{i}\sin[(2k-2i+1)ax] \qquad (m = 2k+1),$$

thus reducing the above integrals to those considered in Item 1°.

3°. Integrals of the form

$$\int P_n(x)\, e^{ax} \sin(bx)\, dx, \qquad \int P_n(x)\, e^{ax} \cos(bx)\, dx$$

can be evaluated by repeated integration by parts.

In particular,

$$\begin{aligned} \int x^n e^{ax} \sin(bx) &= e^{ax} \sum_{k=1}^{n+1} \frac{(-1)^{k+1} n!}{(n-k+1)!\,(a^2+b^2)^{k/2}} x^{n-k+1} \sin(bx+k\theta) + C, \\ \int x^n e^{ax} \cos(bx) &= e^{ax} \sum_{k=1}^{n+1} \frac{(-1)^{k+1} n!}{(n-k+1)!\,(a^2+b^2)^{k/2}} x^{n-k+1} \cos(bx+k\theta) + C, \end{aligned} \qquad (*)$$

where

$$\sin\theta = -\frac{b}{\sqrt{a^2+b^2}}, \qquad \cos\theta = \frac{a}{\sqrt{a^2+b^2}}.$$

▶ Integrals involving power and logarithmic functions.

1°. The formula of integration by parts with $g'(x) = P_n(x)$ is effective in the evaluation of integrals of the form

$$\int P_n(x) \ln(ax)\, dx = Q_{n+1}(x) \ln(ax) - a \int \frac{Q_{n+1}(x)}{x}\, dx,$$

where $Q_{n+1}(x) = \int P_n(x)\, dx$ is a polynomial of degree $n+1$. The integral on the right-hand side is easy to take, since the integrand is the sum of power functions.

Example. Evaluate the integral $\int \ln x\, dx$.

Setting $f(x) = \ln x$ and $g'(x) = 1$, we find $f'(x) = \frac{1}{x}$ and $g(x) = x$. Substituting these expressions into the formula of integration by parts, we obtain $\int \ln x\, dx = x \ln x - \int dx = x \ln x - x + C$.

2°. The easiest way to evaluate integrals of the more general form

$$I = \int \sum_{i=0}^{n} \ln^i(ax) \biggl(\sum_{j=0}^{m} b_{ij} x^{\beta_{ij}} \biggr)\, dx,$$

where the β_{ij} are arbitrary numbers, is to use the substitution $z = \ln(ax)$, so that

$$I = \int \sum_{i=0}^{n} z^i \biggl(\sum_{j=0}^{m} \frac{b_{ij}}{a^{\beta_{ij}+1}} e^{(\beta_{ij}+1)z} \biggr)\, dz.$$

By removing the brackets, one obtains a sum of integrals like $\int x^n e^{ax}\, dx$, which are easy to evaluate by formula (M7.1.6.1).

M7.2. Definite Integral

M7.2.1. Basic Definitions. Classes of Integrable Functions. Geometrical Meaning of the Definite Integral

► **Basic definitions.** Let $y = f(x)$ be a bounded function defined on a finite closed interval $[a, b]$. Let us partition this interval into n elementary subintervals defined by a set of points $\{x_0, x_1, \dots, x_n\}$ such that $a = x_0 < x_1 < \cdots < x_n = b$. Each subinterval $[x_{k-1}, x_k]$ will be characterized by its length $\Delta x_k = x_k - x_{k-1}$ and an arbitrarily chosen point $\xi_k \in [x_{k-1}, x_k]$. Let us make up an *integral sum* (a *Cauchy–Riemann sum*, also known as a *Riemann sum*)

$$s_n = \sum_{k=1}^{n} f(\xi_k)\Delta x_k \qquad (x_{k-1} \le \xi_k \le x_k).$$

If, as $\Delta x_k \to 0$ for all k and, accordingly, $n \to \infty$, there exists a finite limit of the integral sums s_n and it depends on neither the way the interval $[a, b]$ was partitioned, nor the selection of the points ξ_k, then this limit is denoted $\int_a^b f(x)\,dx$ and is called the *definite integral* (also the *Riemann integral*) of the function $y = f(x)$ over the interval $[a, b]$:

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} s_n \qquad \Bigl(\max_{1\le k\le n} \Delta x_k \to 0\Bigr).$$

In this case, the function $f(x)$ is called *integrable* on the interval $[a, b]$.

► **Classes of integrable functions.**

1. If a function $f(x)$ is continuous on an interval $[a, b]$, then it is integrable on this interval.
2. If a bounded function $f(x)$ has finitely many jump discontinuities on $[a, b]$, then it is integrable on $[a, b]$.
3. A monotonic bounded function $f(x)$ on $[a, b]$ is always integrable on $[a, b]$.

► **Geometric meaning of the definite integral.** If $f(x) \ge 0$ on $[a, b]$, then the integral $\int_a^b f(x)\,dx$ is equal to the area of the domain $D = \{a \le x \le b,\ 0 \le y \le f(x)\}$ (the area of the curvilinear trapezoid shown in Fig. M7.1).

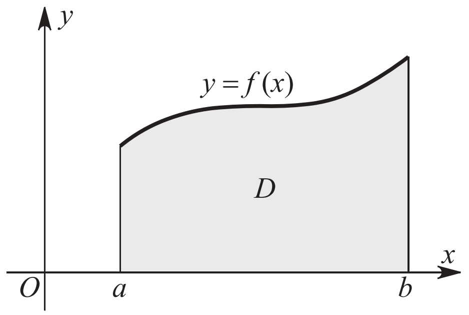

Figure M7.1. The integral of a nonnegative function $f(x)$ on an interval $[a, b]$ is equal to the area of the shaded region.

M7.2.2. Properties of Definite Integrals and Useful Formulas

► Qualitative properties of integrals.

1. If a function $f(x)$ is integrable on $[a,b]$, then the functions $cf(x)$, with $c = \text{const}$, and $|f(x)|$ are also integrable on $[a,b]$.

2. If two functions $f(x)$ and $g(x)$ are integrable on $[a,b]$, then their sum, difference, and product are also integrable on $[a,b]$.

3. If a function $f(x)$ is integrable on $[a,b]$ and its values lie within an interval $[c,d]$, where a function $g(y)$ is defined and continuous, then the composite function $g(f(x))$ is also integrable on $[a,b]$.

4. If a function $f(x)$ is integrable on $[a,b]$, then it is also integrable and on any subinterval $[\alpha,\beta] \subset [a,b]$. Conversely, if an interval $[a,b]$ is partitioned into a number of subintervals and $f(x)$ is integrable on each of the subintervals, then it is integrable on the whole interval $[a,b]$.

5. If the values of a function are changed at finitely many points, this will not affect the integrability of the function and will not change the value of the integral.

► Properties of integrals in terms of identities.

1. *The integral over a zero-length interval is zero*:

$$\int_a^a f(x)\,dx = 0.$$

2. *Antisymmetry under the swap of the integration limits*:

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx.$$

3. *Linearity*. If functions $f(x)$ and $g(x)$ are integrable on an interval $[a,b]$, then

$$\int_a^b \bigl[Af(x) \pm Bg(x)\bigr]\,dx = A\int_a^b f(x)\,dx \pm B\int_a^b g(x)\,dx$$

for any numbers A and B.

4. *Additivity*. If $c \in [a,b]$ and $f(x)$ is integrable on $[a,b]$, then

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

Remark. This property is also valid in the case where $c \notin [a,b]$.

5. *Differentiation with respect to a variable upper limit*. If $f(x)$ is continuous on $[a,b]$, then the function $\Phi(x) = \int_a^x f(t)\,dt$ is differentiable on $[a,b]$, and $\Phi'(x) = f(x)$. This fact can be written as

$$\frac{d}{dx}\left(\int_a^x f(t)\,dt\right) = f(x).$$

6. *Newton–Leibniz formula*:

$$\int_a^b f(x)\,dx = F(x)\Big|_a^b = F(b) - F(a),$$

where $F(x)$ is an antiderivative of $f(x)$ on $[a,b]$.

7. *Integration by parts*. If functions $f(x)$ and $g(x)$ have continuous derivatives on $[a,b]$, then

$$\int_a^b f(x)g'(x)\,dx = \bigl[f(x)g(x)\bigr]\Big|_a^b - \int_a^b f'(x)g(x)\,dx.$$

8. *Repeated integration by parts*:

$$\int_a^b f(x)g^{(n+1)}(x)\,dx = \Bigl[f(x)g^{(n)}(x) - f'(x)g^{(n-1)}(x) + \cdots + (-1)^n f^{(n)}(x)g(x)\Bigr]_a^b$$
$$+ (-1)^{n+1}\int_a^b f^{(n+1)}(x)g(x)\,dx, \qquad n = 0,\ 1,\ \ldots$$

9. *Change of variable (substitution) in a definite integral*. Let $f(x)$ be a continuous function on $[a,b]$ and let $x(t)$ be a continuously differentiable function on $[\alpha,\beta]$. Suppose also that the range of values of $x(t)$ coincides with $[a,b]$, with $x(\alpha) = a$ and $x(\beta) = b$. Then

$$\int_a^b f(x)\,dx = \int_\alpha^\beta f\bigl(x(t)\bigr)\,x'(t)\,dt.$$

Example. Evaluate the integral $\displaystyle\int_0^3 \frac{dx}{(x-8)\sqrt{x+1}}$.

Perform the substitution $x + 1 = t^2$, with $dx = 2t\,dt$. We have $t = 1$ at $x = 0$ and $t = 2$ at $x = 3$. Therefore,

$$\int_0^3 \frac{dx}{(x-8)\sqrt{x+1}} = \int_1^2 \frac{2t\,dt}{(t^2-9)t} = 2\int_1^2 \frac{dt}{t^2-9} = \frac{1}{3}\ln\left|\frac{t-3}{t+3}\right|\Bigg|_1^2 = \frac{1}{3}\ln\frac{2}{5}.$$

M7.2.3. Asymptotic Formulas for the Calculation of Integrals

Below are some general formulas, involving arbitrary functions and parameters, that may be helpful for obtaining asymptotics of integrals.

► **Asymptotic formulas for integrals with weak singularity as $\varepsilon \longrightarrow 0$.** We will consider integrals of the form

$$I(\varepsilon) = \int_0^a \frac{x^{\beta-1}f(x)}{(x+\varepsilon)^\alpha}\,dx,$$

where $0 < a < \infty$, $\beta > 0$, $f(0) \neq 0$, and $\varepsilon > 0$ is a small parameter.

The integral diverges as $\varepsilon \to 0$ for $\alpha \geq \beta$, that is, $\lim\limits_{\varepsilon\to 0} I(\varepsilon) = \infty$. In this case, the leading term of the asymptotic expansion of the integral $I(\varepsilon)$ is given by

$$\begin{aligned} I(\varepsilon) &= \frac{\Gamma(\beta)\Gamma(\alpha-\beta)}{\Gamma(\alpha)} f(0)\varepsilon^{\beta-\alpha} + O(\varepsilon^\sigma) && \text{if} \quad \alpha > \beta, \\ I(\varepsilon) &= -f(0)\ln\varepsilon + O(1) && \text{if} \quad \alpha = \beta, \end{aligned}$$

where $\Gamma(\beta)$ is the gamma function and $\sigma = \min[\beta - \alpha + 1,\ 0]$.

► **Asymptotic formulas for Laplace integrals of special form as $\lambda \longrightarrow +\infty$.** Consider a *Laplace integral of the special form*

$$I(\lambda) = \int_0^a x^{\beta-1}\exp\bigl(-\lambda x^\alpha\bigr)f(x)\,dx,$$

where $0 < a < \infty$, $\alpha > 0$, and $\beta > 0$.

The following formula, called *Watson's asymptotic formula*, holds as $\lambda \to +\infty$:

$$I(\lambda) = \frac{1}{\alpha}\sum_{k=0}^{n}\frac{f^{(k)}(0)}{k!}\Gamma\Bigl(\frac{k+\beta}{\alpha}\Bigr)\lambda^{-(k+\beta)/\alpha} + O\bigl(\lambda^{-(n+\beta+1)/\alpha}\bigr).$$

Remark. Watson's formula also holds for improper integrals with $a = \infty$ if the original integral converges absolutely for some $\lambda_0 > 0$.

► **Asymptotic formulas for Laplace integrals of general form as $\lambda \to +\infty$.** Consider a *Laplace integral of the general form*

$$I(\lambda) = \int_a^b f(x)\exp[\lambda g(x)]\,dx, \qquad \text{(M7.2.3.1)}$$

where $[a, b]$ is a finite interval and $f(x)$, $g(x)$ are continuous functions.

Leading term of the asymptotic expansion of the integral (M7.2.3.1) as $\lambda \to +\infty$. Suppose the function $g(x)$ attains a maximum on $[a, b]$ at only one point $x_0 \in [a, b]$ and is differentiable in a neighborhood of x_0, with $g'(x_0) = 0$, $g''(x_0) \neq 0$, and $f(x_0) \neq 0$. Then the leading term of the asymptotic expansion of the integral (M7.2.3.1), as $\lambda \to +\infty$, is given by

$$\begin{aligned} I(\lambda) &= f(x_0)\sqrt{-\frac{2\pi}{\lambda g''(x_0)}}\exp[\lambda g(x_0)] \quad &&\text{if } a < x_0 < b,\\ I(\lambda) &= \frac{1}{2}f(x_0)\sqrt{-\frac{2\pi}{\lambda g''(x_0)}}\exp[\lambda g(x_0)] \quad &&\text{if } x_0 = a \text{ or } x_0 = b. \end{aligned} \qquad \text{(M7.2.3.2)}$$

Note that the latter formula differs from the former by the factor $1/2$ only.

Under the same conditions, if $g(x)$ attains a maximum at either endpoint, $x_0 = a$ or $x_0 = b$, but $g'(x_0) \neq 0$, then the leading asymptotic term of the integral, as $\lambda \to +\infty$, is

$$I(\lambda) = \frac{f(x_0)}{|g'(x_0)|}\frac{1}{\lambda}\exp[\lambda g(x_0)], \quad \text{where } x_0 = a \text{ or } x_0 = b. \qquad \text{(M7.2.3.3)}$$

For more accurate asymptotic estimates for the Laplace integral (M7.2.3.1), see below.

► **Asymptotic formulas for a power Laplace integral.** Consider the *power Laplace integral*, which is obtained from the exponential Laplace integral (M7.2.3.1) by substituting $\ln g(x)$ for $g(x)$:

$$I(\lambda) = \int_a^b f(x)[g(x)]^\lambda\,dx, \qquad \text{(M7.2.3.6)}$$

where $[a, b]$ is a finite closed interval and $g(x) > 0$. It is assumed that the functions $f(x)$ and $g(x)$ appearing in the integral (M7.2.3.6) are continuous; $g(x)$ is assumed to attain a maximum at only one point $x_0 = [a, b]$ and to be differentiable in a neighborhood of $x = x_0$, with $g'(x_0) = 0$, $g''(x_0) \neq 0$, and $f(x_0) \neq 0$. Then the leading asymptotic term of the integral, as $\lambda \to +\infty$, is expressed as

$$\begin{aligned} I(\lambda) &= f(x_0)\sqrt{-\frac{2\pi}{\lambda g''(x_0)}}\,[g(x_0)]^{\lambda+1/2} \quad &&\text{if } a < x_0 < b,\\ I(\lambda) &= \frac{1}{2}f(x_0)\sqrt{-\frac{2\pi}{\lambda g''(x_0)}}\,[g(x_0)]^{\lambda+1/2} \quad &&\text{if } x_0 = a \text{ or } x_0 = b. \end{aligned}$$

Note that the latter formula differs from the former by the factor $1/2$ only.

Under the same conditions, if $g(x)$ attains a maximum at either endpoint, $x_0 = a$ or $x_0 = b$, but $g'(x_0) \neq 0$, then the leading asymptotic term of the integral, as $\lambda \to +\infty$, is

$$I(\lambda) = \frac{f(x_0)}{|g'(x_0)|}\frac{1}{\lambda}[g(x_0)]^{\lambda+1/2}, \qquad \text{where } x_0 = a \text{ or } x_0 = b.$$

M7.2.4. Mean Value Theorems. Properties of Integrals in Terms of Inequalities

► Mean value theorems.

THEOREM 1. *If $f(x)$ is a continuous function on $[a,b]$, there exists at least one point $c \in (a,b)$ such that*

$$\int_a^b f(x)\,dx = f(c)(b-a).$$

The number $f(c)$ is called the mean value of the function $f(x)$ on $[a,b]$.

THEOREM 2. *If $f(x)$ is a continuous function on $[a,b]$, and $g(x)$ is integrable and of constant sign ($g(x) \geq 0$ or $g(x) \leq 0$) on $[a,b]$, then there exists at least one point $c \in (a,b)$ such that*

$$\int_a^b f(x)g(x)\,dx = f(c)\int_a^b g(x)\,dx.$$

► Properties of integrals in terms of inequalities.

1. *Estimation theorem*. If $m \leq f(x) \leq M$ on $[a,b]$, then

$$m(b-a) \leq \int_a^b f(x)\,dx \leq M(b-a).$$

2. *Inequality integration theorem*. If $\varphi(x) \leq f(x) \leq g(x)$ on $[a,b]$, then

$$\int_a^b \varphi(x)\,dx \leq \int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx.$$

In particular, if $f(x) \geq 0$ on $[a,b]$, then $\int_a^b f(x)\,dx \geq 0$.

Further on, it is assumed that the integrals on the right-hand sides of the inequalities of Items 3–6 exist.

3. *Absolute value theorem* (*integral analogue of the triangle inequality*):

$$\left|\int_a^b f(x)\,dx\right| \leq \int_a^b |f(x)|\,dx.$$

4. *Bunyakovsky's inequality* (*Cauchy–Schwarz–Bunyakovsky inequality*):

$$\left(\int_a^b f(x)g(x)\,dx\right)^2 \leq \int_a^b f^2(x)\,dx \int_a^b g^2(x)\,dx.$$

5. *Cauchy's inequality*:

$$\left(\int_a^b [f(x)+g(x)]^2\,dx\right)^{1/2} \leq \left(\int_a^b f^2(x)\,dx\right)^{1/2} + \left(\int_a^b g^2(x)\,dx\right)^{1/2}.$$

6. *Minkowski's inequality* (generalization of Cauchy's inequality):

$$\left(\int_a^b |f(x)+g(x)|^p\,dx\right)^{\frac{1}{p}} \leq \left(\int_a^b |f(x)|^p\,dx\right)^{\frac{1}{p}} + \left(\int_a^b |g(x)|^p\,dx\right)^{\frac{1}{p}}, \qquad p \geq 1.$$

M7.2.5. Geometric and Physical Applications of the Definite Integral

► **Geometric applications of the definite integral.**

1. The *area of a domain* D bounded by curves $y = f(x)$ and $y = g(x)$ and straight lines $x = a$ and $x = b$ in the xy-plane (see Fig. M7.2 *a*) is calculated by the formula

$$S = \int_a^b \bigl[f(x) - g(x)\bigr]\,dx.$$

If $g(x) \equiv 0$, this formula gives the area of a curvilinear trapezoid bounded by the x-axis, the curve $y = f(x)$, and the straight lines $x = a$ and $x = b$.

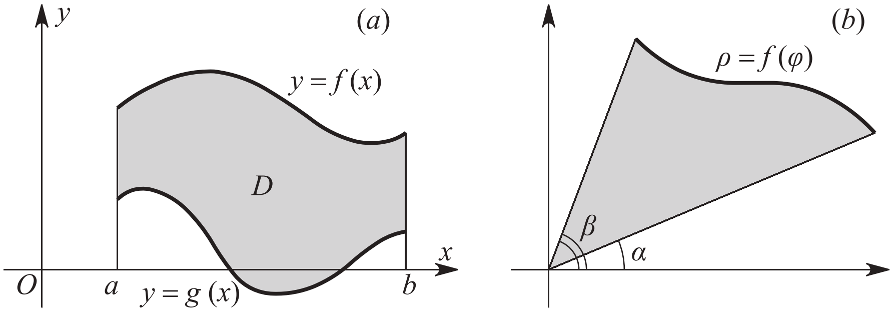

Figure M7.2. (*a*) A domain D bounded by two curves $y = f(x)$ and $y = g(x)$ on an interval $[a, b]$; (*b*) a curvilinear sector.

2. *Area of a domain* D. Let $x = x(t)$ and $y = y(t)$, with $t_1 \le t \le t_2$, be parametric equations of a piecewise-smooth simple closed curve bounding on its left (traced counterclockwise) a domain D with area S. Then

$$S = -\int_{t_1}^{t_2} y(t)x'(t)\,dt = \int_{t_1}^{t_2} x(t)y'(t)\,dt = \frac{1}{2}\int_{t_1}^{t_2} \bigl[x(t)y'(t) - y(t)x'(t)\bigr]\,dt.$$

3. *Area of a curvilinear sector*. Let a curve $\rho = f(\varphi)$, with $\varphi \in [\alpha, \beta]$, be defined in the polar coordinates ρ, φ. Then the area of the curvilinear sector $\{\alpha \le \varphi \le \beta;\ 0 \le \rho \le f(\varphi)\}$ (see Fig. M7.2 *b*) is calculated by the formula

$$S = \frac{1}{2}\int_\alpha^\beta [f(\varphi)]^2\,d\varphi.$$

4. *Area of a surface of revolution*. Let a surface of revolution be generated by rotating a curve $y = f(x) \ge 0$, $x \in [a, b]$, about the x-axis; see Fig. M7.3. The area of this surface is calculated as

$$S = 2\pi \int_a^b f(x)\sqrt{1 + [f'(x)]^2}\,dx.$$

5. *Volume of a body of revolution*. Let a body of revolution be obtained by rotating about the x-axis a curvilinear trapezoid bounded by a curve $y = f(x)$, the x-axis, and straight lines $x = a$ and $x = b$; see Fig. M7.3. Then the volume of this body is calculated as

$$V = \pi \int_a^b [f(x)]^2\,dx.$$

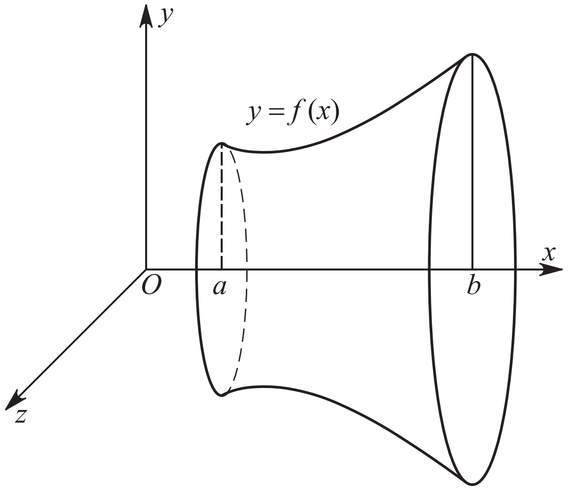

Figure M7.3. A surface of revolution generated by rotating a curve $y = f(x)$.

6. *Arc length of a plane curve defined in different ways.*

(*a*) If a curve is the graph of a continuously differentiable function $y = f(x)$, $x \in [a, b]$, then its length is determined as

$$L = \int_a^b \sqrt{1 + [f'(x)]^2}\, dx.$$

(*b*) If a plane curve is defined parametrically by equations $x = x(t)$ and $y = y(t)$, with $t \in [\alpha, \beta]$ and $x(t)$ and $y(t)$ being continuously differentiable functions, then its length is calculated by

$$L = \int_\alpha^\beta \sqrt{[x'(t)]^2 + [y'(t)]^2}\, dt.$$

(*c*) If a curve is defined in the polar coordinates ρ, φ by an equation $\rho = \rho(\varphi)$, with $\varphi \in [\alpha, \beta]$, then its length is found as

$$L = \int_\alpha^\beta \sqrt{\rho^2(\varphi) + [\rho'(\varphi)]^2}\, d\varphi.$$

7. The *arc length of a spatial curve* defined parametrically by equations $x = x(t)$, $y = y(t)$, and $z = z(t)$, with $t \in [\alpha, \beta]$ and $x(t)$, $y(t)$, and $z(t)$ being continuously differentiable functions, is calculated by

$$L = \int_\alpha^\beta \sqrt{[x'(t)]^2 + [y'(t)]^2 + [z'(t)]^2}\, dt.$$

► Physical applications of the integral.

1. *Work of a variable force*. Suppose a point mass moves along the x-axis from a point $x = a$ to a point $x = b$ under the action of a variable force $F(x)$ directed along the x-axis. The mechanical work of this force is equal to

$$A = \int_a^b F(x)\, dx.$$

2. *Mass of a rectilinear rod of variable density.* Suppose a rod with a constant cross-sectional area S occupies an interval $[0, l]$ on the x-axis and the density of the rod material is a function of x: $\rho = \rho(x)$. The mass of this rod is calculated as

$$m = S\int_0^l \rho(x)\,dx.$$

3. *Mass of a curvilinear rod of variable density.* Let the shape of a plane curvilinear rod with a constant cross-sectional area S be defined by an equation $y = f(x)$, with $a \le x \le b$, and let the density of the material be coordinate dependent: $\rho = \rho(x, y)$. The mass of this rod is calculated as

$$m = S\int_a^b \rho\bigl(x, f(x)\bigr)\sqrt{1 + [y'(x)]^2}\,dx.$$

If the shape of the rod is defined parametrically by $x = x(t)$ and $y = y(t)$, then its mass is found as

$$m = S\int_a^b \rho\bigl(x(t), y(t)\bigr)\sqrt{[x'(t)]^2 + [y'(t)]^2}\,dt.$$

4. The *coordinates of the center of mass of a plane homogeneous material curve* whose shape is defined by an equation $y = f(x)$, with $a \le x \le b$, are calculated by the formulas

$$x_{\rm c} = \frac{1}{L}\int_a^b x\sqrt{1 + [y'(x)]^2}\,dx, \qquad y_{\rm c} = \frac{1}{L}\int_a^b f(x)\sqrt{1 + [y'(x)]^2}\,dx,$$

where L is the length of the curve.

If the shape of a plane homogeneous material curve is defined parametrically by $x = x(t)$ and $y = y(t)$, then the coordinates of its center of mass are obtained as

$$x_{\rm c} = \frac{1}{L}\int_a^b x(t)\sqrt{[x'(t)]^2 + [y'(t)]^2}\,dt, \qquad y_{\rm c} = \frac{1}{L}\int_a^b y(t)\sqrt{[x'(t)]^2 + [y'(t)]^2}\,dt.$$

5. The *coordinates of the center of mass of a homogeneous curvilinear trapezoid* bounded by a curve $y = f(x)$, the x-axis, and the straight lines $x = a$ and $x = b$ (see Fig. M7.1) are given by

$$x_{\rm c} = \frac{1}{S}\int_a^b xf(x)\,dx, \quad y_{\rm c} = \frac{1}{2S}\int_a^b [f(x)]^2\,dx, \qquad S = \int_a^b f(x)\,dx,$$

where S is the area of the trapezoid.

M7.2.6. Improper Integrals with Infinite Integration Limits

An improper integral is an integral with an infinite limit (limits) of integration or an integral of an unbounded function.

► Integrals with infinite limits.

1°. Let $y = f(x)$ be a function defined and continuous on an infinite interval $a \le x < \infty$. If there exists a finite limit $\lim\limits_{b\to\infty}\int_a^b f(x)\,dx$, then it is called a (convergent) *improper integral* of $f(x)$ on the interval $[a, \infty)$ and is denoted

$\int_a^\infty f(x)\,dx$. Thus, by definition

$$\int_a^\infty f(x)\,dx = \lim_{b\to\infty} \int_a^b f(x)\,dx. \qquad \text{(M7.2.6.1)}$$

If the limit is infinite or does not exist, the improper integral is called *divergent*.

The *geometric meaning of an improper integral* is that the integral $\int_a^\infty f(x)\,dx$, with $f(x) \ge 0$, is equal to the area of the unbounded domain between the curve $y = f(x)$, its asymptote $y = 0$, and the straight line $x = a$ on the left.

2°. Suppose an antiderivative $F(x)$ of the integrand function $f(x)$ is known. Then the improper integral (M7.2.6.1) is

(i) *convergent* if there exists a finite limit $\lim_{x\to\infty} F(x) = F(\infty)$;

(ii) *divergent* if the limit is infinite or does not exist.

In case (i), we have

$$\int_a^\infty f(x)\,dx = F(x)\Big|_a^\infty = F(\infty) - F(a).$$

Example 1. Let us investigate the improper integral $I = \int_a^\infty \frac{dx}{x^\lambda}$, $a > 0$.

The integrand $f(x) = x^{-\lambda}$ has an antiderivative $F(x) = \frac{1}{1-\lambda} x^{1-\lambda}$ if $\lambda \ne 1$. Depending on the value of the parameter λ, we have

$$\lim_{x\to\infty} F(x) = \frac{1}{1-\lambda} \lim_{x\to\infty} x^{1-\lambda} = \begin{cases} 0 & \text{if } \lambda > 1, \\ \infty & \text{if } \lambda < 1. \end{cases}$$

Therefore, if $\lambda > 1$, the integral is convergent and is equal to $I = F(\infty) - F(a) = \frac{a^{1-\lambda}}{\lambda - 1}$, and if $\lambda < 1$, the integral is divergent. It is easy to show that the integral is also divergent if $\lambda = 1$.

3°. Improper integrals for other infinite intervals are defined in a similar way:

$$\int_{-\infty}^b f(x)\,dx = \lim_{a\to-\infty} \int_a^b f(x)\,dx,$$
$$\int_{-\infty}^\infty f(x)\,dx = \int_{-\infty}^c f(x)\,dx + \int_c^\infty f(x)\,dx,$$

where c is an arbitrary number. Note that if either improper integral on the right-hand side of the latter relation is convergent, then, by definition, the integral on the left-hand side is also convergent. If at least one of the integrals on the right-hand side is divergent, the integral on the left is called divergent.

4°. Properties 2–4 and 6–9 from Subsection M7.2.2, where a can be equal to $-\infty$ and b can be ∞, apply to improper integrals as well; it is assumed that all quantities on the right-hand sides exist (the integrals are convergent).

► **Sufficient conditions for convergence of improper integrals.** In many problems, it suffices to establish whether a given improper integral is convergent or not and, if yes, evaluate it. The theorems presented below can be useful in doing so.

THEOREM 1 (CAUCHY'S CONVERGENCE CRITERION). *For the integral (M7.2.6.1) to be convergent it is necessary and sufficient that for any* $\varepsilon > 0$ *there exist a number* R *such that the inequality*

$$\left|\int_{\alpha}^{\beta} f(x)\,dx\right| < \varepsilon$$

holds for any $\beta > \alpha > R$.

THEOREM 2. *If* $0 \le f(x) \le g(x)$ *for* $x \ge a$, *then the convergence of the integral* $\int_a^\infty g(x)\,dx$ *implies the convergence of the integral* $\int_a^\infty f(x)\,dx$*; moreover,* $\int_a^\infty f(x)\,dx \le \int_a^\infty g(x)\,dx$. *If the integral* $\int_a^\infty f(x)\,dx$ *is divergent, then the integral* $\int_a^\infty g(x)\,dx$ *is also divergent.*

THEOREM 3. *If the integral* $\int_a^\infty |f(x)|\,dx$ *is convergent, then the integral* $\int_a^\infty f(x)\,dx$ *is also convergent; in this case, the latter integral is called absolutely convergent.*

Example 2. The improper integral $\int_1^\infty \frac{\sin x}{x^2}\,dx$ is absolutely convergent, since $\left|\frac{\sin x}{x^2}\right| \le \frac{1}{x^2}$ and the integral $\int_1^\infty \frac{1}{x^2}\,dx$ is convergent (see Example 1).

THEOREM 4. *Let* $f(x)$ *and* $g(x)$ *be integrable functions on any finite interval* $a \le x \le b$ *and let there exist a limit, finite or infinite,*

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = K.$$

Then the following assertions hold:

1. If $0 < K < \infty$, *both integrals*

$$\int_a^\infty f(x)\,dx, \qquad \int_a^\infty g(x)\,dx \tag{M7.2.6.2}$$

are convergent or divergent simultaneously.

2. If $0 \le K < \infty$, *the convergence of the latter integral in (M7.2.6.2) implies the convergence of the former integral.*

3. If $0 < K \le \infty$, *the divergence of the latter integral in (M7.2.6.2) implies the divergence of the former integral.*

THEOREM 5 (COROLLARY OF THEOREM 4). *Given a function* $f(x)$, *let its asymptotics for sufficiently large* x *have the form*

$$f(x) = \frac{\varphi(x)}{x^\lambda} \qquad (\lambda > 0).$$

Then: (i) if $\lambda > 1$ *and* $\varphi(x) \le c < \infty$, *then the integral* $\int_a^\infty f(x)\,dx$ *is convergent; (ii) if* $\lambda \le 1$ *and* $\varphi(x) \ge c > 0$, *then the integral is divergent.*

THEOREM 6. *Let* $f(x)$ *be an absolutely integrable function on an interval* $[a, \infty)$ *and let* $g(x)$ *be a bounded function on* $[a, \infty)$. *Then the product* $f(x)g(x)$ *is an absolutely integrable function on* $[a, \infty)$.

THEOREM 7 (ANALOGUE OF ABEL'S TEST FOR CONVERGENCE OF INFINITE SERIES). *Let* $f(x)$ *be an integrable function on an interval* $[a, \infty)$ *such that the integral (M7.2.6.1) is* *convergent (maybe not absolutely) and let $g(x)$ be a monotonic and bounded function on $[a,\infty)$. Then the integral*

$$\int_a^\infty f(x)g(x)\,dx \qquad (M7.2.6.3)$$

is convergent.

THEOREM 8 (ANALOGUE OF DIRICHLET'S TEST FOR CONVERGENCE OF INFINITE SERIES). *Let (i) $f(x)$ be an integrable function on any finite interval $[a,A]$ and*

$$\left|\int_a^A f(x)\,dx\right| \le K < \infty \qquad (a \le A < \infty);$$

(ii) $g(x)$ be a function tending to zero monotonically as $x \to \infty$: $\lim\limits_{x\to\infty} g(x) = 0$. Then the integral (M7.2.6.3) is convergent.

Example 3. Let us show that the improper integral $\displaystyle\int_a^\infty \frac{\sin x}{x^\lambda}\,dx$ is convergent for $a > 0$ and $\lambda > 0$.

Set $f(x) = \sin x$ and $g(x) = x^{-\lambda}$ and verify conditions (i) and (ii) of Theorem 8. We have

(i) $\displaystyle\left|\int_a^A \sin x\,dx\right| = |\cos a - \cos A| \le 2;$

(ii) since $\lambda > 0$, the function $x^{-\lambda}$ is monotonically decreasing and tends to zero as $x \to \infty$.

So both conditions of Theorem 8 are met, and therefore the given improper integral is convergent.

M7.2.7. Improper Integrals of Unbounded Functions

► Basic definitions.

1°. Let a function $f(x)$ be defined and continuous for $a \le x < b$, but $\lim\limits_{x\to b-0} f(x) = \infty$. If there exists a finite limit $\lim\limits_{\lambda\to b-0}\int_a^\lambda f(x)\,dx$, it is called the (convergent) *improper integral of the unbounded function* $f(x)$ over the interval $[a,b]$. Thus, by definition

$$\int_a^b f(x)\,dx = \lim_{\lambda\to b-0}\int_a^\lambda f(x)\,dx. \qquad \text{(M7.2.7.1)}$$

If no finite limit exists, the integral is called *divergent*.

If $\lim\limits_{x\to a+0} f(x) = \infty$, then, by definition, it is assumed that

$$\int_a^b f(x)\,dx = \lim_{\gamma\to a+0}\int_\gamma^b f(x)\,dx.$$

Finally, if $f(x)$ is unbounded near a point $c \in (a,b)$ and both integrals $\int_a^c f(x)\,dx$ and $\int_c^b f(x)\,dx$ are convergent, then, by definition,

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

If at least one of the integrals on the right-hand side is divergent, the integral on the left-hand side is called divergent.

2°. The geometric meaning of an improper integral of an unbounded function and also sufficient conditions for convergence of such integrals are similar to those for improper integrals with infinite limit(s).

► **Convergence tests for improper integrals of unbounded functions.** Presented below are theorems for the case where the only singular point of the integrand function is the right endpoint of the interval $[a, b]$.

THEOREM 1 (CAUCHY'S CONVERGENCE CRITERION). *For the integral (M7.2.7.1) to be convergent it is necessary and sufficient that for any* $\varepsilon > 0$ *there exist a number* $\delta > 0$ *such that for any* δ_1 *and* δ_2 *satisfying* $0 < \delta_1 < \delta$ *and* $0 < \delta_2 < \delta$ *the following inequality holds:*

$$\left|\int_{b-\delta_1}^{b-\delta_2} f(x)\,dx\right| < \varepsilon.$$

THEOREM 2. *If* $0 \le f(x) \le g(x)$ *for* $a \le x < b$, *then the convergence of the integral* $\int_a^b g(x)\,dx$ *implies the convergence of the integral* $\int_a^b f(x)\,dx$, *with* $\int_a^b f(x)\,dx \le \int_a^b g(x)\,dx$. *If the integral* $\int_a^b f(x)\,dx$ *is divergent, then the integral* $\int_a^b g(x)\,dx$ *is also divergent.*

Example. For any continuous function $\varphi(x)$ such that $\varphi(1) = 0$, the improper integral $\int_0^1 \frac{dx}{\varphi^2(x) + \sqrt{1-x}}$ is convergent and does not exceed 2, since $\frac{1}{\varphi^2(x) + \sqrt{1-x}} < \frac{1}{\sqrt{1-x}}$, while the integral $\int_0^1 \frac{dx}{\sqrt{1-x}}$ is convergent and is equal to 2.

THEOREM 3. *Let* $f(x)$ *and* $g(x)$ *be continuous functions on* $[a, b)$ *and let the following limit exist:*

$$\lim_{x\to b} \frac{f(x)}{g(x)} = K \qquad (0 < K < \infty).$$

Then both integrals

$$\int_a^b f(x)\,dx, \qquad \int_a^b g(x)\,dx$$

are either convergent or divergent simultaneously.

THEOREM 4. *Let a function* $f(x)$ *be representable in the form*

$$f(x) = \frac{\varphi(x)}{(b-x)^\lambda} \qquad (\lambda > 0),$$

where $\varphi(x)$ *is continuous on* $[a, b]$ *and the condition* $\varphi(b) \ne 0$ *holds.*

Then: (i) if $\lambda < 1$ *and* $\varphi(x) \le c < \infty$, *then the integral* $\int_a^b f(x)\,dx$ *is convergent; (ii) if* $\lambda \ge 1$ *and* $\varphi(x) \ge c > 0$, *then this integral is divergent.*

M7.2.8. Approximate (Numerical) Methods for Computation of Definite Integrals

For approximate computation of an integral like $\int_a^b f(x)\,dx$, let us break up the interval $[a, b]$ into n equal subintervals with length $h = \frac{b-a}{n}$. Introduce the notation: $x_0 = a$, x_1, $\ldots$, $x_n = b$ (the partition points), $y_i = f(x_i)$, $i = 0, 1, \ldots, n$.

1°. *Rectangle rules*:

$$\int_a^b f(x)\,dx \approx h(y_0 + y_1 + \cdots + y_{n-1}),$$

$$\int_a^b f(x)\,dx \approx h(y_1 + y_2 + \cdots + y_n).$$

The error of these formulas, R_n, is proportional to h and is estimated using the inequality

$$|R_n| \le \tfrac{1}{2}h(b-a)M_1, \qquad M_1 = \max_{a\le x\le b}\bigl|f'(x)\bigr|.$$

2°. *Trapezoidal rule*:

$$\int_a^b f(x)\,dx \approx h\Bigl(\frac{y_0 + y_n}{2} + y_1 + y_2 + \cdots + y_{n-1}\Bigr).$$

The error of this formula is proportional to h^2 and is estimated as

$$|R_n| \le \tfrac{1}{12}h^2(b-a)M_2, \qquad M_2 = \max_{a\le x\le b}\bigl|f''(x)\bigr|.$$

3°. *Simpson's rule*:

$$\int_a^b f(x)\,dx \approx \tfrac{1}{3}h[y_0 + y_n + 4(y_1 + y_3 + \cdots + y_{n-1}) + 2(y_2 + y_4 + \cdots + y_{n-2})],$$

where n is even. The error of approximation by Simpson's rule is proportional to h^4:

$$|R_n| \le \tfrac{1}{180}h^4(b-a)M_4, \qquad M_4 = \max_{a\le x\le b}\bigl|f^{(4)}(x)\bigr|.$$

Simpson's rule yields exact results for the case where the integrand function is a polynomial of degree two or three.

M7.3. Double and Triple Integrals

M7.3.1. Definition and Properties of the Double Integral

► **Definition and properties of the double integral.** Suppose there is a bounded set of points defined on the plane, so that it can be placed in a minimal enclosing circle. The diameter of this circle is called the *diameter of the set*. Consider a domain D in the xy-plane. Let us partition D into n nonintersecting subdomains (cells). The largest of the cell diameters is called the *partition diameter* and is denoted $\lambda = \lambda(\mathcal{D}_n)$, where $\mathcal{D}_n$ stands for the partition of the domain D into cells. Let a function $z = f(x, y)$ be defined in D. Select an arbitrary point in each cell (x_i, y_i), $i = 1, 2, \dots, n$, and make up an *integral sum*,

$$s_n = \sum_{i=1}^{n} f(x_i, y_i)\,\Delta S_i,$$

where ΔS_i is the area of the ith subdomain.

If there exists a finite limit, $\mathcal{J}$, of the sums s_n as $\lambda \to 0$ and it depends on neither the partition $\mathcal{D}_n$ nor the selection of the points (x_i, y_i), this limit is denoted $\iint_D f(x, y)\,dx\,dy$ and is called the double integral of the function $f(x, y)$ over the domain D:

$$\iint_D f(x, y)\,dx\,dy = \lim_{\lambda\to 0} s_n.$$

This means that for any $\varepsilon > 0$ there exists a $\delta > 0$ such that for all partitions $\mathcal{D}_n$ such that $\lambda(\mathcal{D}_n) < \delta$ and for any selection of the points (x_i, y_i), the inequality $|s_n - \mathcal{J}| < \varepsilon$ holds. In this case, the function $f(x, y)$ is called integrable over the domain D.

► **Classes of integrable functions.** Further on, it is assumed that D is a closed bounded domain.

1. If $f(x,y)$ is continuous in D, then the double integral $\iint_D f(x,y)\,dx\,dy$ exists.

2. If $f(x,y)$ is bounded and the set of points of discontinuity of $f(x,y)$ has a zero area (e.g., the points of discontinuity lie on finitely many continuous curves in the xy-plane), then the double integral of $f(x,y)$ over the domain D exists.

► **Properties of the double integral.**

1. *Linearity*. If functions $f(x,y)$ and $g(x,y)$ are integrable in D, then

$$\iint_D \bigl[af(x,y) \pm bg(x,y)\bigr]\,dx\,dy = a\iint_D f(x,y)\,dx\,dy \pm b\iint_D g(x,y)\,dx\,dy,$$

where a and b are any numbers.

2. *Additivity*. If the domain D is split into two subdomains D_1 and D_2 that do not have common internal points and if the function $f(x,y)$ is integrable in either subdomain, then

$$\iint_D f(x,y)\,dx\,dy = \iint_{D_1} f(x,y)\,dx\,dy + \iint_{D_2} f(x,y)\,dx\,dy.$$

3. *Estimation theorem*. If $m \le f(x,y) \le M$ in D, then

$$mS \le \iint_D f(x,y)\,dx\,dy \le MS,$$

where S is the area of the domain D.

4. *Mean value theorem*. If $f(x,y)$ is continuous in D, then there exists at least one internal point $(\bar{x},\bar{y}) \in D$ such that

$$\iint_D f(x,y)\,dx\,dy = f(\bar{x},\bar{y})\,S.$$

The number $f(\bar{x},\bar{y})$ is called the *mean value of the function* $f(x,y)$ in D.

5. *Integration of inequalities*. If $\varphi(x,y) \le f(x,y) \le g(x,y)$ in D, then

$$\iint_D \varphi(x,y)\,dx\,dy \le \iint_D f(x,y)\,dx\,dy \le \iint_D g(x,y)\,dx\,dy.$$

In particular, if $f(x,y) \ge 0$ in D, then $\iint_D f(x,y)\,dx\,dy \ge 0$.

6. *Absolute value theorem*

$$\left|\iint_D f(x,y)\,dx\,dy\right| \le \iint_D \bigl|f(x,y)\bigr|\,dx\,dy.$$

► **Geometric meaning of the double integral.** Let a function $f(x,y)$ be nonnegative in D. Then the double integral $\iint_D f(x,y)\,dx\,dy$ is equal to the volume of a cylindrical body with base D in the plane $z=0$ and bounded from above by the surface $z=f(x,y)$; see Fig. M7.4.

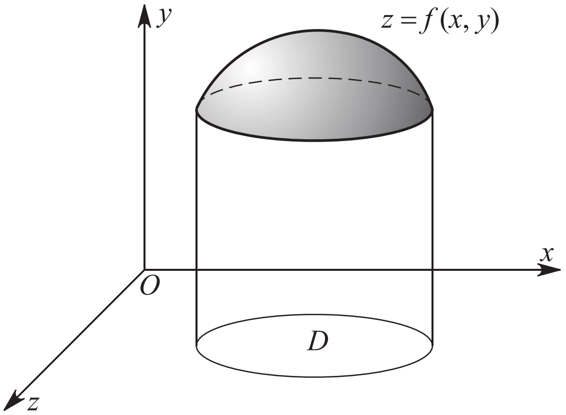

Figure M7.4. A double integral of a nonnegative function $f(x,y)$ over a domain D is equal to the volume of a cylindrical body with base D in the plane $z = 0$ and bounded from above by the surface $z = f(x,y)$.

M7.3.2. Computation of the Double Integral

► Use of iterated integrals.

1°. If a domain D is defined in the xy-plane by the inequalities $a \le x \le b$ and $y_1(x) \le y \le y_2(x)$ (see Fig. M7.5 a), then*

$$\iint_D f(x,y)\,dx\,dy = \int_a^b dx \int_{y_1(x)}^{y_2(x)} f(x,y)\,dy. \qquad \text{(M7.3.2.1)}$$

The expression on the right-hand side is called an *iterated integral*. Note that the variable x in the inner integral is considered constant when integrating.

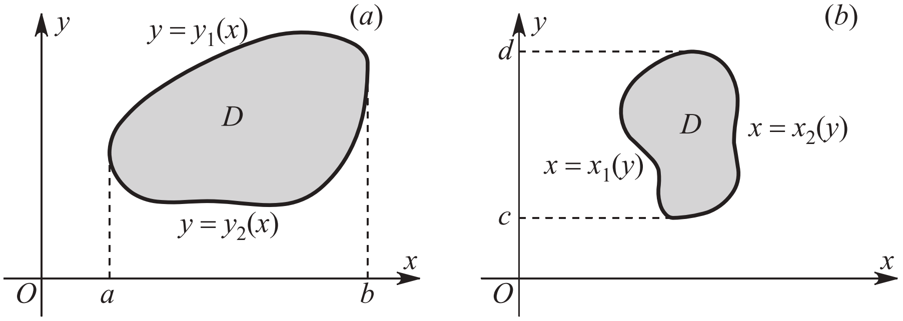

Figure M7.5. Computation of a double integral using iterated integrals: (a) illustration to formula (M7.3.2.1), (b) illustration to formula (M7.3.2.2).

2°. If $D = \{c \le y \le d,\ x_1(y) \le x \le x_2(y)\}$ (see Fig. M7.5 b), then

$$\iint_D f(x,y)\,dx\,dy = \int_c^d dy \int_{x_1(y)}^{x_2(y)} f(x,y)\,dx. \qquad \text{(M7.3.2.2)}$$

Example 1. Compute the integral

$$I = \iint_D \frac{dx\,dy}{(ax+by)^2},$$

where $D = \{0 \le x \le 1,\ 1 \le y \le 3\}$ is a rectangle, $a > 0$, and $b > 0$.

---

* It is assumed that in (M7.3.2.1) and (M7.3.2.2) the double integral on the left-hand side and the inner integral on the right-hand side exist.

Using formula (M7.3.2.2), we get

$$\iint_D \frac{dx\,dy}{(ax+by)^2} = \int_1^3 dy \int_0^1 \frac{dx}{(ax+by)^2}.$$

Compute the inner integral:

$$\int_0^1 \frac{dx}{(ax+by)^2} = -\frac{1}{a(ax+by)}\bigg|_{x=0}^{x=1} = \frac{1}{a}\left(\frac{1}{by} - \frac{1}{by+a}\right).$$

It follows that

$$I = \frac{1}{a}\int_1^3 \left(\frac{1}{by} - \frac{1}{by+a}\right) dy = \frac{1}{ab} \ln \frac{3(a+b)}{a+3b}.$$

3°. Consider a domain D inscribed in a rectangle $\{a \le x \le b,\ c \le y \le d\}$. Let the boundary of D, within the rectangle, be intersected by straight lines parallel to the coordinate axes at two points only, as shown in Fig. M7.6 *a*. Then, by comparing formulas (M7.3.2.1) and (M7.3.2.2), we arrive at the relation

$$\int_a^b dx \int_{y_1(x)}^{y_2(x)} f(x,y)\,dy = \int_c^d dy \int_{x_1(y)}^{x_2(y)} f(x,y)\,dx,$$

which shows how the order of integration can be changed.

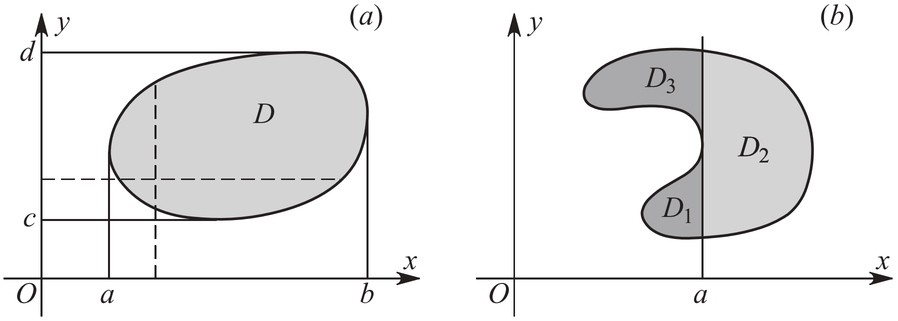

Figure M7.6. Illustrations to the computation of a double integral in a simple (*a*) and a complex (*b*) domain.

4°. In the general case, the domain D is first split into subdomains considered in Items 1° and 2°, and then the property of additivity of the double integral is used. For example, the domain D shown in Fig. M7.6 *b* is divided by the straight line $x = a$ into three subdomains D_1, D_2, and D_3. Then the integral over D is represented as the sum of three integrals over the resulting subdomains.

► Change of variables in the double integral.

1°. Let $x = x(u,v)$ and $y = y(u,v)$ be continuously differentiable functions that map one-to-one a domain D_1 in the uv-plane onto a domain D in the xy-plane, and let $f(x,y)$ be a continuous function in D. Then

$$\iint_D f(x,y)\,dx\,dy = \iint_{D_1} f\bigl(x(u,v),\ y(u,v)\bigr)\,|J(u,v)|\,du\,dv,$$

where $J(u,v)$ is the *Jacobian* (or *Jacobian determinant*) of the mapping of D_1 onto D:

$$J(u,v) = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}.$$

The fraction before the determinant is a common notation for a Jacobian.

The absolute value of the Jacobian characterizes the extension (contraction) of an infinitesimal area element when passing from x, y to u, v.

2°. The Jacobian of the mapping defining the change from the Cartesian coordinates x, y to the polar coordinates ρ, φ,

$$x = \rho\cos\varphi, \qquad y = \rho\sin\varphi, \tag{M7.3.2.3}$$

is equal to

$$J(\rho,\varphi) = \rho. \tag{M7.3.2.4}$$

Example 2. Given a sphere of radius R and a right circular cylinder of radius $a < R$ whose axis passes through the sphere center, find the volume of the figure the cylinder cuts out of the sphere.

The volume of this figure is calculated as

$$V = 2\iint_{x^2+y^2\le a^2} \sqrt{R^2 - x^2 - y^2}\, dx\, dy.$$

Passing in the integral from x, y to the polar coordinates (M7.3.2.3) and taking into account (M7.3.2.4), we obtain

$$V = 2\int_0^{2\pi}\int_0^a \sqrt{R^2-\rho^2}\,\rho\, d\rho\, d\varphi = \frac{4\pi}{3}\bigl[R^3 - (R^2-a^2)^{3/2}\bigr].$$

M7.3.3. Geometric and Physical Applications of the Double Integral

► Geometric applications of the double integral.

1. *Area of a domain D in the xy-plane*:

$$S = \iint_D dx\, dy.$$

2. *Area of a surface defined by an equation $z = f(x,y)$* with $(x,y) \in D$ (the surface is projected onto a domain D in the xy-plane):

$$S = \iint_D \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + 1}\; dx\, dy.$$

3. *Calculation of volumes*. If a domain U of the three-dimensional space is defined by $\bigl\{(x,y) \in D,\ f(x,y) \le z \le g(x,y)\bigr\}$, where D is a domain in the xy-plane, the volume of U is calculated as

$$V = \iint_D \bigl[g(x,y) - f(x,y)\bigr]\, dx\, dy.$$

The three-dimensional domain U is a cylinder with base D bounded by the surface $z = f(x,y)$ from below and the surface $z = g(x,y)$ from above. The lateral surface of this body consists of segments of straight lines parallel to the z-axis.

4. *Area of a surface defined parametrically by equations $x = x(u,v)$, $y = y(u,v)$, $z = z(u,v)$*, with $(u,v) \in D_1$:

$$S = \iint_{D_1} \sqrt{EG - F^2}\, du\, dv.$$

Notation used:

$$E = \left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2 + \left(\frac{\partial z}{\partial u}\right)^2,$$

$$G = \left(\frac{\partial x}{\partial v}\right)^2 + \left(\frac{\partial y}{\partial v}\right)^2 + \left(\frac{\partial z}{\partial v}\right)^2,$$

$$F = \frac{\partial x}{\partial u}\frac{\partial x}{\partial v} + \frac{\partial y}{\partial u}\frac{\partial y}{\partial v} + \frac{\partial z}{\partial u}\frac{\partial z}{\partial v}.$$

5. *Area of a surface defined by a vector equation* $\mathbf{r} = \mathbf{r}(u, v) = x(u, v)\,\mathbf{i} + y(u, v)\,\mathbf{j} + z(u, v)\,\mathbf{k}$, with $(u, v) \in D_1$:

$$S = \iint_{D_1} |\mathbf{n}(u, v)|\,du\,dv,$$

where $\mathbf{n}(u, v) = \mathbf{r}_u \times \mathbf{r}_v$ is a normal vector to the surface; the subscripts u and v denote the respective partial derivatives.

Remark. The formulas from Items 4 and 5 are equivalent—they define one and the same surface in two forms, scalar and vector, respectively.

► **Physical applications of the double integral.** Consider a flat plate that occupies a domain D in the xy-plane. Let $\gamma(x, y)$ be the surface density of the plate material (the case $\gamma = \text{const}$ corresponds to a homogeneous plate).

1. *Mass of a flat plate*:

$$m = \iint_D \gamma(x, y)\,dx\,dy.$$

2. *Coordinates of the center of mass of a flat plate*:

$$x_{\rm c} = \frac{1}{m} \iint_D x\gamma(x, y)\,dx\,dy, \qquad y_{\rm c} = \frac{1}{m} \iint_D y\gamma(x, y)\,dx\,dy,$$

where m is the mass of the plate.

3. *Moments of inertia of a flat plate* about the coordinate axes:

$$I_x = \iint_D y^2\gamma(x, y)\,dx\,dy, \qquad I_y = \iint_D x^2\gamma(x, y)\,dx\,dy.$$

The moment of inertia of the plate about the origin of coordinates is calculated as $I_0 = I_x + I_y$.

M7.3.4. Definition and Properties of the Triple Integral

► **Definition of the triple integral.** Let a function $f(x, y, z)$ be defined in a domain U of the three-dimensional space. Let us break up U into n subdomains (cells) that do not have common internal points. Denote by $\lambda = \lambda(\mathcal{U}_n)$ the *diameter* of the resulting partition $\mathcal{U}_n$, i.e., the maximum of the cell diameters (the diameter of a domain in space is the diameter of the minimal sphere enclosing the domain). Select an arbitrary point, (x_i, y_i, z_i), $i = 1,\ 2,\ \dots,\ n$, in each cell and make up an *integral sum*

$$s_n = \sum_{i=1}^{n} f(x_i, y_i, z_i)\,\Delta V_i,$$

where ΔV_i is the volume of the ith cell. If there exists a finite limit of the sums s_n as $\lambda(\mathcal{U}_n) \to 0$ that depends on neither the partition $\mathcal{U}_n$ nor the selection of the points (x_i, y_i, z_i), then it is called the triple integral of the function $f(x, y, z)$ over the domain U and is denoted

$$\iiint_U f(x, y, z)\,dx\,dy\,dz = \lim_{\lambda \to 0} s_n.$$

► **Properties of the triple integral.** The properties of triple integrals are similar to those of double integrals.

1. *Linearity*. If functions $f(x,y,z)$ and $g(x,y,z)$ are integrable in a domain U, then

$$\iiint_U \bigl[af(x,y,z) \pm bg(x,y,z)\bigr]\,dx\,dy\,dz = a\iiint_U f(x,y,z)\,dx\,dy\,dz \pm b\iiint_U g(x,y,z)\,dx\,dy\,dz,$$

where a and b are any numbers.

2. *Additivity*. If a domain U is split into two subdomains, U_1 and U_2, that do not have common internal points and if a function $f(x,y,z)$ is integrable in either subdomain, then

$$\iiint_U f(x,y,z)\,dx\,dy\,dz = \iiint_{U_1} f(x,y,z)\,dx\,dy\,dz + \iiint_{U_2} f(x,y,z)\,dx\,dy\,dz.$$

3. *Estimation theorem*. If $m \le f(x,y,z) \le M$ in a domain U, then

$$mV \le \iiint_U f(x,y,z)\,dx\,dy\,dz \le MV,$$

where V is the volume of U.

4. *Mean value theorem*. If $f(x,y,z)$ is continuous in U, then there exists at least one internal point $(\bar{x},\bar{y},\bar{z}) \in U$ such that

$$\iiint_U f(x,y,z)\,dx\,dy\,dz = f(\bar{x},\bar{y},\bar{z})\,V.$$

The number $f(\bar{x},\bar{y},\bar{z})$ is called the *mean value of the function* f in the domain U.

5. *Integration of inequalities*. If $\varphi(x,y,z) \le f(x,y,z) \le g(x,y,z)$ in a domain U, then

$$\iiint_U \varphi(x,y,z)\,dx\,dy\,dz \le \iiint_U f(x,y,z)\,dx\,dy\,dz \le \iiint_U g(x,y,z)\,dx\,dy\,dz.$$

6. *Absolute value theorem*:

$$\left|\iiint_U f(x,y,z)\,dx\,dy\,dz\right| \le \iiint_U \bigl|f(x,y,z)\bigr|\,dx\,dy\,dz.$$

M7.3.5. Computation of the Triple Integral. Some Applications. Iterated Integrals and Asymptotic Formulas

► **Use of iterated integrals.**

1°. Consider a three-dimensional body U bounded by a surface $z = g(x,y)$ from above and a surface $z = h(x,y)$ from below, with a domain D being the projection of the body onto the xy-plane. In other words, the domain U is defined as $\{(x,y) \in D : h(x,y) \le z \le g(x,y)\}$. Then

$$\iiint_U f(x,y,z)\,dx\,dy\,dz = \iint_D dx\,dy \int_{h(x,y)}^{g(x,y)} f(x,y,z)\,dz.$$

2°. If, under the same conditions as in Item 1°, the domain D of the xy-plane is defined as $\{a \le x \le b,\ y_1(x) \le y \le y_2(x)\}$, then

$$\iiint_U f(x,y,z)\,dx\,dy\,dz = \int_a^b dx \int_{y_1(x)}^{y_2(x)} dy \int_{h(x,y)}^{g(x,y)} f(x,y,z)\,dz.$$

► Change of variables in the triple integral.

1°. Let $x = x(u, v, w)$, $y = y(u, v, w)$, and $z = z(u, v, w)$ be continuously differentiable functions that map, one to one, a domain Ω of the u, v, w space onto a domain U of the x, y, z space, and let a function $f(x, y, z)$ be continuous in U. Then

$$\iiint_U f(x,y,z)\,dx\,dy\,dz = \iiint_\Omega f\bigl(x(u,v,w),\ y(u,v,w),\ z(u,v,w)\bigr)\,|J(u,v,w)|\,du\,dv\,dw,$$

where $J(u, v, w)$ is the *Jacobian* of the mapping of Ω onto U:

$$J(u,v,w) = \frac{\partial(x,y,z)}{\partial(u,v,w)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix}.$$

The expression in the middle is a common notation for a Jacobian.

The absolute value of the Jacobian characterizes the expansion (or contraction) of an infinitesimal volume element when passing from x, y, z to u, v, w.

2°. The Jacobians of some common transformations in space are listed in Table M7.2.

TABLE M7.2
Some curvilinear coordinates in space and the respective Jacobians.

Name of coordinates	Transformation	Jacobian, J
Cylindrical coordinates ρ, φ, z	$x = \rho\cos\varphi,\ \ y = \rho\sin\varphi,\ \ z = z$	ρ
Generalized cylindrical coordinates ρ, φ, z	$x = a\rho\cos\varphi,\ \ y = b\rho\sin\varphi,\ \ z = z$	$ab\rho$
Spherical coordinates r, φ, θ	$x = r\cos\varphi\,\sin\theta,\ y = r\sin\varphi\,\sin\theta, z = r\cos\theta$	$r^2\sin\theta$
Generalized spherical coordinates r, φ, θ	$x = ar\cos\varphi\,\sin\theta,\ y = br\sin\varphi\,\sin\theta,$ $z = cr\cos\theta$	$abcr^2\sin\theta$
Parabolic cylinder coordinates σ, τ, z	$x = \sigma\tau,\ \ y = \frac{1}{2}(\tau^2 - \sigma^2),\ \ z = z$	$\sigma^2 + \tau^2$

► Some geometric and physical applications of the triple integral.

1. *Volume of a domain* U:

$$V = \iiint_U dx\,dy\,dz.$$

2. *Mass of a body of variable density* $\gamma = \gamma(x, y, z)$ occupying a domain U:

$$m = \iiint_U \gamma\,dx\,dy\,dz.$$

3. *Coordinates of the center of mass*:

$$x_{\rm c} = \frac{1}{m}\iiint_U x\gamma\,dx\,dy\,dz, \quad y_{\rm c} = \frac{1}{m}\iiint_U y\gamma\,dx\,dy\,dz, \quad z_{\rm c} = \frac{1}{m}\iiint_U z\gamma\,dx\,dy\,dz.$$

4. *Moments of inertia* about the coordinate axes:

$$I_x = \iiint_U \rho_{yz}^2\gamma\,dx\,dy\,dz, \quad I_y = \iiint_U \rho_{xz}^2\gamma\,dx\,dy\,dz, \quad I_z = \iiint_U \rho_{xy}^2\gamma\,dx\,dy\,dz,$$

where $\rho_{yz}^2 = y^2 + z^2$, $\rho_{xz}^2 = x^2 + z^2$, and $\rho_{xy}^2 = x^2 + y^2$.

If the body is homogeneous, then $\gamma = \text{const}$.

Example. Given a bounded homogeneous elliptic cylinder,

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1,\qquad 0\le z\le h,$$

find its moment of inertia about the z-axis.

Using the generalized cylindrical coordinates (see the second row in Table M7.2), we obtain

$$\begin{aligned}I_x&=\gamma\iiint_U(x^2+y^2)\,dx\,dy\,dz=\gamma\int_0^h\int_0^{2\pi}\int_0^1\rho^2(a^2\cos^2\varphi+b^2\sin^2\varphi)ab\rho\,d\rho\,d\varphi\,dz\\&=\frac{1}{4}ab\gamma\int_0^h\int_0^{2\pi}(a^2\cos^2\varphi+b^2\sin^2\varphi)\,d\varphi\,dz=\frac{1}{4}ab\gamma\int_0^{2\pi}\int_0^h(a^2\cos^2\varphi+b^2\sin^2\varphi)\,dz\,d\varphi\\&=\frac{1}{4}abh\gamma\int_0^{2\pi}(a^2\cos^2\varphi+b^2\sin^2\varphi)\,d\varphi=\frac{1}{4}\pi ab(a^2+b^2)h\gamma.\end{aligned}$$

5. *Potential of the gravitational field* of a body U at a point (x,y,z):

$$\Phi(x,y,z)=\iiint_U\gamma(\xi,\eta,\zeta)\frac{d\xi\,d\eta\,d\zeta}{r},\qquad r=\sqrt{(x-\xi)^2+(y-\eta)^2+(z-\zeta)^2},$$

where $\gamma=\gamma(\xi,\eta,\zeta)$ is the body density. A material point of mass m is pulled by the gravitating body U with a force **F**. The projections of **F** onto the x-, y-, and z-axes are given, respectively, by

$$\begin{aligned}F_x&=km\frac{\partial\Phi}{\partial x}=km\iiint_U\gamma(\xi,\eta,\zeta)\frac{\xi-x}{r^3}\,d\xi\,d\eta\,d\zeta,\\F_y&=km\frac{\partial\Phi}{\partial y}=km\iiint_U\gamma(\xi,\eta,\zeta)\frac{\eta-y}{r^3}\,d\xi\,d\eta\,d\zeta,\\F_z&=km\frac{\partial\Phi}{\partial z}=km\iiint_U\gamma(\xi,\eta,\zeta)\frac{\zeta-z}{r^3}\,d\xi\,d\eta\,d\zeta,\end{aligned}$$

where k is the gravitational constant.

M7.4. Line and Surface Integrals

M7.4.1. Line Integral of the First Kind

► **Definition of the line integral of the first kind.** Let a function $f(x,y,z)$ be defined on a piecewise smooth curve $\overset{\smile}{AB}$ in the three-dimensional space $\mathbb{R}^3$. Let the curve $\overset{\smile}{AB}$ be divided into n subcurves by points $A=M_0$, M_1, M_2, …, $M_n=B$, thus defining a partition $\mathcal{L}_n$. The longest of the chords M_0M_1, M_1M_2, …, $M_{n-1}M_n$ is called the *diameter of the partition* $\mathcal{L}_n$ and is denoted $\lambda=\lambda(\mathcal{L}_n)$. Let us select on each arc $\overset{\smile}{M_{i-1}M_i}$ an arbitrary point (x_i,y_i,z_i), $i=1$, 2, …, n, and make up an *integral sum*

$$s_n=\sum_{i=1}^{n}f(x_i,y_i,z_i)\,\Delta l_i,$$

where Δl_i is the length of $\overset{\smile}{M_{i-1}M_i}$.

If there exists a finite limit of the sums s_n as $\lambda(\mathcal{L}_n)\to 0$ that depends on neither the partition $\mathcal{L}_n$ nor the selection of the points (x_i,y_i,z_i), then it is called the *line integral of the first kind* of the function $f(x,y,z)$ over the curve $\overset{\smile}{AB}$ and is denoted

$$\int_{AB}f(x,y,z)\,dl=\lim_{\lambda\to 0}s_n.$$

A line integral is also called a *curvilinear integral* or a *path integral*.

If the function $f(x,y,z)$ is continuous, then the line integral exists. The line integral of the first kind does not depend of the direction the path $\overset{\smile}{AB}$ is traced; its properties are similar to those of the definite integral.

► Computation of the line integral of the first kind.

1. If a plane curve is defined in the form $y = y(x)$, with $x \in [a, b]$, then

$$\int_{AB} f(x,y)\,dl = \int_a^b f\bigl(x, y(x)\bigr)\sqrt{1 + (y'_x)^2}\,dx.$$

2. If a curve $\overset{\smile}{AB}$ is defined in parametric form by equations $x = x(t)$, $y = y(t)$, and $z = z(t)$, with $t \in [\alpha, \beta]$, then

$$\int_{AB} f(x,y,z)\,dl = \int_\alpha^\beta f\bigl(x(t), y(t), z(t)\bigr)\sqrt{(x'_t)^2 + (y'_t)^2 + (z'_t)^2}\,dt. \qquad \text{(M7.4.1.1)}$$

If a function $f(x,y)$ is defined on a plane curve $x = x(t)$, $y = y(t)$, with $t \in [\alpha, \beta]$, one should set $z'_t = 0$ in (M7.4.1.1).

Example. Evaluate the integral $\displaystyle\int_{AB} xy\,dl$, where $\overset{\smile}{AB}$ is a quarter of an ellipse with semiaxes a and b.

Let us write out the equations of the ellipse for the first quadrant in parametric form:

$$x = a\cos t, \qquad y = b\sin t \qquad (0 \le t \le \pi/2).$$

We have $\sqrt{(x'_t)^2 + (y'_t)^2} = \sqrt{a^2\sin^2 t + b^2\cos^2 t}$. To evaluate the integral, we use formula (M7.4.1.1) with $z'_t = 0$:

$$\begin{aligned}\int_{AB} xy\,dl &= \int_0^{\pi/2} (a\cos t)\,(b\sin t)\sqrt{a^2\sin^2 t + b^2\cos^2 t}\,dt\\ &= \frac{ab}{2}\int_0^{\pi/2}\sin 2t\sqrt{\frac{a^2}{2}(1-\cos 2t) + \frac{b^2}{2}(1+\cos 2t)}\,dt = \frac{ab}{4}\int_{-1}^1\sqrt{\frac{a^2+b^2}{2} + \frac{b^2-a^2}{2}u}\,du\\ &= \frac{ab}{4}\,\frac{2}{b^2-a^2}\,\frac{2}{3}\left(\frac{a^2+b^2}{2} + \frac{b^2-a^2}{2}u\right)^{3/2}\Bigg|_{-1}^{1} = \frac{ab}{3}\,\frac{a^2+ab+b^2}{a+b}.\end{aligned}$$

► Applications of the line integral of the first kind.

1. *Length of a curve* $\overset{\smile}{AB}$:

$$L = \int_{AB} dl.$$

2. *Mass of a material curve* $\overset{\smile}{AB}$ with a given line density $\gamma = \gamma(x,y,z)$:

$$m = \int_{AB} \gamma\,dl.$$

3. *Coordinates of the center of mass* of a material curve $\overset{\smile}{AB}$:

$$x_{\rm c} = \frac{1}{m}\int_{AB} x\gamma\,dl, \quad y_{\rm c} = \frac{1}{m}\int_{AB} y\gamma\,dl, \quad z_{\rm c} = \frac{1}{m}\int_{AB} z\gamma\,dl.$$

To a material line with uniform density there corresponds $\gamma = \text{const}$.

M7.4.2. Line Integral of the Second Kind

► **Definition of the line integral of the second kind.** Let a vector field

$$\mathbf{a}(x,y,z) = P(x,y,z)\,\mathbf{i} + Q(x,y,z)\,\mathbf{j} + R(x,y,z)\,\mathbf{k}$$

and a piecewise smooth curve $\smile{AB}$ be defined in some domain in $\mathbb{R}^3$. By dividing the curve by points $A = M_0$, M_1, M_2, $\ldots$, $M_n = B$ into n subcurves, we obtain a partition $\mathcal{L}_n$. Let us select on each arc $\overset{\smile}{M_{i-1}M_i}$ an arbitrary point (x_i, y_i, z_i), $i = 1, 2, \ldots, n$, and make up a sum of dot products

$$s_n = \sum_{i=1}^{n} \mathbf{a}(x_i, y_i, z_i) \cdot \overrightarrow{M_{i-1}M_i},$$

called an *integral sum*.

If there exists a finite limit of the sums s_n as $\lambda(\mathcal{L}_n) \to 0$ (λ is the diameter of the partition; see Subsection M7.4.1) that depends on neither the partition $\mathcal{L}_n$ nor the selection of the points (x_i, y_i, z_i), then it is called the *line integral of the second kind* of the vector field $\mathbf{a}(x,y,z)$ along the curve $\overset{\smile}{AB}$ and is denoted

$$\int_{AB} \mathbf{a}\cdot d\mathbf{r}, \quad \text{or} \quad \int_{AB} P\,dx + Q\,dy + R\,dz.$$

The line integral of the second kind depends on the direction the path is traced, so that

$$\int_{AB} \mathbf{a}\cdot d\mathbf{r} = -\int_{BA} \mathbf{a}\cdot d\mathbf{r}.$$

A line integral over a closed contour $\mathcal{C}$ is called a *closed path integral* (or a *circulation*) of a vector field $\mathbf{a}$ around $\mathcal{C}$ and is denoted

$$\oint_{\mathcal{C}} \mathbf{a}\cdot d\mathbf{r}.$$

Physical meaning of the line integral of the second kind: $\int_{AB} \mathbf{a}\cdot d\mathbf{r}$ determines the work done by the vector field $\mathbf{a}(x,y,z)$ on a particle of unit mass when it travels along the arc $\overset{\smile}{AB}$.

► **Computation of the line integral of the second kind.**

1°. For a plane curve $\overset{\smile}{AB}$ defined as $y = y(x)$, with $x \in [a, b]$, and a plane vector field $\mathbf{a}$, we have

$$\int_{AB} \mathbf{a}\cdot d\mathbf{r} = \int_a^b \bigl[P\bigl(x, y(x)\bigr) + Q\bigl(x, y(x)\bigr)y'_x(x)\bigr]\,dx.$$

2°. Let $\overset{\smile}{AB}$ be defined by a vector equation $\mathbf{r} = \mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}$, with $t \in [\alpha, \beta]$. Then

$$\begin{aligned}\int_{AB} \mathbf{a}\cdot d\mathbf{r} &= \int_{AB} P\,dx + Q\,dy + R\,dz\\ &= \int_\alpha^\beta \bigl[P\bigl(x(t), y(t), z(t)\bigr)x'_t(t) + Q\bigl(x(t), y(t), z(t)\bigr)y'_t(t) + R\bigl(x(t), y(t), z(t)\bigr)z'_t(t)\bigr]\,dt. \qquad \text{(M7.4.2.1)}\end{aligned}$$

For a plane curve $\overset{\smile}{AB}$ and a plane vector field $\mathbf{a}$, one should set $z'(t) = 0$ in (M7.4.2.1).

► **Potential and curl of a vector field.**

1°. A vector field $\mathbf{a} = \mathbf{a}(x, y, z)$ is called *potential* if there exists a function $\Phi(x, y, z)$ such that

$$\mathbf{a} = \operatorname{grad}\Phi, \quad \text{or} \quad \mathbf{a} = \frac{\partial\Phi}{\partial x}\mathbf{i} + \frac{\partial\Phi}{\partial y}\mathbf{j} + \frac{\partial\Phi}{\partial z}\mathbf{k}.$$

The function $\Phi(x, y, z)$ is called a *potential* of the vector field $\mathbf{a}$. The line integral of the second kind of a potential vector field along a path $\breve{AB}$ is equal to the increment of the potential along the path:

$$\int_{AB} \mathbf{a}\cdot d\mathbf{r} = \Phi\big|_B - \Phi\big|_A.$$

2°. The *curl* of a vector field $\mathbf{a}(x, y, z) = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ is the vector defined as

$$\operatorname{curl}\mathbf{a} = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\mathbf{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)\mathbf{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathbf{k} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{vmatrix}.$$

The vector $\operatorname{curl}\mathbf{a}$ characterizes the rate of rotation of $\mathbf{a}$ and can also be described as the circulation density of $\mathbf{a}$. Alternative notations: $\operatorname{curl}\mathbf{a} \equiv \nabla\times\mathbf{a} \equiv \operatorname{curl}\mathbf{a}$.

► **Necessary and sufficient conditions for a vector field to be potential.** Let U be a simply connected domain in $\mathbb{R}^3$ (i.e., a domain in which any closed contour can be deformed to a point without leaving U) and let $\mathbf{a}(x, y, z)$ be a vector field in U. Then the following four assertions are equivalent to each other:

(1) the vector field $\mathbf{a}$ is potential;

(2) $\operatorname{curl}\mathbf{a} \equiv \mathbf{0}$;

(3) the circulation of $\mathbf{a}$ around any closed contour $\mathcal{C} \in U$ is zero, or, equivalently, $\oint_{\mathcal{C}} \mathbf{a}\cdot d\mathbf{r} = 0$;

(4) the integral $\int_{AB} \mathbf{a}\cdot d\mathbf{r}$ is independent of the shape of $\breve{AB} \in U$ (it depends only on the initial and final points).

M7.4.3. Surface Integral of the First Kind

► **Definition of the surface integral of the first kind.** Let a function $f(x, y, z)$ be defined on a smooth surface D. Let us break up this surface into n elements (cells) that do not have common internal points and let us denote this partition by $\mathcal{D}_n$. The *diameter*, $\lambda(\mathcal{D}_n)$, of a partition $\mathcal{D}_n$ is the largest of the diameters of the cells (see Paragraph M7.3.4-1). Let us select in each cell an arbitrary point (x_i, y_i, z_i), $i = 1, 2, \ldots, n$, and make up an *integral sum*

$$s_n = \sum_{i=1}^{n} f(x_i, y_i, z_i)\,\Delta S_i,$$

where ΔS_i is the area of the ith element.

If there exists a finite limit of the sums s_n as $\lambda(\mathcal{D}_n) \to 0$ that depends on neither the partition $\mathcal{D}_n$ nor the selection of the points (x_i, y_i, z_i), then it is called the *surface integral of the first kind* of the function $f(x, y, z)$ and is denoted $\iint_D f(x, y, z)\,dS$.

► **Computation of the surface integral of the first kind.**

1°. If a surface D is defined by an equation $z = z(x, y)$, with $(x, y) \in D_1$, then

$$\iint_D f(x, y, z)\, dS = \iint_{D_1} f\bigl(x, y, z(x, y)\bigr) \sqrt{1 + (z'_x)^2 + (z'_y)^2}\, dx\, dy.$$

2°. If a surface D is defined by a vector equation $\mathbf{r} = \mathbf{r}(x, y, z) = x(u, v)\, \mathbf{i} + y(u, v)\, \mathbf{j} + z(u, v)\, \mathbf{k}$, where $(u, v) \in D_2$, then

$$\iint_D f(x, y, z)\, dS = \iint_{D_2} f\bigl(x(u, v), y(u, v), z(u, v)\bigr)\, |\mathbf{n}(u, v)|\, du\, dv,$$

where $\mathbf{n}(u, v) = \mathbf{r}_u \times \mathbf{r}_v$ is a normal to the surface D; the subscripts u and v denote the respective partial derivatives.

► **Applications of the surface integral of the first kind.**

1°. Area of a surface D:

$$S_D = \iint_D dS.$$

2°. Mass of a material surface D with a surface density $\gamma = \gamma(x, y, z)$:

$$m = \iint_D \gamma(x, y, z)\, dS.$$

3°. Coordinates of the center of mass of a material surface D:

$$x_{\rm c} = \frac{1}{m} \iint_D x\gamma\, dS, \quad y_{\rm c} = \frac{1}{m} \iint_D y\gamma\, dS, \quad z_{\rm c} = \frac{1}{m} \iint_D z\gamma\, dS.$$

To the uniform surface density there corresponds $\gamma = \text{const}$.

M7.4.4. Surface Integral of the Second Kind

► **Definition of the surface integral of the second kind.** Let a vector field $\mathbf{a}(x, y, z) = P\, \mathbf{i} + Q\, \mathbf{j} + R\, \mathbf{k}$ be defined on a smooth oriented surface D. Let us perform a partition, $\mathcal{D}_n$, of the surface D into n elements (cells) that do not have common internal points. Also select an arbitrary point $M_i(x_i, y_i, z_i)$, $i = 1, 2, \ldots, n$, for each cell and make up an *integral sum* $s_n = \sum\limits_{i=1}^{n} \mathbf{a}(x_i, y_i, z_i) \cdot \mathbf{n}_i^\circ\, \Delta S_i$, where ΔS_i is area of the ith cell and $\mathbf{n}_i^\circ$ is the unit normal to the surface at the point M_i, the orientation of which coincides with that of the surface.

If there exists a finite limit of the sums s_n as $\lambda(\mathcal{D}_n) \to 0$ (λ is the diameter of the partition, see Subsection M7.4.3) that depends on neither the partition $\mathcal{D}_n$ nor the selection of the points $M_i(x_i, y_i, z_i)$, then it is called the *surface integral of the second kind* (or the *flux of the vector field* **a** *across the oriented surface* D) and is denoted

$$\iint_D \mathbf{a}(x, y, z) \cdot \overrightarrow{dS}, \quad \text{or} \quad \iint_D P\, dy\, dz + Q\, dx\, dz + R\, dx\, dy.$$

Note that the surface integral of the second kind changes its sign when the orientation of the surface is reversed.

► **Computation of the surface integral of the second kind.**

1°. If a surface D is defined by an equation $z = z(x, y)$, with $(x, y) \in D_2$, then the normal $\mathbf{n}(x, y) = \mathbf{r}_x \times \mathbf{r}_y = -z_x\mathbf{i} - z_y\mathbf{j} + \mathbf{k}$ orients the surface D "upward," in the positive direction of the z-axis; the subscripts x and y denote the respective partial derivatives. Then

$$\iint_D \mathbf{a} \cdot \overrightarrow{dS} = \pm \iint_{D_2} \bigl(-z_x P - z_y Q + R\bigr)\, dx\, dy,$$

where $P = P\bigl(x, y, z(x, y)\bigr)$, $Q = Q\bigl(x, y, z(x, y)\bigr)$, and $R = R\bigl(x, y, z(x, y)\bigr)$. The plus sign is taken if the surface has the "upward" orientation, and the minus sign is chosen in the opposite case.

M7.4.5. Integral Formulas of Vector Calculus

► **Ostrogradsky–Gauss theorem (divergence theorem).** Let a vector field $\mathbf{a}(x, y, z) = P(x, y, z)\,\mathbf{i} + Q(x, y, z)\,\mathbf{j} + R(x, y, z)\,\mathbf{k}$ be continuously differentiable in a finite simply connected domain $V \subset \mathbb{R}^3$ and let S denote the surface of V oriented by an outward normal. Then the *Ostrogradsky–Gauss theorem* (or the *divergence theorem*) holds:

$$\iint_S \mathbf{a} \cdot \overrightarrow{dS} = \iiint_V \operatorname{div} \mathbf{a}\; dx\, dy\, dz,$$

where div **a** is the *divergence* of the vector **a**, which is defined as follows:

$$\operatorname{div} \mathbf{a} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}.$$

Thus, the flux of a vector field across a closed surface in the outward direction is equal to the triple integral of the divergence of the vector field over the volume bounded by the surface. In coordinate form, the Ostrogradsky–Gauss theorem reads

$$\iint_S P\, dy\, dz + Q\, dx\, dz + R\, dx\, dy = \iiint_V \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right) dx\, dy\, dz.$$

► **Stokes's theorem (curl theorem).**

1°. Let a vector field $\mathbf{a}(x, y, z)$ be continuously differentiable in a domain of the three-dimensional space $\mathbb{R}^3$ that contains an oriented surface D. The orientation of a surface uniquely defines the direction in which the boundary of the surface is traced; specifically, the boundary is traced counterclockwise when looked at from the direction of the normal to the surface. Then the circulation of the vector field around the boundary $\mathcal{C}$ of the surface D is equal to the flux of the vector curl **a** across D:

$$\oint_{\mathcal{C}} \mathbf{a} \cdot d\mathbf{r} = \iint_D \operatorname{curl} \mathbf{a} \cdot \overrightarrow{dS}.$$

In coordinate notation, *Stokes's theorem* reads

$$\oint_{\mathcal{C}} P\, dx + Q\, dy + R\, dz = \iint_D \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) dy\, dz + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) dx\, dz + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\, dy.$$

2°. For a plane vector field $\mathbf{a}(x, y) = P(x, y)\,\mathbf{i} + Q(x, y)\,\mathbf{j}$, Stokes's theorem reduces to *Green's theorem*:

$$\oint_{\mathcal{C}} P\, dx + Q\, dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\, dy,$$

where the contour $\mathcal{C}$ of the domain D on the xy-plane is traced counterclockwise.

Bibliography for Chapter M7

Adams, R., *Calculus: A Complete Course, 6th Edition*, Pearson Education, Toronto, 2006.

Anton, H., *Calculus: A New Horizon, 6th Edition*, Wiley, New York, 1999.

Anton, H., Bivens, I., and Davis, S., *Calculus: Early Transcendental Single Variable, 8th Edition*, John Wiley & Sons, New York, 2005.

Aramanovich, I. G., Guter, R. S., et al., *Mathematical Analysis (Differentiation and Integration)*, Fizmatlit Publishers, Moscow, 1961.

Borden, R. S., *A Course in Advanced Calculus*, Dover Publications, New York, 1998.

Brannan, D., *A First Course in Mathematical Analysis*, Cambridge University Press, Cambridge, England, 2006.

Bronshtein, I. N. and Semendyayev, K. A., *Handbook of Mathematics, 4th Edition*, Springer-Verlag, Berlin, 2004.

Browder, A., *Mathematical Analysis: An Introduction*, Springer-Verlag, New York, 1996.

Clark, D. N., *Dictionary of Analysis, Calculus, and Differential Equations*, CRC Press, Boca Raton, Florida, 2000.

Courant, R. and John, F., *Introduction to Calculus and Analysis, Vol. 1*, Springer-Verlag, New York, 1999.

Danilov, V. L., Ivanova, A. N., et al., *Mathematical Analysis (Functions, Limits, Series, Continued Fractions)* [in Russian], Fizmatlit Publishers, Moscow, 1961.

Dwight, H. B., *Tables of Integrals and Other Mathematical Data*, Macmillan, New York, 1961.

Edwards, C. H., and Penney, D., *Calculus, 6th Edition*, Pearson Education, Toronto, 2002.

Fedoryuk, M. V., *Asymptotics, Integrals and Series* [in Russian], Nauka Publishers, Moscow, 1987.

Fikhtengol'ts, G. M., *Fundamentals of Mathematical Analysis, Vol. 2*, Pergamon Press, London, 1965.

Fikhtengol'ts, G. M., *A Course of Differential and Integral Calculus, Vol. 2* [in Russian], Nauka Publishers, Moscow, 1969.

Gradshteyn, I. S. and Ryzhik, I. M., *Tables of Integrals, Series and Products, 6th Edition*, Academic Press, New York, 2000.

Kaplan, W., *Advanced Calculus, 5th Edition*, Addison Wesley, Reading, Massachusetts, 2002.

Kline, M., *Calculus: An Intuitive and Physical Approach, 2nd Edition*, Dover Publications, New York, 1998.

Landau, E., *Differential and Integral Calculus*, American Mathematical Society, Providence, Rhode Island, 2001.

Marsden, J. E. and Weinstein, A., *Calculus, 2nd Edition*, Springer-Verlag, New York, 1985.

Mendelson, E., *3000 Solved Problems in Calculus*, McGraw-Hill, New York, 1988.

Polyanin, A. D., Polyanin, V. D., et al., *Handbook for Engineers and Students. Higher Mathematics. Physics. Theoretical Mechanics. Strength of Materials, 3rd Edition* [in Russian], AST/Astrel, Moscow, 2005.

Prudnikov, A. P., Brychkov, Yu. A., and Marichev, O. I., *Integrals and Series, Vol. 1, Elementary Functions*, Gordon & Breach, New York, 1986.

Silverman, R. A., *Essential Calculus with Applications*, Dover Publications, New York, 1989.

Strang, G., *Calculus*, Wellesley-Cambridge Press, Massachusetts, 1991.

Taylor, A. E. and Mann, W. R., *Advanced Calculus, 3rd Edition*, John Wiley, New York, 1983.

Thomas, G. B. and Finney, R. L., *Calculus and Analytic Geometry, 9th Edition*, Addison Wesley, Reading, Massachusetts, 1996.

Widder, D. V., *Advanced Calculus, 2nd Edition*, Dover Publications, New York, 1989.

Zorich, V. A., *Mathematical Analysis*, Springer-Verlag, Berlin, 2004.

Zwillinger, D., *CRC Standard Mathematical Tables and Formulae, 31st Edition*, CRC Press, Boca Raton, Florida, 2002.

Chapter M8

Series

M8.1. Numerical Series and Infinite Products

M8.1.1. Convergent Numerical Series and Their Properties. Cauchy's Criterion

► **Basic definitions.** Let $\{a_n\}$ be a numerical sequence. The expression

$$a_1 + a_2 + \cdots + a_n + \cdots = \sum_{n=1}^{\infty} a_n$$

is called a *numerical series* (*infinite sum*, *infinite numerical series*), a_n is the *generic term of the series*, and

$$s_n = a_1 + a_2 + \cdots + a_n = \sum_{k=1}^{n} a_k$$

is the *nth partial sum of the series*. If there exists a finite limit $\lim\limits_{n\to\infty} s_n = S$, the series is called *convergent*, and S is called the *sum of the series*. In this case, one writes $\sum\limits_{n=1}^{\infty} a_n = S$. If $\lim\limits_{n\to\infty} s_n$ does not exist (or is infinite), the series is called *divergent*. The series $a_{n+1} + a_{n+2} + a_{n+3} + \cdots$ is called the *nth remainder of the series*.

Example 1. Consider the series $\sum\limits_{n=1}^{\infty} aq^{n-1} = a + aq + aq^2 + \cdots$ whose terms form a *geometric progression* with ratio q. This series is convergent for $|q| < 1$ (its sum has the form $S = \frac{a}{1-q}$) and is divergent for $|q| \ge 1$.

► **Necessary condition for a series to be convergent. Cauchy's criterion.**

1. *A necessary condition for a series to be convergent.* For a convergent series $\sum\limits_{n=1}^{\infty} a_n$, the generic term must tend to zero, $\lim\limits_{n\to\infty} a_n = 0$. If $\lim\limits_{n\to\infty} a_n \ne 0$, then the series is divergent.

Example 2. The series $\sum\limits_{n=1}^{\infty} \cos\dfrac{1}{n}$ is divergent, since its generic term $a_n = \cos\dfrac{1}{n}$ does not tend to zero as $n \to \infty$.

The above necessary condition is insufficient for the convergence of a series.

Example 3. Consider the series $\sum\limits_{n=1}^{\infty} \dfrac{1}{\sqrt{n}}$. Its generic term tends to zero, $\lim\limits_{n\to\infty} \dfrac{1}{\sqrt{n}} = 0$, but the series $\sum\limits_{n=1}^{\infty} \dfrac{1}{\sqrt{n}}$ is divergent because its partial sums are unbounded,

$$s_n = \frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \cdots + \frac{1}{\sqrt{n}} > n\frac{1}{\sqrt{n}} = \sqrt{n} \to \infty \quad \text{as} \quad n \to \infty.$$

2. *Cauchy's criterion of convergence of a series.* A series $\sum\limits_{n=1}^{\infty} a_n$ is convergent if and only if for any $\varepsilon > 0$ there exists an $N = N(\varepsilon)$ such that for all $n > N$ and any positive integer k, the following inequality holds: $|a_{n+1} + \cdots + a_{n+k}| < \varepsilon$.

► **Properties of convergent series.**

1. If a series is convergent, then any of its remainders is convergent. Removal or addition of finitely many terms does not affect the convergence of a series.

2. If all terms of a series are multiplied by a nonzero constant, the resulting series preserves the property of convergence or divergence (its sum is multiplied by that constant).

3. If the series $\sum\limits_{n=1}^{\infty} a_n$ and $\sum\limits_{n=1}^{\infty} b_n$ are convergent and their sums are equal to S_1 and S_2, respectively, then the series $\sum\limits_{n=1}^{\infty}(a_n \pm b_n)$ are convergent and their sums are equal to $S_1 \pm S_2$.

4. Terms of a convergent series can be grouped in successive order; the resulting series has the same sum. In other words, one can insert brackets inside a series in an arbitrary order. The inverse operation of opening brackets is not always admissible. Thus, the series $(1-1)+(1-1)+\cdots$ is convergent (its sum is equal to zero), but, after removing the brackets, we obtain the divergent series $1-1+1-1+\cdots$ (its generic term does not tend to zero).

M8.1.2. Convergence Criteria for Series with Positive (Nonnegative) Terms

► **Basic convergence (divergence) criteria for series with positive terms.**

1. *The first comparison criterion.* If $0 \le a_n \le b_n$ (starting from some n), then the convergence of the series $\sum\limits_{n=1}^{\infty} b_n$ implies the convergence of $\sum\limits_{n=1}^{\infty} a_n$; and the divergence of the series $\sum\limits_{n=1}^{\infty} a_n$ implies the divergence of $\sum\limits_{n=1}^{\infty} b_n$.

2. *The second convergence criterion.* Suppose that there exists a finite limit

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \sigma,$$

where $0 < \sigma < \infty$. Then $\sum\limits_{n=1}^{\infty} a_n$ is convergent (resp., divergent) if and only if $\sum\limits_{n=1}^{\infty} b_n$ is convergent (resp., divergent).

Corollary. Suppose that $a_{n+1}/a_n \le b_{n+1}/b_n$ starting from some N (i.e., for $n > N$). Then convergence of the series $\sum\limits_{n=1}^{\infty} b_n$ implies convergence of $\sum\limits_{n=1}^{\infty} a_n$, and divergence of $\sum\limits_{n=1}^{\infty} a_n$ implies divergence of $\sum\limits_{n=1}^{\infty} b_n$.

3. *D'Alembert criterion.* Suppose that there exists the limit (finite or infinite)

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = D.$$

If $D < 1$, then the series $\sum\limits_{n=1}^{\infty} a_n$ is convergent. If $D > 1$, then the series is divergent. For $D = 1$, the d'Alembert criterion cannot be used for deciding whether the series is convergent or divergent.

Example 1. Let us examine the convergence of the series $\sum\limits_{n=1}^{\infty} n^k x^n$ with $x > 0$, using the d'Alembert criterion. Taking $a_n = n^k x^n$, we get

$$\frac{a_{n+1}}{a_n} = \left(1 + \frac{1}{n}\right)^k x \to x \quad \text{as} \quad n \to \infty.$$

Therefore, $D = x$. It follows that the series is convergent for $x < 1$ and divergent for $x > 1$. The series is divergent for $x = 1$, since a_n does not tend to zero as $n \to \infty$.

4. *Cauchy criterion.* Suppose that there exists the limit (finite or infinite)

$$\lim_{n\to\infty} \sqrt[n]{a_n} = K.$$

For $K < 1$, the series $\sum\limits_{n=1}^{\infty} a_n$ is convergent; for $K > 1$, the series is divergent. For $K = 1$, the Cauchy criterion cannot be used to establish convergence of a series.

Remark. The Cauchy criterion is stronger than the d'Alembert criterion, but the latter is, in many cases, simpler than the former.

5. *Gauss's criterion.* Suppose that the ratio of two consecutive terms of a series can be represented in the form

$$\frac{a_n}{a_{n+1}} = \lambda + \frac{\mu}{n} + o\left(\frac{1}{n}\right) \quad \text{as} \quad n \to \infty.$$

The series $\sum\limits_{n=1}^{\infty} a_n$ is convergent if $\lambda > 1$ or if $\lambda = 1$ and $\mu > 1$. The series is divergent if $\lambda < 1$ or if $\lambda = 1$ and $\mu \le 1$.

6. *Maclaurin–Cauchy integral criterion.* Let $f(x)$ be a nonnegative nonincreasing continuous function on the interval $1 \le x < \infty$. Let $f(1) = a_1$, $f(2) = a_2$, $\dots$, $f(n) = a_n$, $\dots$ Then the series $\sum\limits_{n=1}^{\infty} a_n$ is convergent if and only if the improper integral $\int_1^{\infty} f(x)\,dx$ is convergent.

Example 2. The *harmonic series* $\sum\limits_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots$ is divergent, since the integral $\int_1^{\infty} \frac{1}{x}\,dx$ is divergent. In a similar way, one finds that the series $\sum\limits_{n=1}^{\infty} \frac{1}{n^{\alpha}}$ is convergent for $\alpha > 1$ and divergent for $\alpha \le 1$.

M8.1.3. Convergence Criteria for Arbitrary Numerical Series. Absolute and Conditional Convergence

▶ Arbitrary series. Leibniz, Abel, and Dirichlet convergence criteria.

1. *Leibniz criterion.* Suppose that the terms a_n of a series $\sum\limits_{n=1}^{\infty} a_n$ have alternating signs, their absolute values form a nonincreasing sequence, and $a_n \to 0$ as $n \to \infty$. Then this "alternating" series is convergent. If S is the sum of the series and s_n is its nth partial sum, then the following inequality holds for the error $|S - s_n| \le |a_{n+1}|$.

Example 1. The series $1 - \frac{1}{2^2} + \frac{1}{3^3} - \frac{1}{4^4} + \frac{1}{5^5} - \cdots$ is convergent by the Leibniz criterion. Taking $S \approx s_4 = 1 - \frac{1}{2^2} + \frac{1}{3^3} - \frac{1}{4^4}$, we obtain the error less than $a_5 = \frac{1}{5^5} = 0.00032$.

2. *Abel criterion.* Consider the series

$$\sum_{n=1}^{\infty} a_n b_n = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n + \cdots, \qquad \text{(M8.1.3.1)}$$

where a_n and b_n are two sequences or real numbers.

Series (M8.1.3.1) is convergent if the series

$$\sum_{n=1}^{\infty} b_n = b_1 + b_2 + \cdots + b_n + \cdots \qquad \text{(M8.1.3.2)}$$

is convergent and the a_n form a bounded monotone sequence ($|a_n| < K$).

3. *Dirichlet criterion.* Series (M8.1.3.1) is convergent if partial sums of series (M8.1.3.2) are bounded uniformly in n,

$$\left|\sum_{k=1}^{n} b_k\right| \le M \qquad (n = 1,\, 2,\, \dots),$$

and the sequence $a_n \to 0$ is monotone.

Example 2. Consider the series $\sum\limits_{n=1}^{\infty} a_n \sin(nx)$, where $a_n \to 0$ is a monotonically decreasing sequence. Taking $b_n = \sin(nx)$ and using a well-known identity, we find the partial sum

$$s_n = \sum_{k=1}^{n} \sin(kx) = \frac{\cos\left(\frac{1}{2}x\right) - \cos\left[\left(n + \frac{1}{2}\right)x\right]}{2\sin\left(\frac{1}{2}x\right)} \qquad (x \ne 2m\pi;\ \ m = 0, \pm1, \pm2, \dots).$$

This sum is bounded for $x \ne 2m\pi$:

$$|s_n| \le \frac{1}{\left|\sin\left(\frac{1}{2}x\right)\right|}.$$

Therefore, by the Dirichlet criterion, the series $\sum\limits_{n=1}^{\infty} a_n \sin(nx)$ is convergent for any $x \ne 2m\pi$. Direct verification shows that this series is also convergent for $x = 2m\pi$ (since all its terms at these points are equal to zero).

Remark. The Leibniz and the Abel criteria can be deduced from the Dirichlet criterion.

► Absolute and conditional convergence.

1. *Absolutely convergent series.* A series $\sum\limits_{n=1}^{\infty} a_n$ (with terms of arbitrary sign) is called *absolutely convergent* if the series $\sum\limits_{n=1}^{\infty} |a_n|$ is convergent.

Any absolutely convergent series is convergent. In order to establish absolute convergence of a series, one can use all convergence criteria for series with nonnegative terms given in Subsection M8.1.2 (in these criteria, a_n should be replaced by $|a_n|$).

Example 3. The series $1 + \frac{1}{2^2} - \frac{1}{3^2} - \frac{1}{4^2} + \frac{1}{5^2} + \frac{1}{6^2} - \cdots$ is absolutely convergent, since the series with the absolute values of its terms, $\sum\limits_{n=1}^{\infty} \frac{1}{n^2}$, is convergent (see the second series in Example 2 of Subsection M8.1.2 for $\alpha = 2$).

2. *Conditionally convergent series.* A convergent series $\sum\limits_{n=1}^{\infty} a_n$ is called *conditionally convergent* if the series $\sum\limits_{n=1}^{\infty} |a_n|$ is divergent.

Example 4. The series $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$ is conditionally convergent, since it is convergent (by the Leibniz criterion), but the series with absolute values of its terms is divergent (it is a harmonic series; see Example 2 in Subsection M8.1.2).

Any rearrangement of the terms of an absolutely convergent series (in particular, a convergent series with nonnegative terms) neither violates its absolute convergence nor changes its sum. Conditionally convergent series do not possess this property: the terms of a conditionally convergent series can be rearranged in such order that the sum of the new series becomes equal to any given value; its terms can also be rearranged so as to result in a divergent series.

M8.1.4. Multiplication of Series. Some Inequalities

► **Multiplication of series. Cauchy, Mertens, and Abel theorems.** A product of two infinite series $\sum\limits_{n=0}^{\infty} a_n$ and $\sum\limits_{n=0}^{\infty} b_n$ is understood as a series whose terms have the form $a_n b_m$ (n, $m = 0, 1, \ldots$). The products $a_n b_m$ can be ordered to form a series in many different ways. The following theorems allow us to decide whether it is possible to multiply series.

CAUCHY THEOREM. *Suppose that the series* $\sum\limits_{n=0}^{\infty} a_n$ *and* $\sum\limits_{n=0}^{\infty} b_n$ *are absolutely convergent and their sums are equal to* A *and* B*, respectively. Then any product of these series is an absolutely convergent series and its sum is equal to* AB*. The following Cauchy multiplication formula holds*:

$$\left(\sum_{n=0}^{\infty} a_n\right)\left(\sum_{n=0}^{\infty} b_n\right) = \sum_{n=0}^{\infty}\left(\sum_{m=0}^{n} a_m b_{n-m}\right). \qquad \text{(M8.1.4.1)}$$

MERTENS THEOREM. *The Cauchy multiplication formula (M8.1.4.1) is also valid if one of the series,* $\sum\limits_{n=0}^{\infty} a_n$ *or* $\sum\limits_{n=0}^{\infty} b_n$*, is absolutely convergent and the other is (conditionally) convergent. In this case, the product is a convergent series, possibly, not absolutely convergent.*

ABEL THEOREM. *Consider two convergent series with sums* A *and* B*. Suppose that the product of these series in the form of Cauchy (M8.1.4.1) is a convergent series with sum* C*. Then* $C = AB$*.*

► **Inequalities.**

1. *Generalized triangle inequality*:

$$\left|\sum_{n=1}^{\infty} a_n\right| \le \sum_{n=1}^{\infty} |a_n|.$$

2. *Cauchy inequality* (*Cauchy–Schwarz–Bunyakovsky inequality*):

$$\left(\sum_{n=1}^{\infty} a_n b_n\right)^2 \le \left(\sum_{n=1}^{\infty} a_n^2\right)\left(\sum_{n=1}^{\infty} b_n^2\right).$$

3. *Minkowski inequality*:

$$\left(\sum_{n=1}^{\infty} |a_n + b_n|^p\right)^{\frac{1}{p}} \le \left(\sum_{n=1}^{\infty} |a_n|^p\right)^{\frac{1}{p}} + \left(\sum_{n=1}^{\infty} |b_n|^p\right)^{\frac{1}{p}}, \qquad p \ge 1.$$

In all these inequalities it is assumed that the series on the right-hand sides are convergent.

M8.2. Function Series

M8.2.1. Pointwise and Uniform Convergence of Function Series

► **Convergence of a function series at a point. Convergence domain.** A *function series* is a series of the form

$$u_1(x) + u_2(x) + \cdots + u_n(x) + \cdots = \sum_{n=1}^{\infty} u_n(x),$$

where $u_n(x)$ are functions defined on a set $X \subset \mathbb{R}$. The series $\sum\limits_{n=1}^{\infty} u_n(x)$ is called *convergent at a point* $x_0 \in X$ if the numerical series $\sum\limits_{n=1}^{\infty} u_n(x_0)$ is convergent. The set of all $x \in X$ for which the function series is convergent is called its *convergence domain*. The sum of the series is a function of x defined on its convergence domain.

In order to find the convergence domain for a function series, one can use the convergence criteria for numerical series described in Subsections M8.1.2 and M8.1.3 (with the variable x regarded as a parameter).

A series $\sum\limits_{n=1}^{\infty} u_n(x)$ is called *absolutely convergent on a set* X if the series $\sum\limits_{n=1}^{\infty} |u_n(x)|$ is convergent on this set.

Example. The function series

$$1 + x + x^2 + x^3 + \cdots$$

is convergent for $-1 < x < 1$ (see Example 1 in Subsection M8.1.1). Its sum is defined on this interval, $S = \dfrac{1}{1-x}$.

The series $\sum\limits_{k=n+1}^{\infty} u_k(x)$ is called the *remainder of a function series* $\sum\limits_{n=1}^{\infty} u_n(x)$. For a series convergent on a set X, the relation $S(x) = s_n(x) + r_n(x)$, where $s_n(x)$ is the partial sum of the series and $r_n(x)$ is the sum of its remainder, implies that $\lim\limits_{n\to\infty} r_n(x) = 0$ for $x \in X$.

► **Uniformly convergent series. Condition of uniform convergence.** A function series is called *uniformly convergent* on a set X if for any $\varepsilon > 0$ there exists an N (dependent on ε but not on x) such that for all $n > N$, the inequality

$$\left|\sum_{k=n+1}^{\infty} u_k(x)\right| < \varepsilon$$

holds for all $x \in X$.

A *necessary and sufficient condition of uniform convergence of a series.* A series $\sum\limits_{n=1}^{\infty} u_n(x)$ is uniformly convergent on a set X if and only if for any $\varepsilon > 0$ there exists an N (independent of x) such that for all $n > N$ and all $m = 1, 2, \ldots$, the inequality

$$\left|\sum_{k=n+1}^{n+m} u_k(x)\right| < \varepsilon$$

holds for all $x \in X$.

M8.2.2. Basic Criteria of Uniform Convergence. Properties of Uniformly Convergent Series

► **Criteria of uniform convergence of series.**

1. *Weierstrass criterion of uniform convergence.* A function series $\sum\limits_{n=1}^{\infty} u_n(x)$ is uniformly convergent on a set $X \subset \mathbb{R}$ if there is a convergent number series $\sum\limits_{n=1}^{\infty} a_n$ with

nonnegative terms such that $|u_n(x)| \le a_n$ for all sufficiently large n and all $x \in X$. The series $\sum\limits_{n=1}^{\infty} a_n$ is called a *majorant series* for $\sum\limits_{n=1}^{\infty} u_n(x)$.

Example. The series $\sum\limits_{n=1}^{\infty}(-1)^n \dfrac{\sin nx}{n^2}$ is uniformly convergent for $-\infty < x < \infty$, since $\left|(-1)^n \dfrac{\sin nx}{n^2}\right| \le \dfrac{1}{n^2}$, and the numerical series $\sum\limits_{n=1}^{\infty} \dfrac{1}{n^2}$ is convergent (see the second series in Example 2 in Subsection M8.1.2).

2. *Abel criterion of uniform convergence of function series.* Consider a function series

$$\sum_{n=1}^{\infty} u_n(x)v_n(x) = u_1(x)v_1(x) + u_2(x)v_2(x) + \cdots + u_n(x)v_n(x) + \cdots, \qquad \text{(M8.2.2.1)}$$

where $u_n(x)$ and $v_n(x)$ are sequences of functions of the real variable $x \in [a, b]$.

Series (M8.2.2.1) is uniformly convergent on the interval $[a, b]$ if the series

$$\sum_{n=1}^{\infty} v_n(x) = v_1(x) + v_2(x) + \cdots + v_n(x) + \cdots \qquad \text{(M8.2.2.2)}$$

is uniformly convergent on $[a, b]$ and the functions $u_n(x)$ form a monotone sequence for each x and are uniformly bounded (i.e., $|u_n(x)| \le K$ with a constant K independent of n, x).

3. *Dirichlet criterion of uniform convergence of function series.* Series (M8.2.2.1) is uniformly convergent on the interval $[a, b]$ if the partial sums of the series (M8.2.2.2) are uniformly bounded, i.e.,

$$\left|\sum_{k=1}^{n} v_k(x)\right| \le M = \text{const} \qquad (x \in [a, b], \quad n = 1,\ 2,\ \dots),$$

and the functions $u_n(x)$ form a monotone sequence (for each x) that uniformly converges to zero on $[a, b]$ as $n \to \infty$.

▶ **Properties of uniformly convergent series.** Let $\sum\limits_{n=1}^{\infty} u_n(x)$ be a function series that is uniformly convergent on a segment $[a, b]$, and let $S(x)$ be its sum. Then the following statements hold.

THEOREM 1. *If all terms $u_n(x)$ of the series are continuous at a point $x_0 \in [a, b]$, then the sum $S(x)$ is continuous at that point.*

THEOREM 2. *If the terms $u_n(x)$ are continuous on $[a, b]$, then the series admits term-by-term integration:*

$$\int_a^b S(x)\,dx = \int_a^b \left(\sum_{n=1}^{\infty} u_n(x)\right) dx = \sum_{n=1}^{\infty} \int_a^b u_n(x)\,dx.$$

Remark. The condition of continuity of the functions $u_n(x)$ on $[a, b]$ can be replaced by a weaker condition of their integrability on $[a, b]$.

THEOREM 3. *If all terms of the series have continuous derivatives and the function series $\sum\limits_{n=1}^{\infty} u_n'(x)$ is uniformly convergent on $[a, b]$, then the sum $S(x)$ is continuously differentiable on $[a, b]$ and*

$$S'(x) = \left(\sum_{n=1}^{\infty} u_n(x)\right)' = \sum_{n=1}^{\infty} u_n'(x)$$

(i.e., the series admits term-by-term differentiation).

M8.3. Power Series

M8.3.1. Radius of Convergence of Power Series. Properties of Power Series

► **Abel theorem. Convergence radius of a power series.** A *power series* is a function series of the form

$$\sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots \qquad \text{(M8.3.1.1)}$$

(the constants $a_0, a_1, \dots$ are called the *coefficients* of the power series), and also a series of a more general form

$$\sum_{n=0}^{\infty} a_n (x-x_0)^n = a_0 + a_1(x-x_0) + a_2(x-x_0)^2 + a_3(x-x_0)^3 + \cdots ,$$

where x_0 is a fixed point. Below, we consider power series of the first form, since the second series can be transformed into the first by the replacement $\bar{x} = x - x_0$.

ABEL THEOREM. *A power series $\sum_{n=0}^{\infty} a_n x^n$ that is convergent for some $x = x_1$ is absolutely convergent for all x such that $|x| < |x_1|$. A power series that is divergent for some $x = x_2$ is divergent for all x such that $|x| > |x_2|$.*

Remark. There exist series convergent for all x, for instance, $\sum_{n=1}^{\infty} \frac{x^n}{n!}$. There are series convergent only for $x = 0$, for instance, $\sum_{n=1}^{\infty} n!\, x^n$.

For a given power series (M8.3.1.1), let R be the least upper bound of all $|x|$ such that the series (M8.3.1.1) is convergent at point x. Thus, by the Abel theorem, the series is (absolutely) convergent for all $|x| < R$, and the series is divergent for all $|x| > R$. The constant R is called the *radius of convergence* of the power series, and the interval $(-R, R)$ is called its *interval of convergence*. The problem of convergence of a power series at the endpoints of its convergence interval has to be studied separately in each specific case. If a series is convergent only for $x = 0$, the convergence interval degenerates into a point (and $R = 0$); if a series is convergent for all x, then, obviously, $R = \infty$.

► **Formulas for the radius of convergence of power series.**

1°. The radius of convergence of a power series (M8.3.1.1) with finitely many zero terms can be calculated by the formulas

$$R = \lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right| \quad \text{(obtained from the d'Alembert criterion for numerical series),}$$

$$R = \lim_{n\to\infty} \frac{1}{\sqrt[n]{|a_n|}} \quad \text{(obtained from the Cauchy criterion for numerical series).}$$

Example 1. For the power series $\sum_{n=1}^{\infty} \frac{3^n}{n} x^n$, using the first formula for the radius of convergence, we get

$$R = \lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right| = \lim_{n\to\infty} \left| \frac{n+1}{3n} \right| = \frac{1}{3}.$$

Therefore, the series is absolutely convergent on the interval $-\frac{1}{3} < x < \frac{1}{3}$ and is divergent outside that interval. At the left endpoint of the interval, for $x = -\frac{1}{3}$, we have the conditionally convergent series $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$, and at the right endpoint, for $x = \frac{1}{3}$, we have the divergent numerical series $\sum\limits_{n=1}^{\infty} \frac{1}{n}$. Thus, the series under consideration is convergent on the semi-open interval $\left[-\frac{1}{3}, \frac{1}{3}\right)$.

If the number of zero coefficients in a power series is infinite, the above formulas for R are inapplicable. In such cases, one can directly apply the d'Alembert or Cauchy criteria to the series.

Example 2. For the power series $\displaystyle\sum_{n=1}^{\infty} \frac{4^n}{n} x^{2n}$, the d'Alembert criterion gives

$$\lim_{n\to\infty} \left| \frac{a_{n+1} x^{2(n+1)}}{a_n x^{2n}} \right| = \lim_{n\to\infty} \left| \frac{4nx^2}{n+1} \right| = 4x^2.$$

Then the given series is absolutely convergent if $4x^2 < 1$, or on the interval $-\frac{1}{2} < x < \frac{1}{2}$, and hence $R = \frac{1}{2}$. (It is easily seen that the original series diverges at both endpoints of the interval of convergence.)

2°. Suppose that a power series (M8.3.1.1) is convergent at a boundary point of its convergence interval, say, for $x = R$. Then its sum is left-continuous at that point,

$$\lim_{x\to R-0} \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} a_n R^n.$$

Example 3. Having the expansion

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n+1} \frac{x^n}{n} + \cdots \qquad (R = 1)$$

in the domain $-1 < x < 1$ and knowing that the series

$$1 - \frac{1}{2} + \frac{1}{3} - \cdots + (-1)^{n+1} \frac{1}{n} + \cdots$$

is convergent (by the Leibniz criterion for series with terms of alternating sign), we conclude that the sum of the last series is equal to ln 2.

► **Properties of power series.** On any closed segment belonging to the (open) convergence interval of a power series, the series is uniformly convergent. Therefore, on any such segment, the series has all the properties of uniformly convergent series described in Subsection M8.2.2. Therefore, the following statements hold:

1. A power series (M8.3.1.1) admits term-by-term integration on any segment $[0, x]$ for $|x| < R$,

$$\int_0^x \left(\sum_{n=0}^{\infty} a_n x^n \right) dx = \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1}$$
$$= a_0 x + \frac{a_1}{2} x^2 + \frac{a_2}{3} x^3 + \cdots + \frac{a_n}{n+1} x^{n+1} + \cdots .$$

Remark 1. The value of x in this formula may coincide with an endpoint of the convergence interval ($x = -R$ and/or $x = R$), provided that series (M8.3.1.1) is convergent at that point.

Remark 2. The convergence radii of the original series and the series obtained by its term-by-term integration on the segment $[0, x]$ coincide.

2. Inside the convergence interval (for $|x| < R$), the series admits term-by-term differentiation of any order, in particular,

$$\frac{d}{dx} \left(\sum_{n=0}^{\infty} a_n x^n \right) = \sum_{n=1}^{\infty} n a_n x^{n-1}$$
$$= a_1 + 2a_2 x + 3a_3 x^2 + \cdots + n a_n x^{n-1} + \cdots .$$

Remark 1. The sum of a power series is a function that has derivatives of any order inside the interval of convergence.

Remark 2. The convergence radii of the original series and the series obtained by its term-by-term differentiation coincide.

M8.3.2. Taylor and Maclaurin Power Series

► **Basic definitions.** Let $f(x)$ be an infinitely differentiable function at a point x_0. The *Taylor series* for this function is the power series

$$\sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(x_0)(x-x_0)^n = f(x_0) + f'(x_0)(x-x_0) + \frac{1}{2} f''(x_0)(x-x_0)^2 + \cdots,$$

where $0! = 1$ and $f^{(0)}(x_0) = f(x_0)$.

A special case of the Taylor series (for $x_0 = 0$) is the *Maclaurin series*:

$$\sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(0)x^n = f(0) + f'(0)x + \frac{1}{2} f''(0)x^2 + \cdots.$$

A formal Taylor series (Maclaurin series) for a function $f(x)$ may be:

1) divergent for $x \neq x_0$,
2) convergent in a neighborhood of x_0 to a function different from $f(x)$,
3) convergent in a neighborhood of x_0 to the function $f(x)$.

In the last case, one says that $f(x)$ is *expandable in a Taylor series* in the said neighborhood, and one writes

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(x_0)(x-x_0)^n.$$

► **Conditions of expansion in Taylor series.** A *necessary and sufficient condition for a function $f(x)$ to be represented by its Taylor series* in a neighborhood of a point x_0 is that the remainder term in the Taylor formula* should tend to zero as $n \to \infty$ in this neighborhood of x_0.

In order that $f(x)$ could be represented by its Taylor series in a neighborhood of x_0, it suffices that all its derivatives in that neighborhood be bounded by the same constant, $|f^{(n)}(x)| \le M$ for all n.

Uniqueness of the Taylor series expansion. If a function $f(x)$ is representable by the sum of a power series, the coefficients of this series are determined uniquely (since this series is the Taylor series of $f(x)$ and its coefficients have the form $\dfrac{f^{(n)}(x_0)}{n!}$, where $n = 0, 1, 2, \ldots$). Therefore, in problems of representing a function by a power series, the answer does not depend on the method adopted for this purpose.

► **Representation of some functions by the Maclaurin series.** The following representations of elementary functions by Maclaurin series are often used in applications:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots;$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^{n-1} \frac{x^{2n-1}}{(2n-1)!} + \cdots;$$

---

* Different representations of the remainder in the Taylor formula are given in Subsection M6.2.4.

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + \cdots ;$$
$$\sinh x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots + \frac{x^{2n-1}}{(2n-1)!} + \cdots ;$$
$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots + \frac{x^{2n}}{(2n)!} + \cdots ;$$
$$(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!}x^2 + \cdots + \frac{\alpha(\alpha-1)\ldots(\alpha-n+1)}{n!}x^n + \cdots ;$$
$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n+1}\frac{x^n}{n} + \cdots ;$$
$$\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \cdots + (-1)^{n+1}\frac{x^{2n-1}}{2n-1} + \cdots .$$

The first five series are convergent for $-\infty < x < \infty$ ($R = \infty$), and the other series have unit radius of convergence, $R = 1$.

M8.3.3. Operations with Power Series. Summation Formulas for Power Series

► **Addition, subtraction, multiplication, and division of power series.**

1. *Addition and subtraction of power series*. Two series $\sum_{n=0}^{\infty} a_n x^n$ and $\sum_{n=0}^{\infty} b_n x^n$ with convergence radii R_a and R_b, respectively, admit term-by-term addition and subtraction on the intersection of their convergence intervals:

$$\sum_{n=0}^{\infty} a_n x^n \pm \sum_{n=0}^{\infty} b_n x^n = \sum_{n=0}^{\infty} c_n x^n, \qquad c_n = a_n \pm b_n.$$

The radius of convergence of the resulting series satisfies the inequality $R_c \ge \min[R_a, R_b]$.

2. *Multiplication of power series*. Two series $\sum_{n=0}^{\infty} a_n x^n$ and $\sum_{n=0}^{\infty} b_n x^n$, with the respective convergence radii R_a and R_b, can be multiplied on the intersection of their convergence intervals, and their product has the form

$$\biggl(\sum_{n=0}^{\infty} a_n x^n\biggr)\biggl(\sum_{n=0}^{\infty} b_n x^n\biggr) = \sum_{n=0}^{\infty} c_n x^n, \qquad c_n = \sum_{k=0}^{n} a_k b_{n-k}.$$

The convergence radius of the product satisfies the inequality $R_c \ge \min[R_a, R_b]$.

3. *Division of power series*. The ratio of two power series $\sum_{n=0}^{\infty} a_n x^n$ and $\sum_{n=0}^{\infty} b_n x^n$, $b_0 \neq 0$, with convergence radii R_a and R_b can be represented as a power series

$$\frac{\sum_{n=0}^{\infty} a_n x^n}{\sum_{n=0}^{\infty} b_n x^n} = c_0 + c_1 x + c_2 x^2 + \cdots = \sum_{n=0}^{\infty} c_n x^n, \qquad \text{(M8.3.3.1)}$$

whose coefficients can be found, by the method of indefinite coefficients, from the relation

$$(a_0 + a_1x + a_2x^2 + \cdots) = (b_0 + b_1x + b_2x^2 + \cdots)(c_0 + c_1x + c_2x^2 + \cdots).$$

Thus, for the unknown c_n, we obtain a triangular system of linear algebraic equations

$$a_n = \sum_{k=0}^{n} b_k c_{n-k}, \qquad n = 0,\ 1,\ \dots\ ,$$

which is solved consecutively, starting from the first equation:

$$c_0 = \frac{a_0}{b_0}, \quad c_1 = \frac{a_1b_0 - a_0b_1}{b_0^2}, \quad c_n = \frac{a_n}{b_0} - \frac{1}{b_0}\sum_{k=1}^{n} b_k c_{n-k}, \quad n = 2,\ 3,\ \dots$$

The convergence radius of the series (M8.3.3.1) is determined by the formula

$$R_1 = \min\Bigl[R_a, \frac{\rho}{M+1}\Bigr],$$

where ρ is any constant such that $0 < \rho < R_b$; ρ can be chosen arbitrarily close to R_b; and M is the least upper bound of the quantities $|b_m/b_0|\rho^m$ ($m = 1,\ 2,\ \dots$), so that $|b_m/b_0|\rho^m \le M$ for all m.

► **Composition of functions representable by power series.** Consider a power series

$$z = f(y) = a_0 + a_1y + a_2y^2 + \cdots = \sum_{n=0}^{\infty} a_n y^n \tag{M8.3.3.2}$$

with convergence radius R. Let the variable y be a function of x that can be represented by a power series

$$y = \varphi(x) = b_0 + b_1x + a_2x^2 + \cdots = \sum_{n=0}^{\infty} b_n x^n \tag{M8.3.3.3}$$

with convergence radius r. It is required to represent z as a power series of x and find the convergence radius of this series.

Formal substitution of (M8.3.3.3) into (M8.3.3.2) yields

$$z = f\bigl(\varphi(x)\bigr) = \sum_{n=0}^{\infty} a_n\Bigl(\sum_{k=0}^{\infty} b_k x^k\Bigr)^n = A_0 + A_1x + A_2x^2 + \cdots = \sum_{n=0}^{\infty} A_n x^n, \tag{M8.3.3.4}$$

where

$$\begin{aligned}
A_0 &= a_0 + a_1b_0 + a_2b_0^2 + \cdots\ ,\\
A_1 &= a_1b_1 + 2a_2b_0b_1 + 3a_3b_0^2b_1 + \cdots\ ,\\
A_2 &= a_1b_2 + a_2(b_1^2 + 2b_0b_2) + 3a_3(b_0b_1^2 + b_0^2b_2) + \cdots\ ,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots .
\end{aligned}$$

THEOREM ON CONVERGENCE OF SERIES (M8.3.3.4).

(i) If series (M8.3.3.2) is convergent for all y (i.e., $R = \infty$), then the convergence radius of series (M8.3.3.4) coincides with the convergence radius r of series (M8.3.3.3).

(ii) If $0 \le |b_0| < R$, then series (M8.3.3.4) is convergent on the interval $(-R_1, R_1)$, where

$$R_1 = \frac{(R - |b_0|)\rho}{M + R - |b_0|},$$

and ρ is an arbitrary constant such that $0 < \rho < r$; ρ can be chosen arbitrarily close to r; and M is the least upper bound of the quantities $|b_m|\rho^m$ ($m = 1,\ 2,\ \dots$), so that $|b_m|\rho^m \le M$ for all m.

(iii) If $|b_0| > R$, then series (M8.3.3.4) is divergent.

Remark. Case (i) is realized if, for instance, (M8.3.3.2) has finitely many terms.

▶ **Simplest summation formulas for power series.** Suppose that the sum of a power series is known,

$$\sum_{k=0}^{\infty} a_k x^k = S(x). \tag{M8.3.3.6}$$

Then, using term-by-term integration (on the convergence interval), one can find the following sums:

$$\begin{aligned}
&\sum_{k=0}^{\infty} a_k k^m x^k = \left(x\frac{d}{dx}\right)^m S(x);\\
&\sum_{k=0}^{\infty} a_k (nk+m) x^{nk+m-1} = \frac{d}{dx}\bigl[x^m S(x^n)\bigr];\\
&\sum_{k=0}^{\infty} \frac{a_k}{nk+m} x^{nk+m} = \int_0^x x^{m-1} S(x^n)\,dx, \qquad n>0,\ \ m>0;\\
&\sum_{k=0}^{\infty} a_k \frac{nk+s}{nk+m} x^{nk+s} = x\frac{d}{dx}\left[x^{s-m}\int_0^x x^{m-1} S(x^n)\,dx\right], \qquad n>0,\ \ m>0;\\
&\sum_{k=0}^{\infty} a_k \frac{nk+m}{nk+s} x^{nk+s} = \int_0^x x^{s-m}\frac{d}{dx}\Bigl[x^m S(x^n)\Bigr]\,dx, \qquad n>0,\ \ s>0.
\end{aligned} \tag{M8.3.3.7}$$

Example 2. Let us find the sum of the series $\sum\limits_{k=0}^{\infty} kx^{k-1}$.

We start with the well-known formula for the sum of an infinite geometrical progression:

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \qquad (|x|<1).$$

This series is a special case of (M8.3.3.6) with $a_k = 1$, $S(x) = 1/(1-x)$. The series $\sum\limits_{k=0}^{\infty} kx^{k-1}$ can be obtained from the left-hand side of the second formula in (M8.3.3.7) for $m = 0$ and $n = 1$. Substituting $S(x) = 1/(1-x)$ into the right-hand side of that formula, we get

$$\sum_{k=0}^{\infty} kx^{k-1} = \frac{d}{dx}\frac{1}{1-x} = \frac{1}{(1-x)^2} \qquad (|x|<1).$$

M8.4. Fourier Series

M8.4.1. Representation of 2π-Periodic Functions by Fourier Series. Main Results

▶ **Dirichlet theorem on representation of a function by Fourier series.** A function $f(x)$ is said to satisfy the *Dirichlet conditions* on an interval (a, b) if:

1) this interval can be divided into finitely many intervals on which $f(x)$ is monotone and continuous;

2) at any discontinuity point x_0 of the function, there exist finite one-sided limits $f(x_0+0)$ and $f(x_0-0)$.

DIRICHLET THEOREM. *Any 2π-periodic function that satisfies the Dirichlet conditions on the interval $(-\pi,\pi)$ can be represented by its Fourier series*

$$f(x)=\frac{a_0}{2}+\sum_{n=1}^{\infty}\bigl(a_n\cos nx+b_n\sin nx\bigr) \qquad \text{(M8.4.1.1)}$$

whose coefficients are defined by the Euler–Fourier formulas

$$\begin{aligned} a_n&=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\cos nx\,dx, && n=0,\ 1,\ 2,\dots,\\ b_n&=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\sin nx\,dx, && n=1,\ 2,\ 3,\dots \end{aligned} \qquad \text{(M8.4.1.2)}$$

At the points of continuity of $f(x)$, the Fourier series converges to $f(x)$, and at any discontinuity point x_0, the series converges to $\frac{1}{2}[f(x_0+0)+f(x_0-0)]$.

The coefficients a_n and b_n of the series (M8.4.1.1) are called the *Fourier coefficients*.

Remark. Instead of the integration limits $-\pi$ and π in (M8.4.1.2), one can take c and $c+2\pi$, where c is an arbitrary constant.

▶ **Lipschitz and Dirichlet–Jordan convergence criteria for Fourier series.** LIPSCHITZ CRITERION. *Suppose that $f(x)$ is continuous at a point x_0 and for sufficiently small $\varepsilon>0$ satisfies the inequality $|f(x_0\pm\varepsilon)-f(x_0)|\le L\varepsilon^{\sigma}$, where L and σ are constants, $0<\sigma\le 1$. Then the representation (M8.4.1.1)–(M8.4.1.2) holds at $x=x_0$.*

In particular, the conditions of the Lipschitz criterion hold for continuous piecewise differentiable functions.

Remark. The Fourier series of a continuous periodic function with no additional conditions (for instance, of its regularity) may happen to be divergent at infinitely many (even uncountably many) points.

DIRICHLET–JORDAN CRITERION. *Suppose that $f(x)$ is a function of bounded variation on some interval $(x_0-h,x_0+h)\in(-\pi,\pi)$ (i.e., $f(x)$ can be represented as a difference of two monotonically increasing functions). Then the Fourier series (M8.4.1.1)–(M8.4.1.2) of the function $f(x)$ at the point x_0 converges to the value $\frac{1}{2}[f(x_0+0)+f(x_0-0)]$.*

▶ **Asymptotic properties of Fourier coefficients.**

1°. Fourier coefficients of an absolutely integrable function tend to zero as n goes to infinity: $a_n\to 0$ and $b_n\to 0$ as $n\to\infty$.

2°. Fourier coefficients of a continuous 2π-periodic function have the following limit properties:

$$\lim_{n\to\infty}(na_n)=0,\qquad \lim_{n\to\infty}(nb_n)=0,$$

i.e., $a_n=o(1/n)$ and $b_n=o(1/n)$.

3°. If a continuous periodic function is continuously differentiable up to the order $m-1$ inclusively, then its Fourier coefficients have the following limit properties:

$$\lim_{n\to\infty}(n^m a_n)=0,\qquad \lim_{n\to\infty}(n^m b_n)=0,$$

i.e., $a_n=o\bigl(n^{-m}\bigr)$ and $b_n=o\bigl(n^{-m}\bigr)$.

M8.4.2. Fourier Expansions of Periodic, Nonperiodic, Even, and Odd Functions

► **Expansion of 2l-periodic and nonperiodic functions in Fourier series.**

1°. The case of $2l$-periodic functions can be easily reduced to that of 2π-periodic functions by changing the variable x to $z = \dfrac{\pi x}{l}$. In this way, all the results described above for 2π-periodic functions can be easily extended to $2l$-periodic functions.

The Fourier expansion of a $2l$-periodic function $f(x)$ has the form

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \Bigl(a_n \cos\frac{n\pi x}{l} + b_n \sin\frac{n\pi x}{l}\Bigr), \qquad \text{(M8.4.2.1)}$$

where

$$a_n = \frac{1}{l}\int_{-l}^{l} f(x)\cos\frac{n\pi x}{l}\,dx, \qquad b_n = \frac{1}{l}\int_{-l}^{l} f(x)\sin\frac{n\pi x}{l}\,dx. \qquad \text{(M8.4.2.2)}$$

2°. A nonperiodic (aperiodic) function $f(x)$ defined on the interval $(-l, l)$ can also be represented by a Fourier series (M8.4.2.1)–(M8.4.2.2); however, outside that interval, the sum of that series $S(x)$ may differ from $f(x)$*.

► **Fourier expansion of even and odd functions.**

1°. Let $f(x)$ be an even function, i.e., $f(-x) = f(x)$. Then the Fourier expansion of $f(x)$ on the interval $(-l, l)$ has the form of the *cosine Fourier series*:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\frac{n\pi x}{l},$$

where the Fourier coefficients have the form

$$a_n = \frac{2}{l}\int_0^l f(x)\cos\frac{n\pi x}{l}\,dx \qquad (b_n = 0).$$

2°. Let $f(x)$ be an odd function, i.e., $f(-x) = -f(x)$. Then the Fourier expansion of $f(x)$ on the interval $(-l, l)$ has the form of the *sine Fourier series*:

$$f(x) = \sum_{n=1}^{\infty} b_n \sin\frac{n\pi x}{l},$$

where the Fourier coefficients have the form

$$b_n = \frac{2}{l}\int_0^l f(x)\sin\frac{n\pi x}{l}\,dx \qquad (a_n = 0).$$

---

* The sum $S(x)$ is a $2l$-periodic function defined for all x, but $f(x)$ may happen to be nonperiodic or even undefined outside the interval $(-l, l)$.

Example. Let us find the Fourier expansion of the function $f(x) = x$ on the interval $(-\pi, \pi)$.

Taking $l = \pi$ and $f(x) = x$ in the formula for the Fourier coefficients and integrating by parts, we obtain

$$b_n = \frac{2}{\pi}\int_0^\pi x\sin(nx)\,dx = \frac{2}{\pi}\left(-\frac{1}{n}x\cos(nx)\Big|_0^\pi + \frac{1}{n}\int_0^\pi \cos(nx)\,dx\right) = -\frac{2}{n}\cos(n\pi) = (-1)^{n+1}\frac{2}{n}.$$

Therefore, the Fourier expansion of $f(x) = x$ has the form

$$f(x) = 2\sum_{n=1}^{\infty}(-1)^{n+1}\frac{\sin(nx)}{n} \qquad (-\pi < x < \pi).$$

3°. If $f(x)$ is defined on the interval $(0, l)$ and satisfies the Dirichlet conditions, it can be represented by the cosine Fourier series, as well as the sine Fourier series (with the help of the above formulas). The cosine Fourier expansion of $f(x)$ on the interval $(0, l)$ corresponds to the extension of $f(x)$ to the interval $(-l, 0)$ as an even function: $f(-x) = f(x)$. The sine Fourier expansion of $f(x)$ on $(0, l)$ corresponds to the extension of $f(x)$ to the interval $(-l, 0)$ as an odd function: $f(-x) = -f(x)$. Both series on the interval $(0, l)$ give the values of $f(x)$ at points of its continuity and the value $\frac{1}{2}[f(x_0+0) + f(x_0-0)]$ at points of its discontinuity; outside the interval $(0, l)$, these two series represent different functions.

► **Fourier series in complex form.** The complex Fourier expansion of a function $f(x)$ on an interval $(-l, l)$ has the form

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{i\omega_n x},$$

where

$$\omega_n = \frac{n\pi}{l}, \quad c_n = \frac{1}{2l}\int_{-l}^{l} f(x)e^{-i\omega_n x}\,dx; \quad n = 0,\ \pm1,\ \pm2,\ \ldots$$

The expressions $e^{i\omega_n x}$ are called *complex harmonics*, the coefficients c_n are *complex amplitudes*, ω_n are *wave numbers* of the function $f(x)$, and the set of all wave numbers $\{\omega_n\}$ is called the *discrete spectrum* of the function.

M8.4.3. Criteria of Uniform and Mean-Square Convergence of Fourier Series

► **Criteria of uniform convergence of Fourier series.**

LIPSCHITZ CRITERION. *The Fourier series of a function $f(x)$ converges uniformly to that function on an interval $[-l, l]$ if on a wider interval $[-L, L]$ $(-L < -l < l < L)$ the following inequality holds:*

$$|f(x_1) - f(x_2)| \le K|x_1 - x_2|^\sigma \quad \text{for all} \quad x_1, x_2 \in [-L, L],$$

where K and σ are constants, $0 < \sigma \le 1$.

Corollary. The Fourier series of a continuous function $f(x)$ converges uniformly to that function on an interval $[-l, l]$ if on a wider interval the function $f(x)$ has a bounded derivative $f'(x)$.

For any continuously differentiable $2l$-periodic function $f(x)$, its Fourier series [defined by formulas (M8.4.2.1)–(M8.4.2.2)] is uniformly convergent to $f(x)$.

► Fourier series of square-integrable functions. Parseval identity.

1°. For a continuous 2π-periodic function $f(x)$, its Fourier series (M8.4.1.1)–(M8.4.1.2) converges to $f(x)$ in mean square, i.e.,

$$\int_{-\pi}^{\pi}[f(x)-f_n(x)]^2\,dx\to 0\quad\text{as}\quad n\to\infty,$$

where $f_n(x)=\frac{1}{2}a_0+\sum\limits_{k=1}^{n}(a_k\cos kx+b_k\sin kx)$ is a partial sum of the Fourier series.

2°. If $f(x)$ is integrable on the segment $[-\pi,\pi]$ and the integral $\int_{-\pi}^{\pi}f^2(x)\,dx$ exists as an improper integral with finitely many singularities, then the Fourier series (M8.4.1.1)–(M8.4.1.2) is mean-square convergent to $f(x)$.

3°. Let $f(x)\in L^2[-\pi,\pi]$ be a square-integrable function on the segment $[-\pi,\pi]$. Then its Fourier series (M8.4.1.1)–(M8.4.1.2) is mean-square convergent to $f(x)$, and the *Parseval identity* holds:

$$\frac{a_0^2}{2}+\sum_{n=1}^{\infty}(a_n^2+b_n^2)=\frac{1}{\pi}\int_{-\pi}^{\pi}f^2(x)\,dx,$$

where a_n, b_n are defined by (M8.4.1.2). Note that the functions considered in Items 1° and 2° belong to $L^2[-\pi,\pi]$.

Bibliography for Chapter M8

Brannan, D., *A First Course in Mathematical Analysis*, Cambridge University Press, Cambridge, England, 2006.

Bromwich, T. J. I., *Introduction to the Theory of Infinite Series*, American Mathematical Society, Providence, Rhode Island, 2005.

Bronshtein, I. N. and Semendyayev, K. A., *Handbook of Mathematics, 4th Edition*, Springer-Verlag, Berlin, 2004.

Danilov, V. L., Ivanova, A. N., et al., *Mathematical Analysis (Functions, Limits, Series, Continued Fractions)* [in Russian], Fizmatlit Publishers, Moscow, 1961.

Davis, H. F., *Fourier Series and Orthogonal Functions*, Dover Publications, New York, 1989.

Fedoryuk, M. V., *Asymptotics, Integrals, Series* [in Russian], Nauka, Moscow, 1987.

Fichtenholz, G. M., *Functional Series*, Taylor & Francis, London, 1970.

Fikhtengol'ts (Fichtenholz), G. M., *A Course in Differential and Integral Calculus, Vols. 2 and 3* [in Russian], Nauka Publishers, Moscow, 1969.

Gradshteyn, I. S. and Ryzhik, I. M., *Tables of Integrals, Series, and Products, 6th Edition*, Academic Press, New York, 2000.

Hansen, E. R., *A Table of Series and Products*, Prentice Hall, Englewood Cliffs, London, 1975.

Hirschman, I., *Infinite Series*, Greenwood Press, New York, 1978.

Hyslop, J. M., *Infinite Series*, Dover Publications, New York, 2006.

Jolley, L. B. W., *Summation of Series*, Dover Publications, New York, 1961.

Knopp, K., *Infinite Sequences and Series*, Dover Publications, New York, 1956.

Knopp, K., *Theory and Applications of Infinite Series*, Dover Publications, New York, 1990.

Mangulis, V., *Handbook of Series for Scientists and Engineers*, Academic Press, New York, 1965.

Pinkus, A. and Zafrany, S., *Fourier Series and Integral Transforms*, Cambridge University Press, Cambridge, England, 1997.

Prudnikov, A. P., Brychkov, Yu. A., and Marichev, O. I., *Integrals and Series, Vol. 1, Elementary Functions*, Gordon & Breach, New York, 1986.

Zhizhiashvili, L.V., *Trigonometric Fourier Series and Their Conjugates*, Kluwer Academic, Dordrecht, 1996.

Zygmund, A., *Trigonometric Series, 3rd Edition*, Cambridge University Press, Cambridge, England, 2003.

Chapter M9

Functions of Complex Variable

M9.1. Complex Numbers

M9.1.1. Definition of a Complex Number. Arithmetic Operations with Complex Numbers

► **Definition of a complex number. Geometric interpretation.** The set of complex numbers is an extension of the set of real numbers. An expression of the form $z = x + iy$, where x and y are real numbers, is called a *complex number*, and the symbol i is called the *imaginary unit*, which possesses the property $i^2 = -1$. The numbers x and y are called, respectively, the *real* and *imaginary parts* of z and denoted by

$$x = \operatorname{Re} z \quad \text{and} \quad y = \operatorname{Im} z.$$

The complex number $x + i0$ is identified with real number x, and the number $0 + iy$ is denoted by iy and is said to be *pure imaginary*. Two complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ are *equal* if $x_1 = x_2$ and $y_1 = y_2$.

The complex number $\bar{z} = x - iy$ is said to be *conjugate* to the number z.

A complex number $z = x + iy$ can be conveniently represented as a point (x, y) in a two-dimensional Cartesian coordinate system (see Fig. M9.1). The axes OX and OY are called the *real* and *imaginary* axis, respectively, and the plane OXY is called the *complex plane*. The notions of a complex number and a point on the complex plane are identical.

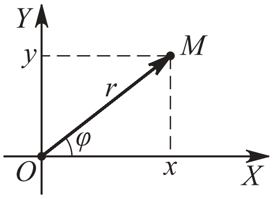

Figure M9.1. Geometric interpretation of a complex number.

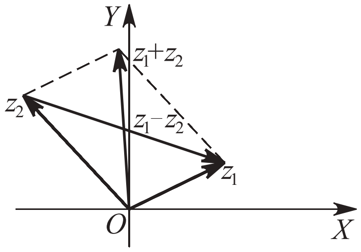

Figure M9.2. The sum and difference of complex numbers.

► **Addition, subtraction, multiplication, and division of complex numbers.** The *sum or difference* of complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ is defined as the number

$$z_1 \pm z_2 = x_1 \pm x_2 + i(y_1 \pm y_2).$$

The *geometric meaning of the operations of addition and subtraction of complex numbers* is as follows: the sum and the difference of complex numbers z_1 and z_2 are the vectors equal to the directed diagonals of the parallelogram spanned by the vectors z_1 and z_2 (Fig. M9.2). The following inequalities hold (Fig. M9.2):

$$|z_1 + z_2| \le |z_1| + |z_2|, \quad |z_1 - z_2| \ge \big||z_1| - |z_2|\big|. \tag{M9.1.2.1}$$

Inequalities (M9.1.2.1) become equalities if and only if the arguments of the complex numbers z_1 and z_2 coincide (i.e., $\arg z_1 = \arg z_2$; see Subsection M9.1.2) or one of the numbers is zero.

The *product* z_1z_2 of complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ is defined to be the number

$$z_1z_2 = (x_1x_2 - y_1y_2) + i(x_1y_2 + x_2y_1).$$

The product of a complex number $z = x + iy$ by its conjugate is always nonnegative:

$$z\bar{z} = x^2 + y^2.$$

If $z_2 \neq 0$, then the *quotient* of z_1 and z_2 is defined as

$$\frac{z_1}{z_2} = \frac{x_1x_2 + y_1y_2}{x_2^2 + y_2^2} + i\frac{x_2y_1 - x_1y_2}{x_2^2 + y_2^2}. \qquad \text{(M9.1.1.1)}$$

Relation (M9.1.1.1) can be obtained by multiplying the numerator and the denominator of the fraction z_1/z_2 by $\bar{z}_2$.

M9.1.2. Trigonometric Form of Complex Numbers. Powers and Radicals

► **Modulus and argument of a complex number.** There is a one-to-one correspondence between complex numbers $z = x + iy$ and points M with coordinates (x, y) on the plane with a Cartesian rectangular coordinate system OXY or with vectors $\overrightarrow{OM}$ connecting the origin O with M (Fig. M9.1). The length r of the vector $\overrightarrow{OM}$ is called the *modulus* (also *magnitude* and *absolute value*) of the number z and is denoted by $r = |z|$, and the angle φ formed by the vector $\overrightarrow{OM}$ and the positive direction of the OX-axis is called the *argument* (also *phase*) of the number z and is denoted by $\varphi = \operatorname{Arg} z$.

The modulus of a complex number is determined by the formula

$$|z| = \sqrt{x^2 + y^2}.$$

The argument $\operatorname{Arg} z$ is determined up to a multiple of 2π, $\operatorname{Arg} z = \arg z + 2k\pi$, where k is an arbitrary integer and $\arg z$ is the *principal value* of $\operatorname{Arg} z$ determined by the condition $-\pi < \arg z \leq \pi$. The principal value $\arg z$ is given by the formula

$$\arg z = \begin{cases} \arctan(y/x) & \text{for } x > 0, \\ \pi + \arctan(y/x) & \text{for } x < 0,\ y \geq 0, \\ -\pi + \arctan(y/x) & \text{for } x < 0,\ y < 0, \\ \pi/2 & \text{for } x = 0,\ y > 0, \\ -\pi/2 & \text{for } x = 0,\ y < 0. \end{cases}$$

For $z = 0$, $\operatorname{Arg} z$ is undefined.

► **Trigonometric form of complex numbers.** Since $x = r\cos\varphi$ and $y = r\sin\varphi$, it follows that the complex number can be written in the *trigonometric* (or *polar*) *form*

$$z = x + iy = r(\cos\varphi + i\sin\varphi).$$

For two complex numbers written in trigonometric form, $z_1 = r_1(\cos\varphi_1 + i\sin\varphi_1)$ and $z_2 = r_2(\cos\varphi_2 + i\sin\varphi_2)$, the following arithmetic rules are valid:

$$z_1z_2 = r_1r_2\big[\cos(\varphi_1 + \varphi_2) + i\sin(\varphi_1 + \varphi_2)\big], \quad \frac{z_1}{z_2} = \frac{r_1}{r_2}\big[\cos(\varphi_1 - \varphi_2) + i\sin(\varphi_1 - \varphi_2)\big].$$

In the latter formula, it is assumed that $z_2 \neq 0$.

► **Powers and radicals.** For any positive integer n, the nth power of z is calculated by *de Moivre's formula*

$$z^n = r^n(\cos n\varphi + i\sin n\varphi),$$

For $z \neq 0$ and positive integer n, there are exactly n distinct values of the nth root of $z = r(\cos\varphi + i\sin\varphi)$, which are determined by

$$\sqrt[n]{z} = z^{1/n} = \sqrt[n]{r}\Big(\cos\frac{\varphi + 2k\pi}{n} + i\sin\frac{\varphi + 2k\pi}{n}\Big) \quad (k = 0, 1, 2, \dots, n-1).$$

Example. Let us find all values of $\sqrt[3]{i}$.

Let us represent the complex number $z = i$ in trigonometric form. We have $r = |z| = 1$ and $\varphi = \arg z = \frac{\pi}{2}$. The distinct values of the cube root are calculated by the formula

$$\omega_k = \sqrt[3]{1}\left(\cos\frac{\frac{\pi}{2} + 2\pi k}{3} + i\sin\frac{\frac{\pi}{2} + 2\pi k}{3}\right) \quad (k = 0, 1, 2),$$

so that

$$\omega_0 = \cos\frac{\pi}{6} + i\sin\frac{\pi}{6} = \frac{\sqrt{3}}{2} + i\frac{1}{2},$$
$$\omega_1 = \cos\frac{5\pi}{6} + i\sin\frac{5\pi}{6} = -\frac{\sqrt{3}}{2} + i\frac{1}{2},$$
$$\omega_2 = \cos\frac{3\pi}{2} + i\sin\frac{3\pi}{2} = -i.$$

The roots are shown in Fig. M9.3.

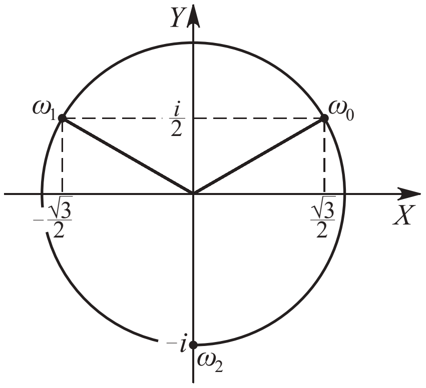

Figure M9.3. The roots of $\sqrt[3]{i}$.

M9.2. Functions of Complex Variables

M9.2.1. Basic Concepts. Differentiation of a Function of a Complex Variable

► **Some concepts and definitions.** A subset D of the complex plane such that each point of D has a neighborhood contained in D (i.e., D is open) and two arbitrary points of D can be connected by a broken line lying in D (i.e., D is connected) is called a *domain* in the complex plane. Each point of D is its *interior point*. A point that does not lie in D but whose arbitrary neighborhood contains points of D is called a *boundary point* of D. The set of all boundary points of D is called the *boundary* of D. The union of a domain D with its boundary is called a *closed domain* and denoted by $\overline{D}$. The boundary of a domain can consist of finitely many closed curves, segments, and points; the curves and cuts are assumed to be piecewise smooth.

The simplest examples of domains are neighborhoods of points on the complex plane. A *neighborhood* of a point a on the complex plane is understood as the set of points z such that $|z-a| < R$, i.e., the interior of the disk of radius $R > 0$ centered at the point a. The *extended complex plane* is obtained by augmenting the complex plane with the fictitious *point at infinity*. A *neighborhood of the point at infinity* is understood as the set of points z such that $|z| > R$ (including the point at infinity itself).

If to each point z of a domain D there corresponds a single point w (resp., a number of points w), then one says that there is a *single-valued* (resp., *multi-valued*) *function* $w = f(z)$ defined on the domain D. If we set $z = x + iy$ and $w = u + iv$, then defining a function $w = f(z)$ of the complex variable z is equivalent to defining two functions $\operatorname{Re} f = u = u(x, y)$ and $\operatorname{Im} f = v = v(x, y)$ of two real variables. If the function $w = f(z)$ is single-valued on D and the images of distinct points of D are distinct, then the mapping determined by this function is said to be *schlicht*. The notions of boundedness, limit, and continuity for single-valued functions of a complex variable do not differ from the corresponding notions for real functions of two real variables.

► **Differentiability. The Cauchy–Riemann conditions.** Let a single-valued function $w = f(z)$ be defined in a neighborhood of a point z. If there exists a limit

$$\lim_{h\to 0} \frac{f(z+h)-f(z)}{h} = f'_z(z),$$

then the function $w = f(z)$ is said to be *differentiable* at the point z and $f'_z(z)$ is called its *derivative* at the point z.

Cauchy–Riemann conditions. If the functions $u(x, y) = \operatorname{Re} f(z)$ and $v(x, y) = \operatorname{Im} f(z)$ are differentiable at a point (x, y), then the *Cauchy–Riemann conditions*

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \tag{M9.2.1.1}$$

are necessary and sufficient for the function $w = f(z)$ to be differentiable at the point $z = x + iy$.

If the function $w = f(z)$ is differentiable, then

$$w'_z = u_x + iv_x = v_y - iu_y = u_x - iu_y = v_y + iv_x,$$

where the subscripts x and y indicate the corresponding partial derivatives.

Remark. The Cauchy–Riemann conditions are sometimes also called the d'Alembert–Euler conditions.

The rules for arithmetic operations on the derivatives and those for taking the derivative of a composite function and the inverse function (if it exists) have exactly the same form as in the case of functions of a real variable:

1. $\bigl[\alpha f_1(z) \pm \beta f_2(z)\bigr]'_z = \alpha[f_1(z)]'_z \pm \beta[f_2(z)]'_z$, where α and β are arbitrary complex constants.
2. $\bigl[f_1(z) f_2(z)\bigr]'_z = [f_1(z)]'_z f_2(z) + f_1(z)[f_2(z)]'_z$.
3. $\left[\dfrac{f_1(z)}{f_2(z)}\right]'_z = \dfrac{[f_1(z)]'_z f_2(z) - f_1(z)[f_2(z)]'_z}{f_2^2(z)} \qquad (f_2(z) \neq 0).$
4. If a function $w = f(z)$ is differentiable at a point z and a function $W = F(w)$ is differentiable at the point $w = f(z)$, then the composite function $W = F(f(z))$ is differentiable at the point z and $W'_z = [F(f(z))]'_z = F'_f(f) f'_z(z)$.

5. If a function $w = f(z)$ is differentiable at a point z and the inverse $z = g(w) \equiv f^{-1}(w)$ exists and is differentiable at the point w, then

$$[f^{-1}(w)]'_w = \frac{1}{f'_z(z)\big|_{z=f^{-1}(w)}} \qquad (f'_z(z) \neq 0).$$

► **Analyticity. The maximum modulus principle. The Liouville's theorem.** A single-valued function differentiable in some neighborhood of a point z_0 is said to be *analytic* (*regular*, *holomorphic*) at this point.

A function $w = f(z)$ is analytic at a point z_0 if and only if it can be represented by a power series

$$f(z) = \sum_{k=0}^{\infty} c_k (z - z_0)^k$$

converging in some neighborhood of z_0.

A function analytic at each point of the domain D is said to be *analytic* in D.

A function $w = f(z)$ is said to be *analytic at the point at infinity* if the function $F(z) = f(1/z)$ is analytic at the point $z = 0$.

A function $w = f(z)$ is analytic at the point at infinity if and only if this function can be represented by a power series

$$f(z) = \sum_{k=0}^{\infty} b_k z^{-k}$$

converging for sufficiently large $|z|$.

If a function $w = f(z)$ is analytic at a point z_0 and $f'_z(z_0) \neq 0$, then $f(z)$ has an analytic inverse function $z(w)$ defined in a neighborhood of the point $w_0 = f(z_0)$. If a function $w = f(z)$ is analytic at a point z_0 and the function $W = F(w)$ is analytic at the point $w_0 = f(z_0)$, then the composite function $W = F[f(z)]$ is analytic at the point z_0. If a function is analytic in a domain D and continuous in $\overline{D}$, then its value at any interior point of the domain is uniquely determined by its values on the boundary of the domain. The analyticity of a function at a point implies the existence and analyticity of its derivatives of arbitrary order at this point.

Single-valued functions, as well as single-valued branches of multi-valued functions, are analytic everywhere on the domains where they are defined. It follows from (M9.2.1.1) that the real and imaginary parts $u(x, y)$ and $v(x, y)$ of a function analytic in a domain are *harmonic* in this domain, i.e., satisfy the Laplace equation

$$\Delta f = f_{xx} + f_{yy} = 0$$

in this domain.

Remark. If $u(x, y)$ and $v(x, y)$ are two arbitrary harmonic functions, then the function $f(z) = u(x, y) + iv(x, y)$ is not necessarily analytic, since for the analyticity of $f(z)$ the functions $u(x, y)$ and $v(x, y)$ must satisfy the Cauchy–Riemann conditions.

Example 1. The function $w = z^2$ is analytic.

Indeed, since $z = x + iy$, we have $w = (x + iy)^2 = x^2 - y^2 + i2xy$, $u(x, y) = x^2 - y^2$, and $v(x, y) = 2xy$. The Cauchy–Riemann conditions

$$u_x = v_y = 2x, \quad u_y = -v_x = -2y$$

are satisfied at all points of the complex plane, so the function $w = z^2$ is analytic.

Example 2. The function $w=\bar{z}$ is not analytic.

Indeed, since $z=x+iy$, we have $w=x-iy$, $u(x,y)=x$, $v(x,y)=-y$. The Cauchy–Riemann conditions are not satisfied,

$$u_x=1\neq -1=v_y,\qquad u_y=-v_x=0,$$

so the function $w=\bar{z}$ is not analytic.

MAXIMUM MODULUS PRINCIPLE. *If a function $w=f(z)$ that is not identically constant is analytic in a domain D and continuous in $\overline{D}$, then its modulus cannot attain a maximum at an interior point of D.*

LIOUVILLE'S THEOREM. *If a function $w=f(z)$ is analytic and bounded in the entire complex plane, then it is constant.*

Remark. The Liouville theorem can be stated in the following form:
if a function $w=f(z)$ is analytic in the extended complex plane, then it is constant.

► Geometric meaning of the derivative.

Geometric meaning of the absolute value of the derivative. Suppose that a function $w=f(z)$ is analytic at a point z_0 and $f'_z(z_0)\neq 0$. Then the value $|f'_z(z_0)|$ determines the dilatation (similarity) coefficient at the point z_0 under the mapping $w=f(z)$. The value $|f'_z(z_0)|$ is called the *dilatation ratio* if $|f'_z(z_0)|>1$ and the *contraction ratio* if $|f'_z(z_0)|<1$.

Geometric meaning of the argument of the derivative. The argument of the derivative $f'_z(z_0)$ is equal to the angle by which the tangent at the point z_0 to any curve passing through z_0 should be rotated to give the tangent to the image of the curve at the point $w_0=f(z_0)$. For $\varphi=\arg f'_z(z)>0$, the rotation is counterclockwise, and for $\varphi=\arg f'_z(z)<0$, the rotation is clockwise.

► Elementary functions.

1°. Consider the *functions $w=z^n$ and $w=\sqrt[n]{z}$* for positive integer n. The function

$$w=z^n$$

is single-valued. It is schlicht in the sectors $2\pi k/n<\varphi<2\pi(k+1)/n$, $k=0,1,2,\dots$, each of which is transformed by the mapping $w=z^n$ onto the plane w with a cut on the positive real semiaxis.

The function

$$w=\sqrt[n]{z}$$

is an n-valued function for $z\neq 0$, and its value is determined by the value of the argument chosen for the point z. If a closed curve C does not surround the point $z=0$, then, as the point z goes around the entire curve C, the point $w=\sqrt[n]{z}$ for a chosen value of the root also moves along a closed curve and returns to the initial value of the argument. But if the curve C surrounds the origin, then, as the point z goes around the entire curve C in the *positive sense* (in the *counterclockwise* direction), the argument of z increases by 2π and the corresponding point $w=\sqrt[n]{z}$ does not return to the initial position. It will return there only after the point z goes n times around the entire curve C. If a domain D does not contain a closed curve surrounding the point $z=0$, then one can single out n continuous single-valued functions, each of which takes only one of the values $w=\sqrt[n]{z}$; these functions are called the *branches* of the multi-valued function $w=\sqrt[n]{z}$. One cannot single out n separate branches of the function $w=\sqrt[n]{z}$ in any neighborhood of the point $z=0$; accordingly, the point $z=0$ is called a *branch point* of this function.

2°. The *Zhukovskii function*

$$w=\frac{1}{2}\left(z+\frac{1}{z}\right)$$

is defined and single-valued for all $z \neq 0$; it is schlicht in any domain that does not simultaneously contain any points z_1 and z_2 such that $z_1 z_2 = 1$.

3°. The *exponential function* $w = e^z$ is defined by the formula

$$w = e^z = e^{x+iy} = e^x(\cos y + i \sin y).$$

The function $w = e^z$ is analytic everywhere. For the exponential function, the usual differentiation rule is preserved:

$$(e^z)'_z = e^z.$$

The basic property of the exponential function is also preserved:

$$e^{z_1} e^{z_2} = e^{z_1+z_2}.$$

For $x = 0$ and $y = \varphi$, the definition of the exponential function implies the *Euler formula* $e^{i\varphi} = \cos\varphi + i \sin\varphi$, which permits one to write any complex number with modulus r and argument φ in the *exponential form*

$$z = r(\cos\varphi + i \sin\varphi) = re^{i\varphi}.$$

The exponential function is periodic with imaginary period $2\pi i$, and the mapping $w = e^z$ is schlicht in the strip $0 \le y < 2\pi$.

4°. The *logarithm* is defined as the inverse of the exponential function: if $e^w = z$, then

$$w = \operatorname{Ln} z.$$

This function is defined for $z \neq 0$. The logarithm satisfies the following relations:

$$\operatorname{Ln} z_1 + \operatorname{Ln} z_2 = \operatorname{Ln}(z_1 z_2), \quad \operatorname{Ln} z_1 - \operatorname{Ln} z_2 = \operatorname{Ln}\frac{z_1}{z_2},$$
$$\operatorname{Ln}(z^n) = n \operatorname{Ln} z, \quad \operatorname{Ln}\sqrt[n]{z} = \frac{1}{n}\operatorname{Ln} z.$$

The exponential form of complex numbers readily shows that the logarithm is infinite-valued:

$$\operatorname{Ln} z = \ln|z| + i \operatorname{Arg} z = \ln|z| + i \arg z + 2\pi k i, \quad k = 0, \pm1, \pm2, \ldots \tag{M9.2.1.2}$$

The quantity $\ln z = \ln|z| + i \arg z$ is taken to be the principal value of this function. Just as with the function $w = \sqrt[n]{z}$, we see that if the point $z = 0$ is surrounded by a closed curve C, then the point $w = \operatorname{Ln} z$ does not return to its initial position after z goes around C in the positive sense, since the argument of w increases by $2\pi i$. Thus if a domain D does not contain a closed curve surrounding the point $z = 0$, then in D one can single out infinitely many continuous and single-valued branches of the multi-valued function $w = \operatorname{Ln} z$; the differences between the values of these branches at each point of the domain are equal to $2\pi k i$, where k is an integer. This cannot be done in an arbitrary neighborhood of the point $z = 0$, and this point is called a *branch point* of the logarithm.

5°. *Trigonometric functions* are defined in terms of the exponential function as follows:

$$\cos z = \frac{e^{iz} + e^{-iz}}{2}, \qquad \sin z = \frac{e^{iz} - e^{-iz}}{2i},$$
$$\tan z = \frac{\sin z}{\cos z} = -i\frac{e^{iz} - e^{-iz}}{e^{iz} + e^{-iz}}, \qquad \cot z = \frac{\cos z}{\sin z} = i\frac{e^{iz} + e^{-iz}}{e^{iz} - e^{-iz}}.$$

These are properties of the functions $\cos z$ and $\sin z$:

1. They are analytic for any z.
2. The usual differentiation rules are valid:

$$(\sin z)'_z = \cos z, \quad (\cos z)'_z = -\sin z.$$

3. They are periodic with real period $T = 2\pi$.
4. $\sin z$ is an odd function, and $\cos z$ is an even function.
5. In the complex plane, they are unbounded.
6. The usual trigonometric relations hold:

$$\cos^2 z + \sin^2 z = 1, \quad \cos 2z = \cos^2 z - \sin^2 z, \quad \text{etc.}$$

The function $\tan z$ is analytic everywhere except for the points

$$z_k = \frac{\pi}{2} + k\pi, \qquad k = 0,\ \pm1,\ \pm2,\ \ldots,$$

and the function $\cot z$ is analytic everywhere except for the points

$$z_k = k\pi, \qquad k = 0,\ \pm1,\ \pm2,\ \ldots$$

The functions $\tan z$ and $\cot z$ are periodic with real period $T = \pi$.

6°. *Hyperbolic functions* are defined by the formulas

$$\cosh z = \frac{e^z + e^{-z}}{2}, \qquad \sinh z = \frac{e^z - e^{-z}}{2},$$
$$\tanh z = \frac{\sinh z}{\cosh z} = \frac{e^z - e^{-z}}{e^z + e^{-z}}, \qquad \coth z = \frac{\cosh z}{\sinh z} = \frac{e^z + e^{-z}}{e^z - e^{-z}}.$$

For real values of the argument, each of these functions coincides with the corresponding real function. Hyperbolic and trigonometric functions are related by the formulas

$$\cosh z = \cos iz, \quad \sinh z = -i \sin iz, \quad \tanh z = -i \tan iz, \quad \coth z = i \cot iz.$$

7°. *Inverse trigonometric and hyperbolic functions* are expressed via the logarithm and hence are infinite-valued:

$$\operatorname{Arccos} z = -i \operatorname{Ln}(z + \sqrt{z^2 - 1}), \qquad \operatorname{Arcsin} z = -i \operatorname{Ln}(iz + \sqrt{1 - z^2}),$$
$$\operatorname{Arctan} z = -\frac{i}{2} \operatorname{Ln} \frac{1 + iz}{1 - iz}, \qquad \operatorname{Arccot} z = -\frac{i}{2} \operatorname{Ln} \frac{z + i}{z - i},$$
$$\operatorname{arccosh} z = \operatorname{Ln}(z + \sqrt{z^2 - 1}), \qquad \operatorname{arcsinh} z = \operatorname{Ln}(z + \sqrt{z^2 - 1}),$$
$$\operatorname{arctanh} z = \frac{1}{2} \operatorname{Ln} \frac{1 + z}{1 - z}, \qquad \operatorname{arccoth} z = \frac{1}{2} \operatorname{Ln} \frac{z + 1}{z - 1}.$$

The principal value of each of these functions is obtained by choosing the principal value of the corresponding logarithmic function.

8°. The *power function* $w = z^\gamma$ is defined by the relation

$$z^\gamma = e^{\gamma \operatorname{Ln} z}, \qquad \text{(M9.2.1.3)}$$

where $\gamma = \alpha + i\beta$ is an arbitrary complex number. Substituting $z = re^{i\varphi}$ into (M9.2.1.3) yields

$$z^\gamma = e^{\alpha \ln r - \beta(\varphi + 2k\pi)} e^{i\alpha(\varphi + 2k\pi) + i\beta \ln r}, \qquad k = 0,\ \pm1,\ \pm2,\ \ldots \qquad \text{(M9.2.1.4)}$$

It follows from relation (M9.2.1.4) that the function $w = z^\gamma$ has infinitely many values for $\beta \neq 0$.

9°. The *general exponential function* is defined by

$$w = \gamma^z = e^{z\,\mathrm{Ln}\,\gamma} = e^{z\ln|\gamma|}e^{zi\,\mathrm{Arg}\,\gamma}, \tag{M9.2.1.5}$$

where $\gamma = \alpha + i\beta$ is an arbitrary nonzero complex number. The function (M9.2.1.5) is a set of separate mutually independent single-valued functions that differ from one another by the factors $e^{2k\pi iz}$, $k = 0, \pm1, \pm2, \dots$

Example 3. Let us calculate the values of some elementary functions at specific points:

1. $\cos 2i = \frac{1}{2}(e^{2ii} + e^{-2ii}) = \frac{1}{2}(e^2 + e^{-2}) = \cosh 2 \approx 3.7622$.
2. $\ln(-2) = \ln 2 + i\pi$, since $|-2| = 2$ and the principal value of the argument is equal to π.
3. $\mathrm{Ln}(-2)$ is calculated by formula (M9.2.1.2):

$$\mathrm{Ln}(-2) = \ln 2 + i\pi + 2\pi ki = \ln 2 + (1 + 2k)i\pi \quad (k = 0, \pm1, \pm2, \dots).$$

4. $i^i = e^{i\,\mathrm{Ln}\,i} = e^{i(i\frac{\pi}{2}+2\pi ki)} = e^{-\frac{\pi}{2}-2\pi k} \quad (k = 0, \pm1, \pm2, \dots)$.

The main elementary functions $w = f(z) = u(x, y) + iv(x, y)$ of the complex variable $z = x + iy$ are given in Table M9.1.

TABLE M9.1

Main elementary functions $w = f(z) = u(x, y) + iv(x, y)$ of the complex variable $z = x + iy$.

No.	Complex function $w = f(z)$	Algebraic form $f(z) = u(x, y) + iv(x, y)$	Zeros of nth order	Singularities
1	z	$x + iy$	$z = 0,\ n = 1$	$z = \infty$ is a first-order pole
2	z^2	$x^2 - y^2 + i\,2xy$	$z = 0,\ n = 2$	$z = \infty$ is a second-order pole
3	$\dfrac{1}{z-(x_0+iy_0)}$ (x_0, y_0 are real numbers)	$\dfrac{x-x_0}{(x-x_0)^2+(y-y_0)^2} + i\dfrac{-(y-y_0)}{(x-x_0)^2+(y-y_0)^2}$	$z = \infty,\ n = 1$	$z = x_0 + iy_0$ is a first-order pole
4	$\dfrac{1}{z^2}$	$\dfrac{x^2-y^2}{(x^2+y^2)^2} + i\dfrac{-2xy}{(x^2+y^2)^2}$	$z = \infty,\ n = 2$	$z = 0$ is a second-order pole
5	$\sqrt{z}$	$\pm\left[\left(\dfrac{x+\sqrt{x^2+y^2}}{2}\right)^{1/2} + i\left(\dfrac{-x+\sqrt{x^2+y^2}}{2}\right)^{1/2}\right]$	$z = 0$ is a branch point	$z = 0$ is a first-order branch point $z = \infty$ is a first-order branch point
6	e^z	$e^x \cos y + ie^x \sin y$	no zeros	$z = \infty$ is an essential singular point
7	$\mathrm{Ln}\,z$	$\ln\|z\| + i(\arg z + 2k\pi)$, $k = 0, \pm1, \pm2, \dots$	$z = 1,\ n = 1$ (for the branch corresponding to $k = 0$)	Logarithmic branch points for $z = 0$, $z = \infty$
8	$\sin z$	$\sin x \cosh y + i \cos x \sinh y$	$z = \pi k,\ n = 1$ $(k = 0, \pm1, \pm2, \dots)$	$z = \infty$ is an essential singular point
9	$\cos z$	$\cos x \cosh y + i(-\sin x \sinh y)$	$z = \frac{1}{2}\pi + \pi k,\ n = 1$ $(k = 0, \pm1, \pm2, \dots)$	$z = \infty$ is an essential singular point
10	$\tan z$	$\dfrac{\sin 2x}{\cos 2x + \cosh 2y} + i\dfrac{\sinh 2y}{\cos 2x + \cosh 2y}$	$z = \pi k,\ n = 1$ $(k = 0, \pm1, \pm2, \dots)$	$z = \frac{1}{2}\pi + \pi k$ $(k = 0, \pm1, \pm2, \dots)$ are first-order poles

► **Analytic continuation.** Let two domains D_1 and D_2 have a common part γ of the boundary, and let single-valued analytic functions $f_1(z)$ and $f_2(z)$, respectively, be given in these domains. The function $f_2(z)$ is called a *direct analytic continuation* of $f_1(z)$ into the domain D_2 if there exists a function $f(z)$ analytic in the domain $D_1 \cup \gamma \cup D_2$ and satisfying the condition

$$f(z) = \begin{cases} f_1(z) & \text{for } z \in D_1, \\ f_2(z) & \text{for } z \in D_2. \end{cases} \qquad \text{(M9.2.1.6)}$$

If such a continuation is possible, then the function $f(z)$ is uniquely determined. If the domains are simply connected and the functions $f_1(z)$ and $f_2(z)$ are continuous in $D_1 \cup \gamma$ and $D_2 \cup \gamma$, respectively, and coincide on γ, then $f_2(z)$ is the direct analytic continuation of $f_1(z)$ into the domain D_2. In addition, suppose that the domains D_1 and D_2 are allowed to have common interior points. A function $f_2(z)$ is called a direct analytic continuation of $f_1(z)$ through γ if $f_1(z)$ and $f_2(z)$ are continuous in $D_1 \cup \gamma$ and $D_2 \cup \gamma$, respectively, and their values on γ coincide. At the common interior points of D_1 and D_2, the function determined by relation (M9.2.1.6) can be double-valued.

M9.2.2. Integration of Functions of Complex Variables

► **Definition and properties of the integral of a function of a complex variable.** Suppose that an oriented curve C connecting points $z = a$ and $z = b$ is given on the complex plane and a function $w = f(z)$ of the complex variable z is defined on the curve. We divide the curve C into n parts, $a = z_0, z_1, \ldots, z_{n-1}, z_n = b$, arbitrarily choose $\xi_k \in [z_k, z_{k+1}]$, and compose the integral sum

$$\sum_{k=0}^{n-1} f(\xi_k)(z_{k+1} - z_k).$$

If there exists a limit of this sum as $\max|z_{k+1} - z_k| \to 0$, independent of the way it is partitioned and the choice of the points ξ_k, then this limit is called the *integral* of the function $w = f(z)$ over the curve C and is denoted by

$$\int_C f(z)\,dz. \qquad \text{(M9.2.2.1)}$$

Properties of the integral of a function of a complex variable:

1. If α, β are arbitrary constants, then $\int_C [\alpha f(z) + \beta g(z)]\,dz = \alpha \int_C f(z)\,dz + \beta \int_C g(z)\,dz$.
2. If $\widetilde{C}$ is the same curve as C but with the opposite sense, then $\int_{\widetilde{C}} f(z)\,dz = -\int_C f(z)\,dz$.
3. If $C = C_1 \cup \cdots \cup C_n$, then $\int_C f(z)\,dz = \int_{C_1} f(z)\,dz + \cdots + \int_{C_n} f(z)\,dz$.
4. If $|f(z)| \le M$ at all points of the curve C, then the following *estimate of the absolute value of the integral* holds: $\left|\int_C f(z)\,dz\right| \le Ml$, where l is the length of the curve C.

If C is a piecewise smooth curve and $f(z)$ is bounded and piecewise continuous, then the integral (M9.2.2.1) exists. If $z = x + iy$ and $w = u(x, y) + iv(x, y)$, then the computation of the integral (M9.2.2.1) is reduced to finding two ordinary curvilinear integrals:

$$\int_C f(z)\,dz = \int_C u(x, y)\,dx - v(x, y)\,dy + i \int_C v(x, y)\,dx + u(x, y)\,dy. \qquad \text{(M9.2.2.2)}$$

Remark. Formula (M9.2.2.2) can also be written in a form convenient for memorizing:

$$\int_C f(z)\,dz = \int_C (u + iv)(dx + i\,dy).$$

If the curve C is given by the parametric equations $x = x(t)$, $y = y(t)$ $(t_1 \le t \le t_2)$, then

$$\int_C f(z)\,dz = \int_{t_1}^{t_2} f(z(t))z_t'(t)\,dt,$$

where $z = z(t) = x(t) + iy(t)$ is the complex parametric equation of the curve C.

If $f(z)$ is an analytic function in a simply connected domain D containing the points $z = a$ and $z = b$, then the *Newton–Leibniz formula* holds:

$$\int_a^b f(z)\,dz = F(b) - F(a),$$

where $F(z)$ is a primitive of the function $f(z)$, i.e., $F_z'(z) = f(z)$ in the domain D.

If $f(z)$ and $g(z)$ are analytic functions in a simply connected domain D and $z = a$ and $z = b$ are arbitrary points of the domain D, then the *formula of integration by parts* holds:

$$\int_a^b f(z)\,dg(z) = f(b)g(b) - f(a)g(a) - \int_a^b g(z)\,df(z).$$

If an analytic function $z = g(w)$ determines a single-valued mapping of a curve $\widetilde{C}$ onto a curve C, then

$$\int_C f(z)\,dz = \int_{\widetilde{C}} f(g(w))g_w'(w)\,dw.$$

► **Cauchy's theorems.**

CAUCHY'S THEOREM FOR A SIMPLY CONNECTED DOMAIN. *If a function $f(z)$ is analytic in a simply connected domain D bounded by a contour C and is continuous in $\overline{D}$, then $\int_C f(z)\,dz = 0$.*

CAUCHY'S THEOREM FOR A MULTIPLY CONNECTED DOMAIN. *If a function $f(z)$ is analytic in a multiply connected domain D bounded by a contour C consisting of several closed curves and is continuous in $\overline{D}$, then $\int_C f(z)\,dz = 0$ provided that the sense of all curves forming C is chosen in such a way that the domain D lies to the same side of the contour.*

► **Cauchy's integral and related integrals. Morera's theorem.** If a function $f(z)$ is analytic in an n-connected domain D and continuous in $\overline{D}$, and C is the boundary of D, then for any interior point z of this domain the *Cauchy integral formula* holds:

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f(\xi)}{\xi - z}\,d\xi. \qquad \text{(M9.2.2.3)}$$

(Here integration is carried out in the positive sense of C; i.e., the contour C is traced so that the domain D lies to the left of the contour.) Under the same assumptions as above, the derivatives of arbitrary order of the function $f(z)$ at any interior point z of the domain are expressed as

$$f_z^{(n)}(z) = \frac{n!}{2\pi i}\int_C \frac{f(\xi)}{(\xi - z)^{n+1}}\,d\xi \qquad (n = 1, 2, \ldots). \qquad \text{(M9.2.2.4)}$$

For an arbitrary smooth curve C, not necessarily closed, and for a function $f(\xi)$ everywhere continuous on C, possibly except for finitely many points at which this function has an integrable discontinuity, the right-hand side of formula (M9.2.2.3) defines a *Cauchy-type* *integral*. The function $F(z)$ determined by a Cauchy-type integral is analytic at any point that does not belong to C. If C divides the plane into several domains, then the Cauchy-type integral generally determines different analytic functions in these domains.

Formulas (M9.2.2.3) and (M9.2.2.4) allow one to calculate the integrals

$$\int_C \frac{f(\xi)}{\xi - z}\,d\xi = 2\pi i f(z), \qquad \int_C \frac{f(\xi)}{(\xi - z)^{n+1}}\,d\xi = \frac{2\pi i}{n!} f_z^{(n)}(z). \tag{M9.2.2.5}$$

Example 1. Let us calculate the integral

$$\int_C \operatorname{Im} z\,dz,$$

where C is the semicircle $|z| = 1$, $0 \le \arg z \le \pi$ (Fig. M9.4).

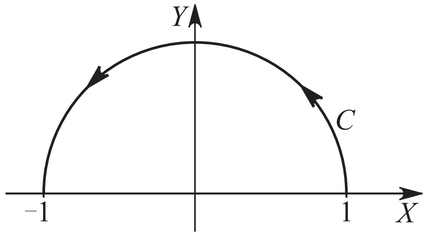

Figure M9.4. The semicircle $|z| = 1$, $0 \le \arg z \le \pi$.

Using formula (M9.2.2.2), we obtain

$$\int_C \operatorname{Im} z\,dz = \int_C y(dx + i\,dy) = \int_C y\,dx + i\int_C y\,dy = \int_1^{-1} \sqrt{1 - x^2}\,dx - i \cdot 0 = -\frac{\pi}{2}.$$

Example 2. Let us calculate the integral

$$\int_C \frac{dz}{z - z_0},$$

where C is the circle of radius R centered at a point z_0 with counterclockwise sense.

Using the integral formula (M9.2.2.5), we obtain

$$\int_C \frac{1}{z - z_0}\,dz = 2\pi i.$$

Example 3. Let us calculate the integral

$$\int_C \frac{dz}{z^2 + 1},$$

where C is the circle of unit radius centered at the point i with counterclockwise sense.

To apply the Cauchy integral formula (M9.2.2.3), we transform the integrand as follows:

$$\frac{1}{1 + z^2} = \frac{1}{(z - i)(z + i)} = \frac{1}{z + i}\,\frac{1}{z - i} = \frac{f(z)}{z - i}, \quad f(z) = \frac{1}{z + i}.$$

The function $f(z) = 1/(z + i)$ is analytic in the interior of the domain under study and on its boundary; hence the Cauchy integral formula (M9.2.2.3) and the first of formulas (M9.2.2.5) hold. From the latter formula, we obtain

$$\int_C \frac{dz}{z^2 + 1} = \int_C \frac{f(z)}{z - i}\,dz = 2\pi i f(i) = 2\pi i \frac{1}{2i} = \pi.$$

Formulas (M9.2.2.3) and (M9.2.2.4) imply the *Cauchy inequalities*

$$|f_z^{(n)}(z)| \le \frac{n!}{2\pi}\left|\int_C \frac{f(\xi)}{(\xi - z)^{n+1}}\,d\xi\right| \le \frac{n!Ml}{2\pi R^{n+1}},$$

where $M = \max_{z\in D}\bigl|f(z)\bigr|$ is the maximum modulus of the function $f(z)$ in the domain D, R is the distance from the point z to the boundary C, and l is the length of the boundary C.

If, in particular, $f(z)$ is analytic in the disk $D = |z - z_0| < R$, and bounded in $\bar{D}$, then we obtain the inequality

$$|f_z^{(n)}(z_0)| \le \frac{n!M}{R^n} \quad (n = 0, 1, 2, \dots).$$

MORERA'S THEOREM. *If a function $f(z)$ is continuous in a simply connected domain D and $\int_C f(z)\,dz = 0$ for any closed curve C lying in D, then $f(z)$ is analytic in the domain D.*

M9.2.3. Taylor and Laurent Series

► **Taylor series.** If a series

$$\sum_{n=0}^{\infty} f_n(z) \qquad \text{(M9.2.3.1)}$$

of analytic functions in a simply connected domain D converges uniformly in this domain, then its sum is analytic in the domain D.

If a series (M9.2.3.1) of functions analytic in a domain D and continuous in $\overline{D}$ converges uniformly in $\overline{D}$, then it can be differentiated termwise any number of times and can be integrated termwise over any piecewise smooth curve C lying in D.

ABEL'S THEOREM. *If the power series*

$$\sum_{n=0}^{\infty} c_n(z-a)^n \qquad \text{(M9.2.3.2)}$$

converges at a point z_0, then it also converges at any point z satisfying the condition $|z - a| < |z_0 - a|$. Moreover, the series converges uniformly in any disk $|z - a| \le q|z_0 - a|$, where $0 < q < 1$.

It follows from Abel's theorem that the domain of convergence of a power series is an open disk centered at the point a; moreover, this disk can fill the entire plane. The radius of this disk is called the *radius of convergence* of a power series. The sum of the power series inside the disk of convergence is an analytic function.

Remark. The radius of convergence R can be found by the *Cauchy–Hadamard formula*

$$\frac{1}{R} = \overline{\lim_{n\to\infty}} \sqrt[n]{|c_n|},$$

where $\overline{\lim}$ denotes the upper limit.

If a function $f(z)$ is analytic in the open disk D of radius R centered at a point $z = a$, then this function can be represented in this disk by its *Taylor series*

$$f(z) = \sum_{n=0}^{\infty} c_n(z-a)^n,$$

whose coefficients are determined by the formulas

$$c_n = \frac{f_z^{(n)}(a)}{n!} = \frac{1}{2\pi i}\int_C \frac{f(\xi)}{(\xi - a)^{n+1}}\,d\xi \qquad (n = 0, 1, 2, \dots),$$

where C is the circle $|z-a| = qR$, $0 < q < 1$. In any closed domain belonging to the disk D, the Taylor series converges uniformly. Any power series expansion of an analytic function is its Taylor expansion. The Taylor series expansions of some elementary functions in powers of z are as follows:

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \dots \quad (|z| < \infty), \tag{M9.2.3.3}$$

$$\cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \dots, \quad \sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \dots \quad (|z| < \infty), \tag{M9.2.3.4}$$

$$\cosh z = 1 + \frac{z^2}{2!} + \frac{z^4}{4!} + \dots, \quad \sinh z = z + \frac{z^3}{3!} + \frac{z^5}{5!} + \dots \quad (|z| < \infty), \tag{M9.2.3.5}$$

$$\ln(1 + z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \dots \quad (|z| < 1), \tag{M9.2.3.6}$$

$$(1 + z)^a = 1 + az + \frac{a(a-1)}{2!}z^2 + \frac{a(a-1)(a-2)}{3!}z^3 + \dots \quad (|z| < 1). \tag{M9.2.3.7}$$

The last two expansions are valid for the single-valued branches for which the values of the functions for $z = 0$ are equal to 0 and 1, respectively.

Remark. Series expansions (M9.2.3.3)–(M9.2.3.7) coincide with analogous expansions of the corresponding elementary functions of the real variable (see Subsection M8.3.2).

To obtain the Taylor series for other branches of the multi-valued function $\mathrm{Ln}(1 + z)$, one has to add the numbers $2k\pi i$, $k = \pm1, \pm2, \dots$ to the expression in the right-hand side:

$$\mathrm{Ln}(1 + z) = 2k\pi i + z - \frac{z^2}{2} + \frac{z^3}{3} - \dots.$$

► **Laurent series.** The domain of convergence of the function series $\sum\limits_{n=-\infty}^{\infty} c_n(z-a)^n$ is a circular annulus $K : r < |z - a| < R$, where $0 \le r < R \le \infty$. The sum of the series is an analytic function in the annulus of convergence. Conversely, in any annulus K where the function $f(z)$ is analytic, this function can be represented by the *Laurent series* expansion

$$f(z) = \sum_{n=-\infty}^{\infty} c_n(z-a)^n$$

with coefficients determined by the formulas

$$c_n = \frac{1}{2\pi i} \int_\gamma \frac{f(\xi)}{(\xi - a)^{n+1}}\, d\xi \quad (n = 0, \pm1, \pm2, \dots), \tag{M9.2.3.8}$$

where γ is the circle $|z-a| = \rho$, $r < \rho < R$. In any closed domain contained in the annulus K, the Laurent series converges uniformly.

The part of the Laurent series with negative numbers,

$$\sum_{n=-\infty}^{-1} c_n(z-a)^n = \sum_{n=1}^{\infty} \frac{c_{-n}}{(z-a)^n},$$

is called its *principal part*, and the part with nonnegative numbers,

$$\sum_{n=0}^{\infty} c_n(z-a)^n,$$

is called the *regular part*. Any expansion of an analytic function in positive and negative powers of $z - a$ is its Laurent expansion.

Example 1. Let us consider Laurent series expansions of the function

$$f(z)=\frac{1}{z(1-z)}$$

in a Laurent series in the domain $0<|z|<1$. This function is analytic in the annulus $0<|z|<1$ and hence can be expanded in the corresponding Laurent series. We write this function as the sum of elementary fractions:

$$f(z)=\frac{1}{z(1-z)}=\frac{1}{z}+\frac{1}{1-z}.$$

Since $|z|<1$, we can use formula (M9.2.3.7) and obtain the expansion

$$\frac{1}{z(1-z)}=\frac{1}{z}+1+z+z^2+\cdots$$

Example 2. Let us consider Laurent series expansions of the function

$$f(z)=e^{1/z}$$

in a Laurent series in a neighborhood of the point $a=0$. To this end, we use the expansion (M9.2.3.3), where we should replace z by $1/z$. Thus we obtain

$$e^{1/z}=1+\frac{1}{1!\,z}+\frac{1}{2!\,z^2}+\cdots+\frac{1}{n!\,z^n}+\cdots\qquad(z\neq 0).$$

M9.2.4. Zeros and Isolated Singularities of Analytic Functions

► **Zeros of analytic functions.** A point $z=a$ is called a *zero* of a function $f(z)$ if $f(a)=0$. If $f(z)$ is analytic at the point a and is not zero identically, then the least order of nonzero coefficients in the Taylor expansion of $f(z)$ centered at a, in other words, the number n of the first nonzero derivative $f^{(n)}(a)$, is called the *order of zero* of this function. In a neighborhood of a zero a of order n, the Taylor expansion of $f(z)$ has the form

$$f(z)=c_n(z-a)^n+c_{n+1}(z-a)^{n+1}+\ldots\qquad(c_n\neq 0,\ n\geq 1).$$

In this case, $f(z)=c_n(z-a)^n g(z)$, where the function $g(z)$ is analytic at the point a and $g(a)\neq 0$. A first-order zero is said to be *simple*. The point $z=\infty$ is a zero of order n for a function $f(z)$ if $z=0$ is a zero of order n for $F(z)=f(1/z)$.

If a function $f(z)$ is analytic in a neighborhood of its zero a and is not identically zero in any neighborhood of a, then there exists a neighborhood of a in which $f(z)$ does not have any zeros other than a.

UNIQUENESS THEOREM. *If functions $f(z)$ and $g(z)$ are analytic in a domain D and their values coincide on some sequence a_k of points converging to an interior point a of the domain D, then $f(z)\equiv g(z)$ everywhere in D.*

ROUCHÉ'S THEOREM. *If functions $f(z)$ and $g(z)$ are analytic in a simply connected domain D bounded by a curve C, are continuous in $\overline{D}$, and satisfy the inequality $|f(z)|>|g(z)|$ on C, then the functions $f(z)$ and $f(z)+g(z)$ have the same number of zeros in D.*

► **Isolated singularities of analytic functions.** A point a is called an *isolated singularity* of a single-valued analytic function $f(z)$ if there exists a neighborhood of this point in which $f(z)$ is analytic everywhere except for the point a itself. The point a is called

1. A *removable singularity* if $\lim\limits_{z\to a} f(z)$ exists and is finite.
2. A *pole* if $\lim\limits_{z\to a} f(z)=\infty$.
3. An *essential singularity* if $\lim\limits_{z\to a} f(z)$ does not exist.

A necessary and sufficient condition for a point a to be a removable singularity of a function $f(z)$ is that the Laurent expansion of $f(z)$ around a does not contain the principal part.

If a function $f(z)$ is bounded in a neighborhood of an isolated singularity a, then a is a removable singularity of this function.

A necessary and sufficient condition for a point a to be a pole of a function $f(z)$ is that the principal part of the Laurent expansion of $f(z)$ around a contains finitely many terms:

$$f(z)=\frac{c_{-n}}{(z-a)^n}+\dots+\frac{c_{-1}}{(z-a)}+\sum_{k=0}^{\infty}c_k(z-a)^k. \tag{M9.2.4.1}$$

The *order of a pole* a of a function $f(z)$ is defined to be the order of the zero of the function $F(z)=1/f(z)$. If $c_{-n}\neq 0$ in expansion (M9.2.4.1), then the order of the pole a of the function $f(z)$ is equal to n. For $n=1$, we have a *simple pole*.

A necessary and sufficient condition for a point a to be an essential singularity of a function $f(z)$ is that the principal part of the Laurent expansion of $f(z)$ around a contains infinitely many nonzero terms.

SOKHOTSKI'S THEOREM. *If a is an essential singularity of a function $f(z)$, then for each complex number A there exists a sequence of points $z_k\to a$ such that $f(z_k)\to A$.*

Example. Let us consider some functions with singular points of different kind.

1°. The function $f(z)=(1-\cos z)/z^2$ has a removable singularity at the origin, since its Laurent expansion about the origin,

$$\frac{1-\cos z}{z^2}=\frac{1}{2}-\frac{z^2}{24}+\frac{z^4}{720}-\dots,$$

does not contain the principal part.

2°. The function $f(z)=1/(1+e^{z^2})$ has infinitely many poles at the points $z=\pm\sqrt{(2k+1)\pi i}$ $(k=0,\pm1,\pm2,\dots)$. All these poles are simple poles, since the function $1/f(z)=1+e^{z^2}$ has simple zeros at these points. (Its derivative is nonzero at these points.)

3°. The function $f(z)=\sin(1/z)$ has an essential singularity at the origin, since the principal part of its Laurent expansion

$$\sin\frac{1}{z}=\frac{1}{z}-\frac{1}{z^3 3!}+\dots$$

contains infinitely many terms.

The following two simplest classes of single-valued analytic functions are distinguished according to the character of singular points.

1. *Entire functions.* A function $f(z)$ is said to be *entire* if it does not have singular points in the finite part of the complex plane. An entire function can be represented by an everywhere convergent power series

$$f(z)=\sum_{n=0}^{\infty}c_n z^n.$$

An entire function can have only one singular point at $z=\infty$. If this singularity is a pole of order n, then $f(z)$ is a polynomial of degree n. If $z=\infty$ is an essential singularity, then $f(z)$ is called an *entire transcendental function*. If $z=\infty$ is a regular point (i.e., $f(z)$ is analytic for all z), then $f(z)$ is constant (*Liouville's theorem*). All polynomials, the exponential function, $\sin z$, $\cos z$, etc. are examples of entire functions. Sums, differences, and products of entire functions are themselves entire functions.

2. *Meromorphic functions.* A function $f(z)$ is said to be *meromorphic* if it does not have any singularities except for poles. The number of these poles in each finite closed domain D is always finite.

Suppose that a function $f(z)$ is analytic in a neighborhood of the point at infinity. The definition of singular points can be generalized to this function without any changes. But the criteria for the type of a singular point at infinity related to the Laurent expansion are different.

THEOREM. *In the case of a removable singularity at the point at infinity, the Laurent expansion of a function $f(z)$ in a neighborhood of this point does not contain positive powers of z. In the case of a pole, it contains finitely many positive powers of z. In the case of an essential singularity, it contains infinitely many powers of z.*

Let $f(z)$ be a multi-valued function defined in a neighborhood D of a point $z = a$ except possibly for the point a itself, and let $f_1(z)$, $f_2(z)$, ... be its branches, which are single-valued continuous functions in the domain where they are defined. The point a is called a *branch point (ramification point)* of the function $f(z)$ if $f(z)$ passes from one branch to another as the point z goes along a closed curve around the point z in a neighborhood of D. If the original branch is reached again after going around this curve m times (in the same sense), then the number $m-1$ is called the *order* of the branch point, and the point a itself is called a *branch point of order* $m-1$.

If all branches $f_k(z)$ tend to the same finite or infinite limit as $z \to a$, then the point a is called an *algebraic* branch point. (For example, the point $z = 0$ is an algebraic branch point of the function $f(z) = \sqrt[m]{z}$.) In this case, the single-valued function

$$F(z) = f(z^m + a)$$

has a regular point or a pole for $z = 0$.

If the limit of $f_k(z)$ as $z \to a$ does not exist, then the point a is called a *transcendental* branch point. For example, the point $z = 0$ is a transcendental branch point of the function $f(z) = \exp(\sqrt[m]{1/z})$.

In a neighborhood of a branch point a of finite order, the function $f(z)$ can be expanded in a *fractional power series* (*Puiseux series*)

$$f(z) = \sum_{k=-\infty}^{\infty} c_k (z-a)^{k/m}. \qquad \text{(M9.2.4.2)}$$

If a new branch is obtained each time after going around this curve (in the same sense), then the point a is called a *branch point of infinite order (a logarithmic branch point)*. For example, the points $z = 0$ and $z = \infty$ are logarithmic branch points of the multi-valued function $w = \operatorname{Ln} z$. A logarithmic branch point is classified as a transcendental branch point.

For $a \neq \infty$, the expansion (M9.2.4.2) contains finitely many terms with negative k (infinitely many in the case of a transcendental point).

M9.2.5. Residues. Calculation of Definite Integrals

► **Residue of an analytic function at an isolated singular point.** The *residue* $\operatorname{res} f(a)$ of a function $f(z)$ at an isolated singularity a is defined as the number

$$\operatorname{res} f(a) = \frac{1}{2\pi i} \oint_C f(z)\,dz, \qquad \text{(M9.2.5.1)}$$

where the integral is taken in the positive sense over a contour C surrounding the point a and containing no other singularities of $f(z)$ in the interior.

Remark. Residues are sometimes denoted by $\operatorname{res}[f(z); a]$ or $\operatorname{res}_{z=a} f(z)$.

The residue $\operatorname{res} f(a)$ of a function $f(z)$ at a singularity a is equal to the coefficient of $(z-a)^{-1}$ in the Laurent expansion of $f(z)$ in a neighborhood of the point a,

$$\operatorname{res} f(a) = \frac{1}{2\pi i} \oint_C f(z)\,dz = c_{-1}.$$

Basic rules for finding the residues:

1. The residue of a function at a removable singularity is zero.
2. If a is a pole of order n, then

$$\operatorname{res} f(a) = \frac{1}{(n-1)!} \lim_{z\to a} \frac{d^{n-1}}{dz^{n-1}} \big[f(z)(z-a)^n\big]. \qquad \text{(M9.2.5.2)}$$

3. For a simple pole ($n = 1$),

$$\operatorname{res} f(a) = \lim_{z\to a} \big[f(z)(z-a)\big].$$

4. If $f(z)$ is the quotient of two analytic functions,

$$f(z) = \frac{\varphi(z)}{\psi(z)},$$

in a neighborhood of a point a and $\varphi(a) \neq 0$, $\psi(a) = 0$, but $\psi'_z(a) \neq 0$ (i.e., a is a simple pole of $f(z)$), then

$$\operatorname{res} f(a) = \frac{\varphi(a)}{\psi'_z(a)}. \qquad \text{(M9.2.5.3)}$$

5. If a is an essential singularity of $f(z)$, then to obtain $\operatorname{res} f(a)$, one has to find the coefficient c_{-1} in the Laurent expansion of $f(z)$ in a neighborhood of a.

► Basic theorems on residues.

A function $f(z)$ is said to be continuous on the boundary C of the domain D if for each boundary point z_0 there exists a limit $\lim\limits_{z\to z_0} f(z) = f(z_0)$ as $z \to z_0$, $z \in D$.

CAUCHY'S RESIDUE THEOREM. *Let $f(z)$ be a function continuous on the boundary C of a domain D and analytic in the interior of D everywhere except for finitely many points $a_1, \ldots, a_n$. Then*

$$\int_C f(z)\,dz = 2\pi i \sum_{k=1}^{n} \operatorname{res} f(a_k), \qquad \text{(M9.2.5.4)}$$

where the integral is taken in the positive sense of C.

The *logarithmic residue* of a function $f(z)$ at a point a is by definition the residue of its logarithmic derivative

$$\big[\ln f(z)\big]'_z = \frac{f'_z(z)}{f(z)}.$$

THEOREM. *The logarithmic derivative $f'_z(z)/f(z)$ has first-order poles at the zeros and poles of $f(z)$. Moreover, the logarithmic residue of $f(z)$ at a zero or a pole of $f(z)$ is equal to the order of the zero or minus the order of the pole, respectively.*

The residue of a function $f(z)$ at infinity is defined as

$$\operatorname{res} f(\infty) = \frac{1}{2\pi i} \oint_\Gamma f(z)\,dz,$$

where Γ is a circle of sufficiently large radius $|z| = \rho$ and the integral is taken in the clockwise sense (so that the neighborhood of the point $z = \infty$ remains to the left of the contour, just as in the case of a finite point).

The residue of $f(z)$ at infinity is equal to minus the coefficient of z^{-1} in the Laurent expansion of $f(z)$ in a neighborhood of the point $z = \infty$,

$$\operatorname{res} f(\infty) = -c_{-1}.$$

THEOREM. *If a function $f(z)$ has finitely many singular points in the extended complex plane, then the sum of all its residues, including the residue at infinity, is zero:*

$$\operatorname{res} f(\infty) + \sum_{k=1}^{n} \operatorname{res} f(a_k) = 0, \qquad \text{(M9.2.5.5)}$$

where $a_1, \dots, a_n$ are finite singular points.

Example 1. Let us calculate the integral

$$\oint_C \frac{\ln(z+2)}{z^2}\,dz,$$

where C is the circle $|z| = \frac{1}{2}$.

In the disk $|z| \le \frac{1}{2}$, there is only one singular point of the integrand, $z = 0$, which is a second-order pole. The residue of $f(z)$ at $z = 0$ is calculated by the formula (M9.2.5.2)

$$\operatorname{res} f(0) = \lim_{z\to 0}\Bigl[z^2 \frac{\ln(z+2)}{z^2}\Bigr]'_z = \lim_{z\to 0}[\ln(z+2)]'_z = \lim_{z\to 0} \frac{1}{z+2} = \frac{1}{2}.$$

Using formula (M9.2.5.1), we obtain

$$\frac{1}{2} = \frac{1}{2\pi i} \oint_C \frac{\ln(z+2)}{z^2}\,dz, \qquad \oint_C \frac{\ln(z+2)}{z^2}\,dz = \pi i.$$

► **Jordan's lemma. Calculation of definite integrals using residues.** Suppose that we need to calculate the integral of a real function $f(x)$ over a (finite or infinite) interval (a, b). Let us supplement the interval (a, b) with a curve Γ that, together with (a, b), bounds a domain D, and then analytically continue the function $f(x)$ into $\overline{D}$. Then the residue theorem can be applied to this analytic continuation of $f(z)$, and by this theorem

$$\int_a^b f(x)\,dx + \int_\Gamma f(z)\,dz = 2\pi i \Lambda,$$

where Λ is the sum of residues of $f(z)$ in D. If $\int_\Gamma f(z)\,dz$ can be calculated or expressed in terms of the desired integral $\int_a^b f(x)\,dx$, then the problem will be solved.

When calculating integrals of the form $\int_{-\infty}^{\infty} f(x)\,dx$, one should apply (M9.2.5.4) to the contour C that consists of the interval $(-R, R)$ of the real axis and the arc C_R of the semicircle $|z| = R$ in the upper half-plane. Sometimes, it is only possible to find the limit as $R \to \infty$ of the integral over the contour C_R rather than to calculate it, and often it turns out that the limit of this integral is equal to zero.

LEMMA. *If a function $f(z)$ is analytic for $|z| > R_0$ and $zf(z) \to 0$ as $|z| \to \infty$ for $y \ge 0$, then*

$$\lim_{R\to\infty} \int_{C_R} f(z)\,dz = 0,$$

where C_R is the arc of the semicircle $|z| = R$ in the upper half-plane.

THEOREM. *Let a function $f(x)$ be defined on the whole real axis $-\infty < x < \infty$ and let it can be analytically continued to the upper half-plane $\operatorname{Im} z > 0$ so that the continuation $f(z)$ satisfies the conditions of the previous lemma. Then the improper integral $\int_{-\infty}^{\infty} f(x)\,dx$ exists and is equal to*

$$\int_{-\infty}^{\infty} f(x)\,dx = 2\pi i \sum_{k=1}^{n} \operatorname{res} f(a_k), \tag{M9.2.5.6}$$

where the a_k are singular points of $f(z)$ in the upper half-plane.

Example 2. Calculate the integral

$$I = \int_{-\infty}^{\infty} \frac{dx}{1+x^4}.$$

The analytic continuation of the integrand into the upper half-plane is $f(z) = (1+z^4)^{-1}$; it satisfies the conditions of the above lemma. The function $f(z)$ has two singular points, $a_1 = e^{i\pi/4}$ and $a_2 = e^{i3\pi/4}$, in the upper half-plane (both points are first-order poles). Using formulas (M9.2.5.3) and (M9.2.5.6), one finds

$$I = 2\pi i \left(\frac{1}{4z^3}\bigg|_{z=e^{i\pi/4}} + \frac{1}{4z^3}\bigg|_{z=e^{i3\pi/4}} \right) = \frac{\pi\sqrt{2}}{2}.$$

JORDAN'S LEMMA. *If a function $g(z)$ tends to zero uniformly with respect to $\arg z$ along a sequence of circular arcs $C_{R_n} : |z| = R_n$, $\operatorname{Im} z > -a$ (where $R_n \to \infty$ and a is fixed), then*

$$\lim_{n\to\infty} \int_{C_{R_n}} g(z) e^{i\lambda z}\,dz = 0$$

for any positive number λ.

Example 3 (Laplace integral). To calculate the integral

$$\int_0^{\infty} \frac{\cos x}{x^2+a^2}\,dx,$$

one uses the auxiliary function

$$f(z) = \frac{e^{iz}}{z^2+a^2} = g(z)e^{iz}, \quad g(z) = \frac{1}{z^2+a^2}$$

and the contour shown in Fig. M9.5. Since $g(z)$ satisfies the inequality $|g(z)| < (R^2 - a^2)^{-1}$ on C_R, it follows that this function uniformly tends to zero as $R \to \infty$, and by Jordan's lemma with $\lambda = 1$ we obtain

$$\int_{C_R} f(z)\,dz = \int_{C_R} g(z)e^{iz}\,dz \to 0$$

as $R \to \infty$.

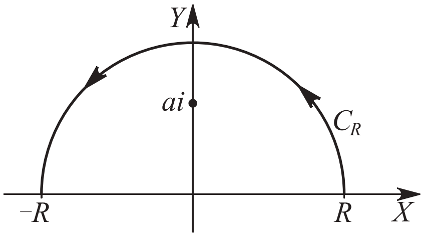

Figure M9.5. The contour for the calculation of the Laplace integral.

By the residue theorem

$$\int_{-R}^{R} \frac{e^{ix}}{x^2+a^2}\,dx + \int_{C_R} f(z)\,dz = 2\pi i\,\frac{e^{-a}}{2ai}$$

for any $R > a$. (The residue at the singular point $z = ai$ of the function $f(z)$, which is a first-order pole and the only singular point of this function lying inside the contour, can be calculated by formula (M9.2.5.3).) In the limit as $R \to \infty$, we obtain

$$\int_{-\infty}^{\infty} \frac{e^{ix}}{x^2+a^2}\,dx = \frac{\pi}{ae^a}.$$

Separating the real part and using the fact that the function is even, we obtain

$$\int_{0}^{\infty} \frac{\cos x}{x^2+a^2}\,dx = \frac{\pi}{2ae^a}.$$

Bibliography for Chapter M9

Ablowitz, M. J. and Fokas, A. S., *Complex Variables: Introduction and Applications* (Cambridge Texts in Applied Mathematics), *2nd Edition*, Cambridge University Press, Cambridge, England, 2003.

Berenstein, C. A. and Roger Gay, R., *Complex Variables: An Introduction* (Graduate Texts in Mathematics), Springer, New York, 1997.

Bieberbach, L., *Conformal Mapping*, American Mathematical Society, Providence, Rhode Island, 2000.

Bronshtein, I. N., Semendyayev, K. A., Musiol, G., and Mühlig, H., *Handbook of Mathematics, 4th Edition*, Springer, New York, 2004.

Brown, J. W. and Churchill, R. V., *Complex Variables and Applications, 7th Edition*, McGraw-Hill, New York, 2003.

Caratheodory, C., *Conformal Representation*, Dover Publications, New York, 1998.

Carrier, G. F., Krock, M., and Pearson, C. E., *Functions of a Complex Variable: Theory and Technique* (Classics in Applied Mathematics), Society for Industrial & Applied Mathematics, University City Science Center, Philadelphia, 2005.

Cartan, H., *Elementary Theory of Analytic Functions of One or Several Complex Variables*, Dover Publications, New York, 1995.

Conway, J. B., *Functions of One Complex Variable I* (Graduate Texts in Mathematics), *2nd Edition*, Springer, New York, 1995.

Conway, J. B., *Functions of One Complex Variable II* (Graduate Texts in Mathematics)*, 2nd Edition*, Springer, New York, 1996.

Dettman, J. W., Applied Complex Variables (Mathematics Series), Dover Publications, New York, 1984.

England, A. H., *Complex Variable Methods in Elasticity, Dover Edition*, Dover Publications, New York, 2003.

Fisher, S. D., *Complex Variables* (Dover Books on Mathematics), *2nd Edition*, Dover Publications, New York, 1999.

Flanigan, F. J., *Complex Variables, Dover Edition*, Dover Publications, New York, 1983.

Greene, R. E. and Krantz, S. G., *Function Theory of One Complex Variable* (Graduate Studies in Mathematics), *Vol. 40, 2nd Edition*, American Mathematical Society, Providence, Rhode Island, 2002.

Ivanov, V. I. and Trubetskov, M. K., *Handbook of Conformal Mapping with Computer-Aided Visualization*, CRC Press, Boca Raton, Florida, 1995.

Korn, G. A and Korn, T. M., *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review*, Dover Edition, Dover Publications, New York, 2000.

Krantz, S. G., *Handbook of Complex Variables*, Birkhäuser, Boston, Massachusetts, 1999.

Lang, S., *Complex Analysis* (Graduate Texts in Mathematics), *4th Edition*, Springer, New York, 2003.

Lavrentiev, M. A. and Shabat, B. V., *Methods of the Theory of Functions of a Complex Variable, 5th Edition [in Russian]*, Nauka Publishers, Moscow, 1987.

LePage, W. R., *Complex Variables and the Laplace Transform for Engineers*, Dover Publications, New York, 1980.

Markushevich, A. I. and Silverman, R. A. (Editor), *Theory of Functions of a Complex Variable, 2nd Rev. Edition*, American Mathematical Society, Providence, Rhode Island, 2005.

Narasimhan, R. and Nievergelt, Y., *Complex Analysis in One Variable, 2nd Edition*, Birkhäuser, Boston, Basel, Stuttgart, 2000.

Needham, T., *Visual Complex Analysis, Rep. Edition*, Oxford University Press, Oxford, 1999.

Nehari, Z., *Conformal Mapping*, Dover Publications, New York, 1982.

Paliouras, J. D. and Meadows, D. S., *Complex Variables for Scientists and Engineers, Facsimile Edition*, Macmillan Coll. Div., New York, 1990.

Pierpont, J., *Functions of a Complex Variable (Phoenix Edition)*, Dover Publications, New York, 2005.

Schinzinger, R. and Laura, P. A. A., *Conformal Mapping: Methods and Applications*, Dover Publications, New York, 2003.

Silverman, R. A., *Introductory Complex Analysis*, Dover Publications, New York, 1984.

Spiegel, M. R., *Schaum's Outline of Complex Variables*, McGraw-Hill, New York, 1968.

Sveshnikov, A. G. and Tikhonov, A. N., The Theory of Functions of a Complex Variable, Mir Publishers, Moscow, 1982.

Wunsch, D. A., *Complex Variables with Applications, 2nd Edition*, Addison Wesley, Boston, Massachusetts, 1993.

Chapter M10

Integral Transforms

M10.1. General Form of Integral Transforms. Inversion Formulas

Normally an integral transform has the form

$$\widetilde{f}(\lambda) = \int_a^b \varphi(x, \lambda) f(x)\, dx.$$

The function $\widetilde{f}(\lambda)$ is called the *transform* of the function $f(x)$ and $\varphi(x, \lambda)$ is called the *kernel* of the integral transform. The function $f(x)$ is called the *inverse transform* of $\widetilde{f}(\lambda)$. The limits of integration a and b are real numbers (usually, $a = 0$, $b = \infty$ or $a = -\infty$, $b = \infty$). For brevity, we rewrite the integral transform as follows: $\widetilde{f}(\lambda) = \mathcal{L}\{f(x)\}$.

General properties of integral transforms (linearity):

$$\mathcal{L}\{kf(x)\} = k\mathcal{L}\{f(x)\},$$
$$\mathcal{L}\{f(x) \pm g(x)\} = \mathcal{L}\{f(x)\} \pm \mathcal{L}\{g(x)\}.$$

Here, k is an arbitrary constant; it is assumed that integral transforms of the functions $f(x)$ and $g(x)$ exist.

In Subsections M10.2–M10.4, the most popular (Laplace, Fourier, Mellin, etc.) integral transforms are described. These subsections also describe the corresponding inversion formulas, which normally have the form

$$f(x) = \int_{\mathcal{C}} \psi(x, \lambda) \widetilde{f}(\lambda)\, d\lambda$$

and make it possible to recover $f(x)$ if $\widetilde{f}(\lambda)$ is given. The integration path $\mathcal{C}$ can lie either on the real axis or in the complex plane.

In many cases, to evaluate the integrals in the inversion formula—in particular, to find the inverse Laplace, Mellin, and Fourier transforms—methods of the theory of functions of a complex variable can be applied, including the theorems about residue and Jordan's lemma, which are outlined in Subsection M9.2.5.

M10.2. Laplace Transform

M10.2.1. Laplace Transform and the Inverse Laplace Transform

► **Laplace transform.** The *Laplace transform* of an arbitrary (complex-valued) function $f(x)$ of a real variable x ($x \geq 0$) is defined by

$$\widetilde{f}(p) = \int_0^\infty e^{-px} f(x)\, dx, \qquad \text{(M10.2.1.1)}$$

where $p = s + i\sigma$ is a complex variable.

The Laplace transform exists for any continuous or piecewise-continuous function satisfying the condition $|f(x)| < Me^{\sigma_0 x}$ with some $M > 0$ and $\sigma_0 \geq 0$. In the following, σ_0 often means the greatest lower bound of the possible values of σ_0 in this estimate; this value is called the *growth exponent* of the function $f(x)$.

For any $f(x)$, the transform $\widetilde{f}(p)$ is defined in the half-plane $\operatorname{Re} p > \sigma_0$ and is analytic there.

For brevity, we shall write formula (M10.2.1.1) as follows:

$$\widetilde{f}(p) = \mathfrak{L}\bigl\{f(x)\bigr\} \qquad \text{or} \qquad \widetilde{f}(p) = \mathfrak{L}\bigl\{f(x),\, p\bigr\}.$$

► **Inverse Laplace transform.** Given the transform $\widetilde{f}(p)$, the function $f(x)$ can be found by means of the inverse Laplace transform

$$f(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \widetilde{f}(p)e^{px}\,dp, \qquad i^2 = -1, \tag{M10.2.1.2}$$

where the integration path is parallel to the imaginary axis and lies to the right of all singularities of $\widetilde{f}(p)$, which corresponds to $c > \sigma_0$.

The integral in inversion formula (M10.2.1.2) is understood in the sense of the Cauchy principal value:

$$\int_{c-i\infty}^{c+i\infty} \widetilde{f}(p)e^{px}\,dp = \lim_{\omega\to\infty}\int_{c-i\omega}^{c+i\omega} \widetilde{f}(p)e^{px}\,dp.$$

In the domain $x < 0$, formula (M10.2.1.2) gives $f(x) \equiv 0$.

Formula (M10.2.1.2) holds for continuous functions. If $f(x)$ has a (finite) jump discontinuity at a point $x = x_0 > 0$, then the right-hand side of (M10.2.1.2) evaluates to $\frac{1}{2}[f(x_0 - 0) + f(x_0 + 0)]$ at this point (for $x_0 = 0$, the first term in the square brackets must be omitted).

For brevity, we write the Laplace inversion formula (M10.2.1.2) as follows:

$$f(x) = \mathfrak{L}^{-1}\bigl\{\widetilde{f}(p)\bigr\} \qquad \text{or} \qquad f(x) = \mathfrak{L}^{-1}\bigl\{\widetilde{f}(p),\, x\bigr\}.$$

There are tables of direct and inverse Laplace transforms (see Sections S2.1 and S2.2), which are handy in solving linear differential and integral equations.

M10.2.2. Main Properties of the Laplace Transform. Inversion Formulas for Some Functions

► **Convolution theorem. Main properties of the Laplace transform.**

1°. The *convolution* of two functions $f(x)$ and $g(x)$ is defined as an integral of the form $\int_0^x f(t)g(x-t)\,dt$, and is usually denoted by $f(x) * g(x)$, so that

$$f(x) * g(x) = \int_0^x f(t)\,g(x-t)\,dt.$$

By performing substitution $x - t = u$, we see that the convolution is symmetric with respect to the convolved functions: $f(x) * g(x) = g(x) * f(x)$.

The *convolution theorem* states that

$$\mathfrak{L}\bigl\{f(x) * g(x)\bigr\} = \mathfrak{L}\bigl\{f(x)\bigr\}\,\mathfrak{L}\bigl\{g(x)\bigr\}$$

and is frequently applied to solve Volterra equations with kernels depending on the difference of the arguments.

2°. The main properties of the correspondence between functions and their Laplace transforms are gathered in Table M10.1.

3°. The Laplace transforms of some functions are listed in Table M10.2; for more detailed tables, see Section S2.1 and the list of references at the end of this chapter.

► **Inverse transforms of rational functions.** Consider the important case in which the transform is a rational function of the form

$$\widetilde{f}(p) = \frac{R(p)}{Q(p)}, \tag{M10.2.2.1}$$

where $Q(p)$ and $R(p)$ are polynomials in the variable p and the degree of $Q(p)$ exceeds that of $R(p)$.

Assume that the zeros of the denominator are simple, i.e.,

$$Q(p) \equiv \text{const}\,(p-\lambda_1)(p-\lambda_2)\dots(p-\lambda_n).$$

Then the inverse transform can be determined by the formula

$$f(x) = \sum_{k=1}^{n} \frac{R(\lambda_k)}{Q'(\lambda_k)} \exp(\lambda_k x), \tag{M10.2.2.2}$$

where the primes denote the derivatives.

If $Q(p)$ has multiple zeros, i.e.,

$$Q(p) \equiv \text{const}\,(p-\lambda_1)^{s_1}(p-\lambda_2)^{s_2}\dots(p-\lambda_m)^{s_m},$$

then

$$f(x) = \sum_{k=1}^{m} \frac{1}{(s_k-1)!} \lim_{p\to s_k} \frac{d^{s_k-1}}{dp^{s_k-1}} \bigl[(p-\lambda_k)^{s_k}\widetilde{f}(p)e^{px}\bigr]. \tag{M10.2.2.3}$$

Example 1. The transform

$$\widetilde{f}(p) = \frac{b}{p^2-a^2} \qquad (a \text{ and } b \text{ real numbers})$$

can be represented as the fraction (M10.2.2.1) with $R(p) = b$ and $Q(p) = (p-a)(p+a)$. The denominator $Q(p)$ has two simple roots, $\lambda_1 = a$ and $\lambda_2 = -a$. Using formula (M10.2.2.2) with $n = 2$ and $Q'(p) = 2p$, we obtain the inverse transform in the form

$$f(x) = \frac{b}{2a}e^{ax} - \frac{b}{2a}e^{-ax} = \frac{b}{a}\sinh(ax).$$

Example 2. The transform

$$\widetilde{f}(p) = \frac{b}{p^2+a^2} \qquad (a \text{ and } b \text{ real numbers})$$

can be written as the fraction (M10.2.2.1) with $R(p) = b$ and $Q(p) = (p-ia)(p+ia)$, $i^2 = -1$. The denominator $Q(p)$ has two simple pure imaginary roots, $\lambda_1 = ia$ and $\lambda_2 = -ia$. Using formula (M10.2.2.2) with $n = 2$, we find the inverse transform:

$$f(x) = \frac{b}{2ia}e^{iax} - \frac{b}{2ia}e^{-iax} = \frac{b}{a}\sin(ax).$$

Example 3. The transform

$$\widetilde{f}(p) = ap^{-n},$$

where n is a positive integer, can be written as the fraction (M10.2.2.1) with $R(p) = a$ and $Q(p) = p^n$. The denominator $Q(p)$ has one root of multiplicity n, $\lambda_1 = 0$. By formula (M10.2.2.3) with $m = 1$ and $s_1 = n$, we find the inverse transform:

$$f(x) = \frac{a}{(n-1)!}x^{n-1}.$$

Remark. Fairly detailed tables of inverse Laplace transforms can be found in Section S2.2.

TABLE M10.1
Main properties of the Laplace transform.

No.	Function	Laplace transform	Operation
1	$af_1(x)+bf_2(x)$	$a\widetilde{f}_1(p)+b\widetilde{f}_2(p)$	Linearity
2	$f(x/a),\ a>0$	$a\widetilde{f}(ap)$	Scaling
3	$f(x-a)$, $f(x-a)\equiv 0$ for $x<a$	$e^{-ap}\widetilde{f}(p)$	Shift of the argument
4	$x^n f(x);\ n=1,2,\ldots$	$(-1)^n\widetilde{f}_p^{(n)}(p)$	Differentiation of the transform
5	$\dfrac{1}{x}f(x)$	$\displaystyle\int_p^\infty \widetilde{f}(q)\,dq$	Integration of the transform
6	$e^{ax}f(x)$	$\widetilde{f}(p-a)$	Shift in the complex plane
7	$f'_x(x)$	$p\widetilde{f}(p)-f(+0)$	Differentiation
8	$f_x^{(n)}(x)$	$p^n\widetilde{f}(p)-\sum\limits_{k=1}^{n}p^{n-k}f_x^{(k-1)}(+0)$	Differentiation
9	$x^m f_x^{(n)}(x),\ m=1,2,\ldots$	$(-1)^m\dfrac{d^m}{dp^m}\Big[p^n\widetilde{f}(p)-\sum\limits_{k=1}^{n}p^{n-k}f_x^{(k-1)}(+0)\Big]$	Differentiation
10	$\dfrac{d^n}{dx^n}\big[x^m f(x)\big],\ \ m\ge n$	$(-1)^m p^n\dfrac{d^m}{dp^m}\widetilde{f}(p)$	Differentiation
11	$\displaystyle\int_0^x f(t)\,dt$	$\dfrac{\widetilde{f}(p)}{p}$	Integration
12	$\displaystyle\int_0^x f_1(t)f_2(x-t)\,dt$	$\widetilde{f}_1(p)\widetilde{f}_2(p)$	Convolution

TABLE M10.2
The Laplace transforms of some functions.

No.	Function, $f(x)$	Laplace transform, $\widetilde{f}(p)$	Remarks
1	1	$1/p$	
2	x^n	$\dfrac{n!}{p^{n+1}}$	$n=1,\ 2,\ \ldots$
3	x^a	$\Gamma(a+1)p^{-a-1}$	$a>-1$
4	e^{-ax}	$(p+a)^{-1}$	
5	$x^a e^{-bx}$	$\Gamma(a+1)(p+b)^{-a-1}$	$a>-1$
6	$\sinh(ax)$	$\dfrac{a}{p^2-a^2}$	
7	$\cosh(ax)$	$\dfrac{p}{p^2-a^2}$	
8	$\ln x$	$-\dfrac{1}{p}(\ln p+\mathcal{C})$	$\mathcal{C}=0.5772\ldots$ is the Euler constant
9	$\sin(ax)$	$\dfrac{a}{p^2+a^2}$	
10	$\cos(ax)$	$\dfrac{p}{p^2+a^2}$	
11	$\operatorname{erfc}\Big(\dfrac{a}{2\sqrt{x}}\Big)$	$\dfrac{1}{p}\exp\big(-a\sqrt{p}\big)$	$a\ge 0$
12	$J_0(ax)$	$\dfrac{1}{\sqrt{p^2+a^2}}$	$J_0(x)$ is the Bessel function

M10.2.3. Limit Theorems. Representation of Inverse Transforms as Convergent Series

▶ **Limit theorems.**

THEOREM 1. *Let* $0 \le x < \infty$ *and* $\widetilde{f}(p) = \mathfrak{L}\bigl\{f(x)\bigr\}$ *be the Laplace transform of* $f(x)$. *If a limit of* $f(x)$ *as* $x \to 0$ *exists, then*

$$\lim_{x\to 0} f(x) = \lim_{p\to\infty} \bigl[p\widetilde{f}(p)\bigr].$$

THEOREM 2. *If a limit of* $f(x)$ *as* $x \to \infty$ *exists, then*

$$\lim_{x\to \infty} f(x) = \lim_{p\to 0} \bigl[p\widetilde{f}(p)\bigr].$$

▶ **Representation of inverse transforms as convergent series.**

THEOREM 1. *Suppose the transform* $\widetilde{f}(p)$ *can be expanded into series in negative powers of* p,

$$\widetilde{f}(p) = \sum_{n=1}^{\infty} \frac{a_n}{p^n},$$

convergent for $|p| > R$, *where* R *is an arbitrary positive number; note that the transform tends to zero as* $|p| \to \infty$. *Then the inverse transform can be obtained by the formula*

$$f(x) = \sum_{n=1}^{\infty} \frac{a_n}{(n-1)!} x^{n-1},$$

where the series on the right-hand side is convergent for all x.

THEOREM 2. *Suppose the transform* $\widetilde{f}(p)$, $|p| > R$, *is represented by an absolutely convergent series,*

$$\widetilde{f}(p) = \sum_{n=0}^{\infty} \frac{a_n}{p^{\lambda_n}}, \qquad \text{(M10.2.3.1)}$$

where $\{\lambda_n\}$ *is any positive increasing sequence,* $0 < \lambda_0 < \lambda_1 < \cdots \to \infty$. *Then it is possible to proceed termwise from series (M10.2.3.1) to the following inverse transform series:*

$$f(x) = \sum_{n=0}^{\infty} \frac{a_n}{\Gamma(\lambda_n)} x^{\lambda_n - 1}, \qquad \text{(M10.2.3.2)}$$

where $\Gamma(\lambda)$ *is the Gamma function. Series (M10.2.3.2) is convergent for all real and complex values of* x *other than zero (if* $\lambda_0 \ge 1$, *the series is convergent for all* x*).*

M10.3. Various Forms of the Fourier Transform

M10.3.1. Fourier Transform and the Inverse Fourier Transform

▶ **Standard form of the Fourier transform.** The *Fourier transform* is defined as follows:

$$\widetilde{f}(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x)\, e^{-iux}\, dx. \qquad \text{(M10.3.1.1)}$$

For brevity, we rewrite formula (M10.3.1.1) as follows:

$$\widetilde{f}(u) = \mathfrak{F}\{f(x)\} \qquad \text{or} \qquad \widetilde{f}(u) = \mathfrak{F}\{f(x), u\}.$$

Given $\widetilde{f}(u)$, the function $f(x)$ can be found by means of the *inverse Fourier transform*

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \widetilde{f}(u)\, e^{iux}\, du. \tag{M10.3.1.2}$$

Formula (M10.3.1.2) holds for continuous functions. If $f(x)$ has a (finite) jump discontinuity at a point $x = x_0$, then the right-hand side of (M10.3.1.2) evaluates to $\frac{1}{2}\big[f(x_0-0) + f(x_0+0)\big]$ at this point.

For brevity, we rewrite formula (M10.3.1.2) as follows:

$$f(x) = \mathfrak{F}^{-1}\{\widetilde{f}(u)\} \qquad \text{or} \qquad f(x) = \mathfrak{F}^{-1}\{\widetilde{f}(u),\, x\}.$$

► Convolution theorem. Main properties of the Fourier transform.

1°. The *convolution* of two functions $f(x)$ and $g(x)$ is defined as

$$f(x) * g(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x-t) g(t)\, dt.$$

By performing substitution $x - t = u$, we see that the convolution is symmetric with respect to the convolved functions: $f(x) * g(x) = g(x) * f(x)$.

The *convolution theorem* states that

$$\mathfrak{F}\big\{f(x) * g(x)\big\} = \mathfrak{F}\big\{f(x)\big\}\, \mathfrak{F}\big\{g(x)\big\}.$$

2°. The main properties of the correspondence between functions and their Fourier transforms are gathered in Table M10.3.

TABLE M10.3
Main properties of the Fourier transform.

No.	Function	Fourier transform	Operation
1	$af_1(x) + bf_2(x)$	$a\widetilde{f}_1(u) + b\widetilde{f}_2(u)$	Linearity
2	$f(x/a),\ a > 0$	$a\widetilde{f}(au)$	Scaling
3	$x^n f(x);\ n = 1, 2, \ldots$	$i^n \widetilde{f}_u^{(n)}(u)$	Differentiation of the transform
4	$f''_{xx}(x)$	$-u^2 \widetilde{f}(u)$	Differentiation
5	$f_x^{(n)}(x)$	$(iu)^n \widetilde{f}(u)$	Differentiation
6	$\int_{-\infty}^{\infty} f_1(\xi) f_2(x-\xi)\, d\xi$	$\widetilde{f}_1(u)\widetilde{f}_2(u)$	Convolution

M10.3.2. Fourier Cosine and Sine Transforms

► Fourier cosine transform.

1°. Let a function $f(x)$ be integrable on the semiaxis $0 \le x < \infty$. The *Fourier cosine transform* is defined by

$$\widetilde{f}_{\mathrm{c}}(u) = \sqrt{\frac{2}{\pi}} \int_0^\infty f(x) \cos(xu)\, dx, \qquad 0 < u < \infty. \tag{M10.3.2.1}$$

Given $\widetilde{f}_{\mathrm{c}}(u)$, the function $f(x)$ can be found by means of the *Fourier cosine inversion formula*

$$f(x) = \sqrt{\frac{2}{\pi}} \int_0^\infty \widetilde{f}_{\mathrm{c}}(u) \cos(xu)\, du, \qquad 0 < x < \infty. \tag{M10.3.2.2}$$

The Fourier cosine transform (M10.3.2.1) is denoted for brevity by $\widetilde{f}_{\mathrm{c}}(u) = \mathfrak{F}_{\mathrm{c}}\{f(x)\}$.

2°. Some other properties of the Fourier cosine transform:

$$\begin{aligned} \mathfrak{F}_{\mathrm{c}}\{x^{2n} f(x)\} &= (-1)^n \frac{d^{2n}}{du^{2n}} \mathfrak{F}_{\mathrm{c}}\{f(x)\}, \quad n = 1, 2, \dots; \\ \mathfrak{F}_{\mathrm{c}}\{f''(x)\} &= -u^2 \mathfrak{F}_{\mathrm{c}}\{f(x)\}. \end{aligned}$$

The function $f(x)$ here is assumed to vanish sufficiently rapidly (exponentially) as $x \to \infty$. In the latter formula, the condition $f'(0) = 0$ is assumed to hold.

Parseval's relation for the Fourier cosine transform:

$$\int_0^\infty \mathfrak{F}_{\mathrm{c}}\{f(x)\} \mathfrak{F}_{\mathrm{c}}\{g(x)\}\, du = \int_0^\infty f(x) g(x)\, dx.$$

There are tables of the Fourier cosine transform (see Section S2.3 and the references listed at the end of the current chapter).

► Fourier sine transform.

1°. Let a function $f(x)$ be integrable on the semiaxis $0 \le x < \infty$. The *Fourier sine transform* is defined by

$$\widetilde{f}_{\mathrm{s}}(u) = \sqrt{\frac{2}{\pi}} \int_0^\infty f(x) \sin(xu)\, dx, \qquad 0 < u < \infty. \tag{M10.3.2.3}$$

For given $\widetilde{f}_{\mathrm{s}}(u)$, the function $f(x)$ can be found by means of the *inverse Fourier sine transform*

$$f(x) = \sqrt{\frac{2}{\pi}} \int_0^\infty \widetilde{f}_{\mathrm{s}}(u) \sin(xu)\, du, \qquad 0 < x < \infty. \tag{M10.3.2.4}$$

The Fourier sine transform (M10.3.2.3) is briefly denoted by $\widetilde{f}_{\mathrm{s}}(u) = \mathfrak{F}_{\mathrm{s}}\{f(x)\}$.

2°. Some other properties of the Fourier sine transform:

$$\mathfrak{F}_\mathrm{s}\bigl\{x^{2n}f(x)\bigr\}=(-1)^n\frac{d^{2n}}{du^{2n}}\mathfrak{F}_\mathrm{s}\bigl\{f(x)\bigr\},\qquad n=1,\,2,\,\dots\,;$$
$$\mathfrak{F}_\mathrm{s}\bigl\{f''(x)\bigr\}=-u^2\mathfrak{F}_\mathrm{s}\bigl\{f(x)\bigr\}.$$

The function $f(x)$ here is assumed to vanish sufficiently rapidly (exponentially) as $x\to\infty$. In the latter formula, the condition $f(0)=0$ is assumed to hold.

Parseval's relation for the Fourier sine transform:

$$\int_0^\infty \mathfrak{F}_\mathrm{s}\bigl\{f(x)\bigr\}\mathfrak{F}_\mathrm{s}\bigl\{g(x)\bigr\}\,du=\int_0^\infty f(x)g(x)\,dx.$$

There are tables of the Fourier sine transform (see Section S2.4 and the references listed at the end of the current chapter).

M10.4. Mellin Transform and Other Transforms

M10.4.1. Mellin Transform and the Inversion Formula

► **Mellin transform.** Suppose that a function $f(x)$ is defined for positive x and satisfies the conditions

$$\int_0^1 |f(x)|\,x^{\sigma_1-1}\,dx<\infty,\qquad \int_1^\infty |f(x)|\,x^{\sigma_2-1}\,dx<\infty$$

for some real numbers σ_1 and σ_2, $\sigma_1<\sigma_2$.

The Mellin transform of $f(x)$ is defined by

$$\hat f(s)=\int_0^\infty f(x)x^{s-1}\,dx,\tag{M10.4.1.1}$$

where $s=\sigma+i\tau$ is a complex variable ($\sigma_1<\sigma<\sigma_2$).

For brevity, we rewrite formula (M10.4.1.1) as follows:

$$\hat f(s)=\mathfrak{M}\{f(x)\}\qquad\text{or}\qquad \hat f(s)=\mathfrak{M}\{f(x),s\}.$$

► **Inverse Mellin transform.** Given $\hat f(s)$, the function $f(x)$ can be found by means of the *inverse Mellin transform*

$$f(x)=\frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty}\hat f(s)x^{-s}\,ds\qquad(\sigma_1<\sigma<\sigma_2),\tag{M10.4.1.2}$$

where the integration path is parallel to the imaginary axis of the complex plane s and the integral is understood in the sense of the Cauchy principal value.

Formula (M10.4.1.2) holds for continuous functions. If $f(x)$ has a (finite) jump discontinuity at a point $x=x_0>0$, then the right-hand side of (M10.4.1.2) evaluates to $\frac12\bigl[f(x_0-0)+f(x_0+0)\bigr]$ at this point (for $x_0=0$, the first term in the square brackets must be omitted).

For brevity, we rewrite formula (M10.4.1.2) in the form

$$f(x)=\mathfrak{M}^{-1}\{\hat f(s)\}\qquad\text{or}\qquad f(x)=\mathfrak{M}^{-1}\{\hat f(s),\,x\}.$$

M10.4.2. Main Properties of the Mellin Transform. Relation Among the Mellin, Laplace, and Fourier Transforms

► **Main properties of the Mellin transform.** The main properties of the correspondence between the functions and their Mellin transforms are gathered in Table M10.4.

TABLE M10.4
Main properties of the Mellin transform.

No.	Function	Mellin transform	Operation
1	$af_1(x)+bf_2(x)$	$a\hat{f}_1(s)+b\hat{f}_2(s)$	Linearity
2	$f(ax),\ a>0$	$a^{-s}\hat{f}(s)$	Scaling
3	$x^a f(x)$	$\hat{f}(s+a)$	Shift of the argument of the transform
4	$f(x^2)$	$\frac{1}{2}\hat{f}\left(\frac{1}{2}s\right)$	Squared argument
5	$f(1/x)$	$\hat{f}(-s)$	Inversion of the argument of the transform
6	$x^\lambda f\left(ax^\beta\right),\ \ a>0, \beta\neq 0$	$\dfrac{1}{\beta}a^{-\frac{s+\lambda}{\beta}}\hat{f}\left(\dfrac{s+\lambda}{\beta}\right)$	Power law transform
7	$f'_x(x)$	$-(s-1)\hat{f}(s-1)$	Differentiation
8	$xf'_x(x)$	$-s\hat{f}(s)$	Differentiation
9	$f_x^{(n)}(x)$	$(-1)^n\dfrac{\Gamma(s)}{\Gamma(s-n)}\hat{f}(s-n)$	Multiple differentiation
10	$\left(x\dfrac{d}{dx}\right)^n f(x)$	$(-1)^n s^n\hat{f}(s)$	Multiple differentiation
11	$x^\alpha\displaystyle\int_0^\infty t^\beta f_1(xt)f_2(t)\,dt$	$\hat{f}_1(s+\alpha)\hat{f}_2(1-s-\alpha+\beta)$	Complicated integration
12	$x^\alpha\displaystyle\int_0^\infty t^\beta f_1\left(\dfrac{x}{t}\right)f_2(t)\,dt$	$\hat{f}_1(s+\alpha)\hat{f}_2(s+\alpha+\beta+1)$	Complicated integration

► **Relation among the Mellin, Laplace, and Fourier transforms.** There are tables of direct and inverse Mellin transforms (see the references listed at the end of the current chapter) that are useful in solving specific integral and differential equations. The Mellin transform is related to the Laplace and Fourier transforms by

$$\mathfrak{M}\{f(x),s\} = \mathfrak{L}\{f(e^x),-s\} + \mathfrak{L}\{f(e^{-x}),s\} = \mathfrak{F}\{f(e^x),is\},$$

which makes it possible to apply much more common tables of direct and inverse Laplace and Fourier transforms.

M10.4.3. Summary Table of Integral Transforms

Table M10.5 summarizes the integral transforms considered above and also lists some other integral transforms; for the constraints imposed on the functions and parameters occurring in the integrand, see the references given at the end of this section.

TABLE M10.5
Summary table of integral transforms.

Integral transform	Definition	Inversion formula
Laplace transform	$\widetilde{f}(p)=\int_0^\infty e^{-px}f(x)\,dx$	$f(x)=\dfrac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}(p)\,dp$
Laplace–Carson transform	$\widetilde{f}(p)=p\int_0^\infty e^{-px}f(x)\,dx$	$f(x)=\dfrac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{px}\dfrac{\widetilde{f}(p)}{p}\,dp$
Two-sided Laplace transform	$\widetilde{f}_*(p)=\int_{-\infty}^\infty e^{-px}f(x)\,dx$	$f(x)=\dfrac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}_*(p)\,dp$
Fourier transform	$\widetilde{f}(u)=\dfrac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-iux}f(x)\,dx$	$f(x)=\dfrac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{iux}\widetilde{f}(u)\,du$
Fourier sine transform	$\widetilde{f}_{\rm s}(u)=\sqrt{\dfrac{2}{\pi}}\int_0^\infty \sin(xu)f(x)\,dx$	$f(x)=\sqrt{\dfrac{2}{\pi}}\int_0^\infty \sin(xu)\widetilde{f}_{\rm s}(u)\,du$
Fourier cosine transform	$\widetilde{f}_{\rm c}(u)=\sqrt{\dfrac{2}{\pi}}\int_0^\infty \cos(xu)f(x)\,dx$	$f(x)=\sqrt{\dfrac{2}{\pi}}\int_0^\infty \cos(xu)\widetilde{f}_{\rm c}(u)\,du$
Hartley transform	$\widetilde{f}_{\rm h}(u)=\dfrac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty (\cos xu+\sin xu)f(x)\,dx$	$f(x)=\dfrac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty (\cos xu+\sin xu)\widetilde{f}_{\rm h}(u)\,du$
Mellin transform	$\widehat{f}(s)=\int_0^\infty x^{s-1}f(x)\,dx$	$f(x)=\dfrac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} x^{-s}\widehat{f}(s)\,ds$
Hankel transform	$\widehat{f}_\nu(w)=\int_0^\infty xJ_\nu(xw)f(x)\,dx$	$f(x)=\int_0^\infty wJ_\nu(xw)\widehat{f}_\nu(w)\,dw$
Y-transform	$F_\nu(u)=\int_0^\infty \sqrt{ux}\,Y_\nu(ux)f(x)\,dx$	$f(x)=\int_0^\infty \sqrt{ux}\,\mathbf{H}_\nu(ux)F_\nu(u)\,du$
Meijer transform (K-transform)	$\widehat{f}(s)=\sqrt{\dfrac{2}{\pi}}\int_0^\infty \sqrt{sx}\,K_\nu(sx)f(x)\,dx$	$f(x)=\dfrac{1}{i\sqrt{2\pi}}\int_{c-i\infty}^{c+i\infty} \sqrt{sx}\,I_\nu(sx)\widehat{f}(s)\,ds$
Kontorovich–Lebedev transform	$F(\tau)=\int_0^\infty K_{i\tau}(x)f(x)\,dx$	$f(x)=\dfrac{2}{\pi^2 x}\int_0^\infty \tau\sinh(\pi\tau)K_{i\tau}(x)F(\tau)\,d\tau$

NOTATION: $i=\sqrt{-1}$; $J_\mu(x)$ and $Y_\mu(x)$ are the Bessel functions of the first and the second kind, respectively; $I_\mu(x)$ and $K_\mu(x)$ are the modified Bessel functions of the first and the second kind, respectively; and $\mathbf{H}_\nu(x)=\sum_{j=0}^\infty \dfrac{(-1)^j(x/2)^{\nu+2j+1}}{\Gamma\bigl(j+\frac{3}{2}\bigr)\Gamma\bigl(\nu+j+\frac{3}{2}\bigr)}$ is the Struve function.

Bibliography for Chapter M10

Bateman, H. and Erdélyi, A., *Tables of Integral Transforms. Vols. 1 and 2*, McGraw-Hill, New York, 1954.

Beerends, R. J., ter Morschem, H. G., and van den Berg, J. C., *Fourier and Laplace Transforms*, Cambridge University Press, Cambridge, England, 2003.

Bellman, R. and Roth, R., *The Laplace Transform*, World Scientific Publishing Co., Singapore, 1984.

Ditkin, V. A. and Prudnikov, A. P., *Integral Transforms and Operational Calculus*, Pergamon Press, New York, 1965.

Oberhettinger, F., *Tables of Fourier Transforms and Fourier Transforms of Distributions*, Springer-Verlag, Berlin, 1980.

Oberhettinger, F. and Badii, L., *Tables of Laplace Transforms*, Springer-Verlag, New York, 1973.

Prudnikov, A. P., Brychkov, Yu. A., and Marichev, O. I., *Integrals and Series, Vol. 4, Direct Laplace Transform*, Gordon & Breach, New York, 1992.

Prudnikov, A. P., Brychkov, Yu. A., and Marichev, O. I., *Integrals and Series, Vol. 5, Inverse Laplace Transform*, Gordon & Breach, New York, 1992.

Sneddon, I., *Fourier Transforms*, Dover Publications, New York, 1995.

Chapter M11

Ordinary Differential Equations

M11.1. First-Order Differential Equations

M11.1.1. General Concepts. The Cauchy Problem. Uniqueness and Existence Theorems

► **Equations solved for the derivative. General solution.** A *first-order ordinary differential equation** solved for the derivative has the form

$$y'_x = f(x,y). \qquad \text{(M11.1.1.1)}$$

Sometimes it is represented in terms of differentials as $dy = f(x,y)\,dx$.

A *solution of a differential equation* is a function $y(x)$ that, when substituted into the equation, turns it into an identity. The *general solution of a differential equation* is the set of all its solutions. In some cases, the general solution can be represented as a function $y = \varphi(x,C)$ that depends on one *arbitrary constant* C; specific values of C define specific solutions of the equation (*particular solutions*). In practice, the general solution more frequently appears in implicit form, $\Phi(x,y,C) = 0$, or parametric form, $x = x(t,C)$, $y = y(t,C)$.

Geometrically, the general solution (also called the general integral) of an equation is a family of curves in the xy-plane depending on a single parameter C; these curves are called *integral curves* of the equation. To each particular solution (particular integral) there corresponds a single curve that passes through a given point (x_0,y_0) in the plane.

For each point (x,y), the equation $y'_x = f(x,y)$ defines a value of y'_x, i.e., the slope of the integral curve that passes through this point. In other words, the equation generates a field of directions in the xy-plane. From the geometrical point of view, the problem of solving a first-order differential equation involves finding the curves, the slopes of which at each point coincide with the direction of the field at this point.

Figure M11.1 depicts the tangent to an integral curve at a point (x_0,y_0); the slope of the integral curve at this point is determined by the right-hand side of equation (M11.1.1.1): $\tan\alpha = f(x_0,y_0)$. The little segments show the field of tangents to the integral curves of the differential equation (M11.1.1.1) at other points.

► **Equations integrable by quadrature.** The process of finding a solution to a differential equation is called integration of this equation. The problem of integration of equation (M11.1.1.1) can often be reduced to the problem of finding indefinite integrals, or quadratures. A solution is expressed as a quadrature if it expressed in terms of elementary functions and the functions appearing in the equation using a finite set of the arithmetic operations, function compositions, and indefinite integrals. An equation is said to be integrable by quadrature if its general solution can be expressed in terms of quadratures.

---

* In what follows, we often call an ordinary differential equation a “differential equation” or, even shorter, an “equation.”

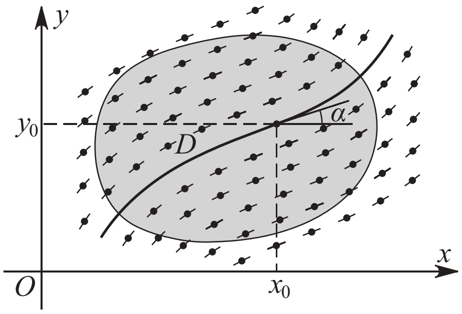

Figure M11.1. The direction field of a differential equation and the integral curve passing through a point (x_0, y_0).

► **Cauchy problem. The uniqueness and existence theorems.** The *Cauchy problem*: find a solution of equation (M11.1.1.1) that satisfies the *initial condition*

$$y = y_0 \quad \text{at} \quad x = x_0, \tag{M11.1.1.2}$$

where y_0 and x_0 are some numbers.

Geometrical meaning of the Cauchy problem: find an integral curve of equation (M11.1.1.1) that passes through the point (x_0, y_0); see Fig. M11.1.

Condition (M11.1.1.2) is alternatively written $y(x_0) = y_0$ or $y|_{x=x_0} = y_0$.

THEOREM (EXISTENCE, PEANO). *Let the function $f(x, y)$ be continuous in an open domain D of the xy-plane. Then there is at least one integral curve of equation (M11.1.1.1) that passes through a point $(x_0, y_0) \in D$; each of these curves can be extended at both ends up to the boundary of any closed domain $D_0 \subset D$ such that (x_0, y_0) belongs to the interior of D_0.*

THEOREM (UNIQUENESS). *Let the function $f(x, y)$ be continuous in an open domain D and have in D a bounded partial derivative with respect to y (or the Lipschitz condition holds: $|f(x, y) - f(x, z)| \le M|y - z|$, where M is some positive number and $(x, z) \in D$). Then there is a unique solution of equation (M11.1.1.1) satisfying condition (M11.1.1.2).*

► **Equations not solved for the derivative. The existence theorem.** A first-order differential equation not solved for the derivative can generally be written as

$$F(x, y, y'_x) = 0. \tag{M11.1.1.3}$$

THEOREM (EXISTENCE AND UNIQUENESS). *There exists a unique solution $y = y(x)$ of equation (M11.1.1.3) satisfying the conditions $y|_{x=x_0} = y_0$ and $y'_x|_{x=x_0} = t_0$, where t_0 is one of the real roots of the equation $F(x_0, y_0, t_0) = 0$ if the following conditions hold in a neighborhood of the point (x_0, y_0, t_0):*

1. *The function $F(x, y, t)$ is continuous in each of the three arguments.*
2. *The partial derivative F_t exists and is nonzero.*
3. *There is a bounded partial derivative with respect to y, $|F_y| \le M$.*

The solution surely exists if $|x - x_0| \le a$, where a is a (sufficiently small) positive number.

► **Singular solutions.** A point (x, y) at which the uniqueness of the solution to equation (M11.1.1.3) is violated is called a *singular* point. If conditions 1 and 3 of the existence and uniqueness theorem hold, then

$$F(x, y, t) = 0, \qquad F_t(x, y, t) = 0 \tag{M11.1.1.4}$$

simultaneously at each singular point. Relations (M11.1.1.4) define a *t-discriminant curve* in parametric form. In some cases, the parameter t can be eliminated from (M11.1.1.4) to give an equation of this curve in implicit form, $\Psi(x, y) = 0$. If a branch $y = \psi(x)$ of the curve $\Psi(x, y) = 0$ consists of singular points and, at the same time, is an integral curve, then this branch is called a *singular integral curve* and the function $y = \psi(x)$ is a *singular solution* of equation (M11.1.1.3).

The singular solutions can be obtained by finding the *envelope of the family of integral curves*, $\Phi(x, y, C) = 0$, of equation (M11.1.1.3). The envelope is part of the *C-discriminant curve*, which is defined by the equations

$$\Phi(x, y, C) = 0, \qquad \Phi_C(x, y, C) = 0.$$

The branch of the C-discriminant curve at which

(a) there exist bounded partial derivatives, $|\Phi_x| < M_1$ and $|\Phi_y| < M_2$, and
(b) $|\Phi_x| + |\Phi_y| \neq 0$

is the envelope.

M11.1.2. Equations Solved for the Derivative. Simplest Techniques of Integration

► Equations with separated or separable variables.

1°. An *equation with separated variables* (a *separated equation*) has the form

$$f(y)y'_x = g(x).$$

Equivalently, the equation can be rewritten as $f(y)\,dy = g(x)\,dx$ (the right-hand side depends on x alone and the left-hand side on y alone). The general solution can be obtained by integration:

$$\int f(y)\,dy = \int g(x)\,dx + C,$$

where C is an arbitrary constant.

2°. An *equation with separable variables* (a *separable equation*) is generally represented by

$$f_1(y)g_1(x)y'_x = f_2(y)g_2(x).$$

Dividing the equation by $f_2(y)g_1(x)$, one obtains a separated equation. Integrating yields

$$\int \frac{f_1(y)}{f_2(y)}\,dy = \int \frac{g_2(x)}{g_1(x)}\,dx + C.$$

Remark. Solutions corresponding to $f_2(y) = 0$ may be lost when dividing the equation by $f_2(y)g_1(x)$. Therefore, the case of $f_2(y) = 0$, should be treated separately.

Many ordinary differential equations are reduced to separable equations.

Example. The equation

$$y'_x = yf(e^{\lambda x}y^k)$$

can be reduced, with the substitution $z = e^{\lambda x}y^k$, to a separable equation: $z'_x = \lambda z + kzf(z)$.

Some other equations reducible to separable equations are considered below.

► **Equation of the form $y'_x = f(ax + by)$.** For $b = 0$, it is an equation with separated variables. For $b \neq 0$, the substitution $z = ax + by$ brings the equation to a separable equation, $z'_x = bf(z) + a$.

► **Homogeneous equation.** A *homogeneous equation* remains the same under simultaneous scaling (dilation) of the independent and dependent variables in accordance with the rule $x \to \alpha x$, $y \to \alpha y$, where α is an arbitrary constant ($\alpha \neq 0$). Such equations can be represented in the form

$$y'_x = f\left(\frac{y}{x}\right).$$

The substitution $u = y/x$ brings a homogeneous equation to a separable one, $xu'_x = f(u) - u$.

Remark. The equations of the form

$$y'_x = f\left(\frac{a_1x + b_1y + c_1}{a_2x + b_2y + c_2}\right)$$

can be reduced to a homogeneous equation. To this end, for $a_1x + b_1y \neq k(a_2x + b_2y)$, one should use the change of variables $\xi = x - x_0$, $\eta = y - y_0$, where the constants x_0 and y_0 are determined by solving the linear algebraic system

$$a_1x_0 + b_1y_0 + c_1 = 0,$$
$$a_2x_0 + b_2y_0 + c_2 = 0.$$

As a result, one arrives at the following equation for $\eta = \eta(\xi)$:

$$\eta'_\xi = f\left(\frac{a_1\xi + b_1\eta}{a_2\xi + b_2\eta}\right).$$

On dividing the numerator and denominator of the argument of f by ξ, one obtains a homogeneous equation whose right-hand side is dependent on the ratio η/ξ only:

$$\eta'_\xi = f\left(\frac{a_1 + b_1\eta/\xi}{a_2 + b_2\eta/\xi}\right).$$

► **Generalized homogeneous equation.**

1°. A *generalized homogeneous equation* remains the same under simultaneous scaling of the independent and dependent variables in accordance with the rule $x \to \alpha x$, $y \to \alpha^k y$, where $\alpha \neq 0$ is an arbitrary constant and k is some number. Such equations can be represented in the form

$$y'_x = x^{k-1} f(yx^{-k}).$$

The substitution $u = yx^{-k}$ brings a generalized homogeneous equation to a separable equation, $xu'_x = f(u) - ku$.

Example. Consider the equation

$$y'_x = ax^2y^4 + by^2. \qquad \text{(M11.1.2.1)}$$

Let us perform the transformation $x = \alpha\bar{x}$, $y = \alpha^k\bar{y}$ and then multiply the resulting equation by α^{1-k} to obtain

$$\bar{y}'_{\bar{x}} = a\alpha^{3(k+1)}\bar{x}^2\bar{y}^4 + b\alpha^{k+1}\bar{y}^2. \qquad \text{(M11.1.2.2)}$$

It is apparent that if $k = -1$, the transformed equation (M11.1.2.2) is the same as the original one, up to notation. This means that equation (M11.1.2.1) is generalized homogeneous of degree $k = -1$. Therefore, the substitution $u = xy$ brings it to a separable equation: $xu'_x = au^4 + bu^2 + u$.

2°. Alternatively, a generalized homogeneous equation can be represented as

$$y'_x = \frac{y}{x} f(x^n y^m).$$

The substitution $z = x^n y^m$ leads to a separable equation: $xz'_x = nz + mzf(z)$.

► **Linear equation.** A first-order *linear equation* is written as

$$y'_x + f(x)y = g(x). \tag{M11.1.2.3}$$

The solution is sought in the product form $y = uv$, where $v = v(x)$ is any function that satisfies the "truncated" equation $v'_x + f(x)v = 0$ [as $v(x)$ one takes the particular solution $v = e^{-F}$, where $F = \int f(x)\,dx$]. As a result, one obtains the following separable equation for $u = u(x)$: $v(x)u'_x = g(x)$. Integrating it yields the general solution:

$$y(x) = e^{-F}\Big(\int e^F g(x)\,dx + C\Big), \qquad F = \int f(x)\,dx,$$

where C is an arbitrary constant.

► **Bernoulli equation.** A *Bernoulli equation* has the form

$$y'_x + f(x)y = g(x)y^a, \qquad a \neq 0,\ 1. \tag{M11.1.2.4}$$

(For $a = 0$ and $a = 1$, it is a linear equation; see above.) The substitution $z = y^{1-a}$ brings it to a linear equation, $z'_x + (1-a)f(x)z = (1-a)g(x)$. With this in view, one can obtain the general integral:

$$y^{1-a} = Ce^{-F} + (1-a)e^{-F}\int e^F g(x)\,dx, \qquad \text{where} \quad F = (1-a)\int f(x)\,dx.$$

► **Equation with exponential nonlinearity reducible to a linear equation.** Consider the equation

$$y'_x = f(x)e^{\lambda y} + g(x).$$

The substitution $u = e^{-\lambda y}$ leads to a linear equation: $u'_x = -\lambda g(x)u - \lambda f(x)$.

M11.1.3. Exact Differential Equations. Integrating Factor

► **Exact differential equations.** An *exact differential equation* has the form

$$f(x,y)\,dx + g(x,y)\,dy = 0, \quad \text{where} \quad \frac{\partial f}{\partial y} = \frac{\partial g}{\partial x}. \tag{M11.1.3.1}$$

The left-hand side of the equation is the total differential of a function of two variables $U(x,y)$. In this case, the general integral is given by

$$U(x,y) = C,$$

where C is an arbitrary constant and the function U is determined from the system

$$\frac{\partial U}{\partial x} = f, \quad \frac{\partial U}{\partial y} = g.$$

Integrating the first equation yields $U = \int f(x,y)\,dx + \Psi(y)$ (while integrating, the variable y is treated as a parameter). On substituting this expression into the second equation, one identifies the function Ψ (and hence, U). As a result, the general integral of an exact differential equation can be represented in the form

$$\int_{x_0}^{x} f(\xi,y)\,d\xi + \int_{y_0}^{y} g(x_0,\eta)\,d\eta = C, \tag{M11.1.3.2}$$

where x_0 and y_0 are any numbers (from the domain of definition of the equation).

TABLE M11.1
An integrating factor $\mu = \mu(x, y)$ for some types of ordinary differential equations $f\,dx + g\,dy = 0$, where $f = f(x, y)$ and $g = g(x, y)$. The subscripts x and y indicate the corresponding partial derivatives.

No.	Conditions for f and g	Integrating factor	Remarks
1	$f = y\varphi(xy),\ g = x\psi(xy)$	$\mu = \frac{1}{xf-yg}$	$xf - yg \not\equiv 0$; $\varphi(z)$ and $\psi(z)$ are any functions
2	$f_x = g_y,\ f_y = -g_x$	$\mu = \frac{1}{f^2+g^2}$	$f + ig$ is an analytic function of the complex variable $x + iy$
3	$\frac{f_y-g_x}{g} = \varphi(x)$	$\mu = \exp\bigl[\int \varphi(x)\,dx\bigr]$	$\varphi(x)$ is any function
4	$\frac{f_y-g_x}{f} = \varphi(y)$	$\mu = \exp\bigl[-\int \varphi(y)\,dy\bigr]$	$\varphi(y)$ is any function
5	$\frac{f_y-g_x}{g-f} = \varphi(x+y)$	$\mu = \exp\bigl[\int \varphi(z)\,dz\bigr],\ \ z = x+y$	$\varphi(z)$ is any function
6	$\frac{f_y-g_x}{yg-xf} = \varphi(xy)$	$\mu = \exp\bigl[\int \varphi(z)\,dz\bigr],\ \ z = xy$	$\varphi(z)$ is any function
7	$\frac{x^2(f_y-g_x)}{yg+xf} = \varphi\bigl(\frac{y}{x}\bigr)$	$\mu = \exp\bigl[-\int \varphi(z)\,dz\bigr],\ \ z = \frac{y}{x}$	$\varphi(z)$ is any function
8	$\frac{f_y-g_x}{xg-yf} = \varphi(x^2+y^2)$	$\mu = \exp\bigl[\frac{1}{2}\int \varphi(z)\,dz\bigr],\ z = x^2+y^2$	$\varphi(z)$ is any function
9	$f_y - g_x = \varphi(x)g - \psi(y)f$	$\mu = \exp\bigl[\int \varphi(x)\,dx + \int \psi(y)\,dy\bigr]$	$\varphi(x)$ and $\psi(y)$ are any functions

Example. Consider the equation

$$(ay^n + bx)y'_x + by + cx^m = 0, \qquad \text{or} \qquad (by + cx^m)\,dx + (ay^n + bx)\,dy = 0,$$

defined by the functions $f(x, y) = by + cx^m$ and $g(x, y) = ay^n + bx$. Computing the derivatives, we have

$$\frac{\partial f}{\partial y} = b, \qquad \frac{\partial g}{\partial x} = b \quad \Longrightarrow \quad \frac{\partial f}{\partial y} = \frac{\partial g}{\partial x}.$$

Hence the given equation is an exact differential equation. Its solution can be found using formula (M11.1.3.2) with $x_0 = y_0 = 0$:

$$\frac{a}{n+1}y^{n+1} + bxy + \frac{c}{m+1}x^{m+1} = C.$$

► **Integrating factor.** An *integrating factor* for the equation

$$f(x, y)\,dx + g(x, y)\,dy = 0$$

is a function $\mu(x, y) \not\equiv 0$ such that the left-hand side of the equation, when multiplied by $\mu(x, y)$, becomes a total differential, and the equation itself becomes an exact differential equation.

An integrating factor satisfies the first-order partial differential equation,

$$g\frac{\partial \mu}{\partial x} - f\frac{\partial \mu}{\partial y} = \left(\frac{\partial f}{\partial y} - \frac{\partial g}{\partial x}\right)\mu,$$

which is not generally easier to solve than the original equation.

Table M11.1 lists some special cases where an integrating factor can be found in explicit form.

M11.1.4. Riccati Equation

► **General Riccati equation.** A *Riccati equation* has the general form

$$y'_x = f_2(x)y^2 + f_1(x)y + f_0(x). \qquad \text{(M11.1.4.1)}$$

If $f_2 \equiv 0$, we have a linear equation, and if $f_0 \equiv 0$, we have a Bernoulli equation (see equation (M11.1.2.4) for $a = 2$), whose solutions were given previously. For arbitrary f_2, f_1, and f_0, the Riccati equation is not integrable by quadrature.

► **Use of particular solutions to construct the general solution.**

1°. Given a particular solution $y_0 = y_0(x)$ of the Riccati equation (M11.1.4.1), the general solution can be written as

$$y = y_0(x) + \Phi(x)\Big[C - \int \Phi(x) f_2(x)\,dx\Big]^{-1}, \qquad \text{(M11.1.4.2)}$$

where C is an arbitrary constant and

$$\Phi(x) = \exp\Big\{\int \big[2f_2(x)y_0(x) + f_1(x)\big]\,dx\Big\}. \qquad \text{(M11.1.4.3)}$$

To the particular solution $y_0(x)$ there corresponds $C = \infty$.

2°. Let $y_1 = y_1(x)$ and $y_2 = y_2(x)$ be two different particular solutions of equation (M11.1.4.1). Then the general solution can be expressed as

$$y = \frac{Cy_1 + U(x)y_2}{C + U(x)}, \quad \text{where} \quad U(x) = \exp\Big[\int f_2(y_1 - y_2)\,dx\Big].$$

To the particular solution $y_1(x)$, there corresponds $C = \infty$; and to $y_2(x)$, there corresponds $C = 0$.

3°. Let $y_1 = y_1(x)$, $y_2 = y_2(x)$, and $y_3 = y_3(x)$ be three distinct particular solutions of equation (M11.1.4.1). Then the general solution can be found without quadrature:

$$\frac{y - y_2}{y - y_1}\,\frac{y_3 - y_1}{y_3 - y_2} = C.$$

► **Some transformations.**

1°. The transformation (φ, ψ_1, ψ_2, ψ_3, and ψ_4 are arbitrary functions)

$$x = \varphi(\xi), \qquad y = \frac{\psi_4(\xi)u + \psi_3(\xi)}{\psi_2(\xi)u + \psi_1(\xi)}$$

reduces the Riccati equation (M11.1.4.1) to a Riccati equation for $u = u(\xi)$.

2°. Let $y_0 = y_0(x)$ be a particular solution of equation (M11.1.4.1). Then the substitution $y = y_0 + 1/w$ leads to a first-order linear equation for $w = w(x)$:

$$w'_x + \big[2f_2(x)y_0(x) + f_1(x)\big]w + f_2(x) = 0.$$

For solution of linear equations, see Subsection M11.1.2.

► **Reduction of the Riccati equation to a second-order linear equation.** The substitution

$$u(x) = \exp\Big(-\int f_2 y\,dx\Big)$$

reduces the general Riccati equation (M11.1.4.1) to a second-order linear equation:

$$f_2 u''_{xx} - \big[(f_2)'_x + f_1 f_2\big]u'_x + f_0 f_2^2 u = 0,$$

which often may be easier to solve than the original Riccati equation.

M11.1.5. Equations Not Solved for the Derivative

► **Method of "integration by differentiation."** In the general case, a first-order equation not solved for the derivative,

$$F(x,y,y'_x)=0, \tag{M11.1.5.1}$$

can be rewritten in the equivalent form

$$F(x,y,t)=0, \qquad t=y'_x. \tag{M11.1.5.2}$$

We look for a solution in parametric form: $x=x(t)$, $y=y(t)$. In accordance with the first relation in (M11.1.5.2), the differential of F is given by

$$F_x\,dx+F_y\,dy+F_t\,dt=0. \tag{M11.1.5.3}$$

Using the relation $dy=t\,dx$, we eliminate successively dy and dx from (M11.1.5.3). As a result, we obtain the system of two first-order ordinary differential equations:

$$\frac{dx}{dt}=-\frac{F_t}{F_x+tF_y},\qquad \frac{dy}{dt}=-\frac{tF_t}{F_x+tF_y}. \tag{M11.1.5.4}$$

By finding a solution of this system, one thereby obtains a solution of the original equation (M11.1.5.1) in parametric form, $x=x(t)$, $y=y(t)$.

Remark 1. The application of the above method may lead to loss of individual solutions (satisfying the condition $F_x+tF_y=0$); this issue requires further investigation.

Remark 2. One of the differential equations of system (M11.1.5.4) can be replaced by the algebraic equation $F(x,y,t)=0$; see equation (M11.1.5.2). This technique is used further for solving some equations.

► **Equations of the form $y=f(y'_x)$.** This equation is a special case of equation (M11.1.5.1), with $F(x,y,t)=y-f(t)$. The method of "integration by differentiation" yields

$$\frac{dx}{dt}=\frac{f'(t)}{t},\qquad y=f(t). \tag{M11.1.5.5}$$

Note the original equation is used here instead of the second equation in system (M11.1.5.4); this is convenient because the first equation in (M11.1.5.4) does not depend on y explicitly.

Integrating the first equation in (M11.1.5.5) yields the solution in parametric form,

$$x=\int\frac{f'(t)}{t}\,dt+C,\qquad y=f(t).$$

► **Equations of the form $x=f(y'_x)$.** This equation is a special case of equation (M11.1.5.1), with $F(x,y,t)=x-f(t)$. The method of "integration by differentiation" yields

$$x=f(t),\qquad \frac{dy}{dt}=tf'(t). \tag{M11.1.5.6}$$

Note the original equation is used here instead of the first equation in system (M11.1.5.4); this is convenient because the second equation in (M11.1.5.4) does not depend on x explicitly.

Integrating the second equation in (M11.1.5.6) yields the solution in parametric form,

$$x=f(t),\qquad y=\int tf'(t)\,dt+C.$$

► **Clairaut's equation $y = xy'_x + f(y'_x)$.** *Clairaut's equation* is a special case of equation (M11.1.5.1), with $F(x, y, t) = y - xt - f(t)$. It can be rewritten as

$$y = xt + f(t), \qquad t = y'_x. \tag{M11.1.5.7}$$

This equation corresponds to the degenerate case $F_x + tF_y \equiv 0$, where system (M11.1.5.4) cannot be obtained. One should proceed in the following way: the first relation in (M11.1.5.7) gives $dy = x\,dt + t\,dx + f'(t)\,dt$; performing the substitution $dy = t\,dx$, which follows from the second relation in (M11.1.5.7), one obtains

$$[x + f'(t)]\,dt = 0.$$

This equation splits into $dt = 0$ and $x + f'(t) = 0$. The solution of the first equation is obvious: $t = C$; it gives the general solution of Clairaut's equation,

$$y = Cx + f(C), \tag{M11.1.5.8}$$

which is a family of straight lines. The second equation generates a solution in parametric form,

$$x = -f'(t), \qquad y = -tf'(t) + f(t), \tag{M11.1.5.9}$$

which is a singular solution and is the envelope of the family of lines (M11.1.5.8).

Remark. There are also "compound" solutions of Clairaut's equation; they consist of part of the curve (M11.1.5.9) joined with the tangents at finite points; these tangents are defined by formula (M11.1.5.8).

► **Lagrange's equation $y = xf(y'_x) + g(y'_x)$.** *Lagrange's equation* is a special case of equation (M11.1.5.1), with $F(x, y, t) = y - xf(t) - g(t)$. In the special case $f(t) \equiv t$, it coincides with Clairaut's equation.

The method of "integration by differentiation" yields

$$\frac{dx}{dt} + \frac{f'(t)}{f(t) - t}x = \frac{g'(t)}{t - f(t)}, \qquad y = xf(t) + g(t). \tag{M11.1.5.10}$$

Here, the original equation is used instead of the second equation in system (M11.1.5.4); this is convenient because the first equation in (M11.1.5.4) does not depend on y explicitly.

The first equation of system (M11.1.5.10) is linear and can easily be integrated to obtain a solution to Lagrange's equation in parametric form.

Remark. With the above method, solutions of the form $y = t_k x + g(t_k)$, where the t_k are roots of the equation $f(t) - t = 0$, may be lost. These solutions can be particular or singular solutions of Lagrange's equation.

M11.1.6. Approximate Analytic Methods for Solution of Equations

► **Method of successive approximations (Picard's method).** The method of successive approximations consists of two stages. At the first stage, the Cauchy problem

$$y'_x = f(x, y) \qquad \text{(equation)}, \tag{M11.1.6.1}$$

$$y(x_0) = y_0 \qquad \text{(initial condition)} \tag{M11.1.6.2}$$

is reduced to the equivalent integral equation:

$$y(x) = y_0 + \int_{x_0}^{x} f(t, y(t))\,dt. \tag{M11.1.6.3}$$

Then a solution of equation (M11.1.6.3) is sought using the formula of successive approximations:

$$y_{n+1}(x) = y_0 + \int_{x_0}^{x} f(t, y_n(t))\,dt; \qquad n = 0,\ 1,\ 2,\ \ldots$$

The initial approximation $y_0(x)$ can be chosen arbitrarily; the simplest way is to take $y_0(x) = y_0$. The iterative process converges as $n \to \infty$, provided the conditions of the theorems in Subsection M11.1.1 are satisfied.

► **Method of Taylor series expansion in the independent variable.** A solution of the Cauchy problem (M11.1.6.1)–(M11.1.6.2) can be sought in the form of the Taylor series in powers of $(x - x_0)$:

$$y(x) = y(x_0) + y'_x(x_0)(x - x_0) + \frac{y''_{xx}(x_0)}{2!}(x - x_0)^2 + \cdots . \qquad \text{(M11.1.6.4)}$$

The first term $y(x_0)$ in solution (M11.1.6.4) is prescribed by the initial condition (M11.1.6.2). The values of the derivatives of $y(x)$ at $x = x_0$ are determined from equation (M11.1.6.1) and its derivative equations (obtained by successive differentiation), taking into account the initial condition (M11.1.6.2). In particular, setting $x = x_0$ in (M11.1.6.1) and substituting (M11.1.6.2), one obtains the value of the first derivative:

$$y'_x(x_0) = f(x_0, y_0). \qquad \text{(M11.1.6.5)}$$

Further, differentiating equation (M11.1.6.1) yields

$$y''_{xx} = f_x(x, y) + f_y(x, y)y'_x. \qquad \text{(M11.1.6.6)}$$

On substituting $x = x_0$, as well as the initial condition (M11.1.6.2) and the first derivative (M11.1.6.5), into the right-hand side of this equation, one calculates the value of the second derivative:

$$y''_{xx}(x_0) = f_x(x_0, y_0) + f(x_0, y_0)f_y(x_0, y_0).$$

Likewise, one can determine the subsequent derivatives of y at $x = x_0$.

Solution (M11.1.6.4) obtained by this method can normally be used in only some sufficiently small neighborhood of the point $x = x_0$.

Example. Consider the Cauchy problem for the equation

$$y'_x = e^y + \cos x$$

with the initial condition $y(0) = 0$.

Since $x_0 = 0$, we will be constructing a series in powers of x. It follows from the equation that $y'_x(0) = e^0 + \cos 0 = 2$. Differentiating the original equation yields $y''_{xx} = e^y y'_x - \sin x$. Using the initial condition and the condition $y'_x(0) = 2$ just obtained, we have $y''_{xx}(0) = e^0 \times 2 - \sin 0 = 2$. Similarly, we find that $y'''_{xxx} = e^y y''_{xx} + e^y (y'_x)^2 - \cos x$, whence $y'''_{xxx}(0) = e^0 \times 2 + e^0 \times 2^2 - \cos 0 = 5$.

Substituting the values of the derivatives at $x = 0$ into series (M11.1.6.4), we obtain the desired series representation of the solution: $y = 2x + x^2 + \frac{5}{6}x^3 + \cdots$.

M11.1.7. Numerical Integration of Differential Equations

► **Method of Euler polygonal lines.** Consider the Cauchy problem for the first-order differential equation

$$y'_x = f(x, y)$$

with the initial condition $y(x_0) = y_0$. Our aim is to construct an approximate solution $y = y(x)$ of this equation on an interval $[x_0, x_*]$.

Let us split the interval $[x_0, x_*]$ into n equal segments of length $\Delta x = \dfrac{x_* - x_0}{n}$. We seek approximate values y_1, y_2, …, y_n of the solution $y(x)$ at the partition points x_1, x_2, …, $x_n = x_*$.

For a given initial value $y_0 = y(x_0)$ and a sufficiently small Δx, the values of the unknown function $y_k = y(x_k)$ at the other points $x_k = x_0 + k\Delta x$ are calculated successively by the formula

$$y_{k+1} = y_k + f(x_k, y_k)\Delta x \qquad (\textit{Euler polygonal line}),$$

where $k = 0, 1, \ldots, n - 1$. The Euler method is a single-step method of the first-order approximation (with respect to the step Δx).

► **Single-step methods of the second-order approximation.** Two single-step methods for solving the Cauchy problem in the second-order approximation are specified by the recurrence formulas

$$y_{k+1} = y_k + f\bigl(x_k + \tfrac{1}{2}\Delta x,\ y_k + \tfrac{1}{2} f_k \Delta x)\Delta x \quad \text{(first method)},$$
$$y_{k+1} = y_k + \tfrac{1}{2}\bigl[f_k + f(x_{k+1},\ y_k + f_k\Delta x)\bigr]\Delta x \quad \text{(second method)},$$

where $f_k = f(x_k, y_k)$; $k = 0,\ 1,\ \dots,\ n-1$.

► **Runge–Kutta method of the fourth-order approximation.** This is one of the widely used methods. The unknown values y_k are successively found by the formulas

$$y_{k+1} = y_k + \tfrac{1}{6}(f_1 + 2f_2 + 2f_3 + f_4)\Delta x,$$

where

$$f_1 = f(x_k,\ y_k), \qquad f_2 = f(x_k + \tfrac{1}{2}\Delta x,\ y_k + \tfrac{1}{2} f_1\Delta x),$$
$$f_3 = f(x_k + \tfrac{1}{2}\Delta x,\ y_k + \tfrac{1}{2} f_2\Delta x), \qquad f_4 = f(x_k + \Delta x,\ y_k + f_3\Delta x).$$

Remark 1. All methods described in Subsection M11.1.7 are special cases of the Runge–Kutta method (a detailed description of this method can be found in the monographs listed at the end of the current chapter).

Remark 2. In practice, calculations are performed on the basis of any of the above recurrence formulas with two different steps Δx, $\frac{1}{2}\Delta x$ and an arbitrarily chosen small Δx. Then one compares the results obtained at common points. If these results coincide within the given order of accuracy, one assumes that the chosen step Δx ensures the desired accuracy of calculations. Otherwise, the step is halved and the calculations are performed with the steps $\frac{1}{2}\Delta x$ and $\frac{1}{4}\Delta x$, after which the results are compared again, etc. (Quite often, one compares the results of calculations with steps varying by a factor of ten or more.)

M11.2. Second-Order Linear Differential Equations

M11.2.1. Formulas for the General Solution. Some Transformations

► **Homogeneous linear equations. Formulas for the general solution.**

1°. Consider a second-order homogeneous linear equation in the general form

$$f_2(x)y''_{xx} + f_1(x)y'_x + f_0(x)y = 0. \tag{M11.2.1.1}$$

The *trivial solution*, $y = 0$, is a particular solution of the homogeneous linear equation.

Let $y_1(x)$, $y_2(x)$ be a fundamental system of solutions (nontrivial linearly independent particular solutions) of equation (M11.2.1.1). Then the general solution is given by

$$y = C_1 y_1(x) + C_2 y_2(x), \tag{M11.2.1.2}$$

where C_1 and C_2 are arbitrary constants.

2°. Let $y_1 = y_1(x)$ be any nontrivial particular solution of equation (M11.2.1.1). Then its general solution can be represented as

$$y = y_1\left(C_1 + C_2\int \frac{e^{-F}}{y_1^2}\,dx\right), \qquad \text{where} \quad F = \int \frac{f_1}{f_2}\,dx. \tag{M11.2.1.3}$$

► **Wronskian determinant and Liouville's formula.** The *Wronskian determinant* (or *Wronskian*) is defined by

$$W(x) = \begin{vmatrix} y_1(x) & y_2(x) \\ y_1'(x) & y_2'(x) \end{vmatrix} \equiv y_1(y_2)'_x - y_2(y_1)'_x,$$

where $y_1(x)$, $y_2(x)$ is a fundamental system of solutions of equation (M11.2.1.1).

Liouville's formula:

$$W(x) = W(x_0)\exp\left[-\int_{x_0}^{x} \frac{f_1(t)}{f_2(t)}\,dt\right].$$

► Simplest second-order linear equations and their solutions.

1°. The *second-order linear equation with constant coefficients*

$$y''_{xx} + ay'_x + by = 0 \qquad \text{(M11.2.1.4)}$$

has the following fundamental system of solutions:

$$\begin{array}{lll}
y_1(x) = \exp\bigl(-\tfrac{1}{2}ax\bigr)\sinh\bigl(\tfrac{1}{2}x\sqrt{a^2-4b}\,\bigr), & y_2(x) = \exp\bigl(-\tfrac{1}{2}ax\bigr)\cosh\bigl(\tfrac{1}{2}x\sqrt{a^2-4b}\,\bigr) & \text{if } a^2 > 4b;\\
y_1(x) = \exp\bigl(-\tfrac{1}{2}ax\bigr)\sin\bigl(\tfrac{1}{2}x\sqrt{4b-a^2}\,\bigr), & y_2(x) = \exp\bigl(-\tfrac{1}{2}ax\bigr)\cos\bigl(\tfrac{1}{2}x\sqrt{4b-a^2}\,\bigr) & \text{if } a^2 < 4b;\\
y_1(x) = \exp\bigl(-\tfrac{1}{2}ax\bigr), & y_2(x) = x\exp\bigl(-\tfrac{1}{2}ax\bigr) & \text{if } a^2 = 4b.
\end{array}$$

Remark. In physics equation (M11.2.1.4) is often called an *equation of damped oscillations.*

2°. The *Euler equation*

$$x^2y''_{xx} + axy'_x + by = 0 \qquad \text{(M11.2.1.5)}$$

is reduced by the change of variable $x = ke^t$ ($k \neq 0$) to the second-order linear equation with constant coefficients $y''_{tt} + (a-1)y'_t + by = 0$, which is treated in Item 1°.

Equation (M11.2.1.5) has the following fundamental system of solutions:

$$\begin{array}{lll}
y_1(x) = |x|^{\frac{1-a}{2}+\mu}, & y_2(x) = |x|^{\frac{1-a}{2}-\mu} & \text{if } (1-a)^2 > 4b,\\
y_1(x) = |x|^{\frac{1-a}{2}}, & y_2(x) = |x|^{\frac{1-a}{2}}\ln|x| & \text{if } (1-a)^2 = 4b,\\
y_1(x) = |x|^{\frac{1-a}{2}}\sin(\mu\ln|x|), & y_2(x) = |x|^{\frac{1-a}{2}}\cos(\mu\ln|x|) & \text{if } (1-a)^2 < 4b,
\end{array}$$

where $\mu = \frac{1}{2}|(1-a)^2 - 4b|^{1/2}$.

► Bessel equation and related equations.

1°. The *Bessel equation* has the form

$$x^2y''_{xx} + xy'_x + (x^2-\nu^2)y = 0.$$

It often arises in numerous applications.

Let ν be an arbitrary noninteger. Then the general solution of the Bessel equation is given by

$$y = C_1J_\nu(x) + C_2Y_\nu(x),$$

where $J_\nu(x)$ and $Y_\nu(x)$ are the *Bessel functions* of the first and second kind:

$$J_\nu(x) = \sum_{k=0}^{\infty}\frac{(-1)^k(x/2)^{2k+\nu}}{k!\,\Gamma(\nu+k+1)}, \qquad Y_\nu(x) = \frac{J_\nu(x)\cos\pi\nu - J_{-\nu}(x)}{\sin\pi\nu}.$$

In the case $\nu = n + \frac{1}{2}$, where $n = 0, 1, 2, \ldots$, the Bessel functions are expressed in terms of elementary functions:

$$J_{n+\frac{1}{2}}(x) = \sqrt{\frac{2}{\pi}}\,x^{n+\frac{1}{2}}\Bigl(-\frac{1}{x}\frac{d}{dx}\Bigr)^n\frac{\sin x}{x}, \qquad J_{-n-\frac{1}{2}}(x) = \sqrt{\frac{2}{\pi}}\,x^{n+\frac{1}{2}}\Bigl(\frac{1}{x}\frac{d}{dx}\Bigr)^n\frac{\cos x}{x},$$
$$Y_{n+\frac{1}{2}}(x) = (-1)^{n+1}J_{-n-\frac{1}{2}}(x).$$

The Bessel functions are described in Section M13.6 in detail.

TABLE M11.2
Some second-order linear equations whose solutions are expressed in terms of Bessel functions and modified Bessel functions.

Equation	General solution	Remarks
$y''_{xx} - ax^k y = 0$	$y = C_1\sqrt{x}\, J_{\frac{1}{2q}}\Big(\frac{\sqrt{\lvert a\rvert}}{q}x^q\Big) + C_2\sqrt{x}\, Y_{\frac{1}{2q}}\Big(\frac{\sqrt{\lvert a\rvert}}{q}x^q\Big)$ if $a < 0$ $y = C_1\sqrt{x}\, I_{\frac{1}{2q}}\Big(\frac{\sqrt{a}}{q}x^q\Big) + C_2\sqrt{x}\, K_{\frac{1}{2q}}\Big(\frac{\sqrt{a}}{q}x^q\Big)$ if $a > 0$	$q = \frac{1}{2}(k+2)$
$y''_{xx} + ay'_x + (bx + c)y = 0$	$y = e^{-ax/2}\sqrt{\xi}\,\big[C_1 J_{1/3}\big(\frac{2}{3}\sqrt{b}\,\xi^{3/2}\big) + C_2 Y_{1/3}\big(\frac{2}{3}\sqrt{b}\,\xi^{3/2}\big)\big]$	$\xi = x + \dfrac{4c - a^2}{4b}$
$xy''_{xx} + ay'_x + by = 0$	$y = x^{\frac{1-a}{2}}\big[C_1 J_\nu\big(2\sqrt{bx}\,\big) + C_2 Y_\nu\big(2\sqrt{bx}\,\big)\big]$ if $bx > 0$ $y = x^{\frac{1-a}{2}}\big[C_1 I_\nu\big(2\sqrt{\lvert bx\rvert}\big) + C_2 K_\nu\big(2\sqrt{\lvert bx\rvert}\big)\big]$ if $bx < 0$	$\nu = \lvert 1 - a\rvert$
$xy''_{xx} + ay'_x + bxy = 0$	$y = x^{\frac{1-a}{2}}\big[C_1 J_\nu\big(\sqrt{b}\,x\big) + C_2 Y_\nu\big(\sqrt{b}\,x\big)\big]$ if $b > 0$ $y = x^{\frac{1-a}{2}}\big[C_1 I_\nu\big(\sqrt{\lvert b\rvert}\,x\big) + C_2 K_\nu\big(\sqrt{\lvert b\rvert}\,x\big)\big]$ if $b < 0$	$\nu = \frac{1}{2}\lvert 1 - a\rvert$
$xy''_{xx} + ay'_x + bx^k y = 0$	$y = x^{\frac{1-a}{2}}\Big[C_1 J_\nu\Big(\frac{2\sqrt{b}}{k+1}x^{\frac{k+1}{2}}\Big) + C_2 Y_\nu\Big(\frac{2\sqrt{b}}{k+1}x^{\frac{k+1}{2}}\Big)\Big]$	$\nu = \dfrac{\lvert 1 - a\rvert}{k+1}$
$x^2 y''_{xx} + xy'_x + (x^2 - \nu^2)y = 0$	$y = C_1 J_\nu(x) + C_2 Y_\nu(x)$	The Bessel equation
$x^2 y''_{xx} + xy'_x - (x^2 + \nu^2)y = 0$	$y = C_1 I_\nu(x) + C_2 K_\nu(x)$	The modified Bessel equation
$y''_{xx} + ae^{\lambda x}y = 0$	$y = C_1 J_0(z) + C_2 Y_0(z)$	$z = \dfrac{2\sqrt{a}}{\lambda}e^{\lambda x/2}$
$y''_{xx} + (ae^x - b)y = 0$	$y = C_1 J_{2\sqrt{b}}\big(2\sqrt{a}\,e^{x/2}\big) + C_2 Y_{2\sqrt{b}}\big(2\sqrt{a}\,e^{x/2}\big)$	
$y''_{xx} + ay'_x + (be^{\lambda x} + c)y = 0$	$y = e^{-ax/2}\big[C_1 J_\nu\big(2\lambda^{-1}\sqrt{b}\,e^{\lambda x/2}\big) + C_2 Y_\nu\big(2\lambda^{-1}\sqrt{b}\,e^{\lambda x/2}\big)\big]$	$\nu = \dfrac{1}{\lambda}\sqrt{a^2 - 4c}$

2°. The *modified Bessel equation* has the form

$$x^2 y''_{xx} + xy'_x - (x^2 + \nu^2)y = 0.$$

It can be reduced to the Bessel equation by means of the substitution $x = i\bar{x}$ ($i^2 = -1$).

The general solution of the modified Bessel equation is given by

$$y = C_1 I_\nu(x) + C_2 K_\nu(x),$$

where $I_\nu(x)$ and $K_\nu(x)$ are *modified Bessel functions* of the first and second kind:

$$I_\nu(x) = \sum_{k=0}^{\infty} \frac{(x/2)^{2k+\nu}}{k!\,\Gamma(\nu + k + 1)}, \qquad K_\nu(x) = \frac{\pi}{2}\,\frac{I_{-\nu}(x) - I_\nu(x)}{\sin \pi\nu}.$$

The modified Bessel functions are described in Subsection M13.7 in detail.

3°. Table M11.2. lists some second-order linear equations whose solutions are expressed in terms of Bessel functions and modified Bessel functions.

► **Nonhomogeneous linear equations. The existence theorem.** A second-order nonhomogeneous linear equation has the form

$$f_2(x)y''_{xx} + f_1(x)y'_x + f_0(x)y = g(x). \qquad \text{(M11.2.1.6)}$$

THEOREM (EXISTENCE AND UNIQUENESS). *On an open interval $a<x<b$, let the functions f_2, f_1, f_0, and g be continuous and $f_2 \neq 0$. Also let*

$$y(x_0) = A, \quad y'_x(x_0) = B$$

be arbitrary initial conditions, where x_0 is any point such that $a < x_0 < b$, and A and B are arbitrary prescribed numbers. Then a solution of equation (M11.2.1.6) exists and is unique. This solution is defined for all $x \in (a, b)$.

► **Formulas for the general solution.**

1°. The general solution of the nonhomogeneous linear equation (M11.2.1.6) is the sum of the general solution of the corresponding homogeneous linear equation (M11.2.1.1) and any particular solution of the nonhomogeneous equation (M11.2.1.6).

2°. Let $y_1 = y_1(x)$, $y_2 = y_2(x)$ be a fundamental system of solutions of the corresponding homogeneous equation, with $g \equiv 0$. Then the general solution of nonhomogeneous equation (M11.2.1.6) can be represented as

$$y = C_1y_1 + C_2y_2 + y_2 \int y_1 \frac{g}{f_2}\frac{dx}{W} - y_1 \int y_2 \frac{g}{f_2}\frac{dx}{W}, \qquad \text{(M11.2.1.7)}$$

where $W = y_1(y_2)'_x - y_2(y_1)'_x$ is the Wronskian determinant.

Remark. Given a nontrivial particular solution $y_1 = y_1(x)$ of the homogeneous equation (with $g \equiv 0$), a second particular solution $y_2 = y_2(x)$ of the homogeneous equation can be calculated from formula (M11.2.1.3). Then the general solution of equation (M11.2.1.6) can be constructed by (M11.2.1.3) and (M11.2.1.7).

3°. Let $\bar{y}_1$ and $\bar{y}_2$ be respective solutions of the nonhomogeneous differential equations $\mathrm{L}[\bar{y}_1] = g_1(x)$ and $\mathrm{L}[\bar{y}_2] = g_2(x)$, which have the same left-hand side but different right-hand sides, where $\mathrm{L}[y]$ is the left-hand side of equation (M11.2.1.6). Then the function $\bar{y} = \bar{y}_1 + \bar{y}_2$ is also a solution of the equation $\mathrm{L}[\bar{y}] = g_1(x) + g_2(x)$.

M11.2.2. Representation of Solutions as a Series in the Independent Variable

► **Equation coefficients are representable in power series form.** Let us consider a homogeneous linear differential equation of the general form

$$y''_{xx} + f(x)y'_x + g(x)y = 0. \qquad \text{(M11.2.2.1)}$$

Assume that the functions $f(x)$ and $g(x)$ are representable, in a neighborhood of a point $x = x_0$, in power series form,

$$f(x) = \sum_{n=0}^{\infty} A_n(x-x_0)^n, \qquad g(x) = \sum_{n=0}^{\infty} B_n(x-x_0)^n, \qquad \text{(M11.2.2.2)}$$

on the interval $|x-x_0|<R$, where R stands for the minimum radius of convergence of the two series in (M11.2.2.2). In this case, the point $x=x_0$ is referred to as an *ordinary point*, and equation (M11.2.2.1) possesses two linearly independent solutions of the form

$$y_1(x)=\sum_{n=0}^{\infty}a_n(x-x_0)^n,\qquad y_2(x)=\sum_{n=0}^{\infty}b_n(x-x_0)^n. \tag{M11.2.2.3}$$

The coefficients a_n and b_n are determined by substituting the series (M11.2.2.2) and (M11.2.2.3) into equation (M11.2.2.1) followed by matching the coefficients of like powers of $(x-x_0)$.*

► **Equation coefficients have poles at some point.** Assume that the functions $f(x)$ and $g(x)$ are representable, in a neighborhood of a point $x=x_0$, in the form

$$f(x)=\sum_{n=-1}^{\infty}A_n(x-x_0)^n,\qquad g(x)=\sum_{n=-2}^{\infty}B_n(x-x_0)^n, \tag{M11.2.2.4}$$

on the interval $|x-x_0|<R$. In this case, the point $x=x_0$ is referred to as a *regular singular point*. Let λ_1 and λ_2 be roots of the quadratic equation

$$\lambda^2+(A_{-1}-1)\lambda+B_{-2}=0.$$

Depending on the values of λ_1 and λ_2, three cases are possible. These cases are considered below.

1. If $\lambda_1\neq\lambda_2$ and $\lambda_1-\lambda_2$ is not an integer, equation (M11.2.2.1) has two linearly independent solutions of the form

$$\begin{aligned}y_1(x)&=|x-x_0|^{\lambda_1}\Bigl[1+\sum_{n=1}^{\infty}a_n(x-x_0)^n\Bigr],\\ y_2(x)&=|x-x_0|^{\lambda_2}\Bigl[1+\sum_{n=1}^{\infty}b_n(x-x_0)^n\Bigr].\end{aligned} \tag{M11.2.2.5}$$

2. If $\lambda_1=\lambda_2=\lambda$, equation (M11.2.2.1) possesses two linearly independent solutions:

$$\begin{aligned}y_1(x)&=|x-x_0|^{\lambda}\Bigl[1+\sum_{n=1}^{\infty}a_n(x-x_0)^n\Bigr],\\ y_2(x)&=y_1(x)\ln|x-x_0|+|x-x_0|^{\lambda}\sum_{n=0}^{\infty}b_n(x-x_0)^n.\end{aligned}$$

3. If $\lambda_1=\lambda_2+N$, where N is a positive integer, equation (M11.2.2.1) has two linearly independent solutions of the form

$$\begin{aligned}y_1(x)&=|x-x_0|^{\lambda_1}\Bigl[1+\sum_{n=1}^{\infty}a_n(x-x_0)^n\Bigr],\\ y_2(x)&=ky_1(x)\ln|x-x_0|+|x-x_0|^{\lambda_2}\sum_{n=0}^{\infty}b_n(x-x_0)^n,\end{aligned}$$

where k is a constant to be determined (it may be equal to zero).

---

* Prior to that, the terms containing the same powers $(x-x_0)^k$, $k=0,1,\ldots$, should be collected.

To construct the solution in each of the three cases, the following procedure should be performed: substitute the above expressions of y_1 (resp., y_2) into the original equation (M11.2.2.1), taking into account (M11.2.2.4), and equate the coefficients of $(x-x_0)^n$ and (resp., $(x-x_0)^n \ln|x-x_0|$) for like values of n to obtain recurrence relations for the unknown coefficients. The desired solution can be found from these recurrence relations.

M11.2.3. Boundary Value Problems

► **First, second, third, and mixed boundary value problems ($x_1 \le x \le x_2$).** We consider the second-order nonhomogeneous linear differential equation

$$y''_{xx} + f(x)y'_x + g(x)y = h(x). \qquad \text{(M11.2.3.1)}$$

1°. *The first boundary value problem*: Find a solution of equation (M11.2.3.1) satisfying the boundary conditions

$$y = a_1 \quad \text{at} \quad x = x_1, \qquad y = a_2 \quad \text{at} \quad x = x_2. \qquad \text{(M11.2.3.2)}$$

(The values of the unknown are prescribed at two distinct points x_1 and x_2.)

2°. *The second boundary value problem*: Find a solution of equation (M11.2.3.1) satisfying the boundary conditions

$$y'_x = a_1 \quad \text{at} \quad x = x_1, \qquad y'_x = a_2 \quad \text{at} \quad x = x_2. \qquad \text{(M11.2.3.3)}$$

(The values of the derivative of the unknown are prescribed at two distinct points x_1 and x_2.)

3°. *The third boundary value problem*: Find a solution of equation (M11.2.3.1) satisfying the boundary conditions

$$\begin{aligned} y'_x + k_1 y &= a_1 \quad \text{at} \quad x = x_1,\\ y'_x + k_2 y &= a_2 \quad \text{at} \quad x = x_2. \end{aligned} \qquad \text{(M11.2.3.4)}$$

4°. *The mixed boundary value problem*: Find a solution of equation (M11.2.3.1) satisfying the boundary conditions

$$y = a_1 \quad \text{at} \quad x = x_1, \qquad y'_x = a_2 \quad \text{at} \quad x = x_2. \qquad \text{(M11.2.3.5)}$$

(The unknown is prescribed at one point, and its derivative at another point.)

Conditions (M11.2.3.2), (M11.2.3.3), (M11.2.3.4), and (M11.2.3.5) are called *homogeneous* if $a_1 = a_2 = 0$.

► **Simplification of boundary conditions. The self-adjoint form of equations.**

1°. Nonhomogeneous boundary conditions can be reduced to homogeneous ones by the change of variable $z = A_2x^2 + A_1x + A_0 + y$ (the constants A_2, A_1, and A_0 are selected using the method of undetermined coefficients). In particular, the nonhomogeneous boundary conditions of the first kind (M11.2.3.2) can be reduced to homogeneous boundary conditions by the linear change of variable

$$z = y - \frac{a_2 - a_1}{x_2 - x_1}(x - x_1) - a_1.$$

2°. On multiplying by $p(x) = \exp\Bigl[\int f(x)\,dx\Bigr]$, one reduces equation (M11.2.3.1) to the self-adjoint form:

$$[p(x)y'_x]'_x + q(x)y = r(x). \qquad \text{(M11.2.3.6)}$$

Without loss of generality, we can further consider equation (M11.2.3.6) instead of (M11.2.3.1). We assume that the functions p, p'_x, q, and r are continuous on the interval $x_1 \le x \le x_2$, and p is positive.

► **Green's function. Linear problems for nonhomogeneous equations.** The *Green's function* of the first boundary value problem for equation (M11.2.3.6) with homogeneous boundary conditions (M11.2.3.2) is a function of two variables $G(x,\xi)$ that satisfies the following conditions:

1°. $G(x,\xi)$ is continuous in x for fixed ξ, with $x_1 \le x \le x_2$ and $x_1 \le \xi \le x_2$.

2°. $G(x,\xi)$ is a solution of the homogeneous equation (M11.2.3.6), with $r = 0$, for all $x_1 < x < x_2$ exclusive of the point $x = \xi$.

3°. $G(x,\xi)$ satisfies the homogeneous boundary conditions $G(x_1,\xi) = G(x_2,\xi) = 0$.

4°. The derivative $G'_x(x,\xi)$ has a jump of $1/p(\xi)$ at the point $x = \xi$, that is,

$$G'_x(x,\xi)\big|_{x\to\xi,\,x>\xi} - G'_x(x,\xi)\big|_{x\to\xi,\,x<\xi} = \frac{1}{p(\xi)}.$$

For the second, third, and mixed boundary value problems, the Green's function is defined likewise except that in 3° one adopts, respectively, the homogeneous boundary conditions (M11.2.3.3), (M11.2.3.4), and (M11.2.3.5), with $a_1 = a_2 = 0$.

The solution of the nonhomogeneous equation (M11.2.3.6) subject to appropriate homogeneous boundary conditions is expressed in terms of the Green's function as follows:*

$$y(x) = \int_{x_1}^{x_2} G(x,\xi)r(\xi)\,d\xi.$$

► **Representation of the Green's function in terms of particular solutions.** We consider the first boundary value problem. Let $y_1(x)$ and $y_2(x)$ be linearly independent particular solutions of the homogeneous equation (M11.2.3.6), with $r = 0$, that satisfy the conditions

$$y_1(x_1) = 0, \qquad y_2(x_2) = 0.$$

(Each of the solutions satisfies one of the homogeneous boundary conditions.)

The Green's function is expressed in terms of solutions of the homogeneous equation as follows:

$$G(x,\xi) = \begin{cases} \dfrac{y_1(x)y_2(\xi)}{p(\xi)W(\xi)} & \text{for } x_1 \le x \le \xi, \\[2ex] \dfrac{y_1(\xi)y_2(x)}{p(\xi)W(\xi)} & \text{for } \xi \le x \le x_2, \end{cases} \tag{M11.2.3.7}$$

where $W(x) = y_1(x)y'_2(x) - y'_1(x)y_2(x)$ is the Wronskian determinant.

Remark. Formula (M11.2.3.7) can also be used to construct the Green's functions for the second, third, and mixed boundary value problems. To this end, one should find two linearly independent solutions, $y_1(x)$ and $y_2(x)$, of the homogeneous equation; the former satisfies the corresponding homogeneous boundary condition at $x = x_1$ and the latter satisfies the one at $x = x_2$.

M11.2.4. Eigenvalue Problems

► **Sturm–Liouville problem.** Consider the second-order homogeneous linear differential equation

$$[p(x)y'_x]'_x + [\lambda\rho(x) - q(x)]y = 0 \tag{M11.2.4.1}$$

---

* The homogeneous boundary value problem, with $r(x) = 0$ and $a_1 = a_2 = 0$, is assumed to have only the trivial solution.

subject to linear boundary conditions of the general form

$$\begin{aligned} s_1 y'_x + k_1 y = 0 \quad &\text{at} \quad x = x_1, \\ s_2 y'_x + k_2 y = 0 \quad &\text{at} \quad x = x_2. \end{aligned} \tag{M11.2.4.2}$$

It is assumed that the functions p, p'_x, ρ, and q are continuous, and p and ρ are positive on an interval $x_1 \le x \le x_2$. It is also assumed that $|s_1| + |k_1| > 0$ and $|s_2| + |k_2| > 0$.

The *Sturm–Liouville problem*: Find the values λ_n of the parameter λ at which problem (M11.2.4.1)–(M11.2.4.2) has a nontrivial solution. Such λ_n are called *eigenvalues* and the corresponding solutions $y_n = y_n(x)$ are called *eigenfunctions* of the Sturm–Liouville problem (M11.2.4.1)–(M11.2.4.2).

► General properties of the Sturm–Liouville problem (M11.2.4.1)–(M11.2.4.2).

1°. There are infinitely (countably) many eigenvalues. All eigenvalues can be ordered so that $\lambda_1 < \lambda_2 < \lambda_3 < \cdots$. Moreover, $\lambda_n \to \infty$ as $n \to \infty$; hence, there can only be a finite number of negative eigenvalues.

2°. The eigenfunctions are defined up to a constant factor. Each eigenfunction $y_n(x)$ has precisely $n - 1$ zeros on the open interval (x_1, x_2).

3°. Any two eigenfunctions $y_n(x)$ and $y_m(x)$, $n \neq m$, are orthogonal with weight $\rho(x)$ on the interval $x_1 \le x \le x_2$:

$$\int_{x_1}^{x_2} \rho(x) y_n(x) y_m(x)\, dx = 0 \quad \text{if} \quad n \neq m.$$

4°. An arbitrary function $F(x)$ that has a continuous derivative and satisfies the boundary conditions of the Sturm–Liouville problem can be decomposed into an absolutely and uniformly convergent series in the eigenfunctions

$$F(x) = \sum_{n=1}^{\infty} F_n y_n(x),$$

where the Fourier coefficients F_n of $F(x)$ are calculated by

$$F_n = \frac{1}{\|y_n\|^2} \int_{x_1}^{x_2} \rho(x) F(x) y_n(x)\, dx, \qquad \|y_n\|^2 = \int_{x_1}^{x_2} \rho(x) y_n^2(x)\, dx.$$

5°. If the conditions

$$q(x) \ge 0, \quad s_1 k_1 \le 0, \quad s_2 k_2 \ge 0 \tag{M11.2.4.3}$$

hold true, there are no negative eigenvalues. If $q \equiv 0$ and $k_1 = k_2 = 0$, the least eigenvalue is $\lambda_1 = 0$, to which there corresponds an eigenfunction $y_1 = \text{const}$. In the other cases where conditions (M11.2.4.3) are satisfied, all eigenvalues are positive.

6°. The following asymptotic formula is valid for eigenvalues as $n \to \infty$:

$$\lambda_n = \frac{\pi^2 n^2}{\Delta^2} + O(1), \qquad \Delta = \int_{x_1}^{x_2} \sqrt{\frac{\rho(x)}{p(x)}}\, dx. \tag{M11.2.4.4}$$

Remark 1. Equation (M11.2.4.1) can be reduced to the case where $p(x) \equiv 1$ and $\rho(x) \equiv 1$ by the change of variables

$$\zeta = \int \sqrt{\frac{\rho(x)}{p(x)}}\, dx, \qquad u(\zeta) = \bigl[p(x)\rho(x)\bigr]^{1/4} y(x).$$

In this case, the boundary conditions are transformed to boundary conditions of a similar form.

TABLE M11.3
Example estimates of the first eigenvalue λ_1 in Sturm–Liouville problems with boundary conditions of the first kind $y(0) = y(1) = 0$ obtained using the Rayleigh–Ritz principle [the right-hand side of relation (M11.2.4.6)].

Equation	Test function	λ_1, approximate	λ_1, exact
$y''_{xx} + \lambda(1 + x^2)^{-2}y = 0$	$z = \sin \pi x$	15.337	15.0
$y''_{xx} + \lambda(4 - x^2)^{-2}y = 0$	$z = \sin \pi x$	135.317	134.837
$y''_{xx} + \lambda(1 + \sin \pi x)y = 0$	$z = \sin \pi x$ $z = x(1 - x)$	$0.54105\,\pi^2$ $0.55204\,\pi^2$	$0.54032\,\pi^2$ $0.54032\,\pi^2$
$(\sqrt{1 + x}\, y'_x)'_x + \lambda y = 0$	$z = \sin \pi x$	11.9956	11.8985

Remark 2. The second-order linear equation

$$\varphi_2(x)y''_{xx} + \varphi_1(x)y'_x + [\lambda + \varphi_0(x)]y = 0$$

can be represented in the form of equation (M11.2.4.1) where $p(x)$, $\rho(x)$, and $q(x)$ are given by

$$p(x) = \exp\left[\int \frac{\varphi_1(x)}{\varphi_2(x)}\,dx\right],\quad \rho(x) = \frac{1}{\varphi_2(x)}\exp\left[\int \frac{\varphi_1(x)}{\varphi_2(x)}\,dx\right],\quad q(x) = -\frac{\varphi_0(x)}{\varphi_2(x)}\exp\left[\int \frac{\varphi_1(x)}{\varphi_2(x)}\,dx\right].$$

► **Problems with boundary conditions of the first kind.** Let us note some special properties of the Sturm–Liouville problem that is the first boundary value problem for equation (M11.2.4.1) with the boundary conditions

$$y = 0 \quad \text{at} \quad x = x_1, \qquad y = 0 \quad \text{at} \quad x = x_2. \tag{M11.2.4.5}$$

1°. For $n \to \infty$, the asymptotic relation (M11.2.4.4) can be used to estimate the eigenvalues λ_n. In this case, the asymptotic formula

$$\frac{y_n(x)}{\|y_n\|} = \left[\frac{4}{\Delta^2 p(x)\rho(x)}\right]^{1/4} \sin\left[\frac{\pi n}{\Delta}\int_{x_1}^{x}\sqrt{\frac{\rho(x)}{p(x)}}\,dx\right] + O\Big(\frac{1}{n}\Big), \qquad \Delta = \int_{x_1}^{x_2}\sqrt{\frac{\rho(x)}{p(x)}}\,dx$$

holds true for the eigenfunctions $y_n(x)$.

2°. If $q \ge 0$, the following upper estimate holds for the least eigenvalue (*Rayleigh–Ritz principle*):

$$\lambda_1 \le \frac{\displaystyle\int_{x_1}^{x_2}\big[p(x)(z'_x)^2 + q(x)z^2\big]\,dx}{\displaystyle\int_{x_1}^{x_2}\rho(x)z^2\,dx}, \tag{M11.2.4.6}$$

where $z = z(x)$ is any twice differentiable function that satisfies the conditions $z(x_1) = z(x_2) = 0$. The equality in (M11.2.4.6) is attained if $z = y_1(x)$, where $y_1(x)$ is the eigenfunction corresponding to the eigenvalue λ_1. One can take $z = (x - x_1)(x_2 - x)$ or $z = \sin\left[\dfrac{\pi(x - x_1)}{x_2 - x_1}\right]$ in (M11.2.4.6) to obtain specific estimates.

It is significant to note that the left-hand side of (M11.2.4.6) usually gives a fairly precise estimate of the first eigenvalue (see Table M11.3).

3°. The extension of the interval $[x_1, x_2]$ results in decreasing the eigenvalues.

4°. Let the inequalities

$$0 < p_{\min} \le p(x) \le p_{\max}, \quad 0 < \rho_{\min} \le \rho(x) \le \rho_{\max}, \quad 0 < q_{\min} \le q(x) \le q_{\max}$$

be satisfied. Then the following bilateral estimates hold:

$$\frac{p_{\min}}{\rho_{\max}} \frac{\pi^2 n^2}{(x_2 - x_1)^2} + \frac{q_{\min}}{\rho_{\max}} \le \lambda_n \le \frac{p_{\max}}{\rho_{\min}} \frac{\pi^2 n^2}{(x_2 - x_1)^2} + \frac{q_{\max}}{\rho_{\min}}.$$

5°. In engineering calculations for eigenvalues, the approximate formula

$$\lambda_n = \frac{\pi^2 n^2}{\Delta^2} + \frac{1}{x_2 - x_1} \int_{x_1}^{x_2} \frac{q(x)}{\rho(x)}\,dx, \qquad \Delta = \int_{x_1}^{x_2} \sqrt{\frac{\rho(x)}{p(x)}}\,dx \tag{M11.2.4.7}$$

may be quite useful. This formula provides an exact result if $p(x)\rho(x) = \text{const}$ and $q(x)/\rho(x) = \text{const}$ (in particular, for constant equation coefficients, $p = p_0$, $q = q_0$, and $\rho = \rho_0$) and gives a correct asymptotic behavior of (M11.2.4.4) for any $p(x)$, $q(x)$, and $\rho(x)$. In addition, relation (M11.2.4.7) gives two correct leading asymptotic terms as $n \to \infty$ if $p(x) = \text{const}$ and $\rho(x) = \text{const}$ [and also if $p(x)\rho(x) = \text{const}$].

6°. Suppose that $p(x) = \rho(x) = 1$ and the function $q(x)$ has a continuous derivative. The following asymptotic relations hold for eigenvalues λ_n and eigenfunctions $y_n(x)$ as $n \to \infty$:

$$\begin{aligned} \sqrt{\lambda_n} &= \frac{\pi n}{x_2 - x_1} + \frac{1}{\pi n} Q(x_1, x_2) + O\Big(\frac{1}{n^2}\Big), \\ y_n(x) &= \sin \frac{\pi n (x - x_1)}{x_2 - x_1} - \frac{1}{\pi n} \Big[(x_1 - x) Q(x, x_2) + (x_2 - x) Q(x_1, x)\Big] \cos \frac{\pi n (x - x_1)}{x_2 - x_1} + O\Big(\frac{1}{n^2}\Big), \end{aligned}$$

where

$$Q(u, v) = \frac{1}{2} \int_u^v q(x)\,dx. \tag{M11.2.4.8}$$

7°. Let us consider the eigenvalue problem for the equation with a small parameter

$$y''_{xx} + [\lambda + \varepsilon q(x)] y = 0 \qquad (\varepsilon \to 0)$$

subject to the boundary conditions (M11.2.4.5) with $x_1 = 0$ and $x_2 = 1$. We assume that $q(x) = q(-x)$.

This problem has the following eigenvalues and eigenfunctions:

$$\begin{aligned} \lambda_n &= \pi^2 n^2 - \varepsilon A_{nn} + \frac{\varepsilon^2}{\pi^2} \sum_{k \ne n} \frac{A_{nk}^2}{n^2 - k^2} + O(\varepsilon^3), \qquad A_{nk} = 2 \int_0^1 q(x) \sin(\pi n x) \sin(\pi k x)\,dx; \\ y_n(x) &= \sqrt{2}\, \sin(\pi n x) - \varepsilon \frac{\sqrt{2}}{\pi^2} \sum_{k \ne n} \frac{A_{nk}}{n^2 - k^2} \sin(\pi k x) + O(\varepsilon^2). \end{aligned}$$

Here, the summation is carried out over k from 1 to ∞. The next term in the expansion of y_n can be found in Nayfeh (1973).

M11.3. Second-Order Nonlinear Differential Equations

M11.3.1. Form of the General Solution. Cauchy Problem

► **Equations solved for the derivative. General solution.** A *second-order ordinary differential equation* solved for the highest derivative has the form

$$y''_{xx} = f(x, y, y'_x). \qquad \text{(M11.3.1.1)}$$

The general solution of this equation depends on two arbitrary constants, C_1 and C_2. In some cases, the general solution can be written in explicit form, $y = \varphi(x, C_1, C_2)$, but more often implicit or parametric forms of the general solution are encountered.

► **Cauchy problem. The existence and uniqueness theorem.** *Cauchy problem*: Find a solution of equation (M11.3.1.1) satisfying the *initial conditions*

$$y(x_0) = y_0, \qquad y'_x(x_0) = y_1. \qquad \text{(M11.3.1.2)}$$

(At a point $x = x_0$, the value of the unknown function, y_0, and its derivative, y_1, are prescribed.)

EXISTENCE AND UNIQUENESS THEOREM. *Let $f(x, y, z)$ be a continuous function in all its arguments in a neighborhood of a point (x_0, y_0, y_1) and let f have bounded partial derivatives f_y and f_z in this neighborhood, or the Lipschitz condition is satisfied: $|f(x, y, z) - f(x, \bar{y}, \bar{z})| \le A\bigl(|y-\bar{y}| + |z-\bar{z}|\bigr)$, where A is some positive number. Then a solution of equation (M11.3.1.1) satisfying the initial conditions (M11.3.1.2) exists and is unique.*

M11.3.2. Equations Admitting Reduction of Order

► **Equations not containing y explicitly.** In the general case, a second-order equation that does not contain y explicitly has the form

$$F(x, y'_x, y''_{xx}) = 0. \qquad \text{(M11.3.2.1)}$$

Such equations remain unchanged under an arbitrary translation of the dependent variable: $y \to y + \text{const}$. The substitution $y'_x = z(x)$, $y''_{xx} = z'_x(x)$ brings (M11.3.2.1) to a first-order equation: $F(x, z, z'_x) = 0$.

► **Equations not containing x explicitly (autonomous equations).** In the general case, a second-order equation that does not contain x explicitly has the form

$$F(y, y'_x, y''_{xx}) = 0. \qquad \text{(M11.3.2.2)}$$

Such equations remain unchanged under an arbitrary translation of the independent variable: $x \to x + \text{const}$. Using the substitution $y'_x = w(y)$, where y plays the role of the independent variable, and taking into account the relations $y''_{xx} = w'_x = w'_y y'_x = w'_y w$, one can reduce (M11.3.2.2) to a first-order equation: $F(y, w, ww'_y) = 0$.

Example 1. Consider the autonomous equation

$$y''_{xx} = f(y),$$

which often arises in the theory of heat and mass transfer and combustion. The change of variable $y'_x = w(y)$ leads to a separable first-order equation: $ww'_y = f(y)$. Integrating yields $w^2 = 2F(y) + C_1$, where $F(y) = \int f(y)\,dy$. Solving for w and returning to the original variable, we obtain the separable equation $y'_x = \pm\sqrt{2F(y) + C_1}$. Its general solution is expressed as

$$\int \frac{dy}{\sqrt{2F(y) + C_1}} = \pm x + C_2, \quad \text{where} \quad F(y) = \int f(y)\,dy.$$

Remark. The equation $y''_{xx} = f(y + ax^2 + bx + c)$ is reduced by the change of variable $u = y + ax^2 + bx + c$ to an autonomous equation, $u''_{xx} = f(u) + 2a$.

► **Equations of the form $F(ax+by, y'_x, y''_{xx}) = 0$.** Such equations are invariant under simultaneous translations of the independent and dependent variables in accordance with the rule $x \to x + bc$, $y \to y - ac$, where c is an arbitrary constant.

For $b = 0$, see equation (M11.3.2.1). For $b \neq 0$, the substitution $bw = ax + by$ leads to the autonomous equation, $F(bw, w'_x - a/b, w''_{xx}) = 0$, that does not contain x explicitly.

► **Homogeneous equations.**

1°. The *equations homogeneous in the independent variable* remain unchanged under scaling of the independent variable, $x \to \alpha x$, where α is an arbitrary nonzero number. In the general case, such equations can be written in the form

$$F(y, xy'_x, x^2 y''_{xx}) = 0. \tag{M11.3.2.3}$$

The substitution $z(y) = xy'_x$ leads to a first-order equation, $F(y, z, zz'_y - z) = 0$.

2°. The *equations homogeneous in the dependent variable* remain unchanged under scaling of the variable sought, $y \to \alpha y$, where α is an arbitrary nonzero number. In the general case, such equations can be written in the form

$$F(x, y'_x/y, y''_{xx}/y) = 0. \tag{M11.3.2.4}$$

The substitution $z(x) = y'_x/y$ leads to a first-order equation, $F(x, z, z'_x + z^2) = 0$.

3°. The *equations homogeneous in both variables* are invariant under simultaneous scaling (dilatation) of the independent and dependent variables, $x \to \alpha x$ and $y \to \alpha y$, where α is an arbitrary nonzero number. In the general case, such equations can be written in the form

$$F(y/x, y'_x, xy''_{xx}) = 0. \tag{M11.3.2.5}$$

The transformation $t = \ln|x|$, $w = y/x$ leads to the autonomous equation, $F(w, w'_t + w, w''_{tt} + w'_t) = 0$, that does not contain t explicitly.

Example 2. The homogeneous equation

$$xy''_{xx} - y'_x = f(y/x)$$

is reduced by the transformation $t = \ln|x|$, $w = y/x$ to the autonomous form: $w''_{tt} = f(w) + w$. For solution of this equation, see Example 1 above (the notation of the right-hand side has to be changed there).

► **Generalized homogeneous equations.**

1°. The *generalized homogeneous equations* remain unchanged under simultaneous scaling of the independent and dependent variables in accordance with the rule $x \to \alpha x$ and $y \to \alpha^k y$, where α is an arbitrary nonzero number and k is some number. Such equations can be written in the form

$$F(x^{-k} y, x^{1-k} y'_x, x^{2-k} y''_{xx}) = 0. \tag{M11.3.2.6}$$

The transformation $t = \ln x$, $w = x^{-k} y$ leads to the autonomous equation,

$$F\big(w, w'_t + kw, w''_{tt} + (2k-1)w'_t + k(k-1)w\big) = 0,$$

that does not contain t explicitly.

2°. The most general form of representation of generalized homogeneous equations is as follows:

$$\mathcal{F}(x^n y^m, xy'_x/y, x^2 y''_{xx}/y) = 0. \qquad \text{(M11.3.2.7)}$$

The transformation $z = x^n y^m$, $u = xy'_x/y$ reduces this equation to the first-order equation

$$\mathcal{F}\big(z, u, z(mu+n)u'_z - u + u^2\big) = 0.$$

Remark. For $m \neq 0$, equation (M11.3.2.7) is equivalent to equation (M11.3.2.6) in which $k = -n/m$. To the particular values $n = 0$ and $m = 0$ there correspond equations (M11.3.2.3) and (M11.3.2.4) homogeneous in the independent and dependent variables, respectively. For $n = -m \neq 0$, we have an equation homogeneous in both variables, which is equivalent to equation (M11.3.2.5).

► **Equations invariant under scaling–translation transformations.**

1°. The equations of the form

$$F(e^{\lambda x} y, e^{\lambda x} y'_x, e^{\lambda x} y''_{xx}) = 0 \qquad \text{(M11.3.2.8)}$$

remain unchanged under simultaneous translation and scaling of variables, $x \to x + \alpha$ and $y \to \beta y$, where $\beta = e^{-\alpha\lambda}$ and α is an arbitrary number. The substitution $w = e^{\lambda x} y$ brings (M11.3.2.8) to the autonomous equation, $F(w,\ w'_x - \lambda w,\ w''_{xx} - 2\lambda w'_x + \lambda^2 w) = 0$, that does not contain x explicitly.

2°. The equation

$$F(e^{\lambda x} y^n, y'_x/y, y''_{xx}/y) = 0 \qquad \text{(M11.3.2.9)}$$

is invariant under the simultaneous translation and scaling of variables, $x \to x + \alpha$ and $y \to \beta y$, where $\beta = e^{-\alpha\lambda/n}$ and α is an arbitrary number. The transformation $z = e^{\lambda x} y^n$, $w = y'_x/y$ brings (M11.3.2.9) to a first-order equation: $F\big(z, w, z(nw + \lambda)w'_z + w^2\big) = 0$.

3°. The equation

$$F(x^n e^{\lambda y}, xy'_x, x^2 y''_{xx}) = 0 \qquad \text{(M11.3.2.10)}$$

is invariant under the simultaneous scaling and translation of variables, $x \to \alpha x$ and $y \to y + \beta$, where $\alpha = e^{-\beta\lambda/n}$ and β is an arbitrary number. The transformation $z = x^n e^{\lambda y}$, $w = xy'_x$ brings (M11.3.2.10) to a first-order equation: $F\big(z, w, z(\lambda w + n)w'_z - w\big) = 0$.

► **Equations of the form $\boldsymbol{F(x, xy'_x - y, y''_{xx}) = 0}$.** The substitution $w(x) = xy'_x - y$ leads to a first-order equation: $F(x, w, w'_x/x) = 0$.

M11.3.3. Methods of Regular Series Expansions with Respect to the Independent Variable

A solution of the Cauchy problem

$$y''_{xx} = f(x, y, y'_x), \qquad \text{(M11.3.3.1)}$$

$$y(x_0) = y_0, \quad y'_x(x_0) = y_1 \qquad \text{(M11.3.3.2)}$$

can be sought in the form of a Taylor series in powers of the difference $(x - x_0)$, specifically:

$$y(x) = y(x_0) + y'_x(x_0)(x - x_0) + \frac{y''_{xx}(x_0)}{2!}(x - x_0)^2 + \frac{y'''_{xxx}(x_0)}{3!}(x - x_0)^3 + \cdots. \qquad \text{(M11.3.3.3)}$$

The first two coefficients $y(x_0)$ and $y'_x(x_0)$ in solution (M11.3.3.3) are defined by the initial conditions (M11.3.3.2). The values of the subsequent derivatives of y at the point $x = x_0$ are determined from equation (M11.3.3.1) and its derivative equations (obtained by successive differentiation of the equation) taking into account the initial conditions (M11.3.3.2). In particular, setting $x = x_0$ in (M11.3.3.1) and substituting (M11.3.3.2), we obtain the value of the second derivative:

$$y''_{xx}(x_0) = f(x_0, y_0, y_1). \qquad \text{(M11.3.3.4)}$$

Further, differentiating (M11.3.3.1) yields

$$y'''_{xxx} = f_x(x, y, y'_x) + f_y(x, y, y'_x)y'_x + f_{y'_x}(x, y, y'_x)y''_{xx}. \qquad \text{(M11.3.3.5)}$$

On substituting $x = x_0$, the initial conditions (M11.3.3.2), and the expression of $y''_{xx}(x_0)$ of (M11.3.3.4) into the right-hand side of equation (M11.3.3.5), we calculate the value of the third derivative:

$$y'''_{xxx}(x_0) = f_x(x_0, y_0, y_1) + y_1 f_y(x_0, y_0, y_1) + f(x_0, y_0, y_1) f_{y'_x}(x_0, y_0, y_1).$$

The subsequent derivatives of the unknown are determined likewise.

The thus obtained solution (M11.3.3.3) can only be used in a small neighborhood of the point $x = x_0$.

Example. Consider the following Cauchy problem for a second-order nonlinear equation:

$$y''_{xx} = yy'_x + y^3; \qquad \text{(M11.3.3.6)}$$

$$y(0) = y'_x(0) = 1. \qquad \text{(M11.3.3.7)}$$

Substituting the initial values of the unknown and its derivative (M11.3.3.7) into equation (M11.3.3.6) yields the initial value of the second derivative:

$$y''_{xx}(0) = 2. \qquad \text{(M11.3.3.8)}$$

Differentiating equation (M11.3.3.6) gives

$$y'''_{xxx} = yy''_{xx} + (y'_x)^2 + 3y^2 y'_x. \qquad \text{(M11.3.3.9)}$$

Substituting here the initial values from (M11.3.3.7) and (M11.3.3.8), we obtain the initial condition for the third derivative:

$$y'''_{xxx}(0) = 6. \qquad \text{(M11.3.3.10)}$$

Differentiating (M11.3.3.9) followed by substituting (M11.3.3.7), (M11.3.3.8), and (M11.3.3.10), we find that

$$y''''_{xxxx}(0) = 24. \qquad \text{(M11.3.3.11)}$$

On substituting the initial data (M11.3.3.7), (M11.3.3.8), (M11.3.3.10), and (M11.3.3.11) into (M11.3.3.3), we arrive at the Taylor series expansion of the solution about $x = 0$:

$$y = 1 + x + x^2 + x^3 + x^4 + \cdots . \qquad \text{(M11.3.3.12)}$$

This geometric series is convergent only for $|x| < 1$. In this case, summing up the series (M11.3.3.12) gives the exact solution of the Cauchy problem (M11.3.3.6)–(M11.3.3.7) of the form $y(x) = \dfrac{1}{1-x}$.

M11.3.4. Perturbation Methods in Problems with a Small Parameter

► **Preliminary remarks.** Perturbation methods are widely used in nonlinear mechanics and theoretical physics for solving problems that are described by differential equations with a small parameter ε. The primary purpose of these methods is to obtain an approximate solution that would be equally suitable at all (small, intermediate, and large) values of the independent variable as $\varepsilon \to 0$.

It is further assumed that the order of the equation remains unchanged at $\varepsilon = 0$.

In many problems of nonlinear mechanics and theoretical physics, the independent variable is dimensionless time t. Therefore, in this subsection we use the conventional t ($0 \le t < \infty$), instead of x.

► **Method of regular (direct) expansion in powers of the small parameter.** We consider an equation of general form with a parameter ε:

$$y''_{tt} + f(t, y, y'_t, \varepsilon) = 0. \tag{M11.3.4.1}$$

We assume that the function f can be represented as a series in powers of ε:

$$f(t, y, y'_t, \varepsilon) = \sum_{n=0}^{\infty} \varepsilon^n f_n(t, y, y'_t). \tag{M11.3.4.2}$$

Solutions of the Cauchy problem and various boundary value problems for equation (M11.3.4.1) with $\varepsilon \to 0$ are sought in the form of a power series expansion:

$$y = \sum_{n=0}^{\infty} \varepsilon^n y_n(t). \tag{M11.3.4.3}$$

One should substitute expression (M11.3.4.3) into equation (M11.3.4.1) taking into account (M11.3.4.2). Then the functions f_n are expanded into a power series in the small parameter and the coefficients of like powers of ε are collected and equated to zero to obtain a system of equations for y_n:

$$y''_0 + f_0(t, y_0, y'_0) = 0, \tag{M11.3.4.4}$$

$$y''_1 + F(t, y_0, y'_0)y'_1 + G(t, y_0, y'_0)y_1 + f_1(t, y_0, y'_0) = 0, \quad F = \frac{\partial f_0}{\partial y'}, \quad G = \frac{\partial f_0}{\partial y}. \tag{M11.3.4.5}$$

Only the first two equations are written out here. The prime denotes differentiation with respect to t. To obtain the initial (or boundary) conditions for y_n, the expansion (M11.3.4.3) should be taken into account.

The success in the application of this method is primarily determined by the possibility of constructing a solution of equation (M11.3.4.4) for the leading term y_0. It is significant to note that the other terms y_n with $n \geq 1$ are governed by linear equations with homogeneous initial conditions.

Example 1. The Duffing equation

$$y''_{tt} + y + \varepsilon y^3 = 0 \tag{M11.3.4.6}$$

with initial conditions

$$y(0) = a, \quad y'_t(0) = 0$$

describes the motion of a cubic oscillator, i.e., oscillations of a point mass on a nonlinear spring. Here, y is the deviation of the point mass from the equilibrium and t is dimensionless time.

For $\varepsilon \to 0$, an approximate solution of the problem is sought in the form of the asymptotic expansion (M11.3.4.3). We substitute (M11.3.4.3) into equation (M11.3.4.6) and initial conditions and expand in powers of ε. On equating the coefficients of like powers of the small parameter to zero, we obtain the following problems for y_0 and y_1:

$$y''_0 + y_0 = 0, \qquad y_0(0) = a, \quad y'_0(0) = 0;$$

$$y''_1 + y_1 = -y_0^3, \qquad y_1(0) = 0, \quad y'_1(0) = 0.$$

The solution of the problem for y_0 is given by

$$y_0 = a\cos t.$$

Substituting this expression into the equation for y_1 and taking into account the identity $\cos^3 t = \frac{1}{4}\cos 3t + \frac{3}{4}\cos t$, we obtain

$$y''_1 + y_1 = -\tfrac{1}{4}a^3(\cos 3t + 3\cos t), \qquad y_1(0) = 0, \quad y'_1(0) = 0.$$

Integrating yields

$$y_1 = -\tfrac{3}{8}a^3 t\sin t + \tfrac{1}{32}a^3(\cos 3t - 3\cos t).$$

Thus the two-term solution of the original problem is given by

$$y = a\cos t + \varepsilon a^3\bigl[-\tfrac{3}{8}t\sin t + \tfrac{1}{32}(\cos 3t - 3\cos t)\bigr] + O(\varepsilon^2).$$

Remark 1. The term $t\sin t$ causes $y_1/y_0 \to \infty$ as $t \to \infty$. For this reason, the solution obtained is unsuitable at large times. It can only be used for $\varepsilon t \ll 1$; this results from the condition of applicability of the expansion, $y_0 \gg \varepsilon y_1$.

This circumstance is typical of the method of regular series expansions with respect to the small parameter; in other words, the expansion becomes unsuitable at large values of the independent variable. Methods that allow avoiding this difficulty are discussed below.

Remark 2. Growing terms as $t \to \infty$, like $t\sin t$, that narrow down the domain of applicability of asymptotic expansions are called *secular*.

► **Method of scaled parameters (Lindstedt–Poincaré method).** This method is usually used for finding periodic solutions to equations of the form

$$y''_{tt} + \omega_0 y = \varepsilon f(y, y'_t), \qquad \text{(M11.3.4.7)}$$

where $\varepsilon \ll 1$.

A solutions is sought in the form

$$t = z\Bigl(1 + \sum_{k=0}^{n}\varepsilon^k\omega_k\Bigr), \qquad y(t) = \sum_{k=0}^{n}\varepsilon^k y_k(z). \qquad \text{(M11.3.4.8)}$$

The constants ω_k and functions $y_k(z)$ are determined; it is assumed that $y_{k+1}/y_k = O(1)$. By choosing appropriate ω_k, one removes the secular terms from the solution.

Example 2. Consider the Duffing equation (M11.3.4.6) once again. Following (M11.3.4.8), one performs the change of variable

$$t = z(1 + \varepsilon\omega_1 + \cdots)$$

to obtain

$$y''_{zz} + (1 + \varepsilon\omega_1 + \cdots)^2(y + \varepsilon y^3) = 0. \qquad \text{(M11.3.4.9)}$$

The solution is sought in the series form (M11.3.4.8), $y = y_0(z) + \varepsilon y_1(z) + \cdots$. Substituting it into equation (M11.3.4.9) and matching the coefficients of like powers of ε, one arrives at the following system of equations for two leading terms of the series:

$$y_0'' + y_0 = 0, \qquad \text{(M11.3.4.10)}$$

$$y_1'' + y_1 = -y_0^3 - 2\omega_1 y_0, \qquad \text{(M11.3.4.11)}$$

where the prime denotes differentiation with respect to z.

The general solution of equation (M11.3.4.10) is given by

$$y_0 = a\cos(z + b), \qquad \text{(M11.3.4.12)}$$

where a and b are constants of integration. Taking into account (M11.3.4.12) and rearranging terms, we reduce equation (M11.3.4.11) to

$$y_1'' + y_1 = -\tfrac{1}{4}a^3\cos\bigl[3(z + b)\bigr] - 2a\bigl(\tfrac{3}{8}a^2 + \omega_1\bigr)\cos(z + b). \qquad \text{(M11.3.4.13)}$$

For $\omega_1 \neq -\frac{3}{8}a^2$, the particular solution of equation (M11.3.4.13) contains a secular term proportional to $z\cos(z + b)$. In this case, the condition of applicability of the expansion, $y_1/y_0 = O(1)$, cannot be satisfied at sufficiently large z. For this condition to be met, one should set

$$\omega_1 = -\tfrac{3}{8}a^2. \qquad \text{(M11.3.4.14)}$$

In this case, the solution of equation (M11.3.4.13) is given by

$$y_1 = \tfrac{1}{32}a^3\cos\bigl[3(z + b)\bigr]. \qquad \text{(M11.3.4.15)}$$

Subsequent terms of the expansion can be found likewise.

With (M11.3.4.12), (M11.3.4.14), and (M11.3.4.15), we obtain a solution of the Duffing equation in the form

$$y = a\cos(\omega t + b) + \tfrac{1}{32}\varepsilon a^3\cos\bigl[3(\omega t + b)\bigr] + O(\varepsilon^2),$$

$$\omega = \bigl[1 - \tfrac{3}{8}\varepsilon a^2 + O(\varepsilon^2)\bigr]^{-1} = 1 + \tfrac{3}{8}\varepsilon a^2 + O(\varepsilon^2).$$

► **Averaging method (Van der Pol–Krylov–Bogolyubov scheme).** The averaging method involves two stages. First, the second-order nonlinear equation (M11.3.4.7) is reduced with the transformation

$$y = a\cos\varphi, \quad y'_t = -\omega_0 a\sin\varphi, \qquad \text{where} \quad a = a(t), \quad \varphi = \varphi(t),$$

to an equivalent system of two first-order differential equations:

$$\begin{aligned} a'_t &= -\frac{\varepsilon}{\omega_0} f(a\cos\varphi, -\omega_0 a\sin\varphi)\sin\varphi, \\ \varphi'_t &= \omega_0 - \frac{\varepsilon}{\omega_0 a} f(a\cos\varphi, -\omega_0 a\sin\varphi)\cos\varphi. \end{aligned} \tag{M11.3.4.16}$$

The right-hand sides of equations (M11.3.4.16) are periodic in φ, with the amplitude a being a slow-varying function of time t. The amplitude and the oscillation character are changing little during the time the phase φ changes by 2π.

At the second stage, the right-hand sides of equations (M11.3.4.16) are being averaged with respect to φ. This procedure results in an approximate system of equations:

$$\begin{aligned} a'_t &= -\frac{\varepsilon}{\omega_0} f_{\rm s}(a), \\ \varphi'_t &= \omega_0 - \frac{\varepsilon}{\omega_0 a} f_{\rm c}(a), \end{aligned} \tag{M11.3.4.17}$$

where

$$\begin{aligned} f_{\rm s}(a) &= \frac{1}{2\pi}\int_0^{2\pi} \sin\varphi\, f(a\cos\varphi, -\omega_0 a\sin\varphi)\,d\varphi, \\ f_{\rm c}(a) &= \frac{1}{2\pi}\int_0^{2\pi} \cos\varphi\, f(a\cos\varphi, -\omega_0 a\sin\varphi)\,d\varphi. \end{aligned}$$

System (M11.3.4.17) is substantially simpler than the original system (M11.3.4.16)—the first equation in (M11.3.4.17), for the oscillation amplitude a, is a separable equation and, hence, can readily be integrated; then the second equation in (M11.3.4.17) can also be integrated.

M11.3.5. Galerkin Method and Its Modifications (Projection Methods)

► **General form of an approximate solution.** Consider a boundary value problem for the equation

$$\mathfrak{F}[y] - f(x) = 0 \tag{M11.3.5.1}$$

with linear homogeneous boundary conditions* at the points $x = x_1$ and $x = x_2$ ($x_1 \le x \le x_2$). Here, $\mathfrak{F}$ is a linear or nonlinear differential operator of the second order (or a higher order operator); $y = y(x)$ is the unknown function and $f = f(x)$ is a given function. It is assumed that $\mathfrak{F}[0] = 0$.

Let us choose a sequence of linearly independent functions (called *basis functions*)

$$\varphi = \varphi_n(x) \qquad (n = 1,\ 2,\ \dots,\ N) \tag{M11.3.5.2}$$

---

* Nonhomogeneous boundary conditions can be reduced to homogeneous ones by the change of variable $z = A_2x^2 + A_1x + A_0 + y$ (the constants A_2, A_1, and A_0 are selected using the method of undetermined coefficients).

satisfying the same boundary conditions as $y = y(x)$. According to all methods that will be considered below, an approximate solution of equation (M11.3.5.1) is sought as a linear combination

$$y_N = \sum_{n=1}^{N} A_n \varphi_n(x), \qquad \text{(M11.3.5.3)}$$

with the unknown coefficients A_n to be found in the process of solving the problem.

The finite sum (M11.3.5.3) is called an *approximation function*. The remainder term R_N obtained after the finite sum has been substituted into the left-hand side of equation (M11.3.5.1),

$$R_N = \mathfrak{F}[y_N] - f(x). \qquad \text{(M11.3.5.4)}$$

If the remainder R_N is identically equal to zero, then the function y_N is the exact solution of equation (M11.3.5.1). In general, $R_N \not\equiv 0$.

► **Galerkin method.** In order to find the coefficients A_n in (M11.3.5.3), consider another sequence of linearly independent functions

$$\psi = \psi_k(x) \qquad (k = 1, 2, \dots, N). \qquad \text{(M11.3.5.5)}$$

Let us multiply both sides of (M11.3.5.4) by ψ_k and integrate the resulting relation over the region $V = \{x_1 \le x \le x_2\}$, in which we seek the solution of equation (M11.3.5.1). Next, we equate the corresponding integrals to zero (for the exact solutions, these integrals are equal to zero). Thus, we obtain the following system of algebraic equations for the unknown coefficients A_n:

$$\int_{x_1}^{x_2} \psi_k R_N \, dx = 0 \qquad (k = 1, 2, \dots, N). \qquad \text{(M11.3.5.6)}$$

Relations (M11.3.5.6) mean that the approximation function (M11.3.5.3) satisfies equation (M11.3.5.1) "on the average" (i.e., in the integral sense) with weights ψ_k. Introducing the scalar product $\langle g, h\rangle = \int_{x_1}^{x_2} gh\,dx$ of arbitrary functions g and h, we can consider equations (M11.3.5.6) as the condition of orthogonality of the remainder R_N to all weight functions ψ_k.

The Galerkin method can be applied not only to boundary value problems, but also to eigenvalue problems (in the latter case, one takes $f = \lambda y$ and seeks eigenfunctions y_n, together with eigenvalues λ_n).

Mathematical justification of the Galerkin method for specific boundary value problems can be found in the literature listed at the end of Chapter M11. Below we describe some other methods that are in fact special cases of the Galerkin method.

Remark. Most often, one takes suitable sequences of polynomials or trigonometric functions as $\varphi_n(x)$ in the approximation function (M11.3.5.3).

► **Bubnov–Galerkin method, the moment method, the least squares method.**

1°. The sequences of functions (M11.3.5.2) and (M11.3.5.5) in the Galerkin method can be chosen arbitrarily. In the case of equal functions,

$$\varphi_k(x) = \psi_k(x) \qquad (k = 1, 2, \dots, N), \qquad \text{(M11.3.5.7)}$$

the method is often called the *Bubnov–Galerkin method*.

2°. The *moment method* is the Galerkin method with the weight functions (M11.3.5.5) being powers of x,

$$\psi_k = x^k. \tag{M11.3.5.8}$$

3°. Sometimes, the functions ψ_k are expressed in terms of φ_k by the relations

$$\psi_k = \mathfrak{F}[\varphi_k] \qquad (k = 1,\, 2,\, \ldots),$$

where $\mathfrak{F}$ is the differential operator of equation (M11.3.5.1). This version of the Galerkin method is called the *least squares method*.

► **Collocation method.** In the collocation method, one chooses a sequence of points x_k, $k = 1, \ldots, N$, and imposes the condition that the remainder (M11.3.5.4) be zero at these points,

$$R_N = 0 \quad \text{at} \quad x = x_k \qquad (k = 1, \ldots, N). \tag{M11.3.5.9}$$

When solving a specific problem, the points x_k, at which the remainder R_N is set equal to zero, are regarded as most significant. The number of collocation points N is taken equal to the number of the terms of the series (M11.3.5.3). This allows one to obtain a complete system of algebraic equations for the unknown coefficients A_n (for linear boundary value problems, this algebraic system is linear).

Note that the collocation method is a special case of the Galerkin method with the sequence (M11.3.5.5) consisting of the Dirac delta functions:

$$\psi_k = \delta(x - x_k).$$

In the collocation method, there is no need to calculate integrals, and this essentially simplifies the procedure of solving nonlinear problems (although usually this method yields less accurate results than other modifications of the Galerkin method).

Example. Consider the boundary value problem for the linear second-order ordinary differential equation with variable coefficients

$$y''_{xx} + g(x)y - f(x) = 0 \tag{M11.3.5.10}$$

subject to the boundary conditions of the first kind

$$y(-1) = y(1) = 0. \tag{M11.3.5.11}$$

Assume that the coefficients of equation (M11.3.5.10) are smooth even functions, so that $f(x) = f(-x)$ and $g(x) = g(-x)$. We use the collocation method for the approximate solution of problem (M11.3.5.10)–(M11.3.5.11).

1°. Take the polynomials

$$y_n(x) = x^{2n-2}(1 - x^2), \qquad n = 1,\, 2,\, \ldots\, N,$$

as the basis functions; they satisfy the boundary conditions (M11.3.5.11), $y_n(\pm 1) = 0$.

Let us consider three collocation points

$$x_1 = -\sigma, \quad x_2 = 0, \quad x_3 = \sigma \qquad (0 < \sigma < 1) \tag{M11.3.5.12}$$

and confine ourselves to two basis functions ($N = 2$), so that the approximation function is taken in the form

$$y(x) = A_1(1 - x^2) + A_2x^2(1 - x^2). \tag{M11.3.5.13}$$

Substituting (M11.3.5.13) into the left-hand side of equation (M11.3.5.10) yields the remainder

$$R(x) = A_1\bigl[-2 + (1 - x^2)g(x)\bigr] + A_2\bigl[2 - 12x^2 + x^2(1 - x^2)g(x)\bigr] - f(x).$$

It must vanish at the collocation points (M11.3.5.12). Taking into account the properties $f(\sigma) = f(-\sigma)$ and $g(\sigma) = g(-\sigma)$, we obtain two linear algebraic equations for the coefficients A_1 and A_2:

$$\begin{gathered} A_1\bigl[-2 + g(0)\bigr] + 2A_2 - f(0) = 0 \qquad (\text{at } x = 0),\\ A_1\bigl[-2 + (1 - \sigma^2)g(\sigma)\bigr] + A_2\bigl[2 - 12\sigma^2 + \sigma^2(1 - \sigma^2)g(\sigma)\bigr] - f(\sigma) = 0 \qquad (\text{at } x = \pm\sigma). \end{gathered} \tag{M11.3.5.14}$$

2°. To be specific, let us take the following functions appearing in equation (M11.3.5.10):

$$f(x) = -1, \quad g(x) = 1 + x^2. \qquad \text{(M11.3.5.15)}$$

On solving the corresponding system of algebraic equations (M11.3.5.14), we find the coefficients

$$A_1 = \frac{\sigma^4 + 11}{\sigma^4 + 2\sigma^2 + 11}, \quad A_2 = -\frac{\sigma^2}{\sigma^4 + 2\sigma^2 + 11}. \qquad \text{(M11.3.5.16)}$$

In Fig. M11.2, the solid line depicts the numerical solution to problem (M11.3.5.10)–(M11.3.5.11), with the functions (M11.3.5.15), obtained by the shooting method (see Subsection M11.3.6). The dashed lines 1 and 2 show the approximate solutions obtained by the collocation method using the formulas (M11.3.5.13), (M11.3.5.16) with $\sigma = \frac{1}{2}$ (equidistant points) and $\sigma = \frac{\sqrt{2}}{2}$ (Chebyshev points*), respectively. It is evident that both cases provide good agreement between the approximate and numerical solutions; the use of Chebyshev points gives a more accurate result.

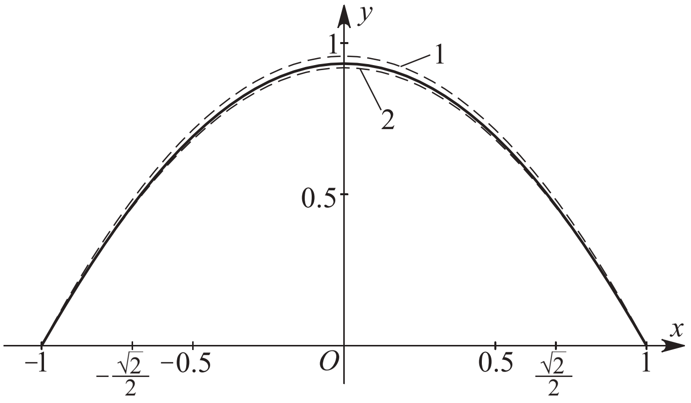

Figure M11.2. Comparison of the numerical solution of problem (M11.3.5.10), (M11.3.5.11), (M11.3.5.15) with the approximate analytical solution (M11.3.5.13), (M11.3.5.16) obtained with the collocation method.

► **Method of partitioning the domain.** The domain $V = \{x_1 \le x \le x_2\}$ is split into N subdomains: $V_k = \{x_{k1} \le x \le x_{k2}\}$, $k = 1, \ldots, N$. In this method, the weight functions are chosen as follows:

$$\psi_k(x) = \begin{cases} 1 & \text{for } x \in V_k, \\ 0 & \text{for } x \notin V_k. \end{cases}$$

The subdomains V_k are chosen according to the specific properties of the problem under consideration and can generally be arbitrary (the union of all subdomains V_k may differ from the domain V, and some V_k and V_m may overlap).

► **Least squared error method.** Sometimes, in order to find the coefficients A_n of the approximation function (M11.3.5.3), one uses the least squared error method based on the minimization of the functional:

$$\Phi = \int_{x_1}^{x_2} R_N^2 \, dx \to \min. \qquad \text{(M11.3.5.17)}$$

For given functions φ_n in (M11.3.5.3), the integral Φ is a function with respect to the coefficients A_n. The corresponding necessary conditions of minimum in (M11.3.5.17) have the form

$$\frac{\partial \Phi}{\partial A_n} = 0 \qquad (n = 1, \ldots, N).$$

This is a system of algebraic equations for the coefficients A_n.

---

* Chebyshev nodes (points) are generally defined by $x_i = \cos\Bigl(\frac{2i-1}{2m}\pi\Bigr)$, $i = 1, \ldots, m$. In this case, $m = 2$.

M11.3.6. Iteration and Numerical Methods

► **Method of successive approximations (Cauchy problem).** The method of successive approximations is implemented in two steps. First, the Cauchy problem

$$y''_{xx} = f(x, y, y'_x) \qquad \text{(equation)}, \qquad \text{(M11.3.6.1)}$$

$$y(x_0) = y_0, \quad y'_x(x_0) = y'_0 \qquad \text{(initial conditions)} \qquad \text{(M11.3.6.2)}$$

is reduced to an equivalent system of integral equations by the introduction of the new variable $u(x) = y'_x$. These integral equations have the form

$$u(x) = y'_0 + \int_{x_0}^{x} f\bigl(t, y(t), u(t)\bigr)\,dt, \qquad y(x) = y_0 + \int_{x_0}^{x} u(t)\,dt. \qquad \text{(M11.3.6.3)}$$

Then the solution of system (M11.3.6.3) is sought by means of successive approximations defined by the following recurrence formulas:

$$u_{n+1}(x) = y'_0 + \int_{x_0}^{x} f\bigl(t, y_n(t), u_n(t)\bigr)\,dt, \qquad y_{n+1}(x) = y_0 + \int_{x_0}^{x} u_n(t)\,dt; \qquad n = 0,\ 1,\ 2,\ \dots$$

As the initial approximation, one can take $y_0(x) = y_0$ and $u_0(x) = y'_0$.

► **Runge–Kutta method (Cauchy problem).** For the numerical integration of the Cauchy problem (M11.3.6.1)–(M11.3.6.2), one often uses the Runge–Kutta method.

Let Δx be sufficiently small. We introduce the following notation:

$$x_k = x_0 + k\Delta x, \quad y_k = y(x_k), \quad y'_k = y'_x(x_k), \quad f_k = f(x_k, y_k, y'_k); \qquad k = 0,\ 1,\ 2,\ \dots$$

The desired values y_k and y'_k are successively found by the formulas

$$y_{k+1} = y_k + y'_k\Delta x + \tfrac{1}{6}(f_1 + f_2 + f_3)(\Delta x)^2,$$

$$y'_{k+1} = y'_k + \tfrac{1}{6}(f_1 + 2f_2 + 2f_3 + f_4)\Delta x,$$

where

$$f_1 = f\bigl(x_k,\ y_k,\ y'_k\bigr),$$

$$f_2 = f\bigl(x_k + \tfrac{1}{2}\Delta x,\ y_k + \tfrac{1}{2}y'_k\Delta x,\ y'_k + \tfrac{1}{2}f_1\Delta x\bigr),$$

$$f_3 = f\bigl(x_k + \tfrac{1}{2}\Delta x,\ y_k + \tfrac{1}{2}y'_k\Delta x + \tfrac{1}{4}f_1(\Delta x)^2,\ y'_k + \tfrac{1}{2}f_2\Delta x\bigr),$$

$$f_4 = f\bigl(x_k + \Delta x,\ y_k + y'_k\Delta x + \tfrac{1}{2}f_2(\Delta x)^2,\ y'_k + f_3\Delta x\bigr).$$

In practice, the step Δx is determined in the same way as for first-order equations (see Remark 2 in Subsection M11.1.7).

► **Shooting method (boundary value problems).** In order to solve the boundary value problem for equation (M11.3.6.1) with the boundary conditions

$$y(x_1) = y_1, \qquad y(x_2) = y_2, \qquad \text{(M11.3.6.4)}$$

one considers an auxiliary Cauchy problem for equation (M11.3.6.1) with the initial conditions

$$y(x_1) = y_1, \qquad y'_x(x_1) = a. \qquad \text{(M11.3.6.5)}$$

(The solution of this Cauchy problem can be obtained by the Runge–Kutta method or some other numerical method.) The parameter a is chosen so that the value of the solution

$y = y(x, a)$ at the point $x = x_2$ coincides with the value required by the second boundary condition in (M11.3.6.4):

$$y(x_2, a) = y_2.$$

First, one finds an a_1 and an a_2 ($a_1 < a_2$) such that

$$[y(x_2, a_1) - y_2][y(x_2, a_2) - y_2] < 0.$$

This implies that the desired a in (M11.3.6.5) belongs to the interval (a_1, a_2). Then a sequence of numbers a_n such that

$$[y(x_2, a_{n-1}) - y_2][y(x_2, a_n) - y_2] < 0$$

is determined numerically, for example, by using the bisection method. The desired a is obtained as: $a = \lim\limits_{n\to\infty} a_n$.

In a similar way one constructs the solution of the boundary value problem with mixed boundary conditions

$$y(x_1) = y_1, \qquad y'_x(x_2) + ky(x_2) = y_2. \tag{M11.3.6.6}$$

In this case, one also considers the auxiliary Cauchy problem (M11.3.6.1), (M11.3.6.5). The parameter a is chosen so that the solution $y = y(x, a)$ satisfies the second boundary condition in (M11.3.6.6) at the point $x = x_2$.

M11.4. Linear Equations of Arbitrary Order and Linear Systems of Equations

M11.4.1. Linear Equations with Constant Coefficients

► **Homogeneous linear equations.** An nth-order homogeneous linear equation with constant coefficients has the general form

$$y^{(n)}_x + a_{n-1} y^{(n-1)}_x + \cdots + a_1 y'_x + a_0 y = 0. \tag{M11.4.1.1}$$

The general solution of this equation is determined by the roots of the characteristic equation

$$P(\lambda) = 0, \qquad \text{where} \quad P(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0. \tag{M11.4.1.2}$$

The following cases are possible:

1°. All roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ of the characteristic equation (M11.4.1.2) are real and distinct. Then the general solution of the homogeneous linear differential equation (M11.4.1.1) has the form

$$y = C_1 \exp(\lambda_1 x) + C_2 \exp(\lambda_2 x) + \cdots + C_n \exp(\lambda_n x).$$

2°. There are m equal real roots $\lambda_1 = \lambda_2 = \cdots = \lambda_m$ ($m \le n$), and the other roots are real and distinct. In this case, the general solution is given by

$$\begin{aligned} y &= \exp(\lambda_1 x)(C_1 + C_2 x + \cdots + C_m x^{m-1}) \\ &\quad + C_{m+1} \exp(\lambda_{m+1} x) + C_{m+2} \exp(\lambda_{m+2} x) + \cdots + C_n \exp(\lambda_n x). \end{aligned}$$

3°. There are m equal complex conjugate roots $\lambda_{1,2} = \alpha \pm i\beta$ ($2m \le n$), and the other roots are real and distinct. In this case, the general solution is

$$\begin{aligned} y &= \exp(\alpha x)\cos(\beta x)(A_1 + A_2 x + \cdots + A_m x^{m-1}) \\ &\quad + \exp(\alpha x)\sin(\beta x)(B_1 + B_2 x + \cdots + B_m x^{m-1}) \\ &\quad + C_{2m+1} \exp(\lambda_{2m+1} x) + C_{2m+2} \exp(\lambda_{2m+2} x) + \cdots + C_n \exp(\lambda_n x), \end{aligned}$$

where $A_1, \ldots, A_m, B_1, \ldots, B_m, C_{2m+1}, \ldots, C_n$ are arbitrary constants.

4°. In the general case, where there are r different roots $\lambda_1, \lambda_2, \ldots, \lambda_r$ of multiplicities $m_1, m_2, \ldots, m_r$, respectively, the left-hand side of the characteristic equation (M11.4.1.2) can be represented as the product

$$P(\lambda) = (\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \ldots (\lambda - \lambda_r)^{m_r},$$

where $m_1 + m_2 + \cdots + m_r = n$. The general solution of the original equation is given by the formula

$$y = \sum_{k=1}^{r} \exp(\lambda_k x)(C_{k,0} + C_{k,1}x + \cdots + C_{k,m_k-1}x^{m_k-1}),$$

where $C_{k,l}$ are arbitrary constants.

If the characteristic equation (M11.4.1.2) has complex conjugate roots $\lambda_s = \alpha_s + i\beta_s$ and $\lambda_{s+1} = \alpha_s - i\beta_s$, then in the above solution, the corresponding functions $\exp(\lambda_s x)$ and $\exp(\lambda_{s+1}x)$ should be replaced with $\exp(\alpha_s x)\cos(\beta_s x)$ and $\exp(\alpha_s x)\sin(\beta_s x)$, respectively, in a similar way to that in Item 3°.

Example 1. Find the general solution of the linear third-order equation

$$y''' + ay'' - y' - ay = 0.$$

Its characteristic equation is $\lambda^3 + a\lambda^2 - \lambda - a = 0$, or, in factorized form,

$$(\lambda + a)(\lambda - 1)(\lambda + 1) = 0.$$

Depending on the value of the parameter a, three cases are possible.

1. Case $a \neq \pm 1$. There are three different roots, $\lambda_1 = -a$, $\lambda_2 = -1$, and $\lambda_3 = 1$. The general solution of the differential equation is expressed as $y = C_1e^{-ax} + C_2e^{-x} + C_3e^{x}$.

2. Case $a = 1$. There is a double root, $\lambda_1 = \lambda_2 = -1$, and a simple root, $\lambda_3 = 1$. The general solution of the differential equation has the form $y = (C_1 + C_2x)e^{-x} + C_3e^{x}$.

3. Case $a = -1$. There is a double root, $\lambda_1 = \lambda_2 = 1$, and a simple root, $\lambda_3 = -1$. The general solution of the differential equation is expressed as $y = (C_1 + C_2x)e^{x} + C_3e^{-x}$.

Example 2. Consider the linear fourth-order equation

$$y''''_{xxxx} - y = 0.$$

Its characteristic equation, $\lambda^4 - 1 = 0$, has four distinct roots, two real and two pure imaginary,

$$\lambda_1 = 1, \quad \lambda_2 = -1, \quad \lambda_3 = i, \quad \lambda_4 = -i.$$

Therefore, the general solution of the equation in question has the form (see Item 3°)

$$y = C_1e^{x} + C_2e^{-x} + C_3\sin x + C_4\cos x.$$

► Nonhomogeneous linear equations. Forms of particular solutions.

1°. An nth-order nonhomogeneous linear equation with constant coefficients has the general form

$$y_x^{(n)} + a_{n-1}y_x^{(n-1)} + \cdots + a_1y'_x + a_0y = f(x). \tag{M11.4.1.3}$$

The general solution of this equation is the sum of the general solution of the corresponding homogeneous equation with $f(x) \equiv 0$ (see equation M11.4.1.1) and any particular solution of the nonhomogeneous equation (M11.4.1.3).

If the roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ of the characteristic equation (M11.4.1.2) are all real and distinct, equation (M11.4.1.3) has the general solution:

$$y = \sum_{\nu=1}^{n} C_\nu e^{\lambda_\nu x} + \sum_{\nu=1}^{n} \frac{e^{\lambda_\nu x}}{P'_\lambda(\lambda_\nu)} \int f(x)e^{-\lambda_\nu x}\,dx. \tag{M11.4.1.4}$$

In the general case, if the characteristic equation (M11.4.1.2) has complex and/or multiple roots, the solution to equation (M11.4.1.3) can be constructed using formula (M11.4.2.5).

TABLE M11.4
Forms of particular solutions to the nonhomogeneous linear equation with constant coefficients $y_x^{(n)} + a_{n-1}y_x^{(n-1)} + \cdots + a_1 y'_x + a_0 y = f(x)$, that correspond to some special forms of the function $f(x)$.

Form of the function $f(x)$	Roots of the characteristic equation $\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = 0$	Form of particular solution
$P_m(x)$	Zero is not a root of the characteristic equation (i.e., $a_0 \neq 0$)	$\widetilde{P}_m(x)$
	Zero is a root of the characteristic equation (multiplicity r)	$x^r \widetilde{P}_m(x)$
$P_m(x)e^{\alpha x}$ (α is a real constant)	α is not a root of the characteristic equation	$\widetilde{P}_m(x)e^{\alpha x}$
	α is a root of the characteristic equation (multiplicity r)	$x^r \widetilde{P}_m(x)e^{\alpha x}$
$P_m(x)\cos\beta x + Q_k(x)\sin\beta x$	$i\beta$ is not a root of the characteristic equation	$\widetilde{P}_\nu(x)\cos\beta x + \widetilde{Q}_\nu(x)\sin\beta x$
	$i\beta$ is a root of the characteristic equation (multiplicity r)	$x^r[\widetilde{P}_\nu(x)\cos\beta x + \widetilde{Q}_\nu(x)\sin\beta x]$
$[P_m(x)\cos\beta x + Q_k(x)\sin\beta x]e^{\alpha x}$	$\alpha + i\beta$ is not a root of the characteristic equation	$[\widetilde{P}_\nu(x)\cos\beta x + \widetilde{Q}_\nu(x)\sin\beta x]e^{\alpha x}$
	$\alpha + i\beta$ is a root of the characteristic equation (multiplicity r)	$x^r[\widetilde{P}_\nu(x)\cos\beta x + \widetilde{Q}_\nu(x)\sin\beta x]e^{\alpha x}$
Notation: P_m and Q_k are polynomials of degrees m and k with given coefficients; $\widetilde{P}_m$, $\widetilde{P}_\nu$, and $\widetilde{Q}_\nu$ are polynomials of degrees m and ν whose coefficients are determined by substituting the particular solution into the basic equation; $\nu = \max\{m, k\}$; and α and β are real numbers, $i^2 = -1$.		

2°. Table M11.4 lists the forms of particular solutions corresponding to some special forms of functions on the right-hand side of the linear nonhomogeneous equation.

3°. Consider the Cauchy problem for equation (M11.4.1.3) subject to the homogeneous initial conditions

$$y(0) = y'_x(0) = \cdots = y_x^{(n-1)}(0) = 0. \qquad \text{(M11.4.1.5)}$$

Let $y(x)$ be the solution of problem (M11.4.1.3), (M11.4.1.5) for arbitrary $f(x)$ and let $u(x)$ be the solution of the auxiliary, simpler problem (M11.4.1.3), (M11.4.1.5) with $f(x) \equiv 1$, so that $u(x) = y(x)|_{f(x)\equiv 1}$. Then the formula

$$y(x) = \int_0^x f(t)u'_x(x-t)\,dt$$

holds. It is called the *Duhamel integral*.

► **Solution of the Cauchy problem using the Laplace transform.** Consider the Cauchy problem for equation (M11.4.1.3) with arbitrary initial conditions

$$y(0) = y_0, \quad y'_x(0) = y_1, \quad \ldots, \quad y_x^{(n-1)}(0) = y_{n-1}, \qquad \text{(M11.4.1.6)}$$

where $y_0, y_1, \ldots, y_{n-1}$ are given constants.

Problem (M11.4.1.3), (M11.4.1.6) can be solved using the Laplace transform based on the formulas (for details, see Section M10.2)

$$\widetilde{y}(p) = \mathfrak{L}\bigl\{y(x)\bigr\}, \quad \widetilde{f}(p) = \mathfrak{L}\bigl\{f(x)\bigr\}, \quad \text{where} \quad \mathfrak{L}\bigl\{f(x)\bigr\} \equiv \int_0^\infty e^{-px} f(x)\,dx.$$

To this end, let us multiply equation (M11.4.1.3) by e^{-px} and then integrate with respect to x from zero to infinity. Taking into account the formula

$$\mathfrak{L}\bigl\{y_x^{(n)}(x)\bigr\} = p^n \widetilde{y}(p) - \sum_{k=1}^{n} p^{n-k} y_x^{(k-1)}(+0)$$

and the initial conditions (M11.4.1.6), we arrive at a linear algebraic equation for the transform $\widetilde{y}(p)$:

$$P(p)\widetilde{y}(p) - Q(p) = \widetilde{f}(p), \tag{M11.4.1.7}$$

where

$$P(p) = p^n + a_{n-1}p^{n-1} + \cdots + a_1 p + a_0, \qquad Q(p) = b_{n-1}p^{n-1} + \cdots + b_1 p + b_0,$$
$$b_k = y_{n-k-1} + a_{n-1}y_{n-k-2} + \cdots + a_{k+2}y_1 + a_{k+1}y_0, \qquad k = 0,\ 1,\ \ldots,\ n-1.$$

The polynomial $P(p)$ coincides with the characteristic polynomial (M11.4.1.2) at $\lambda = p$.

The solution of equation (M11.4.1.7) is given by the formula

$$\widetilde{y}(p) = \frac{\widetilde{f}(p) + Q(p)}{\widetilde{P}(p)}. \tag{M11.4.1.8}$$

On applying the Laplace inversion formula (see in Section M10.2) to (M11.4.1.8), we obtain a solution to problem (M11.4.1.3), (M11.4.1.6) in the form

$$y(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\widetilde{f}(p) + Q(p)}{\widetilde{P}(p)} e^{px}\,dp. \tag{M11.4.1.9}$$

Since the transform $\widetilde{y}(p)$ (M11.4.1.8) is a rational function, the inverse Laplace transform (M11.4.1.9) can be obtained using the formulas from Subsection M10.2.2 or the tables of Section S2.2.

Remark. In practice, the solution method for the Cauchy problem based on the Laplace transform leads to the solution faster than the direct application of general formulas like (M11.4.1.4), where one has to determine the coefficients $C_1, \ldots, C_n$.

Example 3. Consider the following Cauchy problem for a homogeneous fourth-order equation:

$$y''''_{xxxx} + a^4 y = 0; \qquad y(0) = y'_x(0) = y'''_{xxx}(0) = 0, \quad y''_{xx}(0) = b.$$

The Laplace transform reduces this problem to a linear algebraic equation for $\widetilde{y}(p)$: $(p^4 + a^4)\widetilde{y}(p) - bp = 0$. It follows that

$$\widetilde{y}(p) = \frac{bp}{p^4 + a^4}.$$

In order to invert this expression, let us use the table of inverse Laplace transforms S2.2.2 (see row 52) and take into account that a constant multiplier can be taken outside the transform operator to obtain the solution to the original Cauchy problem in the form

$$y(x) = \frac{b}{a^2} \sin\Bigl(\frac{ax}{\sqrt{2}}\Bigr) \sinh\Bigl(\frac{ax}{\sqrt{2}}\Bigr).$$

M11.4.2. Linear Equations with Variable Coefficients

► **Homogeneous linear equations. Structure of the general solution.** The general solution of the nth-order homogeneous linear differential equation

$$f_n(x)y_x^{(n)} + f_{n-1}(x)y_x^{(n-1)} + \cdots + f_1(x)y'_x + f_0(x)y = 0 \qquad \text{(M11.4.2.1)}$$

has the form

$$y = C_1y_1(x) + C_2y_2(x) + \cdots + C_ny_n(x), \qquad \text{(M11.4.2.2)}$$

where the functions $y_1(x)$, $y_2(x)$, …, $y_n(x)$ are a fundamental system of solutions (the y_k are linearly independent particular solutions, $y_k \not\equiv 0$); C_1, C_2, …, C_n are arbitrary constants.

► **Utilization of particular solutions for reducing the order of the equation.**

1°. Let $y_1 = y_1(x)$ be a nontrivial particular solution of equation (M11.4.2.1). The substitution

$$y = y_1(x)\int z(x)\,dx$$

results in a linear equation of order $n-1$ for the function $z(x)$.

2°. Let $y_1 = y_1(x)$ and $y_2 = y_2(x)$ be two nontrivial linearly independent solutions of equation (M11.4.2.1). The substitution

$$y = y_1\int y_2w\,dx - y_2\int y_1w\,dx$$

results in a linear equation of order $n-2$ for $w(x)$.

3°. Suppose that m linearly independent solutions $y_1(x)$, $y_2(x)$, …, $y_m(x)$ of equation (M11.4.2.1) are known. Then one can reduce the order of the equation to $n-m$ by successive application of the following procedure. The substitution $y = y_m(x)\int z(x)\,dx$ leads to an equation of order $n-1$ for the function $z(x)$ with known linearly independent solutions:

$$z_1 = \left(\frac{y_1}{y_m}\right)'_x, \quad z_2 = \left(\frac{y_2}{y_m}\right)'_x, \quad \ldots, \quad z_{m-1} = \left(\frac{y_{m-1}}{y_m}\right)'_x.$$

The substitution $z = z_{m-1}(x)\int w(x)\,dx$ yields an equation of order $n-2$. Repeating this procedure m times, we arrive at a homogeneous linear equation of order $n-m$.

► **Wronskian determinant and Liouville formula.** The *Wronskian determinant* (or simply, *Wronskian*) is the function defined as

$$W(x) = \begin{vmatrix} y_1(x) & \cdots & y_n(x) \\ y'_1(x) & \cdots & y'_n(x) \\ \cdots & \cdots & \cdots \\ y_1^{(n-1)}(x) & \cdots & y_n^{(n-1)}(x) \end{vmatrix}, \qquad \text{(M11.4.2.3)}$$

where $y_1(x)$, …, $y_n(x)$ is a fundamental system of solutions of the homogeneous equation (M11.4.2.1); $y_k^{(m)}(x) = \dfrac{d^my_k}{dx^m}$, $m = 1, \ldots, n-1$; $k = 1, \ldots, n$.

The following *Liouville formula* holds:

$$W(x) = W(x_0)\exp\left[-\int_{x_0}^{x}\frac{f_{n-1}(t)}{f_n(t)}\,dt\right],$$

where x_0 is an arbitrary number.

► **Nonhomogeneous linear equations. Construction of the general solution.**

1°. The general nonhomogeneous nth-order linear differential equation has the form

$$f_n(x)y_x^{(n)} + f_{n-1}(x)y_x^{(n-1)} + \cdots + f_1(x)y_x' + f_0(x)y = g(x). \tag{M11.4.2.4}$$

The general solution of the nonhomogeneous equation (M11.4.2.4) is obtained as the sum of the general solution of the corresponding homogeneous equation (M11.4.2.1) and any particular solution of equation (M11.4.2.4).

2°. Let $y_1(x)$, ..., $y_n(x)$ be a fundamental system of solutions of the homogeneous equation (M11.4.2.1), and let $W(x)$ be the Wronskian determinant (M11.4.2.3). Then the general solution of the nonhomogeneous linear equation (M11.4.2.4) can be represented as

$$y = \sum_{\nu=1}^{n} C_\nu y_\nu(x) + \sum_{\nu=1}^{n} y_\nu(x) \int \frac{W_\nu(x)\,dx}{f_n(x)W(x)}, \tag{M11.4.2.5}$$

where $W_\nu(x)$ is the determinant obtained by replacing the νth column of the matrix (M11.4.2.3) by the column vector with the elements $0, 0, \ldots, 0, g$.

3°. *Superposition principle*. A particular solution of a nonhomogeneous linear equation

$$\mathbf{L}[y] = \sum_{k=1}^{m} g_k(x), \qquad \mathbf{L}[y] \equiv f_n(x)y_x^{(n)} + f_{n-1}(x)y_x^{(n-1)} + \cdots + f_1(x)y_x' + f_0(x)y$$

is determined by adding together particular solutions,

$$y = \sum_{k=1}^{m} y_k,$$

of m (simpler) equations,

$$\mathbf{L}[y_k] = g_k(x), \qquad k = 1, 2, \ldots, m,$$

corresponding to respective nonhomogeneous terms in the original equation.

► **Euler equation.** The nonhomogeneous Euler equation has the form

$$x^n y_x^{(n)} + a_{n-1}x^{n-1}y_x^{(n-1)} + \cdots + a_1 x y_x' + a_0 y = f(x).$$

The substitution $x = be^t$ ($b \neq 0$) leads to a constant coefficient linear equation of the form (M11.4.1.3).

Particular solutions of the homogeneous Euler equation [with $f(x) \equiv 0$] are sought in the form $y = x^k$. If all k are real and distinct, its general solution is expressed as

$$y(x) = C_1|x|^{k_1} + C_2|x|^{k_2} + \cdots + C_n|x|^{k_n}.$$

Remark. To a pair of complex conjugate values $k = \alpha \pm i\beta$ there corresponds a pair of particular solutions: $y = |x|^\alpha \sin(\beta|x|)$ and $y = |x|^\alpha \cos(\beta|x|)$.

M11.4.3. Systems of Linear Equations with Constant Coefficients

► Systems of first-order linear homogeneous equations. The general solution.

1°. In general, a homogeneous linear system of first-order ordinary differential equations with constant coefficients has the form

$$\begin{aligned} y_1' &= a_{11}y_1 + a_{12}y_2 + \cdots + a_{1n}y_n,\\ y_2' &= a_{21}y_1 + a_{22}y_2 + \cdots + a_{2n}y_n,\\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ y_n' &= a_{n1}y_1 + a_{n2}y_2 + \cdots + a_{nn}y_n, \end{aligned} \qquad \text{(M11.4.3.1)}$$

where a prime stands for the derivative with respect to x. In the sequel, all the coefficients a_{ij} of the system are assumed to be real numbers.

The homogeneous system (M11.4.3.1) has the trivial particular solution $y_1 = y_2 = \cdots = y_n = 0$.

Superposition principle for a homogeneous system: any linear combination of particular solutions of system (M11.4.3.1) is also a solution of this system.

The general solution of the system of differential equations (M11.4.3.1) is the sum of its n linearly independent (nontrivial) particular solutions each multiplied by an arbitrary constant.

Remark. System (M11.4.3.1) can be reduced to a single homogeneous linear constant-coefficient nth-order equation.

2°. For brevity (and clarity), system (M11.4.3.1) is conventionally written in vector-matrix form:

$$\mathbf{y}' = \mathbf{A}\mathbf{y}, \qquad \text{(M11.4.3.2)}$$

where $\mathbf{y} = (y_1, y_2, \dots, y_n)^{\mathrm{T}}$ is the column vector of the unknowns and $\mathbf{A} = (a_{ij})$ is the matrix of the equation coefficients. The superscript T denotes the transpose of a matrix or a vector. So, for example, a row vector is converted into a column vector:

$$(y_1, y_2)^{\mathrm{T}} \equiv \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

The right-hand side of equation (M11.4.3.2) is the product of the $n \times n$ square matrix **a** by the $n \times 1$ matrix (column vector) **y**.

Let $\mathbf{y}_k = (y_{k1}, y_{k2}, \dots, y_{kn})^{\mathrm{T}}$ be linearly independent particular solutions* of the homogeneous system (M11.4.3.1), where $k = 1, 2, \dots, n$; the first subscript in $y_{km} = y_{km}(x)$ denotes the number of the solution and the second subscript ($m = 1, \dots, n$) indicates the component of the vector solution. Then the general solution of the homogeneous system (M11.4.3.2) is expressed as

$$\mathbf{y} = C_1\mathbf{y}_1 + C_2\mathbf{y}_2 + \cdots + C_n\mathbf{y}_n. \qquad \text{(M11.4.3.3)}$$

A method for the construction of particular solutions that can be used to obtain the general solution by formula (M11.4.3.3) is presented below.

---

* This means that the condition $\det|y_{km}(x)| \neq 0$ holds.

► **Systems of first-order linear homogeneous equations. Particular solutions.** Particular solutions to system (M11.4.3.1) are determined by the roots of the characteristic equation

$$\Delta(\lambda)=0, \quad \text{where} \quad \Delta(\lambda)\equiv\begin{vmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}-\lambda & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn}-\lambda \end{vmatrix}. \tag{M11.4.3.4}$$

The following cases are possible:

1°. Let λ be a simple real root of the characteristic equation (M11.4.3.4). The corresponding particular solution of the homogeneous linear system of equations (M11.4.3.1) has the exponential form

$$y_1=A_1e^{\lambda x}, \quad y_2=A_2e^{\lambda x}, \quad \dots, \quad y_n=A_ne^{\lambda x}, \tag{M11.4.3.5}$$

where the coefficients A_1, A_2, …, A_n are determined by solving the associated linear homogeneous system of equations:

$$\begin{aligned} &(a_{11}-\lambda)A_1+a_{12}A_2+\dots+a_{1n}A_n=0,\\ &a_{21}A_1+(a_{22}-\lambda)A_2+\dots+a_{2n}A_n=0,\\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\ &a_{n1}A_1+a_{n2}A_2+\dots+(a_{nn}-\lambda)A_n=0. \end{aligned} \tag{M11.4.3.6}$$

The solution of this system is unique to within a constant factor.

If all roots of the characteristic equation λ_1, λ_2, …, λ_n are real and distinct, then the general solution of system (M11.4.3.1) has the form

$$\begin{aligned} &y_1=C_1A_{11}e^{\lambda_1x}+C_2A_{12}e^{\lambda_2x}+\dots+C_nA_{1n}e^{\lambda_nx},\\ &y_2=C_1A_{21}e^{\lambda_1x}+C_2A_{22}e^{\lambda_2x}+\dots+C_nA_{2n}e^{\lambda_nx},\\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\ &y_n=C_1A_{n1}e^{\lambda_1x}+C_2A_{n2}e^{\lambda_2x}+\dots+C_nA_{nn}e^{\lambda_nx}, \end{aligned} \tag{M11.4.3.7}$$

where C_1, C_2, …, C_n are arbitrary constants. The second subscript in A_{mk} indicates a coefficient corresponding to the root λ_k.

2°. For each simple complex root, $\lambda=\alpha+i\beta$, of the characteristic equation (M11.4.3.4), the corresponding particular solution is obtained in the same way as in the simple real root case; the associated coefficients $A_1, A_2, \dots, A_n$ in (M11.4.3.5) will be complex. Separating the real and imaginary parts in (M11.4.3.5) results in two real particular solutions to system (M11.4.3.1); the same two solutions are obtained if one takes the complex conjugate root, $\bar\lambda=\alpha-i\beta$.

3°. Let λ be a real root of the characteristic equation (M11.4.3.4) of multiplicity m. The corresponding particular solution of system (M11.4.3.1) is sought in the form

$$y_1=P_m^1(x)e^{\lambda x}, \quad y_2=P_m^2(x)e^{\lambda x}, \quad \dots, \quad y_n=P_m^n(x)e^{\lambda x}, \tag{M11.4.3.8}$$

where the $P_m^k(x)=\sum\limits_{i=0}^{m-1}B_{ki}x^i$ are polynomials of degree $m-1$. The coefficients of these polynomials result from the substitution of expressions (M11.4.3.8) into equations (M11.4.3.1); after dividing by $e^{\lambda x}$ and collecting like terms, one obtains n equations, each representing a polynomial equated to zero. By equating the coefficients of all resulting polynomials to zero, one arrives at a linear algebraic system of equations for the coefficients B_{ki}; the solution to this system will contain m free parameters.

4°. For a multiple complex root, $\lambda = \alpha + i\beta$, of multiplicity m, the corresponding particular solution is sought, just as in the case of a multiple real root, in the form (M11.4.3.8); here the coefficients B_{ki} of the polynomials $P_m^k(x)$ will be complex. Finally, in order to obtain real solutions of the original system (M11.4.3.1), one separates the real and imaginary parts in formulas (M11.4.3.8), thus obtaining two particular solutions with m free parameters each. The two solutions correspond to the complex conjugate roots $\lambda = \alpha \pm i\beta$.

5°. In the general case, where the characteristic equation (M11.4.3.4) has simple and multiple, real and complex roots (see Items 1°–4°), the general solution to system (M11.4.3.1) is obtained as the sum of all particular solutions multiplied by arbitrary constants.

Example 1. Consider the homogeneous system of two linear differential equations

$$\begin{aligned} y_1' &= y_1 + 4y_2, \\ y_2' &= y_1 + y_2. \end{aligned}$$

The associated characteristic equation,

$$\begin{vmatrix} 1-\lambda & 4 \\ 1 & 1-\lambda \end{vmatrix} = \lambda^2 - 2\lambda - 3 = 0,$$

has distinct real roots:

$$\lambda_1 = 3, \qquad \lambda_2 = -1.$$

The system of algebraic equations (M11.4.3.6) for the solution coefficients becomes

$$\begin{aligned} (1-\lambda)A_1 + 4A_2 &= 0, \\ A_1 + (1-\lambda)A_2 &= 0. \end{aligned} \tag{M11.4.3.9}$$

Substituting the first root, $\lambda = 3$, into system (M11.4.3.9) yields $A_1 = 2A_2$. We can set $A_1 = 2$ and $A_2 = 1$, since the solution is determined to within a constant factor. Thus the first particular solution of the homogeneous system of linear ordinary differential equations (M11.4.3.9) has the form

$$y_1 = 2e^{3x}, \qquad y_2 = e^{3x}. \tag{M11.4.3.10}$$

The second particular solution, corresponding to $\lambda = -1$, is found in the same way:

$$y_1 = -2e^{-x}, \qquad y_2 = e^{-x}. \tag{M11.4.3.11}$$

The sum of the two particular solutions (M11.4.3.10) and (M11.4.3.11) multiplied by arbitrary constants, C_1 and C_2, gives the general solution to the original homogeneous system of linear ordinary differential equations:

$$y_1 = 2C_1e^{3x} - 2C_2e^{-x}, \qquad y_2 = C_1e^{3x} + C_2e^{-x}.$$

Example 2. Consider the system of ordinary differential equations

$$\begin{aligned} y_1' &= -y_2, \\ y_2' &= 2y_1 + 2y_2. \end{aligned} \tag{M11.4.3.12}$$

The characteristic equation

$$\begin{vmatrix} -\lambda & -1 \\ 2 & 2-\lambda \end{vmatrix} = \lambda^2 - 2\lambda + 2 = 0$$

has complex conjugate roots:

$$\lambda_1 = 1 + i, \qquad \lambda_2 = 1 - i.$$

The algebraic system (M11.4.3.6) for the complex coefficients A_1 and A_2 becomes

$$\begin{aligned} -\lambda A_1 - A_2 &= 0, \\ 2A_1 + (2-\lambda)A_2 &= 0. \end{aligned}$$

With $\lambda = 1 + i$, one nonzero solution is given by $A_1 = 1$ and $A_2 = -1 - i$. The corresponding complex solution to system (M11.4.3.12) has the form

$$y_1 = e^{(1+i)x}, \qquad y_2 = (-1-i)e^{(1+i)x}.$$

Separating the real and imaginary parts, taking into account the formulas

$$e^{(1+i)x} = e^x(\cos x + i\sin x) = e^x\cos x + ie^x\sin x,$$
$$(-1-i)e^{(1+i)x} = -(1+i)e^x(\cos x + i\sin x) = e^x(\sin x - \cos x) - ie^x(\sin x + \cos x),$$

and making linear combinations of them, one arrives at the general solution to the original system (M11.4.3.12):

$$y_1 = C_1e^x\cos x + C_2e^x\sin x,$$
$$y_2 = C_1e^x(\sin x - \cos x) - C_2e^x(\sin x + \cos x).$$

► **Nonhomogeneous systems of linear first-order equations.**

1°. In general, a nonhomogeneous linear system of first-order differential equations with constant coefficients has the form

$$\begin{aligned} y_1' &= a_{11}y_1 + a_{12}y_2 + \cdots + a_{1n}y_n + f_1(x),\\ y_2' &= a_{21}y_1 + a_{22}y_2 + \cdots + a_{2n}y_n + f_2(x),\\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ y_n' &= a_{n1}y_1 + a_{n2}y_2 + \cdots + a_{nn}y_n + f_n(x). \end{aligned} \qquad \text{(M11.4.3.13)}$$

For brevity, the conventional vector notation will also be used:

$$\mathbf{y}' = \mathbf{A}\mathbf{y} + \mathbf{f}(x),$$

where $\mathbf{f}(x) = (f_1(x), f_2(x), \ldots, f_n(x))^{\mathrm{T}}$.

The general solution of this system is the sum of the general solution to the corresponding homogeneous system with $f_k(x) \equiv 0$ [see system (M11.4.3.1)] and any particular solution of the nonhomogeneous system (M11.4.3.13).

2°. Let $\mathbf{y}_k = (D_{k1}(x), D_{k1}(x), \ldots, D_{kn}(x))^{\mathrm{T}}$ be particular solutions to the homogeneous linear system of first-order constant-coefficient differential equations (M11.4.3.1) that satisfy the special initial conditions

$$y_k(0) = 1, \quad y_m(0) = 0 \quad \text{for} \quad m \neq k; \quad k,\, m = 1,\, \ldots,\, n.$$

Then the general solution to the nonhomogeneous system (M11.4.3.13) is expressed as

$$y_m(x) = \sum_{k=1}^{n} \int_0^x f_k(t)D_{km}(x-t)\,dt + \sum_{k=1}^{n} C_kD_{km}(x), \qquad m = 1, \ldots, n. \qquad \text{(M11.4.3.14)}$$

The solution of the Cauchy problem for the nonhomogeneous system (M11.4.3.13) with arbitrary initial conditions,

$$y_1(0) = y_1^\circ, \quad y_2(0) = y_2^\circ, \quad \ldots, \quad y_n(0) = y_n^\circ, \qquad \text{(M11.4.3.15)}$$

is determined by formulas (M11.4.3.14) with $C_k = y_k^\circ$, $k = 1, \ldots, n$.

M11.5. Nonlinear Equations of Arbitrary Order

M11.5.1. Structure of the General Solution. Cauchy Problem

► **Equations solved for the highest derivative. General solution.** An *nth-order differential equation solved for the highest derivative* has the form

$$y_x^{(n)} = f(x, y, y_x', \ldots, y_x^{(n-1)}). \qquad \text{(M11.5.1.1)}$$

The general solution of this equation depends on n arbitrary constants $C_1, \dots, C_n$. In some cases, the general solution can be written in explicit form as

$$y = \varphi(x, C_1, \dots, C_n). \qquad \text{(M11.5.1.2)}$$

► **Cauchy problem. The existence and uniqueness theorem.** The *Cauchy problem*: find a solution of equation (M11.5.1.1) with the *initial conditions*

$$y(x_0) = y_0, \quad y'_x(x_0) = y_0^{(1)}, \quad \dots, \quad y_x^{(n-1)}(x_0) = y_0^{(n-1)}. \qquad \text{(M11.5.1.3)}$$

(At a point x_0, the values of the unknown function $y(x)$ and all its derivatives of orders $\le n-1$ are prescribed.)

EXISTENCE AND UNIQUENESS THEOREM. *Suppose the function $f(x, y, z_1, \dots, z_{n-1})$ is continuous in all its arguments in a neighborhood of the point $(x_0, y_0, y_0^{(1)}, \dots, y_0^{(n-1)})$ and has bounded derivatives with respect to y, z_1, $\dots$, z_{n-1} in this neighborhood. Then a solution of equation (M11.5.1.1) satisfying the initial conditions (M11.5.1.3) exists and is unique.*

► **Reduction of an nth-order equation to a system of n first-order equations.** The differential equation (M11.5.1.1) is equivalent to the following system of n first-order equations:

$$y'_0 = y_1, \quad y'_1 = y_2, \quad \dots, \quad y'_{n-2} = y_{n-1}, \quad y'_{n-1} = f(x, y_0, y_1, \dots, y_{n-1}),$$

where the notation $y_0 \equiv y$ is adopted.

M11.5.2. Equations Admitting Reduction of Order

► **Equations not containing $y, y'_x, \dots, y_x^{(k)}$ explicitly.** An equation that does not explicitly contain the unknown function and its derivatives up to order k inclusive can generally be written as

$$F\big(x, y_x^{(k+1)}, \dots, y_x^{(n)}\big) = 0 \qquad (1 \le k+1 < n). \qquad \text{(M11.5.2.1)}$$

Such equations are invariant under arbitrary translations of the unknown function, $y \to y + \text{const}$ (the form of such equations is also preserved under the transformation $u(x) = y + a_k x^k + \dots + a_1 x + a_0$, where the a_m are arbitrary constants). The substitution $z(x) = y_x^{(k+1)}$ reduces (M11.5.2.1) to an equation whose order is by $k+1$ smaller than that of the original equation, $F\big(x, z, z'_x, \dots, z_x^{(n-k-1)}\big) = 0$.

► **Equations not containing x explicitly (autonomous equations).** In general, an equation that does not explicitly contain x has form

$$F\big(y, y'_x, \dots, y_x^{(n)}\big) = 0. \qquad \text{(M11.5.2.2)}$$

Such equations are invariant under arbitrary translations of the independent variable, $x \to x + \text{const}$. The substitution $y'_x = w(y)$ (where y plays the role of the independent variable) reduces by one the order of an autonomous equation. Higher derivatives can be expressed in terms of w and its derivatives with respect to the new independent variable, $y''_{xx} = ww'_y$, $y'''_{xxx} = w^2 w''_{yy} + w(w'_y)^2, \dots$

► **Some other equations admitting reduction of order.** Table M11.5 lists the above nonlinear equations as well as some other equations admitting order reduction. The second column gives simple transformations that allow checking whether the equation is one of this type.

TABLE M11.5
Some ordinary differential equations that admit order reduction by one.

Equation	Transformation preserving form of equation (a is an arbitrary constant)	Transformation reducing order of equation, $u=u(z)$
$F\bigl(x, y'_x, y''_{xx}, \dots, y_x^{(n)}\bigr)=0$	$y=\bar{y}+a$	$z=x,\ u=y'_x$
$F\bigl(y, y'_x, y''_{xx}, \dots, y_x^{(n)}\bigr)=0$	$x=\bar{x}+a$	$z=y,\ u=y'_x$
$F\bigl(\alpha x+\beta y+\gamma, y'_x, y''_{xx}, \dots, y_x^{(n)}\bigr)=0$	$x=\bar{x}+a\beta,\ y=\bar{y}-a\alpha$	$z=\alpha x+\beta y+\gamma,\ u=y'_x$
$F\bigl(x, y'_x/y, y''_{xx}/y, \dots, y_x^{(n)}/y\bigr)=0$	$y=a\bar{y}$	$z=x,\ u=y'_x/y$
$F\bigl(y, xy'_x, x^2y''_{xx}, \dots, x^ny_x^{(n)}\bigr)=0$	$x=a\bar{x}$	$z=y,\ u=xy'_x$
$F\bigl(y/x, y'_x, xy''_{xx}, \dots, x^{n-1}y_x^{(n)}\bigr)=0$	$x=a\bar{x},\ y=a\bar{y}$	$z=y/x,\ u=y'_x$
$F\bigl(e^{\lambda x}y, y'_x/y, y''_{xx}/y, \dots, y_x^{(n)}/y\bigr)=0$	$x=\bar{x}-\frac{1}{\lambda}\ln a,\ y=a\bar{y}$	$z=e^{\lambda x}y,\ u=y'_x/y$
$F\bigl(xe^{\lambda y}, xy'_x, x^2y''_{xx}, \dots, x^ny_x^{(n)}\bigr)=0$	$x=a\bar{x},\ y=\bar{y}-\frac{1}{\lambda}\ln a$	$z=xe^{\lambda y},\ u=xy'_x$
$F\bigl(x^ky, x^{k+1}y'_x, x^{k+2}y''_{xx}, \dots, x^{k+n}y_x^{(n)}\bigr)=0$	$x=a\bar{x},\ y=a^{-k}\bar{y}$	$z=x^ky,\ u=x^{k+1}y'_x$

Bibliography for Chapter M11

Boyce, W. E. and DiPrima, R. C., *Elementary Differential Equations and Boundary Value Problems, 8th Edition*, John Wiley & Sons, New York, 2004.

Cole, G. D., *Perturbation Methods in Applied Mathematics*, Blaisdell Publishing Company, Waltham, Massachusetts, 1968.

Dormand, J. R., *Numerical Methods for Differential Equations: A Computational Approach*, CRC Press, Boca Raton, Florida, 1996.

El'sgol'ts, L. E., *Differential Equations*, Gordon & Breach, New York, 1961.

Ince, E. L., *Ordinary Differential Equations*, Dover Publications, New York, 1964.

Kamke, E., *Differentialgleichungen: Lösungsmethoden und Lösungen, I, Gewöhnliche Differentialgleichungen*, B. G. Teubner, Leipzig, 1977.

Marchenko, V. A., *Sturm–Liouville Operators and Applications*, Birkhäuser Verlag, Basel, 1986.

Murphy, G. M., *Ordinary Differential Equations and Their Solutions*, D. Van Nostrand, New York, 1960.

Nayfeh, A. H., *Perturbation Methods*, Wiley-Interscience, New York, 1973.

Nayfeh, A. H., *Introduction to Perturbation Techniques*, John Wiley & Sons, New York, 1981.

Polyanin, A. D. and Zaitsev, V. F., *Handbook of Exact Solutions for Ordinary Differential Equations, 2nd Edition*, Chapman & Hall/CRC Press, Boca Raton, 2003.

Schiesser, W. E., *Computational Mathematics in Engineering and Applied Science: ODEs, DAEs, and PDEs*, CRC Press, Boca Raton, Florida, 1993.

Tenenbaum, M. and Pollard, H., *Ordinary Differential Equations*, Dover Publications, New York, 1985.

Zwillinger, D., *Handbook of Differential Equations, 3rd Edition*, Academic Press, New York, 1997.

Chapter M12

Partial Differential Equations

M12.1. First-Order Quasilinear Partial Differential Equations

M12.1.1. Characteristic System. General Solution

► **Equations with two independent variables. General solution. Examples.** A *first-order quasilinear partial differential equation with two independent variables* has the general form

$$f(x,y,w)\frac{\partial w}{\partial x}+g(x,y,w)\frac{\partial w}{\partial y}=h(x,y,w). \tag{M12.1.1.1}$$

Such equations are encountered in various applications (continuum mechanics, gas dynamics, hydrodynamics, heat and mass transfer, wave theory, acoustics, multiphase flows, chemical engineering, etc.).

If two independent integrals,

$$u_1(x,y,w)=C_1, \qquad u_2(x,y,w)=C_2, \tag{M12.1.1.2}$$

of the *characteristic system*

$$\frac{dx}{f(x,y,w)}=\frac{dy}{g(x,y,w)}=\frac{dw}{h(x,y,w)} \tag{M12.1.1.3}$$

are known, then the *general solution* of equation (M12.1.1.1) is given by

$$\Phi(u_1,u_2)=0, \tag{M12.1.1.4}$$

where $\Phi(u,v)$ is an arbitrary function of two variables. With equation (M12.1.1.4) solved for u_1 or u_2, we often specify the general solution in the form

$$u_k=\Psi(u_{3-k}),$$

where $k=1$ or 2 and $\Psi(u)$ is an arbitrary function of one variable.

Remark. In the special case $h(x,y,w)\equiv 0$, one of the integrals of the characteristic system is $w=C_1$. Another integral may be determined from the first equation in (M12.1.1.3).

Example. Consider the linear constant coefficient equation

$$\frac{\partial w}{\partial x}+a\frac{\partial w}{\partial y}=b.$$

The characteristic system for this equation is

$$\frac{dx}{1}=\frac{dy}{a}=\frac{dw}{b}.$$

It has two independent integrals:

$$y-ax=C_1, \quad w-bx=C_2.$$

Hence, the general solution of the original equation is given by $\Phi(y-ax,\ w-bx)=0$. On solving this equation for w, one obtains the general solution in explicit form

$$w=bx+\Psi(y-ax),$$

where $\Psi(u)$ is an arbitrary function.

► **Equations with n independent variables. General solution.** A *first-order quasilinear partial differential equation with n independent variables* has the general form

$$f_1(x_1,\dots,x_n,w)\frac{\partial w}{\partial x_1}+\dots+f_n(x_1,\dots,x_n,w)\frac{\partial w}{\partial x_n}=g(x_1,\dots,x_n,w). \qquad \text{(M12.1.1.5)}$$

Let n independent integrals,

$$u_1(x_1,\dots,x_n,w)=C_1,\quad \dots,\quad u_n(x_1,\dots,x_n,w)=C_n,$$

of the characteristic system

$$\frac{dx_1}{f_1(x_1,\dots,x_n,w)}=\dots=\frac{dx_n}{f_n(x_1,\dots,x_n,w)}=\frac{dw}{g(x_1,\dots,x_n,w)}$$

be known. Then the general solution of equation (M12.1.1.5) is given by

$$\Phi(u_1,\dots,u_n)=0,$$

where Φ is an arbitrary function of n variables.

M12.1.2. Cauchy Problem

► **Two formulations of the Cauchy problem.** Consider two formulations of the Cauchy problem.

1°. *Generalized Cauchy problem.* Find a solution $w=w(x,y)$ of equation (M12.1.1.1) satisfying the initial conditions

$$x=h_1(\xi),\quad y=h_2(\xi),\quad w=h_3(\xi), \qquad \text{(M12.1.2.1)}$$

where ξ is a parameter ($\alpha\le\xi\le\beta$) and the $h_k(\xi)$ are given functions.

Geometric interpretation: find an integral surface of equation (M12.1.1.1) passing through the line defined parametrically by equations (M12.1.2.1).

2°. *Classical Cauchy problem.* Find a solution $w=w(x,y)$ of equation (M12.1.1.1) satisfying the initial condition

$$w=\varphi(y)\quad\text{at}\quad x=0, \qquad \text{(M12.1.2.2)}$$

where $\varphi(y)$ is a given function.

It is convenient to represent the classical Cauchy problem as a generalized Cauchy problem by rewriting condition (M12.1.2.2) in the parametric form

$$x=0,\quad y=\xi,\quad w=\varphi(\xi). \qquad \text{(M12.1.2.3)}$$

► **Procedure of solving the Cauchy problem.** The procedure of solving the Cauchy problem (M12.1.1.1), (M12.1.2.1) involves several steps. First, two independent integrals (M12.1.1.2) of the characteristic system (M12.1.1.3) are determined. Then, to find the constants of integration C_1 and C_2, the initial data (M12.1.2.1) must be substituted into the integrals (M12.1.1.2) to obtain

$$u_1\bigl(h_1(\xi),h_2(\xi),h_3(\xi)\bigr)=C_1,\qquad u_2\bigl(h_1(\xi),h_2(\xi),h_3(\xi)\bigr)=C_2. \qquad \text{(M12.1.2.4)}$$

Eliminating C_1 and C_2 from (M12.1.1.2) and (M12.1.2.4) yields

$$\begin{aligned}u_1(x,y,w)&=u_1\bigl(h_1(\xi),h_2(\xi),h_3(\xi)\bigr),\\ u_2(x,y,w)&=u_2\bigl(h_1(\xi),h_2(\xi),h_3(\xi)\bigr).\end{aligned} \qquad \text{(M12.1.2.5)}$$

Formulas (M12.1.2.5) are a parametric form of the solution of the Cauchy problem (M12.1.1.1), (M12.1.2.1). In some cases, one may succeed in eliminating the parameter ξ from relations (M12.1.2.5), thus obtaining the solution in an explicit form.

Example 1. Consider the Cauchy problem for linear equation

$$\frac{\partial w}{\partial x}+a\frac{\partial w}{\partial y}=bw \tag{M12.1.2.6}$$

subjected to the initial condition (M12.1.2.2).

The corresponding characteristic system for equation (M12.1.2.6),

$$\frac{dx}{1}=\frac{dy}{a}=\frac{dw}{bw},$$

has two independent integrals

$$y-ax=C_1,\qquad we^{-bx}=C_2. \tag{M12.1.2.7}$$

Represent the initial condition (M12.1.2.2) in parametric form (M12.1.2.3) and then substitute the data (M12.1.2.3) into the integrals (M12.1.2.7). As a result, for the constants of integration we obtain $C_1=\xi$ and $C_2=\varphi(\xi)$. Substituting these expressions into (M12.1.2.7), we arrive at the solution of the Cauchy problem (M12.1.2.6), (M12.1.2.2) in parametric form:

$$y-ax=\xi,\qquad we^{-bx}=\varphi(\xi).$$

By eliminating the parameter ξ from these relations, we obtain the solution of the Cauchy problem (M12.1.2.6), (M12.1.2.2) in explicit form:

$$w=e^{bx}\varphi(y-ax).$$

Example 2. Consider the Cauchy problem for *Hopf's equation*

$$\frac{\partial w}{\partial x}+w\frac{\partial w}{\partial y}=0 \tag{M12.1.2.8}$$

subject to the initial condition (M12.1.2.2).

First, we rewrite the initial condition (M12.1.2.2) in the parametric form (M12.1.2.3). Solving the characteristic system

$$\frac{dx}{1}=\frac{dy}{w}=\frac{dw}{0}, \tag{M12.1.2.9}$$

we find two independent integrals,

$$w=C_1,\quad y-wx=C_2. \tag{M12.1.2.10}$$

Using the initial conditions (M12.1.2.3), we find that $C_1=\varphi(\xi)$ and $C_2=\xi$. Substituting these expressions into (M12.1.2.10) yields the solution of the Cauchy problem (M12.1.2.8), (M12.1.2.2) in the parametric form

$$w=\varphi(\xi), \tag{M12.1.2.11}$$

$$y=\xi+\varphi(\xi)x. \tag{M12.1.2.12}$$

The straight lines defined by equation (M12.1.2.12) are called *characteristics*. They have the slope $\varphi(\xi)$ and intersect the y-axis at the points ξ. On each characteristic, the function w has the same value equal to $\varphi(\xi)$ (generally, w takes different values on different characteristics).

For $\varphi'(\xi)>0$, different characteristics do not intersect and, hence, formulas (M12.1.2.11) and (M12.1.2.12) define a unique solution.

M12.2. Classification of Second-Order Linear Partial Differential Equations

M12.2.1. Equations with Two Independent Variables

► **Examples of equations encountered in applications.** There are three basic types of linear partial differential equations—*parabolic*, *hyperbolic*, and *elliptic*. The solutions of the equations pertaining to each of the types have their own characteristic qualitative differences.

The simplest example of a *parabolic equation* is the *heat equation*

$$\frac{\partial w}{\partial t}-\frac{\partial^2 w}{\partial x^2}=0, \tag{M12.2.1.1}$$

where the variables t and x play the role of time and the spatial coordinate, respectively. Note that equation (M12.2.1.1) contains only one highest derivative term. Frequently encountered particular solutions of equation (M12.2.1.1) can be found at the end of Subsection M12.4.1.

The simplest example of a *hyperbolic equation* is the *wave equation*

$$\frac{\partial^2 w}{\partial t^2} - \frac{\partial^2 w}{\partial x^2} = 0, \tag{M12.2.1.2}$$

where the variables t and x play the role of time and the spatial coordinate, respectively. Note that the highest derivative terms in equation (M12.2.1.2) differ in sign. The general solution of equation (M12.2.1.2) can be found at the end of Subsection M12.4.1.

The simplest example of an *elliptic equation* is the *Laplace equation*

$$\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} = 0, \tag{M12.2.1.3}$$

where x and y play the role of the spatial coordinates. Note that the highest derivative terms in equation (M12.2.1.3) have like signs. Frequently encountered particular solutions of equation (M12.2.1.3) can be found at the end of Subsection M12.4.1.

Any linear partial differential equation of the second order with two independent variables can be reduced, by appropriate manipulations, to a simpler equation that has one of the three highest derivative combinations specified above in examples (M12.2.1.1), (M12.2.1.2), and (M12.2.1.3).

► **Types of equations. Characteristic equations.** Consider a second-order partial differential equation with two independent variables that has the general form

$$a(x,y)\frac{\partial^2 w}{\partial x^2} + 2b(x,y)\frac{\partial^2 w}{\partial x \partial y} + c(x,y)\frac{\partial^2 w}{\partial y^2} = F\left(x, y, w, \frac{\partial w}{\partial x}, \frac{\partial w}{\partial y}\right), \tag{M12.2.1.4}$$

where a, b, c are some functions of x and y that have continuous derivatives up to the second order inclusive.*

Given a point (x, y), equation (M12.2.1.4) is said to be

$$\begin{array}{ll} \textit{parabolic} & \text{if } b^2 - ac = 0, \\ \textit{hyperbolic} & \text{if } b^2 - ac > 0, \\ \textit{elliptic} & \text{if } b^2 - ac < 0 \end{array}$$

at this point.

In order to reduce equation (M12.2.1.4) to a canonical form, one should first write out the characteristic equation

$$a\,(dy)^2 - 2b\,dx\,dy + c\,(dx)^2 = 0,$$

which splits into two equations

$$a\,dy - \left(b + \sqrt{b^2 - ac}\right)dx = 0 \tag{M12.2.1.5}$$

and

$$a\,dy - \left(b - \sqrt{b^2 - ac}\right)dx = 0, \tag{M12.2.1.6}$$

and then find their general integrals.

---

* The right-hand side of equation (M12.2.1.4) may be nonlinear. The classification and the procedure of reducing such equations to a canonical form are only determined by the left-hand side of the equation.

Remark. The characteristic equations (M12.2.1.5) and (M12.2.1.6) may be used if $a \not\equiv 0$. If $a \equiv 0$, the simpler equations

$$dx = 0,$$
$$2b\,dy - c\,dx = 0$$

should be used; the first equation has the obvious general solution $x = C$.

► **Canonical form of parabolic equations (case $b^2 - ac = 0$).** In this case, equations (M12.2.1.5) and (M12.2.1.6) coincide and have a common general integral,

$$\varphi(x,y) = C.$$

By passing from x, y to new independent variables ξ, η in accordance with the relations

$$\xi = \varphi(x,y), \qquad \eta = \eta(x,y),$$

where $\eta = \eta(x,y)$ is any twice differentiable function that satisfies the condition of nondegeneracy of the Jacobian $\frac{D(\xi,\eta)}{D(x,y)}$ in the given domain, we reduce equation (M12.2.1.4) to the canonical form

$$\frac{\partial^2 w}{\partial \eta^2} = F_1\left(\xi, \eta, w, \frac{\partial w}{\partial \xi}, \frac{\partial w}{\partial \eta}\right). \qquad \text{(M12.2.1.7)}$$

As η, one can take $\eta = x$ or $\eta = y$.

It is apparent that the transformed equation (M12.2.1.7) has only one highest-derivative term, just as the heat equation (M12.2.1.1).

Remark. In the degenerate case where the function F_1 does not depend on the partial derivative $\partial_\xi w$, equation (M12.2.1.7) is an ordinary differential equation for the variable η, in which ξ serves as a parameter.

► **Canonical forms of hyperbolic equations (case $b^2 - ac > 0$).** The general integrals

$$\varphi(x,y) = C_1, \qquad \psi(x,y) = C_2$$

of equations (M12.2.1.5) and (M12.2.1.6) are real and different. These integrals determine two different families of real characteristics.

By passing from x, y to new independent variables ξ, η in accordance with the relations

$$\xi = \varphi(x,y), \qquad \eta = \psi(x,y),$$

we reduce equation (M12.2.1.4) to

$$\frac{\partial^2 w}{\partial \xi \partial \eta} = F_2\left(\xi, \eta, w, \frac{\partial w}{\partial \xi}, \frac{\partial w}{\partial \eta}\right).$$

This is the so-called *first canonical form of a hyperbolic equation*.

The transformation

$$\xi = t + z, \qquad \eta = t - z$$

brings the above equation to another canonical form,

$$\frac{\partial^2 w}{\partial t^2} - \frac{\partial^2 w}{\partial z^2} = F_3\left(t, z, w, \frac{\partial w}{\partial t}, \frac{\partial w}{\partial z}\right),$$

where $F_3 = 4F_2$. This is the so-called *second canonical form of a hyperbolic equation*. Apart from notation, the left-hand side of the last equation coincides with that of the wave equation (M12.2.1.2).

In some cases, reduction of an equation to a canonical form allows finding its general solution.

► **Canonical form of elliptic equations (case $b^2 - ac < 0$).** In this case the general integrals of equations (M12.2.1.5) and (M12.2.1.6) are complex conjugate; these determine two families of complex characteristics.

Let the general integral of equation (M12.2.1.5) have the form

$$\varphi(x,y) + i\psi(x,y) = C, \qquad i^2 = -1,$$

where $\varphi(x,y)$ and $\psi(x,y)$ are real-valued functions.

By passing from x, y to new independent variables ξ, η in accordance with the relations

$$\xi = \varphi(x,y), \qquad \eta = \psi(x,y),$$

we reduce equation (M12.2.1.4) to the canonical form

$$\frac{\partial^2 w}{\partial \xi^2} + \frac{\partial^2 w}{\partial \eta^2} = F_4\left(\xi, \eta, w, \frac{\partial w}{\partial \xi}, \frac{\partial w}{\partial \eta}\right).$$

Apart from notation, the left-hand side of the last equation coincides with that of the Laplace equation (M12.2.1.3).

M12.2.2. Equations with Many Independent Variables

Let us consider a second-order partial differential equation with n independent variables $x_1, \dots, x_n$ that has the form

$$\sum_{i,j=1}^{n} a_{ij}(\mathbf{x}) \frac{\partial^2 w}{\partial x_i \partial x_j} = F\left(\mathbf{x}, w, \frac{\partial w}{\partial x_1}, \dots, \frac{\partial w}{\partial x_n}\right), \tag{M12.2.2.1}$$

where the a_{ij} are some functions that have continuous derivatives with respect to all variables to the second order inclusive, and $\mathbf{x} = \{x_1, \dots, x_n\}$. [The right-hand side of equation (M12.2.2.1) may be nonlinear. The left-hand side only is required for the classification of this equation.]

At a point $\mathbf{x} = \mathbf{x}_0$, the following quadratic form is assigned to equation (M12.2.2.1):

$$Q = \sum_{i,j=1}^{n} a_{ij}(\mathbf{x}_0) \xi_i \xi_j. \tag{M12.2.2.2}$$

By an appropriate linear nondegenerate transformation

$$\xi_i = \sum_{k=1}^{n} \beta_{ik} \eta_k \qquad (i = 1, \dots, n) \tag{M12.2.2.3}$$

the quadratic form (M12.2.2.2) can be reduced to the canonical form

$$Q = \sum_{i=1}^{n} c_i \eta_i^2, \tag{M12.2.2.4}$$

where the coefficients c_i assume the values 1, –1, and 0. The number of negative and zero coefficients in (M12.2.2.4) does not depend on the way in which the quadratic form is reduced to the canonical form.

TABLE M12.1
Classification of equations with many independent variables.

Type of equation (M12.2.2.1) at a point $\mathbf{x} = \mathbf{x}_0$	Coefficients of the canonical form (M12.2.2.4)
Parabolic (in the broad sense)	At least one coefficient of the c_i is zero
Hyperbolic (in the broad sense)	All c_i are nonzero and some c_i differ in sign
Elliptic	All c_i are nonzero and have like signs

Table M12.1 presents the basic criteria according to which the equations with many independent variables are classified.

Suppose all coefficients of the highest derivatives in (M12.2.2.1) are constant, $a_{ij} = \text{const}$. By introducing the new independent variables $y_1, \ldots, y_n$ in accordance with the formulas $y_i = \sum_{k=1}^{n} \beta_{ik} x_k$, where the β_{ik} are the coefficients of the linear transformation (M12.2.2.3), we reduce equation (M12.2.2.1) to the canonical form

$$\sum_{i=1}^{n} c_i \frac{\partial^2 w}{\partial y_i^2} = F_1\left(\mathbf{y}, w, \frac{\partial w}{\partial y_1}, \ldots, \frac{\partial w}{\partial y_n}\right). \tag{M12.2.2.5}$$

The coefficients c_i here are the same as in the quadratic form (M12.2.2.4), and $\mathbf{y} = \{y_1, \ldots, y_n\}$.

Specific equations of parabolic, hyperbolic, and elliptic types will be discussed further in Sections M12.3, M12.5, and M12.7–M12.9.

M12.3. Basic Problems for Linear Equations of Mathematical Physics

M12.3.1. Initial and Boundary Conditions. Cauchy Problem. Boundary Value Problems

Every equation of mathematical physics governs infinitely many qualitatively similar phenomena or processes. This follows from the fact that differential equations have infinitely many particular solutions. The specific solution that describes the physical phenomenon under study is isolated from the set of particular solutions of the given differential equation by means of initial and boundary conditions.

Throughout this section, we consider linear equations in the n-dimensional Euclidean space $\mathbb{R}^n$ or in an open domain $V \in \mathbb{R}^n$ (exclusive of the boundary) with a sufficiently smooth boundary $S = \partial V$.

► **Parabolic equations. Initial and boundary conditions.** In general, a linear second-order partial differential equation of the parabolic type with n independent variables can be written as

$$\frac{\partial w}{\partial t} - L_{\mathbf{x},t}[w] = \Phi(\mathbf{x}, t), \tag{M12.3.1.1}$$

where

$$L_{\mathbf{x},t}[w] \equiv \sum_{i,j=1}^{n} a_{ij}(\mathbf{x}, t) \frac{\partial^2 w}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(\mathbf{x}, t) \frac{\partial w}{\partial x_i} + c(\mathbf{x}, t) w, \tag{M12.3.1.2}$$

$$\mathbf{x} = \{x_1, \dots, x_n\}, \quad \sum_{i,j=1}^{n} a_{ij}(\mathbf{x}, t)\xi_i\xi_j \geq \sigma \sum_{i=1}^{n} \xi_i^2, \quad \sigma > 0.$$

Parabolic equations govern unsteady thermal, diffusion, and other phenomena dependent on time t.

Equation (M12.3.1.1) is called homogeneous if $\Phi(\mathbf{x}, t) \equiv 0$.

Cauchy problem ($t \geq 0$, $\mathbf{x} \in \mathbb{R}^n$). Find a function w that satisfies equation (M12.3.1.1) for $t > 0$ and the initial condition

$$w = f(\mathbf{x}) \quad \text{at} \quad t = 0. \tag{M12.3.1.3}$$

Example 1. The solution to the Cauchy problem for the one-dimensional heat equation (M12.2.1.1) with the initial condition (M12.3.1.3) is given by

$$w(x, t) = \frac{1}{2\sqrt{\pi t}} \int_{-\infty}^{\infty} f(\xi) \exp\left[-\frac{(x - \xi)^2}{4t}\right] d\xi.$$

The derivation of this formula is shown in the example of Subsection M12.6.2.

*Boundary value problem** ($t \geq 0$, $\mathbf{x} \in V$). Find a function w that satisfies equation (M12.3.1.1) for $t > 0$, the initial condition (M12.3.1.3), and the boundary condition

$$\Gamma_{\mathbf{x},t}[w] = g(\mathbf{x}, t) \quad \text{at} \quad \mathbf{x} \in S \qquad (t > 0). \tag{M12.3.1.4}$$

In general, $\Gamma_{\mathbf{x},t}$ is a first-order linear differential operator in the space variables $\mathbf{x}$ with coefficients dependent on $\mathbf{x}$ and t. The basic types of boundary conditions are described in Subsection M12.3.2.

The initial condition (M12.3.1.3) is called homogeneous if $f(\mathbf{x}) \equiv 0$. The boundary condition (M12.3.1.4) is called homogeneous if $g(\mathbf{x}, t) \equiv 0$.

► **Hyperbolic equations. Initial and boundary conditions.** Consider a second-order linear partial differential equation of the hyperbolic type with n independent variables of the general form

$$\frac{\partial^2 w}{\partial t^2} + \varphi(\mathbf{x}, t)\frac{\partial w}{\partial t} - L_{\mathbf{x},t}[w] = \Phi(\mathbf{x}, t), \tag{M12.3.1.5}$$

where the linear differential operator $L_{\mathbf{x},t}$ is defined by (M12.3.1.2). Hyperbolic equations govern unsteady wave processes, which depend on time t.

Equation (M12.3.1.5) is said to be homogeneous if $\Phi(\mathbf{x}, t) \equiv 0$.

Cauchy problem ($t \geq 0$, $\mathbf{x} \in \mathbb{R}^n$). Find a function w that satisfies equation (M12.3.1.5) for $t > 0$ and the initial conditions

$$\begin{aligned} w &= f_0(\mathbf{x}) \quad \text{at} \quad t = 0, \\ \partial_t w &= f_1(\mathbf{x}) \quad \text{at} \quad t = 0. \end{aligned} \tag{M12.3.1.6}$$

Example 2. The solution to the Cauchy problem for the one-dimensional wave equation (M12.2.1.2) with the initial conditions (M12.3.1.6) is given by *d'Alembert's formula*:

$$w(x, t) = \frac{1}{2}[f_0(x + t) + f_0(x - t)] + \frac{1}{2}\int_{x-t}^{x+t} f_1(\xi)\, d\xi.$$

Boundary value problem ($t \geq 0$, $\mathbf{x} \in V$). Find a function w that satisfies equation (M12.3.1.5) for $t > 0$, the initial conditions (M12.3.1.6), and boundary condition (M12.3.1.4).

The initial conditions (M12.3.1.6) are called homogeneous if $f_0(\mathbf{x}) \equiv 0$ and $f_1(\mathbf{x}) \equiv 0$.

Goursat problem. On the characteristics of a hyperbolic equation with two independent variables, the values of the unknown function w are prescribed; for details, see Subsection M12.8.3.

---

* *Boundary value problems* for parabolic and hyperbolic equations are sometimes called *mixed* or *initial-boundary value problems*.

► **Elliptic equations. Boundary conditions.** In general, a second-order linear partial differential equation of elliptic type with n independent variables can be written as

$$-L_{\mathbf{x}}[w] = \Phi(\mathbf{x}), \qquad \text{(M12.3.1.7)}$$

where

$$L_{\mathbf{x}}[w] \equiv \sum_{i,j=1}^{n} a_{ij}(\mathbf{x})\frac{\partial^2 w}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(\mathbf{x})\frac{\partial w}{\partial x_i} + c(\mathbf{x})w, \qquad \text{(M12.3.1.8)}$$

$$\sum_{i,j=1}^{n} a_{ij}(\mathbf{x})\xi_i\xi_j \ge \sigma \sum_{i=1}^{n} \xi_i^2, \quad \sigma > 0.$$

Elliptic equations govern steady-state thermal, diffusion, and other phenomena independent of time t.

Equation (M12.3.1.7) is said to be homogeneous if $\Phi(\mathbf{x}) \equiv 0$.

Boundary value problem. Find a function w that satisfies equation (M12.3.1.7) and the boundary condition

$$\Gamma_{\mathbf{x}}[w] = g(\mathbf{x}) \quad \text{at} \quad \mathbf{x} \in S. \qquad \text{(M12.3.1.9)}$$

In general, $\Gamma_{\mathbf{x}}$ is a first-order linear differential operator in the space variables $\mathbf{x}$. The basic types of boundary conditions are described below in Subsection M12.3.2.

The boundary condition (M12.3.1.9) is called homogeneous if $g(\mathbf{x}) \equiv 0$. The boundary value problem (M12.3.1.7)–(M12.3.1.9) is said to be homogeneous if $\Phi \equiv 0$ and $g \equiv 0$.

M12.3.2. First, Second, Third, and Mixed Boundary Value Problems

For any (parabolic, hyperbolic, and elliptic) second-order partial differential equations, it is conventional to distinguish four basic types of boundary value problems, depending on the form of the boundary conditions (M12.3.1.4) [see also the analogous condition (M12.3.1.9)]. For simplicity, here we confine ourselves to the case where the coefficients a_{ij} of equations (M12.3.1.1) and (M12.3.1.5), with the operator defined by (M12.3.1.2), have the special form

$$a_{ij}(\mathbf{x},t) = a(\mathbf{x},t)\delta_{ij}, \quad \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \ne j. \end{cases}$$

This situation is rather frequent in applications; such coefficients are used to describe various phenomena (processes) in isotropic media.

► **First boundary value problem.** The function $w(\mathbf{x},t)$ takes prescribed values at the boundary S of the domain:

$$w(\mathbf{x},t) = g_1(\mathbf{x},t) \quad \text{for} \quad \mathbf{x} \in S. \qquad \text{(M12.3.2.1)}$$

► **Second boundary value problem.** The derivative along the (outward) normal is prescribed at the boundary S of the domain:

$$\frac{\partial w}{\partial N} = g_2(\mathbf{x},t) \quad \text{for} \quad \mathbf{x} \in S. \qquad \text{(M12.3.2.2)}$$

In heat transfer problems, where w is temperature, the left-hand side of the boundary condition (M12.3.2.2) is proportional to the heat flux per unit area of the surface S.

TABLE M12.2
Boundary conditions for various boundary value problems specified by parabolic and hyperbolic equations in two independent variables ($x_1 \le x \le x_2$).

Type of problem	Boundary condition at $x = x_1$	Boundary condition at $x = x_2$
First boundary value problem	$w = g_1(t)$	$w = g_2(t)$
Second boundary value problem	$\partial_x w = g_1(t)$	$\partial_x w = g_2(t)$
Third boundary value problem	$\partial_x w + \beta_1 w = g_1(t)$ $(\beta_1 < 0)$	$\partial_x w + \beta_2 w = g_2(t)$ $(\beta_2 > 0)$
Mixed boundary value problem	$w = g_1(t)$	$\partial_x w = g_2(t)$
Mixed boundary value problem	$\partial_x w = g_1(t)$	$w = g_2(t)$

► **Third boundary value problem.** A linear relationship between the unknown function and its normal derivative is prescribed at the boundary S of the domain:

$$\frac{\partial w}{\partial N} + k(\mathbf{x}, t)w = g_3(\mathbf{x}, t) \quad \text{for} \quad \mathbf{x} \in S. \qquad \text{(M12.3.2.3)}$$

It is usually assumed that $k(\mathbf{x}, t) = \text{const}$. In mass transfer problems, where w is concentration, the boundary condition (M12.3.2.3) with $g_3 \equiv 0$ describes a surface chemical reaction of the first order.

► **Mixed boundary value problems.** Conditions of various types, listed above, are set at different portions of the boundary S.

If $g_1 \equiv 0$, $g_2 \equiv 0$, or $g_3 \equiv 0$, the respective boundary conditions (M12.3.2.1), (M12.3.2.2), and (M12.3.2.3) are said to be homogeneous.

Boundary conditions for various boundary value problems for parabolic and hyperbolic equations in two independent variables x and t are displayed in Table M12.2. The equation coefficients are assumed to be continuous, with the coefficients of the highest derivatives being nonzero in the range $x_1 \le x \le x_2$ considered.

Remark 1. For elliptic equations, the coefficients a_{ij} as well as the right-hand sides of the boundary conditions (M12.3.2.1)–(M12.3.2.3) are independent of time t.

Remark 2. For elliptic equations, the first boundary value problem is often called the *Dirichlet problem*, and the second boundary value problem is called the *Neumann problem*.

M12.4. Properties and Exact Solutions of Linear Equations

M12.4.1. Homogeneous Linear Equations and Their Particular Solutions

► **Preliminary remarks.** For brevity, in this paragraph a homogeneous linear partial differential equation will be written as

$$\mathcal{L}[w] = 0. \qquad \text{(M12.4.1.1)}$$

For second-order linear parabolic and hyperbolic equations, the linear differential operator $\mathcal{L}[w]$ is defined by the left-hand side of equations (M12.3.1.1) and (M12.3.1.5), respectively. It is assumed that equation (M12.4.1.1) is an arbitrary homogeneous linear partial differential equation of any order in the variables t, x_1, …, x_n with sufficiently smooth coefficients.

A linear differential operator $\mathcal{L}$ possesses the properties

$$\begin{aligned}&\mathcal{L}[w_1+w_2]=\mathcal{L}[w_1]+\mathcal{L}[w_2],\\&\mathcal{L}[Aw]=A\mathcal{L}[w],\qquad A=\text{const}.\end{aligned}$$

An arbitrary homogeneous linear equation (M12.4.1.1) has a trivial solution, $w\equiv 0$.

A function w is called a *classical solution* of equation (M12.4.1.1) if w, when substituted into (M12.4.1.1), turns the equation into an identity and if all partial derivatives of w that occur in (M12.4.1.1) are continuous; the notion of a classical solution is directly linked to the range of the independent variables. In what follows, we usually write “solution” instead of “classical solution” for brevity.

► **Use of particular solutions for the construction of other solutions.** Below are some properties of particular solutions of homogeneous linear equations.

1°. Let $w_1=w_1(\mathbf{x},t)$, $w_2=w_2(\mathbf{x},t)$, …, $w_k=w_k(\mathbf{x},t)$ be any particular solutions of the homogeneous equation (M12.4.1.1). Then the linear combination

$$w=A_1w_1+A_2w_2+\cdots+A_kw_k \qquad \text{(M12.4.1.2)}$$

with arbitrary constants A_1, A_2, …, A_k is also a solution of equation (M12.4.1.1); in physics, this property is known as the *principle of linear superposition*.

Suppose $\{w_k\}$ is an infinite sequence of solutions of equation (M12.4.1.1). Then the series $\sum\limits_{k=1}^{\infty} w_k$, irrespective of its convergence, is called a *formal solution* of (M12.4.1.1). If the solutions w_k are classical, the series is uniformly convergent, and the sum of the series has all the necessary partial derivatives, then the sum of the series is a classical solution of equation (M12.4.1.1).

2°. Let the coefficients of the linear differential operator $\mathcal{L}$ be independent of time t. If equation (M12.4.1.1) has a particular solution $\widetilde{w}=\widetilde{w}(\mathbf{x},t)$, then the partial derivatives of $\widetilde{w}$ with respect to time,*

$$\frac{\partial\widetilde{w}}{\partial t},\quad \frac{\partial^2\widetilde{w}}{\partial t^2},\quad \dots,\quad \frac{\partial^k\widetilde{w}}{\partial t^k},\quad \dots,$$

are also solutions of equation (M12.4.1.1).

3°. Let the coefficients of the linear differential operator $\mathcal{L}$ be independent of the space variables $x_1,\dots,x_n$. If equation (M12.4.1.1) has a particular solution $\widetilde{w}=\widetilde{w}(\mathbf{x},t)$, then the partial derivatives of $\widetilde{w}$ with respect to the space coordinates

$$\frac{\partial\widetilde{w}}{\partial x_1},\quad \frac{\partial\widetilde{w}}{\partial x_2},\quad \frac{\partial\widetilde{w}}{\partial x_3},\quad \dots,\quad \frac{\partial^2\widetilde{w}}{\partial x_1^2},\quad \frac{\partial^2\widetilde{w}}{\partial x_1\partial x_2},\quad \dots,\quad \frac{\partial^{k+m}\widetilde{w}}{\partial x_2^k\partial x_3^m},\quad \dots$$

are also solutions of equation (M12.4.1.1).

If the coefficients of $\mathcal{L}$ are independent of only one space coordinate, say x_1, and equation (M12.4.1.1) has a particular solution $\widetilde{w}=\widetilde{w}(\mathbf{x},t)$, then the partial derivatives

$$\frac{\partial\widetilde{w}}{\partial x_1},\quad \frac{\partial^2\widetilde{w}}{\partial x_1^2},\quad \dots,\quad \frac{\partial^k\widetilde{w}}{\partial x_1^k},\quad \dots$$

are also solutions of equation (M12.4.1.1).

---

* Here and in what follows, it is assumed that the particular solution $\widetilde{w}$ is differentiable sufficiently many times with respect to t and $x_1,\dots,x_n$ (or the parameters).

4°. Let the coefficients of the linear differential operator $\mathcal{L}$ be constant and let equation (M12.4.1.1) have a particular solution $\widetilde{w} = \widetilde{w}(\mathbf{x}, t)$. Then any partial derivatives of $\widetilde{w}$ with respect to time and the space coordinates (including mixed derivatives)

$$\frac{\partial \widetilde{w}}{\partial t}, \quad \frac{\partial \widetilde{w}}{\partial x_1}, \quad \dots, \quad \frac{\partial^2 \widetilde{w}}{\partial x_2^2}, \quad \frac{\partial^2 \widetilde{w}}{\partial t \partial x_1}, \quad \dots, \quad \frac{\partial^k \widetilde{w}}{\partial x_3^k}, \quad \dots$$

are solutions of equation (M12.4.1.1).

5°. Suppose equation (M12.4.1.1) has a particular solution dependent on a parameter μ, $\widetilde{w} = \widetilde{w}(\mathbf{x}, t; \mu)$, and the coefficients of the linear differential operator $\mathcal{L}$ are independent of μ (but can depend on time and the space coordinates). Then, by differentiating $\widetilde{w}$ with respect to μ, one obtains other solutions of equation (M12.4.1.1),

$$\frac{\partial \widetilde{w}}{\partial \mu}, \quad \frac{\partial^2 \widetilde{w}}{\partial \mu^2}, \quad \dots, \quad \frac{\partial^k \widetilde{w}}{\partial \mu^k}, \quad \dots$$

Example 1. A linear *heat equation with a source* of the form

$$\frac{\partial w}{\partial t} = a\frac{\partial^2 w}{\partial x^2} + bw$$

has a particular solution

$$w(x, t) = \exp[\mu x + (a\mu^2 + b)t],$$

where μ is an arbitrary constant. Differentiating this solution with respect to μ yields another solution

$$w(x, t) = (x + 2a\mu t)\exp[\mu x + (a\mu^2 + b)t].$$

Let some constants $\mu_1, \dots, \mu_k$ belong to the range of the parameter μ. Then the sum

$$w = A_1\widetilde{w}(\mathbf{x}, t; \mu_1) + \dots + A_k\widetilde{w}(\mathbf{x}, t; \mu_k), \qquad \text{(M12.4.1.3)}$$

where $A_1, \dots, A_k$ are arbitrary constants, is also a solution of the homogeneous linear equation (M12.4.1.1). The number of terms in the sum (M12.4.1.3) can be both finite and infinite.

6°. Another effective way of constructing solutions involves the following considerations. The particular solution $\widetilde{w}(\mathbf{x}, t; \mu)$, which depends on the parameter μ (as before, it is assumed that the coefficients of the linear differential operator $\mathcal{L}$ are independent of μ), is first multiplied by an arbitrary function $\varphi(\mu)$. Then the resulting expression is integrated with respect to μ over some interval $[\alpha, \beta]$. Thus, one obtains a new function,

$$\int_\alpha^\beta \widetilde{w}(\mathbf{x}, t; \mu)\varphi(\mu)\, d\mu,$$

which is also a solution of the original homogeneous linear equation.

The properties listed in Items 1°–6° enable one to use known particular solutions to construct other particular solutions of homogeneous linear equations of mathematical physics.

► **Multiplicative separable solutions.** Many homogeneous linear partial differential equations have solutions that can be represented as the product of functions depending on different arguments. Such solutions are referred to as *multiplicative separable solutions*; very commonly these solutions are briefly, but less accurately, called just *separable solutions*.

TABLE M12.3
Homogeneous linear partial differential equations that admit multiplicative separable solutions.

No.	Form of equation (M12.4.1.1)	Form of particular solutions
1	Equation coefficients are constant	$w(\mathbf{x},t)=A\exp(\lambda t+\beta_1x_1+\cdots+\beta_nx_n)$, $\lambda,\beta_1,\dots,\beta_n$ are related by an algebraic equation
2	Equation coefficients are independent of time t	$w(\mathbf{x},t)=e^{\lambda t}\psi(\mathbf{x})$, λ is an arbitrary constant, $\mathbf{x}=\{x_1,\dots,x_n\}$
3	Equation coefficients are independent of the coordinates $x_1,\dots,x_n$	$w(\mathbf{x},t)=\exp(\beta_1x_1+\cdots+\beta_nx_n)\psi(t)$, $\beta_1,\dots,\beta_n$ are arbitrary constants
4	Equation coefficients are independent of the coordinates $x_1,\dots,x_k$	$w(\mathbf{x},t)=\exp(\beta_1x_1+\cdots+\beta_kx_k)\psi(t,x_{k+1},\dots,x_n)$, $\beta_1,\dots,\beta_k$ are arbitrary constants
5	$L_t[w]+L_{\mathbf{x}}[w]=0$, operator L_t depends on only t, operator $L_{\mathbf{x}}$ depends on only $\mathbf{x}$	$w(\mathbf{x},t)=\varphi(t)\psi(\mathbf{x})$, $\varphi(t)$ satisfies the equation $L_t[\varphi]+\lambda\varphi=0$, $\psi(\mathbf{x})$ satisfies the equation $L_{\mathbf{x}}[\psi]-\lambda\psi=0$
6	$L_t[w]+L_1[w]+\cdots+L_n[w]=0$, operator L_t depends on only t, operator L_k depends on only x_k	$w(\mathbf{x},t)=\varphi(t)\psi_1(x_1)\dots\psi_n(x_n)$, $\varphi(t)$ satisfies the equation $L_t[\varphi]+\lambda\varphi=0$, $\psi_k(x_k)$ satisfies the equation $L_k[\psi_k]+\beta_k\psi_k=0$, $\lambda+\beta_1+\cdots+\beta_n=0$
7	$f_0(x_1)L_t[w]+\sum\limits_{k=1}^{n}f_k(x_1)L_k[w]=0$, operator L_t depends on only t, operator L_k depends on only x_k	$w(\mathbf{x},t)=\varphi(t)\psi_1(x_1)\dots\psi_n(x_n)$, $L_t[\varphi]+\lambda\varphi=0$, $L_k[\psi_k]+\beta_k\psi_k=0,\quad k=2,\dots,n$, $f_1(x_1)L_1[\psi_1]-\Big[\lambda f_0(x_1)+\sum\limits_{k=2}^{n}\beta_kf_k(x_1)\Big]\psi_1=0$

Table M12.3 presents the most commonly encountered types of homogeneous linear differential equations with many independent variables that admit exact separable solutions. Linear combinations of particular solutions that correspond to different values of the separation parameters, λ, β_1, …, β_n, are also solutions of the equations in question. For brevity, the word "operator" is used below to denote "linear differential operator."

For a constant coefficient equation (see the first row in Table M12.3), the separation parameters must satisfy the algebraic equation

$$D(\lambda,\,\beta_1,\,\dots,\beta_n)=0, \qquad \text{(M12.4.1.4)}$$

which results from substituting the solution into equation (M12.4.1.1). In physical applications, equation (M12.4.1.4) is usually referred to as a *dispersion equation*. Any n of the $n+1$ separation parameters in (M12.4.1.4) can be treated as arbitrary.

Example 2. Consider the *telegraph equation*

$$\frac{\partial^2w}{\partial t^2}+k\frac{\partial w}{\partial t}=a^2\frac{\partial^2w}{\partial x^2}+b\frac{\partial w}{\partial x}+cw.$$

A particular solution is sought in the form

$$w=A\exp(\beta x+\lambda t).$$

This results in the dispersion equation $\lambda^2+k\lambda=a^2\beta^2+b\beta+c$, where one of the two parameters β or λ can be treated as arbitrary.

Remark 1. For stationary equations that do not depend on t, one should set $\lambda=0$, $L_t[w]\equiv0$, and $\varphi(t)\equiv1$ in rows 1, 6, and 7 of Table M12.3.

Remark 2. Multiplicative separable solutions play an important role in the theory of linear partial differential equations; they are used for finding solutions to stationary and nonstationary boundary value problems; see Sections M12.5 and M12.7–M12.9.

► Solutions of some linear equations encountered in applications.

1°. Some particular solutions of the heat equation

$$\frac{\partial w}{\partial t} = a\frac{\partial^2 w}{\partial x^2}$$

are listed below:

$$\begin{aligned}
w(x) &= Ax + B,\\
w(x,t) &= A(x^2 + 2at) + B,\\
w(x,t) &= A(x^3 + 6atx) + B,\\
w(x,t) &= A(x^4 + 12atx^2 + 12a^2t^2) + B,\\
w(x,t) &= x^{2n} + \sum_{k=1}^{n} \frac{(2n)(2n-1)\dots(2n-2k+1)}{k!}(at)^k x^{2n-2k},\\
w(x,t) &= x^{2n+1} + \sum_{k=1}^{n} \frac{(2n+1)(2n)\dots(2n-2k+2)}{k!}(at)^k x^{2n-2k+1},\\
w(x,t) &= A\exp(a\mu^2 t \pm \mu x) + B,\\
w(x,t) &= A\frac{1}{\sqrt{t}}\exp\left(-\frac{x^2}{4at}\right) + B,\\
w(x,t) &= A\frac{x}{t^{3/2}}\exp\left(-\frac{x^2}{4at}\right) + B,\\
w(x,t) &= A\exp(-a\mu^2 t)\cos(\mu x + B) + C,\\
w(x,t) &= A\exp(-\mu x)\cos(\mu x - 2a\mu^2 t + B) + C,\\
w(x,t) &= A\,\mathrm{erf}\left(\frac{x}{2\sqrt{at}}\right) + B,
\end{aligned}$$

where A, B, C, and μ are arbitrary constants, n is a positive integer, and $\mathrm{erf}\, z \equiv \frac{2}{\sqrt{\pi}}\int_0^z \exp(-\xi^2)\,d\xi$ is the error function (probability integral).

2°. The general solution of the wave equation

$$\frac{\partial^2 w}{\partial t^2} = a^2\frac{\partial^2 w}{\partial x^2}$$

is given by

$$w(x,t) = \varphi(x + at) + \psi(x - at),$$

where $\varphi(z_1)$ and $\psi(z_2)$ are arbitrary functions.

3°. Particular solutions of Laplace equation (M12.2.1.3) in the Cartesian coordinate system:

$$\begin{aligned}
w(x,y) &= Ax + By + C,\\
w(x,y) &= A(x^2 - y^2) + Bxy,
\end{aligned}$$

$$
\begin{aligned}
w(x,y) &= A(x^3-3xy^2)+B(3x^2y-y^3),\\
w(x,y) &= \frac{Ax+By}{x^2+y^2}+C,\\
w(x,y) &= \exp(\pm\mu x)(A\cos\mu y+B\sin\mu y),\\
w(x,y) &= (A\cos\mu x+B\sin\mu x)\exp(\pm\mu y),\\
w(x,y) &= (A\sinh\mu x+B\cosh\mu x)(C\cos\mu y+D\sin\mu y),\\
w(x,y) &= (A\cos\mu x+B\sin\mu x)(C\sinh\mu y+D\cosh\mu y),
\end{aligned}
$$

where A, B, C, D, and μ are arbitrary constants.

Particular solutions of Laplace equation (M12.2.1.3) in the polar coordinate system:

$$
\begin{aligned}
w(r) &= A\ln r+B,\\
w(r,\varphi) &= \left(Ar^m+\frac{B}{r^m}\right)(C\cos m\varphi+D\sin m\varphi),
\end{aligned}
$$

where A, B, C, and D are arbitrary constants, and $m = 1, 2, \ldots$

M12.4.2. Nonhomogeneous Linear Equations and Their Particular Solutions

► **Simplest properties of nonhomogeneous linear equations.** For brevity, we write a nonhomogeneous linear partial differential equation in the form

$$\mathcal{L}[w] = \Phi(\mathbf{x},t), \qquad \text{(M12.4.2.1)}$$

where the linear differential operator $\mathcal{L}$ is defined above (see the beginning of Subsection M12.4.1).

Below are the simplest properties of particular solutions of the nonhomogeneous equation (M12.4.2.1).

1°. If $\widetilde{w}_\Phi(\mathbf{x},t)$ is a particular solution of the nonhomogeneous equation (M12.4.2.1) and $\widetilde{w}_0(\mathbf{x},t)$ is a particular solution of the corresponding homogeneous equation (M12.4.1.1), then the sum

$$A\widetilde{w}_0(\mathbf{x},t)+\widetilde{w}_\Phi(\mathbf{x},t),$$

with arbitrary constant A, is also a solution of the nonhomogeneous equation (M12.4.2.1). The following, more general statement holds: The general solution of the nonhomogeneous equation (M12.4.2.1) is the sum of the general solution of the corresponding homogeneous equation (M12.4.1.1) and any particular solution of the nonhomogeneous equation (M12.4.2.1).

2°. Suppose w_1 and w_2 are solutions of nonhomogeneous linear equations with the same left-hand side and different right-hand sides, i.e.,

$$\mathcal{L}[w_1] = \Phi_1(\mathbf{x},t), \qquad \mathcal{L}[w_2] = \Phi_2(\mathbf{x},t).$$

Then the function $w = w_1 + w_2$ is a solution of the equation

$$\mathcal{L}[w] = \Phi_1(\mathbf{x},t)+\Phi_2(\mathbf{x},t).$$

► **Fundamental and particular solutions of stationary equations.** Consider the second-order linear stationary (time-independent) nonhomogeneous equation

$$L_{\mathbf{x}}[w] = -\Phi(\mathbf{x}). \tag{M12.4.2.2}$$

Here, $L_{\mathbf{x}}$ is a linear differential operator of the second (or any) order of general form whose coefficients are dependent on $\mathbf{x}$, where $\mathbf{x} \in \mathbb{R}^n$.

A function $\mathscr{E} = \mathscr{E}(\mathbf{x}, \mathbf{y})$ that satisfies the equation with a special right-hand side

$$L_{\mathbf{x}}[\mathscr{E}] = -\delta(\mathbf{x} - \mathbf{y}) \tag{M12.4.2.3}$$

is called a *fundamental solution* corresponding to the operator $L_{\mathbf{x}}$. In (M12.4.2.3), $\delta(\mathbf{x})$ is an n-dimensional Dirac delta function and the vector quantity $\mathbf{y} = \{y_1, \dots, y_n\}$ appears in equation (M12.4.2.3) as an n-dimensional free parameter. It is assumed that $\mathbf{y} \in \mathbb{R}^n$.

The n-dimensional Dirac delta function possesses the following basic properties:

1. $\delta(\mathbf{x}) = \delta(x_1)\delta(x_2)\dots\delta(x_n)$,
2. $\displaystyle\int_{\mathbb{R}^n} \Phi(\mathbf{y})\delta(\mathbf{x} - \mathbf{y})\,d\mathbf{y} = \Phi(\mathbf{x})$,

where $\delta(x_k)$ is the one-dimensional Dirac delta function, $\Phi(\mathbf{x})$ is an arbitrary continuous function, and $d\mathbf{y} = dy_1 \dots dy_n$.

For constant coefficient equations, a fundamental solution always exists.

The fundamental solution $\mathscr{E} = \mathscr{E}(\mathbf{x}, \mathbf{y})$ can be used to construct a particular solution of the linear stationary nonhomogeneous equation (M12.4.2.2) for arbitrary continuous $\Phi(\mathbf{x})$; this particular solution is expressed as follows:

$$w(\mathbf{x}) = \int_{\mathbb{R}^n} \Phi(\mathbf{y})\mathscr{E}(\mathbf{x}, \mathbf{y})\,d\mathbf{y}. \tag{M12.4.2.4}$$

Remark 1. The fundamental solution $\mathscr{E}$ is not unique; it is defined up to an additive term $w_0 = w_0(\mathbf{x})$, which is an arbitrary solution of the homogeneous equation $L_{\mathbf{x}}[w_0] = 0$.

Remark 2. For constant coefficient differential equations, the fundamental solution possesses the property $\mathscr{E}(\mathbf{x}, \mathbf{y}) = \mathscr{E}(\mathbf{x} - \mathbf{y})$.

Remark 3. The right-hand sides of equations (M12.4.2.2) and (M12.4.2.3) are often prefixed with the plus sign. In this case, formula (M12.4.2.4) remains valid.

Example. For the two- and three-dimensional *Poisson equations*, the fundamental solutions have the forms

Equations		*Fundamental solutions*
$\dfrac{\partial^2 w}{\partial x_1^2} + \dfrac{\partial^2 w}{\partial x_2^2} = -\Phi(x_1, x_2)$	$\Longrightarrow$	$\mathscr{E}(x_1, x_2, y_1, y_2) = \dfrac{1}{2\pi} \ln \dfrac{1}{\rho}$,
$\dfrac{\partial^2 w}{\partial x_1^2} + \dfrac{\partial^2 w}{\partial x_2^2} + \dfrac{\partial^2 w}{\partial x_3^2} = -\Phi(x_1, x_2, x_3)$	$\Longrightarrow$	$\mathscr{E}(x_1, x_2, x_3, y_1, y_2, y_3) = \dfrac{1}{4\pi r}$,

where $\rho = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$ and $r = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}$.

M12.5. Method of Separation of Variables (Fourier Method)

M12.5.1. Solution of Problems for Parabolic Equations

► **Mathematical statement of the problem.** Consider boundary value problems for a linear homogeneous parabolic equation of the form

$$\frac{\partial w}{\partial t} = a(x)\frac{\partial^2 w}{\partial x^2} + b(x)\frac{\partial w}{\partial x} + \big[c(x) + \gamma(t)\big]w \tag{M12.5.1.1}$$

with homogeneous linear boundary conditions,

$$\begin{aligned} s_1\partial_x w + k_1 w = 0 \quad &\text{at} \quad x = x_1,\\ s_2\partial_x w + k_2 w = 0 \quad &\text{at} \quad x = x_2, \end{aligned} \qquad \text{(M12.5.1.2)}$$

and an arbitrary initial condition,

$$w = f_0(x) \quad \text{at} \quad t = 0. \qquad \text{(M12.5.1.3)}$$

We assume that the coefficients of equation (M12.5.1.1) and boundary conditions (M12.5.1.2) meet the following requirements:

$a(x)$, $b(x)$, $c(x)$, and $\gamma(t)$ are continuous functions; $a(x) > 0$, $|s_1| + |k_1| > 0$, $|s_2| + |k_2| > 0$.

Remark. In various applications, equations of the form (M12.5.1.1) may arise with the coefficient $b(x)$ going to infinity at the boundary, $b(x) \to \infty$ as $x \to x_1$, with the other coefficients being continuous. In this case, the first boundary condition in (M12.5.1.2) should be replaced with a condition of boundedness of the solution as $x \to x_1$. This may occur in spatial problems with central or axial symmetry where the solution depends only on the radial coordinate.

► **Derivation of equations and boundary conditions for particular solutions.** The method of separation of variables is based on searching for particular solutions of equation (M12.5.1.1) in the product form

$$w(x,t) = \varphi(x)\,\psi(t). \qquad \text{(M12.5.1.4)}$$

Let us substitute (M12.5.1.4) into (M12.5.1.1) and divide all terms of the equation by $\varphi(x)\,\psi(t)$. Then we move the terms dependent on t to the left-hand side and those dependent on x to the right-hand side to obtain

$$\frac{\psi'_t}{\psi} - \gamma(t) = \frac{a(x)\varphi''_{xx} + b(x)\varphi'_x + c(x)\varphi}{\varphi}$$

This equality is achieved if both sides are equal to the same constant quantity, $-\lambda$. The free parameter λ is called the *separation constant*. Thus, we arrive at the following linear ordinary differential equations for $\varphi = \varphi(x)$ and $\psi = \psi(t)$:

$$a(x)\varphi''_{xx} + b(x)\varphi'_x + [\lambda + c(x)]\varphi = 0, \qquad \text{(M12.5.1.5)}$$

$$\psi'_t + [\lambda - \gamma(t)]\psi = 0. \qquad \text{(M12.5.1.6)}$$

Substituting (M12.5.1.4) into (M12.5.1.2) yields the boundary conditions for $\varphi = \varphi(x)$:

$$\begin{aligned} s_1\varphi'_x + k_1\varphi = 0 \quad &\text{at} \quad x = x_1,\\ s_2\varphi'_x + k_2\varphi = 0 \quad &\text{at} \quad x = x_2. \end{aligned} \qquad \text{(M12.5.1.7)}$$

The homogeneous linear ordinary differential equation (M12.5.1.5) in conjunction with the homogeneous linear boundary conditions (M12.5.1.7) makes up an eigenvalue problem.

► **Solution of eigenvalue problems. Orthogonality of eigenfunctions.** Let $\widetilde{\varphi}_1(x,\lambda)$ and $\widetilde{\varphi}_2(x,\lambda)$ be linearly independent particular solutions of equation (M12.5.1.5). Then the general solution of this equation can be represented as the linear combination

$$\varphi = C_1\widetilde{\varphi}_1(x,\lambda) + C_2\widetilde{\varphi}_2(x,\lambda), \qquad \text{(M12.5.1.8)}$$

where C_1 and C_2 are arbitrary constants.

Substituting solution (M12.5.1.8) into the boundary conditions (M12.5.1.7) yields the following homogeneous linear algebraic system of equations for C_1 and C_2:

$$\begin{aligned}\sigma_{11}(\lambda)C_1+\sigma_{12}(\lambda)C_2&=0,\\ \sigma_{21}(\lambda)C_1+\sigma_{22}(\lambda)C_2&=0,\end{aligned}\qquad\text{(M12.5.1.9)}$$

where $\sigma_{ij}(\lambda)=\bigl[s_i(\widetilde{\varphi}_j)'_x+k_i\widetilde{\varphi}_j\bigr]_{x=x_i}$. For system (M12.5.1.9) to have nontrivial solutions, its determinant must be zero; we have

$$\sigma_{11}(\lambda)\sigma_{22}(\lambda)-\sigma_{12}(\lambda)\sigma_{21}(\lambda)=0.\qquad\text{(M12.5.1.10)}$$

Solving the transcendental equation (M12.5.1.10) for λ, one obtains the *eigenvalues* $\lambda=\lambda_n$, where $n=1,\ 2,\ \dots$ For these values of λ, equation (M12.5.1.5) has nontrivial solutions,

$$\varphi_n(x)=\sigma_{12}(\lambda_n)\widetilde{\varphi}_1(x,\lambda_n)-\sigma_{11}(\lambda_n)\widetilde{\varphi}_2(x,\lambda_n),\qquad\text{(M12.5.1.11)}$$

which are called *eigenfunctions* (these functions are defined up to a constant multiplier).

To facilitate the further analysis, we represent equation (M12.5.1.5) in the form

$$[p(x)\varphi'_x]'_x+[\lambda\rho(x)-q(x)]\varphi=0,\qquad\text{(M12.5.1.12)}$$

where

$$p(x)=\exp\left[\int\frac{b(x)}{a(x)}\,dx\right],\quad q(x)=-\frac{c(x)}{a(x)}\exp\left[\int\frac{b(x)}{a(x)}\,dx\right],\quad \rho(x)=\frac{1}{a(x)}\exp\left[\int\frac{b(x)}{a(x)}\,dx\right].\qquad\text{(M12.5.1.13)}$$

It follows from the adopted assumptions that $p(x)$, $p'_x(x)$, $q(x)$, and $\rho(x)$ are continuous functions, with $p(x)>0$ and $\rho(x)>0$.

The eigenvalue problem for equation (M12.5.1.12) subject to the boundary conditions (M12.5.1.7) is known to possess the following properties:

1. All eigenvalues $\lambda_1,\ \lambda_2,\ \dots$ are real, and $\lambda_n\to\infty$ as $n\to\infty$; consequently, the number of negative eigenvalues is finite. Each eigenvalue is of multiplicity 1.

2. The system of eigenfunctions $\varphi_1(x),\ \varphi_2(x),\ \dots$ is orthogonal on the interval $x_1\le x\le x_2$ with weight $\rho(x)$, i.e.,

$$\int_{x_1}^{x_2}\rho(x)\varphi_n(x)\varphi_m(x)\,dx=0\quad\text{for}\quad n\neq m.\qquad\text{(M12.5.1.14)}$$

3. If

$$q(x)\ge0,\quad s_1k_1\le0,\quad s_2k_2\ge0,\qquad\text{(M12.5.1.15)}$$

there are no negative eigenvalues. If $q\equiv0$ and $k_1=k_2=0$, the least eigenvalue is $\lambda_1=0$ and the corresponding eigenfunction is $\varphi_1=\text{const}$. Otherwise, all eigenvalues are positive, provided that conditions (M12.5.1.15) are satisfied; the first inequality in (M12.5.1.15) is satisfied if $c(x)\le0$.

Subsection M11.2.4 presents some estimates for the eigenvalues λ_n and eigenfunctions $\varphi_n(x)$.

► **Series solutions of boundary value problems for parabolic equations.** The solution of equation (M12.5.1.1) corresponding to the eigenvalues $\lambda = \lambda_n$ and satisfying the normalizing conditions $\psi_n(0) = 1$ has the form

$$\psi_n(t) = \exp\left[-\lambda_n t + \int_0^t \gamma(\xi)\,d\xi\right]. \qquad \text{(M12.5.1.16)}$$

Then the solution of the nonstationary boundary value problem for equation (M12.5.1.1) with the boundary conditions (M12.5.1.2) and initial condition (M12.5.1.3) is sought in the series form

$$w(x,t) = \sum_{n=1}^{\infty} A_n \varphi_n(x)\,\psi_n(t); \qquad \text{(M12.5.1.17)}$$

the A_n are some constant quantities, whose values are determined in the course of solving the problem. The functions $w_n(x,t) = \varphi_n(x)\,\psi_n(t)$ in (M12.5.1.17) are particular solutions of equation (M12.5.1.1) satisfying the boundary conditions (M12.5.1.2). By the principle of linear superposition, series (M12.5.1.17) is also a solution of the original partial differential equation that satisfies the boundary conditions.

To determine the coefficients A_n, we substitute series (M12.5.1.17) into the initial condition (M12.5.1.3), thus obtaining

$$\sum_{n=1}^{\infty} A_n \varphi_n(x) = f_0(x).$$

Multiplying this equation by $\rho(x)\varphi_n(x)$, where the weight function $\rho(x)$ is defined in (M12.5.1.13), then integrating the resulting relation with respect to x over the interval $x_1 \le x \le x_2$, and taking into account the properties (M12.5.1.14), we find

$$A_n = \frac{1}{\|\varphi_n\|^2} \int_{x_1}^{x_2} \rho(x)\varphi_n(x) f_0(x)\,dx, \qquad \|\varphi_n\|^2 = \int_{x_1}^{x_2} \rho(x)\varphi_n^2(x)\,dx. \qquad \text{(M12.5.1.18)}$$

Relations (M12.5.1.17), (M12.5.1.16), (M12.5.1.11), and (M12.5.1.18) give a formal solution of the nonstationary boundary value problem (M12.5.1.1)–(M12.5.1.3).

Example. Consider the first boundary value problem on the interval $0 \le x \le l$ for the heat equation

$$\frac{\partial w}{\partial t} = \frac{\partial^2 w}{\partial x^2} \qquad \text{(M12.5.1.19)}$$

with the general initial condition (M12.5.1.3) and the homogeneous boundary conditions

$$w = 0 \quad \text{at} \quad x = 0, \qquad w = 0 \quad \text{at} \quad x = l. \qquad \text{(M12.5.1.20)}$$

The functions $\psi_n(t)$ in solution (M12.5.1.17) are found from (M12.5.1.16) where $\gamma(t) = 0$:

$$\psi_n(t) = \exp(-\lambda_n t). \qquad \text{(M12.5.1.21)}$$

The functions $\varphi_n(x)$ are determined by solving the eigenvalue problem (M12.5.1.5), (M12.5.1.7) with $a(x) = 1$, $b(x) = c(x) = 0$, $s_1 = s_2 = 0$, $k_1 = k_2 = 1$, $x_1 = 0$, and $x_2 = l$:

$$\varphi''_{xx} + \lambda\varphi = 0; \qquad \varphi = 0 \quad \text{at} \quad x = 0, \qquad \varphi = 0 \quad \text{at} \quad x = l.$$

So we obtain the eigenfunctions and eigenvalues:

$$\varphi_n(x) = \sin\left(\frac{n\pi x}{l}\right), \qquad \lambda_n = \left(\frac{n\pi}{l}\right)^2, \qquad n = 1,\,2,\,\ldots \qquad \text{(M12.5.1.22)}$$

The solution to problem (M12.5.1.19), (M12.5.1.20), (M12.5.1.3) is given by formulas (M12.5.1.17), (M12.5.1.18), (M12.5.1.21), and (M12.5.1.22). Taking into account that $\|\varphi_n\|^2 = l/2$, we obtain

$$w(x,t) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi x}{l}\right) \exp\left(-\frac{n^2\pi^2 t}{l^2}\right), \qquad A_n = \frac{2}{l}\int_0^l f_0(\xi)\sin\left(\frac{n\pi\xi}{l}\right)d\xi. \tag{M12.5.1.23}$$

If the function $f_0(x)$ is twice continuously differentiable and the compatibility conditions (see below) are satisfied, then series (M12.5.1.23) is convergent and admits termwise differentiation, once with respect to t and twice with respect to x. In this case, formula (M12.5.1.23) gives the classical smooth solution of problem (M12.5.1.19, (M12.5.1.20), (M12.5.1.3). [If $f_0(x)$ is not as smooth as indicated or if the compatibility conditions are not met, then series (M12.5.1.23) may converge to a discontinuous function, thus giving only a generalized solution.]

► **Conditions of compatibility of initial and boundary conditions.** Suppose the function w has a continuous derivative with respect to t and two continuous derivatives with respect to x and is a solution of problem (M12.5.1.1)–(M12.5.1.3). Then the boundary conditions (M12.5.1.2) and the initial condition (M12.5.1.3) must be consistent; namely, the following compatibility conditions must hold:

$$[s_1 f_0' + k_1 f_0]_{x=x_1} = 0, \quad [s_2 f_0' + k_2 f_0]_{x=x_2} = 0. \tag{M12.5.1.24}$$

If $s_1 = 0$ or $s_2 = 0$, then the additional compatibility conditions

$$\begin{aligned} &[a(x)f_0'' + b(x)f_0']_{x=x_1} = 0 \quad \text{if} \quad s_1 = 0,\\ &[a(x)f_0'' + b(x)f_0']_{x=x_2} = 0 \quad \text{if} \quad s_2 = 0 \end{aligned} \tag{M12.5.1.25}$$

must also hold; the primes denote the derivatives with respect to x.

M12.5.2. Solution of Problems for Hyperbolic Equations

► **Solution of problems by the method of separation of variables.** Consider a linear homogeneous hyperbolic equation of the form

$$\alpha(t)\frac{\partial^2 w}{\partial t^2} + \beta(t)\frac{\partial w}{\partial t} = a(x)\frac{\partial^2 w}{\partial x^2} + b(x)\frac{\partial w}{\partial x} + \big[c(x) + \gamma(t)\big]w \tag{M12.5.2.1}$$

with homogeneous linear boundary conditions (M12.5.1.2) and arbitrary initial conditions,

$$w = f_0(x) \quad \text{at} \quad t = 0, \tag{M12.5.2.2}$$
$$\partial_t w = f_1(x) \quad \text{at} \quad t = 0. \tag{M12.5.2.3}$$

We assume that the coefficients of equation (M12.5.2.1) and boundary conditions (M12.5.1.2) meet the following requirements:

$$\begin{gathered} \alpha(t),\ \beta(t),\ \gamma(t),\ a(x),\ b(x),\ \text{and}\ c(x)\ \text{are continuous functions},\\ \alpha(t) > 0, \quad a(x) > 0, \quad |s_1| + |k_1| > 0, \quad |s_2| + |k_2| > 0. \end{gathered}$$

Let us use the method of separation of variables to solve boundary value problems for homogeneous linear equations of the hyperbolic type (M12.5.2.1) with homogeneous boundary conditions (M12.5.1.2) and nonhomogeneous initial conditions (M12.5.2.2) and (M12.5.2.3).

The approach is based on searching for particular solutions of equation (M12.5.2.1) in the product form (M12.5.1.4). After separating the variables and rearranging (in much the same way as in Subsection M12.5.1 for parabolic equations), one arrives at the following linear ordinary differential equations for the functions $\varphi = \varphi(x)$ and $\psi = \psi(t)$:

$$a(x)\varphi''_{xx} + b(x)\varphi'_x + [\lambda + c(x)]\varphi = 0, \qquad \text{(M12.5.2.4)}$$

$$\alpha(t)\psi''_{tt} + \beta(t)\psi'_t + [\lambda - \gamma(t)]\psi = 0. \qquad \text{(M12.5.2.5)}$$

Note that equation (M12.5.2.4) coincides with (M12.5.1.5).

For hyperbolic equations, the solution of the boundary value problem (M12.5.2.1)–(M12.5.2.3), (M12.5.1.2) is sought in the series form

$$w(x,t) = \sum_{n=1}^{\infty} \varphi_n(x) \left[A_n \psi_{n1}(t) + B_n \psi_{n2}(t)\right]. \qquad \text{(M12.5.2.6)}$$

The A_n and B_n are some constant quantities, whose values are determined further in the course of solving the problem. The functions $\varphi_n(x)$ and the numbers λ_n are determined by solving the eigenvalue problem (M12.5.2.4), (M12.5.1.7). The functions $\psi_{n1}(t)$ and $\psi_{n2}(t)$ are particular solutions of the linear equation (M12.5.2.5) for ψ (with $\lambda = \lambda_n$) that satisfy the conditions

$$\psi_{n1}(0) = 1, \quad \psi'_{n1}(0) = 0; \qquad \psi_{n2}(0) = 0, \quad \psi'_{n2}(0) = 1. \qquad \text{(M12.5.2.7)}$$

Substituting solution (M12.5.2.6) into the initial conditions (M12.5.2.2) and (M12.5.2.3) yields

$$\sum_{n=1}^{\infty} A_n \varphi_n(x) = f_0(x), \qquad \sum_{n=1}^{\infty} B_n \varphi_n(x) = f_1(x).$$

Multiplying these equations by $\rho(x)\varphi_n(x)$, where the weight function $\rho(x)$ is defined in (M12.5.1.13), then integrating the resulting relations with respect to x over the interval $x_1 \le x \le x_2$, and taking into account the properties (M12.5.1.14), we obtain the coefficients of series (M12.5.2.6) in the form

$$A_n = \frac{1}{\|\varphi_n\|^2} \int_{x_1}^{x_2} \rho(x)\varphi_n(x) f_0(x)\,dx, \quad B_n = \frac{1}{\|\varphi_n\|^2} \int_{x_1}^{x_2} \rho(x)\varphi_n(x) f_1(x)\,dx. \qquad \text{(M12.5.2.8)}$$

The quantity $\|\varphi_n\|$ is defined in (M12.5.1.18).

Relations (M12.5.2.6), (M12.5.1.11), and (M12.5.2.8) together with the solutions of the problems for equation (M12.5.2.5) subject to the initial conditions (M12.5.2.7) give a formal solution of the nonstationary boundary value problem for equation (M12.5.2.1) subject to the initial conditions (M12.5.2.2) and (M12.5.2.3) and boundary conditions (M12.5.1.2).

Example. Consider a mixed boundary value problem on the interval $0 \le x \le l$ for the wave equation

$$\frac{\partial^2 w}{\partial t^2} = \frac{\partial^2 w}{\partial x^2} \qquad \text{(M12.5.2.9)}$$

with the general initial conditions (M12.5.2.2) and (M12.5.2.3) and the homogeneous boundary conditions

$$w = 0 \quad \text{at} \quad x = 0, \qquad \partial_x w = 0 \quad \text{at} \quad x = l. \qquad \text{(M12.5.2.10)}$$

The solution of this mixed boundary value problem for equation (M12.5.2.9) is sought in the series form (M12.5.2.6). The functions $\varphi_n(x)$ are determined by solving the eigenvalue problem (M12.5.2.4), (M12.5.1.7) with $a(x) = 1$, $b(x) = c(x) = 0$, $s_1 = k_2 = 0$, $s_2 = k_1 = 1$, $x_1 = 0$, and $x_2 = l$:

$$\varphi''_{xx} + \lambda\varphi = 0; \qquad \varphi = 0 \quad \text{at} \quad x = 0, \qquad \varphi'_x = 0 \quad \text{at} \quad x = l.$$

So we obtain the eigenfunctions and eigenvalues:

$$\varphi_n(x) = \sin(\mu_n x), \quad \mu_n = \sqrt{\lambda_n} = \frac{\pi(2n-1)}{2l}, \quad n = 1,\ 2,\ \dots \tag{M12.5.2.11}$$

The functions $\psi_{n1}(t)$ and $\psi_{n2}(t)$ are determined by the linear equation [see (M12.5.2.5) with $\alpha(t) = 1$, $\beta(t) = \gamma(t) = 0$, and $\lambda = \lambda_n$]

$$\psi''_{tt} + \lambda\psi = 0$$

with the initial conditions (M12.5.2.7). We find

$$\psi_{n1}(t) = \cos\bigl(\sqrt{\lambda_n}\,t\bigr), \quad \psi_{n2}(t) = \frac{1}{\sqrt{\lambda_n}}\sin\bigl(\sqrt{\lambda_n}\,t\bigr). \tag{M12.5.2.12}$$

The solution to problem (M12.5.2.9), (M12.5.2.10), (M12.5.2.2), (M12.5.2.3) is given by formulas (M12.5.2.6), (M12.5.2.11), (M12.5.2.12), and (M12.5.2.8). Taking into account that $\|\varphi_n\|^2 = l/2$, we have

$$\begin{gathered} w(x,t) = \sum_{n=1}^{\infty}\bigl[A_n\cos(\mu_n t) + B_n\sin(\mu_n t)\bigr]\sin(\mu_n x), \quad \mu_n = \frac{\pi(2n-1)}{2l}, \\ A_n = \frac{2}{l}\int_0^l f_0(x)\sin(\mu_n x)\,dx, \quad B_n = \frac{2}{l\mu_n}\int_0^l f_1(x)\sin(\mu_n x)\,dx. \end{gathered} \tag{M12.5.2.13}$$

If $f_0(x)$ and $f_1(x)$ have three and two continuous derivatives, respectively, and the compatibility conditions are met (see below), then series (M12.5.2.13) is convergent and admits double termwise differentiation. In this case, formula (M12.5.2.13) gives the classical smooth solution of problem (M12.5.2.9), (M12.5.2.10), (M12.5.2.2), (M12.5.2.3).

► **Conditions of compatibility of initial and boundary conditions.** Suppose w is a twice continuously differentiable solution of the problem for equation (M12.5.2.1) subject to the initial conditions (M12.5.2.2) and (M12.5.2.3) and boundary conditions (M12.5.1.2). Then conditions (M12.5.1.24) and (M12.5.1.25) must hold. In addition, the following conditions of compatibility of the boundary conditions (M12.5.1.2) and initial condition (M12.5.2.3) must be satisfied:

$$[s_1 f_1' + k_1 f_1]_{x=x_1} = 0, \quad [s_2 f_1' + k_2 f_1]_{x=x_2} = 0.$$

M12.5.3. Solution of Problems for Elliptic Equations

► **Solution of a special problem for elliptic equations.** Now consider a boundary value problem for the elliptic equation

$$a(x)\frac{\partial^2 w}{\partial x^2} + \alpha(y)\frac{\partial^2 w}{\partial y^2} + b(x)\frac{\partial w}{\partial x} + \beta(y)\frac{\partial w}{\partial y} + \bigl[c(x) + \gamma(y)\bigr]w = 0 \tag{M12.5.3.1}$$

with homogeneous linear boundary conditions (M12.5.1.2) in x and the following mixed (homogeneous and nonhomogeneous) boundary conditions in y:

$$\begin{aligned} \sigma_1\partial_y w + \nu_1 w &= 0 && \text{at} \quad y = y_1, \\ \sigma_2\partial_y w + \nu_2 w &= f(x) && \text{at} \quad y = y_2. \end{aligned} \tag{M12.5.3.2}$$

We assume that the coefficients of equation (M12.5.3.1) and boundary conditions (M12.5.1.2) and (M12.5.3.2) meet the following requirements:

$$\begin{gathered} a(x),\ b(x),\ c(x)\ \alpha(y),\ \beta(y),\ \text{and}\ \gamma(y)\ \text{are continuous functions,} \\ a(x) > 0,\ \alpha(y) > 0,\ |s_1| + |k_1| > 0,\ |s_2| + |k_2| > 0,\ |\sigma_1| + |\nu_1| > 0,\ |\sigma_2| + |\nu_2| > 0. \end{gathered}$$

The method of separation of variables is based on searching for particular solutions of equation (M12.5.3.1) in the product form

$$w(x,y)=\varphi(x)\,\psi(y). \qquad \text{(M12.5.3.3)}$$

As before, we first arrive at the eigenvalue problem (M12.5.1.5), (M12.5.1.7) for the function $\varphi=\varphi(x)$; the solution procedure is detailed in Subsection M12.5.1. Further on, we assume that the λ_n and $\varphi_n(x)$ have been found. The functions $\psi_n=\psi_n(y)$ are determined by solving the linear ordinary differential equation

$$\alpha(y)\psi''_{yy}+\beta(y)\psi'_y+[\gamma(y)-\lambda_n]\psi=0 \qquad \text{(M12.5.3.4)}$$

subject to the homogeneous boundary condition

$$\sigma_1\partial_y\psi+\nu_1\psi=0 \quad \text{at} \quad y=y_1, \qquad \text{(M12.5.3.5)}$$

which follows from the first condition in (M12.5.3.2). The functions ψ_n are determined up to a constant factor.

Taking advantage of the principle of linear superposition, we seek the solution to the boundary value problem (M12.5.3.1), (M12.5.3.2), (M12.5.1.2) in the series form

$$w(x,y)=\sum_{n=1}^{\infty}A_n\varphi_n(x)\psi_n(y); \qquad \text{(M12.5.3.6)}$$

The A_n are some constants, which are to be determined. By construction, series (M12.5.3.6) will satisfy equation (M12.5.3.1) with the boundary conditions (M12.5.1.2) and the first boundary condition (M12.5.3.2). In order to find the coefficients A_n, substitute (M12.5.3.6) into the second boundary condition (M12.5.3.2) to obtain

$$\sum_{n=1}^{\infty}A_nB_n\varphi_n(x)=f(x), \qquad B_n=\sigma_2\frac{d\psi_n}{dy}\bigg|_{y=y_2}+\nu_2\psi_n(y_2). \qquad \text{(M12.5.3.7)}$$

Further, we follow the same procedure as in Subsection M12.5.1. Specifically, multiplying (M12.5.3.7) by $\rho(x)\varphi_n(x)$, then integrating the resulting relation with respect to x over the interval $x_1\le x\le x_2$ and taking into account the properties (M12.5.1.14), we obtain

$$A_n=\frac{1}{B_n\|\varphi_n\|^2}\int_{x_1}^{x_2}\rho(x)\varphi_n(x)f(x)\,dx, \quad \|\varphi_n\|^2=\int_{x_1}^{x_2}\rho(x)\varphi_n^2(x)\,dx, \qquad \text{(M12.5.3.8)}$$

where the weight function $\rho(x)$ is defined in (M12.5.1.13).

Example. Consider the first (Dirichlet) boundary value problem for the Laplace equation

$$\frac{\partial^2 w}{\partial x^2}+\frac{\partial^2 w}{\partial y^2}=0 \qquad \text{(M12.5.3.9)}$$

subject to the boundary conditions

$$\begin{aligned} &w=0 \quad \text{at} \quad x=0, \qquad w=0 \quad \text{at} \quad x=l_1;\\ &w=0 \quad \text{at} \quad y=0, \qquad w=f(x) \quad \text{at} \quad y=l_2 \end{aligned} \qquad \text{(M12.5.3.10)}$$

in a rectangular domain $0\le x\le l_1$, $0\le y\le l_2$.

Particular solutions to equation (M12.5.3.9) are sought in the form (M12.5.3.3). We have the following eigenvalue problem for $\varphi(x)$:

$$\varphi''_{xx} + \lambda\varphi = 0; \qquad \varphi = 0 \quad \text{at} \quad x = 0, \qquad \varphi = 0 \quad \text{at} \quad x = l_1.$$

On solving this problem, we find the eigenfunctions with respective eigenvalues

$$\varphi_n(x) = \sin(\mu_n x), \qquad \mu_n = \sqrt{\lambda_n} = \frac{\pi n}{l_1}, \qquad n = 1,\, 2,\, \ldots \tag{M12.5.3.11}$$

The functions $\psi_n = \psi_n(y)$ are determined by solving the following problem for a linear ordinary differential equation with homogeneous boundary conditions (at this stage, the functions ψ_n are determined up to a constant factor each):

$$\psi''_{yy} - \lambda_n\psi = 0; \qquad \psi = 0 \quad \text{at} \quad y = 0. \tag{M12.5.3.12}$$

It is a special case of problem (M12.5.3.4), (M12.5.3.5) with $\alpha(y) = 1$, $\beta(y) = \gamma(y) = 0$, $\sigma_1 = 0$, and $\nu_1 = 1$. The nontrivial solutions of problem (M12.5.3.12) are expressed as

$$\psi_n(y) = \sinh(\mu_n y), \qquad \mu_n = \sqrt{\lambda_n} = \frac{\pi n}{l_1}, \qquad n = 1,\, 2,\, \ldots \tag{M12.5.3.13}$$

Using formulas (M12.5.3.6), (M12.5.3.8), (M12.5.3.11), (M12.5.3.13) and taking into account the relations $B_n = \psi_n(l_2) = \sinh(\mu_n l_2)$, $\rho(x) = 1$, and $\|\varphi_n\|^2 = l_1/2$, we find the solution of the original problem (M12.5.3.9), (M12.5.3.10) in the form

$$w(x,y) = \sum_{n=1}^{\infty} A_n \sin(\mu_n x)\sinh(\mu_n y), \qquad A_n = \frac{2}{l_1 \sinh(\mu_n l_2)} \int_0^{l_1} f(x)\sin(\mu_n x)\,dx, \qquad \mu_n = \frac{\pi n}{l_1}.$$

► **Generalization to the case of nonhomogeneous boundary conditions.** Now consider the linear boundary value problem for the elliptic equation (M12.5.3.1) with general nonhomogeneous boundary conditions

$$\begin{aligned} &s_1\partial_x w + k_1 w = f_1(y) \quad \text{at} \quad x = x_1, \qquad s_2\partial_x w + k_2 w = f_2(y) \quad \text{at} \quad x = x_2, \\ &\sigma_1\partial_y w + \nu_1 w = f_3(x) \quad \text{at} \quad y = y_1, \qquad \sigma_2\partial_y w + \nu_2 w = f_4(x) \quad \text{at} \quad y = y_2. \end{aligned} \tag{M12.5.3.14}$$

The solution to this problem is the sum of solutions to four simpler auxiliary problems for equation (M12.5.3.1), each corresponding to one nonhomogeneous and three homogeneous boundary conditions in (M12.5.3.14); in these problems, all but one of the functions f_n are zero. Each auxiliary problem is solved using the procedure described above in this subsection, beginning with the search for solutions in the form of the product of functions with different arguments (M12.5.3.3), determined by equations (M12.5.1.5) and (M12.5.3.4). The solution to each of the auxiliary problems is sought in the series form (M12.5.3.6).

M12.6. Integral Transforms Method

Various integral transforms are widely used to solve linear problems of mathematical physics. The Laplace transform and the Fourier transform are in most common use (these and other integral transforms are considered in Chapter M10 in detail).

M12.6.1. Laplace Transform and Its Application to Linear Mathematical Physics Equations

► **Laplace and inverse Laplace transforms. Laplace transforms for derivatives.** The Laplace transform of an arbitrary (complex-valued) function $f(t)$ of a real variable t ($t \ge 0$) is defined by

$$\widetilde{f}(p) = \mathfrak{L}\bigl\{f(t)\bigr\}, \quad \text{where} \quad \mathfrak{L}\bigl\{f(t)\bigr\} \equiv \int_0^{\infty} e^{-pt} f(t)\,dt, \tag{M12.6.1.1}$$

where $p = s + i\sigma$ is a complex variable, $i^2 = -1$.

Given the transform $\widetilde{f}(p)$, the function $f(t)$ can be found by means of the inverse Laplace transform

$$f(t) = \mathfrak{L}^{-1}\bigl\{\widetilde{f}(p)\bigr\}, \quad \text{where} \quad \mathfrak{L}^{-1}\bigl\{\widetilde{f}(p)\bigr\} \equiv \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \widetilde{f}(p)e^{pt}\,dp, \tag{M12.6.1.2}$$

where the integration path is parallel to the imaginary axis and lies to the right of all singularities of $\widetilde{f}(p)$.

In order to solve nonstationary boundary value problems, the following Laplace transform formulas for derivatives will be required:

$$\begin{aligned} \mathfrak{L}\bigl\{f'(t)\bigr\} &= p\widetilde{f}(p) - f(0), \\ \mathfrak{L}\bigl\{f''(t)\bigr\} &= p^2\widetilde{f}(p) - pf(0) - f'(0), \end{aligned} \tag{M12.6.1.3}$$

where $f(0)$ and $f'(0)$ are the initial values.

More details on the properties of the Laplace transform and the inverse Laplace transform can be found in Section M10.2. The Laplace transforms of some functions are listed in Section S2.1. Tables of inverse Laplace transforms are listed in Section S2.2. Such tables are convenient to use when solving linear problems for partial differential equations.

► **Solution procedure for linear problems using the Laplace transform.** Figure M12.1 shows schematically how one can utilize the Laplace transforms to solve boundary value problems for linear parabolic or hyperbolic equations with two independent variables in the case where the equation coefficients are independent of t (the same procedure can be applied to solving linear problems characterized by higher-order equations). Here and henceforth, the short notation $\widetilde{w}(x,p) = \mathfrak{L}\{w(x,t)\}$ will be used; the arguments of $\widetilde{w}$ may be omitted.

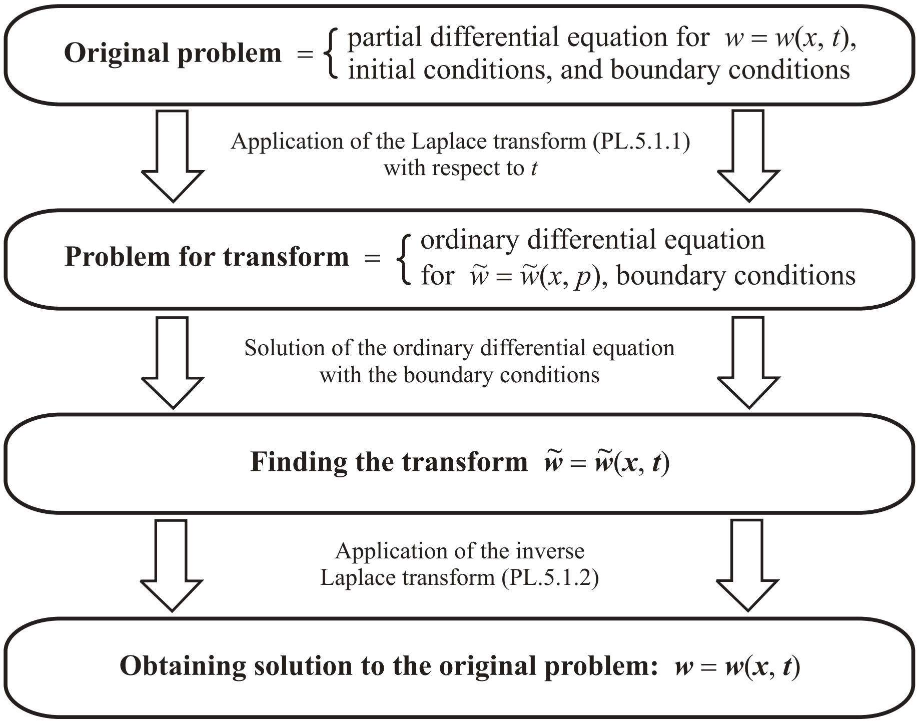
Original problem = { partial differential equation for $w = w(x, t)$, initial conditions, and boundary conditions
Application of the Laplace transform (PL.5.1.1) with respect to t
Problem for transform = { ordinary differential equation for $\widetilde{w} = \widetilde{w}(x, p)$, boundary conditions
Solution of the ordinary differential equation with the boundary conditions
Finding the transform $\widetilde{w} = \widetilde{w}(x, t)$
Application of the inverse Laplace transform (PL.5.1.2)
Obtaining solution to the original problem: $w = w(x, t)$

Figure M12.1. Solution procedure for linear boundary value problems using the Laplace transform.

It is significant that with the Laplace transform, the original problem for a partial differential equation is reduced to a simpler problem for an ordinary differential equation with parameter p; the derivatives with respect to t are replaced by appropriate algebraic expressions taking into account the initial conditions; see formulas (M12.6.1.3).

► **Solving linear problems for parabolic equations using the Laplace transform.** Consider a linear nonstationary boundary value problem for the parabolic equation

$$\frac{\partial w}{\partial t} = a(x)\frac{\partial^2 w}{\partial x^2} + b(x)\frac{\partial w}{\partial x} + c(x)w + \Phi(x,t) \tag{M12.6.1.4}$$

with the initial condition (M12.5.1.3) and the general nonhomogeneous boundary conditions

$$\begin{aligned} s_1\partial_x w + k_1 w &= g_1(t) \quad \text{at} \quad x = x_1, \\ s_2\partial_x w + k_2 w &= g_2(t) \quad \text{at} \quad x = x_2. \end{aligned} \tag{M12.6.1.5}$$

The application of the Laplace transform (M12.6.1.1) results in the problem determined by the ordinary differential equation for x (p is treated as a parameter)

$$a(x)\widetilde{w}''_{xx} + b(x)\widetilde{w}'_x + [c(x) - p]\widetilde{w} + f_0(x) + \widetilde{\Phi}(x,p) = 0 \tag{M12.6.1.6}$$

with the boundary conditions

$$\begin{aligned} s_1\widetilde{w}'_x + k_1\widetilde{w} &= \widetilde{g}_1(p) \quad \text{at} \quad x = x_1, \\ s_2\widetilde{w}'_x + k_2\widetilde{w} &= \widetilde{g}_2(p) \quad \text{at} \quad x = x_2. \end{aligned} \tag{M12.6.1.7}$$

Notation used: $\widetilde{\Phi}(x,p) = \mathfrak{L}\{\Phi(x,t)\}$ and $\widetilde{g}_n(p) = \mathfrak{L}\{g_n(t)\}$ ($n = 1,\ 2$). On solving problem (M12.6.1.6), (M12.6.1.7), one should apply to the resulting solution $\widetilde{w} = \widetilde{w}(x,p)$ the inverse Laplace transform (M12.6.1.2) to obtain the solution, $w = w(x,t)$, of the original problem.

Example. Consider the first boundary value problem for the heat equation:

$$\begin{aligned} &\frac{\partial w}{\partial t} = \frac{\partial^2 w}{\partial x^2} && (x > 0,\ t > 0), \\ &w = 0 \quad \text{at} \quad t = 0 && \text{(initial condition)}, \\ &w = w_0 \quad \text{at} \quad x = 0 && \text{(boundary condition)}, \\ &w \to 0 \quad \text{at} \quad x \to \infty && \text{(boundary condition)}. \end{aligned}$$

We apply the Laplace transform with respect to t. Let us multiply the equation, the initial condition, and the boundary conditions by e^{-pt} and then integrate with respect to t from zero to infinity. Taking into account the relations

$\mathfrak{L}\{\partial_t w\} = p\widetilde{w} - w|_{t=0} = p\widetilde{w}$ (used are first property (M12.6.1.3) and the initial condition),

$\mathfrak{L}\{w_0\} = w_0\mathfrak{L}\{1\} = w_0/p$ (used are the first property of Subsection S2.1.1 and the relation $\mathfrak{L}\{1\} = 1/p$),

we arrive at the following problem for a second-order linear ordinary differential equation with parameter p:

$$\begin{aligned} &\widetilde{w}''_{xx} - p\widetilde{w} = 0, \\ &\widetilde{w} = w_0/p \quad \text{at} \quad x = 0 && \text{(boundary condition)}, \\ &\widetilde{w} \to 0 \quad \text{at} \quad x \to \infty && \text{(boundary condition)}. \end{aligned}$$

Integrating the equation yields the general solution $\widetilde{w} = A_1(p)e^{-x\sqrt{p}} + A_2(p)e^{x\sqrt{p}}$. Using the boundary conditions, we determine the constants, $A_1(p) = w_0/p$ and $A_2(p) = 0$. Thus, we have

$$\widetilde{w} = \frac{w_0}{p}e^{-x\sqrt{p}}.$$

Let us apply the inverse Laplace transform to both sides of this relation. We refer to the table in Subsection S2.2.5, row 20 (where x must be replaced by t and then a by x^2), to find the inverse transform of the right-hand side. Finally, we obtain the solution of the original problem:

$$w = w_0 \operatorname{erfc}\left(\frac{x}{2\sqrt{t}}\right), \qquad \text{where} \quad \operatorname{erfc} z = \frac{2}{\sqrt{\pi}}\int_z^\infty \exp\left(-\xi^2\right) d\xi.$$

M12.6.2. Fourier Transform and Its Application for Linear Equations of Mathematical Physics

► **Fourier transform and its properties.** The Fourier transform is defined as follows:

$$\widetilde{f}(u) = \mathfrak{F}\bigl\{f(x)\bigr\}, \quad \text{where} \quad \mathfrak{F}\bigl\{f(x)\bigr\} \equiv \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(x)e^{-iux}\,dx, \qquad i^2 = -1. \tag{M12.6.2.1}$$

This relation is meaningful for any function $f(x)$ absolutely integrable on the interval $(-\infty, \infty)$.

Given $\widetilde{f}(u)$, the function $f(x)$ can be found by means of the inverse Fourier transform

$$f(x) = \mathfrak{F}^{-1}\bigl\{\widetilde{f}(u)\bigr\}, \quad \text{where} \quad \mathfrak{F}^{-1}\bigl\{\widetilde{f}(u)\bigr\} \equiv \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \widetilde{f}(u)e^{iux}\,du, \tag{M12.6.2.2}$$

where the integral is understood in the sense of the Cauchy principal value.

The main properties of the correspondence between functions and their Fourier transforms are gathered in Table M10.4.

► **Solving linear problems of mathematical physics by the Fourier transform.** The Fourier transform is usually employed to solve problems for linear partial differential equations whose coefficients are independent of the space variable x, $-\infty < x < \infty$.

The scheme for solving linear boundary value problems with the help of the Fourier transform is similar to that used in solving problems with the help of the Laplace transform. With the Fourier transform, the derivatives with respect to x in the equation are replaced by appropriate algebraic expressions; see Property 4 or 5 in Table M10.4. In the case of two independent variables, the problem for a partial differential equation is reduced to a simpler problem for an ordinary differential equation with parameter u. On solving the latter problem, one determines the transform. After that, by applying the inverse Fourier transform, one obtains the solution of the original boundary value problem.

Example. Consider the following Cauchy problem for the heat equation:

$$\begin{aligned} &\frac{\partial w}{\partial t} = \frac{\partial^2 w}{\partial x^2} && (-\infty < x < \infty), \\ &w = f(x) \quad \text{at} \quad t = 0 && \text{(initial condition).} \end{aligned}$$

We apply the Fourier transform with respect to the space variable x. Setting $\widetilde{w} = \mathfrak{F}\{w(x,t)\}$ and taking into account the relation $\mathfrak{F}\{\partial_{xx}w\} = -u^2\widetilde{w}$ (see Property 4 in Table M10.4), we arrive at the following problem for a linear first-order ordinary differential equation in t with parameter u:

$$\begin{aligned} &\widetilde{w}'_t + u^2\widetilde{w} = 0, \\ &\widetilde{w} = \widetilde{f}(u) \quad \text{at} \quad t = 0, \end{aligned}$$

where $\widetilde{f}(u)$ is defined by (M12.6.2.1). On solving this problem for the transform $\widetilde{w}$, we find

$$\widetilde{w} = \widetilde{f}(u)e^{-u^2 t}.$$

Let us apply the inversion formula to both sides of this equation. After some calculations, we obtain the solution of the original problem in the form

$$\begin{aligned} w &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \widetilde{f}(u)e^{-u^2 t}e^{iux}\,du = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(\xi)e^{-iu\xi}d\xi\right]e^{-u^2 t + iux}du \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} f(\xi)\,d\xi\int_{-\infty}^{\infty} e^{-u^2 t + iu(x-\xi)}du = \frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{\infty} f(\xi)\exp\left[-\frac{(x-\xi)^2}{4t}\right]d\xi. \end{aligned}$$

At the last stage we used the relation $\displaystyle\int_{-\infty}^{\infty}\exp\bigl(-a^2u^2 + bu\bigr)\,du = \frac{\sqrt{\pi}}{|a|}\exp\left(\frac{b^2}{4a^2}\right)$.

M12.7. Boundary Value Problems for Parabolic Equations with One Space Variable. Green's Function

M12.7.1. Representation of Solutions via the Green's Function

► **Statement of the problem $(t \geq 0,\ x_1 \leq x \leq x_2)$.** In general, a nonhomogeneous linear differential equation of the parabolic type with variable coefficients with one spatial coordinate can be written as

$$\frac{\partial w}{\partial t} - L_{x,t}[w] = \Phi(x,t), \qquad \text{(M12.7.1.1)}$$

where

$$L_{x,t}[w] \equiv a(x,t)\frac{\partial^2 w}{\partial x^2} + b(x,t)\frac{\partial w}{\partial x} + c(x,t)w, \qquad a(x,t) > 0. \qquad \text{(M12.7.1.2)}$$

Consider the nonstationary boundary value problem for equation (M12.7.1.1) with an initial condition of general form

$$w = f(x) \quad \text{at} \quad t = 0, \qquad \text{(M12.7.1.3)}$$

and arbitrary nonhomogeneous linear boundary conditions

$$s_1\frac{\partial w}{\partial x} + k_1 w = g_1(t) \quad \text{at} \quad x = x_1, \qquad \text{(M12.7.1.4)}$$

$$s_2\frac{\partial w}{\partial x} + k_2 w = g_2(t) \quad \text{at} \quad x = x_2. \qquad \text{(M12.7.1.5)}$$

By appropriately choosing the coefficients s_1, s_2, k_1, and k_2 in (M12.7.1.4) and (M12.7.1.5), one arrives at the first, second, third, and mixed boundary value problems for equation (M12.7.1.1).

► **Representation of the problem solution in terms of the Green's function.** The solution of the nonhomogeneous linear boundary value problem (M12.7.1.1)–(M12.7.1.5) can be represented as

$$\begin{aligned} w(x,t) = &\int_0^t \int_{x_1}^{x_2} \Phi(y,\tau)G(x,y,t,\tau)\,dy\,d\tau + \int_{x_1}^{x_2} f(y)G(x,y,t,0)\,dy \\ &+ \int_0^t g_1(\tau)a(x_1,\tau)\Lambda_1(x,t,\tau)\,d\tau + \int_0^t g_2(\tau)a(x_2,\tau)\Lambda_2(x,t,\tau)\,d\tau. \end{aligned} \qquad \text{(M12.7.1.6)}$$

Here, $G(x,y,t,\tau)$ is the Green's function that satisfies, for $t > \tau \geq 0$, the homogeneous equation

$$\frac{\partial G}{\partial t} - L_{x,t}[G] = 0 \qquad \text{(M12.7.1.7)}$$

with the nonhomogeneous initial condition of a special form

$$G = \delta(x-y) \quad \text{at} \quad t = \tau \qquad \text{(M12.7.1.8)}$$

TABLE M12.4
Expressions of the functions $\Lambda_1(x,t,\tau)$ and $\Lambda_2(x,t,\tau)$ involved in the integrands of the last two terms in solution (M12.7.1.6).

Type of problem	Form of boundary conditions	Functions $\Lambda_m(x,t,\tau)$
First boundary value problem $(s_1=s_2=0,\ k_1=k_2=1)$	$w=g_1(t)$ at $x=x_1$ $w=g_2(t)$ at $x=x_2$	$\Lambda_1(x,t,\tau)=\partial_y G(x,y,t,\tau)\big\|_{y=x_1}$ $\Lambda_2(x,t,\tau)=-\partial_y G(x,y,t,\tau)\big\|_{y=x_2}$
Second boundary value problem $(s_1=s_2=1,\ k_1=k_2=0)$	$\partial_x w=g_1(t)$ at $x=x_1$ $\partial_x w=g_2(t)$ at $x=x_2$	$\Lambda_1(x,t,\tau)=-G(x,x_1,t,\tau)$ $\Lambda_2(x,t,\tau)=G(x,x_2,t,\tau)$
Third boundary value problem $(s_1=s_2=1,\ k_1<0,\ k_2>0)$	$\partial_x w+k_1w=g_1(t)$ at $x=x_1$ $\partial_x w+k_2w=g_2(t)$ at $x=x_2$	$\Lambda_1(x,t,\tau)=-G(x,x_1,t,\tau)$ $\Lambda_2(x,t,\tau)=G(x,x_2,t,\tau)$
Mixed boundary value problem $(s_1=k_2=0,\ s_2=k_1=1)$	$w=g_1(t)$ at $x=x_1$ $\partial_x w=g_2(t)$ at $x=x_2$	$\Lambda_1(x,t,\tau)=\partial_y G(x,y,t,\tau)\big\|_{y=x_1}$ $\Lambda_2(x,t,\tau)=G(x,x_2,t,\tau)$
Mixed boundary value problem $(s_1=k_2=1,\ s_2=k_1=0)$	$\partial_x w=g_1(t)$ at $x=x_1$ $w=g_2(t)$ at $x=x_2$	$\Lambda_1(x,t,\tau)=-G(x,x_1,t,\tau)$ $\Lambda_2(x,t,\tau)=-\partial_y G(x,y,t,\tau)\big\|_{y=x_2}$

and the homogeneous boundary conditions

$$s_1\frac{\partial G}{\partial x}+k_1G=0 \quad \text{at} \quad x=x_1, \tag{M12.7.1.9}$$

$$s_2\frac{\partial G}{\partial x}+k_2G=0 \quad \text{at} \quad x=x_2. \tag{M12.7.1.10}$$

The quantities y and τ appear in problem (M12.7.1.7)–(M12.7.1.10) as free parameters, with $x_1\le y\le x_2$, and $\delta(x)$ is the Dirac delta function.

The initial condition (M12.7.1.8) implies the limit relation

$$f(x)=\lim_{t\to\tau}\int_{x_1}^{x_2} f(y)G(x,y,t,\tau)\,dy$$

for any continuous function $f=f(x)$.

The functions $\Lambda_1(x,t,\tau)$ and $\Lambda_2(x,t,\tau)$ involved in the integrands of the last two terms in solution (M12.7.1.6) can be expressed in terms of the Green's function $G(x,y,t,\tau)$. The corresponding formulas for $\Lambda_m(x,t,\tau)$ are given in Table M12.4 for the basic types of boundary value problems.

It is significant that the Green's function G and the functions Λ_1, Λ_2 are independent of the functions Φ, f, g_1, and g_2 that characterize various nonhomogeneities of the boundary value problem.

If the coefficients of equation (M12.7.1.1)–(M12.7.1.2) are independent of time t, i.e., the conditions

$$a=a(x), \quad b=b(x), \quad c=c(x) \tag{M12.7.1.11}$$

hold, then the Green's function depends on only three arguments,

$$G(x,y,t,\tau)=G(x,y,t-\tau). \tag{M12.7.1.12}$$

In this case, the functions Λ_m depend on only two arguments, $\Lambda_m=\Lambda_m(x,t-\tau)$, $m=1,\ 2$.

Formula (M12.7.1.6) also remains valid for the problem with boundary conditions of the third kind if $k_1=k_1(t)$ and $k_2=k_2(t)$. Here, the relations between Λ_m ($m=1,\ 2$) and the Green's function G are the same as in the case of constant k_1 and k_2; the Green's function itself is now different.

Remark. In the first, second, and third boundary value problems that are considered on the interval $x_1 \le x < \infty$, a condition of boundedness of the solution as $x \to \infty$ is set out. In this case, the solution is calculated by formula (M12.7.1.6) with $\Lambda_2 = 0$ and Λ_1 specified in Table M12.4.

► **Formulas for calculating Green's functions.** Consider the parabolic equation of a special form

$$\frac{\partial w}{\partial t} = a(x)\frac{\partial^2 w}{\partial x^2} + b(x)\frac{\partial w}{\partial x} + \bigl[c(x) + \gamma(t)\bigr]w + \Phi(x,t), \qquad \text{(M12.7.1.13)}$$

which is a special case of equation (M12.7.1.1) with operator (M12.7.1.2) where $a(x,t) = a(x)$, $b(x,t) = b(x)$, and $c(x,t) = c(x) + \gamma(t)$. It is assumed that $a(x) > 0$.

The solution of the nonhomogeneous linear boundary value problem (M12.7.1.13), (M12.7.1.2)–(M12.7.1.5) is found by formula (M12.7.1.6), where the Green's function is given by

$$G(x,y,t,\tau) = \rho(y)\sum_{n=1}^{\infty}\frac{\varphi_n(x)\varphi_n(y)}{\|\varphi_n\|^2}\exp\biggl[-\lambda_n(t-\tau) + \int_\tau^t \gamma(\xi)\,d\xi\biggr], \qquad \text{(M12.7.1.14)}$$

where the λ_n and $\varphi_n(x)$ are the eigenvalues and the corresponding eigenfunctions of the Sturm–Liouville problem for the linear ordinary differential equation (M12.5.1.5) with the homogeneous linear boundary conditions (M12.5.1.7), and

$$\rho(y) = \frac{1}{a(y)}\exp\biggl[\int \frac{b(y)}{a(y)}\,dy\biggr], \qquad \|\varphi_n\|^2 = \int_{x_1}^{x_2}\rho(x)\varphi_n^2(x)\,dx, \qquad \text{(M12.7.1.15)}$$

Table M12.5 lists Green's functions for some problems for the nonhomogeneous heat equation that corresponds to $a(x) = a = \text{const}$ and $b(x) = c(x) = \gamma(t) \equiv 0$ in (M12.7.1.13). The function $G(x,y,t-\tau)$ from this table must be substituted for $G(x,y,t,\tau)$ in formula (M12.7.1.6).

M12.7.2. Problems for Parabolic Equation $\rho(x)\frac{\partial w}{\partial t} = \frac{\partial}{\partial x}\Bigl[p(x)\frac{\partial w}{\partial x}\Bigr] - q(x)w + \Phi(x,t)$

► **General formulas for solving nonhomogeneous boundary value problems.** Consider a linear parabolic equation of the special form

$$\rho(x)\frac{\partial w}{\partial t} = \frac{\partial}{\partial x}\biggl[p(x)\frac{\partial w}{\partial x}\biggr] - q(x)w + \Phi(x,t). \qquad \text{(M12.7.2.1)}$$

It is often encountered in heat and mass transfer theory and chemical engineering sciences. Throughout this subsection, we assume that the functions ρ, p, p'_x, and q are continuous and $\rho > 0$, $p > 0$, and $x_1 \le x \le x_2$.

The solution of equation (M12.7.2.1) under the initial condition (M12.7.1.3) and the arbitrary linear nonhomogeneous boundary conditions (M12.7.1.4) and (M12.7.1.5) can be represented as the sum

$$\begin{aligned} w(x,t) &= \int_0^t\int_{x_1}^{x_2}\Phi(y,\tau)\mathcal{G}(x,y,t-\tau)\,dy\,d\tau + \int_{x_1}^{x_2}\rho(y)f(y)\mathcal{G}(x,y,t)\,dy \\ &\quad + p(x_1)\int_0^t g_1(\tau)\Lambda_1(x,t-\tau)\,d\tau + p(x_2)\int_0^t g_2(\tau)\Lambda_2(x,t-\tau)\,d\tau. \end{aligned} \qquad \text{(M12.7.2.2)}$$

TABLE M12.5
The Green's functions for some boundary value problems for the heat equation.

Type of problem	Green's function, $G(x,y,t)$	Remarks
First boundary value problem ($0\le x<\infty$)	$\frac{1}{2\sqrt{\pi at}}\Bigl\{\exp\Bigl[-\frac{(x-y)^2}{4at}\Bigr]-\exp\Bigl[-\frac{(x+y)^2}{4at}\Bigr]\Bigr\}$	Looking for solution bounded as $x\to\infty$
Second boundary value problem ($0\le x<\infty$)	$\frac{1}{2\sqrt{\pi at}}\Bigl\{\exp\Bigl[-\frac{(x-y)^2}{4at}\Bigr]+\exp\Bigl[-\frac{(x+y)^2}{4at}\Bigr]\Bigr\}$	Looking for solution bounded as $x\to\infty$
First boundary value problem ($0\le x\le l$)	$\frac{2}{l}\sum_{n=1}^{\infty}\sin\Bigl(\frac{n\pi x}{l}\Bigr)\sin\Bigl(\frac{n\pi y}{l}\Bigr)\exp\Bigl(-\frac{an^2\pi^2t}{l^2}\Bigr)$ $=\frac{1}{2\sqrt{\pi at}}\sum_{n=-\infty}^{\infty}\Bigl\{\exp\Bigl[-\frac{(x-y+2nl)^2}{4at}\Bigr]-\exp\Bigl[-\frac{(x+y+2nl)^2}{4at}\Bigr]\Bigr\}$	The first series converges rapidly at large t and the second series at small t
Second boundary value problem ($0\le x\le l$)	$\frac{1}{l}+\frac{2}{l}\sum_{n=1}^{\infty}\cos\Bigl(\frac{n\pi x}{l}\Bigr)\cos\Bigl(\frac{n\pi y}{l}\Bigr)\exp\Bigl(-\frac{an^2\pi^2t}{l^2}\Bigr)$ $=\frac{1}{2\sqrt{\pi at}}\sum_{n=-\infty}^{\infty}\Bigl\{\exp\Bigl[-\frac{(x-y+2nl)^2}{4at}\Bigr]+\exp\Bigl[-\frac{(x+y+2nl)^2}{4at}\Bigr]\Bigr\}$	The first series converges rapidly at large t and the second series at small t
Mixed boundary value problem ($0\le x\le l$); w set at $x=0$ and $\partial_x w$ set at $x=l$	$\frac{2}{l}\sum_{n=0}^{\infty}\sin\Bigl[\frac{\pi(2n+1)x}{2l}\Bigr]\sin\Bigl[\frac{\pi(2n+1)y}{2l}\Bigr]\exp\Bigl[-\frac{a\pi^2(2n+1)^2t}{4l^2}\Bigr]$ $=\frac{1}{2\sqrt{\pi at}}\sum_{n=-\infty}^{\infty}(-1)^n\Bigl\{\exp\Bigl[-\frac{(x-y+2nl)^2}{4at}\Bigr]-\exp\Bigl[-\frac{(x+y+2nl)^2}{4at}\Bigr]\Bigr\}$	The first series converges rapidly at large t and the second series at small t

Here, the modified Green's function is given by

$$\mathcal{G}(x,y,t)=\sum_{n=1}^{\infty}\frac{\varphi_n(x)\varphi_n(y)}{\|\varphi_n\|^2}\exp(-\lambda_n t),\qquad \|\varphi_n\|^2=\int_{x_1}^{x_2}\rho(x)\varphi_n^2(x)\,dx,\quad \text{(M12.7.2.3)}$$

where the λ_n and $\varphi_n(x)$ are the eigenvalues and corresponding eigenfunctions of the following Sturm–Liouville problem for a second-order linear ordinary differential equation:

$$\begin{aligned}&[p(x)\varphi'_x]'_x+[\lambda\rho(x)-q(x)]\varphi=0,\\ &s_1\varphi'_x+k_1\varphi=0\quad\text{at}\quad x=x_1,\\ &s_2\varphi'_x+k_2\varphi=0\quad\text{at}\quad x=x_2.\end{aligned}\qquad \text{(M12.7.2.4)}$$

The functions $\Lambda_1(x,t)$ and $\Lambda_2(x,t)$ that occur in the integrands of the last two terms in solution (M12.7.2.2) are expressed in terms of the Green's function (M12.7.2.3).

Remark. The corresponding formulas for $\Lambda_m(x,t)$ for the basic types of boundary value problems can be obtained from Table M12.4; to do so, one should set $G=\mathcal{G}$ in all formulas of the last column and omit the last argument τ.

► **Properties of Sturm–Liouville problem (M12.7.2.4). Heat equation with a source.** The basic properties of the Sturm–Liouville problem (M12.7.2.4), which coincides with problem (M12.5.1.12), (M12.5.1.7), are outlined in Subsections M12.5.1 and M11.2.4. In particular, there are infinitely many eigenvalues. All eigenvalues are real and different and can be ordered so that $\lambda_1<\lambda_2<\lambda_3<\cdots$, with $\lambda_n\to\infty$ as $n\to\infty$. The different eigenfunctions $\varphi_n(x)$ and $\varphi_m(x)$ are orthogonal with weight $\rho(x)$ on the interval $x_1\le x\le x_2$; see formulas (M12.5.1.14).

Example. Consider the first boundary value problem in the domain $0 \le x \le l$ for the heat equation with a source

$$\frac{\partial w}{\partial t} = a\frac{\partial^2 w}{\partial x^2} - bw$$

under the initial condition (M12.7.1.3) and boundary conditions

$$w = g_1(t) \quad \text{at} \quad x = 0,$$
$$w = g_2(t) \quad \text{at} \quad x = l.$$

The above equation is a special case of equation (M12.7.2.1) with $\rho(x) = 1$, $p(x) = a$, $q(x) = b$, and $\Phi(x,t) = 0$. The corresponding Sturm–Liouville problem (M12.7.2.4) has the form

$$a\varphi''_{xx} + (\lambda - b)\varphi = 0, \qquad \varphi = 0 \quad \text{at} \quad x = 0, \qquad \varphi = 0 \quad \text{at} \quad x = l.$$

The eigenfunctions and eigenvalues are found to be

$$\varphi_n(x) = \sin\left(\frac{\pi n x}{l}\right), \qquad \lambda_n = b + \frac{a\pi^2 n^2}{l^2}, \qquad n = 1,\ 2,\ \dots$$

Using formula (M12.7.2.3) and taking into account that $\|\varphi_n\|^2 = l/2$, we obtain the Green's function

$$\mathcal{G}(x,y,t) = \frac{2}{l}e^{-bt}\sum_{n=1}^{\infty}\sin\left(\frac{\pi n x}{l}\right)\sin\left(\frac{\pi n y}{l}\right)\exp\left(-\frac{a\pi^2 n^2}{l^2}t\right).$$

Substituting this expression into (M12.7.2.2) with $p(x_1) = p(x_2) = \rho(y) = 1$, $x_1 = 0$, and $x_2 = l$ and taking into account the formulas

$$\Lambda_1(x,t) = \partial_y \mathcal{G}(x,y,t)\big|_{y=x_1}, \qquad \Lambda_2(x,t) = -\partial_y \mathcal{G}(x,y,t)\big|_{y=x_2}$$

(see the first row in Table M12.4 and Remark in Subsection M12.7.2), one obtains the solution to the problem in question.

M12.8. Boundary Value Problems for Hyperbolic Equations with One Space Variable. Green's Function. Goursat Problem

M12.8.1. Representation of Solutions in terms of the Green's Function

► **Statement of the problem ($t \ge 0,\ x_1 \le x \le x_2$).** In general, a one-dimensional nonhomogeneous linear differential equation of hyperbolic type with variable coefficients is written as

$$\frac{\partial^2 w}{\partial t^2} + \varphi(x,t)\frac{\partial w}{\partial t} - L_{x,t}[w] = \Phi(x,t), \qquad \text{(M12.8.1.1)}$$

where the operator $L_{x,t}[w]$ is defined by (M12.7.1.2).

Consider the nonstationary boundary value problem for equation (M12.8.1.1) with the initial conditions

$$\begin{aligned} w &= f_0(x) \quad \text{at} \quad t = 0,\\ \partial_t w &= f_1(x) \quad \text{at} \quad t = 0 \end{aligned} \qquad \text{(M12.8.1.2)}$$

and arbitrary nonhomogeneous linear boundary conditions (M12.7.1.4) and (M12.7.1.5).

► **Representation of the problem solution in terms of the Green's function.** The solution of problem (M12.8.1.1), (M12.8.1.2), (M12.7.1.4), (M12.7.1.5) can be represented as the sum

$$\begin{aligned} w(x,t) = & \int_0^t \int_{x_1}^{x_2} \Phi(y,\tau) G(x,y,t,\tau)\,dy\,d\tau \\ & - \int_{x_1}^{x_2} f_0(y) \left[\frac{\partial}{\partial\tau} G(x,y,t,\tau)\right]_{\tau=0} dy + \int_{x_1}^{x_2} \big[f_1(y) + f_0(y)\varphi(y,0)\big] G(x,y,t,0)\,dy \\ & + \int_0^t g_1(\tau) a(x_1,\tau) \Lambda_1(x,t,\tau)\,d\tau + \int_0^t g_2(\tau) a(x_2,\tau)\Lambda_2(x,t,\tau)\,d\tau. \end{aligned} \qquad \text{(M12.8.1.3)}$$

Here, the Green's function $G(x,y,t,\tau)$ is determined by solving the homogeneous equation

$$\frac{\partial^2 G}{\partial t^2} + \varphi(x,t)\frac{\partial G}{\partial t} - L_{x,t}[G] = 0 \qquad \text{(M12.8.1.4)}$$

with the semihomogeneous initial conditions

$$G = 0 \quad \text{at} \quad t = \tau, \qquad \text{(M12.8.1.5)}$$

$$\partial_t G = \delta(x-y) \quad \text{at} \quad t = \tau, \qquad \text{(M12.8.1.6)}$$

and the homogeneous boundary conditions (M12.7.1.9), (M12.7.1.10). The quantities y and τ appear in problem (M12.8.1.4)–(M12.8.1.6), (M12.7.1.9), (M12.7.1.10) as free parameters ($x_1 \le y \le x_2$), and $\delta(x)$ is the Dirac delta function.

The functions $\Lambda_1(x,t,\tau)$ and $\Lambda_2(x,t,\tau)$ involved in the integrands of the last two terms in solution (M12.8.1.3) can be expressed in terms of the Green's function $G(x,y,t,\tau)$. The corresponding formulas for $\Lambda_m(x,t,\tau)$ are given in Table M12.4 for the basic types of boundary value problems.

It is significant that the Green's function G and Λ_1, Λ_2 are independent of the functions Φ, f_0, f_1, g_1, and g_2 that characterize various nonhomogeneities of the boundary value problem.

If the coefficients of equation (M12.8.1.1) are independent of time t, then the Green's function depends on only three arguments, $G(x,y,t,\tau) = G(x,y,t-\tau)$. In this case, one can set $\frac{\partial}{\partial\tau} G(x,y,t,\tau)\big|_{\tau=0} = -\frac{\partial}{\partial t} G(x,y,t)$ in solution (M12.8.1.3).

► **Formulas for the calculation of Green's functions.** Consider the hyperbolic equation of a special form

$$\frac{\partial^2 w}{\partial t^2} + \beta(t)\frac{\partial w}{\partial t} = a(x)\frac{\partial^2 w}{\partial x^2} + b(x)\frac{\partial w}{\partial x} + \big[c(x) + \gamma(t)\big] w + \Phi(x,t), \qquad \text{(M12.8.1.7)}$$

which is a special case of equation (M12.8.1.1) with operator (M12.7.1.2) where $\varphi(x,t) = \beta(t)$, $a(x,t) = a(x)$, $b(x,t) = b(x)$, and $c(x,t) = c(x) + \gamma(t)$. We assume that $a(x) > 0$.

The solution of the nonhomogeneous linear boundary value problem (M12.8.1.7), (M12.8.1.2), (M12.7.1.4), (M12.7.1.5) is found by formula (M12.8.1.3), where the Green's function is given by

$$G(x,y,t,\tau) = \rho(y) \sum_{n=1}^{\infty} \frac{\varphi_n(x)\varphi_n(y)}{\|\varphi_n\|^2} \psi_n(t,\tau), \qquad \text{(M12.8.1.8)}$$

TABLE M12.6
The Green's functions for some boundary value problems for the wave equation $\frac{\partial^2 w}{\partial t^2} = a^2 \frac{\partial^2 w}{\partial x^2}$ in a bounded domain $0 \le x \le l$.

Type of problem	Green's function, $G(x,\xi,t)$	Remarks
First boundary value problem	$\dfrac{2}{a\pi}\sum_{n=1}^{\infty}\dfrac{1}{n}\sin\Bigl(\dfrac{n\pi x}{l}\Bigr)\sin\Bigl(\dfrac{n\pi\xi}{l}\Bigr)\sin\Bigl(\dfrac{n\pi at}{l}\Bigr)$	
Second boundary value problem	$\dfrac{t}{l}+\dfrac{2}{a\pi}\sum_{n=1}^{\infty}\dfrac{1}{n}\cos\Bigl(\dfrac{n\pi x}{l}\Bigr)\cos\Bigl(\dfrac{n\pi\xi}{l}\Bigr)\sin\Bigl(\dfrac{n\pi at}{l}\Bigr)$	
Mixed boundary value problem; w is set at $x=0$ and $\partial_x w$ is set at $x=l$	$\dfrac{2}{al}\sum_{n=1}^{\infty}\dfrac{1}{\lambda_n}\sin(\lambda_n x)\sin(\lambda_n\xi)\sin(\lambda_n at)$	$\lambda_n = \dfrac{\pi(2n+1)}{2l}$

where the λ_n and $\varphi_n(x)$ are the eigenvalues and the corresponding eigenfunctions of the Sturm–Liouville problem for the linear ordinary differential equation (M12.5.1.5) with the homogeneous linear boundary conditions (M12.5.1.7); $\rho(y)$ and $\|\varphi_n\|$ are determined by (M12.7.1.15), and $\psi = \psi_n(t,\tau)$ is the solution of equation (M12.5.2.5) with $\alpha(t) = 1$ and $\lambda = \lambda_n$ that satisfies the initial conditions

$$\psi = 0 \quad \text{at} \quad t = \tau, \qquad \psi'_t = 1 \quad \text{at} \quad t = \tau.$$

In the special case $\beta(t) = \gamma(t) = 0$, we have $\psi_n(t,\tau) = \lambda_n^{-1/2}\sin\bigl[\lambda_n^{1/2}(t-\tau)\bigr]$.

Table M12.6 lists Green's functions for some problems for a nonhomogeneous wave equation, which corresponds to $a(x,t) = a^2 = \text{const}$, $b(x) = c(x) = \beta(t) = \gamma(t) \equiv 0$ in (M12.8.1.7). One should substitute $G(x,y,t-\tau)$ from this table for $G(x,y,t,\tau)$ in formula (M12.8.1.3).

M12.8.2. Problems for Hyperbolic Equation $\rho(x)\frac{\partial^2 w}{\partial t^2} = \frac{\partial}{\partial x}\Bigl[p(x)\frac{\partial w}{\partial x}\Bigr] - q(x)w + \Phi(x,t)$

► **General relations for solving nonhomogeneous boundary value problems.** Consider a linear hyperbolic equation of the special form

$$\rho(x)\frac{\partial^2 w}{\partial t^2} = \frac{\partial}{\partial x}\biggl[p(x)\frac{\partial w}{\partial x}\biggr] - q(x)w + \Phi(x,t). \tag{M12.8.2.1}$$

It is assumed that the functions ρ, p, p'_x, and q are continuous and the inequalities $\rho > 0$, $p > 0$ hold for $x_1 \le x \le x_2$.

The solution of equation (M12.8.2.1) under the general initial conditions (M12.8.1.2) and the arbitrary linear nonhomogeneous boundary conditions (M12.7.1.4)–(M12.7.1.5) can be represented as the sum

$$\begin{aligned} w(x,t) &= \int_0^t\int_{x_1}^{x_2}\Phi(\xi,\tau)\mathcal{G}(x,\xi,t-\tau)\,d\xi\,d\tau \\ &\quad + \frac{\partial}{\partial t}\int_{x_1}^{x_2}\rho(\xi)f_0(\xi)\mathcal{G}(x,\xi,t)\,d\xi + \int_{x_1}^{x_2}\rho(\xi)f_1(\xi)\mathcal{G}(x,\xi,t)\,d\xi \\ &\quad + p(x_1)\int_0^t g_1(\tau)\Lambda_1(x,t-\tau)\,d\tau + p(x_2)\int_0^t g_2(\tau)\Lambda_2(x,t-\tau)\,d\tau. \end{aligned} \tag{M12.8.2.2}$$

Here, the modified Green's function is determined by

$$\mathcal{G}(x,\xi,t)=\sum_{n=1}^{\infty}\frac{\varphi_n(x)\varphi_n(\xi)\sin\bigl(t\sqrt{\lambda_n}\,\bigr)}{\|\varphi_n\|^2\sqrt{\lambda_n}},\qquad \|\varphi_n\|^2=\int_{x_1}^{x_2}\rho(x)\varphi_n^2(x)\,dx,\quad \text{(M12.8.2.3)}$$

where the λ_n and $\varphi_n(x)$ are the eigenvalues and the corresponding eigenfunctions of the Sturm–Liouville problem for the second-order linear ordinary differential equation:

$$\begin{gathered}[p(x)\varphi'_x]'_x+[\lambda\rho(x)-q(x)]\varphi=0,\\ s_1\varphi'_x+k_1\varphi=0\quad\text{at}\quad x=x_1,\\ s_2\varphi'_x+k_2\varphi=0\quad\text{at}\quad x=x_2.\end{gathered}\qquad \text{(M12.8.2.4)}$$

The functions $\Lambda_1(x,t)$ and $\Lambda_2(x,t)$ that occur in the integrands of the last two terms in solution (M12.8.2.2) are expressed in terms of the Green's function of (M12.8.2.3).

Remark. The corresponding formulas for $\Lambda_m(x,t)$ for the basic types of boundary value problems can be obtained using Table M12.4; to do so, one should set $G=\mathcal{G}$ in all formulas of the last column and omit the last argument τ.

► **Properties of the Sturm–Liouville problem. The Klein–Gordon equation.** The general and special properties of the Sturm–Liouville problem (M12.8.2.4) are given in Subsections M12.5.1 and M11.2.5; various asymptotic and approximate formulas for eigenvalues and eigenfunctions can also be found there.

Example. Consider the second boundary value problem in the domain $0\le x\le l$ for the Klein–Gordon equation

$$\frac{\partial^2 w}{\partial t^2}=a^2\frac{\partial^2 w}{\partial x^2}-bw,$$

under the initial conditions (M12.8.1.2) and boundary conditions

$$\begin{aligned}\partial_x w&=g_1(t)\quad\text{at}\quad x=0,\\ \partial_x w&=g_2(t)\quad\text{at}\quad x=l.\end{aligned}$$

The Klein–Gordon equation is a special case of equation (M12.8.2.1) with $\rho(x)=1$, $p(x)=a^2$, $q(x)=b$, and $\Phi(x,t)=0$. The corresponding Sturm–Liouville problem (M12.8.2.4) has the form

$$a^2\varphi''_{xx}+(\lambda-b)\varphi=0,\qquad \varphi'_x=0\quad\text{at}\quad x=0,\qquad \varphi'_x=0\quad\text{at}\quad x=l.$$

The eigenfunctions and eigenvalues are found to be

$$\varphi_{n+1}(x)=\cos\left(\frac{\pi nx}{l}\right),\qquad \lambda_{n+1}=b+\frac{a\pi^2n^2}{l^2},\qquad n=0,\ 1,\ \dots$$

Using formula (M12.8.2.3) and taking into account that $\|\varphi_1\|^2=l$ and $\|\varphi_n\|^2=l/2$ $(n=1,\ 2,\ \dots)$, we find the Green's function:

$$\mathcal{G}(x,\xi,t)=\frac{1}{l\sqrt{b}}\sin\bigl(t\sqrt{b}\,\bigr)+\frac{2}{l}\sum_{n=1}^{\infty}\cos\left(\frac{\pi nx}{l}\right)\cos\left(\frac{\pi n\xi}{l}\right)\frac{\sin\bigl(t\sqrt{(a\pi n/l)^2+b}\,\bigr)}{\sqrt{(a\pi n/l)^2+b}}.$$

Substituting this expression into (M12.8.2.2) with $p(x_1)=p(x_2)=\rho(\xi)=1$, $x_1=0$, and $x_2=l$ and taking into account the formulas

$$\Lambda_1(x,t)=-\mathcal{G}(x,x_1,t),\qquad \Lambda_2(x,t)=\mathcal{G}(x,x_2,t)$$

(see the second row in Table M12.4 and the above remark), one obtains the solution to the problem in question.

M12.8.3. Goursat Problem (a Problem with Initial Data of Characteristics)

► **Statement of the Goursat problem. Basic property of the solution.** Consider the general linear hyperbolic equation in two independent variables which is reduced to the first canonical form (see Subsection M12.2.1):

$$\frac{\partial^2 w}{\partial x \partial y} + a(x,y)\frac{\partial w}{\partial x} + b(x,y)\frac{\partial w}{\partial y} + c(x,y)w = f(x,y), \qquad \text{(M12.8.3.1)}$$

where $a(x,y)$, $b(x,y)$, $c(x,y)$, and $f(x,y)$ are continuous functions.

The *Goursat problem* for equation (M12.8.3.1) is stated as follows: find a solution to equation (M12.8.3.1) that satisfies the conditions at characteristics

$$w(x,y)|_{x=x_1} = g(y), \qquad w(x,y)|_{y=y_1} = h(x), \qquad \text{(M12.8.3.2)}$$

where $g(y)$ and $h(x)$ are given continuous functions that match each other at the point of intersection of the characteristics, so that

$$g(y_1) = h(x_1).$$

Basic properties of the Goursat problem: the value of the solution at any point $M(x_0,y_0)$ depends only on the values of $g(y)$ at the segment AN (which is part of the characteristic $x = x_1$), the values of $h(x)$ at the segment BN (which is part of the characteristic $y = y_1$), and the values of the functions $a(x,y)$, $b(x,y)$, $c(x,y)$, and $f(x,y)$ in the rectangle $NAMB$; see Fig. M12.2. The domain of influence on the solution at the point $M(x_0,y_0)$ is shaded for clarity.

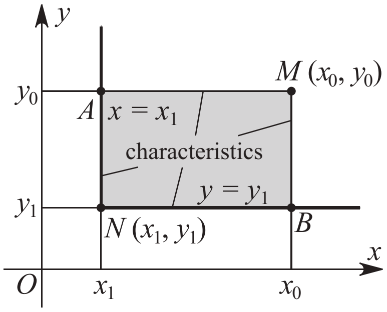

Figure M12.2. Domain of influence of the solution to the Goursat problem at a point M.

► **Riemann function.** A *Riemann function*, $\mathcal{R} = \mathcal{R}(x,y;x_0,y_0)$, corresponding to equation (M12.8.3.1) is defined as a solution to the equation

$$\frac{\partial^2 \mathcal{R}}{\partial x \partial y} - \frac{\partial}{\partial x}\Big[a(x,y)\mathcal{R}\Big] - \frac{\partial}{\partial y}\Big[b(x,y)\mathcal{R}\Big] + c(x,y)\mathcal{R} = 0 \qquad \text{(M12.8.3.3)}$$

that satisfies the conditions

$$\mathcal{R} = \exp\left[\int_{y_0}^{y} a(x_0,\xi)\,d\xi\right] \quad \text{at} \quad x = x_0, \qquad \mathcal{R} = \exp\left[\int_{x_0}^{x} b(\xi,y_0)\,d\xi\right] \quad \text{at} \quad y = y_0 \qquad \text{(M12.8.3.4)}$$

at the characteristics $x = x_0$ and $y = y_0$. Here, (x_0,y_0) is an arbitrary point from the domain of definition of equation (M12.8.3.1). The x_0 and y_0 appear in problem (M12.8.3.3), (M12.8.3.4) as parameters in the boundary conditions only.

THEOREM. *If the functions a, b, c and the partial derivatives a_x, b_y are all continuous, then the Riemann function $\mathcal{R}(x,y;x_0,y_0)$ exists. Moreover, the function $\mathcal{R}(x_0,y_0,x,y)$, obtained by swapping the parameters and the arguments, is a solution to the homogeneous equation (M12.8.3.1), with $f=0$.*

Remark. It is significant that the Riemann function is independent of the initial data on characteristics (M12.8.3.2).

Example 1. The Riemann function for the equation $w_{xy}=0$ is just $\mathcal{R}\equiv 1$.

Example 2. The Riemann function for the equation

$$w_{xy}+cw=0 \qquad (c=\text{const})$$

is expressed in terms of the Bessel function $J_0(z)$ as

$$\mathcal{R}=J_0\bigl(\sqrt{4c(x_0-x)(y_0-y)}\,\bigr).$$

► **Solution representation for the Goursat problem in terms of the Riemann function.** Given a Riemann function, the solution to the Goursat problem (M12.8.3.1), (M12.8.3.2) at any point (x_0,y_0) can be written as

$$w(x_0,y_0)=(w\mathcal{R})_N+\int_N^A \mathcal{R}\bigl(g'_y+bg\bigr)\,dy+\int_N^B \mathcal{R}\bigl(h'_x+ah\bigr)\,dx+\iint_{NAMB} f\mathcal{R}\,dx\,dy.$$

The first term on the right-hand side is evaluated at the point of intersection of the characteristics (x_1,y_1). The second and third terms are integrals along the characteristics $y=y_1$ $(x_1\le x\le x_0)$ and $x=x_1$ $(y_1\le y\le y_0)$; these involve the initial data of (M12.8.3.2). The last integral is taken over the rectangular domain $NAMB$ defined by the inequalities $x_1\le x\le x_0$, $y_1\le y\le y_0$.

The Goursat problem for a hyperbolic equation reduced to the second canonical form (see Subsection M12.2.1) is treated similarly.

Example 3. Consider the Goursat problem for the wave equation

$$\frac{\partial^2 w}{\partial t^2}-a^2\frac{\partial^2 w}{\partial x^2}=0$$

with the boundary conditions prescribed on its characteristics

$$\begin{aligned} w&=f(x) \quad \text{for} \quad x-at=0 \quad (0\le x\le b),\\ w&=g(x) \quad \text{for} \quad x+at=0 \quad (0\le x\le c), \end{aligned} \qquad \text{(M12.8.3.5)}$$

where $f(0)=g(0)$.

Substituting the values set on the characteristics (M12.8.3.5) into the general solution of the wave equation, $w=\varphi(x-at)+\psi(x+at)$, we arrive to a system of linear algebraic equations for $\varphi(x)$ and $\psi(x)$. As a result, the solution to the Goursat problem is obtained in the form

$$w(x,t)=f\left(\frac{x+at}{2}\right)+g\left(\frac{x-at}{2}\right)-f(0).$$

The propagation domain of the solution is the parallelogram bounded by the four straight lines

$$x-at=0, \quad x+at=0, \quad x-at=2c, \quad x+at=2b.$$

M12.9. Boundary Value Problems for Elliptic Equations with Two Space Variables

M12.9.1. Problems and the Green's Functions for Equation $a(x)\dfrac{\partial^2 w}{\partial x^2}+\dfrac{\partial^2 w}{\partial y^2}+b(x)\dfrac{\partial w}{\partial x}+c(x)w=-\Phi(x,y)$

► **Statements of boundary value problems.** Consider two-dimensional boundary value problems for the equation

$$a(x)\frac{\partial^2 w}{\partial x^2}+\frac{\partial^2 w}{\partial y^2}+b(x)\frac{\partial w}{\partial x}+c(x)w=-\Phi(x,y) \qquad \text{(M12.9.1.1)}$$

with general boundary conditions in x,

$$\begin{aligned} s_1\frac{\partial w}{\partial x}-k_1w&=f_1(y) \quad \text{at} \quad x=x_1,\\ s_2\frac{\partial w}{\partial x}+k_2w&=f_2(y) \quad \text{at} \quad x=x_2, \end{aligned} \qquad \text{(M12.9.1.2)}$$

and different boundary conditions in y. It is assumed that the coefficients of equation (M12.9.1.1) and the boundary conditions (M12.9.1.2) meet the requirements

$$a(x),\ b(x),\ c(x) \text{ are continuous } (x_1\le x\le x_2); \quad a>0, \quad |s_1|+|k_1|>0, \quad |s_2|+|k_2|>0.$$

► **Relations for the Green's function.** In the general case, the Green's function can be represented as

$$G(x,y,\xi,\eta)=\rho(\xi)\sum_{n=1}^{\infty}\frac{\varphi_n(x)\varphi_n(\xi)}{\|\varphi_n\|^2}\Psi_n(y,\eta;\lambda_n), \qquad \text{(M12.9.1.3)}$$

where

$$\rho(x)=\frac{1}{a(x)}\exp\left[\int\frac{b(x)}{a(x)}\,dx\right], \quad \|\varphi_n\|^2=\int_{x_1}^{x_2}\rho(x)\varphi_n^2(x)\,dx, \qquad \text{(M12.9.1.4)}$$

and the λ_n and $\varphi_n(x)$ are the eigenvalues and eigenfunctions of the homogeneous boundary value problem for the ordinary differential equation

$$a(x)\varphi''_{xx}+b(x)\varphi'_x+[\lambda+c(x)]\varphi=0, \qquad \text{(M12.9.1.5)}$$

$$s_1\varphi'_x-k_1\varphi=0 \quad \text{at} \quad x=x_1, \qquad \text{(M12.9.1.6)}$$

$$s_2\varphi'_x+k_2\varphi=0 \quad \text{at} \quad x=x_2. \qquad \text{(M12.9.1.7)}$$

The functions Ψ_n for various boundary conditions in y are specified in Table M12.7.

Example. Consider a boundary value problem for the Laplace equation

$$\frac{\partial^2 w}{\partial x^2}+\frac{\partial^2 w}{\partial y^2}=0 \qquad \text{(M12.9.1.8)}$$

in a strip $0\le x\le l$, $-\infty<y<\infty$ with mixed boundary conditions

$$w=f_1(y) \quad \text{at} \quad x=0, \qquad \frac{\partial w}{\partial x}=f_2(y) \quad \text{at} \quad x=l.$$

TABLE M12.7

The functions Ψ_n in (M12.9.1.3) for various boundary conditions. Notation: $\sigma_n = \sqrt{\lambda_n}$.

Domain	Boundary conditions	Function $\Psi_n(y, \eta; \lambda_n)$
$-\infty < y < \infty$	$\lvert w\rvert < \infty$ for $y \to \pm\infty$	$\frac{1}{2\sigma_n} e^{-\sigma_n \lvert y-\eta\rvert}$
$0 \le y < \infty$	$w = 0$ for $y = 0$	$\frac{1}{\sigma_n}\begin{cases} e^{-\sigma_n y}\sinh(\sigma_n\eta) & \text{for } y > \eta, \\ e^{-\sigma_n \eta}\sinh(\sigma_n y) & \text{for } \eta > y \end{cases}$
$0 \le y < \infty$	$\partial_y w = 0$ for $y = 0$	$\frac{1}{\sigma_n}\begin{cases} e^{-\sigma_n y}\cosh(\sigma_n\eta) & \text{for } y > \eta, \\ e^{-\sigma_n \eta}\cosh(\sigma_n y) & \text{for } \eta > y \end{cases}$
$0 \le y < \infty$	$\partial_y w - k_3 w = 0$ for $y = 0$	$\frac{1}{\sigma_n(\sigma_n + k_3)}\begin{cases} e^{-\sigma_n y}[\sigma_n\cosh(\sigma_n\eta) + k_3\sinh(\sigma_n\eta)] & \text{for } y > \eta, \\ e^{-\sigma_n \eta}[\sigma_n\cosh(\sigma_n y) + k_3\sinh(\sigma_n y)] & \text{for } \eta > y \end{cases}$
$0 \le y \le h$	$w = 0$ at $y = 0$, $w = 0$ at $y = h$	$\frac{1}{\sigma_n \sinh(\sigma_n h)}\begin{cases} \sinh(\sigma_n\eta)\sinh[\sigma_n(h-y)] & \text{for } y > \eta, \\ \sinh(\sigma_n y)\sinh[\sigma_n(h-\eta)] & \text{for } \eta > y \end{cases}$
$0 \le y \le h$	$\partial_y w = 0$ at $y = 0$, $\partial_y w = 0$ at $y = h$	$\frac{1}{\sigma_n \sinh(\sigma_n h)}\begin{cases} \cosh(\sigma_n\eta)\cosh[\sigma_n(h-y)] & \text{for } y > \eta, \\ \cosh(\sigma_n y)\cosh[\sigma_n(h-\eta)] & \text{for } \eta > y \end{cases}$
$0 \le y \le h$	$w = 0$ at $y = 0$, $\partial_y w = 0$ at $y = h$	$\frac{1}{\sigma_n \cosh(\sigma_n h)}\begin{cases} \sinh(\sigma_n\eta)\cosh[\sigma_n(h-y)] & \text{for } y > \eta, \\ \sinh(\sigma_n y)\cosh[\sigma_n(h-\eta)] & \text{for } \eta > y \end{cases}$

This equation is a special case of equation (M12.9.1.1) with $a(x) = 1$ and $b(x) = c(x) = \Phi(x,t) = 0$. The corresponding Sturm–Liouville problem (M12.9.1.5)–(M12.9.1.7) is written as

$$u''_{xx} + \lambda u = 0, \qquad u = 0 \quad \text{at} \quad x = 0, \qquad u'_x = 0 \quad \text{at} \quad x = l.$$

The eigenfunctions and eigenvalues are found as

$$\varphi_n(x) = \sin\left[\frac{\pi(2n-1)x}{l}\right], \qquad \lambda_n = \frac{\pi^2(2n-1)^2}{l^2}, \qquad n = 1,\ 2,\ \ldots$$

Using formulas (M12.9.1.3) and (M12.9.1.4) and taking into account the identities $\rho(\xi) = 1$ and $\|y_n\|^2 = l/2$ ($n = 1,\ 2,\ \ldots$) and the expression for Ψ_n from the first row in Table M12.7, we obtain the Green's function in the form

$$G(x,y,\xi,\eta) = \frac{1}{l}\sum_{n=1}^{\infty}\frac{1}{\sigma_n}\sin(\sigma_n x)\sin(\sigma_n \xi)e^{-\sigma_n|y-\eta|}, \qquad \sigma_n = \sqrt{\lambda_n} = \frac{\pi(2n-1)}{l}.$$

M12.9.2. Representation of Solutions to Boundary Value Problems via the Green's Functions

► **First boundary value problem.** The solution of the first boundary value problem for equation (M12.9.1.1) with the boundary conditions

$$w = f_1(y) \quad \text{at} \quad x = x_1, \qquad w = f_2(y) \quad \text{at} \quad x = x_2,$$
$$w = f_3(x) \quad \text{at} \quad y = 0, \qquad w = f_4(x) \quad \text{at} \quad y = h$$

is expressed in terms of the Green's function as

$$\begin{aligned} w(x,y) = {}& a(x_1)\int_0^h f_1(\eta)\left[\frac{\partial}{\partial\xi}G(x,y,\xi,\eta)\right]_{\xi=x_1} d\eta - a(x_2)\int_0^h f_2(\eta)\left[\frac{\partial}{\partial\xi}G(x,y,\xi,\eta)\right]_{\xi=x_2} d\eta \\ &+ \int_{x_1}^{x_2} f_3(\xi)\left[\frac{\partial}{\partial\eta}G(x,y,\xi,\eta)\right]_{\eta=0} d\xi - \int_{x_1}^{x_2} f_4(\xi)\left[\frac{\partial}{\partial\eta}G(x,y,\xi,\eta)\right]_{\eta=h} d\xi \\ &+ \int_{x_1}^{x_2}\int_0^h \Phi(\xi,\eta)G(x,y,\xi,\eta)\,d\eta\,d\xi. \end{aligned}$$

► **Second boundary value problem.** The solution of the second boundary value problem for equation (M12.9.1.1) with the boundary conditions

$$\begin{aligned} &\partial_x w = f_1(y) \quad \text{at} \quad x = x_1, \qquad \partial_x w = f_2(y) \quad \text{at} \quad x = x_2, \\ &\partial_y w = f_3(x) \quad \text{at} \quad y = 0, \qquad \partial_y w = f_4(x) \quad \text{at} \quad y = h \end{aligned}$$

is expressed in terms of the Green's function as

$$\begin{aligned} w(x,y) = &-a(x_1)\int_0^h f_1(\eta)G(x,y,x_1,\eta)\,d\eta + a(x_2)\int_0^h f_2(\eta)G(x,y,x_2,\eta)\,d\eta \\ &-\int_{x_1}^{x_2} f_3(\xi)G(x,y,\xi,0)\,d\xi + \int_{x_1}^{x_2} f_4(\xi)G(x,y,\xi,h)\,d\xi \\ &+\int_{x_1}^{x_2}\int_0^h \Phi(\xi,\eta)G(x,y,\xi,\eta)\,d\eta\,d\xi. \end{aligned}$$

M12.9.3. Solutions of Problems for the Laplace Equation in the Polar Coordinate System

In the polar coordinates, defined by $x = r\cos\varphi$ and $y = r\sin\varphi$, Laplace's equation (M12.9.1.8) becomes

$$\Delta w \equiv \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial w}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 w}{\partial \varphi^2} = 0. \qquad \text{(M12.9.3.1)}$$

► **First boundary value problem ($0 \le r \le R$).** A circle is considered. A boundary condition is prescribed:

$$w = f(\varphi) \quad \text{at} \quad r = R. \qquad \text{(M12.9.3.2)}$$

The solution to equation (M12.9.3.1) subject to the boundary condition (M12.9.3.2) is given by

$$w(r,\varphi) = \frac{1}{2\pi}\int_0^{2\pi} f(\psi)\frac{R^2 - r^2}{r^2 - 2Rr\cos(\varphi - \psi) + R^2}\,d\psi.$$

This formula is conventionally referred to as the *Poisson integral*.

► **Second boundary value problem ($0 \le r \le R$).** A circle is considered. A boundary condition is prescribed:

$$\partial_r w = f(\varphi) \quad \text{at} \quad r = R. \qquad \text{(M12.9.3.3)}$$

The solution to equation (M12.9.3.1) subject to the boundary condition (M12.9.3.3) is expressed as

$$w(r,\varphi) = \frac{R}{2\pi}\int_0^{2\pi} f(\psi)\ln\frac{r^2 - 2Rr\cos(\varphi - \psi) + R^2}{R^2}\,d\psi + C,$$

where C is an arbitrary constant; this formula is known as the *Dini integral*.

Remark. The function $f(\varphi)$ must satisfy the solvability condition $\int_0^{2\pi} f(\varphi)\,d\varphi = 0$.

M12.10. Problems with Many Space Variables. Representation of Solutions via Green's Functions

M12.10.1. Boundary Value Problems for Parabolic Equations

► **Statement of the problem.** In general, a nonhomogeneous linear differential equation of the parabolic type in n space variables has the form

$$\frac{\partial w}{\partial t} - L_{\mathbf{x},t}[w] = \Phi(\mathbf{x},t), \qquad \text{(M12.10.1.1)}$$

where

$$\begin{gathered} L_{\mathbf{x},t}[w] \equiv \sum_{i,j=1}^{n} a_{ij}(\mathbf{x},t)\frac{\partial^2 w}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(\mathbf{x},t)\frac{\partial w}{\partial x_i} + c(\mathbf{x},t)w, \\ \mathbf{x} = \{x_1,\dots,x_n\}, \quad \sum_{i,j=1}^{n} a_{ij}(\mathbf{x},t)\xi_i\xi_j \ge \sigma \sum_{i=1}^{n}\xi_i^2, \quad \sigma > 0. \end{gathered} \qquad \text{(M12.10.1.2)}$$

Let V be some simply connected domain in $\mathbb{R}^n$ with a sufficiently smooth boundary S. We consider the nonstationary boundary value problem for equation (M12.10.1.1) in the domain V with an arbitrary initial condition,

$$w = f(\mathbf{x}) \quad \text{at} \quad t = 0, \qquad \text{(M12.10.1.3)}$$

and nonhomogeneous linear boundary conditions,

$$\Gamma_{\mathbf{x},t}[w] = g(\mathbf{x},t) \quad \text{for} \quad \mathbf{x} \in S. \qquad \text{(M12.10.1.4)}$$

In the general case, $\Gamma_{\mathbf{x},t}$ is a first-order linear differential operator in the space coordinates with coefficients dependent on $\mathbf{x}$ and t.

► **Representation of the problem solution in terms of the Green's function.** The solution of the nonhomogeneous linear boundary value problem defined by (M12.10.1.1)–(M12.10.1.4) can be represented as the sum

$$\begin{aligned} w(\mathbf{x},t) = {} & \int_0^t \int_V \Phi(\mathbf{y},\tau) G(\mathbf{x},\mathbf{y},t,\tau)\, dV_y\, d\tau + \int_V f(\mathbf{y}) G(\mathbf{x},\mathbf{y},t,0)\, dV_y \\ & + \int_0^t \int_S g(\mathbf{y},\tau) H(\mathbf{x},\mathbf{y},t,\tau)\, dS_y\, d\tau, \end{aligned} \qquad \text{(M12.10.1.5)}$$

where $G(\mathbf{x},\mathbf{y},t,\tau)$ is the Green's function; for $t > \tau \ge 0$, it satisfies the homogeneous equation

$$\frac{\partial G}{\partial t} - L_{\mathbf{x},t}[G] = 0 \qquad \text{(M12.10.1.6)}$$

with the nonhomogeneous initial condition of special form

$$G = \delta(\mathbf{x}-\mathbf{y}) \quad \text{at} \quad t = \tau \qquad \text{(M12.10.1.7)}$$

and the homogeneous boundary condition

$$\Gamma_{\mathbf{x},t}[G]=0 \quad \text{for} \quad \mathbf{x}\in S. \qquad \text{(M12.10.1.8)}$$

The vector $\mathbf{y}=\{y_1,\dots,y_n\}$ appears in problem (M12.10.1.6)–(M12.10.1.8) as an n-dimensional free parameter ($\mathbf{y}\in V$), and $\delta(\mathbf{x}-\mathbf{y})=\delta(x_1-y_1)\dots\delta(x_n-y_n)$ is the n-dimensional Dirac delta function. The Green's function G is independent of the functions Φ, f, and g that characterize various nonhomogeneities of the boundary value problem. In (M12.10.1.5), the integration is performed everywhere with respect to $\mathbf{y}$, with $dV_y=dy_1\dots dy_n$.

The function $H(\mathbf{x},\mathbf{y},t,\tau)$ in the integrand of the last term in solution (M12.10.1.5) can be expressed via the Green's function $G(\mathbf{x},\mathbf{y},t,\tau)$. The corresponding formulas for $H(\mathbf{x},\mathbf{y},t,\tau)$ are given in Table M12.8 for the three basic types of boundary value problems; in the third boundary value problem, the coefficient k can depend on $\mathbf{x}$ and t. The boundary conditions of the second and third kinds, as well as the solution of the first boundary value problem, involve operators of differentiation along the conormal of operator (M12.10.1.2); these operators act as follows:

$$\frac{\partial G}{\partial M_x}\equiv\sum_{i,j=1}^{n}a_{ij}(\mathbf{x},t)N_j\frac{\partial G}{\partial x_i},\qquad \frac{\partial G}{\partial M_y}\equiv\sum_{i,j=1}^{n}a_{ij}(\mathbf{y},\tau)N_j\frac{\partial G}{\partial y_i}, \qquad \text{(M12.10.1.9)}$$

where $\mathbf{N}=\{N_1,\dots,N_n\}$ is the unit outward normal to the surface S. In the special case where $a_{ii}(\mathbf{x},t)=1$ and $a_{ij}(\mathbf{x},t)=0$ for $i\neq j$, operator (M12.10.1.9) coincides with the ordinary operator of differentiation along the outward normal to S.

TABLE M12.8
The form of the function $H(\mathbf{x},\mathbf{y},t,\tau)$ for the basic types of nonstationary boundary value problems.

Type of problem	Form of boundary condition (M12.10.1.4)	Function $H(\mathbf{x},\mathbf{y},t,\tau)$
First boundary value problem	$w=g(\mathbf{x},t)$ for $\mathbf{x}\in S$	$H(\mathbf{x},\mathbf{y},t,\tau)=-\frac{\partial G}{\partial M_y}(\mathbf{x},\mathbf{y},t,\tau)$
Second boundary value problem	$\frac{\partial w}{\partial M_x}=g(\mathbf{x},t)$ for $\mathbf{x}\in S$	$H(\mathbf{x},\mathbf{y},t,\tau)=G(\mathbf{x},\mathbf{y},t,\tau)$
Third boundary value problem	$\frac{\partial w}{\partial M_x}+kw=g(\mathbf{x},t)$ for $\mathbf{x}\in S$	$H(\mathbf{x},\mathbf{y},t,\tau)=G(\mathbf{x},\mathbf{y},t,\tau)$

If the coefficient of equation (M12.10.1.6) and the boundary condition (M12.10.1.8) are independent of t, then the Green's function depends on only three arguments, $G(\mathbf{x},\mathbf{y},t,\tau)=G(\mathbf{x},\mathbf{y},t-\tau)$.

Remark. Let S_i ($i=1,\dots,p$) be different portions of the surface S such that $S=\sum_{i=1}^{p}S_i$ and let boundary conditions of various types be set on the S_i,

$$\Gamma^{(i)}_{\mathbf{x},t}[w]=g_i(\mathbf{x},t) \quad \text{for} \quad \mathbf{x}\in S_i, \qquad i=1,\dots,p. \qquad \text{(M12.10.1.10)}$$

Then formula (M12.10.1.5) remains valid but the last term in (M12.10.1.5) must be replaced by the sum

$$\sum_{i=1}^{p}\int_0^t\int_{S_i}g_i(\mathbf{y},\tau)H_i(\mathbf{x},\mathbf{y},t,\tau)\,dS_y\,d\tau. \qquad \text{(M12.10.1.11)}$$

M12.10.2. Boundary Value Problems for Hyperbolic Equations

► **Statement of the problem.** The general nonhomogeneous linear differential hyperbolic equation in n space variables can be written as

$$\frac{\partial^2 w}{\partial t^2} + \varphi(\mathbf{x},t)\frac{\partial w}{\partial t} - L_{\mathbf{x},t}[w] = \Phi(\mathbf{x},t), \qquad \text{(M12.10.2.1)}$$

where the operator $L_{\mathbf{x},t}[w]$ is defined explicitly in (M12.10.1.2).

We consider the nonstationary boundary value problem for equation (M12.10.2.1) in the domain V with arbitrary initial conditions,

$$w = f_0(\mathbf{x}) \quad \text{at} \quad t=0, \qquad \text{(M12.10.2.2)}$$

$$\partial_t w = f_1(\mathbf{x}) \quad \text{at} \quad t=0, \qquad \text{(M12.10.2.3)}$$

and the nonhomogeneous linear boundary condition (M12.10.1.4).

► **Representation of the problem solution in terms of the Green's function.** The solution of the nonhomogeneous linear boundary value problem defined by (M12.10.2.1)–(M12.10.2.3), (M12.10.1.4) can be represented as the sum

$$\begin{aligned} w(\mathbf{x},t) &= \int_0^t \int_V \Phi(\mathbf{y},\tau) G(\mathbf{x},\mathbf{y},t,\tau)\, dV_y\, d\tau - \int_V f_0(\mathbf{y}) \left[\frac{\partial}{\partial \tau} G(\mathbf{x},\mathbf{y},t,\tau)\right]_{\tau=0} dV_y \\ &+ \int_V \bigl[f_1(\mathbf{y}) + f_0(\mathbf{y})\varphi(\mathbf{y},0)\bigr] G(\mathbf{x},\mathbf{y},t,0)\, dV_y \\ &+ \int_0^t \int_S g(\mathbf{y},\tau) H(\mathbf{x},\mathbf{y},t,\tau)\, dS_y\, d\tau. \end{aligned} \qquad \text{(M12.10.2.4)}$$

Here, $G(\mathbf{x},\mathbf{y},t,\tau)$ is the Green's function; for $t > \tau \geq 0$, it satisfies the homogeneous equation

$$\frac{\partial^2 G}{\partial t^2} + \varphi(\mathbf{x},t)\frac{\partial G}{\partial t} - L_{\mathbf{x},t}[G] = 0 \qquad \text{(M12.10.2.5)}$$

with the semihomogeneous initial conditions

$$\begin{aligned} G &= 0 && \text{at} \quad t=\tau, \\ \partial_t G &= \delta(\mathbf{x}-\mathbf{y}) && \text{at} \quad t=\tau, \end{aligned}$$

and the homogeneous boundary condition (M12.10.1.8).

If the coefficients of equation (M12.10.2.5) and the boundary condition (M12.10.1.8) are independent of time t, then the Green's function depends on only three arguments, $G(\mathbf{x},\mathbf{y},t,\tau) = G(\mathbf{x},\mathbf{y},t-\tau)$. In this case, one can set $\frac{\partial}{\partial\tau} G(\mathbf{x},\mathbf{y},t,\tau)\big|_{\tau=0} = -\frac{\partial}{\partial t} G(\mathbf{x},\mathbf{y},t)$ in solution (M12.10.2.4).

The function $H(\mathbf{x},\mathbf{y},t,\tau)$ in the integrand of the last term in solution (M12.10.2.4) can be expressed via the Green's function $G(\mathbf{x},\mathbf{y},t,\tau)$. The corresponding formulas for H are given in Table M12.8 for the three basic types of boundary value problems; in the third boundary value problem, the coefficient k can depend on $\mathbf{x}$ and t.

M12.10.3. Boundary Value Problems for Elliptic Equations

► **Statement of the problem.** In general, a nonhomogeneous linear elliptic equation can be written as

$$-L_{\mathbf{x}}[w]=\Phi(\mathbf{x}),\qquad\text{(M12.10.3.1)}$$

where

$$L_{\mathbf{x}}[w]\equiv\sum_{i,j=1}^{n}a_{ij}(\mathbf{x})\frac{\partial^2 w}{\partial x_i\partial x_j}+\sum_{i=1}^{n}b_i(\mathbf{x})\frac{\partial w}{\partial x_i}+c(\mathbf{x})w.\qquad\text{(M12.10.3.2)}$$

We consider equation (M12.10.3.1) with operator (M12.10.3.2) in a domain V and assume that the equation is subject to the general linear boundary condition

$$\Gamma_{\mathbf{x}}[w]=g(\mathbf{x})\quad\text{for}\quad\mathbf{x}\in S.\qquad\text{(M12.10.3.3)}$$

The solution of the stationary problem (M12.10.3.1)–(M12.10.3.3) can be obtained by passing in (M12.10.1.5) to the limit as $t\to\infty$. To this end, one should start with equation (M12.10.1.1), whose coefficients are independent of t, and take the homogeneous initial condition (M12.10.1.3), with $f(\mathbf{x})=0$, and the stationary boundary condition (M12.10.3.3).

► **Representation of the problem solution in terms of the Green's function.** The solution of the linear boundary value problem (M12.10.3.1)–(M12.10.3.3) can be represented as the sum

$$w(\mathbf{x})=\int_V\Phi(\mathbf{y})G(\mathbf{x},\mathbf{y})\,dV_y+\int_S g(\mathbf{y})H(\mathbf{x},\mathbf{y})\,dS_y.\qquad\text{(M12.10.3.4)}$$

Here, the Green's function $G(\mathbf{x},\mathbf{y})$ satisfies the nonhomogeneous equation of special form

$$-L_{\mathbf{x}}[G]=\delta(\mathbf{x}-\mathbf{y})\qquad\text{(M12.10.3.5)}$$

with the homogeneous boundary condition

$$\Gamma_{\mathbf{x}}[G]=0\quad\text{for}\quad\mathbf{x}\in S.\qquad\text{(M12.10.3.6)}$$

The vector $\mathbf{y}=\{y_1,\dots,y_n\}$ appears in problem (M12.10.3.5), (M12.10.3.6) as an n-dimensional free parameter ($\mathbf{y}\in V$). Note that G is independent of the functions Φ and g characterizing various nonhomogeneities of the original boundary value problem.

The function $H(\mathbf{x},\mathbf{y})$ in the integrand of the second term in solution (M12.10.3.4) can be expressed via the Green's function $G(\mathbf{x},\mathbf{y})$. The corresponding formulas for H for the three basic types of boundary value problems can be obtained using Table M12.8, where one should consider the functions to be independent of t and τ. The boundary conditions of the second and third kinds, as well as the solution of the first boundary value problem, involve operators of differentiation along the conormal of operator (M12.10.3.2); these operators are defined by (M12.10.1.9); in this case, the coefficients a_{ij} depend on $\mathbf{x}$ only.

M12.10.4. Construction of the Green's Functions. General Formulas and Relations

Table M12.9 lists the Green's functions of boundary value problems for second-order equations of various types in a bounded domain V. It is assumed that $L_{\mathbf{x}}$ is a second-order linear (self-adjoint) differential operator of the form

$$L_{\mathbf{x}}[w]\equiv\sum_{i,j=1}^{n}\frac{\partial}{\partial x_i}\left[a_{ij}(\mathbf{x})\frac{\partial w}{\partial x_j}\right],\quad\text{where}\quad\sum_{i,j=1}^{n}a_{ij}(\mathbf{x})\xi_i\xi_j\ge\sigma\sum_{i=1}^{n}\xi_i^2,\quad\sigma>0,$$

TABLE M12.9
The Green's functions of boundary value problems for equations of various types in bounded domains. In all problems, the operators $L_{\mathbf{x}}$ and $\Gamma_{\mathbf{x}}$ are the same; $\mathbf{x} = \{x_1, \dots, x_n\}$.

Equation	Initial and boundary conditions	Green's function
Elliptic equation $-L_{\mathbf{x}}[w] = \Phi(\mathbf{x})$	$\Gamma_{\mathbf{x}}[w] = g(\mathbf{x})$ for $\mathbf{x} \in S$ (no initial condition required)	$G(\mathbf{x},\mathbf{y}) = \sum\limits_{k=1}^{\infty} \dfrac{u_k(\mathbf{x})u_k(\mathbf{y})}{\lVert u_k \rVert^2 \lambda_k}, \quad \lambda_k \neq 0$
Parabolic equation $\partial_t w - L_{\mathbf{x}}[w] = \Phi(\mathbf{x},t)$	$w = f(\mathbf{x})$ at $t = 0$ $\Gamma_{\mathbf{x}}[w] = g(\mathbf{x},t)$ for $\mathbf{x} \in S$	$G(\mathbf{x},\mathbf{y},t) = \sum\limits_{k=1}^{\infty} \dfrac{u_k(\mathbf{x})u_k(\mathbf{y})}{\lVert u_k \rVert^2} \exp(-\lambda_k t)$
Hyperbolic equation $\partial_{tt} w - L_{\mathbf{x}}[w] = \Phi(\mathbf{x},t)$	$w = f_0(\mathbf{x})$ at $t = 0$ $w = f_1(\mathbf{x})$ at $t = 0$ $\Gamma_{\mathbf{x}}[w] = g(\mathbf{x},t)$ for $\mathbf{x} \in S$	$G(\mathbf{x},\mathbf{y},t) = \sum\limits_{k=1}^{\infty} \dfrac{u_k(\mathbf{x})u_k(\mathbf{y})}{\lVert u_k \rVert^2 \sqrt{\lambda_k}} \sin\bigl(t\sqrt{\lambda_k}\bigr)$

in the space variables $x_1, \dots, x_n$, and $\Gamma_{\mathbf{x}}$ is a zeroth- or first-order linear boundary operator that can define a boundary condition of the first, second, or third kind; the coefficients of the operators $L_{\mathbf{x}}$ and $\Gamma_{\mathbf{x}}$ can depend on the space variables but are independent of time t. The coefficients λ_k and the functions $u_k(\mathbf{x})$ are determined by solving the homogeneous eigenvalue problem

$$L_{\mathbf{x}}[u] + \lambda u = 0, \qquad \text{(M12.11.1.1)}$$

$$\Gamma_{\mathbf{x}}[u] = 0 \quad \text{for} \quad \mathbf{x} \in S. \qquad \text{(M12.11.1.2)}$$

It is apparent from Table M12.9 that, given the Green's function in the problem for a parabolic (or hyperbolic) equation, one can easily construct the Green's functions of the corresponding problems for elliptic and hyperbolic (or parabolic) equations. In particular, the Green's function of the problem for an elliptic equation can be expressed via the Green's function of the problem for a parabolic equation as follows:

$$G_0(\mathbf{x},\mathbf{y}) = \int_0^{\infty} G_1(\mathbf{x},\mathbf{y},t)\,dt. \qquad \text{(M12.11.1.3)}$$

It is assumed here that $\lambda = 0$ is not an eigenvalue of problem (M12.11.1.1), (M12.11.1.2).

Remark 1. Formula (M12.11.1.3) can also be used if the domain V is infinite. In this case, one should make sure that the integral on the right-hand side is convergent.

Remark 2. Suppose the equations given in the first column of Table M12.9 contain $-L_{\mathbf{x}}[w] - \beta w$ instead of $-L_{\mathbf{x}}[w]$, with β being a free parameter. Then the λ_k in the expressions of the Green's function in the third column of Table M12.9 must be replaced by $\lambda_k - \beta$; just as previously, the λ_k and $u_k(\mathbf{x})$ were determined by solving the eigenvalue problem (M12.11.1.1), (M12.11.1.2).

M12.11. Duhamel's Principles in Nonstationary Problems

M12.11.1. Problems for Homogeneous Linear Equations

► **Parabolic equations with two independent variables.** Consider the problem for the homogeneous linear equation of parabolic type

$$\frac{\partial w}{\partial t} = a(x)\frac{\partial^2 w}{\partial x^2} + b(x)\frac{\partial w}{\partial x} + c(x)w \qquad \text{(M12.11.1.1)}$$

with the homogeneous initial condition

$$w = 0 \quad \text{at} \quad t = 0 \tag{M12.11.1.2}$$

and the boundary conditions

$$s_1 \partial_x w + k_1 w = g(t) \quad \text{at} \quad x = x_1, \tag{M12.11.1.3}$$

$$s_2 \partial_x w + k_2 w = 0 \quad \text{at} \quad x = x_2. \tag{M12.11.1.4}$$

By appropriately choosing the values of the coefficients s_1, s_2, k_1, and k_2 in (M12.11.1.3) and (M12.11.1.4), one can obtain the first, second, third, and mixed boundary value problems for equation (M12.11.1.1).

The solution of problem (M12.11.1.1)–(M12.11.1.4) with the nonstationary boundary condition (M12.11.1.3) at $x = x_1$ can be expressed by the formula (*Duhamel's first principle*)

$$w(x,t) = \frac{\partial}{\partial t} \int_0^t u(x, t - \tau)\, g(\tau)\, d\tau = \int_0^t \frac{\partial u}{\partial t}(x, t - \tau)\, g(\tau)\, d\tau \tag{M12.11.1.5}$$

in terms of the solution $u(x,t)$ of the auxiliary problem for equation (M12.11.1.1) with the initial and boundary conditions (M12.11.1.2) and (M12.11.1.4), for u instead of w, and the following simpler stationary boundary condition at $x = x_1$:

$$s_1 \partial_x u + k_1 u = 1 \quad \text{at} \quad x = x_1. \tag{M12.11.1.6}$$

Remark. A similar formula also holds for the homogeneous boundary condition at $x = x_1$ and a nonhomogeneous nonstationary boundary condition at $x = x_2$.

Example. Consider the first boundary value problem for the heat equation

$$\frac{\partial w}{\partial t} = \frac{\partial^2 w}{\partial x^2} \tag{M12.11.1.7}$$

with the homogeneous initial condition (M12.11.1.2) and the boundary condition

$$w = g(t) \quad \text{at} \quad x = 0. \tag{M12.11.1.8}$$

(The second boundary condition is not required, since $0 \le x < \infty$ in this case.)

First consider the following auxiliary problem for the heat equation with the homogeneous initial condition and a simpler boundary condition:

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \qquad u = 0 \quad \text{at} \quad t = 0, \qquad u = 1 \quad \text{at} \quad x = 0.$$

This problem has a self-similar solution of the form

$$w = w(z), \qquad z = x t^{-1/2},$$

where the function $w(z)$ is determined by the following ordinary differential equation and boundary conditions:

$$u''_{zz} + \tfrac{1}{2} z u'_z = 0, \qquad u = 1 \quad \text{at} \quad z = 0, \qquad u = 0 \quad \text{at} \quad z = \infty.$$

Its solution is expressed as

$$u(z) = \operatorname{erfc}\left(\frac{z}{2}\right) \quad \Longrightarrow \quad u(x,t) = \operatorname{erfc}\left(\frac{x}{2\sqrt{t}}\right),$$

where $\operatorname{erfc} z = \frac{2}{\sqrt{\pi}} \int_z^\infty \exp(-\xi^2)\, d\xi$ is the complementary error function. Substituting the obtained expression of $u(x,t)$ into (M12.11.1.5), we obtain the solution to the first boundary value problem for the heat equation (M12.11.1.7) with the initial condition (M12.11.1.2) and an arbitrary boundary condition (M12.11.1.8) in the form

$$w(x,t) = \frac{x}{2\sqrt{\pi}} \int_0^t \exp\left[-\frac{x^2}{4(t-\tau)}\right] \frac{g(\tau)\, d\tau}{(t-\tau)^{3/2}}.$$

► **Hyperbolic equations with two independent variables.** Consider the problem for the homogeneous linear hyperbolic equation

$$\frac{\partial^2 w}{\partial t^2} + \varphi(x)\frac{\partial w}{\partial t} = a(x)\frac{\partial^2 w}{\partial x^2} + b(x)\frac{\partial w}{\partial x} + c(x)w \tag{M12.11.1.9}$$

with the homogeneous initial conditions

$$\begin{aligned} w &= 0 \quad \text{at} \quad t = 0,\\ \partial_t w &= 0 \quad \text{at} \quad t = 0, \end{aligned} \tag{M12.11.1.10}$$

and the boundary conditions (M12.11.1.3) and (M12.11.1.4).

The solution of problem (M12.11.1.9), (M12.11.1.10), (M12.11.1.3), (M12.11.1.4) with the nonstationary boundary condition (M12.11.1.3) at $x = x_1$ can be expressed by formula (M12.11.1.5) in terms of the solution $u(x, t)$ of the auxiliary problem for equation (M12.11.1.9) with the initial conditions (M12.11.1.10) and boundary condition (M12.11.1.4), for u instead of w, and the simpler stationary boundary condition (M12.11.1.6) at $x = x_1$.

In this case, the remark made after the boundary condition (M12.11.1.6) remains valid.

M12.11.2. Problems for Nonhomogeneous Linear Equations

► **Parabolic equations.** The solution of the nonhomogeneous linear equation

$$\frac{\partial w}{\partial t} = \sum_{i,j=1}^{n} a_{ij}(\mathbf{x})\frac{\partial^2 w}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(\mathbf{x})\frac{\partial w}{\partial x_i} + c(\mathbf{x})w + \Phi(\mathbf{x}, t)$$

with the homogeneous initial condition (M12.11.1.2) and the homogeneous boundary condition

$$\Gamma_{\mathbf{x}}[w] = 0 \quad \text{for} \quad \mathbf{x} \in S \tag{M12.11.2.1}$$

can be represented in the form (*Duhamel's second principle*)

$$w(\mathbf{x}, t) = \int_0^t U(\mathbf{x}, t - \tau, \tau)\, d\tau. \tag{M12.11.2.2}$$

Here, $U(\mathbf{x}, t, \tau)$ is the solution of the auxiliary problem for the homogeneous equation

$$\frac{\partial U}{\partial t} = \sum_{i,j=1}^{n} a_{ij}(\mathbf{x})\frac{\partial^2 U}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(\mathbf{x})\frac{\partial U}{\partial x_i} + c(\mathbf{x})U$$

with the boundary condition (M12.11.2.1), in which w must be substituted by U, and the nonhomogeneous initial condition

$$U = \Phi(\mathbf{x}, \tau) \quad \text{at} \quad t = 0,$$

where τ is a parameter.

Note that (M12.11.2.1) can represent a boundary condition of the first, second, or third kind; the coefficients of the operator $\Gamma_{\mathbf{x}}$ are assumed to be independent of t.

► **Hyperbolic equations.** The solution of the nonhomogeneous linear equation

$$\frac{\partial^2 w}{\partial t^2}+\varphi(\mathbf{x})\frac{\partial w}{\partial t}=\sum_{i,j=1}^{n}a_{ij}(\mathbf{x})\frac{\partial^2 w}{\partial x_i\partial x_j}+\sum_{i=1}^{n}b_i(\mathbf{x})\frac{\partial w}{\partial x_i}+c(\mathbf{x})w+\Phi(\mathbf{x},t)$$

with the homogeneous initial conditions (M12.11.1.10) and homogeneous boundary condition (M12.11.2.1) can be expressed by formula (M12.11.2.2) in terms of the solution $U=U(\mathbf{x},t,\tau)$ of the auxiliary problem for the homogeneous equation

$$\frac{\partial^2 U}{\partial t^2}+\varphi(\mathbf{x})\frac{\partial U}{\partial t}=\sum_{i,j=1}^{n}a_{ij}(\mathbf{x})\frac{\partial^2 U}{\partial x_i\partial x_j}+\sum_{i=1}^{n}b_i(\mathbf{x})\frac{\partial U}{\partial x_i}+c(\mathbf{x})U$$

with the homogeneous initial and boundary conditions, (M12.11.1.2) and (M12.11.2.1), where w must be replaced by U, and the nonhomogeneous initial condition

$$\partial_t U=\Phi(\mathbf{x},\tau)\quad\text{at}\quad t=0,$$

where τ is a parameter.

Note that (M12.11.2.1) can represent a boundary condition of the first, second, or third kind.

Bibliography for Chapter M12

Akulenko, L. D. and Nesterov, S. V., *High Precision Methods in Eigenvalue Problems and Their Applications*, Chapman & Hall/CRC Press, Boca Raton, Florida, 2004.

Butkovskiy, A. G., *Green's Functions and Transfer Functions Handbook*, Halstead Press–John Wiley & Sons, New York, 1982.

Carslaw, H. S. and Jaeger, J. C., *Conduction of Heat in Solids*, Clarendon Press, Oxford, 1984.

Courant, R. and Hilbert, D., *Methods of Mathematical Physics, Vol. 2*, Wiley-Interscience, New York, 1989.

Ditkin, V. A. and Prudnikov, A. P., *Integral Transforms and Operational Calculus*, Pergamon Press, New York, 1965.

Duffy, D. G., *Transform Methods for Solving Partial Differential Equations, 2nd Edition*, Chapman & Hall/CRC Press, Boca Raton, Florida, 2004.

Farlow, S. J., *Partial Differential Equations for Scientists and Engineers*, John Wiley & Sons, New York, 1982.

Guenther, R. B. and Lee, J. W., *Partial Differential Equations of Mathematical Physics and Integral Equations*, Dover Publications, New York, 1996.

Haberman, R., *Elementary Applied Partial Differential Equations with Fourier Series and Boundary Value Problems*, Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

Kamke, E., *Differentialgleichungen: Lösungsmethoden und Lösungen, II, Partielle Differentialgleichungen Erster Ordnung für eine gesuchte Funktion*, Akad. Verlagsgesellschaft Geest & Portig, Leipzig, 1965.

Leis, R., *Initial-Boundary Value Problems in Mathematical Physics,* John Wiley & Sons, Chichester, 1986.

Mikhlin, S. G. (Editor), *Linear Equations of Mathematical Physics*, Holt, Rinehart and Winston, New York, 1967.

Moon, P. and Spencer, D. E., *Field Theory Handbook, Including Coordinate Systems, Differential Equations and Their Solutions, 3rd Edition*, Springer-Verlag, Berlin, 1988.

Morse, P. M. and Feshbach, H., *Methods of Theoretical Physics, Vols. 1 and 2*, McGraw-Hill, New York, 1953.

Petrovsky, I. G., *Lectures on Partial Differential Equations*, Dover Publications, New York, 1991.

Polyanin, A. D., *Handbook of Linear Partial Differential Equations for Engineers and Scientists*, Chapman & Hall/CRC Press, Boca Raton, Florida, 2002.

Polyanin, A. D., Zaitsev, V. F., and Moussiaux, A., *Handbook of First Order Partial Differential Equations*, Taylor & Francis, London, 2002.

Sneddon, I. N., *Fourier Transformations*, Dover Publications, New York, 1995.

Stakgold, I., *Boundary Value Problems of Mathematical Physics. Vols. 1 and 2*, Society for Industrial & Applied Mathematics, Philadelphia, 2000.

Strauss, W. A., *Partial Differential Equations. An Introduction*, John Wiley & Sons, New York, 1992.
Tikhonov, A. N. and Samarskii, A. A., *Equations of Mathematical Physics*, Dover Publications, New York, 1990.
Vladimirov, V. S., *Equations of Mathematical Physics*, Dekker, New York, 1971.
Zauderer, E., *Partial Differential Equations of Applied Mathematics*, Wiley-Interscience, New York, 1989.
Zwillinger, D., *Handbook of Differential Equations, 3rd Edition*, Academic Press, New York, 1997.

Chapter M13

Special Functions and Their Properties

Throughout Chapter M13 it is assumed that n is a positive integer unless otherwise specified.

M13.1. Some Coefficients, Symbols, and Numbers

M13.1.1. Factorials. Binomial Coefficients. Pochhammer Symbol

► **Factorial.**
$0! = 1! = 1,\ n! = 1\cdot 2\cdot 3\dots(n-1)n,\quad n = 2,\ 3,\ 4,\ \dots$

► **Double factorial.**
$0!! = 1!! = 1,\ (2n)!! = 2\cdot 4\cdot 6\dots(2n),\ (2n+1)!! = 1\cdot 3\cdot 5\dots(2n+1)$, where $n = 1,\ 2,\ 3,\ \dots$

► **Binomial coefficients.**

$$C_n^k = \binom{n}{k} = \frac{n!}{k!\,(n-k)!},\quad \text{where}\quad k = 1,\dots,n;$$

$$C_a^0 = 1,\quad C_a^k = \binom{a}{k} = \frac{a(a-1)\dots(a-k+1)}{k!},\quad \text{where}\quad k = 1,2,\dots$$

Here, a is an arbitrary real number.

Remark. In various texts, the binomial coefficients are also denoted $C_r^n = {}_nC_r = C(n,r)$.

► **Generalization. Some properties.** General case:

$$C_a^b = \frac{\Gamma(a+1)}{\Gamma(b+1)\Gamma(a-b+1)},\quad \text{where}\ \Gamma(x)\ \text{is the gamma function.}$$

Properties:

$$C_a^0 = 1,\quad C_n^k = 0\quad \text{for}\ k = -1,-2,\dots\ \text{or}\ k > n,$$

$$C_a^{b+1} = \frac{a}{b+1}C_{a-1}^b = \frac{a-b}{b+1}C_a^b,\quad C_a^b + C_a^{b+1} = C_{a+1}^{b+1},$$

$$C_{-1/2}^n = \frac{(-1)^n}{2^{2n}}C_{2n}^n = (-1)^n\frac{(2n-1)!!}{(2n)!!},$$

$$C_{1/2}^n = \frac{(-1)^{n-1}}{n2^{2n-1}}C_{2n-2}^{n-1} = \frac{(-1)^{n-1}}{n}\frac{(2n-3)!!}{(2n-2)!!},$$

$$C_{n+1/2}^{2n+1} = (-1)^n 2^{-4n-1}C_{2n}^n,\quad C_{2n+1/2}^n = 2^{-2n}C_{4n+1}^{2n},$$

$$C_n^{1/2} = \frac{2^{2n+1}}{\pi C_{2n}^n},\quad C_n^{n/2} = \frac{2^{2n}}{\pi}C_n^{(n-1)/2}.$$

► **Pochhammer symbol.**

$$(a)_n = a(a+1)\dots(a+n-1) = \frac{\Gamma(a+n)}{\Gamma(a)} = (-1)^n\frac{\Gamma(1-a)}{\Gamma(1-a-n)}.$$

M13.1.2. Bernoulli Numbers

► **Definition.** The *Bernoulli numbers* are defined by the recurrence relation

$$B_0 = 1, \qquad \sum_{k=0}^{n-1} C_n^k B_k = 0, \quad n = 2,\ 3,\ \dots$$

Numerical values:

$$B_0 = 1, \quad B_1 = -\tfrac{1}{2}, \quad B_2 = \tfrac{1}{6}, \quad B_4 = -\tfrac{1}{30}, \quad B_6 = \tfrac{1}{42}, \quad B_8 = -\tfrac{1}{30}, \quad B_{10} = \tfrac{5}{66}, \quad \dots,$$
$$B_{2m+1} = 0 \quad \text{for} \quad m = 1, 2, \dots$$

All odd-numbered Bernoulli numbers but B_1 are zero; all even-numbered Bernoulli numbers have alternating signs.

► **Generating function.**

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} B_n \frac{x^n}{n!}, \qquad |x| < 2\pi.$$

This relation may be regarded as a definition of the Bernoulli numbers.

M13.2. Error Functions. Exponential and Logarithmic Integrals

M13.2.1. Error Function and Complementary Error Function

► **Integral representations.** Definitions:

$$\operatorname{erf} x = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2)\,dt \quad \text{(error function, also called probability integral)},$$
$$\operatorname{erfc} x = 1 - \operatorname{erf} x = \frac{2}{\sqrt{\pi}} \int_x^{\infty} \exp(-t^2)\,dt \quad \text{(complementary error function)}.$$

Properties:

$$\operatorname{erf}(-x) = -\operatorname{erf} x; \quad \operatorname{erf}(0) = 0, \quad \operatorname{erf}(\infty) = 1; \quad \operatorname{erfc}(0) = 1, \quad \operatorname{erfc}(\infty) = 0.$$

► **Expansions as $x \to 0$ and $x \to \infty$. Definite integral.** Expansion of erf x into series in powers of x as $x \to 0$:

$$\operatorname{erf} x = \frac{2}{\sqrt{\pi}} \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{k!\,(2k+1)} = \frac{2}{\sqrt{\pi}} \exp\bigl(-x^2\bigr) \sum_{k=0}^{\infty} \frac{2^k x^{2k+1}}{(2k+1)!!}.$$

Asymptotic expansion of erfc x as $x \to \infty$:

$$\operatorname{erfc} x = \frac{1}{\sqrt{\pi}} \exp\bigl(-x^2\bigr) \left[\sum_{m=0}^{M-1} (-1)^m \frac{\left(\frac{1}{2}\right)_m}{x^{2m+1}} + O\bigl(|x|^{-2M-1}\bigr)\right], \qquad M = 1, 2, \dots$$

Integral:

$$\int_0^x \operatorname{erf} t\,dt = x \operatorname{erf} x - \frac{1}{2} + \frac{1}{2}\exp(-x^2).$$

M13.2.2. Exponential Integral

► **Integral representations.** Definition:

$$\operatorname{Ei}(x) = \int_{-\infty}^{x} \frac{e^t}{t}\,dt = -\int_{-x}^{\infty} \frac{e^{-t}}{t}\,dt \qquad \text{for} \quad x < 0,$$

$$\operatorname{Ei}(x) = \lim_{\varepsilon\to+0}\left(\int_{-\infty}^{-\varepsilon} \frac{e^t}{t}\,dt + \int_{\varepsilon}^{x} \frac{e^t}{t}\,dt\right) \qquad \text{for} \quad x > 0.$$

Other integral representations:

$$\operatorname{Ei}(-x) = -e^{-x}\int_0^\infty \frac{x\sin t + t\cos t}{x^2+t^2}\,dt \quad \text{for} \quad x > 0,$$

$$\operatorname{Ei}(-x) = e^{-x}\int_0^\infty \frac{x\sin t - t\cos t}{x^2+t^2}\,dt \quad \text{for} \quad x < 0.$$

► **Expansions as $x \to 0$ and $x \to \infty$.** Expansion into series in powers of x as $x \to 0$:

$$\operatorname{Ei}(x) = \begin{cases} \mathcal{C} + \ln(-x) + \displaystyle\sum_{k=1}^{\infty} \frac{x^k}{k!\,k} & \text{if } x < 0, \\ \mathcal{C} + \ln x + \displaystyle\sum_{k=1}^{\infty} \frac{x^k}{k!\,k} & \text{if } x > 0, \end{cases}$$

where $\mathcal{C} = 0.5772\ldots$ is the Euler constant.

Asymptotic expansion as $x \to \infty$:

$$\operatorname{Ei}(-x) = e^{-x}\sum_{k=1}^{n}(-1)^k\frac{(k-1)!}{x^k} + R_n, \qquad R_n < \frac{n!}{x^n}.$$

M13.2.3. Logarithmic Integral

► **Integral representations.** Definition:

$$\operatorname{li}(x) = \begin{cases} \displaystyle\int_0^x \frac{dt}{\ln t} & \text{if } 0 < x < 1, \\ \displaystyle\lim_{\varepsilon\to+0}\left(\int_0^{1-\varepsilon} \frac{dt}{\ln t} + \int_{1+\varepsilon}^{x} \frac{dt}{\ln t}\right) & \text{if } x > 1. \end{cases}$$

► **Limiting properties. Relation to the exponential integral.** For small x,

$$\operatorname{li}(x) \approx \frac{x}{\ln(1/x)}.$$

For large x,

$$\operatorname{li}(x) \approx \frac{x}{\ln x}.$$

Asymptotic expansion as $x \to 1$:

$$\operatorname{li}(x) = \mathcal{C} + \ln|\ln x| + \sum_{k=1}^{\infty} \frac{\ln^k x}{k!\,k}, \qquad \mathcal{C} = 0.5772\ldots$$

M13.3. Sine Integral and Cosine Integral. Fresnel Integrals

M13.3.1. Sine Integral

► **Integral representations. Properties.** Definition:

$$\operatorname{Si}(x)=\int_0^x \frac{\sin t}{t}\,dt, \qquad \operatorname{si}(x)=-\int_x^\infty \frac{\sin t}{t}\,dt=\operatorname{Si}(x)-\frac{\pi}{2}.$$

Specific values:

$$\operatorname{Si}(0)=0, \quad \operatorname{Si}(\infty)=\frac{\pi}{2}, \quad \operatorname{si}(\infty)=0.$$

Properties:

$$\operatorname{Si}(-x)=-\operatorname{Si}(x), \quad \operatorname{si}(x)+\operatorname{si}(-x)=-\pi, \quad \lim_{x\to-\infty}\operatorname{si}(x)=-\pi.$$

► **Expansions as $x \to 0$ and $x \to \infty$.** Expansion into series in powers of x as $x \to 0$:

$$\operatorname{Si}(x)=\sum_{k=1}^{\infty}\frac{(-1)^{k+1}x^{2k-1}}{(2k-1)\,(2k-1)!}.$$

Asymptotic expansion as $x \to \infty$:

$$\operatorname{si}(x)=-\cos x\left[\sum_{m=0}^{M-1}\frac{(-1)^m(2m)!}{x^{2m+1}}+O\bigl(|x|^{-2M-1}\bigr)\right]+\sin x\left[\sum_{m=1}^{N-1}\frac{(-1)^m(2m-1)!}{x^{2m}}+O\bigl(|x|^{-2N}\bigr)\right],$$

where $M, N = 1, 2, \dots$

M13.3.2. Cosine Integral

► **Integral representations.** Definition:

$$\operatorname{Ci}(x)=-\int_x^\infty \frac{\cos t}{t}\,dt=\mathcal{C}+\ln x+\int_0^x \frac{\cos t-1}{t}\,dt,$$

where $\mathcal{C}=0.5772\dots$ is the Euler constant.

► **Expansions as $x \to 0$ and $x \to \infty$.** Expansion into series in powers of x as $x \to 0$:

$$\operatorname{Ci}(x)=\mathcal{C}+\ln x+\sum_{k=1}^{\infty}\frac{(-1)^k x^{2k}}{2k\,(2k)!}.$$

Asymptotic expansion as $x \to \infty$:

$$\operatorname{Ci}(x)=\cos x\left[\sum_{m=1}^{M-1}\frac{(-1)^m(2m-1)!}{x^{2m}}+O\bigl(|x|^{-2M}\bigr)\right]+\sin x\left[\sum_{m=0}^{N-1}\frac{(-1)^m(2m)!}{x^{2m+1}}+O\bigl(|x|^{-2N-1}\bigr)\right],$$

where $M, N = 2, 3, \dots$

M13.3.3. Fresnel Integrals

► **Integral representations.** Definitions:

$$S(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \frac{\sin t}{\sqrt{t}}\,dt = \sqrt{\frac{2}{\pi}} \int_0^{\sqrt{x}} \sin t^2\,dt,$$
$$C(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \frac{\cos t}{\sqrt{t}}\,dt = \sqrt{\frac{2}{\pi}} \int_0^{\sqrt{x}} \cos t^2\,dt.$$

► **Expansions as $x \to 0$ and $x \to \infty$.** Expansion into series in powers of x as $x \to 0$:

$$S(x) = \sqrt{\frac{2}{\pi}}\, x \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(4k+3)\,(2k+1)!},$$
$$C(x) = \sqrt{\frac{2}{\pi}}\, x \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(4k+1)\,(2k)!}.$$

Asymptotic expansion as $x \to \infty$:

$$S(x) = \frac{1}{2} - \frac{\cos x}{\sqrt{2\pi x}} P(x) - \frac{\sin x}{\sqrt{2\pi x}} Q(x),$$
$$C(x) = \frac{1}{2} + \frac{\sin x}{\sqrt{2\pi x}} P(x) - \frac{\cos x}{\sqrt{2\pi x}} Q(x),$$
$$P(x) = 1 - \frac{1\times 3}{(2x)^2} + \frac{1\times 3\times 5\times 7}{(2x)^4} - \cdots, \quad Q(x) = \frac{1}{2x} - \frac{1\times 3\times 5}{(2x)^3} + \cdots.$$

M13.4. Gamma Function, Psi Function, and Beta Function

M13.4.1. Gamma Function

► **Integral representations. Simplest properties.** The gamma function, $\Gamma(z)$, is an analytic function of the complex argument z everywhere except for the points $z = 0, -1, -2, \ldots$

For $\operatorname{Re} z > 0$,

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t}\,dt.$$

For $-(n+1) < \operatorname{Re} z < -n$, where $n = 0, 1, 2, \ldots$,

$$\Gamma(z) = \int_0^{\infty} \left[e^{-t} - \sum_{m=0}^{n} \frac{(-1)^m}{m!} \right] t^{z-1}\,dt.$$

Simplest properties:

$$\Gamma(z+1) = z\Gamma(z), \quad \Gamma(n+1) = n!, \quad \Gamma(1) = \Gamma(2) = 1.$$

Fractional values of the argument:

$$\Gamma\Big(\frac{1}{2}\Big) = \sqrt{\pi}, \qquad \Gamma\Big(n+\frac{1}{2}\Big) = \frac{\sqrt{\pi}}{2^n}(2n-1)!!,$$
$$\Gamma\Big(-\frac{1}{2}\Big) = -2\sqrt{\pi}, \quad \Gamma\Big(\frac{1}{2}-n\Big) = (-1)^n\frac{2^n\sqrt{\pi}}{(2n-1)!!}.$$

► **Euler, Stirling, and other formulas.** Euler formula

$$\Gamma(z) = \lim_{n\to\infty}\frac{n!\,n^z}{z(z+1)\dots(z+n)} \qquad (z \neq 0, -1, -2, \dots).$$

Symmetry formulas:

$$\Gamma(z)\Gamma(-z) = -\frac{\pi}{z\sin(\pi z)}, \quad \Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)},$$
$$\Gamma\Big(\frac{1}{2}+z\Big)\Gamma\Big(\frac{1}{2}-z\Big) = \frac{\pi}{\cos(\pi z)}.$$

Multiple argument formulas:

$$\Gamma(2z) = \frac{2^{2z-1}}{\sqrt{\pi}}\Gamma(z)\Gamma\Big(z+\frac{1}{2}\Big),$$
$$\Gamma(3z) = \frac{3^{3z-1/2}}{2\pi}\Gamma(z)\Gamma\Big(z+\frac{1}{3}\Big)\Gamma\Big(z+\frac{2}{3}\Big),$$
$$\Gamma(nz) = (2\pi)^{(1-n)/2}n^{nz-1/2}\prod_{k=0}^{n-1}\Gamma\Big(z+\frac{k}{n}\Big).$$

Asymptotic expansion (*Stirling formula*):

$$\Gamma(z) = \sqrt{2\pi}\,e^{-z}z^{z-1/2}\big[1+\tfrac{1}{12}z^{-1}+\tfrac{1}{288}z^{-2}+O(z^{-3})\big] \qquad (|\arg z| < \pi).$$

M13.4.2. Psi Function (Digamma Function)

Definition:

$$\psi(z) = \frac{d\ln\Gamma(z)}{dz} = \frac{\Gamma'_z(z)}{\Gamma(z)}.$$

The psi function is the logarithmic derivative of the gamma function and is also called the *digamma function*.

Integral representations ($\operatorname{Re} z > 0$):

$$\psi(z) = \int_0^\infty \big[e^{-t}-(1+t)^{-z}\big]t^{-1}\,dt,$$
$$\psi(z) = \ln z + \int_0^\infty \big[t^{-1}-(1-e^{-t})^{-1}\big]e^{-tz}\,dt,$$
$$\psi(z) = -\mathcal{C} + \int_0^1 \frac{1-t^{z-1}}{1-t}\,dt,$$

where $\mathcal{C} = -\psi(1) = 0.5772\dots$ is the Euler constant.

Values for integer argument:

$$\psi(1) = -\mathcal{C}, \qquad \psi(n) = -\mathcal{C} + \sum_{k=1}^{n-1} k^{-1} \quad (n = 2, 3, \dots).$$

Asymptotic expansion as $z \to \infty$ ($|\arg z| < \pi$):

$$\psi(z) = \ln z - \frac{1}{2z} - \frac{1}{12z^2} + \frac{1}{120z^4} - \frac{1}{252z^6} + \cdots = \ln z - \frac{1}{2z} - \sum_{n=1}^{\infty} \frac{B_{2n}}{2nz^{2n}},$$

where the B_{2n} are Bernoulli numbers.

M13.4.3. Beta Function

► **Integral representation. Relationship with the gamma function.** Definition:

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt,$$

where $\operatorname{Re} x > 0$ and $\operatorname{Re} y > 0$.

Relationship with the gamma function:

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

► **Some properties.**

$$\begin{aligned}
&B(x, y) = B(y, x);\\
&B(x, y+1) = \frac{y}{x} B(x+1, y) = \frac{y}{x+y} B(x, y);\\
&B(x, 1-x) = \frac{\pi}{\sin(\pi x)}, \quad 0 < x < 1;\\
&\frac{1}{B(n, m)} = mC_{n+m-1}^{n-1} = nC_{n+m-1}^{m-1} = mC_{n+m-1}^{m} = nC_{n+m-1}^{n},
\end{aligned}$$

where n and m are positive integers.

M13.5. Incomplete Gamma Function

M13.5.1. Integral Representations. Recurrence Formulas

Definitions:

$$\gamma(\alpha, x) = \int_0^x e^{-t} t^{\alpha-1}\,dt, \qquad \operatorname{Re} \alpha > 0,$$

$$\Gamma(\alpha, x) = \int_x^\infty e^{-t} t^{\alpha-1}\,dt = \Gamma(\alpha) - \gamma(\alpha, x).$$

Recurrence formulas:

$$\begin{aligned}
&\gamma(\alpha+1, x) = \alpha\gamma(\alpha, x) - x^\alpha e^{-x},\\
&\gamma(\alpha+1, x) = (x+\alpha)\gamma(\alpha, x) + (1-\alpha)x\gamma(\alpha-1, x),\\
&\Gamma(\alpha+1, x) = \alpha\Gamma(\alpha, x) + x^\alpha e^{-x}.
\end{aligned}$$

Special cases:

$$\gamma(n+1,x) = n!\left[1 - e^{-x}\left(\sum_{k=0}^{n}\frac{x^k}{k!}\right)\right], \qquad n = 0,\ 1,\ \dots;$$

$$\Gamma(n+1,x) = n!\,e^{-x}\sum_{k=0}^{n}\frac{x^k}{k!}, \qquad n = 0,\ 1,\ \dots;$$

$$\Gamma(-n,x) = \frac{(-1)^n}{n!}\left[\Gamma(0,x) - e^{-x}\sum_{k=0}^{n-1}(-1)^k\frac{k!}{x^{k+1}}\right], \quad n = 1,\ 2,\ \dots$$

M13.5.2. Expansions as $x \to 0$ and $x \to \infty$. Relation to Other Functions

Asymptotic expansions as $x \to 0$:

$$\gamma(\alpha,x) = \sum_{n=0}^{\infty}\frac{(-1)^n x^{\alpha+n}}{n!\,(\alpha+n)},$$

$$\Gamma(\alpha,x) = \Gamma(\alpha) - \sum_{n=0}^{\infty}\frac{(-1)^n x^{\alpha+n}}{n!\,(\alpha+n)}.$$

Asymptotic expansions as $x \to \infty$:

$$\gamma(\alpha,x) = \Gamma(\alpha) - x^{\alpha-1}e^{-x}\left[\sum_{m=0}^{M-1}\frac{(1-\alpha)_m}{(-x)^m} + O\big(|x|^{-M}\big)\right], \qquad M = 1,2,\dots;$$

$$\Gamma(\alpha,x) = x^{\alpha-1}e^{-x}\left[\sum_{m=0}^{M-1}\frac{(1-\alpha)_m}{(-x)^m} + O\big(|x|^{-M}\big)\right] \qquad \big(-\tfrac{3}{2}\pi < \arg x < \tfrac{3}{2}\pi\big).$$

Representation of the error function, complementary error function, and exponential integral in terms of the gamma functions:

$$\operatorname{erf} x = \frac{1}{\sqrt{\pi}}\gamma\Big(\frac{1}{2},\ x^2\Big), \quad \operatorname{erfc} x = \frac{1}{\sqrt{\pi}}\Gamma\Big(\frac{1}{2},\ x^2\Big), \quad \operatorname{Ei}(-x) = -\Gamma(0,x).$$

M13.6. Bessel Functions (Cylinder Functions)

M13.6.1. Definitions and Basic Formulas

► **Bessel functions of the first and the second kind.** The *Bessel function of the first kind*, $J_\nu(x)$, and the *Bessel function of the second kind*, $Y_\nu(x)$ (also called the *Neumann function*), are solutions of the Bessel equation

$$x^2 y''_{xx} + x y'_x + (x^2 - \nu^2)y = 0$$

and are defined by the formulas

$$J_\nu(x) = \sum_{k=0}^{\infty}\frac{(-1)^k (x/2)^{\nu+2k}}{k!\,\Gamma(\nu+k+1)}, \quad Y_\nu(x) = \frac{J_\nu(x)\cos\pi\nu - J_{-\nu}(x)}{\sin\pi\nu}. \tag{M13.6.1.1}$$

The formula for $Y_\nu(x)$ is valid for $\nu \neq 0, \pm1, \pm2, \dots$ (the cases $\nu = 0, \pm1, \pm2, \dots$ are discussed in what follows).

The general solution of the Bessel equation has the form $Z_\nu(x) = C_1 J_\nu(x) + C_2 Y_\nu(x)$ and is called a *cylinder function*.

► **Some formulas.**

$$2\nu Z_\nu(x) = x[Z_{\nu-1}(x) + Z_{\nu+1}(x)],$$

$$\frac{d}{dx} Z_\nu(x) = \frac{1}{2}[Z_{\nu-1}(x) - Z_{\nu+1}(x)] = \pm\left[\frac{\nu}{x} Z_\nu(x) - Z_{\nu\pm1}(x)\right],$$

$$\frac{d}{dx}[x^\nu Z_\nu(x)] = x^\nu Z_{\nu-1}(x), \qquad \frac{d}{dx}[x^{-\nu} Z_\nu(x)] = -x^{-\nu} Z_{\nu+1}(x),$$

$$\left(\frac{1}{x}\frac{d}{dx}\right)^n [x^\nu J_\nu(x)] = x^{\nu-n} J_{\nu-n}(x), \qquad \left(\frac{1}{x}\frac{d}{dx}\right)^n [x^{-\nu} J_\nu(x)] = (-1)^n x^{-\nu-n} J_{\nu+n}(x).$$

► **Bessel functions for $\nu = \pm n \pm \frac{1}{2}$, where $n = 0, 1, 2, \ldots$**

$$J_{1/2}(x) = \sqrt{\frac{2}{\pi x}} \sin x, \qquad J_{-1/2}(x) = \sqrt{\frac{2}{\pi x}} \cos x,$$

$$J_{3/2}(x) = \sqrt{\frac{2}{\pi x}} \left(\frac{1}{x} \sin x - \cos x\right), \qquad J_{-3/2}(x) = \sqrt{\frac{2}{\pi x}} \left(-\frac{1}{x} \cos x - \sin x\right),$$

$$J_{n+1/2}(x) = \sqrt{\frac{2}{\pi x}} \left[\sin\left(x - \frac{n\pi}{2}\right) \sum_{k=0}^{[n/2]} \frac{(-1)^k (n+2k)!}{(2k)!\,(n-2k)!\,(2x)^{2k}} \right.$$
$$\left. + \cos\left(x - \frac{n\pi}{2}\right) \sum_{k=0}^{[(n-1)/2]} \frac{(-1)^k (n+2k+1)!}{(2k+1)!\,(n-2k-1)!\,(2x)^{2k+1}}\right],$$

$$J_{-n-1/2}(x) = \sqrt{\frac{2}{\pi x}} \left[\cos\left(x + \frac{n\pi}{2}\right) \sum_{k=0}^{[n/2]} \frac{(-1)^k (n+2k)!}{(2k)!\,(n-2k)!\,(2x)^{2k}} \right.$$
$$\left. - \sin\left(x + \frac{n\pi}{2}\right) \sum_{k=0}^{[(n-1)/2]} \frac{(-1)^k (n+2k+1)!}{(2k+1)!\,(n-2k-1)!\,(2x)^{2k+1}}\right],$$

$$Y_{1/2}(x) = -\sqrt{\frac{2}{\pi x}} \cos x, \qquad Y_{-1/2}(x) = \sqrt{\frac{2}{\pi x}} \sin x,$$

$$Y_{n+1/2}(x) = (-1)^{n+1} J_{-n-1/2}(x), \qquad Y_{-n-1/2}(x) = (-1)^n J_{n+1/2}(x),$$

where $[A]$ stands for the integer part of the number A.

► **Bessel functions for $\nu = \pm n$, where $n = 0, 1, 2, \ldots$** Let $\nu = n$ be an arbitrary integer. The relations

$$J_{-n}(x) = (-1)^n J_n(x), \qquad Y_{-n}(x) = (-1)^n Y_n(x)$$

are valid. The function $J_n(x)$ is given by the first formula in (M13.6.1.1) with $\nu = n$, and $Y_n(x)$ can be obtained from the second formula in (M13.6.1.1) by passing to the limit $\nu \to n$. For nonnegative n, $Y_n(x)$ can be represented in the form

$$Y_n(x) = \frac{2}{\pi} J_n(x) \ln\frac{x}{2} - \frac{1}{\pi} \sum_{k=0}^{n-1} \frac{(n-k-1)!}{k!} \left(\frac{2}{x}\right)^{n-2k} - \frac{1}{\pi} \sum_{k=0}^{\infty} (-1)^k \left(\frac{x}{2}\right)^{n+2k} \frac{\psi(k+1) + \psi(n+k+1)}{k!\,(n+k)!},$$

where $\psi(1) = -\mathcal{C}$, $\psi(n) = -\mathcal{C} + \sum\limits_{k=1}^{n-1} k^{-1}$, $\mathcal{C} = 0.5772\ldots$ is the Euler constant, and $\psi(x) = [\ln \Gamma(x)]'_x$ is the logarithmic derivative of the gamma function, also known as the digamma function (see Subsection M13.4.2).

► **Wronskians.**

$$W(J_\nu, J_{-\nu}) = -\frac{2}{\pi x}\sin(\pi\nu), \qquad W(J_\nu, Y_\nu) = \frac{2}{\pi x},$$

Here, the notation $W(f,g) = fg'_x - f'_x g$ is used.

M13.6.2. Integral Representations and Asymptotic Expansions

► **Integral representations.** The functions $J_\nu(x)$ and $Y_\nu(x)$ can be represented in the form of definite integrals (for $x > 0$):

$$\pi J_\nu(x) = \int_0^\pi \cos(x\sin\theta - \nu\theta)\,d\theta - \sin\pi\nu\int_0^\infty \exp(-x\sinh t - \nu t)\,dt,$$
$$\pi Y_\nu(x) = \int_0^\pi \sin(x\sin\theta - \nu\theta)\,d\theta - \int_0^\infty (e^{\nu t} + e^{-\nu t}\cos\pi\nu)\,e^{-x\sinh t}\,dt.$$

For $|\nu| < \frac{1}{2}$, $x > 0$,

$$J_\nu(x) = \frac{2^{1+\nu}x^{-\nu}}{\pi^{1/2}\Gamma(\frac{1}{2}-\nu)}\int_1^\infty \frac{\sin(xt)\,dt}{(t^2-1)^{\nu+1/2}},$$
$$Y_\nu(x) = -\frac{2^{1+\nu}x^{-\nu}}{\pi^{1/2}\Gamma(\frac{1}{2}-\nu)}\int_1^\infty \frac{\cos(xt)\,dt}{(t^2-1)^{\nu+1/2}}.$$

For $\nu > -\frac{1}{2}$,

$$J_\nu(x) = \frac{2(x/2)^\nu}{\pi^{1/2}\Gamma(\frac{1}{2}+\nu)}\int_0^{\pi/2}\cos(x\cos t)\sin^{2\nu}t\,dt \quad (\textit{Poisson's formula}).$$

For $\nu = 0$, $x > 0$,

$$J_0(x) = \frac{2}{\pi}\int_0^\infty \sin(x\cosh t)\,dt, \qquad Y_0(x) = -\frac{2}{\pi}\int_0^\infty \cos(x\cosh t)\,dt.$$

For integer $\nu = n = 0, 1, 2, \ldots,$

$$J_n(x) = \frac{1}{\pi}\int_0^\pi \cos(nt - x\sin t)\,dt \quad (\textit{Bessel's formula}),$$
$$J_{2n}(x) = \frac{2}{\pi}\int_0^{\pi/2}\cos(x\sin t)\cos(2nt)\,dt,$$
$$J_{2n+1}(x) = \frac{2}{\pi}\int_0^{\pi/2}\sin(x\sin t)\sin[(2n+1)t]\,dt.$$

► **Asymptotic expansions as $|x| \to \infty$.**

$$J_\nu(x) = \sqrt{\frac{2}{\pi x}}\left\{\cos\left(\frac{4x - 2\nu\pi - \pi}{4}\right)\left[\sum_{m=0}^{M-1}(-1)^m(\nu, 2m)(2x)^{-2m} + O(|x|^{-2M})\right]\right.$$

$$-\sin\Bigl(\frac{4x-2\nu\pi-\pi}{4}\Bigr)\biggl[\sum_{m=0}^{M-1}(-1)^m(\nu,2m+1)(2x)^{-2m-1}+O(|x|^{-2M-1})\biggr]\biggr\},$$

$$Y_\nu(x)=\sqrt{\frac{2}{\pi x}}\biggl\{\sin\Bigl(\frac{4x-2\nu\pi-\pi}{4}\Bigr)\biggl[\sum_{m=0}^{M-1}(-1)^m(\nu,2m)(2x)^{-2m}+O(|x|^{-2M})\biggr]$$

$$+\cos\Bigl(\frac{4x-2\nu\pi-\pi}{4}\Bigr)\biggl[\sum_{m=0}^{M-1}(-1)^m(\nu,2m+1)(2x)^{-2m-1}+O(|x|^{-2M-1})\biggr]\biggr\},$$

where $(\nu,m)=\dfrac{1}{2^{2m}m!}(4\nu^2-1)(4\nu^2-3^2)\dots[4\nu^2-(2m-1)^2]=\dfrac{\Gamma(\frac12+\nu+m)}{m!\,\Gamma(\frac12+\nu-m)}$.

For nonnegative integer n and large x,

$$\sqrt{\pi x}\,J_{2n}(x)=(-1)^n(\cos x+\sin x)+O(x^{-2}),$$
$$\sqrt{\pi x}\,J_{2n+1}(x)=(-1)^{n+1}(\cos x-\sin x)+O(x^{-2}).$$

► **Zeros of Bessel functions.** Each of the functions $J_\nu(x)$ and $Y_\nu(x)$ has infinitely many real zeros (for real ν). All zeros are simple, except possibly for the point $x=0$.

The zeros γ_m of $J_0(x)$, i.e., the roots of the equation $J_0(\gamma_m)=0$, are approximately given by

$$\gamma_m=2.4+3.13\,(m-1)\qquad(m=1,2,\dots),$$

with a maximum error of 0.2%.

M13.7. Modified Bessel Functions

M13.7.1. Definitions. Basic Formulas

► **Modified Bessel functions of the first and the second kind.** The *modified Bessel function of the first kind*, $I_\nu(x)$, and the *modified Bessel function of the second kind*, $K_\nu(x)$ (also called the *Macdonald function*), of order ν are solutions of the modified Bessel equation

$$x^2y''_{xx}+xy'_x-(x^2+\nu^2)y=0$$

and are defined by the formulas

$$I_\nu(x)=\sum_{k=0}^{\infty}\frac{(x/2)^{2k+\nu}}{k!\,\Gamma(\nu+k+1)},\qquad K_\nu(x)=\frac{\pi}{2}\frac{I_{-\nu}(x)-I_\nu(x)}{\sin(\pi\nu)},$$

(see below for $K_\nu(x)$ with $\nu=0,1,2,\dots$).

► **Some formulas.** The modified Bessel functions possess the following properties:

$$K_{-\nu}(x)=K_\nu(x);\qquad I_{-n}(x)=(-1)^nI_n(x),\quad n=0,1,2,\dots$$
$$2\nu I_\nu(x)=x[I_{\nu-1}(x)-I_{\nu+1}(x)],\qquad 2\nu K_\nu(x)=-x[K_{\nu-1}(x)-K_{\nu+1}(x)],$$
$$\frac{d}{dx}I_\nu(x)=\frac12[I_{\nu-1}(x)+I_{\nu+1}(x)],\qquad \frac{d}{dx}K_\nu(x)=-\frac12[K_{\nu-1}(x)+K_{\nu+1}(x)].$$

► **Modified Bessel functions for $\nu = \pm n \pm \frac{1}{2}$, where $n = 0, 1, 2, \dots$**

$$I_{1/2}(x) = \sqrt{\frac{2}{\pi x}}\sinh x, \qquad I_{-1/2}(x) = \sqrt{\frac{2}{\pi x}}\cosh x,$$

$$I_{3/2}(x) = \sqrt{\frac{2}{\pi x}}\left(-\frac{1}{x}\sinh x + \cosh x\right), \quad I_{-3/2}(x) = \sqrt{\frac{2}{\pi x}}\left(-\frac{1}{x}\cosh x + \sinh x\right),$$

$$I_{n+1/2}(x) = \frac{1}{\sqrt{2\pi x}}\left[e^x \sum_{k=0}^{n} \frac{(-1)^k (n+k)!}{k!\,(n-k)!\,(2x)^k} - (-1)^n e^{-x} \sum_{k=0}^{n} \frac{(n+k)!}{k!\,(n-k)!\,(2x)^k}\right],$$

$$I_{-n-1/2}(x) = \frac{1}{\sqrt{2\pi x}}\left[e^x \sum_{k=0}^{n} \frac{(-1)^k (n+k)!}{k!\,(n-k)!\,(2x)^k} + (-1)^n e^{-x} \sum_{k=0}^{n} \frac{(n+k)!}{k!\,(n-k)!\,(2x)^k}\right],$$

$$K_{\pm 1/2}(x) = \sqrt{\frac{\pi}{2x}}\,e^{-x}, \qquad K_{\pm 3/2}(x) = \sqrt{\frac{\pi}{2x}}\left(1 + \frac{1}{x}\right)e^{-x},$$

$$K_{n+1/2}(x) = K_{-n-1/2}(x) = \sqrt{\frac{\pi}{2x}}\,e^{-x} \sum_{k=0}^{n} \frac{(n+k)!}{k!\,(n-k)!\,(2x)^k}.$$

► **Modified Bessel functions for $\nu = n$, where $n = 0, 1, 2, \dots$** If $\nu = n$ is a nonnegative integer, then

$$\begin{aligned} K_n(x) = (-1)^{n+1} I_n(x) \ln\frac{x}{2} &+ \frac{1}{2}\sum_{m=0}^{n-1} (-1)^m \left(\frac{x}{2}\right)^{2m-n} \frac{(n-m-1)!}{m!} \\ &+ \frac{1}{2}(-1)^n \sum_{m=0}^{\infty} \left(\frac{x}{2}\right)^{n+2m} \frac{\psi(n+m+1) + \psi(m+1)}{m!\,(n+m)!}; \quad n = 0, 1, 2, \dots, \end{aligned}$$

where $\psi(z)$ is the logarithmic derivative of the gamma function; for $n = 0$, the first sum is omitted.

► **Wronskians.**

$$W(I_\nu, I_{-\nu}) = -\frac{2}{\pi x}\sin(\pi\nu), \qquad W(I_\nu, K_\nu) = -\frac{1}{x},$$

where $W(f, g) = f g'_x - f'_x g$.

M13.7.2. Integral Representations and Asymptotic Expansions

► **Integral representations.** The functions $I_\nu(x)$ and $K_\nu(x)$ can be represented in terms of definite integrals:

$$I_\nu(x) = \frac{x^\nu}{\pi^{1/2} 2^\nu \Gamma(\nu + \frac{1}{2})} \int_{-1}^{1} \exp(-xt)(1 - t^2)^{\nu - 1/2}\,dt \qquad (x > 0,\ \nu > -\tfrac{1}{2}),$$

$$K_\nu(x) = \int_0^\infty \exp(-x\cosh t)\cosh(\nu t)\,dt \qquad (x > 0),$$

$$K_\nu(x) = \frac{1}{\cos\left(\frac{1}{2}\pi\nu\right)} \int_0^\infty \cos(x\sinh t)\cosh(\nu t)\,dt \qquad (x > 0,\ -1 < \nu < 1),$$

$$K_\nu(x) = \frac{1}{\sin\left(\frac{1}{2}\pi\nu\right)} \int_0^\infty \sin(x\sinh t)\sinh(\nu t)\,dt \qquad (x > 0,\ -1 < \nu < 1).$$

For integer $\nu = n$,

$$I_n(x) = \frac{1}{\pi}\int_0^\pi \exp(x\cos t)\cos(nt)\,dt \qquad (n = 0, 1, 2, \dots),$$
$$K_0(x) = \int_0^\infty \cos(x\sinh t)\,dt = \int_0^\infty \frac{\cos(xt)}{\sqrt{t^2+1}}\,dt \qquad (x > 0).$$

► **Asymptotic expansions as $x \to \infty$:**

$$I_\nu(x) = \frac{e^x}{\sqrt{2\pi x}}\left\{1 + \sum_{m=1}^{M}(-1)^m \frac{(4\nu^2-1)(4\nu^2-3^2)\dots[4\nu^2-(2m-1)^2]}{m!\,(8x)^m}\right\},$$
$$K_\nu(x) = \sqrt{\frac{\pi}{2x}}\,e^{-x}\left\{1 + \sum_{m=1}^{M} \frac{(4\nu^2-1)(4\nu^2-3^2)\dots[4\nu^2-(2m-1)^2]}{m!\,(8x)^m}\right\}.$$

The terms of the order of $O(x^{-M-1})$ are omitted in the braces.

M13.8. Degenerate Hypergeometric Functions (Kummer Functions)

M13.8.1. Definitions and Basic Formulas

► **Degenerate hypergeometric functions $\Phi(a, b; x)$ and $\Psi(a, b; x)$.** The *degenerate hypergeometric functions* (*Kummer functions*) $\Phi(a, b; x)$ and $\Psi(a, b; x)$ are solutions of the degenerate hypergeometric equation

$$xy''_{xx} + (b - x)y'_x - ay = 0.$$

In the case $b \neq 0, -1, -2, -3, \dots$, the function $\Phi(a, b; x)$ can be represented as Kummer's series:

$$\Phi(a, b; x) = 1 + \sum_{k=1}^{\infty} \frac{(a)_k}{(b)_k}\frac{x^k}{k!},$$

where $(a)_k = a(a+1)\dots(a+k-1)$, $(a)_0 = 1$.

Table M13.1 presents some special cases where Φ can be expressed in terms of simpler functions.

The function $\Psi(a, b; x)$ is defined as follows:

$$\Psi(a, b; x) = \frac{\Gamma(1-b)}{\Gamma(a-b+1)}\Phi(a, b; x) + \frac{\Gamma(b-1)}{\Gamma(a)}x^{1-b}\Phi(a-b+1,\ 2-b;\ x).$$

Table M13.2 presents some special cases where Ψ can be expressed in terms of simpler functions.

► **Kummer transformation and linear relations.** Kummer transformation:

$$\Phi(a, b; x) = e^x\Phi(b-a, b; -x), \qquad \Psi(a, b; x) = x^{1-b}\Psi(1+a-b, 2-b; x).$$

TABLE M13.1
Special cases of the Kummer function $\Phi(a,b;z)$.

a	b	z	Φ	Notation used
a	a	x	e^x	
1	2	$2x$	$\dfrac{1}{x}e^x\sinh x$	
a	$a+1$	$-x$	$ax^{-a}\gamma(a,x)$	Incomplete gamma function $\gamma(a,x)=\displaystyle\int_0^x e^{-t}t^{a-1}\,dt$
$\dfrac{1}{2}$	$\dfrac{3}{2}$	$-x^2$	$\dfrac{\sqrt{\pi}}{2}\operatorname{erf} x$	Error function $\operatorname{erf} x=\dfrac{2}{\sqrt{\pi}}\displaystyle\int_0^x \exp(-t^2)\,dt$
$-n$	$\dfrac{1}{2}$	$\dfrac{x^2}{2}$	$\dfrac{n!}{(2n)!}\Bigl(-\dfrac{1}{2}\Bigr)^{-n}H_{2n}(x)$	Hermite polynomials $H_n(x)=(-1)^n e^{x^2}\dfrac{d^n}{dx^n}\bigl(e^{-x^2}\bigr)$, $n=0,1,2,\dots$
$-n$	$\dfrac{3}{2}$	$\dfrac{x^2}{2}$	$\dfrac{n!}{(2n+1)!}\Bigl(-\dfrac{1}{2}\Bigr)^{-n}H_{2n+1}(x)$	
$-n$	b	x	$\dfrac{n!}{(b)_n}L_n^{(b-1)}(x)$	Laguerre polynomials $L_n^{(\alpha)}(x)=\dfrac{e^x x^{-\alpha}}{n!}\dfrac{d^n}{dx^n}\bigl(e^{-x}x^{n+\alpha}\bigr)$, $\alpha=b-1$, $(b)_n=b(b+1)\dots(b+n-1)$
$\nu+\dfrac{1}{2}$	$2\nu+1$	$2x$	$\Gamma(1+\nu)e^x\Bigl(\dfrac{x}{2}\Bigr)^{-\nu}I_\nu(x)$	Modified Bessel functions $I_\nu(x)$
$n+1$	$2n+2$	$2x$	$\Gamma\Bigl(n+\dfrac{3}{2}\Bigr)e^x\Bigl(\dfrac{x}{2}\Bigr)^{-n-\frac{1}{2}}I_{n+\frac{1}{2}}(x)$	

Linear relations for Φ:

$$(b-a)\Phi(a-1,b;x)+(2a-b+x)\Phi(a,b;x)-a\Phi(a+1,b;x)=0,$$
$$b(b-1)\Phi(a,b-1;x)-b(b-1+x)\Phi(a,b;x)+(b-a)x\Phi(a,b+1;x)=0,$$
$$(a-b+1)\Phi(a,b;x)-a\Phi(a+1,b;x)+(b-1)\Phi(a,b-1;x)=0,$$
$$b\Phi(a,b;x)-b\Phi(a-1,b;x)-x\Phi(a,b+1;x)=0,$$
$$b(a+x)\Phi(a,b;x)-(b-a)x\Phi(a,b+1;x)-ab\Phi(a+1,b;x)=0,$$
$$(a-1+x)\Phi(a,b;x)+(b-a)\Phi(a-1,b;x)-(b-1)\Phi(a,b-1;x)=0.$$

Linear relations for Ψ:

$$\Psi(a-1,b;x)-(2a-b+x)\Psi(a,b;x)+a(a-b+1)\Psi(a+1,b;x)=0,$$
$$(b-a-1)\Psi(a,b-1;x)-(b-1+x)\Psi(a,b;x)+x\Psi(a,b+1;x)=0,$$
$$\Psi(a,b;x)-a\Psi(a+1,b;x)-\Psi(a,b-1;x)=0,$$
$$(b-a)\Psi(a,b;x)-x\Psi(a,b+1;x)+\Psi(a-1,b;x)=0,$$
$$(a+x)\Psi(a,b;x)+a(b-a-1)\Psi(a+1,b;x)-x\Psi(a,b+1;x)=0,$$
$$(a-1+x)\Psi(a,b;x)-\Psi(a-1,b;x)+(a-c+1)\Psi(a,b-1;x)=0.$$

TABLE M13.2
Special cases of the Kummer function $\Psi(a,b;z)$.

a	b	z	Ψ	Notation used
$1-a$	$1-a$	x	$e^x\Gamma(a,x)$	Incomplete gamma function $\Gamma(a,x)=\int_x^\infty e^{-t}t^{a-1}\,dt$
$\frac{1}{2}$	$\frac{1}{2}$	x^2	$\sqrt{\pi}\,\exp(x^2)\operatorname{erfc}x$	Complementary error function $\operatorname{erfc}x=\frac{2}{\sqrt{\pi}}\int_x^\infty \exp(-t^2)\,dt$
1	1	$-x$	$-e^{-x}\operatorname{Ei}(x)$	Exponential integral $\operatorname{Ei}(x)=\int_{-\infty}^x \frac{e^t}{t}\,dt$
1	1	$-\ln x$	$-x^{-1}\operatorname{li}x$	Logarithmic integral $\operatorname{li}x=\int_0^x \frac{dt}{t}$
$\frac{1}{2}-\frac{n}{2}$	$\frac{3}{2}$	x^2	$2^{-n}x^{-1}H_n(x)$	Hermite polynomials $H_n(x)=(-1)^n e^{x^2}\frac{d^n}{dx^n}\bigl(e^{-x^2}\bigr)$, $n=0,1,2,\dots$
$\nu+\frac{1}{2}$	$2\nu+1$	$2x$	$\pi^{-1/2}(2x)^{-\nu}e^x K_\nu(x)$	Modified Bessel functions $K_\nu(x)$

► **Differentiation formulas and Wronskian.** Differentiation formulas:

$$\frac{d}{dx}\Phi(a,b;x)=\frac{a}{b}\Phi(a+1,b+1;x),\qquad \frac{d^n}{dx^n}\Phi(a,b;x)=\frac{(a)_n}{(b)_n}\Phi(a+n,b+n;x),$$

$$\frac{d}{dx}\Psi(a,b;x)=-a\Psi(a+1,b+1;x),\qquad \frac{d^n}{dx^n}\Psi(a,b;x)=(-1)^n(a)_n\Psi(a+n,b+n;x).$$

Wronskian:

$$W(\Phi,\Psi)=\Phi\Psi'_x-\Phi'_x\Psi=-\frac{\Gamma(b)}{\Gamma(a)}x^{-b}e^x.$$

► **Degenerate hypergeometric functions for $n=0,1,2,\dots$**

$$\Psi(a,n+1;x)=\frac{(-1)^{n-1}}{n!\,\Gamma(a-n)}\biggl\{\Phi(a,n+1;x)\ln x$$
$$+\sum_{r=0}^\infty\frac{(a)_r}{(n+1)_r}\bigl[\psi(a+r)-\psi(1+r)-\psi(1+n+r)\bigr]\frac{x^r}{r!}\biggr\}+\frac{(n-1)!}{\Gamma(a)}\sum_{r=0}^{n-1}\frac{(a-n)_r}{(1-n)_r}\frac{x^{r-n}}{r!},$$

where $n=0,1,2,\dots$ (the last sum is omitted for $n=0$), $\psi(z)=[\ln\Gamma(z)]'_z$ is the logarithmic derivative of the gamma function,

$$\psi(1)=-\mathcal{C},\qquad \psi(n)=-\mathcal{C}+\sum_{k=1}^{n-1}k^{-1},$$

where $\mathcal{C}=0.5772\dots$ is the Euler constant.

If $b < 0$, then the formula

$$\Psi(a, b; x) = x^{1-b}\Psi(a - b + 1,\ 2 - b;\ x)$$

is valid for any x.

For $b \neq 0, -1, -2, -3, \ldots$, the general solution of the degenerate hypergeometric equation can be represented in the form

$$y = C_1\Phi(a, b; x) + C_2\Psi(a, b; x),$$

and for $b = 0, -1, -2, -3, \ldots$, in the form

$$y = x^{1-b}\big[C_1\Phi(a - b + 1,\ 2 - b;\ x) + C_2\Psi(a - b + 1,\ 2 - b;\ x)\big].$$

M13.8.2. Integral Representations and Asymptotic Expansions

► **Integral representations.**

$$\begin{aligned}
\Phi(a, b; x) &= \frac{\Gamma(b)}{\Gamma(a)\,\Gamma(b - a)} \int_0^1 e^{xt} t^{a-1}(1 - t)^{b-a-1}\,dt \quad (\text{for } b > a > 0),\\
\Psi(a, b; x) &= \frac{1}{\Gamma(a)} \int_0^\infty e^{-xt} t^{a-1}(1 + t)^{b-a-1}\,dt \quad (\text{for } a > 0,\ x > 0),
\end{aligned}$$

where $\Gamma(a)$ is the gamma function.

► **Asymptotic expansion as $|x| \to \infty$.**

$$\begin{aligned}
\Phi(a, b; x) &= \frac{\Gamma(b)}{\Gamma(a)} e^x x^{a-b} \left[\sum_{n=0}^{N} \frac{(b - a)_n(1 - a)_n}{n!} x^{-n} + \varepsilon\right], \quad x > 0,\\
\Phi(a, b; x) &= \frac{\Gamma(b)}{\Gamma(b - a)} (-x)^{-a} \left[\sum_{n=0}^{N} \frac{(a)_n(a - b + 1)_n}{n!} (-x)^{-n} + \varepsilon\right], \quad x < 0,\\
\Psi(a, b; x) &= x^{-a} \left[\sum_{n=0}^{N} (-1)^n \frac{(a)_n(a - b + 1)_n}{n!} x^{-n} + \varepsilon\right], \quad -\infty < x < \infty,
\end{aligned}$$

where $\varepsilon = O(x^{-N-1})$.

M13.9. Elliptic Integrals

M13.9.1. Complete Elliptic Integrals

► **Definitions. Properties. Conversion formulas.** *Complete elliptic integral of the first kind*:

$$\mathsf{K}(k) = \int_0^{\pi/2} \frac{d\alpha}{\sqrt{1 - k^2 \sin^2\alpha}} = \int_0^1 \frac{dx}{\sqrt{(1 - x^2)(1 - k^2x^2)}}.$$

Complete elliptic integral of the second kind:

$$\mathsf{E}(k) = \int_0^{\pi/2} \sqrt{1 - k^2 \sin^2\alpha}\,d\alpha = \int_0^1 \frac{\sqrt{1 - k^2x^2}}{\sqrt{1 - x^2}}\,dx.$$

The argument k is called the *elliptic modulus* ($k^2 < 1$).

Notation:

$$k' = \sqrt{1-k^2}, \qquad \mathsf{K}'(k) = \mathsf{K}(k'), \qquad \mathsf{E}'(k) = \mathsf{E}(k'),$$

where k' is the *complementary modulus*.

Properties:

$$\begin{aligned}
&\mathsf{K}(-k) = \mathsf{K}(k), \qquad \mathsf{E}(-k) = \mathsf{E}(k);\\
&\mathsf{K}(k) = \mathsf{K}'(k'), \qquad \mathsf{E}(k) = \mathsf{E}'(k');\\
&\mathsf{E}(k)\,\mathsf{K}'(k) + \mathsf{E}'(k)\,\mathsf{K}(k) - \mathsf{K}(k)\,\mathsf{K}'(k) = \frac{\pi}{2}.
\end{aligned}$$

Conversion formulas for complete elliptic integrals:

$$\begin{aligned}
\mathsf{K}\left(\frac{1-k'}{1+k'}\right) &= \frac{1+k'}{2}\,\mathsf{K}(k),\\
\mathsf{E}\left(\frac{1-k'}{1+k'}\right) &= \frac{1}{1+k'}\big[\mathsf{E}(k) + k'\,\mathsf{K}(k)\big],\\
\mathsf{K}\left(\frac{2\sqrt{k}}{1+k}\right) &= (1+k)\,\mathsf{K}(k),\\
\mathsf{E}\left(\frac{2\sqrt{k}}{1+k}\right) &= \frac{1}{1+k}\big[2\,\mathsf{E}(k) - (k')^2\,\mathsf{K}(k)\big].
\end{aligned}$$

► **Representation of complete elliptic integrals in series form.** Representation of complete elliptic integrals in the form of series in powers of the modulus k:

$$\begin{aligned}
\mathsf{K}(k) &= \frac{\pi}{2}\left\{1 + \left(\frac{1}{2}\right)^2 k^2 + \left(\frac{1\times 3}{2\times 4}\right)^2 k^4 + \cdots + \left[\frac{(2n-1)!!}{(2n)!!}\right]^2 k^{2n} + \cdots\right\},\\
\mathsf{E}(k) &= \frac{\pi}{2}\left\{1 - \left(\frac{1}{2}\right)^2 \frac{k^2}{1} - \left(\frac{1\times 3}{2\times 4}\right)^2 \frac{k^4}{3} - \cdots - \left[\frac{(2n-1)!!}{(2n)!!}\right]^2 \frac{k^{2n}}{2n-1} - \cdots\right\}.
\end{aligned}$$

► **Differentiation formulas. Differential equations.** Differentiation formulas:

$$\frac{d\,\mathsf{K}(k)}{dk} = \frac{\mathsf{E}(k)}{k(k')^2} - \frac{\mathsf{K}(k)}{k}, \qquad \frac{d\,\mathsf{E}(k)}{dk} = \frac{\mathsf{E}(k) - \mathsf{K}(k)}{k}.$$

The functions $\mathsf{K}(k)$ and $\mathsf{K}'(k)$ satisfy the second-order linear ordinary differential equation

$$\frac{d}{dk}\left[k(1-k^2)\frac{d\,\mathsf{K}}{dk}\right] - k\,\mathsf{K} = 0.$$

The functions $\mathsf{E}(k)$ and $\mathsf{E}'(k) - \mathsf{K}'(k)$ satisfy the second-order linear ordinary differential equation

$$(1-k^2)\frac{d}{dk}\left(k\frac{d\,\mathsf{E}}{dk}\right) + k\,\mathsf{E} = 0.$$

M13.9.2. Incomplete Elliptic Integrals

► **Definitions. Properties.** *Elliptic integral of the first kind*:

$$F(\varphi,k)=\int_0^{\varphi}\frac{d\alpha}{\sqrt{1-k^2\sin^2\alpha}}=\int_0^{\sin\varphi}\frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}.$$

Elliptic integral of the second kind:

$$E(\varphi,k)=\int_0^{\varphi}\sqrt{1-k^2\sin^2\alpha}\,d\alpha=\int_0^{\sin\varphi}\frac{\sqrt{1-k^2x^2}}{\sqrt{1-x^2}}\,dx.$$

Elliptic integral of the third kind:

$$\Pi(\varphi,n,k)=\int_0^{\varphi}\frac{d\alpha}{(1-n\sin^2\alpha)\sqrt{1-k^2\sin^2\alpha}}=\int_0^{\sin\varphi}\frac{dx}{(1-nx^2)\sqrt{(1-x^2)(1-k^2x^2)}}.$$

The quantity k is called the *elliptic modulus* ($k^2<1$), $k'=\sqrt{1-k^2}$ is the *complementary modulus*, and n is the *characteristic parameter*.

Complete elliptic integrals:

$$\mathsf{K}(k)=F\Big(\frac{\pi}{2},k\Big),\qquad \mathsf{E}(k)=E\Big(\frac{\pi}{2},k\Big),$$
$$\mathsf{K}'(k)=F\Big(\frac{\pi}{2},k'\Big),\qquad \mathsf{E}'(k)=E\Big(\frac{\pi}{2},k'\Big).$$

Properties of elliptic integrals:

$$F(-\varphi,k)=-F(\varphi,k),\qquad F(n\pi\pm\varphi,k)=2n\,\mathsf{K}(k)\pm F(\varphi,k);$$
$$E(-\varphi,k)=-E(\varphi,k),\qquad E(n\pi\pm\varphi,k)=2n\,\mathsf{E}(k)\pm E(\varphi,k).$$

► **Conversion formulas.** Conversion formulas for elliptic integrals (first set):

$$F\left(\psi,\frac{1}{k}\right)=kF(\varphi,k),$$
$$E\left(\psi,\frac{1}{k}\right)=\frac{1}{k}\big[E(\varphi,k)-(k')^2F(\varphi,k)\big],$$

where the arguments φ and ψ, which may be treated as angles, are related by $\sin\psi=k\sin\varphi$, $\cos\psi=\sqrt{1-k^2\sin^2\varphi}$.

Conversion formulas for elliptic integrals (second set):

$$F\left(\psi,\frac{1-k'}{1+k'}\right)=(1+k')F(\varphi,k),$$
$$E\left(\psi,\frac{1-k'}{1+k'}\right)=\frac{2}{1+k'}\big[E(\varphi,k)+k'F(\varphi,k)\big]-\frac{1-k'}{1+k'}\sin\psi,$$

where φ and ψ are related by $\tan(\psi-\varphi)=k'\tan\varphi$.

Transformation formulas for elliptic integrals (third set):

$$F\left(\psi,\frac{2\sqrt{k}}{1+k}\right)=(1+k)F(\varphi,k),$$
$$E\left(\psi,\frac{2\sqrt{k}}{1+k}\right)=\frac{1}{1+k}\left[2E(\varphi,k)-(k')^2F(\varphi,k)+2k\frac{\sin\varphi\,\cos\varphi}{1+k\sin^2\varphi}\sqrt{1-k^2\sin^2\varphi}\right],$$

where φ and ψ are related by $\sin\psi=\dfrac{(1+k)\sin\varphi}{1+k\sin^2\varphi}$.

M13.10. Orthogonal Polynomials

All zeros of each of the orthogonal polynomials $\mathcal{P}_n(x)$ considered in this section are real and simple. The zeros of the polynomials $\mathcal{P}_n(x)$ and $\mathcal{P}_{n+1}(x)$ are alternating.

M13.10.1. Legendre Polynomials and Legendre Functions

► **Explicit and recurrence formulas for Legendre polynomials and functions.** The *Legendre polynomials* $P_n(x)$ and the *Legendre functions* $Q_n(x)$ are solutions of the second-order linear ordinary differential equation

$$(1-x^2)y''_{xx} - 2xy'_x + n(n+1)y = 0.$$

The *Legendre polynomials* $P_n(x)$ and the Legendre functions $Q_n(x)$ are defined by the formulas

$$P_n(x) = \frac{1}{n!\,2^n}\frac{d^n}{dx^n}(x^2-1)^n,$$
$$Q_n(x) = \frac{1}{2}P_n(x)\ln\frac{1+x}{1-x} - \sum_{m=1}^{n}\frac{1}{m}P_{m-1}(x)P_{n-m}(x).$$

The polynomials $P_n(x)$ can be calculated using the formulas

$$P_0(x) = 1, \quad P_1(x) = x, \quad P_2(x) = \frac{1}{2}(3x^2-1),$$
$$P_3(x) = \frac{1}{2}(5x^3-3x), \quad P_4(x) = \frac{1}{8}(35x^4-30x^2+3),$$
$$P_{n+1}(x) = \frac{2n+1}{n+1}xP_n(x) - \frac{n}{n+1}P_{n-1}(x).$$

The first five functions $Q_n(x)$ have the form

$$Q_0(x) = \frac{1}{2}\ln\frac{1+x}{1-x}, \quad Q_1(x) = \frac{x}{2}\ln\frac{1+x}{1-x} - 1,$$
$$Q_2(x) = \frac{1}{4}(3x^2-1)\ln\frac{1+x}{1-x} - \frac{3}{2}x, \quad Q_3(x) = \frac{1}{4}(5x^3-3x)\ln\frac{1+x}{1-x} - \frac{5}{2}x^2 + \frac{2}{3},$$
$$Q_4(x) = \frac{1}{16}(35x^4-30x^2+3)\ln\frac{1+x}{1-x} - \frac{35}{8}x^3 + \frac{55}{24}x.$$

The polynomials $P_n(x)$ have the explicit representation

$$P_n(x) = 2^{-n}\sum_{m=0}^{[n/2]}(-1)^m C_n^m C_{2n-2m}^n x^{n-2m},$$

where $[A]$ stands for the integer part of a number A.

► **Zeros and orthogonality of the Legendre polynomials.** The polynomials $P_n(x)$ (with natural n) have exactly n real distinct zeros; all zeros lie on the interval $-1 < x < 1$. The zeros of $P_n(x)$ and $P_{n+1}(x)$ alternate with each other. The function $Q_n(x)$ has exactly $n+1$ zeros, which lie on the interval $-1 < x < 1$.

The functions $P_n(x)$ form an orthogonal system on the interval $-1 \le x \le 1$, with

$$\int_{-1}^{1} P_n(x)P_m(x)\,dx = \begin{cases} 0 & \text{if } n \neq m, \\ \dfrac{2}{2n+1} & \text{if } n = m. \end{cases}$$

► **Generating functions.** The generating function for Legendre polynomials is

$$\frac{1}{\sqrt{1-2sx+s^2}} = \sum_{n=0}^{\infty} P_n(x)s^n \qquad (|s| < 1).$$

The generating function for Legendre functions is

$$\frac{1}{\sqrt{1-2sx+s^2}} \ln\left[\frac{x-s+\sqrt{1-2sx+s^2}}{\sqrt{1-x^2}}\right] = \sum_{n=0}^{\infty} Q_n(x)s^n \qquad (|s| < 1,\ x > 1).$$

► **Associated Legendre functions with integer indices. Differential equation.** The *associated Legendre functions* $P_n^m(x)$ of order m are defined by the formulas

$$P_n^m(x) = (1-x^2)^{m/2}\frac{d^m}{dx^m}P_n(x), \qquad n = 1,\ 2,\ 3,\ \dots, \quad m = 0,\ 1,\ 2,\ \dots$$

It is assumed by definition that $P_n^0(x) = P_n(x)$.

Properties:

$$P_n^m(x) = 0 \quad \text{if} \quad m > n, \qquad P_n^m(-x) = (-1)^{n-m}P_n^m(x).$$

The associated Legendre functions $P_n^m(x)$ have exactly $n-m$ distinct real zeros, all of which lie on the interval $-1 < x < 1$.

Some of the associated Legendre functions $P_n^m(x)$ with lower indices are

$$P_1^1(x) = (1-x^2)^{1/2}, \quad P_2^1(x) = 3x(1-x^2)^{1/2}, \quad P_2^2(x) = 3(1-x^2),$$
$$P_3^1(x) = \tfrac{3}{2}(5x^2-1)(1-x^2)^{1/2}, \quad P_3^2(x) = 15x(1-x^2), \quad P_3^3(x) = 15(1-x^2)^{3/2}.$$

The associated Legendre functions $P_n^m(x)$ with $n > m$ are solutions of the linear ordinary differential equation

$$(1-x^2)y''_{xx} - 2xy'_x + \left[n(n+1) - \frac{m^2}{1-x^2}\right]y = 0.$$

► **Orthogonality of the associated Legendre functions.** The functions $P_n^m(x)$ form an orthogonal system on the interval $-1 \le x \le 1$, with

$$\int_{-1}^{1} P_n^m(x)P_k^m(x)\,dx = \begin{cases} 0 & \text{if } n \neq k, \\ \dfrac{2}{2n+1}\dfrac{(n+m)!}{(n-m)!} & \text{if } n = k. \end{cases}$$

The functions $P_n^m(x)$ (with $m \neq 0$) are orthogonal on the interval $-1 \le x \le 1$ with weight $(1-x^2)^{-1}$, that is,

$$\int_{-1}^{1} \frac{P_n^m(x)P_n^k(x)}{1-x^2}\,dx = \begin{cases} 0 & \text{if } m \neq k, \\ \dfrac{(n+m)!}{m(n-m)!} & \text{if } m = k. \end{cases}$$

M13.10.2. Laguerre Polynomials and Generalized Laguerre Polynomials

► **Laguerre polynomials.** The Laguerre polynomials $L_n(x)$ satisfy the second-order linear ordinary differential equation

$$xy''_{xx} + (1-x)y'_x + ny = 0$$

and are defined by the formulas

$$L_n(x) = \frac{1}{n!}e^x \frac{d^n}{dx^n}\big(x^n e^{-x}\big) = \frac{(-1)^n}{n!}\left[x^n - n^2 x^{n-1} + \frac{n^2(n-1)^2}{2!}x^{n-2} + \cdots\right].$$

The first four polynomials have the form

$$L_0(x) = 1, \quad L_1(x) = -x+1, \quad L_2(x) = \tfrac{1}{2}(x^2-4x+2), \quad L_3(x) = \tfrac{1}{6}(-x^3+9x^2-18x+6).$$

To calculate $L_n(x)$ for $n \ge 2$, one can use the recurrence formulas

$$L_{n+1}(x) = \frac{1}{n+1}\big[(2n+1-x)L_n(x) - nL_{n-1}(x)\big].$$

The functions $L_n(x)$ form an orthonormal system on the interval $0 < x < \infty$ with weight e^{-x}:

$$\int_0^\infty e^{-x} L_n(x) L_m(x)\,dx = \begin{cases} 0 & \text{if } n \ne m, \\ 1 & \text{if } n = m. \end{cases}$$

The generating function is

$$\frac{1}{1-s}\exp\Big(-\frac{sx}{1-s}\Big) = \sum_{n=0}^\infty L_n(x)s^n, \qquad |s| < 1.$$

► **Generalized Laguerre polynomials.** The generalized Laguerre polynomials $L_n^\alpha(x)$ $(\alpha > -1)$ satisfy the equation

$$xy''_{xx} + (\alpha+1-x)y'_x + ny = 0$$

and are defined by the formulas

$$L_n^\alpha(x) = \frac{1}{n!}x^{-\alpha}e^x\frac{d^n}{dx^n}\big(x^{n+\alpha}e^{-x}\big) = \sum_{m=0}^n C_{n+\alpha}^{n-m}\frac{(-x)^m}{m!} = \sum_{m=0}^n \frac{\Gamma(n+\alpha+1)}{\Gamma(m+\alpha+1)}\,\frac{(-x)^m}{m!\,(n-m)!}.$$

Notation: $L_n^0(x) = L_n(x)$.
Special cases:

$$L_0^\alpha(x) = 1, \quad L_1^\alpha(x) = \alpha+1-x, \quad L_n^{-n}(x) = (-1)^n\frac{x^n}{n!}.$$

To calculate $L_n^\alpha(x)$ for $n \ge 2$, one can use the recurrence formulas

$$L_{n+1}^\alpha(x) = \frac{1}{n+1}\big[(2n+\alpha+1-x)L_n^\alpha(x) - (n+\alpha)L_{n-1}^\alpha(x)\big].$$

Other recurrence formulas:

$$L_n^\alpha(x)=L_{n-1}^\alpha(x)+L_n^{\alpha-1}(x),\quad \frac{d}{dx}L_n^\alpha(x)=-L_{n-1}^{\alpha+1}(x),\quad x\frac{d}{dx}L_n^\alpha(x)=nL_n^\alpha(x)-(n+\alpha)L_{n-1}^\alpha(x).$$

The functions $L_n^\alpha(x)$ form an orthogonal system on the interval $0 < x < \infty$ with weight $x^\alpha e^{-x}$:

$$\int_0^\infty x^\alpha e^{-x} L_n^\alpha(x) L_m^\alpha(x)\,dx = \begin{cases} 0 & \text{if } n \neq m, \\ \frac{\Gamma(\alpha+n+1)}{n!} & \text{if } n = m. \end{cases}$$

The generating function is

$$(1-s)^{-\alpha-1}\exp\Bigl(-\frac{sx}{1-s}\Bigr) = \sum_{n=0}^\infty L_n^\alpha(x)s^n, \qquad |s| < 1.$$

M13.10.3. Chebyshev Polynomials

► **Chebyshev polynomials of the first kind.** The *Chebyshev polynomials of the first kind* $T_n(x)$ satisfy the second-order linear ordinary differential equation

$$(1-x^2)y''_{xx} - xy'_x + n^2y = 0$$

and are defined by the formulas

$$\begin{aligned} T_n(x) = \cos(n \arccos x) &= \frac{(-2)^n n!}{(2n)!}\sqrt{1-x^2}\,\frac{d^n}{dx^n}\bigl[(1-x^2)^{n-\frac12}\bigr] \\ &= \frac{n}{2}\sum_{m=0}^{[n/2]}(-1)^m\frac{(n-m-1)!}{m!\,(n-2m)!}(2x)^{n-2m} \qquad (n = 0, 1, 2, \dots), \end{aligned}$$

where $[A]$ stands for the integer part of a number A.

An alternative representation of the Chebyshev polynomials:

$$T_n(x) = \frac{(-1)^n}{(2n-1)!!}(1-x^2)^{1/2}\frac{d^n}{dx^n}(1-x^2)^{n-1/2}.$$

The first five Chebyshev polynomials of the first kind are

$$T_0(x) = 1,\quad T_1(x) = x,\quad T_2(x) = 2x^2 - 1,\quad T_3(x) = 4x^3 - 3x,\quad T_4(x) = 8x^4 - 8x^2 + 1.$$

The recurrence formulas:

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \qquad n \geq 2.$$

The functions $T_n(x)$ form an orthogonal system on the interval $-1 < x < 1$ with weight $(1-x^2)^{-1/2}$:

$$\int_{-1}^1 \frac{T_n(x)T_m(x)}{\sqrt{1-x^2}}\,dx = \begin{cases} 0 & \text{if } n \neq m, \\ \frac12\pi & \text{if } n = m \neq 0, \\ \pi & \text{if } n = m = 0. \end{cases}$$

The generating function is

$$\frac{1-sx}{1-2sx+s^2} = \sum_{n=0}^\infty T_n(x)s^n \qquad (|s| < 1).$$

The functions $T_n(x)$ have only real simple zeros, all lying on the interval $-1 < x < 1$.

The normalized Chebyshev polynomials of the first kind, $2^{1-n}T_n(x)$, deviate from zero least of all. This means that among all polynomials of degree n with the leading coefficient 1, the maximum of the modulus $\max\limits_{-1\leq x\leq 1}|2^{1-n}T_n(x)|$ has the least value, with the maximum being equal to 2^{1-n}.

► **Chebyshev polynomials of the second kind.** The *Chebyshev polynomials of the second kind* $U_n(x)$ satisfy the second-order linear ordinary differential equation

$$(1-x^2)y''_{xx} - 3xy'_x + n(n+2)y = 0$$

and are defined by the formulas

$$U_n(x) = \frac{\sin[(n+1)\arccos x]}{\sqrt{1-x^2}} = \frac{2^n(n+1)!}{(2n+1)!}\frac{1}{\sqrt{1-x^2}}\frac{d^n}{dx^n}(1-x^2)^{n+1/2}$$
$$= \sum_{m=0}^{[n/2]}(-1)^m\frac{(n-m)!}{m!\,(n-2m)!}(2x)^{n-2m} \qquad (n=0,\ 1,\ 2,\ \ldots).$$

The first five Chebyshev polynomials of the second kind are

$U_0(x)=1,\quad U_1(x)=2x,\quad U_2(x)=4x^2-1,\quad U_3(x)=8x^3-4x,\quad U_4(x)=16x^4-12x^2+1.$

The recurrence formulas:

$$U_{n+1}(x) = 2xU_n(x) - U_{n-1}(x), \qquad n \ge 2.$$

The generating function is

$$\frac{1}{1-2sx+s^2} = \sum_{n=0}^{\infty} U_n(x)s^n \qquad (|s|<1).$$

The Chebyshev polynomials of the first and second kinds are related by

$$U_n(x) = \frac{1}{n+1}\frac{d}{dx}T_{n+1}(x).$$

M13.10.4. Hermite Polynomials

► **Various representations of the Hermite polynomials.** The *Hermite polynomials* $H_n(x)$ satisfy the second-order linear ordinary differential equation

$$y''_{xx} - 2xy'_x + 2ny = 0$$

and are defined by the formulas

$$H_n(x) = (-1)^n\exp\bigl(x^2\bigr)\frac{d^n}{dx^n}\exp\bigl(-x^2\bigr) = \sum_{m=0}^{[n/2]}(-1)^m\frac{n!}{m!\,(n-2m)!}(2x)^{n-2m}.$$

The first five polynomials are

$H_0(x)=1,\quad H_1(x)=2x,\quad H_2(x)=4x^2-2,\quad H_3(x)=8x^3-12x,\quad H_4(x)=16x^4-48x^2+12.$

Recurrence formulas:

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x), \qquad n \ge 2;$$
$$\frac{d}{dx}H_n(x) = 2nH_{n-1}(x).$$

Integral representation:

$$H_{2n}(x) = \frac{(-1)^n 2^{2n+1}}{\sqrt{\pi}}\exp\bigl(x^2\bigr)\int_0^\infty \exp\bigl(-t^2\bigr)t^{2n}\cos(2xt)\,dt,$$
$$H_{2n+1}(x) = \frac{(-1)^n 2^{2n+2}}{\sqrt{\pi}}\exp\bigl(x^2\bigr)\int_0^\infty \exp\bigl(-t^2\bigr)t^{2n+1}\sin(2xt)\,dt,$$

where $n=0,\ 1,\ 2,\ \ldots$

► **Orthogonality. The generating function. An asymptotic formula.** The functions $H_n(x)$ form an orthogonal system on the interval $-\infty < x < \infty$ with weight e^{-x^2}:

$$\int_{-\infty}^{\infty} \exp\bigl(-x^2\bigr) H_n(x) H_m(x)\,dx = \begin{cases} 0 & \text{if } n \neq m, \\ \sqrt{\pi}\,2^n n! & \text{if } n = m. \end{cases}$$

Generating function:

$$\exp\bigl(-s^2 + 2sx\bigr) = \sum_{n=0}^{\infty} H_n(x) \frac{s^n}{n!}.$$

Asymptotic behavior as $n \to \infty$:

$$H_n(x) \approx 2^{\frac{n+1}{2}} n^{\frac{n}{2}} e^{-\frac{n}{2}} \exp\bigl(x^2\bigr) \cos\bigl(\sqrt{2n+1}\,x - \tfrac{1}{2}\pi n\bigr).$$

► **Hermite functions.** The *Hermite functions* $h_n(x)$ are introduced by the formula

$$h_n(x) = \exp\Bigl(-\frac{1}{2}x^2\Bigr) H_n(x) = (-1)^n \exp\Bigl(\frac{1}{2}x^2\Bigr) \frac{d^n}{dx^n} \exp\bigl(-x^2\bigr), \qquad n = 0, 1, 2, \ldots$$

The Hermite functions satisfy the second-order linear ordinary differential equation

$$h''_{xx} + (2n + 1 - x^2)h = 0.$$

The functions $h_n(x)$ form an orthogonal system on the interval $-\infty < x < \infty$:

$$\int_{-\infty}^{\infty} h_n(x) h_m(x)\,dx = \begin{cases} 0 & \text{if } n \neq m, \\ \sqrt{\pi}\,2^n n! & \text{if } n = m. \end{cases}$$

Bibliography for Chapter M13

Abramowitz, M. and Stegun, I. A. (Editors), *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, National Bureau of Standards Applied Mathematics Series, Washington, D.C., 1964.

Bateman, H. and Erdélyi, A., *Higher Transcendental Functions, Vol. 1 and Vol. 2*, McGraw-Hill, New York, 1953.

Bateman, H. and Erdélyi, A., *Higher Transcendental Functions, Vol. 3*, McGraw-Hill, New York, 1955.

Gradshteyn, I. S. and Ryzhik, I. M., *Tables of Integrals, Series, and Products*, Academic Press, New York, 1980.

Magnus, W., Oberhettinger, F., and Soni, R. P., *Formulas and Theorems for the Special Functions of Mathematical Physics, 3rd Edition*, Springer-Verlag, Berlin, 1966.

McLachlan, N. W., *Bessel Functions for Engineers*, Clarendon Press, Oxford, 1955.

Polyanin, A. D. and Zaitsev, V. F., *Handbook of Exact Solutions for Ordinary Differential Equations, 2nd Edition*, Chapman & Hall/CRC Press, Boca Raton, Florida, 2003.

Slavyanov, S. Yu. and Lay, W., *Special Functions: A Unified Theory Based on Singularities*, Oxford University Press, Oxford, 2000.

Weisstein, E. W., *CRC Concise Encyclopedia of Mathematics, 2nd Edition*, CRC Press, Boca Raton, Florida, 2003.

Zwillinger, D., *CRC Standard Mathematical Tables and Formulae, 31st Edition*, CRC Press, Boca Raton, Florida, 2002.

Chapter M14

Probability Theory

M14.1. Basic Concepts and Simplest Probabilistic Models

M14.1.1. Rules and Formulas of Combinatorics

Below are rules and formulas of combinatorics that are useful in solving problems of probability theory.

► **Rule of sum and rule of product.**

Rule of sum. Suppose there is a collection of objects (of arbitrary nature). If an object α can be chosen from this collection in m different ways and another object β can be chosen in k ways, then one can choose *either object α or object β* in $m+k$ different ways.

Rule of product. If an object α can be chosen in m ways and after that an object β can be chosen in k ways, then one can choose the *ordered pair of objects* (α, β) in mk ways.

Example 1. How many two-digit numbers are there consisting of different digits?

The first digit can be one of 1, 2, ..., 9 and the other, one of the eight remaining plus 0 (i.e., one of nine digits). So, by the rule of product, there are $9\times 9 = 81$ two-digit numbers consisting of different digits.

► **Permutations and combinations.**

Given a set of n (distinct) elements, an arrangement of r elements from this set ($r \le n$) taken in a certain order is called a *permutation*. The total number of different permutations is equal to

$$P_n^r = n(n-1)(n-2)\dots(n-r+1) = \frac{n!}{(n-r)!},$$

which is also denoted $P_r^n = {}_nP_r = P(n,r)$.

Example 2. How many three-digit numbers can be made up from the six digits 0, 1, 2, 3, 4, 5 without repetitions?

Three out of the six digits (in a certain order) can be selected in P_6^3 ways. The triplets that start with 0 (i.e., the number of two-digit numbers consisting of 1, ..., 5) must be subtracted from this. So the desired number of three-digit numbers is $P_6^3 - P_5^2 = 6\cdot 5\cdot 4 - 5\cdot 4 = 100$.

Note that the number of different permutations on a set of n elements (i.e., arrangements that differ from each other by only the order of selection of the n elements) is equal to

$$P_n^n = n! = 1\cdot 2\cdot 3\dots n.$$

Note that $0! = 1$.

Example 3. How many ways can four people be seated at a table?

This can be done in $P_4^4 = 4! = 1\cdot 2\cdot 3\cdot 4 = 24$ ways.

An arrangement of r elements selected from a set of n elements where the order is not important is called a *combination*. The number of combinations is equal to

$$C_n^r = \frac{P_n^r}{P_r^r} = \frac{n(n-1)(n-2)\dots(n-r+1)}{r!} = \frac{n!}{r!\,(n-r)!}$$

and is also denoted $C_r^n = {}_nC_r = C(n,r) = \binom{n}{r}$. Note that the relation $C_n^r = C_n^{n-r}$ holds.

Example 4. How many ways can three balls be selected from a box containing six numbered balls?

The desired number equals the number of combinations of three elements from a set of six elements, or $C_6^3 = \dfrac{6 \cdot 5 \cdot 4}{3 \cdot 2 \cdot 1} = 20$.

M14.1.2. Probabilities of Random Events

► **Random events. Basic definitions.** The simplest mutually exclusive outcomes of an experiment are called *elementary events* ω_i. The set of all elementary outcomes, which we denote by the symbol Ω, is called the *space of elementary events* (or the *sample space*). Any subset of Ω is called a *random event* A (or simply an *event* A). Event A is said to occur when the outcome of the experiment is one of the elementary events that make up A.

Example 1. Suppose the experiment is rolling a six-sided die once. There are six possible mutually exclusive outcomes and the space of elementary events is $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$, where elementary event ω_k stands for rolling a k on the die. If event A is rolling an odd number on the die, then A occurs when the outcome is one of the elementary events ω_1, ω_3, ω_5.

An event A *implies* an event B ($A \subseteq B$) if B occurs when A occurs. Events A and B are said to be *equivalent* ($A = B$) if A implies B and B implies A, i.e., if both events A and B occur or do not occur simultaneously.

The *union* $C = A \cup B = A + B$ of events A and B is the event that at least one of the events A or B occurs. The elementary outcomes of the union $A + B$ are the elementary outcomes that belong to at least one of the events A and B.

The *intersection* $C = A \cap B = AB$ of events A and B is the event that both A and B occur. The elementary outcomes of the intersection AB are the elementary outcomes that simultaneously belong to A and B.

The *difference* $C = A \backslash B = A - B$ of events A and B is the event that A occurs and B does not occur. The elementary outcomes of the difference $A \backslash B$ are the elementary outcomes of A that do not belong to B.

The event that A does not occur is called the *complement* of A, or the *complementary event*, and is denoted by $\overline{A}$. The elementary outcomes of $\overline{A}$ are the elementary outcomes that do not belong to event A.

The set Ω is called a *sure event* or a *certain event* (it will always happen). The empty set $\varnothing$ is called an *impossible event* (it will never happen).

Two complementary events A and $\overline{A}$ are said to be *opposite*; they simultaneously satisfy the following two conditions:

$$A \cup \overline{A} = \Omega, \quad A \cap \overline{A} = \varnothing.$$

Events A and B are said to be *incompatible*, or *mutually exclusive*, if their simultaneous realization is impossible, i.e., if $A \cap B = \varnothing$.

Events $H_1, \dots, H_n$ are said to form a *complete group of events*, or to be *collectively exhaustive*, if at least one of them necessarily occurs for each trial of the experiment, i.e., if

$$H_1 \cup \cdots \cup H_n = \Omega.$$

Events $H_1, \dots, H_n$ form a *complete group of mutually exclusive* or *pairwise incompatible events* if exactly one of the events necessarily occurs for each trial of the experiment, i.e., if

$$H_1 \cup \cdots \cup H_n = \Omega \quad \text{and} \quad H_i \cap H_j = \varnothing \quad (i \neq j).$$

Main properties of random events:

1. $A \cup B = B \cup A$ and $A \cap B = B \cap A$ (commutativity).

2. $(A\cup B)\cap C=(A\cap C)\cup(B\cap C)$ and $(A\cap B)\cup C=(A\cup C)\cap(B\cup C)$ (distributivity).
3. $(A\cup B)\cup C=A\cup(B\cup C)$ and $(A\cap B)\cap C=A\cap(B\cap C)$ (associativity).
4. $A\cup A=A$ and $A\cap A=A$.
5. $A\cup\Omega=\Omega$ and $A\cap\Omega=A$.
6. $A\cup\overline{A}=\Omega$ and $A\cap\overline{A}=\varnothing$.
7. $\overline{\varnothing}=\Omega$, $\overline{\Omega}=\varnothing$, and $\overline{\overline{A}}=A$.
8. $A\setminus B=A\cap\overline{B}$.
9. $\overline{A\cup B}=\overline{A}\cap\overline{B}$ and $\overline{A\cap B}=\overline{A}\cup\overline{B}$ (*de Morgan's laws*).

► **Discrete probability space. Classical definition of probability.** Suppose that $\Omega=\{\omega_1,\dots,\omega_n\}$ is a finite sample space. To each elementary event $\omega_i\in\Omega$ $(i=1,2,\dots,n)$ there corresponds a number $p(\omega_i)$, called the probability of the elementary event ω_i. Thus a real function satisfying the following two conditions is defined on the set Ω:

1. *Nonnegativity condition*: $p(\omega_i)\ge 0$ for any $\omega_i\in\Omega$.
2. *Normalization condition*: $\sum_{i=1}^{n} p(\omega_i)=1$.

The *probability* $P(A)$ of an event A for any subset $A\subset\Omega$ is defined to be the sum of the probabilities of the elementary events that form A; i.e.,

$$P(A)=\sum_{\omega_i\in A} p(\omega_i). \tag{M14.1.2.1}$$

Note that the relations $0\le P(A)\le 1$, $P(\varnothing)=0$, and $P(\Omega)=1$ are always valid.

A special case of the definition of probability (M14.1.2.1) is the *classical definition of probability*, in which all elementary events are equally likely: $p(\omega_1)=\dots=p(\omega_n)=1/n$. Then the probability of the event A is

$$P(A)=\frac{m}{n}, \tag{M14.1.2.2}$$

where m is the number of elementary events making up A (the number *favorable* outcomes).

Example 2. In example 1, the probability of rolling an odd number is

$$P(A)=\tfrac{3}{6}=\tfrac{1}{2}.$$

Example 3. Let two dice be rolled. Under the assumption that the elementary events are equiprobable, find the probability of the event A that the sum of the numbers rolled is greater than 10. Obviously, the sample space can be represented as $\Omega=\{(i,j):i,j=1,2,3,4,5,6\}$, where i is the number shown by the first die and j is that shown by the second die. The total number of elementary events is $|\Omega|=36$. Event A is represented by the subset $A=\{(5,6),(6,5),(6,6)\}$ of Ω. Since $|A|=3$, formula (M14.1.2.2) gives $P(A)=|A|/|\Omega|=3/36=1/12$.

► **Statistical definition of probability.** Suppose an experiment has been repeated N times and an event A has occurred k times. The ratio k/N is called the *relative frequency* of the event A in the given series of experiments. The number about which the relative frequency oscillates in long series of experiments is called the probability $P(A)$ of the event A.

► **Geometric definition of probability.** In the geometric approach to defining probability, one takes an arbitrary set on a straight line, in a plane, or in space to be the sample space Ω. A trial of an experiment is interpreted as a random selection of a point in Ω. The occurrence of an event A is treated as the occurrence of the point in a subdomain A of Ω. It is assumed that the probability of selecting a point in the domain A is proportional to the measure of A (i.e., to its length, area, or volume) and is independent of the position and shape of the domain. The probability of the event A is defined as

$$P(A)=\frac{\text{measure }A}{\text{measure }\Omega}.$$

Example 3. A point is randomly dropped inside a circle of radius $R = 1$. Find the probability of the event that the point lands in the circle of radius $r = \frac{1}{2}$ concentric with the larger one.

Let A be the event that the point lands in the smaller circle. We find the probability $P(A)$ as the ratio of the area of the smaller circle to that of the larger one:

$$P(A) = \frac{\pi r^2}{\pi R^2} = \frac{1}{4}.$$

► **Axiomatic definition of probability.** The *probability of an event* is defined to be a single-valued real function $P(A)$ satisfying the following three axioms:

1. *Nonnegativity*: $P(A) \geq 0$ for any $A \in \Omega$.
2. *Normalization*: $P(\Omega) = 1$.
3. *Additivity*: $P\Bigl(\bigcup\limits_n A_n\Bigr) = \sum\limits_n P(A_n)$, provided that $A_i \cap A_j = \varnothing$ whenever $i \neq j$.

Properties of probability:

1. The probability of an impossible event is zero; i.e., $P(\varnothing) = 0$.
2. The probability of the event $\overline{A}$ opposite to an event A is equal to $P(\overline{A}) = 1 - P(A)$.
3. Probability is a bounded function; i.e., $0 \leq P(A) \leq 1$.
4. If an event A implies an event B ($A \subseteq B$), then $P(A) \leq P(B)$.
5. If events $A_1, \ldots, A_n$ form a complete group of pairwise incompatible events, then $\sum\limits_{i=1}^{n} P(A_i) = 1$.

M14.1.3. Conditional Probability. Total Probability Formula

► **Probability of the union of events.** The probability of realization of at least one of two events H_1 and H_2 is given by the formula (addition theorem)

$$P(H_1 \cup H_2) = P(H_1) + P(H_2) - P(H_1 \cap H_2).$$

In particular, for $H_1 \cap H_2 = \varnothing$, we have

$$P(H_1 \cup H_2) = P(H_1) + P(H_2).$$

The probability of realization of at least one of n events is given by the formula

$$\begin{aligned} P(H_1 \cup \cdots \cup H_n) &= \sum_{k=1}^{n} P(H_k) - \sum_{1\leq k_1<k_2\leq n} P(H_{k_1} \cap H_{k_2}) \\ &\quad + \sum_{1\leq k_1<k_2<k_3\leq n} P(H_{k_1} \cap H_{k_2} \cap H_{k_3}) - \cdots + (-1)^{n-1} P(H_1 \cap \cdots \cap H_n). \qquad \text{(M14.1.3.1)} \end{aligned}$$

Note that $P(H_1 \cup H_2)$ is also written $P(H_1 + H_2)$.

For n pairwise incompatible events $H_1, H_2, \ldots, H_n$ ($H_i \cap H_j = \varnothing$ for $i \neq j$), formula (M14.1.3.1) simplifies to become

$$P(H_1 \cup \cdots \cup H_n) = P(H_1) + P(H_2) + \cdots + P(H_n).$$

It may often be convenient to calculate the probability of the sum of events via the probability of the product of opposite events:

$$P(H_1 \cup \cdots \cup H_n) = 1 - P(\bar{H}_1 \cap \bar{H}_2 \cap \cdots \cap \bar{H}_n).$$

► **Conditional probability.** The *conditional probability* $P(A|H)$ of an event A given the occurrence of some other event H is defined by the formula

$$P(A|H) = \frac{P(A \cap H)}{P(H)}, \quad P(H) > 0. \qquad \text{(M14.1.3.2)}$$

Relation (M14.1.3.2) can be rewritten (multiplication theorem)

$$P(A \cap H) = P(H)\, P(A|H). \qquad \text{(M14.1.3.3)}$$

The probability $P(A \cap H)$ is also written $P(AH)$.

The formula

$$P(A_1 \cap \cdots \cap A_n) = P(A_1)\, P(A_2|A_1)\, P(A_3|A_1 \cap A_2) \ldots P(A_n|A_1 \cap \cdots \cap A_{n-1})$$

is a generalization of (M14.1.3.3).

Example 1. Two guns shoot one and the same target independently, once each. The probability for the first gun to hit the target is 0.8 and that for the second gun is 0.9. It is required to find:

a) the probability that the target is hit only once and

b) the probability that the target is hit at least once.

a) Let A and B denote the events that the first and second guns hit the target, respectively. Then the probability that the target is hit only once is $P(A\bar{B} + \bar{A}B) = P(A\bar{B}) + P(\bar{A}B) = P(A)P(\bar{B}) + P(\bar{A})P(B) = 0.8\,(1 - 0.9) + (1 - 0.8)\,0.9 = 0.26$.

b) The probability of at least one hit is $P(A+B) = P(A) + P(B) - P(A)P(B) = 0.8 + 0.9 - 0.8 \cdot 0.9 = 0.98$. Note that this result is usually obtained as follows: $P(A+B) = 1 - P(\bar{A}\bar{B}) = 1 - P(\bar{A})P(\bar{B}) = 1 - 0.2 \cdot 0.1 = 0.98$.

► **Independence of events.** Two random events A and B are said to be *independent* if the conditional probability of A, given the occurrence of B, coincides with the unconditional probability of A,

$$P(A|B) = P(A).$$

In this case, $P(A \cap B) = P(A)P(B)$.

Random events $A_1, \ldots, A_n$ are jointly independent if the relation

$$P\biggl(\bigcap_{k=1}^{m} A_{i_k}\biggr) = \prod_{k=1}^{m} P(A_{i_k})$$

holds whenever $1 \le i_1 < \cdots < i_m \le n$ and $m \le n$.

The pairwise independence of the events A_i and A_j for all $i \ne j$ $(i, j = 1, 2, \ldots, n)$ does not imply that the events $A_1, \ldots, A_n$ are jointly independent.

Example 2. Suppose that the experiment is to draw one of four balls. Let three of them be labeled by the numbers 1, 2, and 3, and let the fourth ball bear all these numbers. By A_i $(i = 1, 2, 3)$ we denote the event that the chosen ball bears the number i. Are the events A_1, A_2, and A_3 dependent?

Since each number is encountered twice, $P(A_1) = P(A_2) = P(A_3) = 1/2$. Since any two distinct numbers are present only on one of the balls, we have $P(A_1A_2) = P(A_2A_3) = P(A_1A_3) = 1/4$, and hence the events A_1, A_2, and A_3 are pairwise independent. All three distinct numbers are present only on one of the balls, and $P(A_1A_2A_3) = 1/4 \ne P(A_1)\,P(A_2)\,P(A_3) = 1/8$.

Thus we see that the events A_1, A_2, and A_3 are jointly dependent, even though they are pairwise independent.

► **Total probability formula. Bayes's formula.** Suppose that a group of pairwise incompatible events $H_1, \ldots, H_n$ is given and their probabilities $P(H_1), \ldots, P(H_n)$ as well as the conditional probabilities $P(A|H_1), \ldots, P(A|H_n)$ of an event A are known. Then the probability of A can be determined by the *total probability formula*

$$P(A) = \sum_{k=1}^{n} P(H_k)\, P(A|H_k). \qquad \text{(M14.1.3.4)}$$

The events H_k are called *hypotheses* with respect to the event A.

Example 3. There are two urns: the first urn contains a white balls and b black ones, and the second urn contains c white balls and d black ones. We take one ball from the first urn and put it into the second urn. After this, we draw one ball from the second urn. Find the probability of the event that this ball is white.

Let A be the event of drawing a white ball. Consider the following group of events:

H_1, a white ball is taken from the first urn and put into the second urn.

H_2, a black ball is taken from the first urn and put into the second urn. Obviously,

$$P(H_1)=\frac{a}{a+b},\quad P(H_2)=\frac{b}{a+b};\quad P(A|H_1)=\frac{c+1}{c+d+1},\quad P(A|H_2)=\frac{c}{c+d+1}.$$

Now by the total probability formula (M14.1.3.4) we obtain

$$P(A)=P(H_1)\,P(A|H_1)+P(H_2)\,P(A|H_2)=\frac{a}{a+b}\,\frac{c+1}{c+d+1}+\frac{b}{a+b}\,\frac{c}{c+d+1}.$$

If it is known that the event A has occurred as a result of one trial but it is unknown which of the events $H_1,\dots,H_n$ has occurred, then the conditional probabilities of the hypotheses H_k (under the condition that the event A has occurred) are calculated using *Bayes's formula*:

$$P(H_k|A)=\frac{P(H_k)\,P(A|H_k)}{P(A)}\qquad(k=1,2,\dots,n).$$

Example 4. A box contains same parts made by two automats, 40% produced by the first automat and the rest by the second. It is known that 3% of the first automat's products are faulty, while the percentage of faulty products made by the second automat is 2%. Find

a) the probability that a randomly chosen part will be faulty and

b) the probability that a randomly chosen part was made by the first automat if found faulty.

Let A denote the event that a randomly chosen part is faulty and let H_1 and H_2 denote the events that a part is made by the first and second automat, respectively. Then

a) the total probability formula gives

$$P(A)=P(H_1)\,P(A|H_1)+P(H_2)\,P(A|H_2)=0.4\cdot 0.03+(1-0.4)\cdot 0.02=0.024;$$

b) Bayes's formula gives

$$P(H_1|A)=\frac{P(H_1)\,P(A|H_1)}{P(A)}=\frac{0.4\cdot 0.03}{0.024}=0.5.$$

M14.1.4. Sequence of Trials

► **Bernoulli process.** In this case, some event A occurs with probability $p=P(A)$ (the probability of "success") and does not occur with probability $q=P(\overline{A})=1-P(A)=1-p$ (the probability of "failure") in each trial. The probability that in n independent trials, the event A ("success") occurs exactly k times is determined by the *Bernoulli formula (binomial distribution)*:

$$P_n(k)=C_n^k p^k(1-p)^{n-k}\qquad(k=0,1,\dots,n).\qquad\text{(M14.1.4.1)}$$

Note that the relation $\sum\limits_{k=0}^{n}P_n(k)=1$ holds.

Example 1. Find the probability that there will be 5 heads in tossing a coin 10 times.

We have $n=10$, $k=5$, and $p=\frac{1}{2}$. Then, by the Bernoulli formula, the desired probability is $P_{10}(5)=C_{10}^5\bigl(\frac{1}{2}\bigr)^5\bigl(\frac{1}{2}\bigr)^{10-5}=\frac{10\cdot9\cdot8\cdot7\cdot6}{5\cdot4\cdot3\cdot2\cdot1}\cdot\frac{1}{2^{10}}=\frac{63}{256}$.

Example 2 (Banach's problem). A smoker mathematician has two matchboxes on him, each of which initially contains exactly n matches. Each time he needs to light a cigarette, he selects a matchbox at random. Find the probability of the event that as the mathematician takes out an empty box for the first time, precisely k matches will be left in the other box ($k \le n$).

The mathematician has taken matches $2n - k$ times, n out of them from the box that is eventually empty. This scheme corresponds to the scheme of $2n - k$ independent Bernoulli trials with n "successes." The probability of a "success" in a single trial is equal to 0.5. The desired probability can be found by the formula

$$P_{2n-k}(n) = C_{2n-k}^{n} p^{n} (1-p)^{n-k} = C_{2n-k}^{n} \left(\tfrac{1}{2}\right)^{2n-k}.$$

The probability that the event occurs at least m times in n independent trials is calculated by the formula

$$P_n(k \ge m) = \sum_{k=m}^{n} P_n(k) = 1 - \sum_{k=0}^{m-1} P_n(k).$$

The probability that the event occurs at least once in n independent trials is calculated by the formula

$$P_n(k \ge 1) = 1 - (1-p)^n.$$

The number n of independent trials necessary for the event to occur at least once with probability at least P is given by the formula

$$n \ge \frac{\ln(1-P)}{\ln(1-p)}.$$

► **Limit formulas for Bernoulli process.** It is very difficult to use Bernoulli's formula (M14.1.4.1) for large n an k. In this case, one has to use approximate formulas for calculating $P_n(k)$ with desired accuracy.

Poisson formula. If the number of independent trials increases unboundedly ($n \to \infty$) and the probability p simultaneously goes to zero ($p \to 0$) so that their product np is a constant ($np = \lambda = \text{const}$), then the probability $P_n(k)$ satisfies the limit relation

$$\lim_{n\to\infty} P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda}. \qquad \text{(M14.1.4.2)}$$

Local de Moivre–Laplace theorem. Suppose that $n \to \infty$, $p = \text{const}$, $0 < p < 1$, and $0 < c_1 \le x_{n,k} = (k - np)[np(1-p)]^{-1/2} \le c_2 < \infty$; then

$$P_n(k) = \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left[-\frac{(k-np)^2}{2np(1-p)}\right] \left[1 + O(1/\sqrt{n}\,)\right] \qquad \text{(M14.1.4.3)}$$

uniformly with respect to $x_{n,k} \in [c_1, c_2]$.

Integral de Moivre–Laplace theorem. Suppose that $n \to \infty$ and $p = \text{const}$, $0 < p < 1$. The probability that the number of successes k in n independent trials is between k_1 and k_2 can be approximately calculated by the formula

$$P_n\{k_1 \le k \le k_2\} \approx \Phi\left(\frac{k_2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{np(1-p)}}\right). \qquad \text{(M14.1.4.4)}$$

Here

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) dt$$

is the cumulative distribution function of the standard normal distribution (also known as the Laplace function), which is tabulated.

The approximate formula (M14.1.4.2) is normally used for $n \ge 50$ and $np \le 10$. The approximate formulas (M14.1.4.3) and (M14.1.4.4) normally are used for $np(1-p) > 9$.

► **Sequence of n independent trials.** In a series of trials, the trials are said to be *independent* if the occurring events are independent.

Let $p_k = P(A_k)$. Then the probability of the event that n_1 events A_1, n_2 events A_2, …, and n_k events A_k occur in n independent trials is equal to

$$P_n(n_1, \dots, n_k) = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}. \qquad \text{(M14.1.4.5)}$$

Remark. The probability (M14.1.4.5) is the coefficient of $x_1^{n_1} \dots x_k^{n_k}$ in the expansion of the polynomial $(p_1 x_1 + \dots + p_k x_k)^n$ in powers of $x_1, \dots, x_k$.

M14.2. Random Variables and Their Characteristics

M14.2.1. One-Dimensional Random Variables

► **Notion of a random variable. The distribution function of a random variable.** Let $\Omega = \{\omega\}$ be the space of elementary events. A *random variable* X is a real number function $X = X(\omega)$ defined on the set Ω.

Any rule (table, function, graph, or otherwise) that permits one to find the probabilities of events is usually called the distribution law of a random variable. In general, random variables can be *discrete* or *continuous*.

The *cumulative distribution function* of a random variable X is the function $F(x)$ whose value at every point x is equal to the probability of the event $\{X < x\}$:

$$F(x) = P(X < x).$$

Properties of the cumulative distribution function:

1. $0 \le F(x) \le 1$.
2. $\lim\limits_{x\to-\infty} F(x) = F(-\infty) = 0$, $\lim\limits_{x\to+\infty} F(x) = F(+\infty) = 1$.
3. If $x_2 > x_1$, then $F(x_2) \ge F(x_1)$.
4. $P(x_1 \le X < x_2) = F(x_2) - F(x_1)$.
5. $F(x)$ is continuous from the left, i.e., $\lim\limits_{x\to x_0-0} F(x) = F(x_0)$.

► **Discrete random variables.** Let X be a *discrete* random variable assuming the values $x_1, x_2, \dots, x_n, \dots$ with probabilities $p_1, p_2, \dots, p_n, \dots$, so its distribution law can be defined by the table

X	x_1	x_2	…	x_n	…
P	p_1	p_2	…	p_n	…

$$\Bigl(\sum_i p_i = 1\Bigr).$$

(Here and in what follows, it is assumed that the values of a discrete random variable X are arranged in ascending order, so that $x_1 < x_2 < \dots < x_{k-1} < x_k < \cdots$.)

In this case, the cumulative distribution function of a discrete random variable X is the step function defined as the sum

$$F(x) = P\{X < x\} = \sum_{x_n < x} p_n.$$

► **Continuous random variables. The probability density function.** A *continuous random variable* is defined by either a distribution function $F(x)$ or a *probability density function* $p(x)$, which are related by

$$p(x) = F'(x), \quad F(x) = \int_{-\infty}^{x} p(z)\,dz.$$

Sometimes, the distribution function and probability density function of a random variable X are denoted $F_X(x)$ and $p_X(x)$, respectively, rather than $F(x)$ and $p(x)$.

Properties of the probability density function:

1. $p(x) \ge 0$.
2. $p(-\infty) = p(+\infty) = 0$.
3. $\int_{-\infty}^{+\infty} p(x)\,dx = 1$.
4. $P\{a \le X \le b\} = \int_a^b p(x)\,dx$.

The differential $dF(x) = p(x)\,dx \approx P(x \le X \le x + dx)$ is called a *probability element*.

Remark. For continuous random variables, one always has $P(X = x_0) = 0$, but the event $\{X = x_0\}$ is not necessarily impossible.

► **Functions of random variables.** Suppose that a random variable Y is related to a random variable X by a functional dependence $Y = \varphi(X)$. If X is discrete, then, obviously, Y is also discrete. To find the distribution law of the random variable Y, it suffices to calculate the values $\varphi(x_i)$. If there are repeated values among $y_i = \varphi(x_i)$, then these repeated values are taken into account only once, the corresponding probabilities being added.

If X is a continuous random variable with probability density function $p(x)$, then, in general, the random variable Y is also continuous. The cumulative distribution function of Y is given by the formula

$$F_Y(y) = P(Y < y) = P(\varphi(X) < y) = \int_{\varphi(x)<y} p(x)\,dx. \qquad \text{(M14.2.1.1)}$$

If the function $y = \varphi(x)$ is differentiable and monotone on the entire range of the argument x, then the probability density function $p_Y(y)$ of the random variable Y is given by the formula

$$p_Y(y) = p(g(y))\,|g'_y(y)|,$$

where g is the inverse of $\varphi(x)$.

Example 1. Suppose that a random variable X has the probability density

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Find the distribution of the random variable $Y = X^2$.

In this case, $y = \varphi(x) = x^2$. According to (M14.2.1.1), we obtain

$$F_Y(y) = \int_{x^2<y} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{y}}^{\sqrt{y}} e^{-x^2/2}\,dx = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{y}} e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}} \int_0^y \frac{e^{-t/2}}{\sqrt{t}}\,dt.$$

Example 2. Suppose that a random variable X has the probability density

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Bigl[-\frac{(x-a)^2}{2\sigma^2}\Bigr].$$

Find the probability density of the random variable $Y = e^X$.

For $y > 0$, the cumulative distribution function of the random variable $Y = e^X$ is determined by the relations

$$F_Y(y) = P(Y < y) = P(e^X < y) = P(X < \ln y) = F(\ln y).$$

We differentiate this relation and obtain

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF(\ln y)}{dy} = p(\ln y)\frac{1}{y} = \frac{1}{\sqrt{2\pi}\sigma y} \exp\Bigl[-\frac{(\ln y - a)^2}{2\sigma^2}\Bigr] \text{ for } y > 0.$$

The distribution of Y is called the *log-normal distribution*.

Example 3. Suppose that a random variable X has the probability density $p(x)$ for $x \in (-\infty, \infty)$. Then the probability density of the random variable $Y = |X|$ is given by the formula $p_Y(y) = p(x) + p(-x)$ $(y \ge 0)$. In particular, if X is symmetric (i.e., the condition $P(X < -x) = P(X > x)$ holds for all x), then $p_Y(y) = 2p(x)$ $(y \ge 0)$.

M14.2.2. Expectation, Variance and Moments of a Random Variable

► **Expectation.** The *expectation* (*expected value*) $E\{X\}$ *of a discrete or continuous random variable* X is the expression given by the formula

$$E\{X\} = \begin{cases} \displaystyle\sum_i x_i p_i & \text{in the discrete case,} \\ \displaystyle\int_{-\infty}^{+\infty} x p(x)\,dx & \text{in the continuous case.} \end{cases} \qquad \text{(M14.2.2.1)}$$

For the existence of the expectation (M14.2.2.1), it is necessary that the corresponding series or integral converge absolutely.

The expectation is the main characteristic defining the "position" of a random variable, i.e., the number near which its possible values are concentrated.

► **Expectation of function of random variable.** If a random variable Y is related to a random variable X by a functional dependence $Y = \varphi(X)$, then the expectation of the random variable $Y = \varphi(X)$ can be determined by two methods. The first method is to construct the distribution of the random variable Y and then use already known formulas to find $E\{Y\}$. The second method is to use the formulas

$$E\{Y\} = E\{\varphi(X)\} = \begin{cases} \displaystyle\sum_i \varphi(x_i) p_i & \text{in the discrete case,} \\ \displaystyle\int_{-\infty}^{+\infty} \varphi(x) p(x)\,dx & \text{in the continuous case} \end{cases}$$

if these expressions exist in the sense of absolute convergence.

Example. Suppose that a random variable X is uniformly distributed in the interval $(-\pi/2, \pi/2)$, i.e., $p(x) = 1/\pi$ for $x \in (-\pi/2, \pi/2)$ and $p(x) = 0$ for $|x| > \pi/2$. Then the expectation of the random variable $Y = \sin(X)$ is equal to

$$E\{Y\} = \int_{-\infty}^{+\infty} \varphi(x)p(x)\,dx = \int_{-\pi/2}^{\pi/2} \frac{1}{\pi} \sin x\,dx = 0.$$

Properties of the expectation:

1. $E\{a\} = a$ for any real number $a =$ const.
2. $E\{\alpha X + \beta Y\} = \alpha E\{X\} + \beta E\{Y\}$ for any real α and β.
3. $E\{XY\} = E\{X\}E\{Y\}$ for independent random variables X and Y.
4. $|E\{X\}| \le E\{|X|\}$.
5. $E\{X\} \le E\{Y\}$ if $X(\omega) \le Y(\omega)$, $\omega \in \Omega$.
6. $g(E\{X\}) \le E\{g(X)\}$ for convex functions $g(X)$.
7. The *Cauchy–Schwarz inequality* $(E\{|XY|\})^2 \le (E\{X\})^2(E\{Y\})^2$ holds.
8. $E\Big\{\sum\limits_{k=1}^{\infty} X_k\Big\} = \sum\limits_{k=1}^{\infty} E\{X_k\}$ if the series $\sum\limits_{k=1}^{\infty} E\{|X_k|\}$ converges.

► **Moments.** The expectation $E\{(X-a)^k\}$ is called the *kth moment* of the random variable X about the number a. The moments about zero are usually referred to simply as the moments of a random variable. (Sometimes they are called *initial moments*.) The kth moment satisfies the relation

$$\alpha_k = E\{X^k\} = \begin{cases} \displaystyle\sum_i x_i^k p_i & \text{in the discrete case,} \\ \displaystyle\int_{-\infty}^{+\infty} x^k p(x)\,dx & \text{in the continuous case.} \end{cases}$$

If $a = E\{X\}$, then the kth moment of the random variable X about a is called the *kth central moment*. The kth central moment satisfies the relation

$$\mu_k = E\{(X - E\{X\})^k\} = \begin{cases} \displaystyle\sum_i (x_i - E\{X\})^k p_i & \text{in the discrete case,} \\ \displaystyle\int_{-\infty}^{+\infty} (x - E\{X\})^k p(x)\,dx & \text{in the continuous case.} \end{cases}$$

In particular, $\mu_0 = 1$ and $\mu_1 = 0$ for any random variable.

The number $m_k = E\{|X - a|^k\}$ is called the *kth absolute moment* of X about a.

The existence of a kth moment α_k or μ_k implies the existence of the moments α_m and μ_m of all orders $m < k$; if the integral (or series) for α_k or μ_k diverges, then all integrals (series) for α_m and μ_m of orders $m > k$ also diverge.

There is a simple relationship between the central and initial moments:

$$\mu_k = \sum_{m=0}^{k} C_k^m \alpha_m (\alpha_1)^{k-m}, \quad \alpha_0 = 1; \quad \alpha_k = \sum_{m=0}^{k} C_k^m \mu_m (\alpha_1)^{k-m}.$$

The probability distribution is uniquely determined by the moments $\alpha_0, \alpha_1, \ldots$ provided that they all exist and the series $\sum\limits_{m=0}^{\infty} |\alpha_m| t^m / m!$ converges for some $t > 0$.

► **Variance.** The *variance* of a random variable, $\operatorname{Var}\{X\}$, is the measure of the deviation of a random variable X from its expectation $E\{X\}$, determined by the relation

$$\operatorname{Var}\{X\} = E\{(X - E\{X\})^2\}.$$

The variance $\operatorname{Var}\{X\}$ is the second central moment of the random variable X. The variance can be determined by the formulas

$$\operatorname{Var}\{X\} = \begin{cases} \displaystyle\sum_i (x_i - E\{X\})^2 p_i & \text{in the discrete case,} \\ \displaystyle\int_{-\infty}^{+\infty} (x - E\{X\})^2 p(x)\,dx & \text{in the continuous case.} \end{cases}$$

The variance characterizes the spread in values of the random variable X about its expectation.

Properties of the variance:

1. $\operatorname{Var}\{a\} = 0$ for any real number $a = \text{const}$.
2. The variance is nonnegative: $\operatorname{Var}\{X\} \geq 0$.
3. $\operatorname{Var}\{X\} = E\{X^2\} - (E\{X\})^2$.
4. $\operatorname{Var}\{\alpha X + \beta\} = \alpha^2 \operatorname{Var}\{X\}$ for any real numbers α and β.
5. $\operatorname{Var}\{X \pm Y\} = \operatorname{Var}\{X\} + \operatorname{Var}\{Y\}$ for independent random variables X and Y.
6. If X and Y are independent random variables, then

$$\operatorname{Var}\{XY\} = \operatorname{Var}\{X\}\operatorname{Var}\{Y\} + \operatorname{Var}\{X\}(E\{Y\})^2 + \operatorname{Var}\{Y\}(E\{X\})^2.$$

7. $\min\limits_m E\{(X - m)^2\} = \operatorname{Var}\{X\}$ and is attained for $m = E\{X\}$.

► **Other numerical characteristics of random variables.** The *standard deviation (root mean square deviation)* of a random variable X is the square root of its variance,

$$\sigma = \sqrt{\mathrm{Var}\{X\}}.$$

The standard deviation has the same dimension as the random variable itself.

The *coefficient of variation* is the ratio of the standard deviation to the expected value,

$$v = \frac{\sigma}{E\{X\}}.$$

The *asymmetry coefficient*, or *skewness*, is defined by the formula

$$\gamma_1 = \frac{\mu_3}{(\mu_2)^{3/2}}.$$

If $\gamma_1 > 0$, then the distribution curve is more flattened to the right of the mode $\mathrm{Mode}\{X\}$; if $\gamma_1 < 0$, then the distribution curve is more flattened to the left of the mode $\mathrm{Mode}\{X\}$ (see Fig. M14.1).

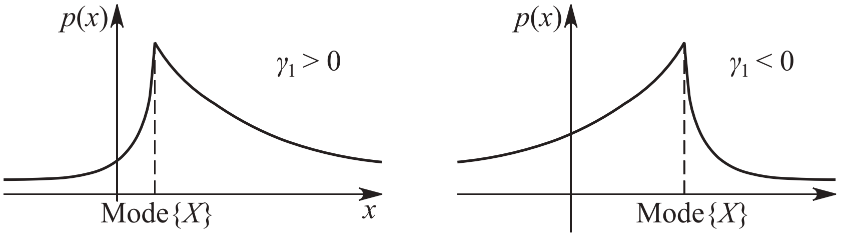

Figure M14.1. Relationship between the distribution curve and the asymmetry coefficient.

The *excess coefficient*, or *excess*, or *kurtosis*, is defined by the formula

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3.$$

One says that for $\gamma_2 = 0$ the distribution has a normal excess, for $\gamma_2 > 0$ the distribution has a positive excess, and for $\gamma_2 < 0$ the distribution has a negative excess.

Remark. The coefficients γ_1^2 and $\gamma_2 + 3$ or $(\gamma_2 + 3)/2$ are often used instead of γ_1 and γ_2.

A *mode* $\mathrm{Mode}\{X\}$ *of a continuous probability distribution* is a point of local maximum of the probability density function $p(x)$. A *mode of a discrete probability distribution* is a value preceded and followed by values having probabilities smaller than $p(\mathrm{Mode}\{X\})$.

Distributions with one, two, or more modes are said to be *unimodal*, *bimodal*, or *multimodal*, respectively.

M14.2.3. Main Discrete Distributions and Their Numerical Characteristics

► **Binomial distribution.** A random variable X has a *binomial distribution* with parameters n and p if

$$P_n(k) = P(X = k) = C_n^k p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n,$$

where $0 < p < 1$ and $n \ge 1$.

The cumulative distribution function has the form

$$F(x) = \begin{cases} 1 & \text{for } x > n, \\ \sum\limits_{k=1}^{m} C_n^k p^k (1-p)^{n-k} & \text{for } m \le x < m+1 \ \ (m = 1, 2, \ldots, n-1), \\ 0 & \text{for } x < 0, \end{cases}$$

and the numerical characteristics are given by the formulas

$$E\{X\} = np, \quad \operatorname{Var}\{X\} = np(1-p), \quad \gamma_1 = \frac{1-2p}{\sqrt{np(1-p)}}, \quad \gamma_2 = \frac{1-6p(1-p)}{np(1-p)}.$$

The binomial distribution is a model of random experiments consisting of n independent identical Bernoulli trials. If $X_1, \ldots, X_n$ are independent random variables, each of which can take only two values 1 or 0 with probabilities p and $q = 1 - p$, respectively, then the random variable $X = \sum\limits_{k=1}^{n} X_k$ has the binomial distribution with parameters (n, p).

See also the limit formulas (M14.1.4.2)–(M14.1.4.4) for the Bernoulli process.

► **Geometric distribution.** A random variable X has a *geometric distribution* with parameter p $(0 < p < 1)$ if

$$P(X = k) = p(1-p)^k, \quad k = 0, 1, 2, \ldots$$

Numerical characteristics for the geometric distribution can be calculated by the formulas

$$E\{X\} = \frac{1-p}{p}, \quad \operatorname{Var}\{X\} = \frac{1-p}{p^2}, \quad \alpha_2 = \frac{(1-p)(2-p)}{p^2}, \quad \gamma_1 = \frac{2-p}{\sqrt{1-p}}, \quad \gamma_2 = 6 + \frac{p^2}{1-p}.$$

The geometric distribution has a maximum at $k = 0$ and decreases monotonically with increasing k.

The geometric distribution describes a random variable X equal to the number of failures before the first success in a sequence of Bernoulli trials with probability p of success in each trial.

The geometric distribution is the only discrete distribution that is *memoryless*, i.e., satisfies the relation

$$P(X > t + s \mid X > t) = P(X > s)$$

for any $s > 0$ and $t > 0$. This property permits one to view the geometric distribution as the discrete analogue of the exponential distribution (see Subsection M14.2.4).

► **Hypergeometric distribution.** A random variable X has a *hypergeometric distribution* with parameters N, p, and n if

$$P(X = k) = \frac{C_{Np}^{k} C_{N(1-p)}^{n-k}}{C_N^n}, \quad k = 0, 1, \ldots, n,$$

where $0 < p < 1$, $0 \le n \le N$, and $N > 0$.

The numerical characteristics are given by the formulas

$$E\{X\} = np, \quad \operatorname{Var}\{X\} = \frac{N-n}{N-1} np(1-p).$$

A typical scheme in which the hypergeometric distribution arises is as follows: n elements are randomly drawn without replacement from a population of N elements containing exactly Np elements of type I and $N(1-p)$ elements of type II. The number of elements of type I in the sample is described by the hypergeometric distribution.

If $n \ll N$ (in practice, $n < 0.1N$), then

$$\frac{C_{Np}^{k} C_{N(1-p)}^{n-k}}{C_{N}^{n}} \approx C_{n}^{k} p^{k} (1-p)^{n-k};$$

i.e., the hypergeometric distribution tends to the binomial distribution.

► **Poisson distribution.** A random variable X has a *Poisson distribution* with parameter λ ($\lambda > 0$) if

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

The cumulative distribution function of the Poisson distribution at the points $k = 0, 1, 2, \dots$ is given by the formula

$$F(k) = \frac{1}{k!} \int_{\lambda}^{\infty} x^k e^{-x}\, dx,$$

and the numerical characteristics are given by

$$\begin{gathered} E\{X\} = \lambda, \quad \operatorname{Var}\{X\} = \lambda, \quad \alpha_2 = \lambda^2 + \lambda, \quad \alpha_3 = \lambda(\lambda^2 + 3\lambda + 1), \\ \alpha_4 = \lambda(\lambda^3 + 6\lambda^2 + 7\lambda + 1), \quad \mu_3 = \lambda, \quad \mu_4 = 3\lambda^2 + \lambda, \quad \gamma_1 = \lambda^{-1/2}, \quad \gamma_2 = \lambda^{-1}. \end{gathered}$$

The sum of independent random variables $X_1, \dots, X_n$ obeying the Poisson distributions with parameters $\lambda_1, \dots, \lambda_n$, respectively, has the Poisson distribution with parameter $\lambda_1 + \dots + \lambda_n$.

The Poisson distribution is the limit distribution for many discrete distributions such as the hypergeometric distribution, the binomial distribution, the negative binomial distribution, distributions arising in problems of arrangement of particles in cells, etc. The Poisson distribution is an acceptable model for describing the random number of occurrences of certain events on a given time interval in a given domain in space.

► **Negative binomial distribution.** A random variable X has a *negative binomial distribution* with parameters r and p if

$$P(X = k) = C_{r+k-1}^{r-1} p^r (1-p)^k, \quad k = 0, 1, \dots, r,$$

where $0 < p < 1$ and $r > 0$.

The numerical characteristics can be calculated by the formulas

$$E\{X\} = \frac{r(1-p)}{p}, \quad \operatorname{Var}\{X\} = \frac{r(1-p)}{p^2}, \quad \gamma_1 = \frac{2-p}{\sqrt{r(1-p)}}, \quad \gamma_2 = \frac{6}{r} + \frac{p^2}{r(1-p)}.$$

The negative binomial distribution describes the number X of failures before the rth success in a Bernoulli process with probability p of success on each trial. For $r = 1$, the negative binomial distribution coincides with the geometric distribution.

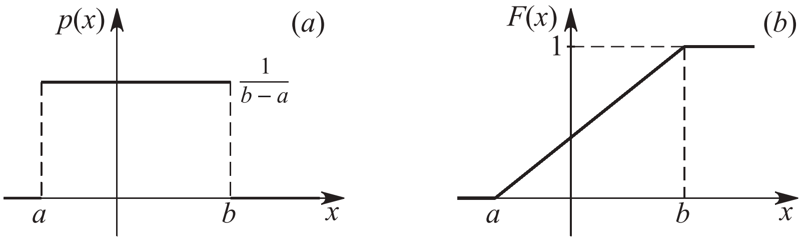

Figure M14.2. Probability density function (*a*) and cumulative distribution function (*b*) of a uniform distribution.

M14.2.4. Main Continuous Distributions and Their Numerical Characteristics

► **Uniform distribution.** A random variable X is *uniformly* distributed on an interval $[a, b]$ (Fig. M14.2 *a*) if

$$p(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } x \in [a, b], \\ 0 & \text{for } x \notin [a, b]. \end{cases}$$

The cumulative distribution function (see Fig. M14.2 *b*) has the form

$$F(x) = \begin{cases} 0 & \text{for } x \le a, \\ \dfrac{x-a}{b-a} & \text{for } a < x \le b, \\ 1 & \text{for } x > b, \end{cases}$$

and the numerical characteristics are given by the expressions

$$E\{X\} = \frac{a+b}{2}, \quad \operatorname{Var}\{X\} = \frac{(b-a)^2}{12}, \quad \gamma_1 = 0, \quad \gamma_2 = -1.2.$$

The uniform distribution does not have a mode.

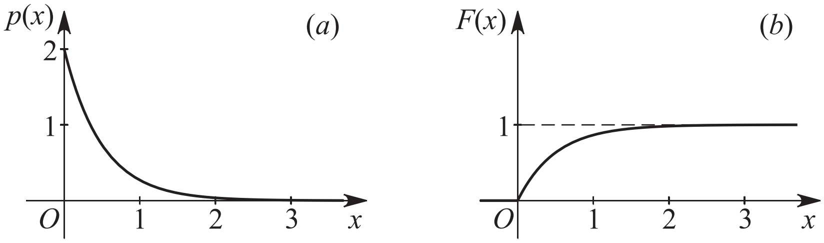

Figure M14.3. Probability density function (*a*) and cumulative distribution function (*b*) of an exponential distribution for $\lambda = 2$.

► **Exponential distribution.** A random variable X has the *exponential distribution* with parameter $\lambda > 0$ (Fig. M14.3 *a*) if

$$p(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x > 0, \\ 0 & \text{for } x \le 0. \end{cases}$$

The cumulative distribution function (see Fig. M14.3 *b*) has the form

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{for } x > 0, \\ 0 & \text{for } x \le 0, \end{cases}$$

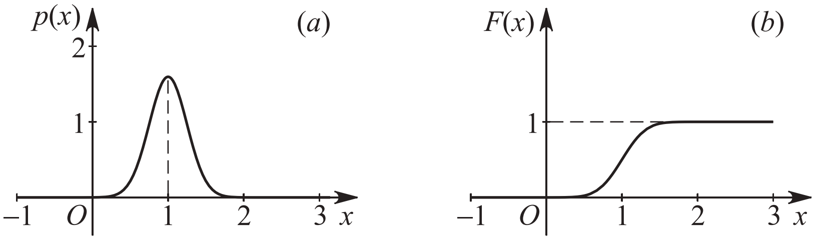

Figure M14.4. Probability density function (*a*) and cumulative distribution function (*b*) of a normal distribution for $a = 1$ and $\sigma = 1/4$.

and the numerical characteristics are given by the formulas

$$E\{X\} = \frac{1}{\lambda}, \quad \operatorname{Var}\{X\} = \frac{1}{\lambda^2}, \quad \alpha_2 = \frac{2}{\lambda^2}, \quad \gamma_1 = 2, \quad \gamma_2 = 6.$$

The exponential distribution is the continuous analogue of the geometric distribution and is memoryless: $P(X > t + s \mid X > t) = P(X > s)$.

► **Normal distribution.** A random variable X has a *normal distribution* with parameters a and σ^2 (see Fig. M14.4 *a*) if its probability density function has the form

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right], \quad x \in (-\infty, \infty).$$

The cumulative distribution function (see Fig. M14.4 *b*) has the form

$$F(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} \exp\left[-\frac{(t-a)^2}{2\sigma^2}\right] dt \equiv \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-a}{\sqrt{2}\,\sigma}\right)\right],$$

where erf z is the error function (see Section M13.2).

The numerical characteristics are given by the formulas

$$E\{X\} = a, \quad \operatorname{Var}\{X\} = \sigma^2, \quad \operatorname{Mode}\{X\} = a, \quad \gamma_1 = 0, \quad \gamma_2 = 0,$$
$$\mu_k = \begin{cases} 0, & k = 2l-1, \ l = 1, 2, \ldots \\ (2k-1)!!\,\sigma^{2k}, & k = 2l, \ l = 1, 2, \ldots \end{cases}$$

The linear transformation $Y = \frac{X-a}{\sigma}$ reduces the normal distribution with parameters (a, σ^2) and cumulative distribution function $F(x)$ to the standard normal distribution with parameters $(0, 1)$ and cumulative distribution function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt. \qquad \text{(M14.2.4.1)}$$

The probability that a random variable X normally distributed with parameters (m, σ^2) lies in the interval (a, b) is given by the formula

$$P(a < x < b) = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right).$$

A normally distributed random variable takes values close to its expectation with large probability; this is expressed by the *sigma rule*

$$P(|X-m|\ge k\sigma)=2[1-\Phi(k)]=\begin{cases}0.3173 & \text{for } k=1,\\ 0.0456 & \text{for } k=2,\\ 0.0027 & \text{for } k=3.\end{cases}$$

The three-sigma rule is most frequently used.

The fundamental role of the normal distribution is due to the fact that, under mild assumptions, the distribution of a sum of random variables is asymptotically normal as the number of terms increases. The corresponding conditions are specified in the central limit theorem.

► **Chi-square distribution.** A random variable $X=\chi^2(n)$ has a *chi-square distribution* with n degrees of freedom if its probability density function has the form

$$p(x)=\begin{cases}\dfrac{1}{2^{n/2}\,\Gamma(n/2)}x^{-1+n/2}e^{-x/2} & \text{for } x>0,\\ 0 & \text{for } x\le 0,\end{cases}\qquad\text{(M14.2.4.2)}$$

where $\Gamma(z)$ is the Gamma function (see Subsection M13.4.1).

The cumulative distribution function can be written as

$$F(x)=\frac{1}{2^{n/2}\Gamma(n/2)}\int_0^x \xi^{-1+n/2}e^{-\xi/2}\,d\xi$$

and the numerical characteristics are given by the formulas

$$E\{\chi^2(n)\}=n,\quad \operatorname{Var}\{\chi^2(n)\}=2n,\quad \alpha_k=n(n+2)\cdot\ldots\cdot[n+2(k-1)],$$
$$\gamma_1=2\sqrt{\frac{2}{n}},\quad \gamma_2=\frac{12}{n},\quad \operatorname{Mode}\{\chi^2(n)\}=n-2\quad (n\ge 2).$$

Main property of the chi-square distribution. For an arbitrary n, the sum

$$X=\sum_{k=1}^{n}X_k^2,$$

of squares of independent random variables obeying the standard normal distribution has the chi-square distribution with n degrees of freedom.

Relationship with other distributions:

1. For $n=1$, formula (M14.2.4.2) gives the probability density function of the square X^2 of a random variable with the standard normal distribution.
2. For $n=2$, formula (M14.2.4.2) gives the exponential distribution with parameter $\lambda=\frac{1}{2}$.
3. As $n\to\infty$, the random variable $X=\chi^2(n)$ has an asymptotically normal distribution with parameters $(n,2n)$.
4. As $n\to\infty$, the random variable $\sqrt{2\chi^2(n)}$ has an asymptotically normal distribution with parameters $(\sqrt{2n-1},1)$.

M14.2.5. Two-dimensional and Multivariate Random Variables

► **Distribution of bivariate random variable.** Suppose that random variables X_1 and X_2 are defined in a probability space Ω_2; then one says that a *two-dimensional random vector* $\mathbf{X} = (X_1, X_2)$ or a *system of two random variables* is given.

The *distribution function* $F(x_1, x_2)$ (also denoted $F_{X_1,X_2}(x_1, x_2)$) *of a two-dimensional random vector* (X_1, X_2), or the *joint distribution function of the random variables* X_1 and X_2, is defined as the probability of the simultaneous occurrence (intersection) of the events $(X_1 < x_1)$ and $(X_2 < x_2)$; i.e.,

$$F(x_1, x_2) = P(X_1 < x_1, X_2 < x_2).$$

Properties of the joint distribution function of random variables X_1 and X_2:

1. The function $F(x_1, x_2)$ is a nondecreasing function of each of the arguments.
2. $F(x_1, -\infty) = F(-\infty, x_2) = F(-\infty, -\infty) = 0$.
3. $F(+\infty, +\infty) = 1$.
4. The probability that the random vector lies in a rectangle with sides parallel to the coordinate axes is

$$P(a_1 \le X_1 < b_1, a_2 \le X_2 < b_2) = F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2).$$

5. The function $F(x_1, x_2)$ is left-continuous in either argument.

► **Discrete bivariate random variables.** A bivariate random variable (X_1, X_2) is said to be *discrete* if each of the random variables X_1 and X_2 is discrete.

If the random variable X_1 takes the values $x_{11}, \ldots, x_{1m}$ and the random variable X_2 takes the values $x_{21}, \ldots, x_{2n}$, then the random vector (X_1, X_2) can take only the pairs of values (x_{1i}, x_{2j}) $(i = 1, \ldots, m,\ j = 1, \ldots, n)$. The entries $p_{ij} = P(X_1 = x_{1i}, X_2 = x_{2j})$ are the probabilities of the simultaneous occurrence of the events $(X_1 = x_{1i})$ and $(X_2 = x_{2j})$; $P_{X_1,i} = p_{i1} + \cdots + p_{in}$ is the probability that the random variable X_1 takes the value x_{1i}; $P_{X_2,j} = p_{1j} + \cdots + p_{mj}$ is the probability that the random variable X_2 takes the value x_{2j}.

The distribution function of a discrete bivariate random variable can be determined by the formula

$$F(x_1, x_2) = \sum_{\substack{x_{1i} < x_1 \\ x_{2j} < x_2}} p_{ij}.$$

► **Continuous bivariate random variables.** A bivariate random variable (X_1, X_2) is said to be *continuous* if its joint distribution function $F(x_1, x_2)$ can be represented as

$$F(x_1, x_2) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} p(y_1, y_2)\, dy_1\, dy_2, \qquad \text{(M14.2.5.1)}$$

where the *joint probability density function* $p(x_1, x_2)$ is piecewise continuous.

The joint probability density function can be expressed in terms of the joint distribution function as follows:

$$p(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F(x_1, x_2). \qquad \text{(M14.2.5.2)}$$

Formulas (M14.2.5.1) and (M14.2.5.2) establish a one-to-one correspondence (up to sets of probability zero) between the joint probability density functions and the joint distribution functions of continuous bivariate random variables. The differential $p(x_1, x_2)\, dx_1\, dx_2$ is called a *probability element*. Up to higher-order infinitesimals, the probability element is equal to the probability for the random variable (X_1, X_2) to lie in the infinitesimal rectangle $(x_1, x_1 + \Delta x_1) \times (x_2, x_2 + \Delta x_2)$.

Properties of the joint probability density function of random variables X_1 and X_2:

1. The function $p(x_1, x_2)$ is nonnegative; i.e., $p(x_1, x_2) \ge 0$.
2. $\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x_1, x_2)\,dx_1\,dx_2 = 1$.
3. $\displaystyle P(a_1 < X_1 < b_1, a_2 < X_2 < b_2) = \int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} p(x_1, x_2)\,dx_2 = \int_{a_2}^{b_2} dx_2 \int_{a_1}^{b_1} p(x_1, x_2)\,dx_1$.
4. The probability for a two-dimensional random variable (X_1, X_2) to lie in a domain $D \subset \mathbb{R}^2$ is numerically equal to the volume of the curvilinear cylinder with base D bounded above by the surface of the joint probability density function:

$$P[(X_1, X_2) \in D] = \iint_{(x_1,x_2)\in D} p_{X_1,X_2}(x_1, x_2)\,dx_1\,dx_2.$$

Random variables X_1 and X_2 are said to be *independent* if the relation

$$P(X_1 \in S_1, X_2 \in S_2) = P(X_1 \in S_1)\,P(X_2 \in S_2)$$

holds for any measurable sets S_1 and S_2.

THEOREM. *Random variables X_1 and X_2 are independent if and only if*

$$F(x_1, x_2) = F_1(x_1)\,F_2(x_2),$$

where F_1 and F_2 are the cumulative distribution functions of X_1 and X_2, respectively.

► **Numerical characteristics of bivariate random variables.** The *expectation* of a function $\varphi(X_1, X_2)$ of a bivariate random variable (X_1, X_2) is defined as the expression computed by the formula

$$E\{\varphi(X_1, X_2)\} = \begin{cases} \displaystyle\sum_i \sum_j \varphi(x_{1i}, x_{2j}) p_{ij} & \text{in the discrete case,} \\ \displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \varphi(x_1, x_2) p(x_1, x_2)\,dx_1\,dx_2 & \text{in the continuous case,} \end{cases}$$

if these expressions exist in the sense of absolute convergence; otherwise, one says that $E\{\varphi(X_1, X_2)\}$ does not exist.

The *moment of order* $r_1 + r_2$ of a two-dimensional random variable (X_1, X_2) about a point (a_1, a_2) is defined as the expectation $E\{(X_1 - a_1)^{r_1}(X_2 - a_2)^{r_2}\}$.

If $a_1 = a_2 = 0$, then the moment of order $r_1 + r_2$ of a two-dimensional random variable (X_1, X_2) is called simply the moment, or the *initial moment*. The initial moment of order $r_1 + r_2$ is usually denoted by α_{r_1,r_2}; i.e., $\alpha_{r_1,r_2} = E\{X_1^{r_1} X_2^{r_2}\}$.

The first initial moments are the expectations of the random variables X_1 and X_2; i.e., $\alpha_{1,0} = E\{X_1^1 X_2^0\} = E\{X_1\}$ and $\alpha_{0,1} = E\{X_1^0 X_2^1\} = E\{X_2\}$. The point $(E\{X_1\}, E\{X_2\})$ on the OXY-plane characterizes the position of the random point (X_1, X_2); this position spreads about the point $(E\{X_1\}, E\{X_2\})$.

If $a_1 = E\{X_1\}$ and $a_2 = E\{X_2\}$, then the moment of order $r_1 + r_2$ of the bivariate random variable (X_1, X_2) is called the *central moment*. The central moment of order $r_1 + r_2$

is usually denoted by μ_{r_1,r_2}; i.e., $\mu_{r_1,r_2} = E\{(X_1 - E\{X_1\})^{r_1}(X_2 - E\{X_2\})^{r_2}\}$. Obviously, the first central moments are zero.

The second central moments are of special interest and have special names and notation:

$$\lambda_{11} = \mu_{2,0} = \operatorname{Var}\{X_1\}, \quad \lambda_{22} = \mu_{0,2} = \operatorname{Var}\{X_2\},$$

$$\lambda_{12} = \lambda_{21} = \mu_{1,1} = E\{(X_1 - E\{X_1\})(X_2 - E\{X_2\})\}.$$

The first two of these moments are the variances of the respective random variables, and the third one is called the *covariance* and will be considered below.

► **Covariance and correlation of two random variables.** The *covariance* (correlation moment, or mixed second moment) $\operatorname{Cov}(X_1, X_2)$ of random variables X_1 and X_2 is defined as the central moment of order (1 + 1):

$$\operatorname{Cov}(X_1, X_2) = \alpha_{1,1} = E\{(X_1 - E\{X_1\})(X_2 - E\{X_2\})\}.$$

Properties of the covariance:

1. $\operatorname{Cov}(X_1, X_2) = \operatorname{Cov}(X_2, X_1)$.
2. $\operatorname{Cov}(X, X) = \operatorname{Var}\{X\}$.
3. If the random variables X_1 and X_2 are independent, then $\operatorname{Cov}(X_1, X_2) = 0$ and if $\operatorname{Cov}(X_1, X_2) \neq 0$, then the random variables X_1 and X_2 are dependent.
4. If $Y_1 = a_1X_1 + b_1$ and $Y_2 = a_2X_2 + b_2$, then $\operatorname{Cov}(Y_1, Y_2) = a_1a_2\operatorname{Cov}(X_1, X_2)$.
5. $\operatorname{Cov}(X_1, X_2) = E\{X_1X_2\} - E\{X_1\}E\{X_2\}$.
6. $|\operatorname{Cov}(X_1, X_2)| \le \sqrt{\operatorname{Var}\{X_1\}\operatorname{Var}\{X_2\}}$. Moreover, $\operatorname{Cov}(X_1, X_2) = \pm\sqrt{\operatorname{Var}\{X_1\}\operatorname{Var}\{X_2\}}$ if and only if the random variables X_1 and X_2 are linearly dependent.
7. $\operatorname{Var}\{X_1 + X_2\} = \operatorname{Var}\{X_1\} + \operatorname{Var}\{X_2\} + 2\operatorname{Cov}(X_1, X_2)$.

If $\operatorname{Cov}(X_1, X_2) = 0$, then the random variables X_1 and X_2 are said to be *uncorrelated*; if $\operatorname{Cov}(X_1, X_2) \neq 0$, then they are *correlated*. Independent random variables are always uncorrelated, but uncorrelated random variables are not necessarily independent in general.

Example. Suppose that we roll two dice. Let X_1 be the number on the first die and let X_2 be the number on the second die. We consider the random variables $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$ (the sum and difference of the points obtained). Then

$$\begin{aligned}\operatorname{Cov}(Y_1, Y_2) &= E\{(X_1 + X_2 - E\{X_1 + X_2\})(X_1 - X_2 - E\{X_1 - X_2\})\}\\ &= E\{(X_1 - E\{X_1\})^2 - (X_2 - E\{X_2\})^2\}\\ &= \operatorname{Var}\{X_1\} - \operatorname{Var}\{X_2\} = 0,\end{aligned}$$

since X_1 and X_2 are identically distributed and hence $\operatorname{Var}\{X_1\} = \operatorname{Var}\{X_2\}$. But Y_1 and Y_2 are obviously dependent; for example, if $Y_1 = 2$ then one necessarily has $Y_2 = 0$.

The covariance of random variables X_1 and X_2 characterizes both their dependence on each other and their spread around the point $(E\{X_1\}, E\{X_2\})$. The covariance of X_1 and X_2 has the dimension equal to the product of dimensions of X_1 and X_2. Along with the covariance of X_1 and X_2, one often uses the coefficient of correlation $\rho(X_1, X_2)$, which is a dimensionless normalized quantity. The *correlation coefficient* of random variables X_1 and X_2 is the ratio of the covariance of X_1 and X_2 to the product of their standard deviations,

$$\rho(X_1, X_2) = \frac{\operatorname{Cov}(X_1, X_2)}{\sigma_1\sigma_2}.$$

The correlation coefficient of random variables X_1 and X_2 indicates the degree of linear dependence between the variables. If $\rho(X_1, X_2) = 0$, then there is no linear relation between the random variables, but there may well be some other relation between them.

Properties of the correlation coefficient:

1. $\rho(X_1, X_2) = \rho(X_2, X_1)$.
2. $\rho(X, X) = 1$.
3. If random variables X_1 and X_2 are independent, then $\rho(X_1, X_2) = 0$. If $\rho(X_1, X_2) \neq 0$, then the random variables X_1 and X_2 are dependent.
4. If $Y_1 = a_1 X_1 + b_1$ and $Y_2 = a_2 X_2 + b_2$, then $\rho(Y_1, Y_2) = \pm\rho(X_1, X_2)$.
5. $|\rho(X_1, X_2)| \le 1$. Moreover, $\rho(X_1, X_2) = \pm 1$ if and only if the random variables X_1 and X_2 are linearly dependent.

The theory of distribution functions of multivariate random variables (random vectors) can be developed in a similar way.

M14.3. Limit Theorems

M14.3.1. Convergence of Random Variables

► **Convergence in probability.** A sequence of random variables X_1, X_2, … is said to *converge in probability* to a random variable $X (X_n \xrightarrow{P} X)$ if

$$\lim_{n\to\infty} P(|X_n - X| \ge \varepsilon) = 0$$

for any $\varepsilon > 0$, i.e., if for any $\varepsilon > 0$ and $\delta > 0$ there exists a number N, depending on ε and δ, such that the inequality

$$P(|X_n - X| \ge \varepsilon) < \delta$$

holds for $n > N$. A sequence of k-dimensional random variables $\mathbf{X}_n$ is said to converge in probability to a random variable $\mathbf{X}$ if each coordinate of the random variable $\mathbf{X}_n$ converges in probability to the respective coordinate of the random variable $\mathbf{X}$.

► **Convergence in the mean.** A sequence of random variables X_1, X_2, … with finite qth initial moments ($E\{X^q\} < \infty$, $q = 1, 2, \ldots$) is said to *converge in the qth mean* to a random variable X if

$$\lim_{n\to\infty} E\{|X_n - X|^q\} = 0.$$

Convergence in the qth mean for $q = 2$ is called *convergence in mean square*. If $X_n \to X$ in the qth mean then $X_n \to X$ in the q_1th mean for all $q_1 \le q$.

Convergence in the qth mean implies convergence in probability. The converse statement is generally not true.

► **Convergence in distribution.** Suppose that a sequence $F_1(x)$, $F_2(x)$, … of cumulative distribution functions converges to a distribution function $F(x)$,

$$\lim_{n\to\infty} F_n(x) = F(x),$$

for every point x. In this case, we say that the sequence X_1, X_2, … of the corresponding random variables converges to the random variable X *in distribution*. The random variables X_1, X_2, … can be defined on different probability spaces.

A sequence $F_1(x)$, $F_2(x)$, … of distribution functions *weakly converges* to a distribution function $F(x)$ ($F_n \to F$) if

$$\lim_{n\to\infty} E\{h(X_n)\} = E\{h(X)\}$$

for any bounded continuous function h.

Convergence in distribution and weak convergence of distribution functions are equivalent.

The weak convergence $F_{X_n} \to F$ for random variables having a probability density function means the convergence

$$\int_{-\infty}^{+\infty} g(x)p_{X_n}(x)\,dx \to \int_{-\infty}^{+\infty} g(x)p(x)\,dx$$

for any bounded continuous function $g(x)$, where $p(x) = \lim\limits_{n\to\infty} p_{X_n}(x)$.

M14.3.2. Limit Theorems

► **Law of large numbers.** The law of large numbers consists of several theorems establishing the stability of average results and revealing conditions for this stability to occur.

The notion of convergence in probability is most often used for the case in which the limit random variable X has the degenerate distribution concentrated at a point a ($P(X = a) = 1$) and

$$X_n = \frac{1}{n}\sum_{k=1}^{n} Y_k,$$

where $Y_1, Y_2, \dots$ are arbitrary random variables.

A sequence $Y_1, Y_2, \dots$ satisfies the *weak law of large numbers* if the limit relation

$$\lim_{n\to\infty} P\Bigl(\Bigl|\frac{1}{n}\sum_{k=1}^{n} Y_k - a\Bigr| \ge \varepsilon\Bigr) \equiv \lim_{n\to\infty} P(|X_n - a| \ge \varepsilon) = 0 \qquad \text{(M14.3.2.1)}$$

holds for any $\varepsilon > 0$.

If the relation

$$P\Bigl(\omega \in \Omega : \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} Y_k = a\Bigr) \equiv P\bigl(\omega \in \Omega : \lim_{n\to\infty} X_n = a\bigr) = 1$$

is satisfied instead of (M14.3.2.1), i.e., the sequence X_n converges to the number a with probability 1, then the sequence $Y_1, Y_2, \dots$ satisfies the *strong law of large numbers*.

Markov's inequality. For any nonnegative random variable X that has an expectation $E\{X\}$, the inequality

$$P(X \ge \varepsilon) \le \frac{E\{X\}}{\varepsilon} \qquad \text{(M14.3.2.2)}$$

holds for any $\varepsilon > 0$. It follows from inequality (M14.3.2.2) that

$$P(X < \varepsilon) \ge 1 - \frac{E\{X\}}{\varepsilon}.$$

Chebyshev's inequality. For any random variable X with finite variance, the inequality

$$P(|X - E\{X\}| \ge \varepsilon) \le \frac{\operatorname{Var}\{X\}}{\varepsilon^2} \qquad \text{(M14.3.2.3)}$$

holds for any $\varepsilon > 0$. It follows from inequality (M14.3.2.3) that

$$P(|X - E\{X\}| < \varepsilon) \ge 1 - \frac{\operatorname{Var}\{X\}}{\varepsilon^2}$$

CHEBYSHEV'S THEOREM. *If* $X_1, X_2, \dots$ *is a sequence of pairwise independent random variables with uniformly bounded finite variances,* $\operatorname{Var}\{X_1\} \le C$, $\operatorname{Var}\{X_2\} \le C$, $\dots$, *then the limit relation*

$$\lim_{n\to\infty} P\biggl(\biggl|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E\{X_k\}\biggr| < \varepsilon\biggr) = 1$$

holds for any $\varepsilon > 0$.

BERNOULLI'S THEOREM. *Let* μ_n *be the number of occurrences of an event* A *(the number of successes) in* n *independent trials, and let* $p = P(A)$ *be the probability of the occurrence of the event* A *(the probability of success) in each of the trials. Then the sequence of relative frequencies* μ_n/n *of the occurrence of the event* A *in* n *independent trials converges in probability to* $p = P(A)$ *as* $n \to \infty$*; i.e., the limit relation*

$$\lim_{n\to\infty} P\biggl(\biggl|\frac{\mu_n}{n} - p\biggr| < \varepsilon\biggr) = 1$$

holds for any $\varepsilon > 0$.

POISSON'S THEOREM. *If in a sequence of independent trials the probability that an event* A *occurs in the* k*th trial is equal to* p_k*, then*

$$\lim_{n\to\infty} P\biggl(\biggl|\frac{\mu_n}{n} - \frac{p_1 + \cdots + p_n}{n}\biggr| < \varepsilon\biggr) = 1.$$

KOLMOGOROV'S THEOREM. *If a sequence of independent random variables* $X_1, X_2, \dots$ *satisfies the condition*

$$\sum_{k=1}^{\infty} \frac{\operatorname{Var}\{X_k\}}{k^2} < +\infty,$$

then it obeys the strong law of large numbers.

The existence of the expectation is a necessary and sufficient condition for the strong law of large numbers to apply to a sequence of independent identically distributed random variables.

► **Central limit theorems.** A random variable X_n with distribution function F_{X_n} is *asymptotically normally distributed* if there exists a sequence of pairs of real numbers m_n, σ_n^2 such that the random variables $(X_n - m_n)/\sigma_n$ converge in probability to a standard normal variable. This occurs if and only if the limit relation

$$\lim_{n\to\infty} P(m_n + a\sigma_n < X_n < m_n + b\sigma_n) = \Phi(b) - \Phi(a),$$

where $\Phi(x)$ is the distribution function of the standard normal law (M14.2.4.1), holds for any a and b $(b > a)$.

LYAPUNOV'S CENTRAL LIMIT THEOREM. *If* $X_1, \dots, X_n, \dots$ *is a sequence of independent random variables having finite expectation values* $E\{X_k\}$ *and finite variances* $\operatorname{Var}\{X_k\}$ *and satisfying Lyapunov's condition*

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} \alpha_3(X_k)}{\sqrt{\sum_{k=1}^{n} \operatorname{Var}\{X_k\}}} = 0,$$

where $\alpha_3(X_k)$ *is the third initial moment of the random variable* X_k*, then the sequence of random variables*

$$Y_n = \frac{\sum_{k=1}^{n} (X_k - E\{X_k\})}{\sqrt{\sum_{k=1}^{n} \operatorname{Var}\{X_k\}}}$$

converges in distribution to the normal law, i.e., the following limit exists:

$$\lim_{n\to\infty} P\left(\frac{\sum_{k=1}^{n}(X_k - E\{X_k\})}{\sqrt{\sum_{k=1}^{n}\operatorname{Var}\{X_k\}}} < t\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-u^2/2}\,du = \Phi(t).$$

LINDEBERG'S CENTRAL LIMIT THEOREM. *Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with finite expectation $E\{X_k\} = m$ and finite variance σ^2. Then, as $n \to \infty$, the random variable $\frac{1}{n}\sum\limits_{k=1}^{n} X_k$ has an asymptotically normal probability distribution with parameters $(m, \sigma^2/n)$.*

Let μ_n be the number of occurrences of an event A (the number of successes) in n independent trials, and let $p = P(A)$ denote the probability of the occurrence of the event A (the probability of success) in each of the trials. Then the sequence of relative frequencies μ_n/n has an asymptotically normal probability distribution with parameters $(p,\ p(1-p)/n)$.

Bibliography for Chapter M14

Bain, L. J. and Engelhardt, M., *Introduction to Probability and Mathematical Statistics, 2nd Edition* (Duxbury Classic), Duxbury Press, Boston, Massachusetts, 2000.

Bertsekas, D. P. and Tsitsiklis, J. N., *Introduction to Probability*, Athena Scientific, Belmont, Massachusetts, 2002.

Burlington, R. S. and May, D., *Handbook of Probability and Statistics With Tables, 2nd Edition*, McGraw-Hill, New York, 1970.

Chung, K. L., *A Course in Probability Theory Revised, 2nd Edition*, Academic Press, Boston, Massachusetts, 2000.

DeGroot, M. H. and Schervish, M. J., *Probability and Statistics, 3rd Edition*, Addison Wesley, Boston, Massachusetts, 2001.

Garcia, L., *Probability and Random Processes for Electrical Engineering: Student Solutions Manual, 2nd Edition*, Addison Wesley, Boston, Massachusetts, 1993.

Gnedenko, B. V. and Khinchin, A. Ya., *An Elementary Introduction to the Theory of Probability, 5th Edition*, Dover Publications, New York, 1962.

Goldberg, S., *Probability: An Introduction*, Dover Publications, New York, 1987.

Grimmett, G. R. and Stirzaker, D. R., *Probability and Random Processes, 3rd Edition*, Oxford University Press, Oxford, 2001.

Hsu, H., *Schaum's Outline of Probability, Random Variables, and Random Processes*, McGraw-Hill, New York, 1996.

Kokoska, S. and Zwillinger, D. (Editors), *CRC Standard Probability and Statistics Tables and Formulae, Student Edition*, Chapman & Hall/CRC, Boca Raton, Florida, 2000.

Lange, K., *Applied Probability*, Springer, New York, 2004.

Ledermann, W., *Probability* (Handbook of Applicable Mathematics), Wiley, New York, 1981.

Lipschutz, S., *Schaum's Outline of Probability, 2nd Edition*, McGraw-Hill, New York, 2000.

Lipschutz, S. and Schiller, J., *Schaum's Outline of Introduction to Probability and Statistics*, McGraw-Hill, New York, 1998.

Montgomery, D. C. and Runger, G. C., *Applied Statistics and Probability for Engineers, Student Solutions Manual, 4th Edition*, Wiley, New York, 2006.

Pitman, J., *Probability*, Springer, New York, 1993.

Ross, S. M., *Introduction to Probability Models, 8th Edition*, Academic Press, Boston, Massachusetts, 2002.

Ross, S. M., *A First Course in Probability, 7th Edition*, Prentice Hall, Englewood Cliffs, New Jersey, 2005.

Rozanov, Y. A., *Probability Theory: A Concise Course*, Dover Publications, New York, 1977.

Seely, J. A., *Probability and Statistics for Engineering and Science, 6th Edition*, Brooks Cole, Stanford, 2003.

Shiryaev, A. N., *Probability, 2nd Edition* (Graduate Texts in Mathematics), Springer, New York, 1996.

Walpole, R. E., Myers, R. H., Myers, S. L., and Ye, K., *Probability & Statistics for Engineers & Scientists, 8th Edition*, Prentice Hall, Englewood Cliffs, New Jersey, 2006.

Zwillinger, D. (Editor), *CRC Standard Mathematical Tables and Formulae, 31st Edition*, Chapman & Hall/CRC, Boca Raton, Florida, 2002.

Part II

Physics

Chapter P1

Physical Foundations of Mechanics

Preliminary remarks. *Mechanical motion* is change in the location of a body with respect to other bodies. This definition implies that mechanical motion is relative. In order to describe motion, one should specify a *frame of reference*, which includes a body of reference, a coordinate system fixed relative to the body, and a set of clocks synchronized with one another. Mechanics studies motions of model objects, a point particle (or a point mass) and a rigid body. The location of these objects is determined by a finite set of independent parameters; the objects are said to have finitely many *degrees of freedom*. *Kinematics* deals with the characterization of motion without finding out its reasons.

P1.1. Kinematics of a Point

P1.1.1. Basic Definitions. Velocity and Acceleration

► **Point particle. Law of motion. Path, distance and displacement.** A body whose dimensions can be neglected in studying its motion (compared to the distances of its movement) is called a *point particle* (or just a *particle*). The position of a point particle at an instant of time t is determined by the *position vector* $\mathbf{r}$ from the origin of some reference frame to the particle (see Fig. P1.1). As the particle moves, the end of the position vector traces a spatial curve, a *path* (also called a *trajectory*). In a rectangular Cartesian reference frame, the position vector is determined by its projections onto the coordinate axis, its x-, y-, and z-coordinates. The motion of a particle is completely determined by specifying its *law of motion*, a single vector function $\mathbf{r}(t)$ or three scalar functions $x(t)$, $y(t)$, $z(t)$. A position vector (or any other vector) can be conveniently written in terms of its projections using unit vectors, $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$, of the respective coordinate axes as follows: $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$. The *distance* traveled by the particle in a given time interval is measured along the curvilinear path. Distance is a scalar quantity; it is nonnegative and nondecreasing with time. The *displacement* of a particle is the vector $\Delta\mathbf{r} = \mathbf{r} - \mathbf{r}_0$ that connects an original position of the particle with a final one and is equal to the difference of the position vectors at the initial and final time.

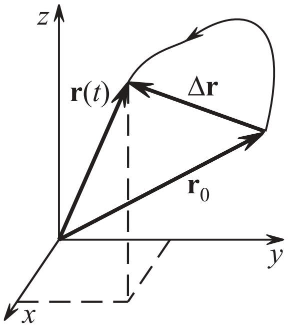

Figure P1.1. Path and displacement.

► **Velocity.** The (instantaneous) *velocity* of a particle is the derivative of its position vector with respect to time:

$$\mathbf{v} = \frac{d\mathbf{r}}{dt}.$$

The velocity is tangent to the path. The (time) average velocity over a finite time interval Δt is defined as the ratio of the displacement to the time interval, $\mathbf{v}_{\text{ave}} = \Delta\mathbf{r}/\Delta t$. (The mean speed is equal to the ratio of the distance traveled to the time interval.) Motion is called *uniform* if $\mathbf{v} = \text{const}$. A uniform motion is a motion along a straight line. A *uniform motion along a given curvilinear path* is a motion with a constant velocity magnitude. (An example of such a motion is a uniform motion in a circle.)

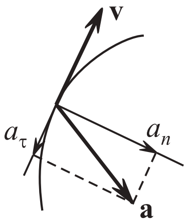

Figure P1.2. Normal and tangential acceleration.

► **Acceleration.** The *acceleration* of a particle is the derivative of the velocity with respect to time:

$$\mathbf{a} = \frac{d\mathbf{v}}{dt}.$$

The vector of acceleration lies in the same plane as the portion of the path where the motion takes places and is directed "inward" the path (if the path is a straight line, the acceleration is directed along it). The projection of the acceleration on the velocity direction is called the *tangential acceleration* and denoted a_τ; it defines the rate at which the velocity magnitude changes with time and equals its time derivative, $a_\tau = dv/dt$. The other component of the acceleration, which is perpendicular to the velocity, is called the *normal acceleration* (Fig. P1.2) and denoted a_n; it characterizes the rate at which the velocity direction changes and equals $a_n = v^2/R$, where R is the radius of curvature of the path (i.e., the radius of the circle that is the best approximation to the path at the given point). The average acceleration over a time interval Δt is defined as $\mathbf{a}_{\text{ave}} = \Delta\mathbf{v}/\Delta t$. A motion is said to be *uniformly accelerated* if $\mathbf{a} = \text{const}$. A *uniformly accelerated motion along a given trajectory* is a motion with $a_\tau = \text{const}$.

P1.1.2. The Direct and Inverse Problems of Kinematics

► **Direct problem.** The problem of determining the characteristics of a motion for a given law of motion is called the *direct problem of kinematics*.

Example 1. Let the law of motion of a particle be given by $x = R\cos\omega t$ and $y = R\sin\omega t$. Find the trajectory equation, velocity, and acceleration of the particle.

Solution. Eliminating time (by making use of the identity $\sin^2\omega t + \cos^2\omega t = 1$) gives the trajectory equation: $x^2 + y^2 = R^2$ (a circle of radius R). Further,

$$\begin{aligned} v_x &= \frac{dx}{dt} = -\omega R\sin\omega t, \\ v_y &= \frac{dy}{dt} = \omega R\cos\omega t \end{aligned} \qquad \text{(velocity components)},$$

$$\left.\begin{aligned} a_x &= \frac{d^2x}{dt^2} = -\omega^2 R\cos\omega t,\\ a_y &= \frac{d^2y}{dt^2} = -\omega^2 R\sin\omega t \end{aligned}\right. \qquad \text{(acceleration components).}$$

The magnitude of the velocity, $v = \sqrt{v_x^2 + v_y^2} = \omega R$, and magnitude of the acceleration, $a = \omega^2 R$, are independent of time and are related by $a = v^2/R$. The acceleration is perpendicular to the velocity and directed towards the center of the circle: $\mathbf{a} = -\omega^2\mathbf{r}$.

► **Inverse problem.** The problem of determining the law of motion for a given acceleration $\mathbf{a}(t)$ is called the *inverse problem of kinematics*. For unique solvability of this problem, one needs to know the *initial conditions*—the position and the velocity of the particle at an initial instant of time.

Example 2. Uniformly accelerated motion. Given the acceleration $\mathbf{a}$ = const of a particle and its initial velocity $\mathbf{v}_0$ and initial position $\mathbf{r}_0$, find the trajectory and the law of motion of the particle.

Solution. Integrating the relation $\frac{d}{dt}\mathbf{v} = \mathbf{a}$ gives the velocity: $\mathbf{v} = \mathbf{v}_0 + \int_0^t \mathbf{a}\,dt = \mathbf{v}_0 + \mathbf{a}t$. Integrating $\frac{d}{dt}\mathbf{r} = \mathbf{v}$ gives the position vector of the particle: $\mathbf{r} = \mathbf{r}_0 + \int_0^t \mathbf{v}\,dt = \mathbf{r}_0 + \mathbf{v}_0 t + \frac{1}{2}\mathbf{a}t^2$.

The particle moves in the plane of the vectors $\mathbf{v}_0$ and $\mathbf{a}$ in a parabola. This becomes apparent in the reference frame where the y-axis is directed along $\mathbf{a}$ and the x-axis is perpendicular to $\mathbf{a}$; we have $x = x_0 + v_{0x}t$ and $y = y_0 + v_{0y}t + \frac{1}{2}at^2$.

► **Other problems.** In the cases where the kinematical problem is neither direct nor inverse and not reduced to a direct or inverse problem, one has to consider a differential equation. Sometimes this equation can be solved by separation of variables.

Example 3. Quite often, when solving dynamical problems, one has to deal with an equation like $dv/dt = -\alpha v^2$ (the body slows down with an acceleration proportional to the velocity squared; α is a constant quantity). Find the speed of the body versus time if its initial value is v_0. Also find the speed versus the displacement s.

Solution. Separating the variables gives $dv/v^2 = -\alpha\,dt$. Integrating the left-hand side from v_0 to v and the right-hand side from 0 to t, one finds $v^{-1} - v_0^{-1} = \alpha t$, and subsequently $v = v_0/(1 + v_0\alpha t)$. Integrating further, one could find the law of motion $s(t)$ and then, having eliminated t, determine $v(s)$. However, the result can be obtained in an easier way: since $v\,dt = ds$, the original separable equation can be rewritten as $dv/v = -\alpha\,ds$. Integrating gives $\ln(v/v_0) = -\alpha s$, and hence $v = v_0\exp(-\alpha s)$.

P1.1.3. Circular Motion

Motion in a circular path can be described using three angular variables: angle of rotation φ, *angular velocity* $\omega = d\varphi/dt$, and *angular acceleration* $\varepsilon = d\omega/dt$. If angles are measured in radians, then the length of the arc corresponding to the angle φ equals $s = \varphi R$. It follows that

$$v = \omega R, \quad a_\tau = \frac{dv}{dt} = \varepsilon R, \quad a_n = \frac{v^2}{R} = \omega^2 R.$$

In the case of a uniform circular motion, the normal acceleration a_n is sometimes called *centripetal*.

The above formulas for the normal and tangential accelerations can be derived by writing the velocity as $\mathbf{v} = v\boldsymbol{\tau}$, where $\boldsymbol{\tau}$ is the unit vector directed along the velocity. Then we have $\mathbf{a} = (dv/dt)\boldsymbol{\tau} + v(d\boldsymbol{\tau}/dt)$. Here we should take into account that $d\boldsymbol{\tau}/dt = \mathbf{n}(d\varphi/dt) = \mathbf{n}(v/R)$, where $\mathbf{n}$ is the unit vector perpendicular to the velocity and directed along the radius. This holds for an arbitrary path as well, provided that the radius of curvature R and the unit normal direction $\mathbf{n}$ are defined by the relation $d\boldsymbol{\tau} = \mathbf{n}(ds/R)$.

P1.1.4. Relativity of Motion. Addition of Velocities

If the motion of a particle is considered relative to two frames of reference, K and K', whose respective axes remain parallel to each other, then the velocities of the particle in these reference frames, $\mathbf{v}$ and $\mathbf{v}'$, satisfy the following velocity-addition formula at any instant of time:

$$\mathbf{v} = \mathbf{v}' + \mathbf{v}_K, \qquad \text{(P1.1.4.1)}$$

where $\mathbf{v}_K$ is the velocity of the reference frame K' relative to K. The accelerations satisfy a similar addition formula: $\mathbf{a} = \mathbf{a}' + \mathbf{a}_K$.

P1.2. Kinematics of a Rigid Body

P1.2.1. Translational and Rotational Motion. Plane Motion

► **Rigid body. Translational and rotational motion**. In mechanics, a *rigid body* is an idealized body the distance between any two points of which does not change (i.e., there is no deformation). Special attention is paid to the description of two simple kinds of motion of a rigid body: translational and rotational. In *translational motion*, the segment connecting any two points of the body moves parallel to itself at all times. Since all points of the body move in the same manner, it suffices to describe the motion of a single point only. In *rotational motion*, all points of the body move in circular paths whose centers lie on the same straight line, called the *axis of rotation*, and the velocities of all points are perpendicular to it. The angular velocities of all points are the same at any instant of time; therefore, it is convenient to introduce a common *vector of angular velocity* $\boldsymbol{\omega}$, directed along the axis of rotation following the right-hand rule. The distribution of linear velocities of points of the body is expressed using cross product*

$$\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}, \qquad \text{(P1.2.1.1)}$$

where the positions vectors are drawn from any point at the axis of rotation. If the axis of rotation is fixed, the vector of angular acceleration, $\boldsymbol{\varepsilon} = d\boldsymbol{\omega}/dt$, is also directed along the axis.

► **Plane motion of a rigid body.** Plane motion of a rigid body is defined as motion in which the velocities of all points of the body are parallel to some plane. If there is a translationally moving reference frame associated with an arbitrary point of the body (or its imaginary continuation), then the relative motion of any other point will be purely rotational about a fixed axis perpendicular to the plane of motion.

Example. Wheel rolling without slipping on a flat surface with a speed v is convenient to represent as the combination of a translation motion with the speed v and a rotational motion with an angular velocity ω (Fig. P1.3). The velocity of any point of the wheel relative to the surface can be found by the velocity-addition law (P1.1.4.1). The speed of the lowest point O' must be zero; it follows that v and ω are related by $v = \omega R$. The accelerations of all points are directed toward the wheel center.

P1.2.2. Instantaneous Axis of Rotation. Addition of Angular Velocities. Motion Relative to a Rotational Frame

► **Instantaneous axis of rotation.** If some point of a body (or its imaginary continuation) is at rest at some instant, then there exists a straight line that passes through this point and is at rest at this instant. This straight line is called an *instantaneous axis of rotation*. The velocity distribution at the instant in question is given by (P1.2.1.1). The instantaneous axis and $\boldsymbol{\omega}$ can change their position both in space and relative to the body. In particular, the velocities of all points of the wheel in the above example can be obtained as the result of pure rotation about the instantaneous axis through contact point O'. The angular acceleration $\boldsymbol{\varepsilon}$ can be nonparallel to the instantaneous axis.

---

* This equation is a special case of the important mathematical formula $d\mathbf{A}/dt = \boldsymbol{\omega} \times \mathbf{A}$ expressing the time derivative of any vector $\mathbf{A}$ rotating with an angular velocity $\boldsymbol{\omega}$.

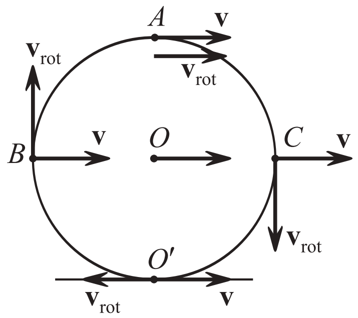
A
v
$\mathbf{v}_{\text{rot}}$
$\mathbf{v}_{\text{rot}}$
B
v
O
C
v
$\mathbf{v}_{\text{rot}}$
O′
$\mathbf{v}_{\text{rot}}$
v

Figure P1.3. Rolling of a wheel can be represented as the sum of its translational and rotational motions or as a pure rotation about the instantaneous axis O'.

► **Motion of a rigid body with a fixed point. Addition of angular velocities.** The motion of a rigid body with a fixed point is, at any instant of time, pure rotation about an instantaneous axis. If this motion can be represented as rotation with an angular velocity $\boldsymbol{\omega}_2$ relative to a reference frame rotating with an angular velocity $\boldsymbol{\omega}_1$, then the resulting angular velocity equals $\boldsymbol{\omega} = \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2$. To prove this, one should generalize the velocity-addition law (P1.1.4.1) to nontranslational motion of the reference frame:

$$\mathbf{v} = \mathbf{v}_{\text{rel}} + \mathbf{v}_{\text{transl}}, \tag{P1.2.2.1}$$

where $\mathbf{v}_{\text{transl}}$ is the *translatory velocity* of the moving reference frame at the point considered. If the reference is rotating, we have

$$\mathbf{v} = \mathbf{v}_{\text{rel}} + \boldsymbol{\omega}_1 \times \mathbf{r}. \tag{P1.2.2.2}$$

If the relative motion is pure rotation, $\mathbf{v}_{\text{rel}} = \boldsymbol{\omega}_2 \times \mathbf{r}$, then $\mathbf{v} = \boldsymbol{\omega}_1 \times \mathbf{r} + \boldsymbol{\omega}_2 \times \mathbf{r} = (\boldsymbol{\omega}_1 + \boldsymbol{\omega}_2) \times \mathbf{r}$ and so the resulting motion is rotation with the angular velocity $\boldsymbol{\omega}_1 + \boldsymbol{\omega}_2$.

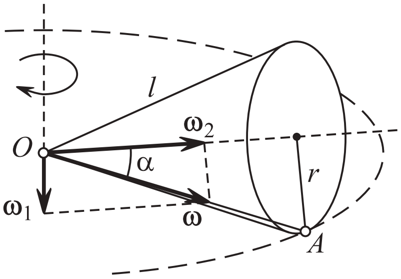

Figure P1.4. A cone rolls on a plane. The sum of the rotational motions is a pure rotation about the axis OA.

Example. A circular cone with a semi-apex angle α is put on its side and rolled without sliding so that its axis rotates with an angular velocity $\boldsymbol{\omega}_1$ (Fig. P1.4). Find the total angular velocity and angular acceleration of the cone.

Solution. In the rotating reference frame where the cone axis is at rest, the cone rotates about this axis with an angular velocity $\boldsymbol{\omega}_2$. The relation between ω_1 and ω_2 can be found from the condition that the point A is at rest; we have $\omega_1 l = \omega_2 r$, or $\omega_1 = \omega_2 \sin\alpha$. It follows that the vector $\boldsymbol{\omega} = \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2$ is horizontal and directed along the line of contact between the cone and the surface; this line is, of course, the instantaneous axis of rotation. The angular acceleration is contributed by only the rotation of the vector $\boldsymbol{\omega}_2$ with the angular velocity $\boldsymbol{\omega}_1$, so that $\boldsymbol{\varepsilon} = d\boldsymbol{\omega}/dt = \boldsymbol{\omega}_1 \times \boldsymbol{\omega}_2$ and $\varepsilon = \omega_1\omega_2 \cos\alpha$.

Formula (P1.2.2.2) is a special case of the formula for the time derivative of an arbitrary vector $\mathbf{A}$:

$$\frac{d\mathbf{A}}{dt} = \left(\frac{d\mathbf{A}}{dt}\right)_{\text{rel}} + \boldsymbol{\omega} \times \mathbf{A}, \tag{P1.2.2.3}$$

where the first term denotes the rate of change of $\mathbf{A}$ relative to the rotating reference frame (see also the footnote to formula (P1.2.1.1)).

In studying motion in noninertial reference frames (see Subsection P1.8.3), we will need the relation between the accelerations of a particle in a fixed reference frame and a rotating one. This relation can be obtained by applying (P1.2.2.3) to either term in (P1.2.2.2):

$$\frac{d\mathbf{v}}{dt}=\frac{d\mathbf{v}_{\rm rel}}{dt}+\boldsymbol{\omega}\times\frac{d\mathbf{r}}{dt}=\left(\frac{d\mathbf{v}_{\rm rel}}{dt}\right)_{\rm rel}+\boldsymbol{\omega}\times\mathbf{v}_{\rm rel}+\boldsymbol{\omega}\times(\mathbf{v}_{\rm rel}+\boldsymbol{\omega}\times\mathbf{r})=\mathbf{a}_{\rm rel}+2\boldsymbol{\omega}\times\mathbf{v}_{\rm rel}+\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r}).$$

The last term, $\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r})$, is the translatory acceleration of the reference frame at the given point (radial or centripetal). Note that along with the relative and translatory accelerations, an additional term arises here. This term, $2\boldsymbol{\omega}\times\mathbf{v}_{\rm rel}$, is called the *Coriolis acceleration*.

P1.3. Dynamics

P1.3.1. Newton's First Law. Mass. Momentum. Force

► **Newton's first law**. Newton's first law states that there are reference frames in which a body not interacting with other bodies is either at rest or in motion in a straight line with constant speed. The reference frames postulated by this law are called *inertial*.

It follows from the velocity-addition law that any reference frame that moves with a constant velocity relative to an inertial reference frame is also inertial. A heliocentric reference frame, related to the Sun and distant stars, can be considered inertial virtually exactly. The Earth can be treated as an inertial reference frame only approximately—the acceleration of the points at its surface amounts to $0.034\,\mathrm{m/s^2}$ due to Earth's daily rotation. In addition, the Earth orbits the Sun but the associated acceleration is three orders of magnitude lower (check this as an exercise).

► **Mass. Momentum**. From Newton's first law it follows that in inertial reference frames, a body's acceleration results from interaction of the body with other bodies. This acceleration depends on the *inertia* of the body, or its ability to resist a change in velocity, and also on the intensity and direction of action of other bodies.

The *mass* of a body, m, is a positive scalar quantity that characterizes the body's inertia. It follows from experiments that when two bodies interact with each other, their accelerations in an inertial reference frame, $\mathbf{a}_1$ and $\mathbf{a}_2$, are opposite to each other in direction and the ratio of their magnitudes, a_1/a_2, is independent of the character and intensity of the interaction. This makes it possible to define the ratio of two masses as the reciprocal of the ratio of the accelerations resulting from the interaction of the bodies: $m_1/m_2=a_2/a_1$. In order to express mass numerically, one has to choose a mass unit. In the International System of Units (SI), the unit of mass is the *kilogram* (kg). Mass possesses the property of additivity: if a body is divided into a number of parts, the sum of masses of all parts equals the mass of the whole body. However, the theory of relativity shows that this property is only valid approximately—it is violated if there is strong interaction between parts of the body. For example, the mass of a nucleus turns out to be less than the sum of masses of its constituting nucleons.

The *momentum* (also referred to as *linear momentum*) of a particle is a vector quantity that equals the product of the mass of the particle by its velocity:

$$\mathbf{p}=m\mathbf{v}. \qquad \text{(P1.3.1.1)}$$

Definition (P1.3.1.1) applies only if $v\ll c$, where c is the speed of light. A more general definition that is valid for any $v<c$ is introduced in the theory of relativity.

► **Force.** A force acting on a body of mass m interacting with another body is the vector quantity $\mathbf{F} = m\mathbf{a}$, where $\mathbf{a}$ is the acceleration imparted by this force to the body m in an inertial reference frame. The more general definition is $\mathbf{F} = d\mathbf{p}/dt$. In Newtonian mechanics, both definitions are equivalent. However, in the theory of relativity, only the latter is valid. In SI, force is measured in *newtons* (*N*); $1\,\mathrm{N} = 1\,\mathrm{kg\,m/s^2}$.

The change of momentum in a time interval Δt, from time 0 to t, under the action of a force $\mathbf{F}$ equals the integral $\Delta\mathbf{p} = \mathbf{p}(t) - \mathbf{p}(0) = \int_0^t \mathbf{F}\,dt$. This quantity is called the *impulse* produced by the force F in the time interval Δt. The average force equals $\mathbf{F}_{\mathrm{ave}} = \Delta\mathbf{p}/\Delta t$.

P1.3.2. Newton's Second and Third Laws

► **Newton's second law.** If a particle is acted upon by N bodies with forces $\mathbf{F}_1, \dots, \mathbf{F}_N$, then the particle acquires an acceleration $\mathbf{a}$ determined by the *vector sum of the forces*:

$$\mathbf{a} = \frac{\mathbf{F}}{m} = \frac{1}{m}\sum_{j=1}^{N}\mathbf{F}_j \qquad \text{or} \qquad \frac{d\mathbf{p}}{dt} = \sum_{j=1}^{N}\mathbf{F}_j.$$

The vector $\mathbf{F} = \sum_{j=1}^{N}\mathbf{F}_j$ is called the *net force* (also known as the *resultant force*) of the forces $\mathbf{F}_1, \dots, \mathbf{F}_N$. Newton's second law is an *equation of motion* of a particle, since it allows (in principle) the calculation of the motion of the particle if the net force applied is a known function of time and the particle position (and velocity). Likewise, given a system of particles, the set of Newton's second laws for each of them allows the calculation of the motion of the particles if the forces applied are known functions of time and mutual positions of the particles.

Outlined below are the properties of forces that are useful in solving mechanical problems.

► **Newton's third law** states that the forces $\mathbf{F}_{AB}$ and $\mathbf{F}_{BA}$ with which two bodies, A and B, act upon each other are directed along the same straight line, opposite in direction, and equal in magnitude:

$$\mathbf{F}_{AB} = -\mathbf{F}_{BA}.$$

P1.3.3. Action at a Distance. Causality. Natural Forces

► **Action at a distance. Causality. Field.** In Newtonian mechanics, Newton's third law is considered to hold regardless of the nature of interaction and of whether the bodies are in direct contact with each other or interact at a distance by means of gravitational or electromagnetic forces. However, the validity of Newton's third law for long-range forces would mean instantaneous transmission of information about changes in body positions. The notion of long-range action contradicts the postulates of the theory of relativity, which forbids information transmission at speeds faster than the speed of light. Modern physics has rejected the use of the concept of action at a distance. Instead, it has introduced a new participant—a *force field* (e.g., gravitational, electric, etc.) that fills the whole of space. A particle at a given point in space is acted upon not by distant bodies but by a field in the vicinity of the point; this field is produced by distant bodies (*field sources*). A change in the field is transmitted from point to point and propagates with a final speed (speed of light). Newtonian mechanics is valid in low-speed approximation, $v \ll c$, and in this, *nonrelativistic limit*, one can use both action at a distance and field.

It is convenient to study motion of a particle (or a system of particles) in an *external stationary field* generated by sources at rest (e.g., Earth's gravitational field). A force field is called *central* if the force exerted by the field on a particle placed in it is directed toward a single point called the *field origin* (*force center*) and the magnitude of the force depends only on the distance r to the origin, so that

$$\mathbf{F}(\mathbf{r}) = F_r(r)\frac{\mathbf{r}}{r},$$

where $F_r(r)$ is the projection of the force onto the radial direction; $F_r > 0$ corresponds to repulsive forces and $F_r < 0$ corresponds to attractive forces. A field is called *uniform* in a region of space if this force is the same (in direction and magnitude) at every point of space: $\mathbf{F} = \text{const}$.

► **Natural forces**. The whole diversity of naturally occurring forces can be reduced to a few fundamental types of interaction: gravitational, weak, electromagnetic, and strong. However, for practical purposes, this approach is unsuitable. Apart from the fundamental long-range forces (gravitational force in mechanics), one needs to know properties of various forces arising when macroscopic bodies are in contact with one another. Below is the list of forces arising in solving mechanical problems.

1. *Gravitational force* $m\mathbf{g}$ (the force with which a body at the Earth's surface is pulled to the Earth's center).

2. *Elastic force*. When a body is deformed in a certain direction, the restoring elastic force is proportional to the deformation (*Hooke's law*): $F_\text{e} = k|\Delta l|$, where k is the spring constant (body stiffness) and Δl is the deformation in this direction. The spring constant of an elastic rod is proportional to its cross-sectional area and inversely proportional to its length, so that

$$F_\text{e} = ES\frac{|\Delta l|}{l},$$

where E is *Young's modulus*, which depends on the rod material. In the limit of infinitely large stiffness, the (normal) reaction force N and the string tension force T arise; these forces are determined from the equations of motion. Such forces are called *constraint reaction forces* or *constraint reactions*.

3. *Dry friction force*. Sliding friction force: $F_\text{fr} = \mu_1 N$, where μ_1 is the sliding friction coefficient. Static friction force: $F_\text{fr} \le \mu_2 N$, where μ_2 is the static friction coefficient. It is usually assumed that $\mu_1 = \mu_2$ in solving problems. The static friction force is also a constraint reaction.

4. *Drag force*. Drag refers to forces that oppose the relative motion of a body through a liquid or gas. It depends on the relative speed and the transverse size of the body. For low relative speeds and small sizes, the drag force is proportional to the speed (viscous drag). For large relative speeds, the drag force is proportional to the cross-sectional area and the speed squared (head-on drag or quadratic drag).

P1.3.4. Galilean Relativity Principle

Galilean relativity (also known as *Galilean invariance*) is a principle that states that all laws of mechanics are the same in all inertial reference frames. Suppose a frame K' moves relative to a frame K with a constant velocity $\mathbf{V}$. Let the respective axes of the two frames be parallel to each other and let the x- and x'-axes be directed along $\mathbf{V}$ (Fig. P1.5). Also let time be counted off from the instant when the frames coincide. Then the coordinates and time in the frame K' are expressed in terms of those in the frame K using the *Galileo transformations*

$$x' = x - Vt, \qquad y' = y, \qquad z' = z, \qquad t' = t. \tag{P1.3.4.1}$$

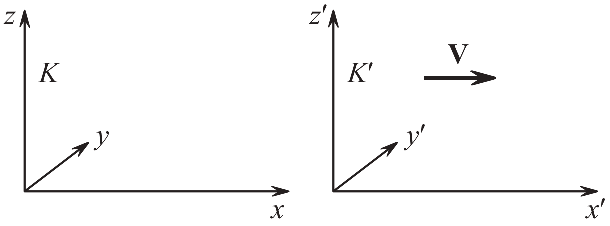

Figure P1.5. Two inertial reference frames. The mechanical laws have the same form in both frames.

Under the Galileo transformations, the difference of the velocities of two particles as well as the distance between them remain unchanged; this means that the forces acting between the two particles also remain the same. Furthermore, the Galileo transformations do not change the accelerations of the particles. Consequently, both sides of the equation of Newton's second law remain unchanged. So the equation of motion has the same form in various reference frames (*invariance under the Galileo transformations*).

The Galileo transformations are based on the assumption that the course of time and the lengths of segments are independent of the reference frame, which were considered to be inherent properties of space and time. The theory of relativity revises the traditional view of space and time and shows that the Galileo transformations are valid for $V \ll c$. The more general transformations are the Lorenz transformations, which are valid for any $V < c$.

P1.4. Law of Conservation of Momentum

P1.4.1. Center of Mass. Momentum of a System of Particles

► **System of particles. Center of mass.** Consider a system of N particles. The forces acting on a jth particle can be divided into *internal* ones, $\mathbf{F}^{\rm i}_{jk}$, exerted by the other particles, and *external* ones, the resultant of which will be denoted $\mathbf{F}^{\rm e}_{j}$. According to Newton's third law, the sum of the internal forces is zero. Hence, the sum of all forces acting on the system equals the sum of external forces. If the particles do not interact with external bodies, then the system is called *closed* (or *isolated*).

The *center of mass* (or *center of inertia*) of a system of N particles with masses $m_1, \ldots, m_N$ and position vectors $\mathbf{r}_1, \ldots, \mathbf{r}_N$ is the point whose position is defined by the vector

$$\mathbf{r}_{\rm cm} = \frac{m_1\mathbf{r}_1 + \cdots + m_N\mathbf{r}_N}{m_1 + \cdots + m_N}. \qquad \text{(P1.4.1.1)}$$

A translationally moving reference frame with origin at the center of mass is called a *center-of-mass frame*.

If the mass of a system is distributed continuously (in space, on surface, or along a line), the system can be mentally divided into a large number of small regions to represent particles. Further, passing to a limit, one can replace summation by integration. The distribution of mass over volume is defined using *mass density* (or simply *density*), $\rho(\mathbf{r})$, such that $dm = \rho(\mathbf{r})\,dV$. Then, for continuously distributed mass, the definition of the center of mass (P1.4.1.1) becomes

$$\mathbf{r}_{\rm cm} = \frac{\int \mathbf{r}\rho(\mathbf{r})\,dV}{\int \rho(\mathbf{r})\,dV}.$$

A body with constant density across the whole volume is called a *uniform body*.

► **Momentum.** The momentum of a system of particles is defined as the sum of momenta of the constituting particles:

$$\mathbf{p} = \sum \mathbf{p}_j.$$

Differentiating equation (P1.4.1.1) with respect to time, one can find that the momentum of the system is expressed in terms of the center-of-mass velocity as

$$\mathbf{p} = m\mathbf{v}_{\text{cm}}, \tag{P1.4.1.2}$$

where m is the total mass of the system. It follows that the momentum of a system of particles equals zero in a center-of-mass frame.

Summing up the Newton's Second Law equations for each particle, $d\mathbf{p}_j/dt = \mathbf{F}_j$, and taking into account that the internal forces are canceled out, one obtains

$$\frac{d\mathbf{p}}{dt} = \sum \mathbf{F}_j^{\text{e}} \tag{P1.4.1.3}$$

This means that the rate of change of the momentum of the system equals the net external force.

P1.4.2. Law of Conservation of Momentum. Motion of Center of Mass

► **Law of conservation of momentum.** From equation (P1.4.1.3) it follows that the *momentum of a closed system is conserved*. The law of conservation of (linear) momentum is a consequence of uniformity of space (there is no distinguished point in space) and is therefore a fundamental law of nature. Although derived within the scope of Newtonian mechanics, this law remains valid beyond it. Even with allowance for a finite speed of signal propagation, which leads to violation of Newton's third law, the law of conservation of momentum is obeyed exactly, provided that the momentum of the transmitter (force field) is taken into account.

The *momentum of an unclosed system* is conserved in the following cases:

1. The net external force is zero.
2. The net external force is perpendicular to some direction. Then the projection of momentum on this direction (not the whole momentum) is conserved.
3. If the interaction last a very short time Δt and the net external force $\mathbf{F}^{\text{e}}$ is bounded,* then the change in the momentum of the system, $\mathbf{F}^{\text{e}}\Delta t$, can be neglected (effectively assuming that $\Delta t \to 0$).

► **Motion of center of mass**. On differentiating equation (P1.4.1.2) with respect to time and taking into account (P1.4.1.3), we arrive at the equation of motion of the center of mass

$$m\mathbf{a}_{\text{cm}} = \sum \mathbf{F}_j^{\text{e}}. \tag{P1.4.2.1}$$

It follows that the center of mass moves in the same manner as though it was a single particle, whose mass is equal to the mass of the system, moving under the action of the net external force.

---

* Under these conditions, the momenta of particles can change considerably under the action of large internal forces (explosion, impact, etc.).

Example. A thin rod of length l and mass m rotates in a horizontal plane about one of its ends with an angular velocity ω. Find the tension force at the middle of the rod.

Solution. The center of mass of the outer half of the rod moves in a circle of radius $\frac{3}{4}l$ under the action of a single force, the unknown tension force. It follows from equation (P1.4.2.1) that $F = \bigl(\frac{1}{2}m\bigr)\omega^2\bigl(\frac{3}{4}l\bigr) = \frac{3}{8}m\omega^2 l$.

Note that the center of mass of a closed system moves with a constant velocity; hence, a center-of-mass frame is inertial.

► **Spacecraft propulsion.** For a spacecraft in outer space (far away from gravitating bodies), a change in its velocity is only possible by exhausting part of mass, rocket propellant. The equation of motion of a spacecraft under the action of an external force **F** and a propulsive jet can be easily obtained by writing down Newton's second law in the form (P1.4.1.3) in an inertial frame associated with the spacecraft at the current instant:

$$m\,\Delta\mathbf{v} + \Delta M\,\mathbf{u} = \mathbf{F}\,\Delta t,$$

where **u** is the jet velocity relative to the spacecraft and ΔM is the mass of the propellant exhausted in time Δt. Dividing by Δt, one arrives at the *Meshcherskii equation*:

$$m\mathbf{a} = \mathbf{F} - \mu\mathbf{u},$$

where $\mu = \Delta M/\Delta t = -\Delta m/\Delta t$ is the *propellant use* in the jet. The latter term on the right-hand side is called the *propulsive force*.

Let us write the Meshcherskii equation for the case of motion in a straight line with no external force applied to obtain $m(dv/dt) = -(dm/dt)u$. Assuming that u is constant and denoting the initial mass of the spacecraft by m_0, we find how the speed of the spacecraft depends on its mass: $v/u = -\ln(m/m_0)$ (*Tsiolkovsky equation*).

P1.5. Law of Conservation of Energy

P1.5.1. Work and Power. Kinetic Energy

► **Mechanical work** is a scalar quantity that is a measure of change in energy. The *work of external forces equals the change in energy*:

$$A = E_2 - E_1. \qquad \text{(P1.5.1.1)}$$

Based on this equation, we will define different kinds of energy. (A refinement of this principle will be given in Chapter P2.)

The work of a force on a small portion of a path is defined as the dot product of the force by the displacement: $\delta A = \mathbf{F}\cdot d\mathbf{r}$. The work on the whole path from position 1 to position 2 equals

$$A = \int_1^2 \mathbf{F}\cdot d\mathbf{r} = \int_1^2 F_\tau\,|d\mathbf{r}| = \int_1^2 F\,|d\mathbf{r}|\cos\alpha,$$

where F_τ is the projection of the force on the direction of motion of the point where the force is applied and α is the angle between the force and this direction. In SI, work and energy are measured in *joules* (*J*) (1 J = 1 N m).

Example 1. The work of a constant force **F**,

$$A = \int_1^2 \mathbf{F}\cdot d\mathbf{r} = \Bigl(\mathbf{F}\cdot\int_1^2 d\mathbf{r}\Bigr) = \mathbf{F}\cdot(\mathbf{r}_2 - \mathbf{r}_1),$$

equals the dot product of the force by the total displacement and is independent of the path of the point at which the force is applied.

Example 2. The work of a central force, the force exerted by a central field (see Subsection P1.3.3), equals

$$A = \int_1^2 \mathbf{F} \cdot d\mathbf{r} = \int_{r_1}^{r_2} F_r(r)\, dr.$$

It has been taken into account that the projection of $d\mathbf{r}$ onto the radial direction is equal to the change in the distance r to the origin. The work depends only on the initial, r_1, and final, r_2, distance to the origin and is independent of the path.

► **Mechanical power** is the rate at which work is performed. The average power is the ratio of the work done to the time interval. The instantaneous work equals the time derivative of the work done:

$$P = \frac{dA}{dt} = \frac{\mathbf{F} \cdot d\mathbf{r}}{dt} = \mathbf{F} \cdot \mathbf{v} = F_\tau v.$$

Power is measured in *watts* (*W*): 1 W = 1 J/s.

► **Kinetic energy.** The kinetic energy of a particle is the energy it possesses due to its motion and depending on its speed. The velocity of a particle changes under the action of a net force **F**, the work of which equals

$$A = \int_1^2 F_\tau\, |d\mathbf{r}| = m\int_1^2 a_\tau v\, dt = m\int_1^2 \frac{dv}{dt} v\, dt = m\int_{v_1}^{v_2} v\, dv = \tfrac{1}{2}mv_2^2 - \tfrac{1}{2}mv_1^2. \quad \text{(P1.5.1.2)}$$

It is apparent that, in accordance with the general principle (P1.5.1.1), kinetic energy can be defined as $E_\text{k} = mv^2/2$. The above equation, stating that the change in the kinetic energy equals the work of the net force, is known as the *work-energy theorem*.

The *kinetic energy of a system of particles* is defined as the sum of the kinetic energies of all constituting particles. The change in the kinetic energy of the system equals the work of all forces acting on its particles. The kinetic energy of a system of particles equals

$$E_\text{k} = \frac{mv_\text{cm}^2}{2} + E_\text{rel}, \quad \text{(P1.5.1.3)}$$

where m is the mass of the system, v_cm is the speed of its center of mass, and E_rel is the kinetic energy relative to a center-of-mass frame. Equation (P1.5.1.3) is known as *König's theorem*.

P1.5.2. Conservative Forces. Potential Energy

► **Conservative forces. Potential field.** A force of interaction between two particles is called *conservative* if the work of this force depends on only the initial and final positions of the particles and does not depend on their paths. Forces that do not satisfy this condition are called *nonconservative*.

An external stationary field is called *potential* if its work on moving a particle depends on only its initial and final positions and does not depend on its path. (It is equivalent to say that the work of the field on moving a particle through a closed path is zero.) A potential field is a field of conservative forces; it is generated by fixed external sources. As follows from example 1, a uniform field is potential. From example 2 it follows that any central field is potential. A superposition of several central fields (generated by several sources) is also a potential field. Hence, an electrostatic field and a stationary gravitational field are also potential. From example 2 it also follows that the elastic force produced by a light spring is conservative.

A system of particles with only conservative forces acting between them is referred to as a *conservative system*.

► **Potential energy** is energy stored within a system; it characterizes interaction between particles and depends on their mutual positions. The *potential energy of a particle in an external potential field* is defined as follows. Let us define the *difference between the potential energies* of the particle as the work performed by the field to move the particle from one position to another:

$$E_\mathrm{p}(\mathbf{r}_1) - E_\mathrm{p}(\mathbf{r}_2) = A_{12}. \qquad \text{(P1.5.2.1)}$$

This work is independent of the path. Hence, the *change in the potential energies*, $E_2 - E_1$, equals the *negative* of the work done by the field. If the particle is moved very slowly by an *external* force, then the work performed by the external force is equal in magnitude and opposite in sign to the work of the field, so $A^\mathrm{e}_{12} = E_2 - E_1$ in accordance with the general principle (P1.5.1.1).

Equation (P1.5.2.1) defines the difference of potential energies; so the potential energy is defined up to an additive constant. To make the definition unambiguous, one should specify the value (usually zero) of the potential energy at some point in space.

Example 1. The work done by the force of gravity $m\mathbf{g}$ on lifting a particle of mass m from height h_1 to height h_2 is equal to $mgh_1 - mgh_2$. Consequently, the potential energy of the particle in the gravitational field is $E_\mathrm{p} = mgh$, where the height is counted off from a chosen zero level. The potential energy of a system of points in the field of gravity equals

$$E_\mathrm{p} = \sum_j m_j g h_j = mg\frac{\sum_j m_j h_j}{m} = mgh_\mathrm{cm},$$

where m is the mass of the system and h_cm is the height of its center of mass.

Example 2. The work done by the elastic force equals $\int_{x_1}^{x_2}(-kx)\,dx = \frac{1}{2}kx_1^2 - \frac{1}{2}kx_2^2$, where x_1 and x_2 are the initial and final points of the spring deformation. Consequently, the potential energy of the elastic spring is $E_\mathrm{p} = \frac{1}{2}kx^2$, where zero energy corresponds to the undeformed spring.

Example 3. The work done by friction or resistive forces is negative on each segment of the path and also along a closed path. Consequently, these forces do not satisfy the condition of being conservative and are nonconservative.

The *potential energy of conservative interation* of two particles can be treated as the potential energy of one particle in the field generated by the other. The result is independent of which particle is considered to be the field source.

► **Relation between force and potential energy.** By writing equation (P1.5.2.1) for two close particles lying on the x-axis and separated by a distance dx, one obtains $E_\mathrm{p}(x) - E_\mathrm{p}(x + dx) = F_x\,dx$. It follows that the projection F_x of the force $\mathbf{F}$ onto the x-axis is determined by the derivative of the potential energy:

$$F_x = -\frac{\partial E_\mathrm{p}}{\partial x}. \qquad \text{(P1.5.2.2)}$$

The partial derivative means that E_p is treated as a function of only x here. The same is obviously valid for projections on any other direction. Consequently, the vector of force equals the opposite of the *gradient* of the potential energy:

$$\mathbf{F} = -\left(\frac{\partial E_\mathrm{p}}{\partial x}\mathbf{i} + \frac{\partial E_\mathrm{p}}{\partial y}\mathbf{j} + \frac{\partial E_\mathrm{p}}{\partial z}\mathbf{k}\right) = -\operatorname{grad} E_\mathrm{p}, \qquad \text{(P1.5.2.3)}$$

where $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are unit vectors in the x-, y-, and z-axis, respectively.

For central field, formula (P1.5.2.2) becomes

$$F_r = -\frac{dE_\mathrm{p}(r)}{dr}. \qquad \text{(P1.5.2.4)}$$

P1.5.3. Mechanical Energy of a System of Particles. Law of Conservation of Mechanical Energy

► **Mechanical energy.** The mechanical energy of a system of particles is defined as the sum of its kinetic energy, the internal potential energy due to interaction between the particles, and the potential energy in an external field:

$$E_{\text{mech}} = \sum_j \frac{m_j v_j^2}{2} + \sum_{j<n} E_{jn} + \sum_j E_j^{\text{e}}. \tag{P1.5.3.1}$$

According to König's theorem (P1.5.1.3), the kinetic energy of a system equals the sum of two terms, the energy of its motion as a whole, $mv_{\text{cm}}^2/2$, and the kinetic energy in a center-of-mass frame, E_{rel}. The sum of E_{rel} and the potential energy of particle interactions is sometimes called the *internal mechanical energy* of the system.

► **Change in mechanical energy.** The change in the kinetic energy of a system equals the total work of all forces applied to the particles (see equation (P1.5.1.2)). The change in the potential energy equals the negative of the work of the conservative forces and external potential fields. Consequently, the change in the mechanical energy of the system equals the work done by the internal nonconservative forces and all external forces, except for the potential fields, whose contribution is included into E_{mech}. So we have

$$\Delta E_{\text{mech}} = A_{\text{nonconservative}} + A^{\text{e}}. \tag{P1.5.3.2}$$

► **Law of conservation of mechanical energy.** The mechanical energy of a closed conservative system does not change. This statement is a special case of the general fundamental principle of conservation of energy: the *total energy of a closed system is conserved*. Apart from the mechanical energy, the total energy also includes the different kinds of internal energy: thermal, chemical, and nuclear. The general principle of conservation of energy goes far beyond the scope of Newtonian mechanics, within the framework of which the law of conservation of the mechanical energy has been obtained. This principle is in close connection with the fundamental postulate of *uniformity of time* (there is no selected instant in time) and is one of the fundamental principles on which modern physics is based.

Note that if a conservative system is in an external potential field, the mechanical energy is also conserved, provided that the potential energy of the particles in the field is included in it.

The condition for a system to be conservative is equivalent to the requirement that both the mechanical and internal components of the total energy must be conserved individually. If, for example, there are friction forces acting in the system, whose work is negative, then the mechanical energy is reduced (see equation (P1.5.3.2)) being converted into the internal energy; in this case, heat is said to be produced in the system. The mechanical energy can change also in the case where there is a mechanism present in the system that can produce work at the expense of the internal energy (e.g., an internal combustion engine, a human being, etc.).

Example 1. Elastic collision. In a central elastic impact of two balls, both the momentum of the system and its mechanical energy are conserved. If m_1 and m_2 are the masses of the balls, v_1 and v_2 are speeds of the balls before the collision, and u_1 and u_2 are those after the collision, we have

$$m_1 v_1 + m_2 v_2 = m_1 u_1 + m_2 u_2,$$

$$\frac{m_1 v_1^2}{2} + \frac{m_2 v_2^2}{2} = \frac{m_1 u_1^2}{2} + \frac{m_2 u_2^2}{2}.$$

Instead of the second equation, it is more convenient to use the condition that the relative speed of the balls does not change in magnitude but changes in sign: $v_{1x} - v_{2x} = u_{2x} - u_{1x}$. This equation can be derived directly from the first two, but it becomes obvious if we consider the motion in a center-of-mass frame (the relative speed remains the same here). In this frame, the total momentum of the system is zero and the speed of either ball simply reverses after the impact; both conservation laws are obeyed. Then, solving the two linear equations, we find the final speeds of the balls:

$$u_1 = \frac{(m_1 - m_2)v_1 + 2m_2v_2}{m_1 + m_2}, \qquad u_2 = \frac{(m_2 - m_1)v_2 + 2m_1v_1}{m_1 + m_2}.$$

If we replace the second ball with a moving wall ($m_2 \gg m_1$), we get

$$u_1 \approx -v_1 + 2v_2, \quad u_2 \approx v_2.$$

Example 2. Inelastic collision. Two balls of masses m_1 and m_2 with velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ undergo a perfectly inelastic collision such that the balls move as one with a velocity $\mathbf{u}$ after the collision. (No rotation occurs if the impact is central in a center-of-mass frame.) The velocities become equal as a result of nonconservative forces, which means that some heat is released in the impact. This can be verified by finding the final velocity using the law of conservation of momentum, $m_1\mathbf{v}_1 + m_2\mathbf{v}_2 = (m_1 + m_2)\mathbf{u}$, and calculating the decrease in the mechanical energy:

$$\frac{m_1v_1^2}{2} + \frac{m_2v_2^2}{2} - \frac{(m_1 + m_2)u^2}{2} = \frac{m_1m_2}{m_1 + m_2}\frac{(\mathbf{v}_1 - \mathbf{v}_2)^2}{2}.$$

It is even easier to obtain this result considering the motion in a center-of-mass frame, where the balls are at rest after the collision.

P1.5.4. Potential Curves. Stability

If we know the potential energy $E_\mathrm{p}(r)$ of a one-dimensional motion in a potential field (an example is shown in Fig. P1.6), then we can find out the following based on relation (P1.5.2.2):

a) The direction of the force ($F > 0$ for $0 < r < r_\mathrm{min}$ and $r > r_\mathrm{max}$; $F < 0$ for $r_\mathrm{min} < r < r_\mathrm{max}$).

b) The points of equilibrium ($F = 0$ at $r = r_\mathrm{min}$ and $r = r_\mathrm{max}$).

c) Stability of equilibrium. In the vicinity of the point $r = r_\mathrm{min}$, the force is directed toward the point, so the equilibrium is stable. At $r = r_\mathrm{max}$, the equilibrium is unstable. A stable equilibrium corresponds to a minimum of the potential energy.

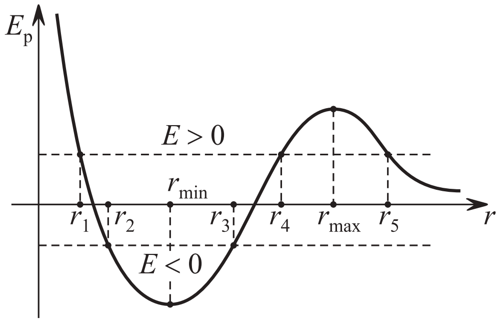

Figure P1.6. Given a potential energy curve, one can determine the direction of the force, the positions of the equilibrium points and their stability, and the character of motion.

d) The character of motion for a given value of the total mechanical energy E. Motion is possible only where $E > E_\mathrm{p}$ (the kinetic energy is nonnegative). If $E < E_\mathrm{p}(\infty)$ (in Fig. P1.6, $E_\mathrm{p}(\infty) = 0$), motion is *finite*, so it can occur in a limited range of r. With $E < 0$ in Fig. P1.6, motion occurs only between the *turning points* r_2 and r_3. For $E > 0$, motion is either *infinite*, where the particle goes to infinity with kinetic energy E after reflecting from the turning point r_5, or locked by a potential barrier, where the particle moves between the turning points r_1 and r_4. In classical mechanics, a potential barrier is impenetrable. In quantum mechanics, there is a nonzero probability that a potential barrier can be penetrated (tunnel effect).

P1.6. Law of Conservation of Angular Momentum

P1.6.1. Moment of Force. Net Force

► **Moment of force.** The moment of a force **F** about a point O is defined as

$$\mathbf{M} = \mathbf{r} \times \mathbf{F}, \tag{P1.6.1.1}$$

where **r** is the vector from O to the point where the force is applied. A moment of force is also called a *torque*. The moment of force does not change if the force is moved along the straight line of its application. The magnitude of the moment of force equals $M = Fr\sin\alpha = Fd$, where α is the angle between the force vector and the radius vector **r**, and d is the distance between the point O and the line of action of the force; the distance d is called the *moment arm* (Fig. P1.7).

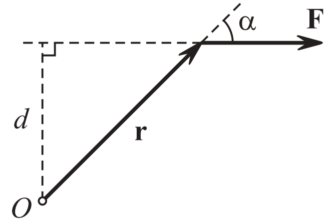

Figure P1.7. The moment of the force **F** is perpendicular to the plane of the figure and directed away from the reader. The magnitude of the moment equals the product of F by the arm d.

The most important property of moment of force is that the sum of the moments of internal forces in an arbitrary system of particles is zero. This fact follows from Newton's third law (see Subsection P1.3.2), since the forces acting between each pair of particles are equal in magnitude, opposite in direction, and directed along the straight line connecting the particles.

Note another property of moment of force: if the sum of forces is zero, then these forces produce the same net moment about any point in space.

► **Net force.** The *net force* of a system of forces is the force that equals the vector sum of them and is applied so that its moment about any point in space equals the net moment of this system of forces.

Example. Compute the net moment of the forces of gravity acting on the particles of a system:

$$\mathbf{M} = \sum \mathbf{r}_j \times m_j\mathbf{g} = \Big(\sum m_j\mathbf{r}_j\Big) \times \mathbf{g} = \mathbf{r}_{\rm cm} \times (m\mathbf{g}),$$

where m is the mass of the system, $\mathbf{r}_{\rm cm}$ is the radius vector of the center of mass. It is apparent that the net force passes through the center of mass; hence, *the center of mass is also the center of gravity*.

The *moment of force about an axis z* is the z-projection, M_z, of the moment of force about any point on the axis. If both **r** and **F** in (P1.6.1.1) are resolved into two components, parallel and perpendicular to the axis z, so that $\mathbf{r} = \mathbf{r}_{\|} + \mathbf{r}_{\perp}$ and $\mathbf{F} = \mathbf{F}_{\|} + \mathbf{F}_{\perp}$, then the components parallel to the axis will not contribute to the projection of the moment onto the axis. It follows that $M_z = (\mathbf{r}_{\perp} \times \mathbf{F}_{\perp})_z$. In the plane perpendicular to the axis, we have the same picture as in Fig. P1.7, except that **r** and **F** must be replaced with $\mathbf{r}_{\perp}$ and $\mathbf{F}_{\perp}$. Accordingly, the magnitude of the moment about an axis is also equal to the product of the force by the moment arm, while the choice of the positive direction along the axis can be replaced with the easier choice of the positive direction of rotation about the axis.

P1.6.2. Angular Momentum of a Particle. Motion in Central Field

► **Angular momentum.** The angular momentum of a particle about a point O is the vector quantity

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}, \tag{P1.6.2.1}$$

where $\mathbf{r}$ is the vector from O to the position of the particle and $\mathbf{p} = m\mathbf{v}$ is the (linear) momentum of the particle. Angular momentum is also known as *moment of momentum*. If the particle moves with a constant velocity, its angular momentum does not change. The magnitude of the angular momentum equals the product of the momentum by the arm. The angular momentum about an axis z is the z-projection, L_z, of the angular momentum $\mathbf{L}$ about any point at the axis. The angular momentum about an axis is determined by the projection of the momentum on a plane perpendicular to the axis: $L_z = (\mathbf{r}_\perp \times \mathbf{p}_\perp)_z$. These properties are similar to those of the moment of force.

The time derivative of the angular momentum of a particle about a point O,

$$\frac{d\mathbf{L}}{dt} = \frac{d\mathbf{r}}{dt} \times \mathbf{p} + \mathbf{r} \times \frac{d\mathbf{p}}{dt} = \mathbf{v} \times \mathbf{p} + \mathbf{r} \times \mathbf{F} = \mathbf{M}, \tag{P1.6.2.2}$$

equals the moment of the net force about O (since $\mathbf{v}$ and $\mathbf{p}$ are parallel to each other, we have $\mathbf{v} \times \mathbf{p} = \mathbf{0}$)

► **Motion in central field.** Since the force acting on a particle in a *central field* (also called a *central-force field*) is directed to the origin of the field, the moment of this force about the origin is identically zero. Hence, the vector angular momentum $\mathbf{L}$ is conserved. It follows from definition (P1.6.2.1) that the motion occurs in one plane perpendicular to $\mathbf{L}$ and the quantity $L = mvr\sin\alpha$ is conserved. The magnitude of the angular momentum is proportional to the rate at which the radius vector $\mathbf{r}$ sweeps out area (Fig. P1.8): $dS/dt = \frac{1}{2}r(v\,dt)\sin\alpha/dt = \frac{1}{2}L/m$. Consequently, the statement that angular momentum is conserved in a central field turns out to be equivalent to *Kepler's second law of planetary motion*.

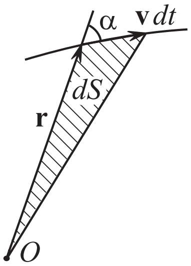

Figure P1.8. The rate at which the radius vector sweeps out area is expressed in terms of the angular momentum.

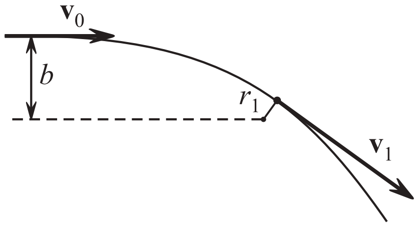

Figure P1.9. Fly-by of a particle near a force field center; b is the aiming parameter.

Example 1. Aiming parameter. Conservation of angular momentum in a central field allows one to relate the aiming parameter b of a particle approaching the field origin from far away (Fig. P1.9) to the distance of closest approach r_1 to the origin. (The *aiming parameter* is the distance from the field origin to the straight line of particle motion in the absence of interaction.) At the instant of closest approach to the origin, the particle velocity is perpendicular to the radius vector, so the law of conservation of angular momentum can be written as

$$mv_0 b = mv_1 r_1,$$

where v_0 is the initial speed of the particle far away from the origin and v_1 is the particle speed at the instant of closest approach. Expressing v_1 and substituting into the law of conservation of energy, $\frac{1}{2}mv_0^2 = \frac{1}{2}mv_1^2 + E_{\rm p}(r_1)$, one can find r_1.

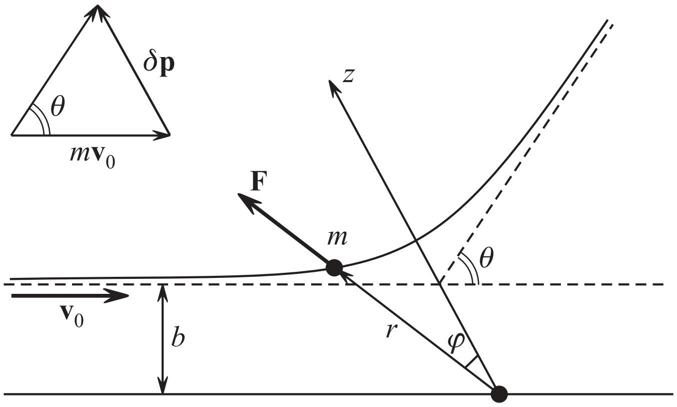

Figure P1.10. To the problem of determining the scattering angle θ in the field $F_r = A/r^2$.

Example 2. Scattering angle. The problem of particle scattering on a Coulomb (or gravitational) center is of significant importance in astrophysics and nuclear physics. Consider a particle of mass m moving toward a (repulsive) force center $F_r = A/r^2$ with the initial kinetic energy $E_\mathrm{k} = \frac{1}{2}mv_0^2$ and aiming parameter b. Find the scattering angle θ.

Solution. From symmetry considerations it is clear that the change in momentum in the whole duration of the motion is directed along the z-axis (see Fig. P1.10) and is equal to

$$|\Delta\mathbf{p}| = \Delta p_z = 2mv_0 \sin(\theta/2).$$

The change in momentum (projected onto the z-axis) in the time dt is

$$dp_z = F_z\,dt = F\cos\varphi\,dt = \frac{A}{r^2}\frac{d\varphi}{v_\varphi/r}\cos\varphi = \frac{mA}{L}\cos\varphi\,d\varphi,$$

where $v_\varphi = r(d\varphi/dt)$ is the azimuthal (perpendicular to the radius vector) velocity component (see Fig. P1.11). The angular momentum of the particle is expressed in terms of v_φ as $L = mv_\varphi r$. Integrating with respect to φ from $-(\pi-\theta)/2$ to $(\pi-\theta)/2$ yields

$$\Delta p_z = \frac{2mA}{L}\cos(\theta/2).$$

Comparing the two expressions of Δp_z and taking into account that $L = mv_0 b$, we find that

$$\tan\frac{\theta}{2} = \frac{A}{2E_\mathrm{k}b}.$$

This formula will be used subsequently in the discussion of Rutherford's experiments.

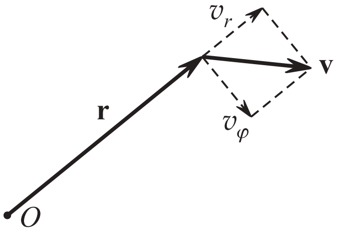

Figure P1.11. Radial and azimuthal components of velocity.

► **Effective potential energy.** With the laws of conservation of energy and angular momentum, one can establish the time dependence of the central distance r by reducing the problem to that of one-dimensional motion. To this end, the particle velocity should be resolved into two mutually perpendicular components (Fig. P1.11), a radial one, $v_r = dr/dt$, and an azimuthal one, v_φ. The angular momentum is expressed in terms of the azimuthal speed: $L = mrv_\varphi$. Hence, the mechanical energy of the particle can be written as

$$E = \frac{m(v_r^2 + v_\varphi^2)}{2} + E_{\rm p}(r) = \frac{m}{2}\left(\frac{dr}{dt}\right)^2 + \left[\frac{L^2}{2mr^2} + E_{\rm p}(r)\right].$$

It is apparent that the dependence $r(t)$ is the same as in a one-dimensional motion with the effective potential energy

$$E_{\rm p}^{\rm eff} = E_{\rm p} + \frac{L^2}{2mr^2}.$$

The second term is sometimes called the *centrifugal energy*. The graph of the effective potential curve is determined by the value of L, which can be evaluated from the initial conditions.

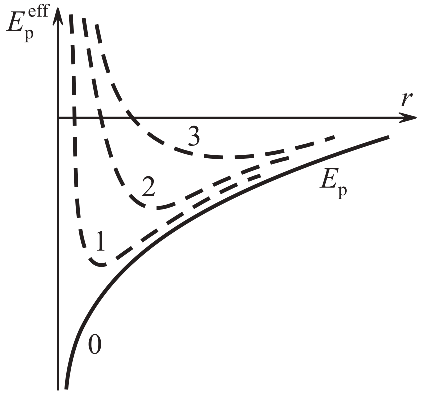

Figure P1.12. The form of the effective potential curve is determined by the angular momentum magnitude L.

Figure P1.12 depicts qualitatively the dependence of the effective potential energy $E_{\rm p}^{\rm eff}$ on the distance r for various L ($L_0 = 0$ and $L_1 < L_2 < L_3$) in the case that $E_{\rm p} = -b/r$. It is apparent that, for any L, the condition of finiteness of the motion remains the same: $E < 0$.

P1.6.3. Conservation of Angular Momentum for a System of Particles

► **Angular momentum of a system of particles.** The angular momentum of a system of particles about a point O is the vector sum of the angular momenta of the constituting particles. It is equal to

$$\mathbf{L} = \mathbf{r}_{\rm cm} \times \mathbf{p} + \mathbf{L}_{\rm rel}, \qquad \text{(P1.6.3.1)}$$

where $\mathbf{L}_{\rm rel}$ is the angular momentum in a center-of-mass frame. One of the consequences of equation (P1.6.3.1) is that if the linear momentum of the system is zero, the angular momentum $\mathbf{L}$ does not depend on the choice of the point O. Summing up equation (P1.6.2.2) over all point of the system and taking into account that the net moment of internal forces is zero, we find that

$$\frac{d\mathbf{L}}{dt} = \mathbf{M}^{\rm e}, \qquad \text{(P1.6.3.2)}$$

or *the time derivative of the angular momentum of the system equals the net moment of external forces.*

► **Law of conservation of angular momentum.** It follows from equation (P1.6.3.2) that *the angular momentum of a closed system is conserved.* The law of conservation of angular momentum is a fundamental law based on the postulate of isotropy of space (there is no selected direction in space). Just as with the laws of conservation of linear momentum and energy, the application of the law of conservation of angular momentum goes beyond the scope of Newtonian mechanics, where it was derived.

The angular momentum of a *unclosed* system is conserved in the following cases:

1. If the net moment of external forces $\mathbf{M}^{\mathrm{e}}$ is zero. (Example: a system of interacting particles in an external central field.)

2. If the moment of external forces about an axis is zero, then the angular momentum about this axis is conserved.

Example. If a weight of mass m attached to one end of a weightless string moves in a horizontal circle, the other end of the string fixed, then the angular momentum about any point at the axis of rotation (except for the center of the circle) is not conserved. However, the angular momentum about the axis of rotation is conserved, since the moment of the force of gravity and that of the tension force about this axis are both zero.

3. If the external forces are bounded, then the change in angular momentum $\Delta\mathbf{L}=\mathbf{M}^{\mathrm{e}}\Delta t$ in the short time of impact Δt can be neglected.

P1.6.4. Two-Body Problem. Reduced Mass

► **Reduced mass.** Consider a closed system consisting of two interacting particles. Solving the problem of motion of these particles (known as the *two-body problem*) means determining the particle positions at any time for given initial conditions. The positions of the particles are expressed in terms of the position of their center of mass $\mathbf{r}_{\mathrm{cm}}(t)$ (see (P1.4.1.1)) and their relative displacement $\mathbf{r}_{12} = \mathbf{r}_1 - \mathbf{r}_2$ as follows:

$$\mathbf{r}_1 = \mathbf{r}_{\mathrm{cm}} + \frac{m_2}{m_1 + m_2}\mathbf{r}_{12}, \qquad \mathbf{r}_2 = \mathbf{r}_{\mathrm{cm}} - \frac{m_1}{m_1 + m_2}\mathbf{r}_{12}.$$

The center of mass moves with a constant velocity, with the initial position and velocity found from the initial conditions (see equations (P1.4.1.1) and (P1.4.1.2)). Thus, solving the two-body problem is reduced to determining $\mathbf{r}_{12}$.

From the conditions of uniformity and isotropy of space and uniformity of time, it follows that the particles interact with central forces such that $\mathbf{F}_{12} = -\mathbf{F}_{21} = \mathbf{F}$, where $\mathbf{F}$ is parallel to $\mathbf{r}_{12}$ and the magnitude F is dependent on $|\mathbf{r}_{12}|$ only. Let us write the equations of motion for each particle:

$$\mathbf{F} = m_1\frac{d^2\mathbf{r}_1}{dt^2}, \qquad -\mathbf{F} = m_2\frac{d^2\mathbf{r}_2}{dt^2}.$$

Subtracting the second equation divided by m_2 from the first equation divided by m_1, we get

$$\left(\frac{1}{m_1} + \frac{1}{m_2}\right)\mathbf{F} = \frac{d^2\mathbf{r}_{12}}{dt^2} \qquad \text{or} \qquad \mathbf{F} = \mu\frac{d^2\mathbf{r}_{12}}{dt^2},$$

where μ is the reduced mass determined by

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2}, \qquad \text{so} \qquad \mu = \frac{m_1 m_2}{m_1 + m_2}. \tag{P1.6.4.1}$$

It is apparent that the vector $\mathbf{r}_{12}$ is determined by solving the problem of motion of a particle of mass μ in a central field.

► **Linear momenta and kinetic energies of two bodies in a center-of-mass frame.** The particle momenta in a center-of-mass frame, $\widetilde{\mathbf{p}}_1$ and $\widetilde{\mathbf{p}}_2$, are equal in magnitude (the total momentum is zero, with $\widetilde{\mathbf{p}}_2 = -\widetilde{\mathbf{p}}_1$ and $|\widetilde{\mathbf{p}}_1| = |\widetilde{\mathbf{p}}_2| = \widetilde{p}$) and are expressed in terms of the particle relative velocity (independent of the reference frame) as

$$|\mathbf{v}_{\text{rel}}| = \left|\left(\frac{\widetilde{\mathbf{p}}_1}{m_1} - \frac{\widetilde{\mathbf{p}}_2}{m_2}\right)\right| = \frac{\widetilde{p}}{\mu},$$

The kinetic energy of the system in the center-of-mass frame can also be expressed in terms of the relative velocity as

$$\widetilde{E}_{\text{k}} = \frac{\widetilde{p}^2}{2m_1} + \frac{\widetilde{p}^2}{2m_2} = \frac{\widetilde{p}^2}{2\mu} = \frac{\mu v_{\text{rel}}^2}{2}.$$

This formula was already encountered in considering inelastic collision (example 2, Subsection P1.5.3).

P1.7. Gravitational Field

P1.7.1. Law of Universal Gravitation. Superposition Principle

► **Newton's law of universal gravitation.** Two particles of masses m_1 and m_2 separated by a distance r are pulled to each other with a force of gravity (gravitational force) equal to

$$F = G\frac{m_1 m_2}{r^2}, \qquad \text{(P1.7.1.1)}$$

where $G \approx 6.673 \times 10^{-11}\,\text{N}\,\text{m}^2/\text{kg}^2$ is the gravitational constant. The gravitational force is a central force; it acts along the line connecting the particles.

The force exerted on a point mass m in a central field of gravity (gravitational field) produced by a fixed mass M can be represented as (see Subsection P1.3.3)

$$\mathbf{F} = -G\frac{mM}{r^2}\frac{\mathbf{r}}{r} \qquad \text{or} \qquad F_r(r) = -G\frac{mM}{r^2}. \qquad \text{(P1.7.1.2)}$$

Thus, the role of the "charge" for a gravitational field is played by the inertial mass m. This property is sometimes formulated as the *equivalence of gravitational and inertial masses*. The potential energy of a particle in a central field of gravity can be found by using the relation between force and potential energy (formula (P1.5.2.4)); we have $-GmM/r^2 = -dE_{\text{p}}/dr$, so $E_{\text{p}} = -GmM/r + \text{const}$. The constant is usually set to zero, thus taking the potential energy to be zero at infinity:

$$E_{\text{p}} = -G\frac{mM}{r}. \qquad \text{(P1.7.1.3)}$$

► **Principle of superposition.** If a field of gravity is generated by several point masses, $M_1, \ldots, M_N$, then the force acting on a particle m and its potential energy are expressed as

$$\begin{aligned} \mathbf{F} &= \sum \mathbf{F}_j = -\sum G\frac{mM_j}{(\mathbf{r}-\mathbf{r}_j)^2}\frac{\mathbf{r}-\mathbf{r}_j}{|\mathbf{r}-\mathbf{r}_j|}, \\ E_{\text{p}} &= \sum E_{\text{p}j} = -\sum G\frac{mM_j}{|\mathbf{r}-\mathbf{r}_j|}, \end{aligned} \qquad \text{(P1.7.1.4)}$$

where $\mathbf{r}$ is the position vector of the particle m and $\mathbf{r}_j$ is the position vector of the mass point M_j. If a gravitational field is generated by a continuously distributed mass, then the summation in (P1.7.1.4) must be replaced by integration.

P1.7.2. Strength and Potential of Gravitational Field

► **Strength and potential of the gravitational field generated by a point mass.** It is apparent from equations (P1.7.1.4) that both the force acting on a particle of mass m in a gravitational field and the potential energy of the particle are proportional to m. Hence, the specific force, $\mathbf{F}/m$, and specific energy, $E_{\rm p}/m$, are independent of m and so characterize the field itself. These are called the *strength* and *potential* of the gravitational field, respectively, and denoted

$$\mathbf{g}=\frac{\mathbf{F}}{m},\qquad \varphi=\frac{E_{\rm p}}{m}.$$

The field strength has a simple physical meaning: it is the *acceleration due to gravity*, or the acceleration acquired by any particle placed at a given point due to the action of the gravitational field only. The strength and potential created by a point mass M are expressed as

$$\mathbf{g}=-G\frac{M}{r^2}\frac{\mathbf{r}}{r},\qquad \varphi=-G\frac{M}{r}.\tag{P1.7.2.1}$$

The strength and potential of the gravitation field generated by several point masses are calculated using the principle of superposition. For example, let us write equations (P1.7.2.1) for the case of a body with distributed mass:

$$\mathbf{g}(\mathbf{r})=-\int G\frac{\rho(\mathbf{r}')}{(\mathbf{r}-\mathbf{r}')^2}\,\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}\,dV',\qquad \varphi(\mathbf{r})=-\int G\frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,dV',\tag{P1.7.2.2}$$

where $\rho(\mathbf{r})$ is the mass density at the position $\mathbf{r}$ and the integration is performed over the volume occupied by the body.

Example 1. Show that the gravitational strength inside a hollow spherical layer is zero.

Solution. For a given point A, consider the contributions to the field strength of the small regions B and C cut out from the sphere by a thin cone with apex at A (Fig. P1.13). The ratio of the areas of these regions, and hence the ratio of their masses, equals the ratio of the squared distances of the regions to the point A. It follows that the strengths produced by the regions B and C at A are equal in magnitude.

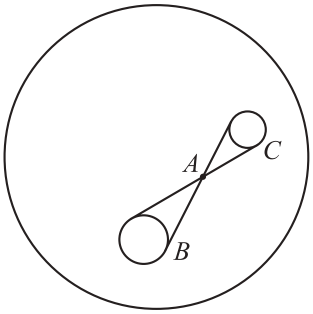

Figure P1.13. The strengths of the gravitational fields generated by opposite regions on a spherical surface cancel each other out.

The field strength produced by a thin hollow sphere of mass M outside the sphere turns out to be equal to the strength produced by a point mass M placed at the sphere center. Cumbersome integration is required to prove this result (Newton was first to do this). In Chapter P3, this statement will be proved using Gauss's theorem. The same result holds true for any spherically symmetric distributed mass and, in particular, for any spherical planet.

Example 2. Suppose a mass M is uniformly distributed along a segment of length l. Compute the gravitational strength and potential at the point lying outside on the continuation of the segment at a distance x from its center.

Solution. The mass of a segment of length dy equals $dm = M\,dy/l$. Integrating yields

$$g(x) = -G\int_{x-l/2}^{x+l/2} \frac{M\,dy}{ly^2} = -G\frac{M}{x^2 - (l/2)^2},$$

$$\varphi(x) = -G\int_{x-l/2}^{x+l/2} \frac{M\,dy}{ly} = -G\frac{M}{l}\ln\frac{x + l/2}{x - l/2}.$$

It is apparent that a symmetric but nonspherical body cannot be replaced by a point mass placed at its center. This example illustrates also that the strength and potential are linked by the relation $g_x = -\partial\varphi/\partial x$, which is similar to (P1.5.2.2).

P1.7.3. Motion in a Central Field of Gravity. Kepler's Laws. Orbital and Escape Velocities

► **Motion in a central field of gravity. Kepler's laws.** Motion in a central field of gravity obeys the general laws of motion in a central field. However, it has some specific features reflected in Kepler's first and third laws initially formulated for planets of the solar system.

Kepler's first law states that a finite motion ($E < 0$) of a particle in a central gravitational field occurs in a closed path—an ellipse with the force center (the sun) located at one of its foci.

Kepler's second law affirms the constancy of the areal velocity, i.e., the rate at which the radius vector of the moving particle sweeps area. This law holds for any central field and follows directly from the law of conservation of momentum (see Subsection P1.6.2).

Kepler's third law states that, for any two particles moving in a central gravitational field, the ratio of the squares of their orbital periods equals the ratio of the cubes of the semimajor axes of their orbits: $T_1^2/T_2^2 = a_1^3/a_2^3$.

Kepler's first law should be supplemented with the statement that an infinite motion of a particle in a central gravitational field occurs in either a parabola ($E = 0$) or a hyperbola ($E > 0$). The third law can be supplemented by the relationship between the specific energy of orbital motion and the semimajor axis: $|E|/m = GM/(2a)$. It is apparent that the orbital period is uniquely determined by the specific energy of the moving particle.

► **Orbital and escape velocities.** The *orbital velocity* (also known as the *first cosmic velocity*) of a small body around a planet is the speed at which the body moves in a circular orbit around the planet near its surface, with the atmospheric drag neglected. The orbital speed is determined from the equation $mg = mv_{\rm o}^2/R$ and equals $v_{\rm o} = \sqrt{gR} = \sqrt{GM/R}$, where M is the mass of the planet. (For Earth, $v_{\rm o} \approx 7.9\,\text{km/s}$.)

The *escape velocity* (also known as the *second cosmic velocity*) of a body is the minimum speed the body must have to "break free" from a planet's gravitational pull and go to infinity. As follows from Subsection P1.6.2, the condition for a motion to be infinite is the inequality $E \ge 0$; hence, the escape velocity is determined by the equation $mv_{\rm e}^2/2 - GmM/R = 0$ and equals $v_{\rm e} = \sqrt{2GM/R} = \sqrt{2gR}$. (For Earth, $v_{\rm e} \approx 11.2\,\text{km/s}$.)

P1.8. Non-Inertial Reference Frames

P1.8.1. Definition of Fictitious Forces

In many cases, it is convenient to solve a dynamics problem in a non-inertial frame of reference (a reference frame that is not inertial) instead of recalculating results first obtained in an inertial frame. To this end, *fictitious forces* (also called *inertial forces*) are introduced as follows. In Newton's second law for a particle of mass m moving under the action of a net force $\mathbf{F}$, the acceleration is first represented as the sum of two components, $\mathbf{a} = \mathbf{a}_{\rm rel} + \mathbf{a}^*$, where $\mathbf{a}_{\rm rel}$ is the acceleration of the particle in the non-inertial frame. Then the other component, multiplied by m, is transferred to the other side of the equation and called the fictitious force. So we have

$$\begin{aligned} \mathbf{F} &= m\mathbf{a}, \\ \mathbf{F} &= m(\mathbf{a}_{\rm rel} + \mathbf{a}^*), \\ \mathbf{F} + (-m\mathbf{a}^*) &= m\mathbf{a}_{\rm rel}, \\ \mathbf{F} + \mathbf{F}_{\rm fict} &= m\mathbf{a}_{\rm rel}. \end{aligned}$$

The fictitious forces are thus defined by

$$\mathbf{F}_{\rm fict} = -m\mathbf{a}^*, \qquad \text{(P1.8.1.1)}$$

where the acceleration $\mathbf{a}^*$ is determined by the kinematic equation

$$\mathbf{a} = \mathbf{a}_{\rm rel} + \mathbf{a}^* \qquad \text{(P1.8.1.2)}$$

and is dependent on the non-inertial frame itself and the position and velocity of the particle in it. Consider some specific cases.

P1.8.2. Translationally Moving Non-Inertial Frames

In this case, $\mathbf{a}^*$ equals the acceleration of the inertial frame K (see formula (P1.1.4.1)). Then the fictitious force is expressed as

$$\mathbf{F}_{\rm fict} = -m\mathbf{a}_K.$$

It is apparent that the fictitious force here is equivalent to the force of gravity. So it is convenient to combine the two forces together in solving problems. With this done, the introduction of a fictitious force turns out to be equivalent to adjusting the acceleration due to gravity (see Subsection P1.7.2), so that $\mathbf{g}$ is effectively replaced by $\mathbf{g}^* = \mathbf{g} + \mathbf{a}^* = \mathbf{g} - \mathbf{a}_K$.

Example. A container filled with a liquid moves with a constant horizontal acceleration $\mathbf{a}$. Find the angle β between surface of the liquid and the horizontal plane.

Solution. In a reference frame associated with the container, the liquid surface is at rest and "horizontal," which means that it is perpendicular to the vector $\mathbf{g}^* = \mathbf{g} + \mathbf{a}^* = \mathbf{g} - \mathbf{a}$. So we need to find the angle between $\mathbf{g}$ and $\mathbf{g} - \mathbf{a}$. We have $\tan\beta = a/g$.

The principle of equivalence of inertial and gravitational forces was laid by Einstein at the basis of his *general theory of relativity*, which is a relativistic theory of gravitation and explains the appearance of gravitational forces by the fact that the space–time continuum is curved by external massive bodies.

P1.8.3. Uniformly Rotating Reference Frame

In this case the kinematic equation for the acceleration (P1.8.1.2) becomes: $\mathbf{a} = \mathbf{a}_{\rm rel} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) + 2\boldsymbol{\omega} \times \mathbf{v}_{\rm rel}$ (see Subsection P1.2.2). The second term is associated with the rotation of the translatory velocity $\boldsymbol{\omega} \times \mathbf{r}$ together with the reference frame. This vector is directed toward the axis of rotation and its magnitude equals $\omega^2 R$ (R is the distance to the axis), and hence it is the normal (centripetal) acceleration of the reference frame at the given point.

The third term (Coriolis acceleration) is due to, first, the rotation of $\mathbf{v}_{\text{rel}}$ together with the reference frame and, second, the change in the translatory velocity $\boldsymbol{\omega}\times\mathbf{r}$ due to the movement of the particle from one point of the rotating non-inertial frame to another. Accordingly, the fictitious force (P1.8.1.1) is the sum of two components, the first called the *centrifugal force* and the second called the *Coriolis force*:

$$\mathbf{F}_{\text{fict}} = \mathbf{F}_{\text{cf}} + \mathbf{F}_{\text{Cor}} = m\omega^2\mathbf{R} + 2m\mathbf{v}_{\text{rel}}\times\boldsymbol{\omega}. \tag{P1.8.3.1}$$

The centrifugal fictitious force is directed outward from the axis of rotation ($\mathbf{R}$ is directed from the axis and is perpendicular to it). Since it is independent of the particle velocity, its action is indistinguishable from a (nonuniform) gravitational field. For example, the force of gravity measured at the Earth's surface is the sum of the gravitational force and the centrifugal inertial force.

The Coriolis force is perpendicular to the particle velocity. For example, in the Northern hemisphere, for motions along the Earth's surface the horizontal component of the Coriolis force is directed to the right of the motion, which exhibits itself in the formation of cyclones (rotate counterclockwise), greater erosion of right river banks, etc.

Example. Find the deflecting action of the Coriolis force onto a body freely falling on the Earth's surface from a height h at the equator.

Solution. Since the deviation is small, the "undisturbed" velocity of free falling $\mathbf{v} = \mathbf{g}t$ can be substituted in the Coriolis force (P1.8.3.1). The vector $\boldsymbol{\omega}$ is perpendicular to $\mathbf{v}$, and hence the Coriolis force is directed eastward, its magnitude equal $2mv\omega$, and the horizontal acceleration imparted to the body is $a_{\text{east}} = 2\omega gt$. Integrating yields the horizontal speed $v_{\text{east}} = g\omega t^2$ and the horizontal displacement $s_{\text{east}} = \frac{1}{3}g\omega t^3$. Substituting the time of falling $t = \sqrt{2h/g}$ gives the final deviation. For example, if the initial height is $h = 300$ m, the final deviation is approximately 10 cm.

P1.9. Dynamics of Rigid Body

P1.9.1. Rotation About a Fixed Axis

For a rigid body consisting of many particles of mass m_j and rotating about a fixed axis z with an angular velocity ω, the magnitude of its angular momentum about the axis (see Subsection P1.6.2) equals

$$|L_z| = \sum_j R_j m_j v_j = \omega\sum_j m_j R_j^2 = I_z\omega, \tag{P1.9.1.1}$$

where R_j is the distance from the particle m_j to the axis; the relation $v_j = \omega R_j$ has been used here. The angular momentum direction coincides with that of $\boldsymbol{\omega}$ (determined by the right hand rule). The quantity

$$I_z = \sum_j m_j R_j^2 = \int R^2\,dm \tag{P1.9.1.2}$$

is called the *moment of inertia* of the body about the axis z. Differentiating (P1.9.1.1) with respect to time and taking into account that $dL_z/dt = M_z$, where M_z is the moment of external forces about the axis of rotation (see equation (P1.6.3.2)), we get

$$M_z = I_z\varepsilon, \tag{P1.9.1.3}$$

where $\varepsilon = d\omega/dt$ is the angular acceleration. This equation is referred to as the *basic equation of rotational dynamics of a rigid body about a fixed axis*. Compute also the kinetic energy of the rotating body,

$$E_\text{k} = \sum_j \frac{m_j v_j^2}{2} = \tfrac{1}{2}\omega^2 \sum_j m_j R_j^2 = \tfrac{1}{2} I_z \omega^2, \tag{P1.9.1.4}$$

and the work done by an external force **F** on rotating the body,

$$A = \int d\mathbf{r} \cdot \mathbf{F} = \int (\boldsymbol{\omega}\, dt \times \mathbf{r}) \cdot \mathbf{F} = \int (\boldsymbol{\omega}\, dt) \cdot (\mathbf{r} \times \mathbf{F}) = \int \mathbf{M} \cdot d\boldsymbol{\varphi},$$

where $d\boldsymbol{\varphi} = \boldsymbol{\omega}\, dt$.

Example. A body of mass m is attached to one end of a weightless rope wound around a pulley of radius R and moment of inertia I. Find the acceleration of the body after it has been released.

Solution. Let us write Newton's second law for the body: $mg - T = ma$. The equation of motion of the pulley (P1.9.1.3) is $TR = I\varepsilon$. Using the kinematic relation $a = \varepsilon R$ (the rope does not slide against the pulley) and solving the above equations for a, we obtain $a = g/(1 + I/mR^2)$.

P1.9.2. Properties of Moment of Inertia

The moment of inertia (P1.9.1.2) is a scalar quantity that characterizes the distribution mass of the body relative to the axis of rotation. It can be seen from equations (P1.9.1.3) and (P1.9.1.4) that the moment of inertia is a measure of the body's inertia with respect to rotational motion; moment of inertia plays the same role as mass in translational motion.

Example. Compute the moment of inertia of a thin disk of mass m and radius R about its axis of symmetry.

Solution. Dividing the disk into thin circular rings and integrating, we obtain

$$I = \int r^2\, dm = \int_0^R r^2 \Bigl(\frac{m}{\pi R^2}\Bigr) 2\pi r\, dr = \frac{mR^2}{2}.$$

The same result holds also for a homogeneous solid cylinder.

► **Parallel axis theorem.** The *parallel axis theorem* (also known as the *Huygens–Steiner theorem*) relates the moment of inertia I about an arbitrary axis to the moment of inertia I_0 about a parallel axis passing through the center of mass of the body:

$$I = I_0 + ma^2, \tag{P1.9.2.1}$$

where m is mass of the body and a is the distance between the axes. For example, the moment of inertia of a disk about an axis perpendicular to its plane and passing through its edge equals $\frac{1}{2}mR^2 + mR^2 = \frac{3}{2}mR^2$. The minimum moment of inertia among all parallel axes is for the axis passing the center of mass.

► **Perpendicular axis theorem.** The *perpendicular axis theorem* asserts that the moment of inertia of a plane body about an arbitrary axis z perpendicular to the body plane equals the sum of the moments of inertia about mutually perpendicular axes x and y lying in the body plane and intersecting z:

$$I_z = I_x + I_y.$$

For example, the moment of inertia of a thin disk about an axis of symmetry lying in its plane equals $I_x = I_y = \frac{1}{2}I_z = \frac{1}{4}mR^2$.

► **Moments of inertia of some bodies.** Listed below are moments of inertia of some bodies of various shape.

1) Thin ring of radius R (about its axis of symmetry): $I = mR^2$. The same result holds for a thin hollow cylinder (without end caps).

2) Thin rod of length l (about an axis through its center and perpendicular to its axis): $I = \frac{1}{12}ml^2$. A thin rectangular plate has the same moment of inertia about the axis through the centers of its opposite sides having the length l. The moment of inertia about a parallel axis through one of the ends of the rod equals $I = \frac{1}{3}ml^2$.

3) A thin rectangular plate (about the axis through its center and perpendicular to its plane): $I = \frac{1}{12}m(a^2 + b^2)$. A rectangular parallelepiped has the same moment of inertia about the axis through the centers of its opposite faces.

4) Thin spherical shell (about its axis of symmetry): $I = \frac{2}{3}mR^2$.

5) Homogeneous solid ball (about its axis of symmetry): $I = \frac{2}{5}mR^2$.

6) Hollow cylinder with inner radius R_1 and outer radius R_2: $I = \frac{1}{2}m(R_1^2 + R_2^2)$.

P1.9.3. Plane Motion of a Rigid Body

Plane motion of a rigid body is the superposition of translational motion of its center of mass and rotational motion in a center-of-mass frame (see Subsection P1.4.2). The motion of the center of mass is described by Newton's second law and determined by the net external force (equation (P1.4.2.1)), where only real external forces must be taken into account, since the moment of fictitious forces relative to the center of mass is zero (just like in the case of the forces of gravity in the example of Subsection P1.6.1). The kinetic energy of plane motion equals $E_k = \frac{1}{2}mv_{cm}^2 + \frac{1}{2}I_0\omega^2$ (see equation (P1.5.1.3)). The angular momentum about a fixed axis perpendicular to the plane of motion is calculated as (see equation (P1.6.1.3)): $L_z = \pm mv_{cm}d \pm I_0\omega$, where d is the arm of the center-of-mass velocity relative to the axis, and the plus or minus sign is determined by the positive direction of rotation selected.

Example 1. Find the acceleration of a round body that rolls down an inclined plane without slipping (Fig. P1.14). The radius of the body is R, its mass is m, and the moment of inertia about the central axis is I. The angle between the inclined plane and the horizontal surface is α.

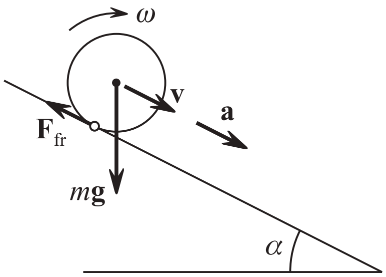

Figure P1.14. Forces acting on a body rolling down an inclined plane.

Solution. The equations of motion are: $mg\sin\alpha - F_{fr} = ma$ and $F_{fr}R = I\varepsilon$. The no-slipping condition ($v = \omega R$) leads to the equation $a = \varepsilon R$ (see the example in Subsection P1.2.1). Solving the above equations gives the acceleration $a = g\sin\alpha/(1 + I/mR^2)$. Since the static friction force does not produce any work, the mechanical energy is conserved.

Example 2. A thin rod of length l and mass M rests on a smooth horizontal plane. A ball of mass m moves in the plane and undergoes an elastic collision with an end of the rod perpendicularly to the line of the rod. The final speed of the ball, u, and the rod center, V, as well as angular velocity of the rod, ω, after the collision are determined from the following three equations:

1) law of conservation of linear momentum: $mv = mu + MV$,

2) law of conservation of energy: $mv^2/2 = mu^2/2 + MV^2/2 + (Ml^2/12)\omega^2/2$,
3) law of conservation of angular momentum (e.g., about the point of impact): $0 = MV(l/2) - (Ml^2/12)\omega$.

P1.9.4. Motion with a Fixed Point

For a rigid body rotating about a fixed point, its angular velocity is directed along the instantaneous axis of rotation and changes its direction both in space and relative to the body itself. The equation of motion

$$\frac{d\mathbf{L}}{dt} = \mathbf{M}^{\mathrm{e}},$$

which is referred to as the *basic equation of motion of a rigid body with a fixed point*, allows us to find out how the angular momentum $\mathbf{L}$ changes in time. Since $\mathbf{L}$ is generally non-parallel to the angular velocity $\boldsymbol{\omega}$, we need an equation that relates $\mathbf{L}$ and $\boldsymbol{\omega}$ in order to close the system of equations of motion.

Example. As an inclined dumbbell rotates about a vertical axis (as shown in Fig. P1.15), its angular momentum $\mathbf{L} = \mathbf{r}_1 \times \mathbf{p}_1 + \mathbf{r}_2 \times \mathbf{p}_2$ is perpendicular to the straight line connecting the masses m_1 and m_2 of the dumbbell and directed at a non-right angle to the axis of rotation. Also the vector $\mathbf{L}$ itself rotates about the vertical axis with the angular velocity $\boldsymbol{\omega}$.

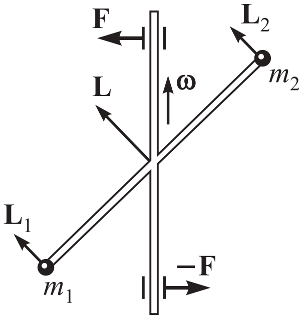

Figure P1.15. The vector of angular momentum $\mathbf{L}$ can be nonparallel to the axis of rotation (to $\boldsymbol{\omega}$).

► **Theorem of principal axes of inertia.** The relationship between $\mathbf{L}$ and $\boldsymbol{\omega}$ can be found using the *theorem of principal axes of inertia* which states that for any rigid body and any point O there are three mutually perpendicular axes through O such that, when the body rotates about any of these axes, the vector $\mathbf{L}$ is parallel to the axis: $\mathbf{L}_i = I_i \boldsymbol{\omega}_i$ $(i = 1, 2, 3)$. The moments of inertia about these axes are called *principal moments of inertia*. If the rotation occurs about an arbitrary axis through O, the angular velocity can be resolved into its projections onto the principal axes, $\boldsymbol{\omega} = \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2 + \boldsymbol{\omega}_3$, and the angular momentum can be expressed as

$$\mathbf{L} = I_1 \boldsymbol{\omega}_1 + I_2 \boldsymbol{\omega}_2 + I_3 \boldsymbol{\omega}_3. \qquad \text{(P1.9.4.1)}$$

(Likewise, $\mathbf{L}$ can be resolved into its projections onto the principal axes and then $\boldsymbol{\omega}$ can be expressed in terms of them.)

If the principal axes are through the center of mass (center of inertia) of the body, they are called *free axes*. When the body rotates about any of the free axes (free rotation), both the linear momentum and the angular momentum of the body are conserved, which means that there is no need to apply an external force or external moment to maintain rotation. (In the above example, the net force is zero; however, there are a couple of forces arising at points of support whose moments provide change of $\mathbf{L}$ with time.) It should be noted that free rotation is stable only about two of the three axes, the ones with the minimum and maximum principal moments of inertia.

Example 1. When is the choice of the free axes unique?

Solution. The free axes are uniquely determined only if the three principal moments of inertia are all different. If two moments are equal (e.g., $I_1 = I_2 = I$), then any axis that lies in the plane of the two free axes is also a free axis, with $\mathbf{L} = I\boldsymbol{\omega}_1 + I\boldsymbol{\omega}_2 = I\boldsymbol{\omega}$. If all three principal moments of inertia are equal (e.g., such is a homogeneous cube), then any axis through the center of mass is a free axis.

Example 2. Given three principal moments of inertia (I_1, I_2, and I_3) of a body, find the moment of inertia I about the axis that makes angles γ_1, γ_2, and γ_3 with the principal axes.

Solution. The kinetic energy of the body rotating with an angular velocity ω about the specified axis can be calculated in two ways. On the one hand, it is equal to

$$E = \frac{I\omega^2}{2}.$$

On the other hand,

$$E = \frac{L_\omega \omega}{2} = \frac{(\mathbf{L}\cdot\boldsymbol{\omega})}{2} = \frac{I_1\omega_1^2}{2} + \frac{I_2\omega_2^2}{2} + \frac{I_2\omega_2^2}{2},$$

where $L_\omega = I\omega$ is the projection of $\mathbf{L}$ onto $\boldsymbol{\omega}$. Since $\omega_i/\omega = \cos\gamma_i$, we find that

$$I = I_1\cos^2\gamma_1 + I_2\cos^2\gamma_2 + I_3\cos^2\gamma_3.$$

Example 3. Free symmetric top. Nutation of the Earth's axis. Let is find out how the position of the axis of rotation changes relative to a rotating body (in a reference frame associated with the body). Let us confine ourselves to the case of a *free symmetric top*, or a freely rotating body with $I_1 = I_2 = I_\perp$ and $I_3 = I_\| \neq I_\perp$. (The Earth is an example of such a top.) For a symmetric top, the vector $\mathbf{L} = I_\|\boldsymbol{\omega}_\| + I_\perp\boldsymbol{\omega}_\perp$ lies in the same plane as the third axis (axis of symmetry) and the vector $\boldsymbol{\omega} = \boldsymbol{\omega}_\| + \boldsymbol{\omega}_\perp$. In an inertial reference frame, the angular momentum $\mathbf{L}$ is conserved. Consequently, in the rotating top-related reference frame, the vector $\mathbf{L}$ performs rotation with an angular velocity $\boldsymbol{\omega}$ at the current time, and hence it obeys the equation

$$\frac{d\mathbf{L}}{dt} = \boldsymbol{\omega}\times\mathbf{L}.$$

This equation is called *Euler's equation* (usually, a system of three scalar equations is written out which is obtained by projecting this vector equation onto axes 1, 2, and 3 taking into account relation (P1.9.4.1)). The complexity of Euler's equation is that the vector $\boldsymbol{\omega}$ is not constant but linked to $\mathbf{L}$ by the linear relation (P1.9.4.1). For a symmetric top, $\boldsymbol{\omega}\times\mathbf{L}$ is perpendicular to the plane ($\boldsymbol{\omega}$, $\mathbf{L}$); hence, $d\mathbf{L}_\|/dt = I_\| d\boldsymbol{\omega}_\|/dt = \mathbf{0}$ ($\omega_\| = \text{const}$). For $\boldsymbol{\omega}_\perp$, we have the equation

$$I_\perp\frac{d\boldsymbol{\omega}_\perp}{dt} = (\boldsymbol{\omega}_\| + \boldsymbol{\omega}_\perp)\times(I_\|\boldsymbol{\omega}_\| + I_\perp\boldsymbol{\omega}_\perp) = (I_\perp - I_\|)\,\boldsymbol{\omega}_\|\times\boldsymbol{\omega}_\perp$$

or

$$\frac{d\boldsymbol{\omega}}{dt} = \boldsymbol{\Omega}\times\boldsymbol{\omega}, \qquad \text{where} \quad \boldsymbol{\Omega} = \frac{(I_\perp - I_\|)}{I_\perp}\boldsymbol{\omega}_\|$$

This equation shows that in the top-related frame, the axis of rotation performs a circular motion about axis 3 with the angular velocity $\boldsymbol{\Omega}$. For the Earth, there is a small relative difference between the moments of inertia $I_\perp$ and $I_\|$, which leads to a slow periodic motion of the Earth's axis (*nutation of the Earth's axis*) with a period of tens of days.

► **Gyroscopes.** A gyroscope is a rigid body that rapidly spins about its axis of symmetry. The problem of motion of the gyroscope's axis can be solved in the so-called *gyroscope approximation*, $\mathbf{L} = I\boldsymbol{\omega}$, where both $\mathbf{L}$ and $\boldsymbol{\omega}$ are directed along the axis of symmetry. A perfectly balanced gyroscope is a gyroscope whose center of mass is fixed. It possesses the property of being *inertialess*: its axis stops moving as soon as the external action is removed ($\mathbf{M} = \mathbf{0}$). This property allows to use such gyroscopes to preserve orientation in space. A *heavy gyroscope* is a gyroscope whose center of mass is displaced by a distance d from the fixed point (Fig. P1.16). The moment of the force of gravity acting on a heavy gyroscope is perpendicular to $\mathbf{L}$. Since $d\mathbf{L}/dt \perp \mathbf{L}$, the angular momentum $\mathbf{L}$ and the gyroscope's axis perform regular rotation about the vertical axis (*precession of a gyroscope*). The end of the vector $\mathbf{L}$ moves in a horizontal circle of radius $L\sin\alpha$ with the angular velocity

$$\Omega = \frac{|d\mathbf{L}/dt|}{L\sin\alpha} = \frac{mgd\sin\alpha}{I\omega\sin\alpha} = \frac{mgd}{I\omega}.$$

The angular velocity of precession is independent of the angle α made by the axis of rotation with the vertical.

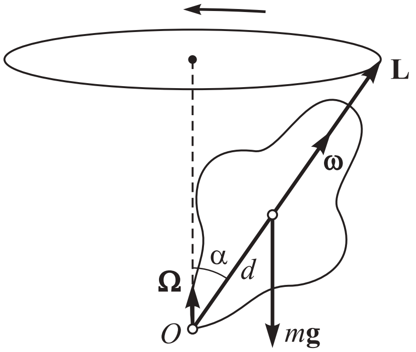

Figure P1.16. Gyroscope precession in the field of gravity.

P1.10. Special Theory of Relativity

P1.10.1. Basics of the Special Theory of Relativity

Einstein's *special theory of relativity* (STR), also known as *special relativity* (SR), extends the limits of classical Newtonian physics, valid for *nonrelativistic velocities* (small compared to the speed of light c), to any velocities, including *relativistic* (comparable with c). In the classical limit $v/c \to 0$, all results of special relativity reduce to results of classical, nonrelativistic physics.

► **Postulates of STR.** The special theory of relativity relies on two postulates:

The *principle of relativity*: all physical laws, both mechanical and electromagnetic, have the same form in all inertial reference frames. This implies that no inertial reference frame can be distinguished from any other by any experiments and taken to be the rest inertial frame. This postulate is an extension of the Galilean relativity principle (see Subsection P1.3.4) to electromagnetic phenomena.

The *principle of invariant light speed*: the speed of light in vacuum is the same for any inertial frame and equals $c \approx 3 \times 10^8$ m/s. This postulate contains two statements: (i) the speed of light is independent of the state of motion of the light source and (ii) the speed of light is independent of which inertial frame is the observer measuring this speed in (i.e., independent of the speed of the receiver).

The invariance of the light speed and its independence of the state of motion of the light source follow from Maxwell's equations of electromagnetic field. It seemed obvious that this statement could hold in only one reference frame. From the classical conceptions of space and time, any other observer moving with a speed v relative to this frame should measure the speed of light as $c+v$ if moving toward the source or $c-v$ if moving away from the source. This result would mean that Maxwell's equations hold in only one reference frame filled with quiescent "ether," through which electromagnetic waves travel. However, the attempt to detect a change in the light speed due to Earth's motion relative to the ether (Michelson–Morley experiment) failed to detect such a change. Einstein put forward the hypothesis that Maxwell's equations, as well as all physics laws, must have the same form in all inertial frames, and hence the light speed must be the same in all inertial frames. This hypothesis led to revision of the basic concepts of space and time.

► **Lorentz transformations.** The Lorentz transformations link the coordinates and time of an event measured in two inertial frames, one of which (K') moves relative to the other (K) with a constant velocity **V**. If the coordinate axes and the initial instant are chosen in the same way as in the Galilean transformations (see formulas (P1.3.4.1)), the Lorentz transformations have the form

$$x' = \frac{x - Vt}{\sqrt{1 - V^2/c^2}}, \quad y' = y, \quad z' = z, \quad t' = \frac{t - (V/c^2)x}{\sqrt{1 - V^2/c^2}}. \qquad \text{(P1.10.1.1)}$$

It is often convenient to use the transformations for the differences of the coordinates and time instants of two events:

$$\Delta x' = \gamma(\Delta x - V\Delta t), \quad \Delta y' = \Delta y, \quad \Delta z' = \Delta z, \quad \Delta t' = \gamma\left[\Delta t - (V/c^2)\Delta x\right]. \quad \text{(P1.10.1.2)}$$

where the notation

$$\gamma = \frac{1}{\sqrt{1 - V^2/c^2}}, \qquad \gamma \ge 1, \qquad \text{(P1.10.1.3)}$$

is used for brevity. The Lorentz transformations reduce to the Galilean transformations for $V \ll c$. The Lorentz transformations are derived from the second pustulate of STR (invariance of light speed) and the requirement of linearity of the transformations (this requirement expresses the condition of uniformity of space). The *inverse transformations*, expressing the coordinates and time in K via those in K', can be obtained from (P1.10.1.1) and (P1.10.1.2) by substituting $-V$ for V:

$$\Delta x = \gamma(\Delta x' + V\Delta t'), \quad \Delta y = \Delta y', \quad \Delta z = \Delta z', \quad \Delta t = \gamma\left[\Delta t' + (V/c^2)\Delta x'\right]. \quad \text{(P1.10.1.4)}$$

P1.10.2. Consequences of the Lorentz Transformations. STR Kinematics

► **Length contraction.** The length of a moving segment is defined as the distance between the points at which the segment ends were at the same time (i.e., $\Delta t = 0$). Consider a rigid body that moves translationally with a speed v and let K' denote a reference frame associated with it. From equation (P1.10.1.2), where one should set $V = v$ and $\Delta t = 0$, one finds that the longitudinal dimension of a moving body is reduced:

$$l = l_0/\gamma = l_0\sqrt{1 - v^2/c^2}, \qquad \text{(P1.10.2.1)}$$

where l_0 is the proper longitudinal dimension measured in frame K', where the body is at rest. The transverse dimensions of a moving body do not change.

Example 1. A square plate moving at the speed $v = 0.8\,c$ along one of its sides becomes a rectangular with the angle between the diagonals equal to arccot $\sqrt{1 - v^2/c^2} = \operatorname{arccot} 0.6 \approx 59^\circ$.

► **Time dilation.** It can be seen from the Lorentz transformations that time passes differently in different inertial frames. In particular, two events occurring simultaneously ($\Delta t = 0$) but at different points in frame K may not be simultaneous in frame K', so that $\Delta t' = -\gamma V\Delta x/c^2$ can be positive or negative. This effect is known as *relativity of simultaneity*. A clock that moves together with frame K' (i.e., that rests relative to K', with $\Delta x' = 0$) shows the *proper time* of this frame. For an observer in frame K, this clock is slow (*time dilation*). Considering two readings of the moving clock as two events, we find from (P1.10.1.4) that

$$\Delta t = \gamma\Delta\tau = \frac{\Delta\tau}{\sqrt{1 - V^2/c^2}}, \qquad \text{(P1.10.2.2)}$$

where $\Delta\tau = \Delta t'$ is the proper time of the moving clock (to be precise, of the associated reference frame). The fact that all reference frames are indistinguishable manifests itself in that a clock resting relative to frame K appears to the observer in frame K' to lag behind the observer's clock in K'. (Note that, to control the moving clock, the stationary observer in K uses *different* clocks at different instants.)

The *twin paradox* consists in that special relativity predicts a change in the age of two twins, one of which stayed on the Earth and the other traveled in deep space at relativistic speeds and then returned to the Earth (the traveler will be younger). It may seem as though this thought experiment violates the principle of indistinguishability of reference frames. In fact, the Earth's twin stayed all the time in one and the same inertial frame but the astronaut changed his or her frame to return to the Earth (the astronaut-associated frame was not inertial).

Example 2. The mean proper lifetime of the muon (an unstable subatomic particle) is $\tau = 2.2\times10^{-6}$ s, with $c\tau \sim 660$ m (c is the speed of light). Due to time dilation, from the viewpoint of an Earth's observer, a space muon flying with a relativistic speed v with $\gamma \gg 1$ has a mean lifetime of $\gamma\tau$ and travels a distance $c\gamma\tau$ from its birth in upper atmosphere, which allows detecting the muon on the Earth's surface.

► **Velocity addition in special relativity.** If a particle moves with velocity $\mathbf{v}'$ relative to frame K', then its velocity relative to frame K can be found by expressing dx, dy, dz, and dt from (P1.10.1.4) and substituting into $v_x = dx/dt$, $v_y = dy/dt$, and $v_z = dz/dt$ to obtain

$$v_x = \frac{v'_x + V}{1 + v'_x V/c^2}, \quad v_y = \frac{v'_y\sqrt{1 - V^2/c^2}}{1 + v'_x V/c^2}, \quad v_z = \frac{v'_z\sqrt{1 - V^2/c^2}}{1 + v'_x V/c^2}. \qquad \text{(P1.10.2.3)}$$

For nonrelativistic speeds, $v_x \ll c$ and $V \ll c$, the above formulas reduce to the nonrelativistic velocity-addition law (see equation (P1.1.4.1)). Formulas (P1.10.2.3) have the important property that if V and v' less than c, then v is also less than c. For example, if the particle is first accelerated to $V = 0.9\,c$ relative to the rest frame and then is further accelerated to $v' = 0.9\,c$ in the moving, particle-associated frame, the resulting speed will be $1.8\,c/1.81 < c$ rather than $1.8\,c$. It is apparent that the speed of light cannot be exceeded. The light speed is the maximum possible speed at which interactions can be transferred in nature.

► **Spacetime interval. Casuality.** The Lorentz transformations preserve neither the time interval nor the length of a spatial segment. However, it can be shown that the following quantity is preserved under the Lorentz transformations:

$$s_{12}^2 = (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2 = (c\Delta t)^2 - (\Delta \mathbf{r})^2. \qquad \text{(P1.10.2.4)}$$

The quantity s_{12} is called the *spacetime interval* between events 1 and 2 ($\Delta t = t_2 - t_1$ and $\Delta\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$).

If $s_{12}^2 > 0$, the spacetime interval is called *time-like*. In this case, there is an inertial frame where $\Delta\mathbf{r} = \mathbf{0}$, which means that the events occur at the same place but at different times. Such events can have a cause–effect relationship.

If $s_{12}^2 < 0$, the spacetime interval between events is called *space-like*. In this case, there is an inertial frame in which $\Delta t = 0$, so that the events occur simultaneously at different locations in space. Such events cannot have a causal relationship. The condition $c|\Delta t| < |\Delta\mathbf{r}|$, or $c|t_2 - T_1| < |\mathbf{r}_2 - \mathbf{v}_1|$, means that a light beam emitted at the instant of the earlier event (at t_1) from the point $\mathbf{r}_1$ cannot reach the point $\mathbf{r}_2$ by the instant t_2. Events separated from event 1 by time-like intervals represent its *absolute past* (if $t_2 - t_1 < 0$) or *absolute future* (if $t_2 - t_1 > 0$); the sequence of events is the same in all inertial frames. The sequence of events separated by a space-like interval can be different in different inertial frames.

► **Lorentz four-vectors.** A quadruple of scalar quantities $(A_x, A_y, A_z, A_\tau) = (\mathbf{A},\ A_\tau)$ that are transformed, when passing from frame K to frame K', in the same way as (x, y, z, ct), or (see (P1.10.1.1))

$$A_x{}' = \frac{A_x - (V/c)A_\tau}{\sqrt{1 - V^2/c^2}}, \quad A_y{}' = A_y, \quad A_z{}' = A_z, \quad A_\tau{}' = \frac{A_\tau - (V/c)A_x}{\sqrt{1 - V^2/c^2}}, \qquad \text{(P1.10.2.5)}$$

is called a *Lorentz four-dimensional vector* (or, for short, *four-vector* or *4-vector*). The numbers A_x, A_y, and A_z are called the spatial components of the 4-vector and A_τ is called its time component. The sum of two 4-vectors and the product of a 4-vector by a scalar number are also 4-vectors. The quantity $A^2 = A_\tau^2 - \mathbf{A}^2$, analogous to the spacetime interval, and also the scalar product $A_\tau B_\tau - \mathbf{A}\cdot\mathbf{B}$ are preserved in changing from one inertial frame to another. A physical equation written as an equality of two 4-vectors remains valid in all inertial frames.

P1.10.3. STR Dynamics

► **Momentum and energy in special relativity.** The velocity components are not transformed in the same way as the components of a 4-vector (cf. equations (P1.10.2.3) and (P1.10.2.5)), since both the numerator and the denominator are transformed in the expression $\mathbf{v} = d\mathbf{r}/dt$. Consequently, the quantity $\sum_j m_j\mathbf{v}_j$, corresponding to the classical definition of momentum, cannot be preserved in all inertial frames. The relativistic 4-vector of momentum is defined as

$$\mathbf{p} = m\frac{d\mathbf{r}}{d\tau} = \frac{m}{\sqrt{1-v^2/c^2}}\frac{d\mathbf{r}}{dt}, \qquad p_\tau = m\frac{d(ct)}{d\tau} = \frac{mc}{\sqrt{1-v^2/c^2}},$$

where $d\tau = dt\,\sqrt{1-v^2/c^2}$ is the infinitesimal change in the proper time of the particle (see (P1.10.2.2)), i.e., the time measured in an inertial frame whose velocity coincides with that of the particle at the given instant ($d\tau$ is independent of which inertial frame the particle is observed from). The spatial components of the 4-vector make up the *relativistic momentum*

$$\mathbf{p} = \frac{m\mathbf{v}}{\sqrt{1-v^2/c^2}}, \tag{P1.10.3.1}$$

while the time component p_τ turns out to be equal to E/c, where E is the *relativistic energy* of the particle:

$$E = \frac{mc^2}{\sqrt{1-v^2/c^2}}. \tag{P1.10.3.2}$$

The 4-vector $(\mathbf{p}, E/c)$ is called the *momentum-energy four-vector* or the *four-momentum*. It is noteworthy that the relativistic energy and momentum are linked by the simple relation

$$\mathbf{p} = \frac{E\mathbf{v}}{c^2}. \tag{P1.10.3.3}$$

In accordance with (P1.10.2.5), in changing from one inertial frame to another, the energy and momentum are transformed as

$$p_x{}' = \frac{p_x - VE/c^2}{\sqrt{1-V^2/c^2}}, \quad p_y{}' = p_y, \quad p_z{}' = p_z, \quad E' = \frac{E - Vp_x}{\sqrt{1-V^2/c^2}}. \tag{P1.10.3.4}$$

The relativistic energy of a particle is nonzero at $v = 0$; it consists of the rest energy, mc^2, and the kinetic energy,

$$E = \frac{mc^2}{\sqrt{1-v^2/c^2}} = mc^2 + E_\mathrm{k}. \tag{P1.10.3.5}$$

Note that for $v/c \ll 1$, the relativistic kinetic energy reduces to the classical kinetic energy $E_\text{k} = mv^2/2$. Since the quantity $(E/c)^2 - p^2$ is conserved, it can be calculated in a frame where the particle is at rest at the given instant:

$$\frac{E^2}{c^2} - p^2 = m^2c^2 \qquad \text{or} \qquad E^2 - p^2c^2 = m^2c^4. \tag{P1.10.3.6}$$

For particles with zero mass (e.g., photons), the energy–momentum relation becomes

$$E = pc \tag{P1.10.3.7}$$

(see also (P1.10.3.3)). Substituting $E = mc^2 + E_\text{k}$ into (P1.10.3.6) yields a convenient relation between momentum and kinetic energy:

$$p^2c^2 = E_\text{k}(E_\text{k} + 2mc^2).$$

► **Inelastic collision. Relation between energy and mass.** Let us write down the law of conservation of relativistic energy for a perfectly inelastic collision of two particles of mass m each that moved toward each other with identical speeds v. We have

$$\frac{mc^2}{\sqrt{1 - v^2/c^2}} + \frac{mc^2}{\sqrt{1 - v^2/c^2}} = Mc^2.$$

It is apparent that the mass M of the combined particle is larger than the total mass of the initial particles. The increase in the internal energy at the inelastic collision by ΔE resulted in a mass increase of $\Delta E/c^2$. This example illustrates the general Einstein relationship between the relativistic energy of a resting body and its mass:

$$E = mc^2. \tag{P1.10.3.8}$$

Relativistic energy includes all kinds of energy.

Example 1. Suppose the energy of a body at rest equals E. Find the momentum of the body in a reference frame moving with a speed $-v$.

Solution. In accordance with the relativistic transformation formulas (P1.10.3.4), the momentum equals

$$\mathbf{p}' = (E'/c^2)\mathbf{v} = \frac{(E/c^2)\mathbf{v}}{\sqrt{1 - v^2/c^2}}.$$

It is seen that we have obtained the formula for relativistic momentum with mass $m = E/c^2$.

Example 2. Accelerators with colliding particle beams. Suppose that in order to achieve the threshold energy of a nuclear reaction, the particles in colliding beams must be accelerated to a kinetic energy αmc^2. To what kinetic energy must a single particle be accelerated if the other particle is in a stationary target?

Solution. In the nonrelativistic case, the energy in a center-of-mass frame equals half the energy of the oncoming particle. It is this energy that can be spent to the reaction, while the other half is inaccessible— it remains as kinetic energy of the system. In the relativistic (an especially ultrarelativistic) limit, the picture is cardinally different. Compute the energy of the moving particle (the other particle is at rest). Write the energy-momentum invariance condition for $E^2 - p^2c^2$. In a center-of-mass frame, we have $E = 2\alpha mc^2 + 2mc^2$ and $p = 0$. In a moving inertial frame where one of the particles is at rest, we have $E = E_1 + mc^2$ and $p = p_1$, with E_1 and p_1 being the energy and momentum of the moving particle. On rearranging and taking into account the identity $E_1^2 - p_1^2c^2 = m^2c^4$, we find the kinetic energy of the oncoming particle as $E_1 - mc^2 = 2mc^2\alpha(\alpha + 2)$. For example, if $\alpha = 100$, we find that one particle must be accelerated to 20,400 mc^2. Approximately, $1/100$ of the particle energy is available for the reaction.

► **Basic law of relativistic dynamics.** The force exerted on a particle equals, just as in the case of classical mechanics, the time derivative of the momentum

$$\mathbf{F} = \frac{d\mathbf{p}}{dt}. \qquad \text{(P1.10.3.9)}$$

However, the relativistic momentum (P1.10.3.1) is different from the classical one. Under the action of the load applied, the momentum can increase unboundedly. However, as seen from definition (P1.10.3.1), the speed will be less than c. The work done by the force (P1.10.3.9),

$$\begin{aligned}\int \mathbf{F}\cdot d\mathbf{r} &= \int \frac{d\mathbf{p}}{dt}\cdot \mathbf{v}\,dt = \int \mathbf{v}\cdot d\mathbf{p}\\ &= \int \sqrt{1-v^2/c^2}\,\frac{\mathbf{p}\cdot d\mathbf{p}}{m} = \int \sqrt{1-v^2/c^2}\,\frac{E\,dE}{mc^2} = \int dE,\end{aligned}$$

equals the change in the kinetic energy. Here we have used formulas (P1.10.3.1) and (P1.10.3.2) and also $\mathbf{p}\cdot d\mathbf{p} = E\,dE/c^2$ (see (P1.10.3.6)).

Bibliography for Chapter P1

Alenitsyn, A. G., Butikov, E. I., and Kondratyev, A. S., *Concise Handbook of Mathematics and Physics*, CRC Press, Boca Raton, Florida, 1998.

Benenson, W., Harris, J. W., Stocker, H., and Lutz, H. (Eds.), *Handbook of Physics*, Springer-Verlag, New York, 2002.

French, A. P., *Special Relativity (M.I.T. Introductory Physics Series)*, W. W. Norton & Company, New York, 1968.

French, A. P., *Newtonian Mechanics (M.I.T. Introductory Physics Series)*, W. W. Norton & Company, New York, 1971.

Goldstein, H., Poole, C. P., and Safko, J. L., *Classical Mechanics (3rd Edition)*, Addison-Wesley, New York, 2002.

Irodov, I. E., *Mechanics. Basic Laws* [in Russian], Fizmatlit, Moscow, 2006

José J. V. and Saletan E. J., *Classical Dynamics: A Contemporary Approach*, Cambridge University Press, Cambridge, England, 1998.

Kibble, T. W. B. and Berkshire, F. H., *Classical Mechanics, 5th edition* , Imperial College Press, London, 2004.

Kittel, C., Knight, W. D., and Ruderman, M. A., *Mechanics (Berkeley Physics Course, Vol. 1), 2nd edition*, McGraw-Hill Book Company, New York, 1973.

Kleppner, D. and Kolenkow, R. J., *Introduction to Mechanics*, McGraw-Hill Inc., Boston, Massachusetts, 1973.

Landau, L. D. and Lifshitz, E. M., *Mechanics, Third Edition: Volume 1 (Course of Theoretical Physics)*, Butterworth-Heinemann, Oxford, England, 1976.

Morin, D., *Introduction to Classical Mechanics With Problems and Solutions*, Cambridge University Press, Cambridge, England, 2008.

Okun, L. B., *Energy and Mass in Relativity Theory*, World Scientific Publishing Company, London, 2009.

Prokhorov, A. M. (Ed.), *Encyclopedia of Physics, Volumes 1–5* [in Russian], Bolshaya Russkaya Encyclopedia, Moscow, 1998.

Prokhorov, A. M. (Ed.), *Encyclopedic Dictionary of Physics* [in Russian], Sovetskaya Encyclopedia, Moscow, 1984 (dic.academic.ru).

Resnick, R., *Introduction to Special Relativity*, John Wiley and Sons, Inc., New York, 1968.

Sivukhin, D. V., *General Physics, Volume 1, Mechanics* [in Russian], Fizmatlit, Moscow, 2005.

Sivukhin, D. V. and Yakovlev, I. A., *A Collection of Problems, Volume 1, Mechanics* [in Russian], Fizmatlit, Moscow, 2006.

Symon, K. R., *Mechanics, 3rd Edition*, Addison-Wesley, Reading, Massachusetts, 1971.

Taylor, J. R., *Classical Mechanics*, University Science Books, Sausalito, 2005.

Thornton, S. T. and Marion, J. B., *Classical Dynamics of Particles and Systems, 5th edition*, Brooks Cole, Belmont, California, 2003.

Woan, G., *The Cambridge Handbook of Physics Formulas*, Cambridge University Press, Cambridge, England, 2003.

Chapter P2

Molecular Physics and Thermodynamics

P2.1. Basic Principles and Definitions

P2.1.1. Two Approaches to Studying Matter

► **Thermodynamic approach.** The *thermodynamic approach* of studying substances consists of establishing links and relations between experimentally determined (phenomenological) parameters (called *thermodynamic parameters*) based on several postulates (*laws of thermodynmics*).

► **Statistical approach.** The *statistical approach* relies on kinetic-molecular postulates on the structure of substance (*fundamentals of kinetic-molecular theory*). These postulates include the following:

1. All bodies consist of a huge number of tiny particles—atoms and molecules.
2. These molecules are in constant random motion.
3. The molecules interact with one another: they experience attractive forces at large distances and repulsive forces at small enough distances.

The thermodynamic parameters are calculated within the framework of a specific model of internal structure of substance (i.e., the model of motion and interaction of atoms and molecules) by averaging over a huge number of states of the system. Statistical physics employs methods of probability theory and mathematical statistics. The classical theory relies on the classical laws of molecular motion, while quantum statistics relies on the laws of quantum mechanics.

► **Amount of substance. Mole.** The amount of substance in a system—or the number of constituting structural units, atoms and molecules—is measured in *moles*. One mole of any substance contains a certain number of molecules, called *Avogadro's number* (or the *Avogadro constant*) and equal to the number of atoms in 12 g of carbon-12. Avogadro's number equals $N_\mathrm{A} \approx 6.02 \times 10^{23}$ mol^{-1}. The number of moles in the system is expressed as

$$\nu = \frac{N}{N_\mathrm{A}} = \frac{m}{\mu}, \qquad \text{(P2.1.1.1)}$$

where N is the number of molecules in the system, $m = m_0 N$ is the mass of the system, m_0 being the mass of a single molecule, and $\mu = m_0 N_\mathrm{A}$ is the molar mass of the substance.

P2.1.2. Equation of State

► **Equilibrium states.** A thermodynamic system that is kept under unchanged external conditions comes to an *equilibrium state*, where there are no fluxes of any kind (e.g., mass or energy fluxes). The thermodynamic parameters of the equilibrium state (pressure p, temperature T, volume V, density ρ, molar mass μ, etc.) are related by an *equation of state*. For example, the *ideal gas law* (also known as the *Clapeyron–Mendeleev equation*), which is the equation of state of an ideal gas, has the form

$$pV = \nu RT \qquad \text{or} \qquad pV = NkT, \qquad \text{(P2.1.2.1)}$$

where $R \approx 8.31\,\mathrm{J/(mol\,K)}$ is the *gas constant* and $k = R/N_\mathrm{A} \approx 1.38 \times 10^{-23}\,\mathrm{J/K}$ is the *Boltzmann constant*. This equation was obtained empirically and it describes well the behavior of thin real gases (see Subsection P2.5.2). The *van der Waals equation*,

$$(V - \nu b)\Big(p + \frac{a\nu^2}{V^2}\Big) = \nu RT, \qquad \text{(P2.1.2.2)}$$

is an equation of state that provides good approximation to the behavior of dense real gases (see Section P2.6).

A process of changing the state of a system is called *equilibrium* (also *quasi-equilibrium* or *quasistatic*) if it occurs so slowly that every intermediate state of the system can be treated as equilibrium. Equilibrium processes are typically *reversible* (i.e., can be made to pass through all their intermediate states in order); however, irreversible equilibrium processes are possible (e.g., for systems with memory). In what follows, when speaking of an equilibrium process we imply a reversible equilibrium process.

► **Simple thermodynamic system.** A homogeneous and isotropic thermodynamic system whose chemical composition does not change is called (*thermodynamically*) *simple*. Examples of simple systems are one-component (pure) gases and liquids. A simple system has two degrees of freedom, which means that only two of its thermodynamic parameters are independent. Equilibrium states of a simple system can be represented by points in a plane (e.g., in coordinates p and V) and equilibrium processes, by lines in this plane. The work done by a simple system against external forces equals

$$\delta A = p\,dV, \qquad A = \int_{V_1}^{V_2} p(V)\,dV, \qquad \text{(P2.1.2.3)}$$

where the relationship $p(V)$ is called a *process equation*. Since the work is not a function of state, the infinitesimal work increment is denoted by δA, in order to distinguish it from the infinitesimal increment of a function (differential).

If any three parameters of a simple system are related to one another, due to an equation of state, their derivatives are also related by an equation. For example, expressing the infinitesimal increment of T in terms of those of p and V, we have $dT = (\partial T/\partial V)_p\,dV + (\partial T/\partial p)_V\,dp$. By applying this expression to isothermal process ($dT = 0$), we obtain

$$\Big(\frac{\partial V}{\partial T}\Big)_p \Big(\frac{\partial T}{\partial p}\Big)_V \Big(\frac{\partial p}{\partial V}\Big)_T = -1.$$

This equation relates three coefficients

$$\alpha = \frac{1}{V}\Big(\frac{\partial V}{\partial T}\Big)_p \quad (\textit{thermal volume expansion coefficient}),$$

$$\beta = \frac{1}{p}\Big(\frac{\partial p}{\partial T}\Big)_V \quad (\textit{thermal pressure coefficient}),$$

$$K = -V\Big(\frac{\partial p}{\partial V}\Big)_T \quad (\textit{isothermal bulk modulus}).$$

► **Nonequilibrium systems.** *Equilibrium thermodynamics* deals with equilibrium states and processes. It can judge the *direction of nonequilibrium processes* between the initial and final equilibrium states but cannot provide quantitative characterization of these processes. Nonequilibrium systems are studied by *physical kinetics* and *nonequilibrium thermodynamics*. Note that the time in which a system comes to equilibrium (*relaxation time*) decreases with decreasing system size, and therefore the term *local equilibrium* in a small portion of a nonequilibrium system makes sense. Consequently, it is meaningful to speak of local thermodynamic parameters (functions of state) that vary from point to point in a nonequilibrium system (e.g., temperature $T(\mathbf{r})$).

An important role in studying nonequilibrium systems is played by the notions of *flux* of a physical quantity (e.g., the number of molecules, mass, momentum, energy, charge, etc.) through a given surface and *flux density* at a given point in space.

For instance, energy flux W equals the amount of energy transferred through a surface per unit time (measured in J/s, or W). A *flux density* is a vector quantity that characterizes the direction and intensity of transfer of a physical quantity through a given point in space. For instance, energy flux density **w** equals in magnitude to the ratio of the energy flux ΔW through a small surface to the area ΔS of this surface (measured in $\mathrm{W/m^2}$) in the limit $\Delta S \to 0$. Among all small surfaces through the given point, one must take one for which the above ratio is maximum. The vector **w** is taken to be perpendicular to this surface and pointing in the direction of the energy transfer. The energy flux through a given surface equals $W = \int \mathbf{w}\,d\mathbf{s}$, where $d\mathbf{s} = \mathbf{n} \cdot ds$, with **n** being the positive unit normal vector at the given point of the surface. Note that in electrodynamics (see Subsection P3.7.1), the flux of electric charge is called the electric current I and the flux density, the current density **j**.

A state of a nonequilibrium system is called *stationary* if the flux densities of all physical quantities at all points are time independent. Recall that in an equilibrium state, there are no fluxes, that is, the flux densities at all points are zero.

P2.2. First Law of Thermodynamics

P2.2.1. Quantity of Heat. Internal Energy

► **Quantity of heat.** Unlike mechanical energy, which can only change through work, internal energy can change both due to work and due to contact with bodies having different temperatures (in the process of heat exchange). The amount of energy absorbed or lost by a system in a thermal contact is called the *quantity of heat* or just *heat* and denoted Q. Heat is considered positive if energy is received by the system and negative if it is given off.

In a thermal contact, energy is transferred from a body with a higher temperature to a body with a lower temperature. Any temperature scale must satisfy this property. Empirical temperature scales are based on indirect measurements, that is, on changes in parameters monotonically dependent on temperature. A gas scale of temperatures relies on the equation of state of the ideal gas (P2.1.2.1). A thermodynamic, or absolute scale, is based on the second law of thermodynamics.

► **Internal energy.** From the viewpoint of classical kinetic-molecular theory, the internal energy U of a thermodynamic system equals the sum of the kinetic energy of motion of all its molecules and the potential energy of their interaction. The internal energy, $U = U(V, T, \dots)$, is a function of state of a thermodynamic system;* it is the most important characteristic of the system.

Thermodynamics must define any quantity phenomenologically, regardless of our knowledge about the internal structure of the substance. The concept of an *adiabatic envelope* can be used to see that this is possible in principle. The adiabatic envelope of thermodynamic system allows the shape of the system to be changed and work on the system to be done but excludes thermal exchange of the system with external bodies. In this case, the work done by the external forces, A^{e}, as the system performs a transition from state 1 to state 2 is independent of the transition process and is only dependent on the initial and final states. This allows us to define the internal energy difference as $U_2 - U_1 = A^{\mathrm{e}}(1 \to 2)$.

P2.2.2. First Law of Thermodynamics

The general law of conservation of energy, which takes into account thermal exchange and internal energy, reads

$$\Delta E_{\text{mech}} + \Delta U = A^{\mathrm{e}} + Q. \qquad \text{(P2.2.2.1)}$$

It states that the change in the total (mechanical plus internal) energy equals the work done by external forces and the quantity of heat received by the system in thermal exchange with external bodies. Sometimes, the law of conservation of energy is formulated as the impossibility of creating a *perpetual motion machine of the first kind* (that would produce energy from nothing). The *first law of thermodynamics* is the special case of the law of conservation of energy as applied to thermodynamic systems, whose mechanical energy does not change. In addition, in thermodynamics it is customary to use the work done by the system against external forces: $A = -A^{\mathrm{e}}$. We have

---

* The relationship $U = U(V, T, \dots)$ is called a *caloric equation of state*, in contract to $p = p(V, T)$, called a *thermal equation of state*.

$$Q = \Delta U + A, \qquad \text{(P2.2.2.2)}$$

which means that the heat supplied to a thermodynamic system is spent to change its internal energy and produce work against external forces.

► **First law of thermodynamics for a simple system. Heat capacity.** Using equation (P2.1.2.3), let us write out the first law of thermodynamics for a simple system that has received an infinitesimal quantity of heat:

$$\delta Q = dU + p\,dV. \qquad \text{(P2.2.2.3)}$$

The ratio of δQ to the temperature increment dT is called the *heat capacity* of the system in the equilibrium process being considered. Heat capacity is measured in J/K and denoted by C. We have

$$C\,dT = dU + p\,dV. \qquad \text{(P2.2.2.4)}$$

Considering U in (P2.2.2.4) to be a function of two variables (V,T), we can write

$$dU = \left(\frac{\partial U}{\partial T}\right)_V dT + \left(\frac{\partial U}{\partial V}\right)_T dV = C_V\,dT + \left(\frac{\partial U}{\partial V}\right)_T dV, \qquad \text{(P2.2.2.5)}$$

where $C_V = \left(\frac{\partial U}{\partial T}\right)_V$ is the *heat capacity at constant volume* (see equation (P2.2.2.4)). So we have

$$C\,dT = C_V\,dT + \left[p + \left(\frac{\partial U}{\partial V}\right)_T\right] dV. \qquad \text{(P2.2.2.6)}$$

It can be seen that the heat capacity for a process $V(T)$ depends on the value of the derivative dV/dT. Since the parameters (V,T) can be varied independently, this derivative and, hence, the heath capacity at a given point can take any value.

► **Enthalpy. The Joule–Thomson process.** If the state of a system is changed as an equilibrium process under a constant pressure, then the heat supplied can be expressed as

$$Q = \Delta U + p\Delta V = \Delta(U + pV) = \Delta H \qquad (p = \text{const}). \qquad \text{(P2.2.2.7)}$$

The quantity $H = U + pV$ is called the *enthalpy* of the system. So the heat supplied equals the enthalpy difference at the final and initial states. Enthalpy is also used to describe the *Joule–Thomson process* (also called a *throttling process*), where a substance is forced through a porous wall without heat exchange with the environment. This process is used to achieve very low temperatures. Enthalpy is conserved in the Joule–Thomson process: $U_1 + p_1V_1 = U_2 + p_2V_2$.

► **Thermodynamic cycle.** A thermodynamic cycle is a series of thermodynamic processes in which a system eventually returns to its initial equilibrium state. At some stage, the system absorbs a quantity of heat Q_1 from higher-temperature bodies (*hot reservoir*) and at another stage, it gives up some heat Q_2 to lower-temperature bodies (*cold reservoir*). (This follows from the second law of thermodynamics; see Section P2.3.) For a heat engine, a device that runs on a thermodynamic cycle, its *thermal efficiency* is the number defined as

$$\eta = \frac{A}{Q_1} = \frac{Q_1 - Q_2}{Q_1} = 1 - \frac{Q_2}{Q_1}, \qquad \text{(P2.2.2.8)}$$

where A is the net work done in one cycle. In a thermodynamic cycle, $\Delta U = 0$, and hence $Q_1 - Q_2 = A$.

P2.2.3. First Law of Thermodynamics for an Ideal Gas

► **Internal energy of an ideal gas**. The internal energy of an ideal gas depends only on its temperature, so that $U(p_1, V_1, T) = U(p_2, V_2, T)$. There is experimental evidence supporting this statement.

1. When a gas expands into a void space inside a rigid adiabatic container, both the work done by the gas, A, and the heat absorbed, Q, are zero, and hence $\Delta U = 0$. It was found (experiments by Gay-Lussac and Joule) that in this non-equilibrium process, the temperature of ideal gases does not change.

2. The temperature of an ideal gas does not change in the Joule–Thomson process.

The second law of thermodynamics allows us to prove that $(\partial U/\partial V)_T = 0$ based on the thermic equation of state (P2.1.2.1).

From equations (P2.2.2.4) and (P2.2.2.5) we find the increment in the internal energy and the first law of thermodynamics for an ideal gas:

$$dU = C_V\,dT, \qquad \delta Q = C_V\,dT + p\,dV. \tag{P2.2.3.1}$$

Experiments show that C_V is independent of T in a wide range of temperatures. Hence, we can write $U = C_V T + \text{const}$.

► **Processes in an ideal gas.** Consider a few examples to illustrate the application of the first law of thermodynamics to specific equilibrium processes in an ideal gas.

Example 1. Isothermal process ($dT = 0$). In this case, $dU = 0$. Hence,

$$\delta Q = p\,dV, \qquad Q = A = \int_{V_1}^{V_2} p(V)\,dV = \nu RT \ln\left(\frac{V_2}{V_1}\right). \tag{P2.2.3.2}$$

The heat capacity of this process can be considered infinitely large.

Example 2. Isobaric process ($dp = 0$). Introducing the *heat capacity at constant pressure* and taking into account the equation of state (P2.1.2.1), we get

$$C_p\,dT = C_V\,dT + p\,dV = C_V\,dT + \nu R\,dT.$$

Dividing by dT gives the relation between the heat capacities (*Mayer's relation*):

$$C_p = C_V + \nu R, \qquad c_p = c_V + \frac{R}{\mu}, \qquad C_p^{\rm m} = C_V^{\rm m} + R, \tag{P2.2.3.3}$$

where $c = C/m$ is the *specific heat capacity* and $C^{\rm m} = C/\nu$ is the *molar heat capacity* of the gas (measured in J/(kg K) and J/(mol K), respectively).

Example 3. Adiabatic process. If an equilibrium process occurs without heat supply ($\delta Q = 0$), then

$$C_V\,dT = -p\,dV \quad\Longrightarrow\quad \nu C_V^{\rm m}\,dT = -\frac{\nu RT}{V}\,dV \quad\Longrightarrow\quad \frac{dT}{T} = -\frac{R}{C_V^{\rm m}}\frac{dV}{V}.$$

Integrating yields the equation of the adiabatic process (*Poisson's equation*):

$$TV^{\gamma-1} = \text{const} \qquad \text{or} \qquad pV^{\gamma} = \text{const}, \tag{P2.2.3.4}$$

where the dimensionless parameter $\gamma = C_p/C_V = 1 + (R/C_V^{\rm m})$ is called the *adiabatic index*. Using Mayer's relation (P2.2.3.3), we can express the molar heat capacities in terms of γ as

$$C_V^{\rm m} = \frac{R}{\gamma - 1}, \qquad C_p^{\rm m} = \frac{\gamma R}{\gamma - 1}.$$

Example 4. Polytropic process. A thermodynamic process with constant heat capacity ($C = \text{const}$) is called polytropic. We have

$$C\,dT = C_V\,dT + p\,dV \quad\Longrightarrow\quad (C_V - C)\,dT = -p\,dV.$$

Proceeding in the same manner as for the adiabatic process, we arrive at the *polytropic process equation*

$$pV^n = \text{const},$$

where $n = 1 + \dfrac{R}{C_V^{\rm m} - C^{\rm m}} = \dfrac{C_p - C}{C_V - C}$ is the *polytropic index.*

It can be verified that the above thermodynamic processes are special cases of a polytropic process—isochoric with $n = \infty$, isobaric with $n = 0$, isothermal with $n = 1$, and adiabatic with $n = \gamma$.

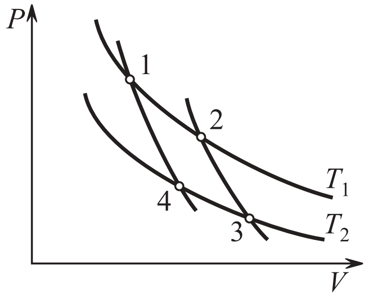

Figure P2.1 A Carnot cycle consists of two isothermal processes (1 to 2 and 3 to 4) and two adiabatic processes (2 to 3 and 4 to 1).

Example 5. Carnot cycle. A Carnot cycle is a thermodynamic cycle consisting of two isothermal processes (at temperatures T_1 and T_2 with $T_1 > T_2$) and two adiabatic processes as shown in Fig. P2.1.

From equation (P2.2.3.2) we have

$$Q_1 = \nu R T_1 \ln\left(\frac{V_2}{V_1}\right), \qquad Q_2 = \nu R T_2 \ln\left(\frac{V_3}{V_4}\right).$$

From Poisson's equation (P2.2.3.4) we have

$$T_1 V_1^{\gamma-1} = T_2 V_4^{\gamma-1},$$
$$T_1 V_2^{\gamma-1} = T_2 V_3^{\gamma-1}.$$

It follows that $V_2/V_1 = V_3/V_4$. Hence, the thermal efficiency is

$$\eta = 1 - \frac{T_2}{T_1}. \qquad \text{(P2.2.3.5)}$$

This means that the thermal efficiency of a Carnot cycle depends only on the cold-to-hot reservoir temperature ratio.

P2.3. Second Law of Thermodynamics

P2.3.1. Formulations

Below are two most common formulations of the second law of thermodynamics:

1. It is impossible to construct an engine that would convert 100% of heat extracted from a heat reservoir into work in a cyclic process (formulation by William Thomson—Lord Kelvin). It is the same to say that it is impossible to create a *perpetual motion machine of the second kind* (that would produce work at the expense of solely the internal energy of a heat reservoir).

2. It is impossible to construct a refrigerator that would produce no other effect than the transfer energy from a cooler body to a hotter one (formulation by Rudolf Clausius).

The Thomson and Clausius formulations are equivalent.

P2.3.2. Carnot's Theorem

In general, a *Carnot cycle* is a thermodynamic cycle in which the working substance (of a heat engine) absorbs heat from a hot reservoir (at a constant temperature T_1) and gives up heat to a cold reservoir (at a constant temperature T_2).

Carnot's theorem states that the thermal efficiency of an arbitrary Carnot cycle cannot exceed that of any reversible Carnot cycle operating at the same temperatures T_1 and T_2. It immediately follows that the thermal efficiency of a reversible Carnot cycle depends on T_1 and T_2 only and is independent of the working substance nature.

Let us sketch the proof of Carnot's theorem. Assume that the thermal efficiency of the reversible heat engine is less than that of the irreversible one. Let us choose the amount of the working substance in the reversible engine such that the work done in one cycle is the same as the work done by the irreversible engine. Using (P2.2.2.8) and the above assumption, we can write the thermal efficiency inequality as $A/Q_1^{\rm rev} < A/Q_1^{\rm irrev}$, or $Q_1^{\rm rev} > Q_1^{\rm irrev}$. Now let us revert the operation of the reversible engine so that its work is consumed by the irreversible engine. In one cycle, the combined engine produces zero work and the hot reservoir receives the energy $Q_1^{\rm rev} - Q_1^{\rm irrev} > 0$ completely extracted from the cold reservoir. We have come to a conclusion that contradicts the Clausius formulation of the second law.

Since we know the thermal efficiency of a Carnot engine, the one that runs on an ideal gas (see (P2.2.3.5)), we can rewrite Carnot's theorem as

$$\frac{Q_1 - Q_2}{Q_1} \le \frac{T_1 - T_2}{T_1} \qquad \text{or} \qquad \frac{Q_1}{T_1} - \frac{Q_2}{T_2} \le 0. \tag{P2.3.2.1}$$

The equality corresponds to a reversible Carnot cycle.

► **Thermodynamic temperature scale.** Carnot's theorem allows us to introduce a temperature scale that does not depend on the properties of specific bodies involved. The temperature ratio of two bodies can be determined by connecting them with a reversible Carnot engine. Since the ratio Q_1/Q_2 depends only on the two temperatures, it can be taken to be equal to the ratio of the thermodynamic temperatures: $\theta_1/\theta_2 = Q_1/Q_2$. As seen from (P2.3.2.1), the ratio of the thermodynamic temperatures equals that of the ideal gas temperatures (in the range where the gas scale is defined).

P2.3.3. Calculation of the Internal Energy

► **Relation between U and V.** The second law of thermodynamics allows us to derive the important formula for the internal energy of a simple system

$$\left(\frac{\partial U}{\partial V}\right)_T = T\left(\frac{\partial p}{\partial T}\right)_V - p, \tag{P2.3.3.1}$$

which cannot be obtained based on the first law only.

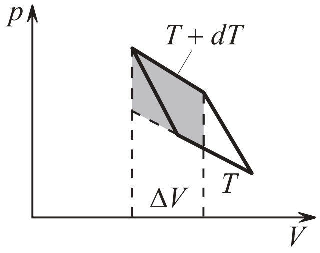

Figure P2.2. Infinitesimal Carnot cycle.

Let us derive formula (P2.3.3.1) from Carnot's theorem. Consider a very small (infinitesimal) reversible Carnot cycle for a simple system and plot it in the pV-coordinates. The work done by the system in one cycle, equal to the area of the small parallelogram (Fig. P2.2), will not change if we replace the adiabatic curves by vertical segments, whose lengths equal $\left(\frac{\partial p}{\partial T}\right)_V \Delta T$. Multiplying by the height ΔV gives $\delta A = \left(\frac{\partial p}{\partial T}\right)_V \Delta T \Delta V$. The heat absorbed at the top isotherm is equal to $\delta Q_1 = \Delta U + p\,dV = \left(\frac{\partial U}{\partial V}\right)_T \Delta V + p\,\Delta V$, where ΔU has been substituted by its expression (P2.2.2.5) at constant temperature. From Carnot's theorem and equation (P2.3.2.1) we get

$$\frac{\delta A}{\delta Q_1} = \frac{\Delta T}{T} \quad\Longrightarrow\quad \frac{\left(\dfrac{\partial p}{\partial T}\right)_V \Delta T \Delta V}{\left[\left(\dfrac{\partial U}{\partial V}\right)_T + p\right]\Delta V} = \frac{\Delta T}{T},$$

which immediately leads to (P2.3.3.1).

Below are a few examples illustrating the application of formula (P2.3.3.1).

Example 1. Internal energy of an ideal gas. Let us substitute the equation of state $p = \frac{1}{V}\nu RT$ into (P2.3.3.1) to obtain $\left(\frac{\partial U}{\partial V}\right)_T = 0$, which means that the internal energy is independent of volume.

Example 2. Internal energy of a van der Waals gas. Expressing pressure from the equation of state (P2.1.2.2) and substituting into (P2.3.3.1), we obtain $\left(\frac{\partial U}{\partial V}\right)_T = \frac{a\nu^2}{V^2}$. It follows that

$$dU = C_V\,dT + \frac{a\nu^2}{V^2}\,dV.$$

Furthermore, we have

$$\left(\frac{\partial C_V}{\partial V}\right)_T = \frac{\partial}{\partial V}\left[\left(\frac{\partial U}{\partial T}\right)_V\right] = \frac{\partial}{\partial T}\left[\left(\frac{\partial U}{\partial V}\right)_T\right] = \frac{\partial}{\partial T}\left(\frac{a\nu^2}{V^2}\right)_V = 0,$$

which means that C_V is independent of volume. For the temperature range where C_V is weakly dependent on T, we can write

$$U = C_V T - \frac{a\nu^2}{V}. \tag{P2.3.3.2}$$

Example 3. General formula for $C_p - C_V$. Substituting (P2.3.3.1) into (P2.2.2.6) and taking pressure to be constant, we find

$$C_p - C_V = T\left(\frac{\partial p}{\partial T}\right)_V\left(\frac{\partial V}{\partial T}\right)_p = -T\left(\frac{\partial p}{\partial V}\right)_T\left(\frac{\partial V}{\partial T}\right)_p^2 \geq 0.$$

(We have used the relation for the derivatives given at the end of Subsection P2.1.2.) This inequality follows from the condition of mechanical stability $\left(\frac{\partial p}{\partial V}\right)_T \leq 0$.

P2.4. Entropy

P2.4.1. Clausius Inequality

Inequality (P2.3.2.1) is a special case of the *Clausius inequality*, which holds for any thermodynamic cycle. If a system absorbs, in a thermodynamic process, some quantities of heat $Q_1, \ldots, Q_N$ from external reservoirs having temperatures $T_1^{\rm e}, \ldots, T_N^{\rm e}$, then the following inequality holds:

$$\sum_{i=1}^{N} \frac{Q_i}{T_i^{\rm e}} \leq 0 \qquad \text{or} \qquad \oint \frac{\delta Q}{T^{\rm e}} \leq 0. \tag{P2.4.1.1}$$

For a reversible process, the inequality becomes equality and the temperature of the reservoir with which the system is in thermal contact at the given point of the cycle is equal to the temperature of the system: $T^{\rm e} = T$. In this case,

$$\oint_{\rm rev} \frac{\delta Q}{T} = 0. \tag{P2.4.1.2}$$

Equation (P2.4.1.2) serves to define another function of state—entropy.

P2.4.2. Definition of Entropy

The Clausius equation for a reversible cycle (P2.4.1.2) can be reformulated as the condition for the integral

$$\int_A^B \frac{\delta Q}{T}$$

to be independent of the equilibrium path by which the system passes from state A to state B. This allows one to define a state variable, *entropy*, as

$$S_B - S_A = \int_A^B \frac{\delta Q}{T}, \qquad dS = \left(\frac{\delta Q}{T}\right)_{\text{rev}}. \tag{P2.4.2.1}$$

To be precise, this relation defines the entropy difference between two states. If there is no heat supply in an equilibrium process, entropy does not change; so an adiabatic process can be called isentropic. For a simple system in an equilibrium process, the first law of thermodynamics (P2.2.2.3) acquires the form

$$T\,dS = dU + p\,dV. \tag{P2.4.2.2}$$

Since $\delta Q = T\,dS$ for any equilibrium process, it is convenient to plot equilibrium processes and cycles in the TS-coordinates, where the area under the curve $T(S)$ equals the quantity of heat received. For example, a Carnot cycle is represented by a rectangle in TS-coordinates.

► Examples of entropy calculations.

Example 1. Process with constant heat capacity. The entropy change corresponding to the temperature change from T_1 to T_2 equals

$$S_2 - S_1 = \int_{T_1}^{T_2} \frac{C\,dT}{T} = C\ln\left(\frac{T_2}{T_1}\right). \tag{P2.4.2.3}$$

Example 2. Entropy of an ideal gas. Substituting the internal energy equation for an ideal gas (P2.2.3.1) into (P2.4.2.2) and taking into account the equation of state $pV = \nu RT$, we obtain

$$\begin{aligned} dS &= \nu C_V^{\text{m}}\frac{dT}{T} + \nu R\frac{dV}{V}, \\ S_2 - S_1 &= \nu\left[C_V^{\text{m}}\ln\left(\frac{T_2}{T_1}\right) + R\ln\left(\frac{V_2}{V_1}\right)\right]. \end{aligned} \tag{P2.4.2.4}$$

Sometimes, entropy is written as $S = \nu(C_V^{\text{m}}\ln T + R\ln V)$, bearing in mind that an entropy difference actually arises in applications.

Example 3. Entropy of a van der Waals gas. Substituting (P2.3.3.2) into (P2.4.2.2) and taking into account the equation of state (P2.1.2.2), we obtain

$$\begin{aligned} dS &= \nu C_V^{\text{m}}\frac{dT}{T} + \frac{a\nu^2}{T}\frac{dV}{V^2} + \frac{p}{T}dV = \nu C_V^{\text{m}}\frac{dT}{T} + \frac{\nu R\,dV}{V - \nu b}, \\ S &= \nu\left[C_V^{\text{m}}\ln T + R\ln(V - \nu b)\right]. \end{aligned}$$

P2.4.3. Direction of Non-Equilibrium Processes in a Thermally Insulated System

Consider a non-equilibrium process that takes a system from an equilibrium state 1 to an equilibrium state 2. Let us set up a cyclic process by returning from 2 to 1 using any equilibrium process. Writing down the Clausius inequality (P2.4.1.1) for this cycle and taking into account the entropy definition (P2.4.2.1) for the equilibrium process, we get

$$\int_1^2 \frac{\delta Q}{T^{\text{e}}} \le S_2 - S_1. \tag{P2.4.3.1}$$

If the non-equilibrium process occurs in an adiabatic envelope, then the left-hand side vanishes and then

$$S_2 \geq S_1, \tag{P2.4.3.2}$$

which means that entropy does not decrease in any process without heat supply. The equilibrium state of a thermally insulated system corresponds to the maximum of entropy.

► Examples of non-equilibrium processes.

Example 1. Heat exchange. Consider heat exchange between two bodies having initial temperatures T_1 and T_2 and identical heat capacities C. From the heat balance equation it follows that the bodies will reach the same equilibrium temperature $T_\mathrm{f} = \frac{1}{2}(T_1 + T_2)$. Moreover, according to (P2.4.2.3), the entropy change of the system will be positive:

$$\Delta S = C \ln\left(\frac{T_\mathrm{f}}{T_1}\right) + C \ln\left(\frac{T_\mathrm{f}}{T_2}\right) = C \ln\left(\frac{T_\mathrm{f}^2}{T_1 T_2}\right) > 0;$$

the argument of the logarithm is greater than one, since the arithmetic mean is greater than the geometric mean.

Example 2. Expansion of an ideal gas into a vacuum. The internal energy and, hence, temperature of an ideal gas remain constant when it expands into a vacuum. The entropy change can be found from (P2.4.2.4):

$$S_2 - S_1 = \nu R \ln\left(\frac{V_2}{V_1}\right). \tag{P2.4.3.3}$$

Entropy increase in adiabatic expansion into a vacuum (volume increase under constant internal energy) can be proved for any simple system. If S in (P2.4.2.2) is treated as a function of U and V, then $\left(\dfrac{\partial S}{\partial V}\right)_U = \dfrac{p}{T} > 0$.

Example 3. Gas mixing. Consider a thermally insulated container divided by a partition into two equal compartments each containing one mole of an ideal gas under the same temperatures and pressures. When the partition is removed, the gases mix with each other. As follows from (P2.4.3.3), the entropy of the system is increased by $2R\ln 2$ (*entropy of mixing*). If the gases are different, the initial state of the system is different from the final state; the gases can be separated again by doing work on special pistons letting through only one sort of molecules. However, if the gases are identical, the initial and final states of the system are the same, and hence the entropy must not change. The absence of a continuous transition when nearly identical molecules are replaced by completely identical ones is called the *Gibbs paradox*.

P2.4.4. Statistical Meaning of Entropy

Boltzmann's entropy formula relates the entropy S of a macroscopic state of a system to the *statistical weight* W of this macrostate (or the number of microstates corresponding to the given macrostate):

$$S = k \ln W, \tag{P2.4.4.1}$$

where k is the *Boltzmann constant*. The greater W, the higher the probability of the macrostate; in non-equilibrium processes, the systems goes from less probable states to more probable states. The logarithm provides for the additivity of entropy: the statistical weight of a system consisting of two independent subsystems equals the product of their statistical weights. (For more details about the statistical meaning of entropy, see Subsection P7.1.1.

Example. Identity of particles and the Gibbs paradox. To illustrate the application of formula (P2.4.4.1), let us calculate the spatial part of the entropy of an ideal gas. Imagine that the whole volume V of the system is divided into small cells of volume v. The corresponding number of independent states equals $W = (V/v)^N$, where N is the number of molecules in the gas. Then the entropy is $S = k\ln W = kN\ln(V/v) = \nu R\ln(V/v)$. Quantum statistics states that two identical particles are indistinguishable in principle, which means that W must be divided by the number of permutations $N!$. Finally, the spatial part of the entropy becomes $S = kN\ln(V/Nv)$. It can be seen that this approach allows us to resolve the Gibbs paradox.

P2.4.5. Third Law of Thermodynamics

The third law of thermodynamics (also known as *Nernst's theorem*) states that as a system approaches absolute zero, its entropy tends to a certain constant value independent of the values of other thermodynamic parameters. The entropy of a system at $T = 0$ is commonly taken to be zero. From the statistical viewpoint this is explained as follows: at $T = 0$, the system is in its lowest energy state whose statistical weight is very low (close to 1) and, by Boltzmann formula (P2.4.4.1), the entropy is negligibly small. It follows from this law that as $T \to 0$, the heat capacities C_V and C_p of the system as well as its thermal volume expansion and pressure coefficients, α and β, tend to zero.

P2.4.6. Thermodynamic Potentials

► **System in a thermostat.** Consider a change in the state of a system that is in a thermal contact with a thermostat having a constant temperature T. The system will have the same temperature at its initial and final equilibrium states. With $T^{\rm e} = \text{const}$ and after a sequence of transformations

$$Q/T \le S_2 - S_1 \quad \Longrightarrow \quad (U_2 - U_1) + A \le TS_2 - TS_1,$$

inequality (P2.4.3.1) becomes

$$(U_1 - TS_1) - (U_2 - TS_2) \ge A \tag{P2.4.6.1}$$

► **Helmholtz free energy.** With the new state variable

$$F = U - TS, \tag{P2.4.6.2}$$

called the *Helmholtz free energy*, inequality (P2.4.6.1) becomes

$$F_1 - F_2 \ge A.$$

If the volume of the system does not change in a non-equilibrium process, then

$$F_2 \le F_1,$$

and so the free energy does not increase. This fact can be rephrased as follows: an equilibrium state of a system of constant volume that is in thermal contact with a thermostat corresponds to the minimum of the free energy. For an equilibrium process, both inequalities above become equalities.

If a system placed in a thermostat and having constant volume is used to produce some "useful" work (e.g., by connecting internal moving parts of the system with external bodies), then the *maximum work* done by the system as it passes from state 1 to state 2 is equal to

$$A_{\rm max} = F_1 - F_2,$$

with the maximum work being done in a completely reversible equilibrium process. This is why this thermodynamic function is called free energy.

From formulas (P2.4.6.2) and (P2.4.2.2) it follows that

$$dF = -S\,dT - p\,dV. \tag{P2.4.6.3}$$

Therefore, the natural variables to express free energy are volume and temperature: $F = F(V,T)$.

► **Gibbs free energy.** Suppose that a thermodynamic system is placed in a thermostat and is in mechanical contact (e.g., through a light piston or a flexible envelope) with the environment that has a fixed pressure p (the system will have the same pressure in the equilibrium state). Then the work done by the system to change its volume equals $A = p(V_2 - V_1)$. From (P2.4.6.1) we get

$$F_1 + pV_1 \geq F_2 + pV_2 \qquad \text{or} \qquad G_1 \geq G_2, \tag{P2.4.6.4}$$

with

$$G = F + pV = U - TS + pV. \tag{P2.4.6.5}$$

The new thermodynamic function G is called the *Gibbs free energy*. It does not increase in non-equilibrium processes and remains constant in completely reversible equilibrium processes. The minimum of $G(p,T)$ corresponds to an equilibrium state. The maximum “useful” work that the system can perform (i.e., the total work minus the work done to change the volume of the system) equals

$$A_{\max} = G_1 - G_2.$$

Since

$$dG = -S\,dT + V\,dp, \tag{P2.4.6.6}$$

the natural variables for the Gibbs free energy are pressure and temperature: $G = G(p,T)$.

► **Chemical potential.** If the pressure and temperature are kept constant, the Gibbs free energy is proportional to the quantity of substance: $G(p,T,m) = mg$ or $G(p,T,N) = N\mu$, where g is the *specific Gibbs free energy* and μ is the thermodynamic potential per molecule, or the *chemical potential*. (Sometimes, the chemical potential per mole is used.)

► **Canonic equations of state. Natural variables.** Although the functions U, H, F, and G, called *thermodynamic potentials*, can be expressed in terms of any independent variables, there are so-called *natural variables* associated with each of them. These are S and V for internal energy, S and p for enthalpy, V and T for Helmholtz free energy, and p and T for the Gibbs free energy. Below are the definitions and the differentials of the above functions listed together:

$$\begin{aligned} & & dU &= T\,dS - p\,dV + \mu\,dN;\\ H &= U + pV; & dH &= T\,dS + V\,dp + \mu\,dN;\\ F &= U - TS; & dF &= -S\,dT - p\,dV + \mu\,dN;\\ G &= U + pV - TS; & dG &= -S\,dT + V\,dp + \mu\,dN. \end{aligned} \tag{P2.4.6.7}$$

Note that each differential has the additive term $\mu\,dN$, which shows the dependence of each function on the number of particles in the system. The expression of a thermodynamic potential in terms of its natural variables is called a *canonic equation of state*. Given a canonic equation, one can obtain both the thermic, $p = p(V,T)$, and caloric, $U(V,T)$, equations of state. For example, $-\left(\frac{\partial F}{\partial V}\right)_T$ gives the expression for $p(V,T)$ and also, having expressed $S(V,T) = -\left(\frac{\partial F}{\partial T}\right)_V$, one finds $U(V,T) = F + TS$.

► **Maxwell's identities.** By equating the cross derivatives of any thermodynamic function with respect to any two variables, one can obtain useful thermodynamic relations, called

Maxwell's identities. Below are Maxwell's identities for the thermodynamic potentials:

$$\begin{aligned}&\left(\frac{\partial T}{\partial V}\right)_S=-\left(\frac{\partial p}{\partial S}\right)_V;\\&\left(\frac{\partial T}{\partial p}\right)_S=\left(\frac{\partial V}{\partial S}\right)_p;\\&\left(\frac{\partial S}{\partial V}\right)_T=\left(\frac{\partial p}{\partial T}\right)_V;\\&\left(\frac{\partial S}{\partial p}\right)_T=-\left(\frac{\partial V}{\partial T}\right)_p.\end{aligned}\tag{P2.4.6.8}$$

Example 1. With Maxwell's identities, it is quite easy to obtain the formula for $\left(\frac{\partial U}{\partial V}\right)_T$ that was derived in Subsection P2.3.3 using an infinitesimally small Carnot cycle. The first equation in (P2.4.6.7) gives

$$\left(\frac{\partial U}{\partial V}\right)_T=T\left(\frac{\partial S}{\partial V}\right)_T-p.$$

Inserting the third identity in (P2.4.6.8), we arrive at the desired formula.

Example 2. Find the quantity of heat required for 1 kg of mercury to be isothermally compressed from 1 atm to 10 atm. The thermal volume expansion coefficient $\alpha = 1.8\times10^{-4}\,\mathrm{K}^{-1}$ and the mercury density $\rho = 13.55\times10^3\,\mathrm{kg\,m}^{-3}$ are assumed to be unchangeable.

Solution. It follows from the last Maxwell's identity in (P2.4.6.8) that

$$Q=T\Delta S=T\left(\frac{\partial S}{\partial p}\right)_T\Delta p=-T\left(\frac{\partial V}{\partial T}\right)_p\Delta p=-TV\frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_p\Delta p=-\frac{m}{\rho}T\alpha\Delta p=-3.6\,\mathrm{J}.$$

P2.5. Kinetic Theory of Ideal Gases

P2.5.1. The Basic Equation of the Kinetic Theory

► **Pressure of an ideal gas.** From the viewpoint of kinetic-molecular theory, a gas is considered ideal if the potential energy of interaction among its molecules is negligible as compared to the kinetic energy and the molecule size is negligible as compared to the average distance between the molecules. The pressure on the walls of the container results from numerous elastic collisions of the molecules with them. (The collisions can be treated as elastic, on the average, since the gas is in thermal equilibrium with the walls.) The *basic equation of the kinetic theory of an ideal gas* expresses pressure in terms of the mean square velocity of the molecules:

$$p=\tfrac{1}{3}n\langle\mathbf{pv}\rangle=\tfrac{1}{3}nm_0\langle v^2\rangle=\tfrac{2}{3}n\langle\varepsilon_{\text{transl}}\rangle,\tag{P2.5.1.1}$$

where $n=N/V$ is the number of molecules per unit volume, m_0 is the mass of one molecule, $\mathbf{p}=m_0\mathbf{v}$ is its momentum, and $\langle\varepsilon_{\text{transl}}\rangle=\langle\tfrac{1}{2}m_0v^2\rangle$ is the mean kinetic energy of translational motion of a molecule.

If the relation between energy and momentum is other than classical, then the expression in (P2.5.1.1) is different. For example, $|\mathbf{p}|=\varepsilon/c$ for a photon (see Subsection P1.10.3). As a result, pressure is expressed as

$$p=\tfrac{1}{3}n\langle\varepsilon\rangle\tag{P2.5.1.2}$$

for a photon gas.

► **Number of collisions against the wall.** The number of molecules, ΔN, hitting a wall region of area ΔS in time Δt equals

$$\frac{\Delta N}{\Delta S \Delta t} = \frac{n\langle v\rangle}{4}. \tag{P2.5.1.3}$$

Calculation formulas. Formulas for the transfer of the number of molecules, momentum, energy, etc. through a small area are obtained by dividing molecules into groups and calculating the contribution of each group. One of the most common approaches is to group molecules having nearly the same velocity, **v**. Consider the group of molecules having velocity magnitudes in the range from v to $v + dv$ in a solid angle $d\Omega$. The number of molecules in this group per unit volume equals $dn(v,\ v+dv;\ d\Omega) = dn(v,\ v+dv)\,d\Omega/4\pi$. In order to find how many molecules in this group pass through a plane region of area ΔS in time Δt, we should construct an oblique cylinder based on this region (Fig. P2.3) whose generatrix has the length $v\Delta t$ and make an angle θ with the x-axis. The molecules from the group concerned that belong to this cylinder only will reach its base in time Δt. Integrating first with respect to angles ($d\Omega = \sin\theta\, d\theta\, d\varphi$) and then with respect to v, one can obtain formulas (P2.5.1.1), (P2.5.1.3), and others. For example, the energy transferred through area ΔS in time Δt is expressed as

$$\begin{aligned}\Delta E &= \int \frac{m_0 v^2}{2}\, dn(v, v+dv)\, \frac{d\Omega}{4\pi}\, \big(\Delta S\, v\Delta t \cos\theta\big)\\ &= \frac{m_0 n}{8} \Delta S\, \Delta t \int_0^\infty \frac{v^3\, dn(v, v+dv)}{n} = \frac{m_0 n\langle v^3\rangle}{8} \Delta S\, \Delta t.\end{aligned} \tag{P2.5.1.4}$$

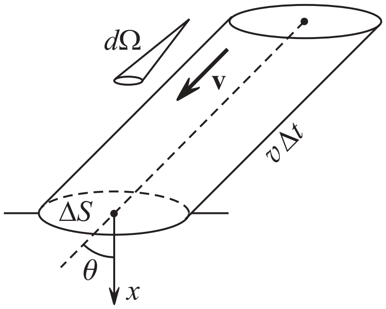

Figure P2.3. To the calculation of the number of molecules passing through a region of area ΔS in time Δt.

P2.5.2. Temperature and Energy of an Ideal Gas

► **Temperature definition in kinetic theory of gases.** It is proved in kinetic theory of gases that if two subsystems (consisting of identical or different molecules) can exchange energy, their mean kinetic energies of translational motion per molecule are equal to each other in equilibrium. Based on this fact, kinetic theory of gases defines temperature as a quantity proportional to the average kinetic energy of translational motion of molecules:

$$\langle \varepsilon_{\text{transl}}\rangle = \frac{3}{2}kT, \tag{P2.5.2.1}$$

where k is the *Boltzmann constant*. The proportionality coefficient is chosen so that the equation of state of an ideal gas,

$$p = nkT = \frac{N}{V}kT, \tag{P2.5.2.2}$$

obtained from the basic equation (P2.5.1.1) and the temperature definition (P2.5.2.1), coincides with equation (P2.1.2.1), where the gas temperature scale is used.

► **Root mean square velocity.** Using (P2.5.1.1) and (P2.5.2.1), one can calculate the root mean square velocity of translational motion of molecules:

$$v_{\rm rms} = \sqrt{\langle v^2\rangle} = \sqrt{\frac{3p}{\rho}} = \sqrt{\frac{3kT}{m_0}} = \sqrt{\frac{3RT}{\mu}}, \qquad \text{(P2.5.2.3)}$$

where $\rho = m_0 n$ is the gas density.

► **Internal energy of an ideal gas.** The *number of degrees of freedom* of a molecule, i, is an important characteristic of an ideal gas. A monatomic molecule has only three degrees of freedom; these correspond to translational motion in three spatial coordinates: $i = i_{\rm transl} = 3$. A rigid diatomic molecule has two additional, rotational degrees of freedom, associated with the rotation about two axes perpendicular to the axis of the molecule: $i = i_{\rm transl} + i_{\rm rot} = 3+2 = 5$. A rigid multiatomic molecule (whose atoms are not collinear) has three rotational degrees of freedom, and therefore $i = 6$. Classical statistical physics proves the *equipartition theorem* (also known as *law of equipartition*). According to this theorem, an average energy of $\frac{1}{2}kT$ per molecule is associated with any degree of freedom to which there corresponds a term like αx^2 or $\beta\dot{x}^2$ in the expression of the molecule energy. Formula (P2.5.2.1) is in complete agreement with this theorem. Accordingly, the average energy per one molecule and the total internal energy of the gas are expressed as

$$\langle\varepsilon\rangle = \tfrac{1}{2} i_{\rm phys} kT, \qquad U = N\langle\varepsilon\rangle = \tfrac{1}{2} i_{\rm phys}\nu RT. \qquad \text{(P2.5.2.4)}$$

The physical number of degrees of freedom, $i_{\rm phys}$, for rigid molecules coincides with the mathematical one. However, at sufficiently high temperatures ($T \sim 10^3$ K), when molecules can no longer be treated as rigid and vibrational energy must be taken into account, an average energy of kT will be associated with each vibrational degree of freedom. As a result, the total number of degrees of freedom becomes $i_{\rm phys} = i_{\rm transl} + i_{\rm rot} + 2i_{\rm vib}$, where $i_{\rm vib}$ is the mathematical number of vibrational degrees of freedom in a molecule. For an s-atom molecule, $i_{\rm vib} = 3s - (i_{\rm transl} + i_{\rm rot})$. In particular, for a nonrigid diatomic molecule, we have $i_{\rm vib} = 6 - (3+2) = 1$ and $i_{\rm phys} = 3 + 2 + 2\times 1 = 7$.

► **Heat capacities of an ideal gas.** Using (P2.5.2.4) and the definitions of heat capacities at constant volume and constant pressure as well as adiabatic index (see Subsection P2.2.3), we find that

$$C_V^{\rm m} = \frac{i_{\rm phys}}{2}RT, \qquad C_p^{\rm m} = \frac{i_{\rm phys}+2}{2}RT, \qquad \gamma = \frac{C_p}{C_V} = \frac{i_{\rm phys}+2}{i_{\rm phys}}. \qquad \text{(P2.5.2.5)}$$

At room temperatures, the heat capacity follows the model of rigid molecules; the rotational degrees of freedom are not excited, or, as is often said, “frozen.” However, at temperature increases to $\sim 10^3$ K, the heat capacity begins to increase, that is, vibrational degrees of freedom “unfreeze.” Furthermore, as temperature decreases to several tens of kelvins, the rotational degrees of freedom “freeze out,” which results in decreasing heat capacity. The phenomenon of *freezing-out of degrees of freedom* is explained by quantum mechanics: if the average energy of thermal motion, kT, is small compared to the distance to the nearest discrete energy level, the kind of motion concerned is not excited.

► **Mixture of ideal gases.** *Dalton’s law* reads: the pressure of a mixture of ideal gases equals the sum of their partial pressures. In particular, for two gases,

$$p = p_1 + p_2 = (n_1 + n_2)kT = (\nu_1 + \nu_2)RT.$$

The internal energy of a mixture equals the sum of the internal energies of its components: $U = \frac{1}{2}i_1\nu_1 RT + \frac{1}{2}i_2\nu_2 RT$. This formula allows us to introduce the so-called effective number of degrees of freedom, $i(\nu_1+\nu_2) = i_1\nu_1 + i_2\nu_2$, effective molar heat capacity, $(\nu_1+\nu_2)C_\mu = \nu_1 C_{\mu 1} + \nu_2 C_{\mu 2}$, and effective molar mass, $(\nu_1+\nu_2)\mu = \nu_1\mu_1 + \nu_2\mu_2$.

P2.5.3. Maxwell's Distribution

► **Definition and properties of distribution functions.** The distribution of molecules over their velocities is characterized by the following functions:

$$\begin{aligned}\varphi(v_x)\,dv_x &= \frac{dn(v_x,\,v_x+dv_x)}{n},\\ f(v)\,dv &= \frac{dn(v,\,v+dv)}{n},\\ \Phi(\mathbf{v})\,dv_x\,dv_y\,dv_z &= \frac{dn(v_x,\,v_x+dv_x;\;v_y,\,v_y+dv_y;\;v_z,\,v_z+dv_z)}{n}.\end{aligned}\qquad\text{(P2.5.3.1)}$$

The definition of each of the distribution functions is based on the assumption that the fraction of molecules that fall (on the average) within a very small interval of a given variable (e.g., velocity, its projection onto an axis, energy, etc.) is proportional to the width of this interval. Note that dv denotes a physically very narrow, rather than mathematically infinitesimal, interval but wide enough to contain a large number of molecules. The mean fraction of molecules having a certain property (e.g., having velocities within a given interval) can be treated as the probability that an arbitrary molecule possesses this property. Therefore, a distribution function is sometimes referred to as a *probability density function*.

Listed below are properties of a distribution function (exemplified by $f(v)$).

1. The fraction of particles (probability) having velocities within a finite interval (v_1, v_2):

$$\frac{\Delta n(v_1,v_2)}{n} = \int_{v_1}^{v_2} f(v)\,dv.$$

2. Normality:

$$\int_0^\infty f(v)\,dv = 1.\qquad\text{(P2.5.3.2)}$$

3. Computation of the average of any function $\chi(v)$:

$$\langle\chi(v)\rangle = \int_0^\infty \chi(v)\,f(v)\,dv.\qquad\text{(P2.5.3.3)}$$

The three distribution functions defined in (P2.5.3.1) are related by

$$f(v) = \Phi(v)4\pi v^2,\qquad \Phi(v_x,v_y,v_z) = \varphi(v_x)\varphi(v_y)\varphi(v_z).$$

The function φ is even, and hence one can write $\varphi = \varphi(v_x^2)$. The function $\Phi(\mathbf{v})$ depends on $v^2 = v_x^2+v_y^2+v_z^2$ only. The relation between Φ and φ is satisfied only if $\varphi(v_x) = A\exp(-\zeta v_x^2)$. The coefficients A and ζ are determined from two conditions: (a) the normality of φ and (b) the requirement that $\langle v_x^2\rangle = \frac{1}{3}\langle v^2\rangle = kT/m_0$; see (P2.5.2.3). Thorough calculations give the answer

$$\begin{aligned}\varphi(v_x) &= \Bigl(\frac{m_0}{2\pi kT}\Bigr)^{1/2}\exp\Bigl(-\frac{m_0v_x^2}{2kT}\Bigr),\\ f(v) &= \Bigl(\frac{m_0}{2\pi kT}\Bigr)^{3/2}\exp\Bigl(-\frac{m_0v^2}{2kT}\Bigr)4\pi v^2.\end{aligned}\qquad\text{(P2.5.3.4)}$$

The latter expression is referred to as *Maxwell's distribution* (Fig. P2.4). The function $f(v)$ attains its maximum at $v_{\rm mp} = (2kT/m_0)^{1/2}$, which is called the *most probable velocity*. The value of $f(v)$ at this point is $f(v_{\rm mp}) = 4e^{-1}(2\pi kT/m_0)^{-1/2}$. For example, it

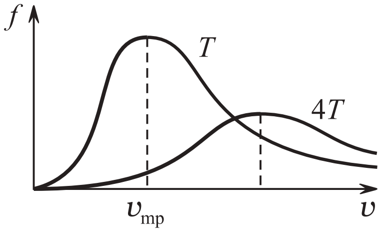

Figure P2.4. Maxwell's distribution function.

follows that if T increases by a factor of 4, the most probable velocity doubles and the value of the function, $f(v_{\rm mp})$, is halved; note that the area under the curve $f(v)$ is equal to one.

The *average velocity* (arithmetic mean) of molecules is calculated by (P2.5.3.3):

$$\langle v\rangle = \int_0^\infty v f(v)\,dv = \sqrt{\frac{8kT}{\pi m_0}} = \sqrt{\frac{8RT}{\pi\mu}}.$$

The distribution of molecules over the energies of their translational motion is determined by

$$\psi(\varepsilon)\,d\varepsilon = \frac{dn(\varepsilon,\,\varepsilon+d\varepsilon)}{n} = \frac{2}{\sqrt{\pi}}\exp\Bigl(-\frac{\varepsilon}{kT}\Bigr)\sqrt{\frac{\varepsilon}{kT}}\,d\Bigl(\frac{\varepsilon}{kT}\Bigr).$$

Useful integrals. In calculating averages using Maxwell's distribution, the following integrals often arise: $I_n = \int_0^\infty x^n \exp(-\beta x^2)\,dx$ $(n = 0, 1, 2, \dots)$. The first two integrals are $I_0 = \frac{1}{2}\sqrt{\pi}\,\beta^{-1/2}$ and $I_1 = \frac{1}{2}\beta^{-1}$. Other integrals can be calculated by differentiating I_0 or I_1 with respect to β. For example, $I_2 = -\dfrac{dI_0}{d\beta} = \frac{1}{4}\sqrt{\pi}\,\beta^{-3/2}$.

P2.5.4. Boltzmann's Distribution

► **Boltzmann's distribution.** If a gas is in an external force field, the concentration of molecules depends on the spatial position. It can be obtained from the condition of mechanical equilibrium of the gas that

$$n(\mathbf{r}) = n(\mathbf{r}_0)\exp\left[-\frac{\varepsilon_{\rm p}(\mathbf{r}) - \varepsilon_{\rm p}(\mathbf{r}_0)}{kT}\right], \qquad \text{(P2.5.4.1)}$$

where $\varepsilon_{\rm p}(\mathbf{r})$ is the potential energy of a molecule in the external field. Formula (P2.5.4.1) is known as *Boltzmann's distribution*. The barometric formula

$$p(h) = n(h)kT = p_0\exp\Bigl(-\frac{m_0gh}{kT}\Bigr), \quad \text{where} \quad n(h) = n_0\exp\Bigl(-\frac{m_0gh}{kT}\Bigr), \qquad \text{(P2.5.4.2)}$$

is a special case of Boltzmann's distribution.

Let us illustrate the derivation of Boltzmann's distribution using the barometric formula as an example. The equilibrium condition for a vertical cylinder with the base area s and height dh in the field of gravity is: $s\,dp = -\rho(s\,dh)g$. Then, using the equations $p = nkT$ and $\rho = m_0n$, we obtain $\dfrac{dn}{dh} = -\dfrac{m_0g}{kT}n$. Integrating yields the barometric formula.

► **Maxwell–Boltzmann distribution.** Both Maxwell's and Boltzmann's distributions, (P2.5.3.4) and (P2.5.4.1), contain the expression $\exp(-\varepsilon/kT)$. The combined Maxwell–Boltzmann distribution expresses the probability that an arbitrary molecule from a container with N molecules is in the spatial region $(x,\ x+dx;\ y,\ y+dy;\ z,\ z+dz)$ and has the velocity within $(v_x,\ v_x+dv_x;\ v_y,\ v_y+dv_y;\ v_z,\ v_z+dv_z)$:

$$\frac{dN}{N}=\frac{dn(\mathbf{r})\,d^3\mathbf{r}}{N}=\frac{n(\mathbf{r})\Phi(\mathbf{v})}{N}\,d^3\mathbf{r}\,d^3\mathbf{v}=A\exp\Bigl(-\frac{\varepsilon}{kT}\Bigr)d^3\mathbf{r}\,d^3\mathbf{v},\qquad\text{(P2.5.4.3)}$$

where $d^3\mathbf{r}=dx\,dy\,dz$, $d^3\mathbf{v}=dv_x\,dv_y\,dv_z$, $\varepsilon=\frac{1}{2}m_0v^2+\varepsilon_\text{p}(\mathbf{r})$ is the mechanical energy of a molecule, A is a coefficient that can be found from the normality condition. The Maxwell–Boltzmann distribution applies to any kinds of molecule energy: rotational, vibrational, potential, angular orientation dependent, and others.

P2.6. Real Gases. Van Der Waals Equation

P2.6.1. Isotherms of a Real Gas

► **Low temperatures, $T<T_\text{c}$.** If the gas temperature is less than the critical temperature (see below), then the following processes occur in equilibrium isothermal volume contraction (Fig. P2.5):

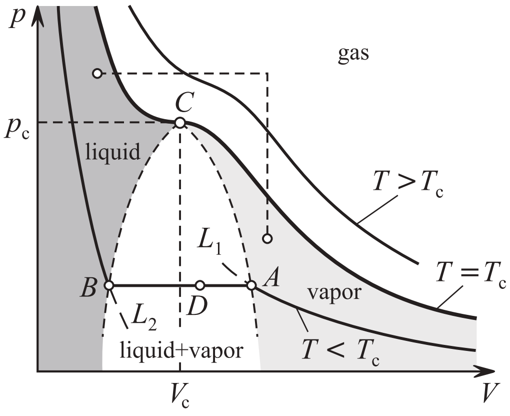

Figure P2.5. Isotherms of a real gas. ACB is the region of two-phase equilibrium.

1. Pressure increases to the point A. If temperature is well below the critical temperature, then the ideal gas law works well up to the point A.

2. Pressure does not change along AB. Some liquid emerges in the container; it is separated from the gas by a clear interphase surface. A gas which is in thermodynamic equilibrium with its liquid phase is called a *saturated vapor*. The liquid-to-gas mass ratio at a point D on the segment AB equals the ratio between the lengths of the segments AD and DB: $m_\text{liq}/m_\text{vap}=AD/DB$. The isotherm portions AL_1 and L_2B correspond to *metastable states*, or states at which the system can remain for some time but then quickly passes to a stable state on the segment AB. A gas and liquid in a metastable state are called a *supercooled vapor* and *superheated liquid* (at constant pressure). The existence of metastable states is accounted for by the fact that the formation of tiny nuclei of a new phase (liquid in vapor or vapor in liquid) requires a considerable amount of energy, known as surface energy (see Subsection P2.8.1).

3. After the point B, pressure sharply increases. There is only the liquid phase in the container.

► **Critical temperature, $T = T_c$.** The higher the temperature, the shorter the horizontal portion of the isotherm. At some *critical temperature*, T_c, it reduces to a mere point, C, the inflection point of the isotherm (Fig. P2.5). The pressure at the *critical point* is called the *critical pressure*, p_c, and the volume of a single mole of the gas is called the *critical molar volume*, V_c.

► **High temperatures, $T > T_c$.** For higher temperatures, $T > T_c$, the isotherm is a smooth monotonic curve. The region above the critical isotherm is called the true gas (it cannot be isothermally transferred to a liquid + vapor state). The region below the critical isotherm consists of three subregions: liquid, unsaturated vapor, and liquid + saturated vapor. It is possible to pass (non-isothermally) from the vapor region to the liquid region without entering the liquid + vapor region. So gas and liquid differ from each other by only quantitative characteristics.

The critical parameters of some substances are shown in the table below.

TABLE P2.1
Critical parameters of some substances.

Substance	T_c, K	p_c, MPa	V_c, cm^3/mol
Helium, He	5.19	0.277	57.4
Hydrogen, H_2	33.24	1.30	65.0
Neon, Ne	44.4	2.654	41.7
Nitrogen, N_2	126.3	3.40	90.1
Oxygen, O_2	154.6	5.04	78
Methane, CH_4	190.66	4.626	99.38
Krypton, Kr	209.39	5.49	91.9
Carbon dioxide, CO_2	304.1	7.38	94.0
Hydrogen sulphide, H_2S	373.6	9.01	97.4
Water, H_2O	647.3	22.12	56
Mercury, Hg	1460	166.1	48
Lithium, Li	3200	68.9	66

P2.6.2. Van Der Waals Equation

► **Van der Waals equation.** The equation of state for one mole of a van der Waals gas is

$$(V_\mu - b)\left(p + \frac{a}{V_\mu^2}\right) = RT. \qquad \text{(P2.6.2.1)}$$

Note that equation (P2.1.2.2) from Section P2.1 for ν moles is obtained by substituting $V_\mu = V/\nu$. The positive constant b reflects the fact that not all of the container volume is available for the motion of molecules due to their own volume. This constant is taken to be the total volume of all molecules multiplied by four: $b = 4N_A \times \frac{1}{6}\pi d^3$, where d is the diameter of a molecule. The term a/V_μ^2 reflects the fact that the total pressure decreases due to mutual interaction of molecules. This can be clearly seen from the expression of the internal energy of a van der Waals gas: $U = C_{\mu V}T - a/V_\mu$.

The average potential energy of interaction among the molecules is characterized by a model expression containing two terms, first corresponding to strong repulsion at small distances ($\sim$ 10^{-10} m) and second corresponding to van der Waals attractive forces at large distances (Fig. P2.6):

$$E_{\rm p} = \frac{\alpha}{r^{12}} - \frac{\beta}{r^6}.$$

The minimum of $E_{\rm p}$ corresponds to the average inter-molecule distance in the absence of thermal motion. For $kT \ll E_{\rm min}$, the motion of molecules reduces to vibrations about the bottom of the potential well (solid state). The case $kT \sim E_{\rm min}$ corresponds to liquid state and the case $kT \gg E_{\rm min}$ corresponds to gaseous state. Since the left branch of the potential curve, corresponding to repulsion, is nearly vertical, the minimum distance to which molecules approach each other before rebounding is weakly dependent on their average translational energy (see also Subsection P1.5.4).

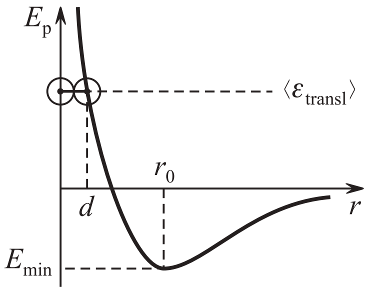

Figure P2.6. Potential energy of interaction of molecules.

► **Van der Waals isotherms.** Figure P2.7 depicts isotherms predicted by the van der Waals equation (P2.6.2.1). At temperatures below the critical value, the isotherms have two extrema, K_1 and K_2. However, for the path AK_1K_2B of a subcritical isotherm, the condition of mechanical stability, $(\partial p/\partial V)_T \le 0$, is violated. In fact, the transition between the points A and B occurs along a horizontal line. According to *Maxwell's rule*, the position of the segment AB is determined by the condition that shaded areas must be equal. Otherwise, the isothermal cycle AK_1K_2BA would perform work, which contradicts the second law of thermodynamics. The portions AK_1 and K_2B correspond to metastable states.

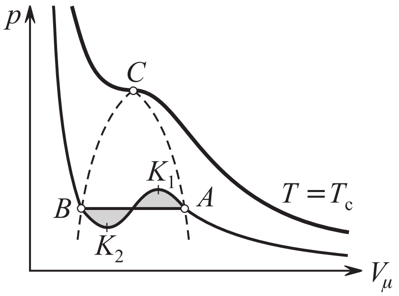

Figure P2.7. Van der Waals isotherms.

► **Calculation of critical parameters.** At the critical point, the following two conditions hold:

$$\left(\frac{\partial p}{\partial V}\right)_T = 0, \qquad \frac{\partial^2 p}{\partial V^2} = 0. \tag{P2.6.2.2}$$

Expressing p from (P2.6.2.1) and substituting into these equations, we find that

$$V_{\rm c} = 3b, \quad p_{\rm c} = \frac{1}{27}\frac{a}{b^2}, \quad T_{\rm c} = \frac{8}{27}\frac{a}{bR}.$$

The critical parameters satisfy the relation $p_c V_c = \frac{3}{8} R T_c$. By introducing the reduced state variables $v = V/V_c$, $\pi = p/p_c$, and $\theta = T/T_c$, one can rewrite the van der Waals equation (P2.6.2.1) as

$$\left(v - \frac{1}{3}\right)\left(\pi + \frac{3}{v^2}\right) = \frac{8}{3}\theta.$$

This equation is the same for all gases. It follows that the van der Waals equation satisfies the experimentally established *principle of corresponding states*. This principle asserts that if two respective reduced state variables for two gases coincide, the third will also coincide. This law is obeyed very well by most real gases.

P2.7. Phase Equilibrium. Phase Transitions

P2.7.1. Conditions for Phase Equilibrium

► **Phase.** A *phase* is a part of a system which is homogeneous in its physical and chemical properties. The same substance can be in different phases, although its chemical composition does not change. Examples of a *two-phase system* are: liquid + saturated vapor, solid + liquid, and solid + vapor.

► **Three equilibrium conditions.** For the phases of a system to be in equilibrium, the following conditions must hold: (1) the pressures must be equal (mechanical equilibrium), (2) the temperatures must be equal (thermal equilibrium), and (3) the chemical potentials must be equal (equilibrium with respect to interphase transitions). The last condition corresponds to the minimum of the Gibbs free energy $G = \mu_1 N_1 + \mu_2 (N - N_1)$ (see Subsection P2.4.6) with respect to the number of molecules in one of the phases, N_1. For a simple system, the condition $\mu_1(p,T) = \mu_2(p,T)$ determines the equation of the *two-phase coexistence curve* $p(T)$; a simple two-phase system has only one degree of freedom. In particular, the pressure of a saturated vapor, determined by the horizontal portion of the isotherm (see Subsection P2.6.1), is a single-valued function of temperature (the *evaporation–condensation curve*). The same applies to the dependence of the melting temperature on the external pressure (the *melting–freezing curve*) and also to the temperature dependence of the vapor pressure over a solid surface (the *sublimation–deposition curve*).

► **Triple point.** Figure P2.7 depicts a schematic phase diagram for water, which includes the three phase-equilibrium curves: evaporation–condensation (I), melting–freezing (II), and sublimation–deposition (III). The point of intersection of the three curves is called the *triple point* (for water, $T_3 = 0.01°\text{C}$ and $p_3 = 610.5$ Pa). The conditions $\mu_{\text{vap}}(p,T) = \mu_{\text{liq}}(p,T)$ and $\mu_{\text{liq}}(p,T) = \mu_{\text{sol}}(p,T)$ must hold simultaneously for a *three-phase system* to exist. The pressure and temperature, identifying the triple point, are determined uniquely by these conditions.

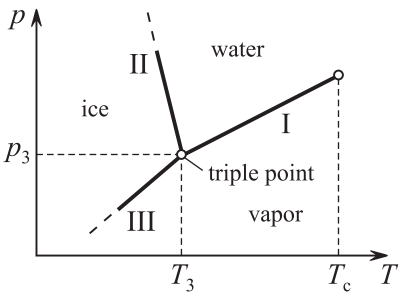

Figure P2.7. Phase diagram for water.

P2.7.2. First-Order Phase Transitions

► **Transition heat and entropy.** In order to transfer a portion of substance of mass m in a two-phase system from one phase to the other, the quantity of heat $Q = mq$ must be supplied (or removed) in an equilibrium isobar-isothermal process. This quantity of heat Q is called the *phase transition heat* or *latent heat*; the quantity q is the *specific phase transition heat* or *specific latent heat*. When a substance changes to a new phase, its entropy changes by $\Delta S = Q/T = mq/T$ and the internal energy, by $\Delta U = mq - mp\Delta v$, where $v = V/m$ is the *specific volume*. The quantity $\Delta s = q/T$ is called the specific entropy of the phase transition, and $\Delta u = q - p\Delta v$ is the change in the specific internal energy at the phase transition.

A transition between phases in a two-phase system in which some heat is absorbed or released and an abrupt change in internal energy, entropy, density, volume, etc. occurs is called a *first-order phase transition*. This name is due to the fact that the first derivatives of the Gibbs free energy, $S = -(\partial G/\partial T)_p$ and $V = (\partial G/\partial p)_T$, experience a jump discontinuity.

► **Clausius–Clapeyron relation.** If the temperature of a two-phase system is changed, then its pressure will adapt so that the chemical potentials (or the specific Gibbs free energies) of the phases remain the same: $dg_1 = dg_2$, or $-s_1\,dT + v_1\,dp = -s_2\,dT + v_2\,dp$ (see Subsection P2.4.6). It follows that the slope of the phase equilibrium curve equals

$$\frac{dp}{dT} = \frac{\Delta s}{\Delta v} = \frac{q}{T\Delta v}. \qquad \text{(P2.7.2.1)}$$

This equation is referred to as the *Clausius–Clapeyron relation*.

Example 1. For the ice–water phase transition at atmospheric pressure, the melting temperature is 273 K, $q = 3.34 \times 10^5$ J/kg, and $v_\mathrm{w} - v_\mathrm{i} = -0.91 \times 10^{-4}$ m^3/kg. Consequently $dp/dT = -1.1 \times 10^7$ Pa/K. The minus sign reflects the anomalous property of water: its density increases in melting.

Example 2. If heat is continuously supplied to a liquid, intensive evaporation begins across the whole volume when the temperature attains the value at which the saturation vapor pressure equals the external pressure. This process is called *boiling*. Equation (P2.7.2.1) determines the dependence between the *boiling temperature* and external pressure. Neglecting the specific volume of the liquid and expressing the specific volume of the vapor from the ideal gas law $v = \dfrac{RT}{\mu p}$, we arrive at the approximate equation $\dfrac{dp}{dT} = \dfrac{\mu q p}{RT^2}$. In particular, for water ($\mu = 18$ g/mol) at atmospheric pressure ($p = 10^5$ Pa), we have the boiling temperature $T = 373$ K and the specific latent heat of evaporation $q = 2.26 \times 10^6$ J/kg. If follows that $\dfrac{dp}{dT} = 3.5 \times 10^3$ Pa/K.

► **Second-order phase transitions.** In a second-order phase transition, density, internal energy, and entropy (the first derivatives of the Gibbs free energy) are changed continuously, while the various derivatives of these thermodynamic functions (such as heat capacity, compressibility, dielectric constant, magnetic permeability, etc.) experience a jump discontinuity or show critical behavior (tend to infinity when approaching the transition point). The name is due to the fact that the second derivatives of the Gibbs free energy experience a discontinuity. Examples of such transitions are the paramagnetic-to-ferromagnetic transition at the Curie point, transition of liquid helium to the superconducting state, superfluid transitions, etc. A second-order phase transition occurs at the whole volume at once and is associated with a qualitative change in the internal structure (such as appearance of additional symmetries or loss of some symmetries).

P2.8. Surface Tension

P2.8.1. Surface Energy

► **Surface tension coefficient**. Molecules at the surface of a liquid possess higher potential energy as compared to molecules in the bulk, since the former have fewer "neighbors," the energy of interaction with which is negative (see Fig. P2.6). At constant temperature, the additional energy is proportional to the surface area, Σ:

$$F_{\text{surf}} = \sigma(T)\Sigma, \tag{P2.8.1.1}$$

where σ is the surface free energy per unit area, called the *surface tension coefficient*.

In an equilibrium isothermal change of a rectangular area on the surface, the work done by external forces equals the change in the Helmholtz free energy (P2.4.6.2). So $A = \Delta F_{\text{surf}} = \sigma\Delta\Sigma = \sigma Lx$, where L is the rectangle side and x is its displacement. It follows that a *surface tension force* acts on the rectangle side which is perpendicular to the side and tangent to the liquid surface:

$$f_{\text{surf}} = \sigma L. \tag{P2.8.1.2}$$

In equilibrium, the Helmholtz free energy is minimum and, hence, a liquid of a liquid film tend to minimize their surface. For example, a drop of liquid tends to a spherical shape at zero gravity.

► **Thermodynamics of surface layer.** Since F_{surf} is a function of T and Σ, then $dF_{\text{surf}} = \sigma\,d\Sigma + \Sigma\dfrac{d\sigma}{dT}\,dT$. It follows, taking into account (P2.4.6.3), that the surface layer entropy and internal energy as well as the heat absorbed by the layer under isothermal area extension are expressed as

$$S = -\frac{d\sigma}{dT}\Sigma, \quad U = F + TS = \left(\sigma - T\frac{d\sigma}{dT}\right)\Sigma, \quad \delta Q = T\,dS = -T\frac{d\sigma}{dT}\,d\Sigma.$$

With increasing temperature T, the surface tension coefficient σ decreases and vanishes at a critical temperature.

P2.8.2. Additional Pressure Under a Curved Surface

► **Spherical surface.** It follows from the condition of mechanical equilibrium of the surface layer that the pressure inside a convex surface must be higher than outside. For a spherical surface of radius R, the additional pressure, Δp, can be found by increasing its radius by dR and equating the work done by external forces with the change in the Helmholtz energy: $\Delta p\,dV = \sigma\,d\Sigma$ and hence $\Delta p\,d(\frac{4}{3}\pi R^3) = \sigma\,d(4\pi R^2)$. Finally, we have

$$\Delta p = \frac{2\sigma}{R}. \tag{P2.8.2.1}$$

► **Young–Laplace formula.** The above equation can be generalized to include two different radii of curvature:

$$\Delta p = \sigma\left(\frac{1}{R_1} + \frac{1}{R_2}\right), \tag{P2.8.2.2}$$

where R_1 and R_2 are the principal radii of curvature of the surface (i.e., the minimum and maximum radii of curvature of the curves obtained by cutting the surface with two mutually perpendicular planes at a given point). Relation (P2.8.2.1) is called the *Young–Laplace formula* or *Young–Laplace equation*.

P2.8.3. Wetting. Capillary Effects

► **Contact angle.** When a liquid and solid are in contact with each other and both are in a gas environment, the liquid surface (liquid–gas interface) and the solid surface (solid–gas interface) meet at an angle, which is called the *contact angle*. The value of the contact angle, θ, depends on the surface tension coefficients at the liquid–gas ($\sigma_{\text{liq–gas}}$), liquid–solid ($\sigma_{\text{liq–sol}}$), and solid–gas ($\sigma_{\text{sol–gas}}$) interfaces. It follows from the condition of mechanical equilibrium at the line where the three phases meet (see Fig. P2.8) that $\sigma_{\text{sol–gas}} = \sigma_{\text{liq–sol}} + \sigma_{\text{liq–gas}} \cos\theta$ (per unit length). Hence,

$$\cos\theta = \frac{\sigma_{\text{sol–gas}} - \sigma_{\text{liq–sol}}}{\sigma_{\text{liq–gas}}}.$$

If $0 < \cos\theta < 1$, or $0 < \theta < 90^\circ$, the liquid is said to *wet* the solid surface. If $-1 < \cos\theta \le 0$, or $90^\circ \le \theta < 180^\circ$, it is said that the liquid does not wet the surface. If the right-hand side of the equation is greater than or equal to 1 ($\theta = 0$), the liquid forms a thin molecular film of the solid surface (this phenomenon is known as *perfect wetting*); in a container, the liquid makes a zero angle with the solid wall. If the right-hand side of the equation is less than or equal to –1 ($\theta = 180^\circ$), this situation is known as *perfect non-wetting*.

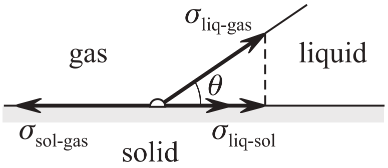
$\sigma_{\text{liq-gas}}$
gas
liquid
θ
$\sigma_{\text{sol-gas}}$
solid
$\sigma_{\text{liq-sol}}$

Figure P2.8. Contact angle.

► **Capillary effects.** If one end of a thin, capillary tube is immersed in a wetting liquid, the liquid will go up the capillary and reach a height h. This height is determined by the mechanical equilibrium condition.

If there is a capillary tube with inner radius r and the radius of curvature of the liquid meniscus is $R = r/\cos\theta$, the pressure under the meniscus (see (P2.8.2.1)) must be less than atmospheric pressure by the value $2\sigma/R = 2\sigma\cos\theta/r$. Equating the pressure difference with ρgh, we find the height of the liquid column: $h = \dfrac{2\sigma\cos\theta}{\rho gr}$.

P2.9. Transfer Phenomena in Gases

P2.9.1. Collision of Molecules in a Gas

► **Collision frequency.** For a given molecule in a gas, its average number of collisions per unit time (collision frequency) with other molecules is equal to

$$z = n\sigma\langle v\rangle\sqrt{2}, \tag{P2.9.1.1}$$

where n is the molecule concentration, $\langle v\rangle$ is the average speed of the molecules, and σ is the *effective cross-section* of elastic collisions. If the molecules are modeled as identical rigid balls of diameter d, we have $\sigma = \pi d^2$. The factor $\sqrt{2}$ accounts for the motion of oncoming molecules; this factor is usually omitted in the qualitative theory of transfer phenomena.

In order to estimate the collision frequency, the polygonal path of a molecule is usually surrounded by a tube of radius d. The moving molecule will collide with any resting molecule whose center is inside the tube. In a unit time, the molecule travels the distance $\langle v\rangle$, the corresponding tube volume equals $\pi d^2\langle v\rangle$, and the number of molecules in the tube is $z = \pi d^2\langle v\rangle n$.

► **Mean free path.** The average distance traveled by a molecule between successive collisions, the *mean free path*, is equal to

$$\lambda = \frac{\langle v\rangle}{z} = \frac{1}{n\sigma\sqrt{2}}. \tag{P2.9.1.2}$$

The total number of collisions among molecules in a unit volume per unit time is

$$\nu = \frac{zn}{2} = \frac{n\langle v\rangle}{2\lambda} = \frac{n^2\sigma\langle v\rangle\sqrt{2}}{2}. \tag{P2.9.1.3}$$

Effective cross-section. The effective cross-section of an interparticle collision process (with respect to a result of interest, such as scattering within a given solid angle, particle absorption, etc.) is determined as follows. Consider a fixed molecule (a scattering center) and an incident beam of molecules with flux $I = nv$ (the number of particles passing through a unit area per unit time, with the area perpendicular to the velocity vector). The effective cross-section of a process is the ratio between the number of collisions per unit time that have led to the desired result, ΔN, and the incident flux:

$$\sigma = \Delta N/I. \tag{P2.9.1.4}$$

The attenuation of the molecule beam traveling through a gas over a distance dx equals

$$dI = -I\sigma n\,dx = -(I/\lambda)\,dx,$$

where σ is the cross-section of elastic scattering, $n\,dx$ is the number of scattering centers along the path dx (per unit area), and $\lambda = 1/(n\sigma)$. Integrating yields the incident flux as a function of the distance traveled:

$$I = I_0\exp(-n\sigma x) = I_0\exp(-x/\lambda). \tag{P2.9.1.5}$$

It follows that the probability for a molecule to experience a collision on the interval $(x,\, x + dx)$ is $|dI|/I_0 = \exp(-x/\lambda)\,d(x/\lambda)$. The mean distance traveled by a molecule to the first collision equals

$$\langle x\rangle = \int_0^\infty x\exp\Bigl(-\frac{x}{\lambda}\Bigr)\frac{dx}{\lambda} = \lambda.$$

P2.9.2. Qualitative Model of Transfer Phenomena in Gases

The phenomena of viscosity, heat conduction, and diffusion are accounted for by the transfer of linear momentum, energy, and concentration due to chaotic thermal motion.

► **Viscosity. Fourier's law.** Viscosity is the appearance of frictional forces between parallel gas layers as they move relative to each other. If the directed flow velocity, $\mathbf{u}$, is parallel to the x-axis and only changes in the y-direction (Fig. P2.9), then a resistive force arises which is proportional to the area and the velocity gradient:

$$|F_x| = \eta S\frac{du}{dy}, \tag{P2.9.2.1}$$

where η is the *coefficient of viscosity* (or simply the *viscosity*). Equation (P2.9.2.1) is known as *Fourier's law*. The appearance of the resistive force is associated with the transfer of the momentum p_x along the y-axis by the molecules crossing the contact area due to their thermal motion. The force per unit area equals the momentum flux density.

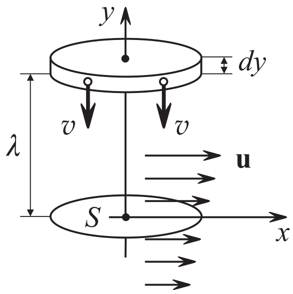

Figure P2.9. A frictional force arises between adjacent gas layers having different relative velocities.

► **Heat conduction. Newton's law of cooling.** If the gas temperature rather than the directed flow velocity changes along the y-axis, then the local heat flux density in the y-direction is expressed as

$$q = -\kappa \frac{dT}{dy}, \tag{P2.9.2.2}$$

where κ is the *thermal conductivity*. This equation is known as Newton's law of cooling.

► **Diffusion. Fick's law.** If the concentration of selected gas molecules, n_1, changes along the y-axis (the total concentration n must be constant), then the flux density of the selected molecules, j, in the y-direction is given by

$$j = -D \frac{dn_1}{dy}, \tag{P2.9.2.3}$$

where D is the *diffusion coefficient* (or *diffusivity*). This equation is known as *Fick's law of diffusion*. The quantity j is called the *diffusion flux density*.

► **Flux calculation formulas.** Let $g(y)$ denote a quantity whose transfer along the y-axis is of interest. Examples of such a quantity are the momentum of directed motion, $g(y) = m_0 u(y)$, for viscosity, the mean energy of thermal motion, $g(y) = \langle\varepsilon\rangle = c_V m_0 T$ (c_V is the specific heat capacity at constant volume), for heat conduction, and the relative concentration of selected molecules, $g(y) = n_1(y)/n$, for diffusion. Let us assume that each molecule transfers a certain (average) amount of the quantity g through a given surface; this amount of g was acquired by the molecule at its latest collision. Then the flux density of g in the y-direction can be calculated using the formula

$$G = -\frac{1}{3} n\langle v\rangle \lambda \frac{dg}{dy}. \tag{P2.9.2.4}$$

For rough estimates, it can be assumed that the particles move with velocity $\langle v\rangle$ in one of the six directions determined by the coordinate axes. Then the particle flux through a unit area is equal to $n\langle v\rangle$. It can also be assumed that the particles crossing the area the y-level experienced their latest collision at an average distance λ (Fig. P2.9). Then we have $G \approx \frac{1}{6} n\langle v\rangle [g(y-\lambda) - g(y+\lambda)]$, and hence $G \approx \frac{1}{3} n\langle v\rangle \lambda (dg/dy)$.

► **Formulas for transfer coefficients in a gas.**

1. Substituting $g = m_0 v_x$ into (P2.9.2.4), gives the momentum flux, which is equal to the frictional force (P2.9.2.1) for a unit area. Then the coefficient of viscosity is expressed as

$$\eta = \tfrac{1}{3} n m_0 \langle v\rangle \lambda = \tfrac{1}{3} \rho \langle v\rangle \lambda. \tag{P2.9.2.5}$$

2. Substituting $g = c_V m_0 T$ into (P2.9.2.4) and comparing with (P2.9.2.2), we find the thermal conductivity:

$$\kappa = \tfrac{1}{3} n m_0 \langle v \rangle c_V \lambda = \tfrac{1}{3} \rho \langle v \rangle c_V \lambda. \qquad \text{(P2.9.2.6)}$$

3. Substituting $g = n_1/n$ into (P2.9.2.4) and comparing with (P2.9.2.3) yields the diffusivity:

$$D = \tfrac{1}{3} \langle v \rangle \lambda. \qquad \text{(P2.9.2.7)}$$

There are two important conclusions that can be drawn from formulas (P2.9.2.5)–(P2.9.2.7):

1) The mean free path is inversely proportional to ρ, which can be seen from equation (P2.9.1.2); hence, the viscosity and thermal conductivity of gases at a given fixed temperature are pressure independent (Maxwell's law). This conclusion holds true as long as the mean free path is small compared to the container size.

2) The ratio between the viscosity and the thermal conductivity is temperature independent.

► **Extremely rarefied gas.** When the mean free path becomes commensurable with the container size (*rarefied gas*), the viscosity and thermal conductivity begin to depend on density. For example, consider the heat transfer between two walls having temperatures T_1 and T_2 and separated by a distance $d \ll \lambda$ (*extremely rarefied gas*). It can be assumed that, after its collision with the wall having the temperature T_1, each molecule acquires (on the average) the energy $c_V m_0 T_1$, then flies without collisions to the opposite wall and gives up the energy $c_V m_0 (T_2 - T_1)$. It follows from (P2.5.1.3) that the number of molecules reaching the wall in time Δt is equal to $\frac{1}{4} n \langle v \rangle S \Delta t$. Consequently, $q = \frac{1}{4} \rho \langle v \rangle c_V \Delta T$, and hence $\kappa = \frac{1}{4} \rho \langle v \rangle c_V d$.

Bibliography for Chapter P2

Alenitsyn, A. G., Butikov, E. I., and Kondratyev, A. S., *Concise Handbook of Mathematics and Physics*, CRC Press, Boca Raton, Florida, 1998.

Ansel'm, A. I., *Basics of Statistical Physics and Thermodynamics* [in Russian], Lan, St. Petersburg, 2007.

Benenson, W., Harris, J. W., Stocker, H., and Lutz, H. (Eds.), *Handbook of Physics*, Springer-Verlag, New York, 2002.

Borgnakke, C. and Sonntag, R. E., *Fundamentals of Thermodynamics, 7th edition*, Wiley, New York, 2008.

Callen, H. B., *Thermodynamics and an Introduction to Thermostatistics, 2nd edition*, John Wiley & Sons, New York, 1985.

Demtröder, W., *Molecular Physics: Theoretical Principles and Experimental Methods*, Wiley-VCH, New York, 2006.

Fermi, E., *Thermodynamics*, Dover Publications, New York, 1956.

Huang K., *Statistical Mechanics, 2nd edition*, Wiley, New York, 1987.

Irodov, I. E., *Physics of Macrosystems. Basic Laws* [in Russian], Fizmatlit, Moscow, 2006.

Kittel, C. and Kroemer, H., *Thermal Physics, 2nd Edition*, W. H. Freeman, San Francisco, California, 1980.

Landau, L. D. and Lifshitz, E. M., *Statistical Physics, Third Edition, Part 1: Volume 5 (Course of Theoretical Physics, Volume 5)*, Butterworth-Heinemann, Oxford, England, 1980.

Moran, J. M. and Shapiro, H. N., *Fundamentals of Engineering Thermodynamics, 6th Edition*, Wiley, New York, 2007.

Pathria, R. K., *Statistical Physics, 2nd edition*, Butterworth-Heinemann, Oxford, UK and Woburn, Massachusetts, 1996.

Pitaevskii, L. P. and Lifshitz, E. M., *Physical Kinetics: Volume 10 (Course of Theoretical Physics, Volume 10)*, Butterworth-Heinemann, Oxford, England, 1981.

Pitaevskii, L. P. and Lifshitz, E. M., *Statistical Physics, Part 2: Volume 9 (Course of Theoretical Physics, Volume 9)*, Butterworth-Heinemann, Oxford, England, 1980.

Plischke, M. and Bergersen, B., *Equilibrium Statistical Physics*, Prentice-Hall, New York, 1989.

Prokhorov, A. M. (Ed.), *Encyclopedia of Physics, Volumes 1–5* [in Russian], Bolshaya Russkaya Encyclopedia, Moscow, 1998.

Prokhorov, A. M. (Ed.), *Encyclopedic Dictionary of Physics* [in Russian], Sovetskaya Encyclopedia, Moscow, 1984 (dic.academic.ru).

Reichl, L. E., *A Modern Course on Statistical Physics*, Wiley-Interscience, New York, 1998.

Schroeder, D. V., *An Introduction to Thermal Physics*, Addison-Wesley, Reading, Massachusetts, 1999.

Sivukhin, D. V., *General Physics, Volume 2. Thermodynamics and Molecular Physics* [in Russian], Fizmatlit, Moscow, 2006.

Sivukhin, D. V. and Yakovlev, I. A., *A Collection of Problems, Volume 2. Thermodynamics and Molecular Physics* [in Russian], Fizmatlit, Moscow, 2006.

Van Ness, H. C., *Understanding Thermodynamics*, Dover Publications, New York, 1983.

Woan, G., *The Cambridge Handbook of Physics Formulas*, Cambridge University Press, Cambridge, England, 2003.

Chapter P3

Electrodynamics

P3.1. Electric Charge. Coulomb's Law

P3.1.1. Electric Charge and Its Properties

► **Electric charge** is a scalar physical quantity that determines the electromagnetic interaction of subatomic particles. In SI, the unit of charge is the *coulomb* (C); it is defined in terms of the unit of electric current, the *ampere* (A): 1 C = 1 A s.

Properties of electric charge:

1. The carriers of electric charge are many subatomic particles such as protons and electrons (as well as antiprotons and positrons) and some metastable particles as π-mesons, K-mesons, muons, etc. Charged particles interact with each other with forces that decay with the distance as slowly as gravitational forces but are much stronger in magnitude.

2. Proton and electron and many other charged subatomic particles have the same electric charge in magnitude. This charge is called the *elementary charge* and denoted e. In SI, $e = 1.602\times 10^{-19}$ C. As experiments have shown, the charge of subatomic particles is independent of their velocity.

3. The charge of subatomic particles can be either positive or negative. Particles with like charges repel one another and particles with unlike charges attract. By convention, the charge of a proton is positive ($+e$) and the charge of an electron is negative ($-e$).

If a macroscopic body includes different numbers of electrons, N_e, and protons, N_p, the body is charged and its charge is always multiple of the elementary charge: $q = e(N_p - N_e)$.

► **Law of conservation of electric charge.** A system is called *electrically closed* if it does not exchange charges with external bodies, which means that charges do not cross its boundary. The total charge of an electrically closed system, or the algebraic sum of all its charges, remains constant. This statement is obvious if there are no elementary particle reactions. However, the law of conservation of charge is fundamental—it holds even if charged particles are created or destroyed.

► **Charge distribution.** The spatial distribution of charge is characterized by its *volume density* $\rho(\mathbf{r})$, measured in $\mathrm{C/m^3}$, *surface density* $\sigma(\mathbf{r})$, measured in $\mathrm{C/m^2}$, and *linear density* $\lambda(\mathbf{r})$, measured in $\mathrm{C/m}$. We have

$$dq = \rho\, dV, \qquad dq = \sigma\, dS, \qquad dq = \lambda\, dl. \tag{P3.1.1.1}$$

► **Dipole. Dipole moment.** A system of charges is called *electrically neutral* if its net charge is zero. An *electric dipole* is a simple example of an electrically neutral system; it consists of two point charges, q and $-q$, separated by a distance l. The *electric dipole moment* is a vector quantity defined as

$$\mathbf{p} = q\mathbf{l}, \tag{P3.1.1.2}$$

where $\mathbf{l} = \mathbf{r}_+ - \mathbf{r}_-$ is the displacement vector from the charge $-q$ to the charge q. The dipole moment is measured in C m.

The following definition is a generalization to the case of an electrically neutral system:

$$\mathbf{p} = \sum_i q_i \mathbf{r}_i = q(\mathbf{r}_+ - \mathbf{r}_-),$$

where $q = \sum_{q_i>0} q_i$ is the sum of all positive charges, $\mathbf{r}_+ = \sum_{q_i>0} q_i \mathbf{r}_i / q$ is the center of positive charges, and $\mathbf{r}_- = -\sum_{q_i<0} q_i \mathbf{r}_i / q$ is the center of negative charges.

P3.1.2. Coulomb's Law

Coulomb's law describes the electrostatic interaction of *point charges*, i.e., charged particles and bodies whose dimensions can be neglected as compared to the distances between them. Coulomb's law states that, in vacuum, the electrostatic force, F, acting on a point charge q_1 due to the presence of a point charge q_2 is directly proportional to each of the charges and inversely proportional to the square of the distance r between them:

$$F = k\frac{|q_1|\,|q_2|}{r^2}, \qquad \text{(P3.1.2.1)}$$

where k is the proportionality coefficient, dependent on the unit system chosen. The force is directed along the straight line connecting the point charges; like charges repel and unlike charges attract.

In SI, the coefficient in (P3.1.2.1) equals $k = 9.0 \times 10^9\ \mathrm{N\,m^2/C^2}$. Also $k = 1/(4\pi\varepsilon_0)$, where $\varepsilon_0 = 8.85 \times 10^{-12}\ \mathrm{F/m}$ is the *electric constant*. In Gaussian units (CGS system), k is taken to be equal to one and the unit of charge is defined by Coulomb's law.

The electrostatic force acting on a charge q due to the presence of many point charges, $Q_1, Q_2, \ldots$, equals the linear vector superposition of forces

$$\mathbf{F} = \sum_i k\frac{qQ_i}{(\mathbf{r} - \mathbf{r}_i)^2}\,\frac{\mathbf{r} - \mathbf{r}_i}{|\mathbf{r} - \mathbf{r}_i|}, \qquad \text{(P3.1.2.2)}$$

where $\mathbf{r}$ is the position vector of the charge q and $\mathbf{r}_i$ is the position vector of the charge Q_i $(i = 1, 2, \ldots)$.

► **On unit conventions in electrostatics.** In what follows, all formulas will be given in SI. To convert to Gaussian units (CGS system), it will be sufficient to set $k = 1$ in most cases; if a formula contains ε_0, one should first replace ε_0 with $1/(4\pi k)$.

P3.2. Electric Field. Strength and Potential

P3.2.1. Strength of Electric Field

► **Electromagnetic field.** The interaction between charged particles is implemented through *electromagnetic field*. This means that: (a) a charged particle generates an electromagnetic field in the surrounding space and (b) the electromagnetic field exerts a force on other charged particles. The field generated by a point charge is proportional to its charge; the force exerted by the field on a particle is proportional to the charge of this particle.

► **Electric field.** An electromagnetic field is a combination of an *electric field* and a *magnetic field*, which are closely related to each other. The action of an electric field on a charged particle does not depend on the particle velocity and that of a magnetic field does. The sources of electric fields are any charged particles, while magnetic fields are generated by moving charges. The proportion between electric and magnetic fields changes when passing from one inertial frame to another. The properties of electric and magnetic fields are completely determined by Maxwell's equations (see Subsection P3.11.1). Maxwell's equations are invariant under relativistic, Lorentz transformations.

► **Test charge. Strength of an electric field.** The characteristics of an electromagnetic field are studied using the concept of a *test charged particle*, which is an idealized small object whose dimensions and charge are small enough so that, when placed into a field of interest, it does not distort the field (does not change the positions of the field sources). The force acting on a stationary test particle of charge q is proportional to the charge and is only determined by the electric field:

$$\mathbf{F}_q = q\mathbf{E}. \qquad \text{(P3.2.1.1)}$$

This formula defines the *strength of an electric field*, **E**. It also answers the question what force the field exerts on any charge q, moving or stationary. In SI, the strength of an electric field is measured in N/C or V/m. For brevity, the term electric field is used to mean the strength of an electric field whenever this does not lead to confusion.

► **Electrostatic field. Principle of superposition.** A stationary electric field generated by a system of stationary charges is called an *electrostatic field*. The strength of an electrostatic field can be determined using Coulomb's law (P3.1.2.2).

1. *Electric field of a point charge.* If a point charge q is placed in the field of a point charge Q, then it follows from (P3.1.2.2) and (P3.2.1.1) that

$$\mathbf{E} = k\frac{Q}{r^2}\frac{\mathbf{r}}{r} = E_r\frac{\mathbf{r}}{r}, \qquad \text{(P3.2.1.2)}$$

where E_r is the projection of **E** onto the radial direction.

2. *Electric field of a system of point charges.* Using the principle of superposition, we get

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 + \cdots = \sum_i k\frac{Q_i}{(\mathbf{r}-\mathbf{r}_i)^2}\frac{\mathbf{r}-\mathbf{r}_i}{|\mathbf{r}-\mathbf{r}_i|}. \qquad \text{(P3.2.1.3)}$$

For a continuously distributed charge, the summation must be replaced by integration (see (P3.1.1.1)). (Compare with the formulas for the gravitational field in Subsection P1.7.2.)

Example 1. There is a charge continuously distributed along an arc of a circle of radius R with a linear density λ (Fig. P3.1). The arc angle is $2\alpha_0$. Find the strength of the electric field generated by the charge, at the circle center.

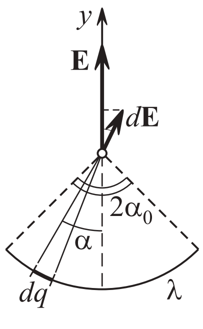

Figure P3.1. Strength of the electric field generated by a charged arc of a circle.

Solution. By symmetry, the field strength is directed along the y-axis, which passes through the midpoint of the arc and the circle center O. The strength of the field generated by the arc element $d\alpha$ at O equals $dE = k\,dq/R^2$, where $dq = \lambda R\,d\alpha$. Projecting onto y and integrating yields

$$E = E_y = \int_{-\alpha_0}^{\alpha_0} k\frac{\lambda R\,d\alpha}{R^2}\cos\alpha = k\frac{2\lambda\sin\alpha_0}{R}.$$

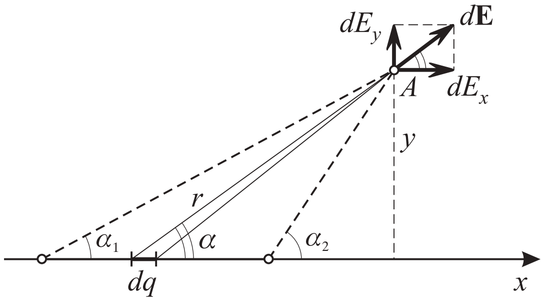

Figure P3.2. Strength of the electric field generated by a charged segment.

Example 2. Field of a segment. A charge is evenly distributed along a segment with a linear density λ. Find the field strength at the point A whose position is determined by the distance y to the straight line of the segment and two angles α_1 and α_2 as shown in Fig. P3.2.

Solution. The contribution of the element dx to the field strength at A equals $dE = k\lambda\,dx/r^2$, where r is the distance from dx to A. The respective projections on the x- and y-axes are $dE\cos\alpha$ and $dE\sin\alpha$. It is convenient to integrate with respect to the angle α. Substituting $r = y/\sin\alpha$ and $x = -y\cot\alpha$, with $dx = y\,d\alpha/\sin^2\alpha$, and integrating, we obtain the following expressions of E_x and E_y:

$$E_x = k\frac{\lambda}{y}(\sin\alpha_2 - \sin\alpha_1), \qquad E_y = k\frac{\lambda}{y}(\cos\alpha_1 - \cos\alpha_2).$$

If $\alpha_1 = \pi - \alpha_2 = \alpha$ (the point A lies on the straight line perpendicular to the segment through its midpoint), then $E_x = 0$ and $E_y = 2k\lambda\cos\alpha/y$. If $\alpha \to 0$ (the segment becomes a straight line), then $E_y = 2k\lambda/y$. If $\alpha_1 = 0$ and $\alpha_2 = \pi/2$ (a ray), then $E_x = E_y = k\lambda/y$.

P3.2.2. Potential of an Electrostatic Field

► **Definition of potential.** An electrostatic field is *potential* (see Subsection P1.5.3) just as any static central field. This means that the work done by the field to move a test charge from one point in space to another is independent of the path. This allows us to define the *potential of an electrostatic field*, $\varphi(\mathbf{r})$. We have

$$A_q(1 \to 2) = W_q(\mathbf{r}_1) - W_q(\mathbf{r}_2) = q\bigl[\varphi(\mathbf{r}_1) - \varphi(\mathbf{r}_2)\bigr], \qquad \text{(P3.2.2.1)}$$

where $A_q(1 \to 2)$ is the work done by the field to transfer the charge q from point $\mathbf{r}_1$ to point $\mathbf{r}_2$, $W_q(\mathbf{r}) = q\varphi(\mathbf{r})$ is the *potential energy of the charge in the electrostatic field* (throughout this chapter, energy will be denoted by W). The scalar quantity $\varphi(\mathbf{r}_1) - \varphi(\mathbf{r}_2)$ is called the *potential difference* between points $\mathbf{r}_1$ and $\mathbf{r}_2$. For $W_q(\mathbf{r})$ and $\varphi(\mathbf{r})$ to be uniquely defined, one has to choose a point where they are zero. In SI, potential is measure in volts (V).

► **Relation between field strength and potential.** Given the field strength $\mathbf{E}(\mathbf{r})$, one can determine the potential difference and potentials as

$$\varphi(\mathbf{r}_1) - \varphi(\mathbf{r}_2) = \int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{E}\,d\mathbf{r}, \qquad \varphi(\mathbf{r}) = \int_{\mathbf{r}}^{\mathbf{r}_0} \mathbf{E}\,d\mathbf{r}, \qquad \text{(P3.2.2.2)}$$

where $\mathbf{r}_0$ is the reference point where the potential is taken to be zero. Given the potential $\varphi(\mathbf{r})$, one can find the projection of the field strength onto any direction l (based on (P3.2.2.2) for two close points):

$$E_l = -\frac{\partial\varphi}{\partial l}. \qquad \text{(P3.2.2.3)}$$

Consequently, we can find the field strength vector:

$$\mathbf{E} = -\operatorname{grad}\varphi = -\left(\mathbf{i}\frac{\partial\varphi}{\partial x} + \mathbf{j}\frac{\partial\varphi}{\partial y} + \mathbf{k}\frac{\partial\varphi}{\partial z}\right), \qquad \text{(P3.2.2.4)}$$

where $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are the unit vectors of the respective coordinate axes.

► Principle of superposition for potential.

1. The *potential of the field of a point charge* Q can be calculated using (P3.2.2.3):

$$E_r = -\frac{d\varphi}{dr} \quad\Longrightarrow\quad \frac{d\varphi}{dr} = -k\frac{Q}{r^2} \quad\Longrightarrow\quad \varphi(r) = k\frac{Q}{r} + \text{const.}$$

Usually, the potential is taken to be zero at infinity. In this case,

$$\varphi(r) = k\frac{Q}{r}. \tag{P3.2.2.5}$$

2. The *potential of the field of a system of point charges* equals the algebraic sum of the potentials of individual charges (principle of superposition for potential):

$$\varphi(\mathbf{r}) = \varphi_1(\mathbf{r}) + \varphi_2(\mathbf{r}) + \cdots, \qquad \varphi(\mathbf{r}) = \sum_i k\frac{Q_i}{|\mathbf{r}-\mathbf{r}_i|}. \tag{P3.2.2.6}$$

For a continuously distributed charge, the summation must be replaced by integration.

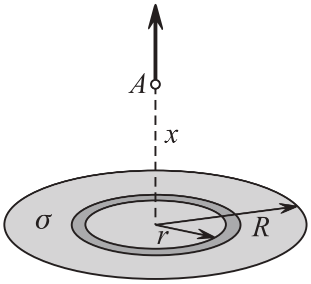

Figure P3.3. To the calculation of the electrostatic field of a charged disk.

Example 1. Field of a disk. A thin disk of radius R is evenly charged with a surface density σ. Find the potential and field strength at the point A located on the disk axis at a distance x from the center (Fig. P3.3).

Solution. The contribution of a thin ring bounded by the circles of radii r and $r + dr$ is

$$d\varphi = k\frac{dq}{\sqrt{x^2+r^2}} = k\frac{\sigma\, 2\pi r\, dr}{\sqrt{x^2+r^2}};$$

note that all points of the ring are essentially at the same distance to A. Integrating yields the desired potential

$$\varphi = \int_0^R k\frac{\sigma\, 2\pi r\, dr}{\sqrt{x^2+r^2}} = 2k\pi\sigma\bigl(\sqrt{x^2+R^2} - x\bigr).$$

Using (P3.2.2.3), we find the field strength (it is directed along x):

$$E_x = -\frac{d\varphi}{dx} = 2\pi k\sigma\left(1 - \frac{x}{\sqrt{x^2+R^2}}\right).$$

In the limit $R \to \infty$, we arrive at the field strength of a plane: $E_x = 2\pi k\sigma = \sigma/(2\varepsilon_0)$.

Example 2. Field of a dipole. Find the potential and field strength far away from a dipole shown in Fig. P3.4.

Solution. The potential at a remote point determined by $\mathbf{r}$ ($r = |\mathbf{r}| \gg l$) equals

$$\varphi(\mathbf{r}) = \frac{kq}{r_1} - \frac{kq}{r_2} \approx \frac{kq}{r-s} - \frac{kq}{r+s} \approx k\frac{(\mathbf{p}\cdot\mathbf{r})}{r^3},$$

where $s = \bigl(\frac{1}{2}\mathbf{l}\cdot\frac{1}{r}\mathbf{r}\bigr)$ and $\mathbf{p} = q\mathbf{l}$ (see Subsection P3.1.1).

The field strength can be found using the identities $\operatorname{grad}(1/r^3) = -3\mathbf{r}/r^5$ and $\operatorname{grad}(\mathbf{p}\cdot\mathbf{r}) = \mathbf{p}$. We have

$$\mathbf{E} = -\operatorname{grad}\varphi = k\left[\frac{3(\mathbf{p}\cdot\mathbf{r})\mathbf{r}}{r^5} - \frac{\mathbf{p}}{r^3}\right]. \tag{P3.2.2.7}$$

It follows that the magnitude of the field strength equals $E = kpr^{-3}\sqrt{3\cos^2\theta + 1}$.

The same expressions hold true for any electrically neutral system with a nonzero dipole moment (see Subsection P3.1.1) at large distances.

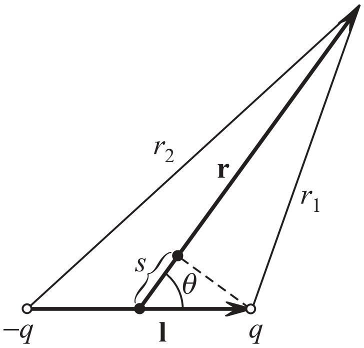

Figure P3.4. To the calculation of the field of a dipole.

► **Dipole in an external field.** A small dipole with dipole moment $\mathbf{p} = q\mathbf{l}$ in an external electrostatic field has the potential energy

$$W_\text{p} = q\varphi(\mathbf{r}_+) - q\varphi(\mathbf{r}_-) = q\frac{\partial\varphi}{\partial l}l = -pE_l = -(\mathbf{p}\cdot\mathbf{E}). \qquad \text{(P3.2.2.8)}$$

This energy changes if the dipole is turned by an angle—the field exerts a rotational moment on the dipole:

$$\mathbf{M} = \mathbf{l}\times(q\mathbf{E}) = \mathbf{p}\times\mathbf{E}, \qquad \text{(P3.2.2.9)}$$

The energy also changes if the dipole moves in a nonuniform field—the field exerts the force

$$\mathbf{F} = q\mathbf{E}(\mathbf{r}_+) - q\mathbf{E}(\mathbf{r}_-) = q\frac{\partial\mathbf{E}}{\partial l}l = p\frac{\partial\mathbf{E}}{\partial l} = p_x\frac{\partial\mathbf{E}}{\partial x} + p_y\frac{\partial\mathbf{E}}{\partial y} + p_z\frac{\partial\mathbf{E}}{\partial z}. \qquad \text{(P3.2.2.10)}$$

An equilibrium orientation with minimum potential energy corresponds to $\mathbf{p}\parallel\mathbf{E}$. A dipole oriented along the field strength vector is pulled into an area where the field is stronger.

P3.3. Gauss's Theorem

P3.3.1. Statement of the Theorem

► **Flux of electric field.** The flux of the electric field $\mathbf{E}$ through a plane area element ds is defined as

$$d\Phi = (\mathbf{E}\cdot d\mathbf{s}) = E_n ds = E\cos\theta\, ds, \qquad \text{(P3.3.1.1)}$$

where $d\mathbf{s} = \mathbf{n}\,ds$, $\mathbf{n}$ is a normal vector to the plane, θ is the angle between $\mathbf{E}$ and $\mathbf{n}$, $E_n = E\cos\theta$ is the projection of $\mathbf{E}$ onto the normal $\mathbf{n}$. The sign of $d\Phi$ depends on how the normal direction is chosen. The flux of the field $\mathbf{E}$ through a finite surface S is determined as

$$\Phi = \int_S E_n\, ds. \qquad \text{(P3.3.1.2)}$$

If the surface is closed, the normal is taken to point outward.

Flux additivity: if an electrostatic field is a superposition of several fields, $\mathbf{E}(\mathbf{r}) = \sum_i \mathbf{E}_i$, then $\Phi = \sum_i \Phi_i$ for any surface.

The flux of the electric field generated by point charge Q through an area ds is

$$d\Phi = E_r\cos\theta\, ds = \pm k\frac{Q}{r^2}\, ds_\perp = \pm kQ\, d\Omega,$$

where $ds_\perp = |ds\cos\theta|$ is the projection of ds onto the plane perpendicular to **r**, and $d\Omega = ds_\perp/r^2$ is the solid angle at which the area ds is seen from the charge Q. The plus sign corresponds to an acute angle between **n** and **r**, and the minus sign corresponds to an obtuse angle. The field flux through a closed surface is equal to $4\pi kQ$ if the charge is inside the surface, and zero if it is outside.

► **Gauss's theorem.** The flux of an electrostatic field through a closed surface is only determined by the net charge, $q_{\rm enc}$, enclosed within the surface:

$$\Phi = 4\pi k q_{\rm enc} = \frac{q}{\varepsilon_0}. \tag{P3.3.1.3}$$

► **Gauss's theorem and electric field lines.** Electric field lines are useful to visualize electric fields. These are drawn using the following rules: (a) the tangent to a field line is directed along the vector **E** at every point in space and (b) the density of field lines is proportional to the magnitude of the field at point of interest. The field flux through a surface is proportional to the number of field lines piercing the surface. (Lines of any vector field can be introduced likewise.) Gauss's theorem implies that the electric field lines must start and finish at charges and also they must be continuous in an empty space. It is noteworthy that Gauss's theorem and continuity of field lines are corollaries of the fact that the Coulomb force decays with distance as $1/r^2$.

Apart from field lines, *equipotential surfaces* are also used for visualization of electric fields. Equipotential surfaces are everywhere perpendicular to field lines (the work done by the field to transfer a test charge along such a surface must be zero).

P3.3.2. Application of Gauss's Theorem

► **Calculating the field using Gauss's theorem.** Suppose, from symmetry considerations, it is possible to construct an imaginary surface such that **E** is perpendicular to it and constant in magnitude on some part of the surface and is parallel to the surface (**E** $\perp$ **n**) on the other part. Such a surface is called a *Gaussian surface*. Gauss's theorem can be applied to this surface to find the field magnitude E.

Example 1. Field of a charged string. Find the electric field generated by an infinitely long, evenly charged straight string (Fig. P3.5).

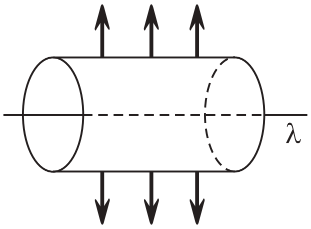

Figure P3.5. To the calculation of the electric field generated by an infinite string using Gauss's theorem.

Solution. The electric field is perpendicular to the string at any point in space. The field magnitude E_r is only dependent on the distance r to the string. Evidently, any cylinder of radius r with generatrix of length l and the axis lying on the string, can be taken as a Gaussian surface. Writing Gauss's theorem (P3.3.1.2), we find

$$E_r\,2\pi rl = \frac{\lambda l}{\varepsilon_0} \quad\Longrightarrow\quad E_r = \frac{\lambda}{2\pi\varepsilon_0 r} = \frac{2k\lambda}{r}$$

(cf. example 2 from Subsection P3.2.1). The potential difference between two points equals

$$\varphi_1 - \varphi_2 = \frac{\lambda}{2\pi\varepsilon_0}\ln\frac{r_2}{r_1}.$$

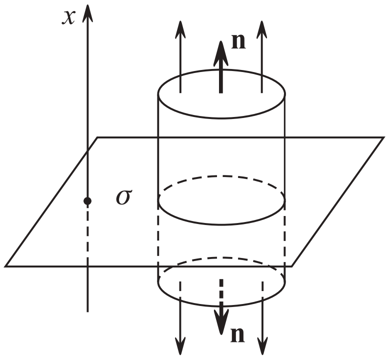

Figure P3.6. To the calculation of the electric field generated by a charged plane using Gauss's theorem.

Example 2. Field of a charged plane. Find the electric field generated by an evenly charged plane (Fig. P3.6).

Solution. The field is everywhere perpendicular to the plane and E_x is only dependent on the distance to the plane. A right cylinder whose bases of area S are symmetric with respect to the plane can be taken as a Gaussian surface. The field flux through this surface equals $E_x(x)S - E_x(-x)S = 2E_xS$. Writing Gauss's theorem, we find that

$$2E_xS = \frac{\sigma S}{\varepsilon_0} \quad\Longrightarrow\quad E_x(x) = -E_x(-x) = \frac{\sigma}{2\varepsilon_0}$$

(see example 1 from Subsection P3.2.2). Taking $\varphi = 0$ at $x = 0$, we have $\varphi(x) = \dfrac{\sigma}{2\varepsilon_0}|x|$.

Example 3. Field of a charged ball. Find the electric field of a uniformly charged ball of radius R with charge q.

Solution. The field of a uniformly charged ball is a central field. A concentric sphere of radius r can be taken as a Gaussian surface. If $r > R$, then the surface encloses the entire ball and so $q_{\rm enc} = q$. Hence, $E_r\,4\pi r^2 = q/\varepsilon_0$. If $r < R$, then $q_{\rm enc} = qr^3/R^3$ and hence $E_r\,4\pi r^2 = qr^3/(\varepsilon_0 R^3)$. Finally, we have

$$\mathbf{E} = \begin{cases} k\dfrac{q}{r^2}\dfrac{\mathbf{r}}{r} = \dfrac{\rho R^3}{3\varepsilon_0 r^2}\dfrac{\mathbf{r}}{r} & \text{for } r > R, \\ k\dfrac{q\mathbf{r}}{R^3} = \dfrac{\rho\mathbf{r}}{3\varepsilon_0} & \text{for } r < R, \end{cases}$$

where ρ is the charge density. The radial component of the field equals $E_r = -d\varphi/dr$ (see Subsection P3.2.2). Hence,

$$\varphi = \begin{cases} k\dfrac{q}{r} & \text{for } r > R, \\ -k\dfrac{qr^2}{2R^3} + \text{const} = -k\dfrac{qr^2}{2R^3} + \dfrac{3}{2}k\dfrac{q}{R} & \text{for } r < R. \end{cases}$$

The unknown constant is determined from the continuity condition for the potential at the ball surface.

► **Earnshaw's theorem.** Earnshaw's theorem is a corollary of Gauss's theorem. It states that a charged particle in vacuum cannot be in a stable equilibrium under the action of electric forces only. Indeed, if this were the case, the electric fields generated by the other charges would all have to point inward a small sphere enclosing the particle (if its charge is positive). But this would contradict Gauss's theorem (all other charges are outside the sphere).

► **Maxwell's equations in electrostatics.** The condition that an electrostatic field must be potential,

$$\oint \mathbf{E}\cdot d\mathbf{r} = 0 \quad\Longrightarrow\quad \operatorname{curl}\mathbf{E} = \mathbf{0},$$

is one of Maxwell's equations (in integral and differential forms) written for the special case of electrostatics.

The second equation is Gauss's theorem (written for a spatially distributed charge):

$$\oint \mathbf{E}\cdot d\mathbf{s} = \frac{1}{\varepsilon_0}\int \rho\, dV \quad\Longrightarrow\quad \operatorname{div}\mathbf{E} = \frac{\rho}{\varepsilon_0}.$$

Substituting $\mathbf{E} = -\operatorname{grad}\varphi$ in the last equation yields *Poisson's equation for electric potential*:

$$\operatorname{div}\operatorname{grad}\varphi = -\frac{\rho}{\varepsilon_0}, \qquad \text{or} \qquad \Delta\varphi = -\frac{\rho}{\varepsilon_0}.$$

P3.4. Conductors and Capacitors

P3.4.1. Electrostatics of Conductors

► **Field in a conductor. The potential of a conductor.** A conductor is a material that is capable of carrying electric current. To this end, the material must contain movable electric charges that can travel across. Electrostatics deals with states where charges are at equilibrium, when there is no directed motion of charged particles—electric current. This means that the electric field must be zero everywhere within the conductor. It follows that:

1. The potential of all point within the conduction has the same value, which is called the *potential of the conductor*.

2. The conductor surface is an equipotential surface. The electric field lines are perpendicular to it.

3. The volume charge density in the conductor is zero (it follows from Gauss's theorem that it must be zero charge in any volume within the conductor). The uncompensated charges must be distributed on the surface.

4. The electric field near the conductor surface is related to the local surface charge density σ by

$$E = \frac{\sigma}{\varepsilon_0}. \qquad \text{(P3.4.1.1)}$$

This formula can be obtained from Gauss's theorem (P3.3.1.3) by taking as a Gaussian surface a small cylinder whose one base is outside the conductor but near its surface and the other is inside. Then we get $ES = \sigma S/\varepsilon_0$.

► **Isolated conductor.** The potential of an isolated conductor is proportional to its charge:

$$\varphi = \frac{1}{C}q, \qquad \text{(P3.4.1.2)}$$

where C is the *capacitance* of the conductor; it depends on the shape and size of the conductor and the dielectric permittivity of the surrounding medium. In SI, the unit of capacitance is the *farad* (F). In CGS, capacitance is measured in centimeters (cm). For example, the potential and capacitance of an isolated ball of radius R equal

$$\varphi = \frac{1}{4\pi\varepsilon_0}\frac{q}{R}, \quad C = 4\pi\varepsilon_0 R. \qquad \text{(P3.4.1.3)}$$

► **Connection of conductors.** When two conductors having different potentials are connected to each other with a conducting wire, their potentials become equal. If the conductors are sufficiently far away from each other, their individual potentials can be calculated as the potential of an isolated conductor. Then the law of conservation of electric charge allows us to find the resulting common potential, φ':

$$C_1\varphi_1 + C_2\varphi_2 = C_1\varphi' + C_2\varphi' \quad\Longrightarrow\quad \varphi' = \frac{C_1\varphi_1 + C_2\varphi_2}{C_1 + C_2}$$

(the capacitance of the wire and its charge may be neglected). If the capacitance of one conductor is much larger than that of the other, its potential remains almost unchanged. For example, if a conductor is connected to the earth (it is said to be earthed), its potential becomes equal to the Earth's potential, which is constant and usually taken to be zero.

► **Conductor in an external electric field.** When any conductor is placed in an external electric field, its movable charges redistribute so that the resulting field (superposition of the external and internal fields) within the conductor is zero. For example, if a positive point charge is brought near an uncharged conductor, then negative charges will gather on the side that is closer to the charge and positive charges will move to the side that is further away. As a result, the charge and conductor will attract. (This phenomenon is called *electrostatic induction*.) Since the charge distribution on the conductor surface is usually unknown, the principle of superposition becomes unsuitable for the calculation of the resulting field. Sometimes, the problem may be solved by indirect techniques.

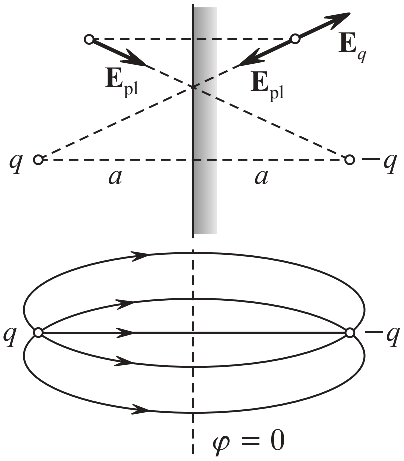

Figure P3.7. Calculation of the field by the electrostatic image method.

Example 1. Let us bring a positive point charge q near the surface of an earthed conductor to a small distance a compared to the conductor size. In this case, the conductor can be replaced by an infinite conducting half-space (Fig. P3.7). There will be induced negative charges at the plane surface. It is required to find the field $\mathbf{E}_{\rm pl}$ produced by these charges at any point of the free space. Note that the charge distribution is unknown!

To begin with, let us find the field $\mathbf{E}_{\rm pl}$ inside the conductor at the point symmetric to an outside point of interest. The field $\mathbf{E}_{\rm pl}$ will compensate the field produced by the charge q at this point. Hence, the field $\mathbf{E}_{\rm pl}$ inside will exactly coincide with that generated by an imaginary charge $-q$ placed at the same point as q. Now we see that the field $\mathbf{E}_{\rm pl}$ produced by the surface charges of the conductor is symmetric relative to the plane. Consequently, $\mathbf{E}_{\rm pl}$ *outside* will exactly coincide with the field produced by the imaginary charge $-q$ placed *symmetrically relative to the charge* q. This symmetric charge is called the *image* of the charge q and the method is known as the *electrostatic image method*. It is now fairly easy to find: (a) the attractive force between the charge and the conductor, $F = kq^2/(2a)^2$, (b) the electric field everywhere outside the conductor (it is equal to the field produced by the charge q and its image $-q$), and (c) the charge distribution of the surface—once the field near the surface is known, we can find the charge density σ using formula (P3.4.1.1).

► **Principle of uniqueness.** The principle of uniqueness is useful in solving problems with unknown charge distribution. This principle states that for a system of conductors, there exists a unique surface charge distribution and a unique field $\mathbf{E}(\mathbf{r})$ such that the net fields inside the conductors vanish and their potentials and charges acquire prescribed values. If one guesses a charge distribution or a field configuration that satisfies the above conditions, it will be the correct solution. For example, one can notice that the equipotential surface $\varphi = 0$ of the field produced by two charges, q and $-q$, coincides with their symmetry plane (see Fig. P3.7). Then it can be inferred that the fields produced by the charge q and it image $-q$ coincides with that produced by the charge q and the induced charge of the conducting plane (see the preceding example).

Example 2. A conducting ball in a uniform field. Consider an uncharged conducting ball of radius R in a unform field $\mathbf{E}_0$. The induced charges on the ball will redistribute so as to generate the field $-\mathbf{E}_0$ inside the ball. In order to obtain this distribution, consider two imaginary balls of the same radius but having a uniform volume charge, one of density ρ and the other of density $-\rho$. (The field of a uniformly charged ball was considered in example 3 from Subsection P3.3.2.) If the center of the positively charged ball is slightly displaced by **l** relative to the center of the negatively charged ball (Fig. P3.8), then the field in the overlapping region is

$$\mathbf{E} = -\frac{\rho\mathbf{r}}{3\varepsilon_0} + \frac{\rho(\mathbf{r}-\mathbf{l})}{3\varepsilon_0} = -\frac{\rho\mathbf{l}}{3\varepsilon_0},$$

and hence is uniform. The charge inside the overlapping region is zero. Now by letting $\mathbf{l} \to \mathbf{0}$ and $\rho \to \infty$ so that $\rho\mathbf{l}/(3\varepsilon_0) = \mathbf{E}_0$, we obtain the desired induced charge distribution: $\sigma = \sigma_0 \cos\theta$, where $\sigma_0 = \rho l = 3\varepsilon_0 E_0$. Also, we find the field the induced charges produce outside the ball. It coincides with that produced by a point dipole with the dipole moment $\mathbf{p} = (\rho V)\mathbf{l} = 4\pi\varepsilon_0 R^3 \mathbf{E}_0$ (see example 2, Subsection P3.2.2).

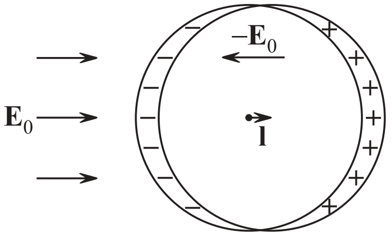

Figure P3.8. To the calculation of induced charges on a conducting ball in a uniform electric field.

P3.4.2. Capacitors

► **Capacitance of a capacitor.** A *capacitor* is an object consisting of a pair of insulated conductors, one of which having a charge $+q$ and the other the charge $-q$. The conductors are often called the *plates* of the capacitor. The charge q is called the *capacitor charge*. The potential difference between the positive and negative plates, U, is called the *capacitor voltage*. The voltage is proportional to the capacitor charge:

$$U = \frac{1}{C}q, \tag{P3.4.2.1}$$

where C is the *electric capacitance* (or just *capacitance*) of the capacitor.

► **Simple capacitors.** A capacitor is called simple if: (a) its electric field is concentrated within a limited region between the plates (so it can be assumed that all field lines start at the positive plate and finish at the negative one), (b) the whole region where the field is concentrated is filled with a dielectric material having a dielectric permittivity ε. (A dielectric weakens the electric field by a factor of ε; see Subsection P3.5.1.)

Example 1. Parallel-plate capacitor. A parallel-plate capacitor is one in which the conductors are parallel plates of area S each separated by a small distance b as compared to the plate size ($b \ll \sqrt{S}$). The electric field between the plates can be regarded as uniform everywhere except for the very edges. The field can be found using formula (P3.4.1.1): $E = q/(\varepsilon_0\varepsilon S)$. The voltage of a parallel-plate capacitor equals $U = Eb$ and its capacitance is

$$C = \frac{\varepsilon_0\varepsilon S}{b}. \tag{P3.4.2.2}$$

Example 2. Spherical capacitor. In a spherical capacitor, its plates are concentric spheres of radii R_1 and R_2. The field between the plates is $E = \dfrac{q}{4\pi\varepsilon_0\varepsilon r^2}$, where r is the distance to the center. The voltage is expressed as $U = \int_{R_1}^{R_2} E(r)\,dr$ and the capacitance equals

$$C = \frac{4\pi\varepsilon_0\varepsilon R_1 R_2}{R_2 - R_1}.$$

For $R_2 \to \infty$, this expression reduces to the capacitance of an isolated sphere (see formula (P3.4.1.3)).

Example 3. Cylindrical capacitor. In a cylindrical capacitor, its plates are concentric cylinders of radii R_1 and R_2 with length $l \gg R_2 - R_1$. The field between the plates is $E = \dfrac{q}{2\pi\varepsilon_0\varepsilon r l}$. The voltage is given by $U = \int_{R_1}^{R_2} E(r)\,dr$ and the capacitance equals is expressed as

$$C = \frac{2\pi\varepsilon_0\varepsilon l}{\ln(R_2/R_1)}.$$

► **Compound capacitors.** By connecting plates of several simple capacitors, one obtains a *compound capacitor*. The capacitance of a compound capacitor is determined by (P3.4.2.1) and is expressed in terms of the capacitances of its constituents.

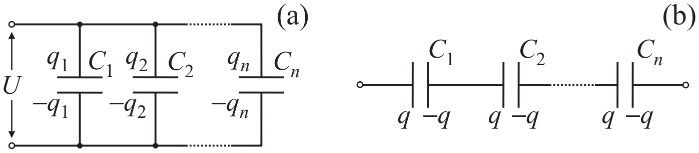

Figure P3.9. Parallel (a) and series (b) connection of capacitors.

1. *Parallel connection*. If n capacitors are connected in parallel with one another as shown in Fig. P3.9(a), they form a compound capacitor for which

$$U = U_i, \quad q = \sum_{i=1}^{n} q_i, \quad C = \sum_{i=1}^{n} C_i.$$

The voltage of all capacitors is the same.

2. *Series connection*. If n capacitors are connected in series as shown in Fig. P3.9(b), they form a compound capacitor for which

$$q = q_i, \quad U = \sum_{i=1}^{n} U_i, \quad \frac{1}{C} = \sum_{i=1}^{n} \frac{1}{C_i}.$$

The total charge at each pair of inner plates is zero.

3. The capacitance of a compound capacitor can often be found by replacing groups of capacitors connected in series or in parallel by an equivalent capacitor, eventually reducing the circuit to a series or parallel one. If the circuit cannot be reduced in this manner, the method of nodal potentials may be applied.

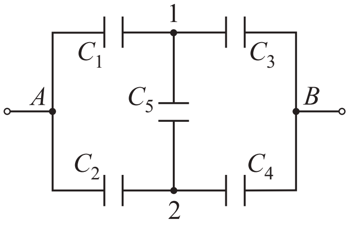

Figure P3.10. A capacitor connection irreducible to a serial or parallel circuit.

Example 4. Consider the circuit shown in Fig. P3.10. Let us take $\varphi_A = 0$. Then $\varphi_B = U$. The three connected plates of the capacitors C_1, C_3, and C_5 must have the same potential, φ_1, and their total charge is zero. The same holds for the capacitors C_2, C_4, and C_5. This allows us to write a system of equations for the potentials φ_1 and φ_2:

$$(\varphi_1 - U)C_3 + (\varphi_1 - \varphi_2)C_5 + \varphi_1 C_1 = 0,$$
$$(\varphi_2 - U)C_4 + (\varphi_2 - \varphi_1)C_5 + \varphi_2 C_2 = 0.$$

Solving the system, we find the charge of the compound capacitor, $q = C_1\varphi_1 + C_2\varphi_2$, and its capacitance $C = q/U$.

P3.5. Electrostatics of Dielectrics

P3.5.1. Polarization of Dielectrics

► **Bound charges. Polarization density.** A dielectric is a material that does not conduct electric current and, hence, has no free charges (or their number is negligibly small). Nevertheless, an electric field becomes weaker inside a dielectric. This indicates that some macroscopic charges arise in the volume and on the surface of a dielectric when placed in an electric field. These charges are due to *polarization* of the dielectric—the phenomenon where charges bound to individual molecules or the crystal lattice are displaced along the field; these charges are called *bound charges*. The average dipole moment of molecules becomes nonzero, which results in a nonzero dipole moment of any volume of the dielectric. The degree of polarization is characterized by the *polarization density vector*, **P**, which is defined as the net dipole moment of the molecules per unit volume:

$$\mathbf{P} = \frac{\sum_i \mathbf{p}_i}{\Delta V} = n\langle \mathbf{p} \rangle,$$

where n is the molecule concentration.

Macroscopic equations of the electric field in a dielectric can be derived using the simplified assumption that all molecules in the small volume ΔV have the same dipole moment $\langle \mathbf{p}\rangle = \mathbf{P}/n = q_0\langle \mathbf{l}\rangle$.

► **Formation of macroscopic bounded charges.** Macroscopic charges distributed over surface and volume can be used to describe the electric field produced by a huge number of oriented microscopic dipoles. The situation where polarization is uniform ($\mathbf{P} = \text{const}$) is easiest to study. In this case, the volume of the dielectric remain electrically neutral and positive surface charges arise on one side of the dielectric and negative surface charges arise on the other side. The vector **P** is directed from the negative to positive charges and the electric field produced will be in *opposite* direction.

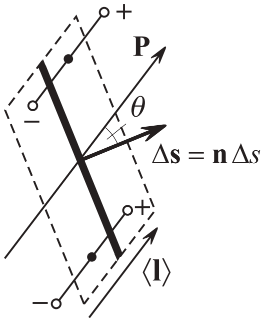

Figure P3.11. Some charges cross a selected area when the dielectric is polarized.

Polarized molecules in the volume of a dielectric are displaced from their original positions so that some charge, Δq, crosses an arbitrary area Δs in the perpendicular direction. This charge equals

$$\Delta q = n(\Delta s\langle l\rangle \cos\theta)q_0 = P_n\Delta s = (\mathbf{P}\cdot\Delta\mathbf{s}) \qquad \text{(P3.5.1.1)}$$

where $\Delta s\langle\, l\,\rangle\cos\theta$ is the volume of a thin cylindrical layer whose molecules cross the area (Fig. P3.11). Integrating over any closed surface yields the net bound charge displaced from the enclosed volume (the net bound charge enclosed within the surface will have the opposite sign):

$$q^{\rm b}_{\rm enc} = -\oint \mathbf{P}\cdot d\mathbf{s} \quad \text{or} \quad \operatorname{div}\mathbf{P} = -\rho^{\rm b}, \tag{P3.5.1.2}$$

where $d\mathbf{s} = \mathbf{n}\,ds$ is the surface element vector directed along the normal toward the surrounding space and $\rho^{\rm b}$ is the volume density of the bound charge. The surface density of the bound charge of (P3.5.1.1) is given by

$$\sigma^{\rm b} = P_n. \tag{P3.5.1.3}$$

► **Electric displacement.** The macroscopic electric field, $\mathbf{E}(\mathbf{r})$, in a polarized dielectric is determined by averaging the microscopic field, $\mathbf{E}_{\rm micro}$ (highly varying at interatomic distances), over a small volume containing sufficiently many molecules: $\mathbf{E}(\mathbf{r}) = \langle \mathbf{E}_{\rm micro}\rangle$. The total electric field is determined by both bound charges and free charges brought to the dielectric from outside:

$$\oint \mathbf{E}\cdot d\mathbf{s} = \frac{1}{\varepsilon_0}(q^{\rm f}_{\rm enc} + q^{\rm b}_{\rm enc}). \tag{P3.5.1.4}$$

Since the bound charge distribution is unknown in advance, it is convenient to introduce a vector quantity

$$\mathbf{D} = \varepsilon_0\mathbf{E} + \mathbf{P}, \tag{P3.5.1.5}$$

which is called the *electric displacement field* or just the *electric displacement*. It follows from equations (P3.5.1.2) and (P3.5.1.4) that the electric displacement is determined by the free charges only:

$$\oint \mathbf{D}\cdot d\mathbf{s} = q^{\rm f}_{\rm enc}. \tag{P3.5.1.6}$$

This equation is known as *Gauss's law for the electric displacement*. In differential form, this law reads

$$\operatorname{div}\mathbf{D} = \rho^{\rm f}.$$

► **Isotropic dielectric.** The polarization density $\mathbf{P}$ at a given point in a dielectric is determined by the external electric field $\mathbf{E}$. For not too strong fields, $\mathbf{P}$ is linearly dependent on $\mathbf{E}$. Furthermore, $\mathbf{P} \parallel \mathbf{E}$ in an *isotropic* dielectric, and hence

$$\mathbf{P} = \varepsilon_0\chi\mathbf{E}, \tag{P3.5.1.7}$$

where χ is the *electric susceptibility* of the dielectric material. Then the electric displacement can be expressed as

$$\mathbf{D} = \varepsilon_0\mathbf{E} + \mathbf{P} = \varepsilon_0\mathbf{E} + \varepsilon_0\chi\mathbf{E} = \varepsilon_0\varepsilon\mathbf{E}, \tag{P3.5.1.8}$$

where $\varepsilon = 1 + \chi$ is the *electric permittivity* of the material.

► **Boundary conditions at dielectric interface.** For two dielectric media meeting along a common interface, the following boundary conditions for tangential and normal components hold:

$$E_{1\tau} = E_{2\tau}, \qquad P_{2n} - P_{1n} = -\sigma^{\rm b}, \qquad D_{2n} - D_{1n} = \sigma^{\rm f}. \tag{P3.5.1.9}$$

The normal $\mathbf{n}$ points from medium 1 to medium 2. If there is no free charge at the interface, the normal components of the displacement are equal to each other. The first condition in (P3.5.1.9) results from the potentiality of the field $\mathbf{E}$ (the work along a closed path is zero). The second and third conditions follow from formulas (P3.5.1.2) and (P3.5.1.6), respectively. For an isotropic dielectric, we have

$$E_{1\tau} = E_{2\tau}, \quad \varepsilon_1 E_{1n} = \varepsilon_2 E_{2n}; \qquad D_{1n} = D_{2n}, \quad \frac{D_{1\tau}}{\varepsilon_1} = \frac{D_{2\tau}}{\varepsilon_2}. \tag{P3.5.1.10}$$

It can be seen that the field lines are refracted at the interface between dielectric media. Note that unlike the electric field, the electric displacement lines are continuous at the interface.

► **Calculation of fields in the presence of dielectrics.** Given the distribution of free charges, the field in the whole space is uniquely determined by the equations for **E** and **D** together with the relation between them (P3.5.1.8) and the boundary conditions (P3.5.1.10). The solution has the following important properties:

1. Using equations (P3.5.1.2) and (P3.5.1.6) and the formula $\mathbf{P} = \frac{\varepsilon - 1}{\varepsilon}\mathbf{D}$, which follows from (P3.5.1.7) and (P3.5.1.8), we find for a homogeneous isotropic dielectric that

$$\rho^{\rm b} = -\frac{\varepsilon - 1}{\varepsilon}\rho^{\rm f}. \qquad \text{(P3.5.1.11)}$$

Consequently, surface bound charge only will arise due to polarization in the absence of volume free charge. Note that for an inhomogeneous dielectric, for which ε depends on the coordinates, volume bound charge arises even if there is no volume free charge.

2. If the boundaries of a homogeneous dielectric coincide with equipotential surfaces of the field $\mathbf{E}_0$ produced by free charges in the absence of the dielectric, then the solution

$$\mathbf{E}(\mathbf{r}) = \frac{\mathbf{E}_0(\mathbf{r})}{\varepsilon} \qquad \text{(P3.5.1.12)}$$

satisfies the boundary conditions.

Example 1. Suppose a point charge q is surrounded by a hollow dielectric sphere with inner radius R_1 and outer radius R_2. The field in the cavity and outside the dielectric will remain the same. Inside the dielectric, the field will change so that

$$E = \frac{1}{4\pi\varepsilon_0\varepsilon}\frac{q}{r^2}, \quad P = \varepsilon_0(\varepsilon - 1)E = \frac{\varepsilon - 1}{\varepsilon}\frac{q}{4\pi r^2},$$
$$\sigma_2^{\rm b} = P(R_2) = \frac{\varepsilon - 1}{\varepsilon}\frac{q}{4\pi R_2^2}, \quad \sigma_1^{\rm b} = -\frac{\varepsilon - 1}{\varepsilon}\frac{q}{4\pi R_1^2}.$$

The bound charges at the inner and outer surfaces are equal to each other in magnitude: $q_2^{\rm b} = -q_1^{\rm b} = q\frac{\varepsilon - 1}{\varepsilon}$.

Example 2. Consider a ball made from a homogeneous dielectric and charged uniformly by a free charge with volume density ρ. The field in the absence of the dielectric was found in example 3, Subsection P3.3.2. The field in the dielectric, for $r < R$, will be weaker by a factor of ε:

$$E = \frac{\rho r}{3\varepsilon_0\varepsilon}, \quad P = \varepsilon_0(\varepsilon - 1)E = \frac{\varepsilon - 1}{\varepsilon}\frac{\rho r}{3},$$
$$\sigma^{\rm b} = P(R) = \frac{\varepsilon - 1}{\varepsilon}\frac{\rho R}{3}, \quad \rho^{\rm b} = -\frac{\varepsilon - 1}{\varepsilon}\rho.$$

It is not difficult to verify that $\sigma^{\rm b}4\pi R^2 = -\rho^{\rm b}\frac{4}{3}\pi R^3$, so the total bound charge is zero, as one would expect.

Example 3. A dielectric slab is placed in a field $\mathbf{E}_0$ so that it enters the slab at the right angle. The electric permittivity of the slab varies from ε_1 at one side to ε_2 at the other side. Then we have

$$D = \varepsilon_0 E_0 = \text{const}, \quad E = \frac{E_0}{\varepsilon(x)}, \quad P = D\frac{\varepsilon - 1}{\varepsilon}, \quad \rho^{\rm b} = -\frac{dP}{dx} = -\frac{D}{\varepsilon^2}\frac{d\varepsilon}{dx},$$
$$\sigma_2^{\rm b} = P_2 = D\frac{\varepsilon_2 - 1}{\varepsilon_2}, \quad \sigma_1^{\rm b} = -P_1 = -D\frac{\varepsilon_1 - 1}{\varepsilon_1}.$$

It is easy to verify that the total bound charge is zero.

Example 4. If a homogeneous dielectric ball is placed in a uniform field $\mathbf{E}_0$, the ball will polarize uniformly ($\mathbf{P} = \text{const}$). Indeed, uniform polarization corresponds to two uniformly charged balls whose centers are displaced by a small vector $\mathbf{l}$, as was shown in example 2, Subsection P3.4.1. The induced polarization field inside the balls is $\mathbf{E}_1 = -\rho\mathbf{l}/3\varepsilon_0 = -(q\mathbf{l}/V)/3\varepsilon_0 = -\mathbf{P}/3\varepsilon_0$. Then the total field in the dielectric is $\mathbf{E} = \mathbf{E}_0 + \mathbf{E}_1 = \mathbf{E}_0 - \mathbf{P}/3\varepsilon_0$. From this equation and the condition $\mathbf{P} = (\varepsilon - 1)\varepsilon_0\mathbf{E}$, we find that

$$\mathbf{E} = \frac{3\mathbf{E}_0}{\varepsilon + 2}, \qquad \mathbf{P} = \frac{3\varepsilon_0(\varepsilon - 1)}{\varepsilon + 2}\mathbf{E}_0.$$

The boundary conditions are all satisfied.

P3.5.2. Types of Dielectric and Polarization Mechanisms

There are three main mechanisms of polarization. Accordingly, three basic types of dielectric are distinguished.

► **1. Dipolar polarization in polar dielectrics.** The molecules of a *polar dielectric* (*polar molecules*) have a permanent dipole moment, p_0, in the absence of electric field. Examples of such substances, are water (H_2O), hydrochloric acid (HCl), ammonia (NH_3), etc. Due to thermal motion, all dipoles are oriented chaotically in the absence of electric field, resulting in a zero average dipole moment. If there is an external field $\mathbf{E}$, dipoles tend to orient along the field (see Subsection P3.2.2). The competition between the chaotic thermal motion and the orienting action of the electric field results in an *average dipole moment* $\langle\mathbf{p}\rangle$. In a weak field ($p_0E \ll kT$), the average dipole moment is proportional to $\mathbf{E}$ and decreases with increasing temperature.

In gaseous dielectrics, the probability for polar molecules to be oriented along $\mathbf{p}$ is proportional to $\exp(-W_\mathrm{p}/kT)$, where $W_\mathrm{p} = -(\mathbf{p}\cdot\mathbf{E})$, which follows from Boltzmann's distribution (see Subsection P2.5.4). It is apparent that, in an external electric field, the average dipole moment must be directed along $\mathbf{E}$ and proportional to $p_0 f(p_0E/kT)$, where $f(x)$ is some function such that $f(0) = 0$ and $f(x) \to 1$ as $x \to \infty$. The case $f(x) \to 1$ corresponds to a strong field, where all dipoles are nearly parallel to $\mathbf{E}$. For small x (a weak field), we have $f(x) \approx \gamma x$ and $\langle\mathbf{p}\rangle \approx \gamma(p_0^2/kT)\,\mathbf{E}$, so the average dipole moment is proportional to $\mathbf{E}$ and inversely proportional to T. The analysis gives $\gamma = 1/3$. Then, $\mathbf{P} = \frac{1}{3}(np_0^2/kT)\,\mathbf{E}$, and hence the electric permittivity evaluates to

$$\varepsilon = 1 + \frac{1}{3}\frac{np_0^2}{\varepsilon_0 kT}.$$

This formula is known as the *Langevin–Debye law*.

► **2. Electronic polarization in nonpolar dielectrics.** In the absence of external electric fields, the dipole moment of a *nonpolar molecule* (e.g., H_2, O_2, N_2, etc.) is zero. So the minimum potential energy corresponds to the state where the center of the negative charge (electronic cloud) of the molecule coincides with that of the positive charge. In the presence of an electric field, the centers of the negative and positive charges are displaced and a force arises that tries to return them to an equilibrium. If the displacement, $\mathbf{l}$, is small, the restoring force is proportional to it: $\mathbf{F} = -\beta\mathbf{l}$ (*quasielastic force*). In equilibrium, $q\mathbf{E} = \beta\mathbf{l}$ and the *induced* dipole moment, $\mathbf{p} = q_0\mathbf{l}$, turns out to be proportional to the field: $\mathbf{p} = \alpha\varepsilon_0\mathbf{E}$. The coefficient α here is called the *molecular polarizability*. The polarization density is $\mathbf{P} = n\mathbf{p} = n\varepsilon_0\alpha\mathbf{E}$, where $\mathbf{E}$ is the field acting on a single molecule. In gases, this field can be regarded as the average field in the dielectric. Consequently, the electric permittivity evaluates to

$$\varepsilon = 1 + \chi = 1 + \alpha n.$$

In dense gases as well as liquid and solid dielectrics, one should take into account that the field acting on a molecule is different from the average field. The molecular field includes the local field produced by the molecule itself, whereas the average field does not.

► **Difference between the local and average fields.** Consider a homogeneous dielectric crystal with nonpolar molecules. In order to find the field acting on a given molecule, consider a sphere with its center at the molecule. If the molecules are arranged in a regular manner (e.g., at the vertices of a cubic lattice), the field produced by the molecules included within the sphere is zero at the center. Hence, the local field at a given point is $\mathbf{E}_\mathrm{loc} = \mathbf{E} + \mathbf{E}_1$, where $\mathbf{E}$ is the average macroscopic field (equal to the sum of the external field and the outer surface charge field) and $\mathbf{E}_1$ is the inner surface charge field of the spherical hollow. As follows from example 4 above, $\mathbf{E}_1 = \mathbf{P}/3\varepsilon_0$, and hence $\mathbf{P} = n\varepsilon_0\alpha\bigl(\mathbf{E} + \frac{1}{3\varepsilon_0}\mathbf{P}\bigr)$, whence

$$\mathbf{P} = \frac{\varepsilon_0 n\alpha\mathbf{E}}{1 - \frac{n\alpha}{3}}.$$

Taking into account that $\mathbf{P} = \varepsilon_0(\varepsilon - 1)\mathbf{E}$, we get

$$\varepsilon - 1 = \frac{n\alpha}{1 - \frac{n\alpha}{3}} \qquad \text{or} \qquad \frac{\varepsilon - 1}{\varepsilon + 2} = \frac{n\alpha}{3}.$$

This result is known as the *Clausius–Mossotti relation*.

► **3. Ionic polarization in solid crystal dielectrics.** Examples of such dielectrics are NaCl, KCl, NaF, KI, etc. In the presence of an external electric field, the sublattice of positive ions (e.g., Na^+) is shifted, as a whole, along the field and the sublattice of negative ions (e.g., Cl^-) is displaced in the opposite direction.

► **Ferroelectrics.** Ferroelectric materials are substances that contain small regions (domains) where all dipoles are spontaneously polarized in the same direction. Ferroelectrics demonstrate this property only below a certain phase transition temperature, T_c, called the *Curie temperature*. These materials are electric analogues of ferromagnetic materials. They have anomalously large electric permittivity, dependent on the electric field and temperature, demonstrate hysteresis properties when polarized, and become usual dielectrics above the Curie temperature. (For details about ferromagnetics, see Subsection P3.9.2.)

P3.6. Energy of Electrostatic Field

P3.6.1. Energy of a System of Charges

► **Potential energy of interaction of charged particles.** The energy of interparticle interaction in a system of charged particles is defined as the work done by external forces to create this system or as the work done by the electric field to destroy this system. The energy of interaction between two charged particles equals the potential energy of one particle in the field produced by the other:

$$W_{12} = q_1\varphi_2(\mathbf{r}_1) = q_2\varphi_1(\mathbf{r}_2) = k\frac{q_1q_2}{r}.$$

The energy of a system of charged particles is

$$W = \sum_{i>j} W_{ij} = \frac{1}{2}\sum_{i\neq j} W_{ij} = \frac{1}{2}\sum_{i} q_i\varphi(\mathbf{r}_i), \qquad \text{(P3.6.1.1)}$$

where $\mathbf{r}_i$ is the position vector of the ith particle and $\varphi(\mathbf{r}_i)$ is the potential created by all other particles at $\mathbf{r}_i$. For continuously distributed charges with volume density $\rho(\mathbf{r})$ and surface density $\sigma(\mathbf{r})$, the above formula becomes

$$W = \frac{1}{2}\int_V \varphi(\mathbf{r})\rho(\mathbf{r})\,dV + \frac{1}{2}\int_S \varphi(\mathbf{r})\rho(\mathbf{r})\,ds. \qquad \text{(P3.6.1.2)}$$

Although formula (P3.6.1.2) is obtained as a formal generalization of formula (P3.6.1.1), there are essential differences between them. In the latter, $\varphi(\mathbf{r}_i)$ is the potential produced by all charges but the ith, whereas in the former, $\varphi(\mathbf{r})$ is the potential of *all* the charges at $\mathbf{r}$ (the contribution of a volume or surface element decreases to zero with the element size). However, the most important difference is this. Formula (P3.6.1.1) considers only the energy of interaction and if the point charges are moved far away from one another, the energy of interaction will be negligibly small but there will still remain the fields of individual charges. Unlike this, formula (P3.6.1.2) calculates the *total energy* of a system of continuously distributed charges and if the system is divided into increasingly smaller parts that are moved away to infinity, the total energy of all these parts will tend to zero.

Example 1. Calculate the energy of a uniformly charged ball of radius R with total charge q.
Solution. Substituting the ball field potential (see example 3, Subsection P3.3.2) into (P3.6.1.2) gives

$$W = \frac{1}{2}\int \rho\varphi\,dV = \frac{\rho}{2}\int_0^R \Bigl(-k\frac{qr^2}{2R^3} + \frac{3}{2}k\frac{q}{R}\Bigr)4\pi r^2 dr = \frac{3}{5}k\frac{q^2}{R}.$$

▶ **Energy of an isolated conductor.** All charges are on the surface and have the same potential, which is the potential of the conductor. From (P3.6.1.2) we have

$$W = \frac{q\varphi}{2} = \frac{C\varphi^2}{2} = \frac{q^2}{2C}, \qquad \text{(P3.6.1.3)}$$

where C is the electric capacitance of the conductor (see Subsection P3.4.1). This result can also be obtained by directly calculating the work done by the field to take the whole charge to infinity: $W = \int \varphi(q)\,dq = C\int_0^\varphi \varphi\,d\varphi = \frac{1}{2}C\varphi^2$.

Example 2. The energy of an isolated conducting sphere equals

$$W = \frac{1}{2}q\cdot\frac{q}{4\pi\varepsilon_0 R} = \frac{q^2}{8\pi\varepsilon_0\varepsilon R},$$

where ε is the electric permittivity of the surrounding medium.

▶ **Capacitor energy.** From (P3.6.1.2) we get

$$W = \frac{q\varphi_1}{2} - \frac{q\varphi_2}{2} = \frac{qU}{2} = \frac{CU^2}{2} = \frac{q^2}{2C}. \qquad \text{(P3.6.1.4)}$$

The field in the capacitor can be destroyed by transferring the charge in small portions from one plate to the other. So $W = \int U(q)\,dq = C\int_0^U U\,dU = \frac{1}{2}CU^2$. For example, the energy of a parallel-plate capacitor is

$$W = \frac{\varepsilon_0\varepsilon S}{2b}(Eb)^2 = \frac{\varepsilon_0\varepsilon E^2}{2}Sb. \qquad \text{(P3.6.1.5)}$$

P3.6.2. Energy of an Electrostatic Field

▶ **Energy density of an electric field.** Within a field approach, one should deal with the energy of the field surrounding charged particles rather than the energy of interparticle interaction. The volume energy density of an electric field, $w = dW/dV$, in vacuum is only dependent on the field strength:

$$w = \tfrac{1}{2}\varepsilon_0 E^2. \qquad \text{(P3.6.2.1)}$$

This expression holds for any charge distribution (e.g., it can be obtained by considering the uniform field of a parallel-plate capacitor).

With the field equations (see Subsection P3.3.2), formula (P3.6.1.2) can be rewritten as

$$W = \frac{1}{2}\int \varepsilon_0 E^2\,dV, \qquad \text{(P3.6.2.2)}$$

where the integration is performed over the whole space.

Formulas (P3.6.1.2) and (P3.6.2.2) allow one to calculate the total energy in the whole space only if: (a) there are no point or linear charges and (b) the charges are limited to a finite region of space. However, formula (P3.6.2.2) calculates not only field energy in the whole space but also energy limited to a finite region of space even if the total energy is infinite.

Example 1. When the charged ball energy was calculated (see example 1) using (P3.6.1.2), the integration was performed of the ball volume (since $\rho = 0$ outside the ball). However, formula (P3.6.2.2) reveals that not all of the field energy is stored within the ball:

$$W_1 = \int \frac{1}{2}\varepsilon_0 E^2\,dV = \int_0^R \frac{1}{2}\varepsilon_0\Big(\frac{qr}{4\pi\varepsilon_0 R^3}\Big)^2(4\pi r^2 dr) = \frac{1}{10}k\frac{q^2}{R}.$$

It makes only 1/6 of the total energy while the 5/6 is stored in the surrounding void space:

$$W_2 = \int_R^\infty \frac{1}{2}\varepsilon_0\Big(\frac{q}{4\pi\varepsilon_0 r^2}\Big)^2(4\pi r^2 dr) = \frac{1}{2}k\frac{q^2}{R}.$$

In a dielectric, the energy density is larger by a factor of ε:

$$w = \tfrac{1}{2}\varepsilon_0\varepsilon E^2 = \tfrac{1}{2}ED = \tfrac{1}{2}\varepsilon_0 E^2 + \tfrac{1}{2}\varepsilon_0(\varepsilon - 1)E^2.$$

The first term in the last expression is the field energy in the dielectric itself while the second term, equal to $\frac{1}{2}\mathbf{E}\cdot\mathbf{P}$, is the work done by the field to polarize the dielectric. This expression can be generalized to the case where $\mathbf{P}$ (and $\mathbf{D}$) are not proportional to $\mathbf{E}$ (e.g., for ferroelectrics and strong fields). To this end, one should look at the process of charging a parallel-plate capacitor filled with such a dielectric:

$$w = \frac{W}{V} = \frac{1}{V}\int U\,dq = \frac{1}{V}\int (Eb)\,d(DS) = \int_0^D \mathbf{E}\cdot d\mathbf{D}.$$

► **Law of conservation of energy. Force calculation.** To calculate the forces, acting on a charged or polarized body in an electric field, one should look at the process of slow isothermal change of the body position. The work done by external forces together with work done by voltage sources are equal to the change in the field energy. Since the body is in mechanical equilibrium, the work done by external forces equals the work done by the force exerted by the field, taken with the opposite sign.

Example 2. A parallel-plate capacitor charged to a voltage U is filled with a liquid or gaseous dielectric with permittivity ε. Find the force acting on either plate. Also treat the case of a solid dielectric.

Solution. Assume that the capacitor is disconnected from the voltage source. When the distance between the plates is increased by dx, the energy increases by $\frac{1}{2}EDS\,dx$ and the work done is $F\,dx$. Hence, the force of attraction equals $F = \frac{1}{2}EDS = \frac{1}{2}Eq$, where $\frac{1}{2}E$ is the field strength in the dielectric produced by either plate ($E = U/b$). However, if the dielectric is a solid slab, then there is an air gap between the dielectric and either plate. Then the energy increment equals $\frac{1}{2}\varepsilon_0 E_0^2 S\,dx$ and the force is

$$F = \frac{1}{2}\varepsilon_0 E_0^2 S = \frac{1}{2}E_0 DS = \frac{1}{2}qE_0,$$

where $\frac{1}{2}E_0$ is the field strength produced by one plate within the air gap ($E_0 = \varepsilon E = \varepsilon U/b$).

Example 3. Find the force at which a dielectric slab is pulled into the gap between the plates of a parallel-plate capacitor charged to a voltage U (see Fig. P3.12).

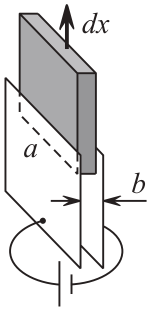

Figure P3.12. To example 3.

Solution. Assume that the respective dimensions of the slab coincide with those of the plates and the slab is slightly thinner than the interplate distance, b. We also assume that the capacitor is connected to a constant voltage source, U. Let us pull the slab out of the capacitor by a distance dx (Fig. P3.12). The field energy within the volume $dV = ab\,dx$ will change by $dW = \left(\frac{1}{2}\varepsilon_0 E^2 - \frac{1}{2}\varepsilon_0\varepsilon E^2\right) ab\,dx$ (a is the slab width and $E = U/b$ is the field). The charge at the plates will change by $dq = \Delta\sigma a\,dx = (\varepsilon_0 E - \varepsilon_0\varepsilon E)a\,dx$ and the work done by the source will be $\delta A_{\text{source}} = U\,dq$, which is twice the field energy increment. Writing the law of conservation of energy, $F\,dx + \delta A_{\text{source}} = dW$, we find that

$$F = \frac{1}{2}\varepsilon_0(\varepsilon - 1)abE^2.$$

Note that this force arises due to field distortion near the plate edges, although the energy calculation was performed without taking into account edge effects.

P3.7. Direct Current

P3.7.1. Electric Current in a Circuit

► **Electric current. Current density.** *Electric current* is an ordered motion of electric charges. *Convection current* is a motion of charges associated with the motion of a charged body. *Conduction current* is a drift motion of *free charges* (*charge carriers*) through a substance (conductor) due to an electric field applied. Examples of free charges are electrons in metals, ions in electrolytes, electrons and ions in plasma, and electrons and "holes" in semiconductors. The positive direction of electric current is conventionally taken to be the direction of positive moving charges.

The charge passing in time dt through an area $d\mathbf{s} = \mathbf{n}\,ds$ inside a conductor in the normal direction $\mathbf{n}$ is determined by the average velocity of free charges, $\langle\mathbf{v}\rangle$:

$$dq = q_0 n\langle\mathbf{v}\rangle\cdot d\mathbf{s}\,dt = \mathbf{j}\cdot\mathbf{s}\,dt, \qquad \text{(P3.7.1.1)}$$

where q_0 is the charge of a single charge carrier, n is the carrier concentration, and $\mathbf{j} = q_0 n\langle\mathbf{v}\rangle$ is the *current density*. The charge passing through the cross-section, S, of a conductor per unit time is called the electric current (or the current) and denoted I:

$$dq = I\,dt = dt\int \mathbf{j}\cdot d\mathbf{s}, \qquad \text{(P3.7.1.2)}$$

where the integration is performed over the cross-section. For a *linear* conductor ($S = \text{const}$), the current density is constant, $\mathbf{j} = \text{const}$, and formula (P3.7.1.2) gives $I = jS$. A current is said to be *constant* if $I = \text{const}$.

The decrease in the total charge within a closed surface equals $dq_{\text{enc}} = -dt \oint \mathbf{j}\cdot d\mathbf{s}$. In differential form, the law of conservation of charge is expressed as $\frac{d\rho}{dt} = -\operatorname{div}\mathbf{j}$. For a constant current, $\oint \mathbf{j}\cdot d\mathbf{s} = 0$, or $\operatorname{div}\mathbf{j} = 0$, so the current flows in continuous lines directed along the conductor.

► **Ohm's law in differential form.** To maintain current through a substance, free charges must be acted upon by an electromagnetic force $\mathbf{F}$, characterized by an electric field $\mathbf{E} = \mathbf{F}/q_0$. As soon as the force is removed, the average velocity of free charges vanishes in a very short time. For not too strong fields, the current density in an isotropic conductor must be proportional to $\mathbf{E}$:

$$\mathbf{j} = \sigma\mathbf{E} = \frac{1}{\rho}\mathbf{E}, \qquad \text{(P3.7.1.3)}$$

where σ is the *conductivity* of the material and ρ is the *resistivity*. It follows from the law of conservation of charge that if $\rho = \text{const}$, then $\oint \mathbf{E}\cdot d\mathbf{s} = \rho\oint \mathbf{j}\cdot d\mathbf{s} = 0$, or $\operatorname{div}\mathbf{E} = 0$. Consequently, the field lines in a *homogeneous conductor* are continuous and in a wire of constant cross-section, they are parallel to the wire surface.

► **Joule–Lenz law in differential form.** As current flows through a conductor, the work done by the field **E** on free charges converts completely into thermal energy due to collisions with ions. The volume density of the thermal power of the current equals the work done by the field on charges per unit time in a unit volume:

$$w_{\text{therm}} = n\langle\mathbf{v}\rangle\cdot\mathbf{F} = q_0 n\langle\mathbf{v}\rangle\cdot\mathbf{E} = \mathbf{j}\cdot\mathbf{E} = \sigma E^2 = \rho j^2. \tag{P3.7.1.4}$$

► **Simple circuit segment. Resistance.** If current flows through a segment of an electrical circuit is due to an electrostatic (Coulomb) field only ($\rho\mathbf{j} = \mathbf{E}_{\text{coulomb}}$), this circuit segment is called *simple*. In a homogeneous conductor ($\rho = \text{const}$), $\oint\mathbf{E}\cdot d\mathbf{s} = \rho\oint\mathbf{j}\cdot d\mathbf{s} = 0$, and, according to Gauss's theorem (see Section P3.3), the field producing charge is only on the surface. In a wire of constant cross-section, the equipotential surfaces coincide with wire cross-sections. The current from a section 1 to section 2 is proportional to the potential difference $\varphi_1 - \varphi_2$:

$$\int_1^2 \mathbf{E}_{\text{coulomb}}\cdot d\mathbf{l} = \int_1^2 \rho\mathbf{j}\cdot d\mathbf{l} = I\int_1^2 \rho\frac{dl}{S} \quad\Longrightarrow\quad \varphi_1 - \varphi_2 = IR. \tag{P3.7.1.5}$$

This result is known as *Ohm's law for a simple circuit segment*. If $I < 0$, then current flows from section 2 to 1. The quantity $R = \int_1^2(\rho/S)\,dl$ is the *electrical resistance of the conductor*. The SI unit of electrical resistance is the ohm (denoted Ω).

For a homogeneous linear conductor, $R = \rho l/S$, where ρ is the electrical resistivity, l is the conductor length, and S is the cross-sectional area. The electrical resistivity is temperature dependent; this dependence is practically linear in a wide range of temperatures: $\rho = \rho_0[1 + \alpha(T - T_0)]$, where ρ_0 is the resistivity at the reference temperature T_0 (e.g., 0° C), α is the *temperature coefficient of resistivity*. For many metals, $\alpha \approx 1/273$ (so $\rho \approx \rho_0\alpha T$); for semiconductors, $\alpha < 0$.

► **Electromotive force.** Although electrostatic forces can maintain current through an open circuit segment, the total work done by these forces around a closed circuit is zero. Consequently, other, nonelectrostatic forces must be present in a closed circuit to compensate for thermal losses; the total work of these forces around a closed circuit is nonzero, and hence these are nonconservative. However, these forces are of electromagnetic nature, $\mathbf{F}_{\text{noncons}} = q_0\mathbf{E}_{\text{noncons}}$, and their work done to transfer a test charge q from point 1 to point 2 is proportional to q:

$$A_q(1\to 2) = q\int_1^2 \mathbf{E}_{\text{noncons}}\cdot d\mathbf{l} = q\mathscr{E}_{12} = \pm q\mathscr{E}, \tag{P3.7.1.6}$$

where $\mathscr{E}$ is the so-called *electromotive force* (or *emf* for short), which has the dimension of voltage and is measured in volts (V). A source of electromotive force is usually regarded as a source of current (such as a battery) and denoted in circuit diagrams by a circle with the minus and plus signs as shown in Fig. P3.13. The plus sign in formula (P3.7.1.6) is taken when the nonconservative forces act in the direction from point 1 to point 2, and the minus sign is taken otherwise. The power of the nonconservative forces (the power of the emf source) equals

$$P_{\text{noncons}} = \frac{A_q(1\to 2)}{t} = I\mathscr{E}_{12}. \tag{P3.7.1.7}$$

The work done by the nonconservative forces in time dt on all charges in the conductor equals the work done to transfer the charge $I\,dt$ from one end of the conductor to the other.

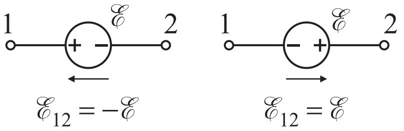
1
$\mathscr{E}$
2
+ −
$\mathscr{E}_{12} = -\mathscr{E}$
1
$\mathscr{E}$
2
− +
$\mathscr{E}_{12} = \mathscr{E}$

Figure P3.13. Sign rule for an emf (voltage) source.

► **Ohm's law for a circuit with a current source.** In a circuit containing a source of current, there are both electrostatic and nonelectrostatic forces acting on charges. In this case, Ohm's law in differential form (P3.7.1.3) reads

$$\mathbf{j} = \frac{1}{\rho}(\mathbf{E}_{\text{coulomb}} + \mathbf{E}_{\text{noncons}}), \tag{P3.7.1.8}$$

Ohm's law in integral form becomes

$$I\int_1^2 \rho\frac{dl}{S} = \int_1^2 (\mathbf{E}_{\text{coulomb}} + \mathbf{E}_{\text{noncons}})\cdot d\mathbf{l} \quad\Longrightarrow\quad IR = (\varphi_1 - \varphi_2) + \mathscr{E}_{12}. \tag{P3.7.1.9}$$

If $I < 0$, current flows from 2 to 1. The quantity $U_{12} = (\varphi_1 - \varphi_2) + \mathscr{E}_{12}$, equal to the work done by all forces to transfer a unit charge from cross-section 1 to 2, is called the *voltage* on the selected segment of the circuit. For a simple segment, the voltage equals the potential difference. Thus, Ohm's law states that $IR = U_{12}$ on any circuit segment.

► **Joule–Lenz law for a circuit segment.** The thermal power of electric current on a circuit segment can be found using formula (P3.7.1.4):

$$P_{\text{therm}} = \int_1^2 wS\,dl = \int_1^2 \rho j^2 S\,dl = I^2\int_1^2 \rho\frac{dl}{S} = I^2R. \tag{P3.7.1.10}$$

P3.7.2. Current in a Closed Circuit

► **Ohm's law for an unbranched circuit.** The current through an unbranched closed circuit is constant, $I = \text{const}$. Adding up equations (P3.7.1.9) for all circuit segments, we get

$$I\sum R_i = \sum \mathscr{E}_{ij}. \tag{P3.7.2.1}$$

For example, if the current source has internal resistance r and the resistance of the remaining circuit is R, we have $I = \mathscr{E}/(R + r)$. The potential difference at the source terminals is $\varphi_+ - \varphi_- = IR = \mathscr{E} - Ir$. The short circuit current of the source ($R = 0$) equals $I_{\text{sc}} = \mathscr{E}/r$. Ohm's law (P3.7.2.1) expresses the law of conservation of energy for an unbranched closed circuit: the power of nonconservative forces equals the power of thermal losses at the resistors in the circuit.

► **Useful power of a current source.** If a current source is used to transmit energy into an external circuit, then the total (spent) power equals the power on nonconservative forces: $P_{\text{total}} = \mathscr{E}I$. The lost power equals the thermal losses at the internal resistance: $P_{\text{lost}} = I^2r$. Then the useful power is $P_{\text{useful}} = \mathscr{E}I - I^2r = I\Delta\varphi$, where $\Delta\varphi$ is the potential difference at the source terminals. It can be seen that the maximum of P_{useful} is attained at $I = \mathscr{E}/(2r)$ and is equal to $P_{\max} = \mathscr{E}^2/(4r)$.

If the potential difference supplied at the ends of a circuit segment is $\varphi_1 - \varphi_2$ and the segment contains a device that performs work against the external forces (e.g., an electric motor), then the total power is $P_{\text{total}} = (\varphi_1 - \varphi_2)I$ and the lost power is $P_{\text{lost}} = I^2R$, where R is the resistance of the motor winding. Hence, the useful power equals $P_{\text{useful}} = (\varphi_1 - \varphi_2)I - I^2R = -\mathscr{E}_{12}I = \mathscr{E}I$. The quantities $\mathscr{E}$ and I depend on the revolution speed of the rotor and the maximum of P_{useful} is attained at $I = (\varphi_1 - \varphi_2)/(2R)$.

► **Analysis of branched circuits. Kirchhoff's circuit laws.** In order to determine the currents at various segments of a branched circuit, one should assign symbols to the unknown currents and specify the current directions (in a voluntary way). One of the following two methods should be used thereafter:

1. *Node potential method*. For a circuit involving N nodes, the nodal potentials are taken to be the unknowns (one of the potentials is set to zero). The currents are expressed in terms of the potentials using (P3.7.1.9). The law of conservation of charge is written for each of the $N-1$ unknown nodal potentials in the form $\sum_i I_i = 0$ (the algebraic sum of the currents flowing into and out of a node is zero; the summation is performed over all currents involved). Solving the resulting linear system, one finds the unknown potentials φ_i and then calculates the currents.

2. *Kirchhoff's laws*. In this method, the currents are taken to be the unknowns. First, the equation $\sum_i I_i = 0$ (known as *Kirchhoff's current law* or *Kirchhoff's first rule*) is written for $N-1$ nodes. Then, one chooses an arbitrary closed path and traces it (either clockwise or counterclockwise) summing up the equations of the form (P3.7.1.9) for each segment. This results in an equation of the form $\sum_i I_i R_i = \sum_{i,j} \mathscr{E}_{ij}$ (known as *Kirchhoff's voltage law* or *Kirchhoff's second rule*). This procedure is repeated for all independent closed paths to obtain the required number of equations. Finally, the resulting linear system of equations is solved for the unknown currents.

For more details and examples, see Chapter E6.

► **Equivalent resistances and sources.** Several resistances R_i connected with one another can be replaced by a single equivalent resistance R.

1) For resistances connected in series, $R = \sum_i R_i$.

2) For resistances connected in parallel, $R^{-1} = \sum_i R_i^{-1}$.

3) In the general case, the equivalent resistance is calculated by analyzing branched circuits.

Several connected current sources with emf $\mathscr{E}_i$ and internal resistance r_i can be replaced by a single equivalent sources with parameters $\mathscr{E}$ and r.

1) For sources connected in series, $\mathscr{E} = \sum_i \mathscr{E}_i$ and $r = \sum_i r_i$.

2) For sources connected in parallel, $\mathscr{E} = \sum_i \mathscr{E}_i r_i^{-1} \big/ \sum_i r_i^{-1}$ and $r^{-1} = \sum_i r_i^{-1}$.

3) In the general case, the equivalent emf and resistance are calculated by analyzing branched circuits.

P3.7.3. Classical Free-Electron Theory of Metals

► **Basic postulates.** Conduction of metals is explained by the behavior of valence electrons. The free valence electrons are assumed to form an electron gas. In the classical free-electron theory of Drude–Lorentz, the electron gas is considered to be a classical gas whose concentration equals the atomic concentration of the metal and temperature equals the metal temperature. Conduction current is calculated by assuming that the electrons acquire the drift velocity as they freely move without collisions and that they lose this velocity completely at each collision with metal ions. In addition, it is assumed that $|\langle \mathbf{v}\rangle| \ll \langle v\rangle \sim \sqrt{kT}$. The average drift velocity is expressed as $\langle \mathbf{v}\rangle \sim a\tau/2 = e\mathbf{E}\tau/(2m_e)$, where the mean time between collisions τ is expressed in terms of the mean free path and the mean thermal speed: $\tau = \lambda/\langle v\rangle$. Using (P3.7.1.1), we get

$$\mathbf{j} = en\langle \mathbf{v}\rangle = \frac{e^2 n\lambda}{2m_e\langle v\rangle}\mathbf{E} \quad \text{or} \quad \rho = \frac{2m_e\langle v\rangle}{e^2 n\lambda}.$$

A thorough analysis leads to the same result except that the 2 is absent from the numerator; however, this is not essential for estimates.

► **Discussion of the result of the classical free-electron theory.**

1. The theory explains the increase of ρ with temperature but predicts the behavior $\rho \sim \sqrt{T}$, which is different from the observed relationship $\rho \sim T$.

2. The theory successfully explains the *Wiedemann–Franz law*: for all metals, the product of the thermal conductivity κ (see Subsection P2.9.2) by the electric resistivity ρ is proportional to temperature T (the product $\kappa\rho$ contains $\langle v\rangle^2$ only).

3. The mean free path of electrons predicted by the classical theory is several orders of magnitude larger than the interatomic distance, which contradicts the basic postulates.

4. The electron contribution to the molar heat capacity (see Subsection P2.5.2) is predicted to be $\frac{3}{2}R$. However, the experimentally found contribution is negligibly small.

P3.8. Constant Magnetic Field

P3.8.1. Lorentz Force and Ampère Force

► **Magnetic field. Lorentz force.** Magnetic field is one of the two components of the electromagnetic field (see Subsection P3.2.1). It acts on moving charged particles, currents, and magnetic moments. The sources of magnetic field are moving charged particles, currents, magnetic moments, and changing electric field. A magnetic field is characterized by a vector quantity **B** called the *magnetic induction* (also called the *magnetic field* or *magnetic flux density*). The vector **B** is determined by the equation expressing the force acting on a moving charged particle in an electromagnetic field:

$$\mathbf{F} = q\mathbf{E} + q\mathbf{v}\times\mathbf{B}. \qquad \text{(P3.8.1.1)}$$

The force $\mathbf{F}_\mathrm{L} = q\mathbf{v}\times\mathbf{B}$ exerted by the magnetic field on a moving charged particle is called the *Lorentz force*. (The electromagnetic force **F** is sometimes also called the Lorentz force or *generalized Lorentz force*.) The Lorentz force perpendicular to the particle velocity and, hence, does not produce work.

In order to determine the magnetic induction **B** by formula (P3.8.1.1), one should do the following:

1) measure the force acting on a static particle to isolate the action of the electric field;

2) find the direction of the velocity **v** for which the magnitude of the magnetic force is maximum at constant v;

3) calculate the magnitude of the magnetic induction as $B = F_\mathrm{max}/(qv)$;

4) for known $\mathbf{F}_\mathrm{max}$ and **v**, find the direction of **B** by the right-hand rule.

The magnetic induction can also be determined by the moment of force the magnetic field acts on a small loop of electric current. The electric field does not affect the electric current loop.

The SI unit of magnetic induction is the *tesla* (T): $1\,\mathrm{T} = 1\,\mathrm{N/(A\,m)} = 1\,\mathrm{kg/(A\,s^2)}$.

Example 1. The Hall effect. A conductor with a current flowing through it is immersed in a transverse constant magnetic field (perpendicular to the current). An electrostatic field perpendicular to the magnetic induction arises within the conductor as shown in Fig. P3.14 (the Hall effect). This field results from displacement of the drifting free charges by the action of the Lorentz force. The displacement continues until the Lorentz force is equalized by the electric force: $q_0\langle\mathbf{v}\rangle\times\mathbf{B} = -q_0\mathbf{E}$. Multiplying by the concentration of free charges, n, we express **E** in terms of the current density to obtain $\mathbf{E} = -\frac{1}{q_0 n}\mathbf{j}\times\mathbf{B}$. It can be seen that the free charge concentration (for charges of the same type) and the sign of the charges can be obtained from the measurement of the magnitude and direction of the electric field **E**.

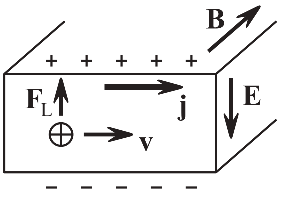

Figure P3.14. A conductor with a current immersed in a magnetic field (the Hall effect).

Example 2. Particle in a uniform magnetic field. Consider a nonrelativistic particle of mass m and charge q moving with a velocity **v** in a uniform magnetic field **B**. Suppose that **v** makes an angle α with **B** at some instant. The Lorentz force is perpendicular to both **v** (hence, v = const) and **B** (hence, the velocity projections $v_{\|}$ and $v_{\perp}$ onto **B** and the plane perpendicular to **B** are conserved). In the projection onto the perpendicular plane, the particle moves in a circle, whose radius can be found from Newton's second law: $qv_{\perp}B = mv_{\perp}^2/R$. The period of revolutions, $T = 2\pi R/v_{\perp} = 2\pi m/qB$, is independent of the speed. So the resulting motion occurs in a helical path with radius R and pitch $h = v_{\|}T = 2\pi mv\cos\alpha/(qB)$.

► **Ampère's law.** The force acting on a current element in a magnetic field (*Ampère's force*) equals the sum of the forces acting on the moving free charges:

$$d\mathbf{F} = q_0[\langle\mathbf{v}\rangle\times\mathbf{B}]\,dN = nq_0[\langle\mathbf{v}\rangle\times\mathbf{B}]\,dV = \mathbf{j}\times\mathbf{B}\,dV = I\,d\mathbf{l}\times\mathbf{B}. \qquad \text{(P3.8.1.2)}$$

This result is known as *Ampère's force law* (in differential form). Note the use of formula (P3.7.1.1), linking the current density to the mean velocity of free charges, to relate the force acting on a volume element dV with current density **j** and a line element $d\mathbf{l}$ with current I. The total force acting of a finite volume with distributed current $\mathbf{j}(\mathbf{r})$ of a linear current-carrying wire is obtained by integrating. For example, the force acting on a wire of length **l** with current I in a uniform magnetic field **B** equals $I\mathbf{l}\times\mathbf{B}$. The force acting on a current-carrying loop is zero: $\mathbf{F} = \oint I\,d\mathbf{l}\times\mathbf{B} = I\bigl(\oint d\mathbf{l}\bigr)\times\mathbf{B} = \mathbf{0}$.

► **Magnetic moment of a current-carrying loop.** The magnetic moment of a wire loop carrying current I is defined as the vector quantity

$$\mathbf{p}_{\rm m} = I\int\mathbf{n}\,ds, \qquad \text{(P3.8.1.3)}$$

where the integration is carried out over any surface spanning the loop and the direction of the normal is determined by the right-hand rule. For a plane loop, we have

$$\mathbf{p}_{\rm m} = IS\mathbf{n} = I\mathbf{S}.$$

The magnetic moment of a loop is a magnetic analogue of the electric dipole moment (see Subsections P3.1.1 and P3.2.2); it determines the magnetic field of the loop at large distances from it and characterizes the behavior of the loop in a magnetic field.

Example 3. Consider a rectangular loop with sides a and b, suspended at the side a is a vertical uniform magnetic field **B** (Fig. P3.15). When current I is supplied, the loop will deflect by an angle β such that the moment of the force of gravity is equalized by the moment of the Ampère force: $mg\frac{1}{2}b\sin\beta = IBab\cos\beta$, or $\tan\beta = 2IBa/(mg)$. Note that the torque exerted by the magnetic field on the loop equals $\mathbf{M} = I\mathbf{S}\times\mathbf{B} = \mathbf{p}_{\rm m}\times\mathbf{B}$. A similar expression was obtained for the torque acting on an electric dipole in an electric field (see Subsection P3.2.2).

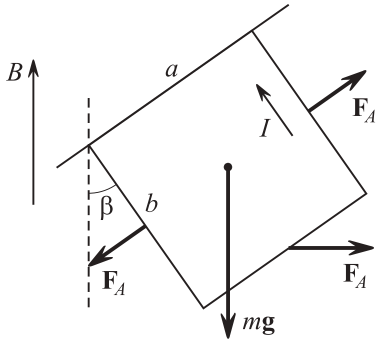

Figure P3.15. A rectangular current-carrying loop in a magnetic electric (to example 3).

P3.8.2. Calculation of Magnetic Induction

► **Biot–Savart–Laplace law.** The magnetic induction created by a current element $I\,d\mathbf{l}$ at a point A equals

$$d\mathbf{B} = \frac{\mu_0}{4\pi}\frac{I\,d\mathbf{l}\times\mathbf{r}}{r^3}, \qquad \text{(P3.8.2.1)}$$

where $\mathbf{r}$ is the radius vector from the current element to the point A and $\mu_0 = 1/(\varepsilon_0 c^2) = 4\pi\times10^{-7}$ T m/A (or H/m) is the magnetic constant. This result is known as the *Biot–Savart law* or the *Biot–Savart–Laplace law*. Equation (P3.8.2.1) can be rewritten for the magnetic induction of a volume current element $\mathbf{j}\,dV$, a surface current element $\mathbf{i}\,ds$, or, on dividing by the number of free current carriers, a nonrelativistic particle $q\mathbf{v}$:

$$d\mathbf{B} = \frac{\mu_0}{4\pi}\frac{\mathbf{j}\times\mathbf{r}}{r^3}dV, \quad d\mathbf{B} = \frac{\mu_0}{4\pi}\frac{\mathbf{i}\times\mathbf{r}}{r^3}ds, \quad d\mathbf{B} = \frac{\mu_0}{4\pi}\frac{q\mathbf{v}\times\mathbf{r}}{r^3}. \qquad \text{(P3.8.2.2)}$$

(The surface density of electric current is the current per unit length of the segment perpendicular to the current: $dI = i\,dl$.) In order to calculate the magnetic induction produced by a finite volume, surface, or wire, one has to perform integration (principle of superposition for magnetic induction).

► **On systems of units in electromagnetism.** Here and henceforth the formulas are given in SI. To convert units to the Gaussian system (CGS), it suffices in most cases to substitute $\mathbf{B}$ by $\mathbf{B}/c$ (and the flux Φ by Φ/c) and μ_0 by $4\pi/c^2$. If a formula involves the Coulomb constant k, it must be substituted by 1. The electric constant ε_0 is substituted by $1/(4\pi)$. It is noteworthy that $\mu_0\varepsilon_0 = 1/c^2$.

Example 1. Find the magnetic force acting between two particle of charges q_1 and q_2 having identical velocities, $\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{v}$ ($v \ll c$), at some instant of time with $\mathbf{v}$ being perpendicular to the line connecting the charges.

Solution. The magnitude of the magnetic induction created by the first particle at the point where the second particle is located is $B = \frac{\mu_0}{4\pi}\frac{q_1 v}{r^2}$, where r is the interparticle distance. The field is perpendicular to $\mathbf{v}$ and $\mathbf{r}$. For charges of like sign, the Lorentz force acts in the direction of particle 1 and its magnitude equals $F = \frac{\mu_0}{4\pi}\frac{q_1 q_2 v^2}{r^2}$. The ratio of the magnetic force to the electrostatic force is $\mu_0\varepsilon_0 v^2 = \frac{v^2}{c^2}$.

Example 2. Magnetic field in the center of a circular current. Find the magnetic induction created by a circular loop of radius R at its center. The current flowing through the loop is I.

Solution. The magnetic induction created by an arc element $d\varphi$ equals

$$dB = \frac{\mu_0}{4\pi}\frac{IR\,d\varphi}{R^2} = \frac{\mu_0 I\,d\varphi}{4\pi R}$$

in magnitude and is directed along the axis (the right-hand rule). At its center, the whole loop creates the field

$$B = \frac{\mu_0 I}{2R}.$$

Example 3. Magnetic field at the axis of a circular current. Find the magnetic induction produced by a circular current (of radius R) on its axis at the point A located at a distance y from the loop center (Fig. P3.16).

Solution. The vector $\mathbf{B}$ is directed along the y-axis and the contribution of an arc element $d\varphi$ of the loop equals

$$dB_y = dB\sin\alpha = \frac{\mu_0}{4\pi}\frac{IR\,d\varphi}{R^2+y^2}\frac{R}{\sqrt{R^2+y^2}}.$$

Integrating with respect to φ yields

$$B = \frac{\mu_0 I R^2}{2(R^2+y^2)^{3/2}}.$$

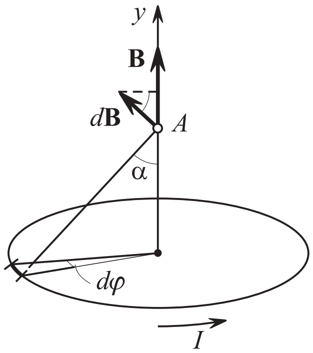

Figure P3.16. Magnetic field at the axis of a circular current.

At large distances from the loop ($y \gg R$), the magnetic induction

$$B \approx \frac{\mu_0 I R^2}{2y^3} = \frac{\mu_0}{4\pi}\frac{2I(\pi R^2)}{y^3}$$

is expressed in terms of the magnetic moment $\mathbf{p}_\mathrm{m} = I\mathbf{S}$ in a similar manner as the electric field strength is expressed via the electric dipole moment (see example 4, Subsection P3.2.2, the case $\mathbf{r} \parallel \mathbf{p}$). I can be shown that at an arbitrary point $\mathbf{r}$, the magnetic induction of a small plane current-carrying loop equals

$$\mathbf{B} = \frac{\mu_0}{4\pi r^3}\left(\frac{3\mathbf{p}_\mathrm{m}\cdot\mathbf{r}}{r^2}\mathbf{r} - \mathbf{p}_\mathrm{m}\right). \qquad \text{(P3.8.2.3)}$$

Example 4. Magnetic field of a linear current segment. Find the magnetic induction produced by a linear wire segment at the point A whose location is determined by its distance y to the straight line of the segment and two angles α_1 and α_2 as shown in Fig. P3.17.

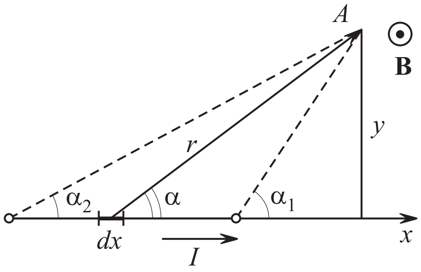

Figure P3.17. Magnetic field of a linear current segment.

Solution. The contribution of the element dx equals $dB = \frac{\mu_0}{4\pi}\frac{I\,dx}{r^2}\sin\alpha$; the field is perpendicular to the plane through the segment and the point A. It is convenient to integrate with respect to the angle α. Substituting $r = y/\sin\alpha$, changing the variable $x = -y\cot\alpha$, $dx = y\,d\alpha/\sin^2\alpha$, and integrating, we obtain

$$B = \frac{\mu_0 I}{4\pi y}(\cos\alpha_1 - \cos\alpha_2). \qquad \text{(P3.8.2.4)}$$

Setting $\alpha_1 = 0$ and $\alpha_2 = \pi$, we find the magnetic induction of an infinitely long straight current-carrying wire:

$$B = \frac{\mu_0 I}{2\pi y}. \qquad \text{(P3.8.2.5)}$$

Let us use (P3.8.2.4) to calculate the magnetic induction at the center of a rectangular loop to obtain

$$B = \frac{\mu_0}{4\pi}\frac{8I\sqrt{a^2+b^2}}{ab},$$

where a and b are the sides of the rectangle.

P3.8.3. Ampère's Circuital Law and Flux of Magnetic Induction

► **Ampère's circuital law.** From the Biot–Savart–Laplace law it can be inferred that the *circulation of magnetic induction* around any closed path is determined by the algebraic sum of the currents enclosed by the path (i.e., by the current passing through a surface spanning the path):

$$\oint \mathbf{B}\cdot d\mathbf{l} = \mu_0 I_{\text{enc}} = \mu_0 \int \mathbf{j}\cdot d\mathbf{s}. \qquad \text{(P3.8.3.1)}$$

The positive direction of the surface normal is determined by the right-hand rule.

► **Gauss's law for magnetism.** The *flux of magnetic induction* (*magnetic flux*) through any closed surface is zero:

$$\oint \mathbf{B}\cdot d\mathbf{s} = 0. \qquad \text{(P3.8.3.2)}$$

This means the magnetic field lines are everywhere continuous, or there are no *magnetic charges*, at which magnetic field lines would start or terminate.

The magnetic flux through any surface S is defined as

$$\Phi = \int_S \mathbf{B}\cdot d\mathbf{s}.$$

The SI unit of magnetic flux is the *weber* (Wb), with $1\,\text{Wb} = 1\,\text{T}\,\text{m}^2$. It is noteworthy that magnetic flux through a surface spanning a closed path does not depend on the shape of the surface. Therefore, it is meaningful to speak about magnetic flux through a closed path.

Example 1. Magnetic field of a solenoid. A solenoid is a helical coil formed by a long thin wire densely wound on a cylinder. A solenoid is characterized by the number of turns per unit length, n. The ideal model of a solenoid is an infinitely long thin cylindrical surface carrying current $i = nI$, with the current perpendicular to the axis. It is not difficult to verify that the magnetic field inside an ideal solenoid is everywhere parallel to the axis (to this end, one should consider the contributions of any two symmetric current elements). There is no magnetic field outside the solenoid. It follows from Ampère's circuital law and Gauss's law that $\mathbf{B} = \text{const}$ inside the solenoid. In order to find the magnitude B, one should apply Ampère's circuital law to a rectangular loop one side of which is inside the solenoid and parallel to its axis and the opposite side is outside as shown in Fig. P3.18. One finds that $Bl = \mu_0 il$. So the magnetic field inside the solenoid is

$$B = \mu_0 i = \mu_0 nI. \qquad \text{(P3.8.3.3)}$$

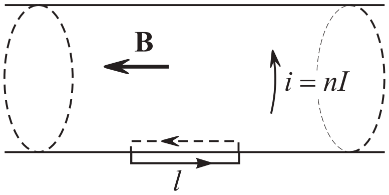

Figure P3.18. The magnetic field inside a solenoid can be found using Ampère's circuital law.

Example 2. Magnetic field of a rod. Find the magnetic field of an infinitely long straight rod of radius R carrying current I (Fig. P3.19). The magnetic lines are circles. Choose a circle of radius r as the closed path for Ampère's circuital law (P3.8.3.1). For $r > R$, we get $B \times 2\pi r = \mu_0 I$. For $r < R$, we find that $B \times 2\pi r = \mu_0 I(\pi r^2/\pi R^2)$. Hence,

$$B = \begin{cases} \dfrac{\mu_0 I}{2\pi r} & \text{for } r > R, \\ \dfrac{\mu_0 I r}{2\pi R^2} & \text{for } r < R. \end{cases}$$

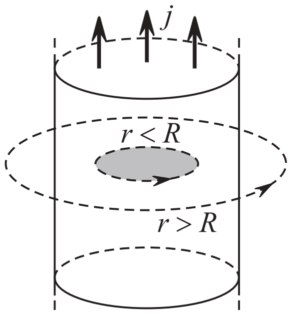

Figure P3.19. Magnetic field of a rod.

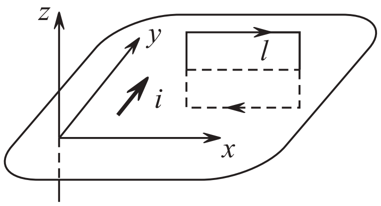

Figure P3.20. Magnetic field of a current plane.

Example 3. Magnetic field of a current plane. Consider an infinite plane (x, y) carrying current in the y-direction with a constant density i. The magnetic field above the plane, $z > 0$, is directed in the positive x-direction and that below the plane, $z < 0$, is in the negative x-direction (this can be verified by considering the contributions of two infinite straight currents symmetric relative to the point of interest). Choose a rectangular path with one side above the plane and in the x-direction and the other being symmetrically located below the plane (Fig. P3.20). Ampère's circuital law gives $2Bl = \mu_0 il$, and hence $B = \frac{1}{2}\mu_0 i$.

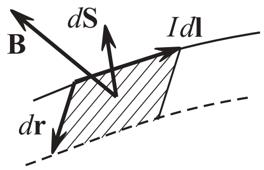

Figure P3.21. Displacement of a current loop element in a magnetic field.

► **Displacement of a current loop in a magnetic field.** When a loop carrying constant current I is moved or deformed in a constant magnetic field, the Ampère force performs work. The elementary work done by the Ampère force acting on a loop element equals

$$\delta A = d\mathbf{F} \cdot d\mathbf{r} = I[d\mathbf{l} \times \mathbf{B}] \cdot d\mathbf{r} = I\mathbf{B} \cdot [d\mathbf{r} \times d\mathbf{l}] = I\mathbf{B} \cdot d\mathbf{s},$$

where $d\mathbf{r}$ is the displacement of the loop element $d\mathbf{l}$, $d\mathbf{s}$ is the area element swept out by $d\mathbf{l}$ (Fig. P3.21). Integrating around the loop yields the elementary mechanical work done by the magnetic field in terms of the increment in the magnetic flux Φ through the loop:

$$\delta A_{\text{mech}} = I\,d\Phi. \qquad \text{(P3.8.3.4)}$$

It follows that a *potential function*, $W_{\text{p}} = -I\Phi$, can be associated with a loop carrying current I in a constant magnetic field. A change in this function equals (with the opposite sign) the work done by the Ampère force:

$$A_{\text{mech}} = -\Delta W_{\text{p}}, \qquad W_{\text{p}} = -I\Phi. \qquad \text{(P3.8.3.5)}$$

For a uniform field or a small plane loop, formula (P3.8.3.5) becomes

$$W_{\text{p}} = -\mathbf{p}_{\text{m}} \cdot \mathbf{B}, \qquad \text{(P3.8.3.6)}$$

where $\mathbf{p}_{\text{m}}$ is the magnetic moment of the current loop (see Subsection P3.8.1). Since this expression coincides with the potential energy of an electric dipole (see Subsection P3.2.2), it is natural to assume that the torque acting on a magnetic dipole and the force acting on a magnetic dipole in a nonuniform field should have the same form:

$$\mathbf{M} = \mathbf{p}_{\text{m}} \times \mathbf{B}, \qquad \mathbf{F} = p_{\text{m}} \frac{\partial \mathbf{B}}{\partial n} = p_{\text{m}x} \frac{\partial \mathbf{B}}{\partial x} + p_{\text{m}y} \frac{\partial \mathbf{B}}{\partial y} + p_{\text{m}z} \frac{\partial \mathbf{B}}{\partial z}, \qquad \text{(P3.8.3.7)}$$

where $\mathbf{n}$ is parallel $\mathbf{p}_{\text{m}}$. The equilibrium orientation of a magnetic dipole corresponds to the minimum potential energy, and hence it corresponds to $\mathbf{p}_{\text{m}} \parallel \mathbf{B}$. A current loop oriented in this manner is pulled into the region of a stronger magnetic field.

▶ **Maxwell's equations for a constant magnetic field.** In differential form, these equations are

$$\operatorname{curl}\mathbf{B} = \mu_0\mathbf{j}, \qquad \operatorname{div}\mathbf{B} = 0.$$

In integral form, these are equations (P3.8.3.1) and (P3.8.3.2).

▶ **Vector potential.** Maxwell's equation $\operatorname{div}\mathbf{B} = 0$ is satisfied if the magnetic field is sought in the form $\mathbf{B} = \operatorname{curl}\mathbf{A}$. The *vector potential* $\mathbf{A}$ is defined up to an arbitrary constant scalar field χ. The potential $\mathbf{A}' = \mathbf{A} + \operatorname{grad}\chi$ generates the same field $\mathbf{B}$ as $\mathbf{A}$; this expression is known as the *calibration transformation for a vector potential*. This transformation allows us to simplify the vector potential equation resulting from substituting $\mathbf{B} = \operatorname{curl}\mathbf{A}$ into the first Maxwell equation: $\operatorname{grad}\operatorname{div}\mathbf{A} - \Delta\mathbf{A} = \mu_0\mathbf{j}$. By choosing a scalar field χ such that $\operatorname{div}\mathbf{A} = 0$, we arrive at the equation

$$\Delta\mathbf{A} = -\mu_0\mathbf{j}.$$

The equation for each projection of $\mathbf{A}$ onto the coordinate axes is mathematically identical with the equation for the electric scalar potential (see Subsection P3.3.2). Consequently, the solution of these equations is given by

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int\frac{\mathbf{j}(\mathbf{r}')\,dV'}{|\mathbf{r}-\mathbf{r}'|}.$$

In particular, one can find the field of a small current loop (see (P3.8.2.3)) or prove the Biot–Savart law using this expression.

P3.9. Magnetic Field in Matter

P3.9.1. Magnetization. Molecular Currents

▶ **Magnetization of a material.** A material is regarded as magnetized if it creates a magnetic field inside itself and in the surrounding space in the absence of conduction currents. The source of magnetic field of a magnetized material are magnetic moments of atoms and molecules (atomic and molecular currents). In the absence of an external magnetic field, microscopic magnetic moments are either absent (in *diamagnetic* materials) or chaotically oriented (in *paramagnetic* materials). In both cases, the macroscopic magnetic field is zero. Constant magnets (made from *ferromagnetic* substances) are an exception; they can have a macroscopic magnetic field (see below for details). When a material is immersed in an external magnetic field, the material is magnetized and an induced volumetric magnetic moment arises. The degree to which a material is magnetized is characterized by the vector of *magnetization* $\mathbf{J}$, which is defined as the net magnetic moment per unit volume:

$$\mathbf{J} = \frac{\sum_i \mathbf{p}_{\mathrm{m}i}}{\Delta V} = n\langle\mathbf{p}_{\mathrm{m}}\rangle, \tag{P3.9.1.1}$$

where n is the molecule concentration. For the construction of macroscopic field equations for a magnetized material, it can be assumed for simplicity that all molecules in a small volume ΔV have the same magnetic moment $\langle\mathbf{p}_{\mathrm{m}}\rangle = \mathbf{J}/n = I_0\mathbf{s}$, or they can be represented as identical microscopic current loops.

▶ **Macroscopic molecular currents.** In characterizing magnetic field in matter, the huge number of oriented microscopic currents (magnetic moments) may be replaced by macroscopic molecular currents distributed across surface and volume. For a simple example, consider a homogeneous magnetized cylinder whose molecular current loops are all oriented along the cylinder axis. The different currents will compensate for one another inside the cylinder and there will be uncompensated current flowing along the surface. The magnetic

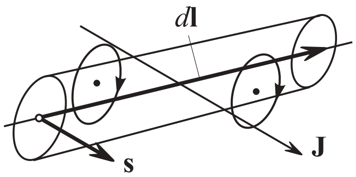

Figure P3.22. Magnetization creates a molecular current piercing the loop.

field of the surface current, $\mathbf{B}_{\text{mol}}$, will have, inside the cylinder, the *same direction* as the magnetization $\mathbf{J}$.

For an arbitrary closed path, its element $d\mathbf{l}$ inside the magnetized material pierces the current loops whose centers fall within the volume $dV = \mathbf{s}\cdot d\mathbf{l}$ (Fig. P3.22). The contribution into the molecular current enclosed by the path (i.e., the current crossing the surface that spans the path) equals $I_0 n\,dV = nI_0\mathbf{s}\cdot d\mathbf{l} = \mathbf{J}\cdot d\mathbf{l}$. Integrating gives

$$\oint \mathbf{J}\cdot d\mathbf{l} = I_{\text{enc}}^{\text{mol}}. \qquad \text{(P3.9.1.2)}$$

In differential form, this equation becomes $\operatorname{curl}\mathbf{J} = \mathbf{j}_{\text{mol}}$. These two equations allow us to express the current density at the surface as

$$\mathbf{i}_{\text{mol}} = \mathbf{J}\times\mathbf{n}, \qquad \text{(P3.9.1.3)}$$

where $\mathbf{n}$ is the outward normal to the surface of the magnetized material.

► **Magnetic field strength.** Macroscopic magnetic induction in a magnetized material results from averaging the microscopic field $\mathbf{B}$, highly varying at interatomic distances, over a small volume, though large enough to contain a large number of molecules: $\mathbf{B}(\mathbf{r}) = \langle\mathbf{B}_{\text{micro}}\rangle$. The magnetic induction is determined by both molecular and nonmolecular currents (such as conduction currents or convection currents):

$$\oint \mathbf{B}\cdot d\mathbf{l} = \mu_0\bigl(I_{\text{enc}}^{\text{nonmol}} + I_{\text{enc}}^{\text{mol}}\bigr). \qquad \text{(P3.9.1.4)}$$

Since the molecular current distribution is usually unknown, it is convenient to introduce a new vector quantity called the *magnetic field strength* and defined as

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{J}, \qquad \text{(P3.9.1.5)}$$

which is only defined by nonmolecular currents:

$$\oint \mathbf{H}\cdot d\mathbf{l} = I_{\text{enc}}^{\text{nonmol}} \qquad \text{(P3.9.1.6)}$$

(Ampère's circuital law for the magnetic field strength). In differential form, this expression becomes: $\operatorname{curl}\mathbf{H} = \mathbf{j}^{\text{nonmol}}$. The SI unit for magnetic field strength is A/m.

► **Boundary conditions at the interface between different magnetic materials.** For the normal and tangential components of $\mathbf{B}$, $\mathbf{H}$, and $\mathbf{J}$, the following boundary conditions hold at the interface between two magnetic media:

$$B_{2n} = B_{1n}, \quad \mathbf{H}_1\times\mathbf{n} - \mathbf{H}_2\times\mathbf{n} = \mathbf{i}_{\text{nonmol}}, \quad \mathbf{J}_1\times\mathbf{n} - \mathbf{J}_2\times\mathbf{n} = \mathbf{i}_{\text{mol}}, \qquad \text{(P3.9.1.7)}$$

where $\mathbf{n}$ is a normal vector pointing from medium 1 to medium 2 and $|\mathbf{H}\times\mathbf{n}| = H_\tau$. If there is no nonmolecular current at the interface, the second condition can be rewritten as $H_{1\tau} = H_{2\tau}$. The first condition follows from Gauss's law (P3.8.3.2) for $\mathbf{B}$, the second follows from (P3.9.1.6), and the third from (P3.9.1.3). There is a formal analogy in boundary conditions between $\mathbf{H}$ and $\mathbf{E}$ and also between $\mathbf{B}$ and $\mathbf{D}$ (see Subsection P3.5.1).

► **Isotropic magnetic material.** The magnetization **J** at a given point inside a magnetic material arises due to an external magnetic field and is determined by the vector **B**. For not too strong fields, **J** depends on **B** linearly and in an *isotropic material*, in addition, $\mathbf{J} \parallel \mathbf{B}$. Then both vectors **J** and **H** are proportional to **B** and to each other. For historical reasons (due to the analogy between **E** and **H** in boundary conditions), the *magnetic susceptibility* is defined by

$$\mathbf{J} = \chi\mathbf{H}, \qquad \mathbf{J} = \frac{\chi}{\mu_0(1+\chi)}\mathbf{B}. \tag{P3.9.1.8}$$

From (P3.9.1.5) we get

$$\mathbf{B} = \mu_0(1+\chi)\mathbf{H} = \mu_0\mu\mathbf{H}, \tag{P3.9.1.9}$$

where μ is the *magnetic permeability* of the material. For paramagnetic materials, $\chi > 0$. For diamagnetic materials, $\chi < 0$. In both cases, $|\chi| \ll 1$ (weak magnetic properties). For ferromagnetic materials, $\chi \gg 1$.

The boundary conditions at the interface between two isotropic magnetic media are

$$B_{1n} = B_{2n}, \quad \frac{B_{1\tau}}{\mu_1} = \frac{B_{2\tau}}{\mu_2}; \qquad \mu_1 H_{1n} = \mu_2 H_{2n}, \quad H_{1\tau} = H_{2\tau}.$$

► **Calculation of fields in the presence of a magnetic material.** The equations for **B** and **H** together with the above boundary conditions determine magnetic field in the whole space if the nonmolecular currents are known. Below are some important properties of solutions:

1. The volume density of the molecular current is proportional to that of the nonmolecular current: $\mathbf{j}^{\text{mol}} = \chi\mathbf{j}^{\text{nonmol}}$. If there are no nonmolecular currents, then surface molecular currents only arise when the material is magnetized.

2. If the boundaries of a homogeneous magnetic material are everywhere tangent to the original ($\mathbf{B}_0$) magnetic lines, then $\mathbf{B} = \mu\mathbf{B}_0$ everywhere in the magnetic material.

Example 1. A long paramagnetic cylinder is immersed in a uniform magnetic field whose induction $\mathbf{B}_0$ is parallel to the cylinder axis. Find the magnetic induction, magnetic field strength, and magnetization inside the cylinder.

Solution. Everywhere inside the cylinder (except for its edges) we have: $\mathbf{B} = \mu\mathbf{B}_0$, $\mathbf{H} = \mathbf{H}_0 = \mathbf{B}_0/\mu_0$, and $\mathbf{J} = (\mu - 1)\mathbf{B}_0/\mu_0$.

Example 2. An infinitely long rod of circular cross-section, made from a material with magnetic permeability $\mu > 1$, carries a constant current with density **j**. Find the magnetic field strength, magnetic induction, and magnetization. Also calculate the surface and volume density of the molecular current.

Solution. The molecular current within the rod volume has the density $\mathbf{j}^{\text{mol}} = (\mu-1)\mathbf{j}$. The field strength H can be found using Ampère's circuital law (P3.9.1.6). We get $H = I/(2\pi r)$ for $r > R$ and $H = Ir/(2\pi R^2)$ for $r < R$ (see example 2, Subsection P3.8.3). The magnetic induction equals $B = \mu_0 H$ for $r > R$ and $B = \mu_0\mu H$ for $r < R$; it has a jump discontinuity at the surface current. The magnetization equals $J = (\mu - 1)Ir/(2\pi R^2)$ ($J = 0$ outside the rod). The magnetic lines are concentric circles and the vectors **H** and **J** are parallel to **B**. The surface molecular current is $\mathbf{i} = \mathbf{J}\times\mathbf{n}$; it equals $i_{\text{mol}} = (\mu - 1)I/(2\pi R)$ in magnitude and is opposite to **j** in direction. The total molecular current is zero, and hence $j_{\text{mol}}\pi R^2 = i_{\text{mol}}2\pi R$.

P3.9.2. Types of Magnetic Materials and Magnetization Mechanisms

Atomic and molecular magnetic moments are created by (a) the motion of electrons in closed orbits (orbital magnetic moments) and (b) the intrinsic magnetic moments of electrons related to the intrinsic mechanical angular momentum—spin (spin magnetic moments). Each type of magnetic moment is proportional to the corresponding mechanical angular momentum: $\mathbf{p}_{\text{m}} = -g(e/2m)\mathbf{L}$, where g is a dimensionless coefficient called the *gyromagnetic ratio*. For orbital moment, $g = 1$. For spin moment, $g = 2$.

► **Diamagnets. Larmor's mechanism of diamagnetic magnetization.** In the absence of external magnetic fields, the magnetic moments of atomic (molecular) electrons in a diamagnetic material (diamagnet) compensate for one another. When a magnetic field is applied, a torque starts to act on each magnetic moment. As a result, the magnetic moment rotates about the magnetic field direction (*Larmor precession*):

$$\frac{d\mathbf{L}}{dt} = \mathbf{p}_\mathrm{m} \times \mathbf{B} = -g\frac{e}{2m}\mathbf{L} \times \mathbf{B} = \left(g\frac{e}{2m}\mathbf{B}\right) \times \mathbf{L}.$$

The vector **L** (and also $\mathbf{p}_\mathrm{m}$) rotates with an angular velocity $\boldsymbol{\omega}_\mathrm{L} = g\frac{e}{2m}\mathbf{B}$ (Fig. P3.23). This rotation gives rise to an additional magnetic moment of each electron, $\Delta\mathbf{p}_\mathrm{m}$, opposing to **B** (independently of the original direction of $\mathbf{p}_\mathrm{m}$).

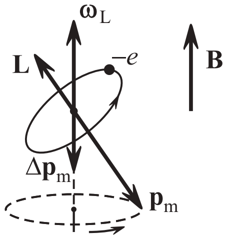

Figure P3.23. Larmor precession in diamagnetic materials.

► **Orientational mechanism of magnetization in paramagnetic materials.** In paramagnets, the magnetic moments of electrons do not compensate for one another even in the absence of external magnetic fields. The weak mechanism of diamagnetic magnetization is superimposed by a stronger orientational mechanism: in the presence of magnetic field, the orientation of the magnetic moment along the field is less costly energetically. In weak fields ($p_\mathrm{m}B \ll kT$), the average magnetic moment is proportional to the magnetic induction. In strong fields, saturation is achieved, where all magnetic moments are oriented along the field and the magnetization virtually does not change. The orientational mechanism of paramagnetic magnetization is similar to that of polar dielectric polarization (see Subsection P3.5.2). In particular, the conclusion about inverse proportionality between the susceptibility and temperature remains valid.

► **Ferromagnets** are materials that show spontaneous magnetization in macroscopic (compared to interatomic distances) regions called *domains*. The domains are small ($\sim 10^{-5}$ to 10^{-4} m) as compared to the size of the sample under examination. Within each domain, spin magnetic moments are parallel to one another, which results from a special quantum interaction among them. In the absence of external magnetic fields, the domains are oriented in such a way that the average magnetization of the material is zero. When an external magnetic field is applied, the domains oriented along the field increase in size and the domains oriented against the field decrease. For stronger fields, the domains tend to turn as a whole to be oriented along the field. In strong enough fields, all domains are oriented along the field and saturation is achieved, when the magnetization reaches its maximum magnitude, J_sat, called the *saturation magnetization*.

The relationship between **J** and **H** is nonlinear (see Fig. P3.24). This means that the magnetic permeability, determined by the formula $\mathbf{J} = (\mu - 1)\mathbf{H}$, depends on H and attains very large values ($\mu \sim 10^5$–10^6). When the external field is removed, some *residual magnetization*, J_res, is observed. The field strength required to completely demagnetize the sample, H_C, is called the *coercive field* or *coercive force*. Ferromagnetic materials possessing a large coercive force are used to make permanent magnets. The dependence $J(H)$ for direct and reverse magnetization to saturation is called a *hysteresis loop*.

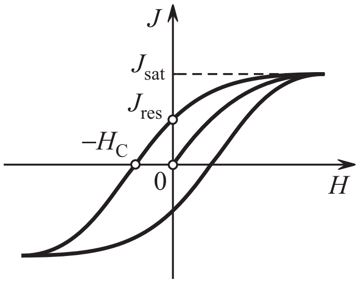

Figure P3.24. Hysteresis loop of a ferromagnetic material.

When heated above the *Curie temperature*, the material loses its ferromagnetic properties (spontaneous magnetization of domains disappears) and becomes a paramagnet. A second-order phase transition occurs at the Curie point.

P3.10. Electromagnetic Induction

P3.10.1. Faraday's Law of Electromagnetic Induction

► **Motion of a wire loop in a constant field.** When a current-carrying wire loop is moved or deformed in a constant magnetic field, the Ampère forces are everywhere perpendicular to the current and, according to (P3.8.3.4), perform mechanical work $\delta A_{\perp} = I\,d\Phi$. Since the total work done by the Lorentz force on any moving particle is zero, the same work (but opposite in sign) is performed by the magnetic field (i.e., the tangential component of the Lorentz force) on free charges moving through the wire. This gives rise to an additional electromotive force (*induced emf*) in the loop:

$$\delta A_{\parallel} = -I\,d\Phi = -(I\,dt)\frac{d\Phi}{dt} = \mathscr{E}\,dq, \qquad \mathscr{E} = -\frac{d\Phi}{dt}. \tag{P3.10.1.1}$$

► **Static loop. Solenoidal electric field.** It follows from Einstein's relativity principle that an induced emf must also arise when a static current-carrying loop is placed in a time varying magnetic field. Consider, for example, a permanent magnet approaching a closed loop at a constant velocity. In the magnet-associated inertial frame of reference, the field produced by the magnet is time independent and the induced emf in the loop is accounted for by the action of the Lorentz force. On the other hand, in the loop-associated inertial frame, the same emf is observed, but it cannot be caused by magnetic forces. Hence, the varying magnetic field must bring about an electric field. Since the work of this field around the loop must be nonzero, it follows that the field is not electrostatic—it is called a *solenoidal* (*eddy*) *electric field*. Thus, an induced emf arises in the loop regardless of the reason for which the magnetic flux through the loop changes. The induced emf equals

$$\mathscr{E} = -\frac{d\Phi}{dt}. \tag{P3.10.1.2}$$

This equation is known as *Faraday's law of electromagnetic induction*. The average induced emf in a finite time Δt equals $\mathscr{E}_{\text{ave}} = -\Delta\Phi/\Delta t$. The charge passed through the loop in this time is

$$q = \int I\,dt = \int \frac{\mathscr{E}}{R}\,dt = -\frac{\Delta\Phi}{R},$$

where R is the resistance of the loop.

► **Lenz's rule.** The direction of the emf is determined by its sign. It should be reminded (see Subsection P3.8.3) that the loop is *oriented*: the positive direction of the normal to the surface, in the formula for the calculation of Φ, and the positive direction of tracing the path, which determines the sign of the emf, are related by the right-hand rule. The minus sign in (P3.10.1.2) is in accordance with *Lenz's rule*. It states that the induced emf must be directed so that the magnetic field produced by the induced current flowing in the same direction, generates a magnetic flux (induced magnetic flux) of opposite sign as compared to the flux change that caused the emf. Lenz's rule is closely associated with the law of conservation of energy.

► **Motion of part of a loop.** When a loop element $d\mathbf{l}$ moves with a velocity $\mathbf{v}$, the work done by the Lorentz force to transfer a test charge q along $d\mathbf{l}$ equals $\delta A = q[\mathbf{v}\times\mathbf{B}]\cdot d\mathbf{l} = q\,d\mathscr{E}$. The electromotive forces induced by an arbitrarily moving portion of the loop in an arbitrary constant field and by a translationally moving portion of the loop in a uniform constant field are, respectively, equal to

$$\mathscr{E} = \int [\mathbf{v}\times\mathbf{B}]\cdot d\mathbf{l}, \qquad \mathscr{E} = [\mathbf{v}\times\mathbf{B}]\cdot \mathbf{l}, \tag{P3.10.1.3}$$

where $\mathbf{l}$ is the vector connecting the beginning and the end of the moving portion. If the wire loop is unclosed, a potential difference will arise between its ends, which is equal to the emf (since the electrostatic force is equalized by the Lorentz force at any point of the wire).

Example 1. A crosspiece of length l and resistance r moves with a speed v along two horizontal rails immersed in a vertical magnetic field (Fig. P3.25). Find the current through the crosspiece.

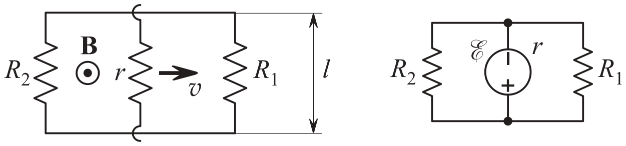

Figure P3.25. To example 1. An equivalent diagram is on the right.

Solution. Suppose that the rails are connected by a resistance R_1 on one side and a resistance R_2 on the other side. In order to find the current through the crosspiece as it moves with the speed v, we use (P3.10.1.3) to calculate the emf induced in the crosspiece: $\mathscr{E} = Bvl$. The fact that the emf is localized within the moving crosspiece enables us to draw an equivalent diagram, shown on the right, and eventually find the current: $I = \mathscr{E}/(r + R)$, where $R = R_1R_2/(R_1 + R_2)$.

► **Equations of solenoidal electric field.** By writing down Faraday's law (P3.10.1.2) for a static loop in a varying magnetic field, we obtain a solenoidal electric field equation in integral form:

$$\oint \mathbf{E}\cdot d\mathbf{l} = -\frac{d}{dt}\int \mathbf{B}\cdot d\mathbf{s}. \tag{P3.10.1.4}$$

Since the circulation of the electrostatic field strength around a closed path is zero, the total electric field satisfies the same equation. The differential form of this equation is

$$\operatorname{curl}\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t}. \tag{P3.10.1.5}$$

Equations (P3.10.1.4) and (P3.10.1.5) represent one of Maxwell's equation (in integral and differential forms).

► **Quasistationary approximation.** If the currents and fields change sufficiently slowly, time delay of processes can be neglected. This means that the varying magnetic field is determined at any point (using the Biot–Savart–Laplace formulas) by the currents at the same instant of time. This also means that the electrostatic field is determined by the charge distribution at the same instant and the solenoidal electric field is calculated from the varying magnetic field using (P3.10.1.4) and (P3.10.1.5). It is this approximation that is used for the analysis of alternating current circuits. For this approximation to be valid it is required that the characteristic current variation time, τ, is large compared to (a) the local relaxation times (e.g., the time in which an equilibrium charge distribution is established in the conductor) and (b) the delay time $\tau_{\rm d} = l/c$, where l is the characteristic size of the system and c is the speed of light. The latter condition suggests that we neglect the displacement current in Maxwell's equations (i.e., we assume that $\operatorname{curl}\mathbf{B} = \mu_0\mathbf{j}$).

Example 2. Let us estimate the time τ of charge redistribution in a conductor with resistivity ρ. We assume that there is a local volume charge somewhere in the conductor. The charge redistribution within an arbitrary volume is governed by the equation $\dfrac{dq}{dt} = -\oint j_n\,ds = -\dfrac{1}{\rho}\oint E_n\,ds = -\dfrac{q}{\varepsilon_0\rho}$, whence $q = q_0\exp\Bigl(-\dfrac{t}{\varepsilon_0\rho}\Bigr)$, and so $\tau = \varepsilon_0\rho$.

P3.10.2. Self-Induction and Mutual Induction

► **Self-induction.** Consider a stationary wire loop carrying current I changing in time. The magnetic induction produced by the current will change at a rate proportional to dI/dt. Hence, the solenoidal electric field arising at every point in space is also proportional to dI/dt. This field induces the emf

$$\mathscr{E}_{\rm self} = -L\frac{dI}{dt} \qquad \text{(P3.10.2.1)}$$

in the loop. This emf is called the *self-induction electromotive force*. The quantity L is called the *coefficient of self-inductance* or the *inductance* of the loop. The SI unit of inductance is the *henry* (H). The quantity $\Phi_{\rm self} = LI$ is the self-induced magnetic flux through the area spanned by the loop. Using (P3.10.2.1), we get

$$\mathscr{E}_{\rm self} = -\frac{d\Phi_{\rm self}}{dt}. \qquad \text{(P3.10.2.2)}$$

This formula remains valid not only if I changes but also if L does.

Note that the definition of L as the coefficient of proportionality between magnetic flux and current meets with difficulties: for a linear wire, the magnetic flux is infinitely large and for a wire of finite thickness, the usual definition of the flux is unsuitable, since it is unclear how to choose a path.

Example 1. Inductance of a solenoid. Consider a solenoid of length l and cross-sectional area S ($l \gg \sqrt{S}$) carrying current with surface density $i = (N/l)I$, where N is the number of turns and I is the current in the coil. The magnetic field in the solenoid is uniform and its magnetic induction (in the ideal solenoid approximation) is given by (P3.8.3.3): $B = \mu_0\mu i$. For a changing current, the circulation of the solenoidal electric field is calculated as $\oint \mathbf{E}\cdot d\mathbf{l} = -S\dfrac{dB}{dt} = -\mu_0\mu S\dfrac{di}{dt}$ and the emf induced along the whole length of the winding equals $\mathscr{E}_c = -N\mu_0\mu S\dfrac{di}{dt} = -\dfrac{N^2\mu_0\mu S}{l}\dfrac{dI}{dt}$. The inductance and the magnetic flux are expressed as

$$L = \frac{N^2\mu_0\mu S}{l}, \qquad \Phi_{\rm self} = LI = NBS. \qquad \text{(P3.10.2.3)}$$

Example 2. Since the self-induction emf is proportional to the current change rate, the presence of inductance in a circuit impedes too rapid changes of the current. Suppose that a resistor R and an induction coil L are in a closed circuit with a source of constant emf $\mathscr{E}$. The current increases from zero to $I_0 = \mathscr{E}/R$ in a finite time and this process is governed by the equation $\mathscr{E} - L(dI/dt) = IR$, whose solution is $I = I_0[1 - \exp(-tR/L)]$. Hence, the characteristic time of current increasing equals $\tau = L/R$.

► **Mutual induction.** If there are two stationary loops, the current change in one of them will bring about in the other loop an electromotive force proportional to dI/dt, called the *mutual induction emf.* We have

$$\mathscr{E}_{21} = -L_{21}\frac{dI_1}{dt}, \qquad \mathscr{E}_{12} = -L_{12}\frac{dI_2}{dt}, \tag{P3.10.2.4}$$

where L_{12} and L_{21} are the so-called *coefficients of mutual inductance*; the *mutuality theorem* (in a non-ferromagnetic medium) states that $L_{12} = L_{21}$. The magnetic flux through loop 1 produced by the current through loop 2 is given by $\Phi_{12} = L_{12}I_2$. This quantity is meaningful for two linear currents also.

In the quasistationary approximation, the vector potential can be calibrated in the same way as in magnetostatics (see Subsection P3.8.3). The magnetic flux can be written as

$$\Phi = \int \mathbf{B}\, d\mathbf{s} = \int \operatorname{curl} \mathbf{A}\, d\mathbf{s} = \oint \mathbf{A}\, d\mathbf{l}.$$

For Φ_{12}, we get

$$\Phi_{12} = \oint \mathbf{A}\, d\mathbf{l}_1 = \oint\oint \frac{\mu_0}{4\pi}\frac{I_2\, d\mathbf{l}_2}{r_{12}} d\mathbf{l}_1.$$

It is apparent that the expression of L_{12} is symmetric with respect to the subscripts 1 and 2.

P3.10.3. Energy of Magnetic Field

► **Energy of a loop.** The energy of a current-carrying loop is the work done by external forces and sources to create the current (or, equivalently, the work of the solenoidal field to destroy the current). The work done by the current source against the self-induction emf of the loop is calculated as

$$W = A = \int_0^\infty L\frac{dI}{dt}(I\,dt) = \frac{LI^2}{2}. \tag{P3.10.3.1}$$

Within a consistent field approach, the energy must be related with the magnetic field rather than the loop current. The easiest way to find the dependence between the energy density and the magnetic induction is to consider a long solenoid with a uniform field in it. We have

$$\frac{W}{V} = \frac{1}{Sl}\frac{\mu_0\mu N^2 S}{l}\frac{I^2}{2} = \left(\frac{\mu_0\mu IN}{l}\right)^2\frac{1}{2\mu_0\mu} = \frac{B^2}{2\mu_0\mu} = \frac{BH}{2}. \tag{P3.10.3.2}$$

If the relationship between **B** and **H** is nonlinear, the energy density is expressed as

$$\frac{W}{V} = \int \mathbf{H}\cdot d\mathbf{B}.$$

Using the mutuality theorem ($L_{12} = L_{21}$), we can generalize (P3.10.3.1) to include two loops:

$$\begin{aligned} W &= \int_0^\infty \left(L_1\frac{dI_1}{dt} + L_{12}\frac{dI_2}{dt}\right)(I_1\,dt) + \int_0^\infty \left(L_2\frac{dI_2}{dt} + L_{21}\frac{dI_1}{dt}\right)(I_2\,dt) \\ &= \frac{L_1I_1^2}{2} + \frac{L_2I_2^2}{2} + \frac{L_{12}+L_{21}}{2}I_1I_2 = \frac{I_1\Phi_1}{2} + \frac{I_2\Phi_2}{2}. \end{aligned} \tag{P3.10.3.3}$$

Likewise, for several loops, we have

$$W=\sum_i \frac{I_i\Phi_i}{2},$$

where Φ_i is the total magnetic flux through the ith loop. It can be seen that the current energy includes the energy of each individual current and the energy of their interaction. With the magnetic field equations, it can be shown that the above expression always converts to

$$W=\int \frac{B^2}{2\mu_0\mu}\,dV=\int \frac{BH}{2}\,dV. \qquad \text{(P3.10.3.4)}$$

► **Law of conservation of energy and force calculation.** If the positions of the bodies and current-carrying loops change sufficiently slowly and isothermally, the total work done by external forces and current sources is spent on changing the magnetic field energy and releasing Joulean heat. Moreover, the work done by external forces on each body equals the work done by the magnetic forces on the body. If the currents are maintained constant, one has only to take into account the additional work done by the current sources against the induction emf; the work done by the original emf equals the heat released and, hence, they cancel each other out in the law of conservation of energy.

Example. Consider two loops carrying currents I_1 and I_2 that are maintained constant. Let us calculate the work A^e done by external forces for a slow movement of the first loop. According to (P3.10.3.3), the changing part of the energy is only the energy of current interaction, which is equal to $W_{12}=L_{12}I_1I_2=I_1\Phi_{12}$, where Φ_{12} is the magnetic flux created by the second current and passing through the first loop. At the same time, the potential function of the loop in an external field is $W_\mathrm{p}=-I_1\Phi_{12}$ (see (P3.8.3.5)), and hence $\delta A^\mathrm{e}=-I_1\,d\Phi_{12}=-I_1I_2\,dL_{12}$. The apparent contradiction is resolved by taking into account the additional work done by the current sources against the induction emf in either loop. This work equals $\delta A^\mathrm{sources}=I_1\,d\Phi_{12}+I_2\,d\Phi_{21}=I_1\,dL_{12}\,I_2+I_2\,dL_{21}\,I_1$, which is twice the field energy change $dW_{12}=I_1I_2\,dL_{12}$, and $\delta A^\mathrm{e}+\delta A^\mathrm{sources}=dW_{12}$.

P3.11. Maxwell's Equations

P3.11.1. Maxwell's Equations in Vacuum and Matter

► **Displacement current.** To generalize the electromagnetic field equations in vacuum to cover varying fields, only one of the four equations (see Subsections P3.3.2, P3.8.3, and P3.10.1) has to be changed. Three equations hold true in all cases but Ampère's circuital law (P3.8.3.1) turns out to be invalid for varying fields and currents. According to (P3.8.3.1), the current $\int \mathbf{j}\cdot d\mathbf{s}$ must be the same for any two surfaces spanning the closed path; if the charge within the volume between the two selected surfaces changes, this statement will contradict the principle of conservation of charge. For example, when a parallel-plate capacitor is being charged (Fig. P3.26), the current through one of the surfaces is $I=dq/dt$ and there in no current through the other surface (passing between the plates).

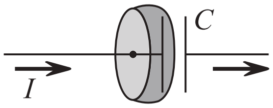

Figure P3.26. To the determination of the displacement current.

To remove the contradiction, Maxwell introduced a *displacement current* in the equation, with its density proportional to the rate of change of the electric field $\mathbf{D}$:

$$\mathbf{j}_\mathrm{d}=\frac{\partial \mathbf{D}}{\partial t}. \qquad \text{(P3.11.1.1)}$$

In a dielectric medium, the displacement current density equals

$$\mathbf{j}_{\rm d} = \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \frac{\partial \mathbf{P}}{\partial t}. \tag{P3.11.1.2}$$

The first term represents the displacement current density in vacuum and the second term is the real current density due to motion of bound charges in changing polarization. The displacement current through a surface equals $d\Phi_D/dt$, where Φ_D is the flux of **D** through the surface. The introduction of the displacement current removes the contradiction with the principle of conservation of charge. For example, when a parallel-plate capacitor is being charged, the displacement current through a surface between the plates, $I_{\rm d} = S(dD/dt) = S(d\sigma/dt) = dq/dt$, equals the current flowing through the wires.

► **System of Maxwell's equations in vacuum.** Once the displacement current has been introduced, the system of Maxwell's equations in differential form becomes:

1) $\operatorname{curl}\mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t}$ (Faraday's law of induction)

2) $\operatorname{curl}\mathbf{B} = \mu_0\mathbf{j} + \mu_0\varepsilon_0\dfrac{\partial \mathbf{E}}{\partial t}$ (Ampère's circuital law with Maxwell's correction)

3) $\operatorname{div}\mathbf{E} = \dfrac{\rho}{\varepsilon_0}$ (Gauss's law)

4) $\operatorname{div}\mathbf{B} = 0$ (Gauss's law for magnetism)

The system of Maxwell's equations in integral form:

$$1)\ \oint \mathbf{E}\cdot d\mathbf{l} = -\frac{d}{dt}\int \mathbf{B}\cdot d\mathbf{s}, \qquad 2)\ \oint \mathbf{B}\cdot d\mathbf{l} = \mu_0\int \mathbf{j}\cdot d\mathbf{s} + \mu_0\varepsilon_0\frac{d}{dt}\int \mathbf{E}\cdot d\mathbf{s},$$

$$3)\ \oint \mathbf{E}\cdot d\mathbf{s} = \frac{1}{\varepsilon_0}\int \rho\, dV, \qquad 4)\ \oint \mathbf{B}\cdot d\mathbf{s} = 0.$$

The charge density and the current density satisfy the relation

$$\oint \mathbf{j}\cdot d\mathbf{s} = -\frac{d}{dt}\int \rho\, dV \qquad \Big(\operatorname{div}\mathbf{j} = -\frac{\partial\rho}{\partial t}\Big),$$

which expresses the law of conservation of charge and is a corollary of Maxwell's equations.

► **Solution of Maxwell's equations in the form of delayed potentials.** The first and fourth Maxwell equations can be satisfied by using the fields in the form

$$\mathbf{B} = \operatorname{curl}\mathbf{A}, \qquad \mathbf{E} = -\operatorname{grad}\varphi - \frac{\partial \mathbf{A}}{\partial t}.$$

The scalar potential φ and vector potential **A** are defined up to the *calibration transformation*:

$$\mathbf{A}' = \mathbf{A} + \operatorname{grad}\chi, \qquad \varphi' = \varphi - \frac{\partial\chi}{\partial t}.$$

The equations for the potentials resulting from the second and third Maxwell equations can be simplified using the calibration transformation. To this end, the potentials are required to satisfy *Lorentz's calibration condition*:

$$\operatorname{div}\mathbf{A} + \frac{1}{c^2}\frac{\partial\varphi}{\partial t} = 0.$$

In this case, the equations for the potentials satisfy *d'Alembert's equation*:

$$\Delta\mathbf{A} - \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} = -\mu_0\mathbf{j}, \qquad \Delta\varphi - \frac{1}{c^2}\frac{\partial^2\varphi}{\partial t^2} = -\frac{\rho}{\varepsilon_0}.$$

If the right-hand side vanishes, d'Alembert's equation becomes the wave equation (see Chapter P4). In the presence of time-varying current and charge sources, d'Alembert's equations have solutions in the form of *delayed potentials*:

$$\mathbf{A}(\mathbf{r},t)=\frac{\mu_0}{4\pi}\int\frac{\mathbf{j}\big(\mathbf{r}',t-\frac{|\mathbf{r}-\mathbf{r}'|}{c}\big)\,dV'}{|\mathbf{r}-\mathbf{r}'|},\qquad \varphi(\mathbf{r},t)=\frac{1}{4\pi\varepsilon_0}\int\frac{\rho\big(\mathbf{r}',t-\frac{|\mathbf{r}-\mathbf{r}'|}{c}\big)\,dV'}{|\mathbf{r}-\mathbf{r}'|}.$$

The potentials at a given point in space and a given time instant are determined by the charges and currents at all other points at preceding instants of time, taking into account the time required for a light signal to deliver information to the given point about the current and charge values at a remote point.

► **Maxwell's equations in a medium** are written as

differential form	integral form
1) $\operatorname{curl}\mathbf{E}=-\dfrac{\partial\mathbf{B}}{\partial t}$,	$\oint\mathbf{E}\cdot d\mathbf{l}=-\dfrac{d}{dt}\int\mathbf{B}\cdot d\mathbf{s}$,
2) $\operatorname{curl}\mathbf{H}=\mathbf{j}+\dfrac{\partial\mathbf{D}}{\partial t}$,	$\oint\mathbf{H}\cdot d\mathbf{l}=\int\mathbf{j}\cdot d\mathbf{s}+\dfrac{d}{dt}\int\mathbf{D}\cdot d\mathbf{s}$,
3) $\operatorname{div}\mathbf{D}=\rho$,	$\oint\mathbf{D}\cdot d\mathbf{s}=\int\rho\,dV$,
4) $\operatorname{div}\mathbf{B}=0$,	$\oint\mathbf{B}\cdot d\mathbf{s}=0$.

These serve to determine the four vector quantities **E**, **D**, **B**, and **H**. Maxwell's equations in a medium must be supplemented by the *material equations* relating **D** to **E** and **H** to **B** and characterizing the electric and magnetic properties of the medium. For isotropic linear media, the equations are

$$\mathbf{D}=\varepsilon_0\varepsilon\mathbf{E},\qquad \mathbf{B}=\mu_0\mu\mathbf{H}.$$

Maxwell's equations can also be used to obtain boundary conditions for **E**, **D**, **B**, and **H** (see Subsections P3.5.1 and P3.9.1).

P3.11.2. Law of Conservation of Energy for Electromagnetic Fields

► **Poynting vector.** The following equation for any volume V bounded by a surface S can be derived from Maxwell's equations:

$$\frac{d}{dt}\int\Big(\frac{\varepsilon_0E^2}{2}+\frac{B^2}{2\mu_0}\Big)dV+\int\big(\mathbf{j}\cdot\mathbf{E}\big)\,dV+\frac{1}{\mu_0}\oint\big[\mathbf{E}\times\mathbf{B}\big]\cdot d\mathbf{s}=0.$$

The first term characterizes the rate of change of electromagnetic field energy in the volume. In general, the formulas obtained before for electrostatic and magnetostatic fields remain valid for the electromagnetic field energy. The second term represents the power of the field work on the charged particles within the volume of interest; it is equal to the rate of change of their kinetic energy. It follows from the law of conservation of energy that the third term has the meaning of (outward) electromagnetic energy flux through the closed surface bounding the volume. The energy flux at a given point in space (the *Poynting vector*) is determined by the vectors **E** and **B** at this point:

$$\mathbf{S}=\frac{1}{\mu_0}\mathbf{E}\times\mathbf{B}=\mathbf{E}\times\mathbf{H}. \qquad \text{(P3.11.2.1)}$$

The last expression represents the electromagnetic energy flux in a substance. The energy density in a substance is expressed as

$$w = \tfrac{1}{2}\mathbf{E}\cdot\mathbf{D} + \tfrac{1}{2}\mathbf{B}\cdot\mathbf{H}. \qquad \text{(P3.11.2.2)}$$

Example. Consider the process of charging a parallel-plate capacitor with circular plates separated by a distance b. The rate of change of the energy in a cylinder of radius r (smaller than the plate radius) equals $\dfrac{d}{dt}\big(\tfrac{1}{2}\varepsilon_0\varepsilon E^2\big)(\pi r^2 b) = \pi r^2 bE\dfrac{dD}{dt}$. The magnetic field strength can be found from the second Maxwell equations: $H\times 2\pi r = \dfrac{dD}{dt}\pi r^2$ (displacement current is on the right-hand side). We find that the rate of energy influx through the lateral surface of the cylinder, $S(2\pi rb) = HE(2\pi rb) = E\pi r^2 b\dfrac{dD}{dt}$, equals the energy change in the volume.

P3.11.3. Relativistic Properties of Fields

► **Transformation of charges and currents.** When changing from one inertial frame of reference to another, both the electromagnetic field sources (charge and current densities) and the fields themselves change. However, Maxwell's equations remain the same. The source transformation formulas for ρ (density of moving charge) and $\mathbf{j} = \rho\mathbf{v}$ have the simplest form. Denoting by ρ_0 the charge density in the inertial frame where $\mathbf{j} = \mathbf{v} = 0$ and taking into account length contraction (see Subsection P1.10.3), we get

$$\rho = \gamma\rho_0, \quad \mathbf{j} = \gamma\rho_0\mathbf{v}; \qquad \gamma = \frac{1}{\sqrt{1 - v^2/c^2}}.$$

Comparing with the momentum-energy 4-vector, we see that $(\mathbf{j}, \rho c)$ also forms a 4-vector, which means that $\mathbf{j}$ and ρc are transformed in the same manner as $(\mathbf{r}, ct)$, by the Lorentz formulas.

► **Transformation of fields.** Once we know the transformation formulas for field sources, it is not difficult to find out how $\mathbf{E}$ and $\mathbf{B}$ are transformed. The formulas are

$$\begin{aligned} \mathbf{E}'_{\parallel} &= \mathbf{E}_{\parallel}, & \mathbf{E}'_{\perp} &= \gamma\big(\mathbf{E}_{\perp} + \mathbf{V}\times\mathbf{B}\big), \\ \mathbf{B}'_{\parallel} &= \mathbf{B}_{\parallel}, & \mathbf{B}'_{\perp} &= \gamma\Big(\mathbf{B}_{\perp} - \frac{1}{c^2}\mathbf{V}\times\mathbf{E}\Big), \end{aligned} \qquad \text{(P3.11.3.1)}$$

where $\mathbf{V}$ is the velocity of inertial frame K' relative to inertial frame K; the transformations are written for the field components parallel and perpendicular to $\mathbf{V}$. The scalar quantities $\mathbf{E}\cdot\mathbf{B}$ and $E^2 - c^2B^2$ are invariant under these transformations:

$$\mathbf{E}\cdot\mathbf{B} = \mathbf{E}'\cdot\mathbf{B}', \qquad E^2 - c^2B^2 = (E')^2 - c^2(B')^2. \qquad \text{(P3.11.3.2)}$$

For $V \ll c$, formulas (P3.11.3.1) simplify to

$$\mathbf{E}' = \mathbf{E} + \mathbf{V}\times\mathbf{B}, \qquad \mathbf{B}' = \mathbf{B} - \frac{1}{c^2}\mathbf{V}\times\mathbf{E}. \qquad \text{(P3.11.3.3)}$$

Example 1. Magnetic field of a nonrelativistic particle. Consider a particle with charge q moving relative to inertial frame K with a constant nonrelativistic velocity $\mathbf{v}$. In the particle associated inertial frame K', there is only the electric field $\mathbf{E}' = kq\mathbf{r}'/r'^3$. To find the fields in frame K, one has to write out the transformation formulas (P3.11.3.3) with $\mathbf{V} = -\mathbf{v}$. Since there is no length contraction in the nonrelativistic limit, we find that

$$\mathbf{E} = \mathbf{E}' = \frac{1}{4\pi\varepsilon_0}\frac{q\mathbf{r}}{r^3}, \quad \mathbf{B} = \frac{\mathbf{v}\times\mathbf{E}'}{c^2} = \frac{\mu_0 q}{4\pi}\frac{\mathbf{v}\times\mathbf{r}}{r^3}$$

for the instant when the particle passes through the origin of coordinates in frame K. In deriving these formulas, we have used the relation $1/c^2 = \mu_0\varepsilon_0$.

Example 2. Polarization of a dielectric moving in a magnetic field. When a dielectric moves through a magnetic field with an nonrelativistic velocity perpendicular to the magnetic field lines, it becomes polarized. In the dielectric-associated inertial frame, there is transverse electric field: $\mathbf{E}' = \mathbf{v}\times\mathbf{B}$. The detailed polarization picture depends on the shape of the dielectric.

Example 3. Electric field of a relativistic particle. Suppose a particle with charge q moves relative to inertial frame K with a constant relativistic velocity $\mathbf{v}$. In the particle associated inertial frame K', there is only the electric field $\mathbf{E}' = kq\mathbf{r}'/r'^3$. To find the fields in frame K, one has to write out the transformation formulas (P3.11.3.1) with $\mathbf{V} = -\mathbf{v}$. Let us write out the formulas for the instant when the particle passes through the origin of coordinates in frame K. For simplicity, consider the plane (x, y). When changing from the coordinates (x', y') to (x, y), we must take into account that $x' = \gamma x$ and $y' = y$ (the coordinates of a point in frame K are measured at the same instant when the particle passes through the origin). We get

$$E_x = E'_x = k\frac{q\gamma x}{[(\gamma x)^2 + y^2]^{3/2}}, \qquad E_y = \gamma E'_y = k\frac{q\gamma y}{[(\gamma x)^2 + y^2]^{3/2}}.$$

It is seen that $\mathbf{E}$ is collinear with $\mathbf{r}$. However, the field at a point on the line of motion will be smaller than that at a point which is at the same distance from the charge but on a straight line perpendicular to $\mathbf{v}$. The magnetic field is obtained as

$$\mathbf{B} = \frac{\gamma}{c^2}\mathbf{v}\times\mathbf{E}' = \frac{1}{c^2}\mathbf{v}\times\mathbf{E} = \frac{\mu_0}{4\pi}\frac{\gamma q\,\mathbf{v}\times\mathbf{r}}{[(\gamma x)^2 + y^2]^{3/2}}.$$

Note that the above electric field is not potential.

Bibliography for Chapter P3

Alenitsyn, A. G., Butikov, E. I., and Kondratyev, A. S., *Concise Handbook of Mathematics and Physics*, CRC Press, Boca Raton, Florida, 1998.

Benenson, W., Harris, J. W., Stocker, H., and Lutz, H. (Eds.), *Handbook of Physics*, Springer-Verlag, New York, 2002.

Cottingham, W. N. and Greenwood, D. A., *Electricity and Magnetism*, Cambridge University Press, Cambridge, England, 1991.

Fitzpatrick, R., *Maxwell's Equations and the Principles of Electromagnetism*, Infinity Science Press, Hingham, Massachusetts, 2008.

Fleisch, D., *A Student's Guide to Maxwell's Equations*, Cambridge University Press, Cambridge, England, 2008.

Griffiths, D. J., *Introduction to Electrodynamics 3rd Edition*, Benjamin Cummings, New York, 1999.

Irodov, I. E., *Electromagnetism. Basic Laws* [in Russian], Fizmatlit, Moscow, 2006.

Jefimenko, O. D., *Electricity and Magnetism: An Introduction to the Theory of Electric and Magnetic Fields, 2nd Edition*, Electret Scientific Co., Star City, West Virginia, 1989.

Landau, L. D. and Lifshitz, E. M., *The Classical Theory of Fields, Fourth Edition: Volume 2 (Course of Theoretical Physics Series)*, Butterworth-Heinemann, Oxford, England, 1980.

Landau, L. D., Pitaevskii, L. P., and Lifshitz, E. M., *Electrodynamics of Continuous Media, Second Edition: Volume 8 (Course of Theoretical Physics)*, Butterworth-Heinemann, Oxford, England, 1984.

Melia F., *Electrodynamics (Chicago Lectures in Physics)*, University of Chicago Press, Chicago, Illinois, 2001.

Panofsky, W. K. H. and Phillips, M., *Classical Electricity and Magnetism: Second Edition*, Addison-Wesley, Reading, Massachusetts, 1962.

Pollack, G. and Stump, D., *Electromagnetism*, Addison-Wesley, Reading, Massachusetts, 2001.

Prokhorov, A. M. (Ed.), *Encyclopedia of Physics, Volumes 1–5* [in Russian], Bolshaya Russkaya Encyclopedia, Moscow, 1998.

Prokhorov, A. M. (Ed.), *Encyclopedic Dictionary of Physics* [in Russian], Sovetskaya Encyclopedia, Moscow, 1984 (dic.academic.ru).

Purcell, E. M., *Electricity and Magnetism, Berkeley Physics Course, Vol. II, 2nd Edition*, McGraw-Hill Science/Engineering/Math, New York, 1984.

Saslow, W. M., *Electricity, Magnetism, and Light*, Academic Press, New York, 2002.

Schwartz, M., *Principles of Electrodynamics*, Dover Publications, New York, 1987.

Sivukhin, D. V., *General Physics, Volume 3, Electricity* [in Russian], Fizmatlit, Moscow, 2004.

Sivukhin, D. V. and Yakovlev, I. A., *A Collection of Problems, Volume 3, Electricity and Magnetism* [in Russian], Fizmatlit, Moscow, 2006.

Tamm, I. E., *Basics of Theory of Electricity* [in Russian], Fizmatlit, Moscow, 2003.

Woan, G., *The Cambridge Handbook of Physics Formulas*, Cambridge University Press, Cambridge, England, 2003.

Chapter P4

Oscillations and Waves

P4.1. Oscillations

► **General definitions.** Oscillations are repetitive variations in the state of a system such that the state parameters vary in time according to a periodic or almost periodic law. If oscillations occur without external action due to system deviation from a stable equilibrium, the oscillations are said to be *free* or *natural*. If the oscillations occur under the action of an external periodic force, then they are said to be *forced*. The oscillations are characterized by their *period* T and *frequency* $\nu = 1/T$ (measured in *hertzs*: $1\,\mathrm{Hz} = 1\,\mathrm{s}^{-1}$). The term *vibrations* is often used in a narrower sense to mean mechanical oscillations, but sometimes is used synonymously with oscillations.

P4.1.1. Harmonic Oscillations. Composition of Oscillations

► **Simple harmonic oscillations.** An oscillation of a quantity x is said to be a *simple harmonic oscillation* (or *simple harmonic motion*) if x varies in time t by the law

$$x = A\cos(\omega t + \varphi_0), \qquad \text{(P4.1.1.1)}$$

where A is the *amplitude*, $\varphi = \omega t + \varphi_0$ is the *phase*, φ_0 is the *initial phase*, and $\omega = 2\pi/T$ is the *angular* or *circular frequency* of the oscillation. The first and second time-derivatives of the quantity x,

$$\begin{aligned} \dot{x} &= -A\omega\sin(\omega t + \varphi_0) = A\omega\cos\bigl(\omega t + \varphi_0 + \pi/2\bigr), \\ \ddot{x} &= -A\omega^2\cos(\omega t + \varphi_0) = A\omega^2\cos(\omega t + \varphi_0 + \pi), \end{aligned} \qquad \text{(P4.1.1.2)}$$

oscillate harmonically with the same frequency but with amplitudes ωA and $\omega^2 A$ and with the phase shifts $\pi/2$ and π, respectively.

Example. If the initial values (at $t = 0$) of the quantity x and its derivative, $x(0) = x_0$ and $\dot{x}(0) = v_0$, are known, then the amplitude and the initial phase of the oscillation can be determined. The equations $x_0 = A\cos\varphi_0$ and $v_0 = -\omega A\sin\varphi_0$ allow one to find $A = \sqrt{x_0^2 + (v_0/\omega)^2}$ and $\tan\varphi_0 = -v_0/(\omega x_0)$.

► **Equation of simple harmonic oscillations.** It follows from (P4.1.1.2) that if x varies by the harmonic law (P4.1.1.1), then x satisfies the *equation of simple harmonic oscillations*:

$$\ddot{x} + \omega^2 x = 0. \qquad \text{(P4.1.1.3)}$$

The converse is also true: if the *equation of motion* of a physical system whose state is determined by a single quantity x can be reduced under certain conditions (usually, for small values of x) to the differential equation $\ddot{x} + \gamma x = 0$, where γ is a positive constant, then x varies according to the law (P4.1.1.1) with $\omega = \sqrt{\gamma}$ (the parameters A and φ_0 are determined by the initial conditions; see the example above).

► **Energy method for determining the frequency.** If the state of a physical system undergoes repetitive variations and the quadratic function

$$E = \tfrac{1}{2}\mu\dot{x}^2 + \tfrac{1}{2}\kappa x^2 \qquad (\mu \text{ and } \kappa \text{ are positive constants}) \tag{P4.1.1.4}$$

remains constant during these variations, then x changes according to the law (P4.1.1.1) with $\omega = \sqrt{\kappa/\mu}$. Indeed, differentiating (P4.1.1.4) with respect to time, we obtain the equation $\mu\ddot{x} + \kappa x = 0$, which is the equation of simple harmonic oscillations. Usually, the quantity E is proportional to the energy of the oscillatory system for small x; accordingly, such an approach is called the *energy method* for determining the oscillation frequency.

► **Complex exponential and vector diagram.** Using Euler's formula $e^{i\varphi} = \cos\varphi + i\sin\varphi$, one can treat the law of harmonic oscillations (P4.1.1.1) as the real part of a complex exponential function, $x(t) = \operatorname{Re}\widetilde{x}$, where

$$\widetilde{x} = \widetilde{A}\exp(i\omega t), \tag{P4.1.1.5}$$

and $\widetilde{A} = A\exp(i\varphi_0)$ is called the *complex amplitude* of oscillations. This technique is especially convenient for studying systems described by linear equations, since, in this case, the real and imaginary parts are transformed independently of each other.

The harmonic law (P4.1.1.1) can be obtained as the projection on the x-axis of a radius vector of length A that uniformly rotates counterclockwise with a constant angular velocity ω from an initial angular position φ_0; in this case, the angle made with the axis x varies by the law $\varphi = \varphi_0 + \omega t$. This technique is called the *method of vector diagrams*; it is especially convenient for the *composition of harmonic oscillations*. With this technique, the addition of functions is replaced by the graphical addition of vectors; note that the projection of the sum of vectors is equal to the sum of their individual projections.

► **Composition of harmonic oscillations of a single direction.** The sum of two harmonic oscillations of the same frequency with amplitudes A_1 and A_2 and the initial phases φ_1 and φ_2 is a harmonic oscillation of the same frequency, whose amplitude and initial phase can be found by the method of vector diagrams (Fig. P4.1):

$$A^2 = A_1^2 + A_2^2 + 2A_1A_2\cos(\varphi_2 - \varphi_1), \qquad \tan\varphi_0 = \frac{A_1\sin\varphi_1 + A_2\sin\varphi_2}{A_1\cos\varphi_1 + A_2\cos\varphi_2}.$$

The parallelogram spanned by the vectors rotates with angular velocity ω as a rigid body. The phase difference of oscillations with the same frequency does not vary in time; such oscillations are said to be *coherent*. If $\varphi_2 - \varphi_1 = \pm 2m\pi$, the amplitude is a maximum, $A = A_1 + A_2$, and if $\varphi_2 - \varphi_1 = \pm(2m+1)\pi$, the amplitude is a minimum, $A = |A_1 - A_2|$.

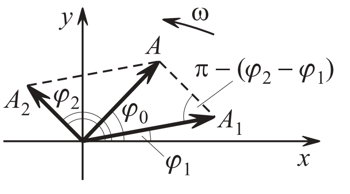

Figure P4.1. Composition of oscillations by the method of vector diagrams.

In the vector addition of incoherent oscillations with different frequencies, the parallelogram spanned by the vectors deforms with time and the magnitude of the resultant vector and its angular velocity vary, and hence the composite motion is not a harmonic oscillation. However, oscillations *with close frequencies* ($\Delta\omega = |\omega_1 - \omega_2| \ll \max\{\omega_1, \omega_2\}$) can be assumed to be approximately coherent on a sufficiently short time interval as compared to the *time of coherence* $\tau_{\rm coh} = 2\pi/\Delta\omega$. The composite oscillation will occur with the angular frequency $\omega = (\omega_1 + \omega_2)/2$ and an amplitude periodically varying between $A_1 + A_2$ and $|A_1 - A_2|$ with period $2\pi/\Delta\omega$. Such oscillations are called *beats*, $\omega_{\rm b} = \Delta\omega$ is the beat angular frequency, and the period of variation in the amplitude $T_{\rm b} = 2\pi/\omega_{\rm b}$ is the beat period. If $A_1 = A_2 = A$, then

$$x(t) = 2A\cos\Bigl(\frac{\omega_2 - \omega_1}{2}t\Bigr)\cos\Bigl(\frac{\omega_2 + \omega_1}{2}t + \varphi_0\Bigr).$$

The beats are an example of *modulated oscillations*, or oscillations obeying the harmonic law (P4.1.1.1) with one of the parameters (amplitude, frequency, or phase) varying periodically in time with a period significantly greater than that of the base oscillations. Amplitude, frequency, and phase modulation techniques are distinguished.

An arbitrary periodic oscillation with period T can be expanded into a *Fourier series* in simple harmonic oscillations with frequencies $\omega_n = n(2\pi/T)$. Such a representation is called the *harmonic analysis* of the periodic oscillation. The Fourier series terms are called the first (fundamental), second, third, etc. *harmonics* of the periodic oscillation. Aperiodic oscillations have, as a rule, a continuous spectrum of frequencies and can be represented as a *Fourier integral* in harmonic oscillations of all frequencies ranging from zero to infinity. All periodic oscillations have a line (discrete) spectrum of frequencies, but oscillations with a line spectrum are not necessarily periodic.

► **Composition of mutually perpendicular harmonic oscillations.** If a point moves in the plane so that its projections on the x- and y-axes oscillate harmonically, this motion is said to be the composition of mutually perpendicular simple harmonic oscillations. If the oscillations in either coordinate have *equal frequencies*, $x = A_1\cos(\omega t + \varphi_1)$ and $y = A_2\cos(\omega t + \varphi_2)$, the path of the point is an inclined ellipse:

$$\frac{x^2}{A_1^2} + \frac{y^2}{A_2^2} - \frac{2xy}{A_1A_2}\cos(\varphi_2 - \varphi_1) = \sin^2(\varphi_2 - \varphi_1).$$

Such a motion is called an *elliptically polarized oscillation*. If $\varphi_2 - \varphi_1 = (2m + 1)\pi/2$, the ellipse axes coincide with the coordinate axes. If $\varphi_2 - \varphi_1 = m\pi$, the ellipse degenerates into a straight line segment; such a motion is called a *linearly polarized oscillation*. The point can move clockwise (for $2m\pi < \varphi_2 - \varphi_1 < 2m\pi + \pi$) or counterclockwise; in these cases, one speaks about the right or left elliptic polarization.

Two-dimensional harmonic motions with frequencies ω_1 and ω_2 occur along curves called *Lissajous figures*. If the frequency ratio ω_1/ω_2 is a rational number, these figures are closed curves. Closed Lissajous figures can be used to determine the frequency ratio.

P4.1.2. Free Undamped Oscillations

► **Free mechanical oscillations with a single degree of freedom.** If the displacement of a conservative mechanical system from a stable equilibrium is described by a single parameter x, then for small x, the potential energy has the form

$$E_{\rm p}(x) = \frac{1}{2}x^2\left(\frac{d^2E_{\rm p}}{dx^2}\right)_{x=0} + \frac{1}{6}x^3\left(\frac{d^3E_{\rm p}}{dx^3}\right)_{x=0} + \cdots \approx \frac{\kappa x^2}{2}. \qquad \text{(P4.1.2.1)}$$

It follows that small oscillations of the system about its stable equilibrium occur harmonically. The potential energy of an oscillatory system is usually reckoned from the stable equilibrium ($E_\text{p} = 0$ for $x = 0$). For small x, the *generalized force*, $F_x = -dE_\text{p}/dx$, is proportional to the displacement: $F_x = -\kappa x$. (If x is a linear coordinate, then the generalized force is the force projection on the displacement direction, and if x is an angle, then F_x is a torque.)

If a mechanical system is a point particle or a translationally moving body, then the kinetic energy of the system has the form $E_\text{k} = \frac{1}{2}m\dot{x}^2$. The equation of motion

$$m\ddot{x} = -\kappa x \qquad \text{(P4.1.2.2)}$$

has the same form as that of a mass attached to a spring. For small x, the *restoring force* acting on the body (i.e., the force that tends to bring the body back to the equilibrium) is called the *quasielastic force*, $F_x = -\kappa x$, and the coefficient κ is the *effective spring constant*. The angular frequency and oscillation period are expressed as

$$\omega = \sqrt{\frac{\kappa}{m}}, \qquad T = 2\pi\sqrt{\frac{m}{\kappa}}. \qquad \text{(P4.1.2.3)}$$

The particle oscillates harmonically, $x = A\cos(\omega t + \varphi_0)$, and its kinetic and potential energies oscillate harmonically with the frequency 2ω:

$$E_\text{k} = E\sin^2(\omega t + \varphi_0) = \tfrac{1}{2}E\bigl[1 - \cos(2\omega t + 2\varphi_0)\bigr],$$
$$E_\text{p} = E\cos^2(\omega t + \varphi_0) = \tfrac{1}{2}E\bigl[1 + \cos(2\omega t + 2\varphi_0)\bigr],$$

where $E = \frac{1}{2}m\omega^2 A^2 = \frac{1}{2}\kappa A^2$ is the total mechanical energy, which is proportional to the amplitude squared. The average kinetic energy is equal to the average potential energy.

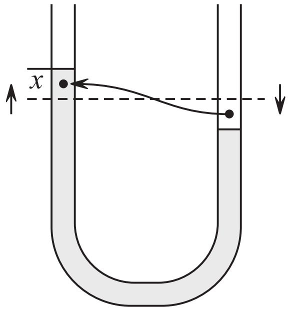

Figure P4.2. To example 1.

Example 1. If the level of a liquid in an U-shaped tube (Fig. P4.2) is displaced by a distance x from equilibrium, the potential energy of the liquid increases by $\Delta mgx = (\rho Sx)gx = (2\rho gS)x^2/2$ (where ρ is the liquid density and S is the tube cross-section) and the kinetic energy is equal to $m\dot{x}^2/2$. Hence the angular frequency of oscillations is equal to $\omega = \sqrt{2\rho gS/m}$.

Example 2. In one of the early models of the hydrogen atom (the Thomson model), it was assumed that the positive charge is uniformly distributed across a ball of radius R, equal to the atomic radius, and the electron was placed at the center of the ball. Let us find the frequency of electron oscillations when it is displaced from the equilibrium. If r denotes the displacement, the attraction force is equal to $F_r = -ke^2 r/R^3$ (see Subsection P3.3.2). Hence, for small displacements, $r < R$, there occur harmonic oscillations with frequency $\omega = \sqrt{ke^2/(mR^3)}$.

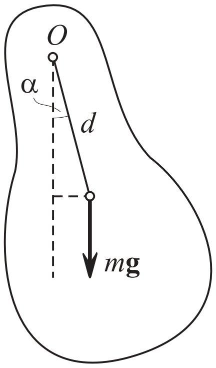

Figure P4.3. Physical pendulum.

Example 3. Physical pendulum. The *physical pendulum* is a rigid body whose horizontal immovable axis of rotation does not pass through its center of mass. For the pendulum deviation from equilibrium by a small angle α (Fig. P4.3), there arises a moment of the gravity force equal to $M = -mgd \sin\alpha$ (d is the distance from the center of mass to the axis of rotation), which tends to return the pendulum to equilibrium. The equation of rotational motion of the rigid body has the form: $I\ddot{\alpha} = -mgd \sin\alpha$, where I is the moment of inertia of the body about the axis of rotation. For small deviation angles $\alpha \ll 1$, this equation becomes the equation of harmonic oscillations with respect to the angle α: $\ddot{\alpha} + (mgd/I)\alpha = 0$. The pendulum angular frequency and oscillation period are equal to

$$\omega = \sqrt{mgd/I}, \qquad T = 2\pi\sqrt{I/(mgd)}.$$

For example, for a homogeneous rod of length l suspended by one of its ends, we obtain $I = ml^2/3$, $d = l/2$, $T = 2\pi\sqrt{2l/(3g)}$. The *reduced length* L of a physical pendulum is the length of a mathematical pendulum such that its period $T = 2\pi\sqrt{L/g}$ is equal to the period of the given physical pendulum: $L = I/(md)$. (For a homogeneous rod suspended by an end, we have $L = 2l/3$.) The point lying at a distance L from the suspension axis on the straight line passing through the center of mass perpendicularly to the suspension axis is called the *center of oscillation*. If a body is suspended on the axis passing through the center of oscillation parallel to the previous axis, then the oscillation period, and hence the pendulum reduced length do not change.

In this example, the generalized force is the moment of gravity force. The frequency can also be determined by the energy method if the dependence of the potential energy on the deviation angle is found: $E_\text{p} = mgd(1-\cos\alpha) \approx mgd\alpha^2/2$. For the total energy, we obtain $E = \frac{1}{2}I\dot{\alpha}^2 + \frac{1}{2}mgd\alpha^2$, which implies the oscillation frequency.

► **Isochronism of harmonic oscillations.** The harmonic oscillations are *isochronous*, which means that their period is independent of the amplitude. The point is that, in the equation of harmonic oscillations $\ddot{x} + \gamma x = 0$, the coefficient γ has dimension s^{-2} and it follows from the dimension considerations that the period cannot depend on anything else except for γ. The oscillations of large amplitude cease to be harmonic and hence isochronous; the period begins to depend on the amplitude. But in several special cases, even small oscillations may not be isochronous. This occurs if the second derivative in formula (P4.1.2.1) for the expansion of E_p in x is zero. Such a situation arises, for example, in the case where a load fixed at the middle of an *unstrained* spring with fixed ends deviates in the direction perpendicular to the spring. For small x, the potential energy is proportional to x^4, and the equation of motion has the form: $\ddot{x} + \gamma x^3 = 0$. In this case, the dimension γ is $\text{m}^{-2}\text{s}^{-2}$, and the oscillation period $T \sim \gamma^{-1/2}A^{-1}$ is inversely proportional to the amplitude.

► **Oscillatory systems with several degrees of freedom.** If the system displacement from the stable equilibrium is characterized by N independent parameters $x_1, \ldots, x_N$, then for an arbitrarily small initial deviation, the dependence $x_i(t)$ cannot be harmonic oscillations. But any motion can be represented as the sum of N special motions called *normal vibrational modes*, in each of which, all x_i harmonically oscillate with a angular

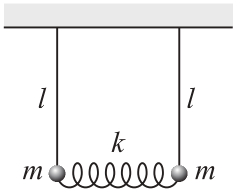

Figure P4.4. Coupled pendulums.

frequency characterizing this mode. In the general case, all the N frequencies can be distinct.

Example 4. Coupled pendulums. We consider two equal mathematical pendulums of length l whose loads are connected by an unstrained spring of rigidity k (Fig. P4.4). The normal modes of this system can be guessed because of the system symmetry. Harmonic oscillations arise for: (a) equal deviations of the pendulums to the same side ($x_1 = x_2$) and (b) for the equal deviations of the pendulums to opposite sides ($x_1 = -x_2$). In the first case: $\omega_1 = \sqrt{g/l}$, in the second case: $\omega_2 = \sqrt{(g/l) + (2k/m)}$. If only the first pendulum deviates to a side, which corresponds to the initial conditions $(x_{10}, x_{20}) = (A, 0)$, then such a deviation can be represented as the sum of two normal deviations: $(A, 0) = \left(\frac{1}{2}A, \frac{1}{2}A\right) + \left(\frac{1}{2}A, -\frac{1}{2}A\right)$. This means that the motion of the first pendulum obeys the law: $x_1 = \frac{1}{2}A\cos(\omega_1 t) + \frac{1}{2}A\cos(\omega_2 t)$, and the motion of the second pendulum obeys the law: $x_2 = \frac{1}{2}A\cos(\omega_1 t) - \frac{1}{2}A\cos(\omega_2 t)$. If $k/m \ll g/l$, then the frequencies ω_1 and ω_2 are close to each other: $\Delta\omega \approx \omega_1 \dfrac{k/m}{g/l}$, and, under superposition of two oscillations with close frequencies, one observes beats (see Subsection P4.1.1): in the time $T_\mathrm{b}/2 = \pi/\Delta\omega$, the second pendulum is in full swing, and the first pendulum stops, and after the same time period, the energy is again transferred to the first pendulum, etc.

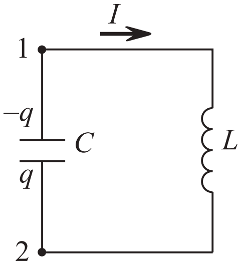

Figure P4.5. Electric oscillatory circuit.

► **Free oscillations in electric circuit.** The oscillatory circuit is a closed circuit consisting of a capacitor of capacitance C and a coil of inductance L (Fig. P4.5). The processes in an oscillatory circuit, just as in alternating current circuits, are studied in the domains of frequencies, where the *quasi-stationarity* condition is satisfied (see Subsection P3.10.1); that is, $\omega \ll c/l$, where l are typical dimensions of the system (in this approximation, in particular, one can neglect the circuit radiation). For the chosen rule of bypassing the circuit, the Ohm law for the subcircuit containing the inductance and for the capacitor charge has the form: $IR = (\varphi_1 - \varphi_2) + \mathscr{E}$, where R is the resistance of the wires and the coil. Taking the relations $I = \dot{q}$ (that holds the chosen signs), $q = (\varphi_2 - \varphi_1)/R$, and $\mathscr{E} = -L\dot{I}$ into account, we obtain $IR = -(q/C) - L\dot{I}$. This implies that

$$\ddot{q} + \frac{R}{L}\dot{q} + \frac{1}{LC}q = 0. \qquad \text{(P4.1.2.3)}$$

If the circuit resistance is negligibly small (which corresponds to $R \ll \sqrt{L/C}\,$), then the capacitor charge varies according to the harmonic law $q = q_m\cos(\omega t + \varphi_0)$ with angular frequency and period equal to

$$\omega = \frac{1}{\sqrt{LC}} \quad \text{(Thomson's formula)}, \qquad T = 2\pi\sqrt{LC}. \qquad \text{(P4.1.2.4)}$$

The current in the circuit varies according to the harmonic law with amplitude $I_m = \omega q_m$ with the phase lead by $\pi/2$.

Thomson's formula (P4.1.2.4) can be obtained by the energy method starting from the expression for the circuit energy:

$$W = \frac{LI^2}{2} + \frac{q^2}{2C},$$

taking into account that $I = \dot{q}$. The energy is transformed from electric into magnetic and conversely: at the moment of maxima capacitor charge, it is equal to $W = q_m^2/(2C)$, and when the capacitor charge is zero, it is equal to $W = LI_m^2/2$. The relations for the time-dependence of the magnetic and electric energy are the same as for the kinetic and potential energies in the case of mechanical oscillations.

P4.1.3. Damped and Forced Oscillations

► **Damped oscillations.** If the energy is lost in an oscillatory system, then the oscillation amplitude decreases with time. If the losses of energy of mechanical oscillations are determined by the force of viscous friction, which is proportional to the velocity for small velocities, then the equation of motion $m\ddot{x} = -\kappa x - b\dot{x}$ can be reduced to the linear differential equation

$$\ddot{x} + 2\beta\dot{x} + \omega_0^2 x = 0, \qquad \text{(P4.1.3.1)}$$

which is called the *equation of damped oscillations*. Here $\beta = b/(2m)$ is the *damping coefficient*, ω_0 is the angular frequency of natural oscillations without damping (as $\beta \to 0$). It is convenient to seek the solution of this equation in exponential form: $\tilde{x} = \tilde{A}\exp(-\gamma t)$. Substituting this into (P4.1.3.1), we obtain the quadratic equation for γ: $\gamma^2 - 2\beta\gamma + \omega_0^2 = 0$. If $\beta < \omega_0$, then the equation has two complex conjugate solutions: $\gamma_{1,2} = \beta \pm i\sqrt{\omega_0^2 - \beta^2}$, which result in equal answers, and hence one can take any of them. Applying the Euler formula (see Subsection P4.1.1), we obtain

$$x(t) = A_0 e^{-\beta t}\cos(\omega t + \varphi_0), \qquad \omega = \sqrt{\omega_0^2 - \beta^2}, \qquad \text{(P4.1.3.2)}$$

where the constants A_0 and φ_0 are determined by the initial conditions.

The function (P4.1.3.2) takes the zero value in equal time intervals $T = 2\pi/\omega$ (Fig. P4.6), and hence ω and T are conditionally called the frequency and period of damped oscillations. If $\beta \ll \omega_0$, then on each time interval $T \ll \tau \ll \beta^{-1}$, the oscillations can be assumed to be harmonic, and $A(t) = A_0\exp(-\beta t)$ has the meaning of the oscillation amplitude on this interval. Therefore (for any $\beta < \omega_0$), $A(t)$ is said to be time-dependent *amplitude of damped oscillations*. In time $\tau = \beta^{-1}$, the amplitude decreases by e times. The amplitude decrease per a period, equal to $A(t)/A(t+T) = \exp(\beta T)$, is called the *damping decrement*, and the

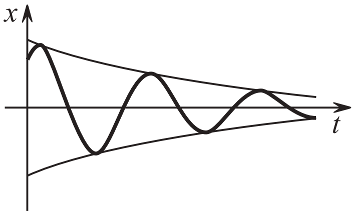

Figure P4.6. Damped oscillations.

logarithm of the decrement

$$\lambda = \ln \frac{A(t)}{A(t+T)} = \beta T \qquad \text{(P4.1.3.3)}$$

is called the *logarithmic damping decrement*. The quantity $1/\lambda$ is the number of total oscillations in which the amplitude decreases by a factor of e.

If $\beta > \omega_0$ (strong damping), then the quadratic equation for γ has two real positive roots: $\gamma_{1,2} = \beta \pm \sqrt{\beta^2 - \omega_0^2}$, and the general solution

$$x(t) = A_1 \exp(-\gamma_1 t) + A_2 \exp(-\gamma_2 t) \qquad \text{(P4.1.3.4)}$$

is of non-oscillatory (anharmonic) character (the parameters A_1 and A_2 are determined by the initial conditions).

► **Damped electric oscillations.** Comparing the equation of oscillations in an electric circuit (P4.1.2.4) with the equation of damped oscillations (P4.1.3.1), we find the expressions for the damping coefficient and the logarithmic damping decrement:

$$\beta = \frac{R}{2L}, \qquad \lambda = \beta T = \pi \frac{R}{\sqrt{L/C}}. \qquad \text{(P4.1.3.5)}$$

The condition for weak damping $\lambda \ll 1$ has the form: $R \ll \sqrt{L/C}$ (the quantity $\sqrt{L/C}$ is called the *wave resistance of the circuit*).

► **Forced mechanical oscillations.** The motion of a system under the action of an external periodic force is called the *forced oscillations*, the external force itself is called the *driving force*. From the equation of motion $m\ddot{x} = -\kappa x - b\dot{x} + F_x(t)$, we obtain the *equation of forced oscillations*:

$$\ddot{x} + 2\beta\dot{x} + \omega_0^2 x = f(t), \qquad \text{(P4.1.3.6)}$$

where the notation $f(t) = F_x(t)/m$ is used.

The general solution of such an *inhomogeneous* (with nonzero right-hand side) equation can be represented as the sum of a particular (that is, any) solution of the inhomogeneous equation and the general solution of the homogeneous equation. (The general solution must contain free parameters, which permit satisfying any initial conditions.) The homogeneous equation is the equation of damped oscillations, its general solution (formulas (P4.1.3.2) and (P4.1.3.4)) exponentially decays in time $\tau \sim 1/\beta$. The damping of natural oscillations means the termination of the regime of transient oscillations and the beginning of the regime of *steady-state* forced oscillations whose characteristics are determined by the functions $f(t)$ and the parameters β, ω_0, but are independent of the initial conditions.

We seek a particular solution of equation (P4.1.3.6) in the form of steady-state oscillations. Because any periodic force $F_x(t)$ can be expanded in the Fourier series, it is natural to study the steady-state forced oscillations under the action of a harmonic driving force $f(t) = F_x(t)/m = (F_0/m)\cos\Omega t$. We seek them in the form of harmonic oscillations of the same frequency but with a phase shift:

$$x = A\cos(\Omega t + \varphi). \qquad \text{(P4.1.3.7)}$$

Substituting (P4.1.3.7) into equation (P4.1.3.6), we obtain

$$\begin{aligned} A &= \frac{F_0}{m\sqrt{(\omega_0^2 - \Omega^2)^2 + 4\beta^2\Omega^2}}, \\ \tan\varphi &= -\frac{2\beta\Omega}{\omega_0^2 - \Omega^2}. \end{aligned} \qquad \text{(P4.1.3.8)}$$

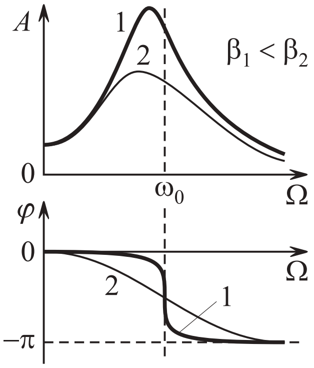

Figure P4.7. The amplitude and phase shift dependence on the frequency of the driving force.

For $\Omega = 0$, we obtain $\varphi = 0$ and $A = F_0/(m\omega_0^2) = F_0/\kappa$, which corresponds to the static displacement of the body following the slowly varying force. For $\Omega \to \infty$, we have $A \to 0$ and $\varphi \to -\pi$. The graphs of $A(\Omega)$ and $\varphi(\Omega)$ are given in Fig. P4.7.

► **Resonance.** The maximum value of the amplitude of steady-state oscillations is attained at the *resonance frequency* $\Omega_\mathrm{r} = \sqrt{\omega_0^2 - 2\beta^2} = \sqrt{\omega^2 - \beta^2}$ and is equal to $A_\mathrm{max} = A(\Omega_\mathrm{r}) = F_0/(2m\beta\omega)$, where $\omega = \sqrt{\omega_0^2 - \beta^2}$ is the angular frequency of damped oscillations. For $\beta \ll \omega_0$, the dependence $A(\Omega)$ contains a sharp and narrow maximum at the resonance frequency, which in this limit is close to the natural frequency of the system oscillations. This phenomenon is called *resonance*, and the curves of the dependence $A(\Omega)$ are the *resonance curves*. The maxima characteristics (for $\beta \ll \omega_0$) are: the ratio of A_max to the static deviation $A(0)$ is equal to $\omega/(2\beta) = \pi/\lambda$ (λ is the logarithmic damping decrement); the quantity $Q = \pi/\lambda$ is called the *quality factor* of the oscillatory system. The width of the resonance curve at the level $A_\mathrm{max}/\sqrt{2}$ is equal to the damping coefficient: $\Delta\Omega \approx \beta$.

The amplitude of steady-state oscillations of the velocity attains the maximum value $F_0/(2m\beta)$ at $\Omega = \omega_0$. At resonance, the velocity oscillations coincide in phase with the driving force oscillations.

We consider the process of establishing the oscillations in the case where the frequency driving force is equal to the resonance frequency (it is assumed that $\beta \ll \omega_0$). If, at the initial moment, the point displacement and velocity are zero, then in the limit under study the initial conditions are satisfied by the solution:

$$x \approx A_\mathrm{max} \sin\omega t - A_\mathrm{max} e^{-\beta t} \sin\omega t = A_\mathrm{max}\bigl(1 - e^{-\beta t}\bigr) \sin\omega t.$$

The obtained dependence is shown in Fig. P4.8. For $t \ll 1/\beta$, the amplitude increases proportionally to time: $A(t) \approx A_\mathrm{max}\beta t = = F_0 t/(2m\omega)$; the damping at this stage does not exhibit any influence. We note that the time of establishment of oscillations is large compared with the period: $1/\beta \gg 2\pi/\omega$. If the frequency Ω is close to ω, but differs from it, then the motion at the initial stage $t \ll 1/\beta$ is the sum of oscillations with close frequencies. If the condition $\omega \gg |\Omega - \omega| \gg 1/\beta$ is satisfied, then in the process of establishment of oscillations, profound beats can be observed (the oscillation amplitude increases almost up to $2A_\mathrm{max}$ and decreases almost to zero with period $T_\mathrm{b} = 2\pi/|\Omega - \omega|$).

Resonance under arbitrary periodic action. If the period of external action $F_x(t)$ is multiple of the period of the system natural oscillations $T = 2\pi/\omega$, then the expansion of $F_x(t)$ in the Fourier series can contain the harmonic with frequency ω. If the Q-factor of an oscillatory system is large, then harmonic oscillations of a noticeable amplitude can arise under the action of this harmonic. For example, the swings can be slightly pushed at each swing (then the first harmonic has the resonance frequency), or at the next nearest swing (then the second harmonic will work), etc.

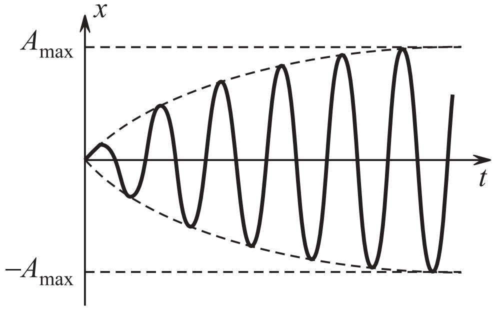

Figure P4.8. Establishment of oscillations at the resonance frequency.

► **Forced electric oscillations.** If the oscillatory circuit (see Subsection P4.2.1) contains alternating emf $\mathscr{E}(t)$, then the forced oscillations are described by the same differential equation as that for a mechanical system:

$$\ddot{q} + 2\beta\dot{q} + \omega_0^2 q = \mathscr{E}(t)/L, \tag{P4.1.3.9}$$

where q is capacitor charge, $\beta = R/(2L)$ is the damping coefficient, and $\omega_0^2 = 1/\sqrt{LC}$ is the angular frequency of the circuit natural oscillations. We consider the steady-state oscillations under the action of a harmonic force $\mathscr{E}(t) = \mathscr{E}_0 \cos\Omega t$. The charge oscillations obey the law $q(t) = q_0\cos(\Omega t + \varphi)$, where the amplitude q_0 and the phase shift φ satisfy expressions (P4.1.3.8). However, in electric circuits, it is of interest to study not the charge oscillations but the oscillations of current $I = \dot{q}$, which occur according to the law $I = I_0\cos(\Omega t - \varphi_1)$ with amplitude and phase shift equal to

$$I_0 = \mathscr{E}_0\left[R^2 + \left(\Omega L - \frac{1}{\Omega C}\right)^2\right]^{-1/2}, \qquad \tan\varphi_1 = \frac{1}{R}\left(\Omega L - \frac{1}{\Omega C}\right). \tag{P4.1.3.10}$$

The maximum of the current oscillation amplitude (resonance) is attained for $\Omega_\mathrm{r} = \omega_0 = 1/\sqrt{LC}$, and it is equal to $I_0(\Omega_\mathrm{r}) = \mathscr{E}_0/R$. At resonance, the current oscillations coincide in phase with emf oscillations.

The voltage oscillations on the resistor, capacitor, and the induction coil are described by the following formulas:

$$\begin{aligned} U_R &= IR = I_0 R\cos\bigl(\Omega t - \varphi_1\bigr),\\ U_C &= \frac{q}{C} = I_0\frac{1}{\Omega C}\cos\Bigl(\Omega t - \varphi_1 - \frac{\pi}{2}\Bigr),\\ U_L &= L\frac{dI}{dt} = I_0(\Omega L)\cos\Bigl(\Omega t - \varphi_1 + \frac{\pi}{2}\Bigr) \end{aligned}$$

(the current amplitude is the same on all the elements, because they are connected in series).

The coefficients of proportionality between the current and voltage amplitudes are called as follows: $x_C = 1/(\Omega C)$ is the *capacitive reactance* and $x_L = \Omega L$ is *inductive reactance*. The quantity $x = x_L - x_C$ is called the *reactance*, R is the *resistance* (the energy is dissipated only at this resistance), and the quantity $z = \sqrt{R^2 + x^2}$ is the *impedance* of the circuit. Formula (P4.1.3.10) can be rewritten as $I_0 = \mathscr{E}_0/z$, $\tan\varphi_1 = x/R$. All these relations become rather obvious if the vector diagrams are used (Fig. P4.9). At resonance, the reactance vanishes, $x = 0$, that is, the voltage oscillations on the capacitance and inductance compensate for each other and the circuit impedance becomes purely *resistive*, $z = R$. If the resistance of a circuit is zero, $R = 0$, the impedance is called purely *reactive*, with $z = x$.

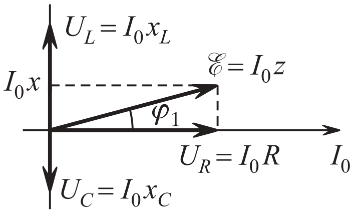

Figure P4.9. Vector diagram of currents and voltages for elements connected in series.

► **Power on a subcircuit.** The *active* or *effective value* of alternating current (voltage, emf) is the value of the constant current at which the quantity of heat released on the resistance in a period coincides with the quantity of heat released by the alternating current (that is, the mean square value over the period):

$$\int_0^T I^2 R\,dt = I_{\text{ef}}^2 RT.$$

For the sinusoidal current, $I_{\text{ef}} = I_0/\sqrt{2}$, and $U_{\text{ef}} = U_0/\sqrt{2}$. If the phase shift between the current and the voltage is φ, then the period-average power is equal to

$$P = \frac{1}{T}\int_0^T U_0 \cos(\Omega t) I_0 \cos(\Omega t - \varphi)\,dt = \frac{1}{2} U_0 I_0 \cos\varphi = I_{\text{ef}} U_{\text{ef}} \cos\varphi.$$

For purely reactive impedance, the average power is zero.

P4.2. Wave Processes

P4.2.1. Elastic and Acoustic Waves

► **Basic definitions.** If a continuum has elastic properties, then the motion of points at one place of the medium (at the *source*) results in propagation of this motion with a certain velocity in the form of an elastic wave. The medium wave motion is described by the function $\boldsymbol{\xi}(\mathbf{r}, t)$ determining the *medium particle* displacement from equilibrium $\mathbf{r}$ at time t. (The medium particle is a sufficiently small *macroscopic* element of the medium, that is, an element containing rather many atoms and molecules.) An elastic wave can propagate in three-dimensional medium, two-dimensional medium (elastic membrane), and one-dimensional medium (elastic rod, stretched rigid fiber, air column). The simplest wave mode is the *plane wave*, where the function $\boldsymbol{\xi} = \boldsymbol{\xi}(x, t)$ varies only in one direction and depends on the coordinate x. If in this case the vector $\boldsymbol{\xi}$ is perpendicular to the wave propagation direction, then the wave is said to be *transverse*, and if it is parallel, then the wave is said to be *longitudinal*.

A medium is said to be *homogeneous* if all its points are equivalent, *isotropic* if all the directions in it are equivalent, *elastic* if all forces arising in it depend only on the displacement (strain), and *linear* if the forces are proportional to the strains. In an absolutely elastic medium, the wave mechanical energy is not dissipated (is not transformed into internal energy). If the medium has only volume elasticity, then only longitudinal waves can propagate in it (liquid, gas); if it also has elasticity of form, then transverse (shear) waves are also possible. In the linear medium, the wave propagation is described by linear differential equations, and the *superposition principle* holds: if two independent sources cause two different waves $\boldsymbol{\xi}_1(\mathbf{r}, t)$ and $\boldsymbol{\xi}_2(\mathbf{r}, t)$, then the common action of the sources results in the wave $\boldsymbol{\xi}(\mathbf{r}, t) = \boldsymbol{\xi}_1(\mathbf{r}, t) + \boldsymbol{\xi}_2(\mathbf{r}, t)$. The same statement holds for a single source whose motion can be decomposed into two motions. Since any function of time can be represented as the Fourier integral in harmonic functions, it follows that the wave motion of a linear medium can be decomposed into *harmonic waves*. Through an arbitrary point of a harmonic wave, one can draw a single surface of constant phase, which is called the *wave surface*. The wave surface coinciding with the leading edge of the wave perturbation is called the *wave front*.

► **Equation of plane harmonic wave.** If the displacement in the source obeys the law $\xi(0,t)=\xi_m\cos\omega t$, then at the points with the coordinate x, the displacement occurs according to the same law, but with the delay by x/v, where v is the wave propagation velocity:

$$\xi(x,t)=\xi_m\cos\Big[\omega\Big(t-\frac{x}{v}\Big)\Big]=\xi_m\cos\Big(\frac{2\pi}{T}t-\frac{2\pi}{\lambda}x\Big)=\xi_m\cos(\omega t-kx),\qquad\text{(P4.2.1.1)}$$

where $\lambda=2\pi v/\omega=vT$ is *wavelength*, and $k=\omega/v=2\pi/\lambda$ is the *wavenumber*. The wavelength is the shortest distance between points exhibiting synchronous oscillations (the phase difference is $\Delta\varphi=2\pi$); the wavelength is the spatial period of the harmonic wave profile at any time moment. The phase difference between arbitrary points of the wave is equal to $\Delta\varphi=2\pi(x_2-x_1)/\lambda$. In equation (P4.2.1.1), it is assumed that the energy losses at the distance x are negligibly small; a small energy dissipation can be taken into account empirically introducing the factor $\exp(-\gamma x)$, where γ is *coefficient of wave damping*, in (P4.2.1.1). If we introduce the *wave vector* **k** directed perpendicularly to the wave front in the direction of wave propagation, then (P4.2.1.1) takes the form invariant under the choice of the coordinate system:

$$\xi(\mathbf{r},t)=\xi_m\cos(\omega t-\mathbf{k}\cdot\mathbf{r}),\qquad\widetilde{\xi}(\mathbf{r},t)=\widetilde{\xi}_m e^{i(\omega t-\mathbf{k}\cdot\mathbf{r})}.\qquad\text{(P4.2.1.2)}$$

The last form of the traveling wave equation is based on the Euler formula; only the real part $\operatorname{Re}\widetilde{\xi}$ has a physical meaning. If the wave velocity is independent of the frequency (there is no dispersion), then the plane wave perturbation of any form propagates without distortion.

► **Spherical and cylindrical diverging waves.** If the wave propagates from a point (spherical) or linear (cylindrical) source isotropically in all directions, then the wave surfaces have the respective shape of spheres or circular cylinders. In these cases, the equations of diverging waves without energy losses have the form:

$$\xi(r,t)=\frac{\xi_0}{r}\cos(\omega t-kr)\qquad\text{(spherical wave)},$$
$$\xi(\mathcal{R},t)=\frac{\xi_0}{\sqrt{\mathcal{R}}}\cos(\omega t-k\mathcal{R})\qquad\text{(cylindrical wave)},$$

where $r=\sqrt{x^2+y^2+z^2}$, $\mathcal{R}=\sqrt{x^2+y^2}$. The amplitude decreases because the wave surface area increases; the energy conservation law demands that the energy flux remain constant (we recall that the energy of oscillations is proportional to the squared amplitude; see also below).

► **Wave equation.** A direct substitution shows that the displacement of the harmonic wave $\xi(\mathbf{r},t)$ satisfies the equation

$$\frac{\partial^2\xi}{\partial x^2}+\frac{\partial^2\xi}{\partial y^2}+\frac{\partial^2\xi}{\partial z^2}=\frac{1}{v^2}\frac{\partial^2\xi}{\partial t^2},\quad\text{or}\quad\Delta\xi=\frac{1}{v^2}\frac{\partial^2\xi}{\partial t^2},\qquad\text{(P4.2.1.3)}$$

which is called the *wave equation*. If the harmonic wave velocity is independent of the frequency (the dependence of ω on k is linear, that is, the dispersion is absent), then any superposition of plane waves satisfies the same equation. The converse statement also holds: if the equation of motion of a particle of the medium can be reduced to (P4.2.1.3), then any wave perturbations of this medium propagate with the velocity v.

In the one-dimensional case, the general solution of equation (P4.2.1.3) has the form

$$\xi = f_1(x - vt) + f_2(x + vt), \tag{P4.2.1.4}$$

where f_1 and f_2 are arbitrary (differential) functions. The first term describes propagation of the plane signal $f_1(x)$ with velocity v without changes in the form in the direction of positive x, and the second term describes propagation of the signal $f_2(x)$ with velocity v in the direction of negative x.

Example 1. Elastic rod. The forces F arising in a thin rod of cross-sectional area S subjected to longitudinal elastic deformations obeys Hooke's law: $\sigma = E\varepsilon$, where $\sigma = F/S$ is the *stress* at a given point of the rod, ε is the strain, and E is Young's modulus. If x is the coordinate along the rod and $\xi(x,t)$ is the longitudinal displacement of the particle with coordinate x at the time t, the strain at the point x is expressed in terms of the partial derivative of the displacement as $\varepsilon = \dfrac{\xi(x+dx) - \xi(x)}{dx} \approx \dfrac{\partial \xi}{\partial x}$. The force acting on the particle of size Δx and mass $\Delta m = \rho\,\Delta x\,S$ is equal to $[\sigma(x+\Delta x) - \sigma(x)]S$. Then the equation of motion of the particle has the form

$$[\sigma(x+\Delta x) - \sigma(x)]S = \Delta m \frac{\partial^2 \xi}{\partial t^2} \quad\Longrightarrow\quad \frac{\partial^2 \xi}{\partial x^2} = \frac{\rho}{E}\frac{\partial^2 \xi}{\partial t^2}.$$

Hence, the wave propagation speed in an elastic rod is equal to $v = \sqrt{E/\rho}$. Longitudinal waves in elastic media propagate with the same speed.

Example 2. Stretched string. A transverse wave can propagate along a string whose tension is equal to T. Suppose that the string profile at a given time moment is described by the function $\xi(x,t)$ (Fig. P4.10). We obtain the equation of motion of the medium particle of dimension Δx under the assumption of low-slope wave: $\partial\xi/\partial x \ll 1$. The vertical projection of the tensile force is equal to $F_\perp(x) = T\sin\alpha \approx T(\partial\xi/\partial x)$ (α is the string inclination angle); the vertical force acting on the medium particle of dimension Δx and mass $\Delta m = \rho_{\rm lin}\Delta x\,S$ is equal to $F_\perp(x+\Delta x) - F_\perp(x)$ ($\rho_{\rm lin}$ is the linear string density). The equation of motion of the medium particle has the form

$$\frac{\partial^2 \xi}{\partial x^2} = \frac{\rho_{\rm lin}}{T}\frac{\partial^2 \xi}{\partial t^2},$$

hence the wave propagation velocity in the string is equal to $v = \sqrt{T/\rho_{\rm lin}}$.

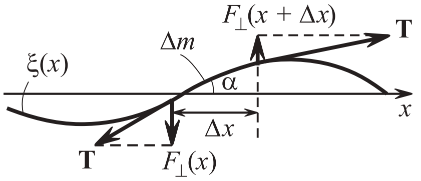

Figure P4.10. Element of a stretched string (to example 2).

Example 3. Air column in a cylinder. The volume elasticity of a gas (or any thermodynamically simple medium) is characterized by the bulk modulus of elasticity $K = -V(\partial p/\partial V) = \rho(\partial p/\partial\rho)$. The value of K depends on the condition under which bulk deformation of the gas occurs. For sound oscillations ($\nu > 16$ Hz), one can neglect the heat exchange between distinct medium particles and assume that K is the adiabatic modulus of elasticity: $K_{\rm ad} = -V\left(\partial p/\partial V\right)_{\rm ad}$. The variation in the pressure at a given point of the cylinder (for small strains) is $\Delta p(x) = -\lim_{V\to 0} K(\Delta V/V) = -\lim_{\Delta x\to 0} K[S\xi(x+\Delta x) - S\xi(x)]/(S\Delta x) = -K(\partial\xi/\partial x)$; the force acting on the air medium particle of dimension Δx and mass $\Delta m = \rho\,\Delta x\,S$ is equal to $\Delta F = [\Delta p(x) - \Delta p(x+\Delta x)]S$. The equation of motion of the medium particle has the form

$$\frac{\partial^2 \xi}{\partial x^2} = \frac{\rho}{K}\frac{\partial^2 \xi}{\partial t^2},$$

and hence the wave propagation velocity in the air cylinder is equal to $v = \sqrt{K/\rho} = \sqrt{\partial p/\partial \rho}$. (These expressions remain valid for any thermodynamically simple medium such as a real gas or a liquid.) For the sound waves in an ideal gas, we obtain: $K = \gamma p$, $v = \sqrt{\gamma p/\rho} = \sqrt{\gamma RT/\mu}$, where γ is the adiabatic exponent (see Subsection P2.2.3).

► **Volume density of energy. Energy flux density.** The densities of the kinetic and potential energies at a given point of the plane elastic longitudinal wave are equal to

$$w_{\text{k}} = \frac{\rho}{2}\left(\frac{\partial \xi}{\partial t}\right)^2, \qquad w_{\text{p}} = \frac{E}{2}\left(\frac{\partial \xi}{\partial x}\right)^2.$$

For a traveling wave of arbitrary form $\xi = f(x \pm vt)$, we have the relation $(\partial\xi/\partial t)^2 = v^2(\partial\xi/\partial x)^2$. Taking into account that $v^2 = E/\rho$ for the elastic longitudinal wave, we see that the densities of the kinetic and potential energies at each point of the traveling wave are equal to each other. This statement holds for traveling waves of any nature. In the case of a harmonic plane wave (P4.2.1.1), the wave energy per unit volume and its time average have the form:

$$w = w_{\text{k}} + w_{\text{p}} = \rho\omega^2\xi_m^2 \sin^2(\omega t - kx), \qquad \langle w\rangle = \tfrac{1}{2}\rho\omega^2\xi_m^2.$$

The energy flux density (the *Umov vector*) at a given point of the traveling wave is equal to

$$\mathbf{U} = w\mathbf{u}, \tag{P4.2.1.5}$$

where **u** is the rate of the energy transfer. In the dispersionless medium, **u** = **v**, where **v** is the wave propagation velocity contained in the wave equation (P4.2.1.4). For harmonic traveling wave (P4.2.1.1), $u = v$, where v is the phase velocity of the wave: $v = \omega/k$. The energy transferred by the traveling wave through the site $d\mathbf{s}$ per unit time (the energy flux) is equal to $d\Phi = \mathbf{U}\cdot d\mathbf{s}$. The magnitude of time-average value of the Umov vector (it is sometimes called the *wave intensity*) for a harmonic plane wave is equal to $I = |\langle\mathbf{U}\rangle| = v\langle w\rangle = \frac{1}{2}\rho v\omega^2\xi_m^2$. In the case of an arbitrary (including the standing) longitudinal elastic wave, one can use the expression for **U** in terms of the power of elastic forces (per unit area):

$$\mathbf{U} = -\sigma\frac{\partial \boldsymbol{\xi}}{\partial t},$$

where $\sigma = E\,\partial\xi/\partial x$ is the stress at a given point of the medium (see example 1). For the transverse wave in the string (example 2), the energy flux has a similar form:

$$P = -F_\perp\frac{\partial\xi}{\partial t} = -T\frac{\partial\xi}{\partial x}\frac{\partial\xi}{\partial t}.$$

► **Dispersion. Group velocity.** The dependence of the harmonic wave velocity on the frequency or the wavelength is called the *dispersion*, and the medium where this phenomenon is observed is called the *dispersion medium*. In the dispersion medium, the nonsinusoidal traveling pulse changes its shape in the process of motion, because the harmonic waves contained in its Fourier expansion move with different velocities. The energy transfer rate contained in expression (P4.2.1.5) for the Umov vector is determined in this case as the velocity of motion of the center of pulse (or as the velocity of the point at which the strain is maximum) and is called the *group velocity* of the wave packet (in contrast to the *phase velocity* of the sinusoidal wave $v = \omega/k$). A finite pulse of dimension Δx is called the *wave packet* because its expansion in the Fourier integral in harmonic waves contains a finite

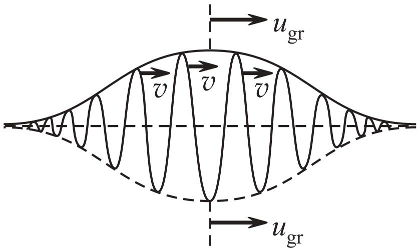

Figure P4.11. Phase and group velocities.

group of waves whose spectral width is determined by the relation: $\Delta x\,\Delta k \sim 1$. The group velocity of such a pulse is equal to

$$u_{\text{gr}} = \frac{d\omega}{dk} = v - \lambda\frac{dv}{d\lambda}, \tag{P4.2.1.6}$$

where $v = \omega/k$ is the phase velocity of the wave. The points of maxima (the "humps") inside the wave packet move with the phase velocity v, and the envelope of these maxima moves with the velocity u_{gr} (Fig. P4.11).

The notion of group velocity of a wave packet can be used only if its spectral width is small so that the dependence $\omega(k)$ can be assumed to be linear.

To illustrate this, we consider the wave that is the superposition of at most two traveling waves with close frequencies and equal amplitude:

$$\begin{aligned}\xi &= \xi_m\cos[(\omega+\Delta\omega)t-(k+\Delta k)x]+\xi_m\cos[(\omega-\Delta\omega)t-(k-\Delta k)x]\\ &= 2\xi_m\cos(\Delta\omega\,t-\Delta k\,x)\cos(\omega t-kx).\end{aligned}$$

This shows that the wave envelope has the spatial period $\Delta x = 2\pi/\Delta k$ and propagates with velocity $u = \Delta\omega/\Delta k$.

Example 4. The group velocity can be both less and greater than the phase velocity. The phase velocity of waves on the surface of a fluid for long (gravitational) waves depends on the wavelength according to the law $v = \sqrt{g\lambda/2\pi}$, and for short (capillary) waves, according to the law $v = \sqrt{2\pi\sigma/\rho\lambda}$. Substituting this in formula (P4.2.1.6), we see that the group velocity in the first case is equal to $u_{\text{gr}} = v/2$, and in the second, to $u_{\text{gr}} = 3v/2$. We also see that the group velocity can be strongly different from the phase velocity.

► **Wave interference.** Two waves are called *coherent* if, at any point of the space, they generate coherent oscillations whose phase difference does not vary in time (see Subsection P4.1.1). The sources of coherent waves are called *coherent sources*. In the case of addition of incoherent waves, the time-average energy of the resultant oscillation is equal to the sum of their average energies. In the case of addition of coherent waves and in the case of particle oscillations that occur in the same or close directions, the phenomenon of *interference* can be observed. Interference of two or more coherent waves is characterized by increased amplitudes of the resulting wave at some points of space (where wave crests meet) and decreased amplitudes at other points (where a crest of a wave meets a trough of another wave). If a point lies at the distance r_1 from one of the coherent sources and at the distance r_2 from the other, then the phase difference between the oscillations at this point is equal to

$$\Delta\varphi = (\varphi_{10}-kr_1)-(\varphi_{20}-kr_2) = 2\pi\frac{r_2-r_1}{\lambda}-(\varphi_{20}-\varphi_{10}) = 2\pi\frac{\Delta}{\lambda}-\Delta\varphi_0.$$

Here $\Delta\varphi_0$ is the phase difference between the source oscillations, $\Delta = (r_2 - r_1)$ is the *wave path difference*. The maximum conditions for the resultant oscillations is $\Delta\varphi = \pm 2\pi m$, the minimum condition is $\Delta\varphi = \pm 2\pi\left(m-\frac{1}{2}\right)$, where m is the order of the interference maximum (minimum). In the case of in-phase sources ($\Delta\varphi_0 = 0$) these conditions take the following especially simple form:

$$\begin{aligned} &\Delta = \pm m\lambda && \text{(maximum)},\\ &\Delta = \pm\bigl(m-\tfrac{1}{2}\bigr)\lambda && \text{(minimum)}. \end{aligned}$$

Example 5. We consider two linear coherent sources of cylindrical waves that are at the distance $\lambda/2$ from each other. For the points at far distances from sources, the wave propagation difference is determined only by the direction of radiation and is equal to $\Delta = \frac{1}{2}\lambda\cos\theta$ (Fig. P4.12). In the case of in-phase sources, the radiation is weakened in the direction $\theta = 0$ and reinforced in the direction $\theta = \pi/2$, and in the case of sources oscillating in opposite phase ($\Delta\varphi_0 = \pi$), the situation is opposite.

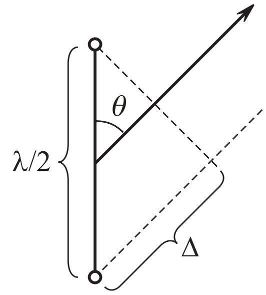

Figure P4.12. To example 5.

► **Standing waves. Natural oscillations.** An important example of interference is the *standing wave*, arising in addition of two equal plane traveling waves propagating towards each other:

$$\xi = \xi_m\cos(\omega t - kx) + \xi_m\cos(\omega t + kx) = 2\xi_m\cos kx\cos\omega t. \qquad \text{(P4.2.1.7)}$$

The oscillation amplitude $A(x) = 2\xi_m\cos kx$ of the wave points varies periodically from zero (at the wave *nodes*) to $2\xi_m$ (at the wave *antinodes*). The distance between neighboring nodes is equal to $\lambda/2$. The oscillations of points between two nodes occur in phase, but the oscillations of points on the opposite sides of the node occur in opposite phase. At the nodes, the velocity oscillation amplitude is zero, but the oscillation amplitude of the strain $\partial\xi/\partial x$ takes the maximum value. Conversely, at the antinodes, the strains are zero, but the velocity oscillations have the maximum amplitude. The time-average energy has the same value at all the points of the standing wave. The Umov vector is zero at both the nodes and antinodes and periodically changes its direction at the intermediate points.

If the end of the rod in oscillatory motion is rigidly fixed, then the displacement of the rod end points is zero; if the rod end is free, then the strain is zero. In the first case, the boundary condition has the form $\xi|_{x=0} = 0$, and in the second case, the form $\partial\xi/\partial x|_{x=0} = 0$, where the value $x = 0$ corresponds to the rod end. (The same conditions arise for the air oscillations in a cylindrical tube in the case of closed or open end of the tube. For transverse oscillations of the stretched string considered in example 2, only the first version in which the string end is fixed is possible.) If all the points of a finite rod oscillate harmonically with a single frequency, then such a motion is called a *natural* or *free oscillation* of the rod. If an infinitely long rod is mentally cut at the point where the standing wave node is located, then the boundary condition for the fixed end is satisfied; and if the cut is at the point of antinode, then the boundary condition for the free end is satisfied. Hence the free oscillations of a rod of length l in the case of two free or two fixed ends must satisfy the condition $l = \frac{1}{2}m\lambda$ or $\omega_m = \pi mv/l$ (the spectrum of natural oscillations, $m = 1, 2, \ldots$),

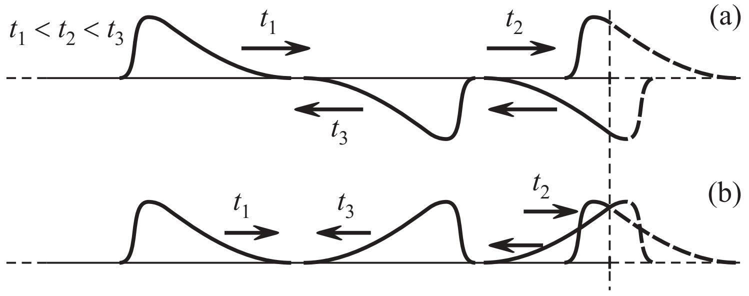

Figure P4.13. Reflection of a wave pulse from an end of a rod: (a) fixed end, (b) free end.

and in the case of one fixed end, the condition $l = \frac{1}{2}\left(m - \frac{1}{2}\right)\lambda$ or $\omega_m = \pi\left(m - \frac{1}{2}\right)v/l$. The natural oscillation with the least possible frequency is called the *fundamental oscillation*, all the other natural oscillations are *overtones* or *harmonics*.

► **Traveling wave reflection.** We consider reflection of traveling pulse from the rod end. If the rod end is fixed, then the boundary condition $\xi|_{x=0} = 0$ and the initial conditions (one incident pulse) are satisfied by the superposition of the incident pulse and the pulse of just the same form traveling towards it in which the point displacements have opposite direction (Fig. P4.13a). But if the rod end is free, then the boundary condition $\partial\xi/\partial x|_{x=0} = 0$ is satisfied by the pulse of displacements of the same sign traveling towards it (Fig. P4.13b). In the case of a harmonic traveling wave, the reflection from the free end occurs without changes in the phase, and a standing wave with antinode at the rod end is formed in this case. The reflection from the fixed end occurs with the phase shift by π, and a standing wave with node at the rod end is formed.

► **Doppler effect in acoustics.** In motion of the wave source and (or) the wave receiver with respect to the medium where the wave propagates, the registered frequency differs from the one under study (if the distance between the source and the receiver varies). We direct the axis x from the receiver to the source. If the source radiates sound with frequency ν_0 and moves along the axis x with velocity $v_x^{\rm S}$ (Fig. P4.14), then the distance between the neighboring maxima of the wave propagating along x with velocity v in the direction of the receiver is equal to $\lambda = vT_0 + v_x^{\rm S}T_0$, and its frequency is equal to v/λ:

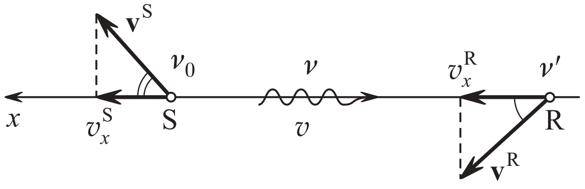

Figure P4.14. Doppler effect arises when the source (S) moves and also when the receiver (R) moves.

$$\nu = \nu_0 \frac{v}{v + v_x^{\rm S}}. \qquad \text{(P4.2.1.8)}$$

If the receiver is fixed, then precisely this frequency is registered. If the receiver moves with velocity $v_x^{\rm R}$, then the time interval between the arrival of the neighboring maxima is equal to $T' = \lambda/(v + v_x^{\rm R})$, and the frequency of the received signal is equal to $1/T'$:

$$\nu' = \nu \frac{v + v_x^{\rm R}}{v} = \nu_0 \frac{v + v_x^{\rm R}}{v + v_x^{\rm S}}. \qquad \text{(P4.2.1.9)}$$

The Doppler effect is observed if the distance between the source and the receiver varies.

P4.2.2. Electromagnetic Waves

► **Properties of electromagnetic waves.** The electromagnetic wave is the propagation of perturbations of an electromagnetic field in empty space or in a medium without sources. The existence of electromagnetic waves follows from the Maxwell theory (see Subsection P3.11.1), which states that an alternating electric field generates an alternating magnetic field and an alternating magnetic field generates an alternating electric field. In the case of a homogeneous isotropic medium, the Maxwell equation have the form

$$\begin{aligned} &1)\ \operatorname{curl}\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, && 2)\ \operatorname{curl}\mathbf{H} = \frac{\partial \mathbf{D}}{\partial t}, \\ &3)\ \operatorname{div}\mathbf{D} = 0, && 4)\ \operatorname{div}\mathbf{B} = 0, \end{aligned} \qquad \text{(P4.2.2.1)}$$

where $\mathbf{D} = \varepsilon\varepsilon_0\mathbf{E}$ and $\mathbf{B} = \mu\mu_0\mathbf{H}$. For the electromagnetic field depending only on the coordinate x, we obtain the equations:

$$\begin{aligned} -\mu\mu_0\frac{\partial H_z}{\partial t} &= \frac{\partial E_y}{\partial x}, & \mu\mu_0\frac{\partial H_y}{\partial t} &= \frac{\partial E_z}{\partial x}, \\ \varepsilon\varepsilon_0\frac{\partial E_y}{\partial t} &= -\frac{\partial H_z}{\partial x}, & \varepsilon\varepsilon_0\frac{\partial E_z}{\partial t} &= \frac{\partial H_y}{\partial x}. \end{aligned} \qquad \text{(P4.2.2.2)}$$

In this case, the equations for longitudinal components have the form $\partial E_x/\partial x = \partial E_x/\partial t = 0$, $\partial H_x/\partial x = \partial H_x/\partial t = 0$, that is, all longitudinal components of the field are constant quantities.

It is easy to see that the transverse components of the electromagnetic field satisfy the wave equation:

$$\Delta E_y = \varepsilon\varepsilon_0\mu\mu_0\frac{\partial^2 E_y}{\partial t^2}, \qquad \Delta H_z = \varepsilon\varepsilon_0\mu\mu_0\frac{\partial^2 H_z}{\partial t^2}$$

(the quantities E_z and H_y satisfy the same equations). It follows from these equations that the electromagnetic perturbations in a medium propagate with the velocity

$$v = \frac{1}{\sqrt{\varepsilon\varepsilon_0\mu\mu_0}} = \frac{c}{\sqrt{\varepsilon\mu}} = \frac{c}{n}, \qquad \text{(P4.2.2.3)}$$

where $c = 1/\sqrt{\varepsilon_0\mu_0} \approx 3\times 10^8\ m/c$ is the *velocity of electromagnetic waves in vacuum*, and $n = \sqrt{\varepsilon\mu}$ is the *index of refraction* of the medium.

We consider the plane electromagnetic wave of a certain frequency (the *monochromatic plane wave*) propagating along the axis x: $E_y = E_{0y}\cos(\omega t - kx)$, $E_z = E_{0z}\cos(\omega t - kx + \varphi)$. From (P4.2.2.2) we find that $H_y = -\sqrt{\varepsilon\varepsilon_0/\mu\mu_0}\,E_z$ and $H_z = \sqrt{\varepsilon\varepsilon_0/\mu\mu_0}\,E_y$. These formulas permit one to obtain the following properties of the plane traveling electromagnetic wave:

1) The electromagnetic wave is *transverse*.

2) Since $\mathbf{E}\cdot\mathbf{H} = 0$, it follows that $\mathbf{E} \perp \mathbf{H}$. The vectors ($\mathbf{E}$, $\mathbf{H}$, $\mathbf{k}$) form the right triple of vectors; that is, the direction of the vector $\mathbf{E}\times\mathbf{H}$ coincides with the direction of the wave propagation.

3) The values of the vectors $\mathbf{E}$ and $\mathbf{H}$ at each time moment are related as $\sqrt{\varepsilon\varepsilon_0}\,E = \sqrt{\mu\mu_0}\,H$ ($E = vB$).

The character of the wave polarization depends on the phase difference φ. For $\varphi = \pm m\pi$, the wave is *plane polarized*; that is, the vector $\mathbf{E}$ oscillates in one plane (this plane is called the *wave polarization plane*), and for another phase difference, the wave is *elliptically polarized*, or (for $\varphi = \pm(m + \frac{1}{2})\pi$ and for $E_{0x} = E_{0y}$), the wave is *circularly polarized* (see Subsection P4.1.1).

► **Energy of electromagnetic waves.** The volume density of the electromagnetic wave energy is calculated by the formula

$$w = \tfrac{1}{2}\varepsilon\varepsilon_0 E^2 + \tfrac{1}{2}\mu\mu_0 H^2 = \varepsilon\varepsilon_0 E^2 = \mu\mu_0 H^2 = \frac{1}{v}EH.$$

In the case of a monochromatic plane polarized wave, we have

$$w = \varepsilon\varepsilon_0 E_0^2 \cos^2(\omega t - kx), \qquad \langle w\rangle = \tfrac{1}{2}\varepsilon\varepsilon_0 E_0^2.$$

For the elliptically polarized wave:

$$w = \varepsilon\varepsilon_0\left[E_{0y}^2\cos^2(\omega t - kx) + E_{0z}^2\cos^2(\omega t - kx + \varphi)\right],$$
$$\langle w\rangle = \tfrac{1}{2}\varepsilon\varepsilon_0\left(E_{y0}^2 + E_{z0}^2\right).$$

For the wave circularly polarized ($E_{0y} = E_{0z} = E_0$, $\varphi = \frac{1}{2}\pi \pm \pi m$), the energy density does not vary in time: $w = \varepsilon\varepsilon_0 E_0^2$.

The energy flux density in an electromagnetic wave is determined by the Poynting vector (see Subsection P3.11.2):

$$\mathbf{S} = w\mathbf{v} = \mathbf{E}\times\mathbf{H},$$

which points in the direction of the wave propagation. The *intensity of an electromagnetic wave* is the magnitude of the time-average of the Poynting vector:

$$I = |\langle\mathbf{S}\rangle| = v\langle w\rangle.$$

► **Wave pressure.** According to the theory of relativity, the energy flux density of electromagnetic field in a vacuum means that there exists a momentum of the electromagnetic field. The volume density of the electromagnetic field momentum is equal (in magnitude) to w/c, the momentum flux density is equal to $\mathbf{S}/c$, and its modulus is equal to w. In the case of wave reflection or absorption, there is a variation in the wave momentum, which must manifest itself in the wave pressure on an obstacle. In the case of normal reflection of a plane wave, the pressure is equal to $p = \langle w\rangle(1+R)$, where R is the fraction of the reflected energy (the *coefficient of reflection*). Within the framework of electrodynamics, the wave pressure is explained by the action of the wave magnetic field on the current excited on reflecting surface by the wave electric field.

► **Radiation of plane current.** The simplest example of a system creating alternating fields and radiating a plane electromagnetic wave is the time-alternating plane current. We consider the current in the plane (x, z) directed along z and varying by the law $i_z = i_0\cos\omega t$ (Fig. P4.15). It follows from the second of Maxwell's equations in integral form (see Subsection P3.11.1) that the displacement current vanishes, and hence, for a narrow loop surrounding the current, we have $B_x(0,t) = \pm\frac{1}{2}\mu_0 i_0\cos\omega t$, where the minus sign corresponds to the field on the right of the plane (for small $y > 0$), and the plus sign corresponds to the field on the left of the plane. Using the relation between E and B in a plane wave, we obtain $E_z(0,t) = -A_E\cos\omega t$, where $A_E = \frac{1}{2}\mu_0 c i_0 = i_0/(2\varepsilon_0 c)$. The electric field is directed opposite to the current, its work over the current is negative; that is, the current gives energy to the wave. The field far from the plane has the form $E(y,t) = -A_E\cos(\omega t - ky)$ for $y > 0$ and $E(y,t) = -A_E\cos(\omega t + ky)$ for $y < 0$.

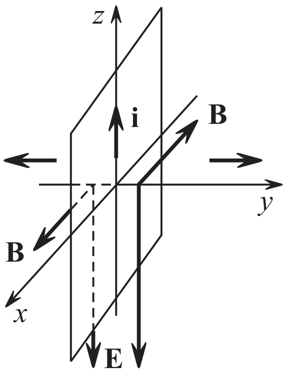

Figure P4.15. Radiation of plane current.

► **Wave radiation by moving charges.** The radiation of electromagnetic waves occurs in accelerated motion of charges. An electric dipole whose dipole moment rapidly varies in time (a *Hertz oscillator* or a *Hertz vibrator*) is a simple radiating system. Since the electromagnetic perturbations propagate with velocity c, the field of radiation at the distance r at time t is determined by the motion of charges in the dipole at time $t' = t - r/c$. The radiation field has a rather simple form only in the *wave zone*; that is, at distances that are large compared with both the dipole dimensions and $\lambda = c\tau$, where τ is the typical time of the dipole moment variation (in the case of harmonic oscillations, λ is the wavelength). The electric and magnetic fields in the wave zones are perpendicular to the radius vector $\mathbf{r}$ drawn from the dipole to the point of observation:

$$\mathbf{E} = \frac{1}{4\pi\varepsilon_0 c^2 r^3}\bigl[(\ddot{\mathbf{p}}\times\mathbf{r})\times\mathbf{r}\bigr]_{t-r/c}, \qquad \mathbf{B} = \frac{\mu_0}{4\pi c r^2}\bigl[\ddot{\mathbf{p}}\times\mathbf{r}\bigr]_{t-r/c}.$$

The mutual location of the vectors $\mathbf{E}$, $\mathbf{H}$, and $\mathbf{r}$ is shown in Fig. P4.16.

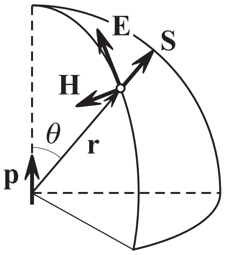

Figure P4.16. Radiation of a moving electric dipole.

The solution of the Maxwell equations in the form or retarded potentials (see Subsection P3.11.1) permits one to write the expressions for the potentials at the point $\mathbf{r}$ at time t. For example, the vector potential has the form:

$$\mathbf{A}(\mathbf{r},t) = \frac{\mu_0}{4\pi}\frac{q\mathbf{v}(t-r/c)}{r} = \frac{\mu_0}{4\pi}\frac{\dot{\mathbf{p}}(t-r/c)}{r}.$$

In calculations of $\mathbf{B} = \operatorname{curl}\mathbf{A}$, the principal role in the wave zone is played by the term arising in the argument differentiation $\dot{\mathbf{p}}$, which results in the above expression for $\mathbf{B}$. Similar but more cumbersome calculations can also be performed for the electric field.

The fields decrease with the distance according to the law $1/r$. The Poynting vector in the wave zone is directed along $\mathbf{r}$; in the direction making an angle θ with the vector $\mathbf{p}$, it is equal to

$$S = \frac{1}{4\pi\varepsilon_0}\frac{\sin^2\theta}{4\pi c^3 r^2}\bigl(\ddot{\mathbf{p}}\big|_{t-r/c}\bigr)^2.$$

The energy flux through a closed surface is precisely the power of the dipole radiation:

$$P = \int_0^\pi S(\theta) 2\pi r^2 \sin\theta \, d\theta = \frac{1}{4\pi\varepsilon_0} \frac{2\ddot{\mathbf{p}}^2}{3c^3}.$$

Assuming that only one of the dipole charges is in motion, we obtain the formula for the power of radiation of a particle in accelerated motion:

$$P = \frac{1}{4\pi\varepsilon_0} \frac{2q^2\ddot{\mathbf{r}}^2}{3c^3}, \qquad \text{(P4.2.2.4)}$$

which is proportional to its squared acceleration.

An electric dipole whose moment varies by the harmonic law, $\mathbf{p} = \mathbf{p}_0 \cos\omega t$, is a *linear harmonic oscillator*. The dependence of the radiated wave intensity on the angle θ has the form

$$I = |\langle\mathbf{S}\rangle| = \frac{1}{4\pi\varepsilon_0} \frac{p_0^2 \omega^4 \sin^2\theta}{8\pi c^3 r^2}.$$

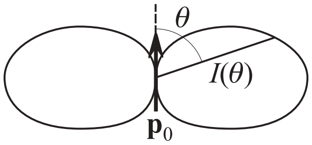

Figure P4.17. Directional pattern of a linear oscillator.

The dependence $I(\theta)$ for $r = \text{const}$ (the *directional pattern*) is shown in Fig. P4.17. The time-average power of radiation of such an oscillator

$$P = \frac{1}{4\pi\varepsilon_0} \frac{\omega^4 \mathbf{p}_0^2}{3c^3} \qquad \text{(P4.2.2.5)}$$

is proportional to the fourth power of its frequency.

Example 1. A free electron (with mass m) in the field of a plane electromagnetic wave with amplitude E_0 exhibits forced oscillations with amplitude $A = eE_0/(m\omega^2)$ (the equation of motion of the electron has the form $m\ddot{x} = -eE_0 \cos\omega t$). One can see that the radiation power of a free electron in the field of an electromagnetic wave is independent of the frequency, and the scattering cross-section (the ratio of the radiation power to the incident wave intensity) is equal to

$$\sigma = \frac{P}{I} = \frac{8}{3} \frac{(ke^2)^2}{m^2 c^4} \sim 10^{-30}\ \text{m}^2.$$

Another situation arises if, in the electromagnetic wave field, there is an atomic electron the natural frequency ω_0 of whose oscillations is large compared to the frequency of the incident (light) wave. Then the oscillation amplitude (see Subsection P4.1.3) is equal to $A = eE_0/(m\omega_0^2)$; that is, it is almost independent of ω. In this case, the radiation power and the light scattering cross-section are proportional to ω^4, which qualitatively explains the blue color of the sky (the blue color scattering is stronger).

Example 2. In a hydrogen atom, let the electron move about the nucleus in a circular orbit with a decreasing radius r and speed v. The electron energy and radiation power are

$$W = \frac{mv^2}{2} - \frac{ke^2}{r} = -\frac{ke^2}{2r}, \qquad P = \frac{2ke^2\ddot{r}}{3c^3} = \frac{2}{3} \frac{(ke^2)^3}{m^2 c^3} \frac{1}{r^4},$$

since $m\ddot{r} = mv^2/r = ke^2/r^2$. It follows from the law of conservation of energy, $P = -dW/dt$, that

$$\frac{dr}{dt} = -\frac{4}{3} \frac{(ke^2)^2}{m^2 c^3} \frac{1}{r^2}.$$

Separating the variables and integrating, we find the time in which the radius decreases from its initial value r_0 to zero (i.e., the time in which the electron falls onto the nucleus, or the *deexcitation time*):

$$t = \frac{1}{4} \frac{m^2 c^3}{(ke^2)^2} r_0^3 \sim 10^{-10}\ \text{s},$$

where $r_0 \sim 10^{-10}$ m is the atomic radius.

► **Doppler effect for electromagnetic waves in vacuum.** The main distinction from the acoustic Doppler effect (see Subsection P4.2.1) is that there is no reference system associated with the medium, where the wave propagates. Therefore, the relation between the radiation frequency ν_0 (measured in the source reference system K_S) and the registered frequency ν (measured in the receiver reference system K_R) can depend only on their relative velocity. Assume that in K_R, the source moves with velocity v directed at the angle θ made with the radius vector drawn from the receiver to the source. Repeating the argument given in Subsection P4.2.1, we obtain

$$\nu = \frac{\widetilde{\nu}_0}{1 + (v/c)\cos\theta},$$

where $\widetilde{\nu}_0$ is the source frequency measured by the clock in the system K_R. Passing to the intrinsic time ($\widetilde{T}_0 = \gamma T_0$, or $\widetilde{\nu}_0 = \nu_0/\gamma$; see Subsection P1.10.2), we obtain

$$\nu = \frac{\nu_0}{\gamma[1 + (v/c)\cos\theta]} = \frac{\nu_0\sqrt{1 - v^2/c^2}}{1 + (v/c)\cos\theta}. \qquad \text{(P4.2.2.6)}$$

One can see that, because of the time transformation, not only the longitudinal but also the transverse Doppler effect is observed (for the source motion perpendicular to the direction to the receiver).

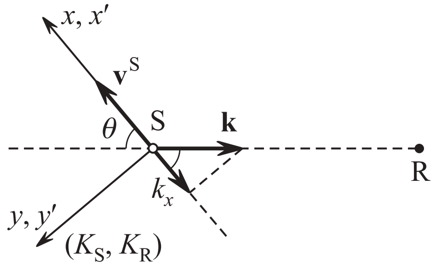

Figure P4.18. To the explanation of the Doppler effect for electromagnetic waves.

It is useful to consider the Doppler effect from a different viewpoint. Since the oscillation phase at a given point of the wave ($\omega t - \mathbf{k}\cdot\mathbf{r}$) must be invariant under the transition to another reference system, the ($\mathbf{k}, \omega/c$) is a four-vector of zero length ($k^2 - \omega^2/c^2 = 0$). If, in the frame K_R, the axis x is directed parallel to the source velocity $\mathbf{v}^\mathrm{S}$ (Fig. P4.18), then, in this reference system, we have $k_x = -k\cos\theta = -(\omega/c)\cos\theta$. Using the Lorentz transformations to pass to the frame K_S, we obtain: $\omega_0/c = \gamma[(\omega/c) - k_x(v/c)]$, or $\omega_0 = \gamma[\omega + \omega(v/c)\cos\theta]$. Expressing ω, we obtain (P4.2.2.6). This method allows us, by writing the Lorentz transformations for the wave vector $\mathbf{k}$, to obtain formulas for the light aberration phenomenon, which is important in astronomy.

► **Scale of electromagnetic waves.** Depending on the way of radiation and registration of electromagnetic waves, several ranges with conditional (overlapping) boundaries are distinguished in the scale of the frequency ν or of the wavelengths in a vacuum $\lambda = c/\nu$:

1. The radio waves ($\lambda > 0.05$ mm).
2. The optical (light) radiation (10 nm $< \lambda <$ 1 mm):
   a) the infrared (IR) radiation (770 nm $< \lambda <$ 1 mm),
   b) the visible light (380 nm $< \lambda <$ 770 nm),
   c) the ultraviolet (UV) radiation (10 nm $< \lambda <$ 380 nm).
3. The X-rays radiation (0.01 pm $< \lambda <$ 100 nm).
4. The gamma radiation (gamma rays) ($\lambda < 0.1$ nm).

Bibliography for Chapter P4

Alenitsyn, A. G., Butikov, E. I., Kondratyev, A. S., *Concise Handbook of Mathematics and Physics*, CRC Press, Boca Raton, Florida, 1998.

Bekefi, G. and Barrett, A. H., *Electromagnetic Vibrations, Waves, and Radiation*, The MIT Press, Cambridge, Massachusetts, 1977.

Benenson, W., Harris, J. W., Stocker, and Lutz, H. (Eds.), *Handbook of Physics*, Springer-Verlag, New York, 2002.

Bloch, I., *The Physics of Oscillations and Waves: With Applications in Electricity and Mechanics*, Springer-Verlag, New York, 2004.

Crawford, F. S., Jr., *Waves and Oscillations. Berkeley Physics Course. Volume 3*, McGraw-Hill, New York, 1968.

Elmore, W. C. and Heald, M. A., *Physics of Waves*, Dover Publications, New York, 1985.

French, A. P., *Vibrations and Waves (M.I.T. Introductory Physics Series)*, W. W. Norton & Company, New York, 1971.

Gorelik, G. S., *Oscillations and Waves. Introduction to Acoustics, Radiophysics and Optics* [in Russian], Fizmatlit, Moscow, 2007.

Ingard, K. U., *Fundamentals of Waves and Oscillations*, Cambridge University Press, Cambridge, England, 1988.

Irodov, I. E., *Wave Processes. Basic Laws* [in Russian], Fizmatlit, Moscow, 2004.

Kneubühl, F. K., *Oscillations and Waves*, Springer, Berlin, 1997.

Pain, H. J., *The Physics of Vibrations and Waves, 6th Edition*, John Wiley & Sons, Chichester, England, 2005.

Prokhorov, A. M. (Ed.), *Encyclopedia of Physics, Volumes 1–5* [in Russian], Bolshaya Russkaya Encyclopedia, Moscow, 1998.

Prokhorov, A. M. (Ed.), *Encyclopedic Dictionary of Physics* [in Russian], Sovetskaya Encyclopedia, Moscow, 1984 (dic.academic.ru).

Sivukhin, D. V., *General Physics, Volume 4, Optics* [in Russian], Fizmatlit, Moscow, 2006.

Sivukhin, D. V., Yakovlev, I. A., *A Collection of Problems, Volume 4, Optics* [in Russian], Fizmatlit, Moscow, 2006.

Smith, W. F., *Waves and Oscillations: A Prelude to Quantum Mechanics* , Oxford University Press, Oxford, England, 2010.

Woan, G., *The Cambridge Handbook of Physics Formulas*, Cambridge University Press, Cambridge, England, 2003.

Chapter P5

Optics

P5.1. Geometric Optics. Radiometry

P5.1.1. Geometric Optics

► **Fundamental laws of geometric optics.** Optics studies the electromagnetic radiation of optic (light) range (see Subsection P4.2.2) as well as phenomena arising in its propagation in space and under interaction with matter. *Geometric optics* does not consider the wave character and polarization of light radiation and deals with notions of *light rays* pointing out the direction of light propagation and narrow *light beams* formed by light rays.

The fundamental laws of geometric optics are listed below.

1. *Law of rectilinear light propagation*.

2. *Law of independence of light beams*. The energy in each beam propagates independently of other beams; illumination of the surface on which several beams are incident, is equal to the sum of illuminations created by each beam separately.

3. *Law of light reflection*. The reflected ray lies in *incidence plane*, formed by the incident ray and the normal to the surface at the point of incidence; the angle of incidence is equal to the angle of reflection. All the angles are counted from the normal.

4. *Law of light refraction*. The refracted ray lies in the plane of incidence; the ratio of the sine of the angle of incidence α_1 to the sine of angle of refraction α_2 depends on the wavelength but is independent of the angle of incidence (Snell law):

$$\frac{\sin\alpha_1}{\sin\alpha_2} = n_{21} = \frac{n_2}{n_1}. \qquad \text{(P5.1.1.1)}$$

The constant quantity n_{21} is called the *relative index of refraction* of the second medium with respect to the first one, which is equal to the ratio of (*absolute*) *indices of refraction* of each of the media (of indices of refraction of the medium with respect to a vacuum). From the viewpoint of wave optics, the absolute index of refraction shows by what factor the phase velocity of the light wave of a given frequency is less than the velocity of this wave in a vacuum:

$$v = c/n, \qquad n = \sqrt{\mu\varepsilon} \approx \sqrt{\varepsilon} \qquad \text{(P5.1.1.2)}$$

(see Subsection P4.2.2). If $n_2 < n_1$, then for $\alpha_1 > \alpha_\text{c}$, where $\sin\alpha_\text{c} = n_{21} < 1$, one observes the *total reflection of light*; that is, the refracted ray is absent. If the light ray propagates in a medium with index of refraction gradually varying along an axis (in a layered medium), then it is convenient to use formula (P5.1.1.1) in the form $n\sin\alpha = \text{const}$, where α is the angle with this axis. One can see that α varies gradually, which means that the ray is bending.

The laws of geometric optics listed above must be supplemented with the *principle of reversibility of light beams*.

► **Limits of applicability of geometric optics.** The laws of geometric optics act where the phenomena of interference, diffraction, and polarization are insignificant. This happens when the wave amplitude and its first spatial derivatives vary just slightly over the wavelength. These conditions are violated at the shadow boundary, near the geometric point of convergence of light rays (focus), where light travels past narrow diaphragms, and also light propagates through media with rapidly varying index of refraction or with strong absorption. For example, when a light beam travels past a circular diaphragm of diameter d (for holes of irregular shape, d is understood as the minimum transverse dimension), it spreads out due to diffraction at a distance $l \sim d^2/\lambda$; one can use the laws of geometric optics at small distances compared to l.

► **Huygens principle and Fermat principle.** The laws of geometric optics are closely related to the *Huygens principle* (Huygens wave construction) and the *Fermat principle*. The Huygens principle is based on abstract-wave concepts and, at each point, permits constructing an auxiliary wave surface the normal to which shows the light ray direction. To construct the wave surface at time $t+\Delta t$, one considers all the points of the wave surface at time t as sources of secondary light waves. The envelope of spherical wave surfaces of secondary waves forms a new wave surface of the principal wave.

The Fermat principle is one of the examples of variational principles that play an important role in physics. It states that the time of light propagation along the true trajectory is extremal (usually, minimum or maximum) compared with all the virtual nearby trajectories. More precisely, this time does not vary (in the first order of magnitude) for a small distortion of the trajectory.

The Huygens and Fermat principles permit obtaining the fundamental laws of geometric optics but do not permit leaving its limits.

► **Optical path length.** It is convenient to express the time of light propagation through a medium with a variable index of refraction in terms of the *optical path length* L:

$$t = \int_1^2 \frac{dl}{c/n} = \frac{L}{c}, \qquad L = \int_1^2 n\,dl. \tag{P5.1.1.3}$$

This notion is also useful in studying the interference.

► **Images in optical systems.** One of the important problems of geometric optics is the construction of *images* formed by optical systems and the study of their properties. An image is the point of intersection of ray transmitted through an optical system. Images are *real* (the rays converge after leaving the system) and *virtual* (emergent rays diverge from the virtual point of intersection of their virtual continuations). The system of reflecting and refracting spherical (and plane) surfaces perpendicular to an axis forms the image of a point source by rays incident at small angles to the axis (*paraxial approximation*). The system axis is called the *principal optic axis*. The point of convergence of rays parallel to the axis is called the *focus of the system*; the plane perpendicular to the axis and drawn through the focus is called the *focal plane*. Like the images, the foci can be real and apparent. Additionally, one introduces the notion of "apparent source" denoting the point of intersection of imaginary continuations of the converging rays incident on the system.

In what follows, we use the important property of images: the optical path lengths of all rays from the source to the image are the same; that is, the optical system does not change the path difference of rays.

The basic elements of optical systems are the simple optical systems: plane and spherical mirrors, plane and spherical refracting surfaces, and thin lenses. The image formed by each of these elements is the source for the subsequent element.

When considering a simple system, one counts all the distances from the point of intersection of this system with the principal optical axis. The distance from this point to the source is denoted by s, the distance to the image, by s', and the distance to the focus, by f (*focal distance*). To real sources, images, and foci there correspond positive values, to apparent sources, negative values.

Below are formulas relating s, s', and f for simple optical systems.

1) **Spherical mirror** of radius R:

$$\frac{1}{s}+\frac{1}{s'}=\frac{1}{f}. \qquad \text{(P5.1.1.4)}$$

The focal distance is equal to $f = R/2$, for a concave mirror, $R > 0$, for a convex mirror, $R < 0$.

2) **Plane mirror** forms the image of a point source located behind the mirror symmetrically to the source. Formally, (P5.1.1.4) holds for it as $R \to \infty$; that is, $s' = -s$ (to a real source there corresponds a virtual image, to an apparent source, there corresponds a real image).

3) **Spherical refracting surface** of radius R with the relative index of refraction $n_{21} = n_2/n_1$:

$$\frac{1}{s}+\frac{n_{21}}{s'}=\frac{n_{21}-1}{R}.$$

For a convex surface, $R > 0$, for a concave surface, $R < 0$.

4) **Plane refracting surface** is obtained from the spherical surface in the limit as $|R| \to \infty$:

$$s' = -n_{21}s.$$

5) **Thin lens** bounded by spherical surfaces of radii R_1 and R_2. Relation (P5.1.1.4) holds, the focal distance of a thin lens is calculated by the formula:

$$\frac{1}{F}=(n_{21}-1)\left(\frac{1}{R_1}+\frac{1}{R_2}\right).$$

The radii R_1, R_2 are assumed to be positive for convex surfaces and negative for concave surfaces. The quantity $D = 1/F$ is called the *focal power*; it is measured in *dioptres* (dpt), with 1 dpt = 1/m.

The theoretical limit of the image dimensions and the resolving power of optical systems is determined by the light diffraction.

P5.1.2. Radiometry and Photometry

Radiometry studies the measurement of electromagnetic radiation at all wavelengths, including visible light. Photometry studies the measurement of light within the wavelength range visible by the human eye.

► **Characteristics of radiation in space.** The radiation power emitted by a radiation source at a given point of space in a particular direction is characterized by the *radiant intensity*, I:

$$d\Phi = I\cos\theta\, ds\, d\Omega. \qquad \text{(P5.1.2.1)}$$

The radiation power $d\Phi$ is called the *radiant flux*; it is the energy transferred per unit time through the area element ds by the light rays contained in the solid angle $d\Omega$. The angle θ is the angle between a selected direction and the normal to the area element. The product $\cos\theta\, ds$ is called the *apparent size* of the area element in the specified direction. The volume density of *radiant energy*, $u = dW/dV$, is calculated as

$$u=\frac{1}{c}\int I\, d\Omega. \qquad \text{(P5.1.2.2)}$$

In the case of isotropic radiation, $u = 4\pi I/c$. For the intensity I and the density u, one can obtain the *spectral decomposition* in frequencies or wavelengths:

$$u = \int_0^\infty u_\nu\, d\nu = \int_0^\infty u_\omega\, d\omega = \int_0^\infty u_\lambda\, d\lambda.$$

Equating the energy contained in the mutually corresponding spectral intervals, one can find the relation between distinct spectral representations. For example, by setting $u_\nu\,|d\nu| = u_\lambda\,|d\lambda|$ with $\lambda = c/\nu$ taken into account, we obtain

$$u_\nu = \left(\frac{c}{\nu^2}\right) u_\lambda.$$

Similarly, we can obtain the decomposition in two mutually perpendicular polarizations.

► **Light and energy characteristics.** The above-introduced *energy characteristics* are used to obtain an objective description of the radiation. The *photometric* or *light* characteristics take account of the sensitivity of the eye to the optical radiation. The radiant power is measured in watts and the photometric unit corresponding to it is called the *luminous flux* and is measured in *lumens* (lm). To pass to the luminous flux, it is necessary to multiply the radiant power by the *luminous efficacy* η (measured in lm/W), which is equal to $\eta_0 = 625\,\mathrm{lm/W}$ for the wavelength $\lambda_0 = 555$ nm and to $\eta_0 V_\lambda$ for other wavelengths, where V_λ is the curve of *visibility*, or *relative spectral sensitivity* of the eye ($V_\lambda = 1$ for $\lambda = 555$ nm and decreases to zero approaching the boundaries of the optical range). Recalculations from arbitrary radiant power to luminous flux can be performed by the formula $\Phi_{\mathrm{lu}} = \eta_0 \int \Phi_\lambda V_\lambda\, d\lambda$. Recalculations from other energy characteristics to luminous characteristics and conversely can be performed similarly (for example, from the radiation intensity to the light intensity).

► **Characteristics of radiation source.** The point source is characterized by the *source energy power* determined as the radiant power in a given direction per unit solid angle: $d\Phi = J\, d\Omega$. The luminous characteristic — that is, the *source luminous intensity* — is determined in terms of the luminous flow. The radiant power unit, *candela*, is one of the basic SI units (1 lm = 1 cd × 1 sr, where “cd” stands for the *candela* and “sr” stands for the *steradian*). If there is no absorption or scattering of radiation, then the radiation intensity in the same direction at the distance r from the source is equal to

$$I = \frac{J}{r^2}.$$

For area sources—that is, for radiating surfaces—the notion of *energy radiance* in a given direction B_θ is introduced (the luminous characteristic is the *brightness*). The energy power emitted by a surface element ds is equal to

$$dJ = B_\theta \cos\theta\, ds \quad \text{or} \quad d\Phi = B_\theta \cos\theta\, ds\, d\Omega, \qquad \text{(P5.1.2.3)}$$

where θ is the angle between the normal to the radiating surface and the radiation direction. The source whose B_θ is independent of the direction is called a *Lambertian source*. A Lambertian source of any shape appears uniformly bright (for example, the solar disk). The *radiant exitance* R (the luminous characteristic is the *luminous exitance*) is determined as the total radiant power in the solid angle 2π per unit area of a radiating surface:

$$R = \int B_\theta \cos\theta\, d\Omega. \qquad \text{(P5.1.2.4)}$$

For a Lambertian source, $R = \pi B$.

► **Irradiance.** The radiant energy flux per unit area of illuminated surface is called the *irradiance* and denoted E. The corresponding light quantity is called the *illuminance* and measured in *luxes* (lx), with $1\,\mathrm{lx} = 1\,\mathrm{lm}/\mathrm{m}^2$. If a point source radiation is incident at an angle θ with the normal, then

$$E = \frac{J}{r^2}\cos\theta.$$

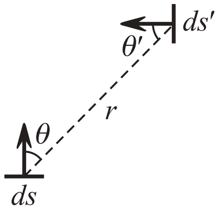

Figure P5.1. To calculations of the irradiance created by an area source.

The irradiance created by a luminous surface S on a site of the irradiated surface ds' is equal to the sum of contributions of all the sites ds. The solid angle from which ds' is viewed at ds is equal to $\cos\theta'\,ds'/r^2$ (Fig. P5.1). For the luminous flux radiated into this solid angle, we have $B_\theta\cos\theta\,ds(\cos\theta'\,ds'/r^2)$. Since $\cos\theta\,ds/r^2$ is equal to the solid angle $d\Omega$ from which the radiating site is viewed at the point of irradiance, it follows that $dE = B_\theta\cos\theta'\,d\Omega$.

Example. Calculate the irradiance created by a Lambertian source of disk shape irradiating a small area located on the disk axis oriented perpendicularly to the axis (Fig. P5.2).

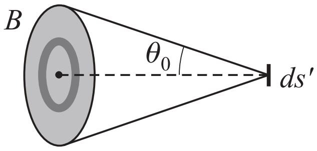

Figure P5.2. Irradiance of a surface element ds' by a Lambertian disk.

Solution. The solid angle element bounded by θ' and $\theta' + d\theta'$ is equal to $d\Omega = 2\pi\sin\theta' d\theta'$. After integration, we obtain $E = \pi B\sin^2\theta_0$, where θ_0 is the cone half-opening angle from which the disk is viewed from the irradiated area element.

P5.2. Wave Properties of Light

P5.2.1. Interference of Light

► **Coherence of light waves.** In the superposition of light waves with a very high frequency of oscillations, one can observe only the time-average energy of oscillations, which is characterized by the *oscillation intensity* $I \sim \langle\mathbf{E}^2\rangle$. When adding two oscillations $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$, we obtain $I = I_1 + I_2 + I_{12}$, where $I_{12} = 2\langle\mathbf{E}_1\cdot\mathbf{E}_2\rangle$ is called the interference term. If $I_{12} = 0$, then the light beams are said to be *incoherent*. The waves polarized in mutually perpendicular directions are always incoherent. Monochromatic waves are coherent only if their phase difference is constant (that is, their frequencies coincide) and if they are parallel polarized (or almost parallel polarized). In superposition of such waves,

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\Delta\varphi,$$

where $\Delta\varphi$ is the oscillation phase difference at a given point. In the special case for $I_1 = I_2 = I_0$, we have $I = 2I_0(1 + \cos\Delta\varphi)$. For in-phase oscillations, $I = (\sqrt{I_1} + \sqrt{I_2}\,)^2$, for anti-phase oscillations, $I = (\sqrt{I_1} - \sqrt{I_2}\,)^2$.

► **Interference of two waves.** In superposition of two plane waves $E_1 = E_{10}\cos(\omega t - \mathbf{k}_1 \cdot \mathbf{r})$ and $E_2 = E_{20}\cos(\omega t - \mathbf{k}_2 \cdot \mathbf{r})$, we obtain $\Delta\varphi = (\mathbf{k}_1 - \mathbf{k}_2)\cdot\mathbf{r}$. The surfaces of constant phase difference are planes perpendicular to the vector $\mathbf{k}_1 - \mathbf{k}_2$. The distance between neighboring planes with maximum intensity is equal to

$$\Delta x = \frac{2\pi}{|\mathbf{k}_1 - \mathbf{k}_2|} = \frac{\pi}{k\sin(\alpha/2)} = \frac{\lambda}{2\sin(\alpha/2)}, \qquad \text{(P5.2.1.1)}$$

where α is the angle between the vectors $\mathbf{k}_1$ and $\mathbf{k}_2$. For small α, we obtain $\Delta x \sim \lambda/\alpha$. If in the region of wave overlap there is a plane screen parallel to $\mathbf{k}_1 - \mathbf{k}_2$, then one can observe parallel alternating dark and light fringes on it.

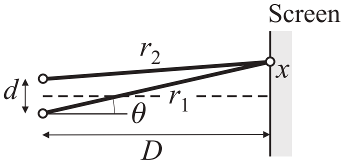

Figure P5.3. Interference of two closely located linear sources.

In superposition of spherical waves from two in-phase point sources, the maximum condition for the intensity of the mth order has the form $r_2 - r_1 = m\lambda$ (this is the equation of a hyperboloid of rotation whose axis passes through the sources). If a screen is placed parallel to this axis, we obtain light and dark fringes of hyperbolic shape. If the distance to the screen D is large compared to the distance d between the sources, then, at the screen center, we obtain equidistant almost parallel fringes. The distance between the fringes is the same as in the *Young experiment*, where the interference is created by two parallel linear sources of coherent light. If x is the distance from a screen point to its center (Fig. P5.3), then $r_1^2 = D^2 + (x + d/2)^2$, $r_2^2 = D^2 + (x - d/2)^2$, and for $d, x \ll D$, we obtain the optical path difference $\Delta = r_1 - r_2 \approx xd/D \approx \alpha x$, where $\alpha \approx d/D$ is the angle from which the source is viewed at the screen center. The intensity varies from $4I$ for $\Delta = \alpha x = m\lambda$ to zero for $\Delta = (m + \frac{1}{2})\lambda$. The distance between the fringes is equal to $\Delta x = \lambda/\alpha = \lambda D/d$.

► **Fraunhofer approximation.** At a large distance from two in-phase linear sources ($r_1, r_2 \gg d$), one can write $r_2 - r_1 \approx d\sin\theta$, where θ is the angle between the normal to the plane of sources and the direction of the interference observation (*the Fraunhofer approximation*). The oscillation amplitudes in interfering waves are assumed to be equal; that is, $I = 2I_0[1 + \cos(kd\sin\theta)]$. The interference maxima are observed at the angles satisfying the relation $\sin\theta = m\lambda/d$.

Let us specify at what distance from the sources the Fraunhofer approximation can be used. We express r_2 according to the cosine theorem:

$$r_2 = (r_1^2 + d^2 - 2dr_1\sin\theta)^{1/2}$$

(the angle in the triangle is equal to 90° – θ), expand it up to the terms of the second order in d/r_1, and obtain

$$r_2 - r_1 \approx d\sin\theta + \frac{d^2}{2r_1} - \frac{d^2}{r_1}\sin^2\theta.$$

The correction can be neglected if $d^2/r_1 \ll \lambda$. We see that the Fraunhofer approximation works in the case where the distance r from the observation point to the sources satisfies the relation

$$r \gg \frac{d^2}{\lambda}. \qquad \text{(P5.2.1.2)}$$

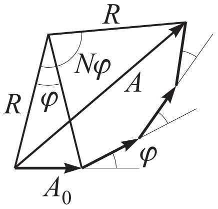

Figure P5.4. Composition of oscillations from several sources.

We note that this inequality is stronger than the simple condition $r \gg d$.

► **Interference of many waves.** We consider the interference of N equal in-phase sources located on the same straight line at the distance d from one another. At a large distance from the sources ($r \gg (Nd)^2/\lambda$) in the direction of θ, the composition of N oscillations of the same amplitude A_0 occurs, and the phase difference between the neighboring oscillations is equal to $\phi = (2\pi/\lambda)d\sin\theta$. The resultant amplitude A is obtained from the equations (see the vector diagram in Fig. P5.4): $A_0 = R\sin(\phi/2)$, $A = R\sin(N\phi/2)$. Eliminating the radius of the circumscribed circle R, we obtain

$$A = A_0\frac{\sin(N\phi/2)}{\sin(\phi/2)}, \qquad I = I_0\frac{\sin^2(N\phi/2)}{\sin^2(\phi/2)}. \tag{P5.1.2.3}$$

► **Interference from natural sources of light.** The radiation of natural (thermal) sources of light is the set of *wave trains* spontaneously radiated by excited atoms in their *luminescence*; that is, when they return to normal state. The train duration is $\sim 10^{-10}$–10^{-8} c, it contains 10^6–10^8 oscillations. Hence two different natural sources are incoherent even if a narrow spectral band is separated in their radiation, because the oscillation phase difference varies very fast and chaotically at each observation point. To observe the interference, it is necessary to split the radiation of one source into two or several beams and to make them to enter the observation point by different paths. In this case, each train interferes with itself, and the maximum or minimum condition is simultaneously satisfied for all the trains of the same frequency radiated from the same point of the source. It is convenient to introduce *optical path difference* of beams

$$\Delta = \int n_2\,dl - \int n_1\,dl,$$

where the integration is performed over the line passed by a given beam from the radiation point to the interference point. The maximum condition is $\Delta = m\lambda$; the minimum condition is $\Delta = \left(m + \frac{1}{2}\right)\lambda$, where m is called the *order of interference*.

We present the classical examples of obtaining two coherent sources.

1. *The Young experiment* (mentioned above). The sunlight is incident on a very narrow slit in the first screen, diverges because of diffraction, and is incident on the two narrow slits in the second screen. Again because of diffraction behind these slits, the light splits and forms overlapping coherent beams.

2. *Fresnel mirrors*. The light from the brightly illuminated slit is incident on two mirrors placed at an angle nearly equal to 180°. The closely located virtual images of the slit form two coherent sources.

3. *Fresnel biprism*. The light from a brightly illuminated slit refracts in two glass prisms with small refraction angles that contact by their bases. As the result of refraction, two closely located virtual images of the slit are formed.

4. *Billet split lens*. A collecting lens is cut into halves and the halves are slightly moved apart. The split lens is illuminated through a narrow split parallel to the cut line. Each of the lens halves forms its own real image of the slit.

5. *Lloyd mirror*. The light from a narrow slit is reflected from the mirror plane and forms the virtual image of the slit. The light from the slit itself and from its image interfere.

► **The influence of the source dimensions. The spatial coherence.** The optical path difference at a given point of the screen has a definite value only in the case of a point source. The path difference varies in transitions from point to point of the area source. If the path difference varies by $\lambda/2$, then the maximum condition becomes the minimum condition; that is, the interference patterns from different sites of the area source overlap and the total interference pattern is smeared.

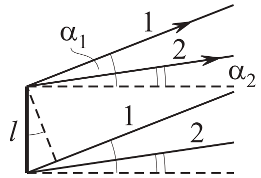

Figure P5.5. Influence of the source dimensions on the interference pattern.

For example, we consider a plane source. We assume that the first of the interfering rays is radiated by the source at an angle α_1 with its normal, and the second, at an angle α_2. We assume that these rays lie in the one plane with the normal and the angles of radiation are almost the same for all point of the source (Fig. P5.5). In transition from one end of the source to the other, the path length of the first ray varies by $l\sin\alpha_1$, and that of the second ray varies by $l\sin\alpha_2$ (l is the dimension of the source in the plane of rays), and the path difference varies by $l(\sin\alpha_2-\sin\alpha_1)$. The interference pattern conservation condition (condition of *spatial coherence*) takes the form

$$l\,|\sin\alpha_2-\sin\alpha_1|<\lambda/2.$$

Example 1. In the Young experiment, l is the slit width in the first screen, and $\alpha_1=-\alpha_2=b/(2L)$, where b is the distance between the slits in the second screen and $L\gg b$ is the distance between the screens; we obtain: $2l(b/2L)<\lambda/2$; that is, $l<\lambda L/(2b)$.

We rewrite this formula as: $b<\lambda/(2\psi)$, where $\psi=l/L$ is the angular dimension of the source or the divergence angle of the rays incident on the slits. For example, in the case of the sun direct illumination of the slits, $\psi\approx 0.009$ rad, $\lambda\approx 500$ nm; that is, the distance between the slits must be less than 0.06 mm. Just therefore, the sun light must first be transmitted through a narrow slit.

► **Influence of the light nonmonochromaticity. Temporal coherence.** A violation of the monochromaticity of the combined waves may result in the interference pattern smearing. We assume that the frequencies of the radiated waves lie in a narrow spectral interval $\Delta\omega$. A strong distortion of the interference pattern occurs if the path difference Δ exceeds a certain critical value L_c, which is called the *coherence length*; L_c has the meaning of the *length of the wave train* radiated by the atom in a single act of radiation (we recall that stable interference occurs when the trains in one of the separated beams are combined with the trains corresponding to them in the other beam).

The time $\tau_c=L_c/c$ that is called the *coherence time* has the meaning of *duration of the wave train*. According to general properties of the Fourier transform, the spectral width of a wave packet is related to its duration as $\Delta\omega\,\Delta t\sim 2\pi$.

Example 2. The "truncated" sinusoid $A(t)\sin\omega_0 t$, where $A=\text{const}$ in the interval $-\tau/2<t<\tau/2$ and $A=0$ outside this interval, has the Fourier transform $a(\omega)\sim\dfrac{\sin[\frac{1}{2}(\omega-\omega_0)\tau]}{(\omega-\omega_0)\tau}$ whose width is equal to $\Delta\omega\sim\dfrac{2\pi}{\tau}$.

We see that the spectral width of interference is related to the coherence duration as $\Delta\omega \sim 2\pi/\tau_c$. The maximum order of the spectrum, where one can observe the interference can be estimated as $N \sim \tau_c/T = \omega/\Delta\omega = \lambda/\Delta\lambda$. It is useful to note that the bands of neighboring orders merge just in this order:

$$(N+1)\lambda = N(\lambda + \delta\lambda)$$

(the trailing edge of the band of order $N+1$ coincides with the leading edge of the band of order N). The coherence length is often expressed in terms of $\delta\lambda$:

$$L_c = N\lambda = \frac{\lambda^2}{\delta\lambda}.$$

► **Interference in thin films.** We consider the interference of rays reflected from the leading and trailing surfaces of a thin film (Fig. P5.6). We assume that the wave front is plane; that is, the source is sufficiently remote. Since the front of refracted wave is perpendicular to the ray, it follows that ray 1 at point D and ray 2 at point D' have equal phases. Hence the optical path difference of rays at point D is equal to $n(D'C + CD)$. Moreover, $\lambda/2$ must be added to the path difference, which permits taking account of the phase variation on π in reflection from a medium with a greater index of refraction (on the air–film interface at point D). After transformations, we obtain:

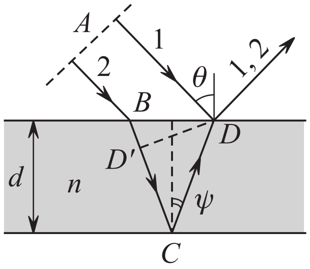

Figure P5.6. Interference in light reflection from a thin film.

$$\Delta = 2dn\cos\psi + \frac{\lambda}{2} = 2d\sqrt{n^2 - \sin^2\theta} + \frac{\lambda}{2},$$

where d is the film width, n is its index of refraction, and θ and ψ are the incidence and refraction angles. The maximum condition for observation in reflected light $\Delta = m\lambda$ is satisfied for certain wavelengths. For very thin films, the maximum condition is satisfied for one or two wavelengths from the visible region, and the film turns out to be colored. The maximum condition for observation in reflected light corresponds to the minimum condition for observation in transmitted light (no reflection occurs from the air–film interface). As always in interference, the energy does not increase but is redistributed.

Now we consider two important cases:

1) *Fringes of equal thickness*. If the rays are incident at an almost constant angle, for example, normally and the film width varies, then the lines of constant thickness are lines of constant path difference. In illumination by monochromatic light, these lines are visible as dark or light fringes. In observation in white light (for films of small thickness) the lines are colored. The interference occurs near the film surface (interference pattern is localized at the surface).

2) *Fringes of equal inclination*. The film thickness is constant, the illumination is by scattered light from a remote source. Changing the angle of observation, we alternatively

obtain the maximum condition and the minimum condition. The interference pattern is localized at infinity (or in the focal plane lens). For extremely thin films, the light can be nonmonochromatic; the observation at a given angle separates the wavelength for which the maximum condition is satisfied.

Example 3. **Newton rings**. If a plane-convex lens is put on the surface of a glass plate (Fig. P5.7) and illuminated by normally incident monochromatic light, then fringes of the shape of rings of equal thickness are observed in the air interval. The air interval thickness is equal to $d \approx \sqrt{R^2 + r^2} - R \approx r^2/(2R)$, where r is the circle radius. The minimum condition has the form $2d + (\lambda/2) = \left(m + \frac{1}{2}\right)\lambda$ or $r_m = \sqrt{m\lambda R}$. At the pattern center, there is a dark spot.

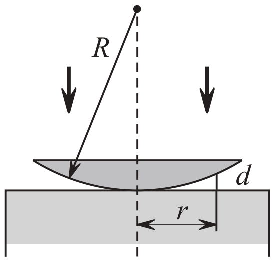

Figure P5.7. Newton rings.

► **Holography principles.** Holography is used to record the volume image of an object. The object is illuminated by laser light with a very high degree of coherence and after reflection goes to a photographic plate. The information about the object shape is carried by the dependence of the object wave phase on the coordinates along the photographic plate. If, simultaneously with light reflected from the object, the photographic plate is illuminated by a reference ray, the same as that used to illuminate the object but going directly to the plate after reflection from a mirror. Then, as the result of interference, a wave is generated whose amplitude and intensity will depend on the phase of the object wave. Since the photographic plate blackening is proportional to the intensity, it preserves the information about the wave phase. Illuminating the developed photographic plate by the light of the same laser, one can reconstruct the initial signal.

Recording the interference signal in the bulk of a thick transparent photographic plate, one obtains a *volume hologram*, which bears information about both the object shape and the laser signal wavelength. If the hologram is illuminated by white light, the waves of other frequencies cancel one another due to the interference, and the image of the object illuminated by the monochromatic laser light arises. If three holograms from lasers with different wavelengths are recorded in one photographic plate, then the volume colored image appears in illumination by white light.

► **Vavilov–Cherenkov radiation.** If a particle moves in a medium with velocity V exceeding the phase velocity $v = c/n$ of light in this medium, then the radiation of electromagnetic waves arises at a certain angle with the direction of the particle motion. The elementary explanation of this phenomenon is quite similar to the explanation of the appearance of shock wave in motion with supersonic speed. As a particle moves, the medium atoms excited by the particle begin to radiate. The path difference of waves radiated by excited atoms in the direction θ from positions A and B (Fig. P5.8) is equal to

$$\Delta = v\frac{L}{V} - L\cos\theta = L\left(\frac{v}{V} - \cos\theta\right).$$

If $V < v$, then for any angle θ, one can choose L such that the radiation of points A and B cancel each other ($\Delta = \lambda/2$). For $V > v$, the same concerns any angle other than

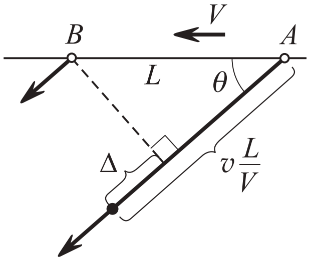

Figure P5.8. Vavilov–Cherenkov radiation.

$\theta_0 = \arccos(v/V)$ (one can attain $\Delta = \pm\lambda/2$). But in the direction θ_0, all the points radiate wave with the same phase; that is, the radiation is mutually reinforced.

P5.2.2. Diffraction

► **Huygens–Fresnel principle.** Deviations from the law of light rectilinear propagation is called diffraction. Approximate calculations of the light intensity behind obstacles or holes in the screen is based on the *Huygens–Fresnel principle* which supplements the geometric Huygens construction (see Subsection P5.1.1) with the condition of interference of secondary rays. According to this principle, all points of an arbitrary surface surrounding the wave sources are sources of coherent secondary waves, and the amplitude of the wave oscillations at an arbitrary point outside this surface can be obtained as the result of interference of secondary waves. Usually, for such a surface one takes the wave front, and then the sources of secondary waves are in phase. The secondary wave amplitude is proportional to the primary wave amplitude and to the area of the surface site:

$$dE = E_0 \frac{K(\alpha)}{r} \exp\bigl[\, i(\omega t - kr)\bigr]\, ds, \qquad \text{(P5.2.2.1)}$$

where E_0 is the (complex) amplitude of the primary wave at the front points, α is the angle between the radiation direction and the normal to the surface, r is the distance from the front of secondary sources to the point of observation, and $K(\alpha)$ is a slowly decreasing function ($K(\pi) = 0$).

We consider the diffraction of a plane monochromatic wave incident normally on a flat screen with a hole of various shape.

Circular hole of radius R_h. The point of observation B lies on the hole axis at a distance l from its center (Fig. P5.9). For the surface of secondary waves we take a part of the front in the plane of the hole. We divide the front into thin circular segments, by cutting circles from point B on it at intervals whose lengths are $l{+}\delta r$, $l{+}2\delta r, \ldots$ ($\delta r \ll \lambda$). Since the optical path length acquires the constant increment δr in transition from segment to segment, it follows that the oscillation phase increments are also the same; moreover, the amplitudes of oscillations from separate circular segments, which are calculated by formula (P5.2.2.1), also coincide. Indeed, from $r^2 = R^2 + l^2$ we obtain $R\delta R = r\delta r$ and $\delta S/r = 2\pi\delta r = \mathrm{const}$ ($\delta S = 2\pi R\delta R$). Hence the vector diagram of combined oscillations (Fig. P5.10) has the form of a slowly twisted spiral (with slow decrease in $K(\alpha)$ taken into account). From the symmetry consideration, one can conclude that the spiral converges to the center of the circle. The sum of the entire spiral corresponds to the free propagation of light in the absence of the screen.

The part of the surface from center to the points, where $r_1 = l + \lambda/2$, is called the *first Fresnel zone*, the part of the surface from the edge of the first zone to $r_2 = l + \lambda$ is the second Fresnel zone, etc. One can see that if only the first Fresnel zone is open, and all the others are closed by a screen, then the oscillation amplitude is twice greater than that without the screen, and the intensity is greater by the factor 4. If the first two zones are open, then the amplitude and intensity at point B are close to zero, etc. Conversely, if a small number of central Fresnel zones is closed by the screen, then this hardly affects the oscillation intensity at point B, it is almost the same as without the screen (this paradoxical fact is known as the "Poisson spot").

If an obstacle of complicated shape is prepared to close the next nearest Fresnel zones (all even zones or all odd zones), then the intensity increases many times at the point of observation. Such an instrument focusing the light like a lens is called a *zone plate* and finds numerous applications.

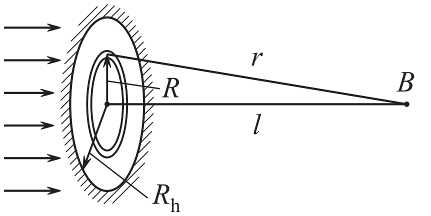

Figure P5.9. Wave front separation into thin circular segments.

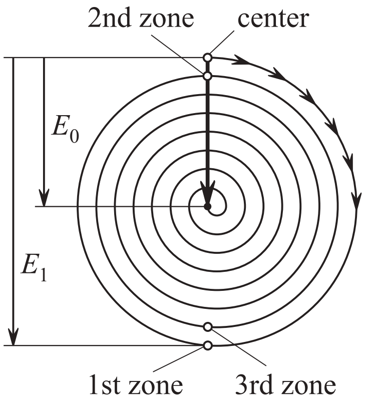

Figure P5.10. Vector diagram for the wave front separation into circular segments.

We find the radius of the mth Fresnel zone from the Pythagorean theorem $r^2 = R^2 + l^2$, substituting $r = l + \Delta$ (where $\Delta \ll l$ is the path difference between the hole edge and the hole center):

$$R(\Delta) = \sqrt{2l\Delta}. \tag{P5.2.2.2}$$

If $\Delta = m\lambda/2$ is substituted, we obtain the radius of the mth Fresnel zone: $R_m = \sqrt{ml\lambda}$.

We also present the formula for the case of incidence at the hole of a spherical wave radiated by a point source lying on the hole axis at the distance a from its center:

$$R_m = \sqrt{\frac{2al\Delta}{a+l}} = \sqrt{\frac{mal\lambda}{a+l}}.$$

As $a \to \infty$, this expression becomes (P5.2.2.2).

The number of Fresnel zones opened by the hole is equal to $m = R_\text{h}^2/l\lambda$. Hence for $l \ll R_\text{h}^2/\lambda$, there are many open Fresnel zones and the light intensity at the point of observation slightly differs from the intensity of the incident wave; that is, the diffraction effects at the hole axis are weak (the geometric optics approximation works). In the converse limit $l \gg R_\text{h}^2/\lambda$, only a small part of the first Fresnel zone is open; that is, all the oscillations at the point of observation are combined in one phase (the Fraunhofer approximation).

For a circular hole, it is more difficult to calculate the diffraction for off-axis points of observation. We will revisit this problem below considering it for the case of Fraunhofer diffraction. Unlike a circular hole, it turns out that the solution of the diffraction problem for all observation points is possible for a long slit-shaped hole.

Slit-shaped hole of width b. The point of observation B lies at a distance l from the screen. In this case, the front must be divided not into circular but into thin rectilinear segments (Fig. P5.11) of the same width δx. The main difference is that the segment areas are proportional not to $R\,\delta R$ but simply to δx. In this case, just as

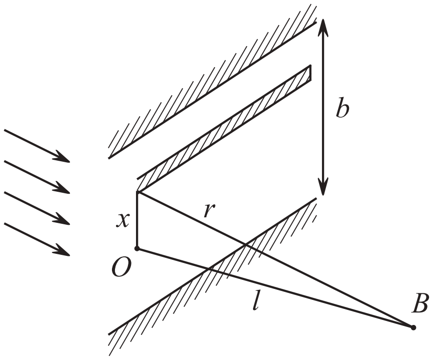

Figure P5.11. Wave front separation into thin rectilinear segments.

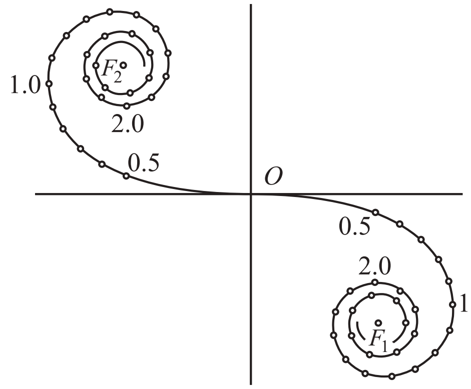

Figure P5.12. Vector diagram for the wave front separation into thin rectilinear segments (Cornu spiral).

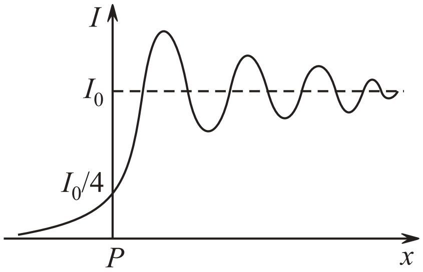

Figure P5.13. Intensity distribution on the screen in diffraction at the half-plane.

before, we have the identity $r^2 = x^2 + l^2$, with $x\,\delta x = r\,\delta r$, and hence $\delta r = x\,\delta s/r$, and the phase increment from segment to segment increases with the distance from point O. This results in that the radius of curvature of the vector diagram gradually decreases and it acquires the shape of spiral called the "Cornu spiral" (Fig. P5.12). The values of the parameter $v = x\sqrt{2/(l\lambda)}$ are marked on this spiral. For $x \ll l$, the path difference is equal to $\Delta = r - l \approx x^2/2l$, and hence $\Delta = v^2\frac{\lambda}{4}$ ($v = 1$ corresponds to $\Delta = \lambda/4$ and $v = 2$ corresponds to $\Delta = \lambda$; see Fig. P5.12). The distance between finite points F_1 and F_2 of the spiral is equal to the amplitude A_0 of oscillations without the obstacle, or to the amplitude of the incident wave. If the slit edges are at the distances x_1 and x_2 from point O, then we calculate the corresponding values v_1 and v_2, obtain the corresponding points on the spiral, and, connecting them, determine the amplitude. Thus, one can obtain the diffraction pattern at the slit not only at the screen center but also aside the center. The case $b \ll \sqrt{l\lambda}$ corresponds to the Fraunhofer diffraction (a small part of the spiral is open); the case $b \gg \sqrt{l\lambda}$ corresponds to the geometric optics approximation. In this case, near the slit edges, one observes the diffraction pattern corresponding to diffraction at the half-plane (Fig. P5.13); precisely opposite the slit edge (point P), the amplitude OF_2 is equal to half the amplitude of the incident wave A_0 (intensity is four times less than I_0), and the width of the diffraction spreading is $\sim \sqrt{l\lambda}$.

Using the Fresnel zone method, one can study the diffraction pattern at a slit graphically using the Cornu spiral obtained numerically. A simple analytic description of diffraction at a slit can be obtained in the Fraunhofer approximation.

► **Fraunhofer diffraction at a slit.** Fraunhofer diffraction is diffraction in parallel rays, where the phase difference of secondary waves radiated from different points of the wave front segment under study can be found under the assumption that they are parallel. In this approximation, the diffraction pattern becomes simpler, and one can calculate the light intensity at different points of the screen. The Fraunhofer approximation acts in the following two cases: (i) if the diffracted light rays are collected by a lens in its focal plane or (ii) if the screen is separated from the diffraction hole by a sufficiently large distance $r \gg R^2/\lambda$, where R is the source dimension (see the discussion of the Fraunhofer condition in Subsection P5.2.1, formula (P5.2.1.2).

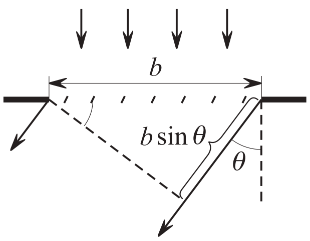

Figure P5.14. Fraunhofer diffraction at a thin slit.

In the Fraunhofer approximation, we calculate the diffraction of a plane monochromatic wave at a slit of width b. We consider only the normal incidence of a wave on a screen with a slit; in this case, all the slit points are sources of secondary in-phase waves. We divide the plane slit into a large number N of equal narrow strips (Fig. P5.14). The light amplitude from each of the strips is equal to E_0/N, where E_0 is the amplitude of light radiated by the slit in the direction $\theta = 0$ (in this case, all the strips radiate in-phase light). In the case of radiation in the direction at an angle θ with the normal, the radiation phase difference between the neighboring strips is equal to $\delta\varphi = (2\pi/\lambda)(b/N)\sin\theta$, the phase difference between the extreme strips is equal to $\varphi = (2\pi/\lambda)b\sin\theta = N\delta\varphi$. Using expression (P5.2.1.3) for the interference of N sources (Subsection P5.2.1), we obtain the amplitude for the slit radiation in the direction θ ($\sin\delta\varphi$ must be replaced by $\delta\varphi$):

$$E_1(\theta) = E_0\frac{\sin(\varphi/2)}{\varphi/2}, \qquad \text{where} \quad \varphi = 2\pi\frac{b}{\lambda}\sin\theta. \tag{P5.2.2.3}$$

The radiation intensity is equal to $I_1 = I_0\sin^2(\varphi/2)/(\varphi/2)^2$, the typical vector diagrams are shown in Fig. P5.15. The minima condition for the radiation has the form:

$$b\sin\theta = m\lambda \qquad (m = 0, \pm1, \pm2, \dots). \tag{P5.2.2.4}$$

the maxima condition is $b\sin\theta \approx (m+\frac{1}{2})\lambda$ $(m = 1, 2, \dots)$. The maxima of intensity are to one another as $I_0 : I_1 : I_2 : \dots \approx 1 : (1.5\pi)^2 : (2.5\pi)^2 : \cdots$. The maximum order of the spectrum is determined by the condition $m \le b/\lambda$.

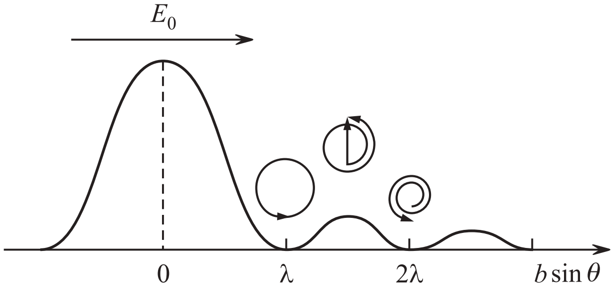

Figure P5.15. Angular distribution of radiation intensity and the vector diagrams for the Fraunhofer diffraction at a slit.

► **Fraunhofer diffraction at a circular hole.** In the case of diffraction at a circular hole, the intensity distribution has the form of concentric dark and light circles around the central light spot. The intensity distribution for small angles θ with hole axis is expressed in terms of the Bessel function of the first order: $I = I_0J_1^2(\pi D\theta/\lambda)$, where D is the hole diameter. In outward appearance, the angular distribution of the intensity slightly differs from the case of diffraction at a slit; the first minimum corresponds to the angle $\theta_1 = 1.22\,\lambda/D$.

The angle θ_1 can be treated as the angle of diffraction expansion of the light beam. The beam spreading is weak at distances l satisfying the condition: $l\theta_1 \ll D$ or $l \ll D^2/\lambda$. This inequality is a condition of the geometric optics applicability (see Subsection P5.1.1).

► **Rayleigh criterion. Resolving power of optical instrument.** When two nearby points are observed in an optical instrument (lens, telescope), their images on the screen are small diffraction circles. It is conditionally assumed that the image points are indistinguishable if the distance between them is so small that the central maximum of one of the circles coincides with the first minimum of the other circle (*Rayleigh criterion*). Hence the angular resolution of such optical instruments (with an open diaphragm of diameter D) is given by the formula

$$\theta_{\min} = 1.22\frac{\lambda}{D}.$$

This limit of an instrument possible resolution is imposed by the wave nature of light and cannot be overcome by any technical improvements.

► **Diffraction grating.** The diffraction grating is a sequence of large number N of equal parallel slits (Fig. P5.16). The width of each slit is denoted by b, and the distance between the corresponding points (for example, centers) of neighboring slits, which is called the *diffraction grating period*, is denoted by d ($d = l/N$, where l is the grating width). In the direction θ, each slit radiates light whose amplitude E_1 is determined by expression (P5.2.2.3). The phase difference between the neighboring slits is equal to $\delta = (2\pi/\lambda)d\sin\theta$, and to calculate the final amplitude, one can again use formula (P5.1.2.3) for the sum of oscillations of N independent sources:

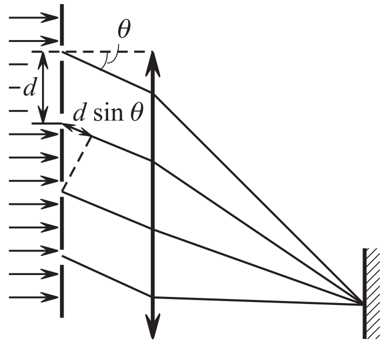

Figure P5.16. Diffraction grating.

$$E = E_1\frac{\sin(N\delta/2)}{\sin(\delta/2)}, \quad I = I_1\frac{\sin^2(N\delta/2)}{\sin^2(\delta/2)}, \quad \text{where} \quad \delta = 2\pi\frac{d}{\lambda}\sin\theta.$$

The *principal maxima of the diffraction grating* are determined by the condition

$$d\sin\theta = m\lambda. \tag{P5.2.2.5}$$

At the principal maxima, the phase difference between the neighboring slits equals $2\pi m$ and the oscillations from all slits are in phase. The vector diagram is a straight line of length NE_1. In the direction of the principal maxima, the radiated wave intensity $I = N^2 I_1$ is completely determined by the radiation intensity of a separate slit in this direction. If the principal maximum is near the minimum of $I_1(\theta)$ determined by the condition $b\sin\theta = m\lambda$, then this maximum is suppressed. The maximum order of the spectrum is determined by the condition $m \le d/\lambda$.

The diffraction pattern has $N-1$ minimum and $N-2$ *additional maxima* between each two principal maxima. The minima are determined by the conditions $d\sin\theta = (m+n/N)\lambda$ ($n = 1, 2, \ldots, N-1$). The first minimum is especially important because it determines the width of the principal maximum. In the vector diagram, the first minimum corresponds to the phase difference between the slits $2\pi/N$, and the vector diagram for N slits is a closed figure, namely, a regular N-gon. One can see that the width of the principal maxima decreases proportionally to $1/N$. The intensity of additional maxima is much less than that of the principal maximum, and it can be neglected in the case of a large number of slits.

If light is incident on the grating at an angle $\theta_0 \ne 0$, then the principal maxima condition takes the form:

$$d(\sin\theta - \sin\theta_0) = m\lambda. \qquad \text{(P5.2.2.6)}$$

For coarse gratings ($d \gg \lambda$), it is efficient to use grazing rays ($\theta_0 \approx \pi/2$), for which the principal maxima condition takes the form: $d\cos\theta_0(\theta-\theta_0) = m\lambda$. One can see that the role of the effective period in this case is played by the small quantity $d\cos\theta_0$.

► **Diffraction grating as a spectral device.** The properties of any spectral device are determined by its *angular dispersion*, *free spectral range*, and *resolving power*.

The angular dispersion is the derivative $D_m = d\theta_m/d\lambda$, where θ_m is the position of the principal maximum of mth order. It follows from the principal maximum condition that

$$D_m = \frac{m}{d\cos\theta}.$$

The free spectral range is the maximum width $\Delta\lambda$ of the spectral interval for which there still is no overlapping of the spectra of neighboring orders. From the equations $d\sin\theta = m(\lambda + \Delta\lambda)$ and $d\sin\theta = (m+1)\lambda$, we have $\Delta\lambda = \lambda/m$. One can see that to study wide parts of the spectrum, it is necessary to use lower orders.

The resolving power is the ratio $R = \lambda/\delta\lambda$, where $\delta\lambda$ is the least wavelength difference that can be resolved by the spectral device. Two lines are assumed to be resolved if the maximum of one line lies on the minimum of the other line (Rayleigh criterion). From the equations $d\sin\theta = (m + 1/N)\lambda$ and $d\sin\theta = m(\lambda + \delta\lambda)$ it follows that $\delta\lambda = \lambda/(mN)$; that is,

$$R = \frac{\lambda}{\delta\lambda} = mN.$$

► **Diffraction of X-rays at a crystal.** The wavelength of X-rays is comparable with the distance between the atoms of a crystal lattice, which is a *spatial diffraction grating* for incident rays. The condition of diffraction maxima consists in simultaneous satisfaction of three equations (P5.2.2.6) for the three mutually perpendicular crystal axes (*Laue conditions*):

$$d_1(\cos\alpha - \cos\alpha_0) = n_1\lambda, \quad d_2(\cos\beta - \cos\beta_0) = n_2\lambda, \quad d_3(\cos\gamma - \cos\gamma_0) = n_3\lambda,$$

where α, β, γ is angles with the axes such that $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$. These three equations cannot be satisfied simultaneously for an arbitrarily chosen direction of the incident ray and for a given wavelength. This means that in the case of the crystal irradiation by monochromatic but scattered X-rays (that is, in all possible directions), there arise diffraction maxima in quite certainly determined directions. An analysis of the *Laue patterns* thus obtained permits obtaining information about the crystal structure. The diffraction maxima are absent if $\lambda/2$ exceeds all the lattice periods; for visible light, the crystal can be treated as a homogeneous medium.

The reflection of X-rays from the crystal surface can also be considered as the interference of rays reflected from a system of successive atomic planes in the crystal. By analogy with the interference in thin films (see Subsection P5.2.1), the path difference between the rays reflected from the neighboring planes is equal to $2d\sin\theta$, where θ is the angle between the incident ray and the atomic plane (not the normal!), and d is the interplanar distance. (We note that the index of refraction of X-rays is almost equal to one; see Subsection P5.3.2.) The reflection is observed only in the directions of the diffraction maxima satisfying the *Wulf–Bragg condition*:

$$2d\sin\theta = m\lambda,$$

where $m = 1, 2, \ldots$ is the order of the diffraction maximum.

P5.3. Interaction Between Light and Matter

P5.3.1. Light Polarization. Fresnel Formulas

► **Polarized and natural light.** A plane wave is said to be *linearly polarized* or *plane polarized* if the vector **E** oscillates in one plane that is perpendicular to the wave front (it is called the *plane of polarization* of the wave). The monochromatic plane wave is either linearly polarized or elliptically polarized or circularly polarized (see Subsection P4.2.2). The elliptically polarized wave is the sum of two mutually perpendicular plane waves such that there is a phase difference between their oscillations. The *natural light* radiated by heated bodies is not polarized, because the direction of the vector **E** oscillations at each point varies rapidly and chaotically. A mixture of natural and polarized light is called *partially polarized* light.

The *polarizer* is a device at whose output the light is linearly polarized in a certain plane called the *transmission plane of polarizer*. The point is that the polarizer completely absorbs the light polarized perpendicularly to the transmission plane. If the natural light is transmitted through a polarizer, then it becomes linearly polarized and its intensity decreases twice (if there is no absorption in the transmission plane of the polarizer). If the linearly polarized light of intensity I_0 is transmitted through a polarizer whose transmission plane makes an angle α with the plane of oscillations of the light wave, then the intensity of the transmitted wave is equal to

$$I = I_0\cos^2\alpha$$

(*Malus's law*). This can be explained by the fact that the linearly polarized light with amplitude E_0 is the sum of two linearly polarized waves: the wave polarized in the transmission plane (its amplitude is equal to $E_0\cos\alpha$) propagates through the polarizer without changes and the second wave is completely absorbed.

► **Reflection and refraction of waves. Fresnel formulas.** The intensity and polarization of the reflected and refracted waves depend on the incident wave polarization. We write the boundary conditions on the interface between two media:

$$E^{\rm i}_\tau + E^{\rm r}_\tau = E^{\rm d}_\tau, \quad n_1^2(E^{\rm i}_n + E^{\rm r}_n) = n_2^2 E^{\rm d}_n, \quad H^{\rm i}_\tau + H^{\rm r}_\tau = H^{\rm d}_\tau, \quad H^{\rm i}_n + H^{\rm r}_n = H^{\rm d}_n.$$

The subscripts τ and n denote the tangential and normal components, and the superscripts i, r, and d corresponds to the incident, reflected, and refracted (deflected) waves, respectively. Moreover, it is necessary to take account of the relations between the electric and magnetic fields in the plane of the electromagnetic wave (see Subsection P4.2.2). For a plane monochromatic wave, we have

$$\mathbf{E}^{\rm i} = \mathbf{E}^{\rm i}_0 \exp\bigl[i(\omega t - \mathbf{k}^{\rm i}\cdot\mathbf{r})\bigr],$$
$$\mathbf{E}^{\rm r} = \mathbf{E}^{\rm r}_0 \exp\bigl[i(\omega t - \mathbf{k}^{\rm r}\cdot\mathbf{r})\bigr],$$
$$\mathbf{E}^{\rm d} = \mathbf{E}^{\rm d}_0 \exp\bigl[i(\omega t - \mathbf{k}^{\rm d}\cdot\mathbf{r})\bigr]$$

The wave vectors (Fig. P5.17) are related by

$$k^{\mathrm{i}}_{\|} = k^{\mathrm{r}}_{\|} = k^{\mathrm{d}}_{\|},$$

or

$$k_1 \sin\alpha = k_1 \sin\beta = k_2 \sin\psi,$$

where $k_1 = \omega n_1/c$ and $k_2 = \omega n_2/c$. We obtain the reflection law $\beta = \alpha$ and refraction law $n_1 \sin\alpha = n_2 \sin\psi$.

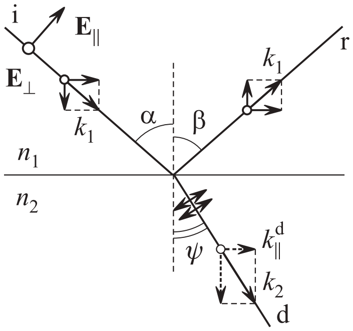

Figure P5.17. Decomposition in polarizations for incident, reflected, and refracted waves.

Special attention is drawn to the case where the light is incident from an optically denser medium ($n_1 > n_2$) at an angle α greater than the *limit angle* α_{c} *of total reflection* ($\sin\alpha_{\mathrm{c}} = n_2/n_1$). In this case, $k_1 \sin\alpha > k_2$, and $k^{\mathrm{d}}_{\perp}$ is imaginary: $k^{\mathrm{d}}_{\perp} = \sqrt{k_2^2 - (k^{\mathrm{d}}_{\|})^2} = i|k^{\mathrm{d}}_{\perp}|$, where $|k^{\mathrm{d}}_{\perp}| = \frac{\omega}{c}\sqrt{n_1^2 \sin^2\alpha - n_2^2}$. This means that the amplitude of the transmitted wave decays exponentially at the distance $\delta \sim 1/|k^{\mathrm{d}}_{\perp}|$, and the amplitude of the reflected wave is equal to the amplitude of the incident wave.

The amplitudes of the transmitted and reflected waves depend on the polarization of the incident wave. We present the result of the reflected waves:

$$E^{\mathrm{r}}_{\|} = -E^{\mathrm{i}}_{\|}\frac{\tan(\alpha - \psi)}{\tan(\alpha + \psi)}, \qquad E^{\mathrm{r}}_{\perp} = -E^{\mathrm{i}}_{\perp}\frac{\sin(\alpha - \psi)}{\sin(\alpha + \psi)}$$

(*Fresnel formulas*). Here the first formula concerns the wave polarized in the incidence plane (it is convenient to derive it from the boundary conditions for **E**), and the second formula concerns the wave polarized in the perpendicular plane (it is convenient to derive it from the boundary conditions for **H**). One can see that, for an angle of incidence satisfying the condition $\alpha + \psi = \pi/2$, the wave polarized in the incidence plane does not reflect at all. Since $\sin\psi = \cos\alpha$ in this case, it follows that the angle of incidence at which the reflected wave is linearly polarized perpendicularly to the incidence plane (the *Brewster angle*) satisfies the relation:

$$\tan\alpha = n.$$

The qualitative explanation is that, in this case, the direction of the dipole oscillations (shown in the figure), generated in the second media by the wave polarized in the incidence plane turns out to be parallel to the direction of the reflected wave (reflected and refracted rays are mutually perpendicular). But the oscillator does not radiate any wave in the direction of its oscillations (see Subsection P4.2.2).

In the case of normal incidence, there is no difference between polarizations:

$$E^{\mathrm{r}} = E^{\mathrm{i}}\frac{n_1 - n_2}{n_1 + n_2}.$$

One can see that in the case of reflection from an optically denser medium ($n_2 > n_1$), the oscillation phase becomes opposite (more precisely, π is added to the phase). By the way, this property follows directly from the Fresnel formulas: for $\alpha > \psi$, the oscillations change the sign.

The ratio of the reflected energy to the flux of the incident energy is called the *reflection index*. In the case of normal incidence, it is equal to

$$R = \frac{(n_1 - n_2)^2}{(n_1 + n_2)^2}. \tag{P5.3.1.1}$$

The transmission coefficient is equal to $D = 1 - R$ (the energy conservation law).

We note that if the index of reflection is equal to the ratio of the energy volume densities of the reflected and incident waves: $R = (E^{\rm r}/E^{\rm i})^2$, then, calculating the transmission coefficient, one must take account of both the difference in the wave velocities in different media and (for incidence at an angle) the variation in the transverse area due to refraction: $S^{\rm d}/S^{\rm i} = \cos\psi/\cos\alpha$ (Fig. P5.18). Finally,

$$D = \frac{v_2 (E^{\rm d})^2 \cos\psi}{v_1 (E^{\rm i})^2 \cos\alpha}.$$

(The energy flux is equal to the product of the energy volume density by the wave velocity and the area of the transverse cross-section.)

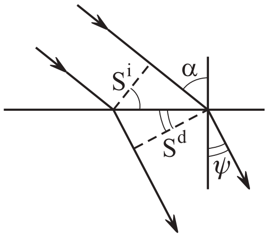

Figure P5.18. In refraction, the beam transverse cross-section is changed.

The coefficients R and D depend only on the *relative index of refraction* $n_{21} = n_2/n_1$ of the two media. The index of reflection is usually not large; for example, for $n_{21} = 1.5$, we obtain $R = 0.04$.

Example. Blooming of optical systems. The reflection index of glasses in optical instruments is not large (of several percents); nevertheless, it is an important problem to decrease the reflection for certain wavelengths. To this end, the surface is covered with a transparent film with the index of refraction $n' = \sqrt{n}$ (n is the index of refraction of glass) and of thickness $\lambda/(4n')$. The optical path difference of the rays reflected from the film surfaces is equal to $\lambda/2$ (the phase variation in reflection need not be taken into account, because it occurs for each of the rays), and the indices of reflection from these surfaces are close to each other (see formula (P5.3.1.1)). As a result, the light reflection is suppressed almost completely.

► **Optically anisotropic media.** When an electromagnetic wave travels through an anisotropic media, the vectors **E** and **D** are generally not parallel to each other. The linear relation between them is of tensor character, which means that each component of the vector **D** can be expressed as a linear combination of all three components of **E**. There exist three mutually perpendicular axes, called the *dielectric axes of the crystal*, for which **D** $\parallel$ **E**: $D_i = \varepsilon_0 \varepsilon_i E_i$ ($i = 1, 2, 3$). The values ε_i are called the *principal dielectric permittivities* of the crystal. We consider only the case of *uniaxial crystals* for which two of the three ε_i coincide ($\varepsilon = \varepsilon_\perp$). The distinguished axis ($\varepsilon = \varepsilon_\parallel$) is called the *optical axis of the crystal*.

If a plane wave propagates in a uniaxial crystal, then the *principal plane of the crystal* is introduced, which is the plane passing through the optical axis and the vector **n** normal to the wave front. It turns out that the propagation of a linearly polarized light wave depends on the direction of its polarization. The wave polarized perpendicularly to the principal plane is said to be *ordinary*. The speed of such a wave propagation $v_\perp = c/\sqrt{\varepsilon_\perp}$ is independent of the direction; the oscillations of the vectors **E** and **D** have the same directions; the direction of energy propagation (i.e., the Poynting vector $\mathbf{S}=\mathbf{E}\times\mathbf{H}$) is perpendicular to the wave front. The wave polarized parallel to principal plane is said to be *extraordinary*. The velocity of its propagation depends on the angle between **n** and the optical axis (for the angle $\pi/2$ between them, it is equal to $v_\parallel = c/\sqrt{\varepsilon_\parallel}$). The vectors **E** and **D** oscillate in different directions, and the Poynting vector $\mathbf{S} = \mathbf{E}\times\mathbf{H}$ is not perpendicular to the wave front (the normal to the wave front is parallel to $\mathbf{D}\times\mathbf{H}$). The difference between the ordinary and extraordinary rays disappears only if the light propagation is parallel to the optical axis.

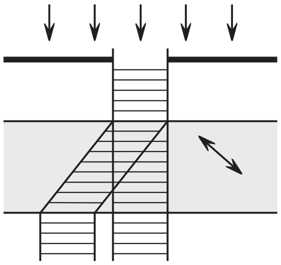

Figure P5.19. Ordinary and extraordinary rays in the case of light normal incidence on the crystal cut at an angle with the optical axis.

If the light is incident on the crystal surface, then it splits into ordinary and extraordinary rays that are linearly polarized perpendicular to each other and have different indices of refraction (*phenomenon of birefringence*). For an extraordinary ray, only the direction of the front propagation obeys the refraction law (see Subsection P5.1.1), but the ray itself can leave the incidence plane. Even in the case of the ray normal incidence on a crystal cut at an angle with the optical axis, the spatial splitting of rays occurs (Fig. P5.19). The fronts are indicated by dashes, the optical axis is indicated by arrow. The extraordinary ray is polarized in the plane of the drawing, the ordinary ray is polarized perpendicularly to it.

To obtain and analyze polarized light, one uses the *polarization prisms* (*nicols*) cut at an angle with the ray propagation direction so that the ordinary ray is totally reflected from the plane of the cut and propagates sidewise, while the extraordinary propagates straightforwardly. Another method for obtaining polarized light is based on the difference in absorption of ordinary and extraordinary rays in some matters (*dichroism phenomenon*). When the light propagates through a *dichroic plate* (tourmaline plate, polaroid), the ordinary ray is absorbed and the linearly polarized extraordinary ray is transmitted.

To analyze the light polarization character, one studies the intensity dependence on the nicol orientation. If the intensity does not vary, then the light is either natural or circularly polarized. To distinguish these cases, one uses a *quarter wave plate* or a *compensator*. The plate thickness d is chosen so that the difference between the ordinary and extraordinary rays $\Delta = \Delta n\, d$ is equal to $\lambda/4$. The phase shift between mutually perpendicular oscillations becomes either zero or π, and the circular polarization becomes the linear polarization.

► **Rotation of the polarization plane.** In the case of propagation of linearly polarized light in several matters (they are said to be *optically active*) the *rotation of polarization plane* can occur. The rotation angle is proportional to the plate thickness: $\chi = \alpha l$, where α is *rotation per unit length*. Depending on the rotation direction, there are right- and left-handed matters. An example is given by a quartz plate cut perpendicularly to the optical axis (quartz can be either levorotary or dextrorotary). In solutions of optically active matter in an inactive solvent, α is proportional to the concentration. The molecules of active matters are antisymmetric with respect to the right- and left-handed rotation of spiral type. The phenomenon of rotation of the polarization plane can be characterized as the circular birefringence. The wave circularly polarized in opposite directions propagates with different velocities; that is, the phase difference between them varies. The sum of two such oscillations is a linear oscillation whose direction depends on the phase difference.

► **Artificial anisotropy.** If many isotropic bodies are placed in a homogeneous electric field, they acquire uniaxial anisotropy with the optical axis oriented parallel to the electric intensity field (the *electro-optical Kerr effect*). The phase difference between the ordinary and extraordinary rays in light propagation perpendicularly to **E** is proportional to the squared electric intensity:

$$\Delta\varphi = 2\pi BlE^2,$$

where l is the thickness of the matter layer and B is called the *Kerr constant*. Artificial anisotropy arises in the cases where the matter molecule polarizability depends on their orientation with respect to the field. A similar effect arise in the case of several matters placed in a magnetic field (the *Cotton–Mouton effect*). It is described by the relation $\Delta\varphi = 2\pi ClB^2$.

In the case of inactive matters placed in a strong magnetic field, the optical activity may arise for light propagating parallel to the vector **B** (*magnetic rotation of polarization plane*). The rotation angle per unit length in this case (for diamagnetic and paramagnetic materials) is proportional to the value of the magnetic induction: $\alpha = RB$, where R is called the *Verdet constant*.

P5.3.2. Light Dispersion and Absorption

► **Classical model of dispersion medium.** When an electromagnetic wave propagates through matter, the charged particles of the medium come into forced oscillatory motion. The amplitude of these oscillations and their phase shift with respect to the wave field strength oscillations depend on the relation between the wave frequency ω and the frequency ω_0 of natural oscillations of particles (see Subsection P4.1.3). The resultant wave perturbation can be considered as the result of interference of the original wave and the waves radiated by the medium particles (such an approach is called the *molecular optics*). But in the case of homogeneous media, it is possible to obtain the wave frequency characteristics semi-phenomenologically taking account of the polarization arising due to the particle displacement, introducing the dielectric susceptibility and permittivity depending on the frequency, and calculating the index of refraction. The wave damping — that is, the transformation of the energy oscillations into thermal energy — is taken into account by introducing the semi-empirical attenuation of oscillators; then the dielectric permittivity and the index of refraction become complex numbers.

First, we consider the medium of equal oscillators. The equation of motion of a charged particle has the form

$$\ddot{\boldsymbol{\xi}} + 2\beta\dot{\boldsymbol{\xi}} + \omega_0^2\boldsymbol{\xi} = \frac{e}{m}\mathbf{E},$$

where **E** is the field acting on the particle (in the optical range, only the electrons are significant). In rare gases, one can neglect the difference between the local and average fields; that is, one can assume that the electrons are under direct action of the wave field $\mathbf{E} = \mathbf{E}_0 \exp\bigl[i(\omega t - \mathbf{k}\cdot\mathbf{r})\bigr]$. We seek the solution of the equation of motion in the form $\boldsymbol{\xi} = \boldsymbol{\xi}_0 \exp(i\omega t)$ and, after the substitution, we obtain

$$\boldsymbol{\xi} = \frac{e}{m(\omega_0^2 - \omega^2 + 2i\beta\omega)}\mathbf{E}$$

(in complex representation, the phase shift is automatically taken into account). The particle displacement results in the molecules acquiring the dipole moments $\mathbf{p} = e\boldsymbol{\xi}$, thus leading to the polarization $\mathbf{P} = N\mathbf{p}$ (N is the particle concentration). From the relation $\mathbf{P} = \varepsilon_0(\varepsilon - 1)\mathbf{E}$ we derive the complex dielectric permittivity

$$\varepsilon = 1 + \frac{Ne^2}{\varepsilon_0 m(\omega_0^2 - \omega^2 + 2i\beta\omega)}. \qquad \text{(P5.3.2.1)}$$

The index of refraction is also complex, $\sqrt{\varepsilon} = n - i\kappa$, as well as the wavenumber, $\widetilde{k} = \frac{\omega}{c}\widetilde{n} = \frac{\omega}{c}(n - i\kappa)$. Consequently, for the argument of the exponential function we have $i(\omega t - \widetilde{k}x) = i(\omega t - kx) - \kappa\frac{\omega}{c}x$, where $k = \frac{\omega n}{c} = \frac{\omega}{v}$, and the attenuation coefficient is equal to $\kappa\frac{\omega}{c}$. Hence,

$$E = E_0 \exp\Bigl(-\frac{\kappa\omega}{c}x\Bigr)\exp[i(\omega t - kx)]. \qquad \text{(P5.3.2.2)}$$

To find $n(\omega)$ and $\kappa(\omega)$, it is necessary to equate the real and imaginary parts in $\varepsilon = (n - i\kappa)^2$. Far from the natural frequency (for $|\omega - \omega_0| \gg \beta$), we obtain

$$\varepsilon = n^2 = 1 + \frac{Ne^2}{\varepsilon_0 m(\omega_0^2 - \omega^2)}.$$

The dependencies $n(\omega)$ and $\kappa(\omega)$ are qualitatively shown in Fig. P5.20 (n_0 is the value of n as $\omega \to 0$, which is called the static value). In the region where the absorption is not large, the index of refraction increases with frequency (*normal dispersion*). In the narrow region of strong absorption, one observes the *anomalous dispersion*.

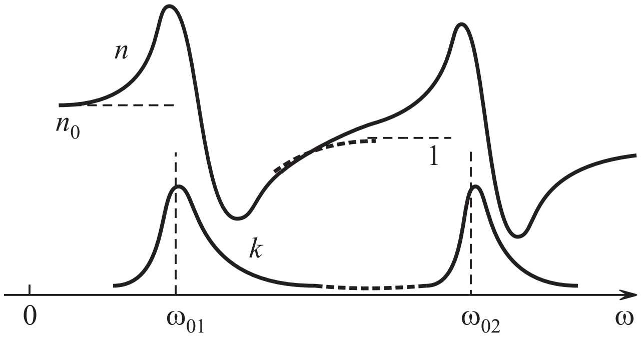

Figure P5.20. Dependence of the index of refraction (upper curve) and of the wave attenuation on the frequency.

A similar situation arises near each natural frequency (ω_{01} and ω_{02} in Fig. P5.20). For example, in the infrared region of the spectrum, the absorption and anomalous dispersion bands related to ion oscillations are observed. The absorption bands in the ultraviolet (sometimes, in the optical) regions of the spectrum can be explained by the electron oscillations on the atomic outer shells (the optical electrons). In the X-ray region of the spectrum, the wave frequency ω is large compared with all natural frequencies and the dependence $n(\omega)$ is determined by electron oscillations, which can be assumed to be free:

$$n^2 = 1 - \frac{Ne^2}{m\varepsilon_0\omega^2}. \qquad \text{(P5.3.2.3)}$$

The index of refraction of X-rays is slightly different from one. The same formula holds for a wave propagating in rarefied plasma containing free electrons.

The phase velocity of a wave in plasma (and also to the right from the absorption band in dielectrics) turns out to be greater than the light speed in a vacuum ($n < 1$). But this does not contradict the theory of relativity, because the wave group velocity $u = d\omega/dk$ (see Subsection P4.2.1) is less than c in this case. Let us verify this fact for a wave in plasma. Using the relation $k^2 = \omega^2 n^2(\omega)/c^2$ and equation (P5.3.2.3), we obtain:

$$c^2 k\,dk = \omega\,d\omega \quad\Longrightarrow\quad \frac{\omega}{k}\frac{d\omega}{dk} = c^2.$$

Hence, in this case, $u = nc < c$.

In the case of polar molecules (for example, of water), a wide band of anomalous dispersion lies in the region of centimeter radio waves, where the amplitude of rotational oscillations of dipoles tending to rotate following the field strength strongly depends on the frequency. Just in this region, $n = \sqrt{\varepsilon}$ decreases from a large static value (for water, $n_0 \approx 9$) to the high-frequency value (for water, $n \approx 1.3$).

Formula (P5.3.2.1) is only valid for n close to unity, or in the regions where the difference of the field acting on the molecule from the average field in the matter can be neglected. A generalization to the case of dense fluids is the *Lorentz–Lorentz formula*:

$$\frac{n^2 - 1}{n^2 + 2} = \frac{Ne^2}{3\varepsilon_0 m(\omega_0^2 - \omega^2)},$$

which follows from the Clausius–Mossotti formula (Subsection P3.5.2). As the matter density varies, the quantity

$$r = \frac{1}{\rho}\frac{n^2 - 1}{n^2 + 2},$$

which is called the *specific refraction*, must remain constant.

► **Light scattering. Wave attenuation.** The wave intensity decreases in a medium not only because of the light absorption, but also because of its scattering. The scattering can be explained by the light emission by atomic oscillators, which occurs in all directions (see Subsection P4.2.2). But in an ideal homogeneous medium, the light scattered at molecules located at the distance $\lambda/2$ from each other would experience the total interference damping, and no attenuation due to scattering would occur in this case. The scattering is observed at small foreign particle (*Tyndall scattering in turbid media*) and at inhomogeneities arising due to the density fluctuations (the *Rayleigh scattering*).

The intensity of light scattered at inhomogeneities whose dimensions are small compared with the wavelength is proportional to λ^{-4} (*Rayleigh law*, also see Subsection P4.2.2). This explains the blue color of the sky (the scattered sunlight) and the yellow-red color of the sun (the transmitted light). The degree of polarization of the scattered natural light depends on the angle of scattering; the light scattered at the angle $\pi/2$ is completely polarized. The qualitative explanation is that only the oscillators whose oscillation direction is perpendicular to the scattering direction radiate in this direction. The scattering at inhomogeneities large compared with the wavelength weakly depend on the frequency; this explains the white color of clouds.

The Rayleigh scattering at the density fluctuations or concentration depend on the temperature. As the critical point is approached, the average values of fluctuations increase sharply and one observes white turbidity of the fluid, which is called the *critical opalescence*.

The light beam attenuation in the case of not large intensity occurs by the exponential law (the *Bouger law*):

$$I = I_0 e^{-\alpha x},$$

where the *wave attenuation coefficient* α is equal to the sum of the absorption coefficient, expressed in terms of the imaginary part of the index of refraction (see formula (P5.3.2.2)), and the scattering coefficient, which describes the wave attenuation due to scattering.

P5.4. Foundations of Quantum Optics

P5.4.1. Thermal Radiation

► **Equilibrium thermal radiation.** The radiation of electromagnetic (radiant) energy by a body due to the energy of chaotic (thermal) motion of its molecules is called the *thermal radiation*. The properties of thermal radiation are determined by the body material and its temperature. If one makes a closed cavity of any material and maintains the temperature of its walls to be constant, then the system (wall + radiation) comes to the state of thermodynamical equilibrium, and the *equilibrium thermal radiation* is attained in the bulk of the cavity. The most important characteristic of equilibrium thermal radiation is that its properties are completely determined by the wall temperature and are independent of their material. This statement follows from the second law of thermodynamics. Moreover, the equilibrium thermal radiation is homogeneous and isotropic.

The main characteristics of both the radiation from a body surface and the radiation in volume were introduced in Subsection P5.1.2. The radiation from a surface is characterized by the radiance B and the radiant exitance R equal to the radiant energy radiated from unit surface per unit time in all directions (that is, in the solid angle 2π). The spectral decompositions of the radiant exitance r_λ, r_ω, r_ν are also introduced, for example, $R = \int_0^\infty r_\lambda\,d\lambda$; the quantities r are called the *radiation capacities of a body*. The volume radiation is characterized by the luminous flux intensity I and the volume density of radiant energy u, as well as by their spectral decompositions. In the case of isotropic radiation, they are related as $u = 4\pi I/c$. The illumination E is determined as the total radiant power through unit site from all directions (from the solid angle 2π); in the case of isotropic radiation,

$$E = \pi I = \tfrac{1}{4}cu.$$

The spectral densities of illumination E are denoted by e_λ, e_ν and e_ω. The recalculation from one spectral characteristic to another is discussed in Subsection P5.1.2.

► **Absorption capacity. Kirchhoff law.** The *absorption capacity* of a body is the fraction of incident radiant energy absorbed by the body for a narrow interval of wavelengths or frequencies:

$$a_\lambda = \frac{d\Phi_{\rm abs}}{d\Phi_{\rm i}}.$$

A body for which $a_\lambda = 1$ in the entire spectral interval is called a *black body*. A black body can be modeled by a closed cavity with a small hole; almost all the rays incident into the cavity through the hole are absorbed as the result of multiple reflections from the internal walls. A body with $a_\lambda = \text{const} < 1$ (usually, $a_\lambda < 0.99$) is called a *grey body*.

Since the equilibrium thermal radiation is in equilibrium with the surface, it follows that for any spectral interval, the quantity of absorbed radiant energy, which is equal to $e_\lambda a_\lambda\,d\lambda$, must be equal to the quantity of the radiated energy, which is equal to $r_\lambda\,d\lambda$. Since the characteristics of the equilibrium bulk radiation are independent of the properties of a specific body, it follows that the ratio of the radiation capacity of any body to its absorption capacity is a universal function of the wavelength and temperature (*Kirchhoff law*):

$$\frac{r_\lambda}{a_\lambda} = e_\lambda(T) = \frac{1}{4}cu_\lambda(T). \tag{P5.4.1.1}$$

Since, for a black body, the absorption capacity is equal to one, it follows that the function on the right-hand side is precisely the *radiation capacity of a black body*, which we denote by r^*_λ:

$$\frac{r_\lambda}{a_\lambda} = r^*_\lambda(T).$$

One can see that the radiation capacity of a black body and its radiant exitance are independent of the method of its production; they are related to the energy volume density as

$$r^*_\lambda(T) = \tfrac{1}{4}cu_\lambda(T), \qquad R^*(T) = \tfrac{1}{4}cu(T). \tag{P5.4.1.2}$$

At a fixed temperature, a black body has the largest radiation capacity and radiant exitance. For example, for a grey body with $a_\lambda = \text{const} = \alpha$, we have $R = \alpha R^*$. We note that since the equilibrium thermal radiation is isotropic, it follows that a black body is a Lambertian source (see Subsection P5.1.2).

► **Stefan–Boltzmann and Wien laws.** The radiation capacity of a black body at a given temperature tends to zero for small and large λ and attains the maximum value for a certain wavelength λ_m, which depends on the temperature. The area under the curve r^*_λ is equal to the radiant exitance $R^*(T)$. The application of general thermodynamical relations to the equilibrium thermal radiation in a cavity allows one to obtain several general relations for it. (The temperature of the equilibrium thermal radiation is assumed to be equal to the temperature of the walls.) The *Stefan–Boltzmann law* states that the radiant exitance of a black body is proportional to the fourth power of temperature:

$$R^* = \sigma T^4, \tag{P5.4.1.3}$$

where $\sigma = 5.67 \times 10^{-8}\ \mathrm{W\,m^{-2}\,K^{-4}}$ is the *Stefan–Boltzmann constant*.

To derive (P5.4.1.3), we use the expression for the pressure of isotropic radiation $p = u/3$ (see Subsection P2.5.1) and the formula $\left(\frac{\partial U}{\partial V}\right)_T = T\left(\frac{\partial p}{\partial T}\right)_V - p$, which is a consequence of the second law of thermodynamics (Subsection P2.3.1). Substituting $U = u(T)V$, we obtain the equation $4u = du/dT$, which implies $u \sim T^4$. Moreover, from the formula for pressure and from the first law of thermodynamics $\left(0 = d(uV) + \frac{1}{3}u\,dV\right)$, one can derive the adiabatic process equation for the equilibrium radiation: $uV^{4/3} = \text{const}$. Taking into account that $u \sim T^4$, we obtain $VT^3 = \text{const}$.

If we consider a slow adiabatic variation in a radiation volume contained in a vessel with mirror walls and apply the Doppler effect formula to the light reflection from a moving mirror (see Subsections P4.2.1 and P4.2.2), then we can prove the *Wien formula*:

$$r^*_\lambda = \lambda^{-5} f_1(\lambda T) \qquad \text{or} \qquad r^*_\lambda = T^5 f_2(\lambda T), \tag{P5.4.1.4}$$

where $f_i(x)$ are unknown functions whose form cannot be found in the framework of thermodynamics. Similar expressions for r^*_ω have the form

$$r^*_\omega = \omega^3 \varphi_1(\omega/T) \qquad \text{or} \qquad r^*_\omega = T^3 \varphi_2(\omega/T). \tag{P5.4.1.5}$$

From the Wien formula (P5.4.1.4) (or (P5.4.1.5)), one can derive the Stefan–Boltzmann law (P5.4.1.3). Moreover, these formulas imply the *Wien displacement law*, which presents the temperature dependence of the position of the maximum of the function r^*_λ (or r^*_ω):

$$\lambda_{\max} T = b \qquad (\omega_{\max}/T = b_1), \tag{P5.4.1.6}$$

where $b = 2.9 \times 10^{-3}\ \mathrm{m\,K}$ is the Wien constant. For example, as the temperature decreases by the factor 2, then the position of the maximum of the function r^*_ω (or u_ω) becomes twice closely to the origin, and the maximum itself is eight times below (Fig. P5.21); in this case, the area under the graph decreases by the factor 16.

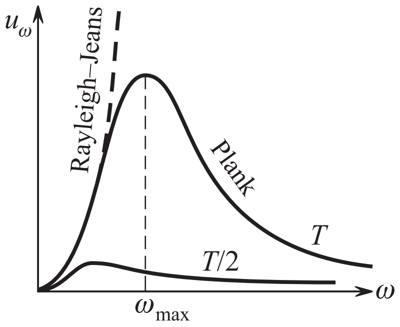

Figure P5.21. Frequency dependence of the energy volume density of the equilibrium thermal radiation.

► **Rayleigh–Jeans law.** Rayleigh and Jeans tried to obtain the form of the function u_ω in the framework of the classical statistical physics. They considered radiation in a cavity as an ensemble of standing electromagnetic waves randomly exchanging energy with the walls and between each other. From the viewpoint of statistics, each independent standing wave with a certain frequency of oscillations is equivalent to the oscillator with the same frequency. The energy calculation is reduced to the two independent problems:

1) What is the number dg of oscillators (standing waves) contained in the frequency interval $d\omega$? The answer must be expressed in the form of the function $G(\omega)$, which is called the *density of states*: $dg = VG(\omega)\,d\omega$, where V is the vessel volume.

To calculate $G(\omega)$, one can consider a vessel of the shape of a rectilinear parallelepiped with sides L_x, L_y, L_z. The boundary conditions (for example, the requirement that the nodes of standing waves lie on the boundaries) lead to the conditions $k_\xi L_\xi = m_\xi \pi$ ($\xi = x, y, z$). Therefore, in the space of wave vectors, the admissible states correspond to the nodes of the lattice with sides π/L_ξ and the cell volume $\delta = \pi^3/(L_x L_y L_z) = \pi^3/V$. The volume of the k-space, corresponding to the variation in the value of the wave vector from k to $k + dk$, is equal to $\frac{1}{8}(4\pi k^2\,dk)$ (the volume of the spherical layer cut by the first quadrant). Dividing by the cell volume, we obtain the number of spatially distinct oscillations in the interval dk: $dg = V\dfrac{k^2\,dk}{2\pi^2}$. It is also necessary to take account of the additional degrees of freedom (in the case of electromagnetic waves, there are two possible states of polarization), which we generally take into account by the additional factor γ.

Let us find the number of states per unit volume:

$$\frac{dg}{V} = \gamma\frac{k^2\,dk}{2\pi^2}. \qquad \text{(P5.4.1.7)}$$

This formula is obtained from the boundary conditions and has a very general character and numerous applications. To pass to ω, we must take the formula $k = \omega/c$ into account. Finally, we obtain ($\gamma = 2$)

$$G(\omega) = \frac{\omega^2}{\pi^2 c^3}. \qquad \text{(P5.4.1.8)}$$

2) What is the average energy of a single oscillator?

The classical physics gives the following answer (see Subsection P2.5.2): to each oscillator, independently of its frequency, one must assign two degrees of freedom and, according to the energy equipartition theorem, the average energy of each oscillator must be equal to kT, where k is the Boltzmann constant.

As a result of this reasoning, the classical physics gives the *Rayleigh–Jeans law*:

$$u_\omega = G(\omega)kT = \frac{\omega^2 kT}{\pi^2 c^3}. \qquad \text{(P5.4.1.9)}$$

Experiments show that the Rayleigh–Jeans law is well satisfied at small frequencies (for $\omega \ll \omega_m$) but is absolutely false at large frequencies (Fig. P5.21). Indeed, although (P5.4.1.9) satisfies the conditions imposed on any possible function u_ω by the Wien formula (P5.4.1.5), one can readily see that the function thus obtained does not have any maximum; it monotonically increases, and the integral $\int_0^\omega u_\omega d\omega$; that is, the total energy of radiation is equal to infinity! This situation is one of the symptoms of the deep crisis in the classical physics and the contemporaries called it the *ultraviolet catastrophe*.

► **Planck's law.** The energy *equipartition theorem* is a consequence of the fact that the oscillator classical energy is proportional to its squared amplitude and can take any even very small values. According to *Planck's quantization assumption*, the oscillator energy (counted from the minimum value) can take only discrete values multiple of a quantity depending on the oscillator frequency:

$$\varepsilon_n(\omega) = n\varepsilon_1(\omega), \qquad n = 0, 1, \dots \tag{P5.4.1.10}$$

To calculate the average energy, one can use the Maxwell–Boltzmann formula (see Subsection P2.5.4), according to which the probability of a state with energy ε is proportional to $\exp\bigl[-\varepsilon/(kT)\bigr]$. We obtain

$$\langle\varepsilon(\omega)\rangle = \frac{\sum\limits_{n=0}^{\infty} n\varepsilon_1 \exp\bigl[-n\varepsilon_1/(kT)\bigr]}{\sum\limits_{n=0}^{\infty} \exp\bigl[-n\varepsilon_1/(kT)\bigr]}.$$

The series in the denominator is simply the sum of a geometric progression, and the numerator is obtained from the denominator by differentiation with respect to $1/(kT)$. After calculations, we obtain

$$\langle\varepsilon(\omega)\rangle = \varepsilon_1\left[\exp\left(\frac{\varepsilon_1}{kT}\right) - 1\right]^{-1}, \quad u_\omega = G(\omega)\langle\varepsilon(\omega)\rangle = \frac{\omega^2\langle\varepsilon(\omega)\rangle}{\pi^2 c^3}. \tag{P5.4.1.11}$$

Comparing with Wien formula (P5.4.1.5), we see that $\varepsilon_1 = \varepsilon_1(\omega)$ must be proportional to ω:

$$\varepsilon_1(\omega) = \hbar\omega, \tag{P5.4.1.12}$$

where $\hbar = h/(2\pi) = 1.05\times10^{-34}$ J s is a universal constant. The constant h is called *Planck's constant* and $\hbar$ is the *reduced Planck constant*, often called "h-bar"; it is convenient to use the former constant in the case of frequency, and the latter, in the case of cyclic frequency: $\varepsilon_1 = h\nu = \hbar\omega$. Since the quantum energy is proportional to the oscillator frequency, it follows, that at a given temperature, the high-frequency oscillations are excited with a very small probability and their contribution to the energy radiation energy is negligibly small. This solves the problem of ultraviolet catastrophe.

After substitution of (P5.4.1.12) into the formula for the average energy, we obtain *Planck's law* for the spectral density of energy:

$$u_\omega = \frac{\hbar\omega^3}{\pi^2 c^3}\left[\exp\left(\frac{\hbar\omega}{kT}\right) - 1\right]^{-1}. \tag{P5.4.1.13}$$

The graph of this function is given in Fig. P5.21. We write Planck's law also in the variables ν and λ:

$$u_\nu = \frac{8\pi h\nu^3}{c^3}\left[\exp\left(\frac{h\nu}{kT}\right) - 1\right]^{-1}, \quad u_\lambda = \frac{8\pi hc}{\lambda^5}\left[\exp\left(\frac{hc}{\lambda kT}\right) - 1\right]^{-1}. \tag{P5.4.1.14}$$

For $\hbar\omega \ll kT$, Planck's law becomes the Rayleigh–Jeans formula (P5.4.1.9). From Planck's law we can obtain the expressions for the Stefan–Boltzmann and Wien constants in terms of the universal constants:

$$\sigma = \frac{\pi^2 k^4}{60\, c^2 \hbar^3}, \qquad b \approx 0.20 \frac{hc}{k}.$$

Planck's law agrees well with experiments in the entire frequency range.

P5.4.2. Light Quanta

► **Photoeffect.** The *external photoelectric effect* (*photoeffect*) is the effect of light-induced emission of electrons from a matter. To study the photoeffect, one employs a vacuum tube with cold cathode (in this case, the thermoelectron emission can be neglected). Radiating the cathode by light of fixed frequency and intensity, one takes the volt-ampere characteristic of the tube (dependence of the current on the anode voltage). The volt-ampere characteristic (Fig. P5.22) permits finding: (a) the number of electrons emitted from the cathode per unit time (it is expressed in terms of the saturation current: $N = I_{\rm sat}/e$) and (b) the maximum kinetic energy of emitted electron; it is expressed in terms of the *stopping voltage*, the anode voltage at which the current is zero:

$$\tfrac{1}{2} m v^2 = e|U_{\rm s}|.$$

At this voltage, even the fastest electrons cannot reach the anode.

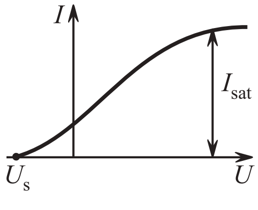

Figure P5.22. Dependence of the photoelectron current on the tube voltage.

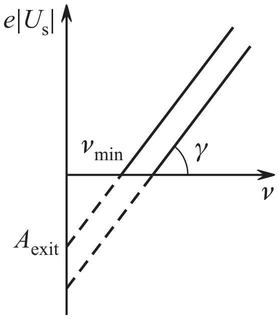

Figure P5.23. The maximum energy of photoelectrons versus the light frequency for different metals.

The first law of photoeffect: the number of electrons emitted from metal by light per unit time is directly proportional to the light wave intensity.

The second law of photoeffect: the maximum kinetic energy of photoelectrons linearly increases with light frequency and is independent of light intensity. If the light frequency is less than a certain minimum frequency $\nu_{\rm min}$ determined for a given material, then the photoeffect is not observed. It was experimentally found that the dependence of $e|U_{\rm s}|$ on ν for a given metal has the form of an inclined straight line, and the inclination of the lines constructed for different metals turned out to be the same (Fig. P5.23). The classical wave theory of light could not explain the second law of photoeffect. Moreover, this theory cannot explain the fact that the *photoeffect is inertialess*; that is, absolutely no delay arises between the beginning of radiation and the appearance of current.

► **Light quanta.** The photoeffect laws were explained by Einstein. He used Planck's quantization assumption (Subsection P5.4.1), but proceeded much further assuming that the light energy quanta are completely absorbed by separate electrons. This means that in the process of absorption, light behaves like a localized particle (it was called *photon*) with energy

$$E = h\nu = \hbar\omega = hc/\lambda. \qquad \text{(P5.4.2.1)}$$

Just as any massless particle moving with the speed of light, a *light quantum* (a *photon*) has the momentum

$$p = E/c = h\nu/c = h/\lambda. \qquad \text{(P5.4.2.2)}$$

The relation between the energy and momentum of a massless particle is given by the theory of relativity (see Subsection P1.10.3).

The quantum properties of light manifest themselves in light radiation, absorption, and scattering. Its wave properties manifest themselves in the phenomena related to light propagation. The light is of dual nature (the wave-particle duality). The same properties are exhibited by all elementary particles.

The *photoeffect* (an act of photoeffect) is the absorption of photon by a particle, for example, by electron. As a result of photoeffect, the light quantum disappears, and the electron acquires additional energy. If a photoelectron is emitted from a matter, then the external photoeffect is observed; if it remains inside, then the *internal photoeffect* takes place. In the case of internal photoeffect, electrons can move from a bound state in a free state, which increases the number of current carriers and hence decreases the resistance.

Photoeffect is used to construct photoelements, photorelay, etc.

Example 1. Can photoeffect occur on a free electron?

Solution. The answer is negative, because, in this case, the energy and momentum conservation laws cannot be satisfied simultaneously. This becomes obvious if we pass to the inertial reference system in which the electron after the photoeffect is at rest. Before the photoeffect, the system contained a light quantum and a moving electron, and after the photoeffect, it contains only an immovable electron; that is, the energy is not conserved.

Absorbing a light quantum, the electron acquires the energy $h\nu$. After escape from a metal, the energy of each electron decreases by a certain value, which is called the *photoelectric work function* A_{exit} (the work required to remove an electron from metal; the photoelectric work function depends on the kind of the matter). The maximum energy of electrons after escape (if there are no other losses) has the form

$$\tfrac{1}{2}mv^2 = h\nu - A_{\text{exit}} \qquad \text{(P5.4.2.3)}$$

(*Einstein equation*). If $h\nu < A_{\text{exit}}$, then the external photoeffect does not occur. Hence the threshold frequency is equal to

$$\nu_{\min} = A_{\text{exit}}/h.$$

It follows from (P5.4.2.3) that the inclination of lines in the graph of the dependence of $e|U_{\text{s}}|$ on ν (Fig. P5.23) is equal to h and the segment cut by the straight line on the ordinate axis is equal to the photoelectric work.

The photon energy and the photoelectric work are usually expressed in *electronvolts* (eV). One electronvolt, which is a non-SI unit, equals the energy acquired by an electron as it accelerates through a potential difference of –1 V: 1 eV = 1.6022×10^{-19} J. If, for example, the stopping voltage is –3.5 V, then the maximum kinetic energy of electrons is equal to 3.5 eV.

Example 2. The existence of the photoelectric work function means that, on the metal boundary, there arise forces retaining the electron inside the metal. How can the electron attraction to an electrically neutral metal be explained?

Solution. A charge particle is attracted by charges of opposite sign induced on the conductor surface. The attraction force is calculated by using the method of electrostatic images (Subsection P3.4.1). The escaping and returning electrons form a charge cloud near the surface, and the charges on the metal surface form a positively charged layer. Between the charged layers, there exists a nonzero average field intensity directed outwards.

► **Boundary of the X-ray spectrum.** If electrons are accelerated in a vacuum tube whose electrodes are charges at several kilovolts, then on impact of electrons on anode, the effect of *slowing down the X-ray radiation* arises. The study of the spectrum of this radiation shows that it does not contain the wavelengths less than a certain value $\lambda_{\rm c}$, which is inversely proportional to the voltage applied to the tube. This fact finds a natural explanation in quantum optics. The radiated photon energy cannot exceed the energy electron kinetic energy: $hc/\lambda \le eU$, which implies

$$\lambda \ge \lambda_{\rm c} = \frac{hc}{eU}.$$

The theoretical value of the coefficient of proportionality between $\lambda_{\rm c}$ and $1/U$ is in good agreement with experiments.

► **Light pressure.** The light pressure was predicted by Maxwell on the basis of the electromagnetic theory and was measured by Lebedev. Lebedev's installation consisted of a light rod suspended in a vacuum on a thin filament. Two thin plates were fixed at the ends of the rod, one of them was a reflecting plate, the other was an absorbing plate. Illuminating the plate and measuring the filament twisting, he calculated the light pressure.

The electromagnetic theory gave the following explanation of the light pressure: the electric field of an electromagnetic wave induces current in the metal and this current is under the action of the Ampère force from the magnetic field of the wave; this force has the same direction as the wave propagation and is the cause of light pressure. The explanation of pressure in the language of light quanta looks much simpler: the photons with the momentum (P5.4.2.2) are absorbed or reflected transferring their momentum to the matter. In the case of photon reflection, the transferred momentum is twice larger than that in the case of absorption (also see Subsection P4.2.2).

► **Compton effect.** As a photon interacts with a free electron, the process of photon absorption is forbidden by conservation laws, but the photon scattering may occur. If the electron was originally at rest, then, as a result of interaction, it acquires a certain velocity. The energy conservation laws require that the photon energy decrease by the value of the electron kinetic energy, which means that its frequency must also decrease. At the same time, from the viewpoint of the wave theory, the frequency of scattered light must coincide with the frequency of incident light. This phenomenon is called the *Compton effect*, it was discovered in X-ray scattering and played an important role in the formation of quantum theory.

The photon scattering on an electron can be considered as an elastic collision of two particles obeying the energy and momentum conservation laws:

$$pc + mc^2 = p'c + E'_e, \qquad \mathbf{p} = \mathbf{p}' + \mathbf{p}'_e,$$

where $\mathbf{p}$ and $\mathbf{p}'$ is the initial and terminal photon momenta, $\mathbf{p}'_e$ and E'_e are the momentum and energy of the recoil electron (Fig. P5.24). We express the energy and momentum of the recoil electron and substitute them into the relation $E'^2_e - \mathbf{p}'^2_e c^2 = m^2c^4$ (see Subsection P1.10.3). After transformations, we obtain

$$mc(p - p') = pp'(1 - \cos\theta),$$

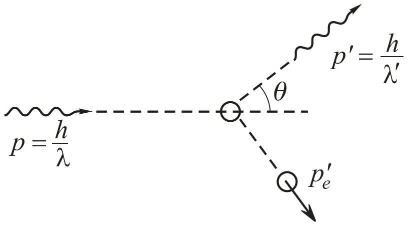

Figure P5.24. Photon scattering on an electron at rest (Compton effect).

where θ is angle of the photon scattering (the angle between the vectors $\mathbf{p}$ and $\mathbf{p}'$). We express the momenta of incident and scattered photon from equation (P5.4.2.2): $p = h/\lambda$, $p' = h/\lambda'$, and obtain the formula for dependence of the wavelength increment on the angle of scattering:

$$\lambda' - \lambda = \lambda_{\rm C}(1 - \cos\theta).$$

The quantity $\lambda_{\rm C} = h/mc = 2.43 \times 10^{-12}$ m is called the electron *Compton wavelength*. The energy of photon with wavelength $\lambda_{\rm C}$ is equal to the energy of electron at rest mc^2. The maximum effect corresponds to the photon scattering at the angle $\theta = \pi$.

► **Number of photons in equilibrium thermal radiation.** Planck's formula (P5.4.1.13) for the density of energy of equilibrium thermal radiation can be written in the language of light quanta as follows:

$$u_\omega = \hbar\omega G(\omega) N_\omega,$$

where $\hbar\omega$ is the energy of a quantum, $G(\omega) = \omega^2/\pi^2 c^3$ is the density of states, and

$$N_\omega = \frac{1}{e^{\hbar\omega/kT} - 1} \qquad \text{(P5.4.2.4)}$$

has the meaning of the number of photons in the state with a certain frequency ω.

P5.4.3. Principle of Laser Operation

► **Spontaneous and forced transitions. Einstein coefficients.** We consider an atom in the field of equilibrium radiation. If the atom is in the state with energy E_i, then, under the action of radiation, it can move into a state with larger energy E_j absorbing a light quantum with frequency $\omega = (E_j - E_i)/\hbar$ (Fig. P5.25). Such a process is called the *forced (induced) absorption*. The probability of such a transition per unit time is proportional to the density of energy radiation at the frequency ω:

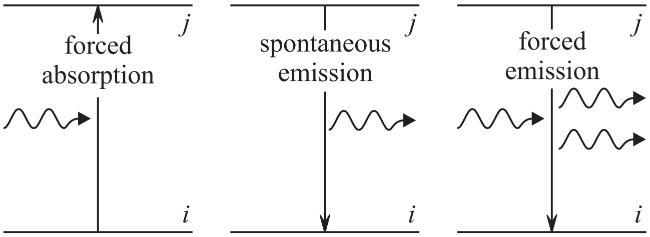

Figure P5.25. Schematic representation of processes of absorption and radiation of light quanta by atoms.

$$P_{i\to j} = B_{ij} u_\omega.$$

The reverse transition from a state with larger energy to a state with lesser energy is related to creation and emission of a quantum of frequency ω. It can occur in two ways. The first type of transitions is called the *spontaneous emission*, it occurs even in the absence of external radiation (for $u_\omega = 0$), and its probability is independent of u_ω:

$$P^{\rm sp}_{j\to i} = A_{ji}.$$

The second type of transitions is called the *forced emission*, it occurs under the action of a quantum of frequency ω. The transition probability per unit time, just as in the case of forced absorption, is proportional to u_ω:

$$P^{\rm ind}_{j\to i} = B_{ji}u_\omega.$$

The coefficients B_{ij}, B_{ji}, and A_{ji} thus derived are called *Einstein coefficients*, and they are independent of the radiation temperature.

► **Equilibrium between atoms and radiation.** We assume that a volume filled with equilibrium radiation at temperature T contains equal atoms. We also assume that the thermal equilibrium is established between atoms and radiation, and in this case the average number of atoms in state i is equal to N_i and that in state j is equal to N_j. These numbers at temperature T are related to each other by the *Boltzmann relation*:

$$N_j = N_i e^{-(E_j-E_i)/kT}.$$

In equilibrium, the number of transitions from i to j must be equal to the number of reverse transitions:

$$N_i B_{ij} u_\omega = N_j (A_{ji} + B_{ji} u_\omega). \tag{P5.4.3.1}$$

It follows from the Boltzmann relation that $N_i/N_j \to 1$ as $T \to \infty$. Taking into account that then $u_\omega \to \infty$, we obtain $B_{ij} = B_{ji}$.

Let us express u_ω from (P5.4.3.1):

$$u_\omega = \frac{A_{ji}/B_{ji}}{e^{\hbar\omega/kT} - 1} \tag{P5.4.3.2}$$

(we took into account that $N_i/N_j = \exp(\hbar\omega/kT)$). One can see that, in fact, we derived Planck's law for equilibrium radiation. The ratio A_{ji}/B_{ji} can be obtained from the condition that the classical Rayleigh–Jeans law must hold at high temperatures:

$$\frac{A_{ji}}{B_{ji}} = \frac{\hbar\omega^3}{\pi^2 c^3}.$$

We look at this relation from a different standpoint. Taking formula (P5.4.2.4) for the number of quanta with a certain frequency into account, we write (P5.4.3.2) as

$$\frac{B_{ji}u_\omega}{A_{ji}} = N_\omega.$$

Hence the ratio of the probability of forced radiation to the probability of spontaneous emission is equal to the number of quanta of frequency ω. In other words, the probability of forced emission calculated per one quantum is equal to the probability of spontaneous emission. One can see that the spontaneous and forced emissions must be of the same nature. The (qualitative) explanation is that even in the absence of light quanta in the system, the electromagnetic field has the so-called *zero-point oscillations* of a given frequency (the oscillator energy in quantum mechanics is equal to $\hbar\omega(N + \frac{1}{2})$). Precisely these oscillations cause spontaneous transitions.

► **Inverse population of levels.** In each act of forced emission, a new quantum is formed, and the created quantum has the same phase as the quantum generating this radiation. In other words, instead of one quantum, two *coherent quanta* appear in the system. Each of these quanta can in turn cause the appearance of a new coherent quantum, etc. Thus, coherent radiation may be amplified. But the newly created quanta are absorbed by atoms that are in the lower of the two energy states. In equilibrium, the number of atoms in the lower state (the population of the lower level) is higher than that in the upper level. Hence absorption prevails over the forced emission, and amplification is impossible.

To make a coherent ray amplification possible, it is necessary to create a metastable state with inverse population of levels. Then amplification of the quantum flux J occurs by the formula

$$dJ = \sigma(n_j - n_i)J\,dx, \qquad \text{or} \qquad J(x) = J(0)e^{\gamma x}, \tag{P5.4.3.3}$$

where n_j and n_i are the atom concentrations, σ is the cross-section of forced absorption or emission, and $\gamma = \sigma(n_j - n_i)$ is the *quantum amplification coefficient* of the medium.

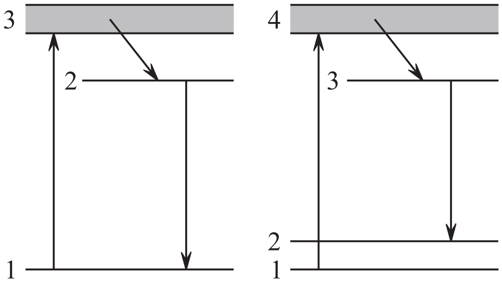

Figure P5.26. Three- and four-level schemes of laser operation.

The inverse population is created by using different methods of pumping. The most widely spread are the three- and four-level schemes (Fig. P5.26). In the three-level scheme (ruby laser), there is intense pumping from the lower level 1 to the wide short-living level 3 from which the long-living (metastable) level 2 is filled. It is necessary to ensure that the population of the upper operation level 2 exceeds the population of the lower operation level 1. The difficulty is that the initial population of level 2 is much less than the population of level 1, and intensive pumping is required. The four-level scheme does not have this drawback. In this scheme, the lower operation level 2 first contains as few atoms as level 3. The filling of level 3 occurs from level 4 which is pumped from the ground level 1. The neodymium laser operates according to this scheme.

► **Construction elements of lasers.** To use the phenomenon of quantum amplification of coherent light for creating the *quantum generator*, it is necessary to ensure positive feedback in order that a beam amplify itself. To this end, mirrors are arranged at the ends of the sample (or a mirror coating is applied). One of the mirrors is partially transparent, which ensures that part of light escapes from the system as a laser beam. The initial beam amplifies itself by multiple reflections from the two mirrors if the following condition is satisfied:

$$2Ln = m\lambda,$$

where n is index of refraction, L is the distance between the mirrors (the sample length), and m is an integer. To ensure coherent light generation, the total amplification coefficient on the entire closed path must be greater than 1:

$$R_1R_2e^{2\gamma L} > 1,$$

where γ is the quantum amplification coefficient (see (P5.4.3.3)) and R_1 and R_2 are the indices of reflection of the mirrors.

Bibliography for Chapter P5

Alenitsyn, A. G., Butikov, E. I., Kondratyev, A. S., *Concise Handbook of Mathematics and Physics*, CRC Press, Boca Raton, Florida, 1998.

Benenson, W., Harris, J. W., Stocker, and Lutz, H. (Eds.), *Handbook of Physics*, Springer-Verlag, New York, 2002.

Bennett, C. A., *Principles of Physical Optics*, John Wiley & Sons, New York, 2008.

Born, M. and Wolf, E., *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light (7th Edition)*, Cambridge University Press, Cambridge, England, 1999.

Fowles, G. R., *Introduction to Modern Optics, 2nd edition*, Dover Publications, New York, 1989.

Fox, M., *Quantum Optics: An Introduction (Oxford Master Series in Physics, 6)*, Oxford University Press, New York, 2006.

Hecht, E., *Optics, 4th Edition*, Addison Wesley, Reading, Massachusetts, 2001.

Irodov, I. E., *Wave Processes. Basic Laws* [in Russian], Fizmatlit, Moscow, 2004.

Landsberg, G. S., *Optics, 6th Edition* [in Russian], Fizmatlit, Moscow, 2006.

Pedrotti, F. L., Pedrotti, L. M. and Pedrotti, L. S., *Introduction to Optics, 3rd Edition*, Benjamin Cummings, New Jersey, 2006.

Prokhorov, A. M. (Ed.), *Encyclopedia of Physics, Volumes 1–5* [in Russian], Bolshaya Russkaya Encyclopedia, Moscow, 1998.

Prokhorov, A. M. (Ed.), *Encyclopedic Dictionary of Physics* [in Russian], Sovetskaya Encyclopedia, Moscow, 1984 (dic.academic.ru).

Saleh, B. E. A. and Teich, M. C., *Fundamentals of Photonics, 2nd edition (Wiley Series in Pure and Applied Optics)*, Wiley-Interscience, New York, 2007.

Saslow, W. M., *Electricity, Magnetism, and Light*, Academic Press, New York, 2002.

Sivukhin, D. V., *General Physics, Volume 4, Optics* [in Russian], Fizmatlit, Moscow, 2006.

Sivukhin, D. V., Yakovlev, I. A., *A Collection of Problems, Volume 4, Optics* [in Russian], Fizmatlit, Moscow, 2006.

Woan, G., *The Cambridge Handbook of Physics Formulas*, Cambridge University Press, Cambridge, England, 2003.

Chapter P6

Quantum Mechanics. Atomic Physics

By the beginning of the twentieth century, classical physics had accumulated many contradictions that could only be resolved by changing the main concepts of space and time (Einstein's theory of relativity, Chapter P1) and the laws of motion of particles on atomic scales (quantum mechanics). The first quantum concepts appeared at the very beginning of the century (Planck's formula for thermal radiation and Einstein's formula for photoeffect, Chapter P5), but quantum mechanics as a closed science was not constructed until the end of the 1920s.

P6.1. Atomic Structure. Bohr Model

P6.1.1. Nuclear Atomic Model

► **Experimental antecedents.** The atomic theory must, first of all, explain the accumulated experimental facts concerning the spectra of radiation of rarefied gases (that is, of separate atoms).

The atomic spectra consist of separate very narrow lines (*line spectra*) whose location is typical of each element. It was noted that any frequency is the difference of some two *terms*: $\nu_{mn} = T_m - T_n$, so that the element can be characterized not by a set of frequencies but by a set of terms. For hydrogen, we have $T_n = R_\nu / n^2$, where R_ν is the *Rydberg constant* (for frequency). The spectral lines for which one of the terms is fixed are called a *spectral series*. For example, in the hydrogen atom spectrum, the Lyman series ($n = 1$), the Balmer series ($n = 2$), etc. are distinguished.

But at the beginning of the twentieth century, there were no clear concepts about the atomic structure, the massive nucleus and electronic orbits. The *Thomson model*, in which the mass and the positive charge are uniformly distributed over the atomic volume and the electrons are contained inside this positive cloud as raisin in a pudding, seemed to be more attractive.

► **Rutherford experiments.** The nuclear atomic model was constructed in *Rutherford's experiments* on scattering of alpha particles at the atoms of thin goldfoil (performed by his collaborators Geiger and Marsden in 1911). It was found that some particles deviate by larger angles and even fly in the opposite direction. The number of such particles was small (approximately one of 8000 particles was scattered by an angle larger than 90°), but the mere fact of the existence of such particles contradicted the Thomson model.

Example 1. Estimate the maximum angle of deviation of alpha particles by the atom in the Thomson model.

Solution. To obtain this estimate, we assume that, in the Thomson model, the maximum deviation is exhibited by the particles flying near the edge of the atom (a ball of radius $R \sim 10^{-10}$ m), where the electric field strength is maximum (see example 3, Subsection P3.3.2). The formula for the angle of the particle deviation by the Coulomb field was derived in Subsection P1.6.2 (example 2):

$$\tan\frac{\theta}{2} = \frac{A}{2E_{\rm k} b}, \qquad \text{(P6.1.1.1)}$$

where $A = 2Ze^2/4\pi\varepsilon_0$ ($Z = 79$), $E_\text{k} \sim 5$–7 MeV is the kinetic energy of alpha particles, and b is the impact parameter. Taking $b \approx R \sim 10^{-10}$ m for the Thomson model, and see that the maximum deviation of alpha particles is $\sim 0.01^\circ$. If the multiple scattering is taken into account, then a greater angle of deviation is obtained, but for the film thickness $d \sim 10^{-7}$–10^{-6} m, the probability of scattering even by several degrees is negligibly small.

Rutherford showed that the experimental results are in good agreement with the formula for the angular distribution of α-particles scattered at a point Coulomb center with charge Ze.

Example 2. Rutherford's formula. We consider a thin target whose unit area contains $N_0 = nd$ nuclei (n is the bulk concentration of nuclei, d is the target width). Assuming that each particle interacts only with a single nucleus, we see that the particles scattered in the angle interval $(\theta, \theta + d\theta)$ are the particles whose impact parameter with respect to any nucleus lies in the interval $[b(\theta), b(\theta + d\theta)]$. The fraction of particles scattered in this angle is equal to the total area of such annuli around all the nuclei contained in the unit area of the target:

$$\frac{dN}{N} = N_0 2\pi b\, db = N_0 2\pi b \frac{db}{d\theta}\, d\theta.$$

Taking into account that the solid angle is equal to $d\Omega = 2\pi \sin\theta d\theta$, we obtain

$$\frac{dN}{N} = N_0 b \frac{db}{d\theta} \frac{d\Omega}{\sin\theta}.$$

Expressing b from formula (P6.1.1.1) for the scattering of alpha particles at the Coulomb center Ze, we obtain the *Rutherford formula*:

$$\frac{dN}{N} = \left(\frac{Ze^2}{8\pi\varepsilon_0 E_\text{k}}\right)^2 \frac{N_0\, d\Omega}{\sin^4(\theta/2)}. \qquad \text{(P6.1.1.2)}$$

Numerous experiments confirmed that this formula is in good agreement with experimental data. (We note that the coefficient at $N_0 d\Omega$ is called the *differential scattering cross-section*; see Subsection P8.2.2 for details.)

► **Atomic structure.** An atom with atomic number Z consists of a heavy nucleus with charge $+Ze$ and electrons in the field of this nucleus. An atom whose number of electrons is not equal to its atomic number Z is called an *ion*. The minimum energy required to remove an electron and form an ion is called the *ionization energy*.

The nuclear model is in unresolvable contradiction with the fundamental concepts of classical physics. The electrons are attracted to the nucleus, and hence they cannot be at rest but must move around the nucleus (therefore, the nuclear model is sometimes called the *planetary model*). But in accelerated motion, the electron radiates electromagnetic waves, and within a time interval of the order of 10^{-10} s, it must lose its entire energy and fall on the nucleus (see example 2, Subsection P4.2.2). Within the framework of classical physics, the atom cannot be stable.

P6.1.2. Bohr Model

► **Bohr postulates.** Bohr's semiclassical model describes the properties of the hydrogen atom and hydrogen-like ions (one electron in the field of the nucleus Ze). The model is based on the *Bohr postulates*:

1) The electrons can move only in certain stationary orbits according to the classical equations of motion but without energy radiation.

2) In this case, the angular momentum can take only discrete values $L_n = n\hbar$. For circular orbits, the quantization rule takes the form

$$mv_n r_n = n\hbar. \qquad \text{(P6.1.2.1)}$$

The number n is called the *principal quantum number*.

3) The radiation or absorption of quanta of electromagnetic radiation occurs only in transition from one stationary orbit to another stationary orbit:

$$\hbar\omega = E_n - E_m. \qquad \text{(P6.1.2.2)}$$

► **Hydrogen atom.** For the motion of an electron in a circular orbit, from the quantization rule (P6.1.2.1) and the equation of motion

$$\frac{ke^2}{r_n^2} = m\frac{v_n^2}{r_n}, \tag{P6.1.2.3}$$

we find the electron velocity in the nth orbit and its radius:

$$v_n = \frac{ke^2}{n\hbar}, \quad r_n = \frac{n^2\hbar^2}{mke^2},$$

and the electron energy

$$E_n = \frac{mv_n^2}{2} - \frac{ke^2}{r_n} = -\frac{ke^2}{2r_n} = -\frac{R}{n^2}, \tag{P6.1.2.4}$$

where

$$R = \frac{m(ke^2)^2}{2\hbar^2} \approx 13.6\ \text{eV} \tag{P6.1.2.5}$$

is the Rydberg constant (for the electron energy). The state with number $n = 1$ and energy $E_1 = -R = -13.6$ eV is called the *ground state of the atom*, the other values correspond to the excited states. The lifetime of excited states is of the order of 10^{-8} s. The radius of the orbit of electron in the ground state in the hydrogen atom is called the *Bohr radius* and is denoted by a_0:

$$a_0 = r_1 = \frac{\hbar^2}{mke^2} \approx 0.53\ \text{Å}.$$

The *energy of ionization* of the hydrogen atom is equal to $|E_1| = R$.

The radiation spectrum of the hydrogen atom is:

$$\frac{1}{\lambda} = R_\lambda\left(\frac{1}{n_2^2} - \frac{1}{n_1^2}\right), \tag{P6.1.2.6}$$

where $R_\lambda = \dfrac{m(ke^2)^2}{4\pi c\hbar^3} = 1.097 \times 10^7\ \text{m}^{-1}$ is the *Rydberg constant* (for the reciprocal wavelength), n_1 is the number of the initial state, n_2 is the number of the final state of electron (Fig. P6.1). The hydrogen lines with $n_2 = 1$ and $n_1 = 2, 3, \ldots$ form the Lyman series (ultraviolet region), the lines with $n_2 = 2$ and $n_1 = 3, 4, \ldots$ form the Balmer (visible region), the lines with $n_2 = 3$ form the Paschen series (infrared region), the further series are Brackett, Pfund, etc. series.

► **Hydrogen-like ions. Taking account of the nucleus motion**. The formulas obtained for the hydrogen atom can easily be generalized to the case of a hydrogen-like ion, or the case of one electron in the field of a charge Ze. It is simply required to replace e^2 by Ze^2 in all formulas. In this case, the Rydberg constant is multiplied by Z^2. For example, in the case of helium one-electron ion He$^+$, the electron energy takes the values $E_n = -4R/n^2$. Obviously, the energy with number $2n$ coincides with the electron energy in the hydrogen atom with number n. This means that the transition, for example, from the fourth level to the second level must correspond to the same frequency as the transition from the second level in the hydrogen atom to the first level. An analysis of the spectra shows that all the corresponding levels of the helium ion lie below the levels of the hydrogen atom

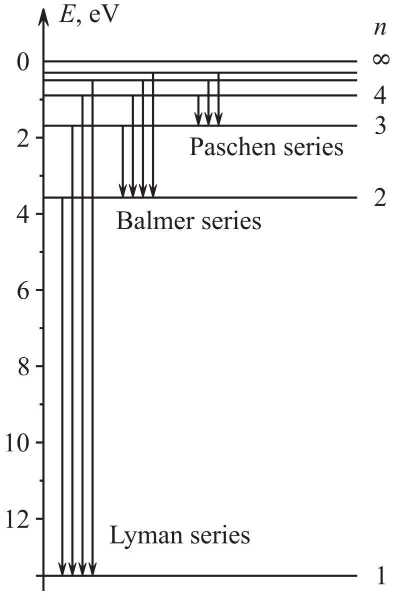

Figure P6.1. Hydrogen atom energy levels and spectral series.

approximately by 0.04 %. This difference can be explained by the fact that the nucleus mass is finite: both the electron and nucleus move about their common center of mass. The simplest way to take the nucleus motion into account is to replace the electron mass m by the reduced mass μ (see Subsection P1.6.4). This results in multiplication of all values of the energy and frequencies by the factor $M/(m+M)$, where M is the nucleus mass. The largest increase in the energy is for the hydrogen atom.

By the same reason, the energy levels of electrons in heavier isotopes lie below (*isotopic effect*). For example, the deuterium levels lie a little below the hydrogen levels, and the tritium levels lie a little below the deuterium levels.

► **The Bohr–Sommerfeld quantization condition.** A generalization of the orbit quantization rule is the *Bohr–Sommerfeld quantization condition*:

$$\oint p\,dq = 2\pi n\hbar, \qquad \text{(P6.1.2.7)}$$

where q and p are the generalized coordinate and momentum. If for the generalized coordinates of motion in a central field we take the angle φ and the angular momentum L, then we obtain the Bohr quantization rule. In the case of one-dimensional motion, the quantization condition (P6.1.2.7) takes the form

$$2\int_{x_1}^{x_2}\sqrt{2m(E_n - U(x))}\,dx = 2\pi n\hbar.$$

where x_1 and x_2 are the turning points, at which the momentum is zero. For example, in the case of oscillator, $U(x) = m\omega^2x^2/2$, and calculating the integral, we obtain

$$E_n = n\hbar\omega,$$

which up to the term $\hbar\omega/2$ (zero-point oscillations) coincides with the exact answer given by quantum mechanics.

The Bohr theory is not a consistent quantum theory. It does not describe the fine structure of hydrogen-like atoms and the properties of many-electron atoms, starting from the helium atom.

P6.2. Elements of Quantum Mechanics

P6.2.1. Wave Function. Uncertainty Relation

▶ **The corpuscular-wave duality. De Broglie waves.** A moving particle of matter can exhibit wave-like properties; also a wave can exhibit particle-like properties. This behavior is known as *corpuscular-wave duality* or *wave-particle duality*. The corresponding wavelengths, wave vector, and cyclic frequency are given by the *de Broglie formulas*:

$$\lambda = \frac{h}{p}, \quad k = \frac{p}{\hbar}, \quad \omega = \frac{E}{\hbar}, \tag{P6.2.1.1}$$

where p is the particle linear momentum and E is its relativistic energy.

An important argument for de Broglie waves was the fact that, with definition (P6.2.1.1) taken into account, the Bohr quantization rule (formula (6.1.2.1)) takes the form of the existence condition for the closed standing wave: $2\pi r_n = n\lambda$ (an integer number of de Broglie waves can be put on the orbit length).

The phase velocity of the de Broglie wave of a free particle moving with velocity v is equal to c^2/v, its group velocity is equal to v. (The phase velocity is not an observable quantity.)

▶ **Wave function.** A particle is associated with a complex *wave function* $\psi(\mathbf{r})$, which has a statistical meaning. The squared modulus of the wave function is equal to the density of probability of finding a particle at a given point in space:

$$dP = |\psi(\mathbf{r})|^2\, dV. \tag{P6.2.1.2}$$

The interference and diffraction phenomena in propagation of particles and in their detection arise due to *superposition principle*: if states ψ_1 and ψ_2 are realized, then the superposition $\psi = c_1\psi_1 + c_2\psi_2$ is also realized. The wave properties of particles were first discovered in experiments on diffraction of electrons at crystals in 1927. (*Davison–Germer experiments*).

▶ **Heisenberg uncertainty relation.** The measurement of a particle coordinates and projections of its momentum is of statistical character. The uncertainties in measurements of these quantities are related as

$$\Delta p_x \Delta x \ge \tfrac{1}{2}\hbar, \quad \Delta p_y \Delta y \ge \tfrac{1}{2}\hbar, \quad \Delta p_z \Delta z \ge \tfrac{1}{2}\hbar. \tag{P6.2.1.3}$$

The momentum and coordinates cannot simultaneously take exact values, and the notion of classical trajectory becomes meaningless.

The electron state with a definite momentum p_x is determined by the wave function in the plane wave form:

$$\psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \exp\Bigl(\frac{i}{\hbar} p_x\, x\Bigr).$$

In this case, the electron is completely delocalized in space. To estimate the uncertainty in measuring the momentum for an electron with localized wave function $\psi(x)$, it is necessary to expand this function in the Fourier integral in wave functions with a definite momentum:

$$\psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int \exp\Bigl(\frac{i}{\hbar} p_x\, x\Bigr) \varphi(p_x)\, dp_x.$$

The probability of that the result of measuring of p_x lies in the interval $(p_x, p_x + dp_x)$ is determined by the squared Fourier transform $|\varphi(p_x)|^2 dp_x$. According to general properties of the Fourier transformation, the function width and its Fourier transform are related as $\Delta x\, \Delta(p_x/\hbar) \ge \frac{1}{2}$ (also see Subsection P5.2.1, discussion of the temporal coherence).

The uncertainty relation for time and energy has the form:

$$\Delta E \Delta t \ge \hbar, \tag{P6.2.1.4}$$

where ΔE is the state energy uncertainty and Δt is the particle lifetime in this state.

Example 1. Impossibility of electron localization in the nucleus.

If we assume that the electron is contained inside a nucleus of dimension $r \sim 10^{-15}$ m, then the uncertainty of its coordinate is of the order of the nucleus dimensions and the uncertainty of the momentum is $\Delta p \sim \hbar/r$. Assuming that the average momentum is of the order of its uncertainty, we obtain the following estimate for the energy: $E = p^2/2m \sim 10^{-8} J \approx 10^5$ MeV, which exceeds the electron binding energy in the atom by many orders of magnitude.

Example 2. Estimate of the ground state energy of the hydrogen atom. Assuming that the coordinate uncertainty is of the order of the atom average dimension and the momentum uncertainty is of the order of the average momentum, we obtain the following relation between momentum and radius:

$$p \sim \frac{\hbar}{r}.$$

Substituting $p(r)$ into expression (P6.1.2.4) for the electron energy, we obtain the energy as a function of one parameter, the orbit radius:

$$E \sim \frac{\hbar^2}{2mr^2} - k\frac{e^2}{r}.$$

This expression takes the minimum value equal (up to a dimensionless factor) to the ground state energy for r equal to the Bohr radius.

► **Applicability criterion for the classical description.** The notion of trajectory can be used only if the coordinate uncertainty Δx is small compared with the typical dimensions R of the motion region and the momentum uncertainty Δp is small compared to the momentum p. In other words, the inequality $pR \gg \hbar$ must be satisfied. In addition to this criterion, one can compare $\hbar$ with other combinations of the system parameters that are of dimension of angular momentum: (momentum × length) or (energy × time).

P6.2.2. The Schrödinger Equation

► **Time-dependent and time-independent Schrödinger equations.** The evolution of the wave function of a particle in the force field $U(\mathbf{r})$ is described by the *Schrödinger equation*:

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\Delta\psi + U\psi. \qquad \text{(P6.2.2.1)}$$

In the state with a certain energy E (*stationary state*), the wave function has the form

$$\psi(\mathbf{r}, t) = \psi(\mathbf{r})\exp\bigl(-i\frac{E}{\hbar}t\bigr). \qquad \text{(P6.2.2.2)}$$

The spatial part of the wave function of stationary state satisfies the *time-independent Schrödinger equation*:

$$\Delta\psi + \frac{2m}{\hbar^2}(E - U)\psi = 0. \qquad \text{(P6.2.2.3)}$$

The wave function of a free particle ($U = 0$) in the state with a certain momentum $\mathbf{p}$ has the form

$$\psi(\mathbf{r}, t) = \frac{1}{(2\pi\hbar)^{3/2}}\exp\bigl[-\frac{i}{\hbar}(Et - \mathbf{p}\,\mathbf{r})\bigr].$$

► **The probability flux density** is determined by the expression

$$\mathbf{S} = -\frac{i\hbar}{2m}\bigl[\psi^*\nabla\psi - \psi\nabla\psi^*\bigr] \qquad \text{(P6.2.2.4)}$$

and in the case of a free particle with momentum $\mathbf{p} = m\mathbf{v}$ is equal to $\mathbf{S} = \mathbf{v}|\psi(\mathbf{r})|^2$. The variation in the probability of finding a particle in a certain region is determined by the probability flux through the boundary of this region.

► **Properties of stationary states.** The solutions of the time-independent Schrödinger equation (P6.2.2.3) describing the *bound states* of a particle in a potential field must satisfy the following conditions:

1) The function and its first derivatives must be continuous.

2) The function must be normed: $\int |\psi|^2\,dV = 1$. These two conditions can be satisfied only for definite values of energy $E_1, E_2, \ldots$ (*discreteness* of energy levels).

3) In the one-dimensional case, the solutions are real and *nondegenerate*; that is, to each energy value E_n there corresponds one function $\psi_n(x)$. The function ψ_n is zero at $n-1$ points (*nodes* of the wave function). If the potential energy is an even function ($U(-x) = U(x)$), then the wave functions with odd numbers (including the wave function of the ground state) are even ($\psi_{2n+1}(-x) = \psi_{2n+1}(x)$), and the wave functions with even numbers are odd ($\psi_{2n}(-x) = -\psi_{2n}(x)$).

The average value of the physical quantity $g(\mathbf{r})$, which is a function of coordinates, is calculated by the formula:

$$\langle g\rangle = \int g|\psi|^2\,dV,$$

where $\psi(\mathbf{r})$ is the normed wave function.

Example 1. Well with infinite walls. If the potential energy has the form

$$U = \begin{cases} 0, & \text{for } 0 \le x \le L; \\ \infty, & \text{for } x < 0 \text{ or } x > L, \end{cases}$$

then the wave function vanishes at edges of the well, at $x = 0$ and $x = L$. The Schrödinger equation (P6.2.2.3) takes the form

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}E\psi = 0$$

and has the solutions

$$\psi_n(x) = \sqrt{\frac{2}{L}}\sin k_n x, \quad k_n = \frac{n\pi}{L}.$$

The well width is equal to an integer number of de Broglie half-waves. The energy of the nth level is equal to $E_n = \dfrac{\hbar^2 k_n^2}{2m} = \dfrac{n^2\pi^2\hbar^2}{2mL^2}$. One can see that the wave functions with odd numbers are symmetric with respect to the well center, and the wave functions with even numbers are antisymmetric.

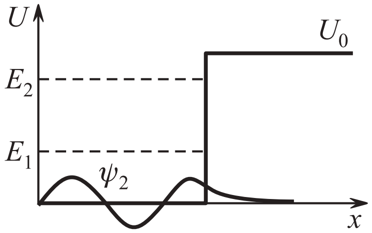

Figure P6.2. The wave function and energy levels in a one-dimensional potential well.

Example 2. Well with finite walls. If the potential energy has the form (Fig. P6.2)

$$U = \begin{cases} 0 & \text{for } 0 \le x \le L; \\ U_0 > 0 & \text{for } x > L; \\ \infty & \text{for } x < 0, \end{cases}$$

then the Schrödinger equation inside the well has the same form as in the preceding example and outside the well, for $x > L$, takes the form

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}(E - U_0)\psi = 0.$$

The localized (bound) states must have energies in the interval $0 < E < U_0$. The wave function vanishes for $x = 0$ and is equal to $\psi_1(x) = A\sin(k_1x)$ inside the well and to $\psi_2(x) = B\exp(-k_2x) + C\exp(k_2x)$ outside the well, where $k_1 = \sqrt{2mE}/\hbar$, $k_2 = \sqrt{2m(U_0 - E)}/\hbar$. It follows from the finiteness and integrability condition for the wave function that $C = 0$. Eliminating the coefficients A and B from the conditions of continuity, $\psi_1(L) = \psi_2(L)$, and smoothness, $\left.\dfrac{d\psi_1}{dx}\right|_{x=L} = \left.\dfrac{d\psi_2}{dx}\right|_{x=L}$, of the wave function, we obtain the equation for the allowed values of the energy: $k_2 = -k_1\cot k_1L$.

The roots must be sought graphically, as points of intersection of the graph $k_2L = -k_1L\cot k_1L$ with the circle $(k_1L)^2 + (k_2L)^2 = 2mU_0L^2/\hbar^2$ in the region $k_1L \ge 0$, $k_2L \ge 0$ (Fig. P6.3). The new bound states arise under the condition $E = U_0$ with $k_1L = \pi(n + 1/2)$; for $U_0L^2 < \pi^2\hbar^2/(8m)$ (a very "narrow" or very "shallow" well), there are no bound states.

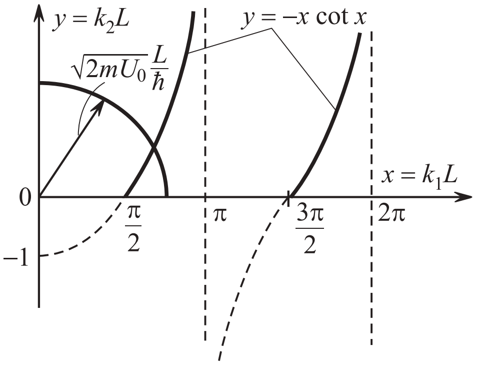

Figure P6.3. Graphical solution of the equations in example 2.

We note that the wave function of a bound state is nonzero for $x > L$ (Fig. P6.2). According to the classical concepts, the particle penetration under the potential barrier is impossible, because for $E < U_0$, the kinetic energy in this region is negative. In quantum mechanics, the probability of finding a particle under the barrier is nonzero, although it decreases exponentially with the distance from the barrier.

Example 3. Harmonic oscillator. The potential energy of harmonic oscillator is $U(x) = m\omega^2x^2/2$. The energies of stationary states are equal to $E_n = \big(n + 1/2\big)\,\hbar\omega$, and the wave functions of stationary states are expressed in terms of special functions, Hermitian polynomials (see Subsection M13.10.4). The energy of the ground state is equal to $\hbar\omega/2$ (the energy of zero-point oscillations).

Passing to dimensionless variables by the change $E = \dfrac{\hbar\omega}{2}\varepsilon$, $x = \sqrt{\dfrac{\hbar}{m\omega}}\xi$, we obtain the Schrödinger equation for the oscillator in the form

$$\frac{d^2\psi}{d\xi^2} + (\varepsilon - \xi^2)\psi = 0.$$

As $\xi \to \infty$, the term ε in parentheses can be neglected, and we obtain the asymptotic for large ξ behavior of the wave function: $\psi \sim \exp(-\xi^2/2)$. We seek the wave function in the form $\psi = \exp(-\xi^2/2)\varphi(\xi)$, and for φ, we obtain the equation

$$\varphi'' - 2\xi\varphi' + (\varepsilon - 1)\varphi = 0.$$

Representing $\varphi(\xi)$ as a series, $\varphi = a_0 + a_1\xi + \cdots$, and substituting it into the equation, we obtain a series of recursive relations: $2a_2 + (\varepsilon - 1)a_0 = 0$, $6a_3 + (\varepsilon - 3)a_1 = 0$, $\ldots$, $(n + 1)(n + 2)a_{n+2} + (\varepsilon - 2n - 1)a_n = 0$, $\ldots$. The wave function $\psi(\xi)$ satisfies the finiteness condition only if the series for $\varphi(\xi)$ contains finitely many terms (otherwise, as $\xi \to \infty$, $\varphi(\xi)$ increases as $\exp(\xi^2)$). The series can be truncated under the condition that it contains either only odd powers (for this, we must set $a_0 = 0$) or only even powers ($a_1 = 0$). The series truncation condition permits obtaining allowed values of the energy: $\varepsilon_n = 2n + 1$ or $E_n = \hbar\omega(n + 1/2)$ $(n = 0, 1, \ldots)$.

Just as in the preceding example, the wave functions of the oscillator stationary states are nonzero in the region, where the oscillator energy is less than its potential energy. The probability of finding a particle in the forbidden region decreases exponentially in the direction inside this region.

We note that in the ground state, the oscillator energy is equal not to zero but to $\hbar\omega/2$; this energy is called the *energy of zero-point oscillations*. The existence of such an energy is inevitable from the viewpoint of the uncertainty relation: a particle cannot be at the lowest point of the well and have zero momentum.

Just as any other quantum system, the oscillator can radiate or absorb photons in transitions from level to level. An analysis of probabilities of such transitions shows that they can occur only between neighboring levels; in other words, the transitions between the oscillator levels obey the *rule of selection*: $\Delta n = \pm 1$. This means that the oscillator radiates photons only of a single frequency ω.

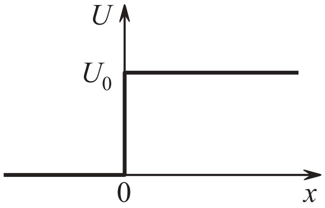

Figure P6.4. A step-like potential barrier (example 4).

Example 4. Reflection from a barrier. We show how to calculate the probability of reflection of the incident particles from an obstacle (barrier). We consider a barrier of the simplest shape, which is like a step of height U_0 (Fig. P6.4):

$$U = \begin{cases} 0 & \text{for } x \le 0; \\ U_0 & \text{for } x > 0. \end{cases}$$

We write the Schrödinger equation as (see example 22)

$$\psi'' + k_1^2\psi = 0 \qquad \text{for } x < 0;$$
$$\psi'' + k_2^2\psi = 0 \qquad \text{for } x > 0.$$

where $k_1 = \sqrt{2mE}/\hbar$, $k_2 = \sqrt{2m(E - U_0)}/\hbar$. The solution of this equation must describe the incident and reflected waves for $x < 0$ and the transmitted wave for $x > 0$:

$$\psi(x) = \begin{cases} A\exp(ik_1x) + B\exp(-ik_1x) & \text{for } x < 0; \\ C\exp(ik_2x) & \text{for } x > 0. \end{cases}$$

(We omit the time-dependent part $\exp(-iEt/\hbar)$ of the wave function.) The densities of the probability flux (formula (P6.2.2.4)) of the incident, reflected, and transmitted waves are equal to $S_1 = |A|^2(\hbar k_1/m)$, $S_2 = |B|^2(\hbar k_1/m)$, and $S_3 = |C|^2(\hbar k_2/m)$. Hence the reflection and transmission coefficients are equal to $R = S_2/S_1 = |B/A|^2$ and $D = S_3/S_1 = (k_2/k_1)|C/A|^2$. From the continuity conditions for the wave function and its derivatives for $x = 0$ we have

$$1 + \frac{B}{A} = \frac{C}{A}, \qquad k_1\left(1 + \frac{B}{A}\right) = k_2\frac{C}{A},$$

which implies

$$R = \left|\frac{k_1 - k_2}{k_1 + k_2}\right|^2, \qquad D = \left|\frac{4k_1k_2}{(k_1 + k_2)^2}\right|. \tag{P6.2.2.5}$$

For $E > U_0$, the numbers k_1 and k_2 are real, the coefficients R and D are nonzero, and $R + D = 1$. We note that, in classical physics, there is no reflection of particles for $E > U_0$; that is, $R = 0$. In quantum physics, R is not only nonzero, but for $E - U_0 \ll U_0$, we obtain $k_2 \ll k_1$, and hence $D \to 0$ and $R \to 1$. Even more amazing purely quantum effect is obtained for $U_0 < 0$ (formally, (P6.2.2.5) remains valid, but $k_2 > k_1$): in the limit $E \ll |U_0|$, we obtain $k_1 \ll k_2$, and hence $R \to 1$ (!).

For $E < U_0$ ($U_0 > 0$), the solution remains formally true if we set $k_2 = i\kappa$, where $\kappa = \sqrt{2m(U_0 - E)}/\hbar$. In this case, $R = 1$, and the solution for $x > 0$ is the real exponential $|\psi|^2 = |C|^2\exp(-2\kappa x)$; the depth of penetration of particles into the classically forbidden region is equal to $1/2\kappa$.

Example 5. Tunnel effect. In the case of incidence on a potential barrier of the de Broglie wave corresponding to a free particle with energy less than the barrier height, there is a nonzero probability of finding the particle on the other side of the barrier. The barrier transparency is the ratio the intensity (the density of the probability flux) of the transmitted wave to that of the incident wave. The potential barrier transparency (Fig. P6.5) is given by the formula

$$D = D_0\exp\left(-\frac{2}{\hbar}\int_{x_1}^{x_2}\sqrt{2m[U(x) - E]}\,dx\right).$$

The particle can be found under the barrier in the region forbidden for motion in classical mechanics.

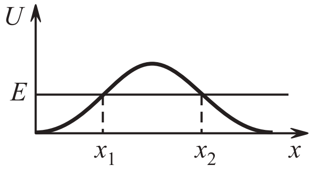

Figure P6.5. Tunnel effect (example 5).

P6.2.3. Operator Approach in Quantum Mechanics

► **Operators of physical quantities.** Any observable physical quantity in quantum mechanics is associated with a Hermitian (self-adjoint) operator acting on the wave function. A physical quantity G has a definite value λ_G only in the states corresponding to *eigenfunctions* of the operator $\widehat{G}$:

$$\widehat{G}\psi = \lambda_G\psi,$$

and the measurable values of the physical quantity in any of these states coincide with the *eigenvalues* λ_G of the operator $\widehat{G}$. The system of eigenfunctions of Hermitian operator is *complete*; that is, any function can be represented as a linear superposition of these functions. If a state is a superposition of eigenfunctions corresponding to different λ, $\psi = \sum_i C_i\psi_i$, then the probability of obtaining the value λ_i in measurements is equal to $|C_i|^2$, and the average value of G in this state is equal to

$$\langle G\rangle = \sum_i \lambda_i|C_i|^2 = \int \psi^*\widehat{G}\psi\,dV. \qquad \text{(P6.2.3.1)}$$

The coordinate operator has the form $\widehat{x}\psi = x\psi$, the momentum projection operator has the form $\widehat{p}_x = -i\hbar\frac{\partial}{\partial x}$, and the energy operator (*Hamiltonian*) is determined by the expression:

$$\widehat{H} = \frac{\widehat{p}^2}{2m} + U = -\frac{\hbar^2}{2m}\Delta + U. \qquad \text{(P6.2.3.2)}$$

The Schrödinger equations (P6.2.2.1) and (P6.2.2.3) become

$$i\hbar\frac{\partial\psi}{\partial t} = \widehat{H}\psi \quad \text{and} \quad \widehat{H}\psi = E\psi; \qquad \text{(P6.2.3.3)}$$

that is, the stationary states are the eigenfunctions of the Hamiltonian, and the energy of these states are the corresponding eigenvalues.

► **Commensurability of physical quantities.** Two physical quantities, A and B, are said to be *commensurable* (can simultaneously have definite values) if they commute: $\widehat{A}\widehat{B} - \widehat{B}\widehat{A} = 0$. For the coordinate and momentum projections, we have

$$\widehat{x}\widehat{p}_x - \widehat{p}_x\widehat{x} = i\hbar. \qquad \text{(P6.2.3.4)}$$

Example 1. Uncertainty principle. Operator approach. In the language of quantum operators, the uncertainty principle becomes an exact mathematical statement. The uncertainty of a physical quantity in the state ψ is $\Delta G = \sqrt{\langle(\widehat{G} - \langle\widehat{G}\rangle)^2\rangle}$ (in the eigenstate of the operator $\widehat{G}$, the uncertainty is zero). The degree of commensurability of two physical quantities $\widehat{A}$ and $\widehat{B}$ is determined by their commutator $\widehat{C} = \widehat{A}\widehat{B} - \widehat{B}\widehat{A}$, which is expressed by the inequality

$$\Delta A\Delta B \ge \tfrac{1}{2}|\langle\widehat{C}\rangle|.$$

► **Conservation condition for a physical quantity.** An important statement of quantum mechanics is that the physical quantity is preserved in time if its operator commutes with the Hamiltonian. Indeed,

$$\frac{d}{dt}\langle\widehat{G}\rangle = \int \frac{\partial\psi^*}{\partial t}\widehat{G}\psi\,dV + \int \psi^*\widehat{G}\frac{\partial\psi}{\partial t}\,dV = \frac{i}{\hbar}\int \psi^*[\widehat{H}\widehat{G}]\psi\,dV = \frac{i}{\hbar}\langle[\widehat{H}\widehat{G}]\rangle,$$

where $\partial\psi/dt$ is expressed by using the Schrödinger equation (P6.2.3.3) (and $\partial\psi^*/dt$ is expressed by using the dual equation).

Example 2. Quantum oscillator. Operator approach. The commutation relations permit obtaining important results by pure algebraic methods. Let us find the eigenvalues and eigenfunctions of the Hamiltonian describing the quantum oscillator:

$$\widehat{H} = \frac{\widehat{p}_x^2}{2m} + \frac{m\omega^2\widehat{x}^2}{2},$$

starting only from the commutation relation (P6.2.3.4) between the coordinate and momentum. Further, we introduce dimensionless operators $\widehat{q}$, $\widehat{p}$, and $\widehat{h}$ using the formulas

$$\widehat{x} = \sqrt{\frac{\hbar}{m\omega}}\,\widehat{q}, \quad \widehat{p}_x = \sqrt{\hbar m\omega}\,\widehat{p}, \quad \widehat{H} = \frac{\hbar\omega}{2}\widehat{h},$$

and $\widehat{p} = -i\frac{d}{dq}$. Then, the dimensionless Hamiltonian acquires the form

$$\widehat{h} = \widehat{p}^2 + \widehat{q}^2,$$

and the dimensionless coordinate and momentum operators satisfy the relation

$$\widehat{q}\widehat{p} - \widehat{p}\widehat{q} = i.$$

We introduce the "raising" and "lowering" operators:

$$\widehat{a}^+ = \widehat{q} - i\widehat{p}, \qquad \widehat{a}^- = \widehat{q} + i\widehat{p},$$

which satisfy the identities

$$\widehat{a}^+\widehat{a}^- = \widehat{h} - 1, \quad \widehat{a}^-\widehat{a}^+ = \widehat{h} + 1, \quad [\widehat{a}^+\widehat{a}^-] = -2.$$

The eigenvalues λ and eigenfunctions ψ_λ of the operator $\widehat{A} = \widehat{a}^+\widehat{a}^-$ have the following properties:

1) The eigenvalues are nonnegative ($\lambda \ge 0$). An eigenvalue is zero if and only if $\widehat{a}^-\psi_0 = 0$.

2) If ψ_λ is a normed eigenfunction with $\lambda > 0$, then $\frac{1}{\sqrt{\lambda}}\widehat{a}^-\psi_\lambda$ is a normed eigenfunction with $\lambda' = \lambda - 2$, and $\frac{1}{\sqrt{\lambda+2}}\widehat{a}^+\psi_\lambda$ is a normed eigenfunction with $\lambda' = \lambda + 2$.

This implies that the eigenvalues of the operator $\widehat{A}$ are even integers, the eigenvalues of the operator $\widehat{h}$ are odd integers, and the eigenvalues of the energy operator are $E_n = \hbar\omega(n + \frac{1}{2})$. The ground state ($n = 0$) satisfies the equation $\widehat{a}^-\psi_0 = 0$ or $\frac{d\psi_0}{dq} = -q\psi_0$. We obtain $\psi_0(q) = \sqrt{\frac{2}{\pi}}\exp(-q^2/2)$. All the other functions can be obtained by successive application of the raising operator: $\psi_1 = \frac{1}{\sqrt{2}}(q - d/dq)\psi_0$, $\psi_2 = \frac{1}{\sqrt{3}}(q - d/dq)\psi_1$, etc.

► **Three-dimensional Schrödinger equation.** Now we consider the three-dimensional Schrödinger equation. This equation is much more complicated than the one-dimensional equation, but in several cases, it can be reduced to three one-dimensional equations by the method of separation of variables. We illustrate this method with an example of the three-dimensional oscillator.

Example 3. Three-dimensional quantum oscillator. Separation of variables. The Schrödinger equation for the three-dimensional oscillator has the form

$$\Delta\psi + \frac{2m}{\hbar^2}\Big(E - \frac{m\omega^2r^2}{2}\Big)\psi = 0. \qquad \text{(P6.2.3.5)}$$

The oscillator potential energy is unique: it is spherically symmetric and simultaneously the sum of three terms each of which depends only on one coordinate and coincides with the potential energy of the one-dimensional oscillator ($U_x = m\omega^2x^2/2$ etc.). This permits seeking the solution in the form

$$\psi(\mathbf{r}) = \psi_1(x)\psi_2(y)\psi_3(z).$$

Substituting ψ in equation (P6.2.3.5) and dividing it by ψ, we obtain

$$\left(\frac{1}{\psi_1}\frac{\partial^2\psi_1}{\partial x^2}-\frac{2m}{\hbar^2}U_x\right)+\left(\frac{1}{\psi_2}\frac{\partial^2\psi_2}{\partial y^2}-\frac{2m}{\hbar^2}U_y\right)+\left(\frac{1}{\psi_3}\frac{\partial^2\psi_3}{\partial z^2}-\frac{2m}{\hbar^2}U_z\right)+\frac{2m}{\hbar^2}E=0.$$

Since only the first term depends on x, it must be a constant that we denote by $2mE_1/\hbar^2$. The obtained equation coincides with the equation for the one-dimensional oscillator, and hence E_1 can take only the values $\hbar\omega(n_1+\frac{1}{2})$. The same concerns the other two terms. Since $E = E_1 + E_2 + E_3$, for the energy of the three-dimensional oscillator we obtain

$$E_{n_1,n_2,n_3}=\hbar\omega(n+\tfrac{3}{2}),$$

where $n = n_1 + n_2 + n_3$, and n_1, n_2, n_3 take the values $0, 1, 2, \ldots$. One can see that the ground state ($n = 0$) is nondegenerate, the level with $n = 1$ is triply degenerate, the level with $n = 2$ is sixfold degenerate, the level with $n = 3$ is tenfold degenerate, etc.

The operator representation permits writing the method of separation of variable more compactly. The Schrödinger equation (P6.2.3.5) takes the form

$$(\widehat{H}_1+\widehat{H}_2+\widehat{H}_3)\psi_1\psi_2\psi_3=E\psi_1\psi_2\psi_3.$$

Dividing by $\psi_1\psi_2\psi_3$, we obtain

$$\frac{\widehat{H}_1\psi_1}{\psi_1}+\frac{\widehat{H}_2\psi_2}{\psi_2}+\frac{\widehat{H}_3\psi_3}{\psi_3}=E.$$

Only the first term depends on x; it is equal to E_1, etc.

P6.2.4. Motion in a Central Field. Hydrogen Atom

The general approach to solving the Schrödinger equation in a spherically symmetric field relies on using the conservation law for the moment of momentum (angular momentum). Such an approach permits separating the variables (in spherical coordinates) for any spherically symmetric potential.

► **Operator of angular momentum.** The *operator of orbital angular momentum* (moment of momentum) of an electron $\widehat{\mathbf{L}}$ with respect to point O is determined by the relations

$$\widehat{L}_x=\widehat{y}\widehat{p}_z-\widehat{z}\widehat{p}_y,\quad \widehat{L}_y=\widehat{z}\widehat{p}_x-\widehat{x}\widehat{p}_z,\quad \widehat{L}_z=\widehat{x}\widehat{p}_y-\widehat{y}\widehat{p}_x, \tag{P6.2.4.1}$$

and the operator of squared angular momentum is determined by the relation $\widehat{L}^2=\widehat{L}_x^2+\widehat{L}_y^2+\widehat{L}_z^2$. The angular momentum projections do not commute with each other:

$$[\widehat{L}_x\widehat{L}_y]=i\hbar\widehat{L}_z,\quad [\widehat{L}_y\widehat{L}_z]=i\hbar\widehat{L}_x,\quad [\widehat{L}_z\widehat{L}_x]=i\hbar\widehat{L}_y, \tag{P6.2.4.2}$$

which means that they cannot simultaneously take definite values. But each of the projections commutes with the squared angular momentum, and hence we can speak about common eigenfunctions of the operators $\widehat{L}_z$ and $\widehat{L}^2$.

The triple of operators satisfying relations (P6.2.4.2) is called the vector of angular momentum even if it is impossible to write (P6.2.4.1) for them. The general notion of angular momentum in quantum mechanics is wider than the notion of classical orbital angular momentum of a particle, it also includes the notion of intrinsic angular momentum (spin) of subatomic particles (see Subsection P6.2.5).

The eigenvalues can be found purely algebraically from the commutation relations (P6.2.4.2). The common eigenfunctions of the operators $\widehat{L}^2$ and $\widehat{L}_z$ can correspond only to the following eigenvalues:

$$\widehat{L}^2\psi_{l,m}=l(l+1)\hbar^2\psi_{l,m},\qquad \widehat{L}_z\psi_{l,m}=m\hbar\psi_{l,m},$$

where l is an integer or half-integer nonnegative number, and for a given l, the number m can take $2l + 1$ values: $m = -l, -l + 1, \ldots, l - 1, l$.

We pass to dimensionless operators: $\widehat{L}_x = \hbar\widehat{l}_x$, $\widehat{L}_y = \hbar\widehat{l}_y$, $\widehat{L}_z = \hbar\widehat{l}_z$, satisfying the commutation relations $[\widehat{l}_x\widehat{l}_y] = i\widehat{l}_z$ etc. We introduce the "raising" and "lowering" operators:

$$\widehat{l}^+ = \widehat{l}_x + i\widehat{l}_y, \qquad \widehat{l}^- = \widehat{l}_x - i\widehat{l}_y,$$

satisfying the relations

$$[\widehat{l}_z\widehat{l}^+] = \widehat{l}^+, \quad [\widehat{l}_z\widehat{l}^-] = -\widehat{l}^-, \quad \widehat{l}^+\widehat{l}^- = \widehat{l}^2 - \widehat{l}_z^2 + \widehat{l}_z, \quad \widehat{l}^-\widehat{l}^+ = \widehat{l}^2 - \widehat{l}_z^2 - \widehat{l}_z.$$

If ψ is an eigenfunction of the operators $\widehat{l}^2$ and $\widehat{l}_z$ with the eigenvalues λ and μ, then $\widehat{l}^+\psi$ is an eigenfunction of these operators with the eigenvalues λ, $\mu + 1$, and $\widehat{l}^-\psi$ is an eigenfunction with the eigenvalues λ, $\mu - 1$. The minimum value μ_1 and maximum value μ_2 satisfy the equations:

$$\lambda - \mu_1^2 + \mu_1 = 0, \quad \lambda - \mu_2^2 - \mu_2 = 0,$$

which implies that $\mu_1 = -\mu_2$, and $\mu_2 - \mu_1 = 2\mu_2 = l$ is an integer. For the eigenvalue of the operator $\widehat{l}^2$, we obtain $\lambda = l(l+1)$.

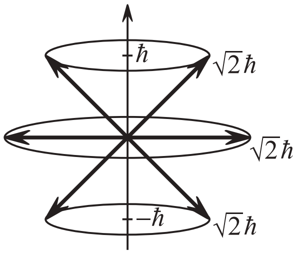

Figure P6.6. Representation of the angular momentum in the vector model.

One can see that the maximum value of the projection of the angular momentum $\hbar l$ is less than its magnitude $\hbar\sqrt{l(l+1)}$. In the *vector model*, the angular momentum is depicted as a vector making a fixed angle with the axis z and rotating about this axis. In this case, the projections L_x and L_y do not have fixed values, which shows that L_z is not commensurable with L_x and L_y. In Fig. P6.6, we present the case $l = 1$.

► **Eigenfunctions of the orbital angular momentum operator.** To find the eigenvalues of the *orbital* angular momentum determined by formulas (P6.2.4.1), it is convenient to pass to spherical coordinates. The orbital angular momentum is a function of only angular variables, and the commuting operators $\widehat{L}_z$ and $\widehat{L}^2$ take the form

$$\begin{aligned} &\widehat{L}_z = -i\hbar\frac{\partial}{\partial\varphi}, \\ &\widehat{L}^2 = -\hbar^2\widehat{\Lambda}, \quad \text{where} \quad \widehat{\Lambda} = \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\Big(\sin\theta\frac{\partial}{\partial\theta}\Big) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}. \end{aligned} \tag{P6.2.4.3}$$

In what follows, it is important that the operator $\widehat{\Lambda}$ coincides with the angular part of the Laplace operator written in spherical coordinates:

$$\Delta = \frac{1}{r^2}\frac{\partial}{\partial r}\Big(r^2\frac{\partial}{\partial r}\Big) + \frac{1}{r^2}\widehat{\Lambda}. \tag{P6.2.4.4}$$

The eigenfunctions can be represented as products of functions of different angular variables: $\psi_{l,m} = \Theta_{l,m}(\theta)\,\Phi_m(\varphi)$. The equation for Φ

$$-i\hbar\frac{d\Phi_m}{d\varphi} = m\hbar\Phi_m$$

has the solution $\Phi_m = A\exp(im\varphi)$, which is a unique function of the angle φ *only for integer values of* m. Hence, in the case of orbital angular momentum, only *integer* values of the numbers l and m are realized. The number l is called the *orbital quantum number* and the number m is called the *magnetic quantum number* (denoted by m_l). Half-integer values are realized in the case of intrinsic angular momentum—*spin*; the corresponding *spin quantum numbers* are denoted by s and m_s.

The equation for the eigenfunctions of the operator $\widehat{\Lambda}$ has solution in the form of *spherical functions* $Y_{lm}(\theta,\varphi) = \Phi_m(\varphi)\Theta_{ml}(\theta)$ with eigenvalues $-l(l+1)$, which are well known in mathematical physics. (The functions $\Theta(\theta)$ are expressed in terms of *associated Legendre functions*; see Subsection M13.10.1). For example,

$$\Phi_m(\phi) = \frac{1}{\sqrt{2\pi}}e^{im\phi},\quad \Theta_{1,0}(\theta) = \sqrt{\frac{3}{2}}\cos\theta,\quad \Theta_{1,\pm1} = \mp\sqrt{\frac{3}{4}}\sin\theta,\quad \Theta_{2,0} = \sqrt{\frac{5}{8}}\,(3\cos^2\theta - 1).$$

Each of the functions is normalized as follows: $\int_0^{2\pi}\Phi^2\,d\varphi = 1$ and $\int_0^{\pi}\Theta^2\sin\theta\,d\theta = 1$. The spherical functions form a complete system of functions of angular variables; that is, any function of the variables θ and φ can be expanded in spherical functions.

► **Addition of angular momenta.** The operator of the sum of two angular momenta $\widehat{\mathbf{L}} = \widehat{\mathbf{L}}_1 + \widehat{\mathbf{L}}_2$, whose values are characterized by quantum numbers l_1 and l_2 (i.e., $\widehat{L}_1^2$ and $\widehat{L}_2^2$ have the eigenvalues $\hbar\sqrt{l_1(l_1+1)}$ and $\hbar\sqrt{l_2(l_2+1)}$), is characterized by the quantum number l, which can take the following values: $|l_1 - l_2| \le l \le l_1 + l_2$. The maximum value corresponds to the parallel orientation of the angular momenta, the minimum value corresponds to their antiparallel orientation. The operator $\widehat{L}^2$ does not commute with the operators L_{1z} and L_{2z}, and hence the eigenfunctions of the operators $\widehat{L}^2$ and $\widehat{L}_z$ are not eigenfunctions of L_{1z} and L_{2z} (but are eigenfunctions of $\widehat{L}_1^2$ and $\widehat{L}_2^2$).

► **Motion in a central field.** In this case, the potential energy $U(\mathbf{r})$ depends only on the modulus of $\mathbf{r}$. Since the orbital angular momentum operators $\widehat{L}_z$ and $\widehat{L}^2$ commute with $\widehat{r}^2$ and $\widehat{p}^2$, they also commute with the Hamiltonian. Hence, just as in classical mechanics, the value of the angular momentum and its projection onto the z-axis are preserved in the motion of a particle in a central field. The wave function of any state $\psi(r,\theta,\varphi)$ can be expanded in the common eigenfunctions ψ_{E,l,m_l} of the three commuting operators $\widehat{H}$, $\widehat{L}_z$ and $\widehat{L}^2$ corresponding to the three simultaneously measurable physical quantities: the energy, the orbital angular momentum projection, and its value. We use the separation of variables; that is, we seek the eigenfunctions in the form: $\psi_{E,l,m_l} = R_{E,l}(r)\Phi_{m_l}(\varphi)\Theta_{l,m_l}(\theta)$. Writing the Hamiltonian in spherical coordinates and taking (P6.2.4.3) and (P6.2.4.4) into account, we obtain

$$\widehat{H} = -\frac{\hbar^2}{2mr^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{\widehat{L}^2}{2mr^2} + U(r). \qquad \text{(P6.2.4.5)}$$

The Schrödinger equation in form (P6.2.3.5) after the action of the Hamiltonian (P6.2.4.5) on the function ψ_{E,l,m_l} becomes the equation for the radial part of the wave function $R_{E,l}(r)$:

$$-\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)R_{E,l} + \left(\frac{\hbar^2 l(l+1)}{2mr^2} + U(r)\right)R_{E,l} = ER_{E,l}. \qquad \text{(P6.2.4.6)}$$

Just as in classical mechanics (see Subsection P1.6.2), the potential energy is supplemented with the *centrifugal energy* $\hbar^2 l(l+1)/2mr^2$. One can see that the magnetic quantum number is not contained in the equation for allowed values of the energy. To each energy level there correspond two quantum numbers: the *orbital number* l and the *radial number* $n_r = 0, 1, 2, \dots$ numbering the allowed discrete values of the energy. The level $E_{n_r,l}$ is $(2l+1)$-fold degenerate with respect to the values of the magnetic quantum number m_l.

► **Magnetic moment.** In classical physics, a charged particle moving in a closed trajectory has a magnetic moment. The value and direction of the magnetic moment completely determine both the magnetic field of the moving charge and its interaction with the external magnetic field (see Section P3.8). For orbital motion of an electron, the ratio of its magnetic moment to its mechanical angular momentum is equal to $-e/2m$. Hence, the projection of the orbital magnetic moment on the axis z takes the values $\mu_z = -(e/2m)\hbar m_l = -\mu_{\rm B} m_l$, and the value of the orbital magnetic moment is equal to $\mu_{\rm B}\sqrt{l(l+1)}$, where $\mu_{\rm B} = e\hbar/2m \approx 9.274\times10^{-24}$ J/T is called the *Bohr magneton*. In the general case, the relation between the magnetic moment and mechanical angular momentum can be written as: $\mu_z = g(e/2m)L_z$ (or $(\mu_z/\mu_{\rm B}) = g(L_z/\hbar)$). The dimensional multiplier g is called the *gyromagnetic ratio*; it is equal to 1 for orbital angular momentum (and denoted g_l) and 2 for spin (denoted g_s).

For an atom in a magnetic field with induction **B** directed along the axis z, the electron acquires additional energy of interaction between the magnetic moment and the magnetic field: $\Delta E = -\mu_z B = \mu_{\rm B} m_l B$. One can see that in the magnetic field, the degeneration in the magnetic quantum number m_l must be removed; that is, instead of one level there arise $2l + 1$ equidistant levels the distance between which is proportional to the magnetic field. In this case, each spectral line must split into three equidistant lines (taking the *rules of selection* discussed below into account, we see that the variation in m_l in transitions between the levels can only be $\Delta m_l = 0, \pm1$). This effect is called the *simple (normal) Zeeman effect*. The fact that the electron has an intrinsic (spin) magnetic moment strongly complicates the pattern of the line splitting in the magnetic field.

► **Hydrogen atom in quantum mechanics.** With the potential energy $U(r) = -ke^2/r$, equation (P6.2.4.6) has the following solutions:

1) The possible values of the energy are determined by the formula

$$E_{n_r,l} = -\frac{R}{(n_r + l + 1)^2},$$

where R is the Rydberg constant (see (P6.1.2.5)). The quantity $n = n_r + l + 1$ is called the *principal quantum number*. One can see that the electron energy is determined only by the principal quantum number. To one and the same principal quantum number n there correspond states with distinct values of l: $l = 0, 1, \dots, n-1$. The total degeneracy of each energy level equals $\sum_{l=0}^{n-1}(2l + 1) = n^2$. The additional degeneracy is a special property of the Coulomb potential. We note that the energy levels coincide with those obtained in the semiclassical Bohr theory (Subsection P6.1.2), but there are profound important distinctions: in Bohr's theory, the orbital angular momentum in the ground state ($n = 1$) is $L = \hbar$, and in quantum theory, the angular momentum of the ground state is zero.

2) The radial functions have the form $R_{n,l} = r^l e^{-r/na_0} P_{n_r}(r)$, where $P_{n_r}(r)$ is a polynomial of degree $n_r = n - l - 1$, and $a_0 = \hbar^2/(mke^2)$ is the radius of the first electron orbit in Bohr's theory (the Bohr radius). For example,

$$R_{1,0} = \frac{2}{\sqrt{a_0^3}}e^{-r/a_0}, \quad R_{2,0} = \frac{1}{\sqrt{8a_0^3}}e^{-r/2a_0}\Bigl(2 - \frac{r}{a_0}\Bigr), \quad R_{2,1} = \frac{1}{2\sqrt{6a_0^3}}\frac{r}{a_0}e^{-r/2a_0}.$$

The radial functions are normed by unity: $\int_0^\infty R^2 r^2\,dr = 1$. The probability of finding the electron at a distance from the nucleus lying in the interval $(r, r + dr)$ is equal to $dP = \int_{(r,r+dr)} |\psi|^2\,dV = R^2(r)r^2\,dr$. In the ground state, the most probable distance from the electron to the nucleus is equal to the Bohr radius a_0.

The electron state is denoted by the Latin letter indicating the value of the orbital quantum number (s, p, d, and f correspond to $l = 0, 1, 2, 3$, respectively, and further in alphabetical order), before which the value of the principal quantum number is indicated. For example, $3d$ corresponds to the electron with $n = 3$ and $l = 2$.

► **The spectra of alkali metals**. In the case of alkali metals (Li, Na, ...), the atomic state is determined by the electron state on the outer shell. This electron is in the centrally symmetric field of the nucleus and $Z - 1$ electrons on the inner shells. The outer shell radius is significantly larger than the inner shell radii ($r_n \sim n^2$), and hence, in the main region of electron existence, the electric field coincides with that of the charge $+e$. The field distortion by inner shell electrons exerts larger influence on a state with lesser orbital number l, because $\psi(r) \sim r^l$ for small r. For the energy levels $E_{n,l}$, where $n = n_r + l + 1$ is the principal quantum number, one usually employs the expression

$$E_{n,l} = -\frac{R}{(n + \alpha_l)^2},$$

where $\alpha_l < 0$ is the *Rydberg correction*. For example, for Li, $\alpha_0 = -0.41$ and $\alpha_1 = -0.04$, and for Na, $\alpha_1 = -0.88$.

The difference between the field where an electron moves from the Coulomb field can be taken into account in the first approximation by adding the term A/r^2 to the potential energy. This additional term takes account of two effects at once: first, the screening field of the electron "core," and second, the polarization of the core by the outer shell electron, which results in that the core acquires a small electric dipole moment directed at each time moment towards this electron. The term A/r^2 can be added to the centrifugal energy (see formula (P6.2.4.6)), and both terms can be written together as the centrifugal energy with redefined $l' = l + \alpha_l$.

To construct the spectrum, one must take account of the fact that the principal quantum number n of the lowest state of the outer shell electron is equal to the shell number in which this electron is located (for lithium, $n = 2$, for potassium, $n = 3$, etc.). Moreover, it is necessary to take account of the rule of selection of the orbital number l: $\Delta l = \pm 1$. In Fig. P6.7, we present the diagram of spectral lines of Li atom and give the names of the first principal series.

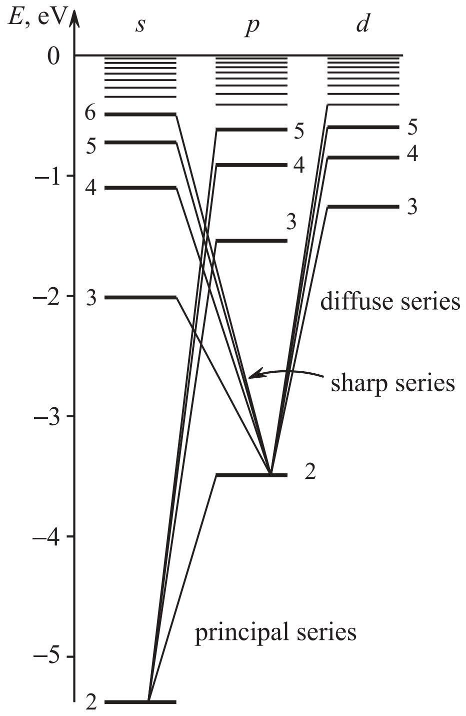

Figure P6.7. Spectra of alkali metals (Li).

P6.2.5. Electron Spin and Fine Structure of the Spectrum

► **Spin of electrons.** Several experimental facts show that to describe the state of electrons in an atom, it is insufficient to have three quantum numbers related to its spatial motion. These facts are:

1) The multiplet (fine) structure of spectral lines of alkali metals. For example, the most intensive line in the principal series; that is, the resonance line corresponding to the transition to the ground s-state from the first excited p-state, has the form of a pronounced doublet.

2) The complicated (anomalous) Zeeman effect. Splitting of atomic lines in a magnetic field cannot be described as the simple Zeeman effect related to the magnetic moment of the orbital motion.

3) The results of the Stern–Gerlach experiments on measuring magnetic moments of silver atoms (1921). A beam of silver atoms was directed into the region of a strongly inhomogeneous magnetic field, where the atoms were acted upon by the force $\mu_z(dB/dz)$. Since a silver atom has a single electron in the outermost shell, the atomic angular momentum and magnetic moment are equal to those of this electron. The orbital magnetic moment projection $\mu_z = m_l\mu_{\rm B}$ can have $2l + 1$ distinct values. Atoms with electrons in the s-state do not deflect in magnetic field; if there is a fraction of atoms with electrons in the p-state, then they must split into three beams. It was discovered that a beam of atoms splits into two beams deflected by the magnetic field to opposite sides by the same distances.

The conclusion is that the electron has an intrinsic angular momentum (called *spin*) that is not related to the electron orbital motion. The spin projection on the axis z can take two values, which implies that the *spin quantum number* s must be equal to $1/2$ (because $2s + 1 = 2$). The value of the spin angular momentum is equal to $\hbar\sqrt{\frac{1}{2}(\frac{1}{2}+1)} = \frac{\sqrt{3}}{2}\hbar$. The spin angular momentum projection on the axis z is equal to $\hbar m_s$, where $m_s = \pm\frac{1}{2}$ is the *magnetic spin quantum number* (also known as the *secondary spin quantum number*). We recall that half-integer values of the angular momentum quantum number do not contradict the commutation relations.

Since the spin projection can take two values, it follows that to the set of quantum numbers n and l, there correspond $2(2l + 1)$ distinct states and to the quantum number n, there correspond $2n^2$ states.

A variation in the value of deviation of atoms in a magnetic field permits finding the value of the intrinsic magnetic moment. It turned out that the intrinsic magnetic moment of an electron has anomalous gyromagnetic ratio $g_s = 2$: the ratio of the spin magnetic moment to the spin angular momentum is equal not to $e/2m$ but to e/m. The magnetic moment projection on the axis z takes the values $(e/m)\hbar m_s = \pm\mu_{\rm B}$. The anomalous values of the gyromagnetic ratio are confirmed by the results of the *Einstein–de Haas experiments* (1914), in which magnetomechanical effects were studied (formation of a mechanical angular momentum in magnetization of an iron sample).

All subatomic particles have an intrinsic angular momentum associated—the spin. The spin quantum number of proton, neutron, and μ-mesons is equal to 1/2, that of photon is equal to 1, and that of π-mesons is 0.

► **Fine structure of spectral lines.** The fine structure of lines of hydrogen-like atoms and alkali metals can be explained by interaction between the spin magnetic moment and the orbital angular momentum. The easiest way to understand the appearance of such an interaction is to use the reference frame associated with the electron. In this reference frame, the moving nucleus creates a magnetic field with which the spin magnetic moment interacts. With the *spin-orbit interaction* (also known as *spin-orbit coupling*) taken into account, the following operators are preserved (commute with the Hamiltonian): the total angular momentum $\widehat{J}^2$, its projection $\widehat{J}_z$, and the value of the orbital angular momentum $\widehat{L}^2$. The position of the levels $E_{n,l,j}$ is determined by three quantum numbers, n, l, and j. (In many-electron atoms, the quantum numbers of the angular momentum are denoted by the capital letters L, S, and J. In spectroscopy, the orbital angular momentum is denoted by the letters S, P, D, F, ... In alkali metals, the atomic state is determined by a single electron, and one can use both capital and small letters.) The degeneration of each level is given by $2j+1$ states with distinct values of the projection of the total angular momentum. For a given l, the total angular momentum number can take two values: $j = (l \pm \frac{1}{2})$, for $l = 0$, there is only one value $j = \frac{1}{2}$. Hence the s-levels do not split (singlet levels), and the levels with $l > 0$ split into two near levels (doublet levels). When constructing the spectra, one must take the rules of selection into account: $\Delta l = \pm 1$, $\Delta j = 0, \pm 1$. For example, in transition from the doublet p-state to the lower singlet s-state, there arise two lines (doublet of the principal series), and the distance between them decreases with increasing n; and in transition from singlet s-levels to the lowest p-level, there arise doublets of the sharp series, which have the same splitting. In transition from the double d-state into the lower doublet p-state, there arise three lines (multiple doublets of diffusion series). The splitting values decreases with increase in both the orbital and principal quantum numbers, but it increases fast in transitions to atoms with larger serial numbers.

In the hydrogen atom and hydrogen-like ions, the spin-orbit interaction is partially compensated by relativistic corrections, and the final answer for the energy levels does not contain any dependence on the orbital quantum number:

$$\Delta E_{n,j} = \alpha^2 R \frac{Z}{n^3} \left(\frac{3}{4n} - \frac{1}{j + 1/2} \right),$$

where $\alpha = ke^2/\hbar c \approx 1/137$ is called the *fine structure constant*. Additional degeneration in the orbital quantum number is a specific characteristic of the Coulomb potential.

P6.3. Structure and Spectra of Complex Atoms

P6.3.1. Pauli Exclusion Principle. Structure of Complex Atoms

► **Self-consistent field.** For many-electron atoms, the problem of describing the atomic states becomes significantly more complicated. In the general case, it is necessary to consider the wave function depending on the coordinates and spin variables of all electrons and to take into account the interaction of electrons not only with the nucleus but also with one another. But a good description is given by the self-consistent field approximation, in which each electron moves in the average field of the nucleus and the other electrons. The self-consistent field potential $\varphi(r)$ is described by the Thomas–Fermi equation, which permits obtaining the energetically most profitable distribution of electrons in the space surrounding the atom.

Solving the Schrödinger equation for an electron in a self-consistent field one can obtain stationary states whose energy $E_{n,l}$ depends on the principal quantum number and the orbital quantum number. First, it is necessary to find the structure of the ground state (with the least energy) of the atom. What one-electron states are occupied by electrons in the ground state? At the first glance, it seems that it is most profitable to place all the electrons in the lowest state $1s$. But such a distribution of electrons is impossible because of the *Pauli exclusion principle* related to the identity property of particles.

► **Particle identity. Pauli principle.** The particle identity principle in quantum mechanics has a stronger meaning than that in classical mechanics. In classical physics, particles are identical but distinguishable from one another; for example, a permutation of two electrons results in a state that must be taken into account as a new state independent of the initial state. Unlike this, in quantum mechanics, particles are identical and indistinguishable, so that the state arising from permutation of particles absolutely identically coincides with the initial state. This means that $|\psi(1,2)|^2 = |\psi(2,1)|^2$. There are two possibilities: either the wave function is symmetric with respect to particle states, so that $\psi(1,2) = \psi(2,1)$, or it is antisymmetric, so that $\psi(1,2) = -\psi(2,1)$. The particles of the first type are called *bosons*, the particles of the second type are called *fermions*. In particle physics, it is proved that bosons are particles with integer spin and the fermions are particles with half-integer spin. If the wave function is constructed of one-particle wave functions, then it must be symmetrized in the case of bosons and antisymmetrized in the case of fermions. For example, in the case of two bosons with one-particle states ψ_1 and ψ_2, the wave function has the form $\psi(\mathbf{r}_1,\mathbf{r}_2) = \frac{1}{\sqrt{2}}[\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2) + \psi_2(\mathbf{r}_1)\psi_1(\mathbf{r}_2)]$, and in the case of two fermions, it has the form $\psi(\mathbf{r}_1,\mathbf{r}_2) = \frac{1}{\sqrt{2}}[\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2) - \psi_2(\mathbf{r}_1)\psi_1(\mathbf{r}_2)]$.

Electrons are fermions, and hence a system of several electrons is described by an antisymmetric (in particle states) wave function. In such a system (just as in a system of any identical fermions), the *Pauli exclusion principle* acts: *two electrons cannot be in absolutely equal quantum states*; indeed, the wave function of such a state must be identically zero.

► **Periodic table (Mendeleev's table).** To describe the electronic structure of the ground state of an atom with atomic number Z, one must place Z electrons in different quantum states with least possible energy. To calculate electronic states, one can choose any set of independent quantum numbers of a single electron; for example, just as in the case of alkali metals, these can be n, l, j, and m_j. But it is more convenient to neglect the spin-orbit interaction and use the numbers n, l, m_l, and m_s for an electron in a centrally symmetric self-consistent field. (To completely description the terms of the ground and excited states, one must choose a set of quantum numbers that describe the state of the entire atom and correspond to quantities that are conserved.)

All the electrons with the same principal quantum number n correspond to one *electron shell*. The maximum number of electrons in the nth shell is equal to $2n^2$. The shell with $n = 1$ is called the K-shell, the shell with $n = 2$ is called the L-shell, and further, by the alphabet. All electrons in a given shell with a fixed number l form an *electron subshell* with $2(2l + 1)$ electrons; $l = 0, \ldots, n - 1$. The electrons of a given shell with a larger value of l have a larger energy due to increasing contribution of the centrifugal energy. The electron state is denoted by the number of the principal quantum number followed by the letter corresponding to its orbital number. If an atom contains several electrons in the same subshell, then the superscript on the right is the number of electrons. We present the electron configurations for several atoms:

K-shell:

$${}_1\mathrm{H}\text{: } 1s, \quad {}_2\mathrm{He}\text{: } 1s^2;$$

L-shell:

$${}_3\mathrm{Li}\text{: } 1s^2\,2s, \quad {}_4\mathrm{Be}\text{: } 1s^2\,2s^2,$$

$${}_5\mathrm{B}\text{: } 1s^2\,2s^2\,2p, \quad {}_6\mathrm{C}\text{: } 1s^2\,2s^2\,2p^2, \quad {}_7\mathrm{N}\text{: } 1s^2\,2s^2\,2p^3,$$

$${}_8\mathrm{O}\text{: } 1s^2\,2s^2\,2p^4, \quad {}_9\mathrm{F}\text{: } 1s^2\,2s^2\,2p^5, \quad {}_{10}\mathrm{Ne}\text{: } 1s^2\,2s^2\,2p^6;$$

two subshells (s and p) in the M-shell:

$${}_{11}\mathrm{Na}\text{: [Ne]}3s, \quad {}_{12}\mathrm{Mg}\text{: [Ne]}3s^2,$$

$${}_{13}\mathrm{Al}\text{: [Ne]}3s^2\,3p, \quad {}_{14}\mathrm{Si}\text{: [Ne]}3s^2\,3p^2, \quad {}_{15}\mathrm{P}\text{: [Ne}3s^2\,3p^3,$$

$${}_{16}\mathrm{S}\text{: [Ne]}3s^2\,3p^4, \quad {}_{17}\mathrm{Cl}\text{: [Ne]}3s^2\,3p^5, \quad {}_{18}\mathrm{Ar}\text{: [Ne]}3s^2\,3p^6;$$

s-subshell in the N-shell:

$$_{19}\text{K: [Ne]}3s^2\,3p^6\,4s, \quad _{20}\text{Ca: [Ne]}3s^2\,3p^6\,4s^2;$$

d-subshell in the M-shell:

$$_{21}\text{Sc: [Ne]}3s^2\,3p^6\,3d\,4s^2, \quad _{22}\text{Ti: [Ne]}3s^2\,3p^6\,3d^2\,4s^2,$$
$$_{23}\text{V: [Ne]}3s^2\,3p^6\,3d^3\,4s^2, \quad _{24}\text{Cr: [Ne]}3s^2\,3p^6\,3d^5\,4s;$$

and so on. The notation [Ne] stands for the electron configuration of $_{10}$Ne ($1s^2\,2s^2\,2p^6$).

One can see that, starting from number 19, the sequential filling of shells is violated: the energy of s-electrons in the N-shell is less than that of d-electrons in the M-shell and the $4s$-subshell is filled before the $3d$-subshell (see also Chapter S7). The chemical properties of elements are determined by the number of electrons in the outermost shell, and this results in the periodicity of chemical properties. For example, the noble gases—helium $_2$He, neon $_{10}$Ne, argon $_{18}$Ar, krypton $_{36}$Kr, xenon $_{54}$Xe, and radon $_{86}$Rn—differ in that their p-subshells are complete, and hence these elements have a large energy of ionization and tend to be chemically inert. The atoms with further numbers (lithium, sodium, potassium, rubidium, cesium, and francium) have complete p-subshells and one electron in the outermost s-subshell. These atoms form the group of alkali metals; they have a small energy of ionization and are all highly reactive.

We note that the total angular momentum and the total magnetic moment of any completely filled shell are zero.

P6.3.2. Spectra of Complex Atoms

► **Addition of orbital and spin angular momenta.** In light and medium atoms, the spin-orbit interaction is weak for separate electrons, and the addition of angular momenta occurs by the following scheme. The orbital angular momenta of electrons are added together to form the total orbital angular momentum $\widehat{\mathbf{L}}$, which is characterized by the quantum number L, and the electron spins are added up to make the total spin angular momentum $\widehat{\mathbf{S}}$ (quantum number S). In turn, the orbital and spin angular momenta add up to form the atom total angular momentum $\widehat{\mathbf{J}}$. Such a scheme is called the normal LS-coupling or the *Russell–Saunders coupling*. The number L is always integer, and the number S is integer if the shell contains an even number of electrons and is half-integer if the number of electrons is odd. We note that one must add up only the angular momenta of electrons in an unfilled shell.

In heavy atoms, the angular momenta are added together according to the scheme called the *jj-coupling*. First, one adds up the orbital and spin angular momenta of each individual electron ($\mathbf{j}_i = \mathbf{l}_i + \mathbf{s}_i$) and then the electron angular momenta are added up to form the total angular momentum of the atom ($\mathbf{J} = \sum_i \mathbf{j}_i$).

► **Classification of terms of complex atoms.** In the case of normal coupling, the total angular momentum is preserved (the value and the projection) and the values of the orbital and spin angular momenta. Each term is characterized by the quantum numbers L, S and J and, in the absence of a magnetic field, it is $(2J+1)$-fold degenerate (the states that differ only in the value of the total angular momentum projection have the same energy). For given L and S, the quantum number J can take $2S+1$ values from $L-S$ to $L+S$, and hence the number $2S+1$ is called the term *mutliplicity*. The difference between the energies of terms in the same multiplet is determined by a comparatively weak spin-orbit interaction. The spectral notation of terms is: the orbital quantum number of an atom is denoted by the capital letter S, P, D, F, G, and further in alphabetical order (to distinguish from one-electron states, where the lowercase letters are used). The subscript on the right shows the number J, which distinguished this term in the multiplet, and the superscript on the left indicates the term multiplicity $2S+1$. (It should be noted that, for $L<S$, the actual multiplicity of a term is equal to $2L+1$, but it is conventional to show $2S+1$.) For example, the term ${}^2P_{3/2}$ corresponds to the quantum numbers $L=1$, $S=1/2$, and $J=3/2$. In addition, in the scheme of spectral transitions, where the excited states must be shown, the principal quantum number of the excited electron is given (the least number is equal to the number of the shell to which the unfilled subshell belongs).

► **The Hund rule.** In the addition of spin and orbital angular momenta of electrons, one can obtain distinct quantum numbers. To understand what quantum numbers correspond to the ground state of the atom, one can use the empirical *Hund rules*. The first rule says that the term with the least energy has the maximum S and, for this S, the value of L is a maximum. (First, one determines the maximum possible m_S and m_L and they are used to determine S and L.) The second rule: $J=|L-S|$ if less than half the subshell is filled, and $J=L+S$ otherwise. The meaning of the first Hund rule is that if electrons have the same spin states, then they have different spatial states, which corresponds to the lesser energy of interaction between the electrons. For example, for carbon with subshell $2p^2$, the total spin is equal to $S=1$, and the orbital number cannot take the maximum value $L=2$, because, in this case, the two electrons would have the same quantum numbers $m_l=1$; hence $L=1$, and since two electrons occupy less than half the p-subshell, where there are 6 places, it follows that $J=|L-S|=0$. The principal term of the carbon atom is ${}^3P_0$. Verify that the principal term of the nitrogen atom is ${}^4S_{3/2}$.

► **Selection rules.** When constructing the scheme of spectral lines of an atom and determining the multiplicity of a radiation line, it is necessary to use both the term multiplicity and the *selection rules* for transitions between the terms. The selection rules for radiation transitions can be obtained by calculating the probability of transitions between the states. But the meaning of the selection rules can be understood using the momentum conservation law and taking into account the fact that the spin of an emitted photon is equal to 1.

The selection rule for the number J is: $\Delta J=0, \pm1$, but the transition between two states with $J=0$ is forbidden. The selection rule for the number m_J is: $\Delta m_J=0,\pm1$. The selection rule for the quantum number S is: $\Delta S=0$. The selection rule for the quantum number L is: $\Delta L=0, \pm1$, but the transition between two states with $L=0$ is forbidden. In atoms with a single electron in an unfilled subshell, the transitions with $\Delta L=0$ are forbidden.

The selection rule for the spin number is related to the fact that the magnetic field of an electromagnetic wave weakly interacts with the magnetic moment of the electron (the action on the orbital angular momentum is determined by electric forces). For example, in atoms with two electrons, one can observe independent series for states with $S=0$ (singlet terms) and for states with $S=1$ (triplet terms). In the case of helium, the data are described using the historical concept about two types of helium: parahelium ($S=0$) and orthohelium ($S=1$).

► **Characteristic X-ray spectra.** In the bombardment of the anode of an electron tube by energetic electrons (with energies $\sim 10^2$–10^4 eV), one can observe the X-ray radiation of two types. The slowing-down radiation is related to radiation of X-ray quanta directly by electrons. A typical characteristic of this radiation is a short-wavelength boundary of the spectrum (see Subsection P5.3.2) and a smooth distribution of the radiation intensity over the spectrum. Against this background, there are several intensive narrow lines of the characteristic radiation. Such a radiation appears in the case of liberation of an electron belonging to one of the inner electron subshells, after which the free place is occupied by an electron from a higher subshell and an X-ray quant is radiated. If one of the two electrons in the K-shell is liberated, then the following notation is used for the X-ray lines: K_α for the electron transition from the L-shell, K_β for the electron transition from the M-shell, etc. If an electron is liberated from the L-shell, then the spectral lines are denoted by L_α, L_β, etc.

For electrons in the inner subshells, the field of the nucleus is nearly unscreened. Therefore, in contrast to electrons in the outer subshells, the energy of inner subshell electrons must strongly depend on the serial number of the nucleus. One can expect that the energy of these electron levels must be described by the formula for hydrogen-like atoms with the screening correction: $E_n = R(Z-\delta)^2/n^2$. But such an approach does not give any good description of experimental facts and the empirical screening correction is introduced directly in the formula for spectral lines. For example, for the line K_α, we have

$$\frac{1}{\lambda} = \frac{3}{4} R_\lambda (Z-\sigma)^2,$$

where the screening correction σ for light atoms is approximately equal to 1. This law was discovered in 1913 by Moseley, who formulated it as the law of linear dependence of $\sqrt{\nu}$ for the characteristic line frequency on the serial number Z. The *Moseley law* allowed one to specify the serial numbers of many elements in the periodic table.

Since the X-ray spectrum is determined by the atomic inner subshell, it is independent of whether the atom is isolated or is contained in a molecule or a crystal. Therefore, the matter composition can be analyzed by using the characteristic X-ray spectra with a high universality.

Similarly to the optical spectra, the X-ray spectra have a fine structure because of the splitting of the levels of inner subshell electrons. But, in this case, one deals not with the normal LS-coupling but with the jj-coupling. For example, an electron in the L-shell can have $l = 0$ and then $j = 1/2$ or $l = 1$ and then the following two versions are possible: $j = 1/2$ or $j = 3/2$. Hence, the L-shell splits into three lines, and the difference in their energies is determined by the spin-orbit interaction. The K-shell contains only one level, the M-shell contains five levels, etc. In the analysis of spectra, one must follow the selection rules $\Delta l = \pm 1$ and $\Delta j = 0, \pm 1$. For example, the K_α line splits into two lines: $K_{\alpha 1}$ and $K_{\alpha 2}$, and there is no transition from the level with $l = 0$.

The study of absorption of the X-ray radiation directly allows one to see the position of energy levels of electrons in inner subshells. As the frequency increases, the absorption coefficients decreases smoothly, but for frequency equal to $\omega_i = E_i/\hbar$, where E_i is the energy of ionization of the electron from one of the inner levels, one observes a sharp increase in this coefficient, which is related to the fact that an addition absorption channel opens. According to the above, one can observe a single peak on the energy of the K-shell, three peaks on the energy of the L-shell, five peaks on the energy of the M-shell, etc.

► **Atom in an external magnetic field.** In sufficiently weak magnetic fields ($B < 1$ Tesla), the energy of the electron interaction with the magnetic field is small compared with the spin-orbit interaction, and the magnetic splitting of terms is small compared with the distance between the terms of the multiplet. In this case, one observes a *complicated (anomalous) Zeeman effect*; that is, the energy level splits into $2J + 1$ sublevels with distinct quantum numbers m_J, and the distance between the sublevels is proportional to the value of the magnetic field and depends on the quantum numbers L, S and J. The value of the sublevel displacement is given by the formula $\Delta E = g_J m_J \mu_\mathrm{B} B$, where μ_B is the Bohr magneton, and

$$g_J = 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)}$$

is called the *Landé splitting factor*.

The complicated character of the Zeeman effect in weak fields can be explained by an anomalous value of the spin magnetic moment for which the gyromagnetic ratio is equal to 2. As a result, the total magnetic moment $\widehat{\boldsymbol{\mu}} = -\frac{e}{2m}(\widehat{\mathbf{L}} + 2\widehat{\mathbf{S}})$ is not proportional to the total angular momentum $\widehat{\mathbf{J}}$. In the vector model, the magnetic moment can be represented as rapidly rotating about the preserved vector of total angular momentum, and, after averaging, only the magnetic moment projection onto the total angular momentum must remain:

$$\boldsymbol{\mu}_J = \mathbf{J}\frac{(\boldsymbol{\mu}\mathbf{J})}{J^2} = -\frac{e}{2m}\mathbf{J}\frac{(\mathbf{L}+2\mathbf{S})(\mathbf{L}+\mathbf{S})}{J^2}.$$

We express $(\mathbf{LS}) = \frac{1}{2}(J^2 - L^2 - S^2)$, replace J^2 by $J(J+1)$, L^2 by $L(L+1)$, and S^2 by $S(S+1)$ and obtain $\boldsymbol{\mu}_J = -\frac{e}{2m}g_J\mathbf{J}$, where g_J is the Landé splitting factor, whose expression was given above.

We note that even for nonzero S, L, and J, the Landé splitting factor can be zero ($S = 3/2$, $L = 2$, $J = 1/2$) and even be negative ($S = 5/2$, $L = 3$, $J = 1/2$). These effects are purely quantum and do not have classical analogs.

To obtain the splitting of spectral lines in a magnetic field, one must take the selection rule into account. For example, the level ${}^2P_{3/2}$ splits into 4 levels with the Landé splitting factor $\frac{4}{3}$, and the singlet level ${}^2S_{1/2}$ splits into two levels with the Landé splitting factor 2 (Fig. P6.8), but one observes not 8 lines but 6, because the transitions from $m_J = 3/2$ to $m_J = -1/2$ and from $m_J = -3/2$ to $m_J = 1/2$ are forbidden by the selection rule $\Delta m_J = 0, \pm 1$.

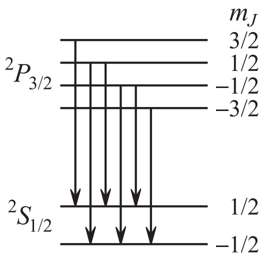

Figure P6.8. Anomalous Zeeman effect.

The simple Zeeman effect (splitting of one line into three) is observed in transition between the levels for which $g_1 = g_2 = g$. In this case, $\Delta\omega = g\Delta m_J B/\hbar$, where $\Delta m_J = 0, \pm 1$. Such a situation appears not only in transitions between two singlet levels ($S = 0$, $g = 1$) but also in some other cases.

If an atom is placed in a strong magnetic field (if the magnetic splitting of terms is much larger than the distance between the multiplet terms), then the fine structure of levels can be neglected—that is, the spin-orbit interaction can be neglected. The level with quantum numbers m_L, m_S is displaced in the magnetic field by $\Delta E = \mu_\mathrm{B} B(m_L + 2m_S)$. The variation in the frequency of radiation is equal to $\hbar\Delta\omega = \mu_\mathrm{B} B(\Delta m_L + \Delta m_S)$. With the selection rule $\Delta m_S = 0$, $\Delta m_L = 0, \pm 1$ taken into account, we see that the each line splits into three—two displaced lines and one undisplaced; hence, the Zeeman effect becomes simple in a strong magnetic field. This phenomenon is called the *Paschen–Buck effect*.

P6.4. Elements of Physics of Molecules

A molecule is a bound state of an electroneutral system consisting of nuclei of several atoms and of electrons surrounding them. The molecules of two atoms are said to be diatomic, the molecules of three and more atoms are said to be polyatomic. The molecule with insufficient or excessive number of electrons is called a molecular ion.

We consider only the simplest diatomic systems such as the hydrogen molecule H_2 and the hydrogen molecular ion H_2^+.

P6.4.1. Classification of Levels of a Molecule

► **Energy hierarchy.** Like an atom, a molecule is a microscopic system and has discrete energy levels corresponding to its ground and bound excited states. But, in contrast with an atom, in addition to the electron energy corresponding to possible states of electrons in the field of stationary nuclei, there also are two types of motion in a molecule: oscillatory motion related to variations in the distance between the nuclei, and rotation of the molecule as a whole about an axis. The molecule energy (counted from the ground state) can be assumed to be equal to the sum of energies corresponding to the above-listed motions:

$$E = E_{\rm el} + E_{\rm vib} + E_{\rm rot}. \qquad \text{(P6.4.1.1)}$$

There exists a pronounced hierarchy of these energies: $E_{\rm el} \gg E_{\rm vib} \gg E_{\rm rot}$. For outermost shell electrons, $E_{\rm el}$ is of the order of several eV, $E_{\rm vib}$ is of the order of 10^{-2} eV, and $E_{\rm rot}$ is of the order of 10^{-4} eV. Just as $E_{\rm rot}/E_{\rm vib}$, the ratio $E_{\rm vib}/E_{\rm el}$ is equal, in the order of magnitude, to $\sqrt{m/M}$, where m is the mass of electron, and M is the mass of the nucleus; that is, for heavy molecules, the energy hierarchy increases.

In the first approximation, all three energies are quantized independently, and to each of them there corresponds its own system of discrete levels. The energy spectrum of a molecule is the set of distant electron levels, more closely located vibrational levels, and even more closely located rotational levels. The system of levels is schematically shown in Fig. P6.9 for two neighboring electron levels, where the letters v and J denote the vibrational and rotational quantum numbers.

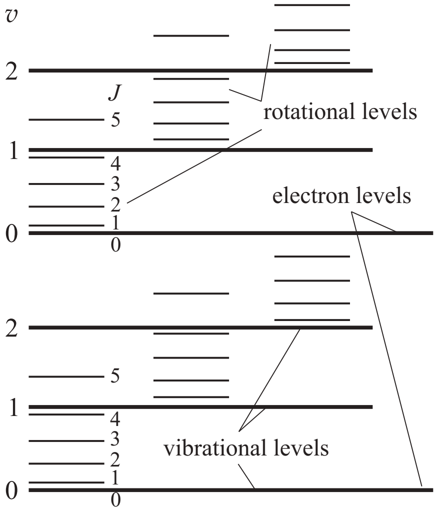

Figure P6.9. Structure of energy levels of a molecule.

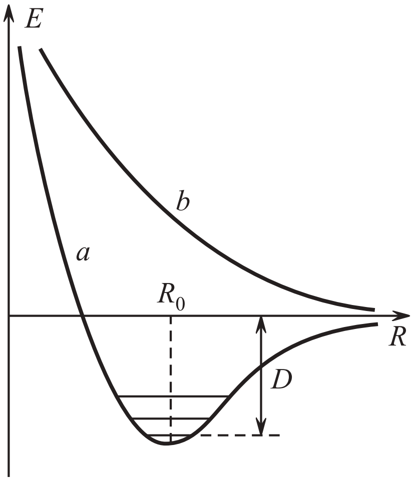

Figure P6.10. Dependence of electron energy of a molecule on the distance between the atoms.

► **The wave function of a molecule.** The wave function of a molecule can be represented as the product of wave functions of motions of three types. Each of these functions, just as the corresponding values of energies, can be obtained by solving the corresponding Schrödinger equation. First, one solves the Schrödinger equation for electrons for a given distance R between the nuclei; this energy depends on R as on a parameter. The energy of interaction between the nuclei is equal to the sum of the electron energy and the energy of the Coulomb repulsion of the nuclei. For some electron levels, this energy has the form of a potential well (curve a in Fig. P6.10), for the other levels (curve b), it has the form of a monotone dependence.

In the first case, there exists a bound state, in the second case, there is no bound state. For the electron ground state, the interaction energy has the form of a potential well, and the depth of this well has the meaning of the molecule dissipation energy. (More precisely, the dissipation energy D is less than the well depth by the energy of the zero-point vibrational level counted from the well depth equal to $\hbar\omega/2$.)

After the dependence of the energy of interaction between the nuclei on the distance R is calculated, one solves the problem about oscillations near the bottom of the obtained potential well and about rotations of a molecule with the distance between the nuclei equal to the average equilibrium distance R_0.

P6.4.2. Electron Energy. Theory of Chemical Bond

► **One electron in the field of two nuclei**. We begin with a molecular ion H_2^+. In this case, the potential of two protons for the motion of a single electron has a cylindrical symmetry with respect to the molecule axis but does not have the spherical symmetry intrinsic to the hydrogen atom. Moreover, the angular momentum projection on the molecule axis is preserved, but the angular momentum magnitude is not. The energy levels are characterized by the magnitude of the magnetic quantum number m_l, which is denoted by the letter λ: $\lambda = |m_l| = 0, 1, \ldots$. The electron states with $\lambda = 0, 1, 2, \ldots$ are denoted by the letters $\sigma, \pi, \delta, \ldots$ and are called *molecular orbitals*. With spin taken into account, the σ-state is twice degenerate, and the states with $\lambda > 0$ are fourfold degenerate. In the σ-state, the electron density is concentrated in the region between the nuclei, the nuclei are attracted to the electron cloud, and thus the attraction between the nuclei described by the potential well a in Fig. P6.10 is realized (for small R, repulsion between the nuclei predominates). The dissipation energy H_2^+ is equal to 2.65 eV. For π-electrons, the electron density has the form of two clouds located between the nuclei symmetrically with respect to a plane passing through the molecule axis, and the electron density in the plane itself is zero.

► **Hydrogen molecule.** In the hydrogen molecule, the chemical bond is realized not by one but by two electrons that are attracted to the nuclei, but repulse from one another. Just as in the case of any fermions, the wave function of two electrons must change the sign in their permutation. If the spin wave function is symmetric with respect to a particle permutation (the spins are parallel, and hence $S = 1$), then the coordinate wave function is antisymmetric and conversely. The state with antiparallel spins ($S = 0$) and a symmetric coordinate wave function has a lesser energy. In this state, the probability for an electron to stay in the space between the nuclei is sufficiently large and the nuclei are attracted to the electron cloud located between them.

Thus, the chemical bond is realized by two electrons with antiparallel spins. The dependence of the potential electrostatic energy on the distance between the nuclei has the form of curve a in Fig. P6.10. The hydrogen molecule dissociation energy is equal to 4.5 eV; this is less than the doubled energy of dissociation of the molecular ion H_2^+ because of the electrostatic repulsion between the electrons. For electrons with parallel spins, the dependence of the energy on the distance between the nuclei has the form of curve b in Fig. P6.10; that is, no bound state is formed.

This mechanism is called the *covalent (homeopolar) chemical bond*. It is realized not only in the hydrogen molecule but by any pair of outermost shell electrons in complex atoms combined in a molecule. For diatomic molecules consisting of different atoms, there exists another mechanism, called the *ionic (heteropolar) chemical bond*, in which one of the atoms gives part of the outermost shell electrons to another atom that has vacancies in the outermost subshell. As a result, two ions are formed that, at large distances, are attracted by the Coulomb law and, at small distances, begin to repulse. Intermediate cases between the covalent and ionic bonds are possible.

► **Exchange energy.** The difference in the energies of the symmetric and antisymmetric states has the general character and is a consequence of the so-called *exchange interaction*, which is a pure quantum phenomenon. The meaning of this phenomenon is reduced to the following. We construct a coordinate wave function of two electrons in the field of two nuclei from the one-electron wave functions $\psi_\alpha(\mathbf{r}_1)$ and $\psi_\beta(\mathbf{r}_2)$. This function can be either symmetric or antisymmetric (depending on the form of the spin function):

$$\varphi(\mathbf{r}_1,\mathbf{r}_2)=\frac{1}{\sqrt{2}}\bigl(\psi_\alpha(\mathbf{r}_1)\psi_\beta(\mathbf{r}_2)\pm\psi_\alpha(\mathbf{r}_2)\psi_\beta(\mathbf{r}_1)\bigr).$$

The average potential energy in these states is equal to

$$\begin{aligned}\langle U\rangle&=\int\varphi^*\widehat{U}\varphi\,d^3\mathbf{r}_1d^3\mathbf{r}_2\\&=\int|\psi_\alpha(\mathbf{r}_1)|^2\widehat{U}|\psi_\beta(\mathbf{r}_2)|^2d^3\mathbf{r}_1d^3\mathbf{r}_2\pm\int\psi_\alpha(\mathbf{r}_1)\psi_\beta^*(\mathbf{r}_1)\widehat{U}\psi_\beta(\mathbf{r}_2)\psi_\alpha^*(\mathbf{r}_2)d^3\mathbf{r}_1d^3\mathbf{r}_2,\end{aligned}$$

where $\widehat{U}$ is the potential energy of interaction of the electrons between each other and with the nuclei. The first term corresponds to the usual classical interaction of electron clouds and the nuclei. The second term is called the *exchange energy*, it is different from zero only in the case where the electron wave functions overlap. In the case of hydrogen molecule, the exchange energy is negative, and therefore, the state with a symmetric coordinate wave function has a lesser energy.

The exchange energy manifests itself in many quantum systems, in particular, it is responsible for the difference between the energies of the singlet and triplet states of electrons in the helium atom.

The chemical bond is usually approximately calculated by one of the two methods. In the *method of electron pairs*, for the initial one-particle wave functions $\psi(\mathbf{r}_1)$ and $\psi(\mathbf{r}_2)$ one takes the atomic wave functions of electrons localized near the first and second nuclei. As the nuclei approach each other, the wave functions begin to overlap, the exchange energy increase and a bound state is formed. In the *method of molecular orbitals*, the electrons successively fill the free σ, π, δ molecular states with different spins and thus form closed electron shells.

P6.4.3. Vibrational and Rotational Spectra

► **Vibrational levels.** The energy of interaction between the nuclei, shown in Fig. P6.10, has the minimum for $R=R_0$. Near this point, the energy can be expanded in a series and represented as

$$U(R)=U(R_0)+\tfrac{1}{2}k(R-R_0)^2+\cdots,$$

where $k=d^2U/dR^2\big|_{R=R_0}$ is a quasi-elastic constant. The cyclic frequency of small oscillations of a molecule near the equilibrium is equal to $\omega_{\rm vib}=\sqrt{k/\mu}$, where $\mu=m_1m_2/(m_1+m_2)$ is the reduced mass of a diatomic molecule (see Subsection P1.6.4). For permitted values of the oscillator energy, quantum mechanics gives values

$$E-U(R_0)=\hbar\omega_{\rm vib}(v+\tfrac{1}{2}),\qquad v=0,1,2,\dots.$$

For large v, the terms neglected in $U(R)$ become essential. If these are taken into account, the distance between the levels begins to decrease.

► **Rotational levels.** The rotational energy levels of a molecule are obtained as eigenvalues of the operator

$$\widehat{H}=\frac{\widehat{L}^2}{2I},\tag{P6.4.3.1}$$

where $\widehat{L}$ is the operator of angular momentum and I is the moment of inertia of the molecule with respect to the center of mass. A rotating system with a constant moment of inertia, which is described by the Hamiltonian (P6.4.3.1), is called a *rotator*. The permitted values of the rotator energy are

$$E = \frac{\hbar^2}{2I} J(J+1), \qquad J = 0, 1, 2, \dots .$$

The frequency of radiation or absorption is obtained from the relation

$$\hbar\omega = \Delta E_{\text{el}} + \Delta E_{\text{rot}} + \Delta E_{\text{vib}}, \qquad \text{(P6.4.3.2)}$$

where the first term corresponds to the transition between distinct electron levels, the second term corresponds to the variation in the vibrational state, and the third term corresponds to the variation in the state of rotation.

If the electron state does not vary in the transition ($\Delta E_{\text{el}} = 0$), then one observes either the pure rotational spectrum (if $\Delta E_{\text{vib}} = 0$) or the vibrational-rotational spectrum. For the pure rotational and vibrational-rotational transitions, one has the selection rule $\delta v = \pm 1$ and $\Delta J = \pm 1$.

For the pure rotational transitions ($\Delta E_{\text{el}} = 0$ and $\Delta E_{\text{vib}} = 0$), the variation in the rotational energy is equal to

$$\Delta E_{\text{rot}} = E(J+1) - E(J) = \frac{\hbar^2}{I}(J+1).$$

One can see that the rotational transitions form a system of close equidistant spectral lines ω_{rot}, $2\omega_{\text{rot}}$, …, where $\omega_{\text{rot}} = \hbar^2/I$ (Fig. P6.11). The rotational lines can be resolved only by devices with a very high resolving power, but usually they are perceived as lines of finite width. The pure rotational transitions correspond to absorption in the far infrared region (hundreds and thousands of μm).

The vibrational transitions (for $\Delta E_{\text{el}} = 0$) are observed in absorption in the near infrared region (of several μm). Since, in this case, one usually has $\Delta E_{\text{rot}} \neq 0$, one observes a vibrational-rotational band of finite width rather than an infinitely thin line. With the selection rules taken into account, we obtain

$$\Delta E_{\text{vib}} + \Delta E_{\text{rot}} = \hbar(\omega_{\text{vib}} \pm n\omega_{\text{rot}}), \qquad (n = 1, 2, \dots).$$

Hence, the distance between neighboring lines at the center of the band is equal to $2\omega_{\text{rot}}$ (Fig. P6.12).

The character of the vibrational-rotational spectra becomes significantly more complicated if the fact that the oscillations are anharmonic is taken into account. As $R - R_0$ increases, the potential energy curve (Fig. P6.10) becomes nonsymmetric. First of all, this results in displacement of vibrational levels; that is, as v increases, the vibrational levels approach one another. Moreover, in the state with larger v, the average distance between the atoms increases, and hence the moment of inertia increases with increasing v, while ω_{vib} decreases. This results in that the lines in the vibrational-rotational band cease to be equidistant.

But if $\Delta E_{\text{el}} \neq 0$, then one observes the electron spectra of emission and absorption in the visible and ultraviolet regions. To each electron transition, there corresponds a system of vibrational-rotational bands. In contrast to the vibrational-rotational transitions, in the case of transitions with variation in the electron state, the selection rules $\Delta v = \pm 1$ is not satisfied; that is, the transitions between levels with any values of vibrational quantum numbers are possible. Therefore, one can observe not a single vibrational-rotational band, but a series of bands separated from one another by ω_{vib}.

We note that diatomic molecules with zero dipole moment (for example, N_2 and O_2) do not have vibrational-rotational spectra. Just for this reason, such molecules do not absorb in the infrared (IR) region of the spectrum, and the absorption by air of the IR radiation occurs because the air contains the molecules H_2O and CO_2.

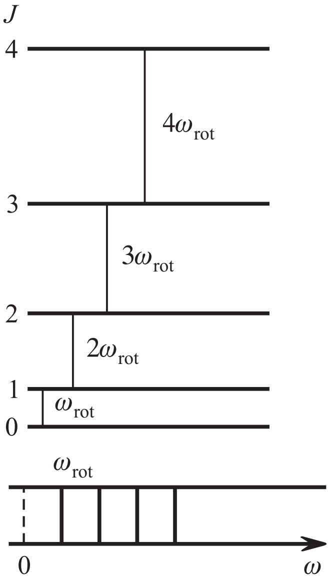

Figure P6.11. Rotational spectrum.

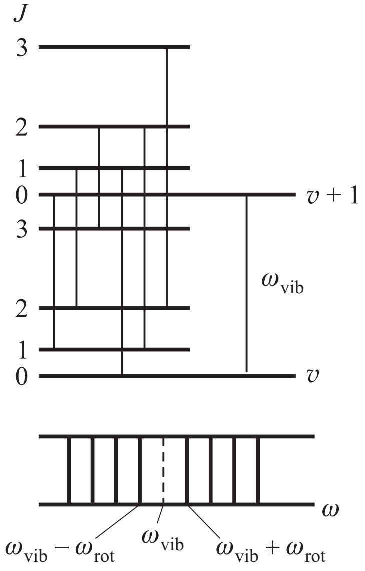
J
3
2
1
0
3
2
1
0
$v+1$
v
$\omega_{\rm vib}$
ω
$\omega_{\rm vib}-\omega_{\rm rot}$
$\omega_{\rm vib}$
$\omega_{\rm vib}+\omega_{\rm rot}$

Figure P6.12. Vibrational-rotational spectrum.

Bibliography for Chapter P6

Alenitsyn, A. G., Butikov, E. I., and Kondratyev, A. S., *Concise Handbook of Mathematics and Physics*, CRC Press, Boca Raton, Florida, 1998.

Benenson, W., Harris, J. W., Stocker, H., and Lutz, H. (Eds.), *Handbook of Physics*, Springer-Verlag, New York, 2002.

Blokhintsev, D. I., *Basics of Quantum Mechanics* [in Russian], Lan, St. Petersburg, 2006.

Bohm, D., *Quantum Theory*, Dover Publications, New York, 1989.

Bransden, B. H. and Joachain, C. J., *Physics of Atoms and Molecules, 2nd Edition*, Benjamin Cummings, New York, 2003.

Budker, D., Kimball, D., and DeMille, D., *Atomic Physics: An Exploration through Problems and Solutions, 2nd Edition*, Oxford University Press, New York, 2008.

Demtröder, W., *An Introduction to Atomic and Molecular Physics*, Springer, Berlin, 2005.

Dirac, P. A. M., *Lectures on Quantum Mechanics*, Dover Publications, New York, 2001.

Foot, C. J., *Atomic Physics (Oxford Master Series in Atomic, Optical and Laser Physics)*, Oxford University Press, New York, 2005.

Griffiths, D. J., *Introduction to Quantum Mechanics (2nd Edition)*, Benjamin Cummings, San Francisco, California, 2004.

Heisenberg, W., *The Physical Principles of the Quantum Theory*, Dover Publications, New York, 1930.

Karlov, N. V. and Kirichenko, N. A., *Introductory Chapters of Quantum Mechanics* [in Russian], Fizmatlit, Moscow, 2006.

Landau, L. D. and Lifshitz, E. M., *Quantum Mechanics Non-Relativistic Theory, Third Edition: Volume 3, (Course of Theoretical Physics Series)*, Butterworth-Heinemann, Oxford, England, 1981.

Liboff, R., *Introductory Quantum Mechanics, 4th Edition*, Addison Wesley, San Francisco, California, 2002.

Mahan G. D., *Quantum Mechanics in a Nutshell*, Princeton University Press, Princeton, New York, 2008.

Prokhorov, A. M. (Ed.), *Encyclopedia of Physics, Volumes 1–5* [in Russian], Bolshaya Russkaya Encyclopedia, Moscow, 1998.

Prokhorov, A. M. (Ed.), *Encyclopedic Dictionary of Physics* [in Russian], Sovetskaya Encyclopedia, Moscow, 1984 (dic.academic.ru).

Sivukhin, D. V., *General Physics, Volume 5, Atomic and Nuclear Physics* [in Russian], Fizmatlit, Moscow, 2006.

Sivukhin, D. V. and Yakovlev, I. A., *A Collection of Problems, Volume 5, Atomic Physics. Nuclear and Particle Physics* [in Russian], Fizmatlit, Moscow, 2006.

Woan, G., *The Cambridge Handbook of Physics Formulas*, Cambridge University Press, Cambridge, England, 2003.

Chapter P7

Quantum Theory of Crystals

Although the laws of quantum mechanics act on an atomic scale, they also affect the macroscopic properties of gases, liquids, and rigid bodies. Quantum mechanics and quantum statistics—that is, statistics of indistinguishable particles—allowed one to obtain correct and consistent descriptions of thermal, magnetic, and many other properties of large ensembles of particles. Here a special place is occupied by quantum theory of crystals, which allowed one to remove several unsolvable contradictions of classical theory, to obtain quantitative description of electromagnetic properties of some classes of matter, to explain the superconductivity and superfluidity, etc.

P7.1. Quantum Statistical Distributions and Their Application

P7.1.1. Elements of Quantum Statistics

Quantum statistics studies the statistical properties of large ensembles of identical indistinguishable particles, fermions and bosons. Statistics of classical (i.e., distinguishable) particles is called the *Maxwell–Boltzmann statistics* (see Subsection P2.5.4). Statistics of *fermions*—that is, indistinguishable particles that cannot share identical states—is called the *Fermi–Dirac statistics*. Statistics of *bosons*—that is, indistinguishable particles that can share identical states—is called the *Bose–Einstein statistics*.

► **Statistical distributions.** Statistics of each type is characterized by its own *statistical distribution* of particles over states. A statistical distribution tells the average number of particles in a given state under the condition that the system is in thermal equilibrium. The most important property of all distributions is that they are functions of a single parameter, the particle energy:* $\langle N\rangle = f(\varepsilon)$.

In quantum mechanics, the notion of a state of a single particle is assumed to be quite definite, but only under certain conditions. For the energy levels of a particle system to be discrete, the system must be localized in a finite region of space. Usually, it is assumed that the system is closed in a box of a sufficiently large volume V. Moreover, the notion of a one-particle state with a certain energy assumes that there is no interparticle interaction. This last difficulty can be overcome by introducing a *self-consistent external field*, which includes the averaged field of the other particles.

For an ensemble of classical particles, one introduces the notion of an *elementary cell of phase space*. The phase space is the space of three coordinates and three momenta. In classical physics, to each state there corresponds one point of the phase space. In this case, the elementary cell corresponding to a single state is introduced for convenience in statistical description, and it can have an arbitrary (sufficiently small) volume. In quantum mechanics, a state of a particle cannot be represented by a point in the phase space, because there are uncertainties in its position and momentum, which are related by the Heisenberg relations $\Delta x \Delta p_x \sim \hbar$, etc. One can see that the volume $\Delta x \Delta y \Delta z \Delta p_x \Delta p_y \Delta p_z \sim \hbar^3$ must correspond to each individual state. A more precise analysis shows that the volume of an elementary cell is equal to $(2\pi\hbar)^3$. To understand how many states there are contained in a finite region of the phase space, it is necessary to divide its volume by the volume of the elementary cell. Such an approach in which the elementary volume of a single state for a classical particle is chosen from quantum considerations is said to be *semiclassical*.

---

* In what follows, the energy of an individual particle is denoted by ε and the energy of the system, by E.

► **Microscopic and macroscopic states. Statistical weight.** To introduce the microscopic state (microstate) of a system means to define the state of all its particles at a given time instant. In classical physics, it is required to determine the coordinates and momenta of all particles—that is, the positions of all particles in the phase space. In the semiclassical approach, one finds the distribution of particles over elementary cells of volume $(2\pi\hbar)^3$ each. In quantum mechanics, one determines the distribution of particles over states with a certain energy (for the motion in a self-consistent field). For convenience of our presentation, we combine the last two cases and also call the states of particles with a certain energy *cells*. Due to interparticle interaction and external forces, the microstate of the system varies all the time and one can observe only the average characteristics in terms of which the macrostate can be determined.

To describe the macrostate quantitatively, let us divide the entire set of cells into *boxes*. Each box contains a large amount of cells with very close energy values lying within a narrow energy interval $(\varepsilon, \varepsilon + \Delta\varepsilon)$. Denote the number of particles in the ith box by N_i. We assume that the macrostate is determined by a set of numbers N_i, and hence all microstates with identical sets of these numbers belong to the same macrostate.

Without loss of generality, we consider a system of particles contained in a box with rigid thermally insulated walls, so that its total energy remains unchanged.

Why does this assumption not restrict the generality of consideration? First of all, we are interested in the properties of the equilibrium macrostate, and in this case, the energy of a large ensemble of particles can be determined with a very high accuracy under any external conditions. For example, for a system of N particles in a thermostat, the relative deviation from the average energy is $1/\sqrt{N}$. Let us explain this with an example of an ideal gas. If the average energy of a single particle is equal to $\langle\varepsilon\rangle$ and the root mean square deviation from the average energy is $\delta\varepsilon = \sqrt{\langle(\varepsilon - \langle\varepsilon\rangle)^2\rangle}$, then the average energy of the system is equal to $\langle E\rangle = N\langle\varepsilon\rangle$, and the root mean square deviation is

$$\sqrt{\bigl\langle (E - \langle E\rangle)^2 \bigr\rangle} = \sqrt{\Bigl\langle \Bigl[\sum_i (\varepsilon_i - \langle\varepsilon\rangle)\Bigr]^2 \Bigr\rangle} = \sqrt{\sum_i \bigl\langle (\varepsilon_i - \langle\varepsilon\rangle)^2 \bigr\rangle} = \sqrt{N}\,\delta\varepsilon;$$

we have taken into account the fact that the deviations of the energies of distinct particles from the average value are mutually independent.

► **The principal postulate of statistical physics** for a system with a fixed energy states that all the microstates admissible for it (that is, those with a given value of total energy) are realized with equal probability. This postulate is closely related to the so-called *ergodic hypothesis* stating that, in the process of time evolution, the system passes through all admissible microstates.

The main conclusion from the principal postulate is the following: the greatest part of time the system is in the macrostates to which there corresponds the greatest number of microstates. The number of microstates corresponding to a given macrostate is called its *statistical weight* and is denoted by W. The macrostate with maximum W is called the *equilibrium macrostate*. According to the Boltzmann postulate (see Subsection P2.4.4), the system entropy in a given macrostate is related to its statistical weight as follows:

$$S = k \ln W.$$

The deviations from the equilibrium state are called *fluctuations*. We note that, because of an enormous number of particles, a large deviation in the system has an extremely small probability; that is, only small fluctuations are practically observable. If the system was initially in a highly non-equilibrium state, then, as the system evolves, it must move from macrostates with lesser statistical weights (lesser entropy) to a macrostate with a larger entropy approaching the equilibrium (the second law of thermodynamics).

► **Equilibrium macrostates.** In order to find out which particle distribution over boxes, $\{N_1, N_2, \dots\}$, corresponds to equilibrium, one must calculate $W(N_1, N_2, \dots)$ or $S/k = \ln W$ and take the maximum of this expression (taking into account the conditions that $\sum_i \varepsilon_i N_i = E$ and $\sum_i N_i = N$). The expressions for the statistical weight and entropy are significantly different for particles with different types of statistical behavior. Below are the expressions of the statistical weight for three different statistical models:

for the Maxwell–Boltzmann statistics,

$$W = \prod_i g_i^{N_i} \frac{N!}{N_1!\, N_2! \,\dots},$$

for the Bose–Einstein statistics,

$$W = \prod_i \frac{(N_i + g_i - 1)!}{N_i!\,(g_i - 1)!},$$

and for in the Fermi–Dirac statistics,

$$W = \prod_i \frac{g_i!}{N_i!\,(g_i - N_i)!},$$

where g_i stands for the number of cells in the ith box.

The equilibrium state $\{N_i\}$ corresponding to the maximum statistical weight under the condition that E and N are fixed can be found using the method of Lagrange multipliers. (It is necessary to find the maximum of the expression $(S/k) - \alpha \sum_i N_i - \beta \sum_i \varepsilon_i N_i$ under the assumption that all N_i are independent. Then the Lagrange multipliers α and β are determined from the equations $\sum_i N_i = N$ and $\sum_i \varepsilon_i N_i = E$.) We obtain the following expressions:

for the Maxwell–Boltzmann statistics,

$$N_i = g_i e^{-\alpha} e^{-\beta \varepsilon_i};$$

for the Bose–Einstein statistics,

$$N_i = \frac{g_i}{e^{\alpha} e^{\beta \varepsilon_i} - 1}$$

and for the Fermi–Dirac statistics,

$$N_i = \frac{g_i}{e^{\alpha} e^{\beta \varepsilon_i} + 1}.$$

The average number of particles per cell is obtained by dividing these expressions by g_i.

The physical meaning of the multipliers α and β can be clarified from thermodynamical considerations. We write the maximum condition for the expression $(S/k) - \alpha N - \beta E$

in the form $T\,dS/kT - \beta\,dE - \alpha\,dN = 0$ and compare with the thermodynamical identity $T\,dS - dU - \mu\,dN = 0$, which holds for variations in the state under constant volume. We obtain $\beta = 1/kT$ and $\alpha = \mu/kT$, where μ is the *chemical potential*.

Finally, the quantum distributions take the form:

Bose–Einstein distribution:

$$\langle N\rangle = f_B(\varepsilon) = \frac{1}{e^{(\varepsilon-\mu)/kT} - 1}. \qquad \text{(P7.1.1.1)}$$

Fermi–Dirac distribution:

$$\langle N\rangle = f_{\rm F}(\varepsilon) = \frac{1}{e^{(\varepsilon-\mu)/kT} + 1}. \qquad \text{(P7.1.1.2)}$$

The chemical potential μ for a system with a fixed number of particles, N, is calculated as

$$N = \sum_i g_i f(\varepsilon_i),$$

where $f_{(}\varepsilon_i)$ is the corresponding distribution (Bose–Einstein or Fermi–Dirac). But if the number of particles in the system is not fixed, then the chemical potential must be set to zero.

For $\varepsilon - \mu \gg kT$, both quantum distributions become the classical *Maxwell–Boltzmann distribution*:

$$\langle N\rangle = Ae^{-\varepsilon/kT}. \qquad (1)$$

This is related to the fact that the average number of particles in a cell is small, and the exclusion rules and the particle identity become insignificant.

Usually, the least possible value of energy (energy of the lowest level) is set to be zero.

► Properties of the Bose–Einstein distribution.

1) Bosons are particles with integer value of the intrinsic angular momentum (spin). Examples of bosons are π-mesons, photons, and phonons.

2) The chemical potential of bosons cannot be positive: $\mu \le 0$.

3) As $T \to 0$, the chemical potential tends to zero and vanishes at $T = 0$.

4) At $T = 0$, all particles are in the lowest energy state with $\varepsilon = 0$. At finite but sufficiently low temperatures, a finite fraction of all particles is in the lowest energy state.

This phenomenon is called *Bose–Einstein condensation*, and the ensemble of particles in the lowest state is called a *Bose–Einstein condensate*. Condensation of bosons begins at a certain critical temperature (the condensation temperature). This phenomenon plays an important role in the explanation of phenomena such as superfluidity and superconductivity.

5) An example of the Bose–Einstein distribution is the Planck formula for the photon distribution over energies (Subsection P5.4.1). The number of photons is not fixed; therefore, $\mu = 0$.

► Properties of the Fermi–Dirac distribution.

1) Fermions are particles with half-integer spin. Examples of fermions are electrons, protons, neutrons, and μ-mesons.

2) The chemical potential of the Fermi gas can be either positive or negative. At low temperatures, the chemical potential is positive. The chemical potential of fermion gas for $T = 0$ is called the *Fermi energy* and is denoted by $\varepsilon_{\rm F}$.

3) At $T = 0$, the energy-dependence of the average number of particles in a cell $\langle N\rangle$ has the form of a sharp step: for $\varepsilon < \varepsilon_{\rm F}$, the cells are filled with particles ($\langle N\rangle = f_{\rm F}(\varepsilon) = 1$), and for $\varepsilon > \varepsilon_{\rm F}$, the cells are free ($f_{\rm F}(\varepsilon) = 0$).

4) At low temperatures ($kT \ll \varepsilon_{\rm F}$), the dependence $f_{\rm F}(\varepsilon)$ has the form of a smoothed step (Fig. P7.1). The spreading width is equal to kT in the order of magnitude. For $\varepsilon = \mu$, the average number of particles is equal to 1/2. For $kT \ll \varepsilon_{\rm F}$, the particle distribution over energies radically differs from the classical distribution; in this case, the gas is said to be *degenerate*. For $kT \gg \varepsilon_{\rm F}$, the fermion gas does not differ noticeably from the classical gas.

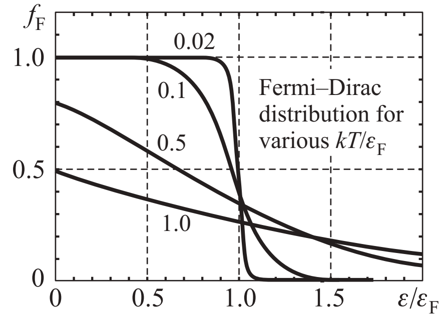

Figure P7.1. Fermi–Dirac distribution for various $kT/\varepsilon_{\rm F}$ (for ideal fermion gas).

► **Energy density of states.** The more specific properties of an ensemble of bosons or fermions depend on the physical properties of the system under study, mainly, of the form of the function $G(\varepsilon)$ characterizing the number of states Δg that correspond to a narrow energy interval $\Delta\varepsilon$ calculated per unit volume:

$$\frac{\Delta g}{V} = G(\varepsilon)\Delta\varepsilon. \qquad \text{(P7.1.1.4)}$$

This function is called the *energy density of states* or simply the *density of states* (cf. Subsection P5.4.1). For example, the number of particles per unit volume is expressed in terms of the density of states:

$$n = \frac{N}{V} = \int_0^\infty f(\varepsilon)G(\varepsilon)\,d\varepsilon \qquad \text{(P7.1.1.5)}$$

as well as the energy of particles per unit volume:

$$\frac{E}{V} = \int_0^\infty f(\varepsilon)\varepsilon G(\varepsilon)\,d\varepsilon. \qquad \text{(P7.1.1.6)}$$

P7.1.2. Electron Gas in Metal

The classical electron theory of metal conductivity originates from the idea that the electrons in metal can be considered as the classical ideal gas (Subsection P3.7.3). One of the main difficulties encountered by this theory was that the electron contribution to the heat capacity predicted by this theory was equal to the heat capacity of a monatomic ideal gas $C_{\rm el} = \frac{3}{2}kN$, where N is the number of free electrons. It is precisely this value that must be the difference between the heat capacities of a metal and a dielectric with similar crystal lattices. But experiment showed that the electron heat capacity at room temperatures is much less than kN (approximately, by a factor of 100) and varies proportionally to the temperature. These and some other facts can naturally be explained by the *Sommerfeld theory*, where the electrons in metal are considered as a quantum gas of fermions.

► **Density of states of electron gas.** We assume that, inside the metal, the electrons are free and the work required for an electron to leave the metal is infinitely large; that is, we treat them as particles in a potential well with infinitely high walls. The boundary conditions for the wave functions $\psi_x = A\sin k_x x$ have the form $k_x L_x = \pi n$, where L_x is the potential well dimension (that is, the metal sample dimension) along the axis x, k_x is the projection of the wave vector related to the electron momentum as $\mathbf{k} = \mathbf{p}/\hbar$, and n is a positive integer (see Subsection P6.2.2). The boundary conditions for the wave vector identically coincide with those arising in the case of consideration of electromagnetic oscillations in a closed cavity (Subsection P5.4.1). Therefore, the density of states in this case has the same form:

$$\frac{dg}{V} = \gamma \frac{k^2\,dk}{2\pi^2},$$

where V is the vessel volume and γ is a multiplier taking the intrinsic degrees of freedom into account. In the case of electrons, $\gamma = 2$ takes account of the two possible projections of spin. To pass to the electron momentum, we must replace $k^2\,dk$ by $p^2\,dp/\hbar^3$:

$$\frac{dg}{V} = \frac{p^2\,dp}{\pi^2\hbar^3}.$$

We rewrite this formula as

$$dg = \gamma \frac{V\,4\pi p^2\,dp}{(2\pi\hbar)^3}.$$

One can see that to determine the number of states, it is necessary to divide the phase space volume by the elementary cell volume $(2\pi\hbar)^3$.

To pass to the energy $\varepsilon = p^2/2m$, it is necessary to make the change $p = \sqrt{2m\varepsilon}$, $p\,dp = m\,d\varepsilon$:

$$\frac{dg}{V} = \frac{m\sqrt{2m}\sqrt{\varepsilon}\,d\varepsilon}{\pi^2\hbar^3}, \qquad G(\varepsilon) = \frac{m\sqrt{2m}\sqrt{\varepsilon}}{\pi^2\hbar^3}.$$

► **Fermi energy of electrons.** At $T = 0$, all the energy levels from zero to $\varepsilon = \varepsilon_{\rm F}$ are filled. To relate the Fermi energy to the electron concentration n, it is necessary to express the total number of electrons per unit volume and equate with the concentration

$$\int_0^{\varepsilon_{\rm F}} G(\varepsilon)\,d\varepsilon = \frac{(2m)^{3/2}\varepsilon_{\rm F}^{3/2}}{3\pi^2\hbar^3} = n.$$

This implied the Fermi energy:

$$\varepsilon_{\rm F} = \frac{\hbar^2(3\pi^2 n)^{2/3}}{2m}. \qquad \text{(P7.1.2.1)}$$

The momentum of an electron with Fermi energy is determined by the relation $\varepsilon_{\rm F} = p_{\rm F}^2/2m$ and is equal to

$$p_{\rm F} = \hbar(3\pi^2 n)^{1/3}. \qquad \text{(P7.1.2.2)}$$

The surface in momentum space formed by states with $\varepsilon = \varepsilon_{\rm F}$ is called the *Fermi surface*. In the case of free electrons, the Fermi surface is a sphere of radius $p_{\rm F}$.

The density of levels on the Fermi surface is equal to

$$G(\varepsilon_{\rm F}) = \frac{dn}{d\varepsilon_{\rm F}} = \frac{3}{2}\frac{n}{\varepsilon_{\rm F}}.$$

The average energy of electrons at $T = 0$ is equal to

$$\langle\varepsilon\rangle = \frac{\int_0^{\varepsilon_\mathrm{F}} \varepsilon G(\varepsilon)\,d\varepsilon}{\int_0^{\varepsilon_\mathrm{F}} G(\varepsilon)\,d\varepsilon} = \frac{\int_0^{\varepsilon_\mathrm{F}} \varepsilon\sqrt{\varepsilon}\,d\varepsilon}{\int_0^{\varepsilon_\mathrm{F}} \sqrt{\varepsilon}\,d\varepsilon} = \tfrac{3}{5}\varepsilon_\mathrm{F}.$$

For $T > 0$, the chemical potential μ is determined by the equation

$$n = \int_0^\infty f_\mathrm{F}(\varepsilon)G(\varepsilon)\,d\varepsilon.$$

As the temperature increases, μ gradually decreases, but for $kT \ll \varepsilon_\mathrm{F}$, its difference from ε_F is very small: $|\mu - \varepsilon_\mathrm{F}|/\varepsilon_\mathrm{F} \approx (\pi^2/12)(kT/\varepsilon_\mathrm{F})^2$. At a certain temperature T_0 ($kT_0 \approx \varepsilon_\mathrm{F}$), the chemical potential is zero, and for higher temperatures, $\mu < 0$ (the chemical potential is determined from the graph in Fig. P7.1; for $\varepsilon = \mu$, we have $f_\mathrm{F}(\varepsilon) = 0.5$).

To simplify the process of determining the temperature region where the electron gas is degenerate, one introduces the *Fermi temperature* by the relation $kT_\mathrm{F} = \varepsilon_\mathrm{F}$. For $T \ll T_\mathrm{F}$, the electron gas is degenerate. For the majority of metals, T_F is equal to several tens of thousands of degrees, which means that the electron gas in metals is strongly degenerate at room temperatures.

► **Electron heat capacity of metals.** In the case of highly degenerate gas, only those electrons will participate in thermal motion whose energy differs from the Fermi energy by a value of the order of kT. Only such electrons can make transitions to a free state with energy greater than ε_F. The fraction of such electrons is $\sim kT/\varepsilon_\mathrm{F} = T/T_\mathrm{F}$ and the variation in their energy is $\sim kT$. Therefore, as the temperature increases from zero to T, the energy of degenerate electron gas increases by $\sim kNT^2/T_\mathrm{F}$, and the contribution of electrons to the heat capacity $C_\mathrm{el} \sim kN(T/T_\mathrm{F})$ varies proportionally to the temperature. (Precise calculations give the expression $C_\mathrm{el} = (\pi^2/2)kN(T/T_\mathrm{F})$.) The number of free electrons per one mole of metal depends on the valency; for example, in the case of an univalent metal, $N \sim N_\mathrm{A}$.

► **Electrical and thermal conduction of metals.** The transport theory gives the following expressions for the electrical and thermal conductivities of electrons (see also Subsections P3.7.3 and P2.9.2):

$$\sigma = \frac{ne^2\lambda}{m\langle v\rangle}, \qquad \kappa = \tfrac{1}{3}C_\mathrm{el}\langle v\rangle\lambda.$$

Here λ is mean free path and C_el is the heat capacity of electrons per unit volume of metal. Participating in collisions with ions will be only those electrons that are located near the Fermi surface; therefore, $\langle v\rangle \approx v_\mathrm{F} = p_\mathrm{F}/m$, which means that the average speed is almost independent of the temperature. Nevertheless, the *Wiedemann–Franz law* remains true: the ratio $\kappa/\sigma \sim \langle v\rangle^2 C_\mathrm{el}/n$ for metals is a linear function of the temperature. But in contrast with the classical electron gas, where C_el/n is independent of the temperature and $\langle v\rangle^2 \sim T$, the linear dependence on the temperature in the Sommerfeld theory is determined by the electron heat capacity.

The Sommerfeld theory cannot explain the dependence of the electrical conduction of pure metals on the temperature, because, in this theory, just as in the Drude theory of the classical electron gas, the mean free path is a pure empirical parameter. The dependence of the mean free path on the temperature is explained in the theory, where the periodicity of the crystal lattice field and the interaction of electrons with the lattice thermal vibrations (with phonons) are taken into account.

► **Richardson–Dashman formula.** The Sommerfeld theory permits explaining the temperature dependence for the current of thermoelectron emission from the metal surface. The metal surface can be left only by the electrons for which $p_z^2/2m > \varepsilon_{\rm F} + A$, where A is *work of liberation* of electron from metal (the axis z is perpendicular to the surface). Omitting the calculations, we present the answer

$$j = CT^2 \exp(-A/kT),$$

where the constant $C = \dfrac{mek^2}{2\pi^2\hbar^3} = 120\,\text{A}/(\text{m}^2\,\text{K}^2)$ is the same for all metals.

P7.1.3. Heat Capacity of Crystal Lattice. Phonons

► **Dulong–Petit law.** The classical statistical physics predicts that the contribution of the lattice vibrations to the molar heat capacity must be equal to $3R$. Indeed, each atom of the lattice exhibits three independent vibrations near the equilibrium and to each vibrational degree of freedom there corresponds the average energy kT. The energy of vibrations of all atoms is equal to $3NkT = 3\nu RT$, and the molar heat capacity of the lattice need not depend on the temperature and be equal to $3R$. Indeed, at room temperature, the molar heat capacity of many dielectrics is close to $3R$ (*Dulong–Petit law*). But as the temperature decreases, in all cases, one observes a significant decrease in the heat capacity, and at low temperatures, it tends to zero as T^3. Such a behavior of the lattice heat capacity is explained by quantum theory.

► **Einstein theory.** Einstein modeled vibrations of atoms at the nodes of the crystal lattice using $3N$ mutually independent equal quantum harmonic oscillators. The average energy of each such oscillator is equal to (see Subsection P5.4.1)

$$\langle\varepsilon\rangle = \frac{\hbar\omega}{2} + \frac{\hbar\omega}{e^{\hbar\omega/kT} - 1}. \qquad \text{(P7.1.3.1)}$$

For $kT \gg \hbar\omega$, the molar heat capacity of a system of quantum oscillators is equal to $3R$, and as $T \to 0$ it exponentially tends to zero. The Einstein theory explained that the heat capacity decreases with decreasing temperature, but not by the law T^3.

► **Elastic waves.** The difficulties in the Einstein theory originate from the fact that the elastic vibrations of noninteracting oscillators of the same frequency were considered as elementary excitations (of normal oscillatory modes). In fact, due to interaction between neighboring atoms, elastic waves of different frequencies and wavelength propagate in crystal. The elementary excitations are waves with definite values of frequency ω and the wave vector $\mathbf{k}$. The average energy per unit wave with frequency ω is given by the formula (P7.1.3.1). To calculate the number of distinct waves in the interval $d\omega$, it is necessary to solve the following two problems:

1) To find the number of independent oscillations per unit volume element $d^3\mathbf{k}$ in the space of wave vectors. This problem can be solved starting from the boundary conditions. The total number of independent oscillations must be equal to the number of degrees of freedom $3N$. For traveling waves, we use *periodic boundary conditions*: $\xi(L_x) = \xi(0)$, with similar relations for y and z. This leads to

$$k_x L_x = 2\pi m, \quad \ldots, \quad \Delta k_x = \frac{2\pi}{L_x}, \quad \ldots, \quad \Delta k_x \Delta k_y \Delta k_z = \frac{(8\pi)^3}{V}.$$

Finally, for the density of states in the **k**-space, we have

$$dg = \gamma \frac{4\pi k^2\, dk}{\Delta k_x \Delta k_y \Delta k_z} = \gamma \frac{k^2\, dk}{2\pi^2} V. \qquad \text{(P7.1.3.2)}$$

So we get the same formula as for standing waves (see Subsection P5.4.1). The multiplier γ is equal to three in the case of sound waves; it takes account of the two longitudinal waves and one transverse wave.

2) To find the law of dispersion, that is, the dependence $\omega(\mathbf{k})$. To solve this problem, it is necessary to write the equation of motion of separate atoms taking account of their interaction with neighboring atoms. The theoretical calculations are possible only for rather simple models, but in the general case, it is possible to find several important features.

For the wavelength larger than the interatomic distances ($ka \ll 1$), good results are given by the *continuum approximation* in which the medium is assumed to be continuous. In this case, the equations of motion of particles are reduced to the wave equation and the law of dispersion of sound waves has the form:

$$\omega = vk, \qquad \text{(P7.1.3.3)}$$

where v is the velocity of the sound wave propagation (for simplicity, we neglect the difference between the longitudinal and transverse waves). The main contribution to the energy is given by the waves with $\hbar\omega \sim k_\mathrm{B}T$, where k_B is the Boltzmann constant (the probability of formation a wave with $\hbar\omega \gg k_\mathrm{B}T$ is exponentially small); hence the contribution of sound waves (for which $ka \ll 1$ or $\omega \ll v/a$) is the main contribution under the condition that $k_\mathrm{B}T \ll \hbar v/a$, that is, at sufficiently low temperatures. Conversely, at high temperatures, the contribution to the energy from each of the elementary waves is equal to $k_\mathrm{B}T$, and the energy is determined only by the common number of degrees of freedom.

► **Debye theory.** To obtain the correct value of the heat capacity at both high and low temperatures in the framework of a simple model, Debye assumed that the density of states (P7.1.3.2) and the dispersion law (P7.1.3.3) hold for all k up to some maximum value k_D and a corresponding frequency ω_D above which the density of states is zero. The *Debye frequency* ω_D is chosen so that the total number of state is equal to $3N$, which ensures that the Dulong–Petit law holds at high temperatures.

The density of the number of states in the Debye model is expressed as

$$G(\omega)\, d\omega = \begin{cases} \dfrac{3\,\omega^2\, d\omega}{2\pi^2 v^3} V & \text{for } \omega < \omega_\mathrm{D}; \\ 0 & \text{for } \omega > \omega_\mathrm{D}. \end{cases}$$

To take the difference between the transverse and longitudinal waves into account, we must define the average velocity by the relation

$$\frac{3}{v^3} = \left(\frac{1}{v_\parallel^3} + \frac{2}{v_\perp^3} \right).$$

We determine the Debye frequency from the equation

$$3N = \int_0^{\omega_\mathrm{D}} \frac{3}{2} \frac{\omega^2\, d\omega}{\pi^2 v^3} V$$

to obtain

$$\omega_{\rm D} = (6\pi^2 n)^{1/3} v, \qquad k_{\rm D} = (6\pi^2 n)^{1/3}. \tag{P7.1.3.4}$$

The energy of the lattice vibrations in the Debye model is given by the expression

$$E_{\rm cr} = \int_0^{\omega_{\rm D}} \frac{\hbar\omega G(\omega)\,d\omega}{e^{\hbar\omega/k_{\rm B}T} - 1}.$$

To separate the regions of high and low temperatures, it is convenient to introduce the *Debye temperature* by the relation

$$k_{\rm B} T_{\rm D} = \hbar\omega_{\rm D}. \tag{P7.1.3.5}$$

Introducing the dimensionless variable $x = \hbar\omega/k_{\rm B}T$, we obtain

$$E_{\rm cr} = 9Nk_{\rm B}T_{\rm D}\left(\frac{T}{T_{\rm D}}\right)^4 \int_0^{T_{\rm D}/T} \frac{x^3\,dx}{e^x - 1}.$$

For $T \gg T_{\rm D}$, expanding the exponential function, we obtain $E_{\rm cr} = 3k_{\rm B}NT = 3\nu RT$, that is, the Dulong–Petit law. For $T \ll T_{\rm D}$, the integration can be continued to infinity, and it becomes equal to $\pi^4/15$. We see that the energy of the lattice vibrations is proportional to T^4, and for the heat capacity, we obtain the law of T^3:

$$C_V = \frac{12}{5}\pi^4 k_{\rm B} N \left(\frac{T}{T_{\rm D}}\right)^3.$$

Example 1. Calculate the energy of zero-point vibrations of the crystal lattice in the Debye model.
Solution. This energy is equal to

$$E_0 = \int_0^{\omega_{\rm D}} \frac{1}{2}\hbar\omega G(\omega) d\omega = \frac{3\hbar\omega_{\rm D}^4}{16\pi^2 v^3} N = \frac{9}{8} N\hbar\omega_{\rm D} = \frac{9}{8} Nk_{\rm B}T_{\rm D}.$$

Example 2. Calculate the energy of the one-dimensional crystal lattice vibrations in the Debye model (only for longitudinal vibrations).

Solution. For the density of states, $G(k) = L/(2\pi)$, we obtain

$$2G(k)\,dk = G(\omega)\,d\omega = \begin{cases} \dfrac{d\omega}{\pi v} L & \text{for } \omega < \omega_{\rm D}, \\ 0 & \text{for } \omega > \omega_{\rm D} \end{cases}$$

(there are two traveling waves for one ω), where L is the length of the chain. Since the total number of states is equal to N, we have

$$\omega_{\rm D} = \pi \frac{N}{L} v = \frac{\pi}{a} v. \tag{P7.1.3.6}$$

For the crystal energy, we obtain

$$E_{\rm cr} = \int_0^{\omega_{\rm D}} \frac{\hbar\omega G(\omega)\,d\omega}{e^{\hbar\omega/k_{\rm B}T} - 1} = Nk_B T_{\rm D}\left(\frac{T}{T_{\rm D}}\right)^2 \int_0^{T_{\rm D}/T} \frac{x\,dx}{e^x - 1}.$$

For $T \ll T_{\rm D}$, the energy of vibrations is proportional to T^2. For $T \gg T_{\rm D}$, $E_{\rm cr} = k_{\rm B}NT$.

► **One-dimensional chain. Dispersion law.** To understand what consequences the discreteness in the mass distribution may lead to, we consider an infinitely long chain of atoms. We assume that the atomic mass is m, the distance between the neighboring atoms in equilibrium is a, and an elastic force with rigidity κ acts between the neighboring atoms. The equation of motion of the sth atom has the form

$$\xi_s'' = \omega_0^2(\xi_{s+1} - 2\xi_s + \xi_{s-1}),$$

where ξ_s is the displacement of the sth atom from equilibrium, $\omega_0 = \sqrt{\kappa/m}$. We seek the solution in the form $\xi_s = A\exp[i(ksa - \omega t)]$. For the relation between ω and k, we obtain

$$\omega = 2\omega_0 \left|\sin\frac{ka}{2}\right|. \qquad \text{(P7.1.3.7)}$$

The velocity of the sound wave ($ka \ll 1$) is equal to $v = \omega_0 a$, and the Debye frequency corresponding to this velocity is $\omega_{\rm D} = \pi\omega_0 = (\pi/2)\omega_{\max}$ ($\omega_{\max} = 2\omega_0$). We note that we obtained a periodic function with period $2\pi/a$ (Fig. P7.2).

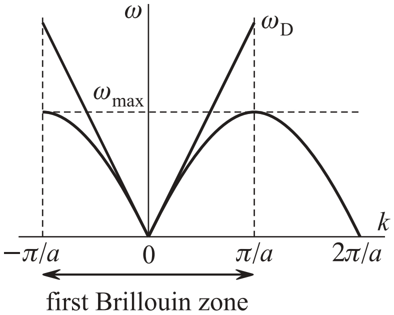

Figure P7.2. Dispersion law and the first Brillouin zone for a one-dimensional chain of atoms.

► **The first Brillouin zone.** We find what values of k correspond to physically distinct states. If we consider the values of the wave vector that differ by $2\pi n/a$: $k' = k + 2\pi n/a$, then we obtain not only equal values of ω but also the identical functions: $\xi_s(k') = \xi_s(k) \cdot \exp(2\pi ins) = \xi_s(k)$. Hence all the physically distinct states are described by wave vectors lying in the interval ($-\pi/a < k \le \pi/a$). This interval is called the *first Brillouin zone*.

On the zone boundary, the traveling wave becomes a standing wave: $A\exp[i(\frac{\pi}{a}as - \omega t)] = A\exp(-i\omega t)(-1)^s$. The neighboring atoms oscillate in antiphase, the wave nodes lie between the atoms. We note that on the zone boundary, the wave group velocity is zero:

$$v_g = \frac{d\omega}{dk} = \omega_0 a\cos\frac{ka}{2} = \frac{a}{2}\sqrt{\omega_{\max}^2 - \omega^2}, \qquad \text{(P7.1.3.8)}$$

where $\omega_{\max} = 2\omega_0$ is the frequency value on the Brillouin zone boundary.

► **The density of states in the Brillouin zone.** The first Brillouin zone contains exactly N distinct values of k (see example 2 above; N is the number of atoms in the chain), which means that all degrees of freedom are exhausted within the framework of the first Brillouin zone (we consider only the longitudinal traveling waves). In the one-dimensional Debye model, the density of states $G(\omega)$ is constant. However, from the dispersion law (P7.1.3.7) we find that

$$G(\omega)\,d\omega = 2G(k)\,dk = 2\frac{L}{2\pi}\frac{dk}{d\omega}\,d\omega = \frac{L}{\pi}\frac{d\omega}{v_{\rm g}}.$$

Using (P7.1.3.7) and (P7.1.3.8), we obtain

$$G(\omega) = \frac{2L}{\pi a}(\omega_{\max}^2 - \omega^2)^{-1/2}.$$

At the Brillouin zone boundary, the density of states $G(\omega)$ becomes infinitely large.

► **Diatomic chain. Optical branches.** If an elementary chain contains several distinct atoms, then in addition to *sound branches*, for which $\omega \sim k$ as $k \to 0$, there arise *optical branches*, on which ω remains finite for all k in the Brillouin zone. For example, for longitudinal oscillations in a chain of alternating atoms with masses m_1 and m_2, there are two branches shown in Fig. P7.3. For small k (near the zone center), the sound branch corresponds to deviations of distinct atoms in the same direction, and the optical branch corresponds to deviations in opposite directions. In the general case, where the cell contains p distinct atoms, there are three sound branches and $3p - 3$ optical branches.

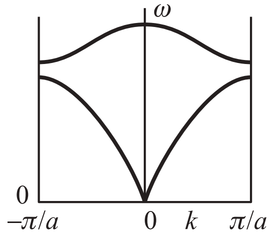

Figure P7.3. Sound and optical branches for a chain of distinct atoms.

► **Phonons.** Just as in the case where oscillations of an electromagnetic field and electromagnetic waves are associated with quantum particles, photons, the lattice elementary vibrations in the form of traveling waves are associated with particles whose momentum and energy are related to the wave vector and frequency by the relations $\mathbf{p} = \hbar\mathbf{k}$ and $\varepsilon = \hbar\omega$. These particles are called *phonons*. Since the phonon energy is a periodic function of their momentum, which is not typical of usual particles, they are called *quasiparticles*, and their momentum is called the *quasi-momentum*. The phonons obey the Bose–Einstein statistics with $\mu = 0$, because the number of phonons in a system is not fixed. The phonons can scatter at impurities and crystal defects and interact with other particles, for example, with electrons and photons. A phonon can also split into two phonons (because of nonlinear effects, which must be taken into account only for sufficiently energetic phonons).

In all these processes, the energy and momentum conservation laws are fulfilled. A nontrivial characteristic of phonons is that, in the final state, the momentum conservation law can contain a phonon that is not in the first Brillouin zone. If we pass to the corresponding phonon from the first zone, then an additional momentum $\hbar\mathbf{G}$, which describes the shift of this Brillouin zone with respect to the first, appears in the momentum conservation law. This momentum corresponds to the motion of the crystal as a whole (which cannot be discovered experimentally, because the mass of the crystal is very large). For example, in the one-dimensional case, $G = 2\pi n/a$, which corresponds to the case of equal displacement of all nodes of the lattice.

The processes of phonons scattering determine their mean free path in terms of which the phonon thermal conduction is expressed:

$$\kappa_{\text{ph}} = \tfrac{1}{3} C_{\text{cr}} \lambda_{\text{ph}} v,$$

where $C_{\text{cr}} \sim T^3$ is volume heat capacity of the lattice, v is the speed of sound. As the temperature increases, the mean free path decreases rapidly.

P7.2. Electrons in a Periodic Field. Energy Bands

Within the framework of the Sommerfeld theory, where electrons in metal are considered as a free fermion gas, it is possible to explain several facts, mainly, the behavior of the electron heat capacity. But many phenomena cannot be explained in this theory. First, this is the problem of origination of metals, semiconductors, and dielectrics and the description of their properties. To clarify these problems, it is necessary to take account of the interaction of electrons with the crystal lattice, which forms a periodic external field for the electrons. The interaction of electrons with the lattice periodic field results in the appearance in their spectrum of *allowed energy bands* separated by *forbidden energy bands*; that is, by the energy intervals that do not contain electron levels. The existence of crystals of different types corresponds to the different character of filling the energy bands by electrons.

P7.2.1. Qualitative Theory of Energy Bands

► **Origination of bands from atomic levels.** Let us track the behavior of electron energy levels in the case of gradual formation of a crystal from isolated atoms. First, we assume that N atoms are at a distance from one another such that the wave functions of electrons do not overlap. In this case, we can assume that each nondegenerate level of a separate atom in the system of N atoms is N-fold degenerate, but if an atomic level is p-fold degenerate, then its degree of degeneration in a system of N atoms is equal to pN. We begin to bring the atoms together to the distance corresponding to the crystal in equilibrium. As a result of interaction between the overlapping electron clouds related to separate atoms, one level yields N very closely located levels contained in a band of width ΔE (Fig. P7.4). The band width is determined by the energy of interaction between the corresponding electrons of neighboring atoms, and it is independent of the number of atoms. For the outer shells, the interaction is strong (the wave functions overlap significantly) and the band width is large, for inner shells, the band width is small. The number of levels in each band is large, so that the spectrum can be assumed to be continuous. Such energy bands are called *allowed bands*, and the energy intervals between the bands, where there are no electron levels, are called *forbidden bands*.

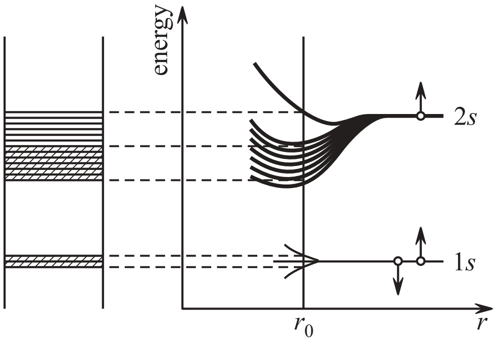

Figure P7.4. Formation of energy bands as the lithium atoms approach each other.

► **Classification of crystals.** The type of the crystal conductivity is defined by the character of filling by electrons (at $T = 0$) of the upper allowed band (above which all the allowed bands are empty). The bands below the upper allowed band are completely filled with electrons. The electrons in these bands at $T = 0$ cannot participate in conductivity, because there are no vacancies where they could move as the external field is switched on. The upper band originating from the atomic outer shell can be filled with electrons either partially or completely.

Metals. If this band is filled partially, then the crystal is a *metal* (or a *semimetal* if the number of electrons in this band is very small). This band is called the *conductivity band*. The Fermi energy of a metal lies in the interior of the conductivity band, the electrons located in this band form a set of free carriers participating in conductivity at any arbitrarily low temperature. The properties of these electrons have much in common with the electron gas in the Sommerfeld model (if the energy is counted from the lower boundary of the band), but there are important distinctions, which we discuss below.

The outer s-shell of atoms of alkali metals contain two electron states (corresponding to different projections of spin), and only one of them is occupied by the electron. Therefore, the conductivity band contains $2N$ free places, and only half of them is filled (see Fig. P7.4 for lithium). For rare-earth metals, both s-states of the outer shell are occupied by electrons, but there is another mechanism of filling the empty band: the energy band originating from the s-shell and containing $2N$ places is overlapped with the band originating from the p-shell containing $6N$ places. The allowed band thus formed contains $8N$ place only the $2N$ of which are filled with electrons (see Fig. P7.5 for beryllium).

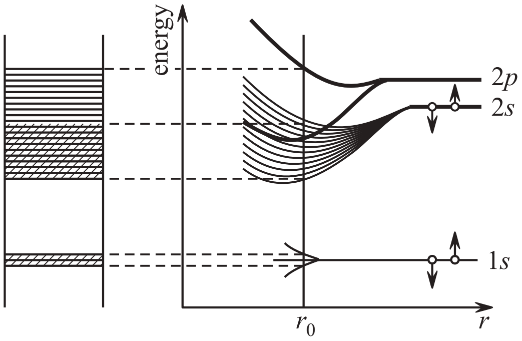

Figure P7.5. Formation of energy bands as beryllium atoms approach each other.

Dielectrics and semiconductors. If the upper band at $T = 0$ is completely filled with electrons, then at $T = 0$, the crystal is a dielectric. For $T > 0$, the conductivity of such a crystal is determined by the width ΔE of the forbidden band lying between the filled band (it is called the *valence band*) and the nearest free band (which in this case is called the *conductivity band*). Because of the thermal motion, the electrons can move from the valence band into the conductivity band. For $k_BT \ll \Delta E$, the electron concentration in the conductivity band is exponentially small ($\sim$ exp(–$\Delta E/k_BT$)). For $k_{\rm B}T \sim \Delta E$, the electron concentration in the conductivity band becomes noticeable and rapidly increases with temperature (the conductivity becomes noticeable already for $k_{\rm B}T \sim 0.1\Delta E$). In this temperature region, the crystal behaves as a *pure semiconductor*. Usually, the semiconductors are crystals whose conductivity becomes noticeable at room temperature.

In a pure semiconductor, simultaneously with the appearance of negative charge carrier—electrons—in the conductivity band, the same number of positive charge carriers—vacant places—arise in the valence band; these positive charge carriers are called *holes*.

The conductivity of pure semiconductors related to the appearance of the equal quantities of electrons and holes is said to be *intrinsic*.

Extrinsic semiconductors. Addition of an impurity results in violation of the symmetry between positive and negative charges. If an impurity whose atom valency is greater than the crystal atom valency is added, there arise excessive electrons that are weakly related to the impurity atoms and become free at the temperatures at which the transitions between the bands are still absent. To these excessive electrons there correspond *donor energy levels* located in the forbidden band near the lower boundary of the conductivity band. The electrons from these levels move into the conductivity band, and since there is still no vacancy in the valence band, the conductivity in this temperature region is realized mainly by electrons. Such extrinsic semiconductors are called *semiconductors of n-type*.

In *semiconductors of p-type*, the valency of impurity is less that the valency of crystal atoms, which results in the appearance of vacant places to which the electrons begin to move at sufficiently low temperatures. To these vacancies there correspond empty energy levels (*acceptor levels*) located near the upper boundary of the valence band. The transition to these levels of electrons from the valence band leads to the formation of holes without appearance of electrons in the conductivity band.

P7.2.2. Quantum Mechanical Description of Energy Regions

Although it is appropriate to use quantum numbers of individual atoms to label energy bands, the state of an electron within an allowed band is described using a completely different quantum number associated with the so-called the *quasi-momentum* of the electron. The quasi-momentum varies quasi-continuously and reflects the translational properties of the Schrödinger equation and the wave function.

► **Bloch waves. Quasi-momentum of electron.** The wave function of a stationary state of an electron in an infinite periodic potential can be written as

$$\psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}\varphi(\mathbf{r}), \qquad \text{(P7.2.2.1)}$$

where $\varphi(\mathbf{r})$ is a periodic function of coordinates (with the same periodic properties as the potential energy of an electron). This statement is called the *Bloch theorem*, and the vector $\hbar\mathbf{k}$ is called the *quasi-momentum of an electron*. The vectors $\mathbf{k}$ and $\mathbf{p}$ for an electron in a crystal are in a sense similar to the wave vector and momentum of the free electron in the Sommerfeld model. Let us clarify the Bloch theorem and the meaning of these notions in the one-dimensional case.

Let the period of one-dimensional lattice be equal to a, and let the potential energy of electron have the same periodicity: $U(x + a) = U(x)$. It is easy to see that if $\psi(x)$ is a solution of the time-independent Schrödinger equation, then $\psi(x + a)$ is also a solution of this equation. Hence $\psi(x+na) = T(na)\psi(x)$, and the discrete function $T(x)$ (determined for $x = na$) has the following properties: $|T(na)| = 1$, $T(na) = [T(a)]^n$. We obtain $T(x) = e^{ikx}$, and the function $\varphi(x) = e^{-ikx}\psi(x)$ is a periodic function of x: $\varphi(x + a) = \varphi(x)$.

For each given value k, from the Schrödinger equation for $\psi(x)$ we obtain a second-order ordinary differential equation for $\varphi(x)$ with the boundary condition $\varphi(a) = \varphi(0)$. This equation has solution only for discrete values $E_n(k)$. If we pass from k to $k' = k + 2\pi/a$, then we obtain the same equation for $\varphi(x)$ and the same values of the energy $E_{k'}$ as for the number k. We conclude that the energy is a periodic function of k with period $G = 2\pi/a$ and the wave functions for the values of k that differ by G are identical. Hence we can reduce all the electron states to the first Brillouin zone. At the well boundaries, there are extrema (minima and maxima) of $E(k)$, which ensure the continuity and smoothness of this function in transitions from one Brillouin zone to another (Fig. P7.6). At the center of the zone (for $k = 0$), $E(k)$ usually also has an extremum.

The above-discussed separation of crystals into metals and dielectrics (at $T = 0$) can be well explained by the dependence $E(k)$ in the first Brillouin zone (Fig. P7.6). Figure P7.6a corresponds to a dielectric, Fig. P7.6b to a metal (or semimetal) in the case of overlapping bands, and Fig. P7.6c to a metal in the case of partially filled band.

► **The number of states in the Brillouin zone.** As k varies in the first Brillouin zone, the energy of the nth level $E_n(k)$ varies within the nth allowed energy band. If we take into account that the crystal is finite (by using, for example, periodic boundary conditions), then in the first Brillouin zone, k takes N values: $0, \pm 2\pi/L, \pm 4\pi/L, \dots, \pi/a$, where $L = Na$. Hence the nth band contains N levels $E_n(k_i)$. With the spin taken into account, we obtain $2N$ electron states.

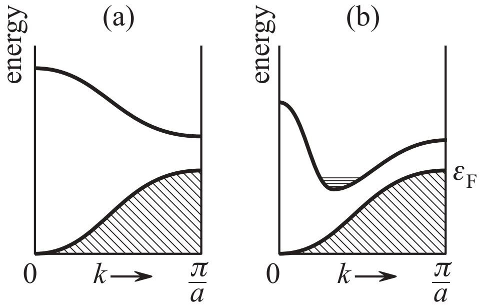

Figure P7.6. Dependence of the electron energy on the quasi-momentum in the first Brillouin zone for the upper energy bands: (a) dielectric, (b) metal or semimetal (overlapping bands), (c) metal (partially filled band).

P7.2.3. Dynamics of Electron in Crystal

► **Equation of motion of an electron in crystal.** The energy spectrum of electrons in the upper bands (the valence band and the conductivity band) bears information about interaction of electrons with the crystal lattice. If this spectrum is used to calculate the group velocity of an electron

$$v_g = \frac{d\omega}{dk} = \frac{1}{\hbar}\frac{d\varepsilon}{dk} = \frac{d\varepsilon}{dp}, \tag{P7.2.3.1}$$

then we can write the equations of motion in the form taking account of the action on electrons of *only external forces*. This equation has the form of usual Newton law

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}, \tag{P7.2.3.2}$$

where **F** is an external force (that does not take account of the interaction with the lattice) and **p** is the quasi-momentum of electron.

For clarity, we present the expression for the rate of variation in the electron energy

$$\frac{d\varepsilon}{dt} = \frac{d\varepsilon}{dp}\frac{dp}{dt} = v_g\frac{dp}{dt},$$

which must be equal to the power of the external force Fv_g.

As an example of the use of equations of motion (P7.2.3.1) and (P7.2.3.2), we consider the behavior of an electron in the empty energy band (where no other electrons are contained) under the action of a constant force. We assume that this electron does not collide with other electrons, phonons, and lattice defects and that, at the initial time, the electron is at rest. In the Sommerfeld model, the free electron would move with constant acceleration away from the point of the beginning of motion. In the band theory, the electron is first at the center of the band (for $p = 0$), then its momentum linearly increases, but the velocity varies differently in this case. It first increases, then decreases approaching the Brillouin zone boundary, and is zero at zone boundary. At this instant, the electron is “transferred” to the left boundary of the zone (if we want to stay in the framework of the first Brillouin zone), and the electron velocity changes the sign. At the time when the electron momentum is again at the center of the zone, its velocity again is zero, and the electron is transferred to the previous place. The electron motion has an oscillatory character. To avoid misunderstanding, we note that the electrons in metal cannot move in such a cycle, because they do not have time between the collisions to go far enough from the Fermi surface and approach the zone boundary.

► **Effective mass.** At the center and the band boundaries, we have $d\varepsilon/dp = 0$, and hence the energy near these points can approximately be written as

$$\varepsilon = \varepsilon_0 + \frac{1}{2}\frac{d^2\varepsilon}{dp^2}\bigg|_{p_0}(p-p_0)^2.$$

where ε_0, p_0 are the energy and quasi-momentum of the center (or boundary) of the band. If in all cases, the momentum and energy are counted from the parabola vertex, then this relation can be written as follows:

$$\varepsilon = \frac{p^2}{2m^*},$$

where

$$\frac{1}{m^*} = \frac{d^2\varepsilon}{dp^2}\bigg|_{p_0}, \tag{P7.2.3.3}$$

and the quantity m^* thus defined is called the *effective mass* of electron. For the electron velocity we obtain the usual relation

$$v_\mathrm{g} = \frac{p}{m^*}.$$

We note that the electron effective mass defined by equation (P7.2.3.3) is positive near the energy band bottom and is negative near the band upper boundary. As we saw, near the band upper boundary, under the action of an external force, the electron velocity does not increase but decreases.

Definition (P7.2.3.3) for m^* acts not only near the band boundaries. Indeed, at any point, we can write

$$\frac{dv_\mathrm{g}}{dt} = \frac{d}{dt}\frac{d\varepsilon}{dp} = \frac{d^2\varepsilon}{dp^2}\frac{dp}{dt},$$

or

$$F = \frac{dp}{dt} = m^*\frac{dv_\mathrm{g}}{dt}. \tag{P7.2.3.4}$$

The effective mass is an important dynamical characteristic of an electron in a crystal. The effective mass of an electron can either be greater than or less than its true mass (sometimes, by several dozens of times). The effective mass stands in the place of the usual mass in the expression for the electrical conduction: $\sigma = e^2 n\tau/m^*$, and in many other physical formulas. The effective mass is especially convenient to describe the properties of semiconductors, where the electron concentration in the conductivity band (and of holes in the valence band) is not large. In anisotropic crystals, the effective mass depends on the direction of motion of the electron.

Although an individual electron near the upper boundary of the band must be assigned a negative effective mass, the absence of such an electron in the completely filled valence band is equivalent to the presence of a single particle with positive mass and positive charge. As was already noted, such charge carriers are called *holes*.

► **Fermi surface for electrons in metals.** In the band theory, the Fermi surface of electrons is not a simple sphere of radius p_F, as it is the case for a free electron. Metals can have quite intricate Fermi surfaces. The determination of electronic Fermi surfaces is one of the most important problems in metal physics. Recall that at room temperatures, the electrons in metals are in a strongly degenerate state ($k_\mathrm{B}T \ll \varepsilon_\mathrm{F}$), and a special role is played by electrons near the Fermi surface. Fermi surfaces are usually depicted within the first

Brillouin zone; however, it is more graphic to display a periodic connected surface covering several Brillouin zones. Figure P7.7 depicts the Fermi surface and the boundaries of the first Brillouin zone for copper; the Fermi surfaces for silver, gold, and other metals of this group look quite similar. In Fig. P7.7, one can clearly see the transitions into neighboring Brillouin zones. (The three-dimensional presentations of the Fermi surfaces for most elements of the periodic table can be found at the Website `http://www.phys.ufl.edu/fermisurface`.)

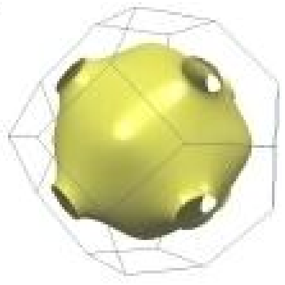

Figure P7.7. Fermi surface and boundaries of the first Brillouin zone for copper.

P7.2.4. Band Theory of Semiconductors

The difference between dielectrics and intrinsic semiconductors is determined by the width of the forbidden band Δ between the filled (at $T=0$) valence band and the empty conductivity band. To semiconductors one usually refers the crystals for which $\Delta < 3$ eV; for example, for germanium, $\Delta = 0.67$ eV, and for silicon, $\Delta = 1.11$ eV. The current carrier concentration is many orders of magnitude less for semiconductors than for metals and it varies quite strongly if impurities, even in negligible concentrations, are added.

► **Fermi energy and conductivity of an intrinsic semiconductor.** The position of the Fermi energy in intrinsic semiconductors is determined by the condition that the electron concentration in the conductivity band and the hole concentration in the valence band coincide. For $k_\mathrm{B}T \ll \Delta$, the Fermi energy is very close to the center of the forbidden band. (A deviation from the center of the band appears because the effective mass of electrons near the conductivity band bottom differs from the effective mass of holes near the top of the valence band). The concentration of charge carriers is calculated from the Fermi distribution, the dependence of the concentration on the temperature is mainly determined by the multiplier $\exp(-\Delta/2k_\mathrm{B}T)$, where $\Delta/2$ is the distance from the Fermi level to the band boundary.

The electron concentration in the conductivity band is calculated by the formula

$$\begin{aligned}\int_{\varepsilon_0}^{\infty} f_\mathrm{F}(\varepsilon)G(\varepsilon)\,d\varepsilon &= \int_{\varepsilon_0}^{\infty} e^{-(\varepsilon-\varepsilon_\mathrm{F})/k_\mathrm{B}T}G(\varepsilon)\,d\varepsilon \\ &= \frac{m^{*3/2}\sqrt{2}}{\pi^2\hbar^3}e^{-(\varepsilon_0-\varepsilon_\mathrm{F})/k_\mathrm{B}T}\int_0^{\infty} e^{-\varepsilon'/k_\mathrm{B}T}\sqrt{\varepsilon'}\,d\varepsilon',\end{aligned}$$

where ε_0 is energy of the conductivity band bottom, $\varepsilon' = \varepsilon - \varepsilon_0$ is the energy counted from the band bottom (it is taken into account that $k_\mathrm{B}T \ll \varepsilon_0 - \varepsilon_\mathrm{F}$).

The intrinsic conductivity of semiconductors depends on the temperature similarly:

$$\sigma = \sigma_0 e^{-\Delta/2k_\mathrm{B}T}.$$

► **Extrinsic semiconductors.** In the case of extrinsic semiconductors, the Fermi energy is displaced to the boundary of the band, where the main charge carriers are located. In the case of semiconductors of n-type with donor conductivity, the Fermi level is located near the conductivity band bottom, in the case of semiconductors of p-type with acceptor conductivity, near the upper boundary of the valence band (Fig. P7.8). At lower temperatures, the number of impurity carriers in the band strongly depends on the temperature, and the conductivity rapidly increases with increasing temperature. Then the impurity levels become "impoverished"; that is, no electrons remain at donor levels (or no free places at acceptor levels), and the conductivity slowly decreases. (Just as in the case of metals, the mean free path decreases because the number of phonons in the system increases.) At higher temperatures, the number of charge carriers begins to increase again, because the mechanism of intrinsic conductivity begins to act. The Fermi energy in this case approaches the center of the forbidden band.

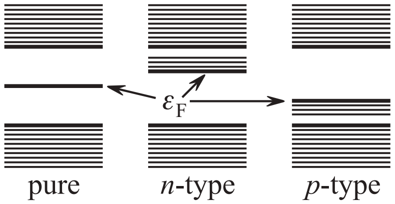

Figure P7.8. Position of the Fermi level in the intrinsic and extrinsic semiconductors.

► **Contact phenomena. p-n transition.** If a semiconductor of n-type comes in contact with a semiconductor of p-type, then a p-n *transition* arises. As a result of diffusion of the majority carriers in the neighboring semiconductor in the transition region, there arises a contact difference of potentials, and the potential of the semiconductor of n-type turns out to be higher. The contact difference of potentials prevents the charge carriers from motion. The potential difference between the semiconductors may attain such a value (Fig. P7.9) that the Fermi energy of the semiconductors becomes equal (the condition that the chemical potentials be equal). In equilibrium, the diffusion current of the majority carrier through the potential barrier is equal to the drift current of minority carriers, to which the contact difference of potentials is no obstacle.

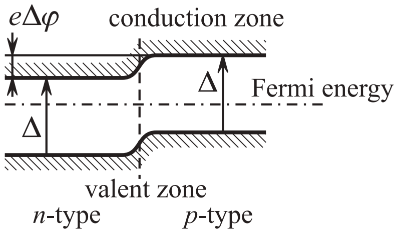

Figure P7.9. Position of energy bands in the region of p-n transition.

If the external field is applied in the forward direction (the positive pole) to the semiconductor of p-type, then the majority carrier current increases in $\exp(eU/k_BT)$ times, because the barrier decreases by eU (U is the applied voltage). If the field is applied in the backward direction, then the forward carrier current decreases (the barrier increases). The applied voltage does not affect the minority carrier current. The volt-ampere characteristic of the p-n transition is determined by the expression

$$I = I_0\left(e^{eU/k_BT} - 1\right),$$

where I_0 is the majority carrier current in the absence of the field (equal to the opposed minority carrier current).

Bibliography for Chapter P7

Alenitsyn, A. G., Butikov, E. I., and Kondratyev, A. S., *Concise Handbook of Mathematics and Physics*, CRC Press, Boca Raton, Florida, 1998.

Anselm, A. I., *Introduction to Physics of Semiconductors* [in Russian], Lan, St. Petersburg, 2008.

Ashcroft, N. W. and Mermin, N. D., *Solid State Physics*, Brooks Cole, Belmont, California, 1976.

Benenson, W., Harris, J. W., Stocker, H., and Lutz, H. (Eds.), *Handbook of Physics*, Springer-Verlag, New York, 2002.

Blakemore, J. S., *Solid State Physics 2nd Edition*, Cambridge University Press, Cambridge, England, 1985.

Chaikin, P. M. and Lubensky, T. C., *Principles of Condensed Matter Physics*, Cambridge University Press, Cambridge, England, 2000.

Ginsburg, I. F., *Introduction to Solid State Physics. Fundamentals of Quantum Mechanics and Statistical Physics with Selected Problems from Solid State Physics* [in Russian], Lan, St. Petersburg, 2007.

Grosso, G. and Parravicini, G. P., *Solid State Physics*, Academic Press, London, 2000.

Hofmann, P., *Solid State Physics: An Introduction*, Wiley-VCH, Berlin, 2008.

Hook, J. R. and Hall, H. E., *Solid State Physics, 2nd Edition*, John Wiley & Sons, Chichester, England, 1991.

Kittel, C., *Introduction to Solid State Physics, 8th Edition*, Wiley, New York, 2004.

Mihaly, L. and Martin, M. C., *Solid State Physics: Problems and Solutions, 2nd Edition*, Wiley-VCH, Berlin, 2009.

Omar, M. A., *Elementary Solid State Physics: Principles and Applications, 4th Edition*, Addison-Wesley, New York, 1994.

Prokhorov, A. M. (Ed.), *Encyclopedia of Physics, Volumes 1–5* [in Russian], Bolshaya Russkaya Encyclopedia, Moscow, 1998.

Prokhorov, A. M. (Ed.), *Encyclopedic Dictionary of Physics* [in Russian], Sovetskaya Encyclopedia, Moscow, 1984 (dic.academic.ru).

Rosenberg, H. M., *The Solid State: An Introduction to the Physics of Crystals for Students of Physics, Materials Science, and Engineering (Oxford Physics Series, 9), 3rd Edition*, Oxford University Press, New York, 1988.

Snoke, D. W., *Solid State Physics: Essential Concepts*, Addison Wesley, New York, 2008.

Woan, G., *The Cambridge Handbook of Physics Formulas*, Cambridge University Press, Cambridge, England, 2003.

Ziman, J. M., *Principles of the Theory of Solids, 2nd Edition*, Cambridge University Press, Cambridge, England, 1979.

Chapter P8

Elements of Nuclear Physics

The idea that practically the entire mass of an atom is concentrated in its positively charged nucleus of infinitesimally small dimensions is due to Rutherford's experiments (see Chapter P6). Since the dimensions of nuclei turned out to be by five orders less than those of atoms, it can be assumed, in the framework of atomic physics, that the nucleus is a point Coulomb center. Actually, the nucleus is a complex structure formed by strongly interacting particles (several ones or hundreds) obeying the laws of quantum mechanics and quantum statistics. The nuclei can undergo radioactive transformations, participate in nuclear reactions, disintegrate, and merge with other nuclei. The characteristic energies in that nuclear world are measured in millions of electron-volts, which explains why nuclei appear as stable objects in atomic processes with energies up to several hundred electron-volts.

P8.1. Basic Properties of Nuclei

P8.1.1. Characteristics of Nuclei

► **Nuclear Composition.** The atomic nucleus consists of protons and neutrons—particles collectively named *nucleons*.

The proton is a subatomic particle with positive charge e and mass $m_p = 1836.15\,m_e$, where m_e is the electron mass. The *neutron* is an electrically neutral particle whose mass is slightly greater than that of the proton: $m_n = 1838.68\,m_e$. In its free state, the neutron is unstable in the sense that within 15.5 minutes, on the average, it turns into a proton after emitting an electron and an antineutrino: $n \rightarrow p + e + \tilde{\nu}$ The spin[*] of both the proton and neutron is equal to $1/2$, which means that they belong to the class of fermions. Both the proton and the neutron possess nonzero magnetic moments: $\mu_p = 2.793\,\mu_\mathrm{N}$ and $\mu_n = -1.91\,\mu_\mathrm{N}$, where $\mu_\mathrm{N} = e\hbar/2m_p$ is the so-called *nuclear magneton* ($\mu_\mathrm{N} = \mu_\mathrm{B}/1836.15$). The nonzero magnetic moment of the neutron suggests that nucleons must have an internal structure involving more elementary charged particles.

The nuclear composition is described by two integers: the *atomic number* Z, equal to the number of protons in the nucleus (and so identical to the *charge number* of the nucleus), and the *mass number* A, equal to the total number of nucleons in the nucleus. The number of neutrons (the *neutron number*) is $N = A - Z$. The charge number of an element coincides with its ordinal number in Mendeleev's table (see Chapter S7) and determines its physical-chemical properties, since these depend on the charge of the nucleus and are almost independent of its mass. The charge and the mass numbers of an element X are specified as indices to the left of its symbol: ${}^A_Z\mathrm{X}$; for example, ${}^3_2\mathrm{He}$, ${}^{13}_6\mathrm{C}$. Since the element symbol uniquely determines its number in the periodic table, which coincides with the charge number Z, the left lower index is sometimes dropped (${}^3\mathrm{He}$, ${}^{13}\mathrm{C}$).

Particular atoms with A nucleons in their nucleus and Z protons, together with their electronic shell containing Z electrons, are called *nuclides*. All nuclides with the same Z

---

[*] Here and henceforth, we often use the word "spin" for short to mean the "spin quantum number."

are called *isotopes* of a given element. The isotopes of hydrogen ^{1}H, ^{2}H, and ^{3}H, whose nuclei are essentially distinct, are respectively called the common hydrogen (sometimes the term *protium* is used), *deuterium* (D), and *tritium* (T), and the nuclei of deuterium and tritium are called *deuteron* (d) and *triton* (t).

There are stable and radioactive isotopes. Radioactive isotopes are characterized by their type of radioactivity (α-decay or β-decay) and their lifetime. (In more detail, radioactive transformations are considered below.) There are many radioactive isotopes occurring in nature, provided that their lifetime is very long (^{235}U, ^{232}Th, and others) or they are constantly produced as a result of decay of such long-lived nuclei (for example, ^{223}Fr). There are no stable isotopes for $Z = 43$ (technetium), $Z = 61$ (promethium) and for all elements with $Z > 83$ (starting from polonium). The other elements possess from 1 to 9 (tin Sn) stable isotopes, and on the average, the number of these is equal to 3. Different stable (or long-lived) isotopes have different occurrence, the observed *atomic weight* is an average over several isotopes. For instance, the atomic weight of natural chlorine is equal to 35.5, and it contains 76% of ^{35}Cl and 24% of ^{37}Cl.

Nuclides with the same mass number A are called *isobars* (for instance, $^{40}_{18}$Ar is $^{40}_{20}$Ca). In the presence of β-decay, isobars can transform into one another. Note that for a stable nuclide, its closest isobars are β-active.

Among stable and long-lived nuclei, the most frequent are even-even ones (even number of protons and even number of neutrons) and the rarest are odd-odd ones (only four stable ones: $^2_1$H, $^6_3$Li, $^{10}_5$B and $^{14}_7$N).

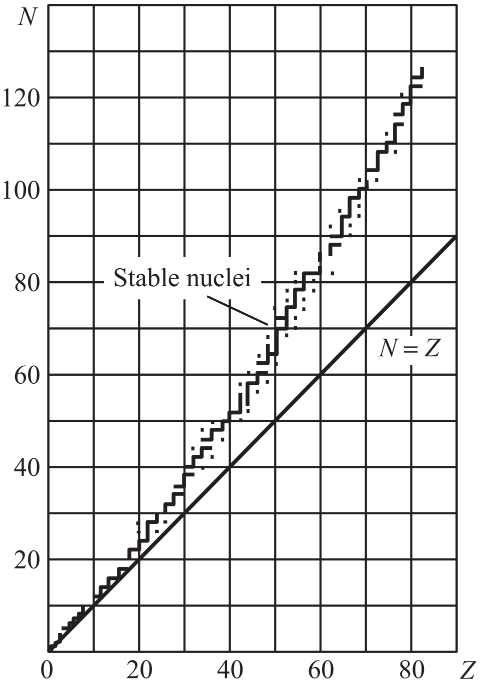

Figure P8.1. Number of neutrons in a nucleus versus its consecutive number.

Figure P8.1 shows the dependence of N on Z for stable nuclides. It can be seen that for the first 15 to 20 elements, the number of protons is approximately the same as that of neutrons ($Z \sim A/2$). With the increase of the consecutive number, there is a clear tendency to the number of neutrons being greater than that of protons (thus, the only stable isotope of bismuth is $^{209}_{83}$Bi).

► **Nuclear spin and magnetic moments.** Nuclear spin I is defined as the total angular momentum of the nucleus, which is the composition of spins and orbital angular momenta of separate nucleons. Nuclei with evenly many nucleons have integer spin, while the spin of nuclei with an odd number of nucleons is half-integer. For nuclei with an even number of both protons and neutrons, the spin of the ground state is equal to zero. This means that angular momenta of separate nucleons compensate one another. It should be mentioned, however, that $I = 1$ for the ground state of deuteron.

As a rule, nuclei with nonzero spin have also a nonzero magnetic moment. Nuclear magnetic moments are usually very small (of the order of nuclear magneton), but their interaction with orbital motion of electrons gives rise to a *hyperfine structure* of electronic levels, which can be observed.

► **Nuclear dimensions and shape.** Nucleon distribution over the nuclear volume has been established in many experiments with scattering of fast electrons and nucleons. (The de Broglie wavelength of scattered particles should be small compared with nuclear dimensions.) In electronic scattering experiments aimed at testing only the electromagnetic interaction one finds spatial distribution of the nuclear charge (i.e., proton distribution), and in neutron scattering experiments one finds spatial distribution of all nucleons. It turns out that these distributions differ but slightly, and in what follows we speak only of the nucleon distribution.

As it happens, the shape of most nuclei is nearly spherical. For nuclei with a fairly large number of nucleons ($A > 12$), one can identify an interior region in which nucleon concentration is constant, and there is a small transition layer over which the concentration drops to zero. If the nuclear radius is defined as the distance from the center where the concentration drops by a factor of two, one arrives at the following dependence on the mass number:

$$R \sim R_0 A^{1/3}\,\text{fm}, \tag{P8.1.1.1}$$

where $R_0 = 1.2\text{–}1.4\,\text{fm}$ and fm (*fermi*) is a non-SI unit of distance ($1\,\text{fm} = 10^{-15}\,\text{m} = 1$ femtometre), common in nuclear physics. It can be seen that nucleon concentration within the nucleus is nearly the same for all nuclei.

Example. Estimate the average nucleon concentration in the nucleus and the average nuclear substance density.

Solution. The average nucleon concentration is equal to $n = A/(\frac{4}{3}\pi R^3) \sim 1.0\text{–}1.5 \times 10^{44}\,\text{m}^{-3}$. This concentration corresponds to the density $\rho = Mn \sim 2 \times 10^{17}\,\text{kg/m}^3$, where M is the nucleon mass.

Note that for many nuclei, one observes a considerable deviation of particle distribution from spherical symmetry (this deviation is especially pronounced in excited states, but also happens in the ground state). Thus, for the simplest compound nucleus, the deuteron, charge distribution in the ground state is nonspherical. (This fact is indicative of a noncentral character of nuclear forces.)

► **Nuclear binding energy.** Nuclear masses measured with a fair degree of accuracy happen to be smaller than the sum of the nucleon masses involved. The nonadditive character of mass in this situation is a direct consequence of the Einstein formula relating mass and energy at rest (see Subsection P1.10.3). Between nucleons, there exist attraction forces by which they are kept within the nucleus in spite of the strong (due to the small distances) Coulomb repulsion, and positive work is required in order to split the nucleus into nucleons. The amount of that work is called *nuclear binding energy*:

$$E_\text{b} = (Zm_p + Nm_n - m_\text{nuc})c^2, \tag{P8.1.1.2}$$

where $N = A - Z$ is the number of neutrons in the nucleus. The huge magnitude of nuclear interactions (compared with, say, atomic or chemical interactions that have a purely electromagnetic nature) made it possible to measure directly the said mass decrease.

Let us make a few remarks with regard to formula (P8.1.1.2).

1. In nuclear physics, nuclear mass and particle mass are expressed either in energy units or in atomic mass units. In the first case, one gives the value of the rest energy mc^2 expressed in electronvolts. Thus, the masses of electron, proton, and neutron in these units are respectively equal to $m_e = 0.511\,\text{MeV}$, $m_p = 938.28\,\text{MeV}$, $m_n = 939.55\,\text{MeV}$. When energy units are used, many formulas become shorter. Thus, formula (P8.1.1.2) becomes

$$E_\text{b} = Zm_p + Nm_n - m_\text{nuc}. \tag{P8.1.1.3}$$

Note that $1\,\text{kg} = 5.61 \times 10^{29}\,\text{MeV}$, $1\,\text{MeV} = 1.78 \times 10^{-30}\,\text{kg}$.

By definition, the atomic mass unit (AMU) is $1/12$ of the mass of carbon nuclide (not nucleus) ^{12}C. Reference tables usually give nuclide passes in AMUs. For example, $m_e = 5.486 \times 10^{-4}\,\text{AMU}$, $m_p = 1.007276\,\text{AMU}$, $m_n = 1.00898\,\text{AMU}$. Note that $1\,\text{AMU} = 931.50\,\text{MeV} = 1.660 \times 10^{-27}\,\text{kg}$.

2. Since reference tables usually list masses of nuclides rather than nuclear masses, it is convenient to rewrite formula (P8.1.1.3) as follows:

$$E_\text{b} = ZM_\text{H} + Nm_n - M_\text{a}, \tag{P8.1.1.4}$$

where M_H is the mass of the hydrogen atom and M_a is the mass of the atom (nuclide) corresponding to the said nucleus. The first term of this formula differs from that of (P8.1.1.3) by Zm_e; and the nuclide mass M_a differs from the nuclear mass m_n by the same amount, if one neglects the energy of electric interaction between electrons and the nucleus.

3. Instead of nuclide mass, tables often list the so-called *mass defect* $\Delta_\text{a} = M_\text{a} - A$ (where M_a is expressed in AMUs). Thus, the mass defect of the nuclide ^{12}C is equal to zero, the mass defect of the hydrogen atom is equal to 0.007825 AMU, the mass defect of the nuclide ^3He is equal to 0.016030 AMU. Binding energy can be expressed in terms of mass defects as follows:

$$E_\text{b} = Z\Delta_\text{H} + N\Delta_n - \Delta_\text{a}. \tag{P8.1.1.5}$$

► **Dependence of binding energy on nuclear composition.** It turns out that in the first rough approximation, binding energy grows in direct proportion to the number of nucleons in the nucleus, i.e., its mass number A. For qualitative and numerical analysis, it is convenient to use *specific binding energy* defined as $\varepsilon = E_\text{b}/A$. Figure P8.2 shows an approximate dependence of ε on the mass number A. Being approximately equal to 8 MeV/nucleon, specific energy first (for $A < 16$) undergoes a sharp increase, then continues to grow smoothly and for $A \sim 60$ (in the region of iron) attains its maximum, 8.8 MeV/nucleon, after which it smoothly decreases to 7.6 MeV/nucleon for the last natural nuclide ^{238}U. Note that due to the growth of ε for small A, fusion of light nuclei is energetically efficient, and due to the decrease of ε for large A, fission of heavy nuclei is energetically efficient. The energy released by fission of heavy nuclei is called *atomic*, and that released by fusion of light nuclei is called *thermonuclear*. Thermonuclear energy per nucleon is several times greater than atomic energy.

Nuclear forces and nuclear models are considered in the next section, but here we make a few brief remarks.

1. The fact that specific binding energy, for most nuclei, is approximately constant and equal to its average value indicates that nuclear forces have a *short-range* character and have the property of *saturation* (each nucleon interacts only with its several closest neighbors).

2. The decrease of ε for small A is explained by an increase of surface effects that depend on the relative fraction of surface particles.

3. The decrease of specific binding energy for large A is explained by an increased relative contribution of Coulomb energy.

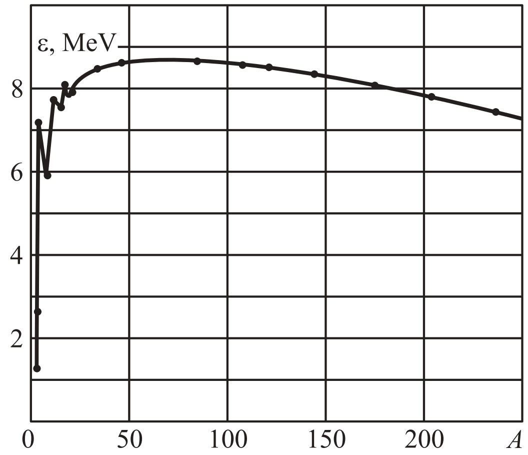

Figure P8.2. Specific binding energy versus mass number.

4. For light elements, the maximum binding energy for nuclides with fixed A (isobars) corresponds to $Z \sim A/2$. (For this reason, such nuclides are usually called β-stable.) This phenomenon is referred to as the *symmetry effect*. The shift to a larger percentage of neutrons observed with the growth of A is explained by an increased role of Coulomb repulsion of protons.

5. Specific binding energy of even-even nuclei is, on the average, greater than that of even-odd nuclei, and specific binding energy of even-odd nuclei is greater than that of odd-odd nuclei. This is explained by *pairing* of identical nucleons in nuclei. The union of two identical nucleons that form a zero-spin pair results in a gain approximately equal to 1 MeV.

6. Most stable (i.e., having the largest ε) among all even-even nuclei are the so-called *magic nuclei* whose number of protons or that of neutrons are equal to one of the *magic numbers*: 2, 8, 20, 28, 50, 82, 126. Especially stable are *double-magic nuclei*; i.e., those with the magic number of both protons and neutrons (e.g., ^{4}He, ^{16}O, ^{40}Ca, ^{208}Pb, etc.). The existence of magic nuclei indicates that nuclei have a shell structure similar to that of atoms.

P8.1.2. Nuclear Forces

► **Properties of nuclear forces.** Between the nucleons in a nucleus there exist nuclear attraction forces which are hundreds of times greater than electromagnetic repulsion forces between protons. These attraction forces ensure nuclear stability. Let us mention some (most important) features of nuclear forces.

1. Nuclear forces have a short-range character with the effective range $r_0 \sim 2.0$–2.5 fm. As the distance between nucleons becomes greater than the effective range, nuclear interaction rapidly decays. If that distance becomes several times smaller than the effective range, attraction is replaced by strong repulsion.

Example. Knowing the deuteron mass, one can calculate binding energy of the simplest nucleus, deuteron. This energy is equal to 2.22 MeV. On the basis of this binding energy, estimate the potential well depth corresponding to nuclear interaction between two nucleons.

Solution. First of all, let us replace the system of two interacting particles of mass M by a single particle with reduced mass $m = M/2$. For simplicity, we assume that potential energy is represented by a well of radius $R = 2.3$ fm and depth U_0. The quantity U_0 can be found from the condition that the ground-state energy in this potential well should be equal to $E_1 = -2.2$ MeV. For the spherically symmetric wave function of the ground state, the Schrödinger equation has the form

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\psi}{dr}\right) + \frac{2m}{\hbar^2}(E-U)\psi = 0,$$

where $U = -U_0$ for $r < R$ and $U = 0$ for $r > R$. Setting $\Psi(r) = r\psi(r)$, we come to the equation

$$\frac{d^2\Psi}{dr^2} + \frac{M}{\hbar^2}(E - U)\Psi = 0,$$

whose solution was obtained in Subsection P6.2.2. Omitting further calculations, we just state the answer: to get the correct ground-state energy, one should have the depth of the potential well about 29 MeV. (For a well of radius 2 fm, the value obtained for the depth is about 35 MeV.)

Note that this attraction value corresponds only to the state of a two-nucleon system in which their total spin is equal to 1. In the state with zero spin, attraction is nearly twice weaker and there is no bound state. This is why there is no stable formation consisting either of two neutrons or two protons (according to Pauli's principle, two identical particles cannot have parallel spins).

There is no final answer with regard to the dependence of nuclear forces on the distance between nucleons. With equal success, different models of that dependence are used for processing experimental data. The best theoretical justification has been given to the Yukawa formula

$$U(r) = A\frac{e^{-r/r_0}}{r}. \qquad \text{(P8.1.2.1)}$$

2. The short-range nature of nuclear forces is closely connected with their *saturation property*. Every nucleon inside the nuclear substance is interacting with a certain number of its closest neighbors. (The nucleons near the surface of a nucleus have a smaller number of neighbors.) For this reason, nuclear binding energy grows in proportion to the number of nucleons, while the average nuclear density remains constant. If nuclear forces were long-range (as, for instance, the Coulomb force), then every nucleon would interact with all the others and the binding energy would grow faster. For example, electric energy of a ball with charge Ze and radius R is proportional to Z^2/R (see Subsection P3.6.1).

3. Nuclear forces have the property of *charge independence*: they do not depend on the type of interacting nucleons (p-p, p-n, or n-n).

4. Nuclear forces are non-central. This is confirmed, in particular, by the non-spherical charge distribution in the deuteron ground state. If the interaction was exactly central, the orbital angular momentum would be conserved and would be strictly equal to zero in the ground state, which corresponds to a spherically symmetric wave function.

5. Nuclear forces depend on spin orientations. (See above with regard to nucleon interaction in the deuteron.)

► **Nature of nuclear forces.** In modern physics, any fundamental interaction is regarded as a result of particle exchange (particles are agents of interaction). Nuclear interaction between nucleons is implemented through the exchange of *π-mesons (pions)*, which are particles with zero spin and mass m_π about 200 times greater than that of the electron.

Around a single nucleon, the so-called *virtual pions* constantly are born and die. The term "virtual" is used here, because these particles exist for such a short time that due to the time-energy uncertainty relation ($\Delta t\,\Delta E \sim \hbar$), it is possible for the energy of the system to change by $\sim m_\pi c^2$. For virtual particles, the common relation between energy and momentum does not hold. These particles have different impulses (of different directions) with the average value $m_\pi c$ and in their lifetime travel the average distance $c\Delta t \sim \hbar/(m_\pi c)$ equal to the Compton wavelength of a π-meson. Sometimes in this situation one speaks of a single nucleon surrounded by a coat of π-mesons whose radius, r_0, is equal to their Compton length. As two nucleons approach one another by the distance r_0, their meson coats begin to overlap and some virtual pions born of one nucleon are absorbed by the other. In 1935, Yukawa showed that upon averaging over the momenta of virtual pions, there arises attraction with potential energy described by (P8.1.2.1) with $r_0 = \hbar/(m_\pi c)$. At that time, π-mesons were still unknown. Yukawa predicted their existence and estimated their mass in terms of the nuclear interaction radius. (It was not until 1947 that π-mesons were discovered in cosmic rays.)

Let it be mentioned here that the carriers of electromagnetic interaction are virtual photons. Photons have zero mass, and therefore, electromagnetic interaction has long-range character ($r_0 = \infty$, and the Yukawa formula (P8.1.2.1) results in the Coulomb interaction). Weak interaction is implemented through the exchange of very massive W-bosons, and therefore, its radius is much smaller than that of nuclear interaction. Note also that the contemporary theory assumes that the basic carriers of strong interaction are *gluons* implementing the interaction between *quarks*, of which all strongly interacting particles are built (see Subsection P8.4.1).

P8.1.3. Nuclear Models

► **Nuclear models.** It should be observed from the start that there is still no universal model of the nucleus that could provide a unified theoretical approach to the calculation of various nuclear characteristics. Each of the existing models is capable of successfully explaining some properties of nuclei, while the other properties remain unaccounted for or their explanation is too complex or questionable. Among these models, a special place is occupied by the *microscopic nuclear model* proposed by A. B. Migdal, who called it the *theory of finite Fermi systems*. This model is based on the theory of quantum Fermi fluid which applies the methods of the quantum field theory to a statistical ensemble of interacting fermions. The major achievements of the microscopic model are that it provides a comparison and a unified justification of outwardly distinct theoretical models (sometimes, it is called a "model of nuclear models").

The other models can be provisionally divided into two groups: collective models and single-particle models. The first group covers, in particular, the theory of infinite nuclear matter and the liquid-drop nuclear model. The second group contains the theories dealing with the motion of separate nucleons either in a plane potential well of finite depth (nuclear Fermi gas) or in some effective potential (shell models). Some sophisticated shell models take into account pairing of identical nucleons.

There are some new models that combine certain features of the said two groups. The interior part of the nucleus corresponding to completely filled up shells is regarded as a drop of nuclear fluid in whose field external nucleons are moving.

► **Liquid-drop (hydrodynamic) nuclear model** treats the nucleus as a charged drop of nuclear liquid with surface tension. This model describes basic properties of the dependence of binding energy on A and Z and characterizes surface vibrations of nuclei and some properties of fission of heavy nuclei.

► **The nuclear Fermi gas model** is used for the explanation of the symmetry effect (Z is approximately equal to N in light nuclei) and its violation in heavy nuclei, and this model allows one to calculate the depth of the effective potential well.

In this model, both protons and neutrons are regarded as two subsystems of identical fermions moving freely in a potential well. On each level, two identical particles with different spins can exist. The more nucleons of a given type are present in the nucleus, the more their Fermi energy. If there is a considerable divergence between the Fermi energies of neutrons and protons, then one nucleon will transform into another with total energy decrease (either $n \to p + e^- + \tilde{\nu}$ or $p \to n + e^+ + \nu$). Therefore, in a stable nucleus, the Fermi energies should be approximately the same. With the Coulomb interaction neglected, the total (common) energy minimum corresponds to approximately the same number of neutrons and protons. This situation corresponds to light nuclei. In heavy nuclei, the Coulomb interaction between protons is increasingly significant and leveling of the Fermi energies occurs when the number of neutrons exceeds that of protons.

Example. Using the mean concentration of nucleons in the nucleus calculated in the example of Subsection P8.1.1, estimate the depth of the potential well occupied by these nucleons.

Solution. To obtain such an estimate, we neglect the Coulomb interaction. The mean kinetic energy of nucleons of one type is equal to (see Subsection P7.1.2)

$$K_{\text{mean}} = \frac{3}{5}E_F = \frac{3}{5}\frac{\hbar^2(3\pi^2 n_1)^{2/3}}{2M} \sim 20\,\text{MeV/nucleon},$$

where $n_1 \sim 0.6 \times 10^{44}\,\text{m}^{-3}$ is the concentration of nucleons of one type equal to one half of the nucleon concentration. The mean specific binding energy is about 8 MeV/nucleon. This energy can be calculated as the difference between the well depth and the mean kinetic energy. We find that the well depth should be about 30 MeV. As will be shown below, the estimate just obtained is well below the actual value.

► **Empirical formula for binding energy.** On the basis of the liquid-drop and the gas models, one can write a fairly simple empirical formula (Weizsäcker's formula) for nuclear binding energy (1935):

$$E_{\text{b}} = \varepsilon A = \alpha_1 A - \alpha_2 A^{2/3} - \alpha_3\frac{Z^2}{A^{1/3}} - \alpha_4\frac{(A-2Z)^2}{A} - \alpha_5\frac{\delta(Z,N)}{A^{3/4}}, \qquad \text{(P8.1.3.1)}$$

where $\delta(Z,N)$ is equal to –1 for even-even nuclei, 0 for even-odd nuclei, and 1 for odd-odd nuclei. The coefficients α_i are chosen from experimental data and at present the following values are adopted: $\alpha_1 = 15.75\,\text{MeV}$, $\alpha_2 = 17.8\,\text{MeV}$, $\alpha_3 = 0.71\,\text{MeV}$, $\alpha_4 = 23.7\,\text{MeV}$, and $\alpha_5 = 34\,\text{MeV}$.

The first three terms correspond to the concepts of the liquid-drop model. The first term reflects approximate proportionality of binding energy to the number of nucleons; the coefficient α_1 can be understood as binding energy associated with a single nucleon in an imaginary nuclear substance (matter) with electric interaction "turned off." The second term describes "surface tension" of a nuclear drop (the nucleons on the surface of the nucleus are in special conditions), and the third term describes its electrostatic energy.

The fourth term is due to the symmetry effect explained in the framework of the nuclear Fermi gas. This term attains its minimum for $N = Z$, since the divergence from total symmetry observed with an increase of A is connected with the Coulomb energy, and this has been accounted for in the third term. The fifth term reflects stepwise change of binding energy at the transition from a given nucleus to nuclei with adjacent values of A or Z.

It is of interest to note a considerable divergence of the coefficient α_1 from the mean value of specific binding energy $\varepsilon \sim 8$ MeV. The point is that ε already contains the Coulomb and the surface energies. If ε is replaced by α_1 in example (when estimating the well depth), then the binding energy acquires the value ~ 40 MeV.

On the basis of Weizsäcker's formula one can perform fairly precise ($\sim 10^{-4}$) calculations of binding energy for many nuclides, especially heavy ones, α-decay energy, fission and fusion energies, and many other characteristics. However, to explain the existence of magic nuclei, to calculate spins and magnetic moments, one cannot do without the shell model.

► **Shell model of the nucleus** (Goeppert-Mayer and Jensen, 1949). This model assumes that each nucleon is moving in a self-consistent field produced by other nucleons. The potential energy of this field is chosen so as to give the best agreement with experimental data for a wide range of nuclei.

In contrast with electrons in an atom, spin-orbit interaction plays a much greater role for nucleons and and can reach 10–15%. Furthermore, the Russel–Saunders coupling common for electrons (i.e., LS-coupling, in which the spins and orbital angular momenta of all particles are first summed up separately and then the total spin is added to the total orbital angular momentum) is suitable for the lightest nuclei only, whereas for most other nuclei the jj-coupling is more appropriate, in which the spin of each individual nucleon is added to its orbital angular momentum. The nucleon energy levels are characterized by the quantum numbers l (denoted by the letters s, p, d, f, g, $\dots$), j (equal to $l-\frac{1}{2}$ or $l+\frac{1}{2}$

and specified as the right subscript), and m_j, and also the number n that enumerates the levels with a given l and can take the values 1, 2, 3, ... To each level, there correspond $2j+1$ states. The nucleon energy levels are grouped into shells; the distances between these groups are significantly greater than the interlevel distances within a shell. The magic numbers correspond to shells being completely filled up:

I: $1s_{1/2}$ (2 nucleons);
II: $2p_{3/2}$, $2p_{1/2}$ (6 nucleons, altogether 8);
III: $1d_{5/2}$, $2s_{1/2}$, $1d_{3/2}$ (12 nucleons, altogether 20);
IV: $1f_{7/2}$ (8 nucleons, altogether 28);
V: $2p_{3/2}$, $1f_{5/2}$, $2p_{1/2}$, $1g_{9/2}$ (22 nucleons, altogether 50);
VI: $1g_{7/2}$, $2d_{5/2}$, $2d_{3/2}$, $3s_{1/2}$, $1h_{11/2}$ (32 nucleons, altogether 82);
VII: $1h_{9/2}$, $2f_{7/2}$, $2f_{5/2}$, $3p_{3/2}$, $3p_{1/2}$, $1i_{13/2}$ (44 nucleons, altogether 126).

Nucleon states are filled up in the order indicated in the above scheme. Using this, one can obtain correct values for the spins of almost all nuclei.

P8.2. Nuclear Transformations

P8.2.1. Radioactivity

Radioactivity is a spontaneous emission of one or several particles from a nucleus accompanied by its transformation into another nucleus (or its transition to another state). The original radioactive nucleus (nuclide) is called *parent nucleus*, and product nuclei are called *daughter nuclei*.

Basic types of radioactivity are the following: α-decay, β-decay, and γ-decay. In spite of different physical processes associated with these types of decay, their time evolution is described by one and the same law.

► **Radioactive decay law.** Since radioactive decay has a random nature, it can be assumed that on a small time-interval dt the fraction of split nuclei is proportional to dt. Denoting by $N(t)$ the number of intact nuclei by the time instant t, we come to the equation

$$\frac{dN}{N} = -\lambda\,dt. \qquad \text{(P8.2.1.1)}$$

The coefficient λ is called the *decay constant*. The solution of equation (P8.2.1.1) has the form

$$N = N_0 e^{-\lambda t}, \qquad \text{(P8.2.1.2)}$$

where N_0 is the number of nuclei for $t = 0$. This relation is called the *radioactive decay law*.

The number of nuclei decaying per unit time is called *radioactive sample activity*:

$$A = \left|\frac{dN}{dt}\right| = \lambda N = A_0 e^{-\lambda t}, \qquad \text{(P8.2.1.3)}$$

where A_0 is the initial activity of the a sample. Activity per unit mass is called *specific activity*. Activity is measured in *becquerels* (1 Bq corresponds to one decay per second) or *curies* (1 Ci = 3.7×10^{10} Bq).

Apart from the constant λ, decay is characterized by the *half-life period* $T_{1/2}$ (often called *half-life*) equal to the time in which half of the nuclei undergo decay ($N = N_0/2$). From (P8.2.1.2), we find that

$$T_{1/2} = \frac{\ln 2}{\lambda} = \frac{0.693}{\lambda}. \qquad \text{(P8.2.1.4)}$$

It is natural to introduce the *mean lifetime* of a nucleus, τ. The probability of the event that the nucleus decays on the time interval $(t, t+dt)$ is equal to $|dN|/N_0 = \lambda N\,dt/N_0 = \lambda e^{-\lambda t}dt$, and therefore,

$$\tau = \int_0^\infty t\lambda e^{-\lambda t}dt = \frac{1}{\lambda} = \frac{T_{1/2}}{\ln 2} = 1.44T_{1/2}. \qquad \text{(P8.2.1.5)}$$

Example. Calculating sample activity. The half-life of the isotope ${}^{90}_{38}\mathrm{Sr}$ is about 28 years. What is the activity of its $m = 1$ g sample?

Solution. The activity is equal to the product of the decay constant $\lambda = \ln 2/T_{1/2}$ and the number of nuclei $N = N_\mathrm{A}(m/\mu)$, where $\mu = 90$ g/mol. We obtain $A = 5.2\times 10^{12}\ \mathrm{c}^{-1} = 140$ curie.

► **Alpha decay.** The alpha decay (or α-decay) of a nucleus X with atomic number Z and mass number A results in the nucleus ${}^4\mathrm{He}$ (known as the *alpha particle* or *α-particle*) and a daughter nucleus Y with atomic number $Z-2$ and mass number $A-4$:

$${}^A_Z\mathrm{X} \to {}^{A-4}_{Z-2}\mathrm{Y} + {}^4_2\mathrm{He}. \qquad \text{(P8.2.1.6)}$$

Alpha decay is observed only for heavy nuclei (mostly with $A > 209$).

The energy released during the α-decay is equal to

$$Q = m_\mathrm{X} - m_\alpha - m_\mathrm{Y}, \qquad \text{(P8.2.1.7)}$$

where the nuclear masses are expressed in energy units. Note that instead of nuclear masses one can substitute masses of nuclides (electron masses are mutually canceled) or their mass defects.

The fact that one observes decay with the emission of the nucleus ${}^4\mathrm{He}$, and not other light nuclei or single nucleons, is explained by the abnormally large binding energy of the nucleus ${}^4\mathrm{He}$. For example, the α-decay of the nucleus ${}^{232}\mathrm{U}$ is accompanied by the emission of 5.4 MeV energy, and for the emission of a proton by this nucleus it should be supplied with 6.1 MeV energy.

Due to the momentum conservation law, the nucleus experiences recoil and carries away part of the energy. For this reason, the registered energy of α-particles is less than the reaction energy Q. Assuming that the parent nucleus were at rest, let us write the momentum conservation law (for) $p_\alpha = p_\mathrm{Y}$ in terms of kinetic energies of the daughter nucleus and the α-particle:

$$m_\alpha K_\alpha = m_\mathrm{Y} K_\mathrm{Y}$$

Expressing K_Y from the above formula and inserting it into the energy conservation law

$$K_\alpha + K_\mathrm{Y} = Q,$$

we obtain

$$K_\alpha = \frac{m_\mathrm{Y}}{m_\mathrm{Y} + m_\alpha}Q = \frac{A-4}{A}Q,$$

where A is the mass number of the parent nucleus. As a rule, we have $A > 210$, and therefore, the kinetic energy of the α-particle is greater than 98% of the decay energy.

Since the average lifetime of most α-active nuclei is great, they are subject to decay from the ground state. Moreover, due to a very “sharp” (strong) dependence of the decay constant on the energy of the outgoing particles, one observes the most intensive transition to the ground state of the parent nucleus. Therefore, in most cases, all α-particles of a given decay have the same energy. In some cases, one observes α-particles that correspond to low (rotational) levels, and this leads to a slight blurring of the α-spectrum. Moreover, for some extremely short-lived nuclei, decay is possible from excited states of the parent nucleus and one observes α-particles with energies considerably larger than the principal value (for example, polonium isotopes ${}^{212}_{84}\mathrm{Po}$ and ${}^{214}_{84}\mathrm{Po}$ emit not only “common” α-particles but those with energies up to 10.5 MeV).

► **Theory of alpha decay.** For a long time, some properties of α-decay looked quite mysterious. In the first place, it is an unusually strong dependence of the half-life on the energy of α-particles. In the principal region of α-activity ($A \geq 210$), α-particles with energies significantly smaller than 4 MeV are never observed (their half-lives are too great), for ${}^{204}_{82}\mathrm{Pb}$, $K_\alpha \approx 4\,\mathrm{MeV}$ and $T_{1/2} = 1.4 \times 10^{17}$ years, and for ${}^{212}_{84}\mathrm{Po}$, $K_\alpha \approx 9\,\mathrm{MeV}$ and $T_{1/2} = 3 \times 10^{-7}$ s. Moreover, one could not understand how it was possible for an α-particle to leave the nucleus at all. Outside the nucleus, its Coulomb potential energy

$$U_{\mathrm{col}} = \frac{2(Z-2)e^2}{4\pi\varepsilon_0 r}$$

attains its maximum on the boundary of the nucleus for $r = R = r_0 A^{1/3}$ (see (P8.1.1.1)), and inside the nucleus it enters a negative potential well of nuclear forces. Hence, the α-particle is locked by a potential barrier (Fig. P8.3) whose height $U_0 \sim 25$–$30\,\mathrm{MeV}$ is considerably greater than the energy of α-particles 4–9 MeV. The alpha particle can penetrate this barrier only through quantum tunneling (see Subsection P6.2.2). The theory of α-decay as a tunneling transition was developed in 1928 by Gamow, Condon, and Gurney and explained all its major properties.

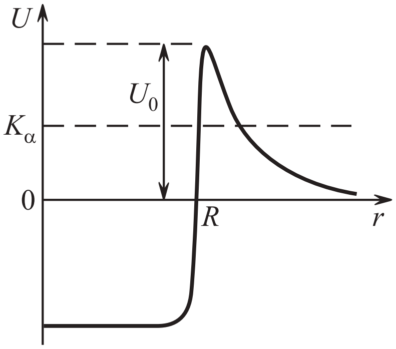

Figure P8.3. Potential energy of an α-particle in the nucleus has the shape of a potential barrier.

In order to calculate the decay constant λ, the probability p of the production of an α-particle inside the nucleus (it is assumed that $p \sim 1$) should be multiplied by the number of its collisions with the barrier $\nu \sim v/2R$ (v is speed of the α-particle) and the transmission coefficient (of passing through the barrier)

$$D \sim \exp\Bigl\{-\frac{2}{\hbar}\int_R^{r_1} \sqrt{2m_\alpha\bigl(U(r) - K_\alpha\bigr)}\,dr\Bigr\}.$$

This transmission coefficient is responsible for the sharp (exponential) dependence of the decay constant (and the half-life) on the energy of the outgoing particle K_α.

► **Beta decay.** Beta decay (or β-decay) is the term applied to radioactive nuclear transformations in which a proton turns into a neutron (or vice versa) and involving electrons (or positrons) and neutrinos (or antineutrinos). Basic types of β-decay are the following:

a) Electron β^--decay (beta-minus):

$${}^A_Z\mathrm{X} \to {}^{A}_{Z+1}\mathrm{Y} + e^- + \tilde{\nu} \qquad \text{(P8.2.1.8)}$$

(for example, decay of tritium ${}^3_1\mathrm{H} \to {}^3_2\mathrm{He} + e^- + \tilde{\nu}$).

b) Positron β^+-decay (beta-plus):

$$ {}^A_Z\mathrm{X} \to {}^{A}_{Z-1}\mathrm{Y} + e^+ + \nu \qquad \text{(P8.2.1.9)} $$

(for example, decay of the artificial radioactive isotope of nitrogen ${}^{13}_{7}\mathrm{N} \to {}^{13}_{6}\mathrm{C} + e^+ + \nu$).

c) Electron capture:

$$ e^- + {}^A_Z\mathrm{X} \to {}^{A}_{Z-1}\mathrm{Y} + \nu \qquad \text{(P8.2.1.10)} $$

(for example, the reaction $e^- + {}^7_4\mathrm{B} \to {}^7_3\mathrm{Li} + \nu$). In electron capture, the nucleus absorbs an electron from an atomic shell, usually from the K-shell, and therefore, this process is called K-capture (one also observes L-capture). Electron capture is accompanied by the characteristic X-ray radiation.

An important distinction of β-decays from other types of radioactivity is that the former are intranucleon (pertaining a nucleon) processes rather than intranuclear (pertaining to the entire nucleus) ones. This means that an elementary act of such a process consists in the transformation of one nucleon into another. The schemes of the transformations listed above are the following:

a) $n \to p + e^- + \tilde{\nu}$
b) $p \to n + e^+ + \nu$;
c) $e^- + p \to n + \nu$.

Such transformations (and therefore, nuclear β-decay) are possible only because of weak interaction. In free space, only the first process (neutron decay) can happen spontaneously. However, within the nucleus the new nucleon turns out surrounded by other nucleons, in the process of its birth the entire nucleus changes its structure, and the energy condition required for the decay is determined by masses of the nuclei (the original and the produced) rather than the nucleon masses.

Spontaneous decay is energetically allowed, if the sum of the masses of the original particles is greater than the sum of the masses of the produced particles. For β-decay, this condition has the form:

$m_\mathrm{X} > m_\mathrm{Y} + m_e$ (β^- and β^+-decay);
$m_\mathrm{X} + m_e > m_\mathrm{Y}$ (electron capture).

It is convenient to rewrite these formulas in terms of nuclide masses. Thus, for β^--decay, one should take $m_\mathrm{X} = M_\mathrm{X} - Zm_e$, $m_\mathrm{Y} = M_\mathrm{Y} - (Z+1)m_e$ and the decay condition becomes

$M_\mathrm{X} > M_\mathrm{Y}$ (β^--decay).

For the other types of decay, we have:

$M_\mathrm{X} > M_\mathrm{Y} + 2m_e$ (β^+-decay);
$M_\mathrm{X} > M_\mathrm{Y}$ (electron capture).

Note that the last two types of decay result in identical daughter nuclei, i.e., are competitive; however, electronic capture is more energetically advantageous.

Weizsäcker's formula (P8.1.3.1) can be utilized to find the parameters (Z, A) for the nuclei most stable with respect to the above types of β-decay. Since the mass number A is unchanged in these types of decay, one should compare the masses (binding energies) of all isobars. Consider, for instance, an odd A (even-odd nuclei, $\Delta(A, Z) = 0$). Minimizing the nuclear energy $E = Zm_p + (A - Z)m_n - E_\mathrm{b}$ with respect to Z, we find Z_β corresponding to the maximum β-stability (Fig. P8.4a). The isobars with $Z < Z_\beta$ should be examined with regard to β^--decay, and those with $Z > Z_\beta$ should be examined with regard to β^+-decay and electron capture. For even A, we obtain two branches of $E(Z)$ (Fig. P8.4b): the upper one

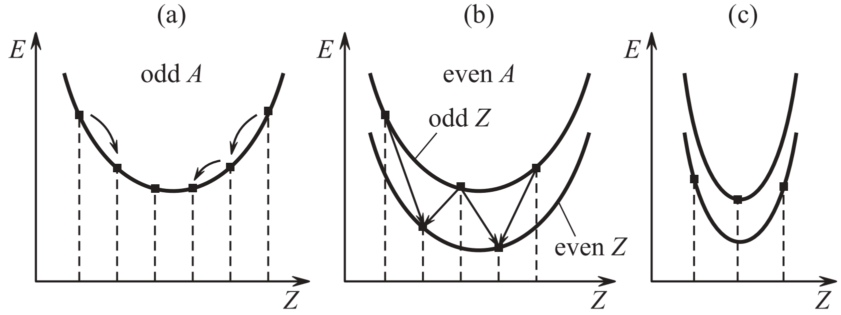

Figure P8.4. Isobaric nuclei energy versus Z and energetically allowed types of β-decay for nuclei with even A (a) and nuclei with odd A (b); the right figure (c) corresponds to a β-stable nucleus with an even A.

for odd-odd nuclei and the lower one for even-even nuclei. The nuclei on the upper branch are usually unstable with respect to transitions to the lower branch, but there are 4 β-stable nuclei corresponding to the situation represented in Fig. P8.4c.

The number of β-active nuclides is much larger than that of α-active nuclides (each element has a β-active isotope). The half-life depends on energy (the greater the energy gain, the smaller $T_{1/2}$), but this dependence is not as strong as in the case of α-decay. Moreover, the greater the divergence between the structures of the final and the original nuclei (for instance, the difference of their spins), the slower the decay. Sometimes, the transition to the ground state of the daughter nucleus has small probability (is a rare event), and as a result of β-decay a nucleus in an excited state is produced.

The neutrino (more precisely, electron-type neutrino, which is one of its three modifications) is a very light (possibly, massless) neutral particle with spin 1/2, which participates only in weak interactions. For this reason, its interaction with matter is extremely weak and its detection is very difficult. Thus, for a neutrino with energy 100 MeV the mean free path in water is of the order 10^{19} m (about 100 light years). A direct experimental proof of the existence of neutrino was not obtained until 1956, when it became possible to utilize a powerful flux of neutrino resulting from uranium fission in nuclear reactor. (Reactions of antineutrino capture $\tilde{\nu} + p \rightarrow n + e^{+}$ were registered.) The existence of neutrino was hypothesized by W. Pauli in 1930, in a rather dramatic situation due to its “invisibility.”

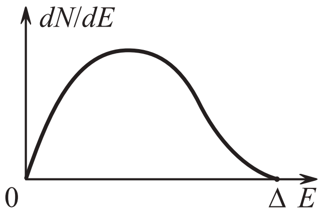

Figure P8.5. Distribution of β-electrons with respect to energies.

It was considered a proven fact that β^{-}-decay results only in two particles: the daughter nucleus and the electron. In that case, as in the case of any two-particle decay (say, α-decay) of a fixed nucleus, the decay products should have certain energies. With the kinetic energy of the daughter nucleus neglected, the electron energy should be equal to $\Delta = m_{\text{X}} - m_{\text{Y}} - m_e = M_{\text{X}} - M_{\text{Y}}$. However, in 1914, J. Chadwick established the continuity of β-spectra. It turned out that the distribution of electrons with respect to energies is like that in Fig. P8.5, and the electron energy averaged with respect to that distribution is as small as $\Delta/3$. The situation looked so paradoxical that in 1924 N. Bohr guessed that the energy conservation law could be violated in micro-processes! As noticed in 1929, the angular momentum conservation law is violated for some types of β-decay. Thus, for $^{210}_{83}\text{Bi} \rightarrow ^{210}_{84}\text{Po} + e^{-}$, the spins of the nuclei are equal to 1 and 0 and the electron carries away a half-integer spin. This deadlock was broken by Pauli's hypothesis about the existence of neutrino. On the basis of this hypothesis, in 1934, E. Fermi developed the theory of weak interactions and calculated the activities of β-decays and the spectra of outgoing electrons. In 1936, N. Bohr accepted the Fermi theory and rejected the idea of non-conservation of energy.

► **Radioactive families.** Most natural radioactive elements belong to four radioactive families. All elements of these families are obtained by consecutive decay from a single parent nuclide and end with some stable element. The decay family of thorium has mass numbers $4n$, starts with ${}^{232}_{90}\mathrm{Th}$ and ends with ${}^{208}_{82}\mathrm{Pb}$. The decay family of neptunium ($A = 4n + 1$) starts with ${}^{237}_{93}\mathrm{Ne}$ and ends with ${}^{209}_{83}\mathrm{Bi}$. The decay family of uranium ($A = 4n + 2$) starts with ${}^{238}_{92}\mathrm{U}$ and ends with ${}^{206}_{82}\mathrm{Pb}$. Finally, the actinium family ($A = 4n + 3$) starts with ${}^{235}_{92}\mathrm{U}$ and ends with ${}^{207}_{82}\mathrm{Pb}$. (The first three terms of the last family were discovered later.) As an example, let us give the sequence of elements of the thorium family:

$$
\begin{aligned}
&{}^{232}_{90}\mathrm{Th} \xrightarrow{\alpha} {}^{228}_{88}\mathrm{Ra} \xrightarrow{\beta} {}^{228}_{89}\mathrm{Ac} \xrightarrow{\beta} {}^{228}_{90}\mathrm{Th} \xrightarrow{\alpha} {}^{224}_{88}\mathrm{Ra} \xrightarrow{\alpha} {}^{220}_{86}\mathrm{Rn} \xrightarrow{\alpha} {}^{216}_{84}\mathrm{Po}\\
&\qquad \xrightarrow{\alpha} {}^{212}_{82}\mathrm{Pb} \xrightarrow{\beta} {}^{212}_{83}\mathrm{Bi} \xrightarrow{\alpha} {}^{208}_{81}\mathrm{Tl} \xrightarrow{\beta} {}^{208}_{82}\mathrm{Pb} \qquad \big({}^{212}_{83}\mathrm{Bi} \xrightarrow{\beta} {}^{212}_{84}\mathrm{Po} \xrightarrow{\alpha} {}^{208}_{82}\mathrm{Pb}\big).
\end{aligned}
$$

The element ${}^{212}_{83}\mathrm{Bi}$ can decay by way of α- and β-radiation, and therefore, on this element the decay sequence branches.

► **Gamma decay.** Gamma radiation is observed in transitions of an excited nucleus to lower energy levels, with unchanged A and Z and emitted γ-quanta. After emitting one or several photons, the nucleus comes to the ground state. Since nuclear energy values are discrete, the spectrum of γ-radiation is also discrete. Excited nuclei are produced in β-decay, if the decay to the ground state is forbidden. For example, β^--decay of ${}^{24}_{11}\mathrm{Na}$ with spin $I = 4$ leads neither to the ground state ${}^{24}_{12}\mathrm{Mg}$ with spin $I = 0$, nor to the first excited level ($I = 2$), but leads to the second level with $I = 4$, upon which the nucleus consecutively emits two γ-quanta with energies 2.7 MeV and 1.4 MeV and comes to the ground state.

A free nucleon cannot emit γ-quanta, since this would contradict the energy and the momentum conservation laws. However, inside the nucleus, the momentum is redistributed between all nucleons and radiation becomes possible. Thus, in contrast to β-decay, the emission of γ-quanta is an intranuclear rather than intranucleon process.

The Coulomb barrier does not obstruct γ-radiation, and therefore, the lifetime of γ-active nuclei is usually small, (10^{-7}–10^{-15} s). However, in the cases when the transition to the ground state is strongly hindered, one observes long-lived excited states called *isomers*. The lifetime of isomers may reach several hours and even more. Some properties of isomers (in particular, probabilities of β-decay) may be observably different from those of the ground states.

A competitor of γ-decay is a process called *internal conversion*, in which an excited nucleus transmits the energy surplus to one of the electrons of the atomic shell. The spectrum of the outgoing electrons, in contrast to that of β-electrons, is discrete. Moreover, internal conversion is accompanied by the characteristic X-ray radiation.

► **Mössbauer effect.** If the energy of an incident photon coincides with the energy of the transition of a system from the ground state to an excited state, then the probability of such a photon being absorbed drastically increases. This phenomenon is called *resonant absorption*. The dependence of the absorption probability on energy has the shape of a sharp peak whose width is equal to that of the excited level ΔE ($\Delta E \sim \hbar/\Delta t$, where Δt is the lifetime of the excited state). The natural scheme of resonant absorption utilizes identical atoms as the emitter and the absorber. Such a scheme is used in optics, where the emission and the absorption are determined by electron transitions. However, observation of resonant absorption of nuclear γ-quanta encounters serious obstacles. The point is that when emitting a γ-quantum the atom experiences a much greater recoil and carries away more energy than in the case of emitting an optical photon. This leads to a more significant *decrease* of the frequency of the emitted photon and its absorption becomes a rare event (has small probability) (moreover, during absorption part of the photon energy is lost on the kinetic energy of the atomic recoil and the energy of the absorbed photon must be not *smaller* but *greater* than the energy of the excited level).

The momentum of nuclear recoil is equal to that of the photon: $p = \hbar\omega/c$, and the nuclear energy is equal to

$$E_{\text{nuc}} = \frac{p^2}{2m_{\text{nuc}}} = \frac{(\hbar\omega)^2}{2m_{\text{nuc}}c^2}. \tag{P8.2.1.11}$$

For resonance absorption, the photon energy shift (coinciding with nuclear energy E_{nuc}) should not exceed the level width ΔE. However, we have $\hbar\omega > 10^4$ eV, and $m_{\text{nuc}}c^2 < 2\times10^{11}$ eV; therefore, $E_{\text{nuc}} > 10^{-3}$ eV. At the same time, the common value of the excited level width is $\Delta E \sim 10^{-8}$–10^{-7} eV. We see that nuclear recoil causes a decrease of the γ-quantum energy by a quantity that exceeds the width of excited nuclear levels by many orders.

In 1958, Mössbauer discovered that at low temperatures resonant absorption is observed on the isotope $^{191}_{77}\text{Ir}$. The essence of the phenomenon known as the Mössbauer effect is the following. During both emission and absorption of γ-quanta by atoms forming a crystal lattice, the recoil momentum can be imparted to the entire crystal as a whole (more precisely, to a group of $N \sim 10^8$ atoms). This means that the recoil energy becomes negligibly small (m_{nuc} in (P8.2.1.11) should be replaced by the crystal mass) and the conditions of resonance absorption hold. There appears the possibility to fix changes of energy of γ-quanta of the level width order, i.e., $\Delta E \sim 10^{-8}$ eV. The important quantity here is not ΔE itself (for electron levels it is of the same order) but its *ratio* to the energy of photons (this ratio characterizes relative measurement errors). Since the energy of nuclear γ-quanta is 10^4–10^8 eV, the precision of measurements based on the Mössbauer effect reaches 10^{-12}–10^{-16}.

One of the first applications of the Mössbauer effect was the measurement of the gravitational red shift in the Earth's field of gravity (Pound and Rebka, 1960). The essence of the phenomenon is that as a photon moves in the gravitational field, its energy should change. The relative variation of the frequency and the energy should be equal to

$$\frac{\Delta E}{E} = \frac{(E/c^2)\Delta\varphi}{E} = \frac{\Delta\varphi}{c^2},$$

where $\Delta\varphi$ is the gravitational potential. If the Mössbauer emitter and receiver are placed on the same vertical line with the distance h between them, then $\Delta\varphi = gh$. For $h = 30$ m, the relative frequency variation $\Delta\omega/\omega = \Delta E/E = 3\times10^{-15}$ is within the precision range of the method.

The change of the frequency shift is measured on the basis of the Doppler effect. The emitter and the receiver being fixed, resonance absorption is absent because of the gravitational shift. Moving the receiver in the vertical direction, one finds the speed v at which there again appears resonant absorption. Then $\Delta\omega/\omega = v/c$.

► **Radiation dose measurement.** The *absorbed dose* of ionizing radiation is defined as the energy absorbed by a medium per unit mass and expressed in *greys* (Gy) (1 Gy = 1 J/kg).

The ionizing effect of X-rays and gamma rays on dry air is characterized by the *exposure dose*, defined as the net charge of all ions of like sign produced in one kilogram of irradiated air and measured in *roentgens* (R); $1\,\text{R} = 2.58\times10^{-4}$ C/kg. Alternatively, 1 R is defined as the amount of radiation required to generate 2.08×10^9 ion pairs in $1\,\text{cm}^3$ of dry air at standard temperature and pressure.

To evaluate the *biological effect* of radiation on living organisms, one uses the so-called *radiation quality factor* Q (also known as the *radiation weighting factor*), which indicates how many times a particular type of radiation is more dangerous than X-rays or gamma rays for the same dose absorbed. For beta radiation, $Q = 1$; for protons with energies less than 10 MeV, $Q = 10$; for thermal neutrons with energies less than 20 keV, $Q = 3$; for neutrons with energies 0.5–10 MeV, $Q = 10$; for alpha radiation with energy less than 10 MeV, $Q = 20$; for heavy nuclei, $Q = 20$. The *equivalent radiation dose* is the product of the absorbed dose and the quality factor. The equivalent dose is measured in *sieverts* (1 Sv is equal to 1 Gy for $Q = 1$). The unit *rem* is also used (roentgen equivalent in man or mammal, a biological equivalent of roentgen): 1 rem = 0.01 Sv.

P8.2.2. Nuclear Reactions

Nuclear reaction is a process of the type

$$a + {}^{A}_{Z}X \to {}^{A'}_{Z'}Y + b, \qquad \text{(P8.2.2.1)}$$

where a is the original (incident) particle, ${}^{A}_{Z}X$ is the original nucleus (target nucleus), ${}^{A'}_{Z'}Y$ is the final nucleus, b is the final (outgoing) particle. One often uses the following symbolic notation:

$${}^{A}X(a,b)^{A'}Y, \qquad \text{(P8.2.2.2)}$$

and the type of the process is briefly denoted by (a,b). For $b = a$ (and $X = Y$), the corresponding process is called *scattering*. If, in addition, the final nucleus Y is in the same state as the original X, then one speaks of *elastic scattering*.

For the reaction (P8.2.2.1) to be possible, it is necessary that conservation laws hold for the electric charge and some other charges (lepton, baryon, etc.), which is discussed in Subsection P8.4.2. Moreover, the states of the resulting particles and the original ones are related by the conservation laws for energy, linear momentum, and angular momentum.

Nuclear reactions are accompanied by changes in the nuclear structure, more precisely, these reactions are possible only if the distances between the reacting particles become smaller than the nuclear interaction radius $r_0 \sim 10^{-15}$ m.

► **Yield and cross-section of nuclear reaction.** Consider a flux of incident particles (a) of given energy hitting a fixed nuclear target X. We start with a thin target of width Δx and the number of nuclei per unit area being $N_0 = n_0\Delta x$ (n_0 is volume concentration of target nuclei). By definition, the *yield of nuclear reaction* is the fraction of incident particles (a) taking part in a given reaction (that is, the probability of the reaction):

$$w = \frac{\Delta N}{N}. \qquad \text{(P8.2.2.3)}$$

Clearly, the reaction yield w is proportional to N_0. The proportionality coefficient has dimensionality of area and is called the *cross-section the nuclear reaction*:

$$w = \sigma N_0 \qquad \sigma = \frac{\Delta N}{N N_0}. \qquad \text{(P8.2.2.4)}$$

Let us make a few remarks about the definition of the cross-section.

1. In Subsection P2.9.1, when discussing transfer phenomena, we introduced the effective cross-section of molecular scattering, whose meaning is similar to that of the nuclear reaction cross-section. Besides, the notion of the differential cross-section of elastic scattering was required in our analysis of Rutherford's experiments (see Subsection P6.1.1).

2. In (P8.2.2.4), the quantity N can be replaced by the particle flux (number of particles per unit time) or by the flux density J (flux per unit area), and the structure of the formula remains the same, since it involves the *relative number of particles*.

3. In the case of a thin target, it can be assumed that each incident particle interacts with only one target nucleus, and therefore, one can consider interaction of incident particles with a single nucleus. Since the number of particles hitting the target is equal to $N = JSt$ (J is flux density, S is target area) and $N_0 S$ is the total number of target nuclei, we find that the cross-section can be defined as the ratio of the number of particles reacting with a single nucleus per unit time to the particle flux density.

4. The yield and the cross-section are introduced for any selected interaction result. For a given input channel $(a + X)$ there are usually several output channels $(Y + b)$. Thus, the result of proton bombardment of nuclei ${}^7_3\mathrm{Li}$ can be ${}^7_3\mathrm{Li} + p$ (elastic and inelastic scattering), $\alpha + \alpha$, ${}^8_4\mathrm{Be} + \gamma$, etc. Some channels are closed for low energies. It should be kept in mind that one of the general statements of quantum mechanics is that any process which is not forbidden by any of the conservation laws is always realized (even if with small probability). In particular, among possible channels, there is always elastic scattering.

5. In order to describe the dependence of the final result of scattering on continuous parameters (such as the exit angle or energy of a particle), one introduces various types of *differential cross-section*. These are defined as the ratio of the number of reactions resulting in the said parameter (say, the angle θ) belonging to a certain narrow interval (say, between θ and $\theta + d\theta$) to the number of incident particles N multiplied by the number of target particles per unit area N_0 and the interval width:

$$\frac{d\sigma}{d\theta} = \frac{dN}{N N_0\, d\theta}$$

or

$$\frac{d\sigma}{d\Omega} = \frac{dN}{N N_0\, d\Omega},$$

where Ω is solid angle, with $d\Omega = \sin\theta\, d\theta\, d\varphi$. Integrating the differential cross-section over all possible values of the parameter (parameters), one obtains the total cross-section of the reaction.

6. To have a more clear idea of the cross-section, one can imagine that the particles are classical and their interaction has such a short range that they can be replaced by balls of radii R_1 and R_2, and the reaction occurs as soon as the particles come into contact. Then, $\sigma = \pi(R_1 + R_2)^2$ (such a model is used in the theory of transfer phenomena). This model has some bearing to reality in the cases of high energy of incident particles, if the de Broglie wavelength λ is small compared with the radius R of the nucleus (and there is no Coulomb interaction). In this situation,

$$\sigma \sim \pi R^2 \sim 10^{-30}\text{–}10^{-28}\ \mathrm{m}^2. \tag{P8.2.2.5}$$

For slow particles ($\lambda \gg R$), the cross-section drastically increases, since it is determined not by the nuclear radius but by the de Broglie wavelength:

$$\sigma \sim \pi\lambda^2.$$

However, this estimate pertains only to neutrons, since a slow charged particle cannot overcome Coulomb repulsion and approach the nucleus.

The estimate (P8.2.2.5) shows why nuclear cross-sections are expressed in *barns* (b) ($1\ \mathrm{b} = 10^{-28}\ \mathrm{m}^2$).

7. If the target is not thin, then the beam intensity decreases inside the target and the cross-section of the reaction is expressed through its yield by a more complex formula than (P8.2.2.4). This can be done, if one knows the total cross-section of the processes leading to beam attenuation. The simplest situation is that of the reaction under consideration being the only possible one. Then, for each layer one can write

$$\frac{dN}{N} = -\sigma n\,dx, \tag{P8.2.2.6}$$

where N is the number of beam particles that have reached the layer with the coordinate x, ndx is the number of target atoms in a layer of thickness dx (per unit area), n is volume concentration of atoms, the symbol (–) indicates that reacted particles abandon the beam. The solution of equation (P8.2.2.6) has the form

$$N = N_0 e^{-\sigma n x}, \tag{P8.2.2.7}$$

where N_0 is the original number of beam particles. The quantity $\alpha = \sigma n$ is called the *absorption coefficient*. It is easy to check that the mean free path is equal to

$$l = \frac{1}{\sigma n} \tag{P8.2.2.8}$$

(see Subsection P2.9.1).

Example. It is required to calculate the free path of neutrino in iron if the neutrino interaction cross-section is $\sigma_\nu \approx 10^{-19}\,\text{b} = 10^{-47}\,\text{m}^2$.

Solution. The iron atomic concentration is equal to $n = \rho N_\text{A}/\mu$, where N_A is Avogadro's number, $\rho = 7.8\times10^3\,\text{kg/m}^3$ is iron density, and $\mu = 58\times10^{-3}\,\text{kg/mol}$ is molar (atomic) mass of iron. We obtain $l = 1/(n\sigma) = 1.2\times10^{18}$ m, which is about 130 light years.

► **Reaction energy.** The energy conservation law for an arbitrary nuclear reaction can be written in the form

$$\sum_i m_i + \sum_i K_i = \sum_i m_i' + \sum_i K_i' \tag{P8.2.2.9}$$

(the masses are expressed in energy units). The *reaction energy* Q is defined as the kinetic energy increment (or mass decrease):

$$Q = \sum_i K_i' - \sum_i K_i = \sum_i m_i - \sum_i m_i' = \sum_i \Delta_i - \sum_i \Delta_i', \tag{P8.2.2.10}$$

where Δ is the mass defect (see Subsection P8.1.1). Note also that reaction energy can be expressed through nuclear binding energy ($Q = \sum_i E_{\text{b}i}' - \sum_i E_{\text{b}i}$), but only if the number of protons is preserved (if charges are carried only by nuclear protons). Reactions with energy release ($Q > 0$) are called *exothermic*, and those with $Q < 0$ are called *endothermic*. Sometimes, reaction energy is indicated in its notation, for example, $n + {}^{10}\text{B} \to {}^{7}\text{Li} + {}^{4}\text{He} + 2.79\,\text{MeV}$ or ${}^{4}_{2}\text{He} + {}^{14}_{7}\text{N} \to {}^{17}_{8}\text{O} + {}^{1}_{1}\text{H} - 1.2\,\text{MeV}$.

Since the last expression in (P8.2.2.10) does not depend on the reference frame, the reaction energy is equal to the kinetic energy increment in any reference frame, in particular, that of the center of mass

$$Q = \sum_i \widetilde{K}_i' - \sum_i \widetilde{K}_i. \tag{P8.2.2.11}$$

If the reaction involves an incident (non-relativistic) particle of mass m_a and a *stationary* target particle of mass M_X, then the kinetic energy of the system in its center-of-mass frame can be linked to the kinetic energy of the incident particle K:

$$\widetilde{K} = \frac{M_X}{m_a + M_X} K. \tag{P8.2.2.12}$$

The kinetic energy in the center-of-mass frame is expressed via the reduced mass $\mu = m_a M_X/(m_a + M_X)$ and the relative particle velocity $\mathbf{v}_{\rm rel} = \mathbf{v}_1 - \mathbf{v}_2$:

$$\widetilde{K} = \frac{\mu \mathbf{v}_{\rm rel}^2}{2} = \frac{M_X}{m_a + M_X} \frac{m_a \mathbf{v}_a^2}{2}$$

(see Subsection P1.6.4). It has been taken into account that the relative velocity does not depend on the reference frame and the target particle is at rest in the laboratory reference frame.

► **Threshold of nuclear reaction.** Endothermic reaction can proceed only if the kinetic energy of the incident particle is greater than some minimum value called *threshold of reaction*. The threshold of reaction $K_{\rm thr}$ does not coincide with $|Q|$, since due to the momentum conservation law the final kinetic energy in the laboratory reference frame differs from zero and $K_{\rm thr} = |Q| + K'$ (see (P8.2.2.10)). However, in the center-of-mass frame, on the reaction threshold the particles in the final state are at rest ($\widetilde{K}' = 0$) and the initial kinetic energy is equal to $|Q|$ (see (P8.2.2.11)). From (P8.2.2.12), we get

$$K_{\rm thr} = \frac{m_a + M_X}{M_X} \widetilde{K} = \frac{m_a + M_X}{M_X} |Q|. \tag{P8.2.2.13}$$

In the situation when the incident particle and decay products cannot be assumed non-relativistic, the threshold of endothermic reaction $m_a + M_X \to m_1 + m_2 + \cdots$ can be expressed through its energy $|Q| = (m_1 + m_2 + \cdots)c^2 - (m_a + M_X)c^2$ as follows. The quantity $E^2 - p^2c^2$ (here, E is the energy of the system and p is its momentum) is preserved during the reaction and is also unchanged, if one passes to another reference frame. Let us equate the initial value of this quantity in the laboratory reference system to its final value calculated in the center-of-mass frame (where $p = 0$):

$$[K_{\rm thr} + (m_a + M_X)c^2]^2 - p^2c^2 = (m_1 + m_2 + \cdots)^2 c^4.$$

Taking into account that the momentum of the target M_X is zero, for particle m_a we can write

$$p^2c^2 = (K_{\rm thr} + m_a c^2)^2 - m_a^2 c^4 = K_{\rm thr}(K_{\rm thr} + 2m_a c^2).$$

After suitable rearrangements, we get

$$(m_a + M_X)^2 c^2 + 2M_X K_{\rm b} = (m_1 + m_2 + \cdots)^2 c^2,$$

and therefore,

$$K_{\rm thr} = \frac{(m_1 + m_2 + \cdots)^2 - (m_a + M_X)^2}{2M_X} c^2 = |Q| \left(\frac{m_a + M_X}{M_X} + \frac{|Q|}{2Mc^2} \right). \tag{P8.2.2.14}$$

This formula can be applied also in the case of the incident particle being the γ-quantum, in which case one should take $m_a = 0$. The photon threshold energy is equal to

$$\varepsilon_{\rm thr} = \frac{(m_1 + m_2 + \cdots)^2 - M_X^2}{2M_X} c^2 = |Q| \left(1 + \frac{|Q|}{2M_X c^2} \right). \tag{P8.2.2.15}$$

Example. Calculate the threshold energy of γ-quantum for the production of an electron-positron pair on a proton and an electron at rest.

Solution. Note that the production of an electron-positron pair by a photon is impossible in empty space (see above), but it is possible in the field of a third particle (e.g., a proton). In this example, we are dealing with the reaction $\gamma + p \to e^- + e^+ + p$. Setting $m_a = 0$, $M_X = m_p$, and $m_1 + m_2 + \cdots = 2m_e + m_p$ in (P8.2.2.14), we obtain

$$\varepsilon_{\rm thr} = \frac{(2m_e + m_p)^2 - m_p^2}{2m_p} c^2 = 2m_e c^2 \left(1 + \frac{m_e}{m_p} \right) \approx 1.02\,\text{MeV};$$

cf. formula (P8.2.2.15) with $|Q| = 2m_e c^2$.

In the case where electron-positron pair production is due photon-electron interaction, we substitute m_p by m_e to obtain $\varepsilon_{\rm thr} = 4m_e c^2 = 2.04\,\text{MeV}$.

► **Mechanisms of nuclear reactions.** Just as there is still no universal theory accounting for all properties of nuclei, there is no unified theory of nuclear reactions. The features of different types of reactions in different energy intervals are described in terms of different models (or *mechanisms*, as they are also called) of nuclear reactions. Let us briefly dwell on some of those.

The mechanism of direct reactions (S. Butler, 1953) is aimed at explaining some features of reactions occurring at high energies, for instance, high anisotropy (maximum probability of forward emission), almost complete energy transfer to the outgoing fragment, etc. Such processes include, in particular, the *stripping reaction* (d,p), (d,n) and the *pickup reaction* (p,d), (n,d). In direct reaction, the incident particle reacts not with the entire nucleus but with its fragment — a nucleon or a small group of nucleons — transmitting energy to that fragment in a short (nuclear) time 10^{-22}–10^{-21} s. (This time is typical of reactions determined exclusively by nuclear forces. *Nuclear time* is estimated as $\tau_{\text{nucl}} = 2R/v \sim 10^{-23}$–$10^{-21}$ s, where $v \sim 10^7$–10^8 m/s is the nucleon speed.)

The mechanism of compound nucleus (N. Bohr, 1936) accounts for many features of reactions occurring at low energies. In the first place, it is the long duration (on a nuclear scale) of many reactions ($\tau \sim 10^{-16}$–10^{-12} s, which is by many orders greater than nuclear time $\tau_{\text{nucl}} \sim 10^{-22}$ s). Secondly, it is the symmetry (in the center-of-mass frame) of forward-backward particle emission (sometimes, even complete isotropy). Thirdly, it is the resonance character of reactions: their cross-sections have sharp maxima for some energy values. The essence of the mechanism of compound nucleus consists in that the reaction occurs in two stages:

$$a + X \to C^* \to Y + b. \tag{P8.2.2.16}$$

On the first stage, the parent nucleus X captures the incident particle and a compound nucleus C (in an excited state, which is indicated by the superscript $*$) is formed. The excitation energy is calculated as the sum of the binding energy of the particle a in the nucleus C (equal to $E_{aC} = m_X + m_a - m_C$) and the initial kinetic energy $\widetilde{K}$ (in center-of-mass frame):

$$E_{\text{exc}} = E_{aC} + \widetilde{K}. \tag{P8.2.2.17}$$

The process of particle capture and distribution of surplus energy between nucleons of the compound nucleus goes on for a short time $\sim \tau_{\text{nucl}}$. Then, the compound nucleus disintegrates in one of the possible ways (with different probabilities): either the surplus energy is randomly concentrated on a single particle (nucleon or a group of nucleons) and it abandons the nucleus, or a γ-quantum is emitted, or there is nuclear fission. All these processes are much slower than nuclear ones, which accounts for the slowness of the reactions occurring with compound nucleus formation.

The symmetry of emission of decay products is explained by the fact that the compound nucleus is capable of "forgetting" the direction of the incident particle. The resonance character of the reaction cross-section is connected with the fact that the probability of particle capture (and therefore, the cross-section of the entire process) increases sharply, if E_{exc} in (P8.2.2.17) coincides with one of the discrete energy levels of the compound nucleus. These values correspond to discrete values of kinetic energy $\widetilde{K}$, through which the energy of the incident particle is expressed (as shown above).

The mechanism of compound nucleus gives an acceptable explanation of the fact that reactions with different initial particles may have similar final results. Thus, the reactions

$$n + {}^{13}_{7}\text{N},\quad {}^{1}_{1}\text{H} + {}^{13}_{6}\text{C},\quad {}^{2}_{1}\text{H} + {}^{12}_{6}\text{C},\quad {}^{3}_{1}\text{H} + {}^{11}_{6}\text{C},\quad {}^{3}_{2}\text{He} + {}^{11}_{5}\text{B},\quad {}^{4}_{2}\text{He} + {}^{10}_{5}\text{B}$$

occur with the formation of one and the same compound nucleus ${}^{14}_{7}\text{N}^*$, and if their excitation energies turn out to be equal, then they will have identical output channels.

► **Statistical model.** For large excitation energies, the spectrum of the compound nucleus becomes practically continuous, and the *statistical model of decay*, or the *nuclear evaporation model*, becomes applicable. In this model, the compound nucleus is regarded as a drop of hot liquid with temperature $T = E_{\text{exc}}/k$ (k is the Boltzmann constant) and particle emission as evaporation from its surface. In the framework of this model, it is clear why within this energy range the angular distribution of emitted particles becomes isotropic (in the center-of-mass frame).

P8.3. Nuclear Power Engineering

Since maximum specific binding energy is characteristic of medium nuclei (with $A \sim$ 50–60), it is natural to expect that fission of heavy nuclei, as well as fusion (synthesis) of light nuclei is accompanied by energy release. The energy released in fission is called *atomic* and that released in fusion *thermonuclear*. The construction in the middle of the twentieth century of atomic and thermonuclear (hydrogen) bombs, in which the said reactions occur uncontrollably, put the very existence of humankind at risk. At the same time, controllable generation of energy at nuclear power plants is of great importance for power industry of many countries (in spite of the many ecological problems involved). The solution of the problem of controlled thermonuclear fusion would give a practically inexhaustible, ecologically safe source of energy. Moreover, reactions of fusion are the main source of energy of the sun and other stars.

P8.3.1. Nuclear Fission

► **Spontaneous fission.** In order to have a qualitative picture of fission phenomena and obtain some quantitative estimates, one can use the liquid-drop nuclear model and the corresponding Weizsäcker's formula (P8.1.3.1). For qualitative understanding, one can consider fission resulting in two identical fragments. Fissure leads to Coulomb energy drop and an increase of surface energy. The ratio of the corresponding terms in (P8.1.3.1) depends on the parameter $\kappa = Z^2/A$, and fission probability also depends on this parameter. It is not difficult to check that fission becomes energetically efficient for

$$\frac{Z^2}{A} > \kappa_1 \approx 18. \qquad \text{(P8.3.1.1)}$$

The substitution of actual values of Z and A shows that fission starts to be efficient for nuclei in the niobium–palladium range, i.e., starting from $A \sim 100$. However, spontaneous fission is observed only for very heavy nuclei (with $A \sim 250$).

This is due to the fact that although the two-fragment state has a smaller energy, the transition to that state requires overcoming a substantial potential barrier. Subject to gradual deformation, the nucleus passes through intermediate states whose energy is greater than that of the initial state. Figure P8.6 schematically represents the dependence of the energy of the deformed nucleus on a conditional parameter γ describing the degree of deformation. The barrier height equal to the difference between the maximum intermediate energy and the initial energy of the nucleus is called *activation energy* E_{act}. If excitation energy received by the nucleus exceeds E_{act}, it quickly splits into two fragments. Activation energy decreases with the growth of Z^2/A and vanishes for

$$\frac{Z^2}{A} > \kappa_2 \approx 50. \qquad \text{(P8.3.1.2)}$$

In the absence of a barrier, the nucleus instantly splits into two fragments without receiving any additional energy. Due to quantum tunneling, spontaneous fission from the ground state is also possible for $Z^2/A < \kappa_2$, but the barrier should be not too high; otherwise, fission probability would be very small. Thus, for ${}^{238}_{92}\mathrm{U}$, we have $Z^2/A = 35.5$, which is considerably less than κ_2, and the and half-life of ${}^{238}_{92}\mathrm{U}$ with respect to spontaneous fission is $T_{1/2} \sim 10^{16}$ years, which is by 6 orders greater than $T_{1/2}$ of this element with respect to α-fission. On the other hand, for the artificial nucleus with $Z = 107$ and $A = 261$, we have $Z^2/A = 43.9$, and its half-life with respect to spontaneous fission is $T_{1/2} \sim 10^{-3}$, which is comparable with $T_{1/2}$ with respect to α-fission.

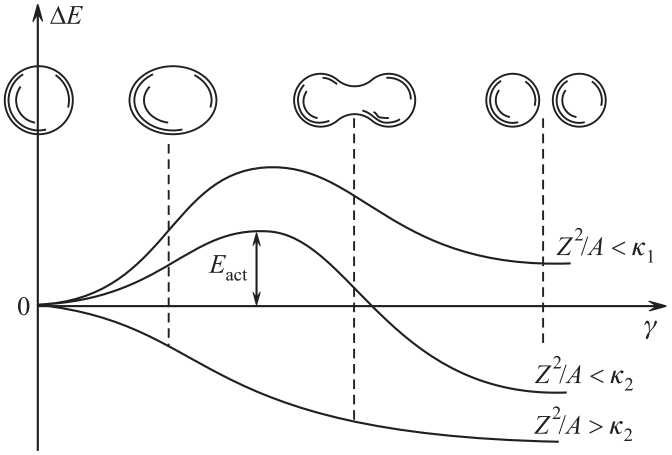

Figure P8.6. Energy of intermediate states of a nucleus in the process if its fission.

Specific binding energy of nuclei with $A \sim 240$ is by 0.9 MeV less than that of fission fragments. This means that in every instance of fission, about 200 MeV energy is released. More than 80% of that energy is carried away as kinetic energy of the fragments. The remaining energy goes for the emission of neutrons and γ- and β-particles by the fragments.

As a rule, fission results in fragments with different masses. Light fragments are grouped around Kr, and heavy ones around Xe. This is explained by the fact that the numbers of neutrons in these nuclei are close to magic numbers 50 and 82.

Since the fraction of neutrons in heavy nuclei is considerably greater than in light nuclei, daughter nuclei have an excess number of neutrons. Immediately after fission, daughter nuclei emit 2 or 3 neutrons with kinetic energy ~ 2 MeV, after which there occur several β-decays that bring the fragments to a stable state. Neutron emission by fission products plays the key role in ensuring the chain reaction of fission.

Of great importance for the *controlled* chain reaction are the so-called *delayed neutrons*. These are produced in the case when a sufficiently slow β-decay of a fission product results in excited nuclear state with small energy of separation of a neutron, upon which the neutron is instantly emitted. The delay is determined by the β-decay time and may reach tens of seconds.

► **Fission of excited nuclei.** When a nucleus receives a sufficient amount of excitation energy (several MeV), it starts to oscillate with its shape periodically changing from oblate to prolate. Even if the excitation energy is less than the activation energy, the probability of fission of an excited nucleus is substantially higher than in the ground state. A competitive process is that of emission of excitation energy in the form of γ-quanta.

An excited nucleus with a considerable fission probability can be obtained by exposing it to different types of particle radiation. The most effective is exposure to neutrons; there is no Coulomb barrier for neutrons and they can approach the nucleus even if their kinetic energy is very small (of the order of 1 keV).

► **Neutron capture.** The principal mechanism of fission of natural uranium isotopes, ^{235}U and ^{238}U, is neutron capture with the formation of compound nuclei ^{236}U and ^{239}U in excited state. It is important that ^{235}U is broken down by any, in particular, thermal neutrons (with energies less than 0.5 keV), while ^{238}U only by fast neutrons (with energies of the order of 2 MeV). The point is that the compound nucleus ^{236}U is of the even-even type and for the same amount of neutron energy, its excitation energy turns out greater than that of the even-odd nucleus ^{239}U.

Excitation energy of a compound nucleus can be found from the equation

$$m_n + K_n + M_A = M_{A+1} + E_{\text{exc}},$$

where K_n is the kinetic energy of the neutron (the kinetic energy of the compound nucleus is neglected). We get

$$E_{\text{exc}} = (m_n + M_A - M_{A+1}) + K_n = \varepsilon_{A+1} + K_n,$$

where ε_{A+1} is the neutron binding energy in the compound nucleus. A compact nucleus undergoes fission, if E_{exc} becomes greater than activation energy. For ^{239}U, we have $E_{\text{act}} = 7.1$ MeV, and $\varepsilon_{A+1} = 5.5$ MeV; therefore, when exposed to neutron radiation, the nucleus ^{238}U requires for fission K_n not less than 1.6 MeV. For ^{236}U, we have $\varepsilon_{A+1} = 6.8$ MeV, which is greater than $E_{\text{act}} = 6.5$ MeV, and thus the nucleus ^{235}U will disintegrate when exposed to any neutrons.

The cross-section of capture of slow neutrons is determined by their de Broglie wavelength, which can be much larger than the nuclear radius. The cross-section of capture of thermal neutrons with energy 0.025 keV is hundreds of times greater than the cross-section of capture of fast neutrons with energy 1 MeV, whence the great importance of fission caused by thermal neutrons.

► **Chain reaction.** The possibility of self-sustained reaction of fission (chain reaction) is due to the fact that upon nuclear fission by a *single neutron*, fission fragments emit *two or three* neutrons, which, in turn, can cause fission of other nuclei. The problem is that some of the produced neutrons drop out of the reaction: are either absorbed without causing fission or just leave the system through its boundary.

One defines the *neutron multiplication coefficient* as the ratio of the number of neutrons in the next generation to that in the preceding one:

$$k = N_{i+1}/N_i.$$

For the chain reaction to start, it is necessary that k be greater than unity. In order to keep the reactor steady, k should be maintained equal to unity.

With the increase of the system's dimensions, the fraction of neutrons abandoning the system through its boundary tends to zero. Therefore, whereas $k_\infty > 1$ for an infinite nuclear medium, the condition $k > 1$ for a finite system of a certain shape will hold, if the mass of that system exceeds some critical value called *critical mass*. The critical mass can be decreased if the active object is surrounded by a shell that reflects neutrons. Thus, for the pure element ^{235}U the critical mass is 47 kg (ball of radius 17 cm), but if ^{235}U is layered by polyethylene films and coated by a beryllium shell, the critical mass reduces to 250 g (ball of radius 3 cm). Next we consider the chain reaction in infinite nuclear medium.

Any nuclear fuel with reasonable production cost contains a high percentage of ^{238}U. Thus, natural uranium contains 99.3% ^{238}U and only 0.7% ^{235}U. Although ^{238}U is also capable of neutron fission, its probability is too small (and for neutrons with energies less than 1.6 MeV this probability is equal to zero), and the rest of the neutrons are absorbed by ^{238}U without fission and with emission of γ-quanta (*radiation capture* (n,γ)). At the same time, the probability of nuclear fission of ^{235}U (and also ^{233}U and ${}^{239}_{94}$Pu) is fairly high, 85–90%, and these elements can be split by both fast and slow neutrons.

High enrichment of uranium (i.e., a substantial increase of ^{235}U concentration) is a very costly process and is used only for the production of nuclear bomb charges (where economical factors are inessential) and the construction of breeders—fast neutron reactors (discussed below).

Chain reactions are mainly obtained with the use of slow (thermal) neutrons. Fission neutrons have mean energy $\sim 2\,\mathrm{MeV}$. For this energy, the cross-sections of neutron capture by nuclei of ^{235}U and ^{238}U are of the same order. However, for thermal neutrons the situation is quite different. For ^{238}U, the capture cross-section remains small (≈ 2.8 b), and for ^{235}U, it sharply increases and is ≈ 650 b (which is much greater than the geometrical cross-section); moreover, the greater part (84%) of absorbed neutrons causes fission (fission cross-section is 550 b).

It is easy to calculate that for thermal neutrons the cross-section of fission of a “medium” nucleus of natural uranium is $550\times 0.007 = 3.9$ b, and the cross-section of radiation capture is $2.8+100\times 0.007 = 3.5$ b. We see that the store of neutrons is not very large (2.5 neutrons are born, on the average, in each fission event, but half of these are lost); therefore, for the realization of chain reaction it is important to decrease the loss of neutrons in the process of their moderation. The best moderator, from the mechanical standpoint, should be hydrogen (contained in common water molecules), since in elastic collision the greatest energy transfer occurs in the case of equal masses of the colliding particles (see Subsection P1.5.3). However, hydrogen atoms easily absorb neutrons (${}^{1}\mathrm{H}+n \to {}^{2}\mathrm{H}+\gamma$), and therefore, deuterium (contained in heavy water), carbon, or beryllium are used. In order to reduce the absorption of moderated neutrons by nuclei of ^{238}U, the nuclear fuel, instead of being mixed with the moderator to form a homogeneous structure, is set in layers that alternate with layers of the moderator to form a heterogeneous structure. Getting into a moderator layer, the neutron slows down without the risk of being absorbed, and then goes back to the region of reaction.

Had the chain reaction been due to merely fission neutrons, such a reaction could hardly be controlled. Indeed, the lifetime of a single generation of fission neutrons is $\sim 10^{-3}$ s. If the multiplication coefficient randomly increases, say, to $k = 1.005$, then in 1 s the number of neutrons increases 150 times, which means that the reaction turns into explosion. The stability of controlled reactions is achieved by way of *delayed neutrons* mentioned above. Even though their fraction is not great ($\sim 0.75\%$), the average lifetime of their generation is ~ 0.1 s, and in the above example the number of neutrons would increase only 1.5 times instead of 150.

► **Transuranic elements.** The only natural fissile element is ^{235}U, but its resources are not too great. Though reactors on fast neutrons require a substantial enrichment of uranium (up to 15%) and are not very powerful, they can be used for a special purpose—to produce a new type of nuclear fuel (this is why they are called *breeders*). Thus, radiation capture of neutrons by ^{238}U, which is so harmful for chain reactions, leads to the production of the artificial transuranic element ${}^{239}_{94}$Pu:

$$
\begin{aligned}
&{}^{238}_{92}\mathrm{U} + n \to {}^{239}_{92}\mathrm{U} + \gamma; && \\
&{}^{239}_{92}\mathrm{U} \to {}^{239}_{93}\mathrm{Ne} + e^{-} + \widetilde{\nu}; && T_{1/2} = 2.3\ \text{min}; \\
&{}^{239}_{93}\mathrm{Ne} \to {}^{239}_{94}\mathrm{Pu} + e^{-} + {}^{\sim}\nu && T_{1/2} = 2.3\ \text{days}.
\end{aligned}
$$

Plutonium is practically stable (its half-life is 24,000 years) and is easily fissionable by neutrons with any energy. Plutonium can be more easily separated from ^{238}U, since their chemical properties are different.

P8.3.2. Fusion Reactions

As mentioned above, for light nuclei, the binding energy ε increases with the growth of A (see Fig. P8.2). Thus, reactions of fusion of two light nuclei with the formation of a heavier nucleus (reactions of synthesis) should occur with energy release. Since the growth of $\varepsilon(A)$ for small A is steeper than its drop for large A, much more energy (per one nucleon) is released in fusion than in fission.

► **Conditions of fusion.** Fusion of two nuclei is possible only if they approach one another by a distance at which nuclear forces start to be active, which is hindered by Coulomb repulsion. To overcome the Coulomb barrier the nuclei must have large kinetic energies. If the substance is heated up to a high temperature, then conditions of self-sustained reaction may arise, and energy release will compensate its loss (for instance, by radiation). For this reason, reactions of fusion are called thermonuclear reactions. Such reactions are the main source of energy of the sun and other stars.

Note that the temperatures at which fusion is possible are so high that the substance acquires the state of fully ionized plasma.

Example. Estimate the temperature at which fusion of two protons to deuteron can occur.

Solution. For this reaction to take place, the protons must approach one another by the distance equal to two effective radii of nuclear forces $r_0 \sim 2\times 10^{-15}$ m (see Subsection P8.1.1). From the energy conservation law

$$2E_{\text{kin}} = \frac{1}{4\pi\varepsilon_0}\frac{e^2}{2r_0}$$

we find the kinetic energy $E_{\text{kin}} \sim 0.1\,\text{MeV}$, and then, using the classical formula $E_{\text{kin}} = \frac{3}{2}k_{\text{B}}T$, estimate the temperature, $T \sim 10^9$ K.

► **Thermonuclear energy of the sun and the stars.** The temperature in the center of the sun is by two orders less than the threshold value obtained above. This means that the *mean* nuclear energy is insufficient for fusion to take place. Still, the reaction occurs, and the energy released by the sun is huge: $\sim 4\times 10^{26}$ W. The point is that, according to the Maxwellian distribution, a ceratin (very small) part of particles possesses energy much larger than the mean value. Moreover, due to quantum tunneling, nuclei with energies less than the Coulomb barrier can participate in the reaction. All this means that there is only a trace amount of colliding nuclei that react. For this reason the specific energy released by the sun per unit time is so small, $\sim 10^{-4}$ W/kg, but owing to its huge mass, the total energy release is so great.

Thermonuclear energy of the sun is released in two cycles of thermonuclear reactions: the *carbon cycle* and the *hydrogen cycle*. Both these cycles result in the transformation of four protons into a helium nucleus (with the emission of positrons, photons, and neutrino), with the total energy release ~ 25 MeV. The carbon cycle requires the presence of carbon nuclei, which play the role of a catalyst, since they are restored as a result of the cycle. The reactions of this cycle are the following:

$$\begin{aligned}
&{}^1_1\text{H} + {}^{12}_{6}\text{C} \to {}^{13}_{7}\text{N} + \gamma;\\
&{}^{13}_{7}\text{N} \to {}^{13}_{6}\text{C} + e^+ + \nu;\\
&{}^1_1\text{H} + {}^{13}_{6}\text{C} \to {}^{14}_{7}\text{N} + \gamma;\\
&{}^1_1\text{H} + {}^{14}_{7}\text{N} \to {}^{15}_{8}\text{O} + \gamma;\\
&{}^{15}_{8}\text{O} \to {}^{15}_{7}\text{N} + e^+ + \nu;\\
&{}^1_1\text{H} + {}^{15}_{7}\text{N} \to {}^{12}_{6}\text{C} + {}^4_2\text{He}.
\end{aligned}$$

The second cycle contains three different reactions: the first and the second take place twice (altogether five reactions):

$$
\begin{aligned}
&{}^1_1\mathrm{H} + {}^1_1\mathrm{H} \to {}^2_1\mathrm{H} + e^+ + \gamma;\\
&{}^1_1\mathrm{H} + {}^2_1\mathrm{H} \to {}^3_2\mathrm{He} + \gamma;\\
&{}^3_2\mathrm{He} + {}^3_2\mathrm{He} \to {}^4_2\mathrm{He} + 2{}^1_1\mathrm{H}.
\end{aligned}
$$

At temperatures up to 1.5×10^7 K, the hydrogen cycle makes a larger contribution to energy release; at higher temperatures, a larger contribution is due to the carbon cycle. For the sun, both cycles are equally important. In less bright stars, the hydrogen cycle dominates and in brighter stars, the carbon cycle does. For giant stars, other cycles are crucial (e.g., the helium and neon cycles).

► **Controlled fusion.** Fusion control would give mankind a practically inexhaustible energy source. In contrast to reactions occurring inside stars, energy should be obtained here from a fairly small amount of nuclear fuel with maximum efficiency. To that end, the substance should be heated up to a temperature at which the probability of nuclear fusion at collisions is sufficiently large. Moreover, one has to find a way to confine incandescent plasma for a sufficiently long time and extract the released energy, transforming it, say, to electricity.

The major technological problem is plasma confinement. In this connection, the most promising method is to place plasma in magnetic field of a certain structure (magnetic trap). In a *tokamak*, plasma is confined by magnetic field produced by the current in the plasma itself (the same current is first used for heating plasma). A fairly long-time plasma confinement is hindered by various types of instabilities that arise and quickly develop in hot plasma placed in magnetic field.

For each reaction, there is a characteristic temperature T_0 to which plasma should be heated in order to ensure the most efficient extraction of energy. The second characteristic parameter is the minimum time τ of plasma confinement necessary for energy gain (more precisely, the product of time and plasma concentration, $n\tau$).

Proton–proton fusion reaction fuels the sun and could be most advantageous as regards energy release; furthermore, protons are abundant on Earth (hydrogen is contained in water). Unfortunately, the cross-section of this reaction is so small that it cannot be considered as a basis for controlled fusion. Equally attractive seem *deuterium–deuterium* reactions

$$
\begin{aligned}
&{}^2_1\mathrm{H} + {}^2_1\mathrm{H} \to {}^3_2\mathrm{He} + n + 3.25\,\mathrm{MeV},\\
&{}^2_1\mathrm{H} + {}^2_1\mathrm{H} \to {}^3_1\mathrm{H} + {}^1_1\mathrm{H} + 4.03\,\mathrm{MeV},
\end{aligned}
$$

since deuterium constitutes 0.015% of all hydrogen resources. This reaction is characterized by the parameters $T_0 \approx 10^9$ K and $n\tau > 10^{22}\,\mathrm{s/m^3}$. However, the next most promising candidate is the *deuterium–tritium* reaction

$$
{}^2_1\mathrm{H} + {}^3_1\mathrm{H} \to {}^4_2\mathrm{He} + n + 17.6\,\mathrm{MeV},
$$

which has a considerably greater power efficiency and better critical parameters: $T_0 \approx 2\times 10^8$ K and $n\tau > 10^{20}\,\mathrm{s/m^3}$, which compensates for the difficulties connected with the production of tritium.

P8.4. Subatomic Particles

In this section, we give some brief reference information with regard to classification of subatomic particles, quantum numbers, conservation laws, and quark structure of hadrons.

P8.4.1. Classes of Subatomic Particles

In modern particle physics, two main types of subatomic particles are distinguished: elementary particles, which are not known consist of other particles, and composite particles. Elementary particles include quarks, leptons, gauge bosons, photons, and gluons. Composite subatomic particles consist of two or more elementary particles and include baryons (e.g., protons and neutrons) and mesons.

Remark. For historical reasons, all subatomic particles are very often called elementary particles and the structureless particles are called truly elementary particles or truly fundamental particles.

► **Fundamental interactions.** Transformations and decay of subatomic particles are governed by the following fundamental interactions: strong, electromagnetic, and weak (gravitational interaction is inessential here). The weaker the interaction, the slower the associated processes. *Strong interaction* has the range 10^{-15} m and the characteristic time of the associated processes is 10^{-23} s. Strong interaction binds nucleons in atomic nuclei. *Electromagnetic interaction* is long-range (it has an infinite range), the characteristic time of the associated processes is not less than 10^{-20} s. A testing ground of this most well-known type interaction is the control of atomic electrons, i.e., of all diversity of substances and chemical reactions. *Weak interaction* has the range 10^{-18} m, the characteristic time of the associated processes (involving neutrino, as a rule) is not less than 10^{-10} s. It is weak interaction that controls fusion of two protons, which is crucial for energy processes in the sun, and it also controls various types of β-decay. Weak interaction is the most universal of the three; it is always present, but its determining character is manifested only in the absence of the other two.

Fundamental interactions are realized by the exchange of special particles that belong to the class of *fundamental vector bosons* (all these particles have unit spin). The interaction carriers are produced by one of the interacting particles and are absorbed by the other. The produced particles are *virtual*, they do not obey the relation $E^2 - p^2c^2 = m^2c^4$, which is allowed by the uncertainty principle $\Delta E\Delta t \sim \hbar$, provided that their lifetime is small. If the mass of the carriers differs from zero, then the interaction radius is equal to their Compton wave-length $\hbar/mc$ (for details, see Subsection P8.1.2). The electromagnetic interaction carrier is a massless particle called a *photon*, the well-known quantum of electromagnetic field. Weak interaction carriers are heavy *intermediate bosons* W^+, W^- and Z, whose masses are almost equal to 100 GeV. Strong interaction carriers are called *gluons*—the quanta of the eight gluon fields that realize interaction between quarks. (In the early π-meson theory of nuclear forces described in Subsection P8.1.2, π-mesons were regarded as strong interaction carriers). The gluon mass is equal to zero, but in contrast to photons, gluons interact with one another and with quarks and cannot go away farther than 10^{-15} m.

► **General properties of particles.** We are going to describe different classes of subatomic particles. First, we list some general properties.

1. Particles can be stable (electron, proton, neutrino) or unstable. In the last case, one indicates the particle lifetime (or width of the energy interval $\Delta E = \hbar/\Delta t$), as well as decay channels. Particles with lifetime $\sim 10^{-23}$ s are called *resonance particles*. Particles whose lifetime is much greater than the said value are called quasi-stable.

2. Particles can be either composite (consisting of other particles; e.g., nucleons consist of quarks) or, for presently accessible energy levels, structureless (photon, electron, neutrino, quarks).

3. A particle can be either a boson (integer spin) or fermion (half-integer spin).

4. Almost any particle has its counterpart—an antiparticle whose charges (all its charges, not only the electric one) have opposite signs. Thus, electron corresponds to positron, proton to antiproton, neutron to antineutron. A particle that coincides with its antiparticle is called *truly neutral* (for example, photon or π^0-meson).

5. All particles are divided into strongly interacting hadrons (baryons and baryon resonances, mesons and meson resonances), leptons (fermions not involved in strong interaction) and fundamental bosons (interaction carriers).

► **Hadrons and quarks.** The term *hadron* refers to all particles involved in strong interaction. Hadrons are divided into two large classes: *baryons* (hadrons with half-integer spin, i.e., fermions) and *mesons* (hadrons with integer spin, i.e., bosons). The class of baryons contains, in particular, nucleons, and that of mesons contains π-mesons.

Hadrons are quite numerous (more than 400 hadrons have been discovered). A systematic classification of hadrons and understanding of many of their properties were not possible until the appearance (in the 1960s) of the hypothesis of *quark structure of hadrons*, according to which all hadrons consist of elementary particles called *quarks*. All quarks have spin 1/2, and therefore, are fermions. There are six varieties of quarks (called *quark flavors*): u, d, s, c, b, and t (from *up*, *down*, *strange*, *charm*, *bottom*, and *top*). The masses of the quarks are: $m_u \approx 5$ MeV, $m_d \approx 7$ MeV, $m_s \approx 150$ MeV, $m_c \approx 1.3$ GeV, $m_b \approx 4.3$ GeV, and $m_t \approx 175$ GeV. The most extraordinary property of quarks is that they possess fractional electric charges. The quarks u, c, and t are called upper quarks and their charge is equal to $+2/3$; the quarks d, s, and b are called lower quarks and their charge is equal to $-1/3$ (in elementary charge units).

All baryons consist of three quarks, and all mesons consist of a quark and an antiquark. Nucleons consist of light quarks u and d: proton contains two u-quarks and one d-quark ($p = uud$), neutron contains two d-quarks and one u-quark ($n = ddu$). These quarks also form π-mesons: $\pi^+ = u\widetilde{d}$ and $\pi^- = d\widetilde{u}$ (a tilde over a quark symbol denotes the corresponding antiquark); π^0 is a quantum superposition of two states: $\pi^0 = (u\widetilde{u} + d\widetilde{d})/\sqrt{2}$. We see that π^+ and π^- are mutual antiparticles, and π^0 coincides with its antiparticle. Particles containing an s-quark are called strange and those with a c-quark are called charmed particles.

It should be immediately observed that, at first sight, the masses of hadrons are much greater than the sum of the masses of their quarks. But the mass of any bound system should be smaller than the sum of the masses of its constituents by the quantity equal to the binding energy. The point is that the quarks in the hadrons are surrounded by a cloud of virtual gluons, sometimes called the "gluon coat." The masses of "coated" quarks are about 350 MeV greater than the masses of "naked" quarks.

Baryons that contain more heavy quarks than u and d are called "hyperons." From quarks u, d, and s, six hyperons with spin $1/2$ can be composed: Λ^0 (uds, $m \approx 1120$ MeV), Σ^+, Σ^0, Σ^- (uus, uds, dds, $m \approx 1200$ MeV), Ξ^0, Ξ^- (uss, dss, $m \approx 1320$ MeV). Decay of all these hyperons is determined by weak interaction, and therefore, their lifetime is $\tau \sim 10^{-10}$ s (except for the Σ^0-hyperon whose decay is determined by electromagnetic interaction, its lifetime is $\tau \sim 10^{-20}$ s). Together with proton and neutron, these hyperons form an octet of baryons with spin $1/2$. (The difference between Λ^0 and Σ^0 is explained below.)

Next, we have the decuplet of baryons with spin $3/2$, among which there are baryons consisting of three identical quarks, for example, Δ^-, Δ^0, Δ^+, Δ^{++} (ddd, udd, uud, uuu, $m \approx 1230$ MeV). The heaviest in the decuplet is the Ω^--hyperon (sss, $m \approx 1670$ MeV). Note that subatomic particles consisting of the same quarks and differing only in spin (or other quantum numbers) often bear different names (e.g., neutron and Δ^0-hyperon).

Further, we briefly describe mesons. In addition to the three π-mesons, there are six mesons having zero spin and orbital momentum and containing an s-quark: the four K^+, K^0, $\widetilde{K}^0$, K^- ($u\widetilde{s}$, $d\widetilde{s}$, $s\widetilde{d}$, $s\widetilde{u}$, $m \approx 500$ MeV) and two truly neutral mesons η and η' (different quantum superpositions of the states $u\widetilde{u}$, $d\widetilde{d}$ and $s\widetilde{s}$). Next, we have nine mesons with spin $S = 1$, among which there are three ρ, four K^*, ω^0, and φ^0. All these mesons (as well as η') are of resonance type $\tau \sim 10^{-23}$ s).

An important event of 1974 was the discovery of the first charmed particle, J/ψ-meson (m = 3.1 GeV); this particle represents one of the states of *charmonium* $c\widetilde{c}$ (charmonium has "hidden charm"). Thereafter, other charmed mesons were found (for example, $D^+ = c\widetilde{d}$, $D^0 = c\widetilde{u}$, and $D_s^+ = c\widetilde{s}$) and charmed baryons (for example, $\Lambda_c^+ = udc$), and also more heavy "beauty" hadrons containing a b-quark.

► **Quark color and gluons.** The existence of baryons consisting of three identical quarks is in contradiction with the Pauli principle, since two of them must have the same spin direction. Moreover, the Ω^--hyperon has spin 3/2; i.e., all its three s-quarks can be in the same state (when the hyperon spin projection is maximum). How can this be explained? It turned out that there are *three* quarks with the same flavor and not just a single such quark. The distinction between these is described by an additional quantum number that can take three values. It was proposed to use the term *color* for that number (or color charge) and ascribe to it three values: R (red), G (green), and B (blue). Accordingly, the colors of antiquarks are denoted by $\widetilde{R}$, $\widetilde{G}$, $\widetilde{B}$ (anti-red, anti-green, and anti-blue). Thus, there are altogether 18 quarks and 18 antiquarks. The mixture of the three colors is regarded as colorless (white). A quark–antiquark pair possesses a "hidden" color.

The color charges of quarks are sources of gluon fields (just as the electric charge is the source of electromagnetic field). The quanta of these fields, called *gluons* (from "glue"), realize strong interaction. (Electromagnetic interaction is realized by the exchange of photons, i.e., quanta of electromagnetic field.) Both gluons and photons are massless particles with spin 1 and negative parity (see below). Photons, however, are electrically neutral, while gluons are "colored"; i.e., take part in strong interactions. Each gluon carries two color charges: one color and one anti-color. There are eight different gluons: six with explicit color, ($g_{R\widetilde{G}}$, $g_{R\widetilde{B}}$, $g_{G\widetilde{R}}$, $g_{G\widetilde{B}}$, $g_{B\widetilde{R}}$, $g_{B\widetilde{G}}$) and two with hidden color.

According to the color charge conservation law, a quark must change its color, when emitting or absorbing a gluon. Thus, when a red R-quark emits a gluon $R\widetilde{G}$, it becomes a G-quark, but its flavor does not change:

$$u_R \to u_G + g_{R\widetilde{G}}.$$

(Quark flavor can change only in weak interactions; i.e., when W- or Z-bosons are emitted.)

Since gluons possess color charges, they can emit and absorb gluons and be involved in mutual strong interaction. Strong interaction, regarded as a result of emission and absorption of gluons by quarks and gluons, is studied by *quantum chromodynamics* (by analogy with *quantum electrodynamics*, which studies electromagnetic interaction as a result exchange of photons). One of the basic statements of quantum chromodynamics is the nonexistence of free colored objects. Free existence is allowed only for such color charge combinations that carry no color charge as a whole. For example, free hadrons can only exist in the "white" state, its three quarks being of different colors at each time instant. (Still, each quark continuously changes its color, emitting or absorbing virtual gluons that have been either emitted by another quark or produced in its "gluon coat"—cloud of virtual gluons. But the total color charge of the hadron remains white.) The same principle applies to mesons. For example, π^+-meson is a quantum superposition of three color states: $u_R\widetilde{d}_{\widetilde{R}}$, $u_G\widetilde{d}_{\widetilde{G}}$, $u_B\widetilde{d}_{\widetilde{B}}$. Each of these states has a hidden color, but their superposition is "white."

► **Confinement.** Why is it impossible to divide a white object into two colored ones (for instance, to remove one quark from a meson or hadron)? First of all, the exchange of colored gluons leads to an interaction whose potential grows with the distance not slower than the linear function (the interaction force does not decrease). A possible explanation is that the gluon field is not scattered in all directions (this would correspond to the Coulomb potential), but is concentrated in a narrow “pipe” (string) joining the quarks. Therefore, infinite energy would be required to split the quarks. In fact, as the quarks are separated from one another by a distance greater than 10^{-15} m, the gluon string is “broken,” a pair quark–antiquark is produced, and the removed quark turns into a white meson. Attempting to obtain two colored objects, we again obtain two white objects. (Note that for large energies of scattered quarks, each of these transforms not into a single meson but into several hadrons, and the so-called “hadron jet” is produced. Observation of hadron jets was an important evidence supporting the existence of quarks.) In this way, quantum chromodynamics explains the failure of all attempts to detect free quarks or knock them out of hadrons, in spite of the fact that the existence of quarks within hadrons was confirmed by experiments with scattering of high energy electrons on hadrons. This behavior of quarks is referred to as “confinement.”

► **Leptons.** Leptons are fermions that do not participate in strong interactions. There are three pairs of leptons: electron e^- and electron-type neutrino ν_e, muon μ^- and muon-type neutrino ν_μ, tau-lepton τ^- and tau-neutrino ν_τ. The spin of all leptons is equal to 1/2. Each charged lepton has its antiparticle: e^+, μ^+ and τ^+. There is no final answer to the question whether neutrinos are truly neutral particles or have antiparticles—antineutrinos, $\tilde{\nu}_e$, $\tilde{\nu}_\mu$ and $\tilde{\nu}_\tau$. Leptons are fundamental particles—no structure has been detected in them. As shown by experiments, each charged lepton takes part in weak interactions coupled to its neutrino, for example:

$$n \to pe^-\tilde{\nu}_e, \qquad \pi^+ \to \mu^+\nu_\mu, \qquad \tau^+ \to \tilde{\nu}_\tau e^+\nu_e.$$

The masses of muon and tau-lepton are respectively equal to 106 MeV and 1784 MeV. These particles are unstable: muon decays along the channel $\mu^- \to e^-\tilde{\nu}_e\nu_\mu$, its lifetime is 2×10^{-6} s; the heavier tau-lepton decays along many channels, its lifetime is 5×10^{-13} s. For the neutrino masses, upper bounds can be given: $m_{\nu_e} < 30$ eV, $m_{\nu_\mu} < 0.5$ MeV, and $m_{\nu_\tau} < 150$ MeV.

► **Fundamental fermions.** Six leptons and six quarks form the class of *fundamental fermions* of which all other subatomic particles are “made.” A special role is played by the “first generation” of fundamental fermions u, d, ν_e, e^-, which compose nucleons and atoms and determine almost all processes observed in nature. The “second generation” of fundamental fermions (c, s, ν_μ, μ^-) and their “third generation” (t, d, ν_τ, τ^-) are artificially obtained in powerful accelerators. However, only the investigation of all fundamental particles in their entirety may lead to the understanding of the structure and the evolution of our universe.

All processes in the world of subatomic particles amount to interactions between fundamental fermions realized by their exchanging several *fundamental vector bosons* (photons, gluons, three intermediate bosons). All these processes are governed by conservation laws.

Recall that the electromagnetic interaction carrier—photon—can be emitted or absorbed only by charged particles, and in this process neither the charge, nor the color, nor the flavor of the particle can change. The carriers of strong interaction—gluons—can be emitted or absorbed by quarks or gluons, and in this process the quark may change its color, but its charge and its flavor remain the same.

Weak interactions are realized in terms of emission and absorption of intermediate vector bosons W^+, W^-, and Z^0. As a charged boson is emitted or absorbed by a charged lepton, the latter transforms into the corresponding neutrino (and vice versa, a neutrino transforms into the corresponding charged lepton). The universal character of weak interactions is manifested in that vector bosons are emitted and absorbed not only by leptons, but also by quarks. (Note that in contrast to gluons, intermediate bosons have no color, i.e., take no part in strong interactions.) After emitting or absorbing a charged boson, the upper quark turns into the lower one and vice versa. It is essential that any upper quark (u, c, t) can transform into *any* lower quark (d, s, b).

For example, neutron decay $n \to pe^{-}\widetilde{\nu}_e$ amounts to d-quark decay $d \to ue^{-}\widetilde{\nu}_e$, which is described as the transformation of the d-quark into an u-quark with emission of a W^{+}-boson, which then turns into the pair $e^{-}\widetilde{\nu}_e$.

An important event in particle physics was the discovery in 1974 of intermediate bosons, whose existence and properties had been predicted by the so-called "standard theory of electroweak interaction" of Glashow, Weinberg, and Salam. The discovery was made in specially constructed proton–antiproton colliders (colliding beam accelerators). Upon collision, one quark and one antiquark turned into an intermediate boson, and the remaining two quarks of each particle turned into powerful hadron jets. In time $\sim 10^{-25}$ s, the produced boson either decayed to quark + antiquark or to two leptons (e.g., electron + antineutrino). It was possible to detect the electrons produced in the W^{-}-boson decay (only those processes were chosen in which the momentum conservation law was violated by the value of the momentum carried away by neutrino, which is equal to the electron momentum). To give an idea of the complexity of the events, it can be mentioned that in the first series of experiments, six reliable events were chosen out of one billion.

P8.4.2. Characteristics of Particles

► **Particle characteristics and conservation laws.** All particle characteristics can be divided into two groups: geometrical characteristics connected with space-time properties (mass, spin, spatial parity) and internal quantum numbers reflecting the symmetry of fundamental interactions (charge, baryon charge, lepton charge, isospin, strangeness, charm, and some others).

Each of these characteristics is associated with some conservation law and it makes sense to discuss them jointly. Some conservation laws are regarded as absolutely precise, others are approximate, i.e., valid for a certain class of reactions. Some approximate conservation laws are violated only by weak interactions (in processes determined by weak interactions), others are violated by both weak and electromagnetic interactions. Conservation laws allow us to predict the possibility or the impossibility of decay processes and reactions, understand the corresponding time lengths, predict the existence of new particles and describe the conditions of their birth. Recall the principle adopted in quantum theory, "what is not forbidden is possible," which makes the application of conservation laws especially effective.

► **Particle mass and law of conservation of energy-momentum.** Particle mass m is its most important characteristic. All subatomic particles have different masses. The Einstein formula relates the mass of a particle to its energy at rest.

The laws of conservation of mass and linear momentum follow from the homogeneity of time and space. These laws are crucial in the examination of reactions. The law of conservation of energy implies, in particular, that decay of an unstable particle can proceed only along such channels in which the sum of the masses of the produced particles is less than the mass of the original particle. Another example: the momentum conservation law forbids annihilation of an electron–positron pair with the emission of one γ-quantum (in center-of-mass frame the pair momentum is zero). The laws of conservation of energy and momentum can be used to calculate the threshold of any reaction (not forbidden by other conservation laws). Such calculations were carried out in Subsection P8.2.2; see formulas (P8.2.2.13)–(P8.2.2.15).

► **Particle spin *I* and angular momentum conservation law.** The spin of a particle determines whether the particle is subject to the Fermi–Dirac statistics (half-integer spin) or the Bose–Einstein statistics (integer spin). The spin of a hadron adds up from spins and orbital angular momenta of its constituent quarks. The spin of a particle determines the unique selected direction in a reference frame in which it is at rest.

The law of conservation of angular momentum is a consequence of the isotropy of space. The application of this law to the analysis of reactions is hindered by the fact that spins are added up by the rule of addition of angular momenta. For example, the total angular momentum of a system of two particles with spin 1 may be zero. However, it is easy to find out whether the total spin is integer or half-integer. This result often suffices for establishing the impossibility of some reaction. Thus, neutrinoless decay of a neutron into two particles—proton and electron—is forbidden by the angular momentum conservation law.

► **Parity.** Parity P of a particle (more precisely, its *spatial parity*) is a purely quantum notion. It indicates the behavior of the particle wave function under *spatial inversion*, i.e., the transformation of coordinates $x \to -x$, $y \to -y$, and $z \to -z$. There are two alternatives: the wave function does not change (positive parity, $P = +1$); the wave function changes its sign (negative parity, $P = -1$). Parity is an internal characteristic of a particle (just as spin, charge, etc.) Electron, proton, and neutron have parity $P = 1$, while photon and π-mesons have parity $P = -1$. Parity of any quark is equal to +1, and that of antiquark is –1. Parity of a compound particle (in the state with zero orbital angular momentum) is equal to the product of parities of its constituent particles.

For a long time, it was assumed that spatial parity is preserved in all reactions. The point is that the parity conservation law is a consequence of the statement that claims the invariance of processes with respect to spatial inversion or, equivalently, with respect to mirror reflection, since inversion reduces to a reflection and a rotation (for instance, reflection with respect to the plane XY and rotation by the angle 180° with respect to the axis Z). It was considered evident that all processes in the common and the "mirror" world should obey the same physical laws. However, later it was found that this statement is incorrect for processes determined by weak interactions. Violation of parity was predicted by Tsung-Dao Lee and Chen-Ning Yang in 1956, and shortly confirmed experimentally by Chien-Shiung Wu. As shown by these experiments, the electrons emitted in β-decay mostly move in the direction opposite to that of nuclear spin. Since nuclear spin changes its direction under reflection, electrons are emitted mostly in the direction of spin in the mirror world.

In the same experiment, one observed violation of another symmetry, the so-called charge symmetry or C-symmetry, when all particles are replaced by the corresponding antiparticles. It turned out that in the "antiworld," positrons would be emitted mostly in the direction of spin. However, under both transformations (i.e., spatial reflection together with the replacement of particles by their antiparticles), the process again transforms into itself. For some time it was assumed that this "combined" CP-symmetry is absolutely precise. But later weak processes were discovered for which the CP-symmetry is violated. At present, only one symmetry is accepted as absolutely precise—the CPT-symmetry, where the above two transformations are supplemented with time reflection. This symmetry is sufficient for proving that a particle and its antiparticle have equal masses and lifetimes.

► **Lepton charges L_e, L_μ, and L_τ.** Experiments show that in any processes the difference between the number of leptons and that of antileptons of a given type is preserved. Accordingly, three types of lepton charges have been introduced: L_e, L_μ, and L_τ. For e and ν_e, it is assumed that $L_e = 1$, for their antiparticles, $L_e = -1$, and for all other particles, $L_e = 0$. In the same way, the other charges are treated.

Lepton charges are preserved in all processes (for energies accessible at present).

The remaining conservation laws pertain to the world of hadrons. These laws are introduced empirically and can be naturally accounted for in the framework of the quark model.

► **Baryon charge B.** As shown by experiments, the difference between the number of baryons and that of antibaryons is preserved in all processes (for energies accessible at present). Accordingly, each baryon is ascribed baryon charge $B = 1$; for any antibaryon, $B = -1$, and for any meson $B = 0$.

If all quarks are ascribed charge $B = 1/3$ (and one takes $B = -1/3$ for antiquarks), then the baryon charge conservation law reduces to the preservation of the number of quarks (more precisely, the difference between the number of quarks and that of antiquarks).

► **Strangeness.** When first hyperons were discovered, some strange properties of their behavior were noticed. First, it was found that although hyperon emission occurred in high energy nucleon collisions (i.e., with strong interactions involved), their decay time was $\sim 10^{-10}$ s, which corresponds to weak interactions. The second strange property was that a hyperon could never be born alone. Thus, in proton collision, a Λ^0-hyperon is born only together with a K^+-meson or Σ^+-hyperon, but never with K^--meson or Σ^--hyperon.

In order to explain these features, particles were ascribed a new quantum number, *strangeness* S, which is preserved in strong and electromagnetic interactions and may change by ±1 in weak interactions. To account for experimental data, the particles Λ^0, Σ^0, Σ^- and K^- were ascribed strangeness $S = -1$, Ξ-hyperons strangeness $S = -2$, and Ω^--hyperon strangeness $S = -3$.

In the quark model of hadrons, the adopted carrier of strangeness is the flavor quark s (for this one takes $S = -1$, and for all other quarks, $S = 0$). The slow rate of decay processes connected with variation of strangeness is explained by that quark flavor change $s \to u$ may occur only in weak interactions with emission of a W^--boson or absorption of a W^+-boson.

► **Charm C, beauty b, and truth t.** In exactly the same way, one introduces the quantum numbers of *charm* C (the c-quark is ascribed $C = 1$) and *beauty* b (the b-quark is ascribed $b = 1$). The last and the heaviest quark, the t-quark, corresponds to the quantum number t, *truth*. Just as in the case of strangeness, the quantum numbers C, b, and t are preserved in strong reactions, but may change in weak reactions.

► **Isotopic spin T.** Isotopic spin was introduced in nuclear physics and played an important role in the classification of the ground and the excited nuclear states. The introduction of this unusual quantum number, which has no analogue in the classical physics, reflects the assumption of *charge independence* of nuclear forces, i.e., the identical coincidence of these forces in the pairs n-n, n-p, p-p in the same spatial and spin states. To describe charge independence, one introduces the *isotopic spin vector* $\mathbf{T}$: its magnitude for both nucleons is equal to $T = 1/2$, and its projections are $T_z = 1/2$ for proton and $T_z = -1/2$ for neutron. Thus, the two nucleons form an *isotopic doublet*. Nuclear interaction is assumed independent of the "direction" of the vector $\mathbf{T}$, i.e., in the absence of electromagnetic interaction, the isotopic spin space is isotropic (properties of the system do not change under rotations of the vector $\mathbf{T}$).

Summation of isotopic spins in a system of several nucleons is carried out by the same rules as that of angular momenta. Nuclei with the same number of nucleons (isobars) and the same values of T form a *nuclear multiplet* of $2T + 1$ nuclei (with different T_z, i.e., different charges), whose members have close properties. For example, the nuclei ${}^{10}_{6}\mathrm{C}$, ${}^{10}_{5}\mathrm{B}$ (in excited state), and ${}^{10}_{4}\mathrm{Be}$ form an isotopic triplet with $T = 1$, and ${}^{4}_{2}\mathrm{He}$ is an isotopic singlet ($T = 0$). A nucleus in its ground state has the minimum possible value of T for a given projection T_z (calculated from its composition); for example, the ground state of ${}^{10}_{5}\mathrm{B}$ has $T = 0$.

Isotopic spin plays an important role in the world of subatomic particles, in particular, for the classification of hadrons. Baryons and mesons are split into multiplets with close properties, and the number of particles in a multiplet is $2T + 1$. Thus, three π-mesons form a triplet with $T = 1$; K^0- and K^+-mesons form a doublet with $T = 1/2$; three Σ-hyperons form a triplet with $T = 1$; four Δ-hyperons have $T = 3/2$; the Λ^0-hyperon constitutes a singlet with $T = 0$. The members of a multiplet differ by their isotopic spin projections T_z.

In the quark model, the presence of multiplets is explained by the fact that the u- and d-quarks have close masses and differ only by their charges. These quarks are ascribed the value $T = 1/2$; for the u-quark, one takes $T_z = 1/2$, and for the d-quark $T_z = -1/2$. For the other quarks, one takes $T = 0$. Isotopic spin of hadrons is calculated by the summation rule for angular momenta. Thus, the doublet of K-mesons has quark composition $u\widetilde{s}$ and $d\widetilde{s}$. Summation of two isotopic spins $1/2$ can result in either $T = 0$ ($T_z = 0$) or $T = 1$ ($T_z = -1$, $T_z = 0$, or $T_z = 1$). The first state corresponds to a hyperon, $\Lambda^0 = uds$ (isotopic singlet), and an isotopic triplet is formed by the hyperons $\Sigma^- = dds$, $\Sigma^0 = uds$, and $\Sigma^+ = uus$.

In reactions determined by strong interactions, one observes the *isotopic spin conservation law*: the quantities T and T_z are preserved. Electromagnetic and weak interactions violate the conservation law for T, but T_z is preserved in electromagnetic interactions and may change only in weak interactions. The point is that T_z can change only if a u-or a d-quark transform into another quark, which is possible only in weak interactions (when a charged W-boson is absorbed or emitted).

P8.4.3. Interaction Unification

► **Electroweak interaction.** In the *standard theory of electroweak interaction*, both interactions, the weak and the electromagnetic, together with their vector bosons, appear in the framework of a unified approach. As the original particles (also called *seed particles*) supposed to realize the unified electroweak interaction one adopts four vector massless particles, two neutral particles, and two charged ones. In the real world, the symmetry between these particles becomes spontaneously broken and three of them acquire mass (intermediate bosons), and one particle remains massless (photon). Many conclusions of the theory (in particular, mass values obtained for intermediate bosons) have had experimental support. The value of the constant of weak interaction happens to be greater than that of electromagnetic interaction, and the slow rate of weak processes (compared with electromagnetic processes) is due to the large mass of W- and Z-bosons ($\sim$ 100 GeV). For energies much larger than these masses, the symmetry is restored (masses can be neglected), and the unified symmetric electroweak interaction becomes active.

► **Grand unification.** The brilliant success of the theory of electroweak interactions opened the way to the creation of new theories that allow for the unification of the three fundamental interactions—strong, weak, and electromagnetic. These theories were called the “Grand Unification.” The boundary energy for this unification is unimaginable, $E \sim 10^{15}$ GeV. At this energy, the constants of the three interactions become equal. For larger energies, the three interactions come forth together as a universal symmetric interaction. The carriers of this interaction are 24 vector bosons: eight gluons, three intermediate bosons, photon, and superheavy X- and Y-bosons with masses $\sim 10^{15}$ GeV. These bosons carry both an electric charge (4/3 for X and 1/3 for Y) and a color charge (taking into account three colors and antiparticles, we have 12 particles). Upon emission or absorption of an X- or a Y-boson, a quark may turn into a lepton and vice versa. Therefore, participation of these bosons in interactions leads to a violation of the laws of conservation of baryon and lepton charges (the electric charge conservation law remains unquestionable). As a result, there is a very small probability (because of the great masses of X and Y) of proton decay into a positron and several π-mesons. The proton lifetime predicted by the theory lies in the interval $\tau \sim 10^{30}$–10^{35} years. Current attempts to detect proton decay (in great masses of water placed in deep mines for the exclusion of background noise) give the estimate $\tau > 10^{32}$ years.

Bibliography for Chapter P8

Alenitsyn, A. G., Butikov, E. I., and Kondratyev, A. S., *Concise Handbook of Mathematics and Physics*, CRC Press, Boca Raton, Florida, 1998.

Benenson, W., Harris, J. W., Stocker, and Lutz, H. (Eds.), *Handbook of Physics*, Springer-Verlag, New York, 2002.

Bertulani, C. A., *Nuclear Physics in a Nutshell*, Princeton University Press, Princeton, New York, 2007.

Bohr, N., Neutron capture and nuclear constitution, *Nature*, Vol. 137, 1936, pp. 344–348.

Butler, S. T. and Hittmair, O. H., *Nuclear Stripping Reactions*, John Wiley and Sons Inc., New York, 1959.

Cottingham, W. N. and Greenwood, D. A., *An Introduction to Nuclear Physics*, Cambridge University Press, Cambridge, England, 2001.

Goeppert-Mayer, M., On closed shells in nuclei, *Physical Review*, Vol. 74, 1948, pp. 235–239.

Goeppert-Mayer, M., On closed shells in nuclei. II, *Physical Review*, Vol. 75, 1949, pp. 1969–1970.

Goeppert-Mayer, M. and Jensen, J. H. D., *Elementary Theory of Nuclear Shell Structure*, John Wiley and Sons, New York, 1955.

Griffiths, D., *Introduction to Elementary Particles, 2nd Edition*, Wiley-VCH Verlag GmbH & Co. KgaA, Weinheim, 2008.

Haxel, O., Jensen, J. H. D., and Suess, H. E., On the “magic numbers” in nuclear structure, *Physical Review*, Vol. 75, 1949, pp. 1766–1766.

Krane, K. S., *Introductory Nuclear Physics*, John Wiley and Sons, New York, 1987.

Lilley, J. S., *Nuclear Physics: Principles and Applications*, Wiley, New York, 2001.

Martin, B. R., *Nuclear and Particle Physics: An Introduction*, John Wiley and Sons, Chichester, England, 2006.

Meier, C. A. (Ed.), *Atom and Archetype: The Pauli/Jung Letters, 1932–1958*, Princeton University Press, Princeton, New Jersey, 2001.

Mukhin, K. N., *Experimental Nuclear Physics, Volumes 1-3, 7th Edition* [in Russian], Lan, Moscow, 2009.

Okun, L. B., *Elementary Introduction to Physics of Elementary Particles* [in Russian], Fizmatlit, Moscow, 2009.

Okun, L. B., *Particle Physics: The Quest for the Substance of Substance (Contemporary Concepts in Physics Series)*, OPA Ltd., Amsterdam, for Harwood Academic Publishers GmbH, 1985.

Okun, L. B., *Physics of Elementary Particles* [in Russian], LKI, Moscow, 2008.

Perkins, D. H., *Introduction to High Energy Physics, 4th edition*, Cambridge University Press, Cambridge, England, 2000.

Pound, R. V. and Rebka, Jr. G. A., Apparent weight of photons, *Physical Review Letters*, Vol. 4, No. 7, 1960, pp. 337–341.

Pound, R. V. and Rebka, Jr. G. A., Gravitational red-shift in nuclear resonance, *Physical Review Letters*, Vol. 3, No. 9, 1959, pp. 439–441.

Prokhorov, A. M. (Ed.), *Encyclopedia of Physics, Volumes 1-5* [in Russian], Bolshaya Russkaya Encyclopedia, Moscow, 1998.

Prokhorov, A. M. (Ed.), *Encyclopedic Dictionary of Physics* [in Russian], Sovetskaya Encyclopedia, Moscow, 1984 (dic.academic.ru).

Sivukhin, D. V., *General Physics, Volume 5, Atomic and Nuclear Physics* [in Russian], Fizmatlit, Moscow,2006.

Sivukhin, D. V. and Yakovlev, I. A., *A Collection of Problems, Volume 5, Atomic Physics. Nuclear and Particle Physics* [in Russian], Fizmatlit, Moscow, 2006.

Williams, W. S. C., *Nuclear and Particle Physics*, Oxford University Press, New York, 1991.

Woan, G., *The Cambridge Handbook of Physics Formulas*, Cambridge University Press, Cambridge, England, 2003.

Part III

Elements of Applied and Engineering Sciences

Chapter E1

Dimensions and Similarity

E1.1. Dimensions

E1.1.1. Basic and Derived Units of Measurement. Systems of Units of Measurement

► **Basic and derived units of measurement.** Mechanical, physical, and physicochemical quantities are characterized by numbers that are determined by direct or indirect comparison with the accepted *units of measurement*. The units of measurement are divided into the *basic* and *derived* units of measurement. The basic units of measurement are defined as artificial and/or natural standards that are chosen from convenience considerations (i.e., rather arbitrarily). The derived units of measurement can be obtained from the basic units in the following two ways:

(i) starting from the definition of the corresponding physical quantity, which prompts the method for its (direct or mental) measurement, or

(ii) by a direct use of formulas relating different quantities by some quantitative relations.

Example 1. The velocity (in uniform motion), by definition, is the ratio of the distance passed in a certain time interval to the value of this time interval. Therefore, in this case, for the unit of velocity it is convenient to take the ratio of the length unit to the time unit.

► **Systems of units of measurement and classes of systems of units of measurement.** A *system of units of measurement* is a set of basic units of measurement that are sufficient to measure numerical characteristics of the class of phenomena under study.

To describe mechanical quantities, it suffices to introduce three independent basic units of measurement, namely, units of mass, length, and time.

Example 2. Nowadays, the most widely used system is the system of units of measurement named the International System of Units (SI), where the unit of mass is 1 kilogram (kg), i.e., the mass of some specially manufactured and thoroughly preserved standard, the unit of length is 1 meter (m), i.e., the length of another standard, and the unit of time is 1 second (s), i.e., the 1/86400 fraction of the mean solar day. (There is also a more precise definition of the second.)

A *class of systems of units of measurement* is a set of systems of units of measurement that differ from one another only in the value of the basic units of measurement.

Example 3. The SI system mentioned in Example 2 belongs to the class of systems of units of measurement denoted by MLT (mass — length — time). The MLT class contains the sometimes used CGS system, where the unit of mass is 1 gram (g) = 0.001 kg (one thousandth of the kilogram), the unit of length is 1 centimeter (cm) = 0.01 m (one hundredth of the metre), and the unit of time is 1 s.

E1.1.2. Dimensional and Dimensionless Quantities. Dimensional Formula

► **Dimensional and dimensionless quantities.** In any practical calculations and in the processing of experimental data whose results are represented as sets of numbers characterizing the properties of phenomena and processes, two types of quantities are encountered.

The quantities whose numerical value depends on the system of units of measurement used within a given class of systems of units of measurement are called *dimensional*. The quantities whose numerical values are independent of the system of units of measurement used within a given class are called *dimensionless*.

► **Dimensional formula. Dimensions of mechanical quantities.** The *dimension of a physical quantity* (in other words, the *dimensional formula*) is a function determining by how many times the numerical value of this quantity varies when passing from the original system of units of measurement to another system contained in the given class. In a dimensional formula, the symbol of unit length is used to be denoted by L, of unit mass, by M, and of unit time, by T; the dimension of any derived quantity A is denoted by the symbol $[A]$. The dimension of a dimensionless quantity is equal to 1.

Example. If the length unit is decreased by a factor of L and the time unit is decreased by a factor of T, then the numerical value of any measured acceleration changes by a factor of LT^{-2}. Thus, for the dimension of acceleration in the class of MLT systems, we have

$$[a] = \mathrm{LT}^{-2}.$$

Similarly, the dimension of force in the MLT class is determined by the formula: $[F] = \mathrm{MLT}^{-2}$. (This formula can also be obtained from Newton's second law $ma = F$.)

Table E1.1 presents the dimensions of basic mechanical quantities in the class of MLT systems.

One can see that for all mechanical quantities given in Table E1.1, the dimension is a power monomial of the form

$$[A] = \mathrm{L}^{\alpha}\mathrm{M}^{\beta}\mathrm{T}^{\gamma}, \qquad \text{(E1.1.2.1)}$$

where α, β, and γ are some constants.

One can prove the following more general statement: *the dimension is always a power monomial*. Such a simple structure of a dimensional formula is necessarily determined by the following physical condition: the ratio of two numerical values of any derived quantity must be independent of the choice of scales for the basic units of measurement.

► **Properties of dimensions.**

1. The dimension of a physical quantity does not vary if it is multiplied by a dimensionless quantity.
2. The dimension of the product of physical quantities is equal to the product of dimensions of these quantities:
$$[A_1A_2\dots A_m] = [A_1][A_2]\dots[A_m].$$
3. The exponent can be brought outside the symbol of dimension:
$$[A^p] = [A]^p.$$
4. The physical quantities can be added (subtracted) only if they have equal dimensions.
5. The exponents contained in the right-hand sides of dimension formulas like (E1.1.2.1) are rational numbers (i.e., they can be represented as the fractions p/q, where p and q are integers).

E1.2. The π-Theorem and Its Practical Use

E1.2.1. The π-Theorem

► **Passing from dimensional quantities to dimensionless quantities. The π-theorem.** Assume that there is a functional dependence between dimensional physical quantities,

$$f(A_1,\dots,A_k,A_{k+1},\dots,A_n) = 0. \qquad \text{(E1.2.1.1)}$$

TABLE E1.1
Dimensions of mechanical quantities in the class of MLT systems.

Quantity	Dimensional formula	Dimension in the SI system	Dimension in the CGS system
Length	L	m	cm
Area	L^2	m^2	cm^2
Volume	L^3	m^3	cm^3
Curvature	L^{-1}	m^{-1}	cm^{-1}
Time	T	s	s
Velocity	LT^{-1}	m/s	cm/s
Acceleration	LT^{-2}	$\mathrm{m/s}^2$	$\mathrm{cm/s}^2$
Angular velocity, frequency	T^{-1}	s^{-1}	s^{-1}
Angular acceleration	T^{-2}	s^{-2}	s^{-2}
Mass	M	kg	g
Density	ML^{-3}	$\mathrm{kg/m}^3$	$\mathrm{g/cm}^3$
Force	MLT^{-2}	$\mathrm{kg{\cdot}m/s}^2$	$\mathrm{g{\cdot}cm/s}^2$
Moment of force, torque	$\mathrm{ML}^2\mathrm{T}^{-2}$	$\mathrm{kg \cdot m^2/s^2}$	$\mathrm{g \cdot cm^2/s^2}$
Momentum	MLT^{-1}	kg·m/s	g·cm/s
Angular momentum	$\mathrm{ML}^2\mathrm{T}^{-1}$	$\mathrm{kg{\cdot}m^2/s}$	$\mathrm{g{\cdot}cm^2/s}$
Pressure, tension	$\mathrm{ML}^{-1}\mathrm{T}^{-2}$	$\mathrm{kg/(m{\cdot}s^2)}$	$\mathrm{g/(cm{\cdot}s^2)}$
Work, energy	$\mathrm{ML}^2\mathrm{T}^{-2}$	$\mathrm{kg{\cdot}m^2/s^2}$	$\mathrm{g{\cdot}cm^2/s^2}$
Power	$\mathrm{ML}^2\mathrm{T}^{-3}$	$\mathrm{kg{\cdot}m^2/s^3}$	$\mathrm{g{\cdot}cm^2/s^3}$
Modulus of elasticity	$\mathrm{ML}^{-1}\mathrm{T}^{-2}$	$\mathrm{kg/(m{\cdot}s^2)}$	$\mathrm{g/(cm{\cdot}s^2)}$
Kinematic viscosity	$\mathrm{L}^2\mathrm{T}^{-1}$	$\mathrm{m^2/s}$	$\mathrm{cm^2/s}$
Dynamic viscosity	$\mathrm{ML}^{-1}\mathrm{T}^{-1}$	kg/(m·s)	g/(cm·s)
Angle	dimensionless quantity	dimensionless quantity	dimensionless quantity

(In the process under study, some of these quantities can be variable, and the other, constant; moreover, here we do not distinguish between dependent and independent variables.) Let $A_1, \ldots, A_k$ ($k \le n-1$) be dimensional quantities of independent dimensions (these quantities are called *constitutive parameters*), and suppose that the dimensions of the other quantities can be expressed via the dimensions of the constitutive parameters:

$$\begin{gathered}[A_{k+1}] = [A_1]^{m_{1\,k+1}}[A_2]^{m_{2\,k+1}}\dots[A_k]^{m_{k\,k+1}},\\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots,\\ [A_n] = [A_1]^{m_{1\,n}}[A_2]^{m_{2\,n}}\dots[A_k]^{m_{k\,n}},\end{gathered} \qquad \text{(E1.2.1.2)}$$

where the $m_{i\,j}$ are constants. Set

$$\Pi_1 = \frac{A_{k+1}}{A_1^{m_{1\,k+1}}A_2^{m_{2\,k+1}}\dots A_k^{m_{k\,k+1}}}, \quad \dots, \quad \Pi_{n-k} = \frac{A_n}{A_1^{m_{1\,n}}A_2^{m_{2\,n}}\dots A_k^{m_{k\,n}}}. \qquad \text{(E1.2.1.3)}$$

It follows from relations (E1.2.1.2) that the quantities Π_1, Π_2, …, Π_{n-k} are dimensionless. Each of the dimensionless quantities Π_i has the form of a power monomial consisting of $\le k+1$ dimensional quantities. (The inequality takes into account the fact that some of the exponents $m_{i\,j}$ can be zero.)

THE π-THEOREM. *The original relation* (E1.2.1.1) *among n dimensional physical quantities can be represented as a relation among $n-k$ dimensionless quantities* (E1.2.1.3),

$$F(\Pi_1, \Pi_2, \dots, \Pi_{n-k}) = 0. \qquad \text{(E1.2.1.4)}$$

It follows from the above that every physical relation between dimensional quantities can be reformulated as a relation between fewer dimensionless quantities. This important fact is a source of useful applications of the π-theorem for the analysis of various phenomena and processes. In particular, for $k = n-1$, the same n dimensional parameters can form only a single dimensionless combination, and hence, by the π-theorem, the initial functional dependence can be represented in the following simple form:

$$A_1^{m_1}A_2^{m_2}\dots A_n^{m_n} = \text{const.} \qquad \text{(E1.2.1.5)}$$

Here the exponents $m_1, m_2, \cdots, m_n$ can readily be determined by dimensional formulas, where it is taken into account that the left-hand side of (E1.2.1.5) is a dimensionless complex and the unknown dimensionless constant on the right-hand side in (E1.2.1.5) can be found either experimentally or theoretically by solving the corresponding mathematical problem (see examples in Section E1.2.2 below).

Remark 1. Without loss of generality, one can assume that one of the exponents m_i in (E1.2.1.5) is equal to 1 (in particular, if $m_1 \ne 0$, then one can set $m_1 = 1$).

Remark 2. If $\Pi_1, \Pi_2, \dots, \Pi_s$ are dimensionless quantities, then the complex $\Pi = \Pi_1^{\beta_1}\Pi_2^{\beta_2}\dots\Pi_s^{\beta_s}$, where $\beta_1, \beta_2, \dots, \beta_s$ are arbitrary constants, is also a dimensionless quantity. This allows one to compose new $n-k$ functionally independent dimensionless quantities from the dimensionless quantities (E1.2.1.3) and write out a relation of the form (E1.2.1.4) for them. A successful choice of the form of the dimensionless parameters permits reducing the number of experiments and simplifying their analysis.

► **Experimental data must be processed in dimensionless variables.** The π-theorem leads to a very important conclusion that experimental data must be processed in dimensionless variables, thus reducing the number of measured and varied quantities.

In practical experiments, it is usually assumed that establishing the dependence of a quantity on some constitutive parameter (argument) means measuring this quantity for at least 10 values of the given argument (of course, the number 10 is here taken at random). Thus, to experimentally determine the quantity A as a function of n constitutive parameters $A_1, \dots, A_n$, one must perform 10^n experiments. But, using the π-theorem, one can reduce the problem to determining a function of $n-k$ dimensionless arguments Π_1, …, Π_{n-k}; to this end, one must perform 10^{n-k} experiments, i.e., 10^k times fewer. In other words, the labouriousness of determining the desired function is reduced by as many orders of magnitude as the number of quantities of independent dimensions contained in the set of constitutive parameters.

E1.2.2. Simplest Examples of the Use of the π-Theorem

In what follows, all dimensions of constitutive parameters are written in the class of MLT systems of units of measurement.

► **The Pythagorean theorem.** Let us describe an amusing version of proving the Pythagorean theorem by analyzing the dimensions.

The area S of a rectangular triangle is determined by the length c of the hypotenuse and, for example, the smaller of the acute angles, β. Using the three quantities S, c, and β whose dimensions have the form $[S] = \mathrm{L}^2$, $[c] = \mathrm{L}$, and $[\beta] = 1$ (the angle β is a dimensionless quantity), one can compose two dimensionless combinations:

$$\Pi_1 = S/c^2, \qquad \Pi_2 = \beta. \tag{E1.2.2.1}$$

We set $n = 3$ and $k = 1$ in formula (E1.2.1.4) and take the explicit form of the dimensionless complexes (E1.2.2.1) into account. We have $F(S/c^2, \beta) = 0$. Resolving the dependence thus obtained for the first argument, we obtain

$$S = c^2\Phi(\beta),$$

where $\Phi(\beta)$ is a function. (The explicit form of this function is not required in the following.)

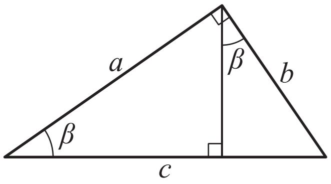

Figure E1.1. A right triangle.

The height perpendicular to the hypotenuse (see Fig. E1.1) divides the main triangle into two right triangles similar to it whose hypotenuses are the corresponding legs a and b of the original triangle. The areas of the new triangles are $S_a = a^2\Phi(\beta)$ and $S_b = b^2\Phi(\beta)$, respectively, where $\Phi(\beta)$ is the same function as above. Since the sum of S_a and S_b is the area of the original triangle, $S = S_a + S_b$, we have $c^2\Phi(\beta) = a^2\Phi(\beta) + b^2\Phi(\beta)$. By canceling $\Phi(\beta)$, we obtain

$$c^2 = a^2 + b^2,$$

as desired.

► **Shock wave front propagation in nuclear explosion.** In nuclear explosion, considerable energy E_0 is released very rapidly (we assume that it is released instantaneously) in a rather small area (which is approximately assumed to be a point). A powerful shock wave propagates from the explosion center, and the pressure behind it initially equals hundreds of thousands of atmospheres. This pressure is much larger than the initial air pressure, whose influence can be neglected at the first stage of the explosion. Thus, after the time interval t, the shock wave front radius r_f depends on E_0, t, and the initial air density ρ_0.

This problem contains four dimensional quantities r_f, E_0, t, and ρ_0, which have the following dimensions:

$$[r_f] = \mathrm{L}, \quad [E_0] = \mathrm{ML}^2\mathrm{T}^{-2}, \quad [t] = \mathrm{T}, \quad [\rho_0] = \mathrm{ML}^{-3}. \tag{E1.2.2.2}$$

In the set of these four quantities, one can choose three quantities with independent dimensions, which corresponds to $k = 3$ and $n = 4$ in the statement of the π-theorem. Therefore, we have $k+1=n$ and can form a single dimensionless complex of the quantities (E1.2.2.2), and the desired functional dependence can be represented by analogy with (E1.2.1.5) as

$$r_f E_0^{m_2} t^{m_3} \rho_0^{m_4} = C = \text{const}, \qquad \text{(E1.2.2.3)}$$

where, without loss of generality, the first exponent m_1 was set equal to 1. The exponents m_2, m_3, and m_4 are determined by the condition that the dimension of the left-hand side of (E1.2.2.3) should be equal to 1 (because the right-hand side is a dimensionless constant). Therefore, we obtain the chain of relations

$$\begin{aligned}[r_f E_0^{m_2} t^{m_3} \rho_0^{m_4}] &= [r_f][E_0]^{m_2}[t]^{m_3}[\rho_0]^{m_4} = L[ML^2T^{-2}]^{m_2}T^{m_3}[ML^{-3}]^{m_4}\\ &= M^{m_2+m_4}L^{2m_2-3m_4+1}T^{-2m_2+m_3} = 1.\end{aligned}$$

For the last relation to be satisfied, it is necessary that all exponents be zero:

$$\begin{aligned} m_2 + m_4 &= 0,\\ 2m_2 - 3m_4 + 1 &= 0,\\ -2m_2 + m_3 &= 0. \end{aligned}$$

The solution of this linear system of algebraic equations has the form

$$m_2 = -\tfrac{1}{5}, \quad m_3 = -\tfrac{2}{5}, \quad m_4 = \tfrac{1}{5}. \qquad \text{(E1.2.2.4)}$$

Substituting the exponents (E1.2.2.4) into (E1.2.2.3) and solving the resulting relation for r_f, we obtain the shock wave propagation law

$$r_f = C\left(\frac{E_0 t^2}{\rho_0}\right)^{1/5}. \qquad \text{(E1.2.2.5)}$$

If the quantities E_0 and ρ_0 are known, then to find the constant C in formula (E1.2.2.5) it suffices to know the shock wave front radius r_f for a single value $t = t_1$ of the time interval. A more detailed study (beyond the framework of the theory of dimensions) shows that $C \approx 1$.

From formula (E1.2.2.5), one can obtain the time dependence of the shock wave front radius:

$$r_f = r_{f1}\bigl(t/t_1\bigr)^{2/5},$$

where $r_{f1} = r_f|_{t=t_1}$.

► **Pendulum oscillations in the field of gravity forces.** We consider a mathematical pendulum, which is a heavy material point of mass m suspended on a weightless unstretchable string of length l; it is assumed that the other end of the string is rigidly fixed. We are interested in the oscillation period T_p of the pendulum that makes an angle φ_0 with the vertical axis at the initial time instant. (The weight velocity is zero at the initial time instant.)

This problem is characterized by the five quantities m, l, T_p, φ_0, and g (g is the acceleration due to gravity), which have the dimensions

$$[m] = \mathrm{M}, \quad [l] = \mathrm{L}, \quad [T_p] = T, \quad [\varphi_0] = 1, \quad [g] = \mathrm{LT}^{-2}; \qquad \text{(E1.2.2.6)}$$

three of these quantities have independent dimensions. Thus, in this case we have $n = 5$ and $k = 3$, and one can compose two dimensionless complexes of these quantities (E1.2.2.6):

$$\Pi_1 = \varphi_0, \quad \Pi_2 = T_p\sqrt{g/l}. \qquad \text{(E1.2.2.7)}$$

Therefore, by the π-theorem, we have the dependence $F(\varphi_0, T_p\sqrt{g/l}\,) = 0$. By solving it for the oscillation period T_p, we obtain

$$T_p = f_1(\varphi_0)\sqrt{l/g}\,. \tag{E1.2.2.8}$$

where $f_1(\varphi_0)$ is a function.

It is important to note that the dimensionless parameters (E1.2.2.7) are independent of the load mass m. Therefore, the oscillation period (E1.2.2.8) is independent of the pendulum mass.

The function $f_1(\varphi)$ can be determined both theoretically and experimentally.

Assuming that the function $f_1(\varphi)$ can be expanded in a Taylor series for small φ, we have

$$f_1(\varphi_0) = a + b\varphi_0 + c\varphi_0^2 + \cdots.$$

Assume that $a \neq 0$. Then, for small pendulum oscillations, by retaining only the leading term of the expansion in (E1.2.2.8), we obtain the simple formula

$$T_p = a\sqrt{l/g}, \tag{E1.2.2.9}$$

where a is a constant.

Thus, for small pendulum oscillations, we can use the theory of dimensions to obtain formula (E1.2.2.9) for the oscillation period up to a constant factor. A theoretical analysis (independent of the theory of dimensions) permits finding the unknown constant $a = 2\pi$.

► **Interphase transport.** When describing the interphase transport in gas–liquid systems, the Higbie–Danckwerts surface renewal model has been widely used. According to this model, turbulent pulsations with frequency $\omega = 1/t_\omega$ in the liquid phase constantly renew the interphase surface. The contact time t_ω is determined experimentally. The diffusion flow J through an element of the renewed surface is assumed to be purely molecular. It depends on the molecular diffusion coefficient D of the gas dissolved in the liquid. The mass transfer coefficient is defined as follows:

$$k = \frac{J}{S(C_\mathrm{s} - C_\infty)},$$

where S is the interphase surface area, C_∞ is the concentration of the transferred component in the depth of the liquid, and C_s is the concentration on the interphase surface.

Experiments show that the mass transfer coefficient k mainly depends only on the contact time t_ω and the diffusion coefficient D. The dimensions of the three above-listed quantities have the form

$$[k] = \frac{\mathrm{L}}{\mathrm{T}}, \quad [D] = \frac{\mathrm{L}^2}{\mathrm{T}}, \quad [t_\omega] = \mathrm{T}.$$

There relations contain only two independent dimensions L and T. Therefore, by the π-theorem, one can construct only one dimensionless complex of the form (E1.2.1.5) from the quantities k, D, and t_ω, which results in the dependence

$$k = C\sqrt{D/t_\omega},$$

where C is a constant.

According to the theoretical results obtained by Higbie and Danckwerts, $C = 2/\sqrt{\pi}$. In the Danckwerts model, it is proposed to determine the value of t_ω by the spectrum of the time of residence of liquid elements on the interphase surface. The exponential form of this spectrum is accepted starting from the concept of statistical independence of turbulent vortices arising in the depth of the liquid and reaching the interphase surface.

► **Equation of state of an ideal gas.** An ideal gas is understood as a gas whose static state is completely determined by the values of pressure P, density ρ, absolute temperature T_a, and the heat capacity c_v.

The dimensions of these quantities have the form

$$[P]=\frac{\mathrm{M}}{\mathrm{LT}^2},\quad [\rho]=\frac{\mathrm{M}}{\mathrm{L}^3},\quad [T_\text{a}]=\mathrm{K},\quad [c_v]=\frac{\mathrm{L}^2}{\mathrm{T}^2\mathrm{K}}. \tag{E1.2.2.10}$$

We take into account the fact that the dimension K of temperature (in degrees Kelvin) is a derived unit of measurement, which can be written in terms of dimensions of the basic units of measurement as follows: $\mathrm{K}=\mathrm{ML}^2/\mathrm{T}^2$. Therefore, relations (E1.2.2.10) for four quantities contain three independent dimensions. Hence, by the π-theorem, the desired functional dependence for P, ρ, T_a, and c_v can be represented in a simple form like (E1.2.1.5). In the case under study, this leads to the Clapeyron—Mendeleev equation

$$\frac{P}{c_v\rho T_\text{a}}=\text{const},$$

which has the more usual form $P=\rho RT_\text{a}$, where R is the gas constant.

► **Flow of a viscous fluid about a solid particle.** Consider a homogenous flow of a viscous incompressible fluid about a solid spherical particle of diameter d. The particle drag force F caused by viscosity depends on the four parameters d, U_i, ν, and ρ, where U_i is the unperturbed flow velocity far from the particle and ν and ρ are the kinematic viscosity and the fluid density, respectively.

The dimensions of all five above-listed quantities (including the force) have the form

$$[F]=\frac{\mathrm{ML}}{\mathrm{T}^2},\ [d]=\mathrm{L},\ [U_\text{i}]=\frac{\mathrm{L}}{\mathrm{T}},\ [\nu]=\frac{\mathrm{L}^2}{\mathrm{T}},\ [\rho]=\frac{\mathrm{M}}{\mathrm{L}^3}.$$

Out of these quantities, one can choose three with independent dimensions and compose two dimensionless complexes

$$\Pi_1=\frac{F}{d^2\rho U_\text{i}^2},\qquad \Pi_2=\frac{dU_\text{i}}{\nu}, \tag{E1.2.2.11}$$

which, by the π-theorem, satisfy the functional relation $F(\Pi_1,\Pi_2)=0$. By solving this relation for Π_1, we find the general structure of the dependence of the drag force on the other parameters:

$$\frac{F}{d^2\rho U_\text{i}^2}=\Phi\Big(\frac{dU_\text{i}}{\nu}\Big). \tag{E1.2.2.12}$$

Thus, dimension analysis permits significantly simplifying experimental studies by decreasing the number of varied variables in the problem from five to two. The complex $\mathsf{Re}=dU_\text{i}/\nu$ on the right-hand side of the expression (E1.2.2.12) is called the *Reynolds number*. (One often takes the particle radius, $a=d/2$, to be the characteristic length when determining the Reynolds number.)

Experiments show that $\Phi\to A$ as $\mathsf{Re}\to\infty$, where A is a constant. This situation in which the drag force is independent of viscosity and quadratically depends on velocity is called the Newton flow mode.

As $\mathsf{Re}\to 0$, we have $\Phi\to B/\mathsf{Re}$. (This result can be obtained theoretically.) In this case, the drag force linearly depends on both the viscosity and the velocity. Such a flow is called a Stokes (or creeping) flow.

Remark. In the case of a homogeneous translational flow of a viscous incompressible fluid about a body of nonspherical shape with typical linear dimension d (for example, for d one can take the diameter of a spherical body with equivalent volume), the dependence (E1.2.2.12) also holds, where the function $\Phi(z)$ depends on the shape of the body.

► **Stirring a liquid.** The power N required to stir a liquid with density ρ and kinematic viscosity ν depends on the mixer diameter d, the angular velocity of stirring ω, and the acceleration due to gravity g. The problem complexity does not allow one to calculate the power by solving the hydrodynamic equations analytically. Therefore, we use dimensional analysis.

We write out the dimensions of the main quantities:

$$[N] = \frac{\mathrm{ML}^2}{\mathrm{T}^3}, \quad [\rho] = \frac{\mathrm{M}}{\mathrm{L}^3}, \quad [\nu] = \frac{\mathrm{L}^2}{\mathrm{T}}, \quad [d] = \mathrm{L}, \quad [\omega] = \frac{1}{\mathrm{T}}, \quad [g] = \frac{\mathrm{L}}{\mathrm{T}^2}.$$

We use the π-theorem to determine the number of dimensionless complexes of the problem:

$$3 = 6(\textit{the number of parameters of the problem}) - 3(\textit{the number of primary dimensions}).$$

Under the assumption that the quantities d, ρ, and ω can enter any dimensionless combination, we have

$$\frac{N}{\rho d^5 \omega^3} = \Phi\Big(\frac{d^2\omega}{\nu}, \frac{d\omega^2}{g}\Big).$$

The structure of the obtained functional dependence allows us to simplify the experimental study significantly by decreasing the number of varied variables in the problem from six to three. Experiments show that for $d^2\omega/\nu \le 10^{-2}$, when the influence of the viscosity is large, the mixer power consumption ceases to depend on the dimensionless complex $d^2\omega/\nu$.

► **Turbulent flow in a tube.** Consider a steady turbulent flow in a circular smooth tube of radius a. Assume that r is the radial coordinate counted off from the tube axis; τ_w is the friction stress* on the tube wall; $\overline{V}$ is the average component of the fluid velocity; $\overline{V}_{\max}$ is the maximum velocity on the tube axis; and ν and ρ are the liquid kinematic viscosity and density, respectively.

It is easily seen that the average velocities $\overline{V}$ and $\overline{V}_{\max}$ of the fluid motion are determined by the following set of parameters:

$$a, \quad \rho, \quad \nu, \quad \tau_w, \quad r = a - y.$$

All the dimensionless quantities depend on the two dimensionless complexes

$$\mathrm{Re}_* = \frac{aU_*}{\nu}, \qquad \frac{r}{a} = 1 - \frac{y}{a},$$

where $U_* = \sqrt{\tau_w/\rho}$ is the velocity of the tangential friction stress on the tube wall.

It follows from the similarity theory that

$$\frac{\overline{V}_{\max} - \overline{V}}{U_*} = F\Big(\mathrm{Re}_*, \frac{a-y}{a}\Big). \qquad \text{(E1.2.2.13)}$$

The difference $\overline{V}_{\max} - \overline{V}$ is called the velocity defect. It characterizes the fluid velocity distribution over the tube cross-section with respect to the motion on the axis.

---

* The dependence $\tau_w = \frac{a}{2}\frac{\Delta P}{L}$ holds, where $\Delta P > 0$ is the pressure difference on the tube length L.

For very large Reynolds numbers Re_*, formula (E1.2.2.13) implies that

$$\frac{\overline{V}_{\max} - \overline{V}}{U_*} = F_1\biggl(\frac{a-y}{a}\biggr), \tag{E1.2.2.14}$$

where $F_1(z) = F(\infty, z)$.

The expression (E1.2.2.14) is called the universal law of velocity distribution for the velocity defect. Experiments show that the law (E1.2.2.14) holds in the entire flow region inside smooth and rough tubes except for a narrow near-wall region. In this case, the function $F_1(z)$ on the right-hand side in formula (E1.2.2.14) is well approximated by the Prandtl expression $F_1(z) = 2{,}5 \ln z$.

E1.3. Similarity and Simulation. General Remarks

E1.3.1. Similarity and Simulation

► **Physical similarity and simulation.** Prior to manufacturing a large and expensive structure (such as ship, aircraft, cooler, chemical reactor, etc.), one often turns to simulation, i.e., model testing, so as to obtain the best characteristics of the structure under its expected operation conditions. In simulation, it is necessary to know how to recalculate the results obtained in experiments with models to the case of actual conditions: if this is unknown, then simulation is useless. This also applies to complicated phenomena and processes. The theory of simulation is based on the notion of similar phenomena.

Consider a class of qualitatively similar physical phenomena or processes that differ only in the numerical values of the constitutive parameters that are used to compose s dimensionless complexes $\Pi_1, \Pi_2, \dots, \Pi_s$. (Here $s = n - k$; see the π-theorem.) Two phenomena or processes from a given class are said to be physically *similar* if they are described by the same dimensionless complexes:

$$\Pi_1' = \Pi_1'', \quad \Pi_2' = \Pi_2'', \quad \dots, \quad \Pi_s' = \Pi_s'', \tag{E1.3.1.1}$$

where the single prime refers to the first phenomenon (called the model phenomenon), and the double prime refers to the second phenomenon (called the full-scale phenomenon). In this definition, the dimensionless complexes $\Pi_1, \Pi_2, \dots, \Pi_s$ are called the *similarity parameters*.

Physical similarity is a generalization of geometric similarity concerned with bodies (objects) whose characteristic dimensions are proportional.

► **Rules for recalculating the measurement results from the model to the case of actual conditions.** If the similarity parameters are determined by formulas (E1.2.1.3), then relations (E1.3.1.1) become

$$\frac{A_q'}{(A_1')^{m_{1q}}(A_2')^{m_{2q}}\dots(A_k')^{m_{kq}}} = \frac{A_q''}{(A_1'')^{m_{1q}}(A_2'')^{m_{2q}}\dots(A_k'')^{m_{kq}}}, \qquad q = k+1, \dots, n \tag{E1.3.1.2}$$

and determine simple rules for recalculating the measurement results from the model to the case of actual conditions (under which it is difficult to perform direct measurements).

Conditions (E1.3.1.2) show how to choose the model constitutive parameters $A_{k+1}', \dots, A_n'$ to ensure the similarity of the model to the case of actual conditions:

$$A_q' = A_q''(A_1'/A_1'')^{m_{1q}}\dots(A_k'/A_k'')^{m_{kq}}, \qquad q = k+1, \dots, n; \tag{E1.3.1.3}$$

the model parameters $A_1', \dots, A_k'$ can be chosen arbitrarily. The quantities $A_1'', \dots, A_n''$ in (E1.3.1.3) are introduced based on required technical (or actually existing) parameters of the process under study.

Example. Suppose a streamlined ship moves on the surface of a liquid. The main contribution to the drag of the rapidly moving ship is due to the surface waves created by it. We consider an ideal situation in which the contribution of the streamlined ship viscous drag can be neglected as the first approximation. The ship drag force F is determined by its characteristic length l, its velocity U, the liquid density ρ, and the acceleration due to gravity g. (The last parameter is essential, because the force of gravity has a determining effect on waves.) The quantities under study have the following dimensions:

$$[F] = \mathrm{MLT}^{-2}, \quad [l] = \mathrm{L}, \quad [U] = \mathrm{LT}^{-1}, \quad [\rho] = \mathrm{ML}^{-3}, \quad [g] = \mathrm{LT}^{-2}.$$

Here $n = 5$ and $k = 3$, and therefore, one can compose two dimensionless complexes

$$\Pi_1 = \frac{U}{\sqrt{lg}}, \quad \Pi_2 = \frac{F}{\rho(lU)^2}. \tag{E1.3.1.4}$$

The dynamic similarity parameter Π_1 is called the *Froude number*, which is denoted by the special symbol Fr. Thus, for the model and in the case of actual conditions, the quantity U^2/l must be the same (with g remaining unchanged), which results in the following rule for recalculating the model velocities to the case of actual conditions:

$$U' = U''\sqrt{\frac{l'}{l''}}. \tag{E1.3.1.5}$$

Using the second similarity parameter in (E1.3.1.4) and assuming that the liquid is the same for the model and the actual conditions (more precisely, $\rho' = \rho''$), we obtain the following rule for recalculating the drag force:

$$F' = F''\left(\frac{l'U'}{l''U''}\right)^2 = F''\left(\frac{l'}{l''}\right)^3, \tag{E1.3.1.6}$$

i.e., the drag force is proportional to the cubed linear scale of simulation. Formula (E1.3.1.6) was derived using relation (E1.3.1.5).

E1.3.2. General Concluding Remarks

The use of dimensionless parameters simplifies analysis and interpretation of experimental data to a large extent and, in numerous cases, allows successful simulation of phenomena and processes that cannot be studied under laboratory conditions. Combining the theory of dimensions and the similarity theory with the conclusions obtained experimentally or by mathematical analysis of the corresponding equations, one can often obtain rather significant and nontrivial results.

Although undoubtedly useful, the theory of dimensions and similarity is in fact a restricted theory, because it does not permit obtaining the explicit form of the functional relations between dimensionless quantities (it only reveals the structure of these dependencies and the number of dimensionless parameters). The desired functional dependencies describing the phenomenon or process under study can be determined both experimentally and theoretically by using the existing investigation methods. For the problems that are stated mathematically as appropriate equations and boundary conditions, the role of methods of the theory of dimensions and similarity is mostly of auxiliary character. (After the problem has been stated mathematically in dimensional variables, it is often reasonable to reformulate it in terms of dimensionless variables.)

Bibliography for Chapter E1

Barenblatt, G. I., *Dimensional Analysis*, Gordon & Breach, New York, 1989.
Bridgman, P. W., *Dimensional Analysis*, Yale Univ. Press, New Haven, Connecticut, 1931.
Danckwerts, P. V., Gas–Liquid Reactions, McGraw-Hill, New York, 1970.
Sedov, L. I., *Similarity and Dimensional Methods in Mechanics*, CRC Press, Boca Raton, Florida, 1993.
Taylor, G. I., The formation of a blast wave by a very intense explosion, *Proc. Roy. Soc., Ser. A,* 1950, Vol. 201, No. 1065, pp. 159–186.

Chapter E2

Mechanics of Point Particles and Rigid Bodies

E2.1. Kinematics

Kinematics is the field of mechanics where the geometric properties of motion of a point or a rigid body are studied independently of their mass and causes of their motion.

E2.1.1. Kinematics of a Point

To specify the motion of a point M is to introduce a rule determining the point position with respect to a reference frame $Oxyz$ at each time instant. The most routine and convenient methods for specifying the motion of a point are the vector, coordinate, and natural methods.

► **The vector method for specifying the motion of a point.** In this method, the position of point M is specified by the *position vector* $\mathbf{r}(t)$ drawn to this point from a stationary center (pole, point). For the pole one can take any point, for example, the origin O of a coordinate system (Fig. E2.1). The position vector is a vector function of a scalar argument—time t—and in the course of time the endpoint of the vector $\mathbf{r}$ traces a curve in space, which is called the *trajectory* (or *path*) of the point.

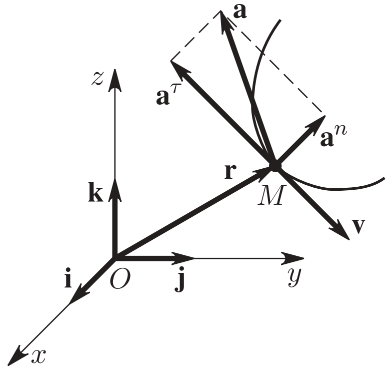

Figure E2.1. Specifying the motion of point M in a Cartesian coordinate system. The trajectory of motion of the point, its velocity, and acceleration.

Along with the position vector, the kinematic state of point M is characterized by the velocity vector $\mathbf{v}(t)$ and the acceleration vector $\mathbf{a}(t)$.

The *velocity* $\mathbf{v} = \dfrac{d\mathbf{r}}{dt} = \dot{\mathbf{r}}$, which is the derivative of the position vector with respect to time,* characterizes the rate and direction of variation in the position of the point in space. The velocity vector is directed along the tangent to the trajectory at point M.

---

* In this chapter, the time derivatives are usually denoted by a dot (which is traditional notation in classical mechanics).

The *acceleration* $\mathbf{a} = \dot{\mathbf{v}} = \ddot{\mathbf{r}}$ is the first time derivative of the velocity vector, or the second derivative of the position vector. It is convenient to represent the acceleration as the sum of two vectors $\mathbf{a}^\tau$ and $\mathbf{a}^n$ lying in the plane spanned by the vectors $\mathbf{v}$ and $\mathbf{a}$ (see Fig. E2.1):

$$\mathbf{a} = \mathbf{a}^\tau + \mathbf{a}^n.$$

The vector $\mathbf{a}^\tau$ is parallel or antiparallel to $\mathbf{v}$ and is called the *tangential* acceleration, and the vector $\mathbf{a}^n$ is perpendicular to $\mathbf{a}^\tau$, points to the center of curvature of the trajectory, and is called the *normal* acceleration.

The projection $\operatorname{comp}_{\mathbf{v}} \mathbf{a}$ of the vector $\mathbf{a}$ on the direction of the vector $\mathbf{v}$, the tangential acceleration magnitude a^τ, and the normal acceleration magnitude a^n can be found by the formulas

$$\operatorname{comp}_{\mathbf{v}} \mathbf{a} = \frac{\mathbf{a} \cdot \mathbf{v}}{v} = \dot{v}, \quad a^\tau = |\dot{v}|, \quad a^n = \sqrt{a^2 - (a^\tau)^2},$$

where $\mathbf{a} \cdot \mathbf{v}$ is the scalar product of the vectors $\mathbf{a}$ and $\mathbf{v}$.

The sign of $\operatorname{comp}_{\mathbf{v}} \mathbf{a}$ determines the character of motion of the point at the time instant under study: if this quantity is positive, then the velocity magnitude v increases, and the motion is said to be *accelerated*; otherwise, the velocity magnitude decreases, and the motion is said to be *decelerated*.

► **The coordinate method for specifying the motion of a point.** The position of point M in space is specified by three functions of time,

$$x = x(t), \qquad y = y(t), \qquad z = z(t), \tag{E2.1.1.1}$$

where x, y, z are the Cartesian coordinates of M. This method is another representation of the vector method, because the position vector $\mathbf{r}$ issuing from the origin can be represented as the sum

$$\mathbf{r} = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k},$$

where $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ are the unit vectors (basis vectors) directed along the coordinate x-, y-, and z-axes (Fig. E2.1).

The projections of the velocity on the coordinate v_x-, v_y-, and v_z-axes can be found by differentiation with respect to time of the corresponding functions (E2.1.1.1):

$$v_x = \dot{x}, \qquad v_y = \dot{y}, \qquad v_z = \dot{z}. \tag{E2.1.1.2}$$

The velocity vector can be found as the vector sum of three vectors,

$$\mathbf{v} = v_x\mathbf{i} + v_y\mathbf{j} + v_z\mathbf{k},$$

and the velocity magnitude is given by the formula

$$v = \sqrt{v_x^2 + v_y^2 + v_z^2}. \tag{E2.1.1.3}$$

The projections a_x, a_y, and a_z of the acceleration onto the coordinate axes, the acceleration $\mathbf{a}$, and its magnitude a can be found by the formulas

$$\begin{gathered} a_x = \dot{v}_x = \ddot{x}, \quad a_y = \dot{v}_y = \ddot{y}, \quad a_z = \dot{v}_z = \ddot{z}, \\ \mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}, \quad a = \sqrt{a_x^2 + a_y^2 + a_z^2}. \end{gathered} \tag{E2.1.1.4}$$

The tangential acceleration is calculated by the formulas

$$\operatorname{comp}_{\mathbf{v}} \mathbf{a} = \dot{v} = \frac{\dot{x}\ddot{x} + \dot{y}\ddot{y} + \dot{z}\ddot{z}}{\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}}; \qquad a^{\tau} = |\dot{v}|. \tag{E2.1.1.5}$$

► **The natural method for specifying the motion of a point.** In this case, the position of point M in space is specified by specifying

(a) The trajectory of the point.

(b) The origin (point O on the trajectory from which the angular position s is counted).

(c) The positive sense of the arc coordinate.

(d) The law of motion $s = s(t)$ of the point, i.e., the time dependence of the arc coordinate.

The arc coordinate s is the length of the trajectory arc counted from point O to point M. If the length is counted in the positive sense, then the arc coordinate is positive; otherwise, it is negative.

The natural coordinate axes are constructed at point M. For a given trajectory of the point, the position vector $\mathbf{r}$ connecting the fixed pole with point M becomes a function of the arc coordinate, $\mathbf{r} = \mathbf{r}(s)$. Its derivative $\boldsymbol{\tau} = \mathbf{r}_s'$ with respect to the arc coordinate is the unit vector *tangent* to the trajectory at point M in the sense of increasing arc coordinate.

The second derivative $\mathbf{r}_{ss}'' = \boldsymbol{\tau}_s' = \mathbf{n}/\rho$ gives the second unit vector $\mathbf{n}$ of the *principal normal*, which is perpendicular to $\boldsymbol{\tau}$ and points to the center of curvature of the trajectory. The geometric meaning of the factor $1/\rho$ will be explained later.

The last, third unit vector $\mathbf{b}$ of the *binormal* is the cross product of the first two unit vectors, $\mathbf{b} = \boldsymbol{\tau} \times \mathbf{n}$.

The right trihedron of unit vectors thus constructed bears the name of the *natural coordinate axes*. The planes passing through them are called as follows: $(\mathbf{nb})$ is the *normal plane*, $(\mathbf{b}\boldsymbol{\tau})$ is the *rectifying plane*, and $(\boldsymbol{\tau}\mathbf{n})$ is the *osculating plane*.

If on the trajectory near the point M we choose a point M' whose arc coordinate $s + \Delta s$ differs by a small value Δs, them, up to terms of higher order of magnitude, the distances from M' to the above-listed planes are as follows: Δs to the normal plane, $(\Delta s)^2/(2\rho)$ to the rectifying plane, and $\beta(\Delta s)^3$ to the osculating plane, where β is a quantity of the order of unity. This means that the part of the trajectory in a small neighborhood of point M "practically" lies in the osculating plane.

Let us construct a family of circles tangent to the trajectory at point M so that their centers are located on the principal normal $\mathbf{n}$. These circles lie in the osculating plane. The circle that is tangent to the trajectory most closely is called the *osculating circle*. The center of this circle, its radius ρ, and the quantity $k = 1/\rho$ are called the *center of curvature*, the *radius of curvature*, and the *curvature* of the trajectory at point M (Fig. E2.2). On flat parts of the trajectory, the values of ρ are larger and the values of k are less than on the curved parts.

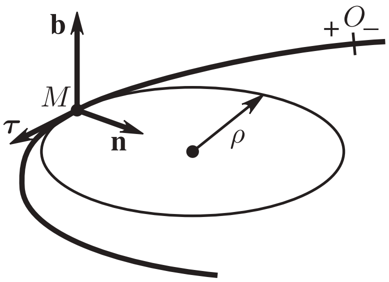

Figure E2.2. Osculating circle, the center of curvature, and the radius of curvature.

Using the natural coordinate axes, one can write the velocity and acceleration vectors of point M as

$$\mathbf{v} = \dot{s}\boldsymbol{\tau}, \qquad \mathbf{a} = \mathbf{a}^{\tau} + \mathbf{a}^{n} = \ddot{s}\boldsymbol{\tau} + (v^2/\rho)\mathbf{n}. \tag{E2.1.1.6}$$

In the last expressions, one must take into account the following:

First, the vector $\mathbf{v}$ is directed along the tangent to the trajectory and $v = |\dot{s}|$; the point velocity $\mathbf{v}$ is directed in the sense of increasing s if $\dot{s} > 0$ and decreasing s if $\dot{s} < 0$.

Second, the vector $\mathbf{a}$ lies in the osculating plane and points in the direction of the trajectory concavity; $\ddot{s} = \operatorname{comp}_{\mathbf{v}}\mathbf{a}$; $|\ddot{s}| = a^{\tau}$; the tangential acceleration vector $\mathbf{a}^{\tau}$ is directed in the sense of increasing s if $\ddot{s} > 0$ and decreasing s if $\ddot{s} < 0$.

Third, if $\dot{s}\ddot{s} > 0$, then, at the time instant under study, the velocity magnitude v increases and the motion is accelerated; for $\dot{s}\ddot{s} < 0$, the velocity magnitude v decreases and the motion is decelerated.

The normal acceleration is zero if the point moves along a rectilinear part of the trajectory, where $k = 1/\rho = 0$. The tangential acceleration of the point is zero if the velocity magnitude remains constant.

A *uniformly varying* motion of a point is a motion where $\ddot{s} = a_0 = \text{const}$. In this case, the values of the velocity and the arc coordinate can be found by the formulas

$$\begin{aligned} \dot{s} &= \dot{s}_0 + a_0(t - t_0), \\ s &= s_0 + \dot{s}_0(t - t_0) + \tfrac{1}{2}a_0(t - t_0)^2. \end{aligned} \tag{E2.1.1.7}$$

Here $\dot{s}_0$ and s_0 are the values of the velocity and the arc coordinate corresponding to the time instant t_0.

A *uniform motion* is a motion of a point in which $\ddot{s} = a_0 = 0$.

In the case of an arbitrary motion, the arc coordinate s and the path L passed over by the point can be found by the formulas

$$s = \int_{t_0}^{t} \dot{s}\,dt, \qquad L = \int_{t_0}^{t} v\,dt.$$

This clearly illustrates the difference between the arc coordinate and the path passed over by the point.

Example. An ellipsograph rod AB of length $2l = 2\,\text{m}$ (Fig. E2.3) moves so that the angle φ (in radians) varies according to the law $\varphi = -\frac{1}{2}t^2 + 2t + \frac{\pi}{6} - \frac{3}{2}$. Furthermore, point B of the rod moves along the horizontal line, and point A moves along the vertical line. Find the trajectory of point K lying at the middle of the rod, its velocity, acceleration, and the radius of curvature of the trajectory for $t = 1$ s.

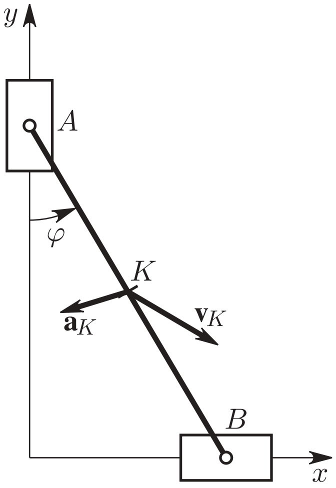

Figure E2.3. Ellipsograph (thin rod) whose endpoints can move along perpendicular straight lines.

Solution. The method used here to specify the motion of the point differs from the three above-mentioned ones. Therefore, we pass to the coordinate method and use the corresponding formulas. We introduce the coordinate axes as is shown in the figure and then find the coordinates of point K: $x = l\sin\varphi$, $y = l\cos\varphi$ (for convenience, the subscript is omitted). Squaring the right and left sides of the equations of motion and adding them termwise, we obtain the trajectory equation

$$x^2 + y^2 = l^2,$$

which implies that point K moves in a circle of radius l. Then, by using formulas (E2.1.1.2)—(E2.1.1.6) and by taking into account the fact that the quantities $\varphi = \pi/6$, $\dot{\varphi} = 1\,\mathrm{s}^{-1}$, and $\ddot{\varphi} = -1\,\mathrm{s}^{-2}$ correspond to time $t = 1$ s, we obtain

$$\begin{aligned}
v_x &= \dot{x} = l\cos\varphi\,\dot{\varphi}; \quad v_x(1) = 0.87\,\mathrm{m/s},\\
v_y &= \dot{y} = -l\sin\varphi\,\dot{\varphi}; \quad v_y(1) = -0.5\,\mathrm{m/s},\\
v &= \sqrt{v_x^2 + v_y^2}; \quad v(1) = 1\,\mathrm{m/s},\\
a_x &= \dot{v}_x = l(-\sin\varphi\,\dot{\varphi}^2 + \cos\varphi\,\ddot{\varphi}); \quad a_x(1) = -1.37\,\mathrm{m/s}^2,\\
a_y &= \dot{v}_y = l(-\cos\varphi\,\dot{\varphi}^2 - \sin\varphi\,\ddot{\varphi}); \quad a_y(1) = -0.37\,\mathrm{m/s}^2,\\
a &= \sqrt{a_x^2 + a_y^2}; \quad a(1) = 1.41\,\mathrm{m/s}^2,\\
\dot{v} &= \frac{v_x a_x + v_y a_y}{v}; \quad \dot{v}(1) = -1.0\,\mathrm{m/s}^2, \quad a^\tau(1) = 1.0\,\mathrm{m/s}^2,\\
a^n &= \sqrt{a^2 - (a^\tau)^2}; \quad a^n(1) = 1.0\,\mathrm{m/s}^2,\\
\rho &= \frac{v^2}{a^n}; \quad \rho(1) = 1.0\,\mathrm{m},
\end{aligned} \tag{E2.1.1.8}$$

and the sign of $\dot{v}$ means that, for $t = 1$s, the motion of the point is decelerated; this result is obvious in advance, as follows from the equation of the trajectory.

E2.1.2. Kinematics of a Rigid Body

► **Some notions, definitions, and theorems.** An *absolutely rigid body* (or a*rigid body*, or a *body*) is a set of points moving in space so that the distance between any two of them remains constant. (In other words, the mutual position of points of a rigid body does not vary in time.)

The problem of kinematics of a rigid body is to obtain formulas for determining the velocity vector $\mathbf{v}$ and the acceleration vector $\mathbf{a}$ for any point of the rigid body.

THEOREM 1. *If a vector* $\mathbf{b}$ *belongs to a moving rigid body* (*connects two of its points*), *then the vector* $\dot{\mathbf{b}}$ (*the time derivative*) *is either zero or orthogonal to* $\mathbf{b}$.

Indeed, by differentiating the relation $\mathbf{b}\cdot\mathbf{b} = \text{const}$ with respect to time, we obtain $\mathbf{b}\cdot\dot{\mathbf{b}} = 0$, which proves the above statement.

THEOREM 2. *The projections of the velocities of any two points of a rigid body onto the straight line connecting them are equal,* $\operatorname{comp}_{\overrightarrow{AB}}\mathbf{V}_A = \operatorname{comp}_{\overrightarrow{AB}}\mathbf{V}_B$ (Fig. E2.4).

This theorem remains valid for an arbitrary motion of the body and states that the projections must be equal in both magnitude and direction.

THEOREM 3. *If the time derivatives* $\dot{\mathbf{b}}_1$ *and* $\dot{\mathbf{b}}_2$ *of two noncollinear vectors* $\mathbf{b}_1$ *and* $\mathbf{b}_2$ *belonging to a rigid body are known, then, for any vector* $\mathbf{b}$ *belonging to this rigid body, its time derivative can be found as the cross product* $\dot{\mathbf{b}} = \boldsymbol{\omega}\times\mathbf{b}$, *where*

$$\begin{aligned}
&\boldsymbol{\omega} = 0 \ \textit{if}\ \dot{\mathbf{b}}_1 = \dot{\mathbf{b}}_2 = 0;\\
&\boldsymbol{\omega} = \eta\cdot\mathbf{b}_2,\ \textit{where}\ \eta = (\dot{\mathbf{b}}_1\cdot\dot{\mathbf{b}}_1)/(\dot{\mathbf{b}}_1\cdot(\mathbf{b}_2\times\mathbf{b}_1))\ \textit{if}\ \dot{\mathbf{b}}_1 \neq 0;\ \dot{\mathbf{b}}_2 = 0;\\
&\boldsymbol{\omega} = (\dot{\mathbf{b}}_1\times\dot{\mathbf{b}}_2)/(\dot{\mathbf{b}}_1\cdot\mathbf{b}_2)\ \textit{if}\ \dot{\mathbf{b}}_1 \neq 0;\ \dot{\mathbf{b}}_2 \neq 0 \quad (\textit{Markuzon's formula}).
\end{aligned} \tag{E2.1.2.1}$$

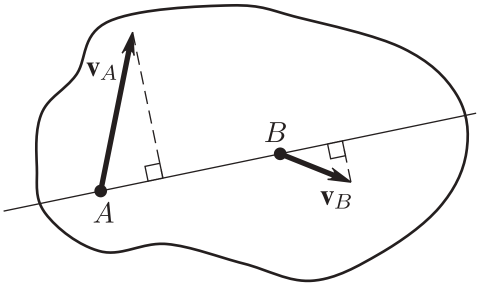

Figure E2.4. Projections of the velocities of two points of a rigid body onto the straight line passing through them.

The vector $\boldsymbol{\omega}$ is called the *vector of angular velocity of the rigid body*; it characterizes the kinematic state of the body at the time instant under study.

Although formulas (E2.1.2.1) make sense for any values of the vectors $\mathbf{b}_1$, $\mathbf{b}_2$, $\dot{\mathbf{b}}_1$, and $\dot{\mathbf{b}}_2$, one has to bear in mind that these vectors in problems of kinematics cannot be introduced arbitrarily; they must satisfy the conditions of the above-mentioned theorems of kinematics of a rigid body.

The vector $\boldsymbol{\varepsilon} = \dot{\boldsymbol{\omega}}$ is called the *vector of angular acceleration* of the rigid body.

We assume that the body moves relative to a fixed reference frame $O\xi\eta\zeta$, and at the time instant under study we know the position of point P of the body, its velocity $\mathbf{v}_P$ and acceleration $\mathbf{a}_P$, and the vectors of the body angular velocity $\boldsymbol{\omega}$ and angular acceleration $\boldsymbol{\varepsilon}$.

In this case, the velocity and acceleration of any point M of the body can be found by differentiating the vector equation $\overrightarrow{OM} = \overrightarrow{OP} + \mathbf{r}$ with respect to time (Fig. E2.5 *a*).

After the first differentiation, we obtain $\dot{\overrightarrow{OM}} = \dot{\overrightarrow{OP}} + \dot{\mathbf{r}}$, or

$$\mathbf{v}_M = \mathbf{v}_P + \boldsymbol{\omega} \times \mathbf{r}; \tag{E2.1.2.2}$$

after the repeated differentiation, we obtain $\dot{\mathbf{v}}_M = \dot{\mathbf{v}}_P + \dot{\boldsymbol{\omega}} \times \mathbf{r} + \boldsymbol{\omega} \times \dot{\mathbf{r}}$, or

$$\mathbf{a}_M = \mathbf{a}_P + \boldsymbol{\varepsilon} \times \mathbf{r} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}). \tag{E2.1.2.3}$$

Formulas (E2.1.2.2) and (E2.1.2.3) in general form solve the problem of kinematics of a rigid body, but they can hardly be used to solve problems frequently encountered in practice. In what follows, we consider special cases of motion of a rigid body.

► **Translational motion.** A motion of a rigid body is said to be *translational* if, in the course of motion, any straight line drawn in the body remains parallel to the initial direction of itself. In this case, $\boldsymbol{\omega} = 0$ and $\boldsymbol{\varepsilon} = 0$, which implies the properties of translational motion: being superimposed on each other, the trajectories of all points of a rigid body coincide (are congruent), and the velocities and accelerations of all points of the body are the same at each time instant:

$$\mathbf{v}_A = \mathbf{v}_B = \mathbf{v}, \qquad \mathbf{a}_A = \mathbf{a}_B = \mathbf{a},$$

where A and B are any points of the body.

Thus, the kinematics of translational motion of a body is reduced to the kinematics of one of its points.

► **Rotation about a fixed axis.** In *rotation of a rigid body about a fixed axis*, which is called the axis of rotation, the points lying on the axis remain fixed. The other points move in circles lying in the planes orthogonal to the rotation axis.

We introduce two reference frames, the fixed reference frame $O\xi\eta\zeta$ whose axis $O\zeta$ coincides with the rotation axis of the body and the moving frame $Oxyz$ rigidly fixed to

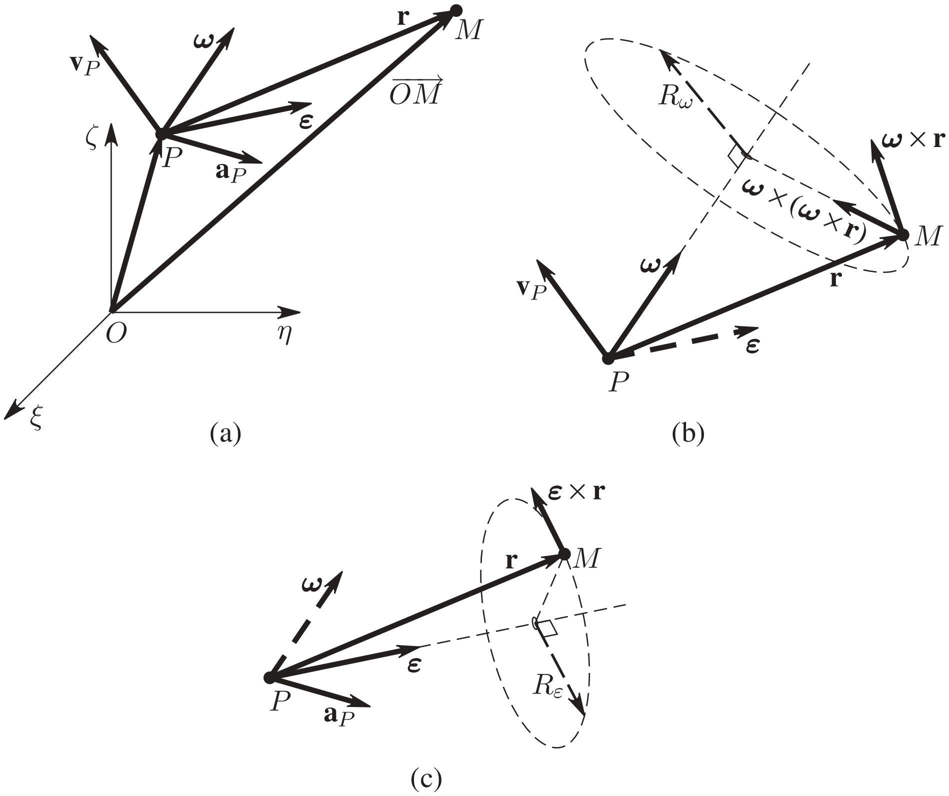

$\mathbf{v}_P$
$\boldsymbol{\omega}$
$\mathbf{r}$
M
$\overrightarrow{OM}$
$\boldsymbol{\varepsilon}$
ζ
P
$\mathbf{a}_P$
O
η
ξ
(a)
R_ω
$\boldsymbol{\omega}\times\mathbf{r}$
$\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r})$
M
$\boldsymbol{\omega}$
$\mathbf{r}$
$\mathbf{v}_P$
$\boldsymbol{\varepsilon}$
P
(b)
$\boldsymbol{\varepsilon}\times\mathbf{r}$
$\mathbf{r}$
M
$\boldsymbol{\omega}$
$\boldsymbol{\varepsilon}$
R_ε
P
$\mathbf{a}_P$
(c)

Figure E2.5. Motion of a body in the fixed reference frame $O\xi\eta\zeta$: (a) the velocity $\mathbf{v}_P$ and acceleration $\mathbf{a}_P$ of point P of the body, and the body angular velocity $\boldsymbol{\omega}$ and angular acceleration ε; (b) the directions of the vectors $\boldsymbol{\omega}\times\mathbf{r}$ and $\boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r})$; (c) the direction of the vector $\varepsilon\times\mathbf{r}$.

the body with axis Oz also directed along the rotation axis. The origin O of both frames is placed at the point of intersection of the rotation axis with the plane containing the point M whose velocity and acceleration are the desired variables.

The unit vectors $\boldsymbol{\xi}^\circ, \boldsymbol{\eta}^\circ, \boldsymbol{\zeta}^\circ$ of the fixed reference frame and $\mathbf{i}^\circ, \mathbf{j}^\circ, \mathbf{k}^\circ$ of the moving frame are shown in Fig. E2.6.

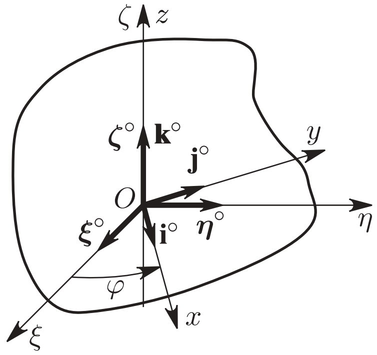

ζ
z
$\boldsymbol{\zeta}^\circ$
$\mathbf{k}^\circ$
$\mathbf{j}^\circ$
y
O
η
$\boldsymbol{\xi}^\circ$
$\mathbf{i}^\circ$
$\boldsymbol{\eta}^\circ$
φ
x
ξ

Figure E2.6. Rotation of a body about a fixed axis.

The dihedral angle φ between the moving plane $\mathbf{i}^\circ\mathbf{k}^\circ$ and the fixed plane $\boldsymbol{\xi}^\circ\boldsymbol{\zeta}^\circ$ is called the angle of rotation of the rigid body; it is measured in radians. The rotation angle is assumed to be positive if it is counted off from the fixed plane towards the moving plane anticlockwise if we look in the positive sense of the axis of rotation. To know the position of the body (and of any of its points) at any time instant, it is necessary to know the dependence of the angle φ on time t,

$$\varphi = \varphi(t).$$

This equation expresses the *law of rotation of a rigid body about a fixed axis*.

Along with the function $\varphi(t)$, its first and second time derivatives characterize the kinematic state of the rigid body at the time instant under study:

$$\omega = \dot{\varphi}, \qquad \varepsilon = \dot{\omega} = \ddot{\varphi},$$

where ω is the angular velocity of the body and ε is its angular acceleration.

The unit vectors of the moving frame are written as functions of the rotation angle φ as follows: $\mathbf{i}^\circ = \boldsymbol{\xi}^\circ \cos\varphi + \boldsymbol{\eta}^\circ \sin\varphi$, $\mathbf{j}^\circ = -\boldsymbol{\xi}^\circ \sin\varphi + \boldsymbol{\eta}^\circ \cos\varphi$, $\mathbf{k}^\circ = \boldsymbol{\zeta}^\circ = \overrightarrow{\text{const}}$. By differentiating these formulas with respect to time, we obtain $\mathbf{i}^\circ = \mathbf{j}^\circ\dot{\varphi}$, $\dot{\mathbf{j}}^\circ = -\mathbf{i}^\circ\dot{\varphi}$, after which the angular velocity vector can be found: $\boldsymbol{\omega} = (\mathbf{i}^\circ \times \dot{\mathbf{j}}^\circ)/(\mathbf{i}^\circ \cdot \mathbf{j}^\circ) = -(\mathbf{j}^\circ \times \mathbf{i}^\circ)\dot{\varphi}/(\mathbf{j}^\circ \cdot \mathbf{j}^\circ) = \boldsymbol{\zeta}^\circ\dot{\varphi}$.

After the differentiation of the last relation, we obtain $\boldsymbol{\varepsilon} = \dot{\boldsymbol{\omega}} = \boldsymbol{\zeta}^\circ\ddot{\varphi}$.

Thus, the vectors of angular velocity and angular acceleration of a body rotating about a fixed axis are directed along this axis. In this case, if we look in the direction opposite to the vector $\boldsymbol{\omega}$, then it seems that the rotation is anticlockwise (the right-hand screw rule). In these cases, the vector $\boldsymbol{\varepsilon}$ is parallel to the vector $\boldsymbol{\omega}$ if the signs of $\dot{\varphi}$ and $\ddot{\varphi}$ coincide. Otherwise, $\boldsymbol{\omega}$ and $\boldsymbol{\varepsilon}$ are antiparallel.

To determine the velocity $\mathbf{v}_M$ and the acceleration $\mathbf{a}_M$ of an arbitrary point M of the body, one can use the vector formulas (E2.1.2.2) and (E2.1.2.3), but here we perform this in a different, more illustrative way.

In the case of rotation of a body about a fixed axis, the trajectory of its arbitrary point M is a circle whose radius R is called the radius of rotation of a point. In Fig. E2.7, this circle is shown from the side of positive sense of the rotation axis. The sign of ω indicates the direction of rotation: if it is positive, then the rotation is anticlockwise. It is convenient to represent the angular velocity and the angular acceleration by using curved arrows with their sign taken into account (in Fig. E2.7, $\omega < 0$ and $\varepsilon > 0$). If ω and ε have the same signs, then one says that the rotation is accelerated, and if the signs are opposite, then the rotation is decelerated.

In the case of rotation of a body about a fixed axis, the magnitude of the velocity of point M can be found by the formula

$$v = |\omega| R; \tag{E2.1.2.4}$$

in this case the velocity is directed along the tangent to the circle, and so is perpendicular to the radius, and its sense is consistent with the sense of the curved arrow ω (Fig. E2.7). The tangential, normal, and total acceleration can be found by the following formulas, which are a special case of formulas (E2.1.2.3):

$$a^\tau = |\varepsilon| R, \quad a^n = \omega^2 R, \quad a = R\sqrt{\varepsilon^2 + \omega^4}. \tag{E2.1.2.5}$$

The tangential acceleration vector $\mathbf{a}^\tau$ is directed along the tangent to the circle, and its sense is consistent with that of the curved arrow ε; the vector $\mathbf{a}^n$ is directed along the radius R towards the center of the circle. The acute angle μ between $\mathbf{a}$ and R can be found by the formula

$$\tan\mu = \frac{a^\tau}{a^n} = \frac{|\varepsilon|}{\omega^2};$$

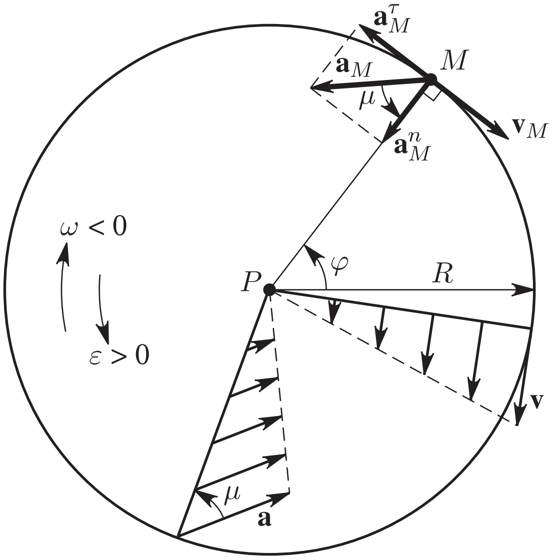

Figure E2.7. The trajectory of an arbitrary point M of a body rotating about a fixed axis; R is the radius of rotation of the point. The curved arrows denote the angular velocity and the angular acceleration.

in this case, the sense of the nearest rotation from **a** to R is consistent with the sense of the curved arrow ε.

It follows from the last two formulas that the vectors **v** and **a** of different points of the body that lie on the same radius are parallel to each other and linearly depend on the distance from the axis of rotation (Fig. E2.7):

$$\omega = \frac{v_B}{BP} = \frac{v_C}{CP} = \frac{v_D}{DP}, \qquad \varepsilon = \frac{a_B}{BP} = \frac{a_C}{CP} = \frac{a_D}{DP}. \tag{E2.1.2.6}$$

A rotation of a rigid body is said to be *uniformly varying* if $\varepsilon = \varepsilon_0 = \text{const}$. In this case, the time dependencies of the rotation angle and the angular velocity are given by the formulas

$$\omega = \omega_0 + \varepsilon_0(t - t_0),$$
$$\varphi = \varphi_0 + \omega_0(t - t_0) + \tfrac{1}{2}\varepsilon_0(t - t_0)^2,$$

which are similar to the formulas for the uniformly varying motion of a point (E2.1.1.7).

The *uniform* rotation of a rigid body is a special case of uniformly varying rotation, namely, the case in which $\varepsilon_0 = \text{const} = 0$.

E2.1.3. Plane-Parallel (Plane) Motion of a Rigid Body

► **Basic notions.** A *plane-parallel* (*plane*) motion is a motion of a body in which the distance from any of its points to some fixed plane remains unchanged. (Each point of the body moves in a plane, but the planes for different points of the body can be different.) The rotation of a body about a fixed axis is a special case of plane motion.

Studying a plane-parallel motion can be reduced to studying the motion of a plane figure in its plane. The rotation angle $\varphi(t)$ of a plane figure is counted off from a fixed straight line to a straight line rigidly fixed to the figure.

The position of a plane figure is completely determined if the coordinates of some of its points (pole, center) P and the rotation angle are known. The relations

$$x_P = x(t), \quad y_P = y(t), \quad \varphi = \varphi(t)$$

determining three functional dependencies on time are called the *equations of motion of a plane figure*.

The angular velocity vector $\boldsymbol{\omega}$ and the angular acceleration vector $\boldsymbol{\varepsilon}$ are introduced by analogy with the case of rotation of a body about a fixed axis; they are orthogonal to the base plane. One has $\omega = \dot{\varphi}$ and $\varepsilon = \dot{\omega} = \ddot{\varphi}$.

Just as in the case of rotation about a fixed axis, when solving practical problems it is convenient to represent the angular velocity and acceleration by curved arrows.

► Theorem on velocities of points of a plane figure. Instantaneous center of velocities.

THEOREM (ON VELOCITIES OF POINTS OF A PLANE FIGURE). *The velocity of any point B of a plane figure is equal to the vector sum of the velocity $\mathbf{v}_P$ of the pole and the velocity $\mathbf{v}_{BP}$ that point B would have in the imaginary rotation of the figure about the fixed pole P at the angular velocity ω* (*Fig.* E2.8):

$$\mathbf{v}_B = \mathbf{v}_P + \mathbf{v}_{BP}. \qquad \text{(E2.1.3.1)}$$

The velocity $\mathbf{v}_{BP}$ is perpendicular to segment BP, its magnitude is calculated by the formula $v_{BP} = |\omega| BP$, and the sense is consistent with the sense of the curved arrow ω.

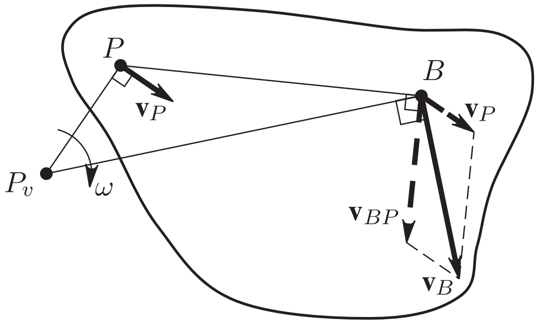

Figure E2.8. Determining the velocity of an arbitrary point of a plane figure (to the theorem on velocities of points of a plane figure).

The point P_v of a plane figure whose velocity at the time instant under study is zero is called the *instantaneous center of velocities*. If this point is taken for the pole, then the velocity of an arbitrary point B of the plane figure is determined by the formula $\mathbf{v}_B = \mathbf{v}_{BP_v}$. The direction of $\mathbf{v}_B$ is consistent with the direction of the curved arrow of the angular velocity, and its magnitude can be found by the formula $v_B = |\omega| BP_v$. The velocities of points of the plane figure at the time instant under study are distributed by analogy with the case of rotation of the figure about a fixed axis passing through the instantaneous center of velocities. Therefore, one can use formula (E2.1.2.6) for rotational motion:

$$\omega = \frac{v_B}{BP_v} = \frac{v_C}{CP_v} = \frac{v_D}{DP_v}. \qquad \text{(E2.1.3.2)}$$

In Fig. E2.9a,b,c, we show several versions of location of the instantaneous center of velocities:

(a) In rolling motion without slip on a fixed surface, the instantaneous center of velocities lies at the point of tangency (Fig. E2.9a).

(b) If $\mathbf{v}_B \parallel \mathbf{v}_C$ and $\mathbf{v}_B \perp BC$, then the position of P_v is determined by the construction shown in Fig. E2.9b.

(c) If $\mathbf{v}_B \parallel \mathbf{v}_C$ and the segment BC is not perpendicular to $\mathbf{v}_B$, then there is no instantaneous center of velocities, $\omega = 0$, and the velocities of all points of the body are the same at a given time instant (Fig. E2.9c).

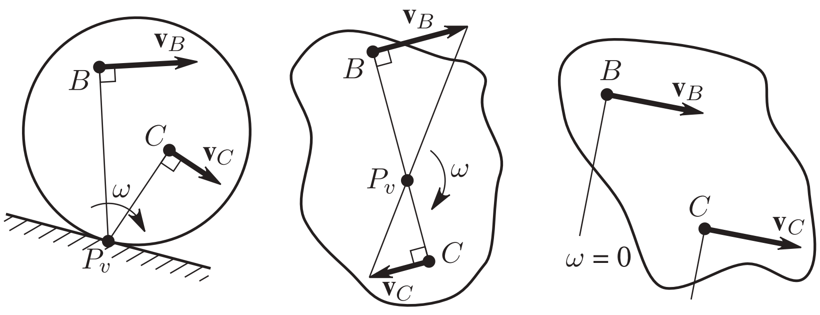

(a) (b) (c)

Figure E2.9. Possible positions of the instantaneous center of velocities: (a) in rolling motion without slip on a fixed surface; (b) the vectors at two points of the body are parallel and perpendicular to the line connecting them; (c) the vectors at two points of the body are parallel but not perpendicular to the line connecting them.

Example. At time $t = 1$ s, find the velocities of points A, B, and K of the ellipsograph rod whose motion is described in the example in Subsection E2.1.1.

Solution. The rod AB is in plane-parallel motion. Drawing perpendiculars to the directions of velocities of points A and B, we find the position of the instantaneous center P_v of velocities (Fig. E2.10). We represent the angular velocity $\omega = \dot{\varphi} = 1\,\mathrm{s}^{-1}$ by a curved arrow with sign taken into account. We direct the velocities of points A, B, and K perpendicular to the straight lines connecting them with P_v so that their senses are consistent with the curved arrow of the angular velocity. The values of the velocities can be found from the relation

$$\omega = \frac{v_A}{AP_v} = \frac{v_B}{BP_v} = \frac{v_K}{KP_v},$$

which implies $v_A = v_K = 1.0\,\mathrm{m/s}$ and $v_B = \sqrt{3} = 1.73\,\mathrm{m/s}$.

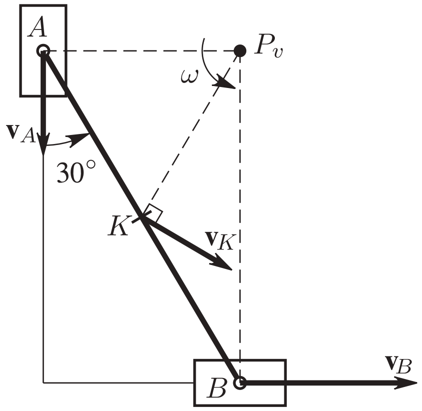

Figure E2.10. Ellipsograph (thin rod) whose endpoints move according to a given law along two perpendicular straight lines.

► Theorem on accelerations of points of a plane figure. Instantaneous center of accelerations.

THEOREM (ON ACCELERATIONS OF POINTS OF A PLANE FIGURE). *The acceleration of any point B of a plane figure is equal to the vector sum of the acceleration of the pole $\mathbf{a}_P$ and the acceleration $\mathbf{a}_{BP}$ that point B would have in the imaginary rotation of the figure about the fixed pole P:*

$$\mathbf{a}_B = \mathbf{a}_P + \mathbf{a}_{BP} = \mathbf{a}_P + \mathbf{a}^{\tau}_{BP} + \mathbf{a}^{n}_{BP}. \qquad \text{(E2.1.3.3)}$$

The magnitudes of the vectors $\mathbf{a}^{\tau}_{BP}$ and $\mathbf{a}^{n}_{BP}$ and their orientation on the plane are as

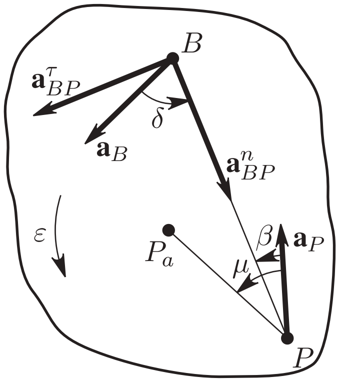

Figure E2.11. Determining the acceleration of an arbitrary point of a plane figure (to the theorem on accelerations of points of a plane figure).

follows: $\mathbf{a}_{BP}^{n} \parallel \overrightarrow{BP}$ and is directed to the pole P; $a_{BP}^{n} = \omega^2 BP$; $\mathbf{a}_{BP}^{\tau} \perp \overrightarrow{BP}$ and is consistent with the sense of the curved arrow ε, $a_{BP}^{\tau} = |\varepsilon| BP$ (Fig. E2.11).

If the vector relation (E2.1.3.3) is projected onto two directions, the parallel direction BP and the perpendicular direction BP, then we obtain the system of two scalar equations

$$a_B \cos\delta = -a_P \cos\beta + \omega^2 BP,$$
$$a_B \sin\delta = -a_P \sin\beta + |\varepsilon| BP$$

containing seven variables a_B, a_P, δ, β, BP, ω, and ε. If five of them are known, then the remaining two can be found by solving the system.

The *instantaneous center of accelerations* of a plane figure is a point P_a of this figure whose acceleration at the time instant under study is zero. The distribution of accelerations of points of the plane figure is the same as in the case of its rotation about a fixed axis passing through the point P_a.

Let us indicate the positions of the instantaneous center of velocities and the instantaneous center of accelerations of a disk that uniformly rolls without slip on a horizontal fixed straight line.

Since the rolling motion is without slip, it follows that the velocities of points of tangency of the disk and of the straight line are the same. And since the straight line is fixed, we conclude that the instantaneous center of velocities of the disk lies at the point of tangency.

The center of the disk moves uniformly, and hence its tangential acceleration is zero. Further, the trajectory of the center of the disk is rectilinear, and hence its normal acceleration is zero. Therefore, the total acceleration of the center, which is equal to the vector sum of the tangential and normal accelerations, is also zero; i.e., the instantaneous center of accelerations of the disk lies at the center of the disk.

The above example shows that the instantaneous center of velocities and the instantaneous center of accelerations are in general different points of a plane figure. But in special cases they can coincide; for example, in the case of rotation of a body about a fixed axis, both P_v and P_a always lie on the axis of rotation.

E2.1.4. Arbitrary Motion of a Rigid Body

The velocity $\mathbf{v}_M$ and the acceleration $\mathbf{a}_M$ of any point of a rigid body in an arbitrary motion can be found if, at the time instant under study, we know the velocity $\mathbf{v}_P$ and the acceleration $\mathbf{a}_P$ of a point P of the body and the vectors $\boldsymbol{\omega}$ of angular velocity and $\boldsymbol{\varepsilon}$ of angular acceleration of the body. The corresponding formulas have the form (see Eqs. (E2.1.2.2) and (E2.1.2.3))

$$\begin{aligned} \mathbf{v}_M &= \mathbf{v}_P + \boldsymbol{\omega}\times\mathbf{r}, \\ \mathbf{a}_M &= \mathbf{a}_P + \boldsymbol{\varepsilon}\times\mathbf{r} + \boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r}). \end{aligned} \tag{E2.1.4.1}$$

In the special case $\mathbf{v}_P = 0$, $\mathbf{a}_P = 0$, which takes place in *spherical motion*, formulas (E2.1.4.1) become

$$\begin{aligned} \mathbf{v}_M &= \boldsymbol{\omega}\times\mathbf{r} \qquad \text{(Euler formula)}, \\ \mathbf{a}_M &= \boldsymbol{\varepsilon}\times\mathbf{r} + \boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r}) = \boldsymbol{\varepsilon}\times\mathbf{r} + \boldsymbol{\omega}\times\mathbf{v}_M. \end{aligned}$$

In this case, the distribution of velocities of points of the body is the same as in the case of rotation about the fixed axis passing through P parallel to the vector $\boldsymbol{\omega}$; therefore, it is called the *instantaneous axis of rotation*. The velocities of all points of the body that lie on this axis at the time instant under study are zero, and a small displacement of the body can be treated as a rotation by a small angle about the instantaneous axis of rotation.

To find the specific features of the spatial orientation of the vectors contained in relations (E2.1.4.1), we rewrite them as

$$\begin{aligned} \mathbf{v}_M &= \mathbf{v}_P + \mathbf{v}_\omega, \\ \mathbf{a}_M &= \mathbf{a}_P + \mathbf{a}_\varepsilon^\tau + \mathbf{a}_\omega^n, \end{aligned} \tag{E2.1.4.2}$$

where we introduce the notation $\mathbf{v}_\omega = \boldsymbol{\omega}\times\mathbf{r}$, $\mathbf{a}_\varepsilon^\tau = \boldsymbol{\varepsilon}\times\mathbf{r}$, and $\mathbf{a}_\omega^n = \boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r})$.

These vectors can be interpreted physically as follows (see Fig. E2.5b,c):

The vector $\mathbf{v}_\omega$ is the velocity of point M of the body in the case of imaginary rotation about the fixed axis directed along the vector $\boldsymbol{\omega}$ and passing through point P. In this imaginary motion, the point moves in a circle of radius R_ω lying in the plane orthogonal to the vector $\boldsymbol{\omega}$. The velocity is directed along the tangent to this circle, and its value can be calculated by formula $v_\omega = \omega R_\omega$.

$\mathbf{a}_\omega^n$ is the vector of normal acceleration of point M in the same imaginary motion of the body. The vector is directed along the radius of the circle towards its center, and its magnitude can be calculated by the formula $a_\omega^n = \omega^2 R_\omega$.

$\mathbf{a}_\varepsilon^\tau$ is the vector of tangential acceleration of point M in the case of *another* imaginary rotation about the fixed axis directed along the vector $\boldsymbol{\varepsilon}$ and passing through point P. In this imaginary motion, the point moves in a circle of radius R_ε lying in the plane orthogonal to the vector $\boldsymbol{\varepsilon}$. The vector of tangential acceleration is directed along the tangent to this circle, and its magnitude can be calculated by the formula $a_\varepsilon^\tau = \varepsilon R_\varepsilon$.

It follows from the above that any small displacement of a rigid body at the time instant under study can be represented as composed of three simultaneous imaginary small motions:

(1) The translational motion with velocity $\mathbf{v}_P$ and acceleration $\mathbf{a}_P$ and with zero magnitudes of $\boldsymbol{\omega}$ and $\boldsymbol{\varepsilon}$.

(2) The rotation about a fixed axis directed along the vector $\boldsymbol{\omega}$ and passing through point P, with zero magnitudes of $\mathbf{v}_P, \mathbf{a}_P$, and $\boldsymbol{\varepsilon}$.

(3) The rotation about a fixed axis directed along the vector $\boldsymbol{\varepsilon}$ and passing through point P, with zero magnitudes of $\mathbf{v}_P, \mathbf{a}_P, \boldsymbol{\omega}$, and t.

We take the scalar products of both sides of Eq. (E2.1.4.1) by the unit vector $\boldsymbol{\omega}/\omega$ and obtain the equation

$$\operatorname{comp}_{\boldsymbol{\omega}} \mathbf{v}_P = \operatorname{comp}_{\boldsymbol{\omega}} \mathbf{v}_M.$$

This means that the projection of the velocity of any point of the body onto the direction of the angular velocity vector is independent of the choice of the point.

Now we find the body points that have the minimum velocity. Without loss of generality, we direct the vector $\boldsymbol{\omega}$ along the axis Pz of the reference frame $Pxyz$, place the velocity vector **v** of point P in the plane xPz, and expand it into two orthogonal components, $\mathbf{v}_{\parallel\omega}$ parallel to the axis Pz and $\mathbf{v}_{\perp\omega}$ parallel to the axis Px. On the axis Py, we choose a point S so that $PS = v_{\perp\boldsymbol{\omega}}/\omega$. A straightforward verification shows that $\mathbf{v}_S \parallel \boldsymbol{\omega}$ and $v_S = v_{\min} = v_{\parallel\omega}$. All the points lying on the axis SZ have the same property.

A system of vectors $\boldsymbol{\omega}, \mathbf{v}$ located on the same straight line is called a *screw* or a *wrench*. The scalar quantity $p = v/\omega$ is called the *parameter* of the screw.

The constant values accompanying some phenomenon or process are usually called *invariants*. Therefore, one can say that, in kinematics of a rigid body, there are two invariants with respect to the choice of the pole P, namely, the vector of angular velocity of a rigid body and the projection of the velocity of a point of the body on the direction of the angular velocity vector.

Thus, in an arbitrary motion of a rigid body, the distribution of velocities of its points at the time instant under study can be one of the following: (1) instantaneous rest, (2) instantaneous translational motion, (3) rotation about the instantaneous axis, and (4) instantaneous screw motion.

Example. A right circular cone with angle 2α at the vertex performs a spherical motion and rolls without slip on the surface of a fixed right circular cone with angle 2β at the vertex (Fig. E2.12). At the time instant under study, the axis of the moving cone, rotating about the vertical axis, has the angular velocity $\boldsymbol{\omega}_1$ and the angular acceleration $\boldsymbol{\varepsilon}_1$.

Find the angular velocity $\boldsymbol{\omega}$ and the angular acceleration $\boldsymbol{\varepsilon}$ of the moving cone at the time instant under study.

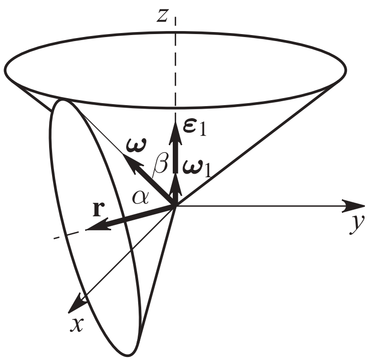

Figure E2.12. Rolling of a circular cone on the surface of another circular cone.

Solution. We introduce fixed Cartesian coordinate axes with origin at the vertex of the cone: the z-axis is directed along the axis of the fixed cone, the x-axis is perpendicular to the z-axis in the plane formed by the cone axes at the initial time instant $t = 0$, and the y-axis is orthogonal to the plane xz, so that the axes thus constructed form a right reference frame.

On the axis of the moving cone, we choose an arbitrary point M and denote the position vector connecting this point with the origin O by **r**.

The moving cone is in spherical motion. The angular velocity vector $\boldsymbol{\omega}$ is directed along the common generator of the cones, because it is an instantaneous axis of rotation. The plane $\boldsymbol{\omega}\mathbf{r}$ rotates about the fixed z-axis, and the vectors $\boldsymbol{\omega}_1$ and $\boldsymbol{\varepsilon}_1$ are directed along it.

The velocity v_M can be found if we first refer it to the cone and then to the plane $\boldsymbol{\omega}\mathbf{r}$:

$$\boldsymbol{\omega} \times \mathbf{r} = \boldsymbol{\omega}_1 \times \mathbf{r} \quad \text{or} \quad (\boldsymbol{\omega} - \boldsymbol{\omega}_1) \times \mathbf{r} = \mathbf{0},$$

which implies that $(\boldsymbol{\omega} - \boldsymbol{\omega}_1) \parallel \mathbf{r} = 0$. This means that the straight line connecting the endpoints of the vectors $\boldsymbol{\omega}$ and $\boldsymbol{\omega}_1$ is parallel to the vector **r**.

Considering the obtained triangle, we obtain $\omega = \omega_1 \sin(\alpha + \beta)/\sin\beta$ from the law of sines.

To find the vector $\boldsymbol{\varepsilon}$, we represent the vector $\boldsymbol{\omega} = \boldsymbol{\omega}(t)$ as the sum of three components parallel to the coordinate axes:

$$\boldsymbol{\omega} = \omega_1 \frac{\sin(\alpha+\beta)\sin\alpha}{\sin\beta}\sin\Bigl(\omega_1 t + \frac{\varepsilon_1 t^2}{2}\Bigr)\mathbf{i} - \omega_1 \frac{\sin(\alpha+\beta)\sin\alpha}{\sin\beta}\cos\Bigl(\omega_1 t + \frac{\varepsilon_1 t^2}{2}\Bigr)\mathbf{j} + \omega_1 \frac{\sin(\alpha+\beta)\cos\alpha}{\sin\beta}\mathbf{k},$$

where $\varphi_1 = (\omega_1 t + \varepsilon_1 t^2/2)$ is the rotation angle of the plane $\boldsymbol{\omega}\mathbf{r}$ in time t. The angular acceleration can be obtained by differentiating the resulting angular velocity vector $\boldsymbol{\varepsilon} = \dot{\boldsymbol{\omega}}$ with respect to time.

Now if the vectors $\boldsymbol{\omega}$ and $\boldsymbol{\varepsilon}$ are known, then one can find the velocity and the acceleration of any point of the moving cone.

The expression for the angular acceleration can be obtained in a simpler way by using the theorems presented below.

E2.1.5. Compound Motion of a Point

► **Basic notions.** Very often, two observers trace one and the same point M, one of them is in the laboratory reference frame $O\xi\eta\zeta$, which is assumed to be fixed, and the other is in a moving reference frame $Pxyz$, which moves relative to the fixed frame according to a given law (Fig. E2.13). In this case, if the law of motion of a point in the moving reference frame is given, then the point position in the fixed reference frame can also be determined at any time instant. The motion of a point defined in this way is called the *compound motion of a point*.

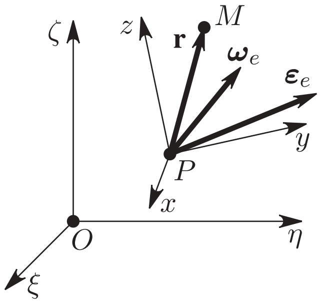

Figure E2.13. Compound motion of a point; the fixed reference frame $O\xi\eta\zeta$ and the moving reference frame $Pxyz$.

This method for specifying the motion of a point differs from those studied above. The velocity of a point and its acceleration can be determined as follows: use appropriate transformations to pass to one of the above-studied methods for determining the motion of a point and then proceed accordingly. But the same problem can be solved differently by using the notions and theorems on the velocities and accelerations in the compound motion of a point, which will be given below.

The *absolute motion of point* M is its motion relative to the fixed (absolute) reference frame. The trajectory of the point, its velocity, and acceleration are called the *absolute trajectory*, *absolute velocity* $\mathbf{v}_\mathrm{a}$ and *absolute acceleration* $\mathbf{a}_\mathrm{a}$, respectively.

The *relative motion of point* M is its motion relative to the moving reference frame. The trajectory of the point, its velocity, and acceleration relative to the moving reference frame are called the *relative trajectory*, *relative velocity* $\mathbf{v}_\mathrm{r}$, and *relative acceleration* $\mathbf{a}_\mathrm{r}$, respectively.

The *frame motion* is the motion of the moving reference frame relative to the fixed reference frame. In this motion, the moving frame “transfers” the set of points rigidly fixed to it.

The *frame velocity* $\mathbf{v}_\mathrm{e}$ and the *frame acceleration* $\mathbf{a}_\mathrm{e}$ of point M are the velocity and acceleration of the point B that is permanently attached to the moving frame and coincides with the moving point M at the given time instant.

If the velocity $\mathbf{v}_P$ and the acceleration $\mathbf{a}_P$ of a point P belonging to the moving frame, as well as the angular velocity $\boldsymbol{\omega}_\mathrm{e}$ and the angular acceleration $\boldsymbol{\varepsilon}_\mathrm{e}$ of the moving frame, are known, then $\mathbf{v}_\mathrm{e}$ and $\mathbf{a}_\mathrm{e}$ can be found by the formulas

$$\mathbf{v}_\mathrm{e} = \mathbf{v}_B = \mathbf{v}_P + \boldsymbol{\omega}_\mathrm{e} \times \mathbf{r}, \qquad \mathbf{a}_\mathrm{e} = \mathbf{a}_B = \mathbf{a}_P + \boldsymbol{\varepsilon}_\mathrm{e} \times \mathbf{r} + \boldsymbol{\omega}_\mathrm{e} \times (\boldsymbol{\omega}_\mathrm{e} \times \mathbf{r}),$$

where $\mathbf{r}$ is the position vector $\overrightarrow{PB}$ of point B of the moving frame at which the point M occurs at the time instant under study.

► Main theorems.

THEOREM ON VELOCITIES IN COMPOUND MOTION OF A POINT. *The absolute velocity vector is equal to the vector sum of the relative and frame velocity vectors,*

$$\mathbf{v}_\mathrm{a} = \mathbf{v}_\mathrm{r} + \mathbf{v}_\mathrm{e}. \tag{E2.1.5.1}$$

THEOREM ON ACCELERATIONS IN COMPOUND MOTION OF A POINT. *The absolute acceleration vector is equal to the vector sum of the relative, frame, and Coriolis acceleration vectors,*

$$\mathbf{a}_\mathrm{a} = \mathbf{a}_\mathrm{r} + \mathbf{a}_\mathrm{e} + \mathbf{a}_\mathrm{C}. \tag{E2.1.5.2}$$

The Coriolis acceleration vector is determined by the formula

$$\mathbf{a}_\mathrm{C} = 2(\boldsymbol{\omega}_\mathrm{e} \times \mathbf{v}_\mathrm{r}) \tag{E2.1.5.3}$$

and is consistent with the rules for calculating the cross product.

To find the vectors $\mathbf{a}_\mathrm{r}$ and $\mathbf{a}_\mathrm{e}$, one has to bear in mind that each of them can be the vector sum of several components, for example, of the tangential and normal accelerations.

Example. Point M moves according to the law $AM = 0.5\,t^2 - 2t + 2.5$ along the ellipsograph rod whose motion was described in the Example in Subsection E2.1.1. Find its velocity and acceleration at time $t = 1$ s.

Solution. We find the velocity and acceleration by applying the theorems on compound motion of a point. Let us introduce a moving reference frame attached to the rod AB. Under such a choice, the relative motion of point M can be introduced in a natural way. For $t = 1$ s, this point is at the middle of the segment.

We use formulas (E2.1.1.6) to obtain the relative velocity and acceleration

$$v_\mathrm{r} = \dot{L} = t - 2; \quad v_\mathrm{r}(1) = -1\ \mathrm{m/s},$$
$$a_\mathrm{r}^\tau = \ddot{L}; \quad a_\mathrm{r}^\tau(1) = 1\ \mathrm{m/s^2}, \quad a_\mathrm{r}^n = v_\mathrm{r}^2/\rho_\mathrm{r} = 0 \quad (\text{because } 1/\rho_\mathrm{r} = 0),$$

where the notation $L = AM$ is used.

In Fig. E2.14, we present the relative velocity and acceleration with signs taken into account.

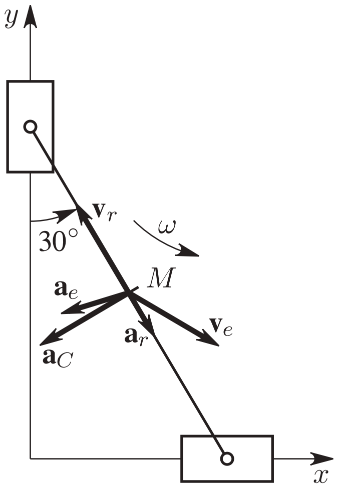

Figure E2.14. Relative and frame velocities and accelerations of point M of the ellipsograph.

The frame velocity and acceleration of point M are equal to the velocity and acceleration of the point on the ellipsograph rod where the point M occurs at time $t = 1\ s$. This point is the rod point K whose velocity and acceleration were obtained by solving the example in Subsection E2.1.1 (see formulas (E2.1.1.8)); we show them in the figure.

Now let us find the Coriolis acceleration. The angular velocity vector of the moving reference frame is directed towards the observer perpendicular to the figure plane, and the relative velocity vector lies in the figure plane. Therefore, the vector $\mathbf{a}_{\rm C}$ lies in the figure plane and is perpendicular to $\mathbf{v}_{\rm r}$, so that if we look towards $\mathbf{a}_{\rm C}$, then the nearest rotation from $\boldsymbol{\omega}$ to $\mathbf{v}_{\rm r}$ seems to be anticlockwise (Fig. E2.14). The magnitude of $\mathbf{a}_{\rm C}$ is equal to

$$a_{\rm C} = 2\omega v_{\rm r} \sin 90^\circ = 2\ \mathrm{m/s^2}.$$

Let us find the absolute velocity of point M:

$$\mathbf{v}_{\rm a} = \mathbf{v}_{\rm r} + \mathbf{v}_{\rm e}.$$

We project both sides of the vector equation onto the x- and y-axes:

$$\begin{aligned} v_{{\rm a}x} &= v_{{\rm r}x} + v_{{\rm e}x} = -1 \sin 30^\circ + 1 \cos 30^\circ = 0.37\ \mathrm{m/s}, \\ v_{{\rm a}y} &= v_{{\rm r}y} + v_{{\rm e}y} = 1 \cos 30^\circ - 1 \sin 30^\circ = 0.37\ \mathrm{m/s}, \end{aligned}$$

which implies that $v_{\rm a} = \sqrt{v_{{\rm a}x}^2 + v_{{\rm a}y}^2} = 0.52\ \mathrm{m/s}$.

Let us find the absolute acceleration of point M:

$$\mathbf{a}_{\rm a} = \mathbf{a}_{\rm r} + \mathbf{a}_{\rm e} + \mathbf{a}_{\rm C}.$$

We project both sides of the last vector relation onto the x- and y-axes:

$$\begin{aligned} a_{{\rm a}x} &= 1 \sin 30^\circ - 1.37 - 2 \cos 30^\circ = -2.60\ \mathrm{m/s^2}, \\ a_{{\rm a}y} &= -1 \cos 30^\circ - 0.37 - 2 \sin 30^\circ = -2.24\ \mathrm{m/s^2}, \end{aligned}$$

which implies that $a_{\rm a} = \sqrt{a_{{\rm a}x}^2 + a_{{\rm a}y}^2} = 3.43\ \mathrm{m/s^2}$.

When solving problems of kinematics by using theorems on the compound motion of a point, the kinematic characteristics of the relative and frame motions can be found only after the moving reference frame is chosen. This choice is not unique, because it depends on the researcher alone. Therefore, the problem on the direction of the Coriolis acceleration makes no sense until the fixed frame, the moving frame, and the law of motion of the latter are chosen.

The theorem on the velocities of a point in compound motion of a point can be stated differently. Indeed, for the point M, the theorem on velocities can be written as (Fig. E2.13)

$$\mathbf{v}_{\rm a} = \mathbf{v}_{\rm r} + \mathbf{v}_{\rm e} \quad \text{or} \quad \frac{d\overrightarrow{OM}}{dt} = \mathbf{v}_{\rm r} + \frac{d\overrightarrow{OP}}{dt} + \boldsymbol{\omega} \times \mathbf{r} \quad \text{or} \quad \frac{d\mathbf{r}}{dt} = \mathbf{v}_{\rm r} + \boldsymbol{\omega} \times \mathbf{r}.$$

Taking into account the fact that $\mathbf{v}_{\rm r}$ is the velocity of point M relative to the moving reference frame, we call it the *local* (*relative*) *derivative* of the vector $\mathbf{r}$ and denote by $\frac{\widetilde{d}\mathbf{r}}{dt}$; the quantity $\frac{d\mathbf{r}}{dt}$ will be called the *absolute derivative*. As a result, the last relation implies that

$$\frac{d\mathbf{r}}{dt} = \frac{\widetilde{d}\mathbf{r}}{dt} + \boldsymbol{\omega} \times \mathbf{r}.$$

This formula is sometimes called the *theorem on the relation between the absolute and local* (*relative*) *derivatives of a vector*.

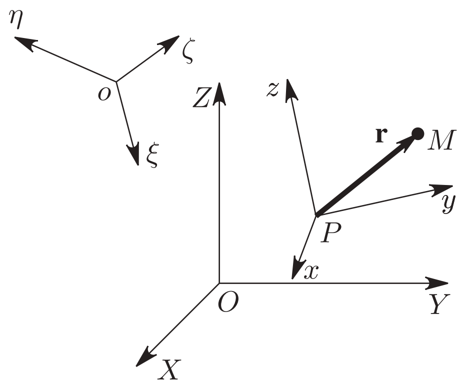

Figure E2.15. Compound motion of a body determined by three reference frames.

E2.1.6. Compound Motion of a Body

► **Absolute, relative, and frame motion.** One often encounters situations in which the position of points of a rigid body is determined by three reference frames: the frame $Pxyz$ rigidly fixed to the body, the moving frame $o\xi\eta\zeta$ in which the motion of the body is given, and the absolute (conventionally, fixed) frame $OXYZ$ relative to which the motion of the moving reference frame is given (Fig. E2.15).

The motion of a body in a moving frame is said to be relative, in the fixed frame, to be absolute, and the motion of the moving frame relative to the fixed frame is called the frame motion. The rigid body motion itself, when given as described above, is said to be *compound*.

► **Addition of translational motions.** Assume that, in a moving frame, a body is in instantaneous translational motion with velocity $\mathbf{v}_\mathrm{r}$, while the moving frame itself is in instantaneous translational motion with velocity $\mathbf{v}_\mathrm{e}$. Using the theorem on the addition of velocities in the compound motion of a point, we obtain the velocity of an arbitrary point M of the body,

$$\mathbf{v}_M = \mathbf{v}_\mathrm{a} = \mathbf{v}_\mathrm{r} + \mathbf{v}_\mathrm{e}.$$

The point M in no way participates in the final result, and this means that the velocity vectors of all points of the body are the same at the time instant under study; that is, in the case of two instantaneous translational motions, the body is in *instantaneous translational motion* with velocity equal to the vector sum of the velocities of these two instantaneous translational motions.

This result can be generalized to the case of a larger number of translational motions: the resulting motion is an instantaneous translational motion with velocity equal to the vector sum of the velocities of the original instantaneous translational motions.

► **Addition of rotations about intersecting axes.** Assume that in the moving frame the body is in instantaneous rotation about an axis with angular velocity $\boldsymbol{\omega}_1$, while the moving frame itself is in instantaneous rotation about another axis with angular velocity $\boldsymbol{\omega}_2$; moreover, the axes intersect at point P (Fig. E2.16).

Using the theorem on the addition of velocities in compound motion of a point, we obtain the velocity of an arbitrary point M of the body:

$$\mathbf{v}_M = \mathbf{v}_\mathrm{a} = \mathbf{v}_\mathrm{r} + \mathbf{v}_\mathrm{e} = \boldsymbol{\omega}_1 \times \mathbf{r} + \boldsymbol{\omega}_2 \times \mathbf{r} = (\boldsymbol{\omega}_1 + \boldsymbol{\omega}_2) \times \mathbf{r} = \boldsymbol{\Omega} \times \mathbf{r}, \quad \text{where} \quad \boldsymbol{\Omega} = \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2.$$

In other words, if the instantaneous angular velocities intersect, then the resulting motion of the body is an instantaneous rotation, with angular velocity equal to their vector sum,

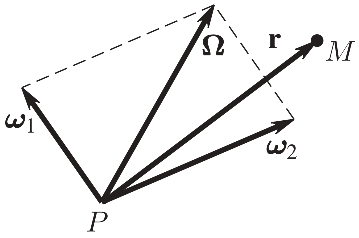

Figure E2.16. Addition of rotations about intersecting axes.

about an axis passing through the point of intersection. This statement holds for any number of instantaneous rotations about axes intersecting at one point.

► **Couple of rotations.** Assume that a body in the moving reference frame is in instantaneous rotation about an axis with angular velocity $\boldsymbol{\omega}'$, while the moving frame itself is in instantaneous rotation about an axis with angular velocity $\boldsymbol{\omega}''$. Moreover, the axes are parallel, and the angular velocities are equal and opposite, $\omega' = \omega'' = \omega$ and $\boldsymbol{\omega}' = -\boldsymbol{\omega}''$ (Fig. E2.17).

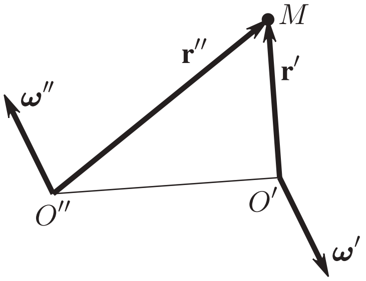

Figure E2.17. Addition of rotations about parallel axes (couple of rotations).

Using the theorem on the addition of velocities in compound motion of a point, we obtain the velocity of an arbitrary point M of the body:

$$\mathbf{v}_M = \mathbf{v}_a = \mathbf{v}_r + \mathbf{v}_e = \boldsymbol{\omega}' \times \mathbf{r}' + \boldsymbol{\omega}'' \times \mathbf{r}'' = \boldsymbol{\omega}' \times (\mathbf{r}' - \mathbf{r}'') = \boldsymbol{\omega}' \times \overrightarrow{O''O'} = \boldsymbol{\omega}'' \times \overrightarrow{O'O''} = \overline{\mathrm{mom}}_{\boldsymbol{\omega}',\boldsymbol{\omega}''},$$

where the symbol $\overline{\mathrm{mom}}_{\boldsymbol{\omega}',\boldsymbol{\omega}''}$ denotes the vector of *moment of the couple of rotations*.

The arbitrary chosen point M of the body does not occur anywhere in the final result, and this means that the velocity vectors of all points of the body are the same at the time instant under study; i.e., in the case of a couple of rotations, the body is in *instantaneous translational motion* with velocity equal to the moment of the couple of rotations.

The result thus obtained can be generalized to the case of several couples of rotations: the resulting motion of the body is an instantaneous translational motion with velocity equal to the vector sum of the moments of couples of rotations.

One can show that

1. A couple of instantaneous rotations can be replaced by another equivalent couple. In this change, it is necessary that the vector of the moment of the new couple of rotations be equal to the vector of the moment of the initial couple.

2. Two couples of rotations are equivalent to one couple whose moment is equal to the vector sum of moments of the initial couples.

It follows from the above that any instantaneous motion of a rigid body can be described by using a *system of sliding angular velocity vectors*.

► **Theorems on sliding vectors.** In the theory of sliding vectors, the following theorems hold:

THEOREM 1. *An arbitrary system of sliding vectors can be reduced to an equivalent system consisting of a sliding vector applied at an arbitrary chosen point P and a couple of vectors; here the vector is equal to the vector sum of vectors of the system (the resultant vector of the system), and the vector of the couple moment is equal to the vector sum of moments of all vectors of the system about the point P (the resultant moment of the system about the point P).*

THEOREM 2. *Two systems of sliding vectors are equivalent if, after being referred to the same point P, their resultant vectors and their resultant moments coincide.*

THEOREM 3. *If a system of sliding vectors is first referred to point P' and then to point P'', then the resultant vectors $\mathbf{\Omega}'$ and $\mathbf{\Omega}''$ and the moments $\mathbf{M}'$ and $\mathbf{M}''$ of the resultant couples are related as*

$$\mathbf{\Omega}'' = \mathbf{\Omega}' = \mathbf{\Omega},$$
$$\mathbf{M}'' = \mathbf{M}' + \overrightarrow{P''P'} \times \mathbf{\Omega}'.$$

THEOREM 4. *It follows from Theorem 3 that the following quantities are independent of the choice of the reference point P:*

$$\mathbf{\Omega}'' = \mathbf{\Omega}' = \mathbf{\Omega},$$
$$\mathbf{\Omega}' \cdot \mathbf{M}' = \mathbf{\Omega}'' \cdot \mathbf{M}''.$$

Therefore, the resultant vector of the system and the scalar product of the resultant vector by the resultant moment are called *invariants of the system of sliding vectors* about the reference center.

The second invariant with the first taken into account can be written in another form:

$$\operatorname{comp}_{\mathbf{\Omega}} \mathbf{M}' = \operatorname{comp}_{\mathbf{\Omega}} \mathbf{M}'';$$

i.e., the projection of the resultant moment onto the direction of the resultant vector is independent of the choice of the center of reference.

THEOREM 5. *The locus of points of reference with coordinates x, y, z at which the vectors $\mathbf{\Omega}$ and $\mathbf{M}$ are parallel is the straight line*

$$\frac{M_{Ox} - (y\Omega_z - z\Omega_y)}{\Omega_x} = \frac{M_{Oy} - (z\Omega_x - x\Omega_z)}{\Omega_y} = \frac{M_{Oz} - (x\Omega_y - y\Omega_x)}{\Omega_z}, \qquad \text{(E2.1.6.1)}$$

where $\Omega_x, \Omega_y, \Omega_z$ and M_{Ox}, M_{Oy}, M_{Oz} are the projections onto the Cartesian coordinate axes $Oxyz$ of the resultant vector $\mathbf{\Omega}$ and the resultant moment $\mathbf{M}_O$ about the origin O.

The straight line (E2.1.6.1) is called the *screw axis* or the *central axis*, and the pair consisting of the resultant vector and the parallel vector of the resultant couple is called the *screw*.

Thus, just as in Subsection E2.1.4, we conclude that for $\mathbf{\Omega} \cdot \mathbf{M} \neq 0$ the motion of the body can be reduced to instantaneous screw motion.

In special cases ($\mathbf{\Omega} \cdot \mathbf{M} = 0$), there may be states of instantaneous rest ($\Omega = 0$ and $M = 0$), of instantaneous translational motion ($\Omega = 0$ and $M \neq 0$), and of rotation about instantaneous axis ($\mathbf{\Omega} \neq 0$ and $M = 0$ or $\mathbf{M} \perp \mathbf{\Omega}$).

E2.2. Basic Notions and Laws of Mechanics

E2.2.1. Basic Notions of Mechanics

► **Mechanical systems. Forces.** A *mechanical system* is defined to be a set of point particles in which the position, velocity, and acceleration of any point depend on the positions, velocities, and accelerations of the other points.

A *constrained mechanical system* is a system of points with constraints imposed on their positions and velocities. The material bodies realizing these constraints are called the *mechanical constraints* imposed on the mechanical system. The relations between the coordinates and velocities of points of the system and time are called *constraint equations*. A material system is said to be *free* if there are no mechanical constraints in the system.

The kinematic state of interacting points (bodies) of a mechanical system is changed. A measure of such interaction is the *force*, depicted in the figures as a directed rectilinear segment. The *line of action of the force* is the straight line on which this segment lies.

A set of forces distinguished according to some property is called a *system of forces* and is denoted by $\{\mathbf{F}_i\}$.

A system of forces is said to be *balanced* if, being applied to a free rigid body at rest, it does not take the body out of this state.

Systems of forces $\{\mathbf{F}_i\}$ and $\{\mathbf{G}_j\}$ are said to be *equivalent* if, being separately applied to a free rigid body, they cause the same changes in its kinematic state.

If there exists a force equivalent to a system of forces applied to a rigid body, then it is called the *resultant* of the system of forces.

The forces of mechanical constraints acting on points of a system are called *constraint reaction forces*. The forces that act on points of a system and are not reaction forces are called *active forces*.

► **Moment of force.** *The moment of a force* $\mathbf{F}$ *about a center* (*point*) P is the vector $\mathbf{M}_P(\mathbf{F})$ equal to the cross product of the position vector $\mathbf{r}$ drawn from the center to the force application point by the vector of force, $\mathbf{M}_P(\mathbf{F}) = \mathbf{r}\times\mathbf{F}$ (Fig. E2.18).

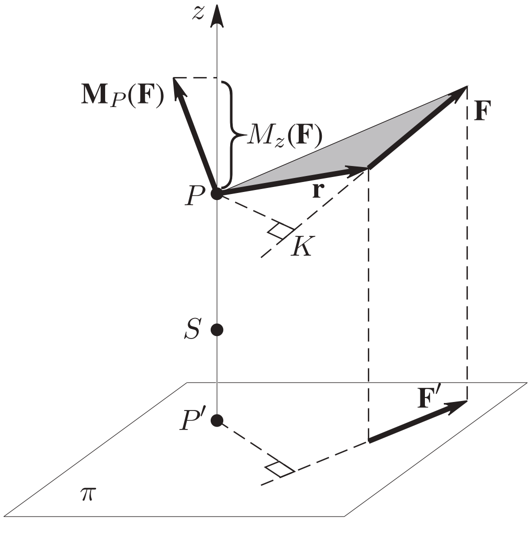

Figure E2.18. Moment of the force $\mathbf{F}$ about the center (point) P.

The magnitude of the moment of force is equal to the product of the magnitude of force by its moment arm about the center. The *moment arm of a force* $\mathbf{F}$ about a center P is the segment of the perpendicular PK drawn from the center to the force action line. The vector $\mathbf{M}_P(\mathbf{F})$ is perpendicular to the plane where the force and the center are located and is directed so that if we look in the direction opposite to this vector, then it seems that rotation of the force about the center is anticlockwise (the right-hand screw rule). If the force is translated along the line of its action, then its moment about the same center does not change.

If two arbitrary centers P and S are chosen on an axis (for example, on the z-axis), then

$$\operatorname{comp}_z \mathbf{M}_S(\mathbf{F}) = \operatorname{comp}_z(\overrightarrow{SP} + \mathbf{r}) \times \mathbf{F} = \operatorname{comp}_z \mathbf{M}_P(\mathbf{F});$$

i.e., the projection of the moment of a force about a center lying on an axis onto this axis is independent of the choice of the center. This scalar quantity is called the *moment of the force* **F** *about the z-axis* and is denoted by $M_z(\mathbf{F})$.

To calculate the moment of force about a given axis, one should

(1) Draw any plane π perpendicular to this axis and mark the point P' of their intersection (Fig. E2.18).

(2) Find the projection $\mathbf{F}'$ of the force onto this plane.

(3) Find the magnitude of the moment of the vector $\mathbf{F}'$ about the center P', which is equal to the product $F'h'$.

(4) Supply the obtained quantity with a sign observing the rotation of $\mathbf{F}'$ about P' from the side of positive sense of the axis, so that the positive sign corresponds to anticlockwise rotation.

The moment of a force about an axis does not vary if the force is transferred along its line of action. It is zero if the force and the axis are in the same plane; in this case, the line of action of the force is either parallel to the axis or intersects it.

The moment of a force **F** about the origin O of a reference frame and the moments of the force about the coordinate axes are calculated by the formulas

$$\mathbf{M}_O(\mathbf{F}) = \mathbf{r} \times \mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x & y & z \\ F_x & F_y & F_z \end{vmatrix} = M_x(\mathbf{F})\,\mathbf{i} + M_y(\mathbf{F})\,\mathbf{j} + M_z(\mathbf{F})\,\mathbf{k};$$
$$M_x(\mathbf{F}) = yF_z - zF_y, \quad M_y(\mathbf{F}) = zF_x - xF_z, \quad M_z(\mathbf{F}) = xF_y - yF_x,$$

where F_x, F_y, and F_z are projections of the force onto the coordinate axes and x, y, and z are the coordinates of the force application point.

The *resultant vector of a system of forces* $\{\mathbf{F}_k\}$ is the vector **R** equal to the vector sum of all forces in the system:

$$\mathbf{R} = \sum_{k=1}^{n} \mathbf{F}_k.$$

The *resultant moment of a system of forces about a center P* is the vector $\mathbf{M}_P$ equal to the vector sum of the vectors of moments of all forces in the system about the center P:

$$\mathbf{M}_P = \sum_{k=1}^{n} \mathbf{M}_P(\mathbf{F}_k) = \sum_{k=1}^{n} \mathbf{r}_k \times \mathbf{F}_k.$$

Here $\mathbf{r}_k$ is the position vector drawn from the center P to the point of application of the force $\mathbf{F}_k$.

► **Couple of forces and its properties.** A couple of forces is a system of two forces $\mathbf{F}'$ and $\mathbf{F}''$ applied to the same rigid body so that they are equal in magnitude, parallel, and have opposite senses: $\mathbf{F}' = -\mathbf{F}''$ (Fig. E2.19).

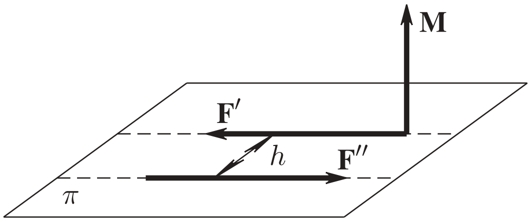

Figure E2.19. Couple of forces.

The shortest distance h between the lines of action of the forces in a couple is called the *arm of the couple*, and the plane where the forces lie is called the *plane of action* of the couple.

Properties of a couple of forces:

1. The resultant vector of a couple of forces is zero.

2. The resultant moment $\mathbf{M}(\mathbf{F}', \mathbf{F}'') = \mathbf{M}$ of a couple of forces is independent of the choice of the center and is called the *torque of the couple of forces*. It is equal to the moment of one of the forces in the couple about the point of application of the other force. The torque of a couple of forces is the measure of its mechanical action on a rigid body. Under the action of a couple of forces, a free rigid body being originally at rest begins to rotate about the axis parallel to the torque of the couple.

E2.2.2. Basic Laws of Mechanics

► **1. Law of inertia** (*Newton's first law*): if no forces act on a free point particle, then it preserves the rest state or a uniform rectilinear motion.

► **2. Law of proportionality between force and acceleration** (*Newton's second law*) can be written as the vector equation

$$m\mathbf{a} = \mathbf{F},$$

according to which a force acting on a free point particle causes point acceleration whose vector is parallel to the force and proportional to its magnitude.

The *point mass* m enters the second law as a scalar coefficient of proportionality. In mechanics, the mass is assumed to be constant; it is a measure of inertia and gravitational properties of a point particle.

The reference frames in which Newton's first and second laws are satisfied are called *inertial reference frames*. If these laws are not satisfied in a reference frame, then it is said to be *noninterial*.

► **3. Law of equality of action and reaction** (*Newton's third law*): to each action, there corresponds an equal and oppositely directed reaction.

The force actions of some material bodies on other bodies cannot be one-sided; they are always reciprocal. The forces do not arise by themselves but are the results of contact or spatial interaction of bodies. The appearance of a force acting on a body assumes that there is another body, which, in turn, is under the action of a force acting from the first body. The law states that these forces are equal in magnitude and opposite in direction, and the lines of their action coincide.

► **4.** A system of forces applied to a point has the resultant $\mathbf{R}^*$ equal to their vector sum. A consequence of this law is the following theorem.

THEOREM (VARIGNON) ON THE MOMENT OF THE RESULTANT OF A SYSTEM OF CONVERGING FORCES. *The moment of the resultant of a system of converging forces about an arbitrary center is equal to the vector sum of the moments of forces of the system about the same center.*

► **5.** The kinematic state of a constrained mechanical system does not vary if the mechanical constraints imposed on the system are replaced by the constraint reaction forces.

E2.3. Statics

E2.3.1. Basic Laws and Theorems of Statics

► **The problems of statics** include

(1) Transforming systems of forces into equivalent systems of forces.
(2) Determining equilibrium conditions for systems of forces acting on a rigid body.

► **Axioms of statics.** The general laws of mechanics are supplemented in statics with the following axioms.

1. *Axiom about the balance of two forces applied to a rigid body*: two forces applied to a rigid body are mutually balanced only if the forces have opposite directions along the common line of action and their magnitudes are equal to each other.

2. *Axiom on inclusion and exclusion of balanced systems of forces*: the action of a system of forces on a rigid body does not vary if a balanced system of forces is added to or removed from the system.

It follows from the last two axioms that a force applied to a rigid body can be transferred along its line of action, whereby the action of the force on the body remains the same.

3. *Axiom on preservation of equilibrium of a mechanical system under the action of additionally imposed mechanical constraints*: if a mechanical system is in equilibrium, then the equilibrium remains preserved if an additional mechanical constraint is imposed on the system. In particular, this implies that the equilibrium of a deformed body under the action of a system of forces is preserved in hardening, because the body hardening process is equivalent to imposing additional mechanical constraints on the system of point particles forming the deformed body.

► **Basic theorems of statics. Several formulas.** The axioms of statics define systems of forces applied to a rigid body as systems of sliding vectors, and therefore, all results of the theory of sliding vectors remain true for systems of forces.

THEOREM ON EQUIVALENCE OF COUPLES OF FORCES. *Two couples of forces are equivalent if their torques coincide.*

This means that, without changing the action of a couple on a rigid body, it can be transferred to any place in the plane of its action; the couple can be turned and the force magnitudes and the arm can be varied so that the rotation direction and the magnitude of the couple torque remain constant. A couple can be transferred into another plane parallel to the initial plane of the couple action. Thus, the main characteristic of a couple is the vector of its torque, and hence it is more convenient to represent the couple by this vector rather than by two antiparallel forces.

THEOREM ON THE RESULTANT COUPLE OF FORCES. *A system of couples of forces applied to a rigid body is equivalent to a single couple of forces whose torque is equal to the vector sum of torques of the original couples.*

THEOREM ON THE PARALLEL TRANSFER OF A FORCE. *One can transfer a force to any point of a rigid body without changing its action on the body provided that one simultaneously adds a couple of forces (an associated couple) whose torque is equal to the moment of the force to be transferred about the point where the force is to be transferred.*

THEOREM ON REDUCING A SYSTEM OF FORCES TO A CENTER. *An arbitrary system of forces applied to a rigid body is equivalent to a single force applied at an a priori chosen center P and a single couple of forces. The force is equal to the resultant vector* $\mathbf{R}$ *of the system of forces, and the torque of the couple is equal to the resultant moment* $\mathbf{M}_P$ *of the system of forces about the chosen center.*

THEOREM ON EQUIVALENCE OF SYSTEMS OF FORCES APPLIED TO A RIGID BODY. *For two systems of forces to be equivalent, it is necessary and sufficient that they have equal resultant vectors and resultant moments about the same center.*

If a system of forces is reduced first to a point P' and then to a point P'', then the resultant vectors $\mathbf{R}'$ and $\mathbf{R}''$ and the resultant moments $\mathbf{M}'$ and $\mathbf{M}''$ satisfy the relations

$$\mathbf{R}'' = \mathbf{R}' = \mathbf{R},$$
$$\mathbf{M}'' = \mathbf{M}' + \overrightarrow{P''P'} \times \mathbf{R}'.$$

The invariants of a system of forces are the resultant vector and the scalar product of the resultant vector by the resultant moment of the system; they are independent of the choice of the reduction center:

$$\mathbf{R}'' = \mathbf{R}' = \mathbf{R},$$
$$\mathbf{R}' \cdot \mathbf{M}' = \mathbf{R}'' \cdot \mathbf{M}''.$$

The second invariant, with the first invariant taken into account, can be represented as

$$\operatorname{comp}_{\mathbf{R}} \mathbf{M}' = \operatorname{comp}_{\mathbf{R}} \mathbf{M}''.$$

For $\mathbf{R}\cdot\mathbf{M} \neq 0$, the system of forces can be reduced to a screw (wrench), and the equation of the screw axis has the form

$$\frac{M_{Ox} - (yR_z - zR_y)}{R_x} = \frac{M_{Oy} - (zR_x - xR_z)}{R_y} = \frac{M_{Oz} - (xR_y - yR_x)}{R_z},$$

where R_x, R_y, R_z and M_{Ox}, M_{Oy}, M_{Oz} are the projections on the Cartesian coordinate axes $Oxyz$ of the resultant vector $\mathbf{R}$ and of the vector $\mathbf{M}_O$ of the resultant moment about the origin O.

Thus, just as in Subsection E2.1.4, we conclude that for $\mathbf{R}\cdot\mathbf{M} \neq 0$ the system of forces can be reduced to a wrench. The special cases where $\mathbf{R}\cdot\mathbf{M} = 0$ include the case of a balanced system of forces ($R = 0$ and $M = 0$), the case of reduction to the resultant couple of forces ($R = 0$ and $\mathbf{M} \neq 0$), and the case of nonzero resultant force $\mathbf{R} \neq 0$.

THEOREM (VARIGNON) ON THE MOMENT OF THE RESULTANT. *If a system of forces has a resultant* $\mathbf{R}^*$*, then its moment about any center* P *is equal to the vector sum of the moments of all forces of the system about the same center:*

$$\mathbf{M}_P(\mathbf{R}^*) = \sum_{k=1}^{n} \mathbf{M}_P(\mathbf{F}_k).$$

Corollary: the moment of the resultant about an arbitrary axis is equal to the algebraic sum of moments of all forces of the system about the same axis.

E2.3.2. Balance Conditions for a System of Forces Applied to a Rigid Body

► **Equations of equilibrium of a rigid body.** For a system of forces applied to a rigid body to be balanced, it is necessary and sufficient that its resultant vector and resultant moment about a center, for example, about the origin O, be zero:

$$\mathbf{R} = \sum_k \mathbf{F}_k = 0; \qquad \mathbf{M}_O = \sum_k \mathbf{M}_O(\mathbf{F}_k) = 0. \tag{E2.3.2.1}$$

In coordinate form, the *balance conditions* for a system of forces applied to a rigid body are obtained from the vector equations and have the form of the system of six equations

$$\begin{aligned} &\sum_k F_{kx} = 0, \qquad \sum_k F_{ky} = 0, \qquad \sum_k F_{kz} = 0,\\ &\sum_k M_x(\mathbf{F}_k) = 0, \quad \sum_k M_y(\mathbf{F}_k) = 0, \quad \sum_k M_z(\mathbf{F}_k) = 0. \end{aligned} \tag{E2.3.2.2}$$

This system is called the *primary system of equations of equilibrium of a rigid body* under the action of an arbitrary system of forces.

The upper three equations are called the *equations of projections* of forces onto the coordinate axes; they reflect the fact that if a rigid body is in equilibrium, then the algebraic sum of projections of all forces applied to the body onto each of the coordinate axes must be zero. The lower three equations are called the *equations of moments of forces* about the coordinate axes. These equations show that if a body is in equilibrium, then the algebraic sum of moments of all forces acting on the body about each of the coordinate axes must be zero.

There are other forms of the system of equilibrium equations, different from the primary system, each of which consists of six equations. When using them, one should verify conditions constraining the choice of axes about which the sums of moments of the forces are calculated.

In each problem on the equilibrium of a rigid body or a structure consisting of several bodies, in addition to some given quantities, there are quantities that must be determined when solving the problem. An equilibrium problem is said to be *statically determinate* if it can be solved completely by methods of statics. In this case, it is necessary that the number of unknowns does not exceed the number of equilibrium equations. A problem is said to be *statically indeterminate* it if cannot be solved by using the equations of statics alone. Thus, any equilibrium problem with seven or more unknowns for a single body is a priori statically indeterminable.

► **Special cases of equilibrium equations for a rigid body.** In special cases of arrangement of forces, one or several equations in the primary system can become identities. For example, (i) if all the forces are perpendicular to the coordinate z-axis, then the equation for the projections onto the z-axis becomes an identity; (ii) if the lines of action of all forces intersect the y-axis, then the equation for the moments about the y-axis becomes an identity. (In these cases, the number of equilibrium equations decreases to five.)

In what follows, we consider systems of forces whose equilibrium is described by three equilibrium equations.

A system of forces is said to be *converging* if the lines of action of the forces intersect at a single point. We choose the origin at this point. The equations of moments of the main system become identities, and the remaining three equations

$$\sum_k F_{kx} = 0, \quad \sum_k F_{ky} = 0, \quad \sum_k F_{kz} = 0 \tag{E2.3.2.3}$$

form the system of equilibrium equations.

A system of forces is said to be *plane* if all the forces lie in the same plane. We direct the coordinate axes so that all the forces lie in the plane xOy; in this case, the third, fourth, and fifth equations of the main system become identities, and the remaining three equations

$$\sum_k F_{kx} = 0, \quad \sum_k F_{ky} = 0, \quad \sum_k M_O(\mathbf{F}_k) = 0 \qquad \text{(E2.3.2.4)}$$

form a system of equilibrium equations of a plane system of forces. Here the equation of moments about the z-axis is written in a different equivalent form: the moment vectors are perpendicular to the plane of action of the forces if the center lies in the same plane. Therefore, the moment of a force about the center can be viewed as a scalar quantity. In a right reference frame, the moment is assumed to be positive if the force tends to rotate anticlockwise about the center.

In the case of a plane system of forces, one can use other forms of equilibrium equations. One of them is

$$\sum_k F_{kx} = 0, \quad \sum_k M_P(\mathbf{F}_k) = 0, \quad \sum_k M_S(\mathbf{F}_k) = 0,$$

where P and S are any points of the plane for which the segment PS is not perpendicular to the x-axis.

Another form of equilibrium equations is

$$\sum_k M_P(\mathbf{F}_k) = 0, \quad \sum_k M_S(\mathbf{F}_k) = 0, \quad \sum_k M_E(\mathbf{F}_k) = 0,$$

where P, S, and E are any points on the plane that do not lie on the same straight line.

A system of forces is called a *system of parallel forces* if the lines of action of the forces are parallel. Let the z-axis be parallel to the lines of action. The first, second, and sixth equations of the main system become identities, and the remaining equations

$$\sum_k F_{kz} = 0, \quad \sum_k M_x(\mathbf{F}_k) = 0, \quad \sum_k M_y(\mathbf{F}_k) = 0$$

form the system of equilibrium equations for a body under the action of parallel forces.

E2.3.3. Solution of the Problems of Statics

► **The general scheme of solution of the equilibrium problem for a body** (or for structures consisting of several bodies) includes several stages. One should

1. Choose a body (or a structure) the study of whose equilibrium allows one to determine the desired quantities. Draw an *assumption diagram*, i.e., a simplified figure that contains only linear dimensions and angles necessary to solve the problem but does not contain insignificant details.
2. In the diagram, draw the active forces given in the assumptions of the problem.
3. In the case of a constrained body, omit the mechanical constraints imposed on it and replace their action by constraint reactions. After such a change, the body becomes free.
4. Verify whether the necessary condition for the problem to be statically determinate is satisfied: the number of unknowns in the assumption diagram should not exceed the number of equilibrium equations for the system of forces under study.
5. If this condition is satisfied, write out the system of equilibrium equations, solve it, and study the results.

When solving the problem, it is desirable to follow the above scheme strictly.

Experience shows that errors most frequently creep in when replacing the constraints by their reaction forces. Consider this issue in more detail.

► **Rules for the alignment of mechanical constraint reaction forces**. Every mechanical constraint is either a body or a mechanical device that imposes some constraints on the body displacements in space. Depending on the type of the constraint, some displacements are prohibited and some displacements are allowed. This permits one to predict some qualitative characteristics even before calculating the numerical values of the constraint forces. The rule that one should stick to when replacing the constraints by reaction forces is the following: the constraint reaction can generally consist of two force factors, namely, a force applied at the point where the constraint is imposed and a couple of forces.

If a constraint prohibits a translational displacement of the body, then there appears a reaction force whose direction is opposite to that of the prohibited displacement. If a constraint prohibits a rotation of the body, then there appears a couple of constraint reaction forces whose torque is directed along the axis of the prohibited rotation.

► **Some types of mechanical constraints**.

1. *Perfectly flexible unstretchable massless string*. In Fig. E2.20, a string attached to a rigid body at point B is shown. The constraint in question is imposed at point B, and since this is a perfectly flexible string, the body is allowed to rotate about any axis passing through this point. This means that no couple of reaction forces appears in this case.

For this type of constraints, small translational displacements of a body are allowed in which point B moves on the surface of a sphere of radius BK centered at point K. Therefore, no reaction force can arise in the direction of allowed displacements. The translational displacement of the body in direction KB is prohibited, because the string is unstretchable. This means that there arises a reaction force $\mathbf{N}_B$ applied to the body at point B in the direction of the straight line BK.

2. One aligns reaction forces in a similar way if the constraint is realized either by *free support of two bodies* such that the surface of one of them is absolutely smooth (Fig. E2.21) or by a *moving hinge* B (Fig. E2.22).

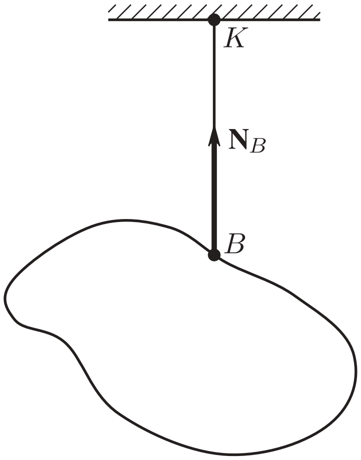

Figure E2.20. Perfectly flexible unstretchable massless string.

3. A *fixed hinge* A (Fig. E2.22) does not prevent beam AB from rotation about point A; therefore, no couple of reaction forces appears at point A. The device under study prohibits any translational displacement; therefore, neither the magnitude nor the direction of the reaction force $\mathbf{R}_A$ is known in advance. In this case, $\mathbf{R}_A$ is usually represented by two components $\mathbf{x}_A$ and $\mathbf{y}_A$ parallel to the coordinate axes.

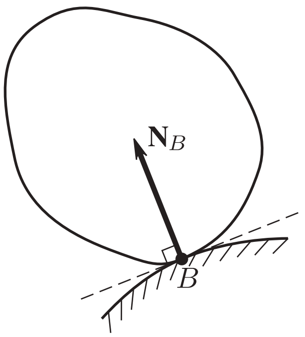

Figure E2.21. Free support of two bodies.

4. *Rigid clamping*. The constraint prohibits any translational displacement of the body; therefore, there arises a constraint reaction force $\mathbf{R}_B$ whose direction is unknown. It is usually represented by three components $\mathbf{x}_B, \mathbf{y}_B, \mathbf{z}_B$ parallel to the coordinate axes. Rigid clamping prohibits rotation about any axis passing through point B; therefore, there arises a couple of reaction forces whose torque $\mathbf{M}_B$ is known in advance neither in magnitude nor in direction. The unknown couple is usually represented by an equivalent system of three couples whose torques $\mathbf{M}_{Bx}, \mathbf{M}_{By}, \mathbf{M}_{Bz}$ are directed along the coordinate axes (Fig. E2.23).

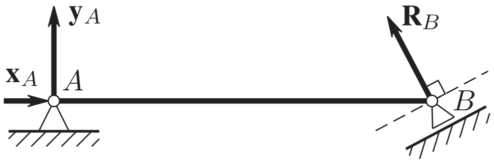

Figure E2.22. Fixed hinge.

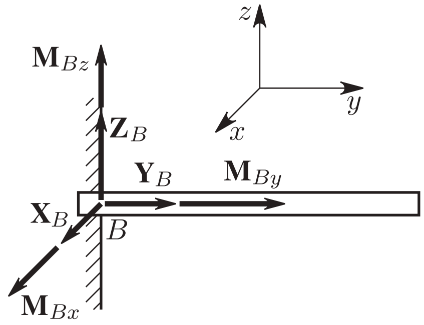

Figure E2.23. Rigid clamping.

If several constraints are imposed on a body, then each of them can be investigated independently of the other constraints and the forces applied to the body.

The unknowns in problems of statics can be not only the constraint reactions but also angles, linear dimensions of structures, and other parameters.

Example 1. Find the reaction of the rigid clamping B of a bent massless beam BDE (Fig. E2.24a) subjected to a force F = 20 kN and a couple of forces with torque M = 2 kN m assuming that α = 30°, BD = 2 m, and DE = 1 m.

Solution. We follow the general scheme for solving body equilibrium problems.

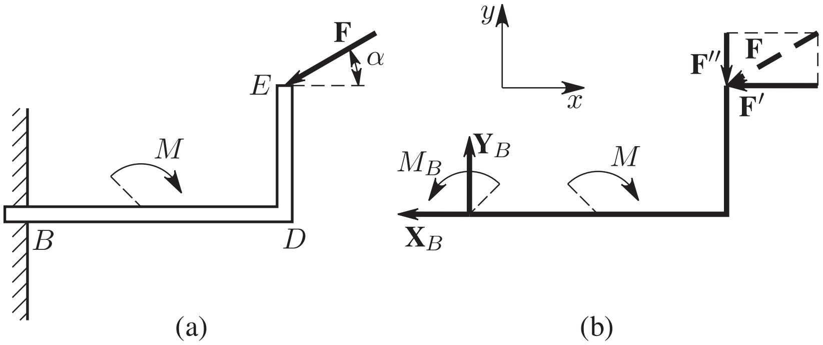

(a) (b)

Figure E2.24. Bent beam with rigidly clamped left end is subjected to a force F and a couple of forces with torque M: (a) active forces; (b) constraint reaction forces.

1. The reaction of the rigid clamping B can be found by studying the beam equilibrium; therefore, for the rigid body we take the beam BDE. We separately draw the assumption diagram (Fig. E2.24b).

2. On the diagram, we align the active force factors acting on the beam, i.e., the force **F** and the couple of forces **M**, which we present as a curved arrow in the direction of rotation of the couple. (In our case, the rotation is clockwise.)

3. The beam is not free, because a constraint is imposed on it; i.e., it is rigidly clamped at point B. We make the beam free by replacing the constraint with constraint reaction forces. The clamping reaction consists of a force $\mathbf{R}_B$ and a couple of forces with torque M_B, for both of which neither the magnitude nor the direction is known.

We introduce the reference frame xy shown in the figure. Since the direction of the force $\mathbf{R}_B$ is unknown, we represent it by two components $\mathbf{X}_B$ and $\mathbf{Y}_B$. The unknown couple M_B of reaction forces is represented by the curved arrow. Note that we need not guess the true directions of the components and the curved arrow of the torque of the couple, because all this will be clear after the problem is solved. (In the assumption diagram, $\mathbf{X}_B$ is shown as if it were directed to the left only because this way it is more noticeable in the figure.)

We replace **F** by an equivalent system of forces consisting of two components $\mathbf{F}'$ and $\mathbf{F}''$ parallel to the coordinate axes,

$$\mathbf{F} = \mathbf{F}' + \mathbf{F}'', \quad F' = F\cos\alpha, \quad F'' = F\sin\alpha.$$

4. Now the assumption diagram represents a free beam subjected to forces lying in the same plane. The system of equilibrium equations consists of three equations. The number of unknowns X_B, Y_B, M_B is also equal to three. This means that the number of unknowns and the number of equations coincide and the necessary conditions for the problem to be statically determinate are satisfied.

5. We use the basic form of equilibrium equations for a plane system of forces, taking the center B as the moment point.

The mishap most frequently encountered when writing out the equilibrium equations is to forget some force or couple of forces; therefore, it is recommended to arrange all forces and all couples of forces applied to the body in a row and then write out the equilibrium equations beneath it:

Forces, couples of forces	$\mathbf{X}_B$	$\mathbf{Y}_B$	M_B	$\mathbf{F}'$	$\mathbf{F}''$	M	
$\sum_k F_{kx} =$	$-X_B$	$+\ 0$	$+\ 0$	$-\ F'$	$+\ 0$	$+\ 0$	$= 0,$
$\sum_k F_{ky} =$	0	$+\,Y_B$	$+\ 0$	$+\ 0$	$-\ F''$	$+\ 0$	$= 0,$
$\sum_k M_B(\mathbf{F}_k) =$	0	$+\ 0$	$+\,M_B$	$+\,F'DE$	$-\,F''BD$	$-\,M$	$= 0.$

By substituting the original data, we obtain the solution of the problem:

$$X_B = -17.1\ \text{kN}, \quad Y_B = 10\ \text{kN}, \quad M_B = 4.9\ \text{kN m}.$$

The signs in the answers show that the direction of the component $\mathbf{X}_B$ is opposite to that shown in the assumption diagram, while the directions of the component $\mathbf{Y}_B$ and of the curved arrow M_B correspond to those in the diagram.

It is not recommended to redraw $\mathbf{X}_B$ in the diagram (Fig. E2.24b) with the sign of the answer taken into account, because should this be done, it would be impossible to verify whether the equations of equilibrium are composed correctly and interpret the results.

The clamping reaction force $\mathbf{R}_B$ is the vector sum of the orthogonal components $\mathbf{X}_B$ and $\mathbf{Y}_B$; therefore, its magnitude can be obtained by the formula

$$R_B = \sqrt{X_B^2 + Y_B^2} = 19.8\,\text{kN}.$$

Example 2. A structure consisting of two massless rods BD and DE is shown in Fig. E2.25. The rods are of the same length 4 m and are hinged to each other and to the base so that $BE = BD = 2a$. The left rod BD is loaded by the horizontal force $F = 20$ kN at the midpoint, and the right rod DE is loaded by the couple of forces $M = 10$ kN m. Find the support reactions and the pressure in the intermediate hinge D.

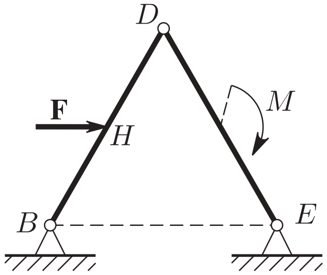

Figure E2.25. Structure consisting of two rods hinged to each other and to the base.

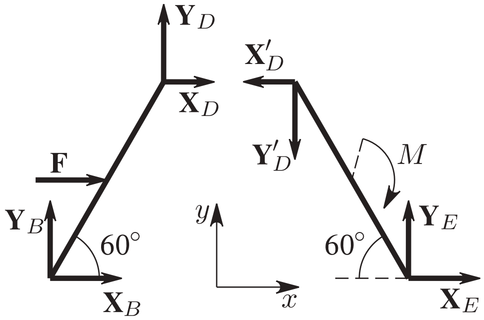

Figure E2.26. Assumption diagram for the two-rod problem.

Solution. We consider the equilibrium of each rod separately, for which we divide the structure into two parts, namely, the rods BD and DE, and align the active forces and the constraint reactions.

After this, the assumption diagram of the problem acquires the form shown in Fig. E2.26, where the reference frame is the same for both rods. The force $\mathbf{R}_D$ acting on the rod BD from the rod ED is represented by two components $\mathbf{X}_D$ and $\mathbf{Y}_D$, and the force $\mathbf{R}'_D$ acting from the rod DB on the rod ED is represented by two components $\mathbf{X}'_D$ and $\mathbf{Y}'_D$. Since the action force $\mathbf{R}_D$ and the reaction force $\mathbf{R}'_D$ must satisfy the action–reaction axiom, it follows that their components also satisfy the relations $\mathbf{X}_D = -\mathbf{X}'_D$ and $\mathbf{Y}_D = -\mathbf{Y}'_D$; i.e., the components are equal and oppositely directed, and this is already taken into account in the diagram. The magnitudes of the components satisfy the relations

$$X_D = X'_D, \quad Y_D = Y'_D. \tag{E2.3.3.1}$$

All in all, there are eight unknowns X_B, Y_B, X_D, Y_D, X'_D, Y'_D, X_E, and Y_E. The equilibrium of each rod is described by three equilibrium equations; together with the last two equations, they form a system of eight equations with eight unknowns. (The necessary conditions for the problem to be statically determinate are satisfied.)

The equilibrium equations read

for rod BD

$$\begin{aligned}
&\textstyle\sum_k F_{kx} = X_B + X_D + F = 0,\\
&\textstyle\sum_k F_{ky} = Y_B + Y_D = 0,\\
&\textstyle\sum_k M_B(\mathbf{F}_k) = -Fa\sin 60^\circ\\
&\qquad - X_D 2a\sin 60^\circ + Y_D 2a\cos 60^\circ = 0,
\end{aligned}$$

for rod ED

$$\begin{aligned}
&\textstyle\sum_k F_{kx} = -X'_D + X_E = 0,\\
&\textstyle\sum_k F_{ky} = -Y'_D + Y_E = 0,\\
&\textstyle\sum_k M_E(\mathbf{F}_k) = X'_D 2a\sin 60^\circ\\
&\qquad + Y'_D 2a\cos 60^\circ - M = 0.
\end{aligned}$$

By solving this system with (E2.3.3.1) taken into account, we obtain $X_B = -16.44$ kN, $Y_B = -11.16$ kN, $X_D = X'_D = -3.56$ kN, $Y_D = Y'_D = 11.16$ kN, $X_E = -3.56$ kN, and $Y_E = 11.16$ kN.

E2.3.4. Center of Parallel Forces. Center of Gravity of a Rigid Body

► **Center of parallel forces.** Assume that a rigid body is under the action of a system $\{\mathbf{F}_k\}$ of parallel forces. The forces can rotate about the points of their application remaining parallel. Since $\mathbf{R}\cdot\mathbf{M}_P = 0$, it follows that the system has a resultant $\mathbf{R}^*$ provided that $\mathbf{R}\neq 0$. One can prove that the line of action of $\mathbf{R}^*$ always passes through the same point C of the body regardless of the force direction. This point is called the *center of parallel forces*, and its position in space is determined by the position vector $\mathbf{r}_C$,

$$\mathbf{r}_C = \frac{\sum_k F_k \mathbf{r}_k}{\sum_k F_k}, \tag{E2.3.4.1}$$

where the $\mathbf{r}_k$ are the position vectors of the points where the forces $\mathbf{F}_k$ are applied and the F_k are the force magnitudes.

The coordinates x_C, y_C, z_C of the center of parallel forces can be determined by the formulas

$$x_C = \frac{\sum_k F_k x_k}{\sum_k F_k}, \quad y_C = \frac{\sum_k F_k y_k}{\sum_k F_k}, \quad z_C = \frac{\sum_k F_k z_k}{\sum_k F_k}. \tag{E2.3.4.2}$$

► **Center of gravity of a body.** The center of gravity of a rigid body is the center of parallel forces that represent elementary forces of gravity of the material particles comprising the body. If the body is on Earth's surface and its dimensions are small compared to Earth's radius, then one can assume that the lines of action of the forces of gravity are parallel and their magnitudes depend only on the body volume V, the material density ρ, and the free fall acceleration g.

The formulas for finding the center of gravity read

$$\mathbf{r}_{\text{cg}} = \frac{1}{G}\int_V \gamma\mathbf{r}\,dV, \quad x_{\text{cg}} = \frac{1}{G}\int_V \gamma x\,dV, \quad y_{\text{cg}} = \frac{1}{G}\int_V \gamma y\,dV, \quad z_{\text{cg}} = \frac{1}{G}\int_V \gamma z\,dV,$$

where $\gamma = \gamma(x,y,z) = g\rho$ is the specific weight and $G = \int_V \gamma\,dV$ is the weight of the body.

Similar formulas hold for the center of gravity of a body that has the shape of a surface or a line.

► **Methods for finding the center of gravity.**

Symmetric bodies. If a body has a plane (axis, center) of material symmetry, then its center of gravity lies in this plane (on the axis, at the center).

Method of partition. Assume that a body consists, for example, of three parts (Fig. E2.27) for each of which we know the weight G_1, G_2, G_3 and the position $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ of the center of gravity. The position vector $\mathbf{r}_{123}$ of the center of gravity of the body consisting of three parts and its coordinates are determined by the formulas

$$\begin{aligned} \mathbf{r}_{123} &= \frac{G_1\mathbf{r}_1 + G_2\mathbf{r}_2 + G_3\mathbf{r}_3}{G_1+G_2+G_3}, \quad & x_{123} &= \frac{G_1x_1+G_2x_2+G_3x_3}{G_1+G_2+G_3},\\ y_{123} &= \frac{G_1y_1+G_2y_2+G_3y_3}{G_1+G_2+G_3}, \quad & z_{123} &= \frac{G_1z_1+G_2z_2+G_3z_3}{G_1+G_2+G_3}. \end{aligned} \tag{E2.3.4.3}$$

Method of negative masses. Now assume that we need to find the position of the center of gravity of a new body consisting of parts *1* and *2* (see Fig. E2.27).

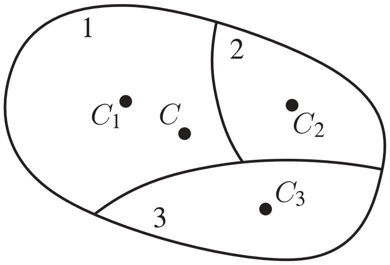

Figure E2.27. Determining the center of gravity of a body consisting of three parts.

The solution can be obtained either by the above formulas if one sets $G_3 = 0$ or by the formulas that, under certain conditions, can be more convenient:

$$\mathbf{r}_{12} = \frac{G\mathbf{r}_{123} - G_3\mathbf{r}_3}{G - G_3}, \quad x_{12} = \frac{Gx_{123} - G_3x_3}{G - G_3},$$
$$y_{12} = \frac{Gy_{123} - G_3y_3}{G - G_3}, \quad z_{12} = \frac{Gz_{123} - G_3z_3}{G - G_3},$$

where $G = G_{123} = G_1 + G_2 + G_3$ is the weight of the body consisting of three parts. The weight G_3 of the cutaway part enters the formulas with negative sign, and just this fact underlies the name of the method.

In Table E4.1 (Section E4.1), we present formulas for the coordinates of the center of gravity for homogeneous bodies of simplest shapes.

E2.3.5. Distributed Forces

In practice, one often encounters the cases in which the body is subjected not to lumped forces but to a load distributed over a volume, surface, or line. The special case of bulk distribution of the force of gravity was considered in the preceding section. The vector quantity $\mathbf{q}$ characterizing the load is called the *load intensity* and is measured in $\mathrm{N/m^3}$, $\mathrm{N/m^2}$, or N/m.

When solving problems of statics, a distributed load is usually replaced by a simpler statically equivalent force (or a system of forces).

1. A *uniformly distributed load* of intensity $\mathbf{q}$ (Fig. E2.28) in a plane problem of statics has the resultant $\mathbf{Q} = \mathbf{q}\,BD$ whose line of action passes through the midpoint of the interval where the load is applied, $BE = ED$.

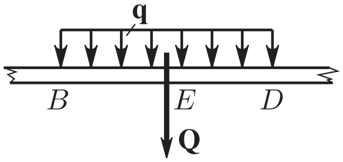

Figure E2.28. Uniformly distributed load.

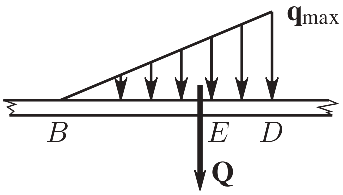

Figure E2.29. Linearly distributed load.

2. A *linearly distributed load* in plane problems of statics (Fig. E2.29) has the resultant $\mathbf{Q} = \frac{1}{2}\mathbf{q}_{\max}\,BD$ passing through point E, and $BE = \frac{2}{3}BD$.

E2.3.6. Friction Laws (Coulomb Laws)

► **Law of sliding friction.** Consider a body of weight G at rest on a rough horizontal plane (Fig. E2.30). An attempt to move the body by applying a horizontal force **F** to it remains unsuccessful until the force attains a certain magnitude $\mathbf{F}^*$. The resultant of the support reaction forces can be represented as the sum of two components, the *normal pressure force* **N** and the *static friction force* $\mathbf{F}_{\text{fr}}$. The equation for the projections of the forces onto the horizontal axis gives $F_{\text{fr}} = F$.

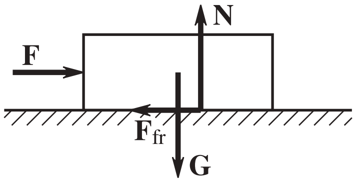

Figure E2.30. Friction force arising in an attempt to move a fixed body on a horizontal plane.

Experiments show that F^* and N satisfy the relation (*Coulomb's dry friction law*)

$$F^* = f_0 N, \qquad \text{(E2.3.6.1)}$$

where f_0 is the *coefficient of static friction*, which depends on the materials of the contacting bodies and the state of their surfaces.

Until $F \le F^*$, the body is at rest. But if a force larger than F^* is applied to it, the body begins to move. In motion, the resistance force can be found by the formula

$$F_{\text{fr}} = fN, \qquad \text{(E2.3.6.2)}$$

where f is the *coefficient of kinetic friction* and F_{fr} is the force of kinetic friction.

Note that the coefficient of kinetic friction is always less than the coefficient of static friction, $f < f_0$.

► **Laws of rolling friction.** Consider a disk of radius R at rest on a nonsmooth horizontal plane (Fig. E2.31). An attempt to roll the disk by applying a horizontal force **F** to its center remains unsuccessful if the force magnitude is less than a certain limit value F^{**}.

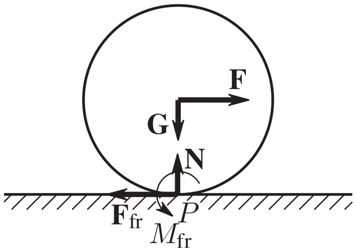

Figure E2.31. Friction force arising in an attempt to roll a circular disk.

According to the theorem on the reduction of a system of forces to a center, the support reaction forces distributed over a small surface near the contact point P can be replaced by an equivalent system, namely, by the normal pressure force **N**, the static friction force $\mathbf{F}_{\text{fr}}$ applied at the contact point P, and a couple of friction forces with torque M_{fr}.

If the disk is in equilibrium, then the equations of moments about the center P imply $FR = M_{\text{fr}}$. Experiments show that F^{**} and N are related by $F^{**} = kN/R$ and

$$M^{**} = kN. \qquad \text{(E2.3.6.3)}$$

The dimensional coefficient k [m] is called the *coefficient of rolling friction*.

A rolling surface is said to be *absolutely rough* if $f \neq 0$ and $k = 0$.

Experiments show that, other conditions being equal, F^* is much larger than F^{**} (usually by two or three orders of magnitude), and hence, in technology, if it is necessary to decrease the friction losses, then one tries to replace sliding by rolling.

E2.4. Dynamics of a Point Particle

E2.4.1. Equations of Motion of a Point Particle

The motion of a point particle with respect to an inertial reference frame is described by Newton's second law:

$$m\mathbf{a} = \sum_{j=1}^{n} \mathbf{F}_j + \sum_{k=1}^{l} \mathbf{R}_k, \qquad \text{(E2.4.1.1)}$$

where m is the mass of the point, $\mathbf{a}$ is its acceleration, and the right-hand side of this equation is the vector sum of all forces applied to the point. The causes of all these forces may be different. Here we distinguish between *active forces* $\mathbf{F}_j$ and *constraint reaction forces* $\mathbf{R}_k$. The active forces can depend on time t and on the position and velocity of the point. The active forces include the forces of gravity, elasticity, viscous friction, aerohydrodynamic drag, etc.

The constraint reaction forces act on a nonfree point particle whose motion is subjected to some mechanical constraints. These forces can only be determined in the course of solution of the problem of dynamics.

We take the Cartesian axes of an inertial reference frame x, y, z, project the vector equality (E2.4.1.1) onto these axes, and obtain

$$\begin{aligned} m\frac{d^2x}{dt^2} &= \sum_j F_{jx} + \sum_k R_{kx}, \\ m\frac{d^2y}{dt^2} &= \sum_j F_{jy} + \sum_k R_{ky}, \\ m\frac{d^2z}{dt^2} &= \sum_j F_{jz} + \sum_k R_{kz}. \end{aligned}$$

These three equations are called the *differential equations of motion of a point particle in Cartesian coordinates*. (From now on, summation indices are sometimes omitted.)

The *differential equations of motion of a point particle in projections onto the natural coordinate axes* have the form

$$\begin{aligned} m\frac{dv}{dt} &= \sum F_{j\tau} + \sum R_{k\tau}, \\ \frac{mv^2}{\rho} &= \sum F_{jn} + \sum R_{kn}, \\ 0 &= \sum F_{jb} + \sum R_{kb}. \end{aligned}$$

Here we have taken into account the fact that $a_\tau = \dfrac{dv}{dt}$, $a_n = \dfrac{v^2}{\rho}$, and $a_b = 0$.

E2.4.2. First and Second Problems of Dynamics

In the equations of motion (see Subsection E2.4.1), unknowns can occur on both left- and right-hand sides. Depending on this, the problems of dynamics are divided into two types, which we consider below.

► **The first problem of dynamics.** The law of motion and the active forces are given, and one needs to find the constraint reaction forces.

Example 1. A load of weight G to which a certain initial velocity was imparted at time $t_0 = 0$ ascends an inclined rough plane (Fig. E2.32). Find the friction force $F_{\rm fr}$ and the normal pressure force N acting on the body under the assumption that the coefficient f of friction of the plane and the angle α of inclination are known.

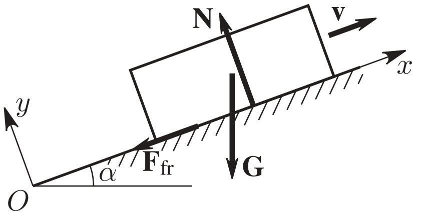

Figure E2.32. Motion of a load on an inclined rough plane.

Solution. We introduce Cartesian coordinate axes and place the origin O at the load position at time $t = 0$. Let us draw the load in an arbitrary position and the forces acting on it. Under the assumption that the load is a point particle, we write out Newton's second law

$$m\mathbf{a} = \mathbf{G} + \mathbf{N} + \mathbf{F}_{\rm fr}.$$

By projecting both sides of this vector equation onto the y-axis, we obtain $0 = -G\cos\alpha + N$ (because the load acceleration is parallel to the x-axis), whence we find $N = G\cos\alpha$. Further, using the Coulomb law, we obtain the friction force $F_{\rm fr} = fN = fG\cos\alpha$.

► **The second problem of dynamics.** The active forces, the mechanical constraint equations, the initial position of the point, and its initial velocity are given, and one needs to find the law of motion of the point and the constraint reaction.

It is recommended to solve the second problem of dynamics successively in several stages listed below.

1. The supposed trajectory of motion is drawn with the point particle shown on it.
2. The forces applied to the point are drawn.
3. Newton's second law is written in vector form.
4. A convenient reference frame is chosen.
5. The equations of motion of the point are written in projections on either the axes of the Cartesian coordinate system or the axes of the natural trihedron. In the first case, it is necessary to express all active forces in terms of t, x, y, z, $\dot{x}$, $\dot{y}$, and $\dot{z}$; in the second case, in terms of t, s, $\dot{s}$.
6. The obtained differential equations are supplemented with the initial conditions, i.e., the values of the coordinates and projections of the velocity of the point at the initial time instant. (They are taken from the conditions of the problem with the introduced reference frame taken into account.)
7. The problem thus posed is solved numerically or analytically.

It is recommended to perform these stages of solution without changing their order.

Example 2. In addition to the conditions of the problem in Example 1, it is assumed that the velocity of the load at time t^* is equal to half the initial velocity. Find the initial velocity v_0 of the load and the path L traveled in time t^*.

Solution. Continuing the process of solving the problem in Example 1, we project both sides of the vector equation onto the x-axis. Using the formulas $G = mg$ and $F_{\rm fr} = fG\cos\alpha$, we obtain

$$\ddot{x} = -g(\sin\alpha + f\cos\alpha).$$

The general solution of the obtained differential equation and the expression for the velocity of the load are given by the formulas

$$x = -\tfrac{1}{2}g(\sin\alpha + f\cos\alpha)t^2 + C_1t + C_2,$$
$$\dot{x} = -g(\sin\alpha + f\cos\alpha)t + C_1.$$

(The details of their derivation are omitted, and their correctness can be verified by differentiation.)

The last two relations must hold at any time instant t and hence at the initial time $t_0 = 0$ and at time t^*; i.e., the relations are satisfied if, instead of t, x, and $\dot{x}$, we first substitute the values 0, 0, v_0 and then the values t^*, L, $v_0/2$. After the substitutions, we obtain the system of four equations

$$\begin{aligned} 0 &= C_2,\\ v_0 &= C_1,\\ L &= -\tfrac{1}{2}g(\sin\alpha + f\cos\alpha)(t^*)^2 + C_1t^* + C_2,\\ \tfrac{1}{2}v_0 &= -g(\sin\alpha + f\cos\alpha)t^* + C_1 \end{aligned}$$

with four unknowns C_1, C_2, v_0, L. By solving this system, we obtain the desired quantities

$$\begin{aligned} v_0 &= 2g(\sin\alpha + f\cos\alpha)t^*,\\ L &= \tfrac{3}{2}g(\sin\alpha + f\cos\alpha)(t^*)^2. \end{aligned}$$

E2.5. General Theorems of Dynamics of a Mechanical System

It is often possible to find important characteristics of motion of a mechanical system without integrating the system of differential equations of motion. This can be done by using general theorems of dynamics.

E2.5.1. Basic Notions and Definitions

► **Internal and external forces.** Any force acting on a point of a mechanical system is necessarily either an active force or a constraint reaction. The entire set of forces acting on the points of the system can also be divided into two classes in a different way: one distinguishes between the *external forces* $\mathbf{F}^{\rm e}$ and the *internal forces* $\mathbf{F}^{\rm i}$. The external forces are the forces acting on the points of the system from points and bodies that are not contained in the system under study. The internal forces are the forces of interaction between the points and bodies contained in the system under study.

This distinction depends on what point particles and bodies are included by the researcher into the mechanical system under study. If the system is extended by including additional points and bodies, then some forces that were external for the original system can become internal for the extended system.

► **Properties of internal forces.** Since these forces are forces of interaction between parts of the system, they are contained in the total system of internal forces in "pairs" formed according to the action–reaction axiom. For each of such "pairs" of forces, the resultant vector and the resultant moment about an arbitrary point are zero. Since the complete system of internal forces consists only of "pairs," it follows that

1. The resultant of the system of internal forces is zero.
2. The resultant moment of the system of internal forces about an arbitrary point is zero.

The *mass m of the system* is the arithmetic sum of masses m_k of all points and bodies that form the system,

$$m = \sum_k m_k.$$

► The **center of mass** (center of inertia) of a mechanical system is the geometric point C whose position vector $\mathbf{r}_C$ and whose coordinates x_C, y_C, z_C are determined by the following formulas similar to the formulas for the center of parallel forces (see Subsection E2.3.4):

$$\mathbf{r}_C = \frac{\sum_k m_k \mathbf{r}_k}{m}, \quad x_C = \frac{\sum_k m_k x_k}{m}, \quad y_C = \frac{\sum_k m_k y_k}{m}, \quad z_C = \frac{\sum_k m_k z_k}{m},$$

where $\mathbf{r}_k$, x_k, y_k, z_k are the position vectors and the coordinates of the points comprising the system.

For a rigid body in a homogeneous field of gravity, the positions of the center of mass and the center of gravity coincide; in other cases, these are different geometric points.

Along with an inertial reference frame, one often considers a noninertial reference frame moving translationally. Its coordinate axes $Cx^*y^*z^*$ (the *König axes*) are chosen so that the origin C constantly coincides with the center of mass of the mechanical system. By definition, the center of mass is stationary in the König axes and is located at the origin.

► **Moments of inertia.** In the space, we choose a plane Π, an axis l, and a point O. The distances from a point of mass m_k to Π, l, and O are denoted by the symbols δ_k, Δ_k, and r_k, respectively. Consider the positive expressions

$$I_\Pi = \sum_k m_k \delta_k^2, \quad I_{ll} = \sum_k m_k \Delta_k^2, \quad I_O = \sum_k m_k r_k^2,$$

where the sum is taken over all the points of the system. These expressions are called the *moments of inertia about the plane* Π*, the axis* l*, and the point* O. In the Cartesian coordinate axes x, y, z with origin at point O, the moments of inertia satisfy the relations

$$\begin{gathered} I_{xx} = I_{yOx} + I_{xOz}, \quad I_{yy} = I_{zOy} + I_{yOx}, \quad I_{zz} = I_{xOz} + I_{zOy}, \\ I_O = I_{xOy} + I_{yOz} + I_{zOx} = \tfrac{1}{2}(I_{xx} + I_{yy} + I_{zz}). \end{gathered}$$

Consider a ray l passing through the origin O with direction cosines α, β, γ (Fig. E2.33). The squared distance Δ_k from the point of mass m_k with coordinates x_k, y_k, z_k to the ray is equal to

$$\Delta^2 = (x^2 + y^2 + z^2)(\alpha^2 + \beta^2 + \gamma^2) - (x\alpha + y\beta + z\gamma)^2.$$

(To be concise, we omit the subscript k.)

The moment of inertia I_{ll} of a material system about the ray is equal to

$$I_{ll} = \sum m\Delta^2 = I_{xx}\alpha^2 + I_{yy}\beta^2 + I_{zz}\gamma^2 - 2I_{yz}\beta\gamma - 2I_{zx}\gamma\alpha - 2I_{xy}\alpha\beta. \qquad \text{(E2.5.1.1)}$$

The quantities

$$I_{yz} = \sum myz, \quad I_{zx} = \sum mzx, \quad I_{xy} = \sum mxy,$$

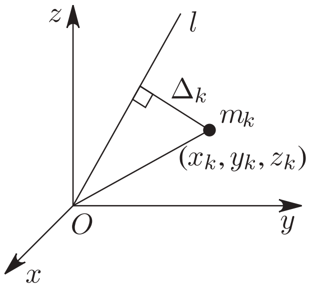

Figure E2.33. Moments of inertia of a point particle with mass m_k about the ray l.

are called the *centrifugal moments of inertia*; they can take both positive and negative values.

On the ray l at a distance $1/\sqrt{I_{ll}}$ from the origin, we mark a point with coordinates $X = \alpha/\sqrt{I_{ll}}$, $Y = \beta/\sqrt{I_{ll}}$, and $Z = \gamma/\sqrt{I_{ll}}$. The equation of the geometric locus of such points with (E2.5.1.1) taken into account has the form

$$I_{xx}X^2 + I_{yy}Y^2 + I_{zz}Z^2 - 2I_{yz}YZ - 2I_{zx}ZX - 2I_{xy}XY = 1.$$

Since the points of the system occupy a bounded domain in space, it follows that the last relation is the equation of an ellipsoid. It is called the *ellipsoid of inertia* of the system with respect to the point O. The principal axes of this ellipsoid are called the *principal axes of inertia* with respect to point O. In these axes, the centrifugal moments of inertia are zero.

If the principal axes of inertia are taken for the coordinate axes, then the expression for the moment of inertia acquires the form

$$I_{ll} = A\alpha^2 + B\beta^2 + C\gamma^2,$$

where $A = I_{xx}$, $B = I_{yy}$, and $C = I_{zz}$.

The principal axes of inertia passing through the center of mass of the system are called the *principal central* axes of inertia.

If a mechanical system is a rigid body, then the summation in the above formulas should be replaced by the integration over the volume V occupied by the body. For example,

$$I_{xOy} = \iiint_V \rho z^2\,dx\,dy\,dz,$$

where $\rho = \rho(x, y, z)$ is the body density.

Table E2.1 presents formulas for calculating the moments of inertia of several figures.

THEOREM (HUYGENS). *The moments of inertia about parallel axes one of which passes through the center of mass C are related by*

$$I_{zz} = I_{Cz} + md^2,$$

where I_{zz} and I_{Cz} are moments of inertia about parallel axes z and Cz, the axis Cz passes through the center of mass, d is the distance between the axes, and m is the mass of the system.

It follows from the Huygens theorem that $I_{Cz} \le I_{zz}$.

TABLE E2.1
The moments of inertia of the simplest homogeneous figures of mass m.

Type of figure	Description of figure	The moment of inertia is calculated about	Moment of inertia, I
Line	Straight line segment of length $2a$	the perpendicular passing through the midpoint of the segment	$\frac{1}{3}ma^2$
Line	Circle arc of radius R supported by angle α	the diameter dividing the arc in half	$\dfrac{1}{2}mR^2\Big(1-\dfrac{\sin\alpha}{\alpha}\Big)$
Surface	Rectangle with sides $2a$ and $2b$	(i) the axis lying in its plane and passing through the center perpendicular to the side $2a$ (ii) the axis perpendicular to the plane of the rectangle and passing through its center	(i) $\frac{1}{3}ma^2$ (ii) $\frac{1}{3}m(a^2+b^2)$
Surface	Ellipse with semiaxes a and b	(i) the axis a (ii) the axis b (iii) the axis perpendicular to the plane of the ellipse and passing through the center	(i) $\frac{1}{4}mb^2$ (ii) $\frac{1}{4}ma^2$ (iii) $\frac{1}{4}m(a^2+b^2)$
Surface	Triangle	the axis lying in its plane and passing through one of the vertices	$\frac{1}{6}m(h_1^2+h_1h_2+h_2^2)$, h_1, h_2 are distances from the other two vertices to the same axis
Surface	Hollow thin-walled cylinder of radius R	the cylinder axis	mR^2
Surface	Thin spherical shell of radius R	the axis passing through the center of the sphere	$\frac{2}{3}mR^2$
Three-dimensional body	Rectangular parallelepiped with sides $2a, 2b, 2c$	the axis passing through the center perpendicular to the face with sides $2a$ and $2b$	$\frac{1}{3}m(a^2+b^2)$
Three-dimensional body	Rectangular pyramid of height h whose base is a rectangle with sides $2a$ and $2b$	(i) the height (ii) the axis passing through the center of gravity and parallel to the side $2a$	(i) $\frac{1}{5}m(a^2+b^2)$ (ii) $\frac{1}{20}m(\frac{3}{4}h^2+4b^2)$
Three-dimensional body	Right circular cylinder of radius R and height h	(i) the cylinder axis (ii) the straight line passing through the center of gravity perpendicular to the cylinder axis	(i) $\frac{1}{2}mR^2$ (ii) $\frac{1}{4}m(\frac{1}{3}h^2+R^2)$
Three-dimensional body	Ball of radius R	the axis passing through the center of the ball	$\frac{2}{5}mR^2$
Three-dimensional body	Triaxial ellipsoid with semiaxes a, b, c	the axis $2a$	$\frac{1}{5}m(b^2+c^2)$
Three-dimensional body	Spherical sector of radius R and height h	the symmetry axis	$\frac{1}{5}m(3Rh-h^2)$

E2.5.2. Theorem on the Motion of the Center of Mass

► Statement of the theorem and some comments.

THEOREM. *The center of mass of a mechanical system moves as a point particle with mass m equal to the mass of the system would move under the action of external forces applied to the system:*

$$m\mathbf{a}_\mathrm{c} = \sum_k \mathbf{F}_k^\mathrm{e}, \tag{E2.5.2.1}$$

where m is the mass of the system and $\mathbf{a}_\mathrm{c}$ is the acceleration of the center of mass.

The mathematical statement of the theorem is similar to that of Newton's second law. Let us explain the statement of the theorem in more detail. In motion of the system, its center of mass moves along a certain trajectory. We assume, for example, that at time t_0 the system is in position B and has a velocity $\mathbf{v}_{\mathrm{c}0}$. Now if a point of mass m is placed in position B at time t_0, the velocity $\mathbf{v}_{\mathrm{c}0}$ is imparted to it, and forces equal to the external forces acting on the system are applied to this point, then the point will move together with the center of mass of the system along the same trajectory with the same velocity and the same acceleration.

The vector equation implies the differential equations of motion of the center of mass in projections onto the axes of the Cartesian coordinate system:

$$m\ddot{x}_\mathrm{c} = \sum_k F_{kx}^\mathrm{e}, \quad m\ddot{y}_\mathrm{c} = \sum_k F_{ky}^\mathrm{e}, \quad m\ddot{z}_\mathrm{c} = \sum_k F_{kz}^\mathrm{e}.$$

► Law of conservation of the velocity of the center of mass of a mechanical system: if the resultant vector of external forces acting on a system is zero, then the center of mass of the system moves at a constant velocity; i.e., if $\sum_k \mathbf{F}_k^\mathrm{e} = 0$, then $\mathbf{v}_\mathrm{c} = \mathrm{const}$.

Note that in this case the velocity vector itself (rather than its magnitude) is constant, and hence the center of mass will move uniformly and rectilinearly.

If the projection of the resultant vector of external forces of the system onto an axis is zero, then the projection of the velocity of the center of mass of the system onto this axis remains constant. For example, if $\sum_k F_{kx} = 0$, then $v_{\mathrm{c}x} = \mathrm{const}$.

Example. Under the action of the force of gravity, a homogeneous disk of mass m_2 rolls down the lateral face of a prism of mass m_1 located on a smooth horizontal plane (Fig. E2.34). The angle between the lateral face and the prism base is α. At the initial time, the velocities of the prism and the disk are zero. Find the distance s_1 traveled by the prism as the disk center travels the distance s_2 along the face.

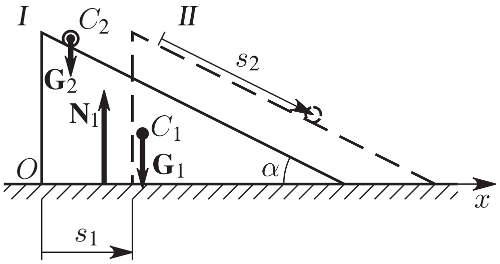

Figure E2.34. Rolling of a disk on the lateral face of a prism.

Solution. Consider the mechanical system consisting of the prism and the disk and draw the external forces, i.e., the active forces of gravity $\mathbf{G}_1$ and $\mathbf{G}_2$ and the reaction force $\mathbf{N}_1$ of the smooth plane. A typical characteristic of the system of external forces is that all of them are perpendicular to the horizontal axis, and hence the sum of their projections onto this axis is zero.

We direct the x-axis horizontally from left to right and place its origin at the point O. In Fig. E2.34, we draw the system in two positions, the initial position I and position II (drawn by dashed lines) of the system at the time by which the disk has traveled the distance s_2 along the prism face. Since the direction and magnitude of the prism displacement are unknown in advance, we put position II on the right of the initial position I without hesitation.

Since $\sum_k F^{\rm e}_{kx} = 0$, it follows from the conservation law for the projection of the velocity of the system center of mass onto the x-axis that $v_{{\rm c}x} = \text{const}$. Since all velocities are zero in position I, we have $v_{{\rm c}x} = 0$. It follows that $x_{\rm c} = \text{const}$, or $x^{\rm I}_{\rm c} = x^{\rm II}_{\rm c}$. We use the formulas for the coordinates of the center of mass to rewrite the last relation as

$$\frac{m_1 x^{\rm I}_1 + m_2 x^{\rm I}_2}{m_1 + m_2} = \frac{m_1 x^{\rm II}_1 + m_2 x^{\rm II}_2}{m_1 + m_2}.$$

Here $x^{\rm I}_1$ and $x^{\rm I}_2$ are the coordinates of the centers of mass of the prism and the disk in position I, and $x^{\rm II}_1$ and $x^{\rm II}_2$ are the respective quantities in position II.

It follows from Fig. E2.34 that $x^{\rm II}_1 = x^{\rm I}_1 + s_1$ and $x^{\rm II}_2 = x^{\rm I}_2 + s_1 + s_2 \cos\alpha$. By substituting these relations into the formula and by performing algebraic manipulations, we obtain $s_1 = -\dfrac{m_2 s_2 \cos\alpha}{m_1 + m_2}$.

The sign of the answer shows that the prism displacement is opposite to that shown in the figure.

E2.5.3. Theorem on the Momentum

► **Momentum.** The *momentum of a point* of mass m moving at a velocity $\mathbf{v}$ is the vector

$$\mathbf{Q} = m\mathbf{v}.$$

The *momentum of a mechanical system* is the resultant of momenta of all points of the system,

$$\mathbf{Q} = \sum_k \mathbf{Q}_k.$$

One can prove that the momentum of a system is equal to the momentum of a fictitious point particle with mass equal to the mass of the system and velocity equal to the velocity of the center of mass, $\mathbf{Q} = m\mathbf{v}_{\rm c}$.

► **Impulse of force.** Suppose that a force $\mathbf{F}$ is applied to a moving point particle. (Along with this force, any other forces can act on the particle, but for now we consider only one of them.)

The *elementary impulse of force* $\mathbf{F}$ in an elementary time interval dt is the vector

$$d\mathbf{S} = \mathbf{F}\,dt.$$

The *impulse of force* $\mathbf{F}$ in a finite time interval from t_0 to t is the vector

$$\mathbf{S} = \int_{t_0}^{t} d\mathbf{S} = \int_{t_0}^{t} \mathbf{F}\,dt.$$

The projections of the impulse of force onto the coordinate axes can be calculated by the formulas

$$S_x = \int_{t_0}^{t} F_x\,dt, \quad S_y = \int_{t_0}^{t} F_y\,dt, \quad S_z = \int_{t_0}^{t} F_z\,dt.$$

► Various statements of the theorem on the momentum.

THEOREM ON THE MOMENTUM IN DIFFERENTIAL FORM. *The time derivative of the momentum of a mechanical system is equal to the resultant vector of external forces acting on the system,*

$$\frac{d\mathbf{Q}}{dt}=\sum_k \mathbf{F}_k^{\mathrm{e}};\qquad (E2.5.3.1)$$

in other words, the differential of the momentum of a system is equal to the vector sum of elementary impulses of all external forces acting on the points of the system,

$$d\mathbf{Q}=\sum_k d\mathbf{S}_k^{\mathrm{e}}.$$

In projections onto the coordinate axes (e.g., onto the x-axis), we obtain

$$\frac{dQ_x}{dt}=\sum_k F_{kx}^{\mathrm{e}}\quad\text{or}\quad dQ_x=\sum_k dS_{kx}^{\mathrm{e}}.$$

THEOREM ON THE MOMENTUM IN INTEGRAL FORM. *The variation in the momentum of a mechanical system in a time interval is equal to the sum of impulses of external forces acting on the system in the same time interval,*

$$\mathbf{Q}-\mathbf{Q}_0=\sum_k \mathbf{S}_k^{\mathrm{e}},\qquad (E2.5.3.2)$$

where the vectors $\mathbf{Q}$ *and* $\mathbf{Q}_0$ *correspond to time instants* t *and* t_0 *and the* $\mathbf{S}_k^{\mathrm{e}}$ *are the impulses of external forces acting on the system in the time interval from* t_0 *to* t.

In projections onto the coordinate axes (e.g., onto the x-axis), we have

$$Q_x-Q_{x0}=\sum_k S_{kx}^{\mathrm{e}}.$$

If the theorem is used in the case of a single point particle, one should remember that any force applied to the point is external.

► The momentum conservation law.

The law of conservation of the momentum of a mechanical system has the form

$$\text{if}\quad \textstyle\sum_k \mathbf{F}_k^{\mathrm{e}}=0,\quad\text{then}\quad \mathbf{Q}=\text{const}.$$

The law of conservation of the momentum projection onto any axis (for example, onto the x-axis) has the form

$$\text{if}\quad \textstyle\sum_k F_{kx}^{\mathrm{e}}=0,\quad\text{then}\quad Q_x=\text{const}.$$

Example. On a smooth horizontal plane, a rectangular parallelepiped of mass m_1 moves at a velocity v_0, and on the upper face of this parallelepiped, there are two propulsion devices with masses m_2 and m_3 (Fig. E2.35). At some time instant, the devices begin to move towards each other, and the laws of their motion with respect to the parallelepiped are determined by the functions $s_2=h_2t^3$ and $s_3=h_3t^3$. Find how the velocity of the parallelepiped depends on time.

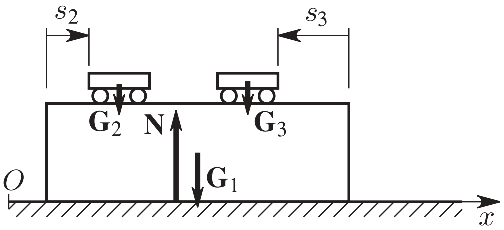

Figure E2.35. Motion of two propulsion devices on the surface of a moving rectangular parallelepiped.

Solution. Consider the system consisting of the three bodies (the parallelepiped and the propulsion devices). Let us draw the external forces, namely, the active forces of gravity $\mathbf{G}_1$, $\mathbf{G}_2$, and $\mathbf{G}_3$ and the plane reaction $\mathbf{N}$. All of them are vertical, and hence we can apply the law of conservation of the system momentum projection onto the x-axis; i.e., we write out Q_{x0} at time t_0 and Q_x at an arbitrary time t and equate the obtained quantities with each other:

$$Q_{x0} = m_1 v_0 + m_2 v_0 + m_3 v_0,$$
$$Q_x = m_1 v_1 + m_2(v_1 + \dot{s}_2) + m_3(v_1 - \dot{s}_3).$$

In the last expression, the velocities of the propulsion devices were calculated according to the theorem on the velocities in the compound motion of a point. By equating the right-hand sides of these equations and by performing algebraic manipulations, we obtain

$$v_1 = v_0 - \frac{3t^2(m_2 h_2 - m_3 h_3)}{m_1 + m_2 + m_3}.$$

E2.5.4. Theorem on the Angular Momentum

► **Moment of momentum and the angular momentum.** The *moment of momentum of a point particle about a fixed center P* is the vector $\mathbf{K}_P$ equal to the cross product of the position vector connecting the center with the point by the momentum of the point:

$$\mathbf{K}_P = \mathbf{M}_P(\mathbf{Q}) = \mathbf{r} \times m\mathbf{v}.$$

The *angular momentum (the net moment of momentum) of a mechanical system about the center P* is the vector sum of moments of momenta of all points of the system about the center:

$$\mathbf{K}_P = \sum_j \mathbf{K}_{Pj}.$$

The *moment of momentum of a point about the x-axis* is the quantity K_x equal to the projection onto this axis of the moment of momentum of the point about any center P belonging to the axis,

$$K_x = M_x(\mathbf{Q}).$$

The *angular momentum of a system about the x-axis* is the projection onto this axis of the angular momentum of the system about any center P belonging to the axis:

$$K_x = \sum_j K_{jx}.$$

The analytic expression for the angular momentum of a system about the coordinate x-axis has the form

$$K_x = \sum_j m_j(y_j \dot{z}_j - \dot{y}_j z_j).$$

The formulas for K_y and K_z are similar.

One can show that the angular momentum of a system about a fixed center P is equal to the sum of the moment of momentum of the center of mass about the center P and the angular momentum of the system about the center of mass C in its relative motion in the König system,

$$\mathbf{K}_P = \mathbf{K}_P + \mathbf{K}_C^*.$$

Here $\mathbf{K}_P = \mathbf{M}_P(m\mathbf{v}_C)$ and $\mathbf{K}_C^*$ is the angular momentum of the system in its motion with respect to the König reference frame.

The angular momentum K_z of a rigid body rotating about the fixed z-axis with angular velocity ω is calculated by the formula

$$K_z = I_{zz}\omega,$$

where I_{zz} is the moment of inertia of the rigid body about the z-axis.

The angular momentum of a rigid body is discussed in more detail in Section E2.8.

► Theorem on angular momentum of a mechanical system.

THEOREM. *The time derivative of the angular momentum about any fixed center P is equal to the resultant moment of external forces of the system about the same center,*

$$\frac{d\mathbf{K}_P}{dt} = \sum_j \mathbf{M}_P(\mathbf{F}_j^{\rm e}). \tag{E2.5.4.1}$$

Projecting the vector expression onto the coordinate axes (e.g., on the x-axis), we obtain the theorem on the angular momentum of the system about a fixed axis:

$$\frac{dK_x}{dt} = \sum_j M_x(\mathbf{F}_j^{\rm e}).$$

If this theorem is used to study the motion of a rigid body rotating about a fixed axis z, then we obtain the *differential equation of rotational motion of a rigid body about a fixed axis*:

$$I_{zz}\ddot{\varphi} = \sum_j M_z(\mathbf{F}_j^{\rm e}),$$

where φ is the rotation angle.

► **Law of conservation of angular momentum.** The law is stated as follows: if the resultant moment of external forces acting on a system about a center P is zero, then the net moment of momentum about this center is constant. For example,

$$\text{if} \quad \textstyle\sum_j \mathbf{M}_P(\mathbf{F}_j^{\rm e}) = \mathbf{0}, \quad \text{then} \quad \mathbf{K}_P = \text{const}.$$

The right-hand side of this relation contains a vector constant; i.e., neither the magnitude nor the direction of the vector depends on time.

If the sum of moments of external forces acting on the system about some fixed axis is zero, then the angular momentum of the system about this axis remains constant. For example,

$$\text{if} \quad \textstyle\sum_j M_x(\mathbf{F}_j^{\rm e}) = 0, \quad \text{then} \quad K_x = \overrightarrow{\text{const}}.$$

Example 1. A point particle moves under the action of a system of forces $\{\mathbf{F}_k\}$ so that the line of action of the resultant $\mathbf{R}^*$ passes through a fixed center O. (We take it for the origin of the reference frame.) Let us find several characteristics of motion of the point.

Since $\sum_k \mathbf{M}_O(\mathbf{F}_k) = \mathbf{M}_O(\mathbf{R}^*) = \mathbf{0}$, it follows that $\mathbf{K}_O = \overrightarrow{\text{const}}$, or $\mathbf{r}\times m\mathbf{v} = \overrightarrow{\text{const}}$.

Taking the scalar products of both sides of the last relation by $\mathbf{r}$, we obtain $\mathbf{r}\cdot\overrightarrow{\text{const}} = 0$, or $C_1x + C_2y + C_3z = 0$, where x, y, z are the coordinates of the moving point and C_1, C_2, C_3 are the coordinates of the vector constant.

The last expression is the equation of a plane passing through the center O, and this means that the trajectory of the point lies in that plane.

Example 2. A massless string with a load of weight G_2 at the end is wound on a homogeneous drum of weight G_1 and radius R (Fig. E2.36). Find the acceleration of the load neglecting the friction forces in the rotation of the drum and assuming that the string unwinds from the drum without slip.

Solution. Let the system consist of the drum, the load, and the string. We draw the external forces, namely, the active forces of gravity $\mathbf{G}_1$ and $\mathbf{G}_2$ and the reaction force $\mathbf{N}_O$ passing through the rotation axis O. The direction of the force $\mathbf{N}_O$ is unknown in advance, therefore we draw it arbitrarily. There is no couple of friction forces on the axis, which follows from the assumptions of the problem.

We apply the theorem on the variations in the angular momentum of a system about the drum rotation axis:

$$\frac{dK_O}{dt} = \sum_j M_O(\mathbf{F}_j^{\rm e}).$$

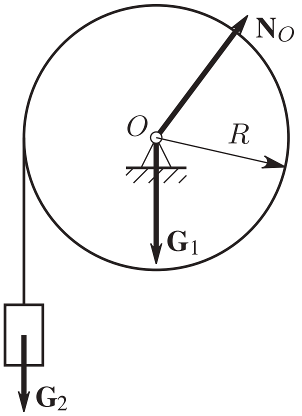

Figure E2.36. Unwinding of a massless string with a load from a rotating drum.

We calculate K_O as the sum of the angular momenta of the drum and of the load, take into account the equation $v_2 = \omega_1 R$, and obtain

$$K_O = v_2 R\left(\tfrac{1}{2}m_1 + M_2\right).$$

We calculate the sum of moments of external forces about the axis:

$$\sum_j M_O(\mathbf{F}_j^{\mathrm{e}}) = m_2 g R.$$

Substituting the right-hand sides of the last formulas into the statement of the theorem, we obtain

$$a_2 = \frac{2m_2}{m_1 + 2m_2} g.$$

E2.5.5. Theorem on the Kinetic Energy

► **Elementary work.** Consider a point B moving under the action of a system of forces. A small displacement of the point along the trajectory is characterized by the vector $d\mathbf{r}$ (Fig. E2.37). We distinguish one force $\mathbf{F}$ in the system.

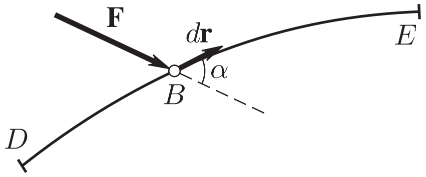

Figure E2.37. Elementary work of the force **F** on the displacement $d\mathbf{r}$.

The *elementary work of the force* **F** on the displacement $d\mathbf{r}$ is the scalar quantity $d'A$ equal to the scalar product of the vectors **F** and $d\mathbf{r}$:

$$d'A = \mathbf{F}\cdot d\mathbf{r} = F\,dr\cos\alpha.$$

In coordinate form, the elementary work is calculated by the formula

$$d'A = F_x\,dx + F_y\,dy + F_z\,dz,$$

where F_x, F_y, F_z and dx, dy, dz are the coordinates of the vectors **F** and $d\mathbf{r}$, respectively, in the Cartesian coordinate system.

Note that the elementary work $d'A$ need not be the total differential of a function depending on the coordinates. (This fact is also reflected in the above representation.)

The sign of the elementary work is determined by the cosine of the angle α: it is positive for $0 \le \alpha < \pi/2$, negative for $\pi/2 < \alpha \le \pi$, and zero for $\alpha = \pi/2$.

► **Calculation of elementary work in special cases.**

1. The elementary work of a force applied to a rigid body rotating about a fixed axis z is calculated by the formula

$$d'A = \pm M_z(\mathbf{F})\,d\varphi.$$

2. The sum of elementary works of a couple of forces applied to a body in arbitrary motion is calculated by the formula $d'A = (\mathbf{M}\cdot\boldsymbol{\omega})\,dt$, and in the cases of rotation about a fixed axis and of plane motion, it can be calculated as follows:

$$d'A = \pm M\,d\varphi.$$

Here M is the torque of a couple of forces, and $d\varphi$ is the angle of rotation of the body. The plus sign is taken if the sense of the curved arrows of the couple torque coincides with the sense of rotation, and the minus sign is taken if they are opposite. (It is assumed that the plane in which the couple acts is parallel to the base plane.)

3. When calculating the elementary work of friction forces applied to a body rolling without slip, it is necessary to take into account the fact that the following forces act at the point of tangency P (Fig. E2.31): the normal pressure force $\mathbf{N}$, the friction force $\mathbf{F}_{\text{fr}}$, and the couple of rolling friction forces with torque $M_{\text{fr}} = Nk$. Since there is no slip, so that the point of tangency is the instantaneous center of velocities and its velocity $\mathbf{v}_P$ is zero, it follows that $d\mathbf{r} = \mathbf{v}_P\,dt = 0$, which implies that

$$d'A_N = dA_{\text{fr}} = 0, \quad d'A_M = -M_{\text{fr}}\,d\varphi = -Nk\,d\varphi.$$

In rolling on an ideally rough surface, $d'A_M = 0$, because $M_{\text{fr}} = 0$.

4. One can prove that the sum of elementary works of forces applied to a rigid body is equal to the sum of elementary works of a statically equivalent system of forces. By the theorem on reducing a system of forces to a given center, an arbitrary system of forces can be replaced by an equivalent system consisting of a force $\mathbf{R}$ applied at an a priori chosen point P and a couple of forces with torque $\mathbf{M}_P$. Therefore, instead of cumbersome calculations of the sum of elementary works of many forces applied to a body, one usually calculates the sum of elementary works of a single force and a single couple.

Example 1. A system of elementary forces of gravity acting on a rigid body always has the resultant equal to the body weight G applied at the center of gravity C. Therefore, the sum of elementary works of the forces of gravity is equal to the elementary work of the force of weight on the displacement of the center of gravity of the body.

Example 2. The sum of elementary works of internal forces applied to points of a rigid body is zero, because the resultant vector and the resultant moment of the system of internal forces are zero.

► **Work of a force. Potential force.** The *work of a force* $\mathbf{F}$ on a finite displacement of a point along a trajectory DE (Fig. E2.37) is given by the curvilinear integral

$$A = \int_{DE} d'A = \int_{DE} \mathbf{F}\cdot d\mathbf{r} = \int_{DE} F_x\,dx + F_y\,dy + F_z\,dz.$$

A *force field* is a part of space where a point particle is under the action of a force $\mathbf{F}$ depending on the position of the point and the time, $\mathbf{F} = \mathbf{F}(\mathbf{r}, t)$, where $\mathbf{r}$ is the position vector of the point. A force field is said to be *nonstationary* (transient) if the force explicitly depends on time and *stationary* if the force is independent of time.

In what follows, we deal only with stationary force fields, where $\mathbf{F} = \mathbf{F}(\mathbf{r})$. In this case, the projections of the force on the coordinate axes are functions of the coordinates of the point, $F_x = F_x(x,y,z)$, $F_y = F_y(x,y,z)$, and $F_z = F_z(x,y,z)$.

In the general case, the work of a force depends on the shape of the trajectory along which the point moves. But there are force fields in which the work depends only on the endpoints of the trajectory. One can show that in these cases the elementary work is the total differential of a function, $d'A = dU(x, y, z)$, and

$$F_x = \frac{\partial U}{\partial x}, \quad F_y = \frac{\partial U}{\partial y}, \quad F_z = \frac{\partial U}{\partial z}.$$

Such a force field is said to be *potential*, and the function U is called the *force function* or the *potential* of the field.

Examples of potential forces are as follows:

1. The force of gravity G; the expression for its work is given by the formula $A = G\,\Delta H$, where G is the weight of a point particle and ΔH is the height difference between the initial and final points of the trajectory.

2. The expression for the sum of works of forces applied to the ends of a weightless linearly elastic spring of rigidity γ is given by the formula $A = -\gamma(\lambda_e^2 - \lambda_s^2)/2$, where L is the spring current length, l is its length in the unstrained state, $\lambda = L - l$ is the length variation, λ_e corresponds to the terminal state of the spring, and λ_s corresponds to the initial state.

The function $\Pi(x, y, z) = -U + \text{const}$ is called the *potential energy* of the field, and the function $E = T + \Pi$ is called the *total mechanical energy*. The potential energy of a point particle is equal to the work of forces of the potential field in the transition from the current state of the point into the zero state. For the zero state we can take any state because of the constant contained in the definition.

► **Kinetic energy.** The *kinetic energy of a point* of mass m moving at a velocity **v** is the scalar quantity T determined by the formula

$$T = \frac{mv^2}{2}.$$

The *kinetic energy of a mechanical system* is the sum of the kinetic energies of all its points:

$$T = \sum_k T_k = \sum_k \frac{m_k v_k^2}{2}.$$

One can prove that the kinetic energy of a system is equal to the sum of the kinetic energy of the center of mass and the kinetic energy of the system in its relative motion in the König reference frame:

$$T = \frac{mv_C^2}{2} + \sum_k \frac{m_k (v_k^*)^2}{2} = T_C + T^*,$$

where $m = \sum_k m_k$ and the v_k^* are the relative velocities of the points.

The formulas for the kinetic energy of a rigid body in the simplest motions are as follows:

(a) $T = \frac{1}{2} m v_C^2$ in translational motion.

(b) $T = \frac{1}{2} I_{zz}\omega^2$ in rotation about the fixed z-axis.

(c)

$$T = \frac{1}{2} m v_C^2 + \frac{1}{2} I_{CC}\omega^2 \qquad \text{(E2.5.5.1)}$$

in plane-parallel motion, where I_{CC} is the moment of inertia of the body about the axis perpendicular to the base plane and passing through the center of mass C.

The kinetic energy of a rigid body is discussed in more detail in Section E2.8.

► Various statements of the theorem on the kinetic energy.

THEOREM ON THE KINETIC ENERGY IN DIFFERENTIAL FORM. *The differential of the kinetic energy of a mechanical system is equal to the sum of elementary works of the forces applied to the points of the system on the elementary displacements of these points:*

$$dT = \sum d'A_j. \tag{E2.5.5.2}$$

THEOREM ON THE KINETIC ENERGY IN INTEGRAL FORM. *The variation in the kinetic energy of a mechanical system on a certain displacement is equal to the sum of works of all forces applied to the points of the system on the displacements of these points:*

$$T - T_0 = \sum A_j.$$

Remark. In contrast to the three above-considered theorems on the dynamics of a system, the last theorem is characterized by the following specific features:

(1) The theorem on the variation in the kinetic energy relates scalar quantities rather than vector quantities.

(2) Both statements of the theorem contain works of all forces, not only external but also internal. (It is also possible to divide the sum of works into the sum of works of active forces and constraint reaction forces.)

(3) The sum of works of internal forces applied to the points of a rigid body is zero.

Example 3. A homogeneous disk under the action of the force of weight G rolls down without slip on an ideally rough plane inclined at an angle α to the horizon (Fig. E2.38). Find the acceleration of the disk center, the friction force magnitude, and the minimum value f^* of the coefficient of friction at which a rolling motion without slip is possible.

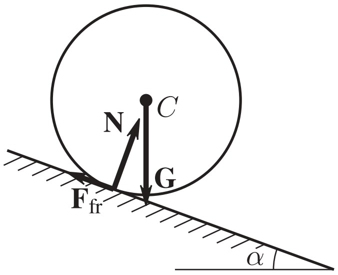

Figure E2.38. Rolling of a disk down an inclined plane under the action of the force of weight.

Solution. For the mechanical system we take the disk itself and study the motion by using the theorem on the kinetic energy in differential form. Calculating the kinetic energy of the disk, we obtain

$$T = \tfrac{1}{2}mv_C^2 + \tfrac{1}{2}I_{CC}\omega^2 = \tfrac{3}{4}mv_C^2.$$

Here we have taken into account the fact that $I_{CC} = \frac{1}{2}mr^2$ and $v_C = \omega r$. Since the system under study is a rigid body, it follows that $\sum d'A^{\rm i} = 0$. We calculate the sum of elementary works of external forces, substitute the obtained expressions for T and $\sum d'A^{\rm e}$ into the statement of the theorem, and obtain

$$d\bigl(\tfrac{3}{4}3mv_C^2\bigr) = mgr\sin\alpha\,d\varphi.$$

Dividing both sides of this relation by dt and taking into account the kinematic relations $\dot v_C = a_C$ and $\dot\varphi = \omega$, we obtain

$$a_C = \tfrac{2}{3}g\sin\alpha.$$

Let us write out the statement of the theorem on the motion of the center of mass of a system as applied to the problem under study:

$$m\mathbf{a}_C = \mathbf{G} + \mathbf{N} + \mathbf{F}_{\rm fr}.$$

We project the obtained vector equation onto two perpendicular axes the first of which is parallel to the normal reaction force **N**:

$$0 = -G\cos\alpha + N,$$
$$ma_C = G\sin\alpha - F_{\rm fr}.$$

Taking into account the value a_C obtained earlier, we obtain the solution of the system of equations: $F_{\rm fr} = \frac{1}{3}G\sin\alpha$, $N = G\cos\alpha$. By substituting these quantities into the Coulomb friction law, we determine the minimum value f^* of the coefficient of friction for which a rolling motion without slip is possible: $f^* = \frac{1}{3}{\rm tg}\alpha$.

A mechanical system whose points are only subjected to potential forces is said to be *conservative*.

The *law of conservation of the total mechanical energy* states that the total mechanical energy of a conservative system remains constant, $E = T + \Pi = T_0 + \Pi_0 = \text{const}$.

E2.6. Elements of Analytical Mechanics

E2.6.1. D'Alembert's Principle

► **Forces of inertia.** Assume that a point of a material system moves under the action of some system of forces (these forces can be divided either into external and internal forces or into active forces and constraint reaction forces). The resultant of this system of converging forces will be denoted by **F**.

The *force of inertia* of a point is the vector $\mathbf{\Phi}$ equal and opposite to the product of the mass of the point by its acceleration,

$$\mathbf{\Phi} = -m\mathbf{a}.$$

The force $\mathbf{\Phi}$ is fictitious; it is not an actual force acting on the point.

D'Alembert's principle: in the motion of a mechanical system (point), any of its states can be considered as an equilibrium if the real forces acting on each point of the system are supplemented with the fictitious forces of inertia.

According to this principle, if each point of the system is supplemented with the force $\mathbf{\Phi}_j = -m_j\mathbf{a}_j$, then the system of forces consisting of the real forces $\mathbf{F}_j$ and the fictitious forces $\mathbf{\Phi}_j$ satisfies all the equations of statics; i.e., the resultant vector of the system of forces and its resultant moment about an arbitrary center P are zero:

$$\begin{aligned} &\sum \mathbf{F}_j + \sum \mathbf{\Phi}_j = 0,\\ &\sum \mathbf{M}_P(\mathbf{F}_j) + \sum \mathbf{M}_P(\mathbf{\Phi}_j) = 0. \end{aligned}$$

In coordinate form, these equations can be written as

$$\begin{aligned} &\sum F_{jx} + \sum \Phi_{jx} = 0, &&\sum M_x(\mathbf{F}_j) + \sum M_x(\mathbf{\Phi}_j) = 0,\\ &\sum F_{jy} + \sum \Phi_{jy} = 0, &&\sum M_y(\mathbf{F}_j) + \sum M_y(\mathbf{\Phi}_j) = 0,\\ &\sum F_{jz} + \sum \Phi_{jz} = 0, &&\sum M_z(\mathbf{F}_j) + \sum M_z(\mathbf{\Phi}_j) = 0. \end{aligned}$$

D'Alembert's principle allows one to transfer the methods for solving problems of statics to problems of dynamics.

Example 1. A load of weight G is in an elevator cabin ascending in decelerated motion with acceleration a (Fig. E2.39). Find the pressure exerted by the elevator cabin floor on the load.

Solution. We assume that the load is a point particle and draw the real forces acting on it: the active force of weight **G** and the force **N** of floor pressure on the load. We supplement these forces with the fictitious force of inertia $\mathbf{\Phi} = -m\mathbf{a}$. (Note that the force $\mathbf{\Phi}$ in the figure is opposite not to the elevator displacement but to the acceleration vector.)

The obtained system of three forces **G**, **N**, and $\mathbf{\Phi}$ is in equilibrium according to d'Alembert's principle. The lines of action of all the forces are directed along a single straight line, and hence the equilibrium of the system of forces is described by the single equation

$$-G + N + \Phi = 0.$$

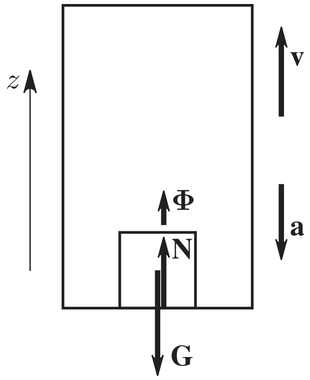

Figure E2.39. Ascent of an elevator with a load.

We substitute the magnitude $ma = \frac{G}{g}a$ of the force of inertia (where g is the free fall acceleration) for Φ and obtain $N = G(g - a)$. One can see that the floor pressure force is less than the weight of the load.

The system of forces of inertia can be very cumbersome in the case of numerous point particles or distributed masses. Using the theorem of statics on reducing a system of forces to a center, one can replace the system of forces $\mathbf{\Phi}_j$ of inertia by an equivalent system consisting of a single force $\mathbf{\Phi}$ applied at a center P given in advance (it is equal to the *resultant vector of forces of inertia*, $\mathbf{\Phi} = \sum \mathbf{\Phi}_j$, and is independent of the choice of the center) and by a single couple of forces whose torque $\mathbf{M}_P^\Phi$ is equal to the *net moment of forces of inertia* about the center, $\mathbf{M}_P^\Phi = \sum \mathbf{M}_P(\mathbf{\Phi}_j)$.

One can show that $\mathbf{\Phi}$ is calculated by the formula

$$\mathbf{\Phi} = -m\mathbf{a}_C,$$

where m is the mass of the system and $\mathbf{a}_C$ is the acceleration of the center of mass.

► **Formulas for the moment of forces of inertia of a body.** Some useful formulas for the net moment of forces of inertia of a rigid body and its projections onto the coordinate axes are given below:

1. $\mathbf{M}_C^\Phi = 0$ in translational motion.
2. $M_z^\Phi = -I_{zz}\varepsilon$ in rotation about the fixed z-axis.
3. $M_C^\Phi = -I_{CC}\varepsilon$ in plane-parallel motion.

Here ε is the angular acceleration of the body, and I_{zz} and I_{CC} are the moments of inertia of the body about the z-axis and the axis passing through the center of mass and perpendicular to the base plane. (The minus signs in the formula mean that the senses of the angular acceleration and the torque of the couple of forces of inertia are opposite.)

Example 2. A homogeneous disk of radius r rolls without slip up a circular arc of radius R (Fig. E2.40). The coefficient of rolling friction is k. Find the acceleration of the disk center and the force of disk pressure on the support at the time instant when the velocity of the disk center is v_0 and the angle between the vertical line and the straight line connecting the centers of the disk and of the arc is α.

Solution. The acceleration of the disk center consists of two components a_C^τ and a_C^n, where the direction of a_C^τ is unknown in advance and $a_C^n = v_0^2/(R-r)$. In the no-slip case, we have $a_C^\tau = \varepsilon r$, where ε is the angular acceleration of the disk.

We reduce the forces of inertia of the disk to the center of mass and decompose the force of inertia into two components, $\mathbf{\Phi} = \mathbf{\Phi}^\tau + \mathbf{\Phi}^n$, where $\mathbf{\Phi}^\tau = -m\mathbf{a}_C^\tau$ and $\mathbf{\Phi}^n = -m\mathbf{a}_C^n$. The torque of the couple of forces of inertia has the magnitude $M_C^\Phi = I_{cc}\varepsilon = mra_C^\tau/2$, and the corresponding curved arrow is opposite to the curved arrow of the assumed angular acceleration.

The system of forces in Fig. E2.40 is in equilibrium because of d'Alembert's principle. Let us write out the equations of equilibrium for a plane system of forces:

$$F_{\text{fr}} + \Phi^\tau - G\sin\alpha = 0,$$

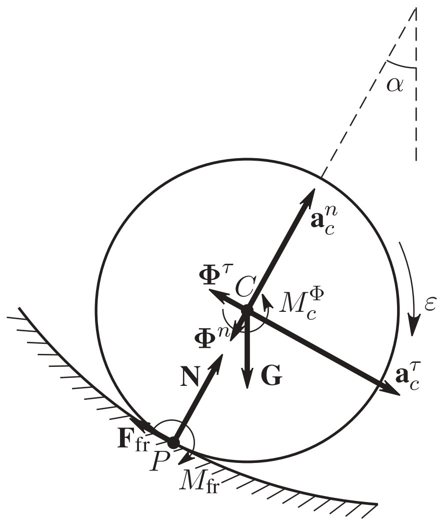

Figure E2.40. A disk rolling on a circular arc.

$$N - \Phi^n - G\cos\alpha = 0,$$
$$-M_{\text{fr}} + M_C^\Phi + \Phi^\tau r - Gr\sin\alpha = 0.$$

Here the first of the equations is the equation for the projections onto the direction of $\mathbf{F}_{\text{fr}}$, the second is the equation for the projections onto the direction of $\mathbf{N}$, and the last is the equation of moments about the center P.

We supplement these equations with Coulomb's law $M_{\text{fr}} = kN$ (see Subsection E2.3.6) and obtain the definitive system of four equations with four unknowns a_C^τ, N, F_{fr}, and M_{fr}. By solving this system, we obtain

$$a_C^\tau = \frac{2}{3r}\left[gr\sin\alpha + k\left(g\cos\alpha + \frac{v_0^2}{R-r}\right)\right], \quad N = m\left(g\cos\alpha + \frac{v_0^2}{R-r}\right).$$

E2.6.2. Classification of Mechanical Constraints. Generalized Coordinates

► **Classification of mechanical constraints.** *Mechanical constraints* are devices (bodies) imposing restrictions on the positions and velocities of points of a mechanical system. These restrictions are always satisfied regardless of the given forces and are written as relations called the *constraint equations*.

Stationary constraints are constraints independent of time; constraints depending on time are said to be *nonstationary*.

Constraints whose equations contain the coordinates of points and time are said to be *geometric*; constraints are said to be *kinematic* (*differential*) if the constraint equations contain not only the coordinates and time, but also the velocities of points.

If a kinematic constraint can be represented as an equivalent geometric constraint, then it is said to be *kinematic integrable*. Otherwise, if a constraint equation cannot be represented as a geometric constraint in principle, then it is called a *nonintegrable constraint*.

Geometric and kinematic integrable constraints are said to be *holonomic*, and kinematic nonintegrable constraints are said to be *nonholonomic*. A mechanical system is said to be *holonomic* if only holonomic constraints are imposed on it and *nonholonomic* if there is at least one nonholonomic constraint.

Constraints are said to be *bilateral* if the restrictions imposed by them on the positions of points, their velocity, and time, can be written as equalities. *Unilateral constraints* are written as inequalities.

► **Virtual displacement of a point.** A *virtual displacement of a point* of a mechanical system is any displacement $\delta\mathbf{r}$ admitted by the imposed constraints from the position occupied by the point at a given time instant. (When constructing such displacements, one should "freeze" the time, so that the nonstationary constraints become fixed, i.e., stationary.) The point does not actually perform virtual displacements but could do so without violating the constraints at a given time instant.

A *virtual displacement of a system* is an arbitrary set of virtual displacements $\delta\mathbf{r}_j$ of the points of the system admitted by all constraints imposed on it. By way of example, consider a point subjected to a nonstationary constraint given by a plane translationally moving at a velocity $\mathbf{v}$ (Fig. E2.41). According to the theorem on the compound motion of a point, its actual displacement $d\mathbf{r}$ is equal to the vector sum of the relative displacement $\delta\mathbf{r}$ and the translational displacement $\mathbf{v}\,dt$. The figure illustrates the difference between the actual and virtual displacements of a point.

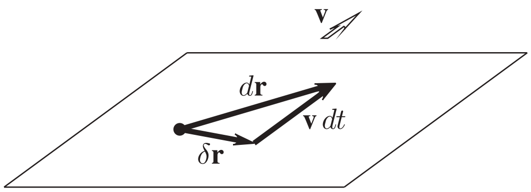

Figure E2.41. Virtual displacement of a point on a moving plane.

For stationary constraints, the actual displacements of points are contained in the set of virtual displacements.

A mechanical system can have quite a few various virtual displacements. But for systems consisting of material rigid bodies and finitely many point particles there exist several mutually independent virtual displacements in terms of which one can express any other virtual displacement. The number of independent displacements is called the *number of degrees of freedom of the mechanical system*.

► **Generalized coordinates.** *Generalized coordinates* are independent parameters that uniquely determine the position of each point of a mechanical system. There are as many degrees of freedom in a holonomic system as there are generalized coordinates, but this is not true for nonholonomic systems, which have fewer degrees of freedom than generalized coordinates.

Consider several specific examples.

1. A free point particle in space is a system with three degrees of freedom.

2. A free rigid body has six degrees of freedom. Indeed, the position of any point of the body in space can be determined if the positions of three points B_1, B_2, B_3 of the body that do not lie on a single straight line are known. The position of each of these points can be determined by three parameters, for example, the coordinates x_j, y_j, z_j ($j = 1, 2, 3$). All in all, there are nine coordinates, but they cannot be specified arbitrarily, because there are three equations saying that the distances s_{12}, s_{23}, s_{31} between the points must remain constant. (The body is rigid.) If, say, the six coordinates x_1, y_1, z_1, x_2, y_2, x_3 are known, then the remaining three coordinates z_2, y_3, z_3 can be found from the constraint equations.

3. A body rotating about a fixed axis has a single degree of freedom, and for the generalized coordinate one can take the rotation angle φ.

4. A rigid body in plane-parallel motion has three degrees of freedom; for generalized coordinates one can take, for example, the rotation angle and two Cartesian coordinates of any point of the body.

5. A rigid body in translational motion has three degrees of freedom; for generalized coordinates one can take three Cartesian coordinates of any point of the body.

6. A system consisting of a prism lying on a plane and a disk rolling without slip on the lateral face of the prism has two degrees of freedom (Fig. E2.42).

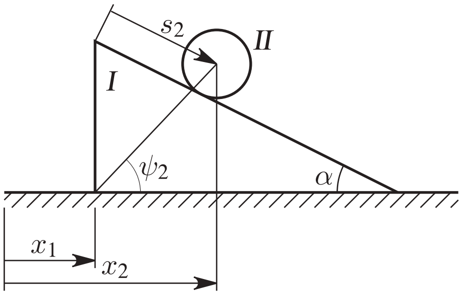

Figure E2.42. System consisting of a prism on a plane and a disk rolling without slip on the lateral face of the prism.

7. A system consisting of two free points has six degrees of freedom.

8. A sewing machine mechanism consisting of numerous rigid bodies has a single degree of freedom.

9. A thin rectilinear rod moving on a plane in such a way that the velocity of the rod center must be parallel to the rod axis has two degrees of freedom.

Of the above examples of mechanical systems, only one system (the last) is nonholonomic, while the others are holonomic.

We again return to the notion of generalized coordinates, which we illustrate by the system in Example 6 (Fig. E2.42). The position of each point of the disk and the prism is known as soon as we specify the values for one of the following pairs of variables: either (x_1, x_2), or (x_1, s_2), or (x_2, ψ_2), or (x_1, ψ_2), or $q_1 = \frac{1}{3}(x_1 + x_2)$, or $q_2 = \frac{1}{2}x_1 - x_2$, etc.

Let us summarize. For the system under study, there are infinitely many choices of generalized coordinates, but each set of these always contains two independent variables. Since the system is holonomic, it follows that the number of generalized coordinates is 2 (i.e., coincides with the number of degrees of freedom). The coordinates are referred to as generalized because they may fail to have a clear geometric meaning. (For example, this is the case for the coordinates (q_1, q_2).)

► **Ideal constraints.** Constraints are said to be *ideal* if the sum of works of their reactions $\mathbf{R}_j$ is zero on any virtual displacement of the system:

$$\sum \delta A_j^{\mathbf{R}} = 0.$$

An example of a system with ideal constraints is a free rigid body. Any complicated mechanism consisting of several rigid bodies can be treated as a mechanical system with ideal constraints if some bodies are connected absolutely rigidly and the other bodies are connected by ideal hinges (without friction) and also by weightless unstretchable perfectly flexible strings. Moreover, the contact surfaces must be either absolutely smooth or perfectly rough when one of the bodies rolls on another without slip so that the moment of rolling friction forces at the point of their tangency is zero and the static friction force is nonzero.

E2.6.3. Principle of Virtual Displacements

Principle of virtual displacements is stated as follows: for a mechanical system with ideal constraints to be in equilibrium, it is necessary and sufficient that the sum of elementary works of active forces $\mathbf{F}_i$ on any virtual displacement of the system be zero.

The mathematical representation of the principle of virtual displacements is

$$\sum \delta A_i^{\mathbf{F}} = 0. \tag{E2.6.3.1}$$

If a system has n degrees of freedom and its position is determined by generalized coordinates $q_j (j = 1, 2, \ldots, n)$, then the last expression can be rewritten as

$$\sum \delta A_i^{\mathbf{F}} = \sum \mathbf{F}_i \cdot \delta \mathbf{r}_i = \sum_{j=1}^{n} Q_j \, \delta q_j = 0,$$

where the δq_j are variations of the generalized coordinates.

The coefficients Q_j are called *generalized forces*. Since the variations of generalized coordinates are independent, we see that for the equation to be satisfied it is necessary that each of the factors multiplying δq_j be zero. Thus, if a system with ideal constraints is in equilibrium, then all generalized forces are zero.

Example. Find the angle φ of deviation from the vertical axis of a heavy homogeneous rod of weight G with a horizontal force F applied to the lower end B (Fig. E2.43).

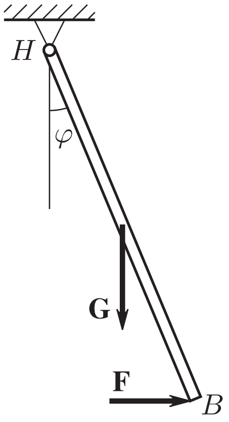

Figure E2.43. Deviation from the vertical axis of a heavy rod under the action of a horizontal force.

Solution. In the mechanical system we include a rod, which is a rigid body. Neglecting the friction in the hinge, we conclude that the constraints imposed on the system are ideal, and its equilibrium can be studied by the principle of virtual displacements.

In the equilibrium characterized by the angle φ, consider a virtual displacement of the system; i.e., we deflect the rod by a small angle $\delta\varphi$ about the hinge H, calculate the sum of elementary works of active forces **G** and **F**, and equate it with zero:

$$-GL \sin\varphi \, \delta\varphi + F2L \cos\varphi \, \delta\varphi = 0$$

(the rod length is $2L$), or, after transformations,

$$\delta\varphi \, (-GL \sin\varphi + F2L \cos\varphi) = 0.$$

Since the virtual displacement $\delta\varphi$ is arbitrary, it follows that, for the product to be zero, it is necessary to equate the expression in parentheses with zero:

$$-GL \sin\varphi + F2L \cos\varphi = 0.$$

This implies the desired angle $\varphi = \arctan(2F/G)$.

E2.6.4. General Equation of Dynamics (d'Alembert–Lagrange Principle)

Statement of the general equation of dynamics: A mechanical system with imposed ideal constraints moves in such a way that, at each time instant, the sum of elementary works of all active forces $\mathbf{F}_j$ and forces of inertia $\mathbf{\Phi}_j$ on any virtual displacement of the system is zero. The mathematical statement of the d'Alembert–Lagrange principle is

$$\sum_j \delta A_j^{\mathbf{F}} + \sum_j \delta A_j^{\Phi} = 0. \qquad \text{(E2.6.4.1)}$$

Remark. Problems of system dynamics can generally be solved by various methods. The general equation of dynamics has the unquestionable advantage that once it is used, the solution process does not involve the unknown constraint reaction forces, which significantly decreases the order of the system of equations of motion.

Example. A rectangular parallelepiped of mass m_1 moves on a smooth horizontal surface, a homogeneous disk of mass m_2 and radius r_2 rolls on the upper ideally rough surface of the parallelepiped, and a constant horizontal force F_2 is applied to the center of the disk (Fig. E2.44a). Find the accelerations a_1 of the parallelepiped and a_2 of the disk center.

Solution. Consider the system consisting of the parallelepiped and the disk. The constraints imposed on the system are ideal, and we can use the general equation of dynamics to study the motion of the system. For the coordinates determining the system position we take the absolute coordinate x_1 of the parallelepiped and the coordinate x_2 characterizing the disk position with respect to the parallelepiped.

The acceleration vectors a_1 of the parallelepiped and a_2 of the disk center are horizontal, their magnitudes are $a_1 = \ddot{x}_1$ and $a_2 = \ddot{x}_1 + \ddot{x}_2$, and the disk angular acceleration is $\varepsilon_2 = \ddot{x}_2/r_2$. The expected directions of the vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ and the corresponding direction of the curved arrow ε_2 are shown in Fig. E2.44b. We supplement the active forces $\mathbf{F}_2$, $\mathbf{G}_1$, $\mathbf{G}_2$ (where $\mathbf{G}_1$ and $\mathbf{G}_2$ are the forces of weight of the parallelepiped and of the disk) with the forces of inertia $\mathbf{\Phi}_1 = -m_1\mathbf{a}_1$, $\mathbf{\Phi}_2 = -m_2\mathbf{a}_2$ and the couple of forces of inertia with torque $M_2^{\Phi} = -I_{2c}\varepsilon_2$ (see Fig. E2.44b).

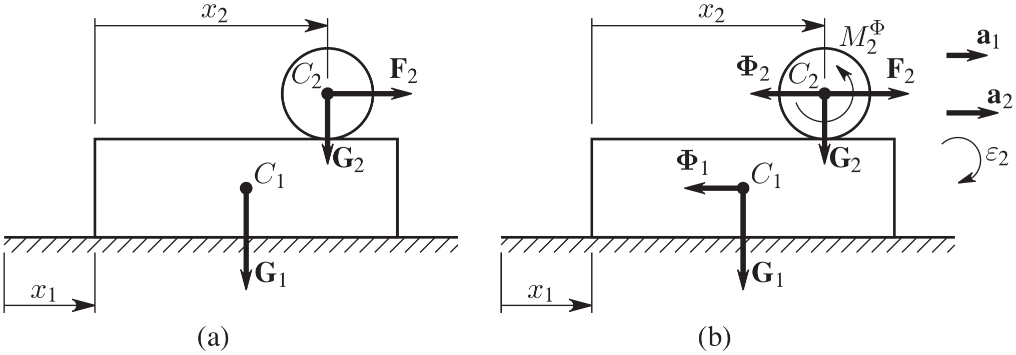

Figure E2.44. Disk rolling on the surface of a moving parallelepiped: (a) coordinates determining the system position and the active forces; (b) vectors of accelerations and forces of inertia.

Consider a virtual displacement of the system increasing the coordinates x_1, x_2 by some values δx_1, δx_2. Then the disk rotates by the angle $\delta\varphi_2 = \delta x_2/r_2$.

Now we calculate the sum of elementary works of active forces and forces of inertia on the virtual displacement and equate it with zero:

$$F_2(\delta x_1 + \delta x_2) - m_1 a_1\,\delta x_1 - m_2 a_2(\delta x_1 + \delta x_2) - I_{2c}\varepsilon_2\,\delta\varphi_2 = 0.$$

On rearranging, we obtain

$$\delta x_1\big[F_2 - m_1\ddot{x}_1 - m_2(\ddot{x}_1 + \ddot{x}_2)] + \delta x_2[F_2 - m_2(\ddot{x}_1 + \ddot{x}_2) - \tfrac{1}{2}m_2\ddot{x}_2\big] = 0.$$

Here the virtual displacements δx_1 and δx_2 can take arbitrary mutually independent values. For the equation to be always satisfied, it is necessary that both factors in square brackets multiplying δx_1 and δx_2 be zero. Thus, the last relation becomes a system of two equations with two unknowns,

$$F_2 - m_1\ddot{x}_1 - m_2(\ddot{x}_1 + \ddot{x}_2) = 0,$$
$$F_2 - m_2(\ddot{x}_1 + \ddot{x}_2) - \tfrac{1}{2}m_2\ddot{x}_2 = 0.$$

By solving the system, we obtain

$$a_1 = \ddot{x}_1 = \frac{F_2}{3m_1 + m_2}, \quad a_2 = \ddot{x}_1 + \ddot{x}_2 = \frac{F_2(2m_1 + m_2)}{m_2(3m_1 + m_2)}.$$

E2.6.5. Lagrange Equations of the Second Kind

To describe the motion of a holonomic mechanical system with n degrees of freedom with imposed ideal constraints, one often uses the *Lagrange equations of the second kind*

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_j}\right) - \frac{\partial T}{\partial q_j} = Q_j, \qquad j = 1, 2, \ldots, n. \tag{E2.6.5.1}$$

Here the q_j are *generalized coordinates* the number of which is equal to the number n of degrees of freedom, the $\dot{q}_j$ are the *generalized velocities* equal to the time derivatives of the generalized coordinates, T is the kinetic energy of the system, and the Q_j are *generalized forces*.

The quantities T and Q_j must be represented as functions of generalized velocities, generalized coordinates, and time:

$$T = T(\dot{q}_1, \ldots, \dot{q}_n, q_1, \ldots, q_n, t),$$
$$Q_j = Q_j(\dot{q}_1, \ldots, \dot{q}_n, q_1, \ldots, q_n, t).$$

The generalized forces are found from the expression for the sum of elementary works of active forces $\mathbf{F}_i$ on a virtual displacement of the system transformed to the form

$$\sum \delta A_i^{\mathbf{F}} = \sum \mathbf{F}_i \cdot \delta \mathbf{r}_i = \sum_{j=1}^{n} Q_j \, \delta q_j.$$

The number of generalized forces is equal to the number of degrees of freedom.

The physical dimension of the generalized force Q_j depends on the dimension of the corresponding generalized coordinate q_j, because the dimension of their product $Q_j \, \delta q_j$ must coincide with the dimension of work of force. Therefore, Q_j may have no clear physical meaning, and just this fact underlies the name “generalized force.”

After the functions T and Q_j are substituted, the Lagrange equations acquire the form of a system of second-order ordinary differential equations, which should be integrated with the initial conditions taken into account.

Example. To visualize the advantages of the above method, we construct the differential equations of motion of the mechanical system considered in the example in the preceding section in the form of Lagrange equations of the second kind.

Solution. Consider a system consisting of a parallelepiped and a disk (Fig. E2.44). The imposed constraints are holonomic and ideal, and hence the Lagrange equations of the second kind can be used. The system has two degrees of freedom; for the generalized coordinates determining the system position we choose the absolute coordinates $q_1 = x_1$ of the parallelepiped and $q_2 = x_1 + x_2$ of the disk center (see Fig. E2.44a).

We write out the expression for the kinetic energy of the system and reduce it to the form of a function of $\dot{q}_1$, $\dot{q}_2$, q_1, q_2, and t:

$$T = T_1 + T_2 = \frac{m_1 v_1^2}{2} + \frac{m_2 v_{2\mathrm{c}}^2}{2} + \frac{I_{2\mathrm{c}}\omega_2^2}{2} = \frac{m_1 \dot{q}_1^2}{2} + \frac{m_2 \dot{q}_2^2}{2} + \frac{m_2(\dot{q}_2 - \dot{q}_1)^2}{2}.$$

In these transformations, we have used the formula for the kinetic energy of the body in plane motion and the kinematic relation $\omega_2 = (\dot{q}_2 - \dot{q}_1)/r_2$.

Because of the simple kinematics and the fortunate choice of generalized coordinates, the expression for the kinetic energy (explicitly) contains neither the generalized coordinates nor time.

Let us calculate the generalized forces. To this end, we calculate the sum of elementary works of active forces $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{F}_2$ on a virtual displacement of the system determined by variations δq_1, δq_2 directed so that the coordinates q_1 and q_2 increase:

$$\sum \delta A_i^{\mathbf{F}} = F_2(\delta x_1 + \delta x_2) = F_2\,\delta q_2 = 0\,\delta q_1 + F_2\,\delta q_2.$$

The coefficients multiplying the variations of generalized coordinates are the desired expressions for generalized forces, whence

$$Q_1 = 0, \quad Q_2 = F_2.$$

It remains to write out the system of equations of motion:

$$\begin{aligned} \frac{d}{dt}\left(\frac{\partial T}{\partial \dot q_1}\right) - \frac{\partial T}{\partial q_1} &= Q_1, \\ \frac{d}{dt}\left(\frac{\partial T}{\partial \dot q_2}\right) - \frac{\partial T}{\partial q_2} &= Q_2. \end{aligned}$$

In view of the obtained expressions for T, Q_1, and Q_2, after appropriate transformations the equations of motion acquire the form

$$\begin{aligned} m_1\ddot q_1 - \tfrac{1}{2}m_2(\ddot q_2 - \ddot q_1) &= 0, \\ m_2\ddot q_2 + \tfrac{1}{2}m_2(\ddot q_2 - \ddot q_1) &= F_2. \end{aligned}$$

The solution thus obtained coincides with the solution of this problem obtained earlier by a different method in the preceding section.

In contrast to the Lagrange equations of the second kind, the general equation of dynamics can also be used to study nonholonomic systems and thus has a wider scope. At the same time, the Lagrange equations of the second kind are more convenient when studying holonomic systems, because they require fewer manipulations, and these manipulations are simpler.

E2.7. Small Oscillations of Mechanical Systems

E2.7.1. Preliminaries

In what follows, we consider a holonomic conservative mechanical system with n degrees of freedom whose position is determined by generalized coordinates $q_1, q_2, \dots, q_n$. Without loss of generality, we assume that the coordinate values $q_1 = q_2 = \dots = q_n = 0$ correspond to an equilibrium of the system. In this case, all generalized forces are zero, $Q_1 = Q_2 = \dots = Q_n = 0$.

The equilibrium $q_j = 0, \dot q_j = 0; j = 1, 2, \dots, n$ of the system is said to be *stable* if, for each $\varepsilon > 0$, there exists a $\delta = \delta(\varepsilon) > 0$ such that if the initial deviations for $t = 0$ are in the δ-neighborhood $|q_j^0| < \delta, |\dot q_j^0| < \delta$, then, for the entire time of motion, the system does not leave the ε-neighborhood of the equilibrium; i.e., $|q_j| < \varepsilon$ and $|\dot q_j| < \varepsilon$ for $t > 0$.

Example. A physical pendulum moving in the vertical plane (Fig. E2.45) has two equilibria corresponding to the angles $\varphi = 0$ and $\varphi = \pi$. It is intuitively clear that the first equilibrium is stable and the second equilibrium is unstable.

The character of equilibrium can be determined by studying the exact solution of the equations of motion of the system. But there exist stability criteria for the system equilibrium, which permit solving this problem more rationally.

One such criterion is given by the following theorem.

THEOREM (LAGRANGE). *If, in a certain position of a conservative system, the potential energy has a strict minimum, then this position is a stable equilibrium of the system.*

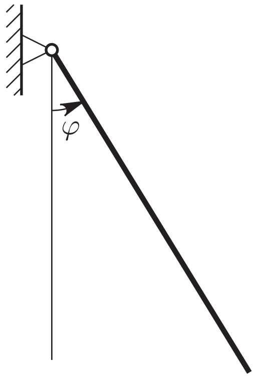

Figure E2.45. Oscillations of a physical pendulum in the vertical plane.

E2.7.2. Small Oscillations of a Conservative System

The kinetic and potential energy of a conservative system with n degrees of freedom can be expressed via the generalized coordinates q_j and generalized velocities $\dot{q}_j$ $(i = 1, \ldots, n)$:

$$T = \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(q_1, \ldots, q_n)\dot{q}_i\dot{q}_j, \qquad \Pi = \Pi(q_1, \ldots, q_n).$$

We assume that:

(1) The state of stable equilibrium of the system corresponds to the coordinates $q_1 = \cdots = q_n = 0$. By the Lagrange theorem, the potential energy of the system has a strict minimum at this point; i.e., $\frac{\partial\Pi}{\partial q_j}(0, 0, \ldots, 0) = 0, j = 1, 2, \ldots, n$.

(2) $\Pi(0, 0, \ldots, 0) = 0$, which can always be assumed because of the arbitrary constant contained in the definition of the potential energy (see Subsection E2.5.5).

(3) In the motion of the system, the values of the coordinates q_j and velocities $\dot{q}_j$ are small.

We expand the functions T and Π in a neighborhood of $q_1 = \cdots = q_n = 0$ and $\dot{q}_1 = \cdots = \dot{q}_n = 0$ into series in powers of the coordinates and velocities preserving the terms of the second order of magnitude in q_j and $\dot{q}_j$:

$$T = \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}\dot{q}_i\dot{q}_j, \qquad \Pi = \frac{1}{2}\sum_{i,j=1}^{n} c_{ij}q_iq_j,$$

where the constants a_{ij} and c_{ij} satisfy the conditions $a_{ij} = a_{ji}$ and $c_{ij} = c_{ji} = \frac{\partial^2\Pi}{\partial q_i\partial q_j}\Big|_{q_k=0}$.

As a result, both functions become quadratic forms whose positive definiteness follows from the physical meaning of the kinetic energy $T \geq 0$ and the fact that the potential energy Π at a point of stable equilibrium has a strict minimum, $\Pi \geq 0$.

The Lagrange equations with the above expressions for T and Π taken into account acquire the form of a system of n second-order ordinary linear equations

$$\sum_{j=1}^{n}(a_{ij}\ddot{q}_j + c_{ij}q_j) = 0 \quad (i = 1, \ldots, n).$$

The above-described process is called the linearization of equations of motion. The problem of closeness of the solutions of the linearized system to those of the original nonlinear problem is important, but we do not discuss it here.

We seek a particular solution of the above system in the form $q_j = u_j \sin(\omega t + \alpha)$ with $j = 1, \ldots, n$. By substituting this expression into the equations of motion and by cancelling $\sin(\omega t + \alpha)$, we obtain a system of linear algebraic equations for the amplitudes u_j. Since some of the amplitudes must be nonzero, it follows that the determinant of the system of homogeneous equations must be zero:

$$\det|c_{ij} - \lambda a_{ij}| = 0, \quad \lambda = \omega^2.$$

The obtained equation is called the *secular equation* or the *frequency equation*. By expanding the determinant, we obtain a polynomial of degree n in λ.

One can prove that in the case under study the secular equation has only real positive roots λ_j associated with real positive frequencies $\omega_j = \sqrt{\lambda_j}$. In this case, the n obtained particular solutions of the system are linearly independent (including the case of multiple roots of the secular equation); they are called the *normal oscillations* of the system.

Since the system of differential equations is linear, we see that any linear combination of normal oscillations with constant coefficients is also a solution of the system.

Example. A double mathematical pendulum consists of two weightless rods of equal length l with heavy point particles of equal mass m fixed at the ends of the rods. The construction moves in the plane of the figure, the upper rod can rotate about the fixed hinge O_1 adjusted to the base, and the second pendulum can rotate about the moving hinge O_2 connecting the rods (Fig. E2.46a). Study the small motions of the system near the stable equilibrium.

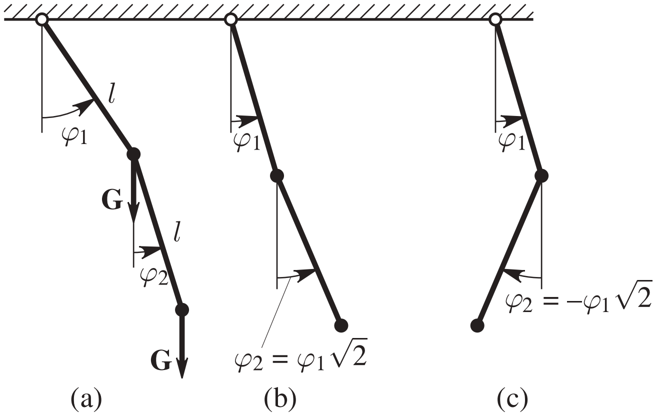
l
φ_1
G
l
φ_2
G
φ_1
$\varphi_2 = \varphi_1\sqrt{2}$
φ_1
$\varphi_2 = -\varphi_1\sqrt{2}$
(a)
(b)
(c)

Figure E2.46. Oscillations of the double mathematical pendulum: (a) sketch of the device; (b) the angles made by the rods with the vertical axis for $C_1 = 0$; (c) the angles made by the rods with the vertical axis for $C_2 = 0$.

Solution. For the generalized coordinates we take φ_1 and φ_2, i.e., the rod rotation angles about the vertical. The expressions for the kinetic and potential energy of the system have the form

$$T = ml^2\dot\varphi_1^2 + 2ml^2\cos(\varphi_2 - \varphi_1)\dot\varphi_1\dot\varphi_2 + ml^2\dot\varphi_2^2,$$
$$\Pi = -2mgl\cos\,\varphi_1 - mgl\cos\,\varphi_2.$$

It follows from the expression for the potential energy that the stable equilibrium of the system corresponds to the coordinate values $\varphi_1 = \varphi_2 = 0$. Assuming that the oscillations are small, we rewrite the expressions for T and Π omitting the terms of the third and higher orders of magnitude:

$$T = ml^2\dot\varphi_1^2 + 2ml^2\dot\varphi_1\dot\varphi_2 + ml^2\dot\varphi_2^2,$$
$$\Pi = mgl\varphi_1^2 + \tfrac{1}{2}mgl\varphi_2^2.$$

We write out the equations of motion of the system as

$$2l\ddot\varphi_1 + l\ddot\varphi_2 + 2g\varphi_1 = 0,$$
$$l\ddot\varphi_1 + l\ddot\varphi_2 + g\varphi_2 = 0.$$

We seek a particular solution of this system in the form

$$\varphi_1 = u_1 \sin(\omega t + \alpha), \quad \varphi_2 = u_2 \sin(\omega t + \alpha).$$

The secular equation is reduced to the form

$$l^2\omega^4 - 4lg\omega^2 + 2g^2 = 0,$$

whose roots are the two angular natural oscillation frequencies

$$\omega_1 = \sqrt{\left(2-\sqrt{2}\right)\frac{g}{l}}, \quad \omega_2 = \sqrt{\left(2+\sqrt{2}\right)\frac{g}{l}}.$$

It follows from the second equation of motion that $u_2 = u_1 \frac{l\omega^2}{g-l\omega^2}$, which implies that $u_2 = u_1\sqrt{2}$ for the frequency ω_1 and $u_2 = -u_1\sqrt{2}$ for the frequency ω_2. Therefore, the general solution of the linearized system of differential equations describing the motion of the double mathematical pendulum has the form

$$\begin{aligned} \varphi_1 &= C_1 \sin(\omega_1 t + \alpha_1) + C_2 \sin(\omega_2 t + \alpha_2), \\ \varphi_2 &= \sqrt{2}\, C_1 \sin(\omega_1 t + \alpha_1) - \sqrt{2}\, C_2 \sin(\omega_2 t + \alpha_2). \end{aligned}$$

Here C_1, C_2, α_1, α_2 are arbitrary constants depending on the initial conditions, i.e., on the values φ_1^0, φ_2^0, $\dot\varphi_1^0$, $\dot\varphi_2^0$ at the initial time instant $t = 0$.

The angles φ_1 and φ_2 made by the rods with the vertical axis for $C_1 = 0$, $C_2 \neq 0$ and $C_2 = 0$, $C_1 \neq 0$ are shown in Fig. E2.46, b and c.

E2.7.3. Normal Coordinates

Obviously, the difficulties encountered when seeking the natural oscillation frequencies of a mechanical system increase with the number of degrees of freedom. Moreover, finding the general characteristics of the system behavior is also difficult, because the characteristic equation is cumbersome. The existence of *normal* or *principal coordinates* in a system permits one to make the analysis of the system less laborious.

THEOREM. *Two quadratic forms at least one of which is positive definite can be simultaneously reduced to "sums of squares" by a nonsingular linear transformation of the variables.*

In the case of oscillations of a conservative system, there are two positive definite quadratic forms: the kinetic energy, which is a quadratic form of the generalized velocities $\dot q_i$, and the potential energy, which is a quadratic form of the generalized coordinates q_i. Since the generalized coordinates and generalized velocities behave in the same way under a linear transformation,

$$q_k = \sum_{s=1}^{n} \gamma_{ks}\theta_s, \quad \dot q_k = \sum_{s=1}^{n} \gamma_{ks}\dot\theta_s, \quad k = 1, \ldots, n; \quad \det(\gamma_{\rm ks}) \neq 0,$$

we see that the kinetic and potential energy of the system in the new variables become

$$T = \frac{1}{2}\sum_{j=1}^{n}\dot\theta_j^2, \qquad \Pi = \frac{1}{2}\sum_{j=1}^{n}\lambda_j\theta_j^2.$$

The Lagrange equations in normal coordinates θ_j have the form

$$\ddot\theta_j + \lambda_j\theta_j = 0, \qquad j = 1, \ldots, n.$$

Each of these equations contains only one unknown function. The general solutions of these equations determine the harmonic oscillations

$$\theta_j = C_j \sin(\omega_j t + \alpha_j), \quad \omega_j = \sqrt{\lambda_j}, \qquad j = 1, \ldots, n,$$

where the C_j and α_j are arbitrary constants.

Thus, we have shown that all the roots of the secular equation for a conservative mechanical system are real and positive. In addition, to n frequencies $\omega_j = \sqrt{\lambda_j}$ there correspond n linearly independent amplitude vectors, and the general solution of the system does not contain secular terms of the form $\theta_j = (C_0 + C_1 t + C_2 t^2 + \cdots)\sin(\omega_j t + \alpha_j)$ in the case of multiple frequencies.

E2.7.4. Influence of Small Dissipative Forces on System Oscillations

Assume that each point of the system is subjected not only to potential forces but also to resistance forces depending on the velocity. Such forces are said to be *dissipative*. In this case, the total mechanical energy of the system does not remain constant but decays with time.

Let us estimate the influence of dissipative forces on system oscillations under the assumption that these forces are small and linearly depend on the velocities of points of the system. In addition, we assume that the conservative system has distinct oscillation frequencies. The expressions for the generalized forces acquire the form

$$Q_i = -\sum_{k=1}^{n} b_{ik}\dot{q}_k - \sum_{k=1}^{n} c_{ik}q_k, \qquad i = 1, \dots, n.$$

By analogy with the potential energy of the system, we introduce the *dissipative Rayleigh function*, specifying it by a positive definite quadratic form R in the generalized velocities $\dot{q}_i$:

$$R = \frac{1}{2}\sum_{i,k=1}^{n} b_{ik}\dot{q}_i\dot{q}_k > 0 \quad \text{for} \quad \sum_{i=1}^{n} \dot{q}_i^2 > 0.$$

In the case under study, the system is *definitely dissipative*, and the equilibrium of the system is asymptotically stable; i.e., $q_i(t) \to 0$ as $t \to \infty$.

In normal coordinates, the equations of motion of the system acquire the form

$$\ddot{\theta}_i + \beta_i\dot{\theta}_i + \omega_i^2\theta_i + \sum_{j\neq i}^{n} \beta_{ij}\dot{\theta}_j = 0 \quad (i, j = 1, \dots, n).$$

Using the fact that the dissipative forces are small—i.e., the coefficients β_i and β_{ij} are small—in the first approximation, one can obtain a system of linearly independent particular solutions ("normal oscillations") in the form (Gantmakher, 1975)

$$\theta_i = A_i e^{-\frac{\beta_i}{2}t}\sin(\omega_i t + \alpha_i), \quad \theta_j = \varepsilon_j A_i e^{-\frac{\beta_i}{2}t}\sin\Big(\omega_i t + \alpha_i + \frac{\pi}{2}\Big),$$
$$\varepsilon_j = -\frac{\beta_{ji}\omega_i}{\omega_j^2 - \omega_i^2}, \quad j = 1, \dots, i-1, i+1, \dots, n,$$

where $i = 1, \dots, n$. This implies that:

(1) The system frequencies are close to those of the corresponding conservative system in the absence of dissipative forces.

(2) The system oscillations are damped as $t \to \infty$.

(3) In the ith normal oscillation, all coordinates θ_j $(j \neq i)$ are small compared with the ith coordinate θ_i and differ from it in phase by a quarter period.

E2.7.5. Forced Oscillations

Assume that the system considered in the preceding section is also subjected to *perturbing forces* Q_i. The equations of motion of such a system acquire the form

$$\sum_{j=1}^{n}(a_{ij}\ddot{q}_j + b_{ij}\dot{q}_j + c_{ij}q_j) = Q_i(t), \qquad i = 1, \dots, n.$$

The general solution $q_i(t)$ of this inhomogeneous system of linear equations is the sum of the general solution $q_i^{\circ}(t)$ of the corresponding homogeneous system and a particular solution $\widetilde{q}_i(t)$ of the inhomogeneous system,

$$q_i(t) = q_i^{\circ}(t) + \widetilde{q}_i(t).$$

Since $q_i^{\circ}(t) \to 0$ as $t \to \infty$, it follows that the system behavior for large values t is determined by the particular solution, $q_i(t) \to \widetilde{q}_i(t)$ as $t \to \infty$.

The system under study is linear, and, because of the linear superposition of particular solutions, the general case of finding the forced oscillations is reduced to studying the solution of the system for the case in which only one of the perturbing forces is nonzero, say, $Q_1 \neq 0$.

We restrict ourselves to the case of a harmonic driving force $Q_1(t) = Ae^{i\Omega t}$, $Q_j = 0$ ($j = 2, \dots, n$), where i is the imaginary unit ($i^2 = -1$). We seek the forced oscillations in the form $q_k = B_k e^{i\Omega t}$ ($k = 1, \dots, n$). By substituting these expressions into the equations of motion, we obtain a system of algebraic equations for B_k:

$$\sum_{j=1}^{n}\bigl[a_{1j}(i\Omega)^2 + b_{1j}(i\Omega) + c_{1j}q_j\bigr]B_j = A,$$
$$\sum_{j=1}^{n}\bigl[a_{kj}(i\Omega)^2 + b_{kj}(i\Omega) + c_{kj}q_j\bigr]B_j = 0, \qquad k = 2, \dots, n.$$

By solving this system, we obtain $B_k = W_{1k}(i\Omega)A$, where $W_{1k}(i\Omega) = \Delta_{1k}(i\Omega)/\Delta(i\Omega)$ is a proper rational function of $i\Omega$ with real coefficients. The hodograph of this function in the complex plane is called the *frequency* or *amplitude-phase characteristic*.

By setting $W_{1k}(i\Omega) = R_{1k}(\Omega)e^{i\Psi_{1k}(\Omega)}$, we rewrite the solution as

$$q_k = R_{1k}(\Omega)Ae^{i[\Omega t + \Psi_{1k}(\Omega)]}, \qquad k = 1, \dots, n,$$

where $R_{1k}(\Omega)$ is the *amplitude* characteristic and $\Psi_{1k}(\Omega)$ is the *phase* characteristic.

In the case of the sinusoidal perturbing force

$$Q_1 = \sin(\Omega t) = \frac{A}{2i}(e^{i\Omega t} - e^{-i\Omega t}),$$

the solution becomes

$$q_k = R_{1k}(\Omega)A\sin[\Omega t + \Psi_{1k}(\Omega)] \quad (k = 1, \dots, n);$$

i.e., a sinusoidal driving force causes a sinusoidal response, where the forces are multiplied by the amplitude characteristic $R_{1k}(\Omega)$ and the phase shift is determined by the phase characteristic $\Psi_{1k}(\Omega)$.

By adjusting the system characteristics, one can suppress oscillations at some "useless" frequencies and increase the amplitudes of these oscillations at the other "required" frequencies. This idea underlies the design of frequency filters.

Example. The equation of forced oscillations of a system with a single degree of freedom and its amplitude characteristic R and phase characteristic Ψ has the form

$$\ddot{q} + b\dot{q} + \omega^2 q = A\sin(\Omega t), \quad R = \frac{1}{\sqrt{(\omega^2 - \Omega^2)^2 + b^2\Omega^2}}, \quad \Psi = \arctan\left(\frac{b\Omega}{\omega^2 - \Omega^2}\right).$$

Here we note the following specific features of the obtained solution: the forced oscillation amplitude $B = AR$ becomes unboundedly large if the dissipative forces are very small and the driving force frequency tends to the natural oscillation frequencies ($B \to \infty$ as $b \to 0$ and $\Omega \to \omega$).

E2.8. Dynamics of an Absolutely Rigid Body

E2.8.1. Rotation of a Rigid Body about a Fixed Axis

Assume that a body of mass m rotates about the vertical z-axis under the action of a system of active forces $\{\mathbf{F}_k\}$. At points O_1 and O_2 of the axis, cylindrical hinges are located and a thrust bearing* is placed at the point O_1 (Fig. E2.47).

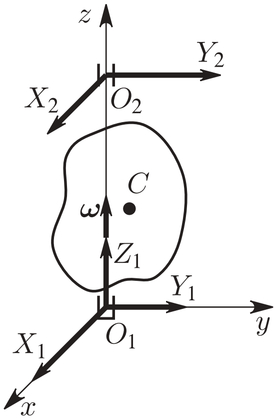

Figure E2.47. Rotation of a rigid body about a vertical axis under the action of active forces.

Theorems on the momentum and angular momentum with respect to the center O_1 imply the following system of six equations with six unknowns:

$$\begin{aligned}
-mx_C\omega^2 - my_C\dot{\omega} &= X_1 + X_2 + \sum F_{kx},\\
-my_C\omega^2 + mx_C\dot{\omega} &= Y_1 + Y_2 + \sum F_{ky},\\
0 &= Z_1 + \sum F_{kz},\\
I_{yz}\omega^2 - I_{zx}\dot{\omega} &= -LY_2 + \sum M_x(\mathbf{F}_k),\\
-I_{zx}\omega^2 - I_{zx}\dot{\omega} &= LX_2 + \sum M_y(\mathbf{F}_k),\\
I_{zz}\dot{\omega} &= \sum M_z(\mathbf{F}_k).
\end{aligned}$$

Here x_C and y_C are the coordinates of the center of mass C of the body, X_1, X_2, Y_1, Y_2, Z_1 are the projections of the constraint reaction forces onto the coordinate axes, and $L = O_1O_2$.

---

* A thrust bearing is a device preventing the body from vertical displacements.

The last equation does not contain reactions; it is called the *equation of motion of a rigid body about a fixed axis*. By integrating this equation, one can find the angular velocity and the angle of rotation as a function of time: $\omega = \omega(t, \omega_0)$, $\varphi = \varphi(t, \omega_0, \varphi_0)$.

If the body is at rest $\omega = \dot{\omega} = 0$, then the constraint reactions are said to be *static*. In the case of body rotation, the reactions are said to be *dynamic*; they depend on the angular velocity and the angular acceleration of the body.

The conditions under which the dynamic reactions are equal to static reactions have the form

$$x_C = y_C = 0, \quad I_{yz} = I_{zx} = 0.$$

This implies that, as a rigid body rotates about a fixed axis, the dynamic constraint reactions are equal to static ones if the rotation axis is one of the principal central axes of inertia.

E2.8.2. Plane-Parallel (Plane) Motion of a Rigid Body

Each motion of a body can be viewed as a motion of the center of mass and the motions of the body with respect to the center of mass.

In the case of plane motion, the body points, including the center of mass, move in planes parallel to a fixed (base) plane. The motion about the center of mass is the rotation about an axis perpendicular to the base plane and passing through the center of mass.

We draw a plane parallel to the base plane through the center of mass C and introduce fixed axes $O\xi\eta$ in this plane (Fig. E2.48). In addition, we introduce two moving reference frames whose common origin is the center of mass C. One of them, Cx^*y^* (the König axes), moves translationally: its axes are parallel to the axes of the fixed system $O\xi\eta$; the second system, Cxy, is rigidly fixed to the body and rotates about C.

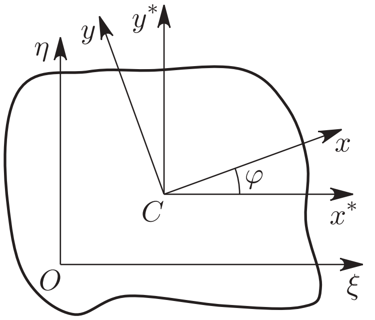

Figure E2.48. Plane-parallel motion of a rigid body.

The body position is determined by the following three parameters: the coordinates x_C, y_C of the center of mass in the fixed system $O\xi\eta$ and the rotation angle φ between the axes Cx^* and Cx.

The theorem on the motion of the center of mass of the body in the fixed system $O\xi\eta$ and the theorem on the angular momentum about the axis perpendicular to the base plane and passing through the center of mass give the following three equations of plane motion of a rigid body:

$$m\frac{d^2x_C}{dt^2} = \sum_k F^{\rm e}_{kx}, \quad m\frac{d^2y_C}{dt^2} = \sum_k F^{\rm e}_{ky}, \quad I_{CC}\frac{d^2\varphi}{dt^2} = \sum_k M_C\mathbf{F}^{\rm e}_k.$$

Here m is the mass of the body and I_{CC} is the moment of inertia of the body about the axis passing through the center of mass and perpendicular to the base plane; the expressions on the right-hand side of the relations contain the projections and moments of the external forces acting on the body.

By integrating the obtained system of equations with the initial conditions taken into account, one can determine x_C, y_C, and φ as functions of time t.

E2.8.3. Motion of a Rigid Body about a Fixed Point

► **Preliminary remarks and definitions.**

An arbitrary motion of a rigid body is the combination of two simultaneous motions: the translational motion together with an arbitrary point of the body and the spherical motion about this point treated as a fixed point.

In several problems of motion of a rigid body, the differential equations of motion determining the translational motion are independent of the equations of motion about the chosen point. Therefore, both systems of equations can be integrated independently of each other. As a rule, the difficulties of analytical integration arise in the second system. Therefore, the problem of motion of a body about a fixed point is most important in mechanics.

Any body with a single fixed point has three degrees of freedom, and its position can be determined by introducing three independent generalized coordinates.

We introduce two reference frames whose origins coincide with the fixed point O of the rigid body. The system $O\xi\eta\zeta$ is fixed, and the system $Oxyz$ is a moving system rigidly fixed to the body. The position of the moving system (i.e., the position of the rigid body) with respect to the fixed system is determined by the *Euler angles* φ, ψ, and θ.

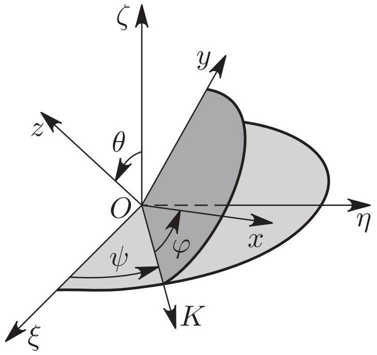

Figure E2.49. Motion of a rigid body about a fixed point: the fixed and moving reference frames $O\xi\eta\zeta$ and $Oxyz$, and the Euler angles φ, ψ, and θ.

The straight line OK of intersection of the planes $O\xi\eta$ and Oxy is called the *nodal line*. Its positive direction is chosen so that the nearest rotation from $O\zeta$ to Oz seems to be anticlockwise if it is observed from the side of the point K.

The *angle* φ *of proper rotation* is the angle between OK and Ox; it is assumed to be positive if, being observed from the positive sense of the axis, the rotation from OK to Ox seems to be anticlockwise. The axis Oz is called the *axis of proper rotation*. If only the angle φ varies, then the body rotates about the fixed axis Oz with the *angular velocity* $\boldsymbol{\dot\varphi} = \dot\varphi\mathbf{z}^\circ$ *of proper rotation* directed along the axis Oz, where $\mathbf{z}^\circ$ is the unit vector of the axis Oz. (We do not show all unit vectors of coordinate axes in Fig. E2.49 so as not to overload it.)

The *precession angle* ψ is the angle between $O\xi$ and OK; it is assumed to be positive if, being observed from the positive sense of the axis $O\zeta$, the rotation from $O\xi$ to OK seems to be anticlockwise. The axis $O\zeta$ is called the *axis of precession*. If only the angle ψ varies, then the body rotates about the fixed axis $O\zeta$ with the *angular precession velocity* $\dot{\boldsymbol{\psi}} = \dot{\psi}\boldsymbol{\zeta}^\circ$ directed along the axis $O\zeta$, where $\boldsymbol{\zeta}^\circ$ is the unit vector of the axis $O\zeta$.

The *nutation angle* θ is the angle between $O\zeta$ and Oz. The nodal line OK is called the *axis of nutation*. If only the angle θ varies, then the body rotates about the fixed axis OK with the *angular nutation velocity* $\dot{\boldsymbol{\theta}} = \dot{\theta}\,\overrightarrow{OK}^\circ$ directed along the axis OK, where $\overrightarrow{OK}^\circ$ is the unit vector of the axis OK.

The superposition of the system $O\xi\eta\zeta$ with the system $Oxyz$ can be realized by three rotations: first, by the angle ψ about the axis $O\zeta$, then by the angle θ about OK, and finally, by the angle φ about the axis Oz.

► **Kinematic Euler equations.** If all three angles vary as the body moves, then the angular velocity of the body $\boldsymbol{\omega}$ according to the theorems on the compound motion of a rigid body is equal to the vector sum of the vectors of angular velocities of the proper rotation, precession, and nutation:

$$\boldsymbol{\omega} = \dot{\boldsymbol{\varphi}} + \dot{\boldsymbol{\psi}} + \dot{\boldsymbol{\theta}}.$$

We denote the projections of the angular velocity $\boldsymbol{\omega}$ on the axes of the moving reference frame $Oxyz$ by p, q, r. The expressions of these quantities via the Euler angles φ, ψ, θ and their derivatives $\dot{\varphi}, \dot{\psi}, \dot{\theta}$ have the form

$$\begin{aligned} p &= \dot{\psi}\sin\theta\sin\varphi + \dot{\theta}\cos\varphi,\\ q &= \dot{\psi}\sin\theta\cos\varphi - \dot{\theta}\sin\varphi,\\ r &= \dot{\psi}\cos\theta + \dot{\varphi}. \end{aligned} \tag{E2.8.3.1}$$

For p', q', r', i.e., the projections of the angular velocity $\boldsymbol{\omega}$ on the axes of the fixed system $O\xi\eta\zeta$, the expressions via the Euler angles and their derivatives have the form

$$\begin{aligned} p' &= \dot{\varphi}\sin\theta\sin\psi + \dot{\theta}\cos\psi,\\ q' &= -\dot{\varphi}\sin\theta\cos\psi + \dot{\theta}\sin\psi,\\ r' &= \dot{\varphi}\cos\theta + \dot{\psi}. \end{aligned} \tag{E2.8.3.2}$$

The last two groups of equations are called the *kinematic Euler equations*.

► **Formulas for the angular momentum. Dynamic Euler equations**. The expressions for the angular momentum $\mathbf{K}_O$ of a rigid body about a fixed point O and its projections K_{Ox}, K_{Oy}, K_{Oz} onto the axes of the moving reference frame have the form (in what follows, the subscript O is omitted for convenience):

$$\begin{aligned} &\mathbf{K} = \sum_j(\mathbf{r}_j \times m_j\mathbf{v}_j) = \sum_j[\mathbf{r}_j \times m_j(\boldsymbol{\omega}\times\mathbf{r}_j)] = \boldsymbol{\omega}\sum_j m_j r_j^2 - \sum_j m_j\mathbf{r}_j(\mathbf{r}_j\cdot\boldsymbol{\omega}),\\ &K_x = pI_{xx} - qI_{xy} - rI_{xz},\\ &K_y = -pI_{yx} + qI_{yy} - rI_{yz},\\ &K_z = -pI_{zx} - qI_{zy} + rI_{zz}. \end{aligned} \tag{E2.8.3.3}$$

The expression for the kinetic energy T of a rigid body moving about a fixed point has the form

$$2T = \sum_j m_j v_j^2 = \sum_j m_j\mathbf{v}_j\cdot(\boldsymbol{\omega}\times\mathbf{r}_j) = \boldsymbol{\omega}\cdot\sum_j(\mathbf{r}_j\times m_j\mathbf{v}_j) = \boldsymbol{\omega}\cdot\mathbf{K},$$

or, in coordinate form,

$$2T = \boldsymbol{\omega} \cdot \mathbf{K} = I_{xx}p^2 + I_{yy}q^2 + I_{zz}r^2 - 2I_{yz}qr - 2I_{zx}rp - 2I_{xy}pq.$$

The *dynamic Euler equations* of motion of a rigid body about a fixed point are obtained from the theorem on the angular momentum about a fixed point O:

$$\frac{d\mathbf{K}_O}{dt} = \sum \mathbf{M}_O\mathbf{F}_j^{\mathrm{e}} \qquad \text{(in the fixed reference frame)}$$

or

$$\frac{\widetilde{d}\mathbf{K}_O}{dt} + \boldsymbol{\omega} \times \mathbf{K}_O = \mathbf{M}_O\mathbf{F}_j^{\mathrm{e}} \quad \text{(in the moving reference frame).}$$

It is inconvenient to project the vector relations on the axes of a fixed reference frame, because the coefficients $I_{\xi\xi}, I_{\eta\eta}, I_{\zeta\zeta}, \dots, I_{\xi\eta}$ contained in the expression for K_O depend on time, and the equations thus obtained are rather cumbersome. The equations of motion become concise if

1. The projections onto the axes of the moving reference frame are taken.
2. For the moving axes one takes the principal axes of inertia of the body for the point O, so that

$$K_{Ox} = I_{xx}p = Ap, \quad K_{Oy} = I_{yy}q = Bq, \quad K_{Oz} = I_{zz}r = Cr.$$

(Here the traditional notation $A = I_{xx}$, $B = I_{yy}$, $C = I_{zz}$ is used.)

As a result, the dynamic Euler equations acquire the form

$$\begin{aligned} A\frac{dp}{dt} + (C-B)qr &= M_x, \\ B\frac{dq}{dt} + (A-C)rp &= M_y, \\ A\frac{dr}{dt} + (B-A)pq &= M_z. \end{aligned} \tag{E2.8.3.4}$$

Along with the kinematic Euler equations (E2.8.3.1), they form a system of six first-order ordinary nonlinear equations for the six unknown functions $p, q, r, \varphi, \psi, \theta$ of time. The general integrals of the system must contain six arbitrary constants, which are determined by the initial conditions $p_0, q_0, r_0, \varphi_0, \psi_0, \theta_0$ for $t = t_0$.

By eliminating p, q, r, one can pass from the system of six equations to a system of three second-order ordinary differential equations for the Euler angles.

The difficulties of analytic integration of the system of nonlinear equations of motion are obvious, because, in the general case, M_x, M_y, and M_z are functions of t, p, q, r, φ, ψ, and θ.

► **Statement of the problem on the motion of a rigid body about a fixed point.** Consider the motion of a heavy rigid body about a fixed point under the action of a single active force of weight $\mathbf{G}$. We neglect the Earth's rotation, assuming that the Earth is fixed. The axis $O\zeta$ of the stationary system $O\xi\eta\zeta$ fixed to the Earth is directed vertically upwards and opposite to the force of gravity. The position vector $\overrightarrow{OC}$ of the center of gravity of a body is denoted by $\mathbf{r}_{\mathrm{c}}(x_{\mathrm{c}}, y_{\mathrm{c}}, z_{\mathrm{c}})$. The expressions for the direction cosines of the angles made by the axis $O\zeta$ with the axes of the moving reference frame become

$$\cos(\widehat{\zeta, x}) = \gamma_1 = \sin\theta \sin\varphi, \quad \cos(\widehat{\zeta, y}) = \gamma_2 = \sin\theta \cos\varphi, \quad \cos(\widehat{\zeta, z}) = \gamma_3 = \cos\theta.$$

Here $\gamma_1, \gamma_2, \gamma_3$ are the projections of the unit vector $\boldsymbol{\zeta}^\circ$ on the moving axes; i.e., $\boldsymbol{\zeta}^\circ = \gamma_1\mathbf{i} + \gamma_2\mathbf{j} + \gamma_3\mathbf{k}$, where $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are the unit vectors of the moving axes. Since the direction of the force of gravity does not vary, it follows that

$$\frac{d\boldsymbol{\zeta}^\circ}{dt} \equiv \frac{\tilde{d}\boldsymbol{\zeta}^\circ}{dt} + \boldsymbol{\omega} \times \boldsymbol{\zeta}^\circ = 0,$$

which implies the *Poisson equations*:

$$\begin{aligned}\frac{d\gamma_1}{dt} &= r\gamma_2 - q\gamma_3,\\ \frac{d\gamma_2}{dt} &= p\gamma_3 - r\gamma_1,\\ \frac{d\gamma_3}{dt} &= q\gamma_1 - p\gamma_2.\end{aligned}$$

In the case under study, with the relations $\mathbf{G} = -G\boldsymbol{\zeta}^\circ$ and $\mathbf{M}_0 = \mathbf{r} \times \mathbf{G}$ taken into account, the dynamic Euler equations acquire the form

$$\begin{aligned}A\frac{dp}{dt} + (C - B)qr &= G(\gamma_2 z_\mathrm{c} - \gamma_3 y_\mathrm{c}),\\ B\frac{dq}{dt} + (A - C)rp &= G(\gamma_3 x_\mathrm{c} - \gamma_1 z_\mathrm{c}),\\ A\frac{dr}{dt} + (B - A)pq &= G(\gamma_1 y_\mathrm{c} - \gamma_2 x_\mathrm{c}).\end{aligned}$$

Here the quantities A, B, C, x_c, y_c, z_c are constants.

The last equations, together with the Poisson equations, form a closed system of six equations with six unknown functions of time $\gamma_1, \gamma_2, \gamma_3, p, q, r$. If they are obtained, then to determine the Euler angles φ, ψ, θ as functions of time it is necessary to find $\psi(t)$ by a single quadrature of any of the kinematic Euler equations, because $\theta(t)$ and $\varphi(t)$ can be found from the expressions for the direction cosines of the angles made by the axis $O\zeta$ with the axes of the moving system.

One can show that the solution of the problem on the motion of a body is reduced to finding only a single integral.

If the system of equations of motion is represented in the canonical form

$$\frac{dp}{P} = \frac{dq}{Q} = \frac{dr}{R} = \frac{\gamma_1}{\Gamma_1} = \frac{\gamma_2}{\Gamma_2} = \frac{\gamma_3}{\Gamma_3} = dt,$$

where

$$\begin{aligned}P &= G(\gamma_2 z_\mathrm{c} - \gamma_3 y_\mathrm{c}) - (C - B)qr, & \Gamma_1 &= r\gamma_2 - q\gamma_3,\\ Q &= G(\gamma_3 x_\mathrm{c} - \gamma_1 z_\mathrm{c}) - (A - C)rp, & \Gamma_2 &= p\gamma_3 - r\gamma_1,\\ R &= G(\gamma_1 y_\mathrm{c} - \gamma_2 x_\mathrm{c}) - (B - A)pq, & \Gamma_3 &= q\gamma_1 - p\gamma_2,\end{aligned}$$

then one can see that

(1) It is possible to integrate the system of the first five equations

$$\frac{dp}{P} = \frac{dq}{Q} = \frac{dr}{R} = \frac{\gamma_1}{\Gamma_1} = \frac{\gamma_2}{\Gamma_2} = \frac{\gamma_3}{\Gamma_3}$$

alone, because they do not contain the time t explicitly.

(2) The structure of this system of five equations is such that, according to the theory of the last Jacobi multiplier for the canonical system of differential equations (the details of this theory are beyond the scope of this book), to reduce the problem to quadratures it suffices to know four rather than five integrals of the system.

(3) One of the integrals, $\gamma_1^2 + \gamma_2^2 + \gamma_3^2 = 1$, is obvious from geometric considerations.

(4) The theorem on the angular momentum of the system about the axis $O\zeta$, $\frac{dK_\zeta}{dt} = M_\zeta = 0$, leads to the integral

$$Ap\gamma_1 + Bq\gamma_2 + Cr\gamma_3 = \text{const.}$$

(5) The theorem on the kinetic energy $dT = -Gd\zeta_\text{c}$ gives one more integral,

$$\frac{1}{2}(Ap^2 + Bq^2 + Cr^2) = -G(x_\text{c}\gamma_1 + y_\text{c}\gamma_2 + z_\text{c}\gamma_3) + \text{const.}$$

Thus, it remains to find one last integral. This problem can be solved only in the following three special cases:

(i) The *Euler–Poinsot case*: motion by inertia, when the moment of external forces about the fixed point is zero: $\mathbf{M}_O = 0$.

(ii) The *Lagrange–Poisson case*: the ellipsoid of inertia of a body about the fixed point is an ellipsoid of rotation, $A = B$, and the center of gravity of the body lies on the axis of rotation of the ellipsoid of inertia.

(iii) The *Kowalewski case* determined by the following two conditions: (a) $A = B = 2C$ and (b) the center of gravity lies in the equatorial plane of the ellipsoid of inertia.

The studies showed that it is only in these three cases that there exists an algebraic integral that holds for any initial conditions, and this reduces the problem to quadratures. The cases of integration described in the literature and differing from the above cases hold only under certain specifically chosen initial conditions.

Example. Under the conditions $A = B$ (the ellipsoid of inertia is an ellipsoid of rotation) and $\mathbf{M}_O = 0$, the solution of the problem on the motion of a body can be solved completely in elementary functions.

The dynamic Euler equations become

$$A\frac{dp}{dt} + (C - A)qr = 0,$$
$$B\frac{dq}{dt} + (A - C)rp = 0,$$
$$A\frac{dr}{dt} = 0.$$

The solution of this system gives the three integrals

$$A(p^2 + q^2) + Cr^2 = h, \quad A^2(p^2 + q^2) + C^2r^2 = K_O^2, \quad r = r_0 = \text{const.}$$

The first of them is the energy integral, the second follows from the law of conservation of the angular momentum $\mathbf{K}_O$, and the third follows directly from the last equation.

To solve the problem completely, it is necessary to express the Euler angles as functions of time. To simplify the solution, we direct the axis $O\zeta$ along the vector $\mathbf{K}_O$, which remains constant in magnitude and direction. The projections of $\mathbf{K}_O$ onto the axes Ox, Oy, and Oz are

$$K_x = Ap = G\sin\theta\sin\varphi, \quad K_y = Aq = G\sin\theta\cos\varphi, \quad K_z = Cr = G\cos\varphi.$$

It follows from the last equation that $\theta = \text{const} = \theta_0$, because $\cos\theta = Cr/G = Cr_0/G = \text{const}$.

After the corresponding substitutions into the kinematic Euler equations, we integrate them and finally obtain the expressions for the Euler angles:

$$\varphi = \Bigl(r_0 - \frac{K_O}{A}\cos\theta_0\Bigr)t + \varphi_0, \quad \psi = \frac{K_O}{A}t + \psi_0, \quad \theta = \theta_0.$$

The motion determined by these equations is called the *regular precession*.

E2.8.4. Elementary Theory of Gyroscope

A gyroscope is a homogeneous body of rotation rapidly rotating about its symmetry axis; the gyro axis can change its direction in space. Usually, the gyro is made as a massive symmetry body fixed so that one of the points of its axis remains immovable. This, for example, can be done by using the gimbal suspension (Fig. E2.50).

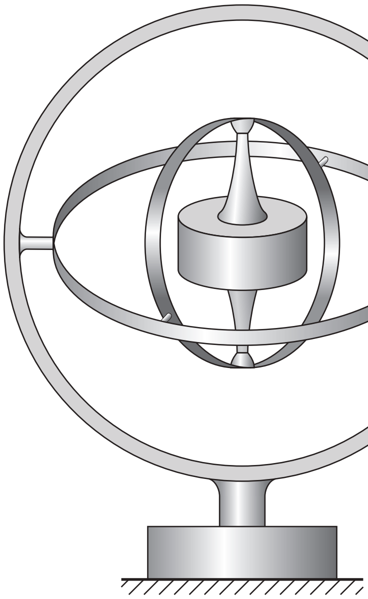

Figure E2.50. Gyroscope.

The exact analytical study of the gyro motion is a very complicated problem. The solution is significantly simplified if the angular velocity of the gyro rotation is sufficiently large.

If $M_O = 0$ for a gyro rotating about the symmetry axis, then the theorem on the angular momentum $d\mathbf{K}_O/dt = \mathbf{M}_O$ implies that $\mathbf{K}_O = I\boldsymbol{\omega}_1 = \mathbf{const}$. This means that the gyro axis preserves its original direction with respect to inertial reference frame, and the angular velocity $\boldsymbol{\omega}_1$ is constant.

If $M_O \neq 0$, then, in addition to proper rotation, the gyro performs precession and nutation motions. Exact studies show that, for a rapidly rotating gyro, the variations in the axis direction due to the nutation oscillations, $\Delta\theta = (\theta_{\max} - \theta_{\min})$, remain in a very small range, and the angular nutation velocity $\dot{\theta}$ is also very small. Therefore, in the approximate theory of gyros, it is assumed that $\theta = \theta_0 = \text{const}$.

The angular precession velocity $\boldsymbol{\omega}_2$ is also small, but it is necessary to take this velocity into account, because it has a constant sign and the gyro axis changes its direction rather noticeably with time.

With the above assumptions taken into account, we have $\boldsymbol{\omega} = \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2 \approx \boldsymbol{\omega}_1$, because $\boldsymbol{\omega}_2 \ll \boldsymbol{\omega}_1$. In what follows, we assume that, at any time instant, the angular velocity and angular momentum vectors are directed along the gyro axis, $\mathbf{K}_O = I\boldsymbol{\omega}_1$ (Fig. E2.51).

Now the theorem on the angular momentum $d\mathbf{K}_O/dt = \mathbf{M}_O$ can be represented as $d(I\boldsymbol{\omega}_1)/dt = \mathbf{M}_O$. The left-hand side of the equation contains the velocity of the endpoint of the vector $I\boldsymbol{\omega}_1$ rotating about the vertical axis at the angular velocity ω_2. Using the Euler formula, we write out the final result:

$$I(\boldsymbol{\omega}_2 \times \boldsymbol{\omega}_1) = \mathbf{M}_O.$$

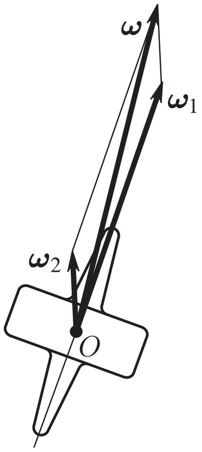

Figure E2.51. Directions of the angular velocity of the gyro proper rotation $\boldsymbol{\omega}_1$ (along the gyro axis) and the angular precession velocity $\boldsymbol{\omega}_2$.

Since the intrinsic angular momentum $I\boldsymbol{\omega}_1$ is known, the last expression allows one to solve the following two problems:

(1) If $\mathbf{M}_O$ is known, find $\boldsymbol{\omega}_2$, i.e., the angular precession velocity.

(2) Conversely, if $\boldsymbol{\omega}_2$ is known, find $\mathbf{M}_O$, i.e., the moment of forces responsible for the precession.

E2.9. Elementary Theory of Impact

As a rule, in the motion of a mechanical system, in each successive small time interval τ, the variations in the velocities of points of the system are also small and have the same order of magnitude. But in several cases, one can observe the phenomenon of *impact* in which the velocities of several points can vary by a constant value in the time interval from t_0 to $t_0+\tau$. This occurs either in the case of instantaneous imposition of a mechanical constraint (for example, when a body falls on a massive plate) or in the case of instantaneous removal of a constraint (for example, when a flying missile bursts).

For such a point, we write the theorem on the momentum in the form $d\,\mathbf{Q} = \sum_k d\,\mathbf{S}_k$ or $d\,(m\mathbf{V}) = \sum_k \mathbf{F}_k\,dt$, perform integration, and obtain

$$\int_{\mathbf{V}_0}^{\mathbf{V}_1} d(m\mathbf{V}) = \int_{t_0}^{t_0+\tau} \sum_k \mathbf{F}_k\,dt \qquad \text{or} \qquad m(\mathbf{V}_1 - \mathbf{V}_0) = \sum_k \widetilde{\mathbf{F}}_k \tau.$$

Here $\mathbf{V}_0$ and $\mathbf{V}_1$ are the velocities of the point before and after the impact, and $\widetilde{\mathbf{F}}_k$ are the average forces acting on the point on impact.

Since the left-hand side of the last equation is a finite quantity, it follows that at least one of the quantities $\widetilde{F}_k$ must be of the order of $1/\tau$ on impact; i.e., since τ is small, it must be much larger than the "ordinary" forces acting on the point. In what follows, such forces are called *impact forces*.

Since the impact processes are very cumbersome, the law of variation of the impact force is, as a rule, unknown. It follows from general considerations that the graph of the impact force magnitude has the shape shown in Fig. E2.52, where the average impact force magnitude is given.

Example 1. Approximate estimates of the impact force in the case of fall of a stone of weight 1 kg on a plate from the height of 10 m give a quantity of the of order of 1.5 tons under the assumptions that the stone

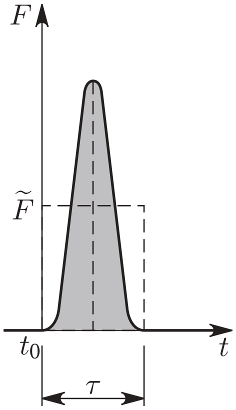

Figure E2.52. Qualitative graph of the impact force magnitude and computation of the average impact force.

after fall does not bounce and $\tau = 0.001$ s. Judging from the figure, the maximum magnitude of the impact force in the time of impact is at least twice larger, i.e., attains the value of the order of 3 tons and exceeds the magnitude of the "ordinary" force of weight approximately by a factor of 3000.

With the above explanations taken into account, the theorem on the momentum of a point can be stated as

$$\mathbf{Q}_1 - \mathbf{Q}_0 = \sum \mathbf{S}_k;$$

i.e., the variation of the momentum of a point in the time of impact is equal to the sum of impact momenta.

The *impact momentum* is the vector $\mathbf{S}_k = \int_{t_0}^{t_0+\tau} \mathbf{F}_k\,dt$. The statement of the theorem contains only the momenta of impact forces, because the momenta of "ordinary" forces can be neglected as being of small order of magnitude.

At the time of impact, the velocity of the point varies according to the relation

$$m\dot{\mathbf{r}} - m\mathbf{V}_0 = \sum \mathbf{S}_k,$$

where $\mathbf{r}$ is the position vector of the point with respect to a fixed center O.

Separating the variables and integrating over the time of impact, we obtain

$$\int_{\mathbf{r}^\circ}^{\mathbf{r}} m\,d\mathbf{r} = \int_{t_0}^{t_0+\tau} \Bigl(m\mathbf{V}_0 + \sum \mathbf{S}_k\Bigr)\,dt,$$

or

$$m(\mathbf{r} - \mathbf{r}^\circ) = \Bigl(m\mathbf{V}_0 + \sum \widetilde{\mathbf{S}}_k\Bigr)\tau,$$

where $\mathbf{r}^\circ$, $\mathbf{r}$, $\widetilde{\mathbf{S}}_k$ are the position vectors of the point at the beginning and at the end of the impact and the average value of the impact momentum. The right-hand side of the equation is a small value of the order of τ, which implies that $\mathbf{r} = \mathbf{r}^\circ$; i.e., the displacement of the point in the time of impact can be neglected.

The last fact allows us to obtain the following theorem.

THEOREM ON THE VARIATION IN THE MOMENT OF MOMENTUM OF A POINT ON IMPACT. *In the time of impact, the variation in the moment of momentum of a point about any fixed center is equal to the sum of moments of impact momenta about the same center:*

$$\mathbf{r} \times \mathbf{Q}_1 - \mathbf{r}^\circ \times \mathbf{Q}_0 = \mathbf{r}^\circ \times \sum \mathbf{S}_k \qquad \textit{or} \qquad \mathbf{K}_o - \mathbf{K}_o^\circ = \sum \mathbf{M}_o \mathbf{S}_k.$$

THEOREM ON THE MOMENTUM OF A MECHANICAL SYSTEM ON IMPACT. *In the time of impact, the variation in the momentum of a system is equal to the sum of external impact momenta:*

$$\mathbf{Q} - \mathbf{Q}^\circ = \sum \mathbf{S}_k^\mathrm{e}.$$

THEOREM ON THE ANGULAR MOMENTUM OF A MECHANICAL SYSTEM ON IMPACT. *In the time of impact, the variation in the angular momentum of the system about any fixed center O is equal to the sum of moments of external impact momenta about the same center:*

$$\mathbf{K}_o - \mathbf{K}_o^\circ = \sum \mathbf{M}_o \mathbf{S}_k^\mathrm{e}.$$

In collision of two bodies, at the point of contact, there arise impact forces applied to each body (Fig. E2.53). These forces (and hence the impact momenta) are equal and opposite, $\mathbf{S}_1 = -\mathbf{S}_2$.

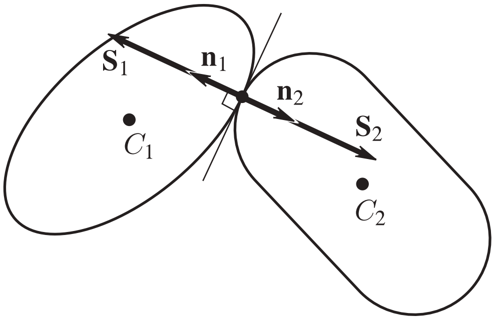

Figure E2.53. Collision of two bodies.

We assume that the forces of friction are negligibly small; in this case, the impact momenta are directed along the common normal to the surfaces of colliding bodies.

As observations show, the hypothesis on the absolute rigidity of bodies under the conditions in question is not satisfied, and therefore, it is necessary to take into account some additional properties.

The impact process can be divided into two phases, the initial and the final. In the initial phase, some constraints are imposed instantaneously, and the deformable bodies approach each other along the line of the common normal. In the final phase of the impact, the constraints are removed instantaneously, the bodies remove from each other and completely or partially restore their original shape.

The complete investigation of the phenomenon under study is far beyond the framework of theoretical mechanics. But here it is also possible to obtain results that are in good agreement with practice if the *Newton's hypothesis* is accepted.

We assume that, in the time of impact, the site of contact of bodies is small and can be assumed to be a point. We measure the relative velocity of approach of the points of contact of the bodies directly before impact and calculate the absolute value of its projection V' on the common normal to the surfaces of the colliding bodies. We perform similar actions when the points of contact separate immediately after impact, and obtain the values of V''.

According to Newton's hypothesis, the quantity $\varepsilon = V''/V'$ characterizes the properties of colliding bodies and is independent of their masses and relative velocity.

This hypothesis can be stated in a different way: the ratio of the magnitudes of momenta in the first and second phases of impact is a constant value for colliding bodies and is a quantity independent of masses and velocities of the bodies.

The physical constant ε is called the *coefficient of elasticity*. It follows from physical considerations that $0 \le \varepsilon \le 1$. For $\varepsilon = 1$, the impact is said to be *absolutely elastic*, for $\varepsilon = 0$, *absolutely inelastic*.

Example 2. The values of the coefficient of elasticity can be obtained experimentally. A heavy ball raised to a height h_0 over a massive horizontal plate is lowered without initial velocity, and the maximal height h_1 of its raising after rebound is measured (Fig. E2.54).

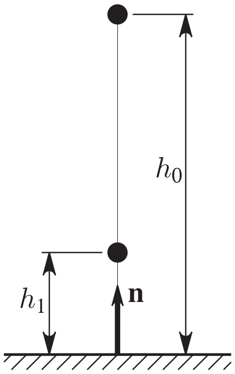

Figure E2.54. Experimental determination of the coefficient of elasticity by measuring the height of the ball rebound from the surface.

Neglecting the air resistance, we determine the velocity of fall on the plate V_0 and the rebound velocity V_1 from the relations $V_0 = \sqrt{2gh_0}$ and $V_1 = \sqrt{2gh_1}$. As a result, the formula for the coefficient of elasticity acquires the form

$$\varepsilon = \frac{V_1}{V_0} = \sqrt{\frac{h_1}{h_2}}.$$

Example 3 (Direct central impact of two bodies). The impact of two bodies in which the vectors of internal impact momenta and the velocities of the centers of mass are directed along the straight line connecting the centers of mass is called a *direct central impact* (Fig. E2.55). We assume that the velocities of the centers of mass of the bodies before impact and the coefficient of elasticity are known. It is required to find the velocities of centers of mass of the bodies after impact and the impact momenta.

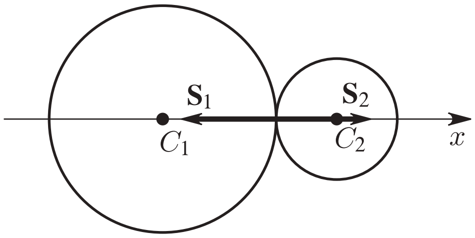

Figure E2.55. Direct central impact of two moving bodies (of different masses).

Solution. We denote the velocities of the centers of mass of the bodies before impact by V_1 and V_2 and after impact by V_1' and V_2'. Since the first body overtakes the second body, we have $V_1 > V_2$, $V_1' \le V_2'$ before impact and $\frac{V_2'-V_1'}{V_1-V_2} = \varepsilon$ after impact.

Since the acting impact momenta are internal, the law of conservation of momentum on impact holds, which, in projections onto the direction of the x-axis, gives the first equation

$$m_1V_1 + m_2V_2 = m_1V_1' + m_2V_2'.$$

In addition, for a known coefficient of elasticity, there is a second equation relating the unknowns V_1' and V_2':

$$\varepsilon(V_1 - V_2) = V_2' - V_1'.$$

By solving the resulting system of two equations, we obtain

$$V_1' = V_1 + \frac{m_2(1+\varepsilon)(V_2 - V_1)}{m_1 + m_2}, \quad V_2' = V_2 + \frac{m_1(1+\varepsilon)(V_1 - V_2)}{m_2 + m_1}.$$

Now we find the impact momentum S_1 applied to the first body:

$$S_1 = m_1(V_1' - V_1) = -\frac{m_1m_2(1+\varepsilon)(V_1 - V_2)}{m_1 + m_2}.$$

The impact momentum S_2 applied to the second body is related to S_1 as $S_2 = -S_1$.

Example 4 (Action of the impact on a body rotating about a fixed axis). To a body rotating with angular velocity ω about the vertical fixed axis AB, the impact momentum $\mathbf{S}$ is applied at point M (Fig. E2.56). In cylindrical hinges A, B and the thrust bearing, there arise impact momenta $\mathbf{S}_{Ax}, \mathbf{S}_{Ay}, \mathbf{S}_{Az}, \mathbf{S}_{Bx}, \mathbf{S}_{By}$, which must be found under given dynamic characteristics of the body.

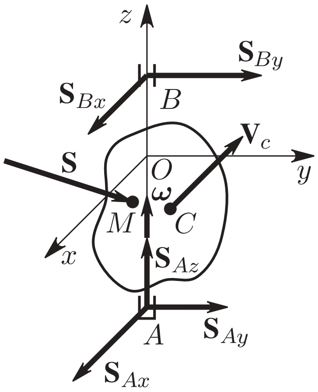

Figure E2.56. Impact on a body rotating about a fixed axis.

Solution. To make the further calculations more convenient, we place the coordinate axes so that the axis Oz coincides with the axis of rotation ($x_A = y_A = x_B = y_B = 0$; $\omega_x = \omega_y = 0$) and the center of mass of the body C lie in the plane yAz ($y_C = 0$); in addition, we choose the origin, i.e., point O, so that the point of application of the impact momentum M lies in the plane xOy ($z_M = 0$).

The theorem on the variation in the momentum of the body and the theorem on the variation in its angular momentum about the center O acquire the form

$$m(\mathbf{V}'_C - \mathbf{V}_C) = \mathbf{S} + \mathbf{S}_A + \mathbf{S}_B,$$
$$\mathbf{K}'_O - \mathbf{K}_O = \mathbf{r}_M \times \mathbf{S} + \mathbf{r}_A \times \mathbf{S}_A + \mathbf{r}_B \times \mathbf{S}_B,$$

where $\mathbf{S}_A$ and $\mathbf{S}_B$ are the momenta of the constraint reaction forces, $\mathbf{r}_M$ is the position vector of point M, and $\mathbf{V}_C$, $\mathbf{K}$ and $\mathbf{V}'_C$, $\mathbf{K}'_A$ are the velocity of the center of mass and the angular momentum of the body before and after the impact.

To two vector equations, there corresponds a system of six scalar equations

$$\begin{aligned} S_x + S_{Ax} + S_{Bx} &= my_C(\omega - \omega'),\\ S_y + S_{Ay} + S_{By} &= 0,\\ S_z + S_{Az} &= 0,\\ -I_{xz}(\omega' - \omega) &= y_M S_z - z_A S_{Ay} - z_B S_{By},\\ -I_{yz}(\omega' - \omega) &= -x_M S_z + z_A S_{Ax} + z_B S_{Bx},\\ I_{zz}(\omega' - \omega) &= x_M S_y - y_M S_x, \end{aligned}$$

with six unknowns ω', S_{Ax}, S_{Ay}, S_{Az}, S_{Bx}, S_{By}.

From the sixth equation, we obtain the body angular velocity after the impact:

$$\omega' = \omega + \frac{1}{I_{zz}}(x_M S_y - y_M S_x),$$

and then find the impact momenta of the constraint reactions.

Now we obtain the conditions under which there are no impact momenta of the constraint reactions. In this case, the point of application of the impact momentum is called the *center of impact*.

If there are no impact momenta of the supports $S_{Ax} = S_{Ay} = S_{Az} = S_{Bx} = S_{By} = 0$, then the system of the above-obtained equations becomes

$$\begin{aligned} S_x &= my_{\tilde{N}}(\omega' - \omega),\\ S_y &= 0,\\ S_z &= 0,\\ -I_{xz}(\omega' - \omega) &= y_M S_z,\\ -I_{yz}(\omega' - \omega) &= -x_M S_z,\\ I_{zz}(\omega' - \omega) &= x_M S_y - y_M S_x. \end{aligned}$$

This implies that the supports do not experience impact loads if

1. The external impact momentum is perpendicular to the plane passing through the center of mass and the axis of rotation of the body, because it follows from the second and third equations that $S_y = 0$ and $S_z = 0$.
2. The axis of the body rotation is a principal axis of inertia for the point O at which it intersects the orthogonal plane containing the point of application of the impact momentum, because it follows from the fourth and fifth equations that $I_{xz} = I_{yz} = 0$.
3. The position of the impact center is determined by the formula $y_M = -I_{zz}.my_C$, which follows from the first and sixth equations.

The theorem on the variation in the kinetic energy of a material system on impact (the Carnot theorem) is formulated for two cases: instantaneous imposition of constraints and instantaneous removal of constraints.

THEOREM (CARNOT).

(i) *The kinetic energy lost by the system in instantaneous imposition of absolutely inelastic constraints on it is equal to the kinetic energy that the system would have if its points moved with the lost velocities:*

$$T - T' = \sum_{\nu} \frac{1}{2} m_\nu (\mathbf{v}_\nu - \mathbf{v}'_\nu)^2;$$

this is the case in which the impact has only the first phase.

(ii) *The kinetic energy acquired by the system in instantaneous removal of constraints is equal to the kinetic energy that the system would have if its points moved with acquired velocities:*

$$T' - T = \sum_{\nu} \frac{1}{2} m_\nu (\mathbf{v}'_\nu - \mathbf{v}_\nu)^2;$$

this is the case where the impact has only the second phase.

Bibliography for Chapter E2

Arnold, V. I., *Mathematical Methods in Classical Mechanics, 2nd Edition*, Springer-Verlag, New York, 1989.
Bat, M. J., Ganelidze, G. U., and Kelson, A.S., *Theoretical Mechanics in Examples and Problems, Vols. 1 and 2* [in Russian], Nauka, Moscow, 1990 and 1991.
Bottema, O. and Roth, B., *Theoretical Kinematics*, Dover Publications, New York, 1990.
Bukhgolts, N. N., *Basic Course of Theoretical Mechanics, Parts 1 and 2* [in Russian], Nauka, Moscow, 1972.
Butenin, N. V. and Fufaev, N. A., *Introduction to Analytical Mechanics* [in Russian], Nauka, Moscow, 1991.
Gantmakher, F. R., *Lectures in Analytical Mechanics*, Mir Publ., Moscow, 1975.
Greenwood, D. T., *Classical Mechanics* [reprint of 1977 ed.], Courier Dover Publ., New York, 1997.
Gregory, D., *Classical Mechanics: An Undergraduate Text*, Cambridge University Press, Cambridge, England, 2006.
Hibbeler, R. C., *Engineering Mechanics, Static, 11th Edition*, Prentice Hall, Upper Saddle River, New Jersey, 2006.
Kibble, T. W. B. and Berkshire, F. H., *Classical Mechanics*, Imperial College Press, England, 2004.
Landau, L. D. and Lifshitz, E. M., *Mechanics, 3rd Edition*, Butterworth-Heinemann, Oxford, England, 1976.
Lurie, A. I., *Analytical Mechanics*, Springer, Berlin, 2002.
Marion, J. B. and Thornton, S. T., *Classical Dynamics of Particles and Systems, 5th Edition*, Brooks Cole, Pacific Grove, California, 2003.
Mescherskii, I. V., *A Collection of Problems of Mechanics* [in Russian], Nauka, Moscow, 1986.
Morin, D., *Introduction to Classical Mechanics: with Problems and Solutions*, Cambridge University Press, Cambridge, England, 2008.
Targ, S., *Theoretical Mechanics: a Short Course*, Gordon and Breach, New York, 1967.
Taylor, J. R., *Classical Mechanics*, University Science Books, Sausalito, California, 2005.

Chapter E3

Elements of Strength of Materials

E3.1. Basic Notions

E3.1.1. Introduction. External and Internal Forces

► **Deformation, strength, and rigidity.** The strength of materials is the field of mechanics where design and analysis of structural members for strength, rigidity, and stability are considered.

The strength of materials is based on the knowledge accumulated in theoretical mechanics. The difference is that the object in theoretical mechanics is an absolutely rigid body, but the strength of materials deals with deformable rigid bodies.

In practice, the actual members of machines and structures experience the action of different forces. Under the action of these forces, the bodies are *deformed*; i.e., the mutual position of the material particles is changed. If the forces are sufficiently large, then the body fracture may occur.

The ability of a body to take loads without fracture and large deformations is called the *strength* and *rigidity*, respectively.

Several equilibria of bodies and structures are *unstable*; i.e., in these states, any negligible actions of random character, as a rule, may lead to significant deviations from these states. But if the deviations are also small, then such equilibria are said to be *stable*.

► **External forces.** The external forces acting on a structure comprise the active forces (loads) and the reactions of external constraints. Several types of loads are distinguished.

Lumped force applied at a point. They are introduced instead of actual forces acting on a small site of the surface of a structural member whose dimensions can be neglected.

Distributed forces. For example, the forces of fluid pressure on the vessel bottom are loads distributed over the surface and are measured in N/m^2, and the weight forces are loads distributed over the volume and measured in N/m^3. In several cases, a load distributed over a line is introduced; the intensity of such a load is measured in N/m.

One version of the load is the *lumped torque* (couple of forces).

► **Internal forces in a rod.** The rod is the most widespread structural member; therefore, the main attention in the strength of materials is paid to the rod.

The longitudinal axis and the transverse cross-section are the basic geometric elements of the rod. It is assumed that the rod transverse cross-sections are perpendicular to the longitudinal axis, and the longitudinal axis passes through the centers of gravity of the transverse cross-sections.

The *internal forces* of a rod are the forces of interaction between its separate parts arising under the action of external forces (it is assumed that, in the absence of external forces, the internal forces are zero).

We consider a rod in equilibrium under the action of a system of external forces (Fig. E3.1a). We mentally draw an arbitrary transverse cross-section dividing the rod into two parts L and R. The right part R is subjected to a system of forces exerted the left part L and distributed over the surface of the transverse cross-section, i.e., a system of internal forces with respect to the entire rod. This system can be reduced to the resultant vector **R** and the resultant moment **M** if the center of gravity of the cross-section, i.e., point O, is taken to be the center of reduction.

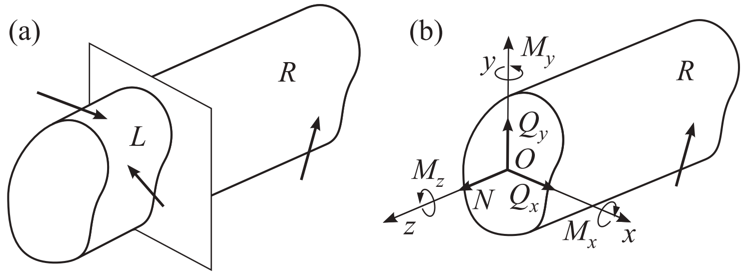

Figure E3.1. Internal forces in a rod: (a) (imaginary) cross-section of the rod, (b) internal forces and moments at the cross-section.

► **Internal force factors.** We choose the coordinate system so that the axes x, y lie in the transverse cross-section and the axis z is perpendicular to it and decompose **R** and **M** into components in these axes: Q_x, Q_y, N and M_x, M_y, M_z (Fig. E3.1b).

These six quantities are called *internal force factors of the rod (or internal forces)* in the cross-section under study. Each of these forces has its own name corresponding to its direction or the specific form of the rod deformation caused by these forces. The forces Q_x and Q_y are called the *transverse* (*cross-axis*) *forces*, and N is called the *normal* (*longitudinal*) *force*. The moments M_x and M_y are called the *bending moments*, and M_z is called the *twisting moment*.

► **Method of sections.** Since the cut-off part of the rod *R* is in equilibrium, it is possible to compose six static equations for the forces acting on this part, which determine all six internal force factors:

$$\sum F_x = 0, \qquad \sum F_y = 0, \qquad \sum F_z = 0,$$
$$\sum m_x(\mathbf{F}) = 0, \quad \sum m_y(\mathbf{F}) = 0, \quad \sum m_z(\mathbf{F}) = 0.$$

This method for determining the internal forces is called the *method of sections*.

On the basis of the law of action and reaction, the right-hand part of the rod acts on the left-hand part with equal but opposite forces; therefore, they can also be determined starting from the fact that the rod part *L* is in equilibrium.

E3.1.2. Stresses and Strains at a Point

► **Stresses.** The *stress vector* p is the intensity of internal forces distributed over the cross-section, at a point of the cross-section (Fig. E3.2). Its components lying in the plane of the cross-section are called *tangential (shear) stresses* τ_x, τ_y, and the component perpendicular to the cross-section is called the *normal stress* σ. The stresses are measured in $\mathrm{N/m^2}$ (Pa) and depend not only on the choice of the point but also on the orientation of the cross-section (or of the site dA, Fig. E3.2) passing through this point. The complete set of stresses at a given point for different sites is called the *stress state* at this point.

If the stresses in the rod cross-section are known, then its internal force factors can be found by the formulas

$$\begin{aligned} &N = \int_A \sigma\, dA, &&Q_x = \int_A \tau_x\, dA, &&Q_y = \int_A \tau_y\, dA,\\ &M_x = -\int_A \sigma y\, dA, &&M_y = -\int_A \sigma x\, dA, &&M_z = \int_A (\tau_y x - \tau_x y)\, dA, \end{aligned} \tag{E3.1.2.1}$$

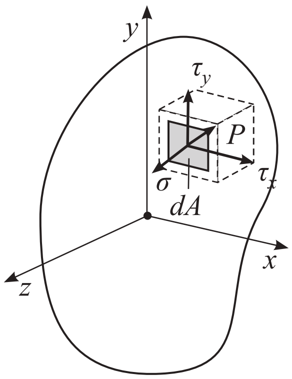

Figure E3.2. Tangential and normal stresses.

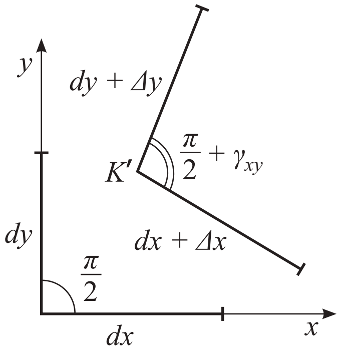

Figure E3.3. Deformation at a point K.

where the integrals are taken over the entire cross-section area A.

► **Strains.** Consider an arbitrary point K of the body in the initial unstrained state and draw two infinitely small segments of lengths dx and dy through this point in the direction of the axes x and y (Fig. E3.3).

Under the action of the load, the body is strained, the point K moves to the point K', the segments dx and dy change their lengths by Δx and Δy, and the angle $\pi/2$ between them varies by γ_{xy}.

The ratios

$$\varepsilon_x = \frac{\Delta x}{dx}, \quad \varepsilon_y = \frac{\Delta y}{dy}$$

are called the *linear strains* at point K in the direction of the axes x and y, respectively, and the quantity γ_{xy} is called the *angular strain* (shear angle) at point K between the axes x and y. The complete set of linear and angular strains in different directions passing through the point under study is called the *strain state* at this point.

E3.1.3. Basic Notions and Hypotheses

► **Elasticity** is the property of a body that is expressed as a unique dependence between the forces acting on the body and the body strains. In particular, an elastic body returns to the initial state after the loads are removed.

In several cases, after the loads are removed, only part of the total strain disappears, and this part is said to be *elastic*. The remaining part, the so-called *residual* (*plastic*) strain, is related to the body property called *plasticity*.

► **Basic hypotheses accepted in the strength of materials:**

1. The bodies are assumed to be solid (without hollows) and homogeneous; i.e., the properties of the body material are the same at different points.

2. The body material is *isotropic*; i.e., its properties are the same in all directions. (In several cases, it is necessary to abandon this assumption for *anisotropic* bodies, whose material properties are different in different directions. For example, the properties of wood in the directions along the fibers and across them are different.)

3. The body strains at each point are directly proportional to the stresses at this point. This property is called the *linear elasticity law* or *Hooke's law*.

4. It is assumed that the strains in the body, as well as the displacements of its particles, are small compared with the geometric dimensions of the body itself.

5. The *principle of independence of forces* (principle of superposition) holds, according to which any quantity a depending on the action of several forces is equal to the sum of

quantities a_i determined for each separate force. This principle follows from hypotheses 3 and 4.

E3.2. Stress–Strain State at a Point

E3.2.1. Types of Stress State

► **Principal stresses and principal directions.** An analysis of the stress state at a point K of the body becomes more convenient if we mentally separate an elementary parallelepiped around this point and consider the stresses acting on its faces. Since the parallelepiped is small, we can assume that the stress state is the same at all of its points (is homogeneous) and coincides with the stress state at point K under study. Varying the orientation of this parallelepiped, we can make all its faces free of shear stresses. The corresponding normal stresses are called the *principal stresses*, and their directions are called the *principal directions* at point K.

The *linear, plane, volumetric (uniaxial, biaxial, and triaxial)* stress states at a point are distinguished depending on whether the parallelepiped oriented in the principal directions experiences extension (or compression) respectively in one, two, or three mutually perpendicular directions.

► **Plane stress state. Reciprocity law.** In what follows, because of its special importance, we consider only the case of plane stress state in which there are no stresses on two opposite faces of an elementary parallelepiped. The other faces, in the general case of their orientation, are under the action of the shear and normal stresses (Fig. E3.4). Since the parallelepiped stress state is homogenous, it follows that the similar stresses on opposite faces are numerically equal. Obviously, the normal forces are mutually balanced on the parallelepiped faces. The tangential forces on the same faces form two couples of forces with moments $(\tau_y\,dy\,dx)\,dz$ and $(\tau_z\,dz\,dx)\,dy$ of opposite direction, where dx, dy, dz—are the parallelepiped dimensions. These moments must be balanced, which implies $\tau_z = \tau_y$.

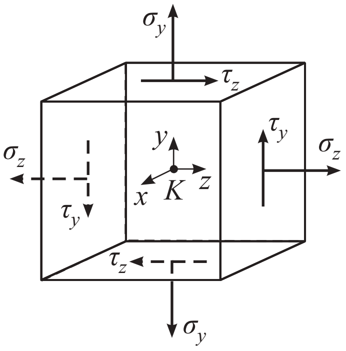

Figure E3.4. Plane stress state at a point.

The last equality expresses the *reciprocity law for shear stresses*: on any mutually perpendicular sites, the shear stresses are equal in value and directed so that they tend to rotate the element in opposite directions (Fig. E3.4). The aforesaid allows one to introduce the unique notation: $\tau = \tau_z = \tau_y$.

E3.2.2. Uniaxial Tension and Compression

► **Modulus of elasticity of the first kind.** We consider an elementary parallelepiped in the state of uniaxial tension or compression with only normal stresses acting on its two opposite faces (Fig. E3.5).

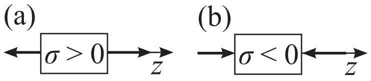

Figure E3.5. (a) Uniaxial tension and (b) uniaxial compression.

Rule of signs: the tensile stresses are assumed to be positive, and the compressing stresses are assumed to be negative. According to Hooke's law, the stress and the linear strain in the direction of the axis z are directly proportional:

$$\sigma = E\varepsilon_z. \tag{E3.2.2.1}$$

The proportionality coefficient E is called the *modulus of elasticity of the first kind* and has the dimension of stress, because ε_z is a dimensionless variable. For $\sigma > 0$ (tension), ε_z is also greater than zero, which corresponds to the parallelepiped elongation in the direction of the axis z. For $\sigma < 0$ (compression), we have $\varepsilon_z < 0$, which corresponds to the parallelepiped shortening.

► **Poisson coefficient.** Because of symmetry, the linear strains in the directions of the axes x and y are equal to each other and also proportional to the stress, or, with (E3.2.2.1) taken into account, to the strain ε_z:

$$\varepsilon_x = \varepsilon_y = -\nu\varepsilon_z = -\frac{\nu\sigma}{E}. \tag{E3.2.2.2}$$

The constant dimensionless proportionality coefficient ν is called the *Poisson ratio*, and the minus sign takes into account the opposite signs of the longitudinal strain ε_z and transverse strains ε_x and ε_y. In particular, for $\sigma > 0$ and $\varepsilon_z > 0$, the quantities ε_x and ε_y are less than zero, and hence the parallelepiped transverse dimensions decrease.

The constant quantities E and ν characterize the elastic properties of a specific material; for example, for steel we have $E = 210$ GPa, $0.25 \le \nu \le 0.30$.

E3.2.3. Simple Shear

► **Simple shear.** Consider a specific form of plane stress state in which the lateral faces of an elementary parallelepiped are only under the action of shear stresses τ (Fig. E3.6a). By the reciprocity law, the stresses on neighboring faces are the same. Such a form of stress state is called the *simple shear*.

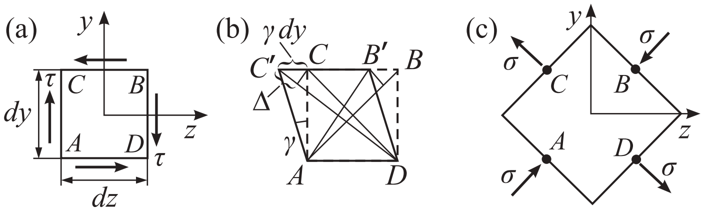

Figure E3.6. Pure shear: (a) initial state (tangential stresses are only acting on the lateral face), (b) deformation of the elementary parallelepiped, (c) normal stresses acting on the elementary parallelepiped rotated by 45°.

The character of the parallelepiped strain is shown in Fig. E3.6b. Initially, the right angles between the lateral faces vary by the quantity γ, which is called the shear angle, and the linear strains ε_x, ε_y, ε_z are zero.

► **Shear modulus.** In this case, Hooke's law has the form

$$\tau = G\gamma. \tag{E3.2.3.1}$$

The proportionality coefficient G is called the *shear modulus* or the *modulus of elasticity of the second kind*. The modulus dimension coincides with the dimension of stresses.

It is easy to see that, for $dy = dz$, the diagonal segment AB in the parallelepiped becomes shorter and the segment CD elongates by the same value $\Delta = \gamma\, dy \sin 45^\circ$ (Fig. E3.6b); therefore, the other elementary parallelepiped, distinguished in a neighborhood of the same point and rotated by the angle 45° with respect to the first parallelepiped, must be under the action of the corresponding normal stresses σ (Fig. E3.6c). In this case, there are already no shear stresses, and hence the normal stresses σ are the principal stresses, and their directions are the principal directions for a given stress state.

► **Relation between G, E, and ν.** Consider the element formed by the diagonal cut of the parallelepiped (see Fig. E3.7). Its faces are under the action of forces whose values are equal to $\tau\, dx\, dy$, $\tau\, dx\, dz$, and $\sigma\, dx\, dz/\cos 45^\circ$.

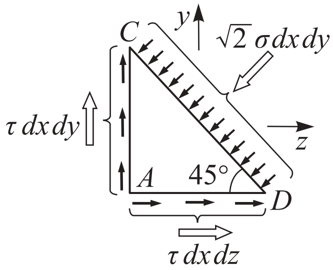

Figure E3.7. Equilibrium of the element formed by the diagonal cut of the parallelepiped.

Since this element is in equilibrium, we obtain

$$\sigma = \tau. \tag{E3.2.3.2}$$

With (E3.2.3.1) and (E3.2.3.2) taken into account, the strain of segment CD (Fig. E3.6b) is equal to $\varepsilon_{CD} = \Delta/CD = \frac{1}{2}\gamma = \frac{1}{2}\tau/G = \frac{1}{2}\sigma/G$.

On the other hand, treating the state in Fig. E3.6c as a combination of two uniaxial stress states (extension and compression in perpendicular directions), we obtain $\varepsilon_{CD} = \dfrac{\sigma}{E} + \dfrac{\nu\sigma}{E}$ form Hooke's law (E3.2.2.1), (E3.2.2.2). This implies $\dfrac{\sigma}{2G} = \dfrac{\sigma}{E} + \dfrac{\nu\sigma}{E}$ or $G = \dfrac{E}{2(1+\nu)}$.

Thus, the shear modulus G is determined via the already introduced constants E and ν.

E3.3. Central Tension and Compression

E3.3.1. Longitudinal Force

The *central tension (compression)* is a type of a rod deformation such that only a longitudinal force N arises in the rod transverse cross-sections and all the other internal forces are zero.

Rule of signs: the tensile longitudinal forces are assumed to be positive, and the compressing longitudinal forces are assumed to be negative.

Example. In Fig. E3.8a we show a rod loaded by two forces F_1 and F_2 lying on its axis. The longitudinal force will be determined by the method of sections (see Subsection E3.1.1). We draw an arbitrary cross-section I on segment "*a*" of the rod and consider the equilibrium of the free cut-off part (Fig. E3.8b). We replace the action of the cut-off part by an unknown force N_1 assuming that it is positive. The equilibrium equation for this cut-off part, stating that the sum of projections of all forces on the axis z is zero, i.e., $N_1 - F_1 = 0$, implies

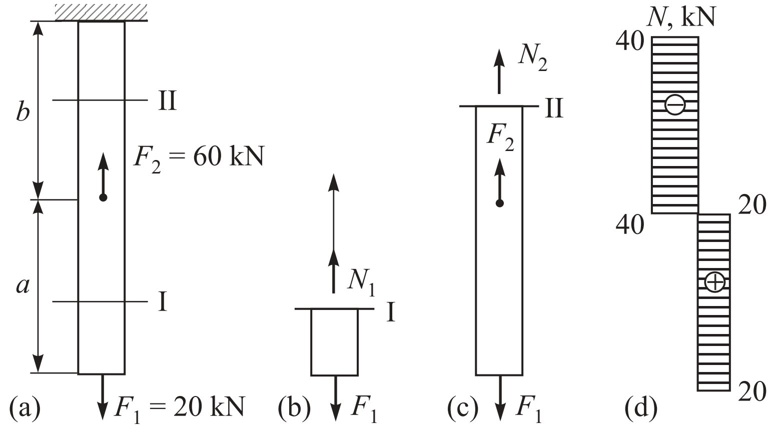

Figure E3.8. Central tension (compression) of a rod: (a) the rod with two forces applied, (b) equilibrium of the cut-off part to cross-section I, (c) equilibrium of the cut-off part to cross-section II, (d) longitudinal force distribution diagram.

that $N_1 = F_1 = 20\,\text{kN}$. Similarly, for the cross-section II on segment "*b*" of the rod, we obtain (Fig. E3.8c) $N_2 + F_2 - F_1 = 0$, which implies $N_2 = F_1 - F_2 = -40\,\text{kN}$.

Thus, the longitudinal force on segments "*a*" and "*b*" of the rod is different. Segment "*a*" of the rod experiences tension ($N > 0$), and segment "*b*" is under compression ($N < 0$). In transition from segment "*a*" to segment "*b*," the force N experiences a jump by the value of the force F_2.

To illustrate the character of variation in the longitudinal force along the rod, one usually constructs the graph of the function $N(z)$, which is called the axial force *diagram*. This graph presents numerical values of the longitudinal (axial) force in typical cross-sections and their signs. For the example under study, the diagram of $N(z)$ is shown in Fig. E3.8d.

E3.3.2. Stresses and Strains under Tension or Compression

► **Flat cross-section hypothesis.** The stresses under tension or compression will be determined by using the *flat cross-section hypothesis*: after strain, the rod transverse cross-sections remain plane and perpendicular to the rod axis. Experiments show that this hypothesis is violated only in the regions of the so-called local stresses, immediately near the points of application of external forces or sites of sharp variation in the area of the transverse cross-section, where the so-called *stress concentration* occurs. Taking them into account is a special problem, and we do not consider it here.

► **Formulas for stresses and strains.** Consider a rod under the action of two tensile forces (Fig. E3.9a). According to the flat cross-section hypothesis and the fact that the longitudinal force is constant along the rod ($N = F$), the stress–strain state of the rod points outside a neighborhood of the endpoints is homogeneous, and the normal stresses are uniformly distributed over the transverse cross-section (Fig. E3.9b).

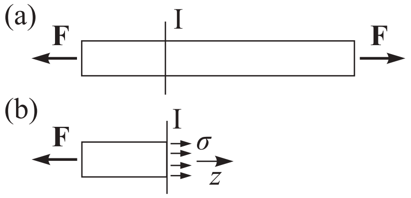

Figure E3.9. (a) A rod under tension by two forces and (b) uniform stress distribution across the cross-section.

It follows from formula (E3.1.2.1) that $N = \int_A \sigma\,dA = \sigma A$, which implies

$$\sigma = \frac{N}{A}. \tag{E3.3.2.1}$$

Formula (E3.3.2.1) permits calculating the normal stresses in the rod transverse cross-sections in the case of central tension or compression from the known longitudinal force N and the known area A of the transverse cross-section.

The elementary parallelepiped separated in the rod is with the conditions of uniaxial tension (Fig. E3.5a). Its strains are determined by Hooke's law (see Subsection E3.2.2). In particular, the longitudinal strain is

$$\varepsilon_z = \frac{\sigma}{E} = \frac{N}{EA}.$$

Summing the elongations of small elements over the entire length of the rod, we obtain its *absolute elongation*

$$\Delta l = \int_0^l \varepsilon_z\,dz = \varepsilon_z l = \frac{Nl}{EA},$$

where l is the rod length.

The quantity EA is called the *rigidity of the rod cross-section under tension and compression.*

E3.3.3. Strength Analysis under Tension and Compression

► **Admissible stress.** To ensure the strength of a rod under tension or compression, it is necessary to bound the maximum stresses of this rod by a value called the *admissible stress* $[\sigma]$. Its value is chosen on the basis of the unsafe stress value, which is introduced in advance for a given material, with the safety margin taken into account. For materials that differently resist tension and compression, the two different admissible stresses $[\sigma^{\rm t}]$ and $[\sigma^{\rm c}]$ are introduced.

► **Problems for strength analysis.** Thus, the strength condition in this case has the form

$$\sigma_{\max} = \frac{|N|}{A} \le [\sigma].$$

Using this condition, one can solve the following problems:

1. From given loads, cross-section dimensions, and the value of admissible stress, verify the rod strength.

2. From given loads and a known value of admissible stress, find the dimensions of the transverse cross-section:

$$A \ge \frac{|N|}{[\sigma]}.$$

3. From given cross-section dimensions and a known value of admissible stress, find the value of admissible longitudinal force:

$$N \le [\sigma]A.$$

Example. For a cast-iron rod (Fig. E3.8a) with cross-section area $A = 5\ \text{cm}^2$, it is necessary to verify the strength conditions for $[\sigma^{\rm t}] = 30\,\text{MPa}$, $[\sigma^{\rm c}] = 100\,\text{MPa}$.

Solution. On segment "*a*," the rod is under the action of the tensile force $N_1 = 20\,\text{kN}$ (see Fig. E3.8d); therefore, the normal stress in the cross-sections on this segment is $\sigma_1 = N_1/A = 40\,\text{MPa}$.

On segment "*b*," the rod is under compression, $N_2 = -40\,\text{kN}$. The corresponding stress is $\sigma_2 = |N_2|/A = 80\,\text{MPa}$.

Since the admissible stresses under tension and compression for cast iron are different, it is necessary to verify two conditions, $\sigma_1 \le [\sigma^{\rm t}]$ and $\sigma_2 \le [\sigma^{\rm c}]$. The first condition is not satisfied, and hence the strength condition is not satisfied in the whole.

E3.4. Torsion

E3.4.1. Twisting Moment

The *torsion* is a type of rod deformation such that only the twisting moment M_z arises in the rod transverse cross-sections and all the other internal force factors are zero.

Rule of signs: if observed on the side of the external normal to the cross-section, the twisting moment seems to be anticlockwise, then it is assumed to be positive.

Example. Consider a rod loaded by two moments m_1 and m_2 such that the planes of their action are perpendicular to the rod axis (Fig. E3.10a). According to the method of sections, we draw an arbitrary transverse cross-section *I* on segment "*a*" and consider the free cut-off part of the rod in equilibrium (Fig. E3.10b). We replace the action of the cut-off part by an unknown twisting moment M_{z1} assuming that its direction is positive. The equilibrium equation for the cut-off part, stating that the sum of moments of all forces about the axis z is zero, i.e., $M_{z1} - m_1 = 0$, implies that $M_{z1} = m_1 = 30\,\text{kN m}$. Similarly, for cross-section *II* we have (Fig. E3.10c) $M_{z2} + m_2 - m_1 = 0$, which implies $M_{z2} = m_1 - m_2 = -10\,\text{kN m}$.

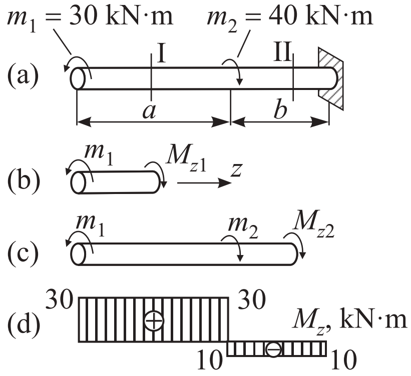

Figure E3.10. Twisting moment: (a) a rod with two moments applied, (b) equilibrium of the cut-off part to cross-section I, (c) equilibrium of the cut-off part to cross-section II, (d) twisting moment diagram.

In Fig. E3.10d, we present the twisting moment diagram, which illustrates the character of variation in M_z along the rod axis. On the boundary between segments "*a*" and "*b*" of the rod, the twisting moment experiences a jump by the value of the moment m_2.

E3.4.2. Stresses and Strains in Torsion

► **Stress state in torsion.** We consider only rods whose transverse cross-sections are of a disk or annulus shape.

To find the law of stress distribution over the cross-section, we assume, which is confirmed experimentally, that the result of deformation is that the rod transverse cross-sections in torsion rotate about the longitudinal axis as rigid disks (annuli). The applicability conditions for this assumption are similar to the applicability conditions of the flat cross-section hypothesis under tension or compression (see Subsection E3.3.2).

Consider a rod experiencing torsion under the action of two moments applied at its ends (Fig. E3.11a). We divide it into coaxial tubes (Fig. E3.11b) and distinguish one of them with the internal radius ρ and infinitely small thickness $d\rho$ (Fig. E3.11c).

By the above assumption about the rod strain character, we assume that an infinitely small element $ACBD$ of the tube (Fig. E3.11c) experiences a strain corresponding to simple shear (see Subsection E3.2.3). This implies that only the shear stresses τ act on the lateral faces of the element (see Fig. E3.6a).

► **Polar moment of inertia.** The shear angle of an element is $\gamma = BB'/dz = \rho\, d\varphi/dz$ (Fig. E3.11c), where $d\varphi/dz$ is the *running angle of torsion* of the tube or the angle of torsion

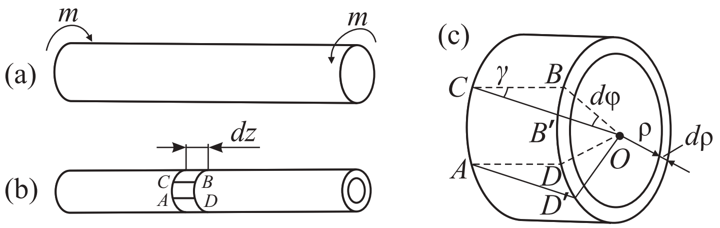

Figure E3.11. Stresses and strains in torsion: (a) a rod with two moments applied, (b) a tube element, (c) deformation of the tube element.

per unit length. It follows from formula (E3.4.2.1) (Hooke's law) that

$$\tau = G\gamma = G\rho\frac{d\varphi}{dz}. \qquad \text{(E3.4.2.1)}$$

The value of the running angle of torsion $d\varphi/dz$ is the same for all coaxial tubes comprising the rod. Therefore, it follows from formula (E3.4.2.1) that the shear stresses in the rod cross-sections are directly proportional to the distance ρ to its axis (Fig. E3.12).

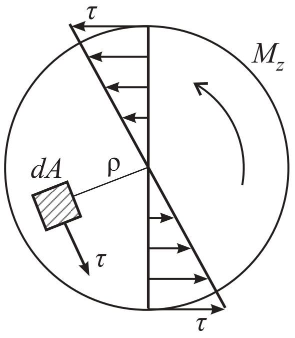

Figure E3.12. Tangential stresses in rod cross-sections.

The twisting moment M_z can be obtained by summing the moments of all stresses distributed over the cross-section about the axis z:

$$M_z = \int_A \tau\rho\,dA = GI_\text{p}\frac{d\varphi}{dz}, \quad \text{where} \quad I_\text{p} = \int_A \rho^2\,dA. \qquad \text{(E3.4.2.2)}$$

The integral I_p over the cross-section surface is called the *polar moment of inertia of the cross-section* and is a geometric characteristic of this cross-section.

For a disk of diameter D, we have $I_\text{p} = \frac{\pi}{32}D^4$.

For an annulus, we have $I_\text{p} = \frac{\pi}{32}(D^4 - d^4)$, where d and D are the outer and inner diameters, respectively.

► **Shear stresses. Angle of torsion.** Using (E3.4.2.2), we obtain

$$\frac{d\varphi}{dz} = \frac{M_z}{GI_\text{p}}. \qquad \text{(E3.4.2.3)}$$

Substituting this expression into formula (E3.4.2.1) for τ, we obtain the basic formula for shear stresses in the rod cross-sections in torsion

$$\tau = \frac{M_z\rho}{I_\text{p}}. \qquad \text{(E3.4.2.4)}$$

The *angle of torsion* of the entire rod in Fig. E3.11a—i.e., the angle of mutual rotation of its end cross-sections—can be found if we know the running angle of torsion (E3.4.2.3) and take into account the fact that the value of M_z is constant along the rod axis ($M_z = m$):

$$\Delta\varphi = \int_0^l \frac{d\varphi}{dz} = \frac{M_z l}{GI_\text{p}},$$

where l is the rod length.

The quantity GI_p is called the *cross-section rigidity in torsion* or *torsional rigidity*.

E3.4.3. Strength Analysis in Torsion

The value of the maximum shear stress in torsion is found from formula (E3.4.2.4) for $\rho = \rho_\text{max}$:

$$\tau_\text{max} = \frac{|M_z|\rho_\text{max}}{I_\text{p}} = \frac{|M_z|}{W_\text{p}}. \qquad \text{(E3.4.3.1)}$$

The quantity $W_\text{p} = I_\text{p}/\rho_\text{max}$ is called the *polar moment of resistance of the cross-section*.

For a disk, we have $W_\text{p} = \frac{\pi}{16}D^3$.

For an annulus, we have $W_\text{p} = \frac{\pi}{16}D^3(1 - d^4/D^4)$.

The strength condition is reduced to the inequality $\tau_\text{max} = |M_z|/W_\text{p} \le [\tau]$, where $[\tau]$ is the admissible stress in torsion.

Just as in case of strength calculated under tension and compression, the following three types of problems are possible, which differ in the form of the strength condition.

1. Checking calculation: $|M_z|/W_\text{p} \le [\tau]$.
2. Choosing the cross-section: $W_\text{p} \ge |M_z|/[\tau]$.
3. Determining the admissible load: $|M_z| \le [\tau]W_\text{p}$.

Example. For a steel rod of circular cross-section (Fig. E3.10a), choose the diameter from the strength condition for $[\tau] = 100$ MPa.

Solution. Since the cross-section is constant along the rod, the cross-sections are unsafe on segment "*a*," where the maximum twisting moment $M_z = 30$ kN m arises (see Fig. E3.10d).

From the strength condition

$$\tau_\text{max} = \frac{M_z}{W_\text{p}} = \frac{16\,M_z}{\pi D^3} \le [\tau]$$

we obtain $D \ge \sqrt[3]{16M_z/(\pi[\tau])} = 0.119$ m.

Rounding up, we finally obtain $D = 12$ cm.

E3.5. Symmetrical Bending

E3.5.1. Bending Moment and Transverse Force

► **Definitions.** The *bending* is a type of rod deformation such that the bending moments M_x and M_y and, perhaps, the transverse forces Q_x and Q_y arise in the rod cross-sections and the other internal force factors (N and M_z) are zero.

If the transverse forces are zero, the bending is said to be *pure*, and if there are transverse forces, then we have the so-called *lateral bending*.

Consider the rods whose cross-sections have the axis of symmetry y. If such a rod is subjected to a system of external forces lying in the plane yz and perpendicular to the rod longitudinal axis (Fig. E3.13), then only two internal force factors—i.e., the bending

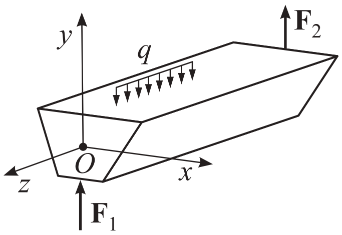

Figure E3.13. Symmetric bending.

moment M_x and the transverse force Q_y—arise in its cross-section. Such a bending is said to be *symmetric* (for the oblique bending, see Subsection E3.6.1).

Rule of signs: the bending moment is assumed to be positive if it results in extension of the lower fibers and compression of the upper fibers; the transverse force is positive if it is directed downwards acting on the cross-section from the right and is directed upwards acting on the cross-section from the left.

As was already noted, the rod internal force factors depend on the longitudinal coordinate of the cross-section under study; hence, they are functions of the variable z.

► **Differential equations of equilibrium of a rod.** Consider a rod under the action of a load q distributed along its axis (Fig. E3.14a). We distinguish a small element by two cross-sections (Fig. E3.14b) and replace the action of the cut-off parts of the rod by the corresponding internal forces with positive directions.

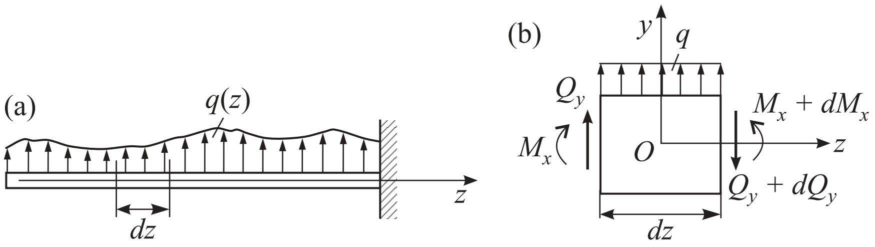

Figure E3.14. Rod equilibrium: (a) a rod under a distributed load (b) equilibrium of a small element.

Calculating the sum of projections of all forces on the axis y and the sum of moments of forces about the point O, we obtain the corresponding equations of equilibrium of the element:

$$q\,dz + Q_y - (Q_y + dQ_y) = 0,$$
$$-M_x + (M_x + dM_x) - Q_y\,dz = 0.$$

These equations imply the relations

$$\frac{dQ_y}{dz} = q, \qquad \frac{dM_x}{dz} = Q_y, \tag{E3.5.1.1}$$

which are called the *differential equations of equilibrium of a rod under bending*. Equation (E3.5.1.1), which must be satisfied by the functions $Q_y(z)$ and $M_x(z)$, is often used to construct diagrams of bending moments and transverse forces, and to verify this construction.

The rods that mainly work under bending are usually called *beams*.

► **Examples of diagram construction.** Consider examples of constructing diagrams of $Q_y(z)$ and $M_x(z)$ for several beams by using the method of sections.

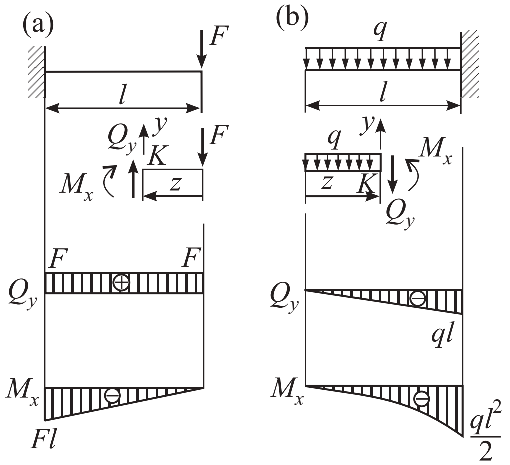

Figure E3.15. Cantilever beam (with clamped end): (a) loaded with a point force, (b) loaded with an evenly distributed force, (c) loaded with a moment.

In Fig. E3.15a, we show a cantilever beam (i.e., a beam with a fixed end) loaded by a lumped force. According to the method of sections, we draw an arbitrary cross-section K at a distance z from the free end of the beam and consider the free cut-off part in equilibrium. We replace the action of the cut-off part by the unknown forces Q_y and M_x assuming that their directions are positive. Calculating the sum of projections of all forces on the axis y and the sum of moments of forces about the point K, we compose the corresponding equations of equilibrium,

$$Q_y - F = 0, \quad -M_x - F_z = 0,$$

and whence obtain $Q_y = F$, $M_x = -F_z$. Thus, the transverse force is constant and the bending moment varies directly proportionally to the distance z taking the values 0 and $-Fl$ at the free ($z = 0$) and fixed ($z = l$) ends of the beam, respectively. In Fig. E3.15a, we present the corresponding diagrams of Q_y and M_x.

The same procedure for a beam uniformly loaded by a distributed load (Fig. E3.15b) gives

$$-qz - Q_y = 0, \quad \tfrac{1}{2}qz^2 + M_x = 0,$$

which implies $Q_y = -qz$ and $M_x = -\frac{1}{2}qz^2$. In this case, the transverse force is a linear function of z, and the bending moment depends on z according to the quadratic parabola law, which agrees with the diagrams of Q_y and M_x in Fig. E3.15b.

Similarly, for a cantilever beam loaded by a moment (Fig. E3.15c), we obtain $Q_y = 0$, $M_x = m$.

For two-support beams (Fig. E3.16) that do not have free ends, it is necessary first to find the support reactions from the equilibrium conditions for the beam as a whole. After this, one can determine the internal forces by analogy with the above procedure.

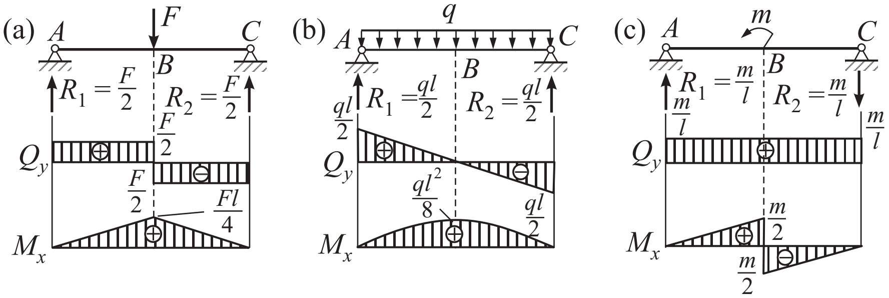

Figure E3.16. A beam hinged at both ends: (a) loaded with a point force, (b) loaded with an evenly distributed force, (c) loaded with a moment.

For an arbitrary cross-section on segment AB of a beam loaded at the center ($AB = BC = l/2$) by a lumped force F (Fig. E3.16a), we obtain $Q_y = R_1 = F/2$, $M_x = R_1 z = Fz/2$, where z is the distance from the left end A. For a cross-section on segment BC, we find $Q_y = R_1 - F = -F/2$, $M_x = R_1 z - F(z - l/2) = F(l - z)/2$. The corresponding diagrams are given at the bottom of Fig. E3.16a.

For the beam in Fig. E3.16b: $Q_y = R_1 - qz = q(l/2 - z)$, $M_x = R_1 z - qz^2/2 = qz(l - z)/2$. At the middle of the beam, we have $M_x = ql^2/8$.

For segment AB of the beam in Fig. E3.16c, we have $Q_y = R_1 = m/l$, $M_x = R_1 z = mz/l$, and for segment BC, we have $Q_y = R_1 = m/l$, $M_x = R_1 z - m = m(z/l - 1)$.

► **Specific characteristics of diagrams of M_x and Q_y.** We briefly formulate the diagram qualitative characteristics, which follows from relations (E3.5.1.1).

1. On the beam part free from any load ($q = 0$), Q_y is constant, and M_x is a linear function of z.

2. On the segment of uniformly distributed load ($q = \text{const}$), Q_y is a linear function, and M_x is a quadratic function (see Figs. E3.15b and E3.16b).

3. At the points of application of the lumped force, Q_y experiences a jump by the value of this force, and the diagram of M_x has a break point (Fig. E3.16a).

4. At the points of application of the lumped moment, M_x experiences a jump by the value of this moment (Fig. E3.16c).

At the extrema points of the function M_x, $Q_y = 0$ (Fig. E3.16b).

The knowledge of qualitative characteristics of diagrams allows one to construct diagrams without finding the functional dependencies $Q_y(z)$ and $M_x(z)$ but calculating the values of Q_y and M_x only on the boundaries of typical segments. The extremum values of M_x are determined after finding the extrema points from the conditions $Q_y = 0$.

E3.5.2. Stresses and Strains under Symmetric Pure Bending

► **Pure bending hypotheses.** Consider a beam in the state of pure bending (Fig. E3.17). In its cross-sections, there arises a single nonzero force factor, i.e., the constant bending moment $M_x = m$.

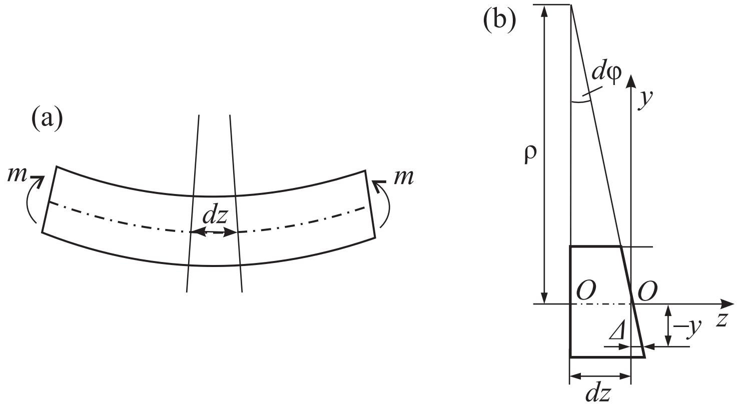

Figure E3.17. Simple pure bending of a beam: (a) deformation in bending, (b) deformation of a small element.

We use the *flat cross-section hypothesis**: after the strain, the beam cross-sections plane before the strain remain plane and perpendicular to the strained longitudinal axis.

According to this hypothesis, two close cross-sections rotate about each other by an angle $d\varphi$ (Fig. E3.17b). As a result of such strain, there arises a *neutral layer* whose longitudinal fibers $O - O$ do not change their length. In the lower part of the cross-section, the longitudinal fibers elongate (in the upper part, become shorter) by the value $\Delta = -y\,d\varphi$. The longitudinal strain of a fiber is equal to $\varepsilon = \Delta/dz = -y\,d\varphi/(\rho\,d\varphi) = -y/\rho$, where ρ is the radius of curvature of the neutral layer.

---

* A similar hypothesis was used in Subsection E3.3.2.

In addition to the flat cross-section hypothesis, we assume that the beam longitudinal fibers do not exert their weight upon one another. This means that they are in the state of uniaxial tension or compression (see Fig. E3.5).

► **Axis curvature. Formula for normal stresses.** According to Hooke's law, we have

$$\sigma = E\varepsilon = -\frac{Ey}{\rho}. \tag{E3.5.2.1}$$

This implies that the normal stresses σ in the beam cross-section vary directly proportionally to the distance y from the neutral layer (Fig. E3.18).

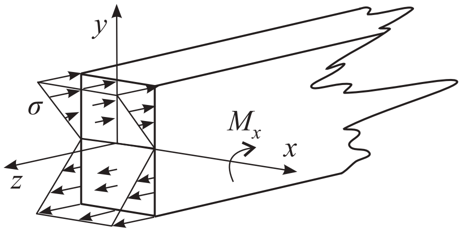

Figure E3.18. Normal stresses in the beam cross-section.

Using formulas (E3.1.2.1) and the fact that the longitudinal force is zero, we obtain

$$N = \int_A \sigma\, dA = -\frac{E}{\rho}\int_A y\, dA = 0,$$

$$M_x = -\int_A \sigma y\, dA = \frac{E}{\rho}\int_A y^2\, dA.$$

It follows from the first relation that the *neutral axis x* of the cross-section passes through the center of gravity, and hence the quantity ρ is the radius of curvature of the deflected axis of the beam. From the second relation, we derive the formula for the axis curvature

$$k = \frac{1}{\rho} = \frac{M_x}{EI_x}, \quad \text{where} \quad I_x = \int_A y^2\, dA. \tag{E3.5.2.2}$$

The quantity I_x is called the *moment of inertia of the cross-section about the axis x* (see Subsection E2.5.1 in the chapter "Mechanics of point particles and rigid bodies"), and the product EI_x is called the *cross-section rigidity in bending* or *flexural rigidity*.

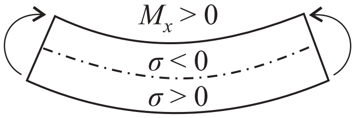

Figure E3.19. Signs of normal stresses in simple bending.

From (E3.5.2.1) and (E3.5.2.2) we obtain the basic formula for the rod normal stresses in the case of symmetric bending

$$\sigma = -\frac{M_x y}{I_x}, \tag{E3.5.2.3}$$

where the "minus" sign is necessary to match the rules of sign for the quantities σ and M_x. For $M_x > 0$ (Fig. E3.19), the upper fibers are compressed ($y > 0, \sigma < 0$), and the lower fibers are extended ($y < 0, \sigma > 0$); for $M_x < 0$, the signs of stresses in the upper and lower parts of the cross-section are changed to the opposite.

E3.5.3. Stresses and Strains in Symmetric Lateral Bending

► **Shear stresses. Zhuravskii formula.** In the case of symmetric lateral bending, in the rod cross-sections there arises not only a bending moment M_x but also a transverse force Q_y related to the shear stresses in the cross-section by the formula

$$Q_y = \int_A \tau_y \, dA.$$

Although there are shear stresses, formulas (E3.5.2.2) and (E3.5.2.3) obtained for the pure bending can also be used in this case.

To determine the shear stresses, consider the beam in Fig. E3.20a. We cut a small element from it by two neighboring cross-sections and one longitudinal cross-section (Fig. E3.20b). The normal stresses calculated by formula (E3.5.2.3) act on its lateral faces. Since the bending moment on the right is greater than the bending moment on the left by the value dM_x, the normal stress on the right is also greater by $d\sigma = dM_x y/I_x$.

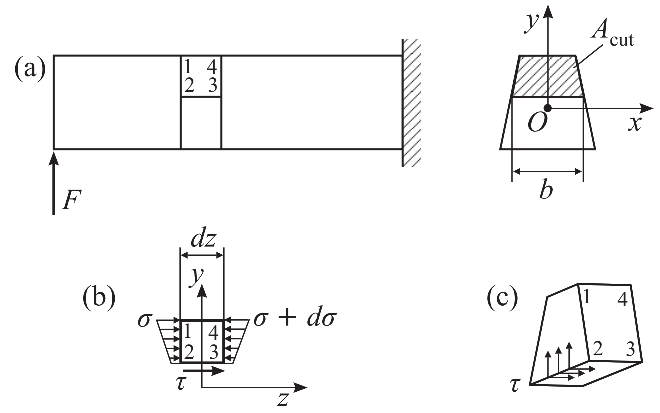

Figure E3.20. Symmetric lateral bending of a cantilever beam: (a) small element and its deformation, (b) normal and tangential stresses at the faces of the small element, (c) normal and tangential stresses in the cross-section.

For this element to be in equilibrium, the shear stresses τ must be applied to its lower face (Fig. E3.20b).

We assume that they are uniformly distributed over the width of the cross-section. From the equilibrium condition for an element stating that the sum of projections of all forces on the axis z is zero,

$$\tau b \, dz - \int_{A_{\text{cut}}} d\sigma \, dA = 0,$$

where the integral is extended to the cut-off part of the rod cross-section, we obtain

$$\tau = \frac{\int_{A_{\text{cut}}} d\sigma \, dA}{b \, dz} = \frac{dM_x}{dz} \frac{\int_{A_{\text{cut}}} y \, dA}{b I_x}.$$

Taking into account the second equation in (E3.5.1.1), we obtain the *Zhuravskii formula* for shear stresses:

$$\tau = \frac{Q_y S_x^{\text{cut}}}{I_x b}, \quad \text{where} \quad S_x^{\text{cut}} = \int_{A_{\text{cut}}} y \, dA. \qquad \text{(E3.5.3.1)}$$

The quantity S_x^{cut} is called the *static moment of the cross-section cut-off part about the axis* x.

According to the reciprocity law for shear stresses (see Subsection E3.2.1), formula (E3.5.3.1) also determines the shear stresses in the rod cross-section (Fig. E3.20c).

► **Cross-section of rectangular shape.** In the special case of the cross-section of rectangular shape (Fig. E3.21), we have

$$I_x = \tfrac{1}{12} bh^3, \quad S_x^{\text{cut}} = y_C\, A_{\text{cut}} = \tfrac{1}{2} b \left(\tfrac{1}{4} h^2 - y^2\right).$$

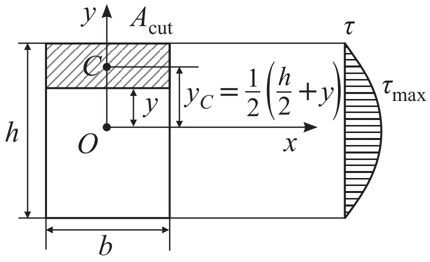

Figure E3.21. Rectangular cross-section and tangential stress diagram.

Substituting these expressions into formula (E3.5.3.1), we obtain

$$\tau = \frac{Q_y S_x^{\text{cut}}}{I_x b} = \frac{6Q_y}{bh^3} \left(\frac{h^2}{4} - y^2\right).$$

One can see that the quantity τ varies over the cross-section height according to the quadratic parabola law and attains the maximum value for $y = 0$ (i.e., on the axis x):

$$\tau_{\max} = \frac{3Q_y}{2bh}. \qquad \text{(E3.5.3.2)}$$

In Fig. E3.21, we present the diagram of shear stresses over the height of the rectangular cross-section.

E3.5.4. Strength Analysis in the Case of Symmetric Bending

► **Maximal stresses.** It follows from formula (E3.5.2.3) that the maximum (in magnitude) normal stresses are attained at the cross-section points most remote from the neutral axis x, i.e., for $|y| = y_{\max}$:

$$\sigma_{\max} = \frac{|M_x| y_{\max}}{I_x} = \frac{|M_x|}{W_x}, \quad \text{where} \quad W_x = \frac{I_x}{y_{\max}}. \qquad \text{(E3.5.4.1)}$$

The quantity W_x is called the *moment of resistance of the cross-section in bending* (axial moment of resistance).

For the cross-section of rectangular shape (Fig. E3.21), we have

$$W_x = \frac{bh^2}{6} \quad \text{and} \quad \sigma_{\max} = \frac{6|M_x|}{bh^2}. \qquad \text{(E3.5.4.2)}$$

We compare the quantities $\sigma_{\max}$ and $\tau_{\max}$ for a beam of rectangular cross-section. In the order of magnitude, $Q_y \sim F$ and $M_x \sim Fl$, where F is the characteristic force acting on the beam and l is the beam length. Then it follows from (E3.5.3.2) and (E3.5.4.2) that

$$\frac{\tau_{\max}}{\sigma_{\max}} \sim \frac{h}{l} \ll 1.$$

Thus, the shear stresses are small compared with the normal stresses. This conclusion also holds for rods with cross-sections of different shapes (an exception is thin-walled rods, in which the appearance of large shear stresses is possible).

► **Beam strength conditions.** As a rule, the strength analysis for beams is performed for the normal stresses using the strength condition in the form

$$\sigma_{\max} = \frac{|M_x|}{W_x} \leq [\sigma].$$

But if the rod material (for example, wood) badly resists the shear strains (cut off), then the calculations for the shear stresses are also necessary:

$$\tau_{\max} = \frac{|Q_y|}{I_x} \left(\frac{S_x^{\text{cut}}}{b} \right)_{\max} \leq [\tau].$$

In the cases where the rod material has different admissible stresses under tension $[\sigma^{\text{t}}]$ and compression $[\sigma^{\text{c}}]$, the strength analysis is performed separately for the tensile and compressing stresses:

$$\sigma^{\text{t}}_{\max} = \frac{|M_x| y^{\text{t}}_{\max}}{I_x} \leq [\sigma^{\text{t}}],$$
$$\sigma^{\text{c}}_{\max} = \frac{|M_x| y^{\text{c}}_{\max}}{I_x} \leq [\sigma^{\text{c}}],$$

where $y^{\text{t}}_{\max}$ and $y^{\text{c}}_{\max}$ are the distances from the axis x to the respective points of the extended and compressed parts of the cross-section that are most remote from this axis.

Just as in the case of tension and compression or torsion, in the strength analysis for rods under bending, one solves the following three problems:

1) checking calculation,
2) choosing the cross-section,
3) determining the admissible load.

All geometric characteristics required in strength analysis of rolled cross-sections (double tee, channel bar, etc.) are given in special tables called the range of rolled steel products.

Example. For a steel double tee beam (No. 55) of length $l = 3.2\,\text{m}$ (see Fig. E3.16b), determine the admissible load $[q]$ for $[\sigma] = 160\,\text{MPa}$ from the strength condition.

Solution. For the beam under study, the unsafe cross-section is that in the middle of the beam span, where the maximum bending moment $M_x = ql^2/8$ arises.

In the range of rolled steel products, we find the resistance moment $W_x = 2000\ \text{cm}^3$ for the double tee.

It follows from the strength condition

$$\sigma_{\max} = \frac{M_x}{W_x} = \frac{ql^2}{8W_x} \leq [\sigma]$$

that $q \leq 8W_x[\sigma]/l^2 = 250\,\text{N/m}$. Thus, $[q] = 250\,\text{kN/m}$.

E3.6. Combined Stress

The *combined stress* includes the types of a rod strain such that at least two nonzero internal force factors arise simultaneously in the rod cross-sections. An exception is the symmetric lateral bending, which is considered as a simple type of strain, although there are two force factors in this case, the bending moment and the transverse force, because in the overwhelming majority of cases, the strength and rigidity analysis are performed without taking into account the transverse forces, i.e., for a single force factor, which is the bending moment.

E3.6.1. Oblique Bending

► **Definition.** When studying the symmetric lateral bending, we assume that the cross-section is symmetric with respect to the axis y. It turns out that all the results of Section E3.5 also hold for cross-sections of arbitrary shape if the axes x and y are the *principal central axes of inertia*; i.e., they are mutually perpendicular axes passing through the cross-section center of gravity and the axial moments of inertia about these axes take extremum values. In what follows, for the axes x and y we take the principal central axes.

The *oblique bending* is the general case of the rod bending (see Subsection E3.5.1) in which two bending moments M_x and M_y and perhaps the transverse forces Q_x and Q_y appear in the rod cross-section.

Using the force superposition principle, we can determine the normal stresses at any point of the cross-section with coordinates x, y by the formula

$$\sigma = -\frac{M_x y}{I_x} - \frac{M_y x}{I_y}, \tag{E3.6.1.1}$$

which follows from formula (E3.5.2.3).

► **Neutral axis in the case of oblique bending.** Just as in the case of symmetric bending, in the cross-section there exists a neutral axis at whose points the stresses are zero. It is determined by the equation

$$\frac{M_x y}{I_x} + \frac{M_y x}{I_y} = 0. \tag{E3.6.1.2}$$

Listed below are several properties of the neutral axis in oblique bending.

1. The neutral axis passes through the center of gravity of the cross-section.

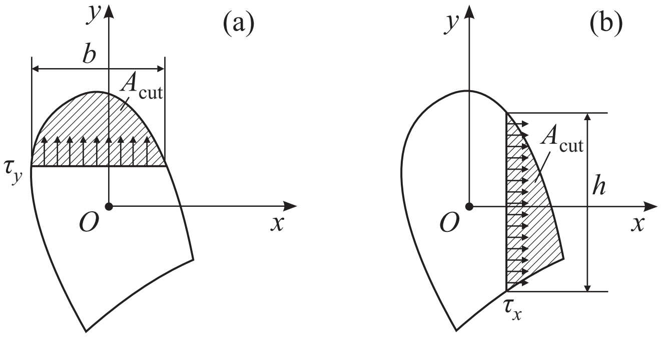

Figure E3.22. Tangential stresses τ_y (a) and τ_x (b) in the cross-section in oblique bending.

2. The neutral axis divides the cross-section into two parts: in one of them, $\sigma > 0$, in the other, $\sigma < 0$.

3. The normal stresses in the cross-section vary directly proportionally to the distance from the neutral axis attaining the maximum values (in magnitude) at the cross-section points most remote from the neutral axis.

The shear stresses in the cross-section (Fig. E3.22) can be determined by formula (E3.5.3.1):

$$\tau_y = \frac{Q_y S_x^{\text{cut}}}{I_x b}, \quad \tau_x = \frac{Q_x S_y^{\text{cut}}}{I_y h}.$$

Just as in the case of symmetric lateral bending, the shear stresses, as a rule, play a secondary role, and hence the strength conditions impose constraints on the maximum normal stresses.

Example. Calculate the maximum normal stresses in the unsafe cross-section of the beam (Fig. E3.23) under oblique bending. The cross-section is of rectangular shape, $b = 6\,$cm, $h = 12\,$cm.

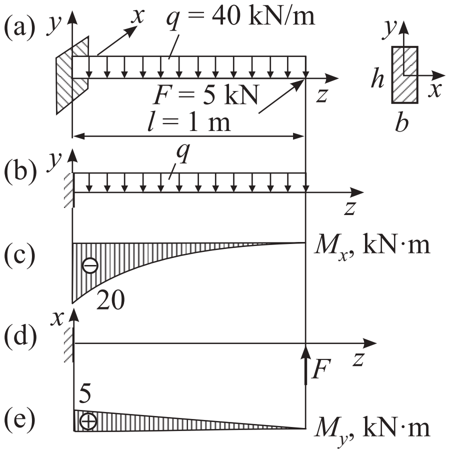

Figure E3.23. Oblique bending: (a) a beam in oblique bending, (b) vertical loading, (c) bending moment diagram for vertical loading, (d) horizontal loading, (e) bending moment diagram for horizontal loading.

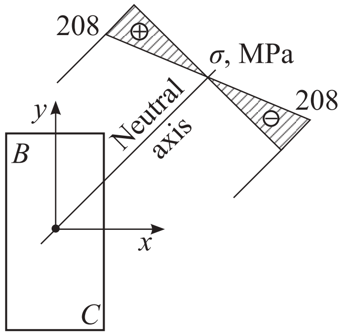

Figure E3.24. Neutral axis and the normal stress diagram in the transversal direction.

Solution. In Fig. E3.23b and d, we show the beam loading diagram in the vertical and horizontal planes, and in Fig. E3.23c and e, the corresponding diagrams of the bending moments M_x, M_y. Judging from these diagrams, the unsafe cross-section is the fixation cross-section, where $M_x = -20\,$kN m, $M_y = 5\,$kN m.

For the rectangular cross-section, $I_x = bh^3/12$, $I_y = hb^3/12$, and the neutral axis equation (E3.6.1.2) for the unsafe cross-section takes the form $-y + x = 0$.

It follows from the position of the neutral line (Fig. E3.24) that the most unsafe points of the cross-section are points B and C; i.e., the points most remote from the neutral axis. Substituting their coordinates into (E3.6.1.1), we calculate the stresses at these points

$$\sigma_B = -\frac{M_x y_B}{I_x} - \frac{M_y x_B}{I_y} = 208 \text{ MPa},$$
$$\sigma_C = -\frac{M_x y_C}{I_x} - \frac{M_y x_C}{I_y} = -208 \text{ MPa}.$$

In Fig. E3.24, we show the diagram for σ, which characterized the variations in the normal stresses in the direction perpendicular to the neutral axis.

E3.6.2. Eccentric Extension or Compression

► The **eccentric extension or compression** is a type of a rod deformation such that the longitudinal force N and bending moments M_x, M_y (and perhaps the transverse forces Q_x, Q_y) arise in the rod cross-section.

The longitudinal force and the bending moments can be treated as the result of action of an eccentrically applied force $F = N$ on the rod (Fig. E3.25). Therefore, such a type of combined stresses is called eccentric extension or compression.

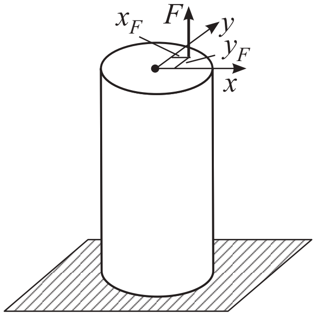

Figure E3.25. Eccentric extension of a rod.

The bending moments are related to the coordinates of the point of application of the force F by the relations $M_x = -Fy_F$, $M_y = -Fx_F$. Therefore, from (E3.6.1.1), formulas (E3.3.2.1), and the force superposition principle for normal stresses at an arbitrary point of any cross-section with coordinates x, y, we obtain

$$\sigma = \frac{N}{A} - \frac{M_x y}{I_x} - \frac{M_y x}{I_y} = F\left(\frac{1}{A} + \frac{y_F y}{I_x} + \frac{x_F x}{I_y}\right). \qquad \text{(E3.6.2.1)}$$

► **Neutral axis in the case of eccentric extension or compression.** In this case, the equation of the cross-section neutral axis at whose points the stresses are zero becomes

$$\frac{1}{A} + \frac{y_F y}{I_x} + \frac{x_F x}{I_y} = 0. \qquad \text{(E3.6.2.2)}$$

It is easy to see that the neutral axis does not pass through the center of gravity of the cross-section. The other properties are the same as in the case of oblique bending. In addition, we mention one more property of the neutral axis in the case of eccentric extension or compression: the neutral axis does not intersect the quarter of the cross-section where the force F is applied.

► **Core of a cross-section.** The neutral axis position, as follows from Eq. (E3.6.2.2), depends on the coordinates of the point of application of the force F. If the point of application of the force F is sufficiently close to the center of gravity of the cross-section, namely, is in the region which is called the *core of the cross-section*, then the neutral axis passes beyond the cross-section; i.e., all the points of the cross-section experience normal stresses of the same sign. In Fig. E3.26, we show the cores for a rectangular and circular cross-sections. The strength conditions in the case of eccentric extension or compression have the form of constraints on the maximum normal stresses.

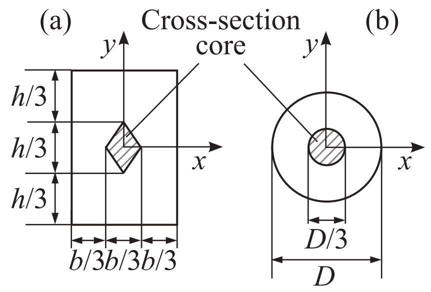

Figure E3.26. (a) Rectangular cross-section core, (b) circular cross-section core.

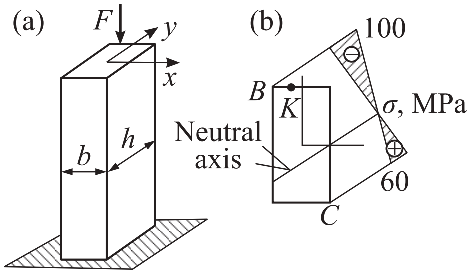

Figure E3.27. (a) Eccentrically compressed beam of rectangular cross-section, (b) force application point and the normal stress diagram.

Example. Calculate the maximum normal stresses in the cross-section of an eccentrically compressed rod of rectangular cross-section for $b = 6\,\text{cm}$, $h = 12\,\text{cm}$ (Fig. E3.27). The point K of application of the force $F = 144$ kN has the coordinates $x_F = -1$ cm, $y_F = 6$ cm (Fig. E3.27b).

Solution. We calculate the geometric characteristics of the cross-section:

$$A = bh = 72\,\text{cm}^2,$$
$$I_x = bh^3/12 = 864\,\text{cm}^4,$$
$$I_y = hb^3/12 = 216\,\text{cm}^4.$$

The neutral axis equation (E3.6.2.2) becomes $1 + y/2 - x/3 = 0$. Its position (Fig. E3.27b) shows that B and C are the most stressed points of the cross-section. Substituting their coordinates into (E3.6.2.1), we find the stresses at these points (the force F must be taken with "minus" sign, because its direction is opposite to that shown in Fig. E3.25):

$$\sigma_B = -F\left(\frac{1}{A} + \frac{y_F y_B}{I_x} + \frac{x_F x_B}{I_y}\right) = -100 \text{ MPa},$$
$$\sigma_C = -F\left(\frac{1}{A} + \frac{y_F y_C}{I_x} + \frac{x_F x_C}{I_y}\right) = 60 \text{ MPa}.$$

In Fig. E3.27b, we construct the diagram of σ, which shows the law of normal stress distribution in the direction perpendicular to the neutral axis.

E3.6.3. Bending with Torsion

► The **bending with torsion** is a type of rod deformation such that the twisting moment M_z, the bending moments M_x and M_y, and, perhaps, the transverse forces Q_x and Q_y arise in its cross-sections.

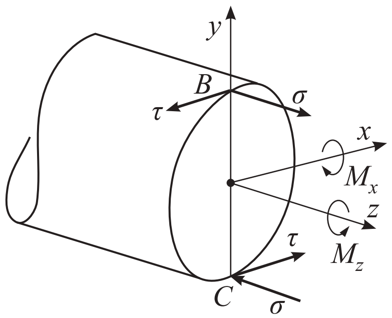

Figure E3.28. Combined bending and torsion of circular beam.

We consider rods of a circular cross-section. Because it is symmetric, any two mutually perpendicular central axes of the cross-section are principal axes. Therefore, they can always be chosen so that, in the cross-section under study, the total bending moment has only a single component, for example, M_x. The maximum (in magnitude) normal stresses in the cross-section are determined by formula (E3.5.4.1) for the symmetric bending

$$\sigma_{\max} = \frac{|M_x|}{W_x} \tag{E3.6.3.1}$$

and are attained at points B and C that are most remote from the neutral axis x (Fig. E3.28).

By formula (E3.4.3.1), we obtain the maximum shear stresses in torsion:

$$\tau_{\max} = \frac{|M_z|}{W_\mathrm{p}}.$$

They appear at the points of the cross-section boundary, including points B and C.

Thus, points B and C are the most unsafe points. In their neighborhood, the plane stress state is realized, which is a combination of uniaxial extension (or compression) and simple shear.

► **The Saint-Venant and von Mises strength conditions.** In contrast to all the strain types considered above, in this case, both σ and τ may be significant, and therefore, the problem of the strength condition arises for such a stress state. There exist different theories of strength answering this question. Most often, one applies the Saint-Venant strength condition

$$\sqrt{\sigma^2 + 4\tau^2} \le [\sigma]$$

or the von Mises strength condition

$$\sqrt{\sigma^2 + 3\tau^2} \le [\sigma].$$

Using the relation between the resistance moments $W_\mathrm{p} = 2W_x$ and taking into account the fact that, in the case of arbitrary orientation of the axes x and y, the quantity M_x in formula (E3.6.3.1) must be replaced by the total bending moment $M_\mathrm{b} = \sqrt{M_x^2 + M_y^2}$, we reduce the strength conditions to the final form:

$$\frac{\sqrt{M_x^2 + M_y^2 + M_z^2}}{W_x} \le [\sigma] \qquad \text{(Saint-Venant)},$$

$$\frac{\sqrt{M_x^2 + M_y^2 + 0.75M_z^2}}{W_x} \le [\sigma] \qquad \text{(von Mises)},$$

where $W_x = \pi D^3/32$ is the axial moment of resistance for a disk of diameter D.

Example. Using the von Mises strength condition, find the diameter of a steel shaft under bending with torsion (Fig. E3.29). The admissible stress is $[\sigma] = 150\,\mathrm{MPa}$.

Solution. The diagram of the shaft loading in the vertical plane and the corresponding diagram of the bending moment M_x are given in Fig. E3.29b and c.

The diagram of the shaft loading by the external moment and the diagram of the arising twisting moments M_z are given in Fig. E3.29d and e.

The most unsafe cross-section is the fixation cross-section, where the maximum bending $M_x = 13\,\mathrm{kN\,m}$ and twisting $M_z = 6\,\mathrm{kN\,m}$ moments arise.

From the strength condition

$$\frac{32\sqrt{M_x^2 + 0.75M_z^2}}{\pi D^3} \le [\sigma]$$

we obtain: $D \ge 0.098\,\mathrm{m}$. Thus, we can round up: $D = 10\,\mathrm{cm}$.

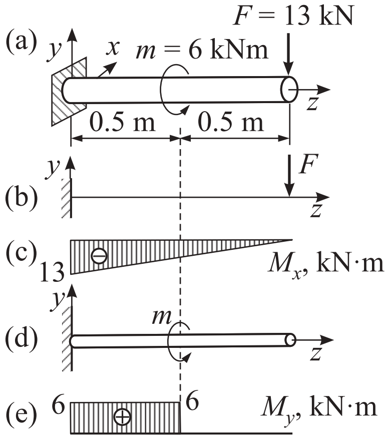

Figure E3.29. A shaft in combined bending and torsion: (a) dimensions and loads, (b) vertical loading, (c) bending moment diagram, (d) torsion loading, (e) twisting moment diagram.

E3.7. Stability of Compressed Rods

E3.7.1. Critical Force

In Subsection E3.1.1, it was already noted that some equilibria of bodies or systems turn out to be unstable, i.e., such that any negligible mechanical actions may lead to significant deviations from these states. Each of these equilibria is characterized by a certain value of the load which separates the regions of stable and unstable equilibria.

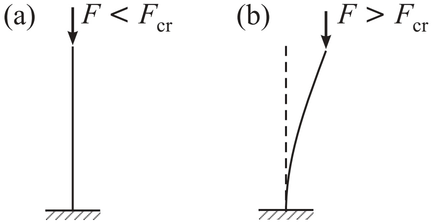

Figure E3.30. A compressed rod: (a) stable state, (b) unstable state.

In the case of a compressed rod (Fig. E3.30), it turns out that if the compressing force F does not exceed a certain value called the *critical force* F_{cr}, then the rectilinear form of the rod equilibrium is stable. But if $F > F_{\text{cr}}$, then the rectilinear form of the rod equilibrium is unstable, and the rod tends to take another stable form of equilibrium with a curvilinear axis.

The compressing stresses $\sigma_{\text{cr}} = F_{\text{cr}}/A$ corresponding to the critical force are called *critical stresses*.

E3.7.2. Euler Formula

Thus, for $F > F_{\text{cr}}$, the solution of the problem of the rod equilibrium is not unique. An analysis of the conditions of this ambiguity permits obtaining the formula for the value of F_{cr}.

Consider a hinged centrally compressed rod in a slightly deflected state (Fig. E3.31). The bending moment in the rod cross-section is equal to $M = -Fv$, where the deflection $v(z)$

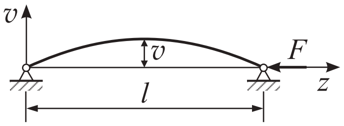

Figure E3.31. A hinged, centrally compressed rod.

determines the rod curvilinear axis. For small deflections of the rod, the curvature k of the deflected axis ($v/l \ll 1, v' \ll 1$) is equal to

$$k = \frac{1}{\rho} = \frac{v''}{(1+v'^2)^{3/2}} \approx \frac{d^2v}{dz^2}.$$

An analysis of the bending strain leads to formula (E3.7.2.3) in Section E3.5 relating the curvature and the bending moment,

$$k = \frac{1}{\rho} = \frac{M}{EI},$$

where the rules of signs agree well for the curvature $k = 1/\rho$ and the moment M.

Thus, we obtain the differential equation for the rod deflected axis:

$$v'' = -a^2 v, \qquad \text{(E3.7.2.1)}$$

where $a^2 = F/(EI)$.

The boundary conditions take into account the fact that the rod ends are fixed:

$$v(0) = v(l) = 0. \qquad \text{(E3.7.2.2)}$$

Equation (E3.7.2.1) with boundary conditions (E3.7.2.2) has the obvious trivial solution $v = 0$ corresponding to the rectilinear form of equilibrium. We are interested in nonzero solutions.

The general integral of Eq. (E3.7.2.1) has the form

$$v = C_1 \sin(az) + C_2 \cos(az).$$

It follows from the first boundary condition that $C_2 = 0$, and from the second, that $C_1 \sin(al) = 0$. Eliminating $C_1 \neq 0$, we obtain the relation

$$\sin(al) = 0,$$

which is satisfied if $al = \pi n$ or

$$F = \frac{\pi^2 n^2 EI}{l^2} \quad (n = 1, 2, \dots). \qquad \text{(E3.7.2.3)}$$

Thus, it was found that if the compression force F takes discrete values determined by formula (E3.7.2.3), then, along with the rectilinear equilibrium, the rod curved equilibria of the sine curve form are possible;

$$v = C_1 \sin\left(\frac{n\pi z}{l}\right).$$

From (E3.7.2.3) with $n = 1$, we obtain the *Euler formula* for the critical force

$$F_{\text{cr}} = \frac{\pi^2 EI}{l^2}. \qquad \text{(E3.7.2.4)}$$

A more exact theory shows that the ambiguity of the equilibrium forms occurs for all forces F exceeding the Euler force (E3.7.2.4).

E3.7.3. Influence of Methods for Fixation of the Rod Ends on the Value of the Critical Force

Similar problems for other cases of rod ends fixation (Fig. E3.32) can also be considered. Their solution shows that the critical force can be determined by the formula similar to (E3.7.2.4):

$$F_{\rm cr} = \frac{\pi^2 EI}{(\mu l)^2}, \tag{E3.7.3.1}$$

introducing the notions of *reduced length* μl and of *reduced length coefficient* μ.

The quantity μl can be treated as the rod length part on which the half-wave of the sine curve corresponding to the rod curved axis can be placed. The values of the coefficient μ in several cases of the rod fixation are given in Fig. E3.32.

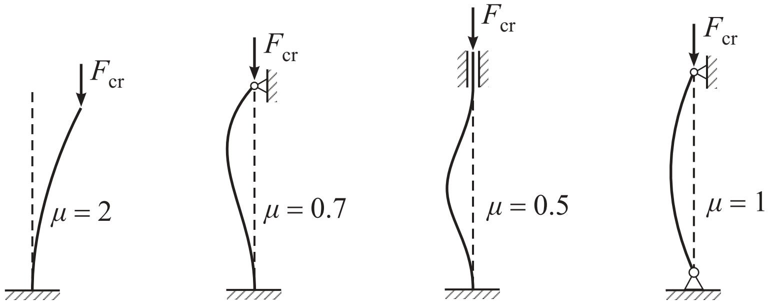

Figure E3.32. Reduced length coefficients μ for various fixing conditions of a rod.

Example. Find the critical force for a steel rod of length $l = 4$ m (Fig. E3.32) one of whose ends is rigidly fixed. The rod cross-section is the double tee No. 22.

Solution. In the range of rolled steel products, we find the moments of inertia $I_x = 2550\,\text{cm}^4$, $I_y = 157\,\text{cm}^4$ for this double tee. Obviously, the rod becomes unstable in plane of minimum rigidity, and therefore, we must take the minimum moment of inertia $I_{\rm min} = I_y = 157\,\text{cm}^4$.

For the given fixation coefficient $\mu = 2$, the modulus of elasticity for steel is $E = 210\,\text{GPa}$. By the Euler formula (E3.7.3.1), we obtain $F_{\rm cr} = \dfrac{\pi^2 EI_{\rm min}}{(\mu l)^2} = 50.8\,\text{kN}$.

E3.7.4. Scope of the Euler Formula

The critical stresses corresponding to the critical force (E3.7.3.1) are equal to

$$\sigma_{\rm cr} = \frac{F_{\rm cr}}{A} = \frac{\pi^2 E}{(\mu l/i)^2},$$

where $i = \sqrt{I/A}$ is the radius of inertia of the cross-section.

We introduce the notation $\lambda = \mu l/i$. Then the preceding formula takes the form

$$\sigma_{\rm cr} = \frac{\pi^2 E}{\lambda^2}. \tag{E3.7.4.1}$$

The quantity λ is called the *rod flexibility*.

It follows from (E3.7.4.1) that as the flexibility decreases, the value of $\sigma_{\rm cr}$ increases unboundedly. It is clear that formula (E3.7.4.1) cannot be used for stresses that are too large, because the rod material under these conditions does not obey Hooke's law.

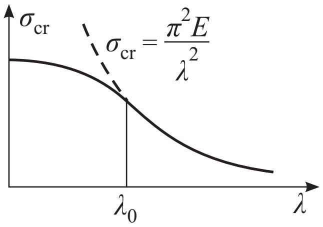

Figure E3.33. Dependence of the critical stress on the rod flexibility.

The value of the maximum stress at which Hooke's law is still satisfied is called the *proportionality limit* σ_{prop}.

Thus, the range of applicability for the Euler formula is determined by the inequality $\sigma_{\text{cr}} < \sigma_{\text{prop}}$ or $\pi^2 E/\lambda^2 < \sigma_{\text{prop}}$. As a result, we obtain $\lambda > \lambda_0 = \pi\sqrt{E/\sigma_{\text{prop}}}$. The dependence of σ_{cr} on λ for $\lambda < \lambda_0$ is usually determined experimentally. In Fig. E3.33, the solid line schematically shows the complete graph of the dependence $\sigma_{\text{cr}}(\lambda)$. For very small flexibilities, the rod loses strength under compression, and therefore, the critical stresses are practically equal to the corresponding strength loss stresses.

E3.7.5. Stability Analysis of Compressed Rods

In practice, it is necessary to ensure the stable operation of compressed rods, and therefore, the stresses in the rod cannot exceed the critical stresses with a certain margin:

$$\sigma = \frac{F}{A} \le \frac{\sigma_{\text{cr}}}{n}, \qquad \text{(E3.7.5.1)}$$

where n is the safety factor.

For convenience, the *coefficient of longitudinal bending* $\varphi = \sigma_{\text{cr}}/(n[\sigma])$ is introduced. Then condition (E3.7.5.1) takes the form

$$\frac{F}{A} \le [\sigma]\varphi. \qquad \text{(E3.7.5.2)}$$

The dependencies $\varphi(\lambda)$ for different materials are given in reference books in the form of tables.

Using condition (E3.7.5.2), one can solve the same problems as those in the strength analysis (see Subsection E3.3.3).

Bibliography for Chapter E3

Beer, F.P., Johnston, E.R. et al., *Mechanics of Materials, 3rd Edition*, McGraw-Hill, New York, 2001.
Birger, I. A. and Mavljutov, R. R., *Strength of Materials* [in Russian], Nauka, Moscow, 1986.
Case, J., Chilver, H., and Ross, C. T. F., *Strength of Materials and Structures, 4th Edition,* Elsevier, Amsterdam, 1999.
Darkov, A. V. and Shapiro, G. S., *Strength of Materials* [in Russian], Vysshaya Shkola, Moscow, 1989.
Den Hartog, J. P., *Strength of Materials*, Dover Publications, New York, 1961.
Drucker, D. C., *Introduction to Mechanics of Deformable Solids*, McGraw-Hill, New York, 1967.
Feodos'ev, V. I., *Strength of Materials* [in Russian], Nauka, Moscow, 1986.
Gastev, V. A., *A Short Course on Strength of Materials* [in Russian], Nauka, Moscow, 1977.
Mott, R. L., *Applied Strength of Materials, 4th Edition*, Prentice-Hall, New York, 2002.
Nash, W., *Schaum's Outline of Strength of Materials, 4th Edition*, McGraw-Hill, New York, 1998.
Popov, E. P., *Engineering Mechanics of Solids*, Prentice Hall, Englewood Cliffs, New Jersey, 1990.
Rabotnov, Yu. N., *Mechanics of Solids* [in Russian], Nauka, Moscow, 1988.

Smith, W. F. and Hashemi, J., *Foundations of Materials Science and Engineering, 4th Edition*, McGraw-Hill, New York, 2006.

Timoshenko, S. P., *Strength of Materials, 3rd Edition*, Krieger Publishing Co., Malabar, Florida, 1983.

Timoshenko, S. P. and Young, D. H., *Elements of Strength of Materials, 5th Edition*, D. Van Nostrand Co., New York, 1968.

Chapter E4

Hydrodynamics*

E4.1. Hydrostatics

E4.1.1. Properties of Incompressible Fluid at Rest. Basic Law of Hydrostatics

► **Preliminaries.** *Fluids* differ from solids in the absence of shape and in light mobility of small volumes. The ability of fluids to take any shape without separation into fractions under the action of small forces is called *fluidity*.

In their mechanical properties, the fluids are separated into the two classes, *liquids* and *gases*.

Liquids are fluids that can form a free surface, namely, a surface of contact (interface) between the liquid and the surrounding gas (another liquid). Fluids have low compressibility. This class of liquids comprises water, gasoline, oil, petroleum, mercury, etc.

Gases are fluids occupying the entire possible volume. Gases are easily compressed and expanded under the action of applied forces.

► **Properties of an incompressible liquids at rest.**

1. *Pascal's law*: pressure applied to a liquid is transferred to any point of the liquid equally in all directions (the hydrostatic pressure on a small plane site is independent of its orientation in space).

Corollary: for a vessel of an arbitrary shape and any dimensions, the liquid pressure is the same at the same depth.

Remark. Pascal's law also holds for gases.

2. The interface between two immiscible liquids (or a liquid and a gas) is a horizontal plane.
3. The *law of communicating vessels*: the level of a homogeneous liquid in communicating vessels is the same.

Example 1. A hydraulic press consists of two cylindrical communicating vessels of different diameters equipped with pistons whose areas are S_1 and S_2. The cylinders are filled with liquid oil. The unloaded pistons are at the same level. A force F_1 acting on the piston with area S_1 creates the additional pressure $p = F_1/S_1$ in the liquid. According to Pascal's law, this pressure is transferred by the liquid in all directions without change. Therefore, the pressure force $F_2 = pS_2 = F_1S_2/S_1$ acts on the second piston. From this relation, we obtain

$$\frac{F_2}{F_1} = \frac{S_2}{S_1}. \qquad \text{(E4.1.1.1)}$$

Hence the forces acting on the hydraulic press pistons are proportional to the areas of these pistons. Therefore, using a hydraulic press, one can win much in the force by choosing $S_2 \gg S_1$.

---

* In this chapter, we consider several problems of hydrodynamics which are often encountered in chemical technology.

Remark. Deriving formula (E4.1.1.1), we neglected the term ρgh determining the pressure in the liquid due to the gravity force (see below).

► **Basic law of hydrostatics.** Consider the equilibrium of an incompressible homogeneous liquid in the field of gravity force. We introduce a rectangular Cartesian coordinate system whose Z-axis is directed vertically upwards (the gravity force is directed vertically downwards). *Basic law of hydrostatics*: the pressure distribution in a given volume of liquid is determined by the formula

$$p = p_0 - \rho g(z - z_0), \qquad \text{(E4.1.1.2)}$$

where p_0 is the pressure for $z = z_0$, ρ is the liquid density, and g is the free fall acceleration.

Example 2. Assume that the coordinate z_0 corresponds to the free surface of the liquid. Then for the liquid pressure at the depth $h = z_0 - z$, we have $p = p_0 + \rho gh$, where p_0 is the atmospheric pressure (the gas pressure) on the free surface.

► **Rotation of a liquid in a cylindrical vessel.** Consider a liquid in a cylindrical vessel rotating about the vertical axis at a constant angular velocity ω. We introduce a rectangular Cartesian coordinate system whose Z-axis is directed along the axis of rotation.

The pressure distribution in the liquid is determined by the formula

$$p = p_0 + \frac{1}{2}\rho\omega^2 r^2 - \rho g(z - z_0), \qquad \text{(E4.1.1.3)}$$

where $r = \sqrt{x^2 + y^2}$ is the radial distance to the cylinder axis. For $\omega = 0$, formula (E4.1.1.3) becomes (E4.1.1.2).

Assume that the coordinate z_0 corresponds to the free surface of the liquid on the axis of rotation. The surface of the liquid becomes a paraboloid of rotation and is described by the equation

$$\frac{1}{2}\omega^2 r^2 = g(z - z_0), \qquad \text{(E4.1.1.4)}$$

which is obtained by substituting the pressure $p = p_0$ into (E4.1.1.3). In (E4.1.1.4), we set $r = R$, where R is the radius of the rotating vessel, and obtain the paraboloid height

$$H = \frac{\omega^2 R^2}{2g},$$

where $H = z_R - z_0$.

E4.1.2. Force of Pressure on a Plane and on a Curved Wall

► **Force of pressure on a plane wall.** Consider the wetted part of a plane wall of area S immersed in a quiescent liquid at an angle α with the horizon. The resultant pressure force $\boldsymbol{P}$ exerted by the liquid on this wall is directed along the normal to its surface and is numerically equal to

$$P = p_c S. \qquad \text{(E4.1.2.1)}$$

Here p_c is the excess pressure at the center of gravity of the wetted part of the wall, which is calculated by the formula

$$p_c = p_0 - p_a + \rho g h_c^\circ, \qquad \text{(E4.1.2.2)}$$

where p_0 is the pressure on the free surface of the liquid, p_a is the external pressure (for example, the atmospheric pressure) on the outer dry side of the wall, and h_c° is depth of immersion of the wall center of gravity with respect to the free surface.

TABLE E4.1
The area, the coordinate of the center of gravity, and the moment of inertia of plane symmetric figures about the axis passing through the center of gravity (the figures are symmetric with respect to the vertical axis and the coordinate z is counted downwards from the upper point of the figure along the axis of symmetry).

Name and description of the figure	Coordinate of the center of gravity, z_0	Area of the figure, S	Moment of inertia, I_c
Rectangle with sides a and b; the side a is parallel to the Z-axis	$\frac{1}{2}a$	ab	$\frac{1}{12}a^3b$
Isosceles triangle with base a and height h	$\frac{2}{3}h$	$\frac{1}{2}ah$	$\frac{1}{36}ah^3$
Isosceles trapezium with bases a and b and height h (a is the upper base)	$\frac{1}{3}h\frac{a+2b}{a+b}$	$\frac{1}{2}h(a+b)$	$\frac{1}{36}h^3\frac{a^2+4ab+b^2}{a+b}$
Disk of radius R	$\frac{1}{2}R$	πR^2	$\frac{1}{4}\pi R^4$
Half-disk of radius R; rectilinear base is above	$\frac{4}{3\pi}R$	$\frac{1}{2}\pi R^2$	$\frac{9\pi^2-64}{72\pi}R^4$
Annular region with radii R and r ($R > r$)	R	$\pi(R^2-r^2)$	$\frac{1}{4}\pi(R^4-r^4)$
Ellipse with semiaxes a and b; semiaxis a is directed vertically	a	πab	$\frac{1}{4}\pi a^3b$

Formula (E4.1.2.2) is often written as $p_c = \rho g(H + h_c^\circ)$, where $H = \dfrac{p_0 - p_a}{\rho g}$ is the *piezometric head*. The surface $z = H$ is called the *piezometric plane*. The sign of the difference $(p_0 - p_a)$ determines the sign of H.

The vertical coordinate of the point of pressure force application to the wetted part of the plane wall of area S is determined by the formula

$$h_p = h_c + \frac{I_c}{h_c S}\sin^2\alpha, \qquad \text{(E4.1.2.3)}$$

where h_p and h_c are the vertical distances from the center of pressure and the center of gravity to the piezometric plane, α is the angle of inclination of the wall to the horizon, and I_c is the moment of inertia of the wall area about the horizontal axis passing through the wall center of gravity. In the case of a horizontal wall (for $\alpha = 0$), the center of gravity and center of pressure of the wall coincide.

Table E4.1 presents formulas for calculation of the moment of inertia of several plane figures about the axis passing through the center of gravity.

For $h_c > 0$, the piezometric plane lies above the center of gravity of the wetted surface, and the center of pressure lies below the center of gravity. For $h_c < 0$, the piezometric plane lies below the center of gravity of the wetted surface, and the center of pressure may be above the center of gravity (if $h_p < 0$). For $h_c = 0$, the piezometric plane passes through the center of gravity of the wetted surface, and the pressure forces are reduced to a couple of forces.

► **Force of pressure on a curved wall.** In the general case of an arbitrary curvilinear surface, the distributed load of the liquid pressure normal at each point of the surface can be reduced to the resultant vector and the principal moment.

For curvilinear walls symmetric with respect to the vertical plane (the most common situation in applications), the sum of elementary pressure forces can be reduced to a single resultant force lying in the plane of symmetry or to a couple of forces lying in the same plane. The value and the direction of the resultant force $\boldsymbol{P}$ are determined by two components, usually, horizontal and vertical.

The horizontal component of the pressure force on the curvilinear wall is equal to the pressure force on the projection of this wall onto the vertical plane of symmetry and is determined by the formula

$$P_1 = \rho g h_c S_1, \tag{E4.1.2.4}$$

where h_c is the depth of immersion of the center of gravity of the above-mentioned wall projection counted from the piezometric plane and S_1 is the area of the wall projection onto the vertical plane. The line of action of the force horizontal component P_1 passes through the center of pressure of the wall projection onto the vertical plane, lies in the plane of symmetry, and is displaced (downwards if $h_c > 0$ or upwards if $h_c < 0$) with respect to the center of gravity of this projection of the wall by the distance

$$\Delta h = \frac{I_c}{h_c S_1},$$

where I_c is the moment of inertia of the above-mentioned projection area of the wall about the horizontal axis passing through the center of gravity of this projection.

The vertical component of the force of pressure on the curvilinear wall is equal to the weight of liquid in the volume V bounded by the wall, the piezometric plane, and the vertical projected surface constructed on the wall contour; it is determined by the formula

$$P_2 = \rho g V. \tag{E4.1.2.5}$$

The vertical component of the pressure force passes through the center of gravity of the volume V (this volume is often called the volume of the body of pressure).

In formulas (E4.1.2.4) and (E4.1.2.5), it is assumed that the liquid is on one side of the wall and the pressure is constant (for example, equal to the atmospheric pressure) on the other (dry) side of the wall. The total pressure force on the wall is the geometric sum of the forces P_1 and P_2, and its absolute value is equal to

$$P = \sqrt{P_1^2 + P_2^2}.$$

The line of action of the total pressure force $\boldsymbol{P}$ passes through the point of intersection of the lines of action of the force components P_1 and P_2. The angle φ of inclination of the resultant force to the horizon is determined by the formula

$$\tan\varphi = \frac{P_2}{P_1}.$$

In the case of walls of constant curvature (cylindrical and spherical walls), the total pressure force passes trough the wall center of curvature.

In the case of liquid excess pressures on the side of the wetted side of the wall, all components and the total force are directed from the liquid to the wall (outside from the inside).

In the case of two-sided action of liquids on the wall, one first determines the horizontal and vertical components on each side of the wall under the assumption of one-sided action of the liquid, and then the total horizontal and vertical components of action of both liquids are determined.

If the curvilinear surface under study is intersected by the vertical axis at more than one point, it is necessary to divide this surface into simple parts (a simple part of a surface is its part intersected by the vertical axis only at a single point) and calculate the vertical component for each of the parts. Then the obtained vertical components are summed algebraically. The same process is used if the surface under study is intersected by the horizontal axis at more than one point.

E4.1.3. Archimedes Principle. Stability of Floating Bodies

► **Archimedes principle.** The body (completely or partly) immersed in a liquid is under the action of the buoyancy force F_A numerically equal to the weight of the liquid in the volume displaced by the body (*Archimedes principle*). Thus, we have

$$F_A = \rho g V_b,$$

where V_b is the volume of the liquid displaced by the body. The force F_A passes through the center of gravity of the volume displaced by the body, which is called the *center of buoyancy*.

► **Stability of floating bodies.** If a floating body is in equilibrium, then its center of gravity and the center of buoyancy lie on the common vertical line called the *axis of buoyancy* (for a body symmetric with respect to a plane, the axis of buoyancy lies in this plane). For a floating body completely immersed in a liquid (underwater floating) to be in stable equilibrium, it is necessary that the body center of gravity be below the center of buoyancy.

If a body is floating on the surface of a liquid (surface floating), then stable equilibrium is possible in several cases where the body center of gravity lies above the center of buoyancy. For a body to be in stable equilibrium in the case of surface floating, it is necessary that, when heeled (the body axis of buoyancy is inclined by an angle θ, see Fig. E4.1), the body *metacenter* M (i.e., the point of intersection of the line of the Archimedes force action with the axis of buoyancy) lie above the center of gravity C. In Fig. E4.1, point B' is the center of buoyancy when the body is heeled, and point B corresponds to the original position of the center of buoyancy on the axis of buoyancy in equilibrium.

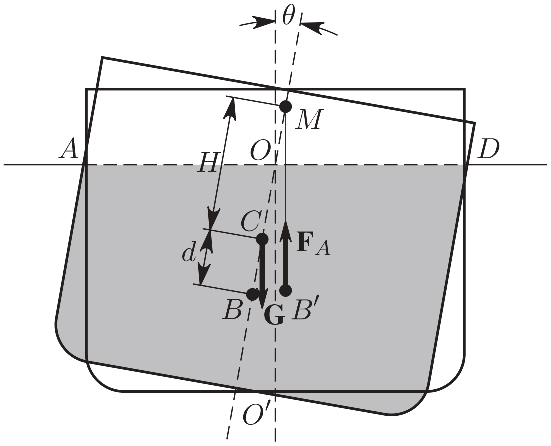

Figure E4.1. Stability of floating bodies. The center of gravity C lies above the center of buoyancy B.

The distance H between points M and C is called the *metacentric height*. For small angles of heeling, the metacentric height is determined by the formula

$$H = \frac{I}{V} - d \qquad (\theta \ll 1),$$

where I is the moment of inertia of the floating section AD about the axis of suspension OO' (see Fig. E4.1), V is the body volume immersed in the liquid, and d is the excess of the center of gravity over the center of buoyancy in equilibrium. For a floating body to be in stable equilibrium, it is necessary that the following condition be satisfied:

$$H = \frac{I}{V} - d > 0. \qquad \text{(E4.1.3.1)}$$

The stability of a floating body must be verified for the axis about which the moment of inertia of the floatation section is minimal.

Example. A wooden bar of square cross-section $a \times a$ and height h and of density ρ_b is floating in a liquid of density ρ_l ($\rho_l > \rho_b$). Find the maximum value of the bar height $h_{\max}$ for which the floating is still stable.

To solve this problem, we use formula (E4.1.3.1) and substitute the minimum value of the metacentric height into this formula:

$$\frac{I}{V} - d = 0. \qquad \text{(E4.1.3.2)}$$

The bar immersion depth x is determined by the Archimedes principle: the body weight must be equal to the weight of the liquid displaced by the body, $\rho_b g a^2 h = \rho_f g a^2 x$. We have

$$x = \frac{\rho_b}{\rho_f} h.$$

Now we calculate the quantities occurring on the left-hand side of (E4.1.3.2):

$$I = \frac{1}{12} a^4, \quad V = a^2 x = \frac{\rho_b}{\rho_f} a^2 h, \quad d = \frac{h}{2} - \frac{x}{2} = \frac{1}{2} h \left(1 - \frac{\rho_b}{\rho_f}\right).$$

(To find the moment of inertia I, we used the data for the rectangle given in Table E4.1 for $a = b$.) Substituting the values thus obtained into formula (E4.1.3.2), we have

$$\frac{1}{12} \frac{\rho_f}{\rho_b} \frac{a^2}{h} - \frac{1}{2} h \left(1 - \frac{\rho_b}{\rho_f}\right) = 0.$$

Resolving this equation for h, we obtain the maximal height of the bar:

$$h_{\max} = \frac{a}{\sqrt{6\delta(1-\delta)}}, \qquad \delta = \frac{\rho_b}{\rho_f}.$$

E4.2. Hydrodynamic Equations and Boundary Conditions

In this section, we present equations and boundary conditions used in solving hydrodynamic problems. Their detailed derivation, as well as an analysis of scope, various physical statements and solutions of related problems, and applied issues can be found, e.g., in the books presented in the literature at the end of this chapter. We consider incompressible fluids with constant density ρ and dynamic viscosity μ.

E4.2.1. Laminar Flows. Navier–Stokes Equations. Euler Equations

► **Preliminaries.** Experiments show that two qualitatively different forms of fluid motion are possible. In particular, at small velocities of the flow in a circular tube, the fluid moves in straight lines parallel to one another and to the tube axis. Such a “smooth" flow without mixing is said to be *laminar*. In the general case of laminar steady-state flow, the fluid elements move along stream lines (a stream line is a line such that the tangent at each of its points coincides with the direction of the fluid particle velocity at this point). For large velocities of motion, an intensive mixing of the fluid is observed in the tube, and this mixing is characterized by unsteady chaotic motion of the fluid elements in the transverse direction (there appear velocity components perpendicular to the tube axis). Such a flow regime is said to be *turbulent*.

In Subsection E4.2.1, equations for laminar flows of fluids are considered, and in Subsection E4.2.3, equations for turbulent flows.

For brevity, in what follows, we often refer to “laminar flows” simply as “flows.”

► **Navier–Stokes equations for a viscous incompressible fluid.** The closed system of equations of motion for a viscous incompressible Newtonian fluid consists of the continuity equation

$$\frac{\partial V_X}{\partial X}+\frac{\partial V_Y}{\partial Y}+\frac{\partial V_Z}{\partial Z}=0 \tag{E4.2.1.1}$$

and three Navier–Stokes equations

$$\begin{aligned}
&\frac{\partial V_X}{\partial t}+V_X\frac{\partial V_X}{\partial X}+V_Y\frac{\partial V_X}{\partial Y}+V_Z\frac{\partial V_X}{\partial Z}\\
&\qquad=-\frac{1}{\rho}\frac{\partial P}{\partial X}+\nu\left(\frac{\partial^2 V_X}{\partial X^2}+\frac{\partial^2 V_X}{\partial Y^2}+\frac{\partial^2 V_X}{\partial Z^2}\right)+g_X,\\
&\frac{\partial V_Y}{\partial t}+V_X\frac{\partial V_Y}{\partial X}+V_Y\frac{\partial V_Y}{\partial Y}+V_Z\frac{\partial V_Y}{\partial Z}\\
&\qquad=-\frac{1}{\rho}\frac{\partial P}{\partial Y}+\nu\left(\frac{\partial^2 V_Y}{\partial X^2}+\frac{\partial^2 V_Y}{\partial Y^2}+\frac{\partial^2 V_Y}{\partial Z^2}\right)+g_Y,\\
&\frac{\partial V_Z}{\partial t}+V_X\frac{\partial V_Z}{\partial X}+V_Y\frac{\partial V_Z}{\partial Y}+V_Z\frac{\partial V_Z}{\partial Z}\\
&\qquad=-\frac{1}{\rho}\frac{\partial P}{\partial Z}+\nu\left(\frac{\partial^2 V_Z}{\partial X^2}+\frac{\partial^2 V_Z}{\partial Y^2}+\frac{\partial^2 V_Z}{\partial Z^2}\right)+g_Z,
\end{aligned} \tag{E4.2.1.2}$$

Equations (E4.2.1.1) and (E4.2.1.2) are written in an orthogonal Cartesian system X, Y, Z in physical space; t is time; g_X, g_Y, and g_Z are the mass force (e.g., the gravity force) density components; $\nu=\mu/\rho$ is the kinematic viscosity of the fluid. The three components V_X, V_Y, V_Z of the fluid velocity and the pressure P are the unknowns.

By introducing the fluid velocity vector $\mathbf{V}=\mathbf{i}_X V_X+\mathbf{i}_Y V_Y+\mathbf{i}_Z V_Z$, where $\mathbf{i}_X$, $\mathbf{i}_Y$, and $\mathbf{i}_Z$ are the unit vectors of the Cartesian coordinate system, and by using the symbolic differential operators

$$\nabla\equiv\mathbf{i}_X\frac{\partial}{\partial X}+\mathbf{i}_Y\frac{\partial}{\partial Y}+\mathbf{i}_Z\frac{\partial}{\partial Z},\qquad \Delta\equiv\frac{\partial^2}{\partial X^2}+\frac{\partial^2}{\partial Y^2}+\frac{\partial^2}{\partial Z^2},$$

one can rewrite system (E4.2.1.1), (E4.2.1.2) in the concise vector form

$$\nabla\cdot\mathbf{V}=0, \tag{E4.2.1.3}$$

$$\frac{\partial\mathbf{V}}{\partial t}+\big(\mathbf{V}\cdot\nabla\big)\mathbf{V}=-\frac{1}{\rho}\nabla P+\nu\,\Delta\mathbf{V}+\mathbf{g}. \tag{E4.2.1.4}$$

The continuity and Navier–Stokes equations in the cylindrical and spherical coordinate systems can be obtained using the expressions of the operators ∇ and Δ given in Section S3.2.

► **Dimensionless form of equations.** To analyze the hydrodynamic equations (E4.2.1.3) and (E4.2.1.4), it is convenient to introduce dimensionless variables and unknown functions as follows:

$$\tau = \frac{Ut}{a}, \quad x = \frac{X}{a}, \quad y = \frac{Y}{a}, \quad z = \frac{Z}{a}, \quad \mathbf{v} = \frac{\mathbf{V}}{U}, \quad p = \frac{P}{\rho U^2},$$

where a and U are the characteristic length and the characteristic velocity, respectively. As a result, we obtain

$$\nabla \cdot \mathbf{v} = 0, \tag{E4.2.1.5}$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\nabla p + \frac{1}{\mathsf{Re}}\Delta \mathbf{v} + \frac{1}{\mathsf{Fr}}\frac{\mathbf{g}}{g}. \tag{E4.2.1.6}$$

In Eq. (E4.2.1.6), the following basic dimensionless state-geometric parameters of the flow are used:

$$\mathsf{Re} = \frac{aU}{\nu} \text{ is the } \textit{Reynolds number}, \qquad \mathsf{Fr} = \frac{U^2}{ga} \text{ is the } \textit{Froude number}.$$

Small values of Reynolds numbers correspond to slow (“creeping”) flows and high Reynolds numbers, to rapid flows. Since these limit cases contain a small or large dimensionless parameter, one can efficiently use various modifications of the perturbation method.

► **Euler equations for an ideal incompressible fluid.** In the special case $V_X = V_Y = V_Z = 0$, the Navier–Stokes equations (E4.2.1.2) become the *equilibrium equations of a quiescent incompressible fluid*, which are independent of the fluid viscosity. From the equilibrium equations of a quiescent incompressible fluid, one can derive the laws of hydrostatics described in Section E4.1.

In the special case $\nu = 0$, the Navier–Stokes equations (E4.2.1.2) become the *Euler equations*,

$$\begin{aligned}
\frac{\partial V_X}{\partial t} + V_X\frac{\partial V_X}{\partial X} + V_Y\frac{\partial V_X}{\partial Y} + V_Z\frac{\partial V_X}{\partial Z} &= -\frac{1}{\rho}\frac{\partial P}{\partial X} + g_X,\\
\frac{\partial V_Y}{\partial t} + V_X\frac{\partial V_Y}{\partial X} + V_Y\frac{\partial V_Y}{\partial Y} + V_Z\frac{\partial V_Y}{\partial Z} &= -\frac{1}{\rho}\frac{\partial P}{\partial Y} + g_Y,\\
\frac{\partial V_Z}{\partial t} + V_X\frac{\partial V_Z}{\partial X} + V_Y\frac{\partial V_Z}{\partial Y} + V_Z\frac{\partial V_Z}{\partial Z} &= -\frac{1}{\rho}\frac{\partial P}{\partial Z} + g_Z,
\end{aligned} \tag{E4.2.1.7}$$

which, together with the continuity equation (E4.2.1.1), describe the flows of ideal (inviscid) incompressible fluid.

We assume that the flow is stationary and of the mass forces only the gravity force is acting, which is characterized by the values $g_X = g_Y = 0$, $g_Z = -g$ (the coordinate Z is counted off from the Earth’s surface, and g is the free fall acceleration). In this case, the Euler equations admit the *Bernoulli integral*; namely, the following conservation law is satisfied on any stream line:

$$\frac{V^2}{2} + \frac{P}{\rho} + gZ = \text{const}, \tag{E4.2.1.8}$$

where $V = \sqrt{V_X^2 + V_Y^2 + V_Z^2}$ is the modulus of the fluid velocity. In the general case, the Bernoulli integral constant on the right-hand side in (E4.2.1.8) is different for different stream lines.

Example. Assume that there is a small hole in the wall of a wide vessel at a depth h. Find the velocity of the liquid flow through the hole.

Consider the stream line issuing from the liquid free surface in the vessel and passing through the hole. For this stream line, we write the Bernoulli integral (E4.2.1.8) in the form

$$\frac{V_1^2}{2}+\frac{P_1}{\rho}+g_1Z=\frac{V_2^2}{2}+\frac{P_2}{\rho}+gZ_2, \tag{E4.2.1.9}$$

where subscript 1 denotes the quantities on the liquid surface and the subscript 2 denotes the quantities on the hole output. The coordinate Z is counted off from the hole. Substituting the values $V_1\approx 0$, $P_1=P_2=P_a$, $Z_1=h$, $Z_2=0$ into (E4.2.1.9), we obtain the velocity of the liquid flow from the hole:

$$V_0=\sqrt{2gh}.$$

This formula is called the *Torricelli formula*.

The motion of an ideal fluid is said to be *potential* if the fluid velocity vector admits the representation

$$\mathbf{V}=\operatorname{grad}\varphi, \tag{E4.2.1.10}$$

where the function φ is called the *potential*. Because of the continuity equation (E4.2.1.1), the potential satisfies the Laplace equation

$$\Delta\varphi=0,$$

and the pressure is determined by the Cauchy–Lagrange integral

$$\frac{\partial\varphi}{\partial t}+\frac{V^2}{2}+\frac{P}{\rho}+gZ=f(t). \tag{E4.2.1.11}$$

where $f(t)$ is an arbitrary function. When writing this integral, it was assumed that of the mass forces only the gravity force is acting. Relation (E4.2.1.11) is satisfied at all points of the region of potential flow of an ideal fluid; to find the function $f(t)$, it suffices to know the left-hand side of the integral as a function of time at any point of the flow.

In the special case of a steady-state potential motion of a fluid, the Cauchy–Lagrange integral (E4.2.1.11) coincides with the Bernoulli integral (E4.2.1.8), where the constant on the right-hand side is the same for all stream lines.

Note that the Euler equations (E4.2.1.7) have a very restricted domain of application (especially in problems of chemical technology).

E4.2.2. Initial Conditions and the Simplest Boundary Conditions

For the solution of system (E4.2.1.1), (E4.2.1.2) to determine the velocity and pressure fields uniquely, we must impose initial and boundary conditions.

► **Initial conditions.** In nonstationary problems, where the terms with partial derivatives with respect to time are retained in the equations of motion, the initial velocity field must be given in the entire flow region and satisfy the continuity equation (E4.2.1.1) there. The initial pressure field need not be given, since the equations do not contain the derivative of pressure with respect to time.

► **Boundary conditions on the surface of a solid body.** As a rule, the region occupied by a moving fluid is not the entire space but only part of it bounded by some surfaces. According to whether the point at infinity belongs to the flow region or not, the problem of finding the unknown functions is called the exterior or interior problem of hydrodynamics, respectively.

On the surface S of a solid body moving in a flow of a viscous fluid, the no-slip condition is imposed. This condition says that the vector $\mathbf{V}|_S$ of the fluid velocity on the surface of the solid is equal to the vector $\mathbf{V}_0$ of the solid velocity. If the solid is at rest, then $\mathbf{V}|_S = \mathbf{0}$. In the projections on the normal $\mathbf{n}$ and the tangent $\boldsymbol{\tau}$ to the surface S, this condition reads

$$V_n\big|_S = 0, \qquad V_\tau\big|_S = 0.$$

More complicated boundary conditions are posed on an interface between two fluids (for example, see boundary conditions (E4.7.1.1)–(E4.7.1.4)).

► **Boundary condition far from the body for the translational flow.** To solve the exterior hydrodynamic problem, one must impose a condition at infinity (that is, far from the body, the drop, or the bubble).

For uniform translational flow with velocity $\mathbf{U}_\mathrm{i}$ around a finite body, the boundary condition far from the body has the form

$$\mathbf{V} \to \mathbf{U}_\mathrm{i} \qquad \text{as} \quad R \to \infty, \tag{E4.2.2.1}$$

where $R = \sqrt{X^2 + Y^2 + Z^2}$.

► **Shear flows.** An arbitrary stationary velocity field $\mathbf{V}(\mathbf{R})$ in an incompressible medium can be approximated near the point $\mathbf{R} = \mathbf{0}$ by two terms of the Taylor series:

$$\begin{gathered} V_k(\mathbf{R}) = V_k(\mathbf{0}) + G_{km} X_m, \\ G_{km} \equiv (\partial V_k/\partial X_m)_{\mathbf{R}=0}, \quad G_{11} + G_{22} + G_{33} = 0. \end{gathered} \tag{E4.2.2.2}$$

Here V_k and G_{km} are the fluid velocity and the shear tensor components in the Cartesian coordinates X_1, X_2, X_3. The sum is taken over the repeated index m; since the fluid is incompressible, it follows that the sum of the diagonal entries G_{mm} is zero.

For viscous flows around particles whose size is much less than the characteristic size of flow inhomogeneities, the velocity distribution (1.1.15) can be viewed as the velocity field far from the particle. The special case $G_{km} = 0$ corresponds to a uniform translational flow. For $V_k(\mathbf{0}) = 0$, Eq. (E4.2.2.2) describes the velocity field in an arbitrary linear shear flow.

Example (Simple shear flow). For the simple shear (Couette) flow, we have

$$V_X = GY, \qquad V_Y = 0, \qquad V_Z = 0. \tag{E4.2.2.3}$$

In this case, G is called the gradient of the flow rate or the shear rate. The Couette flow occurs between two planes one of which is immovable and the other moves at a constant velocity parallel to the first plane.

The expressions (E4.2.2.3) give the exact solution of the Navier–Stokes equations (E4.2.1.1)–(E4.2.1.2) in the case of potential mass forces.

E4.2.3. Turbulent Flows. Reynolds Equations

► **Reynolds equations.** Formally, stationary solutions of the Navier–Stokes equations are possible for any Reynolds numbers. But practically, only stable flows with respect to small perturbations, always present in the flow, can exist. For sufficiently high Reynolds numbers, the stationary solutions become unstable; i.e., the amplitude of small perturbations increases with time. For this reason, stationary solutions can only describe real flows at not too high Reynolds numbers.

The flow in the boundary layer on a flat plate is laminar up to $\mathsf{Re}_X = U_\mathrm{i} X/\nu \approx 3.5\times10^5$ (U_i is the unperturbed fluid velocity far from the plate, and the coordinate X is counted along the plate from the front edge), and that in a circular smooth tube is laminar for $\mathsf{Re} = a\langle V\rangle/\nu < 1500$ (a is the radius of a tube, and $\langle V\rangle$ is the mean flow rate velocity).

For higher Reynolds numbers, the laminar flow loses its stability and a transient regime of development of unstable modes takes place. For $\mathsf{Re}_X > 10^7$ and $\mathsf{Re} > 2500$, a fully developed regime of turbulent flow is established, which is characterized by chaotic variations in the basic macroscopic flow parameters in time and space.

When mathematically describing a fully developed turbulent motion of fluid, it is common to represent the velocity components and pressure in the form

$$V_i = \overline{V_i} + V_i', \quad P = \overline{P} + P', \tag{E4.2.3.1}$$

where the bar and prime denote the time-average and fluctuating components, respectively. The averages of the fluctuations are zero, $\overline{V_i'} = \overline{P'} = 0$.

The representation (E4.2.3.1) of the hydrodynamic parameters of turbulent flow as the sum of the average and fluctuating components followed by the averaging process made it possible, on the basis of the continuity equation (E4.2.1.3) and the Navier–Stokes equations (E4.2.1.4), to obtain (under some assumptions) the Reynolds equations

$$\begin{gathered}\nabla \cdot \overline{\mathbf{V}} = 0,\\ \frac{\partial \overline{\mathbf{V}}}{\partial t} + (\overline{\mathbf{V}} \cdot \nabla)\overline{\mathbf{V}} = -\frac{1}{\rho}\nabla \overline{P} + \nu \Delta \overline{\mathbf{V}} + \mathbf{g} + \frac{1}{\rho}\nabla \cdot \boldsymbol{\sigma}^{\mathrm{t}}\end{gathered} \tag{E4.2.3.2}$$

for the averaged pressure and velocity fields. These equation contain the Reynolds turbulent shear stress tensor $\boldsymbol{\sigma}^{\mathrm{t}}$, whose components are defined as

$$\sigma_{ij}^{\mathrm{t}} = -\rho \overline{V_i' V_j'}. \tag{E4.2.3.3}$$

The variable $\rho \overline{V_i' V_j'}$ is the average rate at which the turbulent fluctuations transfer the jth momentum component along the ith axis.

► **The closure problem. Turbulent viscosity.** Unlike the Navier–Stokes equations completed by the continuity equation, the Reynolds equations form an unclosed system of equations, since these contain the *a priori* unknown turbulent stress tensor $\boldsymbol{\sigma}^{\mathrm{t}}$ with components (E4.2.3.3). Additional hypotheses must be invoked in order to close system (E4.2.3.2).

So far the closure problem for the system of Reynolds equations has not been solved theoretically in a conclusive way. In engineering calculations, various assumptions that the Reynolds stresses depend on the average turbulent flow parameters are often adopted as closure conditions. These conditions are usually formulated on the basis of experimental data, dimensional considerations, analogies with molecular rheological models, etc.

One traditional approach to the closure of the Reynolds equations is outlined below. This approach is based on Boussinesq's model of turbulent viscosity completed by Prandtl's hypothesis. For simplicity, we confine our consideration to the case of simple shear flow, where the transverse coordinate $Y = X_2$ is measured from the wall (the results are also applicable to turbulent boundary layers). According to Boussinesq's model, the only nonzero component of the Reynolds turbulent shear stress tensor and the divergence of this tensor are defined as

$$\sigma_{1,2}^{\mathrm{t}} = \rho \nu_{\mathrm{t}} \frac{\partial \overline{V}}{\partial Y}, \qquad \nabla \cdot \boldsymbol{\sigma}^{\mathrm{t}} = \rho \frac{\partial}{\partial Y}\left(\nu_{\mathrm{t}} \frac{\partial \overline{V}}{\partial Y}\right), \tag{E4.2.3.4}$$

where $\overline{V}$ stands for the longitudinal average velocity component. Formulas (E4.2.3.4) contain the turbulent (eddy) viscosity ν_{t}, which is not a physical constant but is a function of geometric and kinematic flow parameters. It is necessary to specify this function to close the Reynolds equations.

Following Prandtl, we have

$$\nu_{\mathrm{t}} = \kappa^2 Y^2 \left| \frac{\partial \overline{V}}{\partial Y} \right|, \qquad \text{(E4.2.3.5)}$$

where $\kappa = 0.4$ is the von Kármán empirical constant (in the literature, this constant is most frequently taken to be $\kappa = 0.40$ or 0.41, although other values can sometimes be encountered).

There are a number of other methods for closing the Reynolds equations, also based on the notion of turbulent viscosity. These methods, as well as other models and methods of turbulence theory, can be found in the literature listed at the end of this chapter.

E4.3. Hydrodynamics of Thin Films

E4.3.1. Preliminary Remarks. Different Regimes of Film Type Flow

Film type flows are widely used in chemical technology (in contact devices of absorption, chemosorption, and rectification columns as well as evaporators, dryers, heat exchangers, film chemical reactors, extractors, and condensers).

As a rule, the liquid and the gas phase are simultaneously fed into an apparatus where the fluids undergo physical and chemical treatment. Therefore, generally speaking, there is a dynamic interaction between the phases until the flooding mode sets in the countercurrent flows of gas and liquid. However, for small values of gas flow rate one can neglect the dynamic interaction and assume that the liquid flow in a film is due to the gravity force alone.

The value of the Reynolds number $\mathsf{Re} = Q/\nu$, where Q is the volume flow rate per unit film width, determines whether the flow in the gravitational film is laminar, wave, or turbulent. It is well known that laminar flow becomes unstable at the critical value $\mathsf{Re}_* = 2$ to 6. However, the point starting from which the waves actually occur is noticeably shifted downstream. Even in the range $6 \le \mathsf{Re} \le 400$, corresponding to wave flows, a considerable part of the film remains wave-free. Since this part is much larger than the initial part where the velocity profile and the film width reach their steady-state values, we see that for films in which viscous and gravity forces are in balance, the hydrodynamic laws of steady-state laminar flow virtually determine the rate of mass exchange in various apparatuses, like packed absorbing and fractionating columns, widely used in chemical and petroleum industry. In these columns, the films flow over the packing surface whose linear dimensions do not exceed a few centimeters (Raschig rings, Palle rings, Birle seats, etc.).

Paradoxically, the range of flow rates (or Reynolds numbers) for which the assumption of laminar flow can be used in practice is bounded below (rather than above). Indeed, there is a threshold value $Q_{\min}$ of the volume flow rate per unit film width such that for $Q < Q_{\min}$ the flow in separate jets is energetically favorable. It was theoretically established that

$$Q_{\min} = 2.15 \left(\frac{\nu \sigma^3}{\rho^3 g} \right)^{1/5} (1 - \cos\theta)^{3/5},$$

where σ is the surface tension of the liquid and θ is the wetting angle of the wall material and the liquid (see Figure E4.2), determined by Young's fundamental relation

$$\sigma_{\mathrm{gw}} = \sigma \cos\theta + \sigma_{\mathrm{fw}},$$

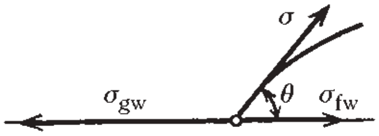

Figure E4.2. Definition of the wetting angle.

where σ_{gw} and σ_{fw} are the specific excess surface energies for the gas–wall and liquid–wall interfaces.

Recently, the criterion of nonbreaking film flow was thermodynamically substantiated with the aid of Prigogine's principle of minimum entropy production, including the case of a double film flow.

In practice, Q_{min} can be reduced by wall hydrophilization — that is, by treating the surface with alcohol, which decreases the wetting angle.

E4.3.2. Films on an Inclined Plane and on a Cylindrical Surface

► **Film flowing by gravity on an inclined plane.** Consider a thin liquid film flowing by gravity on a solid plane surface (Figure E4.3). Let α be the angle of inclination. We assume that the motion is sufficiently slow, so that we can neglect inertial forces (that is, convective terms) compared with the viscous friction and the gravity force. Let the film thickness h (which is assumed to be constant) be much less than the film length. In this case, in the first approximation, the normal component of the liquid velocity is small compared with the longitudinal component, and we can neglect the derivatives along the film surface compared with the normal derivatives.

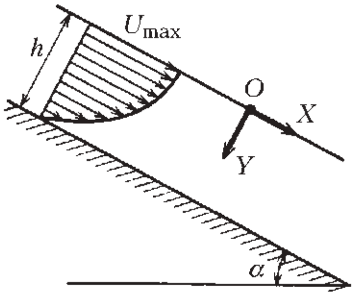

Figure E4.3. Steady waveless laminar flow in thin film on an inclined plane.

These assumptions result in the one-dimensional velocity and pressure profiles $V = V(Y)$ and $P = P(Y)$, where Y is the coordinate measured along the normal to the film surface. The corresponding hydrodynamic equations of thin films expressing the balance of viscous and gravity forces have the form

$$\begin{gathered} \mu\frac{d^2V}{dY^2} + \rho g\sin\alpha = 0,\\ \frac{dP}{dY} - \rho g\cos\alpha = 0. \end{gathered} \tag{E4.3.2.1}$$

To these equations one should add the boundary conditions

$$\begin{aligned} \frac{dV}{dY} &= 0, \quad P = P_0 \qquad \text{at} \quad Y = 0,\\ V &= 0 \qquad\qquad\quad \text{at} \quad Y = h, \end{aligned} \tag{E4.3.2.2}$$

which show that the tangential stress is zero, the pressure is equal to the atmosphere pressure at the free surface, and the no-slip condition is satisfied at the surface of the plane.

The solution of problem (E4.3.2.1), (E4.3.2.2) has the form

$$\begin{aligned} V &= U_{\max}(1-y^2),\\ P &= P_0 + \rho g h \cos\alpha\, y, \end{aligned} \qquad \text{(E4.3.2.3)}$$

where $U_{\max} = \frac{1}{2}(g/\nu)h^2 \sin\alpha$ is the maximum flow velocity (the velocity at the free boundary) and $y = Y/h$ is the dimensionless transverse coordinate.

The volume flow rate per unit width is given by the formula

$$Q = \int_0^h V(Y)\,dY = \frac{gh^3 \sin\alpha}{3\nu} = \tfrac{2}{3}U_{\max}h. \qquad \text{(E4.3.2.4)}$$

The mean flow rate velocity $\langle V\rangle$ is equal to $2/3$ of the maximum velocity,

$$\langle V\rangle = \tfrac{2}{3}U_{\max}.$$

Let us find the Reynolds number for the film flow:

$$\mathsf{Re} = \frac{Q}{\nu} = \frac{gh^3 \sin\alpha}{3\nu^2}.$$

This allows us to express the film thickness via the Reynolds number and the volume flow rate per unit width:

$$h = \left(\frac{3\nu^2}{g\sin\alpha}\mathsf{Re}\right)^{1/3} = \left(\frac{3\nu}{g\sin\alpha}Q\right)^{1/3}.$$

► **Film on a cylindrical surface.** Consider a thin liquid film of thickness h flowing by gravity on the surface of a vertical circular cylinder of radius a. In the cylindrical coordinates $\mathcal{R}$, φ, Z, the only nonzero component of the liquid velocity satisfies the equation

$$\mu\left(\frac{d^2V_Z}{d\mathcal{R}^2} + \frac{1}{\mathcal{R}}\frac{\partial V_Z}{\partial \mathcal{R}}\right) + \rho g = 0. \qquad \text{(E4.3.2.5)}$$

The boundary conditions on the wall and on the free surface can be written as

$$V_Z = 0 \quad \text{at} \quad \mathcal{R} = a, \qquad \frac{dV_Z}{d\mathcal{R}} = 0 \quad \text{at} \quad \mathcal{R} = a + h. \qquad \text{(E4.3.2.6)}$$

The solution of problem (E4.3.2.5), (E4.3.2.6) is given by the formula

$$V_Z(\mathcal{R}) = \frac{\rho g}{4\mu}\left\{a^2 - \mathcal{R}^2 + \left[(a+h)^2 - a^2\right]\frac{\ln(\mathcal{R}/a)}{\ln(1+h/a)}\right\}.$$

E4.4. Fluid Flows in Tubes

E4.4.1. Laminar Flows in Tubes of Various Cross-Sections

Laminar steady-state fluid flows in tubes are often encountered in practice (water-, gas- and oil pipelines, heat exchangers, etc.). It is worth noting that in these cases the corresponding hydrodynamic equations admit an exact closed-form solution. In what follows, we describe the most important results in that direction.

► **Statement of the problem.** Consider a laminar steady-state fluid flow in a rectilinear tube of constant cross-section. The fluid streamlines in such systems are strictly parallel (we neglect the influence of the tube endpoints on the flow). We shall use the Cartesian coordinates X, Y, Z with the Z-axis directed along the flow. Let us take into account the fact that the transverse velocity components of the fluid are zero and the longitudinal component depends only on the transverse coordinates. In this case, the continuity equation (E4.2.1.1) and the first two Navier–Stokes equations in (E4.2.1.2) are satisfied automatically, and it follows from the third equation in (E4.2.1.2) that

$$\frac{\partial^2 V}{\partial X^2}+\frac{\partial^2 V}{\partial Y^2}=\frac{1}{\mu}\frac{dP}{dZ}, \qquad \text{(E4.4.1.1)}$$

where $V \equiv V_Z$ is the longitudinal velocity component.

Equation (E4.4.1.1) must be supplemented by the no-slip condition

$$V=0 \qquad \text{on the tube surface.} \qquad \text{(E4.4.1.2)}$$

The pressure gradient dP/dZ in the steady state is constant along the tube and can be represented in the form

$$\frac{dP}{dZ}=-\frac{\Delta P}{L}, \qquad \text{(E4.4.1.3)}$$

where $\Delta P > 0$ is the total pressure drop along a tube part of length L.

The main flow characteristics are the volume flow rate

$$Q=\int_S V\,dS \qquad \text{(E4.4.1.4)}$$

and the mean flow rate velocity

$$\langle V\rangle=\frac{Q}{S_*}, \qquad \text{(E4.4.1.5)}$$

where S_* is the area of the tube cross-section.

► **Plane channel.** First, consider the flow between two infinite parallel planes at a distance $2h$ from each other. The coordinate X is measured from one of the planes along the normal. Since the fluid velocity is independent of the coordinate Y, we can rewrite (E4.4.1.1) in the form

$$\frac{d^2V}{dX^2}=-\frac{\Delta P}{\mu L}.$$

The solution of this equation under the no-slip boundary conditions on the planes ($V=0$ for $X=0$ and $X=2h$) has the form

$$V=\frac{\Delta P}{2\mu L}X(2h-X). \qquad \text{(E4.4.1.6)}$$

Formula (E4.4.1.6) describes the parabolic velocity field in a *plane Poiseuille flow* symmetric with respect to the midplane $X=h$ of the channel.

The volume flow rate per unit width of the channel can be found by integrating (E4.4.1.6) over the cross-section:

$$Q=\frac{2h^3\Delta P}{3\mu L}. \qquad \text{(E4.4.1.7)}$$

The mean flow rate velocity is

$$\langle V \rangle = \frac{h^2 \Delta P}{3\mu L}. \tag{E4.4.1.8}$$

The maximum velocity is attained on the midplane of the channel:

$$U_{\max} = \frac{h^2 \Delta P}{2\mu L} \qquad (\text{at} \quad X = h).$$

► **Circular tube.** In the case of a circular tube, Eq. (E4.4.1.1), with regard to (E4.4.1.3), acquires the form

$$\frac{1}{\mathcal{R}} \frac{d}{d\mathcal{R}} \left(\mathcal{R} \frac{dV}{d\mathcal{R}} \right) = -\frac{\Delta P}{\mu L}, \qquad \mathcal{R} = \sqrt{X^2 + Y^2}. \tag{E4.4.1.9}$$

The solution of this equation under the no-slip condition on the surface of a tube of radius a ($V = 0$ for $\mathcal{R} = a$) describes an *axisymmetric Poiseuille flow* with parabolic velocity profile:

$$V = \frac{\Delta P}{4\mu L} \left(a^2 - \mathcal{R}^2 \right). \tag{E4.4.1.10}$$

The volume flow rate can be obtained by integrating over the cross-section:

$$Q = 2\pi \int_0^a \mathcal{R} V \, d\mathcal{R} = \frac{\pi a^4 \Delta P}{8\mu L}. \tag{E4.4.1.11}$$

By using (E4.4.1.5), we obtain the mean flow rate velocity

$$\langle V \rangle = \frac{a^2 \Delta P}{8\mu L}. \tag{E4.4.1.12}$$

The maximum fluid velocity is attained at the tube axis:

$$U_{\max} = \frac{a^2 \Delta P}{4\mu L} \qquad (\text{at } \mathcal{R} = 0). \tag{E4.4.1.13}$$

► **Annular channel between two coaxial circular cylinders.** Now consider the flow in an annular channel between two coaxial circular cylinders of radii a_1 and a_2 ($a_1 < a_2$). In this case, Eq. (E4.4.1.9) remains valid. The solution of this equation satisfying the no-slip conditions on the cylinder surfaces,

$$V = 0 \quad \text{at} \quad \mathcal{R} = a_1, \qquad V = 0 \quad \text{at} \quad \mathcal{R} = a_2,$$

has the form

$$V = \frac{\Delta P}{4\mu L} \left[a_2^2 - \mathcal{R}^2 + \frac{a_2^2 - a_1^2}{\ln(a_2/a_1)} \ln \frac{\mathcal{R}}{a_2} \right]. \tag{E4.4.1.14}$$

The volume flow rate is given by the formula

$$Q = \frac{\pi \Delta P}{8\mu L} \left[a_2^4 - a_1^4 - \frac{(a_2^2 - a_1^2)^2}{\ln(a_2/a_1)} \right]. \tag{E4.4.1.15}$$

► **Tube of elliptic cross-section.** Now consider a tube whose cross-section is an ellipse with semiaxes a and b. The surface of this tube is given by the equation

$$\left(\frac{X}{a}\right)^2+\left(\frac{Y}{b}\right)^2=1. \tag{E4.4.1.16}$$

The solution of Eq. (E4.4.1.1) under the no-slip condition on the surface (E4.4.1.16) has the form

$$V=\frac{a^2b^2\Delta P}{2\mu L(a^2+b^2)}\left(1-\frac{X^2}{a^2}-\frac{Y^2}{b^2}\right). \tag{E4.4.1.17}$$

The volume flow rate is

$$Q=\frac{\pi\Delta P}{4\mu L}\frac{a^3b^3}{a^2+b^2}. \tag{E4.4.1.18}$$

By using formula (E4.4.1.5), we find the mean flow rate velocity

$$\langle V\rangle=\frac{\Delta P}{4\mu L}\frac{a^2b^2}{a^2+b^2}. \tag{E4.4.1.19}$$

The maximum velocity is attained at the tube axis

$$U_{\max}=\frac{a^2b^2\Delta P}{2\mu L(a^2+b^2)}\qquad (\text{at } X=Y=0). \tag{E4.4.1.20}$$

In the special case $a=b$, formulas (E4.4.1.17)–(E4.4.1.20) are reduced to the corresponding formulas (E4.4.1.10)–(E4.4.1.13) for a circular tube.

► **Tube of rectangular cross-section.** Now consider a tube of rectangular cross-section with sides a and b. We assume that the flow region is described by the inequalities $0\le X\le a$ and $0\le Y\le b$. The solution of Eq. (E4.4.1.1) under the no-slip conditions on the tube surface has the form

$$\begin{aligned}&V=\frac{\Delta P}{2\mu L}X(a-X)+\sum_{m=1}^{\infty}\sin\left(\frac{\pi mX}{a}\right)\left(A_m\cosh\frac{\pi mY}{a}+B_m\sinh\frac{\pi mY}{a}\right),\\&A_m=\frac{a^2\Delta P}{\pi^3m^3\mu L}[\cos(\pi m)-1],\quad B_m=-A_m\frac{\cosh(\pi mk)-1}{\sinh(\pi mk)},\quad k=\frac{b}{a}.\end{aligned} \tag{E4.4.1.21}$$

By integrating the expression for V, we obtain the volume flow rate

$$\begin{aligned}Q=&\frac{\Delta P}{24\mu L}ab(a^2+b^2)\\&-\frac{8\Delta P}{\pi^5\mu L}\sum_{m=1}^{\infty}\frac{1}{(2m-1)^5}\left[a^4\tanh\left(\pi b\frac{2m-1}{2a}\right)+b^4\tanh\left(\pi a\frac{2m-1}{2b}\right)\right].\end{aligned} \tag{E4.4.1.22}$$

For a tube of square section with side a, this formula acquires the form

$$Q=\frac{a^4\Delta P}{12\mu L}\left[1-\frac{192}{\pi^5}\sum_{m=1}^{\infty}\frac{1}{(2m-1)^5}\tanh\left(\pi\frac{2m-1}{2}\right)\right], \tag{E4.4.1.23}$$

or, after summing the series,

$$Q = 0.0351\frac{a^4\Delta P}{\mu L}. \tag{E4.4.1.23a}$$

It is useful to rewrite the last expression as

$$\frac{Q}{Q_0} = 0.883,$$

where Q_0 is the volume flow rate for a circular tube with the same cross-section area. The volume flow rate for a square tube is smaller, because the cross-section has corners near which the velocity of a viscous fluid decreases noticeably.

► **Tube of triangular cross-section.** Now suppose that the cross-section of the tube is an equilateral triangle with side b. We place the origin at the center of the cross-section and measure the coordinate X along one of the sides of the triangle. In this case, the solution of Eq. (E4.4.1.1) under the boundary condition (E4.4.1.2) has the form

$$V = \frac{\sqrt{3}\,\Delta P}{6\mu bL}\left(Y - \frac{b}{2\sqrt{3}}\right)\left(Y + \sqrt{3}\,X - \frac{b}{\sqrt{3}}\right)\left(Y - \sqrt{3}\,X - \frac{b}{\sqrt{3}}\right).$$

The volume flow rate of this flow is given by the formula

$$Q = \frac{\sqrt{3}}{320}\frac{b^4\Delta P}{\mu L}.$$

It is useful to compare Q with the volume flow rate for a circular tube of the same cross-section area:

$$\frac{Q}{Q_0} = 0.726.$$

This expression shows that the volume flow rate for a tube whose cross-section is an equilateral triangle is substantially lower than for tubes of square or circular cross-section of the same area.

► **Rectilinear tube of arbitrary cross-section.** By using dimensional considerations, one can obtain the following formulas for the volume flow rate and the maximum velocity in a rectilinear tube of constant cross-section of arbitrary shape:

$$Q = \mathcal{K}\frac{S_*^2\Delta P}{\mu L}, \qquad U_{\max} = \mathcal{K}_v\frac{S_*\Delta P}{\mu L}, \tag{E4.4.1.24}$$

where S_* is the tube cross-section area and $\mathcal{K}$ and $\mathcal{K}_v$ are dimensionless coefficients depending on the shape of the cross-section. The coefficients $\mathcal{K}$ and $\mathcal{K}_v$ can be obtained either experimentally or theoretically.

The most important dimensionless geometric parameter characterizing the cross-section shape is the ratio $\sqrt{S_*}/\mathcal{P}$, where $\mathcal{P}$ is the cross-section perimeter. In calculations, it is convenient to use the shape parameter

$$\xi = 2\sqrt{\pi}\frac{\sqrt{S_*}}{\mathcal{P}}, \tag{E4.4.1.25}$$

which always lies in the range $0 \le \xi \le 1$. The value $\xi = 1$ corresponds to a circular tube. This condition is ensured by choosing the proportionality factor $2\sqrt{\pi}$ in formula (E4.4.1.25).

For tubes of convex cross-section that is nearly circular or at least does not deviate very much from the circular shape, it is natural to assume that the coefficients $\mathcal{K}$ and $\mathcal{K}_v$ in (E4.4.1.24) depend only on ξ:

$$\mathcal{K}=\mathcal{K}(\xi),\qquad \mathcal{K}_v=\mathcal{K}_v(\xi). \tag{E4.4.1.26}$$

Then in many cases the function $\mathcal{K}=\mathcal{K}(\xi)$ (which determines the volume flow rate Q in (E4.4.1.24)) is well approximated by the linear function

$$\mathcal{K}=\mathcal{K}_0\xi,\qquad \mathcal{K}_0=\frac{1}{8\pi}\approx 0.0398. \tag{E4.4.1.27}$$

For $\xi=1$, the approximate formula (E4.4.1.27) gives the exact value of $\mathcal{K}_0$ corresponding to a circular tube. For example, formula (E4.4.1.27) can be used for tubes whose cross-section is a regular N-gon ($N=4,5,\dots$). In particular, for a tube of square cross-section, Eq. (E4.4.1.25) gives $\xi=\frac{1}{2}\sqrt{\pi}\approx 0.886$. Substituting this into (E4.4.1.27), we obtain $\mathcal{K}=0.0353$. This differs from the exact value $\mathcal{K}=0.0351$ only by 0.6% (see (E4.4.1.23a)). For a tube of elliptic cross-section with axial ratio $a/b=1.5$, the error in (E4.4.1.27) is about 5%.

► **Hydrodynamic drag coefficient.** Consider the hydrodynamic drag for the laminar flow of a fluid in tubes of various shape. The drag coefficient λ relating the pressure drop and the characteristic pressure head is introduced by the relation

$$\lambda=\frac{d_{\rm e}}{\frac{1}{2}\rho\langle V\rangle^2}\frac{\Delta P}{L}, \tag{E4.4.1.28}$$

where $d_{\rm e}$ is equivalent (or "hydraulic") diameter and $\langle V\rangle$ is the mean flow rate velocity.

Let us introduce the equivalent (or "hydraulic") diameter $d_{\rm e}$ by the formula

$$d_{\rm e}=\frac{4S_*}{\mathcal{P}},$$

where S_* is the area of the tube cross-section and $\mathcal{P}$ is the cross-section perimeter. For tubes of circular cross-section, $d_{\rm e}$ coincides with the diameter, and for a plane channel, $d_{\rm e}$ is twice the height of the channel.

Table E4.2 presents the values of the drag coefficients for tubes with various shapes of the cross-section. The Reynolds number $\mathsf{Re}_d=d_{\rm e}\langle V\rangle/\nu$ can be calculated from the mean flow rate velocity and the equivalent diameter.

For a tube whose cross-section is a regular N-gon, the value $\lambda\,\mathsf{Re}_d$ is given by the approximate formula

$$\lambda\,\mathsf{Re}_d=\frac{64\,N-82}{N-0.95}. \tag{E4.4.1.29}$$

The comparison with Table E4.2 shows that the maximum error in (E4.4.1.29) is less than 0.6% at $N=3,\ 4,\ 6,\ \infty$.

E4.4.2. Turbulent Flows in Tubes

► **Tangential stress. Turbulent viscosity.** A flow of a fluid through a smooth tube of circular cross-section remains laminar while $\mathsf{Re}=a\langle V\rangle/\nu<1500$, where a is the tube radius and $\langle V\rangle$ the mean flow rate velocity of the fluid. For higher Reynolds numbers, the loss of stability of the laminar flow is observed and an intermediate regime occurs. For

TABLE E4.2
Values of the drag coefficients for laminar flow in tubes of various shape.

Tube profile		$\lambda\,\mathsf{Re}_d$	Equivalent diameter d_e
Circular tube of diameter d		64.000	d
Flat tube of width $2h$		96.000	$4h$
Elliptic tube with semiaxes a and b	$b/a =$ 1.00	64.000	$\dfrac{\pi b}{E\big(\sqrt{1-b^2/a^2}\,\big)}$, where $E(\vartheta)$ is the complete elliptic integral of the second kind
	0.80	64.392	
	0.50	67.292	
	0.25	72.960	
	0.125	76.584	
	0.0625	78.144	
	0	78.956	
Tube of rectangular cross-section with sides a and b	$b/a =$ 1.00	58.008	$\dfrac{2ab}{a+b}$
	0.714	58.260	
	0.50	62.192	
	0.25	72.932	
	0.125	82.336	
	0.05	89.908	
	0	96.000	
Equilateral triangle with side a		53.348	$\dfrac{a\sqrt{3}}{3}$
Regular hexagon with side a		60.216	$a\sqrt{3}$
Semi-circle of diameter d		63.068	$\dfrac{\pi d}{\pi+2}$

$\mathsf{Re} > 2500$, a fully developed regime of turbulent flow is established, which is characterized by a chaotic variation of the velocity and pressure in time and space.

In a turbulent flow in a tube, there are two significantly different flow regions. In the first, entry region, the average velocity profile $\overline{V}$ changes dramatically with the distance from the entry cross-section. In the second, stabilized flow region, the average velocity profile is the same at each cross-section. The length of the entry (stabilization) region depends on the Reynolds number and the roughness of the walls and occupies a few dozen diameters (from 25 to 100). For rough estimates, this length is frequently taken to be 50 tube diameters.

In the stabilized flow region, the average fluid velocity $\overline{V}$ is directed along the tube axis and depends only on the distance Y from the tube wall. The integration of the Reynolds equations (E4.2.3.2) yields the following expression for the shear stress:

$$\tau = \tau_{\rm s}(1 - Y/a), \qquad \text{(E4.4.2.1)}$$

where $\tau_{\rm s}$ is the friction stress at the wall. Near the tube axis, as $Y/a \to 1$, it follows from (E4.4.2.1) that $\tau/\tau_{\rm s} \to 0$. Near the wall, as $Y/a \to 0$, we have $\tau/\tau_{\rm s} \to 1$.

The friction stress at the wall for a circular tube is calculated by the formula $\tau_{\rm s} = \frac{1}{2}a(\Delta P/L)$, where $\Delta P > 0$ is the total pressure drop along a tube part of length L.

In accordance with Boussinesq's model (E4.2.3.4), the shear stress can be represented as

$$\tau = \rho(\nu + \nu_{\rm t})\frac{\partial\overline{V}}{\partial Y}, \qquad \text{(E4.4.2.2)}$$

where ν is the kinematic viscosity and $\nu_{\rm t}$ is the turbulent viscosity.

Formulas (E4.4.2.1) and (E4.4.2.2) have formed the basis of most theoretical studies of the average fluid velocity and the drag coefficient in the stabilized region of turbulent flow in a circular tube (and a plane channel of width $2a$). The corresponding results obtained on the basis of Prandtl's relation (E4.2.3.4) and von Kármán's relation for turbulent viscosity can be found in the literature given at the end of this chapter. In what follows, major attention will be paid to empirical and semi-empirical formulas that approximate numerous experimental data quite well.

► **Structure of the flow. Velocity profile in a circular tube.** Experiments show that in the stabilized flow region, two characteristic subregions can be singled out, namely,

$$\begin{aligned} &0 \le Y/a \le \sigma \quad \text{(wall region)},\\ &\sigma \le Y/a \le 1 \quad \text{(core of turbulent flow)}, \end{aligned}$$

where Y is the transverse coordinate measured from the wall and $\sigma = 0.1$ to 0.2.

To describe the turbulent flow in the wall region, one introduces the so-called friction velocity U_* and the dimensionless internal coordinate y^+ according to the formulas

$$U_* = \sqrt{\tau_\text{s}/\rho}, \quad y^+ = YU_*/\nu, \tag{E4.4.2.3}$$

where τ_s is the shear stress at the wall and ρ is the fluid density. According to von Kármán, it is convenient to single out three subdomains in the wall region:

$$\text{wall region} = \begin{cases} \text{viscous sublayer } (\nu_\text{t} \ll \nu),\\ \text{buffer layer } (\nu_\text{t} \sim \nu),\\ \text{logarithmic layer } (\nu_\text{t} \gg \nu). \end{cases}$$

In the viscous sublayer, the turbulent viscosity tends to zero near the wall (ν_t is proportional to Y^3). In the logarithmic layer, the turbulent viscosity depends on the transverse coordinate linearly, $\nu_\text{t} = 0.4U_*Y$.

The approximate ranges of the above subdomains in terms of y^+ are as follows:

$$\begin{aligned} &\text{viscous sublayer} && 0 \le y^+ \le 5,\\ &\text{buffer layer} && 5 \le y^+ \le 30,\\ &\text{logarithmic layer} && 30 \le y^+. \end{aligned}$$

In the viscous sublayer, the average velocity distribution is linear,

$$\frac{\overline{V}(Y)}{U_*} = y^+ \qquad \text{for} \qquad 0 \le y^+ \le 5. \tag{E4.4.2.4}$$

Note that turbulent fluctuations can penetrate into the viscous sublayer, although the turbulent friction is small there.

The average velocity profile in the thickest, logarithmic layer can be described by the formula

$$\frac{\overline{V}(Y)}{U_*} = 2.5 \ln y^+ + 5 \qquad \text{for} \qquad 30 \le y^+ \tag{E4.4.2.5}$$

quite well. This relation is referred to as Prandtl's law of the wall. It is worth noting that the factor 2.5 occurring in Eq. (E4.4.2.5) comes from $1/\kappa$, where $\kappa = 0.4$ is the von Kármán constant. The value 5 of the constant term was obtained on the basis of numerous experimental data.

It should be pointed out that formulas (E4.4.2.4) and (E4.4.2.5) are not only supported by experimental data quite well but also have certain theoretical justification.

For the velocity distribution in the buffer layer, one can use the simple interpolation formula

$$\frac{\overline{V}(Y)}{U_*} = y^+ \cos^{3.24}\xi + (2.5\ln y^+ + 5)\sin^{2.24}\xi, \quad \xi = \frac{\pi}{2}\frac{y^+}{18 + y^+}. \tag{E4.4.2.6}$$

This formula well agrees with the existing experimental data and defines a continuous and smooth profile of the average velocity over the entire wall domain (including the viscous sublayer, the buffer layer, and the logarithmic layer).

Numerous experimental data provide evidence that the average velocity $\overline{V}$ in the core of turbulent flow can be approximated as

$$\frac{U_{\max} - \overline{V}(Y)}{U_*} = 2.5\ln\frac{a}{Y}, \tag{E4.4.2.7}$$

where $U_{\max}$ is the velocity at the flow axis. Equation (E4.4.2.7) is universal and is referred to as von Kármán's velocity defect law;* it is applicable for smooth and rough tubes and any Reynolds numbers corresponding to turbulent flows. Formula (E4.4.2.7) can be somewhat refined by adding the term $0.6\,(1 - Y/a)^2$ to the right-hand side, thus extending the formula to the range $0.01 \le Y/a < 1$.

Darcy's formula

$$\frac{U_{\max} - \overline{V}(Y)}{U_*} = 5.08\left(1 - \frac{Y}{a}\right)^{3/2} \tag{E4.4.2.8}$$

is worth mentioning. It provides a more accurate prediction of the turbulent flow near the flow axis ($Y/a \approx 1$) compared with Eq. (E4.4.2.7) but has a narrower scope, $0.25 \le Y/a \le 1$.

In engineering calculations, it is not uncommon to approximate the average fluid velocity in a turbulent flow by the power law

$$\frac{\overline{V}(Y)}{U_{\max}} = \left(\frac{Y}{a}\right)^{1/n}, \tag{E4.4.2.9}$$

where the parameter n slowly increases with the Reynolds number. The value $n = 7$ suggested by Blasius is most frequently used. In this case, formula (E4.4.2.9) agrees well with experimental data within the range $3 \times 10^3 \le \mathsf{Re}_d \le 10^5$, where $\mathsf{Re}_d = d\langle V\rangle/\nu$ is the Reynolds number for a tube of diameter $d = 2a$.

The average velocity profile for the entire cross-section of a circular tube can be calculated using the unified interpolation formula

$$\frac{\overline{V}(Y)}{U_*} = 2.5\ln(1 + 4\zeta) + 7.5\bigl(1 - e^{-\zeta} - \zeta e^{-3\zeta}\bigr) + 2.5\ln\frac{1.5\,(2 - \eta)}{1 + 2(1 - \eta)^2}, \tag{E4.4.2.10}$$

where $\zeta = 0.1\,YU_*/\nu$ and $\eta = Y/a$. This formula is a bit simpler than Reichardt's formula. The former is obtained from the latter by a slight change in numerical coefficients, which provides a better agreement with the experimental data in the logarithmic layer, with the same accuracy in the other domains.

---

* It is remarkable that Eq. (E4.4.2.7) was first suggested for the logarithmic layer, but it turned out that it can well be extrapolated to almost the entire domain of turbulent core.

► **Drag coefficient of a circular tube.** The drag coefficient λ is expressed via other hydrodynamic parameters as follows:

$$\lambda = \frac{\Delta P}{L}\frac{4a}{\rho\langle V\rangle^2} = \frac{8\tau_\mathrm{s}}{\rho\langle V\rangle^2} = 8\left(\frac{U_*}{\langle V\rangle}\right)^2. \tag{E4.4.2.11}$$

In the region of stabilized turbulent flow in a smooth circular tube, the drag coefficient can be estimated by the Prandtl–Nikuradze implicit formula

$$\frac{1}{\sqrt{\lambda}} = 0.88\ln\bigl(\mathsf{Re}_d\sqrt{\lambda}\bigr) - 0.82, \tag{E4.4.2.12}$$

where Re_d is the Reynolds number determined by the diameter. Within the range $3\times 10^3 \le \mathsf{Re}_d \le 3\times 10^6$, the maximum deviation of the result predicted by Eq. (E4.4.2.12) from experimental data is about 2%. In practice, it is more convenient to use simpler explicit formulas of Blasius and Nikuradze:

$$\lambda = \begin{cases} 0.3164\mathsf{Re}_d^{-0.25} & \text{for } 3\times 10^3 \le \mathsf{Re}_d \le 10^5, \\ 0.0032 + 0.221\,\mathsf{Re}_d^{-0.237} & \text{for } 10^5 \le \mathsf{Re}_d. \end{cases} \tag{E4.4.2.13}$$

These formulas are also accurate within 2%. The first line in Eq. (E4.4.2.13) follows from the assumption that the average velocity profile is given by the power law (E4.4.2.9) with $n = 7$.

In the transient zone, one can calculate the drag coefficient by the formula

$$\lambda = 6.3\times 10^{-4}\sqrt{\mathsf{Re}_d} \qquad (2200 \le \mathsf{Re}_d \le 4000).$$

► **Turbulent flow in a plane channel.** Qualitatively, the picture of stabilized turbulent flow in a plane channel is similar to that in a circular tube. Indeed, in the viscous sublayer adjacent to the channel walls, the velocity distribution increases linearly with the distance from the wall: $\overline{V}(Y)/U_* = y^+$. In the logarithmic layer, the average velocity profile can be described by the expression

$$\frac{\overline{V}(Y)}{U_*} = 2.5\ln y^+ + 5.2.$$

In the flow core, the average velocity distribution in a plane channel of width $2h$ can be approximately described by formulas of the form

$$\frac{U_\mathrm{max} - \overline{V}(Y)}{U_*} = A\left(1 - \frac{Y}{h}\right)^m,$$

where $A = 6.5$ and $m = 1.9$.

In the region of stabilized turbulent flow, the drag coefficient can be determined from the implicit relation

$$\frac{1}{\sqrt{\lambda}} = 0.86\ln\bigl(\mathsf{Re}\sqrt{\lambda}\bigr) - 0.35,$$

where $\lambda = 8\bigl(U_*/\langle\overline{V}\rangle\bigr)^2$ and $\mathsf{Re} = 2h\langle\overline{V}\rangle/\nu$. This relation is in good agreement with experimental data for all $\mathsf{Re} \le 10^4$.

► **Drag coefficient for tubes of other shape.** The drag coefficient λ for turbulent flows in rectilinear tubes of noncircular cross-section can also be computed using relations (E4.4.2.12) and (E4.4.2.13), where the equivalent diameter

$$d_\mathrm{e} = \frac{4S_*}{\mathcal{P}} \qquad \text{(E4.4.2.14)}$$

should be regarded as the characteristic length used to calculate the Reynolds number. In Eq. (E4.4.2.14), S_* is the cross-section area of the tube and $\mathcal{P}$ the cross-section perimeter. The values of the drag coefficient predicted by this approach fairly well agree with experimental data for tubes of rectangular and triangular cross-section.

More detailed information about the structure of turbulent flows in a circular (or noncircular) tube and a plane channel, as well as various relations for determining the average velocity profile and the drag coefficient, can be found in the literature given at the end of this chapter.

E4.5. Hydrodynamic Boundary Layers on a Flat Plate

E4.5.1. Laminar Boundary Layer

► **Preliminary remarks. Flow for large Reynolds numbers.** In practice, one often deals with outside flows around stationary extended equipment elements such as plates, guiding elements, or tubes. In this case, the action of external mass forces can often be neglected, and the hydrodynamic laws are determined by the relation between pressure, viscous, and inertial forces. Then the system of dimensionless steady-state hydrodynamic equations becomes

$$\begin{aligned} \nabla\cdot\mathbf{v} &= 0,\\ (\mathbf{v}\cdot\nabla)\mathbf{v} &= -\nabla p + \frac{1}{\mathsf{Re}}\,\Delta\mathbf{v}. \end{aligned} \qquad \text{(E4.5.1.1)}$$

The system contains a single parameter, the Reynolds number. The solution in the general case is very complicated (the system is nonlinear), but it can be simplified if we consider the passage to the limit as $\mathsf{Re}\to 0$ or $\mathsf{Re}\to\infty$. In this section, we solve the problem on a longitudinal flat-plate flow assuming that $\mathsf{Re}\to\infty$, that is, for a liquid with "vanishing viscosity." The solution is not straightforward: one cannot just disregard the term $\mathsf{Re}^{-1}\Delta\mathbf{v}$, thus obtaining the equation of an ideal fluid. The mathematical difficulty is that the small parameter Re^{-1} occurs in the term with higher derivatives. By neglecting this term, we change the order and the type of the equation, and the solution of the original system need not converge as $\mathsf{Re}^{-1}\to 0$ to the solution of the limit system with $\mathsf{Re}^{-1} = 0$. Here we deal with a *singular perturbation*. Moreover, it is clear from physical considerations that an ideal fluid flow past a body cannot satisfy the no-slip condition on the surface. Actually, the tangential velocity varies from zero on the surface of the body to the velocity of the undisturbed flow remote from the body.

For fluids with low viscosity, this change of velocity occurs in a thin fluid layer adjacent to the surface of the body. Prandtl termed this layer a *boundary layer*. The magnitude of $\Delta\mathbf{v}$ is very large in this layer. Thus, although Re^{-1} is a small parameter, one cannot disregard the term $\mathsf{Re}^{-1}\Delta\mathbf{v}$ in the boundary layer. Nevertheless, since the longitudinal and the transverse coordinates play different roles in the boundary layer, we can simplify the equations. The corresponding formal estimate of terms in the second equation in (E4.5.1.1) can be found in the books listed at the end of this chapter.

► **Statement of the Blasius problem.** Consider the steady-state problem on the longitudinal zero-pressure-gradient flow ($\nabla P \equiv 0$) past a half-infinite flat plate ($0 \le X < \infty$). We assume that the coordinates X and Y are directed along the plate and transverse to the plate, respectively, and the origin is placed at the front edge of the plate. The velocity of the incoming flow is U_i.

Let us write out the final system of boundary layer equations for an incompressible fluid

$$\begin{gathered} V_X\frac{\partial V_X}{\partial X}+V_Y\frac{\partial V_X}{\partial Y}=\nu\frac{\partial^2 V_X}{\partial Y^2},\\ \frac{\partial V_X}{\partial X}+\frac{\partial V_Y}{\partial Y}=0. \end{gathered}\tag{E4.5.1.2}$$

The natural boundary conditions have the form

$$\begin{aligned} &Y=0, && V_X=V_Y=0 && \text{(no-slip condition)},\\ &Y\to\infty, && V_X\to U_\mathrm{i} && \text{(condition far from the plate)}. \end{aligned}\tag{E4.5.1.3}$$

► **Blasius solution. Friction coefficient.** We express the fluid velocity components via the stream function Ψ by the formulas

$$V_X=\frac{\partial\Psi}{\partial Y},\qquad V_Y=-\frac{\partial\Psi}{\partial X}\tag{E4.5.1.4}$$

and substitute them into (E4.5.1.2).

Then we seek the stream function in the form

$$\Psi(X,Y)=\sqrt{\nu XU_\mathrm{i}}\,f(\eta),\qquad \eta=Y\sqrt{\frac{U_\mathrm{i}}{\nu X}},\tag{E4.5.1.5}$$

where η is a self-similar variable.

We obtain the following boundary value problem for the function $f(\eta)$:

$$\begin{aligned} &2f'''+ff''=0;\\ &\eta=0, && f=f'=0;\\ &\eta\to\infty, && f'\to 1. \end{aligned}\tag{E4.5.1.6}$$

Detailed tables containing the numerical solution of this problem can be found, for example, in Schlichting's book (1981).

By using (E4.5.1.4), we can calculate the fluid velocity components as follows:

$$V_X=U_\mathrm{i}f'(\eta),\qquad V_Y=\frac{1}{2}\sqrt{\frac{\nu U_\mathrm{i}}{X}}\big[\eta f'(\eta)-f(\eta)\big].\tag{E4.5.1.7}$$

The obtained solution also allows us to calculate some variables that are of practical interest. For example, the local frictional stress on the wall is

$$\tau_w(X)=\mu\left(\frac{\partial V_X}{\partial Y}\right)_{Y=0}=\mu U_\mathrm{i}\sqrt{\frac{U_\mathrm{i}}{\nu X}}f''(0)=0.332\,\mu U_\mathrm{i}\sqrt{\frac{U_\mathrm{i}}{\nu X}},\tag{E4.5.1.8}$$

and the local friction coefficient is given by

$$c_\mathrm{f}(X)=\frac{\tau_w(X)}{\frac{1}{2}\rho U_\mathrm{i}^2}=0.664\sqrt{\frac{\nu}{U_\mathrm{i}X}}.\tag{E4.5.1.9}$$

The total friction coefficient for a plate of length L is given by the formula

$$\langle c_{\rm f}\rangle = \frac{1}{L}\int_0^L c_{\rm f}(X)\,dX = \frac{1.328}{\sqrt{\mathsf{Re}_L}}, \tag{E4.5.1.10}$$

where $\mathsf{Re}_L = U_{\rm i}L/\nu$ is the Reynolds number for the plate in the flow.

Formula (E4.5.1.9) is known as the Blasius law for the drag in longitudinal flat-plate flow. This formula can be used in laminar flow, that is, for $\mathsf{Re}_L < 3.5\times 10^5$.

Although the boundary layer in this statement of the problem is asymptotic—that is, extends infinitely along the Y-coordinate—one can approximately estimate its thickness if we adopt the convention that the velocity on the boundary of the layer differs from the undisturbed flow velocity at most by 1%. Then the boundary layer thickness is

$$\delta(X) \approx 5\sqrt{\nu X/U_{\rm i}}.$$

The Blasius solution shows that the longitudinal velocity profiles are affinely similar to each other for all cross-sections of the boundary layer.

► **Reversed statement of the Blasius problem.** In applications, one often deals with the "reversed" statement of the Blasius problem, in which a half-infinite plate moves in its plane at a velocity $U_{\rm i}$. In this case, the stream function is also sought in the form (E4.5.1.5), but the boundary value problem (E4.5.1.6) for the function $f = f(\eta)$ is replaced by

$$\begin{aligned}&2f''' + ff'' = 0;\\ &\eta = 0, \qquad f = 0, \quad f' = 1;\\ &\eta \to \infty, \qquad f' \to 0,\end{aligned}$$

In this case, in contrast with (E4.5.1.8), the local friction stress on the wall is given by

$$\tau_w(X) = 0.444\,\mu U_{\rm i}\sqrt{\frac{U_{\rm i}}{\nu X}}.$$

E4.5.2. Turbulent Boundary Layer

► **Structure of the flow. Velocity profile.** The flow in the boundary layer on a flat plate is laminar until $\mathsf{Re}_X = U_{\rm i}X/\nu \approx 3\times 10^5$. On a longer plate, the boundary layer becomes turbulent; that is, its thickness increases sharply and the longitudinal velocity profile alters.

In accordance with Boussinesq's model the turbulence boundary layer on a flat plate is described by the equations

$$\begin{gathered}\overline{V}_X\frac{\partial \overline{V}_X}{\partial X} + \overline{V}_Y\frac{\partial \overline{V}_X}{\partial Y} = \frac{\partial}{\partial Y}\left[(\nu+\nu_t)\frac{\partial \overline{V}_X}{\partial Y}\right],\\ \frac{\partial \overline{V}_X}{\partial X} + \frac{\partial \overline{V}_Y}{\partial Y} = 0,\end{gathered}$$

where $\nu_{\rm t}$ is the turbulent viscosity.

These equations have formed the basis of most theoretical investigations on the determination of the average fluid velocity and the drag coefficient in the stabilized region of turbulent flow on a flat plate. In what follows, major attention will be paid to empirical and semiempirical formulas that approximate numerous experimental data quite well.

Experiments show that the turbulent boundary layer on a flat plate includes two qualitatively different regions, namely, the wall region (adjacent to the plate surface) and the outer region (bordering the unperturbed stream). By analogy with the flow through a circular tube, it is common to subdivide the thin wall region into three subdomains (von Kármán's scheme):

$$\boxed{\text{wall region}} = \boxed{\text{viscous sublayer}} + \boxed{\text{buffer layer}} + \boxed{\text{logarithmic layer}}$$

The friction velocity U_* and the dimensionless internal coordinate y^+ defined by relations (E4.4.2.3) are introduced to describe the turbulent flow of the fluid in the wall region. The transverse coordinate Y in Eq. (E4.4.2.3) is measured from the plate surface.

The velocity profile in the viscous sublayer is linear, $\overline{V}_X(Y)/U_* = y^+$.

In the logarithmic layer and the outer region, the average velocity profile is quite well described by the relations (Monin & Yaglom, 1992)

$$\frac{U_\mathrm{i} - \overline{V}_X(Y)}{U_*} = \begin{cases} -2.5\ln(Y/\delta) + 2.2 & \text{for } Y/\delta \le 0.15, \\ 9.6\,[1-(Y/\delta)]^2 & \text{for } 0.15 \le Y/\delta \le 1, \end{cases}$$

where $\delta = \delta(X)$ is the boundary layer thickness.

Note the Spalding implicit interpolation formula

$$y^+ = 2.5\,V^+ + 0.135\left[\exp(-V^+) - 1 - V^+ - \tfrac{1}{2}(V^+)^2 - \tfrac{1}{6}(V^+)^3\right],$$

where $y^+ = YU_*/\nu$ and $V^+ = 0.4\,\overline{V}(Y)/U_*$. This formula quite well describes the average velocity profile within the entire wall region $0 \le Y/\delta \le 0.15$.

The turbulent boundary layer thickness can be calculated from the formula

$$\delta = 0.33\,X\sqrt{\tfrac{1}{2}c_\mathrm{f}}. \qquad \text{(E4.5.2.1)}$$

► **The local coefficient of friction.** In the case of a two-sided flow past a flat plate, the local coefficient of friction $c_\mathrm{f} = c_\mathrm{f}(X)$ is expressed via other hydrodynamic parameters as

$$c_\mathrm{f} = \frac{\tau_\mathrm{s}}{\frac{1}{2}\rho U_\mathrm{i}^2} = 2\left(\frac{U_*}{U_\mathrm{i}}\right)^2.$$

For the turbulent boundary layer on a flat plate, von Kármán's friction law with modified numerical coefficients

$$\frac{1}{\sqrt{c_\mathrm{f}}} = 1.77\ln\bigl(c_\mathrm{f}\,\mathsf{Re}_X\bigr) + 2.4, \qquad \text{(E4.5.2.2)}$$

is typically used, which quite well agrees with experimental data. In Eq. (E4.5.2.2), $\mathsf{Re}_X = U_\mathrm{i}X/\nu$ is the local Reynolds number. Here it is assumed that the flow turbulization in the boundary layer starts from the front edge of the plate.

Relation (E4.5.2.2) defines the friction coefficient versus the Reynolds number implicitly. For $\mathsf{Re}_X > 10^9$, it is more convenient to use the simpler explicit Falkner's formula

$$c_\mathrm{f} = 0.0262\,\mathsf{Re}_X^{-1/7}, \qquad \text{(E4.5.2.3)}$$

The corresponding mean coefficient of friction for a plate of length L is given by

$$\langle c_\mathrm{f}\rangle = \frac{1}{L}\int_0^L c_\mathrm{f}\,dX = 0.0306\,\mathsf{Re}_L^{-1/7}, \qquad \text{(E4.5.2.4)}$$

where $\mathsf{Re}_L = U_\mathrm{i}L/\nu$.

The comparison of Eq. (E4.5.2.4) with Eq. (E4.5.1.10) reveals that the resistance of a flat plate to turbulent flow is much greater than to laminar flow and decreases with increasing Reynolds number considerably slower.

Note also the Schlichting formulas

$$c_\mathrm{f} = \bigl(2\log_{10}\mathsf{Re}_X - 0.65\bigr)^{-2.3}, \qquad \langle c_\mathrm{f}\rangle = 0.455\bigl(\log_{10}\mathsf{Re}_L\bigr)^{-2.58},$$

which are accurate to within few percent for $10^5 \le \mathsf{Re}_L \le 10^9$.

More detailed information about the structure of turbulent flows on a flat plate, as well as various relations for determining the average velocity profile and the local coefficient of friction, can be found in the bibliography given at the end of this chapter.

E4.6. Spherical Particles and Circular Cylinders in Translational Flow

► **Preliminary remarks.** Solving the problem on the interaction of a solid particle, drop, or bubble with the surrounding continuous phase underlies the design and analysis of many technological processes. The industrial applications of such interaction include classification of suspensions in hydrocyclones, sedimentation of colloids, pneumatic conveyers, fluidization, heterogeneous catalysis in suspension, dissolving solid particles, extraction from drops, absorption, and evaporation into bubbles.

The description of a large variety of meteorological phenomena is also based on the analysis of motion of a collection of drops in air. The recent increase in atmosphere pollution is a serious problem, which requires understanding the transfer of mechanical, chemical, and radioactive particles in the atmosphere.

In rarefied systems of particles, drops, or bubbles, the particle–particle interaction can be neglected in the first approximation; then one deals with the behavior of a single particle moving in fluid. In this case, the streamline pattern depends on the particle shape, the flow type (translational or shear), and a number of other geometric factors.

E4.6.1. Stokes Equations and Their Solution for the Axisymmetric Case

► **Stokes equations.** One of the main approaches to the analysis and simplification of the Navier–Stokes equations is as follows. One assumes that the nonlinear inertia term $\bigl(\mathbf{V}\cdot\nabla\bigr)\mathbf{V}$ is small compared with the linear viscous term $\nu\Delta\mathbf{V}$ and hence can be neglected altogether or taken into account in some special way. This method is well-founded for $\mathsf{Re} = LU/\nu \ll 1$ and is widely used for studying the motion of particles, drops, and bubbles in fluids. Low Reynolds numbers are typical of the following three cases: slow (creeping) flows, highly viscous fluids, and small dimensions of particles.

For steady-state flows of viscous incompressible fluid, by neglecting the inertia terms in (E4.2.1.4) and by including all conservative mass forces in the pressure P, we arrive at the Stokes equations

$$\begin{aligned}\nabla\cdot\mathbf{V} &= 0,\\ \mu\Delta\mathbf{V} &= \nabla P.\end{aligned} \tag{E4.6.1.1}$$

The Stokes equations (E4.6.1.1) are linear and hence much simpler than the nonlinear Navier–Stokes equations. For any two solutions $\{\mathbf{V}_1,\ P_1\}$ and $\{\mathbf{V}_2,\ P_2\}$ of (E4.6.1.1), the sum $\{\alpha\mathbf{V}_1 + \beta\mathbf{V}_2,\ \alpha P_1 + \beta P_2\}$ satisfies the same equations for any constants α and β.

In axisymmetric problems, all the variables in the spherical coordinates R, θ, φ are independent of φ, and the third component of the fluid velocity is zero, $V_\varphi = 0$. The fluid velocity components V_R and V_θ can be expressed via the stream function Ψ as follows:

$$V_R = \frac{1}{R^2 \sin\theta} \frac{\partial \Psi}{\partial \theta}, \qquad V_\theta = -\frac{1}{R \sin\theta} \frac{\partial \Psi}{\partial R}. \tag{E4.6.1.2}$$

Then the first equation in (E4.6.1.1) (the continuity equation) is satisfied automatically. On eliminating the terms containing the pressure from the second vector equation (E4.6.1.1) and replacing the velocity components by the right-hand side of (E4.6.1.2), we obtain the following equation for the stream function:

$$E^2\bigl(E^2\Psi\bigr) = 0, \qquad E^2 \equiv \frac{\partial^2}{\partial R^2} + \frac{\sin\theta}{R^2} \frac{\partial}{\partial \theta}\Bigl(\frac{1}{\sin\theta} \frac{\partial}{\partial \theta}\Bigr). \tag{E4.6.1.3}$$

► **General solution of Stokes equations for the axisymmetric case.** The general solution of Eq. (E4.6.1.3), which gives the finite fluid velocities on the flow axis for $\theta = 0$ and $\theta = \pi$, has the form

$$\Psi(R,\theta) = \sum_{n=2}^{\infty} \bigl(A_n R^n + B_n R^{1-n} + C_n R^{n+2} + D_n R^{3-n}\bigr) \mathcal{J}_n(\cos\theta). \tag{E4.6.1.4}$$

Here A_n, B_n, C_n, and D_n are arbitrary constants, and $\mathcal{J}_n(\zeta)$ are the Gegenbauer functions of the first kind, which can be represented in the form

$$\begin{aligned} \mathcal{J}_n(\zeta) &= \frac{P_{n-2}(\zeta) - P_n(\zeta)}{2n-1} = -\frac{1}{(n-1)!} \Bigl(\frac{d}{d\zeta}\Bigr)^{n-2} \Bigl(\frac{\zeta^2 - 1}{2}\Bigr)^{n-1}, \\ P_n(\zeta) &= \frac{1}{n!\,2^n} \frac{d^n}{d\zeta^n} (\zeta^2 - 1)^n, \end{aligned}$$

where $P_n(\zeta)$ are the Legendre functions. The Gegenbauer functions of the first kind are polynomials; in particular,

$$\mathcal{J}_0(\zeta) = 1, \quad \mathcal{J}_1(\zeta) = -\zeta, \quad \mathcal{J}_2(\zeta) = \tfrac{1}{2}(1-\zeta^2), \quad \mathcal{J}_3(\zeta) = \tfrac{1}{2}\zeta(1-\zeta^2), \quad \mathcal{J}_4(\zeta) = \tfrac{1}{8}(1-\zeta^2)(5\zeta^2-1).$$

The corresponding fluid velocity components and pressure are given by the formulas

$$\begin{aligned} V_R &= -\sum_{n=2}^{\infty} \bigl(A_n R^{n-2} + B_n R^{-n-1} + C_n R^n + D_n R^{1-n}\bigr) P_{n-1}(\cos\theta), \\ V_\theta &= \sum_{n=2}^{\infty} \bigl[n A_n R^{n-2} - (n-1) B_n R^{-n-1} + (n+2) C_n R^n - (n-3) D_n R^{1-n}\bigr] \frac{\mathcal{J}_n(\cos\theta)}{\sin\theta}, \\ P &= -2\mu \sum_{n=2}^{\infty} \Bigl(\frac{2n+1}{n-1} C_n R^{n-1} + \frac{2n-3}{n} D_n R^{-n}\Bigr) P_{n-1}(\cos\theta) + \text{const}. \end{aligned}$$

The force exerted by the fluid on any spherical boundary described by the equation $R = \text{const}$ is given by

$$F = 4\pi\mu D_2.$$

It is of interest that this force is completely determined by only one coefficient of the series (E4.6.1.4).

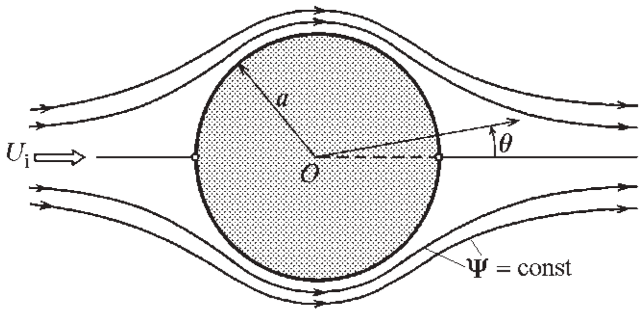

Figure E4.4. Translational Stokes flow past a spherical particle.

E4.6.2. Spherical Particles in Translational Stokes Flow (at Re ⟶ 0)

► **Flow past a spherical particle.** Consider a solid spherical particle of radius a in a translational Stokes flow with velocity U_i and dynamic viscosity μ (Figure E4.4). We assume that the fluid has dynamic viscosity μ. We use the spherical coordinate system R, θ, φ with origin at the center of the particle and with angle θ measured from the direction of the incoming flow (that is, from the rear stagnation point on the particle surface). In view of the axial symmetry, only two components of the fluid velocity, namely, V_R and V_θ, are nonzero, and all the unknowns are independent of the third coordinate φ.

The fluid velocity distribution is given by the Stokes equations (E4.6.1.1) with the no-slip boundary conditions on the surface of the solid sphere,

$$V_R = V_\theta = 0 \qquad \text{at} \quad R = a, \tag{E4.6.2.1}$$

and the boundary conditions at infinity of the form

$$V_R \to U_\mathrm{i}\cos\theta, \quad V_\theta \to -U_\mathrm{i}\sin\theta \qquad \text{as} \quad R \to \infty, \tag{E4.6.2.2}$$

which correspond to condition (E4.2.2.1) of translational unperturbed flow far from the particle.

By passing from the fluid velocity components V_R, V_θ to the stream function Ψ according to formulas (E4.6.1.2), we arrive at Eq. (E4.6.1.3). It follows from the remote boundary conditions (E4.6.2.2) that in the general solution (E4.6.1.4) it suffices to retain only the first term (corresponding to the case $n = 2$). The no-slip conditions (E4.6.2.1) allow us to find the unknown constants A_2, B_2, C_2, and D_2. The resulting expression for the stream function,

$$\Psi = \frac{1}{2}U_\mathrm{i}R^2\left(1 - \frac{3}{2}\frac{a}{R} + \frac{1}{2}\frac{a^3}{R^3}\right)\sin^2\theta, \tag{E4.6.2.3}$$

allows us to find the fluid velocity and pressure in the form

$$\begin{aligned} V_R &= U_\mathrm{i}\left(1 - \frac{3}{2}\frac{a}{R} + \frac{1}{2}\frac{a^3}{R^3}\right)\cos\theta, \\ V_\theta &= -U_\mathrm{i}\left(1 - \frac{3}{4}\frac{a}{R} - \frac{1}{4}\frac{a^3}{R^3}\right)\sin\theta, \\ P &= P_\mathrm{i} - \frac{3\mu U_\mathrm{i} a\cos\theta}{2R^2}, \end{aligned} \tag{E4.6.2.4}$$

where P_i is the unperturbed pressure far from the particle.

The viscous drag acting on the spherical particle is given by the Stokes formula

$$F = 6\pi\mu a U_{\rm i}. \tag{E4.6.2.5}$$

► **Steady-state motion of spherical particles in a fluid.** In chemical technology, one often meets the problem of a steady-state motion of a spherical particle with velocity $U_{\rm i}$ in a quiescent fluid. Since the Stokes equations are linear, the solution of this problem can be obtained from formulas (E4.6.2.4) by adding the terms $\bar{V}_R = -U_{\rm i}\cos\theta$ and $\bar{V}_\theta = U_{\rm i}\sin\theta$, which describe a translational flow with velocity $U_{\rm i}$ in the direction opposite to the incoming flow.

The drag coefficient for a solid spherical particle is calculated by the formula (the Stokes law)

$$c_{\rm f} = \frac{F}{\frac{1}{2}\rho U_{\rm i}^2\pi a^2} = \frac{12}{\mathsf{Re}}, \tag{E4.6.2.6}$$

where $\mathsf{Re} = \rho U_{\rm i} a/\mu$ and ρ is the density of the fluid. The Stokes law is confirmed by experiments for $\mathsf{Re} < 0.1$.

By equating the drag force F of the sphere with the difference $\frac{4}{3}\pi a^3 g\Delta\rho$ between the gravity and buoyancy forces, one can estimate the steady-state velocity of relative motion of phases (the velocity at which a spherical particle falls or rises) as

$$U = \frac{2}{9}\frac{ga^2\Delta\rho}{\mu}, \tag{E4.6.2.7}$$

where $\Delta\rho$ is the difference between the densities of the surrounding fluid and the particle and g is the free fall acceleration.

E4.6.3. Spherical Particles in Translational Flow in a Wide Range of Re

► **Flow past spherical particles in a wide range of Re.** Neglecting the inertia term for the flow past a sphere is adequate to experiments only in the limit case $\mathsf{Re} \to 0$. For $\mathsf{Re} = 0.05$, the error in estimating the drag by formula (E4.6.2.6) is 1.5 to 2%, and for $\mathsf{Re} = 0.5$ the error becomes 10.5 to 11%. Therefore, one can use the estimate $c_{\rm f} = 12/\mathsf{Re}$ for the drag coefficient only for $\mathsf{Re} < 0.2$ (in this case, the maximum error does not exceed 5%).

More exact than the Stokes law (E4.6.2.6) is the two-term Oseen's expansion of the drag coefficient as $\mathsf{Re} \to 0$:

$$c_{\rm f} = \frac{12}{\mathsf{Re}}\left(1 + \frac{3}{8}\mathsf{Re}\right). \tag{E4.6.3.1}$$

The difference between Oseen's approximation (E4.6.3.1) and experimental data is 0 to 1.0% for $\mathsf{Re} \le 0.05$ and 4 to 6% for $\mathsf{Re} = 0.5$.

For $\mathsf{Re} > 0.5$, asymptotic solutions no longer give an adequate description of translational flow of a viscous fluid past a spherical particle.

Numerous available numerical solutions for the Navier–Stokes equations, as well as experimental data, provide a detailed analysis of the flow pattern for increasing Reynolds numbers. For $0.5 < \mathsf{Re} < 10$, there is no flow separation, although the fore-and-aft symmetry typical of inertia-free Stokes flow past a sphere is more and more distorted. Finally, at $\mathsf{Re} \approx 10$, flow separation occurs at the rear of the particle.

The range $10 < \mathsf{Re} < 65$ is characterized by the existence of a closed stable area at the rear, in which there is an axisymmetric recirculating wake. As Re increases, the wake lengthens, and the separation ring θ_s moves forward from the rear point ($\theta_\mathrm{s} = 0°$ at $\mathsf{Re} = 10$) to $\theta_\mathrm{s} = 72°$ at $\mathsf{Re} = 200$ according to the law

$$\theta_\mathrm{s} = 42.5\left(\ln\frac{\mathsf{Re}}{10}\right)^{0.483} \qquad \text{for} \quad 10 < \mathsf{Re} < 200. \tag{E4.6.3.2}$$

Here, as well as in (E4.6.3.3), θ_s is measured in degrees.

At $\mathsf{Re} > 65$, the vorticity region in the rear area ceases to be stable and becomes unsteady. At $65 < \mathsf{Re} < 200$, a long oscillating wake is formed behind the particle, which gradually becomes turbulent for $200 < \mathsf{Re} < 1.5\times10^5$. Simultaneously, the separation point moves upstream according to the law

$$\theta_\mathrm{s} = 102 - 213\,\mathsf{Re}^{-0.37} \qquad \text{for} \quad 200 < \mathsf{Re} < 1.5\times10^5. \tag{E4.6.3.3}$$

► **Formulas for the drag coefficient in a wide range of Re.** We give two simple approximate formulas for the drag coefficient of a spherical particle

$$\begin{aligned} c_\mathrm{f} &= \frac{12}{\mathsf{Re}}\left(1 + 0.241\,\mathsf{Re}^{0.687}\right), && 0 \le \mathsf{Re} \le 400,\\ c_\mathrm{f} &= \frac{12}{\mathsf{Re}}\left(1 + 0.0811\,\mathsf{Re}^{0.879}\right), && 200 \le \mathsf{Re} \le 2500, \end{aligned} \tag{E4.6.3.4}$$

where the Reynolds number is determined with respect to the radius. In formulas (E4.6.3.4), the maximum error does not exceed 5% in the given ranges.

In a wide range of Reynolds numbers, one can use the following more complicated approximation to the drag coefficient:

$$c_\mathrm{f} = \frac{12}{\mathsf{Re}}\left(1 + 0.241\,\mathsf{Re}^{0.687}\right) + 0.42\left(1 + 1.902\times10^4\,\mathsf{Re}^{-1.16}\right)^{-1},$$

whose maximum error does not exceed 6% for $\mathsf{Re} < 1.5\times10^5$.

At $\mathsf{Re} \approx 1.5\times10^5$, one can observe the “drag crisis” characterized by a sharp decrease in the drag coefficient; the boundary layer becomes more and more turbulent; the separation point shifts abruptly to the aft area.

For $\mathsf{Re} \ge 1.7\times10^5$, the drag coefficient can be calculated by the formulas

$$c_\mathrm{f} = \begin{cases} 28.18 - 5.3\log_{10}\mathsf{Re} & \text{for } 1.7\times10^5 \le \mathsf{Re} \le 2\times10^5,\\ 0.1\log_{10}\mathsf{Re} - 0.46 & \text{for } 2\times10^5 < \mathsf{Re} \le 5\times10^5,\\ 0.19 - 4\times10^4\,\mathsf{Re}^{-1} & \text{for } 5\times10^5 < \mathsf{Re}. \end{cases}$$

E4.6.4. Translational Flow Past a Cylinder (the Plane Problem)

In chemical technology and power engineering, equipment containing heat exchanging pipes and various cylindrical links immersed in a moving fluid is often used. The estimation of the hydrodynamic action on these elements is based on the solution of the plane problem on the flow past a cylinder.

► **Flow past a cylinder at low Reynolds numbers.** Consider the translational flow of a viscous incompressible fluid with velocity U_i about a circular cylinder of radius a. The drag coefficient for a circular cylinder at low Reynolds numbers is calculated by the asymptotic formula

$$c_{\mathrm{f}} = \frac{F}{aU_{\mathrm{i}}^2\rho} = \frac{4\pi}{\mathsf{Re}}\bigl(\Delta - 0.87\Delta^3\bigr),$$
$$\Delta = \left(\ln\frac{3.703}{\mathsf{Re}}\right)^{-1}, \quad \mathsf{Re} = \frac{aU_{\mathrm{i}}\rho}{\mu}, \tag{E4.6.4.1}$$

where F is the force per unit length of the cylinder and ρ is the density of the fluid.

Comparison with experimental data shows that formula (E4.6.4.1) can be used for $0 < \mathsf{Re} < 0.4$.

► **Flow past a cylinder at moderate Reynolds numbers.** According to experimental data, a nonseparating flow past a circular cylinder is realized at $\mathsf{Re} \le 2.5$. At such Reynolds numbers, one can use the following approximate formula for calculating the drag coefficient:

$$c_{\mathrm{f}} = 5.65\,\mathsf{Re}^{-0.78}\bigl(1 + 0.26\,\mathsf{Re}^{0.82}\bigr) \qquad \text{for} \quad 0.05 \le \mathsf{Re} \le 2.5;$$

this formula was obtained from experimental and numerical results.

If the Reynolds number becomes larger than the critical value $\mathsf{Re} \approx 2.5$, then the vortex counterflow with closed streamlines arises near the rear point, that is, separation occurs. As the Reynolds number increases, the separation point gradually moves from the axis upward along the cylinder surface. The drag coefficient for a separated flow past a cylinder at moderate Reynolds numbers can be calculated by the empirical formulas

$$c_{\mathrm{f}} = 5.65 \times 10^{-0.78}\bigl(1 + 0.333\,\mathsf{Re}^{0.55}\bigr) \qquad \text{for} \quad 2.5 < \mathsf{Re} \le 20,$$
$$c_{\mathrm{f}} = 5.65 \times 10^{-0.78}\bigl(1 + 0.148\,\mathsf{Re}^{0.82}\bigr) \qquad \text{for} \quad 20 < \mathsf{Re} \le 200.$$

► **Flow past a cylinder at high Reynolds numbers.** With further increase of Re, the rear vortices become longer and then alternative vortex separation occurs (the Kármán vortex street is formed). Simultaneously, the separation point moves closer to the equatorial section. The frequency ν_{f} of vortex shedding from the rear area is an important characteristic of the flow past a cylinder. It can be determined from the empirical formula

$$\mathsf{St} = \frac{0.13}{c_{\mathrm{f}}}\bigl[1 - \exp(-2.38\,c_{\mathrm{f}})\bigr],$$

where $\mathsf{St} = a\nu_{\mathrm{f}}/U_{\mathrm{i}}$ is the Strouhal number.

We also present another useful formula for the vortex separation frequency: $\nu_{\mathrm{f}} = 0.08\,U_{\mathrm{i}}/b$, where b is the half-width of the wake at the point of destruction.

Starting from $\mathsf{Re} \approx 0.5 \times 10^3$, one can speak of a developed hydrodynamic boundary layer. The flow remains laminar in a considerable part of this layer. If the Reynolds number varies in the range $0.5\times10^3 < \mathsf{Re} < 0.5\times10^5$, the separation point θ_s of the laminar boundary layer gradually moves from 71.2° to 95°.

For $\mathsf{Re} > 2000$, the wake becomes totally turbulent at large distances from the body.

The curve $c_{\mathrm{f}}(\mathsf{Re})$ contains two straight-line segments (self-similarity areas), where the drag coefficient is practically constant,

$$c_{\mathrm{f}} = 1.0 \qquad \text{for} \quad 3\times10^2 < \mathsf{Re} < 3\times10^3,$$
$$c_{\mathrm{f}} = 1.1 \qquad \text{for} \quad 4\times10^3 < \mathsf{Re} < 10^5.$$

In the intermediate region between these segments, the drag coefficient monotonically increases with Reynolds number.

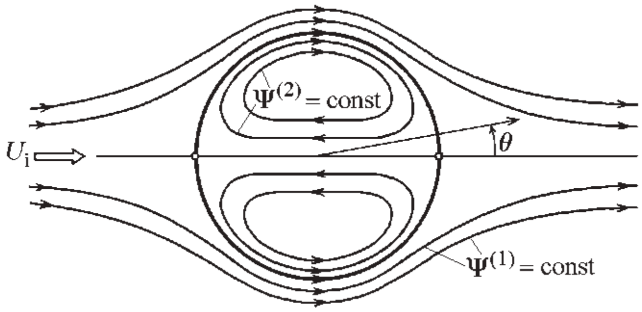

Figure E4.5. Translational Stokes flow past a spherical drop.

► **Developed turbulence in the boundary layer on a cylinder.** Developed turbulence within the boundary layer takes place at higher Reynolds numbers $\mathsf{Re} \approx 10^5$ and is accompanied by the "drag crisis." The cylinder drag first decreases sharply to $c_{\rm f} \approx 0.3$ at $\mathsf{Re} = 3.5\times 10^5$ and then begins to increase and again enters the self-similar regime characterized by the constant value

$$c_{\rm f} = 0.9 \qquad \text{for} \quad \mathsf{Re} > 5\times 10^5.$$

E4.7. Spherical Drops and Bubbles in Translational Flow

E4.7.1. Spherical Drops and Bubbles in Translational Stokes Flow (at Re → 0)

► **Flow past a spherical drop or bubble.** Now consider a spherical drop of radius a in a translational Stokes flow of another fluid with velocity $U_{\rm i}$ (Figure E4.5). We assume that the dynamic viscosities of the outer and inner fluids are equal to μ_1 and μ_2, respectively. The unknown variables outside and inside the drop are indicated by the superscripts (1) and (2), respectively.

To obtain the velocity and pressure profiles for the fluid in each phase, we shall use the Stokes equations (E4.6.1.1). As previously, the condition that the flow is uniform far from the drop has the form (E4.6.2.2).

Let us write out four conditions that must be satisfied on the boundary of a spherical drop.

There is no flow across the interface:

$$V_R^{(1)} = V_R^{(2)} = 0 \qquad \text{at} \quad R = a. \tag{E4.7.1.1}$$

The tangential velocity is continuous across the interface:

$$V_\theta^{(1)} = V_\theta^{(2)} \qquad \text{at} \quad R = a. \tag{E4.7.1.2}$$

The normal stress jump across the interface is equal to the pressure increment due to interfacial tension:

$$P^{(1)} - 2\mu_1 \frac{\partial V_R^{(1)}}{\partial R} + \frac{2\sigma}{a} = P^{(2)} - 2\mu_2 \frac{\partial V_R^{(2)}}{\partial R} \quad \text{at} \quad R = a, \tag{E4.7.1.3}$$

where σ is the interfacial tension.

The tangential stress is continuous across the interface:

$$\mu_1\left(\frac{\partial V_\theta^{(1)}}{\partial R} - \frac{V_\theta^{(1)}}{R}\right) = \mu_2\left(\frac{\partial V_\theta^{(2)}}{\partial R} - \frac{V_\theta^{(2)}}{R}\right) \qquad \text{at} \quad R = a. \tag{E4.7.1.4}$$

We also use the boundedness of the solution at the drop center:

$$V_R^{(2)} < \infty, \quad V_\theta^{(2)} < \infty \qquad \text{at} \quad R = 0. \tag{E4.7.1.5}$$

We introduce the stream function $\Psi^{(m)}$ in each of the phases ($m = 1, 2$) according to formulas (E4.6.1.2). Conditions (E4.7.1.1)–(E4.7.1.5) allow us to determine the constants in the general solutions (E4.7.1.5) inside and outside the drop. As a result, we obtain the Hadamard–Rybczynski solution

$$\begin{aligned}\Psi^{(1)} &= \frac{1}{4}U_\mathrm{i}R^2\left(2 - \frac{2+3\beta}{1+\beta}\frac{a}{R} + \frac{\beta}{1+\beta}\frac{a^3}{R^3}\right)\sin^2\theta,\\ \Psi^{(2)} &= -\frac{U_\mathrm{i}}{4(1+\beta)}R^2\left(1 - \frac{R^2}{a^2}\right)\sin^2\theta, \qquad \beta = \frac{\mu_2}{\mu_1}.\end{aligned} \tag{E4.7.1.6}$$

By using formulas (E4.6.1.2), we calculate the velocity and the pressure outside the drop:

$$\begin{aligned}V_R^{(1)} &= U_\mathrm{i}\left[1 - \frac{2+3\beta}{2(1+\beta)}\frac{a}{R} + \frac{\beta}{2(1+\beta)}\frac{a^3}{R^3}\right]\cos\theta,\\ V_\theta^{(1)} &= -U_\mathrm{i}\left[1 - \frac{2+3\beta}{4(1+\beta)}\frac{a}{R} - \frac{\beta}{4(1+\beta)}\frac{a^3}{R^3}\right]\sin\theta,\\ P^{(1)} &= P_0^{(1)} - \frac{\mu_1 U_\mathrm{i} a(2+3\beta)}{2(1+\beta)}\frac{\cos\theta}{R^2}, \qquad \beta = \frac{\mu_2}{\mu_1}.\end{aligned} \tag{E4.7.1.7}$$

The velocity and the pressure inside the drop are given by

$$\begin{aligned}V_R^{(2)} &= -\frac{U_\mathrm{i}}{2(1+\beta)}\left(1 - \frac{R^2}{a^2}\right)\cos\theta,\\ V_\theta^{(2)} &= \frac{U_\mathrm{i}}{2(1+\beta)}\left(1 - 2\frac{R^2}{a^2}\right)\sin\theta,\\ P^{(2)} &= P_0^{(2)} + \frac{5\mu_2 U_\mathrm{i} R\cos\theta}{a^2(1+\beta)}.\end{aligned} \tag{E4.7.1.8}$$

The constants $P_0^{(1)}$ and $P_0^{(2)}$ in the expressions (E4.7.1.7) and (E4.7.1.8) for the pressure fields are related by

$$P_0^{(2)} - P_0^{(1)} = \frac{2\sigma}{a}. \tag{E4.7.1.9}$$

The drag force acting on the spherical drop is

$$F = 2\pi a U_\mathrm{i}\frac{2\mu_1 + 3\mu_2}{\mu_1 + \mu_2}. \tag{E4.7.1.10}$$

As $\beta = \mu_2/\mu_1 \to \infty$, Eq. (E4.7.1.10) becomes the Stokes formula (E4.6.2.5) for a solid particle. The passage to the limit as $\beta \to 0$ corresponds to a gas bubble.

► **Steady-state motion of drops and bubbles in a fluid.** Consider the problem of a steady-state motion of a spherical drop or bubble with velocity U_i in a quiescent fluid. Since the Stokes equations are linear, the solution of this problem can be obtained from formulas (E4.7.1.7) and (E4.7.1.8) by adding the terms $\bar V_R = -U_\mathrm{i}\cos\theta$ and $\bar V_\theta = U_\mathrm{i}\sin\theta$, which describe a translational flow with velocity U_i in the direction opposite to the incoming flow.

The drag coefficient for a drop is calculated by the formula

$$c_\mathrm{f} = \frac{F}{\frac{1}{2}\rho_1 U_\mathrm{i}^2 \pi a^2} = \frac{4}{\mathsf{Re}}\left(\frac{2+3\beta}{1+\beta}\right), \qquad \text{where} \quad \mathsf{Re} = \frac{\rho_1 U_\mathrm{i} a}{\mu_1}. \tag{E4.7.1.11}$$

By equating the drag force F of the sphere with the difference $\frac{4}{3}\pi a^3 g\Delta\rho$ between the gravity and buoyancy forces, one can estimate the steady-state velocity of relative motion of phases (the velocity at which a spherical drop falls or rises) as

$$U = \frac{2}{3}\frac{ga^2\Delta\rho}{\mu_1}\left(\frac{1+\beta}{2+3\beta}\right), \tag{E4.7.1.12}$$

where $\Delta\rho$ is the difference between the densities of the outer and the inner fluid and g is the free fall acceleration.

Relations (E4.7.1.11) and (E4.7.1.12) cover the entire range $0 \le \beta \le \infty$ of the phase viscosity ratio. In the limit case $\beta \to \infty$, formulas (E4.7.1.11) and (E4.7.1.12) become formulas (E4.6.2.6) and (E4.6.2.7) for a solid spherical particle. In the other limit case $\beta = 0$ corresponding to a spherical gas bubble in a highly viscous liquid, formulas (E4.7.1.11) and (E4.7.1.12) become

$$c_\mathrm{f} = \frac{8}{\mathsf{Re}}, \quad U = \frac{1}{3}\frac{ga^2\Delta\rho}{\mu_1} \qquad \text{(gas bubble)}. \tag{E4.7.1.13}$$

These formulas hold only for extremely pure liquids without any surfactants.

It is known that even very small quantities of surfactants are adsorbed on the bubble surface and lead to its “solidification.” This results in eliminating internal circulation, and hence the bubble rises according to the Stokes law (E4.7.1.11) for solid particles.

E4.7.2. Spherical Drops and Bubbles in Translational Flow at Various Reynolds Numbers

► **Bubble in a translational flow.** For determining the drag coefficient of a spherical bubble in a translational flow in the entire range of Reynolds numbers, one can use the interpolation formula

$$c_\mathrm{f} = \frac{8}{\mathsf{Re}} + \frac{16}{\mathsf{Re}+16}, \quad \mathsf{Re} = \frac{aU_\mathrm{i}}{\nu}, \tag{E4.7.2.1}$$

where a is the bubble radius and ν is the kinematic viscosity of the surrounding fluid. In both limit cases $\mathsf{Re} \to 0$ and $\mathsf{Re} \to \infty$, formula (E4.7.2.1) gives correct asymptotic results; its maximum error for intermediate Reynolds numbers is less than 4.5%. The drag coefficient of a spherical bubble monotonically decreases as the Reynolds number increases.

► **Drop in a translational liquid flow.** For the drag coefficient of a spherical drop in a translational flow at small Reynolds numbers, we have the two-term expansion

$$c_\mathrm{f} = \frac{3\beta+2}{\beta+1}\left(\frac{4}{\mathsf{Re}} + \frac{\mathsf{Re}}{2}\right), \quad \mathsf{Re} = \frac{aU_\mathrm{i}}{\nu}, \tag{E4.7.2.2}$$

where ν is the kinematic viscosity of the fluid surrounding the drop.

The spherical form of a drop or a bubble in a Stokes flow follows from the fact that the flow is inertia-free. However, even for the case in which the inertia forces dominate viscous forces and the Reynolds number cannot be considered small, the drop remains undeformed if the inertia forces are small compared with the capillary forces. The ratio of inertial to capillary forces is measured by the Weber number $\mathrm{We} = \rho_1 U_\mathrm{i}^2 a/\sigma$, where σ is the surface tension at the drop boundary. For small We, a deformable drop will preserve the spherical form.

At the end of Subsection E4.7.1, it was already noted that even a small amount of surfactants in any of the adjacent phases may lead to the "solidification" of the interface, so that the laws of flow around a drop become close to those for a solid particle. This effect often occurs in practice.

The drag coefficient of a spherical drop can be determined by the formula

$$c_\mathrm{f}(\beta, \mathsf{Re}) = \frac{1}{\beta+1} c_\mathrm{f}(0, \mathsf{Re}) + \frac{\beta}{\beta+1} c_\mathrm{f}(\infty, \mathsf{Re}). \qquad \text{(E4.7.2.3)}$$

Here $c_\mathrm{f}(0, \mathsf{Re})$ is the drag coefficient of the spherical bubble, which can be calculated by the formula (E4.7.2.1), and $c_\mathrm{f}(\infty, \mathsf{Re})$ is the drag coefficient of a solid spherical particle, which can be calculated by (E4.6.3.4). The approximate expression (E4.7.2.3) gives three correct terms of the expansion for small Reynolds numbers; for $0 \le \mathsf{Re} \le 50$, the maximum error is less than 5%.

E4.8. Flow Past Nonspherical Particles

E4.8.1. Translational Stokes Flow Past Ellipsoidal Particles

The axisymmetric problem about a translational Stokes flow past an ellipsoidal particle admits an exact closed-form solution. In what follows, we present a brief summary of the corresponding results (see Happel & Brenner, 1965).

► **Oblate ellipsoid of revolution.** Consider an oblate ellipsoid of revolution with semiaxes a and b ($a > b$) in a translational Stokes flow with velocity U_i. We assume that the fluid viscosity is equal to μ.

We pass from the Cartesian coordinates X, Y, Z to the reference frame σ, τ, φ fixed to the oblate ellipsoid of revolution by using the transformations

$$X^2 = m^2(1+\sigma^2)(1-\tau^2)\cos^2\varphi, \quad Y^2 = m^2(1+\sigma^2)(1-\tau^2)\sin^2\varphi, \quad Z = m\sigma\tau,$$
$$\text{where} \quad m = \sqrt{a^2-b^2} \qquad (\sigma \ge 0,\ -1 \le \tau \le 1).$$

As a result, the ellipsoid surface is given by a constant value of the coordinate σ:

$$\sigma = \sigma_0, \qquad \text{where} \quad \sigma_0 = \bigl[(a/b)^2 - 1\bigr]^{-1/2}.$$

Since the problem is axisymmetric, we introduce the stream function as

$$V_\sigma = \frac{1}{m^2\sqrt{(1+\sigma^2)(\sigma^2+\tau^2)}}\frac{\partial\Psi}{\partial\tau}, \quad V_\tau = -\frac{1}{m^2\sqrt{(1-\tau^2)(\sigma^2+\tau^2)}}\frac{\partial\Psi}{\partial\sigma}.$$

Then the Stokes equations (E4.6.1.1) are reduced to one equation for Ψ, which can be solved by the separation of variables. By satisfying the boundary condition for a uniform flow remote from the particle and the no-slip conditions at the particle boundary, we obtain

$$\Psi = \frac{1}{2}m^2 U_\mathrm{i}(1-\tau^2)\left[\sigma^2 + 1 - \frac{(\sigma_0^2+1)\sigma - (\sigma_0^2-1)(\sigma^2+1)\operatorname{arccot}\sigma}{\sigma_0 - (\sigma_0^2-1)\operatorname{arccot}\sigma_0}\right].$$

In the similar problem on the motion of an oblate ellipsoid with velocity U_i in a quiescent fluid, we have

$$\Psi = -\frac{1}{2}m^2U_\mathrm{i}(1-\tau^2)\frac{(\sigma_0^2+1)\sigma-(\sigma_0^2-1)(\sigma^2+1)\operatorname{arccot}\sigma}{\sigma_0-(\sigma_0^2-1)\operatorname{arccot}\sigma_0}. \tag{E4.8.1.1}$$

The force exerted on the ellipsoid by the fluid is

$$F = \frac{8\pi\mu U_\mathrm{i}\sqrt{a^2-b^2}}{\sigma_0-(\sigma_0^2-1)\operatorname{arccot}\sigma_0}.$$

As $\sigma_0 \to 0$, an oblate ellipsoid degenerates into an infinite thin disk of radius a. By passing to the limit in (E4.8.1.1), we can obtain the following expression for the stream function:

$$\Psi = -\frac{1}{\pi}a^2U_\mathrm{i}(1-\tau^2)\bigl[\sigma+(\sigma^2+1)\operatorname{arccot}\sigma\bigr].$$

The disk moving in the direction perpendicular to its plane with velocity U_i in a quiescent fluid experiences the drag force

$$F = 16\mu aU_\mathrm{i}, \tag{E4.8.1.2}$$

which is less than the force acting on a sphere of the same radius (for the sphere, we have $F = 6\pi\mu aU_\mathrm{i}$). Formula (E4.8.1.2) is confirmed by experimental data.

► **Prolate ellipsoid of revolution.** To solve the corresponding problem about an ellipsoidal particle in a translational Stokes flow, we use the reference frame σ, τ, φ fixed to the prolate ellipsoid of revolution. The transformation to the coordinates (σ,τ,φ) is determined by the formulas

$$X^2 = m^2(\sigma^2-1)(1-\tau^2)\cos^2\varphi,\quad Y^2 = m^2(\sigma^2-1)(1-\tau^2)\sin^2\varphi,\quad Z = m\sigma\tau,$$
$$\text{where}\quad m = \sqrt{a^2-b^2}\qquad (\sigma \ge 1 \ge \tau \ge -1).$$

Here, as previously, the larger semiaxis is denoted by a. In this case, the ellipsoid surface is given by the equation

$$\sigma = \sigma_0,\qquad \text{where}\quad \sigma_0 = \bigl[1-(b/a)^2\bigr]^{-1/2}.$$

The fluid velocity is given by

$$V_\sigma = \frac{1}{m^2\sqrt{(\sigma^2-1)(\sigma^2-\tau^2)}}\frac{\partial\Psi}{\partial\tau},\quad V_\tau = -\frac{1}{m^2\sqrt{(1-\tau^2)(\sigma^2-\tau^2)}}\frac{\partial\Psi}{\partial\sigma}$$

in terms of the stream function

$$\Psi = \frac{1}{2}m^2U_\mathrm{i}(1-\tau^2)\left[\sigma^2-1-\frac{(\sigma_0^2+1)(\sigma^2-1)\operatorname{arctanh}\sigma-(\sigma_0^2-1)\sigma}{(\sigma_0^2+1)\operatorname{arctanh}\sigma_0-\sigma_0}\right],$$

where $\operatorname{arctanh}\sigma = \dfrac{1}{2}\ln\dfrac{\sigma+1}{\sigma-1}$.

In the problem about a prolate ellipsoid of revolution moving at a velocity U_i in a quiescent fluid, the corresponding stream function has the form

$$\Psi = -\frac{1}{2}m^2U_\mathrm{i}(1-\tau^2)\frac{(\sigma_0^2+1)(\sigma^2-1)\operatorname{arctanh}\sigma-(\sigma_0^2-1)\sigma}{(\sigma_0^2+1)\operatorname{arctanh}\sigma_0-\sigma_0}.$$

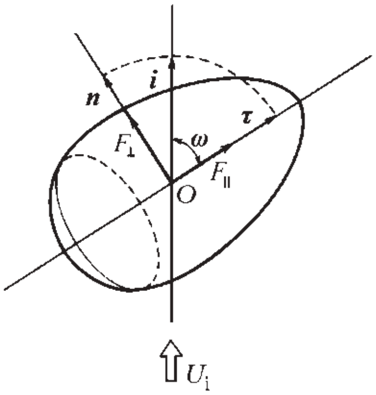

Figure E4.6. Body of revolution in translational flow (arbitrary orientation)

The drag force is calculated as

$$F=\frac{8\pi\mu U_\mathrm{i}\sqrt{a^2-b^2}}{(\sigma_0^2+1)\operatorname{arctanh}\sigma_0-\sigma_0}.$$

If $a\gg b$, then the prolate ellipsoid degenerates into a needle-like rod. In this case, the force acting on the needle of length a and radius b which moves in the direction of its axis at a velocity U_i has the form

$$F=\frac{4\pi\mu aU_\mathrm{i}}{\ln(a/b)+0.193}.$$

E4.8.2. Translational Stokes Flow Past Bodies of Revolution

Consider bodies of revolution of any shape with arbitrary orientation in a translational flow at low Reynolds numbers. We assume that the axis of the body of revolution forms an angle ω with the direction of the fluid velocity at infinity (Figure E4.6). The unit vector $\mathbf{i}$ directed along the flow can be represented as the sum $\mathbf{i}=\boldsymbol{\tau}\cos\omega+\mathbf{n}\sin\omega$, where $\boldsymbol{\tau}$ is the unit vector directed along the body axis and $\mathbf{n}$ the unit vector in the plane of rotation of the body. In the Stokes approximation, the drag force is given by the following expression in the general case:

$$\mathbf{F}=\boldsymbol{\tau}F_{||}\cos\omega+\mathbf{n}F_\perp\sin\omega, \qquad \text{(E4.8.2.1)}$$

where $F_{||}$ and $F_\perp$ are the drag forces of the body of revolution for its parallel ($\omega=0$) and perpendicular ($\omega=\pi/2$) positions in the translational flow.

The projection of the drag force onto the incoming flow direction is equal to the inner product

$$(\mathbf{F}\cdot\mathbf{i})=F_{||}\cos^2\omega+F_\perp\sin^2\omega. \qquad \text{(E4.8.2.2)}$$

It follows from (E4.8.2.1) and (E4.8.2.2) that to calculate the drag force of a body of revolution of any shape with arbitrary orientation in a Stokes flow, it suffices to know the value of this force only for two special positions of the body in space. The "axial" ($F_{||}$) and "transversal" ($F_\perp$) drags can be obtained both theoretically and experimentally. In what follows, we present the expressions for $F_{||}$ and $F_\perp$ for some bodies of revolution of nonspherical shape.

For a thin circular disk of radius a, one has

$$F_{||}=16\mu aU_\mathrm{i},\qquad F_\perp=\tfrac{32}{3}\mu aU_\mathrm{i}.$$

For a dumbbell-like particle consisting of two adjacent spheres of equal radius a, one has

$$F_{\|} = 12\pi\mu a U_{\rm i}\lambda_{\|}, \qquad \lambda_{\|} \approx 0.645,$$
$$F_{\perp} = 12\pi\mu a U_{\rm i}\lambda_{\perp}, \qquad \lambda_{\perp} \approx 0.716.$$

In these formulas, the product $12\pi\mu a U_{\rm i}$ is equal to the sum of drag forces for two isolated spheres of radius a.

For oblate ellipsoids of revolution with semiaxes a and b, one has

$$F_{\|} = 3.77\,(4a + b), \qquad F_{\perp} = 3.77\,(3a + 2b), \tag{E4.8.2.3}$$

where a is the equatorial radius ($a > b$).

For prolate ellipsoids of revolution with semiaxes a and b, one has

$$F_{\|} = 3.77\,(a + 4b), \qquad F_{\perp} = 3.77\,(2a + 3b), \tag{E4.8.2.4}$$

where b is the equatorial radius ($b > a$).

Formulas (E4.8.2.3) and (E4.8.2.4) are approximate. They work well for slightly deformed ellipsoids of revolution. In (E4.8.2.3), the maximum error is less than 6% for any ratio of the semiaxes.

E4.8.3. Translational Stokes Flow Past Particles of Arbitrary Shape

► **Hydrodynamic force and angular momentum acting on a particle.** A particle of arbitrary shape moving in an infinite fluid that is at rest at infinity is subject to the action of the hydrodynamic force and angular momentum due to its translational motion and rotation, respectively:

$$\mathbf{F} = \mu(\mathbf{K}\,\mathbf{U} + \mathbf{S}\,\boldsymbol{\omega}), \tag{E4.8.3.1}$$
$$\mathbf{M} = \mu(\mathbf{S}\,\mathbf{U} + \boldsymbol{\Omega}\,\boldsymbol{\omega}), \tag{E4.8.3.2}$$

where $\mathbf{K}$, $\mathbf{S}$, and $\boldsymbol{\Omega}$ are tensors of rank two depending on the particle geometry.

The symmetric tensor $\mathbf{K} = [K_{ij}]$ is called translational. It characterizes the drag of a body under translational motion and depends only on the size and shape of the body. In the principal axes, the translation tensor is reduced to the diagonal form

$$\mathbf{K} = \begin{bmatrix} K_1 & 0 & 0 \\ 0 & K_2 & 0 \\ 0 & 0 & K_3 \end{bmatrix}, \tag{E4.8.3.3}$$

where K_1, K_2, and K_3 are the principal drags acting on the body as it moves along the major axes. For orthotropic bodies (with three symmetry planes orthogonal to each other), the principal axes of the translational tensor are perpendicular to the corresponding symmetry planes. For axisymmetric bodies,, one of the axes (say, the first) is a major axis, and $K_2 = K_3$. For a sphere of radius a, any three pairwise perpendicular axes are major, and $K_1 = K_2 = K_3 = 6\pi a$.

A symmetric tensor $\boldsymbol{\Omega}$ is called a rotational tensor. It depends both on the shape and size of the particle and on the choice of the origin. The rotational tensor characterizes the drag under rotation of the body and has diagonal form with entries Ω_1, Ω_2, Ω_3 in the principal axes (the positions of the principal axes of the rotational and translational tensors in space are different). For axisymmetric bodies, one of the major axes (for instance, the first) is parallel to the symmetry axis, and in this case $\Omega_2 = \Omega_3$. For a spherical particle, we have $\Omega_1 = \Omega_2 = \Omega_3$.

The tensor **S** is symmetric only at a point O unique for each body; this point is called the center of hydrodynamic reaction. This tensor is called the conjugate tensor and characterizes the combined reaction of the body under translational and rotational motion (the drag moment in the translational motion and the drag force in the rotational motion). For bodies with orthotropic, axial, or spherical symmetry, the conjugate tensor is zero. However, it is necessary to take this tensor into account for bodies with helicoidal symmetry (propeller-like bodies).

In problems of gravity settling of particles, the translational tensor is most important.

► **Principal drags of some nonspherical bodies.**

For a thin circular disk of radius a, we have

$$K_1 = 16a, \qquad K_2 = \tfrac{32}{3}a, \qquad K_3 = \tfrac{32}{3}a. \tag{E4.8.3.4}$$

For needle-like ellipsoids of length l and radius a, one has

$$K_1 = \frac{4\pi l}{2\ln(l/a) - 1}, \quad K_2 = \frac{4\pi l}{\ln(l/a) + 0.5}, \quad K_3 = \frac{4\pi l}{\ln(l/a) + 0.5}. \tag{E4.8.3.5}$$

For thin circular cylinders of length l and radius a, one has

$$K_1 = \frac{4\pi l}{\ln(l/a) - 0.72}, \quad K_2 = \frac{4\pi l}{\ln(l/a) + 0.5}, \quad K_3 = \frac{4\pi l}{\ln(l/a) + 0.5}. \tag{E4.8.3.6}$$

For a dumbbell-like particle consisting of two adjacent spheres of equal radius a, one has

$$K_1 = 24.3\,a, \qquad K_2 = 27.0\,a, \qquad K_3 = 27.0\,a. \tag{E4.8.3.7}$$

► **Axisymmetric bodies of arbitrary shape.** For axisymmetric bodies of arbitrary shape, we introduce the notion of perimeter-equivalent sphere. To this end, we project all points of the surface of the body on a plane perpendicular to its axis. The projection is a circle of radius $a_\perp$. The perimeter-equivalent sphere has the same radius.

We let $\bar{c}_{\|}$ denote the relative coefficient of the axial drag equal to the ratio of the axial drag of an axisymmetric body to the drag of a sphere with equivalent perimeter. The relative coefficients of transversal drag $\bar{c}_\perp$ equal to the ratio of the transversal drag of an axisymmetric body to the drag of a sphere with equivalent perimeter are introduced similarly.

Experimental data and numerical results for principal values of the translational tensor for some axisymmetric and orthotropic bodies (cylinders, doubled cones, parallelepipeds) are well approximated by the following dependence for the relative coefficient of the axial drag:

$$\bar{c}_{\|} = 0.244 + 1.035\,\Sigma - 0.712\,\Sigma^2 + 0.441\,\Sigma^3 \qquad (0.4 \le \Sigma \le 1.2).$$

The values of the perimeter-equivalent factor Σ are equal to the ratio of the surface area of the particle to the surface area of the perimeter-equivalent sphere.

To determine the relative coefficients of transversal drag, one can use the formula

$$\bar{c}_\perp = 0.392 + 0.621\,\Sigma - 0.04\,\Sigma^2,$$

which is well consistent with experimental data for $0.5 \le \Sigma \le 4.5$.

E4.8.4. Sedimentation of Isotropic and Nonisotropic Particles

► **Sedimentation of isotropic particles.** The steady-state rate U_i of particle settling in mass force fields, primarily, in the gravitational field, is an important hydrodynamic characteristic of processes in chemical technology such as settling and sedimentation. Any spherically isotropic body homogeneous with respect to density experiences the same drag under translational motion regardless of its orientation. Such a body is also isotropic with respect to any pair of forces that arise when it rotates around an arbitrary axis passing though its center. If at the initial time such a body has some orientation in fluid and can fall without initial rotation, then this body falls vertically without rotation, preserving its initial orientation.

It is convenient to describe the free fall of nonspherical isotropic particles by using the sphericity parameter

$$\psi = \frac{S_\text{e}}{S_*},$$

where S_* is the surface area of the particle and S_e is the surface area of the volume-equivalent sphere. If the motion is slow, one can calculate the settling rate by the empirical formula

$$\mathbf{U}_\text{i} = \frac{2}{9}\frac{Q\Delta\rho a_\text{e}^2}{\mu}\mathbf{g},$$

where a_e is the radius of the volume-equivalent sphere and

$$Q = 0.843\ln\frac{\psi}{0.065}.$$

We present some values of the sphericity factor ψ for some particles: sphere, 1.000; octahedron, 0.846; cube, 0.806; tetrahedron, 0.670.

► **Sedimentation of nonisotropic particles.** If the velocity of a spherical particle in Stokes settling is always codirected with the gravity force, even for homogeneous axisymmetric particles the velocity is directed vertically if and only if the vertical coincides with one of the principal axes of the translational tensor **K**. If the angle between the symmetry axis and the vertical is φ, then the velocity direction is given by the angle

$$\theta = \pi + \arctan\left(\frac{K_2}{K_1}\tan\varphi\right),$$

where K_1 and K_2 are the axial and transversal principal drags to the translational motion. The velocity is given by

$$U_\text{i} = \frac{V\Delta\rho g}{\mu}(K_1^2\cos^2\theta + K_2^2\sin^2\theta)^{-1/2},$$

where V is the particle volume.

If the settling direction is not vertical, this means that a falling particle is subject to the action of a transverse force, which leads to its horizontal displacement. An additional complication is that the center of hydrodynamic reaction (including the buoyancy force) does not coincide with the particle center of mass. In this case, in addition to the translational motion, the particle is subject to rotation under the action of the arising moment of forces (e.g., the "somersault" of a bullet with displaced center of mass). For axisymmetric particles, this rotation stops when the system "the mass center + the reaction center" becomes stable, that is, the mass center is ahead of the reaction center. In this case, the settling trajectory becomes stable and rectilinear.

However, in the more general case of an asymmetric particle, the combined action of the lateral and rotational forces may lead to motion along a 3D (for instance, spiral, trajectory). At the same time, a steady-state settling trajectory with helicoidal (propeller-like) symmetry remains rectilinear, notwithstanding the body rotation.

Two theorems are useful for estimating the steady-state rate of settling of Stokes nonspherical particles. One theorem, proved by Hill & Power (1956), states that the Stokes drag of an arbitrary body moving in a viscous fluid is larger than the Stokes drag of any inscribed body. Thus, to determine the upper and lower bounds for the Stokes drag of a body of an exotic shape, one can suggest a reasonable set of inscribed and circumscribed bodies with known drags. The other theorem, proved by Weinberger (1972), states that of all particles of different shape but equal volume, the spherical particle has the maximum Stokes settling rate.

► **Mean velocity of nonisotropic particles falling in a fluid.** The mean flow rate velocity $\langle\mathbf{U}\rangle$ of a particle, which is obtained in a large number of experiments when an arbitrarily oriented particle falls in a fluid, is determined for the Stokes flow by the formula:

$$\langle\mathbf{U}\rangle = \frac{V\Delta\rho}{\mu\overline{K}}\mathbf{g}, \tag{E4.8.4.1}$$

where V is the particle volume, $\Delta\rho$ is the difference between the densities of the particle and the fluid, $\mathbf{g}$ is the free fall acceleration, and $\overline{K}$ is the average drag expressed via the principal drags as

$$\frac{1}{\overline{K}} = \frac{1}{3}\left(\frac{1}{K_1} + \frac{1}{K_2} + \frac{1}{K_3}\right). \tag{E4.8.4.2}$$

The average drag force acting on an arbitrarily oriented particle falling in fluid is given by

$$\langle\mathbf{F}\rangle = -\mu\overline{K}\langle\mathbf{U}\rangle. \tag{E4.8.4.3}$$

Formulas (E4.8.4.1)–(E4.8.4.3) are important in view of some problems of Brownian motion.

In the special case of a spherical particle, one must set $V = \frac{4}{3}\pi a^3$ and $\overline{K} = 6\pi a$ in (E4.8.4.1).

Let us calculate the average drag for a thin disk of radius a. To this end, we substitute the principal drags (E4.8.3.4) into (E4.8.4.2). As a result, we obtain

$$\overline{K} = 12a.$$

Substituting the coefficients K_1, K_2, and K_3 from (E4.8.3.5)–(E4.8.3.7) into (E4.8.4.2) and using (E4.8.4.1), one obtains the average settling rate for the above-mentioned nonspherical bodies.

E4.8.5. Flow Past Nonspherical Particles at Various Reynolds Numbers

The Stokes flow around particles of any shape is separation-free—that is, the stream lines come from infinity—bend round the body everywhere closely attaining the body surface, and return to infinity. However, for higher Reynolds numbers, separation occurs, which leads to wake formation behind the body. As the Reynolds number increases, the size of the wake region (the wake length) grows differently for different bodies. With further increase in the Reynolds number, the wake becomes unsteady and completely turbulent and goes to infinity. The force action of the flow on the body is closely related to the wake size and state. The limit asymptotic cases of this action are the Stokes (as $\mathsf{Re} \to 0$) and the Newtonian (as $\mathsf{Re} \to \infty$) regimes of flow. We have already considered the characteristics of the Stokes flow. The Newtonian regime of flow is characterized by the fact that the drag coefficient $c_{\rm f}$ of the body is constant.

In axial flow past disks, which are the limit cases of axisymmetric bodies of small length, the drag coefficients for the entire range of Reynolds numbers calculated with respect to the radius can be determined by approximate formulas, which agree well with numerical results and experimental data:

$$\begin{aligned}
c_{\rm f} &= 10.2\,\mathsf{Re}^{-1}(1+0.318\,\mathsf{Re}) && \text{for} \quad \mathsf{Re} \le 0.005,\\
c_{\rm f} &= 10.2\,\mathsf{Re}^{-1}(1+10^{s}) && \text{for} \quad 0.005 < \mathsf{Re} \le 0.75,\\
c_{\rm f} &= 10.2\,\mathsf{Re}^{-1}(1+0.239\,\mathsf{Re}^{0.792}) && \text{for} \quad 0.75 < \mathsf{Re} \le 66.5,\\
c_{\rm f} &= 1.17 && \text{for} \quad \mathsf{Re} > 66.5,
\end{aligned}$$

where $s = -0.61 + 0.906\log_{10}\mathsf{Re} - 0.025\,(\log_{10}\mathsf{Re})^2$.

The steady-state settling rate of a particle of an arbitrary shape (for Newtonian regime of a flow at high Reynolds numbers) can be obtained by the formula

$$U = 0.69\,\gamma^{1/36}[g a_{\rm e}(\gamma-1)(1.08-\psi)]^{1/2} \quad \text{for} \quad 1.1 < \gamma < 8.6,$$

where γ is the particle-fluid density ratio, $a_{\rm e}$ is the radius of the volume-equivalent sphere, and ψ is the ratio of the area of the volume-equivalent sphere to the particle surface area.

E4.9. Motion of Several Particles. Effective Viscosity of Suspensions

The motion of a particle in infinite fluid creates some velocity and pressure fields. Neighboring particles move in already perturbed hydrodynamic fields. Simultaneously, the first particle itself experiences hydrodynamic interaction with the neighboring particles and neighboring moving or fixed surfaces. Since in a majority of actual disperse systems, the existence of an ensemble of particles and the apparatus walls is inevitable, the consideration of the hydrodynamic interaction between these objects is very important. One method for obtaining the required information about the interaction is based on the construction of exact closed-form solutions. However, even within the framework of Stokes hydrodynamics, to describe the motion of an ensemble of particles is a very complicated problem, which admits an exact closed-form solution only in exceptional cases.

E4.9.1. Motion of Two Spheres. Gravitational Sedimentation of Several Spheres

► **Motion of two spheres along a line passing through their centers.** Consider the axisymmetric problem about the motion of two spheres at the same velocity U. In the Stokes approximation, the force acting on each of the spheres is described by the formula

$$F = 6\pi\mu a U\lambda, \tag{E4.9.1.1}$$

where a is the sphere radius and λ is the correction coefficient depending on the radii and the distance l between the centers of the spheres. In the case of spheres of equal radii, the expression for λ has the form

$$\lambda=\frac{4}{3}\sinh k\sum_{n=1}^{\infty}\frac{n(n+1)}{(2n-1)(2n+3)}\left\{1-\frac{4\sinh^2[(n+\frac{1}{2})k]-(2n+1)^2\sinh^2 k}{2\sinh[(2n+1)k]+(2n+1)\sinh 2k}\right\},\qquad\text{(E4.9.1.2)}$$

where $k=\ln\Bigl[\frac{1}{2}(l/a)+\sqrt{\frac{1}{4}(l/a)^2-1}\,\Bigr]$.

For numerical calculations, it is convenient to use the approximate formula

$$\lambda=\frac{0.88\,a+l}{2.5\,a+l},\qquad\text{(E4.9.1.3)}$$

whose maximal error is less than 1.3% for any a and l.

Since $\lambda\le 1$, it follows from (E4.9.1.1) that the velocity of the steady-state motion of each of the spheres in the ensemble is greater than the velocity of a single sphere.

► **Gravitational sedimentation of several spheres.** Some relations between the drag force F and the sedimentation velocity U averaged with respect to various orientations of particles of equal radius in space were obtained by Happel & Brenner (1965). It was assumed that the maximum distance l between the sphere centers is much larger than their radius a. In all considered cases, the drag force is described by formula (E4.9.1.1), where λ is the correction coefficient depending on the configuration of the system of particles. In what follows, we write out the correction coefficient for some typical configurations of particles.

For a system of two spheres, one has

$$\lambda=\frac{2}{1+(a/l)}.$$

For a system of three spheres arranged in a line,

$$\lambda=\frac{3}{1+\frac{10}{3}(a/l)-\frac{1}{4}(a/l)^2}.$$

For a system of four spheres arranged in a line,

$$\lambda=\frac{4}{1+\frac{13}{2}(a/l)-\frac{9}{8}(a/l)^2}.$$

For a system of four spheres arranged in vertices of a square,

$$\lambda=\frac{4}{1+2.7\,(a/l)-0.04\,(a/l)^2}.$$

For a system of eight spheres arranged in vertices of a cube,

$$\lambda=\frac{8}{1+5.7\,(a/l)-0.34\,(a/l)^2}.$$

E4.9.2. Rate of Suspension Precipitation. Effective Viscosity of Suspensions

► **Cellular model. Rate of suspension precipitation.** The two-phase media with large number of particles are described by a widely used approximate model, namely, the cell model. To each particle of the disperse phase, this model assigns the corresponding volume of the free liquid. Thus, the entire suspension (or emulsion) is divided into a collection of spherical cells of radius b whose centers coincide with the centers of particles of radius a. The geometric parameters of these cells are related to the volume concentration ϕ of the disperse phase as follows:

$$b = a\phi^{-1/3}.$$

The definition of a particle velocity $\mathbf{U}$ determines the axial symmetry of the problem, which is convenient to consider in spherical coordinates. In the Stokes approximation, the general solution of such a problem is determined by formula (E4.6.1.4), where arbitrary constants must be determined from the conditions that the solution is bounded at infinity and the velocity is known on the particle surface and from some other conditions on the cell boundary (for $R = b$). The condition that the normal velocity on this boundary is zero—that is, there is no flow across the cell—is undoubtedly obvious. The second condition, which is necessary for the complete identification of the solution, leads to different opinions. Cunningham postulated that the tangential velocity is zero, thus considering a cell as a container with rigid boundary. Happel proposed to use the condition that the tangential stress is zero, thus postulating that the cell is an isolated force unit (is in equilibrium under the action of forces). Kuwabara proposed to use the condition that the vortex intensity is zero on the cell boundary. Finally, Slobodov and Chepura assumed that the vortex intensity flux vanishes on the cell boundary.

The choice of the boundary condition essentially determines the model of force interaction between the particle at the center of the cell and the other particles. The steady-state velocities of the gravitational sedimentation in suspensions obtained with the help of cell models were compared with numerous experimental data and it was shown that the most precise results can be obtained by using the Slobodov–Chepura model. This model, under the assumption that the particle at the cell center is a drop of a liquid with different viscosity, leads to formula (E4.9.1.1) for the drag forces, where the correction coefficient can be calculated as

$$\lambda = \frac{\beta + \frac{2}{3}}{1 - \frac{3}{5}\phi^{1/3} - \frac{2}{5}\phi^2 + \beta\left(1 - \frac{9}{10}\phi^{1/3} + \frac{1}{2}\phi + \frac{2}{5}\phi^2\right)}, \quad \beta = \frac{\mu_2}{\mu_1}.$$

As $\phi \to 0$ and $\beta \to \infty$, we have $\lambda \to 1$, which corresponds to the Stokes drag law.

► **Effective viscosity of suspensions.** Suspensions of particles in fluid are widely used in various processes of chemical technology. If the dimensions of particles in the suspension are essentially smaller than those of the apparatus, then one can consider the suspension as some continuous medium with properties other than properties of the disperse phase.

Very often, this medium remains Newtonian from the viewpoint of its rheological properties, but the viscosity of this medium is somewhat larger than that of the continuous phase. This viscosity μ_{ef} is called the effective viscosity. In practice, it is convenient to relate the effective viscosity to the viscosity μ of the continuous phase and consider the dimensionless effective viscosity $\bar{\mu} = \mu_{\text{ef}}/\mu$.

The value $\bar{\mu}$ depends primarily on the volume concentration of the disperse phase ϕ. The well-known Einstein formula

$$\bar{\mu} = 1 + 2.5\,\phi, \qquad \text{(E4.9.2.1)}$$

holds for strongly rarefied suspensions of solid spherical particles.

For $0 \le \phi \le 0.4$, the results of experimental and numerical calculations are well approximated by the formula

$$\bar{\mu} = 1 + 2.5\,\phi + 12.5\,\phi^2,$$

which becomes the Einstein formula (E4.9.2.1) as $\phi \to 0$.

The dimensionless effective viscosity of a rarefied emulsion of spherical drops and bubbles moving in a fluid can be determined by the formula

$$\bar{\mu} = 1 + \frac{5\beta + 2}{2\beta + 2}\,\phi, \qquad \text{(E4.9.2.2)}$$

where β is the ratio of the drop viscosity to the fluid viscosity. By passing to the limit as $\beta \to \infty$ in (E4.9.2.2), one obtains the Einstein formula (E4.9.2.1). The value $\beta = 0$ corresponds to gas bubbles.

Bibliography for Chapter E4

Batchelor, G. K., *An Introduction to Fluid Dynamics*, Cambridge Univ. Press, Cambridge, England, 1967.

Blasius, H., Crenzschichten in Flussigkeiten mit Kleiner Reibung, *Zeitschr. für Math. und Phys.,* Bd. 56, Ht. 1, S. 1–37, 1908.

Bradshaw, P., *Turbulence*, Springer-Verlag, Berlin, 1978.

Chhabra, R. P., *Bubbles, Drops, and Particles in Non-Newtonian Fluids*, CRC Press, Boca Raton, Florida, 1993.

Clift, R., Grace, J. R., and Weber, M. E., *Bubbles, Drops and Particles*, Acad. Press, New York, 1978.

Emanuel, G., *Analytical Fluid Dynamics*, CRC Press, Boca Raton, Florida, 1993.

Faber, T. E., *Fluid Dynamics for Physicists*, Cambridge Univ. Press, Cambridge, England, 1995.

Frost, W. and Moulden, T. H., *Handbook of Turbulence*, Plenum Press, New York, 1977.

Hadamard, J. S., Mouvement permanent lent d'une sphere liquide et visqueuse dans un liquide visqueux, *Comp. Rend. Acad. Sci. Paris*, Vol. 152, No. 25, pp. 1735–1739, 1911 and Vol. 154, No. 3, p. 109, 1912.

Happel, J. and Brenner, H., *Low Reynolds Number Hydrodynamics*, Prentice-Hall, Englewood Cliffs, New Jersey, 1965.

Hetsroni, G., *Handbook of Multiphase Systems*, Hemisphere Publ. Corp., Washington, 1982.

Hill, R. and Power, G., Extremum principles for slow viscous flow and approximate calculation of drag, *Quarterly J. Mech. Appl. Math.*, 1956, Vol. 9, No. 3, pp. 313–319.

Hinze, J. O., *Turbulence*, McGraw-Hill, New York, 1975.

Idelchik, I. E., *Handbook of Hydraulic Resistance*, Begell House, New York, 1994.

Kaplun, S. and Lagerstrom, P. A., Asymptotic expansions of Navier–Stokes solutions for small Reynolds numbers, *J. Math. Mech.,* Vol. 6, pp. 585–593, 1957.

Kármán, Th., von, *Mechanische Ahnlichkeit und Turbulenz*, Verhandlg. d. III Intern. Kongress für Techn. Mechanik, Stockholm, Bd. 1, S. 85, 1930.

Kutateladze, S. S., *Heat Transfer and the Hydrodynamical Resistance. A Reference Book* [in Russian], Energoatomizdat, Moscow, 1990.

Lamb, H., *Hydrodynamics*, Dover Publ., New York, 1945.

Landau, L. D. and Lifschitz, E. M., *Fluid Mechanics. Course of Theoretical Physics*, Pergamon Press, Oxford, 1987.

Lesieur, M., *Turbulence in Fluids*, Kluwer Acad. Publ., Dordrecht, 1997.

Levich, V. G., *Physicochemical Hydrodynamics*, Prentice-Hall, Englewood Cliffs, New Jersey, 1962.

Loitsyanskiy, L. G., *Mechanics of Liquids and Gases*, Begell House, New York, 1996.

Mobius, D. and Miller, R., *Drops and Bubbles in Interfacial Research*, Elsevier Science Ltd., Oxford, 1997.

Monin, A. S. and Yaglom, A. M., *Statistical Fluid Mechanics: Theory of Turbulence, Vols. 1 and 2* [in Russian], Gidrometeoizdat, St. Petersburg, 1992 and 1996.

Nikitin N. V., Statistical characteristics of wall turbulence, *Fluid Dynamics*, Vol. 31, No. 3, pp. 361–370, 1996.

Oseen, C. W., Über die Stokes'sche Formel, und über eine verwandte Aufgabe in der Hydrodynamik, *Ark. Math. Astronom. Fys.,* Bd. 6, Ht. 29, 1910.

Paneli, D. and Gutfinger, C., *Fluid Mechanics*, Cambridge Univ. Press, Cambridge, England, 1997.

Petukhov, B. S., *Heat Transfer and Drag in Laminar Flow of Liquids in Tubes* [in Russian], Energiya, Moscow, 1967.

Polyanin, A. D., Kutepov, A. M., Vyazmin, A. V., and Kazenin, D. A., *Hydrodynamics, Mass and Heat Transfer in Chemical Engineering*, Taylor & Francis, London, 2002.

Prandtl, L., The mechanics of viscous fluids, In: *Aerodynamics Theory*, Vol. 3, pp. 34–208; Springer-Verlag, Berlin, 1935.

Proudman, I. and Pearson, J. R. A., Expansions at small Reynolds number for the flow past a sphere and circular cylinder, *J. Fluid Mech.,* Vol. 2, No. 3, pp. 237–262, 1957.

Reid, R. C., Prausnitz, J. M., and Sherwood, T. K., *The Properties of Gases and Liquids*, McGraw-Hill Book Comp., New York, 1977.

Reynolds, A. J., *Turbulent Flows in Engineering*, Wiley, London, 1974.

Rybczynski M. W., Über die fortschreitende Bewegung einer flüssigen Kugel in einem zähen Medium, *Bull. Acad. Sci. Cracovie, Ser. A, Sci. Math.,* Bd. 1, S. 40–46, 1911.

Schlichting, H., *Boundary Layer Theory*, McGraw-Hill, New York, 1981.

Shivamoggi, B., *Theoretical Fluid Dynamics*, Wiley VCH, Chichester, New York, 1998.

Slobodov, E. B. and Chepura I. V., A cellular model of biphasal media, *Theor. Found. Chem. Engng.*, 1982, Vol. 16, No. 3, pp. 235–239.

So, R. M. C., Speciale, C. G., and Launder, B. E., *Near-Wall Turbulent Flows*, Elsevier, Amsterdam, 1993.

Soo, S. L., *Fluid Dynamics of Multiphase Systems*, Blaisdell Publ. Comp., Waltham, 1968.

Stimson, M. and Jeffrey, G. B., The motion of two spheres in a viscous flow, *Proc. Roy. Soc. London*, Vol. A111, No. 757, p. 110, 1926.

Stokes, G. G., On the effect of the internal friction of fluids on the motion of pendulums, *Trans. Camb. Phil. Soc.,* Vol. 9, No. 2, pp. 8–106, 1851.

Taylor, G. I., Viscosity of a fluid, containing small drops of another fluid, *Proc. Roy. Soc. London*, Vol. A138, No. 834, pp. 41–48, 1932.

Townsend, A. A., *The Structure of Turbulent Shear Flow*, Cambridge Univ. Press, London, 1976.

Van Dyke, M. D., *Perturbation Methods in Fluid Mechanics*, Parabolic Press, Stanford California, 1975.

Vargaftic, N. V., Vinogradov, Y. K., and Yargin, V. S., *Handbook of Physical Properties of Liquids and Gases*, Begell House Inc., New York, 1996.

Warsi, Z. U. A., *Fluid Dynamics*, CRC Press, Boca Raton, Florida, 1993.

Weinberger, H. F., Variational properties of steady fall in Stokes flow, *J. Fluid Mech.*, 1972, Vol. 52, Pt. 2, pp. 321–344.

Zapryanov, Z. and Tabakova, S., *Dynamics of Bubbles, Drops and Rigid Particles*, Kluwer Acad. Publ., Dordrecht, 1999.

Zdravkovich, M. M., *Flow Around Circular Cylinders*, Oxford Sci. Publ., Oxford, 1997.

Chapter E5

Mass and Heat Transfer

E5.1. Convective Mass and Heat Transfer. Equations and Boundary Conditions

E5.1.1. Mass and Heat Transfer in Laminar Flows

► **Mass concentration.** So far, we have considered the motion of fluids of homogeneous physical and chemical composition. In practice, one often meets more complicated situations in which the fluid contains dissolved substances (contaminants or reactants) and is a solution or a mixture.

Water solutions of common salt or sugar and water-alcohol mixtures are the simplest examples of such systems.

The composition of solutions and mixtures is usually characterized by the mass density of the substance (the mass of dissolved substance per unit volume) or the dimensionless mass concentration C (the ratio of the mass density of a substance to the total density of the mixture). The latter is normally used in this chapter; for brevity, we refer to it simply as the concentration.

► **Convective mass transfer equation.** Let us write out the main equations and boundary conditions used in the mathematical statement of mass transfer and chemical hydrodynamic problems. We assume that the medium density and viscosity are independent of the concentration and that the concentration distribution does not affect the flow field. This allows one to analyze the hydrodynamic problem about the fluid motion and the diffusion problem of finding the concentration field independently. It is assumed that the information about the fluid velocity field necessary for the solution of the diffusion problem is known. We also assume that the diffusion coefficient is independent of the concentration. For simplicity, we restrict our considerations to the case of two-component solutions.

In Cartesian coordinates X, Y, Z, the solute transfer in absence of homogeneous transformations is described by the equation

$$\frac{\partial C}{\partial t} + V_X \frac{\partial C}{\partial X} + V_Y \frac{\partial C}{\partial Y} + V_Z \frac{\partial C}{\partial Z} = D \left(\frac{\partial^2 C}{\partial X^2} + \frac{\partial^2 C}{\partial Y^2} + \frac{\partial^2 C}{\partial Z^2} \right), \qquad \text{(E5.1.1.1)}$$

where C is the concentration, D is the diffusion coefficient, and V_X, V_Y, and V_Z are the fluid velocity components, which are assumed to be given.

Equation (E5.1.1.1) reflects the fact that the transfer of a substance in a moving medium is due to two distinct physical mechanisms. First, there is molecular diffusion due to concentration difference in a liquid or gas, which tends to equalize the concentrations. Second, the solute is carried along by the moving medium. The combination of these two processes is usually called *convective diffusion*.

► **Initial condition and the simplest boundary conditions.** To complete the statement of the problem, it is necessary to supplement equation (E5.1.1.1) with initial and boundary conditions. As the initial condition, one usually takes the concentration profile in the flow at time $t = 0$. The boundary conditions are, as a rule, given on some surface and far from it, in the bulk of the solution. The latter condition corresponds to prescribing the unperturbed concentration C_i at infinity:

$$\xi_* \to \infty, \quad C \to C_\mathrm{i}, \tag{E5.1.1.2}$$

where ξ_* is the distance measured along the normal to the surface.

In solid dissolution problems, it is usually assumed in the boundary conditions that the concentration is zero in the bulk of the fluid, $C_\mathrm{i} = 0$, and constant on the crystal surface,

$$\xi_* = 0, \quad C = C_\mathrm{s}, \tag{E5.1.1.3}$$

where C_s is given. The boundary conditions (E5.1.1.2) (with $C_\mathrm{i} = 0$) and (E5.1.1.3) are also used in liquid drop evaporation problems.

The boundary conditions on a surface where a chemical reaction occurs depend on the specific physical statement of the problem. In the special case of an “infinitely rapid” heterogeneous chemical reaction, the corresponding boundary condition has the form

$$\xi_* = 0, \quad C = 0 \tag{E5.1.1.4}$$

and means that the reagent is completely taken up in the reaction on the interface. Such a situation is often called the diffusion regime of reaction. Condition (E5.1.1.4) has the following meaning: the chemical reaction on the interface proceeds vigorously, so that all the substance that approaches the interface takes part in the reaction. Note that condition (E5.1.1.4) is a special case of (E5.1.1.3) for $C_\mathrm{s} = 0$.

► **Dimensionless mass transfer equation and boundary conditions.** Let us introduce a characteristic length a (for example, the particle or tube radius) and a characteristic velocity U (for example, the unperturbed flow velocity far from a particle or the fluid velocity on the tube axis). First, consider the boundary conditions (E5.1.1.2) and (E5.1.1.3). Then it is convenient to rewrite equation (E5.1.1.1) for the convective mass transfer in the following dimensionless form. We introduce dimensionless variables by setting

$$\begin{gathered}\tau = \frac{Dt}{a^2}, \quad x = \frac{X}{a}, \quad y = \frac{Y}{a}, \quad z = \frac{Z}{a}, \quad \xi = \frac{\xi_*}{a},\\ v_x = \frac{V_X}{U}, \quad v_y = \frac{V_Y}{U}, \quad v_z = \frac{V_Z}{U}, \quad c = \frac{C_\mathrm{i} - C}{C_\mathrm{i} - C_\mathrm{s}}\end{gathered} \tag{E5.1.1.5}$$

and substitute them into (E5.1.1.1). As a result, we obtain

$$\frac{\partial c}{\partial \tau} + \mathsf{Pe}\left(v_x \frac{\partial c}{\partial x} + v_y \frac{\partial c}{\partial y} + v_z \frac{\partial c}{\partial z}\right) = \frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} + \frac{\partial^2 c}{\partial z^2}. \tag{E5.1.1.6}$$

Here the Peclet number $\mathsf{Pe} = aU/D$ is a dimensionless parameter characterizing the ratio of convective transfer to diffusion transfer and a is a characteristic length (tube radius, film thickness, etc.).

In the new variables (E5.1.1.5), the boundary condition far from the surface (E5.1.1.2) has the form

$$\xi \to \infty, \quad c \to 0. \tag{E5.1.1.7}$$

In a similar way, taking into account (E5.1.1.3) and (E5.1.1.5), we obtain the boundary condition

$$\xi = 0, \quad c = 1. \tag{E5.1.1.8}$$

For brevity, we rewrite the convective diffusion equation (E5.1.1.6) in the following frequently used form:

$$\frac{\partial c}{\partial \tau} + \mathsf{Pe}\,(\mathbf{v}\cdot\nabla)\,c = \Delta c, \qquad \text{(E5.1.1.9)}$$

where ∇ is the Hamilton operator and Δ is the Laplace operator; the explicit form of these operators in the Cartesian coordinates x, y, z follows by comparing (E5.1.1.6) and (E5.1.1.9).

► **Heat transfer equation. Boundary conditions.** We assume that the medium density and viscosity and the thermal conductivity coefficient are independent of temperature and that the temperature distribution does not affect the flow field. This allows one to analyze the hydrodynamic problem about the fluid motion and the heat problem of finding the temperature field independently. It is assumed that the information about the fluid velocity field necessary for the solution of the heat problem is known.

The equation of heat transfer in a moving medium is similar to the convective diffusion equation (E5.1.1.1) and has the form

$$\frac{\partial T_*}{\partial t} + V_X\frac{\partial T_*}{\partial X} + V_Y\frac{\partial T_*}{\partial Y} + V_Z\frac{\partial T_*}{\partial Z} = \chi\left(\frac{\partial^2 T_*}{\partial X^2} + \frac{\partial^2 T_*}{\partial Y^2} + \frac{\partial^2 T_*}{\partial Z^2}\right), \qquad \text{(E5.1.1.10)}$$

where T_* is temperature and χ is the thermal diffusivity. The dissipative heating of fluid is neglected in Eq. (E5.1.1.10).

For nonstationary problems, the temperature distribution in the flow at the initial instant of time must be given.

Far from the surface, one usually has the condition that temperature is constant in the flow region:

$$\xi_* \to \infty, \quad T_* \to T_{\rm i}. \qquad \text{(E5.1.1.11)}$$

For body–medium heat exchange problems, if constant temperature is maintained on the body surface, the second boundary condition is

$$\xi_* = 0, \quad T_* = T_{\rm s}. \qquad \text{(E5.1.1.12)}$$

By using the new dimensionless variables

$$\begin{gathered}\bar\tau = \frac{\chi t}{a^2}, \quad x = \frac{X}{a}, \quad y = \frac{Y}{a}, \quad z = \frac{Z}{a}, \quad \mathsf{Pe}_T = \frac{aU}{\chi},\\ v_x = \frac{V_X}{U}, \quad v_y = \frac{V_Y}{U}, \quad v_z = \frac{V_Z}{U}, \quad T = \frac{T_{\rm i} - T_*}{T_{\rm i} - T_{\rm s}},\end{gathered} \qquad \text{(E5.1.1.13)}$$

one can rewrite Eq. (E5.1.1.10) and the boundary conditions (E5.1.1.11) and (E5.1.1.12) in the form

$$\frac{\partial T}{\partial \bar\tau} + \mathsf{Pe}_T\,(\mathbf{v}\cdot\nabla)\,T = \Delta T; \qquad \text{(E5.1.1.14)}$$

$$\xi \to \infty, \quad T \to 0; \qquad \xi = 0, \quad T = 1. \qquad \text{(E5.1.1.15)}$$

Obviously, from the mathematical viewpoint, the problem for equation (E5.1.1.14) with conditions (E5.1.1.15) describing the body–medium heat exchange is identical to problem (E5.1.1.6)–(E5.1.1.8) describing the flow–particle mass exchange in the case of a diffusion regime of reaction on the particle surface. By virtue of this analogy, in what follows we usually consider only mass transfer problems.

E5.1.2. Basic Dimensionless Parameters. Diffusion Fluxes and the Sherwood Number

► **Basic dimensionless parameters.** The diffusion and thermal Peclet numbers in the convective mass and heat transfer equations (E5.1.1.6) and (E5.1.1.14) are related to the Reynolds number $\mathsf{Re} = aU/\nu$ (where ν is the kinematic viscosity of the fluid) on the right-hand side in the Navier–Stokes equations (1.1.12) by the formulas

$$\mathsf{Pe} = \mathsf{Re}\,\mathsf{Sc}, \qquad \mathsf{Pe}_T = \mathsf{Re}\,\mathsf{Pr}. \tag{E5.1.2.1}$$

Here the Schmidt number $\mathsf{Sc} = \nu/D$ and the Prandtl number $\mathsf{Pr} = \nu/\chi$ are dimensionless values depending only on the physical properties of the continuous phase.

For ordinary gases, the diffusion coefficient and the kinematic viscosity are of the same order of magnitude, which corresponds to Schmidt numbers of the order of one ($\mathsf{Sc} \sim 1$).

For ordinary liquids like water, the kinematic viscosity is several orders of magnitude larger than the diffusion coefficient ($\mathsf{Sc} \sim 10^3$). In extremely viscous liquids like glycerin, the Schmidt number is of the order of 10^6.

The range of the Prandtl number is narrower than that of the Schmidt number. In gases such as air, $\mathsf{Pr} \sim 1$, and in liquids like water, $\mathsf{Pr} \sim 10$. In extremely viscous liquids like glycerin, the Prandtl number is of the order of 10^3. Liquid metals (sodium, lithium, mercury, etc.) are characterized by low Prandtl numbers: $5\times10^{-3} \le \mathsf{Pr} \le 5\times10^{-2}$.

The Reynolds number $\mathsf{Re} = aU/\nu$ is not a physical constant of a medium and depends on geometric and kinematic factors. Therefore, the range of variation of this number may be arbitrary.

It follows from the considered examples and relation (E5.1.2.1) that the Peclet numbers in problems of physicochemical hydrodynamics can vary in a wide range.

Since diffusion processes in fluids are characterized by very large Schmidt numbers, we point out that the Peclet number in convective mass transfer problems in fluid media is also large even for low Reynolds numbers at which the Stokes flow law applies (for a creeping flow).

► **Local and total diffusion fluxes and the Sherwood number.** A local (or differential) diffusion flux of a solute to the surface is determined by

$$j_* = D\rho\left(\frac{\partial C}{\partial \xi_*}\right)_{\xi_*=0} \tag{E5.1.2.2}$$

and in general varies along the surface.

The total (or integral) diffusion flux can be obtained by integrating (E5.1.2.2) over the entire surface S,

$$I_* = \iint_S j_*\,dS. \tag{E5.1.2.3}$$

The total diffusion flux I_* is the amount of substance that reacts on the entire surface per unit time.

In mass transfer problems with the boundary conditions (E5.1.1.2) and (E5.1.1.3), one often replaces (E5.1.2.2) and (E5.1.2.3) by the dimensionless diffusion fluxes

$$j = \frac{aj_*}{D\rho\,(C_\mathrm{i} - C_\mathrm{s})}, \qquad I = \frac{I_*}{aD\rho\,(C_\mathrm{i} - C_\mathrm{s})}. \tag{E5.1.2.4}$$

For the diffusion regime of reaction, which corresponds to the boundary conditions (E5.1.1.2) and (E5.1.1.4), as well as for the finite rate of surface chemical reaction in the case of the boundary conditions (E5.1.1.2) and (E5.1.1.5), one should set $C_{\rm s} = 0$ in (E5.1.2.4).

The main quantity of practical interest is the mean Sherwood number

$$\mathsf{Sh} = \frac{I}{S}, \tag{E5.1.2.5}$$

where $S = S_*/a^2$ is the dimensionless surface area and S_* is the corresponding dimensional area.

The calculation of the diffusion fluxes and the mean Sherwood number is usually carried out in three steps. First, the convective mass transfer problem is solved and the concentration field is determined. Second, the normal derivative $\bigl(\partial C/\partial\xi_*\bigr)_{\xi_*=0}$ on the surface is evaluated. Finally, one applies formulas (E5.1.2.2)–(E5.1.2.5).

For brevity, throughout the book (where it cannot lead to a misunderstanding) we use the terms “concentration” and “diffusion flux” instead of “dimensionless concentration” and “dimensionless diffusion flux.”

E5.1.3. Mass and Heat Transfer in Turbulent Flows

► **Equation of mass and heat transfer.** When mathematically describing the transfer of a passive admixture in a turbulent flow, the admixture concentration and the fluid velocity components are represented as

$$C = \overline{C} + C', \quad V_i = \overline{V}_i + V_i', \tag{E5.1.3.1}$$

where the bar and prime denote the time-average and fluctuating components, respectively. The averages of the fluctuations are zero, $\overline{C'} = \overline{V_i'} = 0$.

The introduction of the average and fluctuating components by relations (E5.1.3.1) followed by the application of the averaging operation made it possible, under some assumptions, to obtain the following equation for the average concentration of the admixture from Eq. (E5.1.1.1):

$$\frac{\partial\overline{C}}{\partial t} + (\overline{\mathbf{V}}\cdot\nabla)\overline{C} = D\Delta\overline{C} + \nabla\cdot(-\overline{C'\mathbf{V}'}). \tag{E5.1.3.2}$$

Here $-\overline{C'\mathbf{V}'}$ is the turbulent (eddy) flux of the admixture. This vector can be determined using some closure hypothesis, based, as a rule, on empirical information.

By analogy with Boussinesq’s ideas about transfer of momentum, it is often assumed for transfer of a substance that

$$-\overline{C'\mathbf{V}'} = \mathbf{D}_{\rm t}\cdot\nabla\overline{C}, \tag{E5.1.3.3}$$

where $\mathbf{D}_{\rm t}$ is the turbulent diffusion tensor.

► **Turbulent Prandtl number.** In a number of problems, it is important to know, as a rule, only one component of turbulent transfer, namely, the component normal to the wall. In this case, the turbulent diffusion strength is characterized by a single scalar quantity $D_{\rm t}$, just as the intensity of turbulent transfer of momentum is characterized by a single scalar quantity $\nu_{\rm t}$. The ratio of the turbulent viscosity coefficient to the turbulent diffusion coefficient,

$$\mathsf{Pr}_{\rm t} = \frac{\nu_{\rm t}}{D_{\rm t}}, \tag{E5.1.3.4}$$

is referred to as the turbulent Prandtl (Schmidt) number.

The simplest way to close equation (E5.1.3.2) is to use the hypothesis that the turbulent Prandtl number for the process examined is constant. Then it readily follows from formula (E5.1.3.4) that the turbulent diffusion coefficient is proportional to the turbulent viscosity, $D_\mathrm{t} = \nu_\mathrm{t}/\mathsf{Pr}_\mathrm{t}$. By using the expression for ν_t borrowed from the corresponding hydrodynamic model, one can obtain the desired value of D_t. In particular, following Prandtl's or von Kármán's model, one can use formula (1.1.21) or (1.1.22) for ν_t.

For a long time, it was assumed without sufficient justification that $\mathsf{Pr}_\mathrm{t} = 1$. With this value of Pr_t, the mechanism of turbulent transfer of momentum turns out to be identical to that of any passive scalar substance (the Reynolds analogy). According to up-to-date knowledge, the Reynolds analogy can be used in some cases for rough estimation of parameters of real flows, but the scope of the predicted results is highly restricted. The turbulent Prandtl number, as well as ν_t and D_t, depends on physical, geometric, and kinematic properties of the turbulent flow. For wall flows, numerous experimental measurements available give the value

$$\mathsf{Pr}_\mathrm{t} \approx 0.85. \qquad \text{(E5.1.3.5)}$$

For jet flows and mixing layers, there are various estimates for Pr_t ranging from 0.5 to 0.75. These estimates are helpful for approximate calculations of turbulent heat and mass transfer and can be used for closing equation (E5.1.3.2).

E5.2. Mass Transfer in Liquid Films

E5.2.1. Mass Exchange Between Gases and Liquid Films

► **Statement of the problem.** Dissolution of gases in flowing liquid films is one of the most important methods for dissolving gases, which is widely used in technology. Film absorbers with irrigated walls are used for obtaining water solutions of gases (e.g., absorption of HCl vapor by water), for separating gas mixtures (e.g., absorption of benzene in the cake and by-product process), for purifying gases from detrimental effluents (e.g., purification of coke-oven from H_2S), etc.

Let us consider the absorption of weakly soluble gases on the free surface of a liquid film in a laminar flow on an inclined plane. It follows from the results of Section E4.3 that for moderate velocities of motion, the steady-state distribution of the velocity inside the film has the form of a semiparabola with maximum velocity U_max on the free surface, which is one and a half the mean flow velocity $\langle V\rangle$:

$$U_\mathrm{max} = \frac{3}{2}\langle V\rangle = \frac{gh^2}{2\nu}\sin\alpha.$$

Here g is the gravitational acceleration, α the angle between the plane and the horizon, and h the film thickness, given by the expression

$$h = \left(\frac{3\nu^2}{g}\,\mathsf{Re}\right)^{1/3},$$

where $\mathsf{Re} = Q/\nu$ is the Reynolds number and Q is the irrigation density (that is, the volume rate flow of the liquid per unit width of the film).

The liquid velocity inside the film has a parabolic profile and is given by the formula

$$V = U_\mathrm{max}(1-y^2), \qquad y = Y/h,$$

where the Y-axis is normal to the film surface (Figure E4.3).

Assume that in the cross-section $X = 0$ the fluid flow contacts with a gas, so that a constant concentration $C = C_\mathrm{s}$ of the absorbed component is attained on the free boundary ($Y = 0$), while no solute is contained in the irrigating liquid. Furthermore, we assume that there is no flow across the wall. We restrict ourselves to the case of high Peclet numbers, so that one can ignore the molecular diffusion along the film.

It follows from these assumptions that the concentration distribution inside the film is described by the equation and the boundary conditions

$$(1-y^2)\frac{\partial c}{\partial x} = \frac{1}{\mathsf{Pe}}\frac{\partial^2 c}{\partial y^2}; \tag{E5.2.1.1}$$

$$x = 0, \quad c = 0 \qquad (0 \le y \le 1); \tag{E5.2.1.2}$$

$$y = 0, \quad c = 1 \qquad (x > 0); \tag{E5.2.1.3}$$

$$y = 1, \quad \partial c/\partial y = 0 \quad (x > 0), \tag{E5.2.1.4}$$

where the following dimensionless variables have been used:

$$x = \frac{X}{h}, \quad y = \frac{Y}{h}, \quad c = \frac{C}{C_\mathrm{s}}, \quad \mathsf{Pe} = \frac{hU_\mathrm{max}}{D}. \tag{E5.2.1.5}$$

Note that for small $x \le O(\mathsf{Pe}^{-1/2})$, near the input section one should consider the complete mass transfer equation with the term $\partial^2 c/\partial y^2$ replaced by Δc.

► **Diffusion boundary layer approximation.** For $x = O(1)$, the concentration mostly varies on the initial interval in a thin diffusion boundary layer near the free boundary of the film. In this region, we expand the transverse coordinate according to the rule

$$y = \zeta/\sqrt{\mathsf{Pe}}. \tag{E5.2.1.6}$$

By substituting (E5.2.1.6) into (E5.2.1.1) and by passing to the limit as $\mathsf{Pe} \to \infty$ (it is assumed that the variables x and ζ and the corresponding derivatives are of the order of 1), we obtain the equation

$$\frac{\partial c}{\partial x} = \frac{\partial^2 c}{\partial \zeta^2}. \tag{E5.2.1.7}$$

In view of (E5.2.1.6), the distance to the wall determined by the coordinate $y = 1$ corresponds to $\zeta = \sqrt{\mathsf{Pe}}$. Therefore, as $\mathsf{Pe} \to \infty$, the value $y = 1$ in the boundary condition (E5.2.1.4) corresponds to $\zeta \to \infty$. This implies that we can rewrite the boundary conditions (E5.2.1.2)–(E5.2.1.4) as follows:

$$x = 0, \quad c = 0; \qquad \zeta = 0, \quad c = 1; \qquad \zeta \to \infty, \quad \partial c/\partial\zeta \to 0. \tag{E5.2.1.8}$$

The solution of problem (E5.2.1.7), (E5.2.1.8) is given by the formula

$$c = \operatorname{erfc}\left(\frac{\zeta}{2\sqrt{x}}\right), \tag{E5.2.1.9}$$

where $\operatorname{erfc} z = \dfrac{2}{\sqrt{\pi}}\displaystyle\int_z^\infty \exp(-t^2)\,dt$ is the complementary error function.

By differentiating (E5.2.1.9), we obtain the dimensionless local diffusion flux to the film surface:

$$j = -\left(\frac{\partial c}{\partial y}\right)_{y=0} = \left(\frac{\mathsf{Pe}}{\pi x}\right)^{1/2}. \tag{E5.2.1.10}$$

TABLE E5.1
The eigenvalues λ_m and the coefficients A_m in the expansion of the solution (E5.2.1.12) for the concentration distribution inside the film with gas absorption on the film surface.

m	λ_m	A_m	m	λ_m	A_m
1	2.2631	1.3382	6	22.3181	–0.1873
2	6.2977	–0.5455	7	26.3197	0.1631
3	10.3077	0.3589	8	30.3209	–0.1449
4	14.3128	–0.2721	9	34.3219	0.1306
5	18.3159	0.2211	10	38.3227	–0.1191

The dimensionless total diffusion flux on the part of the film lying in the interval from 0 to x is equal to

$$I=\int_0^x j\,dx=2\left(\frac{\mathsf{Pe}}{\pi}x\right)^{1/2}. \qquad \text{(E5.2.1.11)}$$

Formulas (E5.2.1.10) and (E5.2.1.11) cannot be used for sufficiently large x, where the diffusion boundary layer “grows” through the entire film. To estimate the scope of these formulas, consider the original problem (E5.2.1.1)–(E5.2.1.4).

► **Exact solution.** For $0\le x<\infty$, the exact solution of problem (E5.2.1.1)–(E5.2.1.4) can be represented in the form of the series

$$c=1-\sum_{m=0}^{\infty}A_m\exp\left(-\frac{\lambda_m^2}{\mathsf{Pe}}x\right)H_m(y), \qquad \text{(E5.2.1.12)}$$

where the functions $H_m(y)$ and the coefficients A_m and λ_m are independent of the Peclet number. The functions $H_m(y)$ are defined by the formulas

$$H_m(y)=y\exp\left(-\tfrac{1}{2}\lambda_m y^2\right)\Phi\left(a_m+\tfrac{1}{2},\,\tfrac{3}{2};\,\lambda_m y^2\right),\quad a_m=\tfrac{1}{4}(1-\lambda_m), \qquad \text{(E5.2.1.13)}$$

where $\Phi(a,b,\xi)=1+\sum\limits_{m=1}^{\infty}\dfrac{a(a+1)\dots(a+m-1)}{b(b+1)\dots(b+m-1)}\dfrac{\xi^m}{m!}$ is the degenerate hypergeometric function. The eigenvalues λ_m can be found by solving the transcendental equation

$$\lambda_m\Phi\left(a_m+\tfrac{1}{2},\,\tfrac{3}{2};\,\lambda_m\right)-\Phi\left(a_m+\tfrac{1}{2},\,\tfrac{1}{2};\,\lambda_m\right)=0,$$

and the coefficients A_m are computed by the formulas

$$A_m=\frac{\displaystyle\int_0^1(1-y^2)H_m(y)\,dy}{\displaystyle\int_0^1(1-y^2)H_m^2(y)\,dy},\qquad \text{where}\quad m=1,\,2,\,\dots \qquad \text{(E5.2.1.14)}$$

Table E5.1 contains the first ten eigenvalues λ_m and the coefficients A_m calculated by Rotem & Neilson (1966).

Taking into account (E5.2.1.12), we obtain the dimensional total diffusion flux to the part of the film boundary from 0 to x in the form

$$I=-\int_0^x\left(\frac{\partial c}{\partial y}\right)_{y=0}dx=\mathsf{Pe}\sum_{m=1}^{\infty}\frac{A_m}{\lambda_m^2}\left(\frac{dH_m}{dy}\right)_{y=0}\left[1-\exp\left(-\frac{\lambda_m^2}{\mathsf{Pe}}x\right)\right]. \qquad \text{(E5.2.1.15)}$$

The substitution of the function (E5.2.1.13) into (E5.2.1.15) yields

$$I = \mathsf{Pe}\sum_{m=1}^{\infty}\frac{A_m}{\lambda_m^2}\left[1 - \exp\left(-\frac{\lambda_m^2}{\mathsf{Pe}}\,x\right)\right].$$

The comparison of this formula with (E5.2.1.11) shows that the diffusion boundary layer approximation can be used in the region $x \le 0.1\,\mathsf{Pe}$.

E5.2.2. Dissolution of a Plate by a Laminar Liquid Film

► **Statement of the problem.** Now consider mass transfer from a solid wall to a liquid film at high Peclet numbers. Such a problem is of serious interest in dissolution, crystallization, corrosion, anodic dissolution of metals in some electrochemical processes, etc. In many practical cases, dissolution processes are rather rapid compared with diffusion. Therefore, we assume that the concentration on the plate surface is equal to the constant C_s and the incoming liquid is pure. As before, we introduce dimensionless variables according to formulas (E5.2.1.5). In this case, the convective mass transfer in the liquid film is described by Eq. (E5.2.1.1), the boundary condition (E5.2.1.2) imposed on the longitudinal variable x, and the following boundary conditions with respect to the transverse coordinate:

$$y = 0, \quad \frac{\partial c}{\partial y} = 0 \quad (x > 0); \tag{E5.2.2.1}$$

$$y = 1, \quad c = 1 \quad (x > 0). \tag{E5.2.2.2}$$

Although this problem differs from the previously studied problem (E5.2.1.1)–(E5.2.1.4) only by a rearrangement of the boundary conditions (E5.2.1.3) and (E5.2.1.4), there is a substantial difference in the solutions of these problems.

► **Diffusion boundary layer approximation.** The concentration mostly varies on the initial interval $x = O(1)$, that is, in the diffusion boundary layer near the film surface. In this region, the asymptotic solution can be obtained by substituting the expanded coordinate

$$\xi = (1 - y)\,\mathsf{Pe}^{1/3} \tag{E5.2.2.3}$$

into Eq. (E5.2.1.1) and by taking the leading term of the concentration expansion as $\mathsf{Pe} \to \infty$. As a result, we arrive at the diffusion boundary layer equation

$$2\xi\frac{\partial c}{\partial x} = \frac{\partial^2 c}{\partial \xi^2}. \tag{E5.2.2.4}$$

By analogy with the already considered problem about gas absorption on the film surface, using (E5.2.1.2), (E5.2.2.1), and (E5.2.2.2), we obtain the boundary conditions for Eq. (E5.2.2.4), which coincide with (E5.2.1.8) up to the replacement $\xi \to \zeta$. The corresponding problem has the solution

$$c = \frac{1}{\Gamma(1/3)}\Gamma\left(\frac{1}{3}, \frac{2\xi^3}{9x}\right), \tag{E5.2.2.5}$$

where $\Gamma(1/3, z)$ is the incomplete gamma function (see Section M13.4).

By differentiating (E5.2.2.5), we calculate the dimensionless local diffusion flux

$$j = \frac{6^{1/3}}{\Gamma(1/3)} \frac{\mathsf{Pe}^{1/3}}{x^{1/3}} \approx 0.678 \frac{\mathsf{Pe}^{1/3}}{x^{1/3}}. \tag{E5.2.2.6}$$

The corresponding dimensionless total diffusion flux to the film boundary is

$$I = \int_0^x j\,dx = 1.02\,\mathsf{Pe}^{1/3} x^{2/3}. \tag{E5.2.2.7}$$

By comparing formulas (E5.2.1.9)–(E5.2.1.11) with (E5.2.2.5)–(E5.2.2.7), we see that for $x \sim 1$ the thickness of the diffusion boundary layer near the free surface of the film, $\delta_0 \sim \mathsf{Pe}^{-1/2}$, is considerably less than that of the boundary layer near the solid surface, $\delta_{\rm sol} \sim \mathsf{Pe}^{-1/3}$. Accordingly, the diffusion flux to the free surface is larger than that to the solid surface. Moreover, the diffusion flux decreases more rapidly on the free surface than on the solid boundary with the increase of the distance from the input cross-section. These effects are due to the fact that the fluid moves much more rapidly near the free surface than near the solid boundary, where the no-slip condition is satisfied.

All facts established for a fluid film remain valid for a majority of problems on the diffusion boundary layer. Namely, near a gas–fluid or fluid–fluid interface, the dimensionless thickness of the layer is proportional to $\mathsf{Pe}^{-1/2}$ (for the diffusion flux we have $j \sim \mathsf{Pe}^{1/2}$), and near the fluid–solid interface the thickness of the boundary layer is proportional to $\mathsf{Pe}^{-1/3}$ (the diffusion flux is $j \sim \mathsf{Pe}^{1/3}$).

► **Exact solution.** The exact solution of the problem for equation (E5.2.1.1) with conditions (E5.2.1.2), (E5.2.2.1), and (E5.2.2.2) in the entire region $0 \le x < \infty$ can be represented in the form of the series (E5.2.1.12), where the eigenfunctions $H_m(y)$ are expressed via the hypergeometric functions by the formulas

$$H_m(y) = \exp\left(-\tfrac{1}{2}\lambda_m y^2\right) \Phi\left(\tfrac{1}{4} - \tfrac{1}{4}\lambda_m,\ \tfrac{1}{2};\ \lambda_m y^2\right) \tag{E5.2.2.8}$$

and the transcendental equation for the eigenvalues λ_m is

$$\Phi\left(a_m,\ \tfrac{1}{2};\ \lambda_m\right) = 0, \qquad \text{where} \quad a_m = \tfrac{1}{4} - \tfrac{1}{4}\lambda_m. \tag{E5.2.2.9}$$

The roots of Eq. (E5.2.2.9) are positive and increase monotonically, so that $\lambda_m \to \infty$ as $m \to \infty$. To determine the eigenvalues λ_m, one can use the approximate formula

$$\lambda_m = 4m + 1.68 \qquad (m = 0,\ 1,\ 2,\ \dots), \tag{E5.2.2.10}$$

whose error is less than 0.2%.

As before, the coefficients A_m in the series (E5.2.1.12) are determined by (E5.2.1.14), where the eigenfunctions $H_m(y)$ are given in (E5.2.2.8). To compute the coefficients A_m, one can use the approximate formulas

$$A_0 = 1.2; \quad A_m = 2.27\,(-1)^m \lambda_m^{-7/6} \qquad \text{for} \quad m = 1,\ 2,\ 3,\ \dots, \tag{E5.2.2.11}$$

where the λ_m are given in (E5.2.2.10). The maximum error of (E5.2.2.11) is less than 0.1%.

The total diffusion flux on the film surface is given by (E5.2.1.15) with $(dH_m/dy)_{y=0}$ replaced by $(dH_m/dy)_{y=1}$. One can use the expressions (E5.2.2.8), (E5.2.2.10), and (E5.2.2.11) for calculating the eigenfunctions $H_m(y)$ and the numbers λ_m and A_m.

E5.3. Heat Transfer to a Flat Plate

E5.3.1. Heat Transfer in Laminar Flow

► **Statement of the problem. Thermal boundary layer.** Consider heat transfer to a flat plate in a longitudinal translational flow of a viscous incompressible fluid with velocity U_i at high Reynolds numbers. We assume that the temperature on the plate surface and far from it is equal to the constants T_s and T_i, respectively. The origin of the rectangular coordinates X, Y is at the front edge of the plate, the X-axis is tangent to the plate, and the Y-axis is normal to it.

Numerous experiments and numerical calculations show that the laminar hydrodynamic boundary layer occurs for $5\times10^2 \le \mathsf{Re}_X \le 5\times10^5$ to 10^6. In this region, the thermal Peclet number $\mathsf{Pe}_X = \mathsf{Pr}\,\mathsf{Re}_X$ is large for gases and common liquids. For liquid metals, there is a range of Reynolds numbers, $10^4 \le \mathsf{Re}_X \le 10^6$, where the Peclet numbers are also large.

Taking into account the previous discussion, we restrict ourselves to the case of high Peclet numbers, for which the longitudinal molecular heat transfer may be neglected. The corresponding thermal boundary layer equations and the boundary conditions have the form

$$\begin{gathered} v_x\frac{\partial T}{\partial x} + v_y\frac{\partial T}{\partial y} = \frac{1}{\mathsf{Pr}}\frac{\partial^2 T}{\partial y^2}; \\ x = 0,\quad T = 0; \qquad y = 0,\quad T = 1; \qquad y \to \infty,\quad T \to 0. \end{gathered} \tag{E5.3.1.1}$$

Here the dimensionless variables are introduced by formulas (E5.1.1.13) with characteristic length $a = \nu/U_\mathrm{i}$ and ν is the kinematic viscosity of the fluid.

The fluid velocity in (E5.3.1.1) is given by the Blasius solution

$$v_x = f'(\eta),\quad v_y = \frac{\eta f' - f}{2\sqrt{x}}, \qquad \text{where}\quad \eta = \frac{y}{\sqrt{x}}. \tag{E5.3.1.2}$$

The function $f = f(\eta)$ was previously described in Section E4.5.1, and the prime stands for the derivative with respect to η.

► **Temperature field.** The solution of problem (E5.3.1.1) in conjunction with relations (E5.3.1.2) will be sought in the self-similar form $T = T(\eta)$. As a result, we obtain the ordinary differential equation

$$\begin{gathered} \frac{d^2T}{d\eta^2} + \frac{1}{2}\mathsf{Pr} f(\eta)\frac{dT}{d\eta} = 0; \\ \eta = 0,\quad T = 1; \qquad \eta \to \infty,\quad T \to 0. \end{gathered} \tag{E5.3.1.3}$$

By taking into account the relation $f = -f'''/f''$, which follows from the equation for f, we can write out the solution of problem (E5.3.1.3) as follows:

$$T = \frac{\displaystyle\int_\eta^\infty \bigl[f''(\eta)\bigr]^{\mathsf{Pr}}\,d\eta}{\displaystyle\int_0^\infty \bigl[f''(\eta)\bigr]^{\mathsf{Pr}}\,d\eta}. \tag{E5.3.1.4}$$

For $\mathsf{Pr} = 1$, this formula implies the simple relation $T(\eta) = 1 - f'(\eta) = 1 - v_x$ between temperature and the longitudinal velocity.

► **Heat flux and the Nusselt number.** By differentiating (E5.3.1.4) and by using the numerical value $f''(0) = 0.332$, we obtain the dimensionless local heat flux to the plate surface:

$$j_T = -\left(\frac{\partial T}{\partial y}\right)_{y=0} = \frac{\varphi(\mathsf{Pr})}{\sqrt{x}}, \quad \text{where} \quad \varphi(\mathsf{Pr}) = \frac{(0.332)^{\mathsf{Pr}}}{\displaystyle\int_0^\infty \left[f''(\eta)\right]^{\mathsf{Pr}} d\eta}. \tag{E5.3.1.5}$$

It is convenient to seek the asymptotics of the function $\varphi(\mathsf{Pr})$ at low and high Prandtl numbers starting from Eq. (E5.3.1.3) with the extended variable $\eta = \zeta/\mathsf{Pr}$. As a result, we obtain the equation $T''_{\zeta\zeta} + f(\zeta/\mathsf{Pr})T'_\zeta = 0$. As $\mathsf{Pr} \to 0$, the argument of the function $f(\zeta/\mathsf{Pr})$ tends to infinity, which corresponds to a constant velocity inside the thermal boundary layer and $f(\eta) \approx \eta$. In the other limit case as $\mathsf{Pr} \to \infty$, the argument of the function $f(\zeta/\mathsf{Pr})$ tends to zero, which corresponds to the linear approximation of the velocity inside the boundary layer and $f(\eta) \approx 0.166\,\eta^2$. By substituting the above-mentioned leading terms of the asymptotic expansion of f into Eq. (E5.3.1.3) and by solving the corresponding problems, we obtain the following expressions for the heat flux (E5.3.1.5):

$$\begin{aligned} &\varphi(\mathsf{Pr}) \to \left(\mathsf{Pr}/\pi\right)^{1/2} \qquad (\mathsf{Pr} \to 0),\\ &\varphi(\mathsf{Pr}) \to 0.339\,\mathsf{Pr}^{1/3} \quad (\mathsf{Pr} \to \infty). \end{aligned} \tag{E5.3.1.6}$$

Both considered limit situations occur in numerous problems of convective heat transfer. In the case $\mathsf{Pr} \to 0$, which approximately takes place for liquid metals (e.g., mercury), one can neglect the dynamic boundary layer in the calculation of the temperature boundary layer and replace the velocity profile $v(x, y)$ by the velocity $v_\infty(x)$ of the inviscid outer flow. As $\mathsf{Pr} \to \infty$, which corresponds to the case of strongly viscous fluids (e.g., glycerin), the temperature boundary layer is very thin and lies inside the dynamic boundary layer, where the velocity increases linearly with distance from the plate surface.

In the entire range of the Prandtl number, the function $\varphi(\mathsf{Pr})$ in formula (E5.3.1.5) is well approximated by the expression

$$\varphi(\mathsf{Pr}) = 0.0817\left[(1 + 72\,\mathsf{Pr})^{2/3} - 1\right]^{1/2}, \tag{E5.3.1.7}$$

whose maximum error is about 0.5%.

Let us write out the local Nusselt number:

$$\mathsf{Nu}_X = -\frac{X}{T_\mathrm{s} - T_\mathrm{i}}\left(\frac{\partial T_*}{\partial Y}\right)_{Y=0} = \sqrt{\mathsf{Re}_X}\,\varphi(\mathsf{Pr}), \tag{E5.3.1.8}$$

where $\mathsf{Re}_X = XU_\mathrm{i}/\nu$ is the local Reynolds number.

E5.3.2. Heat Transfer in Turbulent Flow

► **Temperature profile.** Let us discuss qualitative specific features of convective heat and mass transfer in turbulent flow past a flat plate. Experimental evidence indicates that several characteristic regions with different temperature profiles can be distinguished in the thermal boundary layer on a flat plate. At moderate Prandtl numbers ($0.5 \le \mathsf{Pr} \le 2.0$), it can be assumed for rough estimates that the characteristic sizes of these regions are of the same order of magnitude as those of the wall layer and the core of the turbulent stream, see Section E4.5.2.

For the description of heat and mass transfer in the wall layer, it is common to introduce the friction velocity U_*, friction temperature Θ_*, and dimensionless internal coordinate y^+ by the relations

$$U_* = \sqrt{\frac{\tau_\text{s}}{\rho}}, \quad \Theta_* = \frac{q_\text{s}}{\rho c_p U_*}, \quad y^+ = \frac{YU_*}{\nu}, \tag{E5.3.1.9}$$

where τ_s is the shear stress at the wall, ρ the fluid density, q_s the heat flux at the wall, c_p the specific heat of the fluid, ν the kinematic viscosity, and Y the distance to the plate surface.

In the molecular thermal conduction layer, adjacent to the plate surface, the deviation of the average temperature $\overline{T}$ from the wall temperature T_s depends linearly on the transverse coordinate:

$$\frac{T_\text{s} - \overline{T}(Y)}{\Theta_*} = \mathsf{Pr}\, y^+. \tag{E5.3.1.10}$$

For $\mathsf{Pr} > 1$, the linear law (E5.3.1.10) is satisfied for $0 \le y^+ \le 9\,\mathsf{Pr}^{-1/3}$.

In the logarithmic layer, the average temperature can be determined with the use of the relations*

$$\begin{aligned} &\frac{T_\text{s} - \overline{T}(Y)}{\Theta_*} = 2.12 \ln y^+ + \beta(\mathsf{Pr}), \\ &\beta(\mathsf{Pr}) = (3.85\,\mathsf{Pr}^{1/3} - 1.3)^2 + 2.12 \ln \mathsf{Pr}. \end{aligned} \tag{E5.3.1.11}$$

These formulas are valid within a wide range of Prandtl numbers, $6 \times 10^{-3} \le \mathsf{Pr} \le 10^4$, for $y^+ \ge y^+_\text{log}$. The lower bound of the logarithmic layer, y^+_log, depends on the Prandtl number as follows:

$$y^+_\text{log} \simeq \begin{cases} 12\,\mathsf{Pr}^{-1/3} & \text{if } \mathsf{Pr} \gg 1 \quad \text{(for liquids)}, \\ 30 & \text{if } \mathsf{Pr} \sim 1 \quad \text{(for gases)}, \\ 2/\mathsf{Pr} & \text{if } \mathsf{Pr} \ll 1 \quad \text{(for liquid metals)}. \end{cases}$$

In the stream core and a major part of the logarithmic layer, the average temperature profile can be described by the single formula

$$\frac{\overline{T}(Y) - \overline{T}_\text{m}}{\Theta_*} = -2.12 \ln \eta + 1.5\,(1 - \eta), \tag{E5.3.1.12}$$

where $\overline{T}_\text{m}$ is the average temperature at the stream axis, $\eta = Y/\delta_T$ the dimensionless distance to the plate surface, and δ_T the thermal boundary layer thickness. The results predicted by Eq. (E5.3.1.12) fairly well agree with experimental data provided by various authors for turbulent flows of air, water, and transformer oil for $10^{-2} < Y/\delta_T < 1$.

The thermal boundary layer thickness can be estimated from the relation

$$\delta_T(X) = 0.45\, X \sqrt{\tfrac{1}{2} c_\text{f}}, \tag{E5.3.1.13}$$

where c_f is the local friction coefficient determined by formula (E4.5.2.2) or (E4.5.2.3). The empirical coefficient 0.45 was obtained by processing experimental data for $0.7 \le \mathsf{Pr} \le 64$. It is greater than the similar coefficient in relation (E4.5.2.1) for the hydrodynamic turbulent boundary layer thickness. (Qualitatively, this is accounted for by the fact that the turbulent Prandtl number is less than unity; that is, the hydrodynamic boundary layer lies within the thermal boundary layer.)

---

* The universal constant 2.12 multiplying the logarithm is defined as $\mathsf{Pr}_\text{t}/\kappa$, where $\mathsf{Pr}_\text{t} = 0.85$ is the turbulent Prandtl number for wall flows and $\kappa = 0.4$ the von Kármán constant.

► **Nusselt number.** For determining the local Nusselt number, it is recommended to use the Kader–Yaglom formula

$$\mathsf{Nu}_X = \frac{\mathsf{Pr}\,\mathsf{Re}_X(\frac{1}{2}c_\mathrm{f})^{1/2}}{2.12\ln(\frac{1}{2}c_\mathrm{f}\,\mathsf{Pr}\,\mathsf{Re}_X) + (3.85\,\mathsf{Pr}^{1/3} - 1.3)^2 + 1.5}, \tag{E5.3.1.14}$$

which quite well agrees with numerous experimental data within a wide range of Prandtl numbers ($0.5 \le \mathsf{Pr} \le 100$).

For rough estimates, the simpler formula

$$\mathsf{Nu}_X = 0.0288\,\mathsf{Pr}^{0.4}\mathsf{Re}_X^{0.8}$$

can be used, which is valid for $0.5 \le \mathsf{Pr} \le 50$.

More detailed information about heat and mass transfer in turbulent flows past a flat plate, as well as various relations for determining the temperature profile and Nusselt (Sherwood) numbers, and a lot of other useful information can be found in the references given at the end of this chapter.

E5.4. Convective Heat and Mass Transfer in Tubes and Channels

Many convective mass and heat transfer processes in chemical industry, petroleum chemistry, gas, nuclear, and other branches of industry occur in pipes (water, gas, and oil pipelines, heat exchangers, etc.).

Starting from the classical works by Graetz (1883) and Nusselt (1910), many authors considered the problem about the temperature distribution in a fluid moving in a tube under various assumptions on the type of flow, the tube shape, the form of boundary conditions, the value of the Peclet number, and some other simplifications. In this section, we outline the most important results obtained in this field.

E5.4.1. Heat and Mass Transfer in a Circular Tube with Constant Temperature of the Wall for Laminar Flows

► **Statement of the problem.** Consider a laminar steady-state fluid flow in a circular tube of radius a with Poiseuille velocity profile (see Section E4.4.1). We introduce cylindrical coordinates $\mathcal{R}$, Z with the Z-axis in the flow direction. We assume that for $Z > 0$ the temperature on the wall is equal to the constant T_2. In the entry area $Z < 0$, the temperature on the wall is also constant but takes another value T_1.

The convective heat transfer in a tube is described by the equation

$$\mathsf{Pe}_T(1-\varrho^2)\frac{\partial T}{\partial z} = \frac{\partial^2 T}{\partial \varrho^2} + \frac{1}{\varrho}\frac{\partial T}{\partial \varrho} + \frac{\partial^2 T}{\partial z^2} \tag{E5.4.1.1}$$

and the boundary conditions

$$\varrho = 0, \quad \frac{\partial T}{\partial \varrho} = 0; \qquad \varrho = 1, \quad T = \begin{cases} 0 & \text{for } z < 0; \\ 1 & \text{for } z > 0; \end{cases} \tag{E5.4.1.2}$$

$$z \to -\infty, \quad T \to 0; \qquad z \to \infty, \quad T \to 1. \tag{E5.4.1.3}$$

The following dimensionless variables have been used:

$$\varrho = \frac{\mathcal{R}}{a}, \quad z = \frac{Z}{a}, \quad T = \frac{T_* - T_1}{T_2 - T_1}, \quad \mathsf{Pe}_T = \frac{aU_{\max}}{\chi},$$

where T_* is the fluid temperature, χ the thermal diffusivity, $U_{\max} = a^2\Delta P/(4\mu L)$ the maximum velocity at the center of the tube, ΔP the pressure increment along the distance L, and μ the dynamic viscosity of the fluid.

► **High Peclet numbers (initial region).** As $\mathsf{Pe}_T \to \infty$, the fluid temperature in the region $z < 0$ is constant and equal to the temperature $T \approx 0$ on the wall. In the region $z > 0$ at $z = O(1)$, a thin boundary layer is being formed near the wall of the tube. In this region, on the left-hand side of Eq. (E5.4.1.1) one may retain only the leading term of the velocity expansion as $\varrho \to 1$ and write $v = 1 - \varrho^2 \approx 2\xi$, where $\xi = 1 - \varrho$. Moreover, in contrast with the first term on the right-hand side in (E5.4.1.1), one can neglect the last two terms; that is, $\Delta T \approx \partial^2 T/\partial\xi^2$. Thus, we arrive at an equation that coincides with (E5.2.2.4) up to notation. Taking into account the boundary conditions (E5.4.1.2), we obtain the temperature distribution in the boundary layer,

$$T = \frac{1}{\Gamma(1/3)}\Gamma\left(\frac{1}{3}, \frac{2\,\mathsf{Pe}_T(1-\varrho)^3}{9z}\right), \qquad \text{(E5.4.1.4)}$$

where $\Gamma(1/3, z)$ is the incomplete gamma function (see Section M13.4).

The corresponding dimensionless local and total heat fluxes have the form

$$j_T = -\left(\frac{\partial T}{\partial \varrho}\right)_{\varrho=1} = \frac{1}{\Gamma(1/3)}\left(\frac{6\,\mathsf{Pe}_T}{z}\right)^{1/3}, \qquad \text{(E5.4.1.5)}$$

$$I_T = \int_0^z j_T\,dz = \frac{3(6\,\mathsf{Pe}_T)^{1/3}}{2\Gamma(1/3)} z^{2/3}. \qquad \text{(E5.4.1.6)}$$

The scope of formulas (E5.4.1.4)–(E5.4.1.6) is bounded by the condition $z \ll \mathsf{Pe}_T$. This restriction practically always holds in the similar problem about the diffusion boundary layer.

► **Arbitrary Peclet numbers.** In the general case $0 \le \mathsf{Pe}_T < \infty$, we seek the fluid temperature distribution separately on both sides of the inlet cross-section of the tube in the form of series:

$$T = 1 - \sum_{m=0}^{\infty} A_m \exp\left(-\frac{\lambda_m^2}{\mathsf{Pe}_T} z\right) f_m(\varrho) \qquad \text{for} \quad z > 0, \qquad \text{(E5.4.1.7)}$$

$$T = \sum_{k=0}^{\infty} B_k \exp\left(\frac{\eta_k^2}{\mathsf{Pe}_T} z\right) g_k(\varrho) \qquad \text{for} \quad z < 0. \qquad \text{(E5.4.1.8)}$$

The eigenfunctions f_m and g_k and the eigenvalues λ_m and η_k are determined from two spectral problems that are obtained by substituting the expansions (E5.4.1.7) and (E5.4.1.8) into Eq. (E5.4.1.1) and the boundary conditions (E5.4.1.2) and (E5.4.1.3).

The temperature $T = T(\varrho, z)$ and its derivative must be continuous across the section $z = 0$:

$$T(\varrho, -0) = T(\varrho, +0); \quad \frac{\partial T}{\partial z}(\varrho, -0) = \frac{\partial T}{\partial z}(\varrho, +0). \qquad \text{(E5.4.1.9)}$$

The consistency conditions (E5.4.1.9) allow one to find the coefficients A_m and B_k in the series (E5.4.1.7) and (E5.4.1.8).

In what follows, we give the results of the analysis of this problem only for the region $z > 0$.

The eigenfunctions $f_m(\rho)$ are expressed via the degenerate hypergeometric function $\Phi(a, b; \xi)$ as

$$f_m(\rho) = \exp\left(-\tfrac{1}{2}\lambda_m \varrho^2\right) \Phi(a_m, 1; \lambda_m \varrho^2), \qquad \text{(E5.4.1.10)}$$

$$a_m = \frac{1}{2} - \frac{1}{4}\lambda_m - \frac{\lambda_m^3}{4\,\mathsf{Pe}_T^2},$$

and the eigenvalues λ_m satisfy the transcendental equation

$$\Phi(a_m, 1; \lambda_m) = 0. \qquad \text{(E5.4.1.11)}$$

The coefficients A_m can be computed by the formula

$$A_m = -\frac{2}{\lambda_m \left(\dfrac{\partial f}{\partial \lambda}\right)_{\varrho=1,\ \lambda=\lambda_m}} \qquad \text{(E5.4.1.12)}$$

with the function $f(\rho) = f(\rho; \lambda)$ determined by relations (E5.4.1.10) where the indices m are omitted.

Consider the asymptotic behavior of the eigenvalues λ_m and coefficients A_m in some limit cases.

As $\mathsf{Pe}_T \to 0$, we have

$$\lambda_m = (\gamma_m\,\mathsf{Pe}_T)^{1/2}, \quad A_m = -\big[\gamma_m J_1(\gamma_m)\big]^{-1}, \quad f_m = J_0(\gamma_m \varrho), \qquad \text{(E5.4.1.13)}$$

where $J_0 = J_0(\xi)$ and $J_1 = J_1(\xi)$ are the Bessel functions and the γ_m are the roots of J_0, $J_0(\gamma_m) = 0$.

It is convenient to calculate approximate values of γ_m by using the approximate expression

$$\gamma_m = 2.4 + 3.13\,m \qquad (m = 0,\ 1,\ 2,\ \dots), \qquad \text{(E5.4.1.14)}$$

whose maximum error is less than 0.2%.

As $\mathsf{Pe} \to \infty$, the asymptotics of the eigenvalues λ_m for large m has the form

$$\lambda_m = 4m + \tfrac{8}{3}. \qquad \text{(E5.4.1.15)}$$

This formula constructed under the assumption that $m \gg 1$ can be used for all m. The maximum error in (E5.4.1.15) is attained at $m = 0$ and is equal to 1.4%.

Instead of (E5.4.1.15), it is convenient to use the more precise formula

$$\lambda_m = 4m + 2.7 \qquad (m = 0,\ 1,\ 2,\ \dots), \qquad \text{(E5.4.1.16)}$$

whose maximum error is 0.3%.

The coefficients A_m of the series (E5.4.1.7) can be calculated by the formula

$$A_m = 2.85\,(-1)^m \lambda_m^{-2/3}, \qquad \text{(E5.4.1.17)}$$

whose maximum error is 0.5%.

For large but finite Peclet numbers, the expressions (E5.4.1.16) and (E5.4.1.17) can be used only for a bounded set of eigenvalues such that $\lambda_m \ll \mathsf{Pe}_T$.

The dependence of the ground eigenvalue λ_0 on the Peclet number is nicely approximated by the formula

$$\lambda_0 = 2.7\sqrt{\frac{\exp(0.27\,\mathsf{Pe}_T) - 1}{\exp(0.27\,\mathsf{Pe}_T) - 0.18}},$$

whose maximum error is about 1%.

► **Bulk temperature. Heat flux.** Taking into account the fact that the dimensionless fluid velocity distribution in a tube is $u(\varrho) = 1 - \varrho^2$, we obtain the bulk temperature in an arbitrary cross-section:

$$\langle T \rangle = \frac{\displaystyle\int_0^1 T u(\varrho) 2\pi \varrho \, d\varrho}{\displaystyle\int_0^1 u(\varrho) 2\pi \varrho \, d\varrho} = 4 \int_0^1 T(1 - \varrho^2) \varrho \, d\varrho.$$

By substituting the series (E5.4.1.7) into this formula, we obtain

$$\langle T \rangle = 1 - \sum_{m=0}^{\infty} E_m \exp\left(-\frac{\lambda_m^2}{\mathsf{Pe}_T} z\right), \qquad \text{(E5.4.1.18)}$$

where

$$E_m = 4 A_m \int_0^1 f_m (1 - \varrho^2) \varrho \, d\varrho.$$

By differentiating (E5.4.1.7), we arrive at the dimensionless local heat flux to the tube wall,

$$j_T = \left(\frac{\partial T}{\partial \varrho}\right)_{\varrho=1} = -\sum_{m=0}^{\infty} A_m f_m'(1) \exp\left(-\frac{\lambda_m^2}{\mathsf{Pe}_T} z\right), \qquad \text{(E5.4.1.19)}$$

where

$$f_m'(1) = 2 a_m \lambda_m \exp(-\tfrac{1}{2}\lambda_m) \Phi(a_m + 1,\, 2;\, \lambda_m).$$

► **Nusselt number.** The most practically important variable is the Nusselt number

$$\mathsf{Nu} = \frac{2 j_T}{1 - \langle T \rangle}. \qquad \text{(E5.4.1.20)}$$

Here $1 - \langle T \rangle$ is the temperature head equal to the difference between the wall temperature and the mean flow rate temperature.

It follows from (E5.4.1.18)–(E5.4.1.20) that at large distances from the inlet cross-section (as $z \to +\infty$) the Nusselt number tends to the constant value

$$\mathsf{Nu}_\infty = \frac{-f_0'(1)}{2 \displaystyle\int_0^1 f_0(\varrho)(1 - \varrho^2) \varrho \, d\varrho}. \qquad \text{(E5.4.1.21)}$$

For low Peclet numbers, the limit Nusselt number is

$$\mathsf{Nu}_\infty = \frac{\gamma_0^3}{4} \frac{J_1(\gamma_0)}{J_2(\gamma_0)} \approx 4.16 \qquad (\text{as } \mathsf{Pe}_T \to 0). \qquad \text{(E5.4.1.22)}$$

For high Peclet numbers, the limit Nusselt number is

$$\mathsf{Nu}_\infty = \tfrac{1}{2} \lambda_0^2 \approx 3.66 \qquad (\text{as } \mathsf{Pe}_T \to \infty). \qquad \text{(E5.4.1.23)}$$

In the entire range of the Peclet number, the limit Nusselt number is nicely approximated by the formula

$$\mathsf{Nu}_\infty = \frac{4.16 + 1.15\,\mathsf{Pe}_T}{1 + 0.315\,\mathsf{Pe}_T}, \qquad \text{(E5.4.1.28)}$$

whose maximum error is about 0.6%.

Calculations according to (E5.4.1.20) show that for high Peclet numbers one can conventionally divide the entire length of a heated (cooled) tube into two parts. On the first part, a temperature profile is formed with radial temperature distribution varying from the initial value (at $z = 0$) to some limit value $f_0(\varrho)$. In this region, the number Nu decreases near the inlet cross-section as a power-law function $\mathsf{Nu} \approx 2j_T$, where j_T is described by (E5.4.1.5). On the second part of the tube, the radial distribution of the excess temperature $\delta T = 1 - T$ does not vary along the tube (though the absolute values of temperature do vary), and the number Nu preserves the constant value 3.66. The first part of the tube is called the thermal initial region, and the second one, the region of steady-state heat exchange.

It is conventional to define the length of the thermal initial region as the distance from the inlet cross-section to the point at which the Nusselt number differs from its limit value (E5.4.1.23) by 1%. Calculations show that the dimensional length of the thermal initial region is $l = 0.11a\,\mathsf{Pe}_T$.

E5.4.2. Heat and Mass Transfer in a Circular Tube with Constant Heat Flux at the Wall for Laminar Flows

► **Statement of the problem.** Now let us study the case in which a constant heat flux $q = \kappa(\partial T/\partial \mathcal{R})_{\mathcal{R}=a} = \text{const}$, where κ is the thermal conductivity coefficient of the fluid, is given on the surface of a circular tube for $Z > 0$. The entry part is modeled by the region $Z < 0$, where there is no heat flux across the tube surface and the temperature tends to the constant value T_1 as $Z \to -\infty$.

In this case, it is convenient to introduce the dimensionless temperature as

$$\widetilde{T} = \frac{\kappa(T - T_1)}{aq}, \qquad \text{(E5.4.2.1)}$$

and the other dimensionless variables are determined in the same way as in problem (E5.4.1.1)–(E5.4.1.3).

The considered heat transfer process is described by Eq. (E5.4.1.1) with T replaced by $\widetilde{T}$ and with the boundary conditions

$$\varrho = 0, \quad \frac{\partial \widetilde{T}}{\partial \varrho} = 0; \qquad \varrho = 1, \quad \frac{\partial \widetilde{T}}{\partial \varrho} = \begin{cases} 0 & \text{for } z < 0; \\ 1 & \text{for } z > 0; \end{cases} \qquad \text{(E5.4.2.2)}$$

$$z \to -\infty, \quad \widetilde{T} \to 0. \qquad \text{(E5.4.2.3)}$$

Note that the temperature distribution as $z \to +\infty$ is not known in advance.

► **Temperature field far from the inlet cross-section.** The temperature distribution far from the inlet cross-section has the form

$$\widetilde{T} = 4\frac{z}{\mathsf{Pe}_T} + \varrho^2 - \frac{\varrho^4}{4} + \frac{8}{\mathsf{Pe}_T^2} - \frac{7}{24} \qquad (z \gg 1). \qquad \text{(E5.4.2.4)}$$

The mean flow rate temperature of the fluid and the temperature head in the region of heat stabilization are, respectively, equal to

$$\langle\widetilde{T}\rangle_m = \frac{4}{\mathsf{Pe}_T}z + \frac{8}{\mathsf{Pe}_T^2}, \qquad \widetilde{T}_\text{s} - \langle\widetilde{T}\rangle_m = \frac{11}{24},$$

where $\widetilde{T}_\text{s}$ is the dimensionless temperature on the tube wall.

The limit Nusselt number is

$$\mathsf{Nu}_\infty = \frac{2(d\widetilde{T}/d\varrho)_{\varrho=1}}{\widetilde{T}_\text{s} - \langle\widetilde{T}\rangle} = \frac{48}{11} \approx 4.36.$$

Obviously, Nu_∞ is independent of the Peclet number.

► **Temperature field in the initial region of the tube.** For $z < 0$, we seek the solution of the complete problem (E5.4.1.1), (E5.4.2.2), (E5.4.2.3) in the form of the series (E5.4.1.8) (with T replaced by $\widetilde{T}$). For $z > 0$, the temperature field is constructed on the basis of the asymptotic distribution (E5.4.2.4) as follows:

$$\widetilde{T} = 4\frac{z}{\mathsf{Pe}_T} + \varrho^2 - \frac{\varrho^4}{4} + \frac{8}{\mathsf{Pe}_T^2} - \frac{7}{24} - \sum_{m=0}^{\infty} A_m \exp\left(-\lambda_m^2 \frac{z}{\mathsf{Pe}_T}\right) f_m(\varrho).$$

By substituting these series into (E5.4.1.1), (E5.4.2.2), and (E5.4.2.3) and by separating the variables, we obtain spectral problems for the eigenfunctions f_m and g_k and the eigenvalues λ_m and η_k.

The solution of the problem for f_m is given by (E5.4.1.10), where the numbers λ_m satisfy the transcendental equation

$$\Phi(a_m,\ 1;\ \lambda_m) = 2a_m\Phi(a_m + 1,\ 2;\ \lambda_m).$$

The coefficients of the expansions A_m and B_m are determined by the condition that the temperature and its derivative are continuous across the section $z = 0$ (E5.4.1.9).

If we determine the length l of the thermal initial region on the basis of the formula $\mathsf{Nu} = 1.01\,\mathsf{Nu}_\infty$, then we obtain $l = 0.14\,\mathsf{Pe}a$.

E5.4.3. Heat and Mass Transfer in Plane Channels for Laminar Flows

► **Plane channel with constant temperature of the wall.** We shall study the heat exchange in a laminar fluid flow with parabolic velocity profile in a plane channel of width $2h$. Let us introduce rectangular coordinates X, Y with the X-axis codirected with the flow and lying at equal distances from the channel walls. We assume that the temperature on the walls (at $Y = \pm h$) is constant and is equal to T_1 for $X < 0$ and to T_2 for $X > 0$. Since the problem is symmetric with respect to the X-axis, it suffices to consider half of the flow region, $0 \le Y \le h$.

The temperature distribution T_* is described by the following equation and boundary conditions:

$$\mathsf{Pe}_T(1 - y^2)\frac{\partial T}{\partial x} = \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}; \tag{E5.4.3.1}$$

$$y = 0, \quad \frac{\partial T}{\partial y} = 0; \qquad y = 1, \quad T = \begin{cases} 0 & \text{for } x < 0; \\ 1 & \text{for } x > 0; \end{cases} \tag{E5.4.3.2}$$

$$x \to -\infty, \quad T \to 0; \qquad x \to +\infty, \quad T \to 1, \tag{E5.4.3.3}$$

with the dimensionless variables

$$x=\frac{X}{h},\quad y=\frac{Y}{h},\quad T=\frac{T_*-T_1}{T_2-T_1},\quad \mathsf{Pe}_T=\frac{hU_{\max}}{\chi},\quad U_{\max}=\frac{3}{2}\langle V\rangle,$$

where $U_{\max}$ is the maximum fluid velocity on the flow axis and $\langle V\rangle$ is the mean flow velocity over the cross-section.

By analogy with the case of a circular tube, we seek the solution of problem (E5.4.3.1)–(E5.4.3.3) in the form of a series for separated variables:

$$T=1-\sum_{m=0}^{\infty}A_m\exp\left(-\frac{\lambda_m^2}{\mathsf{Pe}_T}x\right)f_m(y)\qquad\text{for}\quad x>0,\tag{E5.4.3.4}$$

$$T=\sum_{k=0}^{\infty}B_k\exp\left(\frac{\eta_k^2}{\mathsf{Pe}_T}x\right)g_k(y)\qquad\text{for}\quad x<0.\tag{E5.4.3.5}$$

The eigenvalues λ_m, η_k, the eigenfunctions f_m, g_k, and the coefficients A_m and B_k can be obtained by analogy with the case of a circular tube in Section E5.4.1.

In what follows, we present only the basic solutions of the problem for $x>0$.

The eigenfunctions $f_m(y)$ can be written as

$$f_m(y)=\exp\left(-\frac{1}{2}\lambda_m y^2\right)\Phi\left(\frac{1}{4}-\frac{1}{4}\lambda_m-\frac{\lambda_m^3}{4\,\mathsf{Pe}_T^2},\ \frac{1}{2};\ \lambda_m y^2\right).\tag{E5.4.3.6}$$

Here $\Phi(a,b;\xi)$ is the degenerate hypergeometric function and the λ_m satisfy the transcendental equation

$$\Phi\left(a_m,\ \frac{1}{2};\ \lambda_m\right)=0,\quad\text{where}\quad a_m=\frac{1}{4}-\frac{\lambda_m}{4}-\frac{\lambda_m^3}{4\,\mathsf{Pe}_T^2}.\tag{E5.4.3.7}$$

The coefficients A_m are calculated by the formula (E5.4.1.12) with ϱ replaced by y, where the auxiliary function f can be obtained from (E5.4.3.6) after the indices m are omitted.

In the limit case as $\mathsf{Pe}_T\to 0$, we have

$$\lambda_m=\sqrt{\mathsf{Pe}_T\left(\frac{\pi}{2}+\pi m\right)},\quad A_m=\frac{4(-1)^m}{(\pi+2\pi m)^2},\quad f_m=\cos\left[\left(\frac{\pi}{2}+\pi m\right)y\right],\tag{E5.4.3.8}$$

where $m=0,\ 1,\ 2,\ \dots$

As $\mathsf{Pe}_T\to\infty$, the eigenvalues λ_m and the coefficients A_m can be obtained from formulas (E5.2.2.10) and (E5.2.2.11) (since formulas (E5.4.3.6) and (E5.4.3.7) pass as $\mathsf{Pe}_T\to\infty$ into formulas (E5.2.2.8) and (E5.2.2.9), respectively).

► **Mean flow rate temperature. Nusselt number.** The flow rate temperature for a plane channel is given by the formula

$$\langle T\rangle=\frac{3}{2}\int_0^1 T(1-y^2)\,dy.\tag{E5.4.3.9}$$

The local heat flux can be found by using (E5.4.1.19) with z replaced by x and the derivative $f_m'(1)$ calculated from (E5.4.3.6)

Let us substitute formulas (E5.4.3.4), (E5.4.3.9), and (E5.4.1.19) into (E5.4.1.20), and let x tend to infinity. As a result, we obtain the limit Nusselt number

$$\mathsf{Nu}_\infty = \frac{-4f_0'(1)}{3\int_0^1 f_0(y)(1-y^2)\,dy}. \tag{E5.4.3.10}$$

The eigenfunction $f_0(y) = \cos(\pi y/2)$ follows from (E5.4.3.8) for low Peclet numbers. By applying (E5.4.3.10), we obtain

$$\mathsf{Nu}_\infty = \frac{\pi^4}{24} \approx 4.06 \qquad (\text{as } \mathsf{Pe}_T \to 0). \tag{E5.4.3.11}$$

For high Peclet numbers, the limit Nusselt number is

$$\mathsf{Nu}_\infty = \frac{4}{3}\lambda_0^2 \approx 3.77 \qquad (\text{as } \mathsf{Pe}_T \to \infty). \tag{E5.4.3.12}$$

Over the entire range of Peclet numbers, Nu_∞ is nicely approximated by the formula

$$\mathsf{Nu}_\infty = \frac{4.06 + 3.66\,\mathsf{Pe}_T}{1 + 0.97\,\mathsf{Pe}_T}, \tag{E5.4.3.13}$$

whose maximum error is about 0.5%.

► **Plane channel with constant heat flux at the wall.** Now consider the case in which a constant heat flux q is given on the walls of a plane channel for $X > 0$. We assume that for $X < 0$ the walls are thermally insulated and the temperature tends to a constant T_1 as $X \to -\infty$.

For $a \equiv h$, we introduce the dimensionless temperature $\widetilde{T}$ according to (E5.4.2.1). The heat exchange in a plane channel is described by Eq. (E5.4.3.1) (as $T \to \widetilde{T}$) and the boundary conditions (E5.4.2.2), (E5.4.2.3), where z and ϱ are replaced by x and y, respectively.

The temperature distribution far from the inlet cross-section (in the heat stabilization region) has the form

$$\widetilde{T} = \frac{3}{2}\frac{x}{\mathsf{Pe}_T} + \frac{3}{4}y^2 - \frac{1}{8}y^4 + \frac{9}{4\,\mathsf{Pe}_T^2} - \frac{39}{280} \qquad (\text{for } x \gg 1). \tag{E5.4.3.14}$$

From the temperature distribution (E5.4.3.14) far from the input cross-section, one can find the limit Nusselt number

$$\mathsf{Nu}_\infty = 4.12.$$

E5.4.4. Turbulent Heat Transfer in Circular Tube and Plane Channel

► **Temperature profile.** Let us discuss qualitative specific features of convective heat and mass transfer in a turbulent flow through a circular tube and plane channel in the stabilized flow region. Experimental evidence indicates that several characteristic regions with different temperature profiles can be distinguished. At moderate Prandtl numbers ($0.5 \le \mathsf{Pr} \le 2.0$), the structure and sizes of these regions are similar to those of the wall layer and the core of the turbulent stream considered in Section E4.4.2.

In the molecular thermal conduction layer, adjacent to the tube wall, the deviation of the average temperature $\overline{T}$ from the wall temperature T_s satisfies the linear dependence (E5.3.1.10). In the logarithmic layer, the average temperature can be estimated using relations (E5.3.1.11), which are valid for liquids, gases, and liquid metals within a wide range of Prandtl numbers, $6 \times 10^{-3} \le \mathsf{Pr} \le 10^4$.

In the stream core and a major part of the logarithmic layer, the average temperature profile can be described by the single formula

$$\frac{\overline{T}(Y) - \overline{T}_{\rm m}}{\Theta_*} = -2.12 \ln \eta + 0.3\,(1 - \eta^2), \tag{E5.4.4.1}$$

where $\overline{T}_{\rm m}$ is the average temperature at the stream axis, $\eta = Y/a$ the dimensionless distance to the tube wall, and a the tube radius. The results predicted by Eq. (E5.4.4.1) fairly well agree with experimental data provided by various authors for turbulent flows of water, air, and liquid metals through circular tubes and plane channels for $3 \times 10^{-2} < Y/a < 1$.

► **Nusselt number for the thermal stabilized region.** In the thermal stabilization region of a turbulent flow through a smooth tube, for computing the limiting Nusselt number one can use the formula

$$\mathsf{Nu}_\infty = \frac{\mathsf{Pr}\,\mathsf{Re}_d \xi}{2.12 \ln(\mathsf{Pr}\,\mathsf{Re}_d \xi) + (3.85\,\mathsf{Pr}^{1/3} - 1.3)^2 - 4.3 + 6.7\,\xi}. \tag{E5.4.4.2}$$

Here the following notation is used:

$$\mathsf{Nu}_\infty = \frac{\alpha d}{c_p \rho \chi}, \quad \alpha = \frac{q_{\rm s}}{T_{\rm s} - \overline{T}_{\rm m}}, \quad \mathsf{Re}_d = \frac{d\langle V\rangle}{\nu}, \quad \mathsf{Pr} = \frac{\nu}{\chi}, \quad \xi = \sqrt{\tfrac{1}{8}\lambda},$$

where α is the heat transfer coefficient, $d = 2a$ the tube diameter, c_p the specific heat at constant pressure, ρ the fluid density, χ the thermal diffusivity, $q_{\rm s}$ the heat flux at the wall, ν the kinematic viscosity, and $\langle V\rangle$ the mean flow rate velocity. The drag coefficient λ can be found from Eq. (E4.4.2.12) or (E4.4.2.13). Formula (E5.4.4.2) unifies the results of more than fifty experimental studies and applies to liquids and gases within wide ranges of Reynolds and Prandtl numbers, $5 \times 10^3 \le \mathsf{Re}_d \le 2 \times 10^6$ and $0.6 \le \mathsf{Pr} \le 4 \times 10^4$. In the case of mass exchange, the Nusselt number must be replaced by the Sherwood number and the thermal Prandtl number, by the diffusion one.

In engineering calculations, the Nusselt number is often determined from simple two-term (or one-term) formulas like

$$\mathsf{Nu}_\infty = A + B\mathsf{Pr}^n \mathsf{Re}_d^m \tag{E5.4.4.3}$$

valid in a limited range of Prandtl and Reynolds numbers. For liquid metals, in the case of $10^4 < \mathsf{Re}_d < 10^6$ and constant temperature or heat flux at the wall, in Eq. (E5.4.4.3) one should set

$$\begin{aligned} &A = 5, && B = 0.025, && n = m = 0.8 && (\text{if } T_{\rm s} = \text{const}),\\ &A = 6.8, && B = 0.044, && n = m = 0.75 && (\text{if } q_{\rm s} = \text{const}). \end{aligned} \tag{E5.4.4.4}$$

► **Intermediate domain and the entry region of the tube.** In the intermediate domain ($2200 \le \mathsf{Re}_d \le 4000$) at moderate Prandtl numbers, the Nusselt number can be estimated using the formula

$$\mathsf{Nu} = 3 \times 10^{-4} \mathsf{Pr}^{0.35} \mathsf{Re}_d^{1.5} \qquad (\text{for } 0.5 \le \mathsf{Pr} \le 5).$$

For the entry region of the tube in the case of simultaneous development of hydrodynamic and thermal boundary layers, the local Nusselt number $\mathsf{Nu} = \mathsf{Nu}(X)$ at constant temperature or heat flux at the wall can be calculated from the formulas

$$\mathsf{Nu} = \begin{cases} 0.02\,\mathsf{Re}_d^{0.75}\mathsf{Pr}^{0.4}\zeta(\zeta-1)^{-0.2} & \text{if } T_\mathrm{s} = \text{const}, \\ 0.021\,\mathsf{Re}_d^{0.75}\mathsf{Pr}^{0.4}\zeta^{0.8}(\ln\zeta)^{-0.2} & \text{if } q_\mathrm{s} = \text{const}. \end{cases} \tag{E5.4.4.5}$$

Here the following notation is used:

$$\mathsf{Nu} = \frac{\alpha d}{c_p \rho \chi}, \quad \mathsf{Re}_d = \frac{dU_\mathrm{i}}{\nu}, \quad \zeta = (1.18)^b, \quad b = \mathsf{Re}_d^{-0.25}\frac{X}{a},$$

where U_i is the inlet velocity and $\alpha = \alpha(X)$ the local heat transfer coefficient (based on the inlet temperature). Formulas (E5.4.4.5) are valid for $0.5 \le \mathsf{Pr} \le 200$ in the entire entry region, whose length x_L is estimated as $x_L = 1.3\,\mathsf{Re}^{0.25}$.

More detailed information about heat and mass transfer in turbulent flows through a circular tube or plane channel, as well as various relations for determining the temperature profile and Nusselt (Sherwood) numbers, and other useful information can be found in the references given at the end of this chapter.

E5.4.5. Limit Nusselt Numbers for Tubes of Various Cross-Sections

► **Laminar flows.** Below we give some results for the limit Nusselt numbers corresponding to the heat stabilization region for fully developed laminar flow of fluids in tubes of various cross-sections in the case of high Peclet numbers.

Let us introduce the equivalent (or "hydraulic") diameter d_e by the formula

$$d_\mathrm{e} = \frac{4S_*}{\mathcal{P}}, \tag{E5.4.5.1}$$

where S_* is the tube cross-section area and $\mathcal{P}$ is the cross-section perimeter. For tubes of circular cross-section, d_e coincides with the diameter, and for a plane channel, d_e is twice the height of the channel.

Consider a tube of arbitrary shape and denote the contour of the cross-section by Γ. Generally speaking, the Nusselt number varies along the contour Γ. We define the perimeter-average Nusselt number $\overline{\mathsf{Nu}}$ as follows:

$$\overline{\mathsf{Nu}} = \frac{q_\mathrm{s}}{T_\mathrm{s} - \langle T_* \rangle_m}\frac{d_\mathrm{e}}{\kappa}. \tag{E5.4.5.2}$$

Here T_s is the temperature on the wall of the tube, $\langle T_* \rangle_m$ is the mean flow rate temperature of the fluid, κ is the thermal conductivity coefficient, and q_s is the perimeter-average heat flux given by the formula

$$q_\mathrm{s} = -\frac{1}{\mathcal{P}}\kappa \int_\Gamma \left(\frac{\partial T_*}{\partial \xi}\right)_\Gamma d\Gamma, \tag{E5.4.5.3}$$

where $\partial T_*/\partial \xi$ is the derivative of the temperature T_* along the normal to the contour of the tube cross-section.

TABLE E5.2
The values of the limit numbers $\overline{\mathsf{Nu}}_\infty$ for fully developed flow in tubes of various shape for high Peclet numbers (the subscript T corresponds to the constant temperature of the wall and the subscript q, to the constant heat flux).

Tube profile		$\overline{\mathsf{Nu}}_{\infty T}$	$\overline{\mathsf{Nu}}_{\infty q}$	Equivalent diameter d_e
Circular tube of diameter d		3.658	4.364	d
Flat tube of width $2h$		7.541	8.235	$4h$
Elliptic tube with semiaxes a and b	$b/a =$			$\dfrac{\pi b}{E\left(\sqrt{1-b^2/a^2}\right)}$, where $E(\vartheta)$ is the complete elliptic integral of the second kind
	1.00	3.658	4.364	
	0.80	3.669	4.387	
	0.50	3.742	4.558	
	0.25	3.792	4.880	
	0.125	3.725	5.085	
	0.0625	3.647	5.176	
	0	3.488	5.225	
Tube of rectangular cross-section with sides a and b	$b/a =$			$\dfrac{2ab}{a+b}$
	1.00	2.976	3.608	
	0.714	3.077	3.734	
	0.50	3.391	4.123	
	0.25	4.439	5.331	
	0.125	5.597	6.490	
	0.05		7.451	
	0	7.541	8.235	
Equilateral triangle with side a		2.47	3.111	$\dfrac{a\sqrt{3}}{3}$
Regular hexagon with side a		3.34	4.002	$a\sqrt{3}$
Semi-circle of diameter d			4.089	$\dfrac{\pi d}{\pi+2}$

Table E5.2 presents the perimeter-average Nusselt numbers for tubes with various cross-section shapes.

At constant temperature on the wall of a tube of rectangular cross-section with sides a and b, the value $\overline{\mathsf{Nu}}_\infty$ for $a \geq b$ is nicely approximated by the formula

$$\overline{\mathsf{Nu}}_\infty = 7.5 - 17.5\,\epsilon + 23\,\epsilon^2 - 10\,\epsilon^3, \qquad \epsilon = b/a, \tag{E5.4.5.5}$$

whose maximum error is 3%.

At constant temperature on the wall of a tube whose cross-section is a regular N-gon, the limit Nusselt number is given by the approximate formula

$$\overline{\mathsf{Nu}}_\infty = 3.65 - 0.18\,N^{-1} - 10\,N^{-2}. \tag{E5.4.5.6}$$

The comparison with Table E5.2 shows that the maximum error in (E5.4.5.6) is less than 0.5% at $N = 3,\ 4,\ 6,\ \infty$.

► **Turbulent flows.** The strength of heat transfer in turbulent flows of fluids through tubes of noncircular cross-section can be estimated using relations (E5.4.4.2)–(E5.4.4.4) suggested for circular tubes. In this case, the characteristic length on which the Nusselt numbers Nu_∞ are based is taken to be the equivalent diameter, d_e, defined by Eq. (E5.4.5.1).

The quantity $\mathcal{P}$ in this formula is the wettable perimeter, regardless of what part of the perimeter exchanges heat with the fluid.

Detailed information about heat transfer in turbulent flows in tubes and channels, as well as various relations for determining the mean flow rate temperature and the Nusselt number, can be found in the references given at the end of this chapter.

E5.5. Interior Heat Problems for Solid Bodies of Various Shapes

E5.5.1. Statement of the Problem

Consider a class of problems concerning transient heat exchange between convex solid bodies (or a cavity filled with a quiescent fluid) of various shapes and the environment. At the initial time $t = 0$, the temperature is the same throughout the body and is equal to T_i, and for $t > 0$, the temperature on the surface Γ of the body is maintained constant and is equal to T_s. The temperature distribution inside the body is described by the heat equation

$$\frac{\partial T}{\partial \bar{\tau}} = \Delta T, \qquad \text{(E5.5.1.1)}$$

the initial condition

$$T = 0 \quad \text{at} \quad \bar{\tau} = 0, \qquad \text{(E5.5.1.2)}$$

and the boundary condition

$$T = 1 \quad \text{on } \Gamma. \qquad \text{(E5.5.1.3)}$$

The dimensionless variables are introduced according to formulas (E5.1.1.13), where the characteristic length can be taken arbitrarily.

The problem of transient diffusion into a cavity filled with a quiescent medium can be stated in a similar manner.

In this section, attention is chiefly paid to the study of the bulk body temperature

$$\langle T \rangle = \frac{1}{V} \int_v T\, dv, \qquad \text{(E5.5.1.4)}$$

where $V = \int_v dv$ is the dimensionless volume of the body.

E5.5.2. Bulk Temperature for Bodies of Various Shapes

► **Bulk temperature for a sphere, a parallelepiped, and a cylinder of finite length.** First, consider the simplest one-dimensional (with respect to spatial coordinates) heat exchange problem for a solid sphere of radius a. The exact solution of this problem is well known and results in the following expression for the bulk temperature:

$$\langle T \rangle = 1 - \frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k^2} \exp(-\pi^2 k^2 \bar{\tau}), \qquad \bar{\tau} = \frac{\chi t}{a^2}. \qquad \text{(E5.5.2.1)}$$

The exact solution of the corresponding three-dimensional problem (E5.5.1.1)–(E5.5.1.3) for a parallelepiped with sides L_1, L_2, and L_3 can be constructed by separation of variables and results in the following formula for bulk temperature:

$$\langle T\rangle = 1-\left(\frac{8}{\pi^2}\right)^3 \sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\sum_{l=1}^{\infty} \frac{1}{(2k-1)^2(2m-1)^2(2l-1)^2} \times \exp\left\{-\pi^2\left[\frac{(2k-1)^2}{L_1^2}+\frac{(2m-1)^2}{L_2^2}+\frac{(2l-1)^2}{L_3^2}\right]\chi t\right\}. \qquad \text{(E5.5.2.2)}$$

Now consider heat exchange for a cylinder of finite length. Let a be the radius and L the length of the cylinder. By solving problem (E5.5.1.1)–(E5.5.1.3), we obtain the following expression for bulk temperature:

$$\langle T\rangle = 1-\frac{32}{\pi^2}\sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\frac{1}{\vartheta_k^2(2m-1)^2}\exp\left\{-\left[\frac{\vartheta_k^2}{a^2}+\frac{\pi^2(2m-1)^2}{L^2}\right]\chi t\right\}, \qquad \text{(E5.5.2.3)}$$

where the ϑ_k are the roots of the Bessel function of the first kind of index zero: $J_0(\vartheta_k) = 0$. Note that to compute the roots of this transcendental equation one can use the approximate formula

$$\vartheta_k = 2.4 + 3.13\,(k-1) \qquad (k = 1, 2, \dots)$$

with a maximum error of 0.2%.

► **Bulk temperature for a bounded body of arbitrary shape.** For a bounded body of arbitrary shape, the solution of problem (E5.5.1.1)–(E5.5.1.3) tends as $\bar{\tau}\to\infty$ to the limit value (equal to 1) determined by the boundary condition on the surface of the body. By setting $T = 1$ in (E5.5.1.4), we find the asymptotics of bulk temperature for large $\bar{\tau}$:

$$\langle T\rangle \to 1 \quad \text{as} \quad \bar{\tau}\to\infty. \qquad \text{(E5.5.2.4)}$$

One can show that the bulk temperature of a bounded body of arbitrary shape has the following asymptotics for small $\bar{\tau}$:

$$\langle T\rangle \to 2\frac{S}{V}\sqrt{\frac{\bar{\tau}}{\pi}} \quad \text{at} \quad \bar{\tau}\to 0, \qquad \text{(E5.5.2.5)}$$

where S and V are the dimensionless surface area and volume of the body.

For approximately determining the bulk temperature of a bounded convex body of arbitrary shape in the entire time range $0 \le \bar{\tau} < \infty$, one can use the formula

$$\langle T\rangle = 1-\frac{6}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{k^2}\exp\left(-\frac{\pi^2k^2S^2}{9V^2}\bar{\tau}\right).$$

This expression can be rewritten as follows:

$$\langle T\rangle = 1-\frac{6}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{k^2}\exp\left(-\frac{\pi^2}{9}k^2\frac{S_*^2\chi t}{V_*^2}\right), \qquad \text{(E5.5.2.6)}$$

where S_* and V_* are, respectively, the dimensional surface area and volume of the body.

Formula (E5.5.2.6) gives the correct asymptotics (E5.5.2.4) and (E5.5.2.5) for small and large times and is exact for a spherical particle. (Formula (E5.5.2.6) for $S_* = 4\pi a$ and $V_* = \frac{4}{3}\pi a^3$ coincides with (E5.5.2.1).)

The approximation (E5.5.2.6) is compared with the exact bulk temperature (E5.5.2.2) for the parallelepiped in view of the relations $S_* = 2(L_1L_2 + L_1L_3 + L_2L_3)$ and $V_* = L_1L_2L_3$. The maximum error of formula (E5.5.2.6) is about 5% for $0.25 \le L_3/L_1 \le 4.0$ and $L_2/L_1 = 1$.

The exact value (E5.5.2.3) is compared with the approximation (E5.5.2.6), where $S_* = 2\pi a(a + L)$ and $V_* = \pi a^2 L$, for various ratios of the cylinder dimensions. The maximum error in formula (E5.5.2.6) is about 3.5% for $0.25 \le 2a/L \le 4.0$.

For practical calculations, it is expedient to replace the infinite series (E5.5.2.6) by the simpler formula

$$\langle T\rangle = \sqrt{1 - e^{-1.27\,\omega}} + 0.6\,\left(e^{-1.5\,\omega} - e^{-1.1\,\omega}\right), \qquad \omega = \frac{S_*^2\chi t}{V_*^2},$$

whose maximum deviation from (E5.5.2.6) is about 1.7%.

E5.6. Mass and Heat Exchange Between Particles of Various Shapes and a Quiescent Medium

E5.6.1. Stationary Mass and Heat Exchange

► **Statement of the problem.** Consider stationary diffusion to a particle of finite size in a quiescent medium, which corresponds to the case $\mathsf{Pe} = 0$. We assume that the concentration on the surface of the particle and far from it is constant and equal to C_s and C_i, respectively. The concentration field outside the particle is described by the Laplace equation

$$\Delta c = 0 \tag{E5.6.1.1}$$

and the boundary conditions

$$c = 1 \quad \text{on the surface } \Gamma \text{ of the particle,} \tag{E5.6.1.2}$$

$$c = 0 \quad \text{remote from the particle,} \tag{E5.6.1.3}$$

where $c = \dfrac{C_\mathrm{i} - C}{C_\mathrm{i} - C_\mathrm{s}}$ is the dimensionless concentration.

The unknown quantity of most practical interest in these problems is the mean Sherwood number, which is determined by (E5.1.2.5) and is related to the mass transfer coefficient α_c by

$$\mathsf{Sh} = \frac{a\alpha_c}{D}, \tag{E5.6.1.4}$$

where a is the characteristic length.

From the mathematical viewpoint, the diffusion problem (E5.6.1.1)–(E5.6.1.3) is equivalent to the problem on the electric field of a charged conductive body in a homogeneous charge-free dielectric medium. Therefore, the mean Sherwood number in a quiescent fluid coincides with the dimensionless electrostatic capacitance of the body and can be calculated or measured by methods of electrostatics.

► **Shape factor.** In what follows, it is convenient to introduce a shape factor Π, which has the dimension of length, as follows:

$$\Pi = \frac{\alpha_c S_*}{D} = \mathsf{Sh}\frac{S_*}{a}, \tag{E5.6.1.5}$$

where S_* is the dimensional surface area of the particle. Note that sometimes Π is referred to as “conductance.”

TABLE E5.3
Shape factor for particles in quiescent medium.

No	Shape of particle	Shape factor $\Pi = \mathsf{Sh}\dfrac{S_*}{a}$
1	Sphere of radius a	$4\pi a$
2	Oblate ellipsoid of revolution with semiaxes a and b, $\chi = b/a < 1$	$\dfrac{4\pi a\sqrt{1-\chi^2}}{\arccos\chi}$
3	Prolate ellipsoid of revolution with semiaxes a and b, $\chi = b/a > 1$	$\dfrac{4\pi a\sqrt{\chi^2-1}}{\ln\bigl(\chi+\sqrt{\chi^2-1}\bigr)}$
4	Circular cylinder of radius a and length L $(0 \le L/a \le 16)$	$\bigl[8+4.1\,(L/a)^{0.76}\bigr]a$
5	Tangent spheres of equal radius a	$2\ln 2\,(4\pi a)$
6	Tangent spheres of radii a_1 and a_2	$-\dfrac{4\pi a_1a_2}{a_1+a_2}\Bigl[\psi\Bigl(\dfrac{a_1}{a_1+a_2}\Bigr)+\psi\Bigl(\dfrac{a_2}{a_1+a_2}\Bigr)+2\ln\gamma\Bigr]$, where $\psi(x)=\frac{d}{dx}\Gamma(x)$ is the logarithmic derivative of the gamma function and $\ln\gamma=-\psi(1)=0.5772\ldots$ is the Euler constant
7	Orthogonally intersecting spheres with radii a_1 and a_2	$4\pi\left(a_1+a_2-\dfrac{a_1a_2}{\sqrt{a_1^2+a_2^2}}\right)$
8	Cube with edge a	$0.654\,(4\pi a)$
9	Thin rectangular plate with sides L_1 and L_2 $(L_1 \ge L_2)$	$\dfrac{2\pi L_1}{\ln(4L_1/L_2)}$

Table E5.3 shows the values of Π for particles of various shape. It follows from (E5.6.1.5) that the mean Sherwood number can be obtained from the data in this table by dividing the shape factor by the surface area of the particle and then by multiplying by the characteristic length.

One can interpret the table data as follows. Let us project a body of revolution onto a plane perpendicular to the symmetry axis. The projection is a disk of radius a_p. The sphere of radius a_p is called a perimeter-equivalent sphere. We introduce the perimeter-equivalent factor

$$\Sigma = \frac{S_*}{4\pi a_\mathrm{p}^2} = \frac{\text{surface area of particle}}{\text{surface area of perimeter-equivalent sphere}} \qquad \text{(E5.6.1.6)}$$

and consider the corresponding shape factor ratio

$$\widetilde{\Pi} = \frac{\Pi}{4\pi a_\mathrm{p}} = \frac{\text{shape factor of the particle}}{\text{shape factor of the perimeter-equivalent sphere}}. \qquad \text{(E5.6.1.7)}$$

The dimensionless variables (E5.6.1.6) and (E5.6.1.7) are invariant with respect to the choice of the characteristic length. The dependence of $\widetilde{\Pi}$ on Σ for particles of various geometric shapes is fairly well approximated by the formula

$$\widetilde{\Pi} = 0.637 + 0.327\,(2\Sigma-1)^{0.76} \qquad (0.5 \le \Sigma \le 8.5). \qquad \text{(E5.6.1.8)}$$

It is expedient to use this approximate formula for the analysis of mass exchange between particles of complicated shape and a quiescent medium if the solution of problem (E5.6.1.1)–(E5.6.1.3) is unknown.

To calculate the particle shape factor, it is also convenient to use the simple approximate formula

$$\Pi = 5.25\, S_*^{1/4} V_*^{1/6}, \tag{E5.6.1.9}$$

where S_* is the surface area and V_* the volume of the particle.

For spheres and cubes, the error in formula (E5.6.1.9) is at most 0.13%. For circular cylinders of radius a and length $L = a$, $L = 2a$, or $L = 3a$, the error in formula (E5.6.1.9) is respectively 1.1%, 0.6%, and 2.1%.

It is useful to rewrite formula (E5.6.1.9) as follows:

$$\Pi = \sqrt{S_*}\, f(\xi), \qquad f(\xi) = A_s\sqrt{\xi}, \qquad A_s = 2\sqrt{\pi} \approx 3.545,$$

where ξ is a dimensionless geometric parameter,

$$\xi = \lambda \frac{V_*^{1/3}}{S_*^{1/3}}, \qquad \lambda = (36\pi)^{1/6} \approx 2.199.$$

The parameter ξ characterizes the particle shape and ranges in the interval $0 \le \xi \le 1$. The value $\xi = 1$ corresponds to a spherical particle. Formula (E5.6.1.9) gives good results for weakly and moderately deformed convex particles with $0.88 \le \xi \le 1.0$. (For example, this condition is satisfied for cubic particles, regular polyhedra, and circular cylinders of radius a and length L with $a \le L \le 3a$.)

Let us write out some lower and upper bounds for the shape factor. The lower bound for an arbitrary particle is determined by the shape factor of the sphere of the same volume V_*:

$$\Pi \ge (48\pi^2 V_*)^{1/3}. \tag{E5.6.1.10}$$

Another lower bound has the form

$$\Pi \ge 8(S_{\max}/\pi)^{1/2},$$

where $S_{\max}$ is the maximum projected area of the particle. The last formula becomes an equality for a disk. An upper bound is given by the shape factor of any surface (say, a sphere or an ellipsoid of revolution) surrounding the particle.

E5.6.2. Transient Mass and Heat Exchange

Suppose that at the initial time $t = 0$ the concentration in the continuous medium is constant and is equal to C_i and that a constant surface concentration C_s is maintained for $t > 0$. The transient mass exchange between a particle and a quiescent medium is described by the equation

$$\frac{\partial c}{\partial \tau} = \Delta c \tag{E5.6.2.1}$$

with the boundary conditions (E5.6.1.2) and (E5.6.1.3) and the initial condition

$$\tau = 0, \quad c = 0, \tag{E5.6.2.2}$$

where $\tau = tD/a^2$ is dimensionless time.

For a spherical particle, the solution of problem (E5.6.2.1), (E5.6.2.2), (E5.6.1.2), (E5.6.1.3) can be expressed via the complementary error function and has the form

$$c = \frac{1}{r}\,\mathrm{erfc}\left(\frac{r-1}{2\sqrt{\tau}}\right), \tag{E5.6.2.3}$$

where r is the dimensionless radial coordinate normalized to the radius of the particle.

The mean Sherwood number for a sphere is

$$\mathsf{Sh} = 1 + \frac{1}{\sqrt{\pi\tau}}. \tag{E5.6.2.4}$$

For nonspherical particles, the mean Sherwood number can be approximated by the expression

$$\mathsf{Sh} = \mathsf{Sh}_{\rm st} + \frac{1}{\sqrt{\pi\tau}}, \tag{E5.6.2.5}$$

where $\mathsf{Sh}_{\rm st}$ is the Sherwood number corresponding to the solution of the stationary problem (E5.6.1.1)–(E5.6.1.3). In (E5.6.2.5), Sh, $\mathsf{Sh}_{\rm st}$, and τ are assumed to be reduced to dimensionless form by using the same characteristic length.

For nonspherical particles, one can obtain $\mathsf{Sh}_{\rm st}$ by using Table E5.3 and the expression (E5.6.1.8).

E5.7. Mass and Heat Transfer to a Spherical Particle in a Translational Flow

E5.7.1. Statement of the Problem

► **Equation and the boundary conditions.** Consider steady-state diffusion to a solid spherical particle of radius a in a translational flow with velocity $U_{\rm i}$. We assume that the concentration on the particle surface and far from it is constant and equal to $C_{\rm s}$ and $C_{\rm i}$, respectively. The mass transfer process in the continuous medium is described by the equation (E5.1.1.9), which in the stationary case in the spherical coordinate system R, θ, φ (in view of the fact that all variables in this axisymmetric problem are independent of φ) has the form

$$\mathsf{Pe}\left(v_r\frac{\partial c}{\partial r} + \frac{v_\theta}{r}\frac{\partial c}{\partial\theta}\right) = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial c}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial c}{\partial\theta}\right), \tag{E5.7.1.1}$$

and the boundary conditions

$$r = 1, \quad c = 1 \quad \text{(on the surface of the particle)}, \tag{E5.7.1.2}$$

$$r \to \infty, \quad c \to 0 \quad \text{(far from the particle)}, \tag{E5.7.1.3}$$

where the dimensionless variables are introduced according to the formulas

$$r = \frac{R}{a}, \quad v_r = \frac{V_R}{U_{\rm i}}, \quad v_\theta = \frac{V_\theta}{U_{\rm i}}, \quad c = \frac{C_{\rm i} - C}{C_{\rm i} - C_{\rm s}}, \quad \mathsf{Pe} = \frac{aU_{\rm i}}{D}. \tag{E5.7.1.4}$$

► **Mean Sherwood number.** In this section, attention is mainly paid to formulas for the calculation of the mean Sherwood number for a particle in a translational flow at various Reynolds and Peclet numbers. The Reynolds number and the mean Sherwood number, as well as the Peclet number, are determined by the particle radius: $\mathsf{Re} = aU_{\rm i}/\nu$ and

$$\mathsf{Sh} = -\frac{1}{2}\int_0^\pi \sin\theta\left(\frac{\partial c}{\partial r}\right)_{r=1} d\theta. \tag{E5.7.1.5}$$

E5.7.2. Mass Transfer at Low Reynolds Numbers and Various Peclet Numbers

For a solid spherical particle in a translational Stokes flow (low Reynolds numbers, $\mathsf{Re} \to 0$) the fluid velocity components V_R and V_θ in Eq. (E5.7.1.4) are determined by formulas (E4.6.2.4).

► **Low Peclet numbers.** For $\mathsf{Pe} = 0$, which corresponds to a particle in a quiescent fluid, the exact solution of the problem (E5.7.1.1)–(E5.7.1.3) has the form

$$c = \frac{1}{r}. \qquad \text{(E5.7.2.1)}$$

The corresponding mean Sherwood number is equal to unity,

$$\mathsf{Sh} = 1 \qquad (\mathsf{Pe} = 0). \qquad \text{(E5.7.2.2)}$$

The three-term expansion of the mean Sherwood number as $\mathsf{Pe} \to 0$ has the form

$$\mathsf{Sh} = 1 + \tfrac{1}{2}\,\mathsf{Pe} + \tfrac{1}{2}\mathsf{Pe}^2 \ln \mathsf{Pe}, \qquad \text{(E5.7.2.3)}$$

whose error is of the order of Pe^2. The asymptotic formula (E5.7.2.3) with the last term omitted gives a good approximation to the Sherwood number for $\mathsf{Pe} < 0.4$.

► **High Peclet numbers (the diffusion boundary layer).** For high Peclet numbers, a thin diffusion boundary layer is formed near the surface of the sphere. The ratio of the thickness of this layer to the particle radius is of the order of $\mathsf{Pe}^{-1/3}$. In this region, the radial component of molecular diffusion to the particle surface is essential, and tangential diffusion may be neglected. Convective mass transfer due to the motion of the fluid should also be taken into account.

The concentration distribution in the diffusion boundary layer can be expressed as follows via the incomplete gamma function:

$$c = \frac{1}{\Gamma(1/3)}\Gamma\left(\frac{1}{3}, \frac{\mathsf{Pe}}{3}\,\frac{(r-1)^3 \sin^3\theta}{\pi - \theta + \frac{1}{2}\sin 2\theta}\right), \qquad \text{(E5.7.2.4)}$$

where the angle θ is counted off as shown in Figure E4.4.

By differentiating equation (E5.7.2.4), we obtain the dimensionless local diffusion flux to the surface of the sphere in the form

$$j = -\left(\frac{\partial c}{\partial r}\right)_{r=1} = 0.766 \sin\theta \left(\pi - \theta + \frac{1}{2}\sin 2\theta\right)^{-1/3} \mathsf{Pe}^{1/3}. \qquad \text{(E5.7.2.5)}$$

One can see that the local diffusion flux attains its maximum at the front stagnation point on the surface of the sphere (at $\theta = \pi$) and monotonically decreases with the angular coordinate to the minimum value, which is zero and is attained at $\theta = 0$.

The corresponding mean Sherwood number is determined by Levich's formula

$$\mathsf{Sh} = 0.625\,\mathsf{Pe}^{1/3} \qquad (\mathsf{Pe} \to \infty). \qquad \text{(E5.7.2.6)}$$

Let us also present the two-term expansion for the mean Sherwood number,

$$\mathsf{Sh} = 0.625\,\mathsf{Pe}^{1/3} + 0.461 \qquad (\mathsf{Pe} \to \infty). \qquad \text{(E5.7.2.7)}$$

which refines formula (E5.7.2.6). Formula (E5.7.2.7) can be used for practical calculations if $\mathsf{Pe} \geq 10$.

► **Arbitrary Peclet numbers.** The mean Sherwood number for a solid spherical particle in a translational Stokes flow in the entire range of Peclet numbers can be computed with the use of the Clift–Grace–Weber approximate formula

$$\mathsf{Sh} = 0.5 + (0.125 + 0.243\,\mathsf{Pe})^{1/3}. \qquad \text{(E5.7.2.8)}$$

The interpolation formula (E5.7.2.8) gives exact asymptotic results both as $\mathsf{Pe} \to 0$ (E5.7.2.2) and as $\mathsf{Pe} \to \infty$ (E5.7.2.6). The maximum error of formula (E5.7.2.8) in the entire range of Peclet numbers is about 2%. (Comparison was made with available results of finite-difference numerical solutions of problem (E5.7.1.1)–(E5.7.1.3).)

E5.7.3. Mass Transfer at Moderate and High Reynolds Numbers

► **Spherical particle at various Reynolds numbers.** In the case of a spherical particle in a translational flow at $0.5 \le \mathsf{Re} \le 200$ and $0.125 \le \mathsf{Sc} \le 50$, a number of numerical results concerning the mean Sherwood number can be described by the Clift–Grace–Weber approximate formula

$$\mathsf{Sh} = 0.5 + 0.527\,\mathsf{Re}^{0.077}(1 + 2\,\mathsf{Re}\,\mathsf{Sc})^{1/3}, \qquad \text{(E5.7.3.1)}$$

whose error is about 3%.

The analysis of available experimental data on heat and mass transfer to a solid sphere in a translational flow results in the following Clift–Grace–Weber correlations.

Heat exchange with air at $\mathsf{Pr} = 0.7$:

$$\begin{aligned} &\mathsf{Nu} = 0.5 + 0.47\,\mathsf{Re}^{0.47} && \text{for} \quad 50 \le \mathsf{Re} \le 2\times 10^3,\\ &\mathsf{Nu} = 0.5 + 0.2\,\mathsf{Re}^{0.58} && \text{for} \quad 2\times 10^3 \le \mathsf{Re} \le 5\times 10^4. \end{aligned}$$

Mass transfer with a liquid for large Schmidt numbers ($\mathsf{Sc} > 100$):

$$\begin{aligned} &\mathsf{Sh} = 0.5 + 0.5\,\mathsf{Re}^{0.48}\,\mathsf{Sc}^{1/3} && \text{for} \quad 50 \le \mathsf{Re} \le 10^3,\\ &\mathsf{Sh} = 0.5 + 0.31\,\mathsf{Re}^{0.55}\,\mathsf{Sc}^{1/3} && \text{for} \quad 10^3 \le \mathsf{Re} \le 5\times 10^4. \end{aligned}$$

The experimental data for $0.5 < \mathsf{Re} < 50$ are well described by the expression (E5.7.3.1).

► **General correlation for the Sherwood number.** For the calculation of the mean Sherwood number in laminar flows of various types without closed streamlines past a solid spherical particle, one can use the approximate formula

$$\mathsf{Sh} = 0.5 + (0.125 + \mathsf{Sh}_\infty^3)^{1/3}, \qquad \text{(E5.7.3.2)}$$

where the auxiliary value Sh_∞ must be chosen equal to the leading term of the asymptotic expansion of the Sherwood number as $\mathsf{Pe} \to \infty$ (for given $\mathsf{Re} = \text{const}$). Formula (E5.7.3.2) after the substitution of the value of Sh_∞ can be used for the calculation of the Sherwood number for various flows (one can consider, say, shear flows) in the entire range of Peclet numbers. For a solid spherical particle in a translational Stokes flow (as $\mathsf{Re} \to 0$), formula (E5.7.3.2) becomes (E5.7.2.8).

E5.8. Mass and Heat Transfer to Spherical Drops and Bubbles in Translational Flow

E5.8.1. Statement of the Problem. Some Remarks

In this section, attention is mainly paid to formulas for the calculation of the mean Sherwood number for spherical drops and bubbles of radius a in a translational flow with velocity U_i at various Reynolds and Peclet numbers. Consider the exterior problem with limiting diffusion resistance of the continuous phase. We assume that the concentration on the drop or bubble surface and far from it is constant and equal to C_s and C_i, respectively.

The mathematical statement of the problem on a steady-state mass transfer process in a continuous medium outside a drop is described by equation (E5.7.1.1) and the boundary conditions (E5.7.1.2)–(E5.7.1.3), where the fluid velocity components V_R and V_θ in Eq. (E5.7.1.4) are determined from the corresponding hydrodynamic problem. The mean Sherwood number, as in the case of a solid spherical particle, is computed by formula (E5.7.1.5).

The process of convective diffusion to the liquid–liquid (liquid–gas) interface substantially differs from the process of convective diffusion to the fluid–solid interface, which was considered in Section E5.7. This is due to the difference between the hydrodynamic conditions on the interfaces. The fluid velocity on the surface of a solid is always zero by virtue of the no-slip condition. On the contrary, the interface between two fluids can move, and the tangential velocity on the interface differs from zero.

Convective transfer of a substance by a moving fluid to the fluid–solid interface takes place under the condition that the flow is somewhat retarded, so that the transfer rate close to the surface is considerably lower than that in the bulk of the solution. On the contrary, diffusion to a liquid–liquid (liquid–gas) interface takes place under the more favorable conditions of nonretarded flow. That is why convective diffusion of a substance to a liquid–liquid interface is much more intensive than that to a fluid–solid interface.

We denote the mean Sherwood number for a gas bubble by Sh_b and for a solid sphere by Sh_p.

E5.8.2. Mass Transfer at Low Reynolds Numbers and Various Peclet Numbers

For spherical drops and bubbles in a translational Stokes flow (low Reynolds numbers, $\mathsf{Re} \to 0$) the fluid velocity components V_R and V_θ in Eq. (E5.7.1.4) are determined by formulas (E4.7.1.7), where $\beta = \mu_2/\mu_1$ is the drop–ambient medium dynamic viscosity ratio. (The value $\beta = 0$ corresponds to a gas bubble.)

► **Drops and bubbles at low Peclet numbers.** For $\mathsf{Pe} = 0$, which corresponds to a drop (bubble) in a quiescent fluid, the exact solution of problem (E5.7.1.1)–(E5.7.1.3), as in the case of a solid particle, has the form (E5.7.2.1), and the corresponding mean Sherwood number is unity; see Eq. (E5.7.2.2).

The three-term expansion of the mean Sherwood number as $\mathsf{Pe} \to 0$ has the form

$$\mathsf{Sh} = 1 + \frac{1}{2}\,\mathsf{Pe} + \frac{2+3\beta}{6(1+\beta)}\mathsf{Pe}^2 \ln \mathsf{Pe} + O(\mathsf{Pe}^2). \qquad \text{(E5.8.2.1)}$$

For $\beta \to \infty$, formula (E5.8.2.1) passes into formula (E5.7.2.3) for a solid particle.

► **Drops and bubbles at high Peclet numbers (the diffusion boundary layer).** For high Peclet numbers, a thin diffusion boundary layer is formed near the drop (bubble) surface.

The ratio of the thickness of this layer to the drop radius is of the order of $\mathsf{Pe}^{-1/2}$. In this region, the radial component of molecular diffusion to the drop surface is essential, and tangential diffusion may be neglected. Convective mass transfer due to the motion of the fluid should also be taken into account.

For $0 \le \beta \ll \mathsf{Pe}^{1/3}$, the concentration distribution in the diffusion boundary layer can be expressed via the complementary error function as follows:

$$c = \operatorname{erfc}\left(\frac{1}{4}\sqrt{\frac{6\,\mathsf{Pe}}{\beta+1}}\,(r-1)\,\frac{1-\cos\theta}{\sqrt{2-\cos\theta}}\right), \qquad \text{(E5.8.2.2)}$$

where the angle θ is counted off as shown in Figure E4.5.

By differentiating equation (E5.8.2.2), we obtain the dimensionless local diffusion flux to the surface of the spherical drop,

$$j = -\left(\frac{\partial c}{\partial r}\right)_{r=1} = \sqrt{\frac{3\,\mathsf{Pe}}{\pi(\beta+1)}}\,\frac{1-\cos\theta}{\sqrt{2-\cos\theta}}. \qquad \text{(E5.8.2.3)}$$

One can see that the local diffusion flux attains its maximum at the front stagnation point on the drop surface (at $\theta = \pi$) and monotonically decreases with the angular coordinate to the minimum value, which is zero and is attained at $\theta = 0$.

The mean Sherwood number for a spherical drop is given by the Levich formula

$$\mathsf{Sh} = \sqrt{\frac{2\,\mathsf{Pe}}{3\pi(\beta+1)}} = 0.461\left(\frac{\mathsf{Pe}}{\beta+1}\right)^{1/2} \qquad (\mathsf{Pe}\to\infty). \qquad \text{(E5.8.2.4)}$$

Let us also give the two-term expansion for the mean Sherwood number,

$$\mathsf{Sh} = 0.461\left(\frac{\mathsf{Pe}}{\beta+1}\right)^{1/2} + 0.41\left(\frac{3}{4}\beta+1\right) \qquad (\mathsf{Pe}\to\infty), \qquad \text{(E5.8.2.5)}$$

which refines formula (E5.8.2.4).

Formula (E5.8.2.5) can be used for practical calculations at $\mathsf{Pe} \ge 100$ for $0 \le \beta \le 0.82\,\mathsf{Pe}^{1/3} - 1$. The value $\beta = 0$ in (E5.8.2.4) and (E5.8.2.5) corresponds to a gas bubble.

► **Spherical bubbles at any Peclet numbers.** The results of the numerical solution of the problem of mass transfer to a spherical bubble in a translational flow is well approximated by the following expression for the mean Sherwood number:

$$\mathsf{Sh} = 0.6 + (0.16 + 0.213\,\mathsf{Pe})^{1/2} \qquad (\beta = 0), \qquad \text{(E5.8.2.6)}$$

whose maximum error is about 3% in the entire range of Peclet numbers.

► **Spherical drops at any Peclet numbers.** In the range $0 \le \mathsf{Pe} \le 200$, the results of numerical calculations of mean Sherwood numbers for a spherical drop in a translational flow under a limiting resistance of the continuous phase is well described by the approximate formula

$$\mathsf{Sh} = \frac{1}{\beta+1}\mathsf{Sh}_{\rm b} + \frac{\beta}{\beta+1}\mathsf{Sh}_{\rm p}, \qquad \text{(E5.8.2.7)}$$

where β is the ratio of dynamic viscosities of the drop and the ambient fluid ($\beta = 0$ corresponds to a gas bubble, and $\beta \to \infty$, to a solid sphere), and $\mathsf{Sh}_{\rm b}$ and $\mathsf{Sh}_{\rm p}$ are the

Sherwood numbers for a bubble and for a solid particle, which can be calculated by formulas (E5.8.2.6) and (E5.7.2.8), respectively.

It is important to note that the expression (E5.8.2.7) gives three valid terms of the asymptotic expansion of Sh as $\mathsf{Pe} \to 0$ for any β, see formula (E5.8.2.1).

In the interval $200 \le \mathsf{Pe} < \infty$, for any values of the phase viscosities, the mean Sherwood number for a drop can be calculated by solving the cubic equation

$$\mathsf{Sh}^3 - 0.212\frac{\mathsf{Pe}}{\beta+1}\mathsf{Sh} - (0.624)^3\,\mathsf{Pe} = 0. \qquad \text{(E5.8.2.8)}$$

E5.8.3. Mass Transfer at Moderate and High Reynolds Numbers

► **Spherical bubble at any Peclet numbers for $\mathsf{Re} \ge 35$.** For a spherical bubble in a translational flow at moderate and high Reynolds numbers and high Peclet numbers, the mean Sherwood number can be calculated by the approximate formula

$$\mathsf{Sh} = \left(\frac{2}{\pi}\mathsf{Pe}\right)^{1/2}\left(1 - \frac{2}{\sqrt{\mathsf{Re}}}\right)^{1/2}, \qquad \text{(E5.8.3.1)}$$

whose error is less than 7% for $\mathsf{Re} \ge 35$.

For $0 \le \mathsf{Pe} < \infty$ and $\mathsf{Re} \ge 35$, to calculate the mean Sherwood number for a spherical bubble, one can use the formula

$$\mathsf{Sh} = 0.6 + \left[0.16 + 0.637\left(1 - \frac{2}{\sqrt{\mathsf{Re}}}\right)\mathsf{Pe}\right]^{1/2}, \qquad \text{(E5.8.3.2)}$$

which is exact for $\mathsf{Pe} = 0$ and passes into (E5.8.3.1) as $\mathsf{Pe} \to \infty$. For $\mathsf{Re} = \infty$, the error of formula (E5.8.3.2) is about 3%.

► **Spherical drop at high Peclet numbers for $\mathsf{Re} \ge 35$.** For a spherical drop in a translational flow at moderate and high Reynolds numbers and high Peclet numbers, the mean Sherwood number is well approximated by the formula

$$\mathsf{Sh} = \left(\frac{2}{\pi}\mathsf{Pe}\right)^{1/2}\left(1 - \frac{2 + 1.49\,\beta^{0.64}}{\sqrt{\mathsf{Re}}}\right)^{1/2}, \qquad \text{(E5.8.3.3)}$$

which passes into (E5.8.3.1) for $\beta = 0$. Formula (E5.8.3.3) can be applied for $0 \le \beta \le 2$ and $\mathsf{Re} \ge 35$.

E5.8.4. General Correlations for the Sherwood Number

► **Bubbles.** For the calculation of the mean Sherwood number in laminar flows of various types past a spherical bubble, one can use the approximate formula

$$\mathsf{Sh} = 0.6 + (0.16 + \mathsf{Sh}_\infty^2)^{1/2}, \qquad \text{(E5.8.4.1)}$$

where Sh_∞ is the asymptotic value of the mean Sherwood number as $\mathsf{Pe} \to \infty$, which can be calculated by solving the diffusion boundary layer equation in a given flow field. (It is assumed that there are no closed streamlines in the continuous phase.)

Formula (E5.8.4.1) after the substitution of the value of Sh_∞ can be used for the calculation of the Sherwood number in the entire range of Peclet numbers. For a spherical bubble in a translational Stokes flow (as $\mathsf{Re} \to 0$), formula (E5.8.4.1) passes into (E5.8.2.6). The general formula (E5.8.4.1) can also be used in the case of linear shear flows.

► **Drops in the entire range of phase viscosities.** For low and moderate Peclet numbers in an arbitrary laminar flow past a spherical drop under limiting resistance of the continuous phase, it is expedient to calculate the mean Sherwood number by using formula (E5.8.2.7), where Sh_b and Sh_p are the Sherwood numbers for the limit cases of a bubble and a solid particle. These quantities can be calculated by formulas (E5.8.4.1) and (E5.7.3.2). For high Peclet numbers, in the entire range of phase viscosities, the mean Sherwood number can be found by solving the cubic equation

$$\mathsf{Sh}^3 - \mathsf{Sh}_\beta^2\mathsf{Sh} - \mathsf{Sh}_\mathrm{p}^3 = 0, \tag{E5.8.4.2}$$

where Sh_β is the asymptotic value of the mean Sherwood number obtained in the diffusion boundary layer approximation for a drop of moderate viscosity $\beta = O(1)$ as $\mathsf{Pe} \to \infty$, and Sh_p is the corresponding asymptotic value for a solid particle ($\beta = \infty$) as $\mathsf{Pe} \to \infty$. For the special case of a translational Stokes flow past a spherical drop, Eq. (E5.8.4.2) passes into (E5.8.2.8).

For Stokes flows ($\mathsf{Re} \to 0$) at high Peclet numbers, one can use the approximate formula

$$\mathsf{Sh}_\beta = \frac{\mathsf{Sh}_0}{\sqrt{\beta + 1}}, \tag{E5.8.4.3}$$

where Sh_0 is the asymptotic value of the Sherwood number for a gas bubble ($\beta = 0$) as $\mathsf{Pe} \to \infty$. For a translational or arbitrary linear straining Stokes flow past a spherical drop, formula (E5.8.4.3) is exact.

E5.9. Mass Transfer to Nonspherical Particles in Translational Flow

E5.9.1. Mass and Heat Transfer to a Particle of Arbitrary Shape at Low Peclet Numbers

► **Asymptotic formulas for the Sherwood number and diffusion flux.** For low Peclet numbers, the problem of mass exchange between a particle of arbitrary shape and a uniform translational flow were studied by Brenner (1963). The following expression was obtained for the mean Sherwood number up to first-order infinitesimals with respect to Pe:

$$\frac{\mathsf{Sh}}{\mathsf{Sh}_0} = 1 + \frac{1}{8\pi}\mathsf{Pe}_\mathrm{M}, \qquad \mathsf{Pe}_\mathrm{M} = \frac{\Pi U_\mathrm{i}}{D}, \tag{E5.9.1.1}$$

where Sh_0 is the Sherwood number corresponding to a quiescent medium. The influence of the fluid motion is characterized by the modified Peclet number, in which the shape factor Π of the particle plays the role of the characteristic length.

Formula (E5.9.1.1) is quite general and holds for solid particles, drops, and bubbles of arbitrary shape in a uniform translational flow at any Re as $\mathsf{Pe} \to 0$.

It gives a good approximation to the Sherwood number ratio for $\mathsf{Pe}_\mathrm{M} < 5$. In the special case of a spherical particle of radius a, in (E5.9.1.1) one should set $\Pi = 4\pi a$. For nonspherical particles, in (E5.9.1.1) one must use the values of Π from Table E5.3.

For a particle of arbitrary shape in a translational Stokes flow, the first three terms of the asymptotic expansion of the dimensionless total diffusion flux as $\mathsf{Pe} \to 0$ have the form

$$I = I_0 + \frac{1}{8\pi}\mathsf{Pe}\, I_0^2 + \frac{1}{8\pi}\mathsf{Pe}^2 \ln \mathsf{Pe} I_0^2(\mathbf{f}\cdot\mathbf{e}) + O(\mathsf{Pe}^2). \qquad \text{(E5.9.1.2)}$$

Here $I_0 = \Pi/a$ is the total flux to the particle in a quiescent fluid, **f** is the dimensionless vector equal to the ratio of the drag of the particle to the Stokes drag of a solid sphere of radius a (a is the characteristic length that appears in the definitions of the dimensionless variables Pe, I, and I_0), and **e** is the unit vector codirected with the fluid velocity at infinity. The Sherwood number is given by $\mathsf{Sh} = I/S$, where S is the dimensionless surface area of the particle.

To calculate $I_0 = \Pi/a$, one can use the results of Section E5.6.

► **Some special cases.** For an ellipsoid of revolution with semiaxes a and b (a is the equatorial radius) whose symmetry axis is directed along the flow, the Stokes drag is given by the formula

$$(\mathbf{f}\cdot\mathbf{e}) = \begin{cases} \dfrac{4}{3}(\chi^2+1)^{-1/2}\Big[\chi - (\chi^2-1)\operatorname{arccot}\chi\Big]^{-1} & \text{for } a \ge b, \\ \dfrac{8}{3}(\chi^2-1)^{-1/2}\Big[(\chi^2+1)\ln\dfrac{\chi+1}{\chi-1} - 2\chi\Big]^{-1} & \text{for } a \le b, \end{cases}$$

where $\chi = \big|(a/b)^2 - 1\big|^{-1/2}$.

For a body of revolution whose axis is inclined at an angle ω to the incoming flow direction (see Figure E4.6), the following formula holds in the Stokes approximation (as $\mathsf{Re} \to 0$):

$$(\mathbf{f}\cdot\mathbf{e}) = f_{\parallel}\cos^2\omega + f_{\perp}\sin^2\omega, \qquad \text{(E5.9.1.3)}$$

where $f_{\parallel}$ and $f_{\perp}$ are the values of the dimensionless drag of the body for the cases in which its axis is parallel ($\omega = 0$) and perpendicular ($\omega = \pi/2$), respectively, to the flow direction.

In particular, for a thin circular disk, one should set $f_{\parallel} = 8/(3\pi)$ and $f_{\perp} = 16/(9\pi)$ in (E5.9.1.3); for a dumbbell-like particle consisting of two adjacent spheres of equal radius, $f_{\parallel} \approx 0.645$ and $f_{\perp} \approx 0.716$.

The logarithmic term sharply restricts the practical value of the expansion (E5.9.1.2). The two-term expression (E5.9.1.1) holds in a wider range of the Peclet number (although this expression is less accurate for very small Pe).

E5.9.2. Mass Transfer to a Ellipsoidal Particle in Translational Flow

► **High Peclet numbers (the diffusion boundary layer approximations).** Consider diffusion to the surface of a solid ellipsoidal particle in a homogeneous translational Stokes flow ($\mathsf{Re} \to 0$). The particle is an ellipsoid of revolution with semiaxes a and b oriented along and across the flow, respectively. (Here b is the equatorial radius.) We introduce the following notation:

$$\chi = b/a, \quad a_{\mathrm{e}} = a\chi^{2/3}, \quad \mathsf{Pe}_{\mathrm{e}} = a_{\mathrm{e}}U_{\mathrm{i}}/D, \qquad \text{(E5.9.2.1)}$$

where a_{e} is the radius of the volume-equivalent sphere serving as the characteristic length.

The dimensionless total diffusion flux to the surface of an ellipsoidal particle at high Peclet numbers is determined by the formula

$$I = 7.85\, K(\chi)\mathsf{Pe}_{\mathrm{e}}^{1/3}, \qquad \text{(E5.9.2.2)}$$

where the shape coefficient K is given by

$$K(\chi)=\left(\frac{4}{3}\right)^{1/3}\chi^{-2/9}(\chi^2-1)^{1/3}\left(1+\frac{\chi^2-2}{\sqrt{\chi^2-1}}\arctan\sqrt{\chi^2-1}\right)^{-1/3}\quad\text{for}\quad \chi\ge 1,$$
$$K(\chi)=\left(\frac{4}{3}\right)^{1/3}\chi^{-2/9}(1-\chi^2)^{1/3}\left(\frac{2-\chi^2}{2\sqrt{1-\chi^2}}\ln\frac{1+\sqrt{1-\chi^2}}{1-\sqrt{1-\chi^2}}-1\right)^{-1/3}\quad\text{for}\quad \chi\le 1.\tag{E5.9.2.3}$$

For $\chi = 1$, we have $K = 1$, and formula (E5.9.2.2) after the division by 4π gives the Sherwood number (4.6.8) for the solid sphere.

For $0.5 \le \chi \le 3.0$, the shape coefficient is well approximated by the expression

$$K(\chi)=1+\tfrac{2}{45}(\chi-1),\tag{E5.9.2.4}$$

whose maximum error is 0.8%.

The mean Sherwood number is given by the formula $\mathsf{Sh} = I/S$, where S is the dimensionless surface area of the ellipsoid of revolution,

$$S=\frac{2\pi}{\chi^{1/3}}\left(\chi+\frac{1}{2\sqrt{\chi^2-1}}\ln\frac{\chi+\sqrt{\chi^2-1}}{\chi-\sqrt{\chi^2-1}}\right)\quad\text{for}\quad\chi\ge 1,$$
$$S=\frac{2\pi}{\chi^{1/3}}\left(\chi+\frac{1}{\sqrt{1-\chi^2}}\arcsin\sqrt{1-\chi^2}\right)\quad\text{for}\quad\chi\le 1.\tag{E5.9.2.5}$$

The dimensionless quantity S is related to the dimensional surface area S_* of the ellipsoid by $S = S_*/a_{\rm e}^2$.

► **Arbitrary Peclet numbers.** For arbitrary Peclet numbers, the mean Sherwood number (corresponding to the characteristic length $a_{\rm e}$) for a translational Stokes flow past an ellipsoidal particle can be approximated by the formula

$$\mathsf{Sh}=0.5\frac{1}{S}\left(\frac{\Pi}{a_{\rm e}}\right)+\frac{1}{S}\left\{0.125\left(\frac{\Pi}{a_{\rm e}}\right)^3+\left[7.85\,K(\chi)\right]^3\mathsf{Pe}_{\rm e}\right\}^{1/3},\tag{E5.9.2.6}$$

where the shape factor Π is given in the second and third rows of Table E5.3 and the quantities $\mathsf{Pe}_{\rm e}$, K, and S are defined in Eqs. (E5.9.2.1), (E5.9.2.3), and (E5.9.2.5), respectively.

For $0.2 \le \chi \le 5.0$, the maximum error of formula (E5.9.2.6) for an ellipsoidal particle does not exceed 10%. (The comparison used the results of the numerical solution of the corresponding axisymmetric mass and heat transfer problem by the finite-difference method.)

Formulas (E5.9.2.2) and (E5.9.2.4) cannot be used in the case of a strongly oblate ($\chi \gg 1$) or a strongly prolate ($\chi \ll 1$) ellipsoid of revolution.

E5.9.3. Mass Transfer to Particles of Arbitrary Shape

► **Bodies of revolution.** Consider a bounded body of revolution whose axis is inclined at an angle ω to the translation flow velocity at infinity (see Figure E4.6). In this case, the mean Sherwood number can be computed by the approximate formula

$$\mathsf{Sh}=\mathsf{Sh}_{\parallel}\cos^2\omega+\mathsf{Sh}_{\perp}\sin^2\omega,\tag{E5.9.3.1}$$

where $\mathsf{Sh}_{\|}$ and $\mathsf{Sh}_{\perp}$ are the mean Sherwood numbers corresponding to the cases of parallel ($\omega = 0$) and perpendicular ($\omega = \pi/2$) positions of the body in the flow, respectively.

At low Peclet numbers, for the translational Stokes flow past an arbitrarily shaped body of revolution, formula (E5.9.3.1) coincides with the exact asymptotic expression in the first three terms of the expansion. Since (E5.9.3.1) holds identically for a spherical particle at all Peclet numbers, one can expect that for particles whose shape is nearly spherical, the approximate formula (E5.9.3.1) will give good results for low as well as moderate or high Peclet numbers.

► **Particles of arbitrary shape.** For a steady-state viscous flow (without closed streamlines) past arbitrarily shaped smooth particles, one can calculate the mean Sherwood number by the approximate formula

$$\mathsf{Sh} = 0.5\,\mathsf{Sh}_0 + (0.125\,\mathsf{Sh}_0^3 + \mathsf{Sh}_\infty^3)^{1/3}. \qquad \text{(E5.9.3.2)}$$

Here the auxiliary variables Sh_0 and Sh_∞ in (E5.9.3.2) are the leading terms of the asymptotic expansions of the mean Sherwood number at small and large Peclet numbers, respectively. (In (E5.9.3.2), Sh, Sh_0, and Sh_∞ are defined on the basis of the same characteristic length.)

For spherical particles, we have $\mathsf{Sh}_0 = 1$ (the radius is taken as the characteristic length), and (E5.9.3.2) turns into (E5.7.3.2). The substitution of the values of Sh_0 and Sh_∞ corresponding to ellipsoidal particles in a translational Stokes flow into (E5.9.3.2) results in (E5.9.2.6).

For a translational Stokes flow past a convex body of revolution of sufficiently smooth shape with symmetry axis parallel to the flow, the error $\mathcal{E}$ (in percent) in formula (E5.9.3.2) for the mean Sherwood number can be approximately estimated as follows:

$$\mathcal{E} < 2\left(\frac{a}{b} + \frac{b}{a}\right),$$

where a and b are the maximum longitudinal and transverse dimensions of the particle. This estimate agrees well with the results previously described for an ellipsoidal particle.

For a particle of a given shape, the auxiliary quantities Sh_0 and Sh_∞ occurring in (E5.9.3.2) can be determined either theoretically or experimentally. In the last case, the parameter Sh_0 must be found from experiments on diffusion to the particle in a quiescent fluid. (Recall that the value Sh_0 corresponds to the dimensionless capacity of the body; the electrostatic method for measuring this capacity is widely used in electrical engineering.) For a solid particle, the asymptotics of the mean Sherwood number as $\mathsf{Pe} \to \infty$ has the form $\mathsf{Sh}_\infty = B\,\mathsf{Pe}^{1/3}$, where B is a constant. Therefore, to find the parameter B and hence Sh_∞, it suffices to carry out a single experiment at high Peclet numbers. (High Peclet numbers at low Reynolds numbers, $\mathsf{Re} < 0.5$, can easily be achieved in water solutions of glycerin.) Thus, to find Sh_0 and Sh_∞, it suffices to carry out two fairly simple experiments.

E5.9.4. Mass and Heat Transfer to a Circular Cylinder in Translational Flow

Consider diffusion to the surface of a circular cylinder of radius a in a flow with velocity U_i directed along the normal to the cylinder axis. This is a model problem used in chemical engineering for calculating mass transfer to prolate particles; it is used even more widely in mechanics of aerosols for analyzing diffusion sedimentation of aerosols on fibrous filters.

At low Reynolds numbers $\mathsf{Re} = aU_\mathrm{i}/\nu$, the analytical solution of this problem results in the following two-term expansion of the mean Sherwood number (based on the radius of the cylinder) with respect to the high Peclet number $\mathsf{Pe} = aU_\mathrm{i}/D$:

$$\mathsf{Sh} = \frac{0.580}{(2.00 - \ln 2\,\mathsf{Re})^{1/3}}\mathsf{Pe}^{1/3} + 0.0993.$$

If $\mathsf{Sc} > 0.5$, then the mean Sherwood number for cylinders of various cross-sections perpendicular to the flow direction in a wide range of Reynolds numbers can be determined by using the following formula, derived from experimental data (Kutateladze, 1990):

$$\mathsf{Sh} = A\,\mathsf{Sc}^{0.37}\mathsf{Re}^{m},$$

where the coefficients A and m are given below:

The range of Re	A	m
0.05 to 2	0.640	0.305
2 to 4	0.556	0.41
4 to 500	0.381	0.47
500 to 2.5×10^3	0.430	0.47
2.5×10^3 to 2.5×10^4	0.142	0.60
2.5×10^4 to 10^5	0.0168	0.80

E5.10. Transient Mass Transfer to Particles, Drops, and Bubbles in Steady-State Flows

E5.10.1. Statement of the Problem

We consider a laminar steady-state translational flow past a solid spherical particle (drop or bubble) of radius a and study transient mass transfer to the particle surface. At the initial time $t = 0$, the concentration in the continuous phase is constant and equal to C_i, whereas for $t > 0$ a constant concentration C_s is maintained on the particle surface.

In the particle-centered spherical coordinate system (R, θ, φ), the nonstationary problem for the concentration C in dimensionless variables comprises the convective diffusion equation

$$\frac{\partial c}{\partial \tau} + \mathsf{Pe}(\mathbf{v}\cdot\nabla)c = \Delta c \qquad \text{(E5.10.1.1)}$$

and the initial and boundary conditions

$$\tau = 0,\quad c = 0;\qquad r = 1,\quad c = 1;\qquad r \to \infty,\quad c \to 0, \qquad \text{(E5.10.1.2)}$$

where $c = (C_\mathrm{i} - C)/(C_\mathrm{i} - C_\mathrm{s})$, $\tau = Dt/a^2$, $r = R/a$, $\mathsf{Pe} = aU/D$, and U is the characteristic flow velocity (for the translational flow $U = U_\mathrm{i}$, where U_i is the fluid velocity at infinity). The steady-state flow field $\mathbf{v}$ is given.

TABLE E5.4
The expressions for $\mathsf{Sh}_{\rm st}$ in Eq. (E5.10.2.1) for spherical particles, drops, and bubbles; $\mathsf{Pe} = aU_{\rm i}/D$.

No	Disperse phase	Flow mode	$\mathsf{Sh}_{\rm st}$
1	Bubble	Translational Stokes flow ($\mathsf{Re} \ll 1$)	$\left(\dfrac{2\,\mathsf{Pe}}{3\pi}\right)^{1/2}$
2	Bubble	Laminar translational flow at high Reynolds numbers ($\mathsf{Re} \gg 1$)	$\left(\dfrac{2\,\mathsf{Pe}}{\pi}\right)^{1/2}$
3	Drop	Translational Stokes flow ($\mathsf{Re} \ll 1$)	$\left[\dfrac{2\,\mathsf{Pe}}{3\pi(\beta+1)}\right]^{1/2}$
4	Solid particle	Translational Stokes flow ($\mathsf{Re} \ll 1$)	$0.624\,\mathsf{Pe}^{1/3}$

E5.10.2. General Correlations for the Sherwood Number

► **Spherical particles and drops at high Peclet numbers.** In what follows we restrict ourselves to the case of high Peclet numbers and assume that there are no closed streamlines in the flow.

The diffusion boundary layer in problem (E5.10.1.1), (E5.10.1.2) is first adjacent to the particle surface and then rapidly spreads over the flow region with the subsequent exponential relaxation to a steady state. The characteristic relaxation time $\tau_{\rm r}$ is of the order of $\mathsf{Pe}^{-2/3}$ for a solid particle and of the order of Pe^{-1} for bubbles and drops of moderate viscosity.

The time dependence of the mean Sherwood number for an arbitrary steady-state flow past spherical particles, drops, and bubbles can be determined by the approximate formula

$$\frac{\mathsf{Sh}}{\mathsf{Sh}_{\rm st}} = \sqrt{\coth(\pi\,\mathsf{Sh}_{\rm st}^2\tau)}, \qquad \text{(E5.10.2.1)}$$

where $\mathsf{Sh}_{\rm st} = \lim\limits_{\tau\to\infty} \mathsf{Sh}$ is the Sherwood number for the steady-state diffusion mode; $\mathsf{Sh}_{\rm st}$ depends on the Peclet number and can be determined by solving the corresponding stationary problem. (For the translational flow, the stationary problem is determined by the equation and the boundary conditions (E5.7.1.1)–(E5.7.1.3).)

Equation (E5.10.2.1) gives a valid asymptotic result for any flow field in both limit cases $\tau \to 0$ and $\tau \to \infty$.

Table E5.4 presents $\mathsf{Sh}_{\rm st}$ for the translational flow past spherical particles, drops, and bubbles of radius a. The parameter β is the ratio of the dynamic viscosity of the drop to that of the ambient fluid and varies in the range $0 \le \beta \le 2$ (the value $\beta = 0$ corresponds to a gas bubble).

For example, in the case of nonstationary mass transfer in a steady-state translational Stokes flow past a spherical drop with limiting resistance of the continuous phase, the steady-state value $\mathsf{Sh}_{\rm st}$ is presented in the third row of Table E5.4. By substituting this value into (E5.10.2.1), we obtain

$$\mathsf{Sh} = \left[\frac{2\,\mathsf{Pe}}{3\pi(\beta+1)}\coth\left(\frac{2}{3}\,\frac{\mathsf{Pe}\tau}{\beta+1}\right)\right]^{1/2}. \qquad \text{(E5.10.2.2)}$$

The maximum error of Eq. (E5.10.2.2) for $0 \le \tau < \infty$ does not exceed 0.7%.

Formula (E5.10.2.1) gives good results for the mean Sherwood number for linear shear flows past spherical particles, drops, and bubbles at high Peclet numbers. (In the diffusion boundary layer approximation, the maximum error of Eq. (E5.10.2.1) for all τ does not exceed 1.8% in any case.)

► **Spherical particles and drops at arbitrary Peclet numbers.** To calculate the mean Sherwood number for an arbitrary laminar flow past spherical particles, drops, and bubbles in the entire range of Peclet numbers, one can use the interpolation formula

$$\mathsf{Sh} = (\mathsf{Sh}_{\rm st} - 1)\sqrt{\coth\left[\pi(\mathsf{Sh}_{\rm st} - 1)^2\tau\right]} + 1. \qquad \text{(E5.10.2.3)}$$

Consider the behavior of this function in various limit cases. Since $\mathsf{Sh}_{\rm st} \to 1$ as $\mathsf{Pe} \to 0$, we see that Eq. (E5.10.2.3) yields the exact result (E5.6.2.4) for a quiescent medium. As $\mathsf{Pe} \to \infty$, we have $\mathsf{Sh}_{\rm st} \to \infty$, and the expression (E5.10.2.3) passes into (E5.10.2.1). For small τ, Eq. (E5.10.2.3) yields the exact result $\mathsf{Sh} \approx (\pi\tau)^{-1/2}$. As $\tau \to \infty$, we have $\mathsf{Sh} \to \mathsf{Sh}_{\rm st}$, which follows from (E5.10.2.3).

Bibliography for Chapter E5

Acrivos, A. and Goddard, J. D., Asymptotic expansions for laminar forced-convection heat and mass transfer. Part 1. Low speed flows, *J. Fluid Mech.,* Vol. 23, No. 2, pp. 273–291, 1965.

Bird, R. B., Stewart, W. E., and Lightfoot, E. N., *Transport Phenomena*, Wiley, New York, 1965.

Brenner, H., Forced convection-heat and mass transfer at small Peclet numbers from particle of arbitrary shape, *Chem. Eng. Sci.,* Vol. 18, No. 2, pp. 109–122, 1963.

Carslow, H. S. and Jaeger, J. C., *Conduction of Heat in Solids*, Pergamon Press, New York, 1959.

Clift, R., Grace, J. R., and Weber, M. E., *Bubbles, Drops and Particles*, Acad. Press, New York, 1978.

Deavours, C. A., An exact solution for the temperature distribution in parallel plate Poiseuille flow, *Trans. ASME, J. Heat Transfer*, Vol. 96, No. 4, 1974.

De Kee, D. and Chhabra, R. P., *Transport Processes in Bubbles, Drops and Particles*, Hemisphere, New York, 1992.

Eckert, E. R. G. and Drake, R. M., *Analysis of Heat and Mass Transfer*, Hemisphere Publ., New York, 1987.

Friedlander, S. K., Mass and heat transfer to single spheres and cylinders at low Reynolds numbers, *AIChE J.,* Vol. 3, No. 1, pp. 43–48, 1957.

Graetz, L., Über die Warmeleitungsfähigkeit von Flüssigkeiten, *Annln. Phys.,* Bd. 18, S. 79–84, 1883.

Grigull, U. and Sandner, H., *Heat Conduction*, Hemisphere Publ., New York, 1984.

Gupalo, Yu. P., Polyanin, A. D., and Ryazantsev, Yu. S., *Mass and Heat Exchange Between Reacting Particles and Flow* [in Russian], Nauka, Moscow, 1985.

Hewitt, G. F., Shires, G. L., and Bott, T. R., *Process Heat Transfer*, Begell House, New York, 1994.

Hewitt, G. F. and Spalding, D. B., *Encyclopedia of Heat and Mass Transfer*, Hemisphere, New York, 1986.

Incropera, F. P. and Dewitt, D. P., *Fundamentals of Heat and Mass Transfer*, Wiley, New York, 1996.

Kader, B. A. and Yaglom, A. M., Heat and mass transfer laws for fully turbulent wall flows, *Int. J. Heat Mass Transfer*, Vol. 15, No. 12, pp. 2329–2351, 1972.

Kutateladze, S. S., *Heat Transfer and the Hydrodynamical Resistance. A Reference Book* [in Russian], Energoatomizdat, Moscow, 1990.

Levich, V. G., *Physicochemical Hydrodynamics*, Prentice-Hall, Englewood Cliffs, New Jersey, 1962.

Nusselt, W., Abhängigkeit der Wärmeübergangzahl con der Rohrlänge, *VDI Zeitschrift*, Bd. 54, Ht. 28, S. 1154–1158, 1910.

Oellrich, L., Schmidt-Traub, H., and Brauer, H., Theoretische Berechnung des Stofftransport in der Umgebung einer Einzelblase, *Chem. Eng. Sci.,* Vol. 28, No. 3, pp. 711–721, 1973.

Patankar, S. V. and Spalding, D. B., *Heat and Mass Transfer in Boundary Layers*, Morgan-Grampian, London, 1967.

Petukhov, B. S., *Heat Transfer and Drag in Laminar Flow of Liquids in Tubes* [in Russian], Energiya, Moscow, 1967.

Polyanin, A. D. and Dilman, V. V., *Methods of Modeling Equations and Analogies in Chemical Engineering*, CRC Press–Begell House, Boca Raton, Florida, 1994.

Polyanin, A. D., Kutepov, A. M., Vyazmin, A. V., and Kazenin, D. A., *Hydrodynamics, Mass and Heat Transfer in Chemical Engineering*, Taylor & Francis, London, 2002.

Reid, R. C., Prausnitz, J. M., and Sherwood, T. K., *The Properties of Gases and Liquids*, McGraw-Hill Book Comp., New York, 1977.

Rohsenow, W. M., Hartnett, J. P., and Ganic, E. N., *Handbook of Heat Transfer Fundamentals*, McGraw-Hill, New York, 1985.

Rotem, Z. and Neilson, J. E., Exact solution for diffusion to flow down an incline, *Can. J. Chem. Engng.*, Vol. 47, pp. 341–346, 1966.

Sherwood, T. K., Pigford, R. L., and Wilke, C. R., *Mass Transfer*, McGraw-Hill, New York, 1975.

Sparrow, E. M. and Ohadi, M. M., Numerical and experimental studies of turbulent heat transfer in a tube, *Numerical Heat Transfer*, Vol. 11, No. 4, pp. 461–476, 1987.

Vargaftic, N. V., Vinogradov, Y. K., and Yargin, V. S., *Handbook of Physical Properties of Liquids and Gases*, Begell House Inc., New York, 1996.

Zapryanov, Z. and Tabakova, S., *Dynamics of Bubbles, Drops and Rigid Particles*, Kluwer Acad. Publ., Dordrecht, 1999.

Chapter E6

Electrical Engineering

E6.1. Generalities

E6.1.1. Introduction. Basic Notions, Definitions, and Notation

► **Preliminary remarks.** *Electrical engineering* is a technical science studying physical phenomena related to directed motion of charged particles in conducting media at the level of macroprocesses and the problems of practical use of such phenomena. Electrical engineering is traditionally divided into the following fields: foundations of circuit theory, foundations of magnetic circuits, electric devices and transformers, electric machines, and foundations of electronics.

The basic notions of electrical engineering coincide with similar notions in the course of physics. Electrical engineering employs the following physical quantities in the sense explained below.

► **Notions, definitions, and notation.** The *potential* is the work done by electric forces to transfer the unit charge from a given point in space, in particular, a given point of the circuit to the point at infinity (the zero potential point). The potential is denoted by the symbol φ and is measured in volts (V). The potential can be determined for a given point in space (circuit) with respect to the zero potential.

The *voltage* (*potential difference*) is the work done by electric forces to transfer the unit charge from a point in space (circuit) to another point. The voltage is also measured in volts (V) and denoted by the symbol U. The voltage can be determined only between two points in space (circuit).

The *current strength* (*current*) is the electric charge passing through the conductor transverse cross-section per unit time. The current strength is measured in amperes (A) and is denoted by I. The current strength can be determined for a given element of the electric circuit.

The *electromotive force* (EMF) is the work done by forces of nonelectrostatical nature (extraneous forces) to transfer the unit charge from a point in space (circuit) to another point. The EMF is also measured in volts (V) and denoted by E. The EMF can be determined only for two points in space (circuit).

The *resistance* (*electric resistance*) is a physical quantity characterizing the property of bodies or media with current flowing in them to transform the electric energy into the heat energy of particles of these bodies or media. The resistance is measured in ohms (Ω) and is denoted by R. The reciprocal of resistance is called the conductance and is denoted by g. The SI unit of conductance is the siemens (S), 1 S = $1/\Omega$.

The *capacitance* is the physical quantity characterizing the property of bodies to accumulate electric field energy. The SI unit of capacitance is the farad (F).

The *inductance* is the physical quantity characterizing the property of bodies with current flowing in them to accumulate the magnetic field energy. The SI unit of inductance is the henry (H).

The *power* is the work done in electric circuits by force of electrical and nonelectrical nature per unit time. The SI unit of power is the watt (W).

E6.1.2. Perfect Elements. Active and Passive Elements

► **Perfect elements. Notation conventions for perfect elements.** Electrical engineering operates with ideal or perfect elements. The idealization consists in that each of the perfect elements has only one of the properties inherent in actual elements. Any perfect element is an abstraction, i.e., a convenient model that with certain accuracy reflects the processes and phenomena occurring in actual elements.

In electrical engineering, the following perfect elements are studied.

A *conductor* is a perfect element in which the electric energy is neither transformed nor dissipated nor accumulated. The conductor has neither resistance nor inductance. The voltage at the ends of a perfect conductor is always zero. The connection of two points of a perfect conductor is called the *short circuit*.

A *disconnection* is a perfect element in which the electric energy is neither transformed nor dissipated nor accumulated. The disconnection has no capacitance, and the disconnection resistance and inductance can be assumed to be infinitely large. The current through a disconnection is always zero. The disconnection between two points of an electric circuit is called the *open circuit*.

A *switch* is a perfect element having two stable states in one of which it is a perfect conductor and in the other it is a disconnection. The transition from one state into another occurs in an infinitely small time interval.

A *resistor* is a perfect element that has only a finite value of resistance. The inductance of a perfect resistor is zero. In the resistor, the electric energy is dissipated; i.e., it is irreversibly transformed into the heat energy.

A *capacitor* is a perfect element that has only a finite value of capacitance. The resistance of a perfect capacitor is infinitely large, and its inductance is zero. In the capacitor, the electric field energy is accumulated.

A *coil* (*induction coil*) is a perfect element that has only a finite value of inductance. The resistance of ideal induction coil is zero. In the induction coil, the magnetic field energy is accumulated.

An *ideal EMF source* is an active element such that the voltage at its terminals is equal to the value of its electromotive force. This quantity is, in particular, independent of the current flowing through the EMF source. The short circuit problem for an ideal EMF source is ill posed.

The *ideal source of current* is an active element such that the strength of the current flowing through it is always equal to the current of the source. The current of the source is, in particular, independent of its voltage. The open circuit problem for the current source is ill posed.

By default, all the elements are assumed to be perfect. Any perfect element in electrical circuits (networks) is indicated by the symbol of its typical parameter. If it is necessary to distinguish several elements, then the parameters are equipped with corresponding indices.

The notation conventions for perfect elements are shown in Fig. E6.1.

Unless otherwise specified, it is assumed that the geometric dimensions of perfect elements are zero.

The direction from minus to plus at the ideal EMF source shows the direction of extraneous forces or the direction of increase in the potential. The voltage on such a source is equal to E.

A perfect element is said to be linear if its characteristic elements are independent of the current flowing in the element and of its voltage. By default, it is assumed that the elements are linear.

The mathematical model of an actual element can be constructed from perfect elements with any desired accuracy of details.

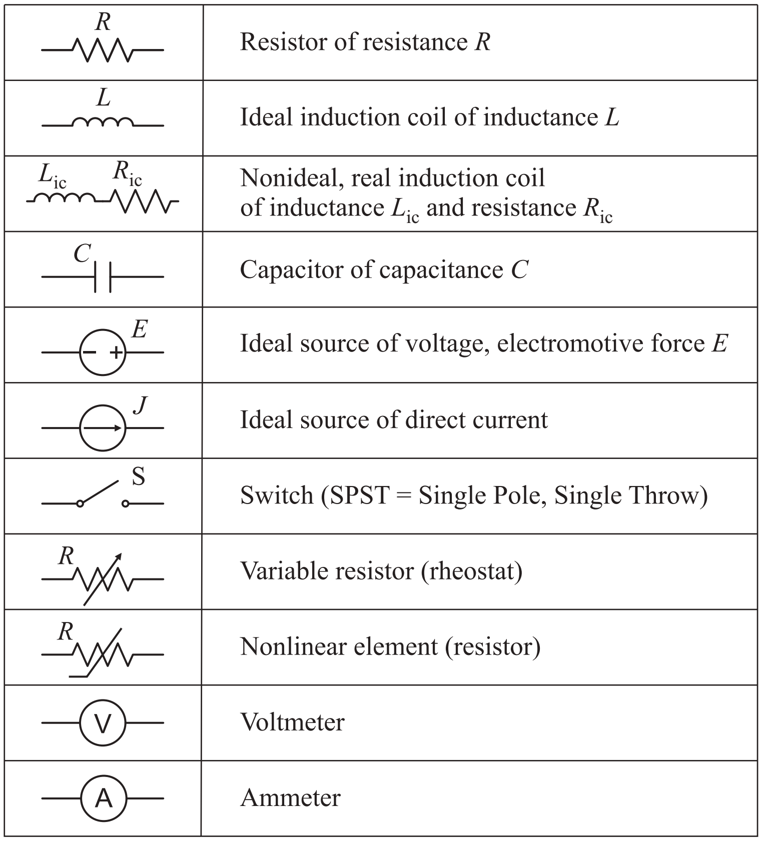

R	Resistor of resistance R
L	Ideal induction coil of inductance L
L_{ic} R_{ic}	Nonideal, real induction coil of inductance L_{ic} and resistance R_{ic}
C	Capacitor of capacitance C
− + E	Ideal source of voltage, electromotive force E
J	Ideal source of direct current
S	Switch (SPST = Single Pole, Single Throw)
R	Variable resistor (rheostat)
R	Nonlinear element (resistor)
V	Voltmeter
A	Ammeter

Figure E6.1. Notation conventions for perfect elements.

► **Active and passive elements.** The perfect elements are divided into active and passive elements according to the character of the energy transformation in them.

In passive elements, only the energy accumulation (inductance and capacitance) or its dissipation (irreversible transformation of electric energy into heat in resistors) may occur. In several passive elements, energy is neither accumulated nor transformed at all (conductor, disconnection, switch).

In active elements, energy is transformed from one type into another. An EMF source and a current source are active elements. The energy transformation in active elements is always reversible.

E6.1.3. Electric Circuits and Networks. Classification of Electric Circuits and Their Operation Modes

► **Electric circuits and networks.** An *electric circuit* is an electrical engineering device that exists or may exist in reality. An electric circuit consists of actual elements.

A *network* (*electrical network*) is a certain graphical representation of an actual device, which is constructed only of perfect elements.

An *equivalent circuit* of an electric circuit consists of a set of different idealized elements chosen so that it were possible to describe the processes in the circuit with a prescribed or desired accuracy.

► **Classification of electric circuits and their operation modes.** The electric circuits are classified according to the characteristics of the elements composing them as follows:

linear and nonlinear circuits, circuits with lumped parameters, and circuits with distributed parameters. For a circuit to be referred to as nonlinear, it suffices to contain at least one nonlinear element. For a circuit to be referred to as a circuit with distributed parameters, it suffices that the linear dimensions of at least one element are commensurable with the wavelength of the electromagnetic field arising in this circuit.

The operation modes of electric circuits are classified according to the character of variations in the input and output signals (currents and voltages) in time. The operation modes are divided into *stationary*, *quasistationary* (*periodic*), and *nonstationary* (*transient*).

E6.1.4. Structure Elements of Network. Notion of the Graph of a Network

► **Structure elements of network.** In any electrical network, the following structural elements can be distinguished: branch, node, loop.

An *branch* is a series connection of several perfect elements. A branch can consist of a single element if it is not a conductor or a disconnection. A series connection is a connection of perfect elements in which the "output" of one element is the "input" of one and only one subsequent element. The branch origin is the "input" of the first of the elements, and the branch end is the "output" of the last of the perfect elements. The origin and the end of a branch are, as a rule, conditional notions. It is convenient to denote all perfect elements belonging to the same branch by the same indices. It follows from the charge conservation law that the current strength is the same for all elements of the branch.

A *node* is a point common for three or more branches. For each of the branches belonging to (incident on) a given node, the node can be only the origin or the end of the branch. It follows from the definition of electric potential that the potentials of all points forming a given node are the same.

An *loop* is any closed path in an electric circuit network such that none of the branches and none of the nodes are encountered twice when going around this path. A loop may be dependent and independent.

► **Graph of a network.** The *graph of a network* is a graph such that the set of its vertices coincides with the set of the network nodes and the set of its edges coincides with the network branches. The set of independent loops in a network is the set of fundamental cycles in the graph of the network. The elements of a graph are branches, vertices, nodes, and loops. A branch is the current path between two vertices, and the current strength in all transverse cross-sections of the branch at a given time moment is constant. A node is a vertex at which three or more branches meet. A loop is a closed path beginning and terminating at the same vertex. The number of loops in a network may be rather large, but still, there is a number N of linearly independent loops, which can be determined by the formula

$$N = B - K + 1,$$

where B is the number of branches and K is the number of nodes.

E6.1.5. Physical Laws for Structure Elements

For the basic structure elements of any electrical network, there exists its own physical law:

Ohm's law	(for branches),
Kirchhoff's first law	(for nodes),
Kirchhoff's second law	(for closed loops).

► **Ohm's law.** *Ohm's law* is formulated as follows: the current strength on a part of the circuit is directly proportional to the voltage applied to this part and is inversely proportional to the resistance of this part.

Ohm's law can be interpreted for a part of the circuit containing an EMF source E (the generalized Ohm law) as follows: the current of a part of the circuit is equal to the algebraic sum of its voltage and EMF divided by the resistance of the part. The EMF and the voltage are taken with positive sign if their directions coincide with the current direction and are taken with negative sign if their directions are opposite to the current direction. The four situations shown in Fig. E6.2 are possible.

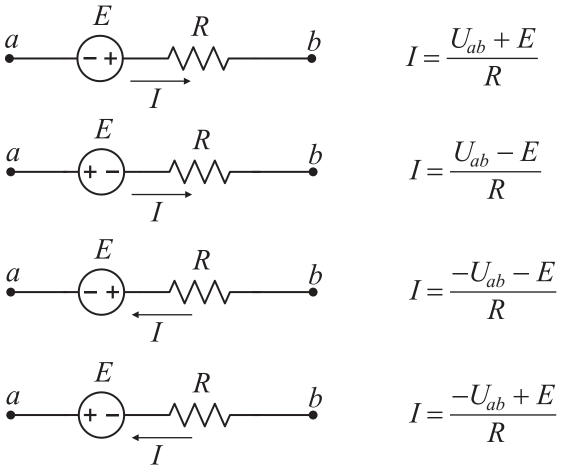

Figure E6.2. Generalized Ohm law for different parts of the circuit.

► **Kirchhoff's laws.** *Kirchhoff's first law* is formulated as follows:

$$\sum_{k=1}^{n} I_k = 0 \qquad \text{(the algebraic sum of currents at a node is zero).}$$

Here it is assumed that the currents flowing into a node are taken with the sign "+", and the currents flowing out of a node, with the sign "–".

Kirchhoff's second law can be formulated as follows:

$$\sum_{k=1}^{n} E_k = \sum_{l=1}^{m} R_l I_l,$$

where $\sum_{k=1}^{n} E_k$ is the algebraic sum of EMF E_k in a closed loop, and a term in this sum is taken with sign "+" if the directions of the source EMF and of an arbitrary chosen path tracing coincide with each other, otherwise, the sign "–" is taken; $\sum_{l=1}^{m} R_l I_l$ is the algebraic sum of voltages in the same loop, and a term of this sum is taken with sign "+" if the directions of current voltage and of an arbitrary chosen path tracing coincide with each other; otherwise, the sign "–" is taken.

E6.1.6. Problems of Analysis and Synthesis of a Network. An Analysis of a Network by Physical Laws. Potential Diagram

► **Problems of analysis and synthesis of a network.** In electrical engineering, the problems of network analysis and synthesis are distinguished. The analysis problem is a problem of calculating currents in branches of a network, of voltages between nodes, of powers, etc. from the well-known structure of the circuit and the parameters of its elements. The analysis problem is generally solved by constructing an equivalent circuit that, with a desired accuracy, reflects the circuit element properties important in the problem under study and the circuit structure. Further, the problem is solved by one of the methods for analyzing the circuits, which are considered below. The choice of any method is based on the following factors: the dimension of the obtained system of equations, the simplicity of algorithmization, and the restricted possibilities of using the method.

The circuit synthesis problem involves the construction of a network ensuring the prescribed values of one or several parameters for one or several elements under certain restrictions.

► **An analysis of a network by Ohm's and Kirchhoff's laws.** An analysis of any correctly described electric circuit can be performed by using only the three physical laws listed in Subsection E6.1.5. In general, to solve the problem, it is necessary to write out the equations of Kirchhoff's second law for all independent loops and of Kirchhoff's first law for all nodes except one arbitrary node. If necessary, the generalized Ohm law is written for separate branches. The system of linear equations thus obtained can be solved by any method.

The order of calculations may be as follows:

1. The positive directions of currents in all branches are chosen and indicated by arrows.
2. The positive directions of path tracing are chosen.
3. The system of equations according to Kirchhoff's laws is constructed: (i) by the first law, we obtain $N_1 = K - 1$ equations, (ii) by the second law, we have $N_2 = B - K + 1$ equations, where B is the number of branches and K is the number of nodes (in the whole, we obtain $N_1 + N_2 = B$ equations).

Example. As an example, consider the electrical network shown in Fig. E6.3. We choose the directions of the currents and the directions of the path tracing as is shown in Fig. E6.3. Then we obtain the system of equations:

$$\begin{aligned} &I_1 - I_2 + I_3 = 0, \\ &E_1 + E_2 = I_1R_1 + I_2R_2, \\ &E_2 + E_3 = I_2R_2 + I_3R_3. \end{aligned}$$

The unknowns in this system of equations may be any three quantities, and the others must be prescribed. Usually, the unknowns are the currents in the circuit branches. If it is necessary to introduce the voltage U_{AB} into the system of equations, then it is also required to write out the generalized Ohm law for any of the three branches.

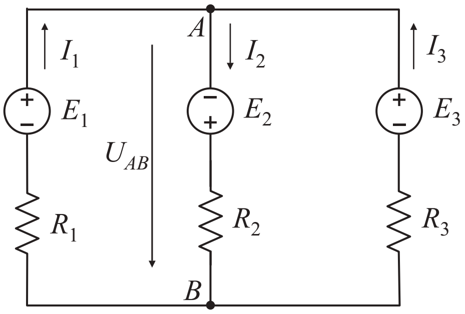

Figure E6.3. Electrical network.

Using Ohm's and Kirchhoff's laws, one can calculate the operation mode of any electric circuit. But the order of the system of equations can be large. The system can be simplified by different methods, and all of them are based on Kirchhoff's and Ohm's laws. A method for solving the problems is chosen starting from the complexity of the network and of given parameters.

► **Distribution of the potential along an unbranched (unramified) electric circuit.** The *potential diagram* is the distribution of electric potentials along parts of the circuit of a closed loop, namely, $\varphi(R)$.

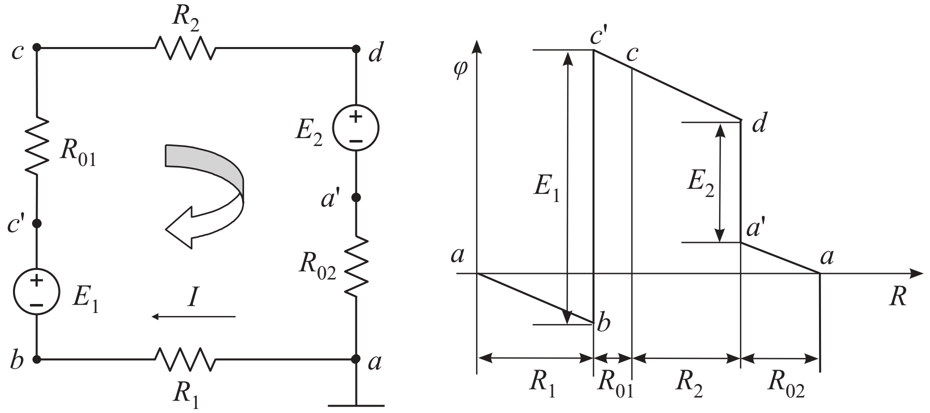

Figure E6.4. Electrical network and the potential diagram; R_{01} and R_{02} stand for the internal resistances of the EMF E_1 and E_2, respectively.

We illustrate the diagram construction with an example of a simple unbranched circuit shown in Fig. E6.4. Assume that the direction of the path tracing coincides with the direction of the current. To construct the potential diagram, we assume that one (any) of the potentials of the closed loop is zero. If the EMF E_1 exceeds the EMF E_2, then, in this case, we have

$$I = \frac{E_1 - E_2}{R_1 + R_{01} + R_2 + R_{02}}.$$

Let $\varphi_a = 0$; then

$$\varphi_b = \varphi_a - R_1 I, \quad \varphi_c = \varphi_b + E_1 - R_{01} I, \quad \varphi_d = \varphi_c - R_2 I, \quad \varphi_a = \varphi_d - E_2 - R_{01} I = 0.$$

In these expressions, the value of the EMF source voltage E is taken with sign "+" if the sense of the path coincides with the direction of E; otherwise, the sign "–" is taken. The value of the resistor voltage is taken with the sign "+" if the sense of the path is opposite to that of the current; otherwise, the sign "–" is taken.

E6.2. Linear Circuits of Direct Current

E6.2.1. Power of Direct Current

To estimate the energy conditions, it is important to know how fast the work is done; i.e, it is important to determine the power: $P = UI$.

For resistive elements, this expression can be transformed using Ohm's law as follows: $P_R = UI = I^2 R = U^2/R$.

For the EMF source whose direction coincides with the direction of the current, the power of extraneous forces is determined as $P_E = EI$. But if the current and EMF directions are opposite (for example, in the case of battery charging), then the power of extraneous forces is: $P_E = -EI$. In this case, the EMF source obtains energy from the external circuit.

The power of extraneous forces of the current source is equal to $P_J = U_{ab}I = U_{ab}J$. If $U_{ab} < 0$, then the source gains energy from the external circuit.

In any electric circuit, the energy balance—i.e., the power balance—must be satisfied, which means that the algebraic sum of powers of all energy sources must be equal to the algebraic sum of powers of all energy receivers:

$$\sum P_E = \sum P_R.$$

Example. Let us compose the energy balance for the network shown in Fig. E6.5. We obtain

$$U_{ab}J + E_1 I_1 = I_1^2 R_1 + I_2^2 R_2.$$

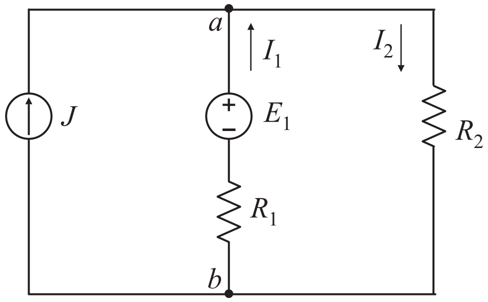

Figure E6.5. Electrical network illustrating the energy balance.

E6.2.2. Notion of Nonideal Source. Equivalent Circuits and External Characteristics

► **Nonideal source. Equivalent circuits.** A nonideal source of energy with known EMF E and internal resistance R_0 can be represented by two basic equivalent networks (equivalent circuits): the voltage source (Fig. E6.6a) or the current source (Fig. E6.6b).

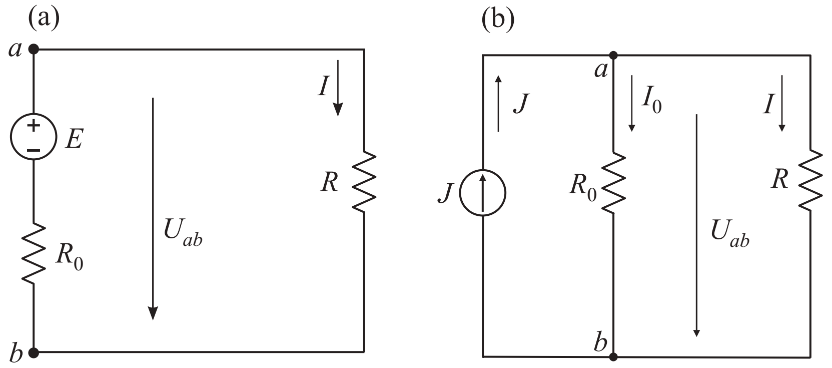

Figure E6.6. (a) Equivalent circuit of voltage source and (b) current source.

The representation of actual electric energy sources in the form of two equivalent circuits is an equivalent representation with respect to the exterior part of the circuit (receiver); i.e., in both cases, the voltages between the source terminals are the same (U_{ab}). But the energy relations in the two equivalent circuits are not the same. There is no equality between the power of the EMF sources and the current source $EI \neq U_{ab}I$ and the loss power $I^2R_0 \neq I_0^2R_0$.

► **External characteristic and operation modes of an EMF source.** According to Ohm's law, the EMF source terminal voltage takes the form

$$U_{ab} = E - IR_0.$$

The dependence between the voltage U and the current I flowing through the EMF source is called the *external characteristic* of the EMF source or the volt-ampere characteristic (VACh). The external characteristic of the source $U(I)$ completely determines the properties of this source and, for a majority of actual sources, this characteristic, can be represented as a straight line (see Fig. E6.7).

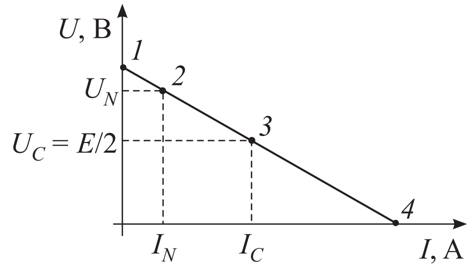

Figure E6.7. External characteristic of the EMF source.

Of all possible operation modes of an EMF source, we note the following four most important modes:

point 1: open circuit mode,
point 2: nominal (typical) operation mode,
point 3: consistent operation mode,
point 4: short circuit mode.

The *open circuit mode* (point 1) is an operation mode in which the receiver resistance is infinitely large ($R = \infty$). In practice, this corresponds to disconnection of an electric circuit, hence the open-circuit current is zero ($I_* = 0$), and hence $U_* = E$. In the open circuit tests, the EMF E can be determined from the readings of a voltmeter connected directly to the open-circuit terminals of the EMF source E.

The *nominal mode* (point 2) of a source is characterized by the fact that its voltage, current, and power correspond to the values calculated by the manufacturing departments. In this case, the best operation conditions are guaranteed for the EMF source (efficiency, operating life, etc.). The quantities determining the nominal mode are usually indicated in the certificate, in the catalog, or on the device panel. In the nominal mode, $U_N = E - R_0 I_N$.

In practice, the nominal mode is attained if U_N lies in the limits $0.90E \le U_N \le 0.95E$.

The *consistent mode* (point 3) is an operation mode in which the source transfers the maximal power $P_{\max}$ to the receiver (to the external circuit). To achieve this operation mode, it is necessary to choose a value of the receiver resistance R equal to the internal resistance of the source R_0, i.e., $R = R_0$.

The *short circuit mode* (point 4) is characterized by the fact that the receiver resistance becomes zero $R = 0$. As a rule, in practice, this is related to closing the electric energy receiver by a conductor of a very small resistance, the so-called industrial short circuit, often caused by a lot of abuse of electrical engineering devices (electric motors, transformers, home equipment, etc.).

For $R = 0$, we have $U' = RI' = 0$, and then the short circuit current is equal to $I' = E/R_0$. Because of a small internal resistance of the EMF source E, the short circuit current (I') can attain very large values significantly exceeding the value of nominal currents ($I' \gg I_N$). Therefore, the short circuit mode, as a rule, is very dangerous and undesirable for both the source and the receiver. But in several cases, the short circuit mode is nominal for electrical engineering devices such as the welding transformers and generators, current transformers, etc.

The parameters of the EMF source E can be determined graphically by using its external characteristic.

E6.2.3. Problem of Maximal Power Transfer. The Source Efficiency Factor

For the receiver to deliver the maximal power $P_{\max}$, it is necessary that the receiver operate with the EMF source E in the consistent operation mode. To achieve the consistent mode, it is necessary to choose the receiver resistance R so that it is equal to the internal resistance R_0 of the EMF source E.

Indeed, the power delivered by the receiver is equal to $P = RI^2$, where $I = \dfrac{E}{R + R_0}$. Therefore,

$$P = \frac{RE^2}{(R + R_0)^2}.$$

Let us analyze this formula. If $R = 0$, then $P = 0$; if $R = \infty$, then $P = 0$. Therefore, the function $P = P(R)$ has an extremum. One can show that this function attains its maximum for $R = R_0$. The corresponding power delivered by the receiver is calculated as

$$P_{\max} = RI_C^2 = \frac{R_0 E^2}{(R_0 + R_0)^2} = \frac{E^2}{4R_0},$$

and the power delivered by the source is equal to

$$P_E = EI_C = E\frac{E}{R_0 + R_0} = \frac{E^2}{2R_0}.$$

The source efficiency factor in this operation mode is equal to

$$\eta = \frac{P_{\max}}{P_E} = 0.5;$$

i.e., 50% of energy the source is transferred to the receiver, and the other 50% are lost in the source itself (due to its internal resistance).

The consistent mode is admissible and desirable in systems of telemechanics, automatics, telephony, etc., where it is necessary to distinguish the maximal possible power $P_{\max}$ in the receiver; the 50% losses in the source can be neglected, because the power consumption in such systems is small of the order of several watts.

In power-supply systems, such an operation mode is impossible, because the powers of the energy sources is of the order of megawatts (MW) and the 50% losses in the sources are extremely large.

E6.2.4. Theorem on Active Impedor (Two-Terminal Network) and Its Application

► **Theorem on active impedor.** The electric circuit (Fig. E6.8) can be divided into two parts: the first contains a branch (e.g., resistance R) in which it is required to find the current I, the second—i.e., the other part of the circuit—contains an active impedor.

The theorem on an active impedor can be formulated as follows: if the active circuit with a branch connected to it is replaced by a source with EMF E equal to the voltage at the terminals of the opened branch and with the resistance equal to the input resistance of active circuit, then the current in this branch does not change.

An active impedor with respect to the resistance R can be treated as an equivalent generator with EMF E and the internal resistance R_0, which permits determining the current flowing through the resistance R as follows: $I = E/(R + R_0)$.

Sometimes, the theorem on active impedor is also called the equivalent generator method or the open-circuit and short-circuit method, the latter name explains the essence of this method.

The unknown values of EMF E and the internal resistance R_0 of an equivalent generator can easily be determined by open circuit and short circuit tests. From the open circuit test we determined U_* and the equivalent generator EMF $E = U_*$.

The impedor internal resistance can be determined in the following two ways. First, one can calculate the resistance of the entire electrical circuit without the resistor R with respect to the terminals of the active impedor for the bridged EMF sources E with their internal resistances taken into account and for the open current sources. Second, one can first find the short circuit current by setting that the resistance R is zero and then determine the internal resistance by the formula $R_0 = E/I'$.

► **Example of applying the theorem on active impedor.** We illustrate the use of the theorem on active impedor by an example for the network shown in Fig. E6.8a. In this network, we choose the positive direction of the desired current I_1. Then we consider the network part connected to the first branch under study (enclosed in the dashed line) as an active impedor with EMF E_* and internal resistance R_0 (see Fig. E6.8b). In the network, the direction of EMF E_* to point A is chosen arbitrarily; then $U_{AB} = E_*$. The full-scale network in open circuit mode is shown in Fig. E6.8c.

In the impedor internal branches, the current is calculated by the formula

$$I_* = \frac{E_2}{R_2 + R_3 + R_{02}}.$$

The open-circuit voltage determines the EMF of the active impedor as follows:

$$U_{AB} = E_* = R_3 I_*.$$

To calculate the impedor internal resistance, we transform the network replacing the source E_2 by a short-circuited part (see Fig. E6.8d). The network input resistance is the internal resistance of the active impedor:

$$R_0 = \frac{(R_2 + R_{02})R_3}{R_2 + R_3 + R_{02}}.$$

Returning to the original network, we determine the desired current by Ohm's law:

$$I_1 = \frac{E_* + E_1}{R_1 + R_{01} + R_0}.$$

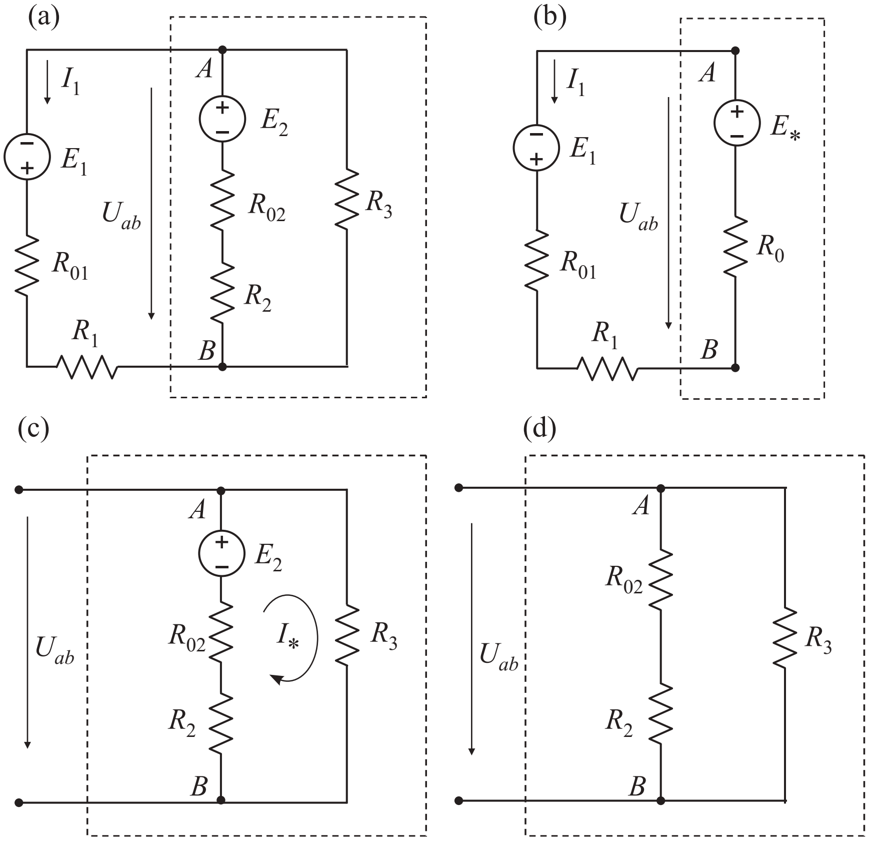
(a)
I_1
E_1
U_{ab}
R_{01}
R_1
A
E_2
R_{02}
R_2
B
R_3
(b)
I_1
E_1
U_{ab}
R_{01}
R_1
A
E_*
R_0
B
(c)
U_{ab}
A
E_2
R_{02}
I_*
R_3
R_2
B
(d)
U_{ab}
A
R_{02}
R_3
R_2
B

Figure E6.8. Electrical networks illustrating the active impedor method.

E6.2.5. Two Compensation Theorems. Properties of Linear Circuits: Linearity Theorem and Reciprocity Theorem

► **Compensation theorems.** The *voltage compensation theorem* (*voltage compensation principle*) says that the network part ab with voltage U_{ab} can be replaced by an equivalent EMF source $E = U_{ab}$ whose direction is opposite to the positive direction of the voltage U_{ab} (see Fig. E6.9). In this case, the current distribution in the network does not change.

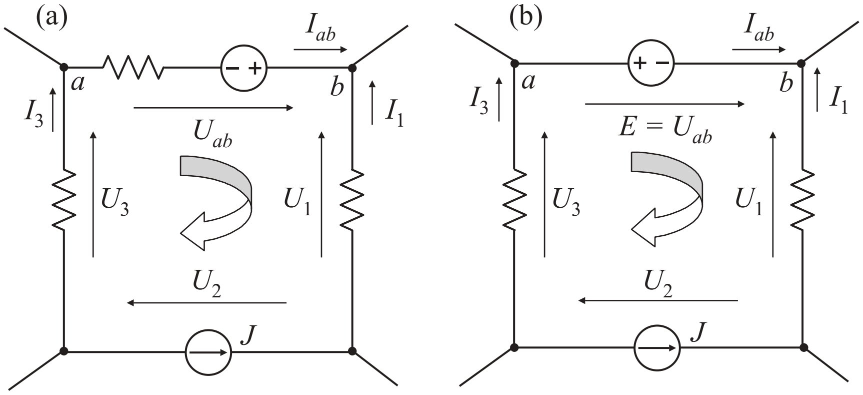
(a)
I_{ab}
a
b
I_3
I_1
U_{ab}
U_3
U_1
U_2
J
(b)
I_{ab}
a
b
I_3
I_1
$E = U_{ab}$
U_3
U_1
U_2
J

Figure E6.9. (a) Initial network and (b) transformed network.

The *current compensation theorem* (*current compensation principle*) says that the network part ab with current I_{ab} can be replaced by an equivalent current source $J = I_{ab}$ whose direction coincides with the positive direction of the current I_{ab}.

► **Linearity theorem and reciprocity theorem.** *Linearity theorem*: if, in any linear electric circuit, there are variations in the EMF E or in the resistance, then any two quantities (currents or voltages) for any two branches of this circuit are related by the linear expression $y = a + bx$, where x is the current or voltage in one branch, and y, in the other branch. If, in the network, the EMF or the resistance vary simultaneously in two branches, then any three quantities in the network (currents or voltages) are related by the linear expression: $y = a + bx + cz$.

The coefficients a, b, and c can be determined experimentally or can be calculated.

So, for example, as the EMF E_m varies in any mth branch, the currents I_k and I_p in any kth and any pth branches are related by the linear expression $I_k = a_k + b_k I_p$. According to the compensation theorem, any branch can be replaced by an appropriately chosen EMF; hence the variation in the resistance in the mth branch also results in variations in the currents and voltages in the branches, and there exists linear relations between these currents. The coefficients a_k and b_k can be determined if the values of the voltages and currents are known in the branches under study for two operation modes in the circuit.

Reciprocity theorem: in any electric circuit, the current I_k in the kth branch caused by the EMF E_m located in the mth branch is equal to the current I_m in the mth branch caused by the EMF E_k located in the kth branch if $E_m = E_k$.

E6.2.6. Equivalent Transformations in Electrical Networks

► **Preliminary remarks.** The analysis of a complicated electric circuit can often be simplified by replacing a group of resistive elements in its equivalent circuit with an equivalent group where the resistive elements are connected in a different way. An expedient transformation of electrical networks leads to a decrease in the number of its branches or nodes and hence in the number of equations determining its electric state. In all cases of replacement of electrical networks by equivalent networks of different form, it is necessary to ensure that the currents and voltages remain unchanged in the network parts that remain untransformed.

► **Series and parallel connection of resistors.** The *series connection of resistors* is a connection such that the current flowing through the resistors is one and the same. For one and the same current to flow through the receivers connected in series, it is necessary that there be no electrical nodes at the points of their connection. Then the equivalent resistance R for N resistors connected in series is determined as

$$R = R_1 + R_2 + \cdots + R_N. \qquad \text{(E6.2.6.1)}$$

The *parallel connection of resistors* is a connection such that the voltage at the resistor terminals (points of connection) is one and the same. For one and the same voltage to be at the terminals of the resistors, it is necessary that they have a single common input (node) and a single common output (node).

For N resistors connected in parallel, we have

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \cdots + \frac{1}{R_N}. \qquad \text{(E6.2.6.2)}$$

► **Useful formulas for several resistors.** From formula (E6.2.6.2), one can obtain the equivalent resistances of two or three resistors connected in parallel:

$$R = \frac{R_1 R_2}{R_1 + R_2} \qquad \text{(two parallel resistors)},$$

$$R = \frac{R_1 R_2 R_3}{R_1 R_2 + R_1 R_3 + R_2 R_3} \qquad \text{(three parallel resistors)}.$$

A Y-network (also know as a star) of resistive elements can be transformed to an equivalent Δ-network; see Fig. E6.10. The inverse transformation is also possible.

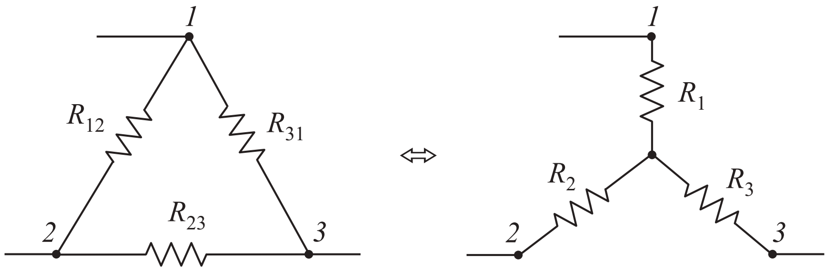

Figure E6.10. Invertible transformation between a Δ-network and a Y-network.

Δ-to-Y transformation formulas:

$$R_1 = \frac{R_{12} R_{31}}{R_{12} + R_{23} + R_{31}}, \quad R_2 = \frac{R_{23} R_{12}}{R_{12} + R_{23} + R_{31}}, \quad R_3 = \frac{R_{31} R_{23}}{R_{12} + R_{23} + R_{31}}.$$

The resistance of each ray in the equivalent Y-network is equal to the product of resistances of the two Δ-network branches adjacent to the corresponding ray divided by the sum of resistances of all branches in the Δ-network.

Y-to-Δ transformation formulas:

$$R_{12} = R_1 + R_2 + \frac{R_1 R_2}{R_3}, \quad R_{23} = R_2 + R_3 + \frac{R_2 R_3}{R_1}, \quad R_{31} = R_3 + R_1 + \frac{R_3 R_1}{R_2}.$$

We also present similar formulas for conductances:

$$g_{12} = \frac{g_1 g_2}{g_1 + g_2 + g_3}, \quad g_{23} = \frac{g_2 g_3}{g_1 + g_2 + g_3}, \quad g_{23} = \frac{g_3 g_1}{g_1 + g_2 + g_3},$$

Thus, the conductance of each branch in the equivalent Δ-network is equal to the product of conductances of the Y-network rays adjacent to the same terminals divided by the sum of conductances of all rays.

E6.2.7. Application of Equivalent Transformations in Network Analysis. Convolution Method

It is convenient to use the method of equivalent transformations mainly for networks with a single energy source, provided that the energy source voltage does not change after the transformation. The convolution method is reduced to the transformation of the electric circuit into the simplest network with a single source and a single equivalent receiver; then the backward calculations are performed.

We illustrate the convolution method by an example of the network given in Fig. E6.11. First, we arbitrarily choose the direction of unknown currents I_1, I_2, I_{34}. The resistors with resistances R_3 and R_4 connected in series are replaced by a resistor with equivalent resistance R_{34} by the formula $R_{34} = R_3 + R_4$. We transform the original network (Fig. E6.11a) into an equivalent network (Fig. E6.11b). The resistors with resistances R_2 and R_{34} connected in parallel are replaced by a resistor with equivalent resistance R_{234} by the formula

$$R_{234} = \frac{R_2 R_{34}}{R_2 + R_{34}}.$$

We replace the transformed network (Fig. E6.11b) by an equivalent network (Fig. E6.11c). It follows from Fig. E6.11c that the resistors with resistances R_1 and R_{234} are connected in series. Therefore, the resistance equivalent to them is calculated by the formula

$$R' = R_1 + R_{234}.$$

Thus, after the above transformations, the original network (Fig. E6.11a) is transformed into an equivalent network (Fig. E6.11d).

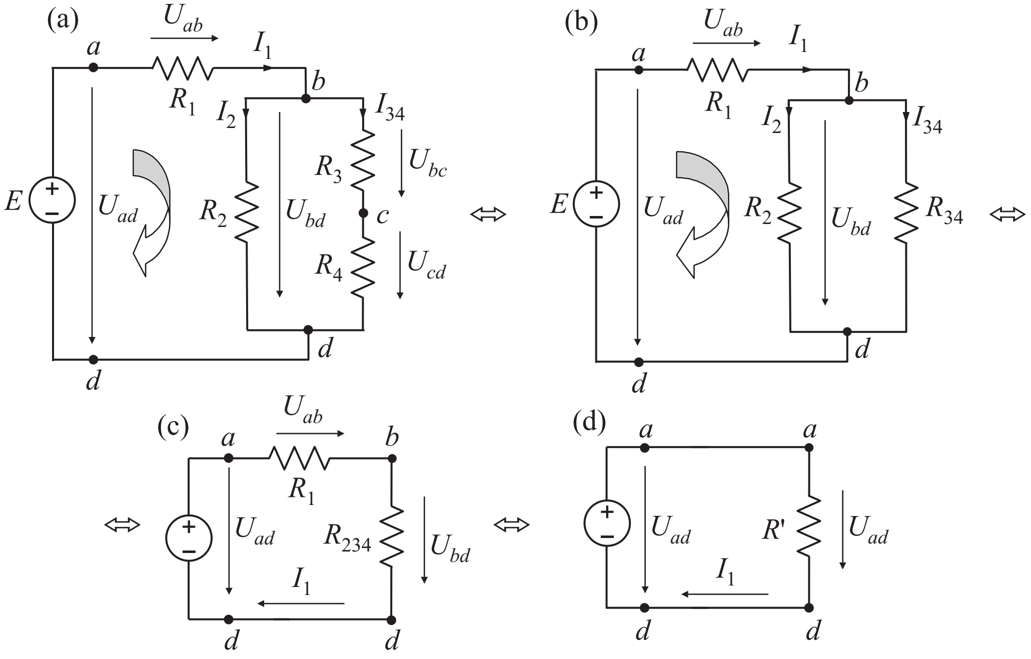

Figure E6.11. (a) Original electrical network and (b, c, d) the transformed networks connected in series.

The current I_1 at the circuit input (Fig. E6.11d) is calculated by Ohm's law:

$$I_1 = \frac{E}{R'} = \frac{U_{ab}}{R'}.$$

To determine the currents in parallel branches, it is necessary to calculate the voltage U_{bd} at them (Fig. E6.11c): $U_{bd} = R_{234} I_1$.

We determine the currents in parallel branches (Fig. E6.11b) by the formula

$$I_2 = \frac{U_{bd}}{R_2}, \qquad I_{34} = \frac{U_{bd}}{R_{34}}.$$

E6.2.8. Superposition Principle. The Notion of Input and Transfer Conductances (Admittances)

► **Superposition principle. Input and transfer conductances.** The following superposition principle holds for linear electric circuits: the current in any branch is equal to the algebraic sum of currents in this branch (partial currents) under the separate action of each source if the other sources are replaced by resistors with resistances equal to the internal resistances of the corresponding sources. An analog of such a principle can be found in mechanics, where the motion of a body under the action of several forces can be considered as a result of addition of actions caused by each separate force.

In the equivalent circuit with m branches, the current in each nth branch is equal to the algebraic sum of partial currents caused by each of the EMF and each of the current sources.

For a network without any current source, one can write an equation for the current in any nth branch of the linear electric circuit in the form

$$I_n = E_1 g_{1n} + E_2 G_{2n} + \cdots + E_m g_{mn} \qquad (n \le m).$$

In this equation, the factors multipying the EMF have the dimension of conductances. Therefore, each of the factors with two equal indices g_{nn} can be called the input conductance of branch n, and any of the factors with two distinct indices g_{mn} can be called the transfer conductance of the branches n and m.

The numerical values of the input and transfer conductances of branches can be calculated as follows. We equate all the EMF E_m with zero except for E_n.

Then $I_n = E_n g_{nn}$, whence $g_{nn} = I_n / E_n$. Therefore, the input conductance of any branch is determined by the current to EMF ratio in this branch provided that the EMF are zero in the other branches. The value of the current in branch m can be determined if we equate all the EMF with zero except for E_n. Then $I_m = E_m n g_{mn}$, whence $g_{mn} = I_m / E_n$. Therefore, the transfer conductance of any two branches is determined by the ratio of the current in one branch to the EMF in the other branch provided that the EMF in the other branches are zero.

The input conductance of a branch has a positive value, because it was assumed that we choose the same direction for the current and for the source EMF. The transfer conductance of two branches can take positive and negative values, and moreover, $g_{mn} = g_{nm}$, which means that the reciprocity principle is satisfied.

► **Electrical networks illustrating the superposition principle.** As an example, consider the network in Fig. E6.12a.

The currents in branches are equal to the sum of partial currents in the networks; see Fig. E6.12b and c.

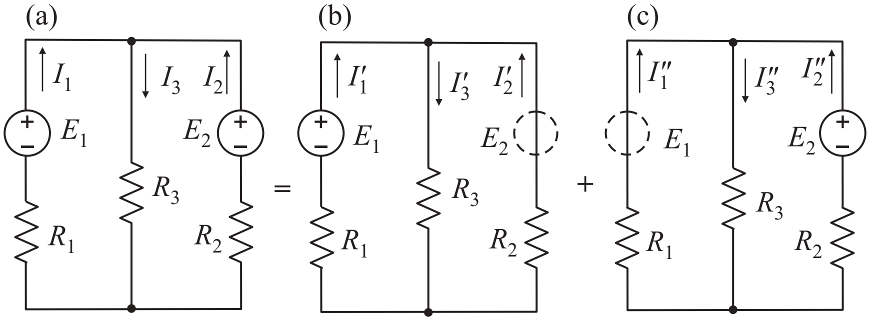

Figure E6.12. Electrical networks illustrating the superposition principle.

We determine the currents in the original network as the sum of partial currents:

$$\begin{aligned}
I_1 &= I_1' + I_1'' = g_{11}E_1 + g_{12}E_2 = \frac{R_1 + R_3}{R^2}E_1 - \frac{R_3}{R^2}E_2,\\
I_2 &= I_2' + I_2'' = g_{21}E_1 + g_{22}E_2 = -\frac{R_3}{R^2}E_1 + \frac{R_1 + R_3}{R^2}E_2,\\
I_3 &= I_3' + I_3'' = g_{31}E_1 + g_{32}E_2 = \frac{R_2}{R^2}E_1 + \frac{R_1}{R^2}E_2,
\end{aligned}$$

where $R = \sqrt{R_1R_2 + R_1R_3 + R_2R_3}$.

In the equivalent circuits with current sources, the partial currents of branches are determined from each of them by eliminating the other current sources by disconnecting the branches containing them.

E6.2.9. Concise Methods for Analyzing Electric Circuits

► **Mesh-current method.** The mesh-current method is based on Kirchhoff's laws and is most convenient for analyzing electric circuits where the number of independent loops is less than the number of electrical nodes.

When solving the problems, it is assumed that, in the independent loops of the electric circuit, there are their own currents, which are denoted for the first, second, etc. loops, respectively.

Then the linear equations for the chosen unknown loop (fictitious) currents can be composed according to Kirchhoff's second law. The number of the composed independent linear equations must correspond to the number of unknown loop currents. When constructing the equations, first, one must compose the simplest independent equations, and second, each subsequent equation must contain at least one branch that does not enter the loops already used. Solving the obtained system of equations for the unknown loop currents, one can determine the unknown (actual) currents.

In the general case, for a network with n independent loops, the system of equations of the mesh-current method can be written as

$$\begin{aligned}
R_{11}I_{11} + R_{12}I_{12} + \cdots + R_{1n}I_{nn} &= E_{11},\\
R_{21}I_{11} + R_{22}I_{12} + \cdots + R_{2n}I_{nn} &= E_{22},\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots&\cdots,\\
R_{n1}I_{11} + R_{n2}I_{12} + \cdots + R_{nn}I_{nn} &= E_{nn}.
\end{aligned} \qquad \text{(E6.2.9.1)}$$

The solution of this system for loop currents can, for example, be found by the Cramer method (see Section M5.3).

Example. We illustrate the use of the mesh-current method by the example of the electrical network shown in Fig. E6.13.

We have the system of equations

$$\begin{aligned}
(R_1 + R_4)I_{11} - R_4I_{22} &= E_1,\\
-R_4I_{11} + (R_3 + R_4 + R_5)I_{22} - R_3I_{33} &= E_3 + E_5,\\
-R_3I_{22} + (R_2 + R_3)I_{33} &= -E_2 - E_3.
\end{aligned} \qquad \text{(E6.2.9.2)}$$

We denote

$$R_{11} = R_1 + R_4, \quad R_{22} = R_3 + R_4 + R_5, \quad R_{33} = R_2 + R_3, \quad R_{12} = R_{21} = -R_4, \quad R_{32} = R_{23} = -R_3,$$
$$E_{11} = E_1, \quad E_{22} = E_3 + E_5, \quad E_{33} = -E_2 - E_3.$$

As a result, system (E6.2.9.2) takes the standard form (E6.2.9.1):

$$\begin{aligned} R_{11}I_{11} + R_{12}I_{22} &= E_{11},\\ R_{21}I_{11} + R_{22}I_{22} + R_{23}I_{33} &= E_{22},\\ R_{32}I_{22} + R_{33}I_{33} &= E_{33}. \end{aligned}$$

Determining the loop currents from this system, we can readily find the actual currents in the circuit by the formulas

$$I_1 = I_{11},\quad I_2 = -I_{33},\quad I_3 = I_{22} - I_{33},\quad I_4 = I_{22} - I_{11},\quad I_5 = I_{22}.$$

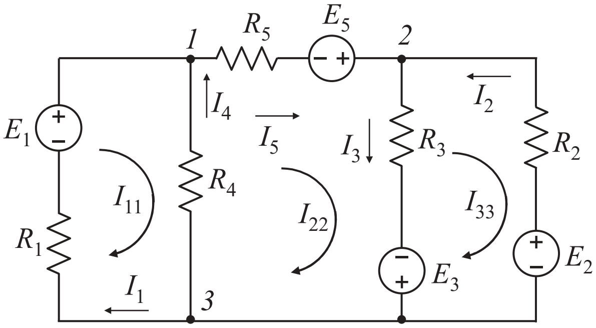

Figure E6.13. Electrical network illustrating the mesh-current method.

► **Node-potential method.** The operation mode of any circuit is completely characterized by the equations constructed on the basis of Kirchhoff's first and second laws, and to determine the currents in all branches, it is necessary to compose and solve a system with the same number of unknowns (unknown node potentials). But the number of equations to be solved can be decreased by using the node-potential method based on the application of Kirchhoff's first law and the generalized Ohm law.

To explain the essence of the method, consider the electrical network shown in Fig. E6.14. Assume that the potential of one of the nodes, for example, of the third, is zero, $\varphi_3 = 0$. Then for nodes 1 and 2 we compose the following equations according to Kirchhoff's first law:

$$I_5 - I_4 - I_1 + I_6 = 0,\quad -I_5 - I_6 - I_2 + I_3 = 0. \tag{E6.2.9.3}$$

The currents in the branches can be determined by using Ohm's law:

$$\begin{aligned} I_6 &= (\varphi_1 - \varphi_2)g_6, & I_1 &= (-\varphi_1 + E_1)g_1,\\ I_4 &= (-\varphi_1)g_4, & I_5 &= (\varphi_1 - \varphi_2 + E_5)g_5,\\ I_2 &= (-\varphi_2 + E_2)g_2, & I_3 &= (\varphi_2 + E_3)g_3, \end{aligned} \tag{E6.2.9.4}$$

where φ_1 and φ_2 are the potentials of the first and second nodes. Substituting the expressions (E6.2.9.4) for the currents into (E6.2.9.3), we obtain

$$\begin{aligned} \varphi_1(g_1 + g_4 + g_5 + g_6) - \varphi_2(g_5 + g_6) &= E_1g_1 - E_5g_5,\\ -\varphi_1(g_5 + g_6) + \varphi_2(g_2 + g_3 + g_5 + g_6) &= E_1g_1 - E_3g_3 + E_5g_5, \end{aligned}$$

or

$$\begin{aligned} \varphi_1 g_{11} - \varphi_2 g_{12} &= E_1g_1 - E_5g_5,\\ -\varphi_1 g_{21} + \varphi_2 g_{22} &= E_1g_1 - E_3g_3 + E_5g_5. \end{aligned}$$

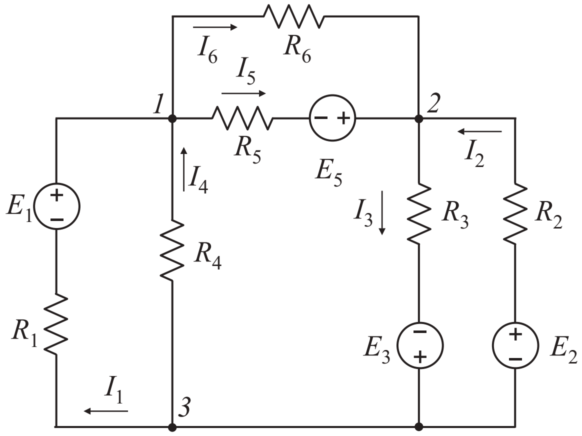

Figure E6.14. Electrical network illustrating the node-potential method.

Here $g_{11} = g_1 + g_4 + g_5 + g_6$ and $g_{22} = g_2 + g_3 + g_5 + g_6$ are the sums of conductances of the branches adjacent to the first and second nodes, respectively, and $g_{12} = g_{21} = g_5 + g_6$ is the sum of conductances of the branches connecting these nodes.

The right-hand side of the last equation is an algebraic sum of products of the EMF E of sources by the conductances of the branches connected to the node of interest. This product is taken with the plus sign if the EMF E is directed towards the node and with the minus sign if the EMF E is directed from the node. (The positive direction of an EMF source is from minus to plus.)

It is remarkable that the current directions in the branches need not be taken into account when composing the equations.

► **Nodal-pair method.** The nodal-pair method is based on the generalized Ohm law and Kirchhoff's first law and is used to calculate the electric circuits with two nodes (i.e., electric circuits with branches connected in parallel). When solving the problems by this method, one first determines the voltage between the two nodes by the formula:

$$U_{12} = \frac{\sum g_i E_i + \sum J_k}{\sum g_j}$$

where g_i are the electric conductances of the corresponding branches, J_k are the values of currents from the current sources, and $g_i E_i$ is the product of electric conductance of a branch by the EMF E of the same branch, which is taken with sign "+" if the EMF E is directed towards node 1; otherwise, the sign "–" is taken.

After the voltage drop between two nodes is determined, one calculates the unknown currents by the generalized Ohm law.

Example. We illustrate the use of the nodal-pair method by anthe example of the electrical network shown in Fig. E6.15.

The voltage between two nodes is equal to

$$U_{12} = \frac{E_1 g_1 - E_2 g_4 + J}{g_1 + g_2 + g_4},$$

where $g_1 = 1/R_1$, $g_2 = 1/(R_2 + R_3)$, and $g_3 = 1/R_4$.

► **Proportional parts method.** Sometimes, the currents in a circuit containing only one source can be found in a relatively simple way by using the proportional parts method. The essence of this method is that in one of the branches, usually in the most remote

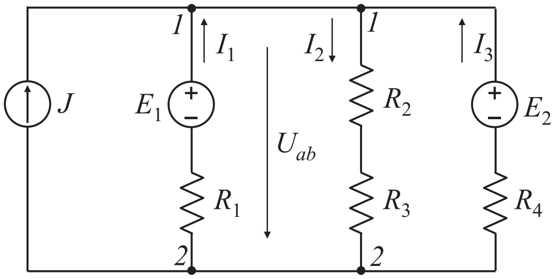

Figure E6.15. Electrical network illustrating the nodal-pair method.

from the energy source, an arbitrary value of the current strength is assumed, usually equal to 1 A. Then, starting from this current, one calculates the currents and voltages in the other branches, as a rule, by using Ohm's law; these calculations are performed until the energy source voltage corresponding to the accepted values of currents in branches is found. After this, the coefficient of proportionality between the actual voltage U_F and the calculated source voltage U' is derived: $k = U_F/U'$. To determine the true values of currents in all the branches, one must multiply the current values obtained earlier by the proportionality coefficient k: $I_m = kI'_m$.

E6.3. Nonlinear Electric Circuits of Direct Current

E6.3.1. Notion of a Nonlinear Element. Several Typical Characteristics of Nonlinear Elements

► **Nonlinear elements and their characteristics.** The electric circuits used in automatics, telemechanics, measurement and computation technology often contain separate passive elements whose electric resistance depends on the value and direction of the flowing current or the applied voltage. Such elements and the electric circuits containing them are said to be *nonlinear* and are used to determine certain functional dependencies between different electric quantities.

The nonlinear elements can be *uncontrolled* or *controlled*. The first ones, impedors, are intended to operate without any actions on them from a controlled factor; the second ones, multiterminal devices, are used under the action of controlled factor on them (transistors, thyristors, etc.).

The electric properties of nonlinear elements are represented by their volt-ampere characteristics (VACh) $I = f(U)$. In the uncontrolled nonlinear elements, the VACh is shown by a single curve, and in the controlled, by a family of curves whose parameter is the control factor. The VACh of nonlinear elements have different shapes and, according to this, can be divided into *symmetric* and *nonsymmetric*. The electric resistance of linear elements is constant and the VACh is a straight line passing through the origin (Fig. E6.16.).

In symmetric nonlinear elements, a change in the direction of the applied voltage U does not change the value of the current I, but changes its direction, and hence the relation $f(U) = -f(-U)$ holds for their VACh (Fig. E6.17a).

In nonsymmetric nonlinear elements, for the same magnitude of the voltage U applied in the opposite directions, the currents I_1 and I_2 are different in values and the inequality $f(U) \neq -f(-U)$ holds for VACh (Fig. E6.17b).

► **Static and differential resistances of a nonlinear element.** For each nonlinear element, the static resistance R_C corresponding to a given VACh point can be distinguished.

It follows from Fig. E6.18 that the value of this resistance for point A is proportional to the tangent of the angle β between the straight line connecting point A with the origin and

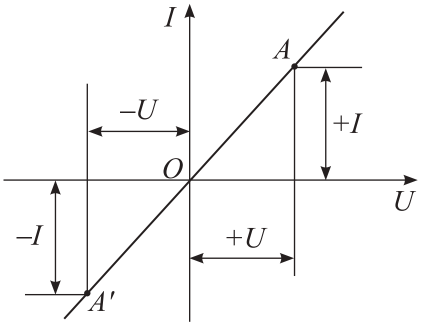

Figure E6.16. The volt-ampere characteristic of a linear element.

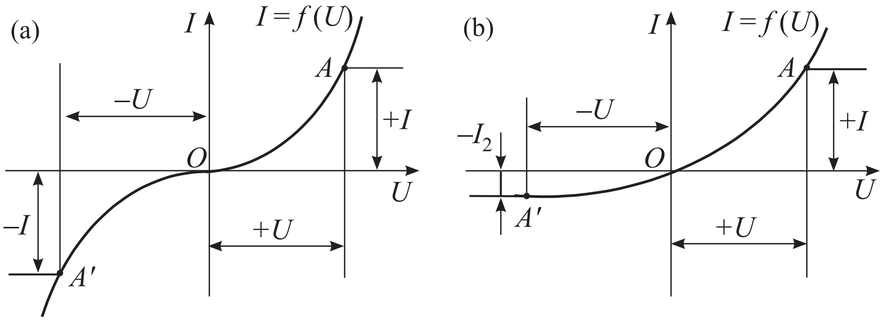

Figure E6.17. The volt-ampere characteristic of a nonlinear symmetric (a) and a nonsymmetric (b) element.

the axis of currents, i.e.,

$$R_S = \frac{U}{I} = \frac{m_U\,OB}{m_I\,BA} = m_R \tan\beta,$$

where m_U, m_I, and m_R are the voltage, current, and resistance scales, respectively.

The limit of the ratio of the voltage increment on the circuit part to the current increment on it, or the derivative of the voltage with respect to the current, determines the differential resistance. The value of this resistance is proportional to the tangent of the angle α between the tangent to VACh at the same point A and the axis of currents, i.e.,

$$R_D = \frac{dU}{dI} = \frac{m_U\,OB}{m_I\,CA} = m_R \tan\alpha.$$

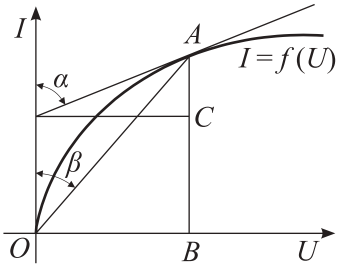

Figure E6.18. Determining the static and differential resistances of a nonlinear element.

The static resistance characterizes the behavior of a nonlinear element in the constant current operation mode, and the differential resistance, in the case of the current small deviations from the steady state.

Consider a closed circuit consisting of a nonlinear element and an active impedor with the external characteristic $U = E - R_0 I$. Point A of intersection of the external characteristic of an active impedor (line 1) with the VACh nonlinear resistive element (curve 2) determined the operation mode circuit (Fig. E6.19).

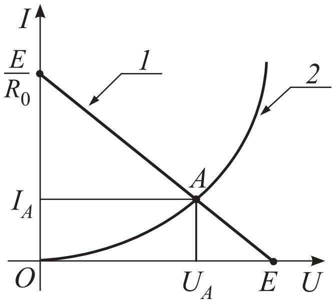

Figure E6.19. The volt-ampere characteristics of an active impedor and a nonlinear resistive element.

E6.3.2. Equivalent Transformations of Characteristics of Nonlinear Elements

If the VACh of a nonlinear element on the operation segment is practically linear, then, in calculation, the nonlinear element can replaced by an equivalent network consisting of a voltage source and a linear resistance R_D. So, for example, the VACh of a nonlinear element shown in Fig. E6.20 on small parts near the operation point A can be replaced by the straight line whose equation has the form

$$U_1 = E_1 + m_R \tan\alpha_1 I \qquad \text{or} \qquad U_1 = E_1 + R_{D1} I. \tag{E6.3.2.1}$$

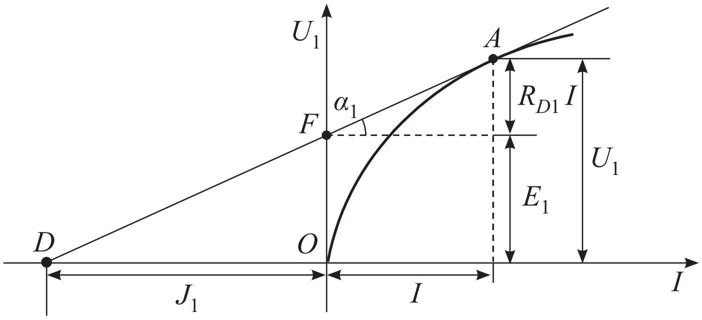

Figure E6.20. The volt-ampere characteristic of a nonlinear element.

At the operating point A, the voltage U_1 and the current I on the nonlinear element (Fig. E6.21a) are related by the expression (E6.3.2.1). The equivalent networks of such an element are shown in Figs. E6.21b and c. In this case, the EMF E_1 is opposite to the current I, because only in the case of this EMF direction the potential of point 1 exceeds

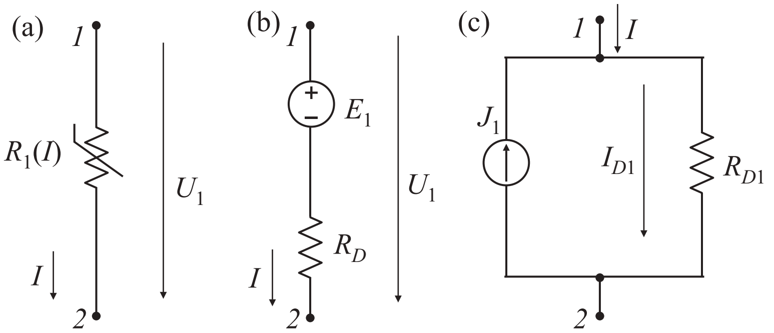

Figure E6.21. Equivalent circuits for replacing a nonlinear element.

the potential of point 2 by U_1. The value of J_1 is equal to the segment OD cut off on the axis of currents by the continuation of the tangent.

If a nonlinear element has the VACh shown in Fig. E6.22, then Eq. (E6.3.2.1) has the form

$$U_2 = -E_2 + m_R \tan\alpha_2\, I \qquad \text{or} \qquad U_2 = -E_2 + R_{D2} I.$$

If the current and voltage of a nonlinear element have the same positive directions (Fig. E6.21a), then the directions of the EMF and of the current in the current source are changed to the opposite in the equivalent networks.

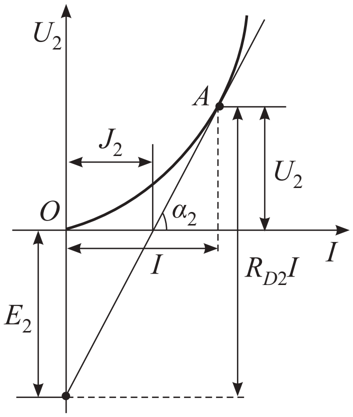

Figure E6.22. The volt-ampere characteristic of a nonlinear element.

E6.3.3. Graphical Method for Analyzing Circuits with Nonlinear Elements

► **Preliminary remarks.** Nonlinear electric circuits are often analyzed by the graphical method based on Kirchhoff's laws and the VACh of some elements contained in the electric circuit. The initial data for calculations (VACh) are given in the form of graphs and tables.

Remark. In the case of nonlinear circuits, one cannot use the theorems and methods based on the assumption that all the elements contained in these circuits are linear. So the superposition principle cannot be used in nonlinear circuits. The linearity theorem, the reciprocity theorem, and the proportional parts method also cannot be used.

We consider several problems with different versions of connection of nonlinear elements in nonlinear circuits whose networks are shown in Fig. E6.23.

► **Series connection of nonlinear elements.** Figure E6.23a shows the series connection of two nonlinear elements $R_1(I_1)$ and $R_2(I_2)$ whose VACh are given in Fig. E6.24. It is required to determine the current I and the voltages U_1 and U_2 of elements for a given voltage U at the circuit terminals.

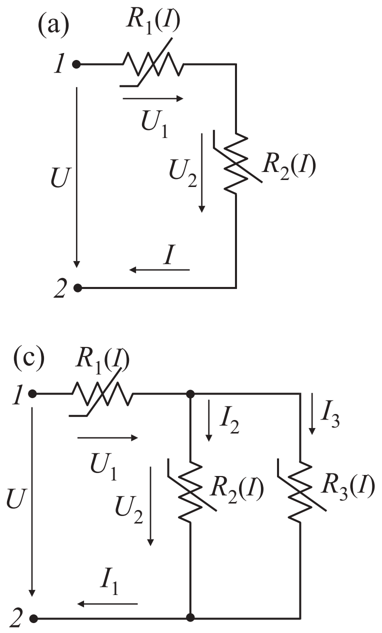

Figure E6.23. Examples of nonlinear circuits.

To calculate the current I and the voltages U_1 and U_2, one must construct an auxiliary characteristic, namely, the dependence of the current on the total voltage $U_1 + U_2$. Since, just as in an unbranched circuit, the current is the same in both nonlinear elements, it follows that, to construct the characteristic $I(U_1 + U_2)$, it is necessary to sum the voltages U_1 and U_2 for equal values of the current I.

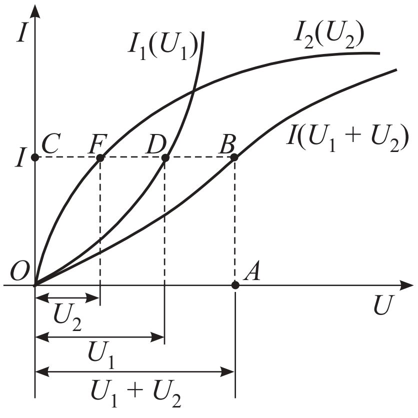

Figure E6.24. The volt-ampere characteristics of nonlinear resistive elements connected in series.

We put the voltage U on the abscissa axis and, from point A, draw the straight line AB parallel to the ordinate axis till the intersection with the curve $I(U_1 + U_2)$; the segment AB thus obtained is equal in scale to the current I. Then, from point B, we draw the straight line BC parallel to the ordinate axis. As a result, we obtain segments CD and CF equal in scale to U_1 and U_2, respectively. Similarly, we can solve the problem of analyzing a circuit consisting of three or more elements with nonlinear characteristics connected in series.

The graphical constructions for analyzing a series circuit can be performed differently. We assume that, in Fig. E6.23a, the ends of a circuit part are connected by the EMF source $E = U$. The voltage U_2 at the terminals of a nonlinear element $R_2(I_2)$ is determined, on one hand, by the VACh of this element, and, on the other hand, by the difference between the EMF $E = U$ and the voltage U_1 at the terminals of the nonlinear element $R_1(I_1)$.

Figure E6.25 presents the characteristic and the curves $I_2(U_2)$ and $I_1(U - U_2)$ at whose point of intersection the currents in the elements coincide.

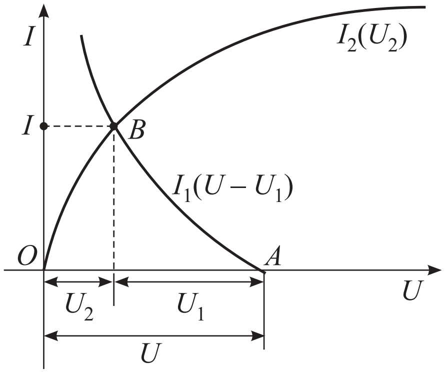

Figure E6.25. Curve $I_2(U_2)$ and curve $I_1(U - U_2)$.

► **Parallel connection of nonlinear elements.** In Fig. E6.23b, we show the parallel connection of two nonlinear elements $R_1(I_1)$ and $R_2(I_2)$ whose VACh are shown in Fig. E6.26. Assume that the voltage at the terminals of the circuit is equal to U. Then, by the VACh $I_1(U_1)$ and $I_2(U_2)$, it is easy to determine the currents I_1 and I_2 in the nonlinear elements and, by the formula $I_1 + I_2 = I$, the current in the unbranched part of the circuit.

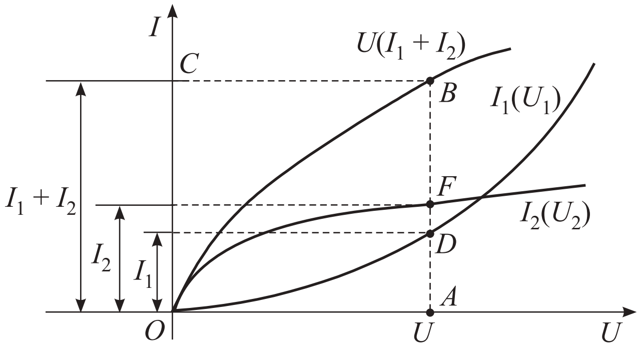

Figure E6.26. The volt-ampere characteristics of nonlinear resistive elements connected in parallel.

If a current I is given, then, to determine the voltage U and the currents I_1 and I_2 in nonlinear elements, one must construct the auxiliary quantity $(I_1+I_2)U$, i.e., the dependence of the total current on the voltage U. Since, in parallel connection, $U_1 = U_2 = U$, it follows that, to construct this quantity, it is necessary to sum the ordinates of the curves $I_1(U_1)$ and $I_2(U_2)$ for the same values of the voltage. Then, on the ordinate axis, we put the interval OC equal in scale to the current I and, from point C, draw the straight line CB parallel to the abscissa axis till the intersection with the characteristic $(I_1 + I_2)U$. The segment $CB = OA$ thus obtained is equal in scale to the voltage U. Finally, the segments AD and AF are equal in scale to the currents I_1 and I_2.

► **Mixed connection of nonlinear elements.** In Figs. E6.23c and d, we show the mixed connection of nonlinear elements (c) and of nonlinear and linear elements (d). To determine the desired currents and voltages, one can also use the graphical method by analogy with calculations of series and parallel circuits.

E6.4. Magnetic Circuits

A set of devices containing ferromagnetic bodies and forming a closed circuit in which a magnetic flux is formed in the presence of a magnetizing force and along which the magnetic flux lines close up, is called a magnetic circuit. The magnetic circuits, just as the electric circuits, can be branched and unbranched.

E6.4.1. Magnetic Field and Its Parameters

► **Magnetic induction vector.** The magnetic field arising when a direct current flows through a conducting medium is graphically shown by magnetic lines of force and is characterized at each point by the magnetic induction vector **B** directed along the tangent to the field lines (Fig. E6.27).

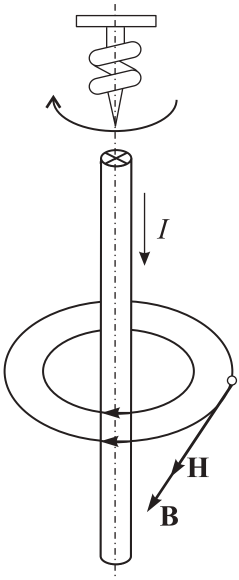

Figure E6.27. Magnetic field of a rectilinear conductor and an induction coil.

The direction of this vector is determined by the right-hand screw rule (corkscrew rule), which means that if the translational motion of such a screw is aligned in direction with the electric current I, then the direction of its rotational motion coincides with that of the magnetic induction vector **B**.

If the magnetic induction **B** has the same value and direction at all points of the field, then such a field is said to be *uniform* or *homogeneous*.

In the international system of units, or the SI system, the magnetic induction (flux density) B is measured in teslas (T) or webers per square meter ($\mathrm{Wb/m^2}$).

► **Magnetic flux.** The second important quantity characterizing the magnetic field is the magnetic flux Φ, i.e., the flux of the magnetic induction vector $\mathbf{B}$ through a surface S (Fig. E6.28a):

$$\Phi = \int \mathbf{B}\cdot d\mathbf{S} = \int B\cos(\mathbf{B}, d\mathbf{S})\,dS.$$

The quantity Φ can take either positive or negative value depending on the value of the angle between the directions of the vectors $\mathbf{B}$ and $d\mathbf{S}$.

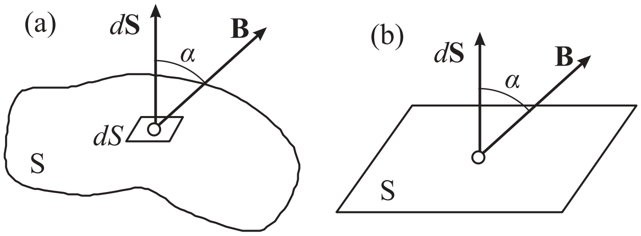

Figure E6.28. Determining the magnetic flux through (a) an arbitrary surface or (b) a plane surface in a uniform magnetic field.

If a magnetic field is uniform (homogeneous), then, for all points of space, the magnetic induction vector is constant, $\mathbf{B} = \mathbf{const}$, and the magnetic flux is determined by the formula (Fig. E6.28b):

$$\Phi = BS\cos\alpha. \qquad \text{(E6.4.1.1)}$$

In SI, the magnetic flux is measured in webers (Wb).

► **Magnetic field strength vector. Absolute and relative magnetic permeability.** When studying magnetic fields and analyzing magnetic devices, one uses the magnetic field strength (intensity) vector $\mathbf{H}$, which coincides in direction with the magnetic induction vector $\mathbf{B}$:

$$\mathbf{B} = \mu_a\mathbf{H} = \mu\mu_0\mathbf{H},$$

where μ_a is the absolute magnetic permeability of the medium, $\mu_0 = 4\pi\cdot 10^{-7}$ H/m (henry/meter) is the magnetic constant (permeability of vacuum), and μ is the (relative) magnetic permeability.

For nonferromagnetic materials and media (wood, paper, copper, aluminum, air), μ_a does not practically differ from μ_0 ($\mu = 1$).

For ferromagnetic materials (steel, cast iron, etc.), μ_a is many times larger than μ_0 and depends on the value of the magnetic field strength H.

In SI system, the magnetic field strength H is measured in amperes per meter (A/m).

E6.4.2. Ampère’s Circuital Law

The calculations of magnetic circuits are based on Ampère’s circuital law, which was obtained on the basis of numerous experiments. This law states that the linear integral of the magnetic field strength vector along any loop is equal to the total current surrounded by this loop (algebraic sum of currents related to the loop):

$$\oint \mathbf{H}\cdot d\mathbf{l} = \oint H\cos\alpha\,dl = \sum I_k,$$

where $\mathbf{H}$ is the magnetic field strength at a given point in space, $H = |\mathbf{H}|$, dl is a length element of the closed loop, and $\cos\alpha$ is the angle between the directions of the vectors $\mathbf{H}$ and $d\mathbf{l}$.

The currents are assumed to be positive if their direction corresponds to the corkscrew rule (is clockwise).

In particular, according to Ampère's circuital law, for the loop in Fig. E6.29, we have

$$\oint \mathbf{H}\cdot d\mathbf{l} = \oint H\cos\alpha\, dl = -I_1 + I_2 + I_3.$$

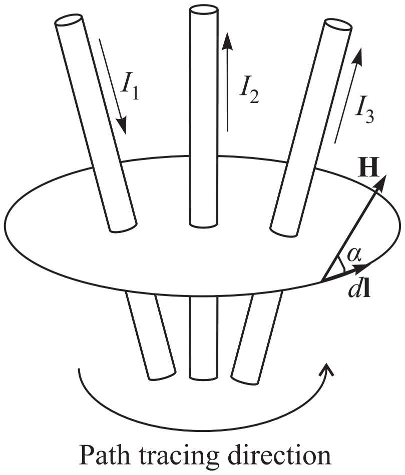

Figure E6.29. Ampère's circuital law.

The quantity $\sum I_k$ is called the magnetomotive force (MMF); in the SI system, the main unit of MMF is ampere (A). The main unit of the magnetic field strength in SI system is ampere per meter (A/m) and does not have any special name. One often uses the "ampere per centimeter" unit, 1 A/cm = 100 A/m.

E6.4.3. Ohm's Law for Magnetic Circuit

To obtain the dependence between the magnetic flux and the MMF, we consider an example of unbranched circuit shown in Fig. E6.30.

We assume that the parts l_1 and l_2 of the magnetic circuit are made of the same ferromagnetic material but have different transverse cross-sections S_1 and S_2.

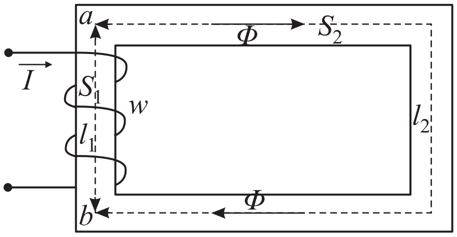

Figure E6.30. Unbranched circuit.

If the lengths of the magnetic circuit parts (l_1 and l_2) significantly exceed the transverse dimensions, then we can assume that the magnetic induction at all points of each part is the same.

The magnetic induction for two parts of the circuit is determined by the formulas

$$B_1 = \frac{\Phi}{S_1}, \qquad B_2 = \frac{\Phi}{S_2}.$$

The magnetic field strength on these parts is calculated as

$$H_1 = \frac{B_1}{\mu_0\mu_1} = \frac{\Phi}{\mu_0\mu_1 S_1}, \qquad H_2 = \frac{B_2}{\mu_0\mu_2} = \frac{\Phi}{\mu_0\mu_2 S_2}. \tag{E6.4.3.1}$$

Let us apply Ampère's circuital law to the loop coinciding with the center magnetic line of length $l = l_1 + l_2$. We obtain

$$F = \oint H\,dl = H_1 l_1 + H_2 l_2 = Iw, \tag{E6.4.3.2}$$

where w is the number of turns of the current loop. In this equation, by analogy with the EMF of an electric circuit, $F = Iw$ is called the magnetomotive force or MMF.

Substituting H_1 and H_2 from (E6.4.3.1) into (E6.4.3.2), we find

$$F = Iw = \frac{\Phi l_1}{\mu_0\mu_1 S_1} + \frac{\Phi l_2}{\mu_0\mu_2 S_2} = \Phi R_{M1} + \Phi R_{M2}.$$

The quantity $l/(\mu_a S) = R_m$ is usually called the *magnetic resistance* of a magnetic circuit (by analogy with the electric resistance $R = l/\gamma S$).

The magnetic resistance of air (gaps) is linear, because $\mu_a = \mu_0 = \text{const}$. The magnetic resistance of the core is nonlinear, because μ_a depends on B.

The product of the magnetic flux Φ by the magnetic resistance R_M is called, by analogy with electric circuit, the *magnetic voltage*:

$$U_{Mab} = \Phi R_{M1},$$

whence we obtain the magnetic flux

$$\Phi = \frac{Iw}{R_{M1} + R_{M2}} = \frac{Iw - U_{Mab}}{R_{M1}} = \frac{U_{Mab}}{R_{M2}}.$$

This equation is usually called *Ohm's law for a magnetic circuit*. It should be noted that this analogy is purely formal, and the physical essence of processes in electric and magnetic circuits is different.

E6.4.4. Ferromagnetic Materials and Their Properties

► **Ferromagnetic materials. Hysteresis loop.** It is known that the magnetic permeability of ferromagnetic materials is a variable quantity and depends on B, and hence the magnetic resistance R_M is also nonconstant, which significantly complicates the calculations of magnetic circuits. Therefore, to calculate the magnetic circuits with ferromagnetic parts, it is necessary to know the magnetization curves, which present the dependence $B = f(H)$. These dependencies are obtained experimentally by testing closed magnetic circuits with distributed winding.

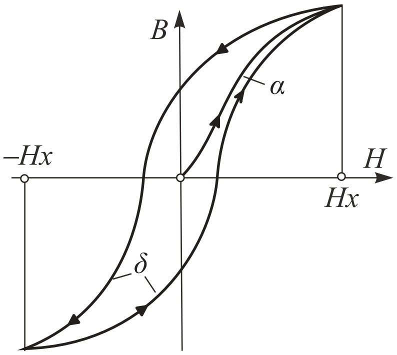

Figure E6.31. Dependence $B = B(H)$ and the hysteresis loop.

The primary magnetization of a specimen is described by the curve μ, which is called the curve of primary magnetization (Fig. E6.31).

If a specimen is cyclically magnetized with the magnetic field strength varying in the limits from Hx to $-Hx$, then the graph is a closed curve, which is known as the hysteresis loop.

If the process of cyclic magnetization is repeated for gradually increasing values of the magnetic field strength, then one can obtain a family of hysteresis loops and the so-called limit hysteresis loop.

To the limit loop there corresponds a variation in the magnetic field strength in the limits from $H_{\max}$ to $-H_{\max}$. Any increase in H beyond $H_{\max}$ does not increase the hysteresis loop area. Such a hysteresis loop determines the value of the residual magnetic induction. The curve connecting the vertices of a hysteresis loops is called the normal magnetization curve. These curves are given in reference books and are used to analyze magnetic circuits.

► **Soft magnetic and hard magnetic materials.** The process of cyclic magnetization reversal requires some energy, which is proportional to the hysteresis loop area.

In this connection, for the magnetic circuits in electrical engineering devices operating under conditions of uninterrupted magnetization reversal (for example, transformers) it is expedient to use ferromagnetic materials with a narrow hysteresis loop (see curves α in Fig. E6.32). Such ferromagnetic materials are said to be *soft magnetic* (electrical grade sheet and several special alloys, for example, permalloy consisting of nickel, iron, and other components).

In production of permanent magnets, it is recommended to use ferromagnetic materials with a wide hysteresis loop (curves δ), which have a large residual induction. Such ferromagnetic materials are said to be *hard magnetic* (several alloys of iron with wolfram, chrome, and aluminum).

E6.4.5. Kirchhoff's Laws for Magnetic Circuits

► **Kirchhoff's first law for magnetic circuits.** The magnetic flux Φ completely closed by a magnetic circuit is called the *main flux*, and the magnetic flux closed partially by parts of a magnetic circuit and partially by the environment is called the *leakage flux*.

It is well known that the lines of the magnetic induction vector are closed. Therefore, the magnetic flux through any closed surface is always zero. Hence, the following equation holds for nodes of magnetic circuits:

$$\sum \Phi_k = 0. \tag{E6.4.5.1}$$

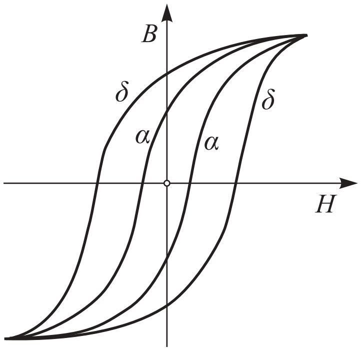

Figure E6.32. Hysteresis loops of soft magnetic (curves α) and hard magnetic (curves δ) materials.

In this equation, the positive signs are taken for the fluxes with the positive direction away from the node, and negative signs are taken for the fluxes with direction towards the node of the magnetic circuit (or vice versa), just by analogy with Kirchhoff's first law for direct current circuits.

Equation (E6.4.5.1) expresses Kirchhoff's first law for magnetic circuits: the algebraic sum of magnetic fluxes at a node of a magnetic circuit is zero.

► **Kirchhoff's second law for magnetic circuits.** Using Ampère's circuital law, one can obtain an equation by analogy with Kirchhoff's second law for electric circuits. In Fig. E6.33, we show a branched magnetic circuit. This circuit consists of six unbranched parts and has the same number of distinct magnetic fluxes. Each part of the magnetic circuit is made of the same material with constant transverse cross-section on its entire length. But different parts of this circuit can be produced of different ferromagnetic materials and can have different transverse cross-sections. Neglecting the leakage fluxes and assuming that the lengths of the parts significantly exceed their transverse dimensions, we can determine the magnetic inductions.

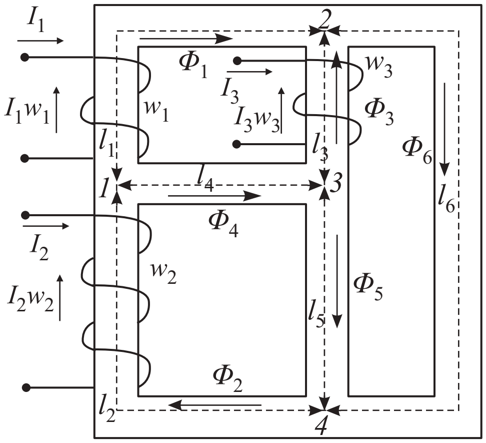

Figure E6.33. Branched magnetic circuit.

The magnetic field strengths on these parts of the magnetic circuit can be found from the magnetization curves under the obtained values of the magnetic induction. We apply Ampère's circuital law, for example, to close the loop 1-2-3-1 of the magnetic circuit (bypassed clockwise) and obtain

$$\sum F_k = \sum I_k w_k = I_1 w_1 - I_3 w_3 = H_1 l_1 - H_3 l_3 - H_4 l_4.$$

Taking into account the fact that

$$H_1 = \frac{\Phi_1}{\mu_0\mu_1 S_1}, \quad H_3 = \frac{\Phi_3}{\mu_0\mu_3 S_3}, \quad H_4 = \frac{\Phi_4}{\mu_0\mu_4 S_4},$$

we have

$$\sum F_k = \frac{\Phi_1 l_1}{\mu_0\mu_1 S_1} - \frac{\Phi_3 l_3}{\mu_0\mu_3 S_3} - \frac{\Phi_4 l_4}{\mu_0\mu_4 S_4} = \Phi_1 R_{M1} - \Phi_3 R_{M3} - \Phi_4 R_{M4}.$$

Generalizing the last formula to any closed loop, we obtain *Kirchhoff's second law for a magnetic circuit*:

$$\sum F_k = \sum I_n w_n = \sum \Phi_p R_{Mp} = \sum H_q l_q;$$

i.e., the algebraic sum of MMF in a loop of a magnetic circuit is equal to the algebraic sum of magnetic voltages in the same loop.

Here it should be specially noted that the resistance of any part of the magnetic circuit cannot usually be considered as a constant variable because of the dependence of the magnetic permeability of ferromagnetic materials on the induction. Therefore, the analysis of magnetic circuits is similar to the analysis of electric circuits with nonlinear VACh.

E6.4.6. Analysis of an Unbranched Magnetic Circuit

When analyzing an unbranched magnetic circuits, we encounter problems of two types.

In some problems, it is necessary to determine the magnetizing current from a given magnetic flux, in other problems, the flux from a given current or a given MMF. In both cases, as a rule, the geometric dimensions of all parts of magnetic circuits, the material of which they are produced, the normal magnetization curves or hysteresis loops, and the number of turns of the induction coil are given. In magnetic circuits composed of soft magnetic materials, hysteresis can usually be neglected; i.e., one can assume that the dependence of the induction on the magnetic field strength is unique and determined by the normal magnetization curve.

► **First problem for unbranched magnetic circuits.** Consider the unbranched magnetic circuit shown in Fig. E6.34, which consists of a π-shaped electromagnetic core and a steel plate closing its terminals. There is an air gap between the electromagnetic core terminals and the plate. It is required to determine the magnetizing current at which the magnetic flux in the air gap has a prescribed value. The electromagnet cross-section is S_1, and the plate cross-section is S_2.

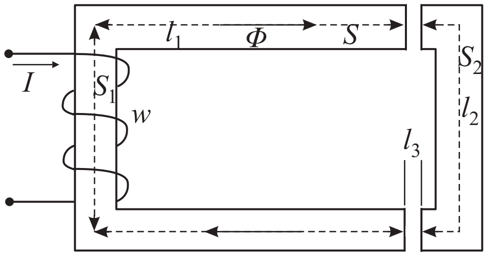

Figure E6.34. Unbranched magnetic circuit.

Under these conditions, we represent an entire magnetic circuit as three parts connected in series: the electromagnetic core, two air gaps, and the steel plate with the same flux Φ. To calculate the magnetic circuit, we determine the lengths of separate parts of the magnetic circuit: l_1, l_2, $2l_3$. For small dimensions of the core transverse cross-sections compared with their length, small inaccuracies in determining the center lengths of the circuit parts do not lead to significant calculational errors. The exact calculations of the magnetic flux distribution in the air gap is a difficult problem, but if the gap length is small and the plane of the ferromagnetic bodies surrounding the gap are parallel, then one can assume that the magnetic field in the gap is homogeneous and its cross-section S_3 is equal to the core cross-section S_1.

We represent the unbranched magnetic circuit as an equivalent network consisting of three series connected resistances R_{M1}, R_{M2}, R_{M3}, and MMF $F = Iw$ (Fig. E6.35a). In this scheme, the resistances R_{M1} and R_{M2} depend on the magnetic flux, and the resistance R_{M3} is a constant quantity. According to Kirchhoff's second law for magnetic circuit, we have

$$F = Iw = H_1 l_1 + H_2 l_2 + 2H_3 l_3.$$

The magnetic inductions B_1, B_2, B_3 are determined by a given value of the magnetic flux by formula (E6.4.1.1)

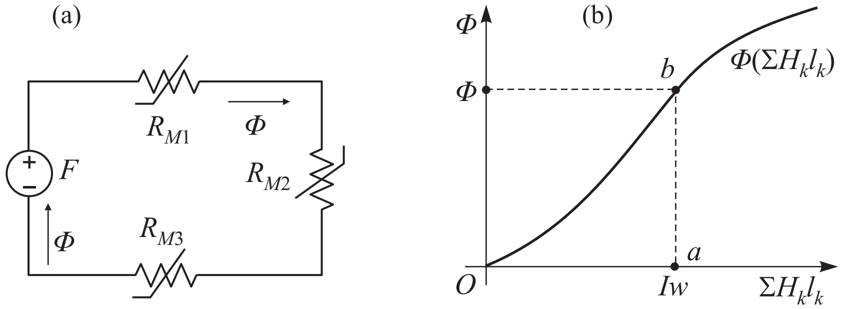

Figure E6.35. (a) Equivalent network of an unbranched magnetic circuit and (b) the auxiliary characteristic $\Phi = f(\sum H_k l_k)$.

From the obtained values B_1 and B_2 of the magnetic induction and the normal magnetization curves for the corresponding materials, we determine the magnetic field strengths H_1 and H_2. For the air gap, we have $H_3 = B_3/\mu_0$. The magnetizing current is calculated as

$$I = \frac{Iw}{w} = \frac{H_1 l_1 + H_2 l_2 + 2H_2 l_3}{w}.$$

► **Second problem for unbranched magnetic circuits.** Now we consider the same magnetic circuit for which it is required to determine the magnetic flux from a given value of MMF. In contrast to the preceding problem, this problem does not have a "direct" solution because of the nonlinear relation between the flux and the magnetizing current.

Such a solution can be obtained, for example, by the following method. First, we prescribe a value of the magnetic flux, for example, Φ'. Then, just as in the preceding problem, we find the MMF $F' = I'w = \sum (H_k l_k)'$. If the obtained value of MMF coincides with the given $F = Iw$, then the problem is solved. But such a coincidence is usually not obtained after the first attempt. Therefore, we must prescribe other values of the magnetic flux Φ'', Φ''', etc. and then find the corresponding values F'', F''', etc. and construct the auxiliary characteristic $\Phi = \Phi(\sum H_k l_k)$ shown in Fig. E6.35b.

Then we put the value of the given MMF $F = Iw$ (point a) on the abscissa axis and draw a straight line from this point parallel to the ordinate axis till the intersection with the curve $\Phi(\sum H_k l_k)$ at point b. As a result, we obtain segment ab determining the desired value of the magnetic flux.

It is necessary to point out that, in practical calculations, it does not make any sense to construct the entire curve $\Phi(\sum H_k l_k)$ starting from the zero value of the magnetic flux. To obtain the first point of the curve, we must equate the given MMF with the magnetic voltage on the site with maximal magnetic resistance $R_M \Phi' = Iw$ and, using the magnetization curve, determine the maximal flux Φ' from this equation.

If there is an air gap in the magnetic circuit, then most often the gap is the site with maximal magnetic resistance. For the air gap, in the case of the circuit in Fig. E6.34, the last equation can be written as

$$\frac{2l_3}{\mu_0 S_3}\Phi' = Iw \qquad \text{or} \qquad \Phi' = \frac{\mu_0 S_3 Iw}{2l_3}.$$

Since the other parts of the same unbranched magnetic circuit, just as the part with maximal magnetic resistance, bound the magnetic flux, its subsequent values necessary to construct the curve $\Phi(\sum H_k l_k)$ must be taken less than Φ'.

Since the magnetic flux is the same in all parts of the unbranched magnetic circuit, the quantity $\Phi(\sum H_k l_k)$ can be constructed by analogy with the unbranched electric circuit by graphical summation of the abscissas of the straight line $\Phi(2H_3 l_3)$ and the curves $\Phi(H_1 l_1)$ and $\Phi(H_2 l_2)$ for the same values of the magnetic flux. The quantity $\Phi(2H_3 l_3)$ is described by the straight line passing through the origin. Its construction becomes easier if the magnetic field strength H_3 is found for each value of Φ, $H_3 = \Phi/(\mu_0 S_3)$.

The quantities $\Phi(H_1 l_1)$ and $\Phi(H_2 l_2)$ are constructed by using the normal magnetization curves for the materials of the first and second parts of the magnetizing circuit. For this, it is necessary to multiply the ordinates of the magnetization curves by the respective cross-sections of the first and second parts ($\Phi = BS$) and to multiply the abscissas by their lengths (Hl).

E6.4.7. Analysis of a Branched Magnetic Circuit

Analysis of branched magnetic circuits is based on the use of Kirchhoff's laws for magnetic circuits. Because of a nonlinear relation between the induction and the magnetic field strength for ferromagnetic materials, the analysis of such circuits is often performed by graphic-analytical methods by analogy with methods used to analyze nonlinear electric circuits.

When analyzing a magnetic circuit, just as in the case of an electric circuit, it is necessary to prescribe the positive direction of MMF and of magnetic fluxes (if they are not already given).

In Fig. E6.36, we present an example of a branched magnetic circuit with a single MMF and its equivalent network.

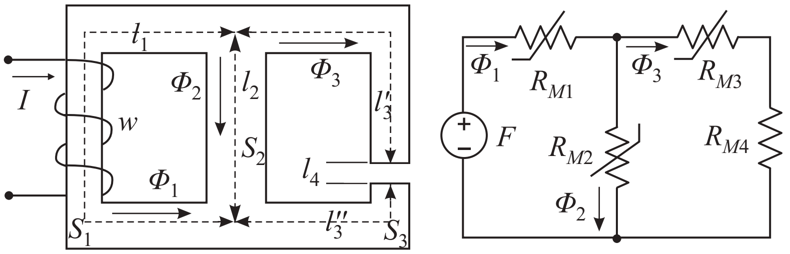

Figure E6.36. Branched magnetic circuit.

For such a magnetic circuit, “direct” computations are possible if the center rod has the same cross-section along the entire length and is produced of the same material as the other rods. The magnetic flux Φ_3 in the air gap must also be given.

From the already known magnetic flux Φ_3, we calculate the induction $B_1 = \Phi/S_1'$; from the magnetization curve, we find the magnetic field strength H_3 and the magnetic field strength in the air gap. The magnetic voltage in the third branch, i.e., between the nodes a and b, is determined as

$$U_{Mab} = \sum H_k l_k = H_3(l_3' + l_3'') + H_4 l_4.$$

Since the second and third branches are connected in parallel, it follows that $U_{Mab} = H_2 l_2$. After H_2 is computed, we find B_2 from the magnetization curve. The fluxes are calculated as follows: $\Phi_2 = B_2 S_2$, $\Phi_1 = \Phi_2 + \Phi_3$. We determine the flux Φ_1, calculate the magnetic induction B_1, and then find the magnetic field strength H_1 from the magnetization curve. According to Kirchhoff’s second law, we have

$$F = Iw = H_1 l_1 + U_{Mab}.$$

A more general problem of calculating all fluxes for a given MMF can be solved by a graphical method by analogy with the calculations of direct current circuits with nonlinear elements. With the magnetic leakage fluxes taken into account, such a problem can be solved by the successive approximation method.

E6.4.8. Analysis of a Branched Homogeneous Magnetic Circuit with Permanent Magnet

The analysis of such a circuit cam be reduced to determining the value of magnetic induction in air gap (Fig. E6.37), which depends on the shape of the magnetization curve $B = f(H_M)$ and the relation between the air gap length l_a and the magnetic circuit ferromagnetic part length l_M, i.e., on the value $N = l_a/l_M$, which is the demagnetization factor if the air gap cross-section coincides with the ferromagnetic part cross-section.

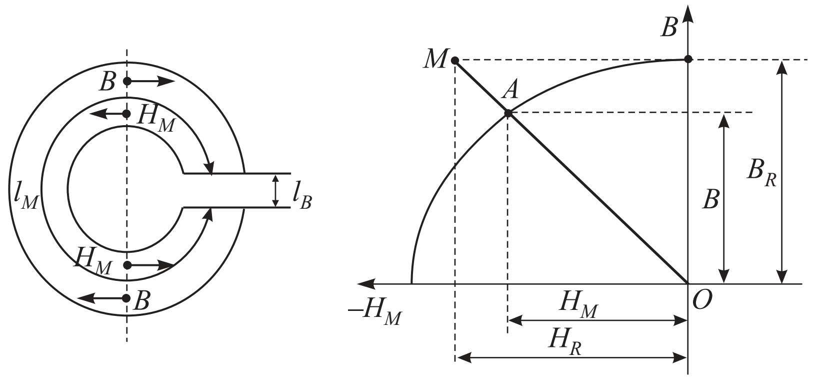

Figure E6.37. Magnetic induction in a magnetic gap and the magnetization curve.

Applying Ampère’s circuital law along the center magnetic line of a permanent magnet and taking into account the fact that there is no winding with magnetizing current, we obtain

$$H_M l_M + H_a \delta = 0,$$

where H_M and H_a are the magnetic field strengths in a ferromagnetic material and in the air gap, respectively; and l_M and δ are the respective lengths of the center magnetic line on the same parts.

It follows from the above that

$$H_M = -\frac{\delta}{l_M} H_a = -\frac{\delta}{\mu_0 l_m} B = -\frac{NB}{\mu_0}, \tag{E6.4.8.1}$$

where B is the value of the magnetic induction in the air gap and in the magnet body, provided that the air gap cross-section is equal in area to magnet cross-section and there is no magnetic leakage flux.

Since the quantities H_M and B are related to each other by the magnetization curve $B = f(H_M)$ and Eq. (E6.4.8.1) of the straight line OM, it follows that the ordinate of point A, which is the point of intersection of these lines, determines the value of the desired magnetic induction B. To construct the straight line OM, it suffices to connect the origin with point M whose ordinate is B_R, which is the residual magnetic induction, and whose abscissa is H_R.

The same calculations can be performed for a small length of the air gap. Otherwise, the pattern of the magnetic field distribution becomes more complicated, and it is necessary to pass from analytical calculations to experimental studies.

E6.5. Single-Phase Harmonic Current Circuits

E6.5.1. Effective and Average Values of Sinusoidal Quantities

► **Effective values of sinusoidal quantities.** The electromotive force e varying in time by a sinusoidal law (see Fig. E6.38) can be obtained by rotating an open rectangular wire loop in a homogeneous magnetic field with constant angular velocity about an axis perpendicular to the field lines.

The equation of sinusoidal EMF has the form

$$e = E_\mathrm{m} \sin(\omega t),$$

where E_m is its maximal value or the *amplitude*.

Under the action of a sinusoidal EMF, in an electric circuit, there arises a sinusoidal or harmonic current whose instantaneous value is determined by the formula

$$i = I_\mathrm{m} \sin(\omega t + \psi).$$

The *sinusoidal current* is a special case of alternating current whose value and direction vary periodically by the sinusoidal law.

The sinusoidal current has significant advantages over the alternating current of any other shape. The derivative and the integral of a sinusoidal function are again sinusoidal functions. Therefore, in the process of transformation of electric energy transferred over a distance, the shape of the curve of sinusoidal current remains the same. The power losses in sinusoidal current circuits are minimal, and the theoretical analysis of sinusoidal current electric circuits are simpler.

A sinusoid is characterized by the following three parameters:

1. The maximal value or amplitude E_m.
2. The *period* T, which is the time of one full wave. The number of waves per second is called the *frequency*. The frequency is the inverse of the period, it is denoted by letter f and is measured in hertzs (Hz). Since the rotation of the wire loop by the angle 2π corresponds to

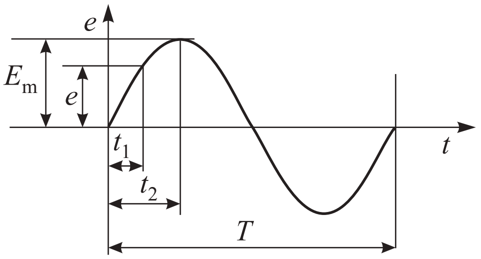

Figure E6.38. Graph of a sinusoidal electromotive force.

the time equal to one period T, it follows that the angular velocity or the *angular frequency* is $\omega = 2\pi f$. The angular frequency is measured in rad/s or s^{-1}.

3. The *phase*, $\omega t + \psi$, is the sine argument. The phase characterizes the state of oscillations at a given time moment t. The angle ψ is equal to the phase at the initial time moment ($t = 0$) and hence is called the *initial phase*. The *phase displacement* between two sinusoidal quantities with equal frequencies is the difference between their initial phases (in other words, the phase shifting, phase shift, phase difference).

► **Average values of sinusoidal quantities.** The average value of a sinusoidally varying quantity is its value averaged over a half-period. For example, the average value of current is determined as

$$\langle I \rangle = \frac{2}{T} \int_0^{T/2} I_m \sin(\omega t)\, dt = \frac{2}{\pi} T_m \approx 0.638\, I_m.$$

The notion of *effective value* of a sinusoidally varying quantity is used very widely. Electric energy transformed into heat energy is proportional to the squared current (the Joule(–Lenz) law). To evaluate the effects of alternating currents, EMF, and voltages, the notion of the quadratic mean over a period, called the *active* or *effective value*, is introduced. For example, the effective value of current is

$$I = \sqrt{\frac{1}{T} \int_0^T i^2\, dt} = \sqrt{\frac{1}{T} \int_0^T I_m^2 \sin^2(\omega t)\, dt} = \frac{I_m}{\sqrt{2}} \approx 0.707\, I_m. \qquad \text{(E6.5.1.1)}$$

As a rule, the electrical measuring instruments of direct estimation show the effective values of the measured quantities.

The effective value of alternating current is numerically equal to the direct current for which, in the time interval equal to one period, the same quantity of heat is released in the resistance R as for the alternating current.

E6.5.2. Representation of Sinusoidal Quantities by Vectors on Complex Plane

► **Sum of two sinusoidal quantities. The vector diagram.** In alternating current circuits, Kirchhoff's laws hold for instantaneous values. Their direct application to sinusoidal current circuits leads to addition and transformation of trigonometric functions, and it is convenient to perform these operations by using complex numbers.

The instantaneous values of any sinusoidal quantities, for example, of the currents $i_1 = I_{m1} \sin(\omega t + \psi_1)$ and $i_2 = I_{m2} \sin(\omega t + \psi_2)$, can be represented at each time instant t by the projections of the amplitude vectors I_{m1} and I_{m2} onto the vertical axis as is shown

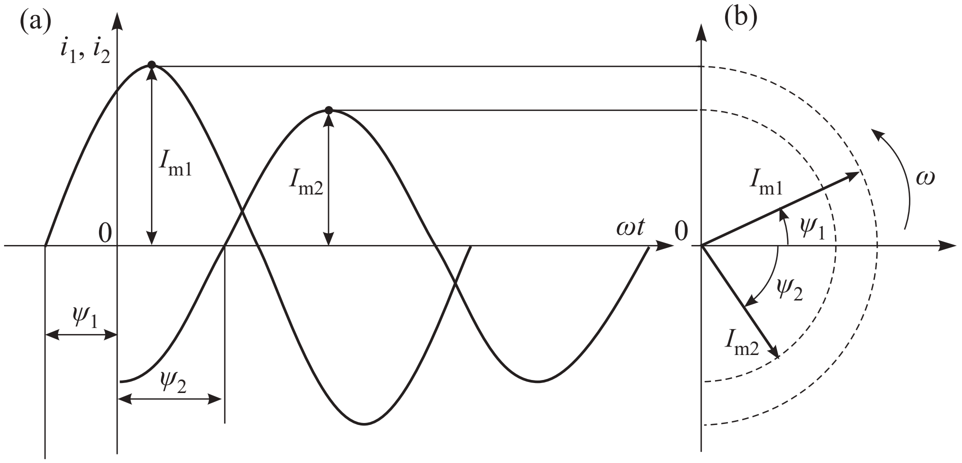

Figure E6.39. Determining the instantaneous value of the sum of two sinusoidal currents (a) by summing the projections of the amplitude vectors (b).

in Fig. E6.39. In this case, the vectors rotate in the plane with constant angular velocity ω about the point 0 anticlockwise. It should be noted that the vectors I_{m1} and I_{m2} rotate about the origin with the same angular velocity ω, and their mutual disposition remains unchanged. Therefore, they are usually represented at time $t = 0$.

The sum of instantaneous values of two sinusoidal currents i_1 and i_2 of the same frequency f can be determined as the projection of the amplitude vector I_m equal to the geometric sum of the summand vectors I_{m1} and I_{m2} and is presented in Fig. E6.40. The entire set of rotating vectors, constructed with preservation of the mutual orientation, is called the vector diagram, permits finding the sum or difference of several harmonic quantities, and reflects the processes in a sinusoidal current electric circuit.

► **Vector of current on complex plane. Complex effective value.** A complex number can be represented as a vector on plane with two mutually perpendicular reference axes for quantities: the real axis (denoted by +1) and the imaginary axis (denoted by $+j$, where $j^2 = -1$). The length of a vector in a chosen scale characterizes the absolute value (or modulus) of the complex number, and the angle of rotation of the vector about the real axis is the argument of the complex number. The positive angle is counted counterclockwise, the negative angle is counted clockwise.

Remark. In electrical engineering, the imaginary unit is denoted by the symbol j, which differs from its common notation i in mathematics (see Chapter M9). This is due to the fact that the symbol i in electrical engineering usually denotes the current.

A complex number can be written in algebraic, exponential, or trigonometric form. For example, to the current $i_m = I_m \sin(\omega t + \psi)$ represented in Fig. E6.41 by a rotating vector there corresponds the complex number

$$a + jb = I_m e^{j(\omega t+\psi)} = I_m \cos(\omega t + \psi) + jI_m \sin(\omega t + \psi),$$

written by using the Euler formula $e^{j\alpha} = \cos\alpha + j\sin\alpha$. The imaginary part b of this complex number is equal to the current i_m.

The amplitude is determined by the formula $I_m = \sqrt{a^2 + b^2}$, and the phase angle can be found from the projections of the vector onto the axes:

$$\sin(\omega t + \psi) = \frac{b}{\sqrt{a^2 + b^2}}, \qquad \cos(\omega t + \psi) = \frac{a}{\sqrt{a^2 + b^2}}.$$

In what follows, the vectors of various sinusoidal quantities are denoted by bold letters.

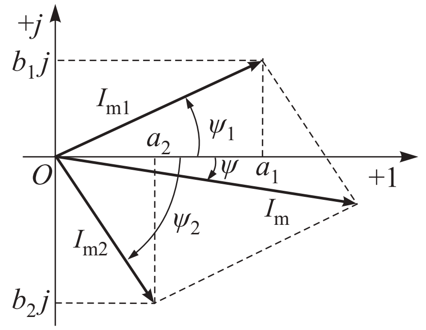

Figure E6.40. Vector summation of current vectors.

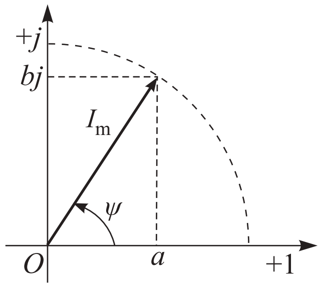

Figure E6.41. Vector of current on complex plane.

Example 1. Let us determine the resultant current i in Fig. E6.40. To this end, we represent each of the given currents in terms of complex amplitudes

$$\mathbf{I}_1 = I_{m1}e^{j\psi_1}, \qquad \mathbf{I}_2 = I_{m2}e^{j\psi_2}.$$

The geometric sum of vectors $\mathbf{I}_1$ and $\mathbf{I}_2$ gives the complex amplitude of the total current, $\mathbf{I} = I_m e^{j\psi}$. The amplitude of the current I_m is determined by the length of the sum vector, and the initial phase ψ is determined by the angle made by this vector with the axis +1.

This example can also be solved analytically by summing the corresponding complex numbers represented in algebraic form:

$$\mathbf{I} = \mathbf{I}_1 + \mathbf{I}_2 = a_1 + jb_1 + a_2 + jb_2 = (a_1 + a_2) + j(b_1 + b_2).$$

The amplitude and the phase angle are determined by the formulas

$$I_m = \sqrt{(a_1 + a_2)^2 + (b_1 + b_2)^2}, \qquad \tan\psi = \frac{b_1 + b_2}{a_1 + a_2}.$$

One can see that, in summation and subtraction of complex numbers, it is convenient to use their algebraic representation, and in multiplication (division), it is more convenient to use the exponential representation.

In the analysis of electric circuits, it is often necessary to pass from the algebraic representation of a complex number to the exponential representation or vice versa. In addition to the complex amplitude, the *complex effective value* is also used, which is introduced as follows:

$$\mathbf{I} = \frac{I_m}{\sqrt{2}}e^{j\psi}.$$

Here $\sqrt{2}$ in the denominator arises by analogy with the expression for the effective current (E6.5.1.1).

We consider two numerical examples of passing from the instantaneous value of current to the complex current and conversely.

Example 2. Express the complex effective value of the following current:

$$i = 3\sin\Bigl(\omega t + \frac{\pi}{4}\Bigr).$$

In this example, $I_m = 3$, and the initial phase is $\psi = \pi/4$. Therefore,

$$\mathbf{I} = \frac{I_m}{\sqrt{2}}e^{j\psi} = \frac{3}{\sqrt{2}}e^{j\pi/4} = 2.1213\Bigl(\cos\frac{\pi}{4} + j\sin\frac{\pi}{4}\Bigr) = 1.5 + 1.5\,j.$$

Example 3. Write the expression of the instantaneous value of current if its complex effective value has the form

$$\mathbf{I} = 2e^{-j\pi/3}.$$

We find the amplitude value $I_m = 2\cdot\sqrt{2} = 2.8284$ and the initial phase $\psi = -\pi/3$. The instantaneous value of the current is $i = 2.8284\sin\Bigl(\omega t - \dfrac{\pi}{3}\Bigr)$.

Remark. In alternating current circuits, one can sum sinusoidal quantities (currents and voltages) that have the same frequency and different initial phases by using the trigonometric formulas

$$\sum_{k=1}^{n} A_k \sin(\omega t+\psi_k)=\sum_{k=1}^{n} A_k[\cos(\psi_k)\sin(\omega t)+\sin(\psi_k)\cos(\omega t)]=A_c\sin(\omega t)+A_s\cos(\omega t)=A\sin(\omega t+\psi),$$

where

$$A_c=\sum_{k=1}^{n} A_k\cos\psi_k,\quad A_s=\sum_{k=1}^{n} A_k\sin\psi_k,\quad A=\sqrt{A_c^2+A_s^2},\quad \sin\psi=\frac{A_s}{A},\quad \cos\psi=\frac{A_c}{A}.$$

E6.5.3. Active Resistance in a Sinusoidal Current Circuit

► **Circuit with active resistance. Vector diagram.** Figure E6.42 presents a circuit with active resistance to which the sinusoidal voltage is applied:

$$u=U_\mathrm{m}\sin(\omega t+\psi_U).$$

The current can be determined using Ohm's law for instantaneous values

$$i=\frac{u}{R}=\frac{U_\mathrm{m}}{R}\sin(\omega t+\psi_U)=I_\mathrm{m}\sin(\omega t+\psi_U).$$

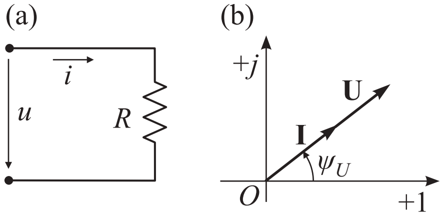

Figure E6.42. (a) A sinusoidal current circuit with active resistance and (b) the vector diagram for such a circuit.

Comparing the expressions for the instantaneous values of voltage and current, one can see that there is no phase shift between voltage and current (Fig. E6.42a), which is determined by the angle equal to the difference of the voltage and current initial phases, i.e.,

$$\varphi=\psi_U-\psi_I=0.$$

Thus, the current and the voltage have the same initial phase (coincide in phase).

► **Instantaneous and average value of power. Ohm's law in complex form.** The instantaneous value of the power of this circuit is equal to the product of instantaneous values of its voltage and current:

$$p=ui=U_\mathrm{m}I_\mathrm{m}\sin^2(\omega t+\psi_U)=UI[1-\cos(2\omega t+2\psi_U)],$$

where $I=I_\mathrm{m}/\sqrt{2}$ and $U=U_\mathrm{m}/\sqrt{2}$ are the effective current and effective voltage. The instantaneous power is measured by doubled angular frequency 2ω and is all the time positive in the limits from zero to the amplitude value.

We calculate the average value of power over a period:

$$P=\frac{1}{T}\int_0^T p\,dt=UI.$$

Thus, the average value of power in this circuit is determined by the product of the effective values of voltage and current, is called the *active power*, and is measured in watts (W).

Ohm's law in complex form for a circuit with active resistance has the form

$$\mathbf{U}=R\mathbf{I}.$$

E6.5.4. Inductance in a Sinusoidal Current Circuit

► **Circuit with inductance. Vector diagram. Inductive reactance.** Consider a circuit (Fig. E6.43) with induction coil of inductance L, whose active resistance is negligibly small ($R_{\text{ric}} \approx 0$). Assume that sinusoidal current flows in this circuit:

$$i = I_{\text{m}} \sin(\omega t - \pi/2) = -I_{\text{m}} \cos(\omega t).$$

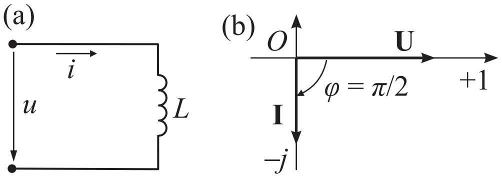

Figure E6.43. (a) A sinusoidal current circuit with inductance and (b) the vector diagram for such a circuit.

The EMF of self-induction induced in the inductor is calculated as

$$e_L = -L\frac{di}{dt} = -\omega L I_{\text{m}} \sin(\omega t) = E_{Lm} \sin(\omega t).$$

Here $E_{Lm} = -\omega L I_{\text{m}}$ is the amplitude.

According to Kirchhoff's second law, we have

$$e_L + u = 0,$$

which implies the instantaneous value of voltage applied to the inductor terminals

$$u_L = -e_L = \omega L I_{\text{m}} \sin(\omega t) = U_{Lm} \sin(\omega t).$$

The effective voltage U and effective current I are related according to Ohm's law as follows:

$$I = \frac{U}{\omega L} = \frac{U}{X_L},$$

where the quantity $X_L = \omega L = 2\pi f L$ is called the *inductive reactance*. The inductive reactance is directly proportional to the coil inductance and the frequency of the current flowing through it.

The angle of phase shift between voltage and current is determined by the difference of their initial phases:

$$\varphi = \psi_U - \psi_I = 0 - (-\pi/2) = \pi/2.$$

It is clear that the inductor voltage phase vector is $\pi/2$ ahead of the current vector.

► **Instantaneous value of power. Ohm's law for a circuit with inductance.** The instantaneous power in a circuit with inductance can be found as follows:

$$p_L = ui = -\frac{1}{2} U_{\text{m}} I_{\text{m}} \sin(2\omega t) = -UI \sin(2\omega t).$$

In the first quarter of the period, where the current i and the voltage u are positive (of the same sign), the power is also positive. The energy is transferred from source to circuit and is consumed to create a magnetic field. At the end of the first quarter of the period, the field energy is maximal. In the second quarter of the period, the current i remains positive but decreases, and the voltage u is negative; then the power p_L is negative. The magnetic field energy is transferred back into the circuit. The average value of power in such a circuit is zero:

$$P_L = \frac{1}{T}\int_0^T p_L\,dt = 0.$$

Thus, in a circuit with inductance, the permanent energy oscillation (exchange) between the circuit and the magnetic field occurs without the source energy consumption.

The amplitude power in an inductive circuit is determined by the formula

$$Q_L = UI = I^2\omega L = 2\pi f L I^2,$$

is called the *reactive inductive power*, is determined by the product of the effective voltage and effective current, and is measured in reactive volt-amperes (RVA).

Ohm's law in complex form for a circuit with inductance has the form

$$\mathbf{U} = j\omega L\mathbf{I} = jX_L\mathbf{I}.$$

E6.5.5. Capacitance in a Sinusoidal Current Circuit

► **Circuit with capacitance. Vector diagram. Capacitive resistance.** Assume that the sinusoidal voltage is applied to a capacitor of capacitance C (Fig. E6.44):

$$u = U_\mathrm{m}\sin(\omega t).$$

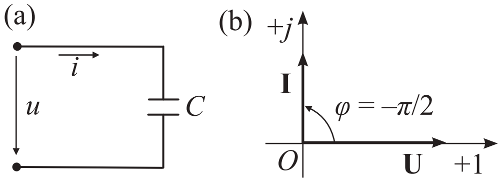

Figure E6.44. (a) A sinusoidal current circuit with capacitance and (b) the vector diagram for such a circuit.

The current flowing through the capacitor is equal to

$$i = C\frac{dU_C}{dt} = \omega C U_{Cm}\cos(\omega t) = I_\mathrm{m}\sin(\omega t + \pi/2),$$

where I_m is the current amplitude,

$$I_\mathrm{m} = \omega C U_{Cm} = \frac{U_{Cm}}{X_C}.$$

The quantity $X_C = \dfrac{1}{\omega C} = \dfrac{1}{2\pi f C}$ is called the *capacitive reactance*. The capacitive resistance is inversely proportional to the frequency of the feeding source and the capacitor capacitance.

The phase shift angle between voltage and current is determined by the difference of their initial phases:

$$\varphi = \psi_U - \psi_I = 0 - \pi/2 = -\pi/2.$$

It is clear that the capacitor voltage phase vector is $\pi/2$ behind the current vector.

► **Instantaneous value of power. Ohm's law for a circuit with capacitance.** The instantaneous power in a circuit with capacitance:

$$p_C = ui = UI\sin(2\omega t).$$

In the first quarter of the period, where the voltage u and the current i have the same sign, the capacitor consumes energy from the feeding source, which created the electric field of the capacitor. In the second quarter of the period, the current i is negative, and the capacitor voltage u, still remaining positive, decreases from its maximum to zero, and the energy stored in the electric field is transferred back to the source, i.e., the instantaneous power is negative. Then the processes repeat. Thus, the period-average in a sinusoidal current circuit with capacitance is zero:

$$P_C = \frac{1}{T}\int_0^T p_C\,dt = 0.$$

The power amplitude in a capacitive circuit is determined by the product of the effective voltage and effective current

$$Q_C = UI = U^2\omega C = 2\pi f C U^2$$

and is called the *reactive capacitive power*; Q_C is measured in reactive volt-amperes (RVA).

Ohm's law in complex form for a circuit with capacitance has the form:

$$\mathbf{U}_C = -jX_C\mathbf{I}.$$

E6.5.6. Unbranched Electric Circuit

► **Kirchhoff's second law. Reactive resistance. Voltage and impedance triangles.** If the elements in an unbranched electric circuit do not affect one another, then the circuit can always be reduced to an equivalent circuit with three elements each of which is characterized by one of the parameters R, L, and C, as is shown in Fig. E6.45.

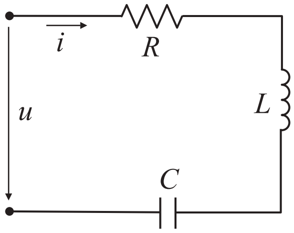

Figure E6.45. Unbranched sinusoidal current circuit.

In the analysis of such circuits, just as in the case of direct current circuits, it is necessary to introduce arbitrary positive directions of voltages and currents.

The instantaneous values of voltages in unbranched circuits always satisfy Kirchhoff's second law equation:

$$u = u_R + u_C + u_L.$$

Kirchhoff's second law equation can be written in complex form:

$$\mathbf{U} = \mathbf{U}_R + \mathbf{U}_C + \mathbf{U}_L. \tag{E6.5.6.1}$$

The voltages on separate elements are calculated by the formulas

$$\mathbf{U}_R = R\mathbf{I}, \quad \mathbf{U}_C = -jX_C\mathbf{I}, \quad \mathbf{U}_L = jX_L\mathbf{I}. \tag{E6.5.6.2}$$

Substituting expressions (E6.5.6.2) into (E6.5.6.1), we obtain the complex voltage at the terminals of a sinusoidal current circuit:

$$\mathbf{U} = R\mathbf{I} - jX_C\mathbf{I} + jX_L\mathbf{I}.$$

This formula can be represented as

$$\mathbf{U} = (R + jX)\mathbf{I},$$

where $X = X_L - X_C$ is the *reactance of the circuit* (reactive resistance), which can be both positive and negative.

The quantity $Z = R + jX$ is called the *circuit impedance* (total complex resistance); it can be represented in exponential form:

$$Z = ze^{j\varphi},$$

where $z = \sqrt{R^2 + X^2}$ is the modulus of the complex number Z and φ is the argument of the complex number Z, which is equal to the phase shift angle in the circuit between voltage and current ($\tan\varphi = X/R$).

In Fig. E6.46, we present the vector diagrams of an unbranched sinusoidal current circuit, which are constructed according to Kirchhoff's second law equation, and in Fig. E6.46a, we consider the case of reactance $X > 0$; i.e., the inductive load predominates in the circuit and the current is by a positive angle φ behind the voltage; in Fig. E6.46b, we consider the case of reactance $X < 0$; i.e., the capacitive load predominates in the circuit and the current is by a negative angle φ ahead of the applied voltage.

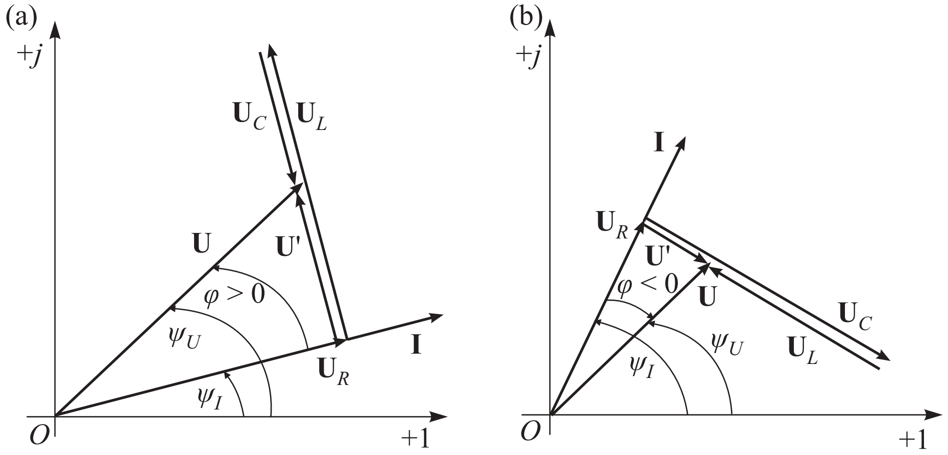

Figure E6.46. Vector diagrams of unbranched sinusoidal current circuit with predominating loads: (a) inductive load; (b) capacitive load.

The rectangular triangle with legs U_R and $U' = U_L + U_C$ and hypotenuse U is called the *voltage triangle*. One can readily pass from the voltage triangle to the similar *impedance triangle* shown in Fig. E6.47 with legs R and jX and hypotenuse Z.

The position of the impedance triangle is independent of the initial phases ψ_U and ψ_I, because the active resistance R is always put on the complex plane in the positive direction

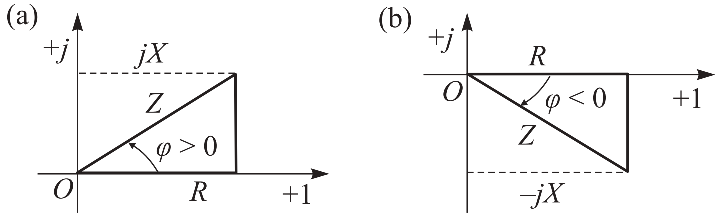

Figure E6.47. Impedance triangles with predominating loads: (a) inductive load; (b) capacitive load.

of the real axis, and the reactance X is put, depending on the sign, in the positive or negative direction of the imaginary axis.

► **Instantaneous value of power, active power, total power, and reactive power.** Assume that the voltage initial phase ψ_U is zero and the instantaneous values of voltage and current are respectively described by the formulas

$$u = U_{\rm m}\sin(\omega t), \qquad i = I_{\rm m}\sin(\omega t-\varphi).$$

We obtain the following expression for the instantaneous value of power

$$\begin{aligned} p = iu = U_{\rm m}I_{\rm m}\sin(\omega t)\sin(\omega t-\varphi) &= \frac{1}{2}U_{\rm m}I_{\rm m}[\cos\varphi-\cos(2\omega t-\varphi)] \\ = UI[\cos\varphi-\cos(2\omega t-\varphi)]. \end{aligned}$$

We find the period-average power (or the active power):

$$P = \frac{1}{T}\int_0^T iu\,dt = UI\cos\varphi,$$

which is determined by half the product of the effective values of voltage, current, and power factor.

The power factor is determined by the formula

$$\cos\varphi = \frac{P}{UI} = \frac{P}{S},$$

where the quantity $S = UI$ is called the *total power*; it is measured in volt-amperes (VA) and is the power for given effective values of voltage and current, when the circuit power factor is equal to one, $\cos\varphi = 1$.

In the analysis of electric circuits, the notion of *reactive power* is also used:

$$Q = UI\sin\varphi,$$

which can be both positive (for $\varphi > 0$) and negative (for $\varphi < 0$), and its absolute value is determined by the formula

$$|Q| = \sqrt{S^2 - P^2}.$$

The *complex total power* is determined as

$$\mathbf{S} = \mathbf{U}\bar{\mathbf{I}} = UIe^{j(\psi_U-\psi_I)} = UIe^{j\varphi} = UI\cos\varphi + jUI\sin\varphi = P + jQ,$$

where $\bar{\mathbf{I}} = Ie^{-j\psi_I}$ is the complex conjugate current.

E6.5.7. Branched Electric Circuit

► **Complex values of currents. The circuit susceptance and the complex admittance.** In the case of parallel connection of elements with parameters L, C, R attached to a source of sinusoidal voltage (Fig. E6.48), the following equation following from Kirchhoff's first law holds:

$$\mathbf{I} = \mathbf{I}_R + \mathbf{I}_C + \mathbf{I}_L;$$

i.e., the complex value of the current in the unbranched part of the circuit is equal to the algebraic sum of complex values of currents in separate branches.

Since the current in the branch with active resistance R coincides in phase with the applied voltage, the current in the branch with inductance L is $\pi/2$ behind the voltage, and the current in the branch with capacitance C is $\pi/2$ ahead of the voltage, we obtain:

$$\mathbf{I}_R = \frac{\mathbf{U}}{R} = g\mathbf{U}, \quad \mathbf{I}_L = -j\frac{\mathbf{U}}{X_L} = -jb_L\mathbf{U}, \quad \mathbf{I}_C = j\frac{\mathbf{U}}{X_C} = jb_C\mathbf{U},$$

where g, b_L, and b_C are, respectively, the (active) conductance, inductive susceptance, and capacitive susceptance of separate branches of the branched sinusoidal current electric circuit under study.

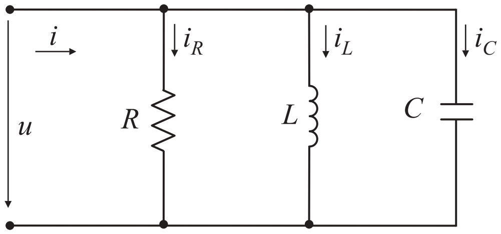

Figure E6.48. Branched alternating current electric circuit.

Let us determine the current in the unbranched part of the circuit:

$$\mathbf{I} = g\mathbf{U} - j(b_L - b_C)\mathbf{U} = (g - jb)\mathbf{U},$$

where $b = b_L - b_C$ is the *circuit susceptance*, which can be both positive and negative.

The quantity

$$Y = g - j(b_L - b_C) = g - jb$$

is called the *complex admittance of a circuit* and can be written in exponential form

$$Y = ye^{-j\varphi},$$

where $y = \sqrt{g^2 + b^2}$ is the modulus of the complex number Y and $\varphi = \arctan(b/g)$ is the angle of phase shift between the applied voltage and the current in the unbranched part of the circuit.

The complex admittance Y of any part of the circuit with parameters R, L, C is related to the complex impedance Z by the inverse relationship:

$$Y = \frac{1}{Z} = \frac{1}{R + jX} = \frac{R}{R^2 + X^2} - j\frac{X}{R^2 + X^2} = \frac{R}{z^2} - j\frac{X}{z^2} = g - jb,$$

where $z = \sqrt{R^2 + X^2}$ is the impedance, $g = R/z^2$ is the conductance, $b = X/z^2$ is the susceptance.

► **Vector diagrams and admittance triangles.** In Fig. E6.49, we present the vector diagrams constructed on the complex plane for an electric circuit with parallel connection of elements with the parameters R, L, C; the first of them corresponds to the case of susceptance $b > 0$, i.e., the inductive load predominates in the circuit and the current in the unbranched part of the circuit is by a positive angle φ behind the voltage, and the second corresponds to the case of susceptance $b < 0$; i.e., the capacitance load predominates in the circuit and the current in the unbranched part of the circuit is by a negative angle φ ahead of the voltage.

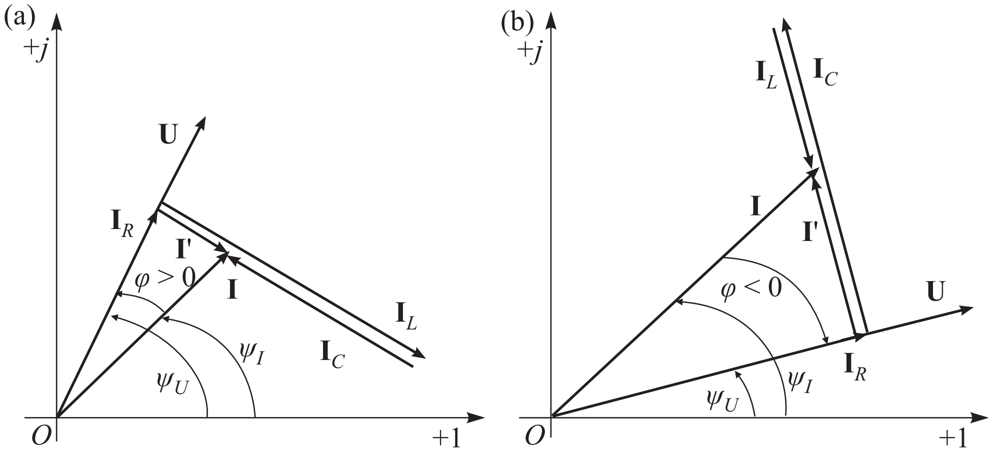

Figure E6.49. Vector diagrams of branched sinusoidal current circuit with predominating loads: (a) inductive load, (b) capacitive load.

Dividing the sides of a rectangular current triangle by the complex voltage, it is easy to obtain the admittance triangle (Fig. E6.50), which is similar to the first but rotated clockwise with respect to it by the angle ψ_I. The admittance triangle position is independent of the initial phases ψ_U and ψ_I, because the conductance g is always put on the complex plane in the positive direction of the real axis, and the susceptance b is put, depending on the sign, in the positive (for $b < 0$) or negative (for $b > 0$) direction of the imaginary axis.

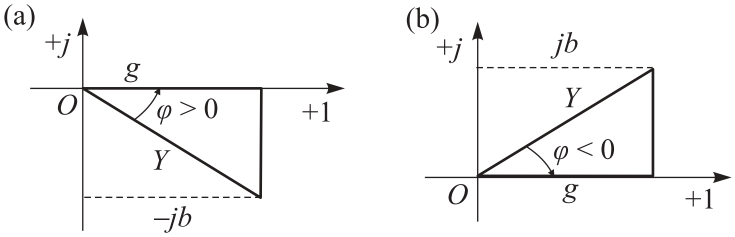

Figure E6.50. Admittance triangles with predominating loads: (a) inductive load, (b) capacitive load.

E6.5.8. Topographic Diagram

In direct current circuits, one constructs graphs of distribution of the potential, which graphically interprets Kirchhoff's second law. In the case of a sinusoidal current circuit, one constructs the topographic vector diagram, which is a set of points on the complex plane representing the complex potentials of the electric network points of the same name with respect to a single point whose potential is conditionally zero. Thus, the order of location of the voltage drop vectors on the diagram strictly corresponds to the order of location of elements in the network.

In Fig. E6.51a, we present the network of an unbranched electric circuit and construct its topographic diagram shown in Fig. E6.51b.

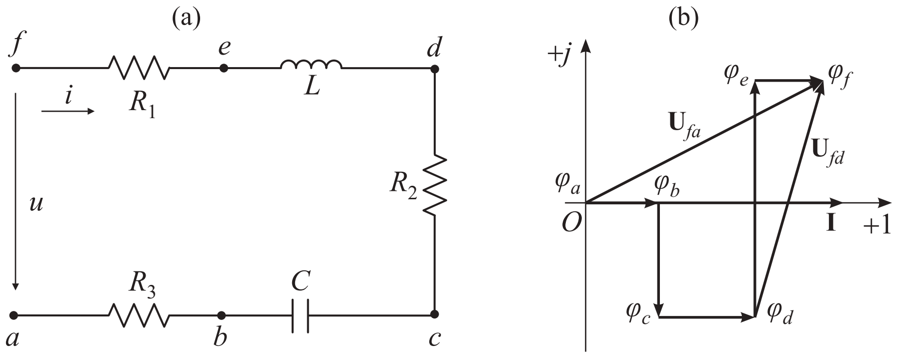

Figure E6.51. Unbranched electric circuit (a) and the topographic diagram (b).

The construction of a topographic diagram starts from calculating the current vector. It is convenient to direct the current along the real axis $\psi_i = 0$. Then one arbitrarily chooses a point whose potential is zero, for example, $\varphi_a = 0$. Thus, point a is at the origin.b

The directions of the voltage vectors on the topographic diagram are shown with arbitrarily chosen direction of the current vector $\mathbf{I}$. The network is traced towards the positive direction of the current $\mathbf{I}$. As we pass to point b, the potential increases by the value of the voltage drop on the resistor R_1 and is equal to

$$\varphi_b = \varphi_a + R_3\mathbf{I} = \mathbf{U}_{ba}.$$

The vector of this voltage drop coincides in phase with the current vector $\mathbf{I}$.

The potential of point c is equal to

$$\varphi_c = \varphi_b + (-jX_C)\mathbf{I}.$$

The voltage $\mathbf{U}_{cb} = (-jX_C)\mathbf{I}$ is by an angle $\pi/2$ behind the current. Similarly, the voltage vectors are constructed for other parts of the circuit. The vector of resultant voltage lies between points a and f.

From the topographic vector diagram, one can readily determine the vector of voltage difference between two arbitrary points of the circuit. For example, to determine the vector of voltage difference between points d and f of the network, it suffices to draw the voltage vector between these points, because

$$\mathbf{U}_{fd} = \varphi_f - \varphi_d.$$

This corresponds to the well-known rule of vector subtraction.

E6.5.9. Resonance Phenomena

► **Preliminary remarks.** The operation mode of an electric circuit containing induction coils and capacitors in which the equivalent (input) reactance or the equivalent (input) susceptance is zero is called the *resonance operation mode*. This operation mode is characterized by the relations

$$X_e = 0, \quad B_e = 0.$$

The electric circuit in resonance behaves as the *pure active resistance* with respect to the external circuit; i.e., the voltage and the current at the circuit input are in phase (coincide in phase).

The following two types of resonance are distinguished: the voltage resonance and the current resonance.

► **Voltage resonance.** The resonance arising in a circuit (Fig. E6.52), where the real induction coil and capacitor connected in series, is called the voltage resonance.

The complex equivalent impedance of such a circuit is

$$Z_{\rm e} = R_{\rm e} + jX_{\rm e} = R_{\rm ic} + jX_{\rm ic} - jX_C = R_{\rm ic} + j(X_{\rm ic} - X_C) = R_{\rm ic} + j\Bigl(\omega L - \frac{1}{\omega C}\Bigr).$$

Equating the imaginary part with zero ($\operatorname{Im} Z_{\rm e} = 0$), we find the voltage resonance condition

$$X_{\rm e} = X_{\rm ic} - X_C = \omega_{\rm r} L - 1/(\omega_{\rm r} C) = 0 \quad\Longrightarrow\quad X_{\rm ic} = X_C \quad \text{or} \quad \omega_{\rm r} L = 1/(\omega_{\rm r} C),$$

where $\omega_{\rm r} = 1/\sqrt{LC}$ is the resonance angular frequency.

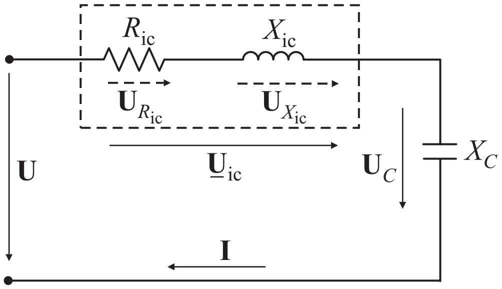
$R_{\rm ic}$
$X_{\rm ic}$
$\mathbf{U}_{R_{\rm ic}}$
$\mathbf{U}_{X_{\rm ic}}$
$\mathbf{U}$
$\underline{\mathbf{U}}_{\rm ic}$
$\mathbf{U}_C$
X_C
$\mathbf{I}$

Figure E6.52. Electric network used to study the voltage resonance.

The voltage resonance can be obtained by changing either the frequency of the power supply voltage or the value of the inductance or capacitance. The first case is called a frequency resonance, and the other two are called parametric resonances.

The inductive (or capacitive) susceptance at the instant of the frequency resonance is called the *characteristic impedance*, is denoted by letter ρ, and is determined as

$$\rho = \omega_{\rm r} L = \frac{1}{\omega_{\rm r} C} = \sqrt{\frac{L}{C}}.$$

The strength of the current flowing in the circuit at the resonance instant is maximal:

$$I_{\rm r} = \max I = \frac{U}{Z_{\rm e}}\bigg|_{X_{\rm e}=0} = \frac{U}{\sqrt{R_{\rm ic}^2 + (X_{\rm ic} - X_C)^2}}\bigg|_{X_{\rm ic}=X_L} = \frac{U}{R_{\rm ic}}.$$

For certain values of capacitance, the capacitor and coil voltages attain their maximal values $\max U_C$ and $\max U_{\rm ic}$, which are determined from the condition that the derivatives are zero:

$$\frac{dU_C}{dC} = 0, \qquad \frac{dU_{\rm ic}}{dC} = 0.$$

If the induction coil is assumed to be ideal (i.e., $R_{\rm ic} = 0$), then the coil and capacitor voltages are equal to each other:

$$X_{\rm ic} I_{\rm r} = X_C I_{\rm r}, \quad U_{\rm ic} = U_C.$$

The ideal voltage resonance is equivalent to the case of shorted input terminals of the circuit (since $I_{\rm r} = U/R_{\rm r} = U/0 = \infty$).

The ratios of the induction coil (or capacitor) voltage to the input voltage in resonance are called the *circuit quality factor* (Q factor):

$$q = \frac{U_{X_{\rm ic}}}{U} = \frac{U_C}{U} = \frac{\rho I_{\rm r}}{R_{\rm ic} I_{\rm r}} = \frac{\rho}{R_{\rm ic}} \qquad \text{or}$$

$$q = \frac{U_{X_{\rm ic}}}{U} = \frac{U_C}{U} = \frac{X_{\rm ic} I_{\rm r}}{R_{\rm ic} I_{\rm r}} = \frac{X_C I_{\rm r}}{R_{\rm ic} I_{\rm r}} = \frac{X_{\rm ic}}{R_{\rm ic}} = \frac{X_C}{R_{\rm ic}}.$$

The quality factor q shows to what extent the induction coil (or capacitor) voltage exceeds the input voltage; it depends on the coil and capacitor parameters.

We write Kirchhoff's second law for this circuit (Fig. E6.52) in complex form:

$$\mathbf{U} = \mathbf{U}_{\rm ic} + \mathbf{U}_C = \mathbf{U}_{R_{\rm ic}} + \mathbf{U}_{X_{\rm ic}} + \mathbf{U}_C = R_{\rm ic}\mathbf{I} + jX_{\rm ic}\mathbf{I} - jX_C\mathbf{I}.$$

We construct the vector diagrams (Fig. E6.53) on the basis of Kirchhoff's second law on the complex plane in the three cases: (a) before resonance ($U_{\rm ic} < U_C$), (b) at the resonance instant ($U_{\rm ic} \cong U_C$), and (c) after resonance ($U_{\rm ic} > U_C$).

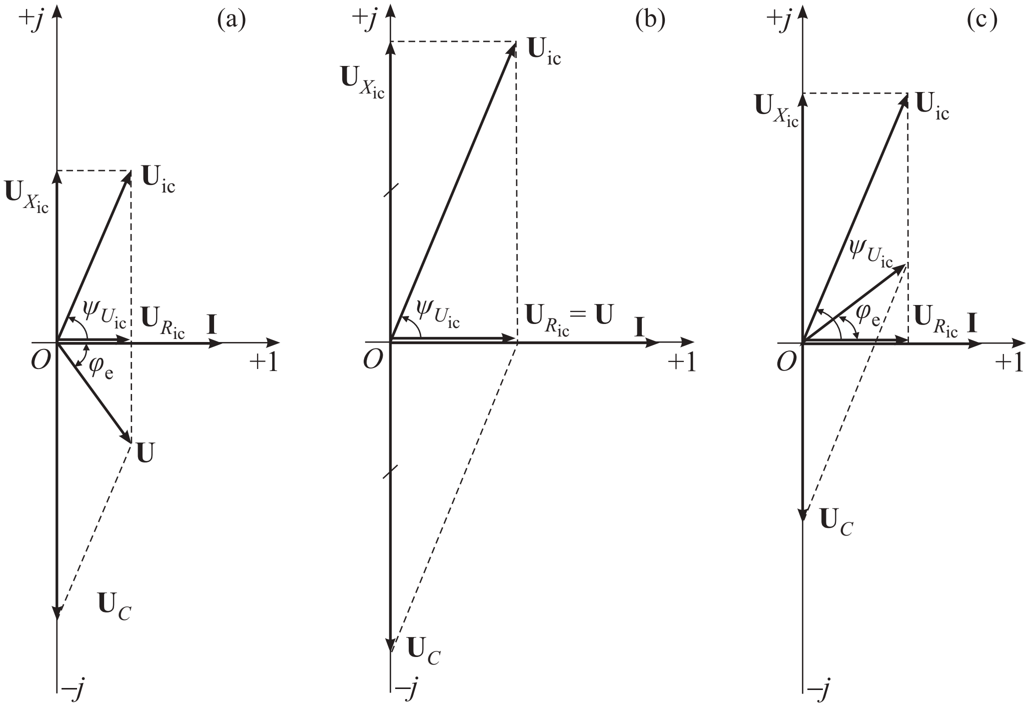

Figure E6.53. Vector diagrams of voltage: (a) before resonance ($U_{\rm ic} < U_C$, $\varphi < 0$), (b) at the resonance instant ($U_{\rm ic} \cong U_C$, $\varphi = 0$), and (c) after resonance ($U_{\rm ic} > U_C$, $\varphi > 0$).

As follows from the vector voltage diagram in the voltage resonance mode, the input current vector I_r coincides in phase with the input voltage vector U; i.e., the phase difference between these vectors is zero:

$$\varphi_\mathrm{e} = \varphi_U - \varphi_I = 0.$$

Here ψ_U is the initial phase of the input voltage, and ψ_I is the initial phase of the input current. In resonance, the equivalent reactive power of the entire circuit is zero:

$$Q_\mathrm{e} = Q_\mathrm{ic} - Q_C = X_\mathrm{ic} I_\mathrm{r}^2 - X_C I_\mathrm{r}^2 = 0,$$

and the equivalent total power of the circuit becomes the pure active power

$$S_\mathrm{e} = \sqrt{P_\mathrm{e}^2 + Q_\mathrm{e}^2} = \sqrt{P_\mathrm{e}^2} = P_\mathrm{e} = R_\mathrm{ic} I_\mathrm{r}^2.$$

In analyzing the parametric voltage resonance, as the capacitance value varies, the resonance curves $I(C)$, $U_\mathrm{ic}(C)$, $U_C(C)$, $\varphi_\mathrm{e}(C)$, $\cos\varphi_\mathrm{e}(C)$ are constructed by the formulas

$$\begin{aligned}
I(C) &= \frac{U}{Z_\mathrm{e}} = \frac{U}{\sqrt{R_\mathrm{ic}^2 + (\omega L - 1/\omega C)^2}},\\
U_\mathrm{ic}(C) &= Z_\mathrm{ic} I = \frac{U\sqrt{R_\mathrm{ic}^2 + (\omega L)^2}}{\sqrt{R_\mathrm{ic}^2 + (\omega L - 1/\omega C)^2}},\\
U_C(C) &= X_C I = \frac{U}{\omega C\sqrt{R_\mathrm{ic}^2 + (\omega L - 1/\omega C)^2}},\\
\varphi_\mathrm{e}(C) &= \arctan\frac{X_\mathrm{e}}{R_\mathrm{e}} = \arctan\frac{\omega L - 1/\omega C}{R_\mathrm{ic}},\\
\cos\varphi_\mathrm{e}(C) &= \frac{R_\mathrm{e}}{Z_\mathrm{e}} = \frac{R_\mathrm{ic}}{\sqrt{R_\mathrm{ic}^2 + (\omega L - 1/\omega C)^2}}.
\end{aligned}$$

► **Current resonance**. The resonance arising in a circuit (Fig. E6.54) where a real induction coil and a capacitor are connected in parallel is called the current resonance.

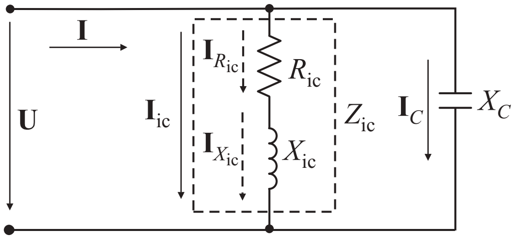

Figure E6.54. Electric network used to study the current resonance.

We find the complex equivalent admittance of such a circuit:

$$\begin{aligned}
\mathbf{Y}_\mathrm{e} = \mathbf{Y}_\mathrm{ic} + \mathbf{Y}_C &= \frac{1}{R_\mathrm{ic} + jX_\mathrm{ic}} + \frac{1}{-jX_C} = \frac{R_\mathrm{ic} - jX_\mathrm{ic}}{(R_\mathrm{ic} + jX_\mathrm{ic})(R_\mathrm{ic} - jX_\mathrm{ic})} + \frac{j}{j(-jX_C)}\\
&= \frac{R_\mathrm{ic}}{R_\mathrm{ic}^2 + X_\mathrm{ic}^2} - j\frac{X_\mathrm{ic}}{R_\mathrm{ic}^2 + X_\mathrm{ic}^2} + j\frac{1}{X_C} = G_\mathrm{ic} - jB_\mathrm{ic} + jB_C = G_\mathrm{ic} - j(B_\mathrm{ic} - B_C),
\end{aligned}$$

where $G_{\rm ic}=\dfrac{R_{\rm ic}}{R_{\rm ic}^2+X_{\rm ic}^2}$, $B_{\rm ic}=\dfrac{X_{\rm ic}}{R_{\rm ic}^2+X_{\rm ic}^2}$, and $B_C=\dfrac{1}{X_C}$ are, respectively, the conductance and the susceptance of the coil and capacitor.

Equating the imaginary part of the obtained expression with zero, $B_{\rm ic}-B_C=0$, we obtain the current resonance condition

$$\frac{X_{\rm ic}}{R_{\rm ic}^2+X_{\rm ic}^2}-\frac{1}{X_C}=\frac{\omega_{\rm r}L_{\rm ic}}{R_{\rm ic}^2+(\omega_{\rm r}L_{\rm ic})^2}-\omega_{\rm r}C=0,$$

where $\omega_{\rm r}$ is the resonance angular frequency.

The current strength at the resonance circuit input at the resonance instant is imaginary

$$I_{\rm r}=\min I=Y_{\rm e}U=\sqrt{G_{\rm ic}^2+(B_{\rm ic}-B_C)^2}\,U=G_{\rm ic}U.$$

If the induction coil is assumed to be ideal ($R_{\rm ic}=0$), then the currents in the coil and capacitor are equal to each other, $I'_{\rm ic}=I'_C$, because $Y'_{\rm ic}=B_CU$ and $B'_{\rm ic}U=B_CU$, where

$$Y'_{\rm ic}=\sqrt{(G'_{\rm ic})^2+(B'_{\rm ic})^2}=\sqrt{0+(B'_{\rm ic})^2}=B'_{\rm ic}=\frac{X_{\rm ic}}{0+X_{\rm ic}^2}=\frac{1}{X_{\rm ic}}.$$

The ideal current resonance is equivalent to the circuit break (open-circuit operation mode), because $I_{\rm r}=G'_{\rm ic}U=0\cdot U=0$.

We write Kirchhoff's first law for the circuit under study in complex form:

$$\begin{aligned}\mathbf{I}&=\mathbf{I}_{\rm ic}+\mathbf{I}_C=\mathbf{Y}_{\rm ic}\mathbf{U}+\mathbf{Y}_C\mathbf{U}=(G_{\rm ic}-jB_{\rm ic})\mathbf{U}+jB_C\mathbf{U}\\&=G_{\rm ic}\mathbf{U}-jB_{\rm ic}\mathbf{U}+jB_C\mathbf{U}=\mathbf{I}_{R_{\rm ic}}+\mathbf{I}_{X_{\rm ic}}+\mathbf{I}_C,\end{aligned}$$

where $\mathbf{I}_{R_{\rm ic}}=G_{\rm ic}\mathbf{U}$ is the active component of the coil current and $\mathbf{I}_{X_{\rm ic}}=-jB_{\rm ic}\mathbf{U}$ is the reactive component of the coil current.

We write Kirchhoff's first law on the complex plane in three cases: (a) before resonance ($I_{\rm ic}>I_C$), (b) at the resonance instant ($I_{\rm ic}\cong I_C$), and (c) after resonance ($I_{\rm ic}<I_C$); i.e., construct the vector diagrams of currents (Fig. E6.55).

As follows from the current diagram (Fig. E6.55), in the current resonance operation mode, the vector of the input voltage coincides in phase with the vector of the input current; i.e., there is no phase shift between these vectors

$$\varphi_{\rm e}=\psi_U-\psi_I=0.$$

In resonance, the equivalent reactive power of the entire circuit is zero

$$Q_{\rm e}=Q_{\rm ic}-Q_C=B_{\rm ic}U^2-B_CU^2=0,$$

and the equivalent total power of the circuit $S_{\rm e}$ is purely active,

$$S_{\rm e}=\sqrt{P_{\rm e}^2+Q_{\rm e}^2}=\sqrt{P_{\rm e}^2}=P_{\rm e}=P_{\rm ic}=G_{\rm ic}U^2.$$

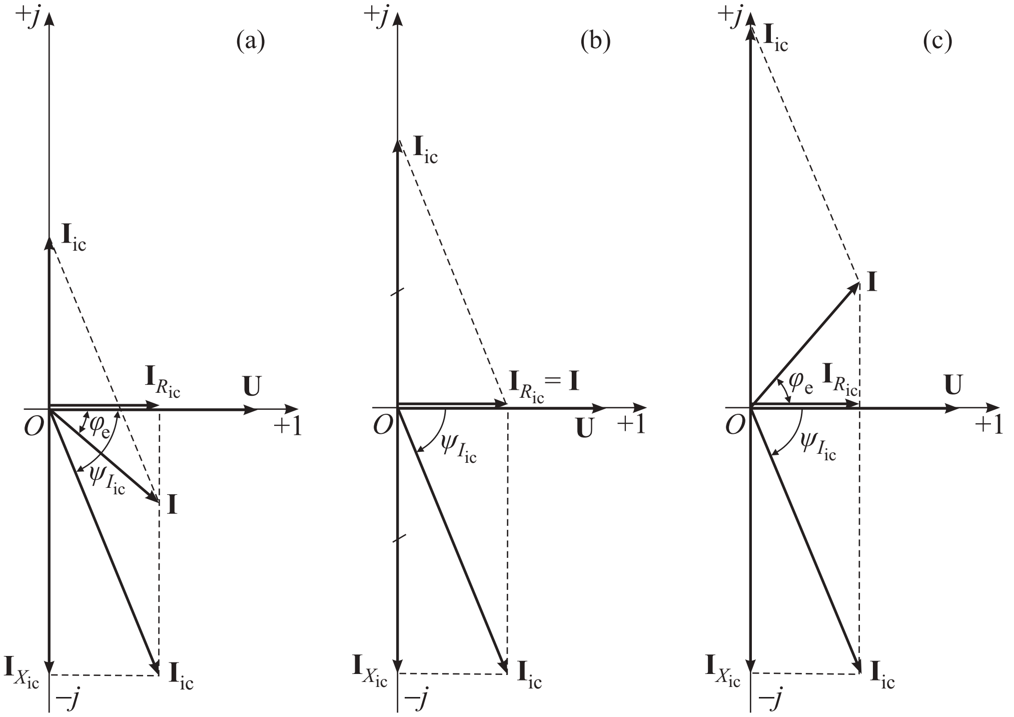
+j
(a)
$\mathbf{I}_{\rm ic}$
$\mathbf{I}_{R_{\rm ic}}$
$\mathbf{U}$
O
+1
$\varphi_{\rm e}$
$\psi_{I_{\rm ic}}$
$\mathbf{I}$
$\mathbf{I}_{X_{\rm ic}}$
$\mathbf{I}_{\rm ic}$
−j
+j
(b)
$\mathbf{I}_{\rm ic}$
$\mathbf{I}_{R_{\rm ic}} = \mathbf{I}$
O
$\mathbf{U}$
+1
$\psi_{I_{\rm ic}}$
$\mathbf{I}_{X_{\rm ic}}$
$\mathbf{I}_{\rm ic}$
−j
+j
$\mathbf{I}_{\rm ic}$
(c)
$\mathbf{I}$
$\varphi_{\rm e}$
$\mathbf{I}_{R_{\rm ic}}$
$\mathbf{U}$
O
+1
$\psi_{I_{\rm ic}}$
$\mathbf{I}_{X_{\rm ic}}$
$\mathbf{I}_{\rm ic}$
−j

Figure E6.55. Vector diagrams of currents: (a) before resonance ($I_{\rm ic} > I_C$, $\varphi_{\rm e} > 0$); (b) at the resonance instant ($I_{\rm ic} \cong I_C$, $\varphi_{\rm e} = 0$); (c) after resonance ($I_{\rm ic} < I_C$, $\varphi_{\rm e} < 0$).

In the analysis of the parametric current resonance, the resonance curves $I_C(C)$, $I_{\rm ic}(C)$, $I(C)$, $\varphi_{\rm e}(C)$, $\cos\varphi_{\rm e}(C)$ are constructed by the formulas

$$I_C(C) = B_C U = \omega C U,$$

$$I_{\rm ic}(C) = Y_{\rm ic} U = \sqrt{G_{\rm ic}^2 + B_{\rm ic}^2}\, U = U\sqrt{\left(\frac{R_{\rm ic}}{R_{\rm ic}^2 + (\omega L)^2}\right)^2 + \left(\frac{\omega L}{R_{\rm ic}^2 + (\omega L)^2}\right)^2} = \text{const},$$

$$I(C) = Y_{\rm e} U = \sqrt{G_{\rm ic}^2 + (B_{\rm ic} - B_C)^2}\, U = U\sqrt{\left(\frac{R_{\rm ic}}{R_{\rm ic}^2 + (\omega L)^2}\right)^2 + \left(\frac{\omega L}{R_{\rm ic}^2 + (\omega L)^2} - \omega C\right)^2},$$

$$\varphi_{\rm e}(C) = \arctan\frac{B_{\rm e}}{G_{\rm e}} = \arctan\left(\frac{\omega L - \omega C[R_{\rm ic}^2 + (\omega L)^2]}{R_{\rm ic}}\right),$$

$$\cos\varphi_{\rm e}(C) = \frac{G_{\rm e}}{Y_{\rm e}} = \frac{R_{\rm ic}}{\sqrt{R_{\rm ic}[R_{\rm ic}^2 + (\omega L)^2] + \{\omega L - \omega C[R_{\rm ic}^2 + (\omega L)^2]\}^2}}.$$

The current resonance, in contrast to the voltage resonance (which causes the overvoltage in electrical devices) is safe for electrical plants and, in particular, can be used to compensate for the reactive power in them. In current resonance, large currents arise in circuits only if the circuit branches have large susceptance; i.e., the circuits contain large capacitor batteries or powerful induction coil.

E6.5.10. Improvement of the Power Factor

The power factor ($\cos\varphi$) is one of the fundamental energy characteristics of electrical engineering devices.

An increase in $\cos\varphi$ decreases the losses in the electric energy transfer from source to receivers and increases the efficiency factor of devices.

As is known, the power factor can be determined from the power triangle by the formula

$$\cos\varphi = \frac{P}{S} = \frac{P}{\sqrt{P^2 + Q^2}} = \frac{P}{\sqrt{P^2 + (Q_L - Q_C)^2}},$$

where φ is the phase shift between the voltage U and current I vectors, S is the total power of the circuit, P is the active power of the circuit, Q_L is the inductive power, Q_C is the capacitive power, $Q = Q_L - Q_C = UI\sin\varphi = UI_X$ is the reactive power of the circuit, and $I_X = I\sin\varphi$ is the reactive component of the actual current I (this current is called the *reactive current*).

As follows from the last formula, to increase $\cos\varphi$, it is necessary to decrease the reactive power Q, but in alternating current motors (asynchronous motors), the torque is obtained by using the stator and conductors rotating field interaction with the rotor current. Hence such machines need an alternating magnetizing reactive current I_X (i.e., the reactive power $Q = UI_X$) to create the torque moment of the field, and this worsens the power factor $\cos\varphi$ in industrial plant devices. A low value of the power factor results in an incomplete use of power in generators, power transmission lines, and transformers. They are uselessly loaded by the reactive current I_X. The reactive current I_X increases the losses $\Delta P_{\rm lin}$ in wires as the electric energy is transferred. These losses consist of the losses arising due to the transfer of the active current I_R and the useless losses due to the transfer of the reactive current I_X. The latter are caused by the energy transmission from magnetic fields of motors to electrical stations and generators and by the reverse transmissions.

The use of the current resonance permits unloading the energy source and the transmitting devices from these useless oscillations of electric energy and hence from the reactive current I_X by closing the electric energy oscillations in the resonance circuit formed by capacitors of capacitance C and a induction coil of inductance L. Practically, this unloading is realized by parallel connection of motors with parameters ($R_{\rm p}$, $X_{\rm p}$) equivalent to those of a battery of capacitors of capacitance C (Fig. E6.56a).

To compensate for the phase shift completely, the reactive (capacitive) power Q_C of the batteries must be equal to the reactive (inductive) power of motors $Q_L = UI_{\rm p}\sin\varphi$.

In a majority of cases, the phase shift is compensated for incompletely, because the presence of a small reactive current I_X is of no importance for $\cos\varphi \ge 0.95$, because $I = \sqrt{I_R^2 + I_X^2}$, and the complete compensation requires additional installation of devices with significant conductance (additional batteries of capacitors), which is often economically unprofitable.

Usually, the value of $\cos\varphi$ that the electrical engineering device must have after compensation is prescribed; if the initial value of the current receiver $I_{\rm p}$ and its $\cos\varphi_{\rm p}$ are known (as a rule, these values are given in the certificates of electrical engineering devices), then the required value of the capacitance C of the battery of capacitors is determined as follows.

To decrease the phase shift $\varphi_{\rm p}$ to the value φ, it is necessary, as the vector diagram shows (Fig. E6.56b), to decrease the resultant reactive current of the device by the quantity $I_{X_{\rm p}} - I_X$; here $I_{X_{\rm p}}$ is the device reactive current before compensation, and I_X is the reactive current after compensation.

The active current I_R is related to the reactive current I_X simply as $I_X = I_R\tan\varphi$; moreover, the active current can be expressed in terms of the active power P and the voltage U of the device (receiver) as

$$I_R = \frac{P}{U} = \frac{UI\cos\varphi}{U} = I\cos\varphi.$$

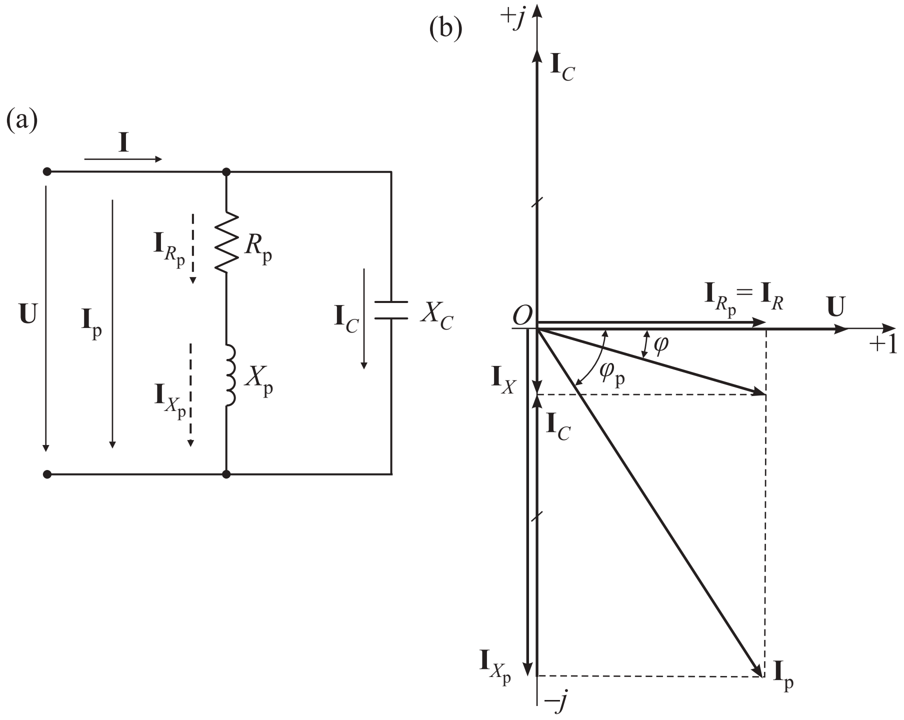

Figure E6.56. Phase shift compensation: (a) network and the vector diagram (b).

The active current does not vary in the course of compensation, $I_{R_\mathrm{p}} = I_R = \mathrm{const}$. After the required substitutions, one can express the desired difference of reactive currents as follows:

$$I_{X_\mathrm{p}} - I_X = I_R(\tan\varphi_\mathrm{p} - \tan\varphi) = \frac{P}{U}(\tan\varphi_\mathrm{p} - \tan\varphi).$$

The capacitive current I_C, necessary to perform the compensation must be numerically equal to this difference:

$$I_C = I_{X_\mathrm{p}} - I_X = \frac{P}{U}(\tan\varphi_\mathrm{p} - \tan\varphi). \qquad \text{(E6.5.10.1)}$$

By Ohm's law, the capacitive current is related to the capacitance as

$$I_C = \frac{U}{X_C} = \frac{U}{1/\omega C} = \omega CU. \qquad \text{(E6.5.10.2)}$$

From (E6.5.10.1) and (E6.5.10.2) we obtain the capacitance necessary to perform the compensation:

$$C = \frac{P}{\omega U^2}(\tan\varphi_\mathrm{p} - \tan\varphi).$$

The improvement of the power factor ($\cos\varphi$) by including capacitors (batteries of capacitors) is called the *artificial improvement of the power factor* in contrast to the *natural improvement of the power factor*, which is attained by complete use of the power of motors and by installing motors that do not consume reactive current (synchronous motors).

Bibliography for Chapter E6

Atabekov, G. I., *Theoretical Foundations of Electrical Engineering. Linear Electrical Networks* [in Russian], Lan, Moscow, 2009.

Bird, J., *Electrical Circuit Theory and Technology, 2nd Edition*, Newnes, 2003.

Bessonov, L. A., *Theoretical Foundations of Electrical Engineering. Electrical Networks* [in Russian], Vysshaya Shkola, Moscow, 2006.

Bobrow, L. S., *Fundamentals of Electrical Engineering*, Oxford University Press, Oxford, England, 1996.

Chen, W. (Editor), *The Electrical Engineering Handbook*, Academic Press, New York, 2004.

Demirchan, K. S., Neiman, L. R., and Korovkin, N. V., *Theoretical Foundations of Electrical Engineering* [in Russian], Piter, Saint-Petersburg, 2009.

Dorf, R. C. (Editor), *The Electrical Engineering Handbook, Third Edition – 6 Volume Set, 3rd Edition*, CRC Press, Boca Raton, Florida, 2006.

Dorf, R. C. and Tallarida, R. J., *Pocket Book of Electrical Engineering Formulas*, CRC Press, Boca Raton, Florida, 1993.

Hambley, A. R., *Electrical Engineering: Principles and Applications, 5rd Edition*, Prentice Hall, New Jersey, 2010.

Kasatkin, A. S., *Electrical Engineering* [in Russian], Vysshaya Shkola, Moscow, 2000.

Kothari, D. P. and Nagrath, I. J., *Basic Electrical Engineering, 3rd Edition*, McGraw-Hill, New York, 2009.

Kubala, T., *Electricity 1, Devices, Circuits & Materials, 9th Edition*, Delmar Cengage Learning, New York, 2008.

Kubala, T., *Electricity 2, Devices, Circuits & Materials, 9th Edition*, Delmar Cengage Learning, New York, 2008.

Mittle, V. N. and Mittal, A., *Basic Electrical Engineering, 2nd Edition*, McGraw-Hill, New York, 2006.

Nagsarkar, T. K. and Sukhija, M. S., *Basic Electrical Engineering*, Oxford University Press, Oxford, England, 2005.

Nilsson, J. W. and Riedel, S. A., *Electric Circuits, 8th Edition*, Prentice Hall, New Jersey, 2008.

Phadke, A. G., *Handbook of Electrical Engineering Calculations*, CRC Press, Boca Raton, Florida, 1999.

Rizzoni, G., *Fundamentals of Electrical Engineering*, McGraw-Hill, New York, 2008.

Stanley W. D., Hackworth, J. R., and Jones, R. L., *Fundamentals of Electrical Engineering and Technology*, Delmar Cengage Learning, New York, 2006.

Chapter E7

Empirical and Engineering Formulas and Criteria for Their Applicability

E7.1. Empirical, Engineering and Interpolation Formulas. Least-Squares Method

E7.1.1. Empirical and Engineering Formulas

► **Preliminaries. Requirements to empirical formulas.** In science and engineering applications, functional dependences are often studied experimentally. The measurement results form a finite set of points, which is usually written as in the form of a table

x_1	x_2	x_3	$\dots$	x_n
y_1	y_2	y_3	$\dots$	y_n

(E7.1.1.1)

Here x_1, x_2, x_3, $\dots$, x_n are the values of the argument, and y_1, y_2, y_3, $\dots$, y_n are the corresponding values of the function.

If the analytic dependence of y on x is unknown (or is known but very cumbersome), then the following practically important problem arises: find an *empirical formula*

$$y = f(x) \qquad \text{(E7.1.1.2)}$$

such that for all x_i ($i = 1, 2, \dots, n$) the values provided by this formula only slightly differ from the experimental data y_i,

$$|y_i - f(x_i)| < \varepsilon \qquad (i = 1, 2, \dots, n).$$

The empirical formula must satisfy the following two fundamental requirements:

(i) It must be sufficiently accurate.

(ii) It must be sufficiently simple. (Of two formulas with the same accuracy, the simpler one is preferable.)

If there is no information about intermediate data, then it is usually assumed that the empirical function is continuous, smooth, and monotone on any interval $[x_i, x_{i+1}]$ (unless $y_i = y_{i+1}$).

Remark. Any experimental data are approximate, because measurements always have certain errors. Therefore, the accuracy of any empirical formula can never be better than that of the experimental data used to construct it.

► **Choice of the form of an empirical formula.** The general form of the desired functional dependence (E7.1.1.2) is usually taken a priori from various considerations. At the initial stage, the basic functional dependence must contain a set of free parameters (indeterminate coefficients) whose values are further determined from the condition that the empirical formula is consistent with the experimental data. Linear, rational-fractional rational, power-law, exponential, and (in the case of periodic processes) trigonometric functions are used most frequently. In what follows, we consider the most widely used types of empirical formulas in Sections E7.1.2 and E7.2.

In many cases, if the argument ranges in a bounded interval, one can use a polynomial

$$y = \sum_{i=0}^{m} a_i x^i$$

to construct an empirical formula. The degree m of the polynomial and the indeterminate coefficients a_i can be found from various considerations, which can take into account a priori known boundary or initial conditions.

THEOREM (WEIERSTRASS). *Let* $f(x)$ *be a continuous function on a finite interval* $[a, b]$. *Then, for any* $\varepsilon > 0$, *there exists a polynomial* $P(x)$ *such that the inequality* $|f(x) - P(x)| < \varepsilon$ *holds at every point of* $[a, b]$.

Assume that there is a class of qualitatively similar problems or phenomena that differ from each other by geometric characteristics (for example, by the shape of the region under study, the flow structure, etc.) and by the numerical values of physical-chemical parameters. If the dependence (E7.1.1.2) of the desired quantity y on the argument x is known for some specific (for example, the simplest) problem (this dependence may be approximate and may be obtained theoretically or experimentally), then the empirical formula describing a similar quantity for another problem of the same class can be sought in the form

$$y = a_1 f(a_2 x + a_3) + a_4, \qquad \text{(E7.1.1.3)}$$

where the numerical values of the parameters a_i are determined, for example, by the least-squares method (see Section E7.1.2). Formula (E7.1.1.3) is obtained from (E7.1.1.2) by combining the translation and scaling operations in both variables x and y. In several cases, it suffices to use two-parameter dependences (E7.1.1.3) in which either $a_2 = 1$ and $a_3 = 0$ or $a_3 = a_4 = 0$.

In certain cases, the form of the empirical formula can be chosen on the basis of additional theoretical considerations concerning the character of the dependence under study. (Such formulas are often said to be *semiempirical*.)

► **Engineering formulas.** The notion of *engineering formula* is wider than that of *empirical formula*. Engineering formulas are approximate formulas used in applications and obtained by either theoretical (approximate, analytic, or numerical) or experimental methods. Various combinations of theoretical and experimental methods can be used as well.

Ideally, the engineering formula must be sufficiently accurate and sufficiently simple and have a wide scope of applicability.

E7.1.2. Least-Squares Method

► **General case.** The *least-squares method* allows approximately determining a function presented in table form (E7.1.1.1) by measurement results.

Consider a functional dependence of the form

$$y = F(x, a_1, a_2, \dots, a_m), \qquad \text{(E7.1.2.1)}$$

where $a_1, a_2, \dots, a_m$ are the desired parameters.

Let us write the sum

$$S = \sum_{i=1}^{n} \bigl[y_i - F(x_i, a_1, a_2, \dots, a_m)\bigr]^2 \qquad \text{(E7.1.2.2)}$$

of squared deviations, where x_i and y_i are the experimental data presented in the table of Eq. (E7.1.1.1).

The numerical values of the parameters a_i are determined by minimizing the sum (E7.1.2.2). Necessary conditions for the minimum of the expression (E7.1.2.2) lead to the following system of m algebraic (or transcendental) equations for the parameters a_i:

$$\frac{\partial S}{\partial a_1} = 0, \quad \frac{\partial S}{\partial a_2} = 0, \quad \dots, \quad \frac{\partial S}{\partial a_m} = 0. \qquad \text{(E7.1.2.3)}$$

These equations may be nonlinear in general.

If system (E7.1.2.3) has a unique solution, then this is the desired solution. If there are several solutions, then one should compare the corresponding values of the sum (E7.1.2.2) and choose the parameter values that provide the minimum of S.

Example 1. For the desired empirical dependence we take the linear function

$$y = ax + b.$$

To calculate the indeterminate coefficients a and b, one minimizes the sum

$$S = \sum_{i=1}^{n} (y_i - ax_i - b)^2.$$

By using the necessary conditions for the minimum of the function $S = S(a, b)$ (i.e., by equating the partial derivatives of S with respect to the parameters a and b with zero), we obtain the system of two linear equations

$$a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i,$$
$$a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i.$$

for a and b. Here the values $x_1, x_2, \dots, x_n$ and $y_1, y_2, \dots, y_n$ are assumed to be taken from table (E7.1.1.1).

Example 2. If for the desired dependence we take the quadratic function

$$y = ax^2 + bx + c,$$

then we obtain the system of three linear equations

$$a\sum_{i=1}^{n} x_i^4 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i^2 y_i,$$
$$a\sum_{i=1}^{n} x_i^3 + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i,$$
$$a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i + cn = \sum_{i=1}^{n} y_i.$$

for the coefficients a, b, and c.

► **Special case** (a generalization of Examples 1 and 2). System (E7.1.2.3) is simplified dramatically if the empirical function (E7.1.2.1) is taken to be linear in the parameters a_i,

$$y = \varphi_0(x) + a_1\varphi_1(x) + \dots + a_m\varphi_m(x). \qquad \text{(E7.1.2.4)}$$

Then system (E7.1.2.3) is the linear algebraic system that can be written as

$$\begin{aligned} a_1(\varphi_1,\varphi_1)+a_2(\varphi_1,\varphi_2)+\cdots+a_m(\varphi_1,\varphi_m)&=(\varphi_1,Y),\\ a_1(\varphi_2,\varphi_1)+a_2(\varphi_2,\varphi_2)+\cdots+a_m(\varphi_2,\varphi_m)&=(\varphi_2,Y),\\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots&\dots\dots\\ a_1(\varphi_m,\varphi_1)+a_2(\varphi_m,\varphi_2)+\cdots+a_m(\varphi_m,\varphi_m)&=(\varphi_m,Y), \end{aligned}$$

where

$$(\varphi_j,\varphi_k)=\sum_{i=1}^n\varphi_j(x_i)\varphi_k(x_i),\quad (\varphi_j,Y)=\sum_{i=1}^n\varphi_j(x_i)Y_i,\quad Y_i=y_i-\varphi_0(x_i).$$

E7.1.3. Refinement of Empirical Formulas

► **Statement of the problem.** Assume that the empirical formula (E7.1.1.2) does not agree well with the experimental data (E7.1.1.1). It is required to modify this formula so as to satisfy the experimental data.

► **The simplest method for refining empirical formulas.** Here we describe the simplest method for refining an empirical formula if it the sum of squared deviations provided by this formula is larger than desired.

Consider the auxiliary function

$$y=f(x)+a, \qquad \text{(E7.1.3.1)}$$

where a is an unknown constant. We determine a from the condition that the sum

$$S=\sum_{i=1}^n[y_i-f(x_i)-a]^2 \qquad \text{(E7.1.3.2)}$$

of squared deviations be minimal. A necessary condition for the minimum of the sum (E7.1.3.2) can be found from the condition $S'_a=0$ and gives

$$a=\frac{1}{n}\sum_{i=1}^n\varepsilon_i,\qquad \varepsilon_i=y_i-f(x_i), \qquad \text{(E7.1.3.3)}$$

which ensures the minimum of the sum under study. Here ε_i is the deviation of the experimental data from the results provided by the empirical formula (E7.1.1.2) at the point x_i.

Thus, if $a\neq 0$, then the addition of the constant a determined by formula (E7.1.3.3) to the right-hand side of the empirical formula (E7.1.1.2) results in a refinement of this formula in the sense that the sum of squared deviations decreases. If $a=0$ or $a\approx 0$, then this method does not give the desired effect. In this case, one can use the more general method described below.

► **A more general method for refining empirical formulas.** Instead of (E7.1.3.1), we take a more general auxiliary function

$$y=f(x)+a\varphi(x), \qquad \text{(E7.1.3.4)}$$

where $\varphi(x)$ is a given function and a is an unknown constant.

Consider the sum

$$S=\sum_{i=1}^{n}[y_i-f(x_i)-a\varphi(x_i)]^2. \qquad \text{(E7.1.3.5)}$$

A necessary condition for the minimum of this function implies the equation

$$\sum_{i=1}^{n}\varphi(x_i)[y_i-f(x_i)-a\varphi(x_i)]=0,$$

which permits finding the desired constant a,

$$a=\frac{\displaystyle\sum_{i=1}^{n}\varepsilon_i\varphi(x_i)}{\displaystyle\sum_{i=1}^{n}\varphi^2(x_i)}, \qquad \varepsilon_i=y_i-f(x_i). \qquad \text{(E7.1.3.6)}$$

Remark 1. Instead of formula (E7.1.3.4), one can refine the empirical formula (E7.1.1.2) by using the more complicated two-parameter dependence

$$y=f(x)+a_1\varphi_1(x)+a_2\varphi_2(x).$$

Remark 2. To refine the empirical formula (E7.1.1.2), one can use the four-parameter dependence (E7.1.1.3).

E7.1.4. Interpolation Formulas

► **Linear interpolation. Lagrange interpolation polynomial.** Interpolation formulas are formulas that give the exact values y_m of a function on a given discrete set of values x_m the argument.

Linear interpolation formulas

$$y=\sum_{i=1}^{n}a_i\varphi_i(x), \qquad \text{(E7.1.4.1)}$$

where $\varphi_1(x)$, …, $\varphi_n(x)$ is a given system of functions, are used most frequently. The coefficients a_i are determined from the condition that the function y coincides with the given values y_m at the respective points x_m,

$$y_m=\sum_{i=1}^{n}a_i\varphi_i(x_m), \qquad m=1,\,2,\,\dots,\,n.$$

For the sequence $\varphi_m(x)=x^{m-1}$ of polynomials, we obtain the function

$$y=L_{n-1}(x)=\sum_{i=1}^{n}y_i\prod_{j\neq i}\frac{x-x_j}{x_i-x_j}, \qquad \text{(E7.1.4.2)}$$

which is called the *Lagrange interpolation polynomial*. The $(n-1)$st-order polynomial (E7.1.4.2) exactly satisfied the data of table (E7.1.1.1).

Let a function $f(x)$ have an nth-order derivative on a finite interval $[a, b]$. Then the error in replacing $f(x)$ with the Lagrange polynomial $L_{n-1}(x)$ such that $f(x_m) = L_{n-1}(x_m)$ $(m = 1, \dots, n)$ is estimated as

$$\begin{aligned}|f(x) - L_{n-1}(x)| &\le \frac{1}{n!} \max_{a\le x\le b} |f^{(n)}(x)|\, |(x - x_1)(x - x_2) \dots (x - x_n)| \\ &\le \frac{1}{n!} \max_{a\le x\le b} |f^{(n)}(x)| \left(\frac{b-a}{n}\right)^n.\end{aligned}$$

► **Two-point linear interpolation.** In applications, the linear interpolation formula with two points x_k and x_{k+1} is used most frequently. This formula reads

$$y = \frac{y_{k+1} - y_k}{x_{k+1} - x_k}(x - x_k) + y_k \qquad (x_k \le x \le x_{k+1}) \tag{E7.1.4.3}$$

and corresponds to the case of $n = 2$ in formula (E7.1.4.2) with x_1 and x_2 redenoted by x_k and x_{k+1}, respectively. Formula (E7.1.4.3) describes the straight line segment passing with endpoints (x_k, y_k) and (x_{k+1}, y_{k+1}) on the (x, y)-plane.

Remark. The interpolation formulas (E7.1.4.2) with $n > 2$ are rather cumbersome and are not used practically to construct empirical formulas on the basis of experimental data (because it does not make any sense to take a function exactly coinciding with the measurement results, which always contain certain errors).

E7.2. Criteria for Applicability of Empirical Formulas

E7.2.1. Direct Method for Verifying Two-Parameter Formulas

► **Method of complete elimination of free parameters**. For simplicity and clarity, we illustrate the method of complete elimination of free parameters by an example of two-parameter functions explicitly defined as

$$y = F(x, a, b). \tag{E7.2.1.1}$$

It is required to examine whether formula (E7.2.1.1) can be used to approximate given experimental data by an appropriate choice of the constants a and b.

By specifying the values of the function at three arbitrary (but distinct) points x_i, x_j, and x_k, we obtain the three relations

$$y_i = F(x_i, a, b), \quad y_j = F(x_j, a, b), \quad y_k = F(x_k, a, b). \tag{E7.2.1.2}$$

From now on, we use the short notation $y_i = y(x_i)$, $y_j = y(x_j)$, and $y_k = y(x_k)$. By eliminating the free parameters a and b from (E7.2.1.2), we obtain the criterion relation. It is convenient to write it out as

$$\Theta = 1, \tag{E7.2.1.3}$$

where the left-hand side contains a function

$$\Theta = \Theta(x_i, x_j, x_k, y_i, y_j, y_k). \tag{E7.2.1.4}$$

It is necessary to satisfy relation (E7.2.1.3) for any values of i, j, k $(1 \le i < j < k \le n$, where n is the total number of experimental points) for the dependence (E7.2.1.1) to hold. For the applicability of formula (E7.2.1.1) as an empirical formula, it is necessary to satisfy relation (E7.2.1.3) with prescribed accuracy.

In practice, it is convenient to use the special case of formula (E7.2.1.4) with $i=j-1$ and $k=j+1$. In this case, it is useful to represent the results of experimental data processing obtained by calculating the function (E7.2.1.4) by points (x_j,Θ_j) on the (x,Θ)-plane and estimate their deviations from the straight line $\Theta=1$ (see the criterion (E7.2.1.3)), which is parallel to the abscissa axis.

We use the above-described method in the case of two-parameter empirical dependences of special form

$$y=af(x)+b, \tag{E7.2.1.5}$$

where $f(x)$ is a given function and a and b are free parameters. We take three distinct points x_i, x_j, and x_k to which there correspond three values y_i, y_j, and y_k of the desired quantity. Under the assumption that the functional dependence (E7.2.1.5) takes place, we have

$$y_i=af(x_i)+b,\quad y_j=af(x_j)+b,\quad y_k=af(x_k)+b. \tag{E7.2.1.6}$$

By eliminating the parameters a and b from these relations, we obtain

$$\frac{y_k-y_i}{y_j-y_i}=\frac{f(x_k)-f(x_i)}{f(x_j)-f(x_i)}. \tag{E7.2.1.7}$$

By introducing the new auxiliary variable

$$\Theta=\frac{y_k-y_i}{y_j-y_i}\,\frac{f(x_j)-f(x_i)}{f(x_k)-f(x_i)}, \tag{E7.2.1.8}$$

we can verify the applicability of the two-parameter approximations of the form (E7.2.1.5), because condition (E7.2.1.3) must be satisfied for them, which follows from relation (E7.2.1.7).

Formula (E7.2.1.8) readily admits computer implementation, and by choosing the function $f(x)$ in (E7.2.1.5) in various ways, one can compare the corresponding approximation errors and then take the dependence with the least error.

Example 1. For the logarithmic dependence

$$y=a\ln x+b\equiv a\ln(cx),\qquad c=\exp(b/a), \tag{E7.2.1.9}$$

formula (E7.2.1.8) becomes

$$\Theta=\frac{y_k-y_i}{y_j-y_i}\,\frac{\ln x_j-\ln x_i}{\ln x_k-\ln x_i}. \tag{E7.2.1.10}$$

For the applicability of two-parameter approximations of the form (E7.2.1.9), relation (E7.2.1.3) must be satisfied under condition (E7.2.1.10).

Example 2. Consider the two-parameter power-law dependence

$$y=ax^m\qquad (x\ge 0,\ y\ge 0). \tag{E7.2.1.11}$$

Taking the logarithms of both sides of this formula, we obtain a dependence of the form (E7.2.1.5) with $f(x)=\ln x$, $b=\ln a$, and $a=m$ and with y replaced by $\ln y$ on the left-hand side. By substituting $f(x)=\ln x$ into (E7.2.1.8) and by replacing y by $\ln y$, we obtain

$$\Theta=\frac{\ln y_k-\ln y_i}{\ln y_j-\ln y_i}\,\frac{\ln x_j-\ln x_i}{\ln x_k-\ln x_i}. \tag{E7.2.1.12}$$

Formula (E7.2.1.12) permits verifying the applicability of power-law functions of the form (E7.2.1.11).

In Table E7.1, we present direct criteria (necessary conditions) for the applicability of two-parameter monotone dependences of the simplest form obtained by the method of complete elimination of free parameters.

TABLE E7.1
Some two-parameter dependences and direct criteria of their applicability of the form (E7.2.1.3) obtained by the method of complete elimination of free parameters.

No.	Form of empirical formula	Form of criterion function Θ	Remarks
1	$y = ax + b$	$\dfrac{(x_j - x_i)(y_k - y_i)}{(x_k - x_i)(y_j - y_i)}$	
2	$y = \dfrac{a}{x} + b$	$\dfrac{x_j(x_k - x_i)(y_j - y_i)}{x_k(x_j - x_i)(y_k - y_i)}$	
3	$y = \dfrac{1}{ax + b}$	$\dfrac{(x_j - x_i)y_j(y_k - y_i)}{(x_k - x_i)y_k(y_j - y_i)}$	
4	$y = \dfrac{x}{ax + b}$	$\dfrac{x_j(x_k - x_i)y_k(y_j - y_i)}{x_k(x_j - x_i)y_j(y_k - y_i)}$	
5	$y = ax^m$	$\dfrac{\ln(x_k/x_i)\ln(y_j/y_i)}{\ln(x_j/x_i)\ln(y_k/y_i)}$	$x > 0$
6	$y = ae^{\lambda x}$	$\dfrac{(x_k - x_i)\ln(y_j/y_i)}{(x_j - x_i)\ln(y_k/y_i)}$	
7	$y = a\ln x + b$	$\dfrac{(y_k - y_i)\ln(x_j/x_i)}{(y_j - y_i)\ln(x_k/x_i)}$	$x > 0$
8	$y = af(x) + b$	$\dfrac{y_k - y_i}{y_j - y_i}\,\dfrac{f(x_j) - f(x_i)}{f(x_k) - f(x_i)}$	$f(x)$ is given
9	$y = a[f(x)]^m$	$\dfrac{\ln(y_k/y_i)}{\ln(y_j/y_i)}\,\dfrac{\ln[f(x_j)/f(x_i)]}{\ln[f(x_k)/f(x_i)]}$	$f(x) \geq 0$ is given
10	$y = a\exp[\lambda f(x)]$	$\dfrac{\ln(y_k/y_i)}{\ln(y_j/y_i)}\,\dfrac{f(x_j) - f(x_i)}{f(x_k) - f(x_i)}$	$f(x)$ is given

Remark 1. For a majority of two-parameter functions (E7.2.1.1), it is impossible to eliminate free parameters from relations (E7.2.1.2); i.e., it is impossible to obtain the criterion relation (E7.2.1.3) explicitly. This fact does not permit using this method efficiently for more general multi-parameter dependences.

Remark 2. The direct method of elimination of free parameters can be used in the case of multi-parameter dependences of the special form

$$y = h(x)F\left(\frac{\sum_{i=1}^{N} a_i f_i(x)}{\sum_{j=1}^{M} b_j g_j(x)}\right), \qquad \text{(E7.2.1.13)}$$

where $f_i(x)$, $g_j(x)$, $h(x)$, and $F(z)$ are given functions and a_i and b_j are free parameters. To this end, we first must rewrite (E7.2.1.13) in the equivalent form

$$F^{-1}\left(\frac{y}{h(x)}\right)\sum_{j=1}^{M} b_j g_j(x) = \sum_{i=1}^{N} a_i f_i(x),$$

where F^{-1} is the inverse function of F. Further, taking the required number of points x_k ($k = 1, 2, \ldots, N+M+1$) and the values y_k corresponding to them, one can eliminate the parameters a_i and b_j from the resulting system of linear algebraic equations.

► **Method of partial elimination of free parameters**. We describe the method of partial elimination of free parameters for two-parameter functions $y = F(x, a, b)$. Prescribing the values of the function at two arbitrary points, we obtain two relations, $y_i = F(x_i, a, b)$ and

$y_j = F(x_j, a, b)$. By eliminating one of the parameters from them and then by solving the obtained relation for the other parameter, we obtain a criterion relation of the form $\varphi(x_i, x_j, y_i, y_j) = a$ or $\psi(x_i, x_j, y_i, y_j) = b$.

Let us apply this method to the two-parameter family of functions (E7.2.1.5). By eliminating the parameter b from the first two relations in (E7.2.1.6), we obtain the relation

$$\frac{y_j - y_i}{f(x_j) - f(x_i)} = a, \tag{E7.2.1.14}$$

which can also be used to verify the applicability of functions of the form (E7.2.1.5). If any two points x_i and x_j and the values y_i and y_j corresponding to them are substituted into the left-hand side of (E7.2.1.14), then one and the same constant a must be obtained. (If Eq. (E7.2.1.14) is satisfied approximately with satisfactory accuracy, then the constant a can readily be verified by the least-squares method.) Another criterion relation for the family of functions (E7.2.1.6) can be derived in a similar way by eliminating the parameter a:

$$\frac{y_i f(x_j) - y_j f(x_i)}{f(x_j) - f(x_i)} = b. \tag{E7.2.1.15}$$

In the special case of the logarithmic law (E7.2.1.9), the first criterion relation (E7.2.1.14) acquires the form

$$\frac{y_j - y_i}{\ln x_j - \ln x_i} = a, \tag{E7.2.1.16}$$

and the second criterion relation (E7.2.1.15) becomes

$$\frac{y_i \ln x_j - y_j \ln x_i}{\ln x_j - \ln x_i} = b. \tag{E7.2.1.17}$$

Now we consider the two-parameter power-law dependence (E7.2.1.11). We take logarithms of both sides of this formula and then apply the resulting equation to two points x_i and x_j. After the elimination of the parameter a, we obtain the first criterion relation

$$\frac{\ln y_j - \ln y_i}{\ln x_j - \ln x_i} = m. \tag{E7.2.1.18}$$

For the power-law dependence (E7.2.1.11), we can also obtain the second criterion relation

$$\frac{\ln x_i \ln y_j - \ln x_j \ln y_i}{\ln x_i - \ln x_j} = \ln a \tag{E7.2.1.19}$$

by eliminating the parameter m.

In Table E7.2, we present the direct criteria (necessary conditions) for the applicability of two-parameter monotone dependences of the simplest form obtained by the method of partial elimination of free parameters (usually, one sets $j = i + 1$).

The method of partial elimination of free parameters leads to simpler criteria than the method of complete elimination of free parameters; in addition, it permits obtaining one of the parameters readily (cf. Table E7.1 and Table E7.2).

TABLE E7.2
Several two-parameter dependences and direct criteria of their applicability obtained by the method of partial elimination of free parameters; f^{-1} is the inverse function of f.

No.	Form of empirical formula	Applicability criterion	Remarks
1	$y=ax+b$	$\dfrac{y_i-y_j}{x_i-x_j}=a$	
2	$y=\dfrac{a}{x}+b$	$\dfrac{x_ix_j(y_j-y_i)}{x_i-x_j}=a$	
3	$y=\dfrac{1}{ax+b}$	$\dfrac{y_j-y_i}{(x_i-x_j)y_iy_j}=a$	
4	$y=\dfrac{x}{ax+b}$	$\dfrac{x_ix_j(y_i-y_j)}{(x_i-x_j)y_iy_j}=b$	
5	$y=ax^m$	$\dfrac{\ln(y_i/y_j)}{\ln x_i-\ln x_j}=m$	$x>0$
6	$y=ae^{\lambda x}$	$\dfrac{\ln(y_i/y_j)}{x_i-x_j}=\lambda$	
7	$y=a\ln x+b$	$\dfrac{y_i-y_j}{\ln x_i-\ln x_j}=a$	$x>0$
8	$y=af(x)+b$	$\dfrac{y_i-y_j}{f(x_i)-f(x_j)}=a$	$f(x)$ is given
9	$y=a[f(x)]^m$	$\dfrac{\ln(y_i/y_j)}{\ln[f(x_i)/f(x_j)]}=m$	$f(x)\ge 0$ is given
10	$y=a\exp[\lambda f(x)]$	$\dfrac{\ln(y_i/y_j)}{f(x_i)-f(x_j)}=\lambda$	$f(x)$ is given
11	$y=f(ax+b)$	$\dfrac{f^{-1}(y_i)-f^{-1}(y_j)}{x_i-x_j}=a$	$f(z)$ is given
12	$y=af(x)+bg(x)$	$\dfrac{y_ig(x_j)-y_jg(x_i)}{f(x_i)g(x_j)-f(x_j)g(x_i)}=a$	$f(x), g(x)$ is given

► **Several generalizations.** The method of partial elimination of free parameters can be used in the case of multi-parameter dependences of special form (E7.2.1.13). For example, for three-parameter dependence in the form of a quadratic polynomial

$$y=ax^2+bx+c,$$

we obtain the criterion

$$\frac{(x_k-x_i)(y_j-y_i)-(x_j-x_i)(y_k-y_i)}{(x_k-x_i)(x_j^2-x_i^2)-(x_j-x_i)(x_k^2-x_i^2)}=a.$$

For a more general three-parameter dependence

$$y=af(x)+bg(x)+ch(x),$$

where $f(x)$, $g(x)$, and $h(x)$ are given functions, we have the criterion

$$\frac{(y_jh_i-y_ih_j)(g_kh_i-g_ih_k)-(y_kh_i-y_ih_k)(g_jh_i-g_ih_j)}{(f_jh_i-f_ih_j)(g_kh_i-g_ih_k)-(f_kh_i-f_ih_k)(g_jh_i-g_ih_j)}=a.$$

Here we use the shorthand notation $f_p = f(x_p)$, $g_p = g(x_p)$, and $h_p = h(x_p)$, where $p = i,\ j,\ k$. The other two criteria can be obtained by cyclic permutations of the constants a, b, and c and the functions f, g, and h.

Remark. Instead of the explicit dependence (E7.2.1.1), one can consider a more general dependence $F(x, y, a, b) = 0$ and, in particular, an implicit dependence of the special form

$$g(x, y) = af(x, y) + b,$$

which generalizes formula (E7.2.1.5). In this case, instead of (E7.2.1.14) we have the criterion

$$\frac{g(x_j, y_j) - g(x_i, y_i)}{f(x_j, y_j) - f(x_i, y_i)} = a.$$

E7.2.2. Three-Parameter Empirical Formulas

► **Three-parameter formula of power-law form.** Consider the power-law dependence

$$y = ax^m + b \qquad \text{(E7.2.2.1)}$$

containing three parameters a, b, and m. We compose the geometric mean for the two extreme points: $x_{\text{gm}} = \sqrt{x_1 x_n}$. Let $x_k \le x_{\text{gm}} \le x_{k+1}$. We use the linear interpolation formula (E7.1.4.3) to obtain the approximate value y_{gm}. Assuming that the points (x_1, y_1), $(x_{\text{gm}}, y_{\text{gm}})$, (x_n, y_n) lie on the curve (E7.2.2.1), we obtain three equations

$$y_1 = ax_1^m + b, \quad y_{\text{gm}} = ax_{\text{gm}}^m + b, \quad y_n = ax_n^m + b.$$

By raising $x_{\text{gm}} = \sqrt{x_1 x_n}$ to the power m and by multiplying by a, we obtain

$$ax_{\text{gm}}^m = \sqrt{ax_1^m ax_n^m} \quad \Longrightarrow \quad y_{\text{gm}} - b = \sqrt{(y_1 - b)(y_n - b)}.$$

By squaring the last relation, we obtain the constant b,

$$b = \frac{y_1 y_n - y_{\text{gm}}^2}{y_1 + y_n - 2y_{\text{gm}}}. \qquad \text{(E7.2.2.2)}$$

Since the coefficient b is now known, the change $Y = y - b$ takes formula (E7.2.2.1) to a formula of the form (E7.2.1.11) with two parameters. To verify its applicability, we can use the criterion relation (E7.2.1.12) or (E7.2.1.18), where y_s must be replaced by $y_s - b$, $s = i, j, k$.

► **Three-parameter formula of exponential form.** Consider the exponential dependence

$$y = ae^{\lambda x} + b \qquad \text{(E7.2.2.3)}$$

containing three parameters a, b, and λ. We compose the arithmetic mean for the two extreme points, $x_{\text{am}} = \frac{1}{2}(x_1 + x_n)$. Let $x_k \le x_{\text{am}} \le x_{k+1}$. We use the linear interpolation formula (E7.1.4.3) to find the approximate value y_{am}. Assuming that the three points (x_1, y_1), $(x_{\text{am}}, y_{\text{am}})$, (x_n, y_n) lie on the curve (E7.2.2.3), we obtain

$$y_1 = ae^{\lambda x_1} + b, \quad y_{\text{am}} = ae^{\frac{1}{2}\lambda(x_1 + x_n)} + b, \quad y_n = ae^{\lambda x_n} + b.$$

This implies that $y_1 - b = ae^{\lambda x_1}$ and $y_n - b = ae^{\lambda x_n}$, and hence $(y_1 - b)(y_n - b) = (y_{\text{am}} - b)^2$. Now we determine the constant b:

$$b = \frac{y_1 y_n - y_{\text{am}}^2}{y_1 + y_n - 2y_{\text{am}}}. \qquad \text{(E7.2.2.4)}$$

Since the coefficient b is now known, the change $Y = y - b$ takes formula (E7.2.2.3) to the formula $Y = ae^{\lambda x}$ with two parameters. To verify its applicability, we can use the criterion dependence given in Tables E7.1 and E7.2 (the 6th row in both cases), where y_s must be replaced by $y_s - b$, $s = i, j, k$.

Remark 1. An exact analytic criterion for the applicability of formula (E7.2.2.3) in the case of equally spaced argument, i.e., for

$$\Delta x_i = x_{i+1} - x_i = h \qquad (i = 1,\ 2,\ \dots,\ n-1),$$

has the form

$$\lambda = \frac{1}{h}\ln\left(\frac{y_{i+2} - y_{i+1}}{y_{i+1} - y_i}\right) = \text{const.}$$

Remark 2. In Subsection E7.2.1, we presented criteria for verifying the applicability of a quadratic polynomial with three parameters. In Subsection E7.2.3, we describe more complicated criteria for verifying three-parameter formulas.

E7.2.3. Combined Method for Verifying Empirical Formulas

► **Description of the method. Three-parameter dependences of power-law form.** In the cases where there are sufficiently many experimental points, empirical formulas can be verified by a combined method. This method is based on the differentiation of the dependence under study, after which, using the original dependence and its derivative calculated at several points, the required number of free parameters is eliminated.

We illustrate the above by an example of the three-parameter power-law dependence

$$y = a(x+b)^m. \tag{E7.2.3.1}$$

By successively setting $x = x_i$, $x = x_j$, $x = x_k$ in (E7.2.3.1), we cannot eliminate the free parameters from the obtained relations; i.e., the direct method cannot be used.

Therefore, we proceed as follows. First, we differentiate (E7.2.3.1) with respect to x,

$$y' = am(x+b)^{m-1}, \tag{E7.2.3.2}$$

and then eliminate the parameter a from (E7.2.3.1) and (E7.2.3.2). We have

$$\frac{y}{y'} = \frac{x+b}{m}. \tag{E7.2.3.3}$$

Relation (E7.2.3.3) contains two parameters m and b and already admits the application of the direct method. By setting $x = x_i$ and $x = x_j$ in (E7.2.3.3) and then by eliminating one of the free parameters, we obtain two criterion relations containing the first derivatives:

$$\frac{(x_i - x_j)y'_i y'_j}{y_i y'_j - y_j y'_i} = m, \tag{E7.2.3.4}$$

$$\frac{x_i y_j y'_i - x_j y_i y'_j}{y_i y'_j - y_j y'_i} = b. \tag{E7.2.3.5}$$

In a similar way, consider the three-parameter dependence

$$y = ax^m + b, \tag{E7.2.3.6}$$

for which it is also impossible to derive criterion relations by the direct method. By differentiating (E7.2.3.6), we obtain

$$y' = amx^{m-1}. \tag{E7.2.3.7}$$

By eliminating the parameter a from (E7.2.3.6) and (E7.2.3.7), we obtain

$$\frac{y-b}{y'} = \frac{x}{m}. \tag{E7.2.3.8}$$

By setting $x = x_i$ and $x = x_j$ in (E7.2.3.8) and then by eliminating the parameter b, we obtain the criterion relation

$$\frac{x_i y'_i - x_j y'_j}{y_i - y_j} = m \tag{E7.2.3.9}$$

containing the first derivatives.

► **Dependences of exponential and logarithmic type.** We present several other simple useful criteria for widely used three-parameter functions.

For the exponential dependence

$$y = ae^{\lambda x} + b,$$

we have the criterion

$$\frac{y'_i - y'_j}{y_i - y_j} = \lambda. \tag{E7.2.3.10}$$

For the logarithmic dependence

$$y = a\ln(bx + c) \equiv a\ln(x + D) + E \qquad (D = c/b,\ E = a\ln b),$$

we have the criterion

$$\frac{y_j - y_i}{\ln(y'_i/y'_j)} = a. \tag{E7.2.3.11}$$

► **Formula for calculation of the derivatives.** Criteria (E7.2.3.4), (E7.2.3.5), (E7.2.3.9), (E7.2.3.10), and (E7.2.3.11) contain derivatives, which can be calculated approximately by using appropriate approximation formulas.

In what follows, we present three difference schemes most widely used in numerical mathematics to calculate the derivatives at the point x_k:

$$y'_k = \frac{y_{k+1} - y_{k-1}}{x_{k+1} - x_{k-1}} \qquad \text{(central difference)},$$

$$y'_k = \frac{y_{k+1} - y_k}{x_{k+1} - x_k} \qquad \text{(right difference)},$$

$$y'_k = \frac{y_k - y_{k-1}}{x_k - x_{k-1}} \qquad \text{(left difference)}.$$

Numerical experiments show that the first scheme based on the central difference provides the most accurate results. It is expedient to use just this scheme for computations.

E7.3. Construction of Engineering Formulas by the Method of Asymptotic Analogies

E7.3.1. Description of the Method of Asymptotic Analogies

► **Preliminary remarks.** The most important stage in the study of specific physical and engineering problems is to find general quantitative laws valid for a class of qualitatively similar problems. In many cases, general results of this type can be obtained by the method of asymptotic analogies. The method is based on the passage from the usual dimensionless variables to special asymptotic coordinates and can be used for constructing wide-scope approximate formulas. (For example, in problems of mass and heat transfer one and the same formula can be used for describing a variety of qualitatively similar problems that differ in surface shape and flow structure.)

Suppose that there is a class of problems that differ in geometric characteristics and depend on a dimensionless parameter x $(0 \le x < \infty)$. We assume that the dependence of the basic desired variable w on the parameter x is known for some specific (say, the simplest) geometry:

$$y = f(x), \tag{E7.3.1.1}$$

where $f(x)$ is a monotone function.

In problems of mass and heat transfer, y is usually the Sherwood (Nusselt) number or the volume-average concentration; the parameter x is dimensionless time, the Peclet number, or the dimensionless rate constant of a chemical reaction.

► **Transition to asymptotic coordinates.** Let us transform (E7.3.1.1) as follows. Let the leading terms of the asymptotic expansions of y for small and large x have the form

$$y \to y_0 \quad (\text{as } x \to 0), \qquad y \to y_\infty \quad (\text{as } x \to \infty), \tag{E7.3.1.2}$$

where y_0 and y_∞ depend in certain way on x,

$$y_0 = \varphi(x), \quad y_\infty = \psi(x), \tag{E7.3.1.3}$$

and are determined from the analysis of formula (E7.3.1.1). We also assume that the condition $\varphi/\psi \neq \text{const}$ is satisfied.

Note that the original dependence (E7.3.1.1), as well as the asymptotics (E7.3.1.2) and (E7.3.1.3), can be determined either theoretically or experimentally.

Using (E7.3.1.1)–(E7.3.1.3), we write the following two relations:

$$\frac{y}{y_0} = \frac{f(x)}{\varphi(x)}, \qquad \frac{y_\infty}{y_0} = \frac{\psi(x)}{\varphi(x)}. \tag{E7.3.1.4}$$

By expressing x from second relation and by substituting it in the first equation (E7.3.1.4), we find the explicit form of y in terms of the asymptotics y_0 and y_∞. As follows from (E7.3.1.4), the structure of this relation is generally written as

$$\frac{y}{y_0} = F\left(\frac{y_\infty}{y_0}\right). \tag{E7.3.1.5}$$

The function f contained in the original formula (E7.3.1.4) and the function F contained in formula (E7.3.1.5) are related as

$$f(x) \equiv \varphi(x) F\left(\frac{\psi(x)}{\varphi(x)}\right).$$

It is seen that, in contrast to formula (E7.3.1.1), formula (E7.3.1.5) remains invariant with respect to the choice of the method for determining the dimensionless value y. We suggest that the range of each of the ratios y/y_0 and y_∞/y_0 be identical for all problems of the class under consideration. The variables y/y_0 and y_∞/y_0 are called *asymptotic coordinates*.

Example. In applications, one usually encounters power-law asymptotics,

$$y_0 = Ax^k \quad \Bigl(\lim_{x\to 0} y/y_0 = 1\Bigr), \qquad y_\infty = Bx^m \quad \Bigl(\lim_{x\to\infty} y/y_\infty = 1\Bigr), \tag{E7.3.1.6}$$

where A, B, k, and m are some constants and $k \neq m$ (for the entire class of considered problems, we assume that the constants k and m remain the same but the parameters A and B can vary).

In the case of power-law asymptotics (E7.3.1.6), formula (E7.3.1.5) can be expressed explicitly via the function f:

$$\frac{y}{y_0} = \frac{1}{A}\left(\frac{A}{B}\frac{y_\infty}{y_0}\right)^{\frac{k}{k-m}} f\left(\left(\frac{A}{B}\frac{y_\infty}{y_0}\right)^{\frac{1}{m-k}}\right), \tag{E7.3.1.7}$$

or, equivalently,

$$\frac{y}{y_\infty} = \frac{1}{B}\left(\frac{A}{B}\frac{y_\infty}{y_0}\right)^{\frac{m}{k-m}} f\left(\left(\frac{A}{B}\frac{y_\infty}{y_0}\right)^{\frac{1}{m-k}}\right). \tag{E7.3.1.8}$$

► **Description of the method.** The basic idea of the method of asymptotic analogies is to use the expression (E7.3.1.5) (or formulas (E7.3.1.7) and (E7.3.1.8)) to approximate similar

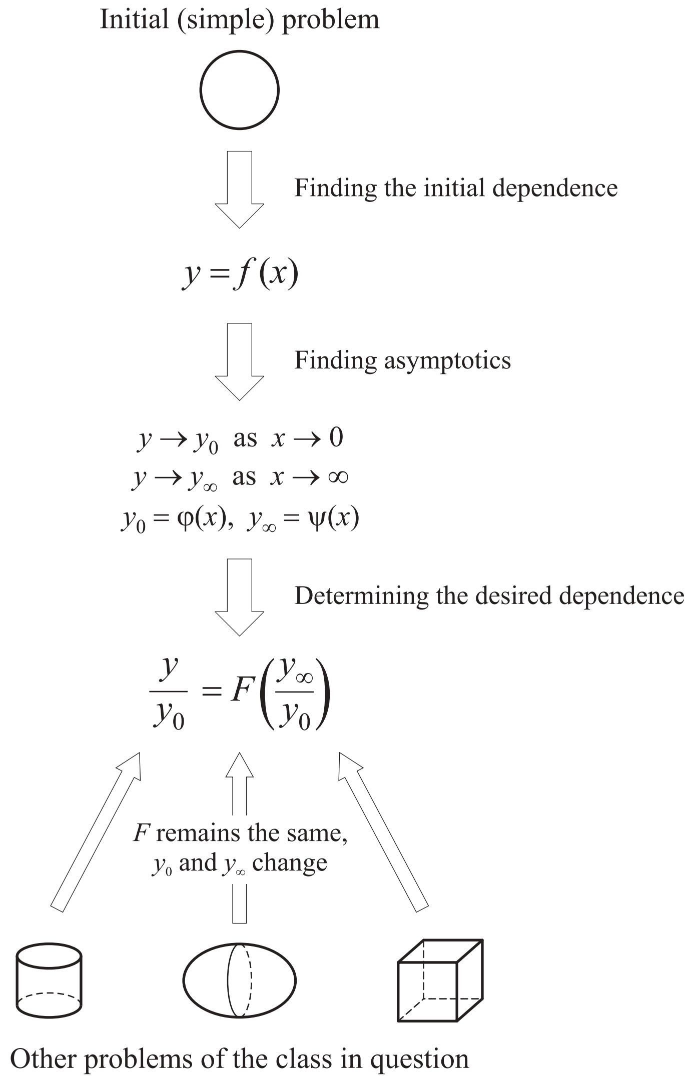
Initial (simple) problem
Finding the initial dependence
$y = f(x)$
Finding asymptotics
$y \to y_0$ as $x \to 0$
$y \to y_\infty$ as $x \to \infty$
$y_0 = \varphi(x)$, $y_\infty = \psi(x)$
Determining the desired dependence
$\frac{y}{y_0} = F\left(\frac{y_\infty}{y_0}\right)$
F remains the same,
y_0 and y_∞ change
Other problems of the class in question

Figure E7.1. Scheme of application of the method of asymptotic analogies for a class of qualitatively similar problems.

characteristics for a wider class of problems describing qualitatively similar phenomena or processes. Specifically, after the relation (E7.3.1.5) has been constructed with the help of (E7.3.1.1) for some specific (say, the simplest) case, we can evaluate y for other problems of this class by finding the asymptotics y_0 (as $x \to 0$) and y_∞ (as $x \to \infty$) and then by substituting these asymptotics into (E7.3.1.5). The approximate formulas thus obtained are asymptotically sharp in both limit cases $x \to 0$ and $x \to \infty$.

Figure E7.1 presents a principle scheme of applying the method of asymptotic analogies.

In the books by Polyanin and Dilman (1994) and Polyanin et al. (2002), the formulas obtained by the method of asymptotic analogies were compared with the already known formulas obtained by exact, numerical, and approximate methods for a large number of specific cases. These investigations confirmed the high accuracy and wide capabilities of the method of asymptotic analogies. In other words, the final functional relation (E7.3.1.5) between y and its asymptotics remains the same (or varies slightly) in a wide class of problems of the same type, and the specific features of geometric distinctions between these problems (like the surface shape and the flow structure) are sufficiently well taken into account by the corresponding asymptotic parameters y_0 and y_∞.

As a result, the scope of the final formula (E7.3.1.5) (or formulas (E7.3.1.7) and (E7.3.1.8)) is substantially wider than that of the original formula (E7.3.1.1). In this sense, one can say that formulas of the type of (E7.3.1.6) are more informative than the original formula (E7.3.1.1).

Remark. The original dependence (E7.3.1.1), as well as the asymptotics (E7.3.1.3) can be determined either theoretically or experimentally.

E7.3.2. Illustrating Examples: Interior Heat Exchange Problems for Bodies of Various Shape

▶ **Statement of the problem.** Consider a class of problems concerning transient heat exchange between convex bodies of various shape and the ambient medium. At the initial time $t=0$, the temperature is the same throughout the body and is equal to T_i, and for $t>0$, the temperature on the surface Γ of the body is maintained constant and is equal to T_s. The general mathematical statement of the problem about the temperature distribution inside the body is given in Section E5.5.1 (see formulas (E5.5.1.1)–(E5.5.1.3)).

In this section, attention is paid to the study of the bulk body temperature

$$\langle T\rangle = \frac{1}{V}\int_v T\,dv, \tag{E7.3.2.1}$$

where T is the dimensionless temperature and $V=\int_v dv$ is the dimensionless volume of the body.

▶ **General formulas for the bulk temperature of the body.** To approximate the dependence of the bulk temperature on time, we use the method of asymptotic analogies. The simplest original problem is taken to be the one-dimensional (with respect to spatial coordinates) heat exchange problem for a sphere of radius a. The exact solution of this problem is well known and results in the following expression for the bulk temperature:

$$\langle T\rangle = 1-\frac{6}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{k^2}\exp(-\pi^2k^2\bar\tau), \qquad \tau=\frac{\chi t}{a^2}, \tag{E7.3.2.2}$$

where χ is the thermal diffusivity.

The asymptotic expressions (for small and large $\bar\tau$) of (E7.3.2.2) have the form

$$\langle T\rangle_0 = 6\pi^{-1/2}\sqrt{\bar\tau}\quad(\bar\tau\to 0);\qquad \langle T\rangle_\infty = 1\quad(\bar\tau\to\infty) \tag{E7.3.2.3}$$

and are a special case of (E7.3.1.6) for $y_0=\langle T\rangle_0$ and $y_\infty=\langle T\rangle_\infty$, where $A=6\pi^{-1/2}$, $B=1$, $k=\frac{1}{2}$, and $m=0$. By substituting these values into (E7.3.1.7) with $f=\langle T\rangle$, we can rewrite (E7.3.2.2) as follows:

$$\frac{\langle T\rangle}{\langle T\rangle_\infty} = 1-\frac{6}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{k^2}\exp\left[-\frac{\pi^3}{36}k^2\left(\frac{\langle T\rangle_0}{\langle T\rangle_\infty}\right)^2\right]. \tag{E7.3.2.4}$$

Following the method of asymptotic analogies, we shall use formula (E7.3.2.4) for the calculation of bulk temperature for nonspherical bodies. To this end, for a body of a given shape, we must first calculate the asymptotics of bulk temperature for small and large $\bar{\tau}$ and then substitute these asymptotics into (E7.3.2.4).

For a bounded body of arbitrary shape, the solution of problem (E5.5.1.1)–(E5.5.1.3) tends as $\bar{\tau} \to \infty$ to the limit value (equal to 1) determined by the boundary condition on the surface of the body. By setting $T = 1$ in (E7.3.2.1), we find the asymptotics of the bulk temperature for large $\bar{\tau}$:

$$\langle T\rangle_\infty = 1. \tag{E7.3.2.5}$$

Now let us consider the initial stage of the process, corresponding to small values of dimensionless time. Let us integrate Eq. (E5.5.1.1) over the volume v occupied by the body. Taking into account the identity $\Delta T = \operatorname{div}(\operatorname{grad} T)$ and applying the Gauss divergence theorem, we replace the volume integral on the right-hand side by a surface integral. As a result, we obtain

$$\frac{\partial}{\partial\bar{\tau}}\int_v T\,dv = -\int_\Gamma \frac{\partial T}{\partial\xi}\,d\Gamma, \tag{E7.3.2.6}$$

where ξ is the coordinate along the inward normal on Γ.

For small $\bar{\tau}$, the temperature mainly varies in a thin region adjacent to the surface. In this region, the derivatives along the surface can be neglected compared with the normal derivatives. Therefore, the temperature distribution as $\bar{\tau} \to 0$ is described by the equation

$$\frac{\partial T}{\partial\bar{\tau}} = \frac{\partial^2 T}{\partial\xi^2} \tag{E7.3.2.7}$$

with the initial and boundary conditions,

$$\bar{\tau} = 0,\ \ T = 0; \qquad \xi = 0,\ \ T = 1,$$

where the value $\xi = 0$ corresponds to the surface of the body.

The solution of problem (E7.3.2.7) is given by the complementary error function

$$T = \operatorname{erfc}\left(\frac{\xi}{2\sqrt{\bar{\tau}}}\right). \tag{E7.3.2.8}$$

By differentiating this formula with respect to ξ and by setting $\xi = 0$, we obtain the asymptotics as $\bar{\tau} \to 0$ of local heat flux to the surface of the body,

$$\left(\frac{\partial T}{\partial\xi}\right)_\Gamma = -\frac{1}{\sqrt{\pi\bar{\tau}}}. \tag{E7.3.2.9}$$

Let us substitute (E7.3.2.9) into (E7.3.2.6). After the integration, we obtain

$$\frac{\partial}{\partial\bar{\tau}}\int_v T\,dv = \frac{1}{\sqrt{\pi\bar{\tau}}}S, \tag{E7.3.2.10}$$

where S is the dimensionless surface area of the body.

Let us integrate both sides of (E7.3.2.10) with respect to $\bar{\tau}$ from 0 to $\bar{\tau}$. With regard to the initial condition (E5.5.1.2) and relation (E7.3.2.1), we obtain the desired asymptotic expression for bulk temperature as $\bar{\tau} \to 0$:

$$\langle T\rangle_0 = 2\frac{S}{V}\sqrt{\frac{\bar{\tau}}{\pi}}. \tag{E7.3.2.11}$$

The substitution of (E7.3.2.5) and (E7.3.2.11) into (E7.3.2.4) gives an approximate dependence of bulk temperature of an arbitrarily shaped body on time:

$$\langle T\rangle = 1-\frac{6}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{k^2}\exp\left(-\frac{\pi^2k^2S^2}{9V^2}\bar{\tau}\right).$$

This expression can be rewritten as follows

$$\langle T\rangle = 1-\frac{6}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{k^2}\exp\left(-\frac{\pi^2}{9}k^2\frac{S_*^2\chi t}{V_*^2}\right), \tag{E7.3.2.12}$$

where S_* and V_* are, respectively, the dimensional surface area and volume of the body.

For practical calculations, it is expedient to replace the infinite series by the simpler formula

$$\langle T\rangle = \sqrt{1-e^{-1.27\,\omega}}+0.6\,\left(e^{-1.5\,\omega}-e^{-1.1\,\omega}\right),\qquad \omega=\frac{S_*^2\chi t}{V_*^2}, \tag{E7.3.2.13}$$

whose maximum deviation from (E7.3.2.12) is about 1.7% (see Table E7.3).

► **Bulk temperature for bodies of various shape.** Let us compare the approximate dependence (E7.3.2.12) with some well-known exact results on heat exchange for nonspherical bodies.

First, consider a parallelepiped with sides L_1, L_2, and L_3. The solution of the corresponding three-dimensional problem (E5.5.1.1)–(E5.5.1.3) can be constructed by separation of variables and results in the following formula for bulk temperature:

$$\begin{aligned}\langle T\rangle = 1-\left(\frac{8}{\pi^2}\right)^3\sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\sum_{l=1}^{\infty}&\frac{1}{(2k-1)^2(2m-1)^2(2l-1)^2}\\ &\times\exp\left\{-\pi^2\left[\frac{(2k-1)^2}{L_1^2}+\frac{(2m-1)^2}{L_2^2}+\frac{(2l-1)^2}{L_3^2}\right]\chi t\right\}.\end{aligned} \tag{E7.3.2.14}$$

Since the surface area and the volume of the parallelepiped are given by the formulas $S_* = 2(L_1L_2 + L_1L_3 + L_2L_3)$ and $V_* = L_1L_2L_3$, we can rewrite (E7.3.2.14) as

$$\begin{aligned}\langle T\rangle = 1-\left(\frac{8}{\pi^2}\right)^3\sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\sum_{l=1}^{\infty}&\frac{1}{(2k-1)^2(2m-1)^2(2l-1)^2}\\ &\times\exp\left[-\frac{\pi^2}{4}\frac{\left(\frac{2k-1}{L_1}\right)^2+\left(\frac{2m-1}{L_2}\right)^2+\left(\frac{2l-1}{L_3}\right)^2}{\left(\frac{1}{L_1}+\frac{1}{L_2}+\frac{1}{L_3}\right)^2}\frac{S_*^2\chi t}{V_*^2}\right].\end{aligned} \tag{E7.3.2.15}$$

In Table E7.3, the approximation (E7.3.2.12) is compared with the exact bulk temperature (E7.3.2.15) for six distinct values of L_1, L_2, and L_3. The maximum error of formulas (E7.3.2.12) and (E7.3.2.13) is about 5% for $0.25 \le L_3/L_1 \le 4.0$ and $L_2/L_1 = 1$.

TABLE E7.3
Comparison of exact and approximate values of bulk temperature $\langle T\rangle$ for bodies of various shape.

Bodies of various shape		Dimensionless time $S_*^2\chi t/V_*^2$							
		0.05	0.1	0.2	0.3	0.5	1.0	1.5	2.0
Sphere, formula (E7.3.2.12)		0.236	0.323	0.438	0.518	0.631	0.795	0.882	0.932
Approximate formula (E7.3.2.13)		0.237	0.324	0.437	0.514	0.623	0.782	0.870	0.923
Paralle-lepiped, formula (E7.3.2.15); $E_\mathrm{i} = L_\mathrm{i}/L_1$	$E_2 = 1$, $E_3 = 0.25$	0.237	0.326	0.443	0.527	0.647	0.821	0.907	0.951
	$E_2 = 1$, $E_3 = 0.5$	0.233	0.318	0.429	0.506	0.615	0.774	0.862	0.915
	$E_2 = 1$, $E_3 = 1$	0.232	0.316	0.425	0.499	0.604	0.757	0.843	0.897
	$E_2 = 1$, $E_3 = 2$	0.232	0.318	0.427	0.503	0.610	0.767	0.854	0.920
	$E_2 = 1$, $E_3 = 4$	0.234	0.320	0.432	0.510	0.620	0.782	0.871	0.952
	$E_2 = 2$, $E_3 = 4$	0.234	0.321	0.435	0.514	0.628	0.794	0.882	0.932
Cylinder, formula (E7.3.2.17); $E = 2a/L$	$E = 0.25$	0.236	0.325	0.440	0.522	0.638	0.807	0.894	0.942
	$E = 0.5$	0.234	0.321	0.434	0.513	0.624	0.787	0.875	0.926
	$E = 1$	0.233	0.319	0.429	0.506	0.613	0.770	0.857	0.910
	$E = 2$	0.234	0.320	0.431	0.509	0.619	0.780	0.868	0.920
	$E = 4$	0.237	0.326	0.444	0.528	0.649	0.823	0.909	0.952

Now consider heat exchange for a cylinder of finite length. Let a be the radius and L the length of the cylinder. By solving problem (E5.5.1.1)–(E5.5.1.3), we obtain the following expression for bulk temperature:

$$\langle T\rangle = 1 - \frac{32}{\pi^2}\sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\frac{1}{\vartheta_k^2(2m-1)^2}\exp\left\{-\left[\frac{\vartheta_k^2}{a^2} + \frac{\pi^2(2m-1)^2}{L^2}\right]\chi t\right\}, \qquad \text{(E7.3.2.16)}$$

where the ϑ_k are the roots of the Bessel function of the first kind of index zero: $J_0(\vartheta_k) = 0$. (The first sixty roots ϑ_k are tabulated in the book by Janke et al. (1960).)

Formula (E7.3.2.16) can be rewritten in the form

$$\langle T\rangle = 1 - \frac{32}{\pi^2}\sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\frac{1}{\vartheta_k^2(2m-1)^2}\exp\left[-\frac{L^2\vartheta_k^2 + \pi^2a^2(2m-1)^2}{4(a+L)^2}\,\frac{S_*^2\chi t}{V_*^2}\right], \qquad \text{(E7.3.2.17)}$$

where $S_* = 2\pi a(a+L)$ is the surface area and $V_* = \pi a^2 L$ the volume of the cylinder.

In Table E7.3, the exact value of the bulk temperature given by (E7.3.2.17) is compared with the approximation (E7.3.2.12) for various ratios of the cylinder dimensions. The maximum error in formula (E7.3.2.12) is about 3.5% for $0.25 \le 2a/L \le 4.0$.

E7.3.3. Illustrating Examples: Mass Exchange Between Bubbles/Particles and Flow

► **Mass transfer to a spherical bubble at low Reynolds numbers.** The problem of mass transfer to a spherical bubble in a translational flow as $\mathsf{Re} \to 0$ was studied in the entire range of Peclet numbers by finite-difference methods. The results for the mean Sherwood number are well approximated by the expression (E5.8.2.6)

$$\mathsf{Sh} = 0.6 + (0.16 + 0.213\,\mathsf{Pe})^{1/2}, \qquad \text{(E7.3.3.1)}$$

where the bubble radius is taken as the characteristic scale.

The maximum deviation of (E7.3.3.1) in the entire range of Peclet numbers is about 3%.

We use formula (E7.3.3.1) as a basis for obtaining a more general dependence, which can be used for nonspherical bubbles and more complicated flows. To this end, we use the method of asymptotic analogies, by setting $x = \mathsf{Pe}$ and $y = \mathsf{Sh}$.

From formula (E7.3.3.1) we find its asymptotics:

$$\mathsf{Sh}_0 = 1 \quad (\mathsf{Pe} \to 0), \qquad \mathsf{Sh}_\infty = (0.213\,\mathsf{Pe})^{1/2} \quad (\mathsf{Pe} \to \infty), \tag{E7.3.3.2}$$

which corresponds to the values of the exponents $k = 0$ and $m = 1/2$ in formulas (E7.3.1.6)–(E7.3.1.8).

Passing from the original variables Pe and Sh to the asymptotic coordinates $\mathsf{Sh}/\mathsf{Sh}_0$ and $\mathsf{Sh}/\mathsf{Sh}_\infty$ in (E7.3.3.1), we obtain the dependence

$$\frac{\mathsf{Sh}}{\mathsf{Sh}_0} = 0.6 + \left[0.16 + \left(\frac{\mathsf{Sh}_\infty}{\mathsf{Sh}_0}\right)^2\right]^{1/2}, \tag{E7.3.3.3}$$

where Sh_∞ is the asymptotic value of the mean Sherwood number as $\mathsf{Pe} \to \infty$, which can be calculated by solving the diffusion boundary layer equation in a given flow field or determined experimentally.

Now we consider an *axisymmetric shear flow*, where the dimensional fluid velocity components remote from the bubble have the following form in the Cartesian coordinates X_1, X_2, X_3:

$$\mathbf{V} = (V_1, V_2, V_3) = (-GX_1,\ -GX_2,\ 2GX_3)\,, \tag{E7.3.3.4}$$

where G is the shear coefficient.

For a spherical bubble in a shear Stokes flow (E7.3.3.4), the asymptotics of the mean Sherwood number has the form (Polyanin et al., 2002)

$$\mathsf{Sh}_0 = 1 \quad (\mathsf{Pe} \to 0), \qquad \mathsf{Sh}_\infty = \left(\frac{3}{\pi}\,\mathsf{Pe}\right)^{1/2} \quad (\mathsf{Pe} \to \infty), \tag{E7.3.3.5}$$

where $\mathsf{Pe} = a^2|G|/D$. By substituting the asymptotics (E7.3.3.5) into formula (E7.3.3.3), we obtain the formula

$$\mathsf{Sh} = 0.6 + (0.16 + 0.955\,\mathsf{Pe})^{1/2},$$

whose error in the entire range of the Peclet numbers does not exceed 3%.

► **Mass transfer to a solid particle at low Reynolds numbers.** The problem of mass transfer to a solid spherical particle in a translational Stokes flow ($\mathsf{Re} \to 0$) was studied in the entire range of Peclet numbers by finite-difference methods. To find the mean Sherwood number for a spherical particle, it is convenient to use the following approximate formula (E5.7.2.8):

$$\mathsf{Sh} = 0.5 + (0.125 + 0.243\,\mathsf{Pe})^{1/3}, \tag{E7.3.3.6}$$

where the particle radius is taken as the characteristic scale.

The maximum deviation of (E7.3.3.6) in the entire range of Peclet numbers is about 2%.

From formula (E7.3.3.6), we find its asymptotics:

$$\begin{aligned} \mathsf{Sh}_0 &= 1 & &\Big(\lim_{\mathsf{Pe}\to 0} \mathsf{Sh}/\mathsf{Sh}_0 = 1\Big), \\ \mathsf{Sh}_\infty &= (0.243\,\mathsf{Pe})^{1/3} & &\Big(\lim_{\mathsf{Pe}\to \infty} \mathsf{Sh}/\mathsf{Sh}_\infty = 1\Big), \end{aligned} \tag{E7.3.3.7}$$

which corresponds to the exponents $k = 0$ and $m = 1/3$ in formulas (E7.3.1.6)–(E7.3.1.8).

By passing from the original variables to the asymptotic coordinates in (E7.3.3.6) with the use of (E7.3.3.7), one can derive the approximate formula

$$\frac{\mathsf{Sh}}{\mathsf{Sh}_0} = 0.5 + \left[0.125 + \left(\frac{\mathsf{Sh}_\infty}{\mathsf{Sh}_0}\right)^3\right]^{1/3}, \tag{E7.3.3.8}$$

which is equivalent to formula (E5.9.3.2). The approximate formula (E7.3.3.8) can be used to calculate the mean Sherwood number for a steady-state viscous flow (without closed streamlines) past smooth particles of nonspherical shape (for example, ellipsoidal particles). The auxiliary value Sh_∞ must be chosen equal to the leading term of the asymptotic expansion of the Sherwood number as $\mathsf{Pe} \to \infty$.

For a particle of a given shape, the auxiliary quantities Sh_0 and Sh_∞ occurring in (E7.3.3.8) can be determined either theoretically or experimentally (see Section E5.9.3).

Remark. For a solid spherical particle in an axisymmetric shear flow (E7.3.3.4), the maximum error of formula (E7.3.3.8) in the entire region of Peclet numbers is about 3%.

► **Transient mass transfer in steady-state flows.** We consider a laminar steady-state flow past a solid spherical particle (drop or bubble) of radius a and study transient mass transfer to the particle surface. At the initial time $t = 0$, the concentration in the continuous phase is constant and equal to C_i, whereas for $t > 0$ a constant concentration C_s is maintained on the particle surface. The mathematical statement of the problem about the concentration distribution in the flow is given in Section E5.10.1 (see formulas (E5.10.1.1) and (E5.10.1.2)).

In what follows, we restrict ourselves to the case of high Peclet numbers, in which there are no closed streamlines in the flow.

For an axisymmetric shear flow past a bubble (the fluid velocity field far from the bubble is determined by formula (E7.3.3.4)), the mean Sherwood number can be calculated by the formula

$$\mathsf{Sh} = \sqrt{\frac{3\,\mathsf{Pe}}{\pi}\coth(3\,\mathsf{Pe}\,\tau)}, \tag{E7.3.3.9}$$

which was obtained in the diffusion boundary layer approximation. In formula (E7.3.3.9), we use the notation $\tau = Dt/a^2$ and $\mathsf{Pe} = a^2G/D$.

From formula (E7.3.3.9), we can find the asymptotics

$$\mathsf{Sh}_0 = (\pi\tau)^{-1/2} \quad (\tau \to 0), \qquad \mathsf{Sh}_\infty = (3\,\mathsf{Pe}/\pi)^{1/2} \quad (\tau \to \infty). \tag{E7.3.3.10}$$

By passing to asymptotic coordinates in (E7.3.3.9) with the use of (E7.3.3.10), we obtain

$$\frac{\mathsf{Sh}}{\mathsf{Sh}_\infty} = \sqrt{\coth\left(\frac{\mathsf{Sh}_\infty}{\mathsf{Sh}_0}\right)^2}. \tag{E7.3.3.11}$$

For an arbitrary steady-state flow past spherical particles, drops, and bubbles, we have $\mathsf{Sh}_0 = (\pi\tau)^{-1/2}$, and formula (E7.3.3.11) becomes

$$\frac{\mathsf{Sh}}{\mathsf{Sh}_\infty} = \sqrt{\coth(\pi\mathsf{Sh}_\infty^2\tau)}, \tag{E7.3.3.12}$$

where $\mathsf{Sh}_\infty = \lim\limits_{\tau\to\infty} \mathsf{Sh}$ is the Sherwood number for the steady-state diffusion mode and $\tau = Dt/a^2$.

Equation (E7.3.3.12) gives a valid asymptotic result for any steady-state flow field in both limit cases $\tau \to 0$ and $\tau \to \infty$.

Table E7.4 presents a comparison of the mean Sherwood numbers calculated according to Eq. (E7.3.3.12) with available data for various flows past spherical drops, bubbles, and solid particles at high Peclet numbers (in this table, we use the abbreviation DBLA for "diffusion boundary layer approximation"). The parameter β is the ratio of the dynamic viscosity of the drop to that of the ambient fluid and varies in the range $0 \le \beta \le 2$. (The value $\beta = 0$ corresponds to a gas bubble.)

Remark. One can find other examples of using the method of asymptotic analogies in the books by Polyanin and Dilman (1994) and Polyanin et al. (2002).

TABLE E7.4
Maximum error of Eq. (E7.3.3.12) for various types of flow past spherical drops, bubbles, and solid particles (according to the data by Polyanin and Dilman, 1994).

No.	Dispersed phase	Flow type	Solution method	Error, %
1	Drop, bubble	Axisymmetric shear Stokes flow	Analytical, DBLA	0
2	Drop, bubble	Translational Stokes flow	Analytical, DBLA	0.7
3	Drop, bubble	Two-dimensional shear Stokes flow	Analytical, DBLA	1.8
4	Bubble	Laminar translational flow at high Reynolds numbers	Analytical, DBLA	0.7
5	Bubble	Axisymmetric shear flow at high Reynolds numbers	Analytical, DBLA	0
6	Drop, bubble	Flow caused by an electric field	Analytical, DBLA	0
7	Solid particle	Translational flow of an ideal (inviscid) liquid	Analytical, DBLA	0.7
8	Solid particle	Translational Stokes flow	Interpolation of numerical and analytical results	1.4
9	Solid particle	Translational Stokes flow	Finite-difference numerical method (at $\mathsf{Pe} = 500$)	4

Bibliography for Chapter E7

Björck, A., *Numerical Methods for Least Squares Problems*, Society for Ind. & Appl. Math., Philadelphia, 1996.

Demidovich, B. P., Maron, I. A., and Shuvalova, E., Z., *Numerical Methods of Analysis* [in Russian], Fizmatlit, Moscow, 1963.

Demidovich, B. P. and Maron, I. A., *Foundations of Comutational Mathematics* [in Russian], Nauka, Moscow, 1970.

Gieck, K. and Gieck, R., *Engineering Formulas, 8th Edition*, McGraw-Hill, New York, 2006.

Janke E., Emde F., Lösch F., *Tafeln Höherer Funktionen*, Teubner Verlogsgesellschaft, Stuttgart, 1960.

Kariya, T. and Kurata, H., *Generalized Least Squares*, Wiley, Hoboken, New Jersey, 2004.

Lawson, C. and Hanson, R., *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, New Jersey, 1974.

Merriman, M., *A Textbook on the Method of Least Squares, 8th Edition*, Cornell University Library, 2009.

Polyanin, A. D. and Dilman, V. V., *Methods of Modeling Equations and Analogies in Chemical Engineering*, CRC Press–Begell House, Boca Raton, Florida, 1994.

Polyanin, A. D., Kutepov, A. M., Vyazmin, A. V., and Kazenin, D. A., *Hydrodynamics, Mass and Heat Transfer in Chemical Engineering*, Taylor & Francis, London, 2002.

Polyanin, A. D., Vyazmina, E. A., and Dilman, V. V., New methods for processing experimental data: Applicability tests for empirical formulas, *Theor. Found. Chem. Eng.,* Vol. 42, No. 4, pp. 354–365, 2008.

Running, T. R., *Empirical Formulas*, Cornell University Library, 2009.

Semendyaev, K. A., *Empirical Formulas* [in Russian], Gos. Tech.-Teor. Izdat., Moscow, 1933.

Weisstein, E. W., *CRC Concise Encyclopedia of Mathematics, 2nd Edition*, CRC Press, Boca Raton, Florida, 2003.

Wolberg, J., *Data Analysis Using the Method of Least Squares: Extracting the Most Information from Experiments*, Springer-Verlag, Berlin, 2006.

Part IV

Supplements

Chapter S1

Integrals

S1.1. Indefinite Integrals

Throughout Section S1.1, the integration constant C is omitted for brevity.

S1.1.1. Integrals Involving Rational Functions

► **Integrals involving $a + bx$.**

1. $\displaystyle\int \frac{dx}{a+bx} = \frac{1}{b}\ln|a+bx|.$

2. $\displaystyle\int (a+bx)^n\,dx = \frac{(a+bx)^{n+1}}{b(n+1)}, \qquad n \neq -1.$

3. $\displaystyle\int \frac{x\,dx}{a+bx} = \frac{1}{b^2}\bigl(a+bx-a\ln|a+bx|\bigr).$

4. $\displaystyle\int \frac{x^2\,dx}{a+bx} = \frac{1}{b^3}\left[\frac{1}{2}(a+bx)^2 - 2a(a+bx) + a^2\ln|a+bx|\right].$

5. $\displaystyle\int \frac{dx}{x(a+bx)} = -\frac{1}{a}\ln\left|\frac{a+bx}{x}\right|.$

6. $\displaystyle\int \frac{dx}{x^2(a+bx)} = -\frac{1}{ax} + \frac{b}{a^2}\ln\left|\frac{a+bx}{x}\right|.$

7. $\displaystyle\int \frac{x\,dx}{(a+bx)^2} = \frac{1}{b^2}\left(\ln|a+bx| + \frac{a}{a+bx}\right).$

8. $\displaystyle\int \frac{x^2\,dx}{(a+bx)^2} = \frac{1}{b^3}\left(a+bx-2a\ln|a+bx| - \frac{a^2}{a+bx}\right).$

9. $\displaystyle\int \frac{dx}{x(a+bx)^2} = \frac{1}{a(a+bx)} - \frac{1}{a^2}\ln\left|\frac{a+bx}{x}\right|.$

10. $\displaystyle\int \frac{x\,dx}{(a+bx)^3} = \frac{1}{b^2}\left[-\frac{1}{a+bx} + \frac{a}{2(a+bx)^2}\right].$

► **Integrals involving $a + x$ and $b + x$.**

1. $\displaystyle\int \frac{a+x}{b+x}\,dx = x + (a-b)\ln|b+x|.$

2. $\displaystyle\int \frac{dx}{(a+x)(b+x)} = \frac{1}{a-b}\ln\left|\frac{b+x}{a+x}\right|,$ $\quad a \neq b$. For $a = b$, see Integral 2 with $n = -2$ in the previous item.

3. $\displaystyle \int \frac{x\,dx}{(a+x)(b+x)} = \frac{1}{a-b}\bigl(a\ln|a+x| - b\ln|b+x|\bigr).$

4. $\displaystyle \int \frac{dx}{(a+x)(b+x)^2} = \frac{1}{(b-a)(b+x)} + \frac{1}{(a-b)^2}\ln\left|\frac{a+x}{b+x}\right|.$

5. $\displaystyle \int \frac{x\,dx}{(a+x)(b+x)^2} = \frac{b}{(a-b)(b+x)} - \frac{a}{(a-b)^2}\ln\left|\frac{a+x}{b+x}\right|.$

6. $\displaystyle \int \frac{x^2\,dx}{(a+x)(b+x)^2} = \frac{b^2}{(b-a)(b+x)} + \frac{a^2}{(a-b)^2}\ln|a+x| + \frac{b^2-2ab}{(b-a)^2}\ln|b+x|.$

7. $\displaystyle \int \frac{dx}{(a+x)^2(b+x)^2} = -\frac{1}{(a-b)^2}\left(\frac{1}{a+x} + \frac{1}{b+x}\right) + \frac{2}{(a-b)^3}\ln\left|\frac{a+x}{b+x}\right|.$

8. $\displaystyle \int \frac{x\,dx}{(a+x)^2(b+x)^2} = \frac{1}{(a-b)^2}\left(\frac{a}{a+x} + \frac{b}{b+x}\right) + \frac{a+b}{(a-b)^3}\ln\left|\frac{a+x}{b+x}\right|.$

9. $\displaystyle \int \frac{x^2\,dx}{(a+x)^2(b+x)^2} = -\frac{1}{(a-b)^2}\left(\frac{a^2}{a+x} + \frac{b^2}{b+x}\right) + \frac{2ab}{(a-b)^3}\ln\left|\frac{a+x}{b+x}\right|.$

► Integrals involving $a^2 + x^2$.

1. $\displaystyle \int \frac{dx}{a^2+x^2} = \frac{1}{a}\arctan\frac{x}{a}.$

2. $\displaystyle \int \frac{dx}{(a^2+x^2)^2} = \frac{x}{2a^2(a^2+x^2)} + \frac{1}{2a^3}\arctan\frac{x}{a}.$

3. $\displaystyle \int \frac{dx}{(a^2+x^2)^3} = \frac{x}{4a^2(a^2+x^2)^2} + \frac{3x}{8a^4(a^2+x^2)} + \frac{3}{8a^5}\arctan\frac{x}{a}.$

4. $\displaystyle \int \frac{dx}{(a^2+x^2)^{n+1}} = \frac{x}{2na^2(a^2+x^2)^n} + \frac{2n-1}{2na^2}\int \frac{dx}{(a^2+x^2)^n}; \quad n = 1,\ 2,\ \ldots$

5. $\displaystyle \int \frac{x\,dx}{a^2+x^2} = \frac{1}{2}\ln(a^2+x^2).$

6. $\displaystyle \int \frac{x\,dx}{(a^2+x^2)^2} = -\frac{1}{2(a^2+x^2)}.$

7. $\displaystyle \int \frac{x\,dx}{(a^2+x^2)^3} = -\frac{1}{4(a^2+x^2)^2}.$

8. $\displaystyle \int \frac{x\,dx}{(a^2+x^2)^{n+1}} = -\frac{1}{2n(a^2+x^2)^n}; \quad n = 1,\ 2,\ \ldots$

9. $\displaystyle \int \frac{x^2\,dx}{a^2+x^2} = x - a\arctan\frac{x}{a}.$

10. $\displaystyle \int \frac{x^2\,dx}{(a^2+x^2)^2} = -\frac{x}{2(a^2+x^2)} + \frac{1}{2a}\arctan\frac{x}{a}.$

11. $\displaystyle \int \frac{x^2\,dx}{(a^2+x^2)^3} = -\frac{x}{4(a^2+x^2)^2} + \frac{x}{8a^2(a^2+x^2)} + \frac{1}{8a^3}\arctan\frac{x}{a}.$

12. $\displaystyle \int \frac{x^2\,dx}{(a^2+x^2)^{n+1}} = -\frac{x}{2n(a^2+x^2)^n} + \frac{1}{2n}\int \frac{dx}{(a^2+x^2)^n}; \quad n = 1,\ 2,\ \ldots$

13. $\displaystyle\int \frac{x^3\,dx}{a^2+x^2} = \frac{x^2}{2} - \frac{a^2}{2}\ln(a^2+x^2).$

14. $\displaystyle\int \frac{x^3\,dx}{(a^2+x^2)^2} = \frac{a^2}{2(a^2+x^2)} + \frac{1}{2}\ln(a^2+x^2).$

15. $\displaystyle\int \frac{x^3\,dx}{(a^2+x^2)^{n+1}} = -\frac{1}{2(n-1)(a^2+x^2)^{n-1}} + \frac{a^2}{2n(a^2+x^2)^n}; \quad n = 2,\ 3,\ \ldots$

16. $\displaystyle\int \frac{dx}{x(a^2+x^2)} = \frac{1}{2a^2}\ln\frac{x^2}{a^2+x^2}.$

17. $\displaystyle\int \frac{dx}{x(a^2+x^2)^2} = \frac{1}{2a^2(a^2+x^2)} + \frac{1}{2a^4}\ln\frac{x^2}{a^2+x^2}.$

18. $\displaystyle\int \frac{dx}{x(a^2+x^2)^3} = \frac{1}{4a^2(a^2+x^2)^2} + \frac{1}{2a^4(a^2+x^2)} + \frac{1}{2a^6}\ln\frac{x^2}{a^2+x^2}.$

19. $\displaystyle\int \frac{dx}{x^2(a^2+x^2)} = -\frac{1}{a^2x} - \frac{1}{a^3}\arctan\frac{x}{a}.$

20. $\displaystyle\int \frac{dx}{x^2(a^2+x^2)^2} = -\frac{1}{a^4x} - \frac{x}{2a^4(a^2+x^2)} - \frac{3}{2a^5}\arctan\frac{x}{a}.$

21. $\displaystyle\int \frac{dx}{x^3(a^2+x^2)^2} = -\frac{1}{2a^4x^2} - \frac{1}{2a^4(a^2+x^2)} - \frac{1}{a^6}\ln\frac{x^2}{a^2+x^2}.$

22. $\displaystyle\int \frac{dx}{x^2(a^2+x^2)^3} = -\frac{1}{a^6x} - \frac{x}{4a^4(a^2+x^2)^2} - \frac{7x}{8a^6(a^2+x^2)} - \frac{15}{8a^7}\arctan\frac{x}{a}.$

23. $\displaystyle\int \frac{dx}{x^3(a^2+x^2)^3} = -\frac{1}{2a^6x^2} - \frac{1}{a^6(a^2+x^2)} - \frac{1}{4a^4(a^2+x^2)^2} - \frac{3}{2a^8}\ln\frac{x^2}{a^2+x^2}.$

► **Integrals involving $a^2 - x^2$.**

1. $\displaystyle\int \frac{dx}{a^2-x^2} = \frac{1}{2a}\ln\left|\frac{a+x}{a-x}\right|.$

2. $\displaystyle\int \frac{dx}{(a^2-x^2)^2} = \frac{x}{2a^2(a^2-x^2)} + \frac{1}{4a^3}\ln\left|\frac{a+x}{a-x}\right|.$

3. $\displaystyle\int \frac{dx}{(a^2-x^2)^3} = \frac{x}{4a^2(a^2-x^2)^2} + \frac{3x}{8a^4(a^2-x^2)} + \frac{3}{16a^5}\ln\left|\frac{a+x}{a-x}\right|.$

4. $\displaystyle\int \frac{dx}{(a^2-x^2)^{n+1}} = \frac{x}{2na^2(a^2-x^2)^n} + \frac{2n-1}{2na^2}\int \frac{dx}{(a^2-x^2)^n}; \quad n = 1,\ 2,\ \ldots$

5. $\displaystyle\int \frac{x\,dx}{a^2-x^2} = -\frac{1}{2}\ln|a^2-x^2|.$

6. $\displaystyle\int \frac{x\,dx}{(a^2-x^2)^2} = \frac{1}{2(a^2-x^2)}.$

7. $\displaystyle\int \frac{x\,dx}{(a^2-x^2)^3} = \frac{1}{4(a^2-x^2)^2}.$

8. $\displaystyle\int \frac{x\,dx}{(a^2-x^2)^{n+1}} = \frac{1}{2n(a^2-x^2)^n}; \quad n = 1,\ 2,\ \ldots$

9. $\displaystyle\int \frac{x^2\,dx}{a^2-x^2} = -x + \frac{a}{2}\ln\left|\frac{a+x}{a-x}\right|.$

10. $\displaystyle\int \frac{x^2\,dx}{(a^2-x^2)^2} = \frac{x}{2(a^2-x^2)} - \frac{1}{4a}\ln\left|\frac{a+x}{a-x}\right|.$

11. $\displaystyle\int \frac{x^2\,dx}{(a^2-x^2)^3} = \frac{x}{4(a^2-x^2)^2} - \frac{x}{8a^2(a^2-x^2)} - \frac{1}{16a^3}\ln\left|\frac{a+x}{a-x}\right|.$

12. $\displaystyle\int \frac{x^2\,dx}{(a^2-x^2)^{n+1}} = \frac{x}{2n(a^2-x^2)^n} - \frac{1}{2n}\int \frac{dx}{(a^2-x^2)^n}; \quad n = 1,\,2,\,\ldots$

13. $\displaystyle\int \frac{x^3\,dx}{a^2-x^2} = -\frac{x^2}{2} - \frac{a^2}{2}\ln|a^2-x^2|.$

14. $\displaystyle\int \frac{x^3\,dx}{(a^2-x^2)^2} = \frac{a^2}{2(a^2-x^2)} + \frac{1}{2}\ln|a^2-x^2|.$

15. $\displaystyle\int \frac{x^3\,dx}{(a^2-x^2)^{n+1}} = -\frac{1}{2(n-1)(a^2-x^2)^{n-1}} + \frac{a^2}{2n(a^2-x^2)^n}; \quad n = 2,\,3,\ldots$

16. $\displaystyle\int \frac{dx}{x(a^2-x^2)} = \frac{1}{2a^2}\ln\left|\frac{x^2}{a^2-x^2}\right|.$

17. $\displaystyle\int \frac{dx}{x(a^2-x^2)^2} = \frac{1}{2a^2(a^2-x^2)} + \frac{1}{2a^4}\ln\left|\frac{x^2}{a^2-x^2}\right|.$

18. $\displaystyle\int \frac{dx}{x(a^2-x^2)^3} = \frac{1}{4a^2(a^2-x^2)^2} + \frac{1}{2a^4(a^2-x^2)} + \frac{1}{2a^6}\ln\left|\frac{x^2}{a^2-x^2}\right|.$

► Integrals involving $a^3 + x^3$.

1. $\displaystyle\int \frac{dx}{a^3+x^3} = \frac{1}{6a^2}\ln\frac{(a+x)^2}{a^2-ax+x^2} + \frac{1}{a^2\sqrt{3}}\arctan\frac{2x-a}{a\sqrt{3}}.$

2. $\displaystyle\int \frac{dx}{(a^3+x^3)^2} = \frac{x}{3a^3(a^3+x^3)} + \frac{2}{3a^3}\int \frac{dx}{a^3+x^3}.$

3. $\displaystyle\int \frac{x\,dx}{a^3+x^3} = \frac{1}{6a}\ln\frac{a^2-ax+x^2}{(a+x)^2} + \frac{1}{a\sqrt{3}}\arctan\frac{2x-a}{a\sqrt{3}}.$

4. $\displaystyle\int \frac{x\,dx}{(a^3+x^3)^2} = \frac{x^2}{3a^3(a^3+x^3)} + \frac{1}{3a^3}\int \frac{x\,dx}{a^3+x^3}.$

5. $\displaystyle\int \frac{x^2\,dx}{a^3+x^3} = \frac{1}{3}\ln|a^3+x^3|.$

6. $\displaystyle\int \frac{dx}{x(a^3+x^3)} = \frac{1}{3a^3}\ln\left|\frac{x^3}{a^3+x^3}\right|.$

7. $\displaystyle\int \frac{dx}{x(a^3+x^3)^2} = \frac{1}{3a^3(a^3+x^3)} + \frac{1}{3a^6}\ln\left|\frac{x^3}{a^3+x^3}\right|.$

8. $\displaystyle\int \frac{dx}{x^2(a^3+x^3)} = -\frac{1}{a^3x} - \frac{1}{a^3}\int \frac{x\,dx}{a^3+x^3}.$

9. $\displaystyle\int \frac{dx}{x^2(a^3+x^3)^2} = -\frac{1}{a^6x} - \frac{x^2}{3a^6(a^3+x^3)} - \frac{4}{3a^6}\int \frac{x\,dx}{a^3+x^3}.$

► **Integrals involving $a^3 - x^3$.**

1. $$\int \frac{dx}{a^3-x^3} = \frac{1}{6a^2}\ln\frac{a^2+ax+x^2}{(a-x)^2} + \frac{1}{a^2\sqrt{3}}\arctan\frac{2x+a}{a\sqrt{3}}.$$

2. $$\int \frac{dx}{(a^3-x^3)^2} = \frac{x}{3a^3(a^3-x^3)} + \frac{2}{3a^3}\int\frac{dx}{a^3-x^3}.$$

3. $$\int \frac{x\,dx}{a^3-x^3} = \frac{1}{6a}\ln\frac{a^2+ax+x^2}{(a-x)^2} - \frac{1}{a\sqrt{3}}\arctan\frac{2x+a}{a\sqrt{3}}.$$

4. $$\int \frac{x\,dx}{(a^3-x^3)^2} = \frac{x^2}{3a^3(a^3-x^3)} + \frac{1}{3a^3}\int\frac{x\,dx}{a^3-x^3}.$$

5. $$\int \frac{x^2\,dx}{a^3-x^3} = -\frac{1}{3}\ln|a^3-x^3|.$$

6. $$\int \frac{dx}{x(a^3-x^3)} = \frac{1}{3a^3}\ln\left|\frac{x^3}{a^3-x^3}\right|.$$

7. $$\int \frac{dx}{x(a^3-x^3)^2} = \frac{1}{3a^3(a^3-x^3)} + \frac{1}{3a^6}\ln\left|\frac{x^3}{a^3-x^3}\right|.$$

8. $$\int \frac{dx}{x^2(a^3-x^3)} = -\frac{1}{a^3x} + \frac{1}{a^3}\int\frac{x\,dx}{a^3-x^3}.$$

9. $$\int \frac{dx}{x^2(a^3-x^3)^2} = -\frac{1}{a^6x} - \frac{x^2}{3a^6(a^3-x^3)} + \frac{4}{3a^6}\int\frac{x\,dx}{a^3-x^3}.$$

► **Integrals involving $a^4 \pm x^4$.**

1. $$\int \frac{dx}{a^4+x^4} = \frac{1}{4a^3\sqrt{2}}\ln\frac{a^2+ax\sqrt{2}+x^2}{a^2-ax\sqrt{2}+x^2} + \frac{1}{2a^3\sqrt{2}}\arctan\frac{ax\sqrt{2}}{a^2-x^2}.$$

2. $$\int \frac{x\,dx}{a^4+x^4} = \frac{1}{2a^2}\arctan\frac{x^2}{a^2}.$$

3. $$\int \frac{x^2\,dx}{a^4+x^4} = -\frac{1}{4a\sqrt{2}}\ln\frac{a^2+ax\sqrt{2}+x^2}{a^2-ax\sqrt{2}+x^2} + \frac{1}{2a\sqrt{2}}\arctan\frac{ax\sqrt{2}}{a^2-x^2}.$$

4. $$\int \frac{dx}{a^4-x^4} = \frac{1}{4a^3}\ln\left|\frac{a+x}{a-x}\right| + \frac{1}{2a^3}\arctan\frac{x}{a}.$$

5. $$\int \frac{x\,dx}{a^4-x^4} = \frac{1}{4a^2}\ln\left|\frac{a^2+x^2}{a^2-x^2}\right|.$$

6. $$\int \frac{x^2\,dx}{a^4-x^4} = \frac{1}{4a}\ln\left|\frac{a+x}{a-x}\right| - \frac{1}{2a}\arctan\frac{x}{a}.$$

S1.1.2. Integrals Involving Irrational Functions

► **Integrals involving $x^{1/2}$.**

1. $$\int \frac{x^{1/2}\,dx}{a^2+b^2x} = \frac{2}{b^2}x^{1/2} - \frac{2a}{b^3}\arctan\frac{bx^{1/2}}{a}.$$

2. $\displaystyle\int \frac{x^{3/2}\,dx}{a^2+b^2x} = \frac{2x^{3/2}}{3b^2} - \frac{2a^2x^{1/2}}{b^4} + \frac{2a^3}{b^5}\arctan\frac{bx^{1/2}}{a}.$

3. $\displaystyle\int \frac{x^{1/2}\,dx}{(a^2+b^2x)^2} = -\frac{x^{1/2}}{b^2(a^2+b^2x)} + \frac{1}{ab^3}\arctan\frac{bx^{1/2}}{a}.$

4. $\displaystyle\int \frac{x^{3/2}\,dx}{(a^2+b^2x)^2} = \frac{2x^{3/2}}{b^2(a^2+b^2x)} + \frac{3a^2x^{1/2}}{b^4(a^2+b^2x)} - \frac{3a}{b^5}\arctan\frac{bx^{1/2}}{a}.$

5. $\displaystyle\int \frac{dx}{(a^2+b^2x)x^{1/2}} = \frac{2}{ab}\arctan\frac{bx^{1/2}}{a}.$

6. $\displaystyle\int \frac{dx}{(a^2+b^2x)x^{3/2}} = -\frac{2}{a^2x^{1/2}} - \frac{2b}{a^3}\arctan\frac{bx^{1/2}}{a}.$

7. $\displaystyle\int \frac{dx}{(a^2+b^2x)^2x^{1/2}} = \frac{x^{1/2}}{a^2(a^2+b^2x)} + \frac{1}{a^3b}\arctan\frac{bx^{1/2}}{a}.$

8. $\displaystyle\int \frac{x^{1/2}\,dx}{a^2-b^2x} = -\frac{2}{b^2}x^{1/2} + \frac{2a}{b^3}\ln\left|\frac{a+bx^{1/2}}{a-bx^{1/2}}\right|.$

9. $\displaystyle\int \frac{x^{3/2}\,dx}{a^2-b^2x} = -\frac{2x^{3/2}}{3b^2} - \frac{2a^2x^{1/2}}{b^4} + \frac{a^3}{b^5}\ln\left|\frac{a+bx^{1/2}}{a-bx^{1/2}}\right|.$

10. $\displaystyle\int \frac{x^{1/2}\,dx}{(a^2-b^2x)^2} = \frac{x^{1/2}}{b^2(a^2-b^2x)} - \frac{1}{2ab^3}\ln\left|\frac{a+bx^{1/2}}{a-bx^{1/2}}\right|.$

11. $\displaystyle\int \frac{x^{3/2}\,dx}{(a^2-b^2x)^2} = \frac{3a^2x^{1/2}-2b^2x^{3/2}}{b^4(a^2-b^2x)} - \frac{3a}{2b^5}\ln\left|\frac{a+bx^{1/2}}{a-bx^{1/2}}\right|.$

12. $\displaystyle\int \frac{dx}{(a^2-b^2x)x^{1/2}} = \frac{1}{ab}\ln\left|\frac{a+bx^{1/2}}{a-bx^{1/2}}\right|.$

13. $\displaystyle\int \frac{dx}{(a^2-b^2x)x^{3/2}} = -\frac{2}{a^2x^{1/2}} + \frac{b}{a^3}\ln\left|\frac{a+bx^{1/2}}{a-bx^{1/2}}\right|.$

14. $\displaystyle\int \frac{dx}{(a^2-b^2x)^2x^{1/2}} = \frac{x^{1/2}}{a^2(a^2-b^2x)} + \frac{1}{2a^3b}\ln\left|\frac{a+bx^{1/2}}{a-bx^{1/2}}\right|.$

► **Integrals involving $(a+bx)^{p/2}$.**

1. $\displaystyle\int (a+bx)^{p/2}\,dx = \frac{2}{b(p+2)}(a+bx)^{(p+2)/2}.$

2. $\displaystyle\int x(a+bx)^{p/2}\,dx = \frac{2}{b^2}\left[\frac{(a+bx)^{(p+4)/2}}{p+4} - \frac{a(a+bx)^{(p+2)/2}}{p+2}\right].$

3. $\displaystyle\int x^2(a+bx)^{p/2}\,dx = \frac{2}{b^3}\left[\frac{(a+bx)^{(p+6)/2}}{p+6} - \frac{2a(a+bx)^{(p+4)/2}}{p+4} + \frac{a^2(a+bx)^{(p+2)/2}}{p+2}\right].$

► **Integrals involving $(x^2+a^2)^{1/2}$.**

1. $\displaystyle\int (x^2+a^2)^{1/2}\,dx = \frac{1}{2}x(a^2+x^2)^{1/2} + \frac{a^2}{2}\ln\bigl[x+(x^2+a^2)^{1/2}\bigr].$

2. $\displaystyle\int x(x^2+a^2)^{1/2}\,dx = \frac{1}{3}(a^2+x^2)^{3/2}.$

3. $\displaystyle\int (x^2+a^2)^{3/2}\,dx = \frac{1}{4}x(a^2+x^2)^{3/2} + \frac{3}{8}a^2x(a^2+x^2)^{1/2} + \frac{3}{8}a^4\ln\bigl|x+(x^2+a^2)^{1/2}\bigr|.$

4. $\displaystyle\int \frac{1}{x}(x^2+a^2)^{1/2}\,dx = (a^2+x^2)^{1/2} - a\ln\left|\frac{a+(x^2+a^2)^{1/2}}{x}\right|.$

5. $\displaystyle\int \frac{dx}{\sqrt{x^2+a^2}} = \ln\bigl[x+(x^2+a^2)^{1/2}\bigr].$

6. $\displaystyle\int \frac{x\,dx}{\sqrt{x^2+a^2}} = (x^2+a^2)^{1/2}.$

7. $\displaystyle\int (x^2+a^2)^{-3/2}\,dx = a^{-2}x(x^2+a^2)^{-1/2}.$

► **Integrals involving $(x^2-a^2)^{1/2}$.**

1. $\displaystyle\int (x^2-a^2)^{1/2}\,dx = \frac{1}{2}x(x^2-a^2)^{1/2} - \frac{a^2}{2}\ln\bigl|x+(x^2-a^2)^{1/2}\bigr|.$

2. $\displaystyle\int x(x^2-a^2)^{1/2}\,dx = \frac{1}{3}(x^2-a^2)^{3/2}.$

3. $\displaystyle\int (x^2-a^2)^{3/2}\,dx = \frac{1}{4}x(x^2-a^2)^{3/2} - \frac{3}{8}a^2x(x^2-a^2)^{1/2} + \frac{3}{8}a^4\ln\bigl|x+(x^2-a^2)^{1/2}\bigr|.$

4. $\displaystyle\int \frac{1}{x}(x^2-a^2)^{1/2}\,dx = (x^2-a^2)^{1/2} - a\arccos\left|\frac{a}{x}\right|.$

5. $\displaystyle\int \frac{dx}{\sqrt{x^2-a^2}} = \ln\bigl|x+(x^2-a^2)^{1/2}\bigr|.$

6. $\displaystyle\int \frac{x\,dx}{\sqrt{x^2-a^2}} = (x^2-a^2)^{1/2}.$

7. $\displaystyle\int (x^2-a^2)^{-3/2}\,dx = -a^{-2}x(x^2-a^2)^{-1/2}.$

► **Integrals involving $(a^2-x^2)^{1/2}$.**

1. $\displaystyle\int (a^2-x^2)^{1/2}\,dx = \frac{1}{2}x(a^2-x^2)^{1/2} + \frac{a^2}{2}\arcsin\frac{x}{a}.$

2. $\displaystyle\int x(a^2-x^2)^{1/2}\,dx = -\frac{1}{3}(a^2-x^2)^{3/2}.$

3. $\displaystyle\int (a^2-x^2)^{3/2}\,dx = \frac{1}{4}x(a^2-x^2)^{3/2} + \frac{3}{8}a^2x(a^2-x^2)^{1/2} + \frac{3}{8}a^4\arcsin\frac{x}{a}.$

4. $\displaystyle\int \frac{1}{x}(a^2-x^2)^{1/2}\,dx = (a^2-x^2)^{1/2} - a\ln\left|\frac{a+(a^2-x^2)^{1/2}}{x}\right|.$

5. $\displaystyle\int \frac{dx}{\sqrt{a^2-x^2}} = \arcsin\frac{x}{a}.$

6. $\displaystyle\int \frac{x\,dx}{\sqrt{a^2-x^2}} = -(a^2-x^2)^{1/2}.$

7. $\displaystyle\int (a^2-x^2)^{-3/2}\,dx = a^{-2}x(a^2-x^2)^{-1/2}.$

► **Integrals involving arbitrary powers. Reduction formulas.**

1. $$\int \frac{dx}{x(ax^n+b)} = \frac{1}{bn}\ln\left|\frac{x^n}{ax^n+b}\right|.$$

2. $$\int \frac{dx}{x\sqrt{x^n+a^2}} = \frac{2}{an}\ln\left|\frac{x^{n/2}}{\sqrt{x^n+a^2}+a}\right|.$$

3. $$\int \frac{dx}{x\sqrt{x^n-a^2}} = \frac{2}{an}\arccos\left|\frac{a}{x^{n/2}}\right|.$$

4. $$\int \frac{dx}{x\sqrt{ax^{2n}+bx^n}} = -\frac{2\sqrt{ax^{2n}+bx^n}}{bnx^n}.$$

The parameters a, b, p, m, and n below in Integrals 5–8 can assume arbitrary values, except for those at which denominators vanish in successive applications of a formula. Notation: $w = ax^n + b$.

5. $$\int x^m(ax^n+b)^p\,dx = \frac{1}{m+np+1}\Bigl(x^{m+1}w^p + npb\int x^m w^{p-1}\,dx\Bigr).$$

6. $$\int x^m(ax^n+b)^p\,dx = \frac{1}{bn(p+1)}\Bigl[-x^{m+1}w^{p+1} + (m+n+np+1)\int x^m w^{p+1}\,dx\Bigr].$$

7. $$\int x^m(ax^n+b)^p\,dx = \frac{1}{b(m+1)}\Bigl[x^{m+1}w^{p+1} - a(m+n+np+1)\int x^{m+n} w^p\,dx\Bigr].$$

8. $$\int x^m(ax^n+b)^p\,dx = \frac{1}{a(m+np+1)}\Bigl[x^{m-n+1}w^{p+1} - b(m-n+1)\int x^{m-n} w^p\,dx\Bigr].$$

S1.1.3. Integrals Involving Exponential Functions

1. $$\int e^{ax}\,dx = \frac{1}{a}e^{ax}.$$

2. $$\int a^x\,dx = \frac{a^x}{\ln a}.$$

3. $$\int xe^{ax}\,dx = e^{ax}\Bigl(\frac{x}{a} - \frac{1}{a^2}\Bigr).$$

4. $$\int x^2e^{ax}\,dx = e^{ax}\Bigl(\frac{x^2}{a} - \frac{2x}{a^2} + \frac{2}{a^3}\Bigr).$$

5. $$\int x^ne^{ax}\,dx = e^{ax}\Bigl[\frac{1}{a}x^n - \frac{n}{a^2}x^{n-1} + \frac{n(n-1)}{a^3}x^{n-2} - \cdots + (-1)^{n-1}\frac{n!}{a^n}x + (-1)^n\frac{n!}{a^{n+1}}\Bigr],$$
$n = 1, 2, \ldots$

6. $$\int P_n(x)e^{ax}\,dx = e^{ax}\sum_{k=0}^{n}\frac{(-1)^k}{a^{k+1}}\frac{d^k}{dx^k}P_n(x),$$ where $P_n(x)$ is an arbitrary polynomial of degree n.

7. $$\int \frac{dx}{a+be^{px}} = \frac{x}{a} - \frac{1}{ap}\ln|a+be^{px}|.$$

8. $$\int \frac{dx}{ae^{px}+be^{-px}} = \begin{cases} \dfrac{1}{p\sqrt{ab}} \arctan\left(e^{px}\sqrt{\dfrac{a}{b}}\right) & \text{if } ab>0, \\ \dfrac{1}{2p\sqrt{-ab}} \ln\left(\dfrac{b+e^{px}\sqrt{-ab}}{b-e^{px}\sqrt{-ab}}\right) & \text{if } ab<0. \end{cases}$$

9. $$\int \frac{dx}{\sqrt{a+be^{px}}} = \begin{cases} \dfrac{1}{p\sqrt{a}} \ln \dfrac{\sqrt{a+be^{px}}-\sqrt{a}}{\sqrt{a+be^{px}}+\sqrt{a}} & \text{if } a>0, \\ \dfrac{2}{p\sqrt{-a}} \arctan \dfrac{\sqrt{a+be^{px}}}{\sqrt{-a}} & \text{if } a<0. \end{cases}$$

S1.1.4. Integrals Involving Hyperbolic Functions

► **Integrals involving $\cosh x$.**

1. $$\int \cosh(a+bx)\,dx = \frac{1}{b}\sinh(a+bx).$$

2. $$\int x\cosh x\,dx = x\sinh x - \cosh x.$$

3. $$\int x^2\cosh x\,dx = (x^2+2)\sinh x - 2x\cosh x.$$

4. $$\int x^{2n}\cosh x\,dx = (2n)!\sum_{k=1}^{n}\left[\frac{x^{2k}}{(2k)!}\sinh x - \frac{x^{2k-1}}{(2k-1)!}\cosh x\right].$$

5. $$\int x^{2n+1}\cosh x\,dx = (2n+1)!\sum_{k=0}^{n}\left[\frac{x^{2k+1}}{(2k+1)!}\sinh x - \frac{x^{2k}}{(2k)!}\cosh x\right].$$

6. $$\int x^p\cosh x\,dx = x^p\sinh x - px^{p-1}\cosh x + p(p-1)\int x^{p-2}\cosh x\,dx.$$

7. $$\int \cosh^2 x\,dx = \tfrac{1}{2}x + \tfrac{1}{4}\sinh 2x.$$

8. $$\int \cosh^3 x\,dx = \sinh x + \tfrac{1}{3}\sinh^3 x.$$

9. $$\int \cosh^{2n} x\,dx = C_{2n}^{n}\frac{x}{2^{2n}} + \frac{1}{2^{2n-1}}\sum_{k=0}^{n-1} C_{2n}^{k}\frac{\sinh[2(n-k)x]}{2(n-k)}, \qquad n=1,\,2,\,\dots$$

10. $$\int \cosh^{2n+1} x\,dx = \frac{1}{2^{2n}}\sum_{k=0}^{n} C_{2n+1}^{k}\frac{\sinh[(2n-2k+1)x]}{2n-2k+1} = \sum_{k=0}^{n} C_{n}^{k}\frac{\sinh^{2k+1} x}{2k+1},$$
$n=1,\,2,\dots$

11. $$\int \cosh^p x\,dx = \frac{1}{p}\sinh x\cosh^{p-1} x + \frac{p-1}{p}\int \cosh^{p-2} x\,dx.$$

12. $$\int \cosh ax\cosh bx\,dx = \frac{1}{a^2-b^2}(a\cosh bx\sinh ax - b\cosh ax\sinh bx).$$

13. $$\int \frac{dx}{\cosh ax} = \frac{2}{a}\arctan\left(e^{ax}\right).$$

14. $$\begin{aligned}\int \frac{dx}{\cosh^{2n} x} &= \frac{\sinh x}{2n-1}\left[\frac{1}{\cosh^{2n-1} x}\right.\\ &\left.+\sum_{k=1}^{n-1}\frac{2^k(n-1)(n-2)\dots(n-k)}{(2n-3)(2n-5)\dots(2n-2k-1)}\frac{1}{\cosh^{2n-2k-1} x}\right], \quad n=1,\,2,\,\dots\end{aligned}$$

15. $$\begin{aligned}\int \frac{dx}{\cosh^{2n+1} x} &= \frac{\sinh x}{2n}\left[\frac{1}{\cosh^{2n} x}\right.\\ &\left.+\sum_{k=1}^{n-1}\frac{(2n-1)(2n-3)\dots(2n-2k+1)}{2^k(n-1)(n-2)\dots(n-k)}\frac{1}{\cosh^{2n-2k} x}\right]+\frac{(2n-1)!!}{(2n)!!}\arctan\sinh x,\end{aligned}$$

$n = 1,\,2,\,\dots$

16. $$\int \frac{dx}{a+b\cosh x} = \begin{cases} -\dfrac{\operatorname{sign} x}{\sqrt{b^2-a^2}}\arcsin\dfrac{b+a\cosh x}{a+b\cosh x} & \text{if } a^2<b^2,\\[2ex] \dfrac{1}{\sqrt{a^2-b^2}}\ln\dfrac{a+b+\sqrt{a^2-b^2}\,\tanh(x/2)}{a+b-\sqrt{a^2-b^2}\,\tanh(x/2)} & \text{if } a^2>b^2.\end{cases}$$

▶ Integrals involving $\sinh x$.

1. $$\int \sinh(a+bx)\,dx = \frac{1}{b}\cosh(a+bx).$$

2. $$\int x\sinh x\,dx = x\cosh x-\sinh x.$$

3. $$\int x^2\sinh x\,dx = (x^2+2)\cosh x-2x\sinh x.$$

4. $$\int x^{2n}\sinh x\,dx = (2n)!\left[\sum_{k=0}^{n}\frac{x^{2k}}{(2k)!}\cosh x-\sum_{k=1}^{n}\frac{x^{2k-1}}{(2k-1)!}\sinh x\right].$$

5. $$\int x^{2n+1}\sinh x\,dx = (2n+1)!\sum_{k=0}^{n}\left[\frac{x^{2k+1}}{(2k+1)!}\cosh x-\frac{x^{2k}}{(2k)!}\sinh x\right].$$

6. $$\int x^p\sinh x\,dx = x^p\cosh x-px^{p-1}\sinh x+p(p-1)\int x^{p-2}\sinh x\,dx.$$

7. $$\int \sinh^2 x\,dx = -\tfrac{1}{2}x+\tfrac{1}{4}\sinh 2x.$$

8. $$\int \sinh^3 x\,dx = -\cosh x+\tfrac{1}{3}\cosh^3 x.$$

9. $$\int \sinh^{2n} x\,dx = (-1)^n C_{2n}^{n}\frac{x}{2^{2n}}+\frac{1}{2^{2n-1}}\sum_{k=0}^{n-1}(-1)^k C_{2n}^{k}\frac{\sinh[2(n-k)x]}{2(n-k)}, \quad n=1,\,2,\,\dots$$

10. $$\begin{aligned}\int \sinh^{2n+1} x\,dx &= \frac{1}{2^{2n}}\sum_{k=0}^{n}(-1)^k C_{2n+1}^{k}\frac{\cosh[(2n-2k+1)x]}{2n-2k+1}\\ &= \sum_{k=0}^{n}(-1)^{n+k}C_n^k\frac{\cosh^{2k+1} x}{2k+1}, \quad n=1,\,2,\,\dots\end{aligned}$$

11. $$\int \sinh^p x\,dx = \frac{1}{p}\sinh^{p-1} x\cosh x - \frac{p-1}{p}\int \sinh^{p-2} x\,dx.$$

12. $$\int \sinh ax \sinh bx\,dx = \frac{1}{a^2-b^2}\bigl(a\cosh ax\sinh bx - b\cosh bx\sinh ax\bigr).$$

13. $$\int \frac{dx}{\sinh ax} = \frac{1}{a}\ln\left|\tanh\frac{ax}{2}\right|.$$

14. $$\int \frac{dx}{\sinh^{2n} x} = \frac{\cosh x}{2n-1}\left[-\frac{1}{\sinh^{2n-1} x} + \sum_{k=1}^{n-1}(-1)^{k-1}\frac{2^k(n-1)(n-2)\dots(n-k)}{(2n-3)(2n-5)\dots(2n-2k-1)}\frac{1}{\sinh^{2n-2k-1} x}\right], \qquad n = 1,\ 2,\ \dots$$

15. $$\int \frac{dx}{\sinh^{2n+1} x} = \frac{\cosh x}{2n}\left[-\frac{1}{\sinh^{2n} x} + \sum_{k=1}^{n-1}(-1)^{k-1}\frac{(2n-1)(2n-3)\dots(2n-2k+1)}{2^k(n-1)(n-2)\dots(n-k)}\frac{1}{\sinh^{2n-2k} x}\right] + (-1)^n\frac{(2n-1)!!}{(2n)!!}\ln\tanh\frac{x}{2},$$

$n = 1,\ 2,\ \dots$

16. $$\int \frac{dx}{a+b\sinh x} = \frac{1}{\sqrt{a^2+b^2}}\ln\frac{a\tanh(x/2)-b+\sqrt{a^2+b^2}}{a\tanh(x/2)-b-\sqrt{a^2+b^2}}.$$

17. $$\int \frac{Ax+B\sinh x}{a+b\sinh x}\,dx = \frac{B}{b}x + \frac{Ab-Ba}{b\sqrt{a^2+b^2}}\ln\frac{a\tanh(x/2)-b+\sqrt{a^2+b^2}}{a\tanh(x/2)-b-\sqrt{a^2+b^2}}.$$

▶ Integrals involving $\tanh x$ or $\coth x$.

1. $$\int \tanh x\,dx = \ln\cosh x.$$

2. $$\int \tanh^2 x\,dx = x - \tanh x.$$

3. $$\int \tanh^3 x\,dx = -\tfrac{1}{2}\tanh^2 x + \ln\cosh x.$$

4. $$\int \tanh^{2n} x\,dx = x - \sum_{k=1}^{n}\frac{\tanh^{2n-2k+1} x}{2n-2k+1}, \qquad n = 1,\ 2,\ \dots$$

5. $$\int \tanh^{2n+1} x\,dx = \ln\cosh x - \sum_{k=1}^{n}\frac{(-1)^k C_n^k}{2k\cosh^{2k} x} = \ln\cosh x - \sum_{k=1}^{n}\frac{\tanh^{2n-2k+2} x}{2n-2k+2},$$

$n = 1,\ 2,\ \dots$

6. $$\int \tanh^p x\,dx = -\frac{1}{p-1}\tanh^{p-1} x + \int \tanh^{p-2} x\,dx.$$

7. $$\int \coth x\,dx = \ln|\sinh x|.$$

8. $$\int \coth^2 x\,dx = x - \coth x.$$

9. $\displaystyle\int \coth^3 x\,dx = -\tfrac{1}{2}\coth^2 x + \ln|\sinh x|.$

10. $\displaystyle\int \coth^{2n} x\,dx = x - \sum_{k=1}^{n} \frac{\coth^{2n-2k+1} x}{2n-2k+1}, \quad n = 1,\ 2,\ \ldots$

11. $\displaystyle\int \coth^{2n+1} x\,dx = \ln|\sinh x| - \sum_{k=1}^{n} \frac{C_n^k}{2k\sinh^{2k} x} = \ln|\sinh x| - \sum_{k=1}^{n} \frac{\coth^{2n-2k+2} x}{2n-2k+2},$
$n = 1,\ 2,\ \ldots$

12. $\displaystyle\int \coth^p x\,dx = -\frac{1}{p-1}\coth^{p-1} x + \int \coth^{p-2} x\,dx.$

S1.1.5. Integrals Involving Logarithmic Functions

1. $\displaystyle\int \ln ax\,dx = x\ln ax - x.$

2. $\displaystyle\int x\ln x\,dx = \tfrac{1}{2}x^2\ln x - \tfrac{1}{4}x^2.$

3. $\displaystyle\int x^p \ln ax\,dx = \begin{cases} \dfrac{1}{p+1}x^{p+1}\ln ax - \dfrac{1}{(p+1)^2}x^{p+1} & \text{if } p \neq -1, \\ \tfrac{1}{2}\ln^2 ax & \text{if } p = -1. \end{cases}$

4. $\displaystyle\int (\ln x)^2\,dx = x(\ln x)^2 - 2x\ln x + 2x.$

5. $\displaystyle\int x(\ln x)^2\,dx = \tfrac{1}{2}x^2(\ln x)^2 - \tfrac{1}{2}x^2\ln x + \tfrac{1}{4}x^2.$

6. $\displaystyle\int x^p(\ln x)^2\,dx = \begin{cases} \dfrac{x^{p+1}}{p+1}(\ln x)^2 - \dfrac{2x^{p+1}}{(p+1)^2}\ln x + \dfrac{2x^{p+1}}{(p+1)^3} & \text{if } p \neq -1, \\ \tfrac{1}{3}\ln^3 x & \text{if } p = -1. \end{cases}$

7. $\displaystyle\int (\ln x)^n\,dx = \frac{x}{n+1}\sum_{k=0}^{n}(-1)^k(n+1)n\ldots(n-k+1)(\ln x)^{n-k}, \quad n = 1,\ 2,\ \ldots$

8. $\displaystyle\int (\ln x)^q\,dx = x(\ln x)^q - q\int (\ln x)^{q-1}\,dx,\ q \neq -1.$

9. $\displaystyle\int x^n(\ln x)^m\,dx = \frac{x^{n+1}}{m+1}\sum_{k=0}^{m}\frac{(-1)^k}{(n+1)^{k+1}}(m+1)m\ldots(m-k+1)(\ln x)^{m-k},$
$n, m = 1,\ 2,\ \ldots$

10. $\displaystyle\int x^p(\ln x)^q\,dx = \frac{1}{p+1}x^{p+1}(\ln x)^q - \frac{q}{p+1}\int x^p(\ln x)^{q-1}\,dx, \quad p,\ q \neq -1.$

11. $\displaystyle\int \ln(a+bx)\,dx = \frac{1}{b}(ax+b)\ln(ax+b) - x.$

12. $$\int x\ln(a+bx)\,dx=\frac{1}{2}\left(x^2-\frac{a^2}{b^2}\right)\ln(a+bx)-\frac{1}{2}\left(\frac{x^2}{2}-\frac{a}{b}x\right).$$

13. $$\int x^2\ln(a+bx)\,dx=\frac{1}{3}\left(x^3-\frac{a^3}{b^3}\right)\ln(a+bx)-\frac{1}{3}\left(\frac{x^3}{3}-\frac{ax^2}{2b}+\frac{a^2x}{b^2}\right).$$

14. $$\int\frac{\ln x\,dx}{(a+bx)^2}=-\frac{\ln x}{b(a+bx)}+\frac{1}{ab}\ln\frac{x}{a+bx}.$$

15. $$\int\frac{\ln x\,dx}{(a+bx)^3}=-\frac{\ln x}{2b(a+bx)^2}+\frac{1}{2ab(a+bx)}+\frac{1}{2a^2b}\ln\frac{x}{a+bx}.$$

16. $$\int\frac{\ln x\,dx}{\sqrt{a+bx}}=\begin{cases}\dfrac{2}{b}\left[(\ln x-2)\sqrt{a+bx}+\sqrt{a}\,\ln\dfrac{\sqrt{a+bx}+\sqrt{a}}{\sqrt{a+bx}-\sqrt{a}}\right] & \text{if } a>0,\\[2ex] \dfrac{2}{b}\left[(\ln x-2)\sqrt{a+bx}+2\sqrt{-a}\,\arctan\dfrac{\sqrt{a+bx}}{\sqrt{-a}}\right] & \text{if } a<0.\end{cases}$$

17. $$\int\ln(x^2+a^2)\,dx=x\ln(x^2+a^2)-2x+2a\arctan(x/a).$$

18. $$\int x\ln(x^2+a^2)\,dx=\tfrac{1}{2}\left[(x^2+a^2)\ln(x^2+a^2)-x^2\right].$$

19. $$\int x^2\ln(x^2+a^2)\,dx=\tfrac{1}{3}\left[x^3\ln(x^2+a^2)-\tfrac{2}{3}x^3+2a^2x-2a^3\arctan(x/a)\right].$$

S1.1.6. Integrals Involving Trigonometric Functions

► **Integrals involving $\cos x$ $(n = 1,\ 2,\ \ldots)$.**

1. $$\int\cos(a+bx)\,dx=\frac{1}{b}\sin(a+bx).$$

2. $$\int x\cos x\,dx=\cos x+x\sin x.$$

3. $$\int x^2\cos x\,dx=2x\cos x+(x^2-2)\sin x.$$

4. $$\int x^{2n}\cos x\,dx=(2n)!\left[\sum_{k=0}^{n}(-1)^k\frac{x^{2n-2k}}{(2n-2k)!}\sin x+\sum_{k=0}^{n-1}(-1)^k\frac{x^{2n-2k-1}}{(2n-2k-1)!}\cos x\right].$$

5. $$\int x^{2n+1}\cos x\,dx=(2n+1)!\sum_{k=0}^{n}\left[(-1)^k\frac{x^{2n-2k+1}}{(2n-2k+1)!}\sin x+\frac{x^{2n-2k}}{(2n-2k)!}\cos x\right].$$

6. $$\int x^p\cos x\,dx=x^p\sin x+px^{p-1}\cos x-p(p-1)\int x^{p-2}\cos x\,dx.$$

7. $$\int\cos^2x\,dx=\tfrac{1}{2}x+\tfrac{1}{4}\sin 2x.$$

8. $$\int\cos^3x\,dx=\sin x-\tfrac{1}{3}\sin^3x.$$

9. $$\int \cos^{2n} x\,dx = \frac{1}{2^{2n}} C_{2n}^{n} x + \frac{1}{2^{2n-1}} \sum_{k=0}^{n-1} C_{2n}^{k} \frac{\sin[(2n-2k)x]}{2n-2k}.$$

10. $$\int \cos^{2n+1} x\,dx = \frac{1}{2^{2n}} \sum_{k=0}^{n} C_{2n+1}^{k} \frac{\sin[(2n-2k+1)x]}{2n-2k+1}.$$

11. $$\int \frac{dx}{\cos x} = \ln\left|\tan\left(\frac{x}{2} + \frac{\pi}{4}\right)\right|.$$

12. $$\int \frac{dx}{\cos^2 x} = \tan x.$$

13. $$\int \frac{dx}{\cos^3 x} = \frac{\sin x}{2\cos^2 x} + \frac{1}{2}\ln\left|\tan\left(\frac{x}{2} + \frac{\pi}{4}\right)\right|.$$

14. $$\int \frac{dx}{\cos^n x} = \frac{\sin x}{(n-1)\cos^{n-1} x} + \frac{n-2}{n-1}\int \frac{dx}{\cos^{n-2} x}, \qquad n > 1.$$

15. $$\int \frac{x\,dx}{\cos^{2n} x} = \sum_{k=0}^{n-1} \frac{(2n-2)(2n-4)\dots(2n-2k+2)}{(2n-1)(2n-3)\dots(2n-2k+3)} \frac{(2n-2k)x\sin x - \cos x}{(2n-2k+1)(2n-2k)\cos^{2n-2k+1} x}$$
$$+ \frac{2^{n-1}(n-1)!}{(2n-1)!!}\bigl(x\tan x + \ln|\cos x|\bigr).$$

16. $$\int \cos ax\,\cos bx\,dx = \frac{\sin\bigl[(b-a)x\bigr]}{2(b-a)} + \frac{\sin\bigl[(b+a)x\bigr]}{2(b+a)}, \qquad a \neq \pm b.$$

17. $$\int \frac{dx}{a + b\cos x} = \begin{cases} \dfrac{2}{\sqrt{a^2-b^2}} \arctan \dfrac{(a-b)\tan(x/2)}{\sqrt{a^2-b^2}} & \text{if } a^2 > b^2, \\ \dfrac{1}{\sqrt{b^2-a^2}} \ln\left|\dfrac{\sqrt{b^2-a^2} + (b-a)\tan(x/2)}{\sqrt{b^2-a^2} - (b-a)\tan(x/2)}\right| & \text{if } b^2 > a^2. \end{cases}$$

18. $$\int \frac{dx}{(a + b\cos x)^2} = \frac{b\sin x}{(b^2-a^2)(a + b\cos x)} - \frac{a}{b^2-a^2}\int \frac{dx}{a + b\cos x}.$$

19. $$\int \frac{dx}{a^2 + b^2\cos^2 x} = \frac{1}{a\sqrt{a^2+b^2}} \arctan \frac{a\tan x}{\sqrt{a^2+b^2}}.$$

20. $$\int \frac{dx}{a^2 - b^2\cos^2 x} = \begin{cases} \dfrac{1}{a\sqrt{a^2-b^2}} \arctan \dfrac{a\tan x}{\sqrt{a^2-b^2}} & \text{if } a^2 > b^2, \\ \dfrac{1}{2a\sqrt{b^2-a^2}} \ln\left|\dfrac{\sqrt{b^2-a^2} - a\tan x}{\sqrt{b^2-a^2} + a\tan x}\right| & \text{if } b^2 > a^2. \end{cases}$$

21. $$\int e^{ax}\cos bx\,dx = e^{ax}\left(\frac{b}{a^2+b^2}\sin bx + \frac{a}{a^2+b^2}\cos bx\right).$$

22. $$\int e^{ax}\cos^2 x\,dx = \frac{e^{ax}}{a^2+4}\left(a\cos^2 x + 2\sin x\cos x + \frac{2}{a}\right).$$

23. $$\int e^{ax}\cos^n x\,dx = \frac{e^{ax}\cos^{n-1} x}{a^2+n^2}(a\cos x + n\sin x) + \frac{n(n-1)}{a^2+n^2}\int e^{ax}\cos^{n-2} x\,dx.$$

▶ Integrals involving $\sin x$ $(n = 1,\ 2,\ \ldots)$.

1. $$\int \sin(a+bx)\,dx = -\frac{1}{b}\cos(a+bx).$$

2. $$\int x\sin x\,dx = \sin x - x\cos x.$$

3. $$\int x^2\sin x\,dx = 2x\sin x - (x^2-2)\cos x.$$

4. $$\int x^3\sin x\,dx = (3x^2-6)\sin x - (x^3-6x)\cos x.$$

5. $$\int x^{2n}\sin x\,dx = (2n)!\left[\sum_{k=0}^{n}(-1)^{k+1}\frac{x^{2n-2k}}{(2n-2k)!}\cos x + \sum_{k=0}^{n-1}(-1)^{k}\frac{x^{2n-2k-1}}{(2n-2k-1)!}\sin x\right].$$

6. $$\int x^{2n+1}\sin x\,dx = (2n+1)!\sum_{k=0}^{n}\left[(-1)^{k+1}\frac{x^{2n-2k+1}}{(2n-2k+1)!}\cos x + (-1)^{k}\frac{x^{2n-2k}}{(2n-2k)!}\sin x\right].$$

7. $$\int x^p\sin x\,dx = -x^p\cos x + px^{p-1}\sin x - p(p-1)\int x^{p-2}\sin x\,dx.$$

8. $$\int \sin^2 x\,dx = \tfrac{1}{2}x - \tfrac{1}{4}\sin 2x.$$

9. $$\int x\sin^2 x\,dx = \tfrac{1}{4}x^2 - \tfrac{1}{4}x\sin 2x - \tfrac{1}{8}\cos 2x.$$

10. $$\int \sin^3 x\,dx = -\cos x + \tfrac{1}{3}\cos^3 x.$$

11. $$\int \sin^{2n} x\,dx = \frac{1}{2^{2n}}C_{2n}^{n}x + \frac{(-1)^n}{2^{2n-1}}\sum_{k=0}^{n-1}(-1)^k C_{2n}^{k}\frac{\sin[(2n-2k)x]}{2n-2k},$$

where $C_m^k = \dfrac{m!}{k!\,(m-k)!}$ are binomial coefficients $(0! = 1)$.

12. $$\int \sin^{2n+1} x\,dx = \frac{1}{2^{2n}}\sum_{k=0}^{n}(-1)^{n+k+1}C_{2n+1}^{k}\frac{\cos[(2n-2k+1)x]}{2n-2k+1}.$$

13. $$\int \frac{dx}{\sin x} = \ln\left|\tan\frac{x}{2}\right|.$$

14. $$\int \frac{dx}{\sin^2 x} = -\cot x.$$

15. $$\int \frac{dx}{\sin^3 x} = -\frac{\cos x}{2\sin^2 x} + \frac{1}{2}\ln\left|\tan\frac{x}{2}\right|.$$

16. $$\int \frac{dx}{\sin^n x} = -\frac{\cos x}{(n-1)\sin^{n-1} x} + \frac{n-2}{n-1}\int\frac{dx}{\sin^{n-2} x}, \qquad n > 1.$$

17. $$\int \frac{x\,dx}{\sin^{2n} x} = -\sum_{k=0}^{n-1}\frac{(2n-2)(2n-4)\ldots(2n-2k+2)}{(2n-1)(2n-3)\ldots(2n-2k+3)}\,\frac{\sin x + (2n-2k)x\cos x}{(2n-2k+1)(2n-2k)\sin^{2n-2k+1} x}$$
$$+\frac{2^{n-1}(n-1)!}{(2n-1)!!}\bigl(\ln|\sin x| - x\cot x\bigr).$$

18. $$\int \sin ax\,\sin bx\,dx = \frac{\sin[(b-a)x]}{2(b-a)} - \frac{\sin[(b+a)x]}{2(b+a)}, \qquad a \neq \pm b.$$

19. $$\int \frac{dx}{a+b\sin x} = \begin{cases} \dfrac{2}{\sqrt{a^2-b^2}}\arctan\dfrac{b+a\tan x/2}{\sqrt{a^2-b^2}} & \text{if } a^2 > b^2, \\ \dfrac{1}{\sqrt{b^2-a^2}}\ln\left|\dfrac{b-\sqrt{b^2-a^2}+a\tan x/2}{b+\sqrt{b^2-a^2}+a\tan x/2}\right| & \text{if } b^2 > a^2. \end{cases}$$

20. $$\int \frac{dx}{(a+b\sin x)^2} = \frac{b\cos x}{(a^2-b^2)(a+b\sin x)} + \frac{a}{a^2-b^2}\int \frac{dx}{a+b\sin x}.$$

21. $$\int \frac{dx}{a^2+b^2\sin^2 x} = \frac{1}{a\sqrt{a^2+b^2}}\arctan\frac{\sqrt{a^2+b^2}\,\tan x}{a}.$$

22. $$\int \frac{dx}{a^2-b^2\sin^2 x} = \begin{cases} \dfrac{1}{a\sqrt{a^2-b^2}}\arctan\dfrac{\sqrt{a^2-b^2}\,\tan x}{a} & \text{if } a^2 > b^2, \\ \dfrac{1}{2a\sqrt{b^2-a^2}}\ln\left|\dfrac{\sqrt{b^2-a^2}\,\tan x+a}{\sqrt{b^2-a^2}\,\tan x-a}\right| & \text{if } b^2 > a^2. \end{cases}$$

23. $$\int \frac{\sin x\,dx}{\sqrt{1+k^2\sin^2 x}} = -\frac{1}{k}\arcsin\frac{k\cos x}{\sqrt{1+k^2}}.$$

24. $$\int \frac{\sin x\,dx}{\sqrt{1-k^2\sin^2 x}} = -\frac{1}{k}\ln\left|k\cos x+\sqrt{1-k^2\sin^2 x}\,\right|.$$

25. $$\int \sin x\sqrt{1+k^2\sin^2 x}\,dx = -\frac{\cos x}{2}\sqrt{1+k^2\sin^2 x} - \frac{1+k^2}{2k}\arcsin\frac{k\cos x}{\sqrt{1+k^2}}.$$

26. $$\int \sin x\sqrt{1-k^2\sin^2 x}\,dx$$
$$= -\frac{\cos x}{2}\sqrt{1-k^2\sin^2 x} - \frac{1-k^2}{2k}\ln\left|k\cos x+\sqrt{1-k^2\sin^2 x}\,\right|.$$

27. $$\int e^{ax}\sin bx\,dx = e^{ax}\left(\frac{a}{a^2+b^2}\sin bx - \frac{b}{a^2+b^2}\cos bx\right).$$

28. $$\int e^{ax}\sin^2 x\,dx = \frac{e^{ax}}{a^2+4}\left(a\sin^2 x - 2\sin x\cos x + \frac{2}{a}\right).$$

29. $$\int e^{ax}\sin^n x\,dx = \frac{e^{ax}\sin^{n-1} x}{a^2+n^2}(a\sin x - n\cos x) + \frac{n(n-1)}{a^2+n^2}\int e^{ax}\sin^{n-2} x\,dx.$$

► Integrals involving $\sin x$ and $\cos x$.

1. $$\int \sin ax\cos bx\,dx = -\frac{\cos[(a+b)x]}{2(a+b)} - \frac{\cos\big[(a-b)x\big]}{2(a-b)}, \qquad a \neq \pm b.$$

2. $$\int \frac{dx}{b^2\cos^2 ax + c^2\sin^2 ax} = \frac{1}{abc}\arctan\left(\frac{c}{b}\tan ax\right).$$

3. $$\int \frac{dx}{b^2\cos^2 ax - c^2\sin^2 ax} = \frac{1}{2abc}\ln\left|\frac{c\tan ax+b}{c\tan ax-b}\right|.$$

4. $$\int \frac{dx}{\cos^{2n} x\sin^{2m} x} = \sum_{k=0}^{n+m-1} C_{n+m-1}^k\,\frac{\tan^{2k-2m+1} x}{2k-2m+1}, \qquad n, m = 1,\ 2,\ \ldots$$

5. $$\int \frac{dx}{\cos^{2n+1} x \sin^{2m+1} x} = C^m_{n+m} \ln|\tan x| + \sum_{k=0}^{n+m} C^k_{n+m} \frac{\tan^{2k-2m} x}{2k-2m}, \qquad n, m = 1,\ 2,\ \ldots$$

► Reduction formulas.

The parameters p and q below can assume any values, except for those at which the denominators on the right-hand side vanish.

1. $$\int \sin^p x \cos^q x\, dx = -\frac{\sin^{p-1} x \cos^{q+1} x}{p+q} + \frac{p-1}{p+q} \int \sin^{p-2} x \cos^q x\, dx.$$

2. $$\int \sin^p x \cos^q x\, dx = \frac{\sin^{p+1} x \cos^{q-1} x}{p+q} + \frac{q-1}{p+q} \int \sin^p x \cos^{q-2} x\, dx.$$

3. $$\int \sin^p x \cos^q x\, dx = \frac{\sin^{p-1} x \cos^{q-1} x}{p+q} \left(\sin^2 x - \frac{q-1}{p+q-2}\right) + \frac{(p-1)(q-1)}{(p+q)(p+q-2)} \int \sin^{p-2} x \cos^{q-2} x\, dx.$$

4. $$\int \sin^p x \cos^q x\, dx = \frac{\sin^{p+1} x \cos^{q+1} x}{p+1} + \frac{p+q+2}{p+1} \int \sin^{p+2} x \cos^q x\, dx.$$

5. $$\int \sin^p x \cos^q x\, dx = -\frac{\sin^{p+1} x \cos^{q+1} x}{q+1} + \frac{p+q+2}{q+1} \int \sin^p x \cos^{q+2} x\, dx.$$

6. $$\int \sin^p x \cos^q x\, dx = -\frac{\sin^{p-1} x \cos^{q+1} x}{q+1} + \frac{p-1}{q+1} \int \sin^{p-2} x \cos^{q+2} x\, dx.$$

7. $$\int \sin^p x \cos^q x\, dx = \frac{\sin^{p+1} x \cos^{q-1} x}{p+1} + \frac{q-1}{p+1} \int \sin^{p+2} x \cos^{q-2} x\, dx.$$

► Integrals involving $\tan x$ and $\cot x$.

1. $$\int \tan x\, dx = -\ln|\cos x|.$$

2. $$\int \tan^2 x\, dx = \tan x - x.$$

3. $$\int \tan^3 x\, dx = \tfrac{1}{2} \tan^2 x + \ln|\cos x|.$$

4. $$\int \tan^{2n} x\, dx = (-1)^n x - \sum_{k=1}^{n} \frac{(-1)^k (\tan x)^{2n-2k+1}}{2n-2k+1}, \qquad n = 1,\ 2,\ \ldots$$

5. $$\int \tan^{2n+1} x\, dx = (-1)^{n+1} \ln|\cos x| - \sum_{k=1}^{n} \frac{(-1)^k (\tan x)^{2n-2k+2}}{2n-2k+2}, \qquad n = 1,\ 2,\ \ldots$$

6. $$\int \frac{dx}{a + b\tan x} = \frac{1}{a^2+b^2} \bigl(ax + b\ln|a\cos x + b\sin x|\bigr).$$

7. $$\int \frac{\tan x\, dx}{\sqrt{a + b\tan^2 x}} = \frac{1}{\sqrt{b-a}} \arccos\left(\sqrt{1 - \frac{a}{b}}\, \cos x\right), \qquad b > a,\ b > 0.$$

8. $$\int \cot x\, dx = \ln|\sin x|.$$

9. $\displaystyle\int \cot^2 x\,dx = -\cot x - x.$

10. $\displaystyle\int \cot^3 x\,dx = -\tfrac{1}{2}\cot^2 x - \ln|\sin x|.$

11. $\displaystyle\int \cot^{2n} x\,dx = (-1)^n x + \sum_{k=1}^{n} \frac{(-1)^k (\cot x)^{2n-2k+1}}{2n-2k+1}, \qquad n = 1,\ 2,\ \ldots$

12. $\displaystyle\int \cot^{2n+1} x\,dx = (-1)^n \ln|\sin x| + \sum_{k=1}^{n} \frac{(-1)^k (\cot x)^{2n-2k+2}}{2n-2k+2}, \qquad n = 1,\ 2,\ \ldots$

13. $\displaystyle\int \frac{dx}{a + b\cot x} = \frac{1}{a^2+b^2}\bigl(ax - b\ln|a\sin x + b\cos x|\bigr).$

S1.1.7. Integrals Involving Inverse Trigonometric Functions

1. $\displaystyle\int \arcsin\frac{x}{a}\,dx = x\arcsin\frac{x}{a} + \sqrt{a^2-x^2}.$

2. $\displaystyle\int \Bigl(\arcsin\frac{x}{a}\Bigr)^2 dx = x\Bigl(\arcsin\frac{x}{a}\Bigr)^2 - 2x + 2\sqrt{a^2-x^2}\,\arcsin\frac{x}{a}.$

3. $\displaystyle\int x\arcsin\frac{x}{a}\,dx = \frac{1}{4}(2x^2-a^2)\arcsin\frac{x}{a} + \frac{x}{4}\sqrt{a^2-x^2}.$

4. $\displaystyle\int x^2\arcsin\frac{x}{a}\,dx = \frac{x^3}{3}\arcsin\frac{x}{a} + \frac{1}{9}(x^2+2a^2)\sqrt{a^2-x^2}.$

5. $\displaystyle\int \arccos\frac{x}{a}\,dx = x\arccos\frac{x}{a} - \sqrt{a^2-x^2}.$

6. $\displaystyle\int \Bigl(\arccos\frac{x}{a}\Bigr)^2 dx = x\Bigl(\arccos\frac{x}{a}\Bigr)^2 - 2x - 2\sqrt{a^2-x^2}\,\arccos\frac{x}{a}.$

7. $\displaystyle\int x\arccos\frac{x}{a}\,dx = \frac{1}{4}(2x^2-a^2)\arccos\frac{x}{a} - \frac{x}{4}\sqrt{a^2-x^2}.$

8. $\displaystyle\int x^2\arccos\frac{x}{a}\,dx = \frac{x^3}{3}\arccos\frac{x}{a} - \frac{1}{9}(x^2+2a^2)\sqrt{a^2-x^2}.$

9. $\displaystyle\int \arctan\frac{x}{a}\,dx = x\arctan\frac{x}{a} - \frac{a}{2}\ln(a^2+x^2).$

10. $\displaystyle\int x\arctan\frac{x}{a}\,dx = \frac{1}{2}(x^2+a^2)\arctan\frac{x}{a} - \frac{ax}{2}.$

11. $\displaystyle\int x^2\arctan\frac{x}{a}\,dx = \frac{x^3}{3}\arctan\frac{x}{a} - \frac{ax^2}{6} + \frac{a^3}{6}\ln(a^2+x^2).$

12. $\displaystyle\int \operatorname{arccot}\frac{x}{a}\,dx = x\operatorname{arccot}\frac{x}{a} + \frac{a}{2}\ln(a^2+x^2).$

13. $\displaystyle\int x\operatorname{arccot}\frac{x}{a}\,dx = \frac{1}{2}(x^2+a^2)\operatorname{arccot}\frac{x}{a} + \frac{ax}{2}.$

14. $\displaystyle\int x^2\operatorname{arccot}\frac{x}{a}\,dx = \frac{x^3}{3}\operatorname{arccot}\frac{x}{a} + \frac{ax^2}{6} - \frac{a^3}{6}\ln(a^2+x^2).$

S1.2. Tables of Definite Integrals

Throughout Section S1.2 it is assumed that n is a positive integer, unless otherwise specified.

S1.2.1. Integrals Involving Power-Law Functions

► **Integrals over a finite interval.**

1. $$\int_0^1 \frac{x^n\,dx}{x+1} = (-1)^n \left[\ln 2 + \sum_{k=1}^{n} \frac{(-1)^k}{k}\right].$$

2. $$\int_0^1 \frac{dx}{x^2 + 2x\cos\beta + 1} = \frac{\beta}{2\sin\beta}.$$

3. $$\int_0^1 \frac{\left(x^a + x^{-a}\right)dx}{x^2 + 2x\cos\beta + 1} = \frac{\pi\sin(a\beta)}{\sin(\pi a)\sin\beta}, \quad |a| < 1,\ \beta \neq (2n+1)\pi.$$

4. $$\int_0^1 x^a(1-x)^{1-a}\,dx = \frac{\pi a(1-a)}{2\sin(\pi a)}, \quad -1 < a < 1.$$

5. $$\int_0^1 \frac{dx}{x^a(1-x)^{1-a}} = \frac{\pi}{\sin(\pi a)}, \quad 0 < a < 1.$$

6. $$\int_0^1 \frac{x^a\,dx}{(1-x)^a} = \frac{\pi a}{\sin(\pi a)}, \quad -1 < a < 1.$$

7. $$\int_0^1 x^{p-1}(1-x)^{q-1}\,dx \equiv B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}, \quad p, q > 0.$$

8. $$\int_0^1 x^{p-1}(1-x^q)^{-p/q}\,dx = \frac{\pi}{q\sin(\pi p/q)}, \quad q > p > 0.$$

9. $$\int_0^1 x^{p+q-1}(1-x^q)^{-p/q}\,dx = \frac{\pi p}{q^2\sin(\pi p/q)}, \quad q > p.$$

10. $$\int_0^1 x^{q/p-1}(1-x^q)^{-1/p}\,dx = \frac{\pi}{q\sin(\pi/p)}, \quad p > 1,\ q > 0.$$

11. $$\int_0^1 \frac{x^{p-1} - x^{-p}}{1-x}\,dx = \pi\cot(\pi p), \quad |p| < 1.$$

12. $$\int_0^1 \frac{x^{p-1} - x^{-p}}{1+x}\,dx = \frac{\pi}{\sin(\pi p)}, \quad |p| < 1.$$

13. $$\int_0^1 \frac{x^p - x^{-p}}{x-1}\,dx = \frac{1}{p} - \pi\cot(\pi p), \quad |p| < 1.$$

14. $$\int_0^1 \frac{x^p - x^{-p}}{1+x}\,dx = \frac{1}{p} - \frac{\pi}{\sin(\pi p)}, \quad |p| < 1.$$

15. $$\int_0^1 \frac{x^{1+p} - x^{1-p}}{1-x^2}\,dx = \frac{\pi}{2}\cot\left(\frac{\pi p}{2}\right) - \frac{1}{p}, \quad |p| < 1.$$

16. $$\int_0^1 \frac{x^{1+p}-x^{1-p}}{1+x^2}\,dx = \frac{1}{p} - \frac{\pi}{2\sin(\pi p/2)}, \quad |p| < 1.$$

17. $$\int_0^1 \frac{dx}{\sqrt{(1+a^2x)(1-x)}} = \frac{2}{a}\arctan a.$$

18. $$\int_0^1 \frac{dx}{\sqrt{(1-a^2x)(1-x)}} = \frac{1}{a}\ln\frac{1+a}{1-a}.$$

19. $$\int_{-1}^1 \frac{dx}{(a-x)\sqrt{1-x^2}} = \frac{\pi}{\sqrt{a^2-1}}, \quad 1 < a.$$

20. $$\int_0^1 \frac{x^n\,dx}{\sqrt{1-x}} = \frac{2\,(2n)!!}{(2n+1)!!}, \quad n = 1,\ 2,\ \ldots$$

21. $$\int_0^1 \frac{x^{n-1/2}\,dx}{\sqrt{1-x}} = \frac{\pi\,(2n-1)!!}{(2n)!!}, \quad n = 1,\ 2,\ \ldots$$

22. $$\int_0^1 \frac{x^{2n}\,dx}{\sqrt{1-x^2}} = \frac{\pi}{2}\,\frac{1\times 3\times\ldots\times(2n-1)}{2\times 4\times\ldots\times(2n)}, \quad n = 1,\ 2,\ \ldots$$

23. $$\int_0^1 \frac{x^{2n+1}\,dx}{\sqrt{1-x^2}} = \frac{2\times 4\times\ldots\times(2n)}{1\times 3\times\ldots\times(2n+1)}, \quad n = 1,\ 2,\ \ldots$$

24. $$\int_0^1 \frac{x^{\lambda-1}\,dx}{(1+ax)(1-x)^\lambda} = \frac{\pi}{(1+a)^\lambda\sin(\pi\lambda)}, \quad 0 < \lambda < 1, \quad a > -1.$$

25. $$\int_0^1 \frac{x^{\lambda-1/2}\,dx}{(1+ax)^\lambda(1-x)^\lambda} = 2\pi^{-1/2}\Gamma\bigl(\lambda+\tfrac{1}{2}\bigr)\Gamma\bigl(1-\lambda\bigr)\cos^{2\lambda}k\,\frac{\sin[(2\lambda-1)k]}{(2\lambda-1)\sin k},$$
$$k = \arctan\sqrt{a}, \quad -\tfrac{1}{2} < \lambda < 1, \quad a > 0.$$

► **Integrals over an infinite interval.**

1. $$\int_0^\infty \frac{dx}{ax^2+b} = \frac{\pi}{2\sqrt{ab}}.$$

2. $$\int_0^\infty \frac{dx}{x^4+1} = \frac{\pi\sqrt{2}}{4}.$$

3. $$\int_0^\infty \frac{x^{a-1}\,dx}{x+1} = \frac{\pi}{\sin(\pi a)}, \quad 0 < a < 1.$$

4. $$\int_0^\infty \frac{x^{\lambda-1}\,dx}{(1+ax)^2} = \frac{\pi(1-\lambda)}{a^\lambda\sin(\pi\lambda)}, \quad 0 < \lambda < 2.$$

5. $$\int_0^\infty \frac{x^{\lambda-1}\,dx}{(x+a)(x+b)} = \frac{\pi(a^{\lambda-1}-b^{\lambda-1})}{(b-a)\sin(\pi\lambda)}, \quad 0 < \lambda < 2.$$

6. $$\int_0^\infty \frac{x^{\lambda-1}(x+c)\,dx}{(x+a)(x+b)} = \frac{\pi}{\sin(\pi\lambda)}\left(\frac{a-c}{a-b}a^{\lambda-1} + \frac{b-c}{b-a}b^{\lambda-1}\right), \quad 0 < \lambda < 1.$$

7. $$\int_0^\infty \frac{x^\lambda\,dx}{(x+1)^3} = \frac{\pi\lambda(1-\lambda)}{2\sin(\pi\lambda)}, \quad -1 < \lambda < 2.$$

8. $$\int_0^\infty \frac{x^{\lambda-1}\,dx}{(x^2+a^2)(x^2+b^2)} = \frac{\pi\left(b^{\lambda-2}-a^{\lambda-2}\right)}{2\left(a^2-b^2\right)\sin(\pi\lambda/2)}, \quad 0<\lambda<4.$$

9. $$\int_0^\infty \frac{x^{p-1}-x^{q-1}}{1-x}\,dx = \pi[\cot(\pi p)-\cot(\pi q)], \quad p,q>0.$$

10. $$\int_0^\infty \frac{x^{\lambda-1}\,dx}{(1+ax)^{n+1}} = (-1)^n \frac{\pi C^n_{\lambda-1}}{a^\lambda \sin(\pi\lambda)}, \quad 0<\lambda<n+1, \quad C^n_{\lambda-1} = \frac{(\lambda-1)(\lambda-2)\ldots(\lambda-n)}{n!}.$$

11. $$\int_0^\infty \frac{x^m\,dx}{(a+bx)^{n+1/2}} = 2^{m+1} m!\,\frac{(2n-2m-3)!!}{(2n-1)!!}\,\frac{a^{m-n+1/2}}{b^{m+1}},$$

$a,b>0, \quad n,m=1,\ 2,\ \ldots, \quad m<b-\frac{1}{2}.$

12. $$\int_0^\infty \frac{dx}{(x^2+a^2)^n} = \frac{\pi}{2}\,\frac{(2n-3)!!}{(2n-2)!!}\,\frac{1}{a^{2n-1}}, \quad n=1,\ 2,\ \ldots$$

13. $$\int_0^\infty \frac{(x+1)^{\lambda-1}}{(x+a)^{\lambda+1}}\,dx = \frac{1-a^{-\lambda}}{\lambda(a-1)}, \quad a>0.$$

14. $$\int_0^\infty \frac{x^{a-1}\,dx}{x^b+1} = \frac{\pi}{b\sin(\pi a/b)}, \quad 0<a\le b.$$

15. $$\int_0^\infty \frac{x^{a-1}\,dx}{(x^b+1)^2} = \frac{\pi(a-b)}{b^2\sin[\pi(a-b)/b]}, \quad a<2b.$$

16. $$\int_0^\infty \frac{x^{\lambda-1/2}\,dx}{(x+a)^\lambda(x+b)^\lambda} = \sqrt{\pi}\left(\sqrt{a}+\sqrt{b}\right)^{1-2\lambda}\frac{\Gamma(\lambda-1/2)}{\Gamma(\lambda)}, \quad \lambda>0.$$

17. $$\int_0^\infty \frac{1-x^a}{1-x^b}\,x^{c-1}\,dx = \frac{\pi\sin A}{b\sin C\sin(A+C)}, \quad A=\frac{\pi a}{b}, \quad C=\frac{\pi c}{b}; \quad a+c<b, \quad c>0.$$

18. $$\int_0^\infty \frac{x^{a-1}\,dx}{(1+x^2)^{1-b}} = \tfrac{1}{2}B\left(\tfrac{1}{2}a, 1-b-\tfrac{1}{2}a\right), \quad \tfrac{1}{2}a+b<1, \quad a>0.$$

19. $$\int_0^\infty \frac{x^{2m}\,dx}{(ax^2+b)^n} = \frac{\pi(2m-1)!!\,(2n-2m-3)!!}{2\,(2n-2)!!\,a^m b^{n-m-1}\sqrt{ab}}, \quad a,b>0, \quad n>m+1.$$

20. $$\int_0^\infty \frac{x^{2m+1}\,dx}{(ax^2+b)^n} = \frac{m!\,(n-m-2)!}{2(n-1)!a^{m+1}b^{n-m-1}}, \quad ab>0, \quad n>m+1\ge 1.$$

21. $$\int_0^\infty \frac{x^{\mu-1}\,dx}{(1+ax^p)^\nu} = \frac{1}{pa^{\mu/p}}B\left(\frac{\mu}{p}, \nu-\frac{\mu}{p}\right), \quad p>0, \quad 0<\mu<p\nu.$$

22. $$\int_0^\infty \left(\sqrt{x^2+a^2}-x\right)^n dx = \frac{na^{n+1}}{n^2-1}, \quad n=2,\ 3,\ \ldots$$

23. $$\int_0^\infty \frac{dx}{\left(x+\sqrt{x^2+a^2}\,\right)^n} = \frac{n}{a^{n-1}(n^2-1)}, \quad n=2,\ 3,\ \ldots$$

24. $$\int_0^\infty x^m\left(\sqrt{x^2+a^2}-x\right)^n dx = \frac{m!\,na^{n+m+1}}{(n-m-1)(n-m+1)\ldots(n+m+1)},$$

$n,m=1,\ 2,\ \ldots, \quad 0\le m\le n-2.$

25. $$\int_0^\infty \frac{x^m\,dx}{\left(x+\sqrt{x^2+a^2}\,\right)^n} = \frac{m!\,n}{(n-m-1)(n-m+1)\ldots(n+m+1)a^{n-m-1}}, \quad n=2,\ 3,\ \ldots$$

S1.2.2. Integrals Involving Exponential Functions

1. $\displaystyle\int_0^\infty e^{-ax}\,dx = \frac{1}{a}, \quad a > 0.$

2. $\displaystyle\int_0^1 x^n e^{-ax}\,dx = \frac{n!}{a^{n+1}} - e^{-a}\sum_{k=0}^{n}\frac{n!}{k!}\,\frac{1}{a^{n-k+1}}, \quad a > 0, \; n = 1,\, 2,\, \ldots$

3. $\displaystyle\int_0^\infty x^n e^{-ax}\,dx = \frac{n!}{a^{n+1}}, \quad a > 0, \; n = 1,\, 2,\, \ldots$

4. $\displaystyle\int_0^\infty \frac{e^{-ax}}{\sqrt{x}}\,dx = \sqrt{\frac{\pi}{a}}, \quad a > 0.$

5. $\displaystyle\int_0^\infty x^{\nu-1}e^{-\mu x}\,dx = \frac{\Gamma(\nu)}{\mu^\nu}, \quad \mu, \nu > 0.$

6. $\displaystyle\int_0^\infty \frac{dx}{1+e^{ax}} = \frac{\ln 2}{a}.$

7. $\displaystyle\int_0^\infty \frac{x^{2n-1}\,dx}{e^{px}-1} = (-1)^{n-1}\left(\frac{2\pi}{p}\right)^{2n}\frac{B_{2n}}{4n}, \quad n = 1,\, 2,\, \ldots;$
the B_m are Bernoulli numbers.

8. $\displaystyle\int_0^\infty \frac{x^{2n-1}\,dx}{e^{px}+1} = (1-2^{1-2n})\left(\frac{2\pi}{p}\right)^{2n}\frac{|B_{2n}|}{4n}, \quad n = 1,\, 2,\, \ldots$

9. $\displaystyle\int_{-\infty}^\infty \frac{e^{-px}\,dx}{1+e^{-qx}} = \frac{\pi}{q\sin(\pi p/q)}, \quad q > p > 0 \text{ or } 0 > p > q.$

10. $\displaystyle\int_0^\infty \frac{e^{ax}+e^{-ax}}{e^{bx}+e^{-bx}}\,dx = \frac{\pi}{2b\cos\left(\dfrac{\pi a}{2b}\right)}, \quad b > a.$

11. $\displaystyle\int_0^\infty \frac{e^{-px}-e^{-qx}}{1-e^{-(p+q)x}}\,dx = \frac{\pi}{p+q}\cot\frac{\pi p}{p+q}, \quad p, q > 0.$

12. $\displaystyle\int_0^\infty \left(1-e^{-\beta x}\right)^\nu e^{-\mu x}\,dx = \frac{1}{\beta}B\left(\frac{\mu}{\beta}, \nu+1\right).$

13. $\displaystyle\int_0^\infty \exp\left(-ax^2\right)dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}, \quad a > 0.$

14. $\displaystyle\int_0^\infty x^{2n+1}\exp\left(-ax^2\right)dx = \frac{n!}{2a^{n+1}}, \quad a > 0, \quad n = 1,\, 2,\, \ldots$

15. $\displaystyle\int_0^\infty x^{2n}\exp\left(-ax^2\right)dx = \frac{1\times 3\times\ldots\times(2n-1)\sqrt{\pi}}{2^{n+1}a^{n+1/2}}, \quad a > 0, \quad n = 1,\, 2,\, \ldots$

16. $\displaystyle\int_{-\infty}^\infty \exp\left(-a^2x^2 \pm bx\right)dx = \frac{\sqrt{\pi}}{|a|}\,\exp\left(\frac{b^2}{4a^2}\right).$

17. $\displaystyle\int_0^\infty \exp\left(-ax^2 - \frac{b}{x^2}\right)dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}\,\exp\left(-2\sqrt{ab}\right), \quad a, b > 0.$

18. $\displaystyle\int_0^\infty \exp\left(-x^a\right)dx = \frac{1}{a}\Gamma\left(\frac{1}{a}\right), \quad a > 0.$

S1.2.3. Integrals Involving Hyperbolic Functions

1. $\displaystyle\int_0^\infty \frac{dx}{\cosh ax} = \frac{\pi}{2|a|}.$

2. $\displaystyle\int_0^\infty \frac{dx}{a + b\cosh x} = \begin{cases} \dfrac{2}{\sqrt{b^2 - a^2}} \arctan \dfrac{\sqrt{b^2 - a^2}}{a + b} & \text{if } |b| > |a|, \\ \dfrac{1}{\sqrt{a^2 - b^2}} \ln \dfrac{a + b + \sqrt{a^2 - b^2}}{a + b - \sqrt{a^2 + b^2}} & \text{if } |b| < |a|. \end{cases}$

3. $\displaystyle\int_0^\infty \frac{x^{2n}\,dx}{\cosh ax} = \left(\frac{\pi}{2a}\right)^{2n+1} |B_{2n}|, \quad a > 0;$ the B_m are Bernoulli numbers.

4. $\displaystyle\int_0^\infty \frac{x^{2n}}{\cosh^2 ax}\,dx = \frac{\pi^{2n}(2^{2n} - 2)}{a(2a)^{2n}} |B_{2n}|, \quad a > 0.$

5. $\displaystyle\int_0^\infty \frac{\cosh ax}{\cosh bx}\,dx = \frac{\pi}{2b\cos\left(\dfrac{\pi a}{2b}\right)}, \quad b > |a|.$

6. $\displaystyle\int_0^\infty x^{2n}\frac{\cosh ax}{\cosh bx}\,dx = \frac{\pi}{2b}\frac{d^{2n}}{da^{2n}}\frac{1}{\cos\left(\frac{1}{2}\pi a/b\right)}, \quad b > |a|, \quad n = 1,\ 2,\ \ldots$

7. $\displaystyle\int_0^\infty \frac{\cosh ax \cosh bx}{\cosh(cx)}\,dx = \frac{\pi}{c}\frac{\cos\left(\dfrac{\pi a}{2c}\right)\cos\left(\dfrac{\pi b}{2c}\right)}{\cos\left(\dfrac{\pi a}{c}\right) + \cos\left(\dfrac{\pi b}{c}\right)}, \quad c > |a| + |b|.$

8. $\displaystyle\int_0^\infty \frac{x\,dx}{\sinh ax} = \frac{\pi^2}{2a^2}, \quad a > 0.$

9. $\displaystyle\int_0^\infty \frac{dx}{a + b\sinh x} = \frac{1}{\sqrt{a^2 + b^2}} \ln \frac{a + b + \sqrt{a^2 + b^2}}{a + b - \sqrt{a^2 + b^2}}, \quad ab \neq 0.$

10. $\displaystyle\int_0^\infty \frac{\sinh ax}{\sinh bx}\,dx = \frac{\pi}{2b}\tan\left(\frac{\pi a}{2b}\right), \quad b > |a|.$

11. $\displaystyle\int_0^\infty x^{2n}\frac{\sinh ax}{\sinh bx}\,dx = \frac{\pi}{2b}\frac{d^{2n}}{dx^{2n}}\tan\left(\frac{\pi a}{2b}\right), \quad b > |a|, \quad n = 1,\ 2,\ \ldots$

12. $\displaystyle\int_0^\infty \frac{x^{2n}}{\sinh^2 ax}\,dx = \frac{\pi^{2n}}{a^{2n+1}}|B_{2n}|, \quad a > 0.$

S1.2.4. Integrals Involving Logarithmic Functions

1. $\displaystyle\int_0^1 x^{a-1}\ln^n x\,dx = (-1)^n n!\, a^{-n-1}, \quad a > 0, \quad n = 1,\ 2,\ \ldots$

2. $\displaystyle\int_0^1 \frac{\ln x}{x + 1}\,dx = -\frac{\pi^2}{12}.$

3. $\displaystyle\int_0^1 \frac{x^n \ln x}{x + 1}\,dx = (-1)^{n+1}\left[\frac{\pi^2}{12} + \sum_{k=1}^n \frac{(-1)^k}{k^2}\right], \quad n = 1,\ 2,\ \ldots$

4. $$\int_0^1 \frac{x^{\mu-1}\ln x}{x+a}\,dx = \frac{\pi a^{\mu-1}}{\sin(\pi\mu)}\bigl[\ln a - \pi\cot(\pi\mu)\bigr], \quad 0<\mu<1.$$

5. $$\int_0^1 |\ln x|^{\mu}\,dx = \Gamma(\mu+1), \quad \mu>-1.$$

6. $$\int_0^{\infty} x^{\mu-1}\ln(1+ax)\,dx = \frac{\pi}{\mu a^{\mu}\sin(\pi\mu)}, \quad -1<\mu<0.$$

7. $$\int_0^1 x^{2n-1}\ln(1+x)\,dx = \frac{1}{2n}\sum_{k=1}^{2n}\frac{(-1)^{k-1}}{k}, \quad n=1,\,2,\,\ldots$$

8. $$\int_0^1 x^{2n}\ln(1+x)\,dx = \frac{1}{2n+1}\left[\ln 4 + \sum_{k=1}^{2n+1}\frac{(-1)^{k}}{k}\right], \quad n=0,\,1,\,\ldots$$

9. $$\int_0^1 x^{n-1/2}\ln(1+x)\,dx = \frac{2\ln 2}{2n+1} + \frac{4(-1)^n}{2n+1}\left[\pi - \sum_{k=0}^{n}\frac{(-1)^k}{2k+1}\right], \quad n=1,\,2,\,\ldots$$

10. $$\int_0^{\infty}\ln\frac{a^2+x^2}{b^2+x^2}\,dx = \pi(a-b), \quad a,b>0.$$

11. $$\int_0^{\infty}\frac{x^{p-1}\ln x}{1+x^q}\,dx = -\frac{\pi^2\cos(\pi p/q)}{q^2\sin^2(\pi p/q)}, \quad 0<p<q.$$

12. $$\int_0^{\infty} e^{-\mu x}\ln x\,dx = -\frac{1}{\mu}(\mathcal{C}+\ln\mu), \quad \mu>0, \quad \mathcal{C}=0.5772\ldots$$

S1.2.5. Integrals Involving Trigonometric Functions

► **Integrals over a finite interval.**

1. $$\int_0^{\pi/2}\cos^{2n}x\,dx = \frac{\pi}{2}\,\frac{1\times3\times\ldots\times(2n-1)}{2\times4\times\ldots\times(2n)}, \quad n=1,\,2,\,\ldots$$

2. $$\int_0^{\pi/2}\cos^{2n+1}x\,dx = \frac{2\times4\times\ldots\times(2n)}{1\times3\times\ldots\times(2n+1)}, \quad n=1,\,2,\,\ldots$$

3. $$\int_0^{\pi/2} x\cos^n x\,dx = -\sum_{k=0}^{m-1}\frac{(n-2k+1)(n-2k+3)\ldots(n-1)}{(n-2k)(n-2k+2)\ldots n}\,\frac{1}{n-2k} + \begin{cases}\dfrac{\pi}{2}\,\dfrac{(2m-2)!!}{(2m-1)!!} & \text{if } n=2m-1,\\[2mm] \dfrac{\pi^2}{8}\,\dfrac{(2m-1)!!}{(2m)!!} & \text{if } n=2m,\end{cases} \qquad m=1,\,2,\,\ldots$$

4. $$\int_0^{\pi}\frac{dx}{(a+b\cos x)^{n+1}} = \frac{\pi}{2^n(a+b)^n\sqrt{a^2-b^2}}\sum_{k=0}^{n}\frac{(2n-2k-1)!!\,(2k-1)!!}{(n-k)!\,k!}\left(\frac{a+b}{a-b}\right)^k, \quad a>|b|.$$

5. $$\int_0^{\pi/2}\sin^{2n}x\,dx = \frac{\pi}{2}\,\frac{1\times3\times\ldots\times(2n-1)}{2\times4\times\ldots\times(2n)}, \quad n=1,\,2,\,\ldots$$

6. $$\int_0^{\pi/2} \sin^{2n+1} x\,dx = \frac{2\times 4\times\ldots\times(2n)}{1\times 3\times\ldots\times(2n+1)}, \quad n = 1,\ 2,\ \ldots$$

7. $$\int_0^{\pi} x\sin^{\mu} x\,dx = \frac{\pi^2}{2^{\mu+1}}\frac{\Gamma(\mu+1)}{\bigl[\Gamma\bigl(\mu+\frac{1}{2}\bigr)\bigr]^2}, \quad \mu > -1.$$

8. $$\int_0^{\pi/2} \frac{\sin x\,dx}{\sqrt{1-k^2\sin^2 x}} = \frac{1}{2k}\ln\frac{1+k}{1-k}.$$

9. $$\int_0^{\pi/2} \sin^{2n+1} x\cos^{2m+1} x\,dx = \frac{n!\,m!}{2(n+m+1)!}, \quad n, m = 1,\ 2,\ \ldots$$

10. $$\int_0^{\pi/2} \sin^{p-1} x\cos^{q-1} x\,dx = \tfrac{1}{2}B\bigl(\tfrac{1}{2}p, \tfrac{1}{2}q\bigr).$$

11. $$\int_0^{2\pi} (a\sin x + b\cos x)^{2n}\,dx = 2\pi\frac{(2n-1)!!}{(2n)!!}\bigl(a^2+b^2\bigr)^n, \quad n = 1,\ 2,\ \ldots$$

12. $$\int_0^{\pi} \frac{\sin x\,dx}{\sqrt{a^2+1-2a\cos x}} = \begin{cases} 2 & \text{if } 0 \le a \le 1, \\ 2/a & \text{if } 1 < a. \end{cases}$$

13. $$\int_0^{\pi/2} (\tan x)^{\pm\lambda}\,dx = \frac{\pi}{2\cos\bigl(\frac{1}{2}\pi\lambda\bigr)}, \quad |\lambda| < 1.$$

► **Integrals over an infinite interval.**

1. $$\int_0^{\infty} \frac{\cos ax}{\sqrt{x}}\,dx = \sqrt{\frac{\pi}{2a}}, \quad a > 0.$$

2. $$\int_0^{\infty} \frac{\cos ax - \cos bx}{x}\,dx = \ln\left|\frac{b}{a}\right|, \quad ab \ne 0.$$

3. $$\int_0^{\infty} \frac{\cos ax - \cos bx}{x^2}\,dx = \tfrac{1}{2}\pi(b-a), \quad a, b \ge 0.$$

4. $$\int_0^{\infty} x^{\mu-1}\cos ax\,dx = a^{-\mu}\Gamma(\mu)\cos\bigl(\tfrac{1}{2}\pi\mu\bigr), \quad a > 0, \quad 0 < \mu < 1.$$

5. $$\int_0^{\infty} \frac{\cos ax}{b^2+x^2}\,dx = \frac{\pi}{2b}e^{-ab}, \quad a, b > 0.$$

6. $$\int_0^{\infty} \frac{\cos ax}{b^4+x^4}\,dx = \frac{\pi\sqrt{2}}{4b^3}\exp\left(-\frac{ab}{\sqrt{2}}\right)\left[\cos\left(\frac{ab}{\sqrt{2}}\right) + \sin\left(\frac{ab}{\sqrt{2}}\right)\right], \quad a, b > 0.$$

7. $$\int_0^{\infty} \frac{\cos ax}{(b^2+x^2)^2}\,dx = \frac{\pi}{4b^3}(1+ab)e^{-ab}, \quad a, b > 0.$$

8. $$\int_0^{\infty} \frac{\cos ax\,dx}{(b^2+x^2)(c^2+x^2)} = \frac{\pi\bigl(be^{-ac} - ce^{-ab}\bigr)}{2bc\bigl(b^2-c^2\bigr)}, \quad a, b, c > 0.$$

9. $$\int_0^{\infty} \cos\bigl(ax^2\bigr)\,dx = \frac{1}{2}\sqrt{\frac{\pi}{2a}}, \quad a > 0.$$

10. $$\int_0^{\infty} \cos\bigl(ax^p\bigr)\,dx = \frac{\Gamma(1/p)}{pa^{1/p}}\cos\frac{\pi}{2p}, \quad a > 0, \quad p > 1.$$

11. $\displaystyle\int_0^\infty \frac{\sin ax}{x}\,dx = \frac{\pi}{2}\operatorname{sign} a.$

12. $\displaystyle\int_0^\infty \frac{\sin^2 ax}{x^2}\,dx = \frac{\pi}{2}|a|.$

13. $\displaystyle\int_0^\infty \frac{\sin ax}{\sqrt{x}}\,dx = \sqrt{\frac{\pi}{2a}}, \quad a > 0.$

14. $\displaystyle\int_0^\infty x^{\mu-1}\sin ax\,dx = a^{-\mu}\Gamma(\mu)\sin\bigl(\tfrac{1}{2}\pi\mu\bigr), \quad a > 0, \quad 0 < \mu < 1.$

15. $\displaystyle\int_0^\infty \sin\bigl(ax^2\bigr)\,dx = \frac{1}{2}\sqrt{\frac{\pi}{2a}}, \quad a > 0.$

16. $\displaystyle\int_0^\infty \sin\bigl(ax^p\bigr)\,dx = \frac{\Gamma(1/p)}{pa^{1/p}}\sin\frac{\pi}{2p}, \quad a > 0, \quad p > 1.$

17. $\displaystyle\int_0^\infty \frac{\sin x\cos ax}{x}\,dx = \begin{cases} \frac{\pi}{2} & \text{if } |a| < 1, \\ \frac{\pi}{4} & \text{if } |a| = 1, \\ 0 & \text{if } 1 < |a|. \end{cases}$

18. $\displaystyle\int_0^\infty \frac{\tan ax}{x}\,dx = \frac{\pi}{2}\operatorname{sign} a.$

19. $\displaystyle\int_0^\infty e^{-ax}\sin bx\,dx = \frac{b}{a^2+b^2}, \quad a > 0.$

20. $\displaystyle\int_0^\infty e^{-ax}\cos bx\,dx = \frac{a}{a^2+b^2}, \quad a > 0.$

21. $\displaystyle\int_0^\infty \exp\bigl(-ax^2\bigr)\cos bx\,dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}\exp\Bigl(-\frac{b^2}{4a}\Bigr).$

22. $\displaystyle\int_0^\infty \cos(ax^2)\cos bx\,dx = \sqrt{\frac{\pi}{8a}}\left[\cos\left(\frac{b^2}{4a}\right) + \sin\left(\frac{b^2}{4a}\right)\right], \quad a, b > 0.$

23. $\displaystyle\int_0^\infty (\cos ax + \sin ax)\cos(b^2x^2)\,dx = \frac{1}{b}\sqrt{\frac{\pi}{8}}\exp\left(-\frac{a^2}{2b}\right), \quad a, b > 0.$

24. $\displaystyle\int_0^\infty \bigl[\cos ax + \sin ax\bigr]\sin(b^2x^2)\,dx = \frac{1}{b}\sqrt{\frac{\pi}{8}}\exp\Bigl(-\frac{a^2}{2b}\Bigr), \quad a, b > 0.$

S1.3. General Reduction Formulas for the Evaluation of Definite Integrals

► *Below are some general formulas, involving arbitrary functions and parameters, that could facilitate the evaluation of integrals.*

S1.3.1. Integrals Involving Functions of a Linear or Rational Argument

1. $\displaystyle\int_{-a}^a f(x)\,dx = 0 \quad \text{if } f(x) \text{ is odd.}$

2. $\displaystyle\int_{-a}^{a} f(x)\,dx = 2\int_0^a f(x)\,dx \quad \text{if } f(x) \text{ is even.}$

3. $\displaystyle\int_0^a [f(x)+f(a-x)]\,dx = 2\int_0^a f(x)\,dx.$

4. $\displaystyle\int_0^a [f(x)-f(a-x)]\,dx = 0.$

5. $\displaystyle\int_0^1 f\bigl(2x\sqrt{1-x^2}\,\bigr)\,dx = \int_0^1 f\bigl(1-x^2\bigr)\,dx.$

S1.3.2. Integrals Involving Functions with Trigonometric Argument

1. $\displaystyle\int_0^{\pi} f(\sin x)\,dx = 2\int_0^{\pi/2} f(\sin x)\,dx.$

2. $\displaystyle\int_0^{\pi/2} f(\sin x)\,dx = \int_0^{\pi/2} f(\cos x)\,dx.$

3. $\displaystyle\int_0^{\pi/2} f(\sin x, \cos x)\,dx = 0 \quad \text{if } f(x,y) = -f(y,x).$

4. $\displaystyle\int_0^{\pi/2} f(\sin 2x)\cos x\,dx = \int_0^{\pi/2} f(\cos^2 x)\cos x\,dx.$

5. $\displaystyle\int_0^{n\pi} xf(\sin x)\,dx = \pi n^2\int_0^{\pi/2} f(\sin x)\,dx \quad \text{if } f(x) = f(-x).$

6. $\displaystyle\int_0^{n\pi} xf(\sin x)\,dx = (-1)^{n-1}\pi n\int_0^{\pi/2} f(\sin x)\,dx \quad \text{if } f(-x) = -f(x).$

7. $\displaystyle\int_0^{2\pi} f(a\sin x+b\cos x)\,dx = \int_0^{2\pi} f\bigl(\sqrt{a^2+b^2}\,\sin x\bigr)\,dx = 2\int_0^{\pi} f\bigl(\sqrt{a^2+b^2}\,\cos x\bigr)\,dx.$

8. $\displaystyle\int_0^{\pi} f\left(\frac{\sin^2 x}{1+2a\cos x+a^2}\right)dx = \int_0^{\pi} f(\sin^2 x)\,dx \quad \text{if } |a| \ge 1.$

9. $\displaystyle\int_0^{\pi} f\left(\frac{\sin^2 x}{1+2a\cos x+a^2}\right)dx = \int_0^{\pi} f\left(\frac{\sin^2 x}{a^2}\right)dx \quad \text{if } 0 < |a| < 1.$

S1.3.3. Integrals Involving Logarithmic Functions

1. $\displaystyle\int_a^b f(x)\ln^n x\,dx = \left[\left(\frac{d}{d\lambda}\right)^n \int_a^b x^{\lambda} f(x)\,dx\right]_{\lambda=0}.$

2. $\displaystyle\int_a^b f(x)\ln^n g(x)\,dx = \left[\left(\frac{d}{d\lambda}\right)^n \int_a^b f(x)[g(x)]^{\lambda}\,dx\right]_{\lambda=0}.$

3. $\displaystyle\int_a^b f(x)[g(x)]^{\lambda}\ln^n g(x)\,dx = \left(\frac{d}{d\lambda}\right)^n \int_a^b f(x)[g(x)]^{\lambda}\,dx.$

S1.3.4. General Reduction Formulas for the Calculation of Improper Integrals

▶ Improper integrals involving power functions.

1. $\displaystyle\int_0^\infty f\left(\frac{a+bx}{1+x}\right)\frac{dx}{(1+x)^2}=\frac{1}{b-a}\int_a^b f(x)\,dx.$

2. $\displaystyle\int_0^\infty \frac{f(ax)-f(bx)}{x}\,dx=\bigl[f(0)-f(\infty)\bigr]\ln\frac{b}{a}$ if $a>0,\ b>0,\ f(x)$ is continuous on $[0,\infty)$, and $f(\infty)=\lim\limits_{x\to\infty}f(x)$ is a finite quantity.

3. $\displaystyle\int_0^\infty \frac{f(ax)-f(bx)}{x}\,dx=f(0)\ln\frac{b}{a}$ if $a>0,\ b>0,\ f(x)$ is continuous on $[0,\infty)$, and the integral $\displaystyle\int_c^\infty \frac{f(x)}{x}\,dx$ exists; $c>0$.

4. $\displaystyle\int_0^\infty f\left(\left|ax-\frac{b}{x}\right|\right)dx=\frac{1}{a}\int_0^\infty f(|x|)\,dx$ if $a>0,\ b>0.$

5. $\displaystyle\int_0^\infty x^2 f\left(\left|ax-\frac{b}{x}\right|\right)dx=\frac{1}{a^3}\int_0^\infty (x^2+ab)f(|x|)\,dx$ if $a>0,\ b>0.$

6. $\displaystyle\int_0^\infty f\left(\left|ax-\frac{b}{x}\right|\right)\frac{dx}{x^2}=\frac{1}{b}\int_0^\infty f(|x|)\,dx$ if $a>0,\ b>0.$

7. $\displaystyle\int_0^\infty f\left(x,\frac{1}{x}\right)\frac{dx}{x}=2\int_0^1 f\left(x,\frac{1}{x}\right)\frac{dx}{x}$ if $f(x,y)=f(y,x).$

8. $\displaystyle\int_0^\infty f\left(x,\frac{a}{x}\right)\frac{dx}{x}=0$ if $f(x,y)=-f(y,x),\ a>0$ (the integral is assumed to exist).

▶ Improper integrals involving logarithmic functions.

1. $\displaystyle\int_0^\infty f\left(\frac{x}{a}+\frac{a}{x}\right)\frac{\ln x}{x}\,dx=\ln a\int_0^\infty f\left(\frac{x}{a}+\frac{a}{x}\right)\frac{dx}{x}$ if $a>0.$

2. $\displaystyle\int_0^\infty f\left(\frac{x^p}{a}+\frac{a}{x^p}\right)\frac{\ln x}{x}\,dx=\frac{\ln a}{p}\int_0^\infty f\left(\frac{x^p}{a}+\frac{a}{x^p}\right)\frac{dx}{x}$ if $a>0,\ p>0.$

3. $\displaystyle\int_0^\infty f(x^a+x^{-a})\frac{\ln x}{1+x^2}\,dx=0$ (a special case of the integral below).

4. $\displaystyle\int_0^\infty f\left(x,\frac{1}{x}\right)\frac{\ln x}{1+x^2}\,dx=0$ if $f(x,y)=f(y,x)$ (the integral is assumed to exist).

5. $\displaystyle\int_0^\infty f\left(x,\frac{1}{x}\right)\frac{\ln x}{x}\,dx=0$ if $f(x,y)=f(y,x)$ (the integral is assumed to exist).

▶ Improper integrals involving trigonometric functions.

1. $\displaystyle\int_0^\infty f(x)\frac{\sin x}{x}\,dx=\int_0^{\pi/2} f(x)\,dx$ if $f(x)=f(-x)$ and $f(x+\pi)=f(x).$

2. $\displaystyle\int_0^\infty f(x)\frac{\sin x}{x}\,dx=\int_0^{\pi/2} f(x)\cos x\,dx$ if $f(x)=f(-x)$ and $f(x+\pi)=-f(x).$

3. $\displaystyle\int_0^\infty \frac{f(\sin x)}{x}\,dx = \int_0^{\pi/2} \frac{f(\sin x)}{\sin x}\,dx \quad \text{if } f(-x) = -f(x).$

4. $\displaystyle\int_0^\infty \frac{f(\sin x)}{x^2}\,dx = \int_0^{\pi/2} \frac{f(\sin x)}{\sin^2 x}\,dx \quad \text{if } f(x) = f(-x).$

5. $\displaystyle\int_0^\infty \frac{f(\sin x)}{x}\cos x\,dx = \int_0^{\pi/2} \frac{f(\sin x)}{\sin x}\cos^2 x\,dx \quad \text{if } f(-x) = -f(x).$

6. $\displaystyle\int_0^\infty \frac{f(\sin x)}{x}\tan x\,dx = \int_0^{\pi/2} f(\sin x)\,dx \quad \text{if } f(-x) = f(x).$

7. $\displaystyle\int_0^\infty \frac{f(\sin x)}{x^2+a^2}\,dx = \frac{\sinh(2a)}{2a}\int_0^{\pi/2} \frac{f(\sin x)\,dx}{\cosh^2 a - \cos^2 x} \quad \text{if } f(-x) = f(x).$

8. $\displaystyle\int_0^\infty f\left(x+\frac{1}{x}\right)\frac{\arctan x}{x}\,dx = \frac{\pi}{4}\int_0^\infty f\left(x+\frac{1}{x}\right)\frac{dx}{x}.$

► **Calculation of improper integrals using analytic functions.** Suppose

$$F(z) = f(r,x) + ig(r,x), \quad z = r(\cos x + i\sin x), \quad i^2 = -1,$$

where $F(z)$ is a function analytic in a circle of radius r, $f(r,x) = \operatorname{Re} F(z)$, and $g(r,x) = \operatorname{Im} F(z)$. Then the following formulas hold:

1. $\displaystyle\int_0^\infty \frac{f(r,x)}{x^2+a^2}\,dx = \frac{\pi}{2a}F(re^{-a}).$

2. $\displaystyle\int_0^\infty \frac{xg(r,x)}{x^2+a^2}\,dx = \frac{\pi}{2}[F(re^{-a}) - F(0)].$

3. $\displaystyle\int_0^\infty \frac{g(r,x)}{x}\,dx = \frac{\pi}{2}[F(r) - F(0)].$

4. $\displaystyle\int_0^\infty \frac{g(r,x)}{x(x^2+a^2)}\,dx = \frac{\pi}{2a^2}[F(r) - F(re^{-a})].$

Bibliography for Chapter S1

Dwight, H. B., *Tables of Integrals and Other Mathematical Data*, Macmillan, New York, 1961.

Gradshteyn, I. S. and Ryzhik, I. M., *Tables of Integrals, Series, and Products, 6th Edition*, Academic Press, New York, 2000.

Prudnikov, A. P., Brychkov, Yu. A., and Marichev, O. I., *Integrals and Series, Vol. 1, Elementary Functions*, Gordon & Breach, New York, 1986.

Prudnikov, A. P., Brychkov, Yu. A., and Marichev, O. I., *Integrals and Series, Vol. 2, Special Functions*, Gordon & Breach, New York, 1986.

Prudnikov, A. P., Brychkov, Yu. A., and Marichev, O. I., *Integrals and Series, Vol. 3, More Special Functions*, Gordon & Breach, New York, 1988.

Zwillinger, D., *CRC Standard Mathematical Tables and Formulae, 31st Edition*, CRC Press, Boca Raton, 2002.

Chapter S2

Integral Transforms*

S2.1. Tables of Laplace Transforms

S2.1.1. General Formulas

No.	Original function, $f(x)$	Laplace transform, $\widetilde{f}(p) = \int_0^\infty e^{-px} f(x)\,dx$
1	$af_1(x) + bf_2(x)$	$a\widetilde{f}_1(p) + b\widetilde{f}_2(p)$
2	$f(x/a),\ \ a > 0$	$a\widetilde{f}(ap)$
3	$\begin{cases} 0 & \text{if } 0 < x < a, \\ f(x-a) & \text{if } a < x \end{cases}$	$e^{-ap}\widetilde{f}(p)$
4	$x^n f(x);\ \ n = 1,\ 2,\ \dots$	$(-1)^n \dfrac{d^n}{dp^n}\widetilde{f}(p)$
5	$\dfrac{1}{x}f(x)$	$\displaystyle\int_p^\infty \widetilde{f}(q)\,dq$
6	$e^{ax}f(x)$	$\widetilde{f}(p-a)$
7	$\sinh(ax)f(x)$	$\frac{1}{2}\bigl[\widetilde{f}(p-a) - \widetilde{f}(p+a)\bigr]$
8	$\cosh(ax)f(x)$	$\frac{1}{2}\bigl[\widetilde{f}(p-a) + \widetilde{f}(p+a)\bigr]$
9	$\sin(\omega x)f(x)$	$-\frac{i}{2}\bigl[\widetilde{f}(p-i\omega) - \widetilde{f}(p+i\omega)\bigr],\ \ i^2 = -1$
10	$\cos(\omega x)f(x)$	$\frac{1}{2}\bigl[\widetilde{f}(p-i\omega) + \widetilde{f}(p+i\omega)\bigr],\ \ i^2 = -1$
11	$f(x^2)$	$\dfrac{1}{\sqrt{\pi}}\displaystyle\int_0^\infty \exp\Bigl(-\frac{p^2}{4t^2}\Bigr)\widetilde{f}(t^2)\,dt$
12	$x^{a-1}f\Bigl(\dfrac{1}{x}\Bigr),\ \ a > -1$	$\displaystyle\int_0^\infty (t/p)^{a/2} J_a\bigl(2\sqrt{pt}\,\bigr)\widetilde{f}(t)\,dt$
13	$f(a\sinh x),\ \ a > 0$	$\displaystyle\int_0^\infty J_p(at)\widetilde{f}(t)\,dt$
14	$f(x+a) = f(x)$ (periodic function)	$\dfrac{1}{1-e^{ap}}\displaystyle\int_0^a f(x)e^{-px}\,dx$
15	$f(x+a) = -f(x)$ (antiperiodic function)	$\dfrac{1}{1+e^{-ap}}\displaystyle\int_0^a f(x)e^{-px}\,dx$
16	$f'_x(x)$	$p\widetilde{f}(p) - f(+0)$
17	$f^{(n)}_x(x)$	$p^n\widetilde{f}(p) - \displaystyle\sum_{k=1}^{n} p^{n-k} f^{(k-1)}_x(+0)$

---

* For definitions and properties of the special functions occurring in this chapter, such as $J_a(z)$, $\operatorname{erfc} z$, $\Gamma(z)$, etc., see Chapter M13.

No.	Original function, $f(x)$	Laplace transform, $\widetilde{f}(p)=\int_0^\infty e^{-px}f(x)\,dx$
18	$x^m f_x^{(n)}(x),\quad m\ge n$	$\left(-\dfrac{d}{dp}\right)^m\left[p^n\widetilde{f}(p)\right]$
19	$\dfrac{d^n}{dx^n}\left[x^m f(x)\right],\quad m\ge n$	$(-1)^m p^n \dfrac{d^m}{dp^m}\widetilde{f}(p)$
20	$\displaystyle\int_0^x f(t)\,dt$	$\dfrac{\widetilde{f}(p)}{p}$
21	$\displaystyle\int_0^x (x-t)f(t)\,dt$	$\dfrac{1}{p^2}\widetilde{f}(p)$
22	$\displaystyle\int_0^x (x-t)^\nu f(t)\,dt,\qquad \nu>-1$	$\Gamma(\nu+1)p^{-\nu-1}\widetilde{f}(p)$
23	$\displaystyle\int_0^x e^{-a(x-t)}f(t)\,dt$	$\dfrac{1}{p+a}\widetilde{f}(p)$
24	$\displaystyle\int_0^x \sinh\left[a(x-t)\right]f(t)\,dt$	$\dfrac{a\widetilde{f}(p)}{p^2-a^2}$
25	$\displaystyle\int_0^x \sin\left[a(x-t)\right]f(t)\,dt$	$\dfrac{a\widetilde{f}(p)}{p^2+a^2}$
26	$\displaystyle\int_0^x f_1(t)f_2(x-t)\,dt$	$\widetilde{f}_1(p)\widetilde{f}_2(p)$
27	$\displaystyle\int_0^x \frac{1}{t}f(t)\,dt$	$\displaystyle\frac{1}{p}\int_p^\infty \widetilde{f}(q)\,dq$
28	$\displaystyle\int_x^\infty \frac{1}{t}f(t)\,dt$	$\displaystyle\frac{1}{p}\int_0^p \widetilde{f}(q)\,dq$
29	$\displaystyle\int_0^\infty \frac{1}{\sqrt{t}}\sin\left(2\sqrt{xt}\right)f(t)\,dt$	$\dfrac{\sqrt{\pi}}{p\sqrt{p}}\widetilde{f}\left(\dfrac{1}{p}\right)$
30	$\displaystyle\frac{1}{\sqrt{x}}\int_0^\infty \cos\left(2\sqrt{xt}\right)f(t)\,dt$	$\dfrac{\sqrt{\pi}}{\sqrt{p}}\widetilde{f}\left(\dfrac{1}{p}\right)$
31	$\displaystyle\int_0^\infty \frac{1}{\sqrt{\pi x}}\exp\left(-\frac{t^2}{4x}\right)f(t)\,dt$	$\dfrac{1}{\sqrt{p}}\widetilde{f}\left(\sqrt{p}\right)$
32	$\displaystyle\int_0^\infty \frac{t}{2\sqrt{\pi x^3}}\exp\left(-\frac{t^2}{4x}\right)f(t)\,dt$	$\widetilde{f}\left(\sqrt{p}\right)$
33	$\displaystyle f(x)-a\int_0^x f\left(\sqrt{x^2-t^2}\right)J_1(at)\,dt$	$\widetilde{f}\left(\sqrt{p^2+a^2}\right)$
34	$\displaystyle f(x)+a\int_0^x f\left(\sqrt{x^2-t^2}\right)I_1(at)\,dt$	$\widetilde{f}\left(\sqrt{p^2-a^2}\right)$

S2.1.2. Expressions Involving Power Functions

No.	Original function, $f(x)$	Laplace transform, $\widetilde{f}(p) = \int_0^\infty e^{-px} f(x)\,dx$
1	1	$\dfrac{1}{p}$
2	$\begin{cases} 0 & \text{if } 0 < x < a, \\ 1 & \text{if } a < x < b, \\ 0 & \text{if } b < x \end{cases}$	$\dfrac{1}{p}\left(e^{-ap} - e^{-bp}\right)$
3	x	$\dfrac{1}{p^2}$
4	$\dfrac{1}{x+a}$	$-e^{ap}\operatorname{Ei}(-ap)$
5	$x^n, \qquad n = 1, 2, \ldots$	$\dfrac{n!}{p^{n+1}}$
6	$x^{n-1/2}, \qquad n = 1, 2, \ldots$	$\dfrac{1 \times 3 \times \ldots \times (2n-1)\sqrt{\pi}}{2^n p^{n+1/2}}$
7	$\dfrac{1}{\sqrt{x+a}}$	$\sqrt{\dfrac{\pi}{p}}\, e^{ap}\operatorname{erfc}\left(\sqrt{ap}\right)$
8	$\dfrac{\sqrt{x}}{x+a}$	$\sqrt{\dfrac{\pi}{p}} - \pi\sqrt{a}\, e^{ap}\operatorname{erfc}\left(\sqrt{ap}\right)$
9	$(x+a)^{-3/2}$	$2a^{-1/2} - 2(\pi p)^{1/2} e^{ap}\operatorname{erfc}\left(\sqrt{ap}\right)$
10	$x^{1/2}(x+a)^{-1}$	$(\pi/p)^{1/2} - \pi a^{1/2} e^{ap}\operatorname{erfc}\left(\sqrt{ap}\right)$
11	$x^{-1/2}(x+a)^{-1}$	$\pi a^{-1/2} e^{ap}\operatorname{erfc}\left(\sqrt{ap}\right)$
12	$x^{\nu}, \qquad \nu > -1$	$\Gamma(\nu+1)p^{-\nu-1}$
13	$(x+a)^{\nu}, \qquad \nu > -1$	$p^{-\nu-1}e^{-ap}\Gamma(\nu+1, ap)$
14	$x^{\nu}(x+a)^{-1}, \qquad \nu > -1$	$ke^{ap}\Gamma(-\nu, ap), \qquad k = a^{\nu}\Gamma(\nu+1)$
15	$(x^2 + 2ax)^{-1/2}(x+a)$	$ae^{ap}K_1(ap)$

S2.1.3. Expressions Involving Exponential Functions

No.	Original function, $f(x)$	Laplace transform, $\widetilde{f}(p) = \int_0^\infty e^{-px} f(x)\,dx$
1	e^{-ax}	$(p+a)^{-1}$
2	xe^{-ax}	$(p+a)^{-2}$
3	$x^{\nu-1}e^{-ax}, \qquad \nu > 0$	$\Gamma(\nu)(p+a)^{-\nu}$
4	$\dfrac{1}{x}\left(e^{-ax} - e^{-bx}\right)$	$\ln(p+b) - \ln(p+a)$
5	$\dfrac{1}{x^2}\left(1 - e^{-ax}\right)^2$	$(p+2a)\ln(p+2a) + p\ln p - 2(p+a)\ln(p+a)$
6	$\exp\left(-ax^2\right), \qquad a > 0$	$(\pi b)^{1/2}\exp\left(bp^2\right)\operatorname{erfc}(p\sqrt{b}), \qquad a = \dfrac{1}{4b}$
7	$x\exp\left(-ax^2\right)$	$2b - 2\pi^{1/2}b^{3/2}p\operatorname{erfc}(p\sqrt{b}), \qquad a = \dfrac{1}{4b}$

No.	Original function, $f(x)$	Laplace transform, $\widetilde{f}(p) = \int_0^\infty e^{-px} f(x)\,dx$
8	$\exp(-a/x), \qquad a \ge 0$	$2\sqrt{a/p}K_1\bigl(2\sqrt{ap}\bigr)$
9	$\sqrt{x}\exp(-a/x), \qquad a \ge 0$	$\frac{1}{2}\sqrt{\pi/p^3}\bigl(1+2\sqrt{ap}\bigr)\exp\bigl(-2\sqrt{ap}\bigr)$
10	$\dfrac{1}{\sqrt{x}}\exp(-a/x), \qquad a \ge 0$	$\sqrt{\pi/p}\exp\bigl(-2\sqrt{ap}\bigr)$
11	$\dfrac{1}{x\sqrt{x}}\exp(-a/x), \qquad a > 0$	$\sqrt{\pi/a}\exp\bigl(-2\sqrt{ap}\bigr)$
12	$x^{\nu-1}\exp(-a/x), \qquad a > 0$	$2(a/p)^{\nu/2}K_\nu\bigl(2\sqrt{ap}\bigr)$
13	$\exp\bigl(-2\sqrt{ax}\bigr)$	$p^{-1}-(\pi a)^{1/2}p^{-3/2}e^{a/p}\operatorname{erfc}\bigl(\sqrt{a/p}\bigr)$
14	$\dfrac{1}{\sqrt{x}}\exp\bigl(-2\sqrt{ax}\bigr)$	$(\pi/p)^{1/2}e^{a/p}\operatorname{erfc}\bigl(\sqrt{a/p}\bigr)$

S2.1.4. Expressions Involving Hyperbolic Functions

No.	Original function, $f(x)$	Laplace transform, $\widetilde{f}(p) = \int_0^\infty e^{-px} f(x)\,dx$
1	$\sinh(ax)$	$\dfrac{a}{p^2-a^2}$
2	$\sinh^2(ax)$	$\dfrac{2a^2}{p^3-4a^2p}$
3	$\dfrac{1}{x}\sinh(ax)$	$\dfrac{1}{2}\ln\dfrac{p+a}{p-a}$
4	$x^{\nu-1}\sinh(ax), \qquad \nu > -1$	$\frac{1}{2}\Gamma(\nu)\bigl[(p-a)^{-\nu}-(p+a)^{-\nu}\bigr]$
5	$\sinh\bigl(2\sqrt{ax}\bigr)$	$\dfrac{\sqrt{\pi a}}{p\sqrt{p}}e^{a/p}$
6	$\sqrt{x}\sinh\bigl(2\sqrt{ax}\bigr)$	$\pi^{1/2}p^{-5/2}\bigl(\frac{1}{2}p+a\bigr)e^{a/p}\operatorname{erf}\bigl(\sqrt{a/p}\bigr)-a^{1/2}p^{-2}$
7	$\dfrac{1}{\sqrt{x}}\sinh\bigl(2\sqrt{ax}\bigr)$	$\pi^{1/2}p^{-1/2}e^{a/p}\operatorname{erf}\bigl(\sqrt{a/p}\bigr)$
8	$\dfrac{1}{\sqrt{x}}\sinh^2\bigl(\sqrt{ax}\bigr)$	$\frac{1}{2}\pi^{1/2}p^{-1/2}\bigl(e^{a/p}-1\bigr)$
9	$\cosh(ax)$	$\dfrac{p}{p^2-a^2}$
10	$\cosh^2(ax)$	$\dfrac{p^2-2a^2}{p^3-4a^2p}$
11	$x^{\nu-1}\cosh(ax), \qquad \nu > 0$	$\frac{1}{2}\Gamma(\nu)\bigl[(p-a)^{-\nu}+(p+a)^{-\nu}\bigr]$
12	$\cosh\bigl(2\sqrt{ax}\bigr)$	$\dfrac{1}{p}+\dfrac{\sqrt{\pi a}}{p\sqrt{p}}e^{a/p}\operatorname{erf}\bigl(\sqrt{a/p}\bigr)$
13	$\sqrt{x}\cosh\bigl(2\sqrt{ax}\bigr)$	$\pi^{1/2}p^{-5/2}\bigl(\frac{1}{2}p+a\bigr)e^{a/p}$
14	$\dfrac{1}{\sqrt{x}}\cosh\bigl(2\sqrt{ax}\bigr)$	$\pi^{1/2}p^{-1/2}e^{a/p}$
15	$\dfrac{1}{\sqrt{x}}\cosh^2\bigl(\sqrt{ax}\bigr)$	$\frac{1}{2}\pi^{1/2}p^{-1/2}\bigl(e^{a/p}+1\bigr)$

S2.1.5. Expressions Involving Logarithmic Functions

No.	Original function, $f(x)$	Laplace transform, $\widetilde{f}(p) = \int_0^\infty e^{-px} f(x)\,dx$
1	$\ln x$	$-\frac{1}{p}(\ln p + \mathcal{C})$, $\mathcal{C} = 0.5772\ldots$ is the Euler constant
2	$\ln(1+ax)$	$-\frac{1}{p} e^{p/a}\,\mathrm{Ei}(-p/a)$
3	$\ln(x+a)$	$\frac{1}{p}\bigl[\ln a - e^{ap}\,\mathrm{Ei}(-ap)\bigr]$
4	$x^n \ln x, \qquad n = 1, 2, \ldots$	$\frac{n!}{p^{n+1}}\bigl(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \ln p - \mathcal{C}\bigr)$, $\mathcal{C} = 0.5772\ldots$ is the Euler constant
5	$\frac{1}{\sqrt{x}} \ln x$	$-\sqrt{\pi/p}\,\bigl[\ln(4p) + \mathcal{C}\bigr]$
6	$x^{n-1/2} \ln x, \qquad n = 1, 2, \ldots$	$\frac{k_n}{p^{n+1/2}}\bigl[2 + \frac{2}{3} + \frac{2}{5} + \cdots + \frac{2}{2n-1} - \ln(4p) - \mathcal{C}\bigr]$, $k_n = 1 \times 3 \times 5 \times \ldots \times (2n-1)\frac{\sqrt{\pi}}{2^n}, \quad \mathcal{C} = 0.5772\ldots$
7	$x^{\nu-1} \ln x, \quad \nu > 0$	$\Gamma(\nu)p^{-\nu}\bigl[\psi(\nu) - \ln p\bigr]$, $\quad \psi(\nu)$ is the logarithmic derivative of the gamma function
8	$(\ln x)^2$	$\frac{1}{p}\bigl[(\ln x + \mathcal{C})^2 + \frac{1}{6}\pi^2\bigr], \quad \mathcal{C} = 0.5772\ldots$
9	$e^{-ax} \ln x$	$-\frac{\ln(p+a) + \mathcal{C}}{p+a}, \qquad \mathcal{C} = 0.5772\ldots$

S2.1.6. Expressions Involving Trigonometric Functions

No.	Original function, $f(x)$	Laplace transform, $\widetilde{f}(p) = \int_0^\infty e^{-px} f(x)\,dx$
1	$\sin(ax)$	$\frac{a}{p^2 + a^2}$
2	$\lvert\sin(ax)\rvert, \qquad a > 0$	$\frac{a}{p^2 + a^2} \coth\left(\frac{\pi p}{2a}\right)$
3	$\sin^{2n}(ax), \qquad n = 1, 2, \ldots$	$\frac{a^{2n}(2n)!}{p\bigl[p^2 + (2a)^2\bigr]\bigl[p^2 + (4a)^2\bigr]\ldots\bigl[p^2 + (2na)^2\bigr]}$
4	$\sin^{2n+1}(ax), \qquad n = 1, 2, \ldots$	$\frac{a^{2n+1}(2n+1)!}{\bigl[p^2 + a^2\bigr]\bigl[p^2 + 3^2a^2\bigr]\ldots\bigl[p^2 + (2n+1)^2a^2\bigr]}$
5	$x^n \sin(ax), \qquad n = 1, 2, \ldots$	$\frac{n!\,p^{n+1}}{\bigl(p^2 + a^2\bigr)^{n+1}} \sum\limits_{0 \le 2k \le n} (-1)^k C_{n+1}^{2k+1} \left(\frac{a}{p}\right)^{2k+1}$
6	$\frac{1}{x} \sin(ax)$	$\arctan\left(\frac{a}{p}\right)$
7	$\frac{1}{x} \sin^2(ax)$	$\frac{1}{4} \ln\bigl(1 + 4a^2p^{-2}\bigr)$

No.	Original function, $f(x)$	Laplace transform, $\widetilde{f}(p)=\int_0^\infty e^{-px}f(x)\,dx$
8	$\frac{1}{x^2}\sin^2(ax)$	$a\arctan(2a/p)-\frac{1}{4}p\ln\bigl(1+4a^2p^{-2}\bigr)$
9	$\sin\bigl(2\sqrt{ax}\,\bigr)$	$\frac{\sqrt{\pi a}}{p\sqrt{p}}e^{-a/p}$
10	$\frac{1}{x}\sin\bigl(2\sqrt{ax}\,\bigr)$	$\pi\operatorname{erf}\bigl(\sqrt{a/p}\,\bigr)$
11	$\cos(ax)$	$\frac{p}{p^2+a^2}$
12	$\cos^2(ax)$	$\frac{p^2+2a^2}{p\bigl(p^2+4a^2\bigr)}$
13	$x^n\cos(ax), \qquad n=1,\,2,\,\ldots$	$\frac{n!\,p^{n+1}}{\bigl(p^2+a^2\bigr)^{n+1}}\sum_{0\le 2k\le n+1}(-1)^kC_{n+1}^{2k}\Bigl(\frac{a}{p}\Bigr)^{2k}$
14	$\frac{1}{x}\bigl[1-\cos(ax)\bigr]$	$\frac{1}{2}\ln\bigl(1+a^2p^{-2}\bigr)$
15	$\frac{1}{x}\bigl[\cos(ax)-\cos(bx)\bigr]$	$\frac{1}{2}\ln\frac{p^2+b^2}{p^2+a^2}$
16	$\sqrt{x}\cos\bigl(2\sqrt{ax}\,\bigr)$	$\frac{1}{2}\pi^{1/2}p^{-5/2}(p-2a)e^{-a/p}$
17	$\frac{1}{\sqrt{x}}\cos\bigl(2\sqrt{ax}\,\bigr)$	$\sqrt{\pi/p}\,e^{-a/p}$
18	$\sin(ax)\sin(bx)$	$\frac{2abp}{\bigl[p^2+(a+b)^2\bigr]\bigl[p^2+(a-b)^2\bigr]}$
19	$\cos(ax)\sin(bx)$	$\frac{b\bigl(p^2-a^2+b^2\bigr)}{\bigl[p^2+(a+b)^2\bigr]\bigl[p^2+(a-b)^2\bigr]}$
20	$\cos(ax)\cos(bx)$	$\frac{p\bigl(p^2+a^2+b^2\bigr)}{\bigl[p^2+(a+b)^2\bigr]\bigl[p^2+(a-b)^2\bigr]}$
21	$\frac{ax\cos(ax)-\sin(ax)}{x^2}$	$p\arctan\frac{a}{x}-a$
22	$e^{bx}\sin(ax)$	$\frac{a}{(p-b)^2+a^2}$
23	$e^{bx}\cos(ax)$	$\frac{p-b}{(p-b)^2+a^2}$
24	$\sin(ax)\sinh(ax)$	$\frac{2a^2p}{p^4+4a^4}$
25	$\sin(ax)\cosh(ax)$	$\frac{a\bigl(p^2+2a^2\bigr)}{p^4+4a^4}$
26	$\cos(ax)\sinh(ax)$	$\frac{a\bigl(p^2-2a^2\bigr)}{p^4+4a^4}$
27	$\cos(ax)\cosh(ax)$	$\frac{p^3}{p^4+4a^4}$

S2.1.7. Expressions Involving Special Functions

No.	Original function, $f(x)$	Laplace transform, $\widetilde{f}(p)=\int_0^\infty e^{-px}f(x)\,dx$
1	$\operatorname{erf}(ax)$	$\frac{1}{p}\exp\bigl(b^2p^2\bigr)\operatorname{erfc}(bp), \qquad b=\frac{1}{2a}$
2	$\operatorname{erf}\bigl(\sqrt{ax}\,\bigr)$	$\frac{\sqrt{a}}{p\sqrt{p+a}}$
3	$e^{ax}\operatorname{erf}\bigl(\sqrt{ax}\,\bigr)$	$\frac{\sqrt{a}}{\sqrt{p}\,(p-a)}$
4	$\operatorname{erf}\bigl(\frac{1}{2}\sqrt{a/x}\,\bigr)$	$\frac{1}{p}\bigl[1-\exp\bigl(-\sqrt{ap}\,\bigr)\bigr]$
5	$\operatorname{erfc}\bigl(\sqrt{ax}\,\bigr)$	$\frac{\sqrt{p+a}-\sqrt{a}}{p\sqrt{p+a}}$
6	$e^{ax}\operatorname{erfc}\bigl(\sqrt{ax}\,\bigr)$	$\frac{1}{p+\sqrt{ap}}$
7	$\operatorname{erfc}\bigl(\frac{1}{2}\sqrt{a/x}\,\bigr)$	$\frac{1}{p}\exp\bigl(-\sqrt{ap}\,\bigr)$
8	$\operatorname{Ci}(x)$	$\frac{1}{2p}\ln(p^2+1)$
9	$\operatorname{Si}(x)$	$\frac{1}{p}\operatorname{arccot} p$
10	$\operatorname{Ei}(-x)$	$-\frac{1}{p}\ln(p+1)$
11	$J_0(ax)$	$\frac{1}{\sqrt{p^2+a^2}}$
12	$J_\nu(ax), \qquad \nu>-1$	$\frac{a^\nu}{\sqrt{p^2+a^2}\,\bigl(p+\sqrt{p^2+a^2}\,\bigr)^\nu}$
13	$x^nJ_n(ax), \qquad n=1,\,2,\,\ldots$	$1\times3\times5\times\ldots\times(2n-1)a^n\bigl(p^2+a^2\bigr)^{-n-1/2}$
14	$x^\nu J_\nu(ax), \qquad \nu>-\frac{1}{2}$	$2^\nu\pi^{-1/2}\Gamma\bigl(\nu+\frac{1}{2}\bigr)a^\nu\bigl(p^2+a^2\bigr)^{-\nu-1/2}$
15	$x^{\nu+1}J_\nu(ax), \qquad \nu>-1$	$2^{\nu+1}\pi^{-1/2}\Gamma\bigl(\nu+\frac{3}{2}\bigr)a^\nu p\bigl(p^2+a^2\bigr)^{-\nu-3/2}$
16	$J_0\bigl(2\sqrt{ax}\,\bigr)$	$\frac{1}{p}e^{-a/p}$
17	$\sqrt{x}J_1\bigl(2\sqrt{ax}\,\bigr)$	$\frac{\sqrt{a}}{p^2}e^{-a/p}$
18	$x^{\nu/2}J_\nu\bigl(2\sqrt{ax}\,\bigr), \qquad \nu>-1$	$a^{\nu/2}p^{-\nu-1}e^{-a/p}$
19	$J_0\bigl(a\sqrt{x^2+bx}\,\bigr)$	$\frac{1}{\sqrt{p^2+a^2}}\exp\bigl(bp-b\sqrt{p^2+a^2}\,\bigr)$
20	$I_0(ax)$	$\frac{1}{\sqrt{p^2-a^2}}$
21	$I_\nu(ax), \qquad \nu>-1$	$\frac{a^\nu}{\sqrt{p^2-a^2}\,\bigl(p+\sqrt{p^2-a^2}\,\bigr)^\nu}$

No.	Original function, $f(x)$	Laplace transform, $\widetilde{f}(p) = \int_0^\infty e^{-px} f(x)\,dx$
22	$x^\nu I_\nu(ax)$, $\quad \nu > -\frac{1}{2}$	$2^\nu \pi^{-1/2}\Gamma\bigl(\nu+\frac{1}{2}\bigr)a^\nu\bigl(p^2-a^2\bigr)^{-\nu-1/2}$
23	$x^{\nu+1} I_\nu(ax)$, $\quad \nu > -1$	$2^{\nu+1}\pi^{-1/2}\Gamma\bigl(\nu+\frac{3}{2}\bigr)a^\nu p\bigl(p^2-a^2\bigr)^{-\nu-3/2}$
24	$I_0\bigl(2\sqrt{ax}\,\bigr)$	$\frac{1}{p}e^{a/p}$
25	$\frac{1}{\sqrt{x}} I_1\bigl(2\sqrt{ax}\,\bigr)$	$\frac{1}{\sqrt{a}}\bigl(e^{a/p}-1\bigr)$
26	$x^{\nu/2} I_\nu\bigl(2\sqrt{ax}\,\bigr)$, $\quad \nu > -1$	$a^{\nu/2}p^{-\nu-1}e^{a/p}$
27	$Y_0(ax)$	$-\frac{2}{\pi}\frac{\operatorname{arcsinh}(p/a)}{\sqrt{p^2+a^2}}$
28	$K_0(ax)$	$\frac{\ln\bigl(p+\sqrt{p^2-a^2}\,\bigr)-\ln a}{\sqrt{p^2-a^2}}$

S2.2. Tables of Inverse Laplace Transforms

S2.2.1. General Formulas

No.	Laplace transform, $\widetilde{f}(p)$	Inverse transform, $f(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}(p)\,dp$
1	$\widetilde{f}(p+a)$	$e^{-ax}f(x)$
2	$\widetilde{f}(ap)$, $\quad a > 0$	$\frac{1}{a}f\Bigl(\frac{x}{a}\Bigr)$
3	$\widetilde{f}(ap+b)$, $\quad a > 0$	$\frac{1}{a}\exp\Bigl(-\frac{b}{a}x\Bigr)f\Bigl(\frac{x}{a}\Bigr)$
4	$\widetilde{f}(p-a)+\widetilde{f}(p+a)$	$2f(x)\cosh(ax)$
5	$\widetilde{f}(p-a)-\widetilde{f}(p+a)$	$2f(x)\sinh(ax)$
6	$e^{-ap}\widetilde{f}(p)$, $\quad a \ge 0$	$\begin{cases} 0 & \text{if } 0 \le x < a, \\ f(x-a) & \text{if } a < x \end{cases}$
7	$p\widetilde{f}(p)$	$\frac{df(x)}{dx}$, $\quad$ if $f(+0) = 0$
8	$\frac{1}{p}\widetilde{f}(p)$	$\int_0^x f(t)\,dt$
9	$\frac{1}{p+a}\widetilde{f}(p)$	$e^{-ax}\int_0^x e^{at}f(t)\,dt$
10	$\frac{1}{p^2}\widetilde{f}(p)$	$\int_0^x (x-t)f(t)\,dt$
11	$\frac{\widetilde{f}(p)}{p(p+a)}$	$\frac{1}{a}\int_0^x \bigl[1-e^{a(x-t)}\bigr]f(t)\,dt$
12	$\frac{\widetilde{f}(p)}{(p+a)^2}$	$\int_0^x (x-t)e^{-a(x-t)}f(t)\,dt$
13	$\frac{\widetilde{f}(p)}{(p+a)(p+b)}$	$\frac{1}{b-a}\int_0^x \bigl[e^{-a(x-t)}-e^{-b(x-t)}\bigr]f(t)\,dt$

No.	Laplace transform, $\widetilde{f}(p)$	Inverse transform, $f(x)=\dfrac{1}{2\pi i}\displaystyle\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}(p)\,dp$
14	$\dfrac{\widetilde{f}(p)}{(p+a)^2+b^2}$	$\dfrac{1}{b}\displaystyle\int_0^x e^{-a(x-t)}\sin\big[b(x-t)\big]f(t)\,dt$
15	$\dfrac{1}{p^n}\widetilde{f}(p),\quad n=1,\,2,\,\ldots$	$\dfrac{1}{(n-1)!}\displaystyle\int_0^x (x-t)^{n-1}f(t)\,dt$
16	$\widetilde{f}_1(p)\widetilde{f}_2(p)$	$\displaystyle\int_0^x f_1(t)f_2(x-t)\,dt$
17	$\dfrac{1}{\sqrt{p}}\widetilde{f}\Big(\dfrac{1}{p}\Big)$	$\displaystyle\int_0^\infty \frac{\cos\big(2\sqrt{xt}\,\big)}{\sqrt{\pi x}}f(t)\,dt$
18	$\dfrac{1}{p\sqrt{p}}\widetilde{f}\Big(\dfrac{1}{p}\Big)$	$\displaystyle\int_0^\infty \frac{\sin\big(2\sqrt{xt}\,\big)}{\sqrt{\pi t}}f(t)\,dt$
19	$\dfrac{1}{p^{2\nu+1}}\widetilde{f}\Big(\dfrac{1}{p}\Big)$	$\displaystyle\int_0^\infty (x/t)^{\nu}J_{2\nu}\big(2\sqrt{xt}\,\big)f(t)\,dt$
20	$\dfrac{1}{p}\widetilde{f}\Big(\dfrac{1}{p}\Big)$	$\displaystyle\int_0^\infty J_0\big(2\sqrt{xt}\,\big)f(t)\,dt$
21	$\dfrac{1}{p}\widetilde{f}\Big(p+\dfrac{1}{p}\Big)$	$\displaystyle\int_0^x J_0\big(2\sqrt{xt-t^2}\,\big)f(t)\,dt$
22	$\dfrac{1}{p^{2\nu+1}}\widetilde{f}\Big(p+\dfrac{a}{p}\Big),\quad -\dfrac{1}{2}<\nu\le 0$	$\displaystyle\int_0^x \Big(\frac{x-t}{at}\Big)^{\nu}J_{2\nu}\big(2\sqrt{axt-at^2}\,\big)f(t)\,dt$
23	$\widetilde{f}\big(\sqrt{p}\big)$	$\displaystyle\int_0^\infty \frac{t}{2\sqrt{\pi x^3}}\exp\Big(-\frac{t^2}{4x}\Big)f(t)\,dt$
24	$\dfrac{1}{\sqrt{p}}\widetilde{f}\big(\sqrt{p}\big)$	$\dfrac{1}{\sqrt{\pi x}}\displaystyle\int_0^\infty \exp\Big(-\frac{t^2}{4x}\Big)f(t)\,dt$
25	$\widetilde{f}\big(p+\sqrt{p}\big)$	$\dfrac{1}{2\sqrt{\pi}}\displaystyle\int_0^x \frac{t}{(x-t)^{3/2}}\exp\Big[-\frac{t^2}{4(x-t)}\Big]f(t)\,dt$
26	$\widetilde{f}\big(\sqrt{p^2+a^2}\,\big)$	$f(x)-a\displaystyle\int_0^x f\big(\sqrt{x^2-t^2}\,\big)J_1(at)\,dt$
27	$\widetilde{f}\big(\sqrt{p^2-a^2}\,\big)$	$f(x)+a\displaystyle\int_0^x f\big(\sqrt{x^2-t^2}\,\big)I_1(at)\,dt$
28	$\dfrac{\widetilde{f}\big(\sqrt{p^2+a^2}\,\big)}{\sqrt{p^2+a^2}}$	$\displaystyle\int_0^x J_0\big(a\sqrt{x^2-t^2}\,\big)f(t)\,dt$
29	$\dfrac{\widetilde{f}\big(\sqrt{p^2-a^2}\,\big)}{\sqrt{p^2-a^2}}$	$\displaystyle\int_0^x I_0\big(a\sqrt{x^2-t^2}\,\big)f(t)\,dt$
30	$\widetilde{f}\big(\sqrt{(p+a)^2-b^2}\,\big)$	$e^{-ax}f(x)+be^{-ax}\displaystyle\int_0^x f\big(\sqrt{x^2-t^2}\,\big)I_1(bt)\,dt$
31	$\widetilde{f}(\ln p)$	$\displaystyle\int_0^\infty \frac{x^{t-1}}{\Gamma(t)}f(t)\,dt$
32	$\dfrac{1}{p}\widetilde{f}(\ln p)$	$\displaystyle\int_0^\infty \frac{x^{t}}{\Gamma(t+1)}f(t)\,dt$
33	$\widetilde{f}(p-ia)+\widetilde{f}(p+ia),\;\; i^2=-1$	$2f(x)\cos(ax)$
34	$i\big[\widetilde{f}(p-ia)-\widetilde{f}(p+ia)\big],\;\; i^2=-1$	$2f(x)\sin(ax)$

No.	Laplace transform, $\widetilde{f}(p)$	Inverse transform, $f(x)=\dfrac{1}{2\pi i}\displaystyle\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}(p)\,dp$
35	$\dfrac{d\widetilde{f}(p)}{dp}$	$-xf(x)$
36	$\dfrac{d^n\widetilde{f}(p)}{dp^n}$	$(-x)^n f(x)$
37	$p^n\dfrac{d^m\widetilde{f}(p)}{dp^m},\quad m\ge n$	$(-1)^m\dfrac{d^n}{dx^n}\bigl[x^m f(x)\bigr]$
38	$\displaystyle\int_p^\infty \widetilde{f}(q)\,dq$	$\dfrac{1}{x}f(x)$
39	$\dfrac{1}{p}\displaystyle\int_0^p \widetilde{f}(q)\,dq$	$\displaystyle\int_x^\infty \frac{f(t)}{t}\,dt$
40	$\dfrac{1}{p}\displaystyle\int_p^\infty \widetilde{f}(q)\,dq$	$\displaystyle\int_0^x \frac{f(t)}{t}\,dt$

S2.2.2. Expressions Involving Rational Functions

No.	Laplace transform, $\widetilde{f}(p)$	Inverse transform, $f(x)=\dfrac{1}{2\pi i}\displaystyle\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}(p)\,dp$
1	$\dfrac{1}{p}$	1
2	$\dfrac{1}{p+a}$	e^{-ax}
3	$\dfrac{1}{p^2}$	x
4	$\dfrac{1}{p(p+a)}$	$\dfrac{1}{a}\bigl(1-e^{-ax}\bigr)$
5	$\dfrac{1}{(p+a)^2}$	xe^{-ax}
6	$\dfrac{p}{(p+a)^2}$	$(1-ax)e^{-ax}$
7	$\dfrac{1}{p^2-a^2}$	$\dfrac{1}{a}\sinh(ax)$
8	$\dfrac{p}{p^2-a^2}$	$\cosh(ax)$
9	$\dfrac{1}{(p+a)(p+b)}$	$\dfrac{1}{a-b}\bigl(e^{-bx}-e^{-ax}\bigr)$
10	$\dfrac{p}{(p+a)(p+b)}$	$\dfrac{1}{a-b}\bigl(ae^{-ax}-be^{-bx}\bigr)$
11	$\dfrac{1}{p^2+a^2}$	$\dfrac{1}{a}\sin(ax)$
12	$\dfrac{p}{p^2+a^2}$	$\cos(ax)$
13	$\dfrac{1}{(p+b)^2+a^2}$	$\dfrac{1}{a}e^{-bx}\sin(ax)$
14	$\dfrac{p}{(p+b)^2+a^2}$	$e^{-bx}\Bigl[\cos(ax)-\dfrac{b}{a}\sin(ax)\Bigr]$

No.	Laplace transform, $\widetilde{f}(p)$	Inverse transform, $f(x) = \dfrac{1}{2\pi i}\displaystyle\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}(p)\,dp$
15	$\dfrac{1}{p^3}$	$\frac{1}{2}x^2$
16	$\dfrac{1}{p^2(p+a)}$	$\dfrac{1}{a^2}\left(e^{-ax}+ax-1\right)$
17	$\dfrac{1}{p(p+a)(p+b)}$	$\dfrac{1}{ab(a-b)}\left(a-b+be^{-ax}-ae^{-bx}\right)$
18	$\dfrac{1}{p(p+a)^2}$	$\dfrac{1}{a^2}\left(1-e^{-ax}-axe^{-ax}\right)$
19	$\dfrac{1}{(p+a)(p+b)(p+c)}$	$\dfrac{(c-b)e^{-ax}+(a-c)e^{-bx}+(b-a)e^{-cx}}{(a-b)(b-c)(c-a)}$
20	$\dfrac{p}{(p+a)(p+b)(p+c)}$	$\dfrac{a(b-c)e^{-ax}+b(c-a)e^{-bx}+c(a-b)e^{-cx}}{(a-b)(b-c)(c-a)}$
21	$\dfrac{p^2}{(p+a)(p+b)(p+c)}$	$\dfrac{a^2(c-b)e^{-ax}+b^2(a-c)e^{-bx}+c^2(b-a)e^{-cx}}{(a-b)(b-c)(c-a)}$
22	$\dfrac{1}{(p+a)(p+b)^2}$	$\dfrac{1}{(a-b)^2}\left[e^{-ax}-e^{-bx}+(a-b)xe^{-bx}\right]$
23	$\dfrac{p}{(p+a)(p+b)^2}$	$\dfrac{1}{(a-b)^2}\left\{-ae^{-ax}+\left[a+b(b-a)x\right]e^{-bx}\right\}$
24	$\dfrac{p^2}{(p+a)(p+b)^2}$	$\dfrac{1}{(a-b)^2}\left[a^2e^{-ax}+b(b-2a-b^2x+abx)e^{-bx}\right]$
25	$\dfrac{1}{(p+a)^3}$	$\frac{1}{2}x^2e^{-ax}$
26	$\dfrac{p}{(p+a)^3}$	$x\left(1-\frac{1}{2}ax\right)e^{-ax}$
27	$\dfrac{p^2}{(p+a)^3}$	$\left(1-2ax+\frac{1}{2}a^2x^2\right)e^{-ax}$
28	$\dfrac{1}{p(p^2+a^2)}$	$\dfrac{1}{a^2}\left[1-\cos(ax)\right]$
29	$\dfrac{1}{p\left[(p+b)^2+a^2\right]}$	$\dfrac{1}{a^2+b^2}\left\{1-e^{-bx}\left[\cos(ax)+\dfrac{b}{a}\sin(ax)\right]\right\}$
30	$\dfrac{1}{(p+a)(p^2+b^2)}$	$\dfrac{1}{a^2+b^2}\left[e^{-ax}+\dfrac{a}{b}\sin(bx)-\cos(bx)\right]$
31	$\dfrac{p}{(p+a)(p^2+b^2)}$	$\dfrac{1}{a^2+b^2}\left[-ae^{-ax}+a\cos(bx)+b\sin(bx)\right]$
32	$\dfrac{p^2}{(p+a)(p^2+b^2)}$	$\dfrac{1}{a^2+b^2}\left[a^2e^{-ax}-ab\sin(bx)+b^2\cos(bx)\right]$
33	$\dfrac{1}{p^3+a^3}$	$\dfrac{1}{3a^2}e^{-ax}-\dfrac{1}{3a^2}e^{ax/2}\left[\cos(kx)-\sqrt{3}\sin(kx)\right]$, $k=\frac{1}{2}a\sqrt{3}$
34	$\dfrac{p}{p^3+a^3}$	$-\dfrac{1}{3a}e^{-ax}+\dfrac{1}{3a}e^{ax/2}\left[\cos(kx)+\sqrt{3}\sin(kx)\right]$, $k=\frac{1}{2}a\sqrt{3}$

No.	Laplace transform, $\widetilde{f}(p)$	Inverse transform, $f(x)=\dfrac{1}{2\pi i}\displaystyle\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}(p)\,dp$
35	$\dfrac{p^2}{p^3+a^3}$	$\frac{1}{3}e^{-ax}+\frac{2}{3}e^{ax/2}\cos(kx),\quad k=\frac{1}{2}a\sqrt{3}$
36	$\dfrac{1}{(p+a)\big[(p+b)^2+c^2\big]}$	$\dfrac{e^{-ax}-e^{-bx}\cos(cx)+ke^{-bx}\sin(cx)}{(a-b)^2+c^2},\quad k=\dfrac{a-b}{c}$
37	$\dfrac{p}{(p+a)\big[(p+b)^2+c^2\big]}$	$\dfrac{-ae^{-ax}+ae^{-bx}\cos(cx)+ke^{-bx}\sin(cx)}{(a-b)^2+c^2},$ $k=\dfrac{b^2+c^2-ab}{c}$
38	$\dfrac{p^2}{(p+a)\big[(p+b)^2+c^2\big]}$	$\dfrac{a^2e^{-ax}+(b^2+c^2-2ab)e^{-bx}\cos(cx)+ke^{-bx}\sin(cx)}{(a-b)^2+c^2},$ $k=-ac-bc+\dfrac{ab^2-b^3}{c}$
39	$\dfrac{1}{p^4}$	$\frac{1}{6}x^3$
40	$\dfrac{1}{p^3(p+a)}$	$\dfrac{1}{a^3}-\dfrac{1}{a^2}x+\dfrac{1}{2a}x^2-\dfrac{1}{a^3}e^{-ax}$
41	$\dfrac{1}{p^2(p+a)^2}$	$\dfrac{1}{a^2}x\big(1+e^{-ax}\big)+\dfrac{2}{a^3}\big(e^{-ax}-1\big)$
42	$\dfrac{1}{p^2(p+a)(p+b)}$	$-\dfrac{a+b}{a^2b^2}+\dfrac{1}{ab}x+\dfrac{1}{a^2(b-a)}e^{-ax}+\dfrac{1}{b^2(a-b)}e^{-bx}$
43	$\dfrac{1}{(p+a)^2(p+b)^2}$	$\dfrac{1}{(a-b)^2}\left[e^{-ax}\left(x+\dfrac{2}{a-b}\right)+e^{-bx}\left(x-\dfrac{2}{a-b}\right)\right]$
44	$\dfrac{1}{(p+a)^4}$	$\frac{1}{6}x^3e^{-ax}$
45	$\dfrac{p}{(p+a)^4}$	$\frac{1}{2}x^2e^{-ax}-\frac{1}{6}ax^3e^{-ax}$
46	$\dfrac{1}{p^2(p^2+a^2)}$	$\dfrac{1}{a^3}\big[ax-\sin(ax)\big]$
47	$\dfrac{1}{p^4-a^4}$	$\dfrac{1}{2a^3}\big[\sinh(ax)-\sin(ax)\big]$
48	$\dfrac{p}{p^4-a^4}$	$\dfrac{1}{2a^2}\big[\cosh(ax)-\cos(ax)\big]$
49	$\dfrac{p^2}{p^4-a^4}$	$\dfrac{1}{2a}\big[\sinh(ax)+\sin(ax)\big]$
50	$\dfrac{p^3}{p^4-a^4}$	$\dfrac{1}{2}\big[\cosh(ax)+\cos(ax)\big]$
51	$\dfrac{1}{p^4+a^4}$	$\dfrac{1}{a^3\sqrt{2}}\big(\cosh\xi\sin\xi-\sinh\xi\cos\xi\big),\quad \xi=\dfrac{ax}{\sqrt{2}}$
52	$\dfrac{p}{p^4+a^4}$	$\dfrac{1}{a^2}\sin\Big(\dfrac{ax}{\sqrt{2}}\Big)\sinh\Big(\dfrac{ax}{\sqrt{2}}\Big)$
53	$\dfrac{p^2}{p^4+a^4}$	$\dfrac{1}{a\sqrt{2}}\big(\cos\xi\sinh\xi+\sin\xi\cosh\xi\big),\quad \xi=\dfrac{ax}{\sqrt{2}}$

No.	Laplace transform, $\widetilde{f}(p)$	Inverse transform, $f(x) = \dfrac{1}{2\pi i}\displaystyle\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}(p)\,dp$
54	$\dfrac{1}{(p^2+a^2)^2}$	$\dfrac{1}{2a^3}\bigl[\sin(ax) - ax\cos(ax)\bigr]$
55	$\dfrac{p}{(p^2+a^2)^2}$	$\dfrac{1}{2a}x\sin(ax)$
56	$\dfrac{p^2}{(p^2+a^2)^2}$	$\dfrac{1}{2a}\bigl[\sin(ax) + ax\cos(ax)\bigr]$
57	$\dfrac{p^3}{(p^2+a^2)^2}$	$\cos(ax) - \frac{1}{2}ax\sin(ax)$
58	$\dfrac{1}{\bigl[(p+b)^2+a^2\bigr]^2}$	$\dfrac{1}{2a^3}e^{-bx}\bigl[\sin(ax) - ax\cos(ax)\bigr]$
59	$\dfrac{1}{(p^2-a^2)(p^2-b^2)}$	$\dfrac{1}{a^2-b^2}\left[\dfrac{1}{a}\sinh(ax) - \dfrac{1}{b}\sinh(bx)\right]$
60	$\dfrac{p}{(p^2-a^2)(p^2-b^2)}$	$\dfrac{\cosh(ax) - \cosh(bx)}{a^2-b^2}$
61	$\dfrac{p^2}{(p^2-a^2)(p^2-b^2)}$	$\dfrac{a\sinh(ax) - b\sinh(bx)}{a^2-b^2}$
62	$\dfrac{p^3}{(p^2-a^2)(p^2-b^2)}$	$\dfrac{a^2\cosh(ax) - b^2\cosh(bx)}{a^2-b^2}$
63	$\dfrac{1}{(p^2+a^2)(p^2+b^2)}$	$\dfrac{1}{b^2-a^2}\left[\dfrac{1}{a}\sin(ax) - \dfrac{1}{b}\sin(bx)\right]$
64	$\dfrac{p}{(p^2+a^2)(p^2+b^2)}$	$\dfrac{\cos(ax) - \cos(bx)}{b^2-a^2}$
65	$\dfrac{p^2}{(p^2+a^2)(p^2+b^2)}$	$\dfrac{-a\sin(ax) + b\sin(bx)}{b^2-a^2}$
66	$\dfrac{p^3}{(p^2+a^2)(p^2+b^2)}$	$\dfrac{-a^2\cos(ax) + b^2\cos(bx)}{b^2-a^2}$
67	$\dfrac{1}{p^n},\quad n = 1,\,2,\,\dots$	$\dfrac{1}{(n-1)!}x^{n-1}$
68	$\dfrac{1}{(p+a)^n},\quad n = 1,\,2,\,\dots$	$\dfrac{1}{(n-1)!}x^{n-1}e^{-ax}$
69	$\dfrac{1}{p(p+a)^n},\quad n = 1,\,2,\,\dots$	$a^{-n}\bigl[1 - e^{-ax}e_n(ax)\bigr],\quad e_n(z) = 1 + \dfrac{z}{1!} + \cdots + \dfrac{z^n}{n!}$
70	$\dfrac{1}{p^{2n}+a^{2n}},\quad n = 1,\,2,\,\dots$	$-\dfrac{1}{na^{2n}}\displaystyle\sum_{k=1}^{n}\exp(a_kx)\bigl[a_k\cos(b_kx) - b_k\sin(b_kx)\bigr],$ $a_k = a\cos\varphi_k,\ b_k = a\sin\varphi_k,\ \varphi_k = \dfrac{\pi(2k-1)}{2n}$
71	$\dfrac{1}{p^{2n}-a^{2n}},\quad n = 1,\,2,\,\dots$	$\dfrac{1}{na^{2n-1}}\sinh(ax) + \dfrac{1}{na^{2n}}\displaystyle\sum_{k=2}^{n}\exp(a_kx)$ $\times\bigl[a_k\cos(b_kx) - b_k\sin(b_kx)\bigr],$ $a_k = a\cos\varphi_k,\ b_k = a\sin\varphi_k,\ \varphi_k = \dfrac{\pi(k-1)}{n}$

No.	Laplace transform, $\widetilde{f}(p)$	Inverse transform, $f(x)=\dfrac{1}{2\pi i}\displaystyle\int_{c-i\infty}^{c+i\infty}e^{px}\widetilde{f}(p)\,dp$
72	$\dfrac{1}{p^{2n+1}+a^{2n+1}}$, $\quad n=0,\,1,\,\ldots$	$\dfrac{e^{-ax}}{(2n+1)a^{2n}}-\dfrac{2}{(2n+1)a^{2n+1}}\displaystyle\sum_{k=1}^{n}\exp(a_kx)$ $\times\bigl[a_k\cos(b_kx)-b_k\sin(b_kx)\bigr]$, $a_k=a\cos\varphi_k,\ b_k=a\sin\varphi_k,\ \varphi_k=\dfrac{\pi(2k-1)}{2n+1}$
73	$\dfrac{1}{p^{2n+1}-a^{2n+1}}$, $\quad n=0,\,1,\,\ldots$	$\dfrac{e^{ax}}{(2n+1)a^{2n}}+\dfrac{2}{(2n+1)a^{2n+1}}\displaystyle\sum_{k=1}^{n}\exp(a_kx)$ $\times\bigl[a_k\cos(b_kx)-b_k\sin(b_kx)\bigr]$, $a_k=a\cos\varphi_k,\ b_k=a\sin\varphi_k,\ \varphi_k=\dfrac{2\pi k}{2n+1}$
74	$\dfrac{Q(p)}{P(p)}$, $P(p)=(p-a_1)\ldots(p-a_n)$; $Q(p)$ is a polynomial of degree $\le n-1$; $\ a_i\ne a_j$ if $\ i\ne j$	$\displaystyle\sum_{k=1}^{n}\frac{Q(a_k)}{P'(a_k)}\exp\bigl(a_kx\bigr)$, (the prime stands for the differentiation)
75	$\dfrac{Q(p)}{P(p)}$, $P(p)=(p-a_1)^{m_1}\ldots(p-a_n)^{m_n}$; $Q(p)$ is a polynomial of degree $< m_1+m_2+\cdots+m_n-1$; $a_i\ne a_j$ if $\ i\ne j$	$\displaystyle\sum_{k=1}^{n}\sum_{l=1}^{m_k}\frac{\Phi_{kl}(a_k)}{(m_k-l)!\,(l-1)!}x^{m_k-l}\exp\bigl(a_kx\bigr)$, $\Phi_{kl}(p)=\dfrac{d^{l-1}}{dp^{l-1}}\left[\dfrac{Q(p)}{P_k(p)}\right],\quad P_k(p)=\dfrac{P(p)}{(p-a_k)^{m_k}}$
76	$\dfrac{Q(p)+pR(p)}{P(p)}$, $P(p)=(p^2+a_1^2)\ldots(p^2+a_n^2)$; $Q(p)$ and $R(p)$ are polynomials of degree $\le 2n-2$; $\ a_l\ne a_j,\ l\ne j$	$\displaystyle\sum_{k=1}^{n}\frac{Q(ia_k)\sin(a_kx)+a_kR(ia_k)\cos(a_kx)}{a_kP_k(ia_k)}$, $P_m(p)=\dfrac{P(p)}{p^2+a_m^2},\quad i^2=-1$

S2.2.3. Expressions Involving Square Roots

No.	Laplace transform, $\widetilde{f}(p)$	Inverse transform, $f(x)=\dfrac{1}{2\pi i}\displaystyle\int_{c-i\infty}^{c+i\infty}e^{px}\widetilde{f}(p)\,dp$
1	$\dfrac{1}{\sqrt{p}}$	$\dfrac{1}{\sqrt{\pi x}}$
2	$\sqrt{p-a}-\sqrt{p-b}$	$\dfrac{e^{bx}-e^{ax}}{2\sqrt{\pi x^3}}$
3	$\dfrac{1}{\sqrt{p+a}}$	$\dfrac{1}{\sqrt{\pi x}}e^{-ax}$
4	$\sqrt{\dfrac{p+a}{p}}-1$	$\frac{1}{2}ae^{-ax/2}\bigl[I_1\bigl(\frac{1}{2}ax\bigr)+I_0\bigl(\frac{1}{2}ax\bigr)\bigr]$
5	$\dfrac{\sqrt{p+a}}{p+b}$	$\dfrac{e^{-ax}}{\sqrt{\pi x}}+(a-b)^{1/2}e^{-bx}\operatorname{erf}\bigl[(a-b)^{1/2}x^{1/2}\bigr]$
6	$\dfrac{1}{p\sqrt{p}}$	$2\sqrt{\dfrac{x}{\pi}}$
7	$\dfrac{1}{(p+a)\sqrt{p+b}}$	$(b-a)^{-1/2}e^{-ax}\operatorname{erf}\bigl[(b-a)^{1/2}x^{1/2}\bigr]$

No.	Laplace transform, $\widetilde{f}(p)$	Inverse transform, $f(x)=\dfrac{1}{2\pi i}\displaystyle\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}(p)\,dp$
8	$\dfrac{1}{\sqrt{p}\,(p-a)}$	$\dfrac{1}{\sqrt{a}}e^{ax}\operatorname{erf}\bigl(\sqrt{ax}\,\bigr)$
9	$\dfrac{1}{p^{3/2}(p-a)}$	$a^{-3/2}e^{ax}\operatorname{erf}\bigl(\sqrt{ax}\,\bigr)-2a^{-1}\pi^{-1/2}x^{1/2}$
10	$\dfrac{1}{\sqrt{p}+a}$	$\pi^{-1/2}x^{-1/2}-ae^{a^2x}\operatorname{erfc}\bigl(a\sqrt{x}\,\bigr)$
11	$\dfrac{a}{p\bigl(\sqrt{p}+a\bigr)}$	$1-e^{a^2x}\operatorname{erfc}\bigl(a\sqrt{x}\,\bigr)$
12	$\dfrac{1}{p+a\sqrt{p}}$	$e^{a^2x}\operatorname{erfc}\bigl(a\sqrt{x}\,\bigr)$
13	$\dfrac{1}{\bigl(\sqrt{p}+\sqrt{a}\,\bigr)^2}$	$1-\dfrac{2}{\sqrt{\pi}}(ax)^{1/2}+(1-2ax)e^{ax}\bigl[\operatorname{erf}\bigl(\sqrt{ax}\,\bigr)-1\bigr]$
14	$\dfrac{1}{p\bigl(\sqrt{p}+\sqrt{a}\,\bigr)^2}$	$\dfrac{1}{a}+\Bigl(2x-\dfrac{1}{a}\Bigr)e^{ax}\operatorname{erfc}\bigl(\sqrt{ax}\,\bigr)-\dfrac{2}{\sqrt{\pi a}}\sqrt{x}$
15	$\dfrac{1}{\sqrt{p}\,\bigl(\sqrt{p}+a\bigr)^2}$	$2\pi^{-1/2}x^{1/2}-2axe^{a^2x}\operatorname{erfc}\bigl(a\sqrt{x}\,\bigr)$
16	$\dfrac{1}{\bigl(\sqrt{p}+a\bigr)^3}$	$\dfrac{2}{\sqrt{\pi}}(a^2x+1)\sqrt{x}-ax(2a^2x+3)e^{a^2x}\operatorname{erfc}\bigl(a\sqrt{x}\,\bigr)$
17	$p^{-n-1/2},\quad n=1,\,2,\,\ldots$	$\dfrac{2^n}{1\times3\times\ldots\times(2n-1)\sqrt{\pi}}\,x^{n-1/2}$
18	$(p+a)^{-n-1/2}$	$\dfrac{2^n}{1\times3\times\ldots\times(2n-1)\sqrt{\pi}}\,x^{n-1/2}e^{-ax}$
19	$\dfrac{1}{\sqrt{p^2+a^2}}$	$J_0(ax)$
20	$\dfrac{1}{\sqrt{p^2-a^2}}$	$I_0(ax)$
21	$\dfrac{1}{\sqrt{p^2+ap+b}}$	$\exp\bigl(-\tfrac{1}{2}ax\bigr)\,J_0\bigl[\bigl(b-\tfrac{1}{4}a^2\bigr)^{1/2}\,x\bigr]$
22	$\bigl(\sqrt{p^2+a^2}-p\bigr)^{1/2}$	$\dfrac{1}{\sqrt{2\pi x^3}}\sin(ax)$
23	$\dfrac{1}{\sqrt{p^2+a^2}}\bigl(\sqrt{p^2+a^2}+p\bigr)^{1/2}$	$\dfrac{\sqrt{2}}{\sqrt{\pi x}}\cos(ax)$
24	$\dfrac{1}{\sqrt{p^2-a^2}}\bigl(\sqrt{p^2-a^2}+p\bigr)^{1/2}$	$\dfrac{\sqrt{2}}{\sqrt{\pi x}}\cosh(ax)$
25	$\bigl(\sqrt{p^2+a^2}+p\bigr)^{-n}$	$na^{-n}x^{-1}J_n(ax)$
26	$\bigl(\sqrt{p^2-a^2}+p\bigr)^{-n}$	$na^{-n}x^{-1}I_n(ax)$
27	$\bigl(p^2+a^2\bigr)^{-n-1/2}$	$\dfrac{(x/a)^nJ_n(ax)}{1\times3\times5\times\ldots\times(2n-1)}$
28	$\bigl(p^2-a^2\bigr)^{-n-1/2}$	$\dfrac{(x/a)^nI_n(ax)}{1\times3\times5\times\ldots\times(2n-1)}$

S2.2.4. Expressions Involving Arbitrary Powers

No.	Laplace transform, $\widetilde{f}(p)$	Inverse transform, $f(x) = \dfrac{1}{2\pi i}\displaystyle\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}(p)\,dp$
1	$(p+a)^{-\nu}, \quad \nu > 0$	$\dfrac{1}{\Gamma(\nu)} x^{\nu-1} e^{-ax}$
2	$\bigl[(p+a)^{1/2} + (p+b)^{1/2}\bigr]^{-2\nu}, \quad \nu > 0$	$\dfrac{\nu}{(a-b)^{\nu}} x^{-1} \exp\bigl[-\tfrac{1}{2}(a+b)x\bigr] I_{\nu}\bigl[\tfrac{1}{2}(a-b)x\bigr]$
3	$\bigl[(p+a)(p+b)\bigr]^{-\nu}, \quad \nu > 0$	$\dfrac{\sqrt{\pi}}{\Gamma(\nu)}\Bigl(\dfrac{x}{a-b}\Bigr)^{\nu-1/2} \exp\Bigl(-\dfrac{a+b}{2}x\Bigr) I_{\nu-1/2}\Bigl(\dfrac{a-b}{2}x\Bigr)$
4	$\bigl(p^2+a^2\bigr)^{-\nu-1/2}, \quad \nu > -\frac{1}{2}$	$\dfrac{\sqrt{\pi}}{(2a)^{\nu}\Gamma(\nu+\frac{1}{2})} x^{\nu} J_{\nu}(ax)$
5	$\bigl(p^2-a^2\bigr)^{-\nu-1/2}, \quad \nu > -\frac{1}{2}$	$\dfrac{\sqrt{\pi}}{(2a)^{\nu}\Gamma(\nu+\frac{1}{2})} x^{\nu} I_{\nu}(ax)$
6	$p\bigl(p^2+a^2\bigr)^{-\nu-1/2}, \quad \nu > 0$	$\dfrac{a\sqrt{\pi}}{(2a)^{\nu}\Gamma\bigl(\nu+\frac{1}{2}\bigr)} x^{\nu} J_{\nu-1}(ax)$
7	$p\bigl(p^2-a^2\bigr)^{-\nu-1/2}, \quad \nu > 0$	$\dfrac{a\sqrt{\pi}}{(2a)^{\nu}\Gamma\bigl(\nu+\frac{1}{2}\bigr)} x^{\nu} I_{\nu-1}(ax)$
8	$\bigl[(p^2+a^2)^{1/2}+p\bigr]^{-\nu} =$ $a^{-2\nu}\bigl[(p^2+a^2)^{1/2}-p\bigr]^{\nu}, \quad \nu > 0$	$\nu a^{-\nu} x^{-1} J_{\nu}(ax)$
9	$\bigl[(p^2-a^2)^{1/2}+p\bigr]^{-\nu} =$ $a^{-2\nu}\bigl[p-(p^2-a^2)^{1/2}\bigr]^{\nu}, \quad \nu > 0$	$\nu a^{-\nu} x^{-1} I_{\nu}(ax)$
10	$p\bigl[(p^2+a^2)^{1/2}+p\bigr]^{-\nu}, \quad \nu > 1$	$\nu a^{1-\nu} x^{-1} J_{\nu-1}(ax) - \nu(\nu+1)a^{-\nu} x^{-2} J_{\nu}(ax)$
11	$p\bigl[(p^2-a^2)^{1/2}+p\bigr]^{-\nu}, \quad \nu > 1$	$\nu a^{1-\nu} x^{-1} I_{\nu-1}(ax) - \nu(\nu+1)a^{-\nu} x^{-2} I_{\nu}(ax)$
12	$\dfrac{\bigl(\sqrt{p^2+a^2}+p\bigr)^{-\nu}}{\sqrt{p^2+a^2}}, \quad \nu > -1$	$a^{-\nu} J_{\nu}(ax)$
13	$\dfrac{\bigl(\sqrt{p^2-a^2}+p\bigr)^{-\nu}}{\sqrt{p^2-a^2}}, \quad \nu > -1$	$a^{-\nu} I_{\nu}(ax)$

S2.2.5. Expressions Involving Exponential Functions

No.	Laplace transform, $\widetilde{f}(p)$	Inverse transform, $f(x) = \dfrac{1}{2\pi i}\displaystyle\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}(p)\,dp$
1	$p^{-1}e^{-ap}, \quad a > 0$	$\begin{cases} 0 & \text{if } 0 < x < a, \\ 1 & \text{if } a < x \end{cases}$
2	$p^{-1}\bigl(1-e^{-ap}\bigr), \quad a > 0$	$\begin{cases} 1 & \text{if } 0 < x < a, \\ 0 & \text{if } a < x \end{cases}$
3	$p^{-1}\bigl(e^{-ap}-e^{-bp}\bigr), \quad 0 \le a < b$	$\begin{cases} 0 & \text{if } 0 < x < a, \\ 1 & \text{if } a < x < b, \\ 0 & \text{if } b < x \end{cases}$
4	$p^{-2}\bigl(e^{-ap}-e^{-bp}\bigr), \quad 0 \le a < b$	$\begin{cases} 0 & \text{if } 0 < x < a, \\ x-a & \text{if } a < x < b, \\ b-a & \text{if } b < x \end{cases}$
5	$(p+b)^{-1}e^{-ap}, \quad a > 0$	$\begin{cases} 0 & \text{if } 0 < x < a, \\ e^{-b(x-a)} & \text{if } a < x \end{cases}$

No.	Laplace transform, $\widetilde{f}(p)$	Inverse transform, $f(x)=\dfrac{1}{2\pi i}\displaystyle\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}(p)\,dp$
6	$p^{-\nu}e^{-ap},\quad \nu>0$	$\begin{cases} 0 & \text{if } 0<x<a, \\ \dfrac{(x-a)^{\nu-1}}{\Gamma(\nu)} & \text{if } a<x \end{cases}$
7	$p^{-1}\bigl(e^{ap}-1\bigr)^{-1},\quad a>0$	$f(x)=n \;\text{ if }\; na<x<(n+1)a;\quad n=0,\,1,\,2,\,\dots$
8	$e^{a/p}-1$	$\sqrt{\dfrac{a}{x}}\,I_1\bigl(2\sqrt{ax}\,\bigr)$
9	$p^{-1/2}e^{a/p}$	$\dfrac{1}{\sqrt{\pi x}}\cosh\bigl(2\sqrt{ax}\,\bigr)$
10	$p^{-3/2}e^{a/p}$	$\dfrac{1}{\sqrt{\pi a}}\sinh\bigl(2\sqrt{ax}\,\bigr)$
11	$p^{-5/2}e^{a/p}$	$\sqrt{\dfrac{x}{\pi a}}\,\cosh\bigl(2\sqrt{ax}\,\bigr)-\dfrac{1}{2\sqrt{\pi a^3}}\,\sinh\bigl(2\sqrt{ax}\,\bigr)$
12	$p^{-\nu-1}e^{a/p},\quad \nu>-1$	$(x/a)^{\nu/2}I_\nu(2\sqrt{ax}\,)$
13	$1-e^{-a/p}$	$\sqrt{\dfrac{a}{x}}\,J_1\bigl(2\sqrt{ax}\,\bigr)$
14	$p^{-1/2}e^{-a/p}$	$\dfrac{1}{\sqrt{\pi x}}\cos\bigl(2\sqrt{ax}\,\bigr)$
15	$p^{-3/2}e^{-a/p}$	$\dfrac{1}{\sqrt{\pi a}}\sin\bigl(2\sqrt{ax}\,\bigr)$
16	$p^{-5/2}e^{-a/p}$	$\dfrac{1}{2\sqrt{\pi a^3}}\,\sin\bigl(2\sqrt{ax}\,\bigr)-\sqrt{\dfrac{x}{\pi a}}\,\cos\bigl(2\sqrt{ax}\,\bigr)$
17	$p^{-\nu-1}e^{-a/p},\quad \nu>-1$	$(x/a)^{\nu/2}J_\nu(2\sqrt{ax}\,)$
18	$\exp\bigl(-\sqrt{ap}\,\bigr),\quad a>0$	$\dfrac{\sqrt{a}}{2\sqrt{\pi}}\,x^{-3/2}\exp\Bigl(-\dfrac{a}{4x}\Bigr)$
19	$p\exp\bigl(-\sqrt{ap}\,\bigr),\quad a>0$	$\dfrac{\sqrt{a}}{8\sqrt{\pi}}\,(a-6x)x^{-7/2}\exp\Bigl(-\dfrac{a}{4x}\Bigr)$
20	$\dfrac{1}{p}\exp\bigl(-\sqrt{ap}\,\bigr),\quad a\ge 0$	$\operatorname{erfc}\Bigl(\dfrac{\sqrt{a}}{2\sqrt{x}}\Bigr)$
21	$\sqrt{p}\exp\bigl(-\sqrt{ap}\,\bigr),\quad a>0$	$\dfrac{1}{4\sqrt{\pi}}\,(a-2x)x^{-5/2}\exp\Bigl(-\dfrac{a}{4x}\Bigr)$
22	$\dfrac{1}{\sqrt{p}}\exp\bigl(-\sqrt{ap}\,\bigr),\quad a\ge 0$	$\dfrac{1}{\sqrt{\pi x}}\exp\Bigl(-\dfrac{a}{4x}\Bigr)$
23	$\dfrac{1}{p\sqrt{p}}\exp\bigl(-\sqrt{ap}\,\bigr),\quad a\ge 0$	$\dfrac{2\sqrt{x}}{\sqrt{\pi}}\exp\Bigl(-\dfrac{a}{4x}\Bigr)-\sqrt{a}\operatorname{erfc}\Bigl(\dfrac{\sqrt{a}}{2\sqrt{x}}\Bigr)$
24	$\dfrac{\exp\bigl(-k\sqrt{p^2+a^2}\,\bigr)}{\sqrt{p^2+a^2}},\quad k>0$	$\begin{cases} 0 & \text{if } 0<x<k, \\ J_0\bigl(a\sqrt{x^2-k^2}\,\bigr) & \text{if } k<x \end{cases}$
25	$\dfrac{\exp\bigl(-k\sqrt{p^2-a^2}\,\bigr)}{\sqrt{p^2-a^2}},\quad k>0$	$\begin{cases} 0 & \text{if } 0<x<k, \\ I_0\bigl(a\sqrt{x^2-k^2}\,\bigr) & \text{if } k<x \end{cases}$

S2.2.6. Expressions Involving Hyperbolic Functions

No.	Laplace transform, $\widetilde{f}(p)$	Inverse transform, $f(x)=\dfrac{1}{2\pi i}\displaystyle\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}(p)\,dp$
1	$\dfrac{1}{p\sinh(ap)}$, $a>0$	$f(x)=2n$ if $a(2n-1)<x<a(2n+1)$; $n=0,\ 1,\ 2,\ \dots$ $(x>0)$
2	$\dfrac{1}{p^2\sinh(ap)}$, $a>0$	$f(x)=2n(x-an)$ if $a(2n-1)<x<a(2n+1)$; $n=0,\ 1,\ 2,\ \dots$ $(x>0)$
3	$\dfrac{\sinh(a/p)}{\sqrt{p}}$	$\dfrac{1}{2\sqrt{\pi x}}\bigl[\cosh\bigl(2\sqrt{ax}\bigr)-\cos\bigl(2\sqrt{ax}\bigr)\bigr]$
4	$\dfrac{\sinh(a/p)}{p\sqrt{p}}$	$\dfrac{1}{2\sqrt{\pi a}}\bigl[\sinh\bigl(2\sqrt{ax}\bigr)-\sin\bigl(2\sqrt{ax}\bigr)\bigr]$
5	$p^{-\nu-1}\sinh(a/p)$, $\nu>-2$	$\frac{1}{2}(x/a)^{\nu/2}\bigl[I_\nu\bigl(2\sqrt{ax}\bigr)-J_\nu\bigl(2\sqrt{ax}\bigr)\bigr]$
6	$\dfrac{1}{p\cosh(ap)}$, $a>0$	$f(x)=\begin{cases}0 & \text{if } a(4n-1)<x<a(4n+1),\\ 2 & \text{if } a(4n+1)<x<a(4n+3),\end{cases}$ $n=0,\ 1,\ 2,\ \dots$ $(x>0)$
7	$\dfrac{1}{p^2\cosh(ap)}$, $a>0$	$x-(-1)^n(x-2an)$ if $2n-1<x/a<2n+1$; $n=0,\ 1,\ 2,\ \dots$ $(x>0)$
8	$\dfrac{\cosh(a/p)}{\sqrt{p}}$	$\dfrac{1}{2\sqrt{\pi x}}\bigl[\cosh\bigl(2\sqrt{ax}\bigr)+\cos\bigl(2\sqrt{ax}\bigr)\bigr]$
9	$\dfrac{\cosh(a/p)}{p\sqrt{p}}$	$\dfrac{1}{2\sqrt{\pi a}}\bigl[\sinh\bigl(2\sqrt{ax}\bigr)+\sin\bigl(2\sqrt{ax}\bigr)\bigr]$
10	$p^{-\nu-1}\cosh(a/p)$, $\nu>-1$	$\frac{1}{2}(x/a)^{\nu/2}\bigl[I_\nu\bigl(2\sqrt{ax}\bigr)+J_\nu\bigl(2\sqrt{ax}\bigr)\bigr]$
11	$\dfrac{1}{p}\tanh(ap)$, $a>0$	$f(x)=(-1)^{n-1}$ if $2a(n-1)<x<2an$; $n=1,\ 2,\ \dots$
12	$\dfrac{1}{p}\coth(ap)$, $a>0$	$f(x)=(2n-1)$ if $2a(n-1)<x<2an$; $n=1,\ 2,\ \dots$
13	$\operatorname{arccoth}(p/a)$	$\dfrac{1}{x}\sinh(ax)$

S2.2.7. Expressions Involving Logarithmic Functions

No.	Laplace transform, $\widetilde{f}(p)$	Inverse transform, $f(x)=\dfrac{1}{2\pi i}\displaystyle\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}(p)\,dp$
1	$\dfrac{1}{p}\ln p$	$-\ln x-\mathcal{C}$, $\mathcal{C}=0.5772\ldots$ is the Euler constant
2	$p^{-n-1}\ln p$	$\bigl(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}-\ln x-\mathcal{C}\bigr)\dfrac{x^n}{n!}$, $\mathcal{C}=0.5772\ldots$ is the Euler constant
3	$p^{-n-1/2}\ln p$	$k_n\bigl[2+\frac{2}{3}+\frac{2}{5}+\cdots+\frac{2}{2n-1}-\ln(4x)-\mathcal{C}\bigr]x^{n-1/2}$, $k_n=\dfrac{2^n}{1\times3\times5\times\ldots\times(2n-1)\sqrt{\pi}}$, $\mathcal{C}=0.5772\ldots$
4	$p^{-\nu}\ln p$, $\nu>0$	$\dfrac{1}{\Gamma(\nu)}x^{\nu-1}\bigl[\psi(\nu)-\ln x\bigr]$, $\psi(\nu)$ is the logarithmic derivative of the gamma function
5	$\dfrac{1}{p}(\ln p)^2$	$(\ln x+\mathcal{C})^2-\frac{1}{6}\pi^2$, $\mathcal{C}=0.5772\ldots$

No.	Laplace transform, $\widetilde{f}(p)$	Inverse transform, $f(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}(p)\,dp$
6	$\frac{1}{p^2}(\ln p)^2$	$x\left[(\ln x + \mathcal{C} - 1)^2 + 1 - \frac{1}{6}\pi^2\right]$
7	$\frac{\ln(p+b)}{p+a}$	$e^{-ax}\left\{\ln(b-a) - \operatorname{Ei}\left[(a-b)x\right]\right\}$
8	$\frac{\ln p}{p^2+a^2}$	$\frac{1}{a}\cos(ax)\operatorname{Si}(ax) + \frac{1}{a}\sin(ax)\left[\ln a - \operatorname{Ci}(ax)\right]$
9	$\frac{p\ln p}{p^2+a^2}$	$\cos(ax)\left[\ln a - \operatorname{Ci}(ax)\right] - \sin(ax)\operatorname{Si}(ax)]$
10	$\ln\frac{p+b}{p+a}$	$\frac{1}{x}\left(e^{-ax} - e^{-bx}\right)$
11	$\ln\frac{p^2+b^2}{p^2+a^2}$	$\frac{2}{x}\left[\cos(ax) - \cos(bx)\right]$
12	$p\ln\frac{p^2+b^2}{p^2+a^2}$	$\frac{2}{x}\left[\cos(bx) + bx\sin(bx) - \cos(ax) - ax\sin(ax)\right]$
13	$\ln\frac{(p+a)^2+k^2}{(p+b)^2+k^2}$	$\frac{2}{x}\cos(kx)(e^{-bx} - e^{-ax})$
14	$p\ln\left(\frac{1}{p}\sqrt{p^2+a^2}\right)$	$\frac{1}{x^2}\left[\cos(ax) - 1\right] + \frac{a}{x}\sin(ax)$
15	$p\ln\left(\frac{1}{p}\sqrt{p^2-a^2}\right)$	$\frac{1}{x^2}\left[\cosh(ax) - 1\right] - \frac{a}{x}\sinh(ax)$

S2.2.8. Expressions Involving Trigonometric Functions

No.	Laplace transform, $\widetilde{f}(p)$	Inverse transform, $f(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}(p)\,dp$
1	$\frac{\sin(a/p)}{\sqrt{p}}$	$\frac{1}{\sqrt{\pi x}}\sinh\left(\sqrt{2ax}\right)\sin\left(\sqrt{2ax}\right)$
2	$\frac{\sin(a/p)}{p\sqrt{p}}$	$\frac{1}{\sqrt{\pi a}}\cosh\left(\sqrt{2ax}\right)\sin\left(\sqrt{2ax}\right)$
3	$\frac{\cos(a/p)}{\sqrt{p}}$	$\frac{1}{\sqrt{\pi x}}\cosh\left(\sqrt{2ax}\right)\cos\left(\sqrt{2ax}\right)$
4	$\frac{\cos(a/p)}{p\sqrt{p}}$	$\frac{1}{\sqrt{\pi a}}\sinh\left(\sqrt{2ax}\right)\cos\left(\sqrt{2ax}\right)$
5	$\frac{1}{\sqrt{p}}\exp\left(-\sqrt{ap}\right)\sin\left(\sqrt{ap}\right)$	$\frac{1}{\sqrt{\pi x}}\sin\left(\frac{a}{2x}\right)$
6	$\frac{1}{\sqrt{p}}\exp\left(-\sqrt{ap}\right)\cos\left(\sqrt{ap}\right)$	$\frac{1}{\sqrt{\pi x}}\cos\left(\frac{a}{2x}\right)$
7	$\arctan\frac{a}{p}$	$\frac{1}{x}\sin(ax)$
8	$\frac{1}{p}\arctan\frac{a}{p}$	$\operatorname{Si}(ax)$
9	$p\arctan\frac{a}{p} - a$	$\frac{1}{x^2}\left[ax\cos(ax) - \sin(ax)\right]$
10	$\arctan\frac{2ap}{p^2+b^2}$	$\frac{2}{x}\sin(ax)\cos\left(x\sqrt{a^2+b^2}\right)$

S2.2.9. Expressions Involving Special Functions

No.	Laplace transform, $\widetilde{f}(p)$	Inverse transform, $f(x) = \dfrac{1}{2\pi i}\displaystyle\int_{c-i\infty}^{c+i\infty} e^{px}\widetilde{f}(p)\,dp$
1	$\exp\bigl(ap^2\bigr)\operatorname{erfc}\bigl(p\sqrt{a}\,\bigr)$	$\dfrac{1}{\sqrt{\pi a}}\exp\Bigl(-\dfrac{x^2}{4a}\Bigr)$
2	$\dfrac{1}{p}\exp\bigl(ap^2\bigr)\operatorname{erfc}\bigl(p\sqrt{a}\,\bigr)$	$\operatorname{erf}\Bigl(\dfrac{x}{2\sqrt{a}}\Bigr)$
3	$\operatorname{erfc}\bigl(\sqrt{ap}\,\bigr), \quad a > 0$	$\begin{cases} 0 & \text{if } 0 < x < a, \\ \dfrac{\sqrt{a}}{\pi x\sqrt{x-a}} & \text{if } a < x \end{cases}$
4	$\dfrac{1}{\sqrt{p}}\operatorname{erfc}\bigl(\sqrt{ap}\,\bigr), \quad a > 0$	$\begin{cases} 0 & \text{if } 0 < x < a, \\ \dfrac{1}{\sqrt{\pi x}} & \text{if } x > a \end{cases}$
5	$e^{ap}\operatorname{erfc}\bigl(\sqrt{ap}\,\bigr)$	$\dfrac{\sqrt{a}}{\pi\sqrt{x}\,(x+a)}$
6	$\dfrac{1}{\sqrt{p}}e^{ap}\operatorname{erfc}\bigl(\sqrt{ap}\,\bigr)$	$\dfrac{1}{\sqrt{\pi(x+a)}}$
7	$\dfrac{1}{\sqrt{p}}\operatorname{erf}\bigl(\sqrt{ap}\,\bigr), \quad a > 0$	$\begin{cases} \dfrac{1}{\sqrt{\pi x}} & \text{if } 0 < x < a, \\ 0 & \text{if } x > a \end{cases}$
8	$\operatorname{erf}\bigl(\sqrt{a/p}\,\bigr)$	$\dfrac{1}{\pi x}\sin\bigl(2\sqrt{ax}\,\bigr)$
9	$\dfrac{1}{\sqrt{p}}\exp(a/p)\operatorname{erf}\bigl(\sqrt{a/p}\,\bigr)$	$\dfrac{1}{\sqrt{\pi x}}\sinh\bigl(2\sqrt{ax}\,\bigr)$
10	$\dfrac{1}{\sqrt{p}}\exp(a/p)\operatorname{erfc}\bigl(\sqrt{a/p}\,\bigr)$	$\dfrac{1}{\sqrt{\pi x}}\exp\bigl(-2\sqrt{ax}\,\bigr)$
11	$p^{-a}\gamma(a, bp), \quad a, b > 0$	$\begin{cases} x^{a-1} & \text{if } 0 < x < b, \\ 0 & \text{if } b < x \end{cases}$
12	$\gamma(a, b/p), \quad a > 0$	$b^{a/2}x^{a/2-1}J_a\bigl(2\sqrt{bx}\,\bigr)$
13	$a^{-p}\gamma(p, a)$	$\exp\bigl(-ae^{-x}\bigr)$
14	$K_0(ap), \quad a > 0$	$\begin{cases} 0 & \text{if } 0 < x < a, \\ (x^2-a^2)^{-1/2} & \text{if } a < x \end{cases}$
15	$K_\nu(ap), \quad a > 0$	$\begin{cases} 0 & \text{if } 0 < x < a, \\ \dfrac{\cosh\bigl[\nu\operatorname{arccosh}(x/a)\bigr]}{\sqrt{x^2-a^2}} & \text{if } a < x \end{cases}$
16	$K_0\bigl(a\sqrt{p}\bigr)$	$\dfrac{1}{2x}\exp\Bigl(-\dfrac{a^2}{4x}\Bigr)$
17	$\dfrac{1}{\sqrt{p}}K_1\bigl(a\sqrt{p}\,\bigr)$	$\dfrac{1}{a}\exp\Bigl(-\dfrac{a^2}{4x}\Bigr)$

S2.3. Tables of Fourier Cosine Transforms

S2.3.1. General Formulas

No.	Original function, $f(x)$	Cosine transform, $\check{f}_c(u)=\int_0^\infty f(x)\cos(ux)\,dx$
1	$af_1(x)+bf_2(x)$	$a\check{f}_{1c}(u)+b\check{f}_{2c}(u)$
2	$f(ax),\quad a>0$	$\frac{1}{a}\check{f}_c\left(\frac{u}{a}\right)$
3	$x^{2n}f(x),\quad n=1,\,2,\,\dots$	$(-1)^n\frac{d^{2n}}{du^{2n}}\check{f}_c(u)$
4	$x^{2n+1}f(ax),\quad n=0,\,1,\,\dots$	$(-1)^n\frac{d^{2n+1}}{du^{2n+1}}\check{f}_s(u),\quad \check{f}_s(u)=\int_0^\infty f(x)\sin(xu)\,dx$
5	$f(ax)\cos(bx),\quad a,b>0$	$\frac{1}{2a}\left[\check{f}_c\left(\frac{u+b}{a}\right)+\check{f}_c\left(\frac{u-b}{a}\right)\right]$

S2.3.2. Expressions Involving Power Functions

No.	Original function, $f(x)$	Cosine transform, $\check{f}_c(u)=\int_0^\infty f(x)\cos(ux)\,dx$
1	$\begin{cases}1 & \text{if } 0<x<a,\\ 0 & \text{if } a<x\end{cases}$	$\frac{1}{u}\sin(au)$
2	$\begin{cases}x & \text{if } 0<x<1,\\ 2-x & \text{if } 1<x<2,\\ 0 & \text{if } 2<x\end{cases}$	$\frac{4}{u^2}\cos u\,\sin^2\frac{u}{2}$
3	$\frac{1}{a+x},\quad a>0$	$-\sin(au)\,\mathrm{si}(au)-\cos(au)\,\mathrm{Ci}(au)$
4	$\frac{1}{a^2+x^2},\quad a>0$	$\frac{\pi}{2a}e^{-au}$
5	$\frac{1}{a^2-x^2},\quad a>0$	$\frac{\pi\sin(au)}{2u}$ (the integral is understood in the sense of Cauchy principal value)
6	$\frac{a}{a^2+(b+x)^2}+\frac{a}{a^2+(b-x)^2}$	$\pi e^{-au}\cos(bu)$
7	$\frac{b+x}{a^2+(b+x)^2}+\frac{b-x}{a^2+(b-x)^2}$	$\pi e^{-au}\sin(bu)$
8	$\frac{1}{a^4+x^4},\quad a>0$	$\frac{1}{2}\pi a^{-3}\exp\left(-\frac{au}{\sqrt{2}}\right)\sin\left(\frac{\pi}{4}+\frac{au}{\sqrt{2}}\right)$
9	$\frac{1}{(a^2+x^2)(b^2+x^2)},\quad a,b>0$	$\frac{\pi}{2}\frac{ae^{-bu}-be^{-au}}{ab(a^2-b^2)}$
10	$\frac{x^{2m}}{(x^2+a)^{n+1}}$, $n,m=1,\,2,\,\dots;\quad n+1>m\ge 0$	$(-1)^{n+m}\frac{\pi}{2n!}\frac{\partial^n}{\partial a^n}\left(a^{1/\sqrt{m}}e^{-u\sqrt{a}}\right)$
11	$\frac{1}{\sqrt{x}}$	$\sqrt{\frac{\pi}{2u}}$
12	$\begin{cases}\frac{1}{\sqrt{x}} & \text{if } 0<x<a,\\ 0 & \text{if } a<x\end{cases}$	$2\sqrt{\frac{\pi}{2u}}\,C(au)$, $C(u)$ is the Fresnel integral

No.	Original function, $f(x)$	Cosine transform, $\check{f}_{\rm c}(u)=\int_0^\infty f(x)\cos(ux)\,dx$
13	$\begin{cases} 0 & \text{if } 0<x<a, \\ \dfrac{1}{\sqrt{x}} & \text{if } a<x \end{cases}$	$\sqrt{\dfrac{\pi}{2u}}\,\bigl[1-2C(au)\bigr]$, $C(u)$ is the Fresnel integral
14	$\begin{cases} 0 & \text{if } 0<x<a, \\ \dfrac{1}{\sqrt{x-a}} & \text{if } a<x \end{cases}$	$\sqrt{\dfrac{\pi}{2u}}\,\bigl[\cos(au)-\sin(au)\bigr]$
15	$\dfrac{1}{\sqrt{a^2+x^2}}$	$K_0(au)$
16	$\begin{cases} \dfrac{1}{\sqrt{a^2-x^2}} & \text{if } 0<x<a, \\ 0 & \text{if } a<x \end{cases}$	$\dfrac{\pi}{2}\,J_0(au)$
17	$(a^2+x^2)^{-1/2}\bigl[(a^2+x^2)^{1/2}+a\bigr]^{1/2}$	$(2u/\pi)^{-1/2}e^{-au}$, $\quad a>0$
18	$x^{-\nu}$, $\quad 0<\nu<1$	$\sin\bigl(\frac{1}{2}\pi\nu\bigr)\Gamma(1-\nu)u^{\nu-1}$

S2.3.3. Expressions Involving Exponential Functions

No.	Original function, $f(x)$	Cosine transform, $\check{f}_{\rm c}(u)=\int_0^\infty f(x)\cos(ux)\,dx$
1	e^{-ax}	$\dfrac{a}{a^2+u^2}$
2	$\dfrac{1}{x}\bigl(e^{-ax}-e^{-bx}\bigr)$	$\dfrac{1}{2}\ln\dfrac{b^2+u^2}{a^2+u^2}$
3	$\sqrt{x}e^{-ax}$	$\frac{1}{2}\sqrt{\pi}\,(a^2+u^2)^{-3/4}\cos\Bigl(\frac{3}{2}\arctan\dfrac{u}{a}\Bigr)$
4	$\dfrac{1}{\sqrt{x}}e^{-ax}$	$\sqrt{\dfrac{\pi}{2}}\Bigl[\dfrac{a+(a^2+u^2)^{1/2}}{a^2+u^2}\Bigr]^{1/2}$
5	x^ne^{-ax}, $\quad n=1,\,2,\,\ldots$	$\dfrac{a^{n+1}n!}{(a^2+u^2)^{n+1}}\sum\limits_{0\le 2k\le n+1}(-1)^kC_{n+1}^{2k}\Bigl(\dfrac{u}{a}\Bigr)^{2k}$
6	$x^{n-1/2}e^{-ax}$, $\quad n=1,\,2,\,\ldots$	$k_nu\dfrac{\partial^n}{\partial a^n}\dfrac{1}{r\sqrt{r-a}}$, where $\;r=\sqrt{a^2+u^2}$, $k_n=(-1)^n\sqrt{\pi/2}$
7	$x^{\nu-1}e^{-ax}$	$\Gamma(\nu)(a^2+u^2)^{-\nu/2}\cos\Bigl(\nu\arctan\dfrac{u}{a}\Bigr)$
8	$\dfrac{x}{e^{ax}-1}$	$\dfrac{1}{2u^2}-\dfrac{\pi^2}{2a^2\sinh^2\bigl(\pi a^{-1}u\bigr)}$
9	$\dfrac{1}{x}\Bigl(\dfrac{1}{2}-\dfrac{1}{x}+\dfrac{1}{e^x-1}\Bigr)$	$-\dfrac{1}{2}\ln\bigl(1-e^{-2\pi u}\bigr)$
10	$\exp\bigl(-ax^2\bigr)$	$\dfrac{1}{2}\sqrt{\dfrac{\pi}{a}}\exp\Bigl(-\dfrac{u^2}{4a}\Bigr)$
11	$\dfrac{1}{\sqrt{x}}\exp\Bigl(-\dfrac{a}{x}\Bigr)$	$\sqrt{\dfrac{\pi}{2u}}\,e^{-\sqrt{2au}}\bigl[\cos\bigl(\sqrt{2au}\,\bigr)-\sin\bigl(\sqrt{2au}\,\bigr)\bigr]$
12	$\dfrac{1}{x\sqrt{x}}\exp\Bigl(-\dfrac{a}{x}\Bigr)$	$\sqrt{\dfrac{\pi}{a}}\,e^{-\sqrt{2au}}\cos\bigl(\sqrt{2au}\,\bigr)$

S2.3.4. Expressions Involving Hyperbolic Functions

No.	Original function, $f(x)$	Cosine transform, $\check{f}_{\mathrm{c}}(u)=\int_0^\infty f(x)\cos(ux)\,dx$
1	$\dfrac{1}{\cosh(ax)}, \quad a>0$	$\dfrac{\pi}{2a\cosh\left(\frac{1}{2}\pi a^{-1}u\right)}$
2	$\dfrac{1}{\cosh^2(ax)}, \quad a>0$	$\dfrac{\pi u}{2a^2\sinh\left(\frac{1}{2}\pi a^{-1}u\right)}$
3	$\dfrac{\cosh(ax)}{\cosh(bx)}, \quad \lvert a\rvert<b$	$\dfrac{\pi}{b}\left[\dfrac{\cos\left(\frac{1}{2}\pi ab^{-1}\right)\cosh\left(\frac{1}{2}\pi b^{-1}u\right)}{\cos\left(\pi ab^{-1}\right)+\cosh\left(\pi b^{-1}u\right)}\right]$
4	$\dfrac{1}{\cosh(ax)+\cos b}$	$\dfrac{\pi\sinh\left(a^{-1}bu\right)}{a\sin b\sinh\left(\pi a^{-1}u\right)}$
5	$\exp\left(-ax^2\right)\cosh(bx), \quad a>0$	$\dfrac{1}{2}\sqrt{\dfrac{\pi}{a}}\exp\left(\dfrac{b^2-u^2}{4a}\right)\cos\left(\dfrac{abu}{2}\right)$
6	$\dfrac{x}{\sinh(ax)}$	$\dfrac{\pi^2}{4a^2\cosh^2\left(\frac{1}{2}\pi a^{-1}u\right)}$
7	$\dfrac{\sinh(ax)}{\sinh(bx)}, \quad \lvert a\rvert<b$	$\dfrac{\pi}{2b}\,\dfrac{\sin\left(\pi ab^{-1}\right)}{\cos\left(\pi ab^{-1}\right)+\cosh\left(\pi b^{-1}u\right)}$
8	$\dfrac{1}{x}\tanh(ax), \quad a>0$	$\ln\left[\coth\left(\frac{1}{4}\pi a^{-1}u\right)\right]$

S2.3.5. Expressions Involving Logarithmic Functions

No.	Original function, $f(x)$	Cosine transform, $\check{f}_{\mathrm{c}}(u)=\int_0^\infty f(x)\cos(ux)\,dx$
1	$\begin{cases}\ln x & \text{if } 0<x<1,\\ 0 & \text{if } 1<x\end{cases}$	$-\dfrac{1}{u}\operatorname{Si}(u)$
2	$\dfrac{\ln x}{\sqrt{x}}$	$-\sqrt{\dfrac{\pi}{2u}}\left[\ln(4u)+\mathcal{C}+\dfrac{\pi}{2}\right]$, $\mathcal{C}=0.5772\ldots$ is the Euler constant
3	$x^{\nu-1}\ln x, \quad 0<\nu<1$	$\Gamma(\nu)\cos\left(\dfrac{\pi\nu}{2}\right)u^{-\nu}\left[\psi(\nu)-\dfrac{\pi}{2}\tan\left(\dfrac{\pi\nu}{2}\right)-\ln u\right]$
4	$\ln\left\lvert\dfrac{a+x}{a-x}\right\rvert, \quad a>0$	$\dfrac{2}{u}\left[\cos(au)\operatorname{Si}(au)-\sin(au)\operatorname{Ci}(au)\right]$
5	$\ln\left(1+a^2/x^2\right), \quad a>0$	$\dfrac{\pi}{u}\left(1-e^{-au}\right)$
6	$\ln\dfrac{a^2+x^2}{b^2+x^2}, \quad a,b>0$	$\dfrac{\pi}{u}\left(e^{-bu}-e^{-au}\right)$
7	$e^{-ax}\ln x, \quad a>0$	$-\dfrac{a\mathcal{C}+\frac{1}{2}a\ln(u^2+a^2)+u\arctan(u/a)}{u^2+a^2}$, see row 2 for $\mathcal{C}$
8	$\ln\left(1+e^{-ax}\right), \quad a>0$	$\dfrac{a}{2u^2}-\dfrac{\pi}{2u\sinh\left(\pi a^{-1}u\right)}$
9	$\ln\left(1-e^{-ax}\right), \quad a>0$	$\dfrac{a}{2u^2}-\dfrac{\pi}{2u}\coth\left(\pi a^{-1}u\right)$

S2.3.6. Expressions Involving Trigonometric Functions

No.	Original function, $f(x)$	Cosine transform, $\check{f}_{\rm c}(u)=\int_0^\infty f(x)\cos(ux)\,dx$
1	$\dfrac{\sin(ax)}{x},\quad a>0$	$\begin{cases}\frac{1}{2}\pi & \text{if } u<a,\\ \frac{1}{4}\pi & \text{if } u=a,\\ 0 & \text{if } u>a\end{cases}$
2	$x^{\nu-1}\sin(ax),\quad a>0,\ \lvert\nu\rvert<1$	$\pi\dfrac{(u+a)^{-\nu}-\lvert u+a\rvert^{-\nu}\operatorname{sign}(u-a)}{4\Gamma(1-\nu)\cos\left(\frac{1}{2}\pi\nu\right)}$
3	$\dfrac{x\sin(ax)}{x^2+b^2},\quad a,b>0$	$\begin{cases}\frac{1}{2}\pi e^{-ab}\cosh(bu) & \text{if } u<a,\\ -\frac{1}{2}\pi e^{-bu}\sinh(ab) & \text{if } u>a\end{cases}$
4	$\dfrac{\sin(ax)}{x(x^2+b^2)},\quad a,b>0$	$\begin{cases}\frac{1}{2}\pi b^{-2}\left[1-e^{-ab}\cosh(bu)\right] & \text{if } u<a,\\ \frac{1}{2}\pi b^{-2}e^{-bu}\sinh(ab) & \text{if } u>a\end{cases}$
5	$e^{-bx}\sin(ax),\quad a,b>0$	$\dfrac{1}{2}\left[\dfrac{a+u}{(a+u)^2+b^2}+\dfrac{a-u}{(a-u)^2+b^2}\right]$
6	$\dfrac{1}{x}\sin^2(ax),\quad a>0$	$\dfrac{1}{4}\ln\left\lvert 1-4\dfrac{a^2}{u^2}\right\rvert$
7	$\dfrac{1}{x^2}\sin^2(ax),\quad a>0$	$\begin{cases}\frac{1}{4}\pi(2a-u) & \text{if } u<2a,\\ 0 & \text{if } u>2a\end{cases}$
8	$\dfrac{1}{x}\sin\left(\dfrac{a}{x}\right),\quad a>0$	$\dfrac{\pi}{2}J_0\left(2\sqrt{au}\right)$
9	$\dfrac{1}{\sqrt{x}}\sin\left(a\sqrt{x}\right)\sin\left(b\sqrt{x}\right),\ a,b>0$	$\sqrt{\dfrac{\pi}{u}}\sin\left(\dfrac{ab}{2u}\right)\sin\left(\dfrac{a^2+b^2}{4u}-\dfrac{\pi}{4}\right)$
10	$\sin\left(ax^2\right),\quad a>0$	$\sqrt{\dfrac{\pi}{8a}}\left[\cos\left(\dfrac{u^2}{4a}\right)-\sin\left(\dfrac{u^2}{4a}\right)\right]$
11	$\exp\left(-ax^2\right)\sin\left(bx^2\right),\quad a>0$	$\dfrac{\sqrt{\pi}}{(A^2+B^2)^{1/4}}\exp\left(-\dfrac{Au^2}{A^2+B^2}\right)\sin\left(\varphi-\dfrac{Bu^2}{A^2+B^2}\right),$ $A=4a,\ B=4b,\ \varphi=\frac{1}{2}\arctan(b/a)$
12	$\dfrac{1-\cos(ax)}{x},\quad a>0$	$\dfrac{1}{2}\ln\left\lvert 1-\dfrac{a^2}{u^2}\right\rvert$
13	$\dfrac{1-\cos(ax)}{x^2},\quad a>0$	$\begin{cases}\frac{1}{2}\pi(a-u) & \text{if } u<a,\\ 0 & \text{if } u>a\end{cases}$
14	$x^{\nu-1}\cos(ax),\quad a>0,\ 0<\nu<1$	$\frac{1}{2}\Gamma(\nu)\cos\left(\frac{1}{2}\pi\nu\right)\left[\lvert u-a\rvert^{-\nu}+(u+a)^{-\nu}\right]$
15	$\dfrac{\cos(ax)}{x^2+b^2},\quad a,b>0$	$\begin{cases}\frac{1}{2}\pi b^{-1}e^{-ab}\cosh(bu) & \text{if } u<a,\\ \frac{1}{2}\pi b^{-1}e^{-bu}\cosh(ab) & \text{if } u>a\end{cases}$
16	$e^{-bx}\cos(ax),\quad a,b>0$	$\dfrac{b}{2}\left[\dfrac{1}{(a+u)^2+b^2}+\dfrac{1}{(a-u)^2+b^2}\right]$
17	$\dfrac{1}{\sqrt{x}}\cos\left(a\sqrt{x}\right)$	$\sqrt{\dfrac{\pi}{u}}\sin\left(\dfrac{a^2}{4u}+\dfrac{\pi}{4}\right)$
18	$\dfrac{1}{\sqrt{x}}\cos\left(a\sqrt{x}\right)\cos\left(b\sqrt{x}\right)$	$\sqrt{\dfrac{\pi}{u}}\cos\left(\dfrac{ab}{2u}\right)\sin\left(\dfrac{a^2+b^2}{4u}+\dfrac{\pi}{4}\right)$
19	$\exp\left(-bx^2\right)\cos(ax),\quad b>0$	$\dfrac{1}{2}\sqrt{\dfrac{\pi}{b}}\exp\left(-\dfrac{a^2+u^2}{4b}\right)\cosh\left(\dfrac{au}{2b}\right)$
20	$\cos\left(ax^2\right),\quad a>0$	$\sqrt{\dfrac{\pi}{8a}}\left[\cos\left(\frac{1}{4}a^{-1}u^2\right)+\sin\left(\frac{1}{4}a^{-1}u^2\right)\right]$
21	$\exp\left(-ax^2\right)\cos\left(bx^2\right),\quad a>0$	$\dfrac{\sqrt{\pi}}{(A^2+B^2)^{1/4}}\exp\left(-\dfrac{Au^2}{A^2+B^2}\right)\cos\left(\varphi-\dfrac{Bu^2}{A^2+B^2}\right),$ $A=4a,\ B=4b,\ \varphi=\frac{1}{2}\arctan(b/a)$

S2.3.7. Expressions Involving Special Functions

No.	Original function, $f(x)$	Cosine transform, $\check{f}_{\rm c}(u) = \int_0^\infty f(x)\cos(ux)\,dx$
1	$\operatorname{Ei}(-ax)$	$-\dfrac{1}{u}\arctan\left(\dfrac{u}{a}\right)$
2	$\operatorname{Ci}(ax)$	$\begin{cases} 0 & \text{if } 0 < u < a, \\ -\frac{\pi}{2u} & \text{if } a < u \end{cases}$
3	$\operatorname{si}(ax)$	$-\dfrac{1}{2u}\ln\left\lvert\dfrac{u+a}{u-a}\right\rvert, \quad u \neq a$
4	$J_0(ax), \quad a > 0$	$\begin{cases} (a^2-u^2)^{-1/2} & \text{if } 0 < u < a, \\ 0 & \text{if } a < u \end{cases}$
5	$J_\nu(ax), \quad a > 0,\ \nu > -1$	$\begin{cases} \dfrac{\cos\left[\nu\arcsin(u/a)\right]}{\sqrt{a^2-u^2}} & \text{if } 0 < u < a, \\ -\dfrac{a^\nu\sin(\pi\nu/2)}{\xi(u+\xi)^\nu} & \text{if } a < u, \end{cases}$ where $\xi = \sqrt{u^2-a^2}$
6	$\dfrac{1}{x}J_\nu(ax), \quad a > 0,\ \nu > 0$	$\begin{cases} \nu^{-1}\cos\left[\nu\arcsin(u/a)\right] & \text{if } 0 < u < a, \\ \dfrac{a^\nu\cos(\pi\nu/2)}{\nu\left(u+\sqrt{u^2-a^2}\right)^\nu} & \text{if } a < u \end{cases}$
7	$x^{-\nu}J_\nu(ax), \quad a > 0,\ \nu > -\frac{1}{2}$	$\begin{cases} \dfrac{\sqrt{\pi}\,\left(a^2-u^2\right)^{\nu-1/2}}{(2a)^\nu\Gamma\left(\nu+\frac{1}{2}\right)} & \text{if } 0 < u < a, \\ 0 & \text{if } a < u \end{cases}$
8	$x^{\nu+1}J_\nu(ax)$, $a > 0,\ -1 < \nu < -\frac{1}{2}$	$\begin{cases} 0 & \text{if } 0 < u < a, \\ \dfrac{2^{\nu+1}\sqrt{\pi}\,a^\nu u}{\Gamma\left(-\nu-\frac{1}{2}\right)(u^2-a^2)^{\nu+3/2}} & \text{if } a < u \end{cases}$
9	$J_0\left(a\sqrt{x}\right), \quad a > 0$	$\dfrac{1}{u}\sin\left(\dfrac{a^2}{4u}\right)$
10	$\dfrac{1}{\sqrt{x}}J_1\left(a\sqrt{x}\right), \quad a > 0$	$\dfrac{4}{a}\sin^2\left(\dfrac{a^2}{8u}\right)$
11	$x^{\nu/2}J_\nu\left(a\sqrt{x}\right),\ a > 0,\ -1 < \nu < \frac{1}{2}$	$\left(\dfrac{a}{2}\right)^\nu u^{-\nu-1}\sin\left(\dfrac{a^2}{4u}-\dfrac{\pi\nu}{2}\right)$
12	$J_0\left(a\sqrt{x^2+b^2}\right)$	$\begin{cases} \dfrac{\cos\left(b\sqrt{a^2-u^2}\right)}{\sqrt{a^2-u^2}} & \text{if } 0 < u < a, \\ 0 & \text{if } a < u \end{cases}$
13	$Y_0(ax), \quad a > 0$	$\begin{cases} 0 & \text{if } 0 < u < a, \\ -(u^2-a^2)^{-1/2} & \text{if } a < u \end{cases}$
14	$x^\nu Y_\nu(ax), \quad a > 0,\ \lvert\nu\rvert < \frac{1}{2}$	$\begin{cases} 0 & \text{if } 0 < u < a, \\ -\dfrac{(2a)^\nu\sqrt{\pi}}{\Gamma\left(\frac{1}{2}-\nu\right)(u^2-a^2)^{\nu+1/2}} & \text{if } a < u \end{cases}$
15	$K_0\left(a\sqrt{x^2+b^2}\right), \quad a, b > 0$	$\dfrac{\pi}{2\sqrt{u^2+a^2}}\exp\left(-b\sqrt{u^2+a^2}\right)$

S2.4. Tables of Fourier Sine Transforms

S2.4.1. General Formulas

No.	Original function, $f(x)$	Sine transform, $\check{f}_{\rm s}(u)=\int_0^\infty f(x)\sin(ux)\,dx$
1	$af_1(x)+bf_2(x)$	$a\check{f}_{1\rm s}(u)+b\check{f}_{2\rm s}(u)$
2	$f(ax),\quad a>0$	$\frac{1}{a}\check{f}_{\rm s}\left(\frac{u}{a}\right)$
3	$x^{2n}f(x),\quad n=1,\,2,\,\dots$	$(-1)^n\frac{d^{2n}}{du^{2n}}\check{f}_{\rm s}(u)$
4	$x^{2n+1}f(ax),\quad n=0,\,1,\,\dots$	$(-1)^{n+1}\frac{d^{2n+1}}{du^{2n+1}}\check{f}_{\rm c}(u),\quad \check{f}_{\rm c}(u)=\int_0^\infty f(x)\cos(xu)\,dx$
5	$f(ax)\cos(bx),\quad a,b>0$	$\frac{1}{2a}\left[\check{f}_{\rm s}\left(\frac{u+b}{a}\right)+\check{f}_{\rm s}\left(\frac{u-b}{a}\right)\right]$

S2.4.2. Expressions Involving Power Functions

No.	Original function, $f(x)$	Sine transform, $\check{f}_{\rm s}(u)=\int_0^\infty f(x)\sin(ux)\,dx$
1	$\begin{cases}1 & \text{if } 0<x<a,\\ 0 & \text{if } a<x\end{cases}$	$\frac{1}{u}\left[1-\cos(au)\right]$
2	$\begin{cases}x & \text{if } 0<x<1,\\ 2-x & \text{if } 1<x<2,\\ 0 & \text{if } 2<x\end{cases}$	$\frac{4}{u^2}\sin u\,\sin^2\frac{u}{2}$
3	$\frac{1}{x}$	$\frac{\pi}{2}$
4	$\frac{1}{a+x},\quad a>0$	$\sin(au)\,\mathrm{Ci}(au)-\cos(au)\,\mathrm{si}(au)$
5	$\frac{x}{a^2+x^2},\quad a>0$	$\frac{\pi}{2}e^{-au}$
6	$\frac{1}{x(a^2+x^2)},\quad a>0$	$\frac{\pi}{2a^2}\left(1-e^{-au}\right)$
7	$\frac{a}{a^2+(x-b)^2}-\frac{a}{a^2+(x+b)^2}$	$\pi e^{-au}\sin(bu)$
8	$\frac{x+b}{a^2+(x+b)^2}-\frac{x-b}{a^2+(x-b)^2}$	$\pi e^{-au}\cos(bu)$
9	$\frac{x}{(x^2+a^2)^n},\quad a>0,\ n=1,\,2,\dots$	$\frac{\pi u e^{-au}}{2^{2n-2}(n-1)!\,a^{2n-3}}\sum_{k=0}^{n-2}\frac{(2n-k-4)!}{k!\,(n-k-2)!}(2au)^k$
10	$\frac{x^{2m+1}}{(x^2+a)^{n+1}},$ $n,m=0,\,1,\dots;\ 0\le m\le n$	$(-1)^{n+m}\frac{\pi}{2n!}\frac{\partial^n}{\partial a^n}\left(a^m e^{-u\sqrt{a}}\right)$
11	$\frac{1}{\sqrt{x}}$	$\sqrt{\frac{\pi}{2u}}$
12	$\frac{1}{x\sqrt{x}}$	$\sqrt{2\pi u}$

No.	Original function, $f(x)$	Sine transform, $\check{f}_{\rm s}(u) = \int_0^\infty f(x)\sin(ux)\,dx$
13	$x(a^2+x^2)^{-3/2}$	$uK_0(au)$
14	$\dfrac{\left(\sqrt{a^2+x^2}-a\right)^{1/2}}{\sqrt{a^2+x^2}}$	$\sqrt{\dfrac{\pi}{2u}}\,e^{-au}$
15	$x^{-\nu}$, $\quad 0<\nu<2$	$\cos\left(\frac{1}{2}\pi\nu\right)\Gamma(1-\nu)u^{\nu-1}$

S2.4.3. Expressions Involving Exponential Functions

No.	Original function, $f(x)$	Sine transform, $\check{f}_{\rm s}(u) = \int_0^\infty f(x)\sin(ux)\,dx$
1	e^{-ax}, $\quad a>0$	$\dfrac{u}{a^2+u^2}$
2	$x^n e^{-ax}$, $\quad a>0,\ n=1,\,2,\,\dots$	$n!\left(\dfrac{a}{a^2+u^2}\right)^{n+1}\sum_{k=0}^{[n/2]}(-1)^k C_{n+1}^{2k+1}\left(\dfrac{u}{a}\right)^{2k+1}$
3	$\dfrac{1}{x}e^{-ax}$, $\quad a>0$	$\arctan\dfrac{u}{a}$
4	$\sqrt{x}e^{-ax}$, $\quad a>0$	$\dfrac{\sqrt{\pi}}{2}(a^2+u^2)^{-3/4}\sin\left(\dfrac{3}{2}\arctan\dfrac{u}{a}\right)$
5	$\dfrac{1}{\sqrt{x}}e^{-ax}$, $\quad a>0$	$\sqrt{\dfrac{\pi}{2}}\,\dfrac{\left(\sqrt{a^2+u^2}-a\right)^{1/2}}{\sqrt{a^2+u^2}}$
6	$\dfrac{1}{x\sqrt{x}}e^{-ax}$, $\quad a>0$	$\sqrt{2\pi}\,\left(\sqrt{a^2+u^2}-a\right)^{1/2}$
7	$x^{n-1/2}e^{-ax}$, $\quad a>0,\ n=1,\,2,\,\dots$	$(-1)^n\sqrt{\dfrac{\pi}{2}}\,\dfrac{\partial^n}{\partial a^n}\left[\dfrac{\left(\sqrt{a^2+u^2}-a\right)^{1/2}}{\sqrt{a^2+u^2}}\right]$
8	$x^{\nu-1}e^{-ax}$, $\quad a>0,\ \nu>-1$	$\Gamma(\nu)(a^2+u^2)^{-\nu/2}\sin\left(\nu\arctan\dfrac{u}{a}\right)$
9	$x^{-2}\left(e^{-ax}-e^{-bx}\right)$, $\quad a,b>0$	$\dfrac{u}{2}\ln\left(\dfrac{u^2+b^2}{u^2+a^2}\right)+b\arctan\left(\dfrac{u}{b}\right)-a\arctan\left(\dfrac{u}{a}\right)$
10	$\dfrac{1}{e^{ax}+1}$, $\quad a>0$	$\dfrac{1}{2u}-\dfrac{\pi}{2a\sinh(\pi u/a)}$
11	$\dfrac{1}{e^{ax}-1}$, $\quad a>0$	$\dfrac{\pi}{2a}\coth\left(\dfrac{\pi u}{a}\right)-\dfrac{1}{2u}$
12	$\dfrac{e^{x/2}}{e^x-1}$	$-\frac{1}{2}\tanh(\pi u)$
13	$x\exp\left(-ax^2\right)$	$\dfrac{\sqrt{\pi}}{4a^{3/2}}\,u\exp\left(-\dfrac{u^2}{4a}\right)$
14	$\dfrac{1}{x}\exp\left(-ax^2\right)$	$\dfrac{\pi}{2}\operatorname{erf}\left(\dfrac{u}{2\sqrt{a}}\right)$
15	$\dfrac{1}{\sqrt{x}}\exp\left(-\dfrac{a}{x}\right)$	$\sqrt{\dfrac{\pi}{2u}}\,e^{-\sqrt{2au}}\left[\cos\left(\sqrt{2au}\right)+\sin\left(\sqrt{2au}\right)\right]$
16	$\dfrac{1}{x\sqrt{x}}\exp\left(-\dfrac{a}{x}\right)$	$\sqrt{\dfrac{\pi}{a}}\,e^{-\sqrt{2au}}\sin\left(\sqrt{2au}\right)$

S2.4.4. Expressions Involving Hyperbolic Functions

No.	Original function, $f(x)$	Sine transform, $\check{f}_{\mathrm{s}}(u) = \int_0^\infty f(x)\sin(ux)\,dx$
1	$\dfrac{1}{\sinh(ax)}, \quad a>0$	$\dfrac{\pi}{2a}\tanh\bigl(\tfrac{1}{2}\pi a^{-1}u\bigr)$
2	$\dfrac{x}{\sinh(ax)}, \quad a>0$	$\dfrac{\pi^2\sinh\bigl(\frac{1}{2}\pi a^{-1}u\bigr)}{4a^2\cosh^2\bigl(\frac{1}{2}\pi a^{-1}u\bigr)}$
3	$\dfrac{1}{x}e^{-bx}\sinh(ax), \quad b>\lvert a\rvert$	$\frac{1}{2}\arctan\Bigl(\dfrac{2au}{u^2+b^2-a^2}\Bigr)$
4	$\dfrac{1}{x\cosh(ax)}, \quad a>0$	$\arctan\bigl[\sinh\bigl(\tfrac{1}{2}\pi a^{-1}u\bigr)\bigr]$
5	$1-\tanh\bigl(\tfrac{1}{2}ax\bigr), \quad a>0$	$\dfrac{1}{u}-\dfrac{\pi}{a\sinh\bigl(\pi a^{-1}u\bigr)}$
6	$\coth\bigl(\tfrac{1}{2}ax\bigr)-1, \quad a>0$	$\dfrac{\pi}{a}\coth\bigl(\pi a^{-1}u\bigr)-\dfrac{1}{u}$
7	$\dfrac{\cosh(ax)}{\sinh(bx)}, \quad \lvert a\rvert<b$	$\dfrac{\pi}{2b}\,\dfrac{\sinh\bigl(\pi b^{-1}u\bigr)}{\cos\bigl(\pi ab^{-1}\bigr)+\cosh\bigl(\pi b^{-1}u\bigr)}$
8	$\dfrac{\sinh(ax)}{\cosh(bx)}, \quad \lvert a\rvert<b$	$\dfrac{\pi}{b}\,\dfrac{\sin\bigl(\frac{1}{2}\pi ab^{-1}\bigr)\sinh\bigl(\frac{1}{2}\pi b^{-1}u\bigr)}{\cos\bigl(\pi ab^{-1}\bigr)+\cosh\bigl(\pi b^{-1}u\bigr)}$

S2.4.5. Expressions Involving Logarithmic Functions

No.	Original function, $f(x)$	Sine transform, $\check{f}_{\mathrm{s}}(u) = \int_0^\infty f(x)\sin(ux)\,dx$
1	$\begin{cases}\ln x & \text{if } 0<x<1,\\ 0 & \text{if } 1<x\end{cases}$	$\dfrac{1}{u}\bigl[\mathrm{Ci}(u)-\ln u-\mathcal{C}\bigr]$, $\mathcal{C}=0.5772\ldots$ is the Euler constant
2	$\dfrac{\ln x}{x}$	$-\frac{1}{2}\pi(\ln u+\mathcal{C})$
3	$\dfrac{\ln x}{\sqrt{x}}$	$-\sqrt{\dfrac{\pi}{2u}}\Bigl[\ln(4u)+\mathcal{C}-\dfrac{\pi}{2}\Bigr]$
4	$x^{\nu-1}\ln x, \quad \lvert\nu\rvert<1$	$\dfrac{\pi u^{-\nu}\bigl[\psi(\nu)+\frac{\pi}{2}\cot\bigl(\frac{\pi\nu}{2}\bigr)-\ln u\bigr]}{2\Gamma(1-\nu)\cos\bigl(\frac{\pi\nu}{2}\bigr)}$
5	$\ln\left\lvert\dfrac{a+x}{a-x}\right\rvert, \quad a>0$	$\dfrac{\pi}{u}\sin(au)$
6	$\ln\dfrac{(x+b)^2+a^2}{(x-b)^2+a^2}, \quad a,b>0$	$\dfrac{2\pi}{u}e^{-au}\sin(bu)$
7	$e^{-ax}\ln x, \quad a>0$	$\dfrac{a\arctan(u/a)-\frac{1}{2}u\ln(u^2+a^2)-e^{\mathcal{C}}u}{u^2+a^2}, \quad \mathcal{C}=0.5772\ldots$
8	$\dfrac{1}{x}\ln\bigl(1+a^2x^2\bigr), \quad a>0$	$-\pi\,\mathrm{Ei}\Bigl(-\dfrac{u}{a}\Bigr)$

S2.4.6. Expressions Involving Trigonometric Functions

No.	Original function, $f(x)$	Sine transform, $\check{f}_{\mathrm{s}}(u)=\int_0^\infty f(x)\sin(ux)\,dx$
1	$\dfrac{\sin(ax)}{x},\quad a>0$	$\dfrac{1}{2}\ln\left\lvert\dfrac{u+a}{u-a}\right\rvert$
2	$\dfrac{\sin(ax)}{x^2},\quad a>0$	$\begin{cases}\frac{1}{2}\pi u & \text{if } 0<u<a,\\ \frac{1}{2}\pi a & \text{if } u>a\end{cases}$
3	$x^{\nu-1}\sin(ax),\quad a>0,\ -2<\nu<1$	$\pi\,\dfrac{\lvert u-a\rvert^{-\nu}-\lvert u+a\rvert^{-\nu}}{4\Gamma(1-\nu)\sin\left(\frac{1}{2}\pi\nu\right)},\quad \nu\neq 0$
4	$\dfrac{\sin(ax)}{x^2+b^2},\quad a,b>0$	$\begin{cases}\frac{1}{2}\pi b^{-1}e^{-ab}\sinh(bu) & \text{if } 0<u<a,\\ \frac{1}{2}\pi b^{-1}e^{-bu}\sinh(ab) & \text{if } u>a\end{cases}$
5	$\dfrac{\sin(\pi x)}{1-x^2}$	$\begin{cases}\sin u & \text{if } 0<u<\pi,\\ 0 & \text{if } u>\pi\end{cases}$
6	$e^{-ax}\sin(bx),\quad a>0$	$\dfrac{a}{2}\left[\dfrac{1}{a^2+(b-u)^2}-\dfrac{1}{a^2+(b+u)^2}\right]$
7	$x^{-1}e^{-ax}\sin(bx),\quad a>0$	$\dfrac{1}{4}\ln\dfrac{(u+b)^2+a^2}{(u-b)^2+a^2}$
8	$\dfrac{1}{x}\sin^2(ax),\quad a>0$	$\begin{cases}\frac{1}{4}\pi & \text{if } 0<u<2a,\\ \frac{1}{8}\pi & \text{if } u=2a,\\ 0 & \text{if } u>2a\end{cases}$
9	$\dfrac{1}{x^2}\sin^2(ax),\quad a>0$	$\frac{1}{4}(u+2a)\ln\lvert u+2a\rvert+\frac{1}{4}(u-2a)\ln\lvert u-2a\rvert-\frac{1}{2}u\ln u$
10	$\exp\left(-ax^2\right)\sin(bx),\quad a>0$	$\dfrac{1}{2}\sqrt{\dfrac{\pi}{a}}\exp\left(-\dfrac{u^2+b^2}{4a}\right)\sinh\left(\dfrac{bu}{2a}\right)$
11	$\dfrac{1}{x}\sin(ax)\sin(bx),\quad a\geq b>0$	$\begin{cases}0 & \text{if } 0<u<a-b,\\ \frac{\pi}{4} & \text{if } a-b<u<a+b,\\ 0 & \text{if } a+b<u\end{cases}$
12	$\sin\left(\dfrac{a}{x}\right),\quad a>0$	$\dfrac{\pi\sqrt{a}}{2\sqrt{u}}J_1\left(2\sqrt{au}\right)$
13	$\dfrac{1}{\sqrt{x}}\sin\left(\dfrac{a}{x}\right),\quad a>0$	$\sqrt{\dfrac{\pi}{8u}}\left[\sin\left(2\sqrt{au}\right)-\cos\left(2\sqrt{au}\right)+\exp\left(-2\sqrt{au}\right)\right]$
14	$\exp\left(-a\sqrt{x}\right)\sin\left(a\sqrt{x}\right),\quad a>0$	$a\sqrt{\dfrac{\pi}{8}}\,u^{-3/2}\exp\left(-\dfrac{a^2}{2u}\right)$
15	$\dfrac{\cos(ax)}{x},\quad a>0$	$\begin{cases}0 & \text{if } 0<u<a,\\ \frac{1}{4}\pi & \text{if } u=a,\\ \frac{1}{2}\pi & \text{if } a<u\end{cases}$
16	$x^{\nu-1}\cos(ax),\quad a>0,\ \lvert\nu\rvert<1$	$\dfrac{\pi(u+a)^{-\nu}-\operatorname{sign}(u-a)\lvert u-a\rvert^{-\nu}}{4\Gamma(1-\nu)\cos\left(\frac{1}{2}\pi\nu\right)}$
17	$\dfrac{x\cos(ax)}{x^2+b^2},\quad a,b>0$	$\begin{cases}-\frac{1}{2}\pi e^{-ab}\sinh(bu) & \text{if } u<a,\\ \frac{1}{2}\pi e^{-bu}\cosh(ab) & \text{if } u>a\end{cases}$
18	$\dfrac{1-\cos(ax)}{x^2},\quad a>0$	$\dfrac{u}{2}\ln\left\lvert\dfrac{u^2-a^2}{u^2}\right\rvert+\dfrac{a}{2}\ln\left\lvert\dfrac{u+a}{u-a}\right\rvert$
19	$\dfrac{1}{\sqrt{x}}\cos\left(a\sqrt{x}\right)$	$\sqrt{\dfrac{\pi}{u}}\cos\left(\dfrac{a^2}{4u}+\dfrac{\pi}{4}\right)$
20	$\dfrac{1}{\sqrt{x}}\cos\left(a\sqrt{x}\right)\cos\left(b\sqrt{x}\right),\ a,b>0$	$\sqrt{\dfrac{\pi}{u}}\cos\left(\dfrac{ab}{2u}\right)\cos\left(\dfrac{a^2+b^2}{4u}+\dfrac{\pi}{4}\right)$

S2.4.7. Expressions Involving Special Functions

No.	Original function, $f(x)$	Sine transform, $\check{f}_{\rm s}(u) = \int_0^\infty f(x)\sin(ux)\,dx$
1	$\operatorname{erfc}(ax), \quad a > 0$	$\dfrac{1}{u}\Bigl[1 - \exp\Bigl(-\dfrac{u^2}{4a^2}\Bigr)\Bigr]$
2	$\operatorname{ci}(ax), \quad a > 0$	$-\dfrac{1}{2u}\ln\Bigl\lvert 1 - \dfrac{u^2}{a^2}\Bigr\rvert$
3	$\operatorname{si}(ax), \quad a > 0$	$\begin{cases} 0 & \text{if } 0 < u < a, \\ -\frac{1}{2}\pi u^{-1} & \text{if } a < u \end{cases}$
4	$J_0(ax), \quad a > 0$	$\begin{cases} 0 & \text{if } 0 < u < a, \\ \dfrac{1}{\sqrt{u^2 - a^2}} & \text{if } a < u \end{cases}$
5	$J_\nu(ax), \quad a > 0,\ \nu > -2$	$\begin{cases} \dfrac{\sin\bigl[\nu \arcsin(u/a)\bigr]}{\sqrt{a^2 - u^2}} & \text{if } 0 < u < a, \\ \dfrac{a^\nu \cos(\pi\nu/2)}{\xi(u+\xi)^\nu} & \text{if } a < u, \end{cases}$ where $\xi = \sqrt{u^2 - a^2}$
6	$\dfrac{1}{x} J_0(ax), \quad a > 0,\ \nu > 0$	$\begin{cases} \arcsin(u/a) & \text{if } 0 < u < a, \\ \pi/2 & \text{if } a < u \end{cases}$
7	$\dfrac{1}{x} J_\nu(ax), \quad a > 0,\ \nu > -1$	$\begin{cases} \nu^{-1} \sin\bigl[\nu \arcsin(u/a)\bigr] & \text{if } 0 < u < a, \\ \dfrac{a^\nu \sin(\pi\nu/2)}{\nu\bigl(u + \sqrt{u^2 - a^2}\bigr)^\nu} & \text{if } a < u \end{cases}$
8	$x^{-1} e^{-ax} J_0(bx), \quad a > 0$	$\arcsin\left(\dfrac{2u}{\sqrt{(u+b)^2 + a^2} + \sqrt{(u-b)^2 + a^2}}\right)$
9	$J_0\bigl(a\sqrt{x}\,\bigr), \quad a > 0$	$\dfrac{1}{u}\cos\Bigl(\dfrac{a^2}{4u}\Bigr)$
10	$\dfrac{1}{\sqrt{x}} J_1\bigl(a\sqrt{x}\,\bigr), \quad a > 0$	$\dfrac{2}{a}\sin\Bigl(\dfrac{a^2}{4u}\Bigr)$
11	$x^{\nu/2} J_\nu\bigl(a\sqrt{x}\,\bigr)$, $a > 0, \quad -2 < \nu < \frac{1}{2}$	$\dfrac{a^\nu}{2^\nu u^{\nu+1}} \cos\Bigl(\dfrac{a^2}{4u} - \dfrac{\pi\nu}{2}\Bigr)$
12	$Y_0(ax), \quad a > 0$	$\begin{cases} \dfrac{2\arcsin(u/a)}{\pi\sqrt{a^2 - u^2}} & \text{if } 0 < u < a, \\ \dfrac{2\bigl[\ln\bigl(u - \sqrt{u^2 - a^2}\,\bigr) - \ln a\bigr]}{\pi\sqrt{u^2 - a^2}} & \text{if } a < u \end{cases}$
13	$K_0(ax), \quad a > 0$	$\dfrac{\ln\bigl(u + \sqrt{u^2 + a^2}\,\bigr) - \ln a}{\sqrt{u^2 + a^2}}$
14	$x^{\nu+1} K_\nu(ax), \quad a > 0,\ \nu > -\frac{3}{2}$	$\sqrt{\pi}\,(2a)^\nu \Gamma\bigl(\nu + \frac{3}{2}\bigr) u (u^2 + a^2)^{-\nu - 3/2}$

Bibliography for Chapter S2

Bateman, H. and Erdélyi, A., *Tables of Integral Transforms. Vols. 1 and 2*, McGraw-Hill, New York, 1954.

Ditkin, V. A. and Prudnikov, A. P., *Integral Transforms and Operational Calculus*, Pergamon Press, New York, 1965.

Oberhettinger, F., *Tables of Fourier Transforms and Fourier Transforms of Distributions*, Springer-Verlag, Berlin, 1980.

Oberhettinger, F. and Badii, L., *Tables of Laplace Transforms*, Springer-Verlag, New York, 1973.

Prudnikov, A. P., Brychkov, Yu. A., and Marichev, O. I., *Integrals and Series, Vol. 4, Direct Laplace Transforms*, Gordon & Breach, New York, 1992.

Prudnikov, A. P., Brychkov, Yu. A., and Marichev, O. I., *Integrals and Series, Vol. 5, Inverse Laplace Transforms*, Gordon & Breach, New York, 1992.

Chapter S3

Orthogonal Curvilinear Systems of Coordinates

S3.1. Arbitrary Curvilinear Coordinate Systems

S3.1.1. General Nonorthogonal Curvilinear Coordinates

► **Metric tensor. Arc length and volume elements in curvilinear coordinates.** The curvilinear coordinates x^1, x^2, x^3 are defined as functions of the rectangular Cartesian coordinates x, y, z:

$$x^1 = x^1(x, y, z), \quad x^2 = x^2(x, y, z), \quad x^3 = x^3(x, y, z).$$

Using these formulas, one can express x, y, z in terms of the curvilinear coordinates x^1, x^2, x^3 as follows:

$$x = x(x^1, x^2, x^3), \quad y = y(x^1, x^2, x^3), \quad z = z(x^1, x^2, x^3).$$

The *metric tensor components* g_{ij} are determined by the formulas

$$\begin{aligned} g_{ij}(x^1, x^2, x^3) &= \frac{\partial x}{\partial x^i}\frac{\partial x}{\partial x^j} + \frac{\partial y}{\partial x^i}\frac{\partial y}{\partial x^j} + \frac{\partial z}{\partial x^i}\frac{\partial z}{\partial x^j}; \\ g_{ij}(x^1, x^2, x^3) &= g_{ji}(x^1, x^2, x^3); \qquad i, j = 1,\ 2,\ 3. \end{aligned}$$

The arc length dl between close points $(x, y, z) \equiv (x^1, x^2, x^3)$ and $(x + dx,\ y + dy,\ z + dz) \equiv (x^1 + dx^1,\ x^2 + dx^2,\ x^3 + dx^3)$ is expressed as

$$(dl)^2 = (dx)^2 + (dy)^2 + (dz)^2 = \sum_{i=1}^{3}\sum_{j=1}^{3} g_{ij}(x^1, x^2, x^3)\, dx^i\, dx^j.$$

The volume of the elementary parallelepiped with vertices at the eight points (x^1, x^2, x^3), $(x^1 + dx^1,\ x^2,\ x^3)$, $(x^1,\ x^2 + dx^2,\ x^3)$, $(x^1,\ x^2,\ x^3 + dx^3)$, $(x^1 + dx^1,\ x^2 + dx^2,\ x^3)$, $(x^1 + dx^1,\ x^2,\ x^3 + dx^3)$, $(x^1,\ x^2 + dx^2,\ x^3 + dx^3)$, $(x^1 + dx^1,\ x^2 + dx^2,\ x^3 + dx^3)$ is given by

$$dV = \frac{\partial(x, y, z)}{\partial(x^1, x^2, x^3)}\, dx^1\, dx^2\, dx^3 = \pm\sqrt{\det|g_{ij}|}\, dx^1\, dx^2\, dx^3.$$

Here the plus sign corresponds to the standard situation where the tangent vectors to the coordinate lines x^1, x^2, x^3, pointing in the direction of growth of the respective coordinate, form a right-handed triple, just as unit vectors **i**, **j**, **k** of a right-handed rectangular Cartesian coordinate system.

► **Vector components in Cartesian and curvilinear coordinate systems.** The unit vectors **i**, **j**, **k** of a rectangular Cartesian coordinate system* x, y, z and the unit vectors $\mathbf{i}_1$, $\mathbf{i}_2$, $\mathbf{i}_3$ of a curvilinear coordinate system x^1, x^2, x^3 are connected by the linear relations

$$\begin{aligned}
\mathbf{i}_n &= \frac{1}{\sqrt{g_{nn}}}\left(\frac{\partial x}{\partial x^n}\mathbf{i} + \frac{\partial y}{\partial x^n}\mathbf{j} + \frac{\partial z}{\partial x^n}\mathbf{k}\right), \qquad n = 1, 2, 3;\\
\mathbf{i} &= \sqrt{g_{11}}\,\frac{\partial x^1}{\partial x}\mathbf{i}_1 + \sqrt{g_{22}}\,\frac{\partial x^2}{\partial x}\mathbf{i}_2 + \sqrt{g_{33}}\,\frac{\partial x^3}{\partial x}\mathbf{i}_3;\\
\mathbf{j} &= \sqrt{g_{11}}\,\frac{\partial x^1}{\partial y}\mathbf{i}_1 + \sqrt{g_{22}}\,\frac{\partial x^2}{\partial y}\mathbf{i}_2 + \sqrt{g_{33}}\,\frac{\partial x^3}{\partial y}\mathbf{i}_3;\\
\mathbf{k} &= \sqrt{g_{11}}\,\frac{\partial x^1}{\partial z}\mathbf{i}_1 + \sqrt{g_{22}}\,\frac{\partial x^2}{\partial z}\mathbf{i}_2 + \sqrt{g_{33}}\,\frac{\partial x^3}{\partial z}\mathbf{i}_3.
\end{aligned}$$

In the general case, the vectors $\mathbf{i}_1$, $\mathbf{i}_2$, $\mathbf{i}_3$ are not orthogonal and change their direction from point to point.

The components v_x, v_y, v_z of a vector **v** in a rectangular Cartesian coordinate system x, y, z and the components v_1, v_2, v_3 of the same vector in a curvilinear coordinate system x^1, x^2, x^3 are related by

$$\begin{aligned}
\mathbf{v} &= v_x\mathbf{i} + v_y\mathbf{j} + v_z\mathbf{k} = v_1\mathbf{i}_1 + v_2\mathbf{i}_2 + v_3\mathbf{i}_3,\\
v_n &= \sqrt{g_{nn}}\left(\frac{\partial x^n}{\partial x}v_x + \frac{\partial x^n}{\partial y}v_y + \frac{\partial x^n}{\partial z}v_z\right), \qquad n = 1, 2, 3;\\
v_x &= \frac{\partial x}{\partial x^1}\frac{v_1}{\sqrt{g_{11}}} + \frac{\partial x}{\partial x^2}\frac{v_2}{\sqrt{g_{22}}} + \frac{\partial x}{\partial x^3}\frac{v_3}{\sqrt{g_{33}}};\\
v_y &= \frac{\partial y}{\partial x^1}\frac{v_1}{\sqrt{g_{11}}} + \frac{\partial y}{\partial x^2}\frac{v_2}{\sqrt{g_{22}}} + \frac{\partial y}{\partial x^3}\frac{v_3}{\sqrt{g_{33}}};\\
v_z &= \frac{\partial z}{\partial x^1}\frac{v_1}{\sqrt{g_{11}}} + \frac{\partial z}{\partial x^2}\frac{v_2}{\sqrt{g_{22}}} + \frac{\partial z}{\partial x^3}\frac{v_3}{\sqrt{g_{33}}}.
\end{aligned}$$

S3.1.2. General Orthogonal Curvilinear Coordinates

► **Orthogonal coordinates. Length, area, and volume elements.** A system of coordinates is orthogonal if

$$g_{ij}(x^1, x^2, x^3) = 0 \quad \text{for} \quad i \neq j.$$

In this case the third invariant of the metric tensor is given by

$$g = \det|g_{ij}| = g_{11}g_{22}g_{33}.$$

The *Lamé coefficients* L_k of orthogonal curvilinear coordinates are expressed in terms of the components of the metric tensor as

$$L_i = \sqrt{g_{ii}} = \sqrt{\left(\frac{\partial x}{\partial x^i}\right)^2 + \left(\frac{\partial y}{\partial x^i}\right)^2 + \left(\frac{\partial z}{\partial x^i}\right)^2}, \qquad i = 1, 2, 3.$$

---

* Here and henceforth the coordinate axes and the respective coordinates of points in space are denoted by the same letters.

Arc length element:

$$dl = \sqrt{(L_1\,dx^1)^2 + (L_2\,dx^2)^2 + (L_3\,dx^3)^2} = \sqrt{g_{11}(dx^1)^2 + g_{22}(dx^2)^2 + g_{33}(dx^3)^2}.$$

The area elements ds_i of the respective coordinate surfaces x^i = const are given by

$$\begin{aligned} ds_1 &= dl_2\,dl_3 = L_2L_3\,dx^2\,dx^3 = \sqrt{g_{22}g_{33}}\,dx^2\,dx^3,\\ ds_2 &= dl_1\,dl_3 = L_1L_3\,dx^1\,dx^3 = \sqrt{g_{11}g_{33}}\,dx^1\,dx^3,\\ ds_3 &= dl_1\,dl_2 = L_1L_2\,dx^1\,dx^2 = \sqrt{g_{11}g_{22}}\,dx^1\,dx^2. \end{aligned}$$

Volume element:

$$dV = L_1L_2L_3\,dx^1\,dx^2\,dx^3 = \sqrt{g_{11}g_{22}g_{33}}\;dx^1\,dx^2\,dx^3.$$

► **Basic differential relations in orthogonal curvilinear coordinates.** In what follows, we present the basic differential operators in the orthogonal curvilinear coordinates x^1, x^2, x^3. The corresponding unit vectors are denoted by $\mathbf{i}_1$, $\mathbf{i}_2$, $\mathbf{i}_3$.

The gradient of a scalar f is expressed as

$$\operatorname{grad} f \equiv \nabla f = \frac{1}{\sqrt{g_{11}}}\frac{\partial f}{\partial x^1}\mathbf{i}_1 + \frac{1}{\sqrt{g_{22}}}\frac{\partial f}{\partial x^2}\mathbf{i}_2 + \frac{1}{\sqrt{g_{33}}}\frac{\partial f}{\partial x^3}\mathbf{i}_3.$$

Divergence of a vector $\mathbf{v} = v_1\mathbf{i}_1 + v_2\mathbf{i}_2 + v_3\mathbf{i}_3$:

$$\operatorname{div}\mathbf{v} \equiv \nabla\cdot\mathbf{v} = \frac{1}{\sqrt{g}}\left[\frac{\partial}{\partial x^1}\left(v_1\sqrt{\frac{g}{g_{11}}}\right) + \frac{\partial}{\partial x^2}\left(v_2\sqrt{\frac{g}{g_{22}}}\right) + \frac{\partial}{\partial x^3}\left(v_3\sqrt{\frac{g}{g_{33}}}\right)\right].$$

Gradient of a scalar f along a vector $\mathbf{v}$:

$$(\mathbf{v}\cdot\nabla)f = \frac{v_1}{\sqrt{g_{11}}}\frac{\partial f}{\partial x^1} + \frac{v_2}{\sqrt{g_{22}}}\frac{\partial f}{\partial x^2} + \frac{v_3}{\sqrt{g_{33}}}\frac{\partial f}{\partial x^3}.$$

Gradient of a vector $\mathbf{w}$ along a vector $\mathbf{v}$:

$$(\mathbf{v}\cdot\nabla)\mathbf{w} = \mathbf{i}_1(\mathbf{v}\cdot\nabla)w_1 + \mathbf{i}_2(\mathbf{v}\cdot\nabla)w_2 + \mathbf{i}_3(\mathbf{v}\cdot\nabla)w_3.$$

Curl of a vector $\mathbf{v}$:

$$\begin{aligned} \operatorname{curl}\mathbf{v} \equiv \nabla\times\mathbf{v} = \mathbf{i}_1&\frac{\sqrt{g_{11}}}{\sqrt{g}}\left[\frac{\partial}{\partial x^2}\left(v_3\sqrt{g_{33}}\right) - \frac{\partial}{\partial x^3}\left(v_2\sqrt{g_{22}}\right)\right]\\ +\,\mathbf{i}_2&\frac{\sqrt{g_{22}}}{\sqrt{g}}\left[\frac{\partial}{\partial x^3}\left(v_1\sqrt{g_{11}}\right) - \frac{\partial}{\partial x^1}\left(v_3\sqrt{g_{33}}\right)\right]\\ +\,\mathbf{i}_3&\frac{\sqrt{g_{33}}}{\sqrt{g}}\left[\frac{\partial}{\partial x^1}\left(v_2\sqrt{g_{22}}\right) - \frac{\partial}{\partial x^2}\left(v_1\sqrt{g_{11}}\right)\right]. \end{aligned}$$

Remark. Sometimes curl **v** is denoted by curl **v**.

Laplace operator of a scalar f:

$$\Delta f \equiv \nabla^2 f = \frac{1}{\sqrt{g}}\left[\frac{\partial}{\partial x^1}\left(\frac{\sqrt{g}}{g_{11}}\frac{\partial f}{\partial x^1}\right) + \frac{\partial}{\partial x^2}\left(\frac{\sqrt{g}}{g_{22}}\frac{\partial f}{\partial x^2}\right) + \frac{\partial}{\partial x^3}\left(\frac{\sqrt{g}}{g_{33}}\frac{\partial f}{\partial x^3}\right)\right].$$

S3.2. Cylindrical and Spherical Coordinate Systems

S3.2.1. Cylindrical Coordinates

► **Transformations of coordinates and vectors. The metric tensor components.** The Cartesian coordinates are expressed in terms of the cylindrical ones as

$$x = \rho\cos\varphi, \quad y = \rho\sin\varphi, \quad z = z$$
$$(0 \le \rho < \infty,\ 0 \le \varphi < 2\pi,\ -\infty < z < \infty).$$

The cylindrical coordinates are expressed in terms of the Cartesian ones as

$$\rho = \sqrt{x^2 + y^2}, \quad \tan\varphi = y/x, \quad z = z \qquad (\sin\varphi = y/\rho).$$

Coordinate surfaces:

$x^2 + y^2 = \rho^2$ (right circular cylinders with their axis coincident with the z-axis),
$y = x\tan\varphi$ (half-planes through the z-axis),
$z = z$ (planes perpendicular to the z-axis).

Direct and inverse transformations of the components of a vector $\mathbf{v} = v_x\mathbf{i} + v_y\mathbf{j} + v_z\mathbf{k} = v_\rho\mathbf{i}_\rho + v_\varphi\mathbf{i}_\varphi + v_z\mathbf{i}_z$:

$$\begin{aligned} v_\rho &= v_x\cos\varphi + v_y\sin\varphi, & v_x &= v_\rho\cos\varphi - v_\varphi\sin\varphi,\\ v_\varphi &= -v_x\sin\varphi + v_y\cos\varphi, & v_y &= v_\rho\sin\varphi + v_\varphi\cos\varphi,\\ v_z &= v_z; & v_z &= v_z. \end{aligned}$$

Metric tensor components:

$$g_{\rho\rho} = 1, \quad g_{\varphi\varphi} = \rho^2, \quad g_{zz} = 1, \quad \sqrt{g} = \rho.$$

► **Basic differential relations.** Gradient of a scalar f:

$$\nabla f = \frac{\partial f}{\partial\rho}\mathbf{i}_\rho + \frac{1}{\rho}\frac{\partial f}{\partial\varphi}\mathbf{i}_\varphi + \frac{\partial f}{\partial z}\mathbf{i}_z.$$

Divergence of a vector $\mathbf{v}$:

$$\nabla\cdot\mathbf{v} = \frac{1}{\rho}\frac{\partial(\rho v_\rho)}{\partial\rho} + \frac{1}{\rho}\frac{\partial v_\varphi}{\partial\varphi} + \frac{\partial v_z}{\partial z}.$$

Gradient of a scalar f along a vector $\mathbf{v}$:

$$(\mathbf{v}\cdot\nabla)f = v_\rho\frac{\partial f}{\partial\rho} + \frac{v_\varphi}{\rho}\frac{\partial f}{\partial\varphi} + v_z\frac{\partial f}{\partial z}.$$

Gradient of a vector $\mathbf{w}$ along a vector $\mathbf{v}$:

$$(\mathbf{v}\cdot\nabla)\mathbf{w} = (\mathbf{v}\cdot\nabla)w_\rho\mathbf{i}_\rho + (\mathbf{v}\cdot\nabla)w_\varphi\mathbf{i}_\varphi + (\mathbf{v}\cdot\nabla)w_z\mathbf{i}_z.$$

Curl of a vector $\mathbf{v}$:

$$\nabla\times\mathbf{v} = \left(\frac{1}{\rho}\frac{\partial v_z}{\partial\varphi} - \frac{\partial v_\varphi}{\partial z}\right)\mathbf{i}_\rho + \left(\frac{\partial v_\rho}{\partial z} - \frac{\partial v_z}{\partial\rho}\right)\mathbf{i}_\varphi + \frac{1}{\rho}\left[\frac{\partial(\rho v_\varphi)}{\partial\rho} - \frac{\partial v_\rho}{\partial\varphi}\right]\mathbf{i}_z.$$

Laplacian of a scalar f:

$$\Delta f = \frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial f}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 f}{\partial\varphi^2} + \frac{\partial^2 f}{\partial z^2}.$$

Remark. The cylindrical coordinates ρ, φ are also used as *polar coordinates* on the plane xy.

S3.2.2. Spherical Coordinates

► **Transformations of coordinates and vectors. The metric tensor components.** The Cartesian coordinates are expressed in terms of the spherical ones as

$$x = r\sin\theta\cos\varphi, \quad y = r\sin\theta\sin\varphi, \quad z = r\cos\theta$$
$$(0 \le r < \infty,\ \ 0 \le \theta \le \pi,\ \ 0 \le \varphi < 2\pi).$$

The spherical coordinates are expressed in terms of the Cartesian ones as

$$r = \sqrt{x^2 + y^2 + z^2}, \quad \theta = \arccos\frac{z}{r}, \quad \tan\varphi = \frac{y}{x} \qquad \biggl(\sin\varphi = \frac{y}{\sqrt{x^2 + y^2}}\biggr).$$

Coordinate surfaces:

$$\begin{aligned} &x^2 + y^2 + z^2 = r^2 && \text{(spheres)},\\ &x^2 + y^2 - z^2\tan^2\theta = 0 && \text{(circular cones)},\\ &y = x\tan\varphi && \text{(half-planes through the } z\text{-axis)}. \end{aligned}$$

Direct and inverse transformations of the components of a vector $\mathbf{v} = v_x\mathbf{i} + v_y\mathbf{j} + v_z\mathbf{k} = v_r\mathbf{i}_r + v_\theta\mathbf{i}_\theta + v_\varphi\mathbf{i}_\varphi$:

$$\begin{aligned} v_r &= v_x\sin\theta\cos\varphi + v_y\sin\theta\sin\varphi + v_z\cos\theta, & v_x &= v_r\sin\theta\cos\varphi + v_\theta\cos\theta\cos\varphi - v_\varphi\sin\varphi,\\ v_\theta &= v_x\cos\theta\cos\varphi + v_y\cos\theta\sin\varphi - v_z\sin\theta, & v_y &= v_r\sin\theta\sin\varphi + v_\theta\cos\theta\sin\varphi + v_\varphi\cos\varphi,\\ v_\varphi &= -v_x\sin\varphi + v_y\cos\varphi; & v_z &= v_r\cos\theta - v_\theta\sin\theta. \end{aligned}$$

The metric tensor components are

$$g_{rr} = 1, \quad g_{\theta\theta} = r^2, \quad g_{\varphi\varphi} = r^2\sin^2\theta, \quad \sqrt{g} = r^2\sin\theta.$$

► **Basic differential relations.** Gradient of a scalar f:

$$\nabla f = \frac{\partial f}{\partial r}\mathbf{i}_r + \frac{1}{r}\frac{\partial f}{\partial\theta}\mathbf{i}_\theta + \frac{1}{r\sin\theta}\frac{\partial f}{\partial\varphi}\mathbf{i}_\varphi.$$

Divergence of a vector $\mathbf{v}$:

$$\nabla\cdot\mathbf{v} = \frac{1}{r^2}\frac{\partial}{\partial r}\bigl(r^2 v_r\bigr) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}\bigl(\sin\theta\, v_\theta\bigr) + \frac{1}{r\sin\varphi}\frac{\partial v_\varphi}{\partial\varphi}.$$

Gradient of a scalar f along a vector $\mathbf{v}$:

$$(\mathbf{v}\cdot\nabla)f = v_r\frac{\partial f}{\partial r} + \frac{v_\theta}{r}\frac{\partial f}{\partial\theta} + \frac{v_\varphi}{r\sin\theta}\frac{\partial f}{\partial\varphi}.$$

Gradient of a vector $\mathbf{w}$ along a vector $\mathbf{v}$:

$$(\mathbf{v}\cdot\nabla)\mathbf{w} = (\mathbf{v}\cdot\nabla)w_r\mathbf{i}_r + (\mathbf{v}\cdot\nabla)w_\theta\mathbf{i}_\theta + (\mathbf{v}\cdot\nabla)w_\varphi\mathbf{i}_\varphi.$$

Curl of a vector $\mathbf{v}$:

$$\nabla\times\mathbf{v} = \frac{1}{r\sin\theta}\left[\frac{\partial(\sin\theta\, v_\varphi)}{\partial\theta} - \frac{\partial v_\theta}{\partial\varphi}\right]\mathbf{i}_r + \frac{1}{r}\left[\frac{1}{\sin\theta}\frac{\partial v_r}{\partial\varphi} - \frac{\partial(rv_\varphi)}{\partial r}\right]\mathbf{i}_\theta + \frac{1}{r}\left[\frac{\partial(rv_\theta)}{\partial r} - \frac{\partial v_r}{\partial\theta}\right]\mathbf{i}_\varphi.$$

Laplacian of a scalar f:

$$\Delta f = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial f}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial f}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 f}{\partial\varphi^2}.$$

Bibliography for Chapter S3

Arfken, G. B. and Weber, H. J., *Mathematical Methods for Physicists, 4th Edition*, Academic Press, New York, 1995.

Korn, G. A. and Korn, T. M., *Mathematical Handbook for Scientists and Engineers, 2nd Edition*, Dover Publications, New York, 2000.

Menzel, D. H., *Mathematical Physics*, Dover Publications, New York, 1961.

Moon, P. and Spencer, D. E., *Field Theory Handbook, Including Coordinate Systems, Differential Equations and Their Solutions, 3rd Edition*, Springer-Verlag, Berlin, 1988.

Morse, P. M. and Feshbach, H., *Methods of Theoretical Physics, Part I*, McGraw-Hill, New York, 1953.

Weisstein, E. W., *CRC Concise Encyclopedia of Mathematics, 2nd Edition*, CRC Press, Boca Raton, Florida, 2003.

Zwillinger, D., *CRC Standard Mathematical Tables and Formulae, 31st Edition*, CRC Press, Boca Raton, Florida, 2002.

Chapter S4

Ordinary Differential Equations

► *In this chapter we shall often use the term "solution" to mean "general solution"; C, C_1, and C_2 are arbitrary constants. Notation like S4.2.1.12 stands for equation 12 from Subsection S4.2.1.*

S4.1. First-Order Equations

S4.1.1. Simplest Equations Integrable in Closed Form

1. $\boldsymbol{y'_x = f(y).}$

Autonomous equation.

Solution: $x = \displaystyle\int \frac{dy}{f(y)} + C.$

Particular solutions: $y = A_k$, where the A_k are roots of the algebraic (transcendental) equation $f(A_k) = 0$.

2. $\boldsymbol{y'_x = f(x)g(y).}$

Separable equation.

Solution: $\displaystyle\int \frac{dy}{g(y)} = \int f(x)\,dx + C.$

Particular solutions: $y = A_k$, where the A_k are roots of the algebraic (transcendental) equation $g(A_k) = 0$.

3. $\boldsymbol{g(x)y'_x = f_1(x)y + f_0(x).}$

Linear equation.

Solution:

$$y = Ce^F + e^F \int e^{-F} \frac{f_0(x)}{g(x)}\,dx, \quad \text{where} \quad F(x) = \int \frac{f_1(x)}{g(x)}\,dx.$$

4. $\boldsymbol{g(x)y'_x = f_1(x)y + f_0(x)y^k.}$

Bernoulli equation. Here, k is an arbitrary number. For $k \neq 1$, the substitution $w(x) = y^{1-k}$ leads to a linear equation: $g(x)w'_x = (1-k)f_1(x)w + (1-k)f_0(x)$.

Solution:

$$y^{1-k} = Ce^F + (1-k)e^F \int e^{-F} \frac{f_0(x)}{g(x)}\,dx, \quad \text{where} \quad F(x) = (1-k)\int \frac{f_1(x)}{g(x)}\,dx.$$

5. $\boldsymbol{y'_x = f(y/x).}$

Homogeneous equation. The substitution $u(x) = y/x$ leads to a separable equation: $xu'_x = f(u) - u$.

6. $\boldsymbol{y'_x = f(ax+by+c)}$.

If $b \neq 0$, the substitution $u(x) = ax + by + c$ leads to a separable equation: $u'_x = bf(u) + a$.

7. $\boldsymbol{y'_x = x^{k-1} f(yx^{-k})}$.

Generalized homogeneous equation. The substitution $u = yx^{-k}$ leads to a separable equation, $xu'_x = f(u) - ku$.

8. $\boldsymbol{y'_x = \dfrac{y}{x} f(x^n y^m)}$.

Generalized homogeneous equation. The substitution $z = x^n y^m$ leads to a separable equation: $xz'_x = nz + mzf(z)$.

9. $\boldsymbol{y'_x = f\Big(\dfrac{ax+by+c}{\alpha x+\beta y+\gamma}\Big)}$.

1°. For $\Delta = a\beta - b\alpha \neq 0$, the transformation $x = u + \dfrac{b\gamma - c\beta}{\Delta}$, $y = v(u) + \dfrac{c\alpha - a\gamma}{\Delta}$ leads to an equation:

$$v'_u = f\Big(\frac{au + bv}{\alpha u + \beta v}\Big).$$

Dividing both the numerator and denominator of the fraction on the right-hand side by u, we obtain a homogeneous equation of the form S4.1.1.5.

2°. For $\Delta = 0$ and $b \neq 0$, the substitution $v(x) = ax + by + c$ leads to a separable equation of the form S4.1.1.2:

$$v'_x = a + bf\Big(\frac{bv}{\beta v + b\gamma - c\beta}\Big).$$

3°. For $\Delta = 0$ and $\beta \neq 0$, the substitution $v(x) = \alpha x + \beta y + \gamma$ also leads to a separable equation:

$$v'_x = \alpha + \beta f\Big(\frac{bv + c\beta - b\gamma}{\beta v}\Big).$$

S4.1.2. Riccati Equation

Preliminary remarks. A Riccati equation has the general form

$$y'_x = f_2(x)y^2 + f_1(x)y + f_0(x). \tag{S4.1.2.1}$$

1°. Given a particular solution $y_0 = y_0(x)$ of the Riccati equation (S4.1.2.1), the general solution can be written as:

$$\begin{aligned} y &= y_0(x) + \Phi(x)\Big[C - \int \Phi(x) f_2(x)\,dx\Big]^{-1}, \\ \Phi(x) &= \exp\Big\{\int \big[2f_2(x)y_0(x) + f_1(x)\big]\,dx\Big\}. \end{aligned} \tag{S4.1.2.2}$$

To the particular solution $y_0(x)$ there corresponds $C = \infty$.

Often only particular solutions will be given for the specific equations presented below in Subsection S4.1.2. The general solutions of these equations can be obtained by formulas (S4.1.2.2).

2°. The substitution

$$u(x) = \exp\Bigl(-\int f_2 y\,dx\Bigr)$$

reduces the general Riccati equation to a second-order linear equation:

$$f_2 u''_{xx} - \bigl[(f_2)'_x + f_1 f_2\bigr] u'_x + f_0 f_2^2 u = 0,$$

which often may be easier to solve than the original Riccati equation. Specific second-order linear equations are outlined in Section S4.2.

3°. Many solvable equations of this form can be found in the book by Polyanin and Zaitsev (2003).

1. $\boldsymbol{y'_x = ay^2 + bx^n.}$

Special Riccati equation, n is an arbitrary number.

1°. Solution for $n \neq -2$:

$$y = -\frac{1}{a}\frac{w'_x}{w}, \qquad w(x) = \sqrt{x}\,\Bigl[C_1 J_{\frac{1}{2k}}\Bigl(\frac{1}{k}\sqrt{ab}\,x^k\Bigr) + C_2 Y_{\frac{1}{2k}}\Bigl(\frac{1}{k}\sqrt{ab}\,x^k\Bigr)\Bigr],$$

where $k = \frac{1}{2}(n+2)$; $J_m(z)$ and $Y_m(z)$ are Bessel functions (see Subsection M13.6).

2°. Solution for $n = -2$:

$$y = \frac{\lambda}{x} - x^{2a\lambda}\Bigl(\frac{ax}{2a\lambda+1}x^{2a\lambda} + C\Bigr)^{-1},$$

where λ is a root of the quadratic equation $a\lambda^2 + \lambda + b = 0$.

2. $\boldsymbol{y'_x = y^2 + f(x)y - a^2 - af(x).}$

Particular solution: $y_0 = a$. The general solution can be obtained by formulas (S4.1.2.2).

3. $\boldsymbol{y'_x = f(x)y^2 + ay - ab - b^2 f(x).}$

Particular solution: $y_0 = b$. The general solution can be obtained by formulas (S4.1.2.2).

4. $\boldsymbol{y'_x = y^2 + xf(x)y + f(x).}$

Particular solution: $y_0 = -1/x$. The general solution can be obtained by formulas (S4.1.2.2).

5. $\boldsymbol{y'_x = f(x)y^2 - ax^n f(x)y + anx^{n-1}.}$

Particular solution: $y_0 = ax^n$. The general solution can be obtained by formulas (S4.1.2.2).

6. $\boldsymbol{y'_x = f(x)y^2 + anx^{n-1} - a^2x^{2n} f(x).}$

Particular solution: $y_0 = ax^n$. The general solution can be obtained by formulas (S4.1.2.2).

7. $\boldsymbol{y'_x = -(n+1)x^n y^2 + x^{n+1} f(x)y - f(x).}$

Particular solution: $y_0 = x^{-n-1}$. The general solution can be obtained by formulas (S4.1.2.2).

8. $\boldsymbol{xy'_x = f(x)y^2 + ny + ax^{2n}f(x).}$

Solution: $$y = \begin{cases} \sqrt{a}\,x^n \tan\Bigl[\sqrt{a}\displaystyle\int x^{n-1}f(x)\,dx + C\Bigr] & \text{if } a > 0, \\ \sqrt{|a|}\,x^n \tanh\Bigl[-\sqrt{|a|}\displaystyle\int x^{n-1}f(x)\,dx + C\Bigr] & \text{if } a < 0. \end{cases}$$

9. $\boldsymbol{xy'_x = x^{2n}f(x)y^2 + [ax^nf(x) - n]y + bf(x).}$

The substitution $z = x^n y$ leads to a separable equation: $z'_x = x^{n-1}f(x)(z^2 + az + b)$.

10. $\boldsymbol{y'_x = f(x)y^2 + g(x)y - a^2f(x) - ag(x).}$

Particular solution: $y_0 = a$. The general solution can be obtained by formulas (S4.1.2.2).

11. $\boldsymbol{y'_x = f(x)y^2 + g(x)y + anx^{n-1} - a^2x^{2n}f(x) - ax^ng(x).}$

Particular solution: $y_0 = ax^n$. The general solution can be obtained by formulas (S4.1.2.2).

12. $\boldsymbol{y'_x = ae^{\lambda x}y^2 + ae^{\lambda x}f(x)y + \lambda f(x).}$

Particular solution: $y_0 = -\dfrac{\lambda}{a}e^{-\lambda x}$. The general solution can be obtained by formulas (S4.1.2.2).

13. $\boldsymbol{y'_x = f(x)y^2 - ae^{\lambda x}f(x)y + a\lambda e^{\lambda x}.}$

Particular solution: $y_0 = ae^{\lambda x}$. The general solution can be obtained by formulas (S4.1.2.2).

14. $\boldsymbol{y'_x = f(x)y^2 + a\lambda e^{\lambda x} - a^2e^{2\lambda x}f(x).}$

Particular solution: $y_0 = ae^{\lambda x}$. The general solution can be obtained by formulas (S4.1.2.2).

15. $\boldsymbol{y'_x = f(x)y^2 + \lambda y + ae^{2\lambda x}f(x).}$

Solution: $$y = \begin{cases} \sqrt{a}\,e^{\lambda x} \tan\Bigl[\sqrt{a}\displaystyle\int e^{\lambda x}f(x)\,dx + C\Bigr] & \text{if } a > 0, \\ \sqrt{|a|}\,e^{\lambda x} \tanh\Bigl[-\sqrt{|a|}\displaystyle\int e^{\lambda x}f(x)\,dx + C\Bigr] & \text{if } a < 0. \end{cases}$$

16. $\boldsymbol{y'_x = y^2 - f^2(x) + f'_x(x).}$

Particular solution: $y_0 = f(x)$. The general solution can be obtained by formulas (S4.1.2.2).

17. $\boldsymbol{y'_x = f(x)y^2 - f(x)g(x)y + g'_x(x).}$

Particular solution: $y_0 = g(x)$.

S4.1.3. Other Equations

1. $\boldsymbol{yy'_x = f(x)y + g(x).}$

Abel equation of the second kind. Many solvable equations of this form can be found in the books by Zaitsev and Polyanin (1994) and Polyanin and Zaitsev (2003).

2. $\boldsymbol{yy'_x = f(x)y^2 + g(x)y + h(x).}$

Abel equation of the second kind. Many solvable equations of this form can be found in the books by Zaitsev and Polyanin (1994) and Polyanin and Zaitsev (2003).

► *In equations S4.1.3.3–S4.1.3.21 below, the functions f, g, and h are arbitrary composite functions whose arguments can depend on both x and y.*

3. $\boldsymbol{y'_x = f(y + ax^n + b) - anx^{n-1}.}$

The substitution $u = y + ax^n + b$ leads to a separable equation: $u'_x = f(u)$.

4. $\boldsymbol{y'_x = -\dfrac{n}{m}\dfrac{y}{x} + y^k f(x)g(x^n y^m).}$

The substitution $z = x^n y^m$ leads to a separable equation: $z'_x = mx^{\frac{n-nk}{m}} f(x) z^{\frac{k+m-1}{m}} g(z)$.

5. $\boldsymbol{y'_x = x^{n-1} y^{1-m} f(ax^n + by^m).}$

The substitution $w = ax^n + by^m$ leads to a separable equation: $w'_x = x^{n-1}[an + bmf(w)]$.

6. $\boldsymbol{[x^n f(y) + xg(y)]y'_x = h(y).}$

This is a Bernoulli equation with respect to $x = x(y)$ (see equation S4.1.1.4).

7. $\boldsymbol{x[f(x^n y^m) + mx^k g(x^n y^m)]y'_x = y[h(x^n y^m) - nx^k g(x^n y^m)].}$

The transformation $t = x^n y^m$, $z = x^{-k}$ leads to a linear equation with respect to $z = z(t)$: $t[nf(t) + mh(t)]z'_t = -kf(t)z - kmg(t)$.

8. $\boldsymbol{x[f(x^n y^m) + my^k g(x^n y^m)]y'_x = y[h(x^n y^m) - ny^k g(x^n y^m)].}$

The transformation $t = x^n y^m$, $z = y^{-k}$ leads to a linear equation with respect to $z = z(t)$: $t[nf(t) + mh(t)]z'_t = -kh(t)z + kng(t)$.

9. $\boldsymbol{x[sf(x^n y^m) - mg(x^k y^s)]y'_x = y[ng(x^k y^s) - kf(x^n y^m)].}$

The transformation $t = x^n y^m$, $w = x^k y^s$ leads to a separable equation: $tf(t)w'_t = wg(w)$.

10. $\boldsymbol{[f(y) + amx^n y^{m-1}]y'_x + g(x) + anx^{n-1}y^m = 0.}$

Solution: $\displaystyle\int f(y)\,dy + \int g(x)\,dx + ax^n y^m = C.$

11. $\boldsymbol{y'_x = e^{-\lambda x} f(e^{\lambda x} y).}$

The substitution $u = e^{\lambda x} y$ leads to a separable equation: $u'_x = f(u) + \lambda u$.

12. $\boldsymbol{y'_x = e^{\lambda y} f(e^{\lambda y} x).}$

The substitution $u = e^{\lambda y} x$ leads to a separable equation: $xu'_x = \lambda u^2 f(u) + u$.

13. $\boldsymbol{y'_x = yf(e^{\alpha x} y^m).}$

The substitution $z = e^{\alpha x} y^m$ leads to a separable equation: $z'_x = \alpha z + mzf(z)$.

14. $\boldsymbol{y'_x = \frac{1}{x} f(x^n e^{\alpha y})}$**.**

The substitution $z = x^n e^{\alpha y}$ leads to a separable equation: $xz'_x = nz + \alpha z f(z)$.

15. $\boldsymbol{y'_x = f(x)e^{\lambda y} + g(x)}$**.**

The substitution $u = e^{-\lambda y}$ leads to a linear equation: $u'_x = -\lambda g(x)u - \lambda f(x)$.

16. $\boldsymbol{y'_x = -\frac{n}{x} + f(x)g(x^n e^y)}$**.**

The substitution $z = x^n e^y$ leads to a separable equation: $z'_x = f(x)zg(z)$.

17. $\boldsymbol{y'_x = -\frac{\alpha}{m} y + y^k f(x)g(e^{\alpha x} y^m)}$**.**

The substitution $z = e^{\alpha x} y^m$ leads to a separable equation:

$$z'_x = m \exp\Bigl[\frac{\alpha}{m}(1-k)x\Bigr] f(x) z^{\frac{k+m-1}{m}} g(z).$$

18. $\boldsymbol{y'_x = e^{\alpha x - \beta y} f(a e^{\alpha x} + b e^{\beta y})}$**.**

The substitution $w = ae^{\alpha x} + be^{\beta y}$ leads to a separable equation: $w'_x = e^{\alpha x}[a\alpha + b\beta f(w)]$.

19. $\boldsymbol{[e^{\alpha x} f(y) + a\beta] y'_x + e^{\beta y} g(x) + a\alpha = 0}$**.**

Solution: $\displaystyle\int e^{-\beta y} f(y)\,dy + \int e^{-\alpha x} g(x)\,dx - a e^{-\alpha x - \beta y} = C$.

20. $\boldsymbol{x[f(x^n e^{\alpha y}) + \alpha y g(x^n e^{\alpha y})] y'_x = h(x^n e^{\alpha y}) - n y g(x^n e^{\alpha y})}$**.**

The substitution $t = x^n e^{\alpha y}$ leads to a linear equation with respect to $y = y(t)$: $t[nf(t) + \alpha h(t)] y'_t = -ng(t)y + h(t)$.

21. $\boldsymbol{[f(e^{\alpha x} y^m) + mxg(e^{\alpha x} y^m)] y'_x = y[h(e^{\alpha x} y^m) - \alpha x g(e^{\alpha x} y^m)]}$**.**

The substitution $t = e^{\alpha x} y^m$ leads to a linear equation with respect to $x = x(t)$: $t[\alpha f(t) + m h(t)] x'_t = m g(t) x + f(t)$.

S4.2. Second-Order Linear Equations

Preliminary remarks. A homogeneous linear equation of the second order has the general form

$$f_2(x) y''_{xx} + f_1(x) y'_x + f_0(x) y = 0.$$

Let $y_0 = y_0(x)$ be a nontrivial particular solution ($y_0 \not\equiv 0$) of this equation. Then the general solution of this equation can be found from the formula

$$y = y_0 \Bigl(C_1 + C_2 \int \frac{e^{-F}}{y_0^2}\,dx\Bigr), \qquad \text{where} \quad F = \int \frac{f_1}{f_2}\,dx.$$

For specific equations described below, sometimes only particular solutions are given, while the general solutions can be obtained with the above formula.

S4.2.1. Equations Involving Power Functions

1. $y''_{xx} + ay = 0.$

Equation of free oscillations.

$$\text{Solution:}\quad y = \begin{cases} C_1 \sinh(x\sqrt{|a|}\,) + C_2 \cosh(x\sqrt{|a|}\,) & \text{if } a < 0, \\ C_1 + C_2 x & \text{if } a = 0, \\ C_1 \sin(x\sqrt{a}\,) + C_2 \cos(x\sqrt{a}\,) & \text{if } a > 0. \end{cases}$$

2. $y''_{xx} - ax^n y = 0.$

1°. For $n = -2$, this is the Euler equation S4.2.1.12 (the solution is expressed in terms of elementary function).

2°. Assume $2/(n+2) = 2m+1$, where m is an integer. Then the solution is

$$y = \begin{cases} x(x^{1-2q}D)^{m+1}\left[C_1 \exp\left(\dfrac{\sqrt{a}}{q}x^q\right) + C_2 \exp\left(-\dfrac{\sqrt{a}}{q}x^q\right)\right] & \text{if } m \geq 0, \\ (x^{1-2q}D)^{-m}\left[C_1 \exp\left(\dfrac{\sqrt{a}}{q}x^q\right) + C_2 \exp\left(-\dfrac{\sqrt{a}}{q}x^q\right)\right] & \text{if } m < 0, \end{cases}$$

where $D = \dfrac{d}{dx}$, $q = \dfrac{n+2}{2} = \dfrac{1}{2m+1}$.

3°. For any n, the solution is expressed in terms of Bessel functions and modified Bessel functions:

$$y = \begin{cases} C_1\sqrt{x}\, J_{\frac{1}{2q}}\left(\dfrac{\sqrt{-a}}{q}x^q\right) + C_2\sqrt{x}\, Y_{\frac{1}{2q}}\left(\dfrac{\sqrt{-a}}{q}x^q\right) & \text{if } a < 0, \\ C_1\sqrt{x}\, I_{\frac{1}{2q}}\left(\dfrac{\sqrt{a}}{q}x^q\right) + C_2\sqrt{x}\, K_{\frac{1}{2q}}\left(\dfrac{\sqrt{a}}{q}x^q\right) & \text{if } a > 0, \end{cases}$$

where $q = \frac{1}{2}(n+2)$. The functions $J_\nu(z)$, $Y_\nu(z)$ and $I_\nu(z)$, $K_\nu(z)$ are described in Sections M13.6 and M13.7 in detail; see also equations S4.2.1.13 and S4.2.1.14.

3. $y''_{xx} + ay'_x + by = 0.$

Second-order constant coefficient linear equation. In physics this equation is called an *equation of damped vibrations.*

$$\text{Solution:}\quad y = \begin{cases} \exp\bigl(-\tfrac{1}{2}ax\bigr)\bigl[C_1 \exp\bigl(\tfrac{1}{2}\lambda x\bigr) + C_2 \exp\bigl(-\tfrac{1}{2}\lambda x\bigr)\bigr] & \text{if } \lambda^2 = a^2 - 4b > 0, \\ \exp\bigl(-\tfrac{1}{2}ax\bigr)\bigl[C_1 \sin\bigl(\tfrac{1}{2}\lambda x\bigr) + C_2 \cos\bigl(\tfrac{1}{2}\lambda x\bigr)\bigr] & \text{if } \lambda^2 = 4b - a^2 > 0, \\ \exp\bigl(-\tfrac{1}{2}ax\bigr)\bigl(C_1 x + C_2\bigr) & \text{if } a^2 = 4b. \end{cases}$$

4. $y''_{xx} + ay'_x + (bx + c)y = 0.$

1°. Solution with $b \neq 0$:

$$y = \exp\bigl(-\tfrac{1}{2}ax\bigr)\sqrt{\xi}\,\bigl[C_1 J_{1/3}\bigl(\tfrac{2}{3}\sqrt{b}\,\xi^{3/2}\bigr) + C_2 Y_{1/3}\bigl(\tfrac{2}{3}\sqrt{b}\,\xi^{3/2}\bigr)\bigr], \qquad \xi = x + \frac{4c - a^2}{4b},$$

where $J_{1/3}(z)$ and $Y_{1/3}(z)$ are Bessel functions.

2°. For $b = 0$, see equation S4.2.1.3.

5. $y''_{xx} + (ax+b)y'_x + (\alpha x^2 + \beta x + \gamma)y = 0.$

The substitution $y = u\exp(sx^2)$, where s is a root of the quadratic equation $4s^2+2as+\alpha=0$, leads to an equation of the form S4.2.1.11: $u''_{xx}+[(a+4s)x+b]u'_x+[(\beta+2bs)x+\gamma+2s]u=0$.

6. $xy''_{xx} + ay'_x + by = 0.$

1°. The solution is expressed in terms of Bessel functions and modified Bessel functions:

$$y = \begin{cases} x^{\frac{1-a}{2}}\left[C_1 J_\nu\left(2\sqrt{bx}\,\right) + C_2 Y_\nu\left(2\sqrt{bx}\,\right)\right] & \text{if } bx > 0, \\ x^{\frac{1-a}{2}}\left[C_1 I_\nu\left(2\sqrt{|bx|}\,\right) + C_2 K_\nu\left(2\sqrt{|bx|}\,\right)\right] & \text{if } bx < 0, \end{cases}$$

where $\nu = |1-a|$.

2°. For $a = \frac{1}{2}(2n+1)$, where $n = 0,\ 1,\ \dots$, the solution is

$$y = \begin{cases} C_1 \dfrac{d^n}{dx^n}\cos\sqrt{4bx} + C_2 \dfrac{d^n}{dx^n}\sin\sqrt{4bx} & \text{if } bx > 0, \\ C_1 \dfrac{d^n}{dx^n}\cosh\sqrt{4|bx|} + C_2 \dfrac{d^n}{dx^n}\sinh\sqrt{4|bx|} & \text{if } bx < 0. \end{cases}$$

7. $xy''_{xx} + ay'_x + bxy = 0.$

1°. The solution is expressed in terms of Bessel functions and modified Bessel functions:

$$y = \begin{cases} x^{\frac{1-a}{2}}\left[C_1 J_\nu\left(\sqrt{b}\,x\right) + C_2 Y_\nu\left(\sqrt{b}\,x\right)\right] & \text{if } b > 0, \\ x^{\frac{1-a}{2}}\left[C_1 I_\nu\left(\sqrt{|b|}\,x\right) + C_2 K_\nu\left(\sqrt{|b|}\,x\right)\right] & \text{if } b < 0, \end{cases}$$

where $\nu = \frac{1}{2}|1-a|$.

2°. For $a = 2n$, where $n = 1,\ 2,\ \dots$, the solution is

$$y = \begin{cases} C_1\left(\dfrac{1}{x}\dfrac{d}{dx}\right)^n \cos\left(x\sqrt{b}\,\right) + C_2\left(\dfrac{1}{x}\dfrac{d}{dx}\right)^n \sin\left(x\sqrt{b}\,\right) & \text{if } b > 0, \\ C_1\left(\dfrac{1}{x}\dfrac{d}{dx}\right)^n \cosh\left(x\sqrt{-b}\,\right) + C_2\left(\dfrac{1}{x}\dfrac{d}{dx}\right)^n \sinh\left(x\sqrt{-b}\,\right) & \text{if } b < 0. \end{cases}$$

8. $xy''_{xx} + ny'_x + bx^{1-2n}y = 0.$

For $n = 1$, this is the Euler equation S4.2.1.12. For $n \neq 1$, the solution is

$$y = \begin{cases} C_1 \sin\left(\dfrac{\sqrt{b}}{n-1}x^{1-n}\right) + C_2 \cos\left(\dfrac{\sqrt{b}}{n-1}x^{1-n}\right) & \text{if } b > 0, \\ C_1 \exp\left(\dfrac{\sqrt{-b}}{n-1}x^{1-n}\right) + C_2 \exp\left(\dfrac{-\sqrt{-b}}{n-1}x^{1-n}\right) & \text{if } b < 0. \end{cases}$$

9. $xy''_{xx} + ay'_x + bx^n y = 0.$

If $n = -1$ and $b = 0$, we have the Euler equation S4.2.1.12. If $n \neq -1$ and $b \neq 0$, the solution is expressed in terms of Bessel functions:

$$y = x^{\frac{1-a}{2}}\left[C_1 J_\nu\left(\frac{2\sqrt{b}}{n+1}x^{\frac{n+1}{2}}\right) + C_2 Y_\nu\left(\frac{2\sqrt{b}}{n+1}x^{\frac{n+1}{2}}\right)\right], \qquad \text{where} \quad \nu = \frac{|1-a|}{n+1}.$$

10. $xy''_{xx} + (b - x)y'_x - ay = 0.$

Degenerate hypergeometric equation.

1°. If $b \neq 0, -1, -2, -3, \ldots$, Kummer's series is a particular solution:

$$\Phi(a, b; x) = 1 + \sum_{k=1}^{\infty} \frac{(a)_k}{(b)_k} \frac{x^k}{k!},$$

where $(a)_k = a(a + 1) \ldots (a + k - 1)$, $(a)_0 = 1$. If $b > a > 0$, this solution can be written in terms of a definite integral:

$$\Phi(a, b; x) = \frac{\Gamma(b)}{\Gamma(a)\,\Gamma(b - a)} \int_0^1 e^{xt} t^{a-1} (1 - t)^{b-a-1}\,dt,$$

where $\Gamma(z) = \int_0^{\infty} e^{-t} t^{z-1}\,dt$ is the gamma function.

If b is not an integer, then the general solution has the form:

$$y = C_1 \Phi(a, b; x) + C_2 x^{1-b} \Phi(a - b + 1,\ 2 - b;\ x).$$

2°. For $b \neq 0, -1, -2, -3, \ldots$, the general solution of the degenerate hypergeometric equation can be written in the form

$$y = C_1 \Phi(a, b; x) + C_2 \Psi(a, b; x),$$

while for $b = 0, -1, -2, -3, \ldots$, it can be represented as

$$y = x^{1-b} \bigl[C_1 \Phi(a - b + 1,\ 2 - b;\ x) + C_2 \Psi(a - b + 1,\ 2 - b;\ x)\bigr].$$

The functions $\Phi(a, b; x)$ and $\Psi(a, b; x)$ are described in Subsection M13.8 in detail.

11. $(a_2 x + b_2)y''_{xx} + (a_1 x + b_1)y'_x + (a_0 x + b_0)y = 0.$

Let the function $\mathcal{J}(a, b; x)$ be an arbitrary solution of the degenerate hypergeometric equation $xy''_{xx} + (b - x)y'_x - ay = 0$ (see equation S4.2.1.10), and let the function $Z_\nu(x)$ be an arbitrary solution of the Bessel equation $x^2 y''_{xx} + xy'_x + (x^2 - \nu^2)y = 0$ (see S4.2.1.13). The results of solving the original equation are presented in Table S4.1.

TABLE S4.1

Solutions of equation S4.2.1.11 for different values of the determining parameters.

<table>
<tr><td colspan="6">Solution: $y = e^{kx} w(z)$, where $z = \dfrac{x - \mu}{\lambda}$</td></tr>
<tr><td>Constraints</td><td>k</td><td>λ</td><td>μ</td><td>w</td><td>Parameters</td></tr>
<tr><td>$a_2 \neq 0$,
$a_1^2 \neq 4a_0a_2$</td><td>$\dfrac{\sqrt{D} - a_1}{2a_2}$</td><td>$-\dfrac{a_2}{2a_2k + a_1}$</td><td>$-\dfrac{b_2}{a_2}$</td><td>$\mathcal{J}(a, b; z)$</td><td>$a = B(k)/(2a_2k + a_1)$,
$b = (a_2b_1 - a_1b_2)a_2^{-2}$</td></tr>
<tr><td>$a_2 = 0$,
$a_1 \neq 0$</td><td>$-\dfrac{a_0}{a_1}$</td><td>1</td><td>$-\dfrac{2b_2k + b_1}{a_1}$</td><td>$\mathcal{J}\bigl(a, \frac{1}{2}; \beta z^2\bigr)$</td><td>$a = B(k)/(2a_1)$,
$\beta = -a_1/(2b_2)$</td></tr>
<tr><td>$a_2 \neq 0$,
$a_1^2 = 4a_0a_2$</td><td>$-\dfrac{a_1}{2a_2}$</td><td>a_2</td><td>$-\dfrac{b_2}{a_2}$</td><td>$z^{\nu/2} Z_\nu\bigl(\beta\sqrt{z}\,\bigr)$</td><td>$\nu = 1 - (2b_2k + b_1)a_2^{-1}$,
$\beta = 2\sqrt{B(k)}$</td></tr>
<tr><td>$a_2 = a_1 = 0$,
$a_0 \neq 0$</td><td>$-\dfrac{b_1}{2b_2}$</td><td>1</td><td>$\dfrac{b_1^2 - 4b_0b_2}{4a_0b_2}$</td><td>$z^{1/2} Z_{1/3}\bigl(\beta z^{3/2}\bigr)$;
see also 2.1.2.12</td><td>$\beta = \dfrac{2}{3}\Bigl(\dfrac{a_0}{b_2}\Bigr)^{1/2}$</td></tr>
<tr><td colspan="6">Notation: $D = a_1^2 - 4a_0a_2$, $B(k) = b_2k^2 + b_1k + b_0$</td></tr>
</table>

12. $x^2y''_{xx} + axy'_x + by = 0.$

Euler equation. Solution:

$$y = \begin{cases} |x|^{\frac{1-a}{2}}\bigl(C_1|x|^{\mu} + C_2|x|^{-\mu}\bigr) & \text{if } (1-a)^2 > 4b, \\ |x|^{\frac{1-a}{2}}(C_1 + C_2\ln|x|) & \text{if } (1-a)^2 = 4b, \\ |x|^{\frac{1-a}{2}}\bigl[C_1\sin(\mu\ln|x|) + C_2\cos(\mu\ln|x|)\bigr] & \text{if } (1-a)^2 < 4b, \end{cases}$$

where $\mu = \frac{1}{2}|(1-a)^2 - 4b|^{1/2}$.

13. $x^2y''_{xx} + xy'_x + (x^2 - \nu^2)y = 0.$

Bessel equation.

1°. Let ν be an arbitrary noninteger. Then the general solution is given by

$$y = C_1J_\nu(x) + C_2Y_\nu(x), \tag{1}$$

where $J_\nu(x)$ and $Y_\nu(x)$ are the Bessel functions of the first and second kind:

$$J_\nu(x) = \sum_{k=0}^{\infty}\frac{(-1)^k(x/2)^{\nu+2k}}{k!\,\Gamma(\nu+k+1)}, \quad Y_\nu(x) = \frac{J_\nu(x)\cos\pi\nu - J_{-\nu}(x)}{\sin\pi\nu}. \tag{2}$$

Solution (1) is denoted by $y = Z_\nu(x)$, which is referred to as the cylindrical function.

The functions $J_\nu(x)$ and $Y_\nu(x)$ can be expressed in terms of definite integrals (with $x > 0$):

$$\begin{aligned} \pi J_\nu(x) &= \int_0^\pi \cos(x\sin\theta - \nu\theta)\,d\theta - \sin\pi\nu\int_0^\infty \exp(-x\sinh t - \nu t)\,dt, \\ \pi Y_\nu(x) &= \int_0^\pi \sin(x\sin\theta - \nu\theta)\,d\theta - \int_0^\infty (e^{\nu t} + e^{-\nu t}\cos\pi\nu)e^{-x\sinh t}\,dt. \end{aligned}$$

2°. In the case $\nu = n + \frac{1}{2}$, where $n = 0, 1, 2, \ldots$, the Bessel functions are expressed in terms of elementary functions:

$$J_{n+\frac{1}{2}}(x) = \sqrt{\frac{2}{\pi}}\,x^{n+\frac{1}{2}}\Bigl(-\frac{1}{x}\frac{d}{dx}\Bigr)^n\frac{\sin x}{x}, \quad J_{-n-\frac{1}{2}}(x) = \sqrt{\frac{2}{\pi}}\,x^{n+\frac{1}{2}}\Bigl(\frac{1}{x}\frac{d}{dx}\Bigr)^n\frac{\cos x}{x},$$

$$Y_{n+\frac{1}{2}}(x) = (-1)^{n+1}J_{-n-\frac{1}{2}}(x).$$

The Bessel functions are described in Section M13.6 in detail.

14. $x^2y''_{xx} + xy'_x - (x^2 + \nu^2)y = 0.$

Modified Bessel equation. It can be reduced to equation S4.2.1.13 by means of the substitution $x = i\bar{x}$ $(i^2 = -1)$.

Solution:

$$y = C_1I_\nu(x) + C_2K_\nu(x),$$

where $I_\nu(x)$ and $K_\nu(x)$ are modified Bessel functions of the first and second kind:

$$I_\nu(x) = \sum_{k=0}^{\infty}\frac{(x/2)^{2k+\nu}}{k!\,\Gamma(\nu+k+1)}, \qquad K_\nu(x) = \frac{\pi}{2}\frac{I_{-\nu}(x) - I_\nu(x)}{\sin\pi\nu}.$$

The modified Bessel functions are described in Section M13.7 in detail.

15. $x^2y''_{xx} + axy'_x + (bx^n + c)y = 0, \qquad n \neq 0.$

The case $b = 0$ corresponds to the Euler equation S4.2.1.12.

For $b \neq 0$, the solution is

$$y = x^{\frac{1-a}{2}}\Big[C_1 J_\nu\Big(\frac{2}{n}\sqrt{b}\,x^{\frac{n}{2}}\Big) + C_2 Y_\nu\Big(\frac{2}{n}\sqrt{b}\,x^{\frac{n}{2}}\Big)\Big],$$

where $\nu = \frac{1}{n}\sqrt{(1-a)^2 - 4c}$; $J_\nu(z)$ and $Y_\nu(z)$ are the Bessel functions of the first and second kind.

16. $x^2y''_{xx} + axy'_x + x^n(bx^n + c)y = 0.$

The substitution $\xi = x^n$ leads to an equation of the form S4.2.1.11:

$$n^2\xi y''_{\xi\xi} + n(n-1+a)y'_\xi + (b\xi + c)y = 0.$$

17. $x^2y''_{xx} + (ax + b)y'_x + cy = 0.$

The transformation $x = z^{-1}$, $y = z^k e^z w$, where k is a root of the quadratic equation $k^2 + (1-a)k + c = 0$, leads to an equation of the form S4.2.1.11:

$$zw''_{zz} + [(2-b)z + 2k + 2 - a]w'_z + [(1-b)z + 2k + 2 - a - bk]w = 0.$$

18. $(1 - x^2)y''_{xx} - 2xy'_x + n(n+1)y = 0, \qquad n = 0,\ 1,\ 2,\ \dots$

Legendre equation.

The solution is given by

$$y = C_1 P_n(x) + C_2 Q_n(x),$$

where the Legendre polynomials $P_n(x)$ and the Legendre functions of the second kind $Q_n(x)$ are given by the formulas

$$P_n(x) = \frac{1}{n!\,2^n}\frac{d^n}{dx^n}(x^2-1)^n, \quad Q_n(x) = \frac{1}{2}P_n(x)\ln\frac{1+x}{1-x} - \sum_{m=1}^{n}\frac{1}{m}P_{m-1}(x)P_{n-m}(x).$$

The functions $P_n = P_n(x)$ can be conveniently calculated using the recurrence relations

$$P_0(x)=1,\ \ P_1(x)=x,\ \ P_2(x)=\frac{1}{2}(3x^2-1),\ \dots,\ \ P_{n+1}(x)=\frac{2n+1}{n+1}xP_n(x)-\frac{n}{n+1}P_{n-1}(x).$$

Three leading functions $Q_n = Q_n(x)$ are

$$Q_0(x) = \frac{1}{2}\ln\frac{1+x}{1-x}, \quad Q_1(x) = \frac{x}{2}\ln\frac{1+x}{1-x} - 1, \quad Q_2(x) = \frac{3x^2-1}{4}\ln\frac{1+x}{1-x} - \frac{3}{2}x.$$

The Legendre polynomials and the Legendre functions are described in Subsection M13.10.1 in more detail.

19. $(1-x^2)y''_{xx} - 2xy'_x + \nu(\nu+1)y = 0.$

Legendre equation; ν is an arbitrary number. The case $\nu = n$ where n is a nonnegative integer is considered in S4.2.1.18.

The substitution $z = x^2$ leads to the hypergeometric equation. Therefore, with $|x| < 1$ the solution can be written as

$$y = C_1 F\Big(-\frac{\nu}{2},\ \frac{1+\nu}{2},\ \frac{1}{2};\ x\Big) + C_2 x F\Big(\frac{1-\nu}{2},\ 1+\frac{\nu}{2},\ \frac{3}{2};\ x\Big),$$

where $F(\alpha, \beta, \gamma; x)$ is the hypergeometric series (see equation S4.2.1.22).

20. $(ax^2 + b)y''_{xx} + axy'_x + cy = 0.$

The substitution $z = \int \frac{dx}{\sqrt{ax^2+b}}$ leads to a constant coefficient linear equation: $y''_{zz} + cy = 0$.

21. $(1-x^2)y''_{xx} + (ax+b)y'_x + cy = 0.$

1°. The substitution $2z = 1 + x$ leads to the hypergeometric equation S4.2.1.22:

$$z(1-z)y''_{zz} + [az + \tfrac{1}{2}(b-a)]y'_z + cy = 0.$$

2°. For $a = -2m-3$, $b = 0$, and $c = \lambda$, the Gegenbauer functions are solutions of the equation.

22. $x(x-1)y''_{xx} + [(\alpha+\beta+1)x - \gamma]y'_x + \alpha\beta y = 0.$

Gaussian hypergeometric equation. For $\gamma \neq 0, -1, -2, -3, \ldots$, a solution can be expressed in terms of the hypergeometric series:

$$F(\alpha, \beta, \gamma; x) = 1 + \sum_{k=1}^{\infty} \frac{(\alpha)_k (\beta)_k}{(\gamma)_k} \frac{x^k}{k!}, \quad (\alpha)_k = \alpha(\alpha+1)\ldots(\alpha+k-1),$$

which, *a fortiori*, is convergent for $|x| < 1$.

For $\gamma > \beta > 0$, this solution can be expressed in terms of a definite integral:

$$F(\alpha, \beta, \gamma; x) = \frac{\Gamma(\gamma)}{\Gamma(\beta)\,\Gamma(\gamma-\beta)} \int_0^1 t^{\beta-1}(1-t)^{\gamma-\beta-1}(1-tx)^{-\alpha}\,dt,$$

where $\Gamma(\beta)$ is the gamma function.

If γ is not an integer, the general solution of the hypergeometric equation has the form:

$$y = C_1 F(\alpha, \beta, \gamma; x) + C_2 x^{1-\gamma} F(\alpha-\gamma+1,\ \beta-\gamma+1,\ 2-\gamma;\ x).$$

In the degenerate cases $\gamma = 0, -1, -2, -3, \ldots$, a particular solution of the hypergeometric equation corresponds to $C_1 = 0$ and $C_2 = 1$. If γ is a positive integer, another particular solution corresponds to $C_1 = 1$ and $C_2 = 0$. In both these cases, the general solution can be constructed by means of the last formula given in the preliminary remarks at the beginning of Section S4.2.

Table S4.2 gives the general solutions of the hypergeometric equation for some values of the determining parameters.

TABLE S4.2
General solutions of the hypergeometric equation for some values of the determining parameters.

α	β	γ	Solution: $y = y(x)$
0	β	γ	$C_1 + C_2 \int \lvert x\rvert^{-\gamma} \lvert x-1\rvert^{\gamma-\beta-1}\,dx$
α	$\alpha+\frac{1}{2}$	$2\alpha+1$	$C_1\bigl(1+\sqrt{1-x}\bigr)^{-2\alpha} + C_2 x^{-2\alpha}\bigl(1+\sqrt{1-x}\bigr)^{2\alpha}$
α	$\alpha-\frac{1}{2}$	$\frac{1}{2}$	$C_1\bigl(1+\sqrt{x}\bigr)^{1-2\alpha} + C_2\bigl(1-\sqrt{x}\bigr)^{1-2\alpha}$
α	$\alpha+\frac{1}{2}$	$\frac{3}{2}$	$\frac{1}{\sqrt{x}}\Bigl[C_1\bigl(1+\sqrt{x}\bigr)^{1-2\alpha} + C_2\bigl(1-\sqrt{x}\bigr)^{1-2\alpha}\Bigr]$
1	β	γ	$\lvert x\rvert^{1-\gamma}\lvert x-1\rvert^{\gamma-\beta-1}\Bigl(C_1 + C_2 \int \lvert x\rvert^{\gamma-2}\lvert x-1\rvert^{\beta-\gamma}\,dx\Bigr)$
α	β	α	$\lvert x-1\rvert^{-\beta}\Bigl(C_1 + C_2 \int \lvert x\rvert^{-\alpha}\lvert x-1\rvert^{\beta-1}\,dx\Bigr)$
α	β	$\alpha+1$	$\lvert x\rvert^{-\alpha}\Bigl(C_1 + C_2 \int \lvert x\rvert^{\alpha-1}\lvert x-1\rvert^{-\beta}\,dx\Bigr)$

23. $\boldsymbol{(1-x^2)^2 y''_{xx} - 2x(1-x^2)y'_x + [\nu(\nu+1)(1-x^2) - \mu^2]y = 0.}$

Legendre equation, ν and μ are arbitrary parameters.

The transformation $x = 1 - 2\xi$, $y = |x^2 - 1|^{\mu/2} w$ leads to the hypergeometric equation S4.2.1.22:

$$\xi(\xi-1)w''_{\xi\xi} + (\mu+1)(1-2\xi)w'_{\xi} + (\nu-\mu)(\nu+\mu+1)w = 0$$

with parameters $\alpha = \mu - \nu$, $\beta = \mu + \nu + 1$, $\gamma = \mu + 1$.

In particular, the original equation is integrable by quadrature if $\nu = \mu$ or $\nu = -\mu - 1$.

24. $\boldsymbol{(x-a)^2(x-b)^2 y''_{xx} - cy = 0, \qquad a \neq b.}$

The transformation $\xi = \ln\left|\dfrac{x-a}{x-b}\right|$, $y = (x-b)\eta$ leads to a constant coefficient linear equation: $(a-b)^2(\eta''_{\xi\xi} - \eta'_{\xi}) - c\eta = 0$. Therefore, the solution is as follows:

$$y = C_1|x-a|^{(1+\lambda)/2}|x-b|^{(1-\lambda)/2} + C_2|x-a|^{(1-\lambda)/2}|x-b|^{(1+\lambda)/2},$$

where $\lambda^2 = 4c(a-b)^{-2} + 1 \neq 0$.

25. $\boldsymbol{(ax^2 + bx + c)^2 y''_{xx} + Ay = 0.}$

The transformation $\xi = \displaystyle\int \frac{dx}{ax^2+bx+c}$, $w = \dfrac{y}{\sqrt{|ax^2+bx+c|}}$ leads to a constant coefficient linear equation of the form S4.2.1.1: $w''_{\xi\xi} + (A + ac - \frac{1}{4}b^2)w = 0$.

26. $\boldsymbol{x^2(ax^n - 1)y''_{xx} + x(apx^n + q)y'_x + (arx^n + s)y = 0.}$

Find the roots A_1, A_2 and B_1, B_2 of the quadratic equations

$$A^2 - (q+1)A - s = 0, \quad B^2 - (p-1)B + r = 0$$

and define parameters c, α, β, and γ by the relations

$$c = A_1, \quad \alpha = (A_1 + B_1)n^{-1}, \quad \beta = (A_1 + B_2)n^{-1}, \quad \gamma = 1 + (A_1 - A_2)n^{-1}.$$

Then the solution of the original equation has the form $y = x^c u(ax^n)$, where $u = u(z)$ is the general solution of the hypergeometric equation S4.2.1.22: $z(z-1)u''_{zz} + [(\alpha+\beta+1)z-\gamma]u'_z + \alpha\beta u = 0$.

S4.2.2. Equations Involving Exponential and Other Functions

1. $y''_{xx} + ae^{\lambda x}y = 0, \qquad \lambda \neq 0.$

Solution: $y = C_1 J_0(z) + C_2 Y_0(z)$, where $z = 2\lambda^{-1}\sqrt{a}\,e^{\lambda x/2}$; $J_0(z)$ and $Y_0(z)$ are Bessel functions.

2. $y''_{xx} + (ae^x - b)y = 0.$

Solution: $y = C_1 J_{2\sqrt{b}}\bigl(2\sqrt{a}\,e^{x/2}\bigr) + C_2 Y_{2\sqrt{b}}\bigl(2\sqrt{a}\,e^{x/2}\bigr)$, where $J_\nu(z)$ and $Y_\nu(z)$ are Bessel functions.

3. $y''_{xx} - (ae^{2\lambda x} + be^{\lambda x} + c)y = 0.$

The transformation $z = e^{\lambda x}$, $w = z^{-k}y$, where $k = \sqrt{c}/\lambda$, leads to an equation of the form S4.2.1.11: $\lambda^2 z w''_{zz} + \lambda^2(2k+1)w'_z - (az+b)w = 0$.

4. $y''_{xx} + ay'_x + be^{2ax}y = 0.$

The transformation $\xi = e^{ax}$, $u = ye^{ax}$ leads to a constant coefficient linear equation of the form S4.2.1.1: $u''_{\xi\xi} + ba^{-2}u = 0$.

5. $y''_{xx} - ay'_x + be^{2ax}y = 0.$

The substitution $\xi = e^{ax}$ leads to a constant coefficient linear equation of the form S4.2.1.1: $y''_{\xi\xi} + ba^{-2}y = 0$.

6. $y''_{xx} + ay'_x + (be^{\lambda x} + c)y = 0.$

Solution: $y = e^{-ax/2}\bigl[C_1 J_\nu\bigl(2\lambda^{-1}\sqrt{b}\,e^{\lambda x/2}\bigr) + C_2 Y_\nu\bigl(2\lambda^{-1}\sqrt{b}\,e^{\lambda x/2}\bigr)\bigr]$, where $\nu = \frac{1}{\lambda}\sqrt{a^2 - 4c}$; $J_\nu(z)$ and $Y_\nu(z)$ are Bessel functions.

7. $y''_{xx} - (a - 2q\cosh 2x)y = 0.$

Modified Mathieu equation. The substitution $x = i\xi$ leads to the Mathieu equation S4.2.2.8:

$$y''_{\xi\xi} + (a - 2q\cos 2\xi)y = 0.$$

For eigenvalues $a = a_n(q)$ and $a = b_n(q)$, the corresponding solutions of the modified Mathieu equation are

$$\mathrm{Ce}_{2n+p}(x,q) = \mathrm{ce}_{2n+p}(ix,q) = \sum_{k=0}^{\infty} A^{2n+p}_{2k+p}\cosh[(2k+p)x],$$

$$\mathrm{Se}_{2n+p}(x,q) = -i\,\mathrm{se}_{2n+p}(ix,q) = \sum_{k=0}^{\infty} B^{2n+p}_{2k+p}\sinh[(2k+p)x],$$

where p can be either 0 or 1, and the coefficients A^{2n+p}_{2k+p} and B^{2n+p}_{2k+p} are specified in S4.2.2.8.

The modified Mathieu functions are discussed in the books by Bateman & Erdélyi (1955, vol. 3) and Abramowitz & Stegun (1964) in more detail.

8. $y''_{xx} + (a - 2q\cos 2x)y = 0.$

Mathieu equation.

1°. Given numbers a and q, there exists a general solution $y(x)$ and a characteristic index μ such that

$$y(x+\pi) = e^{2\pi\mu} y(x).$$

For small values of q, an approximate value of μ can be found from the equation

$$\cosh(\pi\mu) = 1 + 2\sin^2\bigl(\tfrac{1}{2}\pi\sqrt{a}\bigr) + \frac{\pi q^2}{(1-a)\sqrt{a}}\sin\bigl(\pi\sqrt{a}\bigr) + O(q^4).$$

If $y_1(x)$ is the solution of the Mathieu equation satisfying the initial conditions $y_1(0) = 1$ and $y'_1(0) = 0$, the characteristic index can be determined from the relation

$$\cosh(2\pi\mu) = y_1(\pi).$$

The solution $y_1(x)$, and hence μ, can be determined with any degree of accuracy by means of numerical or approximate methods.

The general solution differs depending on the value of $y_1(\pi)$ and can be expressed in terms of two auxiliary periodical functions $\varphi_1(x)$ and $\varphi_2(x)$ (see Table S4.3).

TABLE S4.3
The general solution of the Mathieu equation S4.2.2.8 expressed in terms of auxiliary periodical functions $\varphi_1(x)$ and $\varphi_2(x)$.

Constraint	General solution $y = y(x)$	Period of φ_1 and φ_2	Index
$y_1(\pi) > 1$	$C_1e^{2\mu x}\varphi_1(x) + C_2e^{-2\mu x}\varphi_2(x)$	π	μ is a real number
$y_1(\pi) < -1$	$C_1e^{2\rho x}\varphi_1(x) + C_2e^{-2\rho x}\varphi_2(x)$	2π	$\mu = \rho + \frac{1}{2}i,\ \ i^2 = -1$, ρ is the real part of μ
$\lvert y_1(\pi)\rvert < 1$	$(C_1\cos\nu x + C_2\sin\nu x)\varphi_1(x)$ $+(C_1\cos\nu x - C_2\sin\nu x)\varphi_2(x)$	π	$\mu = i\nu$ is a pure imaginary number, $\cos(2\pi\nu) = y_1(\pi)$
$y_1(\pi) = \pm 1$	$C_1\varphi_1(x) + C_2x\varphi_2(x)$	π	$\mu = 0$

2°. In applications, of major interest are periodical solutions of the Mathieu equation that exist for certain values of the parameters a and q (those values of a are referred to as eigenvalues). The most important solutions are listed in Table S4.4.

The Mathieu functions possess the following properties:

$$\mathrm{ce}_{2n}(x,-q) = (-1)^n\,\mathrm{ce}_{2n}\Bigl(\frac{\pi}{2}-x,\ q\Bigr), \qquad \mathrm{ce}_{2n+1}(x,-q) = (-1)^n\,\mathrm{se}_{2n+1}\Bigl(\frac{\pi}{2}-x,\ q\Bigr),$$
$$\mathrm{se}_{2n}(x,-q) = (-1)^{n-1}\,\mathrm{se}_{2n}\Bigl(\frac{\pi}{2}-x,\ q\Bigr), \qquad \mathrm{se}_{2n+1}(x,-q) = (-1)^n\,\mathrm{ce}_{2n+1}\Bigl(\frac{\pi}{2}-x,\ q\Bigr).$$

Selecting a sufficiently large m and omitting the term with the maximum number in the recurrence relations (indicated in Table S4.4), we can obtain approximate relations for the eigenvalues a_n (or b_n) with respect to parameter q. Then, equating the determinant of the corresponding homogeneous linear system of equations for coefficients A^n_m (or B^n_m) to zero, we obtain an algebraic equation for finding $a_n(q)$ (or $b_n(q)$).

TABLE S4.4
Periodical solutions of the Mathieu equation $\mathrm{ce}_n = \mathrm{ce}_n(x,q)$ and $\mathrm{se}_n = \mathrm{se}_n(x,q)$ (for odd n, the functions ce_n and se_n are 2π-periodical, and for even n, they are π-periodical); certain eigenvalues $a = a_n(q)$ and $b = b_n(q)$ correspond to each value of the parameter q; $n = 0,\ 1,\ 2,\ \dots$.

Mathieu functions	Recurrence relations for coefficients	Normalization conditions
$\mathrm{ce}_{2n}(x,q)=\sum\limits_{m=0}^{\infty} A_{2m}^{2n}\cos(2mx)$	$qA_2^{2n}=a_{2n}A_0^{2n}$; $qA_4^{2n}=(a_{2n}-4)A_2^{2n}-2qA_0^{2n}$; $qA_{2m+2}^{2n}=(a_{2n}-4m^2)A_{2m}^{2n}$ $-qA_{2m-2}^{2n},\quad m\geq 2$	$(A_0^{2n})^2+\sum\limits_{m=0}^{\infty}\big(A_{2m}^{2n}\big)^2$ $=\begin{cases}2 & \text{if } n=0\\ 1 & \text{if } n\geq 1\end{cases}$
$\mathrm{ce}_{2n+1}(x,q)=\sum\limits_{m=0}^{\infty} A_{2m+1}^{2n+1}\cos\big[(2m+1)x\big]$	$qA_3^{2n+1}=(a_{2n+1}-1-q)A_1^{2n+1}$; $qA_{2m+3}^{2n+1}=[a_{2n+1}-(2m+1)^2]$ $\times A_{2m+1}^{2n+1}-qA_{2m-1}^{2n+1},\quad m\geq 1$	$\sum\limits_{m=0}^{\infty}\big(A_{2m+1}^{2n+1}\big)^2=1$
$\mathrm{se}_{2n}(x,q)=\sum\limits_{m=0}^{\infty} B_{2m}^{2n}\sin(2mx)$, $\mathrm{se}_0=0$	$qB_4^{2n}=(b_{2n}-4)B_2^{2n}$; $qB_{2m+2}^{2n}=(b_{2n}-4m^2)B_{2m}^{2n}$ $-qB_{2m-2}^{2n},\quad m\geq 2$	$\sum\limits_{m=0}^{\infty}\big(B_{2m}^{2n}\big)^2=1$
$\mathrm{se}_{2n+1}(x,q)=\sum\limits_{m=0}^{\infty} B_{2m+1}^{2n+1}\sin\big[(2m+1)x\big]$	$qB_3^{2n+1}=(b_{2n+1}-1-q)B_1^{2n+1}$; $qB_{2m+3}^{2n+1}=[b_{2n+1}-(2m+1)^2]$ $\times B_{2m+1}^{2n+1}-qB_{2m-1}^{2n+1},\quad m\geq 1$	$\sum\limits_{m=0}^{\infty}\big(B_{2m+1}^{2n+1}\big)^2=1$

For fixed real $q \neq 0$, the eigenvalues a_n and b_n are all real and different, while:

$$\begin{aligned} &\text{if}\quad q>0\quad\text{then}\quad a_0<b_1<a_1<b_2<a_2<\cdots;\\ &\text{if}\quad q<0\quad\text{then}\quad a_0<a_1<b_1<b_2<a_2<a_3<b_3<b_4<\cdots \end{aligned}$$

The eigenvalues possess the following properties:

$$a_{2n}(-q)=a_{2n}(q),\quad b_{2n}(-q)=b_{2n}(q),\quad a_{2n+1}(-q)=b_{2n+1}(q).$$

The solution of the Mathieu equation corresponding to eigenvalue a_n (or b_n) has n zeros on the interval $0 \leq x < \pi$ (q is a real number).

Listed below are two leading terms of asymptotic expansions of the Mathieu functions $\mathrm{ce}_n(x,q)$ and $\mathrm{se}_n(x,q)$, as well as of the corresponding eigenvalues $a_n(q)$ and $b_n(q)$, as $q \to 0$:

$$\mathrm{ce}_0(x,q)=\frac{1}{\sqrt{2}}\Big(1-\frac{q}{2}\cos 2x\Big),\quad a_0(q)=-\frac{q^2}{2}+\frac{7q^4}{128};$$

$$\mathrm{ce}_1(x,q)=\cos x-\frac{q}{8}\cos 3x,\quad a_1(q)=1+q;$$

$$\mathrm{ce}_2(x,q)=\cos 2x+\frac{q}{4}\Big(1-\frac{\cos 4x}{3}\Big),\quad a_2(q)=4+\frac{5q^2}{12};$$

$$\mathrm{ce}_n(x,q)=\cos nx+\frac{q}{4}\Big[\frac{\cos(n+2)x}{n+1}-\frac{\cos(n-2)x}{n-1}\Big],\quad a_n(q)=n^2+\frac{q^2}{2(n^2-1)}\qquad(n\geq 3);$$

$$\mathrm{se}_1(x,q)=\sin x-\frac{q}{8}\sin 3x,\quad b_1(q)=1-q;$$

$$\mathrm{se}_2(x,q)=\sin 2x-\frac{q}{12}\sin 4x,\quad b_2(q)=4-\frac{q^2}{12};$$

$$\operatorname{se}_n(x,q) = \sin nx - \frac{q}{4}\left[\frac{\sin(n+2)x}{n+1} - \frac{\sin(n-2)x}{n-1}\right], \quad b_n(q) = n^2 + \frac{q^2}{2(n^2-1)} \qquad (n \ge 3).$$

The Mathieu functions are discussed in the books by Bateman & Erdélyi (1955) and Abramowitz & Stegun (1964) in more detail.

9. $y''_{xx} + a\tan x\, y'_x + by = 0.$

1°. The substitution $\xi = \sin x$ leads to a linear equation of the form S4.2.1.21: $(\xi^2-1)y''_{\xi\xi} + (1-a)\xi y'_\xi - by = 0$.

2°. Solution for $a = -2$:

$$y\cos x = \begin{cases} C_1\sin(kx) + C_2\cos(kx) & \text{if } b+1 = k^2 > 0, \\ C_1\sinh(kx) + C_2\cosh(kx) & \text{if } b+1 = -k^2 < 0. \end{cases}$$

3°. Solution for $a = 2$ and $b = 3$: $y = C_1\cos^3 x + C_2\sin x\,(1 + 2\cos^2 x)$.

S4.2.3. Equations Involving Arbitrary Functions

► *Notation:* $f = f(x)$ *and* $g = g(x)$ *are arbitrary functions;* a, b, *and* λ *are arbitrary parameters.*

1. $y''_{xx} + fy'_x + a(f-a)y = 0.$

Particular solution: $y_0 = e^{-ax}$.

2. $y''_{xx} + xfy'_x - fy = 0.$

Particular solution: $y_0 = x$.

3. $xy''_{xx} + (xf + a)y'_x + (a-1)fy = 0.$

Particular solution: $y_0 = x^{1-a}$.

4. $xy''_{xx} + [(ax+1)f + ax - 1]y'_x + a^2xfy = 0.$

Particular solution: $y_0 = (ax+1)e^{-ax}$.

5. $xy''_{xx} + [(ax^2+bx)f + 2]y'_x + bfy = 0.$

Particular solution: $y_0 = a + b/x$.

6. $x^2y''_{xx} + xfy'_x + a(f-a-1)y = 0.$

Particular solution: $y_0 = x^{-a}$.

7. $y''_{xx} + (f + ae^{\lambda x})y'_x + ae^{\lambda x}(f+\lambda)y = 0.$

Particular solution: $y_0 = \exp\left(-\frac{a}{\lambda}e^{\lambda x}\right)$.

8. $y''_{xx} - (f^2 + f'_x)y = 0.$

Particular solution: $y_0 = \exp\left(\int f\,dx\right)$.

9. $\boldsymbol{y''_{xx} + 2fy'_x + (f^2 + f'_x)y = 0.}$

Solution: $y = (C_2x + C_1)\exp\Bigl(-\int f\,dx\Bigr)$.

10. $\boldsymbol{y''_{xx} + (1-a)fy'_x - a(f^2 + f'_x)y = 0.}$

Particular solution: $y_0 = \exp\Bigl(a\int f\,dx\Bigr)$.

11. $\boldsymbol{y''_{xx} + fy'_x + (fg - g^2 + g'_x)y = 0.}$

Particular solution: $y_0 = \exp\Bigl(-\int g\,dx\Bigr)$.

12. $\boldsymbol{fy''_{xx} - af'_xy'_x - bf^{2a+1}y = 0.}$

Solution: $y = C_1e^u + C_2e^{-u}$, where $u = \sqrt{b}\int f^a\,dx$.

13. $\boldsymbol{f^2y''_{xx} + f(f'_x + a)y'_x + by = 0.}$

The substitution $\xi = \int f^{-1}dx$ leads to a constant coefficient linear equation: $y''_{\xi\xi} + ay'_\xi + by = 0$.

14. $\boldsymbol{y''_{xx} - f'_xy'_x + a^2e^{2f}y = 0.}$

Solution: $y = C_1\sin\Bigl(a\int e^f\,dx\Bigr) + C_2\cos\Bigl(a\int e^f\,dx\Bigr)$.

15. $\boldsymbol{y''_{xx} - f'_xy'_x - a^2e^{2f}y = 0.}$

Solution: $y = C_1\exp\Bigl(a\int e^f\,dx\Bigr) + C_2\exp\Bigl(-a\int e^f\,dx\Bigr)$.

S4.3. Second-Order Nonlinear Equations

S4.3.1. Equations of the Form $y''_{xx} = f(x, y)$

1. $\boldsymbol{y''_{xx} = f(y).}$

Autonomous equation.

Solution: $\int\Bigl[C_1 + 2\int f(y)\,dy\Bigr]^{-1/2}dy = C_2 \pm x$.

Particular solutions: $y = A_k$, where A_k are roots of the algebraic (transcendental) equation $f(A_k) = 0$.

2. $\boldsymbol{y''_{xx} = Ax^ny^m.}$

Emden–Fowler equation.

1°. With $m \neq 1$, the Emden–Fowler equation has a particular solution:

$$y = \lambda x^{\frac{n+2}{1-m}}, \quad \text{where } \lambda = \left[\frac{(n+2)(n+m+1)}{A(m-1)^2}\right]^{\frac{1}{m-1}}.$$

2°. The transformation $z = x^{n+2}y^{m-1}$, $w = xy'_x/y$ leads to a first-order (Abel) equation: $z[(m-1)w + n + 2]w'_z = -w^2 + w + Az$.

3°. The transformation $y = w/t,\ x = 1/t$ leads to the Emden–Fowler equation with the independent variable raised to a different power: $w''_{tt} = At^{-n-m-3}w^m$.

4°. The books by Zaitsev and Polyanin (1994) and Polyanin and Zaitsev (2003) present 28 solvable cases of the Emden–Fowler equation (corresponding to some pairs of n and m).

3. $\boldsymbol{y''_{xx} + f(x)y = ay^{-3}}$.

Ermakov's equation. Let $w = w(x)$ be a nontrivial solution of the second-order linear equation $w''_{xx} + f(x)w = 0$. The transformation $\xi = \int \frac{dx}{w^2},\ z = \frac{y}{w}$ leads to an autonomous equation of the form S4.3.1: $z''_{\xi\xi} = az^{-3}$.

Solution: $C_1 y^2 = aw^2 + w^2\Bigl(C_2 + C_1 \int \frac{dx}{w^2}\Bigr)^2$.

► *Further on, f, g, h, and ψ are arbitrary composite functions of their arguments indicated in parentheses after the function name (the arguments can depend on x, y, y'_x).*

4. $\boldsymbol{y''_{xx} = f(ay + bx + c)}$.

The substitution $w = ay + bx + c$ leads to an equation of the form S4.3.1.1: $w''_{xx} = af(w)$.

5. $\boldsymbol{y''_{xx} = f(y + ax^2 + bx + c)}$.

The substitution $w = y + ax^2 + bx + c$ leads to an equation of the form S4.3.1.1: $w''_{xx} = f(w) + 2a$.

6. $\boldsymbol{y''_{xx} = x^{-1} f(yx^{-1})}$.

Homogeneous equation. The transformation $t = -\ln|x|,\ z = y/x$ leads to an autonomous equation: $z''_{tt} - z'_t = f(z)$.

7. $\boldsymbol{y''_{xx} = x^{-3} f(yx^{-1})}$.

The transformation $\xi = 1/x,\ w = y/x$ leads to the equation of the form S4.3.1.1: $w''_{\xi\xi} = f(w)$.

8. $\boldsymbol{y''_{xx} = x^{-3/2} f(yx^{-1/2})}$.

Having set $w = yx^{-1/2}$, we obtain $\frac{d}{dx}(xw'_x)^2 = \frac{1}{2}ww'_x + 2f(w)w'_x$. Integrating the latter equation, we arrive at a separable equation.

Solution: $\int \Bigl[C_1 + \frac{1}{4}w^2 + 2\int f(w)\,dw\Bigr]^{-1/2} dw = C_2 \pm \ln x$.

9. $\boldsymbol{y''_{xx} = x^{k-2} f(x^{-k}y)}$.

Generalized homogeneous equation. The transformation $z = x^{-k}y,\ w = xy'_x/y$ leads to a first-order equation: $z(w - k)w'_z = z^{-1}f(z) + w - w^2$.

10. $\boldsymbol{y''_{xx} = yx^{-2} f(x^n y^m)}$.

Generalized homogeneous equation. The transformation $z = x^n y^m,\ w = xy'_x/y$ leads to a first-order equation: $z(mw + n)w'_z = f(z) + w - w^2$.

11. $\boldsymbol{y''_{xx} = y^{-3} f\left(\dfrac{y}{\sqrt{ax^2+bx+c}}\right).}$

Setting $u(x) = y(ax^2+bx+c)^{-1/2}$ and integrating the equation, we obtain a first-order separable equation:

$$(ax^2+bx+c)^2(u'_x)^2 = (\tfrac{1}{4}b^2 - ac)u^2 + 2\int u^{-3} f(u)\,du + C_1.$$

12. $\boldsymbol{y''_{xx} = e^{-ax} f(e^{ax}y).}$

The transformation $z = e^{ax}y$, $w = y'_x/y$ leads to a first-order equation: $z(w+a)w'_z = z^{-1}f(z) - w^2$.

13. $\boldsymbol{y''_{xx} = y f(e^{ax}y^m).}$

The transformation $z = e^{ax}y^m$, $w = y'_x/y$ leads to a first-order equation: $z(mw+a)w'_z = f(z) - w^2$.

14. $\boldsymbol{y''_{xx} = x^{-2} f(x^n e^{ay}).}$

The transformation $z = x^n e^{ay}$, $w = xy'_x$ leads to a first-order equation: $z(aw+n)w'_z = f(z) + w$.

15. $\boldsymbol{y''_{xx} = \dfrac{\psi''_{xx}}{\psi} y + \psi^{-3} f\left(\dfrac{y}{\psi}\right), \qquad \psi = \psi(x).}$

The transformation $\xi = \displaystyle\int \frac{dx}{\psi^2}$, $w = \dfrac{y}{\psi}$ leads to an equation of the form S4.3.1.1: $w''_{\xi\xi} = f(w)$.

Solution: $\displaystyle\int \left[C_1 + 2\int f(w)\,dw\right]^{-1/2} dw = C_2 \pm \int \frac{dx}{\psi^2(x)}$.

S4.3.2. Equations of the Form $f(x,y)y''_{xx} = g(x,y,y'_x)$

1. $\boldsymbol{y''_{xx} - y'_x = f(y).}$

Autonomous equation. The substitution $w(y) = y'_x$ leads to a first-order equation. For solvable equations of the form in question, see the book by Polyanin and Zaitsev (2003).

2. $\boldsymbol{y''_{xx} + f(y)y'_x + g(y) = 0.}$

Lienard equation. The substitution $w(y) = y'_x$ leads to a first-order equation. For solvable equations of the form in question, see the book by Polyanin and Zaitsev (2003).

3. $\boldsymbol{y''_{xx} + [ay + f(x)]y'_x + f'_x(x)y = 0.}$

Integrating yields a Riccati equation: $y'_x + f(x)y + \frac{1}{2}ay^2 = C$.

4. $\boldsymbol{y''_{xx} + [2ay + f(x)]y'_x + af(x)y^2 = g(x).}$

On setting $u = y'_x + ay^2$, we obtain a first-order linear equation: $u'_x + f(x)u = g(x)$.

5. $\boldsymbol{y''_{xx} = ay'_x + e^{2ax} f(y).}$

Solution: $\displaystyle\int\Big[C_1 + 2\int f(y)\,dy\Big]^{-1/2} dy = C_2 \pm \frac{1}{a}e^{ax}.$

6. $\boldsymbol{y''_{xx} = f(y)y'_x.}$

Solution: $\displaystyle\int \frac{dy}{F(y)+C_1} = C_2 + x,$ where $F(y) = \displaystyle\int f(y)\,dy.$

7. $\boldsymbol{y''_{xx} = \big[e^{\alpha x} f(y) + \alpha\big] y'_x.}$

The substitution $w(y) = e^{-\alpha x} y'_x$ leads to a first-order separable equation: $w'_y = f(y)$.

Solution: $\displaystyle\int \frac{dy}{F(y)+C_1} = C_2 + \frac{1}{\alpha}e^{\alpha x},$ where $F(y) = \displaystyle\int f(y)\,dy.$

8. $\boldsymbol{xy''_{xx} = ny'_x + x^{2n+1} f(y).}$

1°. Solution for $n \neq -1$:

$$\int\Big[C_1 + 2\int f(y)\,dy\Big]^{-1/2} dy = \pm\frac{x^{n+1}}{n+1} + C_2.$$

2°. Solution for $n = -1$:

$$\int\Big[C_1 + 2\int f(y)\,dy\Big]^{-1/2} dy = \pm\ln|x| + C_2.$$

9. $\boldsymbol{xy''_{xx} = f(y)y'_x.}$

The substitution $w(y) = xy'_x/y$ leads to a first-order linear equation: $yw'_y = -w + 1 + f(y)$.

10. $\boldsymbol{xy''_{xx} = \big[x^k f(y) + k - 1\big] y'_x.}$

Solution: $\displaystyle\int \frac{dy}{F(y)+C_1} = C_2 + \frac{1}{k}x^k,$ where $F(y) = \displaystyle\int f(y)\,dy.$

11. $\boldsymbol{x^2 y''_{xx} + xy'_x = f(y).}$

The substitution $x = \pm e^t$ leads to an autonomous equation of the form S4.3.1.1: $y''_{tt} = f(y)$.

12. $\boldsymbol{(ax^2 + b)y''_{xx} + axy'_x + f(y) = 0.}$

The substitution $\xi = \displaystyle\int \frac{dx}{\sqrt{ax^2+b}}$ leads to an autonomous equation of the form S4.3.1.1: $y''_{\xi\xi} + f(y) = 0$.

13. $\boldsymbol{y''_{xx} = f(y)y'_x + g(x).}$

Integrating yields a first-order equation: $y'_x = \displaystyle\int f(y)\,dy + \int g(x)\,dx + C.$

14. $\boldsymbol{xy''_{xx} + (n+1)y'_x = x^{n-1} f(yx^n).}$

The transformation $\xi = x^n$, $w = yx^n$ leads to an autonomous equation of the form S4.3.1.1: $n^2 w''_{\xi\xi} = f(w)$.

15. $gy''_{xx} + \frac{1}{2}g'_x y'_x = f(y), \qquad g = g(x).$

Integrating yields a first-order separable equation: $g(x)(y'_x)^2 = 2\int f(y)\,dy + C_1$.

Solution for $g(x) \ge 0$:

$$\int\Big[C_1 + 2\int f(y)\,dy\Big]^{-1/2} dy = C_2 \pm \int \frac{dx}{\sqrt{g(x)}}.$$

16. $y''_{xx} = -ay'_x + e^{ax} f(ye^{ax}).$

The transformation $\xi = e^{ax}$, $w = ye^{ax}$ leads to the equation $w''_{\xi\xi} = a^{-2} f(w)$, which is of the form of S4.3.1.1.

17. $xy''_{xx} = f(x^n e^{ay})y'_x.$

The transformation $z = x^n e^{ay}$, $w = xy'_x$ leads to the following first-order separable equation: $z(aw + n)w'_z = [f(z) + 1]w$.

18. $x^2 y''_{xx} + xy'_x = f(x^n e^{ay}).$

The transformation $z = x^n e^{ay}$, $w = xy'_x$ leads to the following first-order separable equation: $z(aw + n)w'_z = f(z)$.

19. $yy''_{xx} + (y'_x)^2 + f(x)yy'_x + g(x) = 0.$

The substitution $u = y^2$ leads to a linear equation, $u''_{xx} + f(x)u'_x + 2g(x) = 0$, which can be reduced by the change of variable $w(x) = u'_x$ to a first-order linear equation.

20. $yy''_{xx} - (y'_x)^2 + f(x)yy'_x + g(x)y^2 = 0.$

The substitution $u = y'_x/y$ leads to a first-order linear equation: $u'_x + f(x)u + g(x) = 0$.

21. $yy''_{xx} - n(y'_x)^2 + f(x)y^2 + ay^{4n-2} = 0.$

1°. For $n = 1$, this is an equation of the form S4.3.2.22.

2°. For $n \neq 1$, the substitution $w = y^{1-n}$ leads to Ermakov's equation S4.3.1.5: $w''_{xx} + (1 - n)f(x)w + a(1 - n)w^{-3} = 0$.

22. $yy''_{xx} - n(y'_x)^2 + f(x)y^2 + g(x)y^{n+1} = 0.$

The substitution $w = y^{1-n}$ leads to a nonhomogeneous linear equation: $w''_{xx} + (1-n)f(x)w + (1 - n)g(x) = 0$.

23. $yy''_{xx} + a(y'_x)^2 + f(x)yy'_x + g(x)y^2 = 0.$

The substitution $w = y^{a+1}$ leads to a linear equation: $w''_{xx} + f(x)w'_x + (a + 1)g(x)w = 0$.

24. $yy''_{xx} = f(x)(y'_x)^2.$

The substitution $w(x) = xy'_x/y$ leads to a Bernoulli equation S4.1.4: $xw'_x = w + [f(x) - 1]w^2$.

25. $y''_{xx} - a(y'_x)^2 + f(x)e^{ay} + g(x) = 0.$

The substitution $w = e^{-ay}$ leads to a nonhomogeneous linear equation: $w''_{xx} - ag(x)w = af(x)$.

26. $y''_{xx} - a(y'_x)^2 + be^{4ay} + f(x) = 0.$

The substitution $w = e^{-ay}$ leads to Ermakov's equation S4.3.1.5: $w''_{xx} - af(x)w = abw^{-3}$.

27. $y''_{xx} + a(y'_x)^2 - \frac{1}{2}y'_x = e^x f(y).$

The substitution $w(y) = e^{-x}(y'_x)^2$ leads to a first-order linear equation: $w'_y + 2aw = 2f(y)$.

28. $y''_{xx} + \alpha(y'_x)^2 = \left[e^{\beta x} f(y) + \beta\right] y'_x.$

Solution:

$$\int \frac{e^{\alpha y}\,dy}{F(y) + C_1} = C_2 + \frac{1}{\beta}e^{\beta x}, \qquad \text{where} \quad F(y) = \int e^{\alpha y} f(y)\,dy.$$

29. $y''_{xx} + f(y)(y'_x)^2 + g(y) = 0.$

The substitution $w(y) = (y'_x)^2$ leads to a first-order linear equation: $w'_y + 2f(y)w + 2g(y) = 0$.

30. $y''_{xx} + f(y)(y'_x)^2 - \frac{1}{2}y'_x = e^x g(y).$

The substitution $w(y) = e^{-x}(y'_x)^2$ leads to a first-order linear equation: $w'_y + 2f(y)w = 2g(y)$.

31. $y''_{xx} = xf(y)(y'_x)^3.$

Taking y to be the independent variable, we obtain a linear equation with respect to $x = x(y)$: $x''_{yy} = -f(y)x$.

32. $y''_{xx} = f(y)(y'_x)^2 + g(x)y'_x.$

Dividing by y'_x, we obtain an exact differential equation. Its solution follows from the equation:

$$\ln|y'_x| = \int f(y)\,dy + \int g(x)\,dx + C.$$

Solving the latter for y'_x, we arrive at a separable equation. In addition, $y = C_1$ is a singular solution, with C_1 being an arbitrary constant.

33. $y''_{xx} = f(x)g(xy'_x - y).$

The substitution $w = xy'_x - y$ leads to a first-order separable equation: $w'_x = xf(x)g(w)$.

34. $y''_{xx} = \dfrac{y}{x^2} f\Big(\dfrac{xy'_x}{y}\Big).$

The substitution $w(x) = xy'_x/y$ leads to a first-order separable equation: $xw'_x = f(w) + w - w^2$.

35. $gy''_{xx} + \frac{1}{2}g'_x y'_x = f(y)h\big(y'_x\sqrt{g}\,\big), \qquad g = g(x).$

The substitution $w(y) = y'_x\sqrt{g}$ leads to a first-order separable equation: $ww'_y = f(y)h(w)$.

36. $y''_{xx} = f\big(y'^2_x + ay\big).$

The substitution $w(y) = (y'_x)^2 + ay$ leads to a first-order separable equation: $w'_y = 2f(w) + a$.

Bibliography for Chapter S4

Abramowitz, M. and Stegun, I. A. (Editors), *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, National Bureau of Standards Applied Mathematics, Washington, 1964.

Bateman, H. and Erdélyi, A., *Higher Transcendental Functions*, Vol. 3, McGraw-Hill, New York, 1955.

Kamke, E., *Differentialgleichungen: Lösungsmethoden und Lösungen, I, Gewöhnliche Differentialgleichungen*, B. G. Teubner, Leipzig, 1977.

Murphy, G. M., *Ordinary Differential Equations and Their Solutions*, D. Van Nostrand, New York, 1960.

Polyanin, A. D. and Zaitsev, V. F., *Handbook of Exact Solutions for Ordinary Differential Equations, 2nd Edition*, Chapman & Hall/CRC Press, Boca Raton, Florida, 2003.

Zaitsev, V. F. and Polyanin, A. D., *Discrete-Group Methods for Integrating Equations of Nonlinear Mechanics*, CRC Press, Boca Raton, Florida, 1994.

Chapter S5

Some Useful Electronic Mathematical Resources

arXiv.org (http://arxiv.org). A service of automated e-print archives of articles in the fields of mathematics, nonlinear science, computer science, and physics.

Catalog of Mathematics Resources on the WWW and the Internet (http://mthwww.uwc.edu/wwwmahes/files/math01.htm).

CFD Codes List (http://www.fges.demon.co.uk/cfd/CFD_codes_p.html). Free software.

CFD Resources Online (http://www.cfd-online.com/Links). Software, modeling and numerics, etc.

Computer Handbook of ODEs (http://www.scg.uwaterloo.ca/ ecterrab/handbook_odes.html). An online computer handbook of methods for solving ordinary differential equations.

Deal.II (http://www.dealii.org). Finite element differential equations analysis library.

Dictionary of Algorithms and Data Structures—NIST (http://www.nist.gov/dads/). The dictionary of algorithms, algorithmic techniques, data structures, archetypical problems, and related definitions.

DOE ACTS Collection (http://acts.nersc.gov). The Advanced CompuTational Software (ACTS) Collection is a set of software tools for computation sciences.

EEVL: Internet Guide to Engineering, Mathematics and Computing (http://www.eevl.ac.uk). Cross-search 20 databases in engineering, mathematics, and computing.

EqWorld: World of Mathematical Equations (http://eqworld.ipmnet.ru). Extensive information on algebraic, ordinary differential, partial differential, integral, functional, and other mathematical equations.

FOLDOC—Computing Dictionary (http://foldoc.doc.ic.ac.uk/foldoc/index.html). The free online dictionary of computing is a searchable dictionary of terms from computing and related fields.

Free Software (http://www.wseas.com/software). Download free software packages for scientific-engineering purposes.

FSF/UNESCO Free Software Directory (http://directory.fsf.org).

GAMS: Guide to Available Mathematical Software (http://gams.nist.gov). A cross-index and virtual repository of mathematical and statistical software components of use in computational science and engineering.

Google—Mathematics Websites (http://directory.google.com/Top/Science/Math/). A directory of more than 11,000 mathematics Websites ordered by type and mathematical subject.

Google—Software (http://directory.google.com/Top/Science/Math/Software). A directory of software.

Mathcom—PDEs (http://www.mathcom.com/corpdir/techinfo.mdir/scifaq/q260.html). Partial differential equations and finite element modeling.

Mathematical Atlas (http://www.math-atlas.org). A collection of short articles designed to provide an introduction to the areas of modern mathematics.

Mathematical Constants and Numbers (http://numbers.computation.free.fr/Constants/constants.html). Mathematical, historical, and algorithmic aspects of some classical mathematical constants; easy and fast programs are also included.

Mathematical WWW Virtual Library (http://www.math.fsu.edu/Virtual/index.php). A directory of mathematics-related Websites ordered by type and mathematical subject.

Mathematics Archives (http://archives.math.utk.edu). Combined archive and directory of mathematics Websites, mailing lists, and teaching materials.

Mathematics Genealogy Project (http://www.genealogy.ams.org). A large biographical database of mathematicians.

Mathematics Websites (http://www.math.psu.edu/MathLists/Contents.html). A directory of mathematics Websites ordered by type and mathematical subject.

Math Forum: Internet Mathematics Library (http://mathforum.org). A directory of mathematics Websites ordered by the mathematical subject.

MathGuide: SUB Gottingen (http://www.mathguide.de). An Internet-based subject gateway to mathematics-related Websites.

MathWorld: World of Mathematics (http://www.mathworld.com). An online encyclopedia of mathematics, focusing on classical mathematics. The Web's most extensive mathematical resource.

MGNet (http://www.mgnet.org/mgnet-codes.html). Free software.

Netlib (http://www.netlib.org). A collection of mathematical software, papers, and databases.

Numerical Solutions (http://www.numericalmathematics.com/numerical_solutions.htm). A library of mathematical programs.

PlanetMath.Org (http://planetmath.org). An online mathematics encyclopedia.

Probability Web (http://www.mathcs.carleton.edu/probweb/probweb.html). A collection of probability resources on the World Wide Web; the pages are designed to be especially helpful to researchers and teachers.

Science Oxygen—Mathematics (http://www.scienceoxygen.com/math.html). Topics from various sections of mathematics.

Scilab (http://scilabsoft.inria.fr). A free scientific software package.

Software — Differential Equations (http://www.scicomp.uni-erlangen.de/SW/diffequ.html). General resources and methods for ODEs and PDEs.

S.O.S. Mathematics (http://www.sosmath.com). A free resource for math review material from algebra to differential equations.

Statistics Online Computational Resources (http://socr.stat.ucla.edu). Interactive distributions, statistical analysis, virtual probability-related experiments and demonstrations, computer games, and others.

Stat/Math Center (http://www.indiana.edu/ statmath/bysubject/numerics.html). Numerical computing resources on the Internet.

UW-L Math Calculator (http://www.compute.uwlax.edu/index.php). Calculus, differential equations, numerical methods, statistics, and others.

Wikipedia: Free Encyclopedia—List of Open Source Software Packages (http://en.wikipedia.org/wiki/List_of_open-source_software_packages).

Wikipedia: Free Encyclopedia—Mathematics (http://en.wikipedia.org/wiki/Mathematics). A collection of short articles from various sections of mathematics.

Wolfram Functions Site (http://functions.wolfram.com). More than 87,000 formulas used by mathematicians, computer scientists, physicists, and engineers; more than 10,000 graphs and animations of the functions.

Yahoo—Mathematics Websites (http://dir.yahoo.com/science/mathematics/). A directory of mathematics Websites ordered by type and mathematical subject.

Yahoo—Software (http://dir.yahoo.com/Science/Mathematics/Software). A directory of software.

Chapter S6

Physical Tables

S6.1. Symbols and Units

S6.1.1. Greek Letters

Uppercase	Lowercase	Name	Uppercase	Lowercase	Name
A	α	Alpha	N	ν	Nu
B	β	Beta	Ξ	ξ	Xi
Γ	γ	Gamma	O	o	Omicron
Δ	δ	Delta	Π	π	Pi
E	ε, ϵ	Epsilon	P	ρ, ϱ	Rho
Z	ζ	Zeta	Σ	σ, ς	Sigma
H	η	Eta	T	τ	Tau
Θ	θ, ϑ	Theta	Y, Υ	υ	Upsilon
I	ι	Iota	Φ	ϕ, φ	Phi
K	κ, $\varkappa$	Kappa	X	χ	Chi
Λ	λ	Lambda	Ψ	ψ	Psi
M	μ	Mu	Ω	ω, ϖ	Omega

S6.1.2. SI Prefixes

Factor	Prefix	Symbol	Factor	Prefix	Symbol
10^1	deca	da	10^{-1}	deci	d
10^2	hecto	h	10^{-2}	centi	c
10^3	kilo	k	10^{-3}	milli	m
10^6	mega	M	10^{-6}	micro	μ
10^9	giga	G	10^{-9}	nano	n
10^{12}	tera	T	10^{-12}	pico	p
10^{15}	peta	P	10^{-15}	femto	f
10^{18}	exa	E	10^{-18}	atto	a
10^{21}	zetta	Z	10^{-21}	zepto	z
10^{24}	yotta	Y	10^{-24}	yocto	y

S6.1.3. SI Base Units

Name	Symbol	Quantity
metre	m	length
kilogram	kg	mass
second	s	time
ampere	A	electric current
kelvin	K	thermodynamic temperature
mole	mol	amount of substance
candela	cd	luminous intensity

S6.1.4. SI Derived Units with Special Names

Name	Symbol	Quantity	Expression in other units
hertz	Hz	frequency	$1/\mathrm{s}$
radian	rad	angle	m/m (dimensionless)
steradian	sr	solid angle	$\mathrm{m}^2/\mathrm{m}^2$ (dimensionless)
newton	N	force, weight	$\mathrm{kg\,m/s^2}$
pascal	Pa	pressure, stress	$\mathrm{N/m^2 = m^{-1}\,kg\,s^{-2}}$
joule	J	energy, work, heat	$\mathrm{N\,m = W\,s = C\,V = m^2\,kg\,s^{-2}}$
watt	W	power, radiant flux	$\mathrm{J/s = V\,A = m^2\,kg\,s^{-3}}$
coulomb	C	electric charge, electric flux	$\mathrm{A\,s}$
volt	V	voltage, electromotive force	$\mathrm{W/A = J/C = m^2\,kg\,s^{-3}\,A^{-1}}$
farad	F	electric capacitance	$\mathrm{C/V = m^{-2}\,kg^{-1}\,s^4\,A^2}$
ohm	Ω	electric resistance, impedance, reactance	$\mathrm{V/A = m^2\,kg\,s^{-3}\,A^{-2}}$
siemens	S	electrical conductance	$\mathrm{1/\Omega = m^{-2}\,kg^{-1}\,s^3\,A^2}$
weber	Wb	magnetic flux	$\mathrm{J/A = m^2\,kg\,s^{-2}\,A^{-1}}$
tesla	T	magnetic field strength, magnetic flux density	$\mathrm{V\,s/m^2 = N/(A\,m) = Wb/m^2 = kg\,s^{-2}\,A^{-1}}$
henry	H	inductance	$\mathrm{V\,s/A = Wb/A = m^2\,kg\,s^{-2}\,A^{-2}}$
Celsius	°C, C	temperature	K – 273.15
lumen	lm	luminous flux	$\mathrm{lx\,m^2 = cd\,sr}$
lux	lx	illuminance	$\mathrm{lm/m^2 = m^{-2}\,cd\,sr}$
becquerel	Bq	radioactivity (decays per unit time)	$1/\mathrm{s}$
gray	Gy	absorbed dose of ionizing radiation	$\mathrm{J/kg = m^2/s^2}$
sievert	Sv	equivalent dose (of ionizing radiation)	$\mathrm{J/kg = m^2/s^2}$
katal	kat	catalytic activity	$\mathrm{mol/s}$

S6.1.5. Non-SI Units Accepted for Use with SI

Name	Symbol	Quantity	SI equivalent
minute	min	time	$1\,\text{min} = 60\,\text{s}$
hour	h	time	$1\,\text{h} = 60\,\text{min} = 3600\,\text{s}$
day	d	time	$1\,\text{d} = 24\,\text{h} = 1440\,\text{min} = 86400\,\text{s}$
degree of arc	$^\circ$	angle	$1^\circ = \frac{\pi}{180}\,\text{rad}$
minute of arc	$'$	angle	$1' = \frac{1}{60}{}^\circ = \frac{\pi}{10800}\,\text{rad}$
second of arc	$''$	angle	$1'' = \frac{1}{60}{}' = \frac{1}{3600}{}^\circ = \frac{\pi}{648000}\,\text{rad}$
hectare	ha	area	$1\,\text{ha} = 10000\,\text{m}^2$
litre	l, L	volume	$1\,\text{l} = 1\,\text{dm}^3 = 0.001\,\text{m}^3$
tonne	t	mass	$1\,\text{t} = 1000\,\text{kg}$

S6.1.6. Other Common Non-SI Units

Name	Symbol	Quantity	SI equivalent
fermi	fm	length	$1\,\text{fm} = 10^{-15}\,\text{m}$
ångström	Å	length	$1\,\text{Å} = 0.1\,\text{nm} = 10^{-10}\,\text{m}$
nautical mile	M	length	$1\,\text{M} = 1852\,\text{m}$
astronomical unit	AU	length	$1\,\text{AU} = 1.49598 \times 10^{11}\,\text{m}$
light-year	ly	length	$1\,\text{ly} = 9.46073 \times 10^{15}\,\text{m}$
parsec	pc	length	$1\,\text{pc} = 3.26156\,\text{ly} = 3.08568 \times 10^{16}\,\text{m}$
are	a	area	$1\,\text{a} = 1\,\text{dam}^2 = 100\,\text{m}^2$
barn	b	area	$1\,\text{b} = 10^{-28}\,\text{m}^2$
electronvolt	eV	energy	$1\,\text{eV} = 1.60218 \times 10^{-19}\,\text{J}$
electronvolt	eV	mass	$1\,\text{eV} = 1.78266 \times 10^{-36}\,\text{kg}$
atomic mass unit	AMU, u	mass	$1\,\text{AMU} = 1.66054 \times 10^{-27}\,\text{kg}$
dalton	Da	mass	$1\,\text{Da} = 1\,\text{AMU} = 1.66054 \times 10^{-27}\,\text{kg}$
bar	bar	pressure	$1\,\text{bar} = 10^5\,\text{Pa}$
millibar	mbar	pressure	$1\,\text{mbar} = 1\,\text{hPa} = 100\,\text{Pa}$
atmosphere	atm	pressure	$1\,\text{atm} = 1.01325 \times 10^5\,\text{Pa}$
millimetre of mercury	mmHg	pressure	$1\,\text{mmHg} = 133.322\,\text{Pa}$
decibel	dB	logarithmic ratio	

S6.1.7. Base Planck Units

The base Planck units are those of length, mass, time, electric charge, and temperature; these are derived from five fundamental physical constants: speed of light in vacuum (c), gravitational constant (G), reduced Planck constant ($\hbar$), Coulomb constant ($k_\mathrm{e} = 1/(4\pi\varepsilon_0)$), and Boltzmann constant (k_B).

Name	Expression	Value in SI units
Planck length	$l_\mathrm{P} = \sqrt{\hbar G/c^3}$	1.616×10^{-35} m
Planck mass	$m_\mathrm{P} = \sqrt{\hbar c/G}$	2.176×10^{-8} kg
Planck time	$t_\mathrm{P} = l_\mathrm{P}/c = \sqrt{\hbar G/c^5}$	5.391×10^{-44} s
Planck charge	$q_\mathrm{P} = \sqrt{4\pi\varepsilon_0 \hbar c}$	$1.875\,546 \times 10^{-18}$ C
Planck temperature	$T_\mathrm{P} = m_\mathrm{P} c^2/k_\mathrm{B} = \sqrt{\hbar c^5/(G k_\mathrm{B}^2)}$	1.417×10^{32} K

S6.2. Physical Constants

S6.2.1. Universal Constants

Quantity	Symbol	Value
Gravitational constant	G, γ	$6.674(28 \pm 67) \times 10^{-11}$ N m^2/kg^2
Speed of light in vacuum	c	299,792,458 m/s
Planck constant	h	$6.626\,069 \times 10^{-34}$ J s
Reduced Planck constant	$\hbar = h/(2\pi)$	$1.054\,572 \times 10^{-34}$ J s

S6.2.2. Astronomical Constants

Name	Symbol	Value
Earth's standard surface gravity	g, g_n, g_0	9.80665 m/s^2
Earth mass	$M_\oplus$, M_E	5.9737×10^{24} kg
Mean Earth density	$\rho_\oplus$, ρ_E	5.515×10^3 kg/m^3
Mean Earth radius	$R_\oplus$, R_E	6.371×10^6 m
Mean Earth–Moon distance		3.844×10^8 m
Mean Earth–Sun distance		1.496×10^{11} m = 1 AU
Solar mass	$M_\odot$, M_S	1.989×10^{30} kg
Mean Solar radius	$R_\odot$, R_S	6.96×10^8 m
Lunar mass	$M_☾$, M_L	7.348×10^{22} kg
Mean Lunar radius	$R_☾$, R_L	1.737×10^6 m

S6.2.3. Planetary, Lunar, and Solar Parameters

Body	Equatorial radius, km	Mass, $\times 10^{24}$ kg	Mean density, $\mathrm{g/cm^3}$	Mean distance from Sun, AU	Equatorial surface gravity, $\mathrm{m/s^2}$
Mercury	2,440 (0.383 $R_\oplus$)	0.330 (0.055 $M_\oplus$)	5.427 (0.984 $\rho_\oplus$)	0.3871	3.7 (0.38 g)
Venus	6,052 (0.950 $R_\oplus$)	4.869 (0.815 $M_\oplus$)	5.243 (0.951 $\rho_\oplus$)	0.7233	8.87 (0.91 g)
Earth	6,378 (1.001 $R_\oplus$)	5.9737 ($M_\oplus$)	5.515 ($\rho_\oplus$)	1.0000	9.766 (0.996 g)
Earth's Moon	1,737 (0.273 $R_\oplus$)	0.073 (0.012 $M_\oplus$)	3.341 (0.606 $\rho_\oplus$)	0.002570*	1.622 (0.165 g)
Mars	3,397 (0.533 $R_\oplus$)	0.642 (0.107 $M_\oplus$)	3.933 (0.714 $\rho_\oplus$)	1.5237	3.693 (0.378 g)
Jupiter	71,492 (11.221 $R_\oplus$)	$1,898$ (317.8 $M_\oplus$)	1.326 (0.241 $\rho_\oplus$)	5.2034	20.87 (2.128 g)
Saturn	60,268 (9.460 $R_\oplus$)	568.5 (95.16 $M_\oplus$)	0.70 (0.127 $\rho_\oplus$)	9.5371	10.40 (1.060 g)
Uranus	25,559 (4.012 $R_\oplus$)	86.85 (14.37 $M_\oplus$)	1.30 (0.236 $\rho_\oplus$)	19.191	8.43 (0.860 g)
Neptune	24,764 (3.887 $R_\oplus$)	102.4 (17.15 $M_\oplus$)	1.76 (0.317 $\rho_\oplus$)	30.069	10.71 (1.092 g)
Pluto**	1,151 (0.181 $R_\oplus$)	0.013 (0.002 $M_\oplus$)	2 (0.4 $\rho_\oplus$)	39.482	0.81 (0.083 g)
Sun	695,500 (109 $R_\oplus$)	332,900 $M_\oplus$	1.409 (0.255 $\rho_\oplus$)		274.0 (27.9 g)

* This is the average Moon–Earth distance. The Moon's average distance to the Sun is 1 AU, the same as the Earth's.

** In 2006, Pluto was reclassified as a dwarf planet after the discovery of similar objects.

S6.2.4. Electromagnetic Constants

Quantity	Symbol	Value
Magnetic constant	μ_0	$4\pi \times 10^{-7}\ \mathrm{N/A^2} = 1.256\,637 \times 10^{-6}\ \mathrm{H/m}$
Electric constant	$\varepsilon_0 = 1/(\mu_0 c^2)$	$8.854\,188 \times 10^{-12}\ \mathrm{F/m}$
Coulomb's law coefficient	$k_\mathrm{e} = 1/(4\pi\varepsilon_0),\ k$	$8.987\,552 \times 10^{9}\ \mathrm{m/F}$
Biot–Savart law coefficient	$\mu_0/(4\pi)$	$10^{-7}\ \mathrm{H/m}$
Elementary charge	e	$1.602\,176 \times 10^{-19}\ \mathrm{C}$
Bohr magneton	$\mu_\mathrm{B} = e\hbar/(2m_e)$	$9.274\,009 \times 10^{-24}\ \mathrm{J/T}$
Nuclear magneton	$\mu_\mathrm{N} = e\hbar/(2m_p)$	$5.050\,783 \times 10^{-27}\ \mathrm{J/T}$

S6.2.5. Atomic and Nuclear Constants

Quantity	Symbol	Value
Bohr radius	$a_0 = \hbar^2/(m_e k_\mathrm{e} e^2)$	$5.291\,772 \times 10^{-11}\ \mathrm{m}$
Classical electron radius	$r_e = e^2/(4\pi\varepsilon_0 m_e c^2)$	$2.817\,940 \times 10^{-15}\ \mathrm{m}$
Electron rest mass	m_e	$9.109\,382 \times 10^{-31}\ \mathrm{kg}$
Proton rest mass	m_p	$1.672\,622 \times 10^{-27}\ \mathrm{kg}$
Neutron rest mass	m_n	$1.674\,927 \times 10^{-27}\ \mathrm{kg}$

Quantity	Symbol	Value
Electron rest energy	$m_e c^2$	510.999 keV
Proton rest energy	$m_p c^2$	938.272 MeV
Neutron rest energy	$m_n c^2$	939.566 MeV
Proton–electron mass ratio	m_p/m_e	1836.153
Specific charge of electron	$-e/m_e$	$-1.758\,820 \times 10^{11}$ C/kg
Specific charge of proton	e/m_p	$9.578\,834 \times 10^{7}$ C/kg
Magnetic moment of electron	μ_e	$-1.001\,160\,\mu_\mathrm{B}$
Magnetic moment of proton	μ_p	$2.792\,847\,\mu_\mathrm{N}$
Fine-structure constant	$\alpha = e^2/(4\pi\varepsilon_0\hbar c)$	$7.297\,353 \times 10^{-3}$
Rydberg constant (reciprocal wavelength)	$R_\infty = m_e e^4/(8\varepsilon_0^2 h^3 c)$	$1.097\,373 \times 10^{7}$ m^{-1}
Rydberg constant (electron energy)	$R_E = hcR_\infty$	13.605 69 eV = 1 Ry
Rydberg constant (frequency)	$R_\nu = cR_\infty$	$3.289\,842 \times 10^{15}$ s^{-1}
Rydberg constant (angular frequency)	$R_\omega = 2\pi cR_\infty$	$2.067\,069 \times 10^{16}$ s^{-1}
Compton wavelength of electron	$\lambda_e = h/(m_e c)$	$2.426\,310 \times 10^{-12}$ m
Compton wavelength of proton	$\lambda_p = h/(m_p c)$	$1.321\,410 \times 10^{-15}$ m

S6.2.6. Other Constants

Quantity	Symbol	Value
Avogadro’s number	N_A	$6.022\,14 \times 10^{23}$ mol^{-1}
Gas constant	R	8.3144 J/(mol K)
Boltzmann constant	$k = R/N_\mathrm{A}$, k_B	$1.380\,65 \times 10^{-23}$ J/K
Stefan–Boltzmann constant	$\sigma = \pi^2 k^4/(60\,\hbar^3 c^2)$	5.6704×10^{-8} W/(m^2 K^4)
Wien constant	b	$2.897\,76 \times 10^{-3}$ m K
Hubble constant	H_0	70.8 ± 4.0 (km/s)/Mpc

S6.3. Nuclear Physics

S6.3.1. Masses of Light Nuclides

Z	Nuclide	$M-A$, AMU	Z	Nuclide	$M-A$, AMU
0	n	0.00867	6	^{11}C	0.01143
1	^{1}H	0.00783	6	^{12}C	0
1	^{2}H	0.01410	6	^{13}C	0.00335
1	^{3}H	0.01605	7	^{13}N	0.00574
2	^{3}He	0.01603	7	^{14}N	0.0037
2	^{4}He	0.00260	7	^{15}N	0.00011
3	^{6}Li	0.01513	8	^{15}O	0.00307
3	^{7}Li	0.01601	8	^{16}O	–0.00509
4	^{7}Be	0.01693	8	^{17}O	–0.00087
4	^{8}Be	0.00531	9	^{19}F	–0.00160
4	^{9}Be	0.01219	10	^{20}Ne	–0.00756
4	^{10}Be	0.01354	11	^{23}Na	–0.01023
5	^{10}B	0.01294	11	^{24}Na	–0.00903
5	^{11}B	0.00930	12	^{24}Mg	–0.01496

Notation: Z charge number, M particle mass, A mass number.

S6.3.2. Leptons

Leptons are fermions, i.e., spin-$1/2$ elementary particles. They all have baryon number 0 and form three generations: electronic, comprising electrons (e^-) and electron neutrinos (ν_e), muonic, comprising muons (μ^-) and muon neutrinos (ν_μ), and tauonic, comprising tauons (τ^-) and tauon neutrinos (ν_τ). Every lepton has a corresponding antilepton. The main properties of leptons are listed below.

Lepton	Mass, MeV/c^2	Q	L_ν	L_μ	L_τ	Mean lifetime, s	Common decay modes
ν_e, $\tilde{\nu}_e$	$< 8\times 10^{-6}$	0	±1	0	0	unknown	
ν_μ, $\tilde{\nu}_\mu$	< 0.27	0	0	±1	0	unknown	
ν_τ, $\tilde{\nu}_\tau$	< 31	0	0	0	±1	unknown	
e^-, e^+	0.51100	∓1	±1	0	0	stable	stable
μ^-, μ^+	105.658	∓1	0	±1	0	2.1970×10^{-6}	$\mu^- \to e^- + \tilde{\nu}_e + \nu_\mu$ $\mu^+ \to e^+ + \nu_e + \tilde{\nu}_\mu$
τ^-, τ^+	1,777	∓1	0	0	±1	2.91×10^{-13}	$\tau^- \to e^- + \tilde{\nu}_e + \nu_\tau,$ $\mu^- + \tilde{\nu}_\mu + \nu_\tau,$ $\pi^- + \nu_\tau, \ \ldots$ $\tau^+ \to e^+ + \nu_e + \tilde{\nu}_\tau, \ \ldots$

Notation: Q electric charge (in e), $L_{\nu,\mu,\tau}$ lepton numbers. The tilde denotes an antiparticle.

S6.3.3. Quarks

Quarks combine to form composite particles—hadrons (baryons and mesons). All quarks are fermions and have spin $1/2$ and baryon number $1/3$; they have six flavors and form three generations: up and down (first), charm and strange (second), and top and bottom (third). Every quark flavor has a corresponding antiquark.

Quark	Q	I	I_z	C	S	T	B	Mass, GeV/c^2
u (up)	$+2/3$	$1/2$	$+1/2$	0	0	0	0	0.0015–0.0033
d (down)	$-1/3$	$1/2$	$-1/2$	0	0	0	0	0.0035–0.0060
c (charm)	$+2/3$	0	0	1	0	0	0	1.160–1.340
s (strange)	$-1/3$	0	0	0	–1	0	0	0.0899–0.0949
t (top)	$+2/3$	0	0	0	0	1	0	169.1–173.3
b (bottom)	$-1/3$	0	0	0	0	0	–1	4.130–4.370

Notation: Q electric charge (in e), I isospin, I_z isospin projection, C charm, S strangeness, T topness, B bottomness.

S6.3.4. Elementary Bosons

Elementary bosons are bosonic (integer-spin) elementary particles that act as carriers of the fundamental forces: electromagnetic (photons), weak (Z and W bosons), strong (gluons), and gravitational (gravitons, currently hypothetical). Also the existence of the Higgs boson is theoretically predicted, which is thought to be the mediator of mass.

Particle	Q, e	Mass, GeV/c^2	S	Width, GeV
Photon, γ	0	0	1	stable
Z^0	0	91.19	1	2.5
W^+, W^-	±1	80.4	1	2.1
Gluon, g	0	0	1	
Graviton, g, G	0	0	2	stable
Higgs boson, H^0	0	115–150	0	

Notation: Q electric charge, S spin quantum number.

S6.3.5. Some Baryons

Baryons are the family of composite particles made of three quarks. Baryons have half-integer spins and, hence, are fermions. They have lepton number 0 and baryon number 1. Each baryon has a corresponding antibaryon, where each quark is replaced by the corresponding antiquark.

Particle	Quark content	Mass, MeV/c^2	Q	J^P	S	C	B	I	I_z	Mean lifetime
p	uud	938.272	+1	$1/2^+$	0	0	0	$1/2$	$1/2$	$> 2.1 \times 10^{29}$ yr
n	udd	939.566	0	$1/2^+$	0	0	0	$1/2$	$-1/2$	886 s
Λ^0	uds	1115.7	0	$1/2^+$	–1	0	0	0	0	2.6×10^{-10} s
Σ^+	uus	1189.4	+1	$1/2^+$	–1	0	0	1	1	8.0×10^{-11} s
Σ^0	uds	1192.6	0	$1/2^+$	–1	0	0	1	0	7×10^{-20} s
Σ^-	dds	1197.4	–1	$1/2^+$	–1	0	0	1	–1	1.5×10^{-10} s
Ξ^0	uss	1315	0	$1/2^+$	–2	0	0	$1/2$	$1/2$	2.9×10^{-10} s
Ξ^-	dss	1322	–1	$1/2^+$	–2	0	0	$1/2$	$-1/2$	1.6×10^{-10} s
Δ^-	ddd	1232	–1	$3/2^+$	0	0	0	$3/2$	$-3/2$	5.6×10^{-24} s
Δ^0	udd	1232	0	$3/2^+$	0	0	0	$3/2$	$-1/2$	5.6×10^{-24} s
Δ^+	uud	1232	+1	$3/2^+$	0	0	0	$3/2$	$1/2$	5.6×10^{-24} s
Δ^{++}	uuu	1232	+2	$3/2^+$	0	0	0	$3/2$	$3/2$	5.6×10^{-24} s
Ω^-	sss	1672	–1	$3/2^+$	–3	0	0	0	0	8×10^{-11} s
Λ_c^+	udc	2286	+1	$1/2^+$	0	1	0	0	0	2.0×10^{-13} s
Σ_c^0	ddc	2454	0	$1/2^+$	0	1	0	1	–1	3×10^{-13} s
Ξ_c^0	dsc	2471	0	$1/2^+$	–1	1	0	$1/2$	$-1/2$	10^{-13} s
Ω_c^0	ssc	2697	0	$1/2^+$	–2	1	0	0	0	7×10^{-14} s
Λ_b^0	udb	5620	0	$1/2^+$	0	0	–1	0	0	1.4×10^{-12} s

Notation: Q electric charge (e), J total angular momentum number, S strangeness, C charm, B bottomness, I isospin, I_z isospin projection.

S6.3.6. Some Mesons

Mesons are the family of composite particles made of a quark and an antiquark. Mesons have integer spins and, hence, are bosons. They have lepton number 0 and baryon number 0. Each meson has a corresponding antimeson, where each quark is replaced by the corresponding antiquark.

Particle	Quark content	Mass, MeV/c^2	Q	J^P	S	C	B	I	I_z	Mean lifetime, s
π^+, π^-	$u\widetilde{d}$, $d\widetilde{u}$	139.57	±1	0^-	0	0	0	1	±1	2.60×10^{-8}
π^0	$\frac{1}{\sqrt{2}}(u\widetilde{u} - d\widetilde{d})$	134.98	0	0^-	0	0	0	1	0	8.4×10^{-17}
η	$\frac{1}{\sqrt{6}}(u\widetilde{u} + d\widetilde{d} - 2s\widetilde{s})$	547.8	0	0^-	0	0	0	0	0	5.0×10^{-19}
η'	$\frac{1}{\sqrt{3}}(u\widetilde{u} + d\widetilde{d} + s\widetilde{s})$	957.7	0	0^-	0	0	0	0	0	3.2×10^{-21}

Particle	Quark content	Mass, MeV/c^2	Q	J^P	S	C	B	I	I_z	Mean lifetime, s
K^+, K^-	$u\widetilde{s}$, $\widetilde{u}s$	493.7	±1	0^-	±1	0	0	$1/2$	$\pm 1/2$	1.24×10^{-8}
K^0, $\widetilde{K}^0$	$d\widetilde{s}$, $\widetilde{d}s$	497.7	0	0^-	±1	0	0	$1/2$	$\pm 1/2$	8.9×10^{-11}, 5.2×10^{-8}
ρ^+, ρ^-	$u\widetilde{d}$, $d\widetilde{u}$	775.4	±1	1^-	0	0	0	1	±1	$\sim 4.5\times10^{-24}$
ρ^0	$\frac{1}{\sqrt{2}}(u\widetilde{u}-d\widetilde{d})$	775.5	0	1^-	0	0	0	1	0	$\sim 4.5\times10^{-24}$
ω	$\frac{1}{\sqrt{2}}(u\widetilde{u}+d\widetilde{d})$	782.6	0	1^-	0	0	0	0	0	7.7×10^{-23}
φ	$s\widetilde{s}$	1019.4	0	1^-	0	0	0	0	0	1.5×10^{-22}
K^{*+}, K^{*-}	$u\widetilde{s}$, $\widetilde{u}s$	891.7	±1	1^-	±1	0	0	$1/2$	$\pm 1/2$	$\sim 7.3\times10^{-20}$
K^{*0}, $\widetilde{K}^{*0}$	$d\widetilde{s}$, $\widetilde{d}s$	896.0	0	1^-	±1	0	0	$1/2$	$\pm 1/2$	7.3×10^{-20}
D^+, D^-	$\widetilde{d}c$, $d\widetilde{c}$	1869.6	±1	0^-	0	±1	0	$1/2$	$\mp 1/2$	1.04×10^{-12}
D^0, $\widetilde{D}^0$	$\widetilde{u}c$, $u\widetilde{c}$	1864.8	0	0^-	0	±1	0	$1/2$	$\mp 1/2$	4.10×10^{-13}
D_s^+, D_s^-	$c\widetilde{s}$, $\widetilde{c}s$	1968.5	±1	0^-	±1	±1	0	0	0	5.0×10^{-13}
B^+, B^-	$u\widetilde{b}$, $\widetilde{u}b$	5279	±1	0^-	0	0	±1	$1/2$	$\pm 1/2$	1.64×10^{-12}
B^0, $\widetilde{B}^0$	$d\widetilde{b}$, $\widetilde{d}b$	5280	0	0^-	0	0	±1	$1/2$	$\pm 1/2$	1.53×10^{-12}
η_c	$c\widetilde{c}$	2980	0	0^-	0	0	0	0	0	2.5×10^{-23}
J/ψ	$c\widetilde{c}$	3096.9	0	1^-	0	0	0	0	0	7.1×10^{-21}
Υ	$b\widetilde{b}$	9460	0	1^-	0	0	0	0	0	1.2×10^{-20}

Notation: Q electric charge (e), J total angular momentum number, P parity, S strangeness, C charm, B bottomness, I isospin, I_z isospin projection.

Bibliography for Chapter S6

Benenson, W., Harris, John W., Stocker, H., and Lutz, H. (Eds.), *Handbook of Physics*, Springer, New York, 2002.

Forsythe, W. E., *Smithsonian Physical Tables, 9th Revised Edition*, Smithsonian Institution Press, Washington, 1954; Knovel, New York, 2003.

Griffiths, D., *Introduction to Elementary Particles, 2nd Edition*, Wiley-VCH Verlag GmbH & Co. KgaA, Weinheim, 2008.

Haynes, W. M. (Ed.), *CRC Handbook of Chemistry and Physics, 91st Edition*, CRC Press, Boca Raton, Florida, 2010.

Nordling, C. and Osterman, J.,, *Physics Handbook for Science and Engineering*, Studentlitteratur, Lund, Sweden, 2003.

Prokhorov, A. M. (Ed.), *Encyclopedic Dictionary of Physics* [in Russian], Sovetskaya Encyclopedia, Moscow, 1984 (dic.academic.ru).

Prokhorov, A. M. (Ed.), *Encyclopedia of Physics, Volumes 1–5* [in Russian], Bolshaya Russkaya Encyclopedia, Moscow, 1998.

Chapter S7

Periodic Table

The *periodic table of the chemical elements* (also known as *Mendeleev's table*) illustrates recurring ("periodic") trends in the properties of the elements. Currently (as of May 2010), the standard table contains 118 elements; see next page.

The elements are listed in order of increasing *atomic number* (the number of protons in the atomic nucleus). Each element has a *name* and a *symbol* and is listed together with its *standard atomic weight*—the relative atomic mass, or the ratio of the average mass per atom of the element to 1/12 of the mass of an atom of ^{12}C, calculated for all naturally occurring isotopes of the element. If an element has no stable isotopes, its standard atomic weight is taken to be the atomic mass number (the total number of protons and neutrons in the atomic nucleus) of the longest-lived isotope.

The elements are arranged in 18 groups (vertical columns) and 7 periods (horizontal rows). In most groups, the elements have very similar properties that exhibit a clear trend. Many groups are given unsystematic names such as the *alkali metals* (group 1, except hydrogen), *alkaline earth metals* (group 2), *transition metals* (groups 3–11), *pnictogens* (group 15), *chalcogens* (group 16), *halogens* (group 17), and *noble gases* (group 18). Some regions of the periodic table show horizontal trends in properties, which is true for the transition metals and especially for the *lanthanides* (group 3, period 6) and *actinides* (group 3, period 7).

The chemical properties of an element are primarily determined by its *electron configuration*—the distribution of electrons in atomic orbitals (allowed quantum states). The set of orbitals corresponding to the same principal quantum number, n, is called an *electron shell*. The shells corresponding to $n = 1, 2, 3, \dots$ are denoted by K, L, M, $\dots$ (K is the innermost shell). Each shell is divided into *subshells* corresponding to the azimuthal quantum number, l ($l = 0, \dots, n-1$), and the subshells are labeled s, p, d, f, g, h, etc. Each subshell can accommodate up to a maximum of $2(2l+1)$ electrons and the maximum number of electrons in the nth shell is $2n^2$. The electrons in an atom occupy orbitals starting from the lowest-energy orbital.

The number of electrons in an atom determines its electron configuration. For example, sodium Na has 11 electrons and its electron configuration is written as $1s^2\,2s^2\,2p^6\,3s$, which means that there are 2 electrons in the K-shell, 8 electrons in the L-shell with 2 in the s-subshell and 6 in the p-subshell, and 1 electron in the s-subshell of the M-shell. The electron configuration of Na can be written as [Ne] $3s$ for short, where [Ne] stands for that of Ne ($1s^2\,2s^2\,2p^6$).

The order in which orbitals are filled by the electrons is given by *Madelung's rule* (also known as the $n+l$ rule): (i) orbitals are filled in order of increasing $n+l$, (ii) if two orbitals have the same value of $n+l$, they are filled in order of increasing n. This rule is graphically illustrated in Fig. S7.1. According to the rule, the orbitals are filled in the order $1s\,2s\,2p\,3s\,3p\,4s\,3d\,4p\,5s\,4d\,5p\,6s\,4f\,5d\,6p\,7s\,5f\,6d\,7p$ for the 118 elements of the periodic table. Madelung's rule was established experimentally, it applies only to neutral atoms in their ground state, and it works very well for most of the elements. However, there are a few exceptions; for example, Cr should have four electrons in the $3d$ orbital but has five,

Periodic Table of the Chemical Elements

Atomic number — 4; Atomic weight — 9.0122; Symbol — Be; Element name — BERYLLIUM

43 (98) Tc TECHNETIUM — Atomic mass number of the most stable isotope

Physical state
(300 K, 101 kPa)
Ti solid
Br liquid
Ne gas

Natural occurrence
Mn primordial
Tc from decay
Bh synthetic

Period \ Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 1.0079 H HYDROGEN																	2 4.0026 He HELIUM
2	3 6.941 Li LITHIUM	4 9.0122 Be BERYLLIUM											5 10.811 B BORON	6 12.011 C CARBON	7 14.007 N NITROGEN	8 15.999 O OXYGEN	9 18.998 F FLUORINE	10 20.180 Ne NEON
3	11 22.990 Na SODIUM	12 24.305 Mg MAGNESIUM											13 26.982 Al ALUMINIUM	14 28.086 Si SILICON	15 30.974 P PHOSPHORUS	16 32.065 S SULPHUR	17 35.453 Cl CHLORINE	18 39.948 Ar ARGON
4	19 39.098 K POTASSIUM	20 40.078 Ca CALCIUM	21 44.956 Sc SCANDIUM	22 47.867 Ti TITANIUM	23 50.942 V VANADIUM	24 51.996 Cr CHROMIUM	25 54.938 Mn MANGANESE	26 55.845 Fe IRON	27 58.933 Co COBALT	28 58.693 Ni NICKEL	29 63.546 Cu COPPER	30 65.39 Zn ZINC	31 69.723 Ga GALLIUM	32 72.64 Ge GERMANIUM	33 74.922 As ARSENIC	34 78.96 Se SELENIUM	35 79.904 Br BROMINE	36 83.80 Kr KRYPTON
5	37 85.468 Rb RUBIDIUM	38 87.62 Sr STRONTIUM	39 88.906 Y YTTRIUM	40 91.224 Zr ZIRCONIUM	41 92.906 Nb NIOBIUM	42 95.94 Mo MOLYBDENUM	43 (98) Tc TECHNETIUM	44 101.07 Ru RUTHENIUM	45 102.91 Rh RHODIUM	46 106.42 Pd PALLADIUM	47 107.87 Ag SILVER	48 112.41 Cd CADMIUM	49 114.82 In INDIUM	50 118.71 Sn TIN	51 121.76 Sb ANTIMONY	52 127.60 Te TELLURIUM	53 126.90 I IODINE	54 131.29 Xe XENON
6	55 132.91 Cs CAESIUM	56 137.33 Ba BARIUM	57-71 La-Lu *	72 178.49 Hf HAFNIUM	73 180.95 Ta TANTALUM	74 183.84 W TUNGSTEN	75 186.21 Re RHENIUM	76 190.23 Os OSMIUM	77 192.22 Ir IRIDIUM	78 195.08 Pt PLATINUM	79 196.97 Au GOLD	80 200.59 Hg MERCURY	81 204.38 Tl THALLIUM	82 207.2 Pb LEAD	83 208.98 Bi BISMUTH	84 (209) Po POLONIUM	85 (210) At ASTATINE	86 (222) Rn RADON
7	87 (223) Fr FRANCIUM	88 (226) Ra RADIUM	89-103 Ac-Lr **	104 (267) *Rf* RUTHERFORDIUM	105 (268) *Db* DUBNIUM	106 (271) *Sg* SEABORGIUM	107 (270) *Bh* BOHRIUM	108 (269) *Hs* HASSIUM	109 (278) *Mt* MEITNERIUM	110 (281) *Ds* DARMSTADTIUM	111 (281) *Rg* ROENTGENIUM	112 (285) *Cn* COPERNICIUM	113 (286) *Uut* UNUNTRIUM	114 (289) *Uuq* UNUNQUADIUM	115 (289) *Uup* UNUNPENTIUM	116 (293) *Uuh* UNUNHEXIUM	117 (294) *Uus* UNUNSEPTIUM	118 (294) *Uuo* UNUNOCTIUM

* Lantanoinds	57 138.91 La LANTHANIUM	58 140.12 Ce CERIUM	59 140.91 Pr PRASEODYMIUM	60 144.24 Nd NEODYMIUM	61 (145) Pm PROMETHIUM	62 150.36 Sm SAMARIUM	63 151.96 Eu EUROPIUM	64 157.25 Gd GADOLINIUM	65 158.93 Tb TERBIUM	66 162.50 Dy DYSPROSIUM	67 164.93 Ho HOLMIUM	68 167.26 Er ERBIUM	69 168.93 Tm THULIUM	70 173.04 Yb YTTERBIUM	71 174.97 Lu LUTETIUM
** Actinoinds	89 (227) Ac ACTINIUM	90 232.04 Th THORIUM	91 231.04 Pa PROTACTINIUM	92 238.03 U URANIUM	93 (237) Np NEPTUNIUM	94 (244) Pu PLUTONIUM	95 (243) *Am* AMERICIUM	96 (247) *Cm* CURIUM	97 (247) *Bk* BERKELIUM	98 (251) *Cf* CALIFORNIUM	99 (252) *Es* EINSTEINIUM	100 (257) *Fm* FERMIUM	101 (258) *Md* MENDELEVIUM	102 (259) *No* NOBELIUM	103 (262) *Lr* LAWRENCIUM

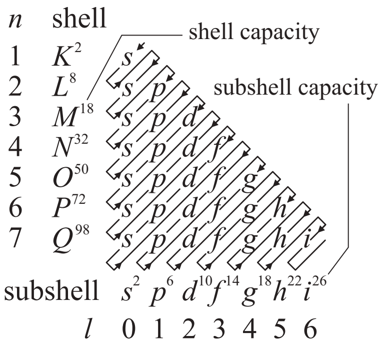

Figure S7.1. Approximate order in which atomic orbitals are filled by electrons (Madelung's rule).

with the electronic configuration [Ar] $3d^5 4s^1$ rather than [Ar] $3d^4 4s^2$ (the former turns out to have a lower energy than the latter). Similarly, Cu takes the configuration [Ar] $3d^{10} 4s^1$ instead of [Ar] $3d^9 4s^2$. There are 18 more known exceptions in periods 5, 6, and 7.

The chemical properties of an element are chiefly determined by the outmost (valence) electrons in the electronic configuration. These electrons have the highest energies and travel farthest from the nucleus and usually belong to the outmost (valence) shell, which is especially true for light atoms. Atoms with complete valence shells (all noble gases) are the most chemically inert, while those with only one electron in their valence shells (alkali metals) or just missing one electron from having a complete shell (halogens) are the most reactive. However, for heavier atoms, electrons in inner subshells can have higher energy than those in the outer subshell; for example, the $3d$ electrons have more energy than the $4s$ electrons and are, therefore, most important in chemical reactions.

The number of valence shell electrons determines the group to which the element belongs. The type of orbital in which the atom's outermost electrons reside determines the *block* to which it belongs. The s-block comprises groups 1 and 2 (plus helium), the p-block comprises the last six groups (13 through 18), the d-block contains all transition metals (groups 3 to 12). The lanthanides (elements 57–71) and actinides (elements 89–103) form the f-block. Elements within each block exhibit some similarity in their properties.

For more details about the periodic table, its applications, and the element properties, see the bibliography below.

Bibliography for Chapter S7

Atkins, P. W., *The Periodic Kingdom: A Journey in the Land of the Chemical Elements*, Basic Books, New York, 1995.

Ball, P., *The Ingredients: A Guided Tour of the Elements*, Oxford University Press, New York, 2002.

Griffiths, D. J., *Introduction to Quantum Mechanics, 2nd Edition*, Prentice Hall, Upper Saddle River, New Jersey, 2004.

Haynes, W. M. (Ed.), *CRC Handbook of Chemistry and Physics, 91st Edition*, CRC Press, Boca Raton, Florida, 2010.

Makarenya, A. A. and Trifonov, D. N., *The Periodic Law of D. I. Mendeleev* [in Russian], Prosveshchenie, Moscow, 1969.

Scerri, E., *The Periodic System, Its Story and Its Significance*, Oxford University Press, New York, 2007.

INDEX

A

Abel convergence criterion for series, 213
Abel criterion of uniform convergence for function series, 217
Abel theorem, 215, 241
Abel equation, 1014, 1015
  on convergence of power series, 218
  convergence radius of power series, 218
Abel–Ruffini theorem, 109
abscissa, 61, 79
abscissa axis, 61
absolute acceleration, 681
absolute convergence, 214
absolute elongation, 752
absolute future, 436
absolute indices of refraction, 535
absolute motion, 681, 684
absolute past, 436
absolute trajectory, 681
absolute value
  of complex number, 230
  of number (modulus of number), 7
absolute value theorem, 187
  for double integral, 196
  for triple integral, 201
absolute velocity, 681
absolutely convergent
  function series, 216
  series, 214
absolutely rough surface, 701
absorbed radiation dose, 631
absorption, 800, 826
absorption, of light, 555
absorption band, 556, 557
absorption capacity, 558
absorption coefficient, 634
absorption of gases on surface of liquid film, 826
acceleration, 406, 668
acceleration due to gravity, 426
acceptor levels, 611, 615
acoustic waves, 521
actinides, 1047
action at distance, 411
active elements, 866, 867
activation energy, 637
active forces, 687, 696, 701
active power, 904
active (effective) value
  of current, 521
  of voltage, 521
acute triangle, 37
addition
  of accelerations, 408
  of angular momenta, 582
  of angular velocities, 408, 409
  of complex numbers, 229
  of hyperbolic functions, 31
  of inverse hyperbolic functions, 34
  of inverse trigonometric functions, 28
  of matrices, 115
  of numbers, 5
  of orbital and spin angular momenta, 588
  of power series, 221
  of real numbers, 5
  of rotations about intersecting axes, 684
  of translational motions, 684
  of trigonometric functions, 23
  of velocities, 407, 409, 436
addition formulas, 24, 31
additional maxima, 550
additional pressure under curved surface, 463
additive inverse of real number, 5
adiabatic envelope, 443
adiabatic index, 445
adiabatic process, 445
adjoint matrix, 117
admissible stress, 752
admittance triangles, 911
affine coordinate system, 134
affine space, 133
aiming parameter, 421
algebra, 103
algebraic branch point, 245
algebraic equation, 109
  binomial, 108
  biquadratic, 107
  cubic, 105
  cubic, complete, 106
  cubic, incomplete, 105
  fourth-degree, 107
  linear, 105
  of degree n, 109
  quadratic, 105
  reciprocal, 107
alkali metals, 1047
alkaline earth metals, 1047
allowed energy bands, 609, 611
alpha decay, 626
alpha particle, 626
altitude
  of pyramid, 53
  of trapezoid, 45
  of triangle, 40
Ampère circuital law, 496, 891
Ampère force, 493
Ampère law, 493
amplitude, 511, 900
amplitude of damped oscillations, 517

analytic continuation, 238
analytic geometry, 61
analytical mechanics, 716
angle
  between planes, 93
  between segments, 62
  between straight line and plane, 94
  between straight lines in space, 91
  between two straight lines, 67
  between vectors, 82
  of torsion, 753–755
  shear, 747, 749, 753
angle bisector of triangle, 40
angular acceleration, 407, 408
angular dispersion, 550
angular frequency, 513, 901
angular momentum, 421–424, 431–433, 429, 580–583, 710
  intrinsic (spin), 500, 582, 585, 588, 600
  of mechanical system about center, 710
  of particle, 421
  of rigid body, 429
  of system about x-axis, 710
  of system of particles, 423
  orbital, 582, 586, 588, 624
  total, 586, 588, 591, 619
angular strain, 747
angular velocity, 407, 408
anisotropic body, 747
anisotropic crystal, 613
anisotropic media, 553
annihilating polynomial, 122
annulus, 49
anomalous dispersion, 556
antiderivative, 171
antinodes (of standing wave), 526
antisymmetric matrix, 114
apex, 57
apex of pyramid, 53
apothem of regular pyramid, 53
apparent size of area element, 537
apparent source, 536
approximate (numerical) methods for computation of definite integrals, 194
approximate analytic methods for solution of equations, 269
approximate solution of equations (root-finding algorithms for continuous functions), 160
approximation function, 288
arbitrary curvilinear coordinate systems, 1005
arbitrary motion of rigid body, 678
arbitrary series (Leibniz, Abel, and Dirichlet convergence criteria), 213
arc coordinate, 669
arccosine, 27
arccotangent, 27
Archimedes principle, 777
arcsine, 27
arctangent, 27
area
  of triangle, 41, 63
  of polygon, 63
  of quadrilaterals, 43
argument, 140
argument of complex number, 230
arithmetic, 3
arithmetic mean, 11
arithmetic operations, 5
arithmetic progression, 10
arithmetic sequence, 10
arm of couple, 689
artificial anisotropy, 555
artificial improvement of power factor, 919
associated Legendre functions, 374, 582
  differential equation, 374
associated linear space of affine space, 133
asymmetry coefficient, 390
asymptote, 145
  of graph of function, 145
  of hyperbola, 72
asymptotic analogies method, 934
asymptotic cone, 97
asymptotic coordinates, 934
asymptotic formulas
  for calculation of integrals, 185
  for diffusion flux, 857
  for integrals with weak singularity, 185
  for Laplace integrals of general form, 186
  for Laplace integrals of special form, 185
  for power Laplace integral, 186
  for Sherwood number, 856
asymptotic properties of Fourier coefficients, 224
asymptotic solution, 803, 829
atom deexcitation time, 531
atom, hydrogen, 514, 531, 569–571, 583
atom in external magnetic field, 590
atom luminescence, 541
atomic energy, 620, 637
atomic mass number, 617–621, 1047
atomic mass unit, 620, 1039
atomic structure, 570
atomic weight, 618, 1047
augmented matrix, 123
autonomous equation, 1011, 1028, 1030
average components, 783
average dipole moment, 484
average drag of particle, 817
average fluid velocity, 783, 792–796, 798
average temperature, 834
average values of sinusoidal quantities, 901
average velocity, 457
average velocity profile, 792–796, 799
averaging method (Van der Pol–Krylov–Bogolyubov scheme), 287
Avogadro constant, 441
Avogadro number, 441
axial drag, 813, 814
  dumbbell-like particle, 812, 813, 857
  oblate ellipsoid, 812
  prolate ellipsoid, 812
  thin circular disk, 811
axial principal drag, 814
axiomatic definition of probability, 382

axioms of affine space, 133
axioms of statics, 690
axis
  of buoyancy, 777
  of rotation, 408
  principal optic, 536
axis curvature, 759
axisymmetric bodies, 813
axisymmetric Poiseuille flow, 788
axisymmetric shear flow, 940
azimuthal angle, 81

B

balance conditions for system of forces, 692
balanced system of forces, 687
ball, 58
Banach problem, 385
band theory of semiconductors, 614
barns, 633
baryon, 644, 1044
baryon charge, 648
base of natural (Napierian) logarithms, 8, 17
basic convergence (divergence) criteria for series with positive terms, 212
basic criteria of uniform convergence, 216
basic differential relations, 1008, 1009
basic equation
  of motion of rigid body with fixed point, 432
  of rotational dynamics of rigid body about fixed axis, 429
  of kinetic theory of ideal gas, 453
basic hypotheses accepted in strength of materials, 747
basic law of hydrostatics, 773, 774
basic law of relativistic dynamics, 439
basic laws of mechanics, 689
basic laws of statics, 690
basic matrix of linear system of algebraic equations, 123
basic operations with matrices, 114
basic properties of inequalities, 5
basic properties of nuclei, 617
basic quadrilaterals, 44
basic theorems of statics, 690
basic theorems on residues, 246
basic units of measurement, 655
basis of linear space, 133
basis vectors, 80
Bayes formula, 383, 384
beam, 756
beam strength conditions, 762
beat, 513
becquerel, 625
bending, 755
bending moment, 746, 755
bending with torsion, 766
Bernoulli equation, 265, 1011
Bernoulli formula, 384
Bernoulli integral, 780
Bernoulli numbers, 356
Bernoulli process, 384
Bernoulli theorem, 400
Bessel equation, 272, 1020
Bessel formula, 364
Bessel function
  of first kind, 272, 362, 1020
  of second kind, 272, 362, 1020
Bessel functions (cylinder functions), 362
beta decay, 627
beta function, 361
beta-minus decay, 627
beta-plus decay, 628
bilateral constraint, 718
billet split lens, 541
bimodal distribution, 390
binormal, 669
binomial algebraic equation, 108
binomial coefficients, 9, 24, 355
binomial distribution, 384, 390
binomial theorem, 9
Biot–Savart–Laplace law (Biot–Savart law), 494
biquadratic equation, 107
birefringence, 554
bisection method, 160
black body, 558
Blasius law, 798
Blasius problem, 797
Blasius solution, 797, 798, 831
Bloch theorem, 611
Bloch waves, 611
blooming of optical systems, 553
Bochner transform, 260
Bohr–Sommerfeld quantization condition, 572
Bohr magneton, 583
Bohr model, 570
Bohr postulates, 570
Bohr radius, 571
boiling, 462
boiling temperature, 462
Boltzmann constant, 442, 450, 454
Boltzmann distribution, 457
Boltzmann entropy formula, 450
Boltzmann relation, 566
Bolzano–Cauchy theorem, 141
Bose–Einstein condensate, 599
Bose–Einstein condensation, 600
Bose–Einstein distribution, 600
Bose–Einstein statistics, 597, 599
boson, 587, 597, 1044
Bouger law, 557
bound charges, 481
bound states, 575
boundary condition, remote from body for translational flow, 782
boundary conditions, 311, 821
  at dielectric interface, 482
  at interface between different magnetic materials, 499
  on surface of solid body, 781
boundary layer, 796, 798, 827, 831
  diffusion, 827–830, 851, 853–857, 861
  laminar, 796, 805

boundary layer (*continued*)
  thermal, 831–833
  turbulent, 798, 804, 806
boundary layer thickness, 799
boundary of X-ray spectrum, 564, 589
boundary point of domain, 231
boundary value problems
  for elliptic equations, 348
  for elliptic equations with two space variables, 342
  for hyperbolic equations, 347
  for hyperbolic equations with one space variable, 336
  for parabolic equations, 345
  for parabolic equations with one space variable, 332
bounded sequence, 141
bounded set, 162
bounds for roots of algebraic equations with real coefficients, 110
Boussinesq model, 783, 792, 797
boxes (in statistics), 598
branch of many-valued function, 234
branch point, 245
  of function, 234, 235
  of infinite order, 245
branched electric circuit, 910
Brewster angle, 552
brightness, 538
Brillouin zone, 607, 611, 612, 614
bubble
  at high Peclet numbers (diffusion boundary layer), 853
  at low Peclet numbers, 853
  in translational flow, 808
Bubnov–Galerkin method, 288
buffer layer, 793, 799
bulk temperature, 837, 845, 846, 936
  for bodies of various shape, 845, 938
  for bounded body of arbitrary shape, 846
  for cylinder of finite length, 845
  for parallelepiped, 845
  for sphere, 845
Bunyakovsky inequality, 187
buoyancy force, 803, 808

C

calibration transformation, 507
  for vector potential, 498
caloric equation of state, 443
candela, 538
canonic equation of state, 452
canonical coefficients, 129
canonical equation
  of circle, 68
  of ellipse, 70
  of hyperbola, 72
  of parabola, 74
  of quadratic curves, 75
  of straight line, 89
canonical form
  for central surfaces, 96
  of elliptic equations, 310
  of hyperbolic equations, 309
  of parabolic equations, 309
  of quadratic curve, 75
  of quadric, 99
canonical representation of quadratic form, 129
capacitance, 477, 479, 865
  in sinusoidal current circuit, 906
  of capacitor, 479
capacitive resistance, 520, 906
capacitor, 479, 866
capacitor charge, 479
capacitor energy, 486
capacitor voltage, 479
capillary effects, 464
carbon cycle, 641
Cardano solution, 105
Carnot cycle, 446
Carnot theorem, 446, 743
Cartesian coordinates, 61
Cartesian coordinate system, 61, 79
Cauchy convergence criterion for series, 213
Cauchy criterion, 211
Cauchy inequality, 11, 187, 215, 240
Cauchy integral, 239
Cauchy integral formula, 239
Cauchy problem, 306, 312
  existence and uniqueness theorem, 262, 281, 302
Cauchy residue theorem, 246
Cauchy series multiplication formula, 215
Cauchy test, 192
Cauchy theorem, 153, 215
  for multiply connected domain, 239
  for simply connected domain, 239
Cauchy–Bunyakovsky inequality, 11
Cauchy–Hadamard formula, 241
Cauchy–Riemann conditions, 232
Cauchy–Riemann sum, 183
Cauchy–Schwarz–Bunyakovsky inequality, 11, 187, 215
Cauchy–Schwarz inequality, 82, 136, 388
Cauchy-type integral, 240
causality, 411, 436
Cayley–Hamilton theorem, 121
cellular model, 818
center
  of buoyancy, 777
  of curvature, 669
  of ellipse, 70
  of gravity, 39, 420, 698
  of gravity of rigid body, 698
  of hydrodynamic reaction, 814
  of hyperbola, 72
  of inertia, 413
  of mass, 413, 704
  of oscillation (of physical pendulum), 515
  of parallel forces, 698, 704
  of pencil of straight lines, 66

center (*continued*)
  of regular polygon, 45
  of sphere, 58
  of surface, 96
center-of-mass frame, 413
central angle, 45, 47
central compression, 750
central limit theorems, 401
central field, 412, 416, 421, 580
central force, 412
central moment of bivariate random variable, 397
central surface, 96
central tension, 750
centrifugal energy, 423, 582
centrifugal force, 429
centripetal acceleration, 407
centroid, 39
certain event, 380
chain reaction, 638–640
chalcogens, 1047
change in mechanical energy, 418
change of variables
  in double integral, 198
  in triple integral, 202
characteristic equation, 76, 292
  for system of ODEs, 299
  of matrix, 121
  of quadratic curves, 76
  of quadratic form, 101
characteristic equations (PDEs), 308
characteristic impedance, 913
characteristic index, 1025
characteristic matrix, 120
characteristic polynomial of matrix, 121
characteristic quadratic form, 101
characteristic system, 305
characteristic X-ray spectra, 589, 590
characteristics, 307
characteristics of nuclei, 617
characteristics of particles, 647
characteristics of radiation source, 538
charge
  baryon, 648
  electric, 469
  lepton, 648
charge carriers, 488
charge distribution, 469
charge independence, 622, 649
charge number, 617
charm (quark flavor), 644
charmonium, 645
Chebyshev inequality, 400
Chebyshev polynomials, 376
  of first kind, 376
  of second kind, 377
Chebyshev theorem, 400
chemical potential, 452, 462, 600, 603
Chi-square distribution, 395
chord, 46, 96
circle, 46, 68
circuit impedance, 520, 908
  reactive, 520
  resistive, 520
circuit quality factor, 914
circuit susceptance, 910
circuit with active resistance, 904
circuit with capacitance, 906
circuit with inductance, 905
circular cylinder in translational flow, 800, 804, 859
circular frequency, 511
circular motion, 407
circular sector, 48
circularly polarized, 528
circulation, 205
circulation of magnetic induction, 496
circumcenter, 40
circumcircle of triangle, 40
circumscribed about circle, polygon, 43
Clairaut equation, 269
Clapeyron–Mendeleev equation, 441
class of systems of units of measurement, 655
classes of subatomic particles, 643
classes of integrable functions, 183, 196
classical definition of probability, 381
classical free-electron theory of metals, 491
classical model of dispersion medium, 555
classical solution, 315
classification
  of crystals, 609
  of electric circuits and their operation modes, 868
  of molecular levels, 592
  of mechanical constraints, 718
  of quadrics, 99
  of second-order linear partial differential equations, 307
  of terms of complex atoms, 588
Clausius–Clapeyron relation, 462
Clausius–Mossotti relation, 485, 557
Clausius inequality, 448
closed domain, 231
closed path integral, 205
closed system of particles, 413
coefficient
  damping, 517, 520
  damping, wave, 522
  Einstein, 565
  of friction, 412, 797–800
  of kinetic friction, 700
  of longitudinal bending, 771
  of mutual inductance, 505
  of reflection, 529, 553, 577
  of rolling friction, 701
  of self-inductance, 504
  of static friction, 412, 700
  of variation, 390
  of viscosity, 465
  of wave damping, 522
  quantum amplification, 567
  wave attenuation, 557
coercive field, 501

coercive force, 501
cofactor expansion of determinant
  along column, 113
  along row, 113
cofactor of element, 113
coherence
  spatial, 542
  temporal, 542
coherence length, 542
coherence of light waves, 539
coherence time, 542
coherent quanta, 567
coherent sources, 525
coherent waves, 525
coil, 866
cold reservoir, 444
collinear vectors, 78
collision frequency, 464
collisions of molecules in gas, 464
collocation method, 289
color (quantum number), 645
column vector, 114
combined method for verifying empirical formulas, 932
combined stress, 763
commensurability of physical quantities, 578
common difference, 10
common logarithm, 8
common ratio, 10
commuting matrices, 116
comparison of infinitely small quantities, 146
compatibility conditions, 324
compensation theorems, 875, 876
compensator, 554
complementary error function, 356, 369, 827, 854, 937
complementary event, 380
complementary modulus, 372
complete cubic equation, 106
complete elliptic integral
  of first kind, 370
  of second kind, 370
complete group of events, 380
complete group of incompatible events, 380
complex admittance of circuit, 910
complex amplitude, 512
complex conjugate matrix, 117
complex effective value, 902, 903
complex exponential diagram, 512
complex linear space, 132
complex number, 229
complex plane, 229
complex total power, 909
complex values of currents, 910
complex vector diagram, 512
complicated (anomalous) Zeeman effect, 590
composite function, 140, 164, 222
composite number, 3
composition of harmonic oscillations, 512
composition of mutually perpendicular harmonic oscillations, 513
compound capacitor, 480
compound motion, 681
  of body, 684
  of point, 681
compound percentage, 7
Compton effect, 564
Compton wavelength, 565
concave functions, 148
concise methods for analyzing electric circuits, 881
condition
  Bohr–Sommerfeld quantization, 572
  for planes to be parallel, 93
  for planes to be perpendicular, 94
  for planes to coincide, 94
  for straight line and plane to be perpendicular, 95
  for three points to be collinear, 66
  for two straight lines in space to meet, 95
  Wulf–Bragg, 551
conditional convergence, 214
conditional extrema of functions of several variables, 169
conditional extremum of function of two variables, 168
conditional maximum, 168
conditional minimum, 168
conditional probability, 382, 383
conditionally convergent series, 214
conditions
  Laue, 550
  for phase equilibrium, 461
  for straight line and plane to be parallel, 94
  for two straight lines to be parallel, 92
  for two straight lines to be perpendicular, 92
  for two planes to be parallel, 93
  for two planes to be perpendicular, 94
  for vectors to be parallel, 83
  for vectors to be perpendicular, 83
  of compatibility of initial and boundary conditions, 324, 326
  of expansion in Taylor series, 220
  of extremum of function of two variables, 167
  of fusion, 641
conducting ball in uniform field, 478
conduction current, 488
conductivity, 488
  intrinsic, 610
conductivity band, 609, 610
conductor, 477, 866
  in external electric field, 478
cone, 57, 97
confinement, 642, 645
congruent polyhedral angles, 51
conical surface, 57
conjugate complex number, 229
conjugate tensor, 813
connection
  of conductors, 477
  of resistors, 491
  of emf sources, 491
consequences of Lorentz transformations, 435

conservation condition for physical quantity, 579
conservation law of angular momentum, 423, 424, 649
conservation of angular momentum for system of particles, 423
conservative force, 416
conservative system, 416
consistency condition for linear system, 123
consistent mode, 873
consistent system of linear algebraic equations, 123
constant
  Avogadro, 441, 1042
  Boltzmann, 442, 450, 454, 1042
  decay, 625
  effective spring, 514
  electric, 470, 1041
  Euler, 360, 363, 369
  fine structure, 586, 1042
  gas, 442
  Kerr, 555
  of separation, 321
  Planck, 561, 1040
  Planck, reduced, 561, 1040
  Rydberg, 569, 571, 583, 1042
  Verdet, 555
  von Kármán, 784, 793, 833
constant magnetic field, 492
constants, physical, 1040
constrained maximum, 168
constrained mechanical system, 687
constrained minimum, 168
constraint, 168
constraint equations, 687, 718
constraint reaction forces, 412, 687, 702
construction of graphs, 158
construction of Green functions, 348
contact angle, 464
contact phenomena (for semiconductors), 615
continuity equation, 779, 801
continuous bivariate random variable, 396
continuous functions, 147, 163
continuous random variable, 386
continuum approximation, 605
contraction ratio, 234
controlled fusion, 642
convection current, 488
convective diffusion, 821, 853
convective diffusion equation, 823, 860
convective mass and heat transfer
  boundary conditions, 822
  equations, 821–823
  in liquid film, 829
  in tubes and channels, 834
convergence
  in distribution, 399
  in mean, 399
  in mean square, 399
  in probability, 398
  of function series, pointwise, 215
  of function series, uniform, 216
  of functions, nonuniform, 149
  of functions, pointwise, 149
  of functions, uniform, 149
  of random variables, 398
convergence criteria for arbitrary numerical series (absolute and conditional convergence), 213
convergence criteria for series with positive (nonnegative) terms, 212
convergence domain of series, 216
convergence tests for improper integrals of unbounded functions, 194
convergent sequence, 141
convergent numerical series, 211
converging system of forces, 690, 692, 716
conversion formulas, 372
convex functions, 148
convex polygon, 43
convex polyhedron, 51
convolution theorem, 252, 256
  for Fourier transform, 256
  for Laplace transform, 252
coordinate axes
  in space, 79
  on plane, 61
coordinate method for specifying motion of point, 668
coordinate of point, 61
coordinate of system, 79
coordinate transformations, 135
coordinates
  asymptotic, 934
  in space, 78
  of element, 132, 133
  of center of mass of body, 203
  of center of mass of flat plate, 200
coplanar vectors, 78
core of cross-section, 765
core of turbulent flow, 793
Coriolis acceleration, 410, 682
Coriolis force, 429
corner of polygon, 43
corpuscular-wave duality, 563, 573
correction coefficient for drag force, 817, 818
correlation of two random variables, 397
correlated random variables, 398
cosine, 21
cosine integral, 358
cotangent, 22
Cotton–Mouton effect, 555
Couette flow, 782
Coulomb law, 470
  dry friction, 700
countable set, 3
couple of forces, 688–691
couple of rotations, 685
coupled pendulums, 516
coupling
  jj, 588, 624
  LS, 588, 624
  spin-orbit, 585
  Russell–Saunders, 588, 624
covalent (homeopolar) chemical bond, 593

covariance, 397
covariance of two random variables, 397
Cramer rule, 124
creeping flow, 662, 780, 800, 824
criteria
  for applicability of empirical formulas, 926
  of mean-square convergence of Fourier series, 226
  of positive and negative definiteness of quadratic form, 128
  of uniform convergence of Fourier series, 226
  of uniform convergence of series, 216
critical force, 768–770
critical mass, 639
critical molar volume, 459
critical opalescence, 557
critical point, 459
critical pressure, 459
critical stresses, 768
critical temperature, 459
cross-section of nuclear reaction, 632
cross-section rigidity in bending, 759
cross-section rigidity in torsion, 755
cross product of two vectors, 83
cube, 52
cubic equations, 105
cumulative distribution function, 386
curie (unit), 625
Curie temperature, 485, 502
curl of vector field, 206
current, 488, 865
  conduction, 488
  convection, 488
  displacement, 506
  in closed circuit, 490
current compensation principle, 876
current compensation theorem, 876
current density, 488
current resonance, 915
current strength, 865
curvature, 657, 669
curve, visibility, 538
curvilinear integral, 204
curvilinear coordinate systems, 1006
cylinder, 56
cylinder function, 362
cylindrical capacitor, 480
cylindrical coordinates, 81, 1008
cylindrical diverging waves, 522
cylindrical function, 1020
cylindrical surface, 55
cylindrical tube, 57

D

D'Alembert equation, 507
D'Alembert–Euler conditions, 232
D'Alembert convergence criterion for series, 212
D'Alembert formula, 312
D'Alembert principle, 716
Dalton law, 455
damped electric oscillations, 518
damped oscillations, 517
damping coefficient, 517
damping decrement, 517
  logarithmic, 518
Darcy formula, 794
daughter nuclei, 625
Davison–Germer experiments, 573
de Broglie formulas, 573
de Broglie waves, 573
de Moivre formula, 231
de Morgan laws, 381
Debye frequency, 605
Debye temperature, 605
Debye theory, 605
decadic logarithm, 8
decay constant, 625
decelerated motion, 668
decreasing function, 140
decreasing sequences, 142
defect of matrix, 118
definite integral, 183
deformation, 745
degenerate Fermi gas, 601
degenerate hypergeometric equation, 1019
degenerate hypergeometric function (Kummer function), 367, 369, 828, 836, 840
degenerate quadratic curves, 75
degenerate quadratic form, 128
degenerate quadric, 99
degree of freedom, 405
delayed neutrons, 638, 640
delayed potentials, 508
density, 413
  energy flux, wave, 529
density of states, 560, 565, 601
  Brillouin zone, 607
  energy, 601
  of electron gas, 602
density, probability, 573
dependence of binding energy on nuclear composition, 620
dependent variable, 140
derived units of measurement, 655
derivative, 150
  of function at point, 232
  of higher orders, 154
  of polynomial, 105
Descartes–Euler solution, 107
Descartes theorem, 111
determinant of matrix, 119
determinant of order 2, 112
determinant of order 3, 112
determined system of linear algebraic equations, 123
deuterium, 618
deuteron, 618
developed turbulence in boundary layer, 806
d-fold degenerate matrix, 118
diagonal matrix, 114
diamagnetic materials (diamagnets), 498, 501
diameter of circle, 46

diameter of set, 195
diameter of sphere, 58
diatomic chain, 608
dichroic plate, 554
dichroism phenomenon, 554
dielectric, 481, 610
difference
  between local and average fields, 484
  of complex numbers, 229
  of events, 380
  of matrices, 115
  of polynomials, 103
  of vectors, 78, 132
differentiable function, 150, 163
differential, 150
differential calculus for functions of single variable, 150
differential constraints, 718
differential cross-section, 570, 633
differential equation of rotational motion of rigid body about fixed axis, 711
differential equations
  of equilibrium of rod, 756
  of equilibrium of rod under bending, 756
  of motion of point in projections onto natural coordinate axes, 701
  of motion of point particle in Cartesian coordinates, 701
differential operators of field theory, 170
differential resistance of nonlinear element, 884
differential scattering cross-section, 570
differentiation formulas, 25, 28, 32, 35
differentiation of functions of several variables, 163
differentiation rules, 152
diffraction, 545
  Fraunhofer, at circular hole, 548
  Fraunhofer, at slit, 547
  Fresnel, at circular hole, 545
  Fresnel, at slit, 546
  of X-rays at crystal, 550
diffraction grating, 549
diffraction grating as spectral device, 550
diffraction grating period, 549
diffusion
  Fick law, 466
diffusion boundary layer, 827, 835, 861
diffusion boundary layer approximation, 827, 829, 941
diffusion coefficient, 466
diffusion flux, 466, 828
  to film surface, 827
diffusion Peclet numbers, 824
diffusion regime of reaction, 822, 824
diffusivity, 466
digamma function, 360, 363
dihedral angle, 50
dilatation ratio, 234
dimension
  of affine space, 133
  of linear space, 133
  of matrix, 114
  of physical quantity, 656
dimensional formula, 656
dimensional quantities, 656–659
dimensionless diffusion flux, 824
dimensionless form of equations, 780
dimensionless internal coordinate, 799
dimensionless mass transfer equation, 822
dimensionless quantities, 656
dimensions, 655–657
Dini integral, 344
dioptre, 537
dipolar polarization, 484
dipole, 469
  in external field, 474
dipole moment, 469
Dirac delta function, 320
direct analytic continuation, 238
direct current, 488
direct method for verifying two-parameter formulas, 926
direct problem of kinematics, 406
direction coefficients of straight line, 89
direction cosine of vector, 80
direction of non-equilibrium processes, 442
  in thermally insulated system, 449
direction of convexity of graph of function, 157
directional derivative, 166
directional pattern, 531
directional radiation pattern, 531
directrix, 55, 57
  of ellipse, 70
  of hyperbola, 72
  of parabola, 74
Dirichlet–Jordan convergence criterion for Fourier series, 224
Dirichlet conditions, 223
Dirichlet convergence criterion for series, 214
Dirichlet criterion of uniform convergence of function series, 217
Dirichlet problem, 314
Dirichlet theorem on Fourier series, 223, 224
disconnection, 866
discontinuities of first kind, 147
discontinuities of second kind, 147
discrete bivariate random variable, 396
discrete probability space, 381
discrete random variables, 386
discreteness, 575
discreteness of energy levels, 575
discriminant of algebraic equation, 110
discriminant of quadratic form, 128
disk, 46, 69
dispersion, 522, 524
dispersion equation, 317
dispersion medium, 524, 555
dispersion of light, 555
  anomalous, 556
  normal, 556
displacement, 405
displacement current, 506

displacement of current loop in magnetic field, 497
dissolution of plate by laminar liquid film, 829
dissolving solid particles, 800
distance, 405
  between elements, 136
  between parallel lines, 68
  between points, 62
  between points on plane, 62
  between straight lines, 95
  between two parallel planes, 95
  from point to plane, 95
  from point to straight line, 67, 95
distributed forces, 699, 745
distribution
  Boltzmann, 457
  Bose–Einstein, 600
  Fermi–Dirac, 600
  Maxwell, 456
  Maxwell–Boltzmann, 456, 600
distribution function of two-dimensional random vector, 395
distribution of bivariate random variable, 395
distribution of potential along unbranched (unramified) electric circuit, 871
divergence, 208
divergence of tensor, 783
divergence theorem, 208
divergent integral, 191
divergent series, 211
divisibility tests, 3
division
  of complex numbers, 229
  of polynomials, 103
  of power series, 221
  of segment in given ratio, 63
domain of function, 140
donor energy levels, 610
Doppler effect for electromagnetic waves in vacuum, 532
Doppler effect in acoustics, 527
dot product, 82, 136
double integral, 195, 197
double factorial, 9, 355
double-magic nuclei, 621
drag coefficient, 795, 804, 808, 816, 842
  for circular cylinder, 805
  for circular tube, 795
  for drop, 808
  for spherical drop, 808
  for spherical particle, 803, 804
  for tubes, 796
drag crisis, 804, 806
drag force, 412
  body of revolution with arbitrary orientation, 811
  circular cylinder, 805
  disk, 810
  spherical drop, 808
drag law for spherical bubbles, 808
driving force, 518
drop
  at high Peclet numbers (diffusion boundary layer), 853
  at low Peclet numbers, 853
  in translational liquid flow, 808
dry friction force, 412
Duffing equation, 285
Duhamel first principle, 350
Duhamel second principle, 351
Duhamel integral, 294
Dulong–Petit law, 604
dumbbell-like particle, 857
dynamic Euler equations, 734
dynamics, 410
  of absolutely rigid body, 730
  of electron in crystal, 612
  of point particle, 701
  of rigid body, 429

E

Earnshaw theorem, 476
eccentric compression, 765
eccentric extension, 765
eccentricity of ellipse, 70
eccentricity of hyperbola, 72
eddy (turbulent) viscosity, 783, 791–793, 798
edge
  of dihedral angle, 50
  of polygon, 43
  of polyhedral angle, 50
  of polyhedron, 51
  of trihedral angle, 50
effect
  capillary, 464
  Cotton–Mouton, 555
  Compton, 564
  Doppler, in acoustics, 527
  Doppler, for electromagnetic waves, 532
  external photoelectric, 562
  Hall, 492
  isotopic, 572
  Kerr, 555
  Mössbauer, 630
  Paschen–Buck, 591
  symmetry, 621
  tunnel, 577
  Zeeman, anomalous, 590
  Zeeman, normal, 583, 591
effective cross-section, 465
effective mass, 613
effective potential energy, 422, 623
effective spring constant, 514
effective value, 521
effective values of sinusoidal quantities, 900
effective viscosity, 818
effective viscosity of suspensions, 818
eigenfunctions, 278, 322
  of orbital angular momentum operator, 581
eigenvalue, 278, 322
eigenvalue of matrix, 120
eigenvalue problems, 277

eigenvalues equation of matrix, 120
eigenvalues
  of angular momentum operator, 581
  of orbital angular momentum operator, 581
  of square matrices, 120
eigenvector of matrix, 120
Einstein coefficients, 566
Einstein–de Haas experiments, 585
Einstein equation, 563
Einstein formula (for volume concentration), 818
Einstein theory (of crystal heat capacity), 604
elastic collision, 418
elastic force, 412
elastic rod, 523
elastic waves, 521, 604
elasticity, 747
electric capacitance, 479
electric charge, 469
electric circuit, 867
electric constant, 470
electric current, 488
electric dipole, 469
electric dipole moment, 469
electric displacement field, 482
electric field, 470
  of point charge, 471
  of relativistic particle, 510
  of system of point charges, 471
electric permittivity, 482
electric resistance, 865
electric susceptibility, 482
electrical conduction of metals, 603
electrical engineering, 865
electrical network, 867, 880
electrical network branch, 868
electrical network loop, 868
electrical network node, 868
electrical resistance of conductor, 489
electrically closed, 469
electrically neutral system of charges, 469
electrodynamics, 469
electromagnetic field, 470
electromagnetic induction, 502
electromagnetic interaction, 643
electromagnetic waves, 527
electromotive force, 489, 865
electron capture, 628
electron configuration, 1047
electron energy, 592
electron gas in metal, 601
electron heat capacity of metals, 603
electron shell, 587, 1047
electron subshell, 587, 1047
electron spin and fine structure of spectrum, 585
electronic mathematical resources, 1035
electronic polarization in nonpolar dielectrics, 484
electrons in periodic field, 608
electronvolt, 563
electrostatic field, 471
electrostatic image method, 478
electrostatic induction, 478
electrostatics of conductors, 477
electrostatics of dielectrics, 481
electroweak interaction, 647, 650
elementary cell of phase space, 597
elementary charge, 469
elementary event, 380
elementary functions, 15, 234
elementary geometry, 37
elementary momentum of force, 708
elementary (subatomic) particles, 642–650
elementary symmetric functions, 110
elementary theory of gyroscope, 737
elementary theory of impact, 738
elementary work, 712
elementary work of force, 712
ellipse, 69
ellipsoid, 96
  of revolution, 857
elliptic equation, 308, 311, 313
elliptic integral, 370
  of first kind, 372
  of second kind, 372
  of third kind, 372
elliptic modulus, 372
elliptic paraboloid, 98
elliptically polarized, 528
elliptically polarized oscillation, 513
Emden–Fowler equation, 1028
emf (electromotive force), 489
emission
  forced, 566
  spontaneous, 566
empirical formula for binding energy, 624
empirical formulas, 921
  criteria for applicability, 926
endothermic reaction, 634
energy
  centrifugal, 423, 582
  Fermi, 600
  Fermi, of electrons, 602
  of electromagnetic waves, 529
  of electrostatic field, 485, 486
  of ideal gas, 454
  of ionization, 570
  of isolated conductor, 486
  of loop, 505
  of magnetic field, 505
  of system of charges, 485
  of zero-point oscillations, 576
energy bands, allowed, 609
energy bands, forbidden, 609, 610, 614
energy characteristics (of radiation), 538
energy density
  of electric field, 486
  of electromagnetic field, 508
  of magnetic field, 505
  of states, 602
energy hierarchy, 592
energy method (for determining frequency), 512
energy radiance, 538
engineering formulas, 922

enthalpy, 444
entire function, 244
entire transcendental function, 244
entropy, 448, 449
  of ideal gas, 449
  of mixing, 450
  of phase transition, 462
  of van der Waals gas, 449
equalities, 5
equation
  Abel, 1014, 1015
  algebraic, 109
  autonomous (ODE), 1011, 1028, 1030
  Bernoulli, 265, 1011
  Bessel, 272, 1020
  binomial, 108
  biquadratic, 107
  caloric, of state, 443
  canonic, of state, 452
  characteristic, of matrix, 121
  characteristic, of quadratic curves, 76
  characteristic, of quadratic form, 101
  Clapeyron–Mendeleev, 441
  Clairaut, 269
  continuity, 779, 801
  convective diffusion, 822, 823, 860
  convective heat transfer, 823
  convective mass transfer, 822
  cubic, 105
  cubic, complete, 106
  cubic, incomplete, 105
  d'Alembert, 507
  degenerate hypergeometric, 1019
  dimensionless mass transfer, 822
  dispersion, 317
  Duffing, 285
  Einstein, for photoeffect, 563
  elliptic (PDE), 308, 311, 313
  Emden–Fowler, 1028
  Ermakov, 1029
  Euler, 272, 297, 780, 1020
  exact (ODE), 265, 267
  first-order quasilinear (PDE), 307
  for average concentration, 825
  for stream function, 801
  fourth-degree, 107
  Gaussian hypergeometric, 1022
  general Riccati, 267, 1012, 1013
  generalized homogeneous, 264, 282, 1012, 1029
  heat, 307
  heat transfer, 823
  heat, with source, 316
  homogeneous linear (ODE), 264, 282, 312, 313, 1011, 1029
  homogeneous linear, of second order (ODE), 1016
  Hopf, 307
  hyperbolic (PDE), 308, 309, 311, 312, 352
  hyperbolic, first canonical form, 309
  hyperbolic, second canonical form, 309
  hypergeometric, 1022
  hypergeometric, degenerate, 1019
  Klein–Gordon, 339
  Lagrange, 269
  Laplace, 308, 847
  Legendre, 1021–1023
  Lienard, 1030
  linear algebraic, 105
  linear, of arbitrary order (ODE), 292
  linear, with constant coefficients (ODE), 292
  linear, with variable coefficients (ODE), 296
  mass transfer, 825
  mass transfer, dimensionless, 822
  Mathieu, 1025
  Meshcherskii, 415
  modified Bessel, 273, 1020
  modified Mathieu, 1024
  of circle in Cartesian coordinate system, 68
  of curve, 64
  of curve in plane, 64
  of damped vibrations, 1017
  of damped oscillations, 272, 517
  of ellipse, 70
  of ellipse in polar coordinates, 71
  of free oscillations, 1017
  of forced oscillations, 518
  of harmonic oscillations, 511
  of hyperbola, 72
  of hyperbola in polar coordinates, 73
  of kinetic theory of ideal gas, 453
  of motion, 411, 511
  of motion of electron in crystal, 612
  of motion of rigid body with fixed point, 432
  of normal, 151
  of parabola, 74
  of parabola in polar coordinates and parametric equations, 75
  of pencil of planes, 89
  of pencil of straight lines, 66
  of plane, 84
  of plane harmonic wave, 522
  of plane passing through point and parallel to another plane, 86
  of plane passing through point and parallel to two straight lines, 87
  of plane passing through point and perpendicular to two planes, 87
  of plane passing through three points, 86
  of plane passing through two points and parallel to straight line, 86
  of plane passing through two points and perpendicular to given plane, 87
  of planes passing through line of intersection of planes, 89
  of projections of forces, 692
  of rotational motion of rigid body about fixed axis, 429, 711
  of quadratic curve, 75
  of simple harmonic oscillations, 511
  of state, 441
  of state of ideal gas, 441, 662
  of straight line in space, 89, 90

equation (*continued*)
  of straight line on plane, 64
  of straight line passing through two given points, 65
  of tangent line, 151
  of tangent of ellipse, 70
  of tangent of hyperbola, 73
  of tangent of parabola, 74
  of tangent plane, 166
  parabolic (PDE), 307, 311, 351
  Poisson, for adiabatic process, 445, 735
  Poisson, three-dimensional, 320
  Poisson, two-dimensional, 320
  process, 442
  quadratic, 105
  reciprocal, algebraic, 107
  Riccati, 267, 269, 1012
  Riccati, general, 267, 1012, 1013
  Riccati, special, 1013
  Schrödinger, 574
  Schrödinger, time-dependent, 574
  Schrödinger, time-independent, 574
  Schrödinger, three-dimensional, 574, 579
  second-order constant coefficient linear, 272, 1017
  separable, 1011
  special Riccati, 1013
  telegraph, 317
  thermal boundary layer, 831
  thermal, of state, 443
  three-dimensional Poisson, 320
  two-dimensional Poisson, 320
  Tsiolkovsky, 415
  Van der Waals, 442, 458, 459
  wave, 308, 522
  Young–Laplace, 463
equations
  admitting reduction of order (ODEs), 281, 302
  arbitrary order (ODEs), 292–303
  arbitrary order, linear (ODEs), 292–302
  arbitrary order, nonlinear (ODEs), 301–303
  characteristic (PDEs), 308
  constraint, 687, 718
  convective mass and heat transfer, 822, 823
  elliptic (PDEs), 308, 310
  equilibrium, for quiescent incompressible fluid, 780
  equilibrium, for rigid body, 692
  Euler, for ideal incompressible fluid, 780
  Euler, dynamic, 734–736
  Euler, kinematic, 733
  first-order ODEs, 261–271, 1011–1016
  first-order quasilinear (PDEs), 305–307
  hydrodynamic, 778
  hyperbolic (PDEs), 308, 309
  integrable by quadrature (ODEs), 261
  invariant under scaling–translation transformations (ODEs), 283
  involving arbitrary functions (ODEs), 1027
  involving exponential and other functions (ODEs), 1024
  involving power functions (ODEs), 1017
  Lagrange, of second kind, 723
  linear (PDEs), 307–353
  Maxwell, 506
  Maxwell, for constant magnetic field, 498
  Maxwell, in electrostatics, 476
  Maxwell, in magnetostatics, 498
  Maxwell, in medium, 508
  Maxwell, in vacuum, 506
  Navier–Stokes, for laminar flow, 779, 787, 824
  nonlinear, of arbitrary order (ODEs), 301
  not solved for derivative (ODEs), 262, 268
  ODEs, 263–304, 1011–1033
  of straight line passing through point and perpendicular to plane, 91
  of straight line passing through two points, 91
  of equilibrium of rigid body, 692
  of equilibrium of rod, 756
  of moments of forces, 692
  of motion of plane figure, 676
  of motion of particle in projections onto natural coordinate axes, 701
  of motion of particle in Cartesian coordinates, 701
  of solenoidal electric field, 503
  parabolic equations (PDEs), 308, 309
  PDEs, 305–353
  Reynolds, for turbulent flow, 782, 783
  second-order (ODEs), 271–292, 1016–1033
  second-order, linear (ODEs), 271–280, 1016–1028
  second-order, linear (PDEs), 307–353
  second-order, nonlinear (ODEs), 283–293, 1028–1033
  solved for derivative (ODEs), 261, 263, 281
  solved for highest derivative (ODEs), 301
  Stokes, 800
  with separable variables (ODEs), 263
  with separated variables (ODEs), 263
equilateral hyperbola, 72
equilateral triangle, 37, 42, 43
equilibrium equations for quiescent incompressible fluid, 780
equilibrium equations for rigid body, 692
equilibrium macrostate, 598, 599
equilibrium process, 442
equilibrium state, 441
equilibrium thermal radiation, 558
equilibrium thermodynamics, 442
equipartition law, 455
equipartition theorem, 455, 561
equipotential surfaces, 475
equivalent (or hydraulic) diameter, 791, 843, 844
equivalent circuit, 867
equivalent events, 380
equivalent infinitely small functions, 146
equivalent resistances, 491
equivalent sources, 491
equivalent systems
  of equations, 124
  of forces, 687

equivalent transformations
  in electrical networks, 877
  of characteristics of nonlinear elements, 886
ergodic hypothesis, 598
Ermakov equation, 1029
error function, 318, 356, 368
  complementary, 356
escape velocity, 427
essential singularity, 243
estimation theorem
  for double integral, 196
  for triple integral, 201
Euclidean space, 136
Euler formula, 25
Euler–Fourier formulas, 224
Euler constant, 360, 363, 369
Euler equation, 272, 297, 433, 780, 1020
Euler equations for ideal incompressible fluid, 780
Euler formula, 25, 235, 360, 679, 769, 770
Euler line, 40
Euler substitutions, 177
Euler theorem, 51
Euler transform of first kind, 260
Euler transform of second kind, 260
evaporation–condensation curve, 461
evaporation, 822
evaporation into bubbles, 800
even function, 140
even number, 3
event, 380
exact differential equations, 265
exact solution, 828, 830
excess coefficient, 390
exchange energy, 594
exchange interaction, 594
existence of nontrivial solutions of homogeneous system, 127
exitance, radiant, 538
exothermic reaction, 634
expansion
  of nonperiodic functions in Fourier series, 225
  of periodic functions in Fourier series, 225
  of polynomial in powers of linear binomial, 104
expansion of ideal gas into vacuum, 450
expectation, 388
  of bivariate random variable, 397
  of continuous random variable, 388
  of discrete random variable, 388
  of function of random variable, 388
expected value of random variable, 388
experimental antecedents of quantum physics, 569
experiments
  Davison–Germer, 573
  Rutherford, 569
explicit formula of Blasius, 795
explicit formula of Nikuradze, 795
exponential distribution, 393
exponential form of complex number, 235
exponential function, 17
exponential integral, 357, 369
exposure radiation dose, 631
extended complex plane, 232
exterior problem of hydrodynamics, 781
external angle, 37
external characteristic, 873
external forces, 413, 703, 745
external photoelectric effect, 562
external stationary field, 411
extraction from drops, 800
extraordinary wave, 554
extremal points, 156
  of functions of several variables, 167
  of functions of three variables, 167
extremely rarefied gas, 467
extremely viscous liquids, 824
extrinsic semiconductors, 610, 615

F

face of dihedral angle, 50
face of polyhedral angle, 50
face of polyhedron, 51
face of trihedral angle, 50
factorial, 9, 355
Falkner formula, 799
false position method, 160
farad, 477
Faraday law of electromagnetic induction, 502
favorable outcome, 381
Fermat principle, 536
fermi, 619
Fermi–Dirac distribution, 600
Fermi–Dirac statistics, 597, 599
Fermi energy, 600, 602, 614
  of electrons, 602
Fermi surface, 602
  of electrons in metal, 614
Fermi temperature, 603
fermion, 587, 597
Ferrari solution, 108
ferroelectrics, 485
ferromagnetic materials (ferromagnets), 498, 501, 893
  hysteresis loop, 501, 893
Fick law of diffusion, 466
fictitious forces, 428
field
  central, 412, 416, 421, 580
  coercive, 501
  electric, 470
  electric, of point charge, 471
  electric, of relativistic particle, 510
  electric, of system of point charges, 471
  electromagnetic, 470
  electrostatic, 471
  external stationary, 411
  gravitational, 425
  in conductor, 477
  magnetic, 470, 492, 890
  magnetic, at axis of circular current, 494

field (*continued*)
  magnetic, in center of circular current, 494
  magnetic, in matter, 498
  magnetic, of linear current segment, 495
  magnetic, of nonrelativistic particle, 509
  magnetic, of rod, 496
  magnetic, of solenoid, 496
  of charged ball, 476
  of charged plane, 476
  of charged string, 475
  of dipole, 473
  of disk, 473
  of segment, 472
  of source, 411
  potential, 416
  solenoidal electric, 502
  temperature, 831
  uniform, 412
film
  flowing by gravity on inclined plane, 785
  on cylindrical surface, 786
  on inclined plane, 785
film thickness, 826
fine structure constant, 586, 1042
fine structure of spectral lines, 585
finite motion, 419
first boundary value problem, 276, 313, 343, 344
first Brillouin zone, 607
first comparison criterion, 212
first cosmic velocity, 427
first Fresnel zone, 545
first law of photoeffect, 562
first law of thermodynamics, 443, 444
  for ideal gas, 445
  for simple system, 444
first noteworthy limit, 144
first-order differential equations, 261, 1011
first-order phase transition, 462
first-order quasilinear partial differential equation, 305
  with n independent variables, 306
first problem for unbranched magnetic circuits, 896
first problem of dynamics, 702
fission of excited nuclei, 638
fixed hinge, 694
flat cross-section hypothesis, 751, 758
flavor (of quark), 644
flexural rigidity, 759
flow
  axisymmetric Poiseuille, 788
  axisymmetric shear, 940
  Couette, 782
  creeping, 662, 780, 800, 824
  in rectilinear tube of constant cross-section, 787
  of viscous liquid about solid particle, 662
  past cylinder, 804, 805
  past nonspherical particles, 809, 815
  past spherical bubble, 806
  past spherical drop, 806
  past spherical particle, 802, 803
  simple shear, 782
  translational Stokes, 802, 806, 809–812, 851–853, 856–861
  translational Stokes, bodies of revolution, 811
  translational Stokes, ellipsoidal particles, 810
  translational Stokes, particles of arbitrary shape, 812
  translational Stokes, spherical bubble, 806
  translational Stokes, spherical drop, 806
  with parabolic velocity profile, 788, 826, 839
flow separation, 803
fluctuating components, 783
fluctuations, 599
fluid, 773
fluid flows in tubes, 787
fluid streamlines, 787
fluidity, 773
flux density, 442, 443
  of momentum, 466
  of heat, 466
  of particles, 466
flux of electric field, 474
flux of magnetic induction, 496
flux of vector field, 208
focal axis of ellipse, 70
focal distance, 537
  of ellipse, 69
  of hyperbola, 71
  of parabola, 74
  of spherical mirror, 537
  of thin lens, 537
focal parameter, 74
  of ellipse, 70
  of hyperbola, 72
focal plane, 536
focal power, 537
focal properties
  of ellipse, 70
  of hyperbola, 72
  of parabola, 74
focal radius, 70, 72
  of parabola, 74
focus
  of ellipse, 69, 70
  of hyperbola, 71, 72
  of optical system, 536
  of parabola, 74
focus-directrix properties
  of ellipse, 70
  of hyperbola, 72
forbidden energy bands, 609
force, 411, 687
  active, 687, 701
  Ampère, 493
  buoyancy, 803, 808
  center, 412
  central, 413
  centrifugal, 429
  coercive, 501
  conservative, 416
  constraint reaction, 412, 687, 702

force (*continued*)
  Coriolis, 429
  critical, 768–770
  distributed, 699, 745
  drag, 412
  driving, 518
  dry friction, 412
  elastic, 412
  electromotive, 489, 865
  external, 413, 703, 745
  fictitious, 428
  field, 411
  generalized, 514, 721, 723, 728
  generalized Lorentz, 492
  gravitational, 412
  gravity, 779, 803, 808
  hydrodynamic, 812
  inertial, 428
  internal, 413, 703, 716, 746
  longitudinal, 746, 750
  Lorentz, 492
  lumped, 745
  natural, 412
  net, 411, 420
  nonconservative, 416
  nuclear, 621
  of inertia, 716, 809
  of pressure on curved wall, 775
  of pressure on plane wall, 774
  of surface tension, 463
  perturbing, 729
  potential, 713, 714
  propulsive, 415
  quasielastic, 484, 514
  restoring, 514
  resultant, 411
  static friction, 700
forced (induced) absorption, 565
forced electric oscillations, 520
forced emission, 566
forced mechanical oscillations, 518
forced oscillations, 518, 729
forced transition, 565
forces of inertia, 716, 809
formal solution, 315
formula
  Bayes, 383, 384
  Bernoulli, 384
  Bessel, 364
  Blasius, 795
  Boltzmann, 450
  Cauchy–Hadamard, 241
  Cauchy integral, 239
  d'Alembert, 312
  Darcy, 794
  de Moivre, 231
  dimensional, 658
  Einstein (for volume concentration), 818
  Euler, 25, 224, 235, 360, 433, 679, 768–771
  Falkner, 799
  for calculation of volume of domain, 203
  for repeated integration by parts, 173
  Fourier cosine, inversion, 257
  Heron, 41
  Jacobi, 130
  Kader–Yaglom, 834
  Leibniz, 155
  Liouville, 271, 296
  Lorentz–Lorentz, 557
  Maclaurin, 156
  Mellin transform, inversion, 258
  Newton–Leibniz, 184
  Newton–Leibniz, for function of complex variable, 239
  Nikuradze, 795
  of finite increments, 153
  of integration by parts, 173, 239
  Planck, 561
  Poisson, 364, 385
  Prandtl–Nikuradze, 795
  Rayleigh–Jeans, 560
  Reichardt, 794
  Richardson–Dashman, 604
  Rutherford, 570
  Spalding, implicit interpolation, 799
  Stirling, 360
  Stokes, 803
  Taylor, 156
  Thomson, 517
  Torricelli, 781
  total probability, 383
  Watson, 186
  Weizsäcker, 624
  Wien, 559
  Young–Laplace, 463
  Zhuravskii, 760
formulas
  addition, 24, 31
  asymptotic, for integrals with weak singularity, 185
  asymptotic, for Laplace integrals of general form, 186
  asymptotic, for Laplace integrals of special form, 185
  asymptotic, for power Laplace integral, 186
  de Broglie, 573
  differentiation, 25, 28, 32, 35
  empirical, 921
  engineering, 922
  Euler–Fourier, 224
  for angular momentum, 733
  for Bernoulli process, 385
  for calculation of Green functions, 334, 337
  for drag coefficient, 791, 795, 796, 803–805, 808, 809, 816
  for moment of forces of inertia of body, 717
  for radius of convergence of power series, 218
  for stresses and strains, 751
  for transfer coefficients in gas, 466
  Fresnel, 552
  interpolation formulas, 925
  involving arbitrary powers, 10
  linear interpolation, 925

formulas (*continued*)
  Mollweide, 38
  three-parameter empirical, 931
  two-parameter empirical, 926
four-momentum, 437
four-vector (4-vector), 437
Fourier law, 465
Fourier coefficients, 224
Fourier cosine inversion formula, 257
Fourier cosine transform, 257, 260
Fourier expansion, 225
Fourier integral, 513
Fourier series, 223, 224, 513
  in complex form, 226
  of square-integrable functions, 226
Fourier sine transform, 255, 257, 260
Fourier transform, 255, 260, 331
  cosine, 257, 260
  sine, 255, 257, 260
fourth-degree equation, 107
fractional integral, 260
fractional power series, 245
frame acceleration, 681
frame motion, 681, 684
frame of reference, 405
frame velocity, 681
Fraunhofer approximation, 540
Fraunhofer diffraction
  at circular hole, 548
  at slit, 547
free axes, 432
free charges, 488
free material system, 687
free mechanical oscillations with single degree of freedom, 513
free oscillations in electric circuit, 516
free spectral range, 550
free support of two bodies, 694
free symmetric top, 433
free undamped oscillations, 513
freezing-out of degrees of freedom, 455
frequency, 511, 900
  of vortex shedding, 805
Fresnel biprism, 541
Fresnel formulas, 552
Fresnel integrals, 359
Fresnel mirrors, 541
Fresnel zones, 545, 546
friction laws (Coulomb laws), 700
friction temperature, 833
friction velocity, 793, 799, 833
fringes of equal inclination, 543
fringes of equal thickness, 543
Froude number, 665, 780
frustum of cone, 58
frustum of pyramid, 54
function
  analytic at point, 233
  analytic at point at infinity, 233
  analytic in domain, 233
  associated Legendre, 374, 582
  beta, 361
  Bessel, of first kind, 272, 362, 1020
  Bessel, of second kind, 272, 362, 1020
  cylinder, 362
  complementary error, 356, 369, 827, 854, 937
  composite, 140, 164, 222
  concave, 148
  continuous, 147, 163
  convex, 148
  cylindrical, 1020
  decreasing, 140
  degenerate hypergeometric (Kummer), 367, 369, 828, 836, 840
  differentiable, 150, 163, 232
  digamma, 360, 363
  Dirac delta, 320
  elementary, 15, 234
  entire, 244
  entire transcendental, 244
  error, 318, 356, 368
  error, complementary, 356
  even, 140
  exponential, 17
  gamma, 359, 1019
  Gegenbauer, 801, 1022
  generating, 356, 374
  Green, 277
  harmonic, 233
  Hermite, 378
  holomorphic at point, 233
  hyperbolic, 29
  implicit, 164
  incomplete gamma, 361, 368, 369, 829, 835, 851
  increasing, 140
  infinitely large, 146
  infinitely small, 146
  integrable, 183, 196
  integrand, 171
  inverse, 141
  inverse hyperbolic, 33
  Lagrange, 168
  Legendre, 373, 801
  Legendre, of second kind, 1021
  logarithmic, 18
  Macdonald, 365
  Mathieu, 1025, 1026
  meromorphic, 245
  modified Bessel, of first kind, 273, 365, 1020
  modified Bessel, of second kind, 273, 365, 1020
  monotone, 140
  multi-valued, 140, 232
  Neumann, 362
  nondecreasing, 140
  nonincreasing, 140
  odd, 140
  of complex variable, 229, 231
  of random variables, 387
  of several variables, 162
  of two or three variables, 162
  periodic, 140

function (*continued*)
  power, 15
  probability density, 386, 456
  psi, 360
  rational, 174
  real-valued, of real variable, 140
  regular at point, 233
  Riemann, 340, 341
  single-valued, 140, 232
  special, 355
  spherical, 582
  stream, 797, 801, 802, 807, 809, 810
  strictly concave, 149
  strictly convex, 148
  Struve, 260
  trigonometric, 19, 26
  wave, 573
  wave, of molecule, 592
  Zhukovskii, 235
function series, 215
fundamental fermions, 646
fundamental interaction, 643
fundamental laws of geometric optics, 535
fundamental oscillation, 527
fundamental solution, 320
fundamental theorem of algebra, 109
fundamental trigonometric substitution, 179
fundamental vector bosons, 643, 646
fusion reactions, 641, 642

G

Galerkin method (projection method), 287, 288
Galilean invariance, 412
Galilean relativity, 412
Galilean relativity principle, 412
Galileo transformations, 412
Gamma decay, 630
gamma function, 359, 1019
gas, 773
gas constant, 442
Gas mixing, 450
Gauss convergence criterion for series, 213
Gauss law
  for electric displacement, 482
  for magnetism, 496
Gauss theorem, 474, 475
  for electric displacement, 482
  for magnetism, 496
Gauss transform, 260
Gaussian hypergeometric equation, 1022
Gaussian method, 125
Gaussian surface, 475
Gegenbauer functions, 801, 1022
general (complete) equation of plane, 84
general correlations for Sherwood number, 852, 855, 861
general equation
  of plane, 84
  of quadratic curve, 75
  of quadric, 99
  of straight line, 65
  of straight line in space, 90
  of straight line passing through given point, 65
  of dynamics (d'Alembert–Lagrange principle), 722
general form of integral transforms, 251
general formulas for solving nonhomogeneous boundary value problems, 334, 338
general fourth-degree equation, 107
general nonorthogonal curvilinear coordinates, 1005
general orthogonal curvilinear coordinates, 1006
general properties of particles, 643
general properties of Sturm–Liouville problem, 278
general reduction formulas for calculation of improper integrals, 972
general reduction formulas for evaluation of definite integrals, 970
general Riccati equation, 267, 1012, 1013
general scheme of solution of equilibrium problem for body, 693
general solution of first-order PDE, 305
general solution of Stokes equations, 801
general square system of linear equations, 124
general theorems of dynamics of mechanical system, 703
general theory of relativity, 428
generalized coordinates, 719, 723
generalized force, 514, 721, 723, 728
generalized homogeneous equation (ODE), 264, 282, 1012, 1029
generalized Laguerre polynomials, 375
generalized Lorentz force, 492
generalized reciprocal (algebraic) equation, 107, 109
generalized triangle inequality, 11, 215
generalized velocities, 723
generating function, 356, 374
generic term of sequence, 141
generic term of series, 211
geometric constraints, 718
geometric definition of probability, 381
geometric distribution, 391
geometric mean, 11
geometric meaning of definite integral, 183
geometric meaning of derivative, 151, 234
geometric meaning of double integral, 196
geometric optics, 535
geometric parameter, 849
geometric progression, 10
geometric sequence, 10
Gergonne point, 40
Gibbs free energy, 452
Gibbs paradox, 450
gluons, 623, 643, 645, 1044
gorge ellipse, 97
Goursat problem, 312, 340
gradient, 417
Gram–Schmidt orthogonalization, 137
Gram determinant, 137
graph of function, 140
graph of network, 868

grating period, 549
gravitational mass, 425
gravitational field, 425
gravitational film, 784
gravitational force, 412
gravitational sedimentation, 817
gravity force, 779, 803, 808
greatest common divisor, 4
Greek alphabet, 1037
Green function, 277
Green theorem, 209
grey body, 558
ground state of atom, 571, 574, 583, 586, 587
group velocity, 524
growth exponent, 252
gyromagnetic ratio, 500, 583, 585
gyroscope, 433
gyroscope approximation, 433

H

Hadamard–Rybczynski solution, 807
hadron, 644
half-life period, 625
Hall effect, 492
halogens, 1047
Hamilton operator, 170, 823
Hamiltonian, 578
Hankel transform, 260
hard magnetic materials, 894
Hardy transform, 260
harmonic function, 233
harmonic mean, 11
harmonic oscillations, 511
harmonic oscillator, 572, 576, 579, 604
harmonic series, 213
harmonic waves, 522
harmonics, 513, 527
Hartley transform, 260
heat, 443
heat and mass transfer
  in circular tube with constant heat flux at wall, for laminar flows, 838
  in circular tube with constant temperature of wall, for laminar flows, 834
  in plane channels, for laminar flows, 839
heat capacity, 444
  at constant pressure, 445
  at constant volume, 444
  of crystal lattice, 604, 606
  of electrons (in crystals), 603
  of ideal gas, 455
heat conduction, 465
heat equation, 307
  with source, 316
heat exchange, 443, 450, 845, 846, 938, 939
heat flux, 832, 833, 842
heat transfer coefficient, 842
heat transfer equation, 823
heat transfer in laminar flow, 831
heat transfer in turbulent flow, 832
heat transfer to flat plate, 831–834
heavy gyroscope, 433
height of pyramid, 53
Heisenberg uncertainty relation, 573, 578
Helmholtz free energy, 451
henry, 504, 1038
Hermite functions, 378
Hermite polynomials, 368, 369, 377
Hermitian matrix, 114
Heron formula, 41
Hertz vibrator, 530
heterogeneous catalysis, 800
higher-order derivatives, 154
higher-order differentials, 154
Hilbert transform, 260
holes, 610, 613
holography, 544
holography principles, 544
holonomic constraint, 718
homogeneous boundary condition, 312
homogeneous conductor, 488
homogeneous equation, 264, 282, 312, 313, 1011, 1029
homogeneous initial condition, 312
homogeneous linear equation of second order, 1016
homogeneous linear equations, 292, 314, 271, 296
homogeneous system of linear algebraic equations, 123
hoof, 57
Hooke law, 412, 747
Hopf equation, 307
Horner scheme, 104
hot reservoir, 444
Hund rule, 589
Huygens–Fresnel principle, 545
Huygens principle, 536
Huygens–Steiner theorem, 430
Huygens theorem on moments of inertia, 705
hydraulic (or equivalent) diameter, 791, 843, 844
hydrodynamic boundary condition far from body, 782
hydrodynamic boundary conditions
  for solid body, 782
  for spherical drop, 806
hydrodynamic boundary layer, 796
  on flat plate, 796
hydrodynamic drag coefficient, 791
hydrodynamic equations, 778
  of thin films, 785
hydrodynamic force, 812
hydrodynamics, 773
  of thin films, 784
hydrogen atom, 571
  in quantum mechanics, 583
hydrogen cycle, 641
hydrogen-like ions, 571
hydrogen molecule, 593
hydrostatics, 773
hyperbola, 71
hyperbolic cosine, 29

hyperbolic cotangent, 29
hyperbolic equation, 308, 311, 352
  boundary and initial conditions, 312
  first canonical form, 309
  second canonical form, 309
  with two independent variables, 351
hyperbolic functions, 29
  of half argument, 32
  of multiple argument, 32
hyperbolic paraboloid, 99
hyperbolic sine, 29
hyperbolic tangent, 29
hyperboloid, 96
hyperboloid of revolution, 97
hyperfine structure, 619
hypergeometric distribution, 391
hypergeometric equation, 1022
hypergeometric series, 1022
hyperons, 644
hypotenuse, 41
hysteresis loop, 501

I

ideal constraints, 720
ideal emf source, 866
ideal gas law, 441
ideal source of current, 866
identity matrix, 114
illuminance, 539
image, 478
images in optical systems, 536
  real, 536
  virtual, 536
imaginary axis, 229
imaginary part, 229
imaginary quadratic curve, 75
imaginary quadric, 99
imaginary semiaxis
  of hyperbola, 72
  of one-sheeted hyperboloid, 96
  of two-sheeted hyperboloid, 97
imaginary unit, 229
impedance, 520
  reactive, 520
  resistive, 520
  triangle, 908
implicit functions, 164
impossible event, 380
improper integrals, 190
  involving logarithmic functions, 972
  involving power functions, 972
  involving trigonometric functions, 972
  of unbounded functions, 193
  with infinite integration limits, 190
improper subspace, 134
improvement of power factor, 917
impulse of force, 411, 708
incenter, 40
incenter point, 40
incidence plane, 535
incircle of triangle, 40
incoherent, 539
incompatible events, 380
incomplete cubic equation, 105
incomplete elliptic integrals, 372
incomplete gamma function, 361, 368, 369, 829, 835, 851
inconsistent system of linear algebraic equations, 123
increasing function, 140
increasing sequences, 142
indefinite integrals, 171, 945
indefinite quadratic form, 128
independent events, 383
independent random variables, 396
independent trials, 386
independent variable, 140
index
  of inertia of quadratic form, 129
  reflection, 553, 567
  refraction, 528, 535
  refraction, absolute, 535
  refraction, relative, 535
indices of refraction, 535
induced emf, 502
inductance, 504, 865
inductance coil, 866
inductance in sinusoidal current circuit, 905
inductance of solenoid, 504
inductive reactance, 520, 905
inelastic collision, 419, 438
inequalities, 5
inequalities for mean values, 11
inertia, 410
inertial forces, 428
inertial reference frames, 410, 689
infinite-dimensional linear space, 133
infinite motion, 419
infinite products, 211
infinitely large function, 146
infinitely large sequence, 143
infinitely rapid heterogeneous reaction, 822
infinitely small function, 146
infinitely small sequences, 143
inflection points, 157
initial condition(s), 311, 407, 781, 821
initial moment of two-dimensional random variable, 397
initial phase, 901
inner product, 136
inscribed angle, 47
inscribed in circle, polygon, 43
instantaneous value of power, 904, 905, 907, 909
instantaneous axis of rotation, 408, 679
instantaneous center of accelerations, 678
instantaneous center of velocities, 676
integer, 3
integer number, 3
integrable function, 183

integral, 171
  absolutely convergent, 192
  Bernoulli, 780
  Cauchy, 239
  Cauchy-type, 240
  complete elliptic,
  complete elliptic, of first kind, 370
  complete elliptic, of second kind, 370
  cosine, 358
  curvilinear, 204
  definite, 183
  Dini, 344
  divergent, 191
  double, 195, 197
  Duhamel, 294
  elliptic, 370
  elliptic, of first kind, 372
  elliptic, of second kind, 372
  elliptic, of third kind, 372
  exponential, 357, 369
  Fourier, 513, 573
  fractional, 260
  Fresnel, 359
  incomplete elliptic, 372
  indefinite, 171, 945
  iterated, 197
  Laplace, 248
  Laplace, general form, 186
  Laplace, special form, 185
  line, of first kind, 204, 205
  line, of second kind, 205
  logarithmic, 357, 369
  Poisson, 344
  power Laplace, 186
  probability, 318, 356
  Riemann, 183
  sine, 358
  surface, of first kind, 207
  surface, of second kind, 208
  triple, 200–202
integral de Moivre–Laplace theorem, 385
integral formulas of vector calculus, 208
integral friction coefficient for plate, 798
integral of differential binomial, 178
integral of function of complex variable, 238
integral sum, 183, 195
integral transforms, 251, 975
integral transforms method, 328
integrals, 171, 945
improper integrals, 190
  involving arbitrary powers, 952
  involving exponential functions, 952, 966
  involving functions of linear or rational argument, 970
  involving functions with trigonometric argument, 971
  involving hyperbolic functions, 953, 967
  involving inverse trigonometric functions, 962
  involving irrational functions, 949
  involving logarithmic functions, 956, 967, 971
  involving power functions, 182, 963
  involving rational functions, 945
  involving trigonometric functions, 957, 968
  over finite interval, 963, 968
  over infinite interval, 964, 969
  with infinite limits, 190
integrand function, 171
integrating factor, 266
integration
  by parts, 173
  of exponential functions, 179
  of expressions involving radicals of linear-fractional functions, 177
  of functions of complex variables, 238
  of hyperbolic functions, 179
  of improper fractions, 176
  of irrational functions, 177
  of polynomials multiplied by elementary functions, 181
  of product of polynomial by exponential functions, 181
  of product of polynomial by trigonometric function, 181
  of proper fraction, 175
  of rational functions, 174
  of trigonometric functions, 179
intensity of electromagnetic wave, 529
intensity of source, luminous, 538
intensity of source, radiant, 537
intensity of wave, 524
interaction between light and matter, 551
interaction unification, 650
intercept equation of plane, 85
intercept-intercept equation of straight line, 66
interfacial tension, 806
interference, 525
  from natural sources of light, 541
  in thin films, 543
  of light, 539
  of many waves, 541
  of two waves, 540
interior angle, 37
  of polygon, 43
interior heat problems for solid bodies, 845
interior problem of hydrodynamics, 781
intermediate bosons, 643
internal conversion, 630
internal energy, 443
  mechanical, of system, 418
  of ideal gas, 445, 448, 455
  of van der Waals gas, 448
internal force factors, 746
  of rod, 746, 753, 755
internal forces, 413, 703, 716, 745
internal mechanical energy of system, 418
internal photoeffect, 563
interphase transport, 661
interpolation formulas, 925
interpolation polynomial, 122
interpolation, two-point linear, 926
intersection of events, 380
intersection of straight line and plane, 95
intersection of subspaces, 134

interval of convergence of power series, 218
invariant of quadratic curve, 76
invariant of quadric, 100
invariants of system of sliding vectors, 686
inverse Fourier transform, 256, 331
inverse function, 141
inverse hyperbolic cosine, 33
inverse hyperbolic cotangent, 34
inverse hyperbolic functions, 33
inverse hyperbolic sine, 33
inverse hyperbolic tangent, 34
inverse Laplace transform, 252
inverse matrix, 119
inverse Mellin transform, 258
inverse population of levels, 567
inverse problem of kinematics, 406, 407
inverse transform, 251
inverse transforms of rational functions, 253
inverse trigonometric functions, 26
invertible matrix, 119
involutive matrix, 120
ion, 570
ionic (heteropolar) chemical bond, 593
ionic polarization in solid crystal dielectrics, 485
ionization energy, 570
irradiance, 539
irrational numbers, 4
irrigation density, 826
isobaric process, 445
isobars, 618
isochronism of harmonic oscillations, 515
isochronous, 515
isolated conductor, 477
isolated singularity, 243
isolated system of particles, 413
isomers, 630
isomorphic Euclidean spaces, 137
isomorphic linear spaces, 133
isosceles trapezoid, 45
isosceles triangle, 37, 42
isothermal bulk modulus, 442
isothermal process, 445
isotherms of real gas, 458
isotopes, 618
isotopic effect, 572
isotopic spin, 649
isotopic spin conservation law, 650
isotopic spin vector, 649
isotropic dielectric, 482
isotropic magnetic material, 500
iterated integral, 197
iteration and numerical methods, 291

J

Jacobi formula, 130
Jacobian, 165, 199
Jacobian determinant, 199
Jensen inequality, 148
jj-coupling, 588, 624
joining face of prism, 51
joint distribution function of random variables, 395
joint probability density function, 396
Jordan–Gauss method, 126
Jordan lemma, 248
Jordan transformation, 126
joule, 415
Joule–Lenz law
  for circuit segment, 490
  in differential form, 489
Joule–Thomson process, 444

K

K-capture, 628
König axes, 704
König theorem, 416
Kader–Yaglom formula, 834
Kármán constant, 784, 793, 833
Kármán hypothesis, 783
Kármán velocity defect law, 794
Kármán vortex, 805
Kepler first law, 427
Kepler second law, 421, 427
Kepler third law, 427
kernel of integral transform, 251
Kerr constant, 555
Kerr effect, 555
kilogram, 410
kinematic constraints, 718
kinematic Euler equations, 733
kinematic integrable constraint, 718
kinematic viscosity, 779
kinematics, 405, 667
  of point, 405, 667
  of rigid body, 408, 671
kinetic energy, 416, 714
  of mechanical system, 714
  of point, 714
  of system of particles, 416
kinetic-molecular theory, 441
kinetic theory of ideal gases, 453
Kirchhoff first rule, 491
Kirchhoff law, 491, 558, 868, 869
  first, 868, 869
  first, for magnetic circuits, 894
  second, 868, 869, 907
  second, for magnetic circuit, 895
  voltage, 491
Kirchhoff second rule, 491
Klein–Gordon equation, 339
Kolmogorov theorem, 400
Kontorovich–Lebedev transform, 260
Kronecker–Capelli theorem, 123
kth absolute moment, 389
kth central moment, 389
kth moment, 388
Kummer series, 367, 1019
Kummer transformation, 367
kurtosis, 390

L

L'Hospital rule, 153
Lagrange method, 129
Lagrange ordinary differential equation, 269
Lagrange equations of second kind, 723
Lagrange function, 168
Lagrange interpolation polynomial, 925
Lagrange multiplier, 168
Lagrange theorem, 153, 724
Laguerre polynomials, 368, 375
  generalized, 375
Lamé coefficients, 1006
Lambertian source, 538
laminar boundary layer, 796, 805, 831
laminar flow, 779, 784, 786, 792, 821, 831, 834, 838, 839, 843
  heat transfer to flat plate, 831, 832
  hydrodynamic boundary layer, 796–798, 804, 805, 831
  in thin films, 785
  in tubes of various cross-sections, 786, 791, 792
  mass and heat transfer, 821–823
  mass and heat transfer in circular tubes, 834–839
  mass and heat transfer in plane channels, 839–841
  mass transfer in liquid films, 826–830
  Navier–Stokes equations, 779
  transient mass transfer, 860–862
Landé splitting factor, 590
Langevin–Debye law, 484
lanthanides, 1047
Laplace equation, 308, 847
Laplace integral, 248
  general form, 186
  special form, 185
Laplace operator, 170, 823
Laplace transform, 251, 260, 328
  inverse, 251, 328
Larmor precession, 501
laser, 565–567
latent heat, 462
lateral bending, 755
lateral face of prism, 51
lateral face of pyramid, 53
Laue conditions, 550
Laue patterns, 550
Laurent series, 241, 242
law
  Ampère, 493
  Ampère circuital, 496, 891
  Ampère force, 493
  Biot–Savart–Laplace (Biot–Savart), 494
  Blasius, 798
  Bouger, 557
  Coulomb, 470, 700
  Dalton, 455
  Dulong–Petit, 604
  equipartition, 455
  Faraday, for electromagnetic induction, 502
  Fick, 465
  Fourier, 465
  for ideal gas, 441
  Gauss, for electric displacement, 482
  Gauss, for magnetism, 496
  Hooke, 412, 523, 747
  isotopic spin conservation, 650
  Joule–Lenz, for circuit segment, 490
  Joule–Lenz, in differential form, 489
  Kármán velocity defect, 794
  Kepler, first, 427
  Kepler, second, 421, 427
  Kepler, third, 427
  Kirchhoff, current, 491, 869
  Kirchhoff, first, 491, 868, 869
  Kirchhoff, for magnetic circuits, 894
  Kirchhoff, for magnetic circuits, first, 894
  Kirchhoff, for magnetic circuits, second, 895
  Kirchhoff, second, 491, 868, 869, 907
  Kirchhoff, thermal radiation, 558
  Kirchhoff, voltage, 491, 869
  Langevin–Debye, 484
  Malus, 551
  Moseley, 590
  Newton, first, 410, 689
  Newton, of cooling, 466
  Newton, of universal gravitation, 425
  Newton, second, 411, 689
  Newton, third, 411, 689
  of communicating vessels, 773
  of conservation of angular momentum, 420, 424, 711
  of conservation of electric charge, 469
  of conservation of energy, 415, 418, 487, 506, 508, 649
  of conservation of energy for electromagnetic fields, 508
  of conservation of mechanical energy, 418
  of conservation of momentum, 413, 414
  of conservation of momentum of mechanical system, 709
  of conservation of momentum projection onto any axis, 709
  of conservation of velocity of center of mass of mechanical system, 707
  of cosines, 38
  of equality of action and reaction, 689
  of hydrostatics, basic, 773, 774
  of independence of light beams, 535
  of inertia, 689
  of inertia of quadratic forms, 129
  of large numbers, 399, 400
  of light refraction, 535
  of motion, 405
  of photoeffect, first, 562
  of photoeffect, second, 562
  of proportionality between force and acceleration, 689
  of rectilinear light propagation, 535
  of relativistic dynamics, basic, 439
  of rotation of rigid body about fixed axis, 674

law (*continued*)
  of sines, 38
  of sliding friction, 700
  of tangents, 38
  of thermodynamics, first, 443
  of thermodynamics, for simple system, 444
  of thermodynamics, for ideal gas, 445
  of thermodynamics, second, 446
  of thermodynamics, third, 451
  of universal gravitation, 425
  Ohm, for circuit with capacitance, 907
  Ohm, for circuit with current source, 490
  Ohm, for circuit with inductance, 905
  Ohm, for magnetic circuit, 892, 893
  Ohm, for simple circuit segment, 489
  Ohm, for unbranched circuit, 490
  Ohm, in differential form, 488
  Pascal, 773
  Planck, 561
  Prandtl, of wall, 793
  radioactive decay, 625
  Rayleigh, 557
  Rayleigh–Jeans, 560
  Snell, 535, 552
  Stefan–Boltzmann, 559
  Stokes, 803
  velocity-addition, 407, 409, 436
  von Kármán friction, 799
  Wiedemann–Franz, 492, 603
  Wien, 559
  Wien displacement, 559
laws
  friction, 700
  Coulomb, 700
  de Morgan, 381
  of geometric optics, 535
  of mechanics, basic, 689
  of rolling friction, 700
  of statics, basic, 690
  of thermodynamics, 441
leakage flux, 894
least common multiple, 4
least squared error method, 290
least squares method, 289, 922
left-hand limit, 144
left rectangular Cartesian coordinate system, 61
leg of isosceles triangle, 42
leg of right triangle, 41
leg of trapezoid, 45
Legendre equation, 1021–1023
Legendre function, 373, 801
  of second kind, 1021
Legendre polynomials, 373, 1021
Leibniz convergence criterion for series, 213
Leibniz formula, 155
length
  contraction, 435
  of coherence, 542
  of element, 136
  of thermal initial region, 838, 839
  of vector, 78
  of wave train, 542
Lenz rule, 503
lepton, 646, 1043
  charges, 648
level line, 162
levels, acceptor, 611
levels, donor, 610
Lienard equation, 1030
light characteristics (of radiation), 538
light absorption, 555–557
light beams, 535
light dispersion, 555
  anomalous, 556
  normal, 556
light, natural, 551
light polarization, 551
light, partially polarized, 551
light pressure, 564
light quanta, 562, 563
light rays, 535
light scattering, 557
limit, 139
limit angle of total reflection, 552
limit formulas for Bernoulli process, 385
limit Nusselt number
  for circular tube, 838, 839
  for plane channel, 841
  for tubes of various cross-section, 843
limit of function at point, 144, 162
limit of sequence, 141
limit theorems, 255, 398, 399
limiting diffusion resistance, 853
limiting resistance of continuous phase, 854, 861
limits of applicability of geometric optics, 535
Lindeberg central limit theorem, 401
line in space, 89
line integral of first kind, 204, 205
line integral of second kind, 205
line of action of force, 687
line spectra, 569
linear circuits of direct current, 871
linear combination of column vectors, 118
linear combination of row vectors, 118
linear combination of vectors, 132
linear density of charge, 469
linear eccentricity
  of ellipse, 70
  of hyperbola, 72
linear elasticity, 747
linear equation, 105
  ODE of arbitrary order, 292
  ODE with constant coefficients, 292
  ODE with variable coefficients, 296
linear harmonic oscillator, 531
linear interpolation formulas, 925
linear momentum, 410
linear shear flow, 782
linear space, 131
linear span, 134
linear strains, 747
linear subspace, 134
linearity theorem, 877

linearly dependent column vectors, 118
linearly dependent elements, 132
linearly dependent row vectors, 118
linearly distributed load, 699
linearly independent column vectors, 118
linearly independent elements, 132
linearly independent row vectors, 118
linearly polarized oscillation, 513
linearly polarized wave, 551
Lindstedt–Poincaré method, 286
Liouville formula, 271, 296
Liouville theorem, 234, 244
Lipschitz convergence criterion for Fourier series, 224
Lipschitz criterion of uniform convergence of Fourier series, 226
liquid, 773
liquid-drop (hydrodynamic) nuclear model, 623
liquid metals, 824, 832, 842
Lissajous figures, 513
Lloyd mirror, 542
load intensity, 699
local coefficient of friction, 799
local (differential) diffusion flux, 824
local de Moivre–Laplace theorem, 385
local dimensionless diffusion flux, 851
  on spherical drop, 854
local equilibrium, 442
local friction coefficient, 797, 833
local frictional stress, 797
local heat flux, 835, 840
local maximum, 156
local minimum, 156
local Nusselt number, 832, 834, 843
local Reynolds number, 799
logarithm, 8
log-normal distribution, 387
logarithmic branch point, 245
logarithmic damping decrement, 518
logarithmic derivative of gamma function, 363, 366
logarithmic function, 18
logarithmic integral, 357, 369
logarithmic layer, 793, 799, 833, 842
logarithmic residue, 246
longitudinal force, 746, 750, 751
Lorentz calibration condition, 507
Lorentz force, 492
Lorentz four-dimensional vector, 437
Lorentz four-vectors, 436
Lorentz–Lorentz formula, 557
Lorentz transformations, 434
lower bound, 139
lower limit of sequence, 143
lower triangular matrix, 114
lumen, 538
luminescence, 541
luminous efficacy, 538
luminous exitance, 538
luminous flux, 537, 538
luminous intensity, 538
lumped force, 745
lumped toque, 745
lux, 539, 1038
Lyapunov central limit theorem, 401

M

Mössbauer effect, 630
Macdonald function, 365
Maclaurin–Cauchy integral criterion of convergence of series, 213
Maclaurin formula, 156
Maclaurin method, 111
Maclaurin power series, 220
Maclaurin series, 220
macroscopic molecular currents, 498
macroscopic state, 598
Madelung rule, 1047
magic nuclei, 621
magic numbers, 621
magnetic charges, 496
magnetic circuits, 890
magnetic field, 470, 492, 890
  at axis of circular current, 494
  flux density, 492, 890
  in center of circular current, 494
  in matter, 498
  induction, 492, 890
  of current plane, 497
  of linear current segment, 495
  of nonrelativistic particle, 509
  of rod, 496
  of solenoid, 496
  strength, 499, 891
magnetic flux, 496, 891
magnetic flux density, 492, 890
magnetic induction, 492, 890
magnetic moment, 493, 498, 583, 585
  induced, 498
  intrinsic, 500, 585
  microscopic, 498
  nuclear, 617, 619
  of current-carrying loop, 493
  of electron, 501, 583, 585, 1042
  of neutron, 617
  of proton, 617, 1042
  orbital, 500, 585
  orbital, projection, 583, 585, 591
  spin, 500, 585
  total, 588, 591
magnetic permeability, 500
magnetic quantum number, 582
magnetic resistance, 893
magnetic rotation of polarization plane, 555
magnetic spin quantum number, 585
magnetic susceptibility, 500
magnetic voltage, 893
magnetization, 498
  of material, 498
magnitude of complex number, 230
magnitude of vector, 78
main continuous distributions, 393

main diagonal, 114
main discrete distributions, 390
main operations over polynomials, 103
main properties
  of operations with vectors, 79
  of Laplace transform, 252
  of Mellin transform, 259
main theorems about differentiable functions, 153
majorant series, 217
Malus law, 551
Markov inequality, 400
mass, 410
  of system, 704
mass and heat exchange between particles and quiescent medium, 847
mass and heat transfer, 821
  convective, 821, 834
  in circular tubes, 834–839, 841
  in laminar flows, 821–823, 831, 834
  in plane channels, 839–842
  in turbulent flows, 825, 832
  stationary, 847
  to circular cylinder in translational flow, 859
  to nonspherical particle in translational flow, 856
  to particle of arbitrary shape at low Peclet numbers, 857
  to spherical drops and bubbles in translational flow, 853
  to spherical particle in translational flow, 850
  transient, 849
mass concentration, 821
mass defect, 620
mass density, 413, 821
mass exchange between gases and liquid films, 826
mass force density, 779
mass fraction of total mass, 821
mass number, 617
mass of body of variable density, 203
mass transfer
  convective, 821, 834
  at low Reynolds numbers, 851, 853
  at moderate and high Reynolds numbers, 852, 855
  in circular tubes, 834–839, 841
  in laminar flows, 821–823, 831, 834
  in liquid films, 826
  in turbulent flows, 825, 832
  stationary, 847
  to ellipsoidal particle in translational flow, 857
  to nonspherical particles in translational flow, 856
  to particles of arbitrary shape, 858
  transient, 849
mass transfer coefficient, 847
mathematical induction, 12
Mathieu equation, 1025
Mathieu functions, 1025, 1026
matrix, 114
matrix entry, 114
matrix of quadratic form, 128
maximum, 156
maximum modulus principle, 234
maximum stresses, 761
Maxwell–Boltzmann distribution, 457, 600
Maxwell–Boltzmann statistics, 597, 599
Maxwell distribution, 456
Maxwell equations, 506
  for constant magnetic field, 498
  in electrostatics, 476
  in magnetostatics, 498
  in medium, 508
  in vacuum, 507
Maxwell identities, 452, 453
Maxwell rule, 460
Mayer relation, 445
mean flow rate temperature, 839, 840
mean flow rate velocity, 786, 788, 826, 840
mean lifetime of nucleus, 626
mean Sherwood number, 825, 847, 850, 858, 861
  cylinders of various cross-section, 860
  spherical bubble, 854, 855, 939
  spherical drop, 854, 856
  spherical particle, 852, 940
mean value of function, 196, 201
mean value theorem, 187
  for double integral, 196
  for integration, first, 187
  for integration, second, 187
  for triple integral, 201
mean values, 11
mean velocity of nonisotropic particles falling in fluid, 815
mechanical constraints, 687, 718
mechanical energy, 418
  of system of particles, 418
mechanical motion, 405
mechanical power, 416
mechanical system, 687
mechanical work, 415
mechanism of compound nucleus, 636
mechanism of direct nuclear reactions, 636
mechanism of nuclear reactions, 635
median of trapezoid, 45
median of triangle, 39
medium particle, 521
Meijer transform (K-transform), 260
Meler–Fock transform, 260
Mellin transform, 258, 260
  inversion formula, 258
melting–freezing curve, 461
meromorphic function, 245
Mertens theorem, 215
mesh-current method, 881
Meshcherskii equation, 415
meson, 644, 1045
metacenter, 777
metacentric height, 778
metal, 609
metastable states, 458

method
  averaging, 287
  bisection, 160
  Bubnov–Galerkin, 288
  collocation, 289
  combined, for verifying empirical formulas, 932
  direct, for verifying two-parameter formulas, 926
  electrostatic image, 478
  energy (for determining frequency), 512
  false position, 160
  Fourier, 321
  Galerkin, 287, 288
  Gaussian, 125
  integral transforms, 328
  iteration and numerical, 291
  Jordan–Gauss, 126
  Lagrange, 129
  least squared error, 290
  least squares, 289, 922
  Lindstedt–Poincaré, 286
  Maclaurin, 111
  mesh-current, 881
  moment, 289
  Newton, 111
  Newton–Raphson, 161
  nodal-pair, 883
  node-potential, 491, 882
  of asymptotic analogies, 934
  of complete elimination of free parameters for verifying empirical formulas, 926
  of electron pairs, 594
  of Euler polygonal lines, 270
  of "integration by differentiation", 268
  of molecular orbitals, 594
  of moments, 289
  of negative masses, 698
  of partial elimination of free parameters for verifying empirical formulas, 928
  of partition, 698
  of partitioning domain, 290
  of regular (direct) expansion in powers of small parameter, 285
  of scaled parameters, 286
  of sections, 746
  of separation of variables, 321
  of successive approximations (for Cauchy problem), 291
  of successive approximations, 269
  of Taylor series expansion in independent variable, 270
  of undetermined coefficients, 174
  of vector diagrams, 512
  perturbation, 780
  Picard, 269
  proportional parts, 883
  regula falsi, 160
  Runge–Kutta, for Cauchy problem, 291
  Runge–Kutta, of fourth-order approximation, 271
  shooting, 291
methods
  for finding center of gravity, 698
  for finding quotient and remainder, 104
  for interpreting indeterminate expressions, 154
  of regular series expansions with respect to independent variable, 283
  perturbation, in problems with small parameter, 284
metric tensor, 1005
metric tensor components, 1005, 1008, 1009
microscopic nuclear model, 623
microscopic state, 598
midline, 38
minimal polynomial, 122
Minkowski inequality, 11, 82, 136, 187, 215
minor, 113, 117
mixed boundary value problem, 276, 314
mixed connection of nonlinear elements, 890
mixture of ideal gases, 455
mode of continuous probability distribution, 390
mode of discrete probability distribution, 390
model
  atomic, nuclear, 570
  atomic, planetary, 570
  atomic, Thomson, 569
  Bohr, 569, 570
  Boussinesq, 783, 792, 798
  cellular, 818
  liquid-drop (hydrodynamic) nuclear, 623
  microscopic nuclear, 623
  nuclear atomic, 569
  nuclear evaporation, 637
  nuclear, Fermi gas, 623
  planetary, 570
  Prandtl, 826
  of dispersion medium, 555
  of transfer phenomena in gases, 465
  shell, of nucleus, 624
  statistical, of decay, 637
  Thomson, 569
  vector, of angular momentum, 580
  von Kármán, 826
modified Bessel equation, 273, 1020
modified Bessel function, 273, 365
  of first kind, 273, 365, 1020
  of second kind, 273, 365, 1020
modified Mathieu equation, 1024
modulated oscillations, 513
modulus
  of complex number, 230
  of number, 7
modulus of elasticity
  of first kind, 748, 749
  of second kind, 750
molar heat capacity, 445
mole, 441
molecular currents, 498
molecular optics, 555
molecular orbitals, 593
molecular physics, 441
molecular polarizability, 484

molecular thermal conduction layer, 841
Mollweide formulas, 38
moment, 388
  arm, 420
  arm of force, 687
  method, 289
  of angular momentum, 710
  of force, 420, 687
  of force about axis, 420
  of force about center, 687
  of inertia, 429
  of inertia of cross-section about axis, 759
  of momentum, 421
  of momentum of point about axis, 710
  of momentum of point particle about fixed center, 710
  of resistance of cross-section under bending, 761
  of two-dimensional random variable, 397
moments of inertia, 704
  of body, 203
  of flat plate, 200
  of some bodies, 430
momentum, 410, 414, 708
  conservation law, 709
  in special relativity, 437
  of mechanical system, 708
  of point particle, 410, 708
  of unclosed system, 414
momentum-energy four-vector, 437
monotone functions, 140
monotone sequences, 142
Morera theorem, 239, 241
Moseley law, 590
most probable velocity, 456
motion
  circular, 407
  in central field, 421, 427, 580, 582
  in central field, hydrogen atom, 580
  in central field of gravity, 427
  of center of mass, 414
  of rigid body about fixed axis, 430, 672, 674, 711, 730
  of rigid body about fixed point, 409, 431, 732, 734
  of part of loop, 503
  of several particles, 816
  of two/several spheres, 817
  of wire loop in constant magnetic field, 502
  plane, 408, 431
  relative to rotational frame, 408
  rotational, 408
  translational, 408
moving hinge, 694
multimodal distribution, 390
multiplication
  of matrix by scalar, 115
  of complex numbers, 229
  of matrices, 116
  of numbers, 5
  of polynomials, 103
  of power series, 221
  of real numbers, 5
  of series, 215
multiplicative inverse of real number, 5
multiplicative separable solution, 317
multiplicity of eigenvalue, 121, 322
multiplicity of root, 293, 294, 299
multiplicity of radiation line, 589
multi-valued function, 140, 232
muon, 647, 1043
mutual induction, 504, 505
mutuality theorem, 505

N

Napierian base, 19
natural coordinate axes, 669
natural forces, 412
natural light, 551
natural logarithms, 142
natural method for specifying motion of point, 669
natural number, 3
natural oscillations, 526
natural variables, 452
nature of nuclear forces, 622
Navier–Stokes equations, 779, 787, 824
n-dimensional coordinate space, 132
n-dimensional linear space, 133
necessary condition
  for series to be convergent, 211
  of extremum, 167
necessary and sufficient conditions
  for existence of extremum, 157
  for vector field to be potential, 206
needle-like rod, 811
negative binomial distribution, 392
negative definite quadratic form, 128
negative index of inertia of quadratic form, 129
neighborhood of point
  at infinity, 232
  on complex plane, 232
Nernst theorem, 451
net force, 411, 420
net moment
  of forces of inertia, 717
  of momentum, 710
network, 867
Neumann function, 362
Neumann problem, 314
neutral axis, 759
  in case of eccentric extension or compression, 765
  in case of oblique bending, 763
neutral layer, 758
neutrino, 629, 643, 646, 647, 1043
neutron, 619, 1041, 1045
neutron capture, 639
neutron multiplication coefficient, 639
newton, 411, 1038
Newton first law, 410, 689
Newton hypothesis, 740

Newton law of cooling, 466
Newton law of universal gravitation, 425
Newton method, 111
Newton rings, 544
Newton second law, 411, 689
Newton third law, 411, 689
Newton–Leibniz formula, 184
  for function of complex variable, 239
Newton–Raphson method, 161
newtonian fluid, 779
newtonian regime of flow, 816
nicol, 554
nilpotent matrix, 120
no-slip conditions, 782, 802, 809
noble gases, 1047
nodal-pair method, 883
node-potential method, 491, 882
nodes
  of standing wave, 526
  of wave function, 575
nominal mode, 873
non-inertial reference frames, 427
nonconservative force, 416
nondecreasing function, 140
nondegenerate matrix, 119
nondegenerate quadratic curves, 75
nondegenerate quadric, 99
nonequilibrium systems, 442
nonequilibrium thermodynamics, 442
nonholonomic constraint, 718
nonhomogeneous linear equations, 293, 297, 319
  existence theorem, 274
nonhomogeneous system
  of linear algebraic equations, 123
  of linear first-order ODEs, 301
nonideal source, 872
nonincreasing function, 140
nonintegrable constraint, 718
noninertial reference frames, 689
nonlinear electric circuits of direct current, 884
nonlinear elements, 884
nonlinear equations of arbitrary order (ODEs), 301
nonnegative quadratic form, 128
nonpolar dielectrics, 484
nonpolar molecule, 484
nonpositive quadratic form, 128
nonrelativistic limit, 411
nonrelativistic velocities, 434
nonsingular matrix, 119
nonstationary constraints, 718
norm of element, 136
normal (longitudinal) force, 746
normal acceleration, 406, 668
normal coordinates, 727
normal dispersion, 556
normal distribution, 394
normal matrix, 117
normal pressure force, 700
normal representation of quadratic form, 131
normal stress, 746
normal vector, 85
normal vibrational modes, 515
normed Euclidean space, 136
n-sided pyramid, 53
nth arithmetic root, 7
nth-order determinant, 112
nth radical, 7
nuclear atomic model, 569
nuclear binding energy, 619
nuclear composition, 617
nuclear dimensions, 619
nuclear evaporation model, 637
nuclear Fermi gas model, 623
nuclear fission, 637
nuclear forces, 620–622
nuclear hydrodynamic model, 623
nuclear magnetic moment, 619
nuclear magneton, 617
nuclear models, 623
nuclear multiplet, 649
nuclear physics, 617
nuclear power engineering, 637
nuclear reaction energy, 634
nuclear reactions, 632
nuclear shape, 619
nuclear spin, 619
nuclear time, 636
nuclear transformations, 625
nucleons, 617
nuclides, 617
null matrix, 114
null subspace, 134
null vector, 78
number e, 8, 17, 19
number of collisions against wall, 454
number of degrees of freedom, 455
  of mechanical system, 719
number sets, 139
numerical eccentricity of ellipse, 70
numerical integration of differential equations, 270
numerical sequence, 141
numerical series, 211
Nusselt number, 834, 837
  for circular tube, 837
  for thermal stabilized region, 842
nutation of Earth axis, 433

O

obelisk, 54
oblate ellipsoid of revolution, 809
oblate spheroid, 96
oblique asymptote, 145
oblique bending, 763
oblique prism, 51
obtuse triangle, 37
odd function, 140
odd number, 3
ODE, *abbreviation for* ordinary differential equation

Ohm law, 868, 869
  for circuit with capacitance, 907
  for circuit with current source, 490
  for circuit with inductance, 905
  for magnetic circuit, 892, 893
  for simple circuit segment, 489
  for unbranched circuit, 490
  in differential form, 488
one-dimensional chain, 606
one-dimensional random variables, 386
one-sheeted hyperboloid, 96
open circuit, 866
open circuit mode, 873
open intervals, 139
operations with fractions, 6
operations with power series, 221
operations with powers and roots, 8
operator
  angular momentum, 580
  coordinate, 578
  differential, 170, 287, 311–320, 779, 1007–1009
  Hamilton, 170, 578, 824
  Hermitian, 578
  Laplace, 170, 581, 824, 1007–1009
  momentum, 578, 579
  orbital angular momentum, 580
  squared angular momentum, 580
operator approach in quantum mechanics, 578
operators
  differential, of field theory, 170, 1007–1009
  of physical quantities, 578
opposite of real number, 5
opposite vector, 78
optical branches, 608
optical path difference, 541
optical path length, 536
optical property
  of hyperbola, 73
  of ellipse, 71
  of parabola, 75
optically active media, 554
optically anisotropic media, 553
optics, 535
  geometric, 535
orbital quantum number, 582
orbital velocity, 427
order of branch point, 245
order of interference, 525, 541
order of pole, 244
order of zero of function, 243
ordered pair of objects, 379
ordinary differential equation, 261, 1011
  Abel, 1014, 1015
  autonomous, 1011, 1028, 1030
  Bernoulli, 265, 1011
  Bessel, 272, 1020
  Clairaut, 269
  degenerate hypergeometric, 1019
  Duffing, 285
  Emden–Fowler, 1028
  Ermakov, 1029
  Euler, 272, 297, 782, 1020
  exact, 265, 267
  Gaussian hypergeometric, 1022
  general Riccati, 267, 1012, 1013
  generalized homogeneous, 264, 282, 1012, 1029
  hypergeometric, 1022
  hypergeometric, degenerate, 1019
  homogeneous, 264, 282, 312, 313, 1011, 1029
  homogeneous linear, of second order, 1016
  Lagrange, 269
  Legendre, 1021–1023
  Lienard, 1030
  linear, of arbitrary order, 292
  linear, with constant coefficients, 292
  linear, with variable coefficients, 296
  Mathieu, 1025
  Meshcherskii, 415
  modified Bessel, 273, 1020
  modified Mathieu, 1024
  of damped oscillations, 272, 517
  of free oscillations, 1017
  of forced oscillations, 518
  of harmonic oscillations, 511
  Riccati, 267, 269, 1012
  Riccati, general, 267, 1012, 1013
  Riccati, special, 1013
  second-order constant coefficient linear, 272, 1017
  separable, 1011
  special Riccati, 1013
ordinary differential equations, 261–303, 1011–1033
  admitting reduction of order, 281, 302
  arbitrary order, 292–303
  arbitrary order, linear, 292–302
  arbitrary order, nonlinear, 301–303
  Euler, dynamic, 733–735
  Euler, kinematic, 733
  first-order, 261–271, 1011–1016
  integrable by quadrature, 261
  invariant under scaling–translation transformations, 283
  involving arbitrary functions, 1027
  involving exponential and other functions, 1024
  involving power functions, 1017
  nonlinear, of arbitrary order, 301
  not solved for derivative, 262, 268
  second-order, 271–292, 1016–1033
  second-order, linear, 271–280, 1016–1028
  second-order, nonlinear, 281–292, 1028–1033
  solved for derivative, 261, 263, 281
  solved for highest derivative, 301
  with separable variables, 263
  with separated variables, 263
ordinary wave, 554
ordinate, 61, 79
ordinate axis, 61
orientational mechanism of magnetization in paramagnetic materials, 501

orientational mechanism of polarization (dipolar polarization), 484
oriented area, 63
origin in space, 79
origin of affine coordinate system, 134
origin on plane, 61
origination of bands from atomic levels, 609
orthocenter of triangle, 40
orthogonal coordinates, 1006
orthogonal curvilinear systems of coordinates, 1005
orthogonal elements, 136
orthogonal matrix, 117
orthogonal polynomials, 373
orthogonality of associated Legendre functions, 374
orthogonalization, 137
orthonormal basis, 137
  in finite-dimensional Euclidean space, 137
orthotropic bodies, 812
oscillating wake, 804
oscillation amplitude, 511
oscillation frequency, 511
  angular, 511
  circular, 511
oscillations, 511
  coherent, 512
  damped, 517
  damped, electric, 518
  forced, 518
  forced, electric, 520
  forced, mechanical, 518
  free, 511
  free, in elastic rod, 526
  free, in electric circuit, 516
  harmonic, 511
  isochronous, 515
  natural, 511, 526
  natural, in elastic rod, 526
oscillator, 530
  Hertz, 530
  linear harmonic, 531, 576
  quantum, 579, 604
  quantum, three-dimensional, 579, 580
oscillatory systems with several degrees of freedom, 515
osculating circle, 669
Ostrogradsky–Gauss theorem, 208
overtones, 527

P

pairing, 621
parabola, 74
parabolic equation, 307, 311, 351
  initial and boundary conditions, 311
  with two independent variables, 349
paraboloid, 97
  of revolution, 98
paradox
  Gibbs, 450
  twin, 436
parallel-plate capacitor, 479
parallel axis theorem, 430
parallel connection
  of capacitors, 491
  of emf sources, 491
  of nonlinear elements, 889
  of resistors, 491, 877
parallel lines, 49, 67, 68
parallel planes, 49
parallel vectors, 78
parallelepiped, 51, 52
parallelogram, 44
parallelogram rule, 78
paramagnetic materials, 498, 501
parameter of ellipse, 70
parameter of hyperbola, 72
parametric equations
  of curve on plane, 64
  of straight line in space, 89
  of straight line on plane, 66
paraxial approximation, 536
parent nucleus, 625
Parseval identity, 227
Parseval relation
  for Fourier cosine transform, 257
  for Fourier sine transform, 258
partial density, 821
partial derivative, 163
partial differential equation
  continuity, 779, 787
  convective diffusion, 860
  convective heat transfer, 823
  convective mass transfer, 821
  dimensionless mass transfer, 822
  elliptic, 308, 311, 313
  first-order quasilinear, 307
  for stream function, 797, 801
  heat, 307
  heat, with source, 316
  heat transfer, 823, 831
  Hopf, 307
  hyperbolic, 308, 309, 311, 312, 352
  hyperbolic, first canonical form, 309
  hyperbolic, second canonical form, 309
  Klein–Gordon, 339
  Laplace, 308, 847
  mass transfer, 825
  mass transfer, dimensionless, 822
  parabolic, 307, 311, 351
  Poisson, 477
  Poisson, three-dimensional, 320
  Poisson, two-dimensional, 320
  Schrödinger, 574
  Schrödinger, time-dependent, 574
  Schrödinger, time-independent, 574
  Schrödinger, three-dimensional, 579
  telegraph, 317
  thermal boundary layer, 831
  wave, 308, 522, 523

partial differential equations, 305–353
  convective mass and heat transfer, 821, 822
  elliptic, 308, 310
  Euler, for ideal incompressible fluid, 780
  first-order quasilinear, 305–307
  first-order quasilinear, with n independent variables, 306
  hydrodynamic, 789
  hyperbolic, 308, 309
  linear, 307–353
  Navier–Stokes, for laminar flow, 779, 787, 824
  parabolic equations, 308, 309
  Poisson, 735
  Reynolds, for turbulent flow, 783
  second-order linear, 307–353
  Stokes, 800
partial fraction decomposition of rational function, 174
partial increment of function, 163
partial limit of sequence, 143
partial sum of series, 211
partially polarized light, 551
particle, 405
particle identity, 586
particle in uniform magnetic field, 493
particle spin, 585, 647
partition diameter, 195
Pascal law, 773
Paschen–Buck effect, 591
passive elements, 866, 867, 884
path difference, optical, 541
path integral, 204
Pauli exclusion principle, 586, 587
PDE, *abbreviation for* partial differential equation
Peclet number, 822, 824, 831, 835–839, 850, 851–862
pencil of straight lines, 66
pendulum oscillations in field of gravity forces, 660
percentage, 6
perfect elements, 866
perfect non-wetting, 463
perfect wetting, 463
perfectly flexible unstretchable massless line, 694
perimeter-average Nusselt number, 843, 844
perimeter-equivalent factor, 813, 848
perimeter-equivalent sphere, 813, 848
period, 900
  diffraction grating, 549
  oscillations, 511
periodic boundary conditions, 604
periodic function, 140
periodic table (Mendeleev table), 587, 1047
permutation, 379
perpendicular axis theorem, 430
perpendicular bisector, 40
perpendicular lines, 67, 68
perpetual motion machine
  of first kind, 444
  of second kind, 446
perturbation method, 780
perturbation methods in problems with small parameter, 284
perturbing forces, 729
phase, initial, 511
phase of harmonic oscillation, 461, 901
phase displacement, 901
phase equilibrium, 460
phase of complex number, 230
phase transition heat, 462
phase transition entropy, 462
phase velocity, 524
phenomenon of birefringence, 554
phonon, 608
photoeffect, 562, 563
  Einstein equation, 563
  external, 562
  first law, 562
  internal, 563
  second law, 562
photometric characteristics of light wave, 538
photometry, 537
photon, 563, 600, 630, 641, 643–646, 648
physical applications of double integral, 200
physical applications of integral, 189
physical constants, 1040
physical foundations of mechanics, 405
physical kinetics, 442
physical laws for structure elements, 868
physical meaning of derivative, 151
physical pendulum, 515
physical tables, 1037
physical similarity, 664
physical simulation, 664
Picard method, 269
piezometric head, 775
piezometric plane, 775
Planck constant, 561, 1040
Planck law, 561
Planck quantization assumption, 561
plane
  in space, 84
  polarization, of wave, 528, 551
  transmission, of polarizer, 551
plane channel
  with constant heat flux at wall, 841
  with constant temperature of wall, 839
plane geometry, 37
plane mirror, 537
plane motion of rigid body, 408, 431
plane of action, of couple, 689
plane Poiseuille flow, 787
plane polarized wave, 528, 551
plane refracting surface, 537
plane stress state, 748
plane system of forces, 693
plane triangle, 37
plane wave, 521
plane-parallel (plane) motion of rigid body, 675, 731
planetary model, 570
plasticity, 747

p-n transition, 615
pnictogens, 1047
Pochhammer symbol, 355
point-slope equation of straight line, 65
point at infinity, 232
point charges, 470
point mass, 689
point of affine space, 133
point of extremum, 167
point of intersection of straight lines, 68
point of local maximum, 167
point of local minimum, 167
point of tangency, 46
point particle, 405
point set, 162
point of discontinuity of function, 147
pointwise convergence of function series, 215
pointwise convergence of functions, 149
Poiseuille flow
  axisymmetric, 788
  plane, 787
Poisson equation
  for adiabatic process, 445, 735
  for electric potential, 477
  three-dimensional, 320
  two-dimensional, 320
Poisson formula, 364, 385
Poisson coefficient, 749
Poisson distribution, 392
Poisson integral, 344
Poisson theorem, 400
polar angle, 62
polar axis, 61
polar coordinates, 61, 62, 1008
polar coordinate system, 61
polar dielectric, 484
polar form of complex number, 230
polar molecules, 484
polar moment, 756
  of inertia, 753
  of inertia of cross-section, 754
  of resistance of cross-section, 755
polar radius, 62
polarization density, 481
polarization density vector, 481
polarization of dielectric, 481
  moving in magnetic field, 510
polarizing prisms, 554
polarizer, 551
pole, 61, 243
polygon, 43
polyhedra, 51
polyhedral angle, 50
polyhedron, 51
polynomial, 103
  Lagrange interpolation, 925
  with matrix argument, 120
polynomial remainder theorem, 104
polytropic index, 445
polytropic process, 445
position vector, 79, 405, 667
  of point, 79
positive definite quadratic form, 128
positive index of inertia of quadratic form, 129
positive integer power of matrix, 120
postulate of statistical physics, 600
postulates
  Bohr, 570
  of special relativity, 434
potential, 206, 865
  of conductor, 477
  of electrostatic field, 472
  of field of point charge, 473
  of field of system of point charges, 473
  of gravitational field, 203, 426
  of vector field, 206
potential curves, 419
potential diagram, 871
potential difference, 865
potential energy, 416, 417
  of charge in electrostatic field, 472
  of conservative interaction, 417
  of interaction of charged particles, 485
  of particle in external potential field, 417
potential field, 416
potential flow of ideal fluid, 781
potential force, 713
potential function of current loop in magnetic field, 497
potential vector field, 206
power, 7, 865
power function, 15
power Laplace integral, 186
power of direct current, 871
power on subcircuit, 521
power series, 218
  expansions of elementary functions, 25, 29, 32, 35
powers, 7, 231
  of hyperbolic functions, 31
  of square matrices, 120
  of trigonometric functions, 24
Poynting vector, 508
Prandtl hypothesis, 783
Prandtl law of wall, 793
Prandtl model, 826
Prandtl–Nikuradze implicit formula, 795
Prandtl number, 824, 832
precession of gyroscope, 433
pressure, 779
  ideal gas, 453
  light, 564
  wave, 529
pressure gradient, 787
primary system of equations of equilibrium of rigid body, 692
prime number, 3
primitive function, 171
principal axes of inertia, 437
principal central axes of inertia, 763
principal dielectric permittivities, 553
principal directions, 748

principal drags, 812, 815
  axial and transversal, 814
  dumbbell-like particle, 813
  needle-like ellipsoids, 813
  nonspherical bodies, 813
  thin circular cylinders, 813
  thin circular disk, 813
principal maxima of diffraction grating, 549
principal moments of inertia, 432
principal normal, 669
principal optic axis, 536
principal part of Laurent series, 242
principal plane of crystal, 553
principal postulate of statistical physics, 598
principal quantum number, 570, 583
principal stresses, 748
principle
  Archimedes, 777
  current compensation, 877
  d'Alembert, 716
  Duhamel, first, 350
  Duhamel, second, 351
  Fermat, 536
  Galilean relativity, 412
  holography, 544
  Huygens, 536
  Huygens–Fresnel, 545
  maximum modulus, 234
  of corresponding states, 461
  of independence of forces, 747
  of invariant light speed, 434
  of laser operation, 565
  of linear superposition, 297, 315, 880
  of relativity, 434
  of superposition, 297, 315, 425, 471, 473, 494, 521, 573, 880
  of superposition, for field strength, 471
  of superposition, for fields, 471, 473, 494
  of superposition, for homogeneous system, 298
  of superposition, for magnetic induction, 494
  of superposition, for potential, 473
  of superposition, for wave functions, 573
  of superposition, for waves, 521
  of uniqueness, 478
  of virtual displacements, 720
  Pauli exclusion, 586, 587
  Rayleigh–Ritz, 279
  uncertainty, 573, 578
  voltage compensation, 876
prism, 51
probability, 381
probability density function, 386, 456
probability element, 387, 396
probability flux density, 574
probability integral, 318, 356
probability multiplication theorem, 383
probability of event, 382
probability of random events, 380
probability of union of events, 382
probability theory, 379
problem
  Banach, 385
  Blasius, 797, 798
  boundary value, first, 276, 313, 343, 344
  boundary value, mixed, 276, 314
  boundary value, second, 276, 313, 344
  boundary value, third, 276, 314
  Cauchy, 262, 281, 291, 302, 306, 312
  Dirichlet, 314
  for unbranched magnetic circuits, first, 896
  for unbranched magnetic circuits, second, 897
  Goursat, 312, 340
  Neumann, 314
  of dynamics, first, 702
  of dynamics, second, 702
  of hydrodynamics, exterior, 781
  of hydrodynamics, interior, 781
  of kinematics, direct, 406
  of kinematics, inverse, 406, 407
  of maximal power transfer, 874
  on motion of rigid body about fixed point, 734
  statically determinate, 692
  statically indeterminate, 692
  Sturm–Liouville, 277, 278, 334–336, 338, 339, 343
  two-body, 424
problems
  boundary value, for elliptic equations, 348
  boundary value, for elliptic equations with two space variables, 342
  boundary value, for hyperbolic equations, 347
  boundary value, for hyperbolic equations with one space variable, 336
  boundary value, for parabolic equations, 345
  boundary value, for parabolic equations with one space variable, 332
  eigenvalue, 277
  for homogeneous linear equations, 349
  for hyperbolic equation, 338
  for nonhomogeneous linear equations, 351
  for parabolic equation, 334
  for strength analysis, 752
  heat, for solid bodies, 845
  of analysis and synthesis of network, 869
  of statics, 690
  with boundary conditions of first kind, 279
  with many space variables (PDEs), 345
procedure of solving Cauchy problem, 306
process equation, 442
process with constant heat capacity, entropy change, 449
product
  of complex numbers, 230
  of hyperbolic functions, 31
  of matrices, 116
  of matrix by scalar, 115
  of polynomials, 103
  of scalar and vector, 131
  of trigonometric functions, 23
  of vector by number, 78
progressions, 10
prolate ellipsoid of revolution (spheroid), 810
prolate spheroid (ellipsoid of revolution), 96

proof by contradiction, 12
proof by counterexample, 13
propellant use, 415
proper subspace, 134
proper time, 435
properties
  of Bose–Einstein distribution, 600
  of concave functions, 149
  of continuous functions, 147
  of convex functions, 149
  of convergent sequences, 141
  of convergent series, 212
  of couple of forces, 689
  of definite integrals, 184
  of determinants, 113
  of dimensions, 656
  of double integral, 196
  of electric charge, 469
  of electromagnetic waves, 528
  of exponential function, 17
  of Fermi–Dirac distribution, 600
  of fractions, 6
  of hyperbolic functions, 30
  of incompressible (dropping) liquid, 773
  of indefinite integral, 172
  of infinitely small and infinitely large functions, 146
  of integrals in terms of identities, 184
  of integrals in terms of inequalities, 187
  of internal forces, 703
  of inverse hyperbolic functions, 34
  of inverse trigonometric functions, 27
  of limits, 144
  of linear equations, 314
  of logarithmic function, 19
  of logarithms, 8
  of moment of inertia, 430
  of nuclear forces, 621
  of positive sequences, 142
  of power function, 16
  of power series, 219
  of quadrilaterals, 43
  of stationary states, 575
  of Sturm–Liouville problem, 335, 339
  of trigonometric functions, 22
  of triple integral, 201
  of uniformly convergent series, 217
proportional parts method, 883
proportionality limit, 771
proportions, 6
propulsive force, 415
protium, 618
proton, 617, 1041, 1042, 1045
psi function (digamma function), 360
Ptolemy theorem, 43
Puiseux power series, 245
pure active resistance, 913
pure bending hypotheses, 758
pure imaginary complex number, 229
pure semiconductor, 610
pyramid, 53
Pythagorean theorem, 42, 136, 659

Q

quadrant, 61
quadratic, 75
quadratic curve, 68, 75
quadratic equation, 105
quadratic form, 127
quadratic form, with n variables, 127
quadratic mean, 11
quadric, 99
quadric surfaces, 96
quadrics
  canonical equations, 96
  general theory, 99
qualitative analysis of functions, 158
qualitative model of transfer phenomena in gases, 465
qualitative properties of integrals, 184
qualitative theory of energy bands, 609
quality factor of oscillatory system, 519
quantity of heat, 443
quantum amplification coefficient, 567
quantum chromodynamics, 645
quantum generator, 567
quantum mechanics, 573
quantum mechanical description of energy regions, 611
quantum number
  magnetic, 582
  magnetic, spin, 585
  orbital, 582
  principal, 570, 582, 587
  radial, 582
  spin, 582, 585
  spin, secondary, 585
  spin, magnetic, 585
quantum oscillator, 572, 576, 579, 580
quantum statistics, 597
quantum statistical distributions, 597
quantum theory of crystals, 597
quark color, 645
quark confinement, 646
quark flavor, 644
quark structure of hadrons, 644, 1045
quarks, 625, 645, 1044
quarter wave plate, 554
quasi-equilibrium process, 442
quasi-momentum of electron, 611
quasi-momentum of phonon, 608
quasielastic force, 484, 514
quasiparticles, 608
quasistatic process, 442
quasistationary approximation (electromagnetic induction), 504
quotient, 103
  of complex numbers, 230

R

radial quantum number, 582
radiant energy, 537

radiant exitance, 538
radiant flux, 537
radiant intensity, 537
radiation
  of moving charges, 530
  of plane current, 529
  Vavilov–Cherenkov, 544
radiation capacities, 558
radiation capacity of black body, 558
radiation capture, 639
radiation dose measurement, 631
radiation quality factor, 631
radiation weighting factor, 631
radicals, 231
radioactive decay law, 625
radioactive families, 630
radioactive sample activity, 625
radioactivity, 625
radius
  of convergence, 218
  of convergence of power series, 218, 241
  of curvature, 669
  of sphere, 58
ramification point, 245
random event, 380
random variable, 386
range of function, 140
rank of matrix, 118
rank of quadratic form, 128
rarefied gas, 467
rate of suspension precipitation, 818
rational function, 174
rational number, 4
rational polynomial function, 174
ratios, 6
Rayleigh criterion, 549
Rayleigh–Jeans law, 560
Rayleigh law, 557
Rayleigh–Ritz principle, 279
Rayleigh scattering, 557
reactance, 520
  capacitive, 520, 906
  inductive, 520, 905
  of circuit, 520, 908
reaction energy, 634
reactive capacitive power, 907
reactive current, 918
reactive inductive power, 906
reactive power, 909
real-valued functions of real variable, 140
real axis, 139, 229
real Euclidean space, 136
real gases, 458
real linear space, 132
real matrix, 117
real number, 3, 4
real part, 229
real semiaxes
  of hyperbola, 72
  of one-sheeted hyperboloid, 96
  of two-sheeted hyperboloid, 97
rear stagnation point, 802
reciprocal (algebraic) equation, 107
reciprocal of real number, 5
reciprocity law for shear stresses, 748
reciprocity theorem, 877
rectangle rules, 195
rectangular Cartesian coordinate system, 61, 79
rectangular hyperbola, 72
rectangular parallelepiped, 52
rectilinear tube of arbitrary cross-section, 790
reduced length, 770
reduced length coefficient, 770
reduced length of physical pendulum, 515
reduced mass, 424
reduction formulas, 22, 961
reduced Planck constant, 561, 1040
reduction of Riccati equation to second-order linear equation, 267
reference frame, 405
reflection from barrier, 577
reflection of waves, 527, 551
reflection of traveling wave, 527
refraction of waves, 551
region of stabilized flow, 792, 795, 798
regula falsi method, 160
regular part of Laurent series, 242
regular polygon, 45
regular polyhedra, 54
regular polyhedron, 54
regular prism, 51
regular pyramid, 53
regular triangle, 37, 43
Reichardt formula, 794
relation
  between energy and mass, 438
  between field strength and potential, 472
  between force and potential energy, 417
relations for Green function, 342
relationship between Cartesian and polar coordinates, 62
relative acceleration, 681
relative frequency, 381
relative index of refraction, 535
relative motion of point, 681–683
relative spectral sensitivity, 538
relative trajectory, 681
relative velocity, 425, 681
relativistic energy, 437
relativistic momentum, 437
relativistic properties of fields, 509
relativity of motion, 407
relativity of simultaneity, 435
relaxation time, 442
remainder, 103
remainder in Taylor formula
  in form of Cauchy, 156
  in form of Lagrange, 156
  in form of Peano, 156
  in form of Roche, 156
  in form of Schlömilch, 156
  in integral form, 156
remainder of function series, 216

remainder of series, 211
removable singularity, 243
repeated integration by parts, 173
representation
  of complete elliptic integrals in series form, 371
  of elementary functions by Maclaurin series, 220
  of Green function in terms of particular solutions, 277
  of inverse transforms as convergent series, 255
  of sinusoidal quantities by vectors on complex plane, 901
  of solutions as series in independent variable, 274
  of solutions to boundary value problems via Green functions, 332, 336, 337, 343, 345, 347, 348
  of vector in terms of basis, 133
residual magnetization, 501
residual (plastic) strain, 749
residue, 245–247
  of analytic function at isolated singular point, 245
resistance, 489, 520, 865
resistivity, 488, 489, 492
resistor, 490, 520, 866
resolved system, 126
resolved unknown, 126
resolving power, 550
  of optical instruments, 549
resonance, 519
resonance curves, 519
resonance frequency, 519
resonance operation mode, 912
resonance particles, 645
resonance phenomena, 912
resonance under arbitrary periodic action, 519
resonant absorption, 630
restoring force, 514
resultant force, 411
resultant moment of system of forces about center, 688
resultant of system of forces, 687
resultant vector
  of forces of inertia, 717
  of system of forces, 688
reversed statement of Blasius problem, 798
reversible process, 442
Reynolds equations, 782, 783
Reynolds number, 662, 780, 796, 824
  for film flow, 786
Reynolds stress, 783
rhombus, 44
Riccati equation, 267, 1012
Richardson–Dashman formula, 604
Riemann function, 340
Riemann integral, 183
Riemann sum, 183
right-angled triangle, 37
right arm of hyperbola, 72
right circular cone, 57, 97
right circular cylinder, 55
right cylinder, 55
right-hand limit, 144
right-handed trihedral, 83
right prism, 51
right rectangular Cartesian coordinate system, 61
right triangle, 37, 41
rigid body, 408
rigid clamping, 695
rigidity, 745
rigidity of rod cross-section under tension or compression, 752
rod flexibility, 770
roentgen (röntgen), 631
Rolle theorem, 153
root mean square, 11
root mean square deviation, 390
root mean square velocity, 455
root of equation, 109
root of multiplicity m, 109
rotation
  about fixed axis, 408, 429, 672
  about fixed point, 732
  about instantaneous axis, 408, 609, 679
  of liquid in cylindrical vessel, 774
  of polarization plane, 554
  of polarization plane, magnetic, 555
  of rigid body about fixed axis, 672, 732
  per unit length, 555
rotational levels, 594
rotational motion, 408
rotational spectra, 594
rotational tensor, 812
rotator, 595
Rouché theorem, 243
row vector, 114
rule
  Cramer, 124
  Hund, 589
  Kirchhoff first, 491
  Kirchhoff second, 491
  L'Hospital, 153
  Lenz, 503
  Maxwell, 460
  of parallelogram, 78
  of product, 379
  of sum, 379
  sigma, 394
  Simpson, 195
  trapezoidal, 195
  triangle, 78, 113
rules
  for alignment of mechanical constraint reaction forces, 694
  for calculating higher-order derivatives, 155
  of combinatorics, 379
  of differentiation, 152
  of rectangle, 195
  selection, 577, 583, 586, 589
Runge–Kutta method for Cauchy problem, 291

Runge–Kutta method of fourth-order approximation, 271
running angle of torsion, 753
Russell–Saunders coupling, 588
Rutherford experiments, 569
Rutherford formula, 570
Rydberg constant, 569, 571, 572
Rydberg correction, 584

S

Saint-Venant strength condition, 767
sample space, 380
Sarrus scheme, 113
saturated vapor, 458
saturation magnetization, 501
saturation property, 622, 624
scalar, 131
scalar product, 82, 136
scalar square, 82
scalar triple product of three vectors, 84
scalar triple product of vectors, 84
scale of electromagnetic waves, 532
scalene ellipsoid, 96
scalene triangle, 37
scattering, 422, 563, 565, 632
  elastic, 465, 632
  light, 557
  light, Rayleigh, 557
  particle, 422, 569, 619
  phonon, 608
  photon, 564, 565
  Tyndall in turbid media, 557
scattering angle, 422, 557
scattering center, 465
scattering coefficient, 557
scattering cross-section, 531, 570
schlicht mapping, 232
Schmidt number, 824
Schrödinger equation, 574
  time-dependent, 574
  time-independent, 574
  three-dimensional, 574, 579
secant, 20, 47
second-order constant coefficient linear equation (ODE), 272, 1017
second-order curve, 75
second-order determinant, 112
second-order differential operators, 170
second-order linear differential equations, 271
second-order linear equations, 1016
second-order nonlinear differential equations, 281
second-order nonlinear equations, 1028
second-order phase transitions, 462
second boundary value problem, 276, 313, 344
second convergence criterion, 212
second cosmic velocity, 427
second differentials, 164
second law of photoeffect, 562
second law of thermodynamics, 446
second noteworthy limit, 144
second partial derivatives, 164
second problem for unbranched magnetic circuits, 897
second problem of dynamics, 702
secondary spin quantum number, 585
sector, 48
secular terms, 286
sedimentation
  of colloids, 800
  of isotropic particles, 814
  of nonisotropic particles, 814
segment, 48
selection rules, 577, 583, 586, 589
self-adjoint matrix, 114
self-consistent field, 586, 599
self-induction, 504
self-induction electromotive force, 504
semi-invariant of quadric, 100
semiaxis of ellipsoid, 96
semiclassical approximation, 598
semiconductor
  of n-type, 610
  of p-type, 611
  pure, 610
semimajor axis of ellipse, 70
semimetal, 609
semiminor axis of ellipse, 70
separable equation, 1011
separable solution, 317
separation constant, 321
sequence of n independent trials, 386
sequence of trials, 384
series, 211
series connection
  of capacitors, 480
  of emf sources, 481
  of nonlinear elements, 888
  of resistors, 491, 877
series solutions of boundary value problems for parabolic equations, 323
sets on plane and in space, 162
settling rate of particle, 816
shape coefficient, 858
shape factor, 848, 856, 858
shear flow, 782
  axisymmetric, 940–942
  linear, 782, 856, 862
  simple, 782, 783
shear modulus, 749, 750
shear rate, 782
shear stress, 754, 760, 792, 833
shear tensor, 782
shell model of nucleus, 624
Sherwood number, 824, 825, 842, 847, 850–862
shock wave front propagation in nuclear explosion, 659
shooting method, for boundary value problems, 291
short circuit, 866
short circuit current, 874, 875
short circuit mode, 873

SI (International System of Units)
  base units, 1038
  derived units, 1038
  prefixes, 1037
sigma rule, 394
signature of quadratic form, 129
similar processes, 664
similarity, 664
similarity parameters, 664
simple (normal) Zeeman effect, 583, 591
simple capacitor, 479
simple circuit segment, 489
simple harmonic oscillations, 511
simple percentage, 6, 7
simple pole, 244
simple shear, 749
simple shear flow, 782, 783
simple thermodynamic system, 442
simple zero, 243
simplest properties of nonhomogeneous linear equations, 319
simplest summation formulas for power series, 223
simplification of boundary conditions, 276
Simpson line, 40
Simpson rule, 195
simulation, 664
sine, 20, 21
sine integral, 358
single-phase harmonic current circuits, 900
single-step methods of second-order approximation, 271
single-valued function, 140, 232
singular solution, 262
sinusoid, sine curve, 21
sinusoidal current, 900
skew lines, 49
skew-symmetric matrix, 114
skewness, 390
slope-intercept equation of straight line, 64
slowing down X-ray radiation, 564
small discriminant, 76
small oscillations
  of conservative system, 725
  of mechanical systems, 724
soft magnetic materials, 894
Sokhotski theorem, 244
solenoidal electric field, 502
solid angle, 51
solid geometry, 49
solids formed by revolution of lines, 55
solution
  Blasius, 797, 798, 831
  Cardano, 105
  classical, 315
  Descartes–Euler, 107
  Ferrari, 108
  formal, 315
  fundamental, 320
  Hadamard–Rybczynski, 807
  multiplicative separable, 317
  of eigenvalue problems, 322
  of first-order PDE, 305
  of Maxwell equations, in form of delayed potentials, 507
  of Stokes equations, 801
  separable, 317
  singular, 262
  trivial, 127
solution of linear problems
  for parabolic equations using Laplace transform, 330
  of mathematical physics by Fourier transform, 331
  using Laplace transform, 329
solution of problems
  for elliptic equations, 326
  for hyperbolic equations, 324
  for Laplace equation in polar coordinate system, 344
  for parabolic equations, 321
  of statics, 693
solution of Cauchy problem using Laplace transform, 294
solution of Goursat problem in terms of Riemann function, 341
Sommerfeld theory, 601
sound branches, 608
source energy power, 538
source luminous intensity, 538
space-like interval, 436
space of elementary events, 380
spacecraft propulsion, 415
spacetime interval, 436
Spalding implicit interpolation formula, 799
spatial coherence, 542
spatial inversion, 648
spatial parity, 648
special functions, 355
special relativity, 434
special Riccati equation, 1013
special theory of relativity, 434
specific activity, 625
specific binding energy, 620
specific excess surface energies, 785
specific Gibbs free energy, 452
specific heat, 833, 842
specific heat capacity, 445
specific latent heat, 462
specific phase transition heat, 462
specific refraction, 557
specific volume, 462
spectra of alkali metals, 584
spectra of complex atoms, 586, 588
spectral decomposition of light, 538
spectral series, 569, 571
spectrum of matrix, 120
sphere, 58
spherical bubble, 800
  at any Peclet numbers, 854, 855
  in translational flow at various Reynolds numbers, 808, 809
  in translational Stokes flow, 806–808

spherical bubble (*continued*)
  mass and heat transfer in translational flow, 853–855
  transient mass transfer in steady-state flow, 860–862
spherical cap, 58
spherical capacitor, 479
spherical coordinates, 81, 1009
spherical diverging waves, 522
spherical drop
  at any Peclet numbers, 855, 862
  at high Peclet numbers, 856, 861
  in translational flow, 808
  in translational flow at various Reynolds numbers, 808, 809
  in translational Stokes flow, 806–808
  mass and heat transfer in translational flow, 853–855
  transient mass transfer in steady-state flow, 860–862
spherical functions, 582
spherical mirror, 537
spherical motion, 679
spherical particle
  at arbitrary Peclet numbers, 862
  at high Peclet numbers, 861
  at various Reynolds numbers, 853
  in translational flow, 802–804
  in translational flow in wide range of Reynolds numbers, 804
  in translational Stokes flow, 802
  mass and heat transfer in translational flow, 853–855
  transient mass transfer in steady-state flow, 860–862
spherical parts, 58
spherical refracting surface, 537
spherical sector, 59
spherical segment, 59
spherical zone, 59
sphericity parameter, 814
spheroid, 96
spin, 500, 582, 585, 588, 600
  of boson, 600
  of fermion, 600
  of electron, 585, 587, 593
  of nucleus, 619
  of neutron, 617
  of neutrino, 629
  of proton, 617
  of particle, 649
spin-orbit interaction (coupling), 585, 624
spin quantum number, 582, 585
spontaneous emission, 566
spontaneous fission, 637
spontaneous transition, 565
square matrix, 114
square system of linear algebraic equations, 123
stability
  of equilibrium, 419
  of compressed rods, 768, 771
  of floating bodies, 777
standard deviation, 390
standard form of Fourier transform, 255
standard theory of electroweak interaction, 650
standing wave, 526
static resistance of nonlinear element, 884
static friction force, 700
static loop, 502
static moment of cross-section cut-off part about axis, 761
statically determinate problem, 692
statically indeterminate problem, 692
statics, 690
stationary constraints, 718
stationary mass and heat exchange, 847
stationary points, 167
stationary state, 574, 575
  of nonequilibrium system, 443
statistical approach, 441
statistical definition of probability, 381
statistical distributions, 597
statistical meaning of entropy, 450
statistical model of decay, 637
statistical weight, 450, 598
statistics
  Bose–Einstein, 599
  Fermi–Dirac, 599
  Maxwell–Boltzmann, 599
steady-state motion
  of bubbles in fluid, 807
  of drops in fluid, 807
  of spherical particles in fluid, 803
steady-state oscillations, 518
Stefan–Boltzmann law, 559
steradian, 538
Stirling formula, 360
stirring liquid, 663
Stokes equations, 800
Stokes formula, 803
Stokes law, 803
Stokes theorem (curl theorem), 208
stopping voltage, 562
STR dynamics, 437
straight line, 49, 64
strain state, 747
strains, 747
  at point, 746
  in symmetric lateral bending, 760
  in symmetric pure bending, 758
  in tension or compression, 752
  in torsion, 753
strangeness, 649
stream function, 797, 801, 803, 807, 810
strength, 745
  in symmetric bending, 761
  in tension and compression, 752
  in torsion, 755
  of electric field, 470, 471
  of gravitational field, 426
  of gravitational field generated by point mass, 426
  of materials, 745

strength condition, 752, 755, 762, 764
  beam, 762
  Saint-Venant, 767
  von Mises, 767
stress, 523, 746
  admissible, 752
stress concentration, 751
stress state, 746
  biaxial, 748
  in torsion, 753
  linear, 748
  plane, 748, 767
  simple shear, 749
  triaxial, 748
  uniaxial, 748, 750
  volumetric, 748
stress–strain state at point, 748
stress vector, 746, 748
stresses, 746
  at point, 746
  critical, 770, 771
  in bending with torsion, 766, 767
  in eccentric tension/compression, 765, 766
  in oblique bending, 763, 764
  in symmetric lateral bending, 760
  in symmetric pure bending, 758
  in tension or compression, 753
  in torsion, 755
  maximum, 761
  normal, 746, 748, 749, 759
  principal, 748
  Reynolds, 783
  shear, 746, 748, 754, 760, 764
  tangential, 746, 749, 754
  tensile, 749
stretched string, 523
strictly concave function, 149
strictly convex function, 148
strong interaction, 643
strong law of large numbers, 400
Strouhal number, 805
structure of complex atoms, 586
structure elements of network, 868
structure of general solution (Cauchy problem), 301
Struve function, 260
Sturm–Liouville problem, 277, 278, 334–336, 338, 339, 343
subdiagonal matrix, 114
sublimation–deposition curve, 461
subsequence, 141
subspaces of linear spaces, 134
subtraction
  of complex numbers, 229
  of inverse hyperbolic functions, 34
  of inverse trigonometric functions, 28
  of matrices, 115
  of power series, 221
  of trigonometric functions, 23
sufficient conditions
  for convergence of improper integrals, 191
  of extremum, 167
sum
  of complex numbers, 229
  of elements of linear space, 131
  of matrices, 114
  of polynomials, 103
  of series, 211
  of subspaces, 134
  of two sinusoidal quantities, 901
  of vectors, 78
summation formulas for power series, 221, 223
supercooled vapor, 458
superdiagonal matrix, 114
superheated liquid, 458
superposition principle, 297, 425, 471, 473, 494, 521, 573
  for field strength, 471
  for fields, 425, 471, 474, 494
  for magnetic induction, 494
  for homogeneous system, 298
  for wave functions, 573
  for waves, 521
sure event, 380
surface density of charge, 469
surface energy, 462
surface integral of first kind, 207
surface integral of second kind, 208
surface tension, 462, 623, 809
surface tension coefficient, 462–464, 784, 809
surface tension force, 463
surfactants, 808
switch, 866
Sylvester criterion, 128
Sylvester inequalities, 118
Sylvester theorem, 121, 123
symbols of order, O and o, 146
symmetric bodies, 698
symmetric matrix, 114
symmetric random variable, 387
symmetrical bending, 755
symmetry effect, 621, 623
system
  in thermostat, 451
  of forces, 687, 690–693, 713, 716, 745
  of linear algebraic equations, 123
  of linear differential equations, with constant coefficients, 298
  of Maxwell equations in vacuum, 507
  of Maxwell equations in medium, 508
  of parallel forces, 693
  of particles, 413
  of two equations with two unknown quantities, 124
  of two random variables, 395
  of units of measurement, 655
system of first-order linear homogeneous equations
  general solution, 298
  particular solutions, 299

T

table
  of basic integrals, 172
  of derivatives of basic elementary functions, 151
  of higher-order derivatives of some elementary functions, 155
  of integral transforms, 259
  periodic (Mendeleev's), 587, 1047
tables
  of definite integrals, 963
  of Fourier cosine transforms, 995
  of Fourier sine transforms, 1000
  of inverse Laplace transforms, 982
  of Laplace transforms, 975
  of physical constants, 1040–1042
  of units of measurement, 1038–1040
tangent, 20, 21, 46
tangent line, 151
tangential (shear) stresses, 746, 748, 754, 760, 764, 791
tangential acceleration, 406, 668
Taylor formula, 156
Taylor power series, 220
Taylor series, 220, 241
telegraph equation, 317
temperature, 441, 443, 445, 447, 454
  boiling, 462
  bulk, 837, 845–847, 936–939
  condensation, 600
  critical, 458, 459, 463, 600
  Curie, 485, 502
  Debye, 606
  Fermi, 603
  friction, 833
  mean flow rate, 839, 840, 843
  melting, 461, 462
  of ideal gas, 454
temperature coefficient of resistivity, 489
temperature field, 823, 831
  far from inlet cross-section, 838
  in initial region of tube, 839
temperature profile, 832–834, 841–843
tesla, 492
test charged particle, 471
theorem
  Abel, 215, 241
  Abel, on convergence of power series, 218
  Abel, on convergence radius of power series, 218
  Abel–Ruffini, 109
  absolute value, 187
  absolute value, for double integral, 196
  absolute value, for triple integral, 201
  Bernoulli, 400
  binomial, 9
  Bloch, 611
  Bolzano–Cauchy, 141
  Carnot, 446, 743
  Cauchy, 153, 215
  Cauchy residue, 246
  Cauchy, for multiply connected domain, 239
  Cauchy, for simply connected domain, 239
  Cayley–Hamilton, 121
  central limit, 401
  Chebyshev, 400
  convolution, 252, 256
  convolution, for Fourier transform, 256
  convolution, for Laplace transform, 252
  current compensation, 877
  Descartes, 111
  Dirichlet, on Fourier series, 223, 224
  divergence, 208
  Earnshaw, 476
  equipartition, 455, 560, 561
  estimation, for double integral, 196
  estimation, for triple integral, 201
  Euler, 51
  fundamental, of algebra, 109
  Gauss, 474, 475
  Green, 209
  Huygens–Steiner, 430
  Huygens, on moments of inertia, 705
  de Moivre–Laplace, 385
  König, 416
  Kolmogorov, 400
  Kronecker–Capelli, 123
  Lagrange, 153, 724
  Lindeberg central limit, 401
  Liouville, 234, 244
  local de Moivre–Laplace, 385
  Lyapunov central limit, 401
  mean value, 187
  mean value, for double integral, 196
  mean value, for integration, first, 187
  mean value, for integration, second, 187
  mean value, for triple integral, 201
  Mertens, 215
  Morera, 239, 241
  mutuality, 505
  Nernst, 451
  of principal axes of inertia, 432
  on accelerations in compound motion of point, 682
  on accelerations of points of plane figure, 677
  on active impedor (two-terminal network), 875
  on angular momentum, 710, 711
  on arrangement of two lines in space, 93
  on basic minor, 118
  on equivalence of couples of forces, 690
  on equivalence of systems of forces, 691
  on kinetic energy, 712
  on kinetic energy in differential form, 715
  on kinetic energy in integral form, 715
  on momentum, 708
  on momentum in differential form, 709
  on momentum in integral form, 709
  on motion of center of mass, 707
  on parallel transfer of force, 691
  on projections, 38
  on reducing system of forces to center, 691
  on resultant couple of forces, 690

theorem (*continued*)
  on variation in kinetic energy of system on impact, 743
  on variation in moment of momentum of point on impact, 739
  on velocities in compound motion of point, 682
  on velocities of points of plane figure, 676
  Ostrogradsky–Gauss, 208
  π, 656
  Poisson, 400
  polynomial remainder, 104
  probability multiplication, 383
  Ptolemy, 43
  Pythagorean, 42, 136, 659
  Rolle, 153
  Rouché, 243
  Sokhotski, 244
  Stokes (curl theorem), 208
  Sylvester, 121, 123
  uniqueness, 243
  Varignon, on moment of resultant, 693
  Viète, 105, 106, 110
  voltage compensation, 876
  work-energy, 416
theorems
  about differentiable functions, 153
  compensation, 876
  limit, 255, 398, 399
  of dynamics of mechanical system, 703
  of statics, basic, 690
  on number of real roots of polynomials, 111
  on residues, basic, 246
  on sliding vectors, 686
theory
  Debye, 605
  Einstein, 604
  of alpha decay, 627
  of finite Fermi systems, 623
  of gyroscope, 737
  of impact, 738
  Sommerfeld, 601
thermal boundary layer, 831–833, 843
thermal boundary layer equation, 831
thermal boundary layer thickness, 833
thermal conduction layer, 833, 841
thermal conduction of metals, 603
thermal conductivity, 464
thermal diffusivity, 823, 835, 842, 936
thermal efficiency, 444
thermal equation of state, 443
thermal initial region, 838, 839
thermal Peclet numbers, 824, 831
thermal pressure coefficient, 442
thermal radiation, 558
  equilibrium, 558
thermal volume expansion coefficient, 442
thermodynamic approach, 441
thermodynamic cycle, 444
thermodynamic parameters, 441
thermodynamic potentials, 451, 452
thermodynamic temperature scale, 447
thermodynamics of surface layer, 463
thermonuclear energy, 620, 637, 641
  of sun and stars, 641
thickness of diffusion boundary layer, 830
thickness of hydrodynamic boundary layer
  in laminar flow, 798
  in turbulent flow, 799
thickness of thermal boundary layer, 833
thin circular disk (hydrodynamic drag), 811, 813, 857
thin lens (focal distance), 537
third boundary value problem, 276, 314
third law of thermodynamics, 451
third-order determinant, 112
Thomson atomic model, 569
Thomson formula, 517
three equilibrium conditions, 461
three-dimensional Poisson equation, 320
three-dimensional quantum oscillator, 579
three-dimensional Schrödinger equation, 579
three-parameter empirical formulas, 931
three-phase system, 461
three subdomains in wall region, 793
threshold of nuclear reaction, 635
throat ellipse, 97
throttling process, 444
time dilation, 435
time-dependent Schrödinger equation, 574
time-independent Schrödinger equation, 574
time-like interval, 436
time of atom deexcitation, 531
time of coherence, 513, 542
tokamak, 642
topographic diagram, 911
torque, 420
torque of couple of forces, 689
Torricelli formula, 781
torsion, 753–755, 766–768
torsional rigidity, 755
torus, 59
total derivative, 164
total diffusion flux, 824, 828, 830, 857
total heat flux, 835
total increment of function, 163
total power, 909
total probability formula, 383
total reflection of light, 535
trace of matrix, 117
trace of square matrix, 117
trajectory (path), 405, 669
transcendental branch point, 245
transfer phenomena in gases, 464
transform, 251
  Bochner, 260
  Euler, of first kind, 260
  Euler, of second kind, 260
  Fourier, 255, 260, 331
  Fourier, inverse, 256, 331
  Fourier cosine, 257, 260
  Fourier sine, 255, 257, 260
  Gauss, 260

transform (*continued*)
  Hankel, 260
  Hardy, 260
  Hartley, 260
  Hilbert, 260
  inverse, 251
  Kontorovich–Lebedev, 260
  Laplace, 251, 252, 260, 328
  Laplace, inverse, 251, 328
  Meijer (K-transform), 260
  Meler–Fock, 260
  Mellin, 258, 260
  Mellin, inverse, 258
  two-sided Laplace, 260
  Weber, 260
  Weierstrass, 260
transforms
  of Fourier cosine, tables, 995
  of Fourier sine transforms, tables, 1000
  of integral, 251, 975
  of integral of general form, 251
  of inverse Laplace transforms, tables, 982
  of Laplace transforms, tables, 975
transformation
  calibration, 507
  calibration, for vector potential, 498
  Jordan, 126
  Kummer, 367
  of charges and currents, 509
  of fields, 509
  of quadratic curves to canonical form, 75
transformations
  equivalent, in electrical networks, 877
  equivalent, of characteristics of nonlinear elements, 886
  Galileo, 412
  Lorentz, 434
  nuclear, 625
  of coordinates, 135, 1008
  of graphs of functions, 159
  of real quadratic forms, 128
  of vectors, 1008
transient mass and heat exchange, 849
transient mass transfer to particles, drops, and bubbles in flow, 860
transient zone, 795
transition metals, 1047
translation tensor, 812
translational flow, 782, 800, 831
  past cylinder, 804, 805, 859
  past particles of arbitrary shape, 812, 856–859
  past spherical bubble, 806–808, 853–856, 861
  past spherical drop, 806–808, 853–856, 861
  past spherical particle, 803, 850–852, 861
translational motion, 408, 672, 684
translational Stokes flow, 852, 854
  past bodies of revolution, 811
  past ellipsoidal particles, 809, 857, 858
  past particles of arbitrary shape, 812, 857, 859
  past spherical bubble, 806, 853, 861
  past spherical drop, 806, 853, 861
  past spherical particle, 802, 851, 861
translationally moving noninertial frames, 428
translatory velocity, 409
transmission plane of polarizer, 551
transpose, 116
transuranic elements, 640
transversal drag, 811
  axisymmetric body, 813
  dumbbell-like particle, 812
  oblate ellipsoid, 812
  prolate ellipsoid, 812
  thin circular disk, 811
transverse (cross-axis) force, 746
transverse wave, 521, 523, 528
trapezoid, 45
trapezoidal rule, 195
traveling wave, 522, 604
traveling wave reflection, 527
triangle, 37
triangle inequality, 38, 136
triangle rule, 78, 113
triangular pyramid, 53
triaxial ellipsoid, 96
trigonometric circle, 19
trigonometric form of complex number, 230
trigonometric functions, 19
  of half argument, 25
  of multiple arguments, 24
trigonometric substitutions, 178
trihedral angle, 50
triple point, 461
triple integral, 200–202
tritium, 618
triton, 618
trivial solution, 127
truly neutral particle, 643
truncated cylinder, 56
truncated prism, 52
Tsiolkovsky equation, 415
tunnel effect, 577
turbulent (eddy) flux of admixture, 825
turbulent boundary layer, 783, 798, 799
turbulent core, 793, 794
turbulent diffusion coefficient, 825
turbulent diffusion tensor, 825
turbulent flow, 779, 782, 791–796, 799, 825, 832, 843, 844
  in plane channel, 795
  in tubes, 665, 793, 796
  Reynolds equations, 783
turbulent heat transfer in circular tube and plane channel, 841
turbulent Prandtl number, 825, 833
turbulent Schmidt number, 825
turbulent (eddy) viscosity, 783, 791–793, 798, 825, 826
turbulent viscosity coefficient, 826
turning points, 419
twin paradox, 436
twisting moment, 746, 753, 754, 766
two-body problem, 424
two-dimensional Poisson equation, 320

two-dimensional random variables, 395
two-dimensional random vector, 395
two-parameter empirical formulas, 926
two-phase coexistence curve, 461
two-phase system, 461
two-point linear interpolation, 926
two-sided Laplace transform, 260
Tyndall scattering in turbid media, 557
types of dielectric, 484
types of equations, 308
types of magnetic materials, 500
types of mechanical constraints, 694
types of stress state, 748

U

ultraviolet catastrophe, 561
Umov vector, 524
unbounded set, 162
unbranched electric circuit, 907
uncertainty principle (relation), 573, 578, 622, 643
unclosed system, 424
uncorrelated random variables, 398
undetermined system of linear algebraic equations, 123
uniaxial crystal, 553
uniaxial compression, 749, 759
uniaxial tension, 748, 749, 759
uniform body, 413
uniform convergence of function series, 216
uniform convergence of functions, 149
uniform distribution, 393
uniform field, 412
uniform motion, 406, 655, 670
  along given curvilinear path, 406
uniform rotation, 675
uniformity of time, 418
uniformly accelerated, 406
uniformly accelerated motion along given trajectory, 406
uniformly convergent series, 216
uniformly distributed load, 699
uniformly rotating reference frame, 428
uniformly varying motion, 670
uniformly varying rotation, 675
unilateral constraints, 718
unimodal distribution, 390
union of events, 380
uniqueness theorem, 243
unit vector, 78
unitary matrix, 117
units (of measurement), 655
  non-SI, 1039
  Planck, 1040
  SI base, 1038
  SI derived, 1038
unsaturated vapor, 459
upper bound, 139
upper limit of sequence, 143
upper triangular matrix, 114
use of particular solutions
  for construction of other solutions, 315
  for construct general solution, 267
useful power of current source, 490
utilization of particular solutions for reducing order of equation, 296

V

valence band, 610
Van der Waals equation, 442, 458, 459
Van der Waals isotherms, 460
Vandermonde determinant, 122
variable parameter, 64
variance, 389
  of random variable, 389
Varignon theorem on moment of resultant, 691
  of system of converging forces, 690
various forms of Fourier transform, 255
various statements of theorem
  on kinetic energy, 715
  on momentum, 709
Vavilov–Cherenkov radiation, 544
vector, 78, 114, 131
vector bosons, 643
vector, 78
  acceleration, 667
  of angular acceleration, 672
  of angular momentum, 580
  of angular velocity, 408
  of angular velocity of rigid body, 672
  of current on complex plane, 902
  of forces of inertia, resultant, 717
  of moment of couple of rotation, 685
  of normal acceleration, 679
  of system of forces, resultant, 688, 707, 716
  of tangential acceleration, 679
  of voltage difference, 912
  position (radius), 79, 667
  Poynting, 508
  Umov, 524
  unit, 78, 405
  velocity, 667
  zero (null), 78
vector components in Cartesian system, 1006
vector diagram, 512, 904–906, 911
vector method for specifying motion of point, 667
vector model, 581
vector potential, 498
vector space, 131
velocity, 406, 667
  absolute, 681
  angular, 407
  average, 406, 457, 488, 605, 792–800
  escape (first cosmic), 427
  frame, 681
  friction, 793, 799, 833
  group, 524, 573
  instantaneous, 406
  mean flow rate, 786, 788, 826, 840
  most probable, 456

velocity (*continued*)
  of electromagnetic waves in vacuum, 528
  orbital (second cosmic), 427
  phase, 524, 535, 544, 573
  relative, 425, 532, 681, 740
  root mean square, 455
  translatory, 409
velocity addition in special relativity, 436
velocity profile in circular tube, 793
velocity profile in turbulent boundary layer, 797
Verdet constant, 555
vertex, 37
  of cone, 57, 97
  of graph of electric network, 868
  of ellipse, 70
  of hyperbola, 72
  of parabola, 74
  of paraboloid, 97
  of polygon, 43
  of polyhedral angle, 50
  of polyhedron, 51
  of pyramid, 50
  of sector, 20
  of spherical sector, 59
  of triangle, 37
  of trihedral angle, 50
vertical asymptote, 145
Viète theorem, 105, 106, 110
vibrational levels, 594
vibrational spectra, 594
vibrations, 511, 604, 606, 608
virtual displacement, 719–724
  of point, 719
  of system, 719
virtual gluons, 644, 645
virtual particles, 622, 643
virtual pions, 622
virtual photons, 623
viscosity, 465, 467, 662, 780, 821
  dynamic, 657, 778, 802, 835
  effective, of suspensions, 816, 818, 819
  kinematic, 657, 663, 779, 792, 808, 824, 842
  turbulent (eddy), 783, 792, 793, 798, 825
viscous sublayer, 793–795, 799
voltage, 479, 487, 490, 865, 873
  capacitor, 479
  complex, 908
  effective, 904, 907
  magnetic, 893
  stopping, 562
voltage compensation principle, 876
voltage compensation theorem, 876
voltage resonance, 913
voltage triangle, 908
volume density
  of bound charge, 482
  of electric charge, 469
  of electric current, 488
  of electric energy, 486
  of energy of elastic waves, 524
  of energy of electromagnetic waves, 529
  of energy of equilibrium thermal radiation, 560
  of magnetic energy, 505
  of molecular current, 500
  of radiant energy, 537
  of thermal power of current, 489
  of wave energy, 524
volume hologram, 544
volume rate of flow, 784, 786–788
von Kármán constant, 784, 793
von Kármán friction law, 799
von Kármán model, 826
von Mises strength condition, 767
vorticity region, 804

W

wall region, 793, 799
Watson formula, 186
watt, 416, 538, 865, 1038
wave
  Bloch, 611
  circularly polarized, 528
  cylindrical, 522
  de Broglie, 573
  electromagnetic, 528
  elliptically polarized, 528
  harmonic, 522
  in air column, 523
  in elastic rod, 523
  in stretched string, 523
  longitudinal, 521
  monochromatic plane, 528
  plane, 521
  plane polarized, 528
  spherical, 522
  standing, 526
  surface, 522
  transverse, 521
wave attenuation, 557
wave attenuation coefficient, 557
wave equation, 308, 318, 338, 522, 523
wave front, 522, 546, 659
wave function, 573–578
  of electron in infinite periodic potential, 611
  of molecule, 592
  spherically symmetric, 621
wave intensity, 524
wave interference, 525
wave packet, 524, 542
wave-particle duality, 563, 573
wave path difference, 525
wave polarization plane, 528
wave pressure, 529, 564
wave processes, 521
wave properties of light, 539
wave radiation by moving charges, 530
wave resistance of circuit, 518
wave surface, 522
wave train, 541
wave vector, 522
wave zone, 530
wavelength, 522

wavenumber, 522
weak convergence, 399
weak interaction, 643
weak law of large numbers, 399
weber, 496, 1038
Weber number, 809
Weber transform, 260
wedge, 54
Weierstrass criterion of uniform convergence, 216
Weierstrass transform, 260
Weizsäcker formula, 624
well with finite walls, 575
well with infinite walls, 575
wetting, 464
wetting angle for wall material, 784
Wiedemann–Franz law, 492, 603
Wien displacement law, 559
Wien formula, 559
Wien law, 559
work, 415
  of Ampère force, 497, 502
  of central force, 416
  of electrostatic field, 472
  of external forces, 415
  of field, 417
  of force, 415, 712, 713
  of liberation, 604
  of Lorentz force, 492, 503
  of nonelectrostatic forces, 489
  of solenoidal electric field, 502
  elementary, 712
  mechanical, 415
  photoelectric, 563
work-energy theorem, 416
Wronskian, 364, 366
  of degenerate hypergeometric functions, 369
Wronskian determinant, 271, 277, 296
Wulf–Bragg condition, 551

Y

yield of nuclear reaction, 632
Young experiment, 540, 541
Young–Laplace equation, 463
Young–Laplace formula, 463
Young modulus, 412
Young relation, 784
Y-transform, 260

Z

Zeeman effect
  anomalous (complicated), 585, 590, 591
  normal (simple), 583, 591
zenithal angle, 81
zero matrix, 114
zero of function, 243
zero-point oscillations, 566, 572, 576
zero vector, 78
zeros of analytic functions, 243
zeros of Bessel functions, 365
zeros of Legendre polynomials, 373
Zhukovskii function, 235
Zhuravskii formula, 760
zone plate, 546

Π

π-mesons (pions), 622
π theorem, 656

Mathematics/Physics/Engineering Sciences

A CONCISE HANDBOOK OF MATHEMATICS, PHYSICS, AND ENGINEERING SCIENCES

- Presents basic notions, formulas, equations, problems, theorems, methods, and laws on each of the subjects in brief form
- First part covers arithmetic, elementary and analytic geometry, algebra, differential and integral calculus, functions of complex variables, integral transforms, ordinary and partial differential equations, special functions, and probability theory
- Second part covers molecular physics and thermodynamics, electricity and magnetism, oscillations and waves, optics, special relativity, quantum mechanics, atomic and nuclear physics, and elementary particles
- Third part covers dimensional analysis and similarity, mechanics of point masses and rigid bodies, strength of materials, hydrodynamics, mass and heat transfer, and electrical engineering, as well as methods for constructing empirical and engineering formulas
- Focuses on formulas, problems, methods, and laws that most frequently occur in scientific and engineering applications and university education
- Includes numerous examples, figures, graphs, and diagrams, with many results represented in tabular form
- Avoids, wherever possible, special terminology and excessively complicated details
- Provides a compact and clear presentation of the material, allowing readers to get quick help on the desired topic and to understand difficult issues
- Offers an extensive table of contents, a detailed index, special font highlighting in the text, and cross-references that help readers find the desired information

This self-contained handbook has no analogs in the literature due to its breadth of coverage and selection of material for engineers, researchers, and graduate and postgraduate students.

K10319
ISBN: 978-1-4398-0639-5
90000
9 781439 806395

CRC Press
Taylor & Francis Group
an Informa business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
270 Madison Avenue
New York, NY 10016
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

PHYSICS, AND ENGINEERING SCIENCES

10445

Haynes TECHBOOK

WELDING MANUAL

Basics of Gas, Arc, MIG, TIG, and Plasma Welding & Cutting

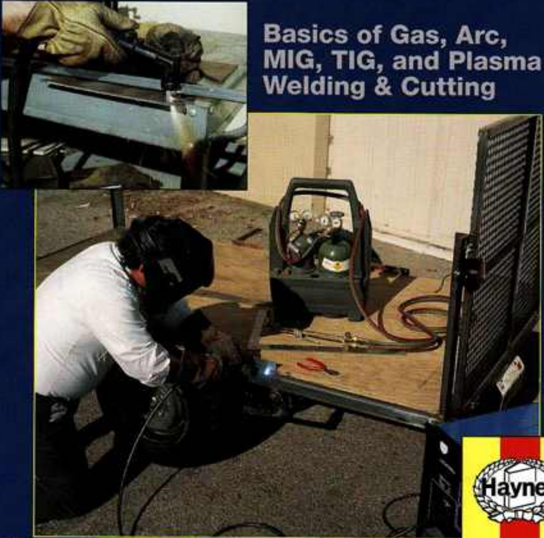

TECHBOOK

The Haynes Welding Manual

by Jay Storer
and John H Haynes
Member of the Guild of Motoring Writers

The Haynes Manual for selecting and using welding equipment

ABCDE
FGHU
KLMNO
PQRS

Haynes Publishing Group
Sparkford Nr Yeovil
Somerset BA22 7JJ England

Haynes North America, Inc
861 Lawrence Drive
Newbury Park
California 91320 USA

Haynes Welding Manual

Acknowledgements

We are grateful for the help and cooperation of welding equipment manufacturers such as Airco, Daytona MIG, HTP America, Lincoln Electric, Miller Electric and the ESAB Group, for their assistance with technical information and illustrations. Thanks to FitzMaurice-Smith Racing in Ventura for assistance with cover photography. We also wish to thank the Eastwood Company, Maguire's Welding and Valley Vintage Rods for assistance.

© Haynes North America, Inc. 1994

With permission from J.H. Haynes & Co. Ltd.

A book in the Haynes Automotive Repair Manual Series

Printed in the U.S.A.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage or retrieval system, without permission in writing from the copyright holder.

ISBN 1 56392110 3

Library of Congress Catalog Card Number 94-74485

While every attempt is made to ensure that the information in this manual is correct, no liability can be accepted by the authors or publishers for loss, damage or injury caused by any errors in, or omissions from, the information given.

94-176

Contents

Chapter 1 Introduction

Definition of welding 1-2
Development of modern welding 1-3
Welding today 1-6

Chapter 2 Types of welding

How it works 2-3
Metal alloys 2-3
Oxy-acetylene gas welding 2-6
Arc welding 2-9
MIG (wire-fed) welding 2-11
TIG (heli-arc) welding 2-14
Duty-cycles 2-17
Plasma-arc welding and cutting 2-18
Practice and training 2-21

Chapter 3 Oxy-acetylene gas welding/cutting

The basic gas process 3-1
The equipment 3-3
Getting started 3-7
Flame adjustment 3-9
Gas welding 3-9
Welding with filler rod 3-11
Brazing 3-16
Oxy-acetylene cutting 3-18
Heating with oxy-acetylene 3-24

Chapter 4 Arc welding

Comparing duty-cycles 4-3
AC, DC or both? 4-4
Rewiring for an arc welder 4-5
The arc process 4-6
Safety considerations 4-7
Beginning arc welding 4-8
Types of joints 4-11
Choosing electrodes 4-13

Chapter 5 MIG welding

Shopping for a MIG welder....5-5
Choosing shielding gas....5-12
Choosing wire....5-14
Learning MIG welding....5-16

Chapter 6 TIG welding

The equipment....6-4
The process in action....6-8
TIG-welding aluminum....6-13

Chapter 7 Plasma cutting/welding

Plasma-arc welding....7-2
Plasma-arc cutting....7-3
Choosing plasma cutting equipment....7-5
Using a plasma cutter....7-8

Chapter 8 Safety and shop equipment....s 1

Chapter 9 Building a utility trailer....9-1

Glossary....QL-I

Sourcelist....sn

Index....IND-1

1 Introduction

Welding is a process critical to our present state of civilization and technical advancement, yet little understood and most often taken for granted. Unless exposed to the building, machinery or automotive trades, the average person never realizes how much we depend on the welding process, which is a fundamental part of the process of building most of what we depend on daily, including vehicles, buildings, appliances, bridges and a great deal more. In fact, once you really start to examine the objects around us, it's hard to imagine our world without the welding process.

Architecturally speaking, we might all be living in one-room wood or adobe-brick houses if it weren't for welding. Certainly all large commercial and residential structures are built with a considerable "skeleton" of welded structural steel, and even most single-family, wood-framed houses are built using some welded components, even down to items like the electrical outlet boxes in the walls. Anyone who has watched the construction progress of a major highway improvement like a bridge or tunnel has seen the helmeted weldors, unsung heroes of the construction process, spraying a shower of sparks from high on a scaffold while they join metals to hold critical loads.

Sitting in an airport terminal recently with some time on our hands gave us something to think about. Virtually everything around us involved welding in some way. There was a large rack of telephone books in stainless-steel racks, each carefully TIG-welded and sanded, a post office box that was made of welded steel, the telephone stall had welded components, and the seats we were sitting on were part of a welded-steel structure that held eight seats. Everywhere you look in the modern world, you'll find examples of how widespread and important is the use of welding techniques and equipment.

Much of our Haynes audience is familiar with the automotive world, and here is a field where construction and repairs made by welding are absolutely essential, in fact essential to virtually all forms of transportation, from bicycles to cars, trucks, trains, aircraft and space vehicles. Even if we could go back before the "horseless carriage" was developed at the end of the 19th century, we would still need some form of welding to return to horse-drawn transportation, in welded brackets for harnesses and wagon components.

1.1 Welding had its ancient origins in the fires of blacksmiths, who could forge two white-hot pieces of metal together with hammer blows and patience. It remained for history to bring us to electricity and bottled gasses for welding techniques to develop any further. Modern-day farrier (blacksmith/horseshoer) Richard Heller illustrates the old ways.

1.2 Metal cutting in the old days was no easier. A white-hot piece of metal was laid over a hardy (like a wedge or chisel point) in the big anvil, and struck hard with a hammer, chopping the softened piece off. Final shaping of most items was by hand with stones and files.

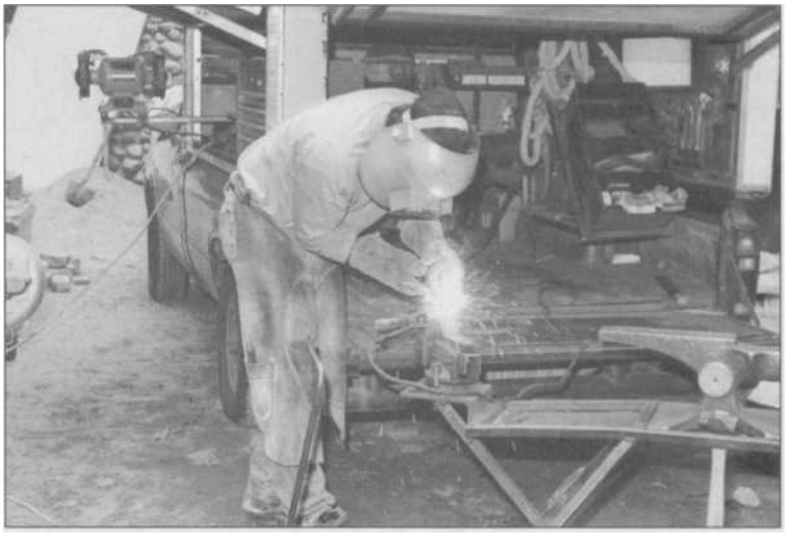

1.3 The farrier of today works out of a truck, filled with equipment like a grinder, drill press and welder, all running on AC power when available; otherwise, horseshoes are shaped in a portable forge. Today's blacksmith has the advantage of tools like this MIG welder that speed up making special horseshoes.

Since we've established that welding is a technical process our present society can't live without, in this book we'll explore more about the process, removing the mystery, and examining the available equipment as it can be useful to us in fabricating with metal or repairing metal items. We'll show all of the most up-to-date equipment, describe the various welding types, make recommendations, show how welding is used in everyday situations, and develop welding projects that illustrate how you can build or repair automotive, farm and household objects.

1.4 Under the center bridge column here are a welder (a machine) and a weldor (a person). He is welding up steel enclosures for the concrete columns, to retrofit more strength in California bridges to meet newer earthquake codes.

Definitions of welding

At one time, the simple definition of welding was "joining metals through heating them to a molten state and fusing them together." As technical progress in welding processes has advanced, the definition has had to change.

There are now two basic forms of welding: fusion and non-fusion. The former is the most common, and it involves the actual melting of the parent metals being joined. Not all welding today involves melting. Non-fusion welding is most commonly represented by soldering and brazing, two processes of joining metals where the parent metal is heated, but not melted, and a second or "filler" metal is melted between them, forming a strong bond when all are cooled.

Pressure and friction alone can weld metals together, such as when a machinist turns down a piece of metal in a lathe. Often, pieces of the metal chips can become welded to the cutting

1.5 Wherever you see buildings go up, the basis for them is always steel, whether they wind up with wood, bricks or cement on the *outside,* it all began with the work of ironworkers and weldors.

1.6 The construction and repair of heavy equipment and farm machinery would be nearly impossible without welding. Arc welding is usually used, as here modifying the front of a dirt scoop on a John Deere, where welding is outdoors and on heavy plates.

tool, which is a simple example of a process that can be used in production work in joining metals. Other kinds of "cold" welding may today involve sound and light, as in sonic-welding or laser-welding.

Today, the term "welding" has even been applied to the processes of joining non-metallic materials, such as plastic-welding which sometimes involves a fusion of materials as a result of heat or chemical action. As kids, we have all played with plastic models that we constructed using "glues" that would react with and actually "melt together" the two pieces we were joining.

For today's definition of welding to be all-encompassing, it would have to read "the joining of metals and plastics without the use of fasteners." This definition covers a lot of ground, but, given the interests and needs of the majority of our readers, this book will concentrate on welding as it applies to metals joined by heat processes produced either by a flame or electrical current.

Semantically speaking, throughout this book we will be illustrating and referring to pieces of welding equipment and to the people who operate them. Sometimes the same term "welder" is used to apply to both the machine and the man, which can become confusing, so for our purposes, we will from now on refer to the machine as a "welder" and to the person operating it as a "weldor."

Development of modern welding

Welding can trace its roots far back in time to the first blacksmiths who heated and shaped metals. At that time, metals were primarily used for tools and weapons, both of vital importance in those days. The blacksmith was an important tradesman in any community, earning a little more respect than most, even by lords and kings who depended on weapons for maintaining power. The "art" of smithing was understood by a select few, and blacksmiths were accorded almost the fear-based respect of a low-level sorcerer. Given the traditional image

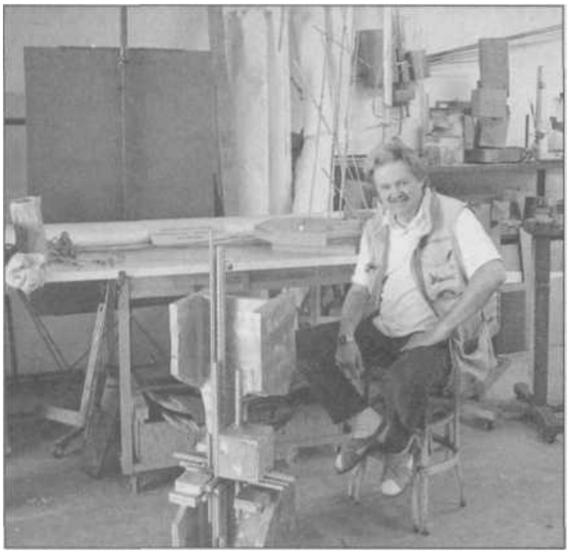

1.7 Welding is even used in many art forms today. Bill McKewen is a metal sculptor who works with rods, tubes, plates, wires and custom castings to create interesting works that combine hard metals with "organic" elements.

1.8 One of Bill's largest works is this 30-foot tall sculpture called "Organic Form 1", which is made entirely of stainless-steel and weighs over 1000 pounds.

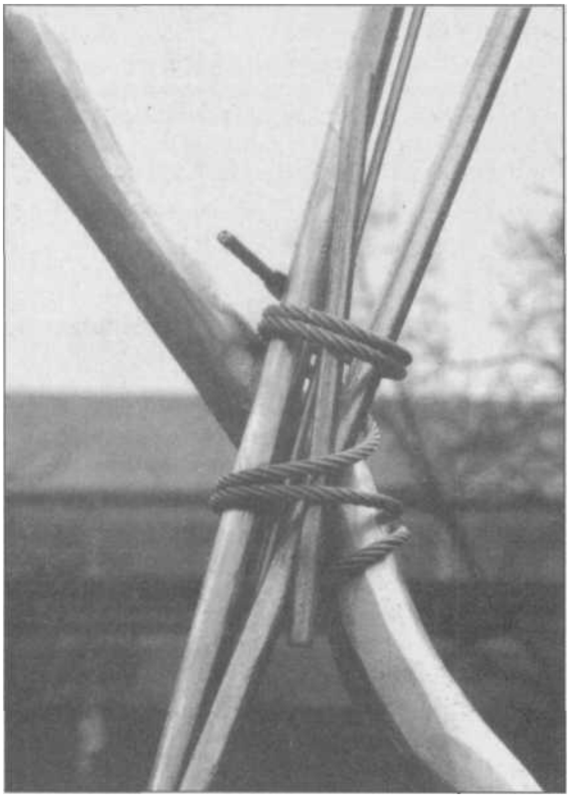

1.9 In a close-up of the "transition point" in Bill's sculpture, we can see how the stainless tubing and solid rod is wrapped with stainless-steel oil field cable and welded to a large shape originally carved of foam and then sandcast in stainless and final-shaped with a body grinder.

of the blacksmith as a large, muscular man covered with soot who works in a fiery, smoky environment, hammering loudly and magically making useful objects out of nothing, it's no wonder he was treated a little differently.

The blacksmith heated metals in a wood fire (coal was used later on) and hammered them into tools and weapons, performed basic heat-treating to harden some areas, and ground sharp edges with a foot-powered stone grinding wheel.

There were times when an object couldn't be made from one piece of metal. Gradually, the techniques developed to join pieces of metal either with bolts, hot rivets, or welding. This first use of welding consisted of heating the objects to a certain color (fairly precise indicator of the temperature) and quickly hammering them together on the anvil. The heat and pressure joined the items, and the process has been called "forge-welding."

As history marched on, larger and larger items had to be made of metal, especially with the industrial revolution of the 19th century. Most machinery was made of cast metal, produced when molten metal was poured into a mold and

1.10 In high-technology areas like nuclear vessels, aviation and here, at a large race car-building shop, quality welding technology is critical to success with metals.

1.11 Not every welder and weldor works indoors. This is an oil field service truck with gas-welding/cutting equipment, arc welder and crane. Many welding shops have portable setups like this for remote jobs.

allowed to solidify. Repairing broken castings was a common procedure in industry and manufacturing. Cast metal is too brittle for forge-welding on an anvil, and the items were too large, so a process of "cast-welding" was developed in which the broken machinery was heated, a temporary mold bolted around the area to be repaired, and molten metal was poured in. Done right, the molten metal bonded with the parent casting and the goldrush mine or cotton gin was back in business.

As the 20th century dawned, electricity came into wider usage, especially for lighting. The early carbon-arc lamps made as much heat as they did light, and someone started using electric carbon-arc rods to fusion-weld metals. Soon after, the simple stick electrode was developed that is similar to arc-welding rods of today, and about that same time oxy-acetylene gas welding was also developing. It's ironic that gas welding also grew out of the advancements in lighting, since acetylene gas had been used for car lights up until just before W.W.I.

1.12 The back of the same oil field welding truck is built as a large, heavy workbench-away-from-a-shop. Big Lincoln ac/dc arc welder has its own engine power and generator to run lights for night work.

Speaking of W.W.I, anyone who is familiar with history knows that many technological advances have resulted from the accelerated development that, unfortunately, only seems to come from a wartime environment. W.W.I saw the further development of gas and stick welding, and soon after that the refinement of X-ray technology aided further industrial use of welding. The exacting inspection of welds made possible by X-rays speeded welding's acceptance.

The rapid development of the aircraft industry between the wars, and the increased use of lightweight metals to replace wood and fabric in fuselages and wings led to further advances in welding. W.W.II saw the burgeoning aircraft industry with increased requirements for joining light metals such as aluminum and magnesium faster, stronger and smoother than with drilling and riveting, as had been used before. Inert-gas welding was invented before the war, but the gasses were considered too expensive at the time, and it took a war-time environment to accelerate its development. After the war, because the lightweight metals were in demand for many military and civilian applications, TIG welding was further refined, and MIG welding was invented around 1948.

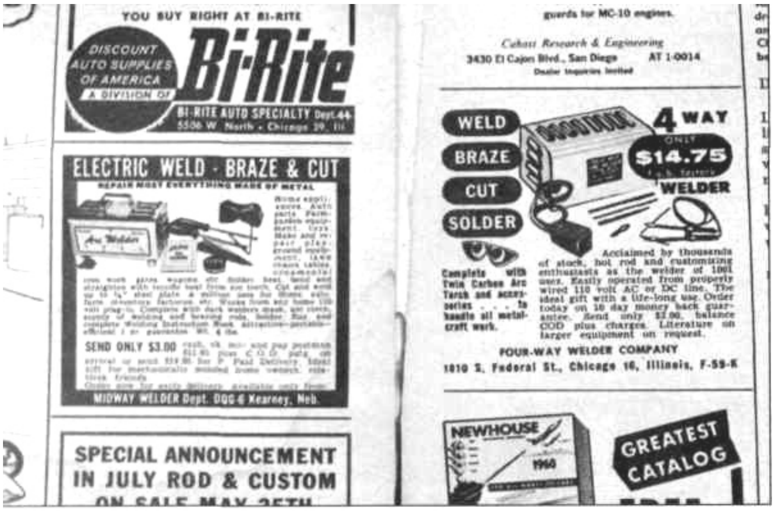
YOU BUY RIGHT AT BI-RITE

DISCOUNT AUTO SUPPLIES OF AMERICA

A DIVISION OF

Bi-Rite

BI-RITE AUTO SPECIALTY Dept.44
5506 W. North • Chicago 39, Ill.

ELECTRIC WELD - BRAZE & CUT

REPAIR MOST EVERYTHING MADE OF METAL

SEND ONLY $3.00

MIDWAY WELDER Dept. DGG-6 Kearney, Neb.

SPECIAL ANNOUNCEMENT
IN JULY ROD & CUSTOM
ON SALE MAY 25TH

guards for MC-10 engines.

Cahast Research & Engineering
3430 El Cajon Blvd., San Diego AT 1-0014
Dealer Inquiries Invited

WELD
BRAZE
CUT
SOLDER

4 WAY
ONLY $14.75
WELDER

Complete with Twin Carbon Arc Torch and accessories . . . to handle all metal-craft work.

Acclaimed by thousands of stock, hot rod and customizing enthusiasts as the welder of 1001 uses. Easily operated from properly wired 110 volt AC or DC line. The ideal gift with a life-long use. Order today on 10 day money back guarantee. Send only $2.00, balance COD plus charges. Literature on larger equipment on request.

FOUR-WAY WELDER COMPANY
1810 S. Federal St., Chicago 16, Illinois, F-59-K

NEWHOUSE
1960

GREATEST CATALOG

1.13 Back in the Fifties and Sixties, car enthusiast magazines were all advertising these "twin carbon-arc" welders, which were the original "Buzzboxes." While some of these did *work,* most of them were tried a few times and then put on a shelf out in the garage. Today they're collector's items from welding history.

SERVING INDUSTR
GENUINE HELI

1.15 If you decide to purchase a welder, visit your local welding supply center to talk to the salespeople and find out what equipment best suits your needs and budget. A well-equipped store will have everything you'll need, from gasses to glasses.

1.14 Today's modern home/shop/garage welding equipment is likely to be a clean, efficient, safer welding system like MIG, or wire-feed, welding. Machines like this are getting less expensive all the time, and some models can be hooked to standard household 110V current, making them quite portable.

Today

In the remaining decades of the 20th century, welding developments have come at a rapid pace, often closely tied to electronics development, as new and better methods of applying cleaner and more controlled heat have come along. In the past decade, we have seen the cost and complexity of welding equipment reduced in some areas, to the point where equipment that was once considered solely the province of the high-production professional shop is now found in garages and hobby shops around the country.

It's been good news for the automotive hobbyist, whether he is restoring an old car or building a race car, as well as good news for the small farmer trying to maintain and repair his equipment. Welders are now being purchased by metal-sculptors and other artists and craftsmen, as well as being used for the most everyday kinds of household jobs such as repairing a bicycle or garden tool, building a firewood storage rack or a barbecue, or even fabricating a small utility trailer at home.

Welding has been compared to playing a musical instrument. In the same way that anyone can pick-up a harmonica and start making sounds with it, most anyone with some mechanical aptitude can, with a little practice, start using popular kinds of modern welding equipment for non-critical jobs. But to make *music* with that harmonica, or good, strong, clean welds with a welder, will take time. As the hip musician once said when asked how to get to Carnegie Hall, "Practice, man, practice!" The more time spent consistently practicing with a welding setup, whatever the type, the better your results will be. If you use your welder only occasionally, don't just pick-up the torch and expect to lay a perfect bead the first time. There is a rhythm to find. Practice first on a scrap piece of metal of the same thickness as the work you intend to weld, and, when you have the rhythm down, repair or fabricate your job.

Though not considered with the same awe as the ancient blacksmith, a professional weldor today can still make a satisfying living, for it seems to be a skill that, far from fading out due to replacement technology, is seeing ever-increasing usage in business and industry. Weldors today find diversified employment in oil field and pipeline work, building construction, bridges and other infrastructure, automotive work from assembly-line welding to body shop repair to race car fabrication and antique auto restoration, to the nuclear-power industry, aviation and aerospace work, defense work, and manufacturing work building products from household appliances to huge boilers and construction equipment.

This book is intended as an overall introduction to the welding process, illustrating most of the common equipment and work techniques for both home and shop welding. While this is not a textbook for the would-be professional weldor, there is enough of an overview here to give a prospective welding student a basic understanding before delving into the more detailed professional textbooks on the subject.

Whether you plan to restore a vintage car, build a race car or experimental aircraft, construct your own wrought-iron fence, sculpt a metal art masterpiece, or go into welding professionally, you'll find this book of interest. The handy Source List and Glossary of Terms at the end of the book will be helpful for future reference.

Notes

2 Types of welding

If you are reading this book, chances are you have had some exposure to welding through watching a repair done such as having a new exhaust system put on your car, through some hobby interest in the metal arts/crafts area or through some industrial exposure to welding as used in manufacturing and building processes. Obviously you have become interested enough in learning about welding to purchase this book which we feel is an excellent introduction to a field where there are lots of involved textbooks for the person pursuing welding as a profession, but few *basic* books for someone getting started at the hobby, farm or home/shop level.

Perhaps your initial exposure to welding has sparked an interest in doing it yourself. If you are involved in automotive work, you already know how valuable the process can be in fabrication and repairs. Once you have seen it performed, you realize how *handy* this capability is. You can join pieces of metal to either repair something that was damaged and otherwise scheduled for replacement or build something entirely new, from a barbecue grille to a race car.

Once you have the basic skills and the right equipment, you'll find many more uses for welding than you had anticipated. Like a good truck or a specialized tool, once you have a welder, you'll wonder how you ever got along without it! You'll probably find yourself building a materials rack, stocking it with various sizes of tubing and plates, and actually looking for new projects to tackle, from building a workbench to last a lifetime, to storage racks, moveable shop carts, engine stands, shelving, and much more.

There are quite a few types of welding processes, and today there are a great many welders to choose from. There are so many, in fact, that just picking the process and the ma-

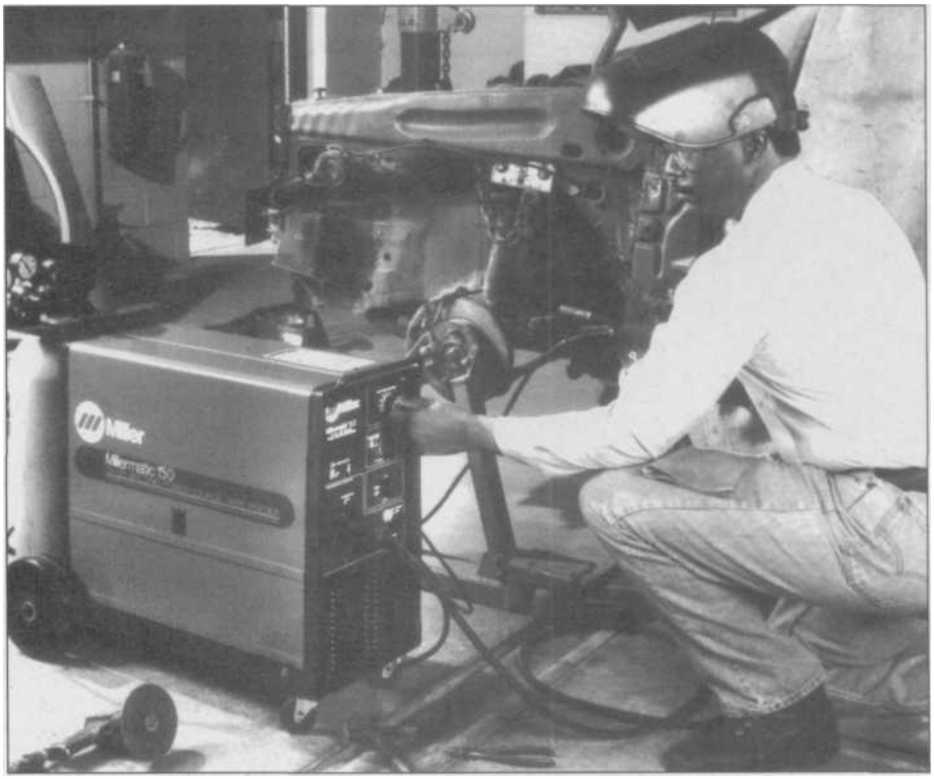

2.1 There are a number of choices today when shopping for welding equipment. The right choice for you depends on the kind of jobs you plan to do, where you plan to do them, your budget and how much time you can devote to training and practice before you become proficient at one of the methods we'll be looking at.

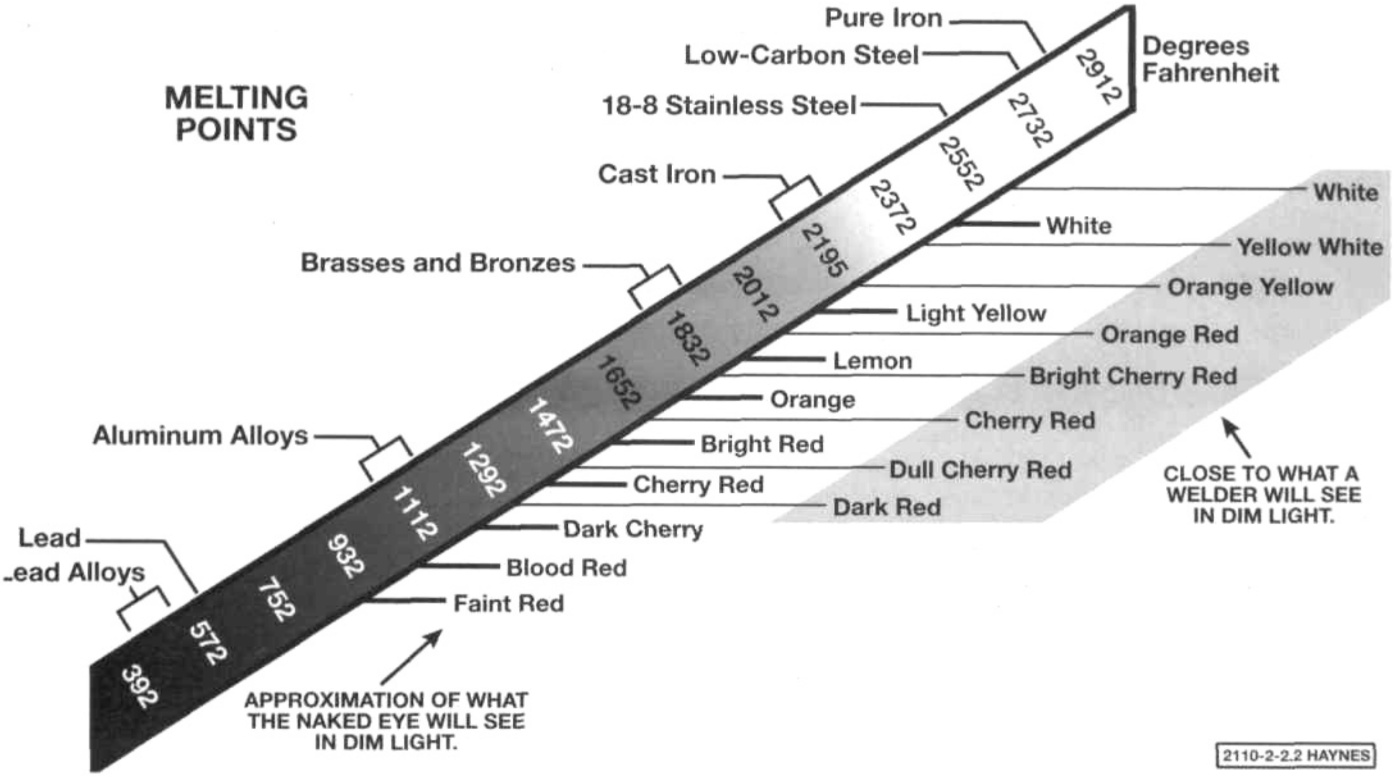

2.2 Welding is a process of *fusion,* in which metal parts are heated to the melting point and fused together, usually with a filler of the same material melted along with the "parent" metal. All metals melt at different temperatures, and this chart shows some interesting comparisons, as well as relating the heated metal's color to its temperature.

chinery which best suit your needs may seem a considerable task. This chapter will give you a brief overview of the various processes along with an analysis of the pros and cons for each type and how best to select equipment based on *your* needs.

How it works

The most basic principle of the welding process is joining two pieces of metal together (or at least two edges of the same piece, in the case of repairing a crack). This is generally accomplished by heating the metals to be joined until they become liquid or molten and the two edges fuse together. Most often, the complete joining of the two metal edges is accomplished by melting new metal into the joint at the same time. The new metal added to form a fused welding joint is called filler metal, while the original pieces being joined are called the parent metal. Together they form a welded "bead" of filler and parent metal that is usually thicker than the parent metal. Depending on the skill of the weldor and the type of welding, two pieces of metal can be joined in such a way that with a little filing or sanding of the bead, the joint is virtually undetectable, a particularly important aspect when making automotive body repairs. The first time you may have observed a professional weldor working, the process may have seemed like a sorcerer doing alchemy with a magic wand. With the proper equipment and practice, you can do a little magic yourself, a magic that can give tremendous personal satisfaction, as well as save you considerable expense compared to having the same work done at a professional fabrication shop.

2.3 The welding corner of this well-equipped race-car prep shop holds a variety of equipment, from a large TIG machine at left (with cooling unit on top), to a heavy-duty plasma cutter, oxy-acetylene cutting/heating/welding outfit, and two small plasma cutters. Not shown here are the two MIG welders in use elsewhere in the shop.

KLY SPECIAL
PER POUND
REMS
$3.00 LB

2.4 You will have to develop some new contacts once you get into welding, one of which will be a reliable local source for metals for your projects. This is ABC Metals (aluminum, brass and copper) in Oxnard, California. They have a lot of aluminum here at scrap prices, but it's more of a gold mine for the hobbyist or car-builder who needs relatively small bits and pieces and not a trainload. Knowledgeable salespeople at a metal yard can be very helpful.

It takes a tremendous amount of localized heat to weld metals together, and heat control is the key to welding properly. Every material has its own specific melting point, and to make a weld you need to heat the material to that point but not beyond it. Visualize an ice cube, which is solid material (when cold). If you heat it to the melting point (above 32 degrees F), the solid becomes a liquid (water), heating it further will vaporize it into steam and for your purposes the material is *gone.* The same changes happen to metal, although at much higher temperatures. Common lead solder such as you might use to solder electrical connections can melt at temperatures from 250-750° F (depending on the alloy), aluminum melts at just below 1250° F, and common mild steel melts at 2750° F.

The heat required to make metal molten enough to fusion-weld can be achieved in several ways, but the most common for home/shop situations will be generated either with a flame or some use of electrical current. The traditional source in welding has been the oxy-acetylene torch, while electricity is now used in most of the other methods, such as arc-welding, MIG-welding, and TIG-welding.

One thing that is common to all the forms of welding is that the filler material must be compatible with the parent metal, and all efforts must be made to make a "clean" weld free of outside contaminants that could weaken the joint. If you are welding aluminum, the filler rod must be aluminum, a stainless filler rod must be used for welding stainless-steel and steel rods are used on steel. In gas welding, the cleanliness of the weld is controlled by the correct adjustment of the torch flame and the cleanliness of the two edges of the parent metal. In electric welding, an inert gas "cloud" is formed right around the welding area that keeps outside oxygen or impurities from contaminating the weld. The shielding gas is generated in several ways, as you'll see as we further describe the various types of equipment.

Metal alloys

The melting point of the metal you work with will vary with the basic nature of the material (iron, steel, aluminum, magnesium, etc.), and the alloy of the metal. Most metals today are not in pure form, they are alloyed or mixed with another metal to give the new material special characteristics. Copper, lead and iron are basic pure metals that have been used by man for tools and other ob-

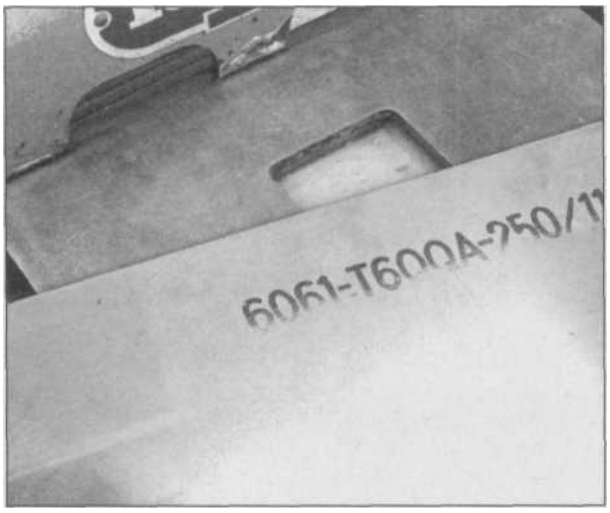

2.5 Non-ferrous metals like aluminum are generally marked with their alloy and heat-treat, such as here on this sheet of 6061-T6. If you need a specific metal for a project, you may have to buy new material rather than remnants, because the smaller scrap pieces may not have the markings on them.

2.6 Weldor Bill Maguire fabricated this recumbent bicycle for himself from 4130 chrome-moly thinwall tubing, which is very strong but light. The idea was to reduce the bike weight, wind resistance and pedal effort. Bill joined the thin tubing with TIG welding.

jects for thousands of years. Mixing various metals together can produce a new metal with new uses. Copper mixed with zinc will make brass, which has strength, reduced cost, and better suitability for machining and casting. The same base copper mixed with tin makes bronze, which was alloyed as far back as 2000 years to make weapons. Gold and silver, precious as they are in their pure state are seldom utilized in their natural form which is quite soft in comparison to other metals.

When alloyed with other metals which add strength or other characteristics, gold and silver can be used for jewelry, coins and many other uses. We commonly describe different gold objects by "carats." While pure gold is 24 carat, 12-carat gold is only half gold and half other metals, and the closer the carat-number is to 24, the more gold is in the object. The other alloys reduce the expense of the pure gold and make it more durable and useful. Were rings and other jewelry to be made of 24-carat pure gold, they would be too soft and not last in normal use.

Most of the metals you will be working with in your welding will be of two kinds, ferrous and non-ferrous. The former includes metals that contain iron, most commonly steel. The most commonly-welded non-ferrous metal is aluminum. Both steel and aluminum can vary considerably in the welding process depending on the alloy. By changing the alloy of either steel or aluminum, different properties can be obtained, to either make the metal more flexible (ability to bend without breaking), malleable (ability to be formed with a hammer), ductile (ability to be drawn out or hammered thin) or to improve its strength for a specific application. Steel is made from refined iron combined with carbon and other elements. How much carbon is added determines the properties of the steel alloy. Most of the steel we might use for projects is relatively low in carbon, called mild steel. With higher levels of carbon, you get medium-carbon steel (used for shafts and axles), high-carbon steel (used for automotive and industrial springs), while very-high-carbon steel is used to make files and sharp-edged cutting tools. The common mild steel we use most often is weldable by virtually all of the techniques described in this book, while the higher-carbon steels have special requirements. Other elements commonly alloyed with steel are manganese, tungsten, nickel and chromium. The latter two combine with steel to make

2.7 Get to know the personnel at your local welding supply store. They can be very helpful when it comes to choosing equipment, and they will have all the supplies you'll need in the future. Most stores carry several brands of equipment, all the safety items, and even the smaller home/shop machines.

2.8 The welding supply store should be able to tell you where to buy steel locally, which you will probably use for most of your welding projects. Most yards have a remnant section as shown, where you can buy short lengths of material at by-the-pound prices. Tubing for a trailer project is being weighed.

stainless-steel, a very useful material that requires somewhat different welding techniques. Anyone familiar with race-car and aircraft construction may have heard of 4130 chrome-moly steel, which is often used in these applications for its high strength relative to its weight. The four-digit number describes the alloy as containing molybdenum, and the amount of carbon. This alloy contains more carbon than mild steel, as well as chromium and molybdenum, both of which add properties of rust-resistance, strength and hardness. Even though this is a higher-carbon steel than mild steel, it really contains only 30/100th of 1% of carbon, which shows how scientific the alloying of metals really is. A tiny change in content can radically affect the properties of the final metal.

In steel and aluminum, not only are there different alloys, but different heat-treat processes. In the simplest terms, heat-treating is a scientific process of heating a metal to a specific temperature and then cooling it, either slowly or quickly, and with or without oil. The heat-treating can affect the hardness and other characteristics of the metal. When aluminum is purchased new in sheets or tubes, it is generally marked with its alloy and heat-treat, such as 3003-T3, which is a sheet aluminum that is considered "half-hard" and is commonly used in making race-car bodywork, where it has to be somewhat strong, but also able to be bent, welded and hammered. On the other end of the spectrum, 7075-T6 aluminum is very hard and strong. Called "aerospace aluminum" in the vernacular, it is often used in making machined aluminum parts and applications where very high strength is required. While it is strong and hard, it doesn't bend. There are volumes of scientific books on the alloying and heat-treating of metals, but, for your purposes as a home weldor, just remember to find out what kind of metal you are welding, and when making a project ask a metals expert to recommend the most suitable material. In general, the higher the carbon content in steel, and the higher the heat-treat on aluminum, the stronger the material will be but tougher to form into a shape, and the tougher metal alloys can be more brittle.

If you do decide to purchase a welder, you will eventually get to know the people at your local source for welding equipment, and they should be of considerable help in answering your questions and getting your setup working well. They will also be able to tell you where locally to purchase the metals you need for your projects. Most metal supply houses have a "scraps" section where you can purchase cut ends, small plates and short lengths of various tubes, all at about half the price of prime material, which is usually sold in large, unwieldy sheets or 20-foot lengths of tubing. Use the scrap pieces for all your experimentation with various types of welds, until you are doing pretty well. You may even be able to build small projects, like a welding cart, using the short lengths of tubing or angle-iron from the scraps section. Once you get to know the people at both the welding supply house and the metal yard, you'll have experts to wade through the technical jargon of alloys and heat-treats, and help you sort out the right material for your future projects.

Oxy-acetylene gas welding

This is perhaps the oldest and most versatile of welding setups. For a long time, it was the only setup recommended for the home/shop use, and has been among the least expensive to get started with. The basic combination in a typical gas-welding package are two high-pressure cylindrical tanks, one for oxygen, one for acetylene, a set of gauges and regulators to control the gas flow out of the tanks, a pair of hoses, and a torch. The torch usually comes with a variety of tips, tip cleaners, a spark lighter, and good sets may include a helmet, gloves and often a cutting-torch.

The latter is what really makes the oxy-acetylene system so versatile. It is one of the few welding systems that can do *cutting* as well as welding. This can be invaluable in both repair and fabrication work. Cutting away damaged or unwanted material is easily done with a properly-used cutting-torch attachment, and you may have many projects where you need to cut an *irregular* shape out of steel plate. The bulk of tubing and angle-iron cutting is usually done with some kind of saw or an abrasive cutoff wheel, but these tools can only make a straight cut; they can't go around corners. If you need to cut out a circle from a steel plate, you can draw the circle on the plate with a compass and a special hard crayon called a soapstone (which leaves a line you can see even when welding), then use your cutting torch to follow the line and you have your part. Any shape can be cut out. If you make up a cardboard template of the piece you need, trace around the pattern with your soapstone onto the plate, and make it.

Cutting with a torch takes skill to closely follow a line, and even then the edges of the metal will require some grinding, filing or sanding to get a smooth edge. Most experienced weldors know just how much to cut *outside* their pattern

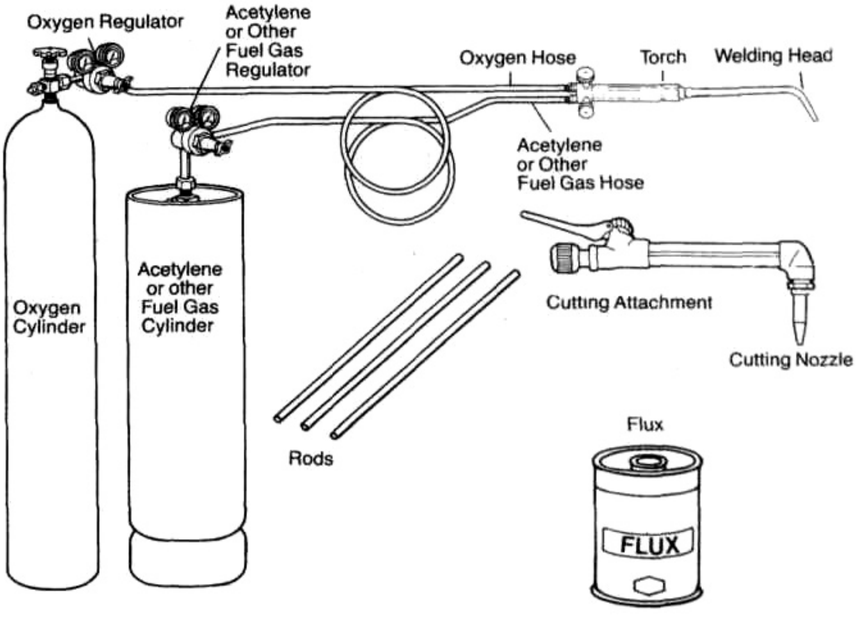
Oxygen Regulator
Acetylene or Other Fuel Gas Regulator
Oxygen Hose
Torch
Welding Head
Acetylene or Other Fuel Gas Hose
Acetylene or other Fuel Gas Cylinder
Oxygen Cylinder
Cutting Attachment
Cutting Nozzle
Rods
Flux
FLUX

2.9 Gas welding with oxygen and acetylene gasses is one of the oldest forms of welding, and is still used today in construction, muffler shops and farm repairs. Its versatility lies in the ability to cut *and* weld, on thick and thin materials, and to do braze-welding as well. The flux shown is only used for brazing. Note the different size and shape of the two gas cylinders.

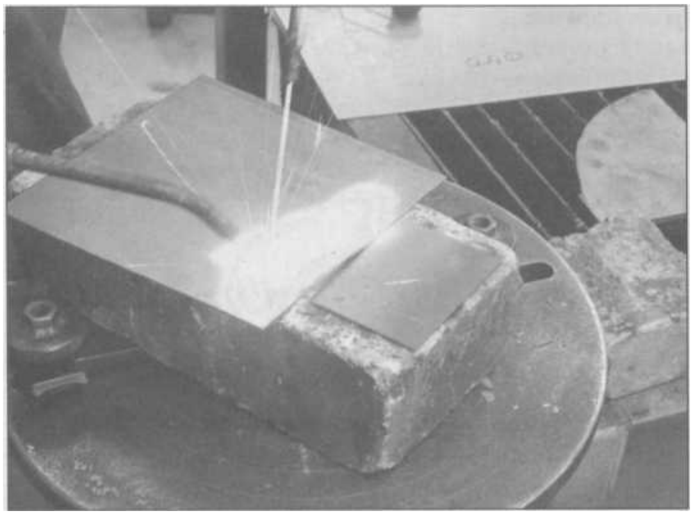

2.10 You'll need a safe area to practice your gas welding. Fire bricks like these are safe to weld on and don't suck the heat out of your parent metal as you weld. The skills you'll have to learn with gas welding are torch movement, even feeding of filler rod with your other hand, and steady, small circles to make consistent puddles.

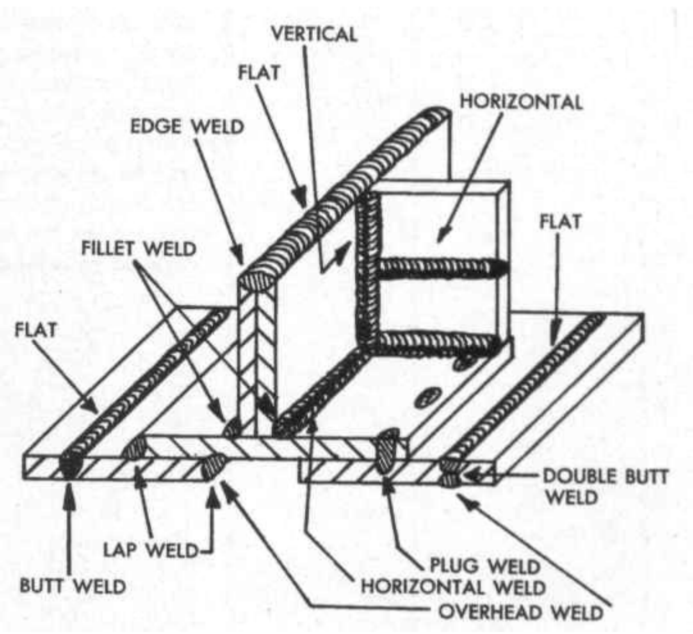

2.11 This illustration shows virtually all of the basic weld types and positions. With each type of welding you practice, you'll first learn to make a steady bead on a steel plate, then do flat butt-welds, then move on to corner welds and tougher ones like overhead and vertical welds. If you first learn torch control with gas welding, you'll be able to pick-up any other welding method more easily.

size to have an exact-size piece after cleaning up the cut edges with a grinder.

The oxy-acetylene setup is still the least expensive welding system if you buy the torches, gauges and hoses, and *lease* the gas cylinders. The tanks are expensive to purchase outright, but can be leased from your local welding supply store with a deposit down and a small monthly fee. Many shops today will take a credit application in lieu of a deposit, and you will open an account there, assuming your credit is good.

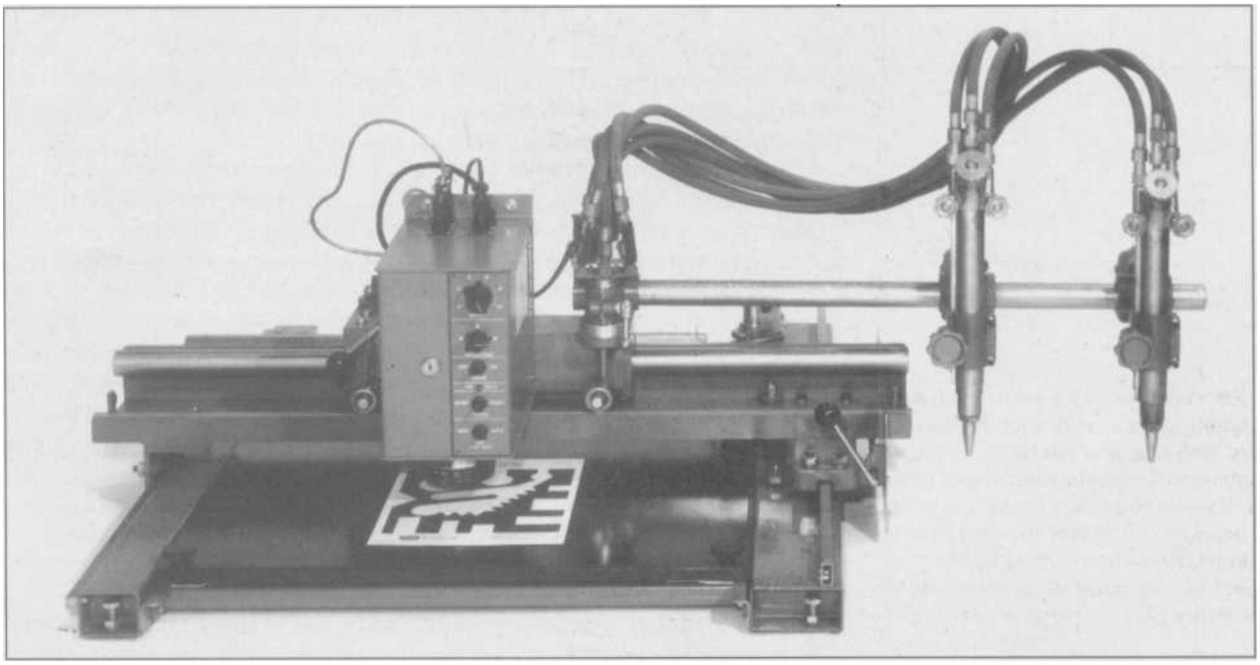

2.12 Oxy-acetylene torch flames can be put to some sophisticated uses. This is a mechanical gas pattern cutter (ESAB Silhouette 500), which can follow a pattern under the stylus and cut out two identical copies with the torch heads at right. Intricate parts can be cut out repeatedly, and with very clean edges because the torch movement is motorized and very smooth.

Leasing tanks is an inexpensive way to get started, but if you do only a small amount of welding infrequently it may not be a practical arrangement for you. After you have been paying rental fees for a year or two, you'll realize you could have purchased the tanks outright for the same amount of money. If you knew you only had to do some gas-welding for a few months, then the leasing deal would be better.

When welding with oxy-acetylene equipment, the basic procedure is to set the proper gas flow to the torch with the regulators, crack open the valves on the torch, light the flame with a friction-sparking lighter and then adjust the ratio of oxygen to acetylene to achieve the proper flame. Changing tip sizes makes a bigger or smaller flame, so you suit the tip size to the thickness of the parent metals you are welding. A smaller flame is used for thinner metals. The torch is brought down to the work area (the weldor is wearing his dark-lensed safety goggles) and the flame is used to heat the two edges to be joined, while your other hand feeds a piece of filler rod into the molten puddle as you move along the joint. Weld joints can be made with or without a filler rod, but a filler rod is used most often.

This is a very simplistic description of the process of gas-welding. It takes considerable practice and good hand/eye coordination to master. Once learned, the skills can be very useful, but it is a process rather difficult to learn from a book. Taking a class or having an experienced friend take you through the process will ease the learning curve considerably.

Besides the versatility of doing both welding and cutting, oxy-acetylene equipment also has many other shop uses in supplying a lot of localized heat. You may have projects where you need to bend a piece of metal. Thin sheet metal can be easily bent with pliers, vise-grips or put into a vise and bent over with a hammer, but thick metal may crack when bent cold. If you have a 1/4-in-thick steel bracket you need to bend at an angle, cold-bending with a hammer and vise may require so much hammer force as to distort the part out of shape, as well as mark the surface up or even damage the vise. If you closely examine the piece of thick plate after a cold bend, you may see the metal in the corner of the bend looking crystallized, which weakens that spot. Heating the metal to the right temperature with a torch before bending it, along the line where the bend should be, allows an easy bend with less disturbance of the metal's integrity along the bend. Another use of gas welding equipment is for brazing ferrous and non-ferrous metals such as copper and brass. In brazing, the parent metal is not made molten, it is heated enough to melt a brass filler rod, which attaches to both pieces of parent metal, making a firm joint.

Besides heating metal for bending, gas equipment is also used in many automotive shops for freeing frozen parts. Metals expand when heated, and when a rusted nut on a fastener is heated, the nut expands and breaks the bond, so the nut can be removed. When working under a car that has seen winter road salts, or when disassembling an old car for restoration, a gas torch can be very handy for getting off rusted nuts without busting your knuckles when your wrench rounds off an old nut. Many a mechanic on older cars would not be without his faithful "smoke wrench." Machine shops also use a torch to heat and expand parts that have a press fit. A gear that fits on a shaft may be heated with a special "rosebud" tip that spreads the flame around a wider area, and when the gear has expanded, it is picked up with tongs and slipped onto the unheated (and not expanded) shaft. When it cools off and contracts, the gear is securely fastened to the shaft, but can be removed at a later time just by reheating it again.

Despite the versatility of the gas-welding equipment, it may not be the ideal equipment for you, depending on your needs. If you have other uses for it besides just welding, then it is definitely a must have, but, if you just need to occasionally weld various thicknesses of steel together, some of today's electric welders may be more suitable for you. Gas welding is harder to learn, there are more safety problems in the shop when using a gas torch (especially when cutting), welding thick metals takes good skills and it is easy to distort the parent metal when welding thin sheet metal. Automotive body men today are using a torch less and less when doing repairs on thin metal. However, one of the advantages of gas-welding equipment over any electric-welder is portability. Your gas welding cart can be moved around anywhere without wires, even carried (when properly secured) in your pick-up for welding at a remote site.

If you have the budget and the need for gas-welding's versatility and portability, then your ideal shop setup would include both an oxy-acetylene rig *and* some kind of electric welder.

2.13 In farm and construction equipment repair, most work is done with oxy-acetylene cutting and arc, or stick, welding. Arc welding is not affected by wind outdoors, and is able to join or repair very thick materials.

Arc welding

Like gas welding, electric arc welding has been around for almost 100 years, and the fact that it is still around today illustrates its continued usefulness. The official acronym for arc welding is SMAW, which stands for Shielded Metal Arc Welding. The basic components of the setup include the machine (the power source), a ground lead you clamp to the work anywhere except where the weld is to be made, an electrode lead which runs from the machine to an electrode holder, which is a handle with a clamp that holds consumable electrodes. The electrodes are metal rods covered with a coating.

In use, the weldor strikes an arc against the parent metal with the electrode, which completes the circuit between the two leads and causes a bright light and concentrated heat. Arc welding uses considerable amperage of electricity to generate the intense arc, which melts the parent metal. The central metal core of the electrode melts as the work progresses, becoming the filler metal, while the fluxed coating produces a shielding gas around the welding area that protects the parent and filler metal from impurities in the air. Arc-welding produces *slag* as you proceed, a thick coating of impurities and deposits left from the rod's coating. This slag must be chipped off with a chipping hammer, which is usually included with the machine.

There are a wide variety of welding rods (electrodes) available to suit almost any purpose. The 12-14-inch-long rods are also called "sticks", and you may often hear arc-wetding referred to as *stick* welding. The rods vary in thickness, according to the thickness of the metal you are welding, and also in alloy and flux-coating content. There are many special-purpose rods, and, because of the variety, rods are usually marked with a number at the beginning of the flux coating, and different colors may also be added to the fluxes for quick identification.

Rods are usually sold in boxes or cans of fairly large quantity, which can be a problem for home/shop use where the welding is infrequent. The coatings on arc-welding rods are very susceptible to moisture in the air, and must be stored in very dry, secure containers to remain effective. You may have seen welding filler rods in gas-welding outfits stored in simple lengths of pipe welded to the welding cart, but this is not suitable for arc rods. If you do purchase an arc machine, invest in several airtight metal containers to store the rods, even using bags of desiccant (moisture-absorbing crystals usually found in small bags packed with cameras or sensitive electronic equipment) in the cans.

2.14 Two endeavors you wouldn't think of together, art and welding, combine today as many sculptors and artists are now working in metals. Small works are usually torch or TIG welded, with larger works like this one utilizing arc or MIG welding equipment.

There are two basic types of arc-welders, relating to the polarity of the electricity they produce, AC or DC. The DC machines are generally larger, industrial units found in production shops, where they are hard-wired in, or mounted in conjunction with an engine for truck-mounted use in mobile field welding. Most shop-type DC machines require shop-type electrical input, such as 440V or *three-phase* 220V, which you will not find in any standard home. They are designed to operate day in and day out without overheating, and are considered the best choice for welding really thick materials, so that is why we see DC arc machines in use building bridges, buildings, ships, etc. There are a few small DC machines for home use, but they have limited amperage and should be used on lighter materials. There are even combination AC/DC machines, but these are usually expensive shop machines.

By far, the most basic and practical home/shop arc-welder is an AC machine often called a "buzz-box" because of the sound it makes when you are welding. This is the least expensive single welding system you can buy, with a good name-brand buzz-box costing about the same as a set of oxy-acetylene torches. However, the arc machine comes with rods and everything you need, while the gas setup also requires filled cylinders which make a ready-to-weld gas setup about twice as expensive as a basic arc box.

The wiring you have in your house or shop will be a factor in choosing the type of equipment best suited for your purposes. To use an arc machine, you'll need 220V availability. If you already have an electric stove in your house (or the house had at least been wired for this) or an electric clothes dryer, you're probably in good shape because many electric stoves and dryers use 220V current. However, unless your dryer outlet is already out in your garage, you may have to have a qualified electrician run this 220V power out to where you'll be doing your welding. If you don't already have a 220V outlet, the cost of running new service to your garage may double the total expense of setting up to do arc-welding at home.

Also, welders do not just plug into the same outlet as your appliances. The welder has a different arrangement of prongs on its plug, and an adapter is required to plug into a 220V appliance outlet. The outlet you use should also have a good 20-30-amp circuit breaker as well. Note that there are some small buzz-boxes that plug into ordinary household 110V power, but they aren't recommended for anything but light and occasional work.

A stick-welder is relatively easy to use. Unlike gas welding, where you have to operate the torch with one hand while feeding the filler rod with the other hand, there is only one piece to control with arc-welding, the rod-holder. However, the rod starts out 12 to 14 inches long but is used up continually (gets shorter) as you weld, making it tough to maintain an exact distance of rod to workpiece, which is critical to a good arc weld. With the rod too close, you burn holes in the metal, and too far away you can lose the arc entirely and have to restart. So the trick in arc welding is control of the tip of the rod, and, because it is always getting shorter, you have to "fine-tune" your wrist movement in the hand working the electrode-holder. For this reason, most arc-welding is done with two hands, especially when learning. Use your other hand to steady and help control your wrist action on the "working" hand. Professional weldors use arc-welding upside down, laying on their back, hanging from scaffolding, or even underwater with special equipment, but for the novice, a comfortable body and hand position is very important to making good welds.

An AC arc-welder is well-suited to working on heavier steel materials, from 1/8-inch to 1/2-inch thick, but is difficult to control on thinner materials. The machine will have an amperage knob, with settings from 30-230 amps (depends on make and model), which you suit to the rod and the work material and thickness. Some of the more expensive shop machines have settings that go as low as 4 amps for light materials, but gas, MIG or TIG welding is- more popular today for thin metals such as most automotive work. For fabricating shop equipment, building a utility trailer, frame repairs or farm equipment maintenance, the buzz-box works fine.

To sum up the pros and cons of an AC arc-welder, the advantages include the low initial cost, easy operation (with practice), versatility (it can be used indoors or out) and it offers a high level of dependability (no moving parts) and quietness of operation. Disadvantages include: it's less practical for thinner metals, your shop may require rewiring to accommodate it, the home arc-welder usually can't be used for welding long seams all at one time, you are restricted to the limits of your power-cord length (as with any electric machine except the generator-driven type), the welds may have considerable spatter and not look as "clean" as other types if that is a consideration (such as in art projects, metal furniture design or street-rod fabricating), and there is the safety consideration of potential skin burns. Compared to gas welding, arc-welding poses more danger due to burns, not just from little spatters of hot metal, but from any skin that is exposed to the UV and infra-red rays. You can get a severe sunburn from exposure (pro weldors wear heavy leather protective clothing) and observing an arc-weld in progress without a helmet on, even for just a second or two, can cause headaches and eye irritation.

The buzz-box was once the most practical home/shop welder, but increasing availability and affordability of home MIG machines in the last ten years has put up a serious challenge to that title.

MIG (wire-feed) welders

This category of welding system has become one of the most popular for today's home/shop use. The initials stand for Metal Inert Gas, but is also listed in technical descriptions as GMAW, for Gas Metal Arc Welding. The basic elements of the setup include a power supply (machine), a torch with a large-diameter cable, a ground wire with clamp, and a bottle of compressed shielding gas. Inside the machine is a roll of relatively-thin wire and a motorized transport system for this wire.

In practice, you weld almost like arc-welding, but the electrode (wire) is constantly fed through the cable to the gun and consumed at the weld. When you pull the trigger of the MIG gun, you start the supply of amperage (when the arc starts on your work), the feeding of the wire electrode, and the flow of shielding gas, which is also routed inside the gun's cable and comes out of the tip all around the electrode, preserving the integrity of the weld like the flux coating does on arc rods.

The advantages of the MIG system includes a much cleaner weld than either gas or arc, good versatility in materials with the ability to do very well on thin metals, and there is no electrode or filler rod to keep replacing. Control is easy because you can set the amperage on the machine and also infinitely adjust the speed of the wire com-

2.15 Anyone who is a fan of auto racing is familiar with various types of welding. In the immaculate shop of Larry Smith Marketing, a crewman is putting a final MIG weld on a complete NASCAR chassis for a car like the Matco Tools entry at left. Race car shops use a variety of welding techniques, from gas to MIG and TIG.

DAYTONA MIG

2.16 Small, portable wire-feed, or MIG, welders have become one of the most popular welding machines for home/shop use. Easy to set up and simple to operate, they produce very clean welds, and models from 100-amp to 140-amp can be operated on household 110V current.

2.17 Most MIG machines operate with a bottle of shielding gas, but some are equipped to run with flux-cored wire which makes its own shielding gas as it arcs, making the process more suitable for outdoor and drafty situations. This home/shop-sized unit shown is a combination machine that can operate with gas shielding/solid wire, or the flux-core wire.

2.18 At the professional shop level, MIG welders are all 220V machines, and offer power up to 250 amps, which is enough to weld most anything. This Millermatic 250 is a popular machine in fabricating shops and automotive shops. There is even a model of this machine with a small computer in which you can store the welding parameters of up to nine different jobs, in five different languages.

ing out of the gun. Various-diameter steel wires are available to weld material from the thinnest sheet metal to 1/4-inch and 3/8-inch, as well as stainless-steel and aluminum wires to weld those materials with. The weld from a MIG gun is much cleaner than arc welding in terms of spatter, and there is no slag coating to chip off.

Originally, wire-feed MIG machines were developed strictly for the professional shop doing long-duration production welding. A roll of wire in one of these machines can last for an entire eight-hour shift without the weldor ever stopping for a new electrode. Because the wire is machine fed into the weld puddle, the heat and wire-feed rate can be tuned so that production welds can be made much faster than with other systems. The MIG machine has been an expensive piece of equipment in the past (and heavy-duty, big-shop models are still fairly expensive), so it never used to be recommended to anyone for home/shop use. However, less-expensive MIG machines have become widely available in the past 5-10 years that have made these an excellent choice for the hobbyist, particularly the automotive hobbyist. These newer machines, mostly imported and built with a few less features and a lower duty-cycle than would be required in a production environment, have not only become very popular with the hobbyists but are also are finding their way into more small fabricating shops, muffler shops and race-car fabrication shops as a less expensive alternative to the big machines, yet with better speed and weld quality than arc or gas.

MIG machines range in size and features from "pocket" machines to large, full-featured machines for shop environments that require 100% duty-cycle. There are a number of smaller machines available today that operate on 110V household current, which makes them very portable, and no garage rewiring is

2.19 If you're lucky, the welding store near you will have a demonstration area, such as here at Altair Gasses & Equipment, Oxnard, California. Here a portable MIG unit is being tested where customers can watch through a large safety lens (behind operator, at right). You can get a feel for different equipment before you decide what machine best suits your needs.

Rod Nationals
street is neat
INK-Credible
Advertising
4442 F

2.20 If you are interested in cars as a hobby, whether it be kit cars, street rods, race cars, or antique restorations, you should become familiar with various welding techniques, and perhaps buy a system. When fabricating parts at home, it's a lot easier if you have your own welder, even if all you do is tack parts together and bring them to someone else for final welding.

necessary. These smaller wire-feed machines usually have 4-7 heat range settings from 30 to 110 or 140 amps, and can handle light sheet metal such as auto body material up to metals as thick as 1/4-inch or 3/8-inch (less maximum thickness is possible on aluminum). They either have wheels on the bottom or have optional, wheeled carts that make then very easy to move around your shop or garage. Many beginning weldors like to build their own welding cart as their first project with a new welding machine.

The shielding gas used with MIG machines can be CO_2, Argon, or a mixture of the two, depending on the materials you are welding. The basic gas used most often is straight CO_2, because it is the least expensive. Various sized gas bottles are available from small, very portable 20-cubic-foot bottles (about two feet high), to 120-cubic-foot bottles (four foot high, six-inches in diameter) that can supply enough gas for eight hours of continuous welding. The small bottles are fine for occasional home/shop use where long seams are not welded regularly.

One of the factors that has made MIG welders so popular with home/shop users is the relatively easy learning curve associated with using them. Given a little instruction and practice, most people can be up and welding a decent bead in an hour or two. We're not suggesting that that person would then be ready for a full-time job as a weldor, or that his/her first welds would pass rigid specifications for nuclear reactors, but the wire-feed machines are easier to learn than most other systems. One of the factors that makes it easier is of course not having to deal with feeding the filler rod in with your other hand. In automotive body work this can be particularly helpful when you have to hold something with your left hand, like the alignment of two pieces of sheet metal, while you tack them together with the one-hand MIG torch. Torch position is very important in all forms of welding, and the fact that the distance of the MIG tip from the work is constant (in many cases the nozzle is actually touching the work to steady it) makes controlling the weld much easier.

After selecting the proper heat setting based on the thickness of the material

2.21 Typical home/shop projects such as a utility trailer, race-car trailer, metals rack, shop cart, engine stand, hoist and much more can be built at home with either MIG (shown) or an AC arc welder.

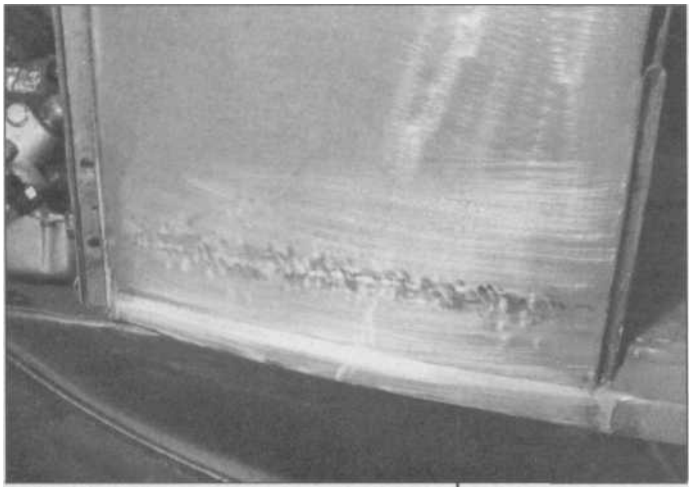

2.22 Street rodding and antique car restoration frequently require welding on old, even rusty sheet metal body panels. Here the lower portion of a '32 Ford cowl has been patched with new steel to replace rust-out area. The panel was MIG-welded in with a 110V welder, ground down with a body-sander, is ready for finishing and another 60 years on the road.

you are welding, the only other variable to control is the speed the wire is fed into the weld puddle. If the speed is wrong, there will be a lot of sputtering and uneven welding and penetration. Most MIG weldors tune the wire-speed in by *sound,* They set up on a scrap piece of metal with a wire speed that may be too slow or too fast, then pull a bead while adjusting the speed. When the right speed is reached, the MIG welder gives a distinctive crackling sound to the procedure. The steady crackling sound means you have the right wire speed.

There are three basic types of small MIG machines. Those that use plain wire with a bottle of shielding gas, those that use flux-coated wire and no gas bottle, and combination machines that can run either fluxed or plain wire. The flux on the wire functions the same as the coatings on arc rods, creating a shielding gas as the wire is consumed. The bottled-gas-shielded MIG welding is the preferred method for indoor welding where clean-looking weld beads are desired. However, the shielding gas around the weld process is subject to being disturbed by air currents, so for outdoor welding (such as farm equipment repair, fence construction, etc.) the fluxed-wire machine is better. The weld isn't quite as clean as with the gas, and fluxed wire doesn't do as good a job on thin sheet metal (under 18-gauge) but these machines are more practical for the outdoor work, which is usually on heavier materials anyway. Also, the fluxed-wire machine never need to have a bottle refilled, so as long as you have plenty of wire inside the machine, you can't have your work delayed by running out of shielding gas.

Considering all of these factors, the modern MIG machine may be the most practical type of machine for many do-it-yourselfers, particularly the automotive hobbyist, who has to make very clean welds, both with thin sheet metal as well as heavier chassis work. The ease of operation, maneuverability around the shop (with the 110V machines), one-hand torch operation and the gentle learning curve have combined to make many first-time MIG weldors tackle their projects with confidence.

TIG (heli-arc) welding

This is unquestionably the "Rolls-Royce" of the welding processes in many ways. The lettering stands for Tungsten Inert Gas, but many weldors refer to it as

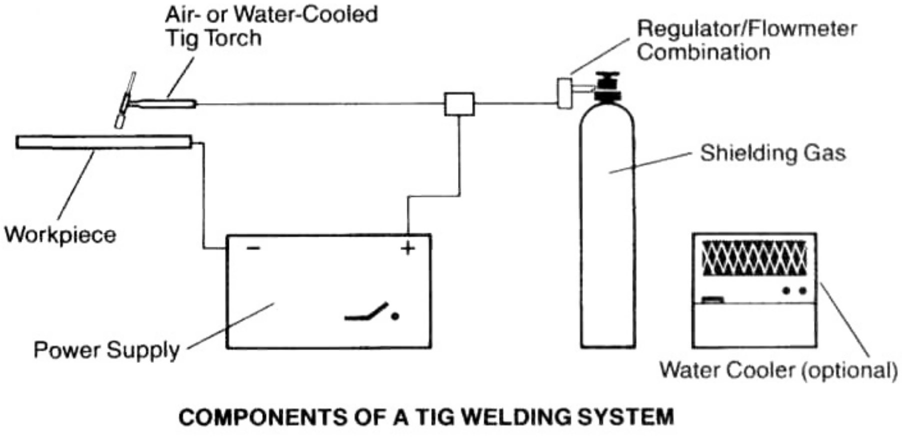
Air- or Water-Cooled Tig Torch
Regulator/Flowmeter Combination
Shielding Gas
Workpiece
−
+
Power Supply
Water Cooler (optional)
COMPONENTS OF A TIG WELDING SYSTEM

2.23 Basic components of a TIG-welding system. Very similar to operating a gas torch and using your other hand to feed filler rod, TIG produces the cleanest welds with the least amount of material warpage, and is the best system for thin, high-strength material like chrome-moly as well as aluminum.

2.24 While TIG or heli-arc welding is considered by many to be the most *controllable* welding technique, it takes more practice than any other. You have to learn to manipulate the torch with one hand, the rod with the other, and the amperage control with your foot on a pedal. There are no sparks, spatter or slag. Here an aluminum frypan's handle mount is being repaired. The heli-arc welding didn't even bother the non-stick coating on the other side!

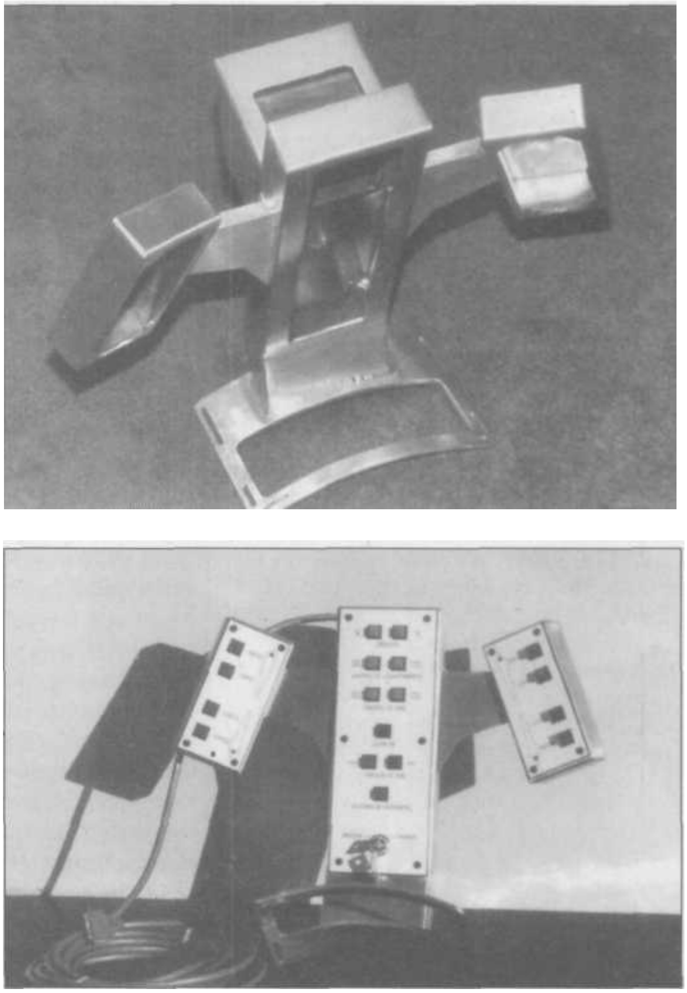

2.25 Where TIG welding really shines is in two tough situations in one job, like thin materials *and* aluminum, such as this part that looks like something from a space ship. This wound up being a belt-pack of controls for a big piece of machinery - the operator strapped this to his waist. Aluminum is tougher to weld than steel because it doesn't change color as you heat it. When it melts it turns shiny, but tarry a little longer with the torch and the material has holes in it. Don't tackle a project like this until you have mastered the process.

heli-arc welding, though this is a trade name established by the Linde Corporation many years ago. The name refers to the fact that back then, helium was used as the shielding gas for the process, though today argon is the most commonly used gas for TIG welding. Either shielding gas has the capacity to exclude foreign materials during the welding process, as well as being unable to combine with other elements to form chemical compounds detrimental to the weld integrity. The process can also been seen abbreviated as GTAW, for Gas Tungsten Arc Welding.

The basic components of the TIG process include a welding machine, a torch with a tungsten electrode, a ground cable, a foot-operated amperage control pedal and a bottle of compressed shielding gas such as argon. It differs basically from stick welding and MIG welding in that the electrode is not consumed, and filler rod is separately hand-fed into the weld puddle when needed. In operation, the machine is turned on, the weldor brings the torch close to the work and depresses the foot pedal to start the arc. Current flows through the tungsten to the workpiece, creating an intense but concentrated heat that melts the parent metal, with the operator controlling the amperage with the foot-pedal while welding, and feeding the appropriate filler rod to the puddle as needed. The tungsten electrode has a very high melting point and is not consumed.

What makes the heli-arc different in practice from other systems is the higher operator skill required, since the weldor has to coordinate with both hands and the foot pedal, all of which affects the quality of the finished weld. Unless you are already an accomplished weldor with an oxy-acetylene gas-welding torch, you'll find the TIG process takes a *lot* of practice.

TIG welding produces the cleanest, flux-free and spatter-free beads with virtually no contamination. For this reason, it is the method of choice for such critical applications as race cars, aircraft, missile construction and nuclear reactor welding. The TIG welds are strong, ductile and corrosion-resistant with virtually no cleanup required, making it popular for welding of surgical, medical and food-service equipment, especially on stainless-steel. There is little smoke produced during TIG welding and no flying sparks, and without the smoke and sparks the weldor can more critically observe the welding action as he goes.

While there is a lot of skill required in TIG welding, its ability to handle all kinds of metals of varying thicknesses, the ability to reach into corners and hard-to-access areas to weld and its outstanding ability to mate thin metals without putting excessive heat and distortion into a wide area on either side of the weld has made the heli-arc method particularly useful in race-car and other critical construction applications. The types of metals that can be welded with TIG machines include: all steel alloys, aluminum alloys, aluminum castings, stainless-steel, copper and nickel alloys, magnesium and exotic metals such as titanium and zirconium. It is one of the few systems that can tackle welding dissimilar metals and joining metals of different thicknesses. It's been said that a skilled TIG weldor could join a razor blade to a railroad track, though there isn't much call for that.

The torch for TIG welding is smaller and lighter than most other welding torches, allowing for greater weldor comfort when welding for long periods. For production welding, most TIG machines are equipped with a water-circulating system to keep the torch cool. In use, a water pump circulates filtered water from a storage tank out through the main cable to the torch head and then returns the heated water back to the tank. Many cooling units employ a small radiator and electric fan to cool the water. Where lots of production welding is done, a large water storage tank is used, or the water is fed in cold from the shop's plumbing and the heated water is plumbed back into the drain system, so heated water is never circulated back through the torch. The TIG machine requires shop-type current input (230V/460V/575V single-phase or 230V/460V three-phase) and connections to and from the water-cooling equipment, making it one of the least portable of welding systems.

Because of the electrical requirements and other hookups, the steep operator learning curve, and the considerable expense of the most shop-type TIG machines, this has never been a system that has penetrated much into the home/shop and hobbyist market. Rare and proud is the automotive hobbyist or racer who has a heli-arc machine at home, and who knows how to use it to its capacity. The process has assumed an almost "mystical" aura to automotive fans because of its use in race-car construction, and even otherwise skilled home weldors will bring certain components like critical steering gear to a professional to have the parts TIG-welded. Street rod builders love it for the quality of weld, which is often cosmetically appealing enough to require no grinding or filling to make the welded part ready for painting or even chrome-plating.

Shops with TIG machines rarely take advantage of this, but most TIG welding machines are capable of also performing stick welding if the job should require it. TIG welding is ideal for thinner materials, but if a job required joining some one-inch-thick plates, the stick-welding mode would be much faster. The arc capability would also come in handy if the job required some welding outdoors or near a drafty doorway, where the draft could blow the argon shielding gas away and cause an erratic weld. Switching to the arc mode simply requires a switch on the machine be thrown to the correct current position for arc, and the arc "torch" electrode cable plugged in. When in the arc mode, the foot control is deactivated and the weldor sets the amperage on the machine with a knob, like a straight arc machine.

While most production shop TIG machines are large, expensive, heavy-duty machines with water-cooling systems for the torch, there have been strides in the past ten years at making a more affordable TIG machine. Today there are several to choose from which are air-cooled instead of water-cooled, reducing cost and complexity, that feature straight 220V, single-phase input connections for simplified electrical hookup, and have lighter-duty, less-expensive internal electrical components. These new machine are one-half to one-third the cost of their bigger conventional TIG brothers.

Because of these factors, there are more small shops that today can afford the benefits of TIG welding, such as high-quality welds on thin materials, and the ability to weld beautifully on non-ferrous materials like aluminum, even cast aluminum. The are often seen in large engine rebuilding facilities and automotive machine shops, where they serve their time repairing cracks in aluminum blocks and heads, expensive parts that would otherwise have to be replaced if such repairs could not be made. However, these "economy" TIG machines can't do everything the big ones can, particularly when it comes to torch cooling. If long welds are to be made, the air-cooled torch will get too uncomfortable to hold. Also, the smaller, less-expensive electrical components inside will not have as long a duty-cycle as the big machines (more on duty-cycles later), but in a small shop fabricating race-car or street rod parts and doing precision chassis work, long, continuous welds are not often encountered.

Some welding machines today offer "convertible" adaptability. Kits can be added to some MIG machines to add heli-arc capability, and there are kits with a different torch and a separate wire-feeder to use the output of the bigger TIG machines as a MIG machine, but this convertibility usually puts the machine out of the home/shop price range.

Even with the introduction of this new generation of more-affordable TIG machines, this is still a system that isn't practical for the average home/shop weldor to get into. As with oxy-acetylene gas welding, the practice time it takes to be good with a heli-arc torch requires consistent, daily operation, and most occasional weldors just can't get enough time on the machine to really do well. Most basic home/shop fabrication and repair projects do not require the "reactor-quality" welds and superior cosmetic appearance of TIG welding, and even the economy TIG machines cost more than twice what other starting welder setups cost, making the gas welder and small MIG machines still the most versatile equipment for home use. A drawback of TIG welding is that it is a *slow* process, so for a project like building a utility trailer, a typical home MIG machine would be much faster.

Duty-cycles

In any discussion of electric welding processes, the term duty-cycle comes up and is a source of some confusion to the beginning weldor. Basically, the bigger, heavier and more expensive pro-shop type welding machines have transformers and other internal electrical components designed to operate for long periods of time without overheating. Smaller, home-type machines can't operate as long, because the internal components had to be made less expensive in order to reduce the overall costs to a "consumer" level.

The official quantifier of such work capability is called the duty-cycle. It is expressed as a percentage, such as a "40% duty-cycle." What this refers to is a ten-minute period in a laboratory welding test. At a 40% duty-cycle, a particular welding machine is capable of welding for *four minutes out of ten.* In other words, if you welded non-stop for four minutes straight, you would have to let the machine "idle" for six minutes before resuming welding, allowing the internal components to cool off. The maximum current that a welding machine can draw from its components is limited by the temperature rise of those parts, and the temperature goes up with the square of the current draw.

Bigger, more professional and more expensive welding machines will have higher duty-cycles because of the nature of the work they have to do. There are some cases where a robotic or machine-driven automatic welder must be capable of 100% duty-cycle because it is operated continuously, but this is unusual.

The advertised duty-cycle of a welding machine is *usually* given at the maximum current draw of the machine, the highest amp setting. This setting of course is required only when welding the thickest material the machine is capable of. As the amperage setting of a welding machine is reduced, the effective duty-cycle is *increased,* so that there is somewhere on the welding machine's "curve" a point where the duty-cycle is close to 100% at some amperage setting.

For the home/shop user who is rarely welding anything thicker than 1/4-inch or 3/8-inch mild steel, the duty-cycle is not a problem. However, it's important to know what the duty-cycle for your machine is at various amperages, and especially to understand the term when shopping for a machine. For instance, one popular introductory-level 110V MIG machine has an amperage range of 30-110 amps. Right away, you know that such a machine is not really designed for long welds on heavy materials, because the maximum amperage is only 110. Further, the advertised duty-cycle for this machine is 95% at the 30-amp setting, but 25% at the highest amp setting, showing the difference in how the current-draw heats up the components. Depending on your average welding needs, such a machine may be just fine, handling short welds occasionally on 1/4-inch steel, and longer periods on thinner material. Most automotive sheet metal, for instance, seldom requires an amperage setting above 30, and even the small MIG machines have a long duty-cycle at this setting.

Generally, welding machine price goes up with maximum amperage and duty-cycle. You can roughly compare the performance of the machines being considered by comparing their duty-cycle at the amperages you will most often be using. A feature to look for on some of the better machines is an "overheat" protection circuit and warning light, which may save you the cost of a burned-out component. Some machines automatically shut down when the overheat circuit kicks, and in others an internal electric fan comes on to speed the cooling process, but like most "bells and whistles" these features usually cost a little more.

Plasma arc welding and cutting

This is another welding/cutting process that has developed in the recent past that was once considered exotic, yet is now filtering down to lower price levels in the welding marketplace. The nomenclature refers to the plasma state of a gaseous material, in which the material is heated so much that the gas conducts electricity. What makes PAW (Plasma Arc Welding) so different is that the flow of superheated *gas* actually makes the welding arc, with the metal electrode

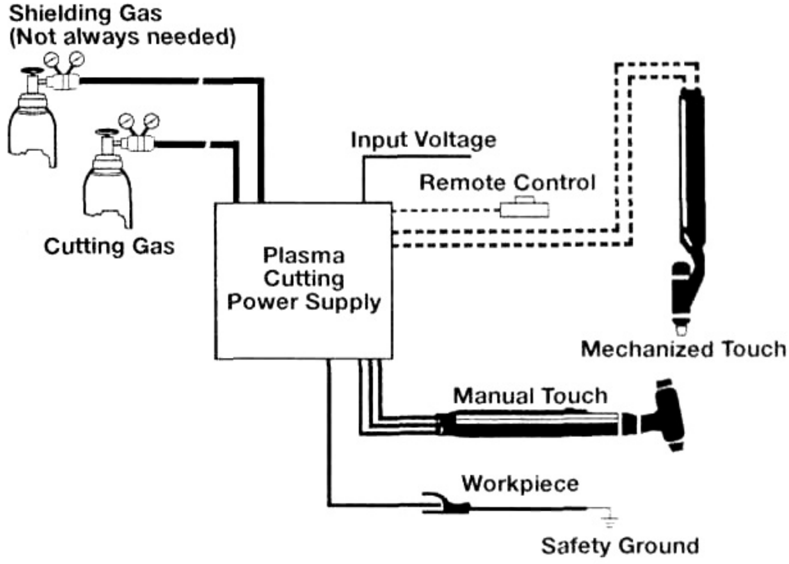

2.26 Plasma cutting is rapidly replacing oxy-acetylene cutting in many home and shop applications. Used either as a mechanized cutter or by hand, the system is simple to operate, and makes a faster, cleaner cut with much less sparking and less heat induced into the parent metal.

safely hidden way up inside the torch body away from the welding action.

The plasma welder makes a hotter and easier-to-control arc than even the much-touted TIG welder. The arc becomes a narrow column of heated gas (usually argon or nitrogen) that can be directed as precisely as a surgical laser. A second gas is plumbed through the torch to maintain a gas shield over the weld area, as is done in most electric welding. The extremely-high temperature of the plasma arc means that higher welding speeds can be used in production applications, and where there is a close fit of the two edges being welded, as is usually the case in a production situation with newly cut materials, often the plasma weld is accomplished with no addition of filler material when welding thinner stocks. This has made it a preferred system for some industrial applications.

In operation, the plasma equipment looks much like TIG equipment, except that there are two bottles of gas, and the torch is designed differently inside. The very dense, very hot "flame" does not require as precise an arc-starting-distance of the torch to the material, making plasma equipment somewhat easier to operate than TIG, especially on projects where no filler metal is needed. In TIG-welding, the weldor must strike an arc with his tungsten electrode With the plasma method, however, the electrode is too far up inside the torch head to get near the parent metal, so most of the plasma machines have high-frequency arc-starting circuitry that makes them easy to get up and running, using the hot gas column to start the arc, after which the proper torch-to-work distance is easily established by the weldor.

While there are specialized applications where plasma-arc welding is the preferred method, its widest area of usage today is in *cutting.* The column of plasma-stage-heated gas is hot enough to cut through virtually any material at almost any thickness. On thin materials, the plasma cutter goes through as fast as the operator can move the torch (speeds up to 100 inches per minute), and even on 1/2-inch-thick steel the moderate-size cutters can cut at 15 inches per minute, which is about one inch every four seconds!

2.27 In fabricating shops, a large 220V (or higher current, three-phase) plasma cutter is used, capable of cutting thick materials and traveling at fast cutting speeds. Most plasma cutters are used without shielding gas, using air from the shop air compressor to cool the torch and form the ionized gas column under the tip that creates the arc.

While the speed of the plasma cutter is exceptional, it is the quality of the cut that has made it so popular today in many applications like automotive, duct repair and fabrication, and medical or food-service equipment work. The plasma cutter will cut virtually any ferrous or non-ferrous material with equal ease, even on aluminum and stainless where ordinary torch cutting is difficult without considerable pre-cut and post-cut cleaning.

Since there need be no direct contact between the torch and the work material, plasma cutting doesn't involve a lot of spatter getting into the torch head, and most materials can be cut with no pre-cut cleaning. This is especially important to automotive users, who can make clean, fast cuts in sheet metal, whether it's painted, rusty or dirty. The concentration of the narrow cut and the speed the plasma torch can travel means that the cut edges are many times cleaner than even a machine-cut edge from an oxy-acetylene torch. There may be little or no cleanup of the cut edges required, depending on the job requirements, and the cut edges are not only smooth but uncontaminated, so that plasma-cut edges can generally be welded together much easier that edges cut with a gas torch where oxidation must be ground off before welding.

Plasma cutters have become particularly popular in machine-cutting operations where the torch head is mounted on a machine that has a pantograph arrangement. The weldor moves an arm around a wood, paper or metal template and the cutting torch cuts that exact shape out of metal. Sometimes, several cut-

DAYTONA MIG
POCKET Turbo 130
PLASMA

2.28 Where the really complete home/shop setup was once an AC arc welder and a set of gas torches, today more shops utilize a MIG welder and plasma cutter. This welding cart features a gas-shielded 130-amp MIG and a plasma cutter capable of cutting up to 3/16-inch material, and both machines operate on household 110V current.

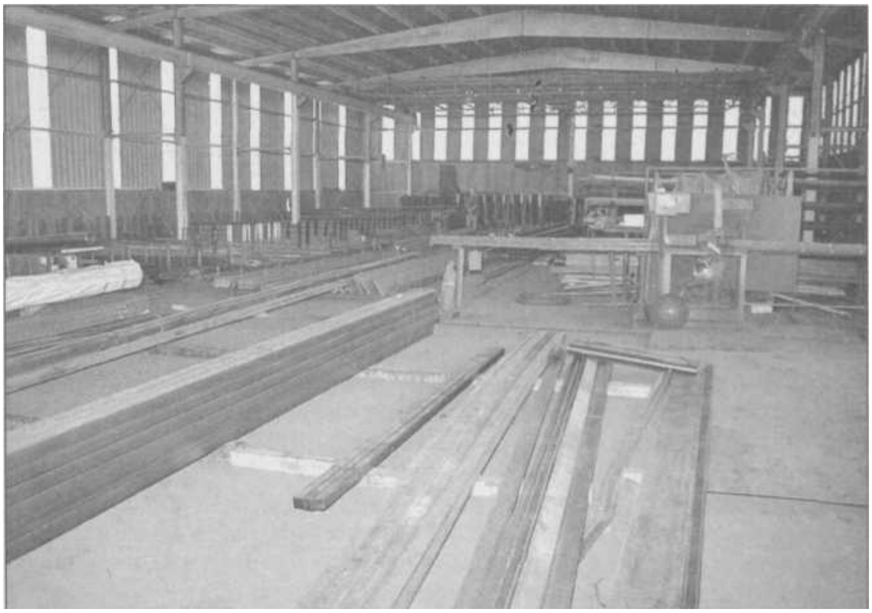

2.29 Eventually, you will find yourself at a steel supply store, shopping for plate, angle-iron or tubing for a project. Steel is usually sold in 20-foot lengths, but it's often cheaper to buy steel here than at a home/lumber type store.

ting heads are operating on a large sheet of metal plate, cutting three or four copies of the template out at the same time. The clean plasma cuts mean less labor is required to make a finished part.

Metal cutting has been done for decades with oxy-acetylene equipment, but, besides the clean-cut edges, a plasma cutter has one other major advantage over the oxy-acetylene torch. Since the main component of the air around us is nitrogen, which make a good inert gas to superheat into a plasma flame, common air can be used as the cutting gas. This is usually supplied simply by hooking the welding machine up to a standard shop air compressor making 75 psi. Not only is the compressed air virtually free, with no bottles to fill, but plasma cutting involves no flammable gasses in dangerous high-pressure cylinders. There is a major savings in bottled-gas costs, convenience, and shop safety.

The nature of the plasma-cutting process is such that very little spark spray comes from the cutting action, and there is virtually no discoloration or distortion of the cut edges. It is almost like "cold" cutting. These attributes make the plasma system the method of choice now in automobile wrecking yards for cutting up cars for dismantling. It's much faster than a gas torch, and the chance of fires starting from flying sparks and droplets of molten metal is greatly reduced. Plasma cutters have been operated in such situations with torch hoses as long as 234 feet, with the welding machine kept indoors in a shed and the operator wandering around the yard with the torch, not having to transport a heavy cart with bottles of compressed gasses. Body shops love it for its clean edges and the ability to cut an exterior body panel with less chance of igniting undercoating, paint or burning a hole in an interior panel with a stray blob of molten metal.

Although the technology sounds complicated, plasma equipment is perhaps easier to maintain than gas, MIG or TIG equipment, and the learning curve is not as steep. As with the modern MIG and TIG machines, recent years have seen the price of plasma-cutting equipment come down considerably, and this has put the technology in the hands of many automotive shops. Most home/shop users will not have enough cutting work to do to justify owning a plasma cutter, but when a lot of cutting needs to be done, and done more safely than with traditional gas, this is definitely the way to go. If you are around a body shop or fabri-

2.30 If you can't carry a 20-foot length of steel home with you, steel houses will generally make one "courtesy" cut for a small fee. Any cuts after that are expensive, so it's best if you have some method at home for cutting the metal for your projects.

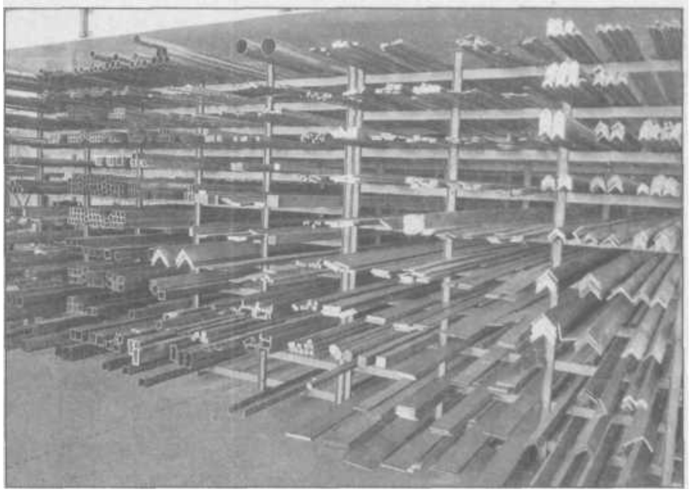

2.31 This steel supply store has a neatly-organized rack for materials, in which each different thickness of material is identified by a different color of paint on the ends, which makes shopping a lot easier for hobbyists without a micrometer.

cation shop that has a plasma cutter, see if they will demonstrate it to you. You'll find it amazing when you see how fast it cuts, how little spark spray there is, and how clean and relatively *cool* the cut edges are.

Practice and training

When it comes to hand-eye coordination, everyone has a little different skill level to start with. How easily each person can pick-up a welding machine, read the instructions and begin joining metals together will differ with that person's mechanical background. Some people seem to flow into the work intuitively; others require considerable practice and often a little coaching from someone who can weld already.

Each of the welding systems we have discussed in this chapter requires a different amount of practice and skill to master, but rest assured that even if you have never done anything like this before *you can learn to weld.* Millions of men and women have learned to do this in the past. During the domestic manpower shortages of W.W.II, there were many housewives who went to work in the defense plants, and even though they had never done anything mechanical they were there riveting, metal-shaping and welding on aircraft, tanks, ships and many other critical projects.

The first step in deciding what kind of equipment you need is to accurately assess the kinds of projects you would do most often if you had welding equipment in your home/shop. The thicknesses of metals you want to work with, the types and alloys of metals, the kind of electrical power you have available and the length of the welds you may need to perform are all factors in choosing the right machine. We've given you enough of a brief outline of the advantages and disadvantages of each type of equipment to aim you in the right direction. You may also want to consult with your local welding shop and local welding supply store before making your decision. If you know someone who already has a welder at home, ask his advice and even see if he will let you try out his equipment on some scrap pieces.

Most people who are serious about picking up welding skills and doing it right can benefit from actual training. Your local community college may have adult evening courses in basic welding, or there may be a trade/technical school nearby. Courses offered at publicly-funded community colleges and trade schools are usually inexpensive, and the hours are often built around a working person's schedule. There is a real difference between a weld that *looks* good, and a weld that *performs* well. If you practice welding on your own, you may achieve welds that look OK to you, but which may fail in actual use. One of the valuable parts of the training you will receive in a legitimate course is the critique from the instructor. He can show you how to test various trial welds for strength and let you know when you're getting it right. We think such training is invaluable if you plan to do anything more involved than repairing a bicycle or garden tool. Good training shortens the learning curve dramatically, and we recommend it to any readers who have it available in their area.

3 Oxy-acetylene gas welding/cutting

Perhaps because of its history going back to the turn of the century, or perhaps due to its appearances in films and other popular imagery, gas welding/cutting has the most "romantic" image of the various welding systems we'll deal with in this book. When most people conjure up a vision of a weldor, it is of a perspiring, leather-clad man wearing dark goggles and wielding a flaming torch with sparks flying everywhere. In truth, this is more likely a picture of someone *cutting* with a torch rather than welding, but it points out the ubiquitous nature of the gas equipment in the overall welding picture.

Although losing some ground to modern electric welders in the most recent decade as the welding system of choice for the average home/shop user, oxy-acetylene equipment is still unarguably the most versatile setup for home, farm or shop. Besides its function as a fusion-joiner of metals, gas equipment also of course is extremely valuable as metal *cutting* equipment, and can also be used for non-fusion metalwork such as brazing, soldering, and the dying art of automotive body lead work. It is also useful equipment in any automotive shop for freeing rusted fasteners, heating machined parts that require a hot shrink-to-fit connection (like a gear on a shaft) and heating metal parts prior to bending them. It requires no water or electrical connections, making it one of the most portable of systems, you can load it into a pick-up truck and take it to any remote location. If the gas hoses are long enough, the weldor can climb up a pole or down a shaft to perform welding or cutting, a flexibility few other welding systems can match.

The basic gas process

When gasses are used in other types of welding, they are usually of the inert kind, like argon, CO_2 or helium, which are involved in the welding process only to keep the molten weld puddle clear of impurities from the air during formation. In oxy-acetylene welding/cutting, however, the

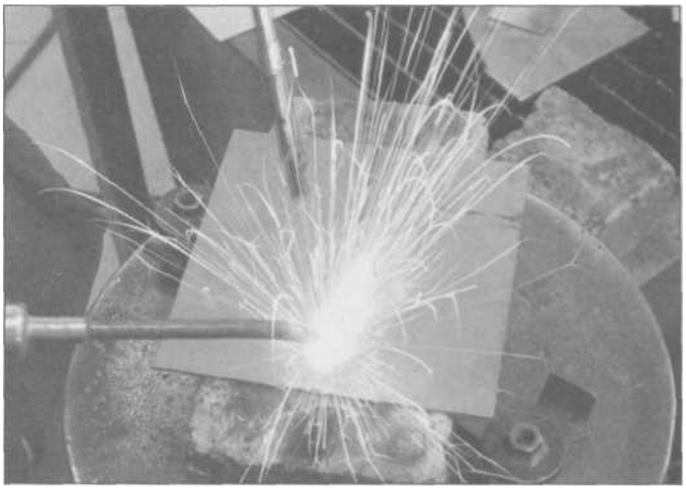

3.1 Versatile gas-welding equipment has been a mainstay of metal fabricating for most of the 20th century, and little has changed with procedures or equipment. The basics, fusing metals at their melting point, are still valid for many uses in farm, industry and home/shop hobby applications.

3.2 You'll find helpful people at your local welding supply store that can set you up with torch equipment, oxygen and acetylene bottles, and other supplies you'll need.

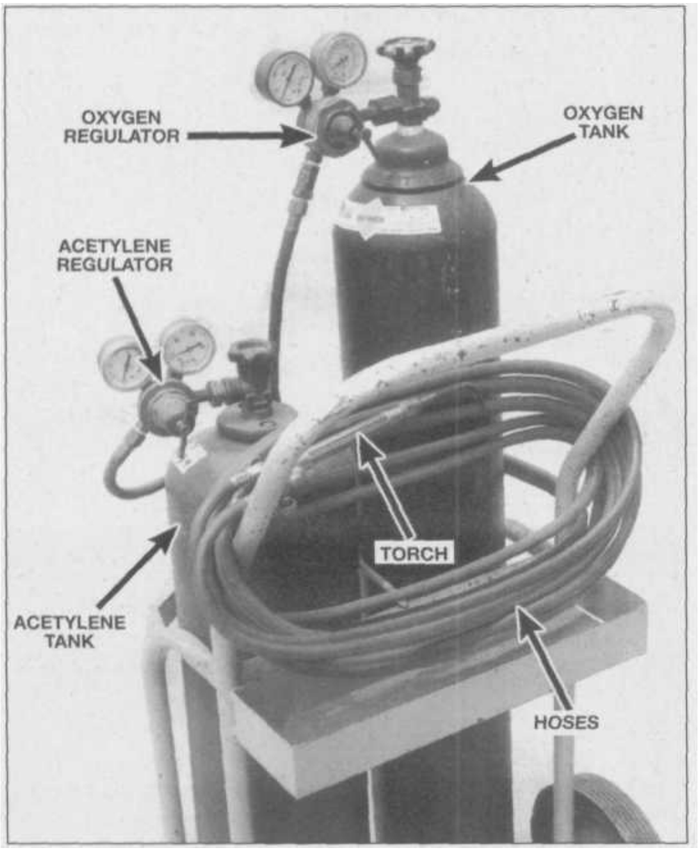

3.3 Look in the corner of almost any automotive or fabricating shop and you'll a setup just like this, with bottles, regulators, hoses and torch neatly corralled in a portable welding cart. Carts can be readily purchased, but most beginners like to make their own cart as their first project.

PUROX
PUROX
PUROX

3.4 This is typical of a good, general-purpose welding, cutting and heating set, ideal for automotive, muffler shops, farm and trade schools. Starter sets have smaller torches and 10-12 foot of hose, bigger sets have larger torches and 25-foot hoses.

gases are what make the flame itself. Acetylene gas is quite flammable, and combined with oxygen, which by itself does not burn but speeds up the oxidation or burning of any other fuel, makes one of the hottest possible gas flames (5600-6300° F),suitable for the rapid welding, cutting or heating of most ferrous and non-ferrous materials. Although all compressed gasses pose some shop hazards because of the pressure inside the bottles, oxygen and acetylene are considerably more dangerous to work around, and require much more caution and close attention to safety rules.

Oxygen, while not exactly flammable by itself, is the gas necessary both foi us to breathe and for any type of combustion to take place. Combustion is really nothing more than very rapid oxidation, and if pure oxygen is directed at some thing flammable, a fire can start very easily. Some inexperienced weldors hav< been know to dust off their work clothes with their unlit gas torch, but the extn oxygen that gets into their clothes can make it so flammable that any tiny sparl could start the clothes on fire. Likewise, oxygen gas must be kept away fron things like oily rags or any petroleum products. Oxygen as used in welding equip ment is generally stored in high-pressure, 1/4-inch-thick-walled steel cylinders s 2200 psi. While oxygen may be important for our lungs to work, it can still be dan gerous if too much is introduced to the bloodstream. For this reason, oxygen ga should be kept away from open cuts such as you might have on your hands.

Acetylene gas, on the other hand is flammable to the point of being explosivt and is also mildly poisonous, causing nausea and headaches if you breathe muc of it. Older readers may remember using a "carbide cannon" on the 4th of Jul when they were kids. A small quantity of pellets were put into a small met;

cannon which held some water. A gas was produced inside that was lit by a flint-sparker mechanism on top, and the result was a tremendous bang. The pellets were calcium-carbide, and the gas produced was acetylene.

Pressure in the acetylene gas cylinder is much less than with oxygen, at 250-325 psi, but the construction of the cylinder is different. Acetylene cylinders are shorter and fatter than oxygen bottles, and are constructed in two halves. Because acetylene is unstable at high pressures, the only way to get sufficient quantities into the standard bottle is to dissolve the acetylene in another medium. In welding tanks, the two halves are filled with an asbestos/cement mixture and then welded together. After baking, the material forms a honeycomb inside the tank. Liquid acetone is put into the tank because it will absorb 25 times its own volume in acetylene gas, thus stabilizing the acetylene.

Because of the differences in chemical action and storage of oxygen and acetylene, there must be no mix-ups between the two. For this reason, the tanks are made in different proportions and different colors; the acetylene bottle has only left-hand threads, and the hoses for each bottle are different, i.e. red hose for acetylene, green hose for oxygen **(see illustration).** Both bottles have a threaded top for filling and connection of the regulators, and due to their high internal pressure, especially the oxygen cylinder, gas bottles should always be stored securely. If a bottle were to fall from a truck onto concrete, for instance, the regulator could be knocked off, suddenly releasing enough gas to propel the heavy bottle around like a kid's balloon, except with potentially deadly force. In addition, the acetylene bottle should always be stored in an upright position. If stored laying down, some of the liquid acetone inside could be drawn out of the bottle into the welding supply, ruining the weld.

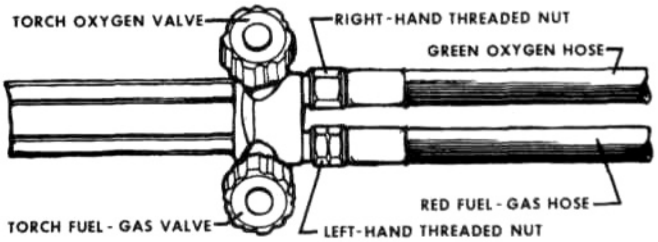

3.5 There must be no mix-up of gasses, so oxygen hoses are colored green, acetylene hoses red, and the acetylene hoses have left-hand threads, identified by a groove cut around the fittings.

The equipment

Besides the two gas bottles, there are gauges/regulators, hoses and the torch itself **(see illustration).** Each gas tank has it's own set of two gauges mounted on a regulator body designed to reduce the high cylinder pressure down to a useable pressure for the torch. On each tank, there is

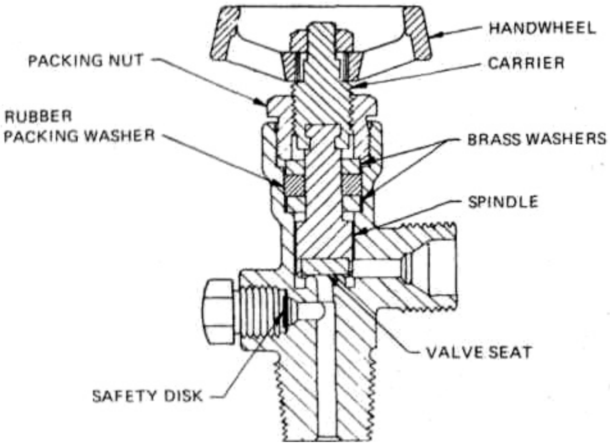

3.6 The gas bottles are fitted with valves, to which regulators are attached. This is a cutaway of an oxygen cylinder valve, showing the way the valve seals internally. Oxygen cylinder valves seal best when either closed or fully open.

PUROX
PUROX
PUROX

3.7 Heavier-duty torch sets like this Trade Master from L-TEC have torches capable of welding up to 3/8 steel, cutting up to 1 1/2-inch steel, and heating capability of 103,000 BTUs per hour.

PREST-O-LITE
Kangaroo Kart

3.8 Highly portable "Kangaroo" outfit has everything you need in a small package, including a polypropylene carry cart. This doesn't have a lot of welding-time capacity, but is popular for maintenance work, metal sculptors, and automotive shops that don't use a torch too often.

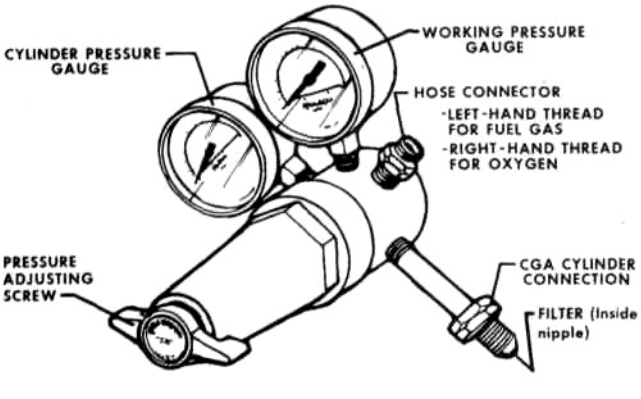
CYLINDER PRESSURE GAUGE
WORKING PRESSURE GAUGE
HOSE CONNECTOR
-LEFT-HAND THREAD FOR FUEL GAS
-RIGHT-HAND THREAD FOR OXYGEN
PRESSURE ADJUSTING SCREW
CGA CYLINDER CONNECTION
FILTER (Inside nipple)

3.10 The gauge/regulator setup screws onto the tank's cylinder valve, and has a fitting for attaching the hose leading to the torch.

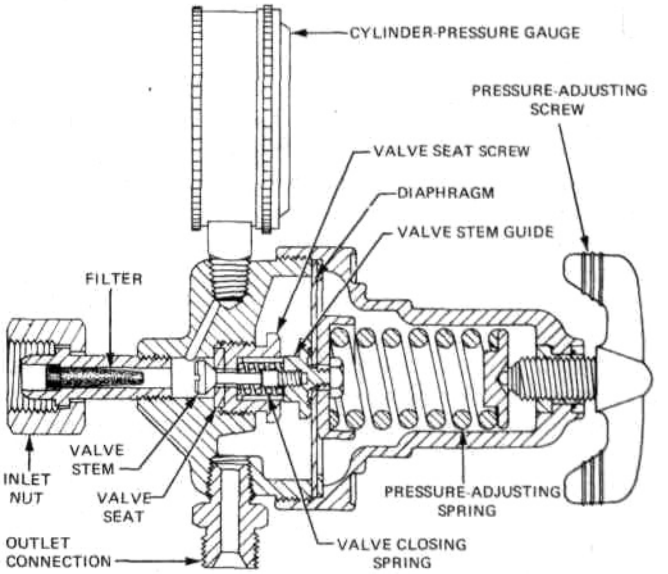
CYLINDER-PRESSURE GAUGE
PRESSURE-ADJUSTING SCREW
VALVE SEAT SCREW
DIAPHRAGM
VALVE STEM GUIDE
FILTER
VALVE STEM
INLET NUT
VALVE SEAT
PRESSURE-ADJUSTING SPRING
OUTLET CONNECTION
VALVE CLOSING SPRING

3.9 Cutaway of a single-stage regulator shows how the cylinder pressure is adjusted down to a usable range for welding. There are two gauges attached, one shows the total pressure left in the cylinder, and one shows the adjusted pressure supplying the torch.

one gauge to indicate the pressure in the tank, while the other gauge reads the pressure *after the regulator,* which is the pressure delivered to the torch and a very important adjustment in oxy-acetylene welding.

There are two basic types of regulators found on gas welding setups, single-stage and two-stage. The former type reduces the cylinder pressure to a working pressure in just one step, say from oxygen's 2200 psi to 3 psi for the torch **(see illustration).** In the two-stage regulator, the first stage reduces pressure to about 30 psi; the second stage adjusts pressure from there down.

A great portion of the expense in a gas-welding setup comes from the gauges/regulators, and better ones simply cost more. The simpler-to-manufacture single-stage regulators cost less and are therefore commonly found on the less-expensive kits for oxy-acetylene welding **(see illustration).** They work fine for most purposes, and their only drawback is

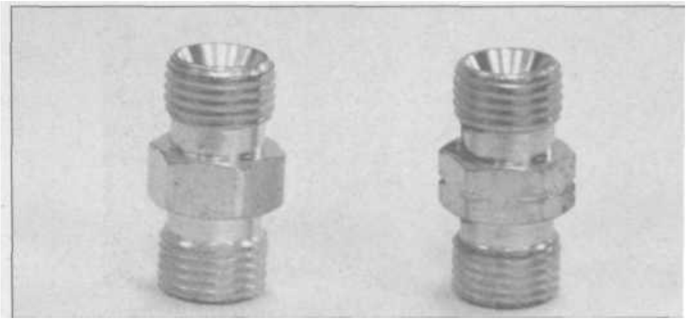

3.11 Connectors are available to add more hose onto a starter set for a longer working distance if you need it. These screw into your hoses and the extra new hoses screw on the other side.

that the regulation depends heavily (excuse the pun) on the pressure inside the tank. As the tank pressure goes down as the gas is used up, the regulation changes and must be adjusted again. Also, when doing a job that requires a lot of gas flow, such as large welds or cutting/heating thick plates, the single-stage regulator may not be able to keep up. The more expensive two-stage regulators maintain gas flow to the torch with less fluctuation, which is important in a professional welding operation but may not be a big factor in home/shop welding if you don't work on heavier materials. Obviously, gas-welding sets with two-stage regulators will cost more initially than the single-stage kits.

The hoses for a gas-welding setup are specially-designed to carry the oxygen and acetylene gasses. While they do not handle the very high tank pressures, they are built in three layers to withstand the *regulated* pressures for many years with proper care and maintenance. As mentioned above, the hose for carrying acetylene is red, while the oxygen hose is green, and the acetylene hose has left-hand threads on its fitting, so there is no chance of mixing them up when making connections **(Note:** *Left-handed fittings generally have a groove cut in around the fitting, to differentiate them from normal, right-hand fittings).* Doing so could be disastrous, since any residual flammable gas or other material in a hose connected to pure oxygen could cause the hose to ignite or explode. For this same reason, it is important that no lubricants or chemical compounds of any kind be used on the hoses or fittings. Gas-welding packages may contain hoses around 10-12 feet long in the economy or light-duty kits, and 20-25 feet long in the heavier, more professional sets. Extension hoses can be added to your starter set if you need to reach areas further from your tanks **(see illustration).**

3.12 In industrial situations, like this oilfield service truck, gas welding setups may be equipped with large cylinders and 100 feet of hose, to handle any situation.

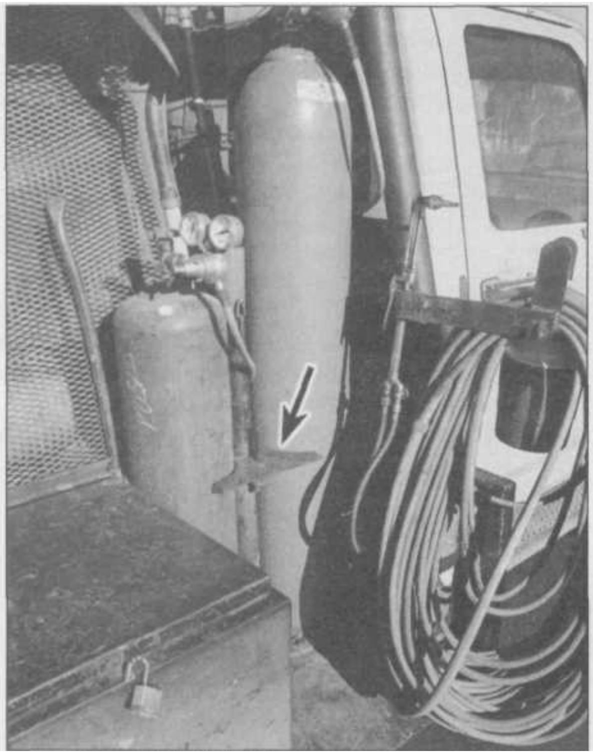

3.13 In this clever setup, the hoses and torch are attached to a swingaway arm for access to the tanks for servicing. Arrow indicates retaining plate with a simple steel wedge rod. Tanks are held securely but can be changed quickly without wrenches.

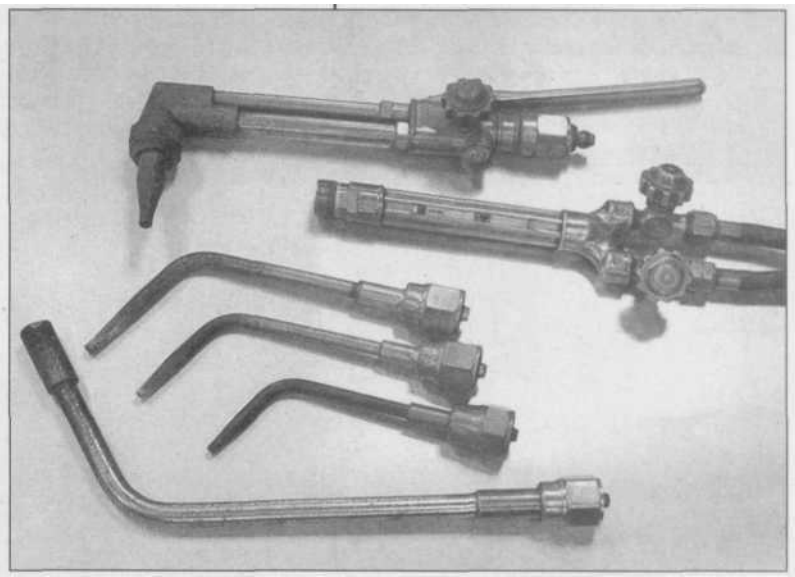

3.14 Basic torch sets include, from top: cutting torch attachment, basic torch head, a variety of welding tips sized for different thicknesses of material, and at bottom a rosebud tip for heating large areas.

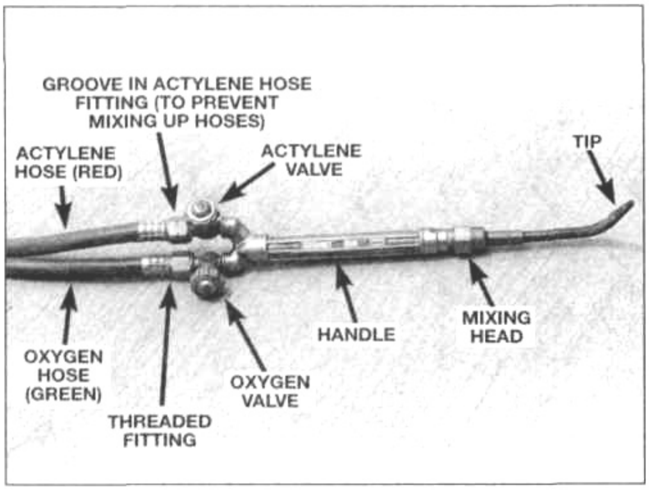
GROOVE IN ACTYLENE HOSE FITTING (TO PREVENT MIXING UP HOSES)
ACTYLENE HOSE (RED)
ACTYLENE VALVE
TIP
OXYGEN HOSE (GREEN)
THREADED FITTING
OXYGEN VALVE
HANDLE
MIXING HEAD

3.15 When assembled with welding tip and hoses, this is what an oxy-acetylene welding torch looks like. The two gasses are controlled by separate knobs, and are mixed inside the torch body.

The torch for oxy-acetylene welding is a precision-made tool, usually constructed with a forged brass body, two knobs controlling the gas flow, and copper tips of varying size **(see illustration).** In use, the two gasses flow through valves controlled by the knobs, and mix inside the torch with the combined gasses coming out of the tip in an oxygen-to-acetylene ratio controlled by the weldor. The weldor changes tips, each of which has a carefully sized opening, to suit the gas-flow and flame shape required for a particular job. The makeup of the cutting torch is different from the welding torch and will be discussed later on.

Gas welding sets can be purchased three ways: as a welding-only setup, a cutting-only setup or a combination set with two torches. The type you choose will depend on your needs. If all you need to do is weld up animal corral fencing made from steel pipes, you may use a cutting-only gas set for cutting pipe to length and shaping the cut ends to fit against the next pipe, and use a generator-driven arc welder on your truck for joining the pipes. However, if you are doing automotive sheet metal work or metal sculptures with relatively-thin materials, you will probably do most of your cutting with an electric shear, hand tin snips, hacksaw, etc., and only require a torch for welding, brazing and occasional heating. A combination set is of course the most versatile. Most sets that you buy can be fitted with a different torch at any time, so if you decide to buy a welding-only setup, you can purchase a cutting torch later on if your need to.

After you purchase your welding package, you will have to buy or rent tanks. The standard-size oxygen and acetylene tanks will hold enough gas to last the home/shop user for a long, long time. The tanks must be *hydrostatically-tested* periodically, which involves filling them with water to a certain pressure to test for safety. When the test is due, your local gas supply cannot refill your tank until it is tested. When you buy new tanks outright, you then become responsible for the testing costs and any cylinder-valve repairs, while if you lease the cylinders, the gas company takes care of that as part of the lease agreement. Leasing the cylinders allows you to get started welding with the minimum investment, because outright purchase may be expensive. However, if you add up the monthly rent

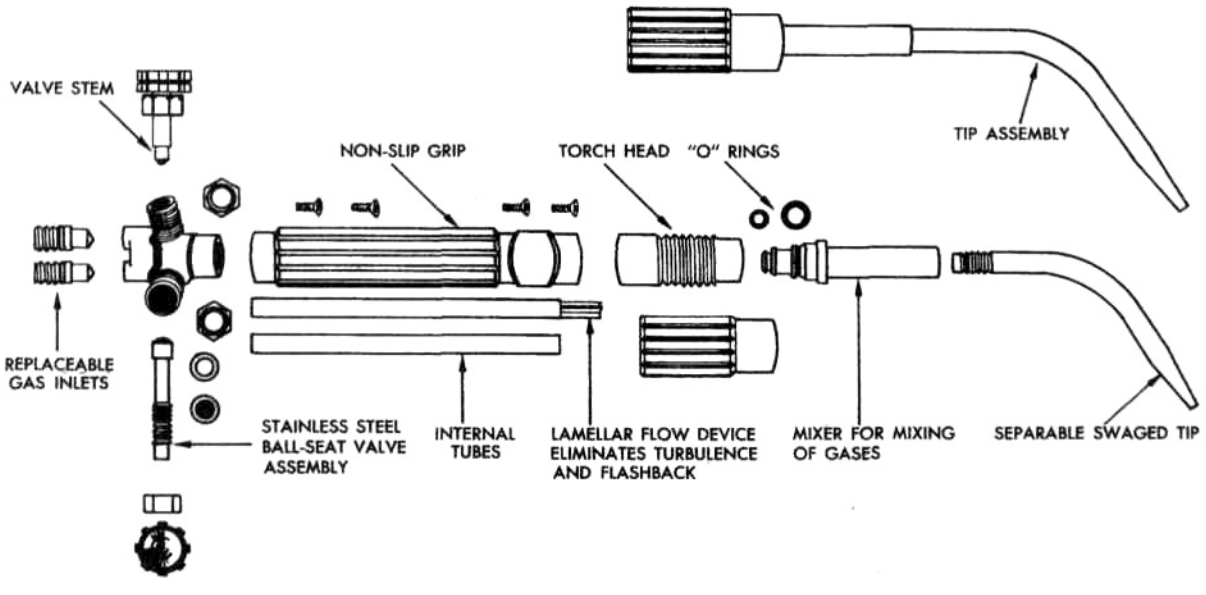
VALVE STEM
NON-SLIP GRIP
TORCH HEAD
"O" RINGS
TIP ASSEMBLY
REPLACEABLE GAS INLETS
STAINLESS STEEL BALL-SEAT VALVE ASSEMBLY
INTERNAL TUBES
LAMELLAR FLOW DEVICE ELIMINATES TURBULENCE AND FLASHBACK
MIXER FOR MIXING OF GASES
SEPARABLE SWAGED TIP

3.16 You'll never have cause to disassemble your torch this far, but these are the internal components. Some torches require wrenches to change tips and heads, while others have knurled, hand-tightened connectors.

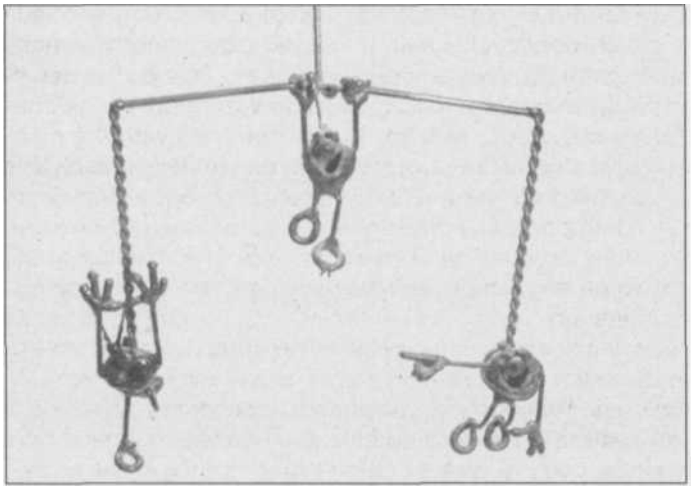

3.17 Each of these little "flea circus" figures are welded up completely of thin welding rod and tiny blobs of molten metal. The artist works with a very small welding tip and tremendous patience, showing what can be done with skill and imagination.

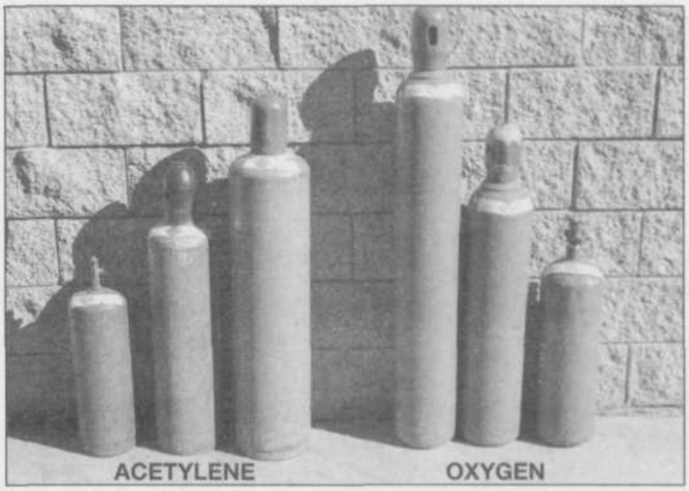

3.18 The small, medium and large welding cylinders contain 80,150 and 275 cubic feet of gas, but some manufacturers label theirs by a letter designation. Ask your welding supply shop approximately how long each size would last in terms of welding, and choose according to your needs.

charges, you'll find that after two years of leasing you could have purchased the tanks outright for the same money. Most leases run from 5-99 years, but if you move to another area where a different gas company is located, you may be able to return your tanks to the original location where you leased them and get back a pro-rated portion, then begin a new lease account in the area you move to. Look in your local telephone Yellow Pages under "Gasses, Industrial" or "Welding Supplies."

There are usually three sizes of tanks, small (80 cu. ft.), medium (150 cu. ft.) and large (275 cu. ft.). The small tanks are the easiest to store and maneuver around the shop and are fine for most occasional users. The medium tanks will be more than enough for most any home/shop use, lasting more than a year under normal use. The largest tanks are heavy, and more difficult to move around the shop.

Getting started with oxy-acetylene

Thoroughly read all the directions that come with your gas-welding setup before you do anything. If you have purchased a used welding set from someone, make sure the instructions are included, or have an experienced person help you get set up. Compressed gasses are dangerous; so is the welding process, and there are many more safety considerations to understand than we can cover here in this book. Obviously, since we are dealing here with a very hot flame and the promise of flying sparks, you should have a clean work area away from any flammable materials and a good fire extinguisher should be handy, as well as a source of water. If a spark gets onto your shoe, don't waste your extinguisher on that. Use a water spray bottle, which is also handy at times for cooling off a piece of metal.

Follow your set's directions for attaching the gauge/regulator sets to your filled bottles of oxygen and acetylene **(see illustration).** The bottles should be secured with a sturdy

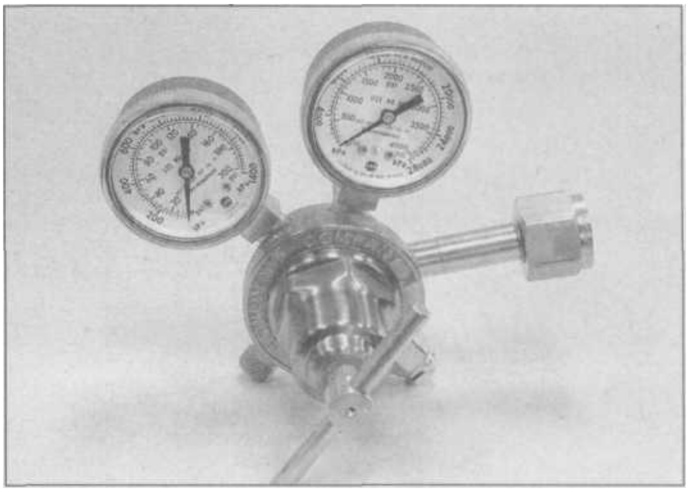

3.19 Most home/shop users of oxy-acetylene will have single-stage regulators on their torch sets. Unless you do a lot of welding/cutting, these are just fine, though you have to readjust the output pressure more frequently than with the professional, two-stage regulators.

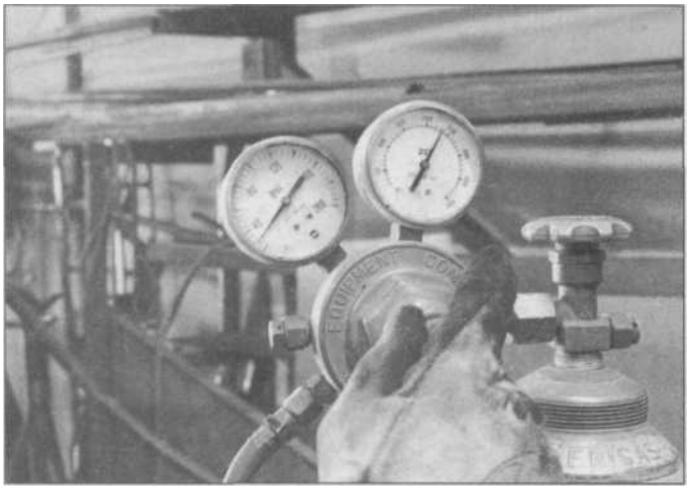

3.20 Gas flow is regulated on the oxygen cylinder to 1-5 psi for welding, depending on the thickness of the material to be welded, and the acetylene pressure is adjusted the same. Higher pressures are set when performing cutting, but acetylene must never exceed 15 psi to the torch.

chain to either your bench or a welding cart that has provisions to keep it from tipping over easily. Before the regulators are attached, blow any dust out of the cylinder valves by *cracking* them. You should stand to the side of the cylinder when doing this, with the outlet pointed away, and very quickly open the valve slightly and close it again. Just a little blast of gas will ensure that the connection is free of dust.

Use a proper-fitting wrench (not an adjustable) on all your regulator and gas hose connections, which are usually brass fittings, and remember that the acetylene connections are *left-hand* threaded. Blow out the regulator by slowly opening the cylinder valve when the pressure-adjusting screw (usually T-shaped) is turned out all the way. One gauge will show you the cylinder pressure, and the other the hose/torch pressure. Attach the hoses and blow them out with a blast of gas. Adjust the regulators to provide 2-3 psi working pressure from each bottle **(see illustration).** The cylinder valves should always be opened slowly, with your body away from the ends, and the oxygen valve should always be opened all the way. It is of a type that seals best either fully closed or fully open. The acetylene valve should be opened only about 1-1/2 turns.

Connect the hoses to your torch with the gas cylinder valves closed. In most cases this involves wrenches, but there are some professional sets that have snap-on connections (like air line hoses) to make quick, wrenchless work of changing torches **(see illustration).** Select a medium tip from your set (your set's instructions will recommend what gas pressure works with which tip and what thickness of metal they are suited for) and attach to the torch. After everything is assembled, you can test the connections with soapy water for leaks.

You will need something to light the torch flame, and a wire-framed metal striker should have been included with your set **(see illustration).** Never use open flames such as a cigarette, cigarette lighter or match to start your torch! Your fingers need to be well out of the way. You should be wearing leather gloves and heavy clothing with leather shoes or boots. Do not tuck your pants into your boots, or you could create a pocket there that could collect sparks. Likewise, do not wear pants with cuffs, for the same reason. Have your colored welding goggles on your head, but not yet over your eyes.

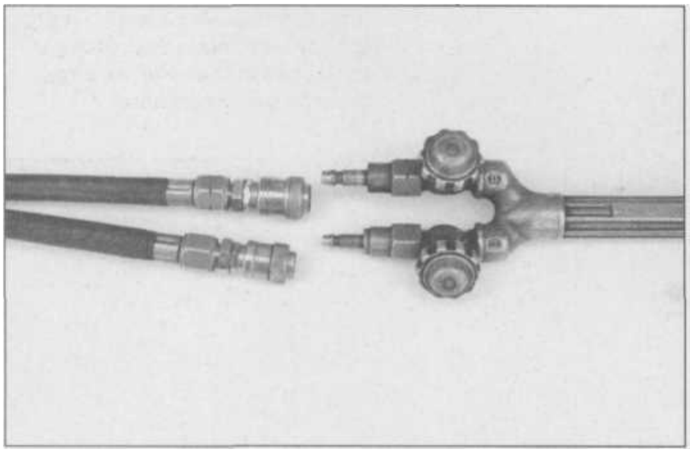

3.21 Some torches are equipped with quick-release fittings as usually seen on air-compressor hoses, making for quick, easy changes.

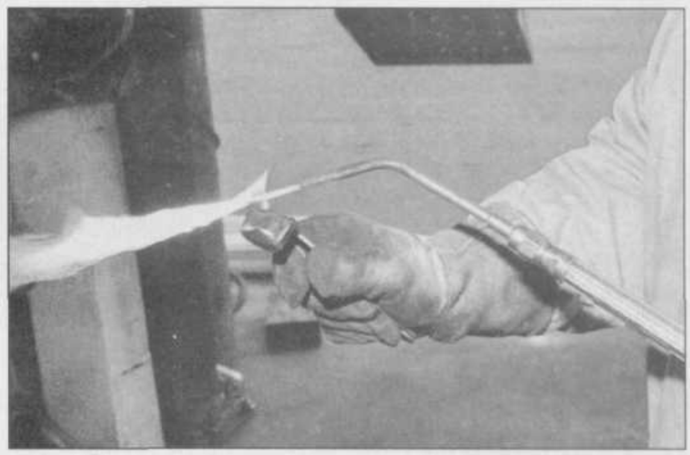

3.22 To start a torch, open the acetylene valve 1/2 turn and light immediately with a striker like this. Never use a match, cigarette or lighter, you could easily burn your hand. Once the sooty acetylene is lit, gradually add oxygen until you achieve a neutral flame.

Flame adjustment

This is a very important step in gas welding, and one you will be making every time you pick-up the torch to weld, braze, cut or solder, so you'll have to practice this to the point where it becomes second nature. Open the cylinder valves slowly as described before, and make sure the regulated pressure is set correctly. Hold the torch in your operating hand (right, if you are right-handed) and the striker in the opposite hand. Both valves on the torch are closed. Open the acetylene valve on the torch about a half-turn and operate the striker near the torch tip. There should be a flame with some sooty flying particles. Quickly add some oxygen and the soot should disappear. Study the illustrations of flame type. What you have right now is a *carburizing* flame, which is overly-rich in acetylene. As you add more oxygen, the flame characteristics will change. While the rich mix gives you a wide, flared-out yellow flame, adding oxygen will make the flame narrower and brighter.

The ideal flame for most gas welding is a *neutral* flame, which is neither carburizing nor oxidizing (too much oxygen). When you have adjusted the torch properly, there will be a large flame graduating from blue at the end away from the tip down to white near the torch and with an inner cone that should be light blue. If you add too much oxygen, the inner cone will be smaller, pale in color, and there may be a hissing sound. You'll need to practice achieving the neutral flame, and when you start welding, you'll see what happens to metals when the flame is incorrect.

The proper method of shutting off the torch is equally important. Close the acetylene valve on the torch first, then the oxygen. If you do it the other way around, there could be a small amount of gas left in the tip that could ignite. If you are through welding for some time, it is good safety practice to then close down the gas cylinder valves as well, but if you are going to weld again in a short while, just shut off the torch valves only.

Gas welding

You'll need a lot of small pieces of clean scrap metal to practice on, of varying thicknesses. Even if you only intend to do light sheet-metal welding, start by practicing with thicknesses up to 1/4-inch to become familiar with different tip sizes and flame adjustment. The cleaner the scrap pieces, the

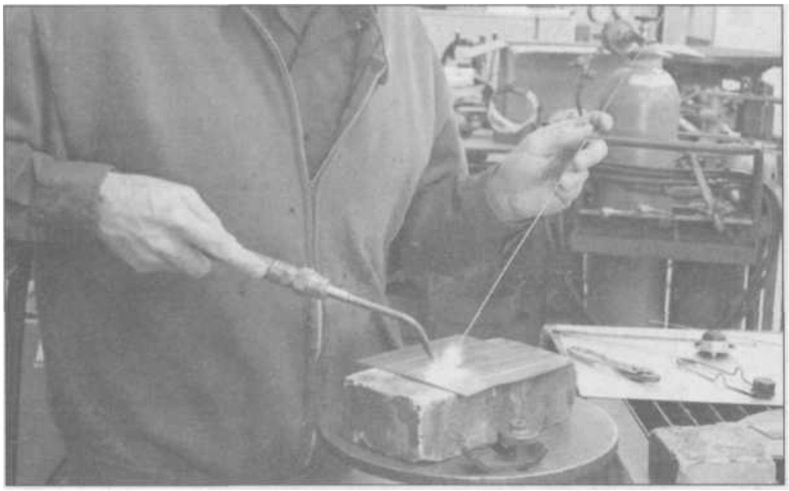

3.24 You don't have to worry about grounding with gas equipment, but there is a dangerous flame involved, so your welding projects and practice sessions are best done over some firebricks like this, which you can get at your welding supply store.

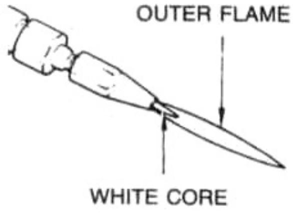

Neutral Flame

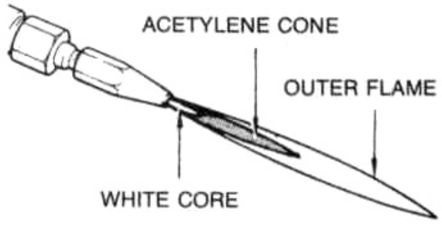

Carburizing Flame

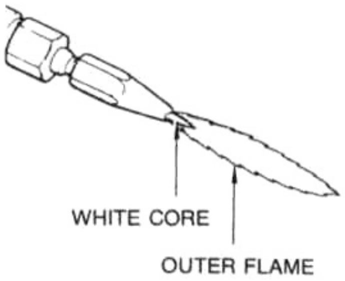

OUTER FLAME

Oxidizing Flame

3.23 The oxidizing flame has too much oxygen, and a feathery blue flame with small white inner flame. The carburizing flame is acetylene rich, and features a large flame, with a large blue inner cone and a longer, colored cone around that. For welding, adjust for a neutral flame.

better your welds will be, and this applies to any type of welding, gas or electric. You should sand or grind the areas to be welded to remove any paint, coatings or rust. You should not use for testing any metals that are galvanized or coated with zinc, or any kind of plating. Not only will welding be dirty and difficult, but the coatings may cause harmful gasses when vaporized in the weld.

Start your first exercise with a piece of steel on a level, non-flammable surface, not a wooden bench. The ideal work area would be a steel table (not just a wooden bench with thin sheet metal nailed onto it) with some firebricks on top. The firebricks can be obtained at your welding supply center, and, in addition to being non-flammable, they do not suck the heat out of your welding sample like a steel table, and there's no chance you're going to weld the sample *to* the brick. It can be very embarrassing when you practice on a steel table and you have various scraps welded - stuck to your table.

With your torch adjusted properly, pull down your goggles and put the torch tip about one inch from the steel and just try to make a long weld puddle of molten metal. The torch should be leaning back at a 45-degree angle, with the flame pointing in the direction the bead is going. When the steel starts melting into a small pool where you started, move the tip along and give it a rhythmic motion of an oval shape as you go, trying to follow a straight line, blending the newly-heated edge of the puddle with the back edge of the puddle. The process will leave a line of rippled ridges. If you are burning holes in the metal, there may be too much oxygen, the torch may be too close to the metal, the tip could be too large or you are possibly advancing the torch too slowly. If you go too fast, have the torch too far away or the tip is too small, your puddle will be very small and it won't penetrate the metal very far. The more the torch is pointed straight down at the metal (no angle) the deeper the penetration will be, and vice-versa. Practice at different angles until you can make a bead like our illustration, with even ripples and proper penetration.

Of course, not all of your welding will be done this way, in fact almost all of your real welding will involve adding filler rod, but for now you need to learn torch control and rhythm. Many of the projects you encounter will involve welding along an irregular line, especially in repair work, where you are welding up a crack. To become more familiar with making a bead along an irregular line (still without filler rod) draw a zigzag or curved line on the steel with soapstone and practice making a good bead. Proper hand control is everything in gas welding. Your hand must be comfortable and steady at all times. In welding classes, students are often challenged to write their first name in script with welding bead on a piece of metal, which challenges you in handling changes in bead direction.

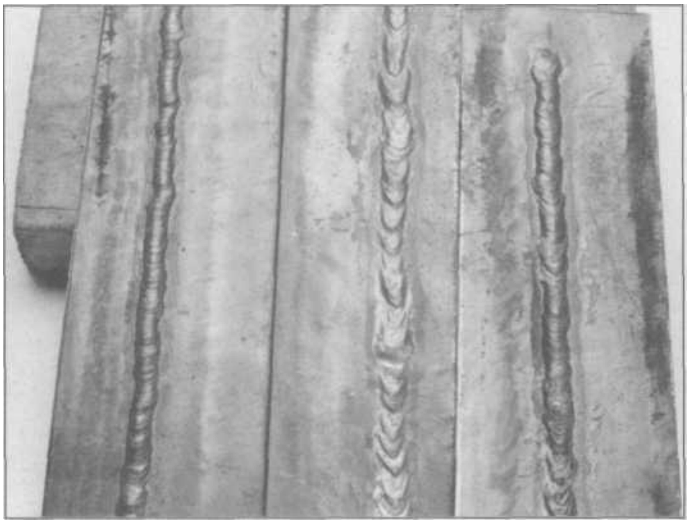

3.25 Heat in gas welding is controlled by matching the gas flow and tip size to the thickness of material. Your welding travel speed and tip-to-work distance will control the puddle size and penetration. With all things right, you should be able to heat your first puddle molten in about six seconds, then move along to make your bead. At left is a bead run too cold (bead stands up), at center is one run too hot and too fast, and at right is a better bead with good width and penetration.

When you are successful with these experiments laying down a good bead on the metal, you're ready to try joining two pieces using a filler rod. This requires not only good torch control, but control of your other hand as well with the filler rod. Generally, the filler rod should be the same material as the base or parent metal for good fusion welding. If you weld aluminum, the filler rod must be aluminum, and steel rods are used on steel. The steel filler rods are coated with copper to keep them from rusting and to make cleaner welds. As a basic rule, select a filler rod that is the same diameter as the thickness of the metal you are welding. Most light metals are welded with 3/32-inch, 1/16-inch or 1/8-inch rod, and you will probably never gas-weld with anything bigger than 1/4-inch rod. To weld thick materials, several passes are required to make a good fusion, and various forms of electric welding are usually much faster than a torch at making multiple-pass seams.

Welding with filler rod

Let's say that you need to join two pieces of 1/8-inch-thick metal in a straight bead from right to left. Position the two pieces such that they are butted together where they are to be joined, and that you have something blocking or clamping them in place so they won't move as you weld. Start your weld puddle at the right end of the seam, and as the puddle develops, use your left hand to insert the tip of the filler rod into the center of the hot puddle, then pull it out right away. Set up a rhythm of alternately putting in the rod and taking it out as the bead progresses to the left. When we describe the puddle, the edge of the puddle closest to the direction you are traveling is the *front* edge, while the part of the puddle you just finished is the rear.

If you have ever done any soldering, you may have a bad habit of heating the solder with your torch or soldering iron and letting the molten lead drip onto the joint. This is a bad habit and one you shouldn't bring with you in your welding practice. The molten puddle should melt the filler rod, not the torch. If the rod sticks, then you are not keeping the puddfe molten enough, or you are adding filler rod too fast, either of which will result in a low-quality weld.

This skill of making a good bead, with evenly overlapping ovals and good penetration will take considerable practice, but if you learn two-handed gas-welding techniques well, any other welding system you pick up later on will be considerably easier to learn. Most of the welding you will do will be *forehand* welding, in which you angle the torch so that the flame is aimed somewhat toward the direction of the desired weld bead. This preheats the metal as you go and makes addition of filler rod to the front edge of the puddle easier because it is truly hot. Sometimes it seems like keeping the proper torch distance, moving at the right speed, circling or ovaling the tip and dipping the rod in at the right time is just too much to concentrate on at one time, but you *will* get the hang of it. This isn't "brain surgery" and millions have learned before you. After a good long practice session, begin to examine things around you that are of welded construction and you'll get an appreciation of the practice it took to lay down those beautiful beads that look like they were done by some machine instead of human hands. Indeed there are some kinds of welding that are done with largely robotic welding machines, but a good weldor can make virtually perfect joints that are stronger than the parent metal.

We mentioned before that cleanliness is critical in making good welds. The two edges to be joined need to be sanded or ground clean to bright metal and should fit as closely as possible. Small gaps are inevitable in some kinds of work and can be bridged with filler rod, but gaps should be kept to a minimum. On thicker materials, you will find that it is easier to make a clean weld with good penetration if the edges of the two pieces are beveled before you weld. This can be done with a file or grinder. Thin materials have their own idiosyncrasies, mostly to do with keeping the two edges in alignment. As you progress with your weld bead, more and more heat travels out into the parent metal, so that after welding for some distance, the metal you now encounter is hotter than the metal you started on, so you may have to adjust the torch away slightly to compensate for the puddle forming faster.

Another problem with thin materials is that the parent metal tends to *distort* with the application of heat. Metal expands when heated, and two edges that were parallel when you started welding probably won't stay that way. Usually, the edges pull away and you have a gap to bridge that wasn't there when you started. The solution is to *tack-weld* at intervals before applying too much heat. A tack-weld is

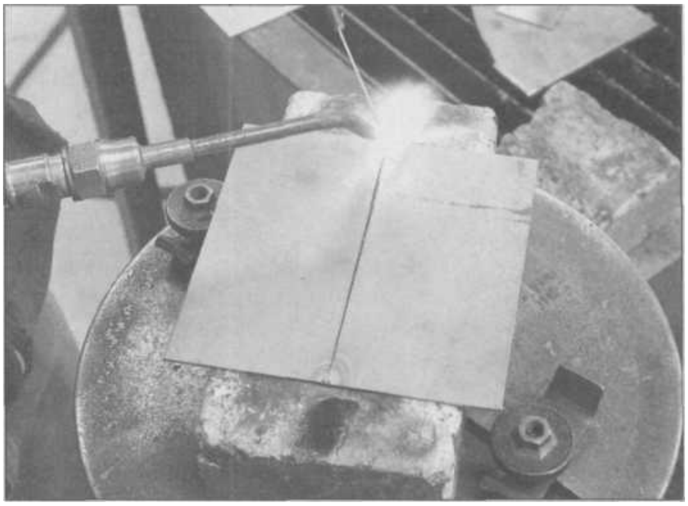

3.26 When practicing basic butt welds, tack the near corner first, then tack the top corner with a slight gap between the parts. This will close up as the plates grow from welding heat.

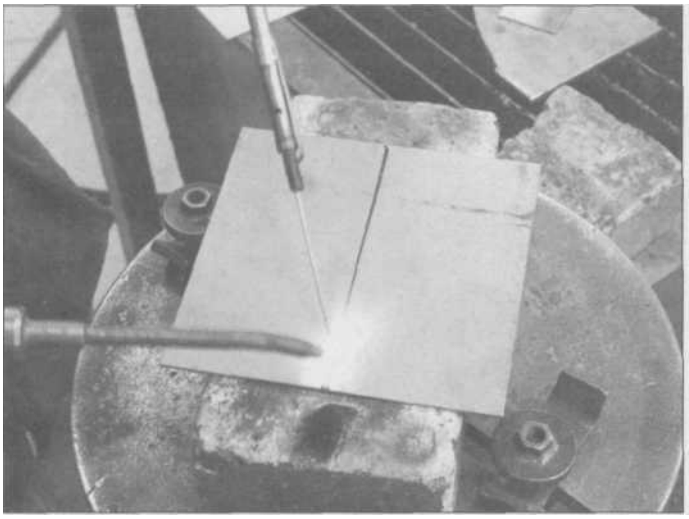

3. 27 Start your bead about an inch beyond that first tack, and proceed toward the end, keeping the torch steady and dropping molten beads off the filler rod by dipping into the puddle quickly and evenly as you go, building an overlapping series of puddles.

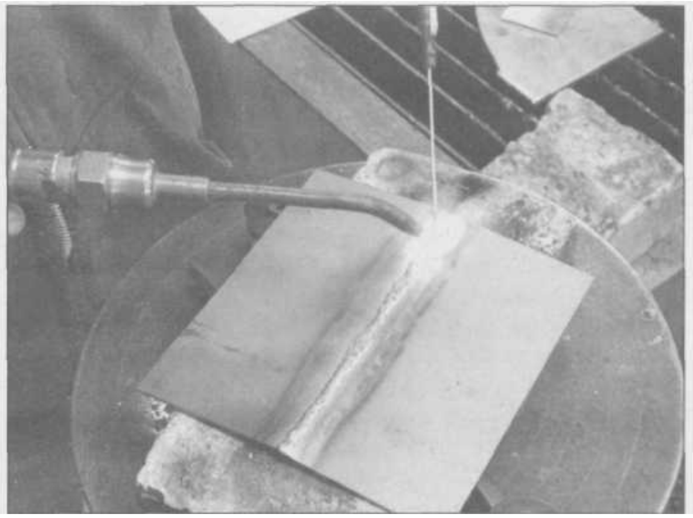

3.28 Before you are done, you will notice some movement or warping of the sample plates. If you plan to do automotive body work with a torch, you'll have to practice a lot on thin metal to get the hang of welding without burning holes or warping the work. Gas welding is not suited to the thin, high-strength body steels used on today's unibody cars.

simply a bead only one or two puddles long, done quickly. To join a long seam in sheet metal, tack the ends first, then the middle, and then put tacks every few inches apart, while alternating the ends you tack **(see illustration).** When you go along later to do the final, full seam welding, the tack puddles can be melted into the main seam as you go, and the pieces will stay aligned the whole way. This is particularly important when joining long sheet-metal seams like welding a new quarter-panel onto a car. In auto bodywork, you need to keep the warpage to a minimum. In some cases when doing the final seam after tacking sheet metal, you may stop after a few inches are welded and quickly apply a body dolly to the backside of the weld area and hammer on the weld area on top with a body hammer. This "hammer-welding" (alternating welding and hammering) technique straightens the seam and the two panels while they are still hot and pliable, and the resultant seam when done by a good body man requires much less grinding or filling to make it ready for primer and paint.

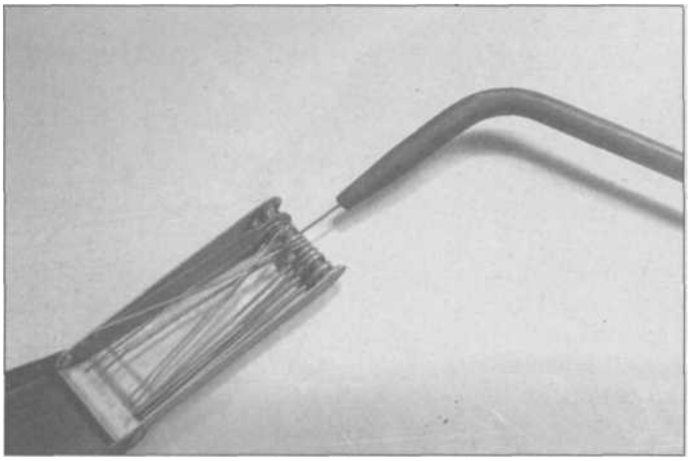

3.29 As you weld, it is inevitable that sparks and spatter will blow back onto your torch tip. If it plugs up, your bead will be spotty and your torch may go out entirely. Keep the tips clean with a set of fine-wire cleaners that come with your gas set.

The basic joint we have illustrated is a buff joint, where two pieces of metal are pushed together and welded along the seam where they meet. There are many kinds of joints, and it is helpful to try welding a few other types after you have practiced enough on butt welds on thick and thin materials. *Corner* welds, where the two edges meet at an angle can be either inside or outside corners **(see illustrations).** The outside corner weld is the easier to perform because you can see the seam so clearly, and the pieces need not be beveled. Just placing them so that the edges form a groove or V will do it. On inside corner welds, depending on the angle involved, the action is harder to observe, being shrouded somewhat by the pieces. After you have practiced on these more difficult welds, try a lap weld. This is where the two pieces lay over each other (overlapping). You run a bead where the joint is, and depending upon the strength required for the application, you may want to weld the seam on the other side as well. In auto bodywork, sometimes panels have to be overlapped instead of butt-welded, but in such cases the backside seam is seldom welded.

A T-weld is where you join a piece of metal to another larger piece at right angles, forming a letter T in profile. This requires some additional practice to perform well, as part of the weld puddle is being made uphill, and the larger piece may take more heat than the smaller part. You have to play the torch back and forth over both pieces and get a feel for how much heat each requires.

Further practice sessions for you may involve trying to weld pieces of tubing and angle-iron together, both typical materials for home/shop projects. Try joining a flat plate to a tube, or making a corner between two pieces of angle-iron. Careful measurement and cutting will be required to make gapless corners for a good weld. Square or rectangular tubing is often used in projects such as shop equipment or a utility trailer, and you should practice making 90-degree corners from two pieces cut at 45-degree angles. The inside welds will be the tougher ones, and you'll find when you stop practicing and assemble a real-world project that you can't always turn the work around to make the seam area flat in front of you on your bench. Often you will be forced to weld in a vertical seam, or even weld upside down underneath something. The latter is difficult and dangerous, ' requiring the best in full-face eye protection (a helmet instead of goggles) and a leather apron over your arms and shoulders to protect you from falling sparks. Welding an exhaust system under a car, especially if you don't have access to a lift, can be extra challenging.

When these don't seem like a challenge anymore, you're doing very well, and you can try making nice-fitting joints between pieces of *round* tubing. To fit two pieces of round tubing together, such as in a roll-cage for a race car, requires a measure-twice, cut-once procedure and lots of patience in cutting and grinding round tubing to join them with full contact. Another procedure to try with round tubing is to make a slip-fit, with one piece of tubing sized to just fit over another. In such cases, you can weld around the joint, or, for high-strength applications, cut the outside tube in a "fish mouth," as if it were being joined at a 90-degree angle to another piece of round tubing. When this piece is welded onto the inner tube, the fish mouth makes for a much longer weld seam between the two and gives greater strength.

Other types of welds you may run across in your projects are spot-welds and rosette welds. The former is found often in auto sheet metal work, where you join two pieces of light-gauge material together without welding the seam entirely. The spot-weld is just a round welding spot, rather than a bead, and is made when one piece is on top of another and you just melt a puddle that fuses the two panels together, then move a distance away and make another spot-weld. Filler rod is often not used in such cases, especially where appearance is more important than strength. Most spot welds lie flat enough that little prep is required to make them ready for paint. Modern cars are welded with thousands of spot-welds, but these are accomplished by special electric welders, usually computer-controlled and robotic.

Where a little more strength than a simple spot-weld is required, a rosette or plug-weld can be made. This is where the panel on top has holes drilled or punched into it at intervals. You weld up the holes, puddling both upper and lower panels together, which joins them securely in that spot. Plug-welds often do require some filler rod.

Sometime in your gas-welding work, the torch will go out and you'll get a loud pop at the torch which will really get your attention because you don't expect it. When welding in tight quarters, you may tend to get the tip too close to the work, which will overheat the tip and cause a tiny explosion as the gasses POP inside the tip, which is called a backfire. This will alarm you greatly when it happens, but the answer is simply to not get so close. You may have been working close to get the pieces hot enough, but if that is the case, you really need a larger tip or more flow of oxygen and acetylene from the tank regulators to maintain the right heat level with your torch further away from the confinement of the weld corner. Welding too close to the work will also dirty your tips. So if you have a backfire, stop work, shut off the gas at the torch and remove the tip for a

FLASH ARRESTORS
TORCH-MOUNTED TYPE
UL-LISTED
STOP DANGEROUS FLASHBACKS
L-TEC
L-TEC
L-TEC
WELDING & CUTTING SYSTEMS
OXYGEN and FUEL GAS PAIR
• Fit on all torches with "B" size inlets
• Oxygen inlet pressures up to 150 psig
• Acetylene inlet pressures up to 15 psig
• LPG, MPS, Nat. Gas inlet pressures up to 50 psig
• Both models incorporate reverse flow check valves
• Delivery pressure must be increased to compensate for pressure drop — see instructions
P/N 20357

3.30 These prevent serious accidents from a flashback, indicated by a loud squealing noise. Flashbacks are caused by gasses burning *inside* the hose. For safety reasons, both hoses should be equipped with flash arrestors, which are reverse-flow check valves. Some readjustment of gauge pressures may be necessary when using check valves.

thorough cleaning. Your welding set will have come with a set of wire tip cleaners. Cool off the tip with a water-soaked rag.

Something akin to a torch backfire but much more dangerous is a "flashback." This is where combustion actually follows the gas trail back into the torch and even the hoses. At the least there is danger of damaging the regulators and hoses, but if the combustion reaches the tanks, there is a possibility of a serious explosion. In a flashback, the torch goes out and there is a loud squealing or hissing noise with black smoke coming out of the torch. If this happens, shut off the oxygen at the torch, then the acetylene at the torch, then the cylinder valves. Your welding supply center should have available "flashback flame arrestors" that install in your two gas lines where the hoses attach to the regulators **(see illustration).** Acting like one-way valves, they are designed to prevent serious accidents caused by flashbacks, and they should be installed on any gas-welding setup, where the hoses attach to the regulators.

When making welds with filler rods, you will run across situations where the rod gets short enough for your left hand to get too close for comfort to the heat of the welding. Before making a bead, practice will have shown you how fast the welding rod gets used up and how long a rod you will need to complete the seam. Start with a rod long enough to do the whole seam. If the seam is longer and will take several rods to complete, you will have to stop welding and get a new filler rod. It will take some time and practice to get the hang of continuing a bead with a new rod where you left off before. The trick is to make the whole bead look like a continuous weld, without little piles of buildup where you stopped and restarted. When a rod gets short, many welders simply tack the short piece onto the end of a new rod and go right on welding. When you get good at it, the bead won't have a chance to cool off too much. In inside-corner welding, you may want to take the welding rod and put a bend into it, such that it puts your hand in a more comfortable position, further away from the heat of the action **(see illustration).**

Checking your welds

Chances are, you won't be making nuclear-vessel welds in your home/shop, or other welds so critical that they require X-ray inspection to pass specifications, as is often done in commercial welding. However, you do want to learn how to make strong joints, and your practice time should include some testing as well. Looks are important, but a good-looking bead may not be a proper weld, especially if it doesn't have good penetration. Checking for good penetration requires looking closely at the backside of the welded joint, which is easy for you to do with your practice sample pieces. In real-world situations, it isn't always practical to get at the backside of something you have welded (like a joint between two rectangular or round tubes), so you should be familiar with the proper heat control and penetration before you consider yourself through with your practice phase.

In looking at the back of the weld, see if there is any buildup on the backside. A small amount of "dropthrough" is acceptable, but 100% penetration can be

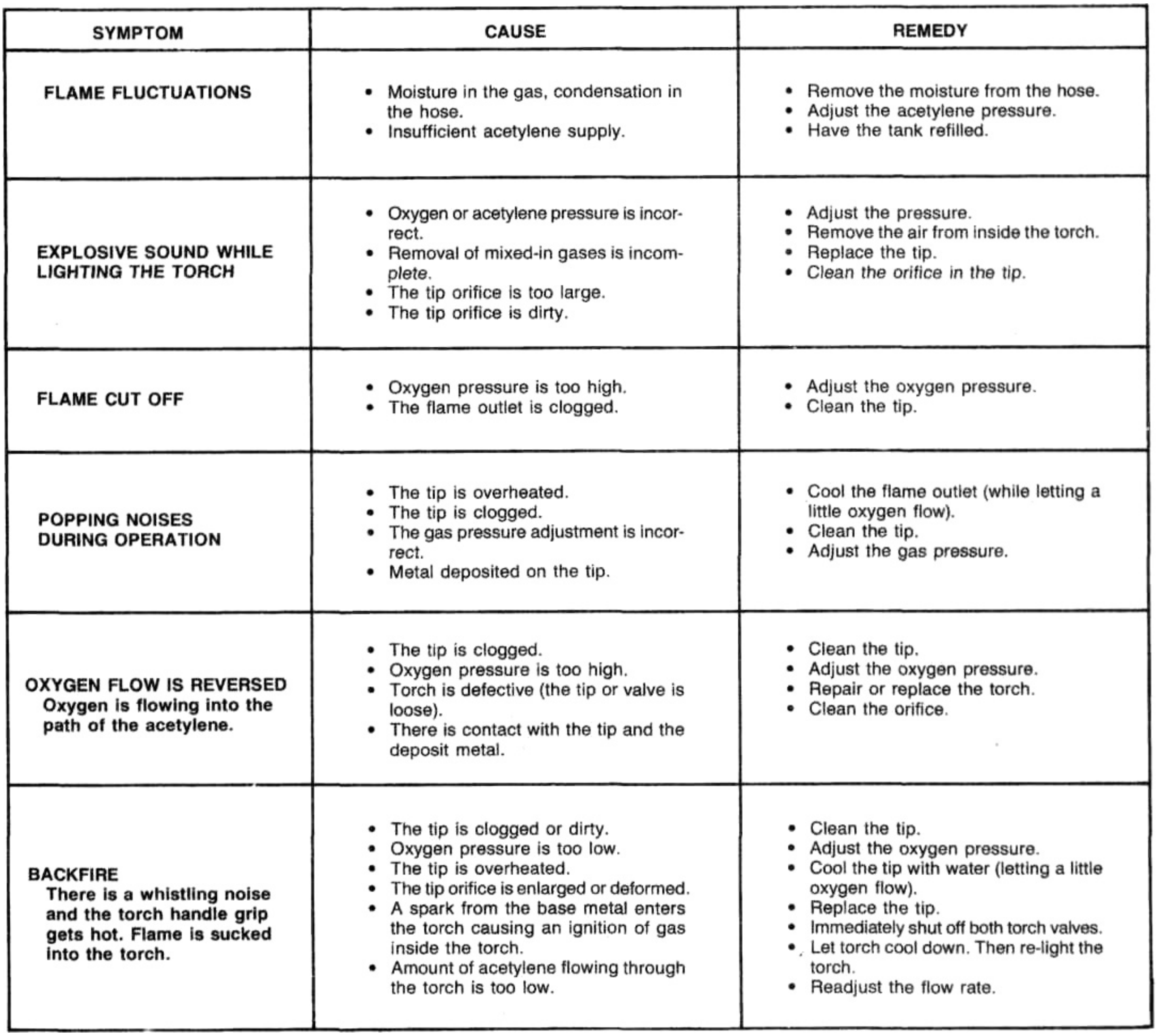

3.31 After you have gained some welding experience, and a good torch weldor isn't trained in one day, you will learn to read the welding conditions based on the flame from the torch. This troubleshooting chart indicates some of the causes **and cures of torch flame problems.**

SYMPTOM	CAUSE	REMEDY
FLAME FLUCTUATIONS	• Moisture in the gas, condensation in the hose. • Insufficient acetylene supply.	• Remove the moisture from the hose. • Adjust the acetylene pressure. • Have the tank refilled.
EXPLOSIVE SOUND WHILE LIGHTING THE TORCH	• Oxygen or acetylene pressure is incorrect. • Removal of mixed-in gases is incomplete. • The tip orifice is too large. • The tip orifice is dirty.	• Adjust the pressure. • Remove the air from inside the torch. • Replace the tip. • Clean the orifice in the tip.
FLAME CUT OFF	• Oxygen pressure is too high. • The flame outlet is clogged.	• Adjust the oxygen pressure. • Clean the tip.
POPPING NOISES DURING OPERATION	• The tip is overheated. • The tip is clogged. • The gas pressure adjustment is incorrect. • Metal deposited on the tip.	• Cool the flame outlet (while letting a little oxygen flow). • Clean the tip. • Adjust the gas pressure.
OXYGEN FLOW IS REVERSED **Oxygen is flowing into the path of the acetylene.**	• The tip is clogged. • Oxygen pressure is too high. • Torch is defective (the tip or valve is loose). • There is contact with the tip and the deposit metal.	• Clean the tip. • Adjust the oxygen pressure. • Repair or replace the torch. • Clean the orifice.
BACKFIRE **There is a whistling noise and the torch handle grip gets hot. Flame is sucked into the torch.**	• The tip is clogged or dirty. • Oxygen pressure is too low. • The tip is overheated. • The tip orifice is enlarged or deformed. • A spark from the base metal enters the torch causing an ignition of gas inside the torch. • Amount of acetylene flowing through the torch is too low.	• Clean the tip. • Adjust the oxygen pressure. • Cool the tip with water (letting a little oxygen flow). • Replace the tip. • Immediately shut off both torch valves. • Let torch cool down. Then re-light the torch. • Readjust the flow rate.

achieved without extra material building up on the backside. Good penetration is particularly important in welding auto body panels, because the outside of the weld bead is probably going to be ground down with a body grinder in preparation for filling and painting. Thus the strength of the topside buildup or puddling will be gone and the joint's strength will rely totally on the fusion of the two pieces below the surface. In your tests of practice welds, it can be helpful to use a hacksaw to cut through the center of one of your welded joints to see just what the penetration is. You can't always tell by looking at the ends of the joints, because of how you may have ended the bead.

We're assuming that your home/shop has a good, sturdy bench vise. It can be useful in testing your welded joints. Your first practice welds may have been straight seams with two pieces butt-welded together. Take your best example and put it in your vise with the weld seam close to the jaws **(see illustration).** Clamp a pair of vise-grip pliers on one edge of the upper piece and start hitting the upper piece with a hammer, while holding it with the pliers so it can't fly off. Hammer the upper piece until it's bent over all the way onto the vise (making a 90-degree angle in the welded pieces). Remove the sample from the vise and examine the weld. If the upper piece broke off during hammering, it obviously wasn't a strong enough weld.

Any of the various joints you may have practiced on can be tested by hammering. Whatever the joint, just secure the piece and hammer on one of the two joined pieces. Hammer from the backside area of the joint and toward the side that you welded on. If good penetration wasn't achieved, the backside will be weaker than the front and bending the metal this direction will really test the backside for good fusion. Even a weld with 100% penetration may not be strong enough if there was any "undercutting" on the top side. Undercutting is a groove along either or both sides of a bead, caused by applying too much heat or holding the torch at the wrong angle. In a stressed situation, the undercut area will be the weak point and the joint may fracture right along the undercut area.

3.32 Brazing with a gas torch has many uses, such as repairing castings and antique-car sheet metal. Here a flux-coated rod is being used. Brazing is not a fusion process, so it is used where higher, melting-point heat might damage the work, or surrounding materials.

Brazing

Brazing, or braze-welding, is not a true fusion process such as we have been discussing. The parent metal does not melt into a puddle. The parent metal is heated and the brazing rod melts into the joint securing the two pieces together. The brazing rod is made of brass (an alloy of copper and zinc) and is used in conjunction with a flux, which may come already on the rod, or in a can that you dip the rod into to coat it as you go. Brazed joints are as strong as welded joints when done properly and with an overlap of the pieces, but brazing is particularly helpful where the higher heat of fusion welding would harm the parent metal or adjacent materials, or where the pieces being joined are made of brass or copper.

The rod material melts at a much lower temperature than the parent metal, so the basic principle is to heat the parent metal just above the melting temperature of the filler rod, so that the parent metal, not the torch, melts the rod. Overheating the rod with the torch will boil the zinc out of the rod and cause poor adhesion in the joint. Brazing is most often done where there is an overlap of two pieces, or in building up a worn or broken area of a casting, that will be re-machined to original shape after brazing.

A joint to be brazed must be thoroughly *cleaned,* by grinding, filing, or sanding. When you have selected the proper rod for the material you are working on (copper/zinc for ferrous metals), you begin by heating the joint with your torch, using a lot less heat that you would for welding. You should pull the torch tip back 2-3 inches, which is much further back than with gas-welding. When the metal is dull red, apply the filler rod to the hot spot on the parent metal and see if it melts in. Heat next to that spot and apply some more fluxed rod, keeping this process going until you have completed your seam. When a long seam is contemplated with brazing, preheat the whole area. When the metal is the correct temperature, the brazing rod will flow nicely into the joint; if the metal is too cold, the braze material will ball up on the surface, and if it is too hot, it will spread out over a wide area. White smoke during the process usually indicates the area is overheated, and the smoke is the zinc boiling off. Not only does this affect the quality of the joint, but the fumes can make you sick if you breathe enough of them. Even if you don't overheat the rod, there are fumes given off during brazing.

Always braze with adequate ventilation, and a good respirator mask is also helpful insurance. Brazing is often used when joining pieces of galvanized metal because it will adhere better than standard gas-welding; however, this combination of brazing fumes and fumes from the plating or galvanizing should be avoided unless you are outdoors.

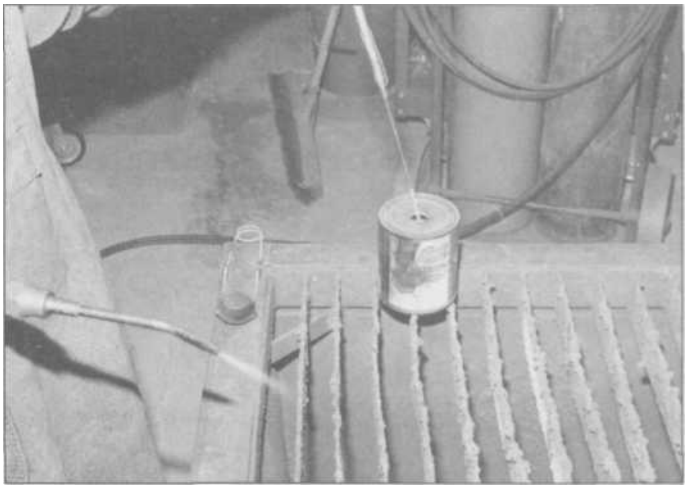

3.33 Most welders use dip-type powdered flux for convenience, since the pre-coated rods can flake if they are exposed to moisture. Plain brazing rod is heated with the torch, then dipped into can of flux and pulled out coated with enough flux for an inch or two of brazing travel.

If you are familiar with basic soldering, you know that good adhesion is ensured by using *flux,* a compound that prepares the metal for a strong bond, and you may also be familiar with the term *tinning.* Tinning is a process (in soldering) of applying a thin coating of lead to the fluxed surface before the main solder is melted in. The procedure is similar in braze-welding, except that, when brazing thin materials, the tinning usually takes place as you are brazing, in one step rather than two. Brazing larger joints requires that a light coating of brass be flowed onto the surfaces, making it easier for the main brazing to stick to. The flux for brazing can be a part of the rod itself (pre-fluxed rods) which is convenient, or you may have a metal can of flux on your welding table. You simply heat the end of your brazing rod and stick it into the flux and withdraw it covered with flux **(see illustration).** Once you have started the brazing process, the rod will always be hot enough to pick up flux as you need more. The pre-fluxed rods are convenient, but the coating is somewhat fragile and won't take rough handling. These rods must also be kept very dry.

When brazing thin sheet metal, the purchase area or overlap between the two panels is the key to the strength of the joint. When there is 3/4-inch or more of overlap for the brazing to adhere to, the joint will actually be stronger than the parent metal. Because the brazing rod melts at only about 1000° F instead of the steel's melting point of 2700° F, brazing will induce much less warpage in thin metal, which is why it has been used for years in many areas of traditional auto-body repair. The overlap in the brazed joint also makes the joint area much stiffer (two thicknesses of metal there), so a seam such as when installing a body patch panel will take less hammer-and-dolly work to prepare for painting **(see illustration).** Brazing will not work on a butt-joint because there isn't enough "captured"

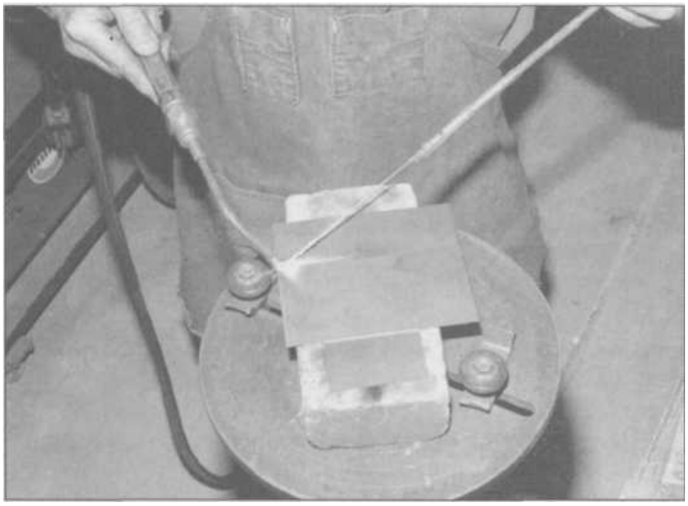

3.34 Lap welds are best for brazing, such as when joining overlapping sheet metal panels where the brass has lots of contact area for adhesion. As with fusion welding, tack the ends before running the whole seam.

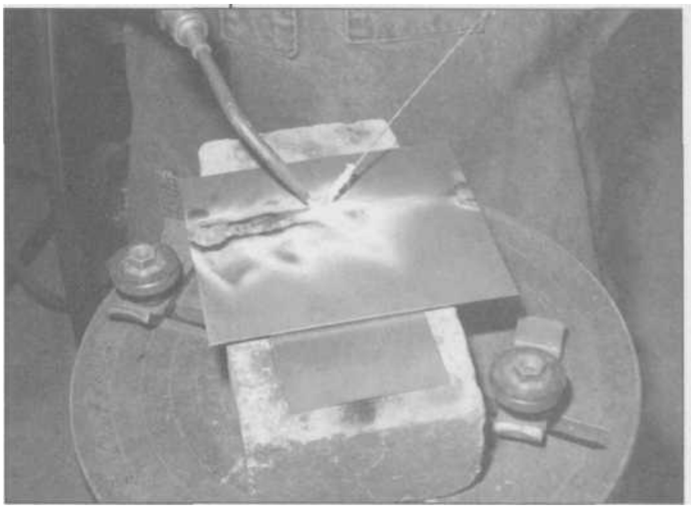

3.35 You should start with very clean metal when brazing, although the flux is designed to surface clean as you go, leaving some minor slag afterwards. You will have to keep the torch further away from the seam than with welding, to keep from overheating the brass.

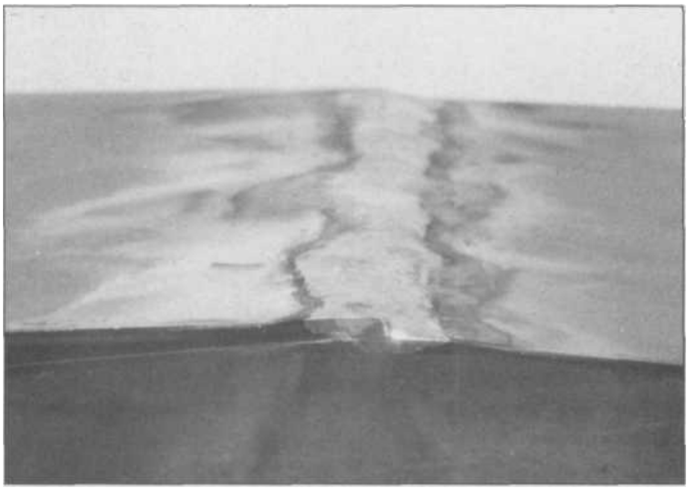

3.36 End view of a brazed lap joint, showing how the brass flows in between the plates with capillary action, much as solder does when applied to a hot copper pipe. The base metal must be heated to the point where *it* melts the rod, not the torch.

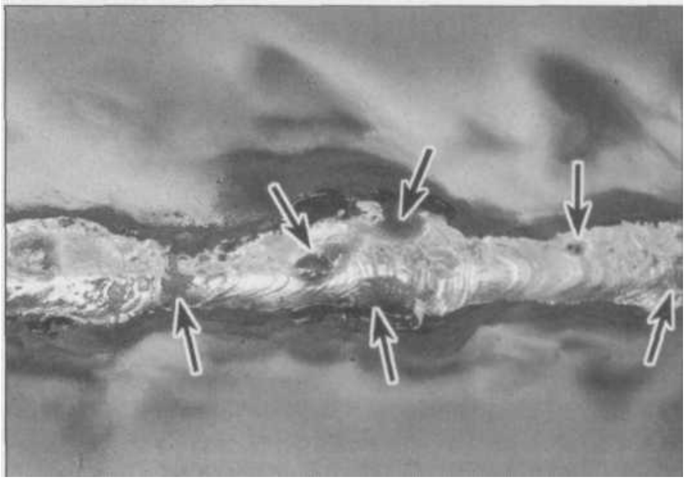

3.37 There is some silicon slag after brazing (arrows), which can be softened with water after the part cools enough so that water will not warp it, then the rest comes clean with wire-brushing.

area for the capillary action to flow the brass into each piece, although you can back up a butt-joint with a strip of similar material, which gives an overlap on both pieces and good bonding area for the brazing.

Although brazing is still a technique used by older body craftsmen in restoration of antique and collectible cars, it should not be used on new cars that are built of low-alloy, high-strength but lightweight steels. In fact, some car manufacturers specify that no torch work at all should be done on their body metal, that only MIG welding is acceptable because of the narrow heated area and faster welding-bread travel.

Other metals than steel can be brazed as well, such as stainless-steel, castings, brass, bronze and even aluminum when a special aluminum-brazing rod is used. If you have special project considerations or a question, ask your local welding supply for advice on choosing the best brazing rod for the job. One advantage of brazing is that dissimilar metals can be joined, which ordinarily can't be done with fusion welding. Also, where two pieces of different thickness are being joined, like a thin pipe to a thick flange, brazing is preferred. In ordinary gas welding, you have to be very good with the torch to play the right amount of heat onto the thicker piece while not burning up the thin piece. With brazing, neither piece needs to be heated beyond the melting point of the brazing rod.

After braze-welding a seam, you will see that the flux remains as a kind of crust on the parent metal. It should be removed with a wire brush, and can be softened with water first, which makes the wire-brushing work easier.

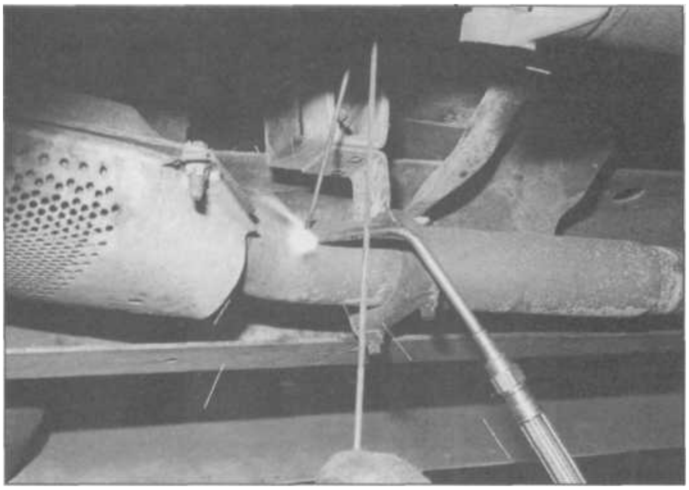

3.38 The versatility of gas welding equipment has always found favor in the exhaust system business. You must be careful not to cut or weld near fuel tanks, fuel lines, electrical components or brake lines, but the equipment is highly portable, and can cut as well as weld. Long welding rod can be bent to curve around exhaust pipes to weld in tight quarters.

Oxy-acetylene cutting

The other side of the dual-purpose nature of gas equipment is the ability to *cut* as well as join metals. The gas torch can make short work of cutting through even large or

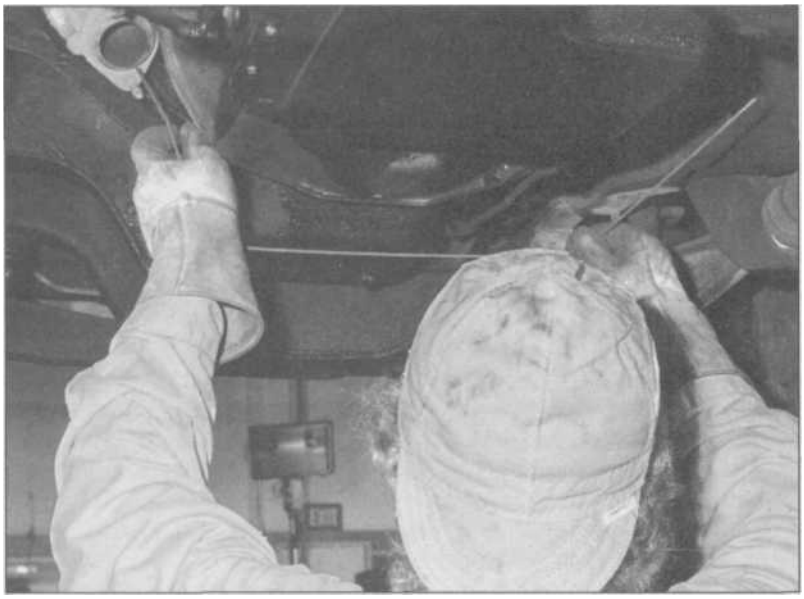

3.39 Muffler installers also find uses for welding rod to simulate the length and required bends for a pipe to be installed. Bent wire is taken over to the hydraulic tube bender where shape is duplicated in pipe.

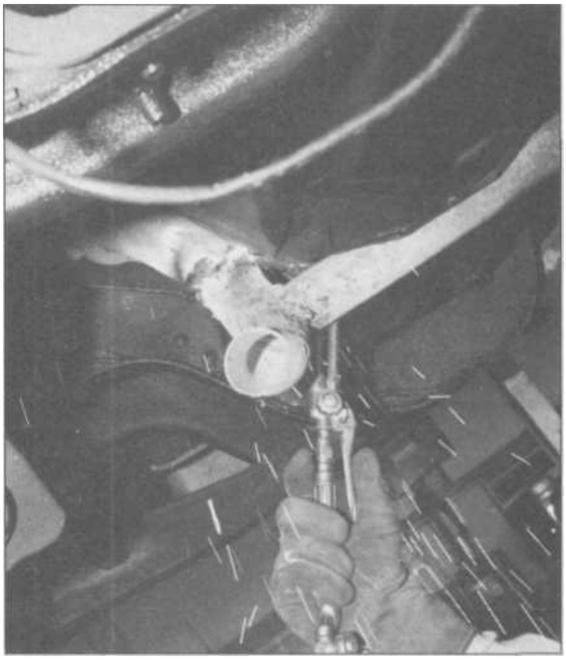

3.40 Sometimes, the oxy-acetylene cutting torch the fastest, easiest way to remove the old exhaust system.

thick pieces of steel. For sheet metal work you will probably continue to use shears and other hand tools to cut light-gauge material, but being faced with cutting through a half-inch steel plate with a hacksaw will easily convince you of the value of a cutting torch.

The only oxy-acetylene cutting component different from your basic welding setup is the torch itself. The cutting torch has one major difference from the welding torch: it has an extra oxygen supply, operated by a lever instead of a valve **(see illustration).** Actually, when you cut steel with a torch, it isn't the flame that is really doing the cutting; it is a strong flow of oxygen that hits a heated part and the oxidation of the metal happens so fast the affected area disintegrates into flying sparks. The basic principle of oxy-acetylene cutting is that the torch tip has several small holes surrounding a central larger hole. The smaller outside holes are where the basic oxy-acetylene flame comes out - actually several small flames. These flames are really used to preheat the metal to be cut. The weldor adjusts the gas and oxygen flow on the torch for the proper flames from the pre-heat holes and plays the torch on the end of the piece to be cut. When the end is heated to a cherry red color, the weldor presses down on the cutting lever, which activates a strong flow of pure oxygen through the large central hole in the tip. This is the oxygen that really cuts through the metal.

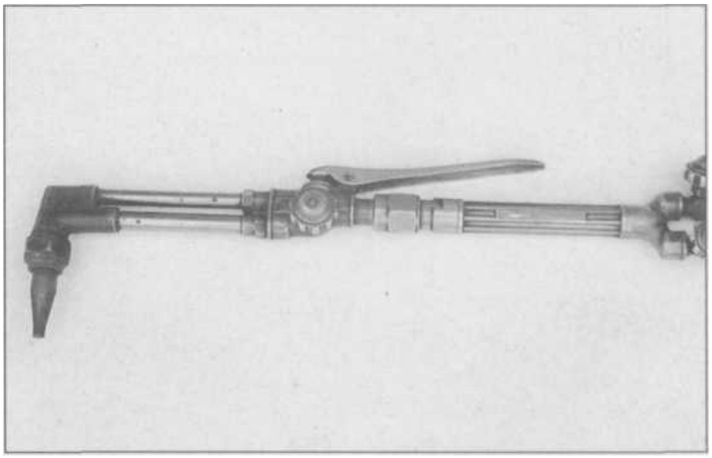

3.41 The cutting torch is adjusted for a good flame, then checked with the extra-oxygen supplied when the cutting lever is depressed. The flame should remain close to neutral. The excess oxygen is what is used to push the molten slag from the kerf as you cut.

Because oxy-acetylene cutting involves a huge shower of sparks from the underside of the metal you're cutting, you must be prepared before cutting. Clear the area around your welding table, have the proper welding attire on (including leather welding gloves), and make sure the area under the test piece is clear of any obstructions. It's best to support your test piece on some firebrick, leaving the area under the cut line open. If there were any restriction to the flow of sparks under your test piece, molten shrapnel could blow back into the weldor's face.

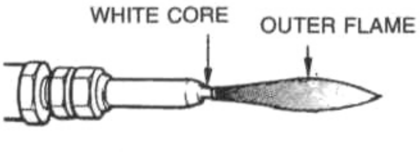

Neutral Flame

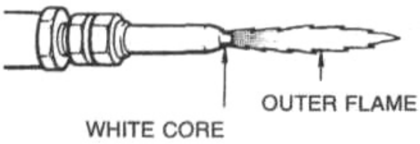

3.42 As in welding, the flame of your cutting torch head should be adjusted to a neutral flame before cutting.

Mark your cut line using a sharpened piece of soapstone and a straightedge. After marking the cut line, find a piece of tubing or angle-iron that you can clamp down onto the workpiece to act as a guide for the torch head **(see illustration).**

Before you begin cutting with your oxy-acetylene torch, you should reset the regulators. Adjust the acetylene gas to 2-3 psi, and the oxygen to 15-20 psi, since a much larger percentage of oxygen will be used in the process. Select a gas pressure and tip size that correlates to the thickness of the material you are cutting. Your setup's instructions will tell you the recommended sizes for that brand of equipment. Light your torch as you have learned to do in gas welding practice and watch for a neutral flame from the preheat holes in the tip. Watching the flame pattern, try cutting in the extra oxygen flow from the cutting handle. It shouldn't affect the quality of the preheat flames. Place the torch flame at the edge of the material you want to cut and hold it there a short distance from the material (1/16-inch or 1/8-inch away). When the metal is cherry red (about 1500° F), angle the torch tip back about 10 degrees (with the tip aiming in the direction of the cut) and cut in the oxygen. You'll find that oxy-acetylene cutting is much easier to learn than the welding process was; the main thing to learn here is the proper travel speed for the tip size, gas pressure and metal thickness you are working with.

When you cut through metal with the torch, the quality of the cut edges will tell you much about the process. At the bottom edges of the cut, slag may collect, leaving a rough edge that must be cleaned up with a grinder later. Also, remember when you are cutting that the torch takes out material as it goes; the part being removed is called the *kerf.* When you draw soapstone lines to cut out shapes, remember to cut *outside* your lines, or the part will wind up being too small. After your have experimented a little with the torch, you'll have an idea of how big the kerf will be with certain tips and certain materials.

If you try to advance the torch too fast, the bottom of the kerf may be irregular or not cut through all the way. When you look at the edge of the cut metal, you'll see vertical lines in the edge of the cut, in a good cut these lines are straight and smooth, in a rough cut the lines are almost like teeth on a saw blade. If you advance the torch too fast, the cutting lines will look curved back toward the direction you started the cut from, instead of straight up and down **(see illustration).** If you move too slowly with the torch, the top of the kerf will be rounded off, and the bottom edge may accumulate too much slag. If you have someone else in the shop with you, have them observe your cutting from a safe distance. When you are making good cuts, the shower of

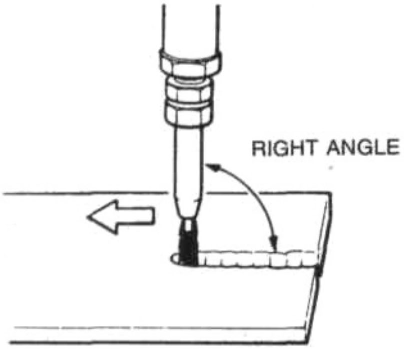

Cutting Thick Panel

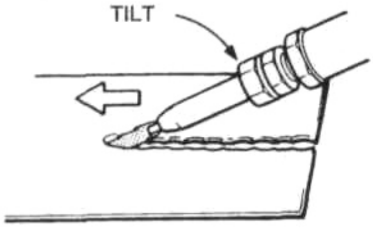

Cutting Thin Panel

3.43 The cutting torch should be held more nearly vertical to the work when cutting heavier metals, and angled back when cutting thinner material like exhaust pipe or sheet metal. Use scrap steel tubing or angle as a brace to guide the torch for steadier cutting.

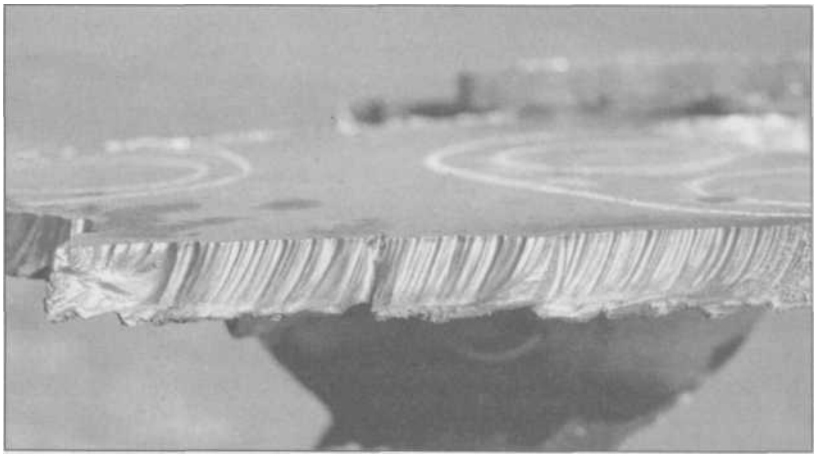

3.44 Examine the cut edge of plates after you have used your cutting torch. The proper heat and travel speed is when there is the least amount of slag along the bottom of the cut. This hand-held cut was moved a little too fast, as evidenced by the angled cut lines.

sparks should come straight down to the floor. Your observer can tell you if the spark flow is going back toward the beginning of the cut, which indicates that the metal isn't being cut all the way until after the torch has moved on, the result is that the uncut area at the bottom directs the sparks back toward the beginning of the cut. Some of the sparks and blobs could also get blown back into your face or clothing, so learn to develop the right cutting speed for clean cuts.

You will find that the steadier you are with the cutting torch, the better the cuts will be, and half the problem of cutting is finding a comfortable position for your hands. Since only one hand is required to operate a cutting torch, you can use your other hand (gloved) to steady the torch and help advance it correctly along your cut line. Keeping the torch tip the same right distance from the work is also important, so find a piece of angle-iron or tubing in your shop sized such that, when you rest your torch neck on it, it puts the tip about the right distance away from a piece of steel the angle or tube is lying on **(see illustration).** Mark and save this piece as your torch rest and it will be invaluable in making straight, clean cuts.

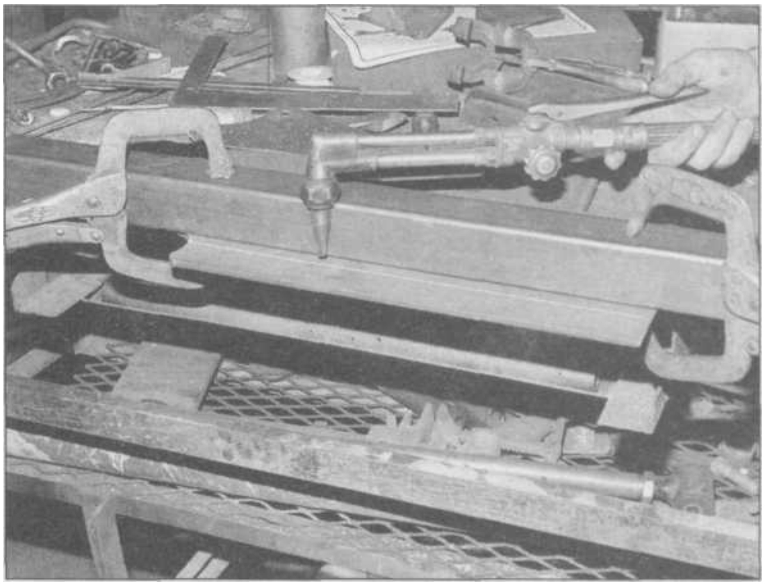

3.45 Clamp a piece of metal to your work that spaces the cutting torch head just the right distance to make your cut-line. The steadier your torch is held, the cleaner the cut. Travel at a very steady speed.

There are many methods for cutting steel for projects, like saws and abrasive discs, but the versatility of the cutting torch is that you can cut out a piece of one-inch plate in the shape of an elephant if you want to. Irregular shapes cut out just as easily as straight lines, as far as the torch part goes, though you will have to develop a rhythm of smooth hand movements with the torch to follow your soapstone lines accurately. As with any aspect of welding, this will take practice. You can also stack several pieces of plate together. Either clamp them together or tack-weld the edges, and when you cut (sizing the tip for the combined thickness) you can cut out as many as a dozen identical pieces. To keep them truly identical if necessary, these parts can later be tack-welded, clamped or bolted together for the final sanding or grinding of the edges, ensuring that they all match in dimensions.

Besides making straight cuts, which are easiest, you may find that some projects require certain shapes to be cut out more than once. Here is where various cutting aids can be used. If you can make the torch tip follow a pattern of some kind, you can cut out the same shape repeatedly. Perhaps you need four round discs of steel six inches in diameter. You can make a steel or wooden pattern that has a circle in it bigger than six inches. How much bigger than the six-inch measurement depends on the size of the kerf with the tip you are using. If you know that you need an extra 3/8-inch to allow for the kerf, make a template with a hole 3/8-inch bigger all around. When you use this as a cutting guide/rest for the torch, you can cut out as many circles as you want and they will all be round and the same size. You clamp the template over your workpiece and move the torch around your cut, always in contact with you template. The smoother your template and the steadier your hand movement, the nicer the cut will be. Watching a pro weldor some time will show you how perfect you can get with what seems like a relatively-crude method of making a part.

In fabrication shops, special equipment is used to make repeated cutout shapes where lots of parts are needed. One of the basic machines you would see in most street rod or race-car shops is a *"flame-cutter"* **(see illustration).** This is basically a cutting table on which is

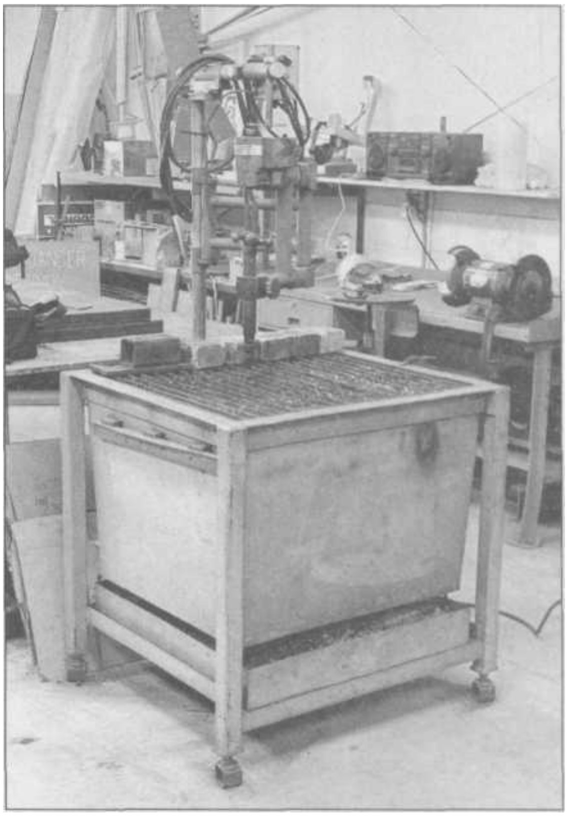

3.46 A tool you will see in most race car, hot rod and fabrication shops is a mechanized cutting torch, often called a flamecutter. It uses a motor-driven torch to travel around a template and cut out exact shapes. Templates can be temporary ones cut from heavy cardboard, to steel templates for frequent use.

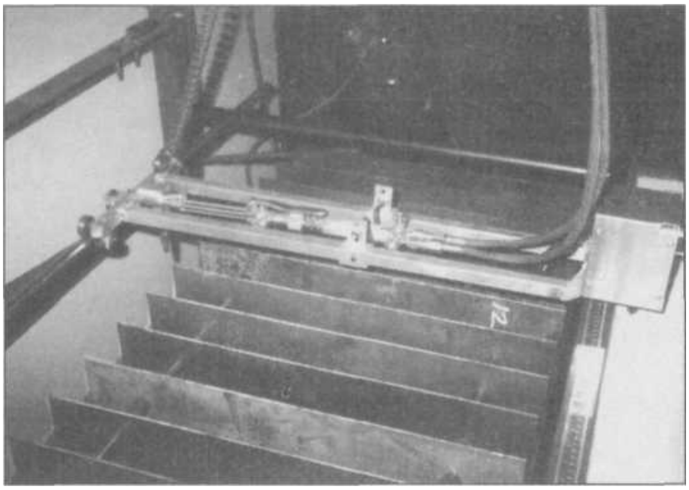

3.47 This mechanized cutting setup from Daytona MIG uses a standard cutting torch, laid down and clamped in place.

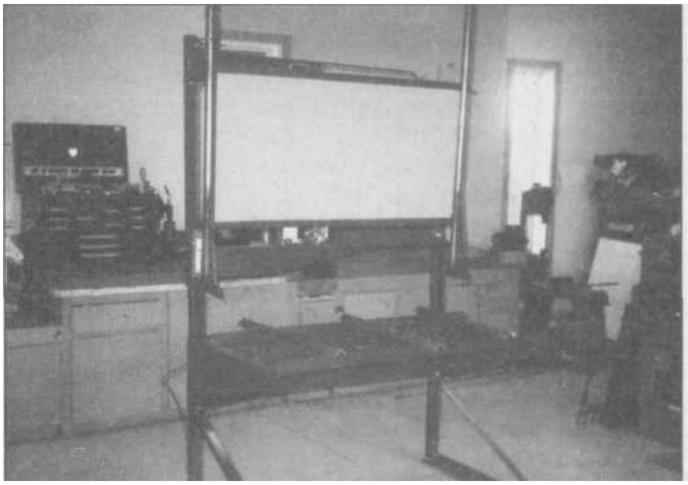

3.48 The complete Daytona MIG shape-cutting table is shown here with the erasable porcelain layout table folded up to take up less space when not being used. This table can be used with a gas torch or plasma cutter.

mounted a framework that holds a special torch, with a straight cutting end instead of a 90-degree end like most hand torches **(see illustration).** There is a series of arms that allow the torch to stay perfectly vertical and yet move to follow any shape on the table, and at the top is a small electric motor. There is a place at the top to mount a template, usually made of heavy-gauge sheet metal, and a magnetized, knurled tip on top of the torch assembly that follows the template exactly. The motor drives the knurled tip around the template at a steady speed (adjustable for the thickness of the material being cut), while below, the torch tip cuts out exactly the shape of the template. Removing the human element from the torch movement means that a flamecutter can make smoother, more repeatable parts. In most shops with a flamecutter, there is usually a large wall behind it on which hang a wide variety of templates for frequently used shapes, such as suspension brackets **(see illustration).** In some shops, the flamecutting setup has *a pantograph* arrangement, which is a series of arms that make the torch cut out a shape dictated by the movement of a stylus around a template on a table next to the flamecutter. Some sophisticated machines have an electronic stylus that will follow the lines of a blueprint or even the lead of a pencil line drawn on paper.

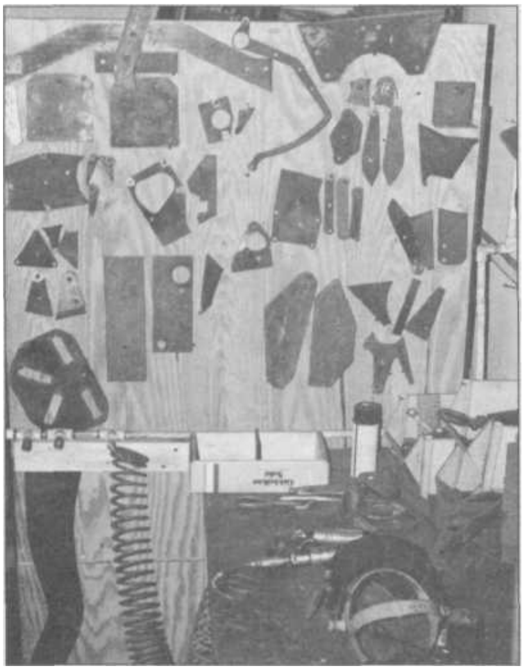

3.49 Wherever you find a flamecutter, chances are you'll find a wall of metal templates nearby, used to easily and precisely duplicate commonly used shapes such as suspension mounting brackets. At lower left is a shape used to make a head for an engine stand.

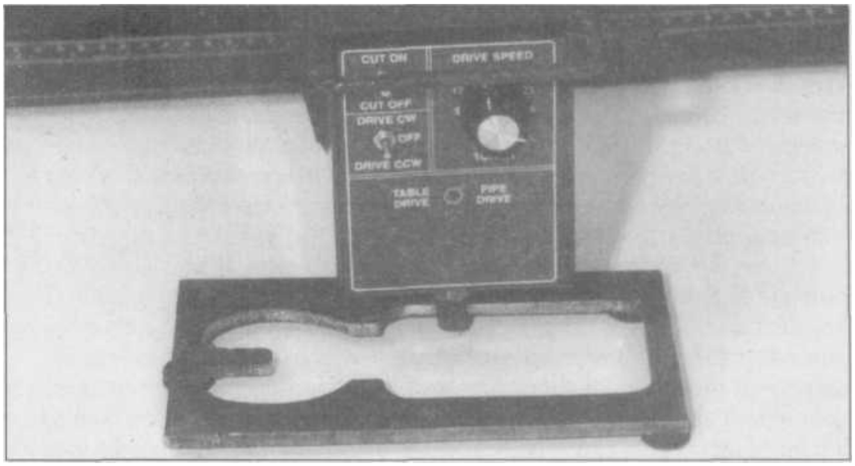

3.50 Most mechanized cutters have a drive head with a magnetized tip for following metal templates, inside or outside a pattern. The cutting speed is infinitely variable to suit the material.

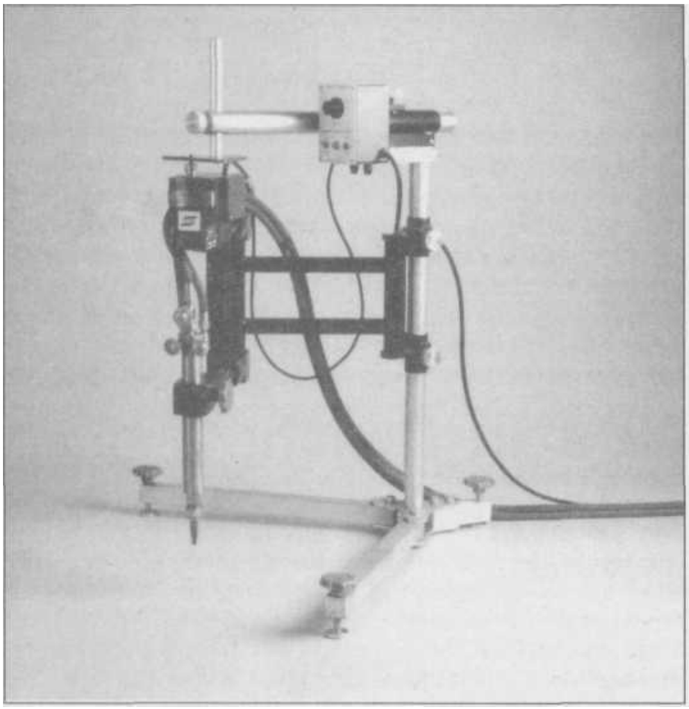

3.51 The ESAB Porta-Graph can be moved anywhere you have a source of electricity, and is useful for cutting shapes out of large steel plates that won't fit under a stationary cutting machine.

3.52 This flamecutter with swing arms can be mounted over a plate, or bolted to a table.

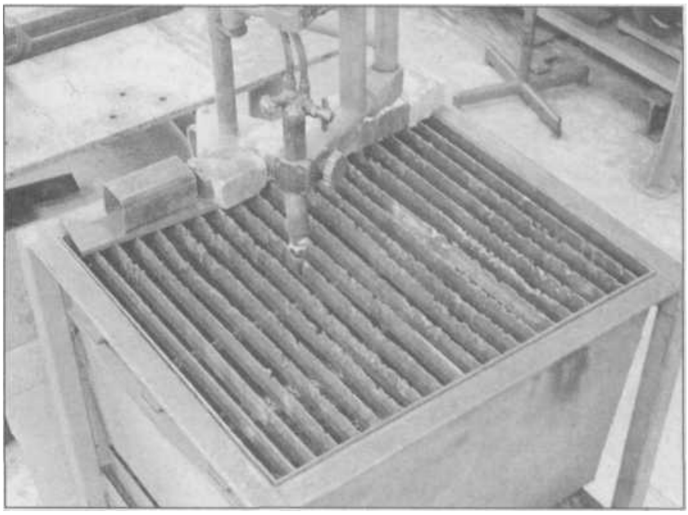

3.53 Under a mechanized cutter, you need a grate like this to support the work while it is being cut. After some usage, the 1/4-inch by two-inch slats will be fairly cut up, but can be pulled out and swapped end for end, top for bottom or simple replaced.

If you decide to cut out shapes with your oxy-acetylene torch, you will probably need to build a cutting table **(see illustration).** On a straight cut, the end of the part can be laid over the edge of your bench; the work is supported and the cutoff piece just falls to the floor. But cutting out shapes larger than a few inches means you need a different surface to work on, or you'll soon be cutting into your bench! Look at the photos of some of the cutting tables pictured and you'll see how to construct one. A framework is made that holds long strips of 1/4-inch steel straps in an edge-up position, spaced an inch or two apart. The framework should be placed over some kind of fireproof container, like an old drum (sandblasted clean of any flammable residues first). This will capture the flying sparks and hot scrap pieces and keep the shop neater and safer. Another safety consideration to watch for is protection of your feet. Make sure when you are cutting out a heavy part that it isn't going to fall right on your feet. Either have a metal deflector under your cutting area, or keep your feet well away from where the very hot part is going to fall.

There will be times when you cannot start a cut at the outside edge of your workpiece, and you will have to cut from *inside* a pattern. This requires a little different technique to get the cut started. Adjust the torch to a neutral flame with the cutting lever depressed. With the lever *off,* start heating a spot on your workpiece along the cut-line. When the molten puddle starts to appear, pull the tip back about a half-inch (because this is a closed spot, there is a chance of molten metal and spark being driven back into the tip if it is too close) and slowly cut in the cutting lever. When the spot becomes a complete *hole,* you can proceed around the rest of your pattern or line.

Some parts cut with a torch need to have a beveled edge if they are going to be joined with another part by welding, such as when joining sections of heavy pipe. Cutting a beveled edge on a round pipe by hand is challenging, and in the pipe-fitting industry they often use a machine that clamps around the pipe. The fixture holds a cutting torch, and a motor drives the assembly around the pipe smoothly and evenly, and is adjustable for speed and the angle of the bevel. There's no place for such equipment anywhere but in the pipe business, but it's interesting to see the ways in which the human element of hand control of the torch can be eliminated with special equipment. When you do make beveled cuts in thicker metals with your torch, you should remember to size the torch tip to the measurement *across the bevel,* which is thicker than the straight thickness of the metal.

3.54 The rosebud tip will be useful for heating metal parts to be bent, freeing press-fitted parts, and for preheating castings before welding, such as here where a cast-aluminum mailbox post is preheated before TIG welding.

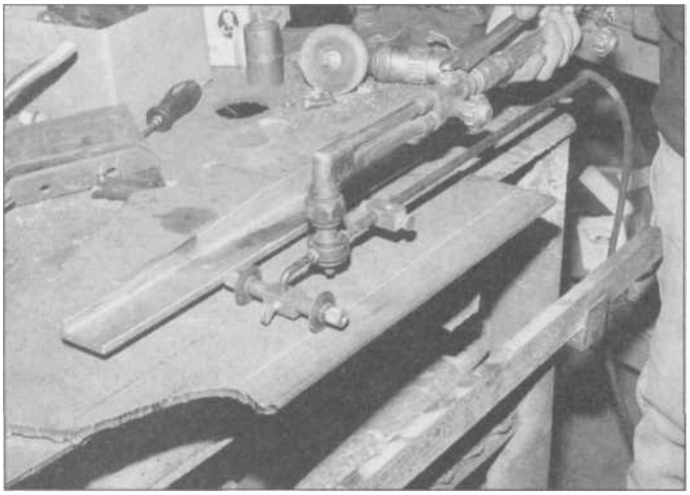

3.55 A great aid to cutting straight lines manually is this "compass" guide. It holds the torch tip firmly, but in a swivel arrangement that allows it to turn as you move along the line, and the torch stays the same distance from the work and much steadier. Here it is being demonstrated against another great tool, a magnetic steel guide that sticks strongly to the work and saves the hassles of setting up a bar or tube with some clamps.

Heating with oxy-acetylene

You will be surprised how easily a piece of heavy steel can be bent into a shape when it is properly heated with a torch. Even a thick piece of steel, heated to a cherry red color, can be bent like a sheet of soft copper. Your bench vise will come in very handy for projects like this. The old-time blacksmiths hand-shaped everything they made the same way, except that they had to heat the part in a coal or wood fire, while you can do the same thing in seconds with an oxy-acetylene torch.

Depending on the thickness of the metal and the area to be heated and bent, you can use a welding torch or cutting torch to heat up the workpiece. It shouldn't take too long to get the area cherry red; if it does, use a larger tip on the torch. There is another type of tip that can be used called a *rosebud,* that is expressly designed for heating. If your oxy-acetylene set was a deluxe model, such a tip may have been included, or you can easily get one at your welding supply center.

The rosebud tip has a much larger end that flows a huge amount of gas. You will have to crank your tank regulators up to perhaps 25 psi on oxygen and 10 psi for acetylene to feed this device (never exceed 15 psi on the acetylene regulator). The big, wide flame of the rosebud tip is ideal for spreading heat, and lots of it, more evenly around a part than with the tighter flame of a welding or cutting torch tip, although those can be used to heat small areas, especially if you are just freeing frozen, rusted fasteners.

The most common usage of the rosebud tip is for heating components that are press-fitted together. Heat expands metal, so logically, if you can heat up the exterior part while not heating the interior part, the heated part will expand and can be removed (with tongs). This is most commonly done in automotive or industrial situations where there is a gear pressed on a shaft, or an axle bearing must be removed. In some situations, however, you cannot simply blast the area with uncontrolled heat, or damage may occur to one of the components. You need to apply only as much heat as it will take to separate the parts, no more. Your welding supply shop should have a variety of paint-marking sticks or crayons that are designed to melt or change color at specific temperatures. One brand, called Thermomelt, is available in 87 different temperature ratings, from 100° F to 2200° F, and in a liquid, stick or pellet form. You use this product on the part you need to protect, and only heat the area until the product melts or changes color, then you work on separating the components without adding any unnecessary heat. These products are also useful when tempering or retempering tools that have been reground, which is a lengthy subject unto itself.

Deciding which type of welding equipment to buy is always a tough decision, but one that is made easiest by carefully defining your needs before buying any equipment. When you have read all the chapters in this book, you should have a good idea what the benefits and drawbacks of each system are, and what is best-suited for your work. Any weldor will tell you that the versatility of the gas-welding/cutting setup becomes valuable when you need it. Even when most of your work is done with some sort of electric-welding equipment, you'll find so many handy uses for oxy-acetylene equipment that if your budget can handle it and you have enough work to utilize it, a great idea is to have the gas equipment and some form of electric welder, too.

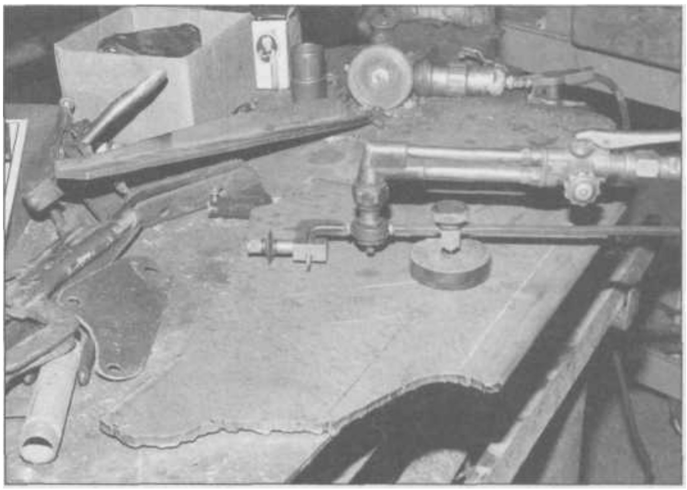

3.56 The same tool can be used, with different attachments, to cut a perfect circle of any size from a few inches to two feet in diameter. Removing as much of the human element of hand control makes for better parts that require less grinding to finish.

Notes

4 Arc welding

One of the oldest forms of welding, arc or "stick" welding offers versatility, strength and the ability to handle big projects and thick materials. Also, introductory equipment can be purchased inexpensively. There are some rather expensive arc-welding machines, but the most common AC-current machines will do fine for the home/shop user doing standard projects and repairs.

If you have thoroughly read Chapter 2, you should have a good understanding of the various kinds of welding equipment and what each type is capable of. If you think the arc welder serves your needs, then this chapter will give you a basic introduction to the equipment and procedures involved. When shopping for a machine, you may see them referred to as SMAW welders, which is the technical description of the process and stands for Shielded Metal Arc Welding.

Basic-yet-rugged arc machines are available today in welding equipment stores, major nationwide stores such as Sears and Montgomery Ward (and their mail-order catalogs) and can sometimes be found at a good savings in large lumber/home supply centers that carry a lot of power tools. Most starter units are sold with a complete setup, including welding gloves, helmet, chipping hammer, sample electrodes, and a basic instruction book. Other than the consumable welding rods, there will be little else you will ever have to buy to continue arc-welding.

4.1 Arc welding will always be with us, as the best welding method for joining heavy plates for commercial, farm and industrial work. The weldor requires considerable protection from the spatter and radiation, usually wearing heavy leather protective gear.

4.2 For home and farm use, the Lincoln 225 is the classic AC buzz-box welder, having been in production relatively unchanged for decades. Most home/shop units are AC only, and require some rewiring to accommodate the required 220V input.

4.3 Another popular AC 220V welding machine for non-industrial use is this Miller Thunderbolt 225. The crank handle on top changes the output amperage, which is read on a scale at the top left of the front panel.

You can find arc machines priced from less than $200 to professional units costing several thousands, depending on the features, but most home/shop projects and farm repairs can be performed with machines at the lower-priced end of the spectrum. There are really no moving parts involved in the machinery, and most of the name-brand equipment is rugged enough to last for many years of service. The larger machines offer certain advantages for application in professional shop use only, such as higher amperage, AC/DC switching and higher duty-cycles.

The principle of arc-welding is to attach a ground cable to the workpiece, set the machine for the correct amperage based on the thickness of the material, fit a consumable electrode (welding rod) in the electrode holder, and with your helmet down strike the rod against the work to start a flow of current, the arc, that produces intense heat and light and welds your seam together. Welding thicker materials requires more heat in the form of higher amperage from the machine. The basic arc machines of interest to you generally have an amperage range of 40-225 amps. The arc process is best suited for thicker materials, but you will probably *never* use your machine at the higher settings. Even when welding on materials one-inch thick (which you will most likely never encounter in your home shop), or when repairing large castings, the seam isn't completed all in one big bead, it requires several passes on overlapping beads to totally join the parts.

At the same time, you will probably never use your AC welder at the lowest settings. For materials thinner than 1/8-inch, there are other welding processes more suitable, such as gas, heli-arc or MIG-welding. Most of the basic welding performed with AC arc-welding machines is done at 90-125 amps, on materials from 1/8-inch to 3/8-inch and sometimes 1/2-inch.

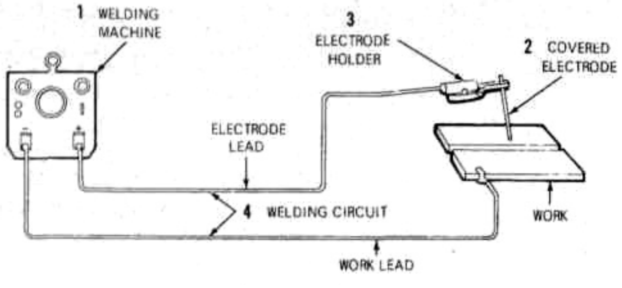

4.4 The layout of the arc-welding setup is simplicity itself. The rod (also called a stick, stinger or electrode) is both the source of the arc and the shielding gas, produced as the rod melts its flux coating.

Comparing duty cycles

What you will look for in an arc machine, therefore, is not the highest amperage it offers, but the *duty-cycle* at the 90-125-amp ranges you will use most often. We discussed the duty-cycle in chapter 2, but it bears repeating in brief. All electric welding machines have a duty-cycle rating, which refers to how long they can weld at a specific output without overheating. The duty cycle is described as a percentage. If a machine has a 100% duty cycle, it means that it can be operated virtually all day, except to stop when you change electrodes. The percentage is actually based on a ten-minute period as a test, meaning that if the duty-cycle was 50%, the machine could be used for five minutes out of ten. You could weld nonstop for five minutes and then "rest" the machine for five before starting again.

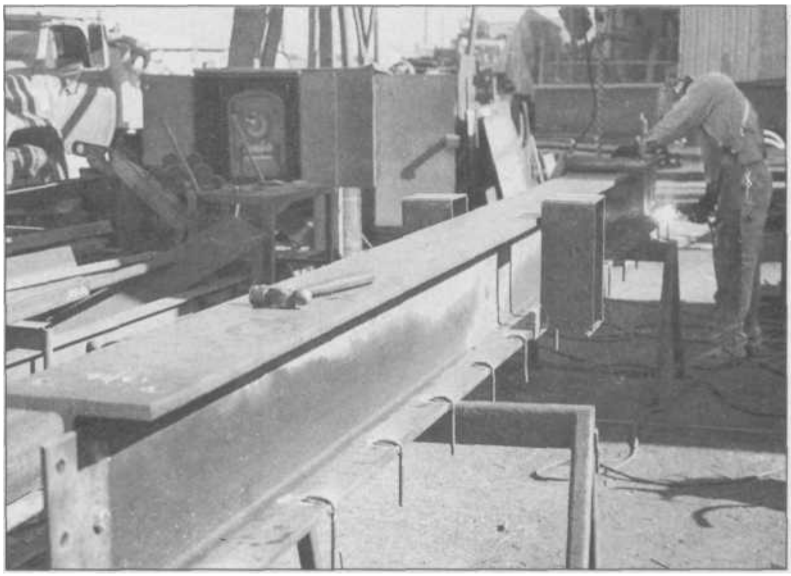

4.5 Virtually any building over single-family-home size requires structural steel in its construction, and that requires arc welding. Here a large beam is being prepared for earthquake retrofit in a public school.

The duty-cycle rating confuses most first-time buyers, because different companies may take their published ratings from different amperage settings. The higher the amperage you weld at (for thicker material), the less the duty cycle will be. Obviously, the machine will overheat quicker at the higher amperages. On every machine's range of amperages, there is probably a point where the duty cycle is 100%. A specific machine may have a 50% duty cycle at its highest setting, yet have a 100% rating at the mid-range settings you will use most. It's important when shopping for a welder to find out what the ratings are for the high range *and* the mid-range. In the larger professional machines designed for welding shops and production-line work, the duty cycle must be close to 100% for any situation, and thus these machines have to be built with much more expensive components and reserve capacity.

Some of the duty-cycle rating comparisons may be academic for the average home/shop user. If the only long seams you weld are on thin materials at lower heat settings, most machines will be fine for your purposes, and there is no reason to spend many times more for the machine to get a higher duty rating. In a typical home/farm project, say where you are welding together steel tubing to make a utility trailer, the *setup* of the work takes half or more of your time anyway. Each welded joint between pieces of tubing may only take one or two minutes of actual welding, and then it may take you four minutes or more longer to set up clamps on the next joint or flip the work over before you're ready to make another weld. The point is that, in this example, a welding machine with a 40% duty-cycle at the amperage you were using would be perfectly adequate for the job. You wouldn't find yourself being slowed down on the job because you had to weld and then wait for the machine to "catch up." However, there is something to be said for having some duty-cycle "cushion," and if you were considering two machines in the same price range and featured similarly, you would take the one with a higher duty-cycle at the heat you would most often use it. Generally, the machine with the higher duty-cycle at a mid-point amperage will also have higher ratings at other settings.

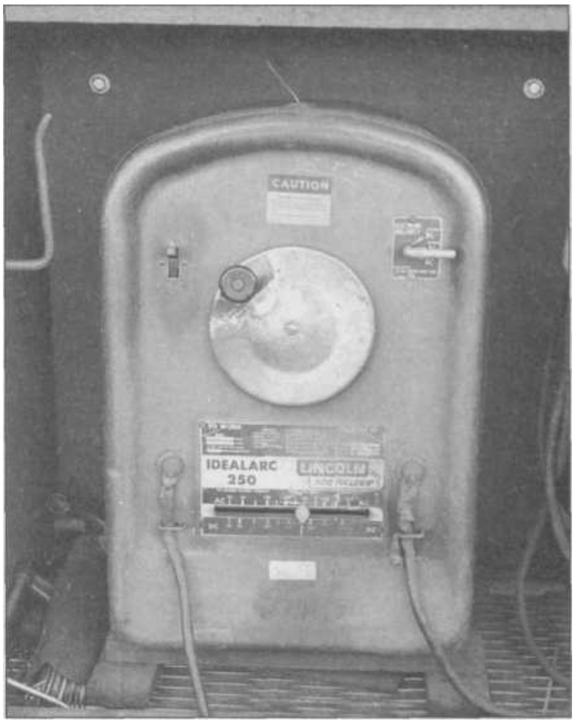

4.6 In industrial use, the arc welding is generally all DC, and you will see these venerable Lincoln welders performing their duties day in and day out for years. Most industrial machines have a 100% duty-cycle for production work. This one is housed outside in a steelyard, enclosed in a weatherproof box.

Miller
XMT 300 CC/CV
DC INVERTER ARC WELDER
AUTO-LINK

4.7 In larger welding shops, a more sophisticated power source is used, such as this 300-amp **Miller DC machine, which has a few more "bells and whistles" than the home-type units.**

AC, DC or both?

Most of the basic machines in stick welding are AC-powered. For those readers who may have taken electricity for granted since high-school physics classes, AC refers to *alternating current,* which is what we have in our houses, businesses, and power lines. The DC designation refers to *direct current,* for which the most common daily usage is in the 12-volt systems in our cars. When electricity was first being used in the 1890s, Thomas Edison, for all his genius in other scientific regards, was insistent on DC current being the standard for home lighting and any other usage. Unfortunately, DC current isn't practical to send any long distance through wires, and with DC every neighborhood would have to have their own power plant. The brilliant Nikola Tesla (the true inventor of the radio and many other breakthroughs) developed the alternating-current system and licensed it to Westinghouse, which became a giant corporation when AC was accepted as the world standard (AC could be sent hundreds of miles along power lines).

How AC "alternates" is by traveling in a wave, alternating in polarity up and down in a repeating cycle. Most electricity in the US alternates at a rate of 60 cycles-per-second. What all of this means to arc-welding is that the less-expensive arc-welding machines are AC-only. The AC welder is very good at producing less welding spatter, at welding heavy plates with large electrodes, requires less electricity to run and usually has less maintenance expense than bigger machines. They take the standard line current, which is high voltage but low in amperage and reduce it through a transformer to a low-voltage/high amperage current for welding. The only drawback to AC arc-welding is that the constant switching of polarity can make for tiny inconsistencies in the weld bead, imperfections you and I would never notice, but something that could be critical in an oil field pipe line, high-rise-building framework or a nuclear reactor. For this reason, most professional arc welders are DC, which produces a much smoother weld, a more stable arc and there is a wider selection of special electrodes (rods) for the DC-type professional arc welder. The AC-only machines are generally used strictly for joining ferrous metals, but the DC machines can also be used for stainless-steel and for hard-surfacing industrial parts. In addition to the two current-specific types of arc welders, there are also combination AC/DC machines, which usually have a rectifier added to a basic AC machine.

Another factor separating the larger professional machines in terms of flexibility is the choice of polarity, and the option of a TIG torch setup. In the DC mode of operation, the operator of some machines can choose between negative or positive polarity, depending on the type of metal he is welding and the rod he is using. Most DC welding is done with reverse polarity, meaning that the rod is positive, and the work clamp is negative. This method keeps the rod very hot and makes for smooth welds and improved out-of-position work (anything other than flat on your welding table). This is not of concern in using AC machines because the nature of the current is switching polarity 60 times a second anyway. Many of the professional machines are versatile power supplies that can perform arc welding, MIG-welding (with the addition of a motorized wire-feeding attachment), or TIG-welding with an optional torch and foot control **(see illustrations).**

This is the kind of equipment you will find in professional welding/fabricating shops, but most shops today don't perform much arc welding, using their power supplies mostly for TIG welding, and using a separate machine for most wire-feed requirements. The arc-welding arena in professional welding is usually in industrial jobs or construction, pipelines, etc.

Not to confuse you any further about electricity, but the larger professional machines are often available for several types of input. The most basic machines, home or shop, require 220V, and industrial units may be set up for 440V or more, and there are different *phases* as well. The type of current we have in most homes is called single-phase. The larger professional arc welders are made in three-phase configuration, which simply means that there are three identical inputs spaced 120 electrical degrees apart. The waves in these inputs overlap, so that voltage never falls completely to zero, making for smoother welds. Many industrial shops have big motors on lathes, mills and other machinery that run on three-phase power, which is smoother and cheaper to operate on large equipment. We will not find three-phase power at home, and some expensive equipment is required to set up a building for industrial three-phase power.

Miller
Maxstar
AMPERAGE ADJUSTMENT
AMPERAGE A
V VOLTS
Miller
Maxstar™ 175
CC-DC INVERTER WELDING POWER SOURCE
RATED OUTPUT

4.8 Many shops utilize a "power source" machine which not only powers arc welding work, but can be set up to take a TIG torch, and some can use an optional wire-drive to perform MIG welding.

Rewiring for an arc welder

Except for a few small, household-current machines capable of 100 amps or less, even the most basic AC arc welders require 220V input. Depending on the present wiring of your house, this may require an additional expense in rewiring to accommodate the welder, a factor to take into consideration when choosing the right system for you. Some homes already have a 220V outlet for hooking up an electric clothes dryer or stove (the most common household appliances to run on this voltage), but this outlet may not be where you need to weld. If your washer/dryer setup is in the garage, you're ahead of the game, and rewiring may not be necessary. Likewise, if you plan to do most of your welding in the kitchen, then your case is simplified.

4.9 Contractors often use an engine-driven arc welder mounted on a truck bed or a small trailer. Besides powering the welder, the gas engine also has a large generator to run power tools in a remote work site.

If you have to run a new circuit in your house to put 220V where you need it for the arc welder, put it on it's own separate circuit with a 30-amp circuit-breaker. The kind of current you may be drawing when welding thick material may blow a standard 15 or 20-amp household breaker. Call a few electricians before you buy an arc welder. Tell them what size electrical box you have (how many amps), how many open spaces there are for new circuits, and how far a new 220V circuit would have to be run to get an outlet in your proposed welding area. They should be able to give you a rough estimate of the rewiring costs. It may cost as much or more than the welder itself, and may ultimately influence

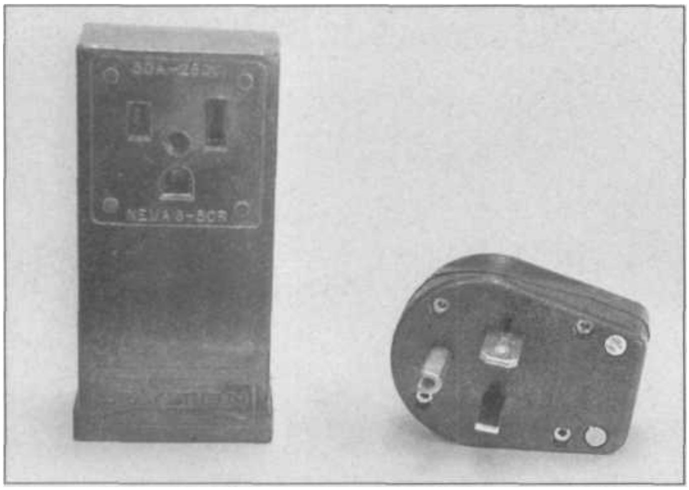

4.10 Even if you do have a 220V circuit in your shop for a clothes dryer, you'll find the outlet prongs are different on a 220V welder's plug. You either have to rewire your outlet box to accept a welding plug, put a household 220V plug on the welder's power cable, or use an adapter.

what kind of welding machine you finally purchase.

Even if you do have an existing 220V outlet you can use, you'll be rudely surprised when you find the welder can't be plugged in, at least not directly. The arrangement of prongs on the welder's power cable is slightly different than the layout of the typical 220V home appliance plug. You can make or purchase an adapter, or have the electrician install an outlet box in your welding area that matches the kind of plug on your arc welder (see **illustration).** Before plugging a new stick welder into a household 220V outlet with an adapter, you should have the system checked to be sure the wire gauge and circuit-breaker are up to the task of handling a welding machine.

The arc process

All electrical-welding processes use the flow of electricity to create heat. The power flows from the torch or electrode to the work, which is grounded to the source at the machine. In arc welding, the consumable electrode or rod makes the connection that creates the arc to the piece being welded. The welding rod is a metal rod coated with a hard flux material. As the arc is created when the tip of the metal comes to the workpiece, the heat generated at the bead is 6000° F or more, which melts both the parent metal and the filler rod, while simultaneously vaporizing the flux coating to create a gas shield around the bead, protecting the solidifying weld from contamination by gasses in the air **(see illustrations).** The flux actually re-solidifies on top of the bead as a hard coating of flux and slag, and when you look at a completed bead, you'll see a dome of ceramic-like material over the weld. At this point, the weld doesn't look very impressive, but when you remove the slag with a chipping hammer, a beautiful, clean bead is revealed (see **illustrations).** Depending on the type of rod and amperage used, there may also be some spatter (tiny beads of metal) stuck alongside the bead. Most of these beads will come off with stiff application of a wire brush, and more stubborn ones can be removed with the

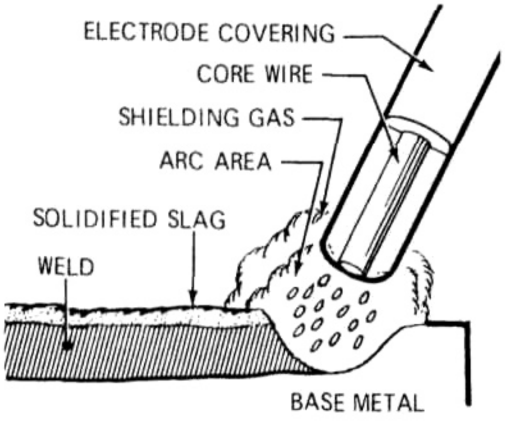

4.11 Arc-welding electrodes or rods are a metal wire, covered with a flux coating. The metal wire must make initial contact with the work to initiate the arc, then it is pulled back to a proper rod-to-work distance, the vaporizing flux makes a shielding gas, but also deposits a layer of slag over the weld as it forms, to further protect the weld from contamination.

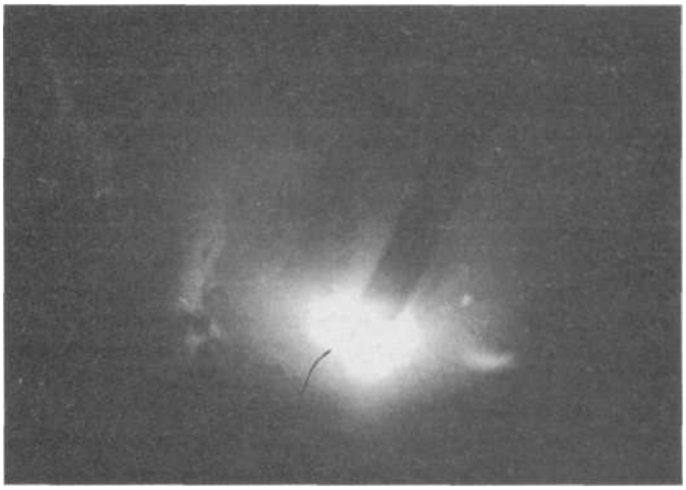

4.12 This is what you will see through your helmet and welding lens. The arc is intensely bright, but the lens cuts this down to just a bright glow around the arc itself, here you can just about see where the shielding gas cloud forms around the arc.

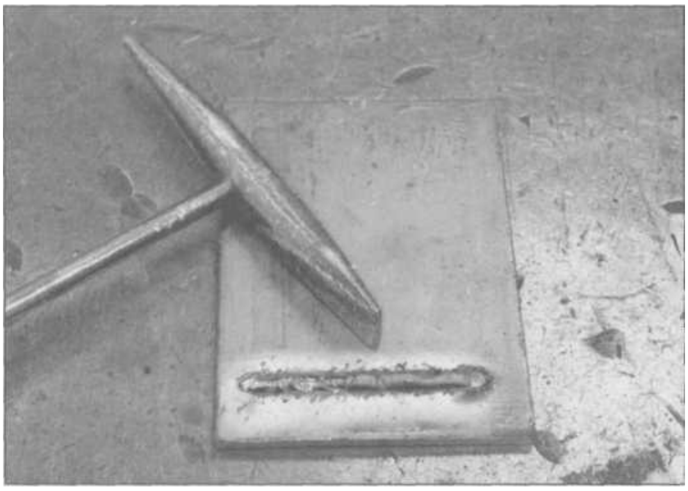

4.13 The slag formed over the weld bead must be removed with a chipping hammer like this, and most spatter and slag residue can be cleaned up with a wire brush after chipping.

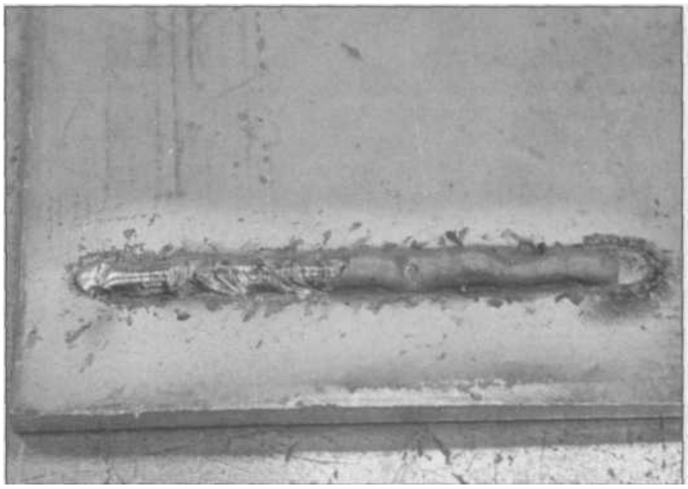

4.14 At right on this bead, you can see the dark gray slag coating covering the weld. The portion at the left has been chipped off to reveal a bright, shiny weld.

chipping hammer or a chisel.

Making your first passes with the AC stick welder will be relatively easy, although really good passes will require considerable practice. The instruction book that comes with your welder will specify the right rods to use for various materials, and the amperage to set for different thicknesses of ferrous metal. The rest of the technique is finding the most comfortable position for your electrode-holder hand, and maintaining the proper arc distance and travel speed to make good joints.

Safety considerations

With every method of welding, safety is of paramount consideration, but each type has precautions that apply to that type of equipment in particular. In all forms of electric welding, including arc welding, high-amperage electrical current is the primary hazard. All of your cables, plugs and leads should be inspected regularly for any signs of defects. Even dirt or paint overspray on connections can cause arcing and poor welds. Water, of course, is a good conductor of electricity, and therefore should be avoided in the work area. Your clothing, equipment and especially the floor must be kept dry to avoid the possibility of electrical shock. Rubber-soled shoes are recommended, but athletic shoes (non-leather) are not. Most experts will tell you not to wear metal jewelry such as watchbands, rings, bracelets, necklaces or belt buckles when welding. If electric-welder power comes into contact with metal articles you are wearing, they can become instantly hot to the point of melting, or can cause electric shock.

Of the electric welding methods, arc welding requires the most protection of your face and body during welding. The intensity of the arc produces strong UV and infrared radiation. Any skin exposed during the welding process can become burned, in severity ranging from mild sunburn to serious burns, with the symptoms not appearing until eight hours after the exposure. Leave the top button unbuttoned on your shift and you'll have a nasty V-shaped burn on your neck after only a short while arc-welding. Likewise, wear fire-resistant, long-sleeved shirts, and keep your sleeves rolled down at all times. Keep these shirts just for welding, and tear off the pockets if they have any, or keep them empty and buttoned. An experienced weldor friend of ours was recently burned painfully when welding overhead with just a shop shirt on — a hot bead of spatter went right into his shirt pocket and burned into his chest. Without the pockets, there's a chance the bead will roll off onto the floor rather than stay in one spot on your shirt. For this

4.15 The arc spatter and radiation are dangerous. Do not allow any bystanders to observe the arc while you are welding. Also, notice how much smoke is around the weldor's helmet here. The fumes are dangerous and you should have an electric fan of some kind to blow it away from you when you are welding in a corner like this.

same reason, your pants should be kept uncuffed, and never tucked into your boots.

If you are going to be doing arc-welding often, we'd recommend you invest in some leather safety clothing, like jackets, vests or pull-on sleeves that go over your regular shirt. Arc-welding is prone to more spattering than other types of welding, and these leather weldor's clothes are highly resistant to arc spatter.

Probably your most sensitive and fragile body parts exposed to welding dangers are your eyes. Even the tiniest bit of spatter in an unprotected eye can have truly long-lasting negative effects. Always wear a full-coverage safety helmet when welding, preferably with a leather flap at the bottom-front that protects your neck area. Especially when welding overhead, like underneath a vehicle, wear a cloth cap backwards (bill to back) to cover your hair and the back of your neck. Your helmet should be equipped with the proper safety lens for the type of welding you are doing, or your eyes could receive overexposure of UV and infrared rays in a very short time. Never observe *anyone else* doing arc-welding unless you are wearing proper eye protection, and make sure that when you are welding that there is no one observing you who could be hurt by watching, particularly children. Watching too much arc will not show immediate effects, but later the affected eyes will be sore, and with a sensation almost like having lots of sand in your eyes. If you do not yet have your own welder, but want to watch someone else work, get your own helmet to observe through. If you do have a welder, you may want to keep a spare helmet around in case someone wants to observe *your* welding prowess.

Your eyes can be permanently damaged by overexposure to arc rays, but they must also be protected when working around most shop equipment, such as grinders, mills, drills and sanders, all equipment that may be involved in your welding project. Keep several pair of good safety glasses around your shop, the kind that have protection all the way around the sides. After arc-welding, you will also want to wear these safety glasses when chipping slag from your welds. The little fragments that break off are like glass. Always keep a very complete first-aid kit accessible in your work area in case of accidents.

A particular hazard with arc welding is the presence of fumes. When the electrode is consumed, the flux is vaporized, creating the shielding gasses that protect the weld from contamination during formation. Depending on the metal being welded, other gasses may be released as the metal is melted. Most welding gasses are colorless, odorless, tasteless and inert, but this is not to imply that they are harmless. Any of the common welding gasses can displace oxygen, and when you are breathing in air that contains less than 18% oxygen, you may experience dizziness, or even lose consciousness. For this reason, arc welding, or any welding process, should be performed only where there is adequate ventilation. In the case of arc welding, there is less chance of the shielding gasses being blown away and causing a bad weld, so if you find yourself welding in one spot too long, or in a confined area, you can use a household fan somewhere in your work area to maintain air circulation.

Beginning arc welding

If you have already read the previous chapter on gas welding, what you will have learned in practicing that mode will help you greatly in learning all other forms of welding, including arc. Stick (arc) welding is easier in some ways than gas-torch welding, and more difficult in others. Practice and more practice will put you on the road to good welding in any mode.

What you will find different at first about arc welding is that you only need one hand. The electrode holder and its electrode or rod is *it,* other than the ground clamp, which should be clamped to your workpiece or your steel welding table. The rod is both the source of filler metal and the shielding gasses, which are generated when the flux covering is vaporized. Have some scrap steel handy and some 1/8-inch rods. There are a great many specialized rods, but one of the most common is an E-6011, which is one of the easier rods to start and maintain an arc with. Although the arc process only involves the one hand-held tool, you may find that the 14-inch-long rods put your hand a much further distance from the weld area than the other welding processes. One end of the rod is bare of any flux coating for about an inch, this is the end you put in the electrode holder.

It's a good idea to practice the arc "setup" *without* turning on the welding machine. Just situate yourself comfortably in relation to the work, and practice holding the electrode 1/4-inch away from your work joint or seam area, following the proposed seam while slightly weaving the electrode tip side to side as you travel along. This will give a feel for what is required in terms of coordination.

Now if you're ready, set the machine for the right amperage for the thickness of your scrap steel, say 1/4-inch steel plates, or perhaps slightly hotter than the instructions recommend to make it easier to learn the starting procedure. One drawback of the arc-welding process is getting the arc *started.* You can't just flip down your helmet or lens and stab the rod against the work.

The starting procedure has been described as similar to striking a match, in which you draw the rod tip across the area you plan to start the bead. At some point in your "scratch" the rod will momentarily contact the work and the arc will start, but the rod must continue moving. Arc welding is essentially a process of creating a short-circuit across the rod and the work, and it can only be started by a momentary contact of the two. Once started, this short-circuit heats the air around the weld and ionizes it to the point where the air conducts electricity and continues the arc without actual metal-to-metal contact. As soon as the arc starts, the rod tip must be pulled back to the suggested tip-to-work distance. All of this sounds tedious and difficult, but you *will* pick the technique up in the first half-hour of practice.

If you touch the rod to the surface for more than a split second, it may stick firmly, in which case the rod can get red-hot for its whole length in a very short span of time. As soon as you feel the rod stick to the surface, squeeze the clamp on your electrode holder to release the rod, which is the only way to stop the rod from melting. The hot rod will stay stuck to your work. You can take it off with some pliers, but when it is really hot, don't try snapping it off with your gloves, it may burn right through the leather. Let that rod cool off and start another rod until your have mastered the arc-starting process. When you are more experienced, you'll react quickly enough to a stuck rod that you can simply break the connection immediately by twisting the electrode holder and rod side to side.

Another method of arc starting favored by some weldors is a "tapping" style, in which you quickly tap the electrode tip to the work to start the arc, not in a "pressing" action, but in a short in-and-out jerk that makes contact at the bottom of the tip's travel toward the work.

Once you understand the starting process, the tip to work distance is next. When the arc starts, you must pull the rod back for a second to make a relatively-long arc (about twice the thickness of the electrode you are using) as a way of inducing some preheat into the metal, then immediately drop the rod closer, to about one rod-thickness away from the work. Keep the rod moving along the seam, and move the rod side-to-side slightly as you travel. Some weldors use a movement like a series of tiny, overlapping ovals, others a zigzag or even a "weaving" pattern. Stay in one place too long or with the rod tip too close to the work, and you'll melt a hole in your work; also, if you pull the rod back too far, you can lose the arc process and have to restart. On small seams in thin materials, you won't need to weave the rod much. When joining thick materials, the joint is usually Vee'd or beveled, and a straight pass is made along the bottom of the joint, followed by one or more passes where the oscillation or weaving of the tip spreads the bead circles out larger in the wider gap at the top of the bevel.

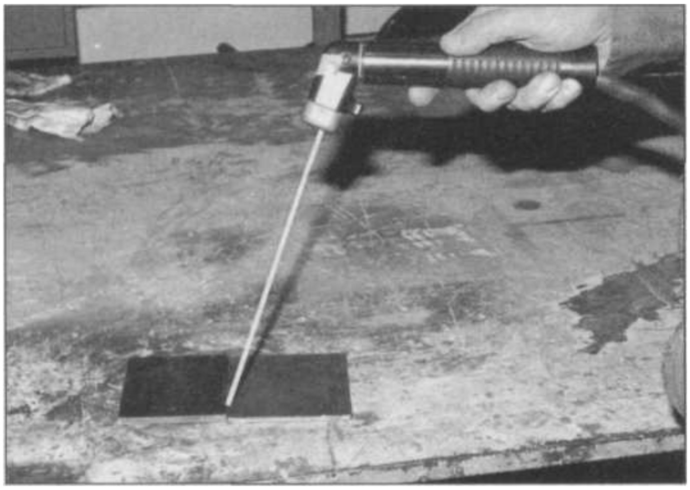

4.16 The electrode is usually held more or less perpendicular to the welding surface to start the arc, then laid back to continue the bead. Some weldors prefer a forehand technique, others a backhand direction, such as here.

The rod should be more or less vertical when the arc is struck initially, where it can really preheat the metal, but should be angled forward (in the direction of travel) 20-30 degrees as you make your pass, i.e. you hold the electrode holder somewhat ahead of the puddle, with the rod 20-30 degrees from vertical **(see illustration).** Try making some straight passes along a flat, horizontal plate, until you get the hang of running a bead. Run a straight bead with the puddles being about twice the diameter of the rod you are using. While you are welding, it is important to remember not to watch the arc, but rather focus on the puddle you are leaving behind it. The shape, size and crown of the puddle are the keys to determine how you are doing.

When you have the travel speed, amperage and rod-to-work distance correct, you'll find out why they used to call an AC arc welder a "buzz box," because you'll get a very satisfying sound, which is a steady, crisp noise something like bacon frying. This sound your welder makes, and the way the bead looks will tell you much about your progress. If you have the arc gap too large, you'll have uneven puddling, a bead that is too wide, and the sound will be uneven. Such a weld will have more than normal spatter. On the other hand, if you have the rod tip too close, it may stick to the work, the bead will be high but not very wide and the sound will be softer.

One of the hard parts to learn here compared to gas welding is that the part you are holding, the electrode holder, must continually be brought closer to the work as the rod gets shorter, while in gas welding the torch stays the same distance from the welding. The hand-eye coordination you must learn involves keeping the tip the right distance from the work, the rod at the right angle, the correct speed of travel, and compensating for the shortening of the rod. Speaking of electrode length, when the rod gets fairly short, it's best to stop and then re-start with a new rod. When the electrodes are too short, a lot of extra heat travels to the electrode holder and your gloved hand.

The speed you travel is almost as important as the rod-to-work distance. If you travel too fast, the resulting bead will be too narrow, and you may not get 100% penetration. If you proceed too slowly, you'll wind up with a large bead,

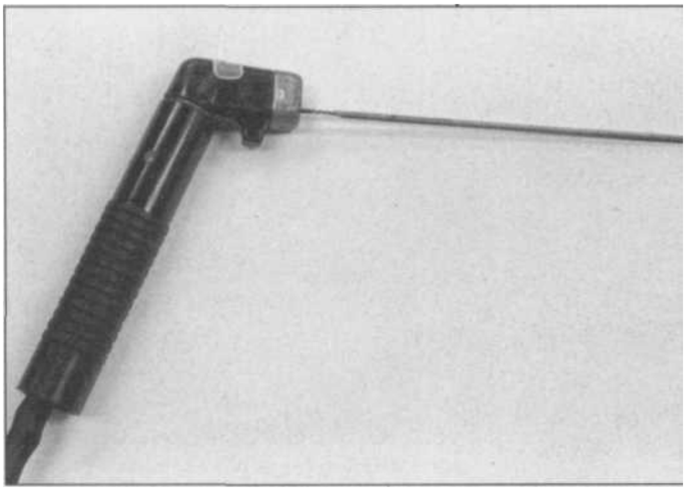

4.17 Arc-welding electrode holders are usually a clamp-like affair, with notches to allow the rod to be clamped at several different angles. Many arc weldors prefer this style, called a short-stub, in which the rod is inserted and the handle is twisted to lock the rod in place.

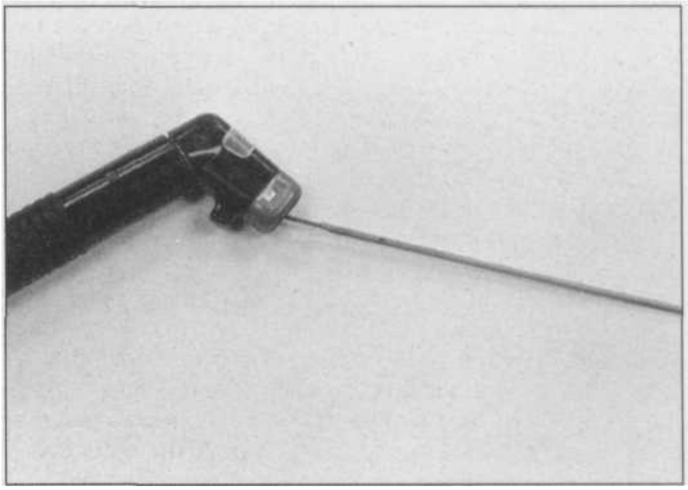

4.18 To achieve a different rod angle with a short-stub electrode-holder, just bend the rod near the holder to the desired angle. In some situation, where you have to reach down inside something to weld, the rod can be bent straight out.

and you may induce excessive heat into the workpiece. After you have practiced speed and arc-to-work distance, practice stopping and starting a bead. In real-world welding, you will encounter seams that take several sticks to complete, but you want the completed weld to look continuous even if it wasn't. To stop a bead, when you have to change electrodes for instance, just pull the rod back up quickly to break the arc. Any time you stop arc welding, you must chip the slag away from the place you last stopped, before continuing with a new rod. When you pick-up again with the new rod, start about 1/2-inch ahead of where the last puddle was and re-strike your arc and proceed. The arc will melt the original "last puddle" and continue the bead without any apparent interruption. This will take some practice. If you have seen professionally-welded seams, they look like they were applied by a continuously-operating machine, so integrated are the stops and starts, and that's what you are shooting for.

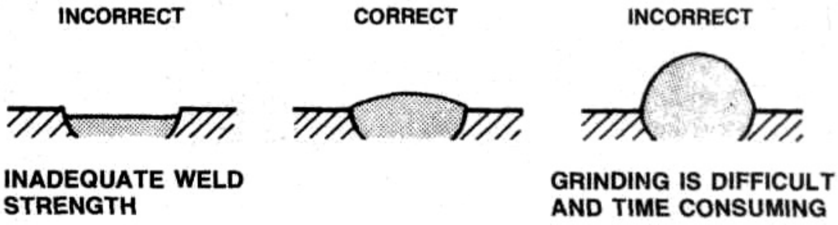

4.19 Too much amperage and the bead will sink too far into the work and leave cratering or undercutting on the top. If your welding speed is too slow, the bead can build up unnecessarily on top.

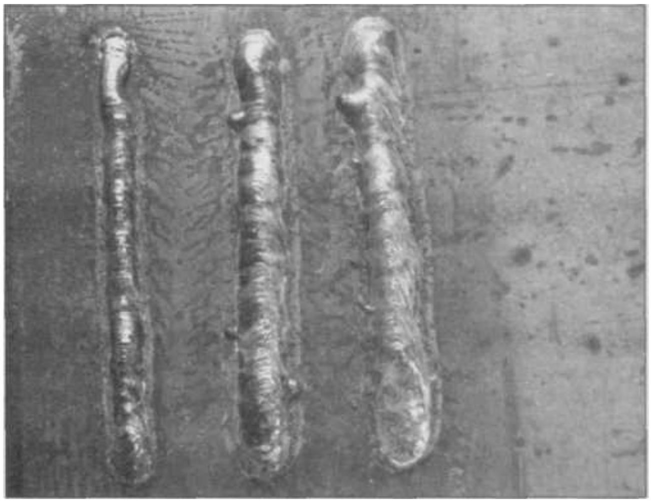

4.20 In these three practice arc-welding beads, the left-hand bead was too "cold", meaning not enough amperage. The center weld is OK, and the right has too much amperage, with increased spatter and too wide a bead for the thickness of the material.

Types of joints

After you are experienced at making straight beads on plates lying flat in front of you, begin to practice on joints between two pieces. The simplest to learn on are butt joints. If the material is 1/4-inch or thicker, you should bevel the edges of both parts before welding. As with any parts to be welded, either ferrous or nonferrous, cleanliness of the work is very important to making a sound weld, so grind the pieces to be joined not just on the bevel, but at least 1/2-inch on either side of the joint, so that impurities don't contaminate the weld.

Most seams are started by tacking the parts together at either end and perhaps several places along the way, depending on how long the seam is. Because of the growth of the parts from the heat of welding, you may need to "build-in" a gap between them during the tacking phase. Some weldors place a small piece of bare copper wire between the two parts, tack one end, and then move the wire "spacer" to where the next tack will be, continuing so that the two parts are tacked slightly apart from each other with an even gap. As you weld up the seam, the parts will "grow" together. Whenever we mention tack-welding in conjunction with arc welding, remember that the slag must be chipped and wire-brushed away from the tacks before you "connect the dots" with continuous welds.

Depending on the thickness of the material you are working on, you may make a two-pass weld, one on the top and one on the bottom. This would only be feasible on plates, not on pipe or tubing, but makes for a very strong joint, and the opposing forces of distortion may keep the parts flatter than if you made only one large pass on one side. Looking at an end view of the plates, the first pass should go into the joint a little more than halfway, then the second weld on the other side should bite into that first bead for a completely-welded joint. Usually, such joints do not require as high an amperage setting on the welder as for making a single-pass weld.

One of the common uses of arc-welding equipment is in farming and ranching, where the equipment being repaired is usually too large to bring into the shop. That means that most welding that requires a bottle of shielding gas may be too susceptible to wind to be effective, and the arc-welding

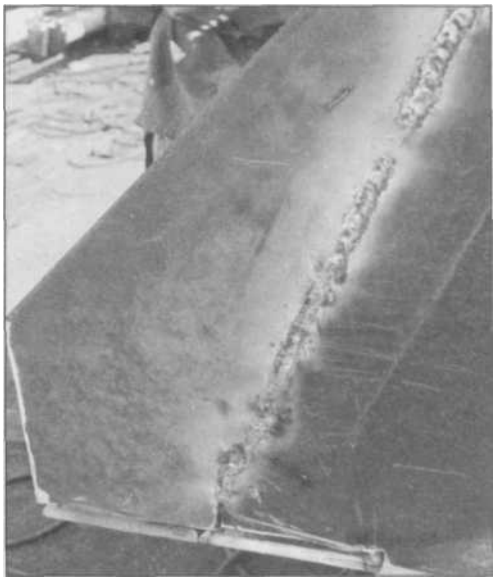

4.21 Welding large plates, such as here where new metal is being added to a John Deere dirt scoop, requires either multi-pass welding, or using a large rod at high amperages. This seam was done with a single pass with 1/4-inch electrodes.

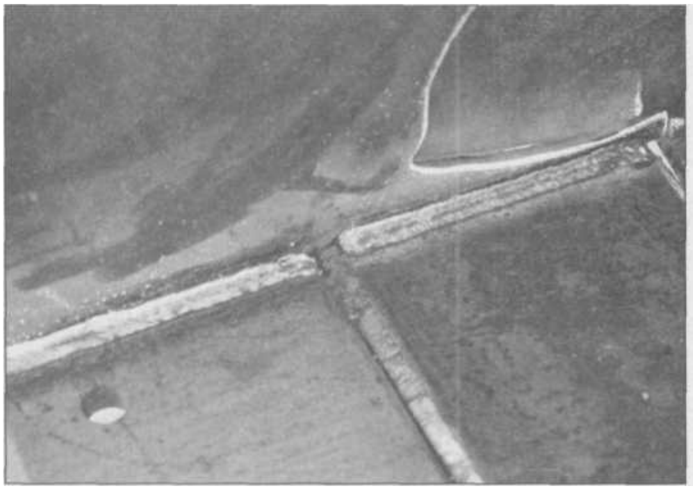

4.22 Here half-inch plate on the sides was welded with one pass, but a gap was left where the one-inch, beveled lower plates join. The bigger plate will be welded with several passes and strength is improved if the joint starts all the way at the ends.

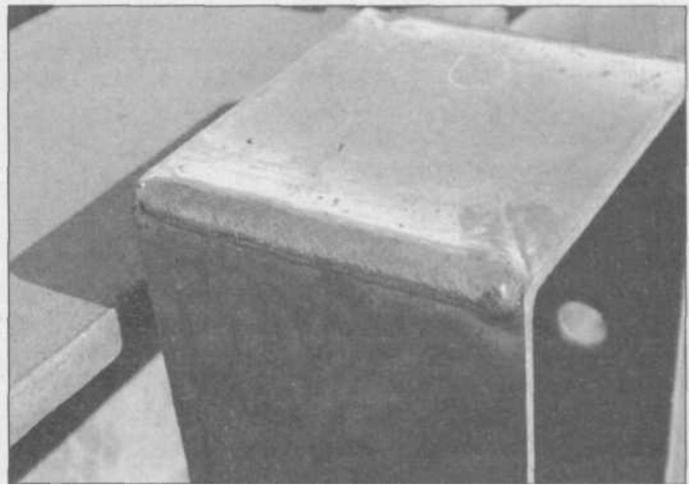

4.23 An outside corner weld on 5/16-inch plates is made with one pass with a wide weaving pattern, spread equal heat on both parts.

process is advantageous here. Also, most farm equipment is rather heavy, with large components, and arc welding's ability to fill large breaks, weld castings, and deposit large quantities of filler metal if needed makes arc a good choice. The visual beauty of the welds is seldom a factor in agricultural repairs: it's getting the equipment back into service that counts. Building up of worn agricultural parts is a common task in which the weldor makes repeated beads right next to each other to literally build a new and higher surface on the part. The work is tedious, but necessary.

Generally, you make the first pass in normal welding mode, then make succeeding passes with the electrode held at a slight side angle to the work. Remember than when arc welding, the hot filler metal is virtually sprayed off the end of the rod like a spray gun, and where you aim the rod is where the metal will be deposited. The slight side angle allows the succeeding passes to bite slightly into the previous bead. Alternate the direction of the beads and overlap them about one-third of the bead width. There are even special rods made that deposit a hard-facing on metal. These can be used to renew or create a hard cutting edge on a part, which can be later ground or filed to an edge that will stay sharper longer than the parent metal. When making any kind of 'buildup" welds, each bead must be chipped clean before you begin the next pass, or slag and impurities could be trapped by the next bead. Also, when you are building up a part, rather than fusing a seam as in regular welding, you can run the bead much wider than we have suggested so far. The idea here is not to fuse two parts but to deposit as much metal as possible onto the surface. In these cases, you can "weave" the rod tip back and forth in a zigzag or other pattern to make a wider bead.

You will find that "out-of-position" welds, welds that are not made with the parts lying flat in front of you on the welding table, are the toughest to learn. Corner joints, vertical seams and overhead seams offer increasing levels of challenge. Corner joints are commonly found when joining two plates where one plate is perpendicular to the other, like a T. The electrode should be at a 45° angle between the two parts, so that equal heat and filler metal is directed onto both pieces. In some cases, you may find you have to place the arc such that it puts more heat on the bottom plate (which is dissipating more heat from the joint) while aiming the electrode "spray" more at the upper plate to avoid "undercutting", which is when you see a slight crevice along one or both edges of a welded joint after you chip the slag away.

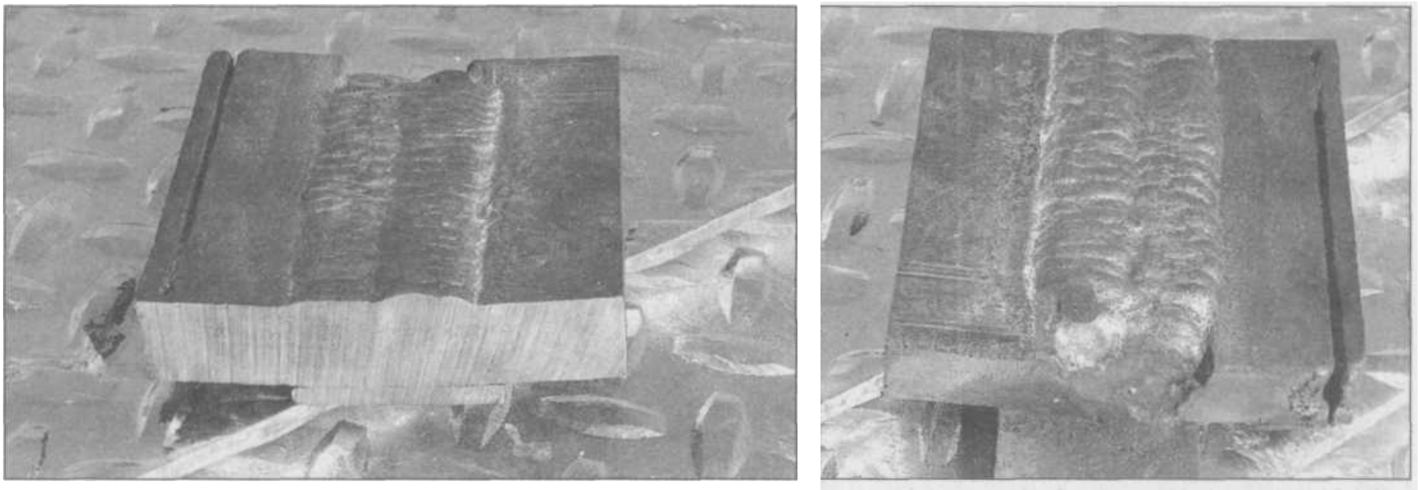

4.24 This sample weld shows how large plates are joined by multiple passes with large rods. Compare this front view (right) where the weld passes began, to the cutaway photo (left) which illustrates how integrated and free of contaminants even such a large weld can be.

If you are joining heavy plates, the fillet weld may take three passes, in which case the first bead is put right into the corner, then a second pass is made alongside that one with the electrode aimed to "spray" slightly more toward the side between this bead and the corner's edge, and the third pass is aimed to the opposite side of the center **(see illustrations).** If, on the other hand, you are welding a thick plate to a thinner one, then favor the angle of the electrode towards the heavier plate to give it more heat. In practicing such welds, it may be advantageous to use plates that are only an inch or two by six inches. This will make for less work when cutting a cross-section through them afterwards to examine your weld for penetration and lack of voids or impurities, especially if you have to hand-hacksaw through them.

While basic, flat butt joints are most common in typical home/shop projects, the Tee or fillet joint and the lap joint make up the majority of welds in industry and construction. The lap joint, where one plate lays on top of another, is approached much like the Tee joint, with the electrode angle at about 45°, aimed at the inside corner where the two plates meet. If they are different thicknesses, then the electrode may need to be aimed to give more heat to the heavier piece. Another difference between industry and home projects is that you will seldom encounter any joint that requires more than one pass to complete in a home/shop project. The use of really thick plates just isn't that common in automotive, arts and crafts or making shop equipment, with the possible exception of building an engine stand, which might have a half-inch plate welded to a tube for the part the engine-holding arms attach to. This is one reason that arc-welding isn't that common anymore as a home/shop tool. Other methods seem better suited to the thinner materials encountered in hobby projects, though farm equipment repair does sometimes require the large-scale abilities of an arc welder.

Choosing electrodes

Your home AC arc-welder will probably come with an assortment of basic welding rods to practice with. After that you'll have to shop at your local welding supply or order from a catalog, and you should have a good idea by then of what type and thicknesses of metals you'll be working on, and thus what size and type of rods to use. Your local supplier should be able to give you good advice on what best suits your purposes. As you read or talk with other people about arc-welding, you may hear the rods termed any of the following: stinger, rod, electrode or stick. They all mean the same thing.

While there is a confusing array of arc-welding electrodes on the market, the bulk of them are specifically designed for industrial and other specialty usage, and you will never have a call for them. There are perhaps only a half dozen rods that would cover your needs. Rods differ in the type and diameter the central metal wire is made of, and in the thickness and composition of the coating of flux as well. In actual use, the flux burns away at a slower rate than the rod inside, which makes for a sort of "collar" around the tip of the rod, further shielding the weld process and helping concentrate and direct the "spray" of metal coming off the wire **(see illustration).**

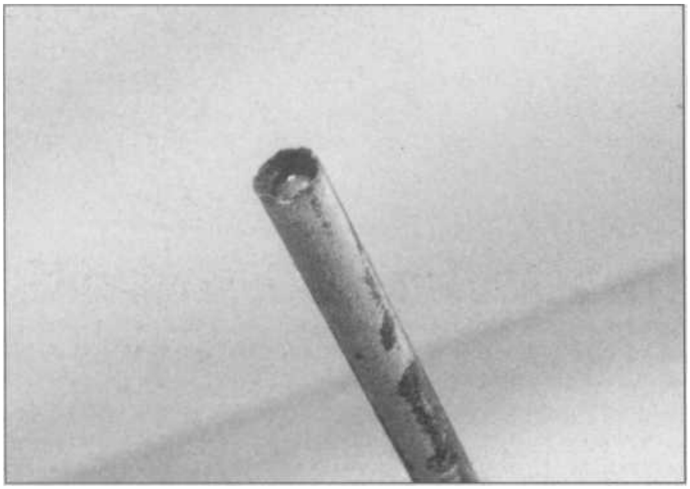

4.25 Some arc rods melt the wire at a faster rate than the flux, leaving the wire recessed inside. During the welding, this makes a more concentrated arc spray. When re-striking an arc with such a rod, the extra flux around the tip must be broken off with your gloved fingers or pliers (with the rod cooled off) to expose the metal for re-striking the arc.

Many of the hundreds of special-purpose arc rods are designed strictly for the professional DC welder, so you won't have to worry about those if you're using a typical 220V home AC welder. For your needs, there are perhaps three types you will use the most. Of these, E-6011 is the designation for what is the most commonly-used electrode. It is one of the easiest for a beginner to master, can be used on AC or DC, in virtually all positions of welding, and can be used in situations where the parent metal hasn't been prepared spotlessly. It's main drawback is that it produces a lot of spatter. Where the *appearance* of the weld is less important than the ability to make a strong joint on rough or dirty materials, such as in farm equipment repairs, this is a good choice. After practicing with this one, you may want to move on to more "sophisticated" electrodes. The E-6013 rod works great in many situations, and, although it requires better surface preparation and cleaning than the 6011 rod, it will produce much better looking welds, is suitable for a variety of positions and handles metals up to 3/16-inch thickness with most home-type buzz-boxes. It can be used for lots of projects like building shop equipment, and there are even variations on this rod from different manufacturers that are specifically for sheet-metal work, in sizes down to 1/16-inch. When the size of a welding rod is mentioned, it refers only to the diameter of the wire *inside.* Obviously, the flux coating makes the rod appear much thicker, and if you need to check a **rod,** measure the bare end meant to go into the electrode holder.

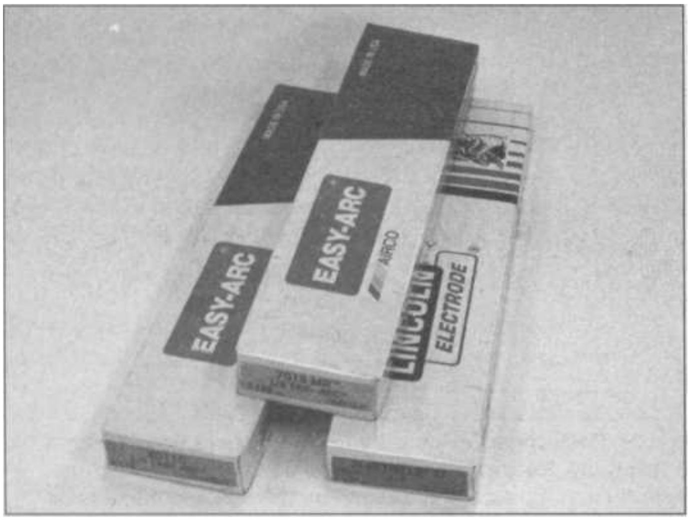

4.26 You'll find a wide variety of electrodes at your welding supply store, but there are only a few rods you will use in normal home/shop projects on mild steel. Don't buy much more than you will need for the next project.

There has been a variety of methods used in the past to identify welding rods, including numbers and color codes, but most rods you find today have both a number and code from the manufacturer, as well as a standard A.W.S. number, such as E-6013 **(see illustration).** These standard designations come from the American Welding Society and should be easy to find somewhere on the boxes of rod you buy. The E stands for electrode (for arc welding), the 60 part is multiplied by 1000 to give you the tensile strength of the wire (in this case, 60,000 psi), the next digit is a code for the type of position the rod is recommended for, and the last digit refers to the type and polarity of the current required. The 1 in our example in the "position" spot indicates the rod is good in all positions, a 2 would mean flat position. The last number indicates that this rod can be used with AC or DC, and on DC can be used with straight or reverse polarity, though straight polarity is seldom used.

Another good rod for starting out is E-7014, which is higher in strength and produces good-looking beads. It and the E-6013 mentioned earlier require a little different arc procedure than what we have described so far. These are called "contact rods" because, instead of maintaining a basic 1/8-inch gap as you go along, you keep the tip just in contact with the parent metal, in a sense dragging the rod lightly along the seam.

A special-purpose rod used in repairing cast-iron, such as automotive blocks, cracked heads, transmission cases, and various farm equipment, is EST, although there are several brand names in the trade just for cast-iron rod. Most rods for this purpose have a high nickel content. Although these rods are available, it doesn't mean that cast-iron welding is easy. Any cast part should be thoroughly preheated to 400° F or better to prevent cracking after the localized heat of welding is induced into the part. Parts, depending on their size, can be heated in an oven or with a rosebud tip on an oxy-acetylene torch, using temperature indicating paints to tell when you have it evenly heated to the proper temperature. Even then, there can be cracking problems. It can be very difficult to get a cracked cast-iron part thoroughly cleaned before welding, usually requiring a deep Vee to make a good repair. Better success on cast iron may be had by welding in short strips 1/2 to 3/4-inch long, with the part cleaned and allowed to cool off in between. In automotive use, broken cast-iron exhaust manifolds are a common repair item, and the high-nickel rods can be used successfully, but a well-used exhaust manifold needs to be sandblasted inside and out beforehand to get rid of as much of the carbon baked into the pores of the casting as possible. Usually, a brand-new casting will take a much-longer-lasting weld, such as when modifying manifolds to accept turbocharger flanges. Most small cast-iron repairs are better made with brazing by an oxy-acetylene torch, where much less heat is introduced, although the part should still be preheated.

Arc-welding electrodes are also marked with color codes. Usually, the flux coating itself is a different color to indicate a type, as well as markings or dabs of colors near the tip. The E-6011 has a white flux coating and a blue spot color, E-6013 is a dark tan with a brown spot, etc.

Proper storage of electrodes is critical to their performance **(see illustration).** They are very susceptible to moisture, and must be stored in a perfectly dry environment. Once they absorb moisture, the flux coating tends to loosen and flake off, and the rod is useless. In large production operations where critical welding is done, the rods are stored in a special oven that keeps them at a constant 100° F or more to ensure that they are dry, and only enough rod that will be used in a few hours is removed at one time from the oven.

In a home/shop situation, buy only as much rod as you think you need to keep around for unexpected projects, a few pounds of each type you normally use. Keep these in either sealable metal cans or plastic bags with desiccant inside. Desiccant is a moisture-absorbing compound, the stuff we always find packed in little "teabags" with electronic components and camera equipment. You can save up these bags whenever you get a new electronic goodie. The bags can be dried out in the oven for a short while, then put into the container holding your welding rods. If you have a bigger project in mind, buy what you need for that project only, fresh from the welding supply, and don't keep any large quantities around. We have seen a variety of methods used by weldors to keep electrodes dry, including racks of custom-built metal tubes with screwcaps.

4.27 This is the bare end of the rod, where it fits into the electrode holder. You can judge the size of a rod from this bare end - don't measure the flux coating. Rods are marked as to their type and color-coded.

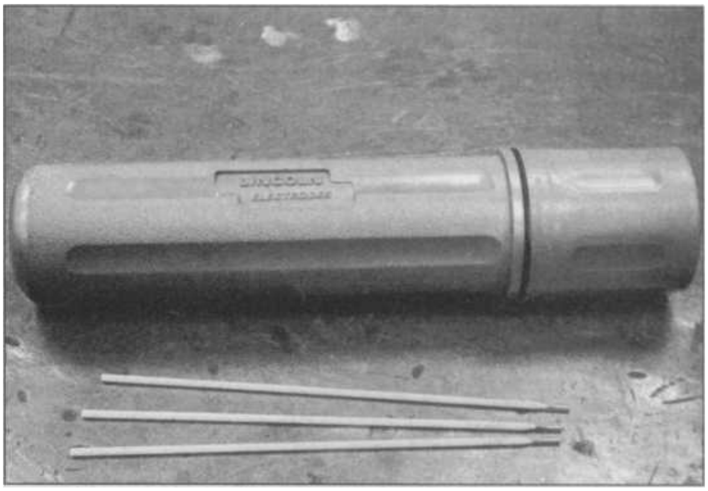

4.28 Rods must be kept dry for the flux coatings to work properly. Keep only as much rod as you need, and store them in a tight container such as this Lincoln plastic rod box with gasketed screw cap.

L-TEC STW 140i
L-TEC STW 90i

4.29 These L-TEC portable arc power sources are lightweight, and offer up to 90 amps on the STW 90i (right) and 130 amps on the STW 140i. Unlike most small arc welders, these are DC, and can even be set up with a torch and power adapter to run a TIG torch. They are available for 110V or 220V input.

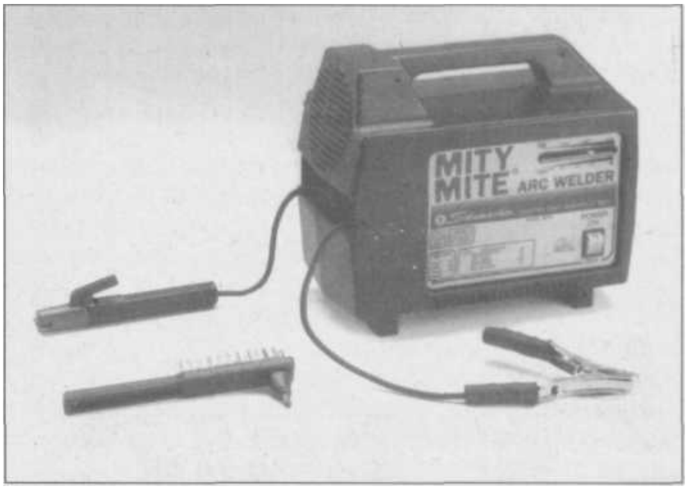

4.30 The Eastwood Company markets a lot of equipment and supplies for auto restoration, and this is their small 110V arc welder. At its maximum of 70 amps, it delivers a 10% duty-cycle, so you won't be welding any bridges together, but it only weighs 26 pounds, so it's extremely portable and welds up to 3/16-inch.

If you have reviewed the above introduction to arc-welding, you should have a good idea by now of whether it will suit your particular welding needs. It can be both easy-to-learn and challenging, simple and complicated, and can produce results from "farm-grade" (where strength is primary) to automotive-quality to atomic-reactor quality. It has the advantage of simple operation, low maintenance, low initial cost (not considering the cost of rewiring for 220V if necessary), and the unmatched ability to weld outdoors without being affected by wind, making it suitable for structural-steel construction, pipelines, iron-fence work, and farm repairs, as well as being the best choice for welding and repairing thick materials.

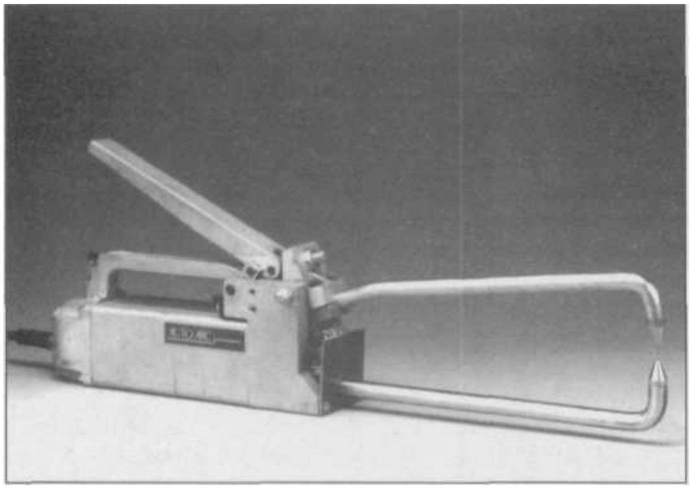

4.31 Auto manufacturers build cars with spot-welding machines, and with Eastwood Company's spot-welder gun, you can repair one the same way. The replaceable electrodes last for 60-100 welds.

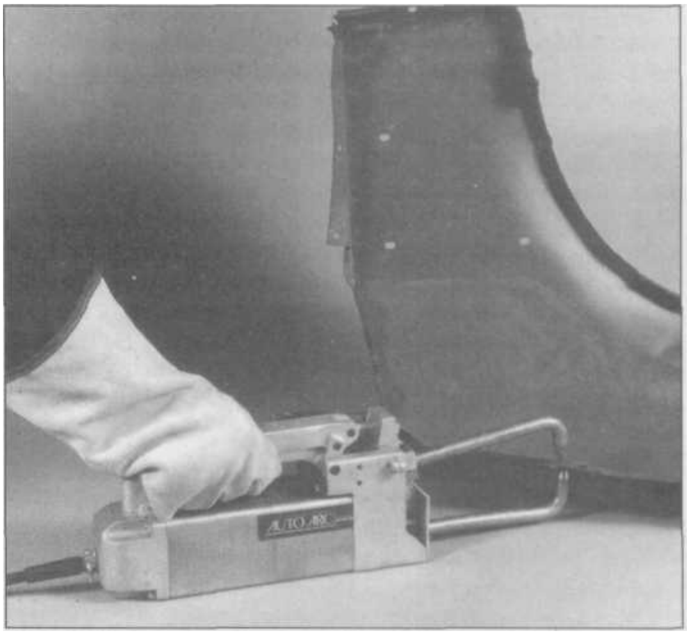

4.32 The spot welder is a self-contained mini-arc-welder, designed for joining sheet-metal panels on auto body work.

5 MIG welding

Welding books written as recently as ten years ago may have recommended oxy-acetylene and arc-welding equipment as the basic tools for home/shop welding. At that time, MIG, or *wire-feed,* welding equipment was considered too expensive for amateur use, despite its advantages. The MIG equipment was recommended, and indeed was originally designed, for high-production shop work only. Much has changed in the intervening years, and the introduction of lower-cost MIG units and the competitiveness of the marketplace has brought wire-feed welding into an affordable range for the home/shop weldor.

An outgrowth of arc welding, the MIG process was conceived as a way to speed up production in industrial applications. The basic difference with MIG, which stands for Metal Inert Gas welding (sometimes also referred to by its technical description as Gas Metal Arc Welding, or GMAW), is that the welding filler rod is automatically fed through the gun whenever welding takes place. The weldor doesn't need to stop to change arc electrodes. The Airco company developed the MIG process

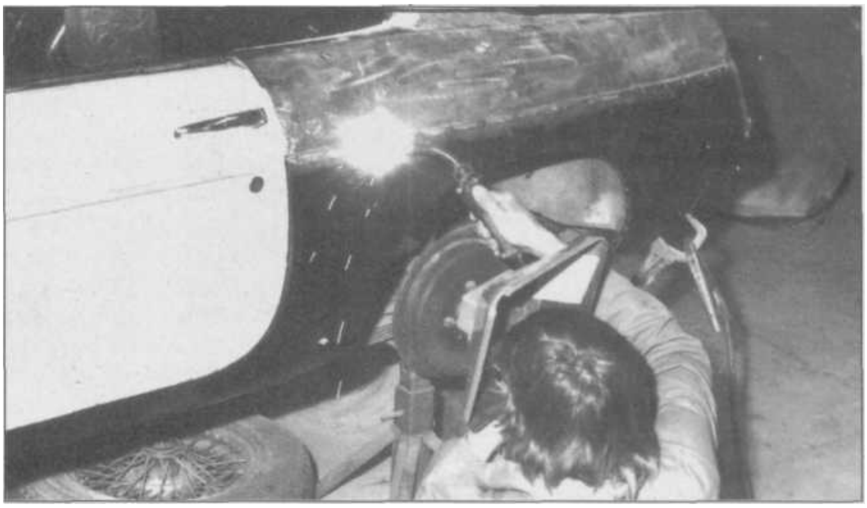

5.1 The use of wire-feed MIG welding has expanded greatly in the last ten years to encompass body shops, muffler shops, race-car builders and home hobbyists as well as the industrial production work it was designed for. It has advantages of speed, simplicity, cleanliness, and the ability to work well on thick and thin materials.

L-TEC
Mig 130
PREST-O-WELD
Mig 130
WIRE FEED SPEED
VOLTAGE

5.2 Among the compact MIG welders in the 110V field is the L-TEC MIG 130, which features 100% duty-cycle at its 30 amp setting for thin material, and has a 25% duty-cycle at 130 amps. It's designed for body shops, hobbyists and light maintenance use.

5.3 **For the larger-scale shop, the heavy-duty 220V machines like this Migmaster 250 offer more amperage, more amp settings, higher duty cycles, and various extra features for timing various weld functions.**

5.4 The Miller Electric Co. makes a variety of larger professional welding machines, but they also offer this Millermatic 130, which has a 20% duty cycle at its maximum 130 amps, putting it right in the midst of the 110V home/shop MIGs.

right after W.W.II, and it has been refined continuously ever since.

The many advantages of the MIG process, besides the basic one of not having to stop to change electrodes, include: being able to weld in all positions; there is no slag removal and the welds are much cleaner and with very little spatter compared to arc welding; the welding can proceed a lot faster than with other methods because the wire-feed rate is automatic and adjustable, which also makes for less distortion of the workpiece because less heat is concentrated on the seam; the process works well on joints that are irregular or have gaps; and the amperage, wire size and wire-feed speed can be tuned down to do ideal, almost distortionless welds on thin sheet metal. There are more advantages that relate to how the weldor uses the equipment, and we'll see these points as we continue.

There are technically three different processes of metal transfer with MIG welding equipment, but the one we are concerned with is called the *short-arc* or short-circuit process. The other types are used in larger industrial welding processes where high amperages and large-diameter welding wire is used. Most of the MIG machines suitable for home/shop use are designed for short-arc process, where the molten end of the welding wire touches the weld puddle and creates a short-circuit, and the wire diameters are from .023-inch to .045-inch. Amperages in these machines are seldom higher than 225 amps yet can handle materials from the thinnest auto-body sheet metal to plates as thick as 1/2-inch, a range that certainly covers everything the home/shop weldor, metal sculptor, body man, farmer or ornamental iron worker will have a call for.

In the short-arc process, the basic components of the MIG equipment consist of the welding machine, which contains an AC-to-DC rectifier with constant voltage potential and a wire-feeding mechanism that holds a large roll of wire, a bottle of shielding gas (in most models), a welding cable that routes wire, power and gas to the torch, a simple trigger-operated torch or gun and a ground or work cable. In operation, the weldor brings the torch down to the work until the bare wire electrode, which sticks out of the gun 1/2 to 3/4-inch, is touching the

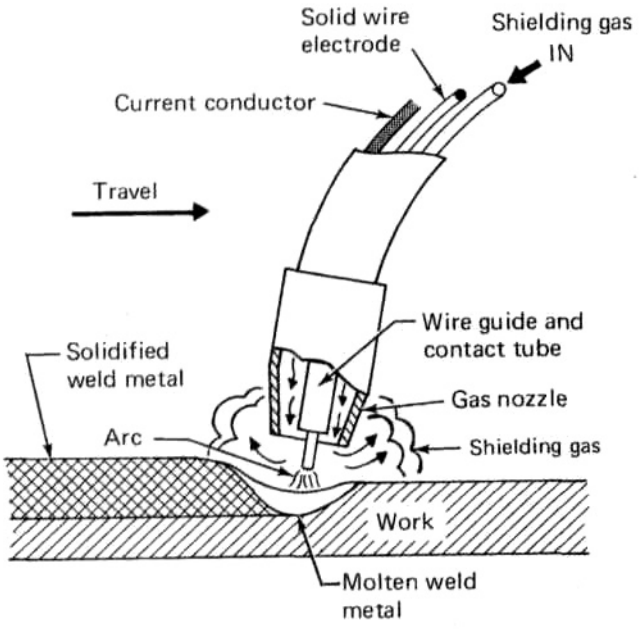

>—Molten weld metal

5.5 The cable leading to the MIG torch has to carry several different materials, from the electrical current, to the wire electrode and the shielding gas. The torch itself is rather simple, feeding the wire out to maintain a short-circuit arc against the work, while keeping the weld under an envelope of shielding gas.

OWER
MIG
130 T
HICAGO
LECTRIC
welding
ystems

5.6 There are very few user controls on the panel of a basic MIG welder, just a switch for the different amperage settings, and a variable knob for controlling the wire-feed speed.

work where the seam is to be; he then flips his helmet or lens down and pulls the trigger. The trigger immediately puts power to the wire, which arcs against the work to create very localized heat while shielding gas is simultaneously released that puts a shielding "envelope" over the forming weld, preventing contamination **(see illustration).** The short-arc process actually is alternately melting and not melting the electrode wire about 90 times a second. Each time the wire touches the work it melts and creates a gap between itself and the work, which is when a drop of molten wire attaches to the work and blends in. The constant arc-on-arc-off is what gives the MIG process its characteristic "sizzling" sound when you weld with the correct wire speed, amperage and travel speed. The current used is DC with reverse polarity (like most of the professional arc-welding machines), which provides for deep penetration. Only a few applications and machines have the straight polarity option, and this is only for shallow-penetration work where the coverage or speed of deposition is most important.

The weldor has few controls to worry about in setting up a MIG machine. There is a voltage knob or switch to control the current, and a wire-speed knob that determines how fast the welding wire comes out of the gun **(see illustration).** Selecting the amperage is based mostly on the thickness of the material to be welded. A rule of thumb from the Miller Electric Company is "one ampere for every .001-inch of plate thickness." For example, an 1/8-inch (.125-inch) plate would require 125 amps. Of course, this doesn't mean that a 1/2-inch plate would take 500 amps, because large plates are generally beveled before welding and joined with a root pass in the bottom and several more passes on top. There are, however, industrial machines that routinely weld large plates, such as on ship decks, with very large wire diameter and high amperages. This is just a general rule, and other factors, such as the wire speed and welding travel speed affect the current requirement. In the smaller home-type MIG machines, the amperage settings are fixed, with perhaps four different positions on the current knob, but the wire-speed knob is infinitely variable on all MIG machines. The

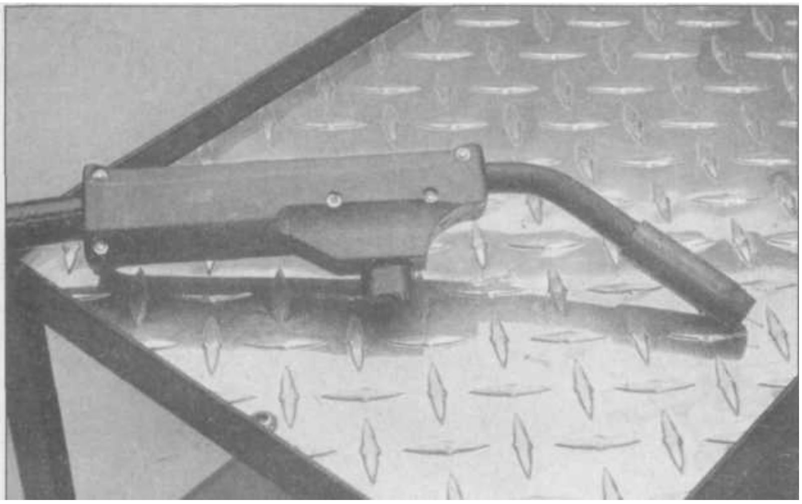

5.7 **The basic MIG torch is quite simple. Inside is a guide tube that the wire comes through, a switch that turns on the power and the shielding-gas flow. The curved neck, called a "swan neck", holds the contact tube and the nozzle.**

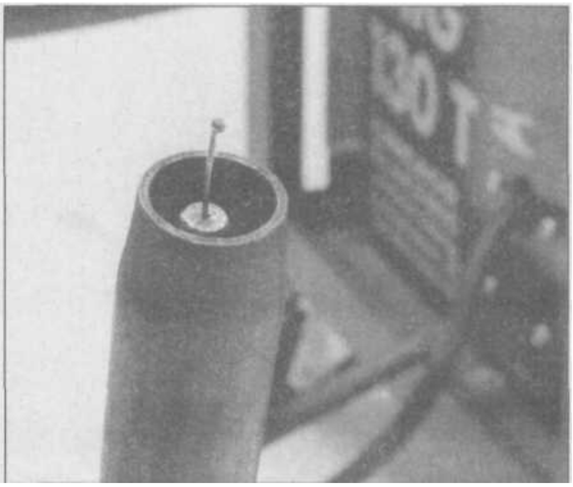

5.8 **The wire must fit the contact tube closely, as this is where power is transmitted to the electrode wire. Note that the contact tube is recessed inside the nozzle, and the wire sticks out 1/2-3/4-inch from the nozzle.**

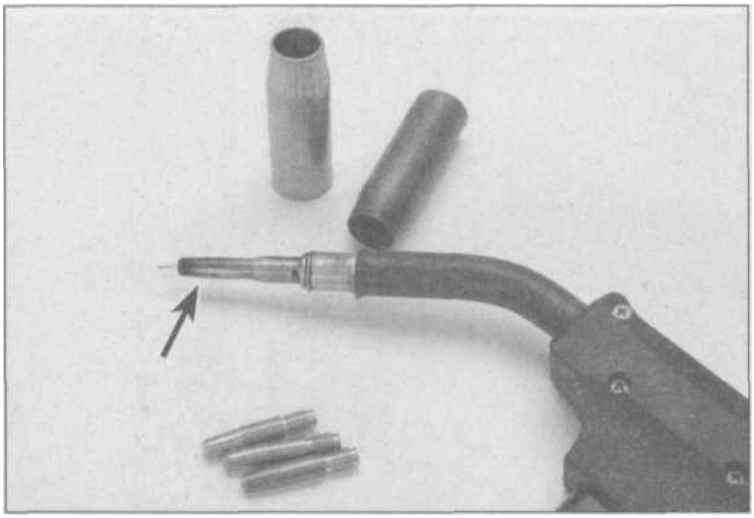

5.9 **Most MIG nozzles just twist off past a snap ring, exposing the contact tube (arrow). You should keep an extra nozzle as a spare, and several spares of the sizes of contact tubes (below) that you use most. Eventually, you will weld the wire to the contact tube, and have to unscrew it to get it off, then clip off the wire and install a new tube, also called a contact *tip.***

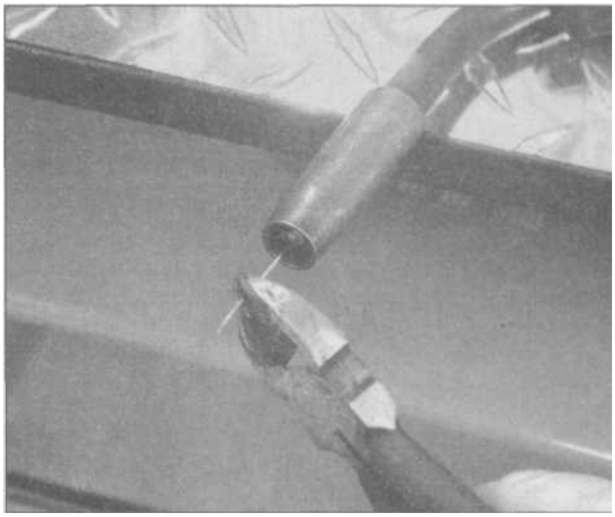

5.10 **It's important to keep a pair of side-cutter pliers handy at all times to your MIG welder. Not only do you need it to maintain the correct protrusion of the wire from the nozzle, but when the wire develops a ball on the end, the arc is easier to start if you snip the ball off.**

more expensive, more professional MIG machines have variable-current output, or at least more positions on the switch, and may have other optional controls for timing weld events.

The torch cooling on almost all MIG machines, certainly the ones we are considering for home/shop use, is by air, although there are some industrial applications which have a water-cooled torch because of the heavy wire and high amperage being used. Looking inside the MIG torch, the main body of the handle contains the switch or trigger, a wire conduit with a metal or plastic-tubing liner, and the curved "swan neck" that comes out holding both a contact or wire guide tube and a nozzle. The contact tube is so called because it actually puts the power to the wire. Usually made of copper, the contact tube also keeps the wire output centered in the gas cup or nozzle **(see illustration).** Since this is where

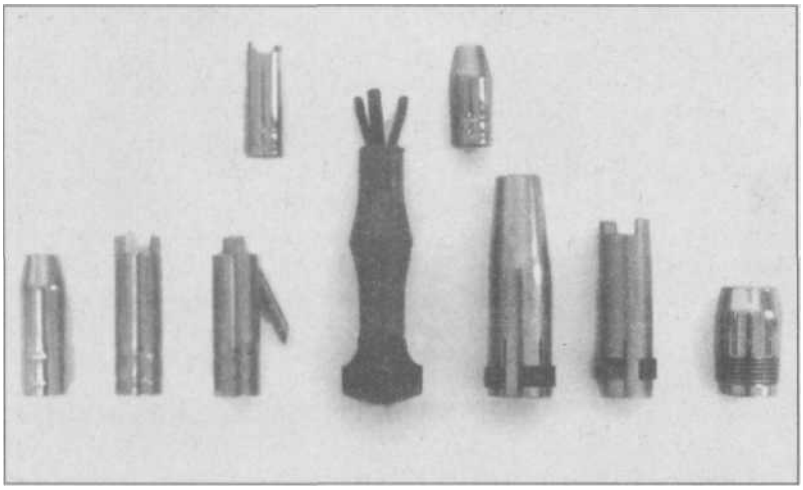

5.11 For some MIGs, there is a variety of different-size nozzles, some with special purposes like spot-welding. At the center here is a tool for scraping debris from inside a nozzle.

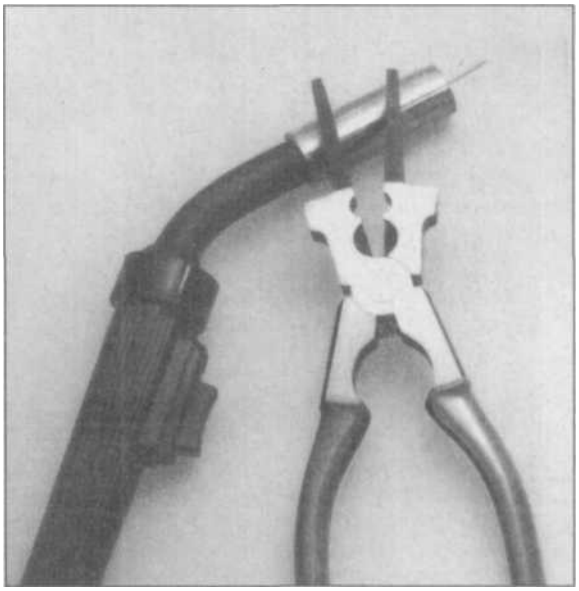

5.12 Eastwood Company makes these multi-function **MIG** pliers, which can be used to clean nozzles, snip wires, and hold contact tips to unscrew them.

the power is transferred to the wire, the contact tube must have a specific size related to the size of the electrode wire being used. Because of the unavoidable wear in the contact tube, and occasional arcing that ruins a contact tube, these are considered consumables in MIG welding, and you should have several spares of any of the sizes you normally use.

The nozzle's main function is to contain the shielding gas in an envelope around the arcing wire. Most MIG machines have only one size nozzle, but in production applications smaller nozzles are used on smaller-amperage work and larger nozzles on bigger jobs **(see illustration).** The nozzle takes some punishment in MIG welding, with spatter and weld debris constantly building up inside. Special tools are available that scrape out the buildup, and most MIG weldors use an anti-spatter or "nozzle-shield" gel or spray to keep down the buildup and allow easy removal of the gook when necessary **(see illustration).** Obviously, if too much debris builds up inside the nozzle, it will affect the gas flow to the weld area and thus cause deterioration and unevenness in the bead. The sprays can be directed inside the nozzle after cleaning, to prevent future buildup, while the gel comes in a jar that is left uncapped while welding. Every so often, you take the hot MIG gun and stick the nozzle directly into the gel, which leaves a coating inside the hot nozzle and on the contact tube.

3M
Weld-Thru Coating
WELDER'S HELPER
ANTI-SPATTER SPRAY
CLEARCO
WELDING JELLY

5.13 These three chemicals are helpful in performing clean welds and maintaining your **MIG** torch in good condition. The anti-spatter spray and gel are used on the nozzle to make removal of debris easier, and the weld-through coating by 3M is used when welding high-strength steel in late-model cars. Unlike the other sprays, it is applied to the area to be welded, providing extra corrosion protection after the weld.

Shopping for a MIG welder

The array of available MIG machines aimed at the consumer and small shop market has greatly increased in the past decade. Many automotive body workers and fabricators in small shops used to dream of having a MIG welder to round out their capabilities, but the machines available at that time were either very expensive self-contained MIG machines or wire-feeding attachments for already-expensive, large-size arc/TIG power supplies found only in welding shops. The hobbyist, auto body man or muffler shop couldn't justify the expense.

That changed when imported MIG machines started making their appearance in the US, aimed at the hobbyist market specifically. The companies advertised in hobbyist publications, attended auto enthusiast shows, and made a presence in the business community. At first, the big domestic welding-equipment manufacturers took little notice, but after a few years they realized there was a gap in their marketing and price structure and began to produce similar equipment to gain back this emerging hobbyist/small shop interest in MIG weld-

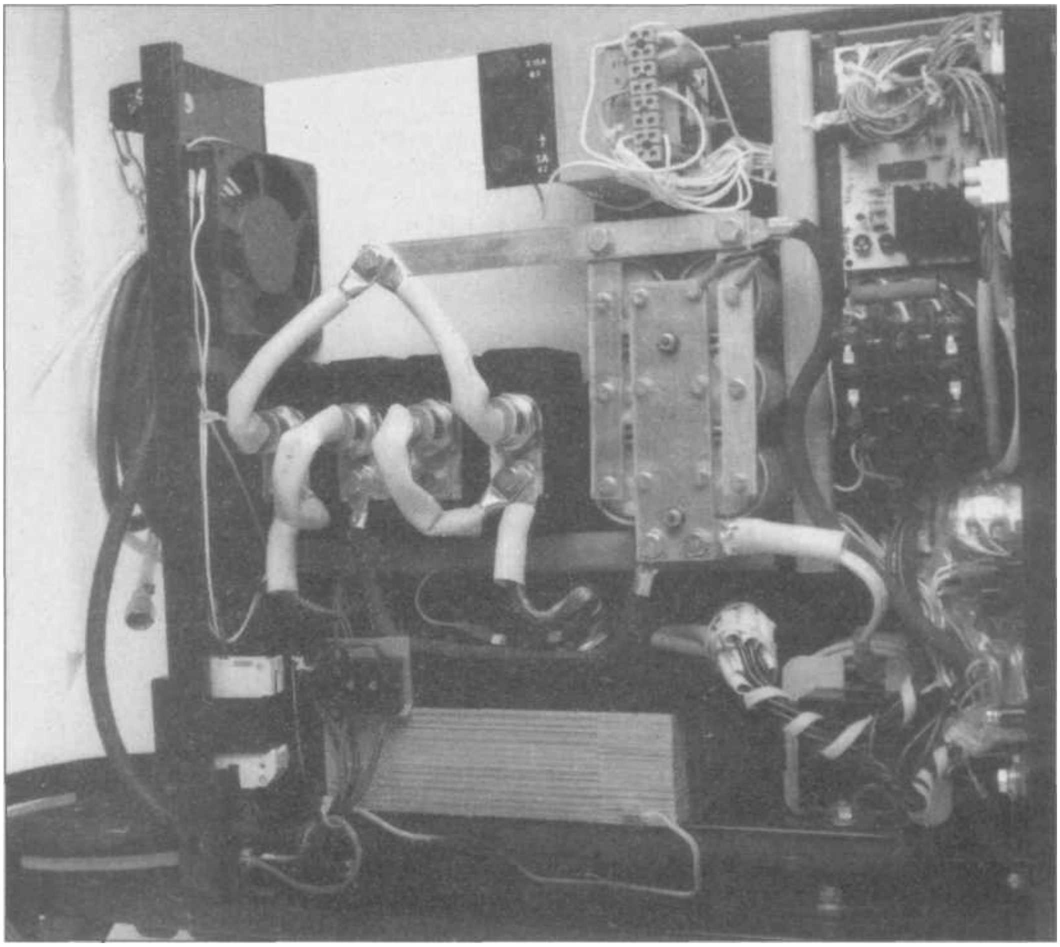

5.14 The bigger and better the machine, the more expensive the internal components. Sometimes you can get an idea of the relative complexity inside a MIG machine by comparing shipping weights when shopping for a welder. The more expensive ones are generally much heavier. This is the inside of a MIG 200 from HTP.

ing. Today there is a wide variety of machines offered from both domestic and imported sources that run the gamut from very small, very portable units that can be powered from a 3000-watt generator if necessary, to larger shop units with all the "bells and whistles."

When the domestic manufacturers started competing with the imports for the lower-price end of the MIG market, there was considerable talk at the local welding supply level about how the imports weren't as good as the domestic machines and that anyone who bought the cheaper imports would have trouble later on finding replacement parts. In the beginning of this marketing struggle, there was some truth to this, but the competition of today's marketplace has pretty much sorted these problems out. There are some very reputable companies selling imported MIG and Plasma machines, and most of them have a full line of consumables and other replacement parts on hand.

As with automobiles, there will always be some debate on import vs. domestic, but today such distinctions are harder to draw than they used to be. Like our cars, some domestic welding machines use components made overseas, and some imported machines use American-made parts. As you shop for a MIG machine or plasma cutter, you may still find resistance to imported equipment at your local welding shop. The answer is to listen to everyone, including especially anyone you know who has purchased and used the equipment you are contemplating. Find out if they have ever needed parts or service for their equipment, and did they have any problems with equipment under warranty. Be sure to compare the usage they put the equipment to, because there is a big difference between what is required in hours of welding on a daily basis in a shop, and your needs of welding a project or two a few times a year.

We have discussed duty-cycle comparisons in the previous chapter. Most manufacturers give you a duty-cycle rating at the highest amperage and the lowest amperage, and you may never have to use either one. If you contemplate a machine with say 140-amp capacity, it may have a duty-cycle rating of 95% at 30 amps and a 30% duty-cycle at its peak of 140 amps. The cycle relates to how many minutes out of a ten-minute period you can weld continuously without overheating the machine. At a 100% rating, you don't need to stop for anything, even if you are welding long seams together. In most home/shop situations however, you are not on a production schedule or a time-clock, and most of your joints will not be long seams. Also, you will probably weld a seam, such as a corner joint on a tubular framework for a utility trailer, that will take you only a minute or less to weld up (doing one side at a time), then you will move to another part of the project, perhaps set up some clamps or fixturing to ready the next joint for welding, and all this time the machine is cooling off so you may never exceed the duty-cycle even at higher amperages. This is where the big difference in cost in MIG machines is. The heavier-duty, shop-type machines have more expensive electrical components inside to handle a higher duty-cycle **(see illustration).** You should buy as much duty-cycle as you can afford to, but don't be too concerned that in your garage you'll spend all your time sitting around waiting for the machine to cool off. You'll find you will not often weld at the highest amperage output anyway, and when you do, even a 30% duty-cycle is livable.

5.15 Sometimes a medium-size machine is all that is called for in a particular shop, like this Miller 150 used in a muffler shop, but the weldor uses a large-sized shielding gas bottle for extra weld time. Note how this weldor has added a tubing arm to the cart, to organize the cables and hold pliers and and-spatter gel.

There are two basic types of smaller MIG machines you'll be looking at, the gas-shielded type that uses bare electrode wire and a bottle of compressed gas, and the flux-core type that uses no shielding-gas bottle but has a special wire with a fluxed core inside that produces the shielding gasses as you weld, much like the vaporizing coating on arc-welding rods. There are advantages and disadvantages to each type, and your choice depends on the kind of work you will be doing. Most of the machines on the market are for use with bottled shielding-gas, and there are some units that can be equipped for either type of wire, in case you have need for both. These machines cost a little more than single-use MIG machines because they have to have a switching device inside to change from reverse polarity (with the torch being positive and the work negative) which is required for the solid wire and straight polarity (with the torch negative) for the flux-cored wires.

The flux-cored wire machines do not make as pretty a weld, if that is a concern to you, and there is some slag to be removed afterward, though it is easier to remove than the slag from arc-welding. With some kinds of flux-cored wire, the slag will peel off the seam in strips. In body shop work, the extra thickness of the flux-cored wire may make them too hot for light-gauge sheet-metal work. Some experts say

5.16 A popular entry-level MIG machine is this 110-V Pocket-MIG from Daytona MIG. Though it weighs only 42 pounds, it has an amp range from 30-110 amps, capable of welding up to 1/4-inch steel. Very portable, it's ideal for home/shop use, especially on sheet metal.

5.17 A 110-V, 130-amp MIG is made by several companies for the home/shop market, this one is by Daytona MIG and welds up to 5/16-inch steel. It features a cooling fan and a 30% duty-cycle at maximum amperage, 95% at 30 amps.

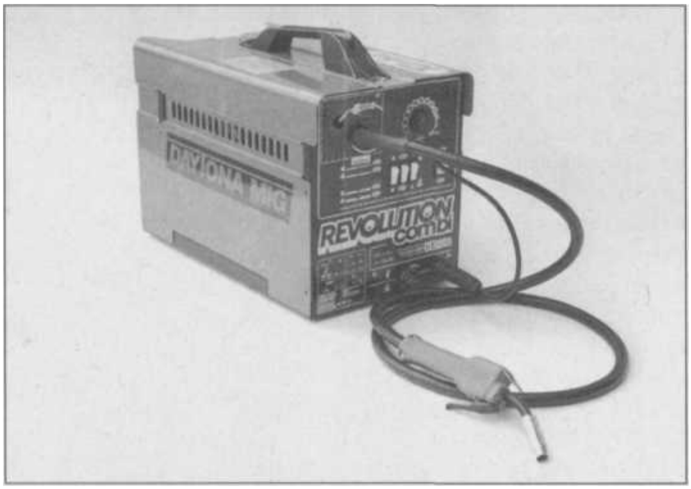

5.18 If you need the flexibility of welding indoors and out, a combination **MIG** that uses either gas-shielding or flux-cored wire fills the bill. This Combi 888 runs on 110V power, with 130-amp capacity.

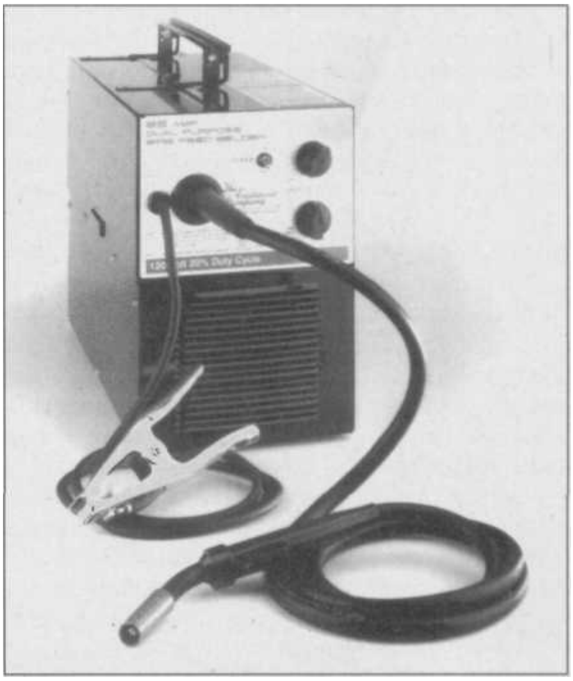

5.19 There are small **MIG** machines that come equipped for use with flux-cored wire, but have an optional kit that can be attached to work with shielding gas and solid wire. Eastwood Company's wire-feeder uses **110V** current, and has a 20% duty-cycle at its maximum amperage of 85. It's fan-cooled and adequate for light-gauge welding.

Gas Conversion Kit for Wire Feed Arc Welders
Operating, Safety, and Installation Instructions

5.20 Eastwood's gas conversion kit for their MIG welder contains the regulator, nozzle, tips and other parts to adapt a shielding-gas bottle to their 85-amp welder.

that flux-core MIG welding is best suited for 18-gauge or thicker metals, and most cars today are using 22- to 24-gauge steel, except for 20-gauge in areas like floor pans. Even replacement steel patch panels for cars are seldom thicker than 20-gauge. Virtually all of the newer cars are using lighter-gauge metal of higher-rated alloy to reduce the overall weight of vehicles to meet tougher fuel economy and emissions standards. The main advantage of the flux-cored machines are in outdoor or drafty situations where wind can blow away the shielding gas from a "bottle-fed" machine, ruining the weld consistency. The flux-cored wire works well for fencing repairs, farm equipment too big to be moved indoors, outsized sculptures and similar projects. These machines also have the advantage of not needing periodic refilling of gas bottles. As long as you have electricity and a spare roll of wire, there's nothing for you to run out of halfway through a project on a Sunday afternoon when you can't get your gas bottle refilled or exchanged.

The advantages of the bottle-fed machines are numerous, the primary benefit being the higher quality of weld and the ability to work on very thin materials without burning through. With shielding gas, there is very little spatter, no slag, and you can even weld body panels that are painted or rusty, though a cleaner parent metal will always make a better weld and produce less hazard from fumes. We have seen a big-rig repairman MIG-welding a new and painted fan guard to a rusty truck radiator-support framework. The budget for the job didn't allow for the time that would be spent in grinding and sanding all of the edges clean - the truck had to be back on the road making money. The result wasn't

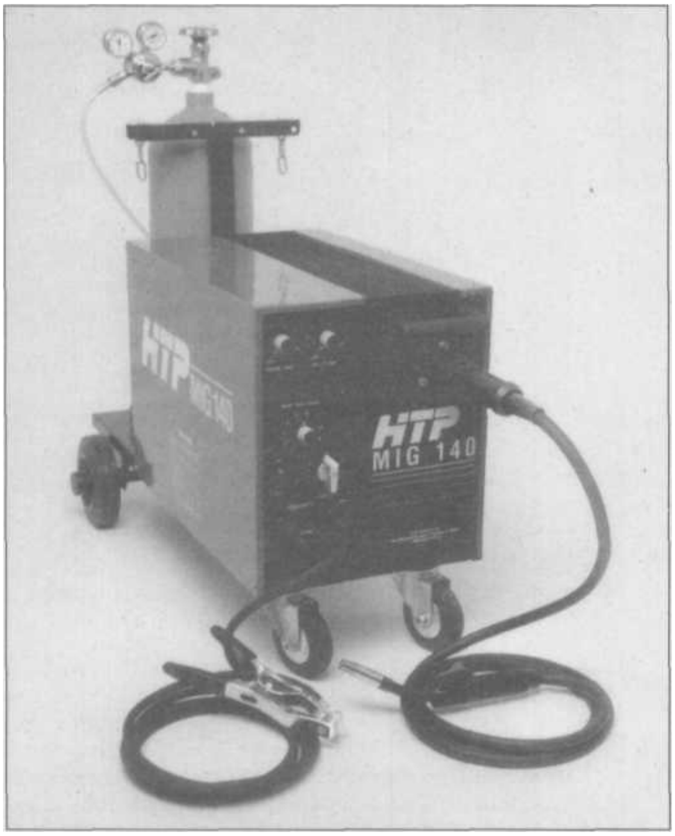

5.21 At the upper end of the 110V **MIG** scale are 140-amp units like this one from **HTP** America. It weighs 135 pounds, welds up to 1/4-inch steel, and has timing features such as stitch-welding and spot-welding.

GRENELL STUD / MIG WELDER
HTP
MIG 160

5.22 Above the 140-ampere range, you are into the 220V category of **MIG** welders, with more features and higher duty capacities. This is HTP's **MIG** 160, which can weld up to 5/16-inch steel, has eight heat settings, and is suitable for street rod shops, muffler shops, or farm use.

pretty, but it was more than strong enough for the application and a testament to what can be done.

As far as running out of gas is concerned, this is an issue, but most home/shop users handle this by just buying a bottle that is twice as big as they think they need. There is, of course, a pair of gauges on the tank that indicate the gas-flow going to the welder and the total pressure left in the bottle, so you should know before any big project just how much welding time you have left. Bigger shops just keep a spare bottle of gas on hand and switch if they run out, then send the empty back to the gas supply company for refilling to become the next spare.

Input current is the next consideration in shopping for a MIG machine. There's a number of home-type machines that can be operated on standard household 110V current, which negates the need for any rewiring of your shop area and also makes them more portable in the sense that you can use one anywhere there is a source of standard power. However, any circuit used for a welding machine should be at least a 20-amp circuit, and, in the case of the bigger 110V machines, a 30-amp circuit is recommended. The 110V (it's confusing, but household current is also referred to as 115-volt at times, but 110V and 115V are the same) machines are available in power abilities with a range of 30 to 110 amps on the smallest home MIGs and a 30- to 140-amp range on the largest. Most of these machines handle wire diameters from .023-inch to .035-inch, though some of the imports are set up for metric wire sizes, which are close enough to US sizes that you can always find replacement wire to suit them.

Above the 140-amp range in maximum power brings you into the bigger shop-type machines that require 220V current input. Most of these machines

MIG
181 C
DAYTONA MIG

5.23 This 220V unit, the 181C from Daytona MIG, offers settings from 20 amps for light sheet metal to 180 amps for heavier work, and has a spot-timer and quick-disconnect torch. This is suitable for light fabrication shops.

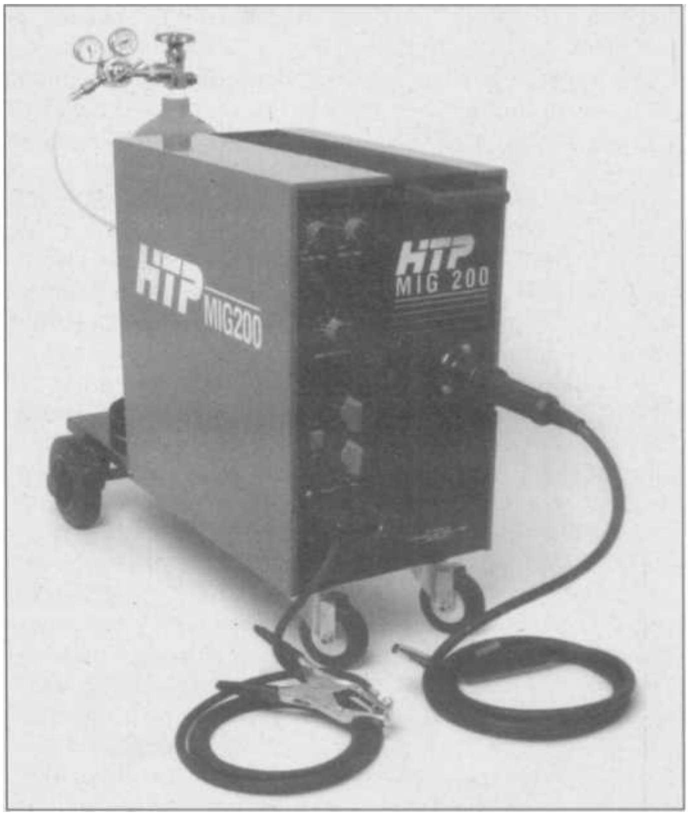

5.24 For welding up to 3/8-inch steel, a **MIG** like HTP's **MIG** 200 will do the job. It has 24 heat settings, up to 200 amps, and is suitable for farm or ranch work, commercial fabrication, or repeated automotive projects like building trailers or engine stands.

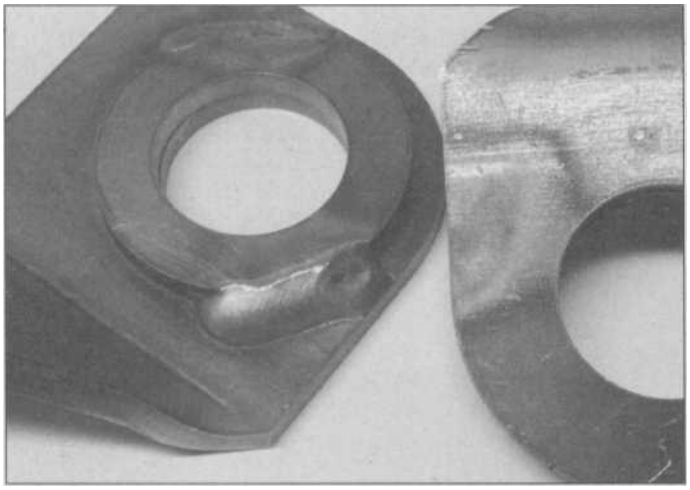

5.25 This close-up of a commercially made street rod part has a thick steel ring MIG-welded in two places to a heavy bracket for suspension mounting. This is a large bead from a good-sized machine. Note how smooth and clean the weld is, the bead looks almost like one piece, rather than overlapping puddles, and look at the backside of the part (at right) to see the indications of penetration. A **MIG** is easy to learn, but it will takes lots of practice to make welds like this.

have an upper limit of 250 amps, which is a *lot,* and they offer a 50% duty-cycle even at this high amperage. These larger machines also offer more "bells and whistles" such as larger wire spools, the ability to handle more different wire sizes, wheels on the bottom of the machine as standard equipment and the capacity and bracketry to hold larger bottles of shielding gas. Designed as they are for bigger jobs, these medium-size machines can handle wire diameters from .023-inch to .0625-inch (1/16-inch), which is bigger than virtually any requirement you will have need for in your home shop unless you take up amateur bridge building. You will see these 250-amp MIG machines hard at work in many shops today, from muffler businesses to street rod shops, race-car shops and general steel fabricators.

Other options you may find on these mid-size machines are controls for timing various weld functions. If you were regularly spot-welding long seams in sheet metal, you could set one of these machines up with a special spot-welding nozzle which has a flat front face with cutouts on each side. You bring such a nozzle directly in contact with the metal and pull the trigger. An adjustable timer on the machine sets the weld time for a perfect spot-weld with little or no crown or buildup to grind off later. Also, you can set one up to do stitch-welding as well, where you join work with a series of spot welds that overlap. The welding machine has a stitch timer that turns the arc on, pauses, arcs, pauses, arcs, etc. This stitch mode isn't used very often but is helpful in joining very thin materials where there is a chance of burring through with continuous welding. In the stitch mode (in which the pause time is variable at the machine), the pause time allows the first puddle to solidify before the next puddle is made, reducing the overall heat.

Mid-size MIG machines also offer more range of heat settings than the entry-level MIGs. While the smaller machines may have four settings from their lowest to their highest amperage, their bigger brothers may also have a "fine-power" adjustment knob, which essentially gives four more positions for each of the four basic settings on the main amperage knob. So instead of four settings, you now have a possible sixteen.

Adapters and special attachments offered on some machines include a metal shrinker and a stud-welding attachment. The metal shrinker works to reduce high spots in sheet metal, much as you would do with an oxy-acetylene torch, but without the flames and with very controlled heat. The basic principle is to heat the center of the high spot (such as from having hammered out a dent from the backside) to cherry-red, then quenching it with a wet rag, which "shrinks" down the high spot of stretched metal: The MIG shrinker attachment screws on in place of your contact tube, the wire is pulled back inside the torch (the wire isn't used) and the wire drive and gas flow are disconnected. The shrinking tip is placed in contact with the metal and the trigger pulled briefly, heating the high spot.

The stud-welding attachment is another body-working tool that allows you to "spot-weld" small studs onto a dented panel. These studs fit a special slide-hammer that pulls on the studs to work the dent out without having to drill holes in the body and use a screw-type puller as has been traditional. The studs can be clipped and ground off easily after the pulling is done. There is also a separate stud-welding tool that operates on 110V current without the need for the welding machine (see the Safety and Shop Equipment chapter). Another stud-welding attachment is made to use a threaded stud as an "electrode" and fuse it in place of a worn-out stud in automotive work **(see illustrations).**

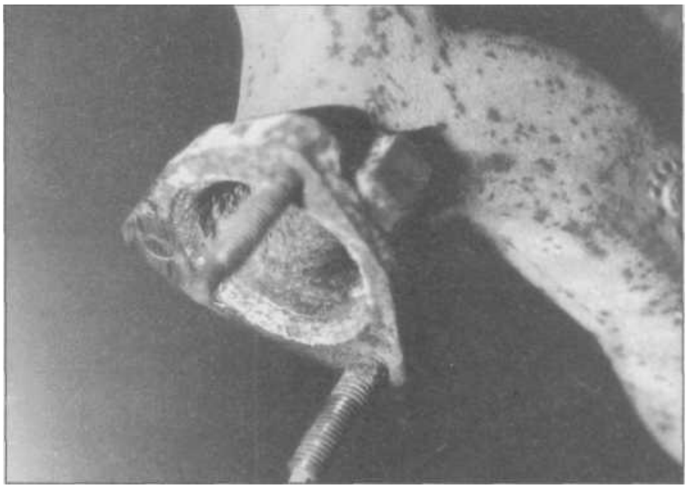

5.26 In exhaust system work, muffler shops deal with broken and rusted-away manifold studs all the time. HTP has a special feature to their MIG 160 that repairs such problems. First the old stud is cut off flush.

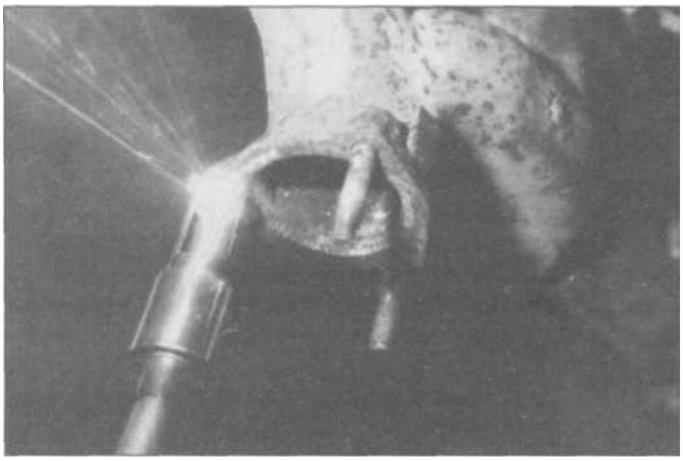

5.27 A special attachment to the MIG gun holds a hollow replacement stud. When the arc is turned on for about four seconds the wire feeds through the stud and welds it to the old stub.

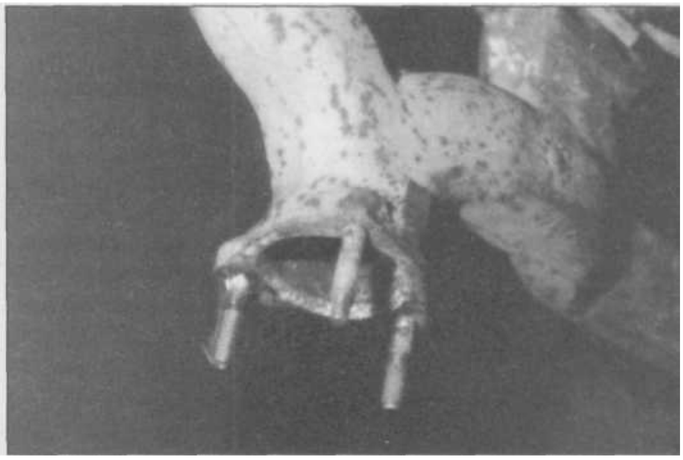

5.28 The new stud is now welded in place, with strength said to be up to 85% as good as original. This feature is a real time-saver for busy muffler businesses.

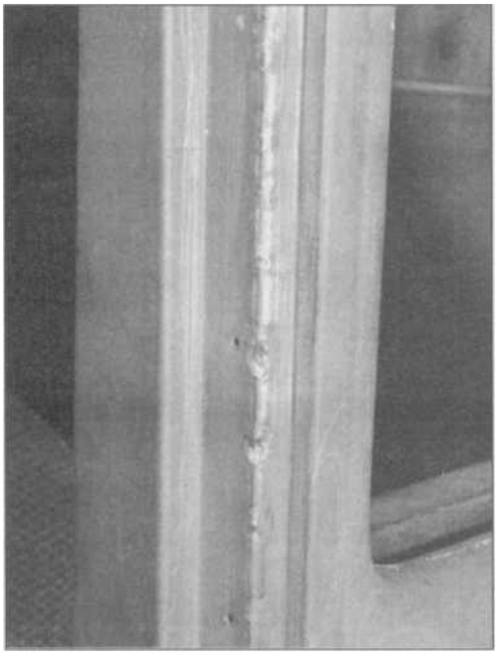

5.29 Spot-welding of sheet-metal panels is easy with a MIG, even a small one. For large panels, welds may be made through holes in the top panel. Here a new door jamb has been added to a 1932 Ford with a series of small spot welds. So little weld is above the surface that no strength is lost when the area is ground for painting.

To go to the top level in MIG machinery, you are now talking about very expensive professional welding machines with large, multi-purpose power supplies. As mentioned in Chapter 2, the bigger welding machines designed strictly for professional shops are often multi-purpose power supplies capable of high amperages which can be set up for Arc welding, TIG and MIG, with any polarity required. The amateur user will never need the high amperages or high duty-cycles of these top-level machines. To use the big machines for MIG welding usually requires that a separate wire-feeding accessory drive be set up, and, with the lower prices of today's MIG market, the big shops prefer to just buy a separate MIG welder for those needs, using the large power supply just for TIG welding.

Choosing shielding gas

By now, you should have an idea of what level of MIG machine will suit your purposes. If you are like most home/shop users who aren't going to be welding every day, your needs may be best served by a typical, entry-level 140-amp, 110V MIG machine using shielding gas. You now have to choose a shielding-gas bottle (not included with most MIG machines), some wire (a small sample roll may be included with your machine), and the type of shielding gas you want to use. This sounds like a lot of choices, but the field is narrowed down by what kind of work you want to do and how often you need to do it. In all cases, the choice of shielding gas and wire type must be matched to the kind of material you will be welding.

Capacity of the gas bottle is important, and most first-time weldors make the mistake of buying less capacity because the bottles are cheaper, only to find themselves someday working on a project over a weekend (when there is little

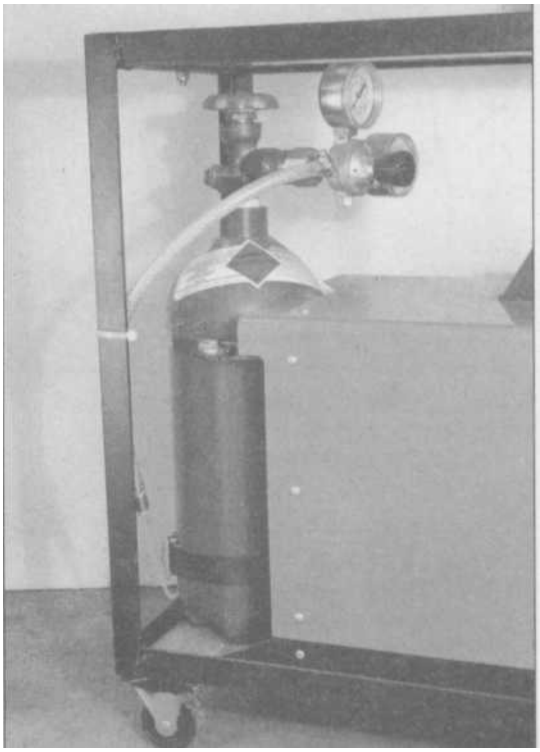

5.31 The smallest bottles are 20 cubic-foot capacity, and easily fit into even a small welding cart like this. The bottle and regulator are totally protected here. This is a good size if you only use the welder infrequently, or on small projects.

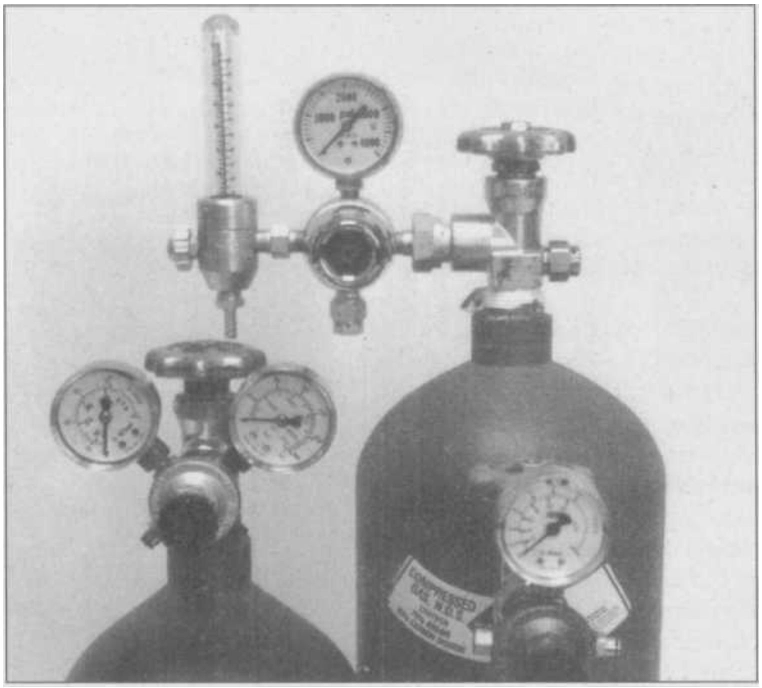

5.30 various size shielding-gas bottles are available for MIG setups. Check with your welding supply store on how long each size might last. Most home/shop users buy relatively small bottles because they want portability in a small garage area.

5.32 Gas-shielded MIGs are not made for outdoor work, but as long as you can keep the wind currents away, they work fine. On this driveway utility-trailer project, pieces of plywood are used as wind shields while welding.

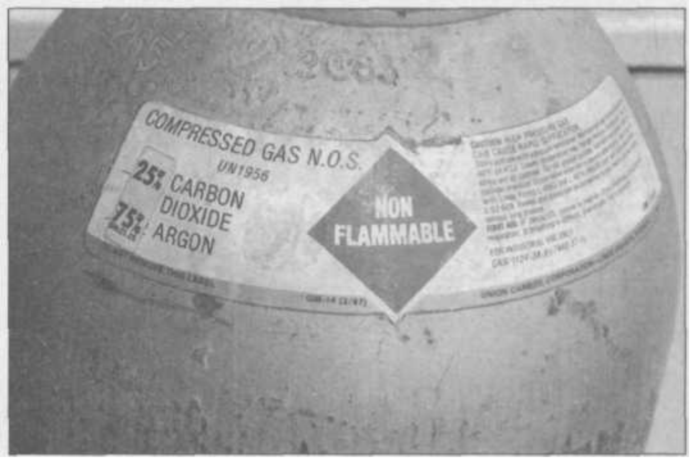

5.33 Most of your welding with shielding gas can be done best with a 75% argon/25% CO_2 mixture, which is excellent for cleanliness and penetration, and can be used on cast-iron and stainless-steel as well.

chance of getting a bottle refilled) and they are out of gas. Gas bottles are available in sizes from 20 cubic feet to 40, 60, 80 and 125 cubic feet of gas. This roughly translates into welding time of from four hours for the 20 cu.ft. bottles to 14-16 hours for the 125 cu.ft. when used on a small machine like the 140-amp unit in our example. There's a number of factors involved in how long a particular bottle may last under different welding conditions, so these are just rough guidelines. Every time you use your MIG machine, you should look at your tank gauge to see how much more gas is left. You may even want to just lease a small cylinder of gas when your first get your machine just to see how long such a tank would last you, and then buy a bigger one if you think you need it.

The small bottles fit nicely under a small welding cart, making for a very portable arrangement for occasional home use **(see illustration).** The larger bottles are heavy and bulky, so you'll have to build a welding cart (most first-time weldors' first project) that can accommodate one. Of course, bottles of compressed gas are dangerous, even if the gas is non-flammable, due to the pressures involved, and any welding gas bottle must be securely mounted with chains or sturdy clamps. It's important to note that welding bottles for electric welding machines should be ideally mounted in an insulated manner, such as with rubber underneath and rubber hose over any restraining chains. This prevents the remote possibility of arcing against the cylinder, which could damage it.

Next there is the choice of shielding gas to use. One of the widely used shielding gasses for MIG work is plain carbon-dioxide, or CO_2. While it is in the bottle, CO_2 is really an inert gas, but technically it turns into carbon monoxide and oxygen under the intense heat of the arc process. The oxygen can combine with other elements in the air and in the parent metal to form undesirable oxides. Thus, plain CO_2 will not usually result in as clean a weld as with other gasses, but it is used often because it is the least expensive shielding gas for welding steel. It will weld very fast, with good penetration, but you should use a welding wire that contains deoxidizing elements to counteract the effect of pure CO_2.

Argon is a versatile inert shielding gas that is often used by itself, or in combination with other gasses to produce certain weld characteristics. For instance, welding of non-ferrous materials like aluminum is usually done with pure argon. It makes for good penetration patterns and a concentrated arc.

To weld ferrous materials, like the mild steel we will use for most home/shop projects, argon is usually mixed with another gas, or sometimes two other gasses, to provide special characteristics. While there are some esoteric mixtures of gasses in different percentages for specific purposes, a mixture of 75% argon and 25% CO_2 has become pretty much the standard for welding mild steel with short-arc MIG machines with wire diameters of .035-inch or less (see illustration). This mixture, often abbreviated as C-25, is more expensive than the plain CO_2 but produces much less spatter and consistently better looking welds, even on materials that exhibit minor rust or scale. It is what we would recommend for most of your welding needs. If you were doing a lot of thicker materials, overhead or other out-of-position welding and pipe welding, you might use a 50-50 mix of argon and CO_2 which offers good wetting and bead shape without the excessive fluidity that causes a bead to droop or fall when doing out-of-position welding.

If you are working in a body shop, you might be interested in a mixture that Airco calls Argoshield LG®, which stands for light gauge. It consists of mostly argon with small additions of CO_2 and oxygen. The combination is designed for metals down to 20-gauge, and produces good penetration, but with a smaller weld bead than normal for less sheet-metal finishing, less spatter and smoke and good arc starting. It is specifically formulated to work well on the thin, low-alloy steels now used on auto sheet metal.

For welding of non-ferrous materials, either straight argon or mixtures of argon and helium are used in various combinations which provide higher heat to a MIG arc. Usually, the thicker the material to be welded, the higher the percentage of helium that is included in the mix, and the HE-75 gas, which is 25% argon and 75% helium is typically used in industry to weld thick aluminum.

If you were interested in welding stainless-steel, you can actually use the C-25 gas we recommended above for mild steel, but a mixture of 90% helium, 7.5% argon and 2.5% CO_2 is widely used in stainless-steel MIG welding because it offers a higher heat for the normally sluggish weld puddle on stainless, as well as offering good stability, penetration and resistance to corrosion.

If you have to pick just one type of gas, the C-25 works great on steel, can be used on stainless-steel and is also useable on cast iron.

Choosing wire

The bare steel wires used in a bottled-shielding-gas MIG machine are all variations of the same basic metal, with varying amounts of deoxidizers added for different applications and shielding gasses. The most common deoxidizer is silicon. Others added may be manganese, aluminum, titanium, or zirconium, with nickel, chromium or molybdenum added to improve mechanical or corrosion-resistance properties. The welding wires with higher levels of deoxidizers generally are better able to weld on rusty or unclean steel surfaces.

Most of the steel wires for MIG welding fall under a designation from the American Welding Society (AWS) of E70S-something, with the last digit being the particular variation. For instance, one of the most common wire specifications is E70S-2, which is deoxidized, and makes a good weld with C-25 gas, even on rusty steel. Its main drawback is that it lacks fluidity; the puddle doesn't want to flow out width-wise and may not stick well in heavy materials.

The E70S-3 wire is one of the most common and least-expensive MIG wires available, with more deoxidizers and a more fluid puddle that makes a wider bead. It has been used successfully for years on cars, farm equipment and appliances.

Next up in terms of deoxidizer content is E70S-4, which is a medium-priced wire suitable for almost all steel welding. It offers good fluidity and better arc characteristics than E70S-3, but has slightly more spatter, and is used on structural steel, ships, piping and boiler vessels.

Probably the high-performance and higher-cost wire available is the E70S-6, which has the highest level of silicon and manganese as deoxidizers. It is suitable for welding almost all steels, from thin mild steel to 1/2-inch plate (in the appro-

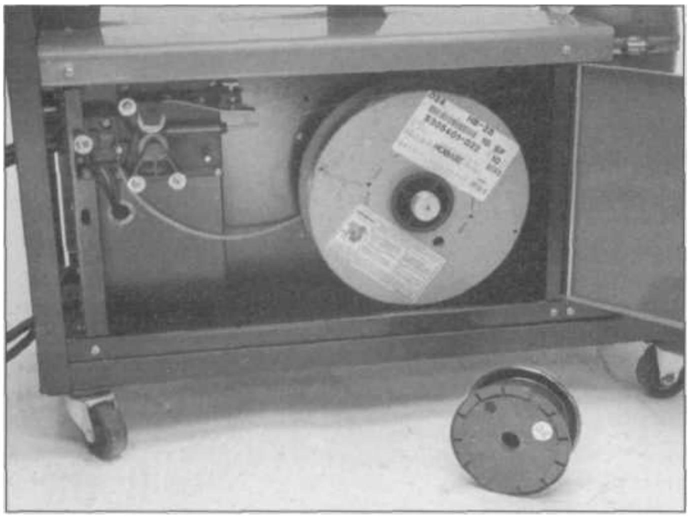

5.34 Inside the wire-drive portion of a typical small MIG machine, this machine comes with a small wire roll such as at lower right, but comes with an adapter to handle the larger rolls such as in the machine. The small wire rolls are inconvenient unless you do only small projects.

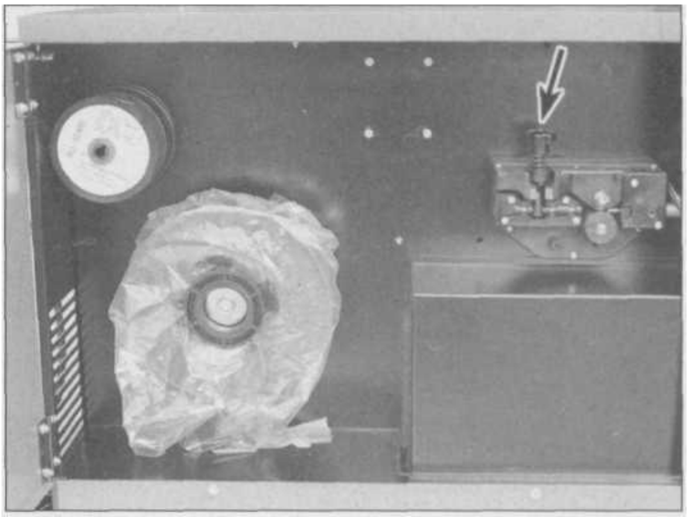

5.35 This is the drive section of a Daytona MIG wire-feed, which has a place inside to store an extra small roll of wire, good for keeping the spare clean if you occasionally switch to another type of wire for specific needs. Arrow indicates the knob for adjusting the wire roller tension.

priate wire diameter) and works with the most popular gas mixtures, offering good puddle fluidity, too **(see illustration).** It is one of the few MIG wires that is completely compatible with today's thin, low-alloy, high-strength body metal in unibody cars.

L-Tec markets a wire meeting the AWS classification for ER70S-2, 3,4 and 6 that they call Easy Grind, which is designed specifically for auto body work. It is said to require less skill to weld thin materials, and to have a reduced carbon content that means easier grinding to prepare for filler and painting. Besides automobiles, it is also used in welding ductwork, metal cabinets, lawn furniture, and is designed to be on metal as thin as 26-gauge with smaller MIG machines.

Non-ferrous metals, of course, require different wires. The basic rule of welding electrode selection is to choose a wire or electrode whose content is much the same as the parent metal you are welding. In the case of stainless-steel, it *can* be welded with the steel wire we have recommended above, but for better corrosion resistance (the main reason for making something from stainless-steel in the first place) stainless-steel wire should be used. There are various alloys of stainless-steel, but one of the more common varieties is type 304, and the best wire for MIG welding type 304 is called ER308, with variations of increasing silicone content in ER308L and ER308LS.

For aluminum welding, you need aluminum wire. Again, there are many alloy variations of aluminum, and several metal blends in aluminum wire, but perhaps the best MIG wire for aluminum is called ER5356, which has a chemical content proven to work in most situations. Aluminum welding is more difficult than steel in many ways, and often the exact alloy of the parent metal is impossible to determine, unless you have bought a large enough piece at a metal yard that is still factory marked as to its alloy and heat-treat. There are even some aluminum alloys which are considered unweldable, like high-strength aluminum such as 7075-T6.

5.36 Your welding supply store will have a variety of MIG electrode wires to choose from. The E70S-6 wire at left is one of the most popular. At right is L-TECs Easy-Grind for body work.

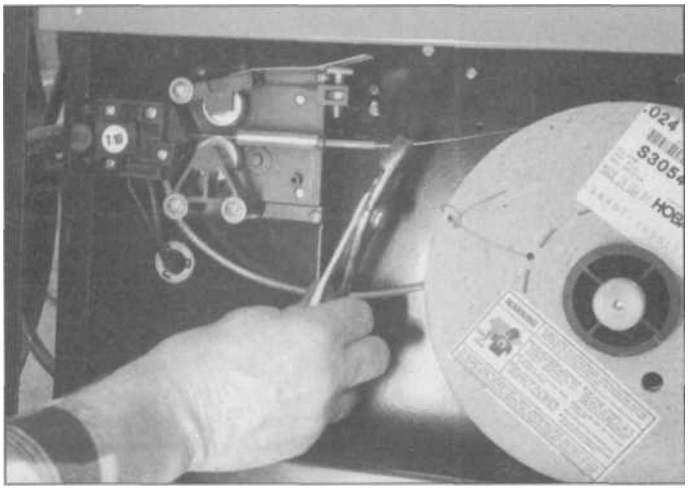

5.37 When starting in a new roll of wire, keep a pair of pliers on the end you take off the reel, or the wire can unroll everywhere. Feed the wire through the guide and over the drive roller.

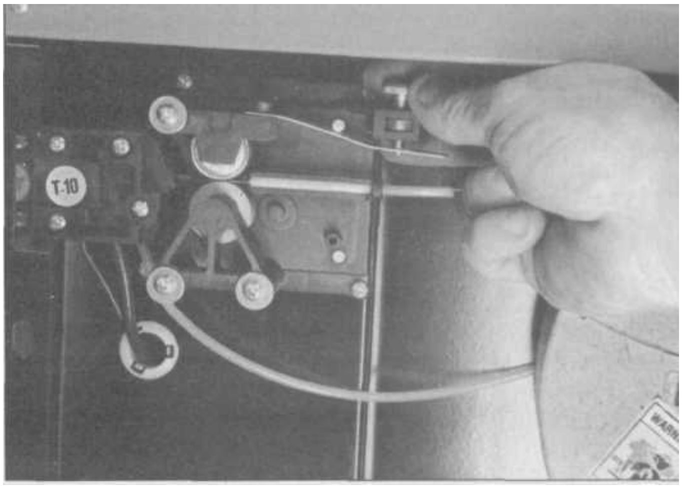

5.38 The wire tension is adjusted with a knob, and proper tension avoids wire-twisting problems inside the gun. The wire should be able to slip when it needs to. Most machines have a feature that allows you to retain the tension setting, once set, by swinging the tension arm away when changing wires.

Learning MIG welding

One of the attractions of MIG welding that has brought these machines into the home/shop market is how easy they are to learn. With a little instruction, virtually anyone can be MIG welding with an hour's practice time. As with anything else, the more you practice, the easier it will be and the better your welds, but before you get overconfident about your MIG welds, try cutting apart a seam you have made to check the penetration, and make the other tests (bending two butt-welded sample plates in a vise) we have previously described to see if the welds you're making are as strong as they are good-looking.

To set up your MIG machine, the first step would be to read all the directions and cautions in the instruction book that came with it. One of the few adjustments you will have to make initially is to put a roll of wire into the machine and set the drive-roller tension. The wire unrolls from the reel through a guide and over a motorized roller, which feeds the wire through another guide and into the cable going to your torch. Lay your torch and cable out on the floor as straight as possible. Mount the wire spool into the machine as per your directions, but be careful when cutting the end of the wire which is usually bent over to lock it into a hole in the side of the reel. There is winding tension on the wire and if the cut end of the wire gets away from you, it can start unraveling all over the place. Don't cut the end loose until you are ready to feed it.

Hold the cut end with some pliers or vise-grips and carefully file or sand the cut end until it is smooth. A sharply-cut end may snag as it travels through the cable or torch. Insert the end into the guide, loosen and swing away the wire-tension adjuster, and feed the wire over the groove in the drive roller and on into the next guide, still holding the wire firmly **(see illustration).** Most machines are set up to handle several sizes of wire, but you must be certain that you use the drive-roller groove that is the right size for the wire you are using. Most rollers have two sizes on them and can be turned around on their shaft to put the other size in line with the wire. There will probably be an extra roller with other sizes on it that comes with your machine.

When six inches or more of wire have been fed into the welding cable, put the wire-tension adjuster back in place over the wire, keeping the wire centered in the groove. Turn on the welder and put the wire-feed speed about one-quarter

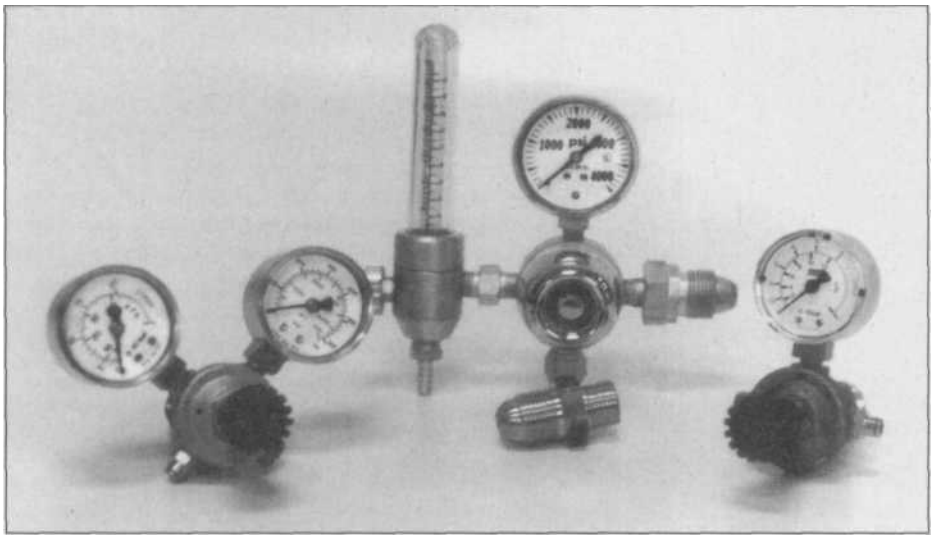

5.39 There are three choices in regulators for MIG shielding gas. The unit at right is a single-gauge, economy unit which indicates only the line flow, not the total tank pressure, at left is a dual-gauge unit that shows both, and in the center is the more professional model that has a floating-ball flow meter easy to read from a distance.

of the way and pull the trigger. At this point, adjust the pressure on the wire tensioner until the rollers are feeding the wire, and then turn it 1/4-turn more. With the nozzle and contact tube unscrewed from the end of your torch, keep the torch cable straight and run the machine until the wire comes out of the torch. Slip the correct-size contact tube over the wire, screw it in and attach the nozzle.

The final check on wire tension is to make sure the wire can "slip" if necessary. Eventually in your MIG practice, you will virtually weld the wire to the contact tube. When this happens, and you are still pulling the trigger, the machine tries to feed more wire but it can't come out of the gun, so it jams up inside in what is called a "birds nest." To prevent this, bend the wire to the side of your torch and put the nozzle right against the shop floor, simulating a stuck wire. While watching the drive mechanism inside the machine, pull the trigger. The drive rollers should have just enough tension to always pull the wire through, but not enough to prevent slippage in case of a jam. If your rollers *do* slip in this test, you're set.

If yours is a flux-cored-wire machine, you are ready to weld, but for the shielding-gas machines, you have to set up the bottle. With the bottle mounted firmly on your cart, attach the regulator to the bottle and attach the reinforced hose (should come with your machine) to the regulator with a hose clamp, connecting the gas to the machine. On some machines, a special adapter connects the reinforced hose to a small plastic hose going into the machine. Keep the two regulator gauges protected from impact, and always keep the gas valve turned off when you are not using the machine. Some cylinder valves only seal well when the valve is either fully off or fully on, so when your machine is on, turn the cylinder valve all the way, not just a few turns. You can check any of your gas connections for leaks using some soapy water solution. Bubbles indicate leaks. Adjust your gas regulator to the proper flow rate (cubic-feet-per-hour rate, according to your instructions) for the material and wire you are working on. **Note:** *One gauge tells you the flow rate and the other tells you how much pressure is still in the bottle. You can't read either one unless the cylinder valve is turned on, so don't look at your bottle one day and notice that the tank gauge looks empty - turn on the valve to really see what's left.*

Now you're ready to practice. Select some 1/8-inch steel, set your amperage according to the suggestions in your instructions, and set your wire speed a little

5.40 The basic MIG units have only a few power settings. Your instruction sheet will tell you the recommended thickness of metal for each setting, and a chart on the front of the machine usually tells you the amperage and duty-cycle of the settings. After some practice on different materials, you will learn instinctively which settings work best for your work.

5.41 The wire speed has a lot to do with the heat and quality of the weld. Generally when you have it right, the MIG machine makes a very satisfying steady "sizzle" sound. Practice by adjusting the feed high and gradually lowering it until the sound and results are satisfactory. If you do repetitive projects, write down the amperage and wire speed settings that have worked for you.

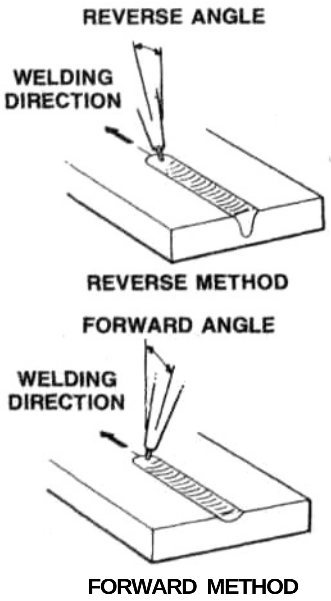

Gun *tip inclination and the direction of the weld can be changed to match different welding conditions*

5.42 Of the two methods of gun travel with MIG welding, beginners often like to use the reverse or backhand method because they can watch the puddling action much better.

higher than what's recommended. The wire speed is key to getting a good weld with proper heat and penetration, and tuning the speed is easy. Listen to the sound the welder is making as you make a straight bead. If the speed is too fast, most of the wire coming out will be red-hot and there will be a loud crackling noise, gradually adjust the speed down until you get a steady sizzling sound. The wire should be melting right at the weld. Practice tuning in the wire speed at different power settings and metal thicknesses.

There are two torch-holding techniques for MIG welding, forehand and backhand **(see illustration).** The first is when the torch is angled and moved such that the weld is taking place ahead of the torch's direction of travel, and the backhand is when the weld takes place behind the torch **(see illustrations).** Most beginner find the backhand (also called "dragging" the weld) technique works best on most steel-welding projects, giving the weldor a better view of the puddling action. In either style, the torch is angled about 35 degrees to the work. If you find yourself doing a vertical seam, you can travel either up or down, but use the forehand technique (sometimes called "pushing" the weld) only when going up, and the backhand technique going down, keeping the arc on the puddle's leading edge at all times for good penetration.

After making good straight beads on top of the plate, try joining two plates. If you are working with materials under 1/8-inch thick, most steels can be MIG-welded without beveling the edges, but you should tack the parts together with a slight gap between them before finishing off the seam with a full weld. This assures good penetration. On 1/8-inch and thicker metals, the edges should be beveled before tacking them. If you find that you are consistently burning through the material, the heat setting is too high, or you are moving too slowly. If it is "cold" appearing and not penetrating, the amperage may be too low, or you may be traveling too fast. You will adjust your travel speed with the torch by watching how the puddle develops, moving the torch along at a slow, *steady* rate. One of the great advantages of the MIG process is that your torch is right down close to the work, and there is much less intense light, heat and spatter as with arc-welding; you can get your head closer to the work to really see the puddle progress. Keep a pair of sharp wire-cutting pliers on your welder at all times. You will use it when you start each time to maintain about 1/2 inch of wire sticking out of the

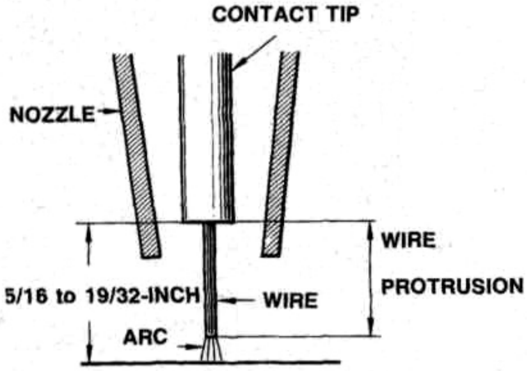

5.43 The tip-to-work distance is very important in maintaining the short-circuit arc. When practicing, watch the arc closely. The white-hot, melting end of the wire should stay about the same short distance from the puddle. Wire speed and welding travel speed control this.

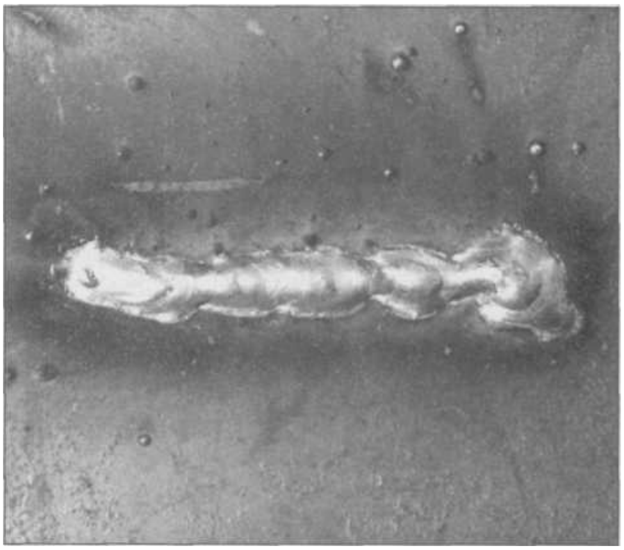

5.44 This is what some of your practice welds will look like. The heat setting was too hot here, and the weldor wasn't very steady with the gun. Use both hands and find a very comfortable position to hold it steady. Note the amount of spatter from making and breaking the arc.

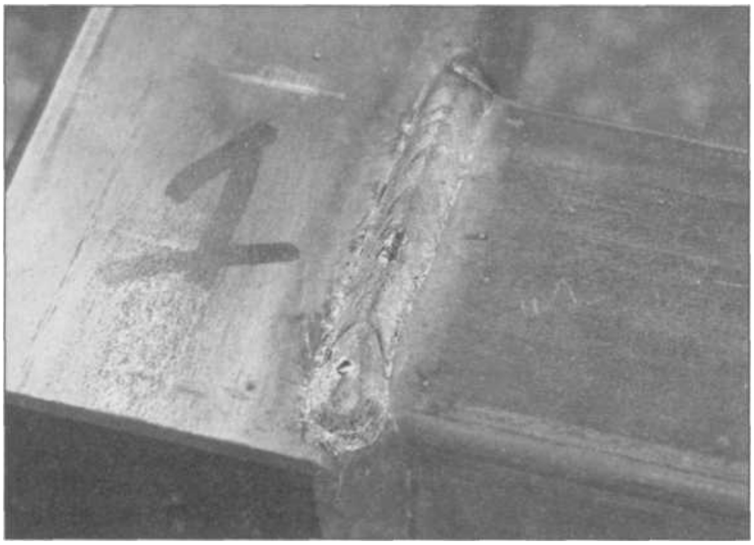

5.45 This production weld by a professional shows how clean and perfect MIG welds should look. The heat-affected area on either side is small, the weld has good penetration, and there is minimal spatter. Heat was evenly played on both pieces of tubing.

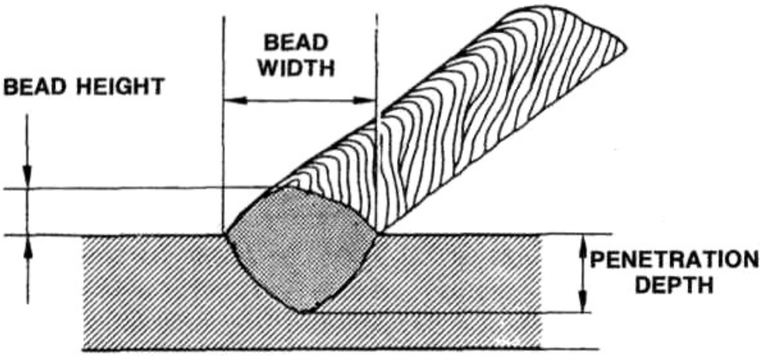

5.46 Welding current affects the penetration of the weld and the size of the bead. Ideally, you are looking for good penetration without a too-tall or too-wide bead.

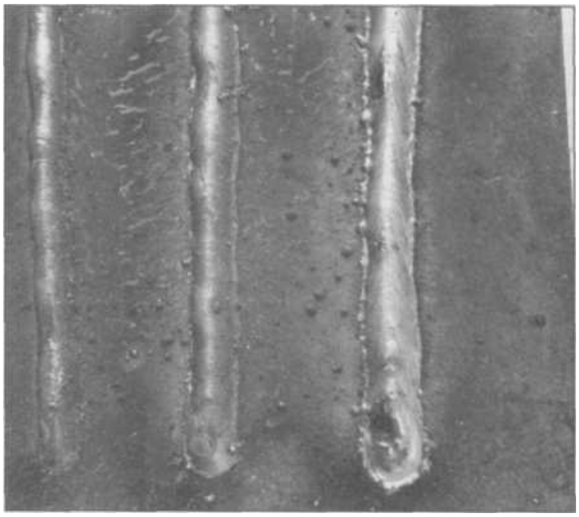

5.47 At left here is a MIG bead that was too low in amperage for the thickness of the material. Penetration is shallow and the bead sticks up too high. At center, the correct amperage was used, and at right, the amps were too high. The right-hand bead may look OK here, but penetration was too deep (dropthrough on the backside) and there was some undercutting on either side of the bead.

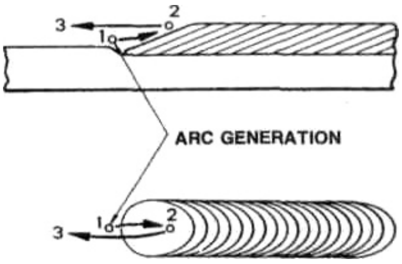
3
1
2
ARC GENERATION
3
1
2

OVERLAPPING BEADS

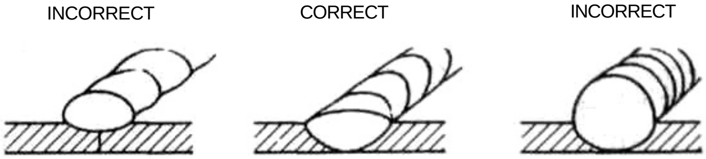
INCORRECT
CORRECT
INCORRECT

Insufficient penetration. Weld strength is poor and the panel could separate when it is finished with a grinder.

Good penetration and easy to grind.

There js good penetration but finish grinding will be both difficult and time consuming.

5.48 After you have practiced sufficiently, you'll develop a natural rhythm for the torch movements to create a good bead of overlapping oval puddles.

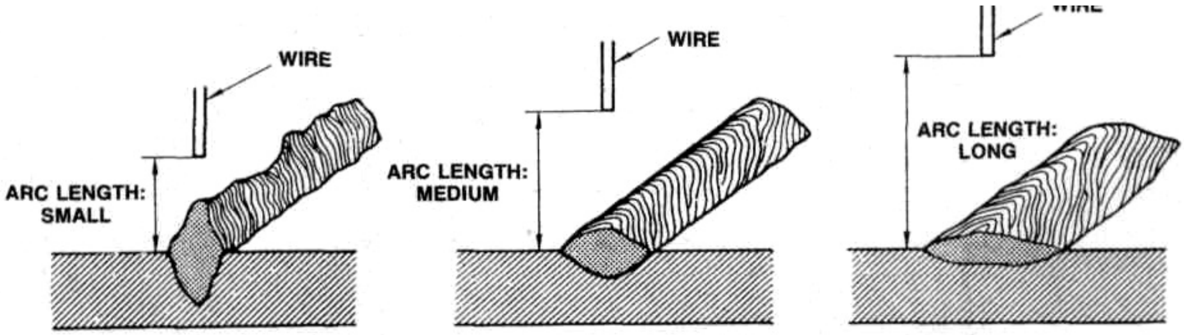
WIRE
ARC LENGTH: SMALL
WIRE
ARC LENGTH: MEDIUM
WIRE
ARC LENGTH: LONG

ARC VOLTAGE: LOW

ARC VOLTAGE: MEDIUM

ARC VOLTAGE: HIGH

5.49 Arc voltage and length affect the shape of the bead's profile and its penetration. Cut a few of your practice welds in half to examine the beads. The shape of the overlapping puddles should have rounded ends exposed, not sharp, triangle ends.

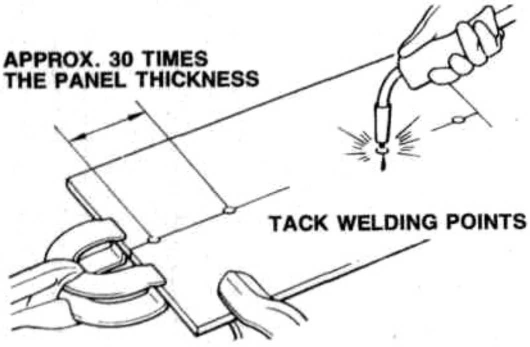
APPROX. 30 TIMES THE PANEL THICKNESS
TACK WELDING POINTS

5.50 The thickness of the panel you are working on determines how far apart your initial tack-welds should be to firmly align your parts. On thin materials, you need a lot more tacks.

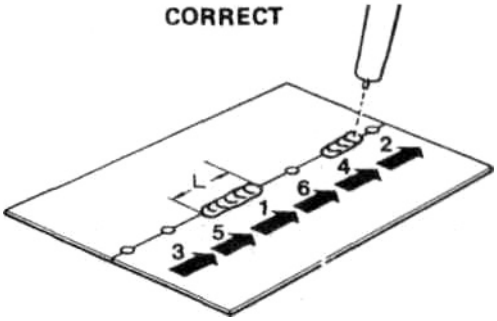
CORRECT
L
2
4
6
1
5
3

5.51 When working on large panels of thin material, such as when doing body work like welding on a new quarter-panel, warpage is a problem unless you keep the heat down by making lots of tack welds, then switching back and forth to the coolest section to add a short weld, then move to another section, and switch back and forth until the whole seam is done. A continuous weld might put too much heat in, and warp the new panel, requiring lots of hammer-work to straighten.

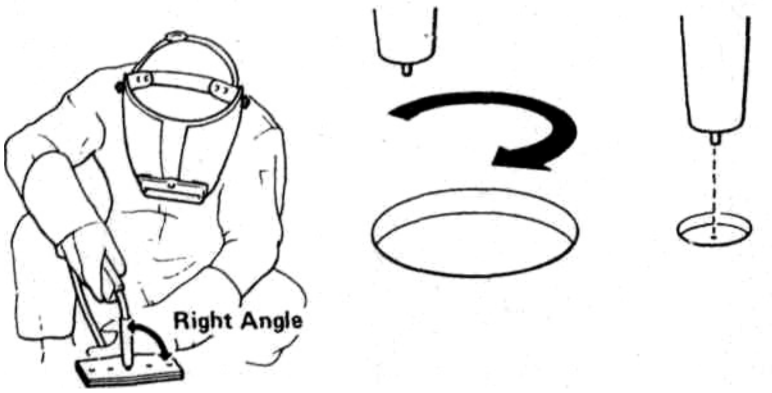

5.52 Instead of a straight butt-weld, panels of thin material are often joined with spot welds, in which the top panel is punched with small holes. You MIG-weld around the hole, joining the top panel to the bottom.

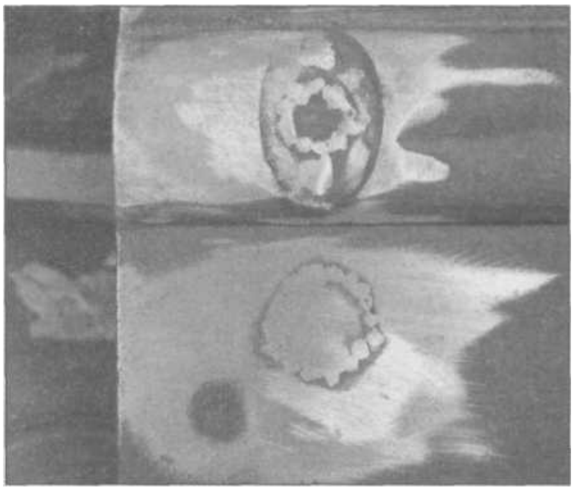

5.53 Door-handle and lock holes have been filled in on this street rod door, using small circles of sheet metal the same thickness as the original metal. In order to avoid warping, the circles were filled by tacking the filler pieces in a few places, waiting for the part to cool off, then adding a few more tacks, until the whole circle was made of tack welds.

gun. If you are welding thicker materials, or two edges that have irregular gaps, you can weave the arc back and forth from side to side as you travel, bridging the distance between the two edges.

Sometimes a start-and-stop technique works well for bridging gaps or filling holes. You watch the arc and weave from one side to the other, then pause (let go of the trigger) for a second with the torch in the same spot, allowing the puddle to solidify, then pull the trigger again, building a new puddle onto the last one. To fill a hole, just keep making tack-welds around the edge of the hole until you have made it smaller **(see illustration).** Let the metal cool, then add an inner row of tacks inside the first "ring" until you close up the hole with filler metal. This is often done in filling sheet-metal holes in body work, but any hole larger than a dime should be filled by welding in a circle of new metal of the same thickness.

The beauty of the MIG process is that the torch is a one-hand operation, which allows you to use your other hand to steady your primary hand and draw a smooth bead. Also the torch stays the same distance from the weld at all times, unlike arc welding where the rod is always getting shorter. MIG welding doesn't require stopping to change electrodes either, so you can concentrate nicely on what you are doing. The one-handedness of MIG welding tempts some weldors to operate the torch with one hand while they actually hold something on the work with their other hand. This is OK for tack-welding, where you have a body panel you need to hold in alignment while you tack, but better seams are made when you hold the torch with both hands. Some users get used to the lower brightness and spatter of the MIG process, and are tempted to tack parts together without using a helmet. Body men in particular are prone to hold two panels in alignment, put the torch where it needs to be, close their eyes and make a tack weld for a second. Although MIG welding does produce less intense light and spatter than arc welding, you are still subject to both radiation burns and some spatter. Don't ever take a chance with your eyesight. One blob of spatter onto your bare eyelid could be really dangerous! A much better procedure (and the only one we recommend) would be to use some clamps or other devices (see the Safety and Shop Equipment chap-

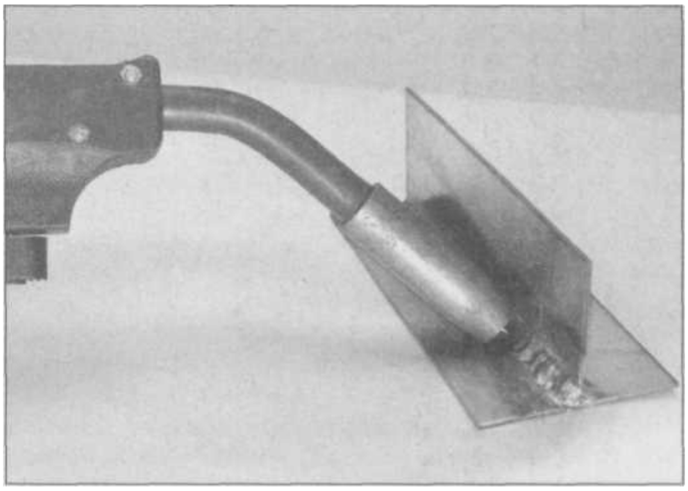

5.54 When welding a corner joint such as a T-section, angle the gun at about a 45-degree angle to aim equal heat and filler on both pieces.

5.55 Don' be tempted by the low spatter of MIG welding to just shut your eyes when tacking body panels. If you need speed and convenience while tacking, use a hand-held safety-shield with lens as shown here. Weldor here has long sleeves, but should have gloves on and a turned-around hat to protect his hair from hot spatter.

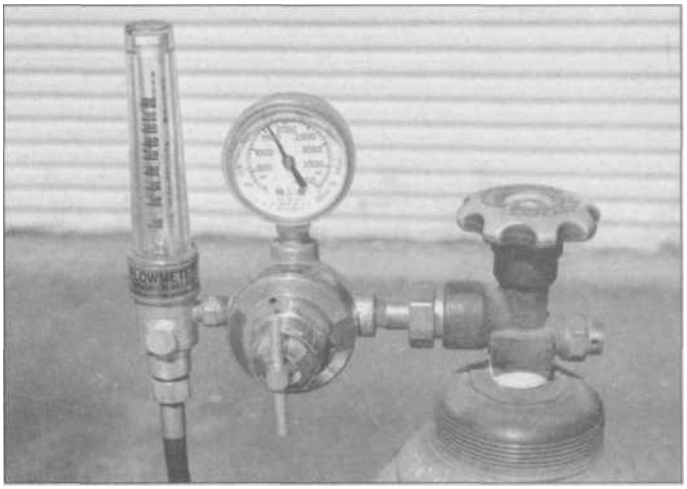

5.56 The best argon flow meters for professional use are the type with a floating ball inside a glass tube. Every time you pull the trigger, the ball floats up on the scale to indicate the gas flow to the machine. At a normal working distance from the machine, the weldor can more easily read this than a dial gauge. It makes it easy to see at a glance if you forgot to turn on the gas at all.

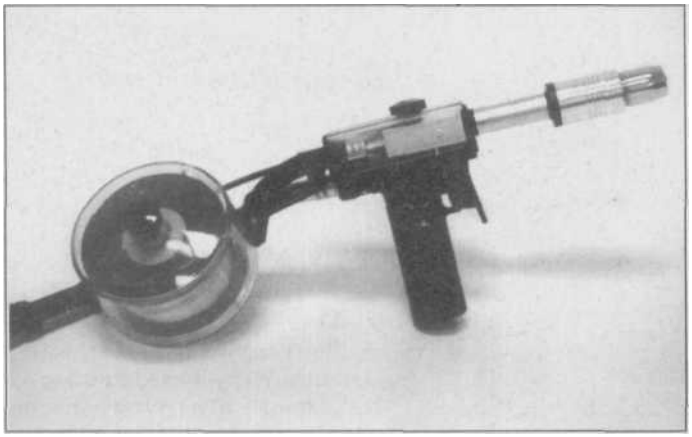

5.57 When MIG-welding aluminum, especially thin aluminum, the aluminum wire itself is fairly flexible, which can cause feeding problems inside the machine, because the drive rollers are *pushing* the soft wire a long distance. This is a spool gun, which holds a small roll of aluminum wire right at the gun for easy feeding.

ter) to hold your panels in alignment, and use a helmet. If you don't like the bother of flipping a big helmet up and down to just make a series of quick tack-welds, then use a hand-held face shield with lens **(see illustration).** It's very easy to hold the shield in your left hand while tacking with your right, and it goes instantly from shielded to an unshielded view of the work.

If you are welding along and all of a sudden the bead goes bad, there are several things to consider. If you have been steady with your arc distance and travel speed, then perhaps a draft temporarily blew the shielding gas away from the weld. If it's not a draft, look to your main torch cable. If there are any bends or kinks in it, in the position in which you hold it when welding, the flow of gas can be restricted easily. Watch your gas bottle gauges while you pull the trigger for a second to see if the flow rate has changed or if the gas bottle is running low. The better equipped MIG machines in most shops have a type of regulator with a flow meter to indicate the flow rate **(see illustration).** It consists of a glass tube with a colored ball inside that floats during gas flow. The numbered line it floats to indicates the flow rate. The advantage of these more expensive regulator/meter setups is that the weldor can see the gas flow from where he is *at the work.* He doesn't have to go back to the machine to read a gauge dial.

Aluminum welding will be more of a challenge than steel, regardless of the welding method used. Aluminum really soaks up heat, and you have to use thicker wire, and more heat, wire and gas flow to weld it. Equipping your machine to weld aluminum means: setting the gas flow almost twice as high as for steel, using aluminum wire, using the right gas mix, replacing the wire liner in the torch with a Teflon liner and using a contact tip that is one wire-size bigger than the aluminum wire. The reason for all this is that the aluminum wire tends to grow more with heat than the steel wire. The wire tension needs to be fairly light, because bird-nesting is more commonplace with the less stiff aluminum wire. In fact, there are special MIG guns made for

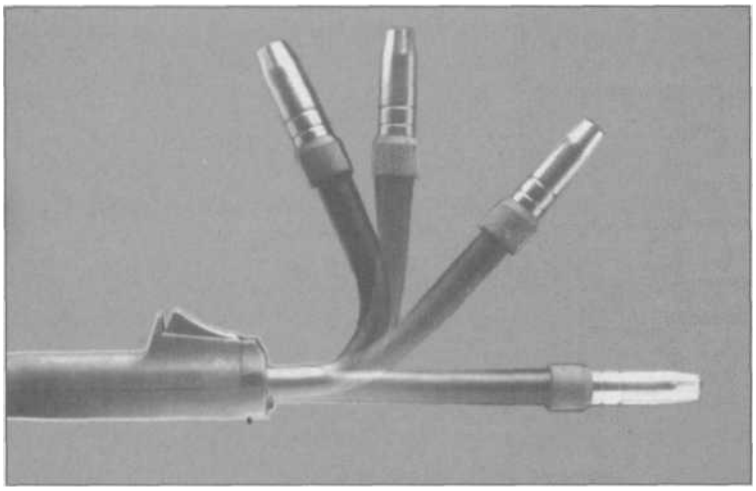

5.58 For tight quarters such as muffler work and other round-tubing situations, HTP makes this MIG gun with a flexible swan neck. It can even be used straight ahead, to reach into a spot where you can't get your hands.

5.59 Muffler shops are finding that the MIG machine is faster and safer to operate under a car than the traditional oxy-acetylene torch equipment. There is no constant re-lighting and adjusting of the flame to worry about when initially tack-welding various pipe sections together.

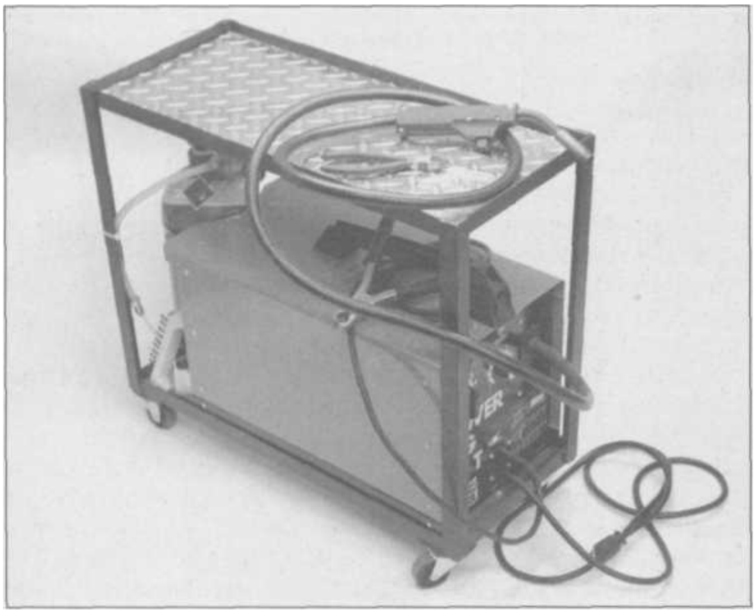

5.60 A handy welding cart for your MIG machine will probably be one of your first home/shop projects. This one is simply-built from one-inch angle-iron, fitted with a steel bottom shelf to hold the welder, bottle and accessories, and the top was made from a piece of polished-aluminum diamond plate purchased from a remnant pile at a scrap yard. Caster wheels were added quickly by welding their mounts to the bottom.

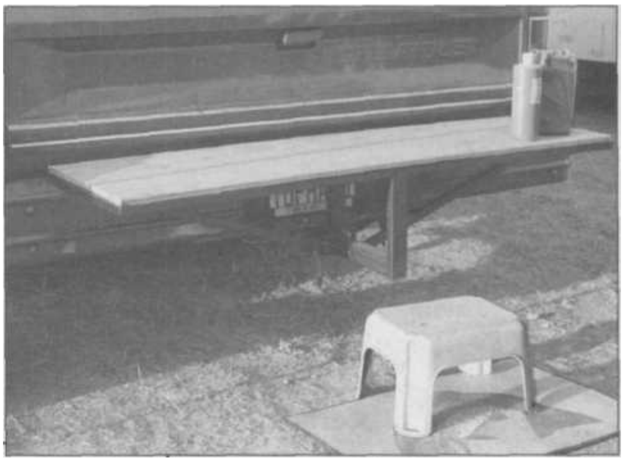

5.61 You'll find many uses for your welder once you get used to using it. This example of a typical home project is a plywood-covered rear bumper extension for a truck. It's used for carrying extra camping gear, and used as a tailgate for working or cooking when camping. Its also a wide step when stepping out of a camper shell on the truck.

production work with thin aluminum wire that hold a spool of wire right at the gun. Called spool guns, they eliminate the wire-feeding problems of trying to push a soft aluminum wire all the way through a cable to your torch **(see illustration).**

The wire speed should be faster than for an equivalent sized steel wire, and the forehand technique is recommend because it puts more preheat ahead of the puddle. There are two extra steps in welding aluminum. The metal must be scrupulously clean and preheated to 350° F to get a good weld. You should keep

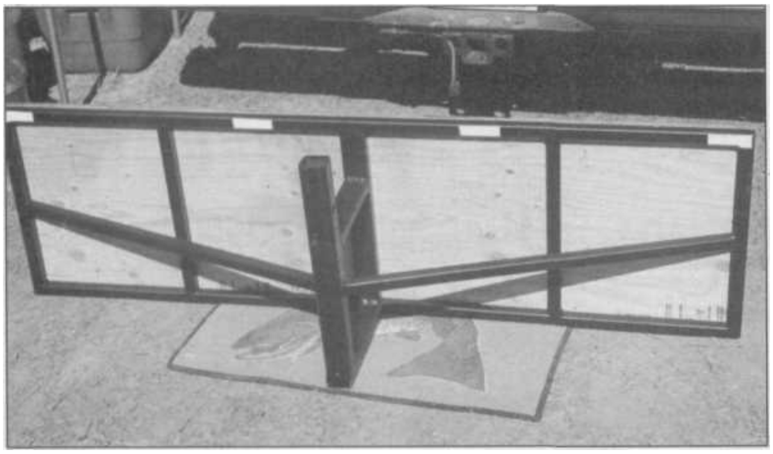

5.62 The camper-step was MIG-welded from one-inch steel tubing and braced to a 2x2-inch tube that fits into the truck's hitch receiver.

5.63 Street rod shops, customizers and restoration shops are using a **MIG** welder more and more, for its convenience and ability to handle delicate thin materials. This 1932 Ford door has had a new skin MIG-welded along the bottom where the original door was rotted away.

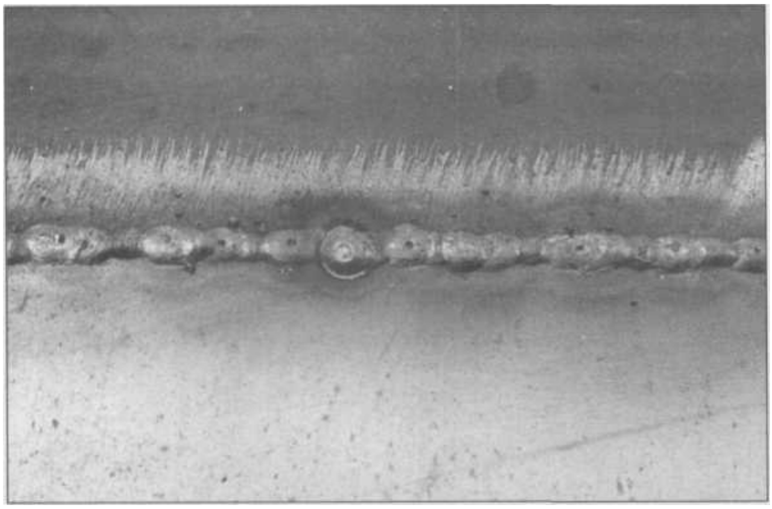

5.64 The door bottom was welded on with a series of very short welds, alternating from one area to another to keep from putting too much heat at one time into an area. Eventually, the whole seam is welded. L-TEC Easy-Grind deoxidized **MIG** wire is perfect for these kinds of jobs, because it grinds off easier, without inducing too much heat in preparing the seam for finishing and paint.

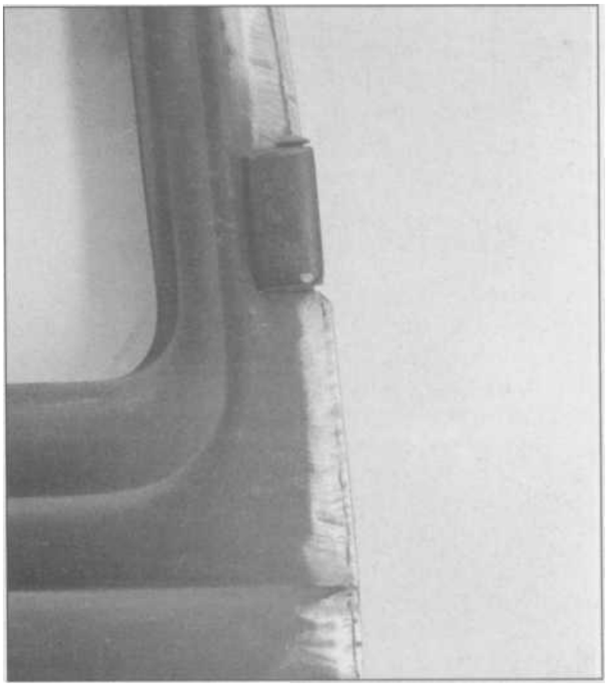

5.65 Customizers and restorers are fond of achieving a perfect gap around body panels on first-class vehicles. Often this involves grinding material off an edge or, as here, adding a sliver of metal along the door edge (at right) to close up a gap. Welding these thin strips is the perfect province of small **MIG** machines.

a stainless-steel wire brush handy that is strictly used for cleaning aluminum. Don't use it for brushing steel or it will later on contaminate your next aluminum weld.

Start welding aluminum with the torch further out than you would for steel, say one inch. Always start an aluminum weld with the end of the wire cut off fresh with snippers. As soon as the arc starts, move the torch closer, down 3/8 to 1/2-

inch from the work. Instead of the torch angle we recommended for steel, hold the gun almost square to the material and inclined only 5 to 15 degrees in the direction of travel.

The puddling will seem much different in aluminum than in steel. The puddle freezes very fast, and the only change you notice is that the molten area is the only place that is shiny. Aluminum has a narrow range of melting before it vaporizes and is highly susceptible to tiny amounts of foreign material. If you see black specks appearing as you go along on a seam, they represent impurities coming to the surface. The short-circuit type of MIG machines we have been discussing are best when the aluminum being welded is relatively thin. Thicker parts are better welded with TIG equipment or the spray-arc-pattern found on larger industrial MIG machines.

There will be a little more maintenance on the nozzle and contact tube when welding aluminum, due to increased spatter. The use of an anti-spatter spray inside the nozzle, or dipping the hot gun tip in nozzle gel, won't produce any less gunk inside the gun, but it will make its removal much easier, as it does with steel welding. If you use an anti-spatter chemical, it can act as a contaminant to the aluminum weld, so, after applying the anti-spatter, briefly use the torch on a piece of scrap aluminum to "burn-in" the chemical so it won't drip off onto your "good" weld.

One of the most common aluminum welding jobs in automotive work is repairing cracked cylinder heads, bellhousings and other large aluminum castings. The cracks should be well ground out to virgin metal and Vee'd before welding. These parts can really soak up heat so they should be thoroughly preheated, either in an oven or with an oxy-acetylene torch fitted with a rosebud tip. The old-timer's method of measuring the preheat was to coat the part with black soot from a overly-rich gas torch. Then with a torch set *properly,* the part was heated until the soot burned off, and that was considered enough preheat. Obviously, too much preheat could require complete re-machining of the part after the welding to correct warpage. Modern weldors use temperature marking sticks to get an exact preheat.

Maintenance on your MIG equipment is simple, but should be performed regularly to keep everything operating smoothly. The most often-maintained items are the gun components. The contact tubes wear out from frequent use as well as become burned such as when you hose kinks and the gas stops during a weld. Sometimes the hot wire welds itself to the contact tip. Often the problem can be fixed by filing on the end of the tip to free the wire's bond, or by pushing the wire to the side with your cutting pliers. If not, you can remove the nozzle, unscrew the contact tip and cut off the wire. Then feed the wire through a new contact tip and screw the tip in place and replace the nozzle. Use anti-spatter agents often, and clean out the nozzle and contact tip before starting a project, during the project, and after the project. When you replace the roll of wire, use a blowgun on your air compressor hose to blow any dust out from inside the cabinet, and blow off the roll of wire before you install it. Keep the drive rollers clean with mineral spirits and lint-free cloth. Often professional weldors will tape piece of lint-free cloth inside their MIG welder such that it drags across the wire as it comes off the roll. The cloth helps to snatch any lint or dirt from the wire before it can collect in your cable or torch.

That's about all there is to MIG maintenance. This is a welding system that has really become the prominent method for the home/shop user in the last decade, and if you have read all of the above, you can see why it has many advantages. Once you start using one, you'll really appreciate it. The lack of paraphernalia, the ease of starting and maintaining the arc, the cleanliness of the welds, reduced shop fire hazards and the ability to weld various materials, including thin sheet metal, have combined to take the MIG process out of the exclusive hands of production weldors and give hobbyists, artists, farmers and restoration technicians a very practical, easy-to-learn welding system.

DEFECT	DEFECT CONDITION	MAIN CAUSE
PORES/PITS	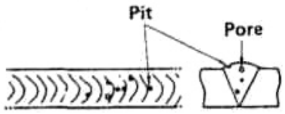	• There is rust or dirt on the base metal. • There is rust or moisture adhering to the wire. • Improper shielding action. (The nozzle is blocked, wind or the gas flow volume is low.) • Weld is cooling off too fast. • Arc length is too long.
UNDERCUT	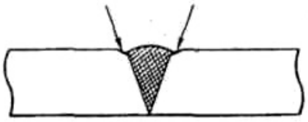	• Arc length is too long. • Gun angle is improper. • Welding speed is too fast.
OVERLAP	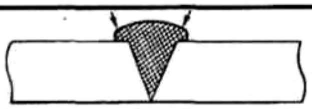	• Welding speed is too slow. • Arc length is too short.
INSUFFICIENT PENETRATION	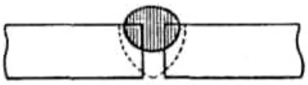	• Welding current is too low. • Arc length is too long. • The end of the wire is not aligned with the butted portion of the panels.
EXCESS WELD SPATTER	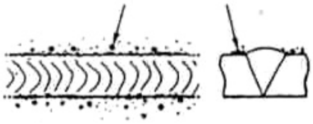	• Arc length is too long. • Rust on the base metal. • Gun angle is too severe.
BEAD NOT UNIFORM	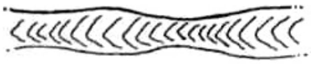	• The contact tip hole is worn or deformed and the wire is oscillating as it comes out of the tip. • The gun is not steady during welding.
BURN THROUGH	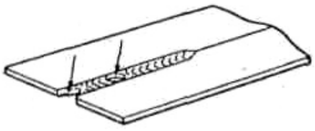	• The welding current is too high. • The gap between the metal is too wide. • The speed of the gun is too slow. • The gun to base metal distance is too short.

5.66 It will take some practice to become competent with any welding equipment, but MIG is the easiest to learn and the most forgiving in most automotive and home/shop applications. Test your practice welds for strength and cut them apart now and again to compare with the illustrations shown in this MIG-welding troubleshooting chart.

6 TIG welding

When those who are interested in welding talk about fusion techniques, the subject of TIG welding is held in a certain reverence. Its reputation in this regard partly stems from its place in history as the method used for construction of a great many famous aircraft. Although developed initially in the 1920s, TIG (Tungsten Inert Gas welding) wasn't used much because the helium shielding gas was too expensive. The intensified research atmosphere of W.W.II spurred further development as aircraft were being made lighter and lighter and TIG became the preeminent method of joining such non-ferrous materials as aluminum and magnesium. The Linde Corporation (now L-TEC) was the first company to capitalize on the technique, and after the war their trademark name "Heli-Arc" became the de facto generic name for TIG welding. Many weldors today still use the term heli-arc more often than TIG as a description of this type of welding, even though a number of other companies have been making TIG equipment for many years.

Besides the romantic beginnings, the TIG process has been considered quite special for other reasons. It does take considerable practice to be good at it, and, because the equipment has always been rather expensive, a weldor who had one and was good with it developed a reputation, particularly in the field of esoteric materials and exotic construction in aircraft and race cars. Basically an outgrowth of

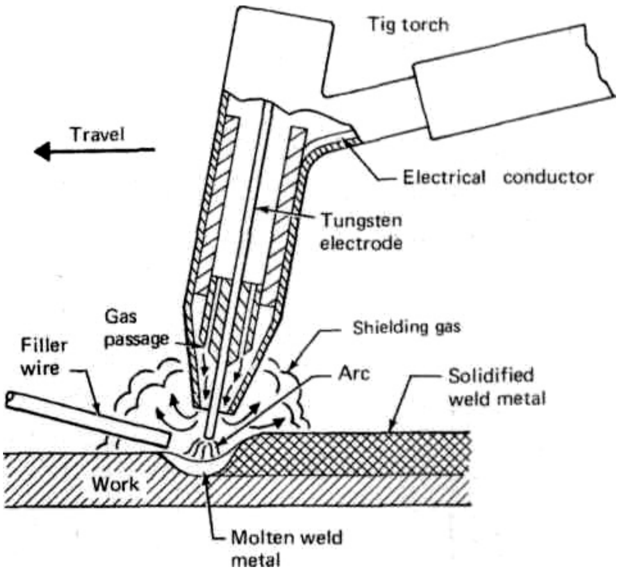

6.1 The **TIG** process is considered the most precise, most controllable and cleanest method of welding. The arc is made between the work and the central tungsten electrode in the torch, while a flow of shielding gas comes in from the nozzle. **Note** the low angle of filler rod addition. TIG or heli-arc **welding** is also one of the most observable welding process, with no spatter and little smoke.

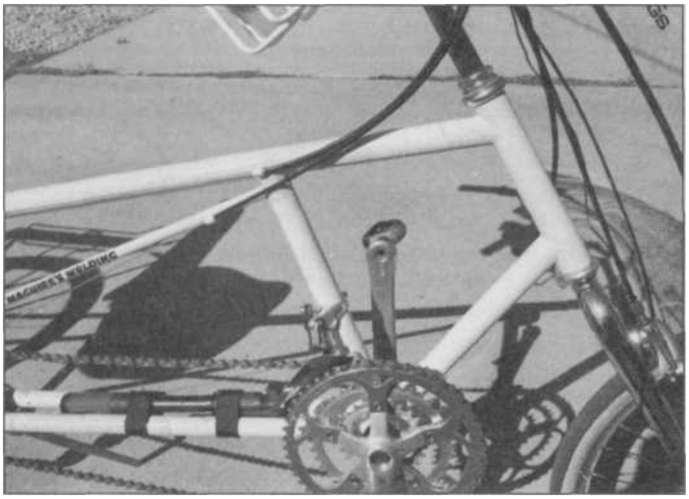

6.2 Where the TIG process is unequaled is on non-ferrous materials or on thin materials. Here a custom bicycle frame was heli-arc-welded from very light, very thin-wall 4130 chrome-moly tubing for strength without a weight penalty. For the same reasons, this kind of material and welding process is use in race cars and aircraft.

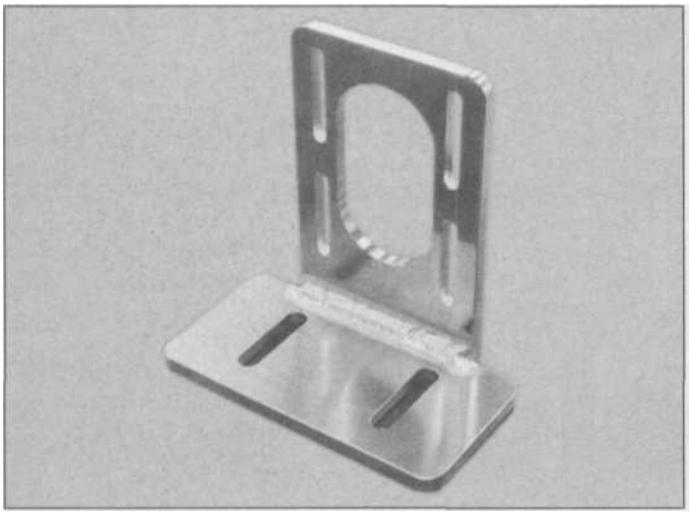

6.3 Where welds on aluminum have to be as pretty as they are strong, TIG is again the top choice. The weld on these two machined and polished parts can easily be wire-brushed to look as good as the polished parts.

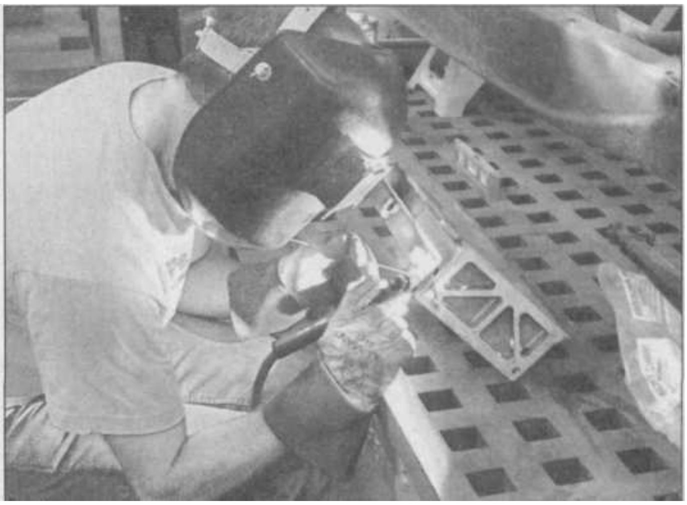

6.4 Unlike arc-welding or MIG-welding, the weldor can get very close to his work with heli-arc equipment. Even when welding on this cast-aluminum dry-sump oil pan, smoke is minimal and the arc not nearly as intense.

arc welding, the TIG process is done with a lightweight torch that uses a tungsten electrode to draw an intense, concentrated arc, shielded by an inert gas. The gas comes out of the torch, all around the electrode and displaces the air from the weld zone to exclude oxygen and nitrogen from contaminating the weld. Although helium was the original gas used, today argon is the most common shielding gas for TIG welding. The tungsten electrode is not considered a "consumable" in the usual sense, since filler metal is supplied by separate, hand-held rods, much like in oxy-acetylene welding. In many ways, heli-arc welding is like a high-tech version of the old-fashioned gas torch.

The concentrated nature of the TIG arc is one of its strong points. Welds of great strength and quality can be made on thin materials, light materials, dissimilar materials, in fact most available metals, and all with minimal distortion or corruption of the adjoining base metal. The process does not generate sparks or spatter, and the welds are as clean as can be, which is why it is the method of choice in aircraft, race-car construction, and stainless and aluminum products made for the medical, dental or food-handling services. There is no slag involved, and, because there is very little smoke generated, the weldor has a good view of the process, allowing him to make very precise joints.

The weldor has greater control with TIG than most other welding methods. The amperage control is infinite because, besides the settings on the power source, the weldor has a foot control he operates as he welds. We recently asked a veteran weldor to make some sample TIG welds to photograph for this book. There was to be one good bead and several bad ones. Even when we set the amperage way too high and way too low, he instinctively tried to produce a good bead, adjusting his torch speed and foot control to compensate. He couldn't make himself do a bad weld, and once the arc was started he tried his best and made fairly good welds even with the machine parameters way off.

With good torch manipulation, a TIG weldor can weld thin materials to thick ones, steel to stainless steel, steel to cast iron, and enough other oddball situations that a good TIG weldor begins to be regarded as some kind of magi-

6.5 Where precision welds are involved, especially in tight quarters such as on this stainless-steel race-car header, TIG welding's wide range of operator control is seen as an advantage.

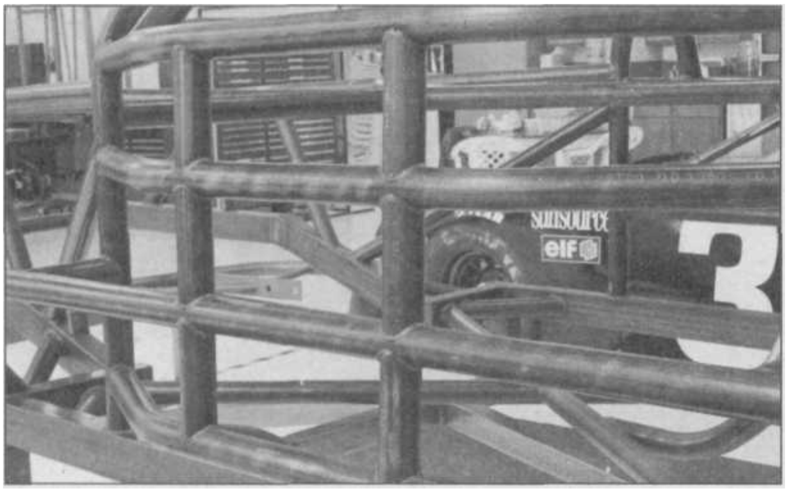

6.6 When thick-wall mild steel tubing is used in a race-car chassis, welding is usually done with MIG equipment, because of its speed. However, this chassis of lightweight, thinwall 4130 chrome-moly tubing was TIG welded. There are a lot of man-hours represented here in precisely-fit joints and welding time.

Miller
Syncrowave 250

6.7 Most shop-type TIG welders are heavy, bulky power sources that require three-phase input power and water connections, making them less portable than other welding machines. This is a Miller Synchrowave 250, a popular unit in many fabrication and race-car shops.

cian. In the time that we were in our friend's fabrication shop, he did a production job heli-arc welding some aluminum brackets, some prototype motorcycle seat frames, repaired a Lamborghini aluminum water-pump casting, welded up a fancy cast-aluminum mailbox post that had been run over, and finally a gunsmith brought in a trigger and barrel from an antique cavalry carbine to be repaired with TIG-welding. He handled them all with equal aplomb, and it was obvious that people from all around brought the tough jobs to him knowing he could handle it.

We've discussed many types of welding systems to this point, and each has been shown to have its advantages and disadvantages. With TIG welding there are only few disadvantages. There are very few jobs a TIG welder can't do, but there are some jobs other techniques are better suited to for practical reasons, not because of a difference in quality or strength of weld. For big jobs, like welding up a car trailer or a big metals rack made from steel tubing, the heli-arc is just too slow - an arc welder or MIG would be much faster. However, there are applications in aircraft and race-car chassis work where a large project may be tediously done with a heli-arc, simply because a lightweight, high-strength material like thinwall 4130 chrome-moly tubing is used where TIG would be the preferred method **(see illustration).** For outdoor work, again an arc welder or flux-core-wire MIG are the best choices because they aren't as affected by air currents.

The cost and complexity of the TIG system has kept it out of the hands of most small shops until the last few years. Most large TIG setups are heavy, some weigh up to 800 pounds and have to be moved with a crane or hydraulic pallet-jack, and most professional units have a water-cooled torch assembly which often involves some plumbing connections in the shop beside the normal electrical hookups **(see illustration).** So the big shop-type TIG machines have always been about the most expensive welding machines available, which has increased their mystique but not widened their market. In the last few years, however, manufacturers have built smaller and lighter, solid-state-electronic TIG machines to make them more attractive to the small fabrication shop, race-car shops and street rod builders. Many of these machines are now air cooled, and, at about one-third the cost of the traditional TIG machines, they are being seen out in the field more and more. As one street rod builder told us, "There isn't a thing on a car that would require me to have a 600-amp welder, and when you design and build a car two brackets at a time, the torch never gets that hot."

That is the main difference between the usage of the large, industrial TIG

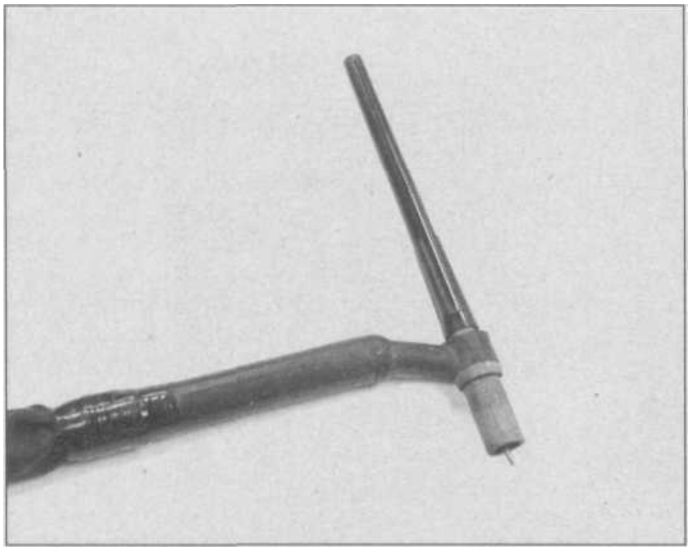

6.8 The basic TIG torch is very light, easy to hold comfortably, and quite durable. The only "consumable" is the filler rods, if used, and to some extent the tungstens.

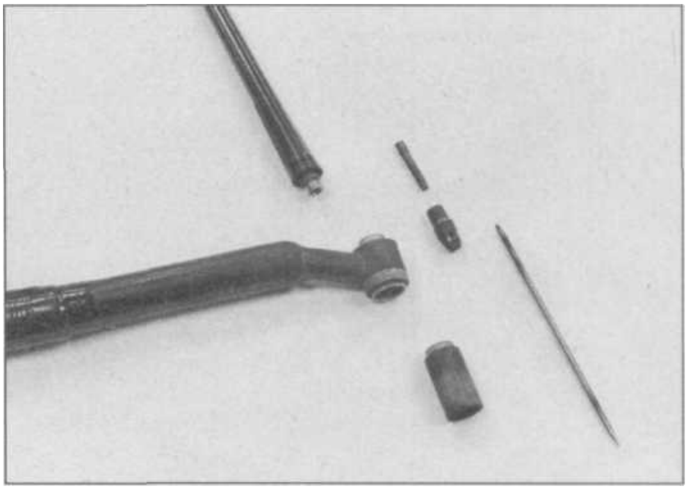

6.9 The exploded view of the typical TIG torch shows the tungsten (right) and from top to bottom: the back cap, collet, collet-chuck, torch body, and ceramic cup or nozzle. To assemble, the collet-chuck is screwed into the torch body on the bottom, the nozzle is screwed in on the bottom, the collet slides in from the top, and then the tungsten is slid in and retained by the back cap.

welding machines and the "economy" TIGs on the market today. The big boys can weld all day long at any amperage they need and never overheat the torch or the electronics, the air-cooled smaller TIGs are fine for anything but long seams or welding big aluminum castings.

The equipment

The basic components of the TIG setup include the power supply, foot control, water circulation system and the torch. The latter differs from the other types of electric-welding torches by being smaller and lighter and holding a central rod of tungsten as the electrode for the arc **(see illustration).** Taking the torch apart, there is a ceramic shield at the business end (called the ceramic, the gas cup or the nozzle), then the collet, which retains the tungsten in the collet body, and the back cap **(see illustration).** There is customization of the torch, depending on the work to be done, in that there are different-sized tungstens, different gas cups with shapes for specific kinds of jobs and different back cap lengths depending on how tight the quarters are where the weld has to be made. For instance, you could use a short tungsten and short back cap to fit into a tight corner better. The shape of the gas cup can create a different shape and size envelope of shielding gas over the weld area. From the larger welding-machine companies, there's a variety of different-size torch bodies with different amperage capacities as well as some special models made for tight confines, including one in which the tungsten and ceramic nozzle come straight out of the handle rather than at an angle like most TIG or MIG torches.

When it comes to the basic power supply, you will not find as great a range of choices as with say, MIG welders. The marketplace for TIG equipment is still pretty much a professional arena, and the newer, less-expensive air-cooled TIGs are at one end, and the big, shop units are at the other, with few choices in the middle. A TIG setup is still not a home/shop type welding process, though there are certainly some small specialty fabrication businesses run out of a home garage that could utilize a TIG machine, particularly if the work involved aluminum, stainless steel or other non-ferrous parts, and high-quality, squeaky-clean welds were important to the products' strength or appearance. If you had a

POCKET PULSE
DAYTONA MIG
CEBORA

6.10 There are only a few small, more-portable TIG units, like this 100-amp "pulse-TIG" from Daytona MIG. It is a 220V, DC-only machine, but is said to be suitable for steel, stainless-steel and chrome-moly up to 3/16-inch thick.

DAYTONA MIG
CEBORA
TIG STAR 16-P
A.C.
D.C.

6.11 Mid-size TIG machines are in the 160-250 amp range, though this fairly-portable Daytona MIG unit is 130-amp, 220V. Mid-range machines offer options like slope control and post-flow. This one has square-wave technology, and AC/DC operation.

small business making stainless-steel rope clips or other items for rock climbers, for instance, you might weigh the cost of an air-cooled TIG versus sending all your parts out to be welded in a shop. If you had the volume of work to justify the one-time costs and the learning process, it could be a good move to do your own. However, it would be a wise move to take some TIG classes and get yourself certified first.

Input voltage on TIG machines begins in the 220V range of single-phase power, and the bigger professional boxes are for 460V three-phase. The professional units range in amperage capacity from 250 amps to 300, 400, 500, and 650, though the larger ones are multi-technique machines on which the high amperages are probably used only on the arc-welding process. It would be pretty rare to use more than 300 amps on a TIG weld, unless you had to weld really thick aluminum plates or castings. When welding at 300 amps or more, tiny bits of the tungsten electrode can come off and become included in the weld, which would weaken the joint. The three-phase current models are definitely for shop use only, where the amount of shop equipment and power used every month makes the expense of transformers and rewiring the building economical over the long run in utility bills. Home shops generally do not have three-phase power.

On the small-shop scale, there are very few TIG-capable units in the 115V range. One we have seen is the Hobart Ultra-Weld 130, which is an arc-welder with TIG capability, but with a limited range for either. In its 115V version, it offers 75 amp capacity with a 40% duty-cycle, and the 220V model offers 130 amps at 60% duty-cycle. It does not include the TIG torch, and its TIG range of material thicknesses would be limited. We expect that as the home TIG market expands, there may be some interesting, smaller 115V and 220V units to come. A few imported TIG units are built with DC-only capability to keep the price down to a starter level, and they are fine for steel (only) up to 3/16-inch **(see illustration).**

Of the medium-range TIG units, there are several of the domestic and imported brands offered in the 160-250-amp category that can handle virtually anything in the small shop environment **(see illustration).** Most of these are 220V or more in input current requirement, but most are single-phase, which means you could easily wire your shop to accommodate them. Be advised though, that even the 160 amp units can draw 36 amps of input power at their highest settings, and the 250-amp machines can draw 60 amps or better on full load. This means that your household 30-amp 220V circuit would blow at any of the higher settings. You would really need a separate circuit and box for the welder, with a 50-amp breaker for the smaller mid-range machines and 100-amp breaker for the bigger ones.

The mid-range machines offer several special features, such as *slope control* **(see illustration).** This refers to how the wave of current cycles when the machine is welding. Picture the wave as a curved line that goes up, down, crosses the baseline and then curves back up and over again in repeated fashion many times a second. This is called a sinusoidal wave in alternating current. TIG machines use both AC and DC current, with the AC used for non-ferrous metals like aluminum or magnesium. One half of the AC wave is actually straight polarity, and the other half is reverse polarity. The straight-polarity portion of the wave builds the heat in the arc, while the reverse current has the effect of cleaning the oxides from the aluminum being welded. So the TIG torch on AC is constantly cycling between a hot arc and a cleaning cycle, but at lower amperages the arc can be obstructed at the changeover point or actually go out. Most of the TIG machines have a special circuit inside that produces a high-frequency, low-power extra current that keeps the process going at all times. The high-frequency control makes the TIG arc easy to start without touching the tungsten to the work, and makes a more stable arc while welding. On many machines the high-frequency system can be set on the front panel of the welder to arc starting, off, or continuous use. On ferrous metal welding, the high frequency is used mainly for arc starting.

The slope control is a feature in which the weldor can adjust the parameters of the wave on the up side or the down side to get the perfect combination of penetration and cleaning action on aluminum. Even a clean piece of aluminum contains some aluminum oxides in the base material, and the slope adjustments allow a good weld to be made without these oxides being included in the bead.

Most of today's TIG machines offer *square-wave* AC. The traditional sinusoidal wave with high-frequency over it creates a lot of radio interference as well as some problems on welding aluminum **(see illustration).** The new machines are all solid-state electronics, with features that make a wave made of up and down squares, rather than half-circles. This means the high-frequency control intervenes less often, and the same-sized electrode can be used at higher currents. This discussion of the current waves may be more than the home/shop user needs to know, but at least you may better understand the language used by manufacturers to describe their equipment. Suffice it to say that the modern square-wave technology and solid-state electronics are desirable features in a full-on TIG machine.

Another feature you may see mentioned is post-flow. When you are making a bead and get to the end of your seam it's a good idea to allow gas to flow over the end of the weld as it cools, to keep out contaminants. On most professional machines, there is an adjustable timer that allows gas to flow for so many seconds after you have stepped off the foot control to quit the arc. You keep the torch in place just where you stopped, and gas keeps flowing for the specified time. This is

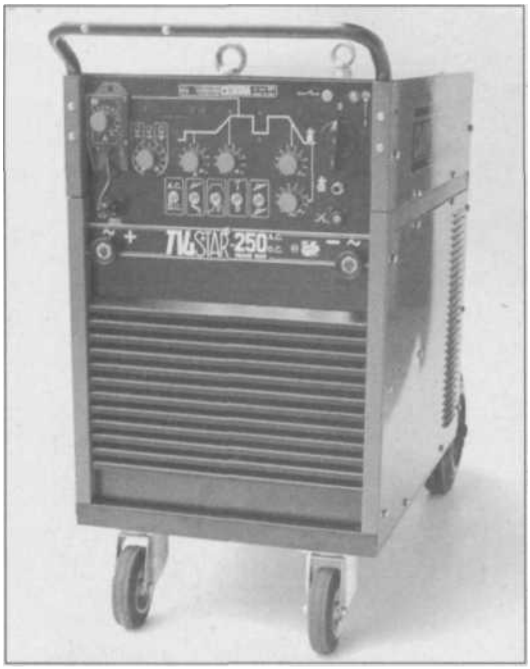

6.12 When you go up in welding machine size, you get more amperage and higher duty cycles. This TIG-Star 250 operates on several single-phase input currents, and has a 60% duty-cycle at 200 amps, with 100% duty-cycle at 140 amps.

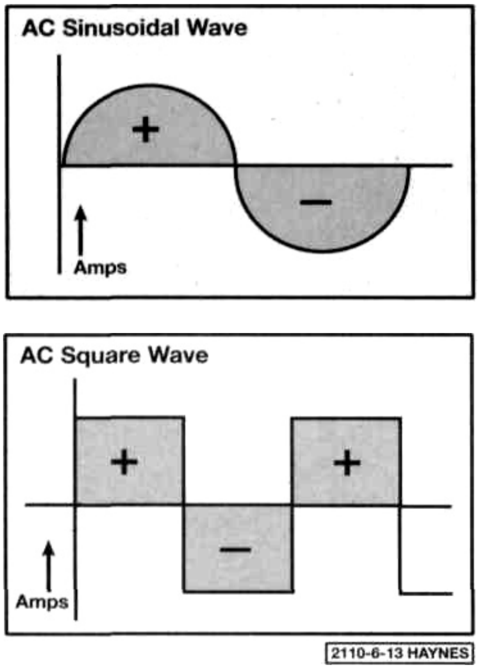

6.13 As a standard "sinusoidal" wave of AC current flows, it is constantly reversing its polarity and its output is always fluctuating as the wave flows. With the new square-wave technology, electrodes can be used at higher currents and have good arc staring and stability, but without the radio interference of traditional high-frequency superimposition. The square-wave design inverts the current much faster, for better welds on non-ferrous materials.

6.14 **Less expensive, air-cooled-torch TIG machines have been available now for some time and are making inroads into shops where TIG was previously considered too expensive. This Miller "Econo-TIG" is being used in a street-rod shop on a home-built cart to make it as portable as any MIG machine.**

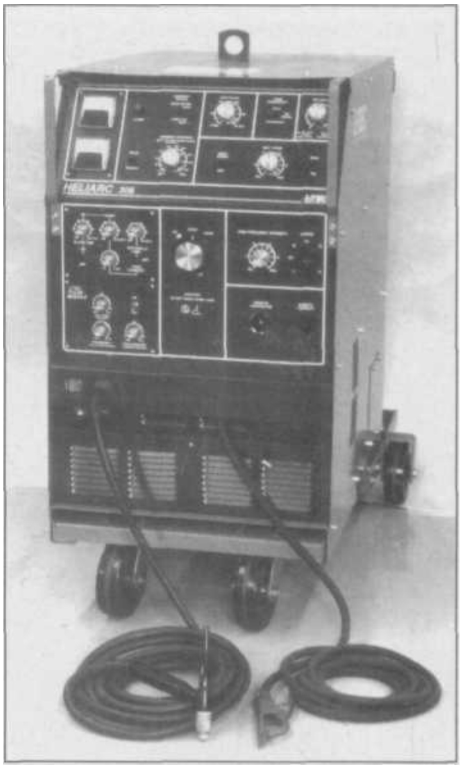

6.15 **Today's state-of-the-art in TIG machines is represented by this 300-amp Heli-arc 306 by L-TEC (the new name for Linde Corporation, the original trademark-holders of the name heli-arc). All solid-state with square-wave technology, it provides high performance and high reliability. The bells and whistles include: slope control, spot-weld control, wave balance control, post-flow, and an "arc force" fine-tuning adjustment.**

helpful on steel, but is essential on aluminum, magnesium and brittle, high-strength steels like the 4130 used in aircraft and race-car construction. It's called post-flow because it is post-welding gas flow. The post-flow of water and gas to the TIG torch also helps to cool the tungsten electrode down to the point where it can't oxidize when regular shop air flows around it after the welding stops.

Most shop-type TIG machines feature a water-cooled torch, but some are air cooled, with the water-cooling system as an option **(see illustration).** Obviously, the extra equipment for a water-cooled torch adds to the total price of the package. In some shops, the water-cooling system consist of a unit added on to the basic power supply, containing 1-5 gallons of water, a water-circulating pump and a small radiator and fan. It's like a miniature automotive cooling system. In shops where the amperages used can be quite high, or where long welds are made, the torch may be hooked to a source of regular cold tap water, and the heated return water flows back into the shop's drain. This system uses more water, but keeps the torch cooler than recirculated water systems **(see illustration).**

The shielding-gas system for TIG welding isn't much different than for MIG welding, as discussed in the last chapter, except that you will use helium, argon or a mixture of the two in your bottle. Most TIG setups use as large a bottle as possible because portability isn't usually a concern in TIG welding **(see illustration).** Shops that do TIG welding usually have the machine set up right near their welding table and don't move them around, although the newer air-cooled units can be wheeled around a shop without too much trouble.

Unlike oxy-acetylene bottles or most MIG-welding gas bottles that use two-gauge regulators, TIG setups have a single round-faced gauge on their tank and a flow meter instead of the other gauge for the line flow. The flow meter has a vertical glass tube with numbered increments marked on it and a colored ball inside. When the gas is

6.16 **Argon bottles are available in a number of sizes, depending on whether you are trying to make your TIG setup portable or not, and how much welding time you need. Shown here, from left, the sizes are: 20, 55, 80, 125, 154 and 248 cubic foot.**

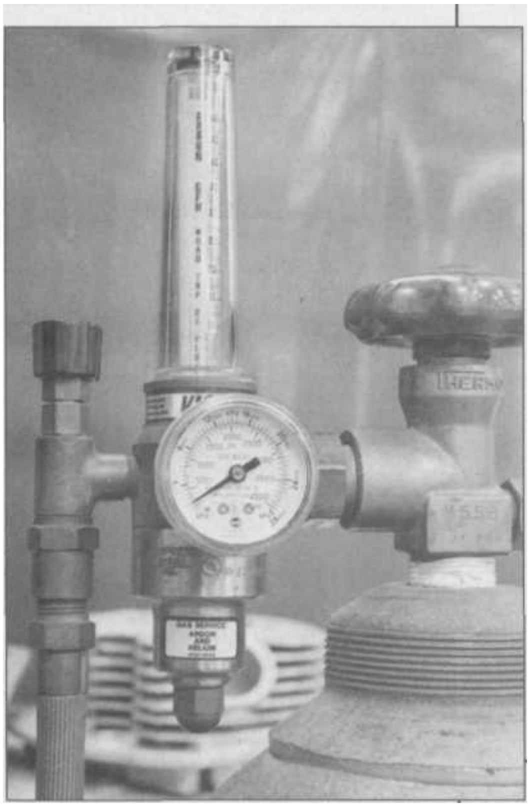

6.17 Flow-meters for argon bottles in TIG use are generally used in place of twin-gauge setups. Inside the glass tube on top is a colored ball that floats along a measured scale to indicate gas flow, but without requiring the weldor to step away from his work.

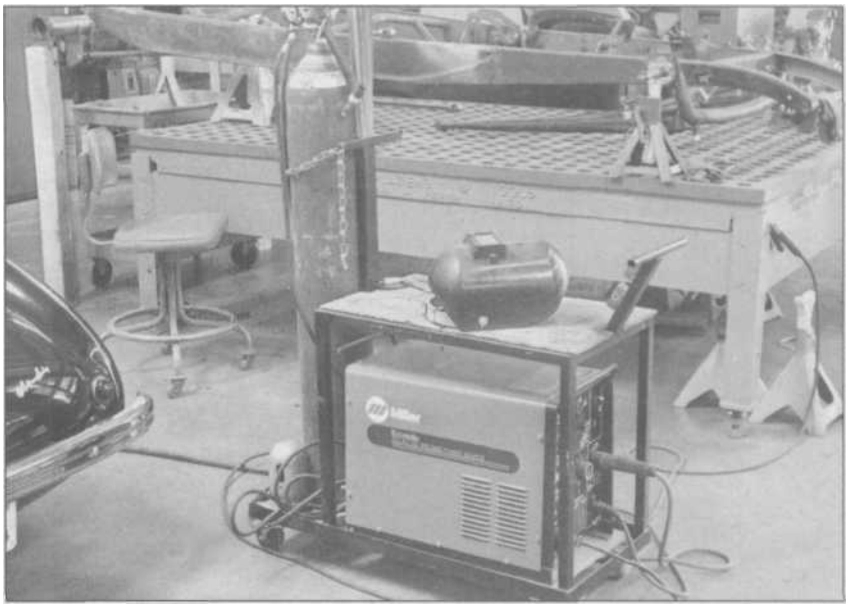

6.18 Argon is versatile enough to be used as the shielding gas for almost all TIG work, simplifying the storage of bottles. Often one large bottle mounted on a welding cart is all that's needed in most smaller shops.

turned on and you click your foot-control on, the ball will float up inside the tube, indicating the level of gas flow in cubic feet of gas per hour **(see illustration).** Professional weldors like the flow meter better than a gauge because that colored ball can be seen floating in the tube from as far away as their welding table - they don't have to get up and look at the face of a gauge on the tank. They can just step on the pedal to start gas flow and see if the flow meter is set right. They can also tell if they have forgotten to turn on the gas bottle before starting.

The correct gas flow is very important in maintaining the integrity of the weld by keeping out contaminants. For most basic steels, a flow rate of 8-10 cubic feet per hour with argon is sufficient up to 1/8-inch material, but the gas flow requirements go up with the amperage used and the thickness of the metal being welded. Stainless-steel requires only a little more gas flow than for plain steel, but if you were welding aluminum, the gas flow recommendation would go up to 15-20 CFH for the same thicknesses, and a larger-diameter tungsten may be required.

Argon is certainly the basic, all-around shielding gas for the TIG process today, and is useable on virtually all metals **(see illustration).** Straight helium is seldom used, but often in production work a mixture of argon and helium is used to get the best properties of both gasses. Argon has the qualities of good arc-starting, cleaning action and best puddle control, while helium makes for a hotter weld for greater penetration and a faster welding speed, such as when welding thicker materials. An argon-helium mix is often used to allow welding of non-ferrous metals at higher speeds. The helium in the mix helps make more heat for aluminum welding in a way that one weldor described as like "adding on a turbocharger."

The process in action

Basically, TIG welding is enough like oxy-acetylene welding that if you have already learned well how to weld with a gas torch, the learning curve for TIG welding will be much shorter than for most other weldors who have never learned

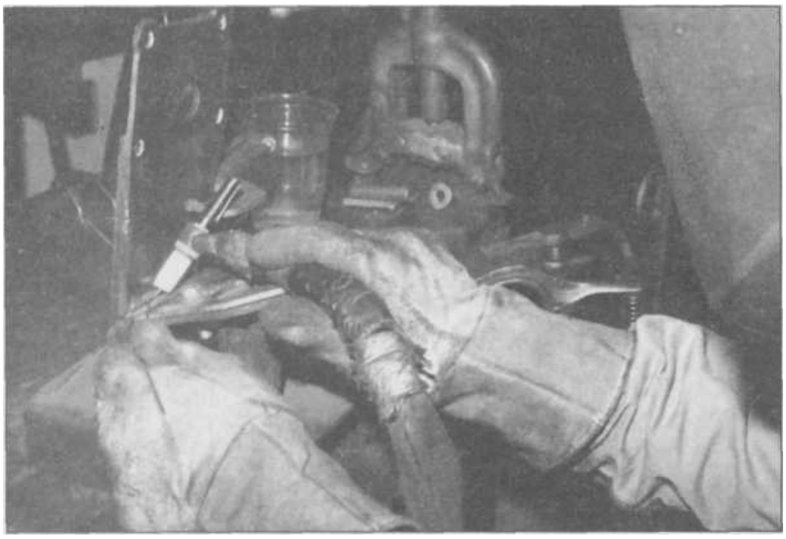

6.19 **Finding a comfortable position for your torch hand is more than half the battle in achieving steadiness with TIG welding. Practice making a bead on your proposed setup *without* turning on the arc, just to find a comfortable position. Note the low angle and position of the filler rod in the weldor's left hand.**

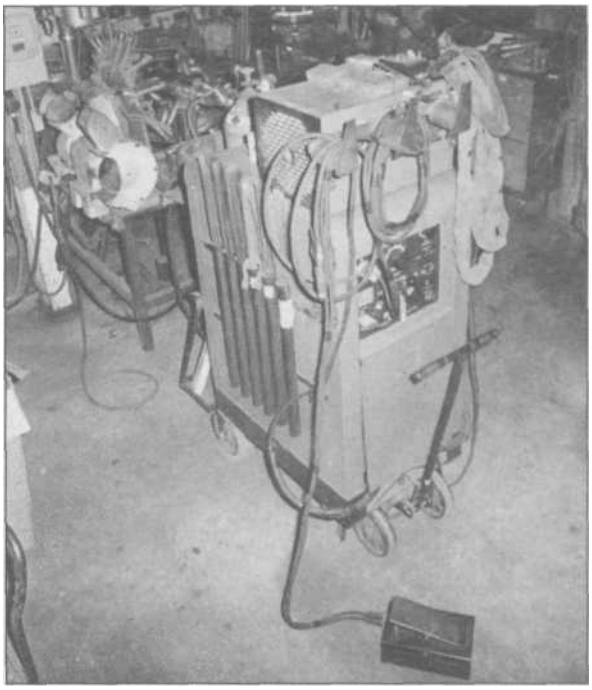

6.20 This is a TIG welder of some years back. Today's current crop of TIGs are half this size. The water-cooling unit here is mounted on top, and at left is a rack of metal tubing holding various welding rods. The foot-pedal amperage control is in front.

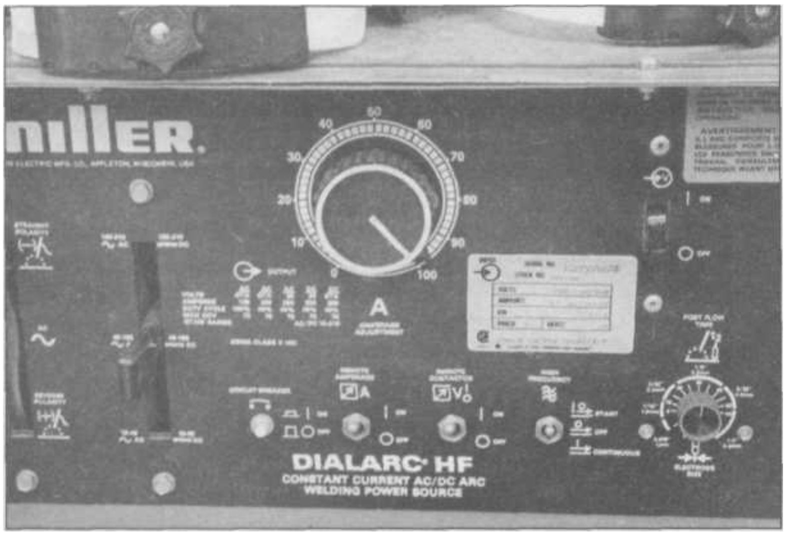

6.21 Layout of this front panel is typical of most TIG power sources, with controls for coarse and fine amperage, current type, high-frequency control and post-timing.

with a torch. It can be as challenging as it is rewarding. Like gas welding, the operator must use both hands, one for the torch and one for the filler rod, each of which must be moved along the weld seam at the same, steady rate **(see illustration).** In addition, the TIG weldor has to operate a foot-pedal-controlled or hand-controlled amperage device (usually clamped directly to the torch body) that is remote from the welding machine **(see illustration).** Using the remote control is one of the operator privileges of the TIG process, allowing the weldor to start an arc, infinitely vary the welding current *as he is welding,* and end the arc without moving the torch. Being able to vary the amperage while making a weld is what makes TIG welding so versatile in handling different thicknesses of metals and the joining of dissimilar metals. Just the right amount of heat can be aimed at each side of a joint, and this also helps make TIG a technique that can work for all positions of welding, too.

TIG machines produce either AC output with high-frequency over it (HFAC) or DC with the high-frequency for arc starting only **(see illustration).** Most TIG machines have a control on the front panel that switches between the types of

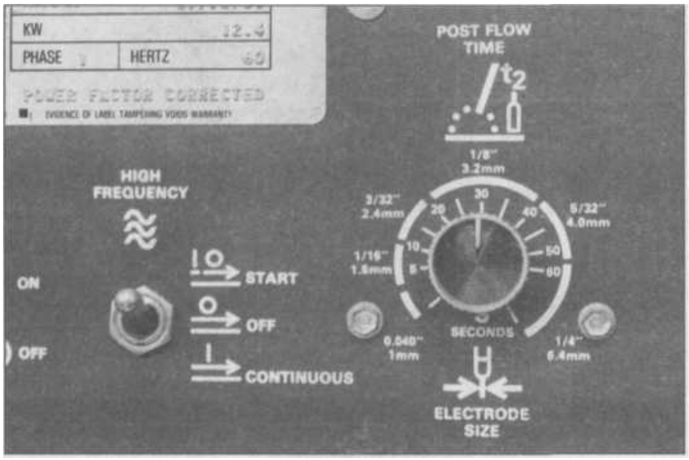

6.22 The three-position switch at left is for the high-frequency, and at right is the post-flow timing as related to various sizes of tungstens. The bigger the tungsten, the more post-flow timing to cool it off.

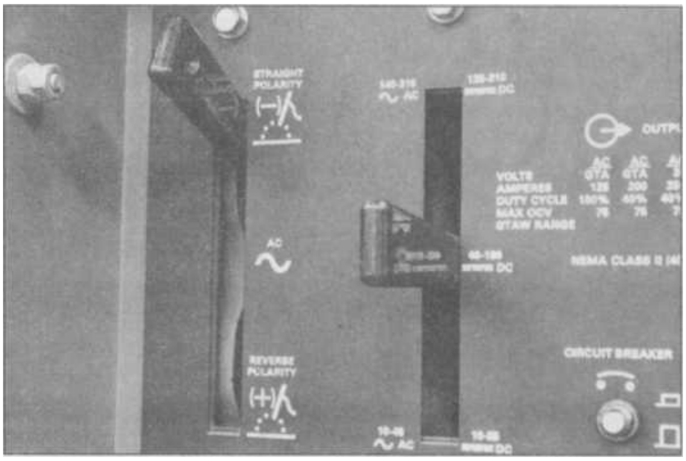

6.23 The lever at left on this Miller TIG sets the polarity, while the right lever is the coarse amperage, with a bottom setting of 10-55 amps, a medium range of 40-190, and a top range of 125-310 amps. Within each range, the fine-amperage knob sets the upper limit for the pedal.

current, a control to select the basic amp range and controls for high-frequency and post-flow **(see illustration).**

The AC is used on non-ferrous metals and usually has the high frequency on during welding, except for the newer square-wave machines that don't require high frequency all the time. The DC current can be either straight polarity or reverse polarity. The straight polarity (DCSP) is used for steel, stainless-steel, silver, silicon copper, brass alloys, cast iron, deoxidized copper and titanium, with the AC mainly for aluminum and magnesium.

Set your Machine to DCSP (sometimes marked DC-), the high-frequency to the start-only position, and set the amperage for the material you are working on. Experienced TIG weldors usually set the three-position amperage lever as a coarse adjustment, then dial in the fine amperage control with a knob (see **illustration).** Often weldors will set the fine control all the way up, using their pedal as the ultimate control. That way, if during the welding the weldor needs a little more heat, he just puts the pedal down further. Whatever you set as the limit on the fine-amperage knob, that's as much as you can get from the pedal **(see illustration).**

There should be a chart with your machine that tells you the correct size tungsten electrode to use with various amperages and material thicknesses. You will probably use a 1/16-inch, 3/32-inch or 1/8-inch for almost everything you do. There are two basic kinds of tungsten materials, pure tungsten or thoriated tungsten. The pure tungsten works under all conditions and metals, but is *required* for AC TIG, while the thoriated tungstens for DCSP welding are used often because they make slightly better starts and penetration.

Tungstens are somewhat expensive, but can be made to last a long time **(see illustration).** The sharpness of the tip is important to good arc starting and penetration on steels, so TIG weldors keep a small supply of sharpened electrodes handy to slip into the torch **(see illustration).** Most weldors will sharpen both ends of a tungsten, so that if they need a new point right away they just loosen the back cap on the torch, slide the tungsten out and reinsert it with the new point down. Good weldors can do this in a

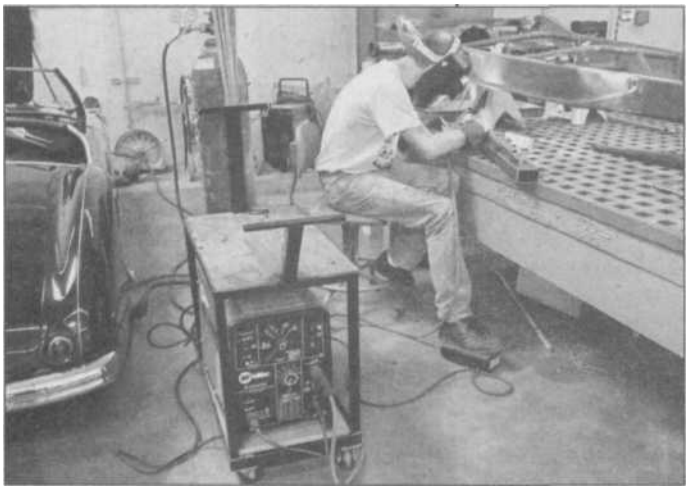

6.24 The pedal is the final weldor control of the amperage in TIG welding. Until you are experienced, you may want to set the amperage knob at the highest you need for a particular job, and know that when you are welding, the pedal can go all the way. It takes some experience to know *how much* pedal to give, if the knob is set to the max for the amp range on the coarse adjustment.

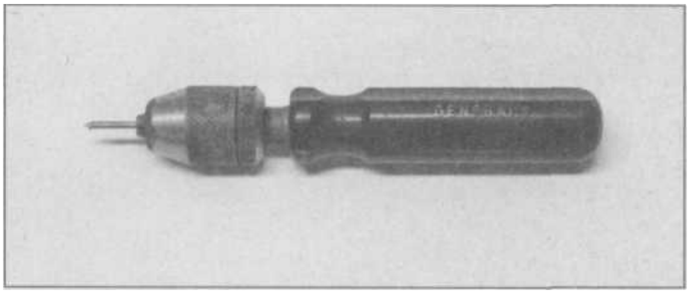

6.25 Make the most of expensive tungsten electrodes by using them until they are too short to fit the torch'. Shorter tungstens can be safely sharpened without burning your fingers, using a tungsten-holder like this.

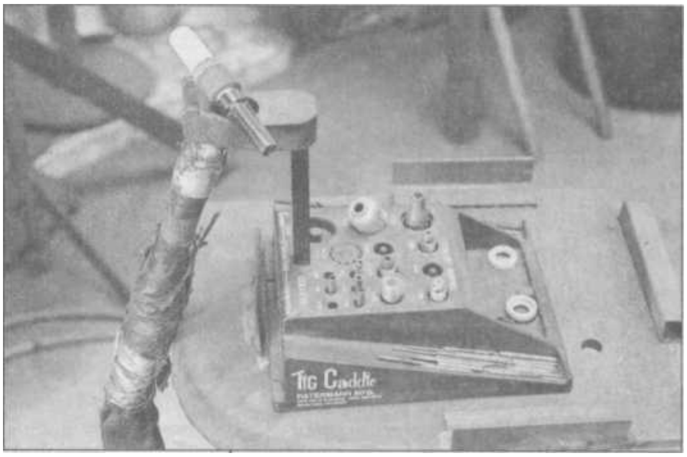

6.26 Spare, sharpened tungstens should be kept handy. This is a TIG caddy, which holds extra tungstens as well as spares for all your torch parts and different size collets and nozzles. The caddy is magnetic, and features a convenient torch rest.

blink of an eye. A belt sander or grinder is usually used for sharpening tungstens, but it should be a dedicated sharpening source **(see illustration).** A grinder used for smoothing steel or aluminum will have deposits on it that can contaminate the tungsten. Sharpen the tip (like sharpening a pencil) only by grinding *lengthwise* on the electrode, grinding from the tip back to the main body. Grinding it held sideways to the grinder will leave a pattern of marks on the tip that may contribute to arc scattering. Electrode tips for use on aluminum with HFAC should be sharpened to be clean but with a much duller point than used for steel **(see illustration).** Theoretically, you should have separate ceramic cups for aluminum work too, but many weldors do use the same cup for both steel and aluminum TIG welding. When tightening the electrode in the torch, push the whole torch straight down against a solid surface (with the welder off) to see if the tungsten will slide in. If it slides, the tungsten wasn't tightened enough **(see illustration).**

At some point, your tungsten tip *will* contact the puddle, this isn't desirable, but it happens to everyone. Pull the pedal back all the way to break the arc, and keep the tungsten in the gas flow for a few seconds until it cools off a little. You cannot continue to weld with this tungsten after it has touched either the filler rod or the puddle because it will

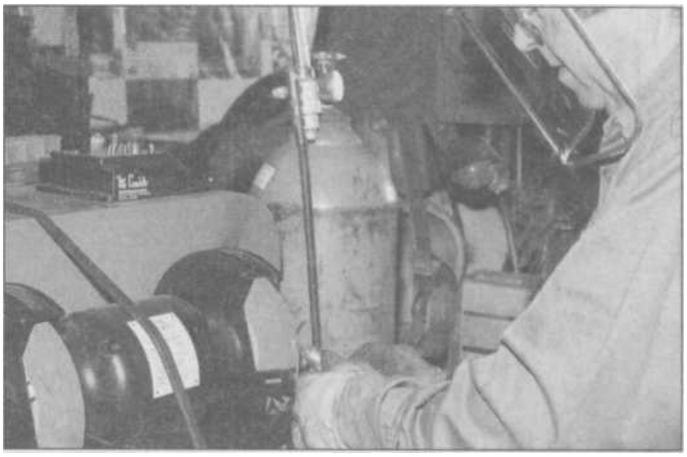

6.27 It's important to sharpen tungstens lengthwise on the tip, so as not to leave any radial grinding marks that could cause arc scatter. Grind points for steel like a pencil, like a crayon for welding aluminum.

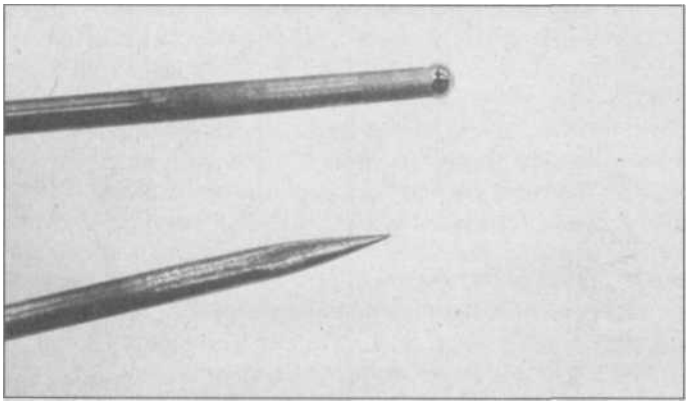

6.28 The tungsten at bottom has been sharpened for use on steel, while the top one is for aluminum, with a small ball-shape on the end. In theory, you could create a ball-end like this by striking an arc on a piece of copper before using the tungsten on aluminum, but a suitably-dull point will turn into a ball like this with a few seconds use on HFAC for aluminum.

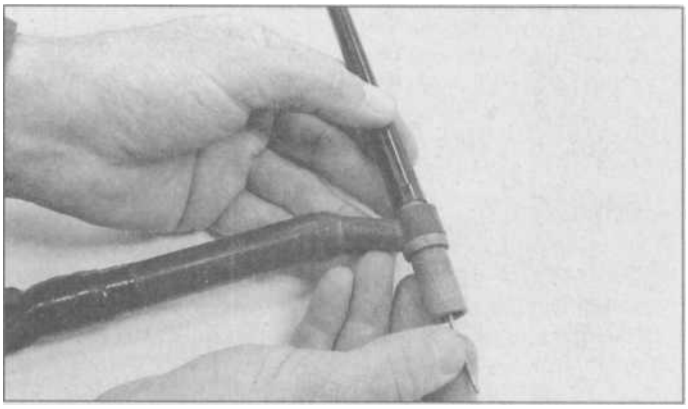

6.29 The sharpened tungsten is inserted into the torch body, and when the tip is sticking out the right distance, you tighten the back cap to hold it in place.

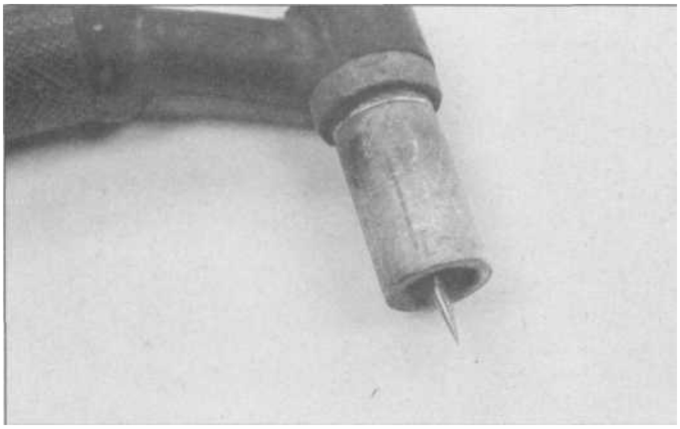

6.30 Although you can crank up the gas flow and extend the tungsten tip out further for reaching into a tight corner, the normal extension of the tip should be 1/8-inch to 3/16-inch from the end of the nozzle.

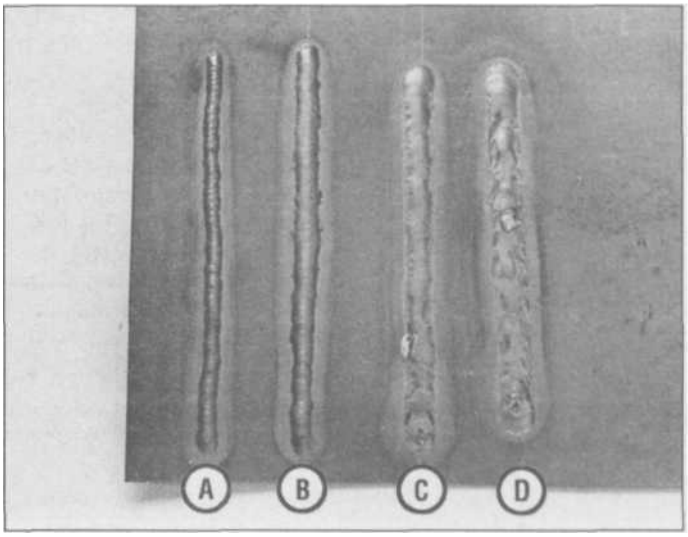

6.31 You'll make a lot of practice beads while learning to TIG weld. Here at left we have: A, a bead with too low an amperage, the bead sits up proud, B, the current control seems about right., C, the amperage is set a little too hot even though the bead seems OK on top, and D. way too hot, with the bead too wide and sunken.

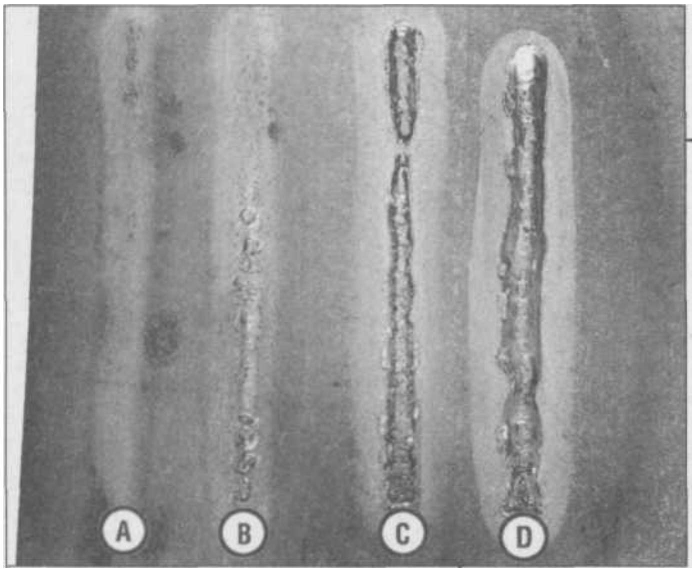

6.32 This is the backside of the same plate as the previous welds. Note how A has only a slight discoloration, B is close to being right, with a tiny bit of dropthrough, and C is too hot, with lots of dropthrough, while D is so hot there's more bead underneath than on top.

be contaminated, and the arc won't be clean and steady. The blackened and pitted section of the tungsten will have to be broken off with a pair of pliers, and the tip will need to be re-sharpened. Although tungstens are not considered consumable in the normal sense of the welding term, they do get shorter every time you break off a piece and every time you re-sharpen them, so eventually they become too short to use. When welding in tight corners, a short back cap can be used with a stubby tungsten, but otherwise use a long cap to protect you from contact with either a sharpened tip (when you have sharpened both ends) sticking out or a hot tungsten. The long back cap covers even a full-length new tungsten.

Turn on your gas and set the flow to 10 CFH, attach the work-lead to a sample plate and run some practice beads. The tungsten should be held in the torch such that 1/8 to 3/16-inch of tungsten extends out past the ceramic cup **(see illustration).** At no time should the tungsten touch the work. When you bring the electrode to within about 1/8-inch of the work and hit the foot pedal, the arc should start and there will be a red-orange glow at the tip of the tungsten. As you move the tip around in small circle in one spot with the torch perpendicular to the work, bring the pedal control down for more amperage. When the puddle is a bright fluid, move the torch along your seam slowly and evenly. The torch should now be tilted back to a 75-degree angle for the rest of the bead. To stop the process for any reason, just back off the foot control all the way to break the arc, but keep the torch over the same spot to cool the weld and tungsten with gas flow.

When adding filler rod as in most joints of any thickness other than thin sheet metal, move the electrode to the rear of the puddle and dip the filler rod to the forward edge of the puddle. Then bring the torch to the front again and make a new puddle, move back, add filler, move again. This may sound like too much work, but really it is very similar to gas-torch welding and the movements become instinctive after enough practice. The filler rod should not touch the electrode, and it should be kept near the forward edge of the seam to keep it hot between "dips." When it is brought to the molten leading edge of the puddle, a drop of molten steel will come off the rod into the puddle and be blended in when the torch advances over it. The filler rod is usually aligned with the direction of weld travel and held down to a shallow angle with the work.

The rate at which you travel will depend on the thickness of the material you are working on, the amperage and the desired height and width of the bead. If you are too cold with the amperage or travel too fast, the bead will tend to stand up more, and even if it looks smooth and uniform, the weld won't pass a bending test because of lack of penetration **(see illustrations).** If you are too hot or moving too slowly, the bead will settle into the parent metal and leave a long crater or at least some undercut on either side. On metals thicker than 1/8-inch, the edges of the parts should be beveled, and you will probably make more than one pass. The bottom pass or *root* pass ties the pieces together, and subsequent passes can be made at a higher amperage. TIG welding is not a process that quickly fills in large gaps or beveled seams on thick plates. MIG or arc welding is better for that.

The filler rods used for TIG welding may look exactly like those for gas welding, but they aren't. The TIG rods for steel welding have deoxidizers added, in fact one brand calls theirs "triple-deoxidized," and if you accidentally use a gas welding rod, you can make a bead but the weld will exhibit porosity. The proper TIG rods are very helpful when welding low-quality steels, which may be the "hot-rolled" steel you find at the salvage pile at your local metal supply. There are a variety of filler rods, including special blends used for the low-alloy, high-strength sheet metals now used in automobiles.

After welding, your tungsten tips should be silvery and relatively clean. If they show signs of contamination, or are rough to touch, it usually indicates that air is leaking in somewhere in the torch or that shielding gas is leaking out somewhere. Check the cables and the fit of the collet and ceramic cup. If the electrode looks like it has a ball on the end like a miniature flagpole, then the amperage may be too high for that diameter of electrode. The tungsten shouldn't have a ball on the end bigger than 1.5 times the diameter of the electrode itself. If the tungsten tip is black or blue, check the timing of the post-flow on your machine. If the tip is black, it means the tungsten is cooling off in atmosphere instead of shielding gas and you need a little more time on the post-flow setting.

Look at the inside of your ceramic cup after welding and check it for contamination. A cup that is too small for the work will show deterioration and even cracking. Cups are usually sized in 1/16-inch increments. A #4 nozzle is four-sixteenths or 1/4-inch, a #8 is 1/2-inch. The general rule is for the size of the cup to measure 4-6 times the size of the electrode being used.

Your practice time before obtaining suitable beads will vary with your skill and previous welding experience. As we have said, the similarities between TIG welding and oxy-acetylene gas welding are such that you can pick-up TIG welding much faster if you are already an accomplished gas weldor. When you have practiced straight beads and flat, butt welds and have those down fairly well, then try some corner or Tee welds, and then vertical seams. When you are comfortable on these more difficult welds and the pieces you have done pass simple bending tests and cutaways to examine the penetration, you may be ready to do some actual projects with a TIG welder, but don't start right away with something critical like an automotive suspension or steering part, or anything that goes on an airplane.

TIG-welding aluminum

Perhaps a little more difficult to learn than welding steel, precision joints in aluminum or magnesium are among the primary reasons for having a TIG welder, so this should be your next area of practice.

You need a well lighted area to do TIG welding because the arc is not as bright as with other electric arc techniques, and on aluminum you really need to see how the puddle is developing. Set your machine to HFAC, with the high-frequency on continuous, or use the square-wave feature if you have it, and adjust the post-flow for 1/8-inch electrodes. Use a clean 1/8-inch electrode (pure tungsten) with the tip shaped for aluminum (not a needle-sharp point), and adjust the flow meter for the argon to 17-20 cubic feet per hour. Set an 1/8-inch-thick aluminum plate that has been cleaned thoroughly down on the welding table. Aluminum tends to age-harden when it is exposed to air, and aluminum oxides will constantly build up on the surface; such oxides are not always visible. Parts to be welded need to be cleaned just before welding to provide the freshest surface. Do not use chemical cleaners for preparing aluminum, unless it is something you have purchased at a legitimate welding supply store. Household or automotive cleaners may contain chemicals that will vaporize during welding and create toxic fumes. Use a stainless-steel wire brush (not steel wool) only, and never use this stainless-steel brush for cleaning plain steel. Since most sandpapers are made from aluminum oxide, and this is exactly the material we are trying to eliminate when cleaning aluminum for heli-arc welding, don't use ordinary sandpapers to pre-clean.

Use an aluminum welding rod, and clean the welding rods before starting, using alcohol and a lint-free cloth to wipe them down. The aluminum part you are welding should be well grounded to your table. If it is lightweight plate, put a heavy weight on part of it to keep it in contact with your grounded table, or attach the work clamp directly to the part. Otherwise, you may find that the bottom of an aluminum part you have welded on your table will have arc burns from intermittent contact with the table. Draw a bead on the aluminum for a short distance, then immediately cut off the arc. Examine this weld and you will probably see that where you stopped there's a small depression at the end of the bead. To avoid this at the end of every weld, don't back off suddenly, but rather ease out of it while slowly moving the tip back over the end puddle, and it should keep from sinking.

Try to maintain a puddle of molten aluminum about 3/8-inch wide. Going too fast will keep it narrower, and vice-versa. You won't see any real changes in the aluminum in the way you are used to seeing steel alter during welding. The only clue that an area has become molten is that it gets very shiny in that spot. You'll have to watch for this change in reflectivity when making a series of puddles that make up a bead, because aluminum melts at 1250°F, much lower than steel and can be overheated easily with the intensity of an arc. When welding steel, you have a whole range of colors that the metal turns at different temperatures. These colors make it easier to tell when you have the metal molten and when it is just starting to cool off. You won't have this advantage with aluminum, and you'll probably burn through a number of samples before you learn temperature control on this metal **(see illustration).**

6.33 When you are learning TIG welding on aluminum, keep your torch angle so that you can get a good, close look at how the puddle develops and the molten area becomes shiny. You need enough heat to form the puddle, but not much more or the aluminum vaporizes.

After the weld is completed and cooled, you'll see your bead is nice and shiny because all of the bead material has been melted and it looks as good as virgin aluminum just out of the foundry, but along each side of the bead you'll see a light-colored band that is contrastingly dull. This is an area where aluminum oxides have been vaporized from the weld bead by the high frequency in the arc and deposited next to the weld. If the aluminum plate you have been practicing on was not entirely clean, you may see some black spots, black "smoke" residue, or a wide dull area on either side of the weld bead, indicating too much aluminum oxide was present. Inspect your welds for uniformity, appearance and cleanliness, and check the backside of your welds for evidence of sufficient penetration.

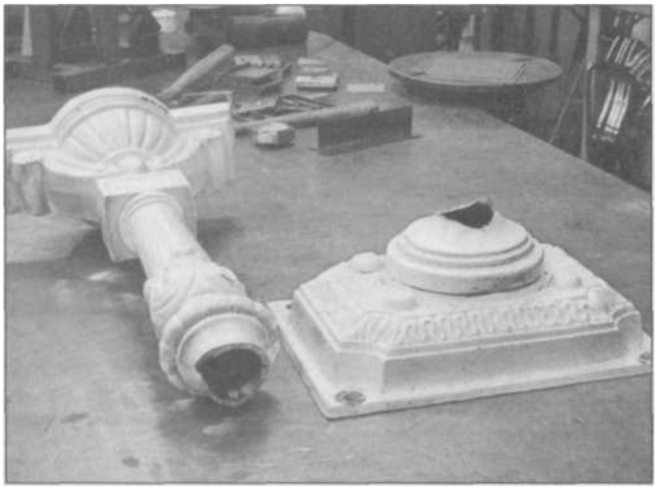

6.34 When welding thick aluminum materials, or aluminum castings like this fancy mailbox post, the edges must be thoroughly cleaned and beveled before welding.

6.35 The rosebud tip on an oxy-acetylene torch is the easiest method for preheating thick and cast aluminum before heli-arc welding. In order not to warp cast-aluminum heads, manifolds and blocks, use heat-markers on the metal while preheating.

When TIG welding any aluminum thicker than 1/8-inch, it is desirable to preheat the aluminum to 350°F before welding **(see illustration).** This is also particularly important when welding aluminum castings such as repairing cylinder heads, water pumps, aluminum engine blocks, bellhousings and other automotive castings. If the casting isn't preheated before welding, it will probably crack again later on, due to the welding section being expanded more than the adjacent metal, and then shrinking. Aluminum castings soak up a lot of heat, so the initial welding is made easier if the part is preheated because the part doesn't draw off the heat you're putting in to make the seam melt **(see illustration).** Preheating on automotive castings also helps bake out any dirt or oils that may be present, although any aluminum castings should be thoroughly cleaned before even attempting to weld them. Even starting with a clean casting and a Veed-out crack, impurities in the casting may keep coming out while you weld.

Ideally, aluminum parts should be evenly heated in a temperature-controlled oven, but some parts won't fit in a regular oven, and perhaps the boss of the kitchen may not allow it anyway. The alternative is to use temperature-sensitive paint sticks from the welding supply. Use one made for 350°F and dab some on various parts of the casting. Heat the whole part with a rosebud tip on an oxy-acetylene torch just until all the temperature indicators have melted **(see illustration).**

TIG-welding stainless steel should prove to be very similar to your practice on mild steel. What makes stainless steel different from ordinary steel is its content of both chromium and nickel, part of the reason for its much higher expense. You will find some characteristics that are different, such as the heat staying in the bead area longer than in plain steel because stainless steel is only about half as conductive of heat as regular steel. Once stainless steel does take on heat, it expands some 50% more than mild steel, which can cause problems of warpage in thinner materials if the pieces aren't clamped or jigged securely during the welding process. In critical welding in industry, jigs may be made of copper because its conductivity helps to carry heat out of the stainless parts quicker.

In terms of setup, stainless steel can be TIG welded with the argon flow set only slightly more than for steel, from 11-13 CFH for thicknesses

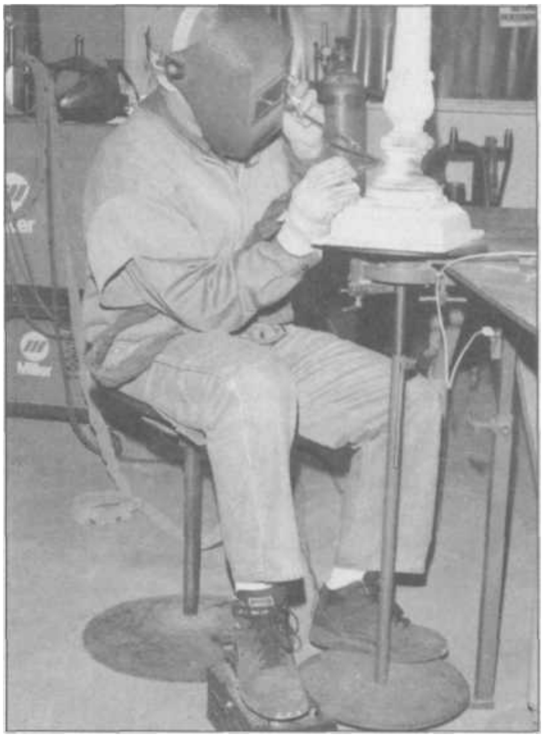

6.36 Weldor Bill Maguire is TIG-welding the heated mailbox post, using his rotating welding stand he invented. At the bottom is a round disc you can spin with your foot, leaving your hands free to continue welding, especially where you have to go *around* something. These should be available for sale at the time of printing.

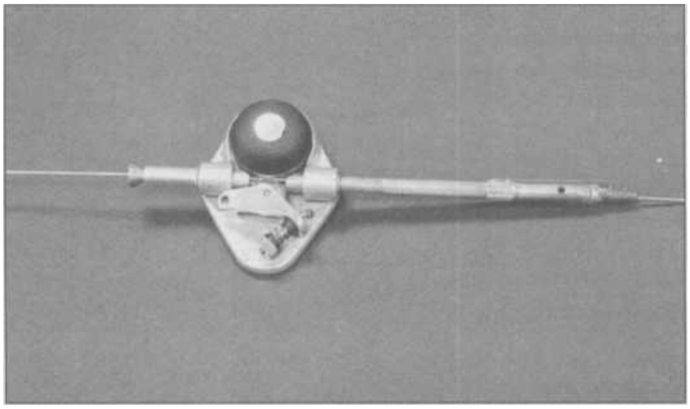

6.37 No one knows better what a weldor would need than a weldor. Another invention of Bill Maguire's is this "Pocket-Feeder", which is patented and for sale through major suppliers. It hand-feeds a welding wire, from .020-in. to 3/32-in., to your weld, while your hand stays the *same,* comfortable distance from the work. Your thumb on the wheel determines the feed speed.

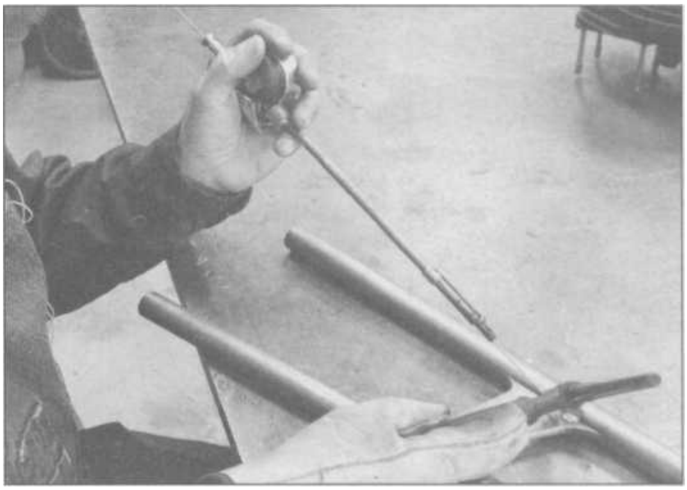

6.38 In use, the Pocket-Feeder can be used with either left or right hands, and you can reload new wire with one hand. Sensitivity on the thumbwheel is good, even with a light glove on.

up to 1/4-inch. Set your current controls for DCSP as you would for steel, with high-frequency for starting only, unless you are working on very thin pieces, in which case you might use HFAC instead because of its lower heat input and lesser tendency to burn through thin material. When using DCSP, the 2%-thoriated tungstens work best, with the pure tungstens good for HFAC on thin material. Of course the welding rods must be stainless steel. Practice on some stainless scraps until you get the hang of handling the different heat-transfer capabilities of stainless. Your puddling and rod-dipping action will be at a different pace than for mild steel.

In industry, critical welds on stainless steel are often made with a setup on the backside of the part to exclude air from the weld area. Of course, the shielding gas on top from the torch takes care there, but on the backside the weld can be contaminated. The pros form a box or shield around the back of the weld area and plumb some extra argon shielding gas in one end and out the other end into a tube to carry it away from the weldor. This will purge any air from the backside of the weld where it could cause any molten stainless-steel to crystallize. This technique is also used in protecting the back of titanium when TIG welding. When they weld critical tubing and can't get at the inside with a shield, they plug each end of the tube or pipe with tape and cardboard and fill the pipe with shielding gas. This is particularly true when welding stainless-steel tubing meant to carry medical ingredients or anything else that must be kept from contamination or oxidation.

Although you will likely never encounter other metals in your home/shop use, there are a variety of special alloys and rarely-used metals that are used often in specialized industries, and for which TIG is the primary or only method found to do the job properly. For instance, in the food-service and canning/bottling industry, there are numerous products that are transported through tubing or pipe. Besides stainless steel, nickel is often used in these applications, for its resistance to alkalines and acids. For the same reasons, it is used extensively in the chemical industry. But nickel-alloy pipe used in these situations cannot tolerate the possibility of a particle of slag, such as from stick welding or MIG welding, becoming included in the weld bead. It would become a weak point when faced with continued chemical attack if high temperatures are involved, until surrounding material is affected and the joint is weakened. For this reason, nickel work must be thoroughly cleaned before welding without using chemical cleaners, or a black nickel oxide coating can form.

6.39 **TIG welding is especially helpful in situations like this race-car suspension upright, where you have thin chrome-moly plate and tubing, and need total heat control. If you burn through on a job like this, the whole part may have to be scrapped.**

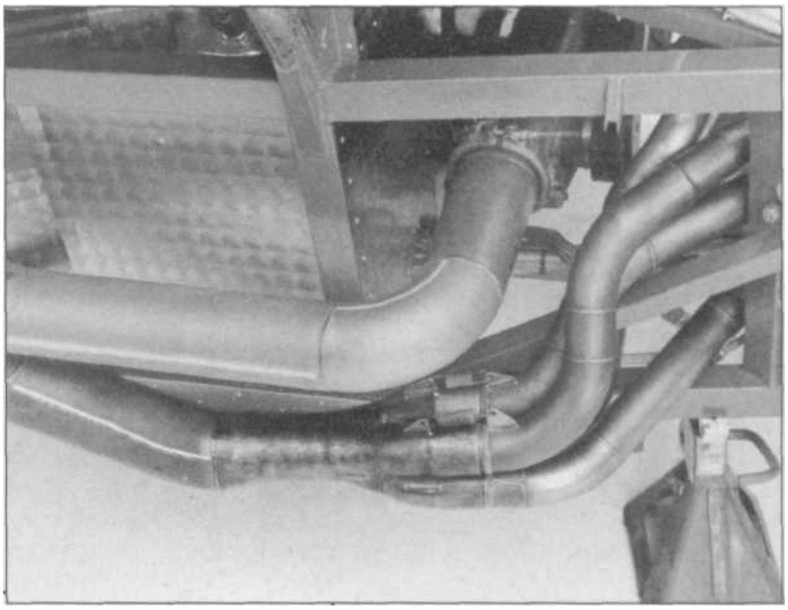

6.40 **Beautiful, precise welds on this race-car exhaust system typify experienced, professional TIG welding.**

For metals like nickel, copper, titanium, and others, TIG welding has proven itself to be the most adaptable method, especially in critical applications. Since there is no flux of any kind used in TIG welding, there is no chance of even minute slag particles being included in the weld bead. Industry is also capitalizing on the fine-grain structure of TIG welded seams (and the heat-affected areas nearest the weld). Helium is most often used as a shielding gas on nickel for its hotter arc, which allows for a higher welding speed, with a gas flow of 8-30 CFH, depending on work thickness. Polarity is usually set at DCSP (except for thin work where HFAC can be used to prevent burn-through), and the high frequency control set for *start-only,* using thoriated tungstens. Filler rods must be nickel or nickel alloy. Oxidation can be kept to a minimum by keeping the filler rod at a low angle to the work and keeping the rod tip in the shielding-gas envelope.

It isn't done often, but TIG welding can be used to repair cast iron. Your welding supply store will have some special rods that are cast of gray iron, called "three-in-one," and these should work for either gas or TIG welding although the same preheating instruction apply as above when we discussed welding cast aluminum parts.

Heli-arc or TIG welding is likely to remain as mostly a shop-only process due to the expense of the equipment and the length of the learning curve, but don't discount it as something you can't try. If you know someone in a shop with TIG equipment, ask them if you can spend a few hours closely observing them (with a suitable spare helmet on, of course) performing TIG welds. The lack of spatter and smoke from TIG welding means you can observe this process better than most other welding methods, and this will tell you much more about how the process works than a chapter in a book. If you feel you have observed enough to get a good idea, see if they will let you try it out, but first offer to pay for any tungstens you ruin in your practice. You may find you have a hidden aptitude in hand-eye coordination, and TIG may be a process that appeals to you.

Notes

7 Plasma cutting/welding

Just as the emergence of lower-cost, higher-efficiency MIG machines has made something of a revolution in small shop and home/shop welding in the last ten years, so has plasma technology changed the face of metal *cutting.* While the technology has been around for a while, it has been seen by most weldors as a high-technology setup that seemed complicated or hard to use. Plasma technology's biggest usage had been in plasma arc welding, or PAW, and it is still used in this form in many industrial applications.

The simpler plasma-arc *cutting* (or PAC) technology got very little play at the time, and it has taken some years for the equipment to filter down to the general welding/cutting marketplace. As with the MIG welders, what was originally considered to be a professionals-only setup has entered the province of the small shop and the amateur fabricator, due mainly to the introduction of less-expensive imported equipment that really opened up new markets with hobbyists. The domestic welding manufacturers suddenly became aware of these newer markets for smaller versions of equipment they had been selling to industry all along and brought increasing competition to the fray. Today there's a wide selection of plasma machines to choose from, from small 110V AC portables to large shop equipment for materials up to one inch thick.

7.1 Plasma cutting can be done outdoors, being very little bothered by wind, makes very clean cuts on ferrous and non-ferrous materials, and produces no smoke and very little sparks. It has become the modern way to cut metal. The operator here should be wearing gloves and a long-sleeved shirt.

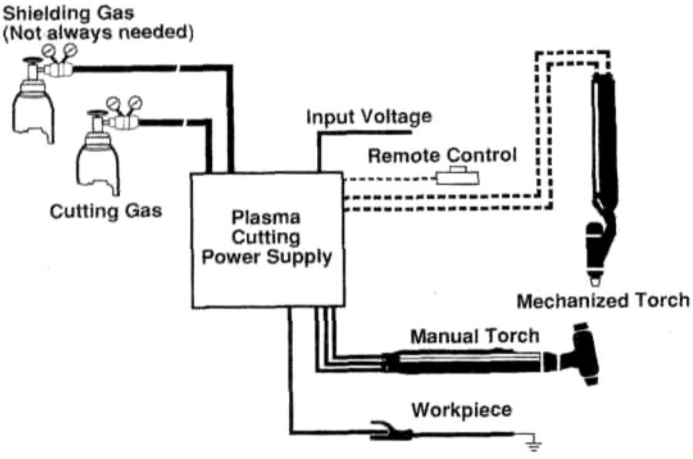

7.2 This illustration shows the basic components of the plasma-arc layout. The two bottles of shielding gas are really only required on the large industrial machines, not the home/shop portables which run on compressed air. The medium and larger sized plasma cutters can be used with a straight (mechanized) torch body for trace cutting.

Usage today is increasingly with the cutting side of plasma technology. With an incredibly easy learning curve and not a lot of special installation costs, the benefits of plasma-arc cutting, usually just called plasma cutting, are being reaped today by body shops, muffler shops, wrecking yards, sheet-metal shops, duct repairmen, street rodders, race-car builders and fabrication shops. Among the benefits are use of plain shop air-compressor air, no dangerous gasses, no flames, smoother cuts, less heat distortion in the edges you cut, little oxidation of the edges and incredible speed.

In basic principle, the plasma process uses a torch something like a TIG torch, but the electrode is deeply recessed up into the torch, where it cannot contact the work and be contaminated, and the arc is maintained as a needle-like point when the air between the electrode and the work is ionized. The air or gasses around the electrode are heated to such a point that they reach the *plasma* state, in which the gasses themselves can *conduct electricity.* While the gas is ionized and making continuity for the arc process between the electrode and the work, another gas is introduced around the nozzle. In the case of plasma welding, the gas is an inert shielding gas; in the case of plasma cutting, it is compressed air. In either case, welding or cutting, the action takes place in a very narrow stream.

Plasma-Arc welding

The original plasma technology was developed in the laboratories of the Linde Air Products Company (a pioneer in many welding processes, now the L-TEC Company) back in the late Fifties. They and other companies have worked on refining the equipment and the process ever since. You won't find plasma welding equipment at most welding supply stores, but it is in a few industry catalogs. This is the kind of equipment usually sold through a company representative. Most of this equipment is used in high-precision joining of very thin materials. Although, up until the last ten years, both plasma welding and cutting have been virtually reserved for industrial purposes only, and most of the plasma-arc welding equipment is still expensive and for limited industrial usage, the process is worth looking at.

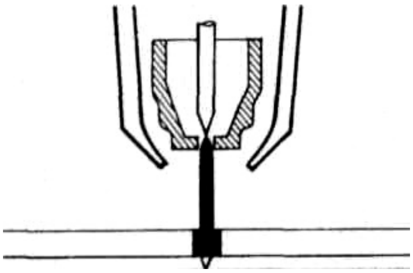

7.3 In the needle-arc version of plasma welding, the narrow-and-tall arc forms a weld that makes 100% penetration, top and bottom. The technique is used mostly in precision industrial applications.

In operation, the plasma torch has a central tungsten electrode tucked up inside a nozzle that is more restrictive than other welding gun nozzles, that is closer to the electrode and increasingly tighter as it gets close to the tip of the electrode, until there is just a hole for the arc to come through **(see illustration).** An "orifice" gas is forced through this nozzle, and it speeds up like air/fuel mix going through a carburetor's venturi. The gas is usually argon.

Inside the welder's power source is a high-frequency arc-starting system not unlike a conventional TIG machine, and this initiates the arc, since the electrode doesn't stick out and can get close enough to the work metal to start the arc itself. This is called the "pilot-arc" circuitry, and it generally cuts out after the arc is established. The current in the electrode superheats the argon until the gas itself conducts electricity, and a long, extremely hot arc is created that welds the work. A second inert gas is introduced in a large nozzle around the inner one, such that a cone of shielding and cooling gas is formed around the long arc column to protect the weld area from contamination, much as in other electric arc-welding processes.

There are several methods of welding with plasma technology, a conventional method akin to TIG welding where filler rod is added, a needle-arc process where parts are fused without filler rod, and a keyhole method where a weld without filler rod can be made that actually fuses the parent metal on the whole mating surface at the same time, top and bottom. The latter produces as pure and uncontaminated a joint as possible. The benefits of using plasma-arc welding over conventional TIG welding is that it is easier to operate as far as maintaining the arc length exactly and the weldor has a better view of the process.

Although you may find one or two PAW units in a catalog, most of the technology is used in industrial situations with mechanized torches and water cooling with amp ranges up to 300 or 400. Most manual-operation plasma welders do not have as much amperage as conventional welders; L-TEC's catalog shows a 110V 15-amp unit, and a 100-amp 220V three-phase machine weighing some 730 pounds that only welds up to 1/8-inch plate, although it has a 100% duty cycle at maximum output.

For those reasons, the plasma-arc welding process is probably not for your home/shop. Much more flexible, lighter and more affordable conventional equipment is available for making welds. However, there is no question but that the practical application of the technology in everyday situations is eminently provided by plasma-arc *cutting* equipment.

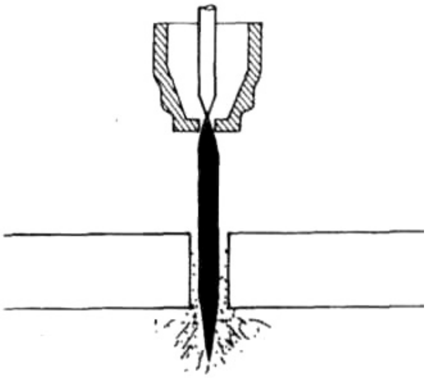

7.4 **In plasma-arc cutting, there is usually only one gas (in all but the biggest machines), and that is air from the shop's air compressor, which is forced around the electrode to blow the kerf out neatly.**

Plasma-Arc cutting

In plasma-arc cutting, the technology is much the same as welding, but the arc is even more constricted, with the plasma temperature so high that the arc can cut *any* metal. The very restricted arc cuts a pencil-thin line through the metal, and a sharp force of gas through the nozzle blows the melted metal from the kerf **(see illustration).** The velocity and steadiness of the cutting action is such that the edges of the cuts are kept very straight and clean.

The benefits of plasma-arc technology in cutting metal are many. The cuts are very clean, with little oxidation of the metal as found in oxy-acetylene cutting, the velocity of the action creates smoother edges that need a great deal less cleanup, and in fact, with the industrial units that are mechanized much like a gas flame-cutter or pattern-cutter, the cuts need no cleanup or preparation for welding **(see illustration).** There is little or no slag found on the bottom edge of the kerf.

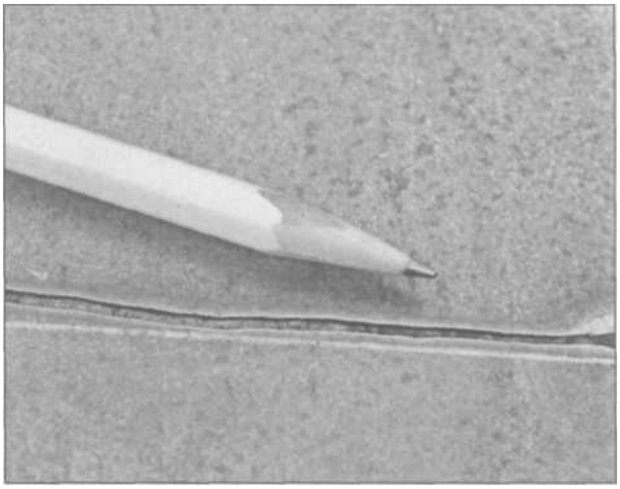

7.5 Plasma cutters have become very popular with auto bodymen for the cleanliness and speed in cutting body panels. These sheet-metal edges need very little prep, and a plasma cutter can cut painted or rusty metal just as easily.

Plasma-arc cutting is considered much safer in many work environments than gas cutting. There are no flammable gasses involved, there is no smoke from the cutting action to bother the weldor, he can watch the process much more closely and there is much less spray of sparks coming from under the cut. If you remember the little sparking toys you had as a kid, where you pushed a plunger and a wheel went around, scraping against a tiny flint making sparks, that is a small-scale version of what you see when plasma cutting. This is in high contrast to gas cutting, where there is a lot of heat generated, and blobs of molten metal are constantly falling on the floor under the cut, creating a fire hazard.

There is *some* heating action in plasma cutting, but in a short time, depending on the thickness of the material you have cut, the parts are ready to be handled for further work. With a gas cutting torch, the parts are still red-hot. Some have described the plasma process as almost "cold cutting." Another benefit of the cooler cutting action is that much less heat is induced into the parent metal, so when working on thin materials there is greatly reduced warpage.

This is what has made the plasma cutter so popular with automotive bodymen and sheet-metal workers. Thin materials can be cut with virtually no distortion. The cuts can be made, as one bodyman described it, "as fast as you could draw a line with a pencil." A few minutes with a file or sandpaper and the edges are ready for some other work, whether to be welded or left as a finished edge. Because of the lower heat and reduced sparks, panel-cutting on cars is much easier, even if there is a secondary panel on the other side because there is less damage to what's on the other side like other painted surfaces that need to be left untouched. Bodymen also like the fact that you can make a cut, lay down the torch, do something and pick up the torch again and start cutting with no other functions to perform other than putting your safety helmet down and pulling the torch trigger. With a gas cutting torch, there is always the danger of what to do *between cuts* with a torch flaming full-bore. If you keep shutting it off each time, you have to go through the process of re-lighting and adjusting it every time.

The last reason is why plasma-arc cutting has become the tool of choice now in wrecking yards across the country. In a yard full of discarded cars with flammables like upholstery still in them and various pools of oil, gasoline and other fluids soaked into the ground all around, oxy-acetylene cutting on cars in wrecking yards has always been a hazardous job. Also, the torches and bottles are a heavy setup and had to be hauled around to various parts of the wrecking yard. With a good plasma machine, the 220V cutter can be plugged in a small shed central to the yard, while a plasma torch cable as long as 234 feet could be used to cut almost anywhere with a lot less fire danger and much faster speeds, 10 or 20 times faster than with acetylene equipment.

Because the plasma machine has arc-starting circuitry built-in, most plasma torches do not need to contact the metal to get started, which is a decided advantage in body work. Once the arc is started and the gas is ionized and conductive, a plasma torch can cut through rusty metal, painted metal, even sections of a car body that have Bondo or undercoating on them. When using an oxy-acetylene torch to cut painted sheet metal, the cut edges are blackened, scarred and blistered, while with plasma the edges have a gray discoloration near the edge and that's it.

In fabricating shops, one of the biggest advantages of plasma-arc cutting equipment is the fact that all metals can be handled equally, from steel to aluminum, to stainless and plated or galvanized metals. It is the nonferrous metals like aluminum and austenitic stainless steel, which are normally hard to cut cleanly with other methods, that make the plasma machines appealing to shops. The author remembers his first experience with a plasma cutter some years back. An aluminum motor plate was being made for a race car, and we needed to cut a 14-inch hole out of the center to clear the flywheel. The material was 1/4-inch-thick aluminum and 7075-T6, which is so hard you can have difficulty even drilling it unless you have a fresh bit. The idea of using a sabre saw or band saw seemed tedious and out of the question. We took it to a shop that had a plasma cutter and within minutes of making a plywood circle template to run the torch around, we had our hole cut with very little heat and perfect edges. It was eye-opening, and when you try one you may be similarly amazed.

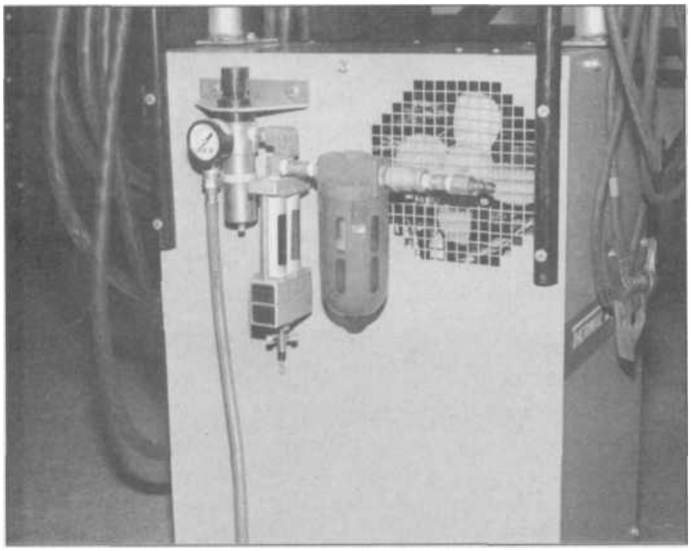

7.6 Shop air-compressor air is all that's required to run most plasma cutters. Here on the back of a Thermal Dynamics machine is the air filter and pressure regulator.

When plasma-arc cutting equipment first came into wider, less-industrial usage, some traditional fabricators and bodymen were intimidated. The technology wasn't widely understood, and the term "plasma" seemed to indicate some highly-technical process akin to esoteric lasers. Once the machines got out into the field, they quickly became accepted for their speed, cleanliness, lack of maintenance, lower fire hazards and the ease with which a weldor can operate them. It's almost a pull-the-trigger-point-and-shoot kind of process.

Perhaps one of the bigger attractions of plasma cutting equipment we haven't mentioned yet is the type of gasses used in the process. While the big industrial plasma machines do use traditional shielding gasses, like argon for the arc "pilot" gas and hydrogen, nitrogen or a gas mixture (like 35% hydrogen/65% argon) as the cutting gas, most portable plasma units today use nothing more than *compressed air* from a shop air compressor **(see illustration).** There are no expensive bottles to buy or periodically refill.

This is how the shop-air works. When the gas in the arc column is ionized, it becomes the arc medium between the work and the electrode. Nitrogen happens to ionize well and makes for a hot arc, and also happens to make up about 75% of the air around us by weight. Most plasma cutters use a high velocity of shop compressed air to become the ionized arc gas, the cutting gas that blows the material out of the cut, and the cooling medium. It can be assumed that all shops already have a heavy-duty air compressor on hand for operating other equipment, so using shop air greatly simplifies the use of a plasma cutter and keeps expenses down. In industrial situations, pure nitrogen or hydrogen or a mix is used for totally clean cuts in precision work, but the shop-air cuts are more than good enough for any automotive or fabricating-shop usage.

Choosing plasma cutting equipment

The large, expensive and heavy industrial plasma cutters and welders are for specific purposes in high-precision and high-production mechanized work, often requiring two gas bottles, three-phase power for 200-600 amps, and use water cooling for the torch as well. For these reasons, we'll concentrate on the wide array of more portable plasma cutters that are available for home and shop use, most of which operate on 220V, single-phase current and use only shop compressed air for cutting and cooling.

At the most basic level, where a home/shop user might be interested, there's a number of domestic and imported plasma cutters that weigh 75 pounds or less and are capable of cutting material thicknesses that an automotive user would come across. If you are comparing plasma cutters to welders, don't be alarmed at the amperages on the smaller plasma machines. Units with 50-amp capability can cut steel up to 3/4-inch thick. An area where there is a comparison with welders is that smaller plasma machines have lower duty-cycles and their cutting speed slows down when thicker materials are cut.

In the 110V category, Daytona Mig offers the Pocket Plasma , which puts out 25 amps at a 20% duty-cycle and cuts sheet metal at 100 inches per minute, though it is said to cut up to 3/16-inch steel material. It requires 70psi air input at a rate of 4-6 cubic feet per minute (cfm). Miller makes several 110V units, with the Spectrum 187D cutting up to 9/64-inch steel material and weighing 47 pounds. Thermal Dynamics starts their extensive line off with the DynaPak 110, which cuts up to 3/16-inch steel, uses all solid-state circuitry, sequential status lights, and has an amazing 80% duty-cycle at 20 amps (see illustrations). With all of the machines available, the maximum cutting thickness is usually given for mild steel, while aluminum and other non-ferrous material take more amps and the same machine can only cut a lesser thickness of these metals. On some machines, a figure may be advertised that gives a maximum thickness for cutting

7.7 The compact end of the plasma cutter market is represented by 110V machines such as this Plasma Pocket from Daytona MIG. It puts out 25 amps, weighs 55 pounds, and requires 4-6 cfm of air at 70 psi.

7.8 Another quite-portable plasma machine is the PCM-VPi from L-TEC, which offers cuts up to 5/8-inch in steel, variable amps from 10-40, and a 60% duty-cycle at its rated maximum. It operates on 220V current.

Miller
Spectrum 187D
WARNING

7.9 One of Miller's smallest plasma cutters, the Spectrum 187D weighs only 47 pounds, yet cuts up to 9/64-inch materials, and runs on 110V house current.

OFF
ON
HTP
Micro Cut
250

7.11 HTP's entry-level plasma cutter takes 220V current, but weighs only 56 pounds and is said to cut up to 1/4-inch steel. It has a 100% copper-wound transformer.

MATCO TOOLS
PLASMA CUTTER

7.10 The popularity of plasma-arc equipment has really jumped in the last five years. Now imported units are sold here, and tool companies like Matco have them, as well as the traditional welding equipment manufacturers.

and another, higher figure for severing. The severing action is a short cut, such as when cutting a length of material.

Beyond the 25-amp range that seems to be the limit for 110V equipment, there are machines for 220V input that are still quite portable. HTP America offers affordable plasma cutters from their Micro-Cut 250 that cuts up to 3/16-inch steel, to the Micro-Cut 375 which can be used on 110V, 220V or 460V input and offers 7.5-35 amps, cutting up to 3/8-inch steel, as well as larger shop units **(see illustration).** L-TEC has a popular unit, the PCM-VPi, which weighs only 46 pounds with its plastic case and carrying handle, but is said to cut up to 5/8-inch steel with its 50 amps of power, and features a 60% duty-cycle at 40 amps output **(see illustration).** It is a good choice for any automotive shop, and can sever up to 3/4-inch material such as rebar for construction sites. Lincoln's entry-level plasma cutter is the Pro-Cut 40, with a 60% duty-cycle at 40 amps, 220V input, cuts up to 3/8-inch steel, and offers a timed pilot arc and timed post-flow. Daytona MIG's medium-sized lineup includes a 35-amp, 220V Plasma Prof 35 that is said to cut up to 3/8 steel quickly, and 1/2-inch steel at 15 inches per minute.

Above the 50-amp range, plasma-arc cutting equipment is in the serious shop equipment arena. Most of these machines are for 220V, *three-phase* power input, which you won't have at most home garages or small shops. Within this large group of choices though, there are cutters with total amperage ranging from 50 amps to 150 amps or more. In this category, you'll find the Lincoln Pro-Cut 125, with 125 amps, 60% duty-cycle, a microprocessor-controlled overshoot limiter, and "Drag-Detect" circuitry that is said to prevent consumable damage. In this machine, let's look at the duty cycles for a professional-grade plasma cutter. Rated at 60% duty-cycle at its highest 125-amp output, it has an 80% duty-cycle at 110 amps, and a 100% duty-cycle at 100 amps or less. At its highest output, this machine can cut 1 1/4-inch steel, and for ordinary 1/2-inch plate, it can cut all day long. That's the difference in the heavier professional machines; they have stronger components to work heavy materials not just a few times a day, but all day, which is why they are overbuilt and too expensive for the home/shop market, even disregarding the three-phase input power required.

Other powerful plasma-arc cutters include: HTP's PCA 65, with 60 amps (220V single-phase), cuts up to 3/4-inch steel (clean cuts on 5/8-inch steel), a 50% duty-cycle at 60 amps, and a weight of 180 pounds. HyperTherm has their 100-amp MAX 100, that will cut 1 1/4-inch steel on 220V or 440V three-phase power, with CC (constant-current) transistorized output, ESAB offers the PCS-90 that cuts up to one-inch steel with three-phase 220V or 440V, and has air pre-flow, post-flow and a low-air-pressure sensor. L-TEC has the PCM-1000i, which can cut one-inch steel at production speeds with its 80 amps, and a 70% duty cycle at that maximum amperage, yet weighs only 80 pounds. Daytona MIG has a full line of heavy-duty plasma cutters, with the Plasma Prof 120 being near the top with 120 amps (220V three-phase), cuts up to 1 1/8-inch steel and 7/8-inch aluminum, with quick-disconnect torches so you can switch to a mechanical torch for a cutting table (see illustrations).

Most of these machines offer such features as infinitely-variable output control, which means that you can dial in the exact requirements for a job, say cutting through an outer skin on a vehicle without cutting panels on the other side. Of course, you can cut thin materials with more power than

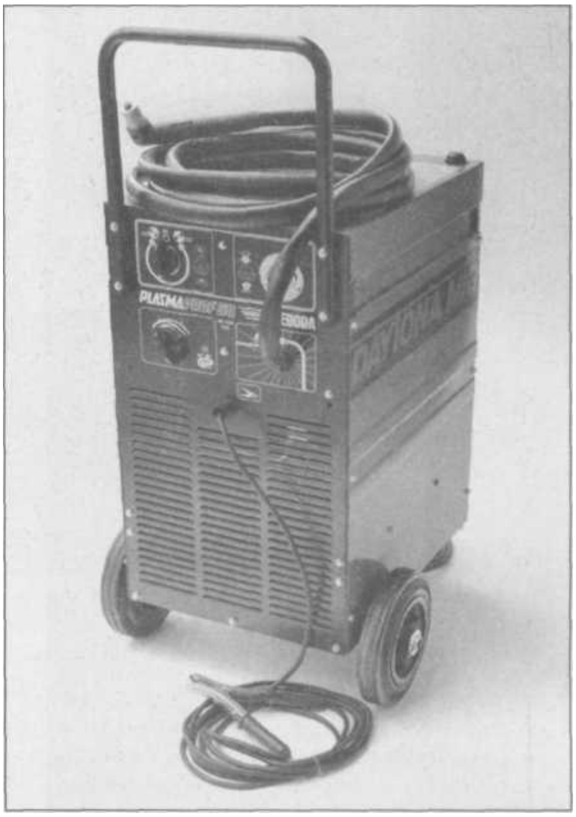

7.12 **The bigger-amperage machines are all in the 220V category, and many, like this Daytona MIG Plasma Prof 50, require shop-type three-phase current, too. This one has a 35% duty-cycle at its rated max of 50 amps and 100% at 30 amps.**

PCM-750i
L-TEC

7.13 **Still light and portable, but powerful enough to cut 3/4-inch steel, The L-TEC PCM-750i has 50 amps with a 40% duty-cycle at that setting, and has its own torch storage spool and spare parts storage compartment.**

PLASMA
150
DAYTONA MIG

7.14 **The larger plasma-arc cutting machines are all in the three-phase power category, like this Plasma Prof 150 from Daytona MIG. It weighs 317 pounds, has a range from 20-150 amps, and can cut 1 1/2-inch steel and 1-inch aluminum at 30 inches per minute.**

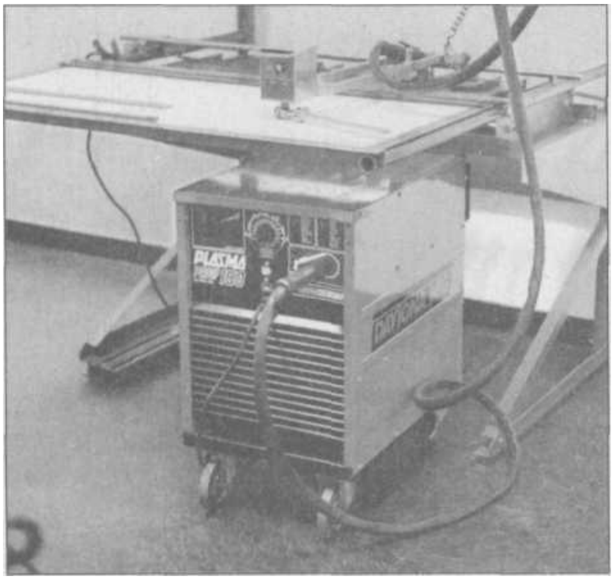

7.15 Larger machines are usually offered with an optional 180-degree (straight) torch for use with shape-cutting machines for industrial work.

is required, but using just the right amount means you have the least danger from flying sparks. On the less-expensive end of the plasma cutter spectrum, some of the machines offer only a knob with several predetermined positions.

Beyond the 80-120 amp range of plasma cutters, where really large plates are cut all day long, the equipment offered is no longer usable with shop air, requiring two inert gas supplies and water-cooling for the torch. However, they offer from 100 amps to as much as 600 amps (the higher amperages are usually found on machines built to be used with mechanized torches), and all with 100% duty-cycle. Obviously, that kind of power doesn't arrive at your doorstep without a hefty invoice.

Using a plasma cutter

If you thought the modern MIG welder brought great advancements in ease of the learning and operating curves, wait until you try plasma-arc cutting. It's as much of an improvement over gas-torch cutting, in both learning curve and reduced maintenance. We watched as a race-car mechanic in a large shop was being trained by another worker to utilize their plasma cutter. The newcomer to the technology made his first 4-inch by 6-inch part from 3/8-inch steel plate after only fifteen minutes of instruction. After experimenting with different speeds, power levels and materials, he was ready to do real jobs after an hour. You can't learn everything there is to know about the process in your first day, but this story is typical of first-time users.

Unlike welders, most plasma cutters of the shop-type variety are ready to use when you uncrate them. Their power cord is ready to plug in (you don't have to put on your own plug), the torch is usually already hooked up to the front of the machine, and all you have to do is connect your shop air supply before making your first cuts.

Although the risk of fire or injury is less with a plasma cutter than with a welder or oxy-acetylene cutting torch, you must still take every precaution before starting. Read over the instruction manual that came with your machine before doing anything. When you are familiar with all the controls, take the torch apart, so you'll understand how it works. Specific design of electrodes and other parts may vary from make to make, but, whatever the make, you'll notice that the electrode is like no other you've seen. Instead of a six-inch-long rod of tungsten as with a TIG torch, the plasma electrode is a machined part.

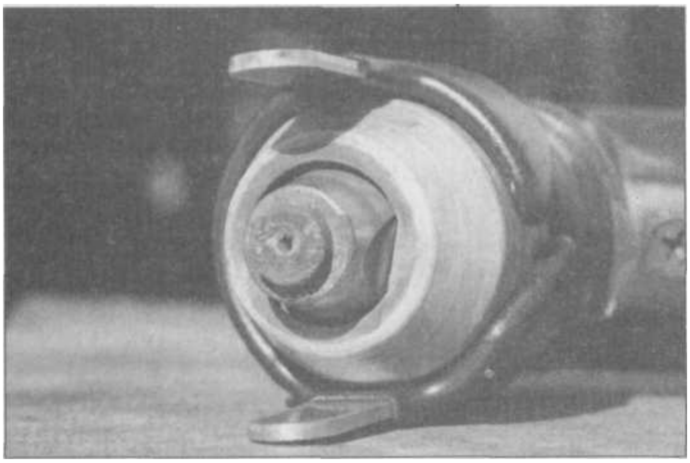

7.16 Eventually, enough material will blow back from cutting to cause a building of slag on the exposed nozzle and orifice of the torch, requiring periodic maintenance.

Although it never comes in contact with the work, the plasma cutter's electrode should still be considered a "semi-consumable" item. Electrodes can last a long time, but keep several spares on hand because they do get worn and contaminated with material blown back from the cutting **(see illustration).** Besides the electrode, you'll find several other pieces inside the plasma torch, including the nozzle (also called a shield cup, retaining cup, and gas diffuser), and a cutting tip, which is what converges the airflow past the electrode **(see illustration).** Different brands of torch may have small springs, insulators or swirl rings, but the basic principle is the same for all brands.

There is much less danger with a plasma cutter than other equipment, yet there are still hot sparks generated. There is also ultraviolet radiation. Because the process doesn't seem as bright or intense as welding, some users

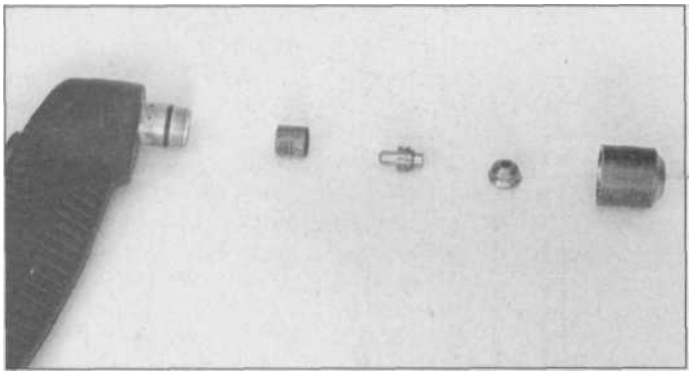

7.17 The main components of a typical plasma-arc cutting torch are, from left: torch body with o-ringed and threaded extension, the swirl ring, the electrode, the cutting nozzle, and the retaining cup. Some brands have a slightly different arrangement of parts inside.

7.18 Hot rodders never leave anything alone. Rugged Thermal Dynamics 50-amp plasma cutter used in a rod shop has been modified with the addition of a tubing rack and box on top to hold nozzle dip and plasma consumables.

are tempted to perform plasma cutting without proper safety wear, but your skin can still be burned. So wear long-sleeved shirts with the collar and cuffs buttoned up, and wear leather welding gloves and a full-face shield. If you ever have to work on your machine, it should be unplugged, even if all you are doing is changing the electrode. The machine's output is high enough voltage to cause severe shock and injury. Also, if you are cutting any painted, plated or galvanized metals, ensure that you have adequate ventilation. The plasma process doesn't create fumes by itself, but what's on the surface of the metal can.

On most machines, there is a high frequency arc-starting circuit that makes it unnecessary to touch the work with the torch to initiate the arc. Set your output voltage at the machine, depending on the thickness of the metal you are cutting. On some of the smallest machines, there may just be an on/off switch, or a switch with two or three amperage positions. Most machines have a variable knob spanning from the machine's minimum amperage to its maximum.

Use a guide of some kind to make straight practice cuts **(see illustration).** Place the torch so that the outer cup rides on your guide with the nozzle the suggested distance from the work and just push the button. The arc will start, it will

7.19 You'll need to keep on hand some spares of each of the torch components, including wrenches for changing the tip and the electrode.

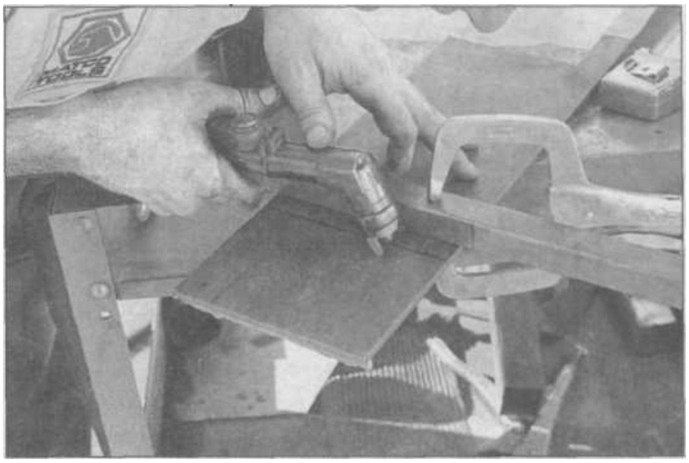

7.20 The plasma cutter makes very clean cuts, so any raggedness on yours means your hand wasn't steady. Use a sturdy guide whenever possible. With a piece of tubing as a straightedge, here a dry run is being made to ensure the torch follows the correct cut line.

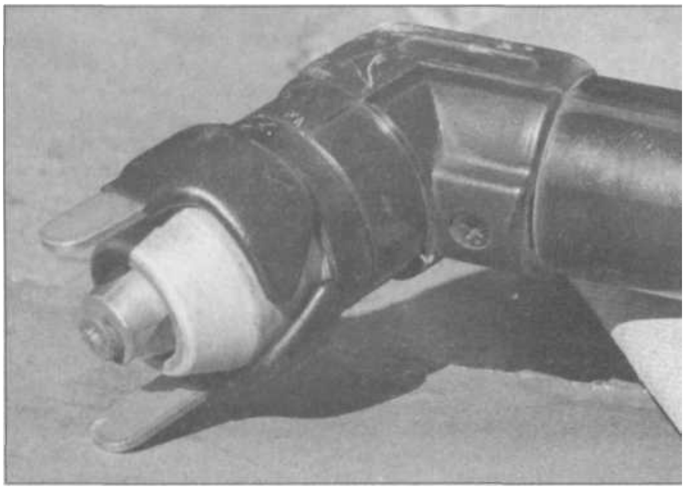

7.21 This clever tool snapped onto a Miller plasma torch holds the torch just the right distance away from the work, so if you have this and a sturdy guide, your cuts should be almost as good as machine-operated torches.

7.22 Same tool as the previous photo, showing how it holds the tip away from the work, and provides a good surface to lightly drag against your pattern or straightedge.

pierce the metal (no need to go to the edge of a plate to get the cut started) and begin blowing the slag out the bottom. Move along at a steady speed and you're *cutting.* You'll develop a feel for the correct cutting speed for various materials, but the instructions with your machine should tell you the recommended speeds for various thicknesses and materials. Practice moving the torch a few times without the arc, just to see how fast or slow 15 inches per minute may be, and what 100 inches per minute feels like.

Examine the cut edges of the material. If the lines in the cut are not straight up and down, your speed was off **(see illustration).** If the lines turn to the direction the cut was made from, your cutting speed was too fast. You can also have someone else observe you from a short distance while you are cutting. They will tell you if the sparks went straight down (proper speed), or sprayed more to the direction the torch came from (too fast). Moving too fast can cause premature failure of the nozzles.

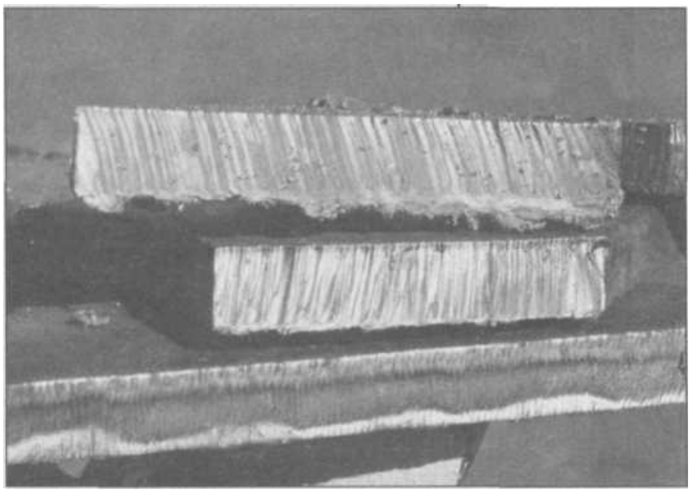

7.23 At top here are two cuts made with a hand-held plasma cutter. In the top one the torch was moved too fast and with not enough amperage. The middle plate had the speed more correct. The bottom plate for comparison only was oxy-acetylene cut with a machine, so this is the smoothness you are trying to duplicate. Only a gas torch cut would have this discoloration of the metal showing.

When the part you have cut is cool enough to pick up, examine the kerf. If your cutting speed and amperage was right, the small amount of slag at the bottom of the kerf can be easily clipped off with wire-cutter pliers, and the rest of the metal should take minimal cleanup.

In cutting shapes, you can use plywood patterns to cut repeated shapes such as round or square holes or anything else. Just make the template the right size for the torch head to follow. For basic circles, most welding supply shops have a circle-cutting tool (sometimes called a circle-cutting "compass") that can be used with a plasma cutter.

By far the most important thing to maintaining and operating your plasma-arc cutter is the condition and quantity of the air it runs on. Read the manufacturer's recommendations for the size of compressor you need and the pressure and cfm requirements. If you try to make a cut when you don't have enough air, most machines will either not start or make a grinding-like noise inside. Besides having *enough* air, plasma cutters also require *clean, dry* air. Most will come equipped with a drier/regulator right at the air-line connection on the back of the machine, but you also need a good drier on your compressor. And, if your compressor is

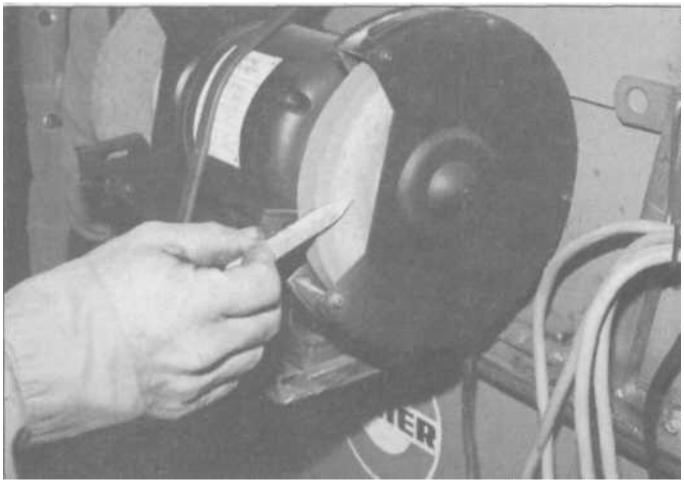

7.24 As with traditional torch cutting, soapstone is useful to mark your cut lines because it won't burn off. With the reduced heat of plasma, you could also use a marker pen. Sharpen the soapstone before drawing.

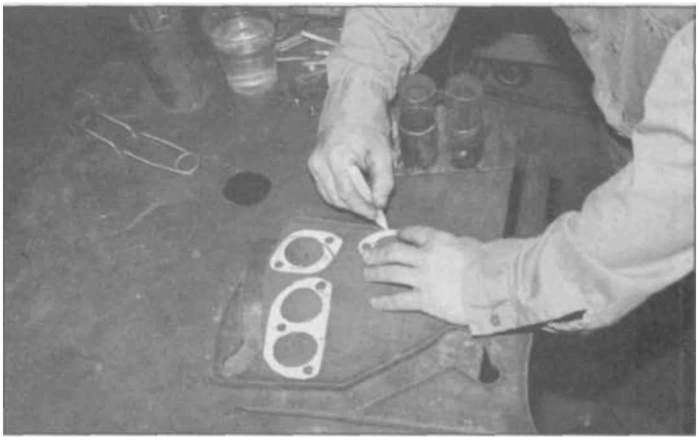

7.25 Here an exhaust gasket is being traced to make thick header flanges from a piece of 3/8-inch steel plate.

very far from where you hook up the flex line to the plasma cutter, it's a good idea to install one on the wall just ahead of the connection for the flex line **(see illustrations).** Your compressor and the driers should all be drained every night to remove dirt and moisture. Moist air will rapidly deteriorate the electrodes and nozzles.

Cutting too fast or trying to cut too thick a metal for the size machine you have will cause some slag buildup inside the nozzle of a plasma cutter, not unlike the buildup inside a MIG gun. In fact, you may use the same anti-spatter spray or dip in your plasma cutter as you use on your MIG torch to make the slag buildup easy to remove. Eventually, even if you take care of your plasma torch, you will have to replace nozzles and electrodes because usage will enlarge the holes, making for a less-concentrated arc. If any slag is "fried" onto the nozzle or electrode, sand it off by holding the part squarely against sandpaper on a flat surface, and then clean out any orifice holes.

7.26 The 50-amp plasma cutter makes short work of cutting out the patterns and can easily pierce the center of a pattern to cut out internal holes and shapes.

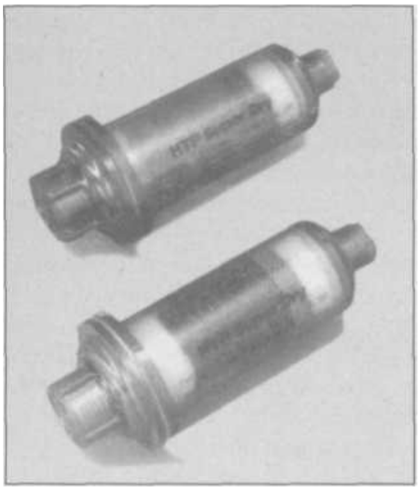

7.27 Dry, clean shop air is extremely important to extending the life of your plasma nozzles and electrodes. These HTP Super-Dry air driers have a bronze filter and desiccant inside that changes color when it needs to be changed.

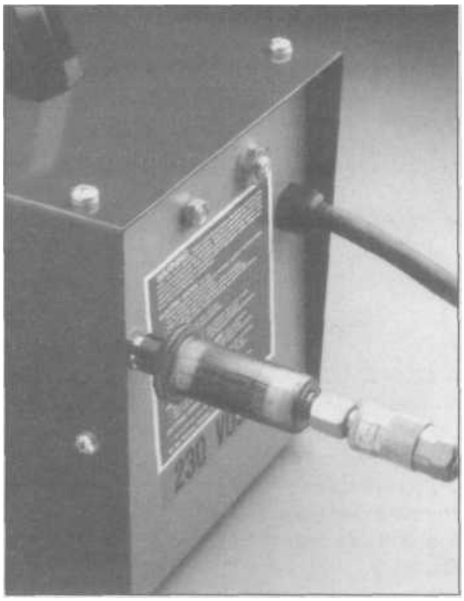

7.28 In most applications the HTP Super-Dry disposable filter/driers can be screwed right onto the rear of your machine.

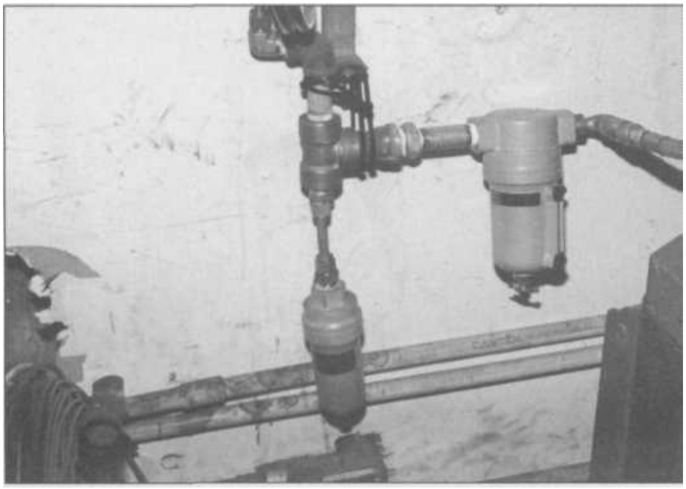

7.29 You can't have the air too dry for use with a plasma cutter. If your compressor is outside or a long distance away, use an extra filter/drier on the wall where the hard line switches to the flexible hose for your plasma cutter. Drain daily.

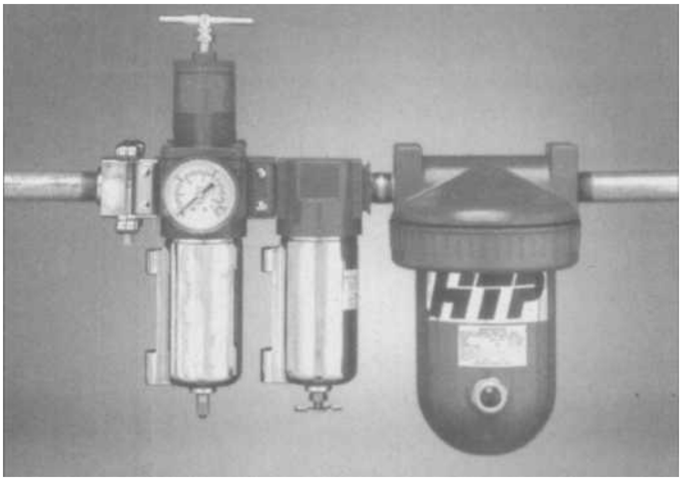

7.30 The Max-Dry from HTP is a three-stage system. Stage one is a self-draining 5-micron filter, stage two is a "coalescing" filter, and the final stage is another filter with a large amount of desiccant. The desiccant changes color when it can't absorb any more humidity, but can be dried out in an oven for reuse.

You'll find there isn't too much to discuss further about how a plasma cutter operates. They are at the same time technologically advanced, and yet simple to operate and maintain. They perform exactly as they are claimed to, in terms of speed, cleanliness of the cuts, lack of cleanup required and the minimal heat induced into the adjacent parts. Plasma-arc cutting is most definitely the preferred modern method of cutting metals, in situations ranging from an air-conditioning duct repairman or an automotive bodyman working on gauge metal to industrial plants where shapes are cut from thick plates all day long. To most automotive enthusiasts and race-car shops, the modern shop combination is a good MIG gun for welding and a companion plasma machine for cutting **(see illustration)**.

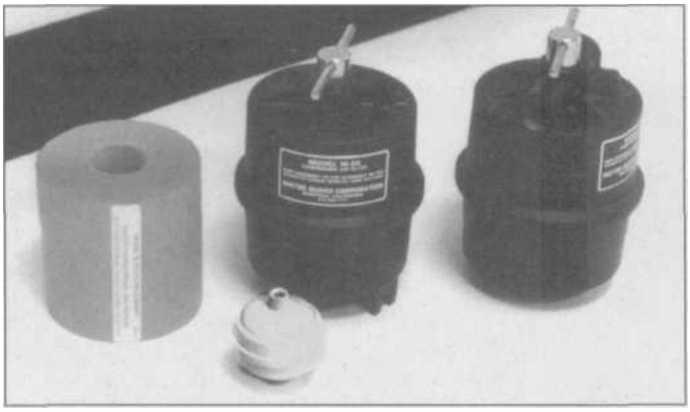

7.31 Daytona MIG offers several air filters, from left: the disposable filter element, a wall-mount filter housing, machine-mount filter housing, and in front a small disposable filter.

POCKET turbo 130
CEBORA
PLASMA

7.32 Given his preferences, today's hot rodder, race-car builder, farmer or fabricator would like to have one of each - a reliable, easy-to-use MIG welder, and a safe, clean and equally easy-to-use plasma cutter.

8 Safety & shop equipment

When it comes to equipment for working with metals, there's a lot more required than just a means of cutting metal and fusion-joining it. Many automotive enthusiasts are "tool addicts" to begin with and welding just adds another dimension to the want list.

Other than your welding/cutting equipment itself, the two main requirements for a good environment are the proper safety gear, particularly the goggles or helmet, and a clean, safe, well-designed welding work area or table. As you'll see, there's a wide variety of helmets available. Choose one that you are comfortable with and that meets your budget. You may have gotten a basic helmet with your welding machine, and you can use that until you save up for one of the more advanced models. Other personal safety equipment would include steel-toed industrial safety shoes (available at most shoe stores), good leather gloves, fire-resistant long-sleeve shirts or, if you are arc-welding, a leather welding jacket and/or apron.

A clean, safe welding table setup is of paramount importance. While you are learning, there will be a lot more sparks and slag generated than when you are advanced, and your welding table should be heavy steel, positioned well away from all stored

LEATHERS
EYEWEAR
LENSES
GOGGLES
HELMETS
TILLMAN

8.1 You'll find most of your welding safety gear and clothing at your local welding supply store, such as helmets, leather protective wear, gloves and safety shields or goggles.

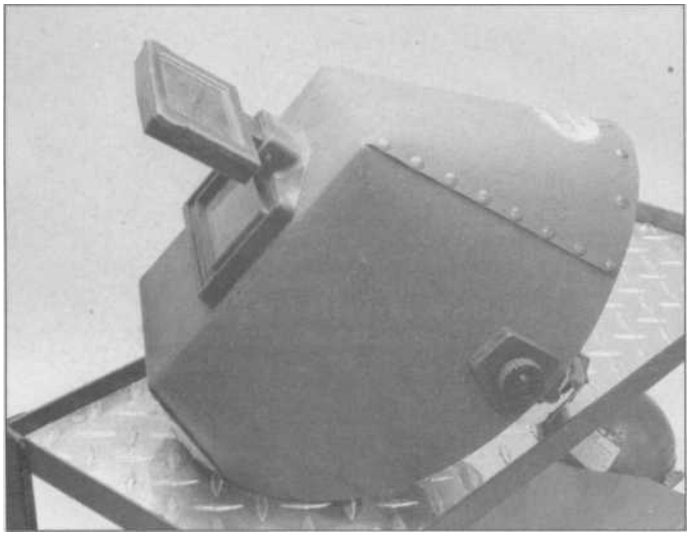

8.2 This Hunstman welding helmet offers full face and chin protection, and features a flip up welding lens. When you are setting up to weld, the helmet can be down on your head and you look through the clear lens, then flip down the dark lens just before welding. Some weldors find these better than fixed lenses where you have to shake your head to flip the whole helmet down.

L-TEC
LIFTORAMA

8.3 This L-TEC helmet from ESAB, called the Liftorama, features a lift-up safety lens, and underneath a shatter-proof clear lens that can be used when grinding a weld.

L-TEC
WELDORAMA

8.4 Another L-TEC helmet, the Weldorama offers a clear sight plate above the welding lens, so the weldor can keep the helmet down, and then just shift his eyes to the dark lens when he is ready to weld.

flammables and with a clear area underneath for any sparks or molten slag to be safely contained without having them scooting across the shop or lodging on top of your shoes. For metal cutting, you need a good "grate" setup as pictured in Chapter 3. You will also need some firebricks, which are helpful to practice welds on because the work won't stick to them. Do not just cover your old wooden work bench with a piece of sheet metal and call it a welding table. Invest in a piece of plate from a scrap yard. When you put a lot of heat into welding projects on a tin-covered wooden surface, enough heat can go through the metal to start the wood smoldering, perhaps even starting a fire after you have left the shop. Also, you must have proper fire extinguishers on hand, preferably an ABC type or halon, that can put out all kinds of fires.

Much of the work of sizing and joining metals for home/shop projects will involve tools you already have, such as a drill press, grinders and sanders, clamps, measuring equipment, squares and levels. A common hacksaw can be used to sever lengths of material but can be tedious on heavier-wall stock. Sheet metal can be cut with aircraft tin-snips, manual or electric nibblers, or a plasma cutter.

Many of the tools shown here are special-purpose tools that you may not have already, and which perform special functions that relate to improving welding work or holding materials to be welded. All of them are useful when the time comes. We all wish we could just put one of each in our home garage, but you may have to collect them over the years as your projects get more involved.

We also show here some tools and equipment that were made by fabricators

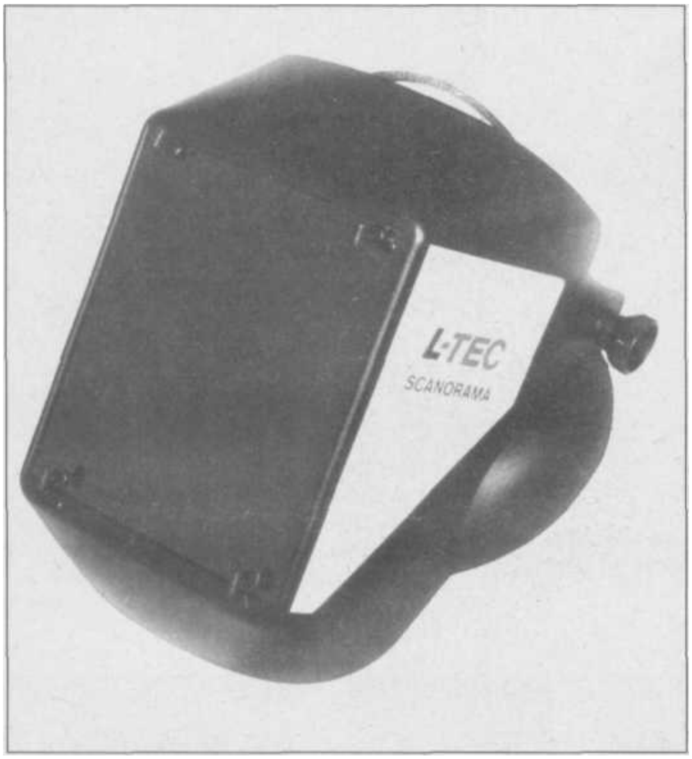

8.5 The Scanorama helmet was designed for MIG and TIG-welding applications, with a large shade 5 filter disc surrounding the welding lens. You can see most details through this area, but your skin is still protected from radiation. You look through the darker portion when welding.

AUTORAMA 2000

8.6 This Darth Vaderish design from ESAB is the Autorama 2000, which one of a modern breed of welding helmet with an automatically-darkening lens. The lens is powered by solar cells, and turns dark as soon as the welding arc begins. These are considered by many weldors to be much easier to use than flipping a helmet up and down to set up each weld.

DQF
SPEED
SHADE

8.7 Daytona MIG offers a variety of modern safety helmets, and a special, clip-on lens (foreground) that attaches to your MIG gun. It allows you to set up a weld, then look through the lens when making spot-welds without the bother of a full helmet.

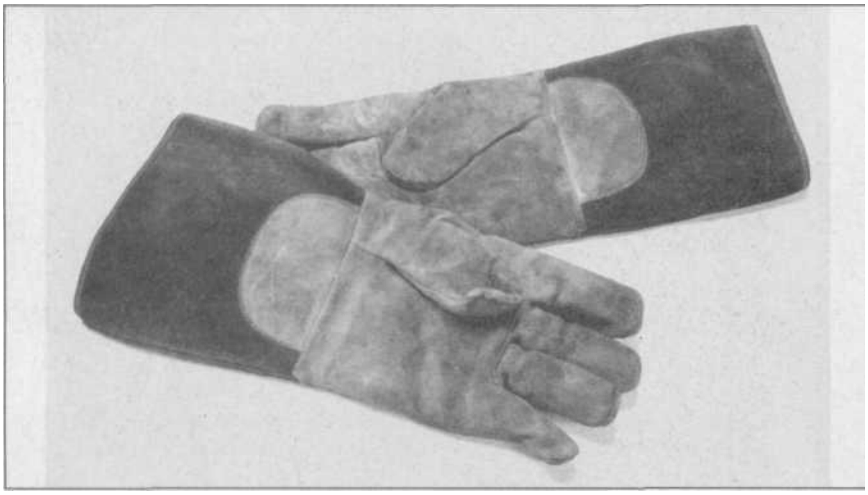

8.8 A good pair of leather welding gloves will last you for many years, but if you are using oxy-acetylene equipment, keep an extra-clean pair for use with this equipment. A spurt of oxygen on a dirty glove could cause combustion, so don't use these good gloves for handling hot or dirty parts.

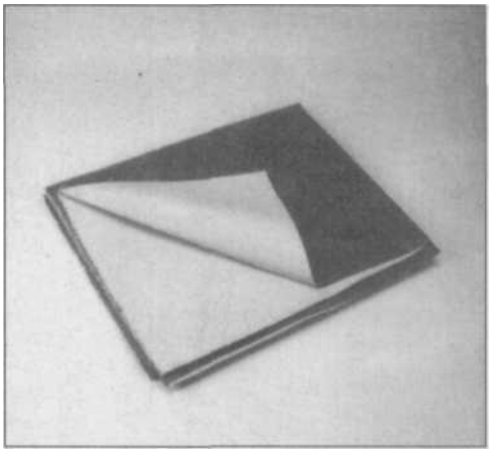

8.9 From HTP is this Super Spark Guard welding blanket. It can take up 2300° F, and can be used when working on a car to protect the window glass, dash or interior. It measures 37 inches by 50 inches.

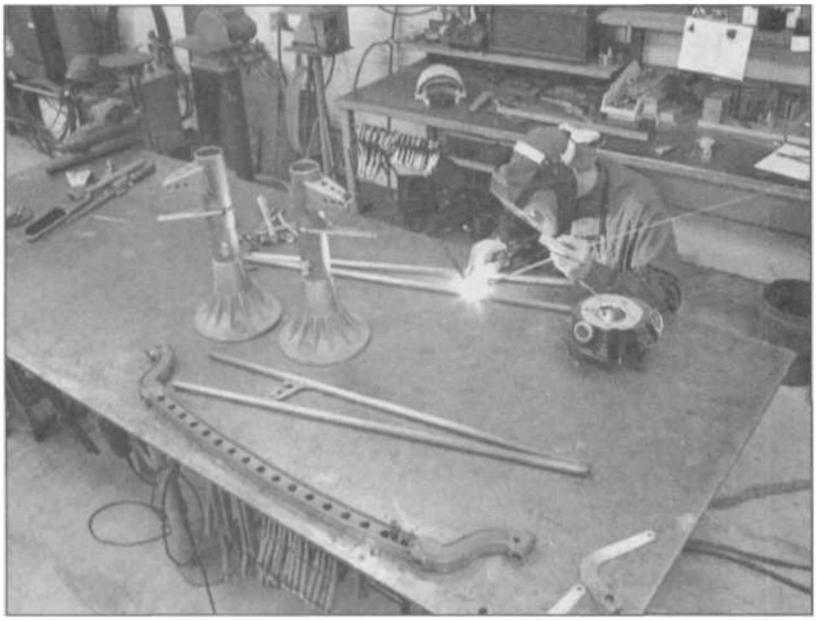

8.10 A clean, safe welding table area promotes good work. Your home shop won't need a steel-plate table this big, but note how uncluttered and organized things are here in a professional shop. A rack under the table holds all the C-clamps.

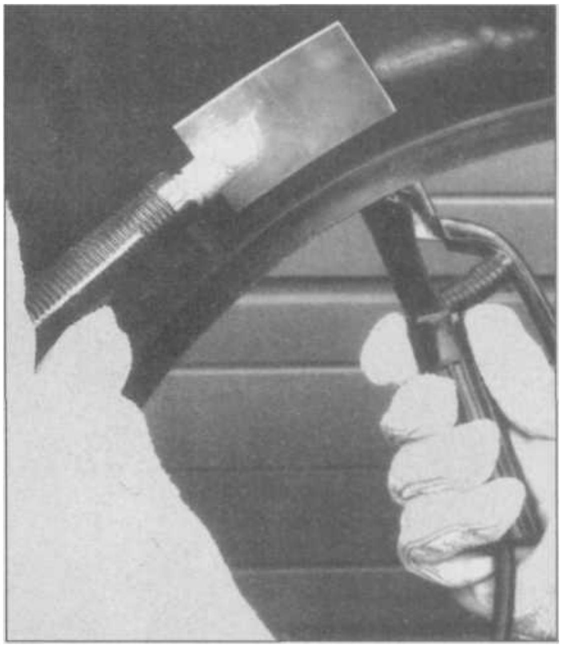

8.11 A very useful tool when welding on thin materials such as newer-car body metal is a copper "spoon." The heat-absorbing spoon can be used to back up butt welds or when filling holes in sheet metal. The welds will not stick to the copper.

and hot rodders for their own use. These are typical of projects you can do once you have practiced your welding skills, and some represent considerable involvement, though on the part of someone who has already "been there."

Sources for the commercial products in this and other chapters can be found in the Sourcelist at the end of this manual.

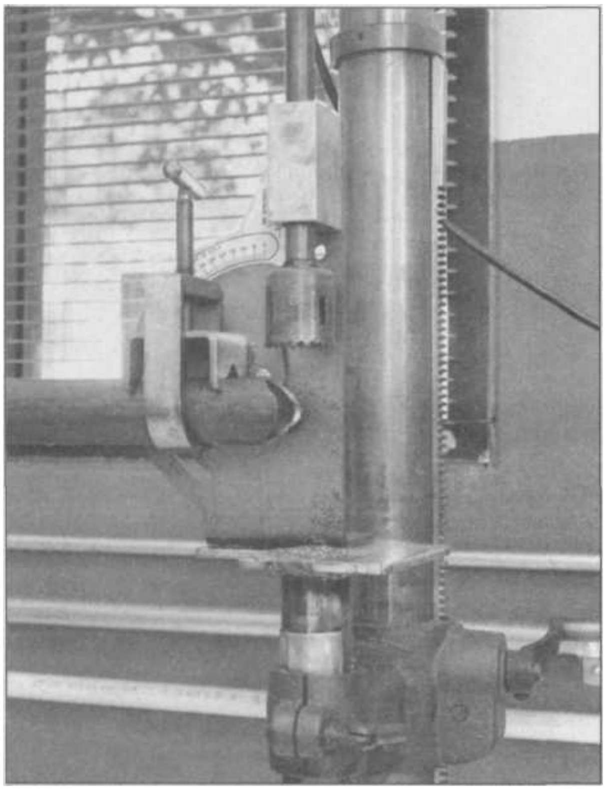

8.12 If you get into a race-car project, you might invest in one of these tubing notchers. Available from Eastwood, Williams Low-Buck Tools, Speedway Motors and other sources. It fits in your drill press and allows precise fish-mouthing of tubing via a captive hole saw.

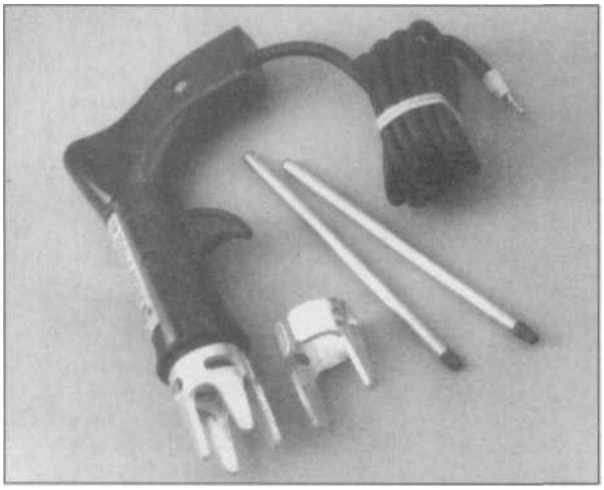

8.15 This unique spot-welding gun for sheet-metal work has a power lead that clamps into the electrode holder of your arc welder. Use the trigger to make the electrode touch the work, pull up for 5 seconds, then pull all the way to break the arc. Get it from Eastwood.

8.16 One of the handiest shop tools we have ever used is the Milwaukee "Porta-Band", which as the name implies is a portable bandsaw. It cuts fast and straight and is very helpful in cutting tubing at angles and other chores associated with welding projects.

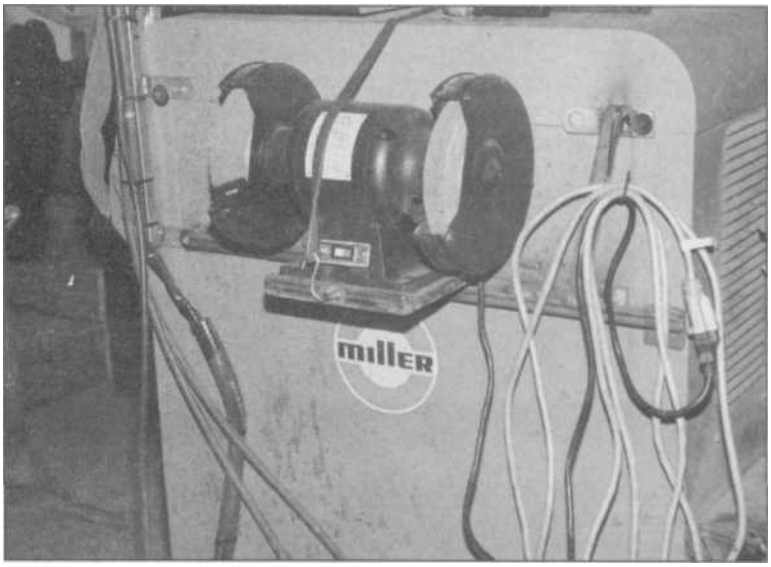

8.13 This welding shop has a dedicated grinder for sharpening tungstens mounted right on the side of the TIG machine. The rubber bungee cord kills the vibration noise when the grinder is slowing down.

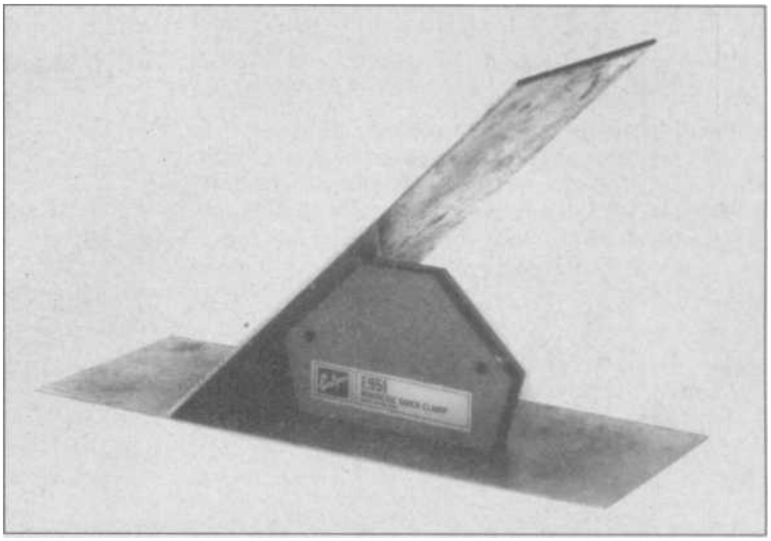

8.14 Eastwood offers this magnetic quick-clamp which functions to hold parts in alignment at various angles for tack-welding.

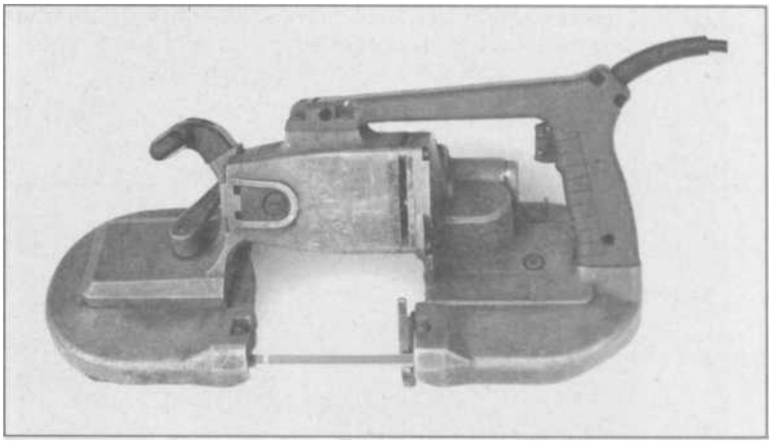

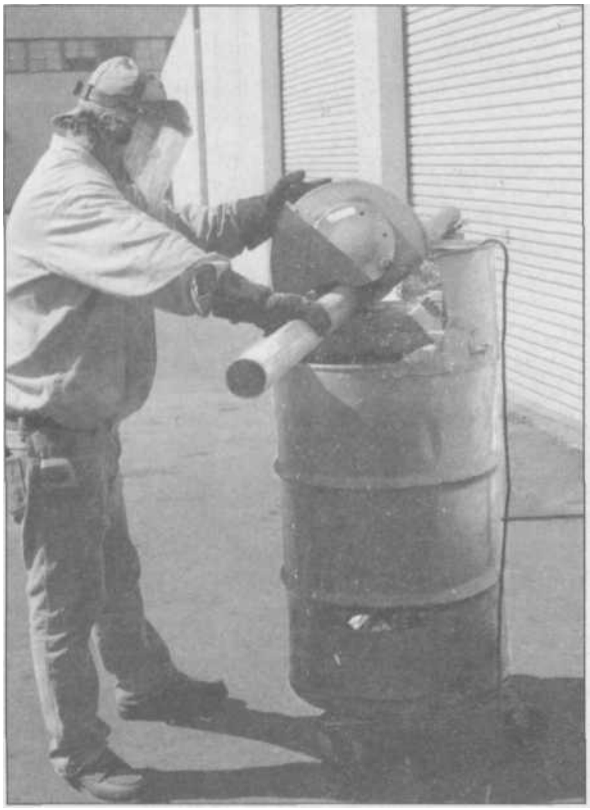

8.17 A "Radiac" or cutoff saw with an abrasive blade is very fast and secure when cutting large amounts of tubing. The work is held in a moveable vise so angle-cuts can be made. This one is cleverly mounted on a 55-gallon drum to capture all the sparks and debris.

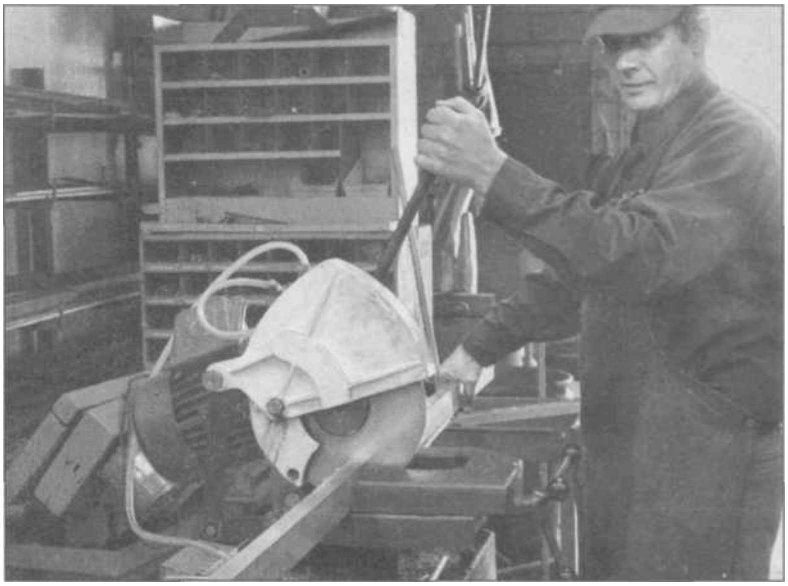

8.18 Other kinds of metal-cutting power saws include this type, which uses a bi-metallic toothed blade, here being used to cut clean through aluminum tubing.

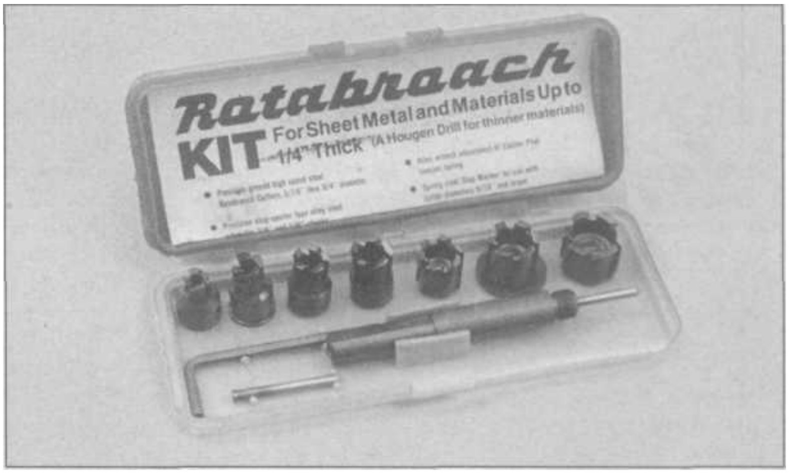
Rotabroach
KIT For Sheet Metal and Materials Up to 1/4" Thick (A Hougen Drill for thinner materials)

8.19 In auto bodywork, a tool like this can be used in your electric drill to cut out factory spot welds to remove panels for replacement.

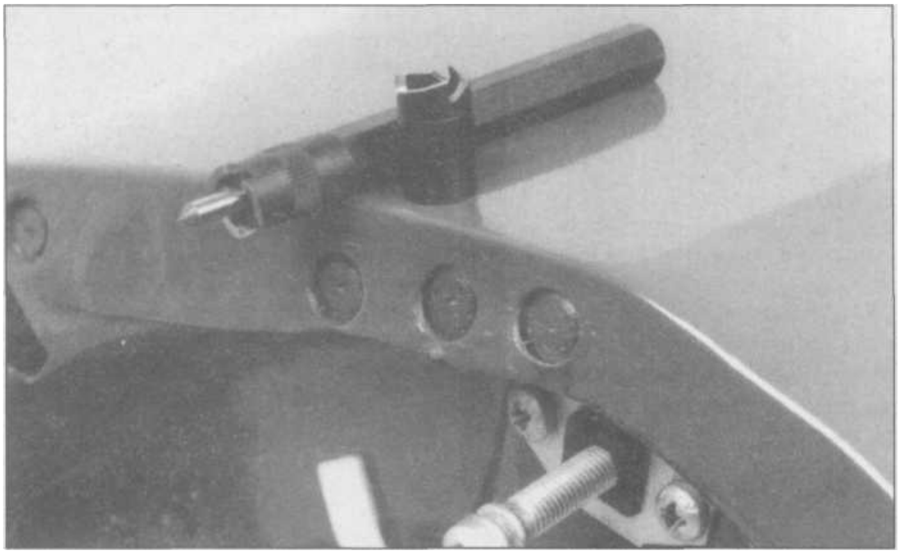

8.20 For Eastwood offers a similar tool for removing spot welds which features a hex-sided drive end, so it can't slip in your drill chuck.

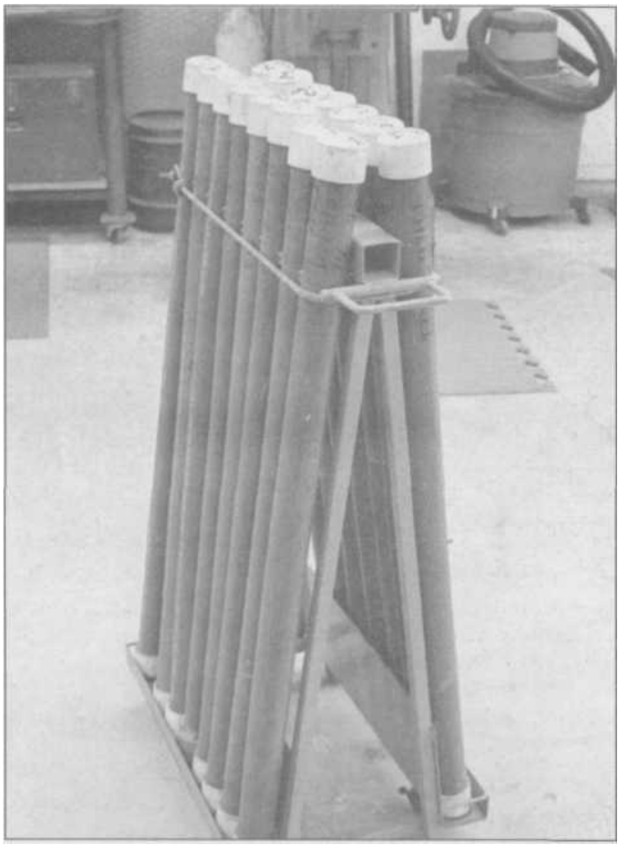

8.21 Weldor Bill Maguire made up this electrode rack for his TIG machine from a rolling platform holding 15 plastic plumbing pipes with caps, each marked for the size and type of electrode. A bungee cord holds them in the rack. Tubes keep the rods clean and dry.

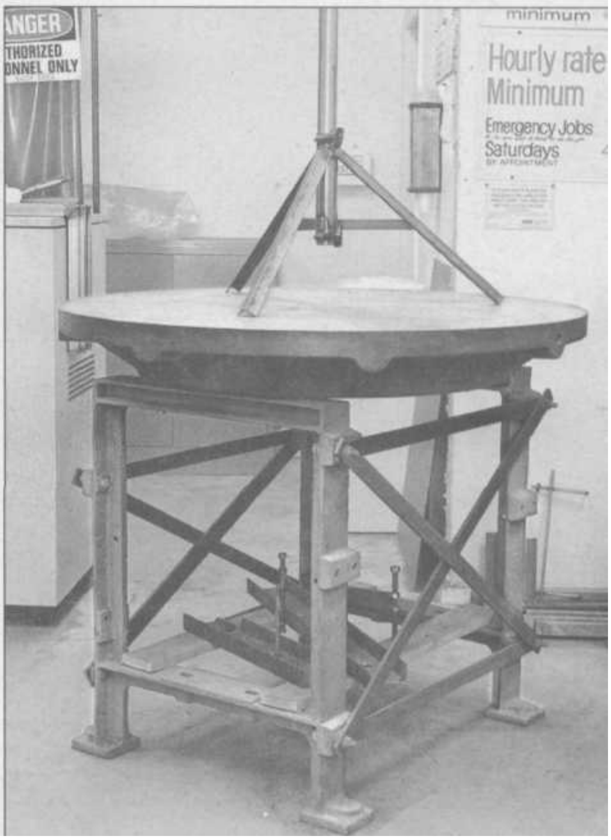

8.22 You never know when you'll find something that comes in handy in the shop. This huge and hefty round plate was found at a metal yard and bought at the per-pound price. It's dead flat and cannot warp, and has mounting holes in it for clamping work to it.

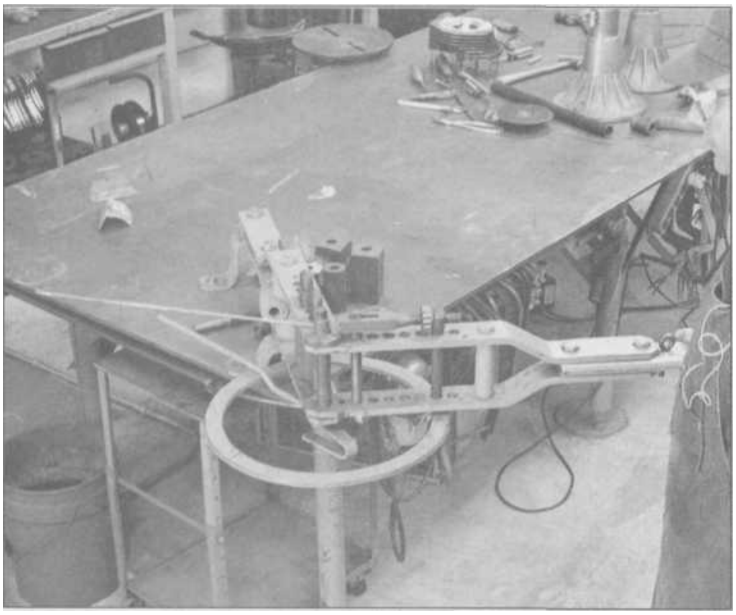

8.23 If you plan to do any ornamental iron work or make a lot of brackets, an ironworker like this is very handy, It can bend solid round, square or hex stock, as well as bend round tubing and flat stock into repeatable shapes.

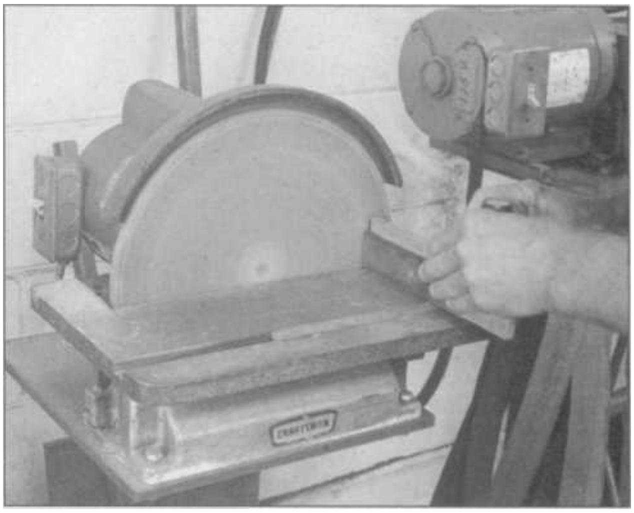

8.24 A disc sander with at least 1/2 horsepower is very handy around the shop for squaring edges of flamecut parts, and taking sharp edges and corners off work.

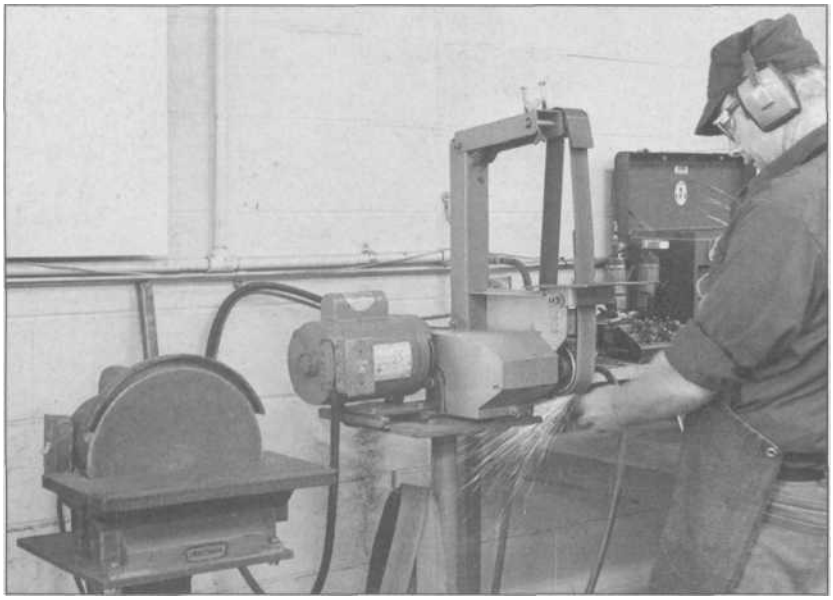

8.25 Ideally, you would have both a disc sander and belt sander for a variety of metal finishing needs. There are combination units, but beware of weak-powered sanders. When you put the rough edge of a 3/8-inch plate up there to sand, they stall easily.

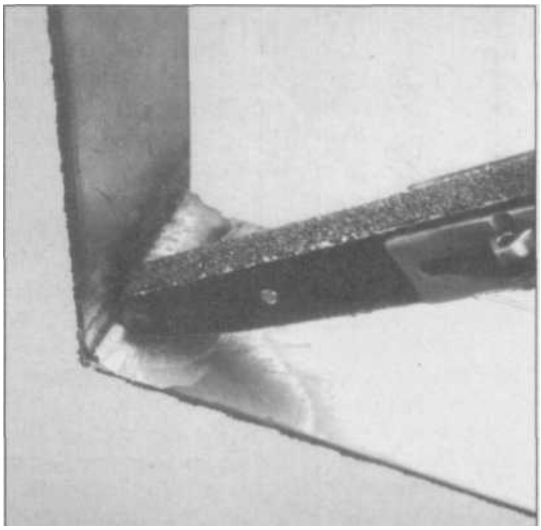

8.26 HTP is marketing this clever belt-sander that can get deep into corners and places where regular sanders and grinders can't reach. The belt guide rotates 160 degrees, and the sander uses 3/8-inch belts.

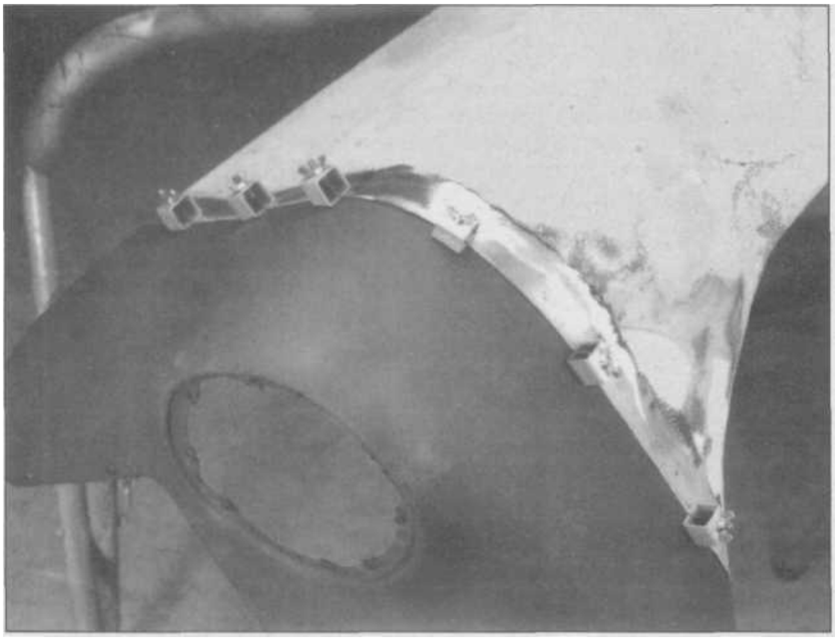

8.27 Something that is always a problem in body work when replacing panels is getting parts aligned properly before making a butt-weld. Eastwoods Intergrip clamps space the two parts .040-inch apart for full penetration. After tacking the panels, you can disassemble the clamps.

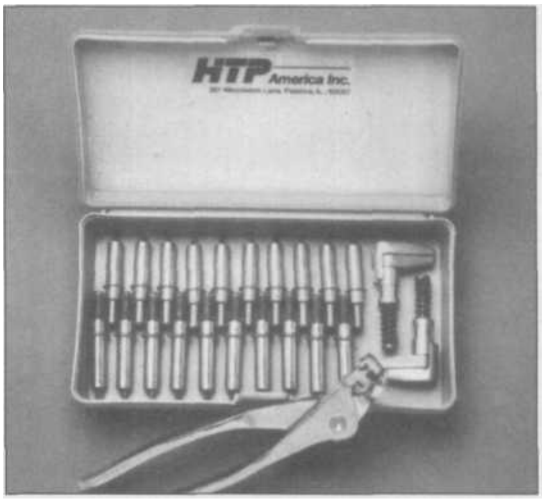

8.28 Nothing beats the original Cleco clamps and pliers for test-assembling sheet-metal parts. Originally developed for aircraft use, these operate like reusable, temporary rivets. In blind holes you insert the fastener with the pliers and the two parts are held together, but come apart easily with the pliers again. HTP offers this set.

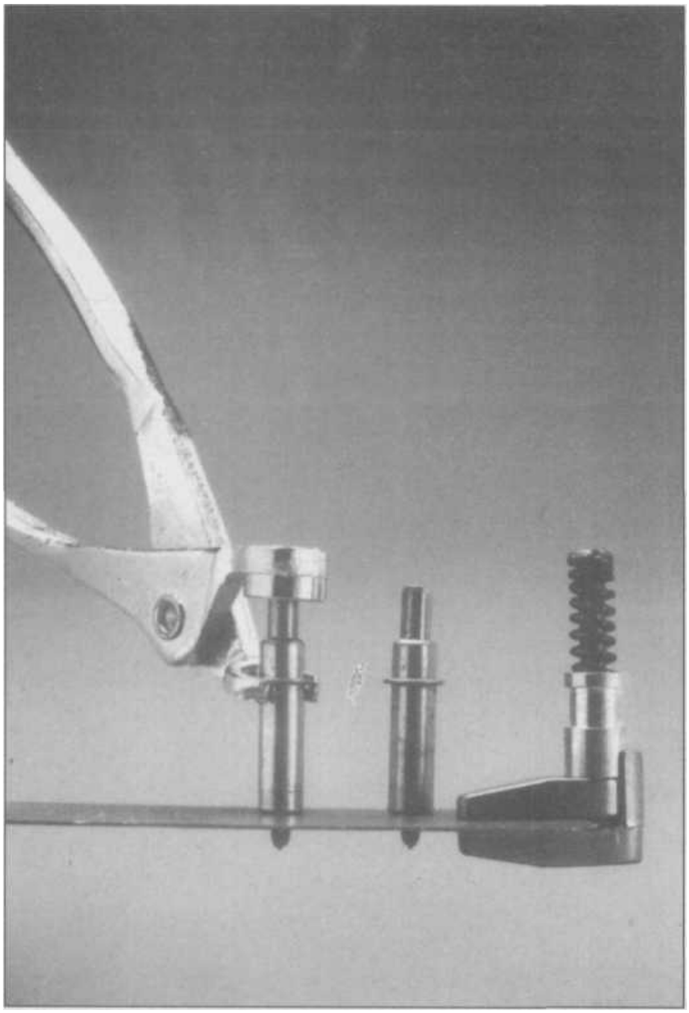

8.29 You can see here how the pliers extend the tip of the Cleco fastener through a set of holes. When the tool is released, the Cleco grips the parts. Also available are Cleco mini-clamps (right) that exert a temporary hold while you tack-weld small parts together.

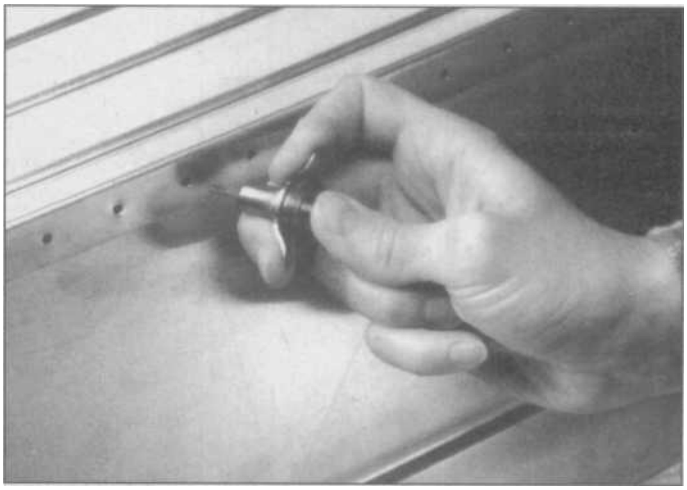

8.30 For small jobs, Eastwood offers these little Cleco-type panel holders that you operate with your fingers. No tool is needed. They have a range of up to 1/4-inch for thickness and fit a 1/8-inch hole.

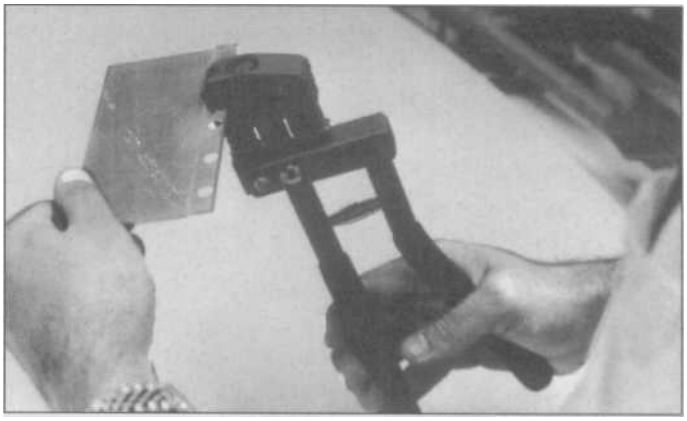

8.31 When you want to spot-weld sheet-metal panels that didn't have them originally, like new parts or repair panels you have made , a manual flange punch like this is a real help. This one is from Daytona MIG.

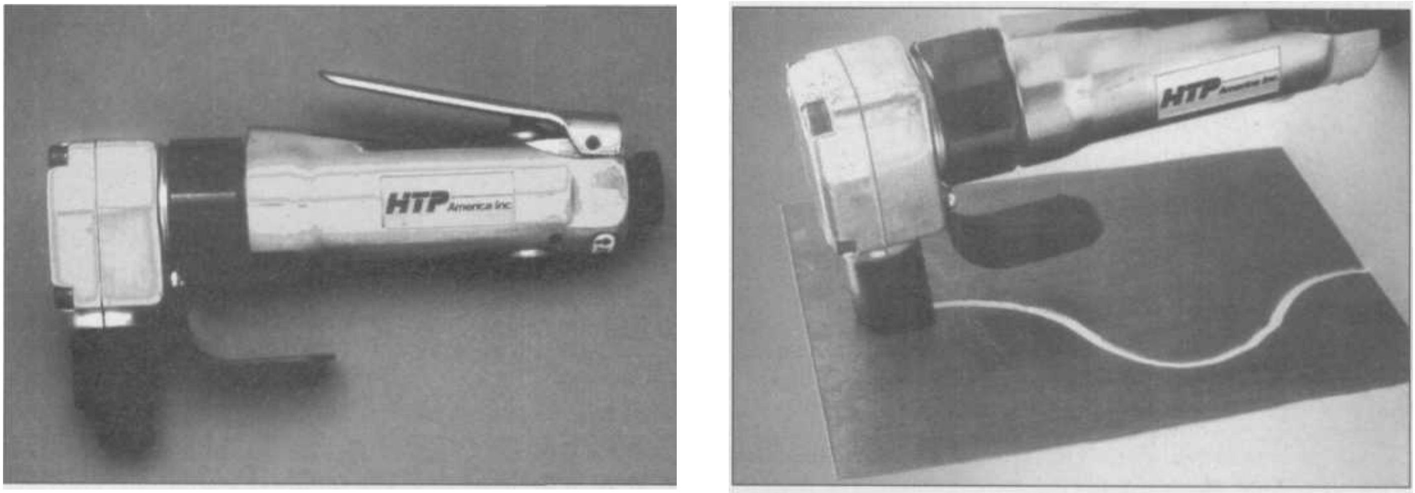

8.32 An air-powered "panibbler" is a fast, clean way to cut sheet metal up to 3/64-inch in steel or 5/64 in aluminum. It will follow a pattern or cutline very closely, and you control the speed with the throttle. This one is from HTP.

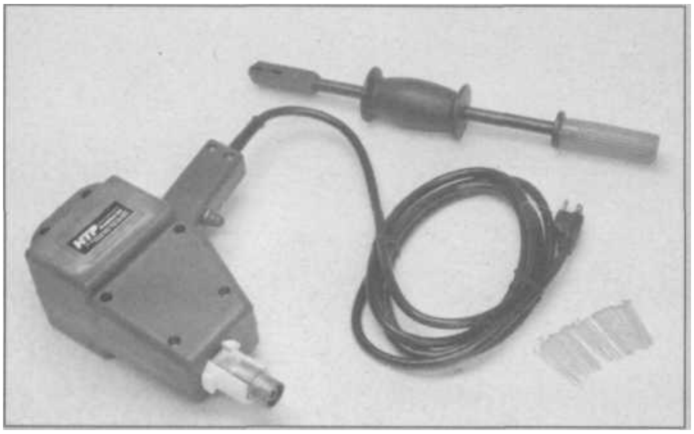

8.33 The HTP stud setter is a great boon to bodymen, who used to have to pull out creases and dents by drilling holes and using slidehammers that distorted the metal even further. This 110V tool saves considerable time and labor.

8.34 Small-headed copper-plated "studs" are inserted into the magnetic head of the stud setter.

8.35 On a dented area that has been ground to clean metal, the tool is pushed in contact with the metal and the trigger is pulled for a half-second.

8.36 The suds are spot-welded in place, and a special slide hammer is used to pull out the dent or crease. Studs hold 500 pounds of pull.

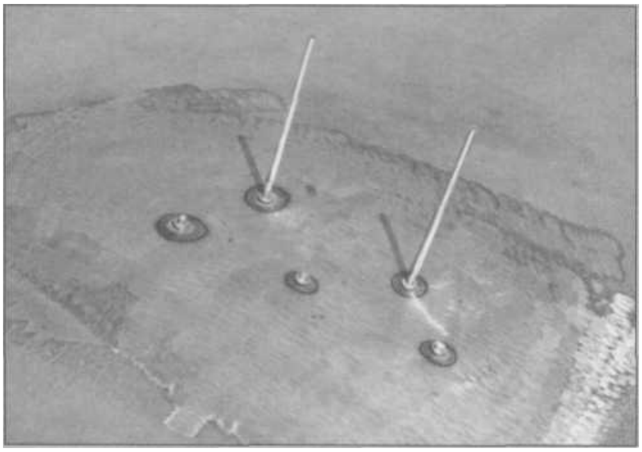

8.37 On this dent, the first three studs pulled out the bulk of the damage, and two more are in place to pull the last bit out. Note how the first three studs have been clipped off with cutting pliers.

8.38 After the pulling, the stud heads are easily ground off without distorting the metal, and there are no holes left to fill in, as with conventional methods.

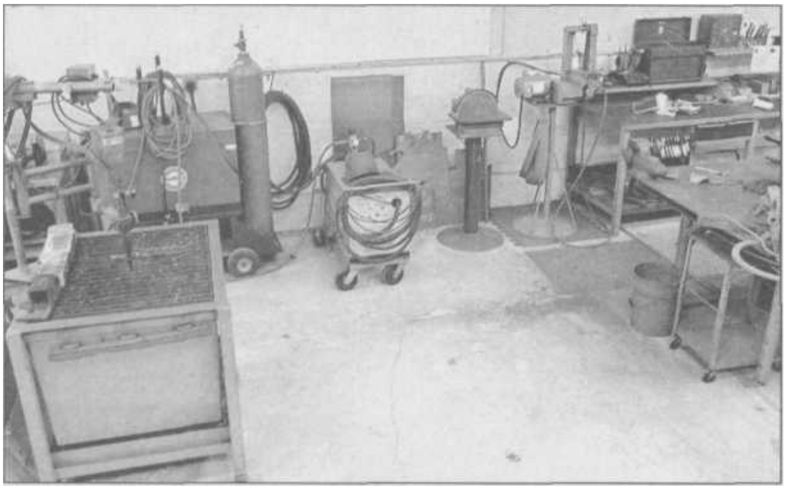

8.39 Visit most street rod shops, fabrication shops and race-car facilities, and you'll find that a lot of the shop equipment they have is homemade, and these are the kinds of things you'll be able to do once you have welding and cutting equipment.

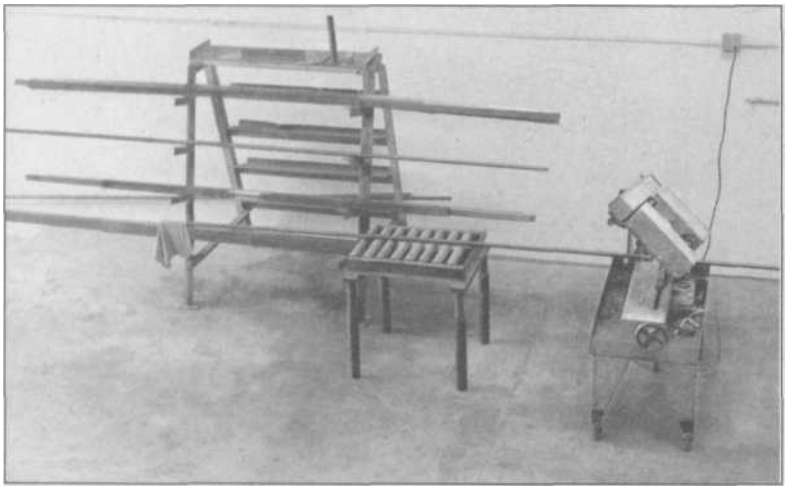

8.40 A simple steel materials rack like this is great for keeping tubing, bar stock and flat stock out of your way, and is easily constructed. In front of this one is an electric, automatic bandsaw great for making unattended cuts. This one was purchased inexpensively at a tool auction. Watch your newspapers for industrial auctions and shop foreclosures.

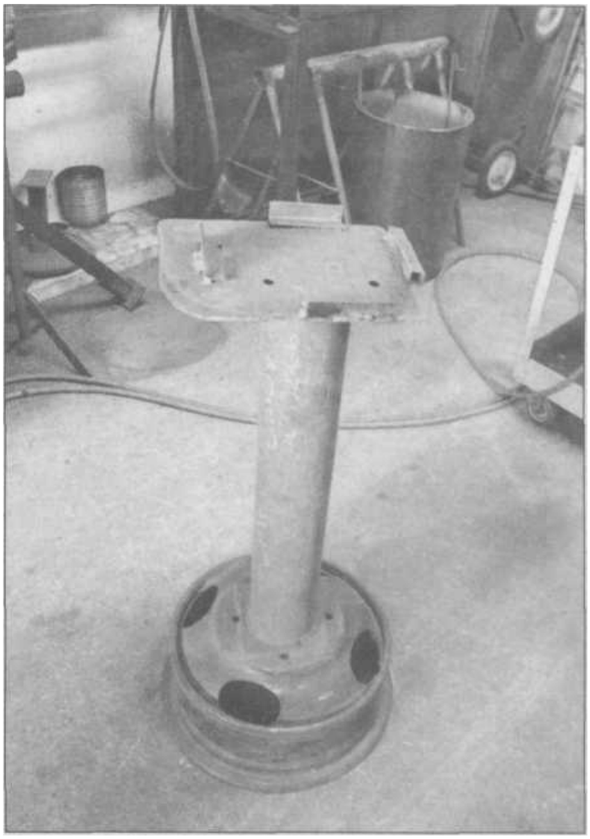

8.41 An old truck rim, a length of round pipe and a piece of scrap plate are combined with welding to make a perfect small welding table or mount for a grinder or sander.

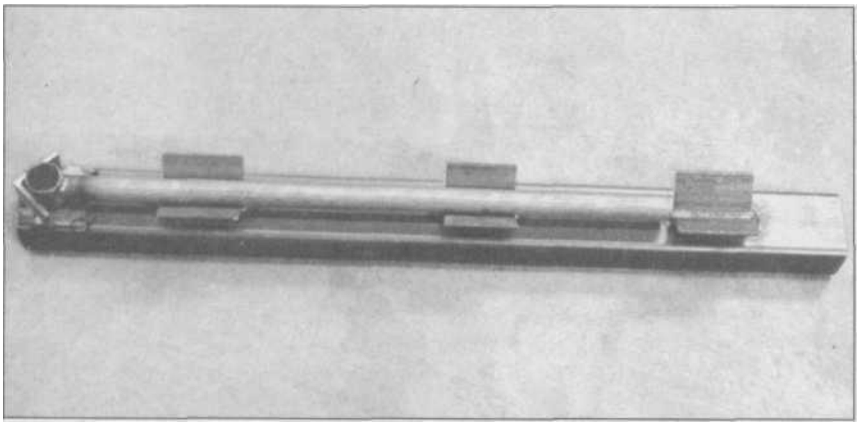

8.42 Leftover pieces of tubing and other stock should be saved in a barrel. You never know when you need a short piece. This fixture for making up a street rod four-link rod is made from short pieces of angle-iron and a length of rectangular tubing.

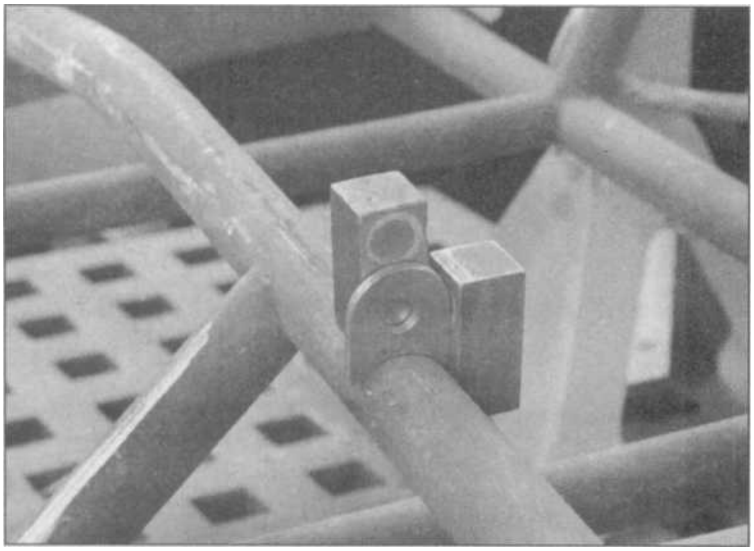

8.43 Ideal for holding small brackets onto tubing for tacking are these magnetic blocks, available at most welding supply stores.

8.45 Another type of angle fixture from HTP is this type, which clamps the two parts at a perfect 90 degree angle for tack-welding.

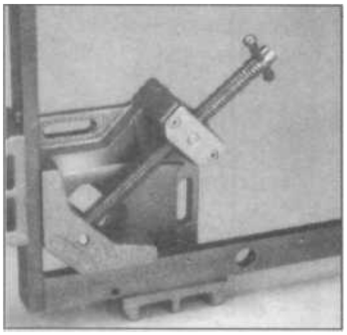

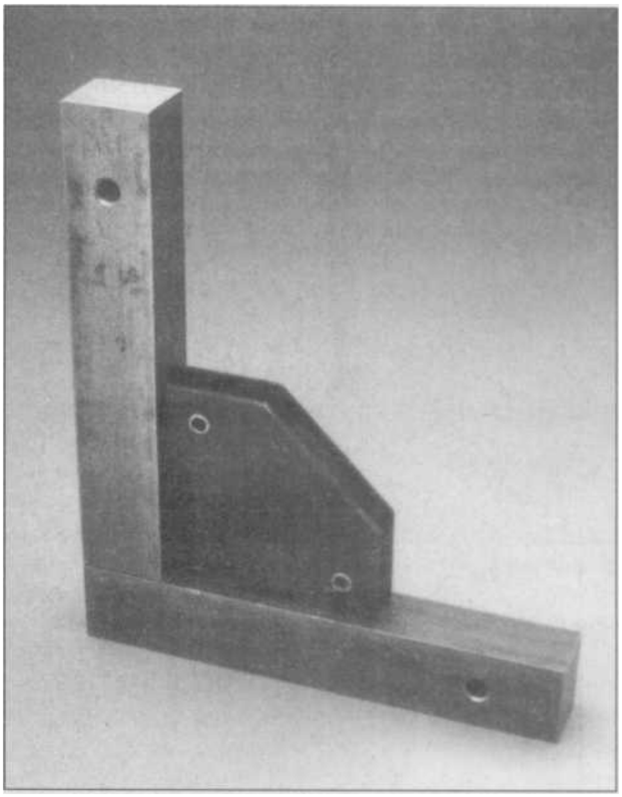

8.44 HTP and other companies offer these strong magnetic holders that can secure two pieces of tubing for angles of 45 degrees, 90 degrees, and 135 degrees. Tack the parts then move the magnets to the next joint.

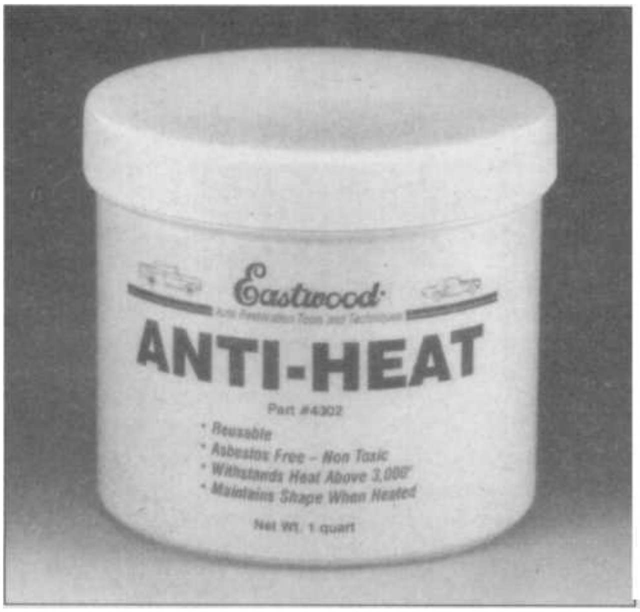
Eastwood
Auto Restoration Tools and Techniques
ANTI-HEAT
Part #4302
• Reusable
• Asbestos Free – Non Toxic
• Withstands Heat Above 3,000°
• Maintains Shape When Heated
Net Wt. 1 quart

8.46 The Eastwood anti-heat is a putty-like paste that can be formed around heat-sensitive parts when you are welding near them. The paste absorbs heat and can be used to protect auto glass when welding near it.

8.47 Daytona MIG sells this handy spot-sandblaster. It is useful for cleaning paint and rust away from a spot your have to weld, like a recess where a regular sander won't reach. Three-quarters of the blasting media is returned to the gun, so it won't make your shop a "sandpit."

9

Building a utility trailer

One of the most common home/shop projects for welding equipment is a trailer of some kind, Which can be used for carrying a race car or off-road vehicle or On a smaller scale, can function as a utility trailer. Each of these is easily within the realm of the hobbyist, though a trailer of any size should not be your *first* welding project. Hone your cutting, fitting and welding skills on other, less critical projects first, like a metals storage rack, a wrought-iron front gate, a fireplace grate, barbecue grille or a rolling cart for your welding equipment. Frankly, the latter is usually the first welding project after the practice sessions. Many are the accomplished weldors we have met who say they welded up their oxy-acetylene equipment cart some 20 years ago and are proud to be still using it.

Since the trailer project is something to be used on public roads and at highway speeds, you want your welding quality to be good before you get started because the welds here have to be strong, not just good-looking. The basic construction details of any trailer are fairly similar, whatever the size. You assemble a rectangle of steel tubing or angle-iron, attach a sus-

9.1 A 4x8-foot utility trailer like this is a good project once you have perfected your welding and fitting techniques. Once you have a trailer like this, you'll wonder how you ever got along without one. It can be used for hauling camping gear, lumber, yard trimmings, a go-kart or furniture.

9.2 Find a good trailer supply store in your area, where you can browse for suspension parts, wiring, lights and hitch components.

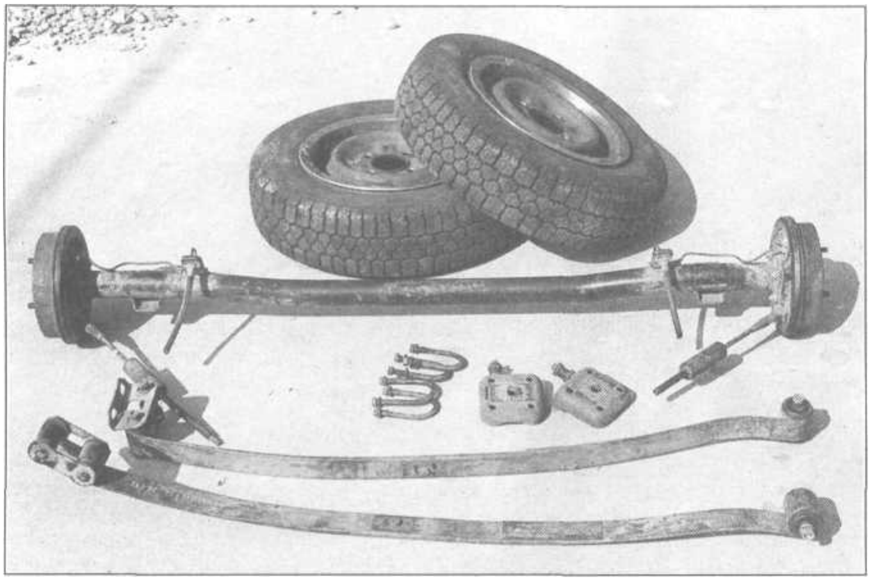

9.3 We purchased this Chevrolet Sprint solid rear axle and springs at a do-it-yourself wrecking yard very inexpensively. New trailer axles and springs can be expensive, but you can often find used ones at swap meets, or you can buy an old, small boat trailer that you can rework into a utility trailer.

pension and axle system, add a floor and sides, taillights, safety chain and a hitch in the front. Most smaller trailers, from 4x8-foot size and down, can be constructed of 2x2 tubing, or 2x2 angle-iron, depending on the hauling requirements. A trailer such as we are illustrating here is capable of carrying motorcycles, a go-kart, personal watercraft, garden tractor, camping equipment, building materials, firewood or landscaping dirt. With a few simple poles and some canvas, you could actually use it to camp in, like a homemade pop-up camper. We built this one of 2x3 tubing and 2x2 tubing of 1/8-inch wall thickness, which is somewhat overbuilt, but these were the materials that were accessible, and the cost of building such a trailer is entirely dependent on your materials. If you were building a car-hauling trailer, you would use two axles, preferably with brakes, and use diamond-plate steel for the floor, with some very sturdy ramps.

If you can get much of the steel from the scrap or remnant pile at your local metal source, scrounge used plywood or surplus diamond-plate steel and get your axle and suspension from a donor car at a wrecking yard, such a trailer can be built for 1/3 to 1/2 the cost of a newly purchased trailer of the same size. You'll need some kind of metal cutting equipment beyond the basic hacksaw or your arm is going to get tired. There's a lot of metal to cut to build even a simple trailer. You could lay out your plan on the driveway or garage floor, mark all your materials and bring them to a shop for cutting if you have to, then bring them back home to weld. Welding such a project can be done with any of the equipment we have discussed in this book, but it goes together fairly quickly with either arc-welding or MIG-welding.

We used inexpensive-grade plywood for our trailer sides, which we made two feet high, and a better grade of 3/4-inch plywood for the floor bolted to the frame with carriage bolts so the floor would remain smooth enough for easy rolling of a furniture dolly into the trailer. Once you have built such a trailer, it's like having a new truck. Everyone you know will want to

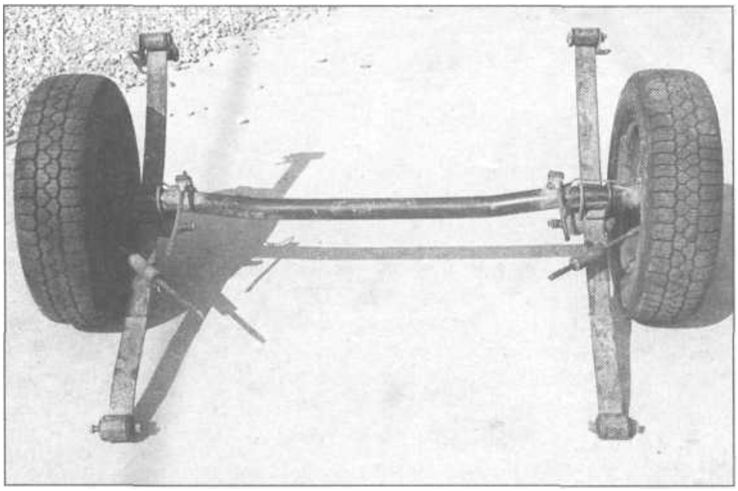

9.4 Bolt your springs, U-bolts, spring plates and shackles together, making sure the springs are square to the axle centerline and measuring the same distance apart front and rear.

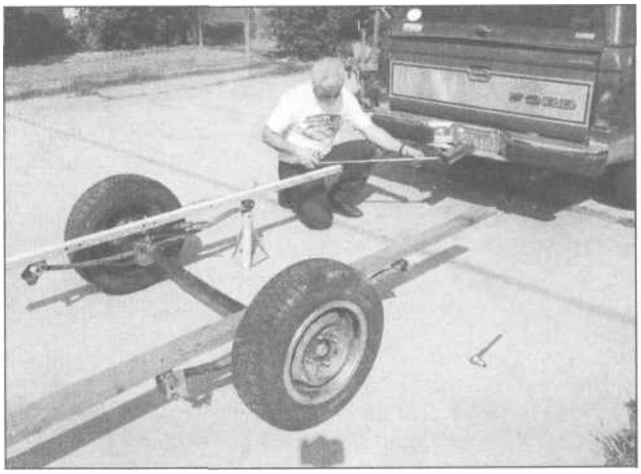

9.5 Attach a hitch to the ball on your tow vehicle, and mock up the assembly, using 2x4s to simulate the rails. This should give you an idea of what size and layout of trailer you want, and where to locate the axle and wheels. Keep more trailer frame in *front* of the axle than behind, to insure you have enough tongue weight.

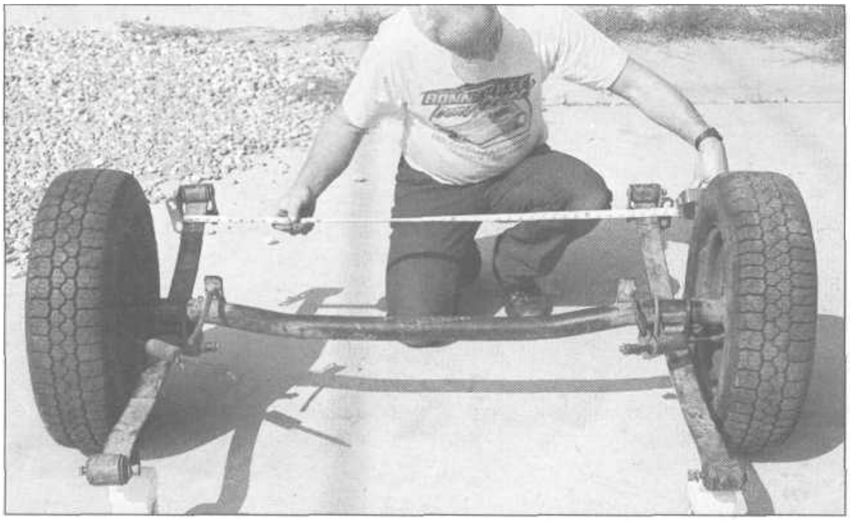

9.6 In our case, we determined that the Sprint axle was too narrow for our purposes, since we wanted a full 4x8-foot sheet of plywood as a floor. This meant widening our tubular axle 7 inches.

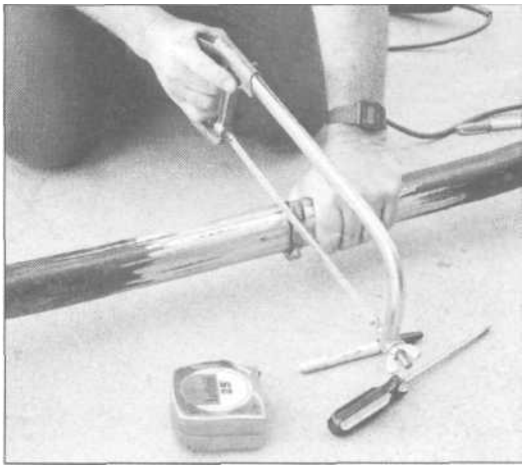

9.7 Grind the paint from the center of the axle, mark a straight centerline across the axle for lining-up purposes later on and cut the axle with a hacksaw. A hose clamp around the axle is used as a saw guide to keep the cut straight.

borrow it to move something. Our tailgate was made from a frame of one-inch-square steel tubing, .063-inch-wall, and we tack-welded to it a section of steel mesh such as is used for barbecue grilles. The hinged tailgate allows wind to flow through the tailgate on the road, reducing fuel consumption from the drag of a solid tailgate.

For our project, we used a solid rear axle and suspension from a compact FWD Chevrolet Sprint, which has single-leaf rear springs. We found later on that we had to add helper springs to accommodate heavier loads with the trailer. The 4x8-foot size allows for a fairly heavy load. We just bought two more Sprint rear springs at the local do-it-yourself wrecking yard, cut the eyes off each end with a friend's plasma cutter and used these as an extra leaf on each side. This increased the load capacity without making

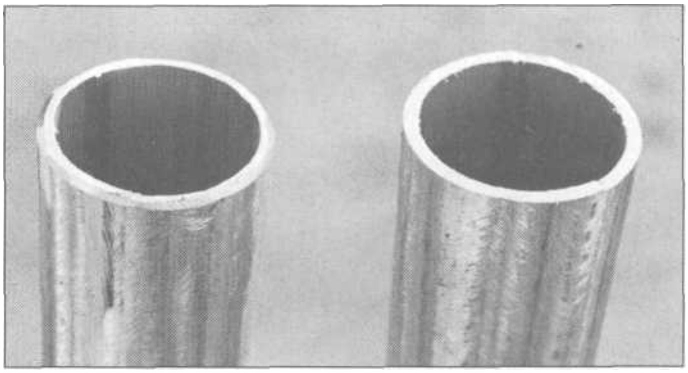

9.8 This axle is heavy-wall tubing, and, since we wanted good weld penetration for strength, the cut ends were beveled with a grinder. Tube at left has been beveled to about half its thickness.

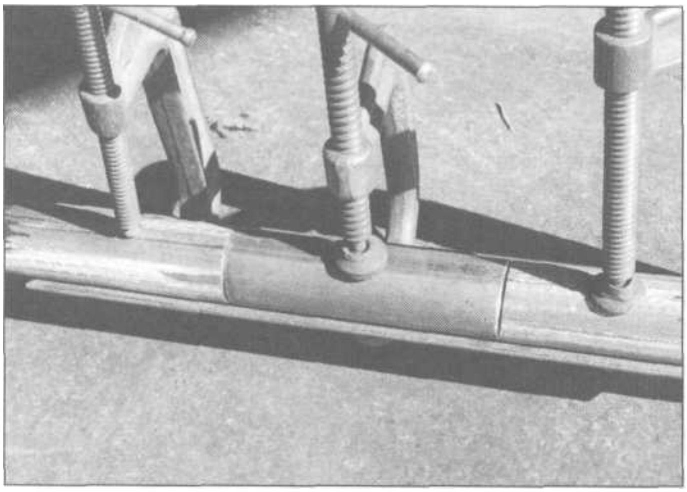

9.9 With the remains of your original centerline in alignment, clamp the axle ends together with a piece of new heavy-wall tubing the same size and keep them all in alignment with a length of wide angle-iron.

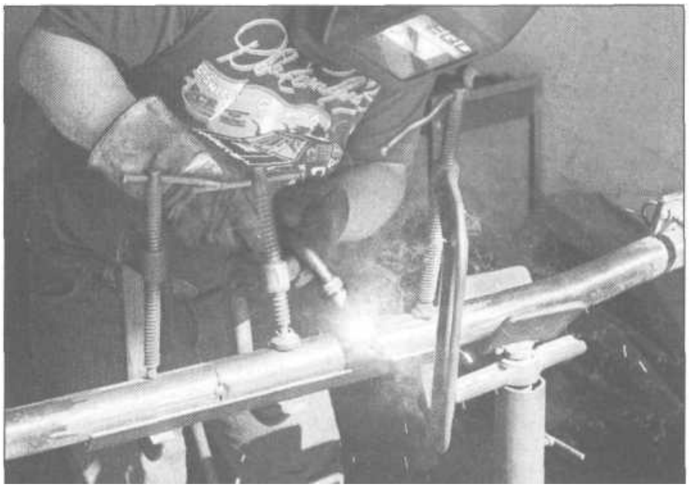

9.10 Here the assembly is tack-welded at several points around each joint with a MIG welder.

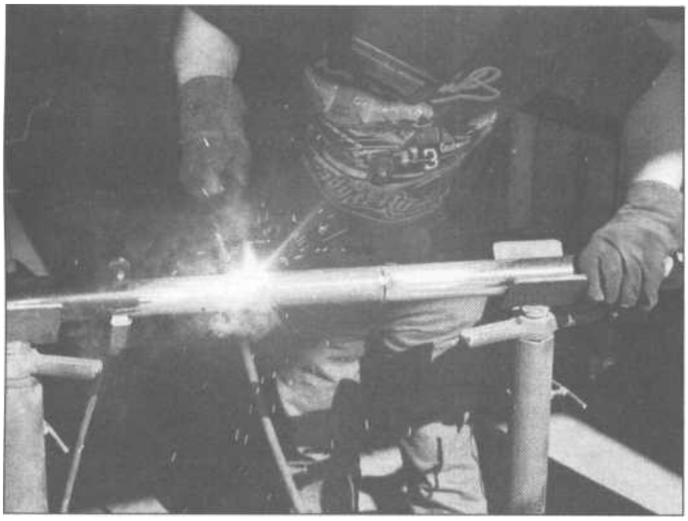

9.11 When the axle is securely tacked at several points, remove the angle-iron and make a strong weld bead all around the axle at each joint. Check for straightness after it has cooled off.

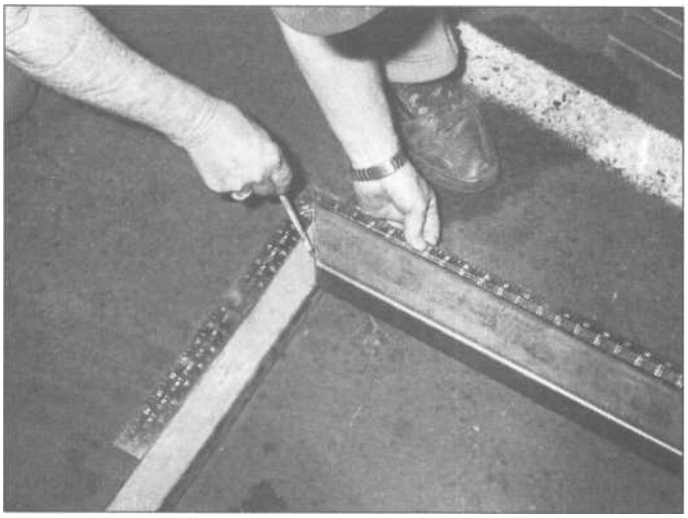

9.13 We used what materials we had, with 2x3-inch tubing for the two main rails and 2x2-inch tubing for the rest. When joining tubing of different sizes, cut one to a standard 45-degree angle (here the 2x3), then lay the next piece in, square it and mark along the second tube for the correct angle to cut it.

9.14 You can cut tubing and angle-stock with a hacksaw, but it is tedious. We borrowed a Milwaukee Port-A-Band hand-held bandsaw that made short, accurate work of fitting our tubing.

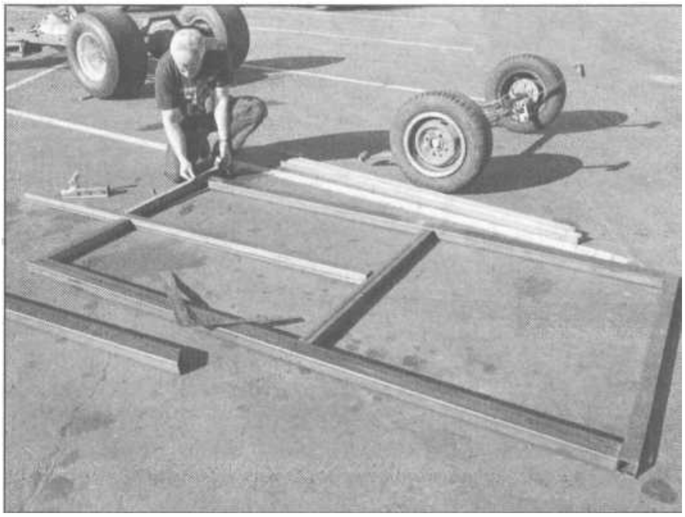

9.12 If your driveway or shop floor is flat enough, lay out your pieces of tubing on the ground and check all your measurements. Cut your joint angles carefully, and, with the pieces laid out, measure diagonally from corner to opposite corner to see if the frame is square. Check this again several times during the tacking and welding phase, too.

it too stiff when unloaded. Since you are using an automotive axle, it comes with brakes and all, and you *could* hook these brakes up to your tow vehicle with parts available from a trailer supply store. For our loads, like firewood, furniture and camping equipment, it didn't seem necessary. You could also buy a regular trailer axle with springs and brakes, but they are rather expensive if purchased new. There's a number of straight rear-axle assemblies available from front-wheel-drive cars in wrecking yards at reasonable prices.

There are few hard-and-fast rules in building a trailer, except that you should design it with the suspension and axle located in the rear 50% of the basic frame, not in the center, to maintain good tongue weight up front. Your tongue weight should be roughly 10% of your total trailer weight statically at ride height. The one thing that will certainly make a trailer handle badly is not having a long

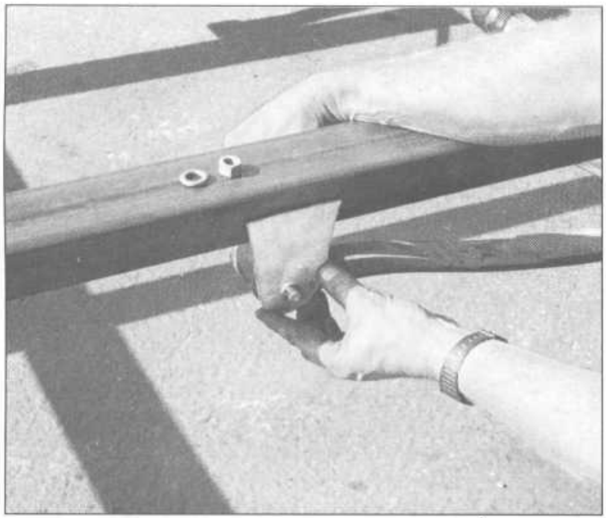

9.20 Put bushings in the front spring eyes and use cardboard to make up a template for four front-spring-eye mounting plates to be welded to the frame

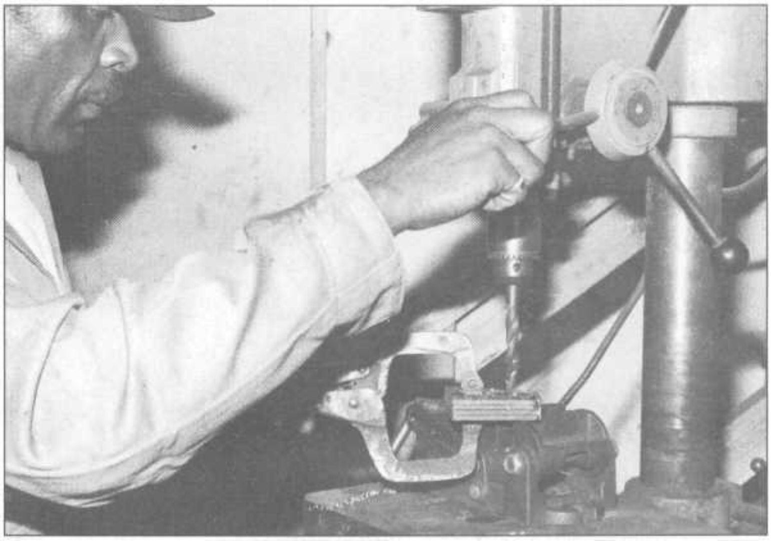

9.21 After flame-cutting, torching or plasma-cutting your plates, sand the edges smooth and stack them together with a vise-grip clamp. You can drill all four plates at the same time, saving time and ensuring that all have the spring-bolt holes in the same place.

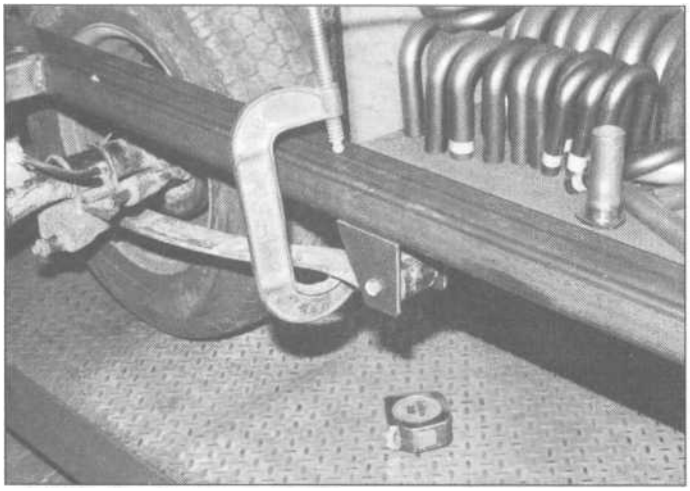

9.22 The front spring-mounting plates are bolted onto either side of the front spring eye, squared to the rails and clamped lightly in position for tacking.

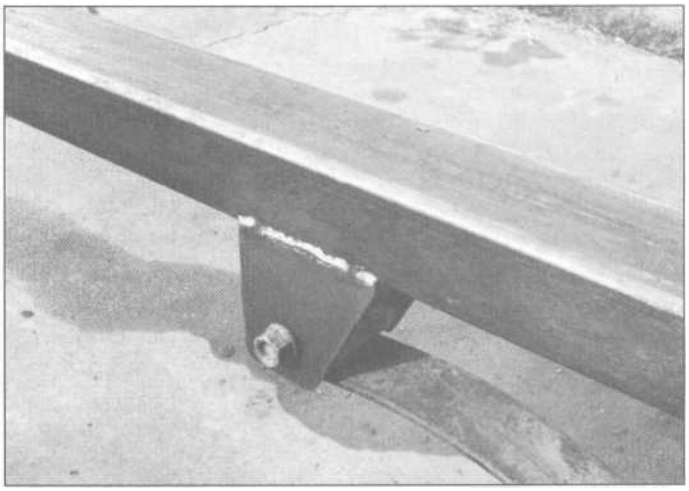

9.23 After some solid tack-welding, pull the springs out so the bushings don't melt, and fully-weld the four front spring mounts. Here the springs are bolted back in.

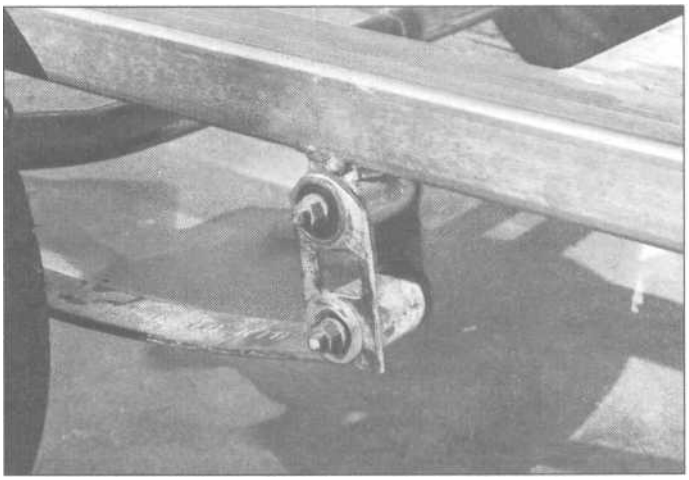

9.24 At the rear, short lengths of heavy-wall tubing, with an inside-diameter that fits the spring bushings, are welded to the underside of the rails (tack first, then remove the bushings to finish welding). Tubes should be positioned to have the spring shackles straight up and down. When the weight of the whole trailer is put on, they will arc to the rear enough for easy spring action.

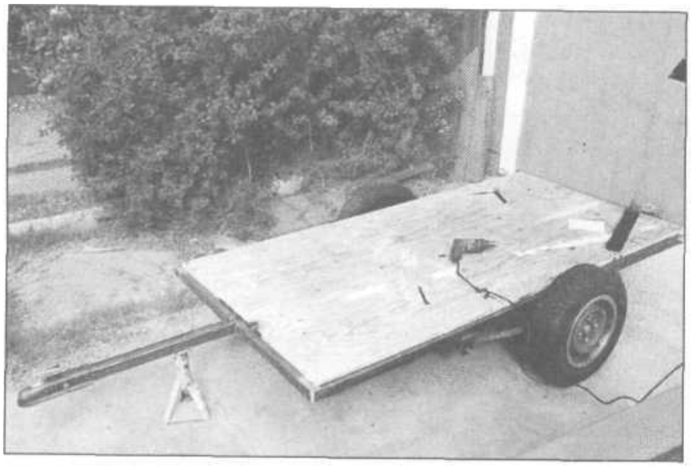

9.25 In our case, the frame was built just the right size to use an uncut 4x8-foot piece of 3/4-inch plywood for the trailer's floor.

9.26 Locate where your tubes are and drill through the flooring and tubing, then bolt the floor down with carriage bolts, with the round heads on top so the floor is relatively smooth and a furniture dolly will go over the floor easily.

enough tongue on the front and, therefore, not enough tongue weight. When loading your utility trailer, try to load the heavier items toward the front of the trailer.

When making joints in tubing, angle and other materials, get as tight a fit as you can before welding. A tight joint will always weld up as a stronger unit than something with gaps to be filled. If you are cutting with a hacksaw, use a carpentry miter-box to start your 45-degree cuts at the right angle. When you're done, give the whole thing a coat or two of rust-resistant paint, get an assigned registration number and license plate at your local Department of Motor Vehicles, and you're on your way.

When your trailer is complete, have a qualified weldor/metals expert inspect it for strength and safety. Then find some things to move!

9.27 Four posts of heavy 1x2-inch angle-iron were welded on, one in each corner, as supports for the plywood sides. After welding, all were checked with a square.

9.28 Having metal around for welding projects can pay off when working with wood, too. Use an extra length of straight angle-iron or tubing to clamp down on a piece of plywood as a saw guide, for perfectly straight cuts.

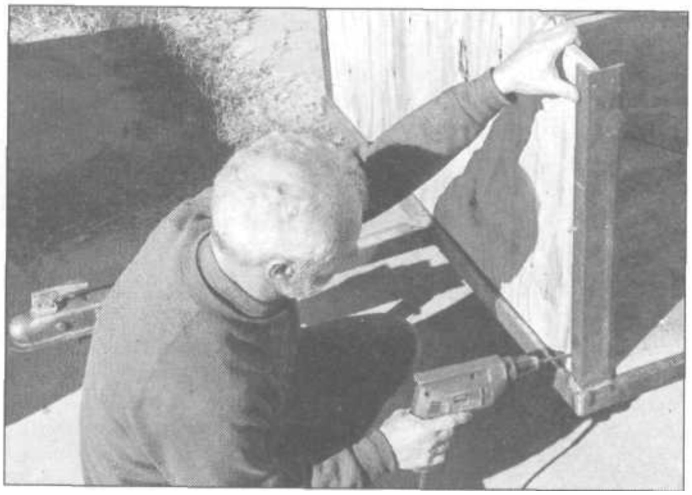

9.29 The front panel is attached with carriage bolts through the angle-iron posts. The posts were pre-drilled before welding on, because they are much easier to drill in a drill-press than once they are welded in place.

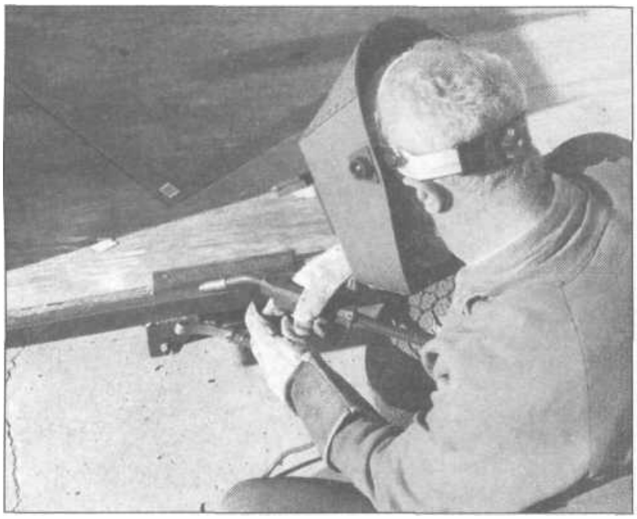

9.30 Short strips of 1/8-inch thick, 2-inch-wide strap metal (pre-drilled) were welded along the floor edges to help attach the bottom of the side panels. When MIG welding outdoors with argon gas shielding, use pieces of plywood around you to form a wind barrier to keep the gas from blowing away as you weld.

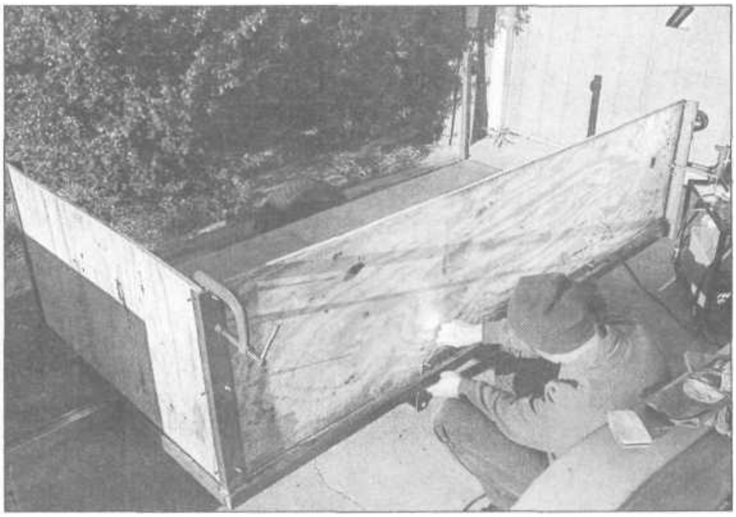

9.31 Plywood sides are bolted to the end pillars and the side plates.

9.32 At the local trailer supply store, we found a wide variety of lights, wiring harnesses, plugs and hitch parts to complete our utility project.

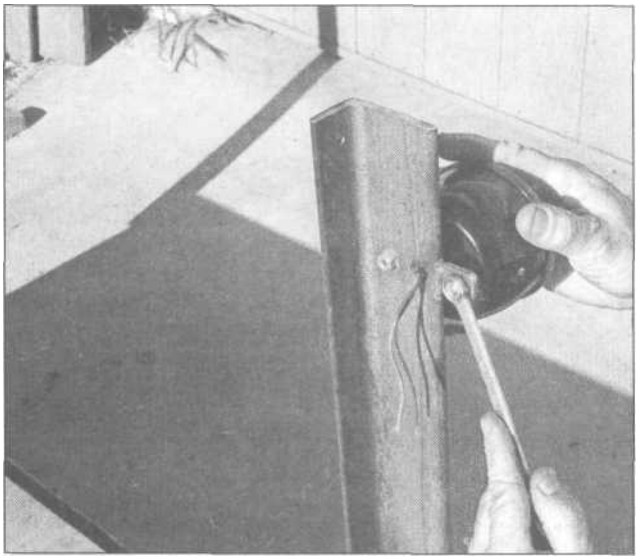

9.33 The taillights were bolted to the rear angle posts, but with a short extra piece of steel added on to move the lights further outboard to clear the tailgate.

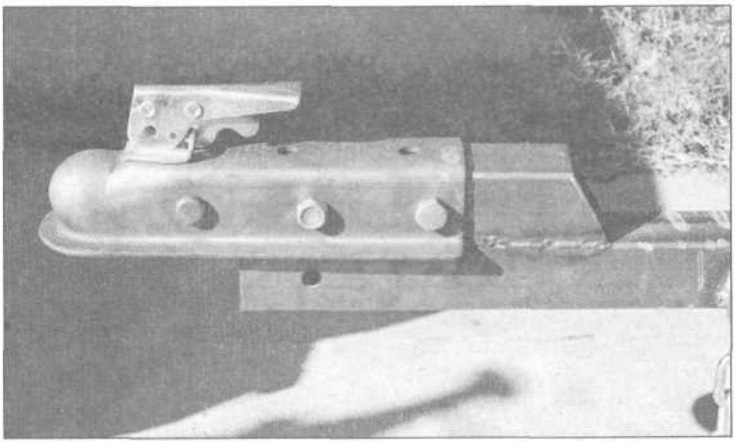

9.34 We originally had the hitch bolted to the main tongue, but found we had a better tongue weight with the tongue down another two inches. Simply welding on another piece of 2x2-inch, heavy-wall tubing set the trailer just right.

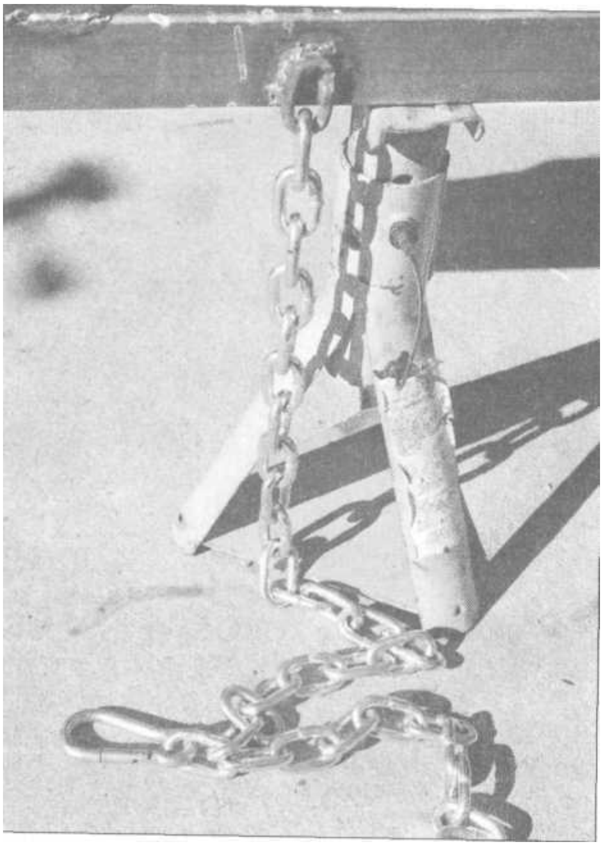

9.35 Every trailer needs a safety chain, check with your local DMV to see if your trailer qualifies for just one, due to light weight, or needs two to be legal. Take one end link, grind it clean of any plating and weld it to the trailer tongue (at top here). Below you'll see the removable snap-link that hooks to our tow vehicle.

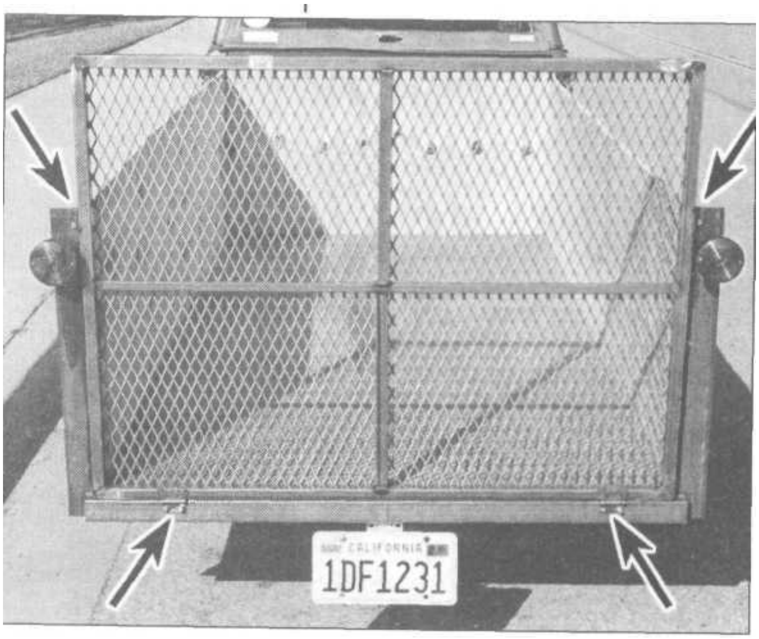

9.36 Our metal tailgate will save fuel because air flows through it on the road, cutting wind resistance. The frame was made of 1x1-inch tubing, to which was added steel mesh. Hinges (bottom arrows) were welded to the tailgate and bolted to the trailer. Top arrows indicate where tabs were welded to the gate for bolting with wingnuts to the trailer structure.

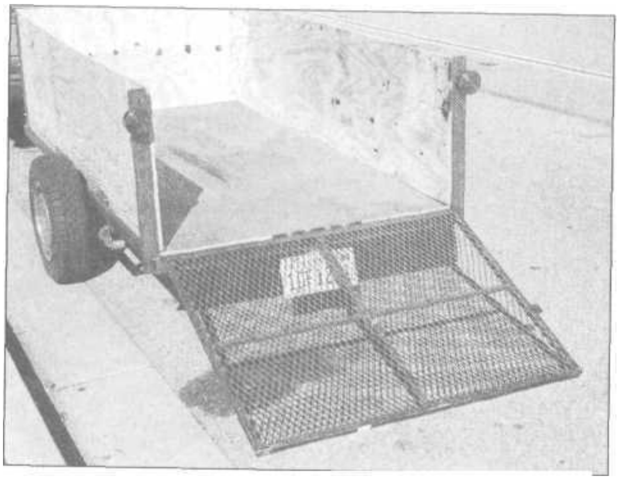

9.37 With the wingnuts undone, the hinged tailgate folds down as a ramp for furniture dollies, lawnmowers, etc. The trailer should be hooked up to the tow vehicle before you step on the tailgate/ramp, or the lightweight trailer could tilt quickly.

9.38 A 4x8x2-foot utility trailer actually hauls a lot of "stuff". We've known people who built a simple trailer just to make a long-distance household move, then sold the trailer when they got there, saving considerably on what a rental trailer would have cost.

Glossary of welding terms

AC Current - Alternating current, such as in homes and businesses. Polarity switches at 60 cycles per second (US).

Acetylene - A highly flammable gaseous hydrocarbon fuel used with oxygen in gas torch welding/brazing/cutting.

Alloy - A modified metal made by combining a base metal with other metals or chemicals to produce a metal with different properties.

Amperage - A measure of electrical power. In electric welding, such as TIG, MIG and arc, the higher the welding amperage, the more heat is generated at the arc.

Annealing - A process of heating a metal and slowly cooling it to soften it and make it more workable.

Arc Welding - An electrical welding process in which a consumable, flux-coated electrode rod makes an arc with the work. Also called stick welding. Arc welding deposits a coating of slag over the seam, which must be chipped off after it cools.

Acetylene bottles

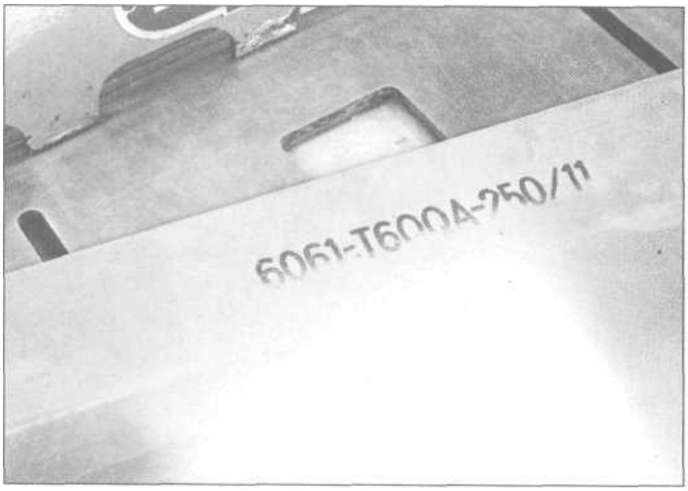

Alloy numbers on aluminum

Arc welding

Argon bottles

Arc Blow - A phenomenon in electrical welding where the magnetic forces in the work can cause a wind-like force that blows shielding gasses away. Usually happens in corner welding and can be avoided by switching polarity or moving the ground clamp.

Argon - A colorless, odorless inert gas often used as a shielding gas in TIG and MIG welding.

Backfire - A loud pop heard at the torch in oxy-acetylene welding that indicates combustion has backed up into the torch. A potentially dangerous phenomenon caused by holding the torch too close.

Base Metal - In alloying, this is the pure metal before alloys are added. In welding, this refers to the work metal, as opposed to the filler metal melted in.

Bead - The seam between two pieces of metal that have been fused by welding. The bead usually looks like a tight series of overlapping circles or ovals of metal.

Brazing - A type of metal-joining where the parent or base metal isn't melted but is heated only enough to melt a filler metal such as brass, which holds the parts together after cooling.

Butane - A fuel gas sometimes used with oxygen for brazing or soldering. Butane does not produce as much heat as acetylene.

Burn-Through - When the welding flame or arc is too hot for the material, holes develop in the seam.

Butt Weld - The most common joint in home/shop welding, where the two edges to be joined are lying flat in front of the weldor.

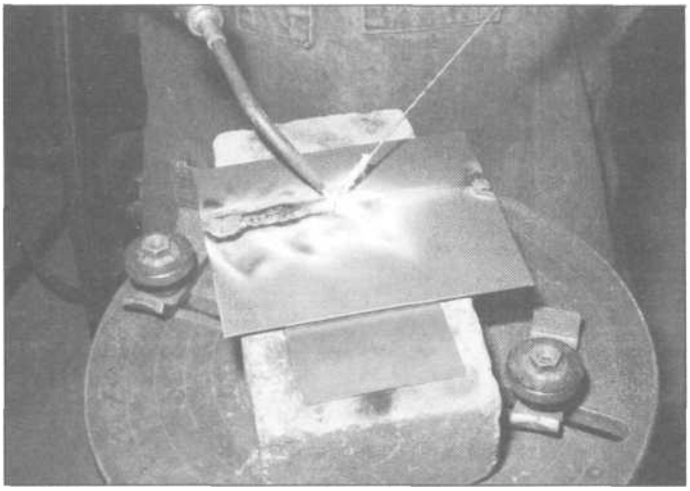

Brazing

Carbon - A non-metallic basic element that is a key ingredient in alloying of steels. Higher amounts of carbon in a steel make it tougher.

Carbon-Arc - A type of welding in which two carbon rods are brought together to create an arc independent of the parent metal. A process virtually obsolete in home/shop welding today.

Carburizing - In welding, an oxy-acetylene flame that is rich in acetylene, used to coat parent metals with soot before annealing. In metal work, a process of treating a surface such as a tool or knife edge by adding carbon to a hot metal, increasing surface hardness.

Cast Iron - A common metal, formed of steel with very high carbon and silicon content, in which soft graphite flakes appear throughout. The graphite makes the material hard to weld, but it is commonly used in large castings such as automotive engine blocks, heads and crankshafts, as it is inexpensive to manufacture and resists vibration damage.

Clad Metals - Usually mild steel alloys coated with another metal, such as galvanized steel, or sandwiched with another metal such as

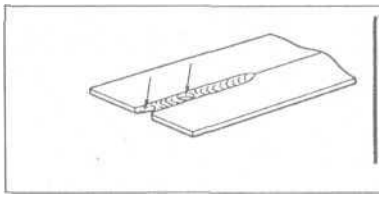

- The welding current is too high.
- The gap between the metal is too wide.
- The speed of the gun is too slow.
- The gun to base metal distance is too short.

Burn-through

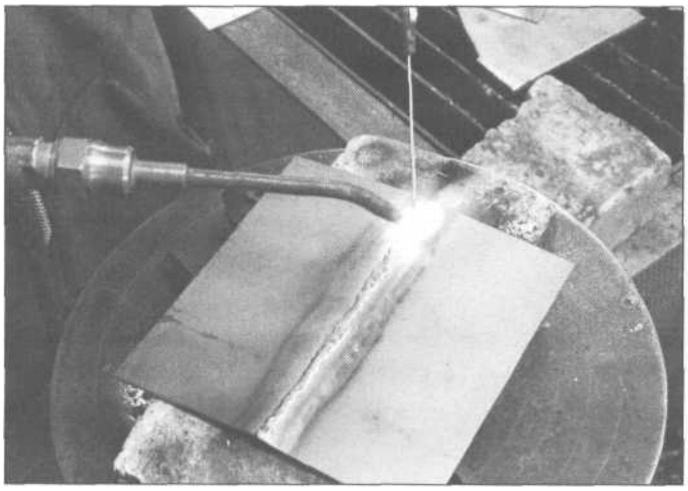

Butt weld

Deposit

kitchenware that has mild steel sandwiched between two layers of stainless-steel.

Copper - A reddish basic metal that is relatively soft, malleable and is one of the best conductors of heat and electricity.

Current - The flow of electricity, usually described by its direction of flow (alternating or direct) and strength (amperage and voltage).

Deposit - The addition of new metal to a welding seam by melting in a filler rod.

Deposition Rate - How fast filler metal is added to a welding seam. Determined by the wire-feed speed in MIG welding, and by the movement of the hand-held filler rod in gas, arc and TIG welding.

Ductility - A property of a metal that indicates how well it can be hammered out thin or drawn into a wire. A welded joint that passes a bending test without breaking is said to have good ductility.

Duty-Cycle - In electric welding processes, this is a way of gauging how long a particular machine can weld before it must be cooled off. It is expressed as a percentage of a ten-minute test period. At a 40% duty-cycle, a welder can be operated for four minutes, then must be cooled for six. At 100% duty-cycle, such as industrial equipment, the welder can be operated all the time.

Electrode - In electric welding, a conductive element that makes the final connection with the work to create an arc of electrical energy. In arc-welding, it is the rod or stick. In MIG it is the welding wire, and in TIG it is a tungsten rod.

Eye-Flash- A eye problem that occurs from observing a welding arc without protective goggles or lenses. Even brief exposure to the intense ultraviolet and infrared radiation can harm your eyes, causing a burning sensation that may show up 6-8 hours after exposure.

Ferrous Metals - Any of several metals made from iron and other elements. Common ferrous metals include wrought-iron, cast-iron, steel and cast-steel. Magnets attract ferrous metals.

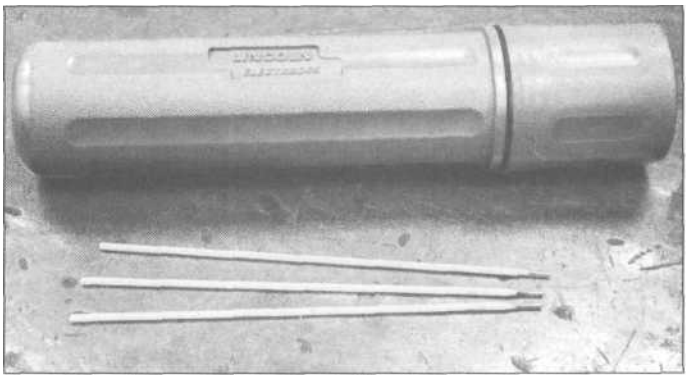

Electrode - arc rods

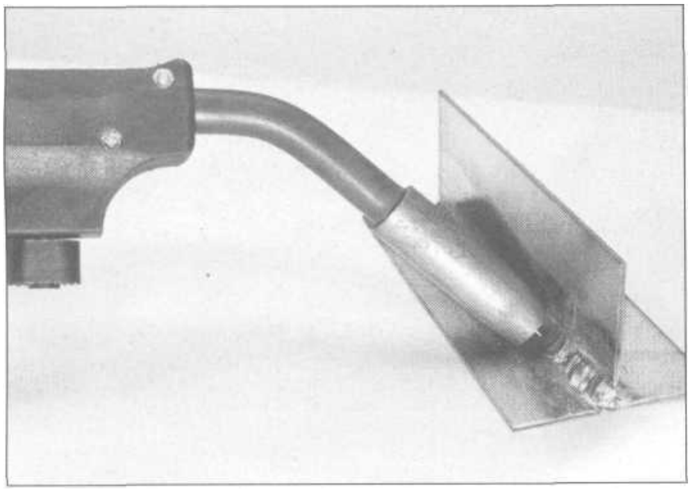

Fillet weld

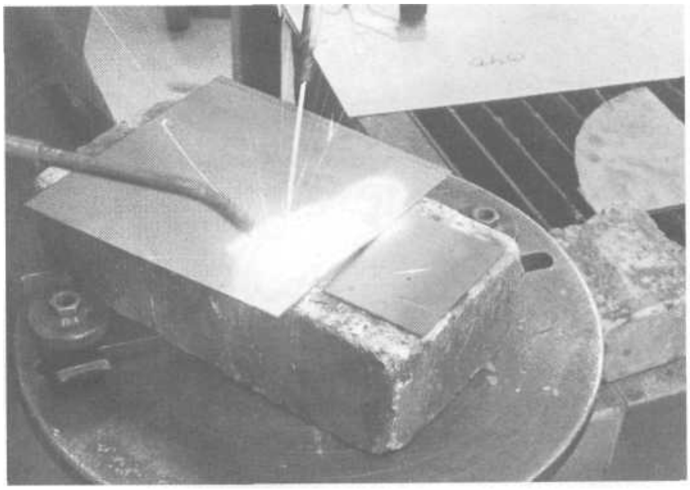

Firebrick

Filler Metal - New metal added to a welded joint by melting in a filler rod or wire. Most weld seams require filler metal.

Fillet Weld - A joint such as a 90-degree or T-joint, where the weld bead is filling a triangular-shaped area. Also describes a multiple-pass weld on thick, beveled plates.

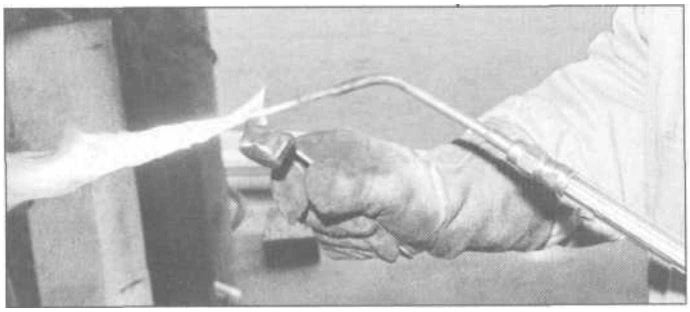

Flame

Firebrick- Special fire-proof bricks used in welding as a non-conductive base on which to set parts to be welded. Usually larger than common masonry bricks. Do *not* use concrete or red brick to weld on, they will crack or explode from trapped moisture inside.

Flame - The heat-creating, working-end in oxy-acetylene welding/cutting/brazing. The flame, not the torch, contacts the work.

Flamecutter - A mechanized gas or plasma torch that is motorized and follows a pattern, to cut out repetitive shapes from metal.

Flashback - A situation in gas-welding in which combustion goes back into the torch. More serious than a backfire, it is signaled by a loud squeal, and should be stopped immediately by turning off the gasses at their tanks.

Flashback Arrestors - One-way valves installed in oxygen and acetylene hoses to prevent flashback. All gas-welding setups should have these.

Flowmeter - A type of gauge for shielding gasses which indicates the flow via a colored ball floating in a clear glass tube which is easier to read from distance than a dial-type gauge.

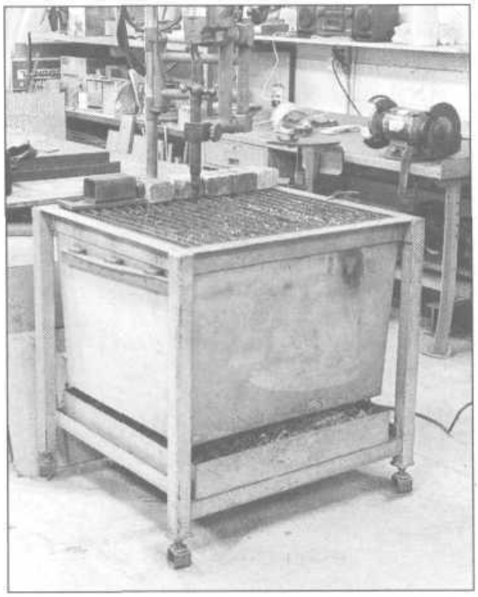

Flamecutter

FLASH ARRESTORS
TORCH-MOUNTED TYPE
UL-LISTED
STOP DANGEROUS FLASHBACKS
L-TEC
L-TEC
L-TEC

Flashback arrestors

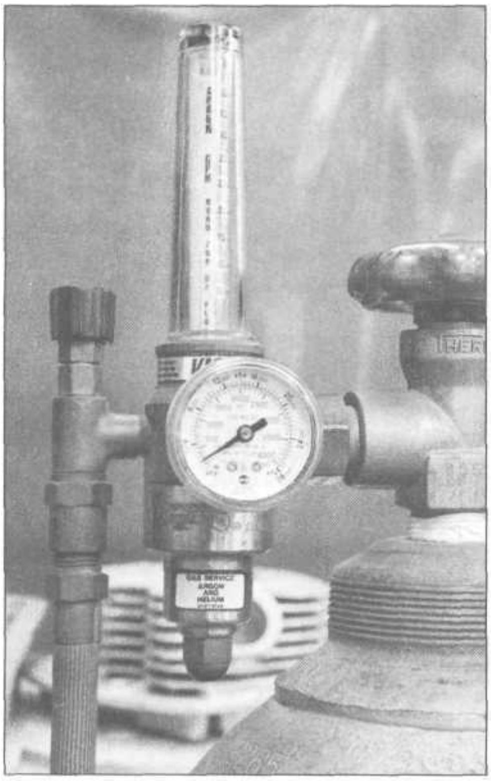

Flowmeter

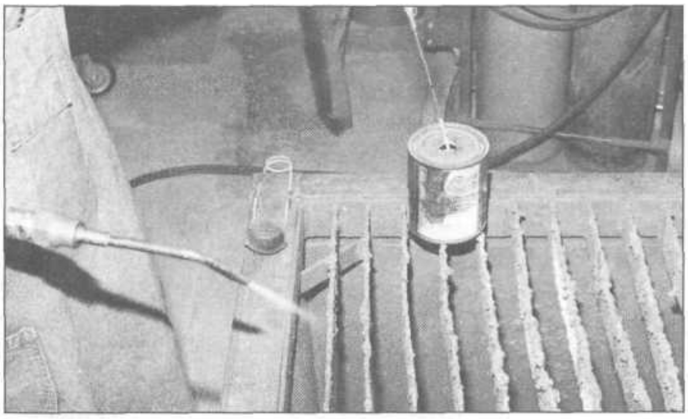

Flux

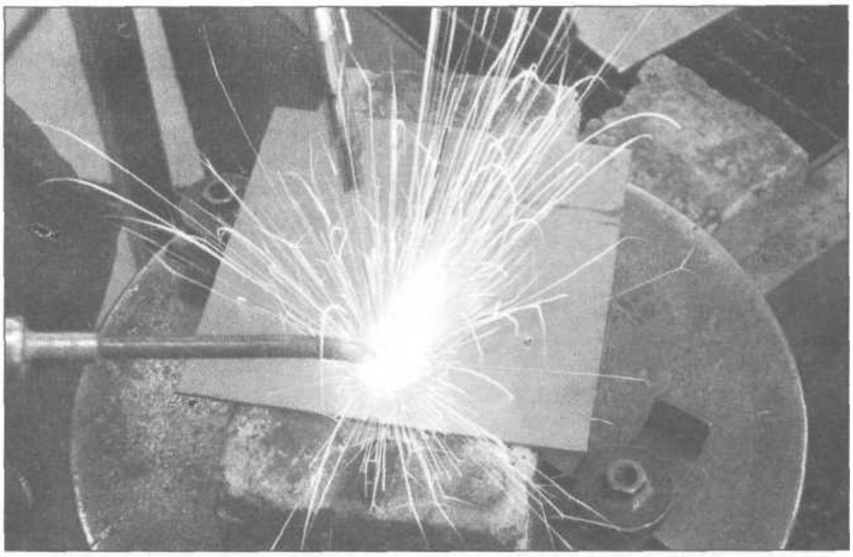

Gas welding

Flux - A chemical compound that cleans base metal under heat. In brazing, flux prepares the base metal for the brass to stick to it. Flux is applied as a coating to arc-welding rods and to wire for flux-cored MIG welding and is consumed in the arc to produce shielding gas.

Forge-Welding - A blacksmithing process where two pieces of metal are heated to the near-molten state and hammered together, resulting in fusion. Until the late 19th-century, all metals were either soldered, brazed or forge-welded.

Fusion Welding - Where the edges of two pieces of metal are brought to a molten state and joined, usually with filler metal of a similar alloy melted in.

Gas Welding - Using a mixture of oxygen and acetylene gasses to form a very hot, concentrated flame to heat and fuse metals.

GMAW - Stands for Gas Metal-Arc Welding, also known as MIG (Metal Inert Gas welding) and commonly called wire-feed welding. The process uses a shielding gas emitted from the torch while the consumable wire is the filler and electrode.

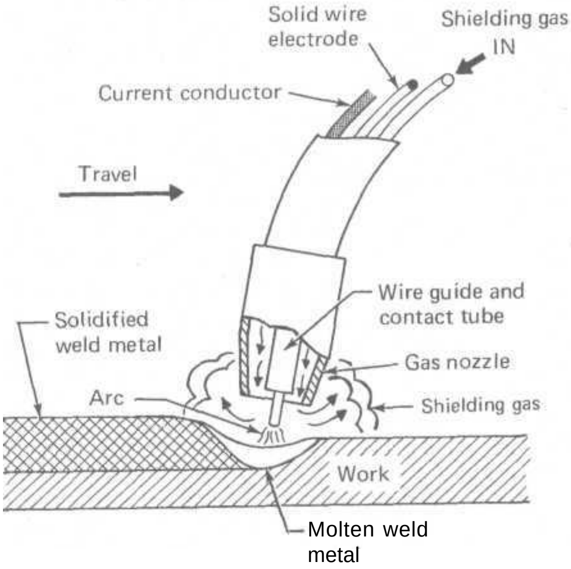
Solid wire electrode
Shielding gas IN
Current conductor
Travel
Wire guide and contact tube
Solidified weld metal
Gas nozzle
Arc
Shielding gas
Work
Molten weld metal

GMAW

GTAW - Stands for Gas Tungsten Arc Welding, also called TIG (Tungsten Inert Gas welding), and may be referred to as heli-arc welding, though this is a tradename. Helium or argon shielding gas emits from the torch, while a non-consumable tungsten electrode makes the arc, and filler is added from a hand-held rod like in gas welding.

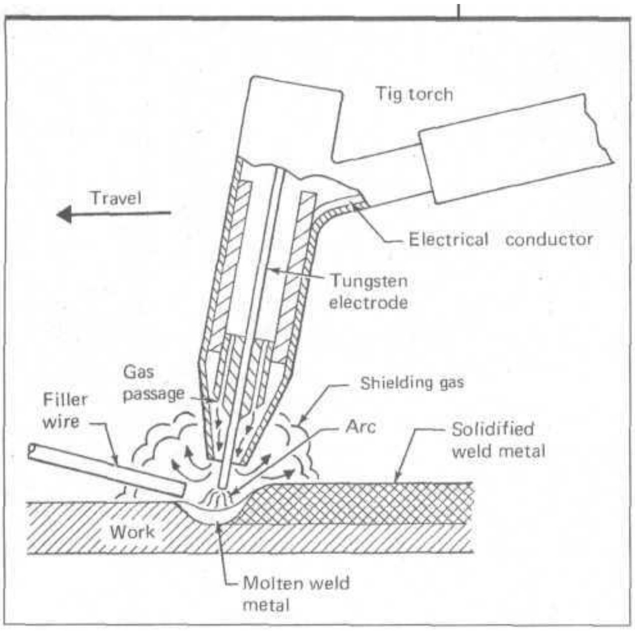

GTAW

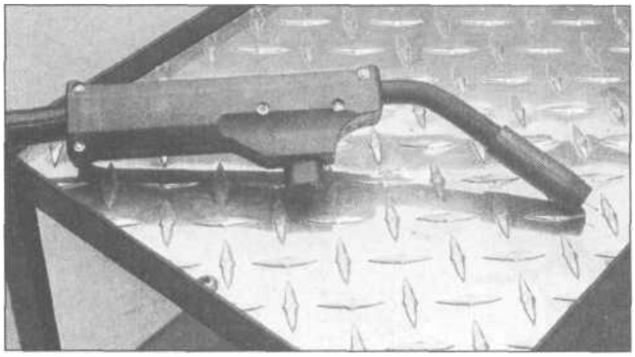

MIG gun

Ground Clamp - In electrical methods of welding, this is a heavy cable from the welding machine that clamps to the work or the conductive metal work table, completing the circuit from the electrode to the work and back to the machine.

Gun - In welding, this usually refers to the "torch," particularly in the MIG welding process. The similarity to a gun is that the wire "shoots" out of it into the weld bead.

Hardness - A metal property that can be tested by the ability of the metal to scratch another surface or be scratched. Sensitive equipment is used to measure metal hardness, and the alloy number of a metal is used as a hardness indicator.

Hard-Surfacing - In arc-welding, a procedure in which many weld beads are made overlapping, covering a large area. Using special rods, a very hard, abrasion-resistant face can be added to tools or farm implements.

Heat-Affected Zone - This is the area along either side of the weld that has been changed by the fusion heat. On ferrous metals, it is indicated by a discoloration. In some cases the metal in this zone may have become weaker than the joint itself.

Heat Sink - Something used behind or around an area to be welded, to protect adjacent areas from heat. Wet rags and moldable, heat-resistant clays are often used around welding on thin sheet metal.

Heat-Treating - Changing a metal's characteristics of strength and formability by scientifically heating and cooling it. The term usually implies making a metal harder.

Heli-Arc - A tradename of the L-TEC company (formerly Linde), for their TIG welding equipment. The original process began with Helium as the shielding gas, and the tradename

Small heat-affected zone on **MIG** weld

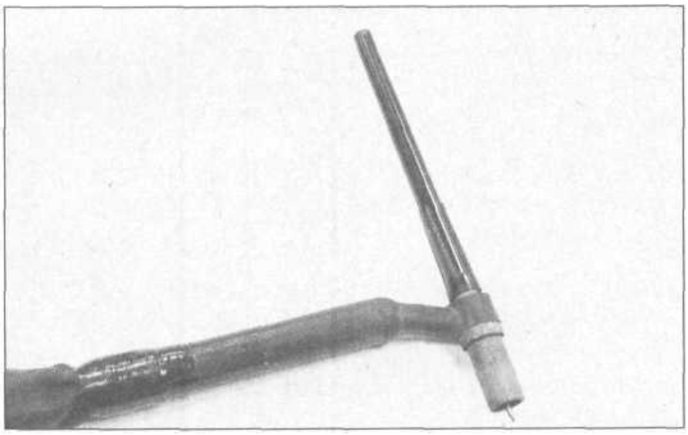

Heli-arc torch

came into such wide usage that it is now generic when used by weldors.

Helium - An extremely light, colorless, nonflammable gas, used in welding as a shielding gas and in everyday use to fill balloons.

Hydrogen - Atomically the simplest of all atoms, hydrogen in its normal state is a colorless, odorless lightweight gas that is extremely flammable. In small amounts it is used in some industrial welding/cutting applications.

Jig - A device used to keep parts in alignment or position during welding.

Joint - Where two parts are joined by welding, or where two sides of a crack are welded.

Kerf - The width of the cut in metal cutting by torch or plasma cutter.

Lap Weld - Where two pieces of metal are overlapped before welding, rather than butted together. Commonly used in brazing, and to stiffen a joint in thin sheet metal.

Lead - Another term for an electrical cable, such as a ground *lead.*

Machine - The power supply for an electrical welding setup.

Machine Welding/Cutting - When the torch or welding/cutting gun is moved by motorized equipment rather than by hand, resulting in a more precise weld or cut.

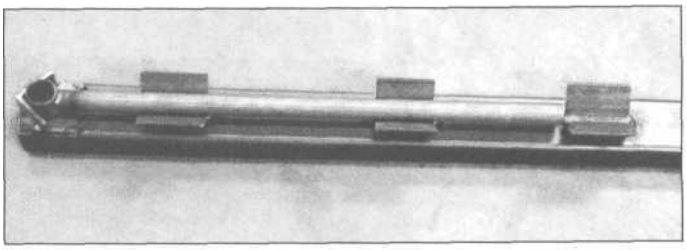

Jig

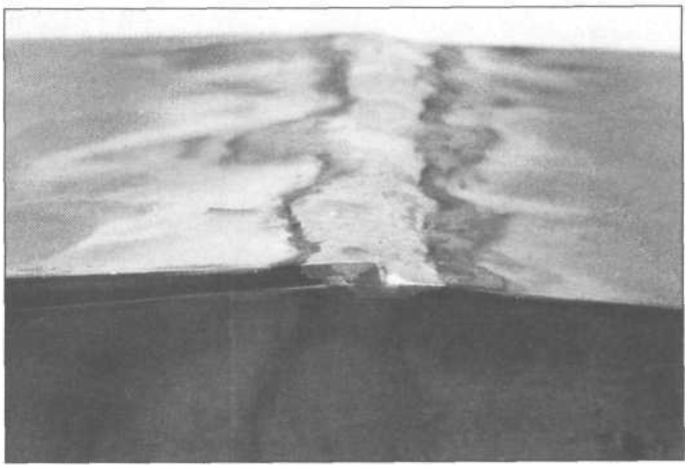

Brazed lap joint

TIG machine

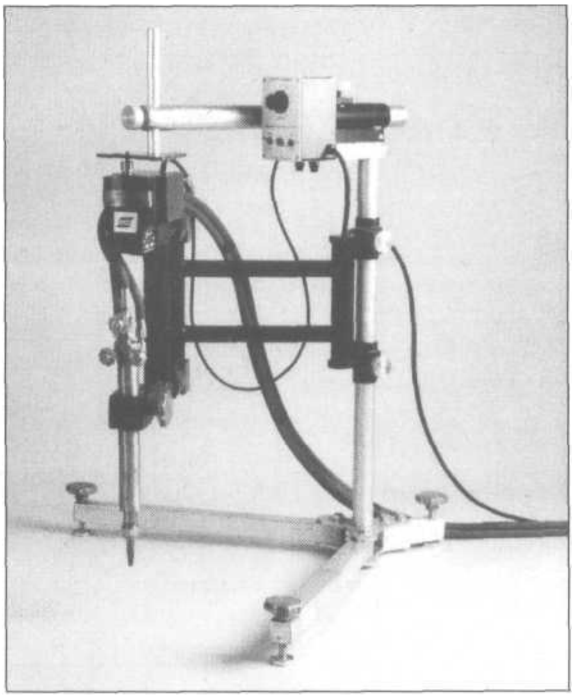

Machine welder/cutter

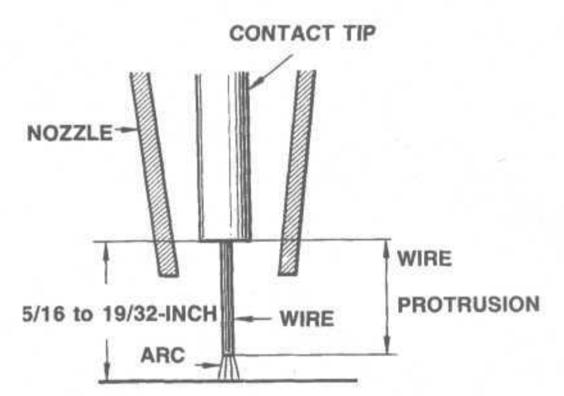

MIG

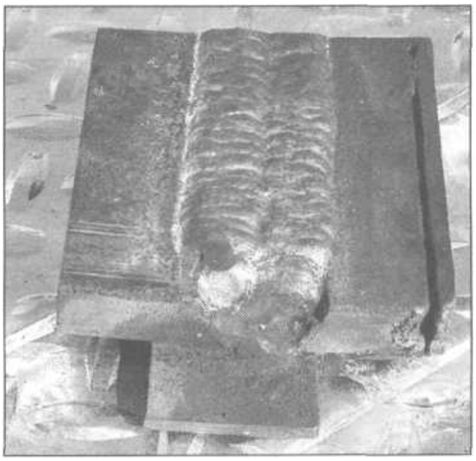

Multiple-pass weld

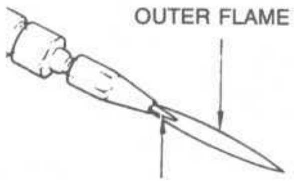

Neutral flame

Malleability - A metal quality indicating how well the material responds to being shaped by a hammer or forming tools.

Manifold System - Where a number of welding torches are fed by a common source of gasses, such as in an industrial situation or a welding school.

MIG - Stands for Metal Inert Gas welding, more commonly called wire-feed welding because the consumable electrode is a metal wire machine-fed through the gun from a roll.

Mild Steel - Synonymous with low-carbon steel, the most commonly found and highly useful material, easy to weld, heat and form.

Molybdenum - A modern metal used in alloying, to make steels and stainless steels stronger and more resistant to corrosion. A well known steel alloy is 4130 chrome-moly which contains molybdenum and chromium and is a strong material used in aircraft and race cars.

Multiple-Pass Weld - When welding thick plates or repairing large cracks in castings, more than one bead must be made. The joint is usually beveled first and the succeeding passes have wider beads until the joint is filled.

Neutral Flame - In oxy-acetylene welding and cutting, this is the desired adjustment of the torch flame, with equal amounts of each gas.

Nickel - A metal used in alloying steel and stainless steels chiefly to increase corrosion resistance and suitability for both very low and very high temperature extremes.

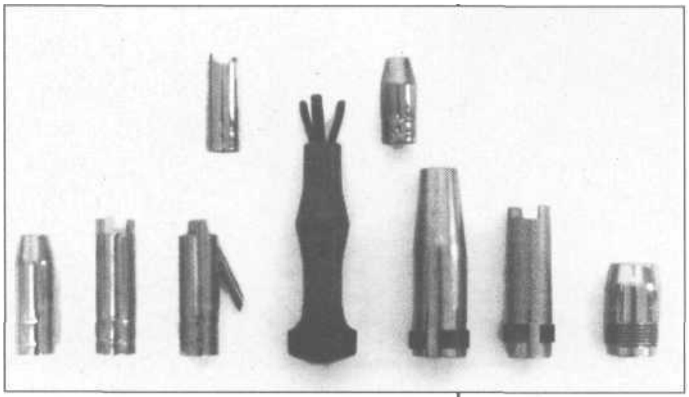

MIG nozzles

Nitriding - A metal treatment process to surface-harden certain steels, castings and forgings.

Nitrogen - The main component (75%) of the air we breathe, nitrogen is an inert gas that can be used as a shielding gas in welding copper, or as the cutting gas in plasma equipment.

Normalizing - A heat-treating process that refines the grain structure in hard steels. It is used when high-strength steels have been welded, to remove stresses.

Nozzle - A component part of the torch in TIG, MIG and plasma welding equipment, designed to focus the shielding gas over the arc.

Out-Of-Position Welding - Welding whenever the parts are not laid flat in front of the weldor, such as welding vertically, overhead or horizontally but perpendicular to the ground.

Oxidation - The deterioration process when metals are exposed to oxygen through air, moisture and natural or manmade harmful elements, On steel it appears as red-brown rust, on aluminum it can be a gray-black. To prevent it, metals must be coated with oils, paints or plating processes.

Oxidizing Flame - In oxy-acetylene welding/cutting, when the flame is adjusted with too much oxygen, identified by a ragged purple outer cone and a hiss from the torch.

Oxygen - A colorless, odorless, tasteless gas that makes up 21 % of the air around us. It is essential to combustion processes, and combined with equal parts acetylene makes a 3315° C flame for welding and cutting.

Penetration - Important in all welding processes, penetration describes how far into the two pieces of a joint the fusing process extends. Maximum penetration makes the strongest joint.

Piercing - In metal cutting, piercing is the technique for starting a cut in the middle of the material (such as cutting an *interior* hole in a piece of plate). It usually requires slightly different torch procedure than starting from an edge.

Oxygen bottles

Pickling - Treating a metal chemically before welding it. For instance, cad plating on bolts, chain and other hardware can create dangerous fumes when welded, plus good contact is hard to make. Plated parts should be stripped in swimming-pool acid before welding.

Plug Weld - Where a piece of tubing fits inside another and is welded to it through holes drilled in the outer tube. Also called a rosette weld.

Polarity - The direction in which electrical current flows in a welding setup. There is AC, which alternates in 60 cycles per second (US), and DC, direct current, which can be either *straight* polarity or *reversed* polarity.

Porosity - An undesirable quality in a weld, with tiny holes caused by welding dirty metal, using damp arc electrodes, or getting a gas torch too close to the metal.

Post-Flow - A setting on a TIG machine allows the weldor to adjust how long after the arc is broken the shielding gas will continue to flow to protect/cool the bead.

Post-Heating - Heating a welded part after the final welding to relieve stresses.

Pre-Heating - When an entire part, or just the weld area, is heated with an oven or torch before welded to prevent thermal stresses and uneven expansion and contraction. Thick metals and aluminum and iron castings are usually preheated before welding.

Propane - A fuel gas commonly used in self-contained torches for heating and soldering plumbing pipes. It can be used with oxygen to perform welding/brazing on thin materials.

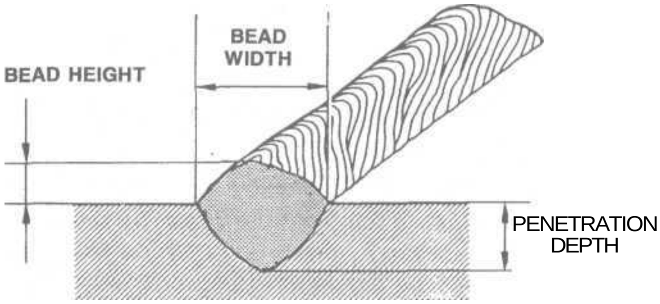

Penetration

Pre-heating

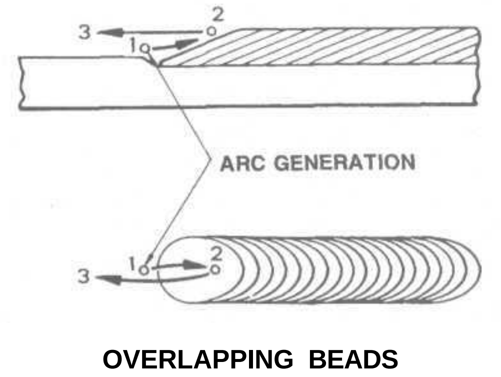
3
1
2
ARC GENERATION
3
1
2
OVERLAPPING BEADS

Puddle

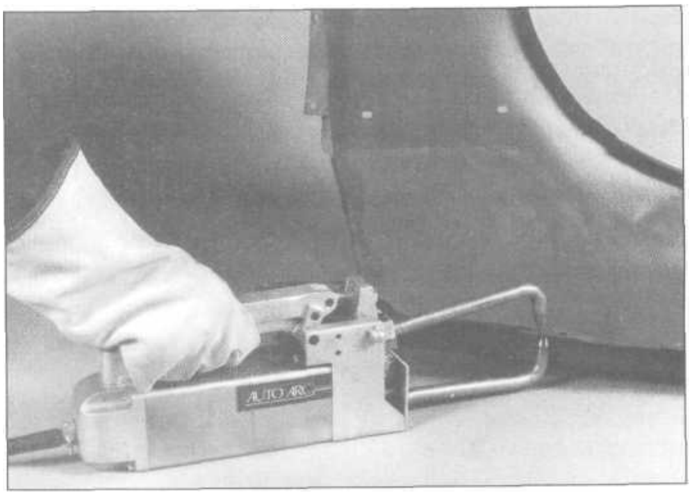

Resistance welding

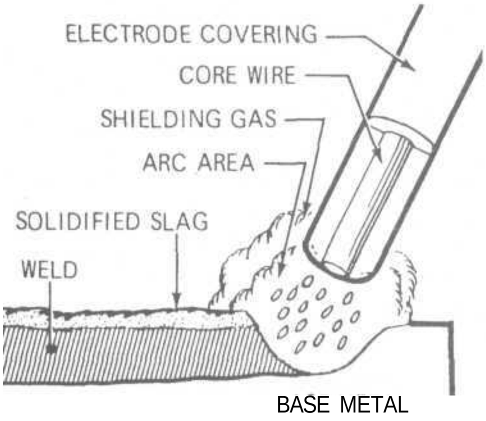
ELECTRODE COVERING
CORE WIRE
SHIELDING GAS
ARC AREA
SOLIDIFIED SLAG
WELD
BASE METAL

Arc welding rod

Puddle - The welding bead in process, as the concentrated area of the parent metal is melted and joined with molten filler rod. The finished bead is a continuous string of overlapping puddles.

Rectifier - A major component of many electric welding machines, the rectifier changes AC input current to DC output.

Resistance Welding - A method of welding where electrodes are placed on either side of the material to be joined, and a brief arc is made, which is enough to fuse the materials together at that point without greatly affecting the area around it. The weld is made by heat and pressure, and the technique is often used as spot-welding on sheet metal.

Rod - The filler rod in welding processes. In arc-welding, rod is another term for the coated, consumable electrode. In gas and TIG welding, the filler rod is usually uncoated and hand-fed into the puddle.

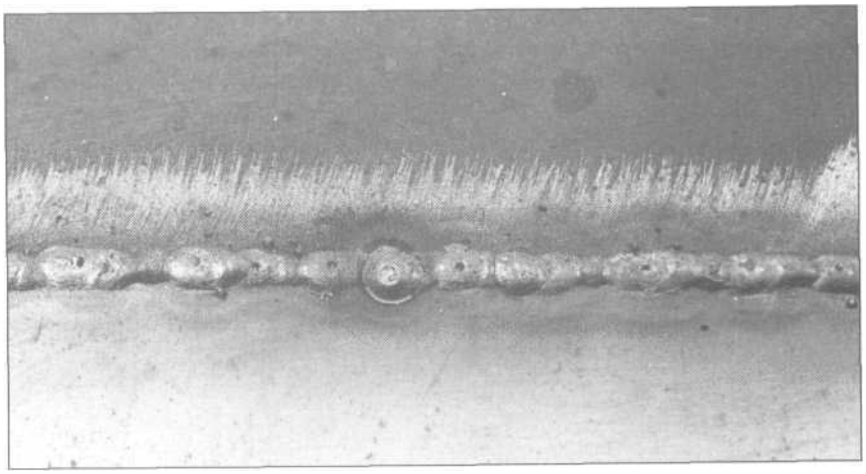

Seam welding

Glossary of welding terms

Root - The bottom of a weld bead viewed in cutaway. In multiple-pass welding oh thick material, the first bead in the bottom is called the *root pass.*

Reverse Polarity - In electric welding, the term refers to a DC setup where the work is made *negative,* and is called DCRP. It is the opposite of DCSP, or *straight* polarity.

Seam Welding - A procedure used on lighter-gauge metals where the bead is composed of overlapping spot welds. Some MIG machines can be timed to perform seam welds. It is also used on machine-made welds to join long seams in sheet metal.

Sheet metal - Thinner metals, such as found in auto bodies, appliances, heating/cooling ducts. Usually refers to metals from 12-gauge to 24-gauge.

Shielding Gas - An inert (nonflammable, non-reactive) gas forced out around a weld procedure to cover the arc area, excluding air and impurities from the forming weld.

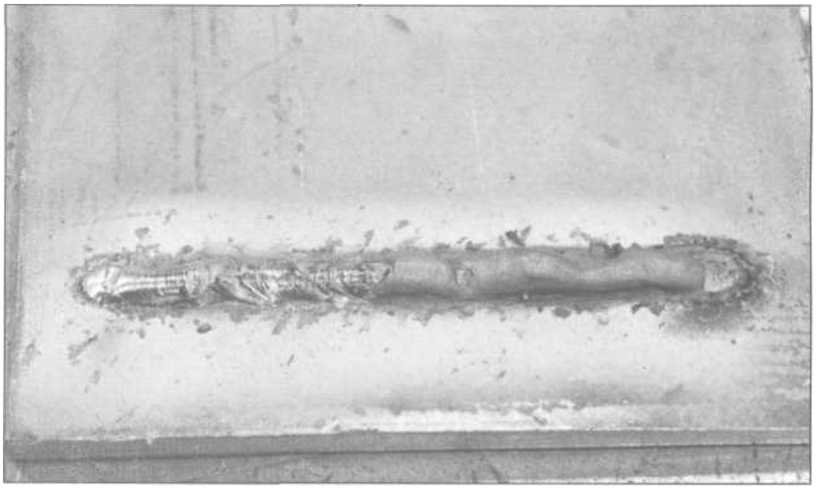

Slag

Shielded Metal Arc Welding - Another term for stick or arc-welding, it is abbreviated as SMAW. The flux coating on the consumable electrodes vaporizes into shielding gas to protect the weld.

Short Circuit - In electrical terms, this is what makes the arc when there is an air gap between positive and negative current flow. When you plug an appliance that is already switched on into the wall outlet and there is a tiny spark at the plug's prongs, this is a small-scale version of the short-circuit that makes electric welding possible.

Slag - Oxidized impurities resulting from welding and cutting. In arc welding, the slag forms a thick, hard coating over the seam, which must be chipped off with a pointed hammer. Slag is also found along the bottom of a cut made by a gas torch.

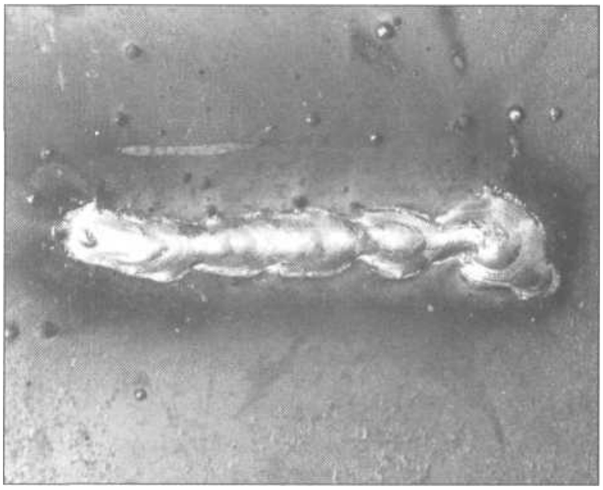

Spatter

Slope Control - In electrical welding, professional-level welders have controls that adjust the shape of the current waves, to weld better on non-ferrous materials. Slope control can also be used to describe controls that gradually increase current at the beginning of a weld, and taper off current at the end, both to avoid cratering.

Spatter - Tiny balls of filler metal stuck to the parent metal around the weld, spatter is produced mostly by incorrect welding technique, but in arc welding spatter is unavoidable. Spatter is also present in smaller amounts with MIG welding.

Spot Welding - Small circular welds used to hold light sheet-metal panels together. Usually made through a hole in the topmost panel and common in auto bodywork (also see resistance welding).

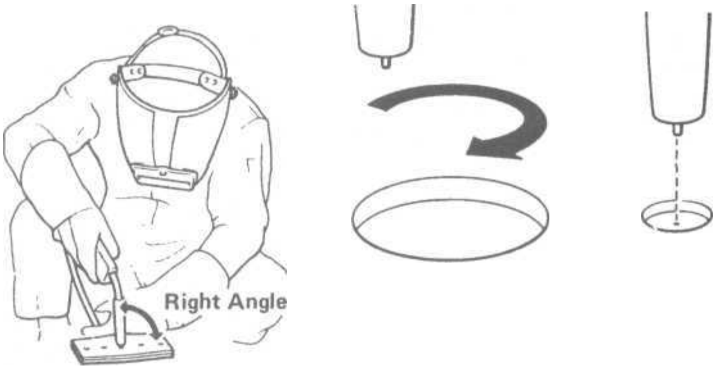

Spot welding

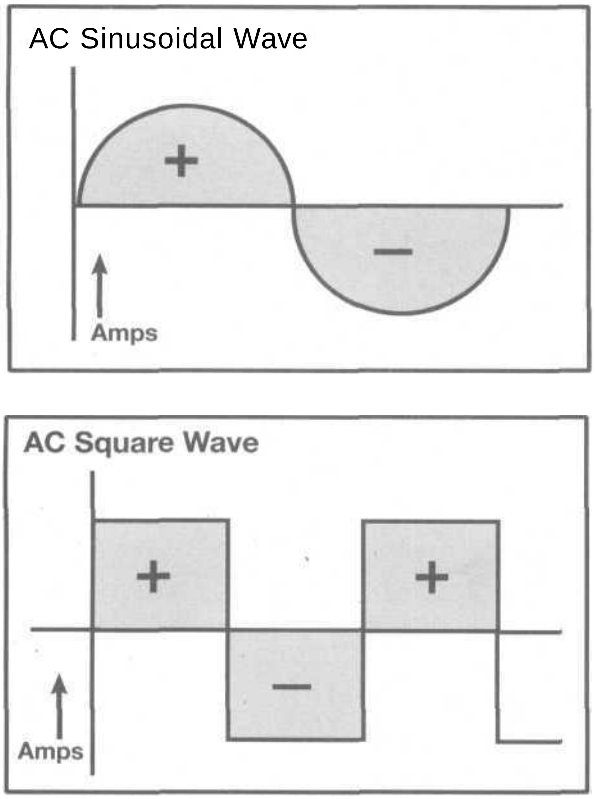

2110-6-13 HAYNES

Square-wave technology

Stick Welding - Another term for arc welding.

Stitch Welding - A seam where welding is done only in short strips with gaps in between where full welding isn't required. Better MIG machines have timer controls to easily do stitch welding.

Stud Welding - An option for some types of MIG welders whereby a special attachment to the gun accepts hollow studs. The studs are pressed onto a surface where the arc is made, and filler wire welds the stud to the surface and fills the inside of the stud. A very fast method of replacing broken studs in exhaust system work.

Square-Wave Technology - An electronic method of changing the shape of the AC current wave for improved welding of non-ferrous metals with TIG equipment.

Submerged-Arc Welding - An industrial process where the seam is buried under a long, thin pile of flux material, and a mechanized welder travels along the seam welding *underneath* the flux pile. The arc is not visible, and the technique is used often on heavy plates such as in shipbuilding.

Synchronous Control - A type of electronic control which ensures that each weld will begin at the same point on the wave cycle, particularly important for consistent spot welds in automatic operation.

Tack Welding - Very short welds made at each end and the middle of a seam, to hold the pieces in alignment before a full seam is welded. When designing and building a project, it's important to tack-weld everything first before finish welding. You may want to change something at the last minute and tack-welds can be easily broken off or cut through to make changes.

Tempering - Steels can be hardened by heat-treating but may become too brittle. Tempering heats a metal to some degree below the transformation temperature and then cools it to bring back some toughness in the metal.

Tensile Strength - A measure of a metal alloy's (or welded seam's) ability to resist being pulled apart.

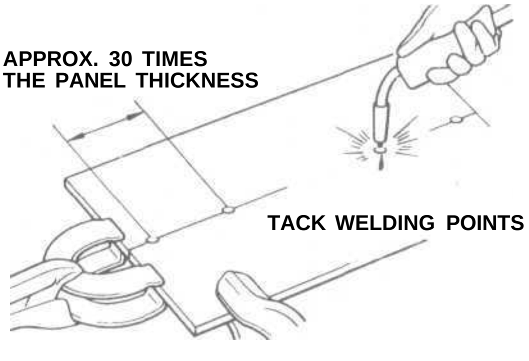

Tack welding

TIG Welding - This stands for Tungsten Inert Gas welding (see heli-arc).

Titanium - An expensive, strong, lightweight material used in aerospace and race cars. Special techniques are required to weld titanium.

Transformer - A major electronic component of electric welding machines, it turns high-voltage, low-amperage line current into low-voltage, high-amperage welding current.

Tungsten - The non-consumable electrode used in TIG welding. It is resistant to corrosion and doesn't melt until 6170° F.

Glossary of welding terms

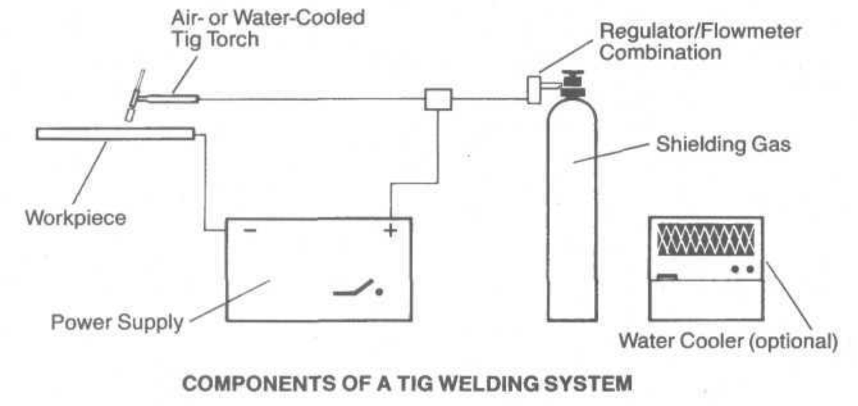
Air- or Water-Cooled
Tig Torch
Regulator/Flowmeter
Combination
Shielding Gas
Workpiece
−
+
Power Supply
Water Cooler (optional)
COMPONENTS OF A TIG WELDING SYSTEM

TIG welding

Undercut - A weld defect where too much heat was applied and at the sides of the bead the parent metal is eroded. Usually accompanied by too much penetration on the backside of the weld.

Voltage - A characteristic of electrical current that equals the amps times the ohms (resistance). In welding machines, input current is usually described in volts, while welder output is usually measured in amps.

Welder - The machine that makes the welded.seam.

Weldor - The person who operates the welder.

Wrought Iron - A soft form of iron that includes some slag in its formation. Traditionally the material chosen for making ornate fences, gates and horseshoes because of its shapability.

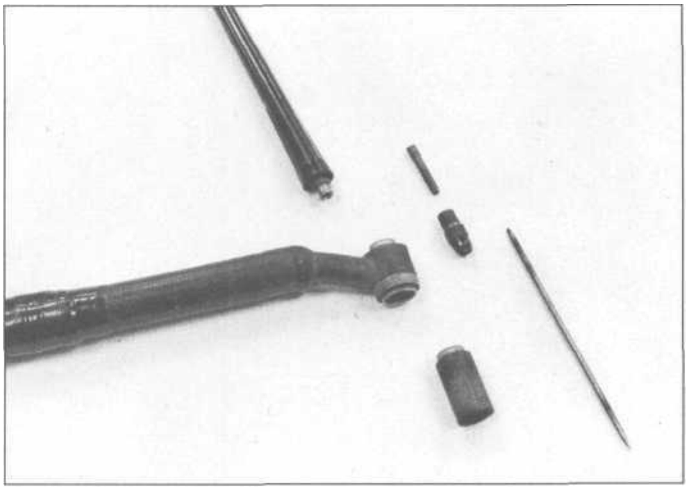

Tungsten

Index

Alloys, metal, 2-3
Angle clamp, 8-12
Anti-heat paste, 8-12
Arc welding, 2-9, 4-1 through 4-16
  arc welders, AC or DC?, 4-4
  beginning, 4-8
  choosing electrodes, 4-13
  safety considerations, 4-7
  the arc process, 4-6
  types of joints, 4-11

B

Bandsaw, portable, 8-5
Beginning arc welding, 4-8
Belt sander, 8-8
Bender, 8-7
Blanket, welding, 8-4
Brazing, 3-16
Building a utility trailer, 9-1 through 9-10

Checking your welds (oxy-acetylene), 3-14
Choosing a MIG welder, 5-5
Choosing electrodes (arc welding), 4-13
Choosing wire (MIG welding), 5-14
Clamp, magnetic, 8-5
Cleco clamps, 8-8, 8-9
Copper spoon, 8-4
Cutoff saws, 8-6
Cutting (with gas cutting torch), 3-18

Definitions of welding, 1-2
Development of modern welding, 1-3
Disc sander, 8-7
Duty cycles, 2-17, 4-3, 5-7

Electrode rack, 8-7

Flame adjustment, oxy-acetylene welding, 3-9
Flange punch, 8-9

Gas welding/cutting, 2-6, 3-1 through 3-26
  checking your welds, 3-14
  cutting, 3-18
  equipment, 3-3
  flame adjustment, 3-9
  heating, 3-24
  how to, 3-9
  the process, 3-1
  welding with filler rod, 3-11
Gloves, welding, 8-4
Grinder, 8-5
Gun, spot welding, 8-5

H
Heli-arc welding (TIG), 6-1 through 6-18
Helmets, welding, 8-2
How welding works, 2-2

I
Introduction, 1-1 through 1-4

M
Magnetic blocks, 8-12
Materials rack, 8-11
Metal alloys, 2-3
MIG (wire-feed) welding, 2-11, 5-1 through 5-26
  choosing a MIG welder, 5-5
  choosing a shielding gas, 5-12
  choosing wire, 5-14
  description, 2-11
  learning, 5-16

N
Nibbler, 8-9
Notcher, tubing, 8-5

Oxy-acetylene gas welding/cutting, 2-6, 3-1 through 3-26
  checking your welds, 3-14
  cutting, 3-18
  equipment, 3-3
  flame adjustment, 3-9
  heating, 3-24
  how to, 3-9
  the process, 3-1
  welding with filler rod, 3-11

Plasma arc welding and cutting, 2-18, 7-1 through 7-12
  choosing plasma cutting equipment, 7-5
  cutting, 7-3
  how to use, 7-8
  welding, 7-2
Post-flow (TIG welding), 6-6
Practice and training, 2-21
Punch, flange, 8-9

Rack
  electrode, 8-7
  materials, 8-11
Rewiring for an arc welder, 4-5

Safety & shop equipment, 8-1 through 8-12
Safety considerations, arc welding, 4-7
Sandblaster, 8-12
Sander
  belt, 8-8
  disc, 8-7
Saws
  bandsaw, portable, 8-5
  cutoff, 8-6
Slope control (TIG welding), 6-6
Spoon, copper, 8-4
Spot weld remover, 8-6
Spot welding gun, 8-5
Stud setter, 8-10

Table, welding, 8-4
TIG {heli-arc) welding, 2-14, 6-1 through 6-18
  aluminum, 6-13
  equipment, 6-4
  how to, 6-8
  post-flow, 6-6
  slope control, 6-6
Trailer, utility, building, 9-1
Training and practice, 2-21
Tubing bender, 8-7
Tubing notcher, 8-5
Types of welding, 2-1 through 2-22

U

Utility trailer, building, 9-1

W

Welding
- Arc welding, 2-9, 4-1 through 4-16
  - arc welders, AC or DC?, 4-4
  - beginning, 4-8
  - choosing electrodes, 4-13
  - safety considerations, 4-7
  - the arc process, 4-6
  - types of joints, 4-11
- blanket, 8-4
- definitions of, 1-2
- development, 1-3
- helmets, 8-2
- how it works, 2-2
- MIG (wire feed) welding, 2-11, 5-1 through 5-26
  - choosing a MIG welder, 5-5
  - choosing a shielding gas, 5-12
  - choosing wire, 5-14
- oxy-acetylene (gas), 2-6
- plasma arc, 2-18
- safety considerations, arc welding, 4-7
- table, 8-4
- TIG (heli-arc) welding, 2-14, 6-1 through 6-18
  - aluminum, 6-13
  - equipment, 6-4
  - how to, 6-8
  - post-flow, 6-6
  - slope control, 6-6
- types of, 2-1 through 2-22
- wire-feed, 2-11
- with filler rod (oxy-acetylene), 3-11

Sourcelist

Airco Gas and Gear
BOC Gasses
575 Mountain Avenue
Murray Hill, NJ 07974

Daytona **MIG**
1821 Holsonback Drive
Daytona Beach, FL 32117

Eastwood Company
580 Lancaster Avenue
Malvern, PA 19355

HTP America
261 Woodwork Lane
Palatine, IL 60067

Lincoln Electric Co.
22801 St. Claire Avenue
Cleveland, OH 44117-1199

L-TEC Welding & Cutting Systems
The ESAB Group
P.O. Box 100545
Florence, SC 29501-0545

Miller Electric Mfg. Co.
718 S. Bounds Street
Appleton, WI 54911

10445 (2110)

Haynes TECHBOOK

WELDING MANUAL

Basics of Gas, Arc, MIG, TIG, and Plasma Welding & Cutting

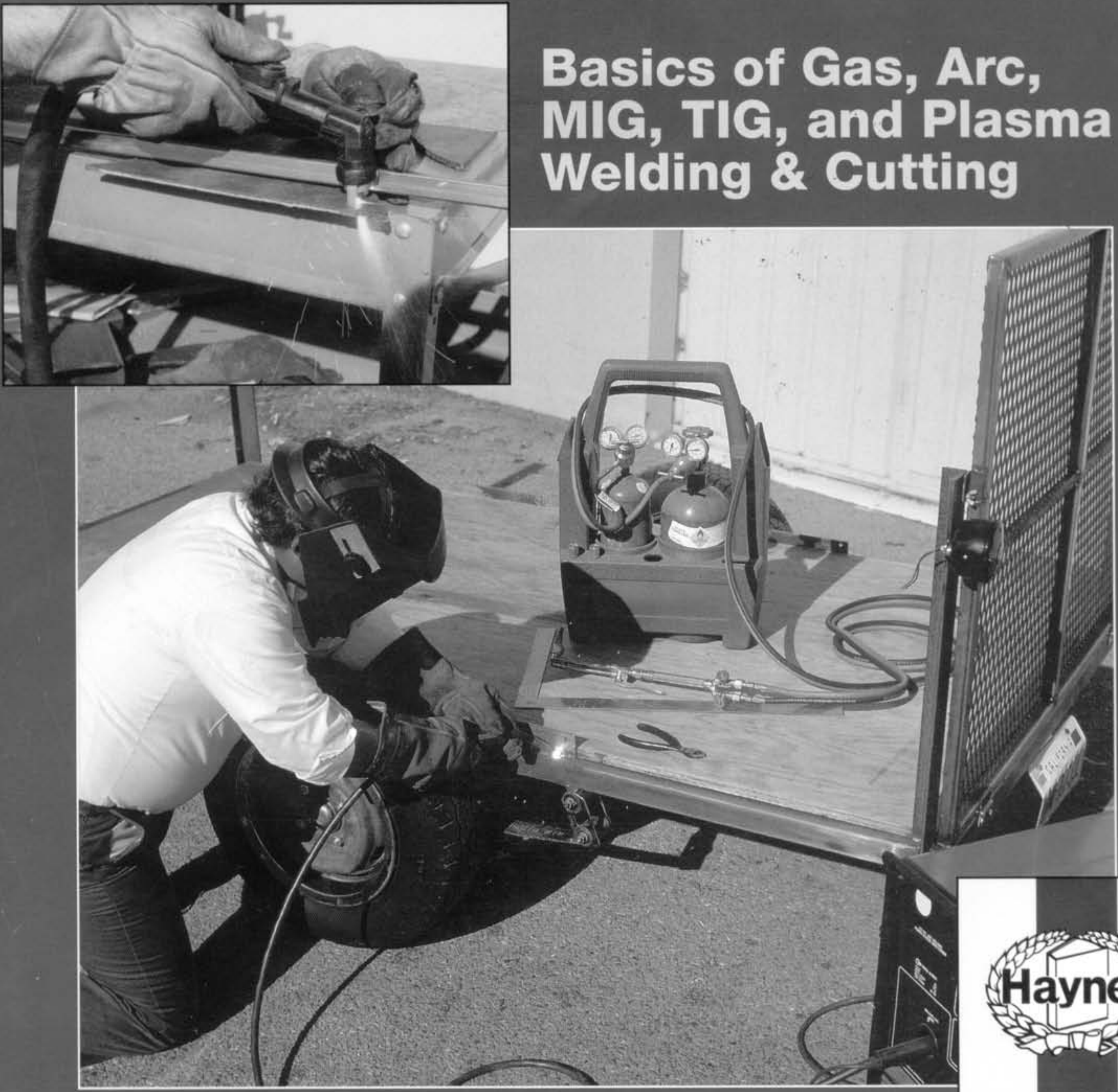

Haynes

TECHBOOK

WELDING MANUAL

Types of welding

Haynes Welding Manual

Building a utility trailer

**The Haynes TECHBOOK Series ...
written for the do-it-yourselfer,
good enough for the pro!**

Included in this manual are:

- **Over 350 photographs**
- **Choosing welding equipment**
- **Oxy-acetylene welding, heating and cutting**
- **Arc welding**
- **MIG (wire-feed) welding**
- **TIG (Heli-arc) welding**
- **Plasma-arc welding and cutting**
- **Safety and shop equipment**
- **Building a utility trailer**

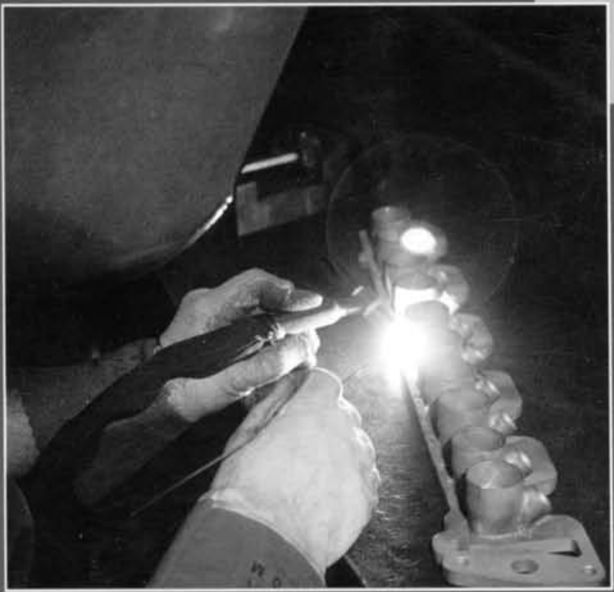

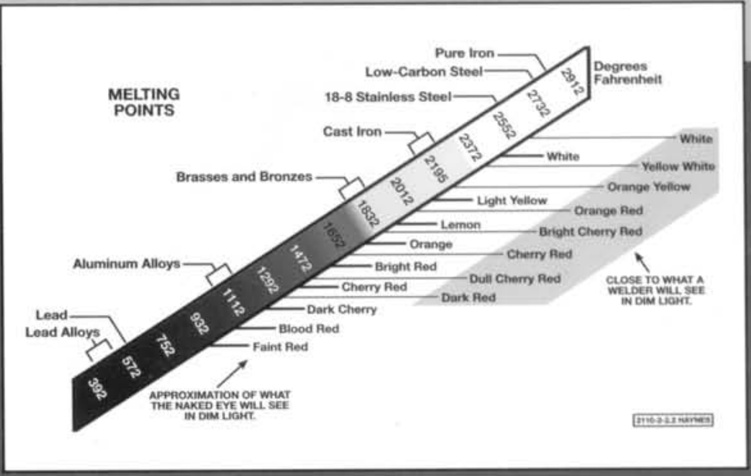
MELTING POINTS
Pure Iron
Low-Carbon Steel
18-8 Stainless Steel
Cast Iron
Brasses and Bronzes
Aluminum Alloys
Lead
Lead Alloys
Degrees Fahrenheit
2912
2732
2552
2372
2195
2012
1832
1652
1472
1292
1112
932
752
572
392
White
Yellow White
Orange Yellow
Orange Red
Bright Cherry Red
Cherry Red
Dull Cherry Red
Dark Red
White
Light Yellow
Lemon
Orange
Bright Red
Cherry Red
Dark Cherry
Blood Red
Faint Red
CLOSE TO WHAT A WELDER WILL SEE IN DIM LIGHT.
APPROXIMATION OF WHAT THE NAKED EYE WILL SEE IN DIM LIGHT.
2110-2-2.2 HAYNES

TECHBOOK

Haynes

ISBN 1 56392 110 3

The Haynes
Welding Manual

by Jay Storer
and John H Haynes

Member of the Guild of Motoring Writers

The Haynes Manual
for selecting and using welding equipment

ABCDE
FGHU
KLMNO
PQRS

Haynes Publishing Group
Sparkford Nr Yeovil
Somerset BA22 7JJ England

Haynes North America, Inc
861 Lawrence Drive
Newbury Park
California 91320 USA

Haynes Welding Manual

Acknowledgements
We are grateful for the help and cooperation of welding equipment manufacturers such as Airco, Daytona MIG, HTP America, Lincoln Electric, Miller Electric and the ESAB Group, for their assistance with technical information and illustrations. Thanks to FitzMaurice-Smith Racing in Ventura for assistance with cover photography. We also wish to thank the Eastwood Company, Maguire's Welding and Valley Vintage Rods for assistance.

© **Haynes North America, Inc. 1994**

With permission from J.H. Haynes & Co. Ltd.

A book in the Haynes Automotive Repair Manual Series

Printed in the U.S.A.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage or retrieval system, without permission in writing from the copyright holder.

ISBN 1 56392110 3

Library of Congress Catalog Card Number 94-74485

While every attempt is made to ensure that the information in this manual is correct, no liability can be accepted by the authors or publishers for loss, damage or injury caused by any errors in, or omissions from, the information given.

94-176

Contents

Chapter 1 Introduction

Definition of welding 1-2
Development of modern welding 1-3
Welding today 1-6

Chapter 2 Types of welding

How it works 2-3
Metal alloys 2-3
Oxy-acetylene gas welding 2-6
Arc welding 2-9
MIG (wire-fed) welding 2-11
TIG (heli-arc) welding 2-14
Duty-cycles 2-17
Plasma-arc welding and cutting 2-18
Practice and training 2-21

Chapter 3 Oxy-acetylene gas welding/cutting

The basic gas process 3-1
The equipment 3-3
Getting started 3-7
Flame adjustment 3-9
Gas welding 3-9
Welding with filler rod 3-11
Brazing 3-16
Oxy-acetylene cutting 3-18
Heating with oxy-acetylene 3-24

Chapter 4 Arc welding

Comparing duty-cycles 4-3
AC, DC or both? 4-4
Rewiring for an arc welder 4-5
The arc process 4-6
Safety considerations 4-7
Beginning arc welding 4-8
Types of joints 4-11
Choosing electrodes 4-13

Chapter 5 MIG welding

Shopping for a MIG welder........5-5
Choosing shielding gas........5-12
Choosing wire........5-14
Learning MIG welding........5-16

Chapter 6 TIG welding

The equipment........6-4
The process in action........6-8
TIG-welding aluminum........6-13

Chapter 7 Plasma cutting/welding

Plasma-arc welding........7-2
Plasma-arc cutting........7-3
Choosing plasma cutting equipment........7-5
Using a plasma cutter........7-8

Chapter 8 Safety and shop equipment........s 1

Chapter 9 Building a utility trailer........9-1

Glossary........QL-I

Sourcelist........sn

Index........IND-1

1 Introduction

Welding is a process critical to our present state of civilization and technical advancement, yet little understood and most often taken for granted. Unless exposed to the building, machinery or automotive trades, the average person never realizes how much we depend on the welding process, which is a fundamental part of the process of building most of what we depend on daily, including vehicles, buildings, appliances, bridges and a great deal more. In fact, once you really start to examine the objects around us, it's hard to imagine our world without the welding process.

Architecturally speaking, we might all be living in one-room wood or adobe-brick houses if it weren't for welding. Certainly all large commercial and residential structures are built with a considerable "skeleton" of welded structural steel, and even most single-family, wood-framed houses are built using some welded components, even down to items like the electrical outlet boxes in the walls. Anyone who has watched the construction progress of a major highway improvement like a bridge or tunnel has seen the helmeted weldors, unsung heroes of the construction process, spraying a shower of sparks from high on a scaffold while they join metals to hold critical loads.

Sitting in an airport terminal recently with some time on our hands gave us something to think about. Virtually everything around us involved welding in some way. There was a large rack of telephone books in stainless-steel racks, each carefully TIG-welded and sanded, a post office box that was made of welded steel, the telephone stall had welded components, and the seats we were sitting on were part of a welded-steel structure that held eight seats. Everywhere you look in the modern world, you'll find examples of how widespread and important is the use of welding techniques and equipment.

Much of our Haynes audience is familiar with the automotive world, and here is a field where construction and repairs made by welding are absolutely essential, in fact essential to virtually all forms of transportation, from bicycles to cars, trucks, trains, aircraft and space vehicles. Even if we could go back before the "horseless carriage" was developed at the end of the 19th century, we would still need some form of welding to return to horse-drawn transportation, in welded brackets for harnesses and wagon components.

1.1 Welding had its ancient origins in the fires of blacksmiths, who could forge two white-hot pieces of metal together with hammer blows and patience. It remained for history to bring us to electricity and bottled gasses for welding techniques to develop any further. Modern-day farrier (blacksmith/horseshoer) Richard Heller illustrates the old ways.

1.2 Metal cutting in the old days was no easier. A white-hot piece of metal was laid over a hardy (like a wedge or chisel point) in the big anvil, and struck hard with a hammer, chopping the softened piece off. Final shaping of most items was by hand with stones and files.

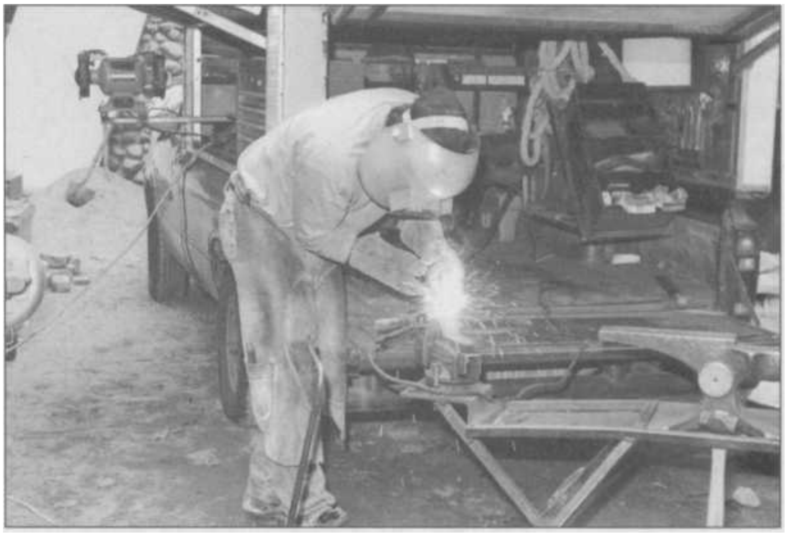

1.3 The farrier of today works out of a truck, filled with equipment like a grinder, drill press and welder, all running on AC power when available; otherwise, horseshoes are shaped in a portable forge. Today's blacksmith has the advantage of tools like this MIG welder that speed up making special horseshoes.

Since we've established that welding is a technical process our present society can't live without, in this book we'll explore more about the process, removing the mystery, and examining the available equipment as it can be useful to us in fabricating with metal or repairing metal items. We'll show all of the most up-to-date equipment, describe the various welding types, make recommendations, show how welding is used in everyday situations, and develop welding projects that illustrate how you can build or repair automotive, farm and household objects.

1.4 Under the center bridge column here are a welder (a machine) and a weldor (a person). He is welding up steel enclosures for the concrete columns, to retrofit more strength in California bridges to meet newer earthquake codes.

Definitions of welding

At one time, the simple definition of welding was "joining metals through heating them to a molten state and fusing them together." As technical progress in welding processes has advanced, the definition has had to change.

There are now two basic forms of welding: fusion and non-fusion. The former is the most common, and it involves the actual melting of the parent metals being joined. Not all welding today involves melting. Non-fusion welding is most commonly represented by soldering and brazing, two processes of joining metals where the parent metal is heated, but not melted, and a second or "filler" metal is melted between them, forming a strong bond when all are cooled.

Pressure and friction alone can weld metals together, such as when a machinist turns down a piece of metal in a lathe. Often, pieces of the metal chips can become welded to the cutting

1.5 Wherever you see buildings go up, the basis for them is always steel, whether they wind up with wood, bricks or cement on the *outside,* it all began with the work of ironworkers and weldors.

1.6 The construction and repair of heavy equipment and farm machinery would be nearly impossible without welding. Arc welding is usually used, as here modifying the front of a dirt scoop on a John Deere, where welding is outdoors and on heavy plates.

tool, which is a simple example of a process that can be used in production work in joining metals. Other kinds of "cold" welding may today involve sound and light, as in sonic-welding or laser-welding.

Today, the term "welding" has even been applied to the processes of joining non-metallic materials, such as plastic-welding which sometimes involves a fusion of materials as a result of heat or chemical action. As kids, we have all played with plastic models that we constructed using "glues" that would react with and actually "melt together" the two pieces we were joining.

For today's definition of welding to be all-encompassing, it would have to read "the joining of metals and plastics without the use of fasteners." This definition covers a lot of ground, but, given the interests and needs of the majority of our readers, this book will concentrate on welding as it applies to metals joined by heat processes produced either by a flame or electrical current.

Semantically speaking, throughout this book we will be illustrating and referring to pieces of welding equipment and to the people who operate them. Sometimes the same term "welder" is used to apply to both the machine and the man, which can become confusing, so for our purposes, we will from now on refer to the machine as a "welder" and to the person operating it as a "weldor."

Development of modern welding

Welding can trace its roots far back in time to the first blacksmiths who heated and shaped metals. At that time, metals were primarily used for tools and weapons, both of vital importance in those days. The blacksmith was an important tradesman in any community, earning a little more respect than most, even by lords and kings who depended on weapons for maintaining power. The "art" of smithing was understood by a select few, and blacksmiths were accorded almost the fear-based respect of a low-level sorcerer. Given the traditional image

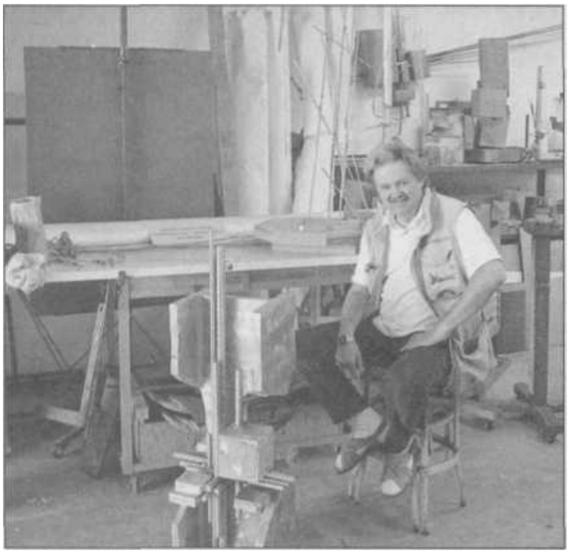

1.7 Welding is even used in many art forms today. Bill McKewen is a metal sculptor who works with rods, tubes, plates, wires and custom castings to create interesting works that combine hard metals with "organic" elements.

1.8 One of Bill's largest works is this 30-foot tall sculpture called "Organic Form 1", which is made entirely of stainless-steel and weighs over 1000 pounds.

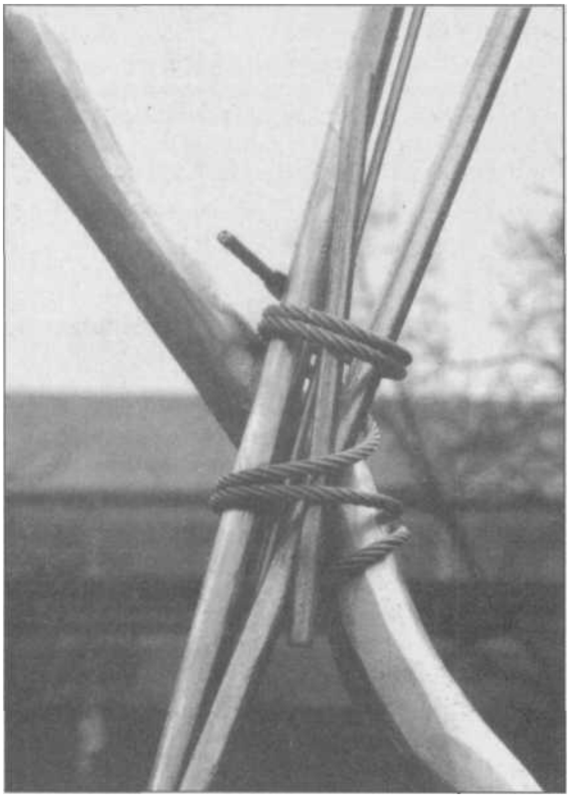

1.9 In a close-up of the "transition point" in Bill's sculpture, we can see how the stainless tubing and solid rod is wrapped with stainless-steel oil field cable and welded to a large shape originally carved of foam and then sandcast in stainless and final-shaped with a body grinder.

of the blacksmith as a large, muscular man covered with soot who works in a fiery, smoky environment, hammering loudly and magically making useful objects out of nothing, it's no wonder he was treated a little differently.

The blacksmith heated metals in a wood fire (coal was used later on) and hammered them into tools and weapons, performed basic heat-treating to harden some areas, and ground sharp edges with a foot-powered stone grinding wheel.

There were times when an object couldn't be made from one piece of metal. Gradually, the techniques developed to join pieces of metal either with bolts, hot rivets, or welding. This first use of welding consisted of heating the objects to a certain color (fairly precise indicator of the temperature) and quickly hammering them together on the anvil. The heat and pressure joined the items, and the process has been called "forge-welding."

As history marched on, larger and larger items had to be made of metal, especially with the industrial revolution of the 19th century. Most machinery was made of cast metal, produced when molten metal was poured into a mold and

1.10 In high-technology areas like nuclear vessels, aviation and here, at a large race car-building shop, quality welding technology is critical to success with metals.

1.11 Not every welder and weldor works indoors. This is an oil field service truck with gas-welding/cutting equipment, arc welder and crane. Many welding shops have portable setups like this for remote jobs.

allowed to solidify. Repairing broken castings was a common procedure in industry and manufacturing. Cast metal is too brittle for forge-welding on an anvil, and the items were too large, so a process of "cast-welding" was developed in which the broken machinery was heated, a temporary mold bolted around the area to be repaired, and molten metal was poured in. Done right, the molten metal bonded with the parent casting and the goldrush mine or cotton gin was back in business.

As the 20th century dawned, electricity came into wider usage, especially for lighting. The early carbon-arc lamps made as much heat as they did light, and someone started using electric carbon-arc rods to fusion-weld metals. Soon after, the simple stick electrode was developed that is similar to arc-welding rods of today, and about that same time oxy-acetylene gas welding was also developing. It's ironic that gas welding also grew out of the advancements in lighting, since acetylene gas had been used for car lights up until just before W.W.I.

1.12 The back of the same oil field welding truck is built as a large, heavy workbench-away-from-a-shop. Big Lincoln ac/dc arc welder has its own engine power and generator to run lights for night work.

Speaking of W.W.I, anyone who is familiar with history knows that many technological advances have resulted from the accelerated development that, unfortunately, only seems to come from a wartime environment. W.W.I saw the further development of gas and stick welding, and soon after that the refinement of X-ray technology aided further industrial use of welding. The exacting inspection of welds made possible by X-rays speeded welding's acceptance.

The rapid development of the aircraft industry between the wars, and the increased use of lightweight metals to replace wood and fabric in fuselages and wings led to further advances in welding. W.W.II saw the burgeoning aircraft industry with increased requirements for joining light metals such as aluminum and magnesium faster, stronger and smoother than with drilling and riveting, as had been used before. Inert-gas welding was invented before the war, but the gasses were considered too expensive at the time, and it took a war-time environment to accelerate its development. After the war, because the lightweight metals were in demand for many military and civilian applications, TIG welding was further refined, and MIG welding was invented around 1948.

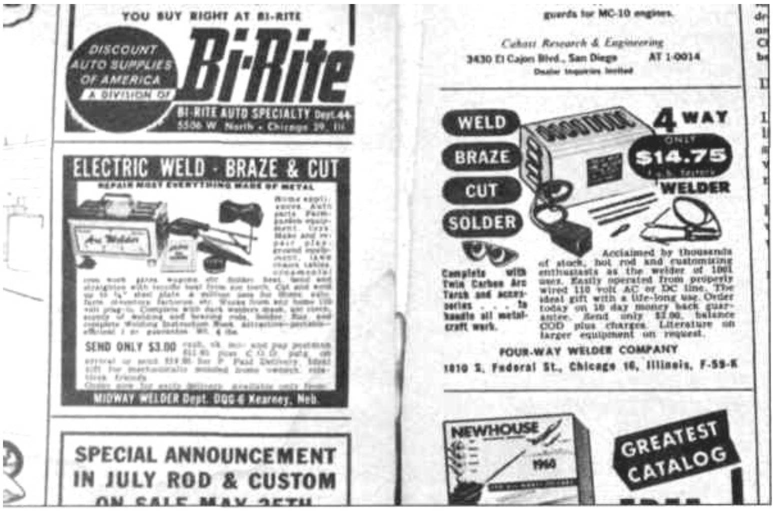

YOU BUY RIGHT AT BI-RITE

DISCOUNT AUTO SUPPLIES OF AMERICA A DIVISION OF Bi-Rite

BI-RITE AUTO SPECIALTY Dept.44
5506 W North • Chicago 39, Ill

ELECTRIC WELD - BRAZE & CUT

REPAIR MOST EVERYTHING MADE OF METAL

SEND ONLY $3.00

MIDWAY WELDER Dept. DGG-6 Kearney, Neb.

SPECIAL ANNOUNCEMENT
IN JULY ROD & CUSTOM
ON SALE MAY 25TH

guards for MC-10 engines.

Cabasi Research & Engineering
3430 El Cajon Blvd., San Diego AT 1-0014
Dealer inquiries invited

WELD
BRAZE
CUT
SOLDER

4 WAY ONLY $14.75 WELDER

Complete with Twin Carbon Arc Torch and accessories . . . to handle all metal-craft work.

Acclaimed by thousands of stock, hot rod and customizing enthusiasts as the welder of 1001 uses. Easily operated from properly wired 110 volt AC or DC line. The ideal gift with a life-long use. Order today on 10 day money back guarantee. Send only $2.00, balance COD plus charges. Literature on larger equipment on request.

FOUR-WAY WELDER COMPANY
1010 S. Federal St., Chicago 16, Illinois, F-59-K

NEWHOUSE 1960

GREATEST CATALOG

1.13 Back in the Fifties and Sixties, car enthusiast magazines were all advertising these "twin carbon-arc" welders, which were the original "Buzzboxes." While some of these did *work,* most of them were tried a few times and then put on a shelf out in the garage. Today they're collector's items from welding history.

SERVING INDUSTR
GENUINE HELI

1.15 If you decide to purchase a welder, visit your local welding supply center to talk to the salespeople and find out what equipment best suits your needs and budget. A well-equipped store will have everything you'll need, from gasses to glasses.

1.14 Today's modern home/shop/garage welding equipment is likely to be a clean, efficient, safer welding system like MIG, or wire-feed, welding. Machines like this are getting less expensive all the time, and some models can be hooked to standard household 110V current, making them quite portable.

Today

In the remaining decades of the 20th century, welding developments have come at a rapid pace, often closely tied to electronics development, as new and better methods of applying cleaner and more controlled heat have come along. In the past decade, we have seen the cost and complexity of welding equipment reduced in some areas, to the point where equipment that was once considered solely the province of the high-production professional shop is now found in garages and hobby shops around the country.

It's been good news for the automotive hobbyist, whether he is restoring an old car or building a race car, as well as good news for the small farmer trying to maintain and repair his equipment. Welders are now being purchased by metal-sculptors and other artists and craftsmen, as well as being used for the most everyday kinds of household jobs such as repairing a bicycle or garden tool, building a firewood storage rack or a barbecue, or even fabricating a small utility trailer at home.

Welding has been compared to playing a musical instrument. In the same way that anyone can pick-up a harmonica and start making sounds with it, most anyone with some mechanical aptitude can, with a little practice, start using popular kinds of modern welding equipment for non-critical jobs. But to make *music* with that harmonica, or good, strong, clean welds with a welder, will take time. As the hip musician once said when asked how to get to Carnegie Hall, "Practice, man, practice!" The more time spent consistently practicing with a welding setup, whatever the type, the better your results will be. If you use your welder only occasionally, don't just pick-up the torch and expect to lay a perfect bead the first time. There is a rhythm to find. Practice first on a scrap piece of metal of the same thickness as the work you intend to weld, and, when you have the rhythm down, repair or fabricate your job.

Though not considered with the same awe as the ancient blacksmith, a professional weldor today can still make a satisfying living, for it seems to be a skill that, far from fading out due to replacement technology, is seeing ever-increasing usage in business and industry. Weldors today find diversified employment in oil field and pipeline work, building construction, bridges and other infrastructure, automotive work from assembly-line welding to body shop repair to race car fabrication and antique auto restoration, to the nuclear-power industry, aviation and aerospace work, defense work, and manufacturing work building products from household appliances to huge boilers and construction equipment.

This book is intended as an overall introduction to the welding process, illustrating most of the common equipment and work techniques for both home and shop welding. While this is not a textbook for the would-be professional weldor, there is enough of an overview here to give a prospective welding student a basic understanding before delving into the more detailed professional textbooks on the subject.

Whether you plan to restore a vintage car, build a race car or experimental aircraft, construct your own wrought-iron fence, sculpt a metal art masterpiece, or go into welding professionally, you'll find this book of interest. The handy Source List and Glossary of Terms at the end of the book will be helpful for future reference.

Notes

2 Types of welding

If you are reading this book, chances are you have had some exposure to welding through watching a repair done such as having a new exhaust system put on your car, through some hobby interest in the metal arts/crafts area or through some industrial exposure to welding as used in manufacturing and building processes. Obviously you have become interested enough in learning about welding to purchase this book which we feel is an excellent introduction to a field where there are lots of involved textbooks for the person pursuing welding as a profession, but few *basic* books for someone getting started at the hobby, farm or home/shop level.

Perhaps your initial exposure to welding has sparked an interest in doing it yourself. If you are involved in automotive work, you already know how valuable the process can be in fabrication and repairs. Once you have seen it performed, you realize how *handy* this capability is. You can join pieces of metal to either repair something that was damaged and otherwise scheduled for replacement or build something entirely new, from a barbecue grille to a race car.

Once you have the basic skills and the right equipment, you'll find many more uses for welding than you had anticipated. Like a good truck or a specialized tool, once you have a welder, you'll wonder how you ever got along without it! You'll probably find yourself building a materials rack, stocking it with various sizes of tubing and plates, and actually looking for new projects to tackle, from building a workbench to last a lifetime, to storage racks, moveable shop carts, engine stands, shelving, and much more.

There are quite a few types of welding processes, and today there are a great many welders to choose from. There are so many, in fact, that just picking the process and the ma-

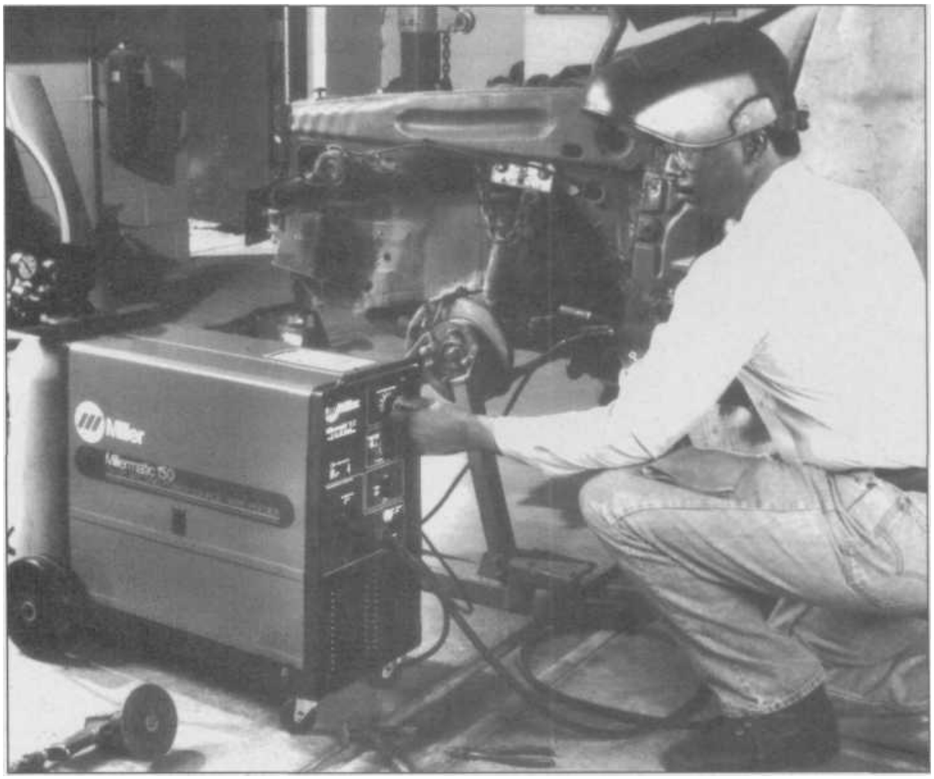

2.1 There are a number of choices today when shopping for welding equipment. The right choice for you depends on the kind of jobs you plan to do, where you plan to do them, your budget and how much time you can devote to training and practice before you become proficient at one of the methods we'll be looking at.

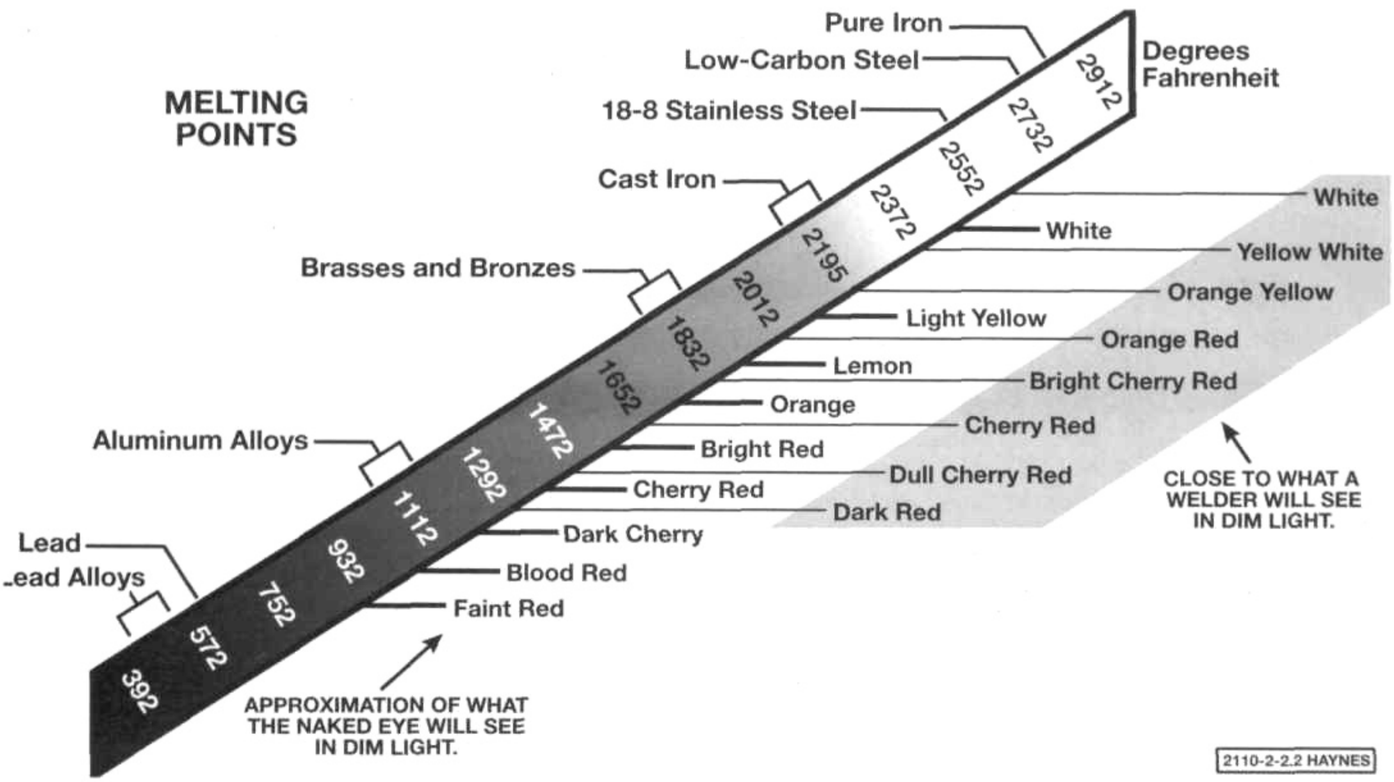

2.2 Welding is a process of *fusion,* in which metal parts are heated to the melting point and fused together, usually with a filler of the same material melted along with the "parent" metal. All metals melt at different temperatures, and this chart shows some interesting comparisons, as well as relating the heated metal's color to its temperature.

chinery which best suit your needs may seem a considerable task. This chapter will give you a brief overview of the various processes along with an analysis of the pros and cons for each type and how best to select equipment based on *your* needs.

How it works

The most basic principle of the welding process is joining two pieces of metal together (or at least two edges of the same piece, in the case of repairing a crack). This is generally accomplished by heating the metals to be joined until they become liquid or molten and the two edges fuse together. Most often, the complete joining of the two metal edges is accomplished by melting new metal into the joint at the same time. The new metal added to form a fused welding joint is called filler metal, while the original pieces being joined are called the parent metal. Together they form a welded "bead" of filler and parent metal that is usually thicker than the parent metal. Depending on the skill of the weldor and the type of welding, two pieces of metal can be joined in such a way that with a little filing or sanding of the bead, the joint is virtually undetectable, a particularly important aspect when making automotive body repairs. The first time you may have observed a professional weldor working, the process may have seemed like a sorcerer doing alchemy with a magic wand. With the proper equipment and practice, you can do a little magic yourself, a magic that can give tremendous personal satisfaction, as well as save you considerable expense compared to having the same work done at a professional fabrication shop.

2.3 The welding corner of this well-equipped race-car prep shop holds a variety of equipment, from a large TIG machine at left (with cooling unit on top), to a heavy-duty plasma cutter, oxy-acetylene cutting/heating/welding outfit, and two small plasma cutters. Not shown here are the two MIG welders in use elsewhere in the shop.

KLY SPECIAL
PER POUND
REMS
$3.00 LB

2.4 You will have to develop some new contacts once you get into welding, one of which will be a reliable local source for metals for your projects. This is ABC Metals (aluminum, brass and copper) in Oxnard, California. They have a lot of aluminum here at scrap prices, but it's more of a gold mine for the hobbyist or car-builder who needs relatively small bits and pieces and not a trainload. Knowledgeable salespeople at a metal yard can be very helpful.

It takes a tremendous amount of localized heat to weld metals together, and heat control is the key to welding properly. Every material has its own specific melting point, and to make a weld you need to heat the material to that point but not beyond it. Visualize an ice cube, which is solid material (when cold). If you heat it to the melting point (above 32 degrees F), the solid becomes a liquid (water), heating it further will vaporize it into steam and for your purposes the material is *gone.* The same changes happen to metal, although at much higher temperatures. Common lead solder such as you might use to solder electrical connections can melt at temperatures from 250-750° F (depending on the alloy), aluminum melts at just below 1250° F, and common mild steel melts at 2750° F.

The heat required to make metal molten enough to fusion-weld can be achieved in several ways, but the most common for home/shop situations will be generated either with a flame or some use of electrical current. The traditional source in welding has been the oxy-acetylene torch, while electricity is now used in most of the other methods, such as arc-welding, MIG-welding, and TIG-welding.

One thing that is common to all the forms of welding is that the filler material must be compatible with the parent metal, and all efforts must be made to make a "clean" weld free of outside contaminants that could weaken the joint. If you are welding aluminum, the filler rod must be aluminum, a stainless filler rod must be used for welding stainless-steel and steel rods are used on steel. In gas welding, the cleanliness of the weld is controlled by the correct adjustment of the torch flame and the cleanliness of the two edges of the parent metal. In electric welding, an inert gas "cloud" is formed right around the welding area that keeps outside oxygen or impurities from contaminating the weld. The shielding gas is generated in several ways, as you'll see as we further describe the various types of equipment.

Metal alloys

The melting point of the metal you work with will vary with the basic nature of the material (iron, steel, aluminum, magnesium, etc.), and the alloy of the metal. Most metals today are not in pure form, they are alloyed or mixed with another metal to give the new material special characteristics. Copper, lead and iron are basic pure metals that have been used by man for tools and other ob-

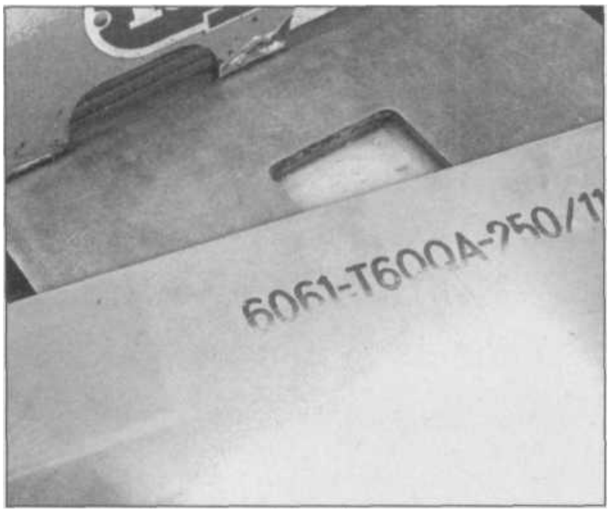

2.5 Non-ferrous metals like aluminum are generally marked with their alloy and heat-treat, such as here on this sheet of 6061-T6. If you need a specific metal for a project, you may have to buy new material rather than remnants, because the smaller scrap pieces may not have the markings on them.

2.6 Weldor Bill Maguire fabricated this recumbent bicycle for himself from 4130 chrome-moly thinwall tubing, which is very strong but light. The idea was to reduce the bike weight, wind resistance and pedal effort. Bill joined the thin tubing with TIG welding.

jects for thousands of years. Mixing various metals together can produce a new metal with new uses. Copper mixed with zinc will make brass, which has strength, reduced cost, and better suitability for machining and casting. The same base copper mixed with tin makes bronze, which was alloyed as far back as 2000 years to make weapons. Gold and silver, precious as they are in their pure state are seldom utilized in their natural form which is quite soft in comparison to other metals.

When alloyed with other metals which add strength or other characteristics, gold and silver can be used for jewelry, coins and many other uses. We commonly describe different gold objects by "carats." While pure gold is 24 carat, 12-carat gold is only half gold and half other metals, and the closer the carat-number is to 24, the more gold is in the object. The other alloys reduce the expense of the pure gold and make it more durable and useful. Were rings and other jewelry to be made of 24-carat pure gold, they would be too soft and not last in normal use.

Most of the metals you will be working with in your welding will be of two kinds, ferrous and non-ferrous. The former includes metals that contain iron, most commonly steel. The most commonly-welded non-ferrous metal is aluminum. Both steel and aluminum can vary considerably in the welding process depending on the alloy. By changing the alloy of either steel or aluminum, different properties can be obtained, to either make the metal more flexible (ability to bend without breaking), malleable (ability to be formed with a hammer), ductile (ability to be drawn out or hammered thin) or to improve its strength for a specific application. Steel is made from refined iron combined with carbon and other elements. How much carbon is added determines the properties of the steel alloy. Most of the steel we might use for projects is relatively low in carbon, called mild steel. With higher levels of carbon, you get medium-carbon steel (used for shafts and axles), high-carbon steel (used for automotive and industrial springs), while very-high-carbon steel is used to make files and sharp-edged cutting tools. The common mild steel we use most often is weldable by virtually all of the techniques described in this book, while the higher-carbon steels have special requirements. Other elements commonly alloyed with steel are manganese, tungsten, nickel and chromium. The latter two combine with steel to make

2.7 Get to know the personnel at your local welding supply store. They can be very helpful when it comes to choosing equipment, and they will have all the supplies you'll need in the future. Most stores carry several brands of equipment, all the safety items, and even the smaller home/shop machines.

2.8 The welding supply store should be able to tell you where to buy steel locally, which you will probably use for most of your welding projects. Most yards have a remnant section as shown, where you can buy short lengths of material at by-the-pound prices. Tubing for a trailer project is being weighed.

stainless-steel, a very useful material that requires somewhat different welding techniques. Anyone familiar with race-car and aircraft construction may have heard of 4130 chrome-moly steel, which is often used in these applications for its high strength relative to its weight. The four-digit number describes the alloy as containing molybdenum, and the amount of carbon. This alloy contains more carbon than mild steel, as well as chromium and molybdenum, both of which add properties of rust-resistance, strength and hardness. Even though this is a higher-carbon steel than mild steel, it really contains only 30/100th of 1% of carbon, which shows how scientific the alloying of metals really is. A tiny change in content can radically affect the properties of the final metal.

In steel and aluminum, not only are there different alloys, but different heat-treat processes. In the simplest terms, heat-treating is a scientific process of heating a metal to a specific temperature and then cooling it, either slowly or quickly, and with or without oil. The heat-treating can affect the hardness and other characteristics of the metal. When aluminum is purchased new in sheets or tubes, it is generally marked with its alloy and heat-treat, such as 3003-T3, which is a sheet aluminum that is considered "half-hard" and is commonly used in making race-car bodywork, where it has to be somewhat strong, but also able to be bent, welded and hammered. On the other end of the spectrum, 7075-T6 aluminum is very hard and strong. Called "aerospace aluminum" in the vernacular, it is often used in making machined aluminum parts and applications where very high strength is required. While it is strong and hard, it doesn't bend. There are volumes of scientific books on the alloying and heat-treating of metals, but, for your purposes as a home weldor, just remember to find out what kind of metal you are welding, and when making a project ask a metals expert to recommend the most suitable material. In general, the higher the carbon content in steel, and the higher the heat-treat on aluminum, the stronger the material will be but tougher to form into a shape, and the tougher metal alloys can be more brittle.

If you do decide to purchase a welder, you will eventually get to know the people at your local source for welding equipment, and they should be of considerable help in answering your questions and getting your setup working well. They will also be able to tell you where locally to purchase the metals you need for your projects. Most metal supply houses have a "scraps" section where you can purchase cut ends, small plates and short lengths of various tubes, all at about half the price of prime material, which is usually sold in large, unwieldy sheets or 20-foot lengths of tubing. Use the scrap pieces for all your experimentation with various types of welds, until you are doing pretty well. You may even be able to build small projects, like a welding cart, using the short lengths of tubing or angle-iron from the scraps section. Once you get to know the people at both the welding supply house and the metal yard, you'll have experts to wade through the technical jargon of alloys and heat-treats, and help you sort out the right material for your future projects.

Oxy-acetylene gas welding

This is perhaps the oldest and most versatile of welding setups. For a long time, it was the only setup recommended for the home/shop use, and has been among the least expensive to get started with. The basic combination in a typical gas-welding package are two high-pressure cylindrical tanks, one for oxygen, one for acetylene, a set of gauges and regulators to control the gas flow out of the tanks, a pair of hoses, and a torch. The torch usually comes with a variety of tips, tip cleaners, a spark lighter, and good sets may include a helmet, gloves and often a cutting-torch.

The latter is what really makes the oxy-acetylene system so versatile. It is one of the few welding systems that can do *cutting* as well as welding. This can be invaluable in both repair and fabrication work. Cutting away damaged or unwanted material is easily done with a properly-used cutting-torch attachment, and you may have many projects where you need to cut an *irregular* shape out of steel plate. The bulk of tubing and angle-iron cutting is usually done with some kind of saw or an abrasive cutoff wheel, but these tools can only make a straight cut; they can't go around corners. If you need to cut out a circle from a steel plate, you can draw the circle on the plate with a compass and a special hard crayon called a soapstone (which leaves a line you can see even when welding), then use your cutting torch to follow the line and you have your part. Any shape can be cut out. If you make up a cardboard template of the piece you need, trace around the pattern with your soapstone onto the plate, and make it.

Cutting with a torch takes skill to closely follow a line, and even then the edges of the metal will require some grinding, filing or sanding to get a smooth edge. Most experienced weldors know just how much to cut *outside* their pattern

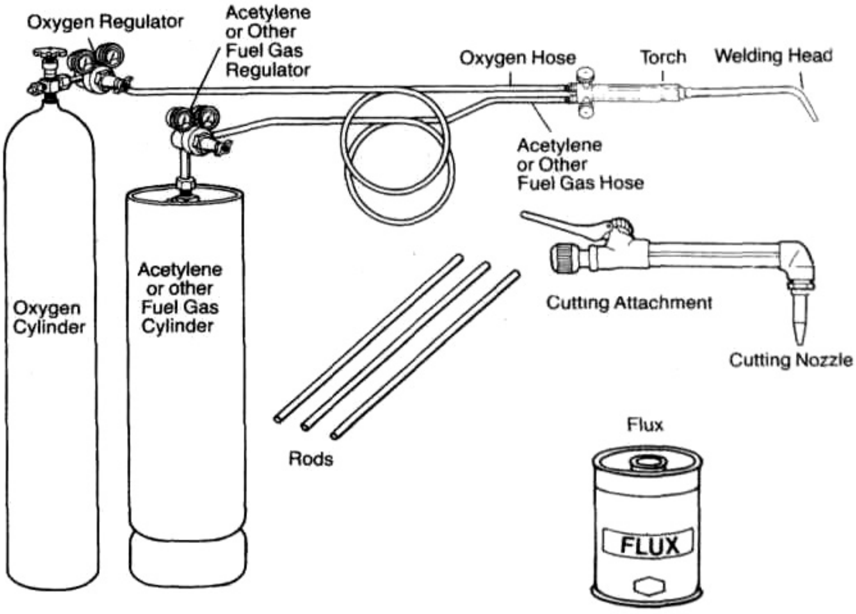
Oxygen Regulator
Acetylene or Other Fuel Gas Regulator
Oxygen Hose
Torch
Welding Head
Acetylene or Other Fuel Gas Hose
Acetylene or other Fuel Gas Cylinder
Oxygen Cylinder
Cutting Attachment
Cutting Nozzle
Rods
Flux
FLUX

2.9 Gas welding with oxygen and acetylene gasses is one of the oldest forms of welding, and is still used today in construction, muffler shops and farm repairs. Its versatility lies in the ability to cut *and* weld, on thick and thin materials, and to do braze-welding as well. The flux shown is only used for brazing. Note the different size and shape of the two gas cylinders.

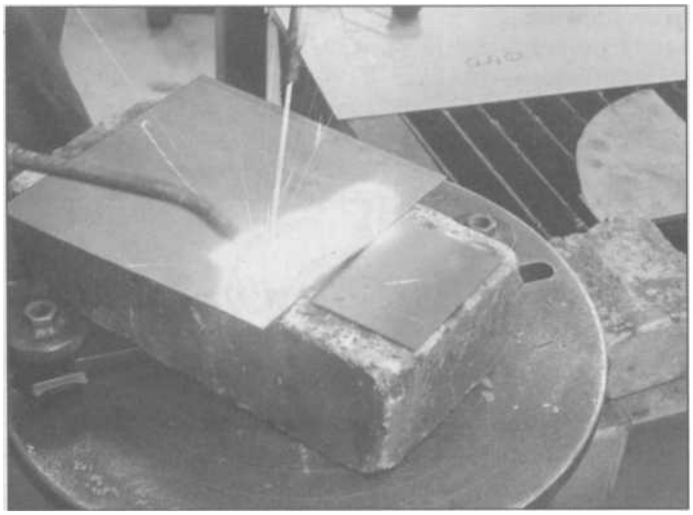

2.10 You'll need a safe area to practice your gas welding. Fire bricks like these are safe to weld on and don't suck the heat out of your parent metal as you weld. The skills you'll have to learn with gas welding are torch movement, even feeding of filler rod with your other hand, and steady, small circles to make consistent puddles.

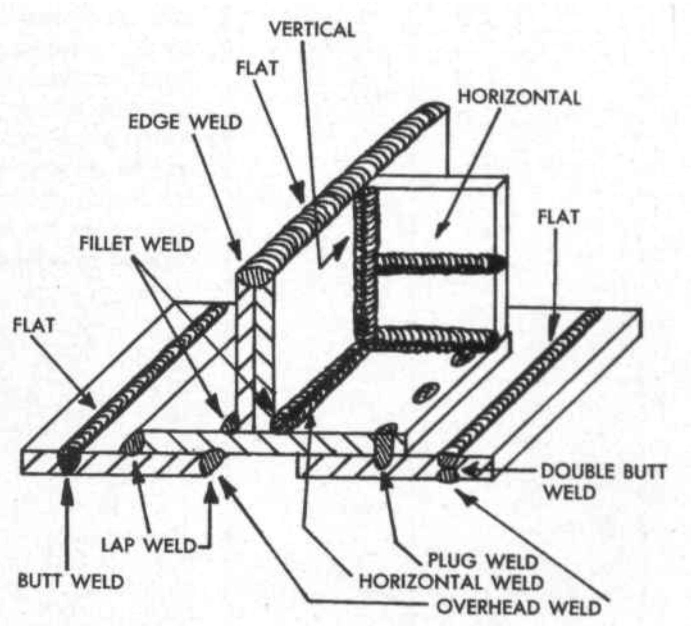

2.11 This illustration shows virtually all of the basic weld types and positions. With each type of welding you practice, you'll first learn to make a steady bead on a steel plate, then do flat butt-welds, then move on to corner welds and tougher ones like overhead and vertical welds. If you first learn torch control with gas welding, you'll be able to pick-up any other welding method more easily.

size to have an exact-size piece after cleaning up the cut edges with a grinder.

The oxy-acetylene setup is still the least expensive welding system if you buy the torches, gauges and hoses, and *lease* the gas cylinders. The tanks are expensive to purchase outright, but can be leased from your local welding supply store with a deposit down and a small monthly fee. Many shops today will take a credit application in lieu of a deposit, and you will open an account there, assuming your credit is good.

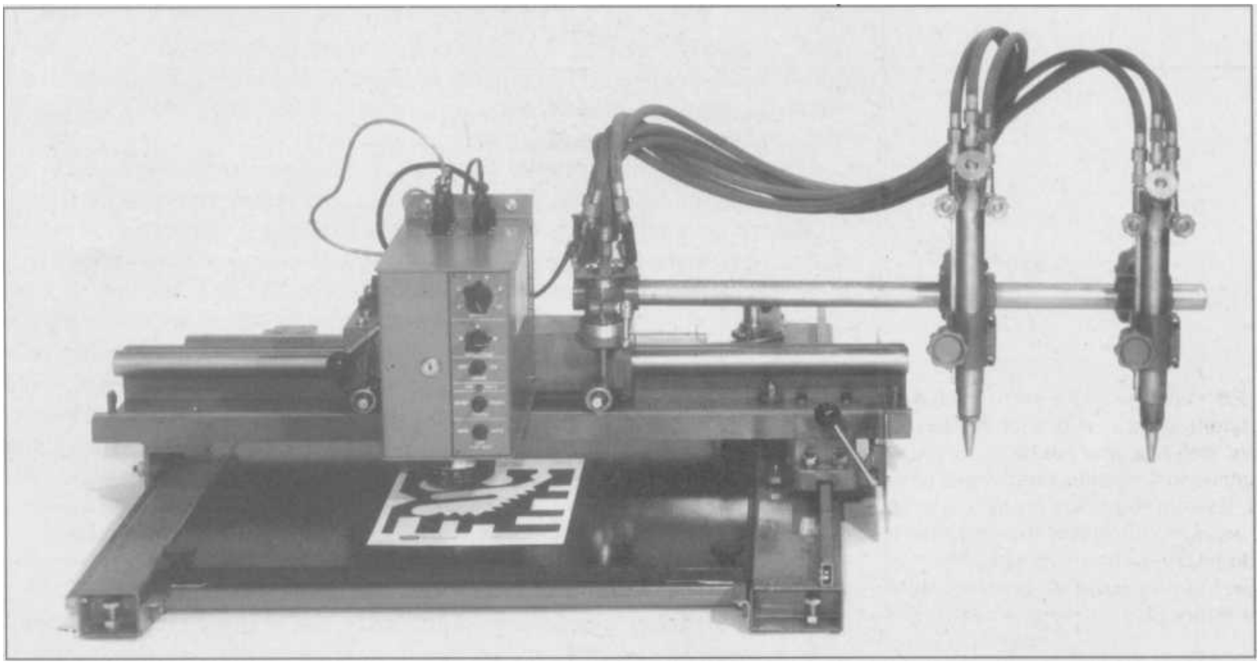

2.12 Oxy-acetylene torch flames can be put to some sophisticated uses. This is a mechanical gas pattern cutter (ESAB Silhouette 500), which can follow a pattern under the stylus and cut out two identical copies with the torch heads at right. Intricate parts can be cut out repeatedly, and with very clean edges because the torch movement is motorized and very smooth.

Leasing tanks is an inexpensive way to get started, but if you do only a small amount of welding infrequently it may not be a practical arrangement for you. After you have been paying rental fees for a year or two, you'll realize you could have purchased the tanks outright for the same amount of money. If you knew you only had to do some gas-welding for a few months, then the leasing deal would be better.

When welding with oxy-acetylene equipment, the basic procedure is to set the proper gas flow to the torch with the regulators, crack open the valves on the torch, light the flame with a friction-sparking lighter and then adjust the ratio of oxygen to acetylene to achieve the proper flame. Changing tip sizes makes a bigger or smaller flame, so you suit the tip size to the thickness of the parent metals you are welding. A smaller flame is used for thinner metals. The torch is brought down to the work area (the weldor is wearing his dark-lensed safety goggles) and the flame is used to heat the two edges to be joined, while your other hand feeds a piece of filler rod into the molten puddle as you move along the joint. Weld joints can be made with or without a filler rod, but a filler rod is used most often.

This is a very simplistic description of the process of gas-welding. It takes considerable practice and good hand/eye coordination to master. Once learned, the skills can be very useful, but it is a process rather difficult to learn from a book. Taking a class or having an experienced friend take you through the process will ease the learning curve considerably.

Besides the versatility of doing both welding and cutting, oxy-acetylene equipment also has many other shop uses in supplying a lot of localized heat. You may have projects where you need to bend a piece of metal. Thin sheet metal can be easily bent with pliers, vise-grips or put into a vise and bent over with a hammer, but thick metal may crack when bent cold. If you have a 1/4-in-thick steel bracket you need to bend at an angle, cold-bending with a hammer and vise may require so much hammer force as to distort the part out of shape, as well as mark the surface up or even damage the vise. If you closely examine the piece of thick plate after a cold bend, you may see the metal in the corner of the bend looking crystallized, which weakens that spot. Heating the metal to the right temperature with a torch before bending it, along the line where the bend should be, allows an easy bend with less disturbance of the metal's integrity along the bend. Another use of gas welding equipment is for brazing ferrous and non-ferrous metals such as copper and brass. In brazing, the parent metal is not made molten, it is heated enough to melt a brass filler rod, which attaches to both pieces of parent metal, making a firm joint.

Besides heating metal for bending, gas equipment is also used in many automotive shops for freeing frozen parts. Metals expand when heated, and when a rusted nut on a fastener is heated, the nut expands and breaks the bond, so the nut can be removed. When working under a car that has seen winter road salts, or when disassembling an old car for restoration, a gas torch can be very handy for getting off rusted nuts without busting your knuckles when your wrench rounds off an old nut. Many a mechanic on older cars would not be without his faithful "smoke wrench." Machine shops also use a torch to heat and expand parts that have a press fit. A gear that fits on a shaft may be heated with a special "rosebud" tip that spreads the flame around a wider area, and when the gear has expanded, it is picked up with tongs and slipped onto the unheated (and not expanded) shaft. When it cools off and contracts, the gear is securely fastened to the shaft, but can be removed at a later time just by reheating it again.

Despite the versatility of the gas-welding equipment, it may not be the ideal equipment for you, depending on your needs. If you have other uses for it besides just welding, then it is definitely a must have, but, if you just need to occasionally weld various thicknesses of steel together, some of today's electric welders may be more suitable for you. Gas welding is harder to learn, there are more safety problems in the shop when using a gas torch (especially when cutting), welding thick metals takes good skills and it is easy to distort the parent metal when welding thin sheet metal. Automotive body men today are using a torch less and less when doing repairs on thin metal. However, one of the advantages of gas-welding equipment over any electric-welder is portability. Your gas welding cart can be moved around anywhere without wires, even carried (when properly secured) in your pick-up for welding at a remote site.

If you have the budget and the need for gas-welding's versatility and portability, then your ideal shop setup would include both an oxy-acetylene rig *and* some kind of electric welder.

2.13 **In farm and construction equipment repair, most work is done with oxy-acetylene cutting and arc, or stick, welding. Arc welding is not affected by wind outdoors, and is able to join or repair very thick materials.**

Arc welding

Like gas welding, electric arc welding has been around for almost 100 years, and the fact that it is still around today illustrates its continued usefulness. The official acronym for arc welding is SMAW, which stands for Shielded Metal Arc Welding. The basic components of the setup include the machine (the power source), a ground lead you clamp to the work anywhere except where the weld is to be made, an electrode lead which runs from the machine to an electrode holder, which is a handle with a clamp that holds consumable electrodes. The electrodes are metal rods covered with a coating.

In use, the weldor strikes an arc against the parent metal with the electrode, which completes the circuit between the two leads and causes a bright light and concentrated heat. Arc welding uses considerable amperage of electricity to generate the intense arc, which melts the parent metal. The central metal core of the electrode melts as the work progresses, becoming the filler metal, while the fluxed coating produces a shielding gas around the welding area that protects the parent and filler metal from impurities in the air. Arc-welding produces *slag* as you proceed, a thick coating of impurities and deposits left from the rod's coating. This slag must be chipped off with a chipping hammer, which is usually included with the machine.

There are a wide variety of welding rods (electrodes) available to suit almost any purpose. The 12-14-inch-long rods are also called "sticks", and you may often hear arc-wetding referred to as *stick* welding. The rods vary in thickness, according to the thickness of the metal you are welding, and also in alloy and flux-coating content. There are many special-purpose rods, and, because of the variety, rods are usually marked with a number at the beginning of the flux coating, and different colors may also be added to the fluxes for quick identification.

Rods are usually sold in boxes or cans of fairly large quantity, which can be a problem for home/shop use where the welding is infrequent. The coatings on arc-welding rods are very susceptible to moisture in the air, and must be stored in very dry, secure containers to remain effective. You may have seen welding filler rods in gas-welding outfits stored in simple lengths of pipe welded to the welding cart, but this is not suitable for arc rods. If you do purchase an arc machine, invest in several airtight metal containers to store the rods, even using bags of desiccant (moisture-absorbing crystals usually found in small bags packed with cameras or sensitive electronic equipment) in the cans.

2.14 **Two endeavors you wouldn't think of together, art and welding, combine today as many sculptors and artists are now working in metals. Small works are usually torch or TIG welded, with larger works like this one utilizing arc or MIG welding equipment.**

There are two basic types of arc-welders, relating to the polarity of the electricity they produce, AC or DC. The DC machines are generally larger, industrial units found in production shops, where they are hard-wired in, or mounted in conjunction with an engine for truck-mounted use in mobile field welding. Most shop-type DC machines require shop-type electrical input, such as 440V or *three-phase* 220V, which you will not find in any standard home. They are designed to operate day in and day out without overheating, and are considered the best choice for welding really thick materials, so that is why we see DC arc machines in use building bridges, buildings, ships, etc. There are a few small DC machines for home use, but they have limited amperage and should be used on lighter materials. There are even combination AC/DC machines, but these are usually expensive shop machines.

By far, the most basic and practical home/shop arc-welder is an AC machine often called a "buzz-box" because of the sound it makes when you are welding. This is the least expensive single welding system you can buy, with a good name-brand buzz-box costing about the same as a set of oxy-acetylene torches. However, the arc machine comes with rods and everything you need, while the gas setup also requires filled cylinders which make a ready-to-weld gas setup about twice as expensive as a basic arc box.

The wiring you have in your house or shop will be a factor in choosing the type of equipment best suited for your purposes. To use an arc machine, you'll need 220V availability. If you already have an electric stove in your house (or the house had at least been wired for this) or an electric clothes dryer, you're probably in good shape because many electric stoves and dryers use 220V current. However, unless your dryer outlet is already out in your garage, you may have to have a qualified electrician run this 220V power out to where you'll be doing your welding. If you don't already have a 220V outlet, the cost of running new service to your garage may double the total expense of setting up to do arc-welding at home.

Also, welders do not just plug into the same outlet as your appliances. The welder has a different arrangement of prongs on its plug, and an adapter is required to plug into a 220V appliance outlet. The outlet you use should also have a good 20-30-amp circuit breaker as well. Note that there are some small buzz-boxes that plug into ordinary household 110V power, but they aren't recommended for anything but light and occasional work.

A stick-welder is relatively easy to use. Unlike gas welding, where you have to operate the torch with one hand while feeding the filler rod with the other hand, there is only one piece to control with arc-welding, the rod-holder. However, the rod starts out 12 to 14 inches long but is used up continually (gets shorter) as you weld, making it tough to maintain an exact distance of rod to workpiece, which is critical to a good arc weld. With the rod too close, you burn holes in the metal, and too far away you can lose the arc entirely and have to restart. So the trick in arc welding is control of the tip of the rod, and, because it is always getting shorter, you have to "fine-tune" your wrist movement in the hand working the electrode-holder. For this reason, most arc-welding is done with two hands, especially when learning. Use your other hand to steady and help control your wrist action on the "working" hand. Professional weldors use arc-welding upside down, laying on their back, hanging from scaffolding, or even underwater with special equipment, but for the novice, a comfortable body and hand position is very important to making good welds.

An AC arc-welder is well-suited to working on heavier steel materials, from 1/8-inch to 1/2-inch thick, but is difficult to control on thinner materials. The machine will have an amperage knob, with settings from 30-230 amps (depends on make and model), which you suit to the rod and the work material and thickness. Some of the more expensive shop machines have settings that go as low as 4 amps for light materials, but gas, MIG or TIG welding is- more popular today for thin metals such as most automotive work. For fabricating shop equipment, building a utility trailer, frame repairs or farm equipment maintenance, the buzz-box works fine.

To sum up the pros and cons of an AC arc-welder, the advantages include the low initial cost, easy operation (with practice), versatility (it can be used indoors or out) and it offers a high level of dependability (no moving parts) and quietness of operation. Disadvantages include: it's less practical for thinner metals, your shop may require rewiring to accommodate it, the home arc-welder usually can't be used for welding long seams all at one time, you are restricted to the limits of your power-cord length (as with any electric machine except the generator-driven type), the welds may have considerable spatter and not look as "clean" as other types if that is a consideration (such as in art projects, metal furniture design or street-rod fabricating), and there is the safety consideration of potential skin burns. Compared to gas welding, arc-welding poses more danger due to burns, not just from little spatters of hot metal, but from any skin that is exposed to the UV and infra-red rays. You can get a severe sunburn from exposure (pro weldors wear heavy leather protective clothing) and observing an arc-weld in progress without a helmet on, even for just a second or two, can cause headaches and eye irritation.

The buzz-box was once the most practical home/shop welder, but increasing availability and affordability of home MIG machines in the last ten years has put up a serious challenge to that title.

MIG (wire-feed) welders

This category of welding system has become one of the most popular for today's home/shop use. The initials stand for Metal Inert Gas, but is also listed in technical descriptions as GMAW, for Gas Metal Arc Welding. The basic elements of the setup include a power supply (machine), a torch with a large-diameter cable, a ground wire with clamp, and a bottle of compressed shielding gas. Inside the machine is a roll of relatively-thin wire and a motorized transport system for this wire.

In practice, you weld almost like arc-welding, but the electrode (wire) is constantly fed through the cable to the gun and consumed at the weld. When you pull the trigger of the MIG gun, you start the supply of amperage (when the arc starts on your work), the feeding of the wire electrode, and the flow of shielding gas, which is also routed inside the gun's cable and comes out of the tip all around the electrode, preserving the integrity of the weld like the flux coating does on arc rods.

The advantages of the MIG system includes a much cleaner weld than either gas or arc, good versatility in materials with the ability to do very well on thin metals, and there is no electrode or filler rod to keep replacing. Control is easy because you can set the amperage on the machine and also infinitely adjust the speed of the wire com-

30
MATCO
GOODYEAR

2.15 Anyone who is a fan of auto racing is familiar with various types of welding. In the immaculate shop of Larry Smith Marketing, a crewman is putting a final MIG weld on a complete NASCAR chassis for a car like the Matco Tools entry at left. Race car shops use a variety of welding techniques, from gas to MIG and TIG.

DAYTONA MIG
MIG 140

2.16 Small, portable wire-feed, or MIG, welders have become one of the most popular welding machines for home/shop use. Easy to set up and simple to operate, they produce very clean welds, and models from 100-amp to 140-amp can be operated on household 110V current.

REVOLUTION combi

2.17 Most MIG machines operate with a bottle of shielding gas, but some are equipped to run with flux-cored wire which makes its own shielding gas as it arcs, making the process more suitable for outdoor and drafty situations. This home/shop-sized unit shown is a combination machine that can operate with gas shielding/solid wire, or the flux-core wire.

Millermatic 250
Miller

2.18 At the professional shop level, MIG welders are all 220V machines, and offer power up to 250 amps, which is enough to weld most anything. This Millermatic 250 is a popular machine in fabricating shops and automotive shops. There is even a model of this machine with a small computer in which you can store the welding parameters of up to nine different jobs, in five different languages.

ing out of the gun. Various-diameter steel wires are available to weld material from the thinnest sheet metal to 1/4-inch and 3/8-inch, as well as stainless-steel and aluminum wires to weld those materials with. The weld from a MIG gun is much cleaner than arc welding in terms of spatter, and there is no slag coating to chip off.

Originally, wire-feed MIG machines were developed strictly for the professional shop doing long-duration production welding. A roll of wire in one of these machines can last for an entire eight-hour shift without the weldor ever stopping for a new electrode. Because the wire is machine fed into the weld puddle, the heat and wire-feed rate can be tuned so that production welds can be made much faster than with other systems. The MIG machine has been an expensive piece of equipment in the past (and heavy-duty, big-shop models are still fairly expensive), so it never used to be recommended to anyone for home/shop use. However, less-expensive MIG machines have become widely available in the past 5-10 years that have made these an excellent choice for the hobbyist, particularly the automotive hobbyist. These newer machines, mostly imported and built with a few less features and a lower duty-cycle than would be required in a production environment, have not only become very popular with the hobbyists but are also are finding their way into more small fabricating shops, muffler shops and race-car fabrication shops as a less expensive alternative to the big machines, yet with better speed and weld quality than arc or gas.

MIG machines range in size and features from "pocket" machines to large, full-featured machines for shop environments that require 100% duty-cycle. There are a number of smaller machines available today that operate on 110V household current, which makes them very portable, and no garage rewiring is

2.19 If you're lucky, the welding store near you will have a demonstration area, such as here at Altair Gasses & Equipment, Oxnard, California. Here a portable MIG unit is being tested where customers can watch through a large safety lens (behind operator, at right). You can get a feel for different equipment before you decide what machine best suits your needs.

Rod Nationals
street is neat
INK-Credible
Advertising
4442 F

2.20 If you are interested in cars as a hobby, whether it be kit cars, street rods, race cars, or antique restorations, you should become familiar with various welding techniques, and perhaps buy a system. When fabricating parts at home, it's a lot easier if you have your own welder, even if all you do is tack parts together and bring them to someone else for final welding.

necessary. These smaller wire-feed machines usually have 4-7 heat range settings from 30 to 110 or 140 amps, and can handle light sheet metal such as auto body material up to metals as thick as 1/4-inch or 3/8-inch (less maximum thickness is possible on aluminum). They either have wheels on the bottom or have optional, wheeled carts that make then very easy to move around your shop or garage. Many beginning weldors like to build their own welding cart as their first project with a new welding machine.

The shielding gas used with MIG machines can be CO_2, Argon, or a mixture of the two, depending on the materials you are welding. The basic gas used most often is straight CO_2, because it is the least expensive. Various sized gas bottles are available from small, very portable 20-cubic-foot bottles (about two feet high), to 120-cubic-foot bottles (four foot high, six-inches in diameter) that can supply enough gas for eight hours of continuous welding. The small bottles are fine for occasional home/shop use where long seams are not welded regularly.

2.21 Typical home/shop projects such as a utility trailer, race-car trailer, metals rack, shop cart, engine stand, hoist and much more can be built at home with either MIG (shown) or an AC arc welder.

One of the factors that has made MIG welders so popular with home/shop users is the relatively easy learning curve associated with using them. Given a little instruction and practice, most people can be up and welding a decent bead in an hour or two. We're not suggesting that that person would then be ready for a full-time job as a weldor, or that his/her first welds would pass rigid specifications for nuclear reactors, but the wire-feed machines are easier to learn than most other systems. One of the factors that makes it easier is of course not having to deal with feeding the filler rod in with your other hand. In automotive body work this can be particularly helpful when you have to hold something with your left hand, like the alignment of two pieces of sheet metal, while you tack them together with the one-hand MIG torch. Torch position is very important in all forms of welding, and the fact that the distance of the MIG tip from the work is constant (in many cases the nozzle is actually touching the work to steady it) makes controlling the weld much easier.

After selecting the proper heat setting based on the thickness of the material

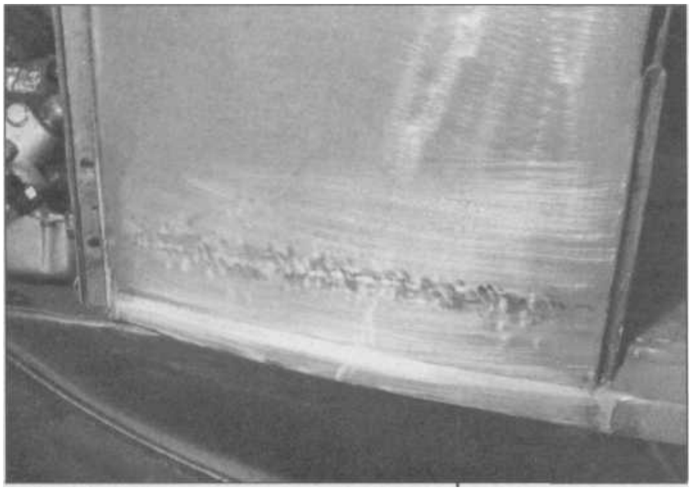

2.22 Street rodding and antique car restoration frequently require welding on old, even rusty sheet metal body panels. Here the lower portion of a '32 Ford cowl has been patched with new steel to replace rust-out area. The panel was MIG-welded in with a 110V welder, ground down with a body-sander, is ready for finishing and another 60 years on the road.

you are welding, the only other variable to control is the speed the wire is fed into the weld puddle. If the speed is wrong, there will be a lot of sputtering and uneven welding and penetration. Most MIG weldors tune the wire-speed in by *sound,* They set up on a scrap piece of metal with a wire speed that may be too slow or too fast, then pull a bead while adjusting the speed. When the right speed is reached, the MIG welder gives a distinctive crackling sound to the procedure. The steady crackling sound means you have the right wire speed.

There are three basic types of small MIG machines. Those that use plain wire with a bottle of shielding gas, those that use flux-coated wire and no gas bottle, and combination machines that can run either fluxed or plain wire. The flux on the wire functions the same as the coatings on arc rods, creating a shielding gas as the wire is consumed. The bottled-gas-shielded MIG welding is the preferred method for indoor welding where clean-looking weld beads are desired. However, the shielding gas around the weld process is subject to being disturbed by air currents, so for outdoor welding (such as farm equipment repair, fence construction, etc.) the fluxed-wire machine is better. The weld isn't quite as clean as with the gas, and fluxed wire doesn't do as good a job on thin sheet metal (under 18-gauge) but these machines are more practical for the outdoor work, which is usually on heavier materials anyway. Also, the fluxed-wire machine never need to have a bottle refilled, so as long as you have plenty of wire inside the machine, you can't have your work delayed by running out of shielding gas.

Considering all of these factors, the modern MIG machine may be the most practical type of machine for many do-it-yourselfers, particularly the automotive hobbyist, who has to make very clean welds, both with thin sheet metal as well as heavier chassis work. The ease of operation, maneuverability around the shop (with the 110V machines), one-hand torch operation and the gentle learning curve have combined to make many first-time MIG weldors tackle their projects with confidence.

TIG (heli-arc) welding

This is unquestionably the "Rolls-Royce" of the welding processes in many ways. The lettering stands for Tungsten Inert Gas, but many weldors refer to it as

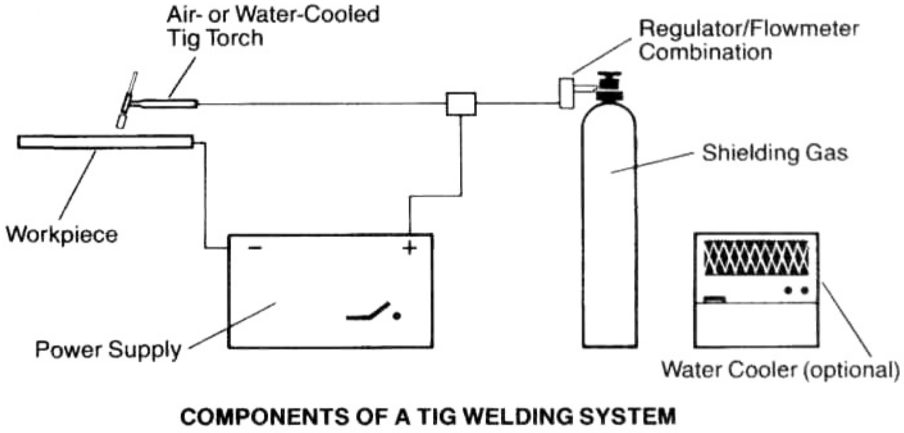
Air- or Water-Cooled Tig Torch
Regulator/Flowmeter Combination
Shielding Gas
Workpiece
−
+
Power Supply
Water Cooler (optional)
COMPONENTS OF A TIG WELDING SYSTEM

2.23 Basic components of a TIG-welding system. Very similar to operating a gas torch and using your other hand to feed filler rod, TIG produces the cleanest welds with the least amount of material warpage, and is the best system for thin, high-strength material like chrome-moly as well as aluminum.

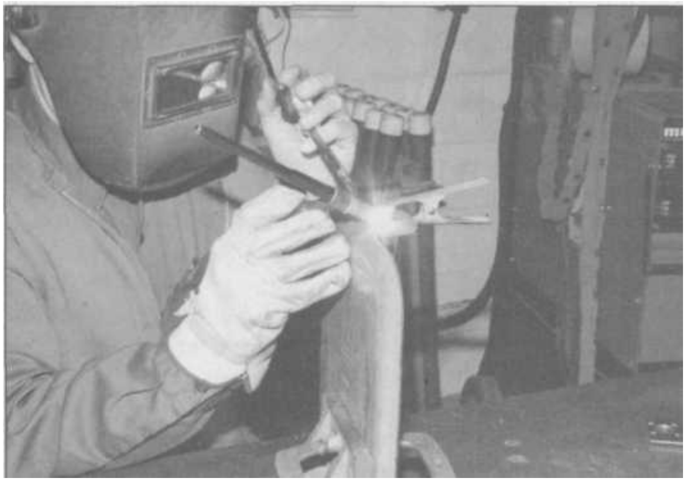

2.24 While TIG or heli-arc welding is considered by many to be the most *controllable* welding technique, it takes more practice than any other. You have to learn to manipulate the torch with one hand, the rod with the other, and the amperage control with your foot on a pedal. There are no sparks, spatter or slag. Here an aluminum frypan's handle mount is being repaired. The heli-arc welding didn't even bother the non-stick coating on the other side!

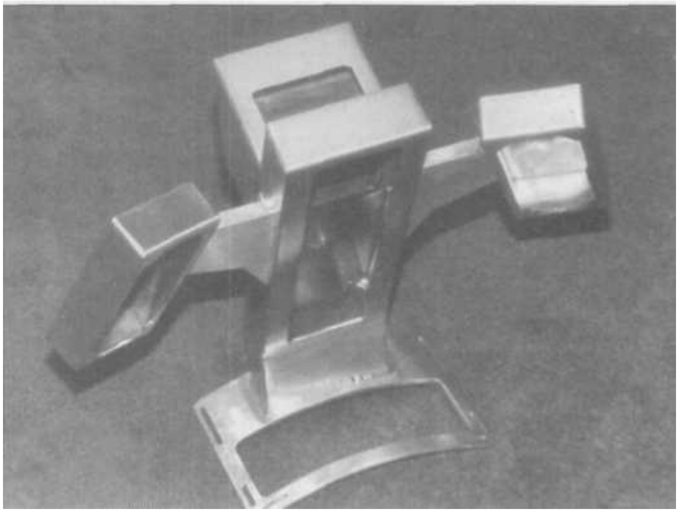

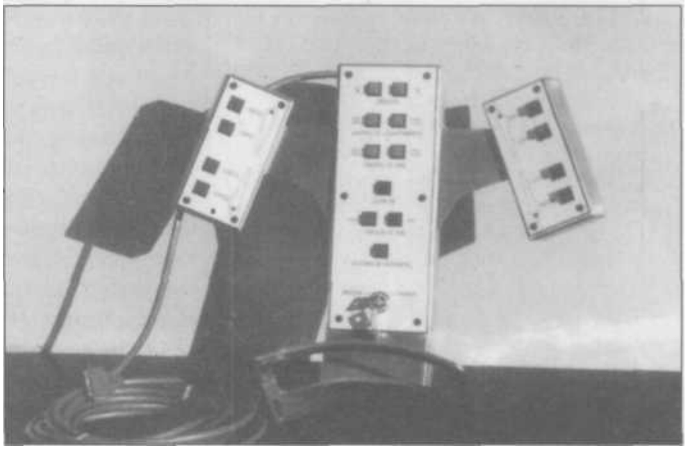

2.25 Where TIG welding really shines is in two tough situations in one job, like thin materials *and* aluminum, such as this part that looks like something from a space ship. This wound up being a belt-pack of controls for a big piece of machinery - the operator strapped this to his waist. Aluminum is tougher to weld than steel because it doesn't change color as you heat it. When it melts it turns shiny, but tarry a little longer with the torch and the material has holes in it. Don't tackle a project like this until you have mastered the process.

heli-arc welding, though this is a trade name established by the Linde Corporation many years ago. The name refers to the fact that back then, helium was used as the shielding gas for the process, though today argon is the most commonly used gas for TIG welding. Either shielding gas has the capacity to exclude foreign materials during the welding process, as well as being unable to combine with other elements to form chemical compounds detrimental to the weld integrity. The process can also been seen abbreviated as GTAW, for Gas Tungsten Arc Welding.

The basic components of the TIG process include a welding machine, a torch with a tungsten electrode, a ground cable, a foot-operated amperage control pedal and a bottle of compressed shielding gas such as argon. It differs basically from stick welding and MIG welding in that the electrode is not consumed, and filler rod is separately hand-fed into the weld puddle when needed. In operation, the machine is turned on, the weldor brings the torch close to the work and depresses the foot pedal to start the arc. Current flows through the tungsten to the workpiece, creating an intense but concentrated heat that melts the parent metal, with the operator controlling the amperage with the foot-pedal while welding, and feeding the appropriate filler rod to the puddle as needed. The tungsten electrode has a very high melting point and is not consumed.

What makes the heli-arc different in practice from other systems is the higher operator skill required, since the weldor has to coordinate with both hands and the foot pedal, all of which affects the quality of the finished weld. Unless you are already an accomplished weldor with an oxy-acetylene gas-welding torch, you'll find the TIG process takes a *lot* of practice.

TIG welding produces the cleanest, flux-free and spatter-free beads with virtually no contamination. For this reason, it is the method of choice for such critical applications as race cars, aircraft, missile construction and nuclear reactor welding. The TIG welds are strong, ductile and corrosion-resistant with virtually no cleanup required, making it popular for welding of surgical, medical and food-service equipment, especially on stainless-steel. There is little smoke produced during TIG welding and no flying sparks, and without the smoke and sparks the weldor can more critically observe the welding action as he goes.

While there is a lot of skill required in TIG welding, its ability to handle all kinds of metals of varying thicknesses, the ability to reach into corners and hard-to-access areas to weld and its outstanding ability to mate thin metals without putting excessive heat and distortion into a wide area on either side of the weld has made the heli-arc method particularly useful in race-car and other critical construction applications. The types of metals that can be welded with TIG machines include: all steel alloys, aluminum alloys, aluminum castings, stainless-steel, copper and nickel alloys, magnesium and exotic metals such as titanium and zirconium. It is one of the few systems that can tackle welding dissimilar metals and joining metals of different thicknesses. It's been said that a skilled TIG weldor could join a razor blade to a railroad track, though there isn't much call for that.

The torch for TIG welding is smaller and lighter than most other welding torches, allowing for greater weldor comfort when welding for long periods. For production welding, most TIG machines are equipped with a water-circulating system to keep the torch cool. In use, a water pump circulates filtered water from a storage tank out through the main cable to the torch head and then returns the heated water back to the tank. Many cooling units employ a small radiator and electric fan to cool the water. Where lots of production welding is done, a large water storage tank is used, or the water is fed in cold from the shop's plumbing and the heated water is plumbed back into the drain system, so heated water is never circulated back through the torch. The TIG machine requires shop-type current input (230V/460V/575V single-phase or 230V/460V three-phase) and connections to and from the water-cooling equipment, making it one of the least portable of welding systems.

Because of the electrical requirements and other hookups, the steep operator learning curve, and the considerable expense of the most shop-type TIG machines, this has never been a system that has penetrated much into the home/shop and hobbyist market. Rare and proud is the automotive hobbyist or racer who has a heli-arc machine at home, and who knows how to use it to its capacity. The process has assumed an almost "mystical" aura to automotive fans because of its use in race-car construction, and even otherwise skilled home weldors will bring certain components like critical steering gear to a professional to have the parts TIG-welded. Street rod builders love it for the quality of weld, which is often cosmetically appealing enough to require no grinding or filling to make the welded part ready for painting or even chrome-plating.

Shops with TIG machines rarely take advantage of this, but most TIG welding machines are capable of also performing stick welding if the job should require it. TIG welding is ideal for thinner materials, but if a job required joining some one-inch-thick plates, the stick-welding mode would be much faster. The arc capability would also come in handy if the job required some welding outdoors or near a drafty doorway, where the draft could blow the argon shielding gas away and cause an erratic weld. Switching to the arc mode simply requires a switch on the machine be thrown to the correct current position for arc, and the arc "torch" electrode cable plugged in. When in the arc mode, the foot control is deactivated and the weldor sets the amperage on the machine with a knob, like a straight arc machine.

While most production shop TIG machines are large, expensive, heavy-duty machines with water-cooling systems for the torch, there have been strides in the past ten years at making a more affordable TIG machine. Today there are several to choose from which are air-cooled instead of water-cooled, reducing cost and complexity, that feature straight 220V, single-phase input connections for simplified electrical hookup, and have lighter-duty, less-expensive internal electrical components. These new machine are one-half to one-third the cost of their bigger conventional TIG brothers.

Because of these factors, there are more small shops that today can afford the benefits of TIG welding, such as high-quality welds on thin materials, and the ability to weld beautifully on non-ferrous materials like aluminum, even cast aluminum. The are often seen in large engine rebuilding facilities and automotive machine shops, where they serve their time repairing cracks in aluminum blocks and heads, expensive parts that would otherwise have to be replaced if such repairs could not be made. However, these "economy" TIG machines can't do everything the big ones can, particularly when it comes to torch cooling. If long welds are to be made, the air-cooled torch will get too uncomfortable to hold. Also, the smaller, less-expensive electrical components inside will not have as long a duty-cycle as the big machines (more on duty-cycles later), but in a small shop fabricating race-car or street rod parts and doing precision chassis work, long, continuous welds are not often encountered.

Some welding machines today offer "convertible" adaptability. Kits can be added to some MIG machines to add heli-arc capability, and there are kits with a different torch and a separate wire-feeder to use the output of the bigger TIG machines as a MIG machine, but this convertibility usually puts the machine out of the home/shop price range.

Even with the introduction of this new generation of more-affordable TIG machines, this is still a system that isn't practical for the average home/shop weldor to get into. As with oxy-acetylene gas welding, the practice time it takes to be good with a heli-arc torch requires consistent, daily operation, and most occasional weldors just can't get enough time on the machine to really do well. Most basic home/shop fabrication and repair projects do not require the "reactor-quality" welds and superior cosmetic appearance of TIG welding, and even the economy TIG machines cost more than twice what other starting welder setups cost, making the gas welder and small MIG machines still the most versatile equipment for home use. A drawback of TIG welding is that it is a *slow* process, so for a project like building a utility trailer, a typical home MIG machine would be much faster.

Duty-cycles

In any discussion of electric welding processes, the term duty-cycle comes up and is a source of some confusion to the beginning weldor. Basically, the bigger, heavier and more expensive pro-shop type welding machines have transformers and other internal electrical components designed to operate for long periods of time without overheating. Smaller, home-type machines can't operate as long, because the internal components had to be made less expensive in order to reduce the overall costs to a "consumer" level.

The official quantifier of such work capability is called the duty-cycle. It is expressed as a percentage, such as a "40% duty-cycle." What this refers to is a ten-minute period in a laboratory welding test. At a 40% duty-cycle, a particular welding machine is capable of welding for *four minutes out of ten.* In other words, if you welded non-stop for four minutes straight, you would have to let the machine "idle" for six minutes before resuming welding, allowing the internal components to cool off. The maximum current that a welding machine can draw from its components is limited by the temperature rise of those parts, and the temperature goes up with the square of the current draw.

Bigger, more professional and more expensive welding machines will have higher duty-cycles because of the nature of the work they have to do. There are some cases where a robotic or machine-driven automatic welder must be capable of 100% duty-cycle because it is operated continuously, but this is unusual.

The advertised duty-cycle of a welding machine is *usually* given at the maximum current draw of the machine, the highest amp setting. This setting of course is required only when welding the thickest material the machine is capable of. As the amperage setting of a welding machine is reduced, the effective duty-cycle is *increased,* so that there is somewhere on the welding machine's "curve" a point where the duty-cycle is close to 100% at some amperage setting.

For the home/shop user who is rarely welding anything thicker than 1/4-inch or 3/8-inch mild steel, the duty-cycle is not a problem. However, it's important to know what the duty-cycle for your machine is at various amperages, and especially to understand the term when shopping for a machine. For instance, one popular introductory-level 110V MIG machine has an amperage range of 30-110 amps. Right away, you know that such a machine is not really designed for long welds on heavy materials, because the maximum amperage is only 110. Further, the advertised duty-cycle for this machine is 95% at the 30-amp setting, but 25% at the highest amp setting, showing the difference in how the current-draw heats up the components. Depending on your average welding needs, such a machine may be just fine, handling short welds occasionally on 1/4-inch steel, and longer periods on thinner material. Most automotive sheet metal, for instance, seldom requires an amperage setting above 30, and even the small MIG machines have a long duty-cycle at this setting.

Generally, welding machine price goes up with maximum amperage and duty-cycle. You can roughly compare the performance of the machines being considered by comparing their duty-cycle at the amperages you will most often be using. A feature to look for on some of the better machines is an "overheat" protection circuit and warning light, which may save you the cost of a burned-out component. Some machines automatically shut down when the overheat circuit kicks, and in others an internal electric fan comes on to speed the cooling process, but like most "bells and whistles" these features usually cost a little more.

Plasma arc welding and cutting

This is another welding/cutting process that has developed in the recent past that was once considered exotic, yet is now filtering down to lower price levels in the welding marketplace. The nomenclature refers to the plasma state of a gaseous material, in which the material is heated so much that the gas conducts electricity. What makes PAW (Plasma Arc Welding) so different is that the flow of superheated *gas* actually makes the welding arc, with the metal electrode

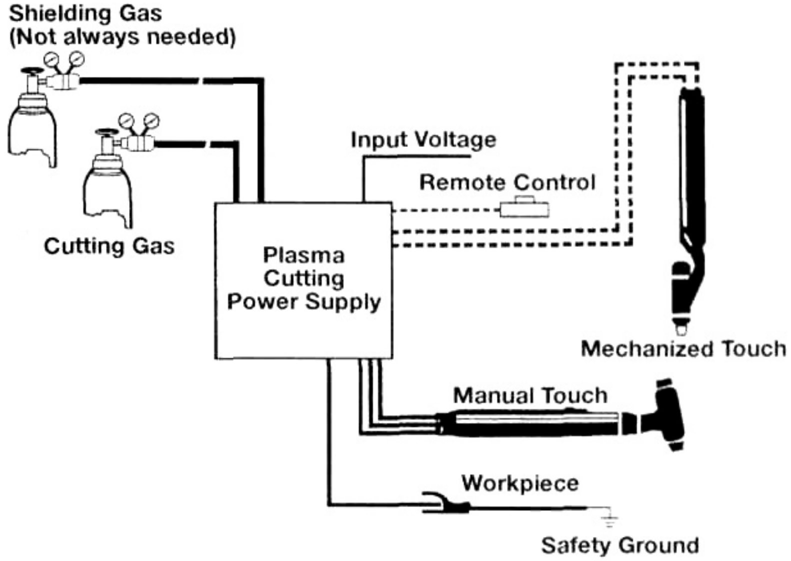

2.26 Plasma cutting is rapidly replacing oxy-acetylene cutting in many home and shop applications. Used either as a mechanized cutter or by hand, the system is simple to operate, and makes a faster, cleaner cut with much less sparking and less heat induced into the parent metal.

safely hidden way up inside the torch body away from the welding action.

The plasma welder makes a hotter and easier-to-control arc than even the much-touted TIG welder. The arc becomes a narrow column of heated gas (usually argon or nitrogen) that can be directed as precisely as a surgical laser. A second gas is plumbed through the torch to maintain a gas shield over the weld area, as is done in most electric welding. The extremely-high temperature of the plasma arc means that higher welding speeds can be used in production applications, and where there is a close fit of the two edges being welded, as is usually the case in a production situation with newly cut materials, often the plasma weld is accomplished with no addition of filler material when welding thinner stocks. This has made it a preferred system for some industrial applications.

In operation, the plasma equipment looks much like TIG equipment, except that there are two bottles of gas, and the torch is designed differently inside. The very dense, very hot "flame" does not require as precise an arc-starting-distance of the torch to the material, making plasma equipment somewhat easier to operate than TIG, especially on projects where no filler metal is needed. In TIG-welding, the weldor must strike an arc with his tungsten electrode With the plasma method, however, the electrode is too far up inside the torch head to get near the parent metal, so most of the plasma machines have high-frequency arc-starting circuitry that makes them easy to get up and running, using the hot gas column to start the arc, after which the proper torch-to-work distance is easily established by the weldor.

While there are specialized applications where plasma-arc welding is the preferred method, its widest area of usage today is in *cutting.* The column of plasma-stage-heated gas is hot enough to cut through virtually any material at almost any thickness. On thin materials, the plasma cutter goes through as fast as the operator can move the torch (speeds up to 100 inches per minute), and even on 1/2-inch-thick steel the moderate-size cutters can cut at 15 inches per minute, which is about one inch every four seconds!

2.27 In fabricating shops, a large 220V (or higher current, three-phase) plasma cutter is used, capable of cutting thick materials and traveling at fast cutting speeds. Most plasma cutters are used without shielding gas, using air from the shop air compressor to cool the torch and form the ionized gas column under the tip that creates the arc.

While the speed of the plasma cutter is exceptional, it is the quality of the cut that has made it so popular today in many applications like automotive, duct repair and fabrication, and medical or food-service equipment work. The plasma cutter will cut virtually any ferrous or non-ferrous material with equal ease, even on aluminum and stainless where ordinary torch cutting is difficult without considerable pre-cut and post-cut cleaning.

Since there need be no direct contact between the torch and the work material, plasma cutting doesn't involve a lot of spatter getting into the torch head, and most materials can be cut with no pre-cut cleaning. This is especially important to automotive users, who can make clean, fast cuts in sheet metal, whether it's painted, rusty or dirty. The concentration of the narrow cut and the speed the plasma torch can travel means that the cut edges are many times cleaner than even a machine-cut edge from an oxy-acetylene torch. There may be little or no cleanup of the cut edges required, depending on the job requirements, and the cut edges are not only smooth but uncontaminated, so that plasma-cut edges can generally be welded together much easier that edges cut with a gas torch where oxidation must be ground off before welding.

Plasma cutters have become particularly popular in machine-cutting operations where the torch head is mounted on a machine that has a pantograph arrangement. The weldor moves an arm around a wood, paper or metal template and the cutting torch cuts that exact shape out of metal. Sometimes, several cut-

DAYTONA MIG
POCKET Turbo 130

2.28 Where the really complete home/shop setup was once an AC arc welder and a set of gas torches, today more shops utilize a MIG welder and plasma cutter. This welding cart features a gas-shielded 130-amp MIG and a plasma cutter capable of cutting up to 3/16-inch material, and both machines operate on household 110V current.

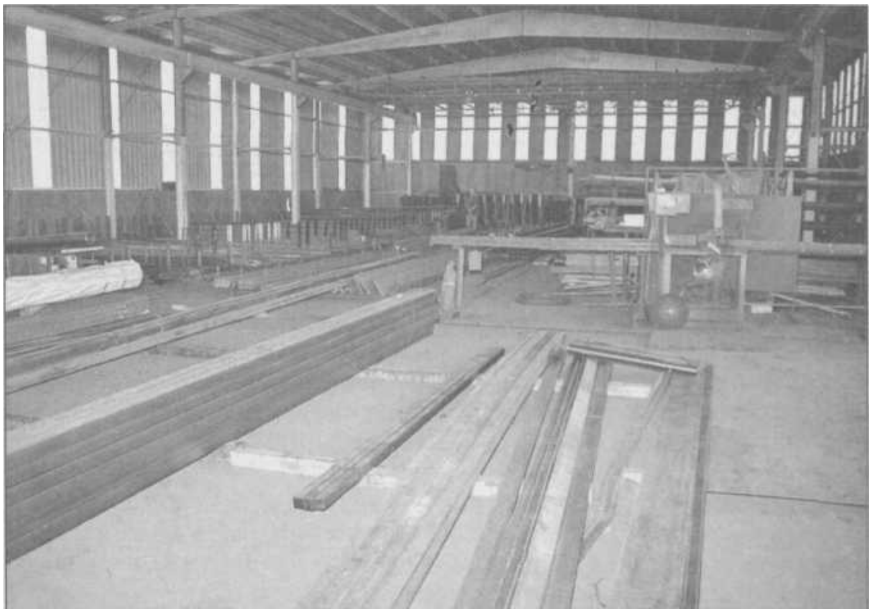

2.29 Eventually, you will find yourself at a steel supply store, shopping for plate, angle-iron or tubing for a project. Steel is usually sold in 20-foot lengths, but it's often cheaper to buy steel here than at a home/lumber type store.

ting heads are operating on a large sheet of metal plate, cutting three or four copies of the template out at the same time. The clean plasma cuts mean less labor is required to make a finished part.

Metal cutting has been done for decades with oxy-acetylene equipment, but, besides the clean-cut edges, a plasma cutter has one other major advantage over the oxy-acetylene torch. Since the main component of the air around us is nitrogen, which make a good inert gas to superheat into a plasma flame, common air can be used as the cutting gas. This is usually supplied simply by hooking the welding machine up to a standard shop air compressor making 75 psi. Not only is the compressed air virtually free, with no bottles to fill, but plasma cutting involves no flammable gasses in dangerous high-pressure cylinders. There is a major savings in bottled-gas costs, convenience, and shop safety.

The nature of the plasma-cutting process is such that very little spark spray comes from the cutting action, and there is virtually no discoloration or distortion of the cut edges. It is almost like "cold" cutting. These attributes make the plasma system the method of choice now in automobile wrecking yards for cutting up cars for dismantling. It's much faster than a gas torch, and the chance of fires starting from flying sparks and droplets of molten metal is greatly reduced. Plasma cutters have been operated in such situations with torch hoses as long as 234 feet, with the welding machine kept indoors in a shed and the operator wandering around the yard with the torch, not having to transport a heavy cart with bottles of compressed gasses. Body shops love it for its clean edges and the ability to cut an exterior body panel with less chance of igniting undercoating, paint or burning a hole in an interior panel with a stray blob of molten metal.

Although the technology sounds complicated, plasma equipment is perhaps easier to maintain than gas, MIG or TIG equipment, and the learning curve is not as steep. As with the modern MIG and TIG machines, recent years have seen the price of plasma-cutting equipment come down considerably, and this has put the technology in the hands of many automotive shops. Most home/shop users will not have enough cutting work to do to justify owning a plasma cutter, but when a lot of cutting needs to be done, and done more safely than with traditional gas, this is definitely the way to go. If you are around a body shop or fabri-

2.30 If you can't carry a 20-foot length of steel home with you, steel houses will generally make one "courtesy" cut for a small fee. Any cuts after that are expensive, so it's best if you have some method at home for cutting the metal for your projects.

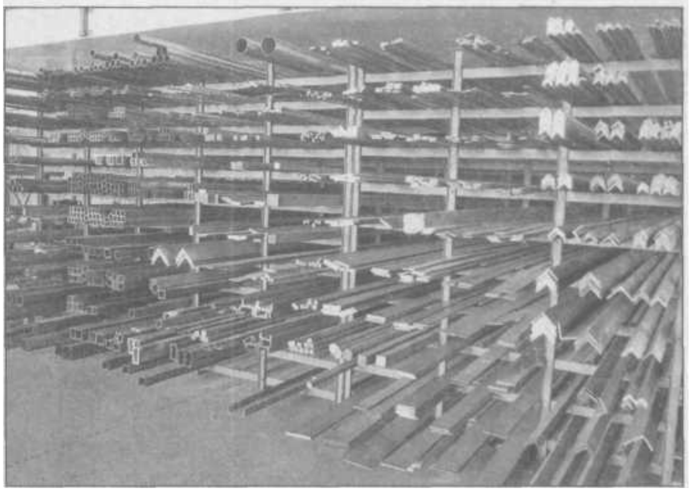

2.31 This steel supply store has a neatly-organized rack for materials, in which each different thickness of material is identified by a different color of paint on the ends, which makes shopping a lot easier for hobbyists without a micrometer.

cation shop that has a plasma cutter, see if they will demonstrate it to you. You'll find it amazing when you see how fast it cuts, how little spark spray there is, and how clean and relatively *cool* the cut edges are.

Practice and training

When it comes to hand-eye coordination, everyone has a little different skill level to start with. How easily each person can pick-up a welding machine, read the instructions and begin joining metals together will differ with that person's mechanical background. Some people seem to flow into the work intuitively; others require considerable practice and often a little coaching from someone who can weld already.

Each of the welding systems we have discussed in this chapter requires a different amount of practice and skill to master, but rest assured that even if you have never done anything like this before *you can learn to weld.* Millions of men and women have learned to do this in the past. During the domestic manpower shortages of W.W.II, there were many housewives who went to work in the defense plants, and even though they had never done anything mechanical they were there riveting, metal-shaping and welding on aircraft, tanks, ships and many other critical projects.

The first step in deciding what kind of equipment you need is to accurately assess the kinds of projects you would do most often if you had welding equipment in your home/shop. The thicknesses of metals you want to work with, the types and alloys of metals, the kind of electrical power you have available and the length of the welds you may need to perform are all factors in choosing the right machine. We've given you enough of a brief outline of the advantages and disadvantages of each type of equipment to aim you in the right direction. You may also want to consult with your local welding shop and local welding supply store before making your decision. If you know someone who already has a welder at home, ask his advice and even see if he will let you try out his equipment on some scrap pieces.

Most people who are serious about picking up welding skills and doing it right can benefit from actual training. Your local community college may have adult evening courses in basic welding, or there may be a trade/technical school nearby. Courses offered at publicly-funded community colleges and trade schools are usually inexpensive, and the hours are often built around a working person's schedule. There is a real difference between a weld that *looks* good, and a weld that *performs* well. If you practice welding on your own, you may achieve welds that look OK to you, but which may fail in actual use. One of the valuable parts of the training you will receive in a legitimate course is the critique from the instructor. He can show you how to test various trial welds for strength and let you know when you're getting it right. We think such training is invaluable if you plan to do anything more involved than repairing a bicycle or garden tool. Good training shortens the learning curve dramatically, and we recommend it to any readers who have it available in their area.

3

Oxy-acetylene gas welding/cutting

Perhaps because of its history going back to the turn of the century, or perhaps due to its appearances in films and other popular imagery, gas welding/cutting has the most "romantic" image of the various welding systems we'll deal with in this book. When most people conjure up a vision of a weldor, it is of a perspiring, leather-clad man wearing dark goggles and wielding a flaming torch with sparks flying everywhere. In truth, this is more likely a picture of someone *cutting* with a torch rather than welding, but it points out the ubiquitous nature of the gas equipment in the overall welding picture.

Although losing some ground to modern electric welders in the most recent decade as the welding system of choice for the average home/shop user, oxy-acetylene equipment is still unarguably the most versatile setup for home, farm or shop. Besides its function as a fusion-joiner of metals, gas equipment also of course is extremely valuable as metal *cutting* equipment, and can also be used for non-fusion metalwork such as brazing, soldering, and the dying art of automotive body lead work. It is also useful equipment in any automotive shop for freeing rusted fasteners, heating machined parts that require a hot shrink-to-fit connection (like a gear on a shaft) and heating metal parts prior to bending them. It requires no water or electrical connections, making it one of the most portable of systems, you can load it into a pick-up truck and take it to any remote location. If the gas hoses are long enough, the weldor can climb up a pole or down a shaft to perform welding or cutting, a flexibility few other welding systems can match.

The basic gas process

When gasses are used in other types of welding, they are usually of the inert kind, like argon, CO_2 or helium, which are involved in the welding process only to keep the molten weld puddle clear of impurities from the air during formation. In oxy-acetylene welding/cutting, however, the

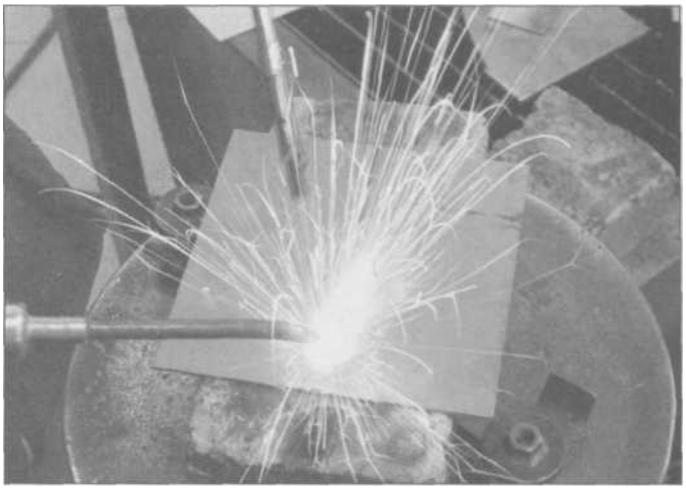

3.1 Versatile gas-welding equipment has been a mainstay of metal fabricating for most of the 20th century, and little has changed with procedures or equipment. The basics, fusing metals at their melting point, are still valid for many uses in farm, industry and home/shop hobby applications.

3.2 You'll find helpful people at your local welding supply store that can set you up with torch equipment, oxygen and acetylene bottles, and other supplies you'll need.

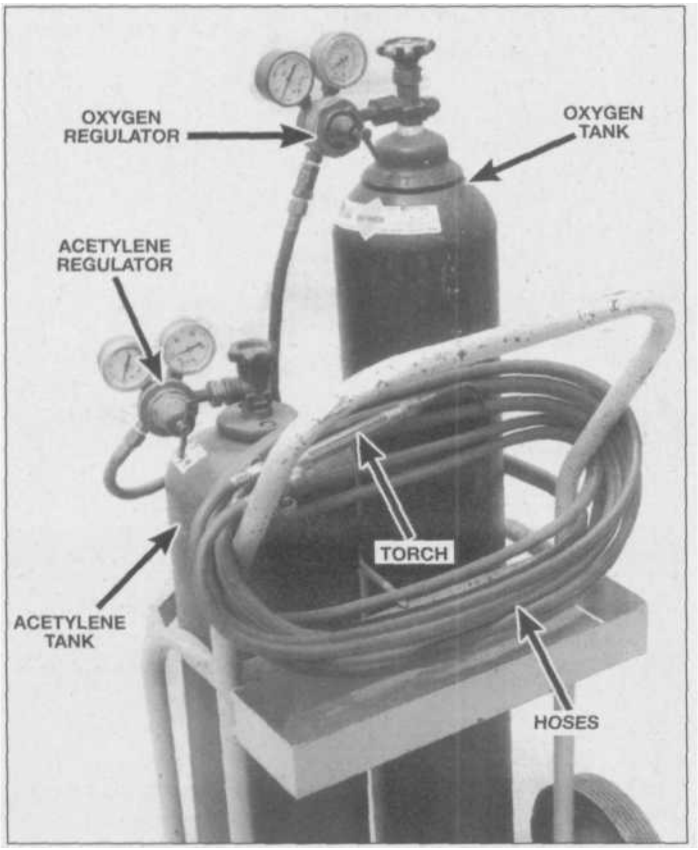

3.3 Look in the corner of almost any automotive or fabricating shop and you'll a setup just like this, with bottles, regulators, hoses and torch neatly corralled in a portable welding cart. Carts can be readily purchased, but most beginners like to make their own cart as their first project.

PUROX
PUROX
PUROX

3.4 This is typical of a good, general-purpose welding, cutting and heating set, ideal for automotive, muffler shops, farm and trade schools. Starter sets have smaller torches and 10-12 foot of hose, bigger sets have larger torches and 25-foot hoses.

gases are what make the flame itself. Acetylene gas is quite flammable, and combined with oxygen, which by itself does not burn but speeds up the oxidation or burning of any other fuel, makes one of the hottest possible gas flames (5600-6300° F),suitable for the rapid welding, cutting or heating of most ferrous and non-ferrous materials. Although all compressed gasses pose some shop hazards because of the pressure inside the bottles, oxygen and acetylene are considerably more dangerous to work around, and require much more caution and close attention to safety rules.

Oxygen, while not exactly flammable by itself, is the gas necessary both foi us to breathe and for any type of combustion to take place. Combustion is really nothing more than very rapid oxidation, and if pure oxygen is directed at something flammable, a fire can start very easily. Some inexperienced weldors hav< been know to dust off their work clothes with their unlit gas torch, but the extn oxygen that gets into their clothes can make it so flammable that any tiny sparl could start the clothes on fire. Likewise, oxygen gas must be kept away fron things like oily rags or any petroleum products. Oxygen as used in welding equipment is generally stored in high-pressure, 1/4-inch-thick-walled steel cylinders s 2200 psi. While oxygen may be important for our lungs to work, it can still be dangerous if too much is introduced to the bloodstream. For this reason, oxygen ga should be kept away from open cuts such as you might have on your hands.

Acetylene gas, on the other hand is flammable to the point of being explosivt and is also mildly poisonous, causing nausea and headaches if you breathe muc of it. Older readers may remember using a "carbide cannon" on the 4th of Jul when they were kids. A small quantity of pellets were put into a small met;

cannon which held some water. A gas was produced inside that was lit by a flint-sparker mechanism on top, and the result was a tremendous bang. The pellets were calcium-carbide, and the gas produced was acetylene.

Pressure in the acetylene gas cylinder is much less than with oxygen, at 250-325 psi, but the construction of the cylinder is different. Acetylene cylinders are shorter and fatter than oxygen bottles, and are constructed in two halves. Because acetylene is unstable at high pressures, the only way to get sufficient quantities into the standard bottle is to dissolve the acetylene in another medium. In welding tanks, the two halves are filled with an asbestos/cement mixture and then welded together. After baking, the material forms a honeycomb inside the tank. Liquid acetone is put into the tank because it will absorb 25 times its own volume in acetylene gas, thus stabilizing the acetylene.

Because of the differences in chemical action and storage of oxygen and acetylene, there must be no mix-ups between the two. For this reason, the tanks are made in different proportions and different colors; the acetylene bottle has only left-hand threads, and the hoses for each bottle are different, i.e. red hose for acetylene, green hose for oxygen **(see illustration).** Both bottles have a threaded top for filling and connection of the regulators, and due to their high internal pressure, especially the oxygen cylinder, gas bottles should always be stored securely. If a bottle were to fall from a truck onto concrete, for instance, the regulator could be knocked off, suddenly releasing enough gas to propel the heavy bottle around like a kid's balloon, except with potentially deadly force. In addition, the acetylene bottle should always be stored in an upright position. If stored laying down, some of the liquid acetone inside could be drawn out of the bottle into the welding supply, ruining the weld.

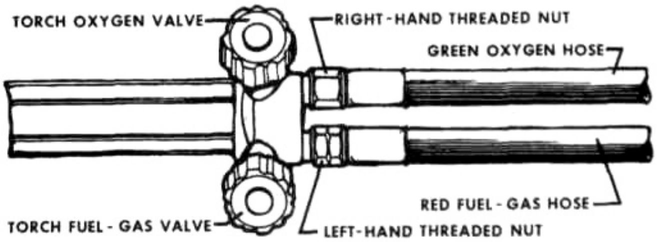

3.5 There must be no mix-up of gasses, so oxygen hoses are colored green, acetylene hoses red, and the acetylene hoses have left-hand threads, identified by a groove cut around the fittings.

The equipment

Besides the two gas bottles, there are gauges/regulators, hoses and the torch itself **(see illustration).** Each gas tank has it's own set of two gauges mounted on a regulator body designed to reduce the high cylinder pressure down to a useable pressure for the torch. On each tank, there is

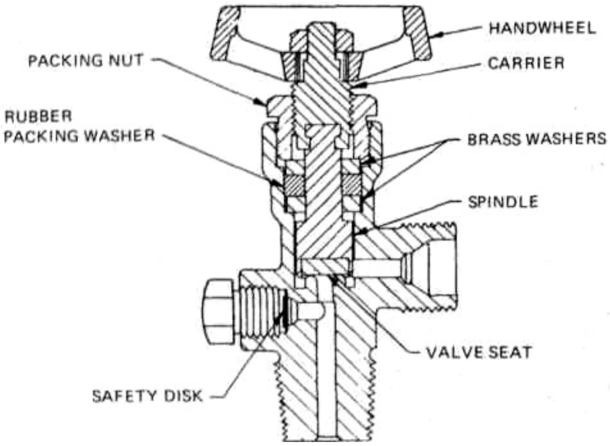

3.6 The gas bottles are fitted with valves, to which regulators are attached. This is a cutaway of an oxygen cylinder valve, showing the way the valve seals internally. Oxygen cylinder valves seal best when either closed or fully open.

PUROX
PUROX
PUROX

3.7 Heavier-duty torch sets like this Trade Master from L-TEC have torches capable of welding up to 3/8 steel, cutting up to 1 1/2-inch steel, and heating capability of 103,000 BTUs per hour.

PREST-O-LITE
Kangaroo Kart

3.8 Highly portable "Kangaroo" outfit has everything you need in a small package, including a polypropylene carry cart. This doesn't have a lot of welding-time capacity, but is popular for maintenance work, metal sculptors, and automotive shops that don't use a torch too often.

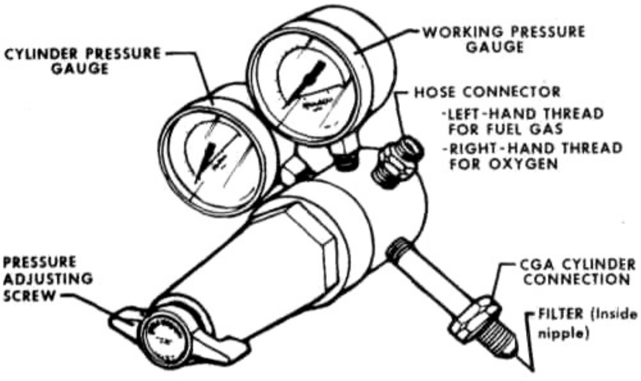
CYLINDER PRESSURE GAUGE
WORKING PRESSURE GAUGE
HOSE CONNECTOR
-LEFT-HAND THREAD FOR FUEL GAS
-RIGHT-HAND THREAD FOR OXYGEN
PRESSURE ADJUSTING SCREW
CGA CYLINDER CONNECTION
FILTER (Inside nipple)

3.10 The gauge/regulator setup screws onto the tank's cylinder valve, and has a fitting for attaching the hose leading to the torch.

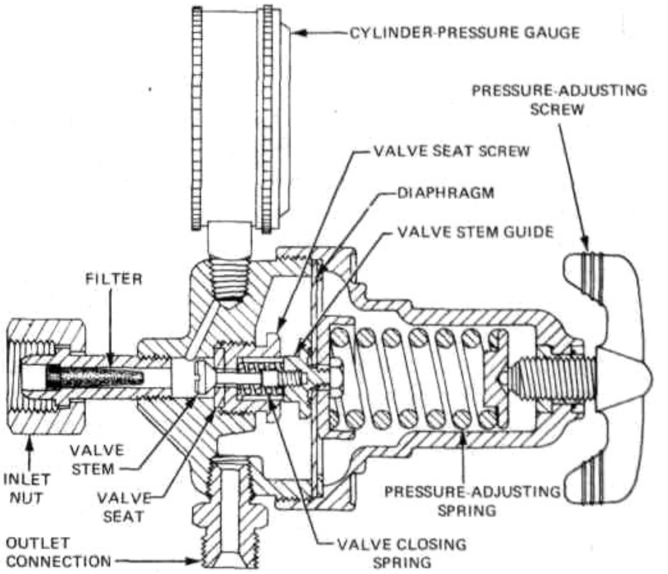
CYLINDER-PRESSURE GAUGE
PRESSURE-ADJUSTING SCREW
VALVE SEAT SCREW
DIAPHRAGM
VALVE STEM GUIDE
FILTER
VALVE STEM
INLET NUT
VALVE SEAT
PRESSURE-ADJUSTING SPRING
OUTLET CONNECTION
VALVE CLOSING SPRING

3.9 Cutaway of a single-stage regulator shows how the cylinder pressure is adjusted down to a usable range for welding. There are two gauges attached, one shows the total pressure left in the cylinder, and one shows the adjusted pressure supplying the torch.

one gauge to indicate the pressure in the tank, while the other gauge reads the pressure *after the regulator,* which is the pressure delivered to the torch and a very important adjustment in oxy-acetylene welding.

There are two basic types of regulators found on gas welding setups, single-stage and two-stage. The former type reduces the cylinder pressure to a working pressure in just one step, say from oxygen's 2200 psi to 3 psi for the torch **(see illustration).** In the two-stage regulator, the first stage reduces pressure to about 30 psi; the second stage adjusts pressure from there down.

A great portion of the expense in a gas-welding setup comes from the gauges/regulators, and better ones simply cost more. The simpler-to-manufacture single-stage regulators cost less and are therefore commonly found on the less-expensive kits for oxy-acetylene welding **(see illustration).** They work fine for most purposes, and their only drawback is

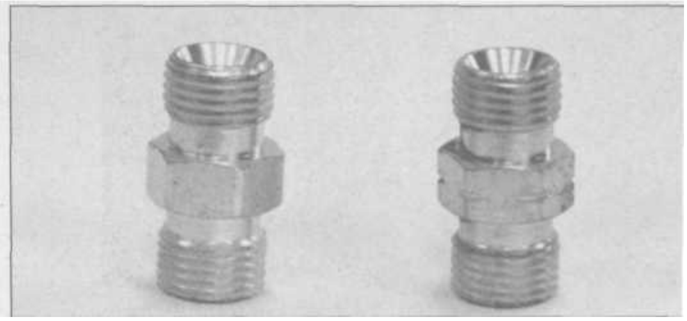

3.11 Connectors are available to add more hose onto a starter set for a longer working distance if you need it. These screw into your hoses and the extra new hoses screw on the other side.

that the regulation depends heavily (excuse the pun) on the pressure inside the tank. As the tank pressure goes down as the gas is used up, the regulation changes and must be adjusted again. Also, when doing a job that requires a lot of gas flow, such as large welds or cutting/heating thick plates, the single-stage regulator may not be able to keep up. The more expensive two-stage regulators maintain gas flow to the torch with less fluctuation, which is important in a professional welding operation but may not be a big factor in home/shop welding if you don't work on heavier materials. Obviously, gas-welding sets with two-stage regulators will cost more initially than the single-stage kits.

The hoses for a gas-welding setup are specially-designed to carry the oxygen and acetylene gasses. While they do not handle the very high tank pressures, they are built in three layers to withstand the *regulated* pressures for many years with proper care and maintenance. As mentioned above, the hose for carrying acetylene is red, while the oxygen hose is green, and the acetylene hose has left-hand threads on its fitting, so there is no chance of mixing them up when making connections **(Note:** *Left-handed fittings generally have a groove cut in around the fitting, to differentiate them from normal, right-hand fittings).* Doing so could be disastrous, since any residual flammable gas or other material in a hose connected to pure oxygen could cause the hose to ignite or explode. For this same reason, it is important that no lubricants or chemical compounds of any kind be used on the hoses or fittings. Gas-welding packages may contain hoses around 10-12 feet long in the economy or light-duty kits, and 20-25 feet long in the heavier, more professional sets. Extension hoses can be added to your starter set if you need to reach areas further from your tanks **(see illustration).**

3.12 In industrial situations, like this oilfield service truck, gas welding setups may be equipped with large cylinders and 100 feet of hose, to handle any situation.

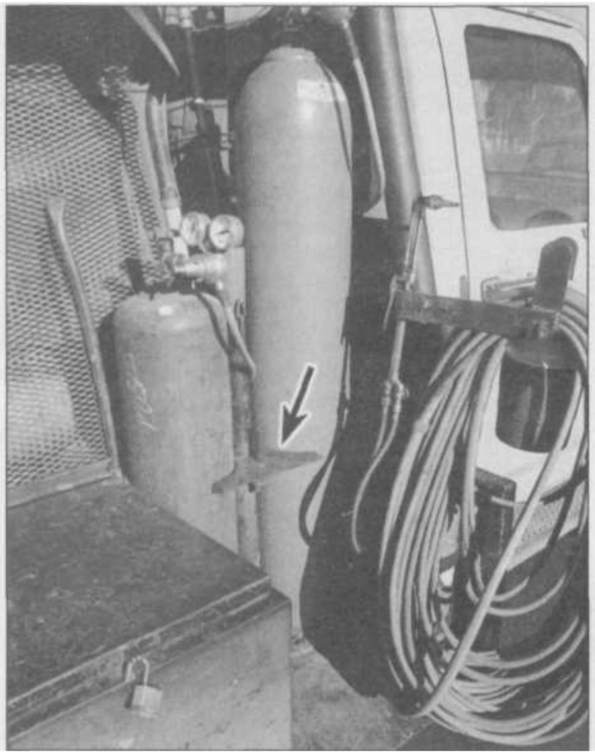

3.13 In this clever setup, the hoses and torch are attached to a swingaway arm for access to the tanks for servicing. Arrow indicates retaining plate with a simple steel wedge rod. Tanks are held securely but can be changed quickly without wrenches.

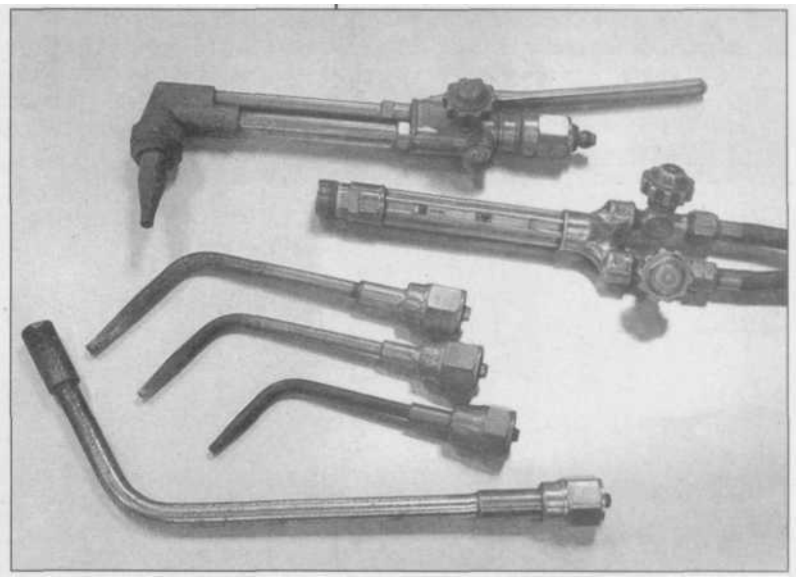

3.14 Basic torch sets include, from top: cutting torch attachment, basic torch head, a variety of welding tips sized for different thicknesses of material, and at bottom a rosebud tip for heating large areas.

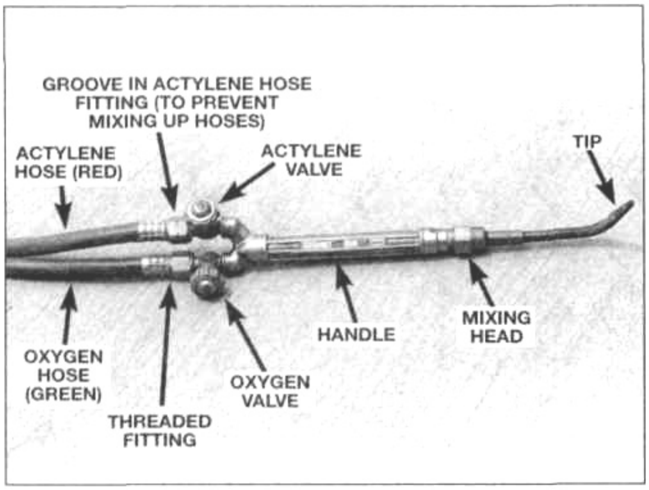
GROOVE IN ACTYLENE HOSE FITTING (TO PREVENT MIXING UP HOSES)
ACTYLENE HOSE (RED)
ACTYLENE VALVE
TIP
OXYGEN HOSE (GREEN)
THREADED FITTING
OXYGEN VALVE
HANDLE
MIXING HEAD

3.15 When assembled with welding tip and hoses, this is what an oxy-acetylene welding torch looks like. The two gasses are controlled by separate knobs, and are mixed inside the torch body.

The torch for oxy-acetylene welding is a precision-made tool, usually constructed with a forged brass body, two knobs controlling the gas flow, and copper tips of varying size **(see illustration).** In use, the two gasses flow through valves controlled by the knobs, and mix inside the torch with the combined gasses coming out of the tip in an oxygen-to-acetylene ratio controlled by the weldor. The weldor changes tips, each of which has a carefully sized opening, to suit the gas-flow and flame shape required for a particular job. The makeup of the cutting torch is different from the welding torch and will be discussed later on.

Gas welding sets can be purchased three ways: as a welding-only setup, a cutting-only setup or a combination set with two torches. The type you choose will depend on your needs. If all you need to do is weld up animal corral fencing made from steel pipes, you may use a cutting-only gas set for cutting pipe to length and shaping the cut ends to fit against the next pipe, and use a generator-driven arc welder on your truck for joining the pipes. However, if you are doing automotive sheet metal work or metal sculptures with relatively-thin materials, you will probably do most of your cutting with an electric shear, hand tin snips, hacksaw, etc., and only require a torch for welding, brazing and occasional heating. A combination set is of course the most versatile. Most sets that you buy can be fitted with a different torch at any time, so if you decide to buy a welding-only setup, you can purchase a cutting torch later on if your need to.

After you purchase your welding package, you will have to buy or rent tanks. The standard-size oxygen and acetylene tanks will hold enough gas to last the home/shop user for a long, long time. The tanks must be *hydrostatically-tested* periodically, which involves filling them with water to a certain pressure to test for safety. When the test is due, your local gas supply cannot refill your tank until it is tested. When you buy new tanks outright, you then become responsible for the testing costs and any cylinder-valve repairs, while if you lease the cylinders, the gas company takes care of that as part of the lease agreement. Leasing the cylinders allows you to get started welding with the minimum investment, because outright purchase may be expensive. However, if you add up the monthly rent

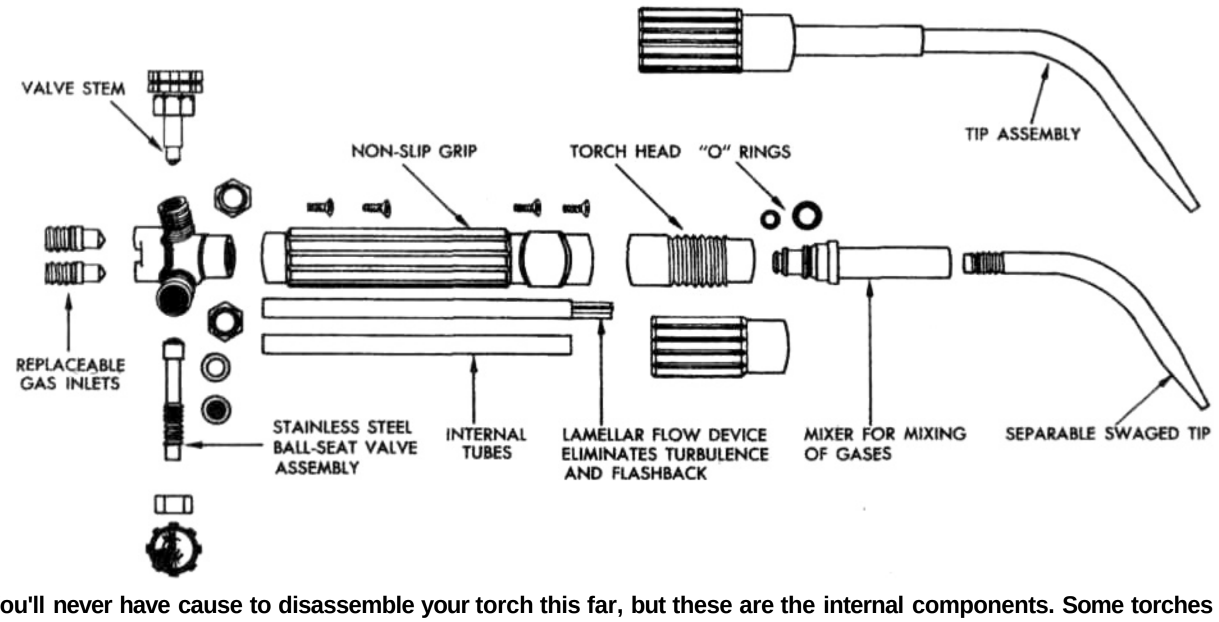
VALVE STEM
NON-SLIP GRIP
TORCH HEAD
"O" RINGS
TIP ASSEMBLY
REPLACEABLE GAS INLETS
STAINLESS STEEL BALL-SEAT VALVE ASSEMBLY
INTERNAL TUBES
LAMELLAR FLOW DEVICE ELIMINATES TURBULENCE AND FLASHBACK
MIXER FOR MIXING OF GASES
SEPARABLE SWAGED TIP

3.16 You'll never have cause to disassemble your torch this far, but these are the internal components. Some torches require wrenches to change tips and heads, while others have knurled, hand-tightened connectors.

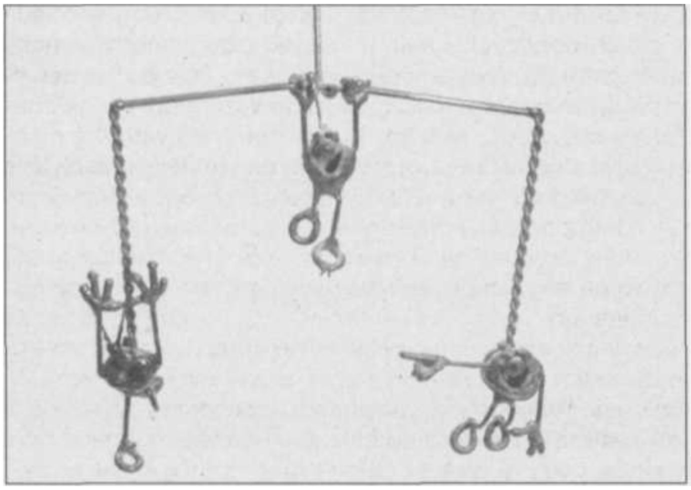

3.17 **Each of these little "flea circus" figures are welded up completely of thin welding rod and tiny blobs of molten metal. The artist works with a very small welding tip and tremendous patience, showing what can be done with skill and imagination.**

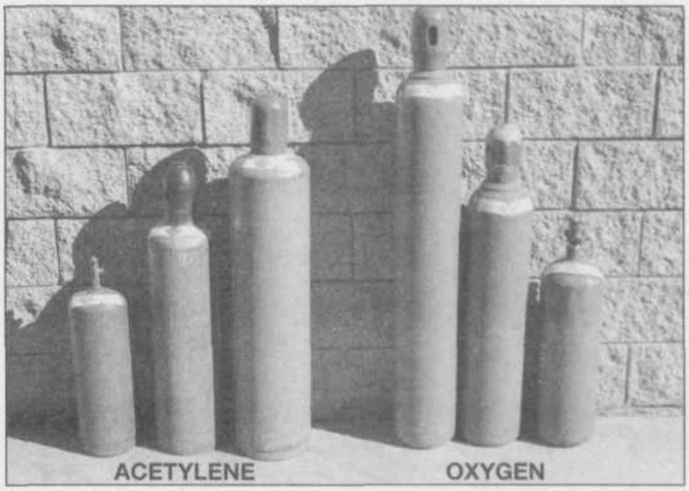

3.18 **The small, medium and large welding cylinders contain 80,150 and 275 cubic feet of gas, but some manufacturers label theirs by a letter designation. Ask your welding supply shop approximately how long each size would last in terms of welding, and choose according to your needs.**

charges, you'll find that after two years of leasing you could have purchased the tanks outright for the same money. Most leases run from 5-99 years, but if you move to another area where a different gas company is located, you may be able to return your tanks to the original location where you leased them and get back a pro-rated portion, then begin a new lease account in the area you move to. Look in your local telephone Yellow Pages under "Gasses, Industrial" or "Welding Supplies."

There are usually three sizes of tanks, small (80 cu. ft.), medium (150 cu. ft.) and large (275 cu. ft.). The small tanks are the easiest to store and maneuver around the shop and are fine for most occasional users. The medium tanks will be more than enough for most any home/shop use, lasting more than a year under normal use. The largest tanks are heavy, and more difficult to move around the shop.

Getting started with oxy-acetylene

Thoroughly read all the directions that come with your gas-welding setup before you do anything. If you have purchased a used welding set from someone, make sure the instructions are included, or have an experienced person help you get set up. Compressed gasses are dangerous; so is the welding process, and there are many more safety considerations to understand than we can cover here in this book. Obviously, since we are dealing here with a very hot flame and the promise of flying sparks, you should have a clean work area away from any flammable materials and a good fire extinguisher should be handy, as well as a source of water. If a spark gets onto your shoe, don't waste your extinguisher on that. Use a water spray bottle, which is also handy at times for cooling off a piece of metal.

Follow your set's directions for attaching the gauge/regulator sets to your filled bottles of oxygen and acetylene **(see illustration).** The bottles should be secured with a sturdy

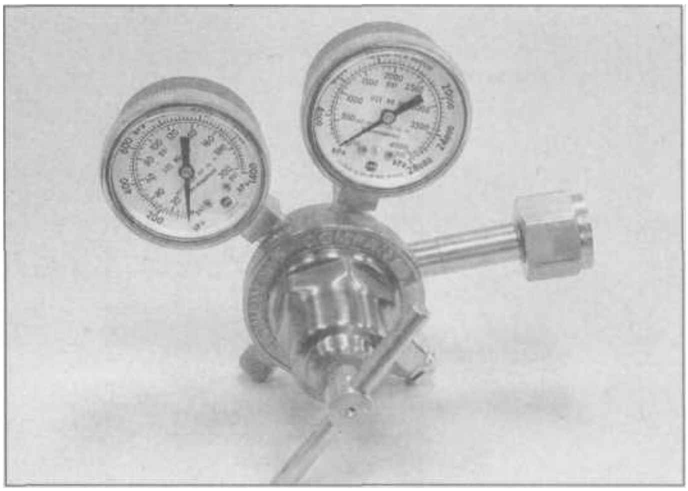

3.19 **Most home/shop users of oxy-acetylene will have single-stage regulators on their torch sets. Unless you do a lot of welding/cutting, these are just fine, though you have to readjust the output pressure more frequently than with the professional, two-stage regulators.**

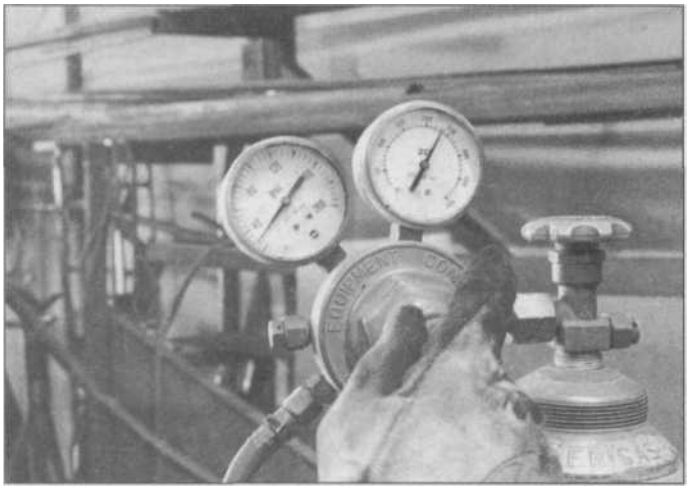

3.20 Gas flow is regulated on the oxygen cylinder to 1-5 psi for welding, depending on the thickness of the material to be welded, and the acetylene pressure is adjusted the same. Higher pressures are set when performing cutting, but acetylene must never exceed 15 psi to the torch.

chain to either your bench or a welding cart that has provisions to keep it from tipping over easily. Before the regulators are attached, blow any dust out of the cylinder valves by *cracking* them. You should stand to the side of the cylinder when doing this, with the outlet pointed away, and very quickly open the valve slightly and close it again. Just a little blast of gas will ensure that the connection is free of dust.

Use a proper-fitting wrench (not an adjustable) on all your regulator and gas hose connections, which are usually brass fittings, and remember that the acetylene connections are *left-hand* threaded. Blow out the regulator by slowly opening the cylinder valve when the pressure-adjusting screw (usually T-shaped) is turned out all the way. One gauge will show you the cylinder pressure, and the other the hose/torch pressure. Attach the hoses and blow them out with a blast of gas. Adjust the regulators to provide 2-3 psi working pressure from each bottle **(see illustration).** The cylinder valves should always be opened slowly, with your body away from the ends, and the oxygen valve should always be opened all the way. It is of a type that seals best either fully closed or fully open. The acetylene valve should be opened only about 1-1/2 turns.

Connect the hoses to your torch with the gas cylinder valves closed. In most cases this involves wrenches, but there are some professional sets that have snap-on connections (like air line hoses) to make quick, wrenchless work of changing torches **(see illustration).** Select a medium tip from your set (your set's instructions will recommend what gas pressure works with which tip and what thickness of metal they are suited for) and attach to the torch. After everything is assembled, you can test the connections with soapy water for leaks.

You will need something to light the torch flame, and a wire-framed metal striker should have been included with your set **(see illustration).** Never use open flames such as a cigarette, cigarette lighter or match to start your torch! Your fingers need to be well out of the way. You should be wearing leather gloves and heavy clothing with leather shoes or boots. Do not tuck your pants into your boots, or you could create a pocket there that could collect sparks. Likewise, do not wear pants with cuffs, for the same reason. Have your colored welding goggles on your head, but not yet over your eyes.

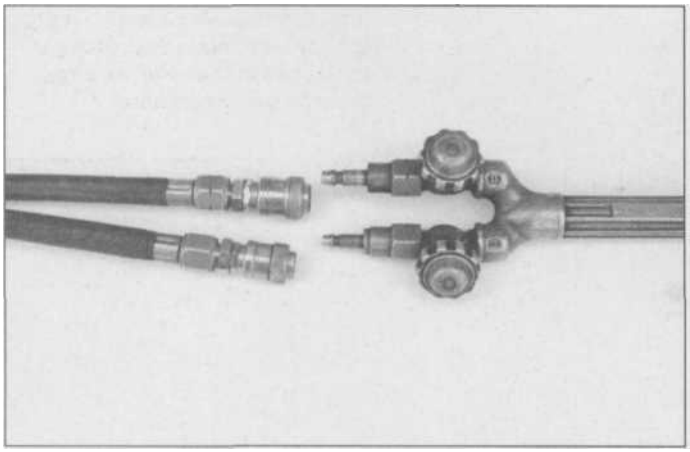

3.21 Some torches are equipped with quick-release fittings as usually seen on air-compressor hoses, making for quick, easy changes.

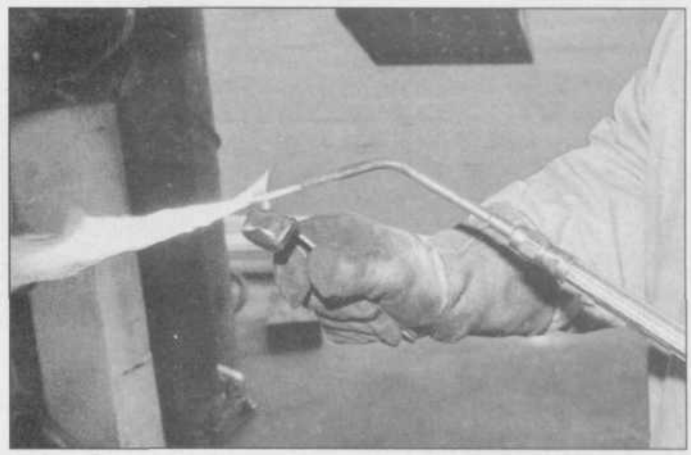

3.22 To start a torch, open the acetylene valve 1/2 turn and light immediately with a striker like this. Never use a match, cigarette or lighter, you could easily burn your hand. Once the sooty acetylene is lit, gradually add oxygen until you achieve a neutral flame.

Flame adjustment

This is a very important step in gas welding, and one you will be making every time you pick-up the torch to weld, braze, cut or solder, so you'll have to practice this to the point where it becomes second nature. Open the cylinder valves slowly as described before, and make sure the regulated pressure is set correctly. Hold the torch in your operating hand (right, if you are right-handed) and the striker in the opposite hand. Both valves on the torch are closed. Open the acetylene valve on the torch about a half-turn and operate the striker near the torch tip. There should be a flame with some sooty flying particles. Quickly add some oxygen and the soot should disappear. Study the illustrations of flame type. What you have right now is a *carburizing* flame, which is overly-rich in acetylene. As you add more oxygen, the flame characteristics will change. While the rich mix gives you a wide, flared-out yellow flame, adding oxygen will make the flame narrower and brighter.

The ideal flame for most gas welding is a *neutral* flame, which is neither carburizing nor oxidizing (too much oxygen). When you have adjusted the torch properly, there will be a large flame graduating from blue at the end away from the tip down to white near the torch and with an inner cone that should be light blue. If you add too much oxygen, the inner cone will be smaller, pale in color, and there may be a hissing sound. You'll need to practice achieving the neutral flame, and when you start welding, you'll see what happens to metals when the flame is incorrect.

The proper method of shutting off the torch is equally important. Close the acetylene valve on the torch first, then the oxygen. If you do it the other way around, there could be a small amount of gas left in the tip that could ignite. If you are through welding for some time, it is good safety practice to then close down the gas cylinder valves as well, but if you are going to weld again in a short while, just shut off the torch valves only.

Gas welding

You'll need a lot of small pieces of clean scrap metal to practice on, of varying thicknesses. Even if you only intend to do light sheet-metal welding, start by practicing with thicknesses up to 1/4-inch to become familiar with different tip sizes and flame adjustment. The cleaner the scrap pieces, the

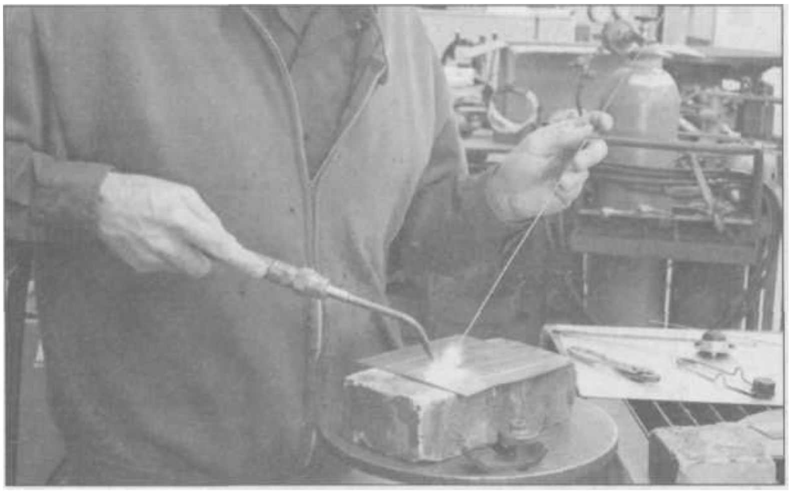

3.24 You don't have to worry about grounding with gas equipment, but there is a dangerous flame involved, so your welding projects and practice sessions are best done over some firebricks like this, which you can get at your welding supply store.

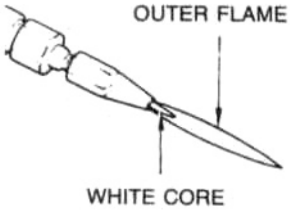

Neutral Flame

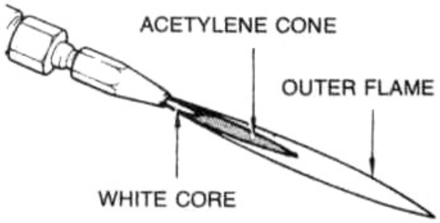

Carburizing Flame

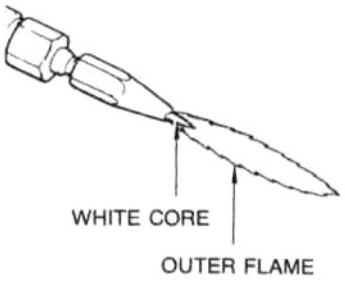

OUTER FLAME

Oxidizing Flame

3.23 The oxidizing flame has too much oxygen, and a feathery blue flame with small white inner flame. The carburizing flame is acetylene rich, and features a large flame, with a large blue inner cone and a longer, colored cone around that. For welding, adjust for a neutral flame.

better your welds will be, and this applies to any type of welding, gas or electric. You should sand or grind the areas to be welded to remove any paint, coatings or rust. You should not use for testing any metals that are galvanized or coated with zinc, or any kind of plating. Not only will welding be dirty and difficult, but the coatings may cause harmful gasses when vaporized in the weld.

Start your first exercise with a piece of steel on a level, non-flammable surface, not a wooden bench. The ideal work area would be a steel table (not just a wooden bench with thin sheet metal nailed onto it) with some firebricks on top. The firebricks can be obtained at your welding supply center, and, in addition to being non-flammable, they do not suck the heat out of your welding sample like a steel table, and there's no chance you're going to weld the sample *to* the brick. It can be very embarrassing when you practice on a steel table and you have various scraps welded - stuck to your table.

With your torch adjusted properly, pull down your goggles and put the torch tip about one inch from the steel and just try to make a long weld puddle of molten metal. The torch should be leaning back at a 45-degree angle, with the flame pointing in the direction the bead is going. When the steel starts melting into a small pool where you started, move the tip along and give it a rhythmic motion of an oval shape as you go, trying to follow a straight line, blending the newly-heated edge of the puddle with the back edge of the puddle. The process will leave a line of rippled ridges. If you are burning holes in the metal, there may be too much oxygen, the torch may be too close to the metal, the tip could be too large or you are possibly advancing the torch too slowly. If you go too fast, have the torch too far away or the tip is too small, your puddle will be very small and it won't penetrate the metal very far. The more the torch is pointed straight down at the metal (no angle) the deeper the penetration will be, and vice-versa. Practice at different angles until you can make a bead like our illustration, with even ripples and proper penetration.

Of course, not all of your welding will be done this way, in fact almost all of your real welding will involve adding filler rod, but for now you need to learn torch control and rhythm. Many of the projects you encounter will involve welding along an irregular line, especially in repair work, where you are welding up a crack. To become more familiar with making a bead along an irregular line (still without filler rod) draw a zigzag or curved line on the steel with soapstone and practice making a good bead. Proper hand control is everything in gas welding. Your hand must be comfortable and steady at all times. In welding classes, students are often challenged to write their first name in script with welding bead on a piece of metal, which challenges you in handling changes in bead direction.

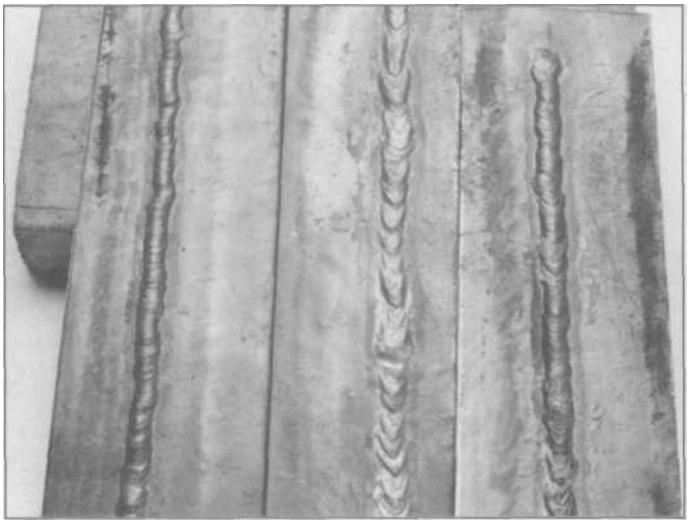

3.25 Heat in gas welding is controlled by matching the gas flow and tip size to the thickness of material. Your welding travel speed and tip-to-work distance will control the puddle size and penetration. With all things right, you should be able to heat your first puddle molten in about six seconds, then move along to make your bead. At left is a bead run too cold (bead stands up), at center is one run too hot and too fast, and at right is a better bead with good width and penetration.

When you are successful with these experiments laying down a good bead on the metal, you're ready to try joining two pieces using a filler rod. This requires not only good torch control, but control of your other hand as well with the filler rod. Generally, the filler rod should be the same material as the base or parent metal for good fusion welding. If you weld aluminum, the filler rod must be aluminum, and steel rods are used on steel. The steel filler rods are coated with copper to keep them from rusting and to make cleaner welds. As a basic rule, select a filler rod that is the same diameter as the thickness of the metal you are welding. Most light metals are welded with 3/32-inch, 1/16-inch or 1/8-inch rod, and you will probably never gas-weld with anything bigger than 1/4-inch rod. To weld thick materials, several passes are required to make a good fusion, and various forms of electric welding are usually much faster than a torch at making multiple-pass seams.

Welding with filler rod

Let's say that you need to join two pieces of 1/8-inch-thick metal in a straight bead from right to left. Position the two pieces such that they are butted together where they are to be joined, and that you have something blocking or clamping them in place so they won't move as you weld. Start your weld puddle at the right end of the seam, and as the puddle develops, use your left hand to insert the tip of the filler rod into the center of the hot puddle, then pull it out right away. Set up a rhythm of alternately putting in the rod and taking it out as the bead progresses to the left. When we describe the puddle, the edge of the puddle closest to the direction you are traveling is the *front* edge, while the part of the puddle you just finished is the rear.

If you have ever done any soldering, you may have a bad habit of heating the solder with your torch or soldering iron and letting the molten lead drip onto the joint. This is a bad habit and one you shouldn't bring with you in your welding practice. The molten puddle should melt the filler rod, not the torch. If the rod sticks, then you are not keeping the puddfe molten enough, or you are adding filler rod too fast, either of which will result in a low-quality weld.

This skill of making a good bead, with evenly overlapping ovals and good penetration will take considerable practice, but if you learn two-handed gas-welding techniques well, any other welding system you pick up later on will be considerably easier to learn. Most of the welding you will do will be *forehand* welding, in which you angle the torch so that the flame is aimed somewhat toward the direction of the desired weld bead. This preheats the metal as you go and makes addition of filler rod to the front edge of the puddle easier because it is truly hot. Sometimes it seems like keeping the proper torch distance, moving at the right speed, circling or ovaling the tip and dipping the rod in at the right time is just too much to concentrate on at one time, but you *will* get the hang of it. This isn't "brain surgery" and millions have learned before you. After a good long practice session, begin to examine things around you that are of welded construction and you'll get an appreciation of the practice it took to lay down those beautiful beads that look like they were done by some machine instead of human hands. Indeed there are some kinds of welding that are done with largely robotic welding machines, but a good weldor can make virtually perfect joints that are stronger than the parent metal.

We mentioned before that cleanliness is critical in making good welds. The two edges to be joined need to be sanded or ground clean to bright metal and should fit as closely as possible. Small gaps are inevitable in some kinds of work and can be bridged with filler rod, but gaps should be kept to a minimum. On thicker materials, you will find that it is easier to make a clean weld with good penetration if the edges of the two pieces are beveled before you weld. This can be done with a file or grinder. Thin materials have their own idiosyncrasies, mostly to do with keeping the two edges in alignment. As you progress with your weld bead, more and more heat travels out into the parent metal, so that after welding for some distance, the metal you now encounter is hotter than the metal you started on, so you may have to adjust the torch away slightly to compensate for the puddle forming faster.

Another problem with thin materials is that the parent metal tends to *distort* with the application of heat. Metal expands when heated, and two edges that were parallel when you started welding probably won't stay that way. Usually, the edges pull away and you have a gap to bridge that wasn't there when you started. The solution is to *tack-weld* at intervals before applying too much heat. A tack-weld is

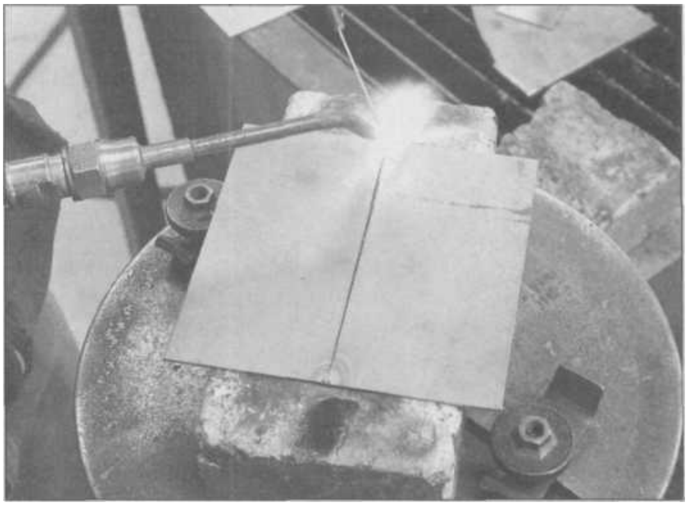

3.26 When practicing basic butt welds, tack the near corner first, then tack the top corner with a slight gap between the parts. This will close up as the plates grow from welding heat.

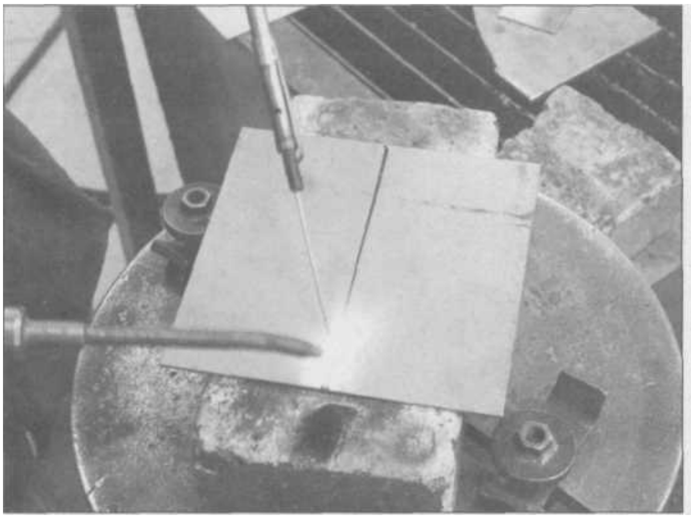

3. 27 Start your bead about an inch beyond that first tack, and proceed toward the end, keeping the torch steady and dropping molten beads off the filler rod by dipping into the puddle quickly and evenly as you go, building an overlapping series of puddles.

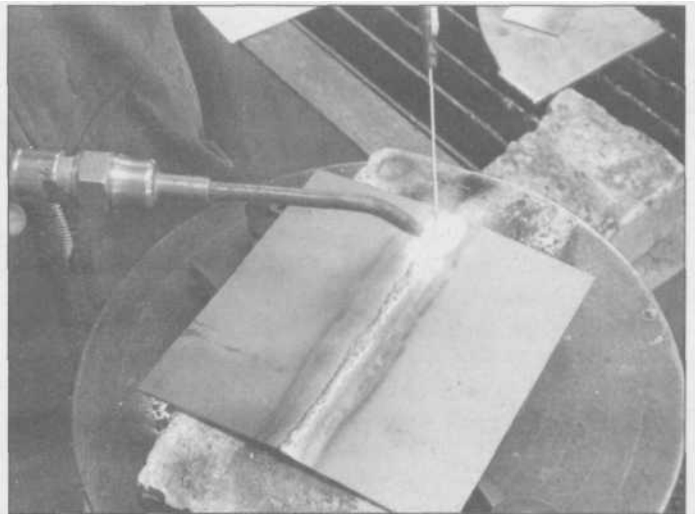

3.28 Before you are done, you will notice some movement or warping of the sample plates. If you plan to do automotive body work with a torch, you'll have to practice a lot on thin metal to get the hang of welding without burning holes or warping the work. Gas welding is not suited to the thin, high-strength body steels used on today's unibody cars.

simply a bead only one or two puddles long, done quickly. To join a long seam in sheet metal, tack the ends first, then the middle, and then put tacks every few inches apart, while alternating the ends you tack **(see illustration).** When you go along later to do the final, full seam welding, the tack puddles can be melted into the main seam as you go, and the pieces will stay aligned the whole way. This is particularly important when joining long sheet-metal seams like welding a new quarter-panel onto a car. In auto bodywork, you need to keep the warpage to a minimum. In some cases when doing the final seam after tacking sheet metal, you may stop after a few inches are welded and quickly apply a body dolly to the backside of the weld area and hammer on the weld area on top with a body hammer. This "hammer-welding" (alternating welding and hammering) technique straightens the seam and the two panels while they are still hot and pliable, and the resultant seam when done by a good body man requires much less grinding or filling to make it ready for primer and paint.

The basic joint we have illustrated is a buff joint, where two pieces of metal are pushed together and welded along the seam where they meet. There are many kinds of joints, and it is helpful to try welding a few other types after you have practiced enough on butt welds on thick and thin materials. *Corner* welds, where the two edges meet at an angle can be either inside or outside corners **(see illustrations).** The outside corner weld is the easier to perform because you can see the seam so clearly, and the pieces need not be beveled. Just placing them so that the edges form a groove or V will do it. On inside corner welds, depending on the angle involved, the action is harder to observe, being shrouded somewhat by the pieces. After you have practiced on these more difficult welds, try a lap weld. This is where the two pieces lay over each other (overlapping). You run a bead where the joint is, and depending upon the strength required for the application, you may want to weld the seam on the other side as well. In auto bodywork, sometimes panels have to be overlapped instead of butt-welded, but in such cases the backside seam is seldom welded.

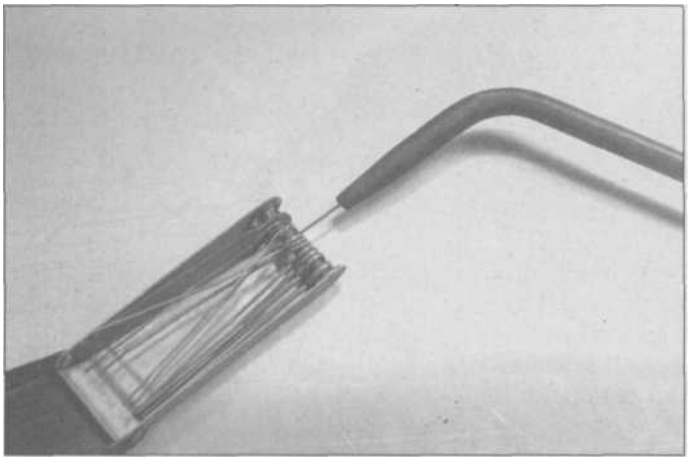

3.29 As you weld, it is inevitable that sparks and spatter will blow back onto your torch tip. If it plugs up, your bead will be spotty and your torch may go out entirely. Keep the tips clean with a set of fine-wire cleaners that come with your gas set.

A T-weld is where you join a piece of metal to another larger piece at right angles, forming a letter T in profile. This requires some additional practice to perform well, as part of the weld puddle is being made uphill, and the larger piece may take more heat than the smaller part. You have to play the torch back and forth over both pieces and get a feel for how much heat each requires.

Further practice sessions for you may involve trying to weld pieces of tubing and angle-iron together, both typical materials for home/shop projects. Try joining a flat plate to a tube, or making a corner between two pieces of angle-iron. Careful measurement and cutting will be required to make gapless corners for a good weld. Square or rectangular tubing is often used in projects such as shop equipment or a utility trailer, and you should practice making 90-degree corners from two pieces cut at 45-degree angles. The inside welds will be the tougher ones, and you'll find when you stop practicing and assemble a real-world project that you can't always turn the work around to make the seam area flat in front of you on your bench. Often you will be forced to weld in a vertical seam, or even weld upside down underneath something. The latter is difficult and dangerous, ' requiring the best in full-face eye protection (a helmet instead of goggles) and a leather apron over your arms and shoulders to protect you from falling sparks. Welding an exhaust system under a car, especially if you don't have access to a lift, can be extra challenging.

When these don't seem like a challenge anymore, you're doing very well, and you can try making nice-fitting joints between pieces of *round* tubing. To fit two pieces of round tubing together, such as in a roll-cage for a race car, requires a measure-twice, cut-once procedure and lots of patience in cutting and grinding round tubing to join them with full contact. Another procedure to try with round tubing is to make a slip-fit, with one piece of tubing sized to just fit over another. In such cases, you can weld around the joint, or, for high-strength applications, cut the outside tube in a "fish mouth," as if it were being joined at a 90-degree angle to another piece of round tubing. When this piece is welded onto the inner tube, the fish mouth makes for a much longer weld seam between the two and gives greater strength.

Other types of welds you may run across in your projects are spot-welds and rosette welds. The former is found often in auto sheet metal work, where you join two pieces of light-gauge material together without welding the seam entirely. The spot-weld is just a round welding spot, rather than a bead, and is made when one piece is on top of another and you just melt a puddle that fuses the two panels together, then move a distance away and make another spot-weld. Filler rod is often not used in such cases, especially where appearance is more important than strength. Most spot welds lie flat enough that little prep is required to make them ready for paint. Modern cars are welded with thousands of spot-welds, but these are accomplished by special electric welders, usually computer-controlled and robotic.

Where a little more strength than a simple spot-weld is required, a rosette or plug-weld can be made. This is where the panel on top has holes drilled or punched into it at intervals. You weld up the holes, puddling both upper and lower panels together, which joins them securely in that spot. Plug-welds often do require some filler rod.

Sometime in your gas-welding work, the torch will go out and you'll get a loud pop at the torch which will really get your attention because you don't expect it. When welding in tight quarters, you may tend to get the tip too close to the work, which will overheat the tip and cause a tiny explosion as the gasses POP inside the tip, which is called a backfire. This will alarm you greatly when it happens, but the answer is simply to not get so close. You may have been working close to get the pieces hot enough, but if that is the case, you really need a larger tip or more flow of oxygen and acetylene from the tank regulators to maintain the right heat level with your torch further away from the confinement of the weld corner. Welding too close to the work will also dirty your tips. So if you have a backfire, stop work, shut off the gas at the torch and remove the tip for a

FLASH ARRESTORS
TORCH-MOUNTED TYPE
UL-LISTED
STOP DANGEROUS FLASHBACKS
L-TEC
L-TEC
L-TEC
WELDING & CUTTING SYSTEMS
OXYGEN and FUEL GAS PAIR
• Fit on all torches with "B" size inlets
• Oxygen inlet pressures up to 150 psig
• Acetylene inlet pressures up to 15 psig
• LPG, MPS, Nat. Gas inlet pressures up to 50 psig
• Both models incorporate reverse flow check valves
• Delivery pressure must be increased to compensate for pressure drop — see instructions
P/N 20357

3.30 These prevent serious accidents from a flashback, indicated by a loud squealing noise. Flashbacks are caused by gasses burning *inside* the hose. For safety reasons, both hoses should be equipped with flash arrestors, which are reverse-flow check valves. Some readjustment of gauge pressures may be necessary when using check valves.

thorough cleaning. Your welding set will have come with a set of wire tip cleaners. Cool off the tip with a water-soaked rag.

Something akin to a torch backfire but much more dangerous is a "flashback." This is where combustion actually follows the gas trail back into the torch and even the hoses. At the least there is danger of damaging the regulators and hoses, but if the combustion reaches the tanks, there is a possibility of a serious explosion. In a flashback, the torch goes out and there is a loud squealing or hissing noise with black smoke coming out of the torch. If this happens, shut off the oxygen at the torch, then the acetylene at the torch, then the cylinder valves. Your welding supply center should have available "flashback flame arrestors" that install in your two gas lines where the hoses attach to the regulators **(see illustration).** Acting like one-way valves, they are designed to prevent serious accidents caused by flashbacks, and they should be installed on any gas-welding setup, where the hoses attach to the regulators.

When making welds with filler rods, you will run across situations where the rod gets short enough for your left hand to get too close for comfort to the heat of the welding. Before making a bead, practice will have shown you how fast the welding rod gets used up and how long a rod you will need to complete the seam. Start with a rod long enough to do the whole seam. If the seam is longer and will take several rods to complete, you will have to stop welding and get a new filler rod. It will take some time and practice to get the hang of continuing a bead with a new rod where you left off before. The trick is to make the whole bead look like a continuous weld, without little piles of buildup where you stopped and restarted. When a rod gets short, many welders simply tack the short piece onto the end of a new rod and go right on welding. When you get good at it, the bead won't have a chance to cool off too much. In inside-corner welding, you may want to take the welding rod and put a bend into it, such that it puts your hand in a more comfortable position, further away from the heat of the action **(see illustration).**

Checking your welds

Chances are, you won't be making nuclear-vessel welds in your home/shop, or other welds so critical that they require X-ray inspection to pass specifications, as is often done in commercial welding. However, you do want to learn how to make strong joints, and your practice time should include some testing as well. Looks are important, but a good-looking bead may not be a proper weld, especially if it doesn't have good penetration. Checking for good penetration requires looking closely at the backside of the welded joint, which is easy for you to do with your practice sample pieces. In real-world situations, it isn't always practical to get at the backside of something you have welded (like a joint between two rectangular or round tubes), so you should be familiar with the proper heat control and penetration before you consider yourself through with your practice phase.

In looking at the back of the weld, see if there is any buildup on the backside. A small amount of "dropthrough" is acceptable, but 100% penetration can be

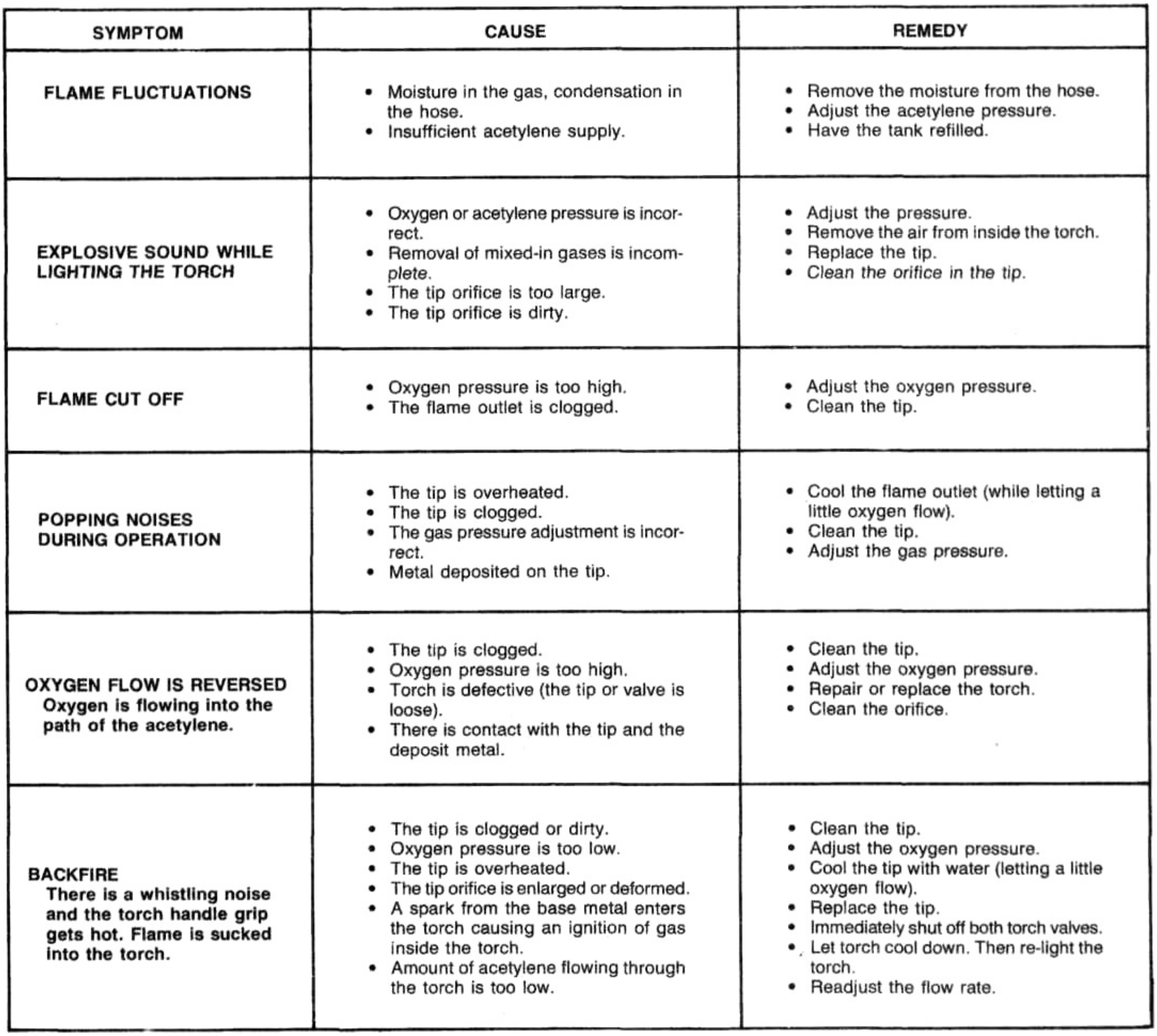

SYMPTOM	CAUSE	REMEDY
FLAME FLUCTUATIONS	• Moisture in the gas, condensation in the hose. • Insufficient acetylene supply.	• Remove the moisture from the hose. • Adjust the acetylene pressure. • Have the tank refilled.
EXPLOSIVE SOUND WHILE LIGHTING THE TORCH	• Oxygen or acetylene pressure is incorrect. • Removal of mixed-in gases is incomplete. • The tip orifice is too large. • The tip orifice is dirty.	• Adjust the pressure. • Remove the air from inside the torch. • Replace the tip. • Clean the orifice in the tip.
FLAME CUT OFF	• Oxygen pressure is too high. • The flame outlet is clogged.	• Adjust the oxygen pressure. • Clean the tip.
POPPING NOISES DURING OPERATION	• The tip is overheated. • The tip is clogged. • The gas pressure adjustment is incorrect. • Metal deposited on the tip.	• Cool the flame outlet (while letting a little oxygen flow). • Clean the tip. • Adjust the gas pressure.
OXYGEN FLOW IS REVERSED **Oxygen is flowing into the path of the acetylene.**	• The tip is clogged. • Oxygen pressure is too high. • Torch is defective (the tip or valve is loose). • There is contact with the tip and the deposit metal.	• Clean the tip. • Adjust the oxygen pressure. • Repair or replace the torch. • Clean the orifice.
BACKFIRE **There is a whistling noise and the torch handle grip gets hot. Flame is sucked into the torch.**	• The tip is clogged or dirty. • Oxygen pressure is too low. • The tip is overheated. • The tip orifice is enlarged or deformed. • A spark from the base metal enters the torch causing an ignition of gas inside the torch. • Amount of acetylene flowing through the torch is too low.	• Clean the tip. • Adjust the oxygen pressure. • Cool the tip with water (letting a little oxygen flow). • Replace the tip. • Immediately shut off both torch valves. • Let torch cool down. Then re-light the torch. • Readjust the flow rate.

3.31 After you have gained some welding experience, and a good torch weldor isn't trained in one day, you will learn to read the welding conditions based on the flame from the torch. This troubleshooting chart indicates some of the causes **and cures of torch flame problems.**

achieved without extra material building up on the backside. Good penetration is particularly important in welding auto body panels, because the outside of the weld bead is probably going to be ground down with a body grinder in preparation for filling and painting. Thus the strength of the topside buildup or puddling will be gone and the joint's strength will rely totally on the fusion of the two pieces below the surface. In your tests of practice welds, it can be helpful to use a hacksaw to cut through the center of one of your welded joints to see just what the penetration is. You can't always tell by looking at the ends of the joints, because of how you may have ended the bead.

We're assuming that your home/shop has a good, sturdy bench vise. It can be useful in testing your welded joints. Your first practice welds may have been straight seams with two pieces butt-welded together. Take your best example and put it in your vise with the weld seam close to the jaws **(see illustration).** Clamp a pair of vise-grip pliers on one edge of the upper piece and start hitting the upper piece with a hammer, while holding it with the pliers so it can't fly off. Hammer the upper piece until it's bent over all the way onto the vise (making a 90-degree angle in the welded pieces). Remove the sample from the vise and examine the weld. If the upper piece broke off during hammering, it obviously wasn't a strong enough weld.

Any of the various joints you may have practiced on can be tested by hammering. Whatever the joint, just secure the piece and hammer on one of the two joined pieces. Hammer from the backside area of the joint and toward the side that you welded on. If good penetration wasn't achieved, the backside will be weaker than the front and bending the metal this direction will really test the backside for good fusion. Even a weld with 100% penetration may not be strong enough if there was any "undercutting" on the top side. Undercutting is a groove along either or both sides of a bead, caused by applying too much heat or holding the torch at the wrong angle. In a stressed situation, the undercut area will be the weak point and the joint may fracture right along the undercut area.

Brazing

3.32 Brazing with a gas torch has many uses, such as repairing castings and antique-car sheet metal. Here a flux-coated rod is being used. Brazing is not a fusion process, so it is used where higher, melting-point heat might damage the work, or surrounding materials.

Brazing, or braze-welding, is not a true fusion process such as we have been discussing. The parent metal does not melt into a puddle. The parent metal is heated and the brazing rod melts into the joint securing the two pieces together. The brazing rod is made of brass (an alloy of copper and zinc) and is used in conjunction with a flux, which may come already on the rod, or in a can that you dip the rod into to coat it as you go. Brazed joints are as strong as welded joints when done properly and with an overlap of the pieces, but brazing is particularly helpful where the higher heat of fusion welding would harm the parent metal or adjacent materials, or where the pieces being joined are made of brass or copper.

The rod material melts at a much lower temperature than the parent metal, so the basic principle is to heat the parent metal just above the melting temperature of the filler rod, so that the parent metal, not the torch, melts the rod. Overheating the rod with the torch will boil the zinc out of the rod and cause poor adhesion in the joint. Brazing is most often done where there is an overlap of two pieces, or in building up a worn or broken area of a casting, that will be re-machined to original shape after brazing.

A joint to be brazed must be thoroughly *cleaned,* by grinding, filing, or sanding. When you have selected the proper rod for the material you are working on (copper/zinc for ferrous metals), you begin by heating the joint with your torch, using a lot less heat that you would for welding. You should pull the torch tip back 2-3 inches, which is much further back than with gas-welding. When the metal is dull red, apply the filler rod to the hot spot on the parent metal and see if it melts in. Heat next to that spot and apply some more fluxed rod, keeping this process going until you have completed your seam. When a long seam is contemplated with brazing, preheat the whole area. When the metal is the correct temperature, the brazing rod will flow nicely into the joint; if the metal is too cold, the braze material will ball up on the surface, and if it is too hot, it will spread out over a wide area. White smoke during the process usually indicates the area is overheated, and the smoke is the zinc boiling off. Not only does this affect the quality of the joint, but the fumes can make you sick if you breathe enough of them. Even if you don't overheat the rod, there are fumes given off during brazing.

Always braze with adequate ventilation, and a good respirator mask is also helpful insurance. Brazing is often used when joining pieces of galvanized metal because it will adhere better than standard gas-welding; however, this combination of brazing fumes and fumes from the plating or galvanizing should be avoided unless you are outdoors.

If you are familiar with basic soldering, you know that good adhesion is ensured by using *flux,* a compound that prepares the metal for a strong bond, and you may also be familiar with the term *tinning.* Tinning is a process (in soldering) of applying a thin coating of lead to the fluxed surface before the main solder is melted in. The procedure is similar in braze-welding, except that, when brazing thin materials, the tinning usually takes place as you are brazing, in one step rather than two. Brazing larger joints requires that a light coating of brass be flowed onto the surfaces, making it easier for the main brazing to stick to. The flux for brazing can be a part of the rod itself (pre-fluxed rods) which is convenient, or you may have a metal can of flux on your welding table. You simply heat the end of your brazing rod and stick it into the flux and withdraw it covered with flux **(see illustration).** Once you have started the brazing process, the rod will always be hot enough to pick up flux as you need more. The pre-fluxed rods are convenient, but the coating is somewhat fragile and won't take rough handling. These rods must also be kept very dry.

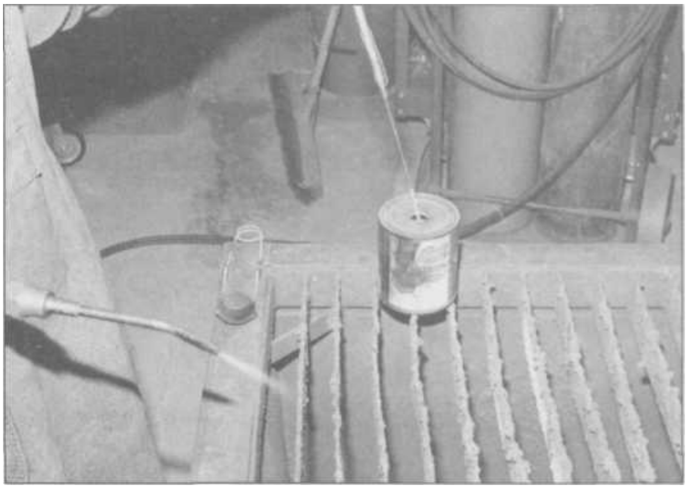

3.33 Most welders use dip-type powdered flux for convenience, since the pre-coated rods can flake if they are exposed to moisture. Plain brazing rod is heated with the torch, then dipped into can of flux and pulled out coated with enough flux for an inch or two of brazing travel.

When brazing thin sheet metal, the purchase area or overlap between the two panels is the key to the strength of the joint. When there is 3/4-inch or more of overlap for the brazing to adhere to, the joint will actually be stronger than the parent metal. Because the brazing rod melts at only about 1000° F instead of the steel's melting point of 2700° F, brazing will induce much less warpage in thin metal, which is why it has been used for years in many areas of traditional auto-body repair. The overlap in the brazed joint also makes the joint area much stiffer (two thicknesses of metal there), so a seam such as when installing a body patch panel will take less hammer-and-dolly work to prepare for painting **(see illustration).** Brazing will not work on a butt-joint because there isn't enough "captured"

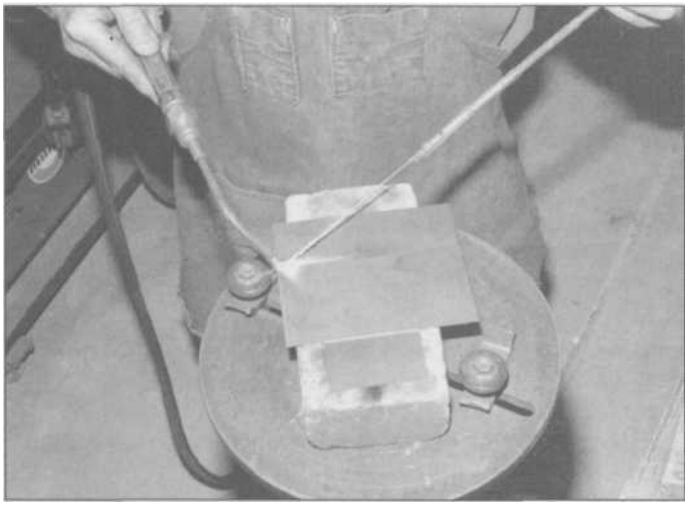

3.34 Lap welds are best for brazing, such as when joining overlapping sheet metal panels where the brass has lots of contact area for adhesion. As with fusion welding, tack the ends before running the whole seam.

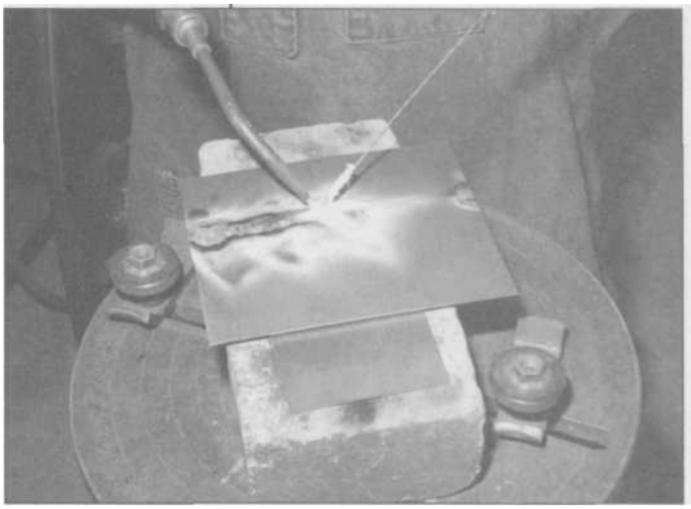

3.35 You should start with very clean metal when brazing, although the flux is designed to surface clean as you go, leaving some minor slag afterwards. You will have to keep the torch further away from the seam than with welding, to keep from overheating the brass.

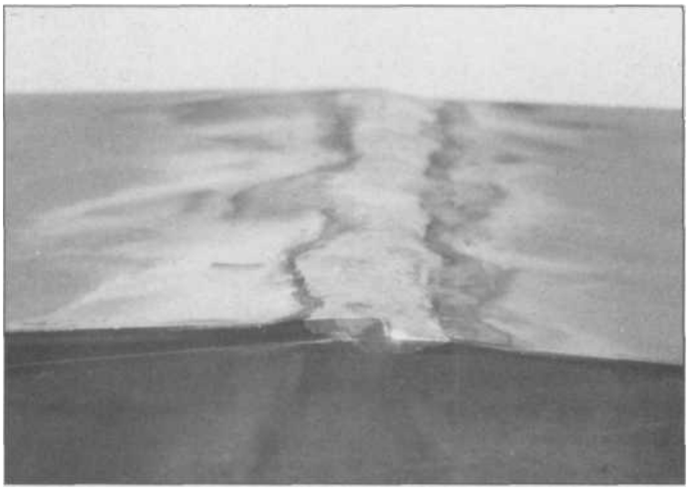

3.36 End view of a brazed lap joint, showing how the brass flows in between the plates with capillary action, much as solder does when applied to a hot copper pipe. The base metal must be heated to the point where *it* melts the rod, not the torch.

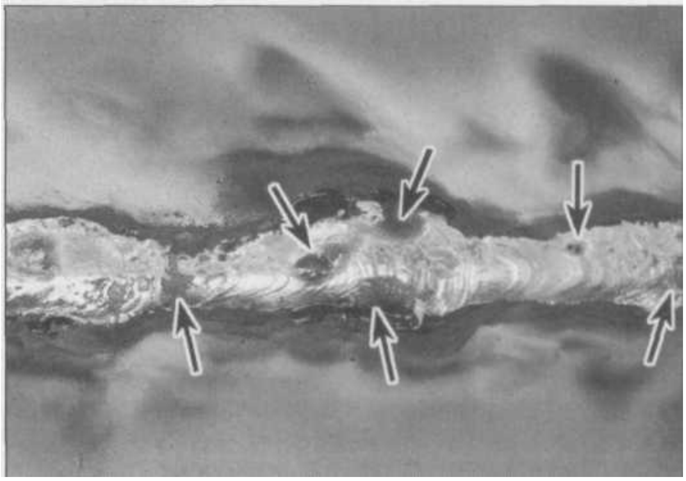

3.37 There is some silicon slag after brazing (arrows), which can be softened with water after the part cools enough so that water will not warp it, then the rest comes clean with wire-brushing.

area for the capillary action to flow the brass into each piece, although you can back up a butt-joint with a strip of similar material, which gives an overlap on both pieces and good bonding area for the brazing.

Although brazing is still a technique used by older body craftsmen in restoration of antique and collectible cars, it should not be used on new cars that are built of low-alloy, high-strength but lightweight steels. In fact, some car manufacturers specify that no torch work at all should be done on their body metal, that only MIG welding is acceptable because of the narrow heated area and faster welding-bread travel.

Other metals than steel can be brazed as well, such as stainless-steel, castings, brass, bronze and even aluminum when a special aluminum-brazing rod is used. If you have special project considerations or a question, ask your local welding supply for advice on choosing the best brazing rod for the job. One advantage of brazing is that dissimilar metals can be joined, which ordinarily can't be done with fusion welding. Also, where two pieces of different thickness are being joined, like a thin pipe to a thick flange, brazing is preferred. In ordinary gas welding, you have to be very good with the torch to play the right amount of heat onto the thicker piece while not burning up the thin piece. With brazing, neither piece needs to be heated beyond the melting point of the brazing rod.

After braze-welding a seam, you will see that the flux remains as a kind of crust on the parent metal. It should be removed with a wire brush, and can be softened with water first, which makes the wire-brushing work easier.

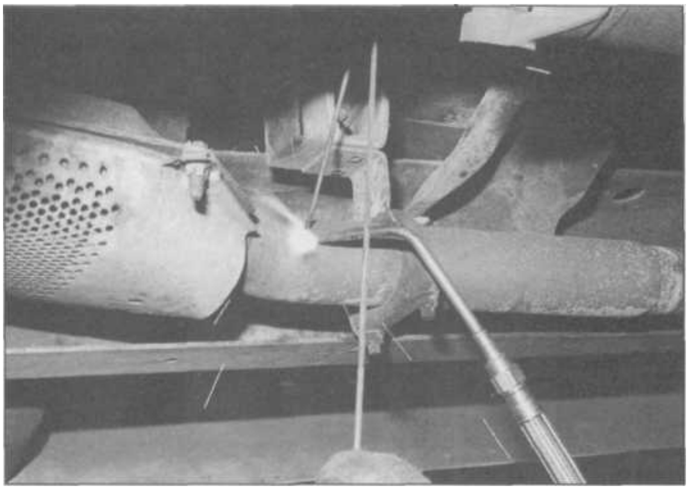

3.38 The versatility of gas welding equipment has always found favor in the exhaust system business. You must be careful not to cut or weld near fuel tanks, fuel lines, electrical components or brake lines, but the equipment is highly portable, and can cut as well as weld. Long welding rod can be bent to curve around exhaust pipes to weld in tight quarters.

Oxy-acetylene cutting

The other side of the dual-purpose nature of gas equipment is the ability to *cut* as well as join metals. The gas torch can make short work of cutting through even large or

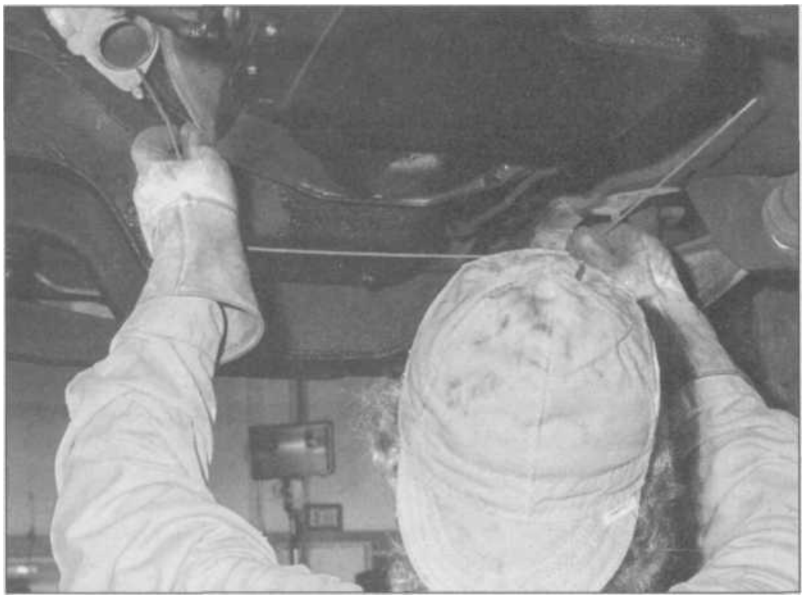

3.39 Muffler installers also find uses for welding rod to simulate the length and required bends for a pipe to be installed. Bent wire is taken over to the hydraulic tube bender where shape is duplicated in pipe.

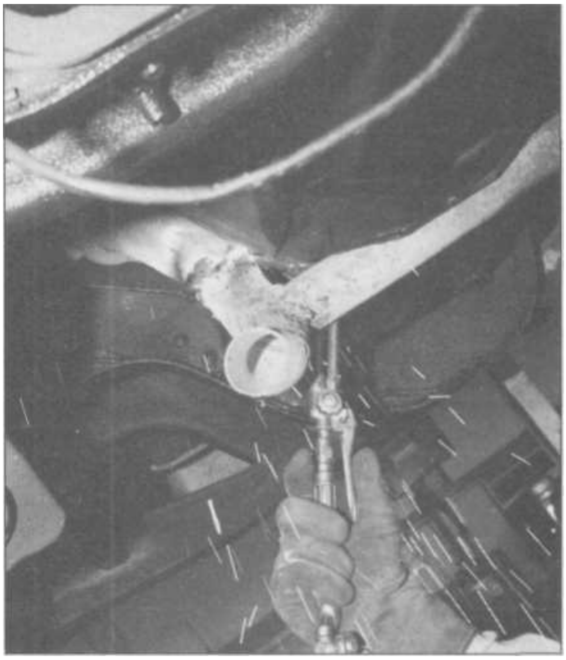

3.40 Sometimes, the oxy-acetylene cutting torch the fastest, easiest way to remove the old exhaust system.

thick pieces of steel. For sheet metal work you will probably continue to use shears and other hand tools to cut light-gauge material, but being faced with cutting through a half-inch steel plate with a hacksaw will easily convince you of the value of a cutting torch.

The only oxy-acetylene cutting component different from your basic welding setup is the torch itself. The cutting torch has one major difference from the welding torch: it has an extra oxygen supply, operated by a lever instead of a valve **(see illustration).** Actually, when you cut steel with a torch, it isn't the flame that is really doing the cutting; it is a strong flow of oxygen that hits a heated part and the oxidation of the metal happens so fast the affected area disintegrates into flying sparks. The basic principle of oxy-acetylene cutting is that the torch tip has several small holes surrounding a central larger hole. The smaller outside holes are where the basic oxy-acetylene flame comes out - actually several small flames. These flames are really used to preheat the metal to be cut. The weldor adjusts the gas and oxygen flow on the torch for the proper flames from the pre-heat holes and plays the torch on the end of the piece to be cut. When the end is heated to a cherry red color, the weldor presses down on the cutting lever, which activates a strong flow of pure oxygen through the large central hole in the tip. This is the oxygen that really cuts through the metal.

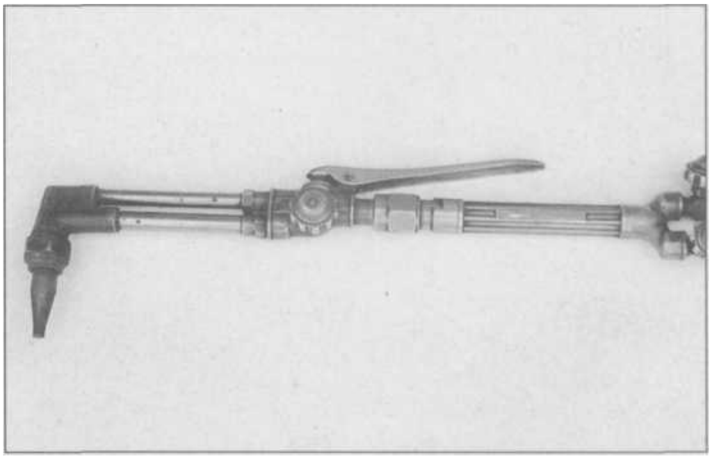

3.41 The cutting torch is adjusted for a good flame, then checked with the extra-oxygen supplied when the cutting lever is depressed. The flame should remain close to neutral. The excess oxygen is what is used to push the molten slag from the kerf as you cut.

Because oxy-acetylene cutting involves a huge shower of sparks from the underside of the metal you're cutting, you must be prepared before cutting. Clear the area around your welding table, have the proper welding attire on (including leather welding gloves), and make sure the area under the test piece is clear of any obstructions. It's best to support your test piece on some firebrick, leaving the area under the cut line open. If there were any restriction to the flow of sparks under your test piece, molten shrapnel could blow back into the weldor's face.

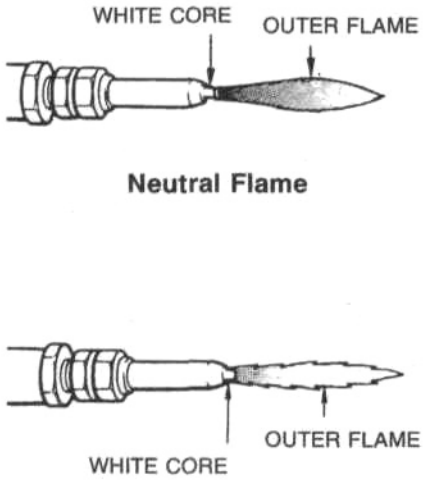

3.42 As in welding, the flame of your cutting torch head should be adjusted to a neutral flame before cutting.

Mark your cut line using a sharpened piece of soapstone and a straightedge. After marking the cut line, find a piece of tubing or angle-iron that you can clamp down onto the workpiece to act as a guide for the torch head **(see illustration).**

Before you begin cutting with your oxy-acetylene torch, you should reset the regulators. Adjust the acetylene gas to 2-3 psi, and the oxygen to 15-20 psi, since a much larger percentage of oxygen will be used in the process. Select a gas pressure and tip size that correlates to the thickness of the material you are cutting. Your setup's instructions will tell you the recommended sizes for that brand of equipment. Light your torch as you have learned to do in gas welding practice and watch for a neutral flame from the preheat holes in the tip. Watching the flame pattern, try cutting in the extra oxygen flow from the cutting handle. It shouldn't affect the quality of the preheat flames. Place the torch flame at the edge of the material you want to cut and hold it there a short distance from the material (1/16-inch or 1/8-inch away). When the metal is cherry red (about 1500° F), angle the torch tip back about 10 degrees (with the tip aiming in the direction of the cut) and cut in the oxygen. You'll find that oxy-acetylene cutting is much easier to learn than the welding process was; the main thing to learn here is the proper travel speed for the tip size, gas pressure and metal thickness you are working with.

When you cut through metal with the torch, the quality of the cut edges will tell you much about the process. At the bottom edges of the cut, slag may collect, leaving a rough edge that must be cleaned up with a grinder later. Also, remember when you are cutting that the torch takes out material as it goes; the part being removed is called the *kerf.* When you draw soapstone lines to cut out shapes, remember to cut *outside* your lines, or the part will wind up being too small. After your have experimented a little with the torch, you'll have an idea of how big the kerf will be with certain tips and certain materials.

If you try to advance the torch too fast, the bottom of the kerf may be irregular or not cut through all the way. When you look at the edge of the cut metal, you'll see vertical lines in the edge of the cut, in a good cut these lines are straight and smooth, in a rough cut the lines are almost like teeth on a saw blade. If you advance the torch too fast, the cutting lines will look curved back toward the direction you started the cut from, instead of straight up and down **(see illustration).** If you move too slowly with the torch, the top of the kerf will be rounded off, and the bottom edge may accumulate too much slag. If you have someone else in the shop with you, have them observe your cutting from a safe distance. When you are making good cuts, the shower of

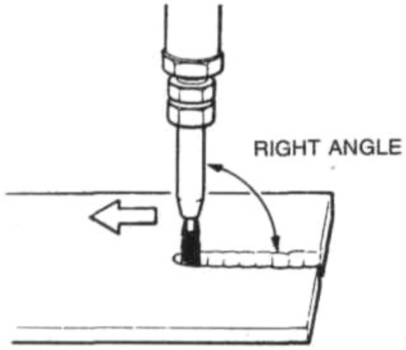

Cutting Thick Panel

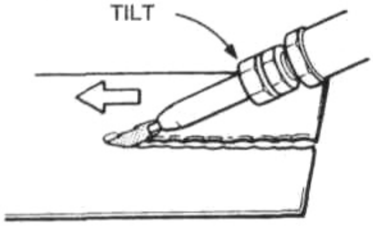

Cutting Thin Panel

3.43 The cutting torch should be held more nearly vertical to the work when cutting heavier metals, and angled back when cutting thinner material like exhaust pipe or sheet metal. Use scrap steel tubing or angle as a brace to guide the torch for steadier cutting.

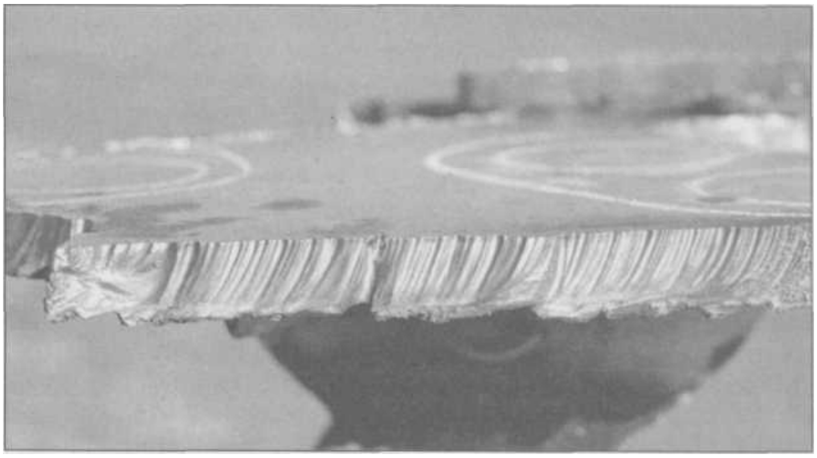

3.44 Examine the cut edge of plates after you have used your cutting torch. The proper heat and travel speed is when there is the least amount of slag along the bottom of the cut. This hand-held cut was moved a little too fast, as evidenced by the angled cut lines.

sparks should come straight down to the floor. Your observer can tell you if the spark flow is going back toward the beginning of the cut, which indicates that the metal isn't being cut all the way until after the torch has moved on, the result is that the uncut area at the bottom directs the sparks back toward the beginning of the cut. Some of the sparks and blobs could also get blown back into your face or clothing, so learn to develop the right cutting speed for clean cuts.

You will find that the steadier you are with the cutting torch, the better the cuts will be, and half the problem of cutting is finding a comfortable position for your hands. Since only one hand is required to operate a cutting torch, you can use your other hand (gloved) to steady the torch and help advance it correctly along your cut line. Keeping the torch tip the same right distance from the work is also important, so find a piece of angle-iron or tubing in your shop sized such that, when you rest your torch neck on it, it puts the tip about the right distance away from a piece of steel the angle or tube is lying on **(see illustration).** Mark and save this piece as your torch rest and it will be invaluable in making straight, clean cuts.

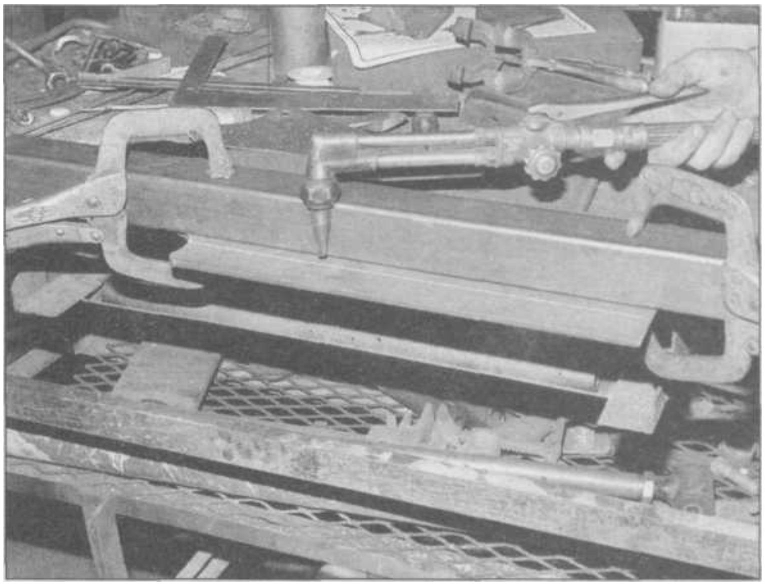

3.45 Clamp a piece of metal to your work that spaces the cutting torch head just the right distance to make your cut-line. The steadier your torch is held, the cleaner the cut. Travel at a very steady speed.

There are many methods for cutting steel for projects, like saws and abrasive discs, but the versatility of the cutting torch is that you can cut out a piece of one-inch plate in the shape of an elephant if you want to. Irregular shapes cut out just as easily as straight lines, as far as the torch part goes, though you will have to develop a rhythm of smooth hand movements with the torch to follow your soapstone lines accurately. As with any aspect of welding, this will take practice. You can also stack several pieces of plate together. Either clamp them together or tack-weld the edges, and when you cut (sizing the tip for the combined thickness) you can cut out as many as a dozen identical pieces. To keep them truly identical if necessary, these parts can later be tack-welded, clamped or bolted together for the final sanding or grinding of the edges, ensuring that they all match in dimensions.

Besides making straight cuts, which are easiest, you may find that some projects require certain shapes to be cut out more than once. Here is where various cutting aids can be used. If you can make the torch tip follow a pattern of some kind, you can cut out the same shape repeatedly. Perhaps you need four round discs of steel six inches in diameter. You can make a steel or wooden pattern that has a circle in it bigger than six inches. How much bigger than the six-inch measurement depends on the size of the kerf with the tip you are using. If you know that you need an extra 3/8-inch to allow for the kerf, make a template with a hole 3/8-inch bigger all around. When you use this as a cutting guide/rest for the torch, you can cut out as many circles as you want and they will all be round and the same size. You clamp the template over your workpiece and move the torch around your cut, always in contact with you template. The smoother your template and the steadier your hand movement, the nicer the cut will be. Watching a pro weldor some time will show you how perfect you can get with what seems like a relatively-crude method of making a part.

In fabrication shops, special equipment is used to make repeated cutout shapes where lots of parts are needed. One of the basic machines you would see in most street rod or race-car shops is a *"flame-cutter"* **(see illustration).** This is basically a cutting table on which is

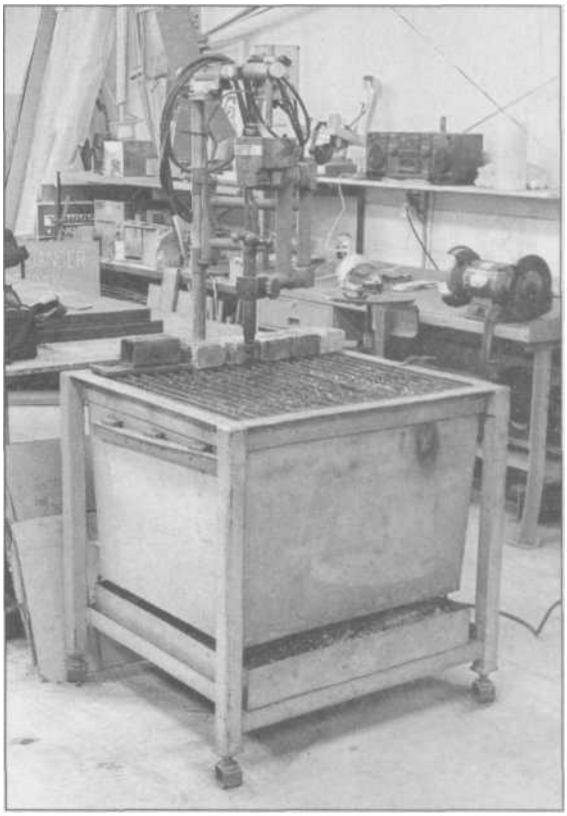

3.46 A tool you will see in most race car, hot rod and fabrication shops is a mechanized cutting torch, often called a flamecutter. It uses a motor-driven torch to travel around a template and cut out exact shapes. Templates can be temporary ones cut from heavy cardboard, to steel templates for frequent use.

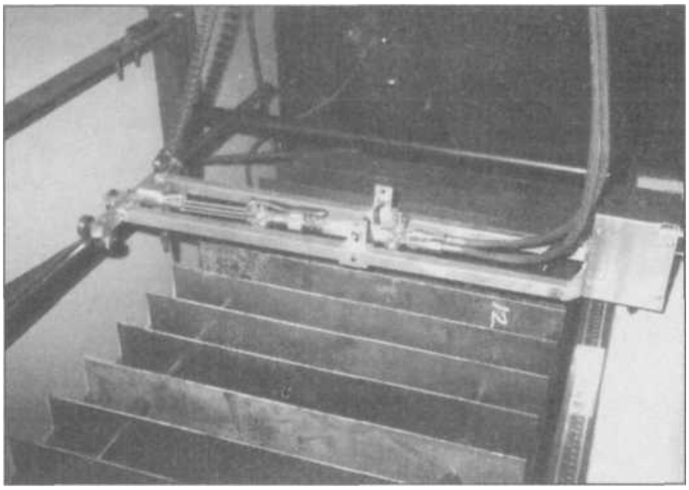

3.47 This mechanized cutting setup from Daytona MIG uses a standard cutting torch, laid down and clamped in place.

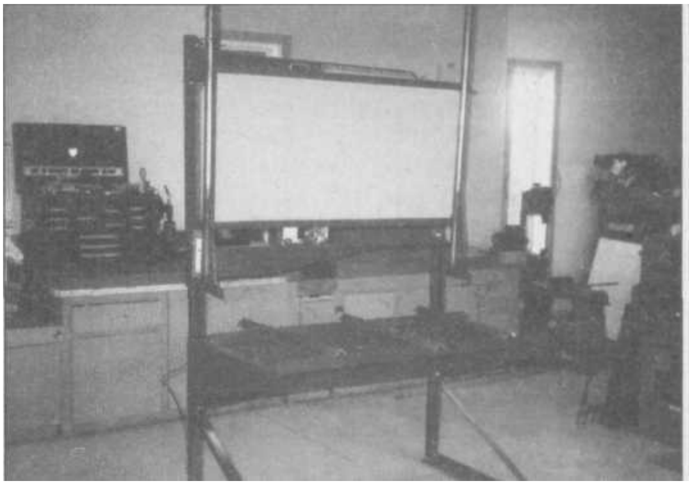

3.48 The complete Daytona MIG shape-cutting table is shown here with the erasable porcelain layout table folded up to take up less space when not being used. This table can be used with a gas torch or plasma cutter.

mounted a framework that holds a special torch, with a straight cutting end instead of a 90-degree end like most hand torches **(see illustration).** There is a series of arms that allow the torch to stay perfectly vertical and yet move to follow any shape on the table, and at the top is a small electric motor. There is a place at the top to mount a template, usually made of heavy-gauge sheet metal, and a magnetized, knurled tip on top of the torch assembly that follows the template exactly. The motor drives the knurled tip around the template at a steady speed (adjustable for the thickness of the material being cut), while below, the torch tip cuts out exactly the shape of the template. Removing the human element from the torch movement means that a flamecutter can make smoother, more repeatable parts. In most shops with a flamecutter, there is usually a large wall behind it on which hang a wide variety of templates for frequently used shapes, such as suspension brackets **(see illustration).** In some shops, the flamecutting setup has *a pantograph* arrangement, which is a series of arms that make the torch cut out a shape dictated by the movement of a stylus around a template on a table next to the flamecutter. Some sophisticated machines have an electronic stylus that will follow the lines of a blueprint or even the lead of a pencil line drawn on paper.

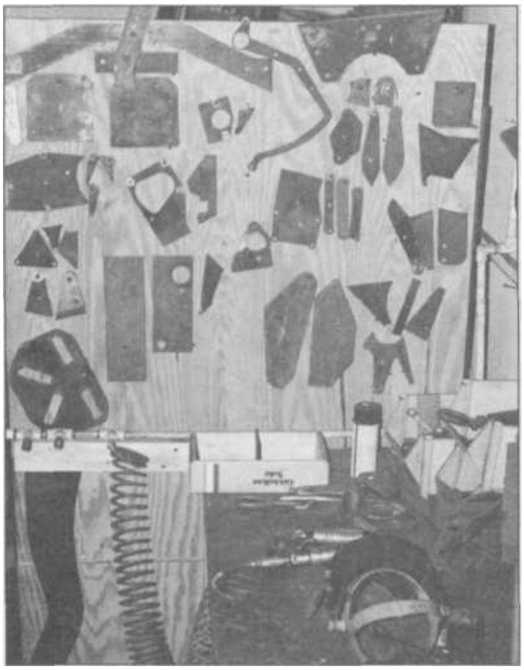

3.49 Wherever you find a flamecutter, chances are you'll find a wall of metal templates nearby, used to easily and precisely duplicate commonly used shapes such as suspension mounting brackets. At lower left is a shape used to make a head for an engine stand.

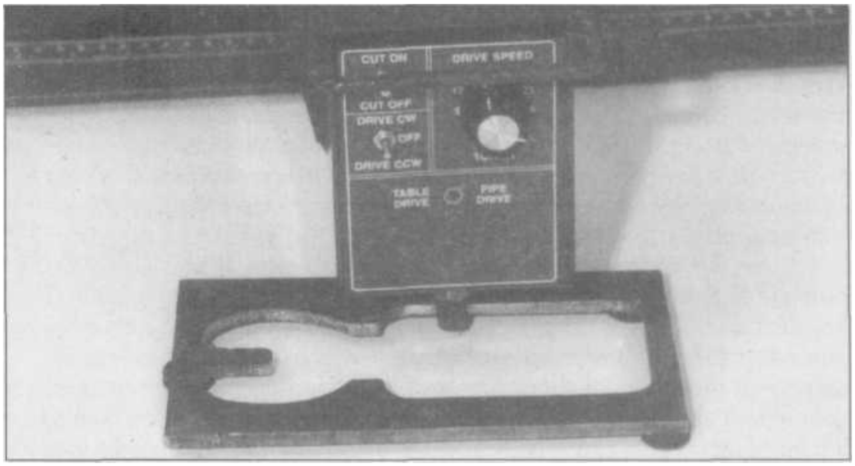

3.50 Most mechanized cutters have a drive head with a magnetized tip for following metal templates, inside or outside a pattern. The cutting speed is infinitely variable to suit the material.

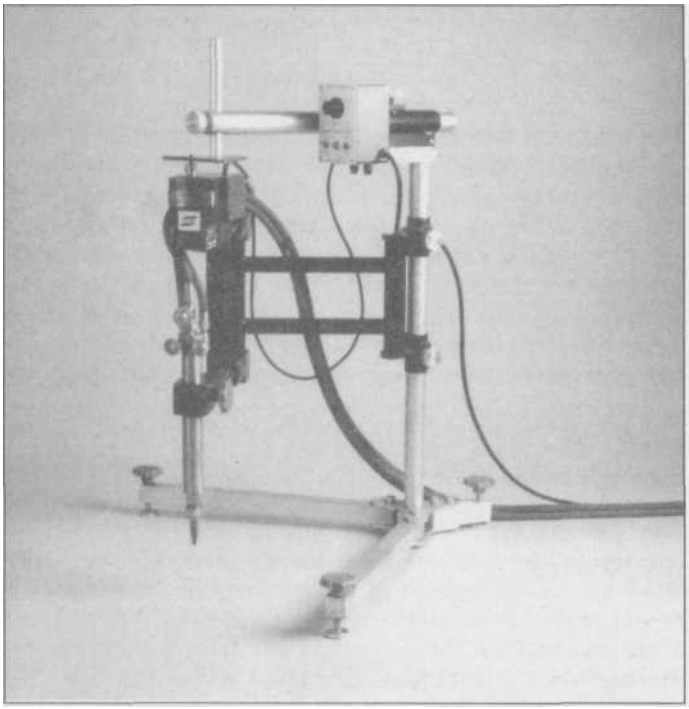

3.51 The ESAB Porta-Graph can be moved anywhere you have a source of electricity, and is useful for cutting shapes out of large steel plates that won't fit under a stationary cutting machine.

3.52 This flamecutter with swing arms can be mounted over a plate, or bolted to a table.

If you decide to cut out shapes with your oxy-acetylene torch, you will probably need to build a cutting table **(see illustration).** On a straight cut, the end of the part can be laid over the edge of your bench; the work is supported and the cutoff piece just falls to the floor. But cutting out shapes larger than a few inches means you need a different surface to work on, or you'll soon be cutting into your bench! Look at the photos of some of the cutting tables pictured and you'll see how to construct one. A framework is made that holds long strips of 1/4-inch steel straps in an edge-up position, spaced an inch or two apart. The framework should be placed over some kind of fireproof container, like an old drum (sandblasted clean of any flammable residues first). This will capture the flying sparks and hot scrap pieces and keep the shop neater and safer. Another safety consideration to watch for is protection of your feet. Make sure when you are cutting out a heavy part that it isn't going to fall right on your feet. Either have a metal deflector under your cutting area, or keep your feet well away from where the very hot part is going to fall.

There will be times when you cannot start a cut at the outside edge of your workpiece, and you will have to cut from *inside* a pattern. This requires a little different technique to get the cut started. Adjust the torch to a neutral flame with the cutting lever depressed. With the lever *off,* start heating a spot on your workpiece along the cut-line. When the molten puddle starts to appear, pull the tip back about a half-inch (because this is a closed spot, there is a chance of molten metal and spark being driven back into the tip if it is too close) and slowly cut in the cutting lever. When the spot becomes a complete *hole,* you can proceed around the rest of your pattern or line.

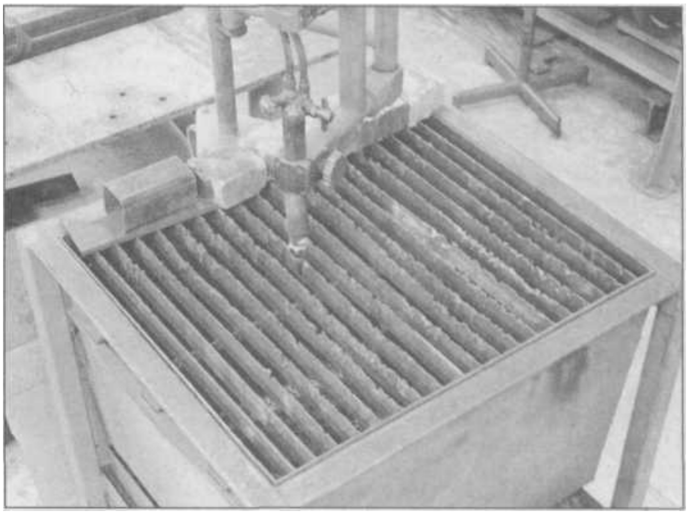

3.53 Under a mechanized cutter, you need a grate like this to support the work while it is being cut. After some usage, the 1/4-inch by two-inch slats will be fairly cut up, but can be pulled out and swapped end for end, top for bottom or simple replaced.

Some parts cut with a torch need to have a beveled edge if they are going to be joined with another part by welding, such as when joining sections of heavy pipe. Cutting a beveled edge on a round pipe by hand is challenging, and in the pipe-fitting industry they often use a machine that clamps around the pipe. The fixture holds a cutting torch, and a motor drives the assembly around the pipe smoothly and evenly, and is adjustable for speed and the angle of the bevel. There's no place for such equipment anywhere but in the pipe business, but it's interesting to see the ways in which the human element of hand control of the torch can be eliminated with special equipment. When you do make beveled cuts in thicker metals with your torch, you should remember to size the torch tip to the measurement *across the bevel,* which is thicker than the straight thickness of the metal.

3.54 The rosebud tip will be useful for heating metal parts to be bent, freeing press-fitted parts, and for preheating castings before welding, such as here where a cast-aluminum mailbox post is preheated before TIG welding.

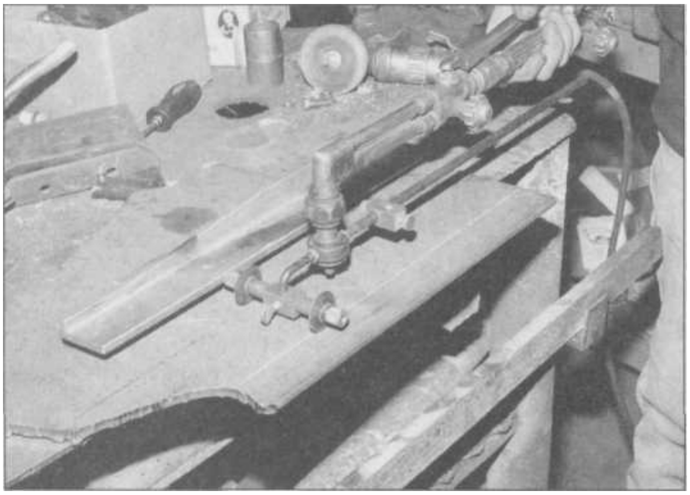

3.55 A great aid to cutting straight lines manually is this "compass" guide. It holds the torch tip firmly, but in a swivel arrangement that allows it to turn as you move along the line, and the torch stays the same distance from the work and much steadier. Here it is being demonstrated against another great tool, a magnetic steel guide that sticks strongly to the work and saves the hassles of setting up a bar or tube with some clamps.

Heating with oxy-acetylene

You will be surprised how easily a piece of heavy steel can be bent into a shape when it is properly heated with a torch. Even a thick piece of steel, heated to a cherry red color, can be bent like a sheet of soft copper. Your bench vise will come in very handy for projects like this. The old-time blacksmiths hand-shaped everything they made the same way, except that they had to heat the part in a coal or wood fire, while you can do the same thing in seconds with an oxy-acetylene torch.

Depending on the thickness of the metal and the area to be heated and bent, you can use a welding torch or cutting torch to heat up the workpiece. It shouldn't take too long to get the area cherry red; if it does, use a larger tip on the torch. There is another type of tip that can be used called a *rosebud,* that is expressly designed for heating. If your oxy-acetylene set was a deluxe model, such a tip may have been included, or you can easily get one at your welding supply center.

The rosebud tip has a much larger end that flows a huge amount of gas. You will have to crank your tank regulators up to perhaps 25 psi on oxygen and 10 psi for acetylene to feed this device (never exceed 15 psi on the acetylene regulator). The big, wide flame of the rosebud tip is ideal for spreading heat, and lots of it, more evenly around a part than with the tighter flame of a welding or cutting torch tip, although those can be used to heat small areas, especially if you are just freeing frozen, rusted fasteners.

The most common usage of the rosebud tip is for heating components that are press-fitted together. Heat expands metal, so logically, if you can heat up the exterior part while not heating the interior part, the heated part will expand and can be removed (with tongs). This is most commonly done in automotive or industrial situations where there is a gear pressed on a shaft, or an axle bearing must be removed. In some situations, however, you cannot simply blast the area with uncontrolled heat, or damage may occur to one of the components. You need to apply only as much heat as it will take to separate the parts, no more. Your welding supply shop should have a variety of paint-marking sticks or crayons that are designed to melt or change color at specific temperatures. One brand, called Thermomelt, is available in 87 different temperature ratings, from 100° F to 2200° F, and in a liquid, stick or pellet form. You use this product on the part you need to protect, and only heat the area until the product melts or changes color, then you work on separating the components without adding any unnecessary heat. These products are also useful when tempering or retempering tools that have been reground, which is a lengthy subject unto itself.

Deciding which type of welding equipment to buy is always a tough decision, but one that is made easiest by carefully defining your needs before buying any equipment. When you have read all the chapters in this book, you should have a good idea what the benefits and drawbacks of each system are, and what is best-suited for your work. Any weldor will tell you that the versatility of the gas-welding/cutting setup becomes valuable when you need it. Even when most of your work is done with some sort of electric-welding equipment, you'll find so many handy uses for oxy-acetylene equipment that if your budget can handle it and you have enough work to utilize it, a great idea is to have the gas equipment and some form of electric welder, too.

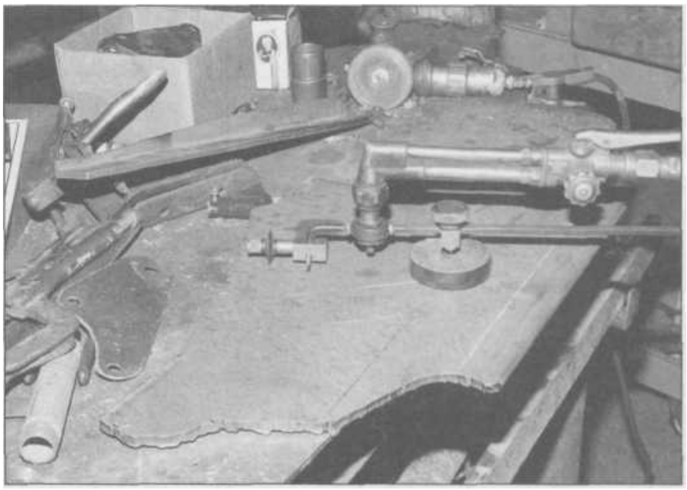

3.56 The same tool can be used, with different attachments, to cut a perfect circle of any size from a few inches to two feet in diameter. Removing as much of the human element of hand control makes for better parts that require less grinding to finish.

Notes

4 Arc welding

One of the oldest forms of welding, arc or "stick" welding offers versatility, strength and the ability to handle big projects and thick materials. Also, introductory equipment can be purchased inexpensively. There are some rather expensive arc-welding machines, but the most common AC-current machines will do fine for the home/shop user doing standard projects and repairs.

If you have thoroughly read Chapter 2, you should have a good understanding of the various kinds of welding equipment and what each type is capable of. If you think the arc welder serves your needs, then this chapter will give you a basic introduction to the equipment and procedures involved. When shopping for a machine, you may see them referred to as SMAW welders, which is the technical description of the process and stands for Shielded Metal Arc Welding.

Basic-yet-rugged arc machines are available today in welding equipment stores, major nationwide stores such as Sears and Montgomery Ward (and their mail-order catalogs) and can sometimes be found at a good savings in large lumber/home supply centers that carry a lot of power tools. Most starter units are sold with a complete setup, including welding gloves, helmet, chipping hammer, sample electrodes, and a basic instruction book. Other than the consumable welding rods, there will be little else you will ever have to buy to continue arc-welding.

4.1 Arc welding will always be with us, as the best welding method for joining heavy plates for commercial, farm and industrial work. The weldor requires considerable protection from the spatter and radiation, usually wearing heavy leather protective gear.

4.2 For home and farm use, the Lincoln 225 is the classic AC buzz-box welder, having been in production relatively unchanged for decades. Most home/shop units are AC only, and require some rewiring to accommodate the required 220V input.

4.3 Another popular AC 220V welding machine for non-industrial use is this Miller Thunderbolt 225. The crank handle on top changes the output amperage, which is read on a scale at the top left of the front panel.

You can find arc machines priced from less than $200 to professional units costing several thousands, depending on the features, but most home/shop projects and farm repairs can be performed with machines at the lower-priced end of the spectrum. There are really no moving parts involved in the machinery, and most of the name-brand equipment is rugged enough to last for many years of service. The larger machines offer certain advantages for application in professional shop use only, such as higher amperage, AC/DC switching and higher duty-cycles.

The principle of arc-welding is to attach a ground cable to the workpiece, set the machine for the correct amperage based on the thickness of the material, fit a consumable electrode (welding rod) in the electrode holder, and with your helmet down strike the rod against the work to start a flow of current, the arc, that produces intense heat and light and welds your seam together. Welding thicker materials requires more heat in the form of higher amperage from the machine. The basic arc machines of interest to you generally have an amperage range of 40-225 amps. The arc process is best suited for thicker materials, but you will probably *never* use your machine at the higher settings. Even when welding on materials one-inch thick (which you will most likely never encounter in your home shop), or when repairing large castings, the seam isn't completed all in one big bead, it requires several passes on overlapping beads to totally join the parts.

At the same time, you will probably never use your AC welder at the lowest settings. For materials thinner than 1/8-inch, there are other welding processes more suitable, such as gas, heli-arc or MIG-welding. Most of the basic welding performed with AC arc-welding machines is done at 90-125 amps, on materials from 1/8-inch to 3/8-inch and sometimes 1/2-inch.

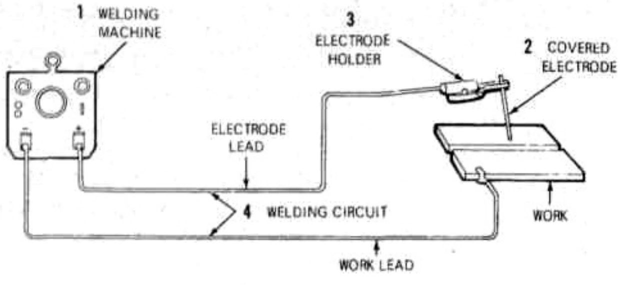

4.4 The layout of the arc-welding setup is simplicity itself. The rod (also called a stick, stinger or electrode) is both the source of the arc and the shielding gas, produced as the rod melts its flux coating.

Comparing duty cycles

What you will look for in an arc machine, therefore, is not the highest amperage it offers, but the *duty-cycle* at the 90-125-amp ranges you will use most often. We discussed the duty-cycle in chapter 2, but it bears repeating in brief. All electric welding machines have a duty-cycle rating, which refers to how long they can weld at a specific output without overheating. The duty cycle is described as a percentage. If a machine has a 100% duty cycle, it means that it can be operated virtually all day, except to stop when you change electrodes. The percentage is actually based on a ten-minute period as a test, meaning that if the duty-cycle was 50%, the machine could be used for five minutes out of ten. You could weld nonstop for five minutes and then "rest" the machine for five before starting again.

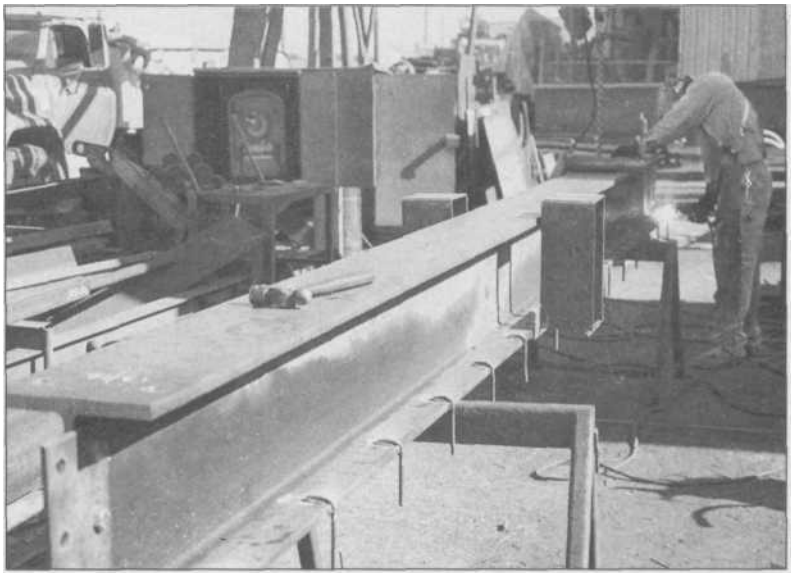

4.5 Virtually any building over single-family-home size requires structural steel in its construction, and that requires arc welding. Here a large beam is being prepared for earthquake retrofit in a public school.

The duty-cycle rating confuses most first-time buyers, because different companies may take their published ratings from different amperage settings. The higher the amperage you weld at (for thicker material), the less the duty cycle will be. Obviously, the machine will overheat quicker at the higher amperages. On every machine's range of amperages, there is probably a point where the duty cycle is 100%. A specific machine may have a 50% duty cycle at its highest setting, yet have a 100% rating at the mid-range settings you will use most. It's important when shopping for a welder to find out what the ratings are for the high range *and* the mid-range. In the larger professional machines designed for welding shops and production-line work, the duty cycle must be close to 100% for any situation, and thus these machines have to be built with much more expensive components and reserve capacity.

Some of the duty-cycle rating comparisons may be academic for the average home/shop user. If the only long seams you weld are on thin materials at lower heat settings, most machines will be fine for your purposes, and there is no reason to spend many times more for the machine to get a higher duty rating. In a typical home/farm project, say where you are welding together steel tubing to make a utility trailer, the *setup* of the work takes half or more of your time anyway. Each welded joint between pieces of tubing may only take one or two minutes of actual welding, and then it may take you four minutes or more longer to set up clamps on the next joint or flip the work over before you're ready to make another weld. The point is that, in this example, a welding machine with a 40% duty-cycle at the amperage you were using would be perfectly adequate for the job. You wouldn't find yourself being slowed down on the job because you had to weld and then wait for the machine to "catch up." However, there is something to be said for having some duty-cycle "cushion," and if you were considering two machines in the same price range and featured similarly, you would take the one with a higher duty-cycle at the heat you would most often use it. Generally, the machine with the higher duty-cycle at a mid-point amperage will also have higher ratings at other settings.

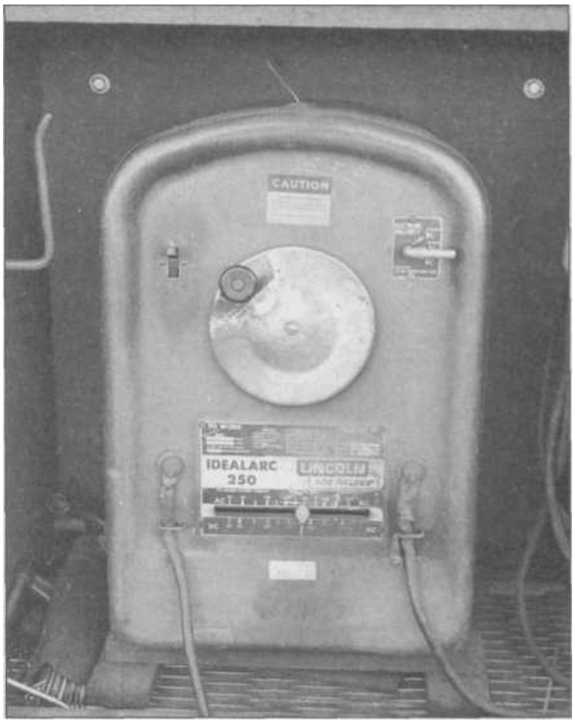

4.6 In industrial use, the arc welding is generally all DC, and you will see these venerable Lincoln welders performing their duties day in and day out for years. Most industrial machines have a 100% duty-cycle for production work. This one is housed outside in a steelyard, enclosed in a weatherproof box.

4.7 In larger welding shops, a more sophisticated power source is used, such as this 300-amp **Miller DC machine, which has a few more "bells and whistles" than the home-type units.**

AC, DC or both?

Most of the basic machines in stick welding are AC-powered. For those readers who may have taken electricity for granted since high-school physics classes, AC refers to *alternating current,* which is what we have in our houses, businesses, and power lines. The DC designation refers to *direct current,* for which the most common daily usage is in the 12-volt systems in our cars. When electricity was first being used in the 1890s, Thomas Edison, for all his genius in other scientific regards, was insistent on DC current being the standard for home lighting and any other usage. Unfortunately, DC current isn't practical to send any long distance through wires, and with DC every neighborhood would have to have their own power plant. The brilliant Nikola Tesla (the true inventor of the radio and many other breakthroughs) developed the alternating-current system and licensed it to Westinghouse, which became a giant corporation when AC was accepted as the world standard (AC could be sent hundreds of miles along power lines).

How AC "alternates" is by traveling in a wave, alternating in polarity up and down in a repeating cycle. Most electricity in the US alternates at a rate of 60 cycles-per-second. What all of this means to arc-welding is that the less-expensive arc-welding machines are AC-only. The AC welder is very good at producing less welding spatter, at welding heavy plates with large electrodes, requires less electricity to run and usually has less maintenance expense than bigger machines. They take the standard line current, which is high voltage but low in amperage and reduce it through a transformer to a low-voltage/high amperage current for welding. The only drawback to AC arc-welding is that the constant switching of polarity can make for tiny inconsistencies in the weld bead, imperfections you and I would never notice, but something that could be critical in an oil field pipe line, high-rise-building framework or a nuclear reactor. For this reason, most professional arc welders are DC, which produces a much smoother weld, a more stable arc and there is a wider selection of special electrodes (rods) for the DC-type professional arc welder. The AC-only machines are generally used strictly for joining ferrous metals, but the DC machines can also be used for stainless-steel and for hard-surfacing industrial parts. In addition to the two current-specific types of arc welders, there are also combination AC/DC machines, which usually have a rectifier added to a basic AC machine.

Another factor separating the larger professional machines in terms of flexibility is the choice of polarity, and the option of a TIG torch setup. In the DC mode of operation, the operator of some machines can choose between negative or positive polarity, depending on the type of metal he is welding and the rod he is using. Most DC welding is done with reverse polarity, meaning that the rod is positive, and the work clamp is negative. This method keeps the rod very hot and makes for smooth welds and improved out-of-position work (anything other than flat on your welding table). This is not of concern in using AC machines because the nature of the current is switching polarity 60 times a second anyway. Many of the professional machines are versatile power supplies that can perform arc welding, MIG-welding (with the addition of a motorized wire-feeding attachment), or TIG-welding with an optional torch and foot control **(see illustrations).**

This is the kind of equipment you will find in professional welding/fabricating shops, but most shops today don't perform much arc welding, using their power supplies mostly for TIG welding, and using a separate machine for most wire-feed requirements. The arc-welding arena in professional welding is usually in industrial jobs or construction, pipelines, etc.

Not to confuse you any further about electricity, but the larger professional machines are often available for several types of input. The most basic machines, home or shop, require 220V, and industrial units may be set up for 440V or more, and there are different *phases* as well. The type of current we have in most homes is called single-phase. The larger professional arc welders are made in three-phase configuration, which simply means that there are three identical inputs spaced 120 electrical degrees apart. The waves in these inputs overlap, so that voltage never falls completely to zero, making for smoother welds. Many industrial shops have big motors on lathes, mills and other machinery that run on three-phase power, which is smoother and cheaper to operate on large equipment. We will not find three-phase power at home, and some expensive equipment is required to set up a building for industrial three-phase power.

Miller
Maxstar
AMPERAGE ADJUSTMENT
POWER
AMPERAGE A
V VOLTS
Miller
Maxstar 175
CC-DC INVERTER WELDING POWER SOURCE
RATED OUTPUT

4.8 Many shops utilize a "power source" machine which not only powers arc welding work, but can be set up to take a TIG torch, and some can use an optional wire-drive to perform MIG welding.

Rewiring for an arc welder

Except for a few small, household-current machines capable of 100 amps or less, even the most basic AC arc welders require 220V input. Depending on the present wiring of your house, this may require an additional expense in rewiring to accommodate the welder, a factor to take into consideration when choosing the right system for you. Some homes already have a 220V outlet for hooking up an electric clothes dryer or stove (the most common household appliances to run on this voltage), but this outlet may not be where you need to weld. If your washer/dryer setup is in the garage, you're ahead of the game, and rewiring may not be necessary. Likewise, if you plan to do most of your welding in the kitchen, then your case is simplified.

4.9 Contractors often use an engine-driven arc welder mounted on a truck bed or a small trailer. Besides powering the welder, the gas engine also has a large generator to run power tools in a remote work site.

If you have to run a new circuit in your house to put 220V where you need it for the arc welder, put it on it's own separate circuit with a 30-amp circuit-breaker. The kind of current you may be drawing when welding thick material may blow a standard 15 or 20-amp household breaker. Call a few electricians before you buy an arc welder. Tell them what size electrical box you have (how many amps), how many open spaces there are for new circuits, and how far a new 220V circuit would have to be run to get an outlet in your proposed welding area. They should be able to give you a rough estimate of the rewiring costs. It may cost as much or more than the welder itself, and may ultimately influence

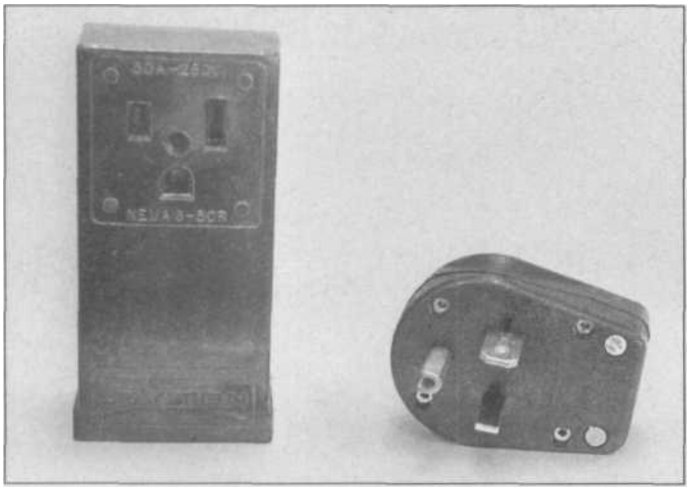

4.10 Even if you do have a 220V circuit in your shop for a clothes dryer, you'll find the outlet prongs are different on a 220V welder's plug. You either have to rewire your outlet box to accept a welding plug, put a household 220V plug on the welder's power cable, or use an adapter.

what kind of welding machine you finally purchase.

Even if you do have an existing 220V outlet you can use, you'll be rudely surprised when you find the welder can't be plugged in, at least not directly. The arrangement of prongs on the welder's power cable is slightly different than the layout of the typical 220V home appliance plug. You can make or purchase an adapter, or have the electrician install an outlet box in your welding area that matches the kind of plug on your arc welder (see **illustration).** Before plugging a new stick welder into a household 220V outlet with an adapter, you should have the system checked to be sure the wire gauge and circuit-breaker are up to the task of handling a welding machine.

The arc process

All electrical-welding processes use the flow of electricity to create heat. The power flows from the torch or electrode to the work, which is grounded to the source at the machine. In arc welding, the consumable electrode or rod makes the connection that creates the arc to the piece being welded. The welding rod is a metal rod coated with a hard flux material. As the arc is created when the tip of the metal comes to the workpiece, the heat generated at the bead is 6000° F or more, which melts both the parent metal and the filler rod, while simultaneously vaporizing the flux coating to create a gas shield around the bead, protecting the solidifying weld from contamination by gasses in the air **(see illustrations).** The flux actually re-solidifies on top of the bead as a hard coating of flux and slag, and when you look at a completed bead, you'll see a dome of ceramic-like material over the weld. At this point, the weld doesn't look very impressive, but when you remove the slag with a chipping hammer, a beautiful, clean bead is revealed (see **illustrations).** Depending on the type of rod and amperage used, there may also be some spatter (tiny beads of metal) stuck alongside the bead. Most of these beads will come off with stiff application of a wire brush, and more stubborn ones can be removed with the

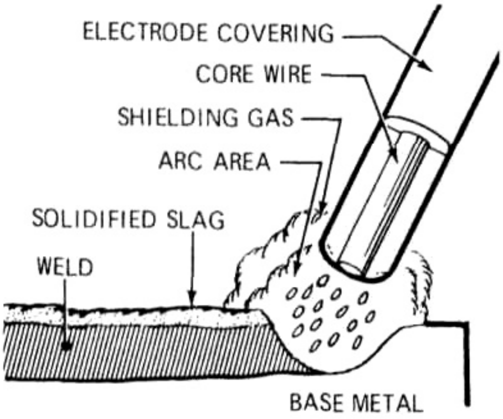

4.11 Arc-welding electrodes or rods are a metal wire, covered with a flux coating. The metal wire must make initial contact with the work to initiate the arc, then it is pulled back to a proper rod-to-work distance, the vaporizing flux makes a shielding gas, but also deposits a layer of slag over the weld as it forms, to further protect the weld from contamination.

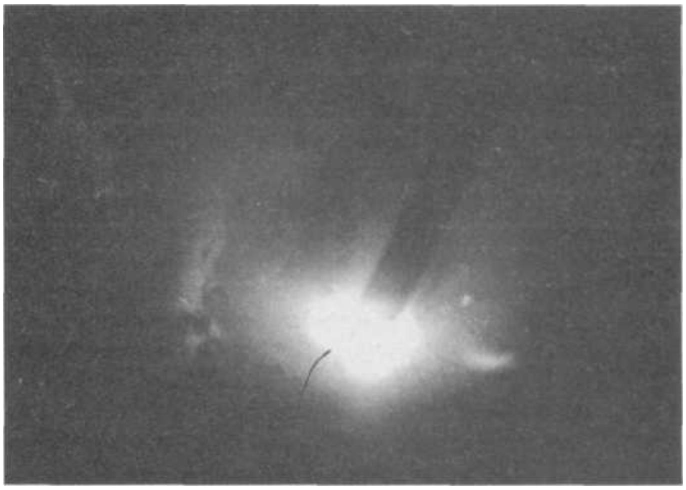

4.12 This is what you will see through your helmet and welding lens. The arc is intensely bright, but the lens cuts this down to just a bright glow around the arc itself, here you can just about see where the shielding gas cloud forms around the arc.

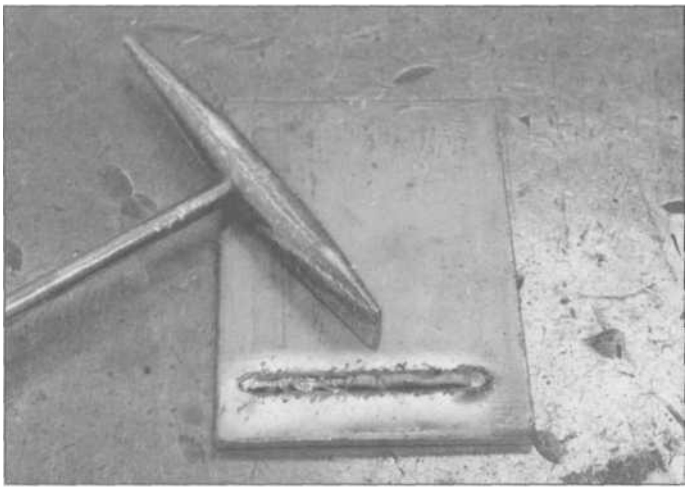

4.13 The slag formed over the weld bead must be removed with a chipping hammer like this, and most spatter and slag residue can be cleaned up with a wire brush after chipping.

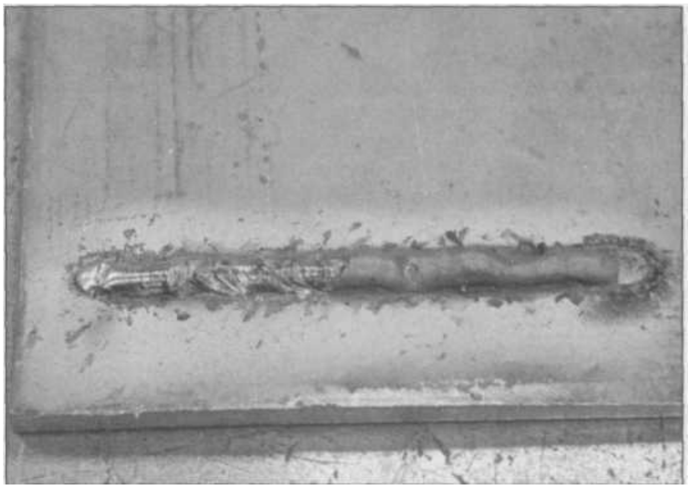

4.14 At right on this bead, you can see the dark gray slag coating covering the weld. The portion at the left has been chipped off to reveal a bright, shiny weld.

chipping hammer or a chisel.

Making your first passes with the AC stick welder will be relatively easy, although really good passes will require considerable practice. The instruction book that comes with your welder will specify the right rods to use for various materials, and the amperage to set for different thicknesses of ferrous metal. The rest of the technique is finding the most comfortable position for your electrode-holder hand, and maintaining the proper arc distance and travel speed to make good joints.

Safety considerations

With every method of welding, safety is of paramount consideration, but each type has precautions that apply to that type of equipment in particular. In all forms of electric welding, including arc welding, high-amperage electrical current is the primary hazard. All of your cables, plugs and leads should be inspected regularly for any signs of defects. Even dirt or paint overspray on connections can cause arcing and poor welds. Water, of course, is a good conductor of electricity, and therefore should be avoided in the work area. Your clothing, equipment and especially the floor must be kept dry to avoid the possibility of electrical shock. Rubber-soled shoes are recommended, but athletic shoes (non-leather) are not. Most experts will tell you not to wear metal jewelry such as watchbands, rings, bracelets, necklaces or belt buckles when welding. If electric-welder power comes into contact with metal articles you are wearing, they can become instantly hot to the point of melting, or can cause electric shock.

Of the electric welding methods, arc welding requires the most protection of your face and body during welding. The intensity of the arc produces strong UV and infrared radiation. Any skin exposed during the welding process can become burned, in severity ranging from mild sunburn to serious burns, with the symptoms not appearing until eight hours after the exposure. Leave the top button unbuttoned on your shift and you'll have a nasty V-shaped burn on your neck after only a short while arc-welding. Likewise, wear fire-resistant, long-sleeved shirts, and keep your sleeves rolled down at all times. Keep these shirts just for welding, and tear off the pockets if they have any, or keep them empty and buttoned. An experienced weldor friend of ours was recently burned painfully when welding overhead with just a shop shirt on — a hot bead of spatter went right into his shirt pocket and burned into his chest. Without the pockets, there's a chance the bead will roll off onto the floor rather than stay in one spot on your shirt. For this

4.15 The arc spatter and radiation are dangerous. Do not allow any bystanders to observe the arc while you are welding. Also, notice how much smoke is around the weldor's helmet here. The fumes are dangerous and you should have an electric fan of some kind to blow it away from you when you are welding in a corner like this.

same reason, your pants should be kept uncuffed, and never tucked into your boots.

If you are going to be doing arc-welding often, we'd recommend you invest in some leather safety clothing, like jackets, vests or pull-on sleeves that go over your regular shirt. Arc-welding is prone to more spattering than other types of welding, and these leather weldor's clothes are highly resistant to arc spatter.

Probably your most sensitive and fragile body parts exposed to welding dangers are your eyes. Even the tiniest bit of spatter in an unprotected eye can have truly long-lasting negative effects. Always wear a full-coverage safety helmet when welding, preferably with a leather flap at the bottom-front that protects your neck area. Especially when welding overhead, like underneath a vehicle, wear a cloth cap backwards (bill to back) to cover your hair and the back of your neck. Your helmet should be equipped with the proper safety lens for the type of welding you are doing, or your eyes could receive overexposure of UV and infrared rays in a very short time. Never observe *anyone else* doing arc-welding unless you are wearing proper eye protection, and make sure that when you are welding that there is no one observing you who could be hurt by watching, particularly children. Watching too much arc will not show immediate effects, but later the affected eyes will be sore, and with a sensation almost like having lots of sand in your eyes. If you do not yet have your own welder, but want to watch someone else work, get your own helmet to observe through. If you do have a welder, you may want to keep a spare helmet around in case someone wants to observe *your* welding prowess.

Your eyes can be permanently damaged by overexposure to arc rays, but they must also be protected when working around most shop equipment, such as grinders, mills, drills and sanders, all equipment that may be involved in your welding project. Keep several pair of good safety glasses around your shop, the kind that have protection all the way around the sides. After arc-welding, you will also want to wear these safety glasses when chipping slag from your welds. The little fragments that break off are like glass. Always keep a very complete first-aid kit accessible in your work area in case of accidents.

A particular hazard with arc welding is the presence of fumes. When the electrode is consumed, the flux is vaporized, creating the shielding gasses that protect the weld from contamination during formation. Depending on the metal being welded, other gasses may be released as the metal is melted. Most welding gasses are colorless, odorless, tasteless and inert, but this is not to imply that they are harmless. Any of the common welding gasses can displace oxygen, and when you are breathing in air that contains less than 18% oxygen, you may experience dizziness, or even lose consciousness. For this reason, arc welding, or any welding process, should be performed only where there is adequate ventilation. In the case of arc welding, there is less chance of the shielding gasses being blown away and causing a bad weld, so if you find yourself welding in one spot too long, or in a confined area, you can use a household fan somewhere in your work area to maintain air circulation.

Beginning arc welding

If you have already read the previous chapter on gas welding, what you will have learned in practicing that mode will help you greatly in learning all other forms of welding, including arc. Stick (arc) welding is easier in some ways than gas-torch welding, and more difficult in others. Practice and more practice will put you on the road to good welding in any mode.

What you will find different at first about arc welding is that you only need one hand. The electrode holder and its electrode or rod is *it,* other than the ground clamp, which should be clamped to your workpiece or your steel welding table. The rod is both the source of filler metal and the shielding gasses, which are generated when the flux covering is vaporized. Have some scrap steel handy and some 1/8-inch rods. There are a great many specialized rods, but one of the most common is an E-6011, which is one of the easier rods to start and maintain an arc with. Although the arc process only involves the one hand-held tool, you may find that the 14-inch-long rods put your hand a much further distance from the weld area than the other welding processes. One end of the rod is bare of any flux coating for about an inch, this is the end you put in the electrode holder.

It's a good idea to practice the arc "setup" *without* turning on the welding machine. Just situate yourself comfortably in relation to the work, and practice holding the electrode 1/4-inch away from your work joint or seam area, following the proposed seam while slightly weaving the electrode tip side to side as you travel along. This will give a feel for what is required in terms of coordination.

Now if you're ready, set the machine for the right amperage for the thickness of your scrap steel, say 1/4-inch steel plates, or perhaps slightly hotter than the instructions recommend to make it easier to learn the starting procedure. One drawback of the arc-welding process is getting the arc *started.* You can't just flip down your helmet or lens and stab the rod against the work.

The starting procedure has been described as similar to striking a match, in which you draw the rod tip across the area you plan to start the bead. At some point in your "scratch" the rod will momentarily contact the work and the arc will start, but the rod must continue moving. Arc welding is essentially a process of creating a short-circuit across the rod and the work, and it can only be started by a momentary contact of the two. Once started, this short-circuit heats the air around the weld and ionizes it to the point where the air conducts electricity and continues the arc without actual metal-to-metal contact. As soon as the arc starts, the rod tip must be pulled back to the suggested tip-to-work distance. All of this sounds tedious and difficult, but you *will* pick the technique up in the first half-hour of practice.

If you touch the rod to the surface for more than a split second, it may stick firmly, in which case the rod can get red-hot for its whole length in a very short span of time. As soon as you feel the rod stick to the surface, squeeze the clamp on your electrode holder to release the rod, which is the only way to stop the rod from melting. The hot rod will stay stuck to your work. You can take it off with some pliers, but when it is really hot, don't try snapping it off with your gloves, it may burn right through the leather. Let that rod cool off and start another rod until your have mastered the arc-starting process. When you are more experienced, you'll react quickly enough to a stuck rod that you can simply break the connection immediately by twisting the electrode holder and rod side to side.

Another method of arc starting favored by some weldors is a "tapping" style, in which you quickly tap the electrode tip to the work to start the arc, not in a "pressing" action, but in a short in-and-out jerk that makes contact at the bottom of the tip's travel toward the work.

Once you understand the starting process, the tip to work distance is next. When the arc starts, you must pull the rod back for a second to make a relatively-long arc (about twice the thickness of the electrode you are using) as a way of inducing some preheat into the metal, then immediately drop the rod closer, to about one rod-thickness away from the work. Keep the rod moving along the seam, and move the rod side-to-side slightly as you travel. Some weldors use a movement like a series of tiny, overlapping ovals, others a zigzag or even a "weaving" pattern. Stay in one place too long or with the rod tip too close to the work, and you'll melt a hole in your work; also, if you pull the rod back too far, you can lose the arc process and have to restart. On small seams in thin materials, you won't need to weave the rod much. When joining thick materials, the joint is usually Vee'd or beveled, and a straight pass is made along the bottom of the joint, followed by one or more passes where the oscillation or weaving of the tip spreads the bead circles out larger in the wider gap at the top of the bevel.

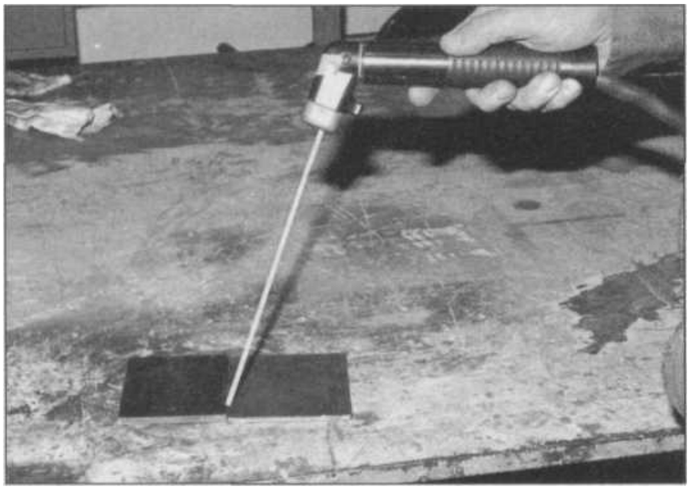

4.16 The electrode is usually held more or less perpendicular to the welding surface to start the arc, then laid back to continue the bead. Some weldors prefer a forehand technique, others a backhand direction, such as here.

The rod should be more or less vertical when the arc is struck initially, where it can really preheat the metal, but should be angled forward (in the direction of travel) 20-30 degrees as you make your pass, i.e. you hold the electrode holder somewhat ahead of the puddle, with the rod 20-30 degrees from vertical **(see illustration).** Try making some straight passes along a flat, horizontal plate, until you get the hang of running a bead. Run a straight bead with the puddles being about twice the diameter of the rod you are using. While you are welding, it is important to remember not to watch the arc, but rather focus on the puddle you are leaving behind it. The shape, size and crown of the puddle are the keys to determine how you are doing.

When you have the travel speed, amperage and rod-to-work distance correct, you'll find out why they used to call an AC arc welder a "buzz box," because you'll get a very satisfying sound, which is a steady, crisp noise something like bacon frying. This sound your welder makes, and the way the bead looks will tell you much about your progress. If you have the arc gap too large, you'll have uneven puddling, a bead that is too wide, and the sound will be uneven. Such a weld will have more than normal spatter. On the other hand, if you have the rod tip too close, it may stick to the work, the bead will be high but not very wide and the sound will be softer.

One of the hard parts to learn here compared to gas welding is that the part you are holding, the electrode holder, must continually be brought closer to the work as the rod gets shorter, while in gas welding the torch stays the same distance from the welding. The hand-eye coordination you must learn involves keeping the tip the right distance from the work, the rod at the right angle, the correct speed of travel, and compensating for the shortening of the rod. Speaking of electrode length, when the rod gets fairly short, it's best to stop and then re-start with a new rod. When the electrodes are too short, a lot of extra heat travels to the electrode holder and your gloved hand.

The speed you travel is almost as important as the rod-to-work distance. If you travel too fast, the resulting bead will be too narrow, and you may not get 100% penetration. If you proceed too slowly, you'll wind up with a large bead,

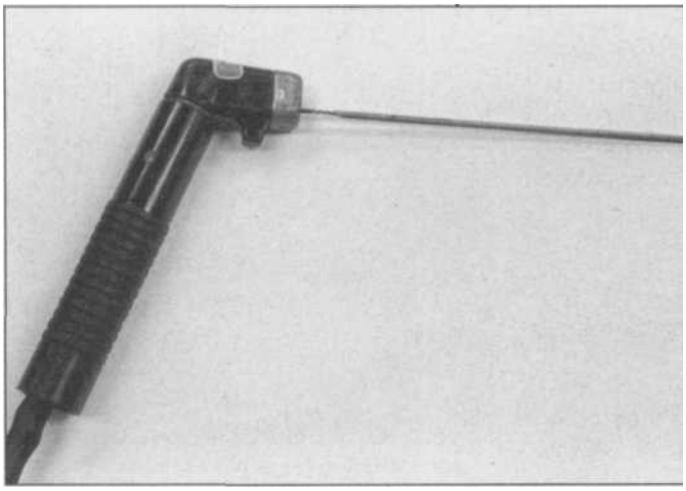

4.17 Arc-welding electrode holders are usually a clamp-like affair, with notches to allow the rod to be clamped at several different angles. Many arc weldors prefer this style, called a short-stub, in which the rod is inserted and the handle is twisted to lock the rod in place.

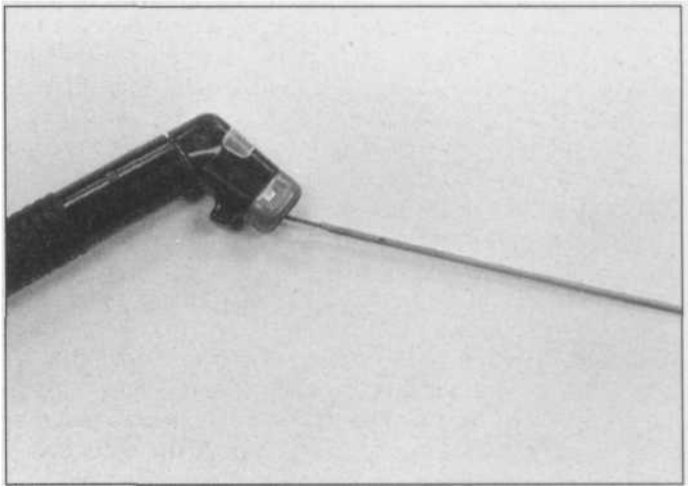

4.18 To achieve a different rod angle with a short-stub electrode-holder, just bend the rod near the holder to the desired angle. In some situation, where you have to reach down inside something to weld, the rod can be bent straight out.

and you may induce excessive heat into the workpiece. After you have practiced speed and arc-to-work distance, practice stopping and starting a bead. In real-world welding, you will encounter seams that take several sticks to complete, but you want the completed weld to look continuous even if it wasn't. To stop a bead, when you have to change electrodes for instance, just pull the rod back up quickly to break the arc. Any time you stop arc welding, you must chip the slag away from the place you last stopped, before continuing with a new rod. When you pick-up again with the new rod, start about 1/2-inch ahead of where the last puddle was and re-strike your arc and proceed. The arc will melt the original "last puddle" and continue the bead without any apparent interruption. This will take some practice. If you have seen professionally-welded seams, they look like they were applied by a continuously-operating machine, so integrated are the stops and starts, and that's what you are shooting for.

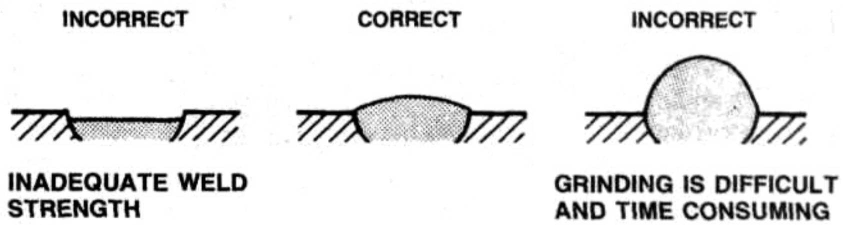

4.19 Too much amperage and the bead will sink too far into the work and leave cratering or undercutting on the top. If your welding speed is too slow, the bead can build up unnecessarily on top.

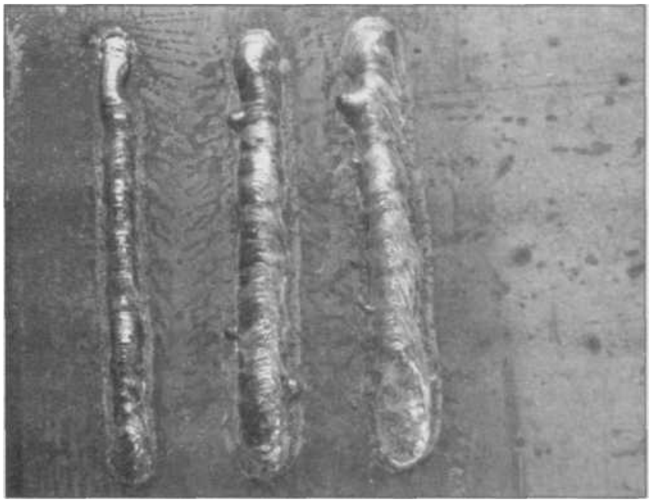

4.20 In these three practice arc-welding beads, the left-hand bead was too "cold", meaning not enough amperage. The center weld is OK, and the right has too much amperage, with increased spatter and too wide a bead for the thickness of the material.

Types of joints

After you are experienced at making straight beads on plates lying flat in front of you, begin to practice on joints between two pieces. The simplest to learn on are butt joints. If the material is 1/4-inch or thicker, you should bevel the edges of both parts before welding. As with any parts to be welded, either ferrous or nonferrous, cleanliness of the work is very important to making a sound weld, so grind the pieces to be joined not just on the bevel, but at least 1/2-inch on either side of the joint, so that impurities don't contaminate the weld.

Most seams are started by tacking the parts together at either end and perhaps several places along the way, depending on how long the seam is. Because of the growth of the parts from the heat of welding, you may need to "build-in" a gap between them during the tacking phase. Some weldors place a small piece of bare copper wire between the two parts, tack one end, and then move the wire "spacer" to where the next tack will be, continuing so that the two parts are tacked slightly apart from each other with an even gap. As you weld up the seam, the parts will "grow" together. Whenever we mention tack-welding in conjunction with arc welding, remember that the slag must be chipped and wire-brushed away from the tacks before you "connect the dots" with continuous welds.

Depending on the thickness of the material you are working on, you may make a two-pass weld, one on the top and one on the bottom. This would only be feasible on plates, not on pipe or tubing, but makes for a very strong joint, and the opposing forces of distortion may keep the parts flatter than if you made only one large pass on one side. Looking at an end view of the plates, the first pass should go into the joint a little more than halfway, then the second weld on the other side should bite into that first bead for a completely-welded joint. Usually, such joints do not require as high an amperage setting on the welder as for making a single-pass weld.

One of the common uses of arc-welding equipment is in farming and ranching, where the equipment being repaired is usually too large to bring into the shop. That means that most welding that requires a bottle of shielding gas may be too susceptible to wind to be effective, and the arc-welding

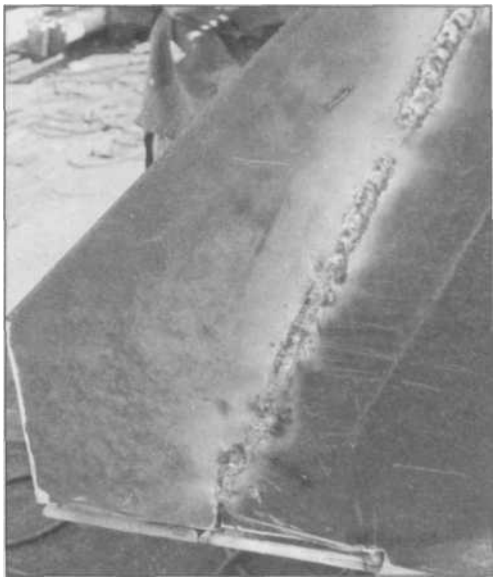

4.21 Welding large plates, such as here where new metal is being added to a John Deere dirt scoop, requires either multi-pass welding, or using a large rod at high amperages. This seam was done with a single pass with 1/4-inch electrodes.

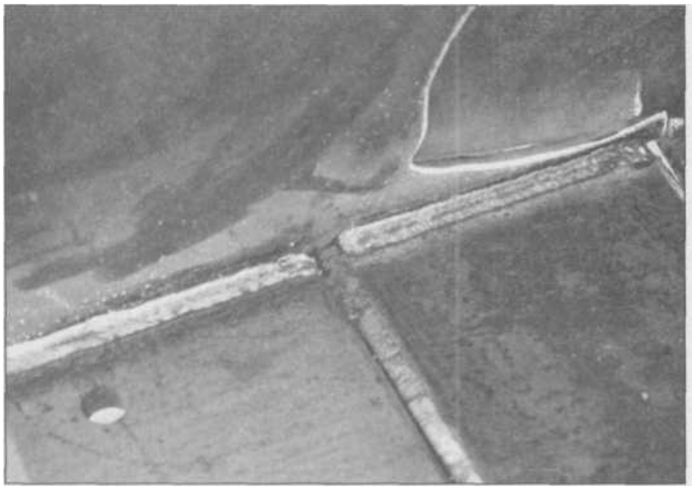

4.22 Here half-inch plate on the sides was welded with one pass, but a gap was left where the one-inch, beveled lower plates join. The bigger plate will be welded with several passes and strength is improved if the joint starts all the way at the ends.

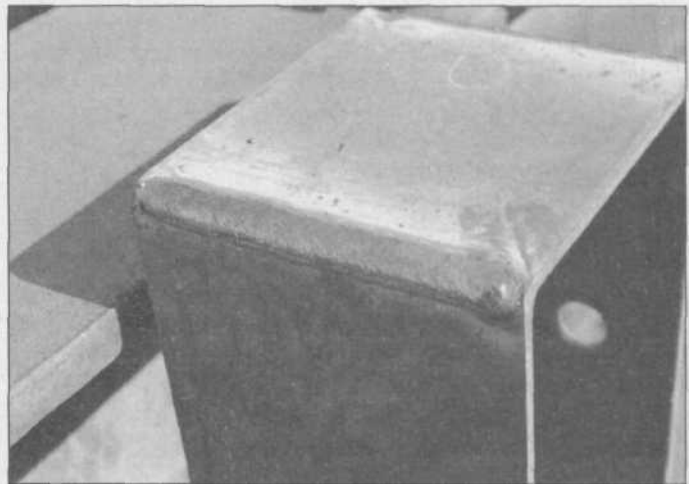

4.23 An outside corner weld on 5/16-inch plates is made with one pass with a wide weaving pattern, spread equal heat on both parts.

process is advantageous here. Also, most farm equipment is rather heavy, with large components, and arc welding's ability to fill large breaks, weld castings, and deposit large quantities of filler metal if needed makes arc a good choice. The visual beauty of the welds is seldom a factor in agricultural repairs: it's getting the equipment back into service that counts. Building up of worn agricultural parts is a common task in which the weldor makes repeated beads right next to each other to literally build a new and higher surface on the part. The work is tedious, but necessary.

Generally, you make the first pass in normal welding mode, then make succeeding passes with the electrode held at a slight side angle to the work. Remember than when arc welding, the hot filler metal is virtually sprayed off the end of the rod like a spray gun, and where you aim the rod is where the metal will be deposited. The slight side angle allows the succeeding passes to bite slightly into the previous bead. Alternate the direction of the beads and overlap them about one-third of the bead width. There are even special rods made that deposit a hard-facing on metal. These can be used to renew or create a hard cutting edge on a part, which can be later ground or filed to an edge that will stay sharper longer than the parent metal. When making any kind of 'buildup" welds, each bead must be chipped clean before you begin the next pass, or slag and impurities could be trapped by the next bead. Also, when you are building up a part, rather than fusing a seam as in regular welding, you can run the bead much wider than we have suggested so far. The idea here is not to fuse two parts but to deposit as much metal as possible onto the surface. In these cases, you can "weave" the rod tip back and forth in a zigzag or other pattern to make a wider bead.

You will find that "out-of-position" welds, welds that are not made with the parts lying flat in front of you on the welding table, are the toughest to learn. Corner joints, vertical seams and overhead seams offer increasing levels of challenge. Corner joints are commonly found when joining two plates where one plate is perpendicular to the other, like a T. The electrode should be at a 45° angle between the two parts, so that equal heat and filler metal is directed onto both pieces. In some cases, you may find you have to place the arc such that it puts more heat on the bottom plate (which is dissipating more heat from the joint) while aiming the electrode "spray" more at the upper plate to avoid "undercutting", which is when you see a slight crevice along one or both edges of a welded joint after you chip the slag away.

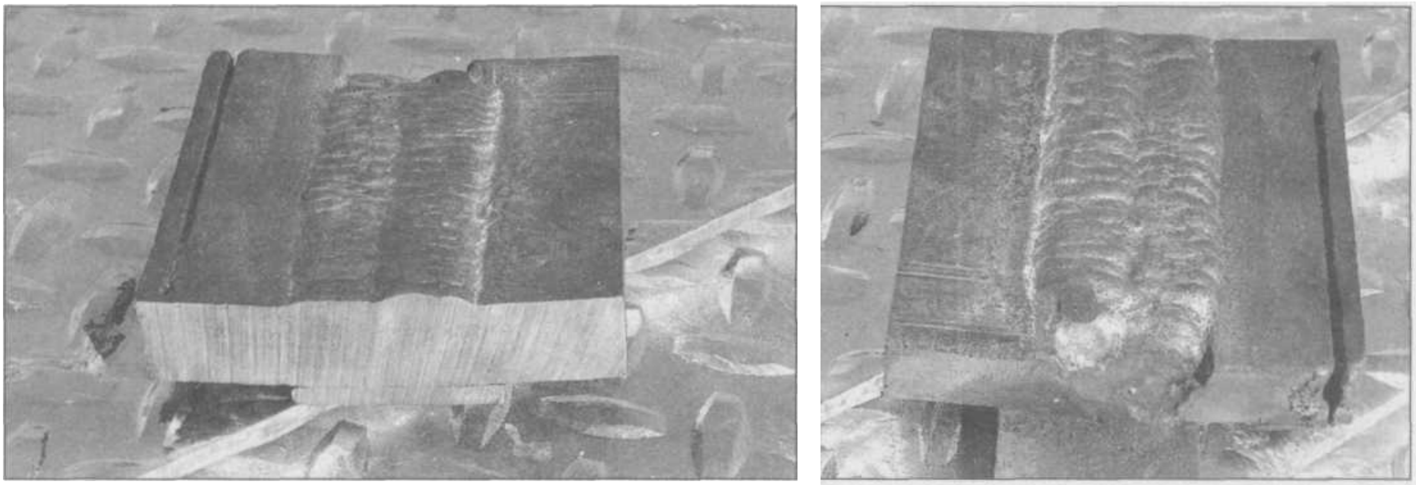

4.24 This sample weld shows how large plates are joined by multiple passes with large rods. Compare this front view (right) where the weld passes began, to the cutaway photo (left) which illustrates how integrated and free of contaminants even such a large weld can be.

If you are joining heavy plates, the fillet weld may take three passes, in which case the first bead is put right into the corner, then a second pass is made alongside that one with the electrode aimed to "spray" slightly more toward the side between this bead and the corner's edge, and the third pass is aimed to the opposite side of the center **(see illustrations).** If, on the other hand, you are welding a thick plate to a thinner one, then favor the angle of the electrode towards the heavier plate to give it more heat. In practicing such welds, it may be advantageous to use plates that are only an inch or two by six inches. This will make for less work when cutting a cross-section through them afterwards to examine your weld for penetration and lack of voids or impurities, especially if you have to hand-hacksaw through them.

While basic, flat butt joints are most common in typical home/shop projects, the Tee or fillet joint and the lap joint make up the majority of welds in industry and construction. The lap joint, where one plate lays on top of another, is approached much like the Tee joint, with the electrode angle at about 45°, aimed at the inside corner where the two plates meet. If they are different thicknesses, then the electrode may need to be aimed to give more heat to the heavier piece. Another difference between industry and home projects is that you will seldom encounter any joint that requires more than one pass to complete in a home/shop project. The use of really thick plates just isn't that common in automotive, arts and crafts or making shop equipment, with the possible exception of building an engine stand, which might have a half-inch plate welded to a tube for the part the engine-holding arms attach to. This is one reason that arc-welding isn't that common anymore as a home/shop tool. Other methods seem better suited to the thinner materials encountered in hobby projects, though farm equipment repair does sometimes require the large-scale abilities of an arc welder.

Choosing electrodes

Your home AC arc-welder will probably come with an assortment of basic welding rods to practice with. After that you'll have to shop at your local welding supply or order from a catalog, and you should have a good idea by then of what type and thicknesses of metals you'll be working on, and thus what size and type of rods to use. Your local supplier should be able to give you good advice on what best suits your purposes. As you read or talk with other people about arc-welding, you may hear the rods termed any of the following: stinger, rod, elec-

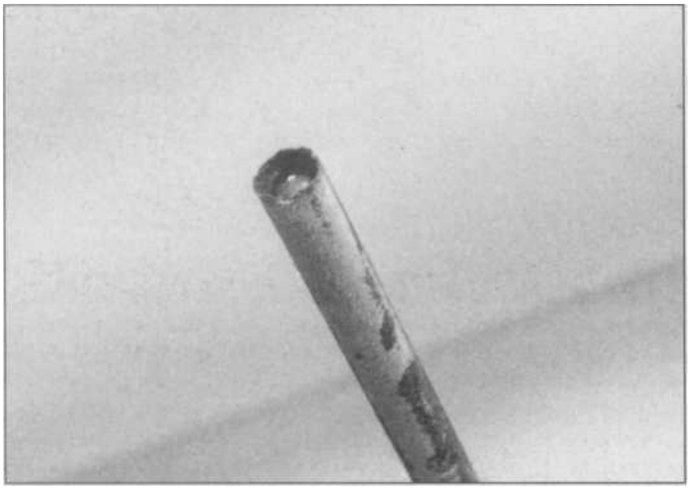

4.25 Some arc rods melt the wire at a faster rate than the flux, leaving the wire recessed inside. During the welding, this makes a more concentrated arc spray. When re-striking an arc with such a rod, the extra flux around the tip must be broken off with your gloved fingers or pliers (with the rod cooled off) to expose the metal for re-striking the arc.

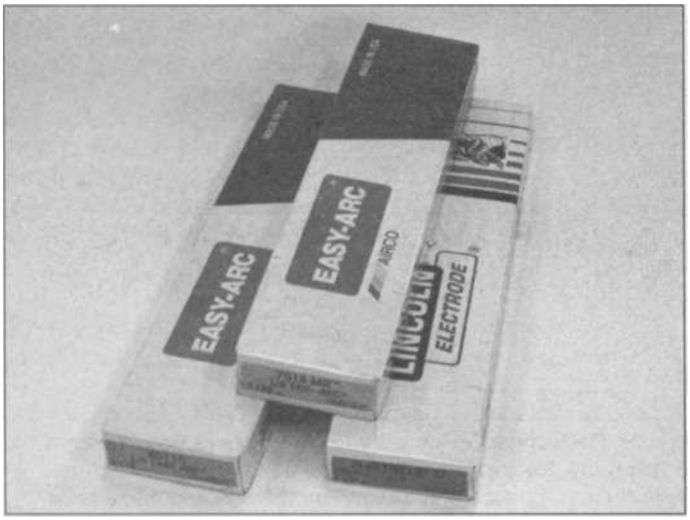

4.26 You'll find a wide variety of electrodes at your welding supply store, but there are only a few rods you will use in normal home/shop projects on mild steel. Don't buy much more than you will need for the next project.

trode or stick. They all mean the same thing.

While there is a confusing array of arc-welding electrodes on the market, the bulk of them are specifically designed for industrial and other specialty usage, and you will never have a call for them. There are perhaps only a half dozen rods that would cover your needs. Rods differ in the type and diameter the central metal wire is made of, and in the thickness and composition of the coating of flux as well. In actual use, the flux burns away at a slower rate than the rod inside, which makes for a sort of "collar" around the tip of the rod, further shielding the weld process and helping concentrate and direct the "spray" of metal coming off the wire **(see illustration).**

Many of the hundreds of special-purpose arc rods are designed strictly for the professional DC welder, so you won't have to worry about those if you're using a typical 220V home AC welder. For your needs, there are perhaps three types you will use the most. Of these, E-6011 is the designation for what is the most commonly-used electrode. It is one of the easiest for a beginner to master, can be used on AC or DC, in virtually all positions of welding, and can be used in situations where the parent metal hasn't been prepared spotlessly. It's main drawback is that it produces a lot of spatter. Where the *appearance* of the weld is less important than the ability to make a strong joint on rough or dirty materials, such as in farm equipment repairs, this is a good choice. After practicing with this one, you may want to move on to more "sophisticated" electrodes. The E-6013 rod works great in many situations, and, although it requires better surface preparation and cleaning than the 6011 rod, it will produce much better looking welds, is suitable for a variety of positions and handles metals up to 3/16-inch thickness with most home-type buzz-boxes. It can be used for lots of projects like building shop equipment, and there are even variations on this rod from different manufacturers that are specifically for sheet-metal work, in sizes down to 1/16-inch. When the size of a welding rod is mentioned, it refers only to the diameter of the wire *inside.* Obviously, the flux coating makes the rod appear much thicker, and if you need to check a **rod,** measure the bare end meant to go into the electrode holder.

There has been a variety of methods used in the past to identify welding rods, including numbers and color codes, but most rods you find today have both a number and code from the manufacturer, as well as a standard A.W.S. number, such as E-6013 **(see illustration).** These standard designations come from the American Welding Society and should be easy to find somewhere on the boxes of rod you buy. The E stands for electrode (for arc welding), the 60 part is multiplied by 1000 to give you the tensile strength of the wire (in this case, 60,000 psi), the next digit is a code for the type of position the rod is recommended for, and the last digit refers to the type and polarity of the current required. The 1 in our example in the "position" spot indicates the rod is good in all positions, a 2 would mean flat position. The last number indicates that this rod can be used with AC or DC, and on DC can be used with straight or reverse polarity, though straight polarity is seldom used.

Another good rod for starting out is E-7014, which is higher in strength and produces good-looking beads. It and the E-6013 mentioned earlier require a little different arc procedure than what we have described so far. These are called "contact rods" because, instead of maintaining a basic 1/8-inch gap as you go along, you keep the tip just in contact with the parent metal, in a sense dragging the rod lightly along the seam.

A special-purpose rod used in repairing cast-iron, such as automotive blocks, cracked heads, transmission cases, and various farm equipment, is EST, although there are several brand names in the trade just for cast-iron rod. Most rods for this purpose have a high nickel content. Although these rods are available, it doesn't mean that cast-iron welding is easy. Any cast part should be thoroughly preheated to 400° F or better to prevent cracking after the localized heat of welding is induced into the part. Parts, depending on their size, can be heated in an oven or with a rosebud tip on an oxy-acetylene torch, using temperature indicating paints to tell when you have it evenly heated to the proper temperature. Even then, there can be cracking problems. It can be very difficult to get a cracked cast-iron part thoroughly cleaned before welding, usually requiring a deep Vee to make a good repair. Better success on cast iron may be had by welding in short strips 1/2 to 3/4-inch long, with the part cleaned and allowed to cool off in between. In automotive use, broken cast-iron exhaust manifolds are a common repair item, and the high-nickel rods can be used successfully, but a well-used exhaust manifold needs to be sandblasted inside and out beforehand to get rid of as much of the carbon baked into the pores of the casting as possible. Usually, a brand-new casting will take a much-longer-lasting weld, such as when modifying manifolds to accept turbocharger flanges. Most small cast-iron repairs are better made with brazing by an oxy-acetylene torch, where much less heat is introduced, although the part should still be preheated.

Arc-welding electrodes are also marked with color codes. Usually, the flux coating itself is a different color to indicate a type, as well as markings or dabs of colors near the tip. The E-6011 has a white flux coating and a blue spot color, E-6013 is a dark tan with a brown spot, etc.

Proper storage of electrodes is critical to their performance **(see illustration).** They are very susceptible to moisture, and must be stored in a perfectly dry environment. Once they absorb moisture, the flux coating tends to loosen and flake off, and the rod is useless. In large production operations where critical welding is done, the rods are stored in a special oven that keeps them at a constant 100° F or more to ensure that they are dry, and only enough rod that will be used in a few hours is removed at one time from the oven.

In a home/shop situation, buy only as much rod as you think you need to keep around for unexpected projects, a few pounds of each type you normally use. Keep these in either sealable metal cans or plastic bags with desiccant inside. Desiccant is a moisture-absorbing compound, the stuff we always find packed in little "teabags" with electronic components and camera equipment. You can save up these bags whenever you get a new electronic goodie. The bags can be dried out in the oven for a short while, then put into the container holding your welding rods. If you have a bigger project in mind, buy what you need for that project only, fresh from the welding supply, and don't keep any large quantities around. We have seen a variety of methods used by weldors to keep electrodes dry, including racks of custom-built metal tubes with screwcaps.

4.27 This is the bare end of the rod, where it fits into the electrode holder. You can judge the size of a rod from this bare end - don't measure the flux coating. Rods are marked as to their type and color-coded.

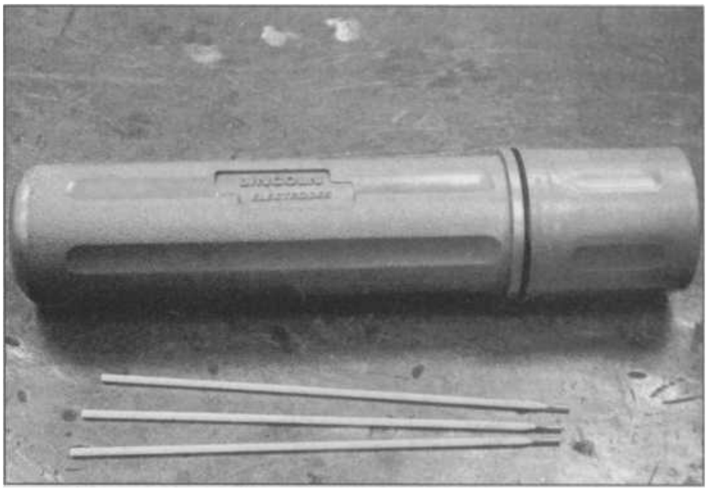

4.28 Rods must be kept dry for the flux coatings to work properly. Keep only as much rod as you need, and store them in a tight container such as this Lincoln plastic rod box with gasketed screw cap.

L-TEC STW 140i
L-TEC STW 90i

4.29 These L-TEC portable arc power sources are lightweight, and offer up to 90 amps on the STW 90i (right) and 130 amps on the STW 140i. Unlike most small arc welders, these are DC, and can even be set up with a torch and power adapter to run a TIG torch. They are available for 110V or 220V input.

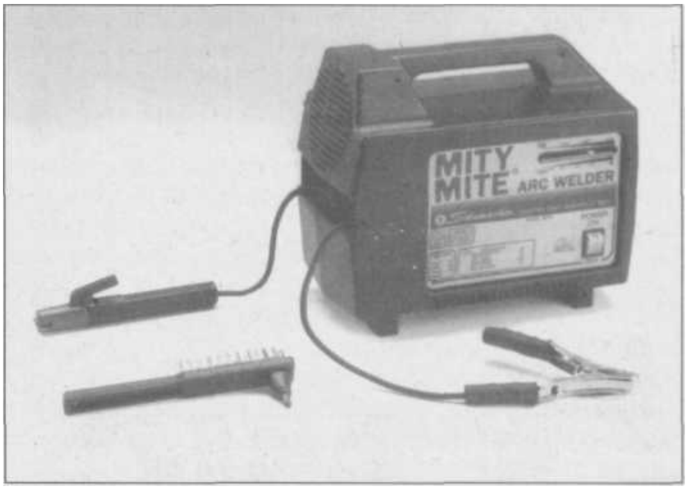

4.30 The Eastwood Company markets a lot of equipment and supplies for auto restoration, and this is their small 110V arc welder. At its maximum of 70 amps, it delivers a 10% duty-cycle, so you won't be welding any bridges together, but it only weighs 26 pounds, so it's extremely portable and welds up to 3/16-inch.

If you have reviewed the above introduction to arc-welding, you should have a good idea by now of whether it will suit your particular welding needs. It can be both easy-to-learn and challenging, simple and complicated, and can produce results from "farm-grade" (where strength is primary) to automotive-quality to atomic-reactor quality. It has the advantage of simple operation, low maintenance, low initial cost (not considering the cost of rewiring for 220V if necessary), and the unmatched ability to weld outdoors without being affected by wind, making it suitable for structural-steel construction, pipelines, iron-fence work, and farm repairs, as well as being the best choice for welding and repairing thick materials.

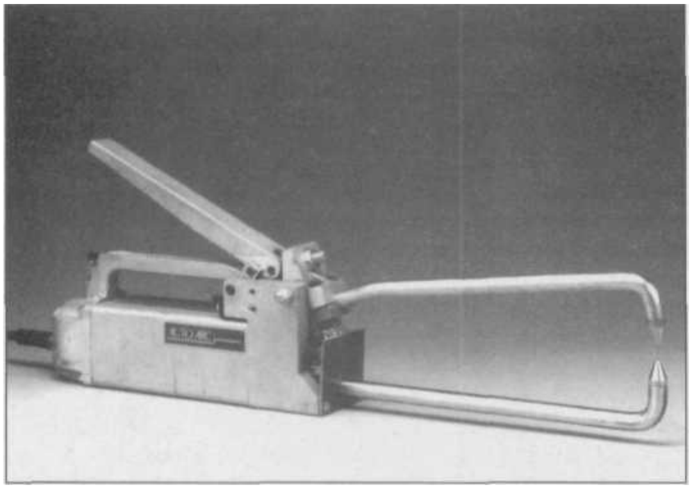

4.31 Auto manufacturers build cars with spot-welding machines, and with Eastwood Company's spot-welder gun, you can repair one the same way. The replaceable electrodes last for 60-100 welds.

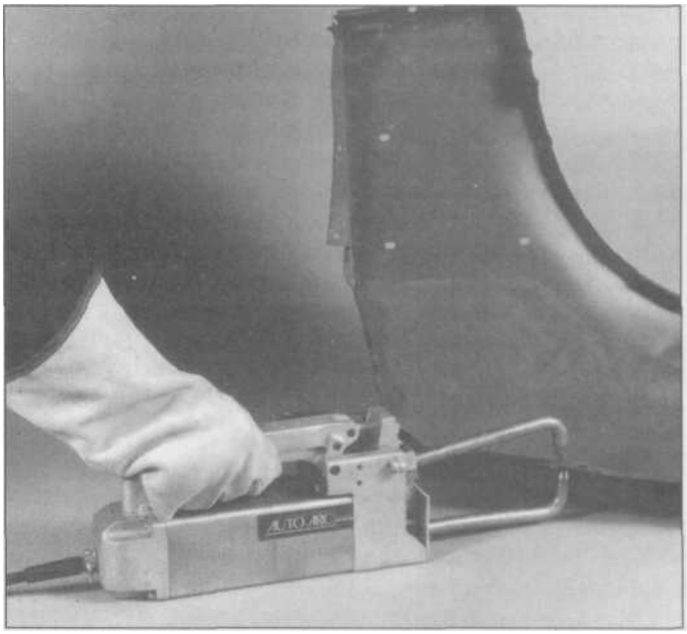

4.32 The spot welder is a self-contained mini-arc-welder, designed for joining sheet-metal panels on auto body work.

5 MIG welding

Welding books written as recently as ten years ago may have recommended oxy-acetylene and arc-welding equipment as the basic tools for home/shop welding. At that time, MIG, or *wire-feed,* welding equipment was considered too expensive for amateur use, despite its advantages. The MIG equipment was recommended, and indeed was originally designed, for high-production shop work only. Much has changed in the intervening years, and the introduction of lower-cost MIG units and the competitiveness of the marketplace has brought wire-feed welding into an affordable range for the home/shop weldor.

An outgrowth of arc welding, the MIG process was conceived as a way to speed up production in industrial applications. The basic difference with MIG, which stands for Metal Inert Gas welding (sometimes also referred to by its technical description as Gas Metal Arc Welding, or GMAW), is that the welding filler rod is automatically fed through the gun whenever welding takes place. The weldor doesn't need to stop to change arc electrodes. The Airco company developed the MIG process

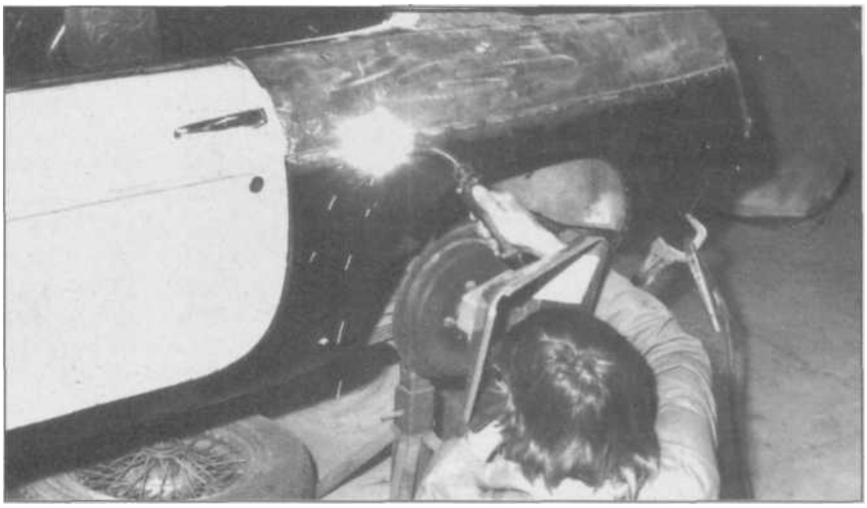

5.1 The use of wire-feed MIG welding has expanded greatly in the last ten years to encompass body shops, muffler shops, race-car builders and home hobbyists as well as the industrial production work it was designed for. It has advantages of speed, simplicity, cleanliness, and the ability to work well on thick and thin materials.

PREST-O-WELD
Mig 130
L-TEC
Mig 130
WIRE FEED SPEED
VOLTAGE

5.2 Among the compact MIG welders in the 110V field is the L-TEC MIG 130, which features 100% duty-cycle at its 30 amp setting for thin material, and has a 25% duty-cycle at 130 amps. It's designed for body shops, hobbyists and light maintenance use.

5.3 For the larger-scale shop, the heavy-duty 220V machines like this Migmaster 250 offer more amperage, more amp settings, higher duty cycles, and various extra features for timing various weld functions.

5.4 The Miller Electric Co. makes a variety of larger professional welding machines, but they also offer this Millermatic 130, which has a 20% duty cycle at its maximum 130 amps, putting it right in the midst of the 110V home/shop MIGs.

right after W.W.II, and it has been refined continuously ever since.

The many advantages of the MIG process, besides the basic one of not having to stop to change electrodes, include: being able to weld in all positions; there is no slag removal and the welds are much cleaner and with very little spatter compared to arc welding; the welding can proceed a lot faster than with other methods because the wire-feed rate is automatic and adjustable, which also makes for less distortion of the workpiece because less heat is concentrated on the seam; the process works well on joints that are irregular or have gaps; and the amperage, wire size and wire-feed speed can be tuned down to do ideal, almost distortionless welds on thin sheet metal. There are more advantages that relate to how the weldor uses the equipment, and we'll see these points as we continue.

There are technically three different processes of metal transfer with MIG welding equipment, but the one we are concerned with is called the *short-arc* or short-circuit process. The other types are used in larger industrial welding processes where high amperages and large-diameter welding wire is used. Most of the MIG machines suitable for home/shop use are designed for short-arc process, where the molten end of the welding wire touches the weld puddle and creates a short-circuit, and the wire diameters are from .023-inch to .045-inch. Amperages in these machines are seldom higher than 225 amps yet can handle materials from the thinnest auto-body sheet metal to plates as thick as 1/2-inch, a range that certainly covers everything the home/shop weldor, metal sculptor, body man, farmer or ornamental iron worker will have a call for.

In the short-arc process, the basic components of the MIG equipment consist of the welding machine, which contains an AC-to-DC rectifier with constant voltage potential and a wire-feeding mechanism that holds a large roll of wire, a bottle of shielding gas (in most models), a welding cable that routes wire, power and gas to the torch, a simple trigger-operated torch or gun and a ground or work cable. In operation, the weldor brings the torch down to the work until the bare wire electrode, which sticks out of the gun 1/2 to 3/4-inch, is touching the

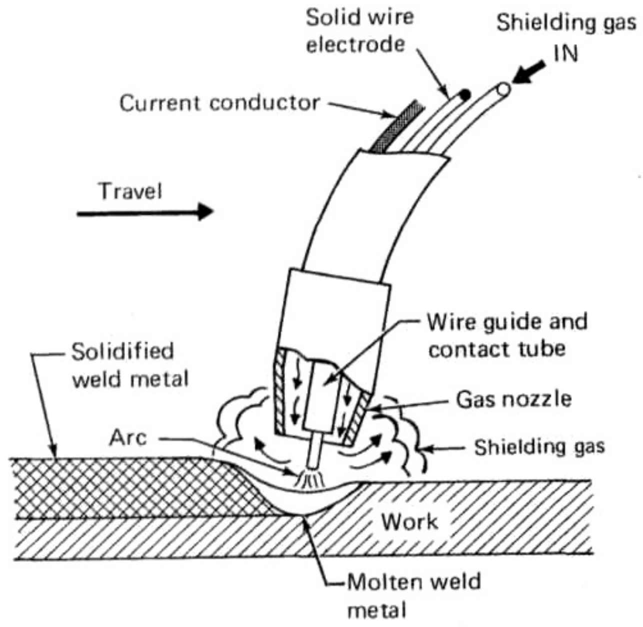

>—Molten weld
metal

5.5 The cable leading to the MIG torch has to carry several different materials, from the electrical current, to the wire electrode and the shielding gas. The torch itself is rather simple, feeding the wire out to maintain a short-circuit arc against the work, while keeping the weld under an envelope of shielding gas.

5.6 There are very few user controls on the panel of a basic MIG welder, just a switch for the different amperage settings, and a variable knob for controlling the wire-feed speed.

work where the seam is to be; he then flips his helmet or lens down and pulls the trigger. The trigger immediately puts power to the wire, which arcs against the work to create very localized heat while shielding gas is simultaneously released that puts a shielding "envelope" over the forming weld, preventing contamination **(see illustration).** The short-arc process actually is alternately melting and not melting the electrode wire about 90 times a second. Each time the wire touches the work it melts and creates a gap between itself and the work, which is when a drop of molten wire attaches to the work and blends in. The constant arc-on-arc-off is what gives the MIG process its characteristic "sizzling" sound when you weld with the correct wire speed, amperage and travel speed. The current used is DC with reverse polarity (like most of the professional arc-welding machines), which provides for deep penetration. Only a few applications and machines have the straight polarity option, and this is only for shallow-penetration work where the coverage or speed of deposition is most important.

The weldor has few controls to worry about in setting up a MIG machine. There is a voltage knob or switch to control the current, and a wire-speed knob that determines how fast the welding wire comes out of the gun **(see illustration).** Selecting the amperage is based mostly on the thickness of the material to be welded. A rule of thumb from the Miller Electric Company is "one ampere for every .001-inch of plate thickness." For example, an 1/8-inch (.125-inch) plate would require 125 amps. Of course, this doesn't mean that a 1/2-inch plate would take 500 amps, because large plates are generally beveled before welding and joined with a root pass in the bottom and several more passes on top. There are, however, industrial machines that routinely weld large plates, such as on ship decks, with very large wire diameter and high amperages. This is just a general rule, and other factors, such as the wire speed and welding travel speed affect the current requirement. In the smaller home-type MIG machines, the amperage settings are fixed, with perhaps four different positions on the current knob, but the wire-speed knob is infinitely variable on all MIG machines. The

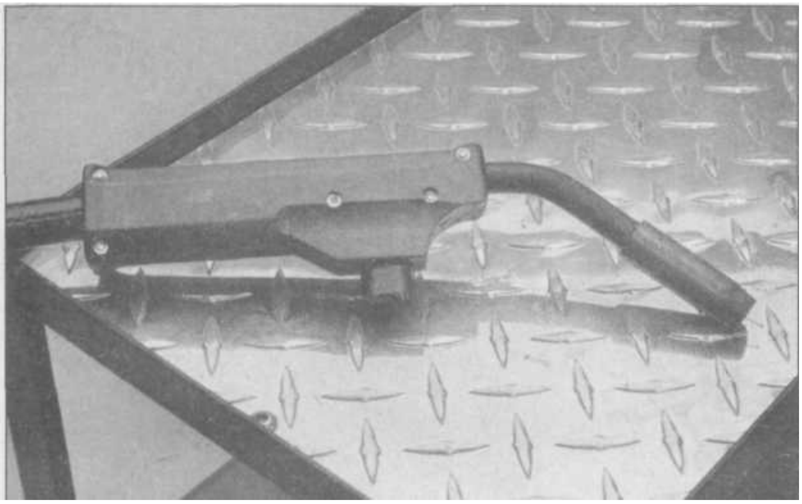

5.7 **The basic MIG torch is quite simple. Inside is a guide tube that the wire comes through, a switch that turns on the power and the shielding-gas flow. The curved neck, called a "swan neck", holds the contact tube and the nozzle.**

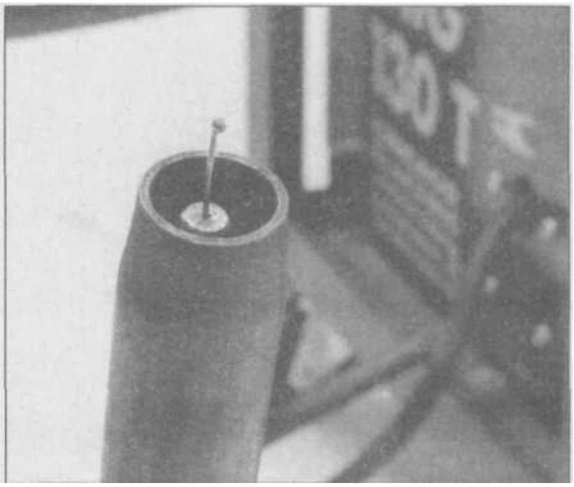

5.8 **The wire must fit the contact tube closely, as this is where power is transmitted to the electrode wire. Note that the contact tube is recessed inside the nozzle, and the wire sticks out 1/2-3/4-inch from the nozzle.**

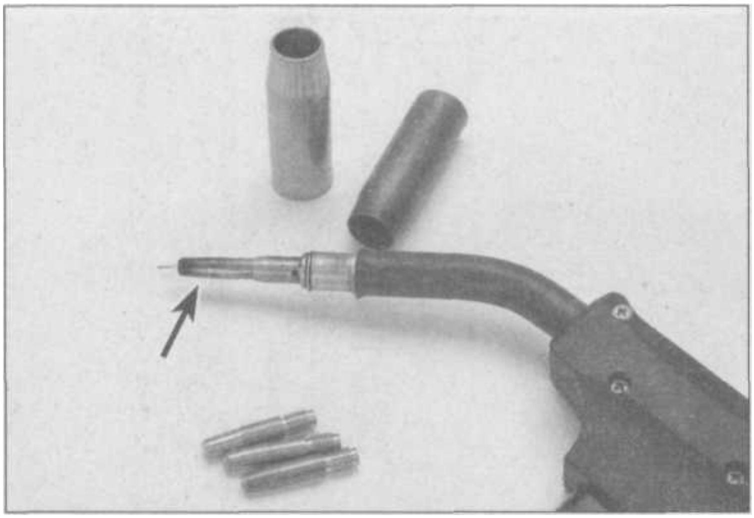

5.9 **Most MIG nozzles just twist off past a snap ring, exposing the contact tube (arrow). You should keep an extra nozzle as a spare, and several spares of the sizes of contact tubes (below) that you use most. Eventually, you will weld the wire to the contact tube, and have to unscrew it to get it off, then clip off the wire and install a new tube, also called a contact *tip.***

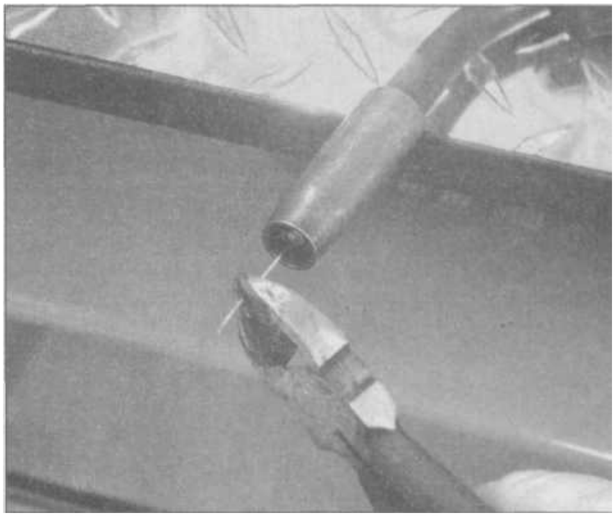

5.10 **It's important to keep a pair of side-cutter pliers handy at all times to your MIG welder. Not only do you need it to maintain the correct protrusion of the wire from the nozzle, but when the wire develops a ball on the end, the arc is easier to start if you snip the ball off.**

more expensive, more professional MIG machines have variable-current output, or at least more positions on the switch, and may have other optional controls for timing weld events.

The torch cooling on almost all MIG machines, certainly the ones we are considering for home/shop use, is by air, although there are some industrial applications which have a water-cooled torch because of the heavy wire and high amperage being used. Looking inside the MIG torch, the main body of the handle contains the switch or trigger, a wire conduit with a metal or plastic-tubing liner, and the curved "swan neck" that comes out holding both a contact or wire guide tube and a nozzle. The contact tube is so called because it actually puts the power to the wire. Usually made of copper, the contact tube also keeps the wire output centered in the gas cup or nozzle **(see illustration).** Since this is where

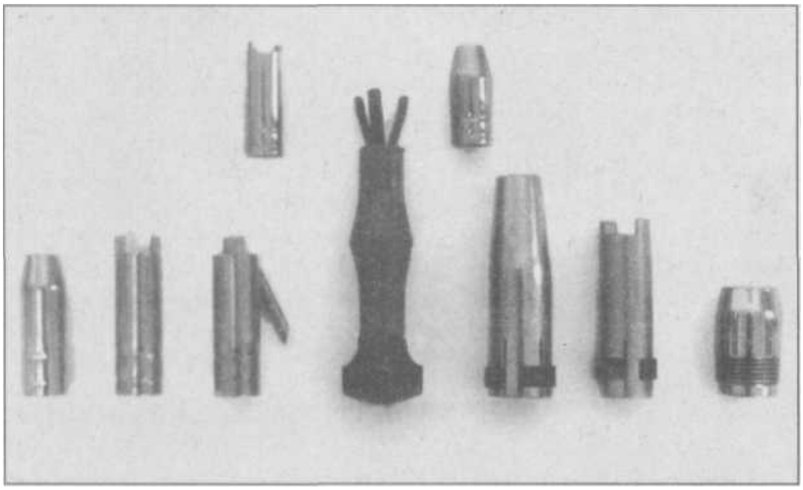

5.11 For some MIGs, there is a variety of different-size nozzles, some with special purposes like spot-welding. At the center here is a tool for scraping debris from inside a nozzle.

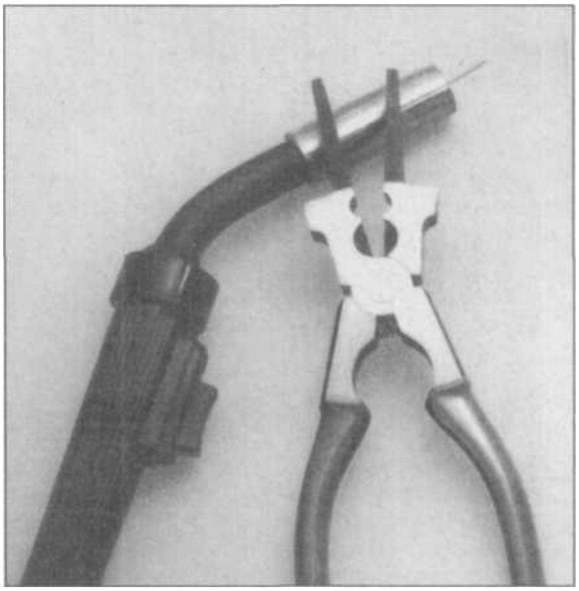

5.12 Eastwood Company makes these multi-function **MIG** pliers, which can be used to clean nozzles, snip wires, and hold contact tips to unscrew them.

the power is transferred to the wire, the contact tube must have a specific size related to the size of the electrode wire being used. Because of the unavoidable wear in the contact tube, and occasional arcing that ruins a contact tube, these are considered consumables in MIG welding, and you should have several spares of any of the sizes you normally use.

The nozzle's main function is to contain the shielding gas in an envelope around the arcing wire. Most MIG machines have only one size nozzle, but in production applications smaller nozzles are used on smaller-amperage work and larger nozzles on bigger jobs **(see illustration).** The nozzle takes some punishment in MIG welding, with spatter and weld debris constantly building up inside. Special tools are available that scrape out the buildup, and most MIG weldors use an anti-spatter or "nozzle-shield" gel or spray to keep down the buildup and allow easy removal of the gook when necessary **(see illustration).** Obviously, if too much debris builds up inside the nozzle, it will affect the gas flow to the weld area and thus cause deterioration and unevenness in the bead. The sprays can be directed inside the nozzle after cleaning, to prevent future buildup, while the gel comes in a jar that is left uncapped while welding. Every so often, you take the hot MIG gun and stick the nozzle directly into the gel, which leaves a coating inside the hot nozzle and on the contact tube.

Shopping for a MIG welder

The array of available MIG machines aimed at the consumer and small shop market has greatly increased in the past decade. Many automotive body workers and fabricators in small shops used to dream of having a MIG welder to round out their capabilities, but the machines available at that time were either very expensive self-contained MIG machines or wire-feeding attachments for already-expensive, large-size arc/TIG power supplies found only in welding shops. The hobbyist, auto body man or muffler shop couldn't justify the expense.

That changed when imported MIG machines started making their appearance in the US, aimed at the hobbyist market specifically. The companies advertised in hobbyist publications, attended auto enthusiast shows, and made a presence in the business community. At first, the big domestic welding-equipment manufacturers took little notice, but after a few years they realized there was a gap in their marketing and price structure and began to produce similar equipment to gain back this emerging hobbyist/small shop interest in MIG weld-

3M
Weld-Thru Coating
WELDER'S HELPER
ANTI-SPATTER SPRAY
CLEARCO WELDING JELLY

5.13 These three chemicals are helpful in performing clean welds and maintaining your **MIG** torch in good condition. The anti-spatter spray and gel are used on the nozzle to make removal of debris easier, and the weld-through coating by 3M is used when welding high-strength steel in late-model cars. Unlike the other sprays, it is applied to the area to be welded, providing extra corrosion protection after the weld.

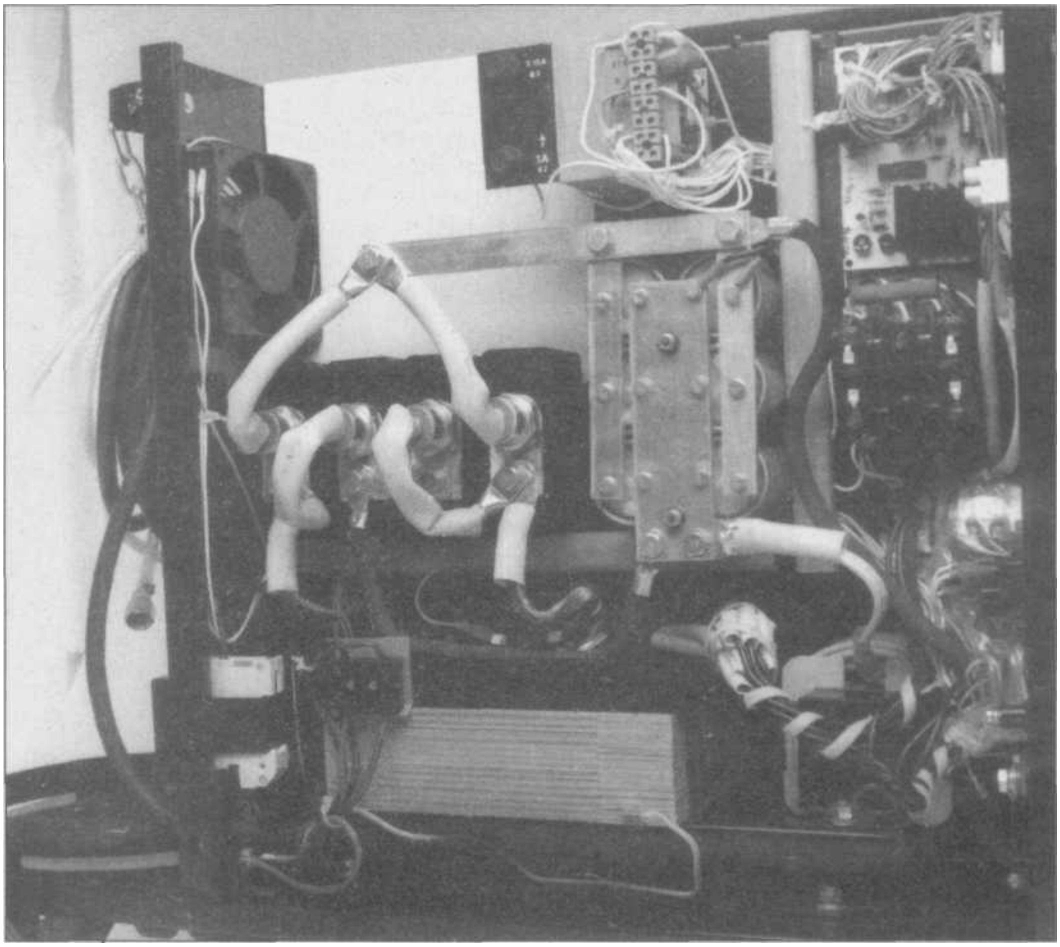

5.14 The bigger and better the machine, the more expensive the internal components. Sometimes you can get an idea of the relative complexity inside a MIG machine by comparing shipping weights when shopping for a welder. The more expensive ones are generally much heavier. This is the inside of a MIG 200 from HTP.

ing. Today there is a wide variety of machines offered from both domestic and imported sources that run the gamut from very small, very portable units that can be powered from a 3000-watt generator if necessary, to larger shop units with all the "bells and whistles."

When the domestic manufacturers started competing with the imports for the lower-price end of the MIG market, there was considerable talk at the local welding supply level about how the imports weren't as good as the domestic machines and that anyone who bought the cheaper imports would have trouble later on finding replacement parts. In the beginning of this marketing struggle, there was some truth to this, but the competition of today's marketplace has pretty much sorted these problems out. There are some very reputable companies selling imported MIG and Plasma machines, and most of them have a full line of consumables and other replacement parts on hand.

As with automobiles, there will always be some debate on import vs. domestic, but today such distinctions are harder to draw than they used to be. Like our cars, some domestic welding machines use components made overseas, and some imported machines use American-made parts. As you shop for a MIG machine or plasma cutter, you may still find resistance to imported equipment at your local welding shop. The answer is to listen to everyone, including especially anyone you know who has purchased and used the equipment you are contemplating. Find out if they have ever needed parts or service for their equipment, and did they have any problems with equipment under warranty. Be sure to compare the usage they put the equipment to, because there is a big difference between what is required in hours of welding on a daily basis in a shop, and your needs of welding a project or two a few times a year.

We have discussed duty-cycle comparisons in the previous chapter. Most manufacturers give you a duty-cycle rating at the highest amperage and the lowest amperage, and you may never have to use either one. If you contemplate a machine with say 140-amp capacity, it may have a duty-cycle rating of 95% at 30 amps and a 30% duty-cycle at its peak of 140 amps. The cycle relates to how many minutes out of a ten-minute period you can weld continuously without overheating the machine. At a 100% rating, you don't need to stop for anything, even if you are welding long seams together. In most home/shop situations however, you are not on a production schedule or a time-clock, and most of your joints will not be long seams. Also, you will probably weld a seam, such as a corner joint on a tubular framework for a utility trailer, that will take you only a minute or less to weld up (doing one side at a time), then you will move to another part of the project, perhaps set up some clamps or fixturing to ready the next joint for welding, and all this time the machine is cooling off so you may never exceed the duty-cycle even at higher amperages. This is where the big difference in cost in MIG machines is. The heavier-duty, shop-type machines have more expensive electrical components inside to handle a higher duty-cycle **(see illustration).** You should buy as much duty-cycle as you can afford to, but don't be too concerned that in your garage you'll spend all your time sitting around waiting for the machine to cool off. You'll find you will not often weld at the highest amperage output anyway, and when you do, even a 30% duty-cycle is livable.

5.15 Sometimes a medium-size machine is all that is called for in a particular shop, like this Miller 150 used in a muffler shop, but the weldor uses a large-sized shielding gas bottle for extra weld time. Note how this weldor has added a tubing arm to the cart, to organize the cables and hold pliers and and-spatter gel.

There are two basic types of smaller MIG machines you'll be looking at, the gas-shielded type that uses bare electrode wire and a bottle of compressed gas, and the flux-core type that uses no shielding-gas bottle but has a special wire with a fluxed core inside that produces the shielding gasses as you weld, much like the vaporizing coating on arc-welding rods. There are advantages and disadvantages to each type, and your choice depends on the kind of work you will be doing. Most of the machines on the market are for use with bottled shielding-gas, and there are some units that can be equipped for either type of wire, in case you have need for both. These machines cost a little more than single-use MIG machines because they have to have a switching device inside to change from reverse polarity (with the torch being positive and the work negative) which is required for the solid wire and straight polarity (with the torch negative) for the flux-cored wires.

The flux-cored wire machines do not make as pretty a weld, if that is a concern to you, and there is some slag to be removed afterward, though it is easier to remove than the slag from arc-welding. With some kinds of flux-cored wire, the slag will peel off the seam in strips. In body shop work, the extra thickness of the flux-cored wire may make them too hot for light-gauge sheet-metal work. Some experts say

5.16 A popular entry-level MIG machine is this 110-V Pocket-MIG from Daytona MIG. Though it weighs only 42 pounds, it has an amp range from 30-110 amps, capable of welding up to 1/4-inch steel. Very portable, it's ideal for home/shop use, especially on sheet metal.

5.17 A 110-V, 130-amp MIG is made by several companies for the home/shop market, this one is by Daytona MIG and welds up to 5/16-inch steel. It features a cooling fan and a 30% duty-cycle at maximum amperage, 95% at 30 amps.

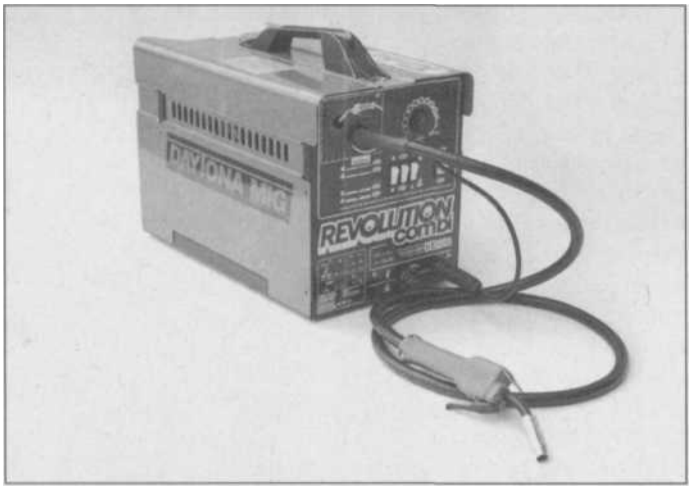

5.18 If you need the flexibility of welding indoors and out, a combination **MIG** that uses either gas-shielding or flux-cored wire fills the bill. This Combi 888 runs on 110V power, with 130-amp capacity.

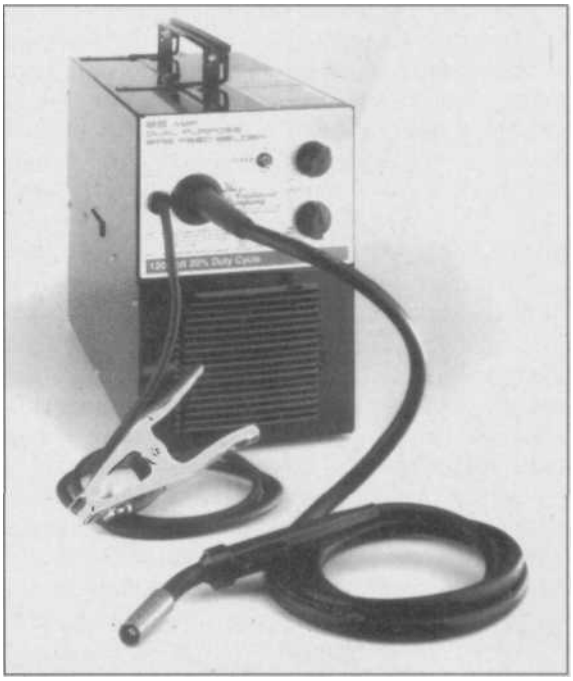

5.19 There are small **MIG** machines that come equipped for use with flux-cored wire, but have an optional kit that can be attached to work with shielding gas and solid wire. Eastwood Company's wire-feeder uses **110V** current, and has a 20% duty-cycle at its maximum amperage of 85. It's fan-cooled and adequate for light-gauge welding.

Gas Conversion Kit for Wire Feed Arc Welders
Operating, Safety, and Installation Instructions

5.20 Eastwood's gas conversion kit for their MIG welder contains the regulator, nozzle, tips and other parts to adapt a shielding-gas bottle to their 85-amp welder.

that flux-core MIG welding is best suited for 18-gauge or thicker metals, and most cars today are using 22- to 24-gauge steel, except for 20-gauge in areas like floor pans. Even replacement steel patch panels for cars are seldom thicker than 20-gauge. Virtually all of the newer cars are using lighter-gauge metal of higher-rated alloy to reduce the overall weight of vehicles to meet tougher fuel economy and emissions standards. The main advantage of the flux-cored machines are in outdoor or drafty situations where wind can blow away the shielding gas from a "bottle-fed" machine, ruining the weld consistency. The flux-cored wire works well for fencing repairs, farm equipment too big to be moved indoors, outsized sculptures and similar projects. These machines also have the advantage of not needing periodic refilling of gas bottles. As long as you have electricity and a spare roll of wire, there's nothing for you to run out of halfway through a project on a Sunday afternoon when you can't get your gas bottle refilled or exchanged.

The advantages of the bottle-fed machines are numerous, the primary benefit being the higher quality of weld and the ability to work on very thin materials without burning through. With shielding gas, there is very little spatter, no slag, and you can even weld body panels that are painted or rusty, though a cleaner parent metal will always make a better weld and produce less hazard from fumes. We have seen a big-rig repairman MIG-welding a new and painted fan guard to a rusty truck radiator-support framework. The budget for the job didn't allow for the time that would be spent in grinding and sanding all of the edges clean - the truck had to be back on the road making money. The result wasn't

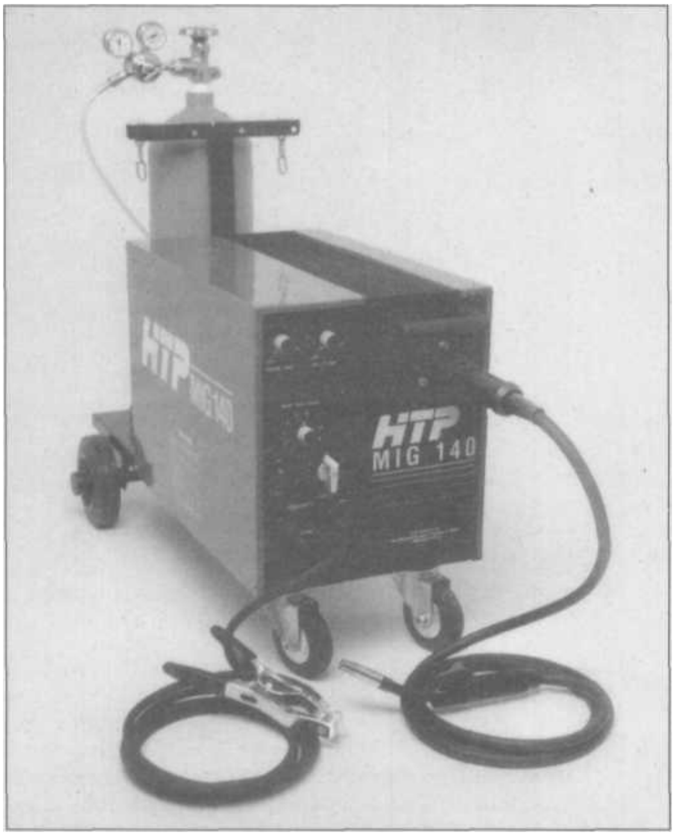

5.21 At the upper end of the 110V **MIG** scale are 140-amp units like this one from **HTP** America. It weighs 135 pounds, welds up to 1/4-inch steel, and has timing features such as stitch-welding and spot-welding.

GRENELL STUD / MIG WELDER
HTP
MIG 160

5.22 Above the 140-ampere range, you are into the 220V category of **MIG** welders, with more features and higher duty capacities. This is HTP's **MIG** 160, which can weld up to 5/16-inch steel, has eight heat settings, and is suitable for street rod shops, muffler shops, or farm use.

pretty, but it was more than strong enough for the application and a testament to what can be done.

As far as running out of gas is concerned, this is an issue, but most home/shop users handle this by just buying a bottle that is twice as big as they think they need. There is, of course, a pair of gauges on the tank that indicate the gas-flow going to the welder and the total pressure left in the bottle, so you should know before any big project just how much welding time you have left. Bigger shops just keep a spare bottle of gas on hand and switch if they run out, then send the empty back to the gas supply company for refilling to become the next spare.

Input current is the next consideration in shopping for a MIG machine. There's a number of home-type machines that can be operated on standard household 110V current, which negates the need for any rewiring of your shop area and also makes them more portable in the sense that you can use one anywhere there is a source of standard power. However, any circuit used for a welding machine should be at least a 20-amp circuit, and, in the case of the bigger 110V machines, a 30-amp circuit is recommended. The 110V (it's confusing, but household current is also referred to as 115-volt at times, but 110V and 115V are the same) machines are available in power abilities with a range of 30 to 110 amps on the smallest home MIGs and a 30- to 140-amp range on the largest. Most of these machines handle wire diameters from .023-inch to .035-inch, though some of the imports are set up for metric wire sizes, which are close enough to US sizes that you can always find replacement wire to suit them.

Above the 140-amp range in maximum power brings you into the bigger shop-type machines that require 220V current input. Most of these machines

MIG
181 C
DAYTONA MIG

5.23 This 220V unit, the 181C from Daytona MIG, offers settings from 20 amps for light sheet metal to 180 amps for heavier work, and has a spot-timer and quick-disconnect torch. This is suitable for light fabrication shops.

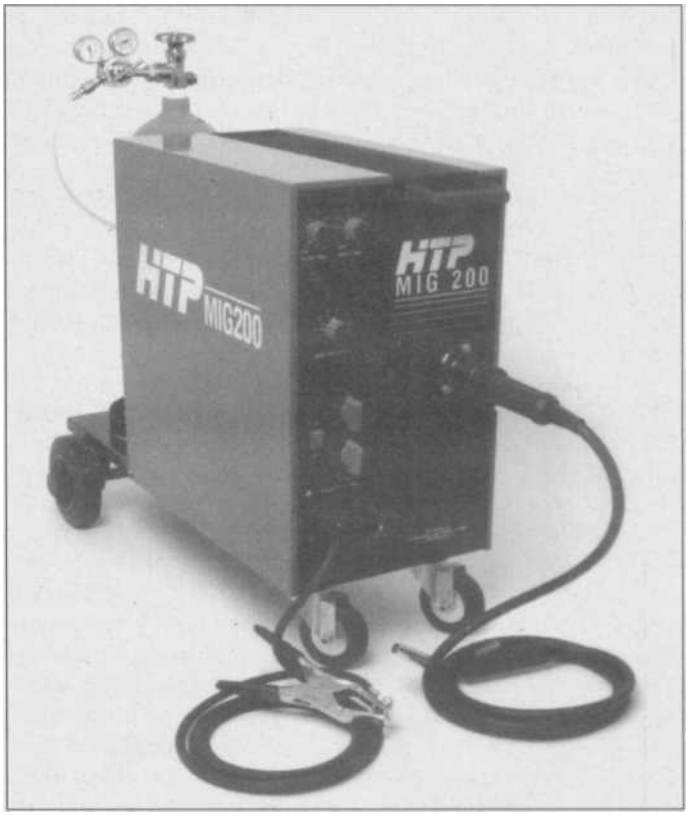

5.24 For welding up to 3/8-inch steel, a **MIG** like HTP's **MIG** 200 will do the job. It has 24 heat settings, up to 200 amps, and is suitable for farm or ranch work, commercial fabrication, or repeated automotive projects like building trailers or engine stands.

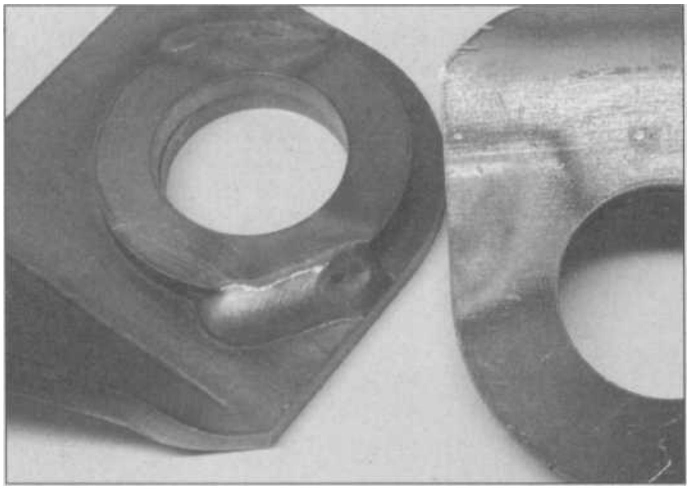

5.25 This close-up of a commercially made street rod part has a thick steel ring MIG-welded in two places to a heavy bracket for suspension mounting. This is a large bead from a good-sized machine. Note how smooth and clean the weld is, the bead looks almost like one piece, rather than overlapping puddles, and look at the backside of the part (at right) to see the indications of penetration. A **MIG** is easy to learn, but it will takes lots of practice to make welds like this.

have an upper limit of 250 amps, which is a *lot,* and they offer a 50% duty-cycle even at this high amperage. These larger machines also offer more "bells and whistles" such as larger wire spools, the ability to handle more different wire sizes, wheels on the bottom of the machine as standard equipment and the capacity and bracketry to hold larger bottles of shielding gas. Designed as they are for bigger jobs, these medium-size machines can handle wire diameters from .023-inch to .0625-inch (1/16-inch), which is bigger than virtually any requirement you will have need for in your home shop unless you take up amateur bridge building. You will see these 250-amp MIG machines hard at work in many shops today, from muffler businesses to street rod shops, race-car shops and general steel fabricators.

Other options you may find on these mid-size machines are controls for timing various weld functions. If you were regularly spot-welding long seams in sheet metal, you could set one of these machines up with a special spot-welding nozzle which has a flat front face with cutouts on each side. You bring such a nozzle directly in contact with the metal and pull the trigger. An adjustable timer on the machine sets the weld time for a perfect spot-weld with little or no crown or buildup to grind off later. Also, you can set one up to do stitch-welding as well, where you join work with a series of spot welds that overlap. The welding machine has a stitch timer that turns the arc on, pauses, arcs, pauses, arcs, etc. This stitch mode isn't used very often but is helpful in joining very thin materials where there is a chance of burring through with continuous welding. In the stitch mode (in which the pause time is variable at the machine), the pause time allows the first puddle to solidify before the next puddle is made, reducing the overall heat.

Mid-size MIG machines also offer more range of heat settings than the entry-level MIGs. While the smaller machines may have four settings from their lowest to their highest amperage, their bigger brothers may also have a "fine-power" adjustment knob, which essentially gives four more positions for each of the four basic settings on the main amperage knob. So instead of four settings, you now have a possible sixteen.

Adapters and special attachments offered on some machines include a metal shrinker and a stud-welding attachment. The metal shrinker works to reduce high spots in sheet metal, much as you would do with an oxy-acetylene torch, but without the flames and with very controlled heat. The basic principle is to heat the center of the high spot (such as from having hammered out a dent from the backside) to cherry-red, then quenching it with a wet rag, which "shrinks" down the high spot of stretched metal: The MIG shrinker attachment screws on in place of your contact tube, the wire is pulled back inside the torch (the wire isn't used) and the wire drive and gas flow are disconnected. The shrinking tip is placed in contact with the metal and the trigger pulled briefly, heating the high spot.

The stud-welding attachment is another body-working tool that allows you to "spot-weld" small studs onto a dented panel. These studs fit a special slide-hammer that pulls on the studs to work the dent out without having to drill holes in the body and use a screw-type puller as has been traditional. The studs can be clipped and ground off easily after the pulling is done. There is also a separate stud-welding tool that operates on 110V current without the need for the welding machine (see the Safety and Shop Equipment chapter). Another stud-welding attachment is made to use a threaded stud as an "electrode" and fuse it in place of a worn-out stud in automotive work **(see illustrations).**

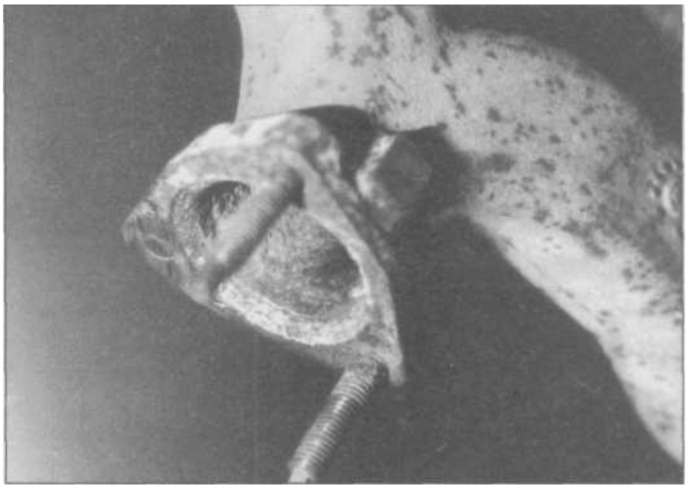

5.26 In exhaust system work, muffler shops deal with broken and rusted-away manifold studs all the time. HTP has a special feature to their MIG 160 that repairs such problems. First the old stud is cut off flush.

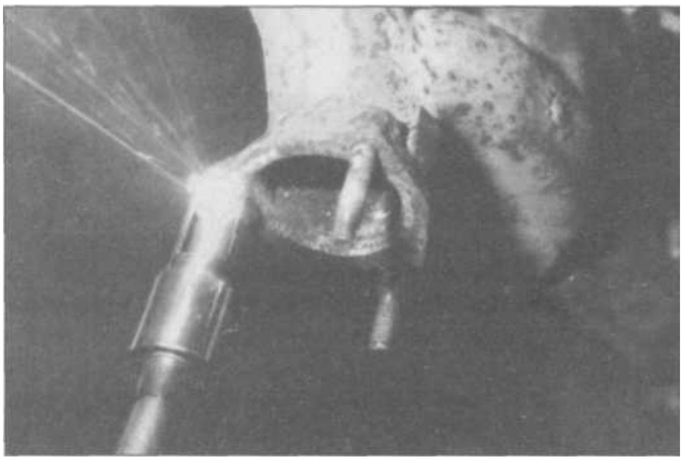

5.27 A special attachment to the MIG gun holds a hollow replacement stud. When the arc is turned on for about four seconds the wire feeds through the stud and welds it to the old stub.

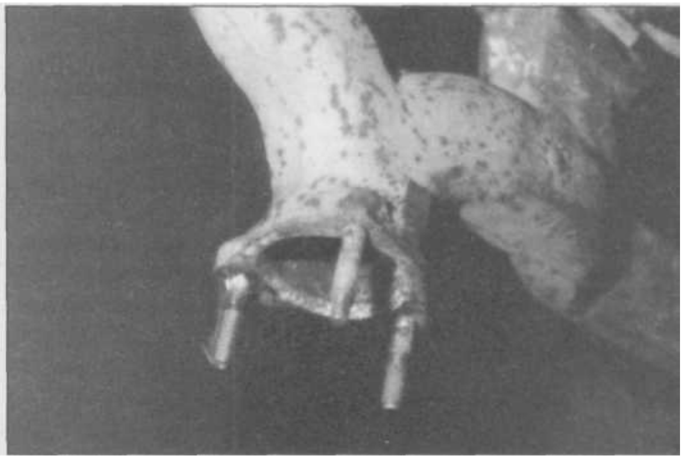

5.28 The new stud is now welded in place, with strength said to be up to 85% as good as original. This feature is a real time-saver for busy muffler businesses.

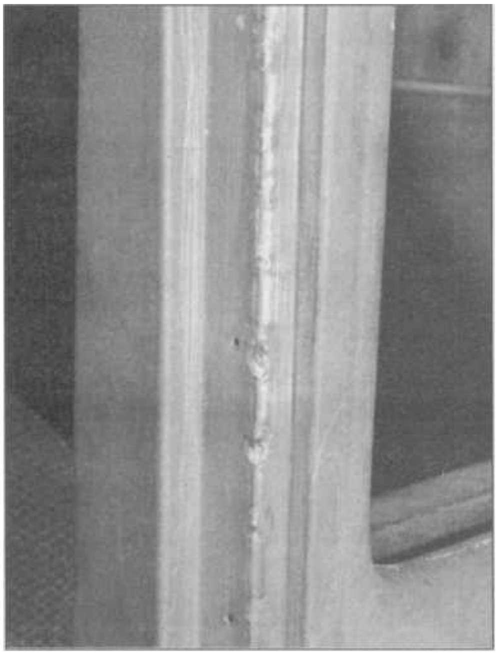

5.29 Spot-welding of sheet-metal panels is easy with a MIG, even a small one. For large panels, welds may be made through holes in the top panel. Here a new door jamb has been added to a 1932 Ford with a series of small spot welds. So little weld is above the surface that no strength is lost when the area is ground for painting.

To go to the top level in MIG machinery, you are now talking about very expensive professional welding machines with large, multi-purpose power supplies. As mentioned in Chapter 2, the bigger welding machines designed strictly for professional shops are often multi-purpose power supplies capable of high amperages which can be set up for Arc welding, TIG and MIG, with any polarity required. The amateur user will never need the high amperages or high duty-cycles of these top-level machines. To use the big machines for MIG welding usually requires that a separate wire-feeding accessory drive be set up, and, with the lower prices of today's MIG market, the big shops prefer to just buy a separate MIG welder for those needs, using the large power supply just for TIG welding.

Choosing shielding gas

By now, you should have an idea of what level of MIG machine will suit your purposes. If you are like most home/shop users who aren't going to be welding every day, your needs may be best served by a typical, entry-level 140-amp, 110V MIG machine using shielding gas. You now have to choose a shielding-gas bottle (not included with most MIG machines), some wire (a small sample roll may be included with your machine), and the type of shielding gas you want to use. This sounds like a lot of choices, but the field is narrowed down by what kind of work you want to do and how often you need to do it. In all cases, the choice of shielding gas and wire type must be matched to the kind of material you will be welding.

Capacity of the gas bottle is important, and most first-time weldors make the mistake of buying less capacity because the bottles are cheaper, only to find themselves someday working on a project over a weekend (when there is little

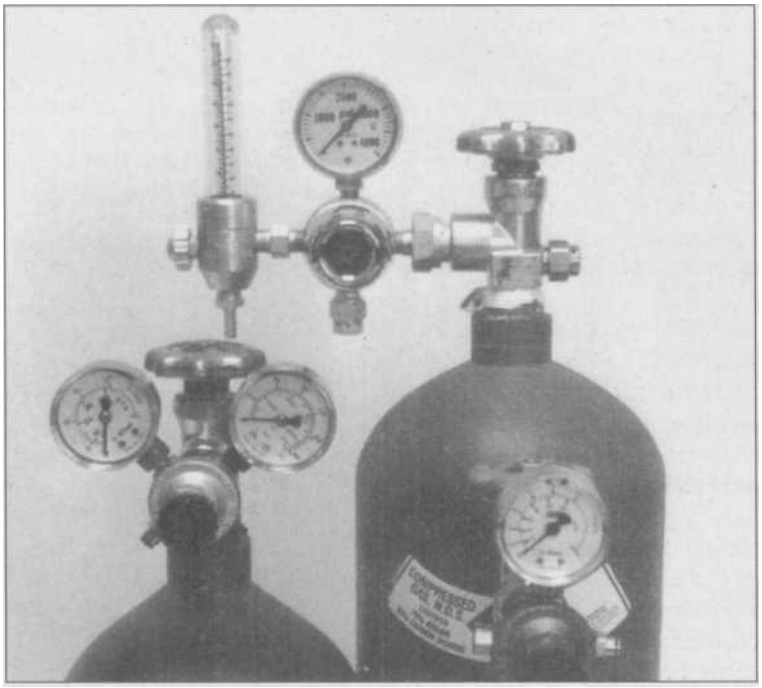

5.30 various size shielding-gas bottles are available for MIG setups. Check with your welding supply store on how long each size might last. Most home/shop users buy relatively small bottles because they want portability in a small garage area.

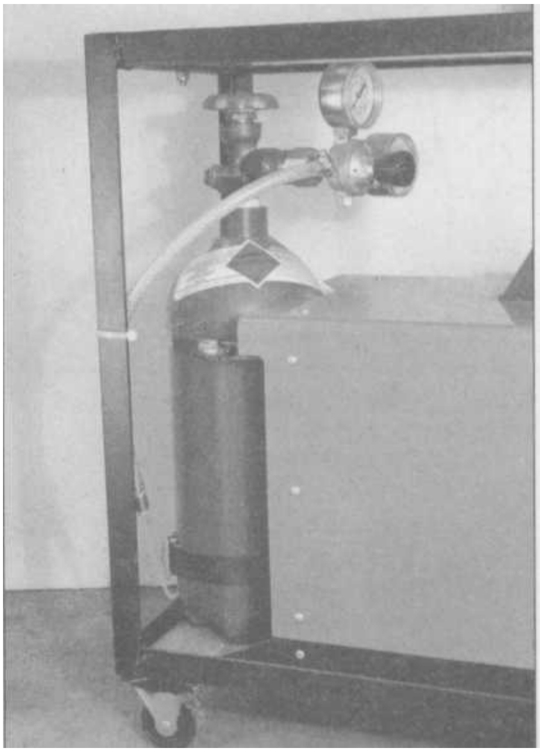

5.31 The smallest bottles are 20 cubic-foot capacity, and easily fit into even a small welding cart like this. The bottle and regulator are totally protected here. This is a good size if you only use the welder infrequently, or on small projects.

5.32 Gas-shielded MIGs are not made for outdoor work, but as long as you can keep the wind currents away, they work fine. On this driveway utility-trailer project, pieces of plywood are used as wind shields while welding.

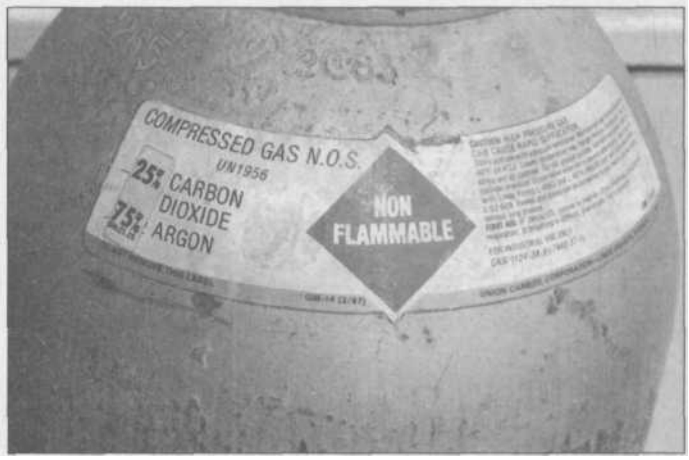

5.33 Most of your welding with shielding gas can be done best with a 75% argon/25% CO_2 mixture, which is excellent for cleanliness and penetration, and can be used on cast-iron and stainless-steel as well.

chance of getting a bottle refilled) and they are out of gas. Gas bottles are available in sizes from 20 cubic feet to 40, 60, 80 and 125 cubic feet of gas. This roughly translates into welding time of from four hours for the 20 cu.ft. bottles to 14-16 hours for the 125 cu.ft. when used on a small machine like the 140-amp unit in our example. There's a number of factors involved in how long a particular bottle may last under different welding conditions, so these are just rough guidelines. Every time you use your MIG machine, you should look at your tank gauge to see how much more gas is left. You may even want to just lease a small cylinder of gas when your first get your machine just to see how long such a tank would last you, and then buy a bigger one if you think you need it.

The small bottles fit nicely under a small welding cart, making for a very portable arrangement for occasional home use **(see illustration).** The larger bottles are heavy and bulky, so you'll have to build a welding cart (most first-time weldors' first project) that can accommodate one. Of course, bottles of compressed gas are dangerous, even if the gas is non-flammable, due to the pressures involved, and any welding gas bottle must be securely mounted with chains or sturdy clamps. It's important to note that welding bottles for electric welding machines should be ideally mounted in an insulated manner, such as with rubber underneath and rubber hose over any restraining chains. This prevents the remote possibility of arcing against the cylinder, which could damage it.

Next there is the choice of shielding gas to use. One of the widely used shielding gasses for MIG work is plain carbon-dioxide, or CO_2. While it is in the bottle, CO_2 is really an inert gas, but technically it turns into carbon monoxide and oxygen under the intense heat of the arc process. The oxygen can combine with other elements in the air and in the parent metal to form undesirable oxides. Thus, plain CO_2 will not usually result in as clean a weld as with other gasses, but it is used often because it is the least expensive shielding gas for welding steel. It will weld very fast, with good penetration, but you should use a welding wire that contains deoxidizing elements to counteract the effect of pure CO_2.

Argon is a versatile inert shielding gas that is often used by itself, or in combination with other gasses to produce certain weld characteristics. For instance, welding of non-ferrous materials like aluminum is usually done with pure argon. It makes for good penetration patterns and a concentrated arc.

To weld ferrous materials, like the mild steel we will use for most home/shop projects, argon is usually mixed with another gas, or sometimes two other gasses, to provide special characteristics. While there are some esoteric mixtures of gasses in different percentages for specific purposes, a mixture of 75% argon and 25% CO_2 has become pretty much the standard for welding mild steel with short-arc MIG machines with wire diameters of .035-inch or less (see illustration). This mixture, often abbreviated as C-25, is more expensive than the plain CO_2 but produces much less spatter and consistently better looking welds, even on materials that exhibit minor rust or scale. It is what we would recommend for most of your welding needs. If you were doing a lot of thicker materials, overhead or other out-of-position welding and pipe welding, you might use a 50-50 mix of argon and CO_2 which offers good wetting and bead shape without the excessive fluidity that causes a bead to droop or fall when doing out-of-position welding.

If you are working in a body shop, you might be interested in a mixture that Airco calls Argoshield LG®, which stands for light gauge. It consists of mostly argon with small additions of CO_2 and oxygen. The combination is designed for metals down to 20-gauge, and produces good penetration, but with a smaller weld bead than normal for less sheet-metal finishing, less spatter and smoke and good arc starting. It is specifically formulated to work well on the thin, low-alloy steels now used on auto sheet metal.

For welding of non-ferrous materials, either straight argon or mixtures of argon and helium are used in various combinations which provide higher heat to a MIG arc. Usually, the thicker the material to be welded, the higher the percentage of helium that is included in the mix, and the HE-75 gas, which is 25% argon and 75% helium is typically used in industry to weld thick aluminum.

If you were interested in welding stainless-steel, you can actually use the C-25 gas we recommended above for mild steel, but a mixture of 90% helium, 7.5% argon and 2.5% CO_2 is widely used in stainless-steel MIG welding because it offers a higher heat for the normally sluggish weld puddle on stainless, as well as offering good stability, penetration and resistance to corrosion.

If you have to pick just one type of gas, the C-25 works great on steel, can be used on stainless-steel and is also useable on cast iron.

Choosing wire

The bare steel wires used in a bottled-shielding-gas MIG machine are all variations of the same basic metal, with varying amounts of deoxidizers added for different applications and shielding gasses. The most common deoxidizer is silicon. Others added may be manganese, aluminum, titanium, or zirconium, with nickel, chromium or molybdenum added to improve mechanical or corrosion-resistance properties. The welding wires with higher levels of deoxidizers generally are better able to weld on rusty or unclean steel surfaces.

Most of the steel wires for MIG welding fall under a designation from the American Welding Society (AWS) of E70S-something, with the last digit being the particular variation. For instance, one of the most common wire specifications is E70S-2, which is deoxidized, and makes a good weld with C-25 gas, even on rusty steel. Its main drawback is that it lacks fluidity; the puddle doesn't want to flow out width-wise and may not stick well in heavy materials.

The E70S-3 wire is one of the most common and least-expensive MIG wires available, with more deoxidizers and a more fluid puddle that makes a wider bead. It has been used successfully for years on cars, farm equipment and appliances.

Next up in terms of deoxidizer content is E70S-4, which is a medium-priced wire suitable for almost all steel welding. It offers good fluidity and better arc characteristics than E70S-3, but has slightly more spatter, and is used on structural steel, ships, piping and boiler vessels.

Probably the high-performance and higher-cost wire available is the E70S-6, which has the highest level of silicon and manganese as deoxidizers. It is suitable for welding almost all steels, from thin mild steel to 1/2-inch plate (in the appro-

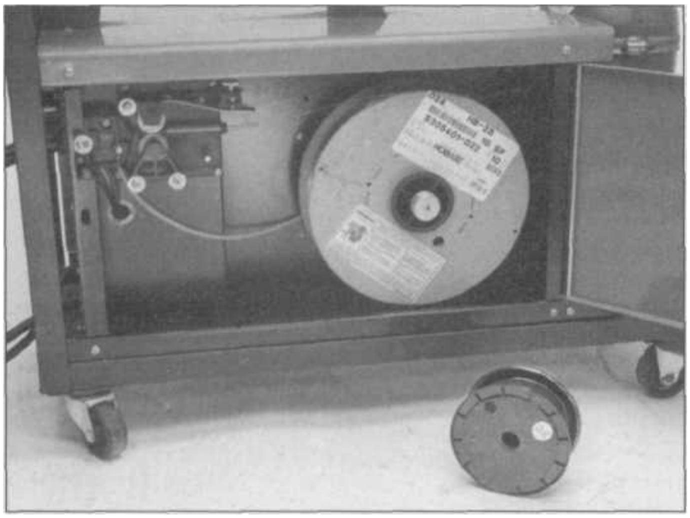

5.34 Inside the wire-drive portion of a typical small MIG machine, this machine comes with a small wire roll such as at lower right, but comes with an adapter to handle the larger rolls such as in the machine. The small wire rolls are inconvenient unless you do only small projects.

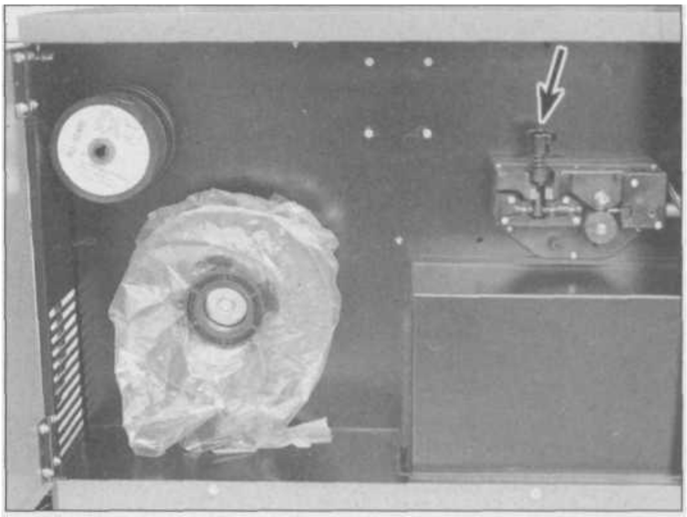

5.35 This is the drive section of a Daytona MIG wire-feed, which has a place inside to store an extra small roll of wire, good for keeping the spare clean if you occasionally switch to another type of wire for specific needs. Arrow indicates the knob for adjusting the wire roller tension.

priate wire diameter) and works with the most popular gas mixtures, offering good puddle fluidity, too **(see illustration).** It is one of the few MIG wires that is completely compatible with today's thin, low-alloy, high-strength body metal in unibody cars.

L-Tec markets a wire meeting the AWS classification for ER70S-2, 3,4 and 6 that they call Easy Grind, which is designed specifically for auto body work. It is said to require less skill to weld thin materials, and to have a reduced carbon content that means easier grinding to prepare for filler and painting. Besides automobiles, it is also used in welding ductwork, metal cabinets, lawn furniture, and is designed to be on metal as thin as 26-gauge with smaller MIG machines.

Non-ferrous metals, of course, require different wires. The basic rule of welding electrode selection is to choose a wire or electrode whose content is much the same as the parent metal you are welding. In the case of stainless-steel, it *can* be welded with the steel wire we have recommended above, but for better corrosion resistance (the main reason for making something from stainless-steel in the first place) stainless-steel wire should be used. There are various alloys of stainless-steel, but one of the more common varieties is type 304, and the best wire for MIG welding type 304 is called ER308, with variations of increasing silicone content in ER308L and ER308LS.

For aluminum welding, you need aluminum wire. Again, there are many alloy variations of aluminum, and several metal blends in aluminum wire, but perhaps the best MIG wire for aluminum is called ER5356, which has a chemical content proven to work in most situations. Aluminum welding is more difficult than steel in many ways, and often the exact alloy of the parent metal is impossible to determine, unless you have bought a large enough piece at a metal yard that is still factory marked as to its alloy and heat-treat. There are even some aluminum alloys which are considered unweldable, like high-strength aluminum such as 7075-T6.

5.36 Your welding supply store will have a variety of MIG electrode wires to choose from. The E70S-6 wire at left is one of the most popular. At right is L-TECs Easy-Grind for body work.

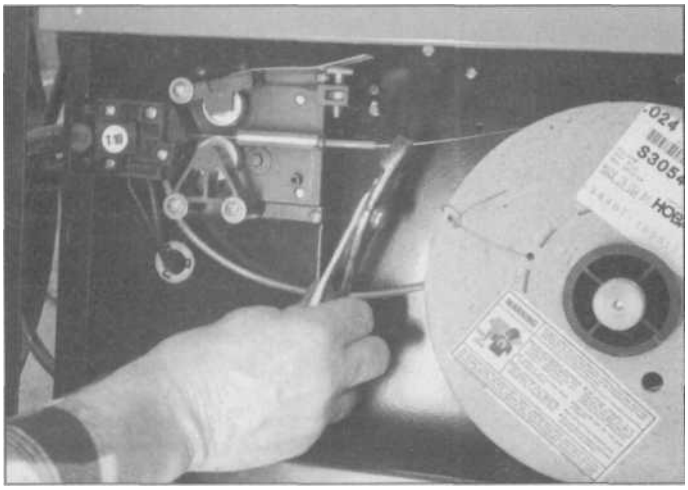

5.37 When starting in a new roll of wire, keep a pair of pliers on the end you take off the reel, or the wire can unroll everywhere. Feed the wire through the guide and over the drive roller.

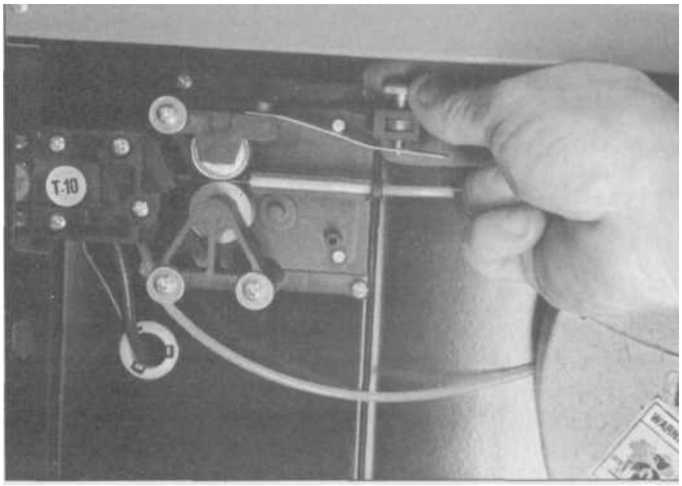

5.38 The wire tension is adjusted with a knob, and proper tension avoids wire-twisting problems inside the gun. The wire should be able to slip when it needs to. Most machines have a feature that allows you to retain the tension setting, once set, by swinging the tension arm away when changing wires.

Learning MIG welding

One of the attractions of MIG welding that has brought these machines into the home/shop market is how easy they are to learn. With a little instruction, virtually anyone can be MIG welding with an hour's practice time. As with anything else, the more you practice, the easier it will be and the better your welds, but before you get overconfident about your MIG welds, try cutting apart a seam you have made to check the penetration, and make the other tests (bending two butt-welded sample plates in a vise) we have previously described to see if the welds you're making are as strong as they are good-looking.

To set up your MIG machine, the first step would be to read all the directions and cautions in the instruction book that came with it. One of the few adjustments you will have to make initially is to put a roll of wire into the machine and set the drive-roller tension. The wire unrolls from the reel through a guide and over a motorized roller, which feeds the wire through another guide and into the cable going to your torch. Lay your torch and cable out on the floor as straight as possible. Mount the wire spool into the machine as per your directions, but be careful when cutting the end of the wire which is usually bent over to lock it into a hole in the side of the reel. There is winding tension on the wire and if the cut end of the wire gets away from you, it can start unraveling all over the place. Don't cut the end loose until you are ready to feed it.

Hold the cut end with some pliers or vise-grips and carefully file or sand the cut end until it is smooth. A sharply-cut end may snag as it travels through the cable or torch. Insert the end into the guide, loosen and swing away the wire-tension adjuster, and feed the wire over the groove in the drive roller and on into the next guide, still holding the wire firmly **(see illustration).** Most machines are set up to handle several sizes of wire, but you must be certain that you use the drive-roller groove that is the right size for the wire you are using. Most rollers have two sizes on them and can be turned around on their shaft to put the other size in line with the wire. There will probably be an extra roller with other sizes on it that comes with your machine.

When six inches or more of wire have been fed into the welding cable, put the wire-tension adjuster back in place over the wire, keeping the wire centered in the groove. Turn on the welder and put the wire-feed speed about one-quarter

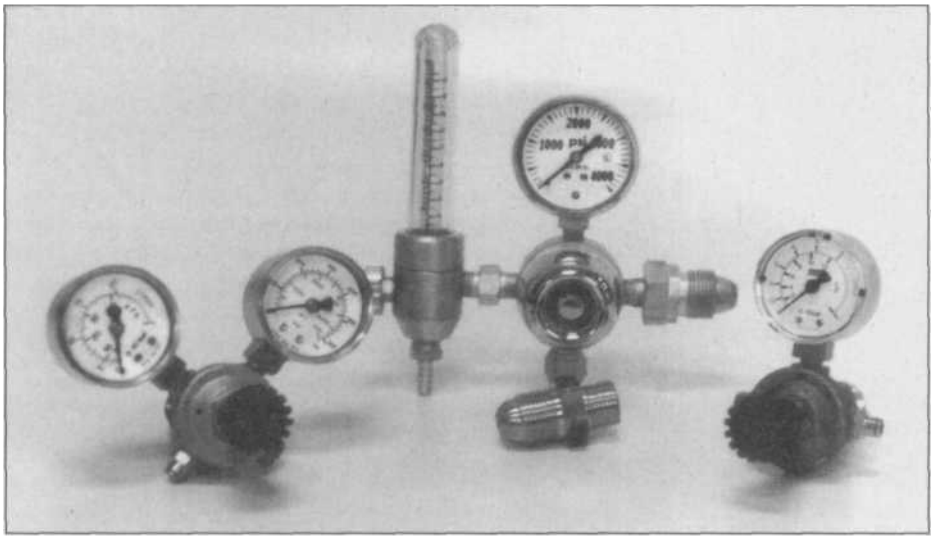

5.39 There are three choices in regulators for MIG shielding gas. The unit at right is a single-gauge, economy unit which indicates only the line flow, not the total tank pressure, at left is a dual-gauge unit that shows both, and in the center is the more professional model that has a floating-ball flow meter easy to read from a distance.

of the way and pull the trigger. At this point, adjust the pressure on the wire tensioner until the rollers are feeding the wire, and then turn it 1/4-turn more. With the nozzle and contact tube unscrewed from the end of your torch, keep the torch cable straight and run the machine until the wire comes out of the torch. Slip the correct-size contact tube over the wire, screw it in and attach the nozzle.

The final check on wire tension is to make sure the wire can "slip" if necessary. Eventually in your MIG practice, you will virtually weld the wire to the contact tube. When this happens, and you are still pulling the trigger, the machine tries to feed more wire but it can't come out of the gun, so it jams up inside in what is called a "birds nest." To prevent this, bend the wire to the side of your torch and put the nozzle right against the shop floor, simulating a stuck wire. While watching the drive mechanism inside the machine, pull the trigger. The drive rollers should have just enough tension to always pull the wire through, but not enough to prevent slippage in case of a jam. If your rollers *do* slip in this test, you're set.

If yours is a flux-cored-wire machine, you are ready to weld, but for the shielding-gas machines, you have to set up the bottle. With the bottle mounted firmly on your cart, attach the regulator to the bottle and attach the reinforced hose (should come with your machine) to the regulator with a hose clamp, connecting the gas to the machine. On some machines, a special adapter connects the reinforced hose to a small plastic hose going into the machine. Keep the two regulator gauges protected from impact, and always keep the gas valve turned off when you are not using the machine. Some cylinder valves only seal well when the valve is either fully off or fully on, so when your machine is on, turn the cylinder valve all the way, not just a few turns. You can check any of your gas connections for leaks using some soapy water solution. Bubbles indicate leaks. Adjust your gas regulator to the proper flow rate (cubic-feet-per-hour rate, according to your instructions) for the material and wire you are working on. **Note:** *One gauge tells you the flow rate and the other tells you how much pressure is still in the bottle. You can't read either one unless the cylinder valve is turned on, so don't look at your bottle one day and notice that the tank gauge looks empty - turn on the valve to really see what's left.*

Now you're ready to practice. Select some 1/8-inch steel, set your amperage according to the suggestions in your instructions, and set your wire speed a little

0
1
2
3
4
CHICAGO
ELECTRIC
welding
systems

5.40 The basic MIG units have only a few power settings. Your instruction sheet will tell you the recommended thickness of metal for each setting, and a chart on the front of the machine usually tells you the amperage and duty-cycle of the settings. After some practice on different materials, you will learn instinctively which settings work best for your work.

5.41 The wire speed has a lot to do with the heat and quality of the weld. Generally when you have it right, the MIG machine makes a very satisfying steady "sizzle" sound. Practice by adjusting the feed high and gradually lowering it until the sound and results are satisfactory. If you do repetitive projects, write down the amperage and wire speed settings that have worked for you.

higher than what's recommended. The wire speed is key to getting a good weld with proper heat and penetration, and tuning the speed is easy. Listen to the sound the welder is making as you make a straight bead. If the speed is too fast, most of the wire coming out will be red-hot and there will be a loud crackling noise, gradually adjust the speed down until you get a steady sizzling sound. The wire should be melting right at the weld. Practice tuning in the wire speed at different power settings and metal thicknesses.

There are two torch-holding techniques for MIG welding, forehand and backhand **(see illustration).** The first is when the torch is angled and moved such that the weld is taking place ahead of the torch's direction of travel, and the backhand is when the weld takes place behind the torch **(see illustrations).** Most beginner find the backhand (also called "dragging" the weld) technique works best on most steel-welding projects, giving the weldor a better view of the puddling action. In either style, the torch is angled about 35 degrees to the work. If you find yourself doing a vertical seam, you can travel either up or down, but use the forehand technique (sometimes called "pushing" the weld) only when going up, and the backhand technique going down, keeping the arc on the puddle's leading edge at all times for good penetration.

After making good straight beads on top of the plate, try joining two plates. If you are working with materials under 1/8-inch thick, most steels can be MIG-welded without beveling the edges, but you should tack the parts together with a slight gap between them before finishing off the seam with a full weld. This assures good penetration. On 1/8-inch and thicker metals, the edges should be beveled before tacking them. If you find that you are consistently burning through the material, the heat setting is too high, or you are moving too slowly. If it is "cold" appearing and not penetrating, the amperage may be too low, or you may be traveling too fast. You will adjust your travel speed with the torch by watching how the puddle develops, moving the torch along at a slow, *steady* rate. One of the great advantages of the MIG process is that your torch is right down close to the work, and there is much less intense light, heat and spatter as with arc-welding; you can get your head closer to the work to really see the puddle progress. Keep a pair of sharp wire-cutting pliers on your welder at all times. You will use it when you start each time to maintain about 1/2 inch of wire sticking out of the

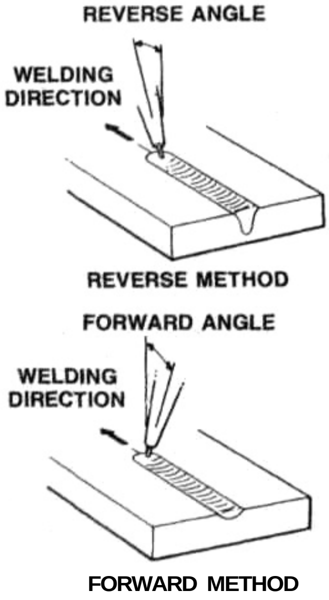

Gun *tip inclination and the direction of the weld can be changed to match different welding conditions*

5.42 Of the two methods of gun travel with MIG welding, beginners often like to use the reverse or backhand method because they can watch the puddling action much better.

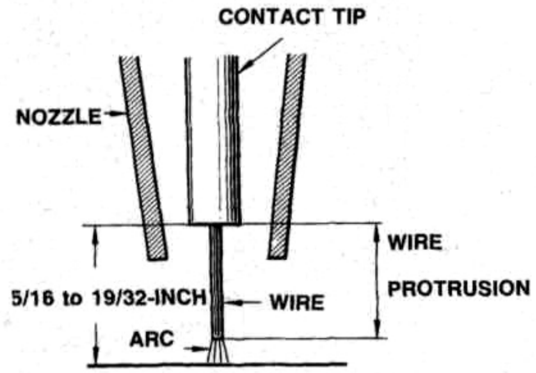

5.43 The tip-to-work distance is very important in maintaining the short-circuit arc. When practicing, watch the arc closely. The white-hot, melting end of the wire should stay about the same short distance from the puddle. Wire speed and welding travel speed control this.

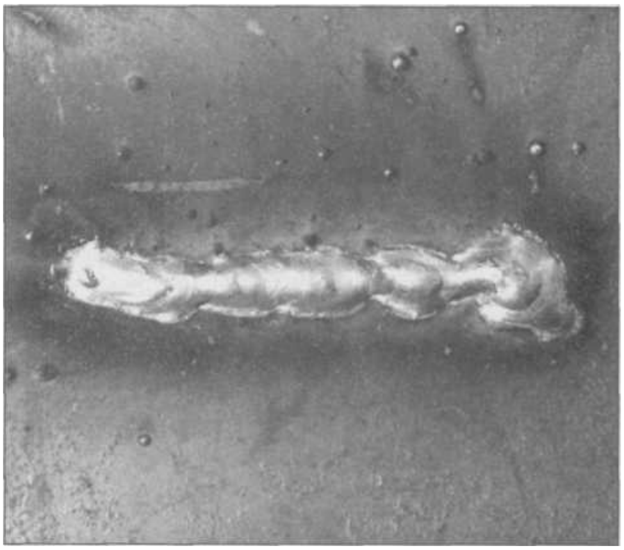

5.44 This is what some of your practice welds will look like. The heat setting was too hot here, and the weldor wasn't very steady with the gun. Use both hands and find a very comfortable position to hold it steady. Note the amount of spatter from making and breaking the arc.

5.45 This production weld by a professional shows how clean and perfect MIG welds should look. The heat-affected area on either side is small, the weld has good penetration, and there is minimal spatter. Heat was evenly played on both pieces of tubing.

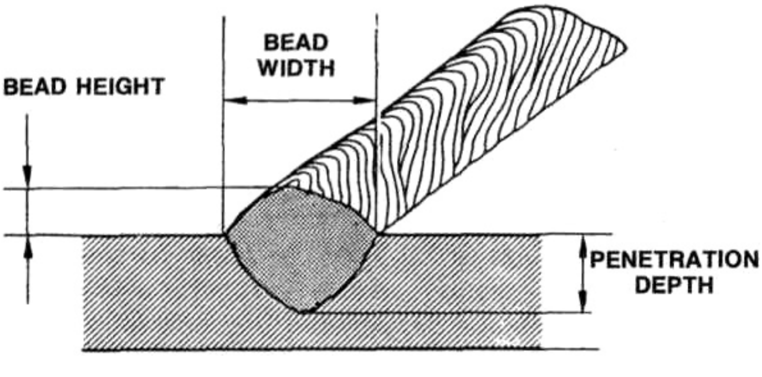

5.46 Welding current affects the penetration of the weld and the size of the bead. Ideally, you are looking for good penetration without a too-tall or too-wide bead.

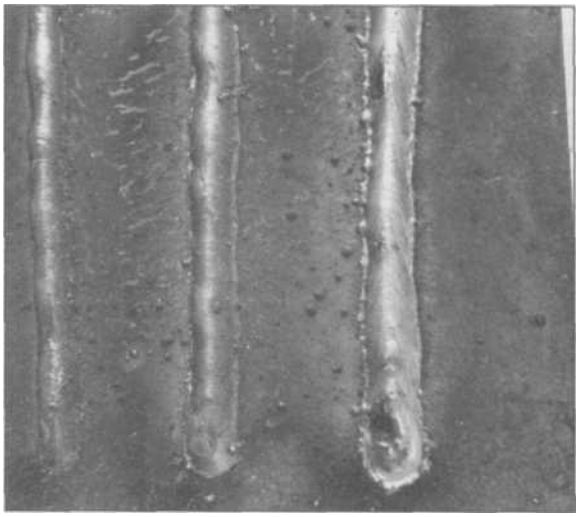

5.47 At left here is a MIG bead that was too low in amperage for the thickness of the material. Penetration is shallow and the bead sticks up too high. At center, the correct amperage was used, and at right, the amps were too high. The right-hand bead may look OK here, but penetration was too deep (dropthrough on the backside) and there was some undercutting on either side of the bead.

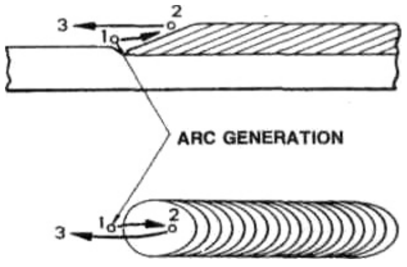
3
2
1
ARC GENERATION
3
1
2

OVERLAPPING BEADS

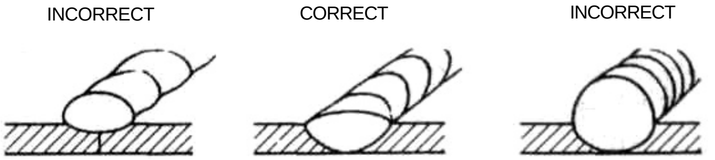
INCORRECT
CORRECT
INCORRECT

Insufficient penetration. Weld strength is poor and the panel could separate when it is finished with a grinder.

Good penetration and easy to grind.

There js good penetration but finish grinding will be both difficult and time consuming.

5.48 After you have practiced sufficiently, you'll develop a natural rhythm for the torch movements to create a good bead of overlapping oval puddles.

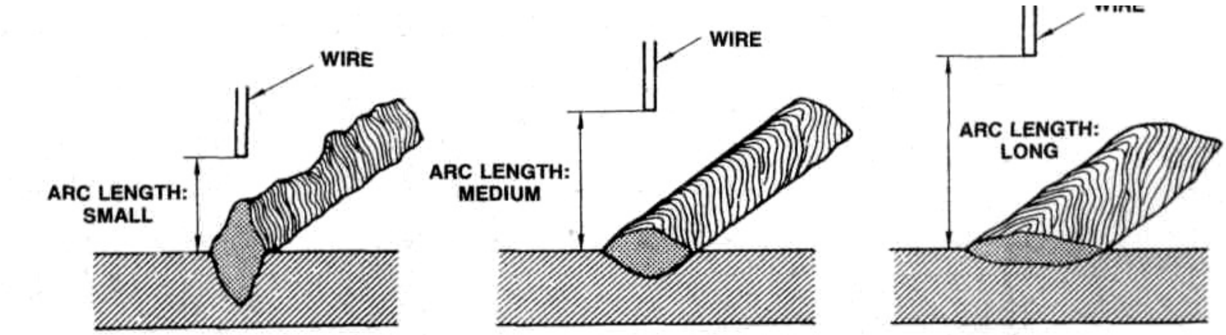
WIRE
ARC LENGTH: SMALL
WIRE
ARC LENGTH: MEDIUM
WIRE
ARC LENGTH: LONG

ARC VOLTAGE: LOW

ARC VOLTAGE: MEDIUM

ARC VOLTAGE: HIGH

5.49 Arc voltage and length affect the shape of the bead's profile and its penetration. Cut a few of your practice welds in half to examine the beads. The shape of the overlapping puddles should have rounded ends exposed, not sharp, triangle ends.

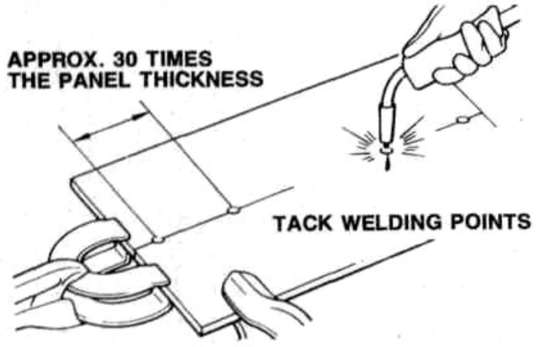
APPROX. 30 TIMES THE PANEL THICKNESS
TACK WELDING POINTS

5.50 The thickness of the panel you are working on determines how far apart your initial tack-welds should be to firmly align your parts. On thin materials, you need a lot more tacks.

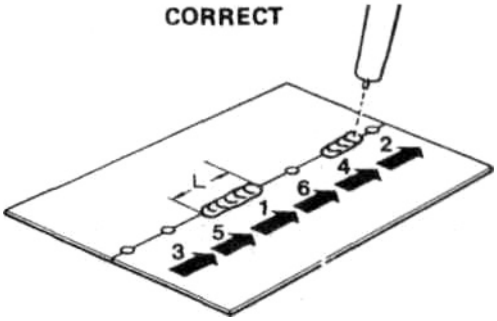
CORRECT
L
1
2
3
4
5
6

5.51 When working on large panels of thin material, such as when doing body work like welding on a new quarter-panel, warpage is a problem unless you keep the heat down by making lots of tack welds, then switching back and forth to the coolest section to add a short weld, then move to another section, and switch back and forth until the whole seam is done. A continuous weld might put too much heat in, and warp the new panel, requiring lots of hammer-work to straighten.

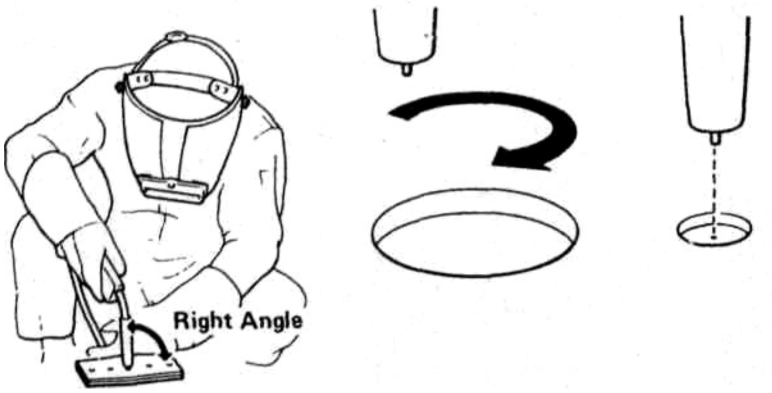

5.52 Instead of a straight butt-weld, panels of thin material are often joined with spot welds, in which the top panel is punched with small holes. You MIG-weld around the hole, joining the top panel to the bottom.

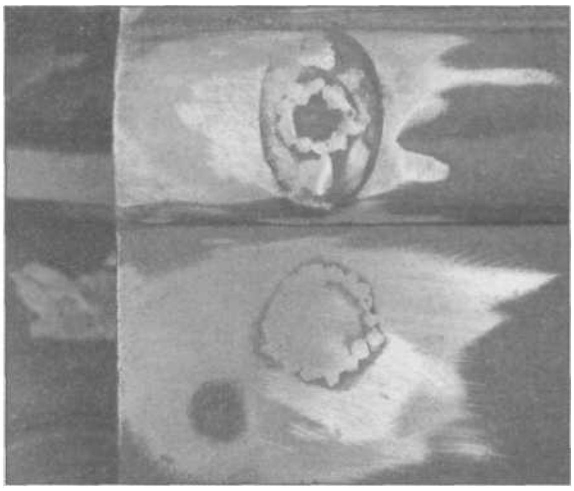

5.53 Door-handle and lock holes have been filled in on this street rod door, using small circles of sheet metal the same thickness as the original metal. In order to avoid warping, the circles were filled by tacking the filler pieces in a few places, waiting for the part to cool off, then adding a few more tacks, until the whole circle was made of tack welds.

gun. If you are welding thicker materials, or two edges that have irregular gaps, you can weave the arc back and forth from side to side as you travel, bridging the distance between the two edges.

Sometimes a start-and-stop technique works well for bridging gaps or filling holes. You watch the arc and weave from one side to the other, then pause (let go of the trigger) for a second with the torch in the same spot, allowing the puddle to solidify, then pull the trigger again, building a new puddle onto the last one. To fill a hole, just keep making tack-welds around the edge of the hole until you have made it smaller **(see illustration).** Let the metal cool, then add an inner row of tacks inside the first "ring" until you close up the hole with filler metal. This is often done in filling sheet-metal holes in body work, but any hole larger than a dime should be filled by welding in a circle of new metal of the same thickness.

The beauty of the MIG process is that the torch is a one-hand operation, which allows you to use your other hand to steady your primary hand and draw a smooth bead. Also the torch stays the same distance from the weld at all times, unlike arc welding where the rod is always getting shorter. MIG welding doesn't require stopping to change electrodes either, so you can concentrate nicely on what you are doing. The one-handedness of MIG welding tempts some weldors to operate the torch with one hand while they actually hold something on the work with their other hand. This is OK for tack-welding, where you have a body panel you need to hold in alignment while you tack, but better seams are made when you hold the torch with both hands. Some users get used to the lower brightness and spatter of the MIG process, and are tempted to tack parts together without using a helmet. Body men in particular are prone to hold two panels in alignment, put the torch where it needs to be, close their eyes and make a tack weld for a second. Although MIG welding does produce less intense light and spatter than arc welding, you are still subject to both radiation burns and some spatter. Don't ever take a chance with your eyesight. One blob of spatter onto your bare eyelid could be really dangerous! A much better procedure (and the only one we recommend) would be to use some clamps or other devices (see the Safety and Shop Equipment chap-

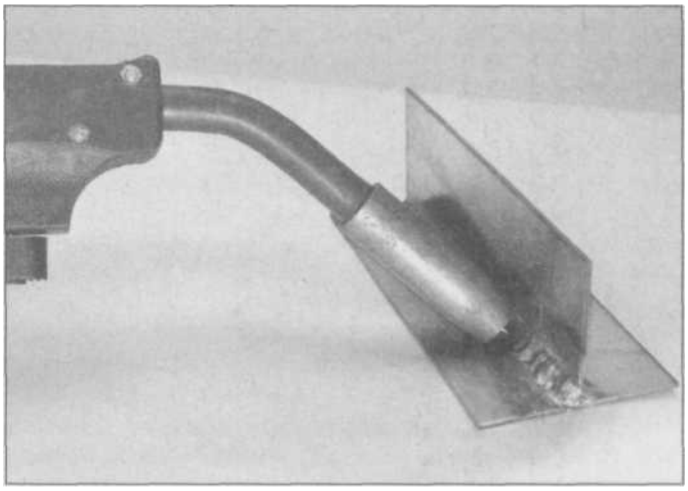

5.54 When welding a corner joint such as a T-section, angle the gun at about a 45-degree angle to aim equal heat and filler on both pieces.

5.55 Don' be tempted by the low spatter of MIG welding to just shut your eyes when tacking body panels. If you need speed and convenience while tacking, use a hand-held safety-shield with lens as shown here. Weldor here has long sleeves, but should have gloves on and a turned-around hat to protect his hair from hot spatter.

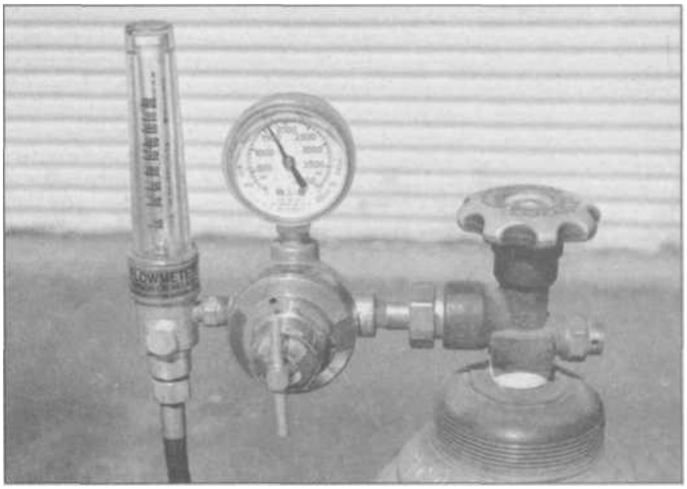

5.56 The best argon flow meters for professional use are the type with a floating ball inside a glass tube. Every time you pull the trigger, the ball floats up on the scale to indicate the gas flow to the machine. At a normal working distance from the machine, the weldor can more easily read this than a dial gauge. It makes it easy to see at a glance if you forgot to turn on the gas at all.

ter) to hold your panels in alignment, and use a helmet. If you don't like the bother of flipping a big helmet up and down to just make a series of quick tack-welds, then use a hand-held face shield with lens **(see illustration).** It's very easy to hold the shield in your left hand while tacking with your right, and it goes instantly from shielded to an unshielded view of the work.

If you are welding along and all of a sudden the bead goes bad, there are several things to consider. If you have been steady with your arc distance and travel speed, then perhaps a draft temporarily blew the shielding gas away from the weld. If it's not a draft, look to your main torch cable. If there are any bends or kinks in it, in the position in which you hold it when welding, the flow of gas can be restricted easily. Watch your gas bottle gauges while you pull the trigger for a second to see if the flow rate has changed or if the gas bottle is running low. The better equipped MIG machines in most shops have a type of regulator with a flow meter to indicate the flow rate **(see illustration).** It consists of a glass tube with a colored ball inside that floats during gas flow. The numbered line it floats to indicates the flow rate. The advantage of these more expensive regulator/meter setups is that the weldor can see the gas flow from where he is *at the work.* He doesn't have to go back to the machine to read a gauge dial.

Aluminum welding will be more of a challenge than steel, regardless of the welding method used. Aluminum really soaks up heat, and you have to use thicker wire, and more heat, wire and gas flow to weld it. Equipping your machine to weld aluminum means: setting the gas flow almost twice as high as for steel, using aluminum wire, using the right gas mix, replacing the wire liner in the torch with a Teflon liner and using a contact tip that is one wire-size bigger than the aluminum wire. The reason for all this is that the aluminum wire tends to grow more with heat than the steel wire. The wire tension needs to be fairly light, because bird-nesting is more commonplace with the less stiff aluminum wire. In fact, there are special MIG guns made for

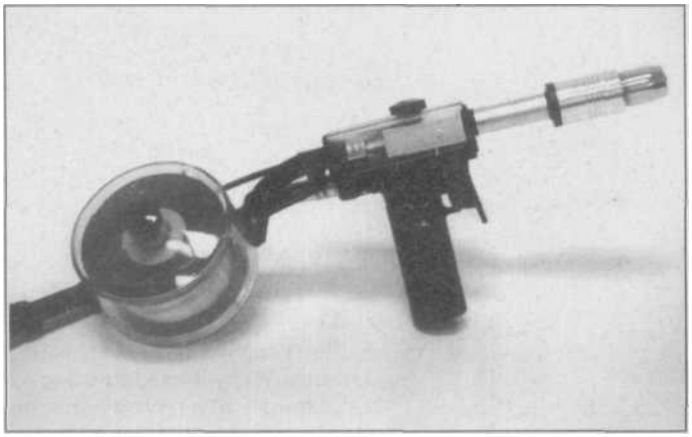

5.57 When MIG-welding aluminum, especially thin aluminum, the aluminum wire itself is fairly flexible, which can cause feeding problems inside the machine, because the drive rollers are *pushing* the soft wire a long distance. This is a spool gun, which holds a small roll of aluminum wire right at the gun for easy feeding.

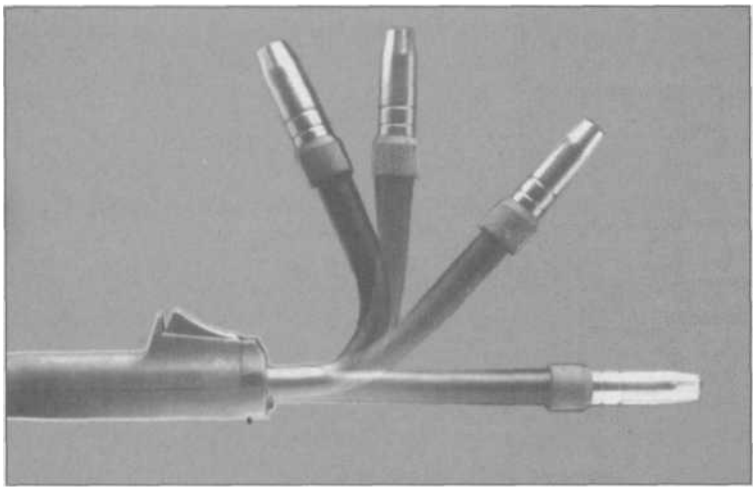

5.58 For tight quarters such as muffler work and other round-tubing situations, HTP makes this MIG gun with a flexible swan neck. It can even be used straight ahead, to reach into a spot where you can't get your hands.

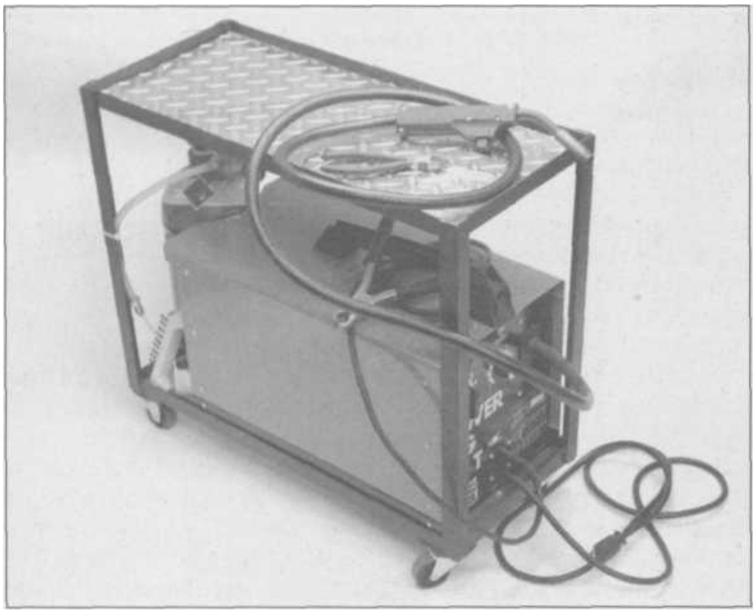

5.60 A handy welding cart for your MIG machine will probably be one of your first home/shop projects. This one is simply-built from one-inch angle-iron, fitted with a steel bottom shelf to hold the welder, bottle and accessories, and the top was made from a piece of polished-aluminum diamond plate purchased from a remnant pile at a scrap yard. Caster wheels were added quickly by welding their mounts to the bottom.

5.59 Muffler shops are finding that the MIG machine is faster and safer to operate under a car than the traditional oxy-acetylene torch equipment. There is no constant re-lighting and adjusting of the flame to worry about when initially tack-welding various pipe sections together.

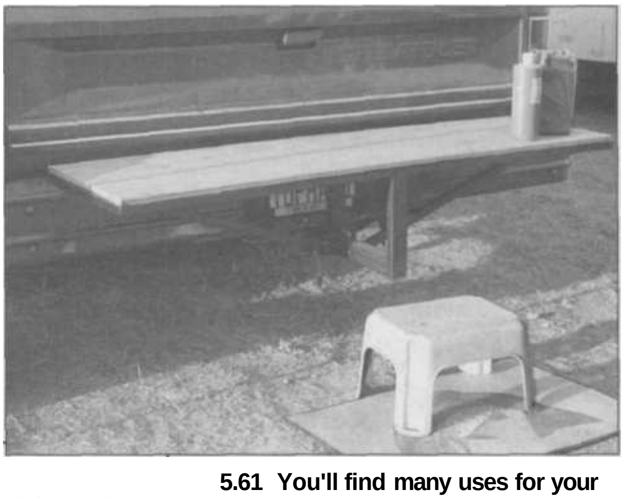

5.61 You'll find many uses for your welder once you get used to using it. This example of a typical home project is a plywood-covered rear bumper extension for a truck. It's used for carrying extra camping gear, and used as a tailgate for working or cooking when camping. Its also a wide step when stepping out of a camper shell on the truck.

production work with thin aluminum wire that hold a spool of wire right at the gun. Called spool guns, they eliminate the wire-feeding problems of trying to push a soft aluminum wire all the way through a cable to your torch **(see illustration).**

The wire speed should be faster than for an equivalent sized steel wire, and the forehand technique is recommend because it puts more preheat ahead of the puddle. There are two extra steps in welding aluminum. The metal must be scrupulously clean and preheated to 350° F to get a good weld. You should keep

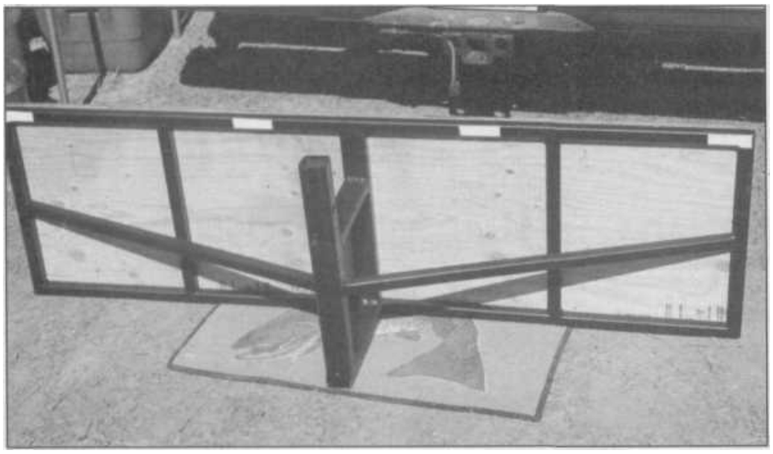

5.62 The camper-step was MIG-welded from one-inch steel tubing and braced to a 2x2-inch tube that fits into the truck's hitch receiver.

5.63 Street rod shops, customizers and restoration shops are using a **MIG** welder more and more, for its convenience and ability to handle delicate thin materials. This 1932 Ford door has had a new skin MIG-welded along the bottom where the original door was rotted away.

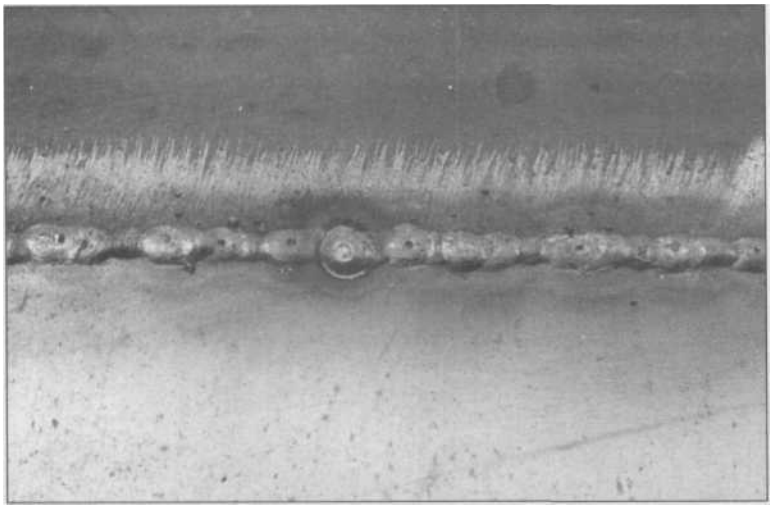

5.64 The door bottom was welded on with a series of very short welds, alternating from one area to another to keep from putting too much heat at one time into an area. Eventually, the whole seam is welded. L-TEC Easy-Grind deoxidized **MIG** wire is perfect for these kinds of jobs, because it grinds off easier, without inducing too much heat in preparing the seam for finishing and paint.

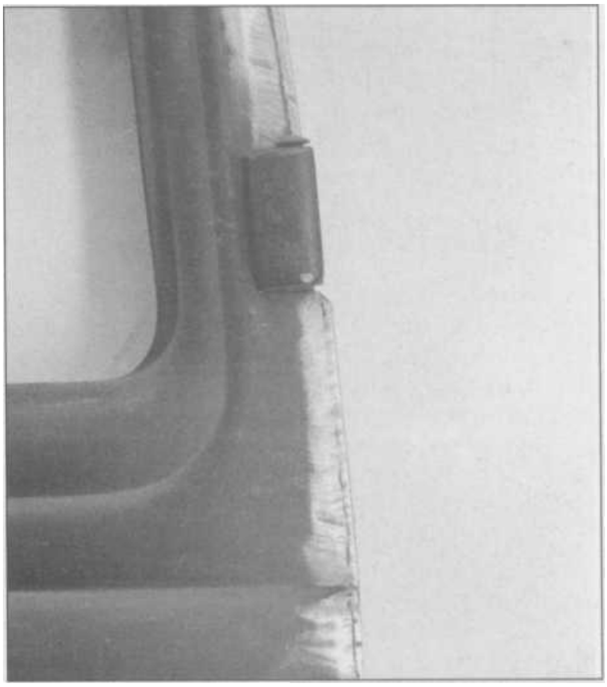

5.65 Customizers and restorers are fond of achieving a perfect gap around body panels on first-class vehicles. Often this involves grinding material off an edge or, as here, adding a sliver of metal along the door edge (at right) to close up a gap. Welding these thin strips is the perfect province of small **MIG** machines.

a stainless-steel wire brush handy that is strictly used for cleaning aluminum. Don't use it for brushing steel or it will later on contaminate your next aluminum weld.

Start welding aluminum with the torch further out than you would for steel, say one inch. Always start an aluminum weld with the end of the wire cut off fresh with snippers. As soon as the arc starts, move the torch closer, down 3/8 to 1/2-

inch from the work. Instead of the torch angle we recommended for steel, hold the gun almost square to the material and inclined only 5 to 15 degrees in the direction of travel.

The puddling will seem much different in aluminum than in steel. The puddle freezes very fast, and the only change you notice is that the molten area is the only place that is shiny. Aluminum has a narrow range of melting before it vaporizes and is highly susceptible to tiny amounts of foreign material. If you see black specks appearing as you go along on a seam, they represent impurities coming to the surface. The short-circuit type of MIG machines we have been discussing are best when the aluminum being welded is relatively thin. Thicker parts are better welded with TIG equipment or the spray-arc-pattern found on larger industrial MIG machines.

There will be a little more maintenance on the nozzle and contact tube when welding aluminum, due to increased spatter. The use of an anti-spatter spray inside the nozzle, or dipping the hot gun tip in nozzle gel, won't produce any less gunk inside the gun, but it will make its removal much easier, as it does with steel welding. If you use an anti-spatter chemical, it can act as a contaminant to the aluminum weld, so, after applying the anti-spatter, briefly use the torch on a piece of scrap aluminum to "burn-in" the chemical so it won't drip off onto your "good" weld.

One of the most common aluminum welding jobs in automotive work is repairing cracked cylinder heads, bellhousings and other large aluminum castings. The cracks should be well ground out to virgin metal and Vee'd before welding. These parts can really soak up heat so they should be thoroughly preheated, either in an oven or with an oxy-acetylene torch fitted with a rosebud tip. The old-timer's method of measuring the preheat was to coat the part with black soot from a overly-rich gas torch. Then with a torch set *properly,* the part was heated until the soot burned off, and that was considered enough preheat. Obviously, too much preheat could require complete re-machining of the part after the welding to correct warpage. Modern weldors use temperature marking sticks to get an exact preheat.

Maintenance on your MIG equipment is simple, but should be performed regularly to keep everything operating smoothly. The most often-maintained items are the gun components. The contact tubes wear out from frequent use as well as become burned such as when you hose kinks and the gas stops during a weld. Sometimes the hot wire welds itself to the contact tip. Often the problem can be fixed by filing on the end of the tip to free the wire's bond, or by pushing the wire to the side with your cutting pliers. If not, you can remove the nozzle, unscrew the contact tip and cut off the wire. Then feed the wire through a new contact tip and screw the tip in place and replace the nozzle. Use anti-spatter agents often, and clean out the nozzle and contact tip before starting a project, during the project, and after the project. When you replace the roll of wire, use a blowgun on your air compressor hose to blow any dust out from inside the cabinet, and blow off the roll of wire before you install it. Keep the drive rollers clean with mineral spirits and lint-free cloth. Often professional weldors will tape piece of lint-free cloth inside their MIG welder such that it drags across the wire as it comes off the roll. The cloth helps to snatch any lint or dirt from the wire before it can collect in your cable or torch.

That's about all there is to MIG maintenance. This is a welding system that has really become the prominent method for the home/shop user in the last decade, and if you have read all of the above, you can see why it has many advantages. Once you start using one, you'll really appreciate it. The lack of paraphernalia, the ease of starting and maintaining the arc, the cleanliness of the welds, reduced shop fire hazards and the ability to weld various materials, including thin sheet metal, have combined to take the MIG process out of the exclusive hands of production weldors and give hobbyists, artists, farmers and restoration technicians a very practical, easy-to-learn welding system.

DEFECT	DEFECT CONDITION	MAIN CAUSE
PORES/PITS	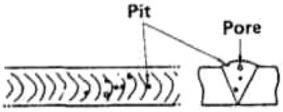	• There is rust or dirt on the base metal. • There is rust or moisture adhering to the wire. • Improper shielding action. (The nozzle is blocked, wind or the gas flow volume is low.) • Weld is cooling off too fast. • Arc length is too long.
UNDERCUT	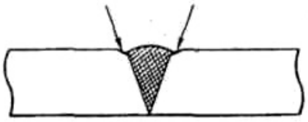	• Arc length is too long. • Gun angle is improper. • Welding speed is too fast.
OVERLAP	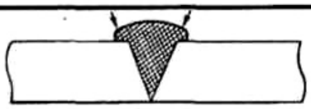	• Welding speed is too slow. • Arc length is too short.
INSUFFICIENT PENETRATION	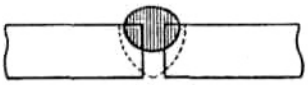	• Welding current is too low. • Arc length is too long. • The end of the wire is not aligned with the butted portion of the panels.
EXCESS WELD SPATTER	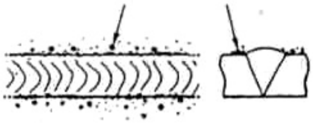	• Arc length is too long. • Rust on the base metal. • Gun angle is too severe.
BEAD NOT UNIFORM	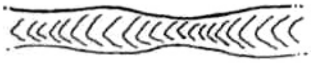	• The contact tip hole is worn or deformed and the wire is oscillating as it comes out of the tip. • The gun is not steady during welding.
BURN THROUGH	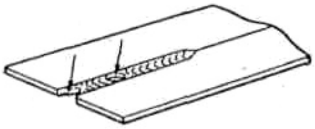	• The welding current is too high. • The gap between the metal is too wide. • The speed of the gun is too slow. • The gun to base metal distance is too short.

5.66 It will take some practice to become competent with any welding equipment, but MIG is the easiest to learn and the most forgiving in most automotive and home/shop applications. Test your practice welds for strength and cut them apart now and again to compare with the illustrations shown in this MIG-welding troubleshooting chart.

6 TIG welding

When those who are interested in welding talk about fusion techniques, the subject of TIG welding is held in a certain reverence. Its reputation in this regard partly stems from its place in history as the method used for construction of a great many famous aircraft. Although developed initially in the 1920s, TIG (Tungsten Inert Gas welding) wasn't used much because the helium shielding gas was too expensive. The intensified research atmosphere of W.W.II spurred further development as aircraft were being made lighter and lighter and TIG became the preeminent method of joining such non-ferrous materials as aluminum and magnesium. The Linde Corporation (now L-TEC) was the first company to capitalize on the technique, and after the war their trademark name "Heli-Arc" became the de facto generic name for TIG welding. Many weldors today still use the term heli-arc more often than TIG as a description of this type of welding, even though a number of other companies have been making TIG equipment for many years.

Besides the romantic beginnings, the TIG process has been considered quite special for other reasons. It does take considerable practice to be good at it, and, because the equipment has always been rather expensive, a weldor who had one and was good with it developed a reputation, particularly in the field of esoteric materials and exotic construction in aircraft and race cars. Basically an outgrowth of

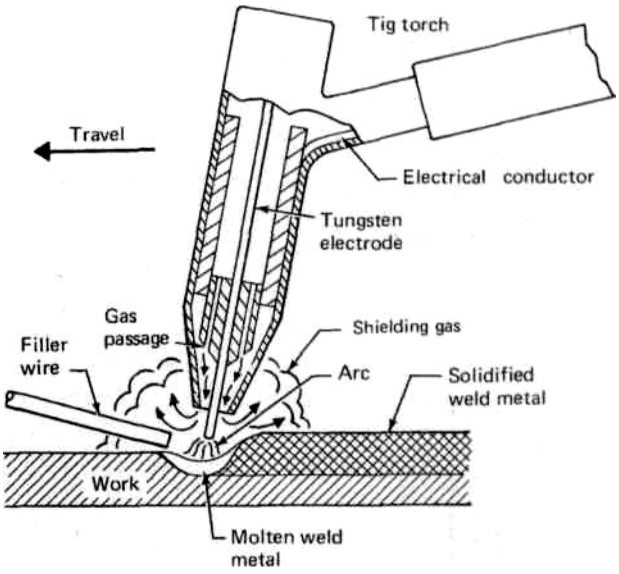

6.1 The **TIG** process is considered the most precise, most controllable and cleanest method of welding. The arc is made between the work and the central tungsten electrode in the torch, while a flow of shielding gas comes in from the nozzle. **Note** the low angle of filler rod addition. TIG or heli-arc **welding** is also one of the most observable welding process, with no spatter and little smoke.

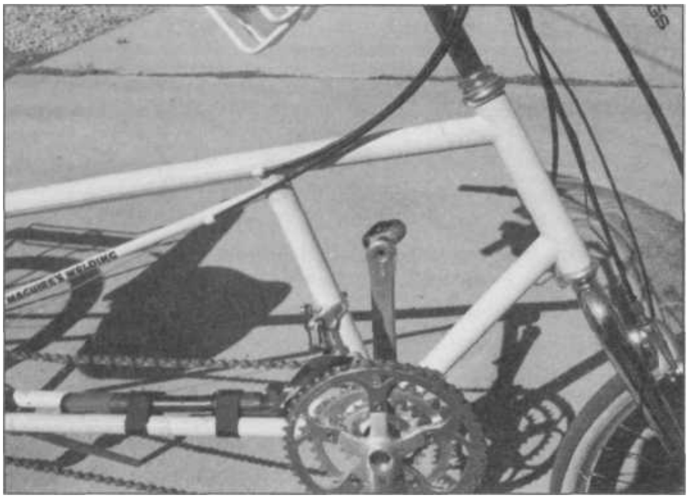

6.2 Where the TIG process is unequaled is on non-ferrous materials or on thin materials. Here a custom bicycle frame was heli-arc-welded from very light, very thin-wall 4130 chrome-moly tubing for strength without a weight penalty. For the same reasons, this kind of material and welding process is use in race cars and aircraft.

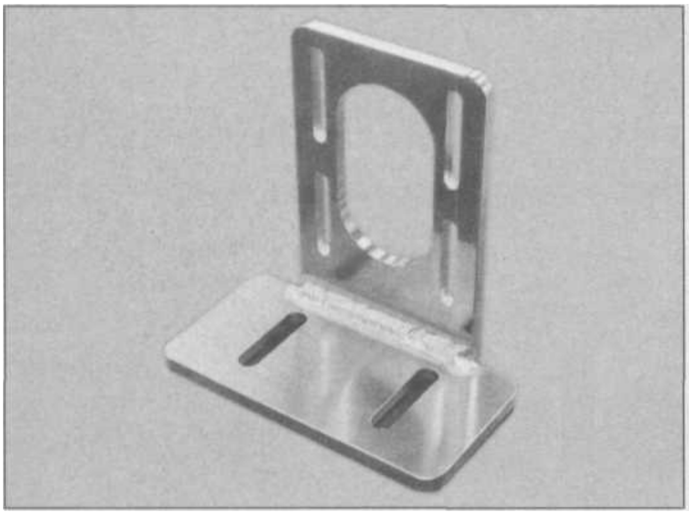

6.3 Where welds on aluminum have to be as pretty as they are strong, TIG is again the top choice. The weld on these two machined and polished parts can easily be wire-brushed to look as good as the polished parts.

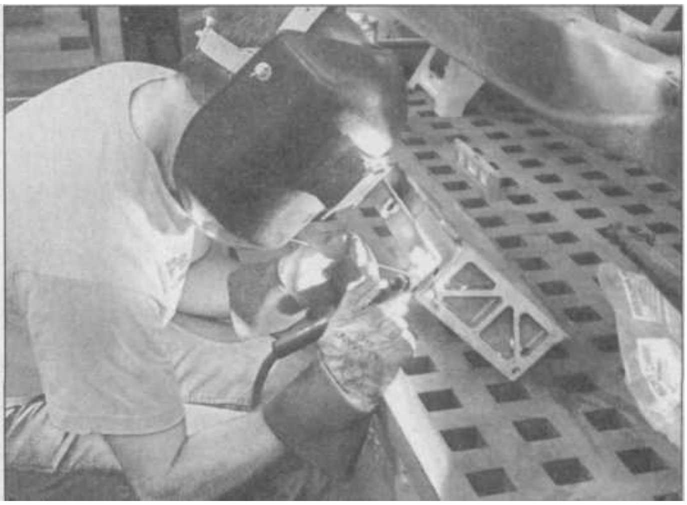

6.4 Unlike arc-welding or MIG-welding, the weldor can get very close to his work with heli-arc equipment. Even when welding on this cast-aluminum dry-sump oil pan, smoke is minimal and the arc not nearly as intense.

arc welding, the TIG process is done with a lightweight torch that uses a tungsten electrode to draw an intense, concentrated arc, shielded by an inert gas. The gas comes out of the torch, all around the electrode and displaces the air from the weld zone to exclude oxygen and nitrogen from contaminating the weld. Although helium was the original gas used, today argon is the most common shielding gas for TIG welding. The tungsten electrode is not considered a "consumable" in the usual sense, since filler metal is supplied by separate, hand-held rods, much like in oxy-acetylene welding. In many ways, heli-arc welding is like a high-tech version of the old-fashioned gas torch.

The concentrated nature of the TIG arc is one of its strong points. Welds of great strength and quality can be made on thin materials, light materials, dissimilar materials, in fact most available metals, and all with minimal distortion or corruption of the adjoining base metal. The process does not generate sparks or spatter, and the welds are as clean as can be, which is why it is the method of choice in aircraft, race-car construction, and stainless and aluminum products made for the medical, dental or food-handling services. There is no slag involved, and, because there is very little smoke generated, the weldor has a good view of the process, allowing him to make very precise joints.

The weldor has greater control with TIG than most other welding methods. The amperage control is infinite because, besides the settings on the power source, the weldor has a foot control he operates as he welds. We recently asked a veteran weldor to make some sample TIG welds to photograph for this book. There was to be one good bead and several bad ones. Even when we set the amperage way too high and way too low, he instinctively tried to produce a good bead, adjusting his torch speed and foot control to compensate. He couldn't make himself do a bad weld, and once the arc was started he tried his best and made fairly good welds even with the machine parameters way off.

With good torch manipulation, a TIG weldor can weld thin materials to thick ones, steel to stainless steel, steel to cast iron, and enough other oddball situations that a good TIG weldor begins to be regarded as some kind of magi-

6.5 Where precision welds are involved, especially in tight quarters such as on this stainless-steel race-car header, TIG welding's wide range of operator control is seen as an advantage.

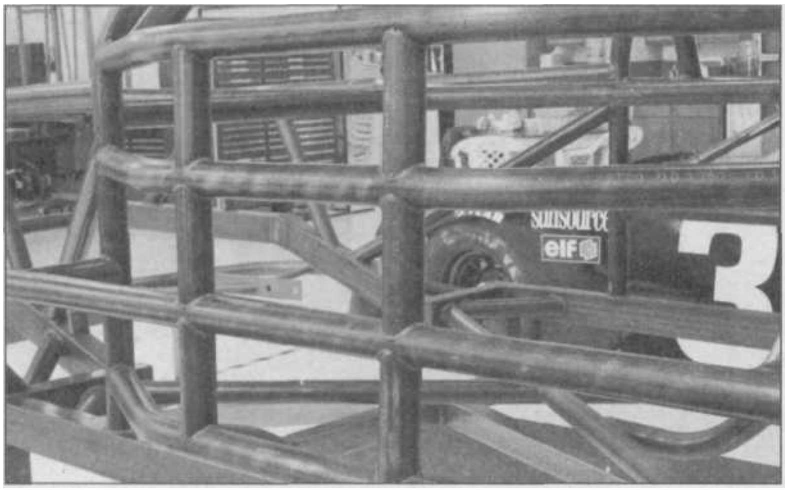

6.6 When thick-wall mild steel tubing is used in a race-car chassis, welding is usually done with MIG equipment, because of its speed. However, this chassis of lightweight, thinwall 4130 chrome-moly tubing was TIG welded. There are a lot of man-hours represented here in precisely-fit joints and welding time.

6.7 Most shop-type TIG welders are heavy, bulky power sources that require three-phase input power and water connections, making them less portable than other welding machines. This is a Miller Synchrowave 250, a popular unit in many fabrication and race-car shops.

cian. In the time that we were in our friend's fabrication shop, he did a production job heli-arc welding some aluminum brackets, some prototype motorcycle seat frames, repaired a Lamborghini aluminum water-pump casting, welded up a fancy cast-aluminum mailbox post that had been run over, and finally a gunsmith brought in a trigger and barrel from an antique cavalry carbine to be repaired with TIG-welding. He handled them all with equal aplomb, and it was obvious that people from all around brought the tough jobs to him knowing he could handle it.

We've discussed many types of welding systems to this point, and each has been shown to have its advantages and disadvantages. With TIG welding there are only few disadvantages. There are very few jobs a TIG welder can't do, but there are some jobs other techniques are better suited to for practical reasons, not because of a difference in quality or strength of weld. For big jobs, like welding up a car trailer or a big metals rack made from steel tubing, the heli-arc is just too slow - an arc welder or MIG would be much faster. However, there are applications in aircraft and race-car chassis work where a large project may be tediously done with a heli-arc, simply because a lightweight, high-strength material like thinwall 4130 chrome-moly tubing is used where TIG would be the preferred method **(see illustration).** For outdoor work, again an arc welder or flux-core-wire MIG are the best choices because they aren't as affected by air currents.

The cost and complexity of the TIG system has kept it out of the hands of most small shops until the last few years. Most large TIG setups are heavy, some weigh up to 800 pounds and have to be moved with a crane or hydraulic pallet-jack, and most professional units have a water-cooled torch assembly which often involves some plumbing connections in the shop beside the normal electrical hookups **(see illustration).** So the big shop-type TIG machines have always been about the most expensive welding machines available, which has increased their mystique but not widened their market. In the last few years, however, manufacturers have built smaller and lighter, solid-state-electronic TIG machines to make them more attractive to the small fabrication shop, race-car shops and street rod builders. Many of these machines are now air cooled, and, at about one-third the cost of the traditional TIG machines, they are being seen out in the field more and more. As one street rod builder told us, "There isn't a thing on a car that would require me to have a 600-amp welder, and when you design and build a car two brackets at a time, the torch never gets that hot."

That is the main difference between the usage of the large, industrial TIG

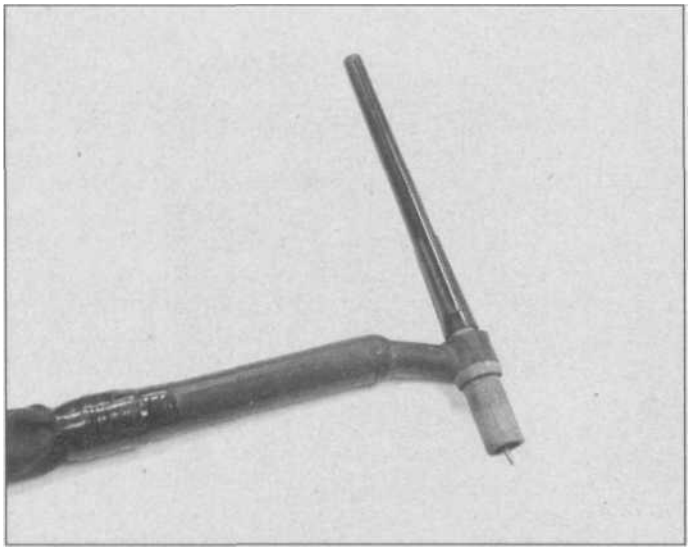

6.8 The basic TIG torch is very light, easy to hold comfortably, and quite durable. The only "consumable" is the filler rods, if used, and to some extent the tungstens.

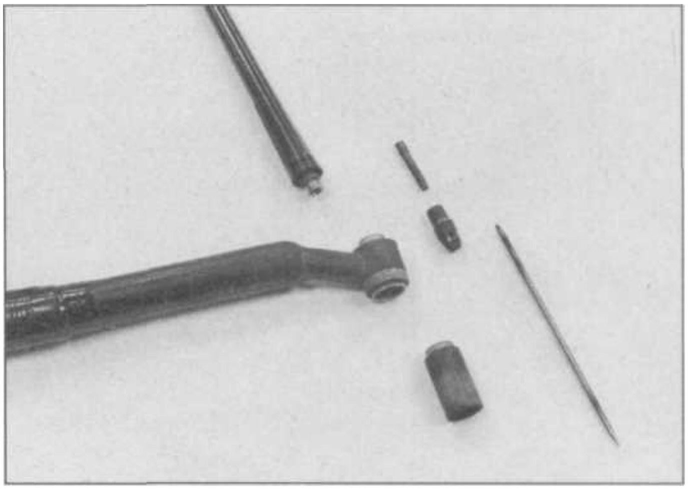

6.9 The exploded view of the typical TIG torch shows the tungsten (right) and from top to bottom: the back cap, collet, collet-chuck, torch body, and ceramic cup or nozzle. To assemble, the collet-chuck is screwed into the torch body on the bottom, the nozzle is screwed in on the bottom, the collet slides in from the top, and then the tungsten is slid in and retained by the back cap.

welding machines and the "economy" TIGs on the market today. The big boys can weld all day long at any amperage they need and never overheat the torch or the electronics, the air-cooled smaller TIGs are fine for anything but long seams or welding big aluminum castings.

The equipment

The basic components of the TIG setup include the power supply, foot control, water circulation system and the torch. The latter differs from the other types of electric-welding torches by being smaller and lighter and holding a central rod of tungsten as the electrode for the arc **(see illustration).** Taking the torch apart, there is a ceramic shield at the business end (called the ceramic, the gas cup or the nozzle), then the collet, which retains the tungsten in the collet body, and the back cap **(see illustration).** There is customization of the torch, depending on the work to be done, in that there are different-sized tungstens, different gas cups with shapes for specific kinds of jobs and different back cap lengths depending on how tight the quarters are where the weld has to be made. For instance, you could use a short tungsten and short back cap to fit into a tight corner better. The shape of the gas cup can create a different shape and size envelope of shielding gas over the weld area. From the larger welding-machine companies, there's a variety of different-size torch bodies with different amperage capacities as well as some special models made for tight confines, including one in which the tungsten and ceramic nozzle come straight out of the handle rather than at an angle like most TIG or MIG torches.

When it comes to the basic power supply, you will not find as great a range of choices as with say, MIG welders. The marketplace for TIG equipment is still pretty much a professional arena, and the newer, less-expensive air-cooled TIGs are at one end, and the big, shop units are at the other, with few choices in the middle. A TIG setup is still not a home/shop type welding process, though there are certainly some small specialty fabrication businesses run out of a home garage that could utilize a TIG machine, particularly if the work involved aluminum, stainless steel or other non-ferrous parts, and high-quality, squeaky-clean welds were important to the products' strength or appearance. If you had a

6.10 There are only a few small, more-portable TIG units, like this 100-amp "pulse-TIG" from Daytona MIG. It is a 220V, DC-only machine, but is said to be suitable for steel, stainless-steel and chrome-moly up to 3/16-inch thick.

6.11 Mid-size TIG machines are in the 160-250 amp range, though this fairly-portable Daytona MIG unit is 130-amp, 220V. Mid-range machines offer options like slope control and post-flow. This one has square-wave technology, and AC/DC operation.

small business making stainless-steel rope clips or other items for rock climbers, for instance, you might weigh the cost of an air-cooled TIG versus sending all your parts out to be welded in a shop. If you had the volume of work to justify the one-time costs and the learning process, it could be a good move to do your own. However, it would be a wise move to take some TIG classes and get yourself certified first.

Input voltage on TIG machines begins in the 220V range of single-phase power, and the bigger professional boxes are for 460V three-phase. The professional units range in amperage capacity from 250 amps to 300, 400, 500, and 650, though the larger ones are multi-technique machines on which the high amperages are probably used only on the arc-welding process. It would be pretty rare to use more than 300 amps on a TIG weld, unless you had to weld really thick aluminum plates or castings. When welding at 300 amps or more, tiny bits of the tungsten electrode can come off and become included in the weld, which would weaken the joint. The three-phase current models are definitely for shop use only, where the amount of shop equipment and power used every month makes the expense of transformers and rewiring the building economical over the long run in utility bills. Home shops generally do not have three-phase power.

On the small-shop scale, there are very few TIG-capable units in the 115V range. One we have seen is the Hobart Ultra-Weld 130, which is an arc-welder with TIG capability, but with a limited range for either. In its 115V version, it offers 75 amp capacity with a 40% duty-cycle, and the 220V model offers 130 amps at 60% duty-cycle. It does not include the TIG torch, and its TIG range of material thicknesses would be limited. We expect that as the home TIG market expands, there may be some interesting, smaller 115V and 220V units to come. A few imported TIG units are built with DC-only capability to keep the price down to a starter level, and they are fine for steel (only) up to 3/16-inch **(see illustration).**

Of the medium-range TIG units, there are several of the domestic and imported brands offered in the 160-250-amp category that can handle virtually anything in the small shop environment **(see illustration).** Most of these are 220V or more in input current requirement, but most are single-phase, which means you could easily wire your shop to accommodate them. Be advised though, that even the 160 amp units can draw 36 amps of input power at their highest settings, and

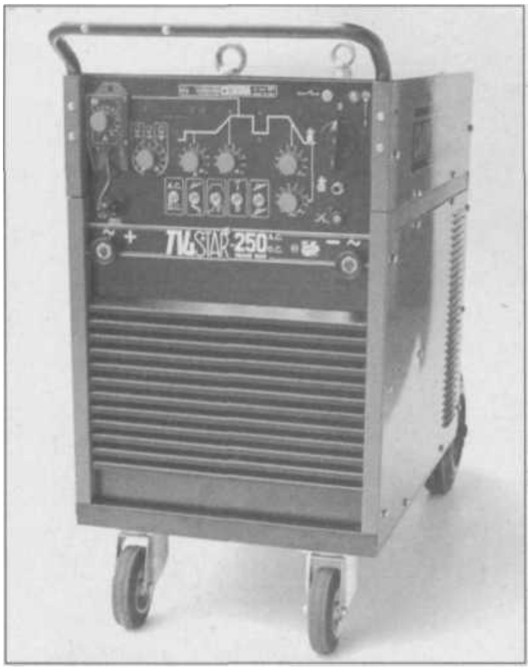

6.12 When you go up in welding machine size, you get more amperage and higher duty cycles. This TIG-Star 250 operates on several single-phase input currents, and has a 60% duty-cycle at 200 amps, with 100% duty-cycle at 140 amps.

the 250-amp machines can draw 60 amps or better on full load. This means that your household 30-amp 220V circuit would blow at any of the higher settings. You would really need a separate circuit and box for the welder, with a 50-amp breaker for the smaller mid-range machines and 100-amp breaker for the bigger ones.

The mid-range machines offer several special features, such as *slope control* **(see illustration).** This refers to how the wave of current cycles when the machine is welding. Picture the wave as a curved line that goes up, down, crosses the baseline and then curves back up and over again in repeated fashion many times a second. This is called a sinusoidal wave in alternating current. TIG machines use both AC and DC current, with the AC used for non-ferrous metals like aluminum or magnesium. One half of the AC wave is actually straight polarity, and the other half is reverse polarity. The straight-polarity portion of the wave builds the heat in the arc, while the reverse current has the effect of cleaning the oxides from the aluminum being welded. So the TIG torch on AC is constantly cycling between a hot arc and a cleaning cycle, but at lower amperages the arc can be obstructed at the changeover point or actually go out. Most of the TIG machines have a special circuit inside that produces a high-frequency, low-power extra current that keeps the process going at all times. The high-frequency control makes the TIG arc easy to start without touching the tungsten to the work, and makes a more stable arc while welding. On many machines the high-frequency system can be set on the front panel of the welder to arc starting, off, or continuous use. On ferrous metal welding, the high frequency is used mainly for arc starting.

The slope control is a feature in which the weldor can adjust the parameters of the wave on the up side or the down side to get the perfect combination of penetration and cleaning action on aluminum. Even a clean piece of aluminum contains some aluminum oxides in the base material, and the slope adjustments allow a good weld to be made without these oxides being included in the bead.

Most of today's TIG machines offer *square-wave* AC. The traditional sinusoidal wave with high-frequency over it creates a lot of radio interference as well as some problems on welding aluminum **(see illustration).** The new machines are all solid-state electronics, with features that make a wave made of up and down squares, rather than half-circles. This means the high-frequency control intervenes less often, and the same-sized electrode can be used at higher currents. This discussion of the current waves may be more than the home/shop user needs to know, but at least you may better understand the language used by manufacturers to describe their equipment. Suffice it to say that the modern square-wave technology and solid-state electronics are desirable features in a full-on TIG machine.

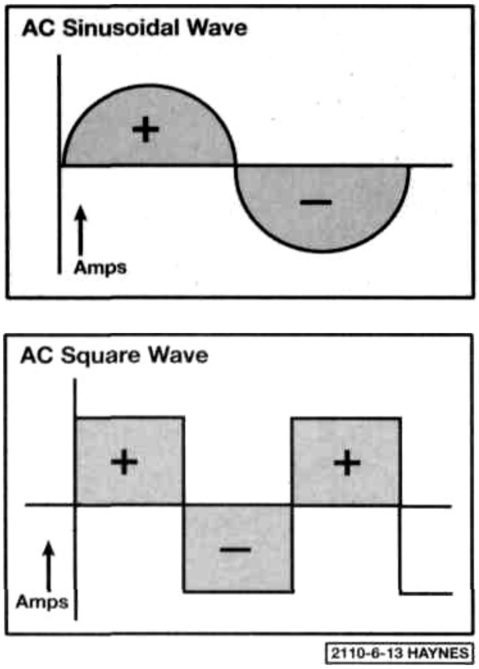

6.13 As a standard "sinusoidal" wave of AC current flows, it is constantly reversing its polarity and its output is always fluctuating as the wave flows. With the new square-wave technology, electrodes can be used at higher currents and have good arc staring and stability, but without the radio interference of traditional high-frequency superimposition. The square-wave design inverts the current much faster, for better welds on non-ferrous materials.

Another feature you may see mentioned is post-flow. When you are making a bead and get to the end of your seam it's a good idea to allow gas to flow over the end of the weld as it cools, to keep out contaminants. On most professional machines, there is an adjustable timer that allows gas to flow for so many seconds after you have stepped off the foot control to quit the arc. You keep the torch in place just where you stopped, and gas keeps flowing for the specified time. This is

6.14 Less expensive, air-cooled-torch TIG machines have been available now for some time and are making inroads into shops where TIG was previously considered too expensive. This Miller "Econo-TIG" is being used in a street-rod shop on a home-built cart to make it as portable as any MIG machine.

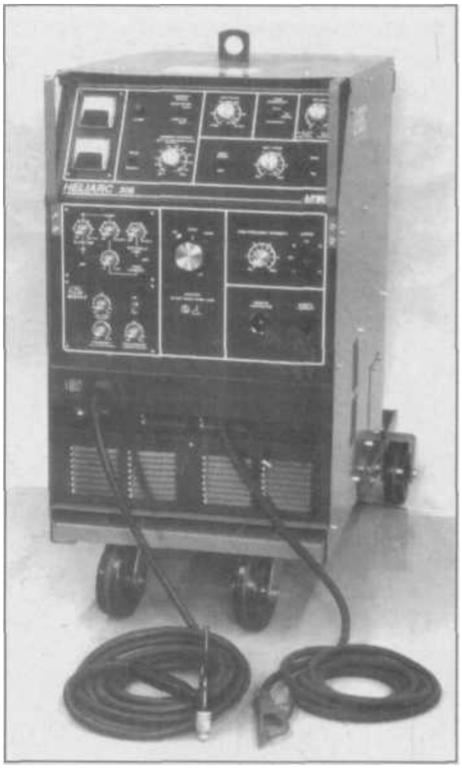

6.15 Today's state-of-the-art in TIG machines is represented by this 300-amp Heli-arc 306 by L-TEC (the new name for Linde Corporation, the original trademark-holders of the name heli-arc). All solid-state with square-wave technology, it provides high performance and high reliability. The bells and whistles include: slope control, spot-weld control, wave balance control, post-flow, and an "arc force" fine-tuning adjustment.

helpful on steel, but is essential on aluminum, magnesium and brittle, high-strength steels like the 4130 used in aircraft and race-car construction. It's called post-flow because it is post-welding gas flow. The post-flow of water and gas to the TIG torch also helps to cool the tungsten electrode down to the point where it can't oxidize when regular shop air flows around it after the welding stops.

Most shop-type TIG machines feature a water-cooled torch, but some are air cooled, with the water-cooling system as an option **(see illustration).** Obviously, the extra equipment for a water-cooled torch adds to the total price of the package. In some shops, the water-cooling system consist of a unit added on to the basic power supply, containing 1-5 gallons of water, a water-circulating pump and a small radiator and fan. It's like a miniature automotive cooling system. In shops where the amperages used can be quite high, or where long welds are made, the torch may be hooked to a source of regular cold tap water, and the heated return water flows back into the shop's drain. This system uses more water, but keeps the torch cooler than recirculated water systems **(see illustration).**

The shielding-gas system for TIG welding isn't much different than for MIG welding, as discussed in the last chapter, except that you will use helium, argon or a mixture of the two in your bottle. Most TIG setups use as large a bottle as possible because portability isn't usually a concern in TIG welding **(see illustration).** Shops that do TIG welding usually have the machine set up right near their welding table and don't move them around, although the newer air-cooled units can be wheeled around a shop without too much trouble.

Unlike oxy-acetylene bottles or most MIG-welding gas bottles that use two-gauge regulators, TIG setups have a single round-faced gauge on their tank and a flow meter instead of the other gauge for the line flow. The flow meter has a vertical glass tube with numbered increments marked on it and a colored ball inside. When the gas is

6.16 Argon bottles are available in a number of sizes, depending on whether you are trying to make your TIG setup portable or not, and how much welding time you need. Shown here, from left, the sizes are: 20, 55, 80, 125, 154 and 248 cubic foot.

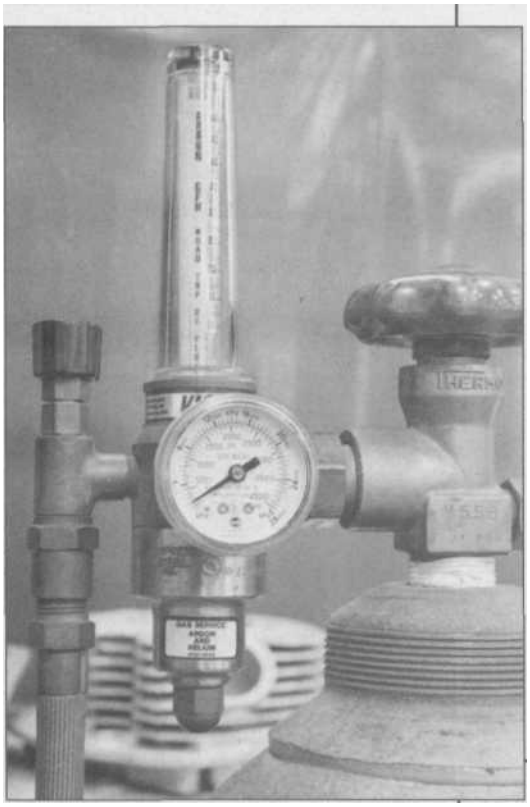

6.17 Flow-meters for argon bottles in TIG use are generally used in place of twin-gauge setups. Inside the glass tube on top is a colored ball that floats along a measured scale to indicate gas flow, but without requiring the weldor to step away from his work.

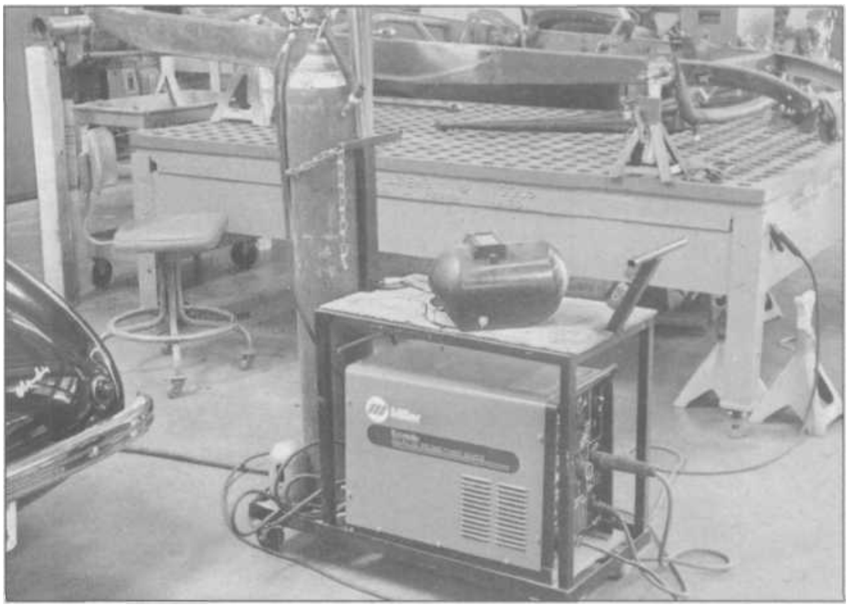

6.18 Argon is versatile enough to be used as the shielding gas for almost all TIG work, simplifying the storage of bottles. Often one large bottle mounted on a welding cart is all that's needed in most smaller shops.

turned on and you click your foot-control on, the ball will float up inside the tube, indicating the level of gas flow in cubic feet of gas per hour **(see illustration).** Professional weldors like the flow meter better than a gauge because that colored ball can be seen floating in the tube from as far away as their welding table - they don't have to get up and look at the face of a gauge on the tank. They can just step on the pedal to start gas flow and see if the flow meter is set right. They can also tell if they have forgotten to turn on the gas bottle before starting.

The correct gas flow is very important in maintaining the integrity of the weld by keeping out contaminants. For most basic steels, a flow rate of 8-10 cubic feet per hour with argon is sufficient up to 1/8-inch material, but the gas flow requirements go up with the amperage used and the thickness of the metal being welded. Stainless-steel requires only a little more gas flow than for plain steel, but if you were welding aluminum, the gas flow recommendation would go up to 15-20 CFH for the same thicknesses, and a larger-diameter tungsten may be required.

Argon is certainly the basic, all-around shielding gas for the TIG process today, and is useable on virtually all metals **(see illustration).** Straight helium is seldom used, but often in production work a mixture of argon and helium is used to get the best properties of both gasses. Argon has the qualities of good arc-starting, cleaning action and best puddle control, while helium makes for a hotter weld for greater penetration and a faster welding speed, such as when welding thicker materials. An argon-helium mix is often used to allow welding of non-ferrous metals at higher speeds. The helium in the mix helps make more heat for aluminum welding in a way that one weldor described as like "adding on a turbocharger."

The process in action

Basically, TIG welding is enough like oxy-acetylene welding that if you have already learned well how to weld with a gas torch, the learning curve for TIG welding will be much shorter than for most other weldors who have never learned

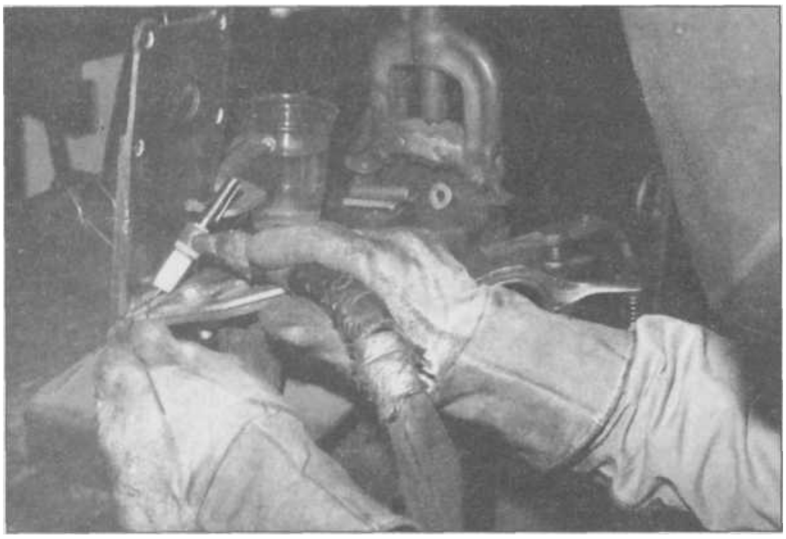

6.19 Finding a comfortable position for your torch hand is more than half the battle in achieving steadiness with TIG welding. Practice making a bead on your proposed setup *without* turning on the arc, just to find a comfortable position. Note the low angle and position of the filler rod in the weldor's left hand.

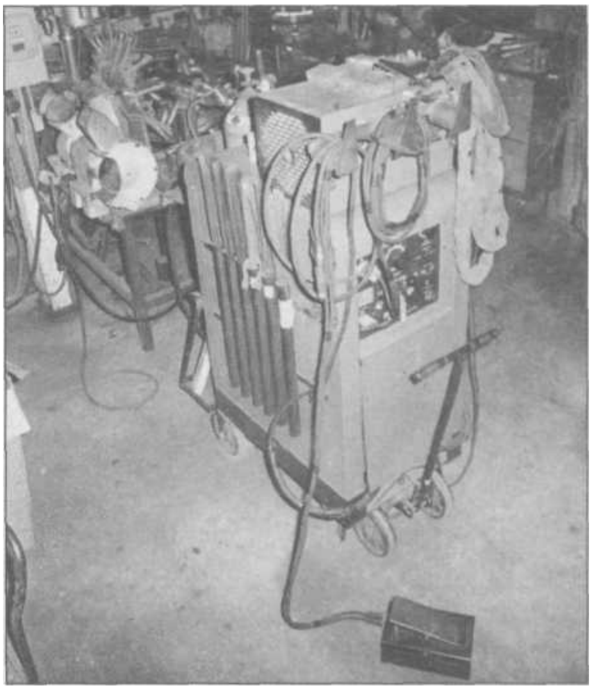

6.20 This is a TIG welder of some years back. Today's current crop of TIGs are half this size. The water-cooling unit here is mounted on top, and at left is a rack of metal tubing holding various welding rods. The foot-pedal amperage control is in front.

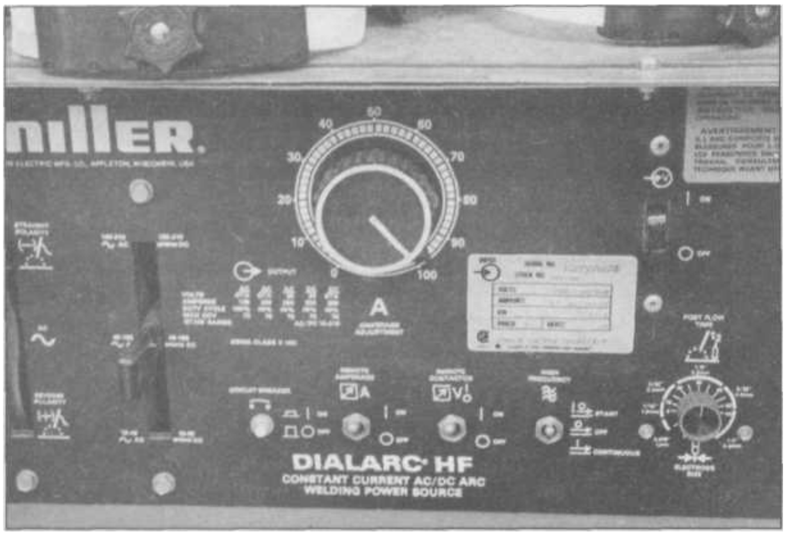

6.21 Layout of this front panel is typical of most TIG power sources, with controls for coarse and fine amperage, current type, high-frequency control and post-timing.

with a torch. It can be as challenging as it is rewarding. Like gas welding, the operator must use both hands, one for the torch and one for the filler rod, each of which must be moved along the weld seam at the same, steady rate **(see illustration).** In addition, the TIG weldor has to operate a foot-pedal-controlled or hand-controlled amperage device (usually clamped directly to the torch body) that is remote from the welding machine **(see illustration).** Using the remote control is one of the operator privileges of the TIG process, allowing the weldor to start an arc, infinitely vary the welding current *as he is welding,* and end the arc without moving the torch. Being able to vary the amperage while making a weld is what makes TIG welding so versatile in handling different thicknesses of metals and the joining of dissimilar metals. Just the right amount of heat can be aimed at each side of a joint, and this also helps make TIG a technique that can work for all positions of welding, too.

TIG machines produce either AC output with high-frequency over it (HFAC) or DC with the high-frequency for arc starting only **(see illustration).** Most TIG machines have a control on the front panel that switches between the types of

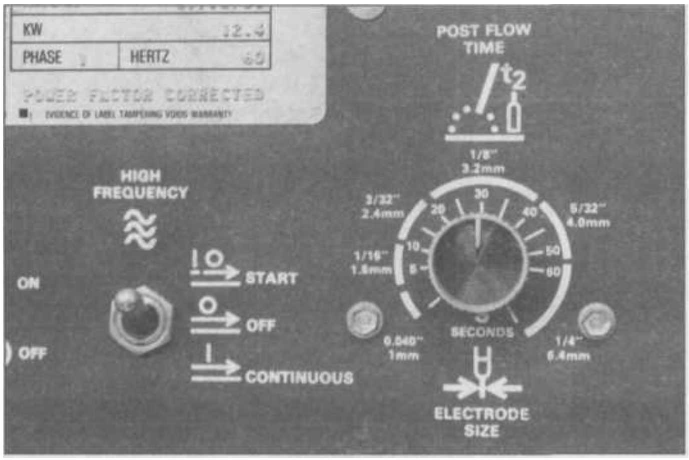

6.22 The three-position switch at left is for the high-frequency, and at right is the post-flow timing as related to various sizes of tungstens. The bigger the tungsten, the more post-flow timing to cool it off.

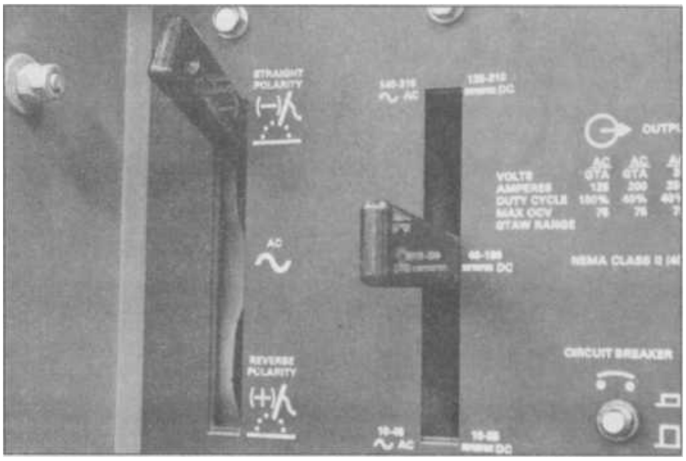

6.23 The lever at left on this Miller TIG sets the polarity, while the right lever is the coarse amperage, with a bottom setting of 10-55 amps, a medium range of 40-190, and a top range of 125-310 amps. Within each range, the fine-amperage knob sets the upper limit for the pedal.

current, a control to select the basic amp range and controls for high-frequency and post-flow **(see illustration).**

The AC is used on non-ferrous metals and usually has the high frequency on during welding, except for the newer square-wave machines that don't require high frequency all the time. The DC current can be either straight polarity or reverse polarity. The straight polarity (DCSP) is used for steel, stainless-steel, silver, silicon copper, brass alloys, cast iron, deoxidized copper and titanium, with the AC mainly for aluminum and magnesium.

Set your Machine to DCSP (sometimes marked DC-), the high-frequency to the start-only position, and set the amperage for the material you are working on. Experienced TIG weldors usually set the three-position amperage lever as a coarse adjustment, then dial in the fine amperage control with a knob (see **illustration).** Often weldors will set the fine control all the way up, using their pedal as the ultimate control. That way, if during the welding the weldor needs a little more heat, he just puts the pedal down further. Whatever you set as the limit on the fine-amperage knob, that's as much as you can get from the pedal **(see illustration).**

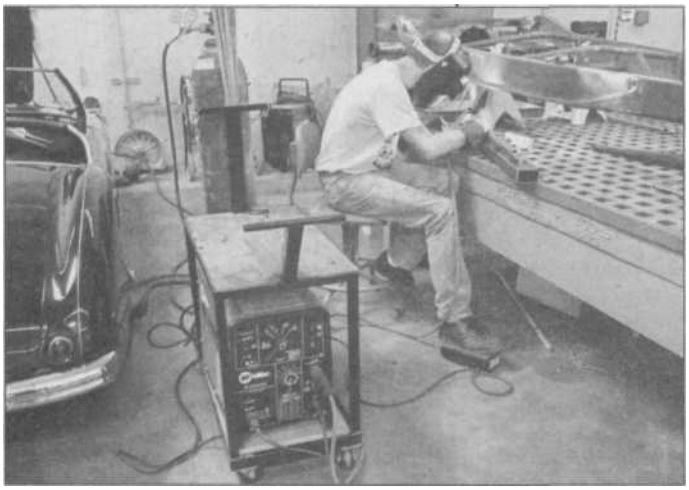

6.24 The pedal is the final weldor control of the amperage in TIG welding. Until you are experienced, you may want to set the amperage knob at the highest you need for a particular job, and know that when you are welding, the pedal can go all the way. It takes some experience to know *how much* pedal to give, if the knob is set to the max for the amp range on the coarse adjustment.

There should be a chart with your machine that tells you the correct size tungsten electrode to use with various amperages and material thicknesses. You will probably use a 1/16-inch, 3/32-inch or 1/8-inch for almost everything you do. There are two basic kinds of tungsten materials, pure tungsten or thoriated tungsten. The pure tungsten works under all conditions and metals, but is *required* for AC TIG, while the thoriated tungstens for DCSP welding are used often because they make slightly better starts and penetration.

Tungstens are somewhat expensive, but can be made to last a long time **(see illustration).** The sharpness of the tip is important to good arc starting and penetration on steels, so TIG weldors keep a small supply of sharpened electrodes handy to slip into the torch **(see illustration).** Most weldors will sharpen both ends of a tungsten, so that if they need a new point right away they just loosen the back cap on the torch, slide the tungsten out and reinsert it with the new point down. Good weldors can do this in a

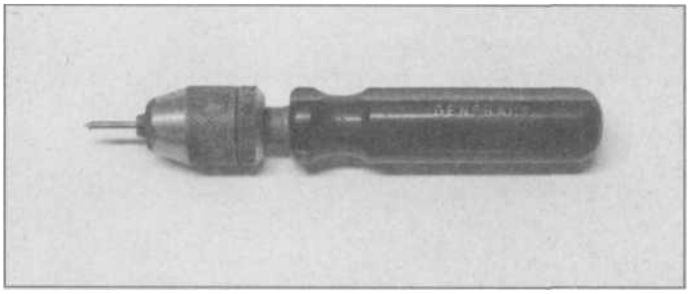

6.25 Make the most of expensive tungsten electrodes by using them until they are too short to fit the torch'. Shorter tungstens can be safely sharpened without burning your fingers, using a tungsten-holder like this.

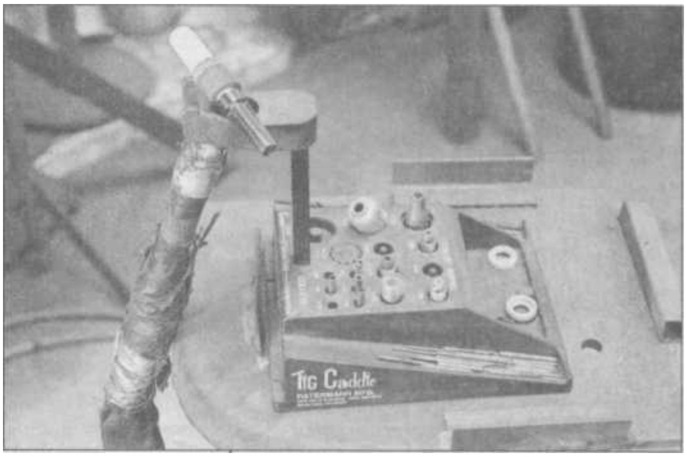

6.26 Spare, sharpened tungstens should be kept handy. This is a TIG caddy, which holds extra tungstens as well as spares for all your torch parts and different size collets and nozzles. The caddy is magnetic, and features a convenient torch rest.

blink of an eye. A belt sander or grinder is usually used for sharpening tungstens, but it should be a dedicated sharpening source **(see illustration).** A grinder used for smoothing steel or aluminum will have deposits on it that can contaminate the tungsten. Sharpen the tip (like sharpening a pencil) only by grinding *lengthwise* on the electrode, grinding from the tip back to the main body. Grinding it held sideways to the grinder will leave a pattern of marks on the tip that may contribute to arc scattering. Electrode tips for use on aluminum with HFAC should be sharpened to be clean but with a much duller point than used for steel **(see illustration).** Theoretically, you should have separate ceramic cups for aluminum work too, but many weldors do use the same cup for both steel and aluminum TIG welding. When tightening the electrode in the torch, push the whole torch straight down against a solid surface (with the welder off) to see if the tungsten will slide in. If it slides, the tungsten wasn't tightened enough **(see illustration).**

At some point, your tungsten tip *will* contact the puddle, this isn't desirable, but it happens to everyone. Pull the pedal back all the way to break the arc, and keep the tungsten in the gas flow for a few seconds until it cools off a little. You cannot continue to weld with this tungsten after it has touched either the filler rod or the puddle because it will

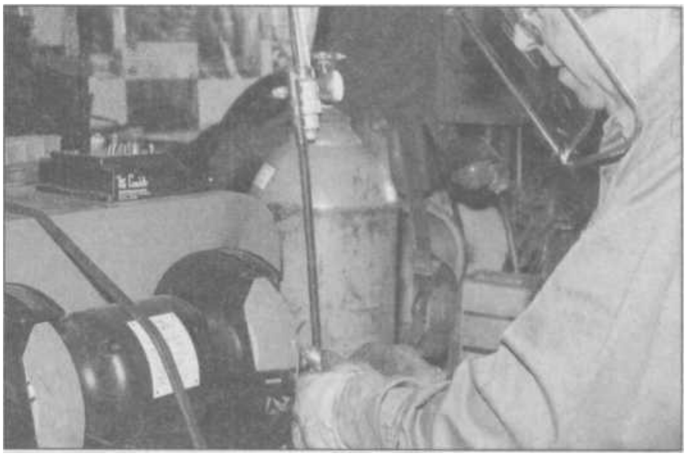

6.27 It's important to sharpen tungstens lengthwise on the tip, so as not to leave any radial grinding marks that could cause arc scatter. Grind points for steel like a pencil, like a crayon for welding aluminum.

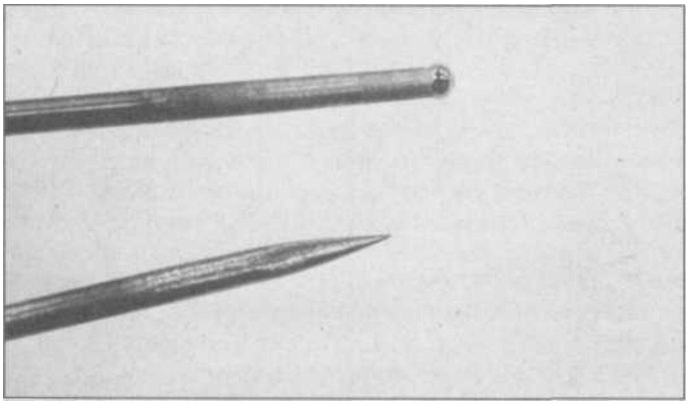

6.28 The tungsten at bottom has been sharpened for use on steel, while the top one is for aluminum, with a small ball-shape on the end. In theory, you could create a ball-end like this by striking an arc on a piece of copper before using the tungsten on aluminum, but a suitably-dull point will turn into a ball like this with a few seconds use on HFAC for aluminum.

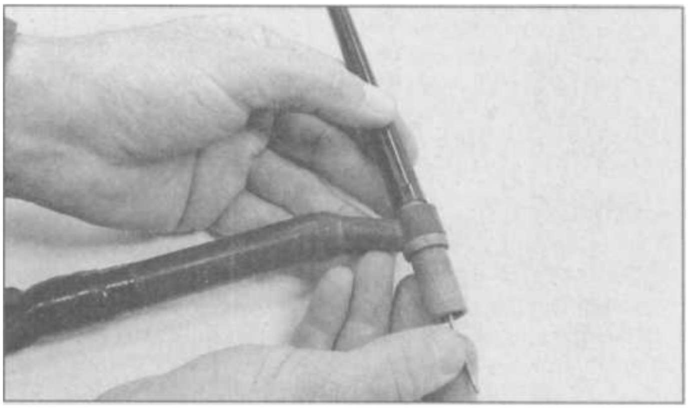

6.29 The sharpened tungsten is inserted into the torch body, and when the tip is sticking out the right distance, you tighten the back cap to hold it in place.

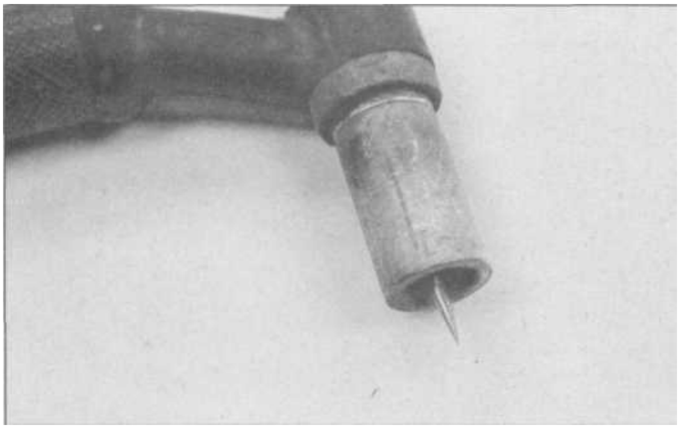

6.30 Although you can crank up the gas flow and extend the tungsten tip out further for reaching into a tight corner, the normal extension of the tip should be 1/8-inch to 3/16-inch from the end of the nozzle.

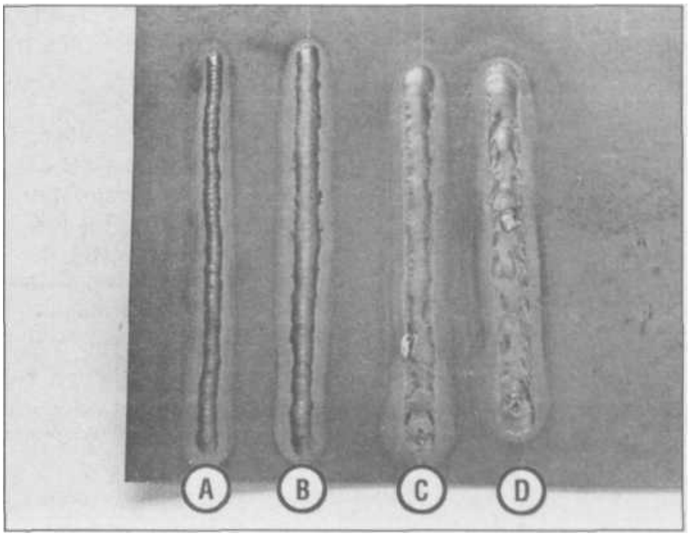

6.31 You'll make a lot of practice beads while learning to TIG weld. Here at left we have: A, a bead with too low an amperage, the bead sits up proud, B, the current control seems about right., C, the amperage is set a little too hot even though the bead seems OK on top, and D. way too hot, with the bead too wide and sunken.

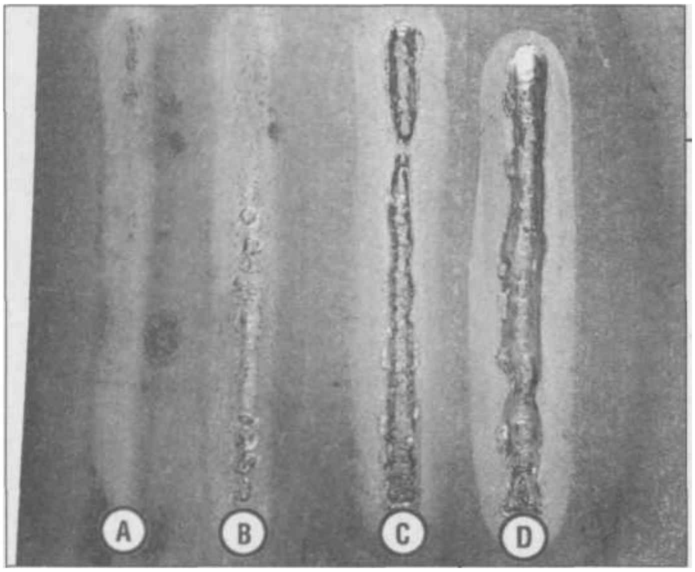

6.32 This is the backside of the same plate as the previous welds. Note how A has only a slight discoloration, B is close to being right, with a tiny bit of dropthrough, and C is too hot, with lots of dropthrough, while D is so hot there's more bead underneath than on top.

be contaminated, and the arc won't be clean and steady. The blackened and pitted section of the tungsten will have to be broken off with a pair of pliers, and the tip will need to be re-sharpened. Although tungstens are not considered consumable in the normal sense of the welding term, they do get shorter every time you break off a piece and every time you re-sharpen them, so eventually they become too short to use. When welding in tight corners, a short back cap can be used with a stubby tungsten, but otherwise use a long cap to protect you from contact with either a sharpened tip (when you have sharpened both ends) sticking out or a hot tungsten. The long back cap covers even a full-length new tungsten.

Turn on your gas and set the flow to 10 CFH, attach the work-lead to a sample plate and run some practice beads. The tungsten should be held in the torch such that 1/8 to 3/16-inch of tungsten extends out past the ceramic cup **(see illustration).** At no time should the tungsten touch the work. When you bring the electrode to within about 1/8-inch of the work and hit the foot pedal, the arc should start and there will be a red-orange glow at the tip of the tungsten. As you move the tip around in small circle in one spot with the torch perpendicular to the work, bring the pedal control down for more amperage. When the puddle is a bright fluid, move the torch along your seam slowly and evenly. The torch should now be tilted back to a 75-degree angle for the rest of the bead. To stop the process for any reason, just back off the foot control all the way to break the arc, but keep the torch over the same spot to cool the weld and tungsten with gas flow.

When adding filler rod as in most joints of any thickness other than thin sheet metal, move the electrode to the rear of the puddle and dip the filler rod to the forward edge of the puddle. Then bring the torch to the front again and make a new puddle, move back, add filler, move again. This may sound like too much work, but really it is very similar to gas-torch welding and the movements become instinctive after enough practice. The filler rod should not touch the electrode, and it should be kept near the forward edge of the seam to keep it hot between "dips." When it is brought to the molten leading edge of the puddle, a drop of molten steel will come off the rod into the puddle and be blended in when the torch advances over it. The filler rod is usually aligned with the direction of weld travel and held down to a shallow angle with the work.

The rate at which you travel will depend on the thickness of the material you are working on, the amperage and the desired height and width of the bead. If you are too cold with the amperage or travel too fast, the bead will tend to stand up more, and even if it looks smooth and uniform, the weld won't pass a bending test because of lack of penetration **(see illustrations).** If you are too hot or moving too slowly, the bead will settle into the parent metal and leave a long crater or at least some undercut on either side. On metals thicker than 1/8-inch, the edges of the parts should be beveled, and you will probably make more than one pass. The bottom pass or *root* pass ties the pieces together, and subsequent passes can be made at a higher amperage. TIG welding is not a process that quickly fills in large gaps or beveled seams on thick plates. MIG or arc welding is better for that.

The filler rods used for TIG welding may look exactly like those for gas welding, but they aren't. The TIG rods for steel welding have deoxidizers added, in fact one brand calls theirs "triple-deoxidized," and if you accidentally use a gas welding rod, you can make a bead but the weld will exhibit porosity. The proper TIG rods are very helpful when welding low-quality steels, which may be the "hot-rolled" steel you find at the salvage pile at your local metal supply. There are a variety of filler rods, including special blends used for the low-alloy, high-strength sheet metals now used in automobiles.

After welding, your tungsten tips should be silvery and relatively clean. If they show signs of contamination, or are rough to touch, it usually indicates that air is leaking in somewhere in the torch or that shielding gas is leaking out somewhere. Check the cables and the fit of the collet and ceramic cup. If the electrode looks like it has a ball on the end like a miniature flagpole, then the amperage may be too high for that diameter of electrode. The tungsten shouldn't have a ball on the end bigger than 1.5 times the diameter of the electrode itself. If the tungsten tip is black or blue, check the timing of the post-flow on your machine. If the tip is black, it means the tungsten is cooling off in atmosphere instead of shielding gas and you need a little more time on the post-flow setting.

Look at the inside of your ceramic cup after welding and check it for contamination. A cup that is too small for the work will show deterioration and even cracking. Cups are usually sized in 1/16-inch increments. A #4 nozzle is four-sixteenths or 1/4-inch, a #8 is 1/2-inch. The general rule is for the size of the cup to measure 4-6 times the size of the electrode being used.

Your practice time before obtaining suitable beads will vary with your skill and previous welding experience. As we have said, the similarities between TIG welding and oxy-acetylene gas welding are such that you can pick-up TIG welding much faster if you are already an accomplished gas weldor. When you have practiced straight beads and flat, butt welds and have those down fairly well, then try some corner or Tee welds, and then vertical seams. When you are comfortable on these more difficult welds and the pieces you have done pass simple bending tests and cutaways to examine the penetration, you may be ready to do some actual projects with a TIG welder, but don't start right away with something critical like an automotive suspension or steering part, or anything that goes on an airplane.

TIG-welding aluminum

Perhaps a little more difficult to learn than welding steel, precision joints in aluminum or magnesium are among the primary reasons for having a TIG welder, so this should be your next area of practice.

You need a well lighted area to do TIG welding because the arc is not as bright as with other electric arc techniques, and on aluminum you really need to see how the puddle is developing. Set your machine to HFAC, with the high-frequency on continuous, or use the square-wave feature if you have it, and adjust the post-flow for 1/8-inch electrodes. Use a clean 1/8-inch electrode (pure tungsten) with the tip shaped for aluminum (not a needle-sharp point), and adjust the flow meter for the argon to 17-20 cubic feet per hour. Set an 1/8-inch-thick aluminum plate that has been cleaned thoroughly down on the welding table. Aluminum tends to age-harden when it is exposed to air, and aluminum oxides will constantly build up on the surface; such oxides are not always visible. Parts to be welded need to be cleaned just before welding to provide the freshest surface. Do not use chemical cleaners for preparing aluminum, unless it is something you have purchased at a legitimate welding supply store. Household or automotive cleaners may contain chemicals that will vaporize during welding and create toxic fumes. Use a stainless-steel wire brush (not steel wool) only, and never use this stainless-steel brush for cleaning plain steel. Since most sandpapers are made from aluminum oxide, and this is exactly the material we are trying to eliminate when cleaning aluminum for heli-arc welding, don't use ordinary sandpapers to pre-clean.

Use an aluminum welding rod, and clean the welding rods before starting, using alcohol and a lint-free cloth to wipe them down. The aluminum part you are welding should be well grounded to your table. If it is lightweight plate, put a heavy weight on part of it to keep it in contact with your grounded table, or attach the work clamp directly to the part. Otherwise, you may find that the bottom of an aluminum part you have welded on your table will have arc burns from intermittent contact with the table. Draw a bead on the aluminum for a short distance, then immediately cut off the arc. Examine this weld and you will probably see that where you stopped there's a small depression at the end of the bead. To avoid this at the end of every weld, don't back off suddenly, but rather ease out of it while slowly moving the tip back over the end puddle, and it should keep from sinking.

Try to maintain a puddle of molten aluminum about 3/8-inch wide. Going too fast will keep it narrower, and vice-versa. You won't see any real changes in the aluminum in the way you are used to seeing steel alter during welding. The only clue that an area has become molten is that it gets very shiny in that spot. You'll have to watch for this change in reflectivity when making a series of puddles that make up a bead, because aluminum melts at 1250°F, much lower than steel and can be overheated easily with the intensity of an arc. When welding steel, you have a whole range of colors that the metal turns at different temperatures. These colors make it easier to tell when you have the metal molten and when it is just starting to cool off. You won't have this advantage with aluminum, and you'll probably burn through a number of samples before you learn temperature control on this metal **(see illustration).**

6.33 When you are learning TIG welding on aluminum, keep your torch angle so that you can get a good, close look at how the puddle develops and the molten area becomes shiny. You need enough heat to form the puddle, but not much more or the aluminum vaporizes.

After the weld is completed and cooled, you'll see your bead is nice and shiny because all of the bead material has been melted and it looks as good as virgin aluminum just out of the foundry, but along each side of the bead you'll see a light-colored band that is contrastingly dull. This is an area where aluminum oxides have been vaporized from the weld bead by the high frequency in the arc and deposited next to the weld. If the aluminum plate you have been practicing on was not entirely clean, you may see some black spots, black "smoke" residue, or a wide dull area on either side of the weld bead, indicating too much aluminum oxide was present. Inspect your welds for uniformity, appearance and cleanliness, and check the backside of your welds for evidence of sufficient penetration.

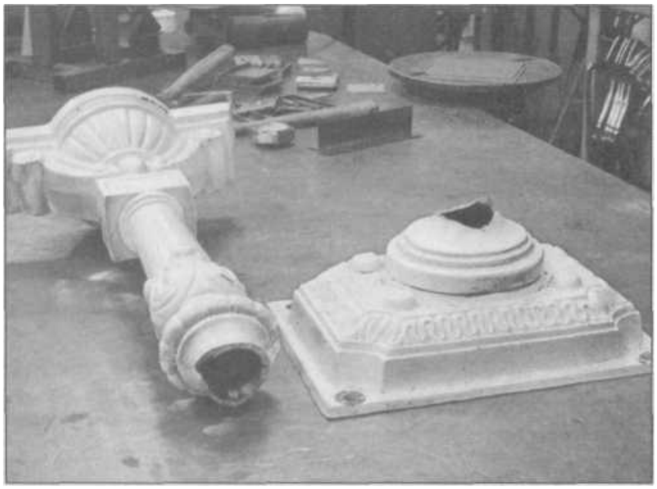

6.34 When welding thick aluminum materials, or aluminum castings like this fancy mailbox post, the edges must be thoroughly cleaned and beveled before welding.

6.35 The rosebud tip on an oxy-acetylene torch is the easiest method for preheating thick and cast aluminum before heli-arc welding. In order not to warp cast-aluminum heads, manifolds and blocks, use heat-markers on the metal while preheating.

When TIG welding any aluminum thicker than 1/8-inch, it is desirable to preheat the aluminum to 350°F before welding **(see illustration).** This is also particularly important when welding aluminum castings such as repairing cylinder heads, water pumps, aluminum engine blocks, bellhousings and other automotive castings. If the casting isn't preheated before welding, it will probably crack again later on, due to the welding section being expanded more than the adjacent metal, and then shrinking. Aluminum castings soak up a lot of heat, so the initial welding is made easier if the part is preheated because the part doesn't draw off the heat you're putting in to make the seam melt **(see illustration).** Preheating on automotive castings also helps bake out any dirt or oils that may be present, although any aluminum castings should be thoroughly cleaned before even attempting to weld them. Even starting with a clean casting and a Veed-out crack, impurities in the casting may keep coming out while you weld.

Ideally, aluminum parts should be evenly heated in a temperature-controlled oven, but some parts won't fit in a regular oven, and perhaps the boss of the kitchen may not allow it anyway. The alternative is to use temperature-sensitive paint sticks from the welding supply. Use one made for 350°F and dab some on various parts of the casting. Heat the whole part with a rosebud tip on an oxy-acetylene torch just until all the temperature indicators have melted **(see illustration).**

TIG-welding stainless steel should prove to be very similar to your practice on mild steel. What makes stainless steel different from ordinary steel is its content of both chromium and nickel, part of the reason for its much higher expense. You will find some characteristics that are different, such as the heat staying in the bead area longer than in plain steel because stainless steel is only about half as conductive of heat as regular steel. Once stainless steel does take on heat, it expands some 50% more than mild steel, which can cause problems of warpage in thinner materials if the pieces aren't clamped or jigged securely during the welding process. In critical welding in industry, jigs may be made of copper because its conductivity helps to carry heat out of the stainless parts quicker.

In terms of setup, stainless steel can be TIG welded with the argon flow set only slightly more than for steel, from 11-13 CFH for thicknesses

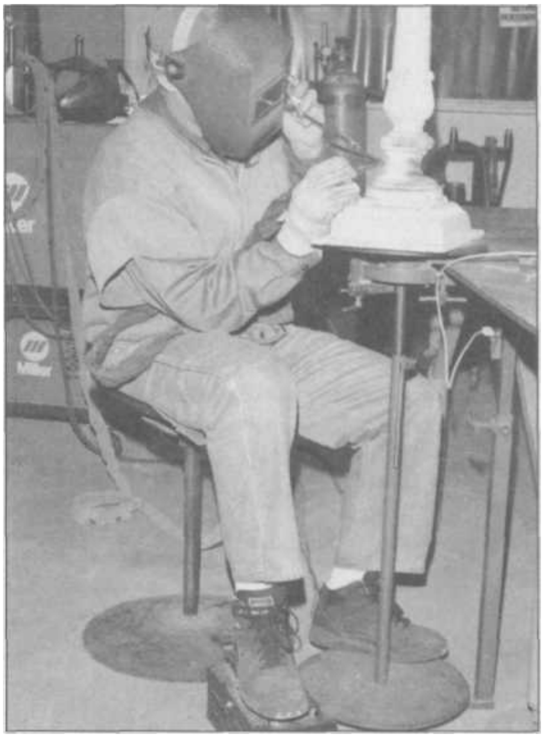

6.36 Weldor Bill Maguire is TIG-welding the heated mailbox post, using his rotating welding stand he invented. At the bottom is a round disc you can spin with your foot, leaving your hands free to continue welding, especially where you have to go *around* something. These should be available for sale at the time of printing.

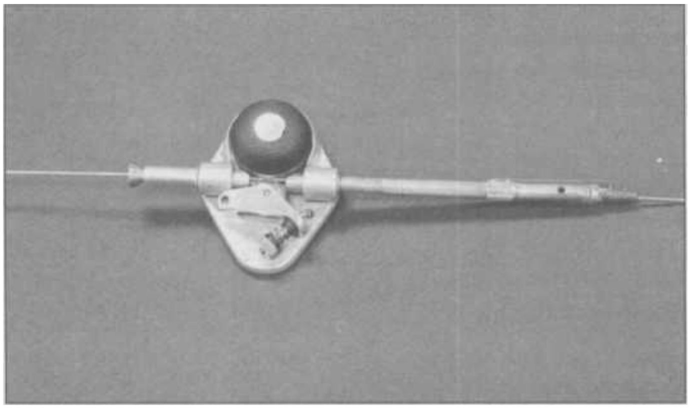

6.37 No one knows better what a weldor would need than a weldor. Another invention of Bill Maguire's is this "Pocket-Feeder", which is patented and for sale through major suppliers. It hand-feeds a welding wire, from .020-in. to 3/32-in., to your weld, while your hand stays the *same,* comfortable distance from the work. Your thumb on the wheel determines the feed speed.

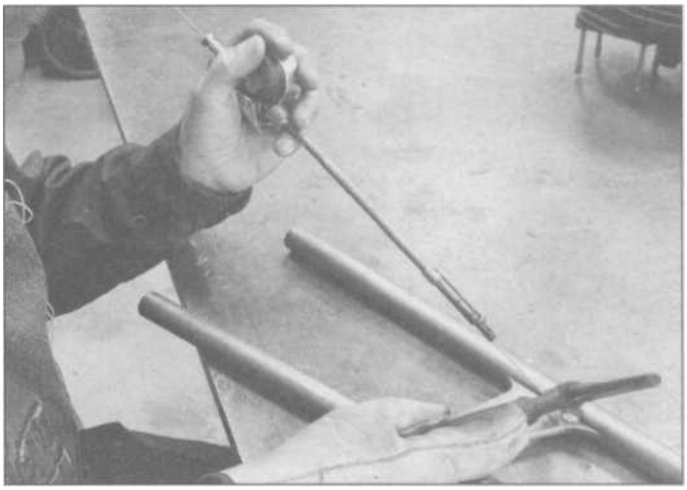

6.38 In use, the Pocket-Feeder can be used with either left or right hands, and you can reload new wire with one hand. Sensitivity on the thumbwheel is good, even with a light glove on.

up to 1/4-inch. Set your current controls for DCSP as you would for steel, with high-frequency for starting only, unless you are working on very thin pieces, in which case you might use HFAC instead because of its lower heat input and lesser tendency to burn through thin material. When using DCSP, the 2%-thoriated tungstens work best, with the pure tungstens good for HFAC on thin material. Of course the welding rods must be stainless steel. Practice on some stainless scraps until you get the hang of handling the different heat-transfer capabilities of stainless. Your puddling and rod-dipping action will be at a different pace than for mild steel.

In industry, critical welds on stainless steel are often made with a setup on the backside of the part to exclude air from the weld area. Of course, the shielding gas on top from the torch takes care there, but on the backside the weld can be contaminated. The pros form a box or shield around the back of the weld area and plumb some extra argon shielding gas in one end and out the other end into a tube to carry it away from the weldor. This will purge any air from the backside of the weld where it could cause any molten stainless-steel to crystallize. This technique is also used in protecting the back of titanium when TIG welding. When they weld critical tubing and can't get at the inside with a shield, they plug each end of the tube or pipe with tape and cardboard and fill the pipe with shielding gas. This is particularly true when welding stainless-steel tubing meant to carry medical ingredients or anything else that must be kept from contamination or oxidation.

Although you will likely never encounter other metals in your home/shop use, there are a variety of special alloys and rarely-used metals that are used often in specialized industries, and for which TIG is the primary or only method found to do the job properly. For instance, in the food-service and canning/bottling industry, there are numerous products that are transported through tubing or pipe. Besides stainless steel, nickel is often used in these applications, for its resistance to alkalines and acids. For the same reasons, it is used extensively in the chemical industry. But nickel-alloy pipe used in these situations cannot tolerate the possibility of a particle of slag, such as from stick welding or MIG welding, becoming included in the weld bead. It would become a weak point when faced with continued chemical attack if high temperatures are involved, until surrounding material is affected and the joint is weakened. For this reason, nickel work must be thoroughly cleaned before welding without using chemical cleaners, or a black nickel oxide coating can form.

6.39 **TIG welding is especially helpful in situations like this race-car suspension upright, where you have thin chrome-moly plate and tubing, and need total heat control. If you burn through on a job like this, the whole part may have to be scrapped.**

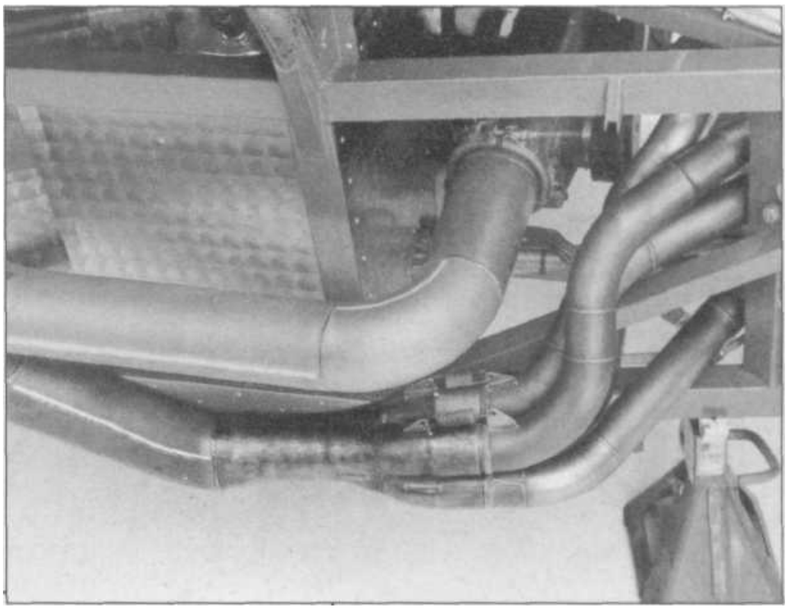

6.40 **Beautiful, precise welds on this race-car exhaust system typify experienced, professional TIG welding.**

For metals like nickel, copper, titanium, and others, TIG welding has proven itself to be the most adaptable method, especially in critical applications. Since there is no flux of any kind used in TIG welding, there is no chance of even minute slag particles being included in the weld bead. Industry is also capitalizing on the fine-grain structure of TIG welded seams (and the heat-affected areas nearest the weld). Helium is most often used as a shielding gas on nickel for its hotter arc, which allows for a higher welding speed, with a gas flow of 8-30 CFH, depending on work thickness. Polarity is usually set at DCSP (except for thin work where HFAC can be used to prevent burn-through), and the high frequency control set for *start-only,* using thoriated tungstens. Filler rods must be nickel or nickel alloy. Oxidation can be kept to a minimum by keeping the filler rod at a low angle to the work and keeping the rod tip in the shielding-gas envelope.

It isn't done often, but TIG welding can be used to repair cast iron. Your welding supply store will have some special rods that are cast of gray iron, called "three-in-one," and these should work for either gas or TIG welding although the same preheating instruction apply as above when we discussed welding cast aluminum parts.

Heli-arc or TIG welding is likely to remain as mostly a shop-only process due to the expense of the equipment and the length of the learning curve, but don't discount it as something you can't try. If you know someone in a shop with TIG equipment, ask them if you can spend a few hours closely observing them (with a suitable spare helmet on, of course) performing TIG welds. The lack of spatter and smoke from TIG welding means you can observe this process better than most other welding methods, and this will tell you much more about how the process works than a chapter in a book. If you feel you have observed enough to get a good idea, see if they will let you try it out, but first offer to pay for any tungstens you ruin in your practice. You may find you have a hidden aptitude in hand-eye coordination, and TIG may be a process that appeals to you.

Notes

7 Plasma cutting/welding

Just as the emergence of lower-cost, higher-efficiency MIG machines has made something of a revolution in small shop and home/shop welding in the last ten years, so has plasma technology changed the face of metal *cutting.* While the technology has been around for a while, it has been seen by most weldors as a high-technology setup that seemed complicated or hard to use. Plasma technology's biggest usage had been in plasma arc welding, or PAW, and it is still used in this form in many industrial applications.

The simpler plasma-arc *cutting* (or PAC) technology got very little play at the time, and it has taken some years for the equipment to filter down to the general welding/cutting marketplace. As with the MIG welders, what was originally considered to be a professionals-only setup has entered the province of the small shop and the amateur fabricator, due mainly to the introduction of less-expensive imported equipment that really opened up new markets with hobbyists. The domestic welding manufacturers suddenly became aware of these newer markets for smaller versions of equipment they had been selling to industry all along and brought increasing competition to the fray. Today there's a wide selection of plasma machines to choose from, from small 110V AC portables to large shop equipment for materials up to one inch thick.

7.1 **Plasma cutting can be done outdoors, being very little bothered by wind, makes very clean cuts on ferrous and non-ferrous materials, and produces no smoke and very little sparks. It has become the modern way to cut metal. The operator here should be wearing gloves and a long-sleeved shirt.**

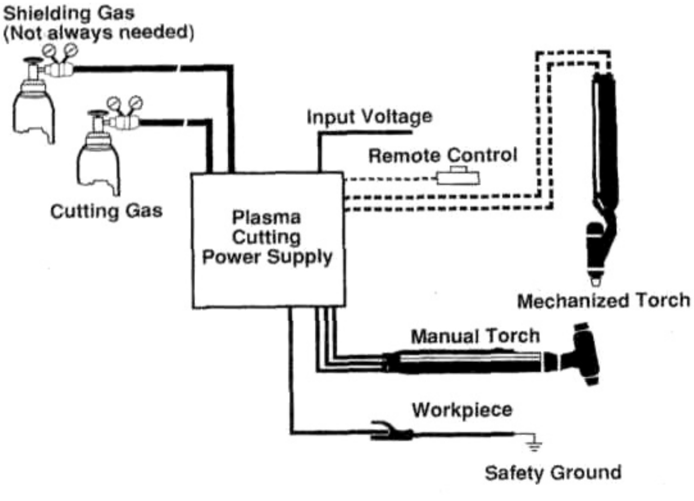

7.2 **This illustration shows the basic components of the plasma-arc layout. The two bottles of shielding gas are really only required on the large industrial machines, not the home/shop portables which run on compressed air. The medium and larger sized plasma cutters can be used with a straight (mechanized) torch body for trace cutting.**

Usage today is increasingly with the cutting side of plasma technology. With an incredibly easy learning curve and not a lot of special installation costs, the benefits of plasma-arc cutting, usually just called plasma cutting, are being reaped today by body shops, muffler shops, wrecking yards, sheet-metal shops, duct repairmen, street rodders, race-car builders and fabrication shops. Among the benefits are use of plain shop air-compressor air, no dangerous gasses, no flames, smoother cuts, less heat distortion in the edges you cut, little oxidation of the edges and incredible speed.

In basic principle, the plasma process uses a torch something like a TIG torch, but the electrode is deeply recessed up into the torch, where it cannot contact the work and be contaminated, and the arc is maintained as a needle-like point when the air between the electrode and the work is ionized. The air or gasses around the electrode are heated to such a point that they reach the *plasma* state, in which the gasses themselves can *conduct electricity.* While the gas is ionized and making continuity for the arc process between the electrode and the work, another gas is introduced around the nozzle. In the case of plasma welding, the gas is an inert shielding gas; in the case of plasma cutting, it is compressed air. In either case, welding or cutting, the action takes place in a very narrow stream.

Plasma-Arc welding

The original plasma technology was developed in the laboratories of the Linde Air Products Company (a pioneer in many welding processes, now the L-TEC Company) back in the late Fifties. They and other companies have worked on refining the equipment and the process ever since. You won't find plasma welding equipment at most welding supply stores, but it is in a few industry catalogs. This is the kind of equipment usually sold through a company representative. Most of this equipment is used in high-precision joining of very thin materials. Although, up until the last ten years, both plasma welding and cutting have been virtually reserved for industrial purposes only, and most of the plasma-arc welding equipment is still expensive and for limited industrial usage, the process is worth looking at.

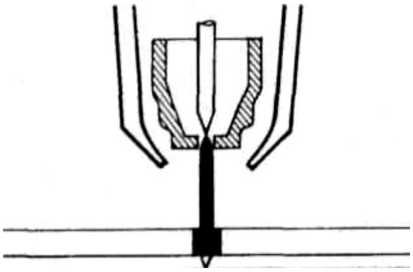

7.3 In the needle-arc version of plasma welding, the narrow-and-tall arc forms a weld that makes 100% penetration, top and bottom. The technique is used mostly in precision industrial applications.

In operation, the plasma torch has a central tungsten electrode tucked up inside a nozzle that is more restrictive than other welding gun nozzles, that is closer to the electrode and increasingly tighter as it gets close to the tip of the electrode, until there is just a hole for the arc to come through **(see illustration).** An "orifice" gas is forced through this nozzle, and it speeds up like air/fuel mix going through a carburetor's venturi. The gas is usually argon.

Inside the welder's power source is a high-frequency arc-starting system not unlike a conventional TIG machine, and this initiates the arc, since the electrode doesn't stick out and can get close enough to the work metal to start the arc itself. This is called the "pilot-arc" circuitry, and it generally cuts out after the arc is established. The current in the electrode superheats the argon until the gas itself conducts electricity, and a long, extremely hot arc is created that welds the work. A second inert gas is introduced in a large nozzle around the inner one, such that a cone of shielding and cooling gas is formed around the long arc column to protect the weld area from contamination, much as in other electric arc-welding processes.

There are several methods of welding with plasma technology, a conventional method akin to TIG welding where filler rod is added, a needle-arc process where parts are fused without filler rod, and a keyhole method where a weld without filler rod can be made that actually fuses the parent metal on the whole mating surface at the same time, top and bottom. The latter produces as pure and uncontaminated a joint as possible. The benefits of using plasma-arc welding over conventional TIG welding is that it is easier to operate as far as maintaining the arc length exactly and the weldor has a better view of the process.

Although you may find one or two PAW units in a catalog, most of the technology is used in industrial situations with mechanized torches and water cooling with amp ranges up to 300 or 400. Most manual-operation plasma welders do not have as much amperage as conventional welders; L-TEC's catalog shows a 110V 15-amp unit, and a 100-amp 220V three-phase machine weighing some 730 pounds that only welds up to 1/8-inch plate, although it has a 100% duty cycle at maximum output.

For those reasons, the plasma-arc welding process is probably not for your home/shop. Much more flexible, lighter and more affordable conventional equipment is available for making welds. However, there is no question but that the practical application of the technology in everyday situations is eminently provided by plasma-arc *cutting* equipment.

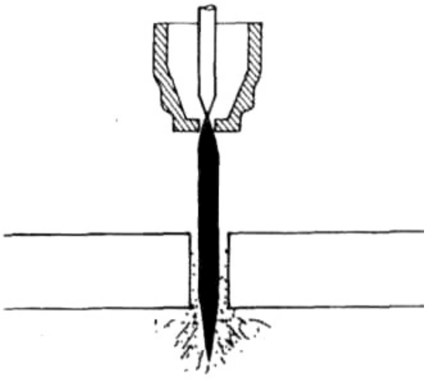

7.4 In plasma-arc cutting, there is usually only one gas (in all but the biggest machines), and that is air from the shop's air compressor, which is forced around the electrode to blow the kerf out neatly.

Plasma-Arc cutting

In plasma-arc cutting, the technology is much the same as welding, but the arc is even more constricted, with the plasma temperature so high that the arc can cut *any* metal. The very restricted arc cuts a pencil-thin line through the metal, and a sharp force of gas through the nozzle blows the melted metal from the kerf **(see illustration).** The velocity and steadiness of the cutting action is such that the edges of the cuts are kept very straight and clean.

The benefits of plasma-arc technology in cutting metal are many. The cuts are very clean, with little oxidation of the metal as found in oxy-acetylene cutting, the velocity of the action creates smoother edges that need a great deal less cleanup, and in fact, with the industrial units that are mechanized much like a gas flame-cutter or pattern-cutter, the cuts need no cleanup or preparation for welding **(see illustration).** There is little or no slag found on the bottom edge of the kerf.

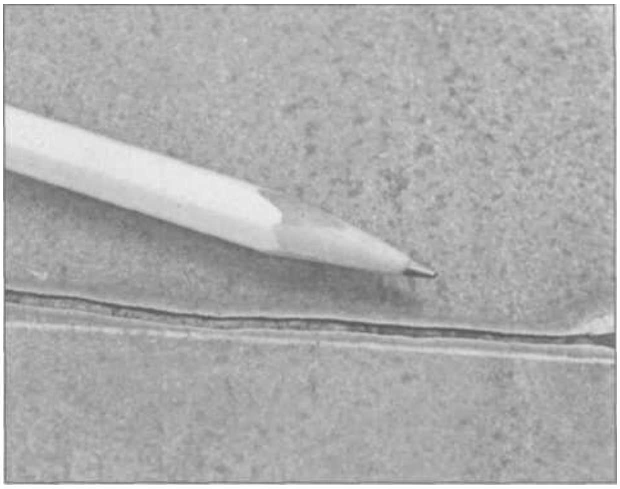

7.5 Plasma cutters have become very popular with auto bodymen for the cleanliness and speed in cutting body panels. These sheet-metal edges need very little prep, and a plasma cutter can cut painted or rusty metal just as easily.

Plasma-arc cutting is considered much safer in many work environments than gas cutting. There are no flammable gasses involved, there is no smoke from the cutting action to bother the weldor, he can watch the process much more closely and there is much less spray of sparks coming from under the cut. If you remember the little sparking toys you had as a kid, where you pushed a plunger and a wheel went around, scraping against a tiny flint making sparks, that is a small-scale version of what you see when plasma cutting. This is in high contrast to gas cutting, where there is a lot of heat generated, and blobs of molten metal are constantly falling on the floor under the cut, creating a fire hazard.

There is *some* heating action in plasma cutting, but in a short time, depending on the thickness of the material you have cut, the parts are ready to be handled for further work. With a gas cutting torch, the parts are still red-hot. Some have described the plasma process as almost "cold cutting." Another benefit of the cooler cutting action is that much less heat is induced into the parent metal, so when working on thin materials there is greatly reduced warpage.

This is what has made the plasma cutter so popular with automotive bodymen and sheet-metal workers. Thin materials can be cut with virtually no distortion. The cuts can be made, as one bodyman described it, "as fast as you could draw a line with a pencil." A few minutes with a file or sandpaper and the edges are ready for some other work, whether to be welded or left as a finished edge. Because of the lower heat and reduced sparks, panel-cutting on cars is much easier, even if there is a secondary panel on the other side because there is less damage to what's on the other side like other painted surfaces that need to be left untouched. Bodymen also like the fact that you can make a cut, lay down the torch, do something and pick up the torch again and start cutting with no other functions to perform other than putting your safety helmet down and pulling the torch trigger. With a gas cutting torch, there is always the danger of what to do *between cuts* with a torch flaming full-bore. If you keep shutting it off each time, you have to go through the process of re-lighting and adjusting it every time.

The last reason is why plasma-arc cutting has become the tool of choice now in wrecking yards across the country. In a yard full of discarded cars with flammables like upholstery still in them and various pools of oil, gasoline and other fluids soaked into the ground all around, oxy-acetylene cutting on cars in wrecking yards has always been a hazardous job. Also, the torches and bottles are a heavy setup and had to be hauled around to various parts of the wrecking yard. With a good plasma machine, the 220V cutter can be plugged in a small shed central to the yard, while a plasma torch cable as long as 234 feet could be used to cut almost anywhere with a lot less fire danger and much faster speeds, 10 or 20 times faster than with acetylene equipment.

Because the plasma machine has arc-starting circuitry built-in, most plasma torches do not need to contact the metal to get started, which is a decided advantage in body work. Once the arc is started and the gas is ionized and conductive, a plasma torch can cut through rusty metal, painted metal, even sections of a car body that have Bondo or undercoating on them. When using an oxy-acetylene torch to cut painted sheet metal, the cut edges are blackened, scarred and blistered, while with plasma the edges have a gray discoloration near the edge and that's it.

In fabricating shops, one of the biggest advantages of plasma-arc cutting equipment is the fact that all metals can be handled equally, from steel to aluminum, to stainless and plated or galvanized metals. It is the nonferrous metals like aluminum and austenitic stainless steel, which are normally hard to cut cleanly with other methods, that make the plasma machines appealing to shops. The author remembers his first experience with a plasma cutter some years back. An aluminum motor plate was being made for a race car, and we needed to cut a 14-inch hole out of the center to clear the flywheel. The material was 1/4-inch-thick aluminum and 7075-T6, which is so hard you can have difficulty even drilling it unless you have a fresh bit. The idea of using a sabre saw or band saw seemed tedious and out of the question. We took it to a shop that had a plasma cutter and within minutes of making a plywood circle template to run the torch around, we had our hole cut with very little heat and perfect edges. It was eye-opening, and when you try one you may be similarly amazed.

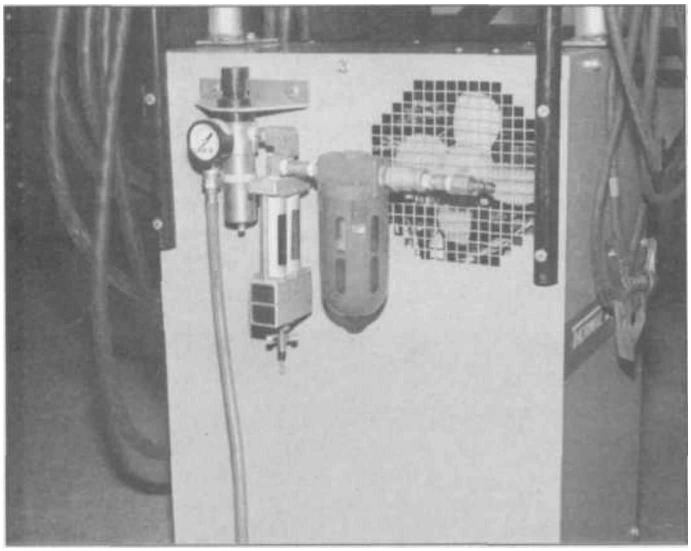

7.6 Shop air-compressor air is all that's required to run most plasma cutters. Here on the back of a Thermal Dynamics machine is the air filter and pressure regulator.

When plasma-arc cutting equipment first came into wider, less-industrial usage, some traditional fabricators and bodymen were intimidated. The technology wasn't widely understood, and the term "plasma" seemed to indicate some highly-technical process akin to esoteric lasers. Once the machines got out into the field, they quickly became accepted for their speed, cleanliness, lack of maintenance, lower fire hazards and the ease with which a weldor can operate them. It's almost a pull-the-trigger-point-and-shoot kind of process.

Perhaps one of the bigger attractions of plasma cutting equipment we haven't mentioned yet is the type of gasses used in the process. While the big industrial plasma machines do use traditional shielding gasses, like argon for the arc "pilot" gas and hydrogen, nitrogen or a gas mixture (like 35% hydrogen/65% argon) as the cutting gas, most portable plasma units today use nothing more than *compressed air* from a shop air compressor **(see illustration).** There are no expensive bottles to buy or periodically refill.

This is how the shop-air works. When the gas in the arc column is ionized, it becomes the arc medium between the work and the electrode. Nitrogen happens to ionize well and makes for a hot arc, and also happens to make up about 75% of the air around us by weight. Most plasma cutters use a high velocity of shop compressed air to become the ionized arc gas, the cutting gas that blows the material out of the cut, and the cooling medium. It can be assumed that all shops already have a heavy-duty air compressor on hand for operating other equipment, so using shop air greatly simplifies the use of a plasma cutter and keeps expenses down. In industrial situations, pure nitrogen or hydrogen or a mix is used for totally clean cuts in precision work, but the shop-air cuts are more than good enough for any automotive or fabricating-shop usage.

Choosing plasma cutting equipment

The large, expensive and heavy industrial plasma cutters and welders are for specific purposes in high-precision and high-production mechanized work, often requiring two gas bottles, three-phase power for 200-600 amps, and use water cooling for the torch as well. For these reasons, we'll concentrate on the wide array of more portable plasma cutters that are available for home and shop use, most of which operate on 220V, single-phase current and use only shop compressed air for cutting and cooling.

At the most basic level, where a home/shop user might be interested, there's a number of domestic and imported plasma cutters that weigh 75 pounds or less and are capable of cutting material thicknesses that an automotive user would come across. If you are comparing plasma cutters to welders, don't be alarmed at the amperages on the smaller plasma machines. Units with 50-amp capability can cut steel up to 3/4-inch thick. An area where there is a comparison with welders is that smaller plasma machines have lower duty-cycles and their cutting speed slows down when thicker materials are cut.

In the 110V category, Daytona Mig offers the Pocket Plasma , which puts out 25 amps at a 20% duty-cycle and cuts sheet metal at 100 inches per minute, though it is said to cut up to 3/16-inch steel material. It requires 70psi air input at a rate of 4-6 cubic feet per minute (cfm). Miller makes several 110V units, with the Spectrum 187D cutting up to 9/64-inch steel material and weighing 47 pounds. Thermal Dynamics starts their extensive line off with the DynaPak 110, which cuts up to 3/16-inch steel, uses all solid-state circuitry, sequential status lights, and has an amazing 80% duty-cycle at 20 amps (see illustrations). With all of the machines available, the maximum cutting thickness is usually given for mild steel, while aluminum and other non-ferrous material take more amps and the same machine can only cut a lesser thickness of these metals. On some machines, a figure may be advertised that gives a maximum thickness for cutting

DAYTONA MIG
PLASMA
CEBORA

7.7 The compact end of the plasma cutter market is represented by 110V machines such as this Plasma Pocket from Daytona MIG. It puts out 25 amps, weighs 55 pounds, and requires 4-6 cfm of air at 70 psi.

L-TEC PCM-VPi
L-TEC

7.8 Another quite-portable plasma machine is the PCM-VPi from L-TEC, which offers cuts up to 5/8-inch in steel, variable amps from 10-40, and a 60% duty-cycle at its rated maximum. It operates on 220V current.

Miller
Spectrum 187D
WARNING
Miller Spectrum 187D

7.9 One of Miller's smallest plasma cutters, the Spectrum 187D weighs only 47 pounds, yet cuts up to 9/64-inch materials, and runs on 110V house current.

MATCO TOOLS
PLASMA CUTTER

7.10 The popularity of plasma-arc equipment has really jumped in the last five years. Now imported units are sold here, and tool companies like Matco have them, as well as the traditional welding equipment manufacturers.

OFF
ON
HTP
Micro Cut
250
HTP

7.11 HTP's entry-level plasma cutter takes 220V current, but weighs only 56 pounds and is said to cut up to 1/4-inch steel. It has a 100% copper-wound transformer.

and another, higher figure for severing. The severing action is a short cut, such as when cutting a length of material.

Beyond the 25-amp range that seems to be the limit for 110V equipment, there are machines for 220V input that are still quite portable. HTP America offers affordable plasma cutters from their Micro-Cut 250 that cuts up to 3/16-inch steel, to the Micro-Cut 375 which can be used on 110V, 220V or 460V input and offers 7.5-35 amps, cutting up to 3/8-inch steel, as well as larger shop units **(see illustration).** L-TEC has a popular unit, the PCM-VPi, which weighs only 46 pounds with its plastic case and carrying handle, but is said to cut up to 5/8-inch steel with its 50 amps of power, and features a 60% duty-cycle at 40 amps output **(see illustration).** It is a good choice for any automotive shop, and can sever up to 3/4-inch material such as re-bar for construction sites. Lincoln's entry-level plasma cutter is the Pro-Cut 40, with a 60% duty-cycle at 40 amps, 220V input, cuts up to 3/8-inch steel, and offers a timed pilot arc and timed post-flow. Daytona MIG's medium-sized lineup includes a 35-amp, 220V Plasma Prof 35 that is said to cut up to 3/8 steel quickly, and 1/2-inch steel at 15 inches per minute.

Above the 50-amp range, plasma-arc cutting equipment is in the serious shop equipment arena. Most of these machines are for 220V, *three-phase* power input, which you won't have at most home garages or small shops. Within this large group of choices though, there are cutters with total amperage ranging from 50 amps to 150 amps or more. In this category, you'll find the Lincoln Pro-Cut 125, with 125 amps, 60% duty-cycle, a microprocessor-controlled overshoot limiter, and "Drag-Detect" circuitry that is said to prevent consumable damage. In this machine, let's look at the duty cycles for a professional-grade plasma cutter. Rated at 60% duty-cycle at its highest 125-amp output, it has an 80% duty-cycle at 110 amps, and a 100% duty-cycle at 100 amps or less. At its highest output, this machine can cut 1 1/4-inch steel, and for ordinary 1/2-inch plate, it can cut all day long. That's the difference in the heavier professional machines; they have stronger components to work heavy materials not just a few times a day, but all day, which is why they are overbuilt and too expensive for the home/shop market, even disregarding the three-phase input power required.

Other powerful plasma-arc cutters include: HTP's PCA 65, with 60 amps (220V single-phase), cuts up to 3/4-inch steel (clean cuts on 5/8-inch steel), a 50% duty-cycle at 60 amps, and a weight of 180 pounds. HyperTherm has their 100-amp MAX 100, that will cut 1 1/4-inch steel on 220V or 440V three-phase power, with CC (constant-current) transistorized output, ESAB offers the PCS-90 that cuts up to one-inch steel with three-phase 220V or 440V, and has air pre-flow, post-flow and a low-air-pressure sensor. L-TEC has the PCM-1000i, which can cut one-inch steel at production speeds with its 80 amps, and a 70% duty cycle at that maximum amperage, yet weighs only 80 pounds. Daytona MIG has a full line of heavy-duty plasma cutters, with the Plasma Prof 120 being near the top with 120 amps (220V three-phase), cuts up to 1 1/8-inch steel and 7/8-inch aluminum, with quick-disconnect torches so you can switch to a mechanical torch for a cutting table (see illustrations).

Most of these machines offer such features as infinitely-variable output control, which means that you can dial in the exact requirements for a job, say cutting through an outer skin on a vehicle without cutting panels on the other side. Of course, you can cut thin materials with more power than

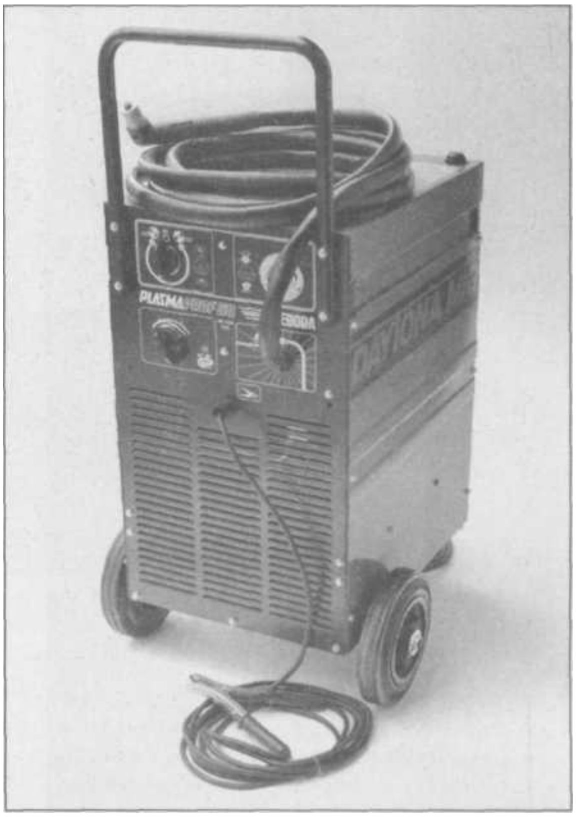

7.12 **The bigger-amperage machines are all in the 220V category, and many, like this Daytona MIG Plasma Prof 50, require shop-type three-phase current, too. This one has a 35% duty-cycle at its rated max of 50 amps and 100% at 30 amps.**

PCM-750i
L-TEC

7.13 **Still light and portable, but powerful enough to cut 3/4-inch steel, The L-TEC PCM-750i has 50 amps with a 40% duty-cycle at that setting, and has its own torch storage spool and spare parts storage compartment.**

PLASMA
150
DAYTONA MIG

7.14 **The larger plasma-arc cutting machines are all in the three-phase power category, like this Plasma Prof 150 from Daytona MIG. It weighs 317 pounds, has a range from 20-150 amps, and can cut 1 1/2-inch steel and 1-inch aluminum at 30 inches per minute.**

is required, but using just the right amount means you have the least danger from flying sparks. On the less-expensive end of the plasma cutter spectrum, some of the machines offer only a knob with several predetermined positions.

Beyond the 80-120 amp range of plasma cutters, where really large plates are cut all day long, the equipment offered is no longer usable with shop air, requiring two inert gas supplies and water-cooling for the torch. However, they offer from 100 amps to as much as 600 amps (the higher amperages are usually found on machines built to be used with mechanized torches), and all with 100% duty-cycle. Obviously, that kind of power doesn't arrive at your doorstep without a hefty invoice.

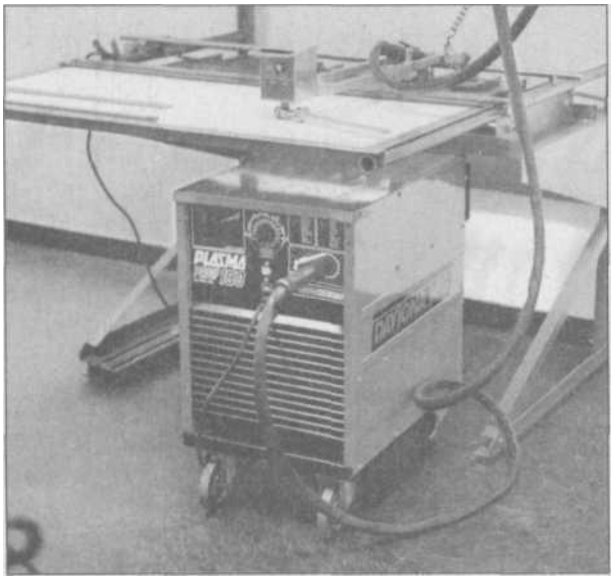

7.15 Larger machines are usually offered with an optional 180-degree (straight) torch for use with shape-cutting machines for industrial work.

Using a plasma cutter

If you thought the modern MIG welder brought great advancements in ease of the learning and operating curves, wait until you try plasma-arc cutting. It's as much of an improvement over gas-torch cutting, in both learning curve and reduced maintenance. We watched as a race-car mechanic in a large shop was being trained by another worker to utilize their plasma cutter. The newcomer to the technology made his first 4-inch by 6-inch part from 3/8-inch steel plate after only fifteen minutes of instruction. After experimenting with different speeds, power levels and materials, he was ready to do real jobs after an hour. You can't learn everything there is to know about the process in your first day, but this story is typical of first-time users.

Unlike welders, most plasma cutters of the shop-type variety are ready to use when you uncrate them. Their power cord is ready to plug in (you don't have to put on your own plug), the torch is usually already hooked up to the front of the machine, and all you have to do is connect your shop air supply before making your first cuts.

Although the risk of fire or injury is less with a plasma cutter than with a welder or oxy-acetylene cutting torch, you must still take every precaution before starting. Read over the instruction manual that came with your machine before doing anything. When you are familiar with all the controls, take the torch apart, so you'll understand how it works. Specific design of electrodes and other parts may vary from make to make, but, whatever the make, you'll notice that the electrode is like no other you've seen. Instead of a six-inch-long rod of tungsten as with a TIG torch, the plasma electrode is a machined part.

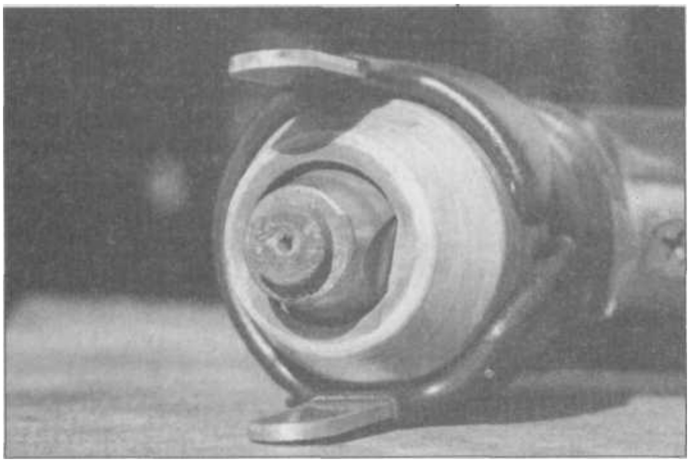

7.16 Eventually, enough material will blow back from cutting to cause a building of slag on the exposed nozzle and orifice of the torch, requiring periodic maintenance.

Although it never comes in contact with the work, the plasma cutter's electrode should still be considered a "semi-consumable" item. Electrodes can last a long time, but keep several spares on hand because they do get worn and contaminated with material blown back from the cutting **(see illustration).** Besides the electrode, you'll find several other pieces inside the plasma torch, including the nozzle (also called a shield cup, retaining cup, and gas diffuser), and a cutting tip, which is what converges the airflow past the electrode **(see illustration).** Different brands of torch may have small springs, insulators or swirl rings, but the basic principle is the same for all brands.

There is much less danger with a plasma cutter than other equipment, yet there are still hot sparks generated. There is also ultraviolet radiation. Because the process doesn't seem as bright or intense as welding, some users

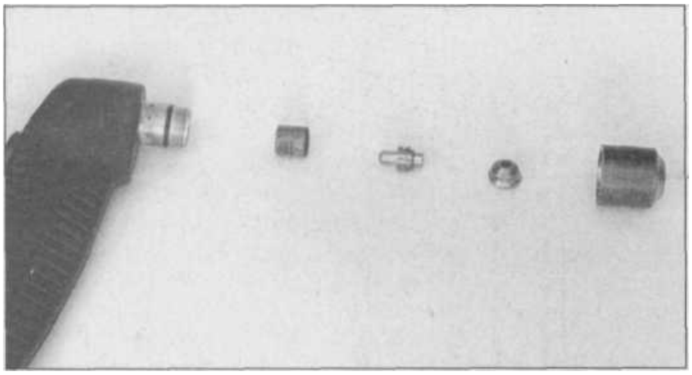

7.17 The main components of a typical plasma-arc cutting torch are, from left: torch body with o-ringed and threaded extension, the swirl ring, the electrode, the cutting nozzle, and the retaining cup. Some brands have a slightly different arrangement of parts inside.

7.18 Hot rodders never leave anything alone. Rugged Thermal Dynamics 50-amp plasma cutter used in a rod shop has been modified with the addition of a tubing rack and box on top to hold nozzle dip and plasma consumables.

are tempted to perform plasma cutting without proper safety wear, but your skin can still be burned. So wear long-sleeved shirts with the collar and cuffs buttoned up, and wear leather welding gloves and a full-face shield. If you ever have to work on your machine, it should be unplugged, even if all you are doing is changing the electrode. The machine's output is high enough voltage to cause severe shock and injury. Also, if you are cutting any painted, plated or galvanized metals, ensure that you have adequate ventilation. The plasma process doesn't create fumes by itself, but what's on the surface of the metal can.

On most machines, there is a high frequency arc-starting circuit that makes it unnecessary to touch the work with the torch to initiate the arc. Set your output voltage at the machine, depending on the thickness of the metal you are cutting. On some of the smallest machines, there may just be an on/off switch, or a switch with two or three amperage positions. Most machines have a variable knob spanning from the machine's minimum amperage to its maximum.

Use a guide of some kind to make straight practice cuts **(see illustration).** Place the torch so that the outer cup rides on your guide with the nozzle the suggested distance from the work and just push the button. The arc will start, it will

7.19 You'll need to keep on hand some spares of each of the torch components, including wrenches for changing the tip and the electrode.

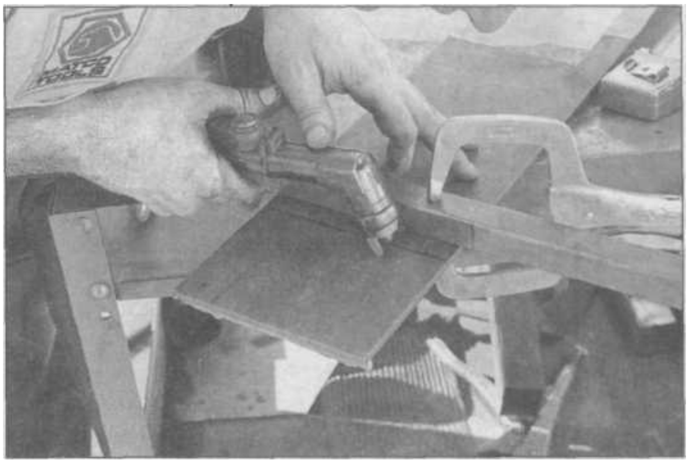

7.20 The plasma cutter makes very clean cuts, so any raggedness on yours means your hand wasn't steady. Use a sturdy guide whenever possible. With a piece of tubing as a straightedge, here a dry run is being made to ensure the torch follows the correct cut line.

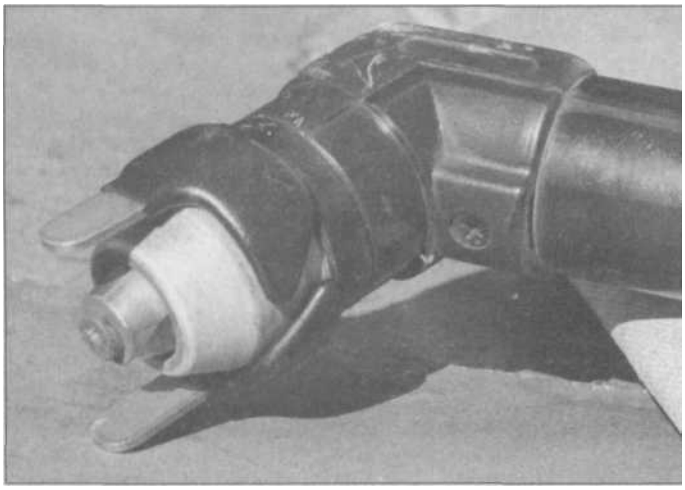

7.21 This clever tool snapped onto a Miller plasma torch holds the torch just the right distance away from the work, so if you have this and a sturdy guide, your cuts should be almost as good as machine-operated torches.

7.22 Same tool as the previous photo, showing how it holds the tip away from the work, and provides a good surface to lightly drag against your pattern or straightedge.

pierce the metal (no need to go to the edge of a plate to get the cut started) and begin blowing the slag out the bottom. Move along at a steady speed and you're *cutting.* You'll develop a feel for the correct cutting speed for various materials, but the instructions with your machine should tell you the recommended speeds for various thicknesses and materials. Practice moving the torch a few times without the arc, just to see how fast or slow 15 inches per minute may be, and what 100 inches per minute feels like.

Examine the cut edges of the material. If the lines in the cut are not straight up and down, your speed was off **(see illustration).** If the lines turn to the direction the cut was made from, your cutting speed was too fast. You can also have someone else observe you from a short distance while you are cutting. They will tell you if the sparks went straight down (proper speed), or sprayed more to the direction the torch came from (too fast). Moving too fast can cause premature failure of the nozzles.

When the part you have cut is cool enough to pick up, examine the kerf. If your cutting speed and amperage was right, the small amount of slag at the bottom of the kerf can be easily clipped off with wire-cutter pliers, and the rest of the metal should take minimal cleanup.

In cutting shapes, you can use plywood patterns to cut repeated shapes such as round or square holes or anything else. Just make the template the right size for the torch head to follow. For basic circles, most welding supply shops have a circle-cutting tool (sometimes called a circle-cutting "compass") that can be used with a plasma cutter.

By far the most important thing to maintaining and operating your plasma-arc cutter is the condition and quantity of the air it runs on. Read the manufacturer's recommendations for the size of compressor you need and the pressure and cfm requirements. If you try to make a cut when you don't have enough air, most machines will either not start or make a grinding-like noise inside. Besides having *enough* air, plasma cutters also require *clean, dry* air. Most will come equipped with a drier/regulator right at the air-line connection on the back of the machine, but you also need a good drier on your compressor. And, if your compressor is

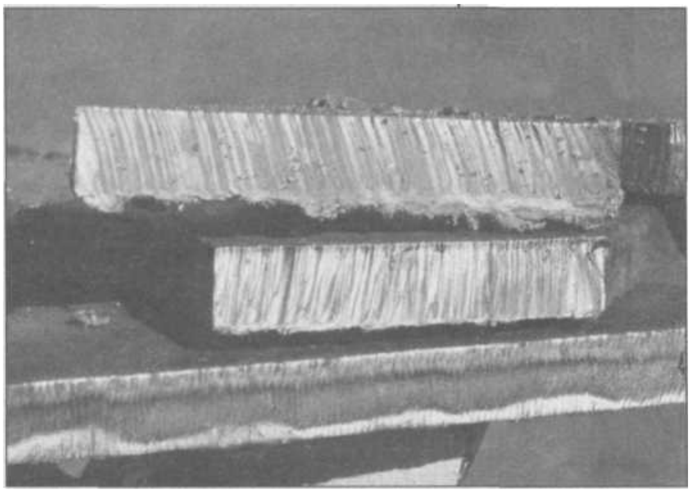

7.23 At top here are two cuts made with a hand-held plasma cutter. In the top one the torch was moved too fast and with not enough amperage. The middle plate had the speed more correct. The bottom plate for comparison only was oxy-acetylene cut with a machine, so this is the smoothness you are trying to duplicate. Only a gas torch cut would have this discoloration of the metal showing.

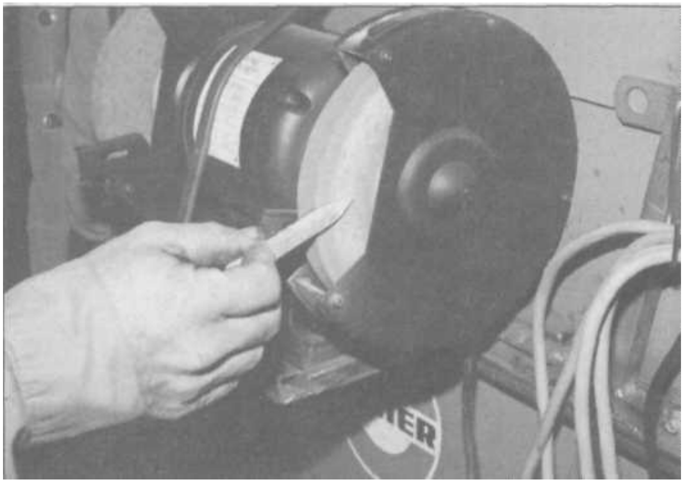

7.24 As with traditional torch cutting, soapstone is useful to mark your cut lines because it won't burn off. With the reduced heat of plasma, you could also use a marker pen. Sharpen the soapstone before drawing.

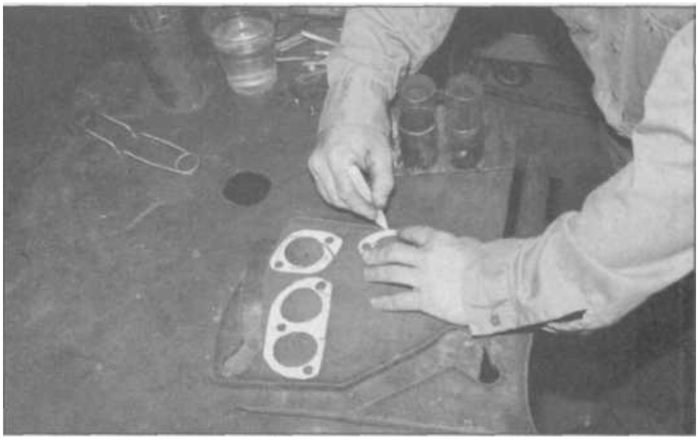

7.25 Here an exhaust gasket is being traced to make thick header flanges from a piece of 3/8-inch steel plate.

very far from where you hook up the flex line to the plasma cutter, it's a good idea to install one on the wall just ahead of the connection for the flex line **(see illustrations).** Your compressor and the driers should all be drained every night to remove dirt and moisture. Moist air will rapidly deteriorate the electrodes and nozzles.

Cutting too fast or trying to cut too thick a metal for the size machine you have will cause some slag buildup inside the nozzle of a plasma cutter, not unlike the buildup inside a MIG gun. In fact, you may use the same anti-spatter spray or dip in your plasma cutter as you use on your MIG torch to make the slag buildup easy to remove. Eventually, even if you take care of your plasma torch, you will have to replace nozzles and electrodes because usage will enlarge the holes, making for a less-concentrated arc. If any slag is "fried" onto the nozzle or electrode, sand it off by holding the part squarely against sandpaper on a flat surface, and then clean out any orifice holes.

7.26 The 50-amp plasma cutter makes short work of cutting out the patterns and can easily pierce the center of a pattern to cut out internal holes and shapes.

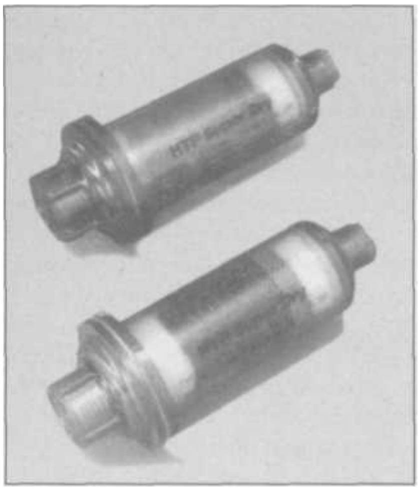

7.27 Dry, clean shop air is extremely important to extending the life of your plasma nozzles and electrodes. These HTP Super-Dry air driers have a bronze filter and desiccant inside that changes color when it needs to be changed.

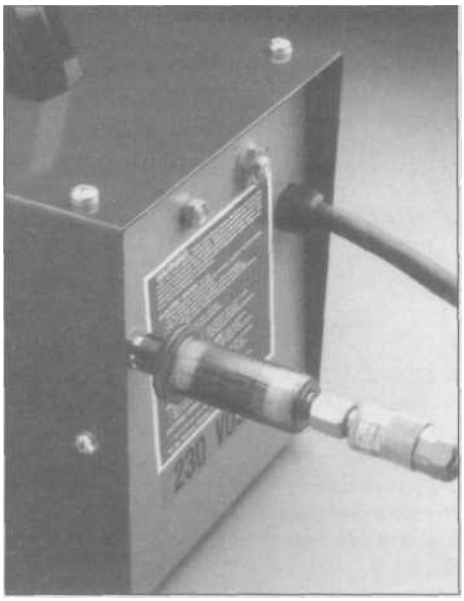

7.28 In most applications the HTP Super-Dry disposable filter/driers can be screwed right onto the rear of your machine.

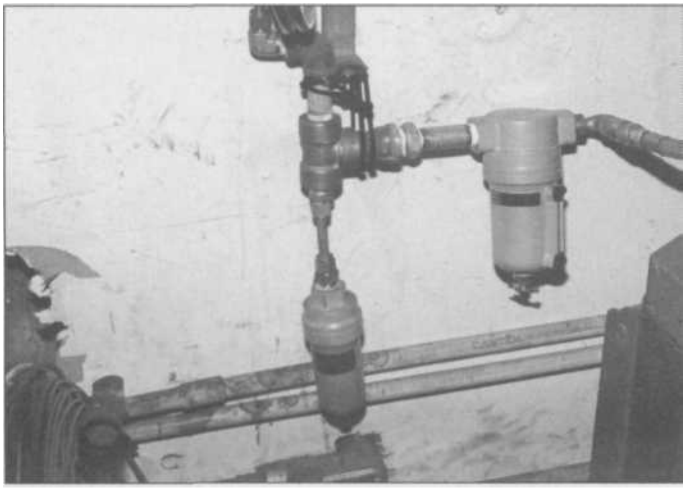

7.29 You can't have the air too dry for use with a plasma cutter. If your compressor is outside or a long distance away, use an extra filter/drier on the wall where the hard line switches to the flexible hose for your plasma cutter. Drain daily.

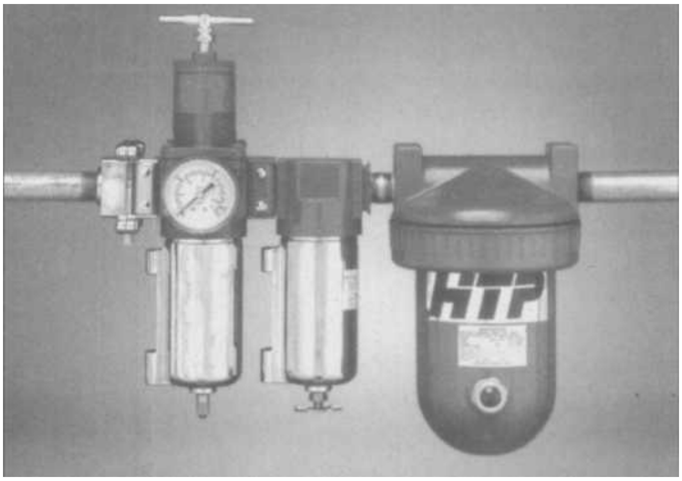

7.30 The Max-Dry from HTP is a three-stage system. Stage one is a self-draining 5-micron filter, stage two is a "coalescing" filter, and the final stage is another filter with a large amount of desiccant. The desiccant changes color when it can't absorb any more humidity, but can be dried out in an oven for reuse.

You'll find there isn't too much to discuss further about how a plasma cutter operates. They are at the same time technologically advanced, and yet simple to operate and maintain. They perform exactly as they are claimed to, in terms of speed, cleanliness of the cuts, lack of cleanup required and the minimal heat induced into the adjacent parts. Plasma-arc cutting is most definitely the preferred modern method of cutting metals, in situations ranging from an air-conditioning duct repairman or an automotive bodyman working on gauge metal to industrial plants where shapes are cut from thick plates all day long. To most automotive enthusiasts and race-car shops, the modern shop combination is a good MIG gun for welding and a companion plasma machine for cutting **(see illustration).**

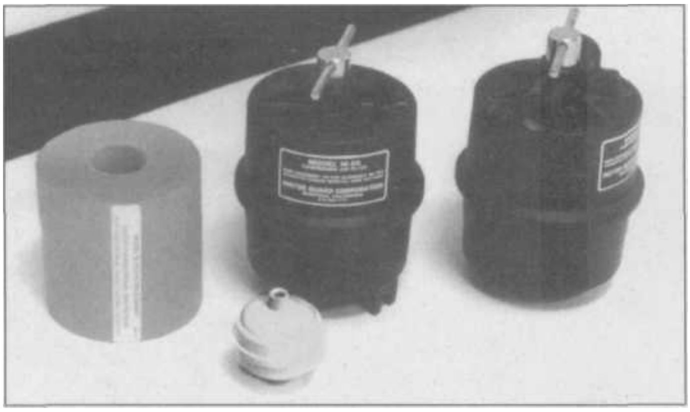

7.31 Daytona MIG offers several air filters, from left: the disposable filter element, a wall-mount filter housing, machine-mount filter housing, and in front a small disposable filter.

POCKET turbo 130
PLASMA

7.32 Given his preferences, today's hot rodder, race-car builder, farmer or fabricator would like to have one of each - a reliable, easy-to-use MIG welder, and a safe, clean and equally easy-to-use plasma cutter.

8 Safety & shop equipment

When it comes to equipment for working with metals, there's a lot more required than just a means of cutting metal and fusion-joining it. Many automotive enthusiasts are "tool addicts" to begin with and welding just adds another dimension to the want list.

Other than your welding/cutting equipment itself, the two main requirements for a good environment are the proper safety gear, particularly the goggles or helmet, and a clean, safe, well-designed welding work area or table. As you'll see, there's a wide variety of helmets available. Choose one that you are comfortable with and that meets your budget. You may have gotten a basic helmet with your welding machine, and you can use that until you save up for one of the more advanced models. Other personal safety equipment would include steel-toed industrial safety shoes (available at most shoe stores), good leather gloves, fire-resistant long-sleeve shirts or, if you are arc-welding, a leather welding jacket and/or apron.

A clean, safe welding table setup is of paramount importance. While you are learning, there will be a lot more sparks and slag generated than when you are advanced, and your welding table should be heavy steel, positioned well away from all stored

LEATHERS
EYEWEAR
LENSES
GOGGLES
HELMETS
TILLMAN

8.1 You'll find most of your welding safety gear and clothing at your local welding supply store, such as helmets, leather protective wear, gloves and safety shields or goggles.

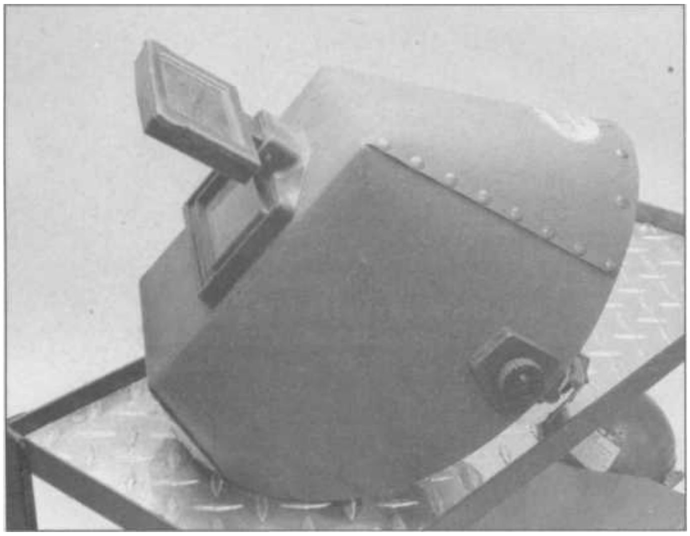

8.2 This Hunstman welding helmet offers full face and chin protection, and features a flip up welding lens. When you are setting up to weld, the helmet can be down on your head and you look through the clear lens, then flip down the dark lens just before welding. Some weldors find these better than fixed lenses where you have to shake your head to flip the whole helmet down.

L-TEC
LIFTORAMA

8.3 This L-TEC helmet from ESAB, called the Liftorama, features a lift-up safety lens, and underneath a shatter-proof clear lens that can be used when grinding a weld.

L-TEC
WELDORAMA

8.4 Another L-TEC helmet, the Weldorama offers a clear sight plate above the welding lens, so the weldor can keep the helmet down, and then just shift his eyes to the dark lens when he is ready to weld.

flammables and with a clear area underneath for any sparks or molten slag to be safely contained without having them scooting across the shop or lodging on top of your shoes. For metal cutting, you need a good "grate" setup as pictured in Chapter 3. You will also need some firebricks, which are helpful to practice welds on because the work won't stick to them. Do not just cover your old wooden work bench with a piece of sheet metal and call it a welding table. Invest in a piece of plate from a scrap yard. When you put a lot of heat into welding projects on a tin-covered wooden surface, enough heat can go through the metal to start the wood smoldering, perhaps even starting a fire after you have left the shop. Also, you must have proper fire extinguishers on hand, preferably an ABC type or halon, that can put out all kinds of fires.

Much of the work of sizing and joining metals for home/shop projects will involve tools you already have, such as a drill press, grinders and sanders, clamps, measuring equipment, squares and levels. A common hacksaw can be used to sever lengths of material but can be tedious on heavier-wall stock. Sheet metal can be cut with aircraft tin-snips, manual or electric nibblers, or a plasma cutter.

Many of the tools shown here are special-purpose tools that you may not have already, and which perform special functions that relate to improving welding work or holding materials to be welded. All of them are useful when the time comes. We all wish we could just put one of each in our home garage, but you may have to collect them over the years as your projects get more involved.

We also show here some tools and equipment that were made by fabricators

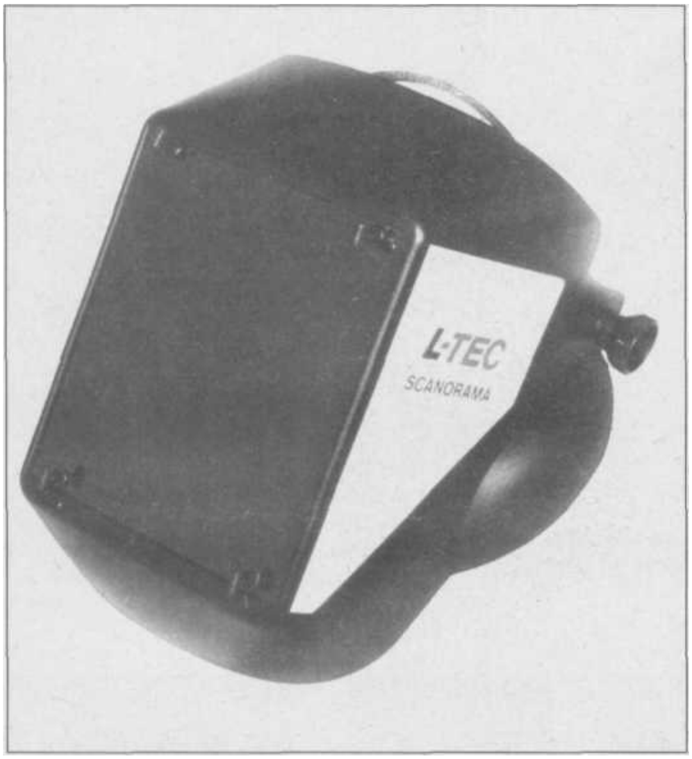

8.5 The Scanorama helmet was designed for MIG and TIG-welding applications, with a large shade 5 filter disc surrounding the welding lens. You can see most details through this area, but your skin is still protected from radiation. You look through the darker portion when welding.

AUTORAMA 2000

8.6 This Darth Vaderish design from ESAB is the Autorama 2000, which one of a modern breed of welding helmet with an automatically-darkening lens. The lens is powered by solar cells, and turns dark as soon as the welding arc begins. These are considered by many weldors to be much easier to use than flipping a helmet up and down to set up each weld.

DQF
SPEED
SHADE

8.7 Daytona MIG offers a variety of modern safety helmets, and a special, clip-on lens (foreground) that attaches to your MIG gun. It allows you to set up a weld, then look through the lens when making spot-welds without the bother of a full helmet.

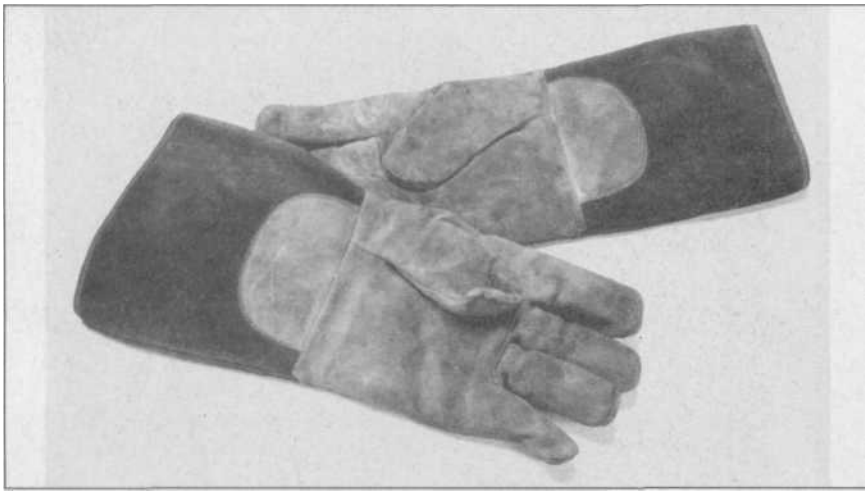

8.8 A good pair of leather welding gloves will last you for many years, but if you are using oxy-acetylene equipment, keep an extra-clean pair for use with this equipment. A spurt of oxygen on a dirty glove could cause combustion, so don't use these good gloves for handling hot or dirty parts.

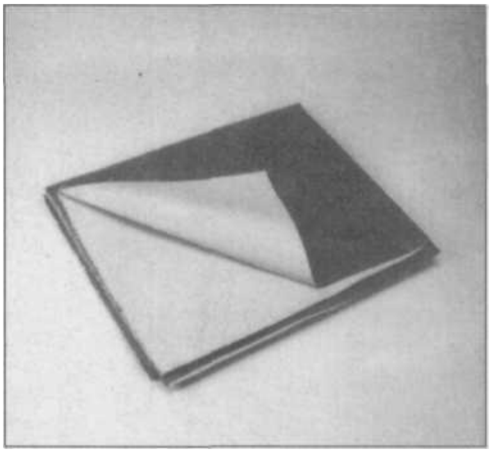

8.9 From HTP is this Super Spark Guard welding blanket. It can take up 2300° F, and can be used when working on a car to protect the window glass, dash or interior. It measures 37 inches by 50 inches.

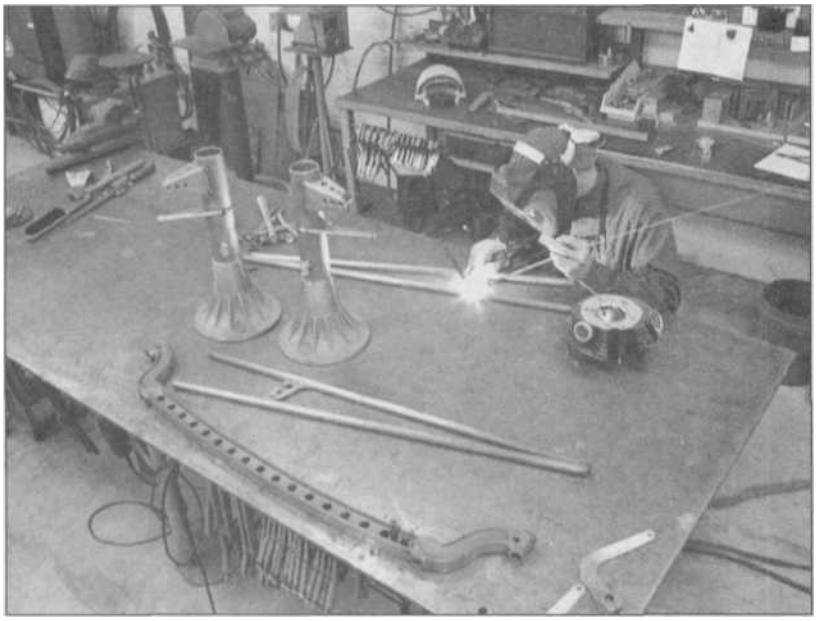

8.10 A clean, safe welding table area promotes good work. Your home shop won't need a steel-plate table this big, but note how uncluttered and organized things are here in a professional shop. A rack under the table holds all the C-clamps.

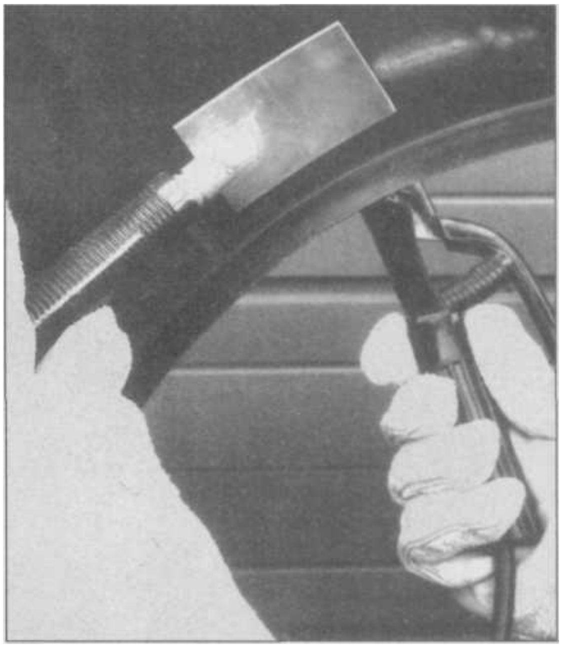

8.11 A very useful tool when welding on thin materials such as newer-car body metal is a copper "spoon." The heat-absorbing spoon can be used to back up butt welds or when filling holes in sheet metal. The welds will not stick to the copper.

and hot rodders for their own use. These are typical of projects you can do once you have practiced your welding skills, and some represent considerable involvement, though on the part of someone who has already "been there."

Sources for the commercial products in this and other chapters can be found in the Sourcelist at the end of this manual.

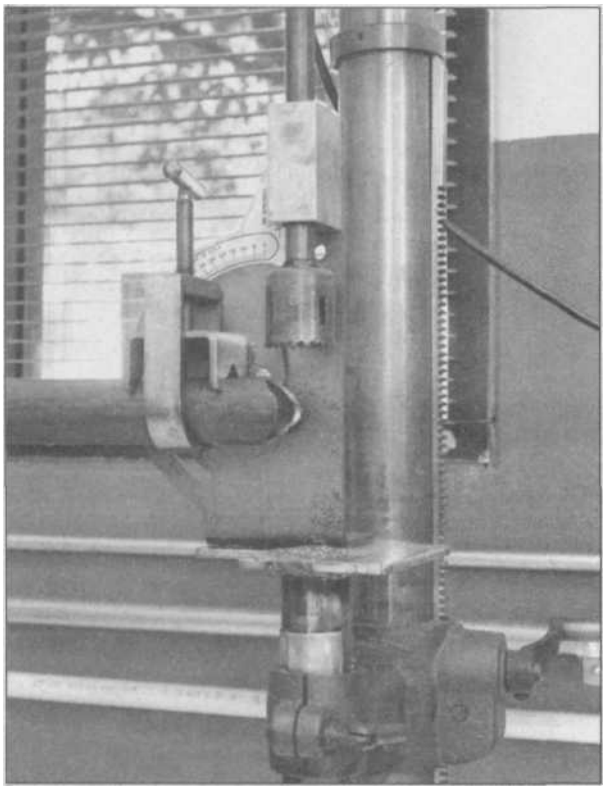

8.12 If you get into a race-car project, you might invest in one of these tubing notchers. Available from Eastwood, Williams Low-Buck Tools, Speedway Motors and other sources. It fits in your drill press and allows precise fish-mouthing of tubing via a captive hole saw.

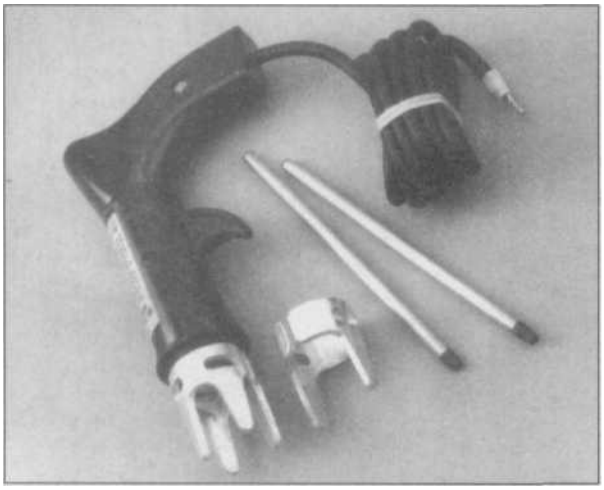

8.15 This unique spot-welding gun for sheet-metal work has a power lead that clamps into the electrode holder of your arc welder. Use the trigger to make the electrode touch the work, pull up for 5 seconds, then pull all the way to break the arc. Get it from Eastwood.

8.16 One of the handiest shop tools we have ever used is the Milwaukee "Porta-Band", which as the name implies is a portable bandsaw. It cuts fast and straight and is very helpful in cutting tubing at angles and other chores associated with welding projects.

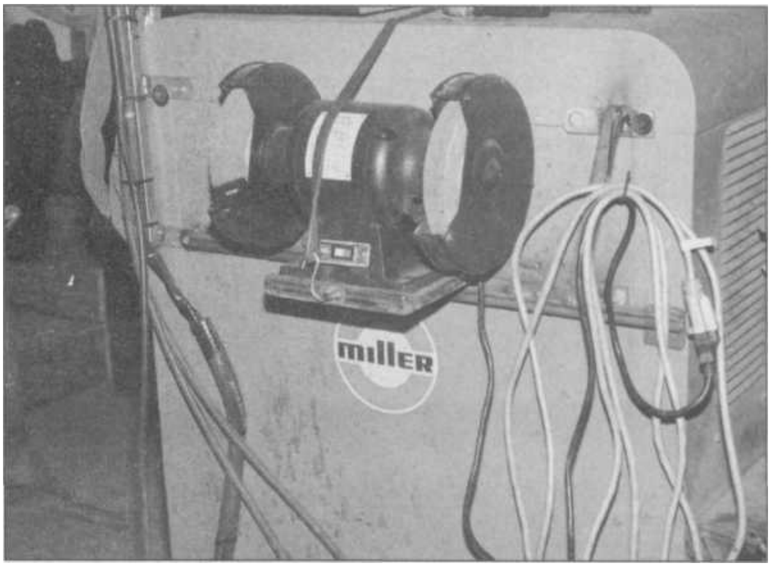

8.13 This welding shop has a dedicated grinder for sharpening tungstens mounted right on the side of the TIG machine. The rubber bungee cord kills the vibration noise when the grinder is slowing down.

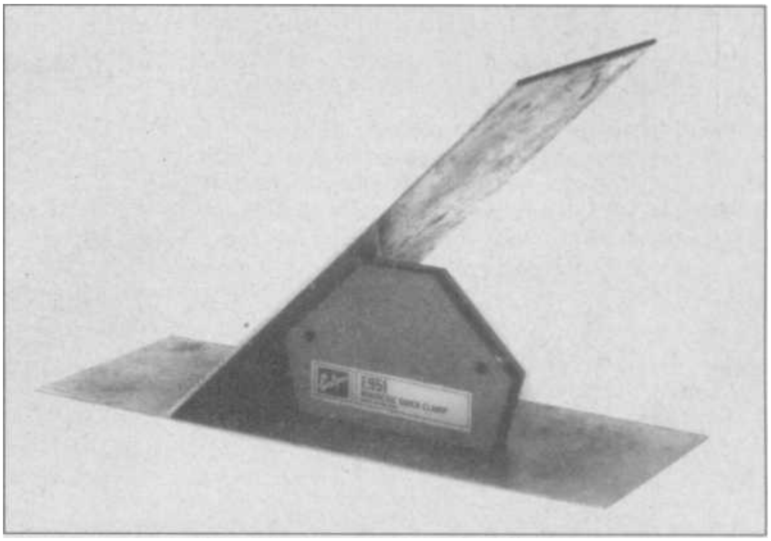

8.14 Eastwood offers this magnetic quick-clamp which functions to hold parts in alignment at various angles for tack-welding.

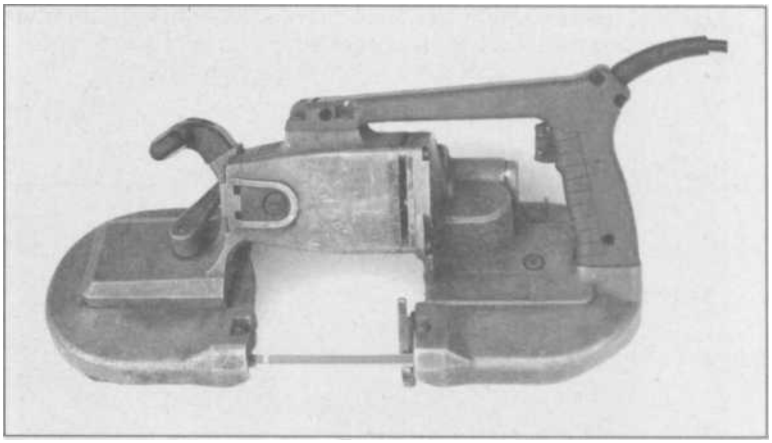

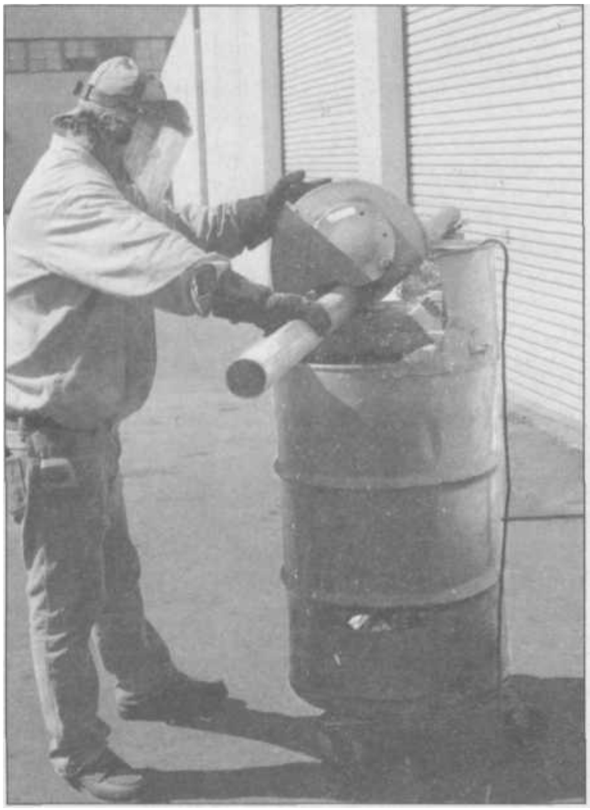

8.17 A "Radiac" or cutoff saw with an abrasive blade is very fast and secure when cutting large amounts of tubing. The work is held in a moveable vise so angle-cuts can be made. This one is cleverly mounted on a 55-gallon drum to capture all the sparks and debris.

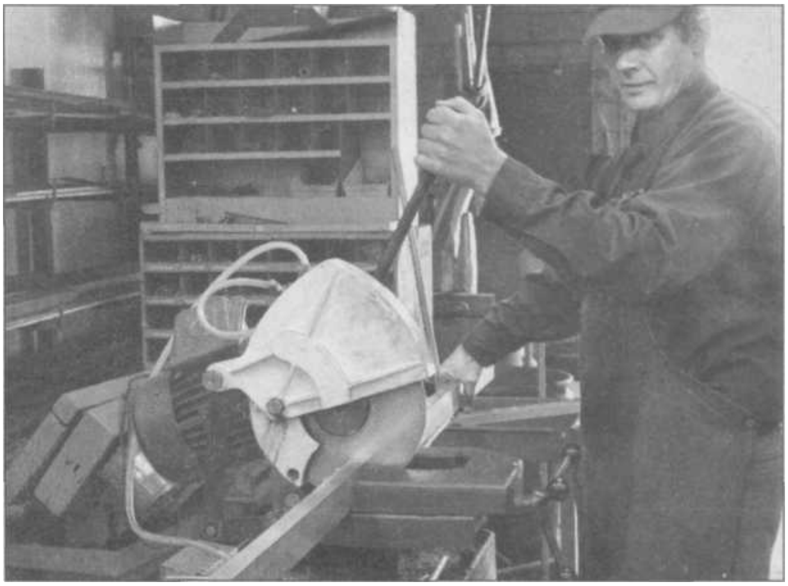

8.18 Other kinds of metal-cutting power saws include this type, which uses a bi-metallic toothed blade, here being used to cut clean through aluminum tubing.

Rotabroach
KIT
For Sheet Metal and Materials Up to 1/4" Thick
(A Hougen Drill for thinner materials)

8.19 In auto bodywork, a tool like this can be used in your electric drill to cut out factory spot welds to remove panels for replacement.

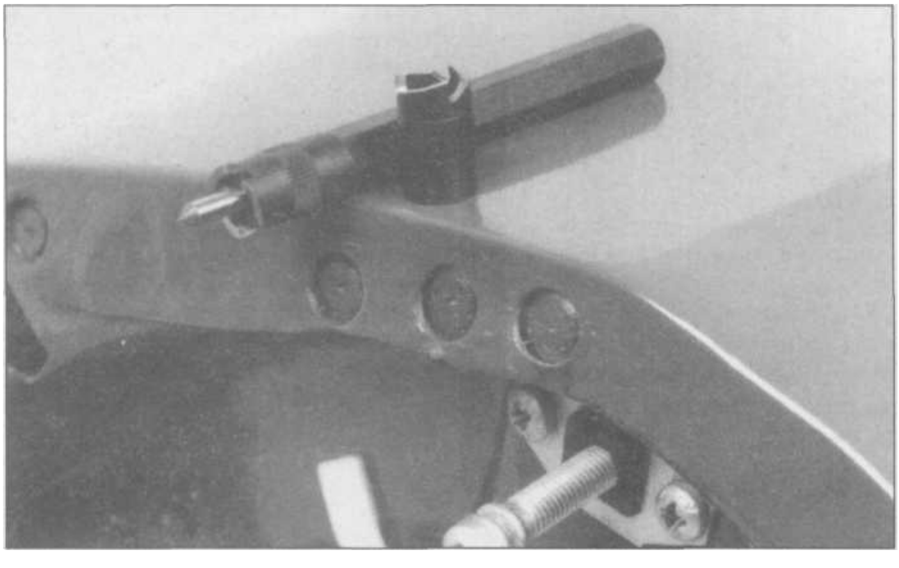

8.20 For Eastwood offers a similar tool for removing spot welds which features a hex-sided drive end, so it can't slip in your drill chuck.

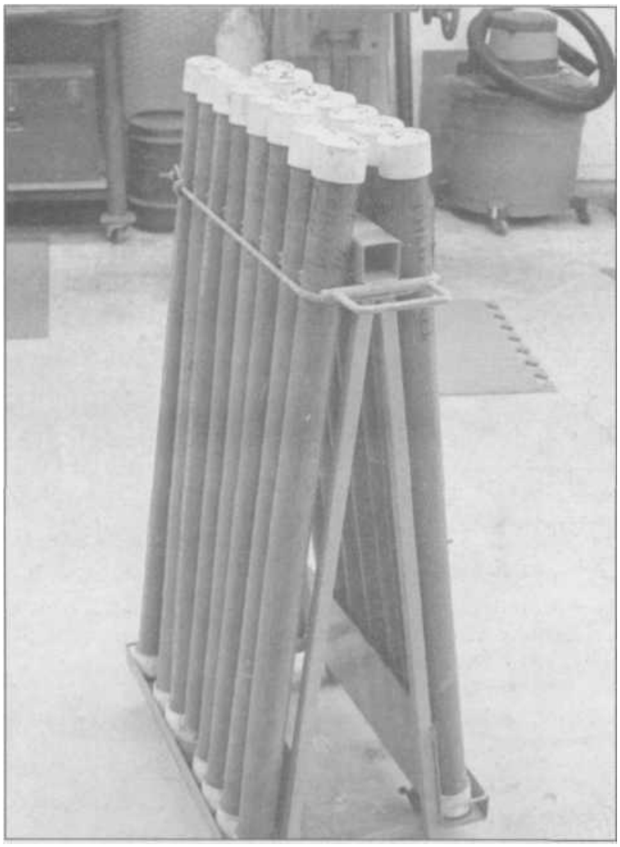

8.21 Weldor Bill Maguire made up this electrode rack for his TIG machine from a rolling platform holding 15 plastic plumbing pipes with caps, each marked for the size and type of electrode. A bungee cord holds them in the rack. Tubes keep the rods clean and dry.

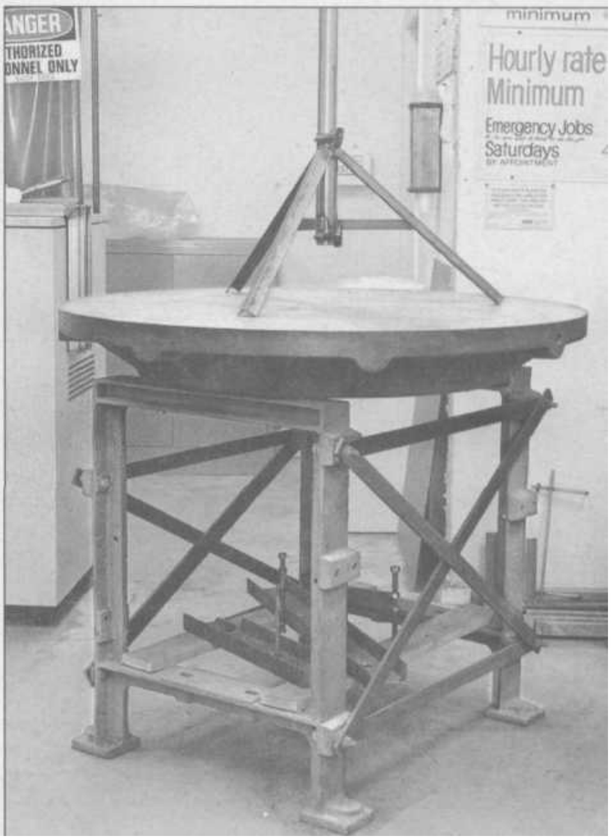

8.22 You never know when you'll find something that comes in handy in the shop. This huge and hefty round plate was found at a metal yard and bought at the per-pound price. It's dead flat and cannot warp, and has mounting holes in it for clamping work to it.

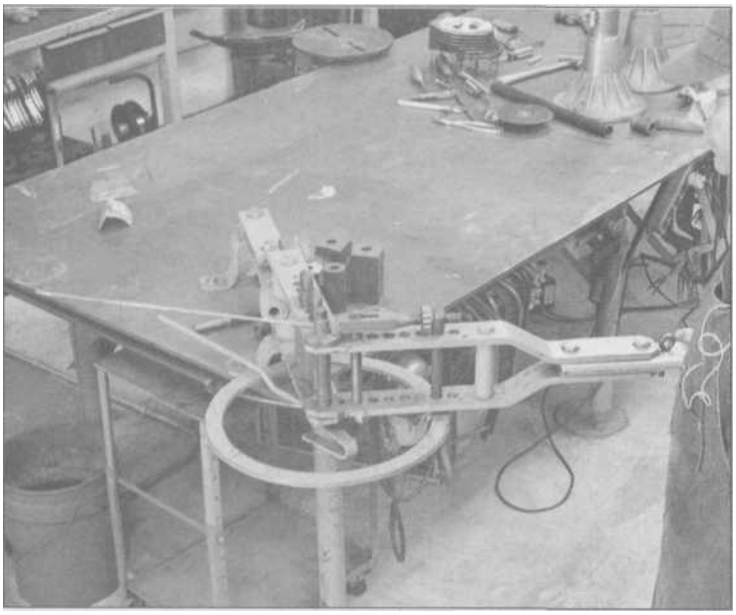

8.23 If you plan to do any ornamental iron work or make a lot of brackets, an ironworker like this is very handy, It can bend solid round, square or hex stock, as well as bend round tubing and flat stock into repeatable shapes.

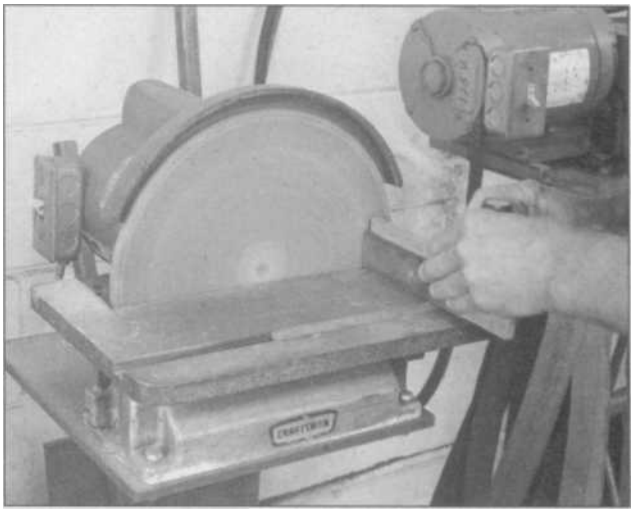

8.24 A disc sander with at least 1/2 horsepower is very handy around the shop for squaring edges of flamecut parts, and taking sharp edges and corners off work.

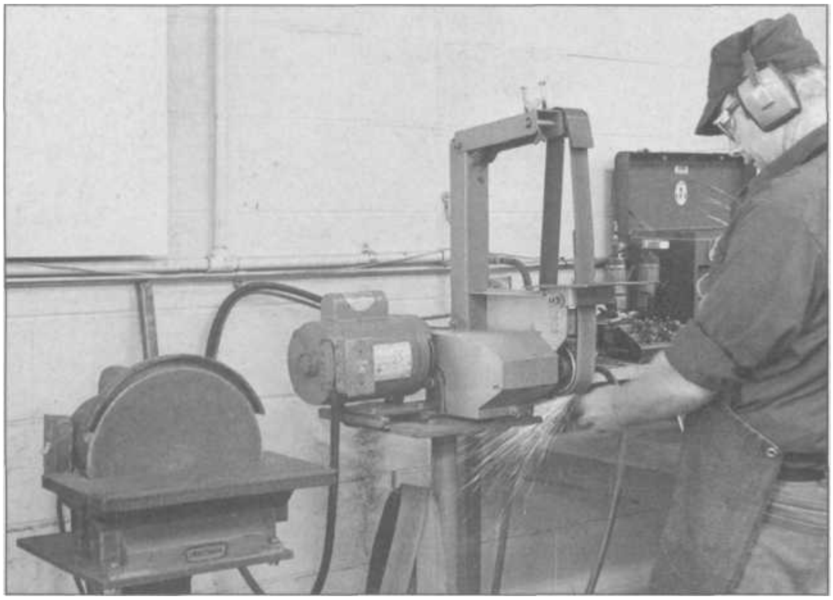

8.25 Ideally, you would have both a disc sander and belt sander for a variety of metal finishing needs. There are combination units, but beware of weak-powered sanders. When you put the rough edge of a 3/8-inch plate up there to sand, they stall easily.

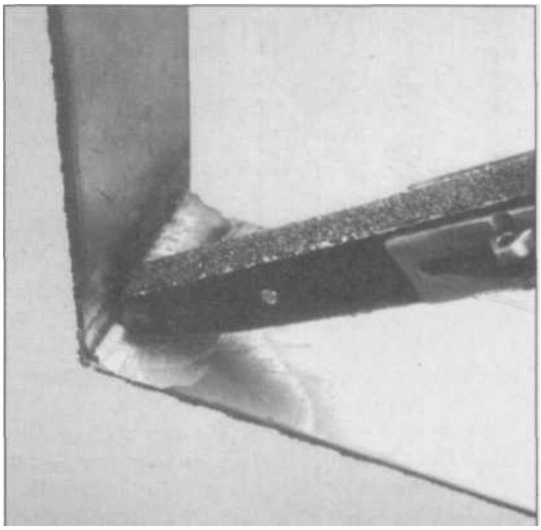

8.26 HTP is marketing this clever belt-sander that can get deep into corners and places where regular sanders and grinders can't reach. The belt guide rotates 160 degrees, and the sander uses 3/8-inch belts.

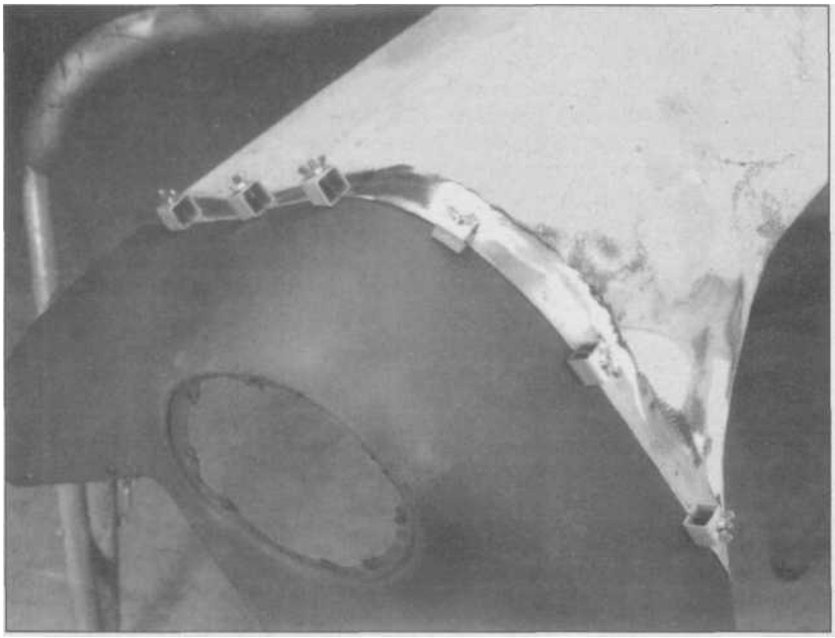

8.27 Something that is always a problem in body work when replacing panels is getting parts aligned properly before making a butt-weld. Eastwoods Intergrip clamps space the two parts .040-inch apart for full penetration. After tacking the panels, you can disassemble the clamps.

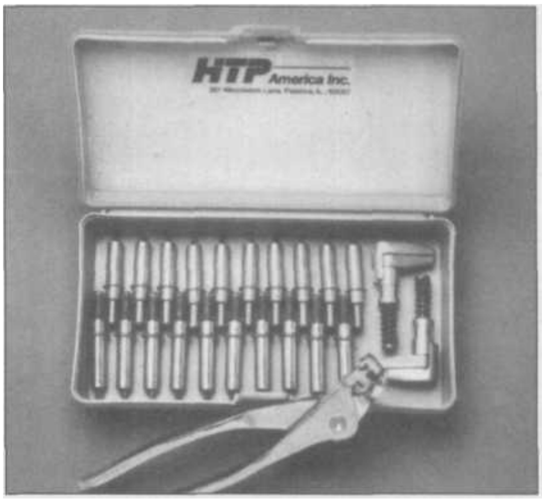

8.28 Nothing beats the original Cleco clamps and pliers for test-assembling sheet-metal parts. Originally developed for aircraft use, these operate like reusable, temporary rivets. In blind holes you insert the fastener with the pliers and the two parts are held together, but come apart easily with the pliers again. HTP offers this set.

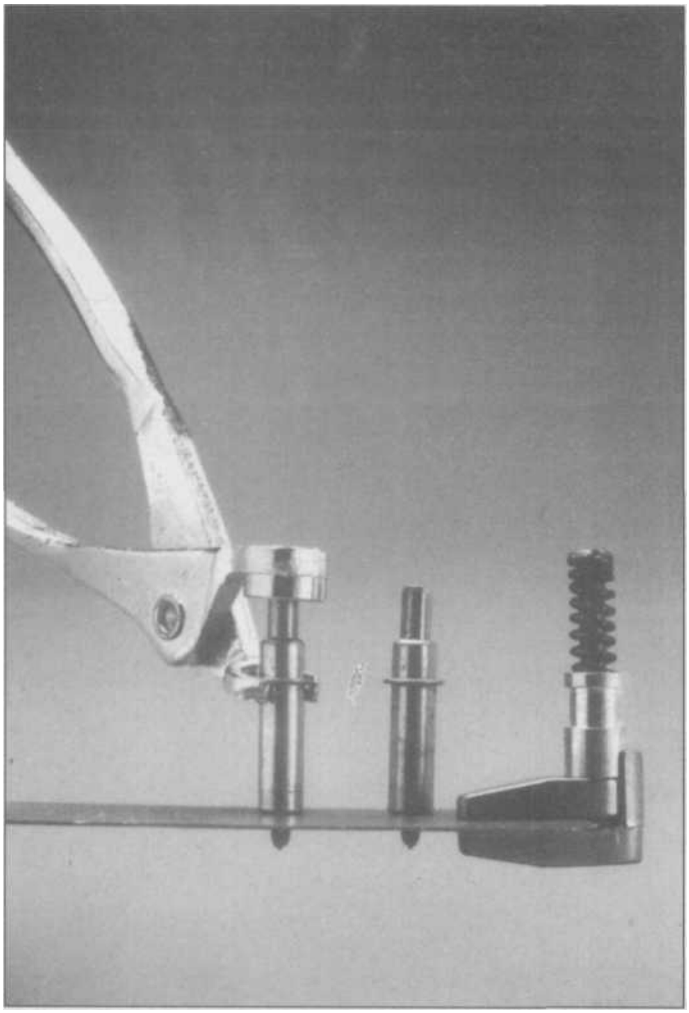

8.29 You can see here how the pliers extend the tip of the Cleco fastener through a set of holes. When the tool is released, the Cleco grips the parts. Also available are Cleco mini-clamps (right) that exert a temporary hold while you tack-weld small parts together.

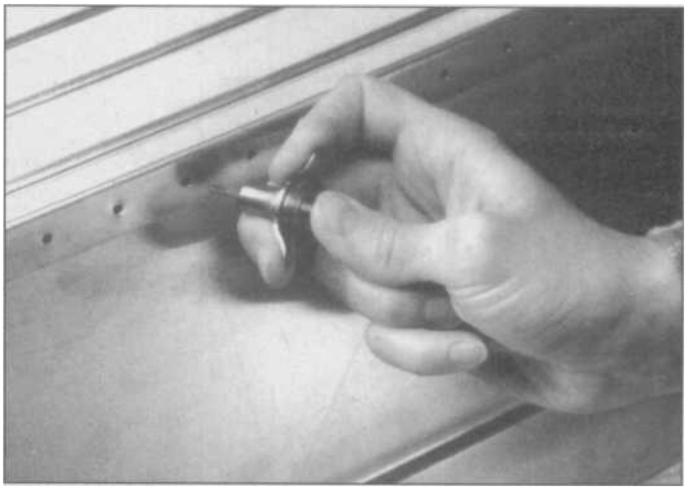

8.30 For small jobs, Eastwood offers these little Cleco-type panel holders that you operate with your fingers. No tool is needed. They have a range of up to 1/4-inch for thickness and fit a 1/8-inch hole.

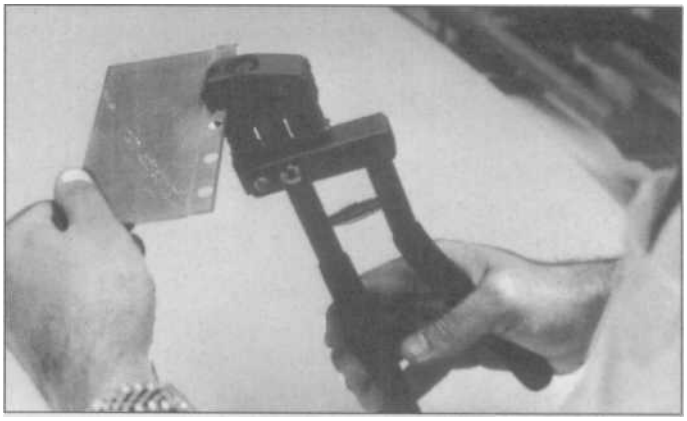

8.31 When you want to spot-weld sheet-metal panels that didn't have them originally, like new parts or repair panels you have made , a manual flange punch like this is a real help. This one is from Daytona MIG.

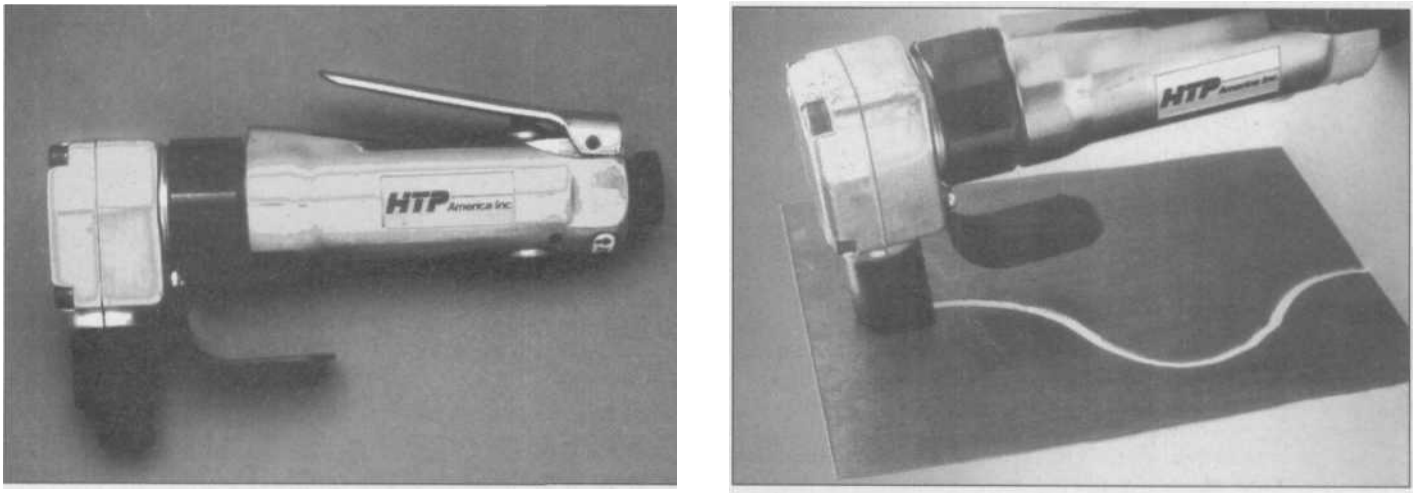

8.32 An air-powered "panibbler" is a fast, clean way to cut sheet metal up to 3/64-inch in steel or 5/64 in aluminum. It will follow a pattern or cutline very closely, and you control the speed with the throttle. This one is from HTP.

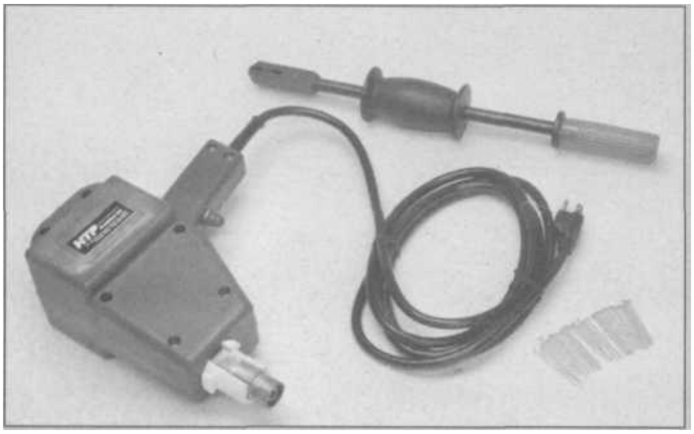

8.33 The HTP stud setter is a great boon to bodymen, who used to have to pull out creases and dents by drilling holes and using slidehammers that distorted the metal even further. This 110V tool saves considerable time and labor.

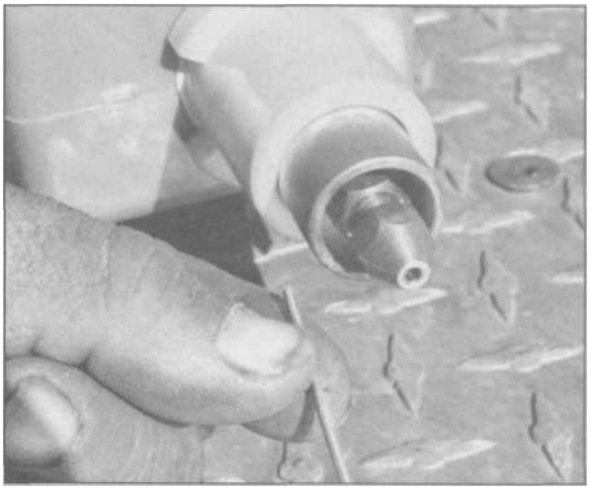

8.34 Small-headed copper-plated "studs" are inserted into the magnetic head of the stud setter.

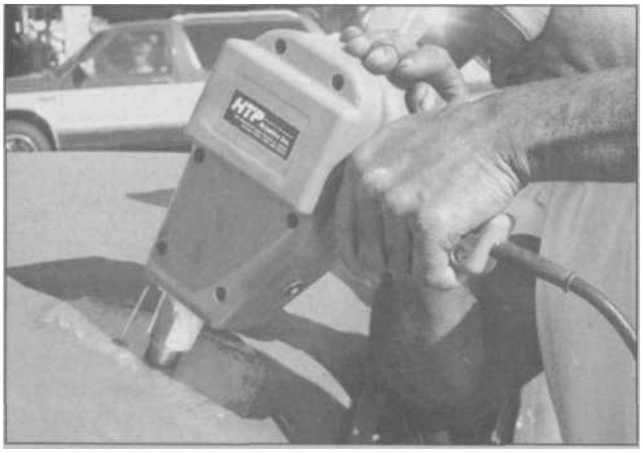

8.35 On a dented area that has been ground to clean metal, the tool is pushed in contact with the metal and the trigger is pulled for a half-second.

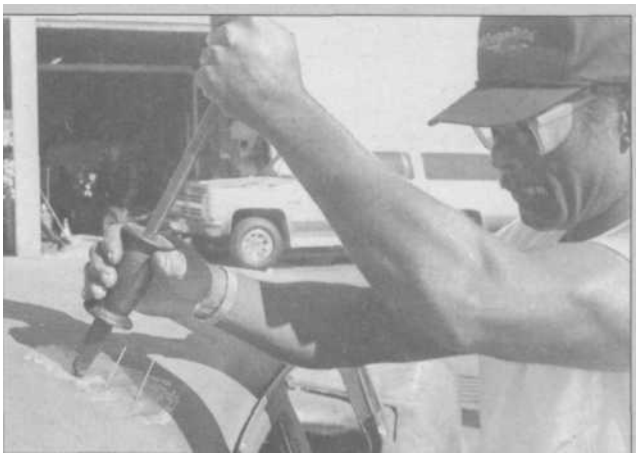

8.36 The suds are spot-welded in place, and a special slide hammer is used to pull out the dent or crease. Studs hold 500 pounds of pull.

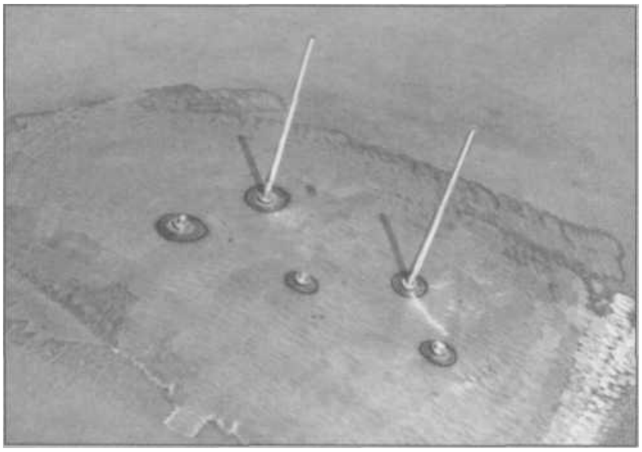

8.37 On this dent, the first three studs pulled out the bulk of the damage, and two more are in place to pull the last bit out. Note how the first three studs have been clipped off with cutting pliers.

8.38 After the pulling, the stud heads are easily ground off without distorting the metal, and there are no holes left to fill in, as with conventional methods.

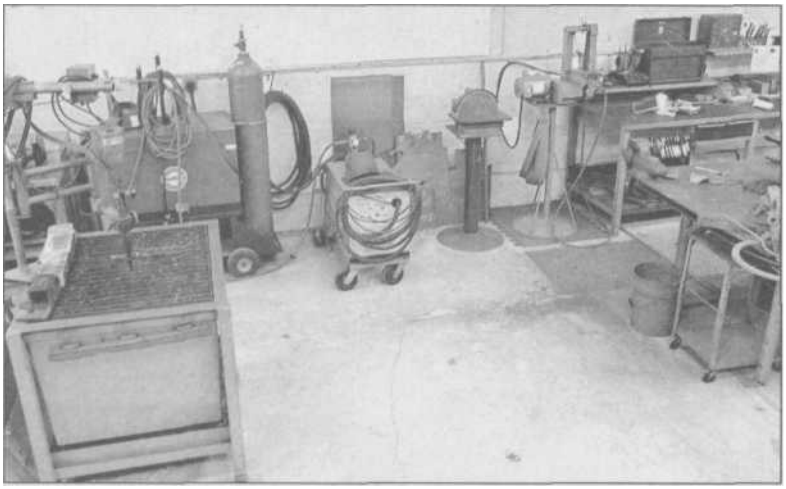

8.39 Visit most street rod shops, fabrication shops and race-car facilities, and you'll find that a lot of the shop equipment they have is homemade, and these are the kinds of things you'll be able to do once you have welding and cutting equipment.

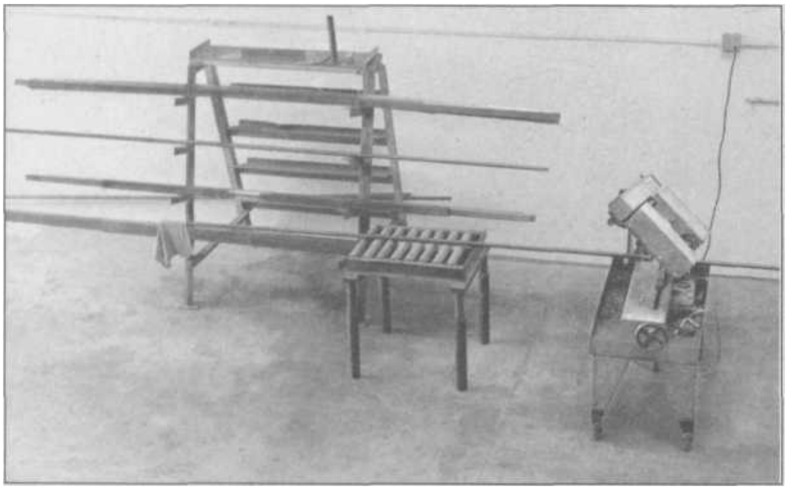

8.40 A simple steel materials rack like this is great for keeping tubing, bar stock and flat stock out of your way, and is easily constructed. In front of this one is an electric, automatic bandsaw great for making unattended cuts. This one was purchased inexpensively at a tool auction. Watch your newspapers for industrial auctions and shop foreclosures.

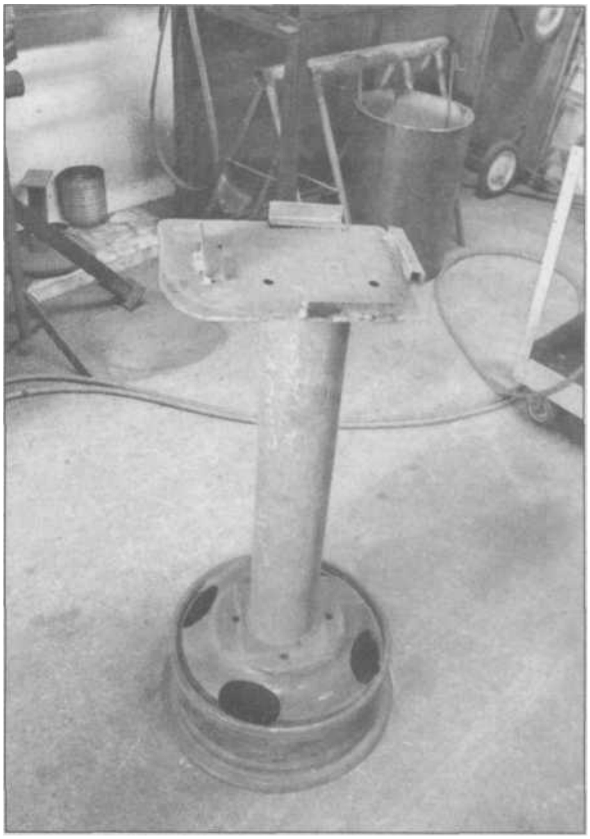

8.41 An old truck rim, a length of round pipe and a piece of scrap plate are combined with welding to make a perfect small welding table or mount for a grinder or sander.

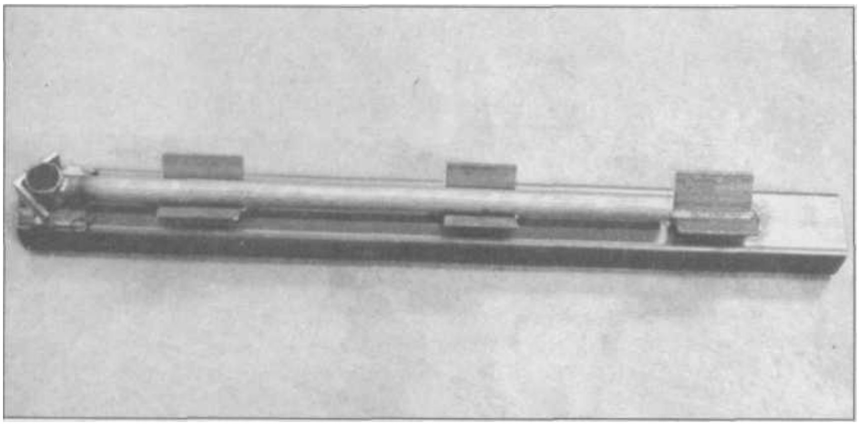

8.42 Leftover pieces of tubing and other stock should be saved in a barrel. You never know when you need a short piece. This fixture for making up a street rod four-link rod is made from short pieces of angle-iron and a length of rectangular tubing.

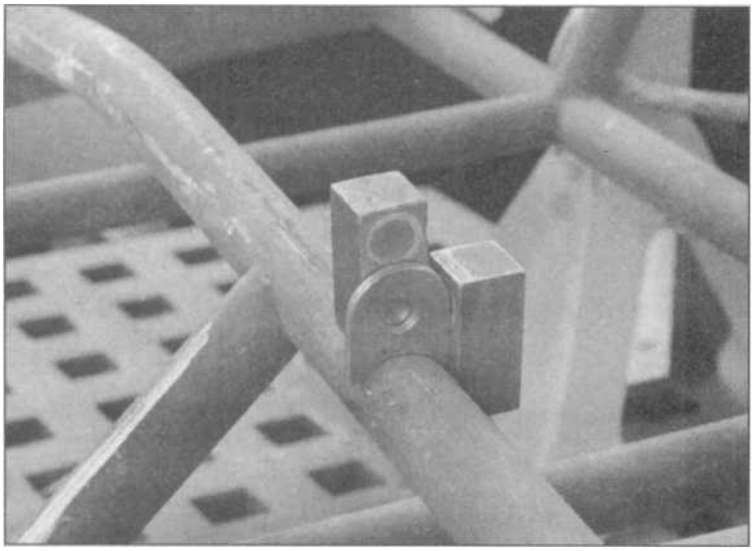

8.43 Ideal for holding small brackets onto tubing for tacking are these magnetic blocks, available at most welding supply stores.

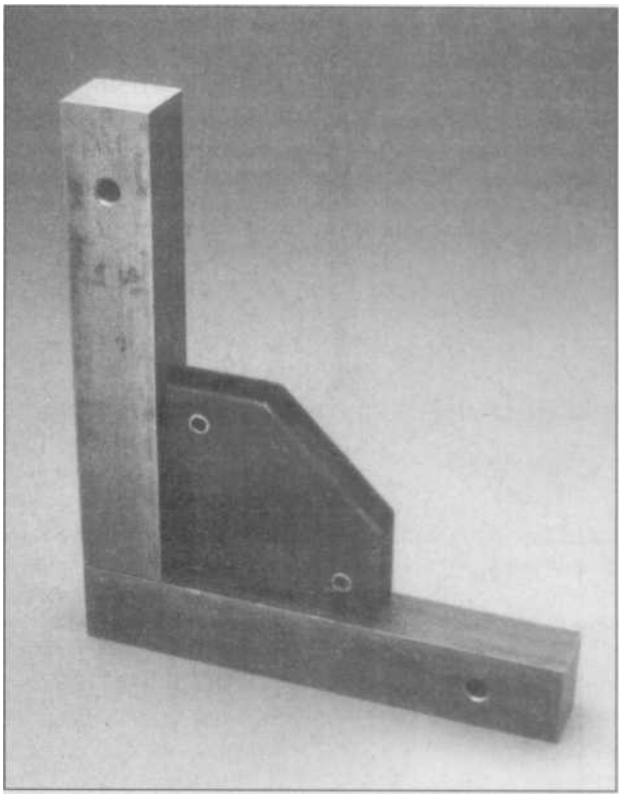

8.44 HTP and other companies offer these strong magnetic holders that can secure two pieces of tubing for angles of 45 degrees, 90 degrees, and 135 degrees. Tack the parts then move the magnets to the next joint.

8.45 Another type of angle fixture from HTP is this type, which clamps the two parts at a perfect 90 degree angle for tack-welding.

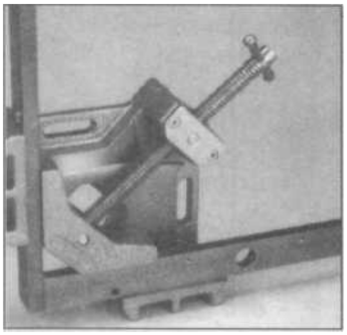

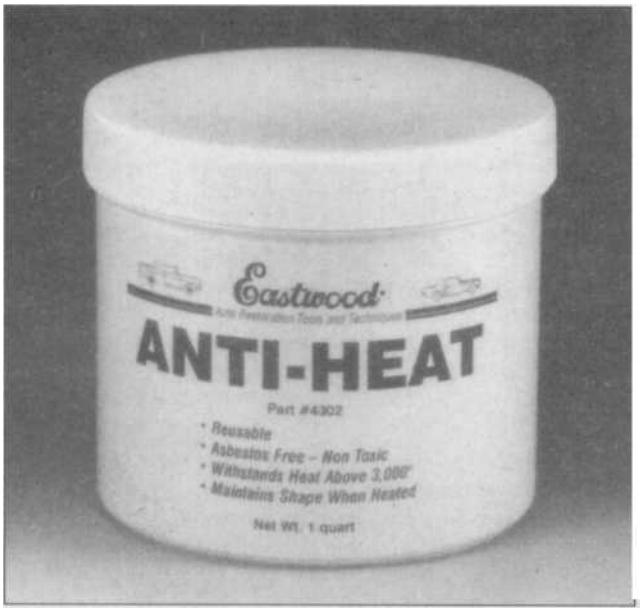
Eastwood
ANTI-HEAT
Part #4302
• Reusable
• Asbestos Free – Non Toxic
• Withstands Heat Above 3,000°
• Maintains Shape When Heated
Net Wt. 1 quart

8.46 The Eastwood anti-heat is a putty-like paste that can be formed around heat-sensitive parts when you are welding near them. The paste absorbs heat and can be used to protect auto glass when welding near it.

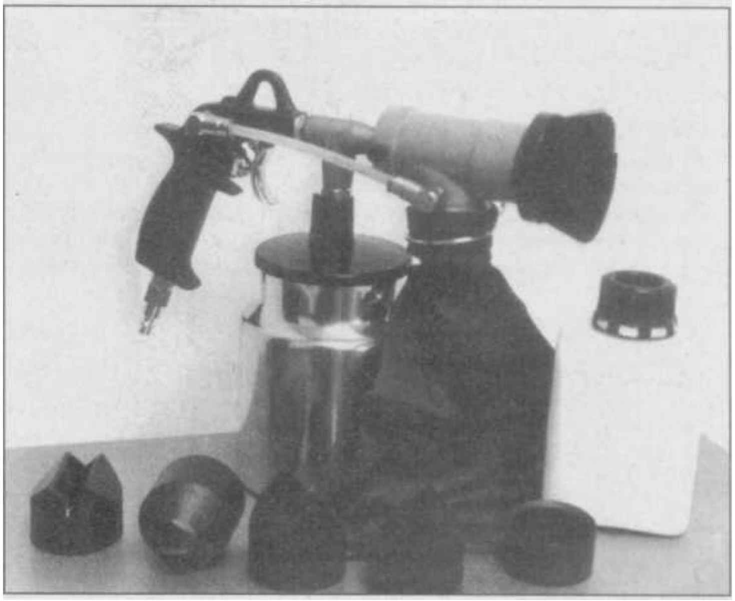

8.47 Daytona MIG sells this handy spot-sandblaster. It is useful for cleaning paint and rust away from a spot your have to weld, like a recess where a regular sander won't reach. Three-quarters of the blasting media is returned to the gun, so it won't make your shop a "sandpit."

9 Building a utility trailer

One of the most common home/shop projects for welding equipment is a trailer of some kind, Which can be used for carrying a race car or off-road vehicle or On a smaller scale, can function as a utility trailer. Each of these is easily within the realm of the hobbyist, though a trailer of any size should not be your *first* welding project. Hone your cutting, fitting and welding skills on other, less critical projects first, like a metals storage rack, a wrought-iron front gate, a fireplace grate, barbecue grille or a rolling cart for your welding equipment. Frankly, the latter is usually the first welding project after the practice sessions. Many are the accomplished weldors we have met who say they welded up their oxy-acetylene equipment cart some 20 years ago and are proud to be still using it.

Since the trailer project is something to be used on public roads and at highway speeds, you want your welding quality to be good before you get started because the welds here have to be strong, not just good-looking. The basic construction details of any trailer are fairly similar, whatever the size. You assemble a rectangle of steel tubing or angle-iron, attach a sus-

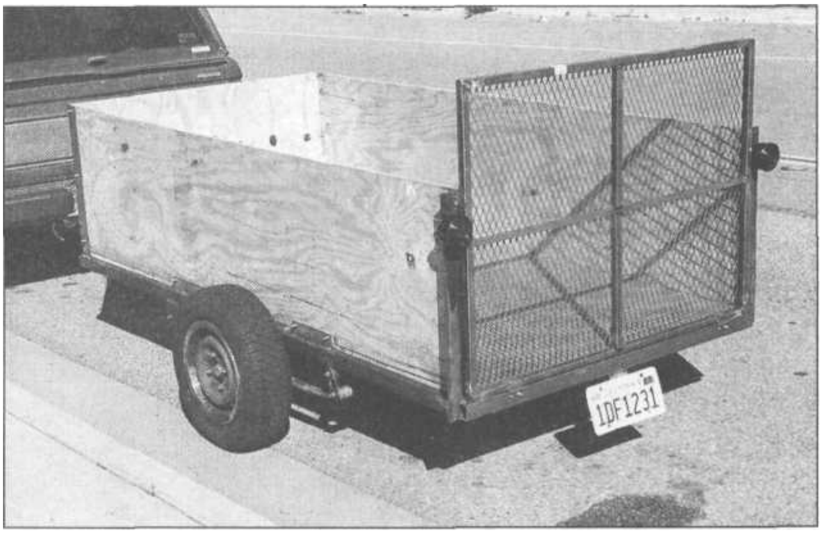

9.1 A 4x8-foot utility trailer like this is a good project once you have perfected your welding and fitting techniques. Once you have a trailer like this, you'll wonder how you ever got along without one. It can be used for hauling camping gear, lumber, yard trimmings, a go-kart or furniture.

9.2 Find a good trailer supply store in your area, where you can browse for suspension parts, wiring, lights and hitch components.

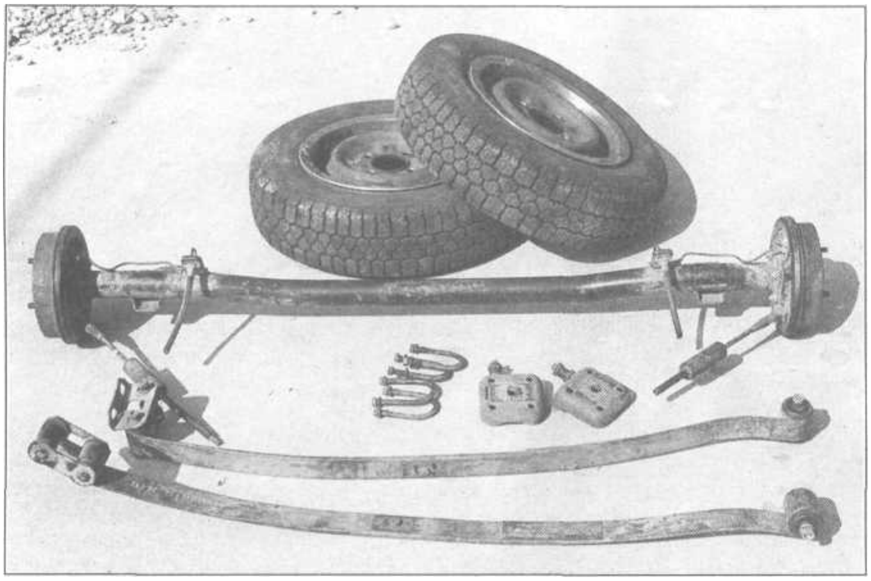

9.3 We purchased this Chevrolet Sprint solid rear axle and springs at a do-it-yourself wrecking yard very inexpensively. New trailer axles and springs can be expensive, but you can often find used ones at swap meets, or you can buy an old, small boat trailer that you can rework into a utility trailer.

pension and axle system, add a floor and sides, taillights, safety chain and a hitch in the front. Most smaller trailers, from 4x8-foot size and down, can be constructed of 2x2 tubing, or 2x2 angle-iron, depending on the hauling requirements. A trailer such as we are illustrating here is capable of carrying motorcycles, a go-kart, personal watercraft, garden tractor, camping equipment, building materials, firewood or landscaping dirt. With a few simple poles and some canvas, you could actually use it to camp in, like a homemade pop-up camper. We built this one of 2x3 tubing and 2x2 tubing of 1/8-inch wall thickness, which is somewhat overbuilt, but these were the materials that were accessible, and the cost of building such a trailer is entirely dependent on your materials. If you were building a car-hauling trailer, you would use two axles, preferably with brakes, and use diamond-plate steel for the floor, with some very sturdy ramps.

If you can get much of the steel from the scrap or remnant pile at your local metal source, scrounge used plywood or surplus diamond-plate steel and get your axle and suspension from a donor car at a wrecking yard, such a trailer can be built for 1/3 to 1/2 the cost of a newly purchased trailer of the same size. You'll need some kind of metal cutting equipment beyond the basic hacksaw or your arm is going to get tired. There's a lot of metal to cut to build even a simple trailer. You could lay out your plan on the driveway or garage floor, mark all your materials and bring them to a shop for cutting if you have to, then bring them back home to weld. Welding such a project can be done with any of the equipment we have discussed in this book, but it goes together fairly quickly with either arc-welding or MIG-welding.

We used inexpensive-grade plywood for our trailer sides, which we made two feet high, and a better grade of 3/4-inch plywood for the floor bolted to the frame with carriage bolts so the floor would remain smooth enough for easy rolling of a furniture dolly into the trailer. Once you have built such a trailer, it's like having a new truck. Everyone you know will want to

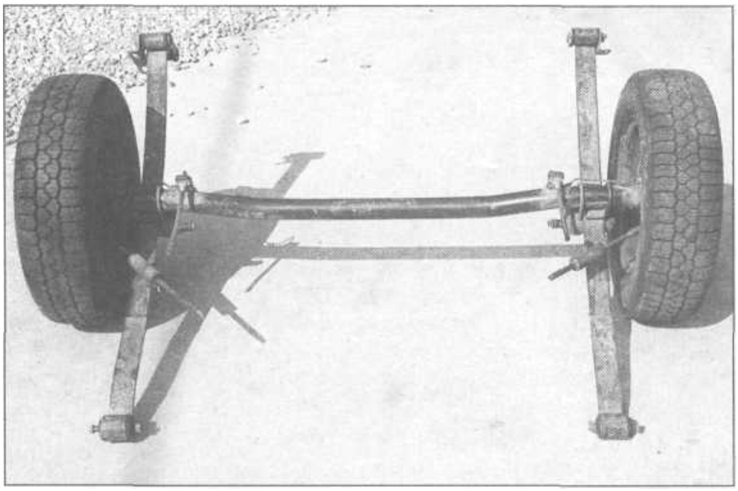

9.4 Bolt your springs, U-bolts, spring plates and shackles together, making sure the springs are square to the axle centerline and measuring the same distance apart front and rear.

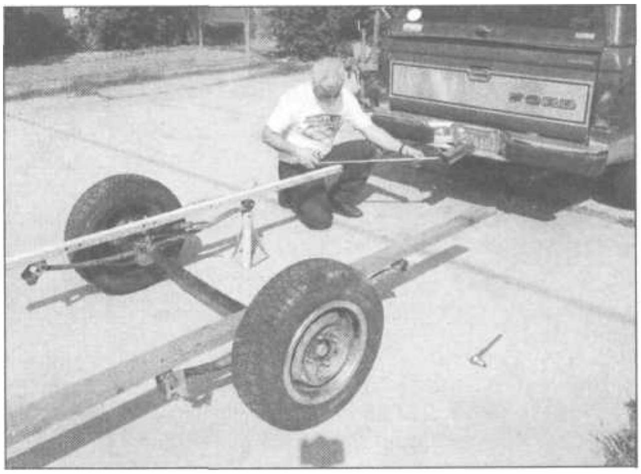

9.5 Attach a hitch to the ball on your tow vehicle, and mock up the assembly, using 2x4s to simulate the rails. This should give you an idea of what size and layout of trailer you want, and where to locate the axle and wheels. Keep more trailer frame in *front* of the axle than behind, to insure you have enough tongue weight.

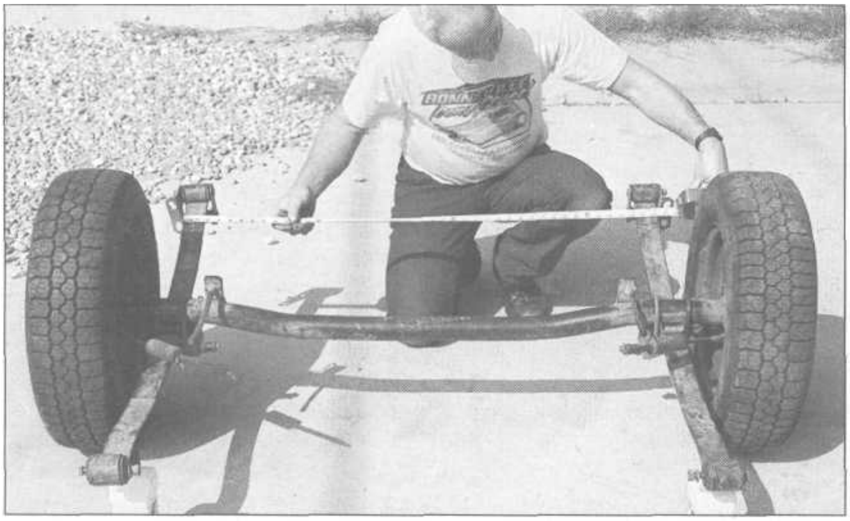

9.6 In our case, we determined that the Sprint axle was too narrow for our purposes, since we wanted a full 4x8-foot sheet of plywood as a floor. This meant widening our tubular axle 7 inches.

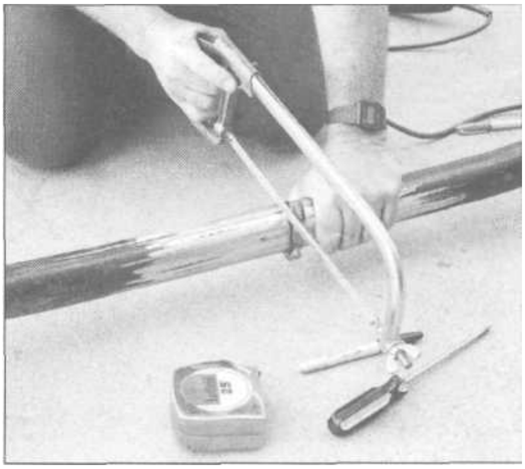

9.7 Grind the paint from the center of the axle, mark a straight centerline across the axle for lining-up purposes later on and cut the axle with a hacksaw. A hose clamp around the axle is used as a saw guide to keep the cut straight.

borrow it to move something. Our tailgate was made from a frame of one-inch-square steel tubing, .063-inch-wall, and we tack-welded to it a section of steel mesh such as is used for barbecue grilles. The hinged tailgate allows wind to flow through the tailgate on the road, reducing fuel consumption from the drag of a solid tailgate.

For our project, we used a solid rear axle and suspension from a compact FWD Chevrolet Sprint, which has single-leaf rear springs. We found later on that we had to add helper springs to accommodate heavier loads with the trailer. The 4x8-foot size allows for a fairly heavy load. We just bought two more Sprint rear springs at the local do-it-yourself wrecking yard, cut the eyes off each end with a friend's plasma cutter and used these as an extra leaf on each side. This increased the load capacity without making

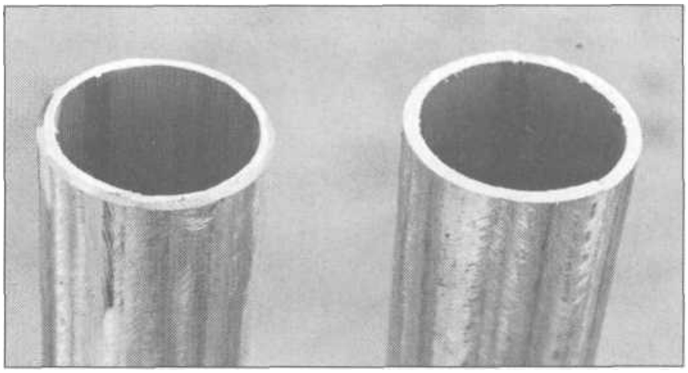

9.8 This axle is heavy-wall tubing, and, since we wanted good weld penetration for strength, the cut ends were beveled with a grinder. Tube at left has been beveled to about half its thickness.

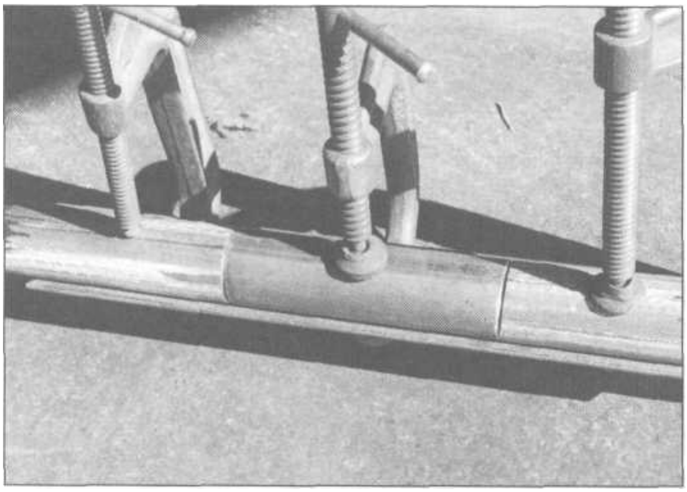

9.9 With the remains of your original centerline in alignment, clamp the axle ends together with a piece of new heavy-wall tubing the same size and keep them all in alignment with a length of wide angle-iron.

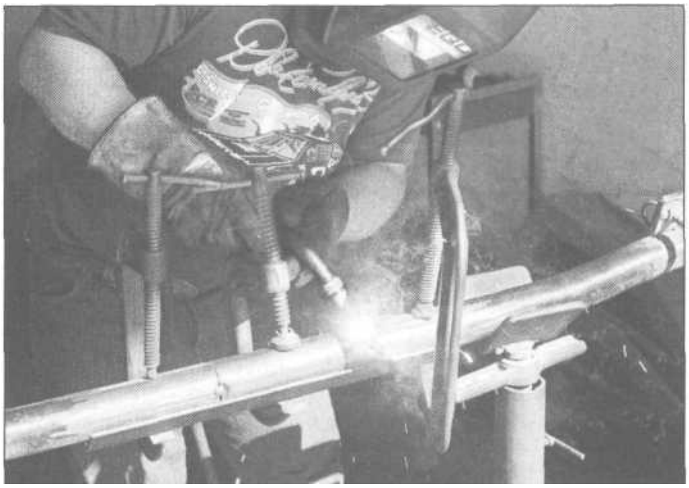

9.10 Here the assembly is tack-welded at several points around each joint with a MIG welder.

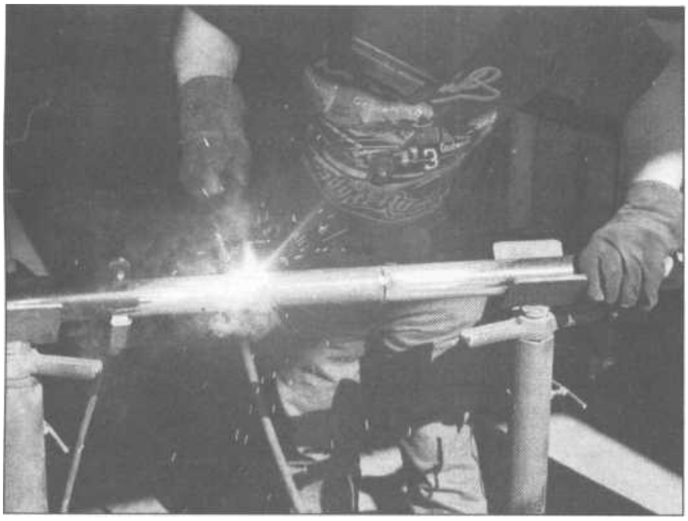

9.11 When the axle is securely tacked at several points, remove the angle-iron and make a strong weld bead all around the axle at each joint. Check for straightness after it has cooled off.

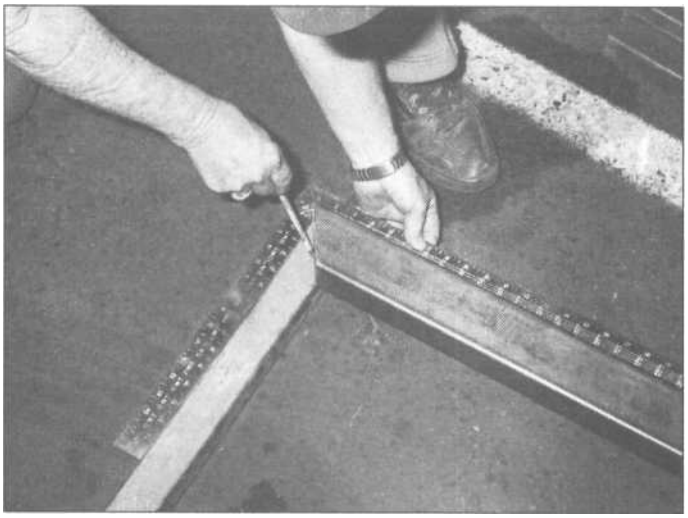

9.13 We used what materials we had, with 2x3-inch tubing for the two main rails and 2x2-inch tubing for the rest. When joining tubing of different sizes, cut one to a standard 45-degree angle (here the 2x3), then lay the next piece in, square it and mark along the second tube for the correct angle to cut it.

9.14 You can cut tubing and angle-stock with a hacksaw, but it is tedious. We borrowed a Milwaukee Port-A-Band hand-held bandsaw that made short, accurate work of fitting our tubing.

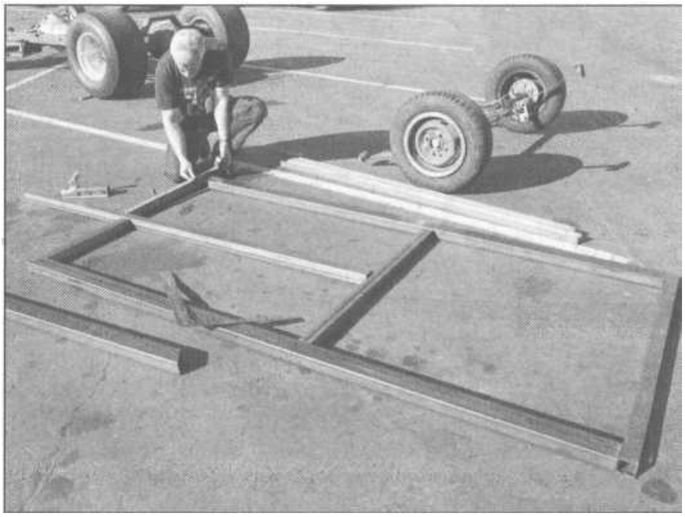

9.12 If your driveway or shop floor is flat enough, lay out your pieces of tubing on the ground and check all your measurements. Cut your joint angles carefully, and, with the pieces laid out, measure diagonally from corner to opposite corner to see if the frame is square. Check this again several times during the tacking and welding phase, too.

it too stiff when unloaded. Since you are using an automotive axle, it comes with brakes and all, and you *could* hook these brakes up to your tow vehicle with parts available from a trailer supply store. For our loads, like firewood, furniture and camping equipment, it didn't seem necessary. You could also buy a regular trailer axle with springs and brakes, but they are rather expensive if purchased new. There's a number of straight rear-axle assemblies available from front-wheel-drive cars in wrecking yards at reasonable prices.

There are few hard-and-fast rules in building a trailer, except that you should design it with the suspension and axle located in the rear 50% of the basic frame, not in the center, to maintain good tongue weight up front. Your tongue weight should be roughly 10% of your total trailer weight statically at ride height. The one thing that will certainly make a trailer handle badly is not having a long

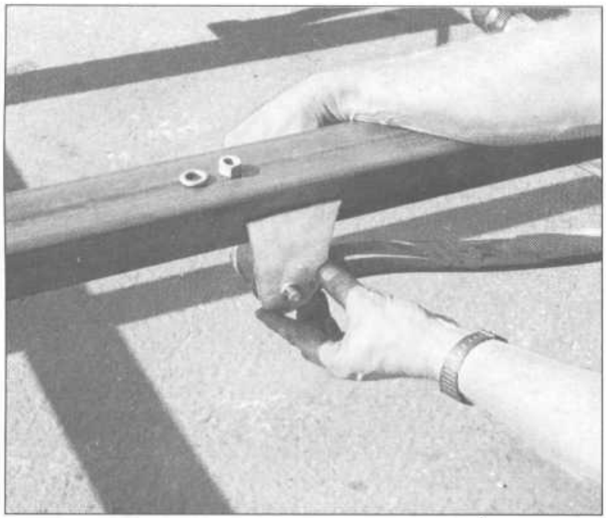

9.20 Put bushings in the front spring eyes and use cardboard to make up a template for four front-spring-eye mounting plates to be welded to the frame

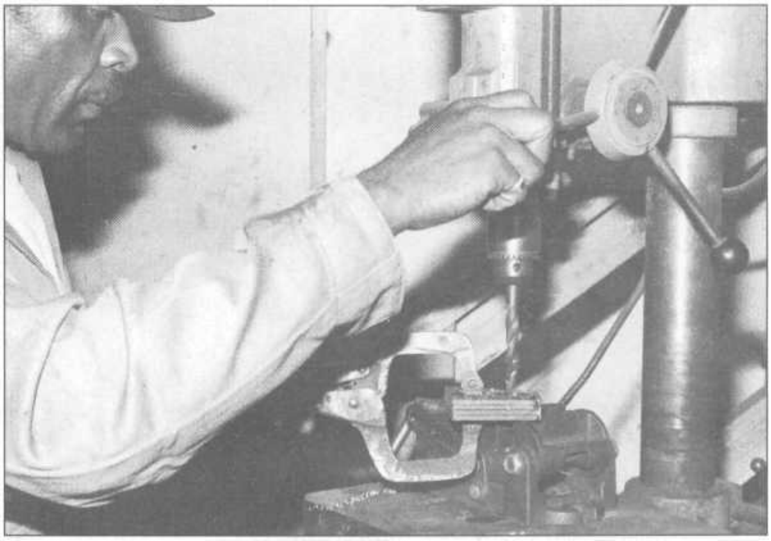

9.21 After flame-cutting, torching or plasma-cutting your plates, sand the edges smooth and stack them together with a vise-grip clamp. You can drill all four plates at the same time, saving time and ensuring that all have the spring-bolt holes in the same place.

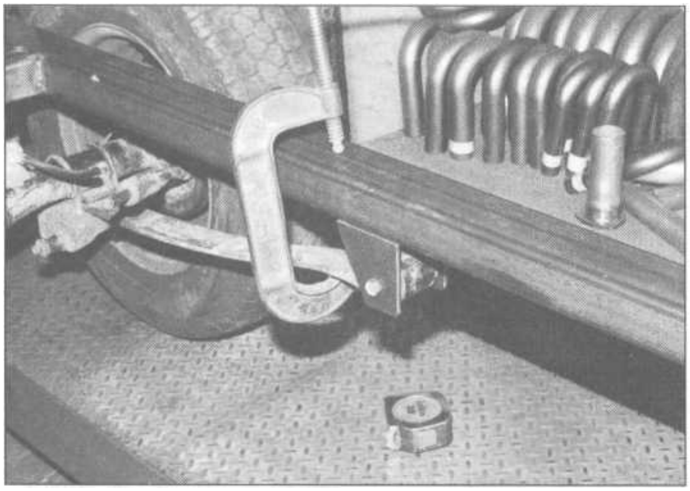

9.22 The front spring-mounting plates are bolted onto either side of the front spring eye, squared to the rails and clamped lightly in position for tacking.

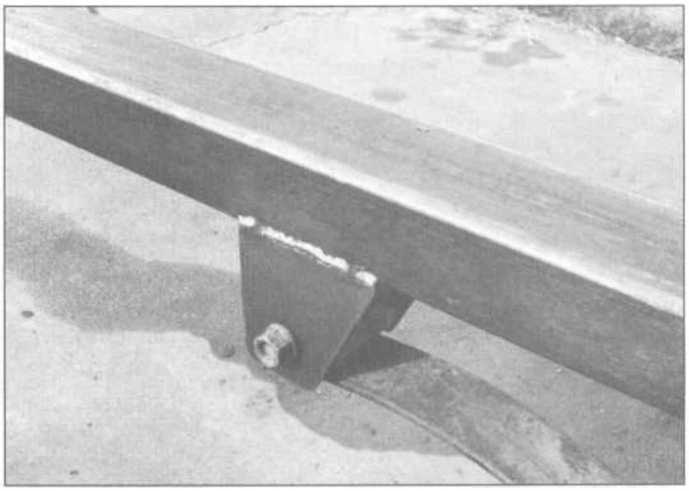

9.23 After some solid tack-welding, pull the springs out so the bushings don't melt, and fully-weld the four front spring mounts. Here the springs are bolted back in.

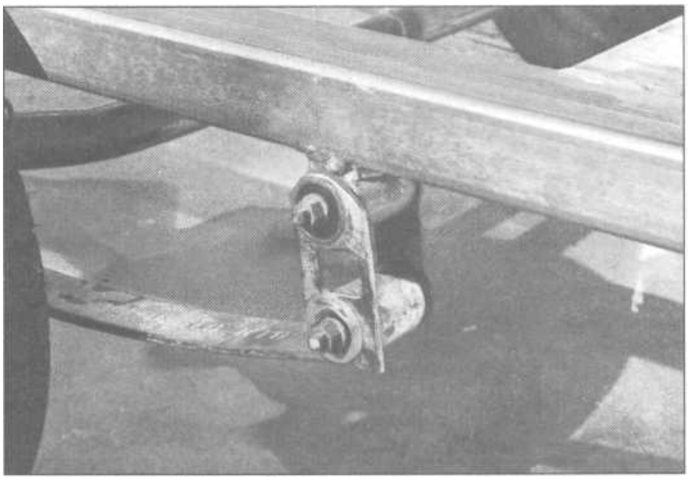

9.24 At the rear, short lengths of heavy-wall tubing, with an inside-diameter that fits the spring bushings, are welded to the underside of the rails (tack first, then remove the bushings to finish welding). Tubes should be positioned to have the spring shackles straight up and down. When the weight of the whole trailer is put on, they will arc to the rear enough for easy spring action.

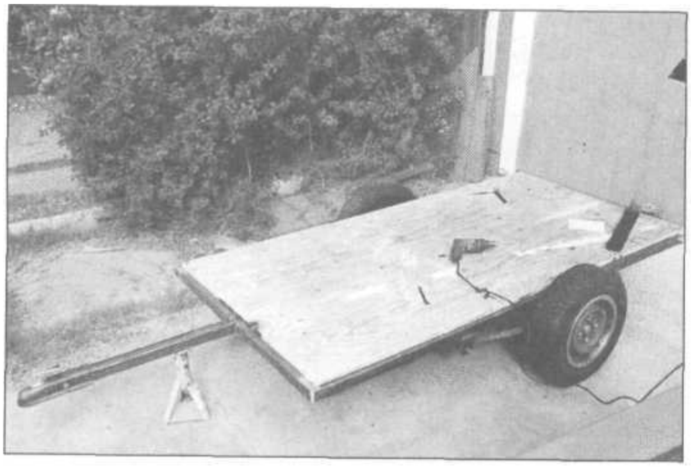

9.25 In our case, the frame was built just the right size to use an uncut 4x8-foot piece of 3/4-inch plywood for the trailer's floor.

9.26 Locate where your tubes are and drill through the flooring and tubing, then bolt the floor down with carriage bolts, with the round heads on top so the floor is relatively smooth and a furniture dolly will go over the floor easily.

enough tongue on the front and, therefore, not enough tongue weight. When loading your utility trailer, try to load the heavier items toward the front of the trailer.

When making joints in tubing, angle and other materials, get as tight a fit as you can before welding. A tight joint will always weld up as a stronger unit than something with gaps to be filled. If you are cutting with a hacksaw, use a carpentry miter-box to start your 45-degree cuts at the right angle. When you're done, give the whole thing a coat or two of rust-resistant paint, get an assigned registration number and license plate at your local Department of Motor Vehicles, and you're on your way.

When your trailer is complete, have a qualified weldor/metals expert inspect it for strength and safety. Then find some things to move!

9.27 Four posts of heavy 1x2-inch angle-iron were welded on, one in each corner, as supports for the plywood sides. After welding, all were checked with a square.

9.28 Having metal around for welding projects can pay off when working with wood, too. Use an extra length of straight angle-iron or tubing to clamp down on a piece of plywood as a saw guide, for perfectly straight cuts.

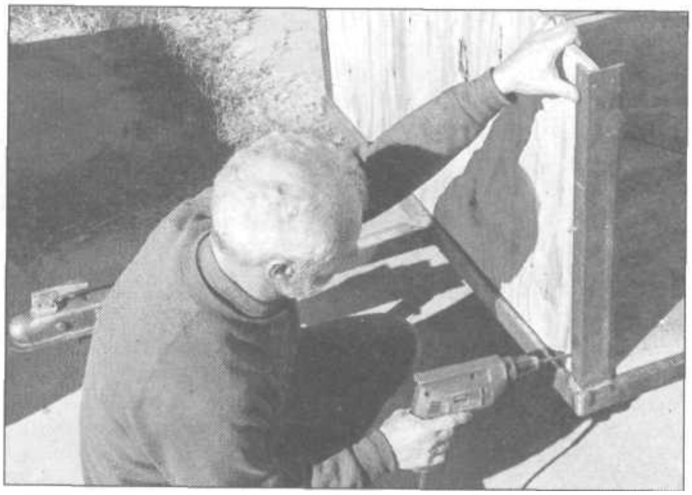

9.29 The front panel is attached with carriage bolts through the angle-iron posts. The posts were pre-drilled before welding on, because they are much easier to drill in a drill-press than once they are welded in place.

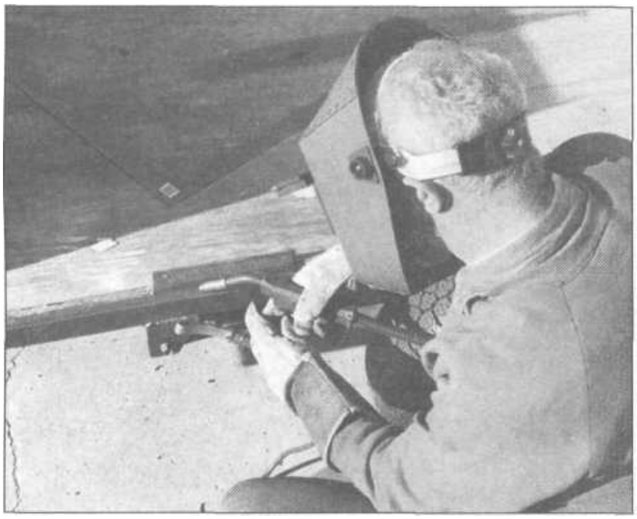

9.30 Short strips of 1/8-inch thick, 2-inch-wide strap metal (pre-drilled) were welded along the floor edges to help attach the bottom of the side panels. When MIG welding outdoors with argon gas shielding, use pieces of plywood around you to form a wind barrier to keep the gas from blowing away as you weld.

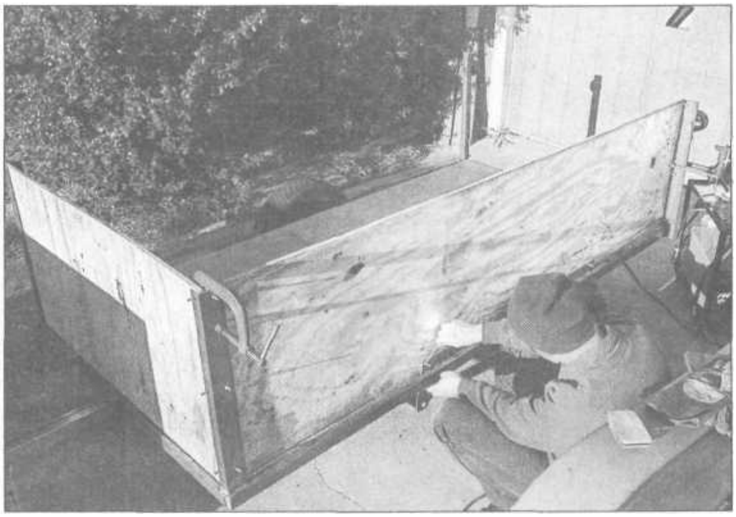

9.31 Plywood sides are bolted to the end pillars and the side plates.

9.32 At the local trailer supply store, we found a wide variety of lights, wiring harnesses, plugs and hitch parts to complete our utility project.

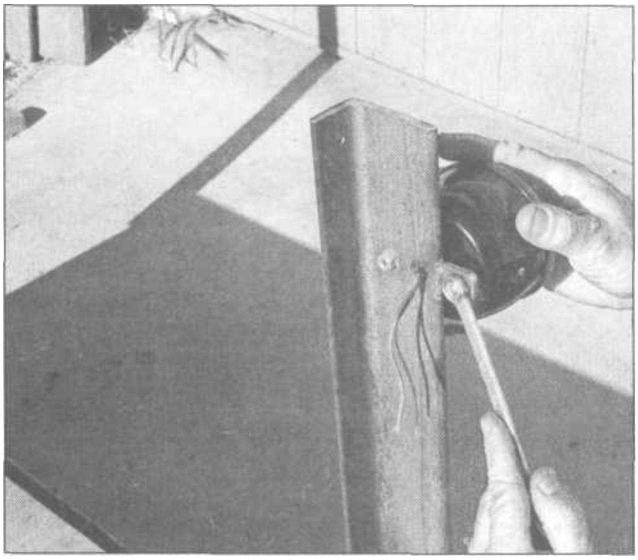

9.33 The taillights were bolted to the rear angle posts, but with a short extra piece of steel added on to move the lights further outboard to clear the tailgate.

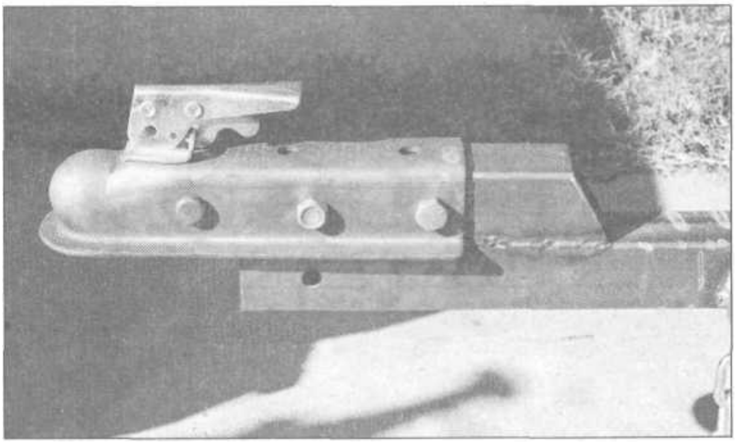

9.34 We originally had the hitch bolted to the main tongue, but found we had a better tongue weight with the tongue down another two inches. Simply welding on another piece of 2x2-inch, heavy-wall tubing set the trailer just right.

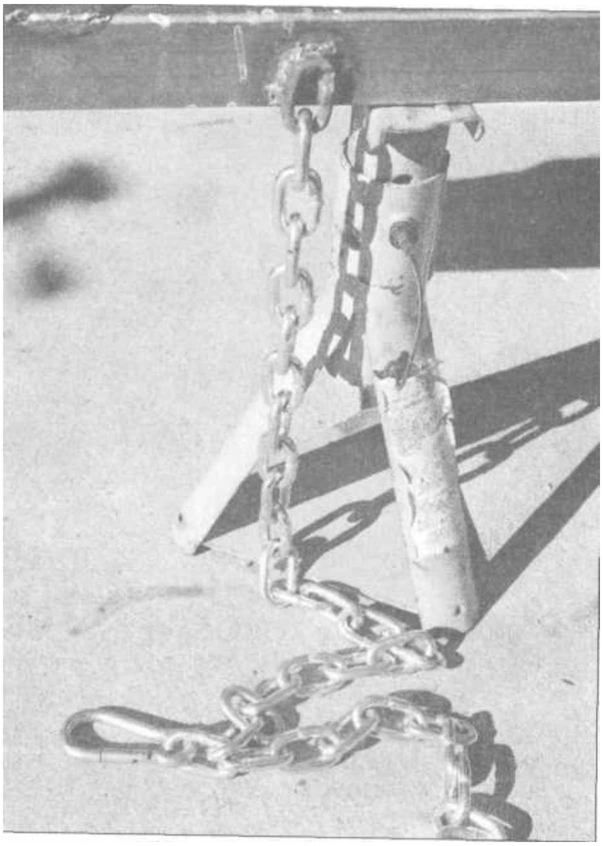

9.35 Every trailer needs a safety chain, check with your local DMV to see if your trailer qualifies for just one, due to light weight, or needs two to be legal. Take one end link, grind it clean of any plating and weld it to the trailer tongue (at top here). Below you'll see the removable snap-link that hooks to our tow vehicle.

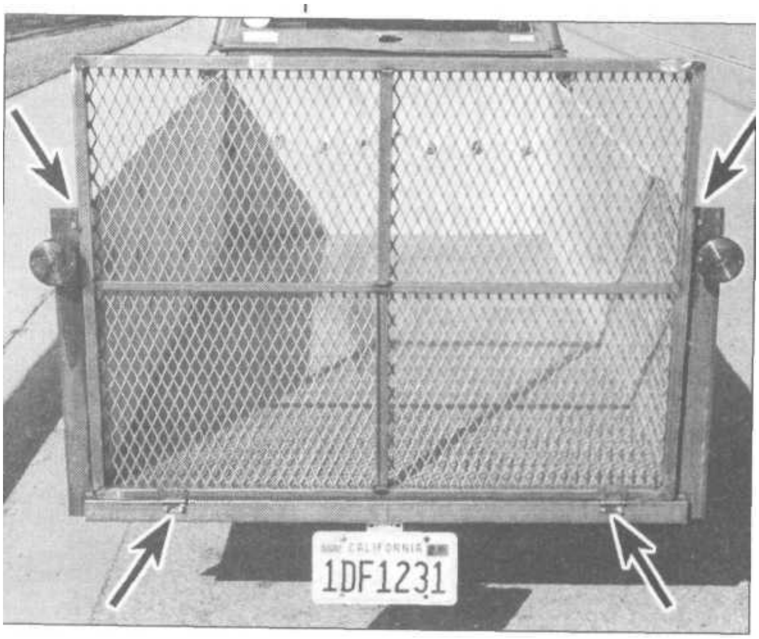

9.36 Our metal tailgate will save fuel because air flows through it on the road, cutting wind resistance. The frame was made of 1x1-inch tubing, to which was added steel mesh. Hinges (bottom arrows) were welded to the tailgate and bolted to the trailer. Top arrows indicate where tabs were welded to the gate for bolting with wingnuts to the trailer structure.

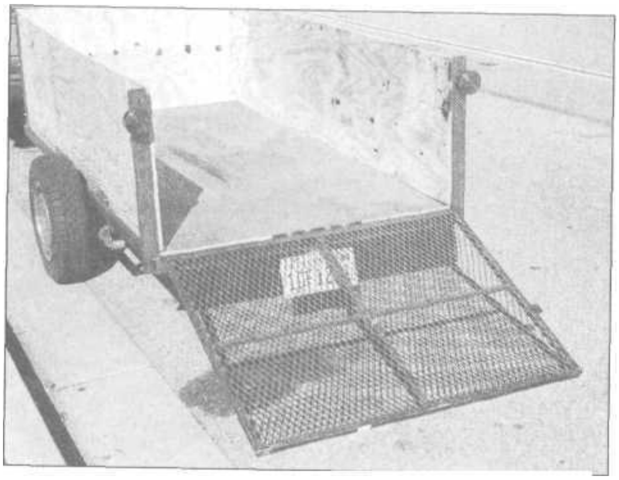

9.37 With the wingnuts undone, the hinged tailgate folds down as a ramp for furniture dollies, lawnmowers, etc. The trailer should be hooked up to the tow vehicle before you step on the tailgate/ramp, or the lightweight trailer could tilt quickly.

9.38 A 4x8x2-foot utility trailer actually hauls a lot of "stuff". We've known people who built a simple trailer just to make a long-distance household move, then sold the trailer when they got there, saving considerably on what a rental trailer would have cost.

Glossary of welding terms

AC Current - Alternating current, such as in homes and businesses. Polarity switches at 60 cycles per second (US).

Acetylene - A highly flammable gaseous hydrocarbon fuel used with oxygen in gas torch welding/brazing/cutting.

Alloy - A modified metal made by combining a base metal with other metals or chemicals to produce a metal with different properties.

Amperage - A measure of electrical power. In electric welding, such as TIG, MIG and arc, the higher the welding amperage, the more heat is generated at the arc.

Annealing - A process of heating a metal and slowly cooling it to soften it and make it more workable.

Arc Welding - An electrical welding process in which a consumable, flux-coated electrode rod makes an arc with the work. Also called stick welding. Arc welding deposits a coating of slag over the seam, which must be chipped off after it cools.

Acetylene bottles

6061-T6004-250/11

Alloy numbers on aluminum

Arc welding

Argon bottles

Arc Blow - A phenomenon in electrical welding where the magnetic forces in the work can cause a wind-like force that blows shielding gasses away. Usually happens in corner welding and can be avoided by switching polarity or moving the ground clamp.

Argon - A colorless, odorless inert gas often used as a shielding gas in TIG and MIG welding.

Backfire - A loud pop heard at the torch in oxy-acetylene welding that indicates combustion has backed up into the torch. A potentially dangerous phenomenon caused by holding the torch too close.

Base Metal - In alloying, this is the pure metal before alloys are added. In welding, this refers to the work metal, as opposed to the filler metal melted in.

Bead - The seam between two pieces of metal that have been fused by welding. The bead usually looks like a tight series of overlapping circles or ovals of metal.

Brazing - A type of metal-joining where the parent or base metal isn't melted but is heated only enough to melt a filler metal such as brass, which holds the parts together after cooling.

Butane - A fuel gas sometimes used with oxygen for brazing or soldering. Butane does not produce as much heat as acetylene.

Burn-Through - When the welding flame or arc is too hot for the material, holes develop in the seam.

Butt Weld - The most common joint in home/shop welding, where the two edges to be joined are lying flat in front of the weldor.

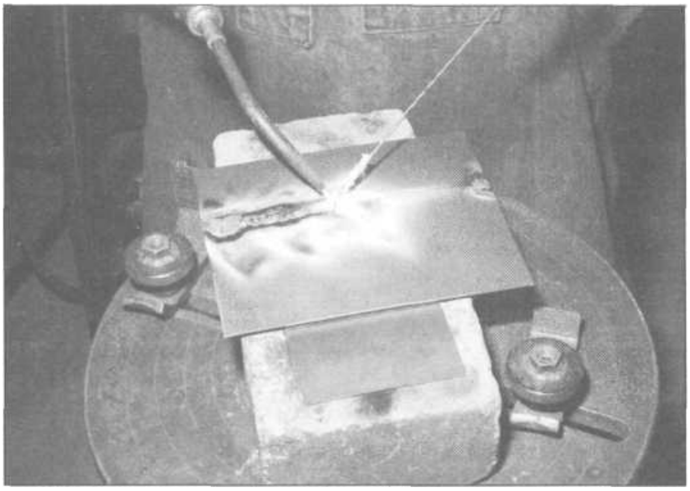

Brazing

Carbon - A non-metallic basic element that is a key ingredient in alloying of steels. Higher amounts of carbon in a steel make it tougher.

Carbon-Arc - A type of welding in which two carbon rods are brought together to create an arc independent of the parent metal. A process virtually obsolete in home/shop welding today.

Carburizing - In welding, an oxy-acetylene flame that is rich in acetylene, used to coat parent metals with soot before annealing. In metal work, a process of treating a surface such as a tool or knife edge by adding carbon to a hot metal, increasing surface hardness.

Cast Iron - A common metal, formed of steel with very high carbon and silicon content, in which soft graphite flakes appear throughout. The graphite makes the material hard to weld, but it is commonly used in large castings such as automotive engine blocks, heads and crankshafts, as it is inexpensive to manufacture and resists vibration damage.

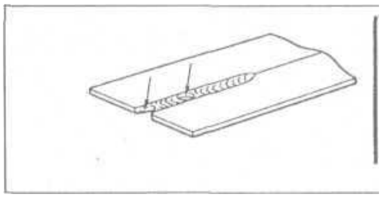

- The welding current is too high.
- The gap between the metal is too wide.
- The speed of the gun is too slow.
- The gun to base metal distance is too short.

Burn-through

Clad Metals - Usually mild steel alloys coated with another metal, such as galvanized steel, or sandwiched with another metal such as

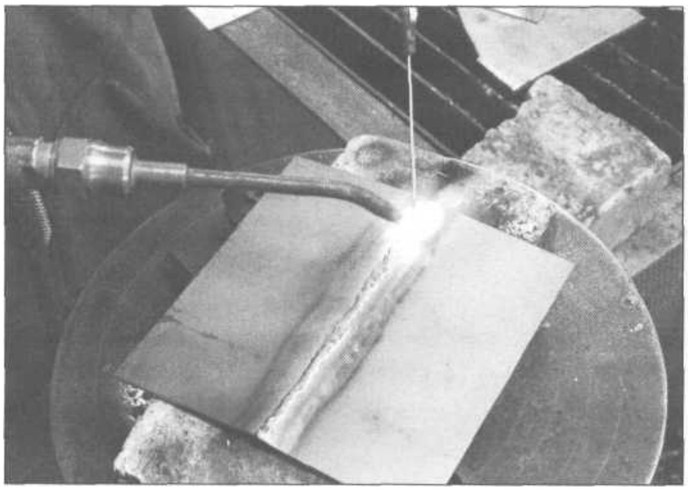

Butt weld

Deposit

kitchenware that has mild steel sandwiched between two layers of stainless-steel.

Copper - A reddish basic metal that is relatively soft, malleable and is one of the best conductors of heat and electricity.

Current - The flow of electricity, usually described by its direction of flow (alternating or direct) and strength (amperage and voltage).

Deposit - The addition of new metal to a welding seam by melting in a filler rod.

Deposition Rate - How fast filler metal is added to a welding seam. Determined by the wire-feed speed in MIG welding, and by the movement of the hand-held filler rod in gas, arc and TIG welding.

Ductility - A property of a metal that indicates how well it can be hammered out thin or drawn into a wire. A welded joint that passes a bending test without breaking is said to have good ductility.

Duty-Cycle - In electric welding processes, this is a way of gauging how long a particular machine can weld before it must be cooled off. It is expressed as a percentage of a ten-minute test period. At a 40% duty-cycle, a welder can be operated for four minutes, then must be cooled for six. At 100% duty-cycle, such as industrial equipment, the welder can be operated all the time.

Electrode - In electric welding, a conductive element that makes the final connection with the work to create an arc of electrical energy. In arc-welding, it is the rod or stick. In MIG it is the welding wire, and in TIG it is a tungsten rod.

Eye-Flash- A eye problem that occurs from observing a welding arc without protective goggles or lenses. Even brief exposure to the intense ultraviolet and infrared radiation can harm your eyes, causing a burning sensation that may show up 6-8 hours after exposure.

Ferrous Metals - Any of several metals made from iron and other elements. Common ferrous metals include wrought-iron, cast-iron, steel and cast-steel. Magnets attract ferrous metals.

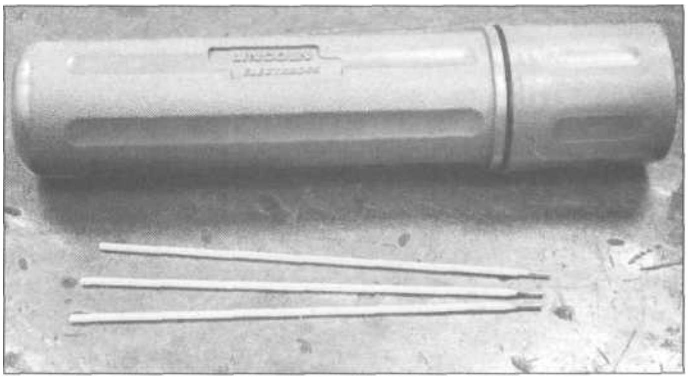

Electrode - arc rods

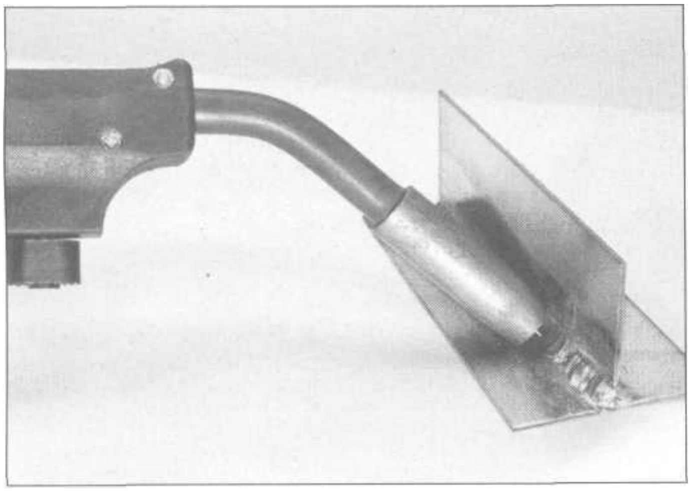

Fillet weld

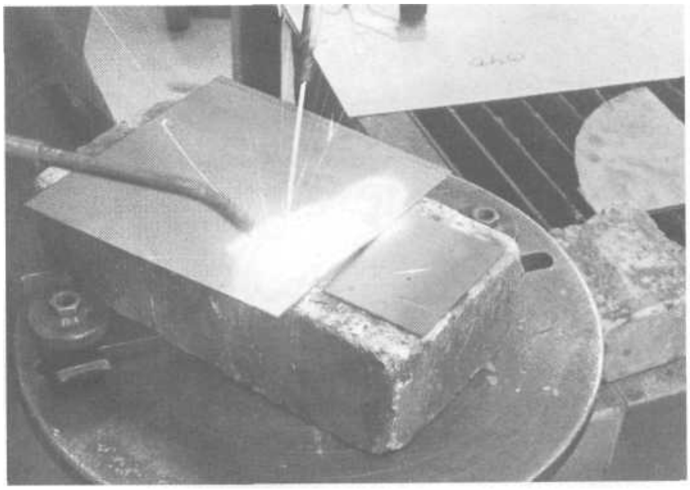

Firebrick

Filler Metal - New metal added to a welded joint by melting in a filler rod or wire. Most weld seams require filler metal.

Fillet Weld - A joint such as a 90-degree or T-joint, where the weld bead is filling a triangular-shaped area. Also describes a multiple-pass weld on thick, beveled plates.

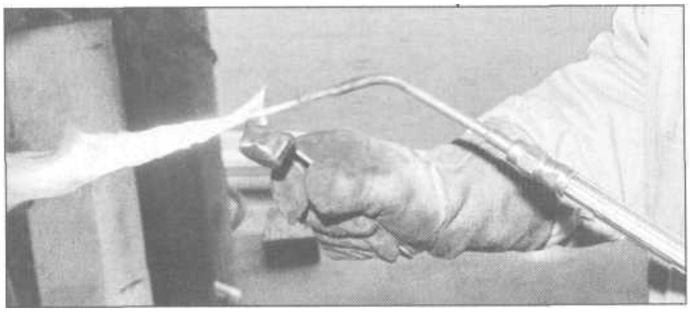

Flame

Firebrick- Special fire-proof bricks used in welding as a non-conductive base on which to set parts to be welded. Usually larger than common masonry bricks. Do *not* use concrete or red brick to weld on, they will crack or explode from trapped moisture inside.

Flame - The heat-creating, working-end in oxy-acetylene welding/cutting/brazing. The flame, not the torch, contacts the work.

Flamecutter - A mechanized gas or plasma torch that is motorized and follows a pattern, to cut out repetitive shapes from metal.

Flashback - A situation in gas-welding in which combustion goes back into the torch. More serious than a backfire, it is signaled by a loud squeal, and should be stopped immediately by turning off the gasses at their tanks.

Flashback Arrestors - One-way valves installed in oxygen and acetylene hoses to prevent flashback. All gas-welding setups should have these.

Flowmeter - A type of gauge for shielding gasses which indicates the flow via a colored ball floating in a clear glass tube which is easier to read from distance than a dial-type gauge.

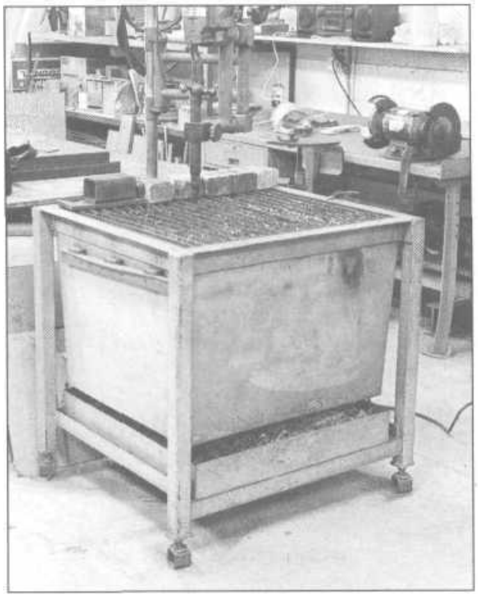

Flamecutter

FLASH ARRESTORS
TORCH-MOUNTED TYPE
UL-LISTED
STOP DANGEROUS FLASHBACKS
L-TEC
L-TEC
L-TEC

Flashback arrestors

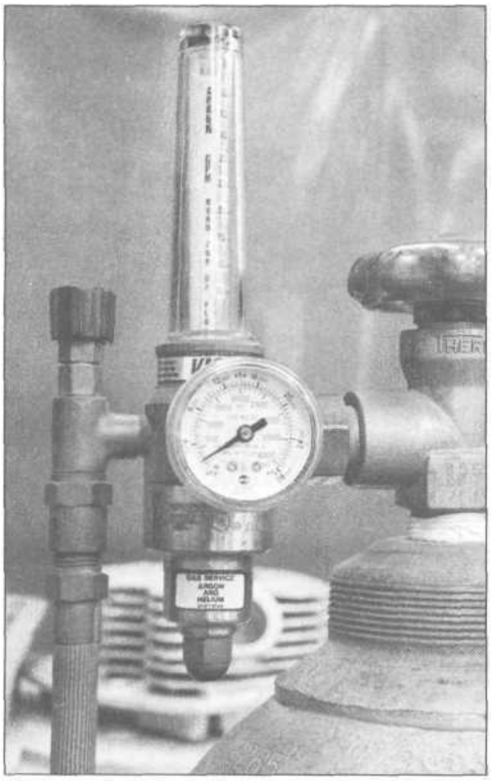

Flowmeter

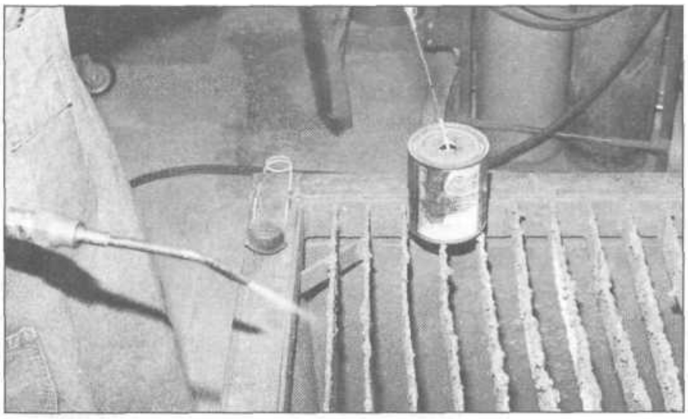

Flux

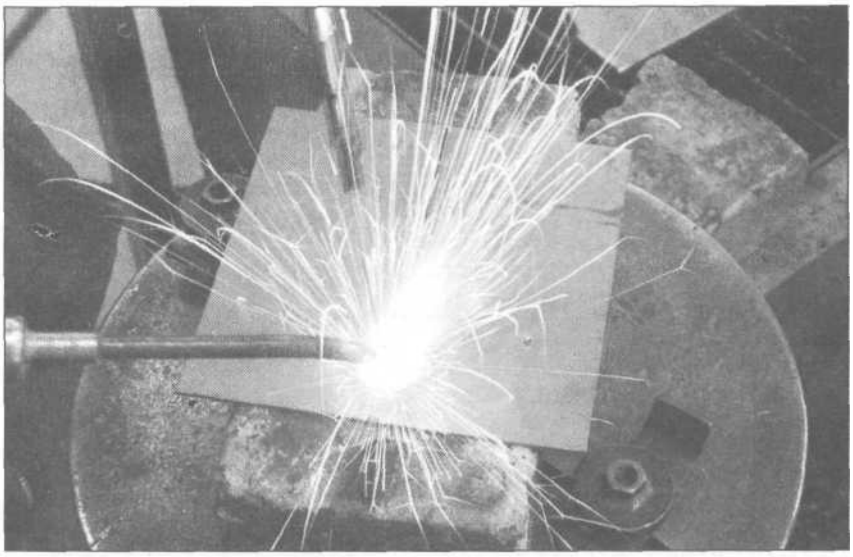

Gas welding

Flux - A chemical compound that cleans base metal under heat. In brazing, flux prepares the base metal for the brass to stick to it. Flux is applied as a coating to arc-welding rods and to wire for flux-cored MIG welding and is consumed in the arc to produce shielding gas.

Forge-Welding - A blacksmithing process where two pieces of metal are heated to the near-molten state and hammered together, resulting in fusion. Until the late 19th-century, all metals were either soldered, brazed or forge-welded.

Fusion Welding - Where the edges of two pieces of metal are brought to a molten state and joined, usually with filler metal of a similar alloy melted in.

Gas Welding - Using a mixture of oxygen and acetylene gasses to form a very hot, concentrated flame to heat and fuse metals.

GMAW - Stands for Gas Metal-Arc Welding, also known as MIG (Metal Inert Gas welding) and commonly called wire-feed welding. The process uses a shielding gas emitted from the torch while the consumable wire is the filler and electrode.

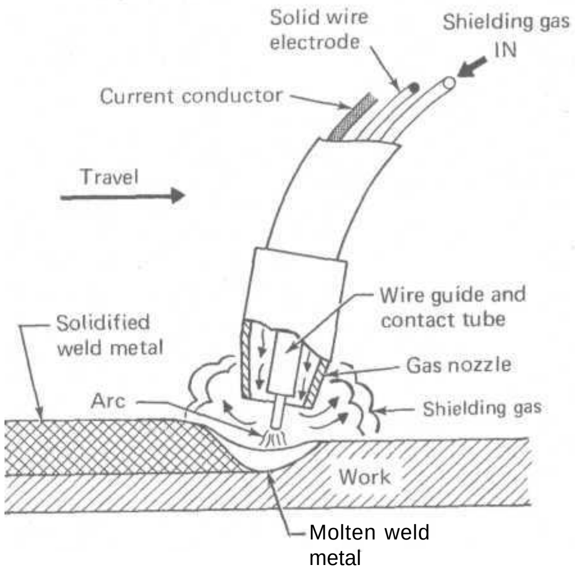
Solid wire electrode
Shielding gas IN
Current conductor
Travel
Wire guide and contact tube
Solidified weld metal
Gas nozzle
Arc
Shielding gas
Work
Molten weld metal

GMAW

GTAW - Stands for Gas Tungsten Arc Welding, also called TIG (Tungsten Inert Gas welding), and may be referred to as heli-arc welding, though this is a tradename. Helium or argon shielding gas emits from the torch, while a non-consumable tungsten electrode makes the arc, and filler is added from a hand-held rod like in gas welding.

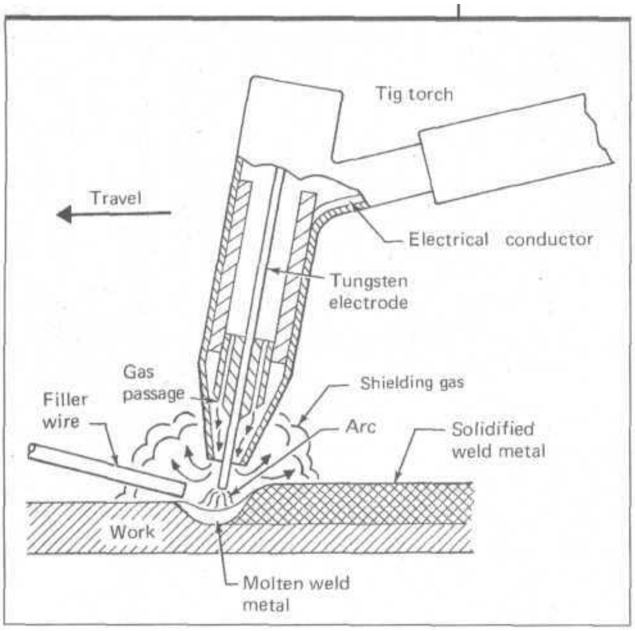

GTAW

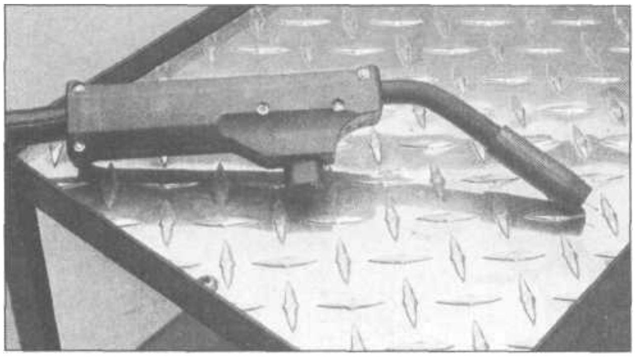

MIG gun

Ground Clamp - In electrical methods of welding, this is a heavy cable from the welding machine that clamps to the work or the conductive metal work table, completing the circuit from the electrode to the work and back to the machine.

Gun - In welding, this usually refers to the "torch," particularly in the MIG welding process. The similarity to a gun is that the wire "shoots" out of it into the weld bead.

Hardness - A metal property that can be tested by the ability of the metal to scratch another surface or be scratched. Sensitive equipment is used to measure metal hardness, and the alloy number of a metal is used as a hardness indicator.

Hard-Surfacing - In arc-welding, a procedure in which many weld beads are made overlapping, covering a large area. Using special rods, a very hard, abrasion-resistant face can be added to tools or farm implements.

Heat-Affected Zone - This is the area along either side of the weld that has been changed by the fusion heat. On ferrous metals, it is indicated by a discoloration. In some cases the metal in this zone may have become weaker than the joint itself.

Heat Sink - Something used behind or around an area to be welded, to protect adjacent areas from heat. Wet rags and moldable, heat-resistant clays are often used around welding on thin sheet metal.

Heat-Treating - Changing a metal's characteristics of strength and formability by scientifically heating and cooling it. The term usually implies making a metal harder.

Heli-Arc - A tradename of the L-TEC company (formerly Linde), for their TIG welding equipment. The original process began with Helium as the shielding gas, and the tradename

Small heat-affected zone on **MIG** weld

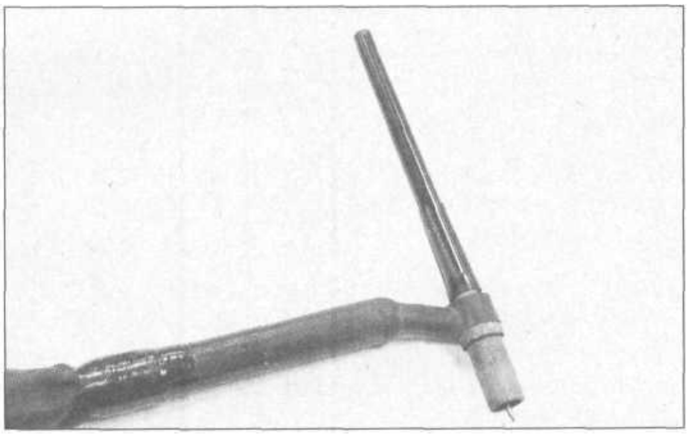

Heli-arc torch

came into such wide usage that it is now generic when used by weldors.

Helium - An extremely light, colorless, nonflammable gas, used in welding as a shielding gas and in everyday use to fill balloons.

Hydrogen - Atomically the simplest of all atoms, hydrogen in its normal state is a colorless, odorless lightweight gas that is extremely flammable. In small amounts it is used in some industrial welding/cutting applications.

Jig - A device used to keep parts in alignment or position during welding.

Joint - Where two parts are joined by welding, or where two sides of a crack are welded.

Kerf - The width of the cut in metal cutting by torch or plasma cutter.

Lap Weld - Where two pieces of metal are overlapped before welding, rather than butted together. Commonly used in brazing, and to stiffen a joint in thin sheet metal.

Lead - Another term for an electrical cable, such as a ground *lead.*

Machine - The power supply for an electrical welding setup.

Machine Welding/Cutting - When the torch or welding/cutting gun is moved by motorized equipment rather than by hand, resulting in a more precise weld or cut.

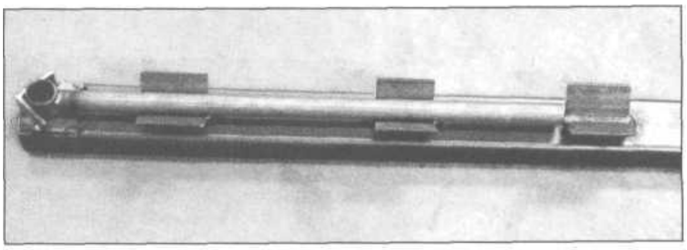

Jig

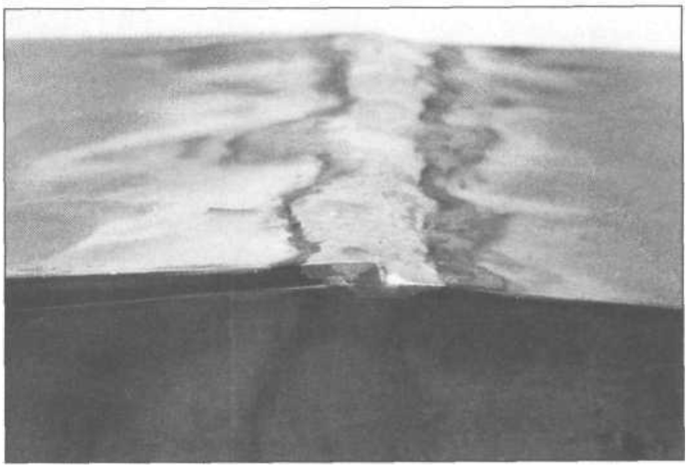

Brazed lap joint

TIG machine

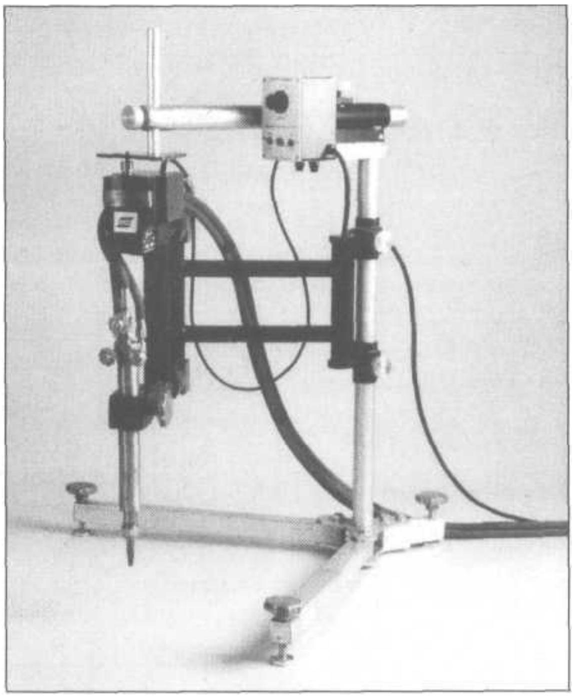

Machine welder/cutter

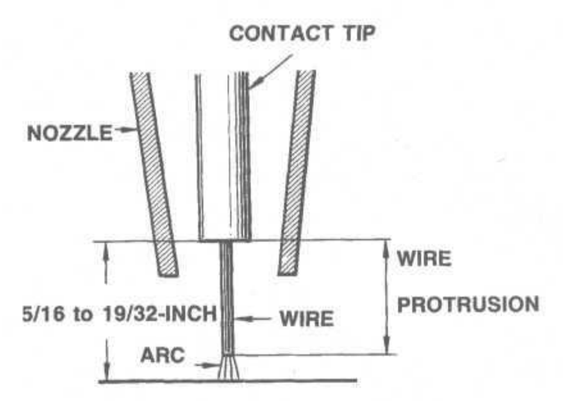

MIG

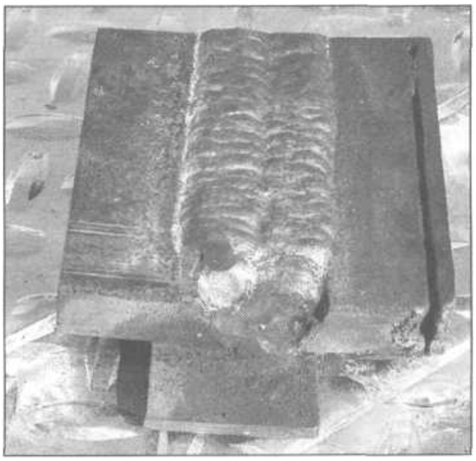

Multiple-pass weld

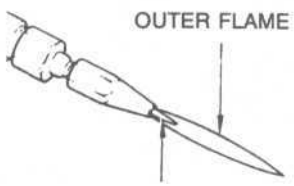

Neutral flame

Malleability - A metal quality indicating how well the material responds to being shaped by a hammer or forming tools.

Manifold System - Where a number of welding torches are fed by a common source of gasses, such as in an industrial situation or a welding school.

MIG - Stands for Metal Inert Gas welding, more commonly called wire-feed welding because the consumable electrode is a metal wire machine-fed through the gun from a roll.

Mild Steel - Synonymous with low-carbon steel, the most commonly found and highly useful material, easy to weld, heat and form.

Molybdenum - A modern metal used in alloying, to make steels and stainless steels stronger and more resistant to corrosion. A well known steel alloy is 4130 chrome-moly which contains molybdenum and chromium and is a strong material used in aircraft and race cars.

Multiple-Pass Weld - When welding thick plates or repairing large cracks in castings, more than one bead must be made. The joint is usually beveled first and the succeeding passes have wider beads until the joint is filled.

Neutral Flame - In oxy-acetylene welding and cutting, this is the desired adjustment of the torch flame, with equal amounts of each gas.

Nickel - A metal used in alloying steel and stainless steels chiefly to increase corrosion resistance and suitability for both very low and very high temperature extremes.

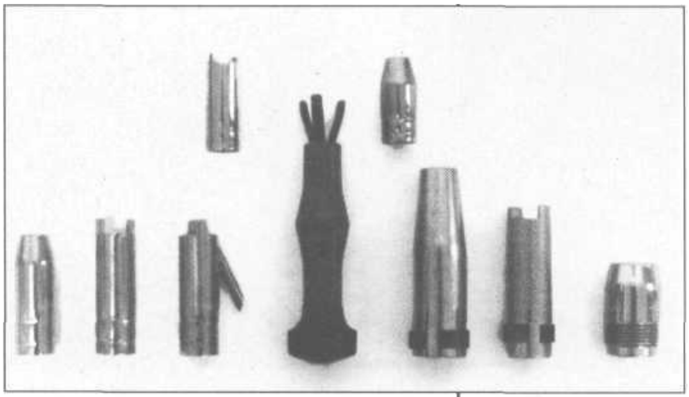

MIG nozzles

Nitriding - A metal treatment process to surface-harden certain steels, castings and forgings.

Nitrogen - The main component (75%) of the air we breathe, nitrogen is an inert gas that can be used as a shielding gas in welding copper, or as the cutting gas in plasma equipment.

Normalizing - A heat-treating process that refines the grain structure in hard steels. It is used when high-strength steels have been welded, to remove stresses.

Nozzle - A component part of the torch in TIG, MIG and plasma welding equipment, designed to focus the shielding gas over the arc.

Out-Of-Position Welding - Welding whenever the parts are not laid flat in front of the weldor, such as welding vertically, overhead or horizontally but perpendicular to the ground.

Oxidation - The deterioration process when metals are exposed to oxygen through air, moisture and natural or manmade harmful elements, On steel it appears as red-brown rust, on aluminum it can be a gray-black. To prevent it, metals must be coated with oils, paints or plating processes.

Oxidizing Flame - In oxy-acetylene welding/cutting, when the flame is adjusted with too much oxygen, identified by a ragged purple outer cone and a hiss from the torch.

Oxygen - A colorless, odorless, tasteless gas that makes up 21 % of the air around us. It is essential to combustion processes, and combined with equal parts acetylene makes a 3315° C flame for welding and cutting.

Penetration - Important in all welding processes, penetration describes how far into the two pieces of a joint the fusing process extends. Maximum penetration makes the strongest joint.

Piercing - In metal cutting, piercing is the technique for starting a cut in the middle of the material (such as cutting an *interior* hole in a piece of plate). It usually requires slightly different torch procedure than starting from an edge.

Oxygen bottles

Pickling - Treating a metal chemically before welding it. For instance, cad plating on bolts, chain and other hardware can create dangerous fumes when welded, plus good contact is hard to make. Plated parts should be stripped in swimming-pool acid before welding.

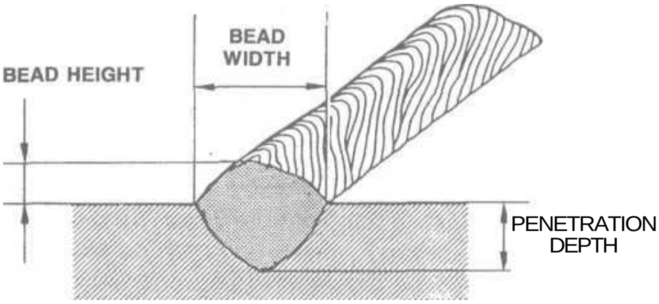

Penetration

Plug Weld - Where a piece of tubing fits inside another and is welded to it through holes drilled in the outer tube. Also called a rosette weld.

Polarity - The direction in which electrical current flows in a welding setup. There is AC, which alternates in 60 cycles per second (US), and DC, direct current, which can be either *straight* polarity or *reversed* polarity.

Porosity - An undesirable quality in a weld, with tiny holes caused by welding dirty metal, using damp arc electrodes, or getting a gas torch too close to the metal.

Post-Flow - A setting on a TIG machine allows the weldor to adjust how long after the arc is broken the shielding gas will continue to flow to protect/cool the bead.

Post-Heating - Heating a welded part after the final welding to relieve stresses.

Pre-Heating - When an entire part, or just the weld area, is heated with an oven or torch before welded to prevent thermal stresses and uneven expansion and contraction. Thick metals and aluminum and iron castings are usually preheated before welding.

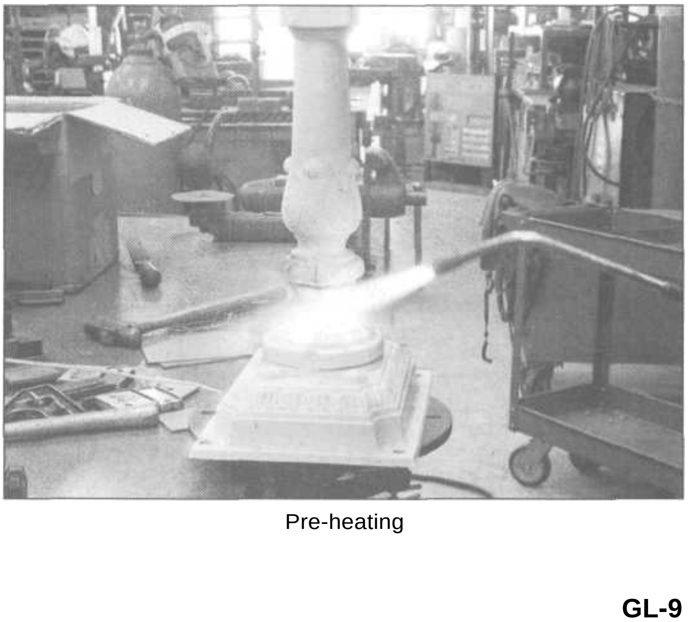

Pre-heating

Propane - A fuel gas commonly used in self-contained torches for heating and soldering plumbing pipes. It can be used with oxygen to perform welding/brazing on thin materials.

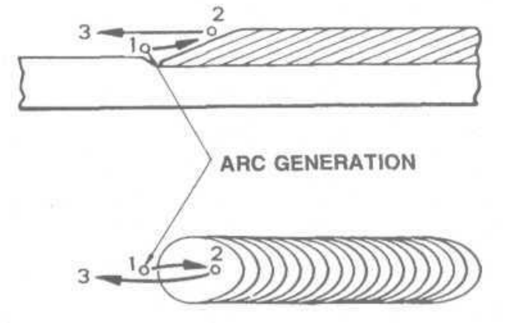
3
1
2
ARC GENERATION
3
1
2

OVERLAPPING BEADS

Puddle

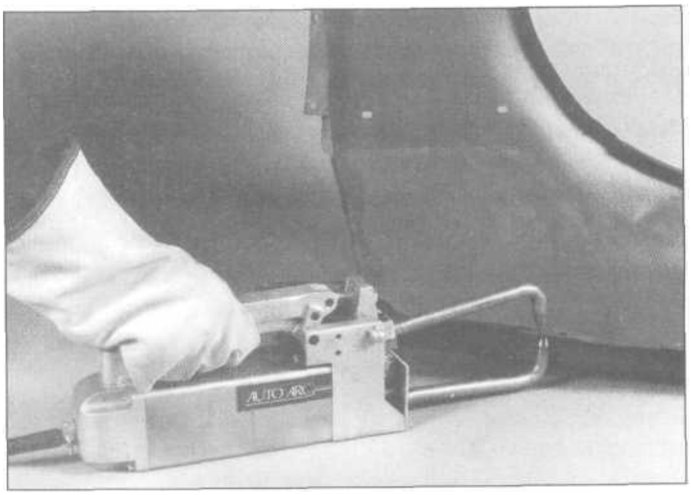

Resistance welding

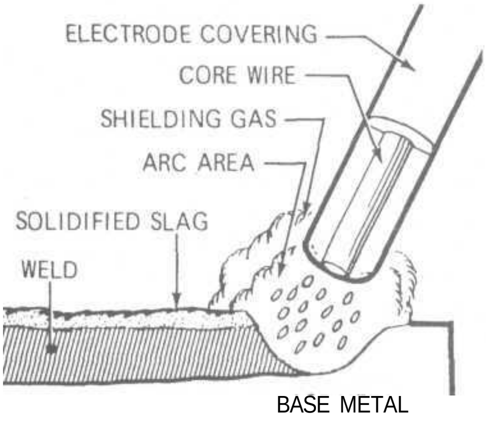
ELECTRODE COVERING
CORE WIRE
SHIELDING GAS
ARC AREA
SOLIDIFIED SLAG
WELD
BASE METAL

Arc welding rod

Puddle - The welding bead in process, as the concentrated area of the parent metal is melted and joined with molten filler rod. The finished bead is a continuous string of overlapping puddles.

Rectifier - A major component of many electric welding machines, the rectifier changes AC input current to DC output.

Resistance Welding - A method of welding where electrodes are placed on either side of the material to be joined, and a brief arc is made, which is enough to fuse the materials together at that point without greatly affecting the area around it. The weld is made by heat and pressure, and the technique is often used as spot-welding on sheet metal.

Rod - The filler rod in welding processes. In arc-welding, rod is another term for the coated, consumable electrode. In gas and TIG welding, the filler rod is usually uncoated and hand-fed into the puddle.

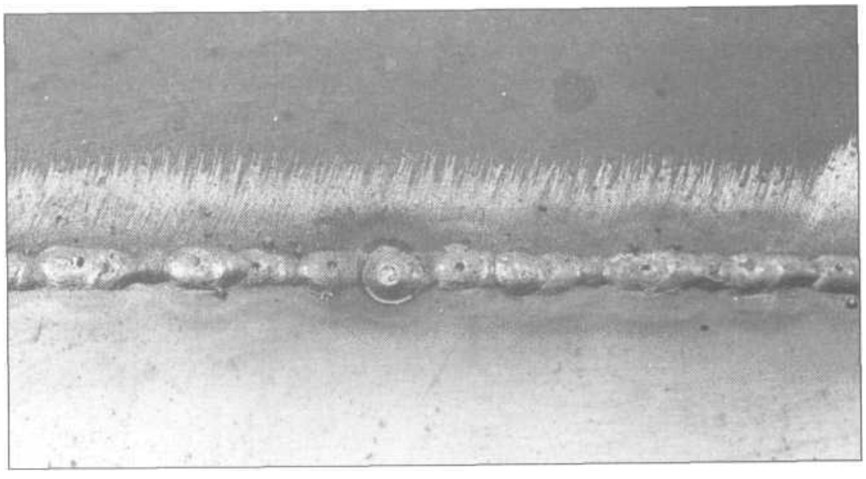

Seam welding

Glossary of welding terms

Root - The bottom of a weld bead viewed in cutaway. In multiple-pass welding oh thick material, the first bead in the bottom is called the *root pass.*

Reverse Polarity - In electric welding, the term refers to a DC setup where the work is made *negative,* and is called DCRP. It is the opposite of DCSP, or *straight* polarity.

Seam Welding - A procedure used on lighter-gauge metals where the bead is composed of overlapping spot welds. Some MIG machines can be timed to perform seam welds. It is also used on machine-made welds to join long seams in sheet metal.

Sheet metal - Thinner metals, such as found in auto bodies, appliances, heating/cooling ducts. Usually refers to metals from 12-gauge to 24-gauge.

Shielding Gas - An inert (nonflammable, non-reactive) gas forced out around a weld procedure to cover the arc area, excluding air and impurities from the forming weld.

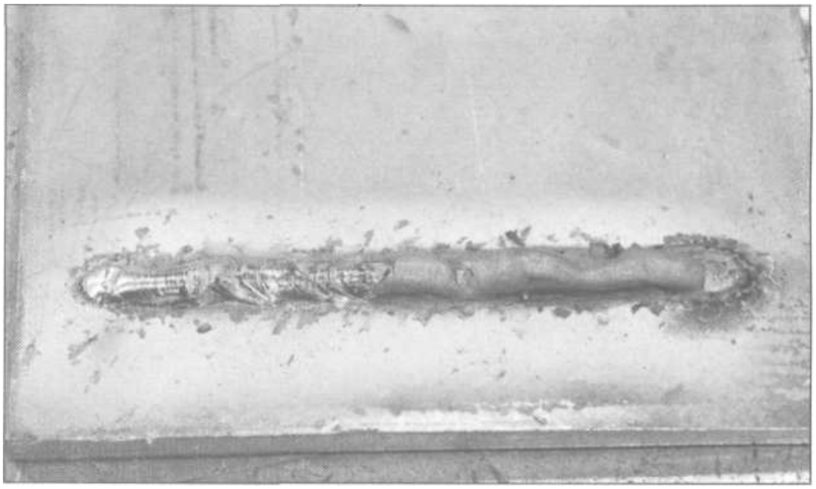

Slag

Shielded Metal Arc Welding - Another term for stick or arc-welding, it is abbreviated as SMAW. The flux coating on the consumable electrodes vaporizes into shielding gas to protect the weld.

Short Circuit - In electrical terms, this is what makes the arc when there is an air gap between positive and negative current flow. When you plug an appliance that is already switched on into the wall outlet and there is a tiny spark at the plug's prongs, this is a small-scale version of the short-circuit that makes electric welding possible.

Slag - Oxidized impurities resulting from welding and cutting. In arc welding, the slag forms a thick, hard coating over the seam, which must be chipped off with a pointed hammer. Slag is also found along the bottom of a cut made by a gas torch.

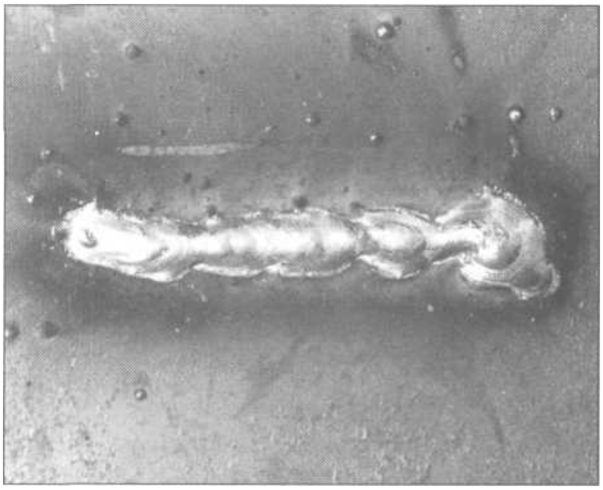

Spatter

Slope Control - In electrical welding, professional-level welders have controls that adjust the shape of the current waves, to weld better on non-ferrous materials. Slope control can also be used to describe controls that gradually increase current at the beginning of a weld, and taper off current at the end, both to avoid cratering.

Spatter - Tiny balls of filler metal stuck to the parent metal around the weld, spatter is produced mostly by incorrect welding technique, but in arc welding spatter is unavoidable. Spatter is also present in smaller amounts with MIG welding.

Spot Welding - Small circular welds used to hold light sheet-metal panels together. Usually made through a hole in the topmost panel and common in auto bodywork (also see resistance welding).

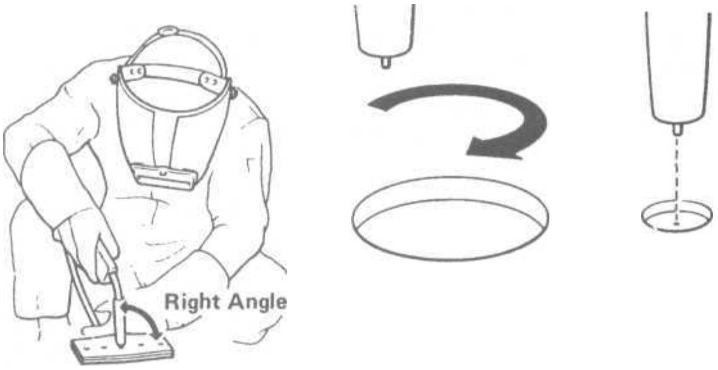

Spot welding

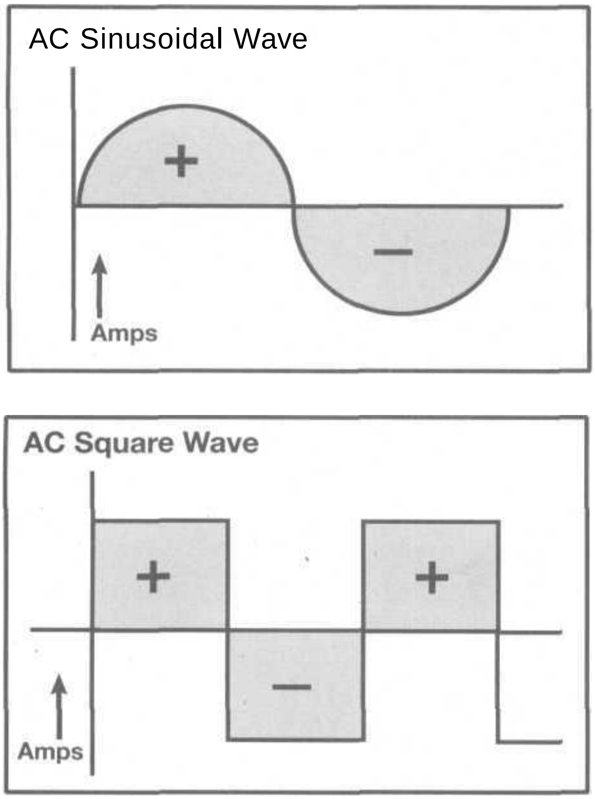

2110-6-13 HAYNES

Square-wave technology

Stick Welding - Another term for arc welding.

Stitch Welding - A seam where welding is done only in short strips with gaps in between where full welding isn't required. Better MIG machines have timer controls to easily do stitch welding.

Stud Welding - An option for some types of MIG welders whereby a special attachment to the gun accepts hollow studs. The studs are pressed onto a surface where the arc is made, and filler wire welds the stud to the surface and fills the inside of the stud. A very fast method of replacing broken studs in exhaust system work.

Square-Wave Technology - An electronic method of changing the shape of the AC current wave for improved welding of non-ferrous metals with TIG equipment.

Submerged-Arc Welding - An industrial process where the seam is buried under a long, thin pile of flux material, and a mechanized welder travels along the seam welding *underneath* the flux pile. The arc is not visible, and the technique is used often on heavy plates such as in shipbuilding.

Synchronous Control - A type of electronic control which ensures that each weld will begin at the same point on the wave cycle, particularly important for consistent spot welds in automatic operation.

Tack Welding - Very short welds made at each end and the middle of a seam, to hold the pieces in alignment before a full seam is welded. When designing and building a project, it's important to tack-weld everything first before finish welding. You may want to change something at the last minute and tack-welds can be easily broken off or cut through to make changes.

Tempering - Steels can be hardened by heat-treating but may become too brittle. Tempering heats a metal to some degree below the transformation temperature and then cools it to bring back some toughness in the metal.

Tensile Strength - A measure of a metal alloy's (or welded seam's) ability to resist being pulled apart.

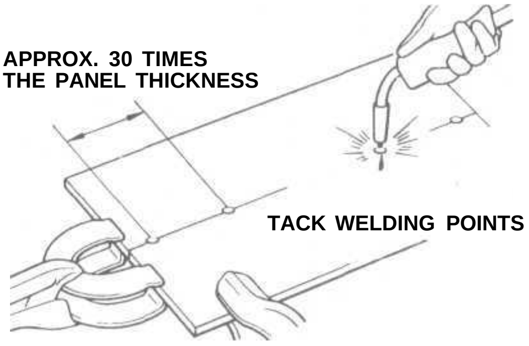
APPROX. 30 TIMES THE PANEL THICKNESS
TACK WELDING POINTS

Tack welding

TIG Welding - This stands for Tungsten Inert Gas welding (see heli-arc).

Titanium - An expensive, strong, lightweight material used in aerospace and race cars. Special techniques are required to weld titanium.

Transformer - A major electronic component of electric welding machines, it turns high-voltage, low-amperage line current into low-voltage, high-amperage welding current.

Tungsten - The non-consumable electrode used in TIG welding. It is resistant to corrosion and doesn't melt until 6170° F.

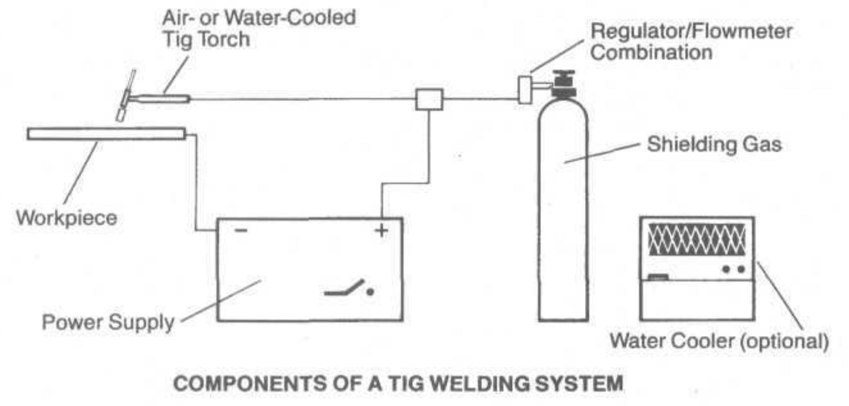
Air- or Water-Cooled
Tig Torch
Regulator/Flowmeter
Combination
Shielding Gas
Workpiece
−
+
Power Supply
Water Cooler (optional)
COMPONENTS OF A TIG WELDING SYSTEM

TIG welding

Undercut - A weld defect where too much heat was applied and at the sides of the bead the parent metal is eroded. Usually accompanied by too much penetration on the backside of the weld.

Voltage - A characteristic of electrical current that equals the amps times the ohms (resistance). In welding machines, input current is usually described in volts, while welder output is usually measured in amps.

Welder - The machine that makes the welded.seam.

Weldor - The person who operates the welder.

Wrought Iron - A soft form of iron that includes some slag in its formation. Traditionally the material chosen for making ornate fences, gates and horseshoes because of its shapability.

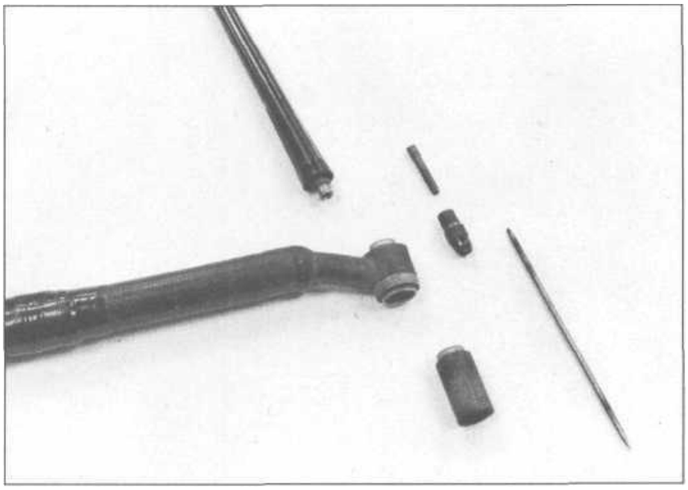

Tungsten

Index

Alloys, metal, 2-3
Angle clamp, 8-12
Anti-heat paste, 8-12
Arc welding, 2-9, 4-1 through 4-16
    arc welders, AC or DC?, 4-4
    beginning, 4-8
    choosing electrodes, 4-13
    safety considerations, 4-7
    the arc process, 4-6
    types of joints, 4-11

B

Bandsaw, portable, 8-5
Beginning arc welding, 4-8
Belt sander, 8-8
Bender, 8-7
Blanket, welding, 8-4
Brazing, 3-16
Building a utility trailer, 9-1 through 9-10

Checking your welds (oxy-acetylene), 3-14
Choosing a MIG welder, 5-5
Choosing electrodes (arc welding), 4-13
Choosing wire (MIG welding), 5-14
Clamp, magnetic, 8-5
Cleco clamps, 8-8, 8-9
Copper spoon, 8-4
Cutoff saws, 8-6
Cutting (with gas cutting torch), 3-18

Definitions of welding, 1-2
Development of modern welding, 1-3
Disc sander, 8-7
Duty cycles, 2-17, 4-3, 5-7

Electrode rack, 8-7

Flame adjustment, oxy-acetylene welding, 3-9
Flange punch, 8-9

- **Gas welding/cutting, 2-6, 3-1 through 3-26**
  - checking your welds, 3-14
  - cutting, 3-18
  - equipment, 3-3
  - flame adjustment, 3-9
  - heating, 3-24
  - how to, 3-9
  - the process, 3-1
  - welding with filler rod, 3-11
- **Gloves, welding, 8-4**
- **Grinder, 8-5**
- **Gun, spot welding, 8-5**

H

- **Heli-arc welding (TIG), 6-1 through 6-18**
- **Helmets, welding, 8-2**
- **How welding works, 2-2**

I

- **Introduction, 1-1 through 1-4**

M

- **Magnetic blocks, 8-12**
- **Materials rack, 8-11**
- **Metal alloys, 2-3**
- **MIG (wire-feed) welding, 2-11, 5-1 through 5-26**
  - choosing a MIG welder, 5-5
  - choosing a shielding gas, 5-12
  - choosing wire, 5-14
  - description, 2-11
  - learning, 5-16

N

- **Nibbler, 8-9**
- **Notcher, tubing, 8-5**

- **Oxy-acetylene gas welding/cutting, 2-6, 3-1 through 3-26**
  - checking your welds, 3-14
  - cutting, 3-18
  - equipment, 3-3
  - flame adjustment, 3-9
  - heating, 3-24
  - how to, 3-9
  - the process, 3-1
  - welding with filler rod, 3-11

- **Plasma arc welding and cutting, 2-18, 7-1 through 7-12**
  - choosing plasma cutting equipment, 7-5
  - cutting, 7-3
  - how to use, 7-8
  - welding, 7-2
- **Post-flow (TIG welding), 6-6**
- **Practice and training, 2-21**
- **Punch, flange, 8-9**

- **Rack**
  - electrode, 8-7
  - materials, 8-11
- **Rewiring for an arc welder, 4-5**

- **Safety & shop equipment, 8-1 through 8-12**
- **Safety considerations, arc welding, 4-7**
- **Sandblaster, 8-12**
- **Sander**
  - belt, 8-8
  - disc, 8-7
- **Saws**
  - bandsaw, portable, 8-5
  - cutoff, 8-6
- **Slope control (TIG welding), 6-6**
- **Spoon, copper, 8-4**
- **Spot weld remover, 8-6**
- **Spot welding gun, 8-5**
- **Stud setter, 8-10**

- **Table, welding, 8-4**
- **TIG {heli-arc) welding, 2-14, 6-1 through 6-18**
  - aluminum, 6-13
  - equipment, 6-4
  - how to, 6-8
  - post-flow, 6-6
  - slope control, 6-6
- **Trailer, utility, building, 9-1**
- **Training and practice, 2-21**
- **Tubing bender, 8-7**
- **Tubing notcher, 8-5**
- **Types of welding, 2-1 through 2-22**

U

Utility trailer, building, 9-1

W

Welding
- Arc welding, 2-9, 4-1 through 4-16
  - arc welders, AC or DC?, 4-4
  - beginning, 4-8
  - choosing electrodes, 4-13
  - safety considerations, 4-7
  - the arc process, 4-6
  - types of joints, 4-11
- blanket, 8-4
- definitions of, 1-2
- development, 1-3
- helmets, 8-2
- how it works, 2-2
- MIG (wire feed) welding, 2-11, 5-1 through 5-26
  - choosing a MIG welder, 5-5
  - choosing a shielding gas, 5-12
  - choosing wire, 5-14
- oxy-acetylene (gas), 2-6
- plasma arc, 2-18
- safety considerations, arc welding, 4-7
- table, 8-4
- TIG (heli-arc) welding, 2-14, 6-1 through 6-18
  - aluminum, 6-13
  - equipment, 6-4
  - how to, 6-8
  - post-flow, 6-6
  - slope control, 6-6
- types of, 2-1 through 2-22
- wire-feed, 2-11
- with filler rod (oxy-acetylene), 3-11

Sourcelist

Airco Gas and Gear
BOC Gasses
575 Mountain Avenue
Murray Hill, NJ 07974

Daytona **MIG**
1821 Holsonback Drive
Daytona Beach, FL 32117

Eastwood Company
580 Lancaster Avenue
Malvern, PA 19355

HTP America
261 Woodwork Lane
Palatine, IL 60067

Lincoln Electric Co.
22801 St. Claire Avenue
Cleveland, OH 44117-1199

L-TEC Welding & Cutting Systems
The ESAB Group
P.O. Box 100545
Florence, SC 29501 -0545

Miller Electric Mfg. Co.
718 S. Bounds Street
Appleton, Wl 54911

HAND SOLDERING TUTORIAL FOR FINE PITCH QFP DEVICES

Scope

This document is intended to help designers create their initial prototype systems using Silicon Lab's TQFP and LQFP devices where surface mount assembly equipment is not readily available. This application note assumes that the reader has at least basic hand soldering skills for through-hole soldering. The example presented will be the removal, cleanup and replacement of a TQFP with 48 leads and 0.5 mm lead pitch.

Safety

Work should be done in a well-ventilated area. Prolonged exposure to solder fumes and solvents can be hazardous. There should be no presence of sparks or flames when solvents are in use.

Materials

The right materials are key to a good solder job. The list below are the recommendations from Silicon Labs. Other materials may work, so the user should feel free to substitute and experiment. The use of organic solder is highly recommended.

Required

1. Wire wrap wire (30 gauge) *
2. Wire strippers for wire wrap wire *
3. Soldering station - variable temperature, ESD-safe. Should support temperatures 800°F (425°C). This example uses a Weller model EC1201A. The soldering wand should have a fine tip no more than 1 mm wide.
4. Solder - 10/18 organic core; 0.2" (0.5 mm) diameter
5. Solder flux - liquid type in dispenser
6. Solder wick - size C 0.075" (1.9 mm)
7. Magnifier - 4X minimum. An inexpensive headset OptiVISOR by Donegan Optical Co. is used for this example.
8. ESD mat or tabletop and ESD wrist strap - both grounded
9. Tweezers with pointed (not flat) tips
10. Isopropyl Alcohol
11. Small stiff bristle brush for cleaning (nylon or other non-metallic material). Cut off hair to approximately 0.25" (6 mm)

* Required for device removal only.

Optional

1. Board vise to hold printed circuit board
2. Dental pick (90 degree bend)
3. Compressed dry air or nitrogen to dry boards

Copyright © 2003 by Silicon Laboratories

4. Optical inspection stereo microscope 30-40X

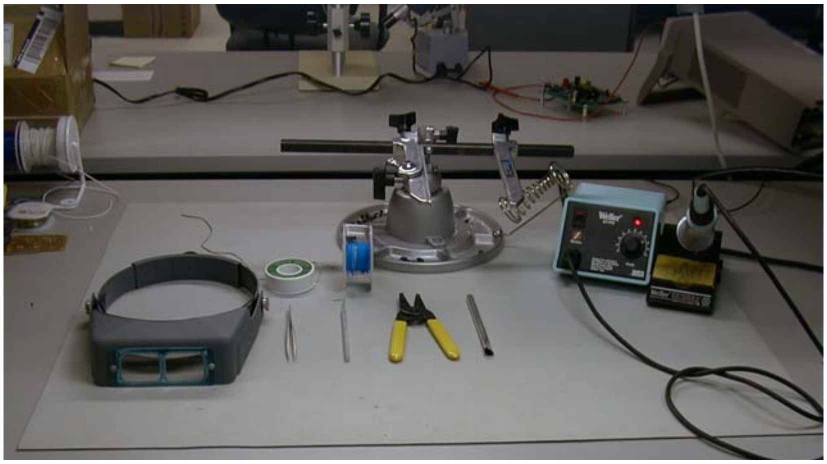

Figure 1. Some of the materials to get started...

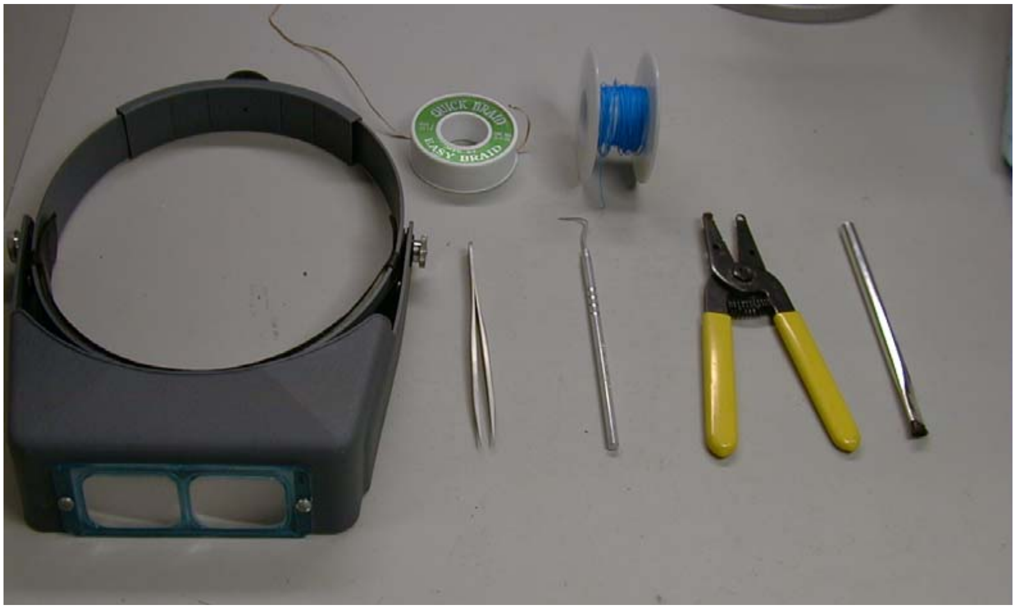

Figure 2. clock-wise from left: 4X magnifying headset, solder wick, wire wrap wire, stiff cleaning brush, wire strippers, and pointed tweezers

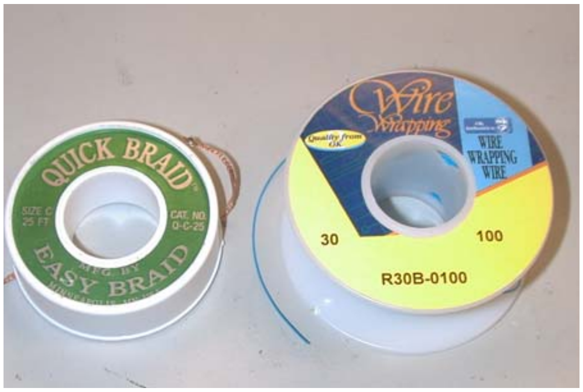
QUICK BRAID
EASY BRAID
Wire Wrapping
WIRE WRAPPING WIRE
30
100
R30B-0100

Figure 3. Solder wick and wire wrap wire

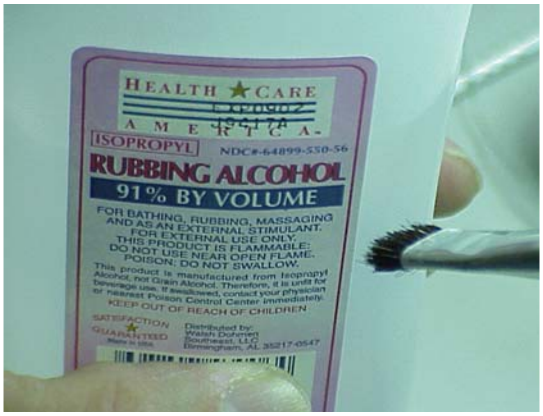
HEALTH ★ CARE
AMERICA
ISOPROPYL
RUBBING ALCOHOL
91% BY VOLUME

Figure 3. Isopropyl Alcohol

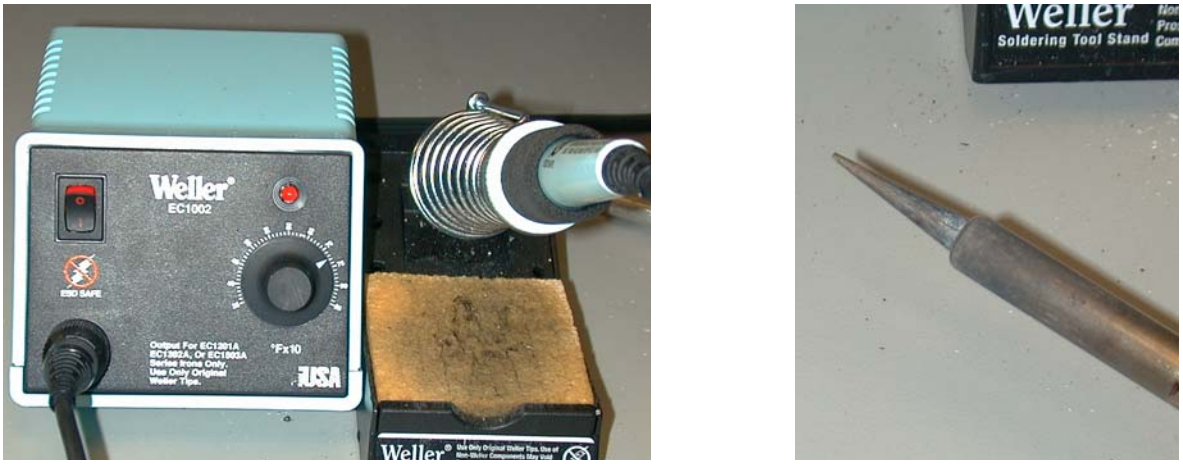
Weller
EC1002
ESD SAFE
°Fx10
USA
Weller
Soldering Tool Stand

Figure 4. ESD safe soldering station with fine tip wand. This one is Weller model EC1201A

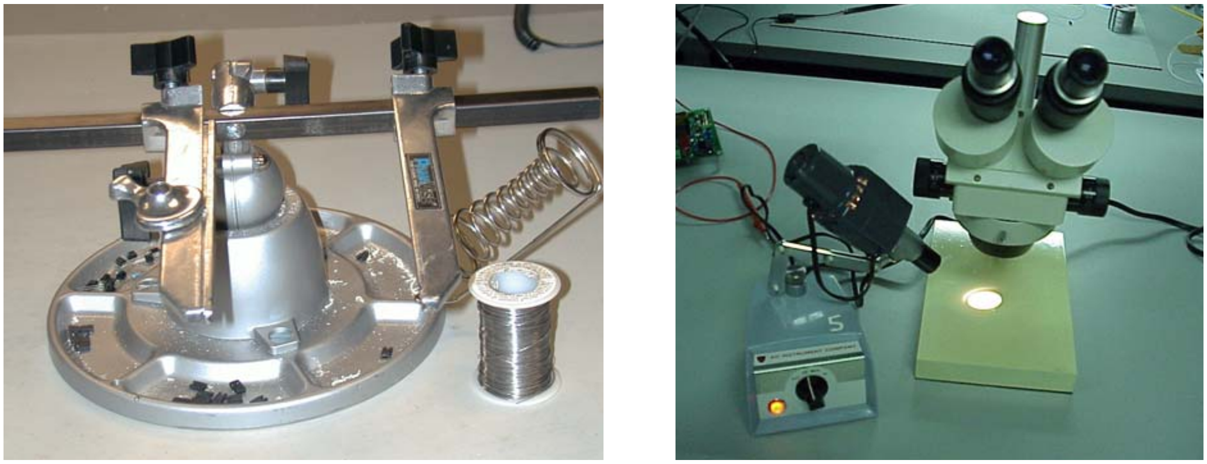

Figure 5. Optional equipment includes a PCB vise and an inspection microscope 7-40X

Procedure

The following procedures cover the replacement of a TQFP 48 pin device with 0.5 mm lead pitch. The lead shape is the standard gullwing associated with JEDEC standard QFPs. This procedure section is divided into three parts:

A. Part Removal

B. Board Cleanup

C. Soldering a new device.

If you are soldering parts to a new printed circuit board, skip part A and refer to the new board cleanup section in part B.

A. Part Removal

Preparation:

- Board with IC to be removed is mounted in a holder or vise. A PCB holder/vise is optional but it is required that the PCB is held steady for the part removal.
- The soldering station is warmed up to 800°F (422°C) and the solder tip is clean.
- ESD precautions have been taken.

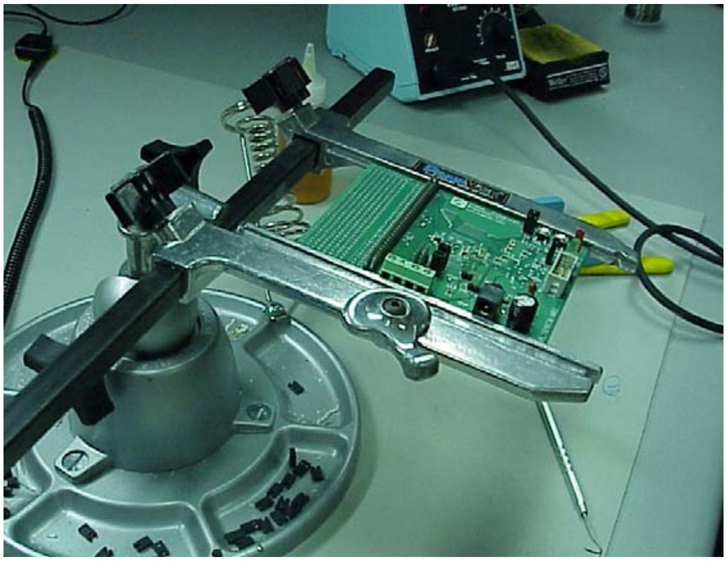

Figure 6. Ready to get started

Begin by wetting all the leads with flux to enhance the initial solder wicking cleanup. Wick up solder as much as possible from the QFP leads. Be careful not to scorch the PCB board with prolonged solder heat.

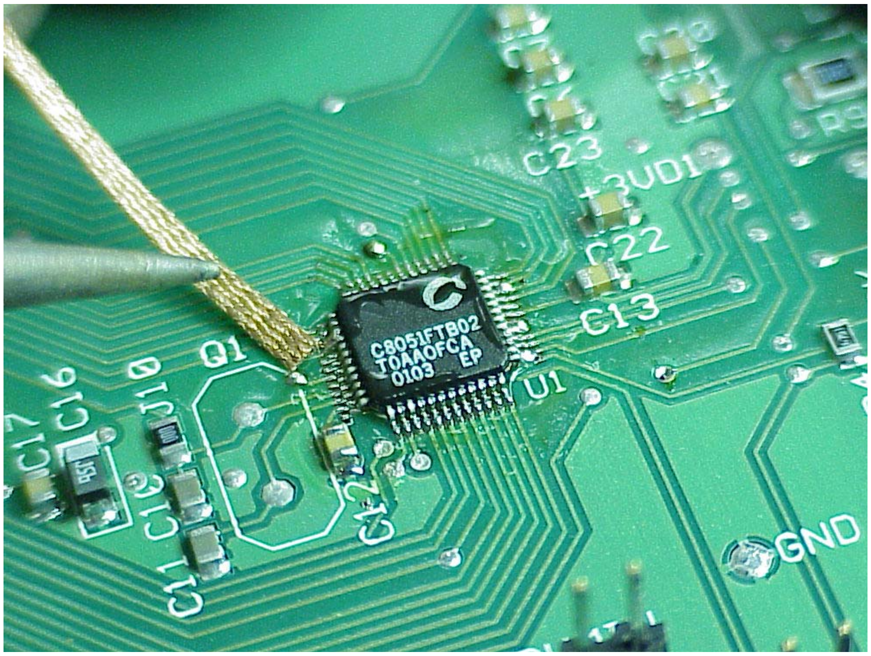
C8051FTB02
T0AAOFCA
0103 EP
Q1
U1
C13
C22
C23
C17
C16
J10
C11
C12
GND

Figure 7. Apply flux; wick excess solder from pins

Next, strip off approximately 3 inches of insulation from a piece of the 30 gauge wire wrap wire. Cut the wire at a comfortable 12 inch length or so.

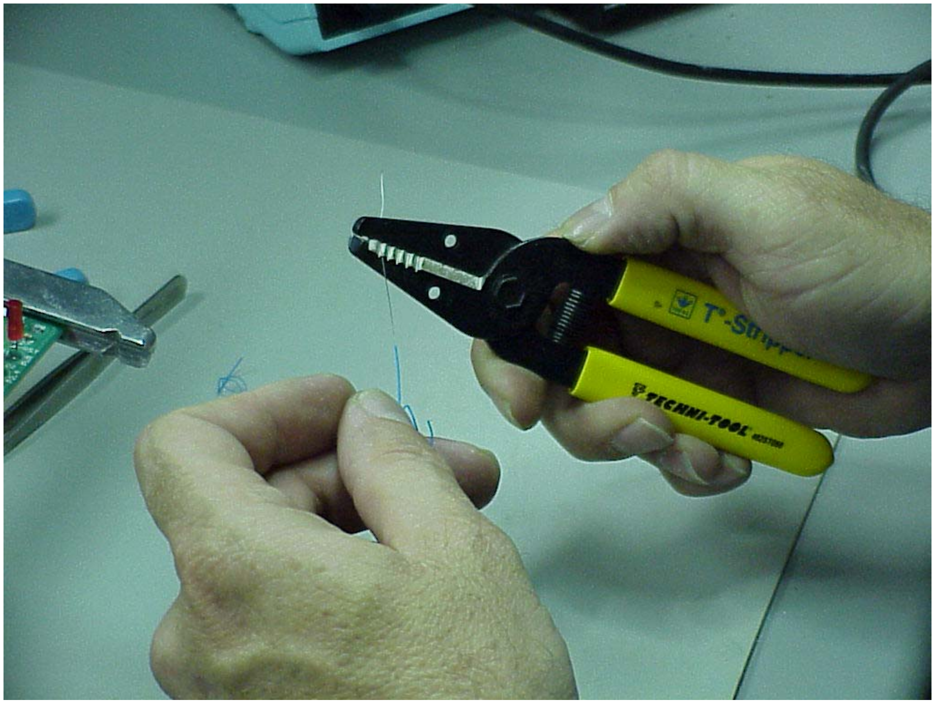
T'-Stripper
TECHNI-TOOL

Figure 8. Wire stripping

Feed the wire behind and under the leads on one side of the IC as shown in Figure 9.

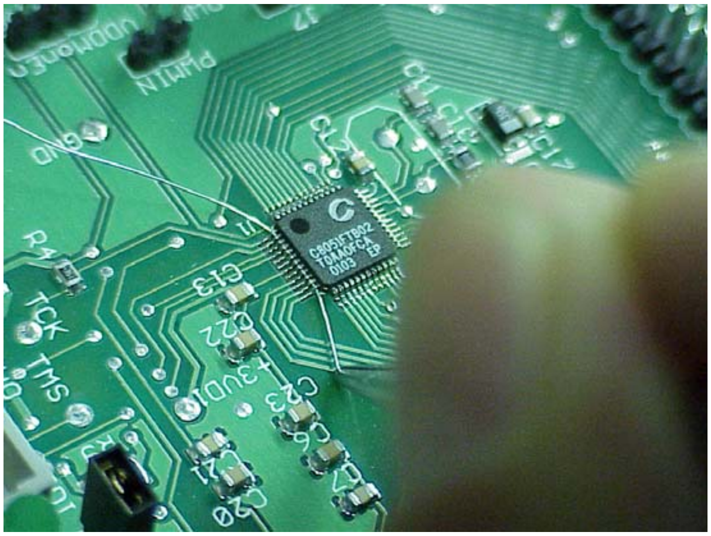

Figure 9. Feeding wire under QFP leads

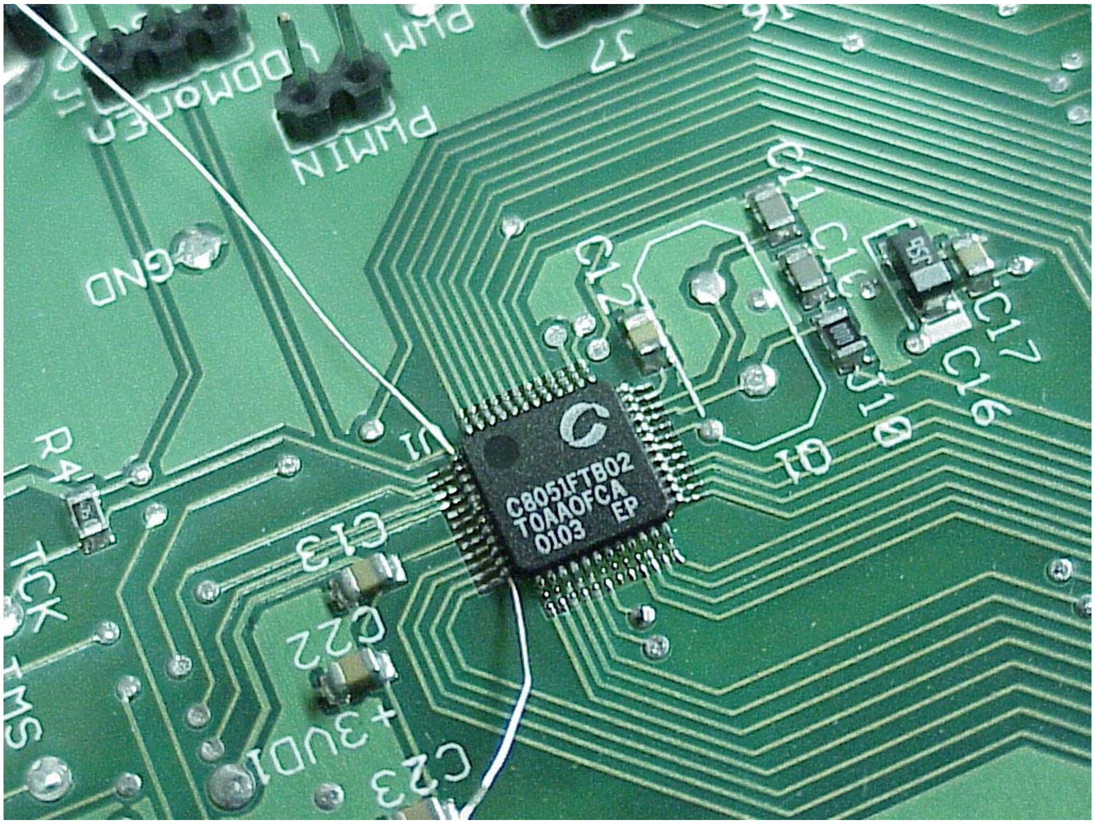

Figure 9. Wire with one side anchored to nearby component

Solder tack (anchor) one end of the 3 inch wire to a nearby via or component on the PCB. The anchor point should be in a location similar to that should in Figure 10.

Dispense a small amount of liquid flux across the leads.

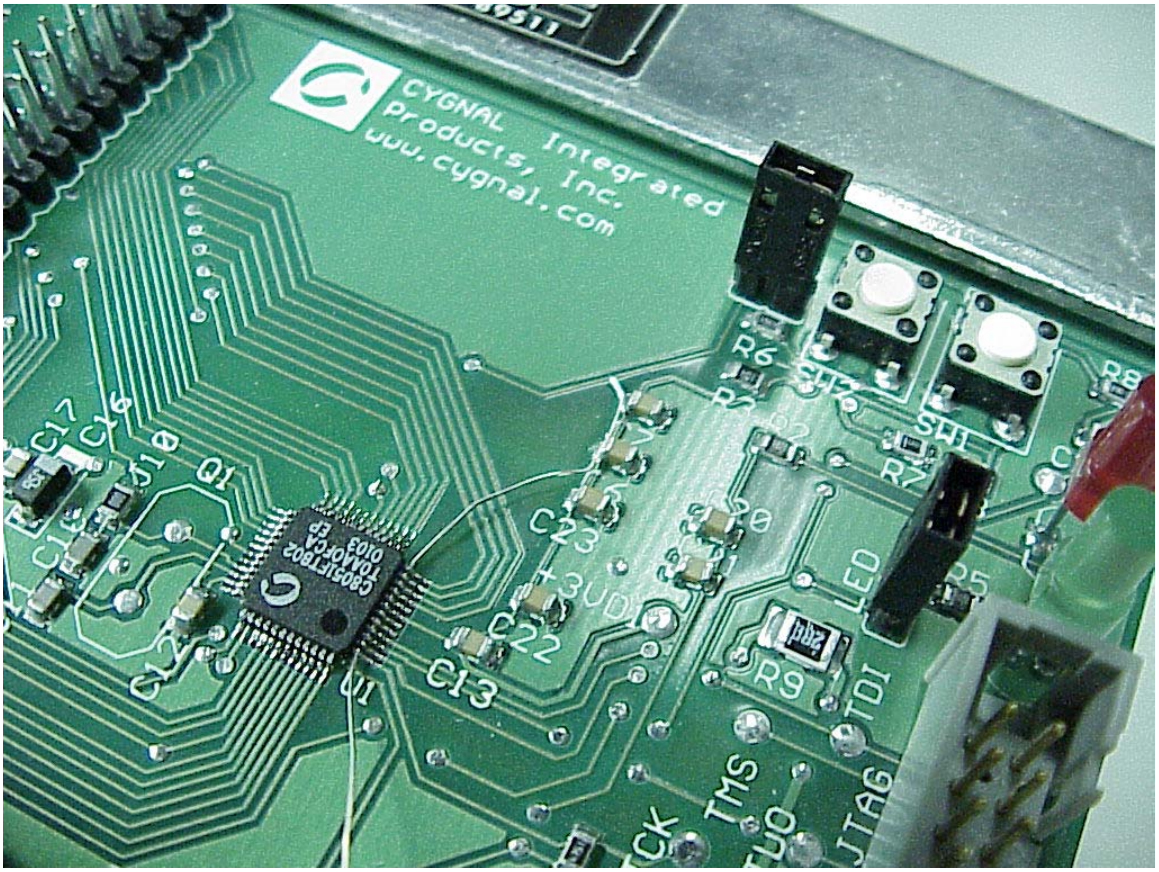
CYGNAL Integrated
Products, Inc.
www.cygnal.com
C8051F302
C17
CX6
J10
Q1
U1
C23
C22
C13
+3VD
R6
SW2
SW1
R7
R9
LED
TDI
TMS
JTAG
R8

Figure 10. Bus wire is anchored on C6

Hold the loose (non-anchored) end of the wire with tweezers in close proximity to the device as shown in Figure 11.

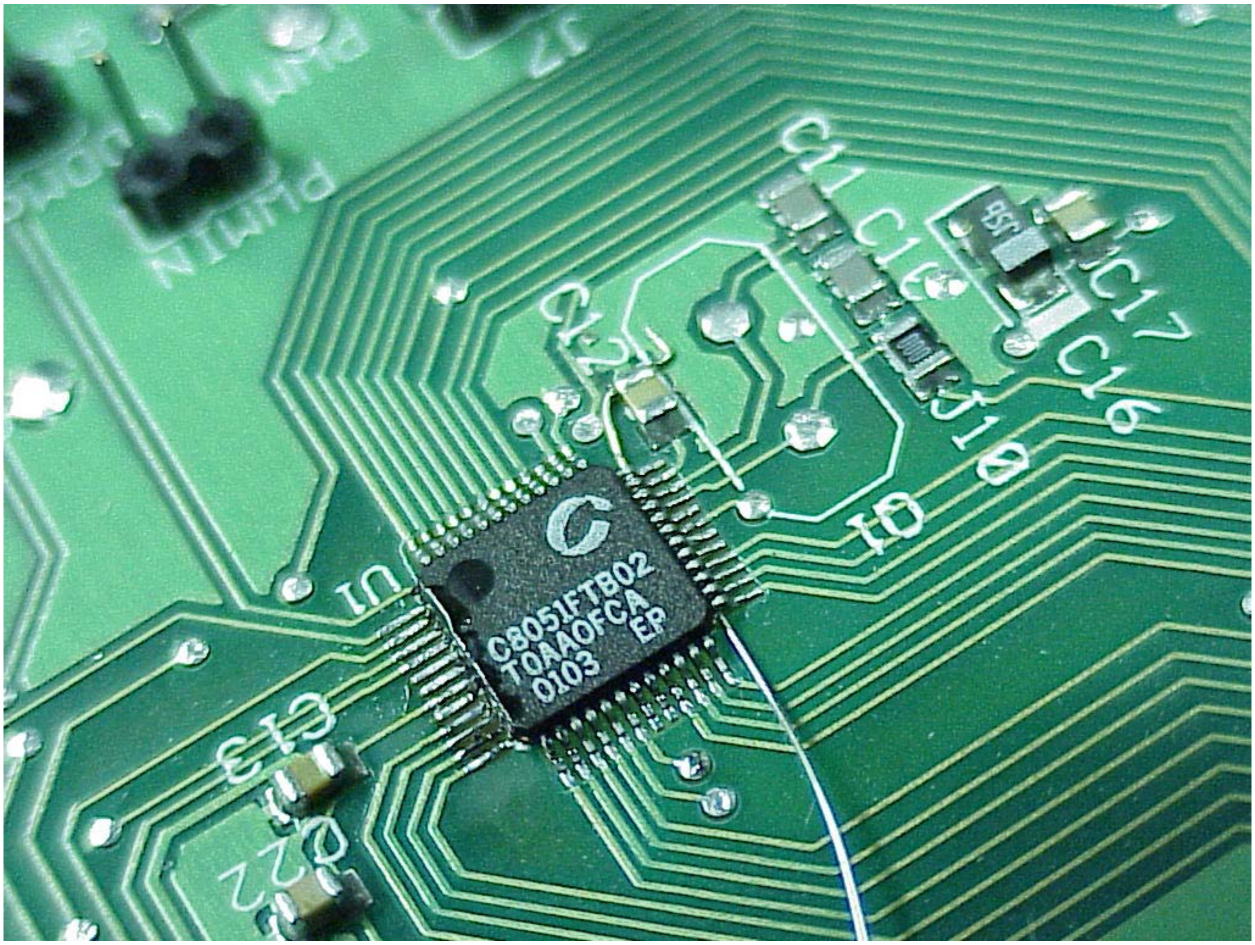

Figure 11. Side 2 anchored and ready for solder heat

You will now need to simultaneously apply solder heat and pull the wire away from the QFP, pulling at a slight upward angle from the board surface. Apply solder heat beginning at the lead closest to your tweezers. As the solder melts, gently pull the wire away from the QFP while continuing to move the solder heat from pin to pin to the right. You should not pull very hard. Pull as the solder melts. Do not leave the solder heat on any lead for more than necessary. The first lead will take the longest to heat, and after the wire gets hot, solder on the other leads will melt quickly. Excessive heat will damage the IC device and the PCB pad. The removal of 12 pins from a 48TQFP should take about 5 seconds total. Signs of excessive heat are:

- Melted plastic on the IC device
- PCB pads that lift off
- Brown scorch marks on the PCB

With one side of the QFP completed, repeat the same procedure on the other three sides of the QFP. Cut off the dirty part of the wire wrap wire or use a new piece for each side. Dispense flux again for each side.

Note that in the following pictures, the old IC device is not being saved. There is slightly more heat being applied here than necessary to speed up the process. This result is some melted plastic and missing gull-wing leads. These are visible in the pictures that follow. If you are trying to save the IC being removed, then you must be very careful to apply as little heat as possible during the removal process such that the QFP leads remain intact in the plastic QFP body. This will require some experimentation with solder heat settings and timing.

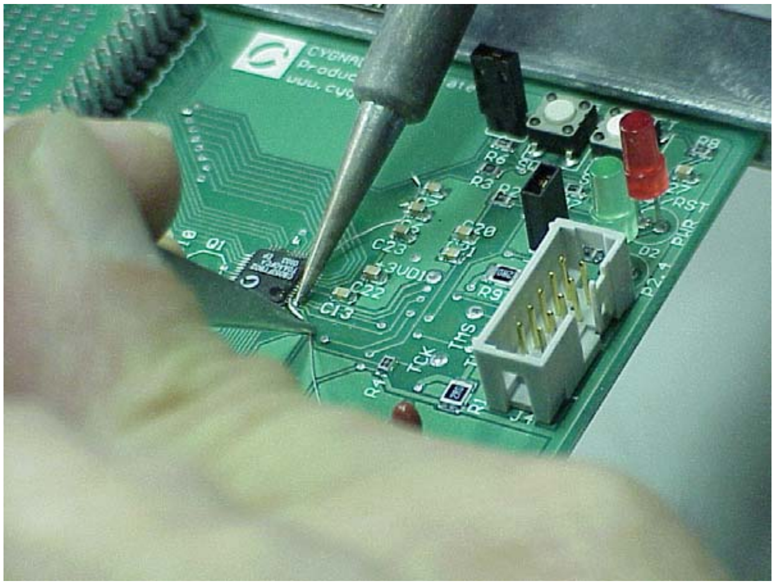

Figure 12. Hold tweezers close to device

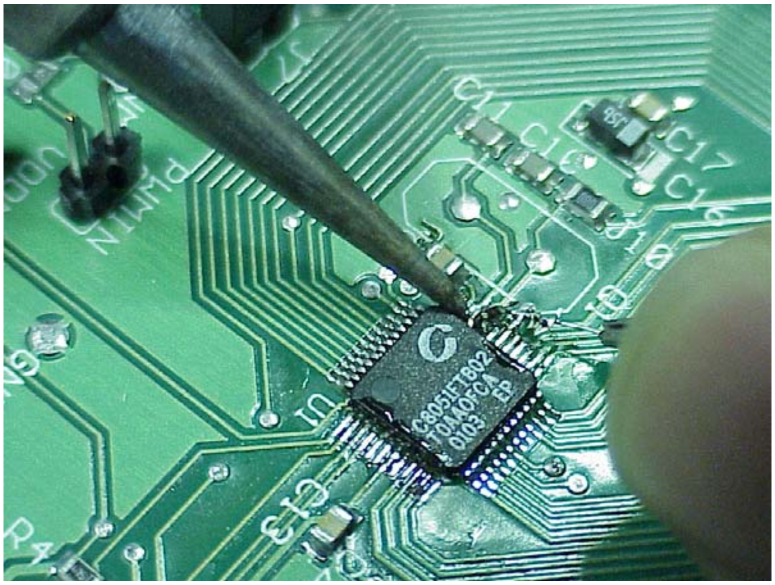

Figure 14. Side 2 almost complete

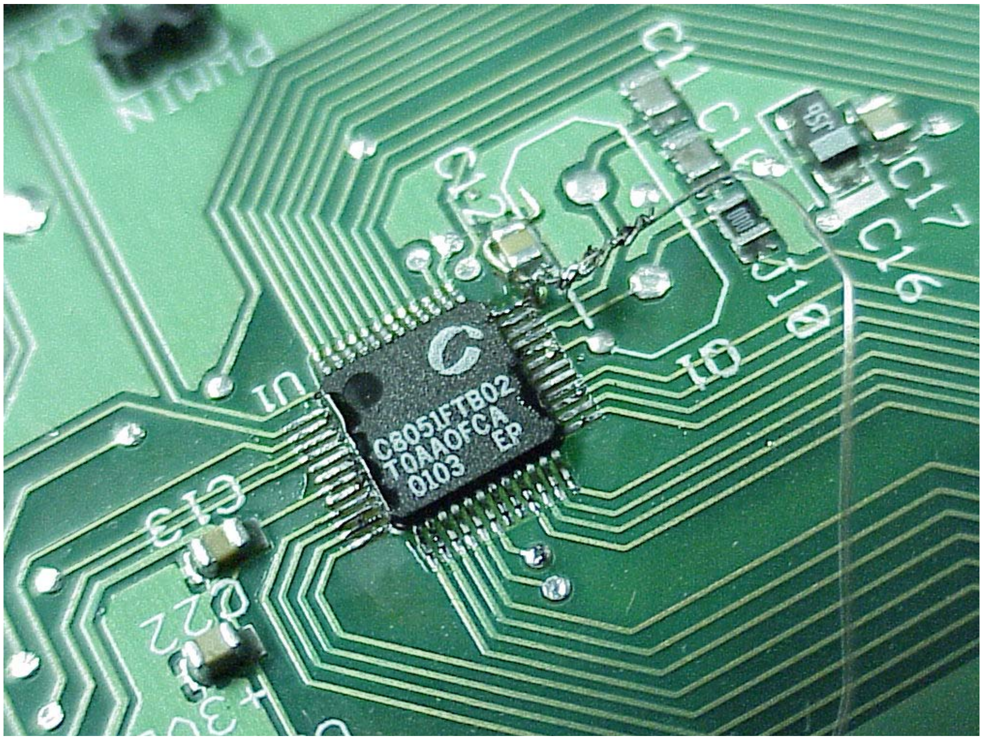

Figure 15. Side 2 is done

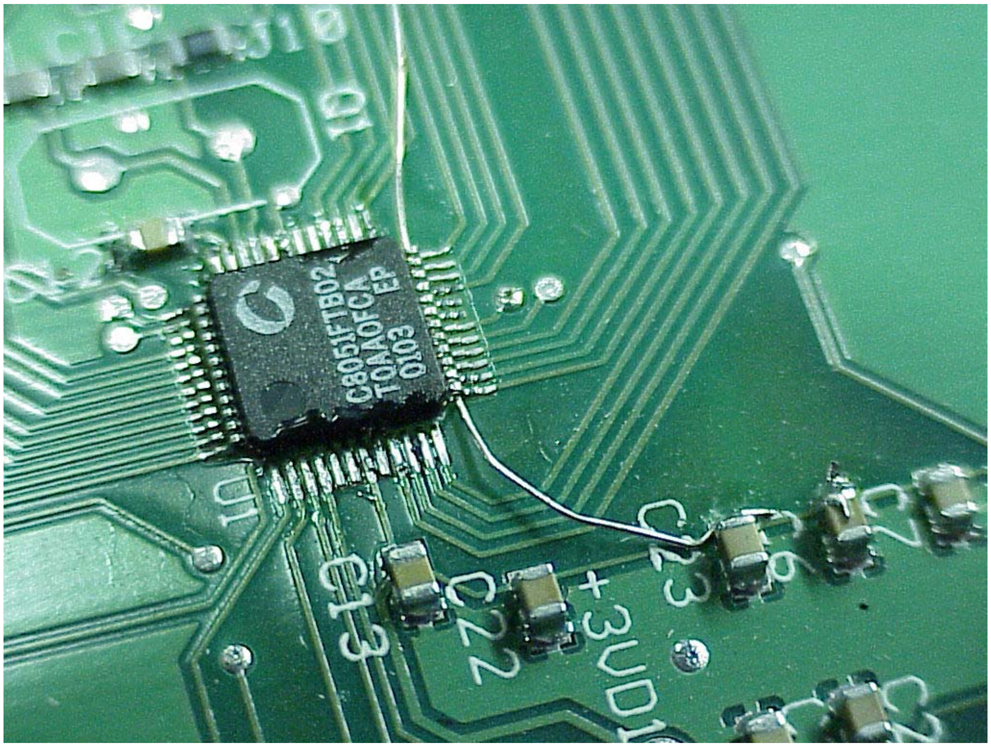

Figure 16. Side 3 anchored and ready

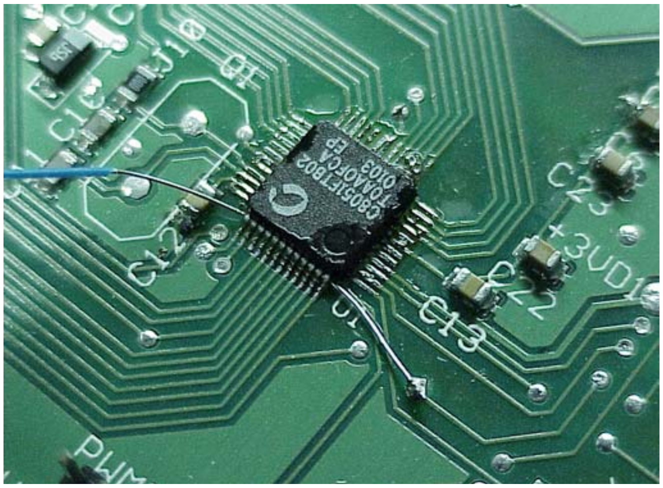

Figure 17. Side 4 anchored to a via

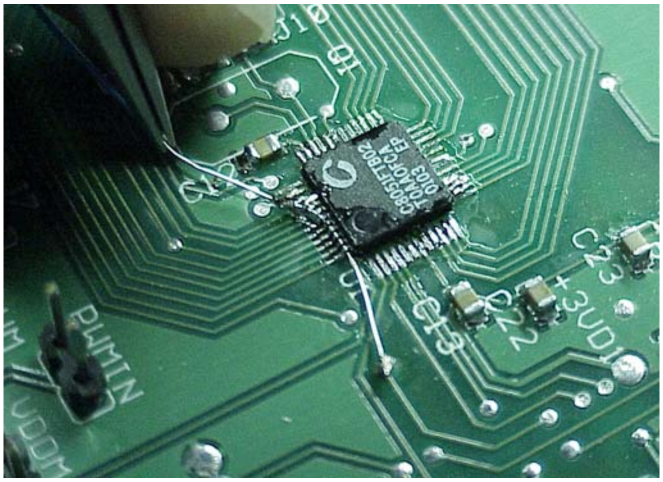

Figure 18. Side 4 removal started

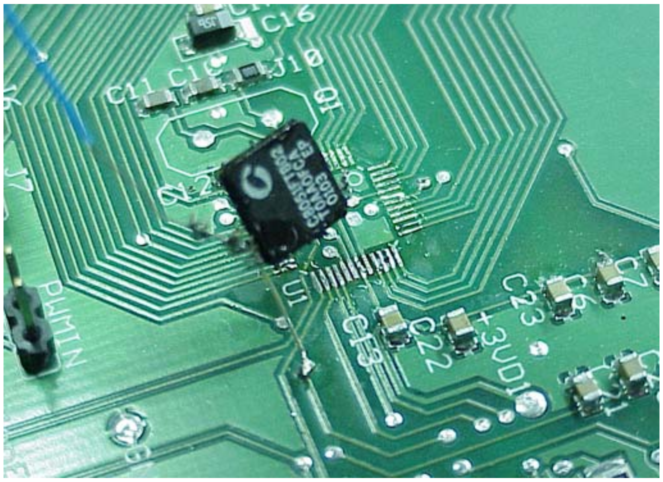

Figure 19. QFP removal a second before completion

B. Board Cleanup

New PCBs

For mounting a device to a new PCB, the amount of cleanup should be minimal. On a new PCB, there should be no solder on the pads. Brushing the pads with isopropyl alcohol (Figure 33) and drying the board should be enough preparation to begin the mounting procedure.

Reworked PCBs

The following section is the cleanup sequence that follows the QFP removal in the previous section. After removing the device, the solder pads will need cleanup. The idea is to clean the pads so that they are flat and free or solder and flux. Solder wick the pads until they are flat and dull. A clean pad appearance should be a dull silver color.

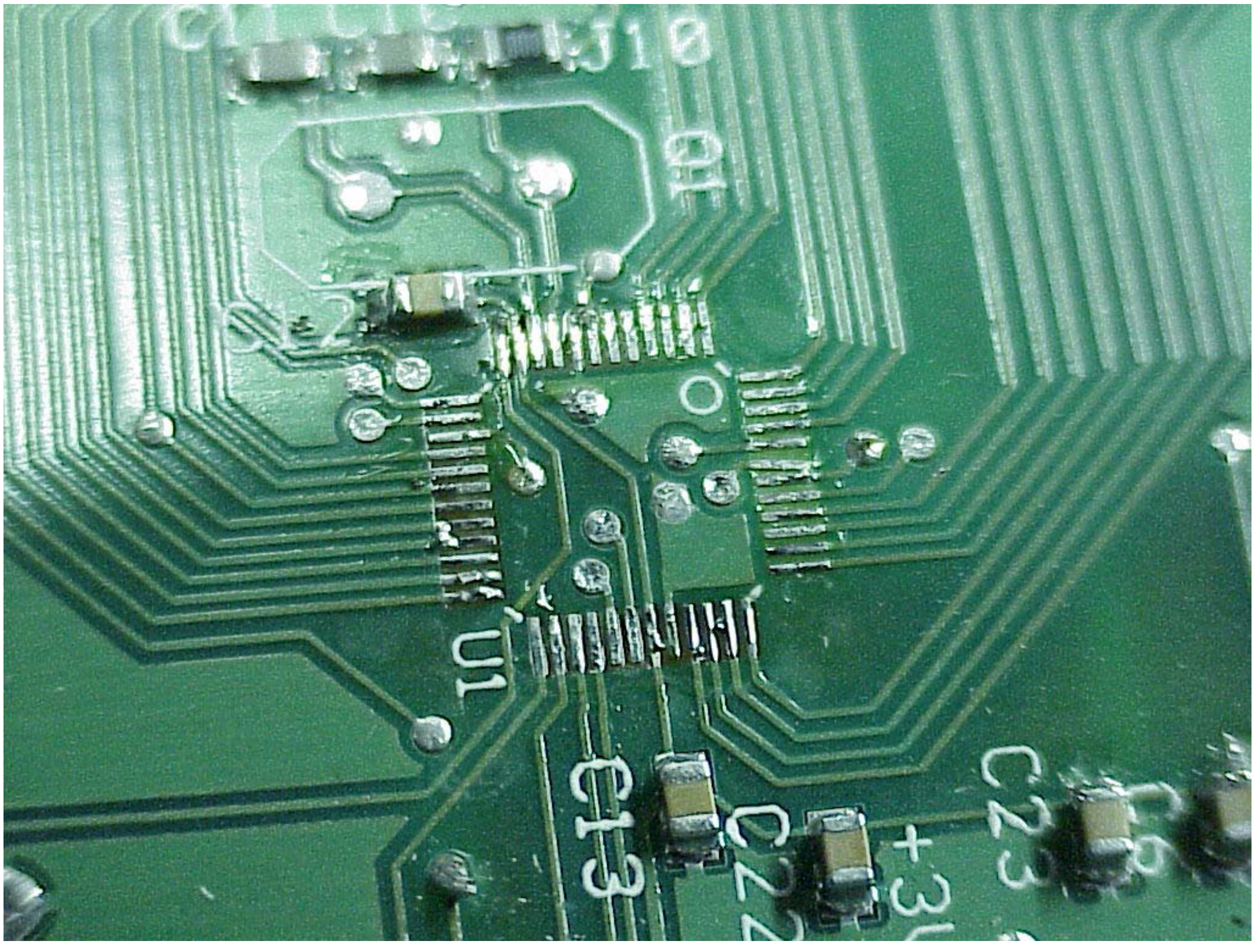

Figure 20. Pads after QFP removal procedure

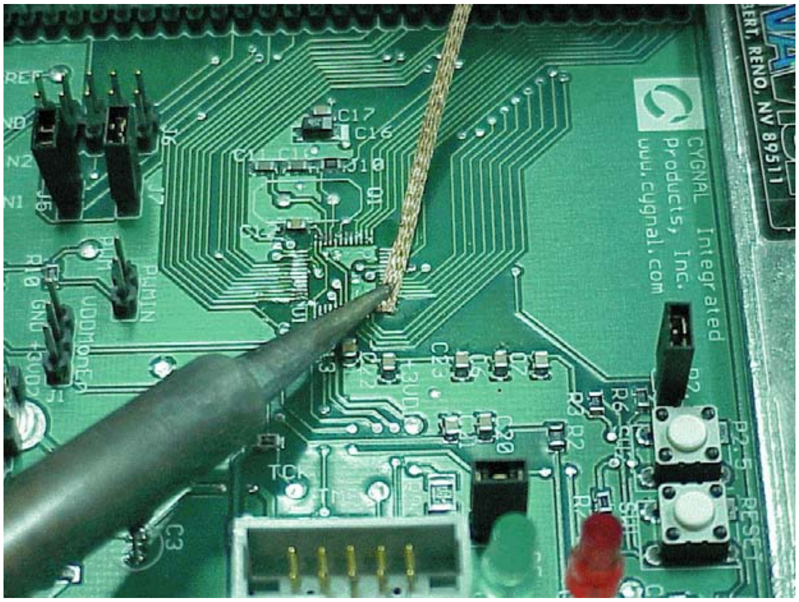

Figure 21. Solder wick to remove solder on the pads

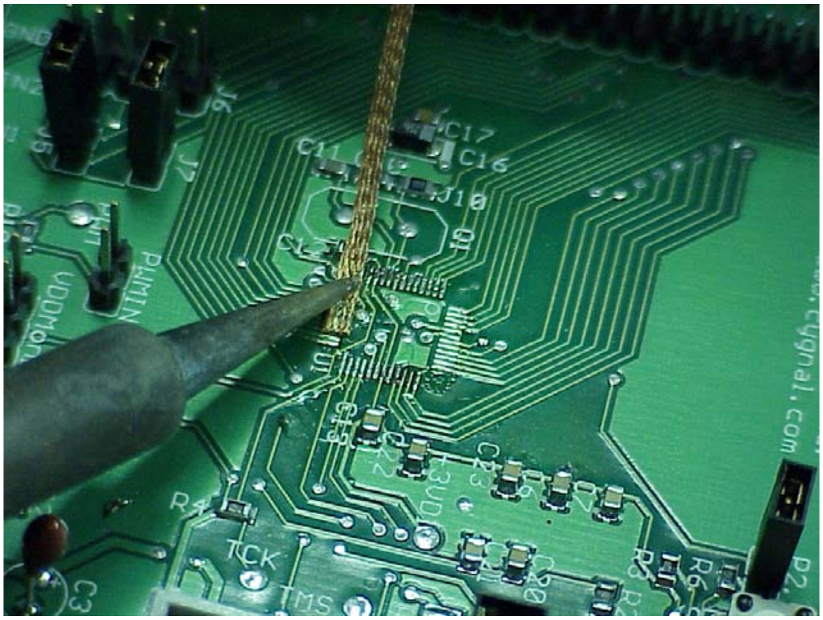

Figure 22. Repeat for all pads

If any pads become loose from the PCB, use a dental pick or other pointed object to re-align the pad (Figure 23 and Figure 24).

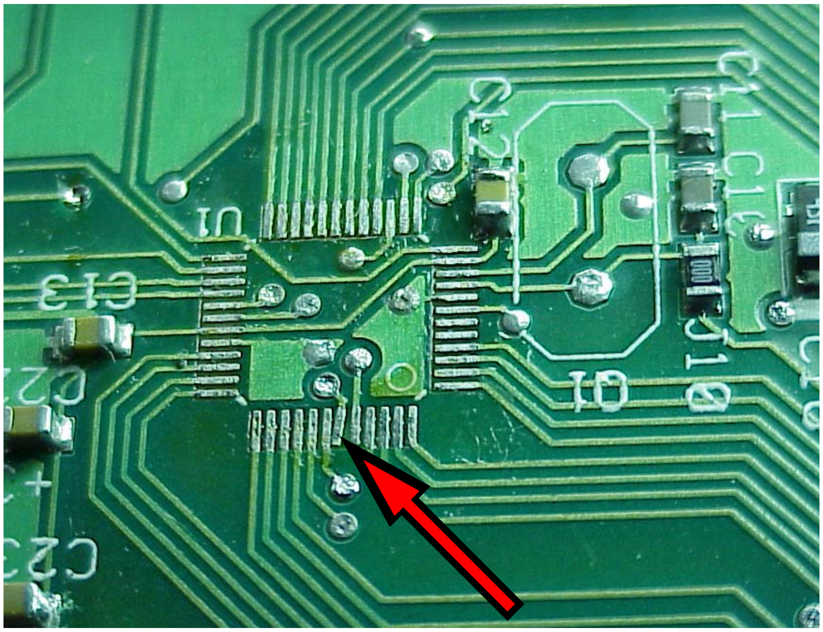
C12
C11
C10
U1
C13
Q1
J10
C22
C23

Figure 23. Clean pads but one pad is slightly crooked

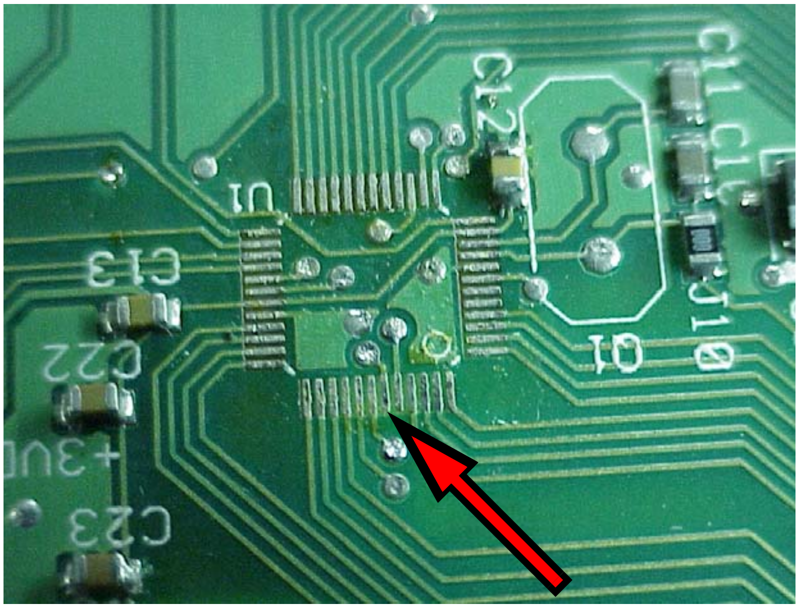
C11
C12
C10
U1
C13
C22
+3V
C23
Q1
J10

Figure 24. Pad straightened out

B. Soldering a new QFP

The pads on the PCB should be clean and free of any solder.

Carefully place the new QFP device on the PCB using tweezers or another safe method. Make sure the part is not dropped as the leads can be easily damaged.

Align the part over the pads using a small pick or similar tool to push the part. Get the alignment as accurate as you can. Also, make sure that the part is oriented correctly (pin 1 orientation).

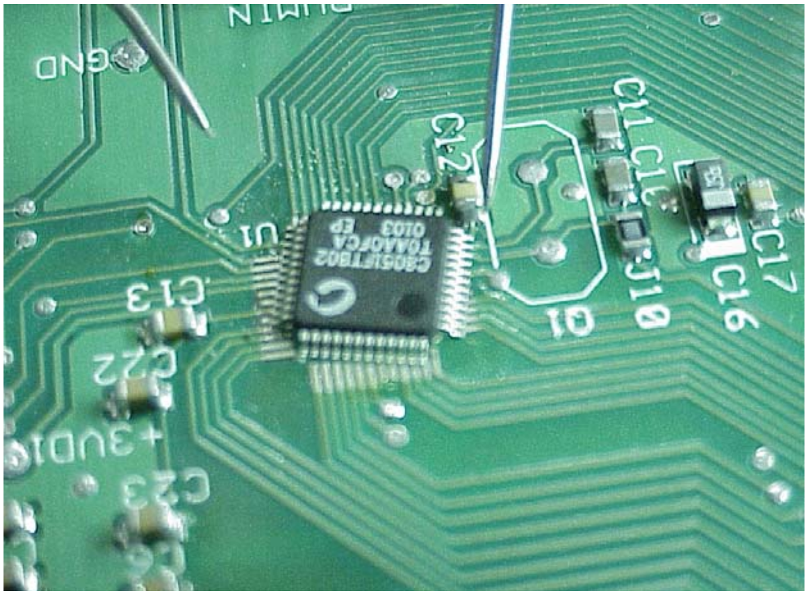

Figure 25. New QFP near pads, ready for alignment

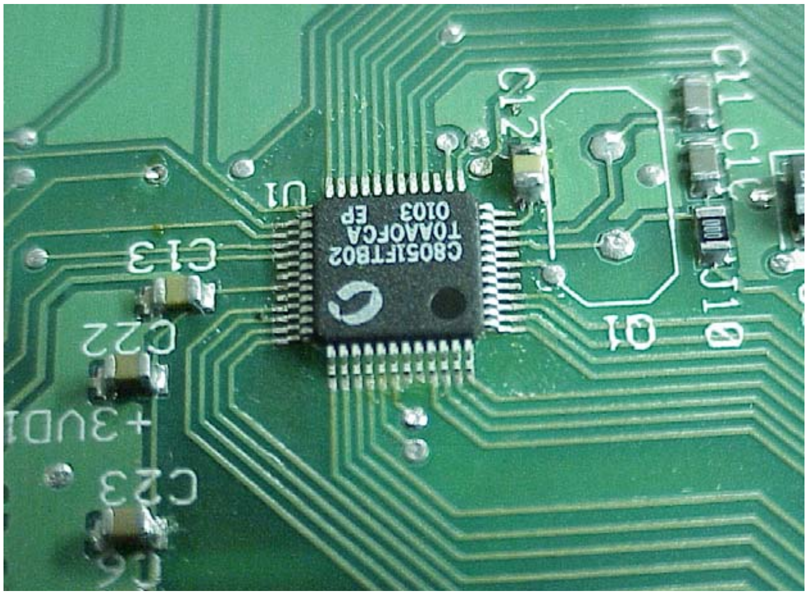

Figure 26. QFP aligned

Adjust the soldering station temperature to 725°F (385°C). Put a small amount of solder on the tip of the solder iron. While holding down the aligned QFP with a pick or other pointed tool, add a small amount of solder flux to the corner leads in two opposite corners. While still holding the part down with the pick, solder down two opposite corner leads on the QFP. Do not worry about excess solder or shorts between adjacent leads at this time. The idea is to anchor down the aligned QFP with solder so it does not move.

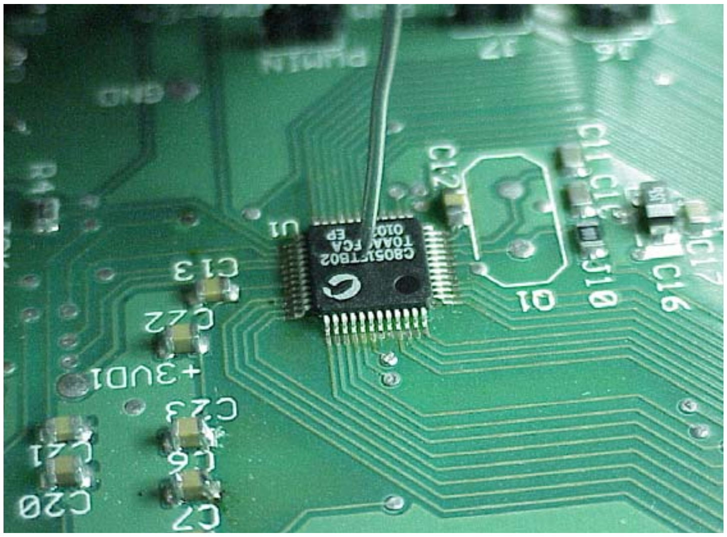

Figure 27. Aligned QFP ready for solder anchoring

Re-check the QFP alignment after soldering the corners. If necessary, make adjustments or remove and start over to get good QFP to PCB alignment.

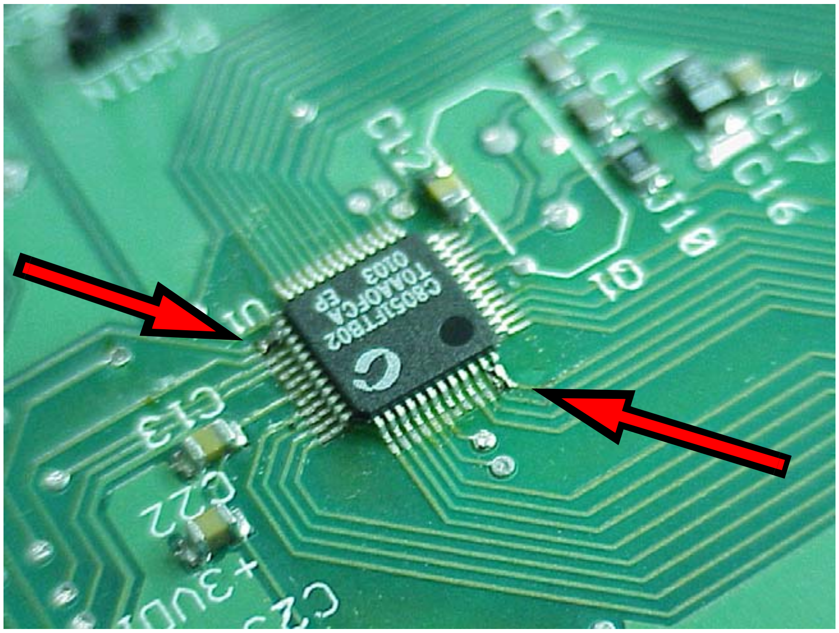

Figure 28. Aligned QFP with corners soldered down

Now you are ready to solder all the leads. Add solder to the tip of the soldering iron. Dispense flux over all the leads to keep them wet.

Touch the solder iron tip to the end of each QFP lead until the solder is seen running up the lead. Repeat for all the leads. Add small amounts of solder as needed to the soldering iron tip. Again, do not worry if you see some solder bridging as you will clean that up in the next step.

When soldering, keep the soldering iron tip parallel with the pins being soldered to prevent excessive solder shorts.

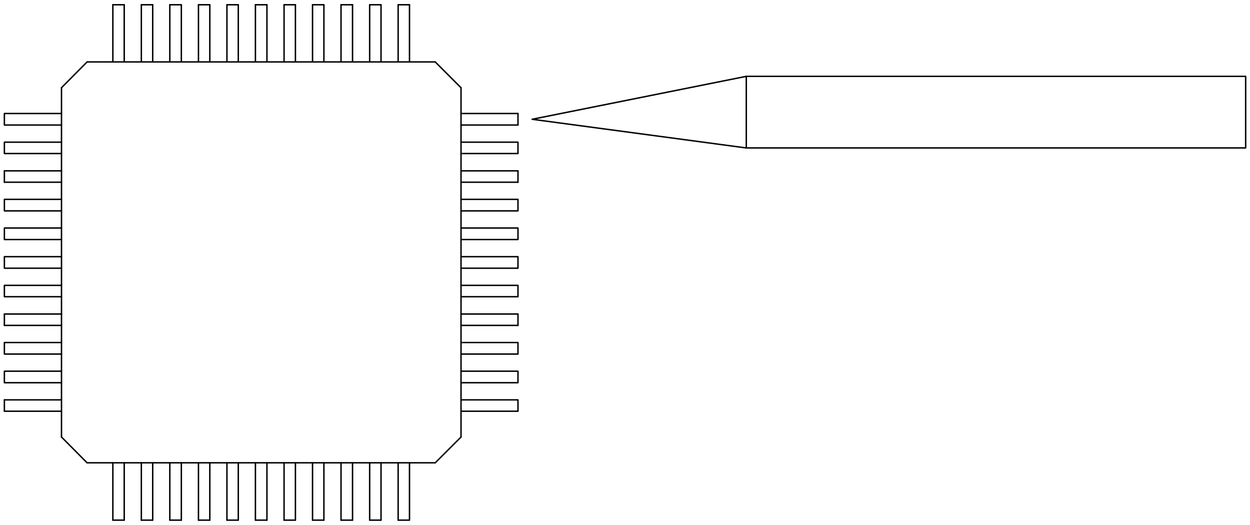

Figure 29. Keep iron tip parallel to pins being soldered

After soldering all the leads, wet all the leads with flux to enhance the solder wicking cleanup. Wick up solder where needed to eliminate any shorts/bridging.

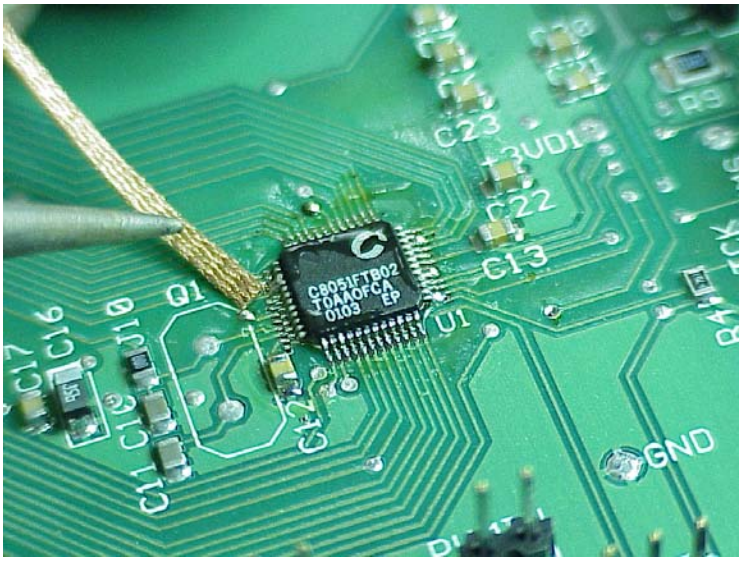
C23
+3VD1
C22
C13
Q1
U1
C17
C16
J10
C11
C12
GND
R9
C8051FTB02
TOAAOFCA
0103 EP

Figure 30. Wicking Solder #1

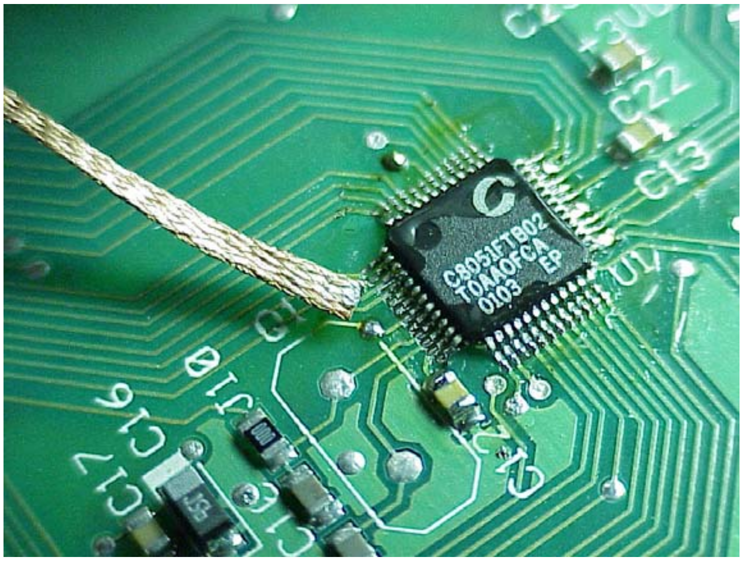
C22
C13
U1
Q1
J10
C16
C17
C8051FTB02
TOAAOFCA
0103 EP

Figure 31. Wicking Solder #2

Hand inspect the board using 4X magnification (or higher) for shorts or marginal solder joints. Solder joints should have a smooth melt transition between each device pin and the PCB. Rework any pins as needed

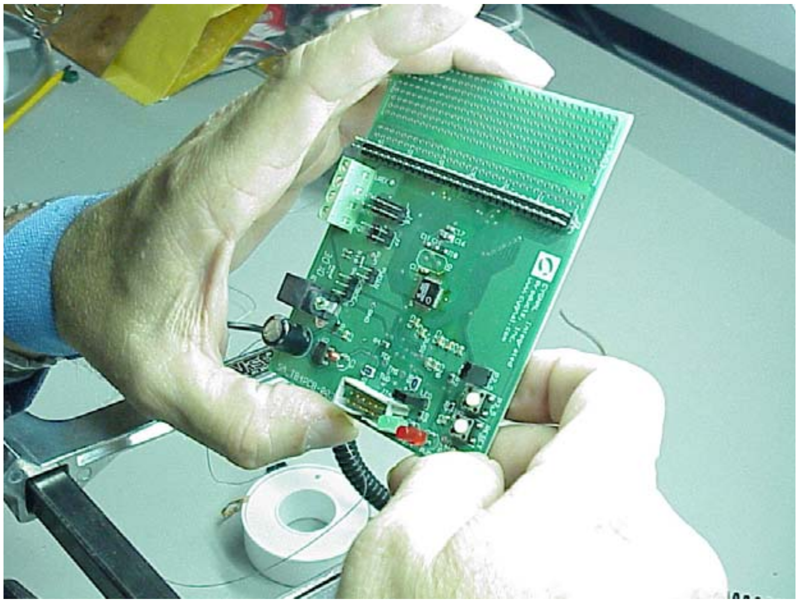

Figure 32. Visual Inspection

After the inspection passes, it is time to clean the flux off the board. Dip the stiff bristle brush into alcohol and wipe in the direction of the leads. Use moderate, but not excessive pressure. Use liberal amounts of alcohol and brush well between the QFP leads until the flux disappears.

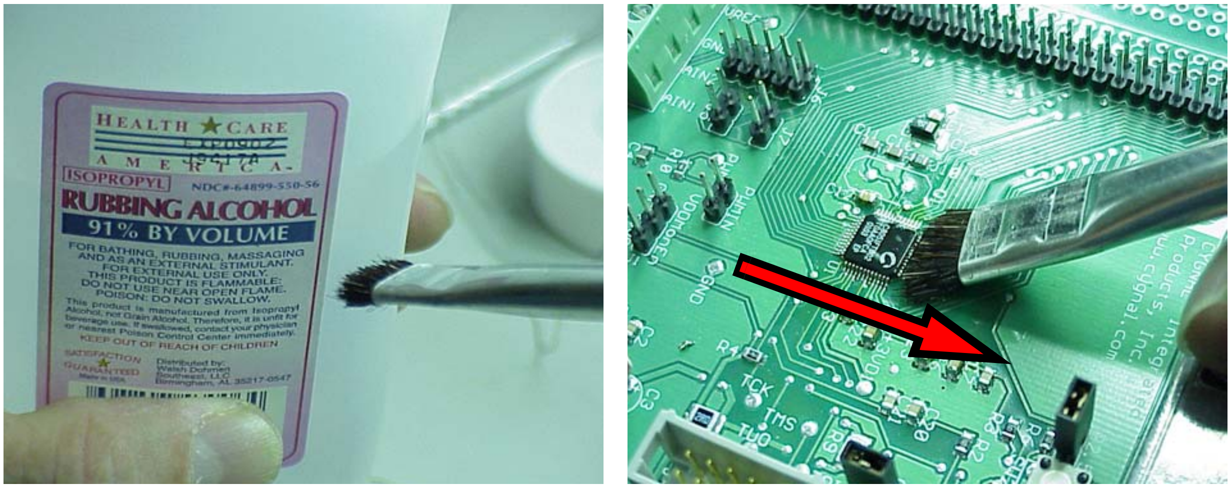

Figure 33. Isopropyl Alcohol and stiff brush used for cleanup. Brush only in direction of leads

Dry the board with compressed dry air or nitrogen. If this is not available, let the board dry for 30 minutes or more to let the alcohol evaporate under the QFP. The QFP leads should look bright and there should be no flux residue.

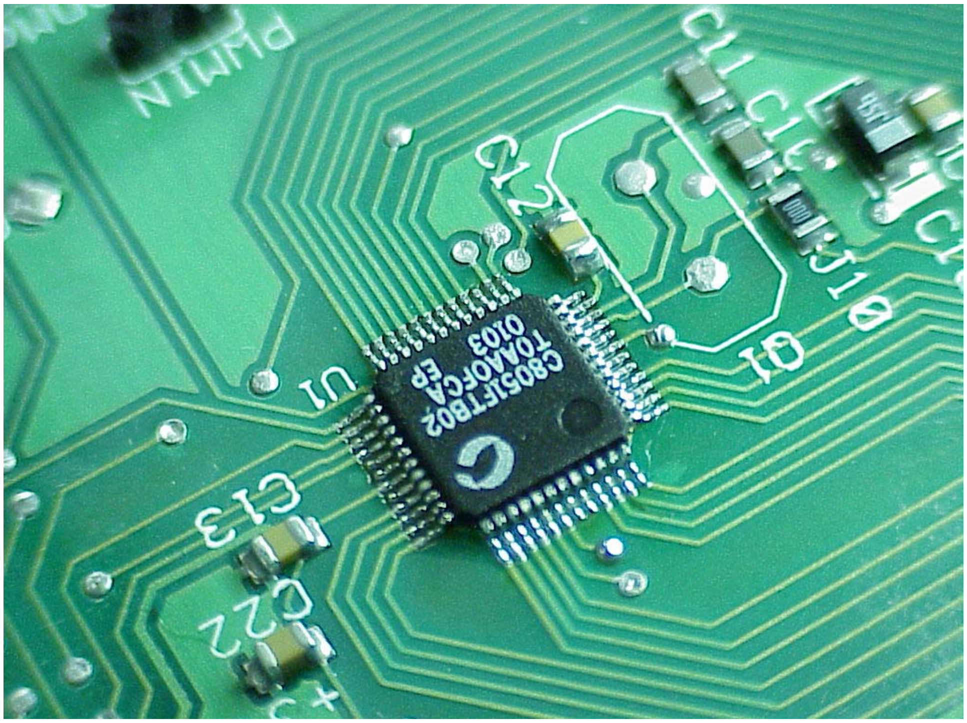
PWMIN
C11
C12
C1
J10
Q1
U1
C8051F0B02
TQNAOFCA
EP
0103
C13
C22

Figure 34. Clean and Pristine

Re-inspect the board for workmanship. Rework any leads if needed.

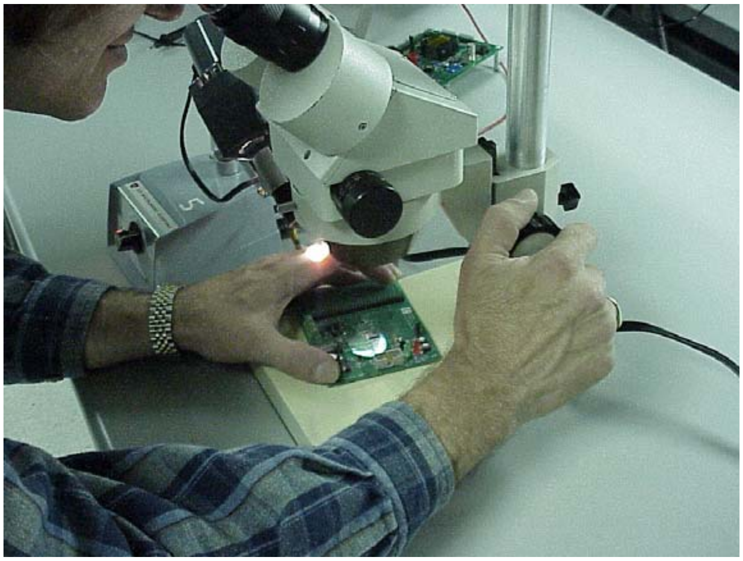

Figure 35. Stereo zoom inspection station (7X to 40X magnification) helps to inspect solder joint workmanship

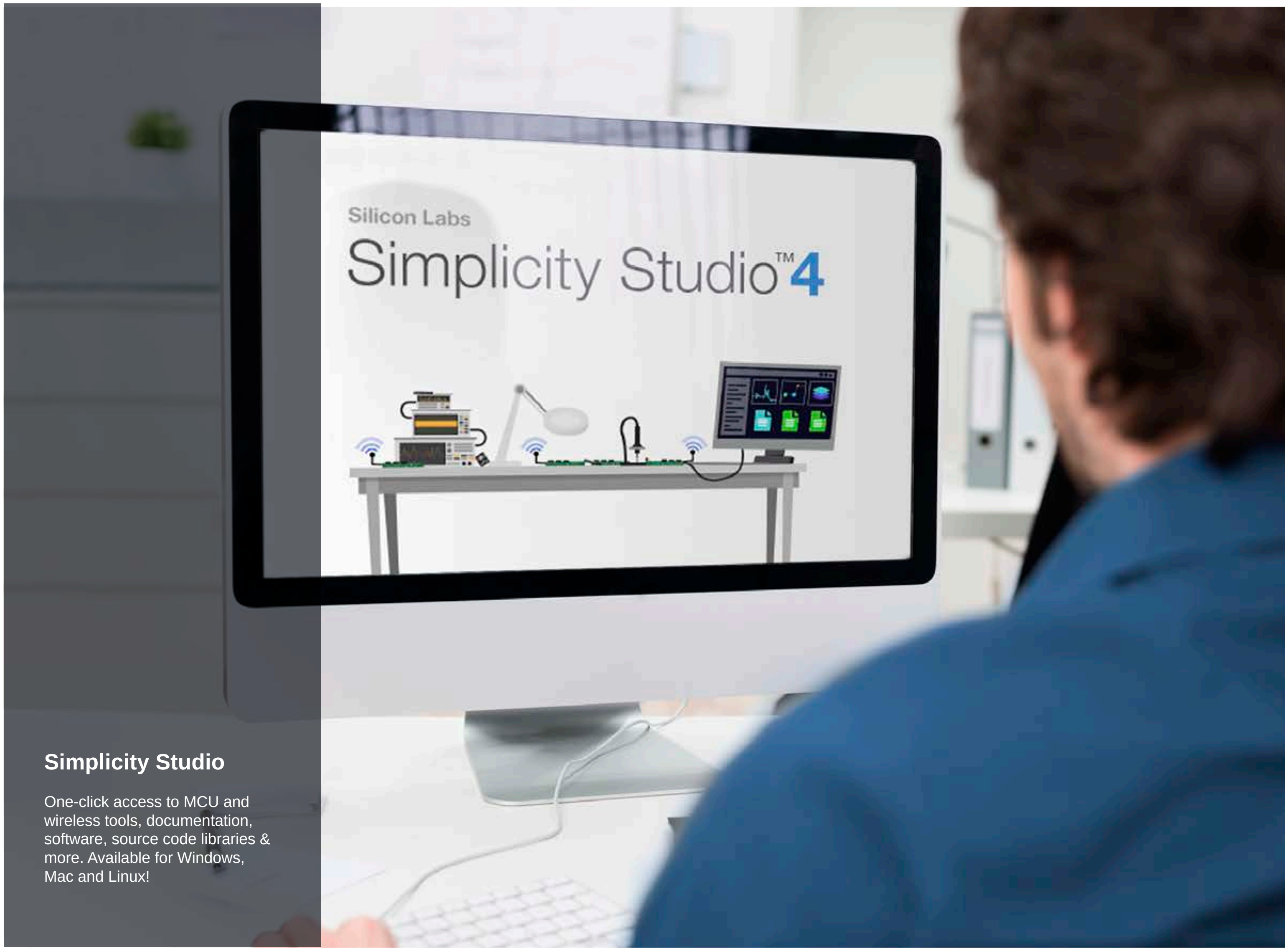

Simplicity Studio

One-click access to MCU and wireless tools, documentation, software, source code libraries & more. Available for Windows, Mac and Linux!

IoT Portfolio
www.silabs.com/IoT

SW/HW
www.silabs.com/simplicity

Quality
www.silabs.com/quality

Support and Community
community.silabs.com

Disclaimer

Silicon Labs intends to provide customers with the latest, accurate, and in-depth documentation of all peripherals and modules available for system and software implementers using or intending to use the Silicon Labs products. Characterization data, available modules and peripherals, memory sizes and memory addresses refer to each specific device, and "Typical" parameters provided can and do vary in different applications. Application examples described herein are for illustrative purposes only. Silicon Labs reserves the right to make changes without further notice and limitation to product information, specifications, and descriptions herein, and does not give warranties as to the accuracy or completeness of the included information. Silicon Labs shall have no liability for the consequences of use of the information supplied herein. This document does not imply or express copyright licenses granted hereunder to design or fabricate any integrated circuits. The products are not designed or authorized to be used within any Life Support System without the specific written consent of Silicon Labs. A "Life Support System" is any product or system intended to support or sustain life and/or health, which, if it fails, can be reasonably expected to result in significant personal injury or death. Silicon Labs products are not designed or authorized for military applications. Silicon Labs products shall under no circumstances be used in weapons of mass destruction including (but not limited to) nuclear, biological or chemical weapons, or missiles capable of delivering such weapons.

Trademark Information

Silicon Laboratories Inc.® , Silicon Laboratories®, Silicon Labs®, SiLabs® and the Silicon Labs logo®, Bluegiga®, Bluegiga Logo®, Clockbuilder®, CMEMS®, DSPLL®, EFM®, EFM32®, EFR, Ember®, Energy Micro, Energy Micro logo and combinations thereof, "the world’s most energy friendly microcontrollers", Ember®, EZLink®, EZRadio®, EZRadioPRO®, Gecko®, ISOmodem®, Precision32®, ProSLIC®, Simplicity Studio®, SiPHY®, Telegesis, the Telegesis Logo®, USBXpress® and others are trademarks or registered trademarks of Silicon Labs. ARM, CORTEX, Cortex-M3 and THUMB are trademarks or registered trademarks of ARM Holdings. Keil is a registered trademark of ARM Limited. All other products or brand names mentioned herein are trademarks of their respective holders.

SILICON LABS

Silicon Laboratories Inc.
400 West Cesar Chavez
Austin, TX 78701
USA

http://www.silabs.com

STAN GIBILISCO

THE ILLUSTRATED DICTIONARY OF Electronics

EIGHTH EDITION

Audio/Video
Consumer Electronics
Wireless Technology

The Illustrated Dictionary of Electronics

This page intentionally left blank.

The Illustrated Dictionary of Electronics

Eighth Edition

Stan Gibilisco
Editor-in-Chief

McGraw-Hill
New York Chicago San Francisco Lisbon London Madrid
Mexico City Milan New Delhi San Juan Seoul
Singapore Sydney Toronto

Cataloging-in-Publication Data is on file with the Library of Congress

McGraw-Hill
A Division of The ***McGraw·Hill*** *Companies*

Copyright © 2001 by The McGraw-Hill Companies, Inc. All rights reserved. Printed in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a data base or retrieval system, without the prior written permission of the publisher.

1 2 3 4 5 6 7 8 9 0 AGM/AGM 0 9 8 7 6 5 4 3 2 1

P/N 0-07-137237-7

ISBN 0-07-137236-9

The sponsoring editor for this book was Scott Grillo, and the production supervisor was Pamela Pelton. It was set in Bookman by Techbooks.

Printed and bound by Quebecor/Martinsburg.

McGraw-Hill books are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. For more information, please write to the Director of Special Sales, McGraw-Hill, Two Penn Plaza, New York, NY 10121-2298. Or contact your local bookstore

Information in this book has been obtained by the publisher from sources believed to be reliable. However, neither the publisher nor the authors guarantee the accuracy or completeness of any information published herein. Neither the publisher nor the authors shall be responsible for any errors, omissions, or damages arising out of use of this information. This work is published with the understanding that the publisher and authors are supplying information but are not attempting to render professional services in any way, shape or form. If such services are required, the assistance of an appropriate professional should be sought.

This book is printed on recycled, acid-free paper containing a minimum of 50 percent recycled de-inked fiber.

To Tony, Tim, and Samuel
from Uncle Stan

This page intentionally left blank.

Contents

Preface *ix*
Acknowledgments *xi*

Dictionary 1
Appendix A Schematic Symbols 773
Appendix B Tables and Data 787

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

This page intentionally left blank.

Preface

The Illustrated Dictionary of Electronics—8th Edition has been revised, clarified, and updated, reflecting technological advances of recent years. New definitions have been added in the fields of wireless technology, robotics, and artificial intelligence. Every effort has been made to be concise and accurate, without "talking down" to the reader.

Many definitions contain cross references (indicated in ALL CAPITALS); these provide recommended additional information or allow comparison with related terms. Expressions of special significance are printed in italics. Electronics abbreviations are included in the text; the full terms are stated as definitions.

While an effort has been made to avoid superfluous mathematics, equations are sometimes necessary to completely and effectively define a term. Mathematics beyond the high-school level has not been used.

Appendix A contains the standard symbols used in electrical and electronic diagrams. These symbols are used in illustrations throughout this dictionary. Appendix B contains the following data tables:

1. Conversion between electrical systems
2. Greek alphabet
3. Mathematical functions and operations
4. Prefix multipliers
5. Resistor color code

Suggestions for future editions are welcome.

Stan Gibilisco
Editor-in-Chief

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

This page intentionally left blank.

Acknowledgments

Illustrations in this book were generated with *CorelDRAW*. Some clip art is courtesy of Corel Corporation, 1600 Carling Avenue, Ottawa, Ontario, Canada K1Z 8R7.

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

This page intentionally left blank.

The
Illustrated Dictionary of Electronics

This page intentionally left blank.

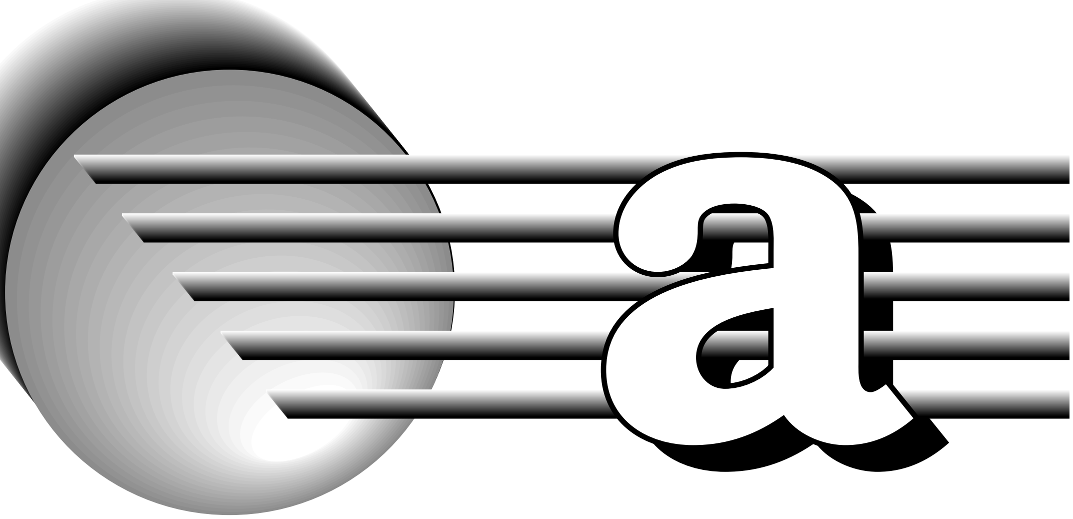

A **1.** Symbol for GAIN. **2.** Symbol for AREA. **3.** Symbol for AMPERE (SI unit for current).

A– Symbol for negative terminal of filament-voltage source in a vacuum-tube circuit.

A+ Symbol for positive terminal of filament-voltage source in a vacuum-tube circuit.

a **1.** Abbreviation of ATTO- (prefix). **2.** Abbreviation of AREA. **3.** Abbreviation of ACCELERATION. **4.** Abbreviation of ANODE. **5.** Obsolete abbreviation of cgs prefix AB-.

aA **1.** Abbreviation of *attoampere*. **2.** Obsolete for ABAMPERE.

AAAS Abbreviation for *American Association for the Advancement of Science.*

AAC Abbreviation of *automatic aperture control* (NASA).

AAS Abbreviation of *advanced antenna system* (NASA).

AASR Abbreviation of *airport and airways surveillance radar.*

AB Abbreviation of *acquisition beacon* (NASA).

A-B In sound and acoustics, the direct comparison of two sources of sound by alternately turning on one and the other.

ab- **1.** Prefix that transforms the name of a practical electrical unit to that of the equivalent electromagnetic cgs unit (e.g., ABAMPERE, ABOHM, ABVOLT). See individual entries of such cgs units. **2.** Abbreviation for ABSOLUTE.

abac A graphic device for the solution of electronics problems. Also see ALIGNMENT CHART.

abampere The unit of current in the cgs electromagnetic system. One abampere equals 10 amperes and corresponds to 1 abcoulomb per second.

Abbe condenser **1.** In microscopy, a special two-piece lens that has enhanced light-gathering power. **2.** A similar focusing device in an electromagnetic antenna.

abbreviated dialing In telephone systems, special circuits requiring fewer-than-normal dialing operations to connect subscribers.

abc **1.** Abbreviation of AUTOMATIC BASS COMPENSATION, a system for boosting the volume of bass sounds at low amplifier gain. **2.** Abbreviation of AUTOMATIC BIAS CONTROL. **3.** Abbreviation of AUTOMATIC BRIGHTNESS CONTROL. **4.** Abbreviation of AUTOMATIC BRIGHTNESS COMPENSATION.

abcoulomb The unit of electrical quantity in the cgs electromagnetic system. One abcoulomb equals 10 coulombs and is the quantity of electricity that flows past any point in a circuit in one second when the current is one abampere.

aberration **1.** Distortion from perfect shape in a lens or reflecting mirror or antenna dish. **2.** A small error in the determination of the direction of a source of electromagnetic energy, on account of the motion of the source and/or the detecting apparatus. **3.** A small displacement in the apparent positions of the stars from month to month on account of the earth's orbital motion.

ABETS Acronym for *airborne beacon electronic test set* (NASA).

abfarad The unit of capacitance in the cgs electromagnetic system. One abfarad equals 10^9 farads and is the capacitance across which a charge of 1 abcoulomb produces a potential of 1 abvolt.

abhenry The unit of inductance in the cgs electromagnetic system. One abhenry equals 10^{-9} henry

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

and is the inductance across which a current that changes at the rate of 1 abampere per second induces a potential of 1 abvolt.

ABL Abbreviation of *Automated Biology Laboratory* (NASA).

abmho The obsolete unit of conductance and of conductivity in the cgs electromagnetic system. Replaced with ABSIEMENS.

abnormal dissipation Power dissipation higher or lower than the customary level, usually an overload.

abnormal oscillation **1.** Oscillation where none is desired or expected, as in an amplifier. **2.** Oscillation at two or more frequencies simultaneously when single-frequency operation is expected. **3.** Oscillation at an incorrect frequency. **4.** Parasitic oscillation.

abnormal propagation **1.** The chance shifting of the normal path of a radio wave, as by displacements in the ionosphere, so that reception is degraded. **2.** Unintentional radiation of energy from some point other than the transmitting antenna. **3.** Propagation over a path or in a direction not expected.

abnormal reflections Sharp, intense reflections at frequencies higher than the critical frequency of the ionosphere's ionized layer.

abnormal termination The shutdown of a running computer program or other process. Caused by the detection of an error by the associated hardware that indicates that some ongoing series of actions cannot be executed correctly.

abnormal triggering The false triggering or switching of a circuit or device, such as a flip-flop, by some undesirable source instead of the true trigger signal. Electrical noise pulses often cause abnormal triggering.

abohm The unit of resistance and of resistivity in the cgs electromagnetic system. One abohm equals 10^{-9} ohms and is the resistance across which a steady current of 1 abampere produces a potential difference of 1 abvolt.

abort To deliberately terminate an operation, experiment, process, or project before it has run its normal course.

AB power pack **1.** A portable dry-cell or wet-cell array containing both A and B batteries in one package. **2.** An ac-operated unit in one package for supplying A and B voltages to equipment normally operated from batteries.

abrasion machine An instrument for determining the abrasive resistance of a wire or cable.

abrasion resistance A measure of the ability of a wire or wire covering to resist mechanical damage.

ABS A basic programming abbreviation for the absolute value (of a number, variable, or expression).

abscissa **1.** The independent variable in a function. **2.** The axis (usually horizontal) on the graph of a function that indicates the independent variable.

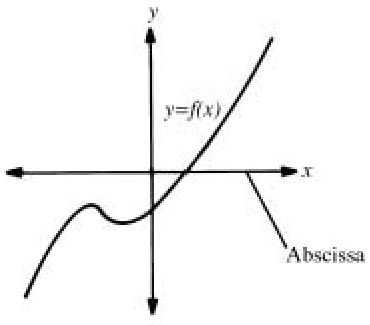

abscissa

absence-of-ground searching selector A rotary switch that searches for an ungrounded contact in a dial telephone system.

absiemens The unit of conductance or conductivity in the cgs electromagnetic system. One absiemens equals 10^9 siemens and is the conductance through which a potential of 1 abvolt forces a current of 1 abampere.

absolute **1.** A temperature scale in which zero represents the complete absence of heat. Units of measure are same as units on Celsius and Fahrenheit scales. See ABSOLUTE SCALE. **2.** Independent of any arbitrarily assigned units of measure or value.

absolute accuracy The full-scale accuracy of a meter with respect to a primary (absolute) standard.

absolute address In a digital computer program, the location of a word in memory, as opposed to location of the word in the program.

absolute code A computer code in which the exact address is given for storing or locating the reference operand.

absolute coding In computer practice, coding that uses absolute addresses.

absolute constant A mathematical constant that has the same value wherever it is used.

absolute delay The time elapsing between the transmission of two synchronized signals from the same station or from different stations, as in radio, radar, or loran. By extension, the time interval between two such signals from any source, as from a generator.

absolute digital position transducer A digital position transducer whose output signal indicates absolute position. (See ENCODER.)

absolute efficiency The ratio X_x/X_s, where X_x is the output of a given device, and X_s is the output of an ideal device of the same kind under the same operating conditions.

absolute encoder system A system that permits the encoding of any function (linear, nonlinear, continuous, step, and so on) and supplies a nonambiguous output.

absolute error The difference indicated by the approximate value of a quantity minus the actual value. This difference is positive when the approximate value is higher than the exact value, and it is negative when the approximate value is lower than the exact value. Compare RELATIVE ERROR.

absolute gain Antenna gain for a given orientation when the reference antenna is isolated in space and has no main axis of propagation.

absolute humidity The mass of water vapor per unit volume of air. Compare RELATIVE HUMIDITY.

absolute instruction A computer instruction that states explicitly and causes the execution of a specific operation.

absolute magnitude For a complex number quantity, the vector sum of the real and imaginary components (i.e., the square root of the sum of the squares of those components). Also see ABSOLUTE VALUE and IMPEDANCE.

absolute maximum rating The highest value a quantity can have before malfunction or damage occurs.

absolute maximum supply voltage The highest supply voltage that can be applied to a circuit without permanently altering its characteristics.

absolute measurement of current Measurement of a current directly in terms of defining quantities. **1.** TANGENT GALVANOMETER method: Current is proportional to the tangent of the angle of deflection of the needle of this instrument. Deflection depends on torque, resulting from the magnetic field produced by current in the galvanometer coil acting against the horizontal component of the earth's magnetic field. **2.** ELECTRODYNAMOMETER method: With this 2-coil instrument, current is determined from the observed deflection, the torque of the suspension fiber of the movable coil, and the coil dimensions.

absolute measurement of voltage Measurement of a voltage directly in terms of defining quantities. **1.** CALORIMETRIC method: A current-carrying coil immersed in water raises the temperature of the water. The difference of potential that forces the current through the coil then is determined in terms of the equivalent heat energy. **2.** Disk-electrometer method: In this setup, a metal disk attached to one end of a balance beam is attracted by a stationary disk mounted below it, the voltage being applied to the two disks. The other end of the beam carries a pan into which accurate weights are placed. At balance, the voltage is determined in terms of the weight required to restore balance, the upper-disk area, and the disk separation.

absolute minimum resistance The resistance between the wiper and the nearer terminal of a potentiometer, when the wiper is as close to that terminal as physically possible. All potentiometers have two such specifications, one for each end terminal.

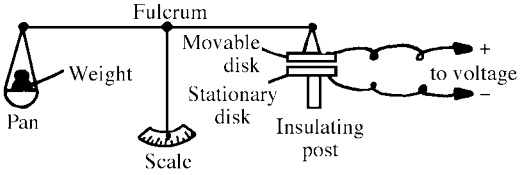

absolute measurement of voltage, 2

absolute Peltier coefficient The product of the absolute Seebeck coefficient and absolute temperature of a material.

absolute pitch A tone in a standard scale, determined according to the rate of vibration, independent of other tones in the range of pitch.

absolute pressure Pressure (force per unit area) of a gas or liquid determined with respect to that of a vacuum (taken as zero).

absolute-pressure transducer A transducer actuated by pressure from the outputs of two different pressure sources, and whose own output is proportional to the difference between the two applied pressures.

absolute scale **1.** A scale in which the zero value indicates the lowest physically possible value that a parameter can attain. **2.** A standard scale for measurement of a quantity. **3.** A universally agreed-upon scale for the determination of a variable quantity. **4.** The Kelvin temperature scale. **5.** The Rankine temperature scale.

absolute Seebeck coefficient The quotient, as an integral from absolute zero to the given temperature, of the Thomson coefficient of a material divided by its absolute temperature.

absolute spectral response The frequency output or response of a device in absolute power units (such as milliwatts) as opposed to relative units (such as decibels).

absolute system of units A system of units in which the fundamental (ABSOLUTE) units are those expressing length (l), mass (m), charge (q), and time (t). All other physical units, including practical ones, are then derived from these absolute units.

absolute temperature Temperature measured on either the Kelvin or Rankine scales, where zero represents the total absence of heat energy.

absolute temperature scale **1.** The Kelvin temperature scale, in which the divisions are equal in size to 1° Celsius, and the zero point is absolute zero, the coldest possible temperature, approximately –273.16° Celsius. **2.** The Rankine temperature scale, in which the divisions are equal in size to 1° Fahrenheit, and the zero point is absolute zero or approximately –459.7° Fahrenheit.

absolute tolerance The value of a component as it deviates from the specified or nominal value. It is usually expressed as a percentage of the specified value.

absolute units Fundamental physical units (see ABSOLUTE SYSTEM OF UNITS) from which all others are derived. See, for example, AMPERE, OHM, VOLT, and WATT.

absolute value The magnitude of a quantity without regard to sign or direction. The absolute value of a is written $|a|$. The absolute value of a positive number is the number itself; thus, $|10|$ equals 10. The absolute value of a negative number is the number with its sign changed: $|-10|$ equals 10.

absolute-value circuit A circuit that produces a unipolar signal in response to a bipolar input and in proportion to the absolute value of the magnitude of the input.

absolute-value computer A computer in which data is processed in its absolute form; i.e., every variable maintains its full value. (Compare to INCREMENTAL COMPUTER.)

absolute-value device In computer practice, a device that delivers a constant-polarity output signal equal in amplitude to that of the input signal. Thus, the output signal always has the same sign.

absolute zero The temperature –273.16°C (−459.7°F and 0 Kelvin). The coldest possible temperature, representing the complete absence of heat energy.

absorbed wave A radio wave that dissipates in the ionosphere as a result of molecular agitation. This effect is most pronounced at low and medium frequencies.

absorptance The amount of radiant energy absorbed in a material; equal to 1 minus the transmittance.

absorption The taking up of one material or medium by another into itself, as by sucking or soaking up. Also, the retention of one medium (or a part of it) by another medium, through which the first one attempts to pass. See, for example, ABSORBED WAVE, ABSORPTION COEFFICIENT, DIELECTRIC ABSORPTION. Compare ADSORPTION.

absorption band See ABSORPTION SPECTRUM.

absorption circuit A circuit that absorbs energy from another circuit or from a signal source—especially a resonant circuit, such as a wavemeter or wavetrap.

absorption current In a capacitor, the current resulting from absorption of energy by the dielectric material.

absorption dynamometer A power-measuring instrument in which a brake absorbs energy from a revolving shaft or wheel.

absorption fading Fading of a radio wave, resulting from (usually) slow changes in the absorption of the wave in the line of propagation.

absorption frequency meter See WAVEMETER.

absorption line See ABSORPTION SPECTRUM.

absorption loss **1.** Transmission loss caused by dissipation of electrical energy, or conversion of it into heat or other forms of energy. **2.** Loss of all or part of a skywave because of absorption by the ionosphere. Also called *ionospheric absorption* or *atmospheric absorption.*

absorption marker A small blip introduced onto an oscilloscope trace to indicate a frequency point. It is so called because it is produced by the action of a frequency-calibrated tuned trap, similar to an absorption wavemeter.

absorption modulation Amplitude modulation of a transmitter or oscillator by means of an audio-frequency-actuated absorber circuit. In its simplest form, the modulator consists of a few turns of wire coupled to the transmitter tank coil and connected to a carbon microphone. The arrangement absorbs energy from the transmitter at a varying rate as the microphone changes its resistance in accordance with the sound waves it receives.

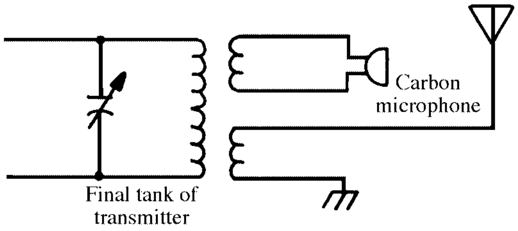

absorption modulation

absorption spectrum For electromagnetic waves, a plot of absorption coefficient (of the medium of propagation) versus frequency. Also called EMISSION SPECTRUM.

absorption trap See WAVETRAP.

absorption wavemeter A resonant-frequency indicating instrument that is inductively coupled to the device under test.

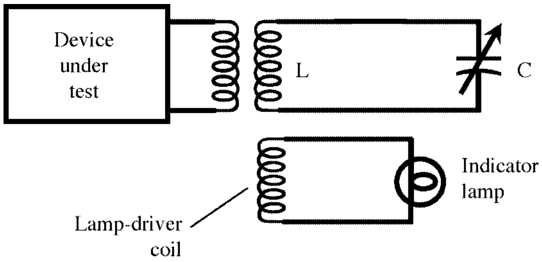

absorption wavemeter

absorptivity In audio and microwave technologies, a measure of the energy absorbed by a given volume of material.

A-B test Comparison of two sounds by reproducing them in alternating succession.

abvolt The unit of potential difference in the cgs electromagnetic system. One abvolt equals 10^{-8} V and is the difference of potential between any two points when 1 erg of work is required to move 1 abcoulomb of electricity between them.

abwatt The unit of power in the cgs electromagnetic system. One abwatt equals 10^{-7} W and is the power corresponding to 1 erg of work per second.

ac **1.** Abbreviation of ALTERNATING CURRENT. **2.** Abbreviation of ATTITUDE CONTROL. **3.** Abbreviation of AERODYNAMIC CENTER. **4.** A suffix meaning AUTOMATIC CALCULATOR or AUTOMATIC COMPUTER.

a/c **1.** Abbreviation of AIRCRAFT. **2.** Abbreviation of AIR CONDITIONING.

Ac Symbol for ACTINIUM.

ACA Abbreviation of *automatic circuit analyzer.*

ac base current Symbol, $I_{B(ac)}$. The ac component of base current in a bipolar transistor.

ac base resistance Symbol, $R_{B(\mathrm{ac})}$. The dynamic base resistance in a bipolar transistor.

ac base voltage Symbol, $V_{B(\mathrm{ac})}$. The ac component of base voltage in a bipolar transistor. It is the ac input signal voltage in a common-emitter amplifier or emitter-follower amplifier.

ac bias In a tape recorder, the high-frequency current that passes through the recording head to linearize operation.

acc **1.** Abbreviation of AUTOMATIC CHROMINANCE CONTROL. **2.** Abbreviation of AUTOMATIC COLOR COMPENSATION. **3.** Abbreviation of ACCELERATION.

ac cathode current Symbol, $I_{K(\mathrm{ac})}$. The ac component of cathode current in an electron tube.

ac cathode resistance Symbol, $R_{K(\mathrm{ac})}$. The dynamic cathode resistance in an electron tube. $R_{K(\mathrm{ac})}$ equals $\mathrm{d}V_K/\mathrm{d}I_K$ for a constant value of V_G.

ac cathode voltage Symbol, $V_{K(ac)}$. The ac component of cathode voltage in an electron tube. It is the ac output signal voltage in cathode-follower and grounded-grid amplifiers.

accelerated life test A test program that simulates the effects of time on devices or apparatus, by artificially speeding up the aging process.

accelerated service test A service or bench test in which equipment or a circuit is subjected to an extreme condition in an attempt to simulate the effects of average use over a long time.

accelerating conductor or relay A conductor or relay that prompts the operation of a succeeding device in a starting mode according to established conditions.

accelerating electrode In a cathode-ray tube or klystron, the electrode to which the accelerating voltage is applied.

accelerating time The elapsed time that starts when voltage is applied to a motor, and ends when the motor shaft reaches maximum speed.

accelerating voltage A positive high voltage applied to the accelerating electrode of a cathode-ray tube to increase the velocity of electrons in the beam.

acceleration at stall The angular acceleration of a servomotor at stall, determined from the stall torque and the moment of inertia of the motor's rotor.

acceleration derivative Acceleration (a) expressed as the second derivative of distance (s) with respect to time (t): a equals d^2s/dt^2.

acceleration potential See ACCELERATING VOLTAGE.

acceleration switch A switch that operates automatically when the acceleration of a body to which it is attached exceeds a predetermined rate in a given direction.

acceleration time The time required by a computer to take in or deliver information after interpreting instructions. Compare ACCESS TIME.

acceleration torque During the accelerating period of a motor, the difference between the torque demanded and the torque actually produced by the motor.

acceleration voltage The potential between accelerating elements in a vacuum tube, the value of which determines average electron velocity.

accelerometer A transducer whose output voltage is proportional to the acceleration of the moving body to which it is attached.

accentuation The emphasis of a desired band of frequencies, usually in the audio-frequency spectrum.

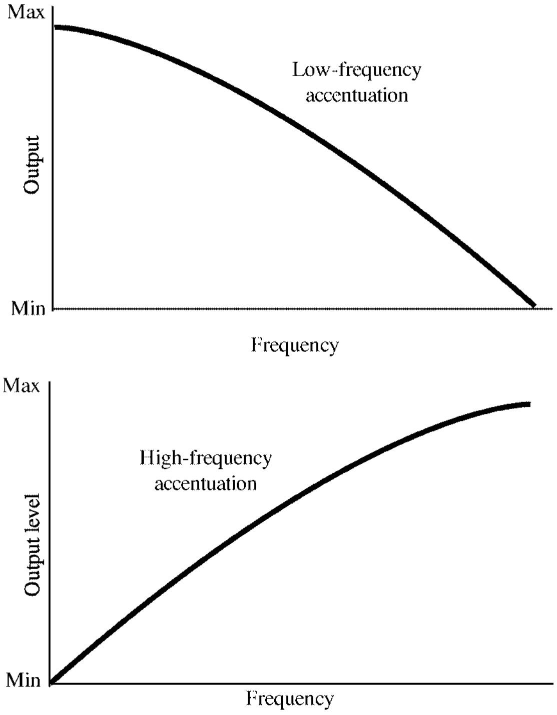

accentuation

accentuator A circuit or device, such as a filter, tone control, or equalizer, used to emphasize a band of frequencies, usually in the audio-frequency spectrum. Also see ACCENTUATION.

acceptable-environmental-range test A test to disclose the environmental conditions that equipment can endure while maintaining at least the minimum desired reliability.

acceptable quality level Abbreviation, AQL. A percentage that represents an acceptable average of defective components allowable for a process, or the lowest quality that a supplier is permitted to regularly present for acceptance.

acceptance sampling plan A probabilistic method of sampling a quantity of units from a lot, and determining from the sample whether to accept the lot, reject the lot, or perform another sampling.

acceptance test A test performed on incoming equipment or on submitted samples to determine if they meet tester's or supplier's specifications.

acceptor **1.** Any device or circuit, such as a series-resonant circuit, that provides relatively easy transmission of a signal, in effect accepting the signal. **2.** A hole-rich impurity added to a semiconductor to make the latter p-type. It is so called because its holes can accept electrons. Compare DONOR.

acceptor circuit See ACCEPTOR, **1.**

acceptor impurity See ACCEPTOR, **2.**

access **1.** To gain entrance to something, such as the interior of the cabinet of a high-fidelity amplifier. **2.** In a computer, the action of going to a specific memory location for the purpose of data retrieval. **3.** A port or opening into a piece of equipment, placed there to make the equipment easy to maintain and repair.

access arm A mechanical device that positions the read/write mechanism in a computer storage unit.

access control register A register that is part of a computer protection system that prevents interference between different software modules.

access method A method of transferring information or data from main storage to an input/output unit.

access right The access status given to computer system users that indicates the method of access permitted (e.g., read a file only or write to a file).

access time The time required by a computer to begin delivering information after the memory or storage has been interrogated.

accidental error An unintentional error committed by a person making measurements and recording data.

accidental triggering The undesired chance-operation of a flip-flop or other switching circuit caused by a noise pulse or other extraneous signal.

ac collector current Symbol, $I_{C(ac)}$. The ac component of collector current in a bipolar transistor.

ac collector resistance Symbol, $R_{C(ac)}$. The dynamic collector resistance of a bipolar transistor. $R_{C(ac)}$ equals dV_C/dI_C for a constant value of base current I_B (in a common-emitter circuit) or emitter current I_E (in a common-base circuit).

ac collector voltage Symbol, $V_{C(ac)}$. The ac component of collector voltage in a bipolar transistor. The ac output signal voltage in a common-emitter or common-base amplifier.

accompanying audio channel The RF signal that supplies television sound. Also called *Cochannnel sound frequency.*

ac component In a complex wave (i.e., one containing both ac and dc), the alternating, fluctuating, or pulsating part of the combination. Compare DC COMPONENT.

accordion A printed-circuit connector contact with a Z-shaped spring that allows high deflection with low fatigue.

ac-coupled flip-flop A flip-flop that is operated by the rise or fall of a clock pulse.

ac coupling Transformer coupling or capacitive coupling, which transmit ac, but not dc. Compare DIRECT COUPLING.

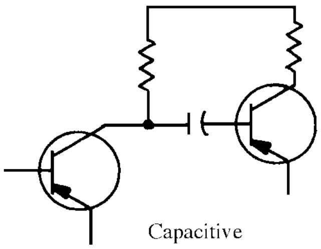

Capacitive

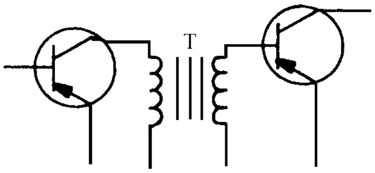

Inductive

ac coupling

accumulator **1.** In a digital computer, a circuit or register device that receives numbers, totals them, and stores them. **2.** Storage battery.

accuracy **1.** Precision in the measurement of quantities and in the statement of physical characteristics. **2.** Degree of precision. Usually expressed, in terms of error, as a percentage of the specified value (e.g., 10 V plus or minus 1%), as a percentage of a range (e.g., 2% of full scale), or as parts (e.g., 100 parts per million).

accuracy rating The maximum error in an instrument, given as a percentage of the full-scale value.

accw Abbreviation of ALTERNATING-CURRENT CONTINUOUS WAVE.

ac/dc Abbreviation of ALTERNATING CURRENT/DIRECT CURRENT. Pertains to equipment that will operate from either ac utility power or a dc power source. A notebook computer is a good example.

ac directional overcurrent relay A relay that works on a specific value of alternating overcurrent that is rectified for a desired polarity.

ac drain current Symbol, $I_{D(ac)}$. The ac component of drain current in a field-effect transistor.

ac drain resistance Symbol, $R_{D(ac)}$. The dynamic drain resistance in a field-effect transistor; $R_{D(ac)}$ equals dV_D/dI_D for a constant value of gate voltage V_G.

ac drain voltage Symbol, $V_{D(ac)}$. The ac component of drain voltage in a field-effect transistor. The ac output signal voltage in a common-source FET amplifier.

ac dump The removal of all ac power from a system or component.

ac emitter current Symbol, $I_{E(ac)}$. The ac component of emitter current in a bipolar transistor.

ac emitter resistance Symbol, $R_{E(ac)}$. The dynamic emitter resistance of a bipolar transistor; $R_{E(ac)}$ equals dV_E/dI_E for a constant value of base current I_B (in an emitter-follower circuit) or collector voltage V_{CC} (in a common-base circuit).

ac emitter voltage Symbol, $V_{E(ac)}$. The ac component of emitter voltage in a bipolar transistor. The ac input signal voltage in a common-base amplifier; the ac output signal voltage in an emitter-follower amplifier.

ac equipment An apparatus designed for operation from an ac power source only. Compare DC EQUIPMENT and AC/DC.

ac erasing In tape recording, the technique of using an alternating magnetic field to erase material already recorded on the tape.

ac erasing head Also called *ac erase head.* In tape and wire recording, a head that carries alternating current to erase material already recorded on the tape or wire. Also see AC ERASING.

acetate Cellulose acetate, a tough thermoplastic material that is an acetic acid ester of cellulose. It is used as a dielectric and in the manufacture of photographic films.

acetate base **1.** The cellulose acetate film that served as the base for the magnetic oxide coating in early recording tape. Most such tapes today are of polyester base. **2.** The cellulose acetate substrate onto which certain photosensitive materials are deposited for lithographic reproduction. Also see ACETATE and ANCHORAGE.

acetate tape Recording tape consisting of a magnetic oxide coating on a cellulose acetate film. Also see ACETATE BASE.

ac gate voltage Symbol, $V_{G(ac)}$. The ac component of gate voltage in a field-effect transistor. The ac input signal voltage.

ac generator **1.** A rotating electromagnetic machine that produces alternating current (e.g., a dynamo or alternator). **2.** An oscillator or combination of an oscillator and an output amplifier.

ac grid voltage Symbol, $V_{G(ac)}$. The ac component of control grid voltage in an electron tube. The ac input signal voltage in a common-cathode amplifier or cathode follower.

A channel The left channel of a two-channel stereo system.

achieved reliability A statement of reliability based on the performance of mass-produced parts or systems under similar environmental conditions. Also called OPERATIONAL RELIABILITY.

achromatic **1.** Without color. In a TV image, the tones from black through gray to white. The term occasionally refers to black-and-white television, although MONOCHROMATIC is more often used in this sense.

achromatic locus Also called *achromatic region.* An area on a chromaticity diagram that contains all points, representing acceptable reference white standards.

achromatic scale A musical scale without accidentals.

ACIA Abbreviation of *asynchronous communications interface adapter.*

acicular Pertaining to the shape of magnetic particles on recording tape. Under magnification, these particles look like thin rods.

acid A substance that dissociates in water solution and forms hydrogen (H) ions (e.g., sulfuric acid). Compare BASE, **2.**

acid depolarizer Also called *acidic depolarizer.* An acid, in addition to the electrolyte, used in some primary cells to slow the process of polarization.

ac line A power line that delivers alternating current only.

ac line filter A filter designed to remove extraneous signals or electrical noise from an ac power line, while causing virtually no reduction of the power-line voltage or power.

ac line voltage The voltage commonly delivered by the commercial power line to consumers. In the United States, the two standards are 117 V and 234 V (~ about 5 percent). The lower voltage is used by most appliances; the higher voltage is intended for appliances and equipment that draws high power, such as electric ovens, cooking ranges, clothes dryers, and amateur-radio amplifiers. In Europe, 220 V is the common standard.

aclinic line Also called *magnetic equator.* An imaginary line drawn on a map of the world or of an area that connects points of zero inclination (dip) of the needle of a magnetic compass.

ACM Abbreviation for *Association for Computing Machinery.*

ac magnetic bias See AC BIAS.

ac meter A meter that is intended to work only on alternating current or voltage. Such meters include iron-vane and rectifier types.

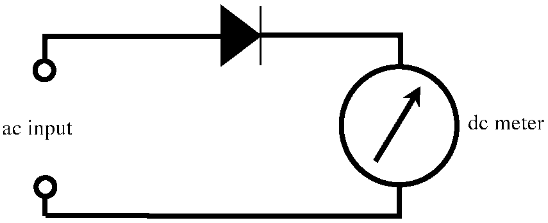

ac meter

ac noise **1.** Electromagnetic interference originating in the ac power lines. **2.** Electrical noise of a rapidly alternating or pulsating nature.

ac noise immunity In computer practice, the ability of a logic circuit to maintain its state, despite excitation by ac noise.

acous Abbreviation for ACOUSTIC.

acoustic Pertaining to audible sound disturbances, usually in air (versus audio-frequency currents or voltages).

acoustic absorption The assimilation of energy from sound waves passing through or reflected by a given medium.

acoustic absorption loss That portion of sound energy lost (as by dissipation in the form of heat) because of ACOUSTIC ABSORPTION.

acoustic absorptivity The ratio of sound energy absorbed by a material to sound energy striking the surface of the material.

acoustic attenuation constant The real-number component of the complex acoustical propagation constant, expressed in nepers per unit distance.

acoustic burglar alarm An alarm that receives the noise made by an intruder. The alarm device responds to the impulses from concealed microphones.

acoustic capacitance The acoustic equivalent of electrical capacitance.

acoustic clarifier In a loudspeaker system, a set of cones attached to the baffle that vibrate to absorb and suppress sound energy during loud bursts.

acoustic communication Communications by means of sound waves. This can be through the atmosphere, or it can be through solids or liquids, such as a taut wire, a body of water, or the earth.

acoustic compliance COMPLIANCE in acoustic transducers, especially loudspeakers. It is equivalent to electrical capacitive reactance.

acoustic consonance An effect that occurs when two objects are near each other but not in physical contact, and both have identical or harmonically related resonant frequencies. An example is shown by two tuning forks with identical fundamental frequencies. If one fork is struck and then brought near the other, the second fork will begin vibrating. If the second fork has a fundamental frequency that is a harmonic of the frequency of the first fork, the second fork will vibrate at its own resonant frequency. See HARMONIC, RESONANCE.

acoustic coupling Data transfer via a sound link between a telephone and a pickup/reproducer. Was once common in computer terminals and facsimile machines. This scheme has been largely replaced by hard wiring and optical coupling.

acoustic damping The deadening or reduction of the vibration of a body to eliminate (or cause to die out quickly) sound waves arising from it.

acoustic delay line Any equivalent of a special transmission line that introduces a useful time delay between input and output signals. In one form, it consists of a crystal block or bar with an input transducer at one end and an output transducer at the other. An electrical input signal in the first transducer sets up sound waves that travel through the interior of the crystal; the piezoelectric reaction of the crystal to sound vibrations sets up an output voltage in the second transducer. The delay is caused by the time required for the acoustic energy to travel the length of the crystal bar.

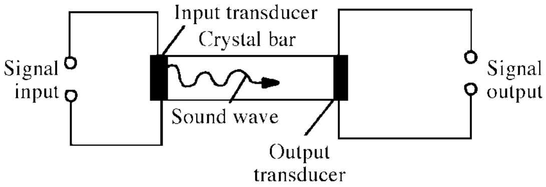

acoustic delay line

acoustic depth finder A direct-reading device for determining the depth of a body of water, or for locating underwater objects via sonic or ultrasonic waves transmitted downward and reflected back to the instrument.

acoustic dispersion Variation of the velocity of sound waves, depending on their frequency.

acoustic elasticity **1.** In a loudspeaker enclosure, the compressibility of air behind the vibrating cone of the speaker. **2.** In general, the compressibility of any medium through which sound passes.

acoustic electric transducer A transducer, such as a microphone or hydrophone, that converts sound energy into electrical energy. Compare ELECTRICAL/ACOUSTIC TRANSDUCER. Also see ACOUSTIC TRANSDUCER.

acoustic feedback A usually undesirable effect that occurs when sound waves from a loudspeaker (or other reproducer) reach a microphone (or other input transducer) in the same system.

This can cause an amplifier to oscillate, with a resultant rumbling, howling, or whistling.

acoustic filter Any sound-absorbing or transmitting arrangement, or combination of the two, that transmits sound waves of desired frequency while attenuating or eliminating others.

acoustic frequency response The sound-frequency range as a function of sound intensity. A means of describing the performance of an acoustic device.

acoustic generator A device that produces sound waves of a desired frequency and/or intensity. Examples are electrical devices (headphones or loudspeakers operated from a suitable oscillator, buzzer, bell, or flame) and mechanical devices (tuning forks, bells, string, or whistles).

acoustic grating A set of bars or slits that are parallel to one another and arranged a fixed distance apart so that an interference pattern forms as sound passes through. Used to determine the wavelength of acoustic waves.

acoustic homing system **1.** A system that uses a sound signal for guidance purposes. **2.** A guidance method in which a missile homes in on noise generated by a target.

acoustic horn A tapered tube (round or rectangular, but generally funnel-shaped) that directs sound and, to some extent, amplifies it. So called to distinguish it from a microwave horn.

acoustic howl See ACOUSTIC FEEDBACK.

acoustician **1.** A person skilled in acoustics (an acoustics technician). **2.** An AUDIOLOGIST.

acoustic impedance Unit, ACOUSTIC OHM. The acoustic equivalent of electrical impedance. Like the latter, acoustic impedance is the total opposition encountered by acoustic force. Also like electrical impedance, acoustic impedance has resistive and reactive components: ACOUSTIC RESISTANCE and ACOUSTIC REACTANCE.

acoustic inductance Also called *inertance.* The acoustic equivalent of electrical inductance.

acoustic inertance See ACOUSTIC INDUCTANCE.

acoustic inhibition See AUDITORY INHIBITION.

acoustic intensity See SOUND INTENSITY.

acoustic interferometer An instrument that evaluates the frequency and velocity of sound waves in a liquid or gas, in terms of a standing wave set up by a transducer and reflector as the frequency or transducer-to-reflector distance varies.

acoustic labyrinth A loudspeaker enclosure whose internal partitions form a maze-like path or "tube" lined with sound-absorbing material. The tube effectively runs from the back of the speaker down to where it terminates in a MOUTH or PORT that opens at the front of the enclosure. The labyrinth provides an extremely efficient reproduction system because of its excellent acoustic impedance-matching capability.

acoustic lens A system of barriers that refracts sound waves the way that an optical lens does with light waves.

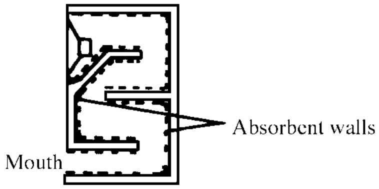

acoustic labyrinth

acoustic line Baffles or other such structures within a speaker that act as the mechanical equivalent of an electrical transmission line to enhance the reproduction of very low bass frequencies.

acoustic load A device that serves simultaneously as the output load of an amplifier and as a transducer of electrical energy into acoustic energy (e.g., headphones or a loudspeaker).

acoustic memory In a computer, a volatile memory element employing an acoustic delay line, often incorporating quartz or mercury as the transmission and delay element.

acoustic mirage A type of sound distortion in which the listener experiences the illusion of two sound sources when there is only one. The phenomenon is caused by the effect of a large temperature gradient in the air or water through which the sound passes.

acoustic mode Crystal-lattice vibration without producing an oscillating dipole.

acoustic noise Interferential (usually disagreeable) sounds carried by the air (or other propagation medium) to the ear or to an acoustic transducer. This is in contrast to electrical noise, which consists of extraneous current or voltage impulses and is inaudible until converted into sound.

acoustic ohm The unit of acoustic resistance, reactance, or impedance. One acoustic ohm equals the volume velocity of 1 cm/s produced by a sound pressure of 1 microbar (0.1 Pa). Also called *acoustical ohm.*

acoustic phase constant The imaginary-number component of the complex acoustic propagation constant expressed in radians per second or radians per unit distance.

acoustic phase inverter A bass reflex loudspeaker enclosure.

acoustic pressure **1.** The acoustic equivalent of electromotive force, expressed in dynes per square centimeter; also called *acoustical pressure.* **2.** Sound pressure level.

acoustic propagation The transmission of sound waves, or subaudible or ultrasonic waves, as a disturbance in a medium, rather than as an electric current or electromagnetic field.

acoustic radiator A device that emits sound waves. Examples are the cone of a loudspeaker, the diaphragm of a headphone, and the vibrating reed of a buzzer.

acoustic radiometer An instrument for measuring the intensity of a sound wave (see SOUND INTENSITY) in terms of the unidirectional steady-state pressure exerted at a boundary as a result of absorption or reflection of the wave.

acoustic reactance Unit, ACOUSTIC OHM. The imaginary-number component of ACOUSTIC IMPEDANCE. It can take the form of ACOUSTIC CAPACITANCE or ACOUSTIC INDUCTANCE.

acoustic reflectivity The ratio F_r/F_i, where F_r is the rate of flow of sound energy reflected from a surface and F_i is the rate of flow of sound energy incident to the surface.

acoustic refraction The deflection of sound waves being transferred obliquely between media that transmit sound at different speeds.

acoustic regeneration See ACOUSTIC FEEDBACK.

acoustic resistance Unit, ACOUSTIC OHM. The real-number component of ACOUSTIC IMPEDANCE. The opposing force that causes acoustic energy to be dissipated in the form of heat. It is attributed to molecular friction in the medium through which sound passes. See ACOUSTIC OHM.

acoustic resonance In an enclosed chamber with walls that reflect sound waves, resonance that occurs at certain wavelengths because the echoes combine in and out of phase. Speaker enclosures almost always have resonance at certain frequencies. This effect can be used to an advantage when it is necessary to get good bass (low-frequency) response from a relatively small speaker.

acoustic resonator **1.** A chamber, such as a box, cylinder, or pipe, in which an air column resonates at a particular frequency. **2.** A piezoelectric, magnetostrictive, or electrostrictive body that vibrates at a resonant audio frequency that is governed by the mechanical dimensions of the body when an audio voltage at that frequency is applied.

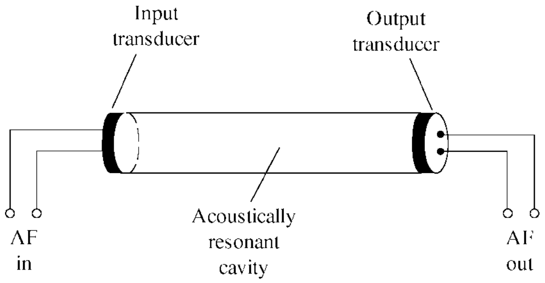

acoustic resonator, 1

acoustics **1.** The physics of sound. The study and applications of acoustic phenomena. **2.** The qualities of an enclosure or sound chamber (room, auditorium, or box) that describe how sound waves behave in it.

acoustic scattering The spreading of a sound wave in many directions as a result of diffraction, reflection, or refraction.

acoustic suspension A loudspeaker design that allows exceptional low-frequency reproduction for a fairly small physical size. An airtight enclosure is used to increase the tension on the speaker cone.

acoustic system **1.** A coordinated array of acoustic components (e.g., acoustic filters, resonators, etc.) that responds to sound energy in a predetermined manner. **2.** An audio-frequency system in which sound energy is converted into electrical energy, processed, and then reconverted into sound energy for a clearly defined purpose.

acoustic telegraph A telegraph that gives audible signals, as opposed to visual signals or printed messages.

acoustic transducer **1.** Any device, such as headphones or a loudspeaker, for converting audio-frequency electrical signals into sound waves. **2.** Any device, such as a microphone, for converting sound waves into alternating, pulsating, or fluctuating currents.

acoustic transmission The direct transmission of sound energy without the intermediary of electric currents.

acoustic transmission system A set of components designed to generate acoustic waves.

acoustic transmissivity Also called *acoustic transmitivity.* The ratio e_t/e_i, where e_t is the sound energy transmitted by a medium, and e_i is the incident sound energy reaching the surface of the medium. Acoustic transmissivity is proportional to the angle of incidence.

acoustic treatment Application of sound-absorbing materials to the interior of an enclosure or room to control reverberation.

acoustic wave The traveling vibration, consisting of molecular motion, via which sound is transmitted through a gas, liquid or solid. Usually refers to sound waves in air.

acoustic wave filter See ACOUSTIC FILTER.

acoustoelectric effect The generation of a voltage across the faces of a crystal by sound waves traveling longitudinally through the crystal.

acoustoelectronics A branch of electronics concerned with the interaction of sound energy and electrical energy in devices, such as surface-wave filters and amplifiers. In such devices, electrically induced acoustic waves travel along the surface of a piezoelectric chip and generate electrical energy. Also called *praetersonics* and *microwave acoustics.*

ac plate current Symbol, $I_{P(ac)}$. The ac component of plate current in a vacuum tube.

ac plate resistance Symbol, $R_{P(ac)}$. The dynamic plate resistance of an electron tube. $R_{P(ac)}$ equals dE_P/dI_P, where E_P is the plate voltage and I_P is the plate current, for a constant value for grid voltage E_G.

ac plate voltage Symbol, $E_{P(ac)}$. The ac component of plate voltage in an electron tube. The ac output-signal voltage in a common-cathode amplifier.

ac power Symbol, P_{ac}. Unit, watt (W). The power acting in an ac circuit, P_{ac} equals $EI \cos q$, where E is in volts, I in amperes, and q is the phase angle. Compare DC POWER. Also see POWER.

ac power supply A power unit that supplies ac only (e.g., ac generator, vibrator-transformer, oscillator, or inverter). Compare DC POWER SUPPLY.

acquisition **1.** The gathering of data from transducers or a computer. **2.** Locating the path of an orbiting body for purposes of collecting telemetered data. **3.** Orienting an antenna for optimum pickup of telemetered data.

acquisition and tracking radar An airborne or ground radar, which locks in on a strong signal and tracks the body that reflects (or transmits) the signal.

acquisition radar A radar that spots an oncoming target and supplies position data regarding the target to a fire-control or missile-guidance radar, which then tracks the target.

acr **1.** Abbreviation of AUDIO CASSETTE RECORDER. **2.** Abbreviation of AUDIO CASSETTE RECORDING SYSTEM.

ac reclosing relay The controlling component in an alternating-current circuit breaker. It causes the breaker to reset after a specified period of time.

ac relay A relay designed to operate on alternating current without chattering or vibrating.

ac resistance Pure resistance in an ac circuit. Unlike reactance and impedance, which are also forms of opposition to the flow of current, ac resistance introduces no phase shift.

acronym A word formed from letters or syllables taken from other applicable words of a multiword term. Acronyms are convenient for naming new devices and processes in electronics. Usually, a term is considered an acronym only when it is spelled in all-capital letters; once the term is accepted and popularized, it is written as a conventional word and is no longer thought of as an acronym. For example, *LASER* was once an acronym for light amplification by the stimulated emission of radiation. By the popularization process, the acronym became a conventional word from which other terms (such as the verb “lase”) were derived.

acrylic resin A synthetic resin used as a dielectric and in electronic encapsulations. It is made from acrylic acid or one of its derivatives.

ACS Abbreviation of *automatic control system.*

ac source current Symbol, $I_{S(ac)}$. The ac component of source current in a field-effect transistor.

ac source resistance Symbol, $R_{S(ac)}$. The dynamic source resistance in a field-effect transistor; $R_{S(ac)}$ equals dV_S/dI_S for a constant value of V_G.

ac source voltage Symbol, $V_{S(ac)}$. The ac component of source voltage in a field-effect transistor. The ac output-signal voltage in a source-follower (grounded-drain) FET amplifier.

acss Abbreviation of *analog computer subsystem.*

ac time overcurrent relay A device with a certain time characteristic, which breaks a circuit when the current exceeds a certain level.

actinic rays Short-wavelength light rays in the violet and ultraviolet portion of the spectrum that give conspicuous photochemical action.

actinism The property whereby radiant energy (such as visible and ultraviolet light, X-rays, etc.) causes chemical reactions.

actinium Symbol, Ac. A radioactive metallic element. Atomic number, 89. Atomic weight, 227.

actinodielectric Exhibiting a temporary rise in electrical conductivity during exposure to light.

actinoelectric effect The property whereby certain materials (such as selenium, cadmium sulfide, germanium, and silicon) change their electrical resistance or generate a voltage on exposure to light. Also see ACTINODIELECTRIC.

actinometer An instrument for measuring the direct heating power of the sun’s rays or the actinic power of a light source.

action current A small transient current that flows in a nerve in the human body as a result of stimulation.

activate To start an operation, usually by applying an appropriate enabling signal.

activation **1.** Supplying electrolyte to a battery cell to prepare the cell for operation. **2.** Causing the acceleration of a chemical reaction.

activation time In the activation of a battery cell (see ACTIVATION, **1**), the interval between addition of the electrolyte and attainment of full cell voltage.

activator A substance added to an accelerator (see ACCELERATOR, **3**) to speed the action of the accelerator.

active Pertaining to a circuit or device that requires a power supply for its operation. This differs from a passive circuit or device, which operates with no external source of power.

active antenna An antenna that uses a small whip, loop, or ferrite loopstick with a high-gain amplifier for receiving at very-low, low, medium, and high radio frequencies (approximately 9 kHz to 30 MHz).

active area The forward-current-carrying portion of the rectifying junction of a metallic rectifier.

active arm See ACTIVE LEG.

active balance In telephone repeater operation, the sum of return currents at a terminal network balanced against the local circuit or drop resistance.

active chord mechanism Abbreviation, ACM. In robots, an electromechanical gripper capable of conforming to irregular objects. It has a structure similar to the human spine, with numerous small, rigid links connected by hinges.

active communications satellite A satellite containing receivers (which pick up beamed electromagnetic signals from a ground point and amplify them) and transmitters (which send signals back to the surface of the earth). Also called *active comsat.* Compare PASSIVE COMMUNICATIONS SATELLITE.

active component **1.** A device capable of some dynamic function (such as amplification, oscillation, or signal control) that usually requires a power supply for its operation. Examples include bipolar transistors, field-effect transistors, and integrated circuits. Compare PASSIVE COMPONENT. **2.** In an ac circuit, a quantity that contains no reactance so that the current is in phase with the voltage.

active component of current See ACTIVE CURRENT.

active computer A computer in an installation or network that is processing data.

active comsat See ACTIVE COMMUNICATIONS SATELLITE.

active control system A device or circuit that compensates for irregularities in the operating environment.

active current In an ac circuit, the current component that is in phase with the voltage. This is in contrast to reactive current, which is not in phase with the voltage, and is “inactive,” with respect to power in the circuit. The active current is equal to the average power divided by the effective voltage.

active decoder An automatic ground-station device that gives the number or letter designation of a received radio beacon reply code.

active device **1.** An electronic component, such as a transistor that needs a power supply, and/or that is capable of amplifying. **2.** Broadly, any device (including electromechanical relays) that can switch (or amplify) by application of low-level signals.

active electric network A network containing one or more active devices or components, usually amplifiers or generators, in addition to passive devices or components.

active element The driven or RF-excited element in a multielement antenna or antenna array.

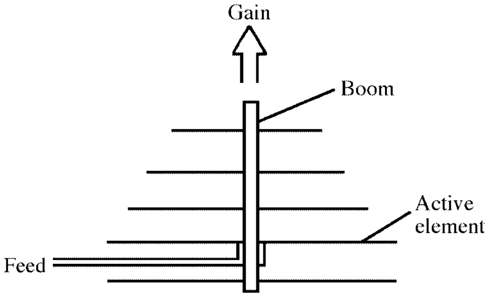

active element

active file A computer file in use (i.e., one that is being updated or referred to).

active filter A bandpass, bandstop, highpass or lowpass filter, consisting of resistors, capacitors, and operational amplifiers, arranged to pass a desired frequency response. Commonly used at audio frequencies.

active infrared detection Detection of infrared rays reflected from a target to which they were beamed.

active jamming Transmission or retransmission of signals for the purpose of disrupting communications.

active junction A pn junction in a semiconductor device that has been created by a diffusion process.

active leg An element within a transducer that changes one or more of its electrical characteristics in response to the input signal of the transducer. Also called *active arm.*

active lines In a U.S. television picture, the lines (approximately 488) that make up the picture. The remaining 37 of the 525 available lines are blanked and are called INACTIVE LINES.

active material **1.** In a storage cell, the chemical material in the plates that provides the electrical action of the cell, as distinguished from the supporting material of the plates themselves. **2.** A radioactive substance. **3.** The phosphor coating of a cathode-ray tube screen. **4.** The material used to coat an electron-tube cathode.

active mixer A signal mixer using one or more active components, such as transistors or integrated circuits. An active circuit provides amplification, input-output isolation, and high input impedance, in addition to the mixing action. Compare PASSIVE MIXER.

active modulator A modulator using one or more active components, such as transistors or integrated circuits. An active circuit provides gain, input-output isolation, and high input impedance, in addition to modulation. Compare PASSIVE MODULATOR.

active network See ACTIVE ELECTRIC NETWORK.

active pressure The electromotive pressure that produces a current in an ac circuit.

active pull-up An arrangement using a transistor as a pull-up resistor replacement in an integrated circuit, providing low output impedance and low power consumption.

active RC network **1.** A resistance-capacitance (RC) circuit that contains active components (transistors or integrated circuits), as well as passive components (capacitors and resistors). **2.** An RC network in which some or all of the resistors and capacitors are simulated by the action of active components.

active repair time The time during which maintenance is done on a system and the system is out of operation.

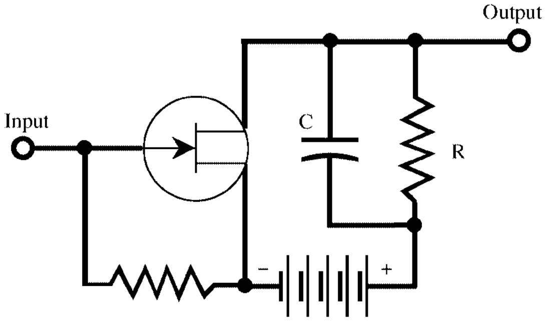

active RC network

active satellite See ACTIVE COMMUNICATIONS SATELLITE.

active sensor In an electronic security system, a transducer that generates an electromagnetic field or acoustic-wave field, and detects changes in the field resulting from the presence or movement of objects in the vicinity.

active substrate In an integrated circuit, a substrate consisting of single-crystal semiconductor material into which the components are formed; it acts as some or all of the components. This is in contrast to a substrate consisting of a dielectric, where the components are deposited on the surface.

active system A radio and/or radar system that requires transmitting equipment to be carried in a vehicle.

active tracking system A system in which a transponder or responder on board a vehicle retransmits information to tracking equipment (e.g., *azusa, secor*).

active transducer **1.** A transducer that contains an active device, such as a transistor or integrated circuit, for immediate amplification of the sensed quantity. **2.** A transducer that is itself an active device.

active wire In the armature of a generator, a wire experiencing induction and, therefore, is delivering voltage.

activity **1.** Intensity of, as well as readiness for, oscillation in a piezoelectric crystal. **2.** Radioactive intensity. **3.** Intensity of thermal agitation. **4.** Thermionic emission of electrons.

activity ratio The ratio of active to inactive records in a computer file.

ac transducer A transducer that either requires an ac supply voltage or delivers an ac output signal—even when operated from a dc supply.

ac transmission The use of an alternating voltage to transfer power from one point to another, usually from generators to a distribution center, and generally over a considerable distance.

actual ground The ground as “seen” by an antenna. The actual ground surface is not necessarily in the same physical location as the true ground surface (i.e., the earth itself). An actual ground can be an artificial ground plane, such as that provided in some antenna structures. Actual ground can also be modified by nearby rooftops, buildings, guy wiring, and utility wiring.

actual height The highest altitude where radio wave refraction actually occurs.

actual power Also called active or AVERAGE POWER. Symbol, P_{avg}. In a resistive circuit under sine-wave conditions, average power is the product of the rms voltage and the rms current. It is also equal to half the product of the maximum current and maximum voltage.

actuating device A device or component that operates electrical contacts to affect signal transmission.

actuating system **1.** An automatic or manually operated system that starts, modifies, or stops an operation. **2.** A system that supplies energy for ACTUATION.

actuating time Also called *actuation time.* The time interval between generation of a control signal, or the mechanical operation of a control device, and the resulting ACTUATION.

actuation **1.** The starting, modification, or termination of an operation or process. **2.** Activation of a mechanical or electromechanical switching device.

actuator An electromechanical device that uses electromagnetism to produce a longitudinal or rotary thrust for mechanical work. It is often the end (load) device of a servosystem.

ACU Abbreviation of *automatic calling unit.*

ac voltage A voltage, the average value of which is zero, that periodically changes its polarity. In one cycle, an ac voltage starts at zero, rises to a maximum positive value, returns to zero, rises to a maximum negative value, and finally returns to zero. The number of such cycles per second is termed the *ac frequency.*

ac voltmeter See AC METER.

acyclic machine Also called ACYCLIC GENERATOR. A dc generator in which voltage induced in the active wires of the armature is always of the same polarity.

A/D Abbreviation for ANALOG-TO-DIGITAL. See ANALOG-TO-DIGITAL CONVERSION.

Ada A microcomputer language designed primarily for use in multi-computer systems, where each small computer communicates with the others, providing some of the advantages of a larger computer.

Adam A communications code word sometimes used for phonetic verbalizing of the letter A. More commonly, ALPHA is used.

adapter **1.** A fitting used to change either the terminal scheme or the size of a jack, plug, or socket to that of another. **2.** A fitting used to provide a transition from one type or style of conductor to another (e.g., waveguide to coaxial line). **3.** An

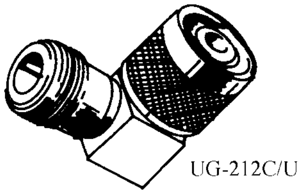

adapter

auxiliary system or unit used to extend the operation of another system (e.g., a citizens-band adapter for a broadcast receiver).

adaptive communication A method of communication that adjusts itself according to the particular requirements of a given time.

adaptive suspension vehicle Abbreviation, ASV. A specialized robot that moves on mechanical legs, rather than on wheels. It generally has six legs and resembles an insect. It is designed to move over extremely irregular or rocky terrain, and to carry a human passenger.

adaptivity The ability of a system to respond to its environment by changing its performance characteristics.

adc Abbreviation of ANALOG-TO-DIGITAL CONVERTER.

Adcock antenna A directional antenna system consisting of two vertical antennas, spaced in such a way that the whole array behaves like a loop antenna. Its members are connected and positioned so that it discriminates against horizontally polarized waves, and delivers output that is proportional to the vector difference of signal voltages induced in the two vertical arms.

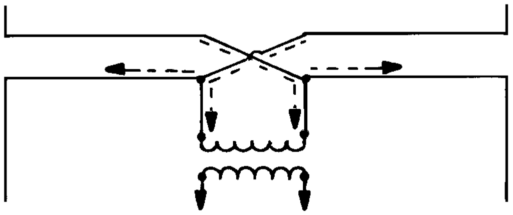

Adcock antenna

Adcock direction finder A radio direction-finding system based on the directivity of the ADCOCK ANTENNA.

Adcock radio range A radio range system with four ADCOCK ANTENNAS situated at the corners of a square, and a fifth antenna at the center of the square.

add-and-subtract relay A stepping relay that can be switched either uprange (add) or downrange (subtract).

addend In a calculation, any number to be added to another. Compare AUGEND.

addend register In a digital computer, the register that stores the addend.

adder **1.** In a digital computer, the device or circuit that performs binary addition. A HALF ADDER is a two-input circuit that can produce a sum output and a carry output, but it cannot accommodate a carry signal from another adder. A FULL ADDER can accommodate a carry input, as well as two binary signals to be added. Also see ANALOG ADDER. **2.** A circuit in a color TV receiver that amplifies the receiver primary matrix signal.

additive **1.** The character or characters added to a code to encipher it. **2.** In a calculation, an item that is to be added. **3.** An ingredient, usually in a small quantity, added to another material to improve the latter in quality or performance.

additive color A color formed by combining the rays from two or three primary-colored lights onto a single neutral surface. For example, by projecting a red and a green beam onto a neutral screen, a yellow additive color results.

additive primaries Primary colors that form other colors in a mixing of light (see ADDITIVE COLOR), but are not themselves formed by mixing other additive primaries. For example, red, green, and blue are the additive primaries used in color television. Through appropriate mixing, these colors can be used to generate an unlimited variety of other colors. Compare SUBTRACTIVE PRIMARIES, which form the color spectrum by mixing pigments rather than lights. In additive systems, each superimposed primary color increases the total light output from the reflecting (viewing) surface; in subtractive systems, each superimposed primary decreases the total reflectivity. Thus, equal combination of additive primaries produces gray or white, and equal combination of subtractive primaries produces gray or black.

addition record An extra data store created in a computer during processing.

address **1.** In computer operations, a usually numerical expression designating the location of material within the memory or the destination of such material. **2.** The accurately stated location of information within a computer; a data point within a grid, matrix, or table; a station within a network. **3.** In computer operations, to select the location of stored information.

address comparator A device that ensures that the address being read is correct.

address computation In digital computer operations, the technique of producing or modifying only the address part of an instruction.

address field In a computer, the part of the instruction that gives the address of a bit of data (or a word) in the memory.

address generation The programmed generation of numbers or symbols used to retrieve records from a randomly stored direct-access file.

address indirect An address that specifies a storage location that contains another address.

address memory The memory sections in a digital computer that contain each individual register.

address modification In computer operations, altering only the address portion of an instruction; if the command or instruction routine is then repeated, the computer will go to the new address.

address part In a digital computer instruction, the part of an expression that specifies the location. Also called ADDRESS FIELD.

address register In a computer, a register in which an address is stored.

add/subtract time In a computer, the time required to perform addition or subtraction, excluding the time required to get the quantities from storage and to enter the sum or difference into storage.

add time In computer operations, the time required to perform addition, excluding the time required to get the quantities from storage and to enter the sum into storage.

a/d converter A device that changes an analog quantity into a digital signal. See ANALOG-TO-DIGITAL CONVERSION.

ADF Abbreviation of AUTOMATIC DIRECTION FINDER.

ADI Abbreviation of ALTERNATE DIGIT INVERSION.

adiabatic damping In an accelerator (see ACCELERATOR, **1**), reduction of beam size as beam energy is increased.

adiabatic demagnetization A technique using a magnetic field to keep a substance at a low temperature, sometimes within a fraction of a degree of absolute zero.

adjacency A character-recognition condition in which the spacing reference lines of two characters printed consecutively in line are closer than specified.

adjacent- and alternate-channel selectivity The selectivity of a receiver or radio-frequency (RF) amplifier, with respect to adjacent-channel and alternate-channel signals. That is, the extent to which a desired signal is passed, and nearby unwanted signals are rejected.

adjacent audio channel See ADJACENT SOUND CHANNEL.

adjacent channel The channel (frequency band) immediately above or below the channel of interest.

adjacent-channel attenuation The reciprocal of the *selectivity ratio* of a radio receiver. The selectivity ratio is the ratio of the sensitivity of a receiver (tuned to a given channel) to its sensitivity in an adjacent channel or on a specified number of channels removed from the original.

adjacent-channel interference In television or radio reception, the interference from stations on adjacent channels. A common form arises from the picture signal in the next higher channel and the sound signal in the next lower channel.

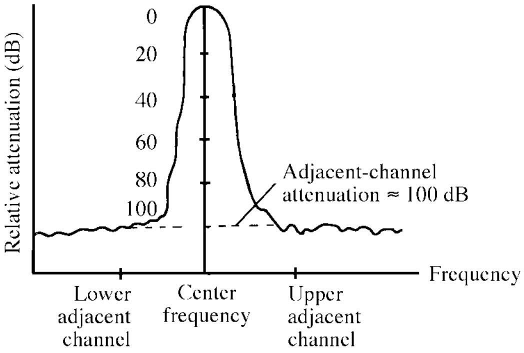

adjacent-channel attenuation

adjacent-channel selectivity The extent to which a receiver or tuned circuit can receive on one channel and reject signals from the nearest outlying channels.

adjacent sound channel In television, the radio-frequency (RF) channel containing the sound modulation of the next lower channel.

adjacent video carrier In television, the radio-frequency (RF) carrier containing the picture modulation of the next higher channel.

adjustable component Any circuit component whose main electrical value can be varied at will (e.g., a variable capacitor, inductor, resistor, or load).

adjustable instrument **1.** An instrument whose sensitivity, range, or response can be varied at will (e.g., multirange meter or wideband generator). **2.** An instrument that requires adjustment or manipulation to measure a quantity (e.g., bridge, potentiometer, or attenuator).

adjustable motor tuning An arrangement that allows the motor tuning of a receiver to be confined to a portion of the frequency spectrum.

adjustable resistor A wirewound resistor in which the resistance wire is partially exposed to allow varying the component's value.

adjustable voltage divider A wirewound resistor with terminals that slide on exposed resistance wire to produce various voltage values.

adjusted circuit A circuit in which leads that are normally connected to a circuit breaker are shunted so that current can be measured under short-circuit conditions without breaker tripping.

adjusted decibels Noise level (in decibels) above a reference noise level (designated arbitrarily as zero decibels) measured at any point in a system with a noise meter that has previously been adjusted for zero (at reference), according to specifications.

admittance Symbol, *Y*. Unit, siemens (formerly mho). The property denoting the comparative ease with which an alternating current flows through a circuit or device. Admittance is the reciprocal of impedance (Z): $Y = 1/Z$.

adp **1.** Abbreviation of AMMONIUM DIHYDROGEN PHOSPHATE, a piezoelectric compound used for sonar crystals. **2.** Abbreviation of AUTOMATIC DATA PROCESSING.

adsorption Adhesion of a thin layer of molecules of one substance to the surface of another without absorption. An example is adsorption of water to the surface of a dielectric. This term is often confused with ABSORPTION because the spellings of the two words are almost identical. Compare ABSORPTION.

adu Abbreviation of *automatic dialing unit.*

advanced-class license An amateur-radio license conveying all operating privileges, except for a few small bands that are allocated to extra-class licensees. The second-highest class of amateur license.

advance information Data published prior to the actual production or availability of a manufactured component, circuit, or system. Advance information is often only an approximate reflection of the expected characteristics of a device.

advance wire A resistance wire used in thermocouples and precision applications. It is an alloy of copper and nickel, which has high resistivity and a negligible temperature coefficient of resistance.

aeolight A glow lamp using a cold cathode and a mixture of inert gases. Because its illumination can be regulated with an applied signal voltage, it is sometimes used as a modulation indicator for motion-picture sound recording.

aerial See ANTENNA.

aerial cable A wire or cable run through the air, using support structures, such as towers or poles.

aerodiscone antenna A miniature discone antenna designed for use on aircraft.

aerodynamics The science dealing with forces exerted by air and other gases in motion—especially upon bodies (such as aircraft) moving through these gases.

aerogram See RADIOGRAM.

aeromagnetic Pertaining to terrestrial magnetism, as surveyed from a flying aircraft.

aeronautical advisory station A civil defense and advisory communications station in service for the use of private aircraft stations.

aeronautical broadcasting service The special service that broadcasts information regarding air navigation and meteorological data pertinent to aircraft operation.

aeronautical broadcast station A station of the aeronautical broadcasting service.

aeronautical fixed service A fixed radio service that transmits information regarding air navigation and flight safety.

aeronautical fixed service station A station that operates in the aeronautical fixed service.

aeronautical ground station A land station that provides communication between aircraft and ground stations.

aeronautical marker-beacon signal A distinctive signal that designates a small area above a beacon transmitting station for aircraft navigation.

aeronautical marker-beacon station A land station that transmits an aeronautical marker-beacon signal.

aeronautical mobile service A radio service consisting of communications between aircraft, and between aircraft and ground stations.

aeronautical radio-beacon station An aeronautical radio-navigation land station that transmits signals used by aircraft and other vehicles to determine their position.

aeronautical radionavigation services Services provided by stations transmitting signals used in the navigation of aircraft.

aeronautical radio service A service that encompasses aircraft-to-aircraft, aircraft-to-ground, and ground-to-aircraft communications important to the operation of aircraft.

aeronautical station A station on land, and occasionally aboard ship, operating in the aeronautical mobile service.

Aeronautical Telecommunication Agency The agency that administers the operation of stations in the aeronautical radio service.

aeronautical telecommunications Collectively, all of the electronic and nonelectronic communications used in the aeronautical service.

aeronautical utility land station A ground station in an airport control tower that provides communications having to do with the control of aircraft and other vehicles on the ground.

aeronautical utility mobile station At an airport, a mobile station that communicates with aeronautical utility land stations and with aircraft and other vehicles on the ground.

aerophare See RADIO BEACON.

aerospace **1.** The region encompassing the earth's atmosphere and extraterrestrial space. **2.** Pertaining to transport and travel in the earth's atmosphere and in outer space. This includes aircraft, orbiting space vessels, and interplanetary spacecraft.

AES Abbreviation for *Audio Engineering Society.*

AEW Abbreviation of *airborne (or aircraft) early warning.*

aF Abbreviation of ATTOFARAD.

AF Abbreviation of AUDIO FREQUENCY.

AFC **1.** Abbreviation of AUTOMATIC FREQUENCY CONTROL. **2.** Abbreviation of AUDIO-FREQUENCY CHOKE.

affirmative In voice communications, a word often used for "yes"—especially when interference is present or signals are weak.

AFIPS Acronym for *American Federation of Information Processing Societies.*

afpc Abbreviation of *automatic frequency/phase control.*

AFSK Abbreviation of AUDIO-FREQUENCY-SHIFT KEYING.

afterglow The tendency of the phosphor of a cathode-ray-tube screen to glow for a certain time after the cathode-ray beam has passed. Also see PERSISTENCE.

afterpulse An extraneous pulse in a multiplier phototube (photomultiplier), induced by a preceding pulse.

AF transformer See AUDIO-FREQUENCY TRANSFORMER.

a/g Abbreviation of AIR-TO-GROUND.

AGC Abbreviation of AUTOMATIC GAIN CONTROL.

AGE Abbreviation of AEROSPACE GROUND EQUIPMENT.

agent An active force, condition, mechanism, or substance that produces or sustains an effect. Thus, a sudden voltage rise is a triggering agent in certain bistable circuits; arsenic is a doping agent in semiconductor processing; the slow cooling of a heated metal to improve ductility is an ANNEALING AGENT.

aging **1.** An initial run of a component or circuit over a certain period of time shortly after manufacture to stabilize its characteristics and performance. **2.** The changing of electrical characteristics or of chemical properties over a protracted period of time.

agonic line An imaginary line connecting points on the earth's surface at which a magnetic needle shows zero declination (i.e., points to true geographic north).

AGREE Acronym for *Advisory Group on Reliability of Electronics Equipment.*

Ah Abbreviation of AMPERE-HOUR. Depending on the standard used, the abbreviation can be amp-hr, a-h, a-hr, or A-h.

aH Abbreviation of ATTOHENRY.

aided tracking In radar and fire control, a system in which manual correction of target tracking error automatically corrects the rate of movement of the tracking mechanism.

AIEE Abbreviation for *American Institute of Electrical Engineers*, now consolidated with the IRE, forming the IEEE.

AIP Abbreviation for *American Institute of Physics.*

air The mixture of gases that constitutes the earth's atmosphere and figures prominently in the manufacture and operation of numerous electronic devices. By volume, air contains about 21 percent oxygen, 78 percent nitrogen, and lesser amounts of argon, carbon dioxide, helium, hydrogen, krypton, neon, and xenon. It also contains varying amounts of water vapor, and in smoggy areas, carbon monoxide and the oxides of sulfur and nitrogen.

airborne intercept radar A type of short-range radar used aboard fighter and interceptor aircraft for tracking their targets.

airborne long-range input Equipment aboard aircraft, for the purpose of facilitating the use of long-range missiles.

airborne noise See ACOUSTIC NOISE.

airborne radar platform Surveillance and altitude-finding radar used aboard aircraft.

air capacitor A capacitor in which air is the dielectric between two sets of conductive plates. Also called *air-dielectric capacitor.*

aircarrier aircraft station On an aircraft, a radio station that is involved in carrying people for hire or in transporting cargo.

air cell A primary electrochemical cell in which the positive electrode is depolarized by reduced oxygen in the air.

air cleaner See DUST PRECIPITATOR.

air column The open space inside an acoustic chamber, pipe, or horn.

air-cooled component A component, such as a power transistor, that is cooled by circulating air, compared with one cooled by a circulating liquid, such as water or oil.

air-cooled transistor A transistor (particularly a power transistor) from which the heat of operation is drawn away, through radiation and convection, into the surrounding air. The transistor is usually mounted on a heatsink or fitted with fins.

air-cooled tube An electron tube from which heat is drawn away, mainly via convection, into the surrounding air. A device called a *chimney* can be placed around the tube, through which air is blown by a fan. Cool air enters through the bottom of the assembly, and hot air escapes from the top.

air-core inductor A coil of wire wound around a hollow cylindrical form or in a loop, designed to introduce inductive reactance into a circuit or system. In practice, the maximum attainable inductance is approximately 1 mH. This type of inductor is used in some wireless transmitters, receivers, and antenna networks. The component can be designed for high current-carrying capacity by using heavy-gauge wire and a large winding radius. The magnetic lines of flux extend considerably beyond the interior of the coil, especially along the winding axis. This increases the likelihood of mutual inductance between the coil and surrounding electrical components, devices, or circuits.

air-core transformer A transformer without a ferromagnetic core, so called because air is the only material at the center of (and immediately surrounding) the transformer coils.

aircraft bonding The practice of solidly connecting, for electrical purposes, the metal parts of an aircraft, including the engine.

aircraft flutter Rapid, repetitive fading and intensifying of a received radio or television signal, resulting from reflections of the signal by passing aircraft.

aircraft station A nonautomatic radio communications station installed on an aircraft.

air-dielectric coax A special type of COAXIAL CABLE designed to have minimum loss. The space between inner and outer conductors is mostly empty (i.e., air-filled). Some such cables are sealed and filled with an inert gas. The inner conductor is held away from the inner wall of the outer conductor by beads, washers, or a spiral-wound filament of high-grade dielectric material, such as polyethylene.

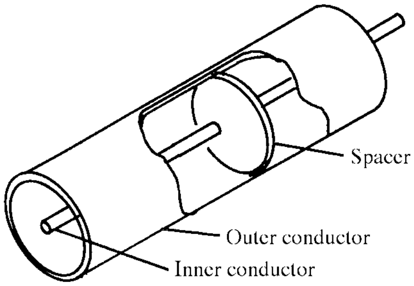

air-dielectric coax

air environment Pertaining to communications equipment aboard aircraft.

airflow The path or movement of air in, through, or around an electronic device or piece of equipment—especially pertaining to an AIR-COOLED COMPONENT.

air gap **1.** A narrow space between two parts of a magnetic circuit (e.g., the gap in the core of a filter choke). Often, this gap is filled with a nonmagnetic material, such as plastic, for mechanical support. **2.** The space between two or more magnetically coupled or electrostatically coupled components. **3.** A device that gets its name from the narrow gap between two small metal balls, needle points, or blunt rod tips therein. When an applied voltage is sufficiently high, a spark discharges across the gap.

air/ground control radio station A station for aeronautical telecommunications related to the operation and control of local aircraft.

air-insulated line **1.** An open-wire feeder or transmission line. Typically, the line consists of two parallel wires held apart by separators (bars or rods of high-grade dielectric material) situated at wide intervals. **2.** AIR-DIELECTRIC COAX.

air-moving device A mechanical device, such as a specially designed fan or blower, used to facilitate air cooling of electronic components.

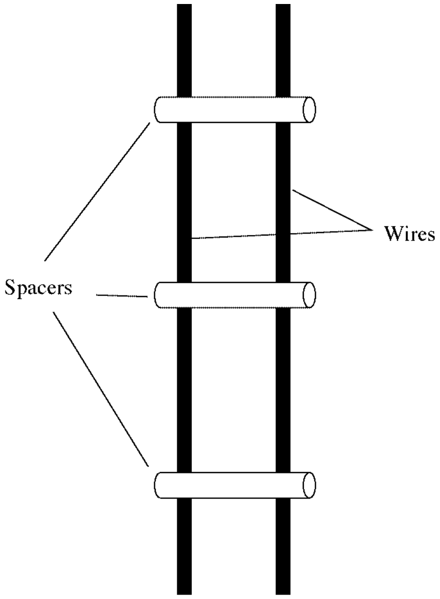

air-insulated line, 1

airport beacon A radio or light beacon that marks the location of an airport.

airport control station A station that provides communications between an airport control tower and aircraft in the vicinity.

airport surveillance radar An air-traffic-control radar that scans the airspace within about 60 miles (approximately 100 kilometers) of an airport, and displays in the control tower the location of all aircraft below a certain altitude and all obstructions in the vicinity.

air-position indicator An airborne computer system that, using airspeed, aircraft heading, and elapsed time, furnishes a continuous indication of the position of the aircraft. The indication is affected by high-altitude winds. Compare GROUND-POSITION INDICATOR.

air-to-air communication Radio transmission from one aircraft to another in flight. Compare AIR-TO-GROUND COMMUNICATION and GROUND-TO-AIR COMMUNICATION.

air-to-ground communication Radio transmission from an aircraft in flight to a station located on the ground. Compare AIR-TO-AIR COMMUNICATION and GROUND-TO-AIR COMMUNICATION.

air-to-ground radio frequency The carrier frequency, or band of such frequencies, allocated for transmissions from an aircraft to a ground station.

airwaves **1.** Radio waves. The term is slang, but is widely used. It probably came from the public's mistaken notion that radio signals are propagated by the air. **2.** Skywaves.

Al Symbol for ALUMINUM.

alabamine See ASTATINE.

alacratized switch A mercury switch in which the tendency of the mercury to stick to the parts has been reduced.

alarm **1.** An electronic security system. **2.** A silent and/or audible alert signal transmitted by an electronic security system when an intrusion occurs. **3.** A silent and/or audible signal that informs personnel of the occurrence of an equipment malfunction.

alarm circuit A circuit that alerts personnel to a system malfunction, a detected condition, or an intruder.

alarm condition **1.** An intrusion or equipment malfunction that triggers an alarm circuit. **2.** The operation of an alarm circuit that occurs in response to an intrusion or equipment malfunction.

alarm hold A device that keeps an alarm sounding once it has been actuated.

alarm output The signal sent from an alarm circuit to a siren, buzzer, computer, or other external device to alert personnel to an ALARM CONDITION.

alarm relay A relay that is actuated by an alarm device.

A-law A form of companding law frequently used in European electronics (the mu-law is more often used in North America). A nonlinear transfer characteristic in companding circuits. It can be continuous, or can be a piecewise linear approximation of a continuous function.

A-law companded Companding by means of an 8-bit binary code following the A-LAW, a specific companding function.

albedo For an unpolished surface, the ratio of reflected light to incident light. It can vary from 0.0 to 1.0, or from 0 to 100 percent.

albedograph An instrument for measuring the albedo of planets.

ALC Abbreviation of AUTOMATIC LEVEL CONTROL.

alerting device An audible alarm that includes a self-contained solid-state audio oscillator. Powered from the ac line or a battery, the device produces a raucous noise when actuated.

Alexanderson antenna A very-low-frequency (VLF) and low-frequency (LF) vertically polarized antenna, designed to minimize ground losses in structures of manageable height. It usually consists of several wires, each quarter-wave resonant with a loading coil, and all connected together at the apex of a tower. The antenna is fed between the ground and the base of one of the wires.

Alford antenna A loop antenna, in a square configuration, with the corners bent toward the center to lower the impedance at the current nodes.

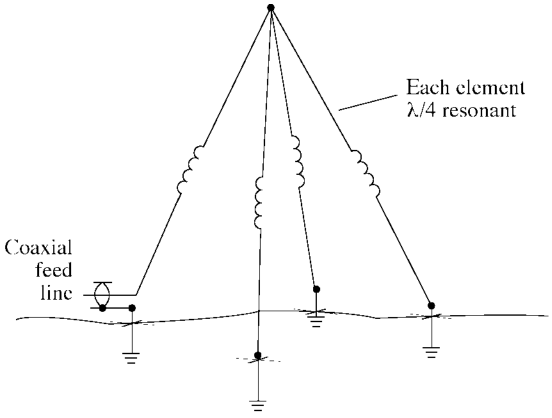

Alexanderson antenna

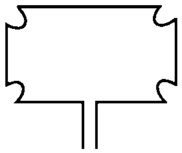

Alford antenna

algebraic adder In computer operations, an adder that provides the algebraic sum, rather than the arithmetic sum, of the entered quantities.

algebraic operation A form of electronic calculator operation, in which the keystrokes proceed in an intuitive sequence, following the way in which the calculation is written down. Compare REVERSE POLISH NOTATION.

algebraic sum The sum of two or more quantities with consideration of their signs. Compare ARITHMETIC SUM.

algorithm A step-by-step procedure for solving a problem, (e.g., the procedure for finding the square root of a number). It can be expressed in a line-by-line instruction set or as a flowchart.

algorithmic language A computer language used to describe a numeral or algebraic process.

alias A label that is an alternate term for items of the same type; a label and several aliases can identify the same data element in a computer program.

aliasing **1.** In analog-to-digital (A/D) conversion, a false output signal that results from a sampling rate that is too slow. Ideally, the sampling rate is at least twice the highest input signal frequency. **2.** Sawtooth-like irregularities, also called *jaggies*, which are sometimes introduced into a bit-mapped computer image when it is changed in size.

aliasing noise A form of signal distortion caused by a signal with an excessive bandwidth.

align **1.** To adjust (i.e., to preset) the circuits of an electronic system, such as a receiver, transmitter, or test instrument, for predetermined response. **2.** To arrange elements in a certain precise orientation and spacing, relative to each other, as in a Yagi antenna. **3.** To orient antennas so that they are in line of sight, with respect to each other.

alignment The process of ensuring that equipment, components, or systems are adjusted, both physically and electronically, for the most efficient possible performance.

alignment chart A line chart for the simple solution of electronic problems. It is so called because its use involves aligning numerical values on various scales, the lines intersecting at the solution on another scale. Also called *nomograph.*

alignment pin A pin or protruding key, usually in the base of a removable or plug-in component, to ensure that the latter will be inserted correctly into a circuit. Often, the pin mates with a keyway, notch, or slot.

alignment tool A specialized screwdriver or wrench (usually nonmagnetic) used to adjust padder or trimmer capacitors or inductor cores.

alive See LIVE.

alkali See BASE, **2.**

alkali metals Metals whose hydroxides are bases (alkalis). The group includes cesium, francium, lithium, potassium, rubidium, and sodium.

alkaline battery **1.** A battery composed of alkaline cells and characterized by a relatively flat discharge curve under load.

alkaline cell A common non-rechargeable electrochemical cell that employs granular zinc for the negative electrode, potassium hydroxide as the electrolyte, and a device called a *polarizer* as the positive electrode. Produces approximately 1.5 volts under no-load conditions. The geometry of construction is similar to that of the zinc–carbon cell, but it can deliver current effectively at lower temperatures. Cells of this type have shelf lives longer than zinc–carbon cells; they also have greater energy-storage capacity per unit volume, but they are more expensive than zinc–carbon cells. They are used in calculators, transistor radios, and cassette tape and compact-disc players. Compare ZINC–CARBON CELL.

alkaline-earth metals The elemental metals barium, calcium, strontium, and sometimes beryllium, magnesium, and radium, some of which are used in vacuum tubes.

alkaline earths Substances that are oxides of the alkaline-earth metals. Some of these materials are used in vacuum tubes.

all-diffused A type of INTEGRATED CIRCUIT in which both active and passive elements have been fabricated by diffusion and related processes.

Allen screw A screw fitted with a six-sided (hexagonal) hole.

Allen wrench A tool used to tighten or loosen an Allen screw. It is a hexagonal rod and is available in various sizes.

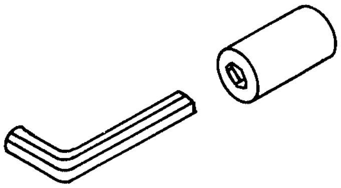

Allen screw and wrench

alligator clip A spring-loaded clip with jagged teeth, designed to be used for temporary electrical connections.

allocate **1.** To assign (especially through legislation) operating frequencies or other facilities or conditions needed for scientific or technical activity; see, for example, ALLOCATION OF FREQUENCIES. **2.** In computer practice, to assign locations in the memory or registers for routines and subroutines.

allocated channel A frequency channel assigned to an individual or group.

allocated-use circuit **1.** A circuit in which one or more channels have been authorized for the exclusive use of one or more services. **2.** A communications link assigned to users needing it.

allocation of frequencies See RADIO SPECTRUM.

allocator A telephone system distributor associated with the finder control group relay assembly. It reserves an inactive line-finder for another call.

allophone A variation in the sound of a phoneme, depending on what comes before and/or after the phoneme in the course of speech. Important in speech recognition and synthesis. There are 128 different phoneme variations in the English language. See PHONEME.

alloter relay A telephone system line-finder relay that reserves an inactive line-finder for the next incoming call from the line.

allotropic Pertaining to a substance existing in two forms.

alloy A metal that is a mixture of several other metals (e.g., brass from copper and zinc), or of a metal and a nonmetal.

alloy deposition In semiconductor manufacture, depositing an alloy on a substrate.

alloy-diffused transistor A transistor in which the base is diffused and the emitter is alloyed. The collector is provided by the semiconductor substrate into which alloying and diffusion are affected. Compare ALLOY TRANSISTOR and DIFFUSE TRANSISTOR.

alloy diode A junction-type semiconductor diode in which a suitable substance (such as p-type) is alloyed into a chip of the opposite type (such as n-type) to form the junction. Also called *alloy-junction diode.*

alloy junction In a semiconductor device, a positive/negative (pn) junction formed by alloying a suitable material (such as indium) with the semiconductor (silicon or germanium).

alloy transistor A transistor whose junctions are created by alloying. Also see ALLOY JUNCTION.

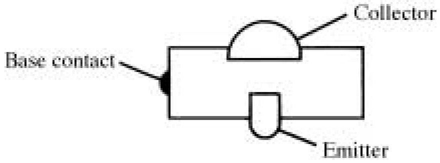

alloy transistor

all-pass filter Also called *all-pass network.* A filter that (ideally) introduces a desired phase shift or time delay, but has zero attenuation at all frequencies.

all-relay central office In telephone service, an automatic central-office switchboard that uses relay circuits to make line interconnections.

all-wave Pertaining to a wide operating-frequency range. Few systems are literally all-wave. For example, a so-called "all-wave radio receiver" might cover 500 kHz to 30 MHz only.

all-wave antenna An antenna that can be operated over a wide frequency range with reasonable efficiency and preferably without needing readjustment. Examples are the DISCONE ANTENNA and the LOG-PERIODIC ANTENNA.

all-wave generator A signal generator that will supply output over a wide range of frequencies.

all-wave receiver A radio receiver that can be tuned over a very wide range of frequencies, such as 10 kHz to 70 MHz.

allyl plastics Plastics, sometimes used as dielectrics or for other purposes in electronics, based on resins made by polymerization of monomers (such as diallyl phthalate) that contain allyl groups.

alnico Coined from the words aluminum, nickel, and cobalt. An alloy used in strong permanent magnets, it contains the constituents noted plus (sometimes) copper or titanium.

alpha **1.** Symbol, α. The current gain of a common-base-connected bipolar transistor. It is the ratio of the differential of collector current to the differential of emitter current; $\alpha = dI_C/dI_E$. For a junction transistor, alpha is always less than unity, but very close to it. **2.** In voice communications, the phonetic representation of the letter A.

alphabet The set of all characters in a natural language.

alphabetic coding In computer practice, an abbreviation system for coding information to be fed into the computer. The coding contains letters, words, and numbers.

alphabetic-numeric Also called *alphabetical-numerical* and *alphanumeric.* In computer operations, pertaining to letters of the alphabet and special characters, and to numerical digits.

alpha cutoff frequency Also called *alpha cutoff.* In a bipolar transistor circuit, the frequency at which the alpha (current gain) becomes 0.707 (70.7 percent) of its value at 1 kHz. A bipolar transistor can have considerable gain at its alpha cutoff. This specification denotes how rapidly a transistor loses gain as the frequency increases, an important consideration in the design of radio-frequency (RF) amplifiers. See ALPHA. Compare GAIN BANDWIDTH PRODUCT.

alpha decay The decay of a substance in which the nuclei of the atoms emit alpha particles, resulting in a change of the atomic number and atomic weight of the substance over a period of time.

alphanumeric See ALPHABETIC-NUMERIC.

alphanumeric code In computer operations or in communications, a code composed of, or using, both letters and numbers.

alphanumeric readout A type of digital readout that displays both letters and numerals.

alpha particle A nuclear particle bearing a positive charge. Consisting of two protons and two neutrons, it is given off by certain radioactive substances. Compare BETA RAYS and GAMMA RAYS.

alpha system An alphabetic code-signaling system.

alphatron An ionizing device in which the radiation source is an emitter of alpha particles.

alteration An inclusive-OR operation.

alternate channel In communications, a channel situated two channels higher or lower than a given channel. Compare ADJACENT CHANNEL.

alternate-channel interference Interference caused by a transmitter operating in the channel beyond an adjacent channel. Compare ADJACENT-CHANNEL INTERFERENCE.

alternate digit inversion In multiplex equipment, a method of switching the binary signals to the opposite state, in accordance with A-law companding.

alternate frequency A frequency allocated as an alternative to a main assigned frequency and used under certain specified conditions.

alternate-mark inversion signal A signal that conveys bits in which the successive signals are of opposite polarity (positive, then negative, then positive, etc.). They are equal in absolute value amplitude.

alternate mode The technique of displaying several signals on an oscilloscope screen by rapidly switching the signals in sequence at the end of each sweep.

alternate routing A secondary, or backup, communications path, used when primary (normal) routing is impossible.

alternating-charge characteristic In a nonlinear capacitor, the relationship between the instantaneous charge and the instantaneous value of an alternating voltage.

alternating current Abbreviation, ac. A current that periodically reverses its direction of flow. In one cycle, an alternation starts at zero, rises to a maximum positive level, returns to zero, rises to a maximum negative level, and again returns to zero. The number of such cycles completed per second is termed the *ac frequency.* Also see CURRENT.

alternating-current continuous wave An amplitude-modulated signal resulting from the operation of an oscillator or RF amplifier with raw ac voltage.

alternating current/direct current See AC/DC.

alternating-current erasing head See AC ERASING HEAD.

alternating-current pulse A short-duration ac wave.

alternating-current transmission **1.** The propagation of alternating currents along a length of conductor—especially for power-transfer purposes. **2.** A means of picture transmission in which a given signal strength produces a constant value of brightness for a very short time.

alternating voltage Also called *alternating-current voltage.* See AC VOLTAGE.

alternation In ac practice, a half cycle. In a complete cycle, there are two alternations, one in the positive direction and one in the negative direction.

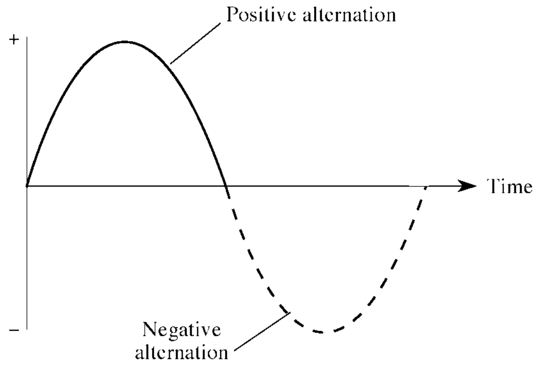

alternation

alternative denial A NOT-AND operation.

alternator Any mechanically driven machine for generating ac power. Sometimes specifically one having a permanent-magnet rotor, such as a *magneto.*

altimeter station An airborne transmitter whose signals are used to determine the altitude of aircraft.

altitude **1.** The vertical distance of an object above sea level. **2.** The vertical distance of an object above the earth's surface. **3.** The angle, measured in degrees, with respect to the horizon, at which a highly directional antenna is pointed.

altitude delay In a plan-position-indicating type of radar, the sync delay introduced between transmission of the pulse and start of the trace on the indicator screen to eliminate the altitude circle in the display.

ALU Abbreviation of ARITHMETIC AND LOGIC UNIT.

alumel An alloy used in the construction of one type of THERMOCOUPLE. It is composed of nickel (three parts) and aluminum (one part).

alumina An aluminum-oxide ceramic used in electron tube insulators and as a substrate in the fabrication of thin-film circuits.

aluminum Symbol, Al. An elemental metal. Atomic number, 13. Atomic weight, 26.98. Aluminum is widely used in electronics, familiar instances being chassis, wire, shields, semiconductor doping, and electrolytic-capacitor plates.

aluminum antimonide Formula, AlSb. A crystalline compound useful as a semiconductor dopant.

aluminized screen A television picture-tube screen with a thin layer of aluminum deposited on its back to brighten the image and reduce ion-spot formation.

Am Symbol for AMERICIUM.

A/m Abbreviation of *ampere per meter:* the SI unit of magnetic field strength.

AM **1.** Abbreviation of *amplitude modulator.* **2.** Abbreviation of AMPLITUDE MODULATION.

amalgam An alloy of a metal and mercury. Loosely, any combination of metals.

amateur **1.** A nonprofessional, usually noncommercial devotee of any technology (i.e., a hobbyist). **2.** A licensed radio operator legally authorized to operate a station in the AMATEUR SERVICE.

amateur band Any band of radio frequencies assigned for noncommercial use by licensed radio amateurs (see AMATEUR, **2**). In the United States, numerous such bands are above 1.8 MHz (160 meters). Also see AMATEUR SERVICE and AMATEUR STATION.

amateur call letters Call letters assigned by a government licensing authority—especially to amateur stations. Call-letter combinations consist of a letter prefix denoting the country in which the station is situated, plus a number designating the location within the country, and two or more letters identifying the particular station. For example: W6ABC: W (or K) = United States, 6 = California, and ABC = identification of individual licensee (issued alphabetically, except under special circumstances).

amateur callsign See AMATEUR CALL LETTERS.

amateur extra-class license The highest class of amateur-radio operator license in the United States. It conveys all operating privileges.

amateur radio **1.** A general term, referring to the practice of operation, experimentation, and other work in and related to the amateur service. **2.** The hardware that comprises an amateur radio station. **3.** A radio receiver, transmitter, or transceiver that is specifically designed for operation in the amateur bands.

amateur radio operator Also called *radio ham* or *ham radio operator.* An individual licensed to transmit radio signals in the amateur service.

amateur service A two-way radio service, existing purely for hobby purposes (i.e., without pecuniary interest).

amateur station A radio station licensed in the AMATEUR SERVICE.

amauroscope An electronic aid to the blind, in which photocells in a pair of goggles receive light images. Electric pulses proportional to the light are impressed upon the visual receptors of the brain through electrodes in contact with nerves above each eye.

amber A yellow or brown fossil resin that is historically important in electronics. It is the first material reported to be capable of electrification by rubbing (Thales, 600 BC). Also, the words *electricity*, *electron*, and *electronics* are derived from the Greek name for amber, *elektron.*

ambience The acoustic characteristic of a room, in terms of the total amount of sound reaching a listener from all directions.

ambient An adjective meaning "surrounding." Often used as a noun in place of the adjective-noun combination (thus, "10 degrees above ambient," instead of "10 degrees above ambient temperature").

ambient humidity The amount of moisture in the air at the time of measurement or operations in which dampness must be accounted for.

ambient level The amplitude of all interference (acoustic noise, electrical noise, illumination, etc.) emitted from sources other than that of a signal of interest.

ambient light Also called *ambient illumination.* Room light or outdoor light incident to a location at the time of measurement or operations.

ambient-light filter In a television receiver, a filter mounted in front of a picture-tube screen to minimize the amount of ambient light reaching the screen.

ambient noise **1.** In electrical measurements and operation, background electrical noise. **2.** In acoustical measurements and operations, audible background noise.

ambient pressure Surrounding atmospheric pressure.

ambient temperature The temperature surrounding apparatus and equipment (e.g., room temperature).

ambient-temperature range **1.** The range over which ambient temperature varies at a given location. **2.** The range of ambient temperature that will cause no malfunction of, or damage to, a circuit or device.

ambiguity **1.** Any unclear, illogical, or incorrect indication or result. **2.** The seeking of a false null by a servo. **3.** In digital computer operations, an error resulting from improper design of logic.

ambiguous count In digital counters, a clearly incorrect count. See ACCIDENTAL TRIGGERING.

ambisonic reproduction A close approximation of the actual directional characteristics of a sound in a given environment. The reproduced sound almost exactly duplicates the sound in the actual environment in which it was recorded.

American Morse code (Samuel F. B. Morse, 1791–1872). Also called *Railroad Morse.* A telegraph code, at one time used on wire telegraph lines in the United States. It differs from the *Continental code*, also called the *International Morse Code*, which is used in radiotelegraphy. Compare CONTINENTAL CODE.

American National Standards Institute Acronym, ANSI. An industrial group in the United States that encourages companies to manufacture devices and equipment in accordance with certain standards. The objective is to minimize hardware incompatibility problems.

American Radio Relay League A worldwide organization of amateur radio operators, headquartered in Newington, Connecticut. The official publications are the monthly magazines, *QST* and *QEX.* They also publish numerous books and other educational materials.

American Standards Association Abbreviation, ASA. At one time, the name of the national association in the U.S. devoted to the formation and dissemination of voluntary standards of dimensions, performance, terminology, etc. See ANSI.

American wire gauge Abbreviation, AWG. Also called *Brown and Sharpe gauge* or *B & S gauge.* The standard American method of designating wire sizes. Wire is listed according to gauge number from 0000 (460 mils diameter) to 40 (3.145 mils diameter).

americium Symbol, Am. A radioactive elemental metal first produced artificially in the 1940s. Atomic number, 95. Atomic weight, 243.

AM/FM receiver A radio set that can receive either amplitude-modulated or frequency-modulated signals. Usually, a band switch incorporates the demodulation-selection circuitry so that as the frequency range is changed, the appropriate detector is accessed.

AM/FM transmitter A radio transmitter whose output signal can be frequency- or amplitude-modulated by a panel selector switch.

AM/FM tuner A compact radio receiver unit that can handle either amplitude- or frequency-modulated signals, and delivers low-amplitude output to a high-fidelity audio power amplifier. Compare AM TUNER and FM TUNER.

Character	*Symbol*	*Character*	*Symbol*
A	.—	U	..—
B	—...	V	...—
C	.. .	W	.——
D	—..	X	.—..
E	.	Y	.. ..
F	.—.	Z	... .
G	——.	1	.——.
H	....	2	..—..
I	..	3	...—.
J	—.—.	4	....—
K	—.—	5	———
L	⸺	6	......
M	——	7	——..
N	—.	8	—....
O	. .	9	—..—
P	.....	0	⸺
Q	..—.	period	..——..
R	. ..	comma	.—.—
S	...	question mark	—..—.
T	—		

American Morse Code

American Wire Gauge (AWG) Diameters

AWG	Millimeters	Inches	AWG	Millimeters	Inches
1	7.35	0.289	21	0.723	0.0285
2	6.54	0.257	22	0.644	0.0254
3	5.83	0.230	23	0.573	0.0226
4	5.19	0.204	24	0.511	0.0201
5	4.62	0.182	25	0.455	0.0179
6	4.12	0.163	26	0.405	0.0159
7	3.67	0.144	27	0.361	0.0142
8	3.26	0.128	28	0.321	0.0126
9	2.91	0.115	29	0.286	0.0113
10	2.59	0.102	30	0.255	0.0100
11	2.31	0.0909	31	0.227	0.00894
12	2.05	0.0807	32	0.202	0.00795
13	1.83	0.0720	33	0.180	0.00709
14	1.63	0.0642	34	0.160	0.00630
15	1.45	0.0571	35	0.143	0.00563
16	1.29	0.0508	36	0.127	0.00500
17	1.15	0.0453	37	0.113	0.00445
18	1.02	0.0402	38	0.101	0.00398
19	0.912	0.0359	39	0.090	0.00354
20	0.812	0.0320	40	0.080	0.00315

AMI See ALTERNATE-MARK INVERSION SIGNAL.

A-minus Also, A-. The negative terminal of an A battery, or pertaining to the part of a circuit connected to that terminal.

ammeter An instrument used to measure the amount of current (in amperes) flowing in a circuit.

ammeter shunt A resistor connected in parallel with an ammeter to increase its current range. Also see AYRTON-MATHER GALVANOMETER SHUNT.

ammeter-voltmeter method The determination of resistance or power values from the measurement of voltage (E) and current (I). For resistance, $R = E/I$; for power, $P = EI$.

ammonium chloride Formula, NH_4Cl. The electrolyte in the carbon-zinc type of primary cell. Also called SAL AMMONIAC.

AMNL Abbreviation of AMPLITUDE-MODULATION NOISE LEVEL.

amortisseur winding **1.** A winding that acts against pulsation of the magnetic field in an electric motor. **2.** A winding that acts to prevent oscillation in a synchronous motor.

amorphous substance A noncrystalline material.

amp **1.** Slang for AMPERE. **2.** Slang for AMPLIFIER—especially in audio high-fidelity applications.

ampacity Current-carrying capacity expressed in amperes.

amperage The strength of an electric current (i.e., the number of amperes).

ampere (Andre Marie Ampere, 1775-1836). Abbreviations, A (preferred), a, amp. The SI base unit of current intensity (I). The ampere is the constant current that, if maintained in two straight parallel conductors of infinite length and of negligible circular cross section and placed 1 meter apart in a vacuum, would produce between the conductors a force of 2×10^{-7} newton per meter. One ampere flows through a 1-ohm resistance when a potential of 1 volt is applied; thus $I = E/R$. Also see MICROAMPERE, MILLIAMPERE, NANOAMPERE, and PICOAMPERE.

ampere balance A device consisting of two conductors in which the force between them (caused by current) is balanced against the gravitational force exerted on an object in the gravitational field of the earth. Used for the precise determination of current of large dimension, or of the size of the ampere.

ampere-hour Abbreviations: Ah, amp-hr. The quantity of electricity that passes through a circuit in one hour when the rate of flow is one ampere. Also see BATTERY CAPACITY.

ampere-hour meter An instrument for measuring ampere-hours. It contains a small motor driven by the current being measured and which moves a point on an ampere-hour scale. The motor speed is proportional to the current. The position of the pointer is proportional to current and elapsed time.

Ampere's law Current flowing in a wire generates a magnetic flux that encircles the wire in the clockwise direction when the current is moving away from the observer.

ampere-turn Symbol, NI. A unit of magnetomotive force equal to 1 ampere flowing in a single-turn coil. The ampere-turns value for any coil is obtained by multiplying the current (in amperes) by the number of turns in the coil.

Amperian whirl The stream of electrons in a single-turn, current-conducting wire loop acting as an elementary electromagnet.

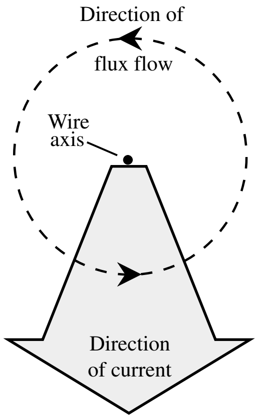

Ampere's Law

amp-hr One style of abbreviating AMPERE-HOUR. Also, Ah.

amplidyne A dynamo-like rotating dc machine that can act as a power amplifier because the response of the output voltage to changes in field excitation is quite rapid. Used in servo systems.

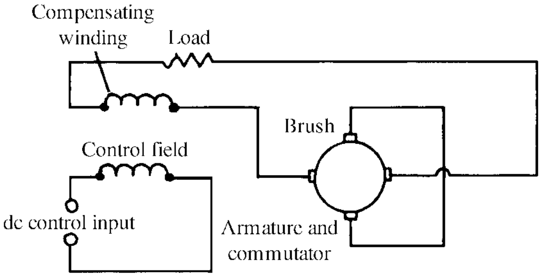

amplidyne

amplification **1.** The process of increasing the magnitude of a signal. This entails an input signal controlling a local power supply to produce a larger output signal. Depending on the kind of input and output signals, amplification can be categorized as CURRENT, VOLTAGE, POWER, or some combination of these. **2.** The qualitative signal increase resulting from the process in **1. 3.** The quantitative signal increase (resulting from the process in **1**), expressed as a factor (such as 100) or in terms of *decibels* (dB). See AMPLIFICATION FACTOR and DECIBEL.

amplification factor **1.** The ratio of the output voltage, current, or power to the input voltage, current, or power of an AMPLIFIER circuit. For voltage or current, this ratio has meaning only when the input and output impedances are identical. **2.** The number of decibels by which an AMPLIFIER circuit increases the amplitude of a signal. For voltage or current, this figure has meaning only when the input and output impedances are identical. See DECIBEL. **3.** The ALPHA or BETA of a bipolar transistor. **4.** In the operation of an electron tube, the ratio of the derivative (instantaneous rate of change) of the plate voltage to the derivative of the grid voltage, for zero change in plate current.

amplified ALC Abbreviation, AALC. An automatic-level-control (ALC) system that uses the amplification of the fed-back control signal. It is used in RF power amplifiers, particularly single-sideband (SSB) linear amplifiers, to prevent overmodulation and nonlinearity.

amplified back bias A declining voltage developed across a fast-time-constant circuit in an amplifier stage and fed back into a preceding stage.

amplifier Any device that increases the magnitude of an applied signal. It receives an input signal and delivers a larger output signal that, in addition to its increased amplitude, is a replica of the input signal. Also see CURRENT AMPLIFIER, POWER AMPLIFIER, and VOLTAGE AMPLIFIER.

amplifier diode Any semiconductor that can provide amplification in a suitable circuit or microwave system. See DIODE AMPLIFIER.

amplifier distortion A change in the waveform of a signal, arising within an amplifier that is operated in compliance with specified conditions.

amplifier input **1.** The terminals and section of an amplifier that receive the signal to be amplified. **2.** The signal to be amplified.

amplifier noise Collectively, all extraneous signals present in the output of an amplifier when no working signal is applied to the amplifier input terminals.

amplifier nonlinearity A condition in which the amplifier output signal does not exhibit a linear relationship to the corresponding input signal. Some amplifiers are designed to operate in a linear manner at all times, but many amplifier types need not function in this manner to be effective. Also see AMPLIFIER DISTORTION and LINEAR AMPLIFIER.

amplifier output **1.** The terminals and section of an amplifier that deliver the amplified signal for external use. **2.** The amplified signal.

amplifier power The power level of the output signal delivered by an amplifier (also called OUTPUT POWER), or the extent to which the amplifier increases the power of the input signal (also called POWER AMPLIFICATION).

amplifier response The performance of an amplifier throughout a specified frequency band. Factors usually included are gain, distortion, amplitude versus frequency, and power output.

amplify To perform the functions of amplification (see AMPLIFICATION, **1**).

amplifying delay line A delay line that causes amplification of signals in a circuit intended for pulse compression.

amplistat A self-saturating magnetic amplifier.

amplitron A backward-wave amplifier used in microwave circuits.

amplitude The extent to which an alternating or pulsating current or voltage swings, positively and negatively, from zero or from a mean value.

amplitude-controlled rectifier A thyratron- or thyristor-based rectifier circuit.

amplitude density distribution A mathematical function giving the probability that, at a given instant in time, a fluctuating voltage has a certain value.

amplitude distortion In an amplifier or network, the condition in which the output-signal amplitude exhibits a nonlinear relationship to the input-signal amplitude.

amplitude error **1.** The error in measuring the amplitude of a signal, normally expressed as a percentage of signal amplitude or as a percentage of full scale. **2.** The frequency at which the output amplitude of a signal is in error by 1% with amplitude at 10% of full scale.

amplitude factor For an ac wave, the ratio of the peak value to the rms value. The amplitude factor of a sine wave is equal to the square root of 2 = 1.4142136.

amplitude fading In the propagation of electromagnetic waves, a condition in which the amplitudes of all components of the signal (i.e., carrier and sidebands) increase and decrease uniformly. Compare SELECTIVE FADING.

amplitude/frequency response Performance of an amplifier throughout a specified range, as exhibited by a plot of output-signal amplitude versus frequency for a constant-amplitude input signal.

amplitude gate A transducer that transmits only those portions of an input wave that lie within two close-spaced amplitude boundaries; also called *slicer.*

amplitude limiter A circuit, usually with automatic gain control (AGC), that keeps an amplifier output signal from exceeding a certain level, despite large variations in input-signal amplitude. A dc-biased diode performs passive limiting action via *clipping.*

amplitude-modulated generator A signal generator whose output is amplitude modulated. Usually, this instrument is an RF generator that is modulated at an audio frequency.

amplitude-modulated transmitter A radiofrequency transmitter whose carrier is varied in amplitude, according to the rate of change of some data-containing signal (such as voice, music, facsimile, television pictures, control signals, or instrument readings).

amplitude modulation Abbreviation, AM. A method of conveying intelligence in wireless communications and broadcasting. The modulating-signal energy appears at sideband frequencies above and below, and very close to, the carrier frequency. These sideband signals carry all the information. The extent of modulation is expressed as a percentage, from 0, which represents an unmodulated carrier, to 100, which represents full modulation. In a signal modulated 100 percent, one-third of the power is used to convey the data; the other two-thirds is consumed by the carrier. This form of modulation is essentially outmoded, although it is still used in the standard broadcast band from 535 to 1605 kHz. See FREQUENCY MODULATION, PHASE MODULATION, SINGLE SIDEBAND.

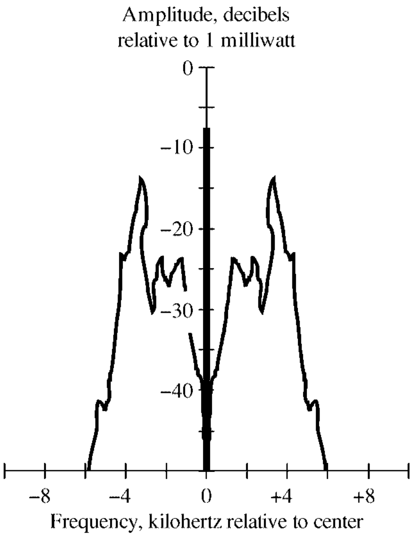

amplitude modulation

amplitude-modulation noise Spurious amplitude modulation of a carrier wave by extraneous signals and random impulses, rather than by the intended data-containing signal.

amplitude noise In radar, amplitude fluctuations of an echo returned by a target. This noise limits the precision of the system.

amplitude of noise The level of random noise in a system. The amplitude of noise is measured in the same way that signal amplitude is measured.

amplitude range The maximum-to-minimum amplitude variation of a signal. It can be expressed as a direct numerical ratio or in decibels.

amplitude response The maximum output obtainable at various frequencies over the range of an instrument operating under rated conditions.

amplitude selection The selection of a signal, according to its correspondence to a predetermined amplitude or amplitude range.

amplitude separator In a television receiver, a circuit that separates the control pulses from the composite video signal.

amplitude suppression ratio The ratio of an undesired output of a frequency-modulated (FM) receiver to the desired output, when the test signal is amplitude modulated and frequency modulated simultaneously.

amplitude-versus-frequency distortion Distortion resulting from varying gain or attenuation of an amplifier or network, with respect to signal frequency.

AMTOR A form of amateur-radio data communications, in which the accuracy of a group of characters in a message is checked periodically by the receiving station. If an error appears likely, then the receiving station sends an instruction to the transmitting station to retransmit that particular group of characters. Characters are sent in bunches with pauses for possible inquiries from the receiving station.

AM tuner A compact radio receiver unit that handles amplitude-modulated signals and delivers low-amplitude audio output to a high-fidelity amplifier. Compare AM/FM TUNER and FM TUNER.

amu Abbreviation of *atomic mass unit.*

amusement robot An electromechanical robot, often computer-controlled, that is intended for use as a toy.

AN- A prefix designator used by American military services to indicate commonality.

anacoustic Pertaining to the lack of sound or absence of reverberation or transmission of sonic waves.

analog **1.** A quantity that corresponds, point for point or value for value, to an otherwise unrelated quantity. Thus, voltage is the analog of water pressure, and current is the analog of water flow. **2.** Varying over a continuous range and, therefore, capable of attaining an infinite number of values or levels. Compare DIGITAL.

analog adder An analog circuit or device that receives two or more inputs and delivers an output equal to their sum.

analog adder/subtracter An analog circuit or device that receives two or more inputs and delivers an output equal to their sum or difference (in any combination), as desired.

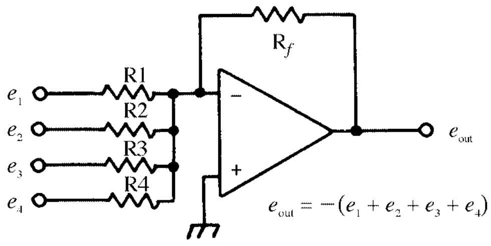

analog adder (inverting)

analog channel In an ANALOG COMPUTER, an information channel in which the extreme limits of data magnitude are fixed, and the data can have any value between the limits.

analog communications Any form of communications in which a carrier, generally an electromagnetic wave or high-frequency current, is varied in a continuous and controlled way by a data-containing signal. See ANALOG, **2.**

analog computer A computer in which input and output quantities are represented as points on continuous (or small-increment) scales. To represent these quantities, the computer uses voltages or resistances that are proportional to the numbers to be worked on. When the quantities are nonelectrical (such as pressure or velocity), they are made analogous by proportional voltages or resistances.

analog data **1.** Data represented in a quantitatively analogous way. Examples are the deflection of a movable-coil meter, the positioning of a slider on a slide rule, and the setting of a variable resistor to represent the value of a nonelectrical quantity. Also see ANALOG. **2.** Data displayed along a smooth scale of continuous values (as by a movable-coil meter), rather than in discrete steps (as by a digital meter).

analog differentiator An analog circuit or device whose output waveform is the derivative of the input-signal waveform, with respect to time.

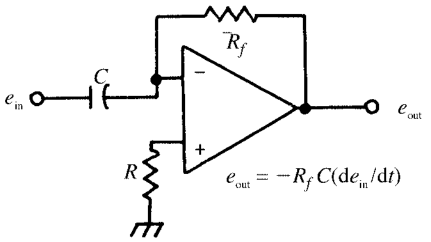

analog differentiator

analog divider An analog circuit or device that receives two inputs and delivers an output equal to their quotient.

analog electronics Electronic techniques and equipment that is based on uniformly changing signals, such as sine waves, and often having continuous-scale indicators, such as D'Arsonval meters. Compare DIGITAL ELECTRONICS.

analog information Approximate numerical information, as opposed to digital information, which is assumed to be exact.

analog integrator An analog circuit or device whose output waveform is the integral of the input signal waveform, with respect to time.

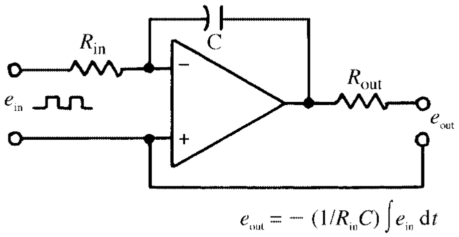

analog integrator

analog inverting adder An analog adder that delivers a sum with the opposite sign to that of the input quantities.

analog meter An indicating instrument that uses a movable-coil arrangement or the equivalent, causing a rotating pointer to indicate a particular value on a graduated printed scale. Compare DIGITAL METER.

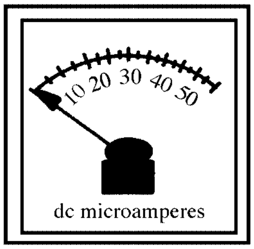

analog meter

analog multiplexer **1.** A multiplexer used with analog signals (see MULTIPLEXER). **2.** An analog time-sharing circuit.

analog multiplier An analog circuit or device that receives two or more inputs and delivers an output equal to their product.

analog network A circuit that permits mathematical relationships to be shown directly by electric or electronic means.

analogous pole In a PYROELECTRIC MATERIAL, the end or face having the positive electric charge.

analog output An output quantity that varies smoothly over a continuous range of values, rather than in discrete steps.

analog record Also called *analog recording.* A record or recording method in which some property of the recorded material, such as displacement or magnetization, varies over a continuous range that is relative to time and/or physical position.

analog recorder Any recorder, such as a recording oscillograph, potentiometric recorder, electroencephalograph, electrocardiograph, or lie detector, that produces an analog record. The counterpart is a digital recorder, which produces a readout in discrete numbers (printed or visually displayed).

analog representation Representation of information within a smooth, continuous range, rather than as separate (discrete) steps or points.

analog signal A signal that attains an infinite number of different amplitude levels, as opposed to one that can attain only a finite number of levels as a function of time.

analog subtracter An analog circuit or device that receives two inputs and delivers an output equal to their difference.

analog summer See ANALOG ADDER.

analog switch A switching device that will only pass signals that are faithful analogs of transducer parameters.

analog-to-digital conversion **1.** A process in which an analog signal (such as a voice waveform) is changed into a digital or binary signal that conveys the same information. This process is commonly used in digital computers to encode sounds and images. It is also used in communications systems to improve efficiency, minimize the necessary bandwidth, and optimize the signal-to-noise ratio. **2.** A process in which continuous mechanical motion is encoded into a digital or binary electronic signal.

analog-to-digital converter Any circuit or device that performs ANALOG-TO-DIGITAL CONVERSION.

analysis **1.** The rigorous determination of the constants and modes of operation for electronic equipment. Compare SYNTHESIS. **2.** A branch of mathematics dealing with point sets, relations, and functions.

analytical engine A primitive mechanical calculating machine, invented in 1833 by Charles Babbage.

analyzer **1.** Any instrument that permits analysis through close measurements and tests (e.g., distortion analyzer, WAVE ANALYZER, or gas analyzer). **2.** A computer program used for debugging purposes; it analyzes other programs and summarizes references to storage locations. **3.** An analysis interface to an oscilloscope.

anastigmatic yoke Also called *full-focus yoke.* In a television (TV) receiver, a deflection yoke with a cosine winding for better focus at the edges of the picture.

anchorage In plastic recording tape, the adhesion of the magnetic oxide coating to the surface of the tape.

ancillary equipment Equipment that does not directly enter into the operation of a central system. Examples are input/output components of a computer and test instruments attached to a system.

AND circuit In digital systems and other switching circuits, a logic gate whose output is high (logic 1) only when all input signals are high. Otherwise the output is low (logic 0). Compare OR CIRCUIT.

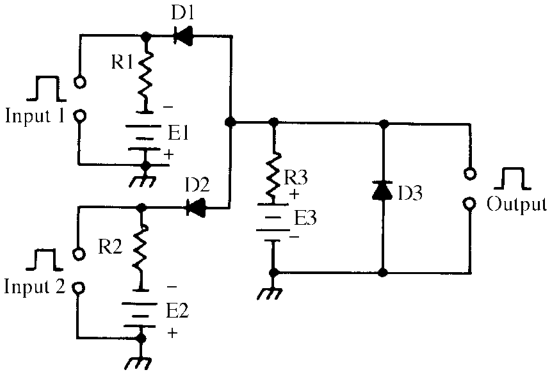

AND circuit

Anderson bridge An ac bridge circuit with six impedances, permitting the value of an unknown inductance to be determined in terms of a standard capacitance.

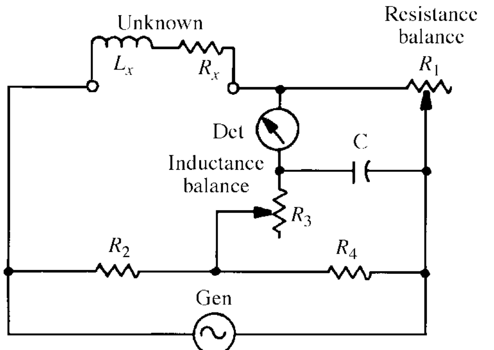

$$L_x = C[R_3(1 + R_2/R_4) + R_2]$$
$$R_x = (R_1 R_2)/R_4$$

Anderson bridge

AND gate **1.** AND circuit. **2.** In a TV receiver, an AND circuit that holds the keyed-AGC signal off until a positive horizontal flyback pulse and a horizontal sync pulse appear simultaneously at the input.

android A sophisticated robot built in humanoid form. Usually, it propels itself by rolling on wheels or on a track drive. A rotatable head contains position sensors, a machine vision system, and/or a machine hearing system. Mechanical arms are equipped with end effectors to perform various tasks. The most advanced androids have self-contained computer control systems.

anechoic Pertaining to the absence of echoes. Examples: ANECHOIC CHAMBER, *anechoic enclosure,* or *anechoic room.*

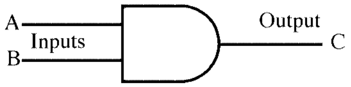

Truth Table

A	B	C
0	0	0
0	1	0
1	0	0
1	1	1

AND gate

anechoic chamber An enclosure that does not reflect sound waves that approach its walls. Such a chamber is used to test certain audio devices.

anemograph An electromechanical device that produces a recording of wind speed versus time. Generally, it consists of an ANEMOMETER connected to a PEN-AND-INK RECORDER via a suitable electronic interface.

anemometer An instrument that measures or indicates wind speed, or speed and direction (velocity).

angel **1.** An extraneous image, usually of short duration, on a cathode-ray-tube (CRT) display. The term applies particularly to anomalies in a radar image caused by low-atmospheric reflection, birds, or other mobile objects. **2.** Air-deployed metallic debris, also known as *chaff,* designed to create radar echoes as a decoy or diversion tactic.

angle jamming A radar jamming technique in which the return echo is jammed with a signal containing improper azimuth or elevation angle components.

angle modulation Variation of the angle of a sine-wave carrier in response to the modulating source, as in FREQUENCY MODULATION and PHASE MODULATION.

angle noise In radar reception, the interference resulting from variations in the angle at which an echo arrives from the target.

angle of arrival The angle which the line of propagation of an incoming radio wave makes with the surface of the earth. Compare ANGLE OF DEPARTURE.

angle of azimuth The horizontal angle between the viewer and object or target, usually measured clockwise from north.

angle of beam The angle enclosing most of the transmitted energy in the radiation from a directional antenna. It is usually measured between the *half-power points* in the main lobe of the directional pattern. This angle can be measured in the horizontal (azimuth) plane or in the vertical (elevation) plane.

angle of conduction **1.** Also called *angle of flow.* The number of degrees of an excitation-signal cycle during which output (drain, collector or plate) current flows in an amplifier circuit. **2.** The number of degrees of any sine wave at which conduction of a device (e.g., a diode) begins.

angle of convergence **1.** In any graphical representation, the angle formed by any two lines or plots that come together at a point. **2.** The angle formed by the light paths of two photocells focused on the same object.

angle of declination The angle between the horizon and a descending line. Compare ANGLE OF ELEVATION.

angle of deflection In a cathode-ray tube, the angle between the electron beam at rest and a new position resulting from deflection.

angle of departure The angle, relative to the horizon, made by the line of propagation of a transmitted radio wave. Compare ANGLE OF ARRIVAL.

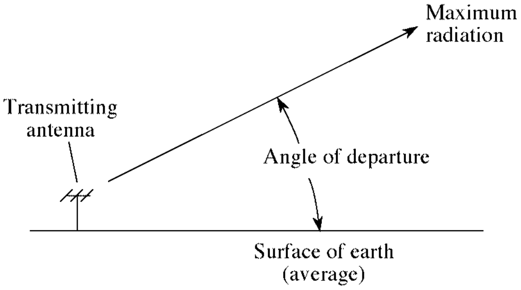

angle of departure

angle of depression See ANGLE OF DECLINATION.

angle of divergence In a cathode-ray tube, the angle formed by the spreading of an undeflected electron beam as it extends from the gun to the screen.

angle of elevation The angle that an ascending line subtends, with respect to the horizon. Compare ANGLE OF DECLINATION.

angle of flow See ANGLE OF CONDUCTION.

angle of incidence The angle, measured relative to the perpendicular (orthogonal) to a surface or boundary, subtended by an approaching ray. Compare ANGLE OF REFLECTION and ANGLE OF REFRACTION.

angle of lag The phase difference (in degrees or radians) whereby one component follows another in time, both components being of the same frequency. Compare ANGLE OF LEAD. Also see PHASE ANGLE.

angle of lead The phase difference (in degrees or radians) whereby one component precedes another in time, both components being of the same frequency. Compare ANGLE OF LAG. Also see PHASE ANGLE.

angle of radiation **1.** The angle, measured with respect to the horizon, at which the principal lobe of an electromagnetic wave leaves a transmitting antenna. **2.** The angle, measured relative to the horizon, of a receiving or transmitting antenna's optimum sensitivity.

angle of reflection The angle, measured relative to the perpendicular (orthogonal) to a surface, subtended by a ray leaving the surface after having been reflected from it. Compare ANGLE OF INCIDENCE.

angle of refraction The angle, measured relative to the perpendicular (orthogonal) to a boundary between two different media, subtended by a ray leaving the boundary after having been refracted thereat. Compare ANGLE OF INCIDENCE.

angle tracking noise Noise in a servo system that results in a tracking error.

angstrom (Anders J. Angstrom, 1814–1874). A unit of length used to describe certain extremely short waves and microscopic dimensions; 1 angstrom equals 10^{-4} microns (10^{-10} meters).

angular deviation loss The ratio of microphone or loudspeaker response on the principal axis of response to the response at a designated angle from that axis. Expressed in decibels.

angular difference See PHASE ANGLE.

angular displacement In an ac circuit, the separation, in degrees, between two waves. See PHASE ANGLE.

angular frequency The frequency of an ac signal, expressed in *radians per second (rad/sec)* and approximately equal to $6.28f$, where f is the frequency in Hertz.

angular length Length, as along the horizontal axis of an ac wave or along the standing-wave pattern on an antenna, expressed as the product of radians and wavelength.

angular-mode keys On a calculator or computer, the DEG, RAD, and GRAD keys for expressing or converting angles in DEGREES, RADIANS, and GRADS, respectively.

angular phase difference For two sinusoidal waves, the phase difference, expressed in degrees or radians.

angular rate In navigation, the rate of bearing change, expressed in degrees or radians.

angular resolution The ability of a radar to distinguish between two targets by angular measurement.

angus pen recorder An instrument that makes a permanent record of the time whenever a channel is used.

anharmonic oscillator An oscillating device in which the force toward the balance point is not linear, with respect to displacement.

anhysteresis The magnetization of a material by a unidirectional field containing an alternating field component of gradually decreasing amplitude.

anhysteretic state The condition of a substance after it has been subjected to a strong magnetic field, the intensity of which alternates in direction and diminishes gradually to zero.

animism A belief or philosophy, held especially in Eastern civilizations, such as Japan, that all things contain an essence of life. This theory renders irrelevant the question of whether or not machines, such as computers and robots can be "alive."

anion A negative ion. Also see ION.

anisotropic Pertaining to the tendency of some materials to display different magnetic and other physical properties along different axes.

ANL Abbreviation of AUTOMATIC NOISE LIMITER.

anneal To heat a metal to a predetermined temperature and let it cool slowly. The operation prevents brittleness and often stabilizes electrical characteristics.

annealed laminations Core laminations for transformers or choke coils that have been annealed.

annealed shield A magnetic shield for cathode-ray tubes, that has been processed by annealing.

annealed wire Soft-drawn wire that has been subjected to annealing.

annotations **1.** Marking on copies of original engineering-installation documents to show changes made during the installation. **2.** Any set of comments or notes accompanying a program, an equipment or system, or a process.

annular **1.** Pertaining to the region between two concentric circles that lie in the same plane; ring-shaped. **2.** Pertaining to two or more concentric circles that lie in a common plane.

annular conductor A number of wires stranded in three concentric layers of alternating twists around a hemp core.

annular transistor A mesa transistor in which the base and collector take the form of concentric rings around a central emitter.

annulling network A subcircuit that shunts a filter to cancel reactive impedance at the extreme ends of the pass band of the filter.

annunciation relay A relay that indicates whether or not a circuit is carrying current.

annunciator A device that produces loud sound and/or conspicuous light to attract attention (e.g., the electronic siren in an automotive security system).

anode **1.** The positive electrode of a vacuum tube or solid-state device (i.e., the electrode toward which electrons move during current flow). **2.** In an electrochemical cell, the electrode that loses electrons by oxidation. This is usually the negative electrode.

anode balancing coil Mutually coupled windings used to maintain equal currents in parallel anodes operating from a common transformer terminal.

anode current Current flowing in the anode circuit of a device.

anode efficiency Also called *plate efficiency.* In a power amplifier using an electron tube, the ratio P_o/P_i, where P_o is the output power in watts and P_i is the dc anode power input in volt-amperes.

anode power input Symbol, $P_{A(input)}$. The product of anode current and anode voltage.

anode power supply The ac or positive dc power supply unit that delivers current and voltage to the anode of a device.

anode saturation The point beyond which a further increase in anode voltage does not produce an increase in anode current.

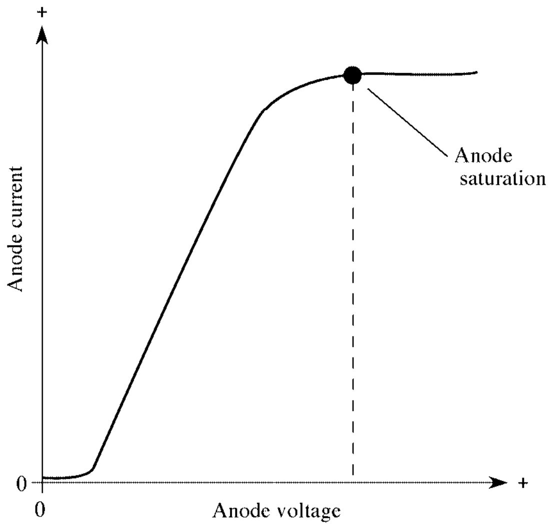

anode saturation

anode strap In a multicavity magnetron, a metal strap connecting the anodes.

anode terminal **1.** In a diode, the terminal to which a positive dc voltage must be applied for forward bias. Compare CATHODE TERMINAL. **2.** In a diode, the terminal at which a negative dc voltage appears when the device is used as an ac rectifier. Compare CATHODE TERMINAL. **3.** The terminal that is connected internally to the anodic element of any device.

anode voltage Symbol, E_A or V_A. The difference in potential between the anode and cathode of a device.

anodic Pertaining to the anode of a device, or to anode-like effects.

anodizing An electrolytic process in which a protective oxide film is deposited on the surface of a metallic body acting temporarily as the anode of the electrolytic cell.

anomalous dispersion Dispersion of electromagnetic radiation that is characterized by a decrease in refractive index with increase in frequency.

anomalous propagation **1.** The low-attenuation propagation of UHF or microwave signals through atmospheric layers. **2.** Unusual, bizarre, or unexplainable electromagnetic-wave propagation (e.g., apparent F-layer ionospheric effects in the FM broadcast band). **3.** Rapid fluctuation of a sonar echo because of variations in propagation.

anoxemia toximeter An electronic instrument for measuring or alerting against the onset of anoxemia (deficiency of oxygen in the blood)—especially in airplane pilots.

AN radio range A navigational facility entailing four zones of equal signal strength. When the aircraft deviates from course, an aural Morse-code signal, A (DIT DAH) or N (DAH DIT) is heard; but when the aircraft is on course, a continuous tone is heard.

ANSI Acronym for *American National Standards Institute.*

AN signal The signal provided by an AN radio range to apprise aircraft pilots of course deviation.

answerback The automatic response of a terminal station to a remote-control signal.

answer cord In a telephone system, the cord used for answering subscribers' calls and incoming trunk calls.

answering machine A device that automatically answers a telephone and records an audio message from the caller.

answer lamp A telephone switchboard lamp that lights when an answer cord is plugged into a line jack; it switches off when the telephone answers and lights when the call is completed.

ant Abbreviation of ANTENNA.

antenna In a communications system, a specialized transducer that converts incoming electromagnetic fields into alternating electric currents having the same frequencies (*receiving antenna*), or converts an alternating current at a specific frequency into an outgoing electromagnetic field at the same frequency (*transmitting antenna*). An antenna can be a simple wire or rod, or a complicated structure. Thousands of geometries and specifications are possible. The optimum antenna type for a given situation depends on the communications frequency, the distance to be covered, and various other factors.

antenna ammeter An RF ammeter, usually of the thermocouple type, employed to measure current flowing to a transmitting antenna.

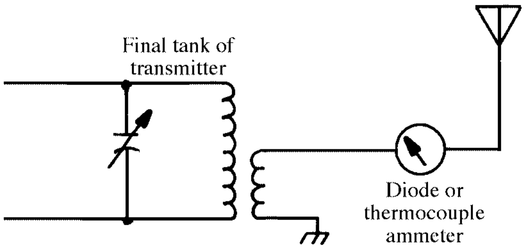

antenna ammeter

antenna amplifier **1.** A radio-frequency amplifier, often installed at the antenna, used to boost signals before they reach a receiver (also called an *RF preamplifier*). **2.** Occasionally, the first RF amplifier stage of a receiver, also known as the *front end.*

antenna array See ARRAY.

antenna bandwidth The frequency range throughout which an antenna will operate at a specified efficiency without needing alteration or adjustment.

antenna beamwidth A measure of the extent to which a directional antenna focuses a transmitted electromagnetic field, or focuses its response to incoming electromagnetic fields. Expressed as the angle in degrees between opposite *half-power points* in the main lobe of the directional pattern. Usually determined in the horizontal plane, but occasionally in the vertical plane.

antenna coil The primary coil of the input RF transformer of a receiver, or the secondary coil of the output RF transformer of a transmitter.

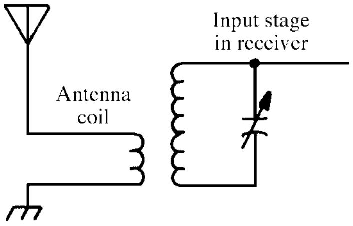

antenna coil

antenna coincidence The condition in which two directional antennas are pointed directly toward each other.

antenna-conducted interference Extraneous signals generated in a transmitter or receiver and presented to the antenna, from which they are radiated.

antenna core A ferrite rod or slab around which a coil of wire is wound to act as a self-contained antenna, usually in a miniature receiver.

antenna coupler A device consisting of an inductor, RF transformer, or a combination of inductor(s) and capacitor(s), used to match the impedance of an antenna to that of a transmitter or receiver. Also known as a *transmatch* or *antenna tuner.*

antenna coupling Inductive and/or capacitive coupling used to optimize the transfer of energy from an antenna to a receiver, or from a transmitter to an antenna.

antenna current **1.** Radio-frequency current flowing from a transmitter into an antenna. **2.** Radio-frequency current flowing from a receiving antenna into a receiver.

antenna detector A circuit that warns aircraft personnel that they are being observed by radar. It picks up the radar pulses and actuates a warning light or other device.

antenna diplexer A coupling device that permits several transmitters to share one antenna without troublesome interaction. Compare ANTENNA DUPLEXER.

antenna directivity The directional characteristics of a transmitting or receiving antenna, usually expressed qualitatively (e.g., *omnidirectional*, *bidirectional*, or *unidirectional*). A more precise expression is ANTENNA BEAMWIDTH.

antenna director In a directional antenna, a PARASITIC ELEMENT situated in front of the radiator and separated from it by an appropriate fraction of a wavelength. Its function is to intensify radiation in the direction of transmission. Compare ANTENNA RADIATOR and ANTENNA REFLECTOR.

antenna duplexer A circuit or device permitting one antenna to be shared by two transmitters without undesirable interaction.

antenna effect The tendency of wires or metallic bodies to act as antennas (i.e., to radiate or receive radio waves).

antenna efficiency The ratio of radio-frequency energy supplied to a wireless transmitting antenna, to the energy radiated into space. Electrically, the radiation resistance of the antenna (R_R) appears in series with loss resistance (R_L). The efficiency Eff of the antenna can be determined by the following formula:

$$Eff = R_R/(R_R + R_L)$$

As a percentage,

$$Eff_{\%} = 100\ (R_R/(R_R + R_L)$$

The efficiency is always less than 1 (100 percent) because, in practice, the loss resistance can never be reduced to zero.

antenna factor A factor (in decibels) added to an RF voltmeter reading to find the true open-circuit voltage induced in an antenna.

antenna field The electromagnetic field immediately surrounding an antenna.

antennafier Low-profile antenna/amplifier device, sometimes used with portable communications systems. Also called an *active antenna.*

antenna front-to-back ratio For a directional antenna, the ratio of field strength in front of the antenna (i.e., directly forward in the line of maximum directivity) to field strength in back of the antenna (i.e., 180 degrees from the front), as measured at a fixed distance from the radiator. It is usually specified in decibels.

antenna gain For a given antenna, the ratio of signal strength (received or transmitted) to that obtained with a comparison antenna, such as a simple dipole. Generally specified in decibels.

antenna ground system The earth, counterpoise, guy wires, radials, and/or various conducting objects in the vicinity of an antenna which, taken together, form the radio-frequency (RF) ground system against which the antenna operates. Some antennas require an extensive ground system to function efficiently; others need no ground system.

antenna/ground system An arrangement embodying both an antenna and a low-resistance connection to the earth, as opposed to an antenna system that involves no connection to earth.

antenna height **1.** The height of an antenna above the surface of the earth immediately beneath the driven element(s). **2.** The height of an antenna above the effective radio-frequency (RF) ground immediately beneath the driven element(s). **3.** The height of an antenna above average terrain, determined against the mean altitude of a number of points on the earth's surface that lie within a certain radius of the antenna structure. Also called *height above average terrain (HAAT).*

antenna impedance The complex-number impedance that an antenna presents to a transmission line. It can vary over a tremendous range, and depends on the antenna type, antenna size, antenna height, operating frequency, and various other factors.

antenna-induced potential Also called *antenna-induced microvolts.* The voltage across the open-circuited terminals of an antenna.

antenna lens Also called *lens antenna.* A radiator designed to focus microwave energy in much the same manner that an optical lens focuses light rays. Lens antennas are made from dielectric materials and/or metals.

antenna loading **1.** The insertion of inductance in antenna elements to lower the resonant frequency of the system without necessarily making the system physically larger or the elements longer. **2.** The insertion of capacitance in antenna elements to raise the resonant frequency of the system without necessarily making the system physically smaller or the elements shorter.

antenna lobe A well-defined region in the radiation pattern of an antenna in which radiation is most intense, or in which reception is strongest. Also see ANTENNA PATTERN.

antenna matching The technique of establishing a satisfactory relationship between the antenna impedance and the transmission-line or transmitter-output impedance, for maximum transfer of power into the antenna. Also, the matching of antenna impedance to receiver-input impedance, for delivery of maximum energy to the receiver.

antennamitter An antenna/oscillator combination that serves as a low-power transmitter.

antenna pattern A polar plot of antenna performance that shows field strength versus angle of azimuth, with the antenna at the center. It is usually specified in the horizontal plane.

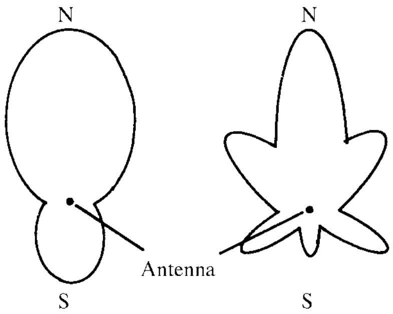

antenna patterns

antenna polarization The orientation of electric lines of flux, with respect to the surface of the earth, for which an antenna is most efficient. A vertical antenna radiates and receives vertically polarized waves. A horizontal antenna radiates and receives horizontally polarized waves broadside to itself, and vertically polarized waves at high elevation angles off its ends. In other directions, the polarization is slanted at various angles.

antenna power Symbol, P_{ant}. The RF power developed in an antenna by a transmitter; P_{ant} equals I^2R, where I is the antenna current and R is the antenna resistance at point I is measured.

antenna power gain The ratio of the maximum effective radiated power (ERP) from a wireless transmitting antenna to the ERP from a reference antenna, expressed in decibels (dB). If the ERP from an antenna under test is P_T watts and the ERP from the reference antenna is P_R watts, then the gain G_{dB} is:

$$G_{dB} = 10 \log_{10} (P_T/P_R)$$

Power gain is always measured in the direction in which the test antenna performs the best. The reference antenna, usually a dipole, is chosen with a gain assumed to be unity, or 0 dB. Gain relative to a dipole is expressed in dBd (decibels relative to a dipole). Alternatively, the reference antenna can be an isotropic radiator, in which case the gain is expressed in dBi (decibels relative to an isotropic radiator). Gain figures in dBd and dBi differ by a constant amount as follows:

$$G_{dBi} = 2.15 + G_{dBd}$$

antenna preamplifier A highly sensitive amplifier used to enhance the gain of a receiver. It is usually used at the very high frequencies and above.

antenna radiation The propagation of radio waves by a transmitting antenna.

antenna radiator The element of an antenna that receives RF energy from the transmitter and radiates waves into space. Also known as the *driven element.* Compare ANTENNA DIRECTOR and ANTENNA REFLECTOR.

antenna range **1.** The frequency band, communication distance characteristically covered, or other continuum of values that specify the operating limits of an antenna. **2.** The region immediately surrounding an antenna in which tests and measurements usually are made. Sometimes called ANTENNA FIELD.

antenna reflector In a directional antenna, a PARASITIC ELEMENT situated behind the radiator and separated from the latter by an appropriate fraction of a wavelength. Its function is to intensify radiation in the direction of transmission. Compare ANTENNA DIRECTOR and ANTENNA RADIATOR.

antenna relay In a radio station, a low-loss, heavy-duty relay that enables the antenna to be switched between transmitter and receiver.

antenna resistance The resistive component of ANTENNA IMPEDANCE.

antenna resonant frequency The frequency, or narrow band of frequencies, at which an antenna's impedance appears resistive.

antenna stage **1.** The first RF amplifier stage of a receiver. **2.** Occasionally, the final RF amplifier of a transmitter.

antenna switch In a radio station, a low-loss, heavy-duty switch that enables the antenna to be connected to transmitter, receiver, or safety ground.

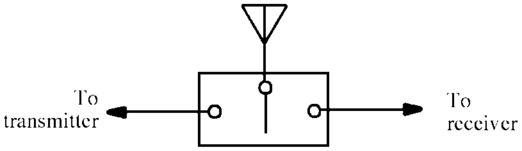

antenna switch
(single-pole, double-throw)

antenna system Collectively, an antenna and all of the auxiliary electrical and mechanical devices needed for its efficient operation, including couplers, tuners, transmission lines, supports, insulators, and rotator.

antenna terminals **1.** The points at which a transmission line is attached to an antenna. **2.** The signal input terminals of a receiver. **3.** The signal output terminals of a transmitter.

antennaverter An antenna and converter combined into a single circuit, intended for connection to the antenna terminals of a receiver to allow operation on frequencies outside the band for which the receiver has been designed.

antenna wire **1.** The radiator element of a wire-type antenna. **2.** A strong solid or stranded wire (e.g., hard-drawn copper, copper-clad steel, or phosphor-bronze) used for antennas.

anthropomorphism The perception, by people, of machines as having human qualities. This can lead to emotional attachment to hardware, such as computers and robots. The more sophisticated the apparatus, in general, the more powerful this perception can become.

antialiasing filter A low-pass or bandpass filter that limits the bandwidth of an input signal to prevent *aliasing* and its effects. See ALIASING, **1**.

anticapacitance switch A switch whose members are thin blades and stiff wires widely separated to minimize capacitance between them.

anticathode The target electrode of an X-ray tube.

Anticipatory Sciences A group of *futurists*, people who attempt to predict the course of technology. Some futurists believe that progress will continue until, for example, homes become fully automated and artificial intelligence reaches a level comparable to human intelligence. Other futurists believe that such things are highly improbable.

anticlutter circuit A supplementary circuit in a radar receiver that minimizes the effect of extraneous reflections that would obscure the image of the target.

anticlutter gain control In a radar receiver, a circuit that automatically raises the gain of the receiver slowly to maximum after each transmitter pulse to reduce the effect of clutter-producing echoes.

anticoincidence Noncoincidental occurrence of two or more signals. Compare COINCIDENCE.

anticoincidence circuit In computers and control systems, a circuit that delivers an output signal only when two or more input signals are not received simultaneously. Compare COINCIDENCE CIRCUIT. Also see NAND CIRCUIT.

anticoincidence operation An exclusive-OR operation.

anticollision radar A vehicular radar system that is used to minimize the probability of a collision with another vehicle, whether or not that other vehicle has a similar system.

antiferroelectric **1.** Pertaining to the property wherein the polarization curve of certain crystalline materials shows two regions of symmetry. **2.** A material that exhibits the aforementioned property.

antiferromagnetic Pertaining to the behavior of materials in which, at low temperatures, the magnetic moments of adjacent atoms point in opposite directions.

antihunt The condition in which *hunting* is counteracted, usually by removing overcorrection in automatic control or compensation systems.

antihunt circuit **1.** A circuit that minimizes or eliminates hunting. Also see ANTIHUNT. **2.** In a television (TV) receiver, a circuit that stabilizes an automatic frequency control (afc) system.

antijamming Pertaining to communications systems that are resistant to, or that counteract, the effects of *jamming.*

antilogarithm Abbreviated, antilog or $\log^{-1}$. The number corresponding to a given logarithm. For example, $\log 10{,}000 = \log 10^4 = 4$, and thus antilog $4 = 10^4 = 10{,}000$.

antilogous pole In a PYROELECTRIC MATERIAL, the end that becomes negatively charged as the temperature rises.

antimagnetic Pertaining to materials having extremely low RETENTIVITY.

antimatter Pertaining to particles that are the counterparts of conventional particles (i.e., positrons instead of electrons, antineutrons instead of neutrons, and antiprotons instead of protons). When a particle meets its antiparticle, the two annihilate, releasing energy. Also see ANTIPARTICLE.

antimicrophonic See NONMICROPHONIC.

antimony Symbol, Sb. A metalloidal element. Atomic number, 51. Atomic weight, 121.76. Often used as n-type dopant in semiconductor manufacture.

antineutrino The antiparticle of the NEUTRINO, emitted as a result of radioactive decay.

antineutron An uncharged particle with a mass equal to that of the neutron, but with a magnetic moment in the direction opposite that of the neutron.

antinode A point of maximum amplitude in a standing wave.

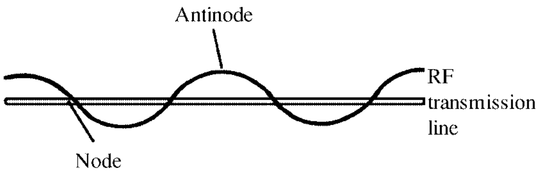

antinode

antinoise carrier-operated circuit A circuit that cuts off the audio output of a receiver while the station transmitter is in use. This can be accomplished in the automatic-gain-control (AGC) circuit of the receiver, or in the speaker or audio line. The circuit is actuated by energy from the transmitted signal.

antinoise microphone Any microphone that discriminates against acoustic noise (e.g., a *lip microphone* or *throat microphone*).

antinucleon A particle with the mass of a nucleon, but with the opposite electrical charge and direction of magnetic moment. Compare NUCLEON.

antioxidant A material, such as a lacquer coat or an inactive oxide layer, that prevents or slows oxidation of a material exposed to air.

antiparticle A subatomic particle opposite in character to conventional particles, such as electrons, neutrons, protons. Antiparticles constitute *antimatter.* Also see ANTINEUTRINO, ANTINEUTRON, ANTINUCLEON, ANTIPROTON, and POSITRON.

antiphase The property of being in phase opposition (180 degrees out of phase).

antipincushioning magnets In some television (TV) receivers, a pair of corrective magnets in the deflection assembly on the picture tube that eliminate *pincushion distortion* (disfigurement of the raster so that it resembles a pincushion—a rectangle with its sides bowed in).

antiproton A subatomic particle with a mass equal to that of the proton, but with opposite electrical charge.

antiquark An ANTIPARTICLE of a QUARK.

antirad substance A material that protects against damage caused by atomic radiation.

antiresonance **1.** Parallel resonance. **2.** The condition of being detuned from a resonant frequency.

antiresonant circuit See PARALLEL-RESONANT CIRCUIT.

antiresonant frequency **1.** The resonant frequency of a parallel-resonant circuit. **2.** In a piezoelectric crystal, the frequency at which impedance is maximum (as in a parallel-resonant circuit).

antisidetone Pertaining to the elimination in telephone circuits of interference between the microphone and earphone of the same telephone.

antistickoff voltage The low voltage applied to the coarse synchro control transformer rotor winding in a dual-speed servo system to eliminate ambiguous behavior in the system.

antitransmit/receive switch Abbreviated *ATR.* In a radar installation, an automatic device to prevent interaction between transmitter and receiver.

antivirus program A computer program or utility designed to detect and eliminate viruses and Trojan horses in a computer system.

antivoice-operated transmission Radio communications that use a voice-activated circuit as a transmitter interlock during reception on the companion receiver.

apc **1.** Abbreviation of *automatic picture control.* **2.** Abbreviation of AUTOMATIC PHASE CONTROL.

aperiodic Characterized by a lack of predictable repetitive behavior. For example, the *sferics* or "static" electromagnetic interference caused by lightning.

aperiodic current The unidirectional current that follows an electromagnetic disturbance in an LCR circuit, in which *R* is equal to or higher than the critical circuit resistance.

aperiodic damping Damping of such a high degree that the damped system, after disturbance, comes to rest without oscillation or hunting.

aperiodic discharge A discharge in which current flowing in an LCR circuit is unidirectional, rather than oscillatory. For this condition, $1/LC$ is less than or equal to $R^2/4L^2$.

aperiodic function A nonrepetitive function (e.g., a hyperbolic trigonometric function).

aperture **1.** The larger, normally open end of a horn antenna or horn loudspeaker. **2.** An opening in an opaque disk or mask that passes a predetermined amount of light or other radiant energy. **3.** The portion of a directional antenna through which most of the radiated energy passes.

aperture angle For an antenna or telescope or microscope, the half angle formed by the radius of the detecting instrument, as viewed from the source.

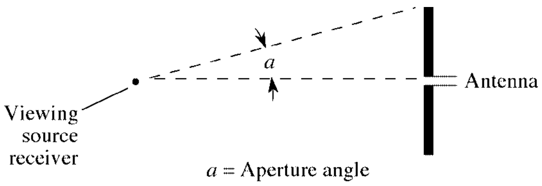

aperture angle

aperture antenna An antenna whose beamwidth depends on the size of a horn, reflector, or lens.

aperture compensation In a television (TV) camera, the minimizing of APERTURE DISTORTION by widening the video-amplifier passband.

aperture distortion In a television (TV) camera tube, a form of distortion that occurs when the scanning beam covers several mosaic elements simultaneously. This condition, caused by excessive beam thickness, results in poor image resolution.

aperture mask In a three-gun color picture tube, a thin, perforated sheet mounted behind the viewing screen to ensure that a particular color phosphor will be excited only by the beam for that color. Also called *shadow mask.*

aperture synthesis In telescopes, a method of obtaining high resolution using several small antennas separated by great distances. The small antennas are moved around to simulate the resolving power of a much larger antenna that would, in practice, be impossible or impractical to construct.

aphelion **1.** The point at which a solar-orbiting satellite attains its highest altitude. It occurs once for every complete orbit. At this point, the satellite travels slower than at any other point in the orbit. **2.** The altitude, measured from the sun's surface or the sun's center, of a solar-orbiting satellite at its most distant point.

APL Abbreviation for *A Programming Language.* A high-level computer language designed for ease of use, and characterized by the requirement for a special character set.

apl **1.** Abbreviation of *average picture level.* **2.** Abbreviation of *automatic phase lock.*

A plus Also, A+. The positive terminal of an A battery. Also, pertaining to the part of a circuit connected to that terminal.

apogee **1.** The point at which an earth-orbiting satellite attains its highest altitude. It occurs once for every complete orbit. At this point, the satellite travels slower than at any other point in the orbit. **2.** The altitude, measured from the earth's surface or the earth's center, of an earth-orbiting satellite at its most distant point.

A power supply A term sometimes used to denote the unit that supplies energy to a vacuum-tube filament. Compare B POWER SUPPLY.

apparent bearing In radio-direction finding, the uncorrected direction from which a signal appears to arrive.

apparent power In an ac circuit, the power value obtained by multiplying the current by voltage (*P* equals *IE*), with no consideration of the effects of phase angle. Compare TRUE POWER.

apparent power loss The loss in an ammeter or voltmeter, caused by the imperfection of the instrument. At full scale, the ammeter has a certain voltage across its terminals; the apparent power loss is the current multiplied by this voltage. A voltmeter carries a small current; the apparent power loss is the product of the current and the indicated voltage.

appearance potential The potential through which an electron must move to produce a certain ion from the atom with which it is associated.

applause meter An instrument consisting essentially of a microphone, audio amplifier, and indicating meter (reading directly in sound level). It is so called because of its familiar use in measuring audience response, as indicated by loudness of applause.

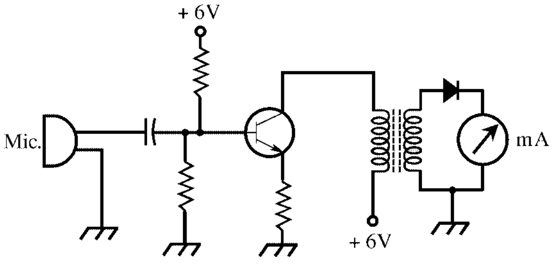

applause meter

Applegate diagram For a velocity-modulated tube, a plot of the positions of electron bunches in the drift space versus time.

Appleton layer Collectively, the F1 and F2 layers of the ionosphere, at a height between 150 and 400 kilometers above the surface of the earth.

apple tube A color picture tube, used in television, with the red, blue, and green phosphor in vertical strips.

appliance Electrical equipment in general. This might include any home-operated device.

application A task or job for which an electronic device or system is used. It especially pertains to personal-computer software that has practical usefulness.

application factor A factor involved in determining the failure rate of a circuit or system affected by unusual operating conditions.

application schematic diagram A diagram of pictorial symbols and lines that illustrate the interrelationship of functional circuit blocks in a specific program mode.

applicators **1.** In dielectric heating, the electrodes between which the dielectric body is placed and the electrostatic field developed. **2.** In medical electronics, the electrodes applied to a patient undergoing diathermy or ultrasonic therapy.

applied voltage The voltage presented to a circuit point or system input, as opposed to the voltage drop resulting from current flow through an element.

applique circuit A circuit for adapting equipment to a specialized job.

approach-control radar A radar installation serving a ground-controlled approach (GCA) system.

approximate data **1.** Data obtained through physical measurements. Such data can never be exact; all measurements are subject to error. **2.** Loosely estimated data or imprecise calculations.

AQL Abbreviation of ACCEPTABLE QUALITY LEVEL. A statistically defined quality level, defined in terms of percent defective, accepted on an average of 95 percent of the time.

Aquadag A tradename for a material that consists of a slurry of fine particles of graphite. Aquadag forms a conductive coating on the inside and outside walls of some cathode-ray tubes.

aqua pura Pure water; in most instances, distilled water. Formula, H_2O. Pure water is a nonconductor with a dielectric constant of about 81.

Ar Symbol for ARGON.

arbitrary function fitter A circuit or device, such as a potentiometer, curve changer, or analog computer element, providing an output current or voltage that is some preselected function of the input current or voltage.

arc **1.** A luminous sustained discharge between two electrodes. Because it is sustained, rather than intermittent, an arc is distinguished from a spark discharge, the latter being a series of discharges (sparks)—even when it appears continuous. **2.** In graphical presentations, a section of curved line, as of a circle.

arc angle The angle in degrees traced out by a circular arc if the center point of the circle is considered to be the vertex of an angle formed by two rays intersecting the arc at designated points.

arc cosecant Abbreviated arc csc or $\csc^{-1}$. **1.** The inverse of the *cosecant* function. **2.** The angle, in

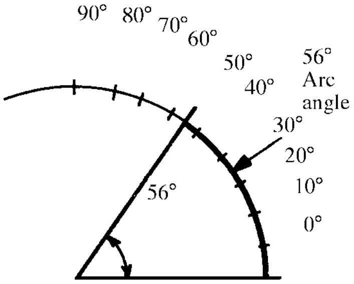

arc angle

radians or degrees, corresponding to a given cosecant.

arc cosine Abbreviated arc cos or cos^{-1}. **1.** The inverse of the *cosine* function. **2.** The angle, in radians or degrees, corresponding to a given cosine.

arc cotangent Abbreviated arc cot or cot^{-1}. **1.** The inverse of the *cotangent* function. **2.** The angle, in radians or degrees, corresponding to a given cotangent.

arc failure **1.** Damage to, and/or failure of, insulation or a dielectric as a result of ARCOVER. **2.** Failure of make-and-break contacts through damage caused by arcover.

arc function An *inverse trigonometric function.* See ARC COSECANT, ARC COSINE, ARC COTANGENT, ARC SECANT, ARC SINE, and ARC TANGENT.

arc furnace A high-temperature electric furnace in which heat is produced by one or more electrical arcs.

architecture The functional design elements of a computer—especially the components of the *central processing unit (CPU)* and the manner in which these elements interact.

archived file A computer file stored on some backup medium, such as magnetic tape, diskette, or CD-ROM (compact disk, read-only memory), rather than being held on the hard disk. Such a file will be apart from the operating system's catalog of current files, but can be reconstituted as needed.

archives A complete, periodically updated set of ARCHIVED FILES.

arcing See ARCOVER.

arcing contacts Make-and-break contacts between which an arc occurs when they are separated.

arcing ring A metal ring placed around an insulator in a high-voltage electrical system. This keeps an arc from charring or breaking the insulator.

arcing time The elapsed time between the breaking of contacts and the end of the arc between the contacts.

arc lamp An electric lamp in which a brilliant arc jumps between the tips of two rods (originally carbon).

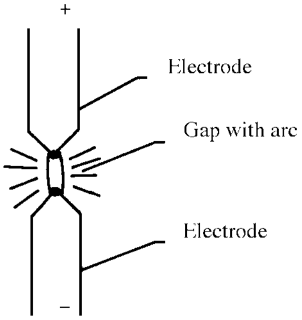

arc lamp

arc length The length along a given arc, usually a part of the circumference of a circle. If the circle has circumference C and the arc measures x degrees, then the arc length is $Cx/360$ units.

arc minute See MINUTE.

arc oscillation Oscillations that can occur when opening relay contacts arc.

arcover The occurrence of an electrical ARC between electrodes, contacts, or capacitive plates.

arcover voltage The voltage at which disruptive discharge occurs, typically accompanied by an arc.

arc resistance The ability of a material, usually a dielectric, to resist damage from arcing. This property is commonly expressed as the length of time between the start of the arc and the establishment of a conductive path through the material.

arc secant Abbreviated arc sec or sec^{-1}. **1.** The inverse of the *secant* function. **2.** The angle, in radians or degrees, that corresponds to a given secant.

arc second See SECOND.

arc sensor A device for detecting visible arcs and excessive reflected power in microwave systems.

arc sine Abbreviated arc sin or sin^{-1}. **1.** The inverse of the *sine* function. **2.** The angle, in radians or degrees, that corresponds to a given sine.

arc suppression Extinguishing an arc discharge. Disruptive arcs in electronic circuits are suppressed by means of auxiliary diodes or resistor-capacitor networks.

arc-suppressor diode A semiconductor diode used to prevent arcing between make-and-break contacts.

arc tangent Abbreviated arc tan or tan^{-1}. **1.** The inverse of the *tangent* function. **2.** The angle, in radians or degrees, corresponding to a given tangent.

arcthrough The puncturing of a material by an arc.

area code In the United States, a three-digit number that indicates the location, according to specified assigned districts, of a telephone subscriber. When making a long-distance call, the area code of the desired station must be given in addition to the seven-digit telephone number.

area protection Coverage of a defined region, in terms of area or volume, by an alarm system.

area redistribution A scheme to determine the effective duration of an irregularly shaped pulse. A rectangle is constructed whose height is equal to the peak height of the pulse, as displayed on an oscilloscope. The rectangle width is adjusted until the area of the rectangle is the same as the area under the curve representing the pulse. The width of the rectangle then represents the effective duration of the pulse.

area search The scanning of a large group of computer records for those of a major category or class.

area sensor A transducer, used with an alarm system, that protects a defined region or volume, such as an office or bedroom.

Argand diagram Named after Jean Robert Argand, (1768-1822) of Geneva, for his work on the graphical representation of complex numbers. A graphical illustration of a complex number in the form *A* + *jB*, where the real-number (*A*) axis is perpendicular to the imaginary-number (*jB*) axis. The value *j* is the square root of -1, the unit imaginary number. The axes are perpendicular, usually with the *A* axis horizontal. The length of the line from the point (0,0) to the point (*A*,*jB*) is the amplitude of the vector **X** = *A* + *jB*. The direction is specified as the angle, in degrees or radians, of the vector measured counterclockwise from the *A* axis.

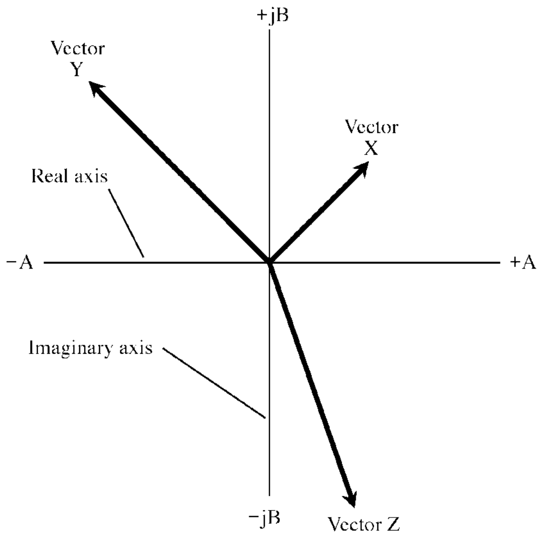

Argand diagram

argon Symbol, Ar. An inert gaseous element. Atomic number, 18. Atomic weight, 39.94. Argon, present in small amounts in the earth's atmosphere, is used in various specialized devices, such as lasers.

argon laser A laser whose tube is filled with argon gas. It generates coherent light at specific wavelengths that are characteristic of elemental argon.

argument **1.** The direction angle of a polar vector. **2.** An independent variable whose value determines the value of a function.

arithmetic address An address obtained by performing an arithmetic operation on another address.

arithmetic and logic unit Abbreviation, ALU. The part of a digital computer containing the circuits that perform calculations and logic operations; distinguished from mass storage, input/output, and peripheral units.

arithmetic circuit Also called *arithmetic element.* In a digital computer, a circuit that is involved in the execution of calculations. Included are adders, storage registers, accumulators, subtracters, and multipliers.

arithmetic mean The average of a group of quantities, obtained by dividing their sum by the number of quantities.

arithmetic operation In digital computer practice, a numerical process performed: addition, subtraction, multiplication, division, comparison.

arithmetic progression A mathematical series in which each term following the first is obtained by adding a constant quantity to the preceding one. For example, *S* = 1, 2, 3, 4, . . . *n*. Compare GEOMETRIC PROGRESSION.

arithmetic shift In a digital computer, the multiplication or division of a quantity by a power of the base used in the notation.

arithmetic sum The sum of two or more quantities disregarding their signs. Compare ALGEBRAIC SUM.

arithmetic symmetry A filter response that is exactly symmetrical about the center frequency when the frequency scale is linear.

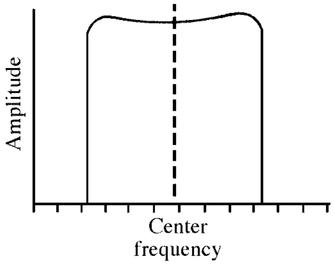

arithmetic symmetry

arm **1.** Any of the distinct branches of a circuit or network. Also called *leg.* **2.** A movable element in a device, usually containing a contact for switching.

armature **1.** The rotating member of a motor. **2.** The rotating member of some types of electro-mechanical generator. **3.** The movable member of a relay, bell, buzzer, or gong. **4.** The movable member of an actuator. **5.** The soft-iron *keeper* placed across the poles of a permanent magnet to conserve power.

armature coil A coil of insulated wire wound on a ferromagnetic core to provide the electromagnetic properties of an armature. In a motor or generator, the armature coil is distinguished from the FIELD COIL.

armature core The ferromagnetic core upon which the armature coil of a motor or generator is wound.

armature gap **1.** In a motor or generator, the space between an armature core and the pole of a field magnet. **2.** In a relay, the space between the armature and the relay-coil core.

armature hesitation A momentary delay in the movement of a relay.

armature-hesitation contact chatter Undesired (usually rapid, repetitive) making and breaking of relay contacts. Generally caused by armature hesitation.

armature-impact contact chatter Undesired (usually rapid, repetitive) making and breaking of relay contacts, caused by *contact bounce* when the armature strikes the relay core (closure) or backstop (opening).

armature relay A relay that uses an electromagnet to pull a lever toward or away from a set of fixed contacts.

armature travel The distance traveled by an armature during relay operation.

armor A protective metal cable covering.

Armstrong FM system (Edwin H. Armstrong, 1890–1954). A phase-shift method of frequency modulation. See PHASE MODULATION.

armature voltage control A means of controlling motor speed by changing the applied armature winding voltage.

armchair copy An amateur radio term for reception of exceptionally clear signals.

arming the oscilloscope sweep Enabling an oscilloscope to trigger on the next pulse by closing a switch.

Armstrong oscillator (Edwin H. Armstrong, 1890–1954). An oscillator circuit that uses inductive feedback between the output and input. Either the output coil or the input coil can be tuned to set the oscillator frequency. The amount of positive feedback is controlled by varying the coupling between the coils.

Armstrong superheterodyne circuit See SUPERHETERODYNE CIRCUIT.

Armstrong superregenerative circuit See SUPERREGENERATIVE CIRCUIT.

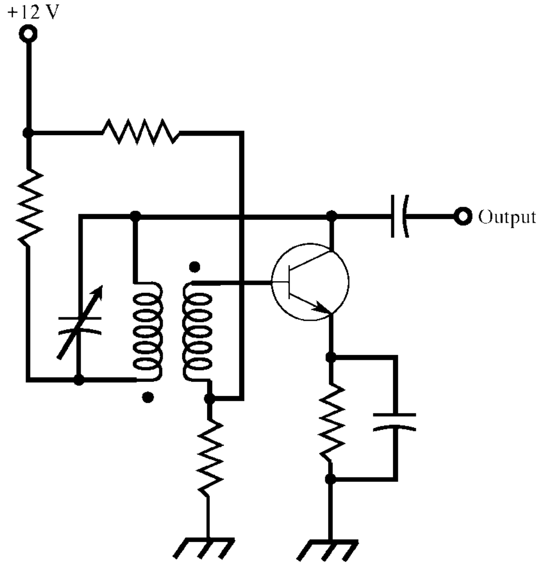

Armstrong oscillator

ARPA Acronym for *Advanced Research Projects Agency,* a subsidiary of the U.S. Department of Defense.

array **1.** A directive antenna that consists of an assembly of properly dimensioned and spaced elements, such as radiators, directors, and reflectors. **2.** A coordinated group or matrix of components, such as diodes, resistors, memory cells, etc., often enclosed in one capsule. **3.** Subscripted variables representing data arranged so that a program can examine the array and extract data relevant to a particular subscript.

array device A group of similar or identical components that are connected together in a certain fashion, to perform a specific task.

arrester **1.** A device used to protect an installation from lightning. It consists of a varistor or an air gap connected between an antenna or power line and an earth ground. The device passes little or no current under ordinary conditions, but passes heavy current to ground during a lightning stroke. Also called LIGHTNING ARRESTER. **2.** A self-restoring protective device used to reduce voltage surges on power lines.

ARRL Abbreviation for *American Radio Relay League.*

arrowhead A wideband, log-periodic antenna with linear polarization.

ARS Abbreviation of *Amateur Radio Service.*

arsenic Symbol, As. A metalloidal element. Atomic number, 33. Atomic weight, 74.91. Arsenic is familiar as an n-type dopant in semiconductor processing.

ARSR Abbreviation of *air route surveillance radar.*

articulation A measure of the effectiveness of voice communications, expressed as the percentage of speech units understood by the listener when the effect of context is negligible.

artificial antenna See DUMMY ANTENNA.

artificial ear A microphone-type sensor, equivalent to the human ear, used to measure sound pressures.

artificial echo **1.** In radar practice, the reflections of a transmitted pulse returned by an artificial target. **2.** A signal from a pulsed radio-frequency (RF) generator, delayed to simulate an echo.

artificial ground The effective ground provided by the radials or disk of a ground-plane antenna, as opposed to actual ground (the earth itself). Compare TRUE GROUND.

artificial horizon In aircraft instrumentation, a device that displays lines showing the position of the aircraft in flight, with reference to the horizon.

artificial intelligence Abbreviation, AI. **1.** A specialized field of computer science overlapping with electronics, biology, physiology, and other sciences, concerning attempts to develop advanced computer systems that can emulate the processes of the human mind. **2.** The ability of a computer to learn from its mistakes, refine its own processes, and perhaps ultimately reason in a humanlike manner.

artificial ionization An artificial reflecting layer that is created in the atmosphere to provide a skip condition.

artificial language A language that is not commonly used, but has been devised for efficiency in a particular situation—especially in a computer system.

artificial life **1.** The ultimate endpoint of ARTIFICIAL INTELLIGENCE, wherein machines acquire qualities, such as wisdom and the capability to feel emotions. The state of the art is currently nowhere near this point. **2.** A hypothetical machine or set of machines with lifelike qualities, including human-level intelligence, wisdom, and emotion.

artificial stimulus An electronic method of robot guidance and navigation using radar, sonar, vision systems, edge detection, and/or beacons.

artificial transmission line A network of capacitors and inductors with characteristics similar to those of the more bulky transmission line it replaces in tests and measurements. It also serves as a time-delay or phase-shift device and as a pulse-forming network.

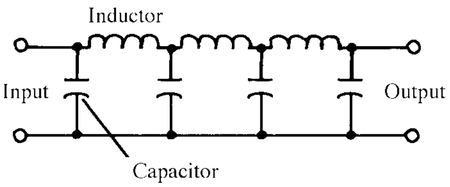

artificial transmission line

artificial voice A device used to test and calibrate noise-canceling microphones, consisting essentially of a small loudspeaker that has a baffle whose acoustical properties simulate those of the human head.

artos stripper A machine that cuts and strips wire for the fabrication of multiconductor cables.

artwork **1.** In the manufacture of printed circuits, the scaled drawings from which the mask or etch pattern is obtained photographically. **2.** Collectively, the illustrations depicting an electronic circuit, device, or system.

As Symbol for ARSENIC.

ASA Abbreviation of AMERICAN STANDARDS ASSOCIATION.

asbestos A nonflammable fibrous material consisting of calcium and magnesium silicates that is used for high-temperature insulation.

A-scan A radar-screen presentation in which the horizontal time axis displays distance or range, and the vertical axis displays the amplitude of signal pulse and echo pulses.

ascending node For a satellite orbiting the earth or another planet, any point at which the groundtrack crosses the equator as it moves from the southern hemisphere into the northern hemisphere. This node generally changes for each succeeding orbit, because the earth or planet rotates underneath the orbit of the satellite. Compare DESCENDING NODE.

ascending pass For a specific point on the earth's surface, the time during which an artificial communications satellite is accessible when its latitude is moving northward. The duration of accessibility depends on the altitude of the satellite, and on how close its groundtrack comes to the earth-based point. Compare DESCENDING PASS.

ASCII Acronym (pronounced "*ask*-ee") for *American Standard Code for Information Interchange.*

ASI Abbreviation for *American Standards Institute.*

A-scope A radar system that displays an A-SCAN.

Askarel A synthetic, nonflammable liquid dielectric.

aspect ratio The width-to-height ratio of a video image, generally three units high by four units wide.

asperities On the surface of an electrode, tiny points at which the electric field is intensified and from which discharge is highly probable.

ASR **1.** Abbreviation of AIRBORNE (or AIRPORT) surveillance radar. **2.** Abbreviation of AUTOMATIC SEND/ RECEIVE.

ASRA Acronym for *automatic stereophonic recording amplifier.*

assemble **1.** To gather subprograms into a complete digital computer program. **2.** To translate a

Symbols for ASCII teleprinter code

First four signals	*Last three signals* 000	*001*	*010*	*011*	*100*	*101*	*110*	*111*
0000	NUL	DLE	SPC	0		P	/	p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EOT	DC4	$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	'	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	,	<	L	/	l	/
1101	CR	GS	–	=	M	]	m	}
1110	SO	RS	.	>	N		n	~
1111	SI	US	|	?	O	-	o	DEL

ACK: acknowledge
BEL: bell
BS: back space
CAN: cancel
CR: carriage return
DC1: device control no. 1
DC2: device control no. 2
DC3: device control no. 3
DC4: device control no. 4
DEL: delete
DLE: data link escape
ENQ: enquiry
EM: end of medium
EOT: end of transmission
ESC: escape
ETB: end of transmission block
ETX: end of text

FF: form feed
FS: file separator
GS: group separator
HT: horizontal tab
LF: line feed
NAK: do not acknowledge
NUL: null
RS: record separator
SI: shift in
SO: shift out
SOH: start of heading
SPC: space
STX: start of text
SUB: substitute
SYN: synchronous idle
US: unit separator
VT: vertical tab

symbolic program language into a machine (binary) language program by substituting operation codes and addresses.

assembly **1.** A finished unit that can be either a practical working model or a dummy, a prototype, or a final model; an integrated aggregation of subunits. **2.** A low-level computer source-code language that uses crude mnemonics that are easier to remember than the *machine-language* equivalents.

assembly language A source code that uses mnemonic instructions. (See ASSEMBLY, **2.**)

assembly program The program that operates on a symbolic-language program to produce a machine language program in the process of assembly. Also called *assembler.*

assembly robot A form of *industrial robot* that puts hardware together. Such a robot is generally a component of an *automated integrated manufacturing system (AIMS).* The robot can do repetitive work at high speed and precision for long periods of time.

assign To reserve part of a computing system for a specific purpose, normally for the duration of a program run.

assigned frequency The radio carrier frequency or band of frequencies designated for a transmitting station by a licensing authority. Also see RADIO SPECTRUM.

associative memory Computer memory in which locations are identified by content, rather than by specific address.

assumed decimal point A decimal point that does not occupy an actual computer storage space, but is used by the computer to align values for calculation; the decimal point is assumed to be at the right unless otherwise specified.

astable Having two temporary states; BISTABLE.

astable circuit A circuit that has two unstable states, and whose operation is characterized by alternation between those states at a frequency determined by the circuit constants.

astable multivibrator A free-running multivibrator. The common circuit uses two bipolar or field-effect transistors, their inputs and outputs being cross coupled. Conduction switches alternately between the two.

astatic **1.** Without fixed position or direction. **2.** In a state of neutral equilibrium.

astatic galvanometer A galvanometer with a movable element consisting of two identical magnetized needles mounted nonparallel on the same suspension. Each needle is surrounded by a coil. The coils are wound in opposite directions, and are connected in series to the current source. A large permanent magnet provides the field against which the needle assembly rotates. The instrument functions independently of the geomagnetic field.

astatine Symbol, At. A radioactive elemental halogen produced from radioactive decay. Atomic number, 85. Atomic weight, 210. Formerly called *alabamine.*

A station One of the two stations in the transmitting system of LORAN (long-range navigation).

astigmatism A focusing fault in a cathode-ray tube (CRT), in which electrons in different axial planes focus at different points.

ASTM Abbreviation for *American Society for Testing and Materials.*

astrionics The design, production, and application of electronic devices and systems for use in space vehicles and space navigation.

astronomical unit Abbreviation, AU. A unit of distance equal to 1.496×10^8 kilometers (9.296×10^7 miles). Approximately equal to the mean distance between the earth and the sun.

A supply See A POWER SUPPLY.

asymmetrical cell A photocell exhibiting ASYMMETRICAL CONDUCTIVITY.

asymmetrical communications **1.** Two-way communications in which the volume of transmitted data is much greater in one direction than in the other. **2.** Two-way communications in which the speed of transmitted data is much greater in one direction than in the other. Compare SYMMETRICAL COMMUNICATIONS.

asymmetrical conductivity A condition in which a device conducts well in one direction, but poorly in the other direction. A rectifier diode is a common example of a component that exhibits this effect.

asymmetrical distortion In a binary system, lengthening or shortening of one of the states, by comparison to the theoretical or ideal duration.

asymmetrical FET A FIELD-EFFECT TRANSISTOR in which the source and drain cannot be interchanged without degrading performance.

asymmetrical multivibrator An unbalanced multivibrator (i.e., one in which the circuit halves are not identical). If the time constants of the halves are different, the output pulses will be short and widely separated.

asymmetrical sideband See VESTIGIAL SIDEBAND.

asymmetrical sideband transmission See VESTIGIAL SIDEBAND TRANSMISSION.

asymmetrical wave A wave whose upper (positive half-cycle) and lower (negative half-cycle) portions have different amplitudes or shapes. Also called *asymmetric wave.*

asymmetry control An adjustment in a device intended for measuring the pH (acidity/alkalinity). This corrects the inaccuracies that results from the differences between the electrodes.

asymptote In analytical geometry, a fixed straight line or ray L with a special relationship to a curve or part of a curve K that recedes to infinity. As the distance from the origin (0,0) increases without limit, the separation between K and L approaches zero, but K and L never actually meet.

asymptotic breakdown voltage A voltage that will cause *dielectric breakdown* if applied continuously for a sufficiently long time.

asymptotic expression An expression having a very small error in terms of percentage.

asynchronous **1.** Not synchronous, i.e., nonrecurrent (as in out-of-phase waves). **2.** A mode of computer operation in which the completion of one operation starts another.

asynchronous device A device not regulated by the system in which it is used, as far as its operating frequency or rate is concerned.

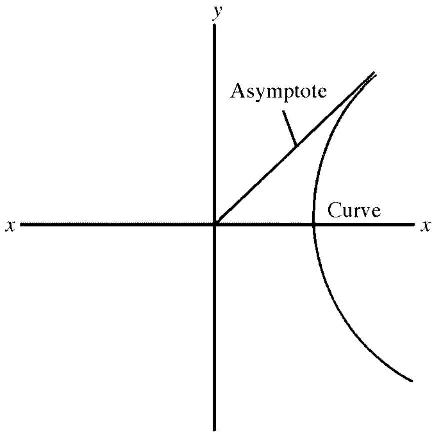

asymptote

asynchronous input In digital circuitry, any flip-flop input at which a pulse can affect the output independently of the clock.

asychronous motor An ac motor whose speed is not proportional to the supply frequency.

asynchronous transmission Data transmission in which each character or symbol begins with a start signal and ends with a stop signal. This eliminates the need for the data to be sent at a uniform speed.

asynchronous vibrator In a vibrator-type portable power supply, a vibrator that only makes and breaks the primary circuit of the step-up transformer. This is in contrast to the synchronous vibrator, which also makes and breaks the secondary circuit in synchronism with the primary. Also called NONSYNCHRONOUS VIBRATOR.

AT A quartz crystal cut wherein the angle between the x-axis and the crystal face is 35 degrees.

At Symbol for astatine.

AT-cut crystal A piezoelectric crystal cut at a 35-degree angle, with respect to the optical axis of the quartz. The frequency of such a crystal does not appreciably change with variations in temperature.

atmosphere **1.** The gas surrounding a planet, particularly the air sheathing the earth. **2.** Abbreviation, atm. A unit of pressure equal to 1.013×10^6 dynes per square centimeter (about 14.7 pounds per square inch).

atmospheric absorption **1.** The conversion of electromagnetic energy into heat, with resulting loss, as the energy passes through the earth's atmosphere. The extent of this effect depends on the wavelength. **2.** See ABSORPTION LOSS, **2.**

atmospheric absorption noise Noise, principally above 1 GHz, resulting from atmospheric absorption (see ABSORPTION LOSS, **2**).

atmospheric bending The refraction or reflection of electromagnetic waves by the troposphere or ionosphere. See ATMOSPHERIC REFLECTION.

atmospheric duct A tropospheric stratum, often associated with temperature inversions, lake effects, or weather fronts, through which electromagnetic energy at ultra-high and microwave frequencies is efficiently propagated for long distances.

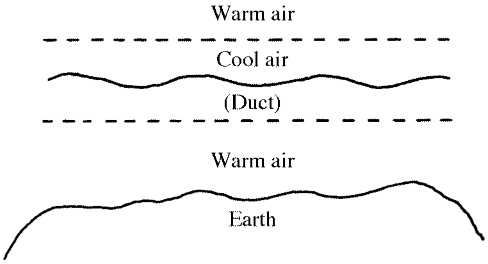

atmospheric duct

atmospheric electricity Static electricity present in the atmosphere, which evidences itself in disturbance of radio communications and in displays of lightning.

atmospheric noise Receiver noise resulting from ATMOSPHERIC ELECTRICITY. Also called *sferics* or *static.*

atmospheric pressure Abbreviation, atm press. **1.** The pressure exerted by the earth's atmosphere, as indicated by a barometer at sea level; normally between 29 and 31 inches of mercury. **2.** A pressure of 1.013×10^6 dynes per square centimeter. See ATMOSPHERE, **2.**

atmospheric radio wave See SKYWAVE.

atmospheric radio window The band of frequencies (approximately 10 MHz to 10 GHz), including radio waves that can penetrate the earth's troposphere and ionosphere.

atmospheric reflection The return of a radio wave to earth, resulting from reflection by an ionized portion of the atmosphere.

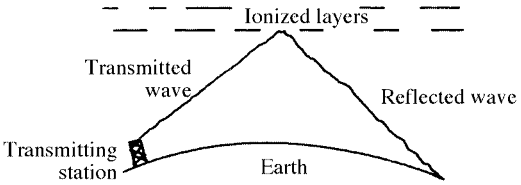

atmospheric reflection

atmospheric refraction **1.** Downward bending of radio waves as a result of variations in the dielectric constant of the troposphere. **2.** Downward bending of radio waves in the ionosphere, resulting in long-range propagation at high frequencies.

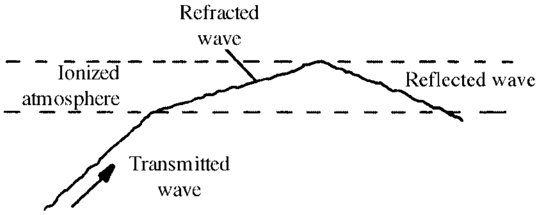

atmospheric refraction

atmospheric scatter **1.** The scattering of very-high frequency (VHF) and ultra-high frequency (UHF) radio waves by the lower atmosphere. **2.** Communication via scattering of VHF and UHF radio waves in the lower atmosphere.

atmospherics See ATMOSPHERIC NOISE.

atom **1.** The smallest material particle that displays the unique characteristics and properties of an element. Atoms consist of a dense, positively charged central nucleus, around which less-massive, negatively charged electrons "swarm" at definite levels called *shells.* Also see BOHR ATOM and RUTHERFORD ATOM. **2.** In a computer-compiling operation, an operator or operand.

atomechanics The physics of electron movement.

atomic battery A battery in which atomic energy is converted into electrical energy.

atomic charge The electrification (i.e., the electron charge) exhibited by an ion.

atomic clock Also called *atomic time standard.* A highly accurate electronic clock, driven by the characteristic oscillations of certain atoms.

atomic energy Energy released by the FUSION or FISSION of atomic nuclei. Also see ATOMIC POWER.

atomic fission See FISSION.

atomic frequency The natural vibration frequency of an atom.

atomic fusion See FUSION.

atomic pile See REACTOR, **2.**

atomic mass unit Abbreviated amu. A unit that expresses the relative mass of an elemental *isotope.* One amu is equal to 1/12 of the atomic mass of carbon 12 (C12). A neutron has a mass of roughly one amu.

atomic migration The transfer or "wandering" of a valence electron between or among atoms in a single molecule.

atomic number The number of protons in the nucleus of an atom. Also, the number of electrons if the atom is electrically neutral. For example, the atomic number for copper is 29, indicating 29 protons in the nucleus. An electrically neutral atom of copper has 29 electrons. The atomic number uniquely identifies an element.

atomic radiation The emission of radiant energy by radioactive substances.

atomic reactor See REACTOR, **2.**

atomic theory The scientific theory that all matter is composed ultimately of *atoms*, which are the smallest particles retaining the identity of an element. Atoms combine to form *molecules*, the smallest particles that retain the identity of a compound. Atoms themselves contain minute *subatomic particles*, some of which carry electric charges. See BOHR ATOM and RUTHERFORD ATOM.

atomic time **1.** A means of time determination that makes use of the resonant vibrations of certain substances, such as cesium. **2.** Synchronized astronomical time, as determined by an ATOMIC CLOCK.

atomic unit of energy In a hydrogen atom, the potential energy of the electron in the lowest-energy shell, as averaged over a certain length of time. The shell represents the mean energy of the electron.

atomic weight **1.** The mass of a particular atom in ATOMIC MASS UNITS (amu). **2.** A number characterizing the average mass of individual atoms for a specific isotope of an element. Thus, carbon 12 (C12) has an atomic weight of 12, oxygen 16 (O16) has an atomic weight of approximately 16, and uranium 238 (U238) has an atomic weight of about 238.

atomistics The science of the atom and atomic energy. Also called *atomics.*

attack **1.** The rise of a pulse from zero to maximum amplitude. **2.** The time required for a pulse to rise from zero to maximum amplitude. **3.** The initialization of a circuit voltage or current for a certain purpose, such as an automatic gain control. **4.** The rise of a musical note from zero to full volume.

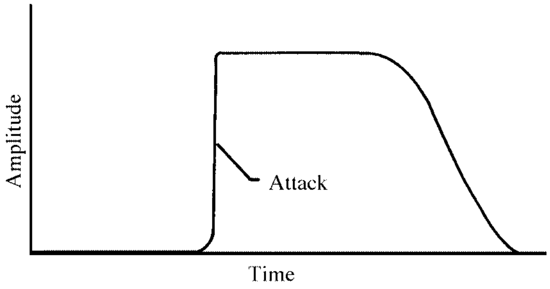

attack

attack time The time required for an applied signal that suddenly increases in amplitude to reach 63.2 percent of its final, stable value.

attemperator An automatic temperature-controlling device; a *thermostat.*

attention display A computer-generated chart or graph, displayed as an alert signal concerning a particular situation.

attenuate To reduce in amplitude.

attenuation A reduction of signal amplitude.

attenuation characteristic Also called *attenuation constant.* **1.** In an amplifier, network, or component, the decrease in signal amplitude as a function of frequency, usually expressed in decibels per octave. **2.** In a transmission line, the decrease in signal amplitude per unit length. Usually expressed in decibels per 100 feet, decibels per mile, or decibels per kilometer.

attenuation constant See ATTENUATION CHARACTERISTIC.

attenuation distortion A type of distortion characterized by variation of attenuation with frequency within a given frequency range.

attenuation equalizer An equalizer that stabilizes the transfer impedance between two ports at all frequencies within a specified frequency band.

attenuation-frequency distortion Distortion characterized by the attenuation of the frequency components in a complex waveform. Frequency-sensitive RC networks (such as a Wien bridge) exhibit this type of distortion when they attenuate a fundamental and each harmonic unequally.

attenuation network A combination of components (R, C, or L singly or in any necessary combination) that provide constant signal attenuation with negligible phase shift throughout a frequency band.

attenuation ratio The ratio indicating a relative current, voltage, power or energy decrease. For example, for voltage, $E_{input}/E_{output} = 6/2 = 3:1 = 3$.

attenuator A device for reducing signal amplitude in precise, predetermined steps, or smoothly over a continuous range. A network of resistors, capacitors, or both. The simplest attenuator consists of one or more noninductive resistors.

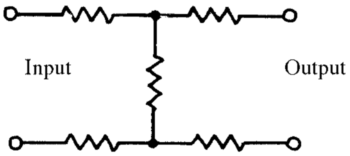

attenuator

attitude The position of an aircraft or space vehicle relative to a (usually terrestrial) reference point, often determined with electronic instruments.

atto- Abbreviated, a. A prefix meaning 10^{-18} or multiplication by 10^{-18}.

attofarad Abbreviation, aF. An extremely small unit of low capacitance; 1 aF equals 10^{-18} F.

attracted-disk electrometer A device to measure potential difference consisting of two parallel metal disks—one of which is connected to a tension spring. The force between the disks indicates the magnitude of the electric field.

attraction The drawing together or pulling toward, as in the attraction between electric charges or magnetic poles. Dissimilar charges and poles attract each other (electric plus to minus, magnetic north to south). Compare REPULSION.

ATV Abbreviation of *amateur television*, used in the Amateur Radio Service.

AU Abbreviation of ASTRONOMICAL UNIT.

Au Symbol for GOLD.

audibility The quality of being detectable by the human ear. In a healthy listener, the *threshold of audibility* is extremely low; at the threshold, the pressure of a sound wave varies from normal by approximately 10^{-4} pascal. The frequency range of human audibility extends roughly from 20 Hz to 20 kHz.

Sound	*Audibility (dB)*
Threshold of hearing	0
Whisper	10–20
Electric fan at 10 feet	30–40
Running water at 10 feet	40–60
Speech at 5 feet	60–70
Vacuum cleaner at 10 feet	70–80
Passing train at 50 feet	80–90
Jet at 1000 feet altitude	90–100
Rock band on stage	110–120
Air hammer at 5 feet	130–140

audibility table

audibility curve A graph (such as the *Fletcher-Munson curve*) that depicts the range of human hearing in terms of frequency versus the sound pressure at the threshold of AUDIBILITY.

audible Detectable by the human ear.

audible alarm device An ANNUNCIATOR that produces an easily identifiable sound in response to an ALARM CONDITION in a security system.

audible frequency See AUDIO FREQUENCY.

audible tone A vibration of air molecules that can be detected by the human ear, and with periodic properties, such as a sine-wave vibration.

audio **1.** Pertaining to the spectrum of frequencies corresponding to the human hearing range (about 20 Hz to 20 kHz), or to equipment or performance associated with that spectrum. **2.** Any disturbance, such as a current or compression wave, falling within the range of about 20 Hz to 20 kHz. **3.** AUDIO FREQUENCY.

audio amplifier See AUDIO-FREQUENCY AMPLIFIER.

audio band The range (band) of audio frequencies.

audio channel **1.** The portion of a complex signal or waveform used to convey audio information exclusively. **2.** The audio-frequency section of a transmitter or receiver (as opposed to the radio-frequency section). **3.** A radio channel of fixed frequency that is reserved for voice communications.

audio clipping Brute-force limiting of the amplitude of an audio signal, usually accomplished using semiconductor diodes to prevent the positive and negative peak amplitudes from exceeding a certain level.

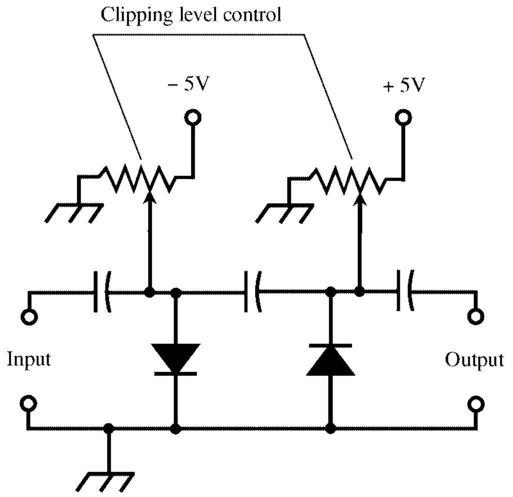

audio clipping

audio component The audio-frequency portion of any wave or signal.

audio converter A circuit in which a received radio-frequency (RF) signal is heterodyned with a local RF oscillator signal to produce an audio-frequency (AF) beat-note output. The beat note is then amplified by an AF amplifier. It is used especially by amateur radio operators in the reception of continuous-wave (CW) radiotelegraphy, radioteletype, and packet radio at high frequencies.

audio frequency A frequency lying within the audible spectrum. Abbreviated AF. See AUDIO-FREQUENCY SPECTRUM.

audio-frequency amplifier An amplifier that operates in part or all of the frequency range 20 Hz to 20 kHz. High-fidelity amplifiers function over a somewhat wider range (e.g., 10 Hz to 50 kHz).

audio-frequency choke An inductor (usually having a ferromagnetic core) that blocks audio-frequency current, but passes direct current.

audio-frequency feedback **1.** Electrical FEEDBACK (positive and/or negative) that affects audio-frequency circuits. **2.** ACOUSTIC FEEDBACK.

audio-frequency filter A filter of any type that operates on any part of the frequency range 20 Hz to 20 kHz.

audio-frequency meter An instrument to measure frequencies in the audio-frequency spectrum (approximately 20 Hz to 20 kHz). Three types are commonly used:

- ***Analog*** Gives direct indications of frequency on the scale of a D'Arsonval meter; the usual range is 20 Hz to 100 kHz.
- ***Digital*** Gives direct indications of frequency by means of readout lamps; the usual range is 1 Hz to 15 MHz. This instrument is useful also as a radio-frequency meter.
- ***Bridge*** Consists of a frequency-sensitive bridge, such as a Wien bridge, with a null-indicating meter. The operator balances the bridge and reads the unknown frequency from the dial of the balance control.

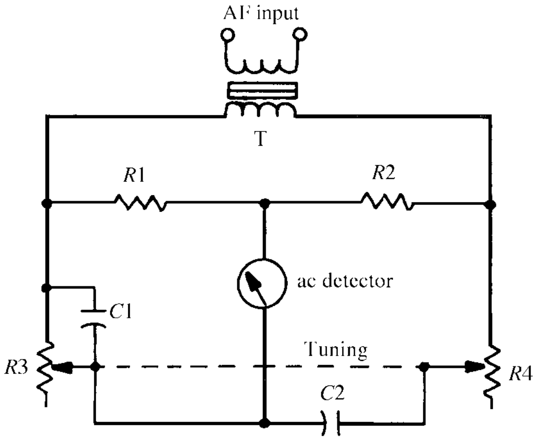

$C1 = C2$, $R3 = R4$
At null, $f = 1/(2\pi R2C1)$

audio-frequency meter

audio-frequency noise Any electrical noise signal causing interference within the audio-frequency spectrum.

audio-frequency oscillator See AUDIO OSCILLATOR.

audio-frequency peak limiter Any circuit or device, such as a biased diode, that performs the function of audio limiting.

audio-frequency-shift keying Abbreviation, AFSK. Frequency-shift keying that is done at audio frequencies (below approximately 20 kHz) rather than at radio frequencies. There are two audio sine-wave signals, one for logic 1 (*high* or *mark*) pulses and the other for the logic 0 (*low* or *space*) pulses. This scheme is commonly used with telephone modems where the signal bandwidth is severely limited by circuit characteristics. At typical data speeds in twisted-pair telephone lines (usually 28.8 or 57.6 kbps), signals of this type sound like a hiss or roar. Compare FREQUENCY-SHIFT KEYING.

audio-frequency-shift modulator A modulator for audio-frequency-shift keying of a signal.

audio-frequency spectrum The band of frequencies extending from roughly 20 Hz to 20 kHz. High-fidelity component specifications extend this range somewhat in both directions (e.g., from 10 Hz to 50 kHz).

audio-frequency transformer Abbreviation, AF transformer. A device used for the purpose of matching impedances at frequencies within the range of human hearing (up to approximately 20 kHz). This ensures the most efficient possible transfer of power between stages of audio amplification, between an amplifier and a speaker or headset, or between a microphone and an audio preamplifier. These transformers are available with various power ratings and impedance-matching ratios. Some devices are tailored to have a certain attenuation-versus-frequency response. At audio frequencies, transformers are physically similar to the alternating-current transformers used in power supplies. They are wound on laminated or powdered-iron cores. Compare RADIO-FREQUENCY TRANSFORMER.

audio-frequency transistor A transistor that is usually used only at audio frequencies.

audiogram A graph used to rate hearing, used by audiologists and audiometrists.

audio image In a direct-conversion receiver, a response to a signal on one side of (above or below) the local-oscillator (LO) frequency, when the operator is listening to a signal on the other side of the LO frequency. These responses are reduced or eliminated in single-signal receivers.

audio-level meter An ac meter for monitoring signal amplitude in an audio-frequency system. It can indicate in volts, decibels, volume units (VU), or arbitrary units, and is often permanently connected in the circuit.

audio limiter A limiter or clipper operated in the audio-frequency (AF) channel of a receiver or transmitter to hold the output-signal amplitude constant, or to minimize the effect of noise peaks.

audiologist A person skilled in testing hearing (i.e., in using audiometers and other electronic instruments) and evaluating their indications for the fitting of hearing aids.

audiometer An instrument used for hearing tests, which consists of a specialized audio-frequency (AF) amplifier with calibrated attenuators, output meter, and signal source.

audiometrist A person skilled in the use of audiometers and other electronic instruments that measure sound and human hearing, and who deals with attendant health and behavior problems. Compare ACOUSTICIAN and AUDIOLOGIST.

audio mixer An amplifier circuit for blending two or more audio-frequency (AF) signals, such as those delivered by microphones or receivers.

audio oscillator **1.** An oscillator that delivers an output signal in the frequency range 20 Hz to 20 kHz. **2.** An audio-frequency (AF) signal generator. Some instruments of this type operate above and below the limits of the common audio-frequency spectrum (e.g., 1 Hz to 1 MHz).

audio output The output of an audio-frequency oscillator or amplifier. It can be measured in terms of peak or rms volts, amperes, or watts.

audiophile A sound-reproduction hobbyist.

audio power Alternating-current power at frequencies roughly between 20 Hz and 20 kHz. When used in connection with transmitters and other modulated radio-frequency (RF) equipment, the term refers to modulator power output.

audio response unit A device that links digitized responses, held in computer storage, to a telephone set or line to answer incoming calls and inquiries.

audio signal generator See AUDIO OSCILLATOR, **2.**

audio spectrum The range of sine-wave frequencies detectable by the human ear when they occur as acoustic vibrations. This range is about 20 Hz to 20 kHz.

audio squelch A squelch circuit that operates only on the audio channel of a receiver.

audio system **1.** The portion of any electronic assembly that is used to process sound. **2.** Special computer equipment capable of storing and processing digitized audio-frequency (AF) data.

audiotape Magnetic tape for the recording and reproduction of data in the audio-frequency (AF) range.

audio taper In potentiometers, a semilogarithmic variation of resistance versus rotation. Used in volume and tone controls for audio circuits. At midposition (the halfway point), the counterclockwise portion of the device has about 1/10 the resistance of the clockwise portion. A listener will hear sound at half-volume because of the logarithmic nature of the human audibility curve.

audio-visual Pertaining to a combination of sound and sight (e.g., television and sound motion pictures).

auditory backward inhibition A subjective phenomenon, in which a sound is erased from the memory of a listener by a second sound arriving about 60 milliseconds later.

auditory inhibition The tendency of sound waves to be partially or totally canceled by the ears/mind of a listener, depending on the waves' intensity, relative phase, and/or direction of impact.

auditory mirage See ACOUSTIC MIRAGE.

audit trail A history of the processes relating to a record, transaction, or file in a computer system. Created during the routine processing of data, the trail is stored as a file. The audit trail allows auditing of the system or the subsequent recreation of files.

augend In a calculation, the number to which another is to be added. Compare ADDEND.

augend register In a digital computer, the register that stores the augend. Compare ADDEND REGISTER.

aural Pertaining to sound actually heard, as opposed to sound that exists only as audio-frequency currents or waves.

aurora A phenomenon sometimes called the *northern lights* or *southern lights*, as seen in the night sky. In the Northern Hemisphere, it is known as *Aurora Borealis*; in the Southern Hemisphere, it is called *Aurora Australis.* It generally occurs a few hours after a solar flare, when charged particles, emitted from the sun, arrive at the earth, and are accelerated in the vicinity of the the geomagnetic poles.

auroral absorption Radio wave absorption by an aurora.

auroral flutter Rapid fading of a signal at high or very high frequencies, so-called because it often imparts a fluttering quality to the signal that is caused by phase distortion and Doppler shift when the waves are reflected from the aurora.

auroral interference **1.** Interference to high-frequency radio propagation and also occasionally to medium-frequency and low-frequency propagation, caused by the activity of the aurora. **2.** Auroral flutter on a signal.

auroral opening A condition in which radio communication becomes possible via AURORAL PROPAGATION. It can occur when communication between two points is normally impossible at a certain frequency. Auroral openings allow long-distance communication well into the very-high-frequency (VHF) spectrum.

auroral propagation Reflection of radio signals from aurora that occur during geomagnetic storms. Theoretically, auroral propagation is possible when the aurora are active, between any two points on the earth's surface from which the same part of the aurora lie on a line of sight. This type of propagation seldom occurs when one end of the

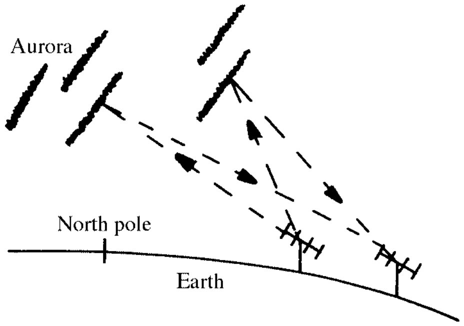

auroral propagation

circuit is at a latitude less than 35 degrees north or south of the equator. Auroral propagation can take place at frequencies well above 30 MHz. It is characterized by deep, rapid fading and random phase modulation of reflected signals.

auroral reflection The return of electromagnetic waves that have been beamed toward an aurora. Most often observed between 15 MHz and 150 MHz.

authorized access switch A device that disables a security system in a defined region or volume so that authorized personnel can enter without triggering an alarm condition.

authorized channel The carrier frequency or band assigned to a transmitting station by a licensing authority. Also see RADIO SPECTRUM.

autoalarm A device that is actuated from a received signal to alert a radio or computer network operator to the existence of a message.

autobaud **1.** In digital communications, a function that allows the equipment to adjust itself to the speed of the terminal. **2.** Any digital communications equipment capable of automatically adjusting to the speed of the terminal.

autocondensation The application of radio-frequency (RF) energy to the human body for medical purposes. The living organs serve as an impedance or load, across which the RF is applied.

autoconduction The application of radio-frequency (RF) currents into the body, by placing the living organ inside a coil and supplying the coil with RF. Used for medical purposes.

autocorrelation function A measure of the similarity between delayed and undelayed versions of a signal, expressed as a delay function.

autodyne reception Radio reception of cw signals by means of an oscillating detector. This is in contrast to heterodyne reception, in which a local oscillator (LO) generates an audio beat note with the cw signal in a separate detector.

autoionization A two-phase process of atomic ionization. The atom is excited beyond its ionization potential, and then it is allowed to deionize, causing the emission of an electron. The result is a positively charged atom (positive ion).

automated communications The transfer of data without the use of operating personnel; generally done with computers connected to communications equipment.

automated guided vehicle Abbreviation, AGV. A robot cart that runs without a driver. It uses an electric engine and is guided by the magnetic field produced by a current-carrying wire embedded in the floor or pavement. Alternatively, the robot can run on a track.

automated home A residence in which many, or most, of the routine chores are done by computers and/or robots. Examples of such tasks are dishwashing, doing the laundry, mowing the lawn, blowing snow, and vacuuming the floors.

automated integrated manufacturing system Acronym, AIMS. An assembly line or factory that uses robots, often controlled by one or more computers, to perform specific tasks that result in the production of various hardware items.

automatic Self-regulating, independent of human intervention. Some periodic adjustment might be needed.

automatic base bias A method of obtaining base bias in a bipolar transistor, where a resistor develops a voltage drop because of the current flowing through it. The resistor is usually placed in the emitter circuit, raising the emitter above ground potential.

automatic bass compensation Also called *bass boost.* In audio high-fidelity systems, a resistor-capacitor (RC) network that increases the relative amplitude of the bass at low volume levels. This compensates for the ear's inefficiency at low frequencies. The function can be automatically actuated by the setting of the volume control, or it can be switched manually on and off.

automatic bias In an amplifier, dc base/gate/grid bias obtained from the voltage drop produced by collector/drain/plate current flowing through a resistor common to the input and output. This resistor is usually shunted by a capacitor and placed in the emitter/source/cathode circuit.

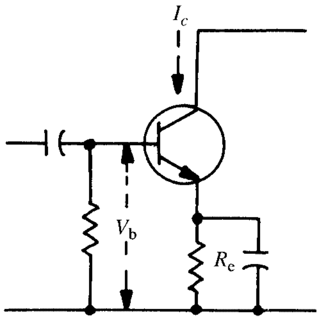

automatic bias

automatic brightness control A circuit that uses the same principles used in AUTOMATIC GAIN CONTROL (AGC) to hold steady the average brightness of a television (TV) picture.

automatic carriage Typewriters, automatic key punches, and other devices that can control automatically the spacing and feeding of paper, cards, and forms.

automatic check **1.** In a digital computer, the automatic inspection of operation and performance by a self-contained subsystem. **2.** The circuit or device for performing this inspection.

automatic chrominance control In a color television (TV) receiver, a subcircuit that controls the gain of the chrominance bandpass amplifier by automatically adjusting its bias.

automatic circuit breaker Any device that opens a circuit automatically when the flow of current becomes excessive. The breaker generally resets automatically after a specified length of time, or after power has been temporarily removed from the circuit.

automatic coding The use of a computer to determine the steps for solving a problem, before the actual program for the problem is written. This can help software engineers develop long and/or complex computer programs.

automatic contrast control A circuit that automatically adjusts the gain of the video IF and RF stages of a television (TV) receiver to preserve good picture contrast.

automatic controller In servo systems, any of several circuits or devices that samples a variable signal, compares it with a standard (reference) signal, and delivers a control or correction signal to an actuator.

automatic crossover **1.** Current limiting in a power supply. **2.** A device that switches a circuit from one operating mode to another automatically when conditions change in a predetermined manner.

automatic current limiter A circuit or device for holding the output current of a power supply to a safe value during overload.

automatic current regulator A circuit or device that holds the output current of a generator or power supply to a predetermined value, in spite of wide variations in load resistance.

automatic cutout A device that shuts down a circuit or system when the safe limits of operation are exceeded. A circuit breaker is an example of such a device, as is a thermostat in a power amplifier.

automatic data processing Abbreviation, ADP. The use of computers and accessories for calculations and tabulations using data gathered automatically by the system.

automatic degausser A system for automatically demagnetizing the picture tube in a color television (TV) receiver.

automatic dialing unit Abbreviated, ADU. A device that automatically generates dialing digits. Many telephone sets have these devices, some of which can be programmed for several different telephone numbers, including country codes and area codes.

automatic dictionary A computer system component that substitutes codes for words and phrases in information retrieval systems. In language-translating systems, it provides word-for-word substitutions.

automatic direction finder Abbreviated ADF. A specialized receiver/antenna combination for automatically showing the direction from which a signal arrives.

automatic error correction A technique of correcting transmission errors using error-detecting and error-correcting codes and, usually, automatic retransmission.

automatic exchange A transmission exchange in which interterminal communications are accomplished without operators.

automatic focusing A method of focusing a picture tube automatically, in which a resistor connects the focusing anode to the cathode; thus, no external focusing voltage is necessary.

automatic frequency control Abbreviation, AFC. A system that keeps a circuit automatically tuned to a desired signal frequency. A detector (such as a discriminator) operated from the tuned circuit delivers a dc output voltage only when the circuit is operating above or below the signal frequency; otherwise, it has zero dc output. The dc output, when present, alters the capacitance of a varactor in the tuned circuit to retune the stage to the desired frequency.

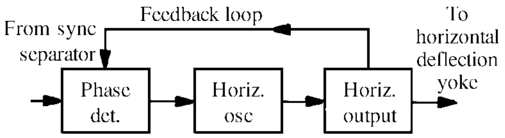

automatic frequency control (used in typical TV receiver)

automatic gain control Abbreviated AGC. A system that holds the output of a receiver or amplifier substantially constant despite input-signal amplitude fluctuations. A rectifier samples the ac signal output and delivers a dc signal proportional to that output. The dc signal is filtered, and the smoothed-out voltage is applied in correct polarity as bias to one or more preceding stages to reduce their gain. The stronger the signal entering the system, the greater the reduction in gain. As a result, weak signals are amplified much more than strong ones. Various forms of this scheme are used in many types of amplifiers and communications systems.

automatic gate bias A method of obtaining gate bias in a FET, where a resistor develops a voltage drop because of the current flowing through it. The resistor is usually placed in the source circuit, raising the source above ground potential.

automatic height control In a television (TV) receiver, a system that automatically maintains the height of the picture, despite signal-amplitude fluctuations, power-line voltage changes, and gain variations.

automatic intercept A telephone answering machine. It allows messages to be recorded when the subscriber is not able to answer the telephone.

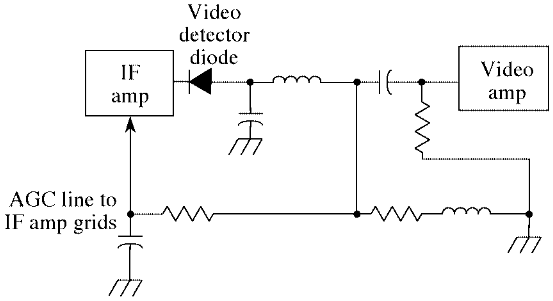

automatic gain control (used in typical TV receiver)

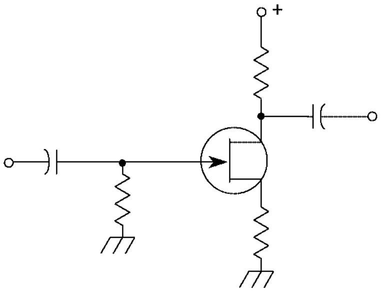

automatic gate bias

automatic interrupt A program interruption caused by hardware or software acting in response to some event independent of the program.

automatic level compensation See AUTOMATIC GAIN CONTROL.

automatic level control Abbreviation, ALC. **1.** A circuit that adjusts the input gain of a magnetic-tape recording device to compensate for changes in the loudness of the sound reaching the microphone. **2.** A form of AUTOMATIC GAIN CONTROL used in single-sideband (SSB) radio transmitters to maintain linearity while increasing the level of the average power relative to the peak power.

automatic line feed In the digital transmission of printed matter, the automatic insertion of a line feed (LF) character immediately following every carriage return (CR) character.

automatic modulation control Abbreviation, AMC. In a frequency-modulated (FM) radio transmitter, a form of AUTOMATIC GAIN CONTROL that regulates the gain of the audio amplifiers to compensate for fluctuating audio input amplitude. This prevents overdeviation while optimizing signal intelligibility.

automatic noise limiter Abbreviation, ANL. Any of several circuits for clipping noise peaks exceeding a predetermined maximum received-signal amplitude.

automatic phase control In a color television (TV) receiver, a circuit that synchronizes the burst signal with the 3.58-MHz color oscillator.

automatic pilot An electronic device, often computer-controlled, that automatically keeps a ship, airplane, or space vehicle on course.

automatic polarity In an electronic metering device, a means of automatically switching the input polarity of the instrument when the input signal polarity is shifted. Also called *bipolar operation.*

automatic programming See AUTOMATIC CODING.

automatic protective device A circuit or device (such as a fuse, circuit breaker, limiter, or regulator) that protects another circuit or device by automatically removing, reducing, or increasing the current or voltage during overload or underload.

automatic radio compass See AUTOMATIC DIRECTION FINDER.

automatic ranging In a metering device, the automatic adjustment or optimization of the full-scale range to compensate for large changes in the input parameter.

automatic regulation **1.** *Voltage regulation.* In a power supply, the automatic holding of the output voltage to a constant value, despite variations in the input voltage or load resistance. **2.** *Current regulation.* In a power supply, the automatic holding of the output current to a constant value, despite variations in the input voltage or load resistance.

automatic relay The relaying of messages automatically from one station to another via intermediate points, without the need for human operators.

automatic repeater station A station that receives signals and simultaneously retransmits them, usually on a different frequency.

automatic reset **1.** The self-actuated restoration of a circuit or device to a given state (e.g., the state of rest). **2.** A circuit or device that restores another circuit or device to a given state.

automatic scanning **1.** The automatic (usually repetitive) tuning or adjustment of a circuit or system throughout a given frequency range. In a radio receiver, the system can be programmed to pause or stop at occupied channels, passing over vacant ones; or it can be programmed to pause or stop at vacant channels, passing over occupied ones. **2.** The repetitive sweep of a cathode-ray-tube (CRT) electron beam.

automatic scanning receiver Also called PANORAMIC RECEIVER. A radio receiver that is automatically tuned (usually repetitively) over a frequency band. Such a receiver either homes in on a signal when one is found, or displays on a cathode-ray-tube (CRT) screen the distribution of signals in the band.

automatic secure voice communications A wideband and narrowband voice-digitizing application to a security network that provides encoded voice communications.

automatic send/receive set A teletypewriter or terminal that is capable of receiving and transmitting.

automatic sensitivity control **1.** A self-actuating circuit using principles similar to those used in AUTOMATIC GAIN CONTROL. It varies the sensitivity of the radio-frequency (RF) and intermediate-frequency (IF) sections of a receiver in inverse proportion to the strength of a received signal. **2.** In a bridge null detector, a circuit similar to the one described in **1**, which operates ahead of the detector, varying the sensitivity of the latter automatically.

automatic sequencing The ability of a digital computer to perform successive operations without additional instructions from the operator.

automatic short-circuiter A device that automatically short-circuits the commutator bias in some single-phase commutator motors.

automatic short-circuit protection A circuit that allows the output of a power supply to be short-circuited without damage to the components in the supply. It usually consists of a current-limiting device.

automatic shutoff A switching arrangement that automatically shuts off a device or circuit under certain specified conditions.

automatic switch center A telephone-switching network that routes calls to their destinations without the need for a human operator.

automatic target control For a vidicon television camera tube, a circuit that automatically adjusts the target voltage in proportion to brightness of the scene.

automatic telegraph reception Telegraph reception providing a direct printout of the received information, without intervention by an operator.

automatic telegraph transmission Telegraph transmission originating from tapes, disks, or other records, rather than from a hand-operated key.

automatic telegraphy Communications that utilize automatic telegraph transmission and reception.

automatic time switch A time-dependent circuit or device that opens or closes another circuit at the end of a predetermined time interval.

automatic tracking A method of keeping a radar beam automatically fixed on a target.

automatic trip A circuit breaker that automatically opens a circuit.

automatic tuning A process whereby a circuit tunes itself to a predetermined frequency upon receiving a command signal.

automatic voltage regulator A circuit that keeps the output of a power supply constant, despite the load resistance or input voltage to the supply.

automatic volume control Abbreviated AVC. The use of AUTOMATIC GAIN CONTROL in an audio amplifier system.

automatic zero In an electronic meter, a means of automatically setting the indicator to zero in the absence of an input signal.

automation **1.** The control of machines or processes by self-correcting electronic systems. See ROBOT. **2.** The use of robots and/or computers, rather than human beings, to perform repetitive tasks. **3.** The use of robots and/or computers to assist human beings in industrial, office, governmental, and educational work.

automaton A simple robot that performs a task or set of tasks without artificial intelligence (AI). These machines have existed for decades. Compare ANDROID.

automonitor In digital computer operations, to require the machine to supply a record of its information-handling operations. Also, the program for such instructions.

automotive battery A set of four or eight rechargeable lead-acid cells connected in series and housed in a common enclosure. The electrolyte is a free-flowing liquid acid. Provides approximately 6 volts (four cells) or 12 volts (eight cells) under no-load conditions when fully charged. The high mass results in large energy-storage capacity. These batteries must be handled with care and always kept in an upright position to prevent spillage of the acid. See LEAD-ACID BATTERY, LEAD-ACID CELL.

autonomous robot A self-contained robot with an independent computerized control system. It moves under its own power, usually by rolling on wheels or a track drive. Compare INSECT ROBOT.

autopatch A remotely controllable device that interconnects a radio-communications system into a telephone network.

autopilot A self-correcting control and guidance device for the automatic management of an aircraft or missile.

autoranging See AUTOMATIC RANGING.

autosyn A device or system that operates on the principle of the synchronous ac motor, in which the position of the rotor in one motor (the transmitter) is assumed by the rotor in a distant motor (the receiver) to which the first is connected.

auto tracking A method of controlling the output voltages of many different power supplies simultaneously.

autotransducter A type of magnetic amplifier whose power windings serve also as control windings.

autotransformer A single-winding transformer in which the primary coil is a fraction of the entire winding for voltage step-up, or the secondary coil is a fraction of the entire winding for voltage step-down.

auxiliary circuit A circuit that is supplementary to the main system.

auxiliary contacts In switches and relays, contacts that are supplementary to the main contacts and are usually actuated with them.

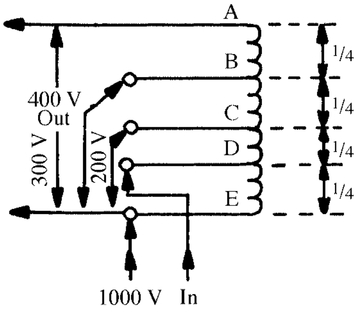

autotransformer

auxiliary equipment **1.** Also known as *peripherals.* An apparatus not directly governed by the central processing unit of a digital computer, such as a printer or personal robot. **2.** Peripheral equipment in any system. **3.** Backup equipment.

auxiliary memory In a digital computer, a unit that is supplementary to the main memory, which it augments.

auxiliary receiver Also called *standby receiver.* In a radio communications system, a receiver that is available for use if the main receiver fails.

auxiliary relay **1.** A standby relay. **2.** A relay whose operation supports that of another relay. **3.** A relay that is actuated by the operation of another relay.

auxiliary switch **1.** A standby switch. **2.** A switch wired in series or parallel with another switch. **3.** A switch that is operated by another switch.

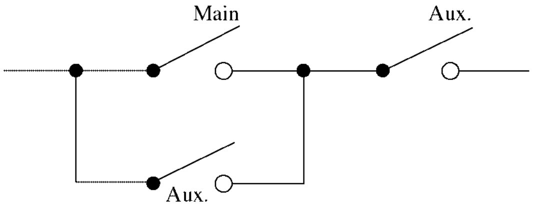

auxiliary switch, 2

auxiliary transmitter Also called *standby transmitter.* In a radio communications system, a transmitter that is available for use if the main transmitter fails.

a/v Abbreviation of AUDIO-VISUAL.

aV Abbreviation of *attovolt.*

availability The proportion of time during which an apparatus is operating correctly. It is usually given as a percentage.

available conversion gain The ratio of the input power to the output power of a transducer or converter. It is generally given in decibels.

available gain The ratio P_o/P_i, where P_i is the available power at the input of a circuit and P_o is the available power at the output.

available line The percentage of the length of a facsimile scanning line that is usable for picture signals.

available power The mean square of the open-circuit terminal voltage of a linear source divided by four, times the resistive component of the source impedance. The available power is the maximum power delivered to a load impedance, equal to the conjugate of the internal impedance of the power source.

available power gain In a power transistor, the ratio of available transistor output power to the power available from the generator. It depends on the generator resistance, but not on the transistor load resistance.

available signal-to-noise ratio The ratio P_s/P_n, where P_s is the available signal power at a given point in a system and P_n is the available random-noise power at that point.

available time **1.** The time during which a computer is available and ready for immediate use. **2.** The amount of time a computer is available to an individual.

avalanche The phenomenon in semiconductors operated at high reverse bias voltage, whereby carriers acquire sufficient energy to produce new electron-hole pairs as they collide with atoms. The action causes the reverse current to increase sharply.

avalanche breakdown In a semiconductor P-N junction, a condition that occurs when the reverse bias voltage exceeds a certain value. If the electric field in the vicinity of the junction becomes strong enough, charge carriers are dislodged from the atoms and the carriers (electrons and holes) flow freely across the P-N junction in the opposite direction from normal. The mini-

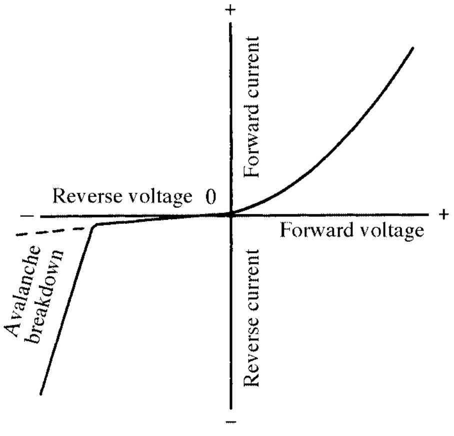

avalanche breakdown

mum reverse-bias voltage required to cause this phenomenon varies among different kinds of diodes. Some diodes are manufactured to have precise avalanche voltages. See ZENER DIODE.

avalanche conduction In a semiconductor junction, the enhanced reverse-bias conduction caused by a condition of AVALANCHE.

avalanche current The high current that flows through a semiconductor junction when AVALANCHE occurs.

avalanche diode See ZENER DIODE.

avalanche impedance The reduced impedance of a diode during avalanche.

avalanche noise Electrical noise generated in a junction diode operated at the point at which avalanche just begins.

avalanche transistor A transistor that operates at a high value of reverse-bias voltage, causing the pn junction between the emitter and base to conduct because of avalanche breakdown.

avalanche voltage In a semiconductor P-N junction, the minimum applied reverse-bias voltage that produces AVALANCHE BREAKDOWN.

AVC Abbreviation of *automatic volume control.*

avdp Abbreviation for *Avoirdupois*, a weight-measurement scheme that is used in English-speaking countries and is based on the pound.

average absolute pulse amplitude The average (disregarding algebraic sign) of the absolute amplitudes of a pulse, taken over the duration of the pulse.

average brightness The average brilliance of a television (TV) picture, cathode-ray-tube (CRT) computer display, or oscilloscope image.

average calculating operation The operating time considered typical for a computer calculation (i.e., one that is longer than an addition and shorter than a multiplication); it is frequently taken as the average of nine additions and one multiplication.

average current Abbreviation, I_{avg}. The average value of alternating current flowing in a circuit. Taking polarity into account, this value is zero for a pure sine wave. For other waveforms, it can vary. When polarity is not considered, the sine-wave value of I_{avg} is equal to 0.637 times I_{pk}, the peak value of current; $I_{avg} = 0.637\ I_{pk}$.

average life See MEAN LIFE.

average noise figure The ratio of the total noise output from a circuit to the thermal noise output at 290 degrees Kelvin. It is usually expressed in decibels, with the noise taken at all frequencies.

average power The average value of power in an ac circuit. In a resistive circuit, it is the square of the effective (rms) current times the resistance; $P_{avg} = (I_{rms})^2R$ (for sine waves).

average pulse amplitude Also called *effective pulse amplitude.* The value obtained by integrating the pulse amplitude, with respect to time, from the start of the pulse to its end, then dividing this integral by the pulse duration.

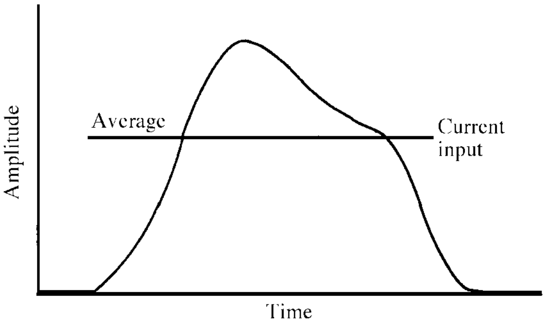

average pulse amplitude

average rectified current Abbreviation, I_{avg}. The average value of rectifier output current before filtering. For a full-wave rectifier with a sine wave input and a resistive load, I_{avg} is equal to the maximum current I_m multiplied by 0.637.

average rectified voltage Abbreviation, E_{avg}. The average value of rectifier output voltage before filtering. For a full-wave rectifier with a sine-wave input and a resistive load, E_{avg} is maximum voltage E_m multiplied by 0.637.

average value **1.** The *arithmetic mean* of two or more quantities. **2.** The *geometric mean* of two or more quantities. **3.** The *harmonic mean* of two or more quantities. **4.** In ac operation, the average current, voltage, or power.

average voltage Abbreviation, E_{avg}. The average value of ac voltage in a circuit. Taking polarity into account, this value is zero for a pure sine wave. For other waveforms, it can vary. When polarity is not considered, the sine-wave value of E_{avg} is equal to 0.637 times E_{pk}, the peak value of voltage.

avg Abbreviation of *average.*

aviation channels Frequency channels assigned to the AVIATION SERVICES.

Aviation services The radio-communication services used by aeronautical-mobile and radio navigation personnel.

avigation Acronym for *aviation navigation.* Aircraft navigation by means of electronic equipment.

avionics Acronym for *aviation electronics.* The design, production, and application of electronic devices and systems for use in aviation, navigation, and astronautics.

Avogadro's constant (Amedeo Avogadro, 1776–1856.) Symbol, NA. The number of molecules in a kilogram-molecular weight of any substance; NA equals 6.025×10^{26} (kg-mole)$^{-1}$.

A voltage The filament voltage in a vacuum-tube circuit.

aW Abbreviation of *attowatt.*

AWG Abbreviation of AMERICAN WIRE GAUGE.

AX.25 A signal format used in some digital communications systems, notably amateur *packet radio.*

axial leads The centrally located leads emanating from the ends of cylindrical components, such as resistors and diodes.

axial ratio The ratio of the minor to major axes of a waveguide's polarization ellipse.

axis **1.** A coordinate in a graphical presentation or display (e.g., horizontal and vertical axes in a rectangular coordinate system). **2.** The real or imaginary straight line around which a body rotates, or the line that passes through the center of a symmetrical arrangement (line of symmetry).

axis of abscissas The horizontal axis (x-axis) of a rectangular-coordinate graph or screen. Compare AXIS OF ORDINATES.

axis of imaginaries The vertical axis of the complex plane in which rectangular vectors lie. Compare AXIS OF REALS.

axis of ordinates The vertical (y-axis) of a rectangular-coordinate graph or screen. Compare AXIS OF ABSCISSAS.

axis of reals The horizontal axis of the complex plane in which rectangular vectors lie. Compare AXIS OF IMAGINARIES.

Ayrton-Mather galvanometer shunt A step-adjustable universal shunt resistor for varying the sensitivity of a galvanometer. It has the virtue of keeping the galvanometer critically damped. The shunt is also useful in multirange milliammeters, microammeters, and ammeters. The sensitive meter movement is never without a shunting resistor during range switching.

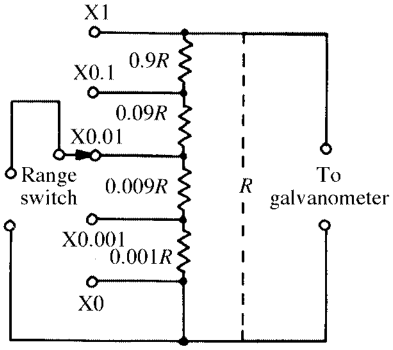

Ayrton-Mather galvanometer shunt

Ayrton-Perry winding A noninductive winding comprising two inductors conducting current in opposite directions; the opposing flow cancels the magnetic field.

azel display A plan-position display that incorporates two different radar traces on a single cathode-ray tube (CRT), one giving bearing, the other elevation.

azimuth Also called *compass direction.* Angular measurement in the horizontal plane, clockwise from north. It is important in radio and television communications, navigation, direction finding, land surveying, and radar.

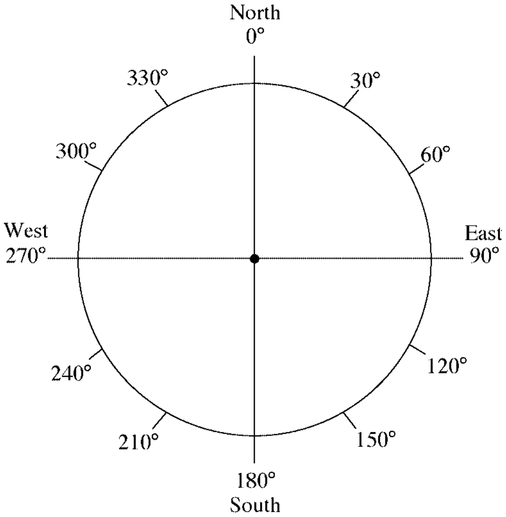

azimuth

azimuth alignment In a tape recorder, the alignment of record and playback head gaps so that their centerlines are parallel.

azimuth blanking In a radar system, blacking-out of the image as the antenna sweeps across a specified range of azimuth angles. Effectively eliminates nuisance echoes from stationary, permanent objects (such as tall buildings or communications towers).

azimuth resolution In a radar system, the minimum azimuth separation of two targets whose range (distance from the station) are equal that is required for the system to show two echoes, rather than one. It is generally measured in degrees.

azusa An electronic tracking system, in which a single station provides slant range and two direction cosines for a distant airborne object. This accurately defines the coordinates of the distant object in three-dimensional space.

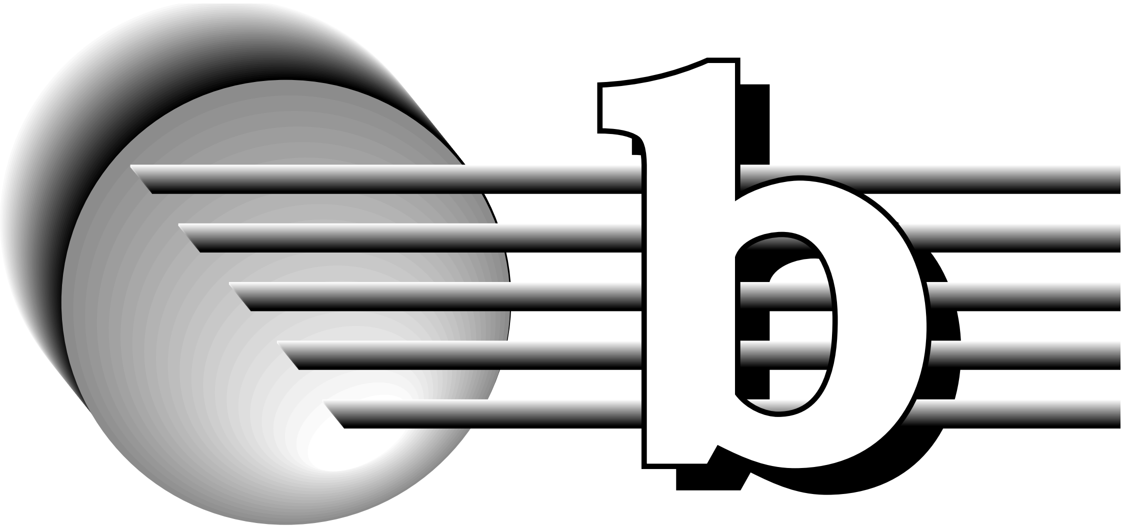

B **1.** Symbol for SUSCEPTANCE. **2.** Symbol for FLUX DENSITY. **3.** Abbreviation of BATTERY. **4.** Symbol for BORON. **5.** Symbol for base of transistor (see BASE, **1**). **6.** Abbreviation of BASS. **7.** Abbreviation of BEL. **8.** Anode voltage or main operating voltage in any circuit (when used with sign). Also see B VOLTAGE.

b **1.** Symbol for SUSCEPTANCE. **2.** Symbol for base of transistor (see BASE, **1**). **3.** Abbreviation of BASS. **4.** Symbol for BARN.

B&S See AMERICAN WIRE GAUGE.

B5-cut crystal A piezoelectric plate cut from a quartz crystal in such a way that the face of the plate is at an angle, with respect to the z-axis of the crystal. This type of crystal has good frequency stability under conditions of changing temperature.

BA Abbreviation of BATTERY. Also see B and BAT.

Ba Symbol for BARIUM.

babbit A relatively soft, tin-base alloy of various compositions. One composition contains 7.4% antimony, 3.7% copper, and 88.9% tin.

babble Interference caused by *crosstalk* from a number of channels.

babble signal A jamming signal containing babble components. See BABBLE and JAMMING.

BABS Abbreviation of BLIND-APPROACH BEACON SYSTEM.

baby monitor A short-range radio transmitter and receiver that can be used to listen at a distance to the sounds in an infant's room. The transmitter contains a sensitive microphone, a whip antenna, and a power supply. The unit can be placed on a table or desk, or even on the floor near the baby's crib. The receiver is similar to a handheld "walkie-talkie." It is battery-powered and can be carried around. It has an inductively loaded, short "rubber duckie" antenna similar to the antennas on cordless telephone sets. The receiver can pick up signals from the transmitter at distances of up to about 200 feet. The radio-frequency signals pass easily through walls, ceilings, and floors.

back bias **1.** A feedback signal (negative or positive). **2.** Reverse bias (also see BIAS). **3.** A reverse bias voltage, obtained from a voltage divider connected between a source and ground.

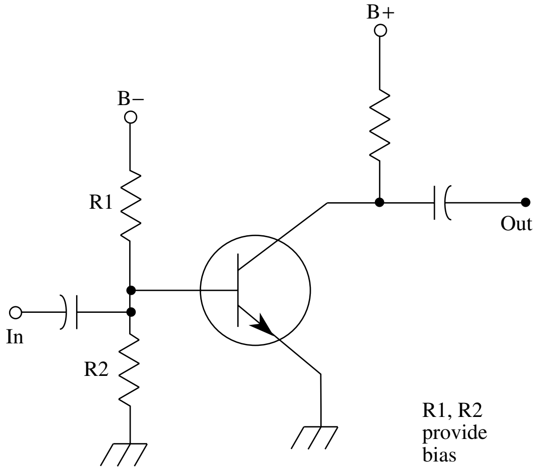

back bias, 3

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

backbone A form of transmission line with capacitive connections between the generator and the load.

back conduction Conduction of current in the reverse direction, as across a semiconductor junction that is reverse-biased.

back contact A contact that closes a circuit when a relay, switch, or jack is in its normal rest position.

back current Symbol, I_b. The normally small current flowing through a reverse-biased pn semiconductor junction. Also called *reverse current* and *inverse current.* Compare FORWARD CURRENT.

back diode A semiconductor diode that is normally back-biased (reverse-biased).

back echo An echo resulting from the rear lobe of an antenna radiation pattern.

back emf See BACK VOLTAGE.

Back-Goudsmit effect See ZEEMAN EFFECT.

background **1.** The context or supporting area of a picture (e.g., the background of a television picture). **2.** Background noise.

background control In a color television receiver, a potentiometer used to set the dc level of the color signal at one input of the three-gun picture tube.

background count Residual response of a radioactivity counter in an environment as free as practicable of radioactivity. This background is caused largely by cosmic rays and inherent radioactivity of surrounding buildings and other bodies.

background job A low-priority, relatively long-running computer program that can be interrupted so that a higher-priority program can be run.

background noise Electrical noise inherent to a particular circuit, system, or device that remains when no other signal is present.

background processing In a computer, the running of programs having low priority.

background radiation Nuclear radiation from materials in the environment. Also see BACKGROUND COUNT.

background response The response of a radiation detector to background radiation.

backing store In a computer, a device that stores large amounts of information. In most small computers, this is done via MAGNETIC DISK and/or MAGNETIC TAPE. A backing store can also be an optical storage medium, such as CD-ROM (compact disk, read-only memory).

backlash **1.** Slack or lag in action of moving parts. Example: delay between initial application of a force (such as that required to turn a knob) and movement of a part or device (e.g., a potentiometer or variable capacitor). **2.** On a mechanical analog tuning dial, an arc within which slack or lag is discernible.

backloaded horn A loudspeaker enclosure in which the front of the speaker cone feeds sound directly into the listening area, and the rear of the cone feeds sound into the same area through a folded horn.

backloading In a cascaded series of amplifiers, the tendency of loading effects to be passed to earlier stages. A change in the output impedance of a final amplifier circuit, for example, could also result in a change in the output impedance of the driver circuit, and perhaps even in a change in the output impedance of the predriver.

back lobe In the pattern of a directional antenna, the lobe directly opposite the major lobe, representing the radiation or response in or from a direction 180 degrees from that in which the gain is greatest.

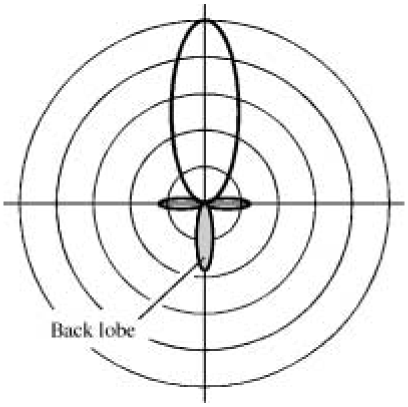

back lobe

backplate A flat electrode in a television (TV) camera tube that receives the stored-charge image via capacitive coupling.

back porch In a television (TV) horizontal sync pulse, the time interval between the end of the rise of the blanking pedestal and the beginning of the rise of the sync pulse. That portion of the flat top of the blanking pedestal behind the sync pulse. Compare FRONT PORCH.

back-porch effect In transistor operation, the continuation of collector-current flow for a short time after the input signal has fallen to zero.

back-porch tilt The departure of the top edge of a back porch from true horizontal.

back pressure sensor A device that detects and measures the torque that a motor is applying, and produces a signal whose amplitude is proportional to the torque. This signal can be used for various purposes. In a robotic device, for example, the sensor output can be fed back to the motor control to limit the applied force.

back resistance Symbol, R_b. The resistance of a reverse-biased pn semiconductor junction. Also called REVERSE RESISTANCE.

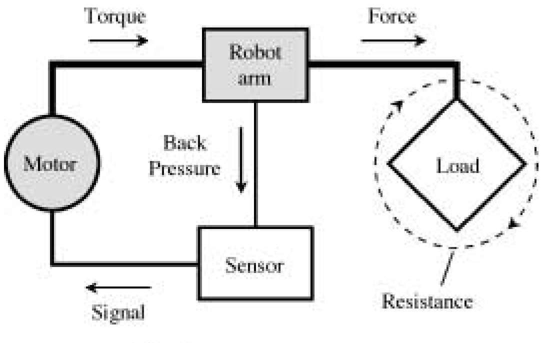

back pressure sensor

back scatter Scattering of a wave back toward a radio transmitter from points beyond the skip zone. This phenomenon is caused by ionospheric reflection. Compare FORWARD SCATTER.

backstop A contact or barrier (such as a screw or post) that serves to limit the BACKSWING of the armature of a relay.

backswing **1.** The tendency of a pulse to overshoot, or reverse direction after completion. Backswing is measured in terms of the overshoot amplitude as a percentage of the maximum amplitude of the pulse. **2.** The extent to which a relay armature moves back from a contact when the relay contacts are open.

back-to-back connection The connection of diodes or rectifiers in reverse parallel (i.e., the anode of one to the cathode of the other) across a signal line to pass both half cycles of ac in certain control circuits.

back-to-back sawtooth A symmetrical sawtooth wave in which the rise slope is equal to the fall slope. Also called *triangular wave* and *pyramidal wave.*

backup **1.** An element, such as a circuit component, that is used to replace a main component, in case of main-component failure. **2.** Any process or scheme that serves to maintain operation of a system in case of main-component failure. **3.** A battery that maintains volatile memory data stored in one or more integrated circuits. **4.** A computer file, or set of files, stored in a nonvolatile medium, such as diskettes or magnetic tape, to prevent catastrophic data loss in the event of hard-disk failure. **5.** A battery or alternative power source that keeps an alarm system operational in the event of a utility power failure.

backup battery **1.** In a computer or microcomputer-controlled electronic device, a source of voltage to preserve volatile memory data if the power is removed. **2.** A battery used for powering a system in the event that the main power source should fail.

backup facility In an electrical or communications system, a facility that is intended for use when the primary, or main, facility is not operational.

back voltage **1.** Voltage induced in an inductor by the flow of current through the inductor, so called because its polarity is opposite to that of the applied voltage. Also called *counter emf.* **2.** A voltage used to obtain bucking action (e.g., the voltage used to zero the meter in an electronic voltmeter circuit). **3.** Reverse voltage applied to a semiconductor junction.

backwall In a pot core, the plate or disk that connects the sleeve and center post to close the magnetic circuit.

backward diode A semiconductor diode manufactured in such a way that its high-current flow occurs when the junction is reverse biased. Such a diode is also a negative-resistance device.

backward-wave oscillator Abbreviation, BWO. A microwave oscillator tube similar to the *traveling-wave tube.* Like the traveling-wave tube, the BWO contains a helical transmission line. In the electron beam, electron bunching results from interaction between the beam and the electromagnetic field, and reflection occurs at the collector. The wave moves backward from collector to cathode, and oscillation is sustained because the backward wave is in phase with the input. Output is taken from the cathode end of the helix.

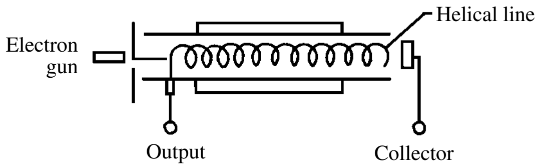

backward-wave oscillator

back wave The oscillator signal present in an amplifier-keyed, continuous-wave (CW), Morse-code transmitter. Normally, this signal is at the same frequency as the transmitter output, but is not sufficiently strong to be radiated over the air.

back-wave radiation The condition wherein a back wave is strong enough to be heard on a continuous-wave (CW) keyed signal at the receiving station. This results from ineffective amplifier keying.

baffle A board on which a loudspeaker is mounted to separate acoustic radiation from the back of the cone from radiation emanating from the front. The baffle improves bass response by increasing the wavelength (lowering the frequency) at which phase cancellation occurs.

baffle plate **1.** See BAFFLE. **2.** A metal plate mounted in a waveguide to reduce the cross-sectional area.

bail A wire loop or chain that holds one member of a two-member assembly to prevent loss (e.g., the short chain holding the dust cap of a jack).

Bakelite The trade name for a specialized plastic dielectric material. Its chemical composition is *phenol-formaldehyde* resin.

baker An obsolete phonetic alphabet code word for letter B. BRAVO is commonly used instead.

baking-out In the process of evacuating a system, the procedure of heating the system to a high temperature to drive out gases occluded in the glass and metal parts.

balance **1.** See BRIDGE. **2.** To null a bridge or similar circuit. **3.** To equalize loads, voltages, or signals between two circuits or components. **4.** In a high-fidelity stereo sound system, a control or set of controls that adjusts the relative loudness of the left and right channels. **5.** Alignment of a *balanced modulator* for minimum carrier output amplitude. **6.** A condition in which two branches of a circuit have identical impedances, relative to ground.

balance coil **1.** A type of autotransformer that enables a three-wire ac circuit to be supplied from a two-wire line. A series of taps around the center of the winding enables the circuit to be compensated for unequal loads. **2.** See BALANCING COIL.

balance control A variable component, such as a potentiometer or variable capacitor, that is used to balance bridges, null circuits, or loudspeakers.

balanced Having identical impedances, with respect to ground.

balanced amplifier Any amplifier with two branches that have identical impedances, with respect to ground. Usually, the two branches are in phase opposition (180 degrees out of phase).

balanced antenna An antenna system where two halves are exact replicas of each other, geometrically and electrically. Such an antenna normally must either be fed with a balanced transmission line or with a coaxial cable and balun.

balanced antenna system A balanced antenna, fed with a balanced transmission line, that has currents of equal magnitude in each side. An example is a half-wave dipole at uniform height above electrical ground, fed at the center with parallel-wire line. It is important that the transmission line runs away from the antenna at a right angle for at least ¼ wavelength, preferably ½ wavelength or more, to prevent line imbalance caused by currents induced from the radiated field.

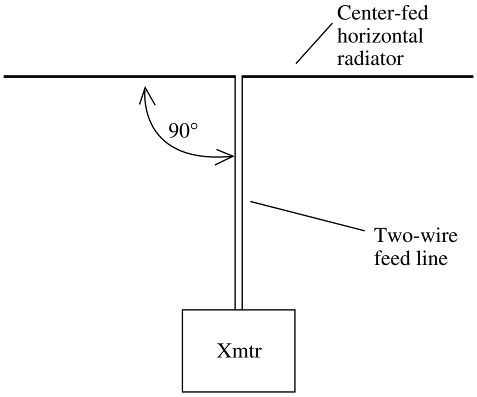

balanced antenna system

balanced bridge Any four-leg bridge circuit in which all legs are identical in all electrical respects.

balanced circuit **1.** A circuit that has its electrical midpoint grounded, as opposed to the *single-ended circuit*, which has one side grounded. **2.** A bridge circuit in the condition of null.

balanced converter See BALUN.

balanced currents Currents with the same value. In the two conductors of a balanced transmission line, these currents are equal in amplitude and opposite in phase at every point along the line.

balanced delta A set of coils or generators in a three-phase system, connected so that the currents in any two coils differ in phase by 120 degrees.

balanced detector A symmetrical demodulator, such as a full-wave diode detector or a discriminator.

balanced electronic voltmeter An electronic voltmeter circuit in which two matched transistors are connected in a four-arm bridge arrangement. The drift in one half of the circuit opposes that in the other half; the resulting drift of the zero point is virtually eliminated.

balanced filter A filter consisting of two identical sections, one in each branch of a balanced system, such as a parallel-wire transmission line.

balanced input An input circuit whose electrical midpoint is grounded. Compare SINGLE-ENDED INPUT.

balanced input transformer An input transformer in which the center tap of the primary winding is grounded.

balanced line A pair of parallel wires that possesses a uniform characteristic impedance. The two conductors are of the same material and have identical diameters. The distance between them is constant. In a balanced two-wire line, the currents in the two conductors are of equal amplitude and opposite phase.

balanced lines In high-fidelity audio systems, a cable that consists of two parallel conductors surrounded by a single braid. The parallel wires carry the audio-frequency (AF) signals, and the braid is grounded for shielding.

balanced loop antenna A loop antenna with a grounded electrical midpoint, determined by the junction of two identical series-connected capacitors shunting the loop.

balanced low-pass filter A low-pass filter used in a balanced system or balanced transmission line.

balanced method A system of instrumentation in which a zero-center scale is used. The reading can be either side of the zero reading.

balanced modulator A symmetrical modulator circuit using bipolar transistors, field-effect transistors, an integrated circuit, or diodes as principal components, that delivers an output signal containing the sidebands, but not the carrier. It is commonly used to generate a double-sideband (DSB) signal that can be filtered to obtain a single-sideband (SSB) signal.

balanced multivibrator A switching oscillator circuit in which the two halves are identical in configuration, and as nearly identical as practicable in performance.

balanced network Any network intended to be used with a balanced system or balanced transmission line. It is characterized by a pair of terminals, each of which shows the same impedance with respect to ground.

balanced oscillator A PUSH-PULL OSCILLATOR.

balanced output Output balanced against ground (e.g., where the electrical midpoint of the output circuit is grounded).

balanced output transformer **1.** A push-pull output transformer with a center-tapped primary winding. **2.** An output transformer with a grounded center tap on its secondary winding.

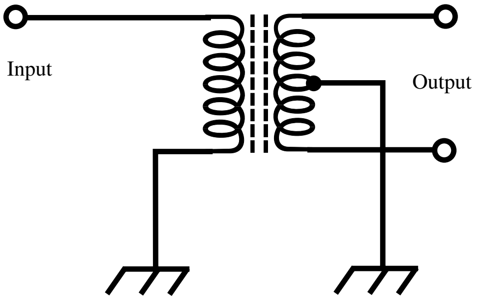

balanced output transformer, 2.

balanced probe A probe, such as one for an electronic voltmeter or oscilloscope, that has a balanced input and (usually) a single-ended output.

balanced-tee trap A wavetrap constructed in a T configuration, with a resonant section in each conductor of a balanced transmission line.

balanced telephone line A telephone transmission line that has two sides, similar to a balanced radio-frequency transmission line. Either side has the same impedance, with respect to ground.

balanced termination A load device (or the practice of using such a device) in which the sections provide identical termination for each of the sections or conductors of a balanced system, such as a balanced line.

balanced-to-unbalanced transformer See BALUN.

balanced transmission line See BALANCED LINE.

balanced varactor tuning A two-varactor, back-to-back circuit for adjusting the value of a capacitor using an applied dc voltage. This arrangement has an advantage over a single-varactor (unbalanced) circuit, because high-tuned-circuit *Q* is maintained and harmonic generation is reduced.

balanced voltages In any symmetrical system, such as a balanced line or push-pull circuit, two or more input or output voltages that are adjusted to have the same amplitude and (usually) opposite phase.

balanced-wire circuit A circuit or conductor system with identical halves that are symmetrical, with respect to ground and to other conductors.

balancing circuit See BUCKING CIRCUIT.

balancing coil In a receiver, a center-tapped antenna coil that is balanced to ground to eliminate MARCONI EFFECT.

ballast **1.** A component that is used to stabilize the current flow through, or operation of, a circuit, stage, or device. **2.** An iron-core choke connected in series with one of the electrodes in a fluorescent or other gas-discharge lamp.

ballast resistor **1.** A nonlinear inductive power resistor whose voltage-current (EI) characteristic is such that current through the resistor is independent of voltage over a useful range. This feature enables the ballast resistor to act as an automatic voltage regulator when it is simply connected in series with a power supply and load. **2.** A small (usually high-resistance) resistor operated in series with a glow lamp, such as a neon lamp, to prevent overload.

ballast transformer A misnomer often used in place of BALLAST, **2.**

ballistic galvanometer An undamped galvanometer that is used particularly to observe electric charges by noting the single throw resulting from the momentary flow of current through the galvanometer coil.

ballistics The electronics-supported science concerned with the motion of projectiles and similar bodies in air or space.

balloon antenna A vertical antenna consisting of a wire or wires held aloft by a captive balloon. Occasionally, used by radio amateurs and shortwave listeners at low and medium frequencies. A potentially dangerous antenna because of large static-electric buildup, a tendency to attract lightning, the possibility of its breaking loose, and the risk of accidental contact with high-voltage power lines.

balop Contraction of BALOPTICON.

balopticon An opaque-picture projecting system in which the picture is viewed by a television (TV) camera, such as a vidicon, and displayed by a picture tube. Also called *balop.*

balun A specialized impedance-matching radio-frequency (RF) transformer. It is a wideband device, usually providing a 1:1 or 1:4 impedance ratio and available in several different forms. It is so called because it has an unbalanced input suitable for coaxial transmission lines, and a balanced output suitable for dipole, Yagi, and quad antennas.

banana jack The female half of a two-part quick-connector combination. Splicing of a circuit is completed by inserting a BANANA PLUG into this jack.

banana plug The male half of a two-part quick-connector combination, with sides usually composed of flat springs that ensure contact with the female BANANA JACK into which it is inserted.

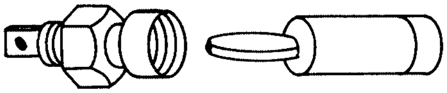

banana jack and plug

band **1.** A continuous range of radio or television communications frequencies or wavelengths, usually designated by the lowest and highest frequencies, or the approximate wavelength (e.g., the 20-meter amateur radio band). **2.** A set of discrete radio or television frequency channels within a specified range (e.g., the standard AM broadcast band). **3.** A range of wavelengths for infrared, visible, ultraviolet, X-ray, or gamma-ray energy. **4.** A range of energy levels. **5.** A colored stripe on a resistor or capacitor that forms part of the code that indicates component value and tolerance.

band center **1.** In a given radio or television communications band, the arithmetic mean of the lowest and highest frequencies. **2.** In a given band, the geometric mean of the longest and shortest wavelengths.

band-elimination filter See BAND-REJECTION FILTER.

band gap In any atom, the difference in electron energy between the conduction and valence bands.

bandpass **1.** The frequency limits between which a BANDPASS FILTER or BANDPASS AMPLIFIER transmits ac energy with negligible loss. **2.** The ability to allow passage of signals at a given frequency or band of frequencies while blocking other signals. Compare BANDSTOP.

bandpass amplifier An amplifier that is tuned to pass only those frequencies between preset limits.

bandpass coupling A coupling circuit with a flat-topped frequency response so that a band of frequencies, rather than a single frequency, is coupled into a succeeding circuit. Also see BANDPASS, **1.**

bandpass filter Any resonant circuit, or combination of resonant circuits, designed to discriminate against all frequencies except a specific frequency f_0, or a band of frequencies between two limiting frequencies f_0 and f_1. In a parallel inductance-capacitance (LC) circuit, the device exhibits high impedance at the desired frequency or frequencies and a low impedance at unwanted frequencies. In a series configuration, the filter has a low impedance at the desired frequency or frequencies and a high impedance at unwanted frequencies. Compare BAND-REJECTION FILTER, HIGH-PASS FILTER, LOW-PASS FILTER.

bandpass flatness The degree to which a bandpass device's attenuation-versus-frequency curve is a straight line with zero slope within the passband.

band pressure level The net acoustic pressure of a sound source within a specified frequency range (band).

band-rejection filter Also called a *band-stop filter*. Any resonant circuit, or combination of resonant circuits designed to discriminate against a specific frequency f_0, or a band of frequencies between two limiting frequencies f_0 and f_1. In a parallel inductance-capacitance (LC) circuit, the device exhibits high impedance at the desired frequencies, and a low impedance at the unwanted frequency or range of frequencies. In a series configuration, the filter has a low impedance at the desired frequencies and a high impedance at the unwanted frequency or range of frequencies. Compare BANDPASS FILTER, HIGH-PASS FILTER, LOW-PASS FILTER, NOTCH FILTER.

band selector Any switch or relay that facilitates switching the frequency of a radio transmitter, receiver, or transceiver among various bands.

bandset capacitor In some older communications receivers, a variable capacitor is used to preset the tuning range in each band to correspond to graduations on the tuning dial. This capacitor is a trimmer or padder operated in conjunction with the main tuning capacitor.

bandspreading In some older communications receivers, the process of widening the tuning range within a given frequency band to cover the entire dial. Otherwise, the band would occupy only a portion of the dial, and tuning would be difficult. It is usually accomplished with a BANDSPREAD TUNING CONTROL whose range is preset via the main tuning control and/or a BANDSET CAPACITOR.

bandspread tuning control An analog adjustment in some older communications receivers that allows continuous tuning over a desired band of frequencies. This control is separate from the main tuning control.

bandstop **1.** The frequency limits between which a BAND-REJECTION FILTER blocks, or greatly attenuates, ac energy. **2.** The ability to suppress or block signals of a given frequency or band of frequencies, while allowing signals of other frequencies to pass with little or no attenuation. Compare BANDPASS.

bandstop filter See BAND-REJECTION FILTER.

band suppression **1.** The property of blocking, or greatly attenuating, signals within a specific frequency band. **2.** The frequency limits between which a device or circuit rejects or blocks ac energy, while passing energy at other frequencies with negligible loss.

band-suppression filter See BAND-REJECTION FILTER.

bandswitch A low-reactance selector switch (usually rotary) that facilitates changing the tuning range of a radio receiver, transmitter or transceiver from one band of frequencies to another.

bandswitching In a receiver, transmitter, or test instrument, the process of switching self-contained tuned circuits to change from one frequency spectrum to another within the range of the device's intended operation.

bandwidth **1.** For a communications or data signal, a measure of the amount of spectrum space the signal occupies. Usually, it is given as the difference between the frequencies at which the signal amplitude is nominally 3 dB down with respect to the amplitude at the center frequency. These frequencies represent the half-power points of the amplitude-versus-frequency function. In general, the bandwidth increases as the data rate (in bits per second, baud, or words per minute) increases. **2.** Also called NECESSARY BANDWIDTH. The minimum amount of spectrum space normally required for effective transmission and reception of a communications or data signal. **3.** See BANDPASS, **1.**

bank A collection of usually similar components used in conjunction with each other, usually in a parallel configuration. Some examples are resistor bank, lamp bank, and transformer bank.

banked transformers Parallel-operated transformers.

bankwound coil A coil wound in such a way that most of its turns are not side by side, thus reducing the inherent distributed capacitance.

bar **1.** Abbreviation, b. The cgs unit of pressure, in which 1 b = 10^5 pascals per square centimeter. **2.** A horizontal or vertical line produced on a television (TV) screen by a bar generator and used to check linearity. **3.** A thick plate of piezoelectric crystal. **4.** A solid metal conductor, usually uninsulated, of any cross section. **5.** A silicon ingot from which semiconductor devices can be fabricated.

BAR Abbreviation of BUFFER ADDRESS REGISTER.

bar code A printed pattern that contains data that can be recovered by laser scanning. It is commonly used for the pricing and identification of store merchandise. It can also be used by an assembly or maintenance robot as an aid to identifying tools.

bar-code reader A laser scanning device that recovers the data from a tag that contains a BAR CODE. The laser beam moves across the tag. The beam is reflected from the white regions between the lines, but is absorbed by the dark lines themselves. This produces modulation of the reflected beam by the data contained in the tag.

bare conductor A conductor with no insulating covering, a common example being bare copper wire.

bar generator A special type of radio-frequency signal generator that produces horizontal or vertical bars on the screen of a television receiver. It is used in adjustment of horizontal and vertical linearity.

bar graph A graphical presentation of data, in which numerical values are represented by horizontal bars of width that correspond to the values. This type of graph is nonstandard in the sense that the ordinate is horizontal, whereas it is usually vertical. Compare COLUMNAR GRAPH.

bar-graph meter See BAR METER.

barium Symbol, Ba. An elemental metal of the alkaline-earth group. Atomic number, 56. Atomic weight, 137.36. It is present in some compounds used as dielectrics (e.g., barium titanate).

barium-strontium oxides The combined oxides of barium and strontium used as coatings of vacuum-tube cathodes to increase electron emission at relatively low temperatures.

barium strontium titanate A compound of barium, strontium, oxygen, and titanium that is used as a ceramic dielectric material. It exhibits ferroelectric properties and is characterized by a high dielectric constant.

barium titanate Formula, $BaTi0_2$. A ceramic used as the dielectric in ceramic capacitors. It exhibits high dielectric constant and some degree of ferroelectricity.

Barkhausen effect The occurrence of minute jumps in the magnetization of a ferromagnetic substance as the magnetic force is increased or decreased over a continuous range.

Barkhausen interference Interference that results from oscillation because of the BARK-HAUSEN EFFECT.

bar magnet A relatively long permanent magnet in the shape of a bar with a rectangular or square cross section.

bar meter A digital meter that displays a quantity, such as signal strength, incrementally, using a set of LEDs or LCDs arranged in a straight line. Its main advantage is that it has no moving parts, yet (unlike direct-readout digital meters) gives the viewer some impression of the way a rapidly fluctuating quantity changes. Its chief

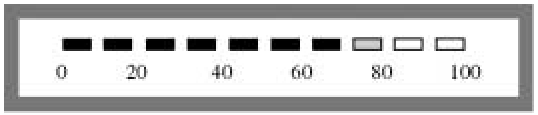

bar meter

disadvantage is that it does not provide a precise indication.

barn Symbol, b. A non-SI unit of nuclear cross section equal to 100 square femtometers or 10^{-24} square centimeters. This unit is approved as compatible with SI (International System of Units).

Barnett effect The development of a small amount of magnetization in a long iron cylinder that is rotated rapidly about its longitudinal axis.

barograph A recording barometer, using either a drum recorder (pen recorder) or a computer to store the data as a function of atmospheric pressure versus time.

barometer An instrument for measuring atmospheric pressure.

barometer effect A relation that appears to exist between the intensity of *cosmic rays* and the atmospheric pressure. It is an inverse relation; that is, increasing pressure seems to correlate with reduced intensity of cosmic rays. It is said to be approximately to 1 or 2% per centimeter of mercury.

barometric pressure The atmospheric pressure, usually given in inches of mercury. The average barometric pressure at the surface of the earth is just under 30 inches of mercury.

bar pattern A series of spaced lines or bars (horizontal, vertical, or both) produced on a television picture screen by means of a BAR GENERATOR. It is useful in adjusting horizontal and vertical linearity of the picture.

barrage array An antenna array in which a string of collinear elements are vertically stacked. The end quarter wavelength of each string is bent in to meet the end quarter wavelength of the opposite radiator to improve balance.

barrage jamming The jamming of many frequencies, or an entire band, at the same time.

barrell distortion Television picture distortion consisting of horizontal and vertical bulging.

barrier **1.** The carrier-free space-charge region in a semiconductor pn junction. **2.** An insulating partition placed between two conductors or terminals to lengthen the dielectric path.

barrier balance The state of near equilibrium in a semiconductor pn junction (after initial junction forming), entailing a balance of majority and minority charge carrier currents.

barrier capacitance **1.** The capacitance in a bipolar transistor between the emitter and collector. It varies with changes in applied voltage, and also with the junction temperature. **2.** The capacitance across any pn junction that is reverse-biased.

barrier height The difference in voltage between opposite sides of a barrier in a semiconductor material.

barrier layer See BARRIER, **1.**

barrier-layer cell A photovoltaic cell, such as the copper oxide or selenium type, in which photons striking the barrier layer produce the potential difference.

barrier potential The apparent internal dc potential across the barrier (see BARRIER, **1**) in a pn junction.

barrier strip A terminal strip having a barrier (see BARRIER, **2**) between each pair of terminals.

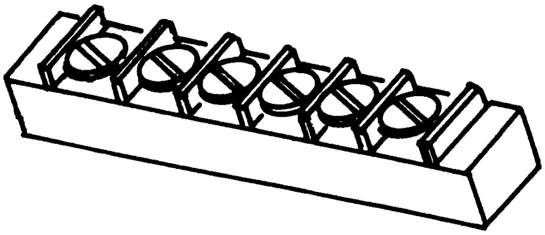

barrier strip

barrier voltage The voltage required for the initiation of current flow through a pn junction.

Bartlett force See EXCHANGE FORCE.

baryon A subatomic particle made up of three quarks.

base **1.** In a bipolar transistor, the intermediate region between the emitter and collector, which usually serves as the input or controlling element of transistor operation. **2.** A substance that dissociates in water solution and forms hydroxyl (OH) ions. For example, sodium hydroxide. **3.** The constant figure upon which logarithms are computed (10 for common logs, 2.71828 for natural logs). **4.** The radix of a number system (e.g., base 10 for the decimal system, base 8 for the octal system, base 16 for the hexadecimal system, and base 2 for the binary system). **5.** A fixed nonportable radio communications installation.

base address The number in a computer address that serves as the reference for subsequent address numbers.

baseband The frequency band of the modulating signal in a communications, broadcast, or data transmitter. For voice communications, this is generally the range of voice frequencies necessary for intelligible transmission. For high-fidelity music broadcasting, it is approximately the range of human hearing. For fast-scan television, it ranges up to several megahertz. It can be restricted or expanded, depending on the nature of the transmitted signal. See BASEBAND FREQUENCY RESPONSE.

baseband frequency response **1.** The amplitude-versus-frequency characteristic of the audio-frequency (AF) or composite video section of a transmitter that defines the BASEBAND, or range of modulating frequencies. **2.** The range of frequencies over which a radio transmitter can be modulated to convey information. For single sideband (SSB), it is approximately 300 Hz to 3 kHz; for high-fidelity, frequency-modulated (FM) music transmission, it is about 10 Hz to 20 kHz or 30 kHz; for fast-scan television, it consists of frequencies up to several megahertz. This range is determined by bandpass and/or lowpass filters in the AF or composite video section of the transmitter.

base bias The steady dc voltage applied to the base electrode of a transistor to determine the operating point along the transistor characteristic curve.

base-bulk resistance The resistance of the semiconductor material in the base layer of a bipolar transistor.

base-charging capacitance In the common-emitter connection of a bipolar transistor, the internal capacitance of the base-emitter junction.

base current Symbol, I_B. Current flowing through the base electrode of a bipolar transistor. Also see AC BASE CURRENT and DC BASE CURRENT.

base electrode See BASE, **1.** Also called *base element.*

base element **1.** Base electrode. **2.** One of the basic metals, such as iron or tin, that are not generally considered precious (as opposed to NOBLE).

base-e logarithm See NAPIERIAN LOGARITHM.

base film The plastic substrate of a magnetic recording tape.

base frequency **1.** The frequency of the principal, or strongest, component in a complex signal or waveform; also called *basic frequency.* **2.** The frequency of operation of a base-station transmitter when the receiver is tuned to a second channel.

base-input circuit A common-collector circuit, common-emitter circuit, or emitter follower.

base insulator A stout dielectric insulator, used to support a heavy conducting element and keep the conductor isolated from other possible conductors or conductive paths.

base line In visual alignment procedures involving an oscilloscope and radio-frequency (RF) sweep generator, a zero-voltage reference line developed by the generator as a horizontal trace on the oscilloscope screen.

baseline stabilizer A clamping circuit that holds the reference voltage of a waveform to a predetermined value. Also called DC RESTORER.

base-loaded antenna A usually vertical antenna or radiating element, the electrical length of which is adjusted by means of a loading coil or tuned circuit in series with, and positioned at the bottom of, the antenna or radiator.

base material In printed circuits, the dielectric material used as a substrate for the metal pattern. Also called *base medium.*

base notation The numbering or radix system used in any application (as octal, decimal, binary, or hexadecimal).

base number See BASE, **4.**

base pin One of the straight prong-like terminals on an electrical or electronic component; it is used to provide support for the device and to allow a physical connection between the socket terminal, into which it fits, and one of the internal electrodes of the device.

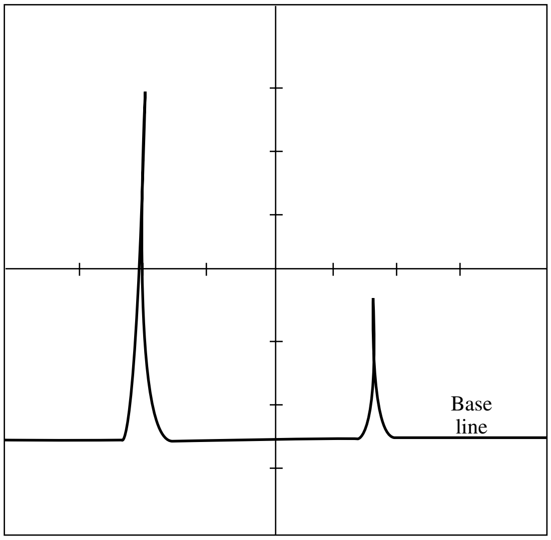

base line

base plate The chassis plate upon which components are mounted before wiring.

base potential See BASE VOLTAGE.

base region See BASE, **1.**

base resistance Symbol, R_B. Resistance associated with the base electrode of a bipolar transistor. Also see AC BASE RESISTANCE and DC BASE RESISTANCE.

base resistor The external resistor connected to the base of a bipolar transistor. In the common-emitter circuit, the base resistor is analogous to the gate resistor of a field-effect transistor (FET) circuit.

base spreading resistance Symbol, r_{BB}. In a bipolar transistor, the bulk-material resistance of the base region between the collector junction and emitter junction.

base station The head station or fixed home station in a communication network.

base-10 logarithm Abbreviation, $\log_{10}$. A logarithm based on the decimal number 10. If $\log_{10}(x) = y$, then $10^y = x$. Base-10 logarithms are commonly used in engineering. Compare NAPIERIAN LOGARITHM.

base voltage Symbol, V_B. The voltage at the base electrode of a bipolar transistor. Also see AC BASE VOLTAGE and DC BASE VOLTAGE.

BASIC Acronym for BEGINNER'S ALL-PURPOSE SYMBOLIC INSTRUCTION CODE, a relatively primitive, but versatile and easy-to-learn computer language developed at Dartmouth College.

basic frequency **1.** The FUNDAMENTAL FREQUENCY of a signal, as opposed to one of its harmonics. **2.** See BASE FREQUENCY, **1.**

basic protection Devices and procedures essential to minimize the risk of damage to electronic equipment, and/or injury or death to its operators, as a result of lightning. Hardware provisions include a substantial earth ground, heavy-gauge grounding wire, lightning arrestors for antennas, and transient suppressors for power connections. The safest procedure is to disconnect and ground all antennas, and unplug all equipment from utility outlets, during electrical storms and/or when the apparatus is not in use. Radio communications equipment with outdoor antennas, in particular, should not be operated during thunderstorms.

basket The structure that supports the cone in an acoustic loudspeaker.

basket-weave coil A type of single-layer inductor in which adjacent turns do not parallel each other around the circumference, but zigzag oppositely as a strand does in the woven pattern of a basket. This reduces distributed capacitance.

bass Low audio frequencies (AF) corresponding to low-frequency musical notes or sounds.

bass boost **1.** The special emphasis given to low audio frequencies (the bass notes) by selective circuits in audio systems. **2.** The technique of increasing the loudness of the bass, relative to the higher audio frequencies, to render a more faithful reproduction of sound at low volume levels.

bass compensation See BASS BOOST, **2.**

bass control **1.** A manually variable potentiometer for adjusting bass boost of an amplifier or sound system. **2.** The arrangement of components that are required to achieve amplitude variation of bass in an audio signal.

bass port In a loudspeaker, a hole in the cabinet that enhances the low-frequency (bass) sound output. Used in high-fidelity audio systems.

bass-reflex enclosure A loudspeaker cabinet with a critically dimensioned duct or port that allows back waves to be radiated in phase with front waves, thus averting unwanted acoustic phase cancellation.

bass-reflex loudspeaker A loudspeaker mounted in a bass reflex enclosure. Also see ACOUSTICAL PHASE INVERTER.

bass-resonant frequency The low frequency at which a loudspeaker or its enclosure displays resonant vibration.

bass roll-off **1.** The attenuation of the low-frequency (bass) component in a high-fidelity audio signal. **2.** A control that allows adjustable attenuation of the low-frequency component in a high-fidelity audio signal.

bass suppression In speech transmission, the removal of all frequencies below about 300 Hz, on the assumption that those frequencies contribute little to intelligibility. This suppression permits the speech level to be increased without overmodulating a transmitter. It also allows smaller audio transformers to be used because transformer core size must increase as the frequency it passes decreases.

bassy In audio and high-fidelity applications, a sound in which the low-frequency components, below about 500 Hz, are overly predominant.

BAT Abbreviation of BATTERY.

batch fabrication process The manufacture of devices in a single batch from materials of uniform grade. Particularly, the manufacture of a large number of semiconductor devices from one batch of semiconductor material by means of carefully controlled, identical processes.

batch processing In digital-computer operations, the processing of quantities of similar information during a single run.

bat-handle switch A toggle switch, the lever of which is relatively long and thick, and is shaped like a baseball bat.

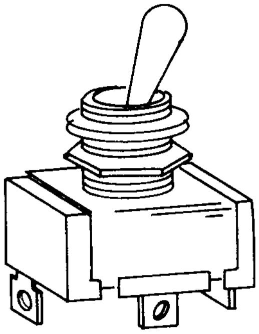

bat-handle switch

bathtub capacitor A (usually oil-filled) capacitor housed in a metal can that looks like a miniature bathtub.

bathyconductorgraph An instrument that is used to measure the electrical conductivity of seawater.

bathythermograph An instrument that plots a graph of temperature versus depth in a body of water, such as a lake or an ocean.

batten Supporting bars or braces that hold a loudspeaker in place within its cabinet, and/or that hold the cabinet panels in place.

battery Abbreviations, B, BA. BAT. A device consisting of two or more interconnected electrochemical or photovoltaic cells that generate dc electricity. The cells can be connected in series to supply a desired voltage, in parallel to supply a desired current-delivering capability, or in series-parallel to obtain a desired voltage and current-delivering capability. Also see CELL, EDISON BATTERY, LEAD-ACID BATTERY, PHOTO-

VOLTAIC CELL, PRIMARY BATTERY, and STORAGE BATTERY.

battery acid **1.** A chemical acid, such as sulfuric acid, used as the electrolyte of a battery. **2.** Colloquially, any cell or battery electrolyte, whether acid, base, or salt.

battery capacity The current-supplying capability of a battery, usually expressed in ampere-hours (Ah).

battery cell See CELL, **1.**

battery charger **1.** A specialized dc power supply, usually embodying a stepdown transformer, rectifier, and filter. It is used to charge a storage battery from an ac power line. **2.** A motor-generator combination used to charge a storage battery from an ac power line. **3.** A combination of solar cells, generators, or other voltaic transducers, that are used to charge a storage battery with dc obtained from a nonelectrical energy source.

battery clip **1.** A heavy-duty metallic clamp that is used for quick, temporary connection to a large cell terminal, such as that of a lead-acid storage battery. **2.** A small connector of the snap-fastener type, used for quick connection to a small power source, such as a transistor-radio battery.

battery eliminator A specialized dc power supply, usually embodying a transformer, rectifier, and filter, that permits battery-powered equipment to be operated from an ac power line.

battery holder **1.** A case or container of any kind for holding a cell or battery. **2.** A shelf for holding a cell or battery. **3.** A small, metal bracket-type device for holding a cell or battery between two contacts.

battery life **1.** The ampere-hour or watt-hour capacity of a battery. **2.** The number of times that a rechargeable electrochemical battery can be cycled before it becomes unusable. **3.** The nominal length of time (e.g., hours, days, or weeks) that an electrochemical battery will function effectively in a given application before it must be discarded or recharged.

battery memory See MEMORY DRAIN.

battery receiver A usually portable radio or television receiver operated from self-contained batteries.

battery substitute See BATTERY ELIMINATOR.

bat wing On a television (TV) or frequency-modulation (FM) broadcast receiving antenna, a metallic element with a shape that resembles that of a bat's wing.

baud A unit of communications processing speed in telegraphy and digital data communications systems. Often confused with *bits per second* (bps). Baud refers to the number of times per second that a signal changes state. The speed in bps is generally higher than the speed in baud, sometimes by a factor of several times. Compare BITS PER SECOND.

Baudot code A machine communications code that uses five parallel binary digits of equal length, the interpretation of which depends on the history of the previous transmission or an additional case bit.

baud rate **1.** A colloquial expression for data speed in BAUD. **2.** Colloquial, and technically inaccurate, expression for data speed in BITS PER SECOND.

Baume (Antione Baume, 1728–1804). Abbreviation, Be. Pertaining to the *Baume scales* for hydrometers. The two such scales are for liquids heavier than water and for liquids lighter than water.

bay One of several sections of a directional antenna array.

bayonet base The insertable portion of a plug-in component (e.g., a lamp) that has a projecting pin that fits into a slot or keyway in the shell of the socket into which the component is inserted.

bayonet socket A socket with a suitably slotted shell for receiving the bayonet base of a plug-in component.

bazooka A linear BALUN, in which a quarter wavelength of metal sleeving surrounds a coaxial feeder, and is shorted to the outer conductor of the feeder to form a shorted quarter-wave section.

bb Abbreviation of BLACKBODY.

BBC Abbreviation of *British Broadcasting Corporation.*

BBM Abbreviation of BREAK BEFORE MAKE.

b-box The index register of a computer.

BC Abbreviation of BROADCAST.

BCD Abbreviation of BINARY-CODED DECIMAL.

BCFSK Abbreviation of BINARY CODE FREQUENCY-SHIFT KEYING.

B channel One of the channels of a two-channel stereophonic system. Compare A CHANNEL.

BCI Abbreviation of BROADCAST INTERFERENCE.

BCL Amateur radio abbreviation of BROADCAST LISTENER.

BCN Abbreviation of BEACON.

BCO Abbreviation of BINARY-CODED OCTAL.

BCST Abbreviation of BROADCAST.

BDC Abbreviation of BINARY DECIMAL COUNTER.

B display A radar display in which the target is represented by a bright spot on a rectangular-coordinate screen. Compare A DISPLAY and J DISPLAY.

Be Symbol for BERYLLIUM.

Be Abbreviation of BAUME.

beacon **1.** A beam of radio waves, or a radio signal, that is used for navigation and/or direction finding. **2.** A transmitter that radiates a beam of radio waves, or a radio signal, as an aid in navigation and/or direction finding. **3.** A signal transmitted continuously on a specific frequency, to help radio operators ascertain propagation conditions. **4.** A station or transmitter that generates and radiates a signal to help radio operators determine propagation conditions. **5.** In robotics, a device or system that aids in navigation. For example, tricorner reflectors can be positioned in strategic locations, and a mobile robot equipped with a scanning infrared laser. The robot controller determines the distance to any given reflector by measuring the time required for the laser beam to return. In this way, two mirrors can allow the robot to locate its position in two dimensions; three mirrors can facilitate position determination in three-dimensional space.

beacon direction finder A direction finder using a signal received from a beacon station.

beacon receiver A receiver that is specially adapted for the reception of beacon signals (see BEACON, **1** and **3**).

beacon station **1.** A station broadcasting beacon signals (see BEACON, **1** and **3**) for direction finding, navigation, and/or determination of radio-wave propagation conditions. **2.** Sometimes, a radar transmitting station.

beacon transmitter A transmitter specially adapted for the transmission of beacon signals (see BEACON, **1** and **3**).

bead **1.** A small ferromagnetic ring that is used as a passive decoupling choke by slipping it over the input power leads of a circuit or stage, or around a coaxial transmission line. **2.** A magnetic memory element in a ferrite-core matrix.

beaded coax A low-loss, coaxial transmission line, in which the inner conductor is separated from the outer conductor by means of spaced dielectric beads.

beaded support A plastic or dielectric bead that is used to support the inner conductor of an air-insulated transmission line of coaxial construction.

bead thermistor A thermistor consisting essentially of a small bead of temperature-sensitive resistance material into which two leads are inserted.

beam **1.** The more-or-less narrow pattern of radiation from a directional antenna. **2.** A directional antenna—especially a YAGI ANTENNA. **3.** The stream or cloud of electrons emitted by the cathode in an electron tube—especially a BEAM POWER TUBE.

beam alignment **1.** The lining-up of a directional transmitting antenna with a directional receiving antenna for maximum signal transfer. **2.** In a television (TV) camera tube, the lining-up of the electron beam so that it is perpendicular to the target. **3.** In a cathode-ray tube, the positioning of the electron rays so that they converge properly on the screen, regardless of the deflection path.

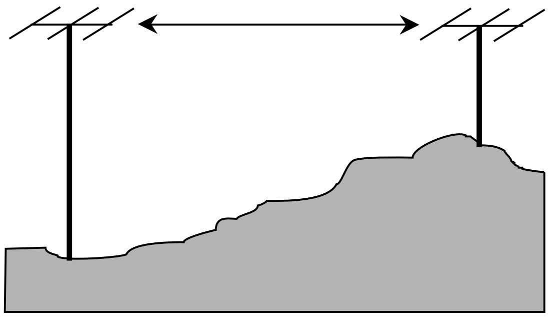

beam alignment

beam angle In the radiation from an antenna, the direction of most intense radiation, the side limits of which are determined by the points at which the field strength drops to half the value in the principal direction.

beam antenna **1.** A multielement directional antenna, consisting of a half-wave driven dipole and one or more parasitic elements. See YAGI ANTENNA. **2.** Any directional antenna used for transmitting and receiving radio-frequency (RF) signals.

beam bender **1.** In a television (TV) picture tube, the ion-trap magnet. **2.** Deflection-plate correction device or circuit.

beam bending Deflection of an electron beam by electric or magnetic fields.

beam blanking See BLANK, **2.**

beam convergence The meeting, at a shadow-mask opening, of the three electron beams in a three-color television picture tube. See BEAM ALIGNMENT, **3.**

beam coupling A method of producing an alternating current between two electrodes by passing a density-modulated beam of electrons between the electrodes. This, in effect, demodulates the electron beam, recovering the information.

beam crossover Either of the half-power points in the beam of a directional antenna, usually in the horizontal plane. The reference point is considered to be the direction of maximum radiation.

beam current The current represented by the flow of electrons in the beam of a cathode-ray tube.

beam cutoff In an oscilloscope or television picture tube, the complete interruption of the electron beam, usually as a result of highly negative control-grid bias.

beam deflector A deflection plate in an oscilloscope tube.

beam efficiency In a cathode-ray tube, the ratio of the number of electrons generated by the gun to the number reaching the screen. The efficiency is high in electromagnetic-deflection tubes and lower in electrostatic-deflection tubes.

beam lead In an integrated circuit, a relatively thick and strong lead that is deposited in contact with portions of the thin-film circuit. It provides stouter connections than continuations of the thin film would provide.

beam-lead isolation In an integrated circuit, reduction of distributed capacitance and other interaction through use of beam leads.

beam modulation See INTENSITY MODULATION.

beam parametric amplifier A PARAMETRIC AMPLIFIER in which the variable-reactance component is supplied by a modulated electron beam.

beam-positioning magnet In a three-gun color television picture tube, a permanent magnet that is used to position one of the electron beams correctly, with respect to the other two.

beam power tube A tetrode or pentode vacuum tube, in which special deflector plates concentrate the electrons into beams in their passage from cathode to plate. The beam action greatly increases plate current at a given plate voltage. It is used in some radio-frequency (RF) power amplifiers.

beam-rider control system A missile-guidance system in which a control station sends a radio beam to a missile. The beam is moved in such a way that as the missile stays within the beam, it hits the target.

beam-rider guidance **1.** An aircraft landing guidance system, in which the aircraft follows a radio beam in its glide path. **2.** The circuitry in a guided missile using a beam-rider control system.

beam splitter A device used to divide a light beam (as by a transparent mirror) into two components, one transmitted and the other reflected; hence, a BEAM-SPLITTING MIRROR.

beam splitting In radar, a method of calculating the mean azimuth of a target from the azimuth at which the target is first revealed by one scan, and the azimuth at which the target information ceases.

beam-splitting mirror In an oscilloscope-camera system, a tilted, transparent mirror that allows rays to pass horizontally from the oscilloscope screen to the camera and to be reflected vertically to the viewer's eye.

beamwidth of antenna The angular width of the main lobe of the pattern of radiation from a directional antenna. Generally, it is measured between the half-power points in the horizontal plane. Occasionally, it is measured in the vertical plane.

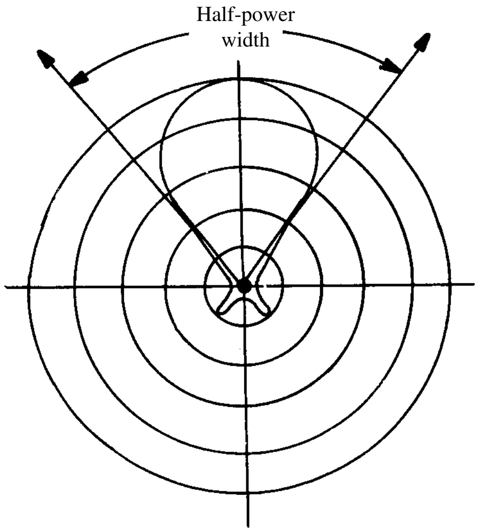

beamwidth of antenna

bearing The direction of an object or point expressed in degrees within a 360° horizontal clockwise boundary, with the center of the circle serving as the observation point.

bearing resolution In radar operations, the minimum horizontal separation of two targets, in degrees, that permits the individual targets to be displayed as two echoes, rather than one.

beat Any one of the series of pulsations constituting a beat note, which results from heterodyning one signal against another.

beat frequency Either of two frequencies f_{C1} and f_{C2} resulting from the mixing of two signals of different frequencies f_A and f_B. Frequency f_{C1} is the sum of the two input frequencies; $f_{C1} = f_A + f_B$. Frequency f_{C2} is the difference; $f_{C2} = f_A - f_B$ when f_A is the higher of the two input frequencies.

beat-frequency oscillator Abbreviation, BFO. An oscillator used to set up audible beat frequencies with an incoming received signal and installed in the intermediate-frequency (IF) stages of a superheterodyne communications receiver. For single-sideband (SSB) reception, the BFO is set at the frequency of the received suppressed carrier. In continuous-wave (CW) Morse code reception, the BFO is set at a frequency that differs from that of the incoming signal by about 400 to 1000 Hz. The resulting tone has an audio frequency equal to the difference between the BFO frequency and the received signal carrier frequency. For reception of frequency-shift-keyed (FSK) signals, the BFO is set to such a frequency that the resulting audio beat notes are appropriate for the mark and space inputs of a terminal unit or modem.

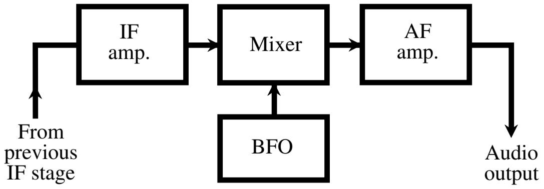

beat-frequency oscillator

beating **1.** Also called *heterodyning.* The combination of signals of different frequencies resulting in sum and difference frequencies. **2.** The fluttering noise heard when two audio tones, very close in frequency and very similar in amplitude, are emitted at the same time.

beat marker In the visual (oscilloscopic) alignment of a tuned circuit, a marker pip that results from the beat note between the sweep-generator signal and the signal from a marker oscillator.

beat note The sum or difference frequency that results from the heterodyning of two signals or, under some conditions, of more than two signals.

beat-note reception **1.** Reception in which a radio-frequency carrier is made audible by heterodyning it with a beat-frequency oscillator (BFO) to produce an audible beat note. **2.** Superheterodyne reception (see SUPERHETERODYNE CIRCUIT).

beat tone A beat note in which the frequency is within the range of hearing.

beaver tail A flat or elongated radar beam, wide in the azimuth plane. Primarily used to determine the altitude of a target. The beam is moved up and down to find the target elevation.

Becquerel effect A phenomenon in which a voltage is produced when radiant energy, such as infrared, visible light, ultraviolet, or X-rays, falls on one electrode in an electrolytic cell.

bedspring A directional antenna consisting of a broadside array with a flat reflector and one or more helical driven elements.

beep A test or control signal, usually of single tone and short duration.

beeper **1.** A pocket- or hand-carried transceiver—especially one for maintaining two-way contact with personnel who are away from their base. **2.** An acoustic transducer that produces a beep in response to an input signal.

beetle A urea formaldehyde plastic used as a dielectric material and as a container material.

bel Abbreviation, B. The basic logarithmic unit (named for Alexander Graham Bell) for expressing gain or loss ratios. One bel is equivalent to a power gain of 10. Also see DECIBEL.

bell An electric alarm device consisting of a metallic gong that emits a ringing sound when it is struck by an electrically vibrated clapper.

Bellini-Tosi direction finder A direction finder in which the sensing element consists of two triangular vertical antennas crossed at right angles, the antennas being open at the top and accordingly not acting as conventional coil antennas.

bell-shaped curve A statistical curve (so called from its characteristic shape) that exhibits a *normal distribution* of data. Typically, the curve describes the distribution of errors of measurement around the real value.

bell transformer A (usually inexpensive) stepdown transformer that operates an electric bell or similar alarm or signaling device from the ac power line.

bell wire Insulated 18-gauge (AWG) solid copper wire, so called because of its principal early use in the wiring of electric-bell circuits.

belt generator Also known as a *Van de Graaff generator.* A very-high-voltage electrostatic generator, a principal part of which is a fast-traveling endless belt of dielectric material. At the lower end, charges of one sign are sprayed on the belt at 10 to 100 kV dc and are carried to the inside of a hollow metal sphere at the upper end, where they are removed and spread to the surface of the sphere, which they raise to a potential up to several million volts.

benchmark A test standard to measure product performance.

benchmark routine A routine designed to evaluate computer software and/or hardware, producing a good indication of how well the software or hardware will perform in real-life situations. In particular, tests *instructions per second* and *throughput*, thereby producing an indication of the overall computer power in applications, such as word processing, database, spreadsheet, graphics, animation, and mathematical calculations.

bench test An extensive checkout of a piece of equipment in the test laboratory—either to find an intermittent problem, or to check for reliability.

bend An angular shift in the lengthwise direction of a waveguide.

bending effect **1.** The downward refraction of a radio wave by the ionosphere. **2.** The low-atmosphere turning of a radio wave downward by temperature discontinuity and atmospheric inversions.

Benito A continuous-wave method of measuring the distance of an aircraft from the ground, involving the transmission of an audio-modulated signal from ground and the retransmission back to ground by the aircraft. The phase shift between the two signals is proportional to the distance to the aircraft.

bent antenna An antenna that has its driven element bent, usually near the ends and at right angles, to conserve space.

bent gun A television picture tube neck arrangement having an electron gun that is slanted to direct the undesired ion beam toward a positive electrode, but which allows the electron beam to pass to the screen. This prevents the ion beam from "burning" a permanent spot on the phosopor of the screen.

BeO Formula for beryllium oxide. Also see BERYLLIA.

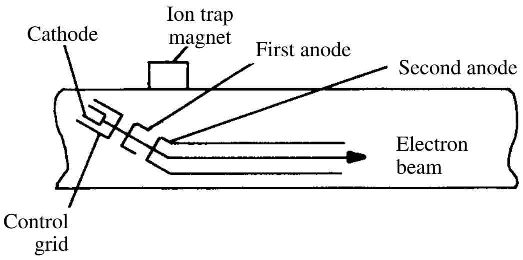

bent gun

berkelium Symbol, Bk. A radioactive elemental metal produced artificially. Atomic number, 97. Atomic weight, 247.

beryllia Formula, Be0. Beryllium oxide, used in various forms as an insulator and structural element (as in resistor cores).

beryllium Symbol, Be. An elemental metal. Atomic number, 4. Atomic weight, 9.01218. Beryllium is present in various dielectrics and alloys used in electronic components.

Bessel functions Sophisticated mathematical functions for dealing with periodic electronic phenomena in which the waveform often displays decrement. Also called *cylindrical functions.*

beta Symbol, β. The current gain of a common-emitter bipolar transistor stage. It is the ratio of the induced change of collector current to the applied change of base current: $\beta = dIC/dIB$.

beta circuit The output-input feedback circuit in an amplifier.

beta cutoff frequency The frequency at which the current amplification of a bipolar transistor falls to 70.7% of its low-frequency value.

beta particles Minute radioactive subatomic bits identical to the electron or positron, and emitted by some radioactive materials. Also see BETA RAYS.

beta rays Rays emitted by the nuclei of radioactive substances, consisting of a stream of beta particles (i.e., electrons or positrons) that move at velocities up to 299.8 million meters per second. Compare ALPHA PARTICLE and GAMMA RAYS.

beta-to-alpha conversion For a bipolar transistor, the conversion of current amplification expressed as beta (β) to current amplification expressed as alpha (α): $\alpha = \beta/(\beta + 1)$.

betatron A particle accelerator in which injected electrons are given extreme velocity by being propelled in circular paths in a doughnut-shaped glass container. The term comes from the fact that high-speed electrons constitute BETA PARTICLES.

beta videocassette recorder The earliest scheme for videocassette recording, developed by Sony corporation in the 1970s. Compare *VHS videocassette recorder.*

beta zinc silicate phosphor Formula, $(Zn0 + Si0_2)$:Mn. A phosphorescent substance used to coat the screen of a cathode-ray tube. The fluorescence is green-yellow.

BeV Abbreviation of *billion electronvolts.* Also see ELECTRONVOLT, GEV, MEV, and MILLION ELECTRONVOLTS. This abbreviation has been supplanted by the SI (International System of Units) abbreviation GeV, for GIGAELECTRONVOLTS.

bevatron An accelerator (see ACCELERATOR, **1**) similar to the synchrotron, which accelerates particles to levels greater than 10 GeV.

Beverage antenna (Harold H. Beverage.) A nonresonant, directional long-wire antenna, erected a few feet above ground and run in a straight line for one to several wavelengths. It is generally used for reception at low and medium frequencies, the best response is to vertically polarized signals arriving from one or both directions in line with the wire. It can be left unterminated for bidirectional response, or it can be terminated at its far end by a noninductive resistor of about 600 ohms for a unidirectional response.

beyond-the-horizon propagation See FORWARD SCATTER.

bezel A faceplate for an electronic instrument, usually having a fitted rim and cutouts for knobs, switches, jacks, etc.

BFO Abbreviation of BEAT-FREQUENCY OSCILLATOR.

BG Abbreviation of BIRMINGHAM WIRE GAUGE. Also abbreviated BWG.

B-H curve A plot showing the *B* and *H* properties of a magnetic material. Magnetizing force *H* is plotted along the horizontal axis, and flux density *B* is plotted along the vertical axis.

B-H loop See BOX-SHAPED LOOP.

B-H meter Any instrument for displaying or evaluating the hysteresis loop of a magnetic material.

bhp Abbreviation of *brake horsepower.*

Bi Symbol for BISMUTH.

bias **1.** Any parameter of which the value is set to a predetermined level to establish a threshold or operating point. Although it is common to think of bias currents and bias voltages, other parameters (e.g., capacitance, resistance, illumination, magnetic intensity, etc.) can serve as biases. **2.** In a high-fidelity audio system, a circuit in a tape recorder/player that optimizes performance for a particular type of recording tape.

bias current A steady, constant current that presets the operating threshold or operating point of a circuit or device, such as a transistor, diode, or magnetic amplifier. Compare BIAS VOLTAGE.

bias current drift The ratio of a change in input bias current to a change in ambient temperature, generally expressed in nanoamperes per degree Celsius.

bias distortion Distortion caused by operation of a tube or transistor with incorrect bias so that the response of the device is nonlinear.

biased diode A diode having a dc voltage applied in either forward or reverse polarity. Current flows readily through the forward-biased diode; the reverse-biased diode appears as an open circuit. The biased diode is the basis of clippers, limiters, slicers, and similar circuits.

biased off In a circuit or device, the state of cutoff caused by application of a control-electrode bias. Examples include collector-current cutoff (when the dc base bias of a bipolar transistor reaches a critical value), and drain-current cutoff (when the dc gate bias reaches a critical value in a field-effect transistor).

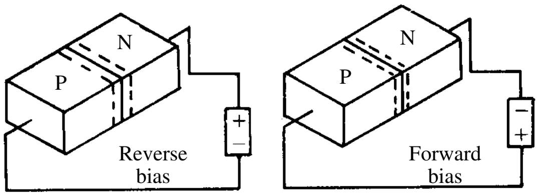

biased diodes

biased search A scheme that a mobile robot can use to find its way to a destination or target, by deliberately searching off to the side and then homing in as the approach progresses. It is so called because the general nature of the initial error (bias) is known, although its exact extent need not be known.

bias oscillator In a magnetic recorder, an oscillator operated at a frequency in the 40-kHz to 100-kHz range to erase prerecorded material and bias the system magnetically for linear recording.

bias resistor A usually fixed resistor, such as the source resistor in a field-effect-transistor (FET) circuit or the emitter resistor in a bipolar-transistor circuit, across which a desired bias voltage is developed by current flowing through the resistor.

bias set A control, such as a potentiometer or variable autotransformer, that facilitates manual adjustment of the dc bias of a circuit.

bias stabilization **1.** The maintenance of a constant bias voltage, despite variations in load impedance or line voltage. It is usually accomplished by means of automatic voltage regulation. **2.** The stabilization of transistor dc bias voltage by means of resistance networks or through the use of barretters, diodes, or thermistors.

bias supply **1.** Batteries that provide bias voltage or current for bipolar or field-effect transistors. **2.** A line-operated unit for supplying dc bias and consisting of a transformer, rectifier, and high-grade filter.

bias voltage A steady voltage that presets the operating threshold or operating point of a circuit or device, such as a transistor. Compare BIAS CURRENT.

bias windings The dc control windings of a saturable reactor or magnetic amplifier.

biconical antenna A form of broadband antenna, consisting of two conical sections joined at the apexes. The cones are at least ¼ wavelength in diagonal height. The vertex angles of the cones can vary, although the apex angle is usually the same in each cone. The vertex angle affects the feedpoint impedance. Such an antenna radiates, and responds optimally to, signals with polarization parallel to the axis of the cones.

biconical horn antenna A double-horn microwave antenna that radiates along relatively sharp front and back beams.

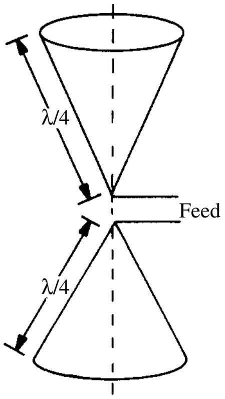

biconical antenna

bidecal base The 20-pin base of a cathode-ray tube. Also see DIHEPTAL, DUODECAL, and MAGNAL.

bidirectional Radiating or receiving (usually equally) from opposite directions (e.g., front-and-back radiation from an antenna or loudspeaker, or front-and-back pickup with an antenna or microphone).

bidirectional antenna An antenna with a directional pattern that consists of maximum lobes 180 degrees apart.

bidirectional bus In computers, a data path over which both input and output signals are routed.

bidirectional bus driver In a microcomputer, a signal-driving device that permits direct connection of a buffer-to-buffer arrangement on one end (the interface to I/O, memories, etc.) and data inputs and outputs on the other. This device permits bidirectional signals to pass and provides drive capability in both directions.

bidirectional counter A counter that can count consecutively up from a given number or down from that number. Also called UP-DOWN COUNTER.

bidirectional current A current that flows in both directions. Utility alternating current (ac) is a common example.

bidirectional loudspeaker A loudspeaker that delivers sound waves to the front and rear.

bidirectional microphone A microphone that picks up sound waves equally well from the front and rear.

bidirectional transistor A symmetrical transistor (i.e., one in which the two main current-carrying electrodes can be interchanged without influencing device performance). Some field-effect transistors (FETs) are of this type; the drain and the source can be interchanged.

bifilar electrometer An electrometer in which the sensitive element consists of two long platinized-quartz fibers. When an electric potential is applied, the fibers separate by a distance proportional to the voltage.

bifilar resistor A wirewound resistor with two oppositely wound filaments. The nature of the winding tends to cancel the inductance, making the device useful at a much higher frequency than an ordinary wirewound resistor.

bifilar transformer A transformer in which unity coupling is approached by interwinding the primary and secondary coils (i.e., the primary and secondary turns are wound side by side and in the same direction).

bifilar winding **1.** A method of winding a coil (such as a resistor coil) in the shape of a coiled hairpin so that the magnetic field is self-canceling and the inductance is minimized. **2.** A method of winding transformers to minimize leakage reactance.

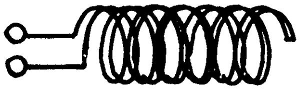

bifilar winding

bifurcated contact A forked contact whose parts act as two contacts in parallel for increased reliability.

bilateral amplifier An amplifier that transmits or receives in either direction equally well (i.e., the input and output can be exchanged at will).

bilateral antenna A bidirectional antenna, such as a loop antenna or a half-wave dipole.

bilateral element A circuit element or component (as a capacitor, resistor, or inductor) that transmits energy equally well in either direction. Compare UNILATERAL ELEMENTS.

bilateral network A network, usually passive and either balanced or unbalanced, that has BILATERAL SYMMETRY. Thus, the input and output terminals can be exchanged without affecting the performance of the network in any way.

bilateral symmetry **1.** Exhibiting symmetry, with respect to a vertical line or plane. **2.** For a network, having the property that if the input and output are reversed, the circuit behavior remains precisely the same. See BILATERAL NETWORK. **3.** For an amplitude-versus-frequency response curve, having the property that the right-hand and left-hand halves are mirror images of each other.

billboard antenna A phased group of dipole antennas that lie in one plane. A reflector might be used behind the entire array.

bilobe pattern An antenna radiation pattern consisting of two major lobes in a given plane, usually the horizontal plane. Often the lobes exist in opposite directions relative to each other, as in a half-wave dipole. But they can be at varying angles, as in a long-wire antenna.

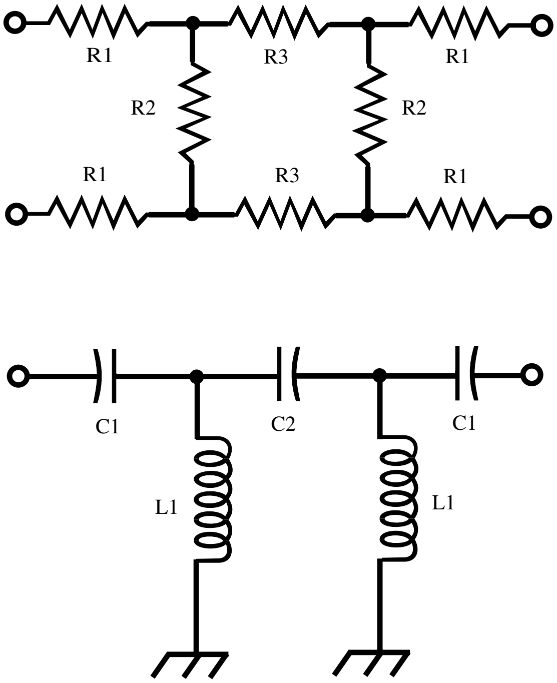

bilateral network

bimetal A union of two dissimilar metals—especially those having a different temperature coefficient of expansion. The two are usually welded together over their entire surface.

bimetallic element A strip or disk of bimetal. When the element is heated, it bends in the direction of the metal that has the lower temperature coefficient of expansion; when cooled, it unbends. Usually, an electrical contact is made at one extreme or the other so that the element can serve as a thermostat.

bimetallic switch A temperature-sensitive switch based on a bimetallic element.

bimetallic thermometer A thermometer based on a bimetallic element that is mechanically coupled (as through a lever and gear system) to a pointer that moves over a temperature scale.

bimetallic thermostat A thermostat in which a bimetallic element closes or opens a pair of switch contacts.

bimorphous cell A piezoelectric transducer that consists of two crystal plates, such as Rochelle salt, bound intimately face to face. In a crystal microphone, vibration of the transducer results in a voltage output; in a crystal headphone, an ac signal voltage impressed on the transducer causes vibratory mechanical motion.

BiMOS A combination of bipolar and MOSFET transistors in an integrated circuit. Thus, a typical BiMOS device can have MOSFET input for high impedance and bipolar output for low impedance.

binant electrometer An electrometer in which a thin platinum vane ("the needle") is suspended within two halves of a metal pillbox-shaped container. The halves or binants are biased with a dc voltage of 1 to 12 V, and the unknown voltage is applied to the vane. It is also called DUANT ELECTROMETER and HOFFMAN ELECTROMETER.

binary **1.** Pertaining to the base-2 number system. Thus, binary arithmetic uses two digits: 0 and 1. **2.** Pertaining to two-element chemical compounds.

binary arithmetic Mathematical operations performed using only the digits 0 and 1.

binary cell In a computer memory, an element that can display either of two stable states.

binary chain A cascade of binary elements, such as flip-flops, each unit of which affects the stable state of the succeeding unit in sequence.

binary channel Any channel whose use is limited to two symbols.

binary code A system of numbers representing quantities by combinations of 1 and 0; a binary-number system.

binary-coded decimal notation In digital computer operations, a system of notation in which each digit of a decimal number is represented by its binary equivalent. Thus, the decimal number 327 in BCD notation becomes 0011 0010 0111. (By contrast, in pure binary notation, 327 is 101000111.)

binary-coded octal notation A method of numbering in which each base-8 digit is represented by a binary number from 000 to 111.

binary-controlled gate circuit A gate circuit controlled by a binary stage. An example is a gating transistor that receives its on/off pulses from a flip-flop.

binary counter A counter circuit consisting of a cascade of bistable stages. Each stage is a scale-of-two counter because its output is on for every second input pulse. At any instant, the total binary count in a multistage counter thus is shown by the on and off states of the various stages in sequence.

binary decoder A device or stage that accepts binary signals on its input lines, and provides a usually exclusive output (representing a decimal digit, for example).

binary digit See BIT.

binary number system The base-two system of notation. This system uses only two symbols, 0 and 1, and accordingly is easily applied to two-position switches, relays, and flip-flops.

binary preset switch In a binary counter or binary control circuit, a selector switch that allows the circuit to be preset to deliver an output pulse only after a predetermined number of input pulses.

binary relay See BISTABLE RELAY.

binary scaler In its simplest form, a single two-stage device, such as a flip-flop, which functions as a divide-by-two counter, because one output pulse results from every two input pulses. Higher-order scaling is obtained by cascading stages.

binary search A system of search entailing the successive division of a set of items into two parts and the rejection of one of the two until all items of the sought-for kind are isolated.

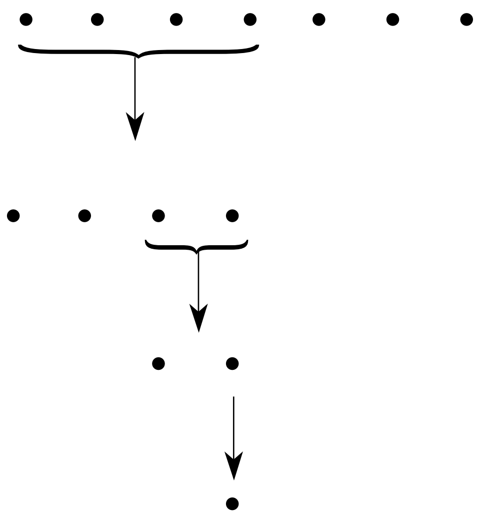

binary search

binary signal Any signal that can attain either of two states. Such a signal is always a digital signal.

binary-to-decimal conversion **1.** The automatic conversion of a number represented by a series of binary pulses into the corresponding decimal number, which then is displayed by a readout device. **2.** The arithmetic operation of converting a binary number into a decimal number; this can be done by noting the powers of 2 represented by the various binary digits in a number, and then adding the decimal values of these powers.

binary word A binary numeral that has a particular meaning, agreed upon by convention. For example, the letters A through Z can be represented by binary numbers 00001 through 11010; a word can be represented by several blocks of five digits.

binaural Literally, two-eared. In sound recording and reproduction, the transcription of a broad sound source using two microphones spaced at approximately the distance between the ears on a human head, and played back using headphones to re-create the stereo effect. The technique evolved into multichannel stereophonic reproduction.

binaural machine hearing Also called *stereo machine hearing.* The ability of a machine, such as a robot, to sense the direction and distance to a source of sound, using two acoustic transducers and a computer to process their output signals. The machine determines the location of the sound source by comparing the relative amplitude and phase of the signals from the two transducers. It functions according to the same principle as human hearing, in which a person can determine the general direction and distance to a sound source by subconsciously comparing the relative amplitude and phase of the sounds arriving at the left and right ears.

binaural sound The equivalent of a listener hearing a concert through a pair of earholes; it takes earphones to reproduce the signal. If speakers are substituted for the earphones, the listener hears monophonically, as if standing back several feet from the earholes.

binder A material (such as lacquer) that acts as a holder and cohesive medium for the particles of another material. It is used in carbon resistors, ceramic dielectric bodies, powder cores, and resistive and metallic paints.

binding energy A property of the nucleus of an atom. The binding energy of a nucleus is equal to the difference between the nuclear weight and the sum of the weights of the lighter particles making up the nucleus. The nucleus is stable when the binding energy is high.

binding force Any one of the electrostatic forces that bind crystals together.

binding post A screw-type terminal of various styles, often having a hole into which a wire or tip can be inserted and gripped. It is used for temporary indoor connections only.

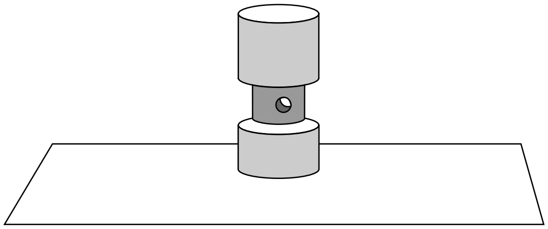

binding post

binistor A semiconductor switching device that exhibits two stable states and also negative resistance.

binocular machine vision Also called *stereoscopic machine vision.* The ability of a machine vision system to provide depth and perspective data. Uses two optical sensors spaced a fixed distance apart. The left sensor sees a slightly different image than the right sensor. These two images are combined and processed by a computer, allowing the machine (such as a mobile robot) to determine the distances to various objects in its environment. Functions on the same principle as stereoscopic human vision.

bin picking In robotics, the selection of a particular object from a container (bin) in which there are many objects. Can be done using object recognition, bar coding, or passive transponders. It requires a sensor, operating in conjunction with a computer that processes the sensed data and controls the movements of the robot.

binomial An algebraic expression containing two terms joined by a plus or minus sign. Examples: $a^2 + b^2$, $3x^3 - 6x$.

binomial theorem The theorem, proven by Isaac Newton, permits a binomial to be raised to any desired power without performing the multiplications. In electronics, *power series* are convenient for expressing such expressions.

biochemical cell A fuel-cell energy source in which electricity is generated chemically through the oxidation of biological substances. Also called *biochemical fuel cell.*

biochip **1.** A natural, living organism with a physical structure that in some way resembles that of an electronic *integrated circuit (IC).* **2.** A theoretical possibility, according to some scientists, but not yet a practical reality: An IC manufactured by a laboratory process that mimics the way in which nature builds living organisms. A form of *artificial life,* harnessed for electronic and/or computing applications.

bioelectricity **1.** Electric currents in living tissues, generated by the organism and not applied by external means. **2.** The science or study of such currents.

bioelectrogenesis The study and application of electricity generated by living animals, including humans, in the powering and control of electronic devices.

bioelectronics Electronics in relation to the life sciences—especially the electronic instrumentation of biological experiments.

bioengineering **1.** The engineering of equipment, such as electron microscopes, electroencephalographs, centrifuges, irradiators, etc., for study and experimentation in the life sciences. **2.** The engineering of equipment, such as pacemakers, hearing aids, X-ray apparatus, shock-therapy units, etc., for aid or support-of-life processes.

biofeedback A technique in which changes in skin temperature and resistance are detected and displayed by an electronic device.

biofeedback monitor A system that provides an indication of skin temperature and resistance to a user. Because skin temperature and resistance are affected by emotions, such as fear, nervousness, anger, etc., these monitors might be of value to people who wish to gain improved control of their emotions, and thus perhaps minimize the physiological effects of stress.

biological robot Believed by some researchers to be possible, but not yet a practical reality: A living organism created by biological cloning, whose brain has been programmed exactly as a computer is programmed.

biological shield An absorbent shield that blocks or attenuates ionizing radiation to protect personnel working near radioactive materials.

bioluminescence **1.** The emission of light by a living organism. **2.** The light itself so produced by living organisms.

biomechanism An electromechanical device that simulates the workings of some part of a living being's body. Examples are electromechanical hands, arms, and legs. Such a device is often difficult to distinguish from its biological counterpart when obscured by clothing.

biomechatronics A contraction of the words *bio*logy, *mecha*nics and elec*tronics.* Research, development and manufacturing that encompasses aspects of all three fields. This is especially important in robotics.

biometrics Mathematics, and in particular, statistics and probability, applied to biology.

biometric security system An advanced intrusion-prevention system that measures biological characteristics of the people who are authorized to enter a property. Such a machine can employ vision systems, object recognition, and/or pattern recognition to check a person's face. The machine might use speech recognition to identify people by the waveforms of their voices. It might record a hand print, a fingerprint, or an iris print, or a combination of all these things. A powerful computer analyzes the data obtained by the sensors and determines whether the person is authorized to enter the premises.

bionics The study, design, and application of microelectronic systems that simulate the functions of living organisms.

biotelemetry The use of telemetry to collect data from living organisms or to direct their movement.

biotelescanner An instrument that monitors body functions via radio, from a great distance.

Biot/Savart law A principle of electromagnetism that expresses the intensity of magnetic field H in the vicinity of a long, straight wire carrying a steady current I. The basic formula is $H = 2I/r$, where H is in oersteds, I is in amperes, and r is the distance in centimeters from the wire.

bip Abbreviation of *binary image processor.*

biphase half-wave rectifier An alternative term for FULL-WAVE RECTIFIER; also, each leg of a two-diode full-wave rectifier.

BIPM Abbreviation of *International Bureau of Weights and Measures.*

bipolar The condition of possessing two pole sets. In a conventional (non-FET) transistor, one pole set exists between the base and collector, and another pole set exists between the base and emitter.

bipolar driving unit A magnetic headphone or loudspeaker in which both poles (north and south) of a magnet actuate a diaphragm or lever.

bipolar operation See AUTOMATIC POLARITY.

bipolar transistor A two-junction transistor whose construction takes the form of a pnp or an npn "sandwich." Such devices are current-operated, compared with field-effect transistors, which are voltage-operated. The bipolar transistor (of which the familiar npn and pnp types are examples) uses both electron and hole conduction.

biquinary code A variety of binary-coded-decimal notation in which seven bits are used to represent each decimal digit. A number is written in two groups of bits: a two-bit group followed by a five-bit group. The positional values are 5 and 0 for the two-bit group, and 4, 3, 2, 1, and 0 for the five-bit group.

biquinary decade A decade counter that consists of a binary stage, followed by a quinary stage.

bird **1.** Slang for orbiting SATELLITE. **2.** Slang for *guided missile.*

birdie **1.** A spurious beat note in a superheterodyne receiver. So called because of the characteristic chirping sound it makes as the operator tunes by the frequency on which it occurs. **2.** A parasitic oscillation in a radio transmitter, also called a *spurious emission* or *spur.*

Birmingham wire gauge Abbreviation, BWG. Also called *Stubs gauge.* A method of designating the various sizes of solid wire. BWG diameters are somewhat larger than corresponding AMERICAN WIRE GAUGE diameters for a given wire-size designator.

Birmingham Wire Gauge (BWG) Diameters

BWG	Millimeters	Inches
1	7.62	0.300
2	7.21	0.284
3	6.58	0.259
4	6.05	0.238
5	5.59	0.220
6	5.16	0.203
7	4.57	0.180
8	4.19	0.165
9	3.76	0.148
10	3.40	0.134
11	3.05	0.120
12	2.77	0.109
13	2.41	0.095
14	2.11	0.083
15	1.83	0.072
16	1.65	0.064
17	1.47	0.058
18	1.25	0.049
19	1.07	0.042
20	0.889	0.035

bismuth Symbol, Bi. A metallic element. Atomic number, 83. Atomic weight, 209.

bismuth flux meter A flux meter in which the sensor contains a length of bismuth wire, which acts as a magnetoresistor.

bismuth thermocouple A thermocouple that uses the junction between bismuth and antimony wires. Used in thermocouple-type meters.

bistable Having two stable states.

bistable device Any device, such as a flip-flop, the operation of which exhibits two stable states and which can be switched at will from one state to the other.

bistable multivibrator A multivibrator, the operation of which exhibits two stable states. More commonly known as a FLIP-FLOP. These circuits are abundant in digital electronic equipment. Compare ASTABLE MULTIVIBRATOR and MONOSTABLE MULTIVIBRATOR.

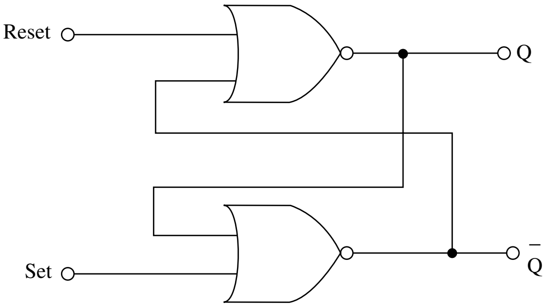

bistable multivibrator

bistable relay A relay that has two stable states: open and closed. Successive actuating pulses open and close the relay, two consecutive pulses being required to return the relay to a given state. Also called *binary relay*, *relay flip-flop*, and *electromechanical flip-flop.*

bistatic radar A radar set in which the transmitting and receiving antennas are separate.

bistate Having two states. Example: the performance of a FLIP-FLOP.

bit An acronym formed from the words *binary digit.* The smallest or elementary unit of data in digital electronics. Represented either by logic 0 (low) or logic 1 (high). These states can be represented by any dichotomy, such as off/on, false/true, minus/plus, dark/bright, red/green, etc.

BIT Abbreviation of *built-in test.*

bit density The number of digital bits per unit area or volume, as the number of bits per square centimeter of magnetic tape.

BITE Abbreviation of built-in test equipment.

bit rate The speed in BITS PER SECOND (bps) at which digital data bits are transmitted or handled.

bit-slice processor A microprocessor whose word or byte capacity is achieved through the use of interrelated smaller capacity processors (e.g., a 16-bit unit derived from eight 2-bit “slices”).

bits per second Abbreviation, bps. An expression of digital data speed. Commonly used in computer communications. This unit is often confused with, and improperly called, the *baud.* There is generally a difference between the speed of a signal in baud, and the speed of the same signal in bps. Compare BAUD.

bitter pattern A pattern produced in a suspension of ferromagnetic powder in the presence of an imperfection in a magnet. The pattern appears as an irregularity that is easy to see.

Bjerknes’ equation An expression for the total (primary plus secondary) decrement of a tuned circuit, based on measurements of the tank current at the resonant frequency and at a frequency near resonance.

BK **1.** Radiotelegraph signal for BREAK. **2.** Abbreviation of BREAK-IN.

Bk Symbol for BERKELIUM.

black-and-white Also called *monochrome* and *gray-scale.* Any system of image reproduction, transmission, or reception in which the image is composed of opaque elements (black) and white or bright areas, as in noncolor television reception.

black area An area in which there is only an encrypted signal.

blackboard system A method via which computers can recognize, and to some extent determine the meaning of, spoken words and visual images. Incorporates machine vision and/or machine hearing in conjunction with artificial intelligence (AI). Incoming voices and/or images are digitized and entered into a large-capacity random-access memory (RAM). The data is evaluated by sophisticated software to determine the most logical or probable interpretations of the sounds and images.

blackbody An ideal surface or object, that completely absorbs energy of any wavelength that strikes it. Such an object is a theoretically perfect radiator of energy at all wavelengths.

blackbody radiation Electromagnetic radiation from a heated ideal BLACKBODY. This radiation is conceived as covering the entire ELECTROMAGNETIC FREQUENCY SPECTRUM. It can be expressed graphically as a characteristic curve with a peak at a wavelength that depends on the absolute temperature of the object. As the absolute temperature increases, the peak occurs at progressively shorter wavelengths (higher frequencies). This enables radio astronomers to get a reasonably good idea of the temperatures of distant celestial objects, such as planets.

black box **1.** Any “box” or “block” that can be included in an analysis or synthesis based upon the BLACK-BOX CONCEPT. **2.** Any functional unit (such as a module) whose operating characteristics are known, and that can be inserted into a system in development or maintenance operations. **3.** Any subcircuit or stage that can be specified in total as required in a system, in terms of its known or prescribed performance, but whose internal structure need not be known.

black-box concept A technique for development of equivalent circuits and of considering their operation. The "box" has a pair of input terminals and a pair of output terminals; one input terminal is often common to one output terminal. The contents of the box need not be known, but from the input and output current and voltage relationships, its nature can be determined. Moreover, from the available input signal and desired output signal, the internal circuit of the box can be specified. Integrated circuits (ICs) are often treated as black boxes by engineers designing complex electronic equipment.

black compression Attenuation of the level of dark areas in a television picture.

blacker than black The video-signal amplitude region above the level that just darkens the screen. Signal information (such as control pulses) in this region are therefore not seen.

black light **1.** Ultraviolet radiation—especially when used to cause visible fluorescence in certain materials. **2.** A lamp that produces a principal portion of its radiation in the ultraviolet region, causing visible fluorescence of certain substances. Such lamps are used in some scientific experiments, and also for creating special effects at presentations or parties. It is hazardous to look directly at the output of such a lamp with unprotected eyes.

blackout **1.** A complete interruption of ac utility power to numerous customers at the same time. **2.** A complete cessation of ionospheric radio-wave propagation, such as might be caused by a solar flare. **3.** Complete blanking of the screen of an oscilloscope or picture tube.

black reference In a television signal, the blanking level of pulses, beyond which the sync pulse is in the *blacker-than-black* region.

black reference level In a television signal, the voltage threshold of the BLACK REFERENCE (i.e., its level above zero volts).

black transmission A system of picture or facsimile transmission in which the maximum copy darkness corresponds to the greatest amplitude (in an amplitude-modulated transmitter) or the lowest instantaneous frequency (in a frequency-modulated transmitter). Compare WHITE TRANSMISSION.

blank **1.** A piezoelectric plate cut from a quartz crystal, but not yet finished to operate at a desired frequency. **2.** To obscure or interrupt a signal or electron beam (usually momentarily), as in z-axis blanking in an oscilloscope. **3.** A silicon wafer cut from a large slab, containing dopants only. **4.** A magnetic diskette or tape on which nothing is recorded. **5.** An optical diskette on which nothing is recorded. **6.** A location (such as a symbol or space) that is used to verify proper data character grouping and values.

blanketing A form of radio interference accompanied by severe degradation of reception, virtually unaffected by tuning, over a wide range of frequencies. An example is ac line noise caused by an arcing power transformer or electrical appliance in the vicinity of a receiving antenna. It tends to occur most often at low, medium, and high frequencies.

blanking Obscuring or momentary elimination of a signal (see BLANK, **2**).

blanking interval The short period during which the electron beam of a cathode-ray tube is cut off so that the beam can return to its start position without creating a trace on the screen.

blanking level The discrete, predetermined level (usually a threshold voltage) at which BLANKING occurs.

blanking pedestal In the horizontal pulse of a television signal, the lower portion between zero volts and the blanking level.

blanking pulse A pulse that produces momentary blanking (see BLANK, **2**).

blanking time The time interval during which the electron beam of a cathode-ray tube is interrupted by a blanking signal.

blank tape **1.** Magnetic tape that has never been subjected to the recording process and that is substantially free from noise. **2.** Magnetic tape from which all preexisting information has been erased.

blasting **1.** Severe overloading of a sound system, usually caused by setting the volume control at or near maximum and then applying a significant input signal to the amplifier. Accompanied by distortion, in its worst form, it can cause damage to speakers and/or headsets. **2.** In a communications receiver, the result of a strong signal coming in unexpectedly when the automatic gain control (AGC) has been switched off, and the audio-frequency (AF) and radio-frequency (RF) gain controls are set high for reception of weak signals.

bleeder A resistor or group of resistors, used permanently to drain current from charged capacitors. It establishes the predetermined initial load level for a power supply or signal source, and it serves as a safety device in high-voltage power supplies.

bleeder current The current normally flowing through a bleeder.

bleeder divider A network of resistors, series-strung across the output of a power supply or its regulator. As a load resistor, the bleeder improves regulation and protects against no-load voltage surges. The resistor junctions allow various voltages to be drawn from the supply.

bleeder power Power dissipated as heat in a bleeder.

bleeder resistor See BLEEDER.

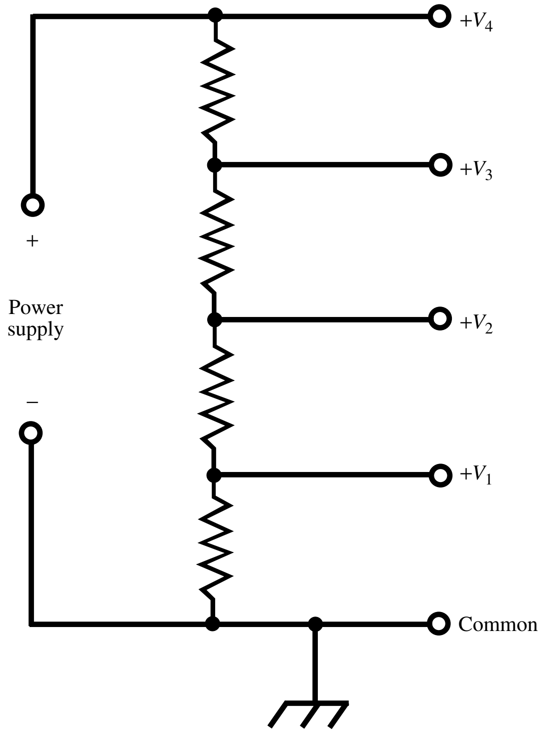

bleeder divider

bleeder temperature The operating temperature in a bleeder. It is generally high because of power dissipation in the form of heat.

bleeding whites A flowing of the white areas of a television picture into the black areas; an overload condition.

blemish See BURN.

blind flight The flying of aircraft entirely by means of instruments and electronic communications.

blind landing Landing of an aircraft entirely by means of instruments and electronic commnications.

blind zone **1.** In radar operations, an area that gives no echoes. **2.** Skip zone (see ZONE OF SILENCE).

blip **1.** The pulse-like figure on a radar scan, indicating the transmission or reflection (see A-SCAN and J-SCAN). Also called PIP. **2.** In visual alignment of a tuned circuit using a sweep generator and marker generator, the pulse or dot produced on the response curve by the marker signal. **3.** A short, momentary signal pulse, such as a single Morse dot.

BLIP Abbreviation for *background-limited infrared photoconductor.*

blip-scan ratio The number of radar scans necessary to show a visible blip, or echo, on a radar screen.

Bloch functions Solutions of the *Schrodinger wave equation* for a single electron surrounded by an electric field. The field varies periodically with distance from the source.

Bloch wall The transition layer between adjacent ferromagnetic domains (see DOMAIN).

block **1.** A group of data words or digits. **2.** A group of memory storage spaces. **3.** A circuit that operates as an identifiable unit. **4.** The symbol for a circuit, stage, unit, or device in a BLOCK DIAGRAM.

block diagram A simplified diagram of an electronic system, in which circuits, stages, units, or devices are shown as two-dimensional boxes with the internal wiring and detail circuitry omitted. This makes it possible to clearly show the interconnection among circuits, stages, units or devices. It also provides a concise rendition of the overall functional concept of the system.

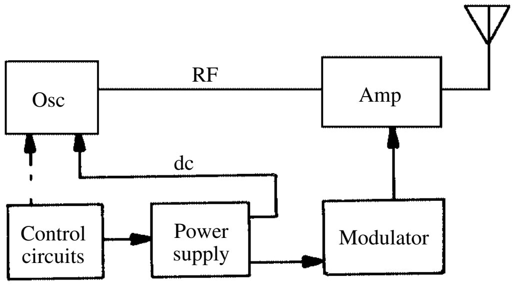

block diagram
(of a radio transmitter)

blocked impedance The input impedance of a transducer, whose output load is a theoretically infinite impedance.

blockette In a computer, the subdivision of a character block that is handled as a unit during data transfer.

blocking action Obstruction of circuit action, usually abrupt, through internal action or by the application of an external signal. Thus, the operation of an amplifier can be blocked (output reduced to zero) by an input signal or by excessive feedback, either of which overloads the input.

blocking capacitor A capacitor inserted into a circuit to prevent the passage of direct current while easily passing alternating current.

blocking choke Any inductor, such as a choke coil, that is used to prevent the flow of an alternating current while allowing direct current to pass with little resistance.

blocking interference Radio interference from signals strong enough to reduce the receiver output through blocking action.

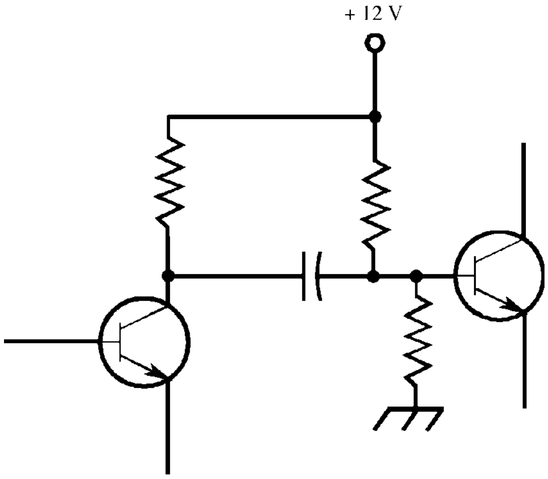

blocking capacitor

blocking oscillator An oscillator that turns itself off after one or more cycles. It does this as a result of an accumulation of negative charge on its input electrode (base of a bipolar transistor or gate of a field-effect transistor). The action is repetitive. In the self-pulsing type of blocking oscillator, a series of pulses consisting of trains of sine waves with intervening spaces is generated. In the single-swing type of blocking oscillator, the output consists of a series of single cycles with long intervals between them.

blocking oscillator synchronization **1.** In the BLOCKING OSCILLATOR used in the vertical deflection circuit of a television receiver, the oscillator is synchronized with vertical sync pulses arriving in the video signal. **2.** Synchronization of the repetition rate of any blocking oscillator with a suitable external control signal.

blocking system In a telephone system, a method of dealing with the condition of having more subscribers than connection paths. Allocation is made on a demand basis. If all channels are in use, it is impossible to make new calls. This prevents excessive degradation of the quality of existing connections.

block length The number of characters, bits, or words that compose a defined unit word or character group.

block transfer The conveyance of a word or character grouping in a computer register to another register or a peripheral device.

blooming On a cathode-ray-tube (CRT) screen, an enlargement of the electron-beam spot, caused by poor focusing. This results in poor image resolution.

blooper **1.** A radio receiver that is in oscillation, and is transmitting a signal that causes interference. **2.** A parasitic oscillation in a radio transmitter. **3.** In broadcasting, a statement in which a radio or television announcer makes an embarrassing error or breach of etiquette.

blow The opening of a fuse or circuit breaker as a result of excessive current.

blower A fan used to remove heat from electronic circuits. These are often used in tube-type radio-frequency (RF) power amplifiers, where much heat is generated, and in computers to cool the microprocessor and surrounding components.

blowout **1.** An alternate term for BURNOUT. **2.** The forceful opening of a circuit breaker. **3.** The extinguishing of an arc.

blowout coil An electromagnet that provides a field to extinguish an arc.

blowout magnet A permanent magnet that provides a field to extinguish an arc.

blst Abbreviation of *ballast.*

blue-beam magnet In a color television picture-tube assembly using three electron guns, a small permanent magnet to adjust the static convergence of the beam for blue phosphor dots.

blue box An accessory device (sometimes unlawfully used) that generates tones that switch a telephone circuit in the placing of calls.

blue glow **1.** In a neon lamp, a bluish light that results from high-voltage arcing. **2.** The normal color of the gas discharge in an argon glow lamp. **3.** The bluish glow between anode and cathode of a gassy vacuum tube. **4.** The normal color of the discharge that fills a mercury-vapor tube.

blue gun The electron in a three-gun color picture tube, the beam from which strikes the blue phosphor dots.

blueprint **1.** A type of contact-print reproduction in which a sheet of sensitized paper is exposed to an image on a translucent or transparent film, under strong light, and is then developed and fixed. Although this process is still used to reproduce electronic illustrations and typescripts, it has been superseded largely by other (dry) processes. **2.** Loosely, any plan or design for the development of a system.

blue restorer In a three-gun color television circuit, the dc restorer in the blue channel.

blue ribbon program A computer program that has been hand-prepared and debugged completely before its first computer run.

blue video voltage The signal voltage presented to the grid of the blue gun of a three-gun color picture tube.

blurring **1.** BLOOMING. **2.** A defocusing of a television picture or oscilloscope trace. **3.** An obscuring of a signal by echoes or trailing (e.g., the slow decrement of a Morse code signal element).

B-minus Also called *B-negative.* The negative terminal of a B-power supply.

BNC Abbreviation of *bayonet Neill-Concelman.* A type of coaxial connector that can be quickly con-

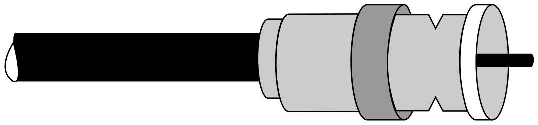

BNC

nected and disconnected. It is commonly used with test equipment.

B-negative Alternative expression for B-MINUS.

BNL Abbreviation of *Brookhaven National Laboratory.*

BO Abbreviation of *beat oscillator.* Also abbreviated BFO.

board **1.** A panel containing patch jacks. **2.** A printed circuit.

boat A type of crucible in which a semiconductor material is melted and sometimes processed. The material of which the boat is made (e.g., graphite) does not react with or contaminate the semiconductor material.

bobbin **1.** A usually nonmetallic spool on which a coil is wound. **2.** The form onto which the voice coil of a loudspeaker is wound.

Bode plot A pair of curves plotted to the same frequency axis, one showing the gain of a network or amplifier and the other showing its phase shift. Phase and amplitude of active and passive networks can be exhibited. Also called *Bode curve* and *Bode diagram.*

body-antenna effect The tendency of the human body to act as a receiving antenna when a finger is touched to the antenna input terminal of a receiver or when a hand (or the whole body) is brought close enough to the circuit to provide capacitive coupling.

body capacitance Capacitance between the body of the operator (as one plate of an equivalent capacitor) and a piece of electronic equipment (as the other plate). This phantom capacitance is often the cause of detuning and of the injection of interfering signals and noise because the body acts as a pickup antenna.

body electrode **1.** An electrode attached to the human body (or to the body of a laboratory animal) to conduct body-generated currents to an instrument, as in cardiography, electroencephalography, and myography. **2.** An electrode attached to the human body (or to the body of a laboratory animal) to conduct currents into the body, as in shock therapy and skin-resistance measurement.

body leakage Leakage of current through the bulk or body of a dielectric material, as opposed to SURFACE LEAKAGE.

body temperature In a thermistor, a rating that represents the temperature measured on the surface of the device. It is any combination of ambient temperature, power dissipation, and operation of the internal heater element (if the thermistor has one).

bof Abbreviation of *barium oxide ferrite.*

boffle A loudspeaker enclosure consisting of stretched screens that are sound absorbing and elastic.

bogie Also called *bogey.* **1.** The exact value of a specified characteristic. Thus, if resistance is given as 1 kΩ ±0.5%, the bogie value is 1 kΩ. **2.** The average value (i.e., the ARITHMETIC MEAN). **3.** A false or unidentified echo on a radar screen.

Bohr atom The concept of the nature of the atom, proposed by Niels Bohr in 1913 partly to explain why the electrons in the Rutherford atom do not fly off into space or fall into the nucleus. The Bohr theory places the electrons in permissible orbits where they cannot radiate energy (see BOHR RADIUS). They can radiate or absorb energy, however, if they go to a lower orbit or to a higher orbit, respectively. Compare RUTHERFORD ATOM.

bohrium Symbol, Bh. Also called *unnilseptium* (Uns). Atomic number, 107. The most common isotope has atomic weight 262. Classified as a transition metal. It is human-made and is not known to occur in nature.

Bohr radius Symbol, a_0. A physical constant whose value is approximately 5.291772×10^{-11} meter.

boiling point Abbreviation, bp. The temperature at which a liquid vaporizes. The boiling point of water in air at a pressure of one atmosphere is 100°C or 212°F.

bolometer Any device that is essentially a small, nonrectifying, temperature-sensitive resistor that can be used for heat sensing, radio-frequency power measurement, curve changing, demodulation, circuit protection, etc. Included in this category are the BARRETTER, the THERMISTOR, and the wire-type FUSE.

bolometer bridge A dc bridge in which a bolometer is one of the four arms. The bridge is balanced first with the bolometer cold. The bolometer then is excited with a radio-frequency (RF) current, whereupon the resultant heating changes the bolometer resistance. The bridge is rebalanced for the new resistance. The RF power driving the bolometer is determined according to a predetermined function of bridge settings versus RF input power.

Boltzmann constant Symbol, k. A figure that enters into the calculation of thermionic emission and of thermal noise factor. It represents the temperature equivalent of work function, in electron volts per Kelvin (eV/K) or joules per Kelvin (J/K). The values are approximately:

$$k = 8.617 \times 10^{-5} \text{ eV/K} = 1.38 \times 10^{-23} \text{ J/K}$$

Boltzmann's principle A description of the statistical distribution of large numbers of tiny particles under the influence of a force, such as an electric or magnetic field. When the system is in statistical equilibrium, the number of particles in any portion of the field is given by:

$$N_{\mathrm{E}} = N_{0e}^{-E/kT}$$

where E is the potential energy of a particle in the observed area, N_0 is the number of particles per unit volume in a part of the field where E is zero, k is the BOLTZMANN CONSTANT, T is the absolute temperature of the system of particles, and e is approximately equal to 2.718.

bombardment The usually forceful striking of a target with rays or a stream of particles.

bond **1.** An area in which two or more items are securely and intimately joined. **2.** The attractive force that holds an atomic or subatomic particle or particle group together.

bonded-barrier transistor A bipolar transistor in which the connection at the base region is alloyed.

bonded negative-resistance diode A diode that displays a negative-resistance characteristic over part of its current curve. This results from avalanche breakdown.

bond energy In a molecule, the energy necessary to break an atomic bond.

bonding **1.** The formation of bonds between adjacent atoms in a crystalline material, such as a semiconductor. See specifically COVALENT BINDING FORCES, IONIC BINDING FORCES, and METALLIC BINDING FORCES. **2.** The secure fastening together of conducting surfaces, as by soldering or brazing, to produce a high-conductance, leak-free continuum.

bond strength The minimum stress required to separate a material from another to which it is bonded.

bone-conduction transducer A device used in place of the earphone in a hearing aid to convey sound energy to the bone structure of the head.

Bongard problem A method of evaluating how well a machine vision system can differentiate among patterns. Similarities and differences are noted between objects in two sets of boxes. It was developed for object-recognition systems, mainly for use in intelligent robots.

book capacitor A variable capacitor in which the metal plates are bonded along one edge and separated from each other by means of mica sheets. The capacitance is varied by opening and closing the assembly book fashion. It is used as a padder or trimmer.

Boolean algebra A system of symbolic logic. Statements are represented as symbols, usually variables such as x, y, and z. The logical AND operation is represented by multiplication; the logical inclusive OR operation is represented by addition; the logical NOT operation is represented by a minus sign or a line over the element symbol. The system has rules, definitions and axioms via which *theorems* can be derived. Used by engineers in the design of digital electronic circuits.

Boolean truth table

		AND	*NOT*	*OR*
x	y	xy	x	$x + y$
0	0	0	1	0
0	1	0	1	1
1	0	0	0	1
1	1	1	0	1

Boolean function In mathematical logic, a function that makes use of BOOLEAN ALGEBRA.

Boolean theorems

1. $x + 0 = x$ (additive identity)
2. $x1 = x$ (multiplicative idenity)
3. $x + 1 = 1$
4. $x0 = 0$
5. $x + x = x$
6. $xx = x$
7. $(x')' = x$ (double negation)
8. $x + x' = 1$
9. $x'x = 0$
10. $x + y = y + x$ (commutativity of addition)
11. $xy = yx$ (commutativity of multiplication)
12. $x + xy = x$
13. $xy' + y = x + y$
14. $x + y + z = (x + y) + z = x + (y + z)$ (associativity of addition)
15. $xyz = (xy)z = x(yz)$ (associativity of multiplication)
16. $x(y + z) = xy + xz$ (distributivity)
17. $(x + w)\,(y + z) = xy + xz + wy + wz$ (distributivity)

boom **1.** A horizontal support for a microphone, enabling the microphone to be suspended over a sound source, but out of the sight of a camera. **2.** A horizontal support for a small antenna that is undergoing tests or sampling the field of another antenna. **3.** The supporting element in a Yagi, quad, or log-periodic antenna. It establishes the center of gravity and directional axis of the radiation pattern. The driven element(s) and parasitic element(s) are attached, usually at right angles.

boost capacitor In the damper circuit of a television receiver, the capacitor that is used to boost the B-plus voltage. Also called *booster capacitor.*

boost charge A high-current, short-interval charge used to revitalize a storage battery quickly. Also called *booster charge.*

booster **1.** Any device used to increase the amplitude of a signal (e.g., as an amplifier or preamplifier) or of an energy source (e.g., to boost the output of a power supply). **2.** A radio-frequency preamplifier used ahead of a television receiver.

booster battery **1.** A battery used to forward bias a diode detector into a favorable region of its conduction curve, or to bias a bolometer into the square-law region of its response. **2.** A battery supplying power to a booster.

booster gain The amplification (usually in terms of voltage gain) provided by a booster (see especially BOOSTER, **2**).

boot **1.** The powering-up routine in a digital computer, in which the machine executes a series of programs to get itself ready for use. **2.** The resetting of a computer, by pressing certain keyboard keys (e.g., CTRL-ALT-DEL), pressing a reset button, or by powering-down, waiting about two minutes, and then powering-up again. **3.** To install a computer diskette and instruct the computer to execute one or more programs on the diskette. **4.** A usually flexible protective nipple or jacket pulled over a cable or connector, so called from its resemblance to a foot boot.

boot loader A form of computer program that operates on the BOOTSTRAP ROUTINE.

bootstrap A technique for making a device or process achieve a condition through its own actions; see BOOTSTRAP CIRCUIT, for example.

bootstrap circuit A specialized form of follower circuit that presents very high input impedance. Its chief feature is the return of the control-element resistor to a tap on the source or emitter resistor. The technique takes its name from the figurative notion that such a circuit "lifts its input impedance by its own bootstraps."

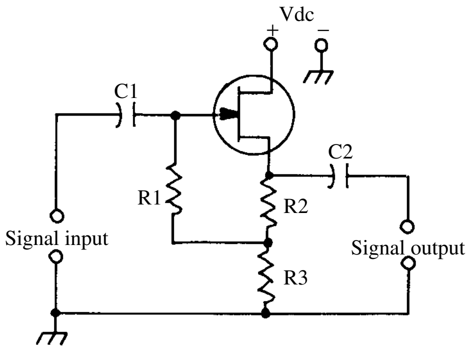

bootstrap circuit
(with junction-type field-effect transistor)

bootstrap routine **1.** Also called *bootstrap program.* In a digital computer, and especially in a personal computer, the routine that the machine follows when first powered-up. See BOOT, **1.** **2.** In a digital computer, a routine in which the first few instructions put in storage are later used to complete the routine, as supplemented by some operator instruction. **3.** A portion of a computer program that is used to establish an alternate version of the program.

borax-aluminum cell An electrolytic cell that consists essentially of an aluminum electrode and a lead electrode in a saturated solution of sodium tetraborate (borax). After electroforming, such a cell can be used either as a rectifier or as an electrolytic capacitor.

boric acid Formula, H_3BO_3. A compound used variously in electronics—especially as the electrolyte in electrolytic capacitors.

bornite Formula, Cu_5FeS_4. A natural mineral that is a sulfide of copper and iron. Its crystalline structure made it important in early semiconductor diodes (crystal detectors).

boron Symbol, B. A metalloidal element. Atomic number, 5. Atomic weight, 10.82. It is used as a dopant in semiconductor processing.

bot **1.** Abbreviation for *beginning of tape.* **2.** Abbreviation of *bottom.*

bottoming Excessive movement of the cone of a loudspeaker or the diaphragm of a headphone so that the magnet or supporting structure is struck by the moving-coil piston assembly. It produces a clapping sound, particularly on bass (low-frequency) audio peaks.

bounce **1.** The springback or vibration of the armature of a relay on closure. **2.** An abnormal, abrupt change in the brightness of the image in a television receiver or cathode-ray-tube (CRT) computer monitor.

boundary **1.** In a polycrystalline substance, the area of contact between adjacent crystals. **2.** The area of meeting of two regions (such as n and p) in a semiconductor.

boundary defect A condition in which a piezoelectric crystal has two regions, intersecting in a plane, with different polarizations.

boundary effect In audio systems, a pheno-menon in which the proximity of an acoustic transducer to a flat surface enhances the pickup and/or transmission of sound. Occurs because of reflection of acoustic waves from the surface.

bound charge The portion of the electric charge on a conductor that does not escape to ground when the conductor is grounded. This occurs because of induction from neighboring charge carriers. Compare FREE CHARGE.

bound electron An electron held tightly in its orbit within an atom so that it is not ordinarily free to drift between atoms and contribute to electric current flow.

bow-tie antenna A center-fed antenna in which the two horizontal halves of the radiator are triangular plates that resemble a bow tie. A flat reflector consisting of closely spaced horizontal wires is mounted behind the triangles.

bow-tie test An oscilloscope-display checkout of a single-sideband (SSB) signal, in which the appearance of the display indicates the signal quality. The transmitter output signal is fed to the vertical deflection plates of the oscilloscope. The exciter audio output is fed to the horizontal sweep input of the scope.

boxcars Long pulses with short separating spaces between them.

box-shaped loop The characteristic square-loop hysteresis curve (B-H loop) that result when a sine wave of current is used to magnetize a sample of magnetic material. In this plot, which covers all four quadrants, the horizontal axis (H) displays magnetizing force, and the vertical axis (B) displays magnetization. Also see HYSTERESIS.

Boys radiomicrometer A detector for radiant energy. The device consists of a thermocouple and a galvanometer. When energy falls on the thermocouple, a voltage is produced, and this is measured by the galvanometer.

bp **1.** Abbreviation of BOILING POINT. **2.** Abbreviation of BANDPASS.

bpi Abbreviation of *bits per inch.*

B-plus Also called *b-positive.* **1.** Symbol, B+. The positive dc voltage required for certain electrodes of vacuum tubes, transistors, etc. **2.** The positive terminal of a B power supply.

B positive See B-PLUS.

B power supply A name used sometimes for the unit that supplies high-voltage dc energy to a vacuum tube plate or screen circuit.

bps Abbreviation of BITS PER SECOND.

Br Symbol for BROMINE.

bracketing A troubleshooting routine characterized by isolating progressively smaller areas in a circuit or chain of stages until the defective subcircuit or stage is located.

Bradley detector A locked-oscillator circuit that was once used as an FM detector.

braid **1.** A woven network of fine metal wires used for grounding purposes. It is usually made of fine copper conductors. The increased surface-area-to-volume ratio improves the conductivity, at radio frequencies, over a single conductor that has the same cross-sectional area. Braid can be tinned (saturated with solder) to retard corrosion. **2.** It is also called a *shield.* The outer conductor in prefabricated coaxial cable.

braided wire A length of braid. Used for grounding or shielding purposes.

brain waves Alternating or pulsating voltages that are caused by electrical activity in the brain of an animal or human being. The voltages can be picked up by electrodes attached to the scalp, and amplified to be viewed on a cathode-ray-tube (CRT) screen, heard by headphones or speakers, or traced by an electroencephalograph.

branch **1.** Any one of the separate paths of a circuit. With respect to the layout of its components, a branch can be series, parallel, series-parallel, parallel-series, or any combination of these. It is also called a LEG. **2.** See BRANCH CIRCUIT.

branch circuit In electrical wiring, a group of outlets served through a single cutout from a source of power-line ac voltage. The source can be a distribution center, subdistribution center, main, or submain. Interior lighting circuits are usually branch circuits because many lights are connected to one circuit controlled by a single fuse or circuit breaker.

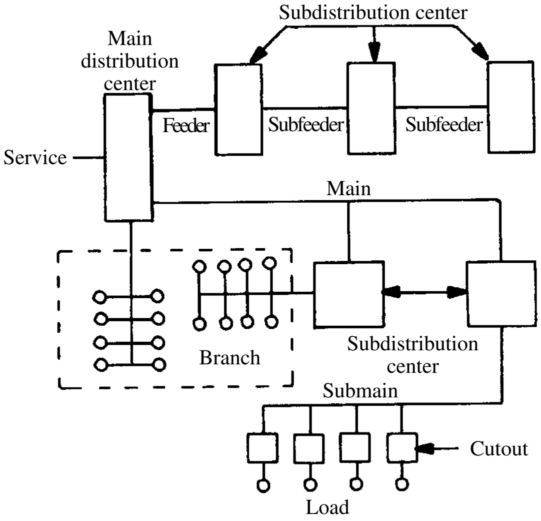

branch circuit
(enclosed in broken lines)

branch current Current flowing through a branch of a circuit, whose magnitude, with respect to the total current of the circuit depends on the nature of the branch.

branched In molecular polymers, the condition of side chains being attached to the main chain.

branched windings Forked windings of a polyphase transformer.

branching In robotics and artificial intelligence (AI), a set of routines or programs containing points at which a computer must select from among two or more alternatives. Such routines are used in critical processes, such as the manufacture of precision equipment.

branch point See JUNCTION POINT.

branch voltage The voltage, or voltage drop, across a branch of a circuit.

brass **1.** An alloy of copper and zinc that is widely used in electronics. Compared to annealed copper, this metal has four times the resistivity (or ¼ the conductivity), half the temperature coefficient, more than twice the tensile strength, and a lower melting point (900°C). **2.** A colloquialism for an old-fashioned, straight telegraph key.

brass pounder **1.** Colloquialism for telegraph operator or radiotelegraph operator. **2.** A radio amateur who handles large amounts of message traffic, particularly via Morse code. **3.** A radio amateur proficient in Morse code operation.

Braun electroscope An electroscope consisting essentially of a fixed metal vane to which a movable needle is fastened at a pivot. The repulsion between the two, when an electric charge is applied, causes the needle to move over a calibrated scale.

bravo Phonetic representation of the letter B.

brazing The joining of two metal (usually iron or steel) parts together with a suitable melted copper-alloy metal. Compare SOLDERING.

breadboard **1.** A perforated board, a chassis, or any basic framework on which electronic components can be mounted and quickly wired for the preliminary test of a circuit. It is so called because the first such foundation units of this sort actually were wooden breadboards. **2.** Any preproduction electronic prototype circuit. **3.** To set up a circuit on a breadboard.

breadboard model **1.** The preliminary model of an electronic device, often built on a breadboard (see BREADBOARD, **1**). **2.** Loosely, any prototype.

break **1.** An open circuit. **2.** To open a circuit. **3.** In communications, a word indicating a desire to transmit on a wavelength already occupied by radio traffic. **4.** See BREAK-IN, **1**.

break-before-make contacts Contacts, especially in a rotary selector switch, that open one circuit before closing the next one.

breakdown **1.** Failure of a circuit or device, caused mainly by excessive voltage, current, or power. A sudden high current, however, does not always indicate failure. **2.** AVALANCHE BREAKDOWN. **3.** The separation of an electronics problem or project into its constituent parts for an easier solution.

breakdown diode See ZENER DIODE.

breakdown region The region, in a pn junction, in which avalanche breakdown occurs.

breakdown strength See DIELECTRIC STRENGTH.

breakdown voltage **1.** The voltage at which current suddenly passes in destructive amounts through a dielectric. **2.** The voltage at which a gas suddenly ionizes, as in a gas tube. **3.** The voltage at which the reverse current of a semiconductor junction suddenly rises to a high value (nondestructive if the current is limited). See AVALANCHE BREAKDOWN.

break-in **1.** A technique of radio communication in which one station interrupts a transmission from another station, rather than waiting until the end of the latter's transmission. **2.** Also called *full break-in.* In a radio communications transceiver or transmitter/receiver combination, extremely rapid transmit/receive switching, approaching *full duplex* communications. Every pause in transmission, even of only a few milliseconds, creates a "receive window" allowing reception between spoken words or Morse code elements. **3.** BURN-IN.

breaking current The momentary current that flows when the contacts of a switch or relay are broken.

break-in keying A system of radiotelegraph keying in which the receiver is in operation whenever the key is open. See BREAK-IN, **2.**

break-in operation In radiotelegraph or single-sideband (SSB) communications, the practice of interrupting at any time to "talk back" to the other transmitting station. This operation is made possible by high-speed transmit/receive switching. See BREAK-IN, **2.**

break-in relay An electromechanical or solid-state relay that enables break-in operation. Largely supplanted by solid-state switching devices.

breakover point In a silicon-controlled rectifier, the source-voltage value at which the load current is suddenly triggered to its steep climb. Also called TRIGGERING POINT.

breakover voltage In a silicon-controlled rectifier with open gate circuit, the anode voltage at which anode current is initiated.

breakpoint A point in a computer program when, for the purpose of obtaining information for the program's analysis, the sequence of operations is interrupted by an operator or a monitor program.

breakpoint frequencies The upper- and lower-frequency points at which the gain-versus-frequency response of an amplifier or network departs from flatness.

breakpoint instruction An instruction that stops a computer.

breakthrough **1.** A new discovery, insight, or solution to a problem that results in an advancement in the state of the art. **2.** See PUNCHTHROUGH. **3.** See BREAKDOWN, **1. 4.** See AVALANCHE BREAKDOWN.

break time The time taken for a relay to drop out completely or a switch to open. Compare MAKE TIME.

breathing Slow, rhythmic pulsations of a quantity, such as current, voltage, brightness, beat note, etc.

breezeway In a sync pulse in NTSC color television, the part of the back porch between the trailing edge of the pulse and the color burst.

B-register An index register in a computer for storing words that are used to change an instruction before it is executed by the program.

Bremsstrahlung radiation The radiation emitted by a charged particle whose speed is altered when it passes through the electric field in the vicinity of an atomic nucleus.

brevity code A code not intended to conceal information, but to shorten the number of characters in a message or data file. The Q SIGNALS are an example of a brevity code used in communications. In computer data transfer and communications, brevity codes allow compression, speeding up the transfer rate and reducing the storage space for a given amount of data.

Brewster angle From BREWSTER'S LAW, the polarizing angle at which the reflected and refracted rays of incident light are perpendicular to each other.

Brewster's law (Sir David Brewster, 1781–1868). For any dielectric reflector, the relationship in which the refractive index is equal to the tangent of the polarizing angle.

bridge **1.** A network, usually consisting of four branches, connected so that an input signal can be applied between two opposite points and the output taken between the other two opposite points. When the component values are in a certain ratio, the voltage between the output points is zero, and the bridge is said to be balanced or set to null. **2.** A circuit such as that described in (**1**) used for electrical measurements. **3.** An audio or servo amplification system in which the load is driven from two outputs having opposite polarity, neither of which are at ground potential. **4.** A communications path between or among two or more networks. This allows the subscribers in any network to obtain data from, or send data to, any other network, in effect creating a network of networks.

bridge balance control A potentiometer, variable capacitor, or variable inductor that is used to adjust a bridge circuit to balance.

bridge-connected amplifier **1.** A dc amplifier stage in which the transistors and resistors are connected in a four-arm bridge circuit, with respect to dc. When the bridge is initially balanced, all dc is eliminated in the output load. The input signal unbalances the bridge, which results in an amplified output signal in the load. **2.** An amplifier pair having opposing outputs across which a load can be bridged to obtain twice the power output of either amplifier alone.

bridged differentiator See HALL NETWORK.

bridge detector The output-indicating device (e.g., meter, oscilloscope, or headphones) that indicates whether a bridge is balanced or unbalanced. Also called *null detector* or *null indicator.*

bridged integrator A null network that consists of a two-stage resistance-capacitance (RC) integrator circuit bridged by a capacitor. This network produces a shallow null at a single frequency determined by the *R* and *C* values in the integrator. Compare HALL NETWORK.

bridged-tee attenuator An attenuator consisting of a tee section, between the input and output of which is bridged a single-series arm.

bridged-tee circuit Any circuit (of resistors, capacitors, inductors, or a combination of these) that consists of a tee section, bridged by a single-series section, from input to output.

bridged-tee null network A bridged-tee circuit of resistance (*R*) and capacitance (*C*), proportioned so that at some setting of the *R* and *C* values, the output of the circuit is zero.

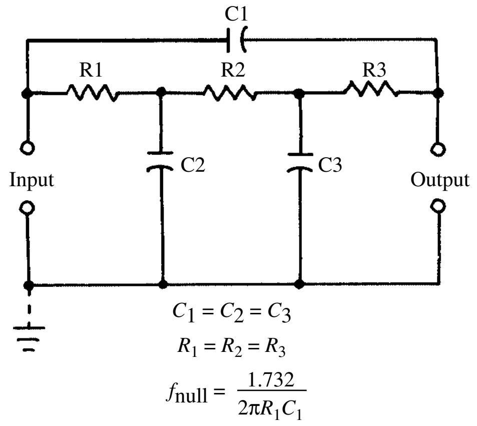

bridged integrator

bridged-tee oscillator A low-distortion oscillator circuit whose frequency is determined by a bridged-tee null network inserted into the negative-feedback path of the circuit.

bridge feedback A combination of current feedback and voltage feedback around an amplifier circuit. It is so called because, in the feedback circuit, the resistors and the output resistance of the amplifier form a four-arm bridge.

bridge generator The power source (e.g., a battery or oscillator) that supplies the signal to a BRIDGE used for electrical measurements.

bridge indicator See BRIDGE DETECTOR.

bridge oscillator See BRIDGE GENERATOR.

bridge rectifier A full-wave rectifier circuit in which four rectifying diodes are connected in a bridge configuration. Each half-cycle of ac input is rectified by a pair of diodes in opposite quarters of the bridge and in series with each other. The bridge does not require a transformer with a center-tapped secondary, as does the FULL-WAVE, CENTER-TAP RECTIFIER circuit.

bridge source See BRIDGE GENERATOR.

bridge-type meter A frequency-sensitive bridge (such as the Wien bridge) that can be used to measure audio frequency. Because the bridge can be balanced at only one frequency at a time,

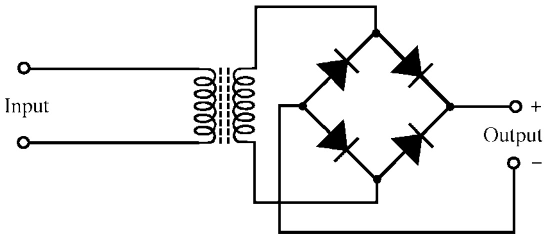

bridge rectifier

its adjustable arm can be calibrated to read the frequency directly.

bridge-type impedance meter An impedance-measuring circuit in which unknown impedance *Z* is connected in series with a calibrated variable resistor R. An ac voltage is applied to the series circuit. The separate voltage drops across the resistor and impedance are measured successively as the value of *R* is varied. When the two voltage drops are identical, *Z* equals *R*, and the impedance can be read from a calibrated dial on the variable-resistor control.

bridge-type oscillator A resistance-capacitance (RC) tuned oscillator in which a Wien bridge is used as the frequency-determining circuit in the feedback loop.

bridge-type power meter **1.** See BOLOMETER BRIDGE. **2.** A four-arm bridge specially designed to operate at radio frequencies. At null, the impedance of the unknown is read directly from the balancing dial or calculated from bridge constants. This instrument is used to measure the impedance of circuit components, antennas, and transmission lines.

bridge-type SWR meter A four-arm bridge that is specially designed to operate at radio frequencies. At null, the standing-wave ratio (SWR) is calculated from the bridge resistance values or read from a direct-reading scale on the null-indicating meter.

bridging amplifier An amplifier whose input impedance is so high that it can be considered infinite for practical purposes. Thus, the amplifier can be connected across a load or line without significantly affecting the operation of the system.

bridging coupler A voltage-dependent resistor that permits an occasionally used device (such as a bell) to be connected permanently across a regularly used device (such as a telephone) without continuously short-circuiting the latter. Thus, the bridging coupler ordinarily has very high resistance; but when the line voltage is momentarily raised, the resistance lowers and the occasionally used device is actuated (e.g., the bell rings).

bridging gain The gain of a bridging amplifier expressed as the ratio (in decibels) of the power developed in the amplifier load to the power in the load to which the input terminals of the amplifier are connected.

bridging loss The loss that results from the shunting of a speaker, microphone, earphone, or other transducer by a resistor, capacitor, or inductor. Generally, the loss is expressed as a power ratio in decibels.

Briggsian logarithm (Henry Briggs, 1556-1631). A base-10 logarithm, generally known as a COMMON LOGARITHM. Compare NAPIERIAN LOGARITHM.

brightness SI unit, candela per square meter (cd/m^2); cgs unit, lambert (L). The quantity of light, per unit area, emitted or reflected perpendicular to a light-emitting surface.

brightness control **1.** In a computer monitor, television receiver, or oscilloscope, a potentiometer that varies the negative bias voltage on the control grid of the cathode-ray tube (CRT). The brightness of the image is inversely proportional to this negative bias voltage. **2.** The control of the brightness of an illuminated area.

brilliance See BRIGHTNESS.

brilliance control **1.** The BRIGHTNESS CONTROL in a television receiver or computer monitor. **2.** The brightness control in a cathode-ray oscilloscope. **3.** A control for adjusting the level of the tweeter output in a speaker system.

British Standard wire gauge Abbreviation, NBS SWG. A classification of wire sizes sometimes used in England, Australia, and New Zealand. The higher the number, the thinner the wire. The designator does not take into account any coatings on the wire, such as enamel, rubber, or plastic insulation. In the United States, the *American wire gauge* is more often used. See AMERICAN WIRE GAUGE.

British Standard Wire Gauge (NBS SWG) Diameters

NBS SWG	Millimeters	Inches
1	7.62	0.300
2	7.01	0.276
3	6.40	0.252
4	5.89	0.232
5	5.38	0.212
6	4.88	0.192
7	4.47	0.176
8	4.06	0.160
9	3.66	0.144
10	3.25	0.128
11	2.95	0.116
12	2.64	0.104
13	2.34	0.092
14	2.03	0.080
15	1.83	0.072
16	1.63	0.064
17	1.42	0.056
18	1.22	0.048
19	1.02	0.040
20	0.91	0.036
21	0.81	0.032
22	0.71	0.028
23	0.61	0.024
24	0.56	0.022
25	0.51	0.020
26	0.46	0.018
27	0.42	0.0164
28	0.38	0.0148
29	0.345	0.0136
30	0.315	0.0124
31	0.295	0.0116
32	0.274	0.0108

NBS SWG	Millimeters	Inches
33	0.254	0.0100
34	0.234	0.0092
35	0.213	0.0084
36	0.193	0.0076
37	0.173	0.0068
38	0.152	0.0060
39	0.132	0.0052
40	0.122	0.0048

British thermal unit Abbreviation, Btu. The amount of heat required to raise the temperature of a pound of water by one degree Fahrenheit, in an ambient environment of slightly greater than 39°F.

broadband Also called *wideband.* Possessing a characteristic wide bandwidth or range of operating frequencies. This term can be applied at audio frequencies (AF) or radio frequencies (RF), and is frequently used to describe the performance of oscillators, amplifiers, antennas, and various types of networks. The term can also be applied to describe the nature of electromagnetic emissions or noise. Examples are given in the following several definitions. Compare NARROWBAND.

broadband amplifier An amplifier that has very wide frequency response, such as 10 Hz to 10 MHz. Examples are an instrument amplifier and a video amplifier.

broadband antenna An antenna that operates satisfactorily over a comparatively wide band of frequencies without requiring retuning at individual frequencies. Examples are the log-periodic and discone antennas.

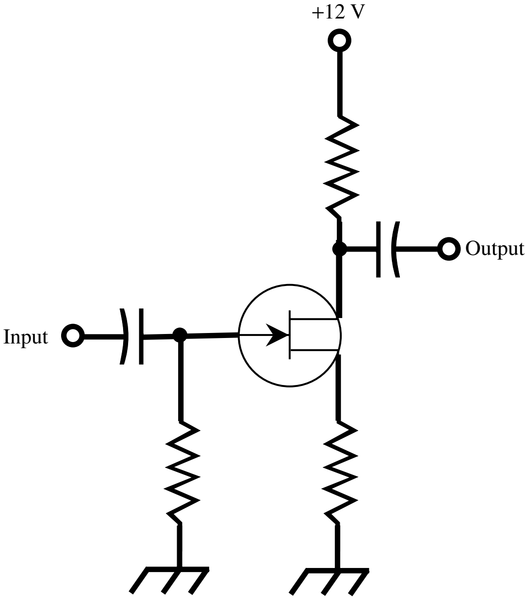

broadband amplifier

broadband electrical noise Electrical noise that is present over a wide frequency spectrum (e.g., 3 kHz to 30 MHz).

broadband I-F An intermediate-frequency (IF) amplifier or amplifier chain. The wide frequency response is important when an increased bandpass is preferred to high selectivity, as in high-fidelity radio tuners.

broadband interference Interference, other than noise, that is present over a wide band of frequencies. An example is over-the-horizon shortwave radar, recognizable by its characteristic "woodpecker" sound in communications receivers at high frequencies.

broadband Klystron A Klystron oscillator with a broadbanded tuned circuit.

broadband tuning Receiver tuning characterized by a selectivity curve having a pronounced flat top or broad nose that passes a wide band of frequencies. Also called *broadband response.*

broadcast **1.** A radio-frequency transmission of an intelligence-bearing signal that is directed to numerous unspecified receiving stations. **2.** The transmission or dissemination of signals to a large, unspecified number of receiving stations.

broadcast band Any band of frequencies allocated for broadcasting (see BROADCAST SERVICE, **1**), but particularly the U.S. standard amplitude-modulation (AM) and frequency-modulation (FM) radio broadcast bands at 535 to 1605 kHz (AM) and 88 to 108 MHz (FM).

broadcasting The dissemination of signals for reception by the general public, not for communications purposes.

broadcast interference Abbreviation, BCI. Interference to normal reception by broadcast receivers, usually arising from signals emitted by other stations.

broadcast receiver A receiver intended primarily to pick up standard broadcast stations. Also see BROADCAST BAND.

broadcast service **1.** Any radio transmitting service (including television) that exists for the purpose of sending out electromagnetic signals for general reception, rather than addressing them to specific receiving stations. **2.** The service provided by a station operating in the broadcast band.

broadcast station Any station in the broadcast service, but especially one assigned to operate in the standard U.S. broadcast bands. Also called *broadcasting station.*

broadcast transmitter A radio transmitter designed specifically for, and operated in, the broadcast service.

broad response Slow deflection of an indicator, such as a meter, over a relatively wide range of values of the input quantity.

broadside In a perpendicular direction; for example, broadside radiation from an antenna.

broadside antenna See BROADSIDE ARRAY.

broadside array Also called *broadside antenna.* A phased group of antennas arranged so maximum radiation occurs in directions perpendicular to the plane containing the driven elements. This requires that all of the antennas be fed in phase. The elements can be half-wave dipoles or full-wave, center-fed conductors. Full-wave elements have a slight gain over half-wave elements. At high frequencies, this type of array is usually constructed from two driven antennas. At very-high and ultra-high frequencies there can be several driven antennas. The antennas can each consist of a single element, or they can be Yagis, loops, or other systems with individual directive properties. In general, the larger the number of elements in the entire array, the greater the gain and directivity.

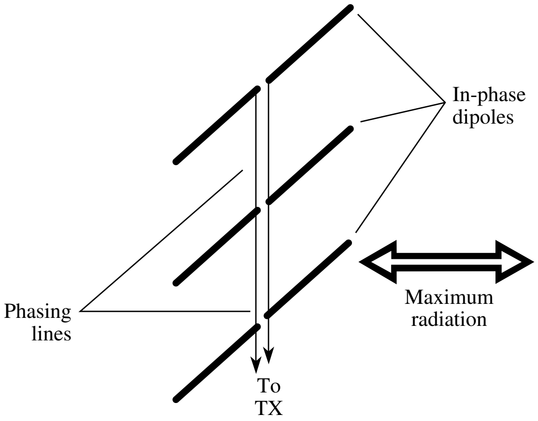

broadside array

broad tuning Tuning that is characterized by pronounced signal width, often resulting in adjacent-channel interference. A common cause of such impaired selectivity is low Q in the tuned circuit(s).

Broca galvanometer A device consisting of an astatic magnetic arrangement, with a coil enclosing central consequent poles. The device is characterized by fast response and high sensitivity.

bromine Symbol, Br. A nonmetallic element of the halogen family. Atomic number, 35. Atomic weight, 79.90.

bronze An alloy of copper and tin that has various uses in electronics. Also see PHOSPHOR BRONZE.

Brown and Sharpe gauge See AMERICAN WIRE GAUGE.

Brownian movement (Robert Brown, 1773–1858). Random movement of microscopic particles—especially in solutions. It occurs because of collisions of molecules with the particles. Einstein showed, in his early work, a connection between this movement and the *Boltzmann constant.*

brownout A deliberate lowering of line voltage by a power company to reduce load demands. Minor events of this type often pass unnoticed by the average consumer. More pronounced events produce observable effects, such as shrinkage of television and cathode-ray-tube (CRT) computer-display images.

Bruce antenna A vertical collinear array that consists of several resonant sections connected by short, rigid, parallel-conductor stubs. The currents in the radiating sections are in phase. Maximum radiation and response occur broadside to the antenna (omnidirectional in the horizontal plane). Polarization is vertical. The antenna produces gain at low radiation and response angles, and is commonly used in repeater installations and fixed communications stations at very-high frequencies (VHF) and ultra-high frequencies (UHF).

brush A usually metal or carbon strip, blade, or block, that slides in contact with another part, as in a motor commutator.

brush discharge Also called *Saint Elmo's fire.* A cloud of repelled ions around the tip of a pointed conductor charged to a high voltage. It often produces a visible glow in the air.

brush holder The housing for a brush in a motor, generator, rheostat, slip-ring junction in a rotating data-transmission system, etc.

brute force **1.** The transmission of a signal of excessive or unnecessary power. **2.** An inefficient approach to a problem, which might solve the problem, but requires far more energy, effort, or computer memory/storage space than the minimum needed to accomplish the same result.

brute-force filter A pi-type lowpass dc power supply filter, so called because of the extremely large inductances and capacitances that are generally used.

brute supply An unregulated power supply.

B-scope A cathode-ray tube (CRT), used in radar, that presents a B DISPLAY.

B service A teletype communication system operated by the Federal Aviation Administration (FAA).

B-supply The dc power supply that provides anode operating voltages, such as plate and screen voltages in a vacuum-tube radio-frequency (RF) power amplifier.

BT-cut crystal A piezoelectric plate cut from a quartz crystal at an angle of rotation (relative to the x-axis) of -49°. It has a zero temperature coefficient of frequency at approximately 25°C. Also see CRYSTAL AXES and CRYSTAL CUTS.

Btu Abbreviation of BRITISH THERMAL UNIT.

BuAer Abbreviation of *Bureau of Aeronautics.*

bubble memory In digital-computer practice, a special type of static magnetic memory. The magnetic material is divided into regions that are magnetized in different directions. So called because the flux lines of the tiny magnetized regions are shaped somewhat like, and move around after the fashion of, bubbles on the surface of a glass of soda.

bubble shift register A shift register that uses a magnetic bubble (see BUBBLE MEMORY) that can be moved sequentially from electrode to electrode on a wafer.

bubbling See MOTORBOATING.

bucket A computer memory or a designated location in such a memory.

bucking The process of counteracting one quantity, such as a current or voltage, via series or parallel application of a similar quantity that has opposite polarity (180 degrees out of phase).

bucking circuit **1.** A circuit used to obtain bucking action. The simplest form is a battery and potentiometer that supply a variable voltage of polarity opposite to that of the voltage to be bucked. A more sophisticated form is an ac transformer, the secondary of which is connected in series and out of phase with the ac utility line. **2.** The zero-set circuit in an electronic voltmeter.

bucking coil A coil placed and positioned so that its magnetic field partially or completely cancels the field of another coil. Troublesome hum fields sometimes are neutralized with such a coil.

bucking voltage See BACK VOLTAGE, **2.**

buckling The warping of storage-battery plates, usually resulting from excessive charge or discharge.

buckshot In an amplitude-modulated (AM) or single-sideband (SSB) radio transmission, broadband signal splatter caused by excessive modulation, or detuned multiplier circuits.

buffer **1.** An amplifier used principally to match two dissimilar impedance points and isolate one stage from a succeeding one in a cascaded system, and thus to prevent undesirable interaction between the two. **2.** In a digital computer, a storage site used temporarily during data transfers to compensate for differences in data flow rates. **3.** In digital-computer operations, a follower stage that is used to drive a number of gates without overloading the preceding stage.

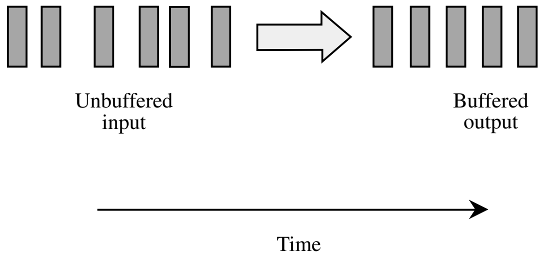

buffer, 2.

buffer amplifier See BUFFER, **1.**

buffer capacitor A high-voltage fixed capacitor that is placed across a transformer secondary to suppress voltage spikes and sharp waveforms—especially when the input is a square wave.

buffer circuit **1.** In a data system that uses a keyboard, an electronic circuit that allows the operator to type ahead of the data output. **2.** See BUFFER, **1**, **2** and **3.**

buffered output An output (power, signal, etc.) that is delivered from the generating device through an isolating stage, such as a buffer amplifier. This arrangement protects the device from variations in the external load. Compare UNBUFFERED OUTPUT.

buffer storage **1.** A buffer that is used to interface between data systems with different rates of transmission. **2.** See BUFFER, **2.**

bug **1.** Slang for WIRETAP, **1. 2.** Slang for circuit fault, **1**. **3.** A semiautomatic key that some radiotelegraph operators use to send Morse code.

bug key See BUG, **3.**

building-block technique The process of assembling electronic equipment by quickly connecting together already completed stages (in the form of boxes or blocks) and supplying power and signals to the setup. Also called *modular technique* and *modular construction.*

building-out circuit A short section of transmission line shunting another line; it is used for impedance matching. Also called *building-out section.*

buildup **1.** The process whereby the voltage of a rotating generator starts at a point that is determined by the residual magnetism of the machine, and gradually increases to a voltage representing the point at which the resistance line crosses the magnetization curve. **2.** The (usually gradual) accumulation of a quantity (e.g., the buildup of charge in a capacitor).

bulb A globe-like container having any of a number of characteristic shapes from spherical to tubular and usually evacuated, for enclosing the elements of an electron device, such as a vacuum tube, gas tube, photocell, or lamp.

bulge **1.** A nonlinear attenuation-versus-frequency curve in a transmission line. **2.** A localized nonlinearity in a function.

bulk The body or mass of a semiconductor specimen, as opposed to junctions within the specimen. Current flows through a junction, but it can also flow, more or less, through the mass of semiconductor wafer into which the junction has been formed.

bulk effect An effect, such as current, resistance, or resistivity, observed in the overall body of a sample of material, as opposed to a region within the material or on its surface. Thus, a silicon diode can display junction resistance (i.e., resistance offered by a junction processed in a wafer of silicon), as well as bulk resistance (i.e., the effective resistance of all paths around the junction, through the mass of the wafer). Compare SURFACE EFFECT.

bulk-erased tape Recording tape whose signal content has been removed via a bulk eraser.

bulk-erase noise **1.** The residual magnetic impulses that remain on a magnetic tape after it has been bulk-erased. **2.** Noise generated by bulk-erased tape when the latter passes through deenergized record or erase heads in a tape machine.

bulk eraser A type of power-line-frequency degausser that erases an entire reel of magnetic tape without requiring that the tape be unreeled and passed continuously under an erase head. This saves considerable time, but often leaves some BULK-ERASE NOISE on the tape. Also called BULK DEGAUSSER.

bulletin board In personal computing or amateur packet communications, a system that allows subscribers to leave messages for each other for access via a *modem* or *terminal node controller.*

bulletin station A station intended for the transmission of bulletins of interest to certain parties, such as military personnel or amateur radio operators. An example is W1AW in Newington, Connecticut, an amateur radio station that transmits bulletins and code practice.

buncher In a Klystron, a cavity resonator that contains two grids mounted parallel to the electron stream. The electrostatic field of the grids alternately accelerates and retards the electrons, velocity-modulating the stream into bunches.

buncher grids In a Klystron, the closely spaced grids that velocity-modulate the electron beam into successive bunches.

buncher resonator In a velocity-modulated tube, such as a Klystron, the input cavity resonator.

buncher voltage The radio-frequency (RF) grid-to-grid voltage in the buncher resonator of a Klystron.

bunching The production of electron bunches in a velocity-modulated tube, such as a Klystron. Also see BUNCHER.

bunch stranding A technique for combining several thin wires into a single thick wire. Often used in guy wires and electrical conductors to improve tensile strength and flexibility. At radio frequencies, bunch stranding also improves electrical conductivity by increasing the ratio of surface area to cross-sectional area. This minimizes losses caused by *skin effect.*

Bunet's formula A formula for calculating the inductance of a multilayer air-core coil that has a diameter less than three times the length:

$$L = a^2N^2/(9a + 10l + 8.4c + 3.2cl/a)$$

where N is the number of turns, a is the average coil radius, c is the winding thickness, and l is the length of the coil.

Bunsen cell A cell consisting of a zinc rod in a sulfuric acid solution contained in a porous pot, which is in a nitric acid solution. The zinc rod serves as the negative pole; the positive pole is a piece of hard carbon. The cell produces about 1.9 volts and delivers relatively high current.

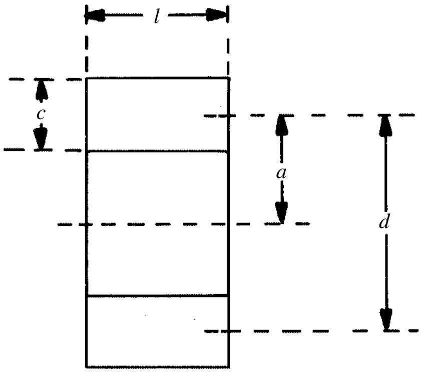

Bunet's formula

burden See VOLTAGE BURDEN.

burn **1.** A blemish on the screen of a cathode-ray tube (CRT), caused by destruction of the phosphor there. This results from prolonged focusing of an intense electron beam in one spot. **2.** A blemish on the screen of a television picture tube, usually resulting from ions that reach the screen when the ion trap is not working correctly.

burn-in A long, thorough, carefully controlled preliminary test of a component, device, or system, to stabilize its electrical characteristics after manufacture and to ensure that it will function according to rated specifications. An important test for equipment whose reliability must be guaranteed, such as an emergency communications transceiver.

burnout **1.** Failure of a conductor or component caused by overheating from excess current or voltage. **2.** The open-circuiting of a fuse. **3.** Electrical failure of any type.

burst **1.** The abrupt ionization of the gas in an ionization chamber by cosmic rays. **2.** An abrupt increase in the amplitude of a signal. Also, the type of signal that results from burst action. **3.** See COLOR BURST.

burst amplifier In a color-television receiver, the amplifier that separates the burst pulse from the video signals and amplifies the former. See COLOR BURST.

burst gate timing In a color-television receiver, the timing of the gating pulse with the input signal of the burst amplifier.

burst generator A signal generator delivering a burst output (see BURST, **2**) for testing various types of equipment. Its output is intermediate between sine waves and square waves, and is convenient for rapidly appraising the performance of such devices as amplifiers, filters, electronic switches, transducers, and loudspeakers.

burst transmission A short transmission at high speed. This method of transmission saves time, but increases the necessary bandwidth of a signal by the same factor as the ratio of the high speed to the original speed.

bus **1.** A main conductor in a circuit. A bus can be high in the sense that its potential is above or below ground, or it can be low or at ground reference. **2.** In computer operations, a common group of paths over which input and output signals are routed.

bus driver A buffering device designed to increase the driving capability of a microprocessor, which itself might be capable of driving no more than a single load.

business machine Any piece of electronic or electromechanical equipment used mainly, or entirely, for business purposes. Examples are photocopiers, facsimile (fax) machines, printers, and computers.

busing The parallel interconnection of circuits.

busy test A check conducted to find out whether or not a certain telephone subscriber line is in use.

busy tone Also called *busy signal.* An intermittent tone that indicates that the subscriber line being called is in use.

Butler oscillator An oscillator that consists of a two-stage amplifier with a quartz crystal in the positive-feedback path from output to input.

butterfly capacitor A plate-type variable capacitor that has two stator sections and a single rotor section common to the two stators. External connections are made to the stators only. Thus, no wiping contact is required to the rotor, and the troubles associated with such a contact are avoided. The butterfly capacitor is actually two variable capacitors in series. The unit is so called from the shape of its rotor.

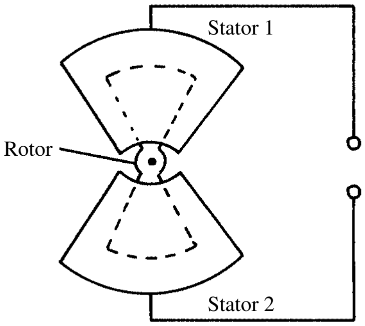

butterfly capacitor

butterfly circuit A combination of a butterfly capacitor and a ring, of which the stator plates of the capacitor are an integral part. The resulting structure is a compact variable-frequency tuned circuit. The ring supplies the inductance, and the butterfly supplies the capacitance. It is also called *butterfly tank* and *butterfly resonator.*

Butterworth filter A high-pass, low-pass, bandpass or band-rejection filter, characterized by a flat passband (absence of passband ripple) and high attenuation at frequencies far removed from the passband.

Butterworth function A mathematical function that is used in the design of a BUTTERWORTH FILTER.

button **1.** Usually, a small switch that is actuated by finger pressure. It is also called pushbutton and pushbutton switch. Sometimes, the term is applied only to the insulated knob or pin which is pushed to operate the switch. **2.** A tiny lump of impurity material, placed on the surface of a semiconductor wafer for alloying with the wafer to form a junction. See ALLOY JUNCTION. **3.** The carbon element(s) in a BUTTON MICROPHONE.

button capacitor A button-shaped ceramic or silvered-mica fixed capacitor. Because of its disk shape and mode of terminal connection, it offers very low internal inductance.

button microphone A microphone in which a button-shaped carbon element is attached to a diaphragm, which is set into vibration by sound waves. This motion causes the button resistance to vary, modulating a direct current that passes

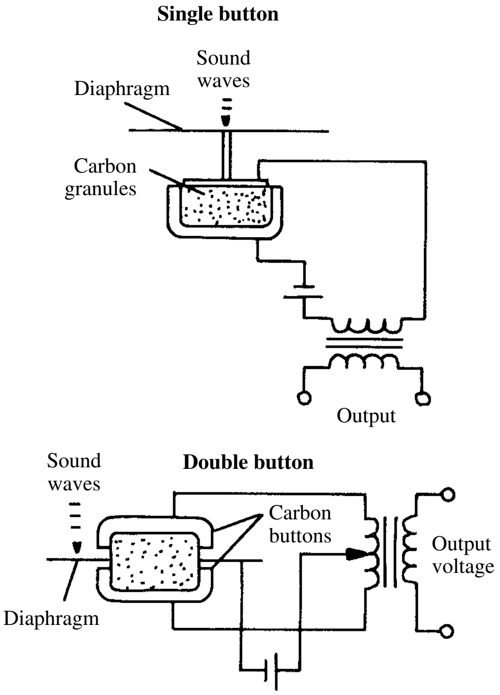

button microphone

through the button. A single-button microphone has only one button, whereas a double-button microphone has two—one mounted on each side of the center of the diaphragm.

buzz **1.** A low-pitched rough sound with high-frequency components, usually the result of electrical interference from nonsinusoidal voltages generated by neighboring equipment or devices. **2.** The waveform associated with such a sound. **3.** Fastening two conducting surfaces by a KELLIE BOND.

buzzer A nonringing device used principally to generate sound other than that achievable with sine waves. In an electromechanical vibrating-reed buzzer, the reed acts as an armature, which is mounted close to the core of an electromagnet. At quiescence, the reed rests against a stationary contact. When voltage is applied to the electromagnet, the reed is attracted to the core, moving away from the contact; but this breaks the circuit, the magnetism ceases, and the reed springs back to the contact. The action is repeated continuously at a frequency that depends on the reed dimensions and its distance from the core.

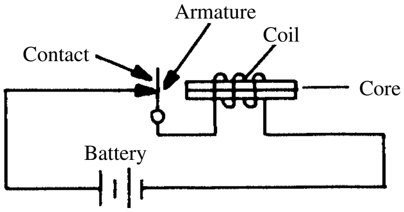

buzzer

BV Abbreviation of BREAKDOWN VOLTAGE.

B voltage The dc voltage required by certain electrodes of vacuum tubes and transistors. It especially pertains to voltages required by the plate and screen of a vacuum tube, as opposed to the filament voltage and control-grid voltage.

bw **1.** Abbreviation of *bandwidth.* 2. Abbreviation of *black-and-white.*

BWA Abbreviation of *backward-wave amplifier.*

BWG Abbreviation of BIRMINGHAM WIRE GAUGE.

BWO Abbreviation of BACKWARD-WAVE OSCILLATOR.

BX Symbol and abbreviation for armored and insulated flexible electrical cable.

bypass A route (either intended or accidental) through which current easily flows around a component or circuit instead of through it.

bypass capacitor A capacitor that is used to conduct an alternating current around a component or group of components. Often the ac is removed from an ac/dc signal, the dc being free to pass through the bypassed component.

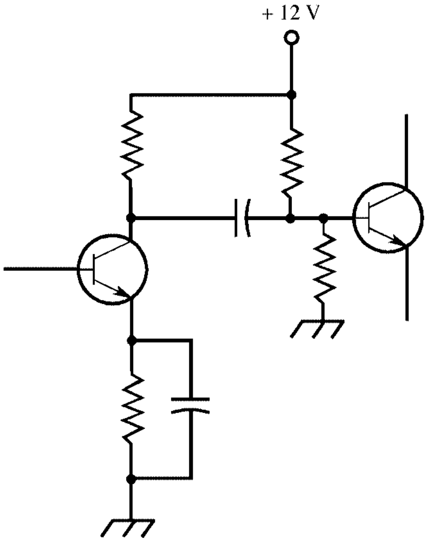

bypass capacitor

B-Y signal In a color television receiver, the color-difference signal which, when combined with a luminance (Y) signal, forms a blue primary signal for the three-gun picture tube.

byte In digital-computer and data-communications operations, a unit of data consisting of eight contiguous bits. In packet communications, the term *octet* is often used.

C

C **1.** Abbreviation of CAPACITANCE. **2.** Symbol for COLLECTOR of a transistor. **3.** Symbol for CARBON. **4.** Abbreviation of CELSIUS. **5.** Symbol for COULOMB. **6.** Abbreviation of CALORIE.

c **1.** Abbreviation of CENTI. **2.** Abbreviation of CENTS. **3.** Symbol for CAPACITANCE. **4.** Symbol for SPEED OF LIGHT in a vacuum.

Ca Symbol for CALCIUM.

cabinet An enclosure for a piece of apparatus. It might or might not incorporate electromagnetic shielding.

cable **1.** A usually flexible (but sometimes rigid) medium, via which electrical power or signals are transferred. Although the term is occasionally applied to a single conductor, especially when it is a braid or weave of a number of wires, cable usually means a bundle of separate, insulated wires or strands of fiberoptic material. **2.** CABLEGRAM.

cable address A code word that specifies the recipient of a CABLEGRAM.

cable assembly A special-purpose cable with connectors.

cable attenuation Reduction of signal intensity along a cable, usually expressed in decibels per foot, hundred feet, mile, etc.

cable capacitance Capacitance between conductors in a cable or between conductors and the outer sheath of a cable. **2.** Sometimes, capacitance between a cable and earth.

cable clamp A support device for cable runs in equipment and systems.

cable communications Telegraphy or telegraphy via a (usually undersea) cable.

cable connector A connector, such as a coaxial fitting, that joins cable circuits or connects a cable to a device.

cabled wiring Insulated leads connecting circuit points; they are tied together with lacing cord or with spaced fasteners.

cablegram A (usually printed) message transmitted or received via undersea cable. Compare RADIOGRAM and TELEGRAM.

cable loss See CABLE ATTENUATION.

cable run The path taken by a cable.

cable splice **1.** An electrical attachment between two sections of cable that has identical or similar construction, with or without the use of connectors. **2.** To electrically attach two sections of cable that have identical or similar construction, with or without the use of connectors.

cable tie A short piece of wire or plastic that holds wires or cables in a bundle.

cable TV See COMMUNITY-ANTENNA TELEVISION.

cache memory A short-term, high-speed, high-capacity computer memory. Similar to a scratchpad or read-write memory.

CAD Acronym for *computer-aided design.*

CAD/CAM Acronym for *computer-aided design and manufacturing.*

cadmium Symbol, Cd. A metallic element. Atomic number, 48. Atomic weight, 112.41. Many electronic structures are cadmium plated for protection.

cadmium borate phosphor Formula, ($CdO + B_2O_3$): Mn. A substance used as a phosphor coating on the screen of cathode-ray tubes. The characteristic fluorescence is green-orange.

cadmium cell Also called *Weston standard cell.* An electrochemical standard cell used as a reference voltage source. Produces 1.0186 volt at 20°C.

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

cadmium plating The process of coating a conductor or component with cadmium to increase its resistance to corrosion.

cadmium selenide photocell A photoconductive cell in which cadmium selenide is the light-sensitive material.

cadmium silicate phosphor Formula, ($CdO + SiO_2$). A substance used as a phosphor coating on the screen of cathode-ray tubes; the characteristic fluorescence is orange-yellow.

cadmium standard cell See STANDARD CELL.

cadmium sulfide photocell A photoconductive cell in which cadmium sulfide is the light-sensitive material.

cadmium tungstate phosphor Formula, $CdO + WO_3$. A substance used as a phosphor coating on the screen of cathode-ray tubes; the characteristic fluorescence is light blue.

cage A completely shielded enclosure, such as a screen room, which is covered with a grounded fine-mesh conductive screen on all sides.

cage antenna An antenna, usually center-fed and balanced, that consists of multiple parallel conductors arranged in a cylindrical cage configuration. The cage results in a much broader bandwidth than is the case with an antenna made up of a single conductor. Cage antennas are typically used at frequencies between about 10 and 200 MHz.

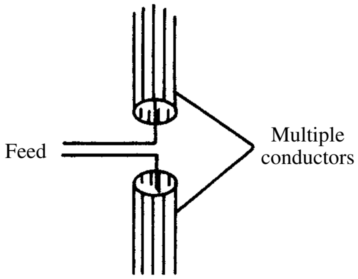

cage antenna

CAI Abbreviation for *computer-assisted instruction.*

CAL An acronym for *conversional algebraic language*, a general-purpose problem-oriented computer programming language used in time-sharing systems.

calcium Symbol, Ca. A metallic element of the alkaline-earth group. Atomic number, 20. Atomic weight, 40.08.

calcium phosphate phosphor Formula, $Ca_3(PO_4)_2$. A substance used as a phosphor coating on the screen of long-persistence cathode-ray tubes; the characteristic fluorescence is white, as is the phosphorescence.

calcium silicate phosphor Formula, ($CaO + SiO_2$): Mn. A substance used as a phosphor coating on the screen of cathode-ray tubes; the characteristic fluorescence ranges from green to orange.

calcium tungstate phosphor Formula, $CaWO_4$. A substance used as a phosphor coating on the screen of short-persistence cathode-ray tubes; the characteristic fluorescence is blue, as is the phosphorescence.

calculate To perform the steps of an intricate mathematical operation. Compare COMPUTE.

calculating punch A data-processing peripheral that reads punched cards, makes calculations, and punches new data into those cards or new cards.

calculator A machine that performs mathematical operations, especially arithmetic. Typically, the device is a small box with buttons and a miniature numeric display. Used only in mathematical applications. In contrast, a COMPUTER can be used for a much wider variety of jobs, such as word processing, graphics, and data-base. Many personal computers have calculator programs; the “buttons” are actuated by pointing and clicking with a mouse.

calculus **1.** The symbology and rules comprising a system of logic, such as BOOLEAN ALGEBRA. **2.** A branch of mathematical analysis concerned with rates of change and accumulation. See DIFFERENTIAL CALCULUS and INTEGRAL CALCULUS.

calendar age The age of a piece of equipment, measured since the date of manufacture. Specified in years, months, and days. The actual manufacture date might alternatively be given.

calendar time The time available in a working period [i.e., a 40-hour work week represents a calendar time of 120 hours (five days times 24 hours per day)].

calibrate To compare and bring into agreement with a standard.

calibrated measurement **1.** A measurement made with an instrument that has been calibrated with a standard reference source. **2.** A measurement that is corrected for instrument error.

calibrated meter An analog or digital meter that has been adjusted to agree as closely as possible with a reference source.

calibrated scale **1.** A scale whose graduations have been carefully checked for accuracy (i.e., they correspond to the true values of the quantity that they represent). The scale is graduated to read directly in units of the quantity, such as milliamperes, kilohertz, volts, etc. **2.** A scale with fixed, plain numeric graduations (e.g., 0 to 100) that do not directly indicate the magnitude of a quantity, but that can be converted to various quantities via graphs, nomographs, tables, or charts. See CALIBRATION CURVE.

calibrated sweep In an oscilloscope, a sweep circuit calibrated to indicate sweep frequency or time at all control settings.

calibrated triggered sweep In an oscilloscope, a triggered sweep circuit calibrated in terms of sweep time or frequency.

calibration **1.** Determining the accuracy with which an instrument indicates a quantity. **2.** Determining the degree to which the response of a circuit or device corresponds to desired performance. **3.** Marking a scale to show actual values of a quantity in the form of a direct readout. For example, the scale of an analog meter might be calibrated in milliamperes (mA) from 0 to 50 in increments of 1 mA.

calibration accuracy **1.** A quantitative expression of the agreement between the value of a quantity, as indicated by an instrument, and the true value. Usually expressed as the maximum percentage of the true value by which the indicated value can be expected to deviate in either direction (e.g., ±0.5 percent). **2.** The precision of a direct-reading meter in terms of its full-scale deflection (e.g., ±2.0 percent of full scale).

calibration curve A graph showing the relation between the actual values of a quantity and the setting or indication of an instrument or component. Usually plotted in rectangular coordinates.

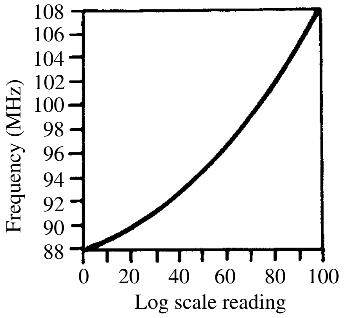

calibration curve

calibration marker A pip or blip, superimposed on a pattern displayed on a cathode-ray-tube (CRT) screen, to identify a point closely as to frequency, voltage, distance, or some similar term.

calibrator A device used to perform a calibration (e.g., a signal generator).

calibrator crystal A highly accurate and stable quartz crystal, used in an oscillator as a frequency standard. An example is the 100-kHz crystal oscillator and harmonic generator used in some communications receivers.

californium Symbol, Cf. A radioactive element produced artificially. Atomic number, 98. Atomic weight, 251.

call **1.** In communications, a transmission by a station for the purpose of either alerting a particular receiving station for which there is a message, or alerting all receiving stations to prepare them for a general broadcast message. **2.** In a computer program, a branch to a closed subroutine; also, to branch to such a subroutine.

call direction code Abbreviation, CDC. In telegraph networks, a special code that, when transmitted to a terminal, causes the teleprinter to be automatically switched on.

calling sequence **1.** Computer program instructions needed to establish the conditions for a call (see CALL, **2**). **2.** Subroutine instructions providing a link to the main program.

call instruction A computer program instruction that makes a program controller branch to a subroutine; it also locates and identifies the parameters needed for the subroutine's execution. Also known as *subroutine call.*

call letters Letters and/or numbers assigned to, and used to identify, licensed radio stations.

calorie Abbreviation, cal or C. The amount of heat energy, at a pressure of 1 atmosphere, that will raise the temperature of 1 gram of water by 1 degree Celsius.

calorimeter An instrument for measuring heat energy. By adaptation, a calorimeter can be used to measure radio-frequency (RF) power—especially at microwave frequencies (see CALORIMETRIC POWER METER).

calorimeter system See CALORIMETRIC POWER METER.

calorimetric power meter A specialized form of wattmeter, in which the power to be measured is dissipated in an oil or water bath that has a known and fixed mass. The wattage is determined indirectly, by measuring the extent to which the temperature of the liquid increases in a certain amount of time.

CAM **1.** Abbreviation of *computer-aided manufacturing.* **2.** Abbreviation of *content-addressed memory.*

cambric Finely woven cotton or linen used for insulation. One type of spaghetti (conductor insulation), for example, is varnished cambric tubing.

camera cable A multiwire cable that conducts the video signal from a television camera to control equipment.

camera chain In television, the camera and the equipment immediately associated with it, excluding the transmitter and its peripherals.

camera signal The output signal delivered by a television camera.

camera tube Any video pickup tube, such as an iconoscope or orthicon, that converts light reflected by a scene into a corresponding television signal.

Campbell bridge A circuit that is used for comparing mutual inductance with capacitance.

camp-on In a telephone system, a method of engaging a line that is busy until it becomes available for use.

CAN Abbreviation of CANCEL CHARACTER.

can **1.** A metal enclosure or container roughly resembling a tin can (though not necessarily cylindrical), used for shielding or potting components. **2.** Colloquial expression for HEADPHONE.

Canada balsam A transparent cement derived from the turpentine distilled from balsam fir resin. It is useful in optical technology and in certain areas of electro-optics.

Canadian Standards Association The Canadian equivalent of the *National Bureau of Standards* in the United States. An agency that publishes agreed-on standards for industries.

cancel character **1.** IGNORE CHARACTER. **2.** A control character indicating that the associated data is erroneous.

cancellation The elimination of one quantity by another, as when a voltage is reduced to zero by another voltage of equal magnitude and opposite sign.

candela Symbol, cd. The SI unit of luminous intensity; 1 cd represents 1⁄60 of the radiating power of one square centimeter of a perfect radiator at the temperature of freezing platinum.

candle Abbreviation, c. Also called *international candle.* A unit of light intensity that is the value of emission by the flame of a sperm-whale-oil candle burning at the rate of 7.776 grams per hour.

candle power Abbreviation, cp. Luminous intensity in international candles: the luminous intensity resulting from the burning of a sperm-whale-oil candle at 7.776 grams per hour.

candoluminescence White light produced without extreme heat.

cannibalization The deliberate use of parts from operational equipment to temporarily repair or maintain other equipment. It is a last-resort, emergency measure.

cap **1.** Abbreviation of CAPACITANCE. **2.** Abbreviation of CAPACITOR.

capacimeter See CAPACITANCE METER.

capacitance Symbol, C. Unit, farad. The property exhibited by two conductors separated by a dielectric, whereby an electric charge becomes stored between the conductors. Capacitance is thought of as analogous to mechanical elasticity. Also see FARAD.

capacitance bridge A four-arm ac bridge for gauging capacitance against a standard capacitor. In its simplest form, it has a standard capacitor in one arm and resistors in the other three.

capacitance coupling The transfer of ac energy between two circuits or devices by a capacitor or capacitance effect. Also see COUPLING.

capacitance diode See VARACTOR.

capacitance divider An alternating-current voltage divider that uses capacitors, rather than resistors. It is used in certain oscillators, such as the Colpitts type.

capacitance filter A filter consisting of only a high-capacitance capacitor. Because the capacitor cannot discharge instantaneously, it tends to maintain its voltage and smooth out the ripples in the voltage applied to it.

capacitance-inductance bridge A combination ac bridge that can be used for either capacitance or inductance measurement. Both capacitance and inductance can be measured in terms of a standard capacitance; however, some of these bridges use standard inductors in the inductance-measuring mode.

capacitance meter A direct-reading meter for measuring capacitance. In most available types, a stable ac voltage is applied to the meter circuit, to which an unknown capacitor is connected in series; meter deflection is roughly proportional to the reactance of the capacitor. Also called MICROFARAD METER.

capacitance ratio In a variable capacitor, the ratio of maximum to minimum capacitance.

capacitance relay A relay circuit that operates from a small change in its own capacitance. It consists of an RF oscillator whose tank capacitance is very low. When a finger is brought near the circuit's short pickup antenna, the attendant increase in capacitance detunes the oscillator, activating the relay. Also called PROXIMITY RELAY and PROXIMITY SWITCH.

capacitance-resistance bridge A combination ac bridge that can be used for either capacitance or resistance measurement. The unknown resistance is measured against a standard resistor; the unknown capacitance against a standard capacitor.

capacitance sensor See CAPACITANCE TRANSDUCER.

capacitive amplifier See DIELECTRIC AMPLIFIER.

capacitive attenuator An ac attenuator whose elements are capacitors in any desired combination of fixed and/or variable units. The desired attenuation is afforded by the capacitance ratio.

capacitive coupling A means of coupling between circuits that uses a series capacitor for direct-current blocking. The signal passes through the capacitor, but the blocking effect allows different bias voltages to be applied to the two stages.

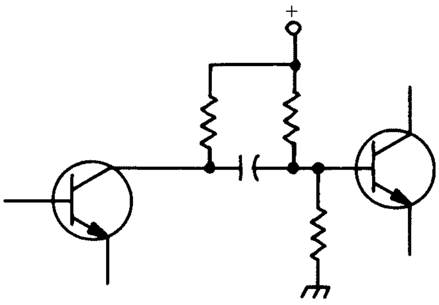

capacitive coupling

capacitive diaphragm A metal plate deliberately placed in a waveguide to introduce capacitive reactance and, thereby, cancel an inductive reactance.

capacitive-discharge ignition An electronic ignition system for automotive engines. Provides nearly constant high voltage, regardless of engine speed. A dc-to-dc step-up converter charges a large capacitor (typically to 300 volts) when the distributor breaker points are closed; when they are open, the capacitor discharges through the ignition coil, thereby generating an ignition pulse of several thousand volts.

capacitive division Reduction of an ac voltage by a capacitive voltage divider.

capacitive feedback Feeding energy back from the output to the input of an amplifier or oscillator through a capacitor.

capacitive-input filter A smoothing filter for ac power supplies, in which the element closest to the rectifier is a capacitor, regardless of the components or circuits placed subsequently.

capacitive load A load consisting of a capacitor or a predominantly capacitive circuit.

capacitive loading In an antenna, the addition of capacitance in series with the element(s). This raises the resonant frequency for a radiator having a given physical length. It can also serve to increase the physical length required for a radiator having a specified resonant frequency. Compare INDUCTIVE LOADING.

capacitive post A protrusion inside a waveguide for the purpose of introducing capacitive reactance to cancel an inductive reactance.

capacitive potentiometer See CAPACITIVE VOLTAGE DIVIDER.

capacitive pressure sensor A pressure sensor that uses a radio-frequency oscillator and a pair of metal plates separated by dielectric foam. The circuit is designed so a change in the capacitance between the plates causes the oscillator frequency to change. This change is sensed. A signal is sent to an analog-to-digital converter (ADC) and then to a microcomputer that calculates the extent of the pressure.

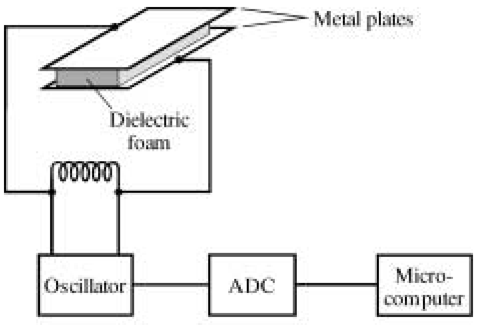

capacitive pressure sensor

capacitive proximity sensor A transducer used in mobile robots that detects the presence of certain kinds of objects. It consists of an oscillator whose frequency is determined by an inductance-capacitance (LC) circuit to which a metal plate is connected. When a conducting or partially conducting object comes near the plate, the mutual capacitance changes the oscillator frequency. This change is detected and sent to the robot controller.

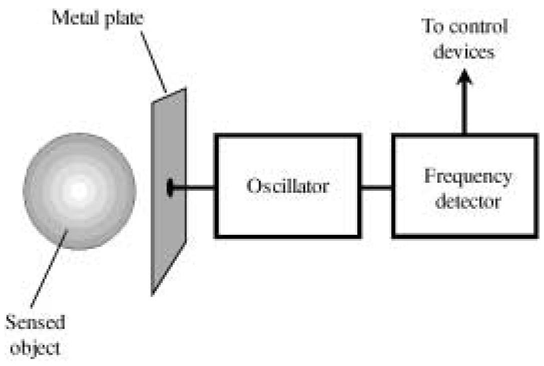

capacitive proximity sensor

capacitive reactance Symbol, X_C. Unit, ohm. The reactance exhibited by an ideal capacitor, considered as a negative pure-imaginary quantity; $X_C = -j/(6.28fC)$, where f is the frequency in hertz, C is the capacitance in farads, and j is the unit imaginary number (the square root of –1). Alternatively, f can be specified in megahertz and C in microfarads. In a pure capacitive reactance, current leads voltage by 90 degrees. Also see CAPACITANCE, CAPACITOR, and REACTANCE.

capacitive speaker See ELECTROSTATIC SPEAKER.

capacitive transducer A transducer consisting essentially of a refined variable capacitor whose value is varied by a quantity under test, such as pressure, temperature, liquid level, etc.

capacitive tuning Variable-capacitor tuning of a circuit.

capacitive voltage divider A capacitive attenuator usually consisting of two series-connected capacitors whose values are such that an applied ac voltage is divided across them in the desired ratio.

capacitive welding An electronic welding system in which energy stored in a capacitor is discharged through the joint to be welded. This develops the heat necessary for the operation.

capacitive window A pair of capacitive diaphragms used in a waveguide to introduce capacitive reactance.

capacitor A passive electronic-circuit component consisting of, in basic form, two metal electrodes or plates separated by a dielectric (insulator).

capacitor amplifier See DIELECTRIC AMPLIFIER.

capacitor antenna See CONDENSER ANTENNA.

capacitor bank A network of capacitors connected in combination, yielding a desired characteristic.

capacitor braking The connection of a capacitor to the winding of a motor after the removal of power, to speed up the process of braking.

capacitor color code See COLOR CODE.

capacitor decade See DECADE CAPACITOR.

capacitor-discharge ignition CAPACITIVE-DISCHARGE IGNITION.

capacitor filter In a direct-current power supply, a filter consisting simply of a capacitor connected in parallel with the rectifier output.

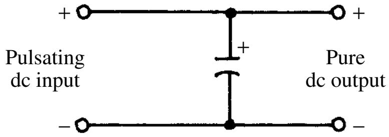

capacitor filter

capacitor-input filter A filter whose input component is a capacitor. The capacitor-input power-supply filter is distinguished by its relatively high dc output voltage, but somewhat poorer voltage regulation, compared with the CHOKE-INPUT FILTER.

capacitor leakage Direct current flowing through the dielectric of a capacitor. In a good nonelectrolytic capacitor, this current is normally less than 1 microampere. In an electrolytic capacitor, it can be up to several milliamperes, depending on the capacitance and the applied voltage.

capacitor loudspeaker See ELECTROSTATIC SPEAKER.

capacitor microphone See CONDENSER MICROPHONE.

capacitor motor An ac motor that uses a capacitor in series with an auxiliary field winding for starting purposes. Initially out-of-phase current in the auxiliary field (starting winding) causes a rotating field that turns the rotor. When the rotor reaches a safe speed, a centrifugal switch disconnects the capacitor and auxiliary field, and the motor continues running as an induction motor.

capacitor series resistance The ohmic loss in a capacitor. It results partly from conductor losses, and partly from losses in the dielectric material.

capacitor substitution box An enclosed assortment of selected-value capacitors arranged to be switched one at a time to a pair of terminals. In troubleshooting and circuit development, any of several useful fixed capacitance values can be thus obtained.

capacitor voltage **1.** The voltage at the terminals of a capacitor. **2.** The maximum voltage rating of a capacitor.

capacitor voltmeter See ELECTROSTATIC VOLTMETER.

capacity **1.** A measure of a cell's or battery's ability to supply current during a given period. **2.** CAPACITANCE. **3.** The number of bits or bytes a computer storage device can hold. **4.** The limits of numbers that a register can process.

capacity lag In an automatic control system, a delay caused by the storing of energy by the components. For example, in a heating system, capacity lag results from the time taken to heat the air or fluid after the thermostat turns on the heat.

capillary electrometer A sensitive voltage indicator, consisting of a column of mercury in a transparent capillary tube, in which is suspended a small drop of acid. When a voltage is applied to both ends of the mercury column, the acid drop moves toward the low-potential end of the column over a distance proportional to the voltage.

capstan The driven spindle or shaft of a magnetic tape recorder or transport.

capture area The effective ability of a radio antenna to pick up electromagnetic signals. The larger the capture area, the greater the antenna gain.

capture effect **1.** In frequency-modulation (FM) radio receivers, the effect of domination by the stronger of two signals, or by the strongest of several signals, on the same frequency. **2.** In an automatic-frequency-control system, the tendency of the receiver to move toward the strongest of several signals near a given frequency. **3.** In general, the tendency of one effect to totally predominate over other effects of lesser amplitude.

capture ratio A measure of frequency-modulation (FM) tuner selectivity: The amplitude difference, in decibels, between unwanted signals and the one being tuned in.

carbon Symbol, C. A nonmetallic element. Atomic number, 6. Atomic weight, 12.011. Carbon, besides being an invaluable material in electronics, is an important constituent of organic compounds.

carbon arc The arc between two electrified pencils of carbon or, as in an arc converter, between a carbon pencil and a metal electrode.

carbon brush A contact made of carbon or some mixture of carbon and another material, used in motors, generators, variable auto-transformers, rheostats, and potentiometers.

carbon-button amplifier An audio-frequency amplifier having as the active component an earphone whose diaphragm is attached to a carbon microphone button. The input signal applied to the earphone makes its diaphragm vibrate. The vibrating button modulates a local direct current. Amplification results from the large ratio of modulated local current to input-signal current.

carbon-composition resistor A non-inductive resistor made from a mixture of finely powdered carbon with a non-conductive substance, usually phenolic. The resulting clay-like material is pressed into a cylindrical shape, and wire leads are inserted in the ends. The resistance depends on the ratio of carbon to the non-conducting material, and on the physical distance between the wire leads. This type of resistor is useful from direct current to ultra-high radio frequencies. Compare FILM RESISTOR, WIREWOUND RESISTOR.

carbon/disk rheostat A rheostat consisting of a stack of carbon disks or washers, arranged so that a controllable pressure can be exerted on the stack. As a knob is turned, a screw increases or decreases the pressure, varying the total resistance of the stack.

carbon-film resistor A stable resistor whose resistance element is a film of carbon, vacuum-deposited on a substrate, such as a ceramic.

carbonization The application of a coat of carbon onto an electrode, either by electroplating or by any other means.

carbon microphone A microphone that includes one or two carbon buttons. See BUTTON MICROPHONE.

carbon-paper recorder A recorder in which a signal-actuated stylus writes, by impression only, through a sheet of carbon paper onto a plain sheet underneath. This eliminates the need for an ink-carrying stylus.

carbon-pile regulator A voltage regulator in which a stack of carbon disks or washers is in series with the shunt field. The pile resistance and field current depend on pressure applied to the pile by a wafer spring acting through a movable iron armature. Voltage drops increase the pressure and voltage rises decrease the pressure, thus regulating the generator with which it is associated.

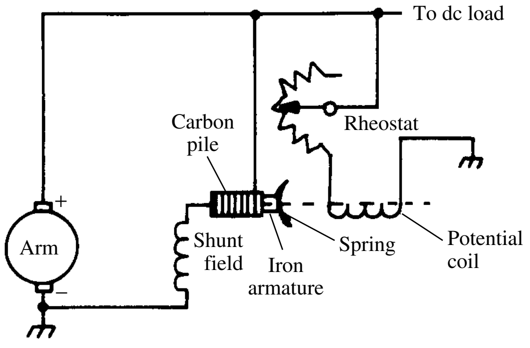

carbon-pile regulator

carbon-pile rheostat See CARBON-DISK RHEOSTAT.

carbon recording 1. A record made with a CARBON-PAPER RECORDER. **2.** The use of a carbon-paper recorder in data acquisition, facsimile, communications, and similar applications.

carbon resistor A resistor made from carbon, graphite, or some composition that contains carbon.

carbon/silicon-carbide thermocouple A thermocouple that is a junction between carbon and silicon carbide.

carbon transfer recording A method of facsimile reception in which the image is reproduced by carbon particles sprayed on the paper, a process controlled by the received signal.

carbon-zinc cell See ZINC-CARBON CELL.

Carborundum Formula, SiC. Trade name for a synthetic silicon carbide used as a semiconductor, refractory, or abrasive. Also see SILICON CARBIDE.

Carborundum crystal Trade name for a characteristically superhard crystal of silicon carbide.

Carborundum varistor A voltage-dependent resistor made from Carborundum.

carcinotron A special kind of oscillator tube used at ultra-high and microwave frequencies.

card 1. A usually thin, rectangular board containing a PRINTED CIRCUIT, often equipped with an edge connector that makes it easy to install, remove, or replace. Common in electronic and computer equipment having modular construction. **2.** The usually flat, thin insulating strip on which a resistor element is wound.

cardiac monitor An electronic device that displays or records electrical impulses from the heart for medical observation or diagnosis.

cardiac pacemaker An electrical cardiac stimulator that causes the heart to beat at certain intervals. Used when the patient has heart disease that prevents the heart from regulating itself.

cardiac stimulator An electronic device (sometimes implanted in the subject) that supplies electric pulses to stimulate heart action. Also called DEFIBRILLATOR and PACEMAKER.

card image In memory storage, the data contained on a single card.

cardiogram ELECTROCARDIOGRAM.

cardiograph ELECTROCARDIOGRAPH.

cardioid diagram A polar response curve in the shape of a cardioid pattern.

cardioid microphone A microphone with a (roughly) heart-shaped sound-field pickup pattern.

cardioid pattern A radiation/response pattern with one sharp null in the direction opposite the single main lobe. The lobe is extremely broad. In two dimensions, the curve is shaped somewhat like a "Valentine" heart.

cardiotachometer A device that indicates the pulse rate.

cardistimulator See CARDIAC STIMULATOR.

Carey-Foster bridge A special version of the slide-wire bridge that is useful for measuring an unknown resistance, whose value is close to that of a standard resistance.

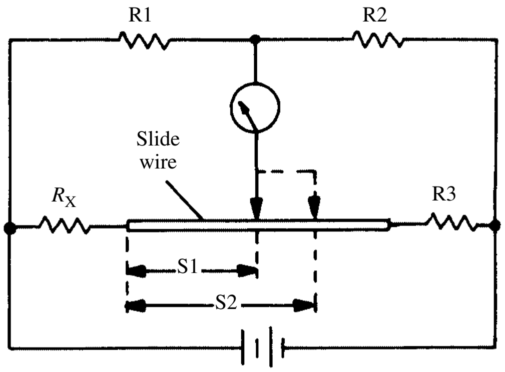

Carey-Foster bridge

Carey-Foster mutual inductance bridge An ac bridge that permits the measurement of mutual inductance in terms of a standard capacitor.

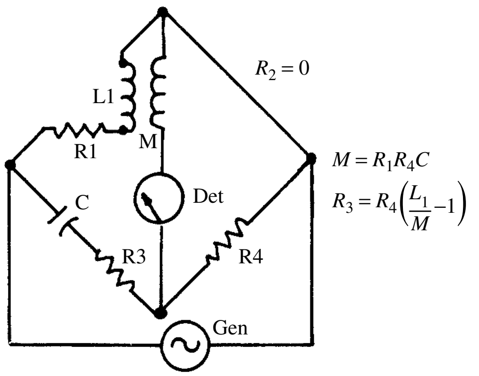

Carey-Foster mutual inductance bridge

carnauba wax A wax obtained from the Brazilian wax palm. Used as an electrical insulator, and as the dielectric in some electrets.

Carnot theorem In thermodynamics, the proposition that in a reversible cycle, all available energy is converted into mechanical work. Also called *Carnot's principle.*

carrier **1.** See CARRIER WAVE. **2.** See CHARGE CARRIER.

carrier amplifier See DIELECTRIC AMPLIFIER.

carrier beating **1.** The mixing of two radio-frequency carriers that are separated by a small amount of frequency, resulting in an audible tone in a receiver. **2.** A heterodyne in a facsimile or television signal, resulting in a pattern of cross hatches in the received image.

carrier choke A radio-frequency (RF) choke, inserted in a line to block a carrier component.

carrier chrominance signal For conveying color television information, sidebands of a modulated chrominance subcarrier.

carrier color signal For conveying color information in color television transmission, the sidebands of a modulated chrominance subcarrier (plus the unsuppressed chrominance subcarrier) added to the monochrome signal.

carrier concentration In a semiconductor material, the number of charge carriers per unit volume.

carrier control **1.** The modification, adjustment, or switching of a carrier wave. **2.** Adjustment of a circuit or device by means of a carrier wave.

carrier current The current component of a carrier wave, or the amplitude of that current. Compare CARRIER POWER and CARRIER VOLTAGE.

carrier-current communication See WIRED WIRELESS.

carrier-current control **1.** Control of the current component in a carrier wave. **2.** Remote control by means of wired wireless.

carrier-current receiver See WIRED-RADIO RECEIVER.

carrier-current relay A radio-frequency (RF) relay circuit, operated over a wire line by means of a transmitter.

carrier-current transmitter See WIRED-RADIO TRANSMITTER.

carrier deviation See CARRIER SWING.

carrier dispersion In a semiconductor, the spreading out of electrons and holes that leave the emitter simultaneously, but arrive at the collector at various times after following different paths.

carrier frequency The center frequency of a CARRIER WAVE.

carrier-frequency pulse A pulse that contains radio-frequency oscillation.

carrier-frequency range The band of carrier frequencies over which a transmitter or signal generator can operate.

carrier injection The apparent emission (injection) of electrons or holes into a semiconductor when a voltage is applied to the junction.

carrier leak **1.** A point at which carrier-wave energy escapes a circuit or enclosure. **2.** The residual carrier voltage present in the output of a carrier-suppressing circuit.

carrier level The amplitude of an unmodulated carrier wave.

carrier lifetime In a semiconductor, the interval before an injected current carrier (see CARRIER INJECTION) recombines with an opposite carrier and ceases to be mobile.

carrier line In carrier-current systems (see WIRED WIRELESS), the line or cable conducting the carrier-wave energy.

carrier mobility Symbol, μ. In a semiconductor material, the average drift velocity of electrons and holes per unit electrostatic field.

carrier noise Modulation of a carrier when there is no input from the modulator itself; unwanted modulation.

carrier noise level The noise signal amplitude that results from unintentional fluctuations of an unmodulated carrier.

carrier-on-light transmission A form of transmission in which many different signals are sent simultaneously by modulating a beam of light at multiple frequencies.

carrier-on-microwave transmission A form of transmission in which many different signals are sent simultaneously by modulating a microwave signal at multiple lower frequencies.

carrier-on-wire transmission A form of transmission in which many different signals are sent at the same time over a wire, by using radio-frequency carriers. Also called CARRIER-CURRENT COMMUNICATIONS or WIRED RADIO.

carrier oscillator In a single-sideband receiver, the radio-frequency (RF) oscillator that supplies the missing CARRIER WAVE.

carrier power The actual power represented by a radio-frequency (RF) carrier applied to an antenna, measured by either the direct or indirect method. The direct method involves determination of power according to the formula $P = I^2R$, where I is antenna current and R is antenna resistance at the point of current measurement. The indirect method involves determination of power according to the formula $P = EIF$, where E and I are antenna voltage and current, and F is a factor less than 1.0, whose value depends on the type of modulation used.

carrier power-output rating The power delivered by an unmodulated transmitter or generator to the normal load or its equivalent.

carrier shift In an amplitude-modulated transmitter or generator, the undesired change of average carrier voltage during modulation.

carrier-shift indicator An instrument for detecting carrier shift. It usually contains only a pickup coil, semiconductor diode, and dc milliammeter in series. Meter deflection is steady until carrier shift is detected; then, the needle fluctuates.

carrier signaling In wire telephony, the use of carrier-wave signals to operate such functions as dialing, ringing, busy signal, etc.

carrier storage In a semiconductor device, the tendency of mobile carriers to stay near a junction for a short time after the junction voltage has been removed or reversed in polarity.

carrier suppression The elimination of the carrier in an amplitude-modulated signal so that only the sideband energy remains.

carrier swing In frequency-modulated or phase-modulated transmission, the total deviation (lowest to highest instantaneous frequency) of the carrier wave.

carrier system The transmission of many signals over one circuit, accomplished by modulating various different carriers at different frequencies. Different signals can use different modulation methods.

carrier telegraphy **1.** Continuous-wave telegraphy by WIRED WIRELESS. **2.** Wired-wireless telegraphy in which a radio-frequency carrier is modulated by an audio-frequency keying wave.

carrier telephony Telephone communication by WIRED WIRELESS.

carrier terminal **1.** At each end of a carrier-current line or cable, the equipment for generating, modifying, or utilizing the carrier energy. **2.** In a balanced modulator, the point of carrier insertion.

carrier-to-noise ratio The ratio of carrier amplitude to noise-voltage amplitude.

carrier transmission Transport of information by a carrier, as by an amplitude-modulated radio wave that carries the low-frequency information as the AF modulation envelope and delivers it to the demodulator at the receiving station.

carrier-type dc amplifier A high-frequency ac amplifier, ahead of which is operated a generator and transducer. A dc voltage applied to the transducer modulates the carrier supplied by the generator; the amplifier boosts the modulated wave, and the resultant output is rectified at a level higher than that of the dc input signal.

carrier voltage The voltage component of a carrier wave; also, the amplitude of this component. Compare CARRIER CURRENT and CARRIER POWER.

carrier wave A sine wave that is modulated to convey information in wireless and cable communications systems. The lowest frequency normally used for wireless signal transmission is 9 kHz, corresponding to a wavelength of approximately 33 km. The highest frequency is less well defined; some systems make use of visible light waves, whose wavelengths are as short as approximately 4×10^{-7} m. For modulation to work effectively, the carrier must have a frequency at least 10 times the highest frequency of the modulating signal.

carry **1.** In adding a column of figures, the digit added to the column at the left when the sum exceeds one less than the radix value. **2.** In digital computers and counters, a pulse that corresponds to the arithmetic operation in which a figure is carried to the next column in addition.

carrying capacity The ability of a conductor, such as copper wire, to carry current safely (expressed in maximum amperes).

carry-complete signal In an arithmetic computation by a computer, an adder-produced signal indicating that the pertinent carries have been generated.

carry system A communications system in which several carries occupy one circuit.

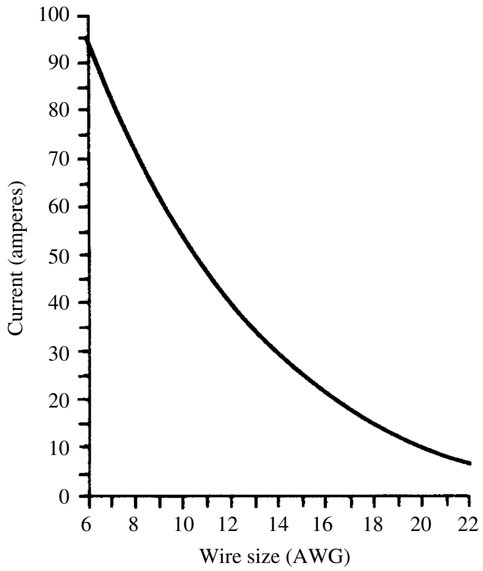

carrying capacity (of some AWG sizes of copper wire)

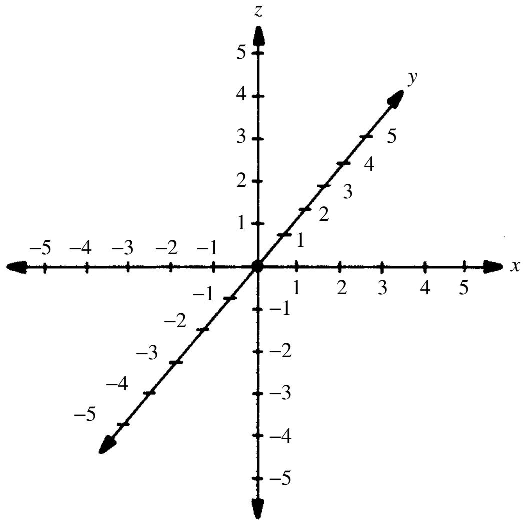

Cartesian three-space

carry time The time taken for a digital computer or counter to perform a carry operation (See CARRY, **2**).

Cartesian coordinate geometry Also called *rectangular coordinate geometry.* In robotic systems, a movement scheme in two or three dimensions. The position of the robot arm is determined by linear coordinates, relative to an origin point. These coordinates are specified along linear axes—each of which is perpendicular to the others at the origin. See CARTESIAN COORDINATES, CARTESIAN PLANE, and CARTESIAN THREE-SPACE.

Cartesian coordinates Also called *rectangular coordinates.* A mathematical system that uniquely defines the position of a point on a plane, in space, or in general, in an n-dimensional hyperspace when n is a whole number greater than 3. There are n axes for n dimensions, each axis intersects all the others at a single point, called the origin. The axes are mutually perpendicular at this origin. The axes are scaled in units with the origin having coordinate values that are all equal to zero (usually). Positive values go along the axes in one direction; negative numbers go in the opposite direction for each axis. Usually, the axes are graduated in equal-sized units. The system gets its name from the mathematician Rene Descartes.

Cartesian plane A linear, two-dimensional coordinate plane commonly used for graphing equations in one variable.

Cartesian three-space A linear, three-dimensional graph-coordinate system used for rendering equations in one or two variables.

Cartesian three-space graph A three-dimensional graph that shows an equation in one or two variables. Three-space graphs are often displayed more clearly by means of computer graphics, in which the entire display can be rotated to show the characteristics of the surface resulting from a given equation or function.

Cartesian n-space The coordinate space defined by a Cartesian system of n coordinates, where n is a whole number of 2 or greater.

cartridge **1.** The replaceable transducer assembly of a microphone. **2.** A magnetic-tape magazine. Also see TAPE CARTRIDGE. **3.** A removable computer mass-storage medium, containing a tape, magnetic diskette, or optical diskette. **4.** An insulating tube housing a fuse, semiconductor component, resistor, capacitor, or other part.

cartridge fuse A fuse consisting of a fusible wire enclosed in a cartridge, having a ferrule at each end for plug-in connection.

cascadable Capable of, or designed for, being connected in cascade with other similar or identical components.

cascade **1.** Components or stages connected and operated in sequence, as in a three-stage amplifier. The components or stages are often but not necessarily identical. **2.** To form a cascade.

cascade control **1.** In an automatic control system, a controller whose setting is varied by the output of another controller. **2.** An automatic control system in which the control units are connected in stages, so that one unit must operate before the next one can function.

cascaded amplifier A multistage amplifier in which the stages are forward-coupled in succession.

cascaded carry In digital computer practice, a system of performing the carry operation (see CARRY) in which the $n + 1$ place receives a carry pulse only when the nth place has received carry information to generate the pulse.

cascade thermoelectric device A thermoelectric component or circuit that consists of several cascaded sensors (see CASCADE, **1**).

cascade voltage doubler A voltage-doubler circuit (see VOLTAGE DOUBLER) consisting of two diode-capacitor combinations in cascade. Unlike the conventional voltage-doubler circuit with two capacitors in the output, the cascade voltage doubler has one in the input and one in the output.

cascode A high-gain, low-noise, high-input-impedance amplifier circuit, consisting of a grounded-emitter or grounded-source input stage coupled directly to a grounded-base or grounded-gate output stage.

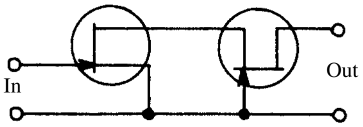

cascode (field-effect transistor arrangement)

case temperature The temperature at a designated point on the outside surface of a component's case or housing.

Cassegrain antenna A dish antenna that uses CASSEGRAIN FEED.

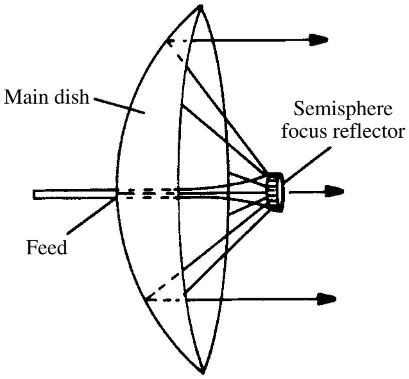

Cassegrain antenna

Cassegrain feed A dish-antenna feed system in which the feed point is located at the center of the dish itself. For transmission, the radio-frequency energy emerges from a waveguide and is directed to a small convex reflector at the focal point of the dish. The small reflector directs the signal back to the dish, spreading the energy out to cover the entire surface of the dish. The dish reflects the energy again and collimates it in the desired direction of propagation. For reception, the process is reversed; the dish focuses the energy on the small reflector, which propagates it back to the feed point.

cassette **1.** A holder (magazine) of reels of magnetic tape that is itself a mechanical subassembly, which can be easily inserted into and removed from a tape deck. **2.** A lightweight holder of photographic film or X-ray plates (before, during, and after exposure).

castor oil A viscous insulating oil extracted from castor beans. Highly refined castor oil is used as an impregnant in some oil-filled capacitors. Dielectric constant, 4.3 to 4.7. Dielectric strength, 380 V/mil.

CAT Abbreviation of COMPUTERIZED AXIAL TOMOGRAPHY.

catalysis The process whereby an agent, called a catalyst, enhances a chemical reaction without entering into the reaction. Catalysts are used in electronics, for example, to promote the setting of resins in potting and encapsulating operations.

catalytic agent A substance that accomplishes catalysis.

cataphoresis As caused by the influence of an electrostatic field, the migration toward the cathode of particles suspended in a liquid.

catastrophic failure **1.** Sudden, unexpected failure of a component or circuit. **2.** Failure that can result in the breakdown of an entire system. Also called *catastrophic breakdown.*

catcher In a Klystron, the second reentrant cavity. (See KLYSTRON.)

catcher diode A diode that is connected to regulate the voltage at the output of a power supply. The cathode is connected to a source of reference voltage. If the anode, connected to the source to be regulated, becomes more positive than the cathode, the diode conducts and prevents the regulated voltage from rising more than 0.3 volt above the reference voltage (for germanium diodes) or 0.6 volt above the reference voltage (for silicon diodes).

catcher grids In a Klystron, the grids through which the bunched electrons pass on their way from the buncher to the collector. Catcher grids absorb energy from the bunched electrons and present it to the collector circuit.

category In a computer system, a group of magnetic disk volumes containing information related by a common application.

category storage A computer-file storage section that contains a number of categories and used by an operating system.

catenation See CONCATENATION.

cathode **1.** The negative electrode of a device (i.e., the electrode from which electrons move when a current passes through the device). **2.** In an electrochemical cell, the electrode that gains electrons. This is generally the positive electrode. **3.** In a vacuum tube, the electron-emitting electrode (filament or indirectly heated cathode sleeve).

cathode current Symbol I_K. The current flowing in the cathode circuit of a tube. Cathode current is the total of grid, plate, screen, and suppressor currents, and can have an ac and a dc component.

cathode dark current The electron emission from the photocathode of a camera tube when there is no illumination.

cathode element In a vacuum tube, an indirectly heated emitter of electrons. Also see CATHODE, **2.**

cathode emission **1.** The giving up of electrons by the cathode element of a device, such as a vacuum tube. Electrons can be emitted by either hot or cold cathodes, depending on the tube. **2.** Collectively, electrons released by a cathode.

cathode heating time The time required for the temperature of a tube cathode to increase from cold to its maximum specified operating temperature after the cathode current has been initiated. Also called cathode warmup time.

cathode luminous sensitivity For a photomultiplier tube, the cathode's sensitivity to light. This sensitivity figure is the ratio of photocathode current to incident light flux.

cathode-ray oscillograph An instrument that provides a permanent record, by photographic or other means, of the image on the screen of a cathode-ray tube.

cathode-ray oscilloscope See OSCILLOSCOPE.

cathode rays Invisible rays emanating from the cathode element of an evacuated tube operated with a high voltage between the anode and cathode. Cathode rays (electrons) cause certain substances, PHOSPHORS, to glow upon striking them.

cathode-ray scanning tube Any tube in which an electron beam is deflected horizontally and vertically to scan an area. These include oscilloscope tubes, some computer monitors, radar displays, and television camera tubes.

cathode-ray tube **1.** An evacuated tube containing an anode and cathode that generates cathode rays when operated at high voltage. **2.** An oscilloscope tube. **3.** A picture tube.

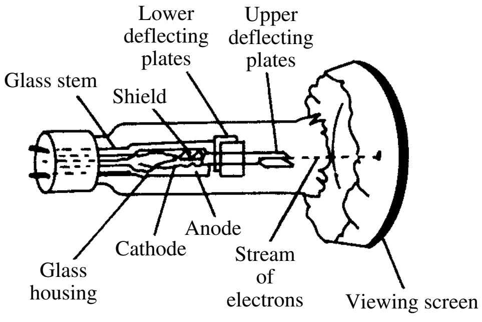

cathode-ray tube

cathode terminal **1.** In a diode (semiconductor or tube), the terminal to which a negative dc voltage must be applied for forward-biasing the diode. Compare ANODE TERMINAL. **2.** In a diode, the terminal at which a positive dc voltage appears when the diode acts as an ac rectifier. Compare ANODE TERMINAL. **3.** The terminal connected internally to the cathode element of device. **4.** In a vacuum tube, an indirectly heated electron emitter.

cathode voltage Symbol, E_K. The voltage between ground (or B-minus) and the cathode of a tube; it can have both ac and dc components.

cathodic protection A method of preventing corrosive galvanic action in underground metal pipes or the submerged hulls of ships. The part to be protected is used as the cathode of a circuit through which a direct current is passed in the direction opposite to that which caused the corrosion, thus counteracting it.

cathodofluorescence Fluorescence resulting from a material's exposure to cathode rays.

cathodoluminescence In a vacuum chamber in which a metal target is bombarded with high-velocity electrons (cathode rays), the emission of radiation of a wavelength characteristic of the metal.

cation A positive ion. Also see ION.

CAT scanner The X-ray apparatus for COMPUTERIZED AXIAL TOMOGRAPHY.

CATV Abbreviation of COMMUNITY-ANTENNA TELEVISION (usually cable television).

caustic soda electrolyte Symbol, NaOH. Sodium hydroxide solution, as used in some secondary cells and experimental devices.

cavitation The local formation of cavities in a fluid used in ultrasonic cleaning because of the reduction in pressure at those points.

cavitation noise In an ultrasonic cleaner, the noise resulting from the collapse of bubbles produced by cavitation.

cavity A metallic chamber (can) in which energy is allowed to reflect, sometimes resulting in resonance.

cavity filter A microwave (usually band rejection) filter consisting of a resonant cavity and associated coupling devices.

cavity frequency meter See CAVITY WAVEMETER.

cavity impedance The impedance across a cavity at a particular frequency. At resonance, the cavity impedance is purely resistive.

cavity laser A laser that employs a resonant cavity filled with gas, such as helium/neon or argon, and a pair of reflectors. Resonance occurs between the reflectors, one of which is totally reflective and the other of which is approximately 95 percent reflective. Output is from the partially reflective end of the device.

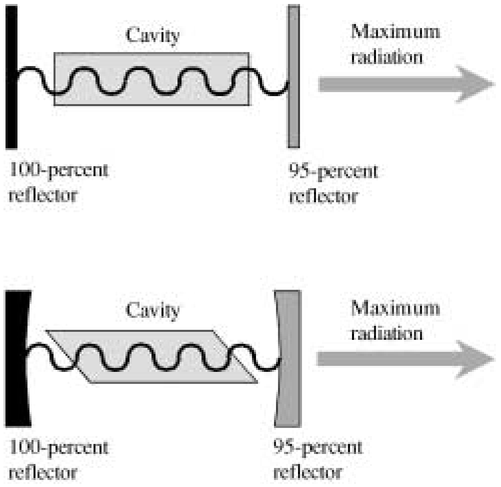

cavity laser

cavity magnetron A magnetron whose anode is a series of resonant cavities.

cavity oscillator An oscillator with a cavity-tuned circuit.

cavity radiation Energy radiated from a tiny hole in an otherwise sealed chamber. The radiation occurs at all electromagnetic wavelengths; the greater the temperature within the chamber, the greater the frequency at which the radiation has its maximum amplitude.

cavity resonance The phenomenon whereby a hollow cavity resonates; specifically, resonance in small metal cavities at microwave frequencies.

cavity resonator See RESONANT CAVITY.

cavity wavemeter An absorption wavemeter whose adjustable element is a tunable resonant cavity into which radio-frequency (RF) energy is injected through a waveguide or coaxial cable. Such an instrument is useful at microwave frequencies.

CB Abbreviation of CITIZENS BAND.

Cb Symbol for COLUMBIUM.

C_B Symbol for BASE CAPACITANCE of a transistor.

C band The band of radio frequencies between 3.9 and 6.2 GHz.

C_C Symbol for collector capacitance of a transistor.

cc **1.** Alternative abbreviation of *cubic centimeter*. The International Organization for Standardization recommends cm^3. **2.** Abbreviation of COTTON-COVERED.

CCA Abbreviation of CURRENT-CONTROLLED AMPLIFIER.

CCD Abbreviation of CHARGE-COUPLED DEVICE.

CCIS Abbreviation of COMMON-CHANNEL INTERFACE SIGNALING.

CCIR Abbreviation of *Comite Consultatif International des Radiocommunications (International Radio Consultative Committee).*

CCIT Abbreviation of *Comite Consultatif International Telegrafique (International Telegraph Consultative Committee).*

CCITT Abbreviation of *Comite Consultatif International Telegrafique et Telephonique (International Telegraph and Telephone Consultative Committee).*

CCS **1.** Abbreviation of CONTINUOUS COMMERCIAL SERVICE. **2.** Abbreviation of *common-channel signaling.*

CCTV Abbreviation of CLOSED-CIRCUIT TELEVISION.

CCTV monitor A video monitor that receives a signal from a CCTV transmitter.

CCTV signal The picture signal in a CCTV system. It can be either a modulated radio-frequency signal or a composite video signal.

ccw Abbreviation of COUNTERCLOCKWISE.

CD Abbreviation of COMPACT DISK.

Cd Symbol for CADMIUM.

cd Abbreviation of CANDELA.

CD-4 A method of obtaining quadraphonic reproduction on a phonograph disk using modulated carriers with frequencies above the human hearing range.

CDI Abbreviation of CAPACITOR-DISCHARGE IGNITION.

C display A radar display showing the target as a dot whose coordinates represent the bearing (horizontal) and angle of elevation (vertical). Compare A DISPLAY, J DISPLAY, and K DISPLAY.

cd/m^2 Candelas per square meter, the SI unit of luminance.

CD-ROM Abbreviation of COMPACT-DISK READ-ONLY MEMORY.

Ce Symbol for CERIUM.

C_e Symbol for EMITTER CAPACITANCE of a transistor.

ceiling **1.** The maximum possible power output from a transmitter. **2.** The maximum possible current or voltage that a circuit can deliver. **3.** In aviation, the level of the cloud base.

ceilometer An instrument for measuring ceiling (cloud height).

cel In animated graphics, an individual image or frame.

cell **1.** A single (basic) unit for producing dc electricity by electrochemical or photovoltaic action, as in a battery or a solar panel. Also see PRIMARY CELL, STANDARD CELL, STORAGE CELL. **2.** An addressable, one-word-capacity storage element in a computer memory. **3.** The geographic region covered by a specified repeater in a cellular communications network. See CELLULAR COMMUNICATIONS. **4.** An electrostatic charge dipole in the atmosphere, usually occurring in or near thunderstorms. **5.** A thunderstorm.

cell constant The surface area of the electrodes in a cell divided by the distance between them. The basic linear units must be the same: for example, square centimeters for surface area and centimeters for distance.

cell counter A bioelectronic instrument used to count blood cells and other minute particles.

cell reversal A condition that can occur in some rechargeable electrochemical cells and batteries, such as *nickel-cadmium batteries.* It most often results from neglecting to recharge the cell or battery when it has become fully discharged.

cell-type enclosure A room designed to prevent the entrance or escape of radio-frequency (RF) electromagnetic fields, characterized by double-walled copper-mesh shielding.

cellular coil A coil having a crisscross (usually multilayer) winding. Examples: lattice-wound coil, honeycomb coil, basket-weave coil.

cellular communications A radio, telephone, or television communications network that makes use of numerous fixed repeaters. Subscribers use mobile or portable transceivers that are always within range of at least one repeater. The most common form is known as *cellular telephone* or *cellular mobile radio telephone.*

celluloid A thermoplastic dielectric material that is a blend of cellulose nitrate and camphor. Dielectric constant, 4 to 7. Dielectric strength, 250 to 780 V/mil.

cellulose acetate A plastic dielectric material used as a substrate for magnetic tapes, photographic film, and similar applications. Dielectric constant, 6 to 8. Dielectric strength, 300 V to 1 kV/mil. Also see ACETATE.

cellulose acetate base See ACETATE BASE.

cellulose acetate butyrate A thermoplastic dielectric material that is an acetic and butyric acid ester of cellulose.

cellulose acetate tape See ACETATE TAPE.

cellulose nitrate The nitric acid ester of cellulose, a plastic insulating material.

cellulose propionate A thermoplastic molding material that is a propionic acid ester of cellulose.

Celsius scale A temperature scale in which 0 degrees is the freezing point of water, and 100 degrees the boiling point of water. Also called CENTIGRADE SCALE. Compare ABSOLUTE SCALE, and FAHRENHEIT SCALE.

cent An audio-frequency interval of 0.01 (1⁄100) of a half step. A half step is the frequency difference between two immediately adjacent keys on a piano.

center channel In high-fidelity stereo, a phantom sound source that appears to exist midway between the left and right speakers or earpieces. The effect is caused by identical, or nearly identical, signals in the left and right channels.

center-fed antenna An antenna in which the feeders are connected to the center of the radiator.

center feed **1.** Attaching a feeder or transmission line to the center of the radiator of an antenna. **2.** Connection of signal-input terminals to the center of a coil. **3.** Descriptive of paper tape whose feed holes are aligned with character hole centers. Compare ADVANCE FEED TAPE.

center frequency **1.** The frequency, in a communications receiver, that is midway between the lower and upper 3-dB-down amplitude points. **2.** The average frequency of a modulated carrier. **3.** The carrier frequency of a modulated signal, whether or not the carrier is suppressed.

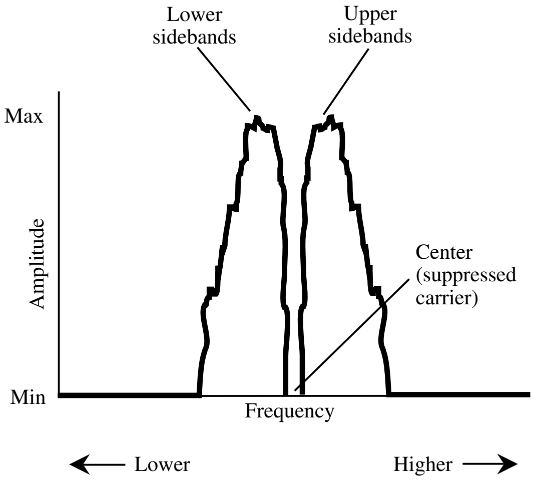

center frequency, 3.

centering control In an oscilloscope circuit, a potentiometer used to position the image on the screen (particularly in the center). Separate controls are provided for horizontal and vertical centering.

center loading In an inductively loaded antenna, placement of the loading coil(s) at or near a point or points midway between the feed point and the end(s) of the radiating element.

center of beam **1.** In a directional antenna system, the direction, denoted by a straight ray, where the signal strength or response is the greatest. **2.** In a beam of visible light, the geometric center of the spot produced when the beam strikes a surface perpendicular to the beam. **3.** In a beam of visible light, the axis within the beam where the intensity is greatest.

center of channel The frequency that is midway between the lowest and highest frequency components of a communications channel.

center of radiation The point from which the energy radiated by an object appears to arrive.

center tap A connection made to the centermost turn of a coil or to the center-value point of a resistor, filament, or capacitor pair.

center-tapped coil See CENTER-TAPPED WINDING.

center-tapped filament A tube or lamp filament that has a tap at its center.

center-tapped inductor An inductor that has a tap at half the total number of turns (the physical center of the winding).

center-tapped potentiometer A potentiometer that has a tap at half the total resistance of the resistance element.

center-tapped resistor A fixed resistor that has a tap at half the total resistance.

center-tapped transformer A transformer that has one or more center-tapped windings.

center-tapped winding A winding that has a tap at half the total number of turns (the physical center of the winding).

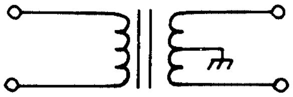

center-tapped winding

center tracking frequency In three-frequency alignment (tracking) of a circuit, the frequency between the upper and lower frequency limits (alignment or tracking points of the circuit).

center-zero meter A meter that has its zero point at the center of the scale (e.g., a dc galvanometer).

centi- Abbreviation, c. Prefix meaning hundredth(s) (10^{-2}).

centigrade scale See CELSIUS SCALE.

centimeter Abbreviation, cm. A unit of length equal to 10^{-2} meter, or 0.3937 inch.

centimeter-gram-second system Abbreviation, cgs. A system of units, now seldom used, in which the centimeter is the fundamental unit of length, the gram is the fundamental unit of mass, and the mean solar second is the fundamental unit of time. Electrical units in the cgs system fall into two categories: *electrostatic* and *electromagnetic.* The names of cgs electrostatic units have the prefix *stat-* (e.g., STATAMPERE, STATVOLT, etc.). Cgs electromagnetic units have the prefix *ab-* (e.g., ABAMPERE, ABVOLT, etc.).

centimetric waves See MICROWAVES.

centipoise A cgs measure of the dynamic viscosity of liquids. Equal to 10^{-2} poise.

central office In telephone systems, a switching network at which numerous circuits or subscriber lines converge.

central processing unit Abbreviation, CPU. In a digital computer, the section containing the arithmetic and logic unit (ALU), control circuits, and internal memory circuits. Also called *central processor.*

Central Radio Propagation Laboratory A government laboratory that studies radio propagation and collects, correlates, and analyzes data for predicting propagation conditions. The organization also studies methods of measuring propagation.

centrifugation potential An electric potential that occurs in a colloidal solution when the solution is centrifuged.

centrifugal switch A switch actuated by rotational motion (e.g., the automatic disconnection switch in a capacitor motor).

centripetal force The force that draws the mass of a rotating body toward the axis of rotation.

ceramal See CERMET.

ceramet seal See CERAMIC-TO-METAL SEAL.

ceramic-based microcircuit A tiny circuit printed or deposited on a ceramic substrate.

ceramic capacitor A component made with sheets of metal stacked alternately with wafers of ceramic. This material, like mica, has low loss, and therefore allows for high efficiency. For low values of capacitance, only one layer of ceramic is needed, and two metal plates can be glued to a disk of porcelain, one on each side. Alternatively, a tube or cylinder of ceramic can be employed, and metal ink applied to the inside and outside of the tube. These capacitors have values ranging from a few picofarads to about 0.5 µF. Their voltage ratings are comparable to those of paper capacitors. Compare ELECTROLYTIC CAPACITOR, MICA CAPACITOR, PAPER CAPACITOR, PLASTIC-FILM CAPACITOR, TANTALUM CAPACITOR.

ceramic dielectric **1.** A ceramic used as a dielectric in capacitors. Examples: barium titanate, barium strontium titanate, and titanium dioxide. Ceramic dielectrics provide high dielectric constant. **2.** A ceramic used as an insulator. Examples: isolantite, porcelain, and steatite.

ceramic filter A resonant filter similar to a crystal filter, but using a piezoelectric ceramic material.

ceramic magnet A permanent magnet made of a magnetic ceramic material, such as mixtures of barium oxide and iron oxide.

ceramic microphone A microphone that uses a CERAMIC PIEZOELEMENT to convert sound waves into electrical impulses.

ceramic piezoelement A component that uses a piezoelectric ceramic material. Examples: ceramic filter, ceramic microphone, ceramic phono pickup, ceramic transducer, and electrostrictive transducer. Also called PIEZOELECTRIC CERAMIC.

ceramic resistor A carborundum resistor whose value is voltage-dependent. It usually displays a negative temperature coefficient of resistance (but a positive coefficient is available) and a negative voltage coefficient of resistance.

ceramics **1.** Clay-based materials used as dielectrics and insulators in electronics. Examples: barium titanate, titanium dioxide, porcelain, isolantite, and steatite. **2.** The science and art of using and developing ceramics.

ceramic-to-metal seal A bond in which ceramic and metal bodies are joined, for example, the bonding of a metal lead to a ceramic disk, through which it passes to provide a leak-proof seal. Also called *ceramet seal.*

ceramic transducer A transducer that uses a CERAMIC PIEZOELEMENT to translate such parameters as pressure and vibration into electrical pulses.

ceramic tube A high-temperature vacuum tube that uses a ceramic material, instead of glass, as the envelope; the tube offers low losses at high frequencies.

Cerenkov radiation Light emanating from a transparent material that is traversed by charged particles, whose speed is higher than the speed of light through the material.

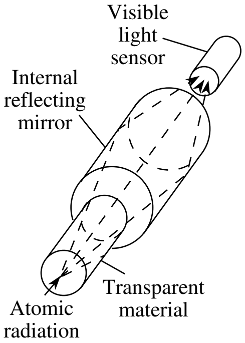

Cerenkov radiation

Cerenkov rebatron device An apparatus for generating radio-frequency energy by passing an electron beam through a piece of dielectric having a small aperture.

ceresin wax A yellow or white wax obtained by refining *ozocerite.* Used as an insulant and sealant against moisture. Dielectric constant, 2.5 to 2.6.

cerium Symbol, Ce. A metallic element of the rare-earth group. Atomic number, 58. Atomic weight, 140.13.

cerium metals A group of metals belonging to the rare-earth group: cerium, lanthanum, neodymium, praseodymium, promethium, and samarium.

cermet An alloy of a ceramic, such as titanium carbide, and nickel, a metal. A thin film of cermet is used as a resistive element in some microcircuits. Cermet is an acronym for *ceramic metal.*

certified tape A magnetic recording tape that has been thoroughly checked and found to have no flaws.

cesium Symbol, Cs. A metallic element of the alkali-metal group. Atomic number, 55. Atomic weight, 132.91. The oscillations of this element's atoms have been used as atomic time standards. The element is used in some phototubes as the light-sensitive material, and in some arc lamps.

cesium-vapor lamp A low-voltage arc lamp used as an infrared source.

Cf Symbol for CALIFORNIUM.

cgs Abbreviation of CENTIMETER-GRAM-SECOND.

chad The punched-out particle(s) constituting refuse from paper-tape punching.

chadded tape Punched paper tape in which the chad is left partially attached to the tape's punched holes.

chadless tape Punched paper tape without CHAD.

chafe **1.** An area that has been abraded by rubbing or scraping. **2.** To produce a chafe.

chaff Strips of metal foil used to create radar interference or ambiguity in locating a target by multiple reflections of the beam. Also called MIRROR.

chain broadcasting Simultaneous transmissions from a number of broadcast transmitters connected together in a network by wire line, coaxial cable, or microwave link.

chain calculation As performed by a calculator, a calculation that can be entered as it would normally be written (i.e., without the need for regrouping operands).

chain printer In the readout channel of a digital computer, a high-speed printer carrying printer's type on a revolving chain.

chain radar system A number of radar stations along a missile-flight path that are connected in a communications or control network.

chain reaction A reaction (as in nuclear fission) that is self-sustaining or self-repeating. Unless controlled from outside, such a reaction runs to destruction.

chain switch A switch that is actuated by pulling a light metal chain. Successive pulls turn the switch alternatively on and off.

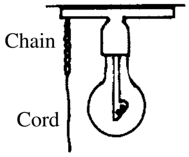

chain switch

change dump In computer operation (especially in debugging), the display of the names of locations that have changed following a specific event.

change file See TRANSACTION FILE.

change of control In a sequence of computer records being processed, a logical break that initiates a predetermined action, after which processing continues.

changer In a high-fidelity disk player, a device that allows several disks to be played, one after the other, without the need for manually exchanging the disks.

change record A computer record that changes information in a related master record. Also called *transaction record.*

change tape See TRANSACTION TAPE.

channel **1.** A frequency (or band of frequencies) assigned to a radio or television station. **2.** See KEYWAY. **3.** A subcircuit in a large system [e.g., the radio-frequency (RF) channel of a receiver, the vertical-amplifier channel of an oscilloscope, or the modulator channel of a radio transmitter]. **4.** The end-to-end electrical path through the semiconductor body in a field-effect transistor. **5.** One of the independent audio circuits in a stereo sound system (e.g., the left channel or the right channel).

channel analyzer A (usually multiband) continuously tunable instrument, similar to a tuned radio receiver, used in troubleshooting radio communications circuits by substituting a perfect channel for one that is out of order.

channel balance The state in which the apparent amplitude of two or more channels is identical.

channel bank In a transmission system, the terminal equipment used for the purpose of multiplexing the individual channels.

channel capacity The fullest extent to which a channel can accommodate the information (frequencies, bits, words, etc.) to be passed through it.

channel designator A name, number, or abbreviation given to a channel in a communications system.

channel effect The possible current flow through a high impedance between the collector and emitter in a bipolar transistor.

channel frequency The CENTER FREQUENCY of a communications channel.

channeling Multiplex transmission in which separate carriers within a sufficiently wide frequency band are used for simultaneous transmission.

channelizing The subdivision of a relatively wide frequency band into a number of separate subbands.

channel reliability **1.** The proportion of time, usually expressed as a percentage, that a communications channel is useful for its intended purpose. **2.** The relative ease with which communications can be carried out over a particular channel.

channel reversal In stereo reproduction, interchanging the left and right channels.

channel-reversing switch In a stereo system, a switch that allows channel reversal without the need for reorienting speaker cables or connectors.

channel sampling rate The rate at which individual channels are sampled. For example, in the electronic switching of an oscilloscope, the number of times per second each input-signal channel is switched to the instrument.

channel selector A switch or relay used to put any of a series of channels into functional status in a system.

channel separation **1.** The spacing between communications channels, expressed in kilohertz. **2.** In stereo reproduction, the degree to which the information on one channel is separate from the other; usually expressed in decibels.

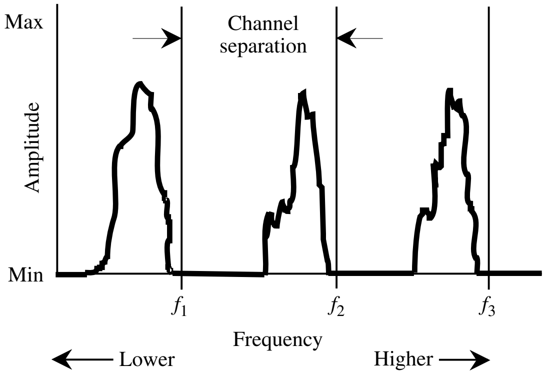

channel separation, 1.

channel slot On a carrier modulated by numerous signals, the position or frequency of a specific modulating signal.

channel shift The interchange of communications channels (e.g., the shift from a calling frequency to a working frequency).

channel strip A fixed-channel amplifier for a television receiver.

channel time slot In a frame of transmitted information, such as a television picture, a time interval designated to a channel for the transmission of a character signal or other information.

channel-to-channel connection A device, such as a channel adapter, used to transfer data rapidly between any two channels of two digital computers, at the data speed of the slower channel.

channel-utilization index An indication of the extent to which channel capacity is used. For a given channel, the index is the ratio of information rate to channel capacity, each expressed in units per second.

channel wave An acoustic wave that travels within a region or layer of a substance because of a physical difference between that layer and the surrounding material. An example of a channel wave is the propagation of sound over a still lake.

channel width In a frequency channel, the difference $f_2 - f_1$, where f_1 is the lower-frequency limit and f_2 is the upper-frequency limit of the channel.

chapter A self-contained computer program section.

character **1.** One of the symbols in a code. **2.** In computer operations, a digit, letter, or symbol used alone or in some combination to express information, data, or instructions.

character code In a communications or computer system, the combination of elements (e.g., bits) representing characters.

character crowding A reduction of the time interval between successive characters—especially those read from tape.

character density The number of characters that can be stored in a given length or surface area of a medium. On a magnetic tape, it might be specified in characters per millimeter; on a magnetic disk, it might be specified in characters per square millimeter.

character emitter A coded-pulse generator in a digital computer.

character generator A device that converts coded information into readable alphanumeric characters.

characteristic **1.** A quantity that characterizes (typifies) the operation of a device or circuit. Examples are emitter current, output power, and frequency deviation. **2.** In floating point notation, the exponent.

characteristic curve A curve showing the relationship between an independent variable and a dependent variable, with respect to the parameter(s) for a device or circuit. Example: the collector voltage-collector current characteristic curve of a transistor.

characteristic distortion **1.** In a digital signal, pulse distortion caused by the effects of the previous pulse or pulses. **2.** Distortion in the characteristic curve of a component or device.

characteristic frequency The frequency peculiar to a given channel, service, or response.

characteristic impedance Symbol, Z_0. **1.** Theoretically, the impedance that would be simulated by a given two-conductor or coaxial line of uniform construction, if that line were of infinite length. This value is determined by the materials used for the two conductors, the dielectric used to insulate the two conductors, the diameters of the conductors, and the spacing between them. **2.** In practice, for a transmission line or waveguide terminated with a load that produces no standing waves on the line, the ratio of radio-frequency (RF) voltage to RF current. This ratio is the same at all points along the length of a perfectly matched line, and depends on the physical construction of the line. Coaxial lines typically have Z_0 between 50 and 100 ohms. Twinlead is available with 75-ohm and 300-ohm Z_0 values. Open-wire line has Z_0 between 300 and 600 ohms, depending on the spacing between the conductors, and also on the type of dielectric (insulating material) employed to keep the spacing constant between the conductors. **3.** Experimentally, the value of impedance that, if it terminates a transmission line or waveguide, results in no reflected power from the load end of line. This is always a pure resistance; that is, it contains no reactance.

characteristic overflow In floating-point arithmetic, the condition that occurs when a characteristic exceeds the upper limit specified by a program or computer.

characteristic spread The range of values over which a characteristic extends. For example, if an amplifier's output ranges from 15 W to 25 W, its characteristic spread is 10 W.

characteristic underflow In floating-point arithmetic, the condition that occurs when a characteristic exceeds the lower limit specified by a program or computer.

character modifier In address modification, a constant (compare VARIABLE) that refers to a specific character's location in memory.

character-oriented A computer in which character locations, rather than words, can be addressed.

character printer A computer output device that prints matter in the manner of a conventional typewriter.

character reader Also called an *optical scanner.* In a digital computer, an input device that can read printing and script directly.

character recognition The reading of a written or printed character by a computer, including its identification and encoding.

character sensing The detection of characters by a computer input device. This can be done galvanically, electrostatically, magnetically, or optically.

character set The set of characters in a complete language, or in a communications system.

character signal The set of elements or bits representing a character in a digital transmission system. The signal can also represent the quantizing value of a sample.

characters per minute An expression of the speed of transmission of a digital signal. The number of characters (on average) transmitted in a period of one minute. In Morse code (CW) transmission, this is generally taken as the number of times the word *paris* plus the subsequent space, multiplied by six (five letters and one space following), can be sent in one minute.

characters per second An expression of the speed of transmission of a digital signal. The number of characters (on average) transmitted in a period of one second.

character string A one-dimensional character array [i.e., a list of characters that, when printed or displayed, would appear in a row or column, but not both (as in a matrix)].

character subset A classification of characters within a set.

Charactron A cathode-ray readout tube that displays letters, numbers, and symbols on its screen. More commonly called a *monitor.*

charcoal tube In a system for producing a high vacuum, a trap containing activated charcoal, which is heated to dull red, then cooled by liquid air to absorb gases.

charge **1.** A quantity of electricity associated with a space, particle, or body. **2.** To electrify a space, particle, or body (i.e., to give an electric charge). **3.** To store electricity, as in a storage battery or capacitor. Compare DISCHARGE.

charge carrier **1.** An ELECTRON whose movement constitutes a flow of electric current. **2.** An electron deficiency (HOLE) whose movement constitutes a flow of electric current. **3.** Any particle, such as a charged atom (ION), PROTON, ALPHA PARTICLE, or BETA PARTICLE, whose movement constitutes a flow of electric current.

charge-coupled device Abbreviation, CCD. A form of analog-to-digital converter that generates a digital signal output representing an analog image input. The transfer of stored charges provides the method of operation. Used in machine vision systems and in numerous scientific applications.

charge density The degree of charge or current-carrier concentration in a region.

charged particle **1.** See CHARGE CARRIER. **2.** See ION.

charged voltage **1.** The voltage across a fully charged capacitor. **2.** The terminal voltage of a fully charged storage cell.

charge holding See CHARGE RETENTION.

charge of electron The negative electric charge carried by a single electron. Approximately equal to 1.602×10^{-19} coulombs.

charger **1.** See BATTERY CHARGER. **2.** Any device or circuit that charges a capacitor.

charge retention **1.** The holding of an electric charge by a cell or battery when no current is being drawn from it. **2.** A measure of the ability of a cell or battery to maintain an electric charge when no current is drawn from it. Often specified in terms of *shelf life.* **3.** The holding of a charge by a capacitor.

charge-storage tube A cathode-ray tube that holds a display of information on its screen until the operator removes it by pressing an erase button.

charge-to-mass The ratio of the electric charge to the mass of a subatomic particle.

charge-to-mass ratio of electron The ratio of the charge (e) of the electron to the mass (m_e) of the electron, in coulombs per kilogram (C/kg). For an electron at rest, e/m_e is approximately equal to 1.602×10^{-19} C divided by 9.11×10^{-31} kg = 1.76×10^{11} C/kg.

charge transfer **1.** The switching of an electric charge from one capacitor to another. **2.** The capture of an electron by a positive ion from a neutral atom of the same kind, resulting in the ion becoming a neutral atom, and the previously neutral atom becoming a positive ion.

charge transfer device A semiconductor in which an electric charge is moved from location to location. Applications include delay lines, video signal processing, and signal storage.

charging **1.** The process of storing electrical energy in a capacitor. **2.** The process of storing electrochemical energy in a storage cell or battery.

charging current **1.** The current flowing into a capacitor. **2.** The current flowing into a previously discharged storage cell.

charging rate **1.** The rate at which charging current flows into a storage cell or battery, expressed in amperes or milliamperes. For most cells and batteries, the rate is greatest initially, when the cell or battery is depleted or nearly depleted; the rate decreases as the cell or battery becomes charged. **2.** The instantaneous rate at which charging current flows into a capacitor or capacitance-resistance circuit, expressed in amperes, milliamperes, or microamperes.

charged voltage **1.** The voltage across a fully charged capacitor. **2.** The terminal voltage of a fully charged storage cell.

Charlie Phonetic alphabet code word for the letter C.

chassis A (usually metal) foundation on which components are mounted and wired.

chassis ground A ground connection made to the metal chassis on which the components of a circuit are mounted. When several ground connections are made to a single point on the chassis, a COMMON GROUND results.

chatter **1.** A rapidly repetitive signal, caused by interruption or variation of a current (usually interference). **2.** Extraneous vibration, as of the armature in a relay.

chatter time The interval between the instant that contacts close (for example, in a relay) and the instant at which chatter ends.

cheater cord An extension cord used to conduct power to a piece of equipment (especially a television receiver) by temporarily bypassing the safety switch or interlock. Use of such a cord presents a potentially fatal shock hazard to personnel using, or working on, the equipment.

Chebyshev filter Also spelled *Tschebyscheff* or *Tschebysheff.* A form of inductance-capacitance (LC) lowpass, highpass, bandpass, or band-rejection filter, characterized by an attenuation-versus-frequency curve with ripple in the passband.

check **1.** A test generally made to verify condition, performance, state, or calculations; specifically, in computer operations, it applies to operands or results. **2.** The usually abrupt halting of an action.

check bit A binary CHECK DIGIT.

check character In a group of characters, one whose value depends on the other characters, which it checks when the group is stored or transferred.

check digit Also called *check number.* In computer operations, a number added to a group of digits, forming a code that identifies entities in the system (including personnel) and can be used for verification. The check digit is the remainder when the number code (for example, 459) is divided by a fixed number (for example, 5); in this case, the check digit (the remainder of 459/5) is 4, and the amended code number is 4594.

check indicator An indication, made via a video display, that something has been shown to be invalid according to a check.

checking program Also called *checking routine.* For debugging purposes, a diagnostic computer program capable of detecting errors in another program.

checkout A test routine that ascertains whether or not a circuit or system is functioning according to specifications.

checkout routine A routine used by programmers to debug programs.

checkpoint A point in a digital-computer program at which sufficient information has been stored to allow restarting the computation from that point.

checkpoint dump The process of recording details of a computer program run. This process might be necessary in the event of a system failure that requires reconstruction of a program or programs.

checkpointing The writing of a computer program in such a manner that, during a program run, information is frequently dumped as insurance against possible loss in the event of a system failure.

check problem A presolved problem used to check the operation of a digital computer or program.

check register In some digital computers, a register in which transferred information is stored so that it can be checked against the same information as it is received a second time.

check routine A special program designed to ascertain if a program or computer is operating correctly. Also see CHECK PROBLEM.

checksum Used as part of a summation check, a sum derived from the digits of a number. For example, the checksum of 23,335 is 16. Also called HASH TOTAL.

check symbol For a specific data item, a digit or digits obtained by performing an arithmetic check on the item, which it then accompanies through processing stages for the purpose of checking it.

check total See CONTROL TOTAL.

check word A check symbol in the form of a word added to, and containing data from, a block of records.

chelate Pertaining to cyclic molecular structure in which several atoms in a ring hold a central metallic ion in a COORDINATION COMPLEX.

chemical deposition The coating of a surface with a substance resulting from *chemical reduction* of a solution. In mirror making, for example, formaldehyde reduces a solution of silver nitrate, and deposits metallic silver on the surface of polished glass. Also see CHEMICALLY DEPOSITED PRINTED CIRCUIT and CHEMICAL REDUCTION.

chemical detector See ELECTROLYTIC DETECTOR.

chemical effect An alteration in the chemical makeup of a substance or solution, resulting from the passage of an electric current through it. Examples include electrolysis, electroplating, and the reduction of ores.

chemical energy Energy that is stored in the chemical bonds of a material or solution. An example is the stored energy in terms of watt hours in an electrolytic cell.

chemical load An arrangement of a chemical material or device for the passage of electricity through it. Examples: electroplater, electrolytic cell for the production of hydrogen gas, and storage battery.

chemically deposited printed circuit A printed circuit in which the pattern of metal lines and areas are chemically deposited on a substrate.

chemically pure Abbreviation, CP. Free from impurities.

chemical rectifier See ELECTROLYTIC CELL.

chemical reduction The process of making a chemical compound (usually in solution) into a metal, by removing the nonmetallic component from the compound. For example, when copper oxide is heated in the presence of hydrogen (a reducting agent), the oxygen (the nonmetallic component) is driven out, and copper (along with some water) remains.

chemical resistor See ELECTROLYTIC RESISTOR.

chemical switch See ELECTROCHEMICAL SWITCH.

CHIL Abbreviation for *current-hogging injection logic.* A form of bipolar digital logic technology.

chip **1.** An INTEGRATED CIRCUIT. **2.** A small slab, wafer, or die of dielectric or semiconductor material, on which a subminiature component or circuit is formed or deposited.

chip capacitor A subminiature capacitor formed on a chip.

chip resistor A subminiature resistor formed on a chip.

chip tray A chad receptacle located at a card or paper tape punching site.

Chireix-Mesny antenna A high-frequency (HF) beam antenna, in which each dipole section constitutes one side of a diamond. Cophased horizontal and vertical components of current flow in each of the diagonals, and radiation is broadside to the plane of the driven element.

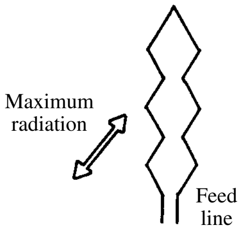

Chireix-Mesny antenna

chirp A rapid change in the frequency of a continuous-wave Morse-code signal. The chirp usually occurs at the beginning of each dot or dash, and can go up or down in frequency. Chirp occurs because of a change in the output impedance of an oscillator as it is keyed. Modern code transmitters do not exhibit significant chirp.

chirp modulation A form of modulation in which the frequency of a signal is deliberately changed in a systematic way. Used in some radar systems.

chirp radar A radar system that uses CHIRP MODULATION.

Chladni's plates Conducting plates that are used to evaluate the nature of a vibration in a solid material. The plates are clamped to the material, and sand is sprinkled on the surface. This produces patterns that indicate the nature of the vibrations.

chlorinated diphenyl A synthetic organic substance used as an impregnant in some oil-filled capacitors.

chlorinated naphthalene See HALOWAX.

chlorine Symbol, Cl. A gaseous element of the halogen family. Atomic number, 17. Atomic weight, 35.453.

choke **1.** To restrict or curtail passage of a particular current or frequency by means of a discrete component, such as a choke coil. **2.** See CHOKE COIL.

choke air gap A fractional-inch opening in the iron core of a filter choke, usually filled with wood or plastic. The gap prevents saturation of the core when the choke coil carries maximum rated direct current.

choke coil **1.** A large-value inductor that provides a high impedance to alternating current (ac), while offering virtually no opposition to direct current (dc). **2.** In radio-frequency (RF) applications, an inductor that provides a high impedance to RF signals while showing low impedance for audio-frequency (AF) signals and direct currents (dc).

choke-coupled modulation An amplitude-modulation (AM) scheme, in which the modulator is coupled to the radio-frequency (RF) amplifier through a shared iron-core choke coil.

choke flange At the end of a waveguide, a flange in which a groove forms a CHOKE JOINT.

choke-input filter A filter whose input component is an inductor (choke). The choke-input power-supply filter is distinguished by its superior voltage regulation, compared with the CAPACITOR-INPUT FILTER.

choke joint A joint connecting two waveguide sections and permitting efficient energy transfer without requiring electrical contact with the inside wall of the waveguide.

chopped dc See INTERRUPTED DC.

chopped mode In a single-gun cathode-ray-tube (CRT) oscilloscope, a technique for sequentially displaying several signals that are not referenced to the oscilloscope sweep.

chopped signal An ac or dc signal that is periodically interrupted, as by means of a CHOPPER.

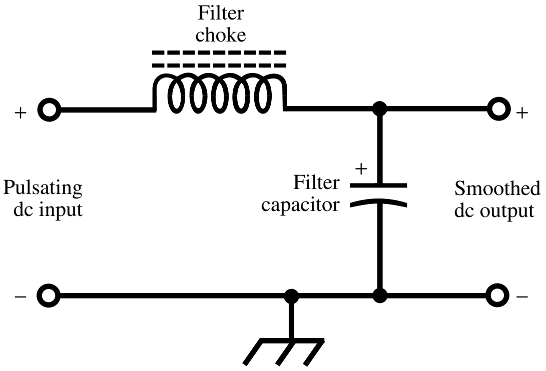

choke-input filter

chopper A device or circuit that interrupts a direct current (dc) at some predetermined rate. Ideally, such a device is characterized by distinct on and off operation.

chopper amplifier A circuit that amplifies the output of a CHOPPER. Used in conjunction with a CHOPPER CONVERTER in dc amplification.

chopper converter A device that interrupts a direct current (dc), and changes it to a pulsating, rectangular-wave current or voltage that can be handled by a stable ac amplifier and rectified to supply amplified dc.

chopper power supply Also called *power inverter.* A circuit that delivers high-voltage ac from a dc source. The input is typically 12 volts dc, and the output is usually 117 volts rms ac. These devices facilitate the use of small appliances such as computers, television sets, and communications radios in portable and mobile environments. The output of a low-cost power inverter is generally not a good sine wave. More sophisticated inverters produce good sine waves and have a frequency close to 60 Hz.

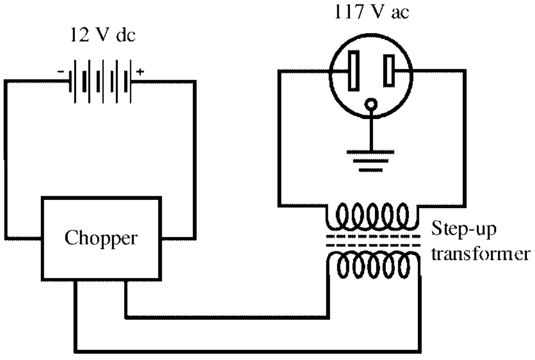

chopper power supply

chopper stabilization **1.** Stabilization of direct-current (dc) amplification by using a CHOPPER CONVERTER ahead of a stable ac amplifier, and rectifying the amplifier output. **2.** In a regulated power supply, use of a CHOPPER AMPLIFIER at the control-circuit input to improve regulation.

chopper-stabilized amplifier See CHOPPER AMPLIFIER and CHOPPER STABILIZATION, **1.**

chopper transistor A transistor that provides rapid and repeated on/off switching of direct current (dc), in the manner of an electromechanical interrupter. See CHOPPER.

chopping frequency The frequency at which a chopper interrupts a signal.

chord **1.** A harmonious mixture of musical tones of various frequencies. **2.** A straight line that joins two points on a curve (such as an arc of a circle). **3.** The width of an airfoil.

chord organ An electronic organ that will sound a musical chord when a key is pressed (see CHORD, **1**).

choreographer program A computer program similar to one originally written by Charles Lecht of Lecht Sciences, Inc. The computer operator gives commands that cause a human form, portrayed on the display screen, to make various movements. Used in animated computer graphics.

chorus Signals at very low radio frequencies (VLF), natural in origin, that sweep upward in frequency. Believed to result from lightning-generated electromagnetic fields that circulate in the magnetosphere (earth's magnetic field). The term is derived from the sound the signals make in high-gain audio-frequency (AF) amplifiers connected directly to VLF receiving antennas.

Christiansen antenna A radio-telescope antenna for obtaining high resolution. Two straight arrays are placed at an angle, intersecting approximately at their centers. The resulting interference pattern has extremely narrow lobes.

Christmas tree A tree-like pattern on the screen of a television receiver, caused by loss of horizontal synchronization.

chroma The quality of a color: *hue* and *saturation.*

chroma circuit In color television, one of several circuits whose ultimate purpose is to produce a color component on the screen.

chroma-clear raster In color television reception, the clear raster resulting from a white video signal, or from operation of the chroma circuits of the receiver (as if they were receiving a white transmission). Also called *white raster.*

chroma control In a color television receiver, a rheostat or potentiometer that permits adjustment of color saturation through variation of the chrominance-signal amplitude before demodulation.

chromatic fidelity See COLOR FIDELITY.

chromaticity **1.** The state of being chromatic (see CHROMA). **2.** A quantitative assessment of a color in terms of dominant or complementary wavelength and purity.

chromaticity coordinate For a color sample, the ratio of any one of the three tristimulus values (primary colors) to the sum of the three.

chromaticity diagram A rectangular-coordinate graph in which one of the three CHROMATICITY COORDINATES of a three-color system is plotted against another coordinate.

chromaticity flicker Flicker caused entirely by chromaticity fluctuation (see CHROMATICITY, **2**).

chromel A nickel-chromium alloy with some iron content, used in thermocouples.

chromel-alumel junction A thermocouple that uses wires of the alloys *chromel* and *alumel.*

chromel-constantan thermocouple A thermocouple consisting of a junction between wires or strips of *chromel* and *constantan.* Typical output is 6.3 mV at 100°C.

chrome plating The process of coating a metal with chromium. Generally protects against corrosion.

chrome recording tape Also called *chrome tape* or *chromium tape.* Tape that is manufactured from the compound chromium dioxide. Noted for its ability to faithfully record and reproduce music.

chrominance In color television, the difference between a reproduced color and a standard reference color of the same luminous intensity.

chrominance amplifier In a color television circuit, the amplifier separating the chrominance signal from the total video signal.

chrominance cancellation On a black-and-white picture tube screen, cancellation of the fluctuations in brightness caused by a chrominance signal.

chrominance-carrier reference In color television, a continuous signal at the frequency of the chrominance subcarrier; it is in fixed phase with the color burst and provides modulation or demodulation phase reference for carrier-chrominance signals.

chrominance channel In color television, a circuit devoted exclusively to the color function, as opposed to audio and general control channels.

chrominance component In the NTSC color television systems, either of the components (I-signal or Q-signal) of the complete chrominance signal.

chrominance demodulator In a color television receiver, a demodulator that extracts video-frequency chrominance components from the chrominance signal, and a sine wave from the chrominance subcarrier oscillator.

chrominance gain control A rheostat or potentiometer in the red, green, and blue matrix channels of a color television receiver, used to adjust the primary-signal amplitudes.

chrominance modulator In a color television transmitter, a device that generates the chrominance signal from the I and Q components and the chrominance subcarrier.

chrominance primary One of the transmission primaries (red, green, and blue) upon which the chrominance of a color depends.

chrominance signal The signal component in color television that represents the hues and saturation levels of the colors in the picture.

chrominance subcarrier In color television, the 3579.545-kHz signal that serves as a carrier for the I- and Q-signals.

chrominance-subcarrier oscillator In a color television receiver, a crystal-controlled oscillator that generates the subcarrier signal (see CHROMINANCE SUBCARRIER).

chrominance video signals Output signals from the red, green, and blue channels of a color television camera or receiver matrix.

chromium Symbol, Cr. A metallic element. Atomic number, 24. Atomic weight, 51.996. Commonly used as a plating for metals to improve resistance to corrosion.

chronistor An elapsed-time indicator in which current, flowing during a given time interval, electroplates an electrode. The duration of the interval is determined from the amount of deposit.

chronograph **1.** An instrument that provides an accurate time base along the horizontal axis of its permanent record. **2.** Stopwatch.

chronometer A precision clock. Electronic chronometers often use a highly accurate and stable crystal oscillator, followed by a string of multivibrators to reduce the crystal frequency to an audio frequency (such as 1 kHz) that drives the clock motor.

chronoscope An instrument for precisely measuring small time intervals.

CHU Call letters of the Canadian time-signal station whose primary frequency is 7.335 MHz.

Ci Symbol for INPUT CAPACITANCE.

CIE Abbreviation for *International Commission on Illumination.*

cinching In a reel of magnetic tape, the slipping of tape as force is applied.

cinematograph See KINEMATOGRAPH.

cipher A code used for the purpose of preventing interception of a message by third parties.

circ **1.** Abbreviation of *circuit.* **2.** Abbreviation of *circular.*

circle graph Also called a *pie graph.* A representational device consisting of a disk subdivided into various triangular areas (radiating from the center of the circle), which are proportional to represented quantities.

circle of confusion A circular image of a point source of light, resulting from an aberration in an optical system.

circle of declination The graduated circular scale of a *declinometer.*

circlotron amplifier A high-powered microwave amplifier of the one-port, cross-field, nonlinear type using a magnetron.

circuit **1.** A closed path through which current flows from a generator, through various components, and back to the generator. (An electronic circuit is often a combination of interconnected subcircuits.) **2.** The wiring diagram of an electronic device or system.

circuit analysis The careful determination of the nature and behavior of a circuit and its various parts. It can be theoretical, practical, or both. Compare CIRCUIT SYNTHESIS.

circuit analyzer See CIRCUIT TESTER.

circuit board A panel, plate, or card on which electronic components are mounted and interconnected to provide a functional unit.

circuit breaker A resettable fuse-like device that is designed to protect a circuit against overloading. In a typical circuit breaker, the winding of an electromagnet is connected in series with the load circuit and with the switch contact points. Excessive current through the magnet winding causes the switch to be opened.

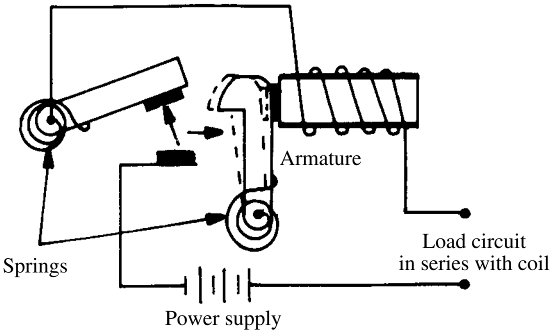

circuit breaker

circuit capacitance The total capacitance (lumped, distributed, and stray) present in a circuit.

circuit capacity **1.** The ability of a circuit to handle a quantity (such as current, voltage, frequency, power, etc.) safely and efficiently. **2.** The maximum value of some parameter at which a circuit can function safely and efficiently (e.g., a circuit capacity of 50 A). **3.** The number of channels that can be accommodated simultaneously by a circuit.

circuit component **1.** Any of the electronic devices or parts (capacitors, resistors, transistors, etc.) that are connected through wiring to form a circuit. **2.** An electrical quantity required for, or arising from, circuit operation. Examples: input voltage, feedback current, stray capacitance, and circuit noise.

circuit diagram A drawing in which symbols and lines represent the components and wiring of an electronic circuit. Also called CIRCUIT

SCHEMATIC, SCHEMATIC DIAGRAM, and WIRING DIAGRAM.

circuit dropout A momentary interruption of circuit operation, often caused by a break in the circuit.

circuit efficiency A quantitative measure of the effectiveness of circuit operation, customarily expressed as the ratio of the useful output power to the total input power.

circuit element See CIRCUIT COMPONENT, **1.**

circuit engineer An electronics engineer who specializes in circuit analysis, circuit synthesis, or both.

circuit fault **1.** Malfunction of a circuit. **2.** An error in circuit wiring.

circuit hole A perforation within the conductive area of a printed-circuit board, for the insertion and connection of a pigtail, terminal, etc., or for connecting the conductors on one side of the board with those on the other.

circuit loading Intentionally or unintentionally drawing power from a circuit.

circuit noise **1.** Electrical noise generated by a circuit in the absence of an applied signal. **2.** In wire telephony, electrical noise as opposed to acoustic noise.

circuit noise level The ratio of circuit-noise amplitude to reference-noise amplitude, expressed in decibels above the reference amplitude.

circuit-noise meter A meter that measures the intensity of the noise generated within a circuit.

circuit parameter See CIRCUIT COMPONENT, **2.**

circuit protection Automatic safeguarding of a circuit from damage from overload, excessive drive, heat, vibration, etc. Protection is afforded by various devices and subcircuits, ranging from the common fuse to sophisticated limiters and breakers.

circuit reliability A quantitative indication of the ability of a circuit to provide dependable operation as specified. See MEAN TIME BEFORE FAILURE and MEAN TIME BETWEEN FAILURES.

circuitry **1.** Collectively, electronic and electrical circuits. **2.** A detailed plan of a circuit and its subcircuits. **3.** Collectively, the components of a circuit.

circuit schematic See CIRCUIT DIAGRAM.

circuit simplification **1.** In circuit analysis, the reduction of a complex circuit to its simplest representation to minimize labor and to promote clarity. Thus, through application of Kirchhoff's laws, a complicated circuit could theoretically be reduced to a single generator in series with a single impedance. **2.** In circuit synthesis, the arrangement of a circuit so as to provide desired performance with the fewest components and least-complex wiring.

circuit switching In telephony, a method of connection in which a single circuit is maintained between two subscribers for the entire duration of the call. The signal path does not change. The connection is maintained even during periods of silence (no data transmitted by either subscriber). Compare PACKET SWITCHING.

circuit synthesis The development of a circuit under the guidance of theoretical or practical knowledge of basic electronics principles and component parameters. Compare CIRCUIT ANALYSIS.

circuit tester An instrument for checking the performance of electronic circuits. Often consists of a specialized continuity tester, but occasionally it includes a dynamic performance tester.

circuit tracking The alignment and/or pretuning of circuits for identical or optimum response. It applies especially to cascaded circuits, whose variable elements, such as tuned inductance-capacitance (LC) networks, must follow each other in step when ganged together.

circular angle The angle described by a radius vector as it rotates counterclockwise around a circle.

circular antenna A half-wave horizontally polarized antenna, whose driven element is a rigid conductor bent into a circle with a break opposite the feed point. Also called *halo antenna.* Used primarily at very-high frequencies (VHF).

circular electric wave An electromagnetic wave with circular electric lines of flux. An example is the field in the immediate vicinity of a CIRCULAR ANTENNA.

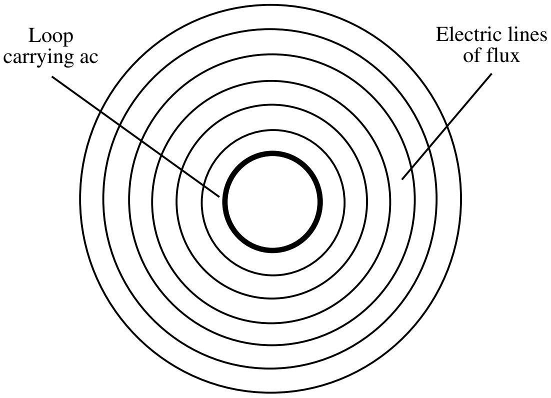

circular electric wave

circular functions Trigonometric functions of the angle described by a vector rotating counterclockwise around a circle. Also see COSINE, COSECANT, COTANGENT, SECANT, SINE, and TANGENT.

circular magnet See RING MAGNET.

circular magnetic wave An electromagnetic wave in which the magnetic lines of flux are circular. An example is the field in the immediate vicinity of a straight-conductor antenna.

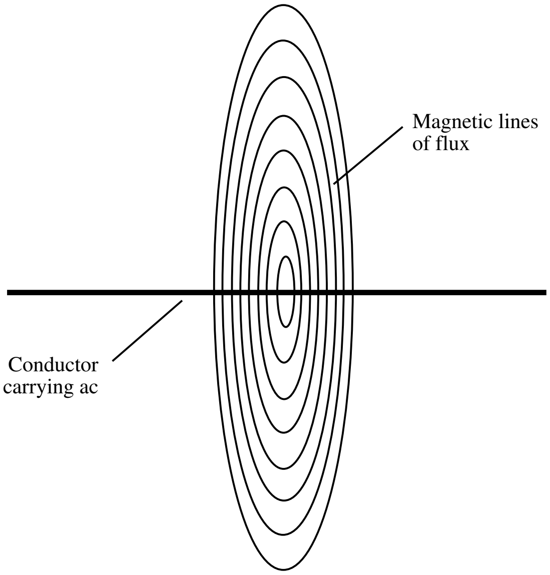

circular magnetic wave

circular mil A unit of cross-sectional area equivalent to 0.785 millionths of a square inch, or the area of a circle having a diameter of 0.001 inch. Generally, the circular mil is used to specify the cross-sectional area of a conductor, such as wire.

circular mil foot A unit of volume in which the length is 1 foot and the cross-sectional area is 1 circular mil.

circular polarization A form of electromagnetic-wave polarization in which the orientation of the electric flux rotates continuously and uniformly as the wave propagates through space. Circular polarization can occur in either a clockwise or counterclockwise sense.

circular radian The angle enclosed by two radii of a unit circle and subtended by a unit arc. Equal to about 57.296 angular degrees.

circular scan A radar scan in which the electron-beam spot describes a circle centered around the transmitting antenna.

circular sweep In an oscilloscope, a sweep obtained when the horizontal and vertical sinusoidal deflecting voltages have the same amplitude and frequency, but are out of phase by 90 degrees (1⁄4 cycle).

circular trace An oscilloscope pattern consisting of a circle obtained with a circular sweep of the electron beam.

circular waveguide A waveguide with a circular cross section.

circulating register In a digital computer, a register in which digits are taken from locations at one end and returned to those at the other end.

circulating tank current The alternating current that oscillates between the capacitor and inductor within a tank circuit.

circulator A multi-terminal coupler in which microwave energy is transmitted in a particular direction from one terminal to the next.

circumvention In a security or alarm system, the evasion of detection. Can be done by physically avoiding regions of coverage, or by defeating the system electronically.

cis A prefix meaning “on this side of.” For example, the *cislunar field* is the field on this side of the moon.

Citizen Band Abbreviation, CB. A band of radio frequencies allocated for two-way communication between private citizens (apart from amateur and commercial services).

Citizens Radio Service Two-way radio communication in a CITIZEN BAND. In the United States, the FCC licenses users of this service without requiring them to take an examination.

C/kg Abbreviation of *coulombs per kilogram*, the unit for electron charge-to-mass ratio.

C/kmol Abbreviation of *coulombs per kilomole*, the unit for the Faraday constant.

ckt Abbreviation of CIRCUIT.

Cl Symbol for CHLORINE.

cl Abbreviation of CENTILITER.

cladding The bonding of one metal to another to minimize or prevent corrosion. A common example is copper-clad steel wire, ideal for use in radio-frequency antenna systems. The copper provides excellent conduction, and the steel provides high tensile strength with a minimum of wire stretching.

clamper A device that restricts a wave to a predetermined dc level. Also called DC RESTORER.

clamping **1.** Fixing the operation of a device at a definite dc level. Also see CLAMPER. **2.** In television, establishing a fixed level for the picture signal at the start of each scanning line.

clamping circuit See CLAMPER.

clamping diode A diode used to fix the voltage level of a signal at a particular reference point.

clapper In a bell, the ball or hammer that strikes the bell; in an electric bell, it is affixed to the vibrating armature.

Clapp-Gouriet oscillator A *Colpitts oscillator* in which a capacitor is connected in series with the inductor. The circuit offers high frequency stability in the presence of input and output capacitance variations.

Clapp oscillator A series-tuned hybrid *Colpitts oscillator,* having a tuning capacitor in series with the inductor, rather than in parallel with the inductor. The circuit allows the use of a smaller tuning capacitor, resulting in improved stability.

Clark cell See ZINC STANDARD CELL.

class-A amplifier An amplifier whose bias is set at approximately the midpoint of the characteristic curve. Output electrode current flows during the

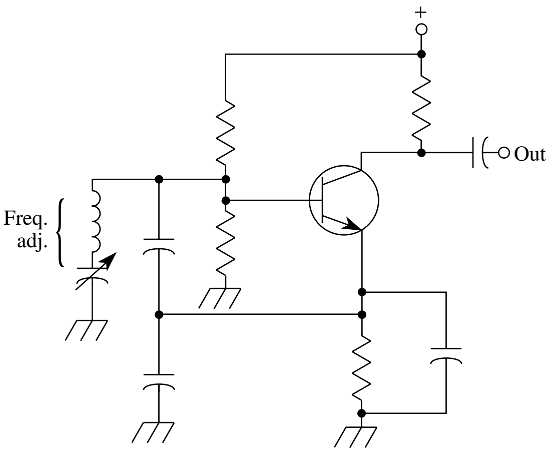

Clapp oscillator

complete ac driving-voltage cycle. The input signal never drives the device into the nonlinear portion of the characteristic curve.

class-AB amplifier Either a CLASS-AB$_1$ AMPLIFIER or a CLASS-AB$_2$ AMPLIFIER.

class-AB$_1$ amplifier An amplifier whose bias is adjusted to a level between that of a class-A amplifier and that of a class-AB$_2$ amplifier. Output electrode current flows during the entire ac driving-voltage cycle. The input signal drives the device into the nonlinear portion of the characteristic curve during part of the cycle.

class-AB$_2$ amplifier An amplifier whose bias is adjusted to a level between that of a class-AB$_1$ amplifier and that of a class-B amplifier. Output electrode current flows during more than 50 percent, but less than 100 percent, of the input signal cycle.

class-AB modulator A modulator whose output stage is a class-AB$_1$ or class-AB$_2$ amplifier.

class-A modulator A circuit for obtaining amplitude-modulated signals; essentially a class-A amplifier with two inputs, one for the carrier and the other for the modulating signal.

class-A operation The operation of a transistor, field-effect transistor, or vacuum tube, in which the collector, drain, or plate current flows during the entire signal cycle.

class-B amplifier An amplifier whose bias is adjusted to operate at the cutoff point in the characteristic curve. Output current flows during approximately 50 percent of the input signal cycle. Efficiency is higher than that of a class-A amplifier.

class-B modulator A push-pull modulator whose output stage is a class-B amplifier.

class-B operation The operation of a transistor, field-effect transistor, or vacuum tube, in which the collector, drain, or plate current flows for approximately half the signal cycle.

class-C amplifier An amplifier whose input-electrode bias is adjusted for operation at a point considerably beyond cutoff. Output current flows during less than half of the input signal cycle. Such an amplifier requires comparatively high driving power, but is capable of excellent efficiency. Commonly used in continuous-wave (CW), amplitude-modulated (AM), and frequency-modulated (FM) radio transmitters.

class-C operation The operation of a transistor, field-effect transistor, or vacuum tube, in which the collector, drain, or plate current flows for significantly less than half the signal cycle.

class-D telephone A telephone restricted to use by emergency services, such as fire departments and guard alarm installations.

classical electron radius Abbreviated r_e. The quantity expressed as $e^2/(m_e c^2)$, where e is the electron's charge in electrostatic units, m_e is its rest mass, and c is the speed of light. The value r_e is equal to approximately 2.82×10^{-13} cm or 2.82×10^{-15} m.

clean room A room for the assembly or testing of critical electronic equipment. The term is derived from the extraordinary steps taken to remove dust and other contaminating agents. The personnel wear carefully cleaned garments (or disposable clothing), gloves, caps, and masks; in some situations, they are required to walk between ceiling and floor ducts of a vacuum system upon entering the room.

cleanup process In the process of electron tube evacuation, a technique used to remove residual and occluded gases from the vacuum apparatus and from the device being evacuated.

clear **1.** In computer operations, to restore a switching element (e.g., a flip-flop) or a memory element to its standard (e.g., zero) state. **2.** In computer practice, an asynchronous input.

clearance The distance between two live terminals, or between one live terminal and ground.

clear band In optical character recognition, the part of a document that must remain unprinted.

clear channel **1.** A channel in the standard amplitude-modulation (AM) broadcast band that is designated to only one station within the area covered by the signal from that station. **2.** In television broadcasting, a channel for which there are no restrictions on the nature of the programming.

clear memory A function in a calculator or small computer that erases the contents of the memory.

clear raster The raster on the screen of a television picture tube in the absence of a signal, noise, or faulty beam deflection.

cleavage In a crystalline substance, the quality of splitting along definite planes. Also, a fragment resulting from such a cleft.

click filter See KEY-CLICK FILTER.

click method An emergency technique for rendering an electric current audibly detectable, by making and breaking the circuit carrying the current to a headset or earphone. A single click results from each make and each break. Also see TIKKER.

click suppressor See KEY-CLICK FILTER.

climate chamber A test chamber that provides accurately controlled temperature, humidity, and/or barometric pressure, for evaluating the performance of electronic components and circuits. Also called ENVIRONMENTAL TEST CHAMBER.

climatometer An instrument incorporating a hygrometer and bimetallic thermometer, whose dial pointers intersect to indicate comfort zones (best temperature-to-humidity ratio).

clinometer An electromechanical device that measures the steepness of a slope. When the device is level (horizontal), the output voltage is zero. If the device is tipped in one direction, a negative voltage is produced; if it is tipped in the other direction, a positive voltage is produced. The output voltage is proportional to the angle at which the device is tipped. Used in mobile robots.

clip A pinch-type connector whose jaws are normally held closed by a spring.

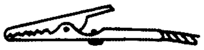

clip

clipped-noise modulation Modulation of a jamming signal through clipping action to increase the sideband energy and resulting interference.

clipper A circuit whose output voltage is fixed at a value for all input voltages higher than a predetermined value. Clippers can flat-top the positive, negative, or both positive and negative peaks of an input voltage.

clipper amplifier An amplifier operated so that the positive, negative, or both positive and negative peaks are clipped in the output signal. The clipping action results from feeding a regular symmetric waveform into an amplifier so that on negative excursion extremes, the stage is cut off; on positive excursion extremes, the amplifier is driven into saturation.

clipper limiter A device that delivers an output signal whose amplitude range corresponds to input-signal voltages between two predetermined limits. It can be used as a *noise limiter* with an element or elements that clip all pulses whose amplitudes are greater than the signal being processed.

clipping **1.** Leveling off (flat-topping) a signal peak at a predetermined level. Also see CLIPPER. **2.** In audio practice, the loss of syllables or words because of cutoff periods in the operation of the circuit (usually caused by overdriving a stage).

clock In a digital computer or controller, the device or circuit that supplies timing pulses to pace the operation of the system.

clocked flip-flops A master-slave arrangement of direct-coupled flip-flops. Information entered into the master unit when the input-trigger pulse amplitude is high is transferred to the slave unit when the amplitude is low.

clock frequency In a digital computer or control, the reciprocal of the period of a single cycle, expressed in terms of the number of cycles occurring in one second of time (hertz, kilohertz, or megahertz).

clock generator A test-signal generator that supplies a chain of pulses identical to those supplied by the clock of a digital computer.

clock module A complete plug-in or wire-in digital unit whose readout indicates time of day or elapsed time. Connected to a suitable power supply, it serves as either a clock or timer.

clock pulse A time-base pulse supplied by the clock of a digital computer, expressed as a period whose reciprocal is frequency.

clock rate See CLOCK FREQUENCY.

clock track On a magnetic tape or disk for data storage, a track containing read or write control (clock) pulses.

clockwise Abbreviation, cw. Rotation in a right-hand direction around a circle, starting at the top. Compare COUNTERCLOCKWISE.

clockwise-polarized wave An elliptically polarized electromagnetic wave whose electric-intensity vector rotates clockwise, as observed from the point of propagation. Compare COUNTERCLOCKWISE-POLARIZED WAVE.

clone A machine manufactured by a relatively unknown company that performs all the same functions, in basically the same way, as another machine manufactured by a well-known, major corporation. The term is used especially in reference to computers and computer peripherals. If a device is compatible with a certain computer, then clones of that device are generally compatible with that computer. Also, the device is likely to be compatible with all clones of the computer.

close coupling Also called *tight coupling.* In a transformer, the placement of the primary and secondary coils as close together as possible for maximum energy transfer. Compare LOOSE COUPLING.

closed capacitance The value of a variable capacitor whose rotor plates are completely meshed with the stator plates. Compare OPEN CAPACITANCE.

closed circuit A continuous unbroken circuit (i.e., one in which current can flow without interruption). Compare OPEN CIRCUIT.

closed-circuit cell A primary cell, such as the early gravity cell, designed for heavy and polarization-free service.

closed-circuit communication Communication between units only within a defined, hard-wired system, not extending to other units or systems.

closed-circuit security system An electronic security or alarm system, consisting of subsystems interconnected so that a disturbance anywhere in the circuit will result in an alarm signal pinpointing the location of the disturbance.

closed-circuit signaling Signaling accomplished by raising or lowering the level of a signaling current flowing continuously in a circuit.

closed-circuit television Abbreviation, CCTV. A usually in-plant television system, in which a transmitter feeds one or more receivers through a cable.

closed core A magnetic core generally constructed in an "O" or "D" configuration to confine the magnetic path to the core material. Compare OPEN CORE.

closed-core choke A choke coil wound on a CLOSED CORE. Also called CLOSED-CORE INDUCTOR.

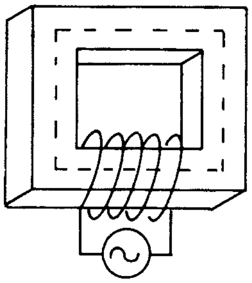

closed-core choke

closed-core transformer A transformer wound on a CLOSED CORE.

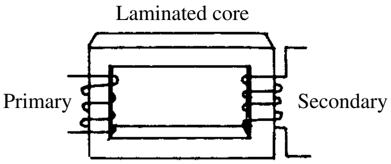

closed-core transformer

closed loop **1.** The feedback path in a self-regulating control system. An oscillator, for example, is a closed-loop amplifier. **2.** A loop within a program that would continue indefinitely, except for an external exit command.

closed-loop bandwidth The frequency at which the gain of a closed-loop circuit (see CLOSED LOOP, **1**) drops 3 decibels from the direct-current or midband value.

closed-loop control system A control system in which self regulation is obtained by means of a feedback path (see CLOSED LOOP). An example is a voltage regulator, in which a rise in output voltage is fed back to the input. This changes the input voltage and reduces the output voltage to its correct value. Compare OPEN-LOOP CONTROL SYSTEM.

closed-loop input impedance The input impedance of an amplifier that has feedback.

closed-loop output impedance The output impedance of an amplifier that has feedback.

closed-loop voltage gain The voltage gain of an amplifier that has feedback.

closed magnetic circuit A magnetic circuit in which the flux is uninterrupted, as in a ferromagnetic core, which has no air gap. Also see CLOSED CORE.

closed subroutine In a digital computer program, a subroutine that can be accessed and left by branch instructions, such as GOSUB and RETURN in the high-level language BASIC.

close-spaced array A beam antenna in which the elements (radiator, director, and reflector) are spaced less than a quarter-wavelength apart.

close-talk microphone A microphone that must be placed close to the mouth. Such a microphone is less susceptible to background noises than an ordinary microphone, and is useful in environments where the ambient noise level is high.

closing rating A specification for closure conditions in a relay, including duty cycle and contact life (total guaranteed closures before contact failure).

closure **1.** The act of closing or being closed (e.g., *switch closure* or *relay closure*). **2.** Circuit completion (i.e., the elimination of all discontinuities).

cloud The mass of electrons constituting the space charge in a vacuum tube.

cloverleaf antenna An omnidirectional transmitting antenna in which numerous horizontal, four-element radiators (stacked vertically, a quarter-wavelength apart) are arranged in the shape of a four-leaf clover.

C/L ratio See LC RATIO.

clutter Extraneous echoes that interfere with the image on a radar display.

clutter gating In radar operations, a switching process that causes the normal video to be displayed in regions free of clutter, and the video indicating target movement to be displayed only in cluttered areas.

Cm Symbol for CURIUM.

cm Abbreviation of CENTIMETER.

c.m. Abbreviation of CIRCULAR MIL.

cm^2 Abbreviation of *square centimeter.*

cm^3 Abbreviation of *cubic centimeter.*

C_{max} Abbreviation of *maximum capacitance.*

C meter See CAPACITANCE METER.

C_{min} Abbreviation of *minimum capacitance.*

CML Abbreviation of CURRENT-MODE LOGIC.

CMOS Abbreviation of COMPLEMENTARY METAL-OXIDE SEMICONDUCTOR.

CMR See COMMON-MODE REJECTION.

CMRR See COMMON-MODE REJECTION RATIO.

CMV See COMMON-MODE VOLTAGE.

C network A circuit with three impedances connected in series, the free leads being connected to a pair of terminals and the two internal junctions, to another pair of terminals.

Co Symbol for COBALT.

C$_o$ Symbol for OUTPUT CAPACITANCE.

coalesce In computer operations, to create one file from several.

coarse adjustment Adjustment of a quantity in large increments. Compare FINE ADJUSTMENT.

coarse-chrominance primary See Q SIGNAL.

coastal bending A change in the horizontal direction of a line-of-sight radio wave when it crosses a coastline.

coast station In the Maritime Mobile Radio Service, a land station that communicates with shipboard stations.

coating **1.** The application of a substance to another substance by means of electroplating, electrophoresis, or similar process, for the purpose of protecting the material, isolating it from the environment, or improving the conductivity of an electrical connection to some other object. **2.** The magnetic material on a recording tape. **3.** In a computer system, the magnetic material on a magnetic diskette or hard disk.

coating thickness On magnetic tape or magnetic disks, the depth of the magnetic coating applied to the base.

coax Abbreviation of COAXIAL CABLE or COAXIAL LINE.

coaxial antenna A half-wave vertical antenna that is center-fed by coaxial cable. The cable runs upward through a 1⁄4-wave section of tubing that composes the lower half of the antenna. The outer conductor of the cable is connected to this tubing through a shorting disk at the top. The inner conductor of the cable is connected to a 1⁄4-wave vertical radiator that is insulated from, and that extends upward from the top of, the lower section.

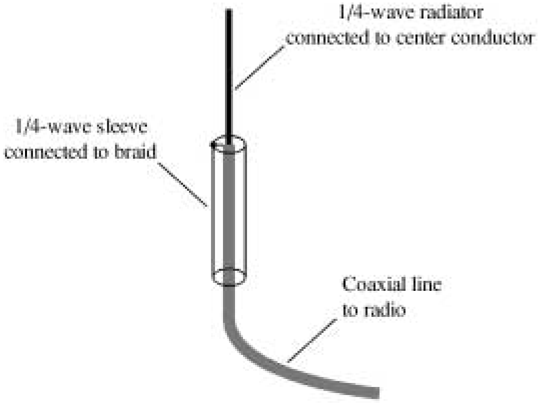

coaxial antenna

coaxial cable An unbalanced cable consisting of two concentric conductors: an inner wire and an outer, braided sleeve. The inner and outer conductors are separated by a dielectric, usually solid or foamed polyethylene. The outer conductor is generally grounded while the inner conductor carries the signals. This cable is used in community-antenna television (CATV) networks, and as a transmission line connecting antennas to radio transmitters, receivers, and transceivers at low, medium, high, and very-high frequencies. It is also used in some high-fidelity sound systems—especially to connect microphones, compact-disc players, tape players, tuners, and speakers to audio amplifiers.

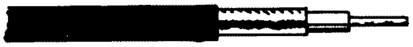

coaxial cable
(From left to right: insulating jacket, woven outer conductor, low-loss insulating sleeve, inner conductor.)

Characteristics of prefabricated coaxial transmission lines.

Type	*Characteristic impedance, (ohms)*	*Velocity factor*	*Outside dia. (in.)*	*Picofarads per foot*
RG-8/U	52	0.66	0.41	29.5
RG-9/U	51	0.66	0.42	30.0
RG-11/U	75	0.66	0.41	20.6
RG-17/U	52	0.66	0.87	29.5
RG-58/U	54	0.66	0.20	28.5
RG-59/U	73	0.66	0.24	21.0
RG-174/U	50	0.66	0.10	30.8
hard line	50	0.81	0.50	25.0
(½-inch)	75	0.81	0.50	16.7
hard line	50	0.81	0.75	25.0
(¾-inch)	75	0.81	0.75	16.7

coaxial capacitor **1.** A somewhat uncommon, but highly effective, capacitor that uses two telescoping sections of tubing. It works because there is a certain effective surface area between the inner and the outer tubing sections. A sleeve of plastic dielectric is placed between the sections of tubing. This allows the capacitance to be adjusted by sliding the inner section in or out of the outer section. Coaxial capacitors are especially useful in antenna systems for tuning and/or impedance matching. Their values are generally from a few picofarads up to about 100 pF. **2.** A short length of coaxial cable that is used as a capacitor rather than a transmission line because of the inherent capacitance between its center conductor and braid. See COAXIAL CABLE.

coaxial cavity A cavity consisting of a cylindrical metal chamber housing a central rod. The cavity can be tuned to resonance by means of a piston.

coaxial connector A device used to splice coaxial line or to connect a coaxial line to a transmitter, receiver, or other piece of apparatus.

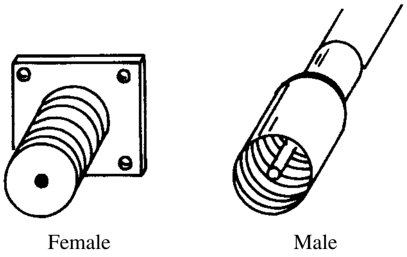

coaxial connector

coaxial diode A semiconductor diode housed in a cylindrical metal shell acting as one contact, and provided with a recessed, concentrically mounted end pin, which serves as the other contact.

coaxial driver See COAXIAL SPEAKER.

coaxial filter **1.** A filter that uses a coaxial cable as a tuned circuit. **2.** A filter designed to be used in a coaxial transmission line.

coaxial jack A female receptacle or connector, whose concentric terminals have the same spacing as a male coaxial-cable connector designed to fit it.

coaxial line A signal transmission line consisting of COAXIAL CABLE.

coaxial-line frequency meter A microwave absorption wavemeter (see WAVEMETER) with input and output receptacles for insertion into a coaxial line.

coaxial-line oscillator See CONCENTRIC-LINE OSCILLATOR.

coaxial loudspeaker See COAXIAL SPEAKER.

coaxial plug A male connector whose concentric terminals have the same spacing as a female coaxial cable connector designed to fit it.

coaxial receptacle A coaxial connector, such as a coaxial jack or plug. Receptacles are installed in equipment, whereas plugs are usually attached to the end of coaxial cables.

coaxial relay A relay designed to connect and disconnect, or to interchange, coaxial cables in a transmission line without disturbing the characteristic impedance of the line.

coaxial speaker Also called *coaxial driver* and *coaxial loudspeaker.* A large low-frequency speaker and a small high-frequency speaker mounted concentrically, the smaller within the larger. When used with a crossover network, this arrangement provides fairly good wide-range audio-frequency response, and saves physical space, compared with the use of separate speakers.

coaxial stub **1.** A length of coaxial cable acting as a branch to another coaxial cable. Commonly used for impedance matching. **2.** A length of coaxial cable, usually cut to 1/4 or 1/2 wavelength, and connected across a coaxial transmission line to act as a WAVETRAP. Commonly used to reject strong interfering signals.

coaxial switch A switch designed to connect and disconnect, or to interchange, coaxial cables in a transmission line without disturbing the characteristic impedance of the line.

coaxial tank A tank circuit consisting of a rod within a cylinder. The tank is usually tuned by a small variable capacitor connected between the rod and cylinder at one end of the combination. Generally used at ultra-high frequencies (UHF).

coaxial-tank oscillator A stable, self-excited oscillator that uses a COAXIAL TANK. Also see CONCENTRIC-LINE OSCILLATOR.

coaxial transistor A transistor in which a semiconductor wafer is mounted centrally in a metal cylinder (the base connection) and is contacted on opposite faces by the emitter and collector whiskers, which are axially mounted.

coaxial transmission line A transmission line that is a COAXIAL CABLE.

coaxial wavemeter A type of absorption wavemeter in which the tunable element is a section of coaxial line (i.e., a metal cylinder surrounding a metal rod). An internal short-circuiting disk is moved along the cylinder to connect its inner wall to selected points along the rod's length, thereby varying the resonant frequency. The instrument is useful at microwave frequencies.

cobalt Symbol, Co. A metallic element. Atomic number, 27. Atomic weight, 58.94.

cochannel interference Interference between similar signals transmitted on the same channel.

Cockcroft-Walton accelerator A proton accelerator in which nuclei of hydrogen atoms are given high velocity through a straight tube by a high dc voltage.

codan Any of several muting (SQUELCH) systems. In particular, a squelch circuit that suppresses noise in a sensitive receiver equipped with automatic gain control (AGC). The receiver is quiet until a carrier of predetermined strength is received. The name is an acronym for *carrier-operated device antinoise.*

codan lamp A lamp that alerts a radio operator that a signal of satisfactory strength is being received. Also see CODAN.

code **1.** A set of symbols for communications (e.g., the Morse code of radiotelegraphy and wire telegraphy in which dots and dashes correspond to letters, numbers, and marks of punctuation). **2.** In a computer program, symbolically represented instructions. **3.** ENCODE.

codec In encoding and decoding equipment, a *coder/decoder*, usually in a single package and operating at 8 kHz for an input signal with a passband of 3100 Hz (300 to 3400 Hz).

code character **1.** The representation of character in a particular code form. **2.** A sequence of dots and dashes in the Morse code.

code conversion The translation of a coded signal from one form of code to another.

coded decimal digit A number expressed in binary form (computer code), that is, in terms of zeros and ones only.

code-directing characters Characters added to a message to indicate how and where it is going.

coded program See PROGRAM.

coded signal **1.** A wire- or radiotelegraph signal in which secrecy is achieved by using letters in cipher groups, instead of straight language. **2.** SCRAMBLED SIGNAL.

coded stop See PROGRAMMED HALT.

code elements The smallest identifiable parts that compose a digital code. For example, in computer code, the elements are ones and zeroes (*high* and *low* logic states); in Morse code, they are *dots* and *dashes.*

code holes In a punched card or tape, holes representing data.

code line A written computer program instruction.

code machine Any one of several devices for recording or reproducing code signals.

code position The part of a data medium (e.g., card row) reserved for data.

code-practice oscillator A simple keyed audio oscillator intended for practicing Morse code.

coder **1.** In computer operations, a person who prepares instructions from flow charts and procedures devised by a programmer. **2.** A device that delivers coded signals.

code receiver A radiotelegraph receiver.

code ringing A method of ringing a telephone subscriber in a predetermined manner to convey a certain message.

code segment The instruction part of computer storage associated with a process. Compare DATA SEGMENT and DUMP SEGMENT.

code set The collection of codes representing all of the characters in a language.

code speed See KEYING SPEED.

code transmitter **1.** A radiotelegraph transmitter. **2.** A tape-operated keyer for wire telegraphy or radiotelegraphy.

code word See PHONETIC ALPHABET CODE WORD.

coding **1.** Performing the service of a CODER. **2.** Writing instructions for a digital computer; a part of programming.

coding check A pencil-and-paper verification of a routine's validity.

coding sheet A form on which program instructions are written prior to input.

codiphase radar A radar system that uses beam forming, signal processing, and a phased-array antenna.

codistor A voltage-regulating semiconductor device.

coefficient **1.** A factor in an indicated product. Thus, in 4y, 4 is the coefficient of y. **2.** A parameter that indicates a specific characteristic of some component or device (e.g., COEFFICIENT OF COUPLING or COEFFICIENT OF REFLECTION).

coefficient of coupling Symbol k. The ratio of MUTUAL INDUCTANCE between two inductors to the maximum possible (theoretical) value of mutual inductance. This ratio is always greater than or equal to 0 (no coupling between inductors), and less than or equal to 1 (perfect coupling between inductors).

coefficient of current detection See CURRENT-DETECTION COEFFICIENT.

coefficient of reflection A measure of the amount of electromagnetic field reflected in a transmission line from the load feed point. The coefficient of reflection is equal to the square root of the reflected power divided by the forward power.

coercive force The demagnetizing force required to remove residual magnetism from a material.

coercivity See COERCIVE FORCE.

cogging Nonuniform rotation of a motor armature. The velocity increases as an armature coil enters the magnetic field and decreases as it leaves the field.

coherence In electromagnetic radiation, a condition in which all the wavefronts are in phase. This results in high energy concentration, and makes possible the long-distance transmission of infrared, visible light, and ultraviolet, because the rays are almost perfectly parallel. It also makes possible the extreme radiation intensity characteristic of some LASER devices.

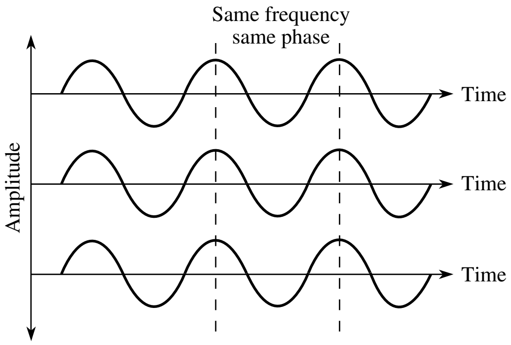

coherence

coherent bundle A bundle of optical fibers, such that the individual fibers are in the same relative positions at either end of the bundle.

coherent carrier A carrier that agrees in frequency and phase with a reference signal.

coherent electroluminescent device See LASER DIODE.

coherent light Visible light in which the phase relationship between successive waves is such that the beam consists of parallel rays that provide a high concentration of energy. Also see LASER.

coherent-light radar See COLIDAR.

coherent oscillator In a radar system, an oscillator that provides a COHERENT REFERENCE.

coherent-pulse operation Pulse operation characterized by a fixed phase relationship between pulses.

coherent radiation Radiation characterized by COHERENCE.

coherent reference A stable reference frequency with which other signals are phase locked for coherence.

coherent transponder A transponder in which the frequency and phase of the input and output signals have a fixed relationship.

coil A long conductor or group of conductors wound into a tight helical package, often in several layers on a cylindrical form. This takes advantage of the resulting concentration of magnetic flux, maximizing the inductance that can be obtained in a component of limited physical size. Further increases in inductance can be realized by the use of ferromagnetic core materials. See also INDUCTOR.

coil antenna See LOOP ANTENNA.

coil checker An alternating-current (ac) meter or simple bridge for checking inductors. Such instruments usually only indicate inductance values, but some list readings of resistance or approximate inductor Q factor.

coil dissipation The power wasted in a coil as heat. Generally, this dissipation or loss is proportional to the resistance of the coil, and to the square of the current passing through the coil.

coil form The insulating support around which an air-core coil is wound.

coil loading The insertion of one or more inductors into a transmission line or antenna element, for the purpose of impedance matching, alteration of the resonant frequency, or both.

coil magnification factor The Q factor of an inductor. Generally given by the ratio X_L/R_L, where X_L is the inductive reactance of the coil in ohms, and R_L is the resistance of the coil in ohms.

coil neutralization See INDUCTIVE NEUTRALIZATION.

coil resistance The resistance of a coil (inductor), as distinct from its reactance. It is almost entirely the result of ohmic loss in the wire from which the coil is manufactured.

coilshield A metal can designed to provide efficient electrostatic and electromagnetic shielding of a coil, preventing unwanted inductive coupling to other components.

coincidence The simultaneous occurrence of two or more signals. Compare ANTICOINCIDENCE.

coincidence amplifier An amplifier that delivers an output signal only when two or more input signals occur simultaneously.

coincidence circuit See AND CIRCUIT.

coincidence counter A circuit or device, such as a gate, that delivers an output pulse only when two or more input pulses occur simultaneously; the output pulses go to a device that counts them.

coincidence detector See AND CIRCUIT.

coincidence gate See AND GATE.

coincident-current selection Selection of a magnetic core (in a core memory or similar device) by applying two or more currents simultaneously.

coin shooting Searching for coins and similar small, buried metallic objects using a METAL LOCATOR.

coke A porous material obtained from the destructive distillation of coal. It is valued for the production of carbon components for electronics, such as dry-cell electrodes and motor brushes.

cold **1.** Pertaining to an electrical circuit, component, or terminal that is at ground potential. **2.** A term denoting a bad solder joint. **3.** Pertaining to an unheated electrode or element. See COLD CATHODE.

cold alignment The alignment of a tracking system (especially of its tuned circuits) when the system is not in operation, as when transistor power is off. Also called QUIET ALIGNMENT.

cold cathode **1.** In an electron tube, a cathode that emits electrons without being heated. **2.** A cathode electrode operated at a temperature below ambient temperature.

cold chamber An enclosure in which electronic equipment can be tested at selected, precise low temperatures. Compare OVEN.

cold flow The (usually gradual) change in the dimensions of a material, such as plastic in a molded part.

cold junction In a thermocouple system, an auxiliary thermocouple connected in series with the hot thermocouple, and immersed in ice or operated at ambient temperature.

cold light Light produced without significant heat, as from the ionization of a gas by a high voltage (as in neon bulbs and fluorescent lamps), or by electroluminescence, bioluminescence, cathodoluminescence, or a similar phenomenon.

cold pressure welding Welding sometimes used in the fabrication of electronic equipment, in which the metal parts to be joined are pressed together tightly to the point of deformation, whereupon they become welded.

cold resistance The resistance of an unheated electronic component. Compare HOT RESISTANCE.

cold rolling A method of manufacturing an inductor core so that the magnetic grains are all arranged lengthwise.

cold solder joint A solder joint in which insufficient heat has been applied, resulting in a bad connection.

cold spot **1.** An area of a circuit or component whose temperature is ordinarily lower than that of the surrounding area. **2.** A node of current or voltage. Compare HOT SPOT.

cold weld A welded joint produced by means of COLD PRESSURE WELDING.

colidar An optical radar system using unmodulated, coherent (laser-produced) light. The term is an acronym for *coherent light detection and ranging.*

collate In data processing, to produce an ordered set from two or more similarly ordered sets (as punched cards).

collator In a punched-card system, a device that collates (see COLLATE) punched cards.

collector **1.** In a bipolar transistor, the electrode toward which emitted current carriers travel. **2.** In a Klystron, the final electrode toward which electrons migrate after passing through the buncher and catcher. **3.** In an iconoscope, a cylindrical electrode around the circumference of the tube, which gathers and conducts away the electrons released by the mosaic. **4.** The final (target) electrode in a backward-wave or traveling-wave tube. **5.** A computer program segment that collates compiled segments so that they can be loaded into the computer.

collector capacitance **1.** Symbol, C_C. The capacitance of the collector junction in a bipolar transistor. **2.** The capacitance of the collector electrode in a Klystron, iconoscope, backward-wave tube, or traveling-wave tube.

collector current **1.** Symbol, I_C. The current flowing in the collector circuit of a bipolar transistor. Also see AC COLLECTOR CURRENT and DC COLLECTOR CURRENT. **2.** Current flowing in the collector circuit of a Klystron, iconoscope, backward-wave tube, or traveling-wave tube.

collector-current cutoff See COLLECTOR CUTOFF.

collector cutoff In a bipolar transistor, the condition in which the collector current is cut off (i.e., reduced to the residual value). Also see CUTOFF CURRENT.

collector cutoff current See CUTOFF CURRENT.

collector-diffusion isolation A method of making integrated circuits that contain bipolar transistors. Provides electrical separation of the transistors in a semiconductor integrated circuit.

collector dissipation Symbol, P_C. In a bipolar transistor, the power dissipation of the collector electrode. The collector dc power dissipation is the product of collector current and collector voltage: $P_C = V_C I_C$.

collector efficiency In a bipolar transistor circuit, the ratio of ac power output to dc collector-power input.

collector family For a bipolar transistor, a group of collector current versus collector voltage curves. Each is plotted for a particular value of base current (common-emitter circuit) or emitter current (common-base circuit).

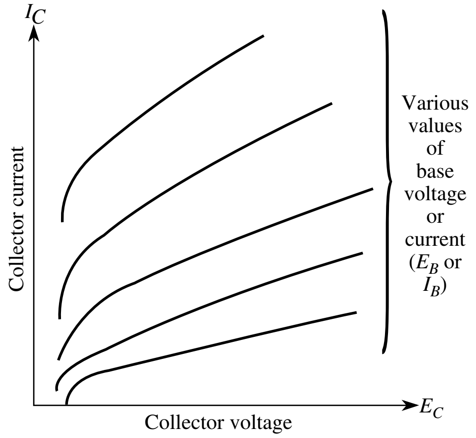

collector family

collector junction In a bipolar transistor, the junction between collector and base layers.

collector mesh In a cathode-ray storage tube, a flat, fine wire screen that attracts and conducts away the secondary electrons knocked out of the storage mesh by the electron beam.

collector multiplication In a bipolar transistor, an increase in the number of electrons at the collector electrode, caused by a momentary alteration of the charge density of the collector junction by injected carriers reaching the junction.

collector resistance In a bipolar transistor, the internal resistance of the collector junction. See AC COLLECTOR RESISTANCE and DC COLLECTOR RESISTANCE.

collector ring **1.** A rotating, brush-contacted ring electrode connected to one end of a coil in an ac generator. **2.** A similar ring which, with a brush, serves as a connection to a rotating element, as in a signal-gathering system. **3.** The collector electrode in an iconoscope.

collector transition capacitance The capacitance between the collector and base of a bipolar transistor under normal operating conditions. This capacitance has a limiting effect on the operating frequency of a bipolar device.

collector voltage Symbol, V_C. In a bipolar transistor, the voltage on the collector electrode. See AC

COLLECTOR VOLTAGE and DC COLLECTOR VOLTAGE.

collimated rays Electromagnetic waves made parallel or nearly parallel. This can be done by means of a reflector, a lens, or a laser.

collimation **1.** The process of rendering electromagnetic rays parallel. **2.** Adjustment of the line of sight of an instrument, such as a level or transit.

collimation equipment Optical-alignment equipment.

collimator A device for producing parallel rays of light or other radiation. In electronics, the most common example is a *dish antenna.*

collinear antenna A broadside directional antenna consisting of two or more half-wave radiators; the current is kept in phase in each section by quarter-wave stubs between each radiating section. The radiators are stacked end to end horizontally or vertically. Also called FRANKLIN ANTENNA.

Collins coupler A single-section, pi-filter circuit, used to match a radio transmitter to a wide range of antenna impedances. Also called *pi coupler* and *Collins network.*

collodion A viscous solution of pyroxylin and a solvent (such as acetone, alcohol, or ether) sometimes used as a binding agent for coils and other components.

cologarithm Abbreviation, colog. The logarithm of the reciprocal of a number; colog $x = \log (1/x) = \log x^{-1} = -\log x$.

color A perceived characteristic, and a direct function, of visible-light wavelength. Seen by the human eye as a spectrum of hues, ranging from red at the longest visible wavelengths, through orange, yellow, green, blue, indigo, and finally violet at the shortest visible wavelengths. See HUE.

coloration In audio applications, a blending of sounds as a result of mixing among components at different frequencies. Sometimes this is done deliberately; in other instances, it is undesirable.

color balance In a color television receiver, adjustment of the beam intensities of the individual guns of a three-gun picture tube. Compensates for the difference in light emissivity of the red, green, and blue phosphors on the tube screen.

color bar-dot generator A radio-frequency (RF) signal generator that produces a bar or dot pattern on the screen of a color television picture tube. Used for testing and alignment.

color-bar pattern A color television test pattern of vertical bars—each of a different color.

color breakup A transient separation of a color television picture into its red, green, and blue components, as a result of a sudden disturbance of viewing conditions (blinking of eyes, moving of head, intermittent blocking of screen, etc.).

color burst As a phase reference for the 3.579545-MHz oscillator in a color television receiver, approximately nine cycles of the chrominance subcarrier added to the back porch of the horizontal blanking pedestal in the composite color signal.

color carrier See CHROMINANCE SUBCARRIER.

colorcast A color television broadcast.

color code **1.** A system that uses colored stripes or dots to mark the nominal values and other characteristics on capacitors, resistors, and other components. **2.** A code that represents the various frequencies being used by radio-control modelers in competition, and used on flags attached to transmitters, for example, as a safeguard against jamming.

color coder See COLOR ENCODER.

color contamination In a color television system, faulty color reproduction resulting from incomplete separation of the red, green, and blue channels.

color-coordinate transformation In a color television system, the computation (performed electrically in the system) of the tristimulus (primary) values with reference to one set of primaries, from the same colors derived from another set of primaries.

color depth An expression for the extent to which an image can accurately render color. Generally expressed in bits or in number of colors. Some systems can reproduce millions of different colors.

color-difference signal Designated B-Y, G-Y, and R-Y. The signal resulting from reducing the amplitude of a color signal by an amount equal to the luminance-signal amplitude. Also see B-Y SIGNAL, G-Y SIGNAL, and R-Y SIGNAL.

color dot **1.** A phosphor spot on the screen of a color television picture tube. **2.** One of the spots stamped on a capacitor, indicating the capacitance, voltage, and tolerance (see COLOR CODE, **1**). **3.** A spot stamped on a resistor, indicating the number of zeros to be added to the value indicated by the color bands.

color edging In a color television picture, an aberration consisting of false color at the boundaries between areas of different color.

color encoder In a color television transmitter, the circuit or channel in which the camera signals and the chrominance subcarrier are combined into the color-picture signal.

color equation A mathematical means of determining the resultant color obtained by adding primary colors in various proportions.

color fidelity The faithfulness with which a color television system, lens, or film reproduces the colors of a scene.

color filter A transparent plate or film that transmits light of a desired color, and eliminates or attenuates all other colors.

color flicker In a color television system, image instability that occurs when the luminance and chromaticity both fluctuate.

color fringing In a color television picture, false color around objects, sometimes causing them to appear separated into different colors.

color generator A special radio-frequency (RF) signal generator to adjust or troubleshoot a color television receiver. The color signals it delivers are identical to those produced by a broadcast station.

color graphics Computer graphics displayed in color on a cathode-ray tube (CRT) or liquid-crystal display (LCD).

colorimeter A device used to quantitatively measure the color intensity of a sample relative to a standard.

colorimetric A characteristic of visible light, representing the wavelength concentration. Refers to the perceived color of a light beam.

colorimetry The science and art of color measurement.

color killer In a color television receiver, a circuit that, in the absence of a color signal, delivers a negative bias to cut off the bandpass amplifier.

color match In photometry, the condition in which color agreement exists between the halves of an area. Also see COLOR MATCHING.

color matching The art of selecting colors that are identical in hue, saturation, and brilliance. This can be done with the unaided eye or with the help of an instrument.

color media Substances that transmit essentially one color of visible light while blocking other colors.

color meter A photoelectric instrument for measuring color values, and comparing and matching colors.

color mixture An additive combination of two or more colors. Thus, red + yellow = orange, blue + red = violet, red + blue + green = white, etc.

color oscillator The oscillator in a color television receiver that coordinates the color response. This oscillator is operated at 3.579545 MHz, to within plus or minus 10 Hz.

color palette In a color video image, the total number of possible colors that can be displayed.

color phase In color television, the phase difference between an I or Q chrominance primary signal and the chrominance carrier reference.

color-phase diagram In color television, a quadrant diagram showing (for each of the three primary and complementary colors) the difference in phase between the color-burst signal and the chrominance signal, as well as the peak amplitude of the chrominance signal. Also shown are the peak amplitude and polarity of both in-phase and quadrature components required for the chrominance signals. For color TV receiver adjustment, the color-phase diagram is displayed, in effect, by a VECTORSCOPE when a suitable signal from a color generator is applied to the receiver.

color picture signal **1.** In color television and/or computer graphics, an electrical signal containing components corresponding to the hue, saturation, and brilliance of a fixed or changing visual image. **2.** In color television, the combination of chrominance and luminance signals minus blanking and sync signals.

color picture tube A specialized type of cathode-ray tube (CRT), used in color television receivers and computer displays. Three different images are produced: one in red, one in blue, and one in green. The three monochrome images are combined to form a complete color image.

color primaries **1.** Also called *additive primaries* or *primary colors.* In color television, the hues red (R), green (G), and blue (B). When these colors are mixed in various ratios, any visual color can result. **2.** Also called *subtractive primaries* or *primary pigments.* In color printing, the hues magenta (M), cyan (C), and yellow (Y). These roughly correspond to red (R), blue (B) and yellow (Y). Sometimes black (K) is also included. When these pigments are mixed in various ratios, any visual pigment can result.

color purity The ratio of wanted to unwanted components in a color. In a pure color, there are no components other than those required to produce the color. Color, in this context, includes white, black, and all shades of gray.

color-purity magnet A permanent magnet on the neck of a color television picture tube, used to help ensure color purity by maintaining proper displacement of the electron beam.

color registration In color television reception, the precise superimposition of red, green, and blue so that the composite is free from COLOR EDGING.

color rendering index A mathematical expression defining the effect of the color of a light source on an object. For example, in red light, a blue object appears nearly black.

color sampling rate The number of times per second that each primary color is sampled in a color television receiver.

color saturation A measure of the purity of a hue. The extent to which a hue is without a white component; 100% saturation indicates a complete absence of white.

color sensing In machine vision systems, the ability to distinguish between light of different wavelengths. Usually done with red, green and blue color filters and three separate cameras.

color sensitivity **1.** The degree of which a photosensitive device, such as a photocell or camera tube, responds to various colors of light. **2.** The degree to which photographic film responds to various colors of light.

color signal See COLOR PICTURE SIGNAL.

color spectrum The band of electromagnetic energy containing visible light; it extends from red (at the longest wavelengths) to violet (at the shortest). Commonly measured in *nanometers (nm),* where 1 nm = 10^{-9} m. Also expressed in *Angstroms*, where 1 Angstrom = 10^{-10} m = 0.1 nm. In order of decreasing wavelength, the colors are red at 750 to 700 nm (7500 to 7000

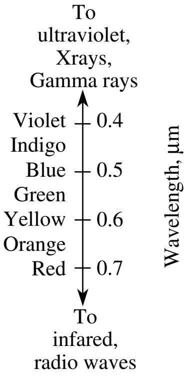

color spectrum

Angstroms), orange, yellow, green, blue, indigo, and violet at 410 to 390 nm (4100 to 3900 Angstroms).

color subcarrier A modulated monochrome signal whose sidebands convey color information.

color-sync signal See COLOR BURST.

color system Also called *RGB color model.* A means of representing a color in terms of mathematical coordinates. This can be done in three dimensions because there are three COLOR PRIMARIES. Each color primary is represented by an axis. Any COMPOSITE COLOR can be represented by a unique vector. The relative amount of each color primary is given by the length of the composite-color vector components along each axis.

color television Television in which the picture approximates natural color. It operates on the basis of mixing three primary colors (red, blue, and green) of phosphors on the picture tube screen.

color television receiver A television receiver designed to reproduce color pictures.

color television signal The signal transmitted by a color television transmitter, containing all of the information needed to reproduce a complete, full-color, moving image.

color transmission The television transmission of a picture in color.

color triad On the screen of a color picture tube, one of the color cells, each of which contains one of the three phosphor dots: red, green, and blue.

color triangle A triangle that can be inscribed on a chromaticity diagram to reveal the chromaticity range resulting from adding the three color primaries.

color TV signal The complete signal (video, color, and sync components) required for transmitting a picture in color.

color weather radar A computer-enhanced radar rendition of weather patterns, usually showing various intensities of precipitation as different colors. Commonly, areas of precipitation show up as violet, blue, green, yellow, orange, and red, in order of increasing intensity.

Colpitts oscillator A radio-frequency (RF) oscillator that uses a single, untapped inductor. A combination of two fixed capacitors in series is connected in parallel with the inductor. The feedback is controlled by the ratio of capacitances. A permeability-tuned coil or a roller inductor can be used to obtain variable-frequency operation. Stability is enhanced when the output of the oscillator is taken from the emitter or source portion of the circuit. To prevent the output signal from being short-circuited to ground, an RF choke is connected in series with the emitter or source. Compare HARTLEY OSCILLATOR.

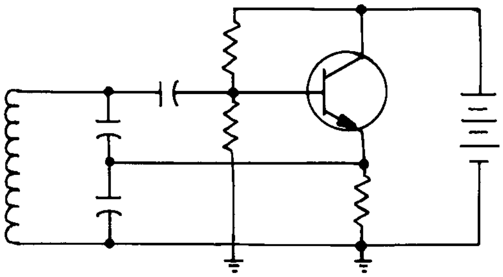

Colpitts oscillator

columbium Symbol, Cb. The former name of the metallic element *niobium.* Atomic number, 41. Atomic weight, 92.9064.

column See CARD COLUMN.

columnar graph A graphical presentation of data, in which the ordinates are represented by vertical columns whose height depends on the value. Commonly used in *presentation graphics*, but less common in *analytical graphics.*

column binary Binary number representation on punched cards, wherein consecutive digits correspond to consecutive column punching positions.

column speaker An acoustic speaker with a long cabinet, so that a large column of air is used for resonating or reinforcing purposes. This type of speaker radiates over a wide azimuth angle, while providing a narrow beam in the elevation plane.

column split On a punched card machine, the device for reading, as two separate characters or codes, two parts of a single column.

COM **1.** Abbreviation for *communications port.* **2.** Abbreviation for *computer output on microfilm.*

coma An aberration that causes the beam spot on the screen of a cathode-ray tube to resemble a comet.

coma lobes An aberration in the radiation or response pattern of a dish antenna that occurs when the radiating element is not exactly at the focal point of the reflector. When the directional pattern is altered by moving the driven element, rather than turning the entire antenna, these lobes appear.

comb amplifier An arrangement of several sharply tuned bandpass amplifiers whose inputs are connected in parallel and whose outputs are separate; the amplifiers separate various frequencies from a multifrequency input signal. The name is derived from the comb-like appearance of the response pattern of various output peaks displayed along a frequency-base axis.

comb filter A selective device that passes several narrow bands of frequencies within a larger band, while rejecting frequencies in between the narrow bands. So called because its frequency-response curve resembles the teeth of a comb when observed on a spectrum analyzer. Also see COMB AMPLIFIER.

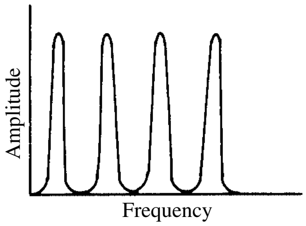

comb filter response

comb generator **1.** A signal generator that provides outputs at evenly spaced frequencies. So called because, on a spectrum analyzer, its output looks like the teeth of a comb. **2.** A transmitter with many spurious signals at its output.

combination **1.** A functional, usually stationary, installation consisting of two or more pieces of equipment. Examples: transmitter/receiver combination, motor/generator combination, and tuner/amplifier combination. **2.** In mathematics, a selection of several factors from a group, without regard to order. Thus, from the group ABC, the three possible combinations are AB, AC, and BC. Compare PERMUTATION.

combinational circuit Two or more basic logic circuits, combined in such a way that the output state depends entirely on the input states.

combination bridge A bridge that affords two or more classes of measurement, usually selectable by means of a function switch. Examples: capacitance-inductance bridge, and capacitance-resistance bridge.

combination cable A cable that has conductors grouped in pairs, threes, quads, or similar arrangements.

combination feedback See CURRENT-VOLTAGE FEEDBACK.

combination microphone Two or more microphones combined into one unit.

combination speaker Two or more loudspeakers combined into one (e.g., a COAXIAL SPEAKER).

combination tone An acoustic tone resulting from the combination of two other acoustic tones. If the original tones have frequencies f_1 and f_2 (where f_1 is higher than f_2), then the first-order combination frequencies are $f_1 + f_2$ and $f_1 - f_2$. Higher-order combination tones can result from mixing among the original tones and the first-order combination tones.

combinatorial logic A form of digital logic, in which the output states depend on the input states, but on no other factor.

combined head See READ-WRITE HEAD.

combined reactance The net reactance (X) in a circuit, obtained by vectorially adding the inductive reactance (X_L) and the capacitive reactance (X_C).

combiner A circuit or device for mixing various signals to form a new signal. Also see MIXER.

combiner circuit In a color television camera, the circuit that combines the chroma and luminance with the sync.

comeback A spurious response in a bandpass or band-rejection filter, at a frequency well above or below the passband or stopband.

command **1.** In computer operations, the group of selected pulses or other signals that cause the computer to execute a step in its program. **2.** Instruction.

command chain Part of a computer operation carried out independently as a series of input/output instructions.

command control In automation, electronic control, and computer operations, the performance of functions in response to a transmitted signal.

command destruct signal A signal for instigating the destruction of a missile in flight.

command guidance system A system in which a guided missile and its target are both tracked by radar.

command language A computer language made up of command operators.

command link In a command guidance system, the section that transmits missile-steering commands.

command network A radio communications network in which the chain of command is rigorously defined and followed.

command reference The current or voltage to which a feedback signal is referenced in a control system or servomechanism.

comment A statement written into a computer program for a documentation, rather than implementation (e.g., to describe the purpose of a step or subroutine).

comment field A record or file in which instructions or explanations are given.

commercial data processing A commercial (rather than industrial, scientific, or personal) application of data processing.

commercial-level security See LEVEL-2 SECURITY.

commercial killer A usually remote-controlled, electronic relay for disabling a radio or television receiver during advertisements.

commercial language A computer programming language for commercial applications (payroll, for example).

common **1.** Grounded. **2.** Pertaining to a connection shared by several different points in a circuit or system. **3.** See COMMON GROUND.

common area A computer storage area usable by several programs or segments within a program.

common-base circuit A bipolar transistor circuit in which the transistor base is the common (or grounded) electrode. Also called *grounded-base circuit.*

common battery **1.** A battery shared by two or more different circuits or pieces of equipment. **2.** In wire telephony, a central office battery that supplies the entire system.

common-battery office In wire telephony, a central office that provides a common battery.

common business-oriented language See COBOL.

common-capacitor coupling The process of coupling one tuned circuit to another by means of a capacitor that is common to both circuits.

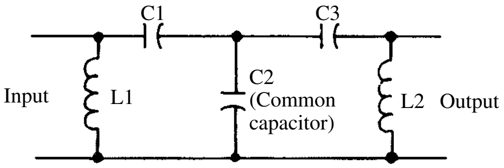

common-capacitor coupling

common-carrier fixed station A fixed radio station that provides public service.

common-cathode circuit A tube circuit in which the cathode is the common (or grounded) electrode. Also called *grounded-cathode circuit.*

common-channel interference Radio interference resulting from two stations transmitting on the same channel. It is characterized principally by beat-note (heterodyne whistle) generation, and suppression or capture of the weaker signal by the stronger one.

common-collector circuit A bipolar-transistor circuit in which the collector is the common (or grounded) electrode. Also called *grounded-collector circuit* and EMITTER FOLLOWER.

common communications carrier A communications company authorized by the licensing agency to furnish public communications.

common-component coupling See COMMON-CAPACITOR COUPLING, COMMON-INDUCTOR COUPLING, and COMMON-RESISTOR COUPLING.

common-drain circuit A field-effect transistor circuit in which the drain terminal is the common (or grounded) electrode. Also called *grounded-drain circuit* and SOURCE FOLLOWER.

common-emitter circuit A bipolar transistor circuit in which the emitter is the common (or grounded) electrode. Also called *grounded-emitter circuit.*

common-gate circuit A field-effect transistor circuit in which the gate is the common (or grounded) electrode. Also called *grounded-gate circuit.*

common-grid circuit A tube circuit in which the control grid is the common (or grounded) electrode. Also called *grounded-grid circuit.*

common ground A single ground-point connection shared by several portions of a circuit.

common impedance A single impedance shared by parts of a circuit. Because currents from the various parts flow through this impedance simultaneously, coupling (desired or undesired) can occur between them.

common-impedance coupling See COMMON-CAPACITOR COUPLING, COMMON-INDUCTOR COUPLING, and COMMON-RESISTOR COUPLING.

common-inductor coupling The process of coupling one tuned circuit to another by means of an inductor that is common to both circuits.

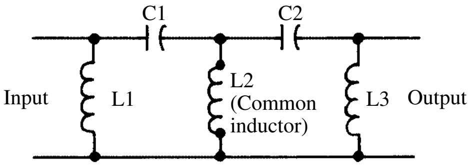

common-inductor coupling

common language A language recognized by all the equipment in a data processing system.

common logarithm Abbreviation, $\log_{10}$. Also called *base-10 logarithm.* A logarithm in which the base number is 10. Also see LOGARITHM.

common mode Pertaining to signals or signal components that are identical in amplitude and duration.

common-mode characteristics In an operational amplifier, characteristics denoting amplifier performance when a common signal is applied to inverting and noninverting inputs.

common-mode gain The voltage gain of a differential amplifier with a common-mode input.

common-mode impedance input The impedance between ground and one of the inputs of a differential amplifier. Compare COMMON-MODE INPUT IMPEDANCE.

common-mode input capacitance In a differential amplifier, the internal capacitance of the common-mode input circuit.

common-mode input circuit In a differential amplifier, the input circuit between ground and the inputs connected together.

common-mode input impedance In a differential amplifier, the open-loop impedance between ground and the inputs connected together. Compare COMMON-MODE IMPEDANCE INPUT.

common-mode input signal A signal applied to the common-mode input circuit of a differential amplifier (i.e., to both inputs connected together). Compare COMMON-MODE SIGNAL.

common-mode input voltage In a differential amplifier, the maximum voltage that can be applied safely between ground and the inputs connected together.

common-mode interference A form of interference that occurs across the terminals of a grounded system.

common-mode rejection The extent to which a differential amplifier will reject a signal presented simultaneously to both inputs in phase, or of two signals identical in amplitude, frequency, and phase applied separately to the two inputs. Also see COMMON-MODE REJECTION RATIO.

common-mode rejection ratio In a differential amplifier, the extent to which the amplifier cancels undesired signals. It is the ratio of the differential gain to the common-mode gain. Also see COMMON-MODE REJECTION.

common-mode signal The algebraic average of two signals applied simultaneously to the two ends of a balanced circuit, such as a differential amplifier. Compare COMMON-MODE INPUT SIGNAL.

common-mode voltage The part of the input that is common to both inputs of a differential amplifier circuit. It is quantitatively defined as the arithmetic mean of the voltages at the inputs.

common-mode voltage gain See COMMON-MODE GAIN.

common-mode voltage range The range limited by the maximum nonsaturating input voltage that can be applied to both inputs of an operational amplifier.

common pool An assigned memory store, utilized by two or more circuits or systems.

common-resistor coupling The process of coupling one circuit to another by means of a resistor that is common to both circuits.

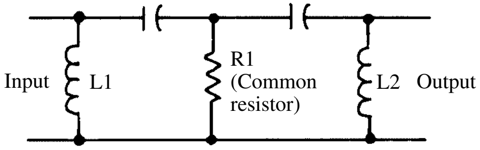

common-resistor coupling

common-source circuit A field-effect transistor circuit in which the source terminal is the common (or grounded) electrode. Also called *grounded-source circuit.*

common-user channels Communication channels open to all licensees in a particular service.

communication band A band of frequencies whose use is authorized expressly for communications, rather than for other services (such as broadcasting, education, remote control, etc.).

communication channel **1.** In radio or wire service, a (usually auxiliary) channel for direct exchange of information between units of the service (e.g., a "talking circuit" between a broadcast studio and the transmitter house). **2.** A data transmission channel between two points (e.g., a remote terminal and a central computer system).

communication link **1.** Collectively, the equipment providing a communication channel between two transmitters. **2.** Data terminal equipment.

communication protocol The specifications of a digital signal, including the speed in bits per second (bps) or bauds, the code type, the bit duration, the mark-to-space ratio, etc.

communications The science and art of using and developing electronic equipment and processes for the transmission and reception of information.

communications common carrier An organization licensed to provide public communication services.

communications network An organization of transmitting and receiving stations for the reliable exchange of intelligence. Also called *net.*

communications receiver A general-coverage or multiband radio receiver, designed primarily for listening to amateur, weather, or other nonbroadcast stations. Compare BROADCAST RECEIVER.

communications satellites Satellites in earth orbit that provide propagation paths (e.g., by reflection or retransmission) for radio waves between terrestrial transmitters and receivers. Also see ACTIVE COMMUNICATIONS SATELLITE and PASSIVE COMMUNICATIONS SATELLITE.

community-antenna television Abbreviation, CATV. A system in which an advantageously located receiving station receives television signals, amplifies them if necessary, and distributes them in the community served by the system. Commonly called *cable TV.*

commutating capacitor **1.** In a flip-flop circuit, a capacitor connected in parallel with the cross-coupling resistor to accelerate the transition from one stable state to the other. Also called *speedup capacitor.* **2.** A capacitor connected in parallel between silicon-controlled rectifier (SCR) stages to momentarily reverse the current going through the SCR, thereby causing the SCR to go into the cutoff condition.

commutation **1.** In a direct-current (dc) generator, periodic reversal of the current in the armature coils as the coils alternately pass the north and south poles of the magnetic field. When the ends of each coil are connected to opposite bars of the commutator, the electrical polarity at the commutator brushes remains constant. **2.** In a thyratron or silicon-controlled rectifier (SCR) circuit, momentarily reversing the polarity to cut the device off.

commutator **1.** In a direct-current (dc) motor, generator, or rotating selector, an arrangement of parallel metal bars or strips on a rotating drum. As the drum turns, the bars contact one or more brushes that are in sliding contact with the commutator. **2.** An electronic circuit that switches a single input sequentially to a series of output terminals, or switches a number of inputs sequentially to a single pair of output terminals.

commutator ripple The pulsating voltage superimposed on the direct-current (dc) voltage delivered by an unfiltered dc generator.

compact disc Abbreviation, CD. A digital, high-density optical disc, used in high-fidelity stereo sound systems. Also used to store computer data. The information is encoded as tiny pits on the surface of the disc, and is recovered by a laser, a sensor and a digital-to-analog (D/A) converter. These disks have largely superseded magnetic tapes, and have rendered long-playing vinyl disks and turntables obsolete. See also COMPACT-DISK READ-ONLY MEMORY.

compact-disk read-only memory Abbreviation, CD-ROM. A digital COMPACT DISC used for the long-term storage of computer data and/or software programs. Usually the same size as a high-fidelity stereo disk, it can hold over 600 megabytes of data. Although data can be read from the disk, it cannot be overwritten.

compander Term for *compressor/expander.* In the transmission and reception of audio-frequency (AF) intelligence, a system that uses an amplitude compressor at the transmitter and an amplitude expander at the receiver. The compressor reduces the dynamic range before transmission, and the expander restores it after reception. Provides improved signal-to-noise ratio under marginal communications conditions. Also increases the ratio of average power to peak power. See COMPANDING.

companding A process in which a signal is compressed at the transmitting end of a circuit and expanded at the receiving end, yielding a signal like the original at the receiver output. Signals are more efficiently transmitted when they are compressed because the average power increases, relative to the peak power. This improves the average signal-to-noise ratio for weak signals. See COMPANDER.

companding law The mathematical function used for companding. It is an output-amplitude versus input-amplitude function for the compression at the transmitter, and the inverse of this function for the expansion at the receiver.

companion keyboard An auxiliary keyboard connected to a regular keyboard and operated remotely.

companionship machine A computer or robot with sufficient machine intelligence to provide entertainment and mental stimulation for humans.

comparator **1.** An integrated circuit (IC) with two inputs, called A and B. The device compares the voltages that appear at these inputs. If the input voltage at A is significantly greater than the input voltage at B, the output is about +5 V. If the input voltage at A is not greater than the input voltage at B, the output voltage is about +2 V. These ICs are used to actuate, or trigger, other devices such as relays and electronic switching circuits. **2.** In general, any circuit that compares some characteristic of two input signals and produces an output that depends on the relationship between the inputs. **3.** An instrument for checking the condition of a component by comparing it directly with an identical component of known quality has a scale reading in percentage deviation, or simply "GO/NO-GO." Examples: capacitor comparator, resistor comparator, coil comparator.

compare In computer operations, a relational test performed on two quantities to determine their relative magnitude, including an indication of the test result and, sometimes, the taking of action. Example: the process and acton resulting from execution of the statement "IF A > B THEN GO TO LINE 250."

comparison **1.** An expression of the relationship between two voltages, currents, phase angles, component values, or other quantities in an electrical or electronic circuit or system. **2.** An examination of different data bits or items, which results in a conclusion about some aspect of their relationship.

comparison bridge A bridge designed especially for the quick comparison of components (e.g., the comparison of resistors with a standard resistor, inductors with a standard inductor, and capacitors with a standard capacitor).

comparison measurement A measurement in which a quantity or component is compared with a known, similar quantity or component value, rather than having the measurement displayed directly by a meter. Examples: bridge measurements, potentiometric measurements, and frequency matching.

compass **1.** Any of several instruments for determining direction on the earth's surface [e.g., magnetic (mariner's) compass and gyrocompass]. **2.** A radio direction finder. **3.** An instrument for drawing circles.

compatibility **1.** A desirable condition in which devices or systems can function efficiently together, without any modification of equipment. **2.** In computer operations, a desirable condition in which different computers can run the same software, without any modification of hardware or software.

compatible color television A color-television system whose transmissions can be received in black and white by any ordinary monochromatic receiver.

compatible integrated circuit A hybrid integrated circuit (IC) that has an active element inside the integrated structure and a passive element deposited on its insulated outer surface.

compensated amplifier A wideband amplifier whose frequency range is extended by special components and circuit modifications. Also see COMPENSATING CAPACITOR and COMPENSATING COIL.

compensated diode detector A diode detector in which a positive dc voltage from the automatic-gain-control (AGC) rectifier is applied to the diode anode. The voltage is always proportional to the signal carrier. The arrangement allows the diode to handle a heavily modulated AM signal without producing excessive distortion.

compensated-impurity resistor A resistor consisting of a diffused semiconductor material to which are added controlled amounts of n- or p-type dopants (impurities).

compensated-loop direction finder A direction finder whose loop antenna is complemented by another antenna for polarization-error compensation.

compensated semiconductor A doped semiconductor material in which the acceptor impurity cancels the effects of the donor impurity.

compensated volume control A combination volume-tone control that provides bass boost at low volume levels to compensate for the ear's deficiency at low frequencies.

compensating capacitor **1.** A capacitor that has a temperature coefficient of capacitance numerically equal to, but having the opposite sign from, that of another capacitor in a tank or other circuit. When the capacitors are connected in parallel, a temperature-induced value change in the main capacitor is balanced by an equal and opposite change in the compensating capacitor; the net capacitance of the circuit does not change. This greatly reduces frequency drift. **2.** In a video amplifier, a large capacitance connected between ground and a tap on the collector or drain resistor to boost low-frequency response. Compare COMPENSATION COIL. **3.** A usually low-capacitance capacitor of known temperature coefficient, operated in combination with a main capacitor to reduce capacitance/temperature drift of the latter to zero or to some desired positive or negative value.

compensating diode A junction diode used to temperature-stabilize a transistor circuit. It is usually forward-biased in the base-bias network of the transistor.

compensating filter **1.** A selective filter used for the purpose of eliminating some irregularity in the frequency distribution of received energy. **2.** A filter used to change the wavelength distribution of electromagnetic energy.

compensating resistor **1.** A low-value resistor of known temperature coefficient, connected in series with a main resistor to reduce the resistance/temperature drift to zero, or to some desired positive or negative value. **2.** See TRIMMER RESISTOR.

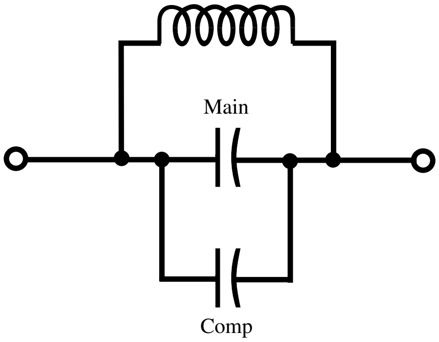

compensating capacitor, 1.

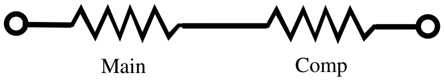

compensating resistor, 1.

compensation Adjusting a quantity, manually or automatically, to obtain precise values, or to counteract undesired variations. Example: temperature compensation of electronic components. For illustration, see COMPENSATING CAPACITOR, **1**.

compensation coil In a video amplifier, an inductor connected in series with the collector or drain resistor, or in the coupling path between stages, or both, to boost high-frequency response.

compensation filter See COMPENSATING CAPACITOR, **2**.

compensation signal A signal recorded on a tape track containing computer data, to ensure that the tape plays back at exactly the correct speed at all times.

compensation theorem An impedance (*Z*) in a network can be replaced by a generator having zero internal impedance, and whose generated voltage equals the instantaneous potential difference produced across *Z* by the current flowing through it. Compare MAXIMUM POWER TRANSFER THEOREM, NORTON'S THEOREM, RECIPROCITY THEOREM, SUPERPOSITION THEOREM, and THEVENIN'S THEOREM.

compensator A device or circuit that facilitates the adjustment of a quantity, manually or automatically, to obtain precise values, or to counteract undesired variations.

compilation time The period during which a program is compiled, as distinct from RUN TIME.

compile **1.** To unify computer subroutines into an all-encompassing program. **2.** To gather information or data together into a single file or file set.

compiler In computer operations, a program that changes a HIGH-LEVEL LANGUAGE, such as BASIC, C, C++, COBOL, or FORTRAN, into MACHINE LANGUAGE. A compiler must be written especially for the high-level language being used.

compiler language Any computer language that serves as an interface between the operator and the computer.

compiler program A program that converts compiler language into machine language.

compiling routine In digital computer operation, a routine permitting the computer itself to construct a program to solve a problem.

complement **1.** The difference between a number and the radix (modulus or base) of the number system. For example, the complement of 7 is equal to 3 (because 10 – 7 = 3) in the decimal (radix-10) number system. **2.** Also called *ones complement.* In computer operations, a representation of the negative value of a binary number. All the available digits are set to 1, and then the number in question is subtracted. For example, the complement of 101 is equal to 010 (because 111 – 101 – 010); the complement of 1001 is equal to 0110 (because 1111 – 1001 = 0110).

complementary A Boolean operation whose result is the same as that of another operation, but with the opposite sign; thus, OR and NOR operations are complementary.

complementary colors **1.** In the additive color system, two colors that produce light gray or white when combined. **2.** In the subtractive color system, two pigments that produce dark gray or black when combined. **3.** Colors or pigments that are opposite each other on the color wheel.

complementary constant-current logic A form of bipolar logic with high operating speed and high component density.

complementary metal-oxide semiconductor Also sometimes called *complementary metal-oxide silicon.* Acronym, CMOS (pronounced "seamoss"). A digital integrated-circuit (IC) technology, in which logic gates are formed by n-channel/p-channel pairs of metal-oxide-semiconductor field-effect transistors (MOSFETs) fabricated on a substrate. Noted for high speed and low current drain.

complementary operator The logical negation (NOT) operation.

complementary pushpull circuit See COMPLEMENTARY-SYMMETRY CIRCUIT.

complementary rectifier In the output circuit of a magnetic amplifier, nonsaturating half-wave rectifier elements.

complementary silicon-controlled rectifier A silicon-controlled rectifier that has polarity opposite from the usual silicon-controlled rectifier.

complementary-symmetry circuit A bipolar-transistor circuit that uses an npn and pnp transistor. The transistors conduct during opposite half-cycles of the input signal, the result being that push-pull output is provided with a single-ended input; no phase-splitting input circuit is required. The complementary-symmetry circuit offers very low output impedance, permitting a loudspeaker voice coil (or other low-impedance load) to be operated directly without a coupling transformer.

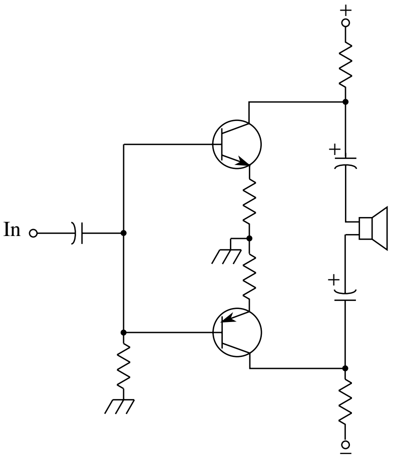

complementary-symmetry circuit

complementary tracking A control system in which several secondary (slave) devices are controlled by a primary (master) device.

complementary transistors A transistor pair of opposite polarity operated in a complementary-symmetry circuit or its equivalent.

complementary wave An electromagnetic wave in a transmission line that occurs as a result of reflection. Any impedance discontinuity will result in complementary waves.

complementer A logic circuit that provides an output pulse when there is no input pulse, and vice versa. Also called INVERTER and NOT CIRCUIT.

complement number In a base-*n* number system, for a given positive integer *p* less than *n*, the positive integer *m* such that $m + p = n$. For example, in the decimal (base-10) system, the complement of 4 is 6, the complement of 7 is 3, and the complement of 9 is 1. In the hexadecimal (base-16) number system, the complement of 4 is 12, the complement of 7 is 9, and the complement of 9 is 7.

complement-number handling A computer system in which the operations are carried out via the complements of the input numbers.

complement-setting technique A process of determining the number of pulses required to complete the switching of a counter circuit when it is started at some state higher than full zero. The number of pulses required for completion is equal to the number that represents the starting state's complement.

complete carry In digital computer operation, a system permitting all carries to generate carries.

complete circuit See CONTINUOUS CIRCUIT.

complete modulation Modulation to the maximum extent possible while maintaining acceptable circuit or system operation.

complete operation In computer operations, the condition in which the machine rigorously follows program instructions.

complete routine A vendor-supplied computer program that is usable without modification.

complex function **1.** A mathematical function of a complex-number variable. **2.** An integrated circuit (IC) containing two or more subcircuits that perform an operation more complicated that of any one of the circuits alone.

complex notation Notation taking into consideration both the real-number and imaginary-number components of a quantity. Thus, impedance (Z) is a complex quantity that includes a resistive (real) component (R) and a reactive (imaginary) component (jX). See COMPLEX NUMBER and COMPLEX OPERATOR.

complex number A number expressed in complex notation (e.g., $a + jb$, where a and b are real numbers and j is the COMPLEX OPERATOR). Can also be expressed as a point or a vector in an ARGAND DIAGRAM.

complex operator The unit imaginary number, represented as j by engineers and as i by mathematicians. This number is defined mathematically as the positive square root of –1.

complex parallel permeability An expression of the permeability of an inductor core under actual operating conditions, assuming zero loss in the conductors of the coil winding. A parallel combination of reactance and resistance.

complex periodic wave A periodic wave composed of a sine-wave fundamental and certain harmonics in specific proportions.

complex permeability An expression of inductor-core permeability, obtained from the mathematical ratio of the magnitudes of the vectors representing the induction and electromagnetic field strength within the core.

complex plane A Cartesian coordinate system with real numbers on the horizontal axis and imaginary numbers on the vertical axis. Used for vectorial representation of complex numbers. See ARGAND DIAGRAM.

complex quantity A quantity containing both real and imaginary components. Example: Impedance (Z) is a complex combination of resistance R (a real component) and reactance jX (an imaginary component): $Z = R + jX$.

complex radar target A radar target that is large enough in theory to be detected by radar, but, because of its geometry, cannot be detected. This effect is the result of phase combinations of signal components reflected from various surfaces on the target.

complex series permeability An expression of complex permeability of an inductor core under actual operating conditions, assuming zero loss in the conductors of the coil winding. A series combination of reactance and resistance.

complex steady-state vibration Periodic vibration with more than one sine-wave component.

complex tone An audio tone made up of more than one sine-wave component.

complex variable A variable having real and imaginary parts.

complex waveform The shape of a COMPLEX PERIODIC WAVE. It is the resultant of the individual sine-wave components (i.e., of the fundamental and the harmonics).

complex-wave generator A signal generator whose output signal is any of several selectable waveforms and frequencies (or repetition rates). Also see FUNCTION GENERATOR.

compliance **1.** The ease with which a material can be flexed or bent, an important characteristic of transducers (such as loudspeakers). Expressed in cm/dyne, compliance is the reciprocal of stiffness, and is the acoustical or mechanical equivalent of capacitance. **2.** A measure of the output impedance of a switched-current signal source. Generally given as maximum current for a certain change in the voltage.

compliance range The voltage range required to maintain a constant current throughout a load-resistance range.

compliance voltage The range over which the output voltage of a constant-current power supply must swing in order to maintain a steady current in a varying load.

compliance-voltage range The output voltage range of a constant-current power supply.

component **1.** A device or part used in a circuit to obtain some desired electrical action [e.g., a resistor (passive component) or an integrated circuit (active component)]. Also see ACTIVE COMPONENT and PASSIVE COMPONENT. **2.** An attribute inherent in a device, circuit, or performance (e.g., the REACTIVE COMPONENT of an inductor). **3.** A specified quantity or term (e.g., the WATTLESS COMPONENT of ac power). **4.** A piece of equipment in a high-fidelity sound system.

component density The number of components (see COMPONENT, **1**) in an electronic assembly of a given physical volume.

component failure rate **1.** The percentage of components, out of a specified group, that can be expected to fail within a specified length of time. **2.** The frequency with which a given component, in a certain application, can be expected to fail.

component layout The mechanical arrangement of components (see COMPONENT, **1**) in an electronic assembly.

component stress The electrical or mechanical strain to which a component is subjected. In general, the greater the stress, the higher the component failure rate.

composite cable A cable containing other cables of different types.

composite circuit A circuit handling telegraphy and telephony simultaneously without causing mutual interference.

composite color A color that is not one of the COLOR PRIMARIES, but instead, consists of a combination of the three color primaries.

composite color signal The complete color television signal, including all picture, color, and control components.

composite conductor A set of wires connected in parallel. The wires are often, but not necessarily, of identical size and constitution.

composite current A current having both alternating-current (ac) and direct-current (dc) components; an alternating current superimposed on a direct current. Also called *fluctuating current.*

composite curve A curve or pair of curves showing two modes of operation, as of biased and unbiased conditions.

composite filter A filter consisting of more than one section. The sections might be, but often are not, identical.

composite video signal The television picture signal containing picture information and sync pulses.

composite-video-signal distortion Distortion of the composite video signal as evidenced by overshooting, ringing, and sync-pulse shortening.

composite voltage A voltage having both alternating-current (ac) and direct-current (dc) components; an ac voltage superimposed on a dc voltage. Also called *fluctuating voltage.*

composite wave filter Two or more wave filters (not necessarily of the same type) operated in cascade.

composition resistor A resistor made from a mixture of materials, usually finely powdered carbon and a binder.

compound A substance in which the atoms of two or more elements have united chemically to form a molecule. For example, an atom of cadmium (Cd) and one of sulfur (S) combine to form a molecule of cadmium sulfide (CdS).

compound connection A direct connection of two transistors, the amplified output of the first being further amplified by the second. The connection provides extremely high current gain. Also called DARLINGTON PAIR.

compound generator A generator that has both series and shunt fields. Also called *compound-wound generator.*

compound horn A horn reflector used for transmission of microwave energy. The faces of the horn approach four geometric plane surfaces as the distance from the center increases.

compound modulation A system of successive modulation, the modulated wave from one step becoming the modulating wave in the next. Also called *multiple modulation.*

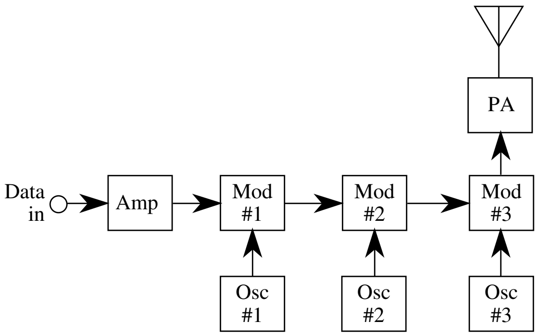

compound modulation

compound motor An electric motor having both series and shunt fields. Also called *compound-wound motor.*

compound transistor Two or more transistors directly coupled in the same envelope for increased amplification. Also see COMPOUND CONNECTION.

compound-wound generator See COMPOUND GENERATOR.

compound-wound motor See COMPOUND MOTOR.

compress **1.** In communications, to reduce or minimize the bandwidth of a signal. **2.** In communications, the processing of a signal to increase low-level components and thereby raise the average power level relative to the peak power level. **3.** In computer operations, to reduce or minimize the number of bits in a digital signal or file, while still retaining all the essential information. Compare EXPAND.

compressed-air capacitor A high voltage air-dielectric capacitor enclosed in a case in which the air pressure is held at several atmospheres. The device exploits the dielectric strength of compressed air, which is higher than that of air at normal pressure.

compressed-air speaker A speaker that uses an airtight chamber to enhance the acoustic reproduction at certain frequencies.

compression **1.** In communications, the reduction or minimization of signal bandwidth. **2.** In communications, the processing of a signal to increase low-level components and thereby raise the average power level relative to the peak power level. Usually, a logarithmic function. **3.** In computer operations, the reduction or minimization of the number of bits in a digital signal or file, while still retaining all the essential information. Compare EXPANSION.

compression ratio In a system using COMPRESSION, the ratio A_1/A_2, where A_1 is the gain (or transmission) at a reference-signal level and A_2 is the gain (or transmission) at a specified higher signal level.

compression wave A wave disturbance that travels via longitudinal motion of particles in a medium. Sound waves through air are the most common example.

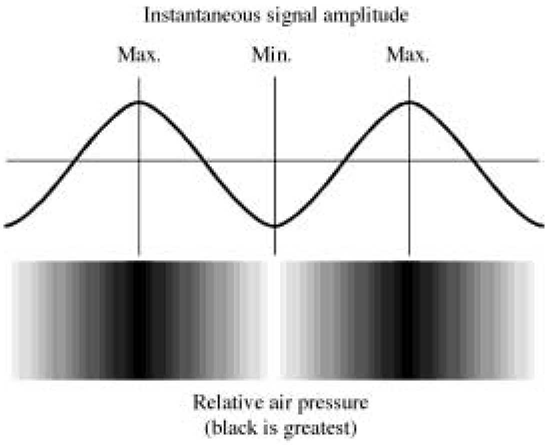

compression wave

compressor A circuit or device that limits the amplitude of its output signal to a predetermined value, despite wide variations in input signal amplitude.

compressor driver unit A loudspeaker that works into an air space connected by a throat to a horn, rather than by driving a diaphragm.

Compton diffusion An effect that occurs when a photon and electron collide. Some of the energy from the photon is transferred to the electron. On a large scale, such collisions result in diffusion of electromagnetic waves.

Compton effect The increase in wavelength (decrease in frequency) of X-rays scattered by the electrons of lighter atoms bombarded with the X-rays.

Compton shift See COMPTON EFFECT.

compute To perform a mathematical operation by means of a relatively simple process. Thus, a digital computer solves intricate problems using simple arithmetic steps. Compare CALCULATE.

computer A device or machine for performing mathematical operations on data, and producing the results as information or control signals. There are numerous types, the most common being the *digital computer.*

computer-aided design Abbreviation, CAD. The use of computers in conceiving, developing, and perfecting new products.

computer-aided manufacturing Abbreviation, CAM. The use of automated manufacturing systems, such as assembly lines, that are partially or totally controlled by computers.

computer antibody Also called *vaccine.* A small subprogram designed to eliminate viruses from computer systems.

computer-assisted instruction Abbreviation, CAI. The use of computers as teaching and training aids.

computer code See MACHINE LANGUAGE.

computer consciousness The degree to which a machine can be considered aware of its own existence. Until recently, this idea was considered ridiculous. But as microprocessor power continues to grow, some researchers now consider it worth thinking about.

computer-controlled catalytic converter A microprocessor-controlled system for automatically supervising gaseous emissions exhausted by a motor vehicle. An oxygen sensor monitors the exhaust stream, and the associated electronic system adjusts the air-to-fuel ratio of the carburetor to reduce smog-producing pollutants in the exhaust.

computer diode A semiconductor diode having low capacitance and fast RECOVERY TIME, thus suiting it to rapid switching in computer circuits and to very-high-frequency applications.

computer engineer A person skilled in the theory and application of computers, related equipment, and associated mathematics.

computer file See FILE.

computer game See VIDEO GAME.

computer graphics **1.** The use of computers to assist in drawing and drafting, and in the processing of video images such as photographs. **2.** Broadly, any computer-generated or computer-processed image.

computer instruction See INSTRUCTION.

computer interfacing apparatus The equipment used to connect a computer to other systems, and to peripherals.

computerized axial tomography Abbreviation, CAT. A multiple X-ray system that enables the observer to obtain cross-sectional images of the internal organs of the body.

computer map A blueprint, used in conjunction with machine vision, sonar, radar or beacons, that a mobile robot can use as a navigational aid. One or more such blueprints are stored in the robot controller's main memory.

computer music **1.** Music that is composed by a computer. **2.** See MUSICAL INSTRUMENT DIGITAL INTERFACE.

computer program See PROGRAM.

computer programmer A person skilled in devising and/or writing the routines that a digital computer uses to solve problems or process data.

computer storage tube A cathode-ray tube in which the electron beam scans and stores information in thousands of memory cells on a target. A cell "remembers" by acquiring and holding an electrostatic charge when it is struck by the beam from the writing gun. Information taken is read out of a cell by a second beam from the reading gun.

computer system A central processor and its associated online and offline peripherals, such as a monitor, modem, printer, optical scanner, magnetic disk drives, CD-ROM drive, and tape backup.

computer technician A professional skilled in building, repairing, and maintaining computers, and who, occasionally, designs them. Usually works under the supervision of a computer engineer.

computer terminal **1.** A teleprinter or video display unit and keyboard, used by human operator(s) of a computer. **2.** An interface between a computer and its human operator(s).

computer/TV interface A device or circuit for delivering the output of a digital computer to a standard television receiver so that the latter can serve as a GRAPHIC TERMINAL.

computer virus A deliberately created and disseminated subprogram or piece of programming code, that electronically spreads through computer systems and hinders operation. Usually diverts the computer(s) from intended functions; sometimes causes a catastrophic malfunction. Often exists undetected, being transferred from one computer to another by means of diskettes or software.

computer word See WORD.

computing amplifier See OPERATIONAL AMPLIFIER.

computing machine See COMPUTER.

concatenation **1.** A method of speed control for a 3-phase motor in which two induction motors are operated with their shafts coupled together. The stator of the first motor is connected to the 3-phase supply, and the slip rings of this motor are connected to the field of the second motor. The slip rings of the second motor are connected to the three ganged sections of a Y-rheostat used for adjusting the speed. **2.** Arrangement of a set into a series.

concentrated-arc lamp A brilliant low-voltage lamp, containing nonvaporizing electrodes in an inert-gas atmosphere. An arc across the electrodes creates the light source.

concentrated winding A coil winding that has a large number of turns in a small space.

concentration cell An electrolytic cell in which two electrodes are immersed in solutions of the same compound but having different combinations. The voltage is usually very small, 0.1 volt or less.

concentration gradient Between points in a semiconductor, the difference in electron or hole concentration.

concentric cable See COAXIAL CABLE.

concentric capacitor A fixed or variable capacitor whose plates are concentric cylinders. Also called *concentric-plate capacitor.*

concentric jack See COAXIAL JACK.

concentric line See COAXIAL LINE.

concentric-line oscillator A stable, self-excited oscillator whose frequency-determining tank consists principally of a section of concentric (coaxial) line. Used primarily at ultra-high frequencies (UHF).

concentric plug See COAXIAL PLUG.

concentric receptacle See COAXIAL RECEPTACLE.

concentric tank See COAXIAL TANK.

concentric-wound coil A combination of two or more coils wound on top of, and insulated from, each other.

conceptual modeling A technique for solving problems by devising a mathematical model based on the results of an experiment; experiments performed on the model are used to verify its validity.

concurrent conversion In computer operations, running conversion and conventional programs together. Also see CONVERSION PROGRAM.

concurrent processing See MULTIPROGRAMMING.

condenser **1.** An obsolete term for CAPACITOR. **2.** A mirror or lens for concentrating light (on an object, for example). **3.** Something that condenses a gas or vapor. **4.** See CONDENSER MICROPHONE.

condenser antenna A two-wire horizontal antenna system in which the radiator is a wire situated above a counterpoise.

condenser microphone Also called *capacitor microphone.* A microphone in which a tightly stretched metal diaphragm forms one plate of an

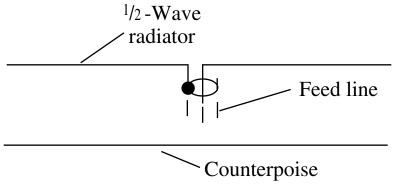

condenser antenna

air capacitor, and a closely situated metal plug forms the other plate. A dc bias voltage is applied to the arrangement. Impinging sound waves cause the diaphragm to vibrate, varying the capacitance and causing the output current to fluctuate accordingly.

condensing routine In computer operations, a program that compresses data. See COMPRESSION, **3**.

condensite A plastic insulating material whose base is phenol formaldehyde resin.

conditional Pertaining to a quantity or phenomenon that depends on some external factor, and is therefore subject to change.

conditional branch A point in a computer program where a relational test is performed, and the statement line in which the test is made is left so that an out-of-sequence instruction can be implemented. Such a branch might be made, for example, following a statement, such as "IF Z = Y THEN GO TO LINE 380."

conditional branch instruction The instruction in a computer program that causes a CONDITIONAL BRANCH.

conditional implication operation A Boolean operation in which the result of operand values X and Y are such that the output is high only if input X is high and input Y is low. Also called *inclusion* or *if-then operation.*

conditional jump See CONDITIONAL BRANCH.

conditional stop instruction In a computer program, an instruction that can cause a halt in the run, as dictated by some specified condition.

conditional transfer See CONDITIONAL BRANCH.

condition code A set of constraints for a computer program; sets limits on what can be done with the computer under certain circumstances.

conditioning **1.** The process of making equipment compatible for use with other equipment. Generally involves some design or installation changes. **2.** Interfacing.

Condor A continuous-wave navigational system that produces a cathode-ray-tube display for automatically determining the bearing and distance from a ground station.

conductance Symbol, G. Unit, siemens. The ability of a circuit, conductor, or device to conduct electricity. Conductance in siemens is the reciprocal of resistance: $G = 1/R$, where R is the resistance in ohms.

conducted heat Heat transferred by conduction through a material substance, as opposed to *convection* and *radiation.* A heatsink conducts dissipated energy away from a transistor, whereupon convection and radiation allow heat to escape from the sink.

conduction **1.** The propagation of energy through a medium, depending on the medium for its travel. **2.** The transfer of electrons through a wire. **3.** The transfer of holes through a P-type semiconductor material. **4.** Heat transfer through a material object (see CONDUCTED HEAT).

conduction angle See ANGLE OF CONDUCTION.

conduction band In the arrangement of energy levels within an atom, the band in which a free electron can exist; it is above the *valence band* in which electrons are bound to the atom. In a metallic atom, conduction and valence bands overlap; but in semiconductors and insulators, they are separated by an energy gap.

conduction current **1.** The electromagnetic-field flow that occurs in the direction of propagation. A measure of the ease with which the field is propagated. **2.** Current in a wire or other conductor.

conduction-current modulation In a microwave tube, cyclic variations in the conduction current; also, the method of producing such modulation.

conduction electron See FREE ELECTRON.

conduction error In a temperature-acutated transducer, error caused by conduction of heat between the sensor and the mounting.

conduction field An energy field that exists in the vicinity of an electric current.

conductive coating A conducting layer applied to the glass envelope of a cathode-ray tube, such as an oscilloscope tube or picture tube.

conductive coupling See DIRECT COUPLING.

conductive material See CONDUCTOR.

conductive pattern The pattern of conductive lines and areas in a printed circuit.

conductivity Unit, S/m (siemens per meter). An expression of conductance per unit length of a material; the reciprocal of *resistivity.*

conductivity meter A device for measuring electrical conductivity. Generally, such a device is calibrated in siemens.

conductivity modulation In a semiconductor, the variation in conductivity that results from a variation of charge-carrier density.

conductivity-modulation transistor A transistor in which the bulk resistivity of the semiconductor material is modulated by minority carriers.

conductor **1.** A material that allows charge carriers (usually electrons) to move with ease among atoms. Examples are metals, electrolytes, and ionized gases. Substances vary widely in their suitability as conductors; the conductivity of commercial copper, for example, is almost twice that of aluminum. Compare INSULATOR. **2.** An individual conducting wire in a cable, insulated or uninsulated.

conduit A hollow tube, made of plastic or metal, through which wires, cables, and other transmission media are fed.

cone The conical diaphragm of a (usually dynamic) loudspeaker.

cone antenna An antenna in which the radiator is a sheet-metal cone or a conical arrangement of rods or wires.

Conelrad An early amplitude-modulation (AM) broadcast protocol, intended for use in the event of a nuclear war. Now replaced by the EMERGENCY BROADCAST SYSTEM.

cone marker A UHF marker beacon whose conical energy lobe radiates vertically from a radio-range beacon station. Aircraft in flight use such markers to accurately locate the beacon station.

cone of protection The zone surrounding a lightning rod, in which the chances of a lightning strike are greatly reduced. The cone has an apex angle of 45°, relative to the rod. Objects entirely within this cone are unlikely to be struck (although it is still possible).

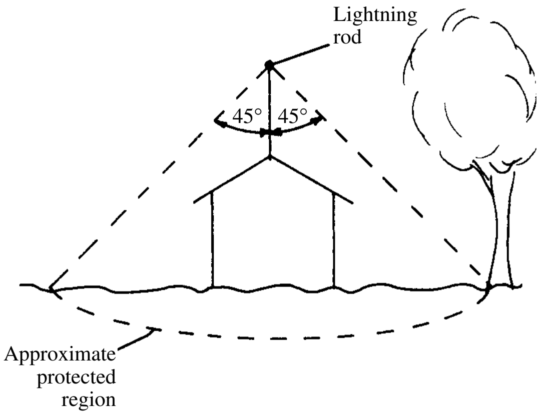

cone of protection

cone of silence A small zero-signal zone directly over a low-frequency radio-range beacon. The zone is the product of the combined directive properties of the beacon transmitting antenna and the antenna on an aircraft.

cone speaker A loudspeaker having a sound-producing cone (diaphragm) made of specially treated paper or other material, as opposed to a loudspeaker having a flat diaphragm.

confetti On a color TV screen, color spots caused by chrominance-amplifier noise.

confidence The probability that a predicted result will occur.

confidence factor Confidence, expressed either as a fraction (between 0 and 1) or as a percentage.

confidence interval The range over which a parameter can vary so that a given confidence factor is maintained.

confidence level See CONFIDENCE FACTOR.

confidence limitations The maximum and minimum points of a confidence interval. Outside the confidence-limitation points, the confidence factor drops below the required minimum.

configuration **1.** The characteristic arrangement of components in an electronic assembly, or of the equipment symbols in the corresponding circuit diagram. **2.** Computer system.

configuration state In a computer system, an expression of the availability status of a device for a given application. A *configured-in* device is available; a *configured-out* device is available, but is restricted to certain users; a *configured-off* device is unavailable.

configuration table Within a computer's operating system, a table that provides the configuration state for various system units.

configured-in See CONFIGURATION STATE.

configured-off See CONFIGURATION STATE.

configured-out See CONFIGURATION STATE.

conformance The degree to which a quantity or variable corresponds to a standard or to expectations.

conformance error The extent (usually expressed as a percentage) to which conformance is lacking.

conical antenna See CONE ANTENNA.

conical horn A horn (antenna, loudspeaker, or sound pickup) having the general shape of a cone: the cross-sectional area varies directly as the square of the horn's axial length.

conical monopole antenna An unbalanced broadband antenna that derives its name from its shape. It is usually constructed from wire and must be operated against a good radio-frequency (RF) ground.

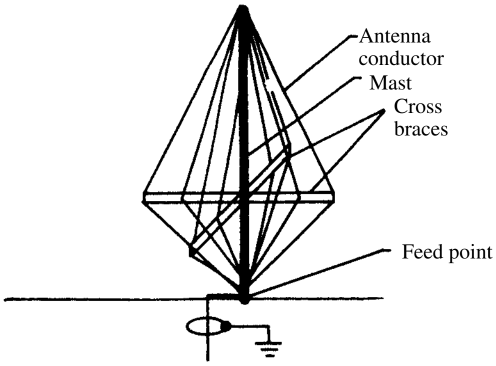

conical monopole antenna

conical scanning In radar transmission, a method of scanning in which the beam describes a cone, at the apex of which is the antenna.

conic sections The geometric plane figures that result from the intersection of a cone with a plane. These figures are the *circle*, the *ellipse*, the *parabola*, and the *hyperbola.*

conjugate For a given complex number $A + jB$, the quantity $A - jB$. When complex conjugates are multiplied together, the result is $A^2 + B^2$.

conjugate branches In a network, two branches of such a nature that a signal in one has no effect on the other.

conjugate bridge A bridge in which the detector and generator occupy positions opposite to those in a conventional bridge of the same general type.

conjugate impedance For a given complex impedance, $R + jX$, where R is the resistive component and jX is the reactive component, the impedance: $R - jX$. The resistance is identical; the reactance is of equal magnitude, but opposite sign (capacitive as opposed to inductive, or vice versa).

conjunction The logical AND operation.

connect To provide an electrical path between two points.

connection The point at which two conductors are physically joined.

connective An operation symbol written between operands.

connector **1.** A device that provides electrical connection. **2.** A fixture (either male or female) attached to a cable or chassis for quickly making and breaking one or more circuits. **3.** A symbol that connects points on a flowchart.

conoscope A device that uses focused polarized light to examine crystals (as in checking the optical axis of a quartz crystal).

consequent poles The poles of an equivalent single magnet that is formed when two magnets are aligned with their two identical poles together. Thus, when the two north poles are placed together, the consequent poles are a south pole at each end and a north pole at the center.

conservation of energy **1.** The preservation of the potential for work by a given quantity of energy—even when it undergoes a change in form within a system. **2.** The *law of conservation of energy*, which states that energy can be neither created nor destroyed, but only changed in form.

console **1.** The main station or position for the control of electronic and/or computer equipment. **2.** The equipment at a fixed location. **3.** An equipment-containing cabinet that stands on the floor. **4.** Equipment permitting communication with a computer. Also called *dumb terminal.*

consonance **1.** Harmony between audio tones. **2.** Acoustical or electrical resonance between bodies or circuits that are not physically connected.

constant **1.** A quantity whose value remains fixed, such as the speed of light in a vacuum. Compare VARIABLE. **2.** The value of a component specified for use in a particular electronic circuit. **3.** An electronic component, particularly a capacitance or inductance. **4.** In a computer program, data items that remain unchanged for each run.

constant-amplitude recording In sound recording, the technique of holding the maximum amplitude of the signal steady as the frequency changes.

constantan An alloy of copper and nickel used in some thermocouples and standard resistors.

constantan-platinum thermocouple A thermocouple that uses the junction between constantan and platinum wires, which is contained in thermocouple-type meters.

constant area As allocated by a computer program, an area of memory that holds constants.

constant bandwidth In a broadband tuned circuit, bandwidth that does not change with frequency.

constant current A current that undergoes no change in value as it flows through a changing resistance. Compare CONSTANT VOLTAGE.

constant-current characteristic A condition in which the current through a circuit remains constant—even if the voltage across the circuit increases or decreases.

constant-current curve A graph in which the dependent variable is an electric current that levels off at, or approaches, a specific maximum. An example is the collector-current versus collector-voltage curve for a bipolar transistor.

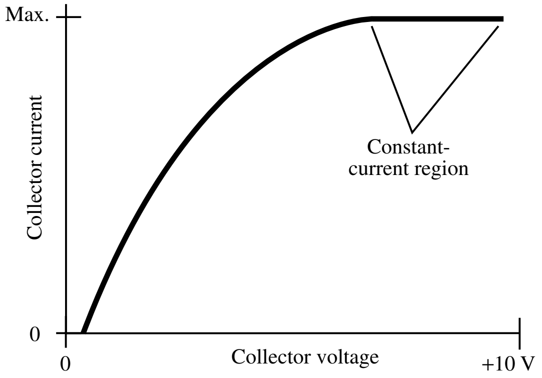

constant-current curve

constant-current drive Driving power obtained from a constant-current source.

constant-current modulation See CHOKE-COUPLED MODULATION.

constant-current power supply See CONSTANT-CURRENT SOURCE.

constant-current sink See CURRENT SINK.

constant-current source A power supply whose current remains steady during variations in load resistance. Also called *constant-current supply* and *current-regulated supply.* Compare CONSTANT-VOLTAGE SOURCE.

constant-current supply See CONSTANT-CURRENT SOURCE.

constant-current transformer A transformer supplied from a constant-voltage source that automatically delivers a constant current to a varying secondary load.

constant-k filter Also called a *Zobel filter.* A filter section in which Z_1Z_2 equals k^2 at all frequencies, where Z_1 is the impedance of the series element and Z_2 is the impedance of the shunt element.

constant-power dissipation line A line connecting points on a family of current-voltage characteristic curves, the points corresponding to the maximum power that can safely be dissipated by the device to which the curves apply.

constant-resistance network A circuit of resistors that, when terminated in a resistance load, presents a constant resistance to a driving source under various conditions of operation.

constant-speed motor **1.** Also called a *shunt motor.* A motor whose speed varies little, or not at all, with variations in the armature current. **2.** A motor that runs at an unvarying speed through the action of associated automatic electronic control circuitry.

constant voltage A voltage that does not change as the load resistance varies. Compare CONSTANT CURRENT.

constant-voltage, constant-current supply A combination current-regulated and voltage-regulated power supply; delivers constant current to low load resistances and constant voltage to high load resistances.

constant-voltage drive Driving power obtained from a CONSTANT-VOLTAGE SOURCE.

constant-voltage source A power supply whose output voltage remains steady during variations in load current. Also called *constant-voltage supply* and *voltage-regulated supply.*

constant-voltage transformer A special transformer used to reduce variations in power-line voltage. A capacitor in the device causes a winding to resonate at the line frequency (e.g., 60 Hz). This tends to maintain a more constant current than would be the case in an ordinary transformer.

construct A source (user's) computer program statement that, when implemented, produces a predetermined effect.

consumer reliability risk **1.** The chance a consumer takes when buying a component or piece of equipment that has not been subjected to quality-assurance/quality-control (QA/QC) testing. **2.** An expression of the failure rate for a consumer item.

contact **1.** A conducting body (such as a button, disk, or blade) that serves to close an electric circuit when pressed against another conductor. Example: switch contact and spring contact. **2.** The state of being touched together, as when two conductors are brought into contact to close a circuit.

contact arc The arc that initially occurs when current-carrying contacts are separated.

contact area **1.** The face of an electrical contact. **2.** The common area shared by two conductors in mutual contact.

contact bounce The springing apart or vibration of contacts upon making or breaking.

contact chatter The abnormal vibration of mating contacts, caused by contact bounce or by an extraneous alternating current.

contact-closure input The input circuit of a device, such as a control-system amplifier, that is actuated by the closing of switching contacts. Compare CONTACT-OPEN INPUT.

contact combination The set of contacts on a switch or electronic relay.

contact detector A rectifier or demodulator, composed of two dissimilar materials in contact with each other. Semiconductor diodes are of this general type. Some contact-detector action can be obtained with two dissimilar fine wires (such as copper and iron) by touching their tips lightly together.

contact EMF Short for *contact electromotive force*; also called *contact potential.* A low direct-current (dc) voltage that is sometimes generated by the contact of two dissimilar materials.

contact follow The tendency of relay contacts to follow the actuating signals.

contact force **1.** The force with which relay contacts close with a given amount of coil current. **2.** The force with which a pair of relay contacts are held together when current flows through the coil. **3.** In a mercury-wetted relay, the force exerted by the mercury on the contacts as the relay closes.

contact gap The distance between contacts when they are open.

contact load **1.** The power dissipated by a load that is connected to a power supply through a closed set of contacts. **2.** The current passing through a set of closed contacts.

contact microphone A microphone placed in direct contact with a vibrating surface for pickup. Actuated by the vibration of a solid, rather than by the movement of air molecules.

contact miss **1.** The improper alignment of contacts in a switch or relay. **2.** The condition of relay contacts not lining up properly.

contact modulator An electromechanical CHOPPER.

contact-open input The input circuit of a device, such as a control-system amplifier, that is actuated by the opening of switching contacts. Compare CONTACT-CLOSURE INPUT.

contactor A switch used for frequent opening or closing of a circuit. An example is a relay contactor used for keying a transmitter.

contactor noise **1.** Electrical noise that is the product of make-and-break contact action or fluctuations in conduction when the contacts are

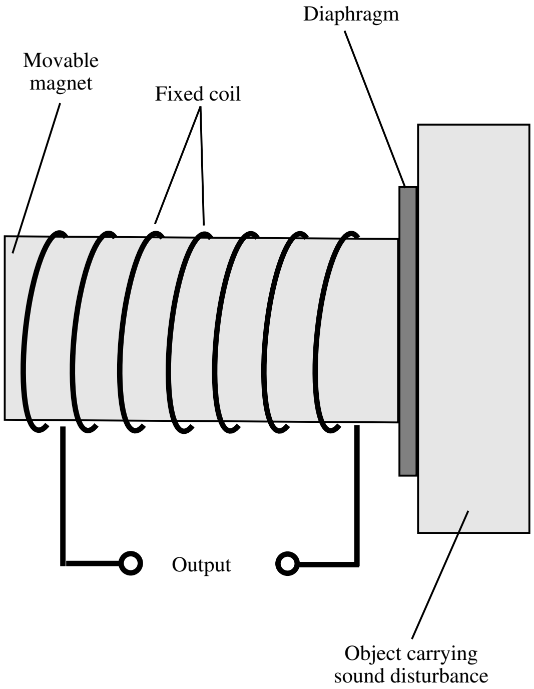

contact microphone

closed. **2.** Sounds coming directly from contacts that are opening and closing.

contact potential The small direct-current (dc) voltage that results from the bombardment of an electrode by electrons, when the electrode has no external voltage applied to it.

contact pressure The pressure that holds contacts together.

contact protector A component (such as a diode, capacitor, resistor, or combination of these) that serves to suppress contact arcing.

contact rating The maximum current, voltage, and/or power specified for a given set of contacts.

contact rectifier A rectifier consisting of two dissimilar materials in direct contact. Examples: copper and copper oxide, magnesium and copper sulfide, selenium and aluminum, and germanium and indium.

contact resistance The resistance of the closed contacts of switches, relays, and other similar devices. Normally, this is a very small resistance.

contact separation See CONTACT GAP.

contact strip See TERMINAL STRIP.

contact switch An electromechanical switch that uses contacts to make and break a circuit, as compared with an electronic switch that uses semiconductor devices.

contact travel The distance over which a relay or switch contact must move to close a circuit.

contact wetting The use of mercury (a conducting liquid) to improve the action of a relay contact or contacts.

contact wipe A sliding motion between closed contacts. Helps to establish a good connection and to keep the contact surfaces clean.

container file See CONTROLLING FILE.

contaminated material **1.** A semiconductor material containing some undesired substance. **2.** A material unintentionally made radioactive.

contamination **1.** The presence of an impurity in a substance. **2.** The addition of a radioactive material to a substance. **3.** In a coaxial cable, the tendency for the jacket material to bleed through the outer braid into the dielectric, resulting in increased loss.

content-addressed storage In a computer, memory- or data-storage locations identified by content (see CONTENTS), instead of by address. Also called *associative storage.*

contention The result of interference among more than one transmitting station on the same communications channel.

contents **1.** The data in a computer random-access memory (RAM). **2.** The data in a specific storage location, such as on a hard disk, diskette, or CD-ROM.

context **1.** The environment in which a word is used in a natural language (such as English, Spanish, or Russian). Important in speech recognition and speech synthesis. **2.** The environment in which a string of characters, composing a data unit or word, is used in a computer program.

Continental code A version of the Morse code used internationally in radiotelegraphy. Also called *International Morse code* and *general service code.* Compare AMERICAN MORSE CODE.

continuity A condition of being uninterrupted—especially pertaining to current flowing in an electrical or electronic circuit.

continuity test A test of the completeness of an electrical path. Ideally, the only concern is whether the circuit is open or closed, but sometimes circuit resistance is also of interest.

continuity tester A device (such as an ohmmeter, battery and buzzer, and battery and lamp) with which a continuity test can be made.

continuity writer The person who prepares copy for a radio or television broadcaster.

continuous carrier A medium (such as a radio-frequency wave) that will convey information (as when the carrier is modulated) with no disruption of the medium itself.

continuous circuit An uninterrupted circuit.

Continuous Commercial Service Abbreviation, CCS. A category in which safe operating parameters are listed for electronic components and communications equipment operated over long, uninterrupted periods. Compare INTERMITTENT COMMERCIAL AND AMATEUR SERVICE.

Character	*Symbol*
A	•–
B	–•••
C	–•–•
D	–••
E	•
F	••–•
G	––•
H	••••
I	••
J	•–––
K	–•–
L	•–••
M	––
N	–•
O	–––
P	•––•
Q	––•–
R	•–•
S	•••
T	–
U	••–
V	•••–
W	•––
X	–••–
Y	–•––
Z	––••
0	–––––
1	•––––
2	••–––
3	•••––
4	••••–
5	•••••
6	–••••
7	––•••
8	–––••
9	––––•
Period	•–•–•–
Comma	––••––
Query	••––••
Slash	–••–•
Dash	–••••–
Break (pause)	–•••–
Semicolon	–•–•–•
Colon	–––•••

Continental code

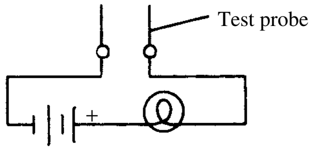

continuity tester

continuous duty The requirement of a device to sustain a 100-percent duty cycle for a prolonged period of time.

continuous-duty rating A maximum current, voltage, or power rating for equipment operated for extended periods at a 100-percent duty cycle.

continuous load A load that requires a continuous feed for a prolonged period of time.

continuous memory See NONVOLATILE MEMORY.

continuous-path motion In robotics, machine movement that occurs in a smooth fashion, rather than in discrete steps. Allows precise positioning of a mechanical arm or gripper.

continuous power The maximum sine-wave power that an amplifier can deliver for 30 seconds.

continuous recorder An instrument that provides an uninterrupted recording.

continuous recording A record made on a continuous sheet or tape, instead of on separate sheets or tapes. An example is a continuous-playing tape used for repeated public announcements.

continuous spectrum **1.** The range of all electromagnetic frequencies between a specified lower limit f_1 and a specified upper limit f_2. **2.** A range of electromagnetic frequencies that exhibits similar behavior between its lower and upper limits.

continuous stationery Also called *fanfold paper.* The pack of paper a line printer uses. It consists of sheets connected by perforated or tear-off edges, folded in accordion fashion. It usually has tear-off perforated strips along either side to facilitate feed through the printer mechanism.

continuous variable A variable that can attain any value within a specific range of values. An example is a frequency within the 75- to 80-meter amateur radio band, from 3.5 to 4.0 MHz.

continuous wave Abbreviation, CW. **1.** A periodic wave, such as a radio-frequency (RF) carrier, that is not interrupted at any point between its normal start and termination, and that is unmodulated. **2.** An RF carrier that is interrupted digitally with a keying device according to some code (such as Morse), for the purpose of conveying information.

continuous-wave laser See CW LASER.

continuous-wave radar See CW RADAR.

contour A control on an audio reproduction system that increases the base and treble amplitudes at low levels to compensate for the ear's natural losses in these ranges. Alternatively, this control can attenuate signals in the 3-kHz region, where the human ear is most sensitive.

contours of equal loudness See AUDIBILITY CURVES.

CONTRAN A computer language that requires no compiler, or translating, interface between the operator and the machine. The programming is done in a language similar to machine language.

contrast **1.** In a video image, the degree to which adjacent areas of a picture are differentiated. Insufficient contrast makes for a "flat" picture; ex-

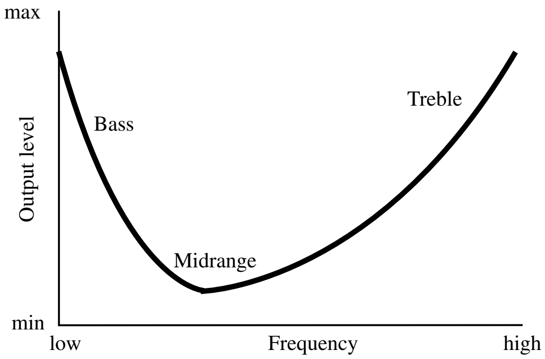

contour

cessive contrast, a "hard" picture. **2.** In optical character recognition, the degree to which a character is distinguishable from its background.

contrast control A potentiometer for adjusting the gain of the video in a television receiver or cathode-ray-tube (CRT) computer display and, accordingly, the image contrast.

contrast range In an image or pattern, the brightness range from the lightest to the darkest parts.

contrast ratio In a video image, the ratio of maximum to minimum luminance.

control **1.** An adjustable component, such as a rheostat, potentiometer, variable capacitor, or variable inductor, that allows some quantity to be varied at will. **2.** A test or experiment conducted simultaneously with another similar test conducted under conditions lacking the factor under consideration. Thus, if 100 resistors coated with a special varnish are tested at 120°F, 100 identical unvarnished resistors could be tested (as a control) under the same conditions; in this way, the effect of the varnish would be ascertainable. **3.** As a computer function, understanding and implementing instructions or carrying out tasks, according to specific conditions.

control ampere-turns The ampere-turns of the control winding in a magnetic amplifier.

control block A storage block for control information in a computer.

control bus In a digital computer, the electrical conductors linking the central-processing-unit (CPU) control register to the memory circuits.

control card A card that provides control information for a computer.

control character A character (bit group) used to start the control of a peripheral.

control characteristic A representation (such as a collector-current versus collector-voltage curve) depicting the extent to which the value of one quantity affects or controls the value of another.

control circuit **1.** A circuit in which one signal or process is made to control another signal or process. **2.** In a digital computer, a circuit that handles and interprets instructions and commands, particularly in the arithmetic and logic unit (ALU).

control computer A computer that receives signals concerning the parameters in some process, and responds with signals that control those parameters.

control counter See CONTROL REGISTER.

control data **1.** In a computer record having a key, information used to put the records in some sequence. **2.** Information affecting a routine's selection or modification.

control electrode An electrode to which an input signal can be applied to control an output signal. Common examples are the base of a bipolar transistor, the gate of a field-effect transistor, and the inputs of a logic gate.

control field **1.** In direct-current generators of the amplifying type, an auxiliary field winding used for feedback and regulation, in contrast to the self-excited field winding (which is the conventional field winding of the generator). **2.** A computer record field containing control data.

control flux In an amplidyne, magnetic flux generated by current flowing through the control winding.

control grid See GRID, **1**.

control-grid bias The negative dc voltage applied between ground and the control grid of a vacuum tube to establish the operating point.

control language Within the operating system of a computer, the command set that the operator or programmer uses to control the running of a program or the operation of peripherals. Also called *job control language* or *system control language.*

control language interpreter See CONTROL LANGUAGE and INTERPRETER.

controlled avalanche diode Also called *avalanche diode* or *Zener diode.* A diode that has a well-defined avalanche voltage. Used primarily for voltage regulation in power supplies.

controlled-carrier modulation See QUIESCENT CARRIER OPERATION.

controlled-carrier transmission See QUIESCENT CARRIER OPERATION.

controlled rectifier A rectifier whose dc output can be varied by adjusting the voltage or phase of a signal applied to the control element. See SILICON-CONTROLLED RECTIFIER.

controller **1.** The control signal of an electronic control (or servo), system. **2.** A device, such as a specialized variable resistor, used to adjust current or voltage. **3.** A computer that oversees and controls the operation of a robot or fleet of robots.

controller function The control of the movements of a servo system.

controlling file A computer storage area encompassing several complete magnetic disk cylinders; its size can be changed to accommodate a number of files.

control loop See CONTROL TAPE.

control mark See TAPE MARK.

control panel **1.** An accessible surface on which are mounted switches, buttons, potentiometers, meters, digital indicators, monitoring devices, and other apparatus essential to regulating and supervising an electronic system. **2.** The console that a computer operator or programmer uses to communicate with the central processing unit (CPU).

control plate The metallic plate or disk that serves as the antenna of a CAPACITANCE RELAY or TOUCHPLATE RELAY.

control program A program that arranges computer-operation programs in a certain order. Puts information in the computer memory for later use.

control rectifier A semiconductor diode device, used for the purpose of switching large currents. A small control signal can provide switching of high-power devices.

control register In a computer, the register that stores the address of the next instruction in the program being run.

control sequence The order in which instructions are executed in a digital computer.

control stack In a computer system, a unit of hardware having storage locations and used to perform arithmetic, assist in allocating memory to programs, and to control internal processes.

control statement In a programming language, an instruction that causes some action to be taken, as specified by a condition; it is also applicable to source program statements that affect the compiler's operation without modifying the machine code.

control tape Punched paper or plastic tape in the form of a closed loop and used to control printing devices. Also called *control loop.*

control total For a file or record group, a total derived during an operation; it is used to verify that all the records have been processed similarly.

control transfer The situation in which the control unit of a digital computer leaves the main sequence of instructions and takes its next instruction from an out-of-sequence address.

control transfer instruction See BRANCH INSTRUCTION.

control-voltage winding In a servomotor, the winding that receives a varying voltage of a phase different from that applied to the fixed-voltage windings.

control winding In a magnetic amplifier, the winding that conducts the control-signal current.

control word A word (a bit group) stored in a computer memory and used for a control function.

convection The flow of a gas or liquid that results in the transfer of heat from one location to another.

convection cooling The removal of excess heat from a component, such as a power vacuum tube or transistor, via upward movement of surrounding air that has been heated by the component.

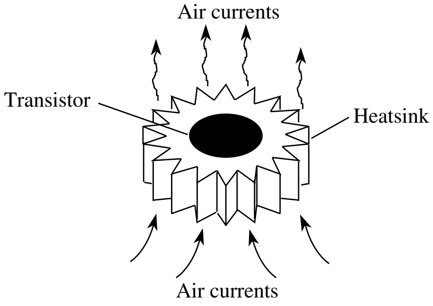

convection cooling

convection current **1.** The motion of current carriers or a charge across the surface of a conductor or dielectric. **2.** Air currents rising above a heat source or heated body.

convective discharge The continuous high-voltage current discharge across a spark gap.

convectron A device that indicates the angle, with respect to the vertical, based on convection cooling of a straight wire. The temperature difference is greatest when the angle is 0 degrees (the wire is vertical); the temperature difference decreases as the angle increases, reaching a minimum at 90 degrees (when the wire is horizontal).

convenience outlet **1.** In North America, a wall outlet providing a nominal 117 volts alternating current (ac) at 60 Hz for common household appliances. **2.** An outlet in a laboratory that provides power for a certain application.

conventional current The notion that current flows from the positive pole to the negative pole in an electric circuit. This representation is used most often by physicists. Electron flow is opposite to conventional current flow; positively charged particles, such as holes, move in the same direction as the conventional current.

convergence **1.** The eventual meeting of values or bodies at some point (sometimes at infinity, as in certain mathematical series). **2.** The intersection point of the beams from separate electron guns in a cathode-ray tube (CRT).

convergence coil One of a pair of coils used in a color television receiver to produce dynamic beam convergence (see CONVERGENCE, **2**).

convergence control In a color television receiver, a potentiometer in the high-voltage circuit for convergence adjustment (see CONVERGENCE, **2**).

convergence electrode An electrode that provides an electrostatic field for converging electron beams. Compare CONVERGENCE MAGNET.

convergence frequency The frequency of the last member of a spectrum series.

convergence magnet An assembly that provides a magnetic field to converge electron beams. Compare CONVERGENCE ELECTRODE.

convergence phase control In a three-gun color picture tube, a variable resistor or variable inductor used to adjust the phase of the dynamic convergence voltage.

convergence plane **1.** In a color picture tube, the plane in which the red, green, and blue beams all focus. **2.** In a cathode-ray tube, the plane in which the electron beam reaches its sharpest focus.

convergent series A mathematical series that approaches a specific, finite numerical value as the number of terms increases. Thus, the series 0.3 + 0.03 + 0.003 + . . . approaches a limiting value of ⅓. Compare DIVERGENT SERIES and INFINITE SERIES.

converging lens A lens having a real focus for parallel rays; generally a *convex lens.*

conversational compiler In computer operations, a compiler that, using the CONVERSATIONAL MODE of operation, shows the programmer whether or not each statement entered into the computer is valid, and whether or not to proceed with the next instruction.

conversational mode High-level computer operation or programming, in which the computer gives responses to the operator's input.

conversion **1.** The deliberate mixing of radio-frequency (RF) signals to produce signals at the sum and/or difference frequencies. **2.** The process of changing direct current (dc) to alternating current (ac). **3.** The process of changing low-voltage dc to high-voltage dc. **4.** The changing of a computer file to another format and, possibly, transferring it to a different storage medium (e.g., from tape to internal memory). **5.** The processing of a program or file written for one computer or application into a form suitable for another computer or application.

conversion efficiency In a converter (see CONVERTER, **1**), the ratio of output-signal amplitude to input-signal amplitude. For example, in a superheterodyne converter, a large intermediate-frequency (IF) output for a low radio-frequency (RF) input indicates high conversion efficiency.

conversion equipment In a computer system, an offline device for transferring data from one medium to another [e.g., a disk-to-tape converter (tape drive)]. Also called CONVERTER.

conversion exciter An exciter for transmitters, in which an output signal of a desired frequency is obtained by beating the output of a variable-frequency self-excited oscillator with the output of a fixed-frequency oscillator (such as a crystal oscillator).

conversion gain Amplification as a byproduct of conversion. See CONVERSION EFFICIENCY.

conversion loss Conversion gain of less than 1.

conversion program In computer operations, a program for data conversion (see CONVERSION, **4** and **5**).

conversion rate Also called *sampling rate.* The number of samples per second taken by an ANALOG-TO-DIGITAL CONVERTER.

conversion time In digital computer operation, the time required for the machine to read out all the digits in a coded word.

conversion transconductance See CONVERSION EFFICIENCY.

convert **1.** To perform frequency conversion (see CONVERSION, **1**). **2.** To perform voltage conversion (see CONVERSION, **2** and **3**). **3.** In computer operations, to change information from one number base to another. **4.** To perform data conversion (see CONVERSION, **4** and **5**).

converter **1.** A heterodyne mixer in which two input signals of different frequency are mixed to yield a third (output) signal of yet a different frequency. **2.** A machine for converting direct current (dc) to alternating current (ac) (e.g., a chopper converter). **3.** A transistor circuit for converting a low-voltage dc to higher-voltage dc. **4.** Conversion equipment. **5.** A circuit or device that changes analog data to digital data or vice versa.

converter amplifier See CHOPPER AMPLIFIER.

converter stage A circuit used principally to mix two signals (such as a received signal and local-oscillator signal in a superheterodyne receiver), and deliver the resultant signal.

convexo-concave Pertaining to a lens having a convex face of greater curvature than its concave face.

coolant A liquid (often water or oil) used to remove heat from an electronic component.

Coolidge X-ray tube An X-ray tube containing a heated filament (with focusing shield) and a slanting tungsten target embedded in a heavy copper anode.

cooling Maintenance of the operating temperature of an electronic component or system at a safe level. Common devices for cooling are heatsinks, circulating or forced air, and circulating liquid.

coordinate bond A covalent bond that consists of a pair of electrons supplied by only one of the atoms joined by the bond.

Coordinated Universal Time Abbreviation, UTC. Astronomical time at the Greenwich meridian (zero degrees longitude). The UTC day begins at 0000 hours and ends at 2400 hours. Based on the mean, or average, synodic (sun-based) rotational period of the earth. The earth is slightly behind UTC near June 1, and is slightly ahead near October 1.

coordinate digitizer A device or circuit that encodes a coordinate graph into digital signals for storage or transmission.

coordinate of chromaticity See CHROMATICITY COORDINATE.

coordinates A set of axes with points that can be uniquely defined or located on a line, in a plane, or in space. See CARTESIAN COORDINATES and POLAR COORDINATES.

coordinate system A mathematical means of uniquely defining or locating a point on a line, in a plane, or in space. The most common coordinates are CARTESIAN COORDINATES (also called *rectangular coordinates*), consisting of numbered lines intersecting at right angles.

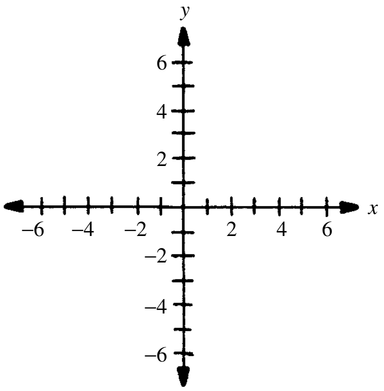

coordinate system (Cartesian)

coordination complex An ion or compound having a central (usually metallic) ion combined by coordinate bonds with a definite number of surrounding groups, ions, or molecules.

coplanar array A set of antennas that lie in the same plane, and are fed by a common transmission line.

copper Symbol, Cu. A metallic element. Atomic number, 29. Atomic weight, 63.546. An excellent conductor of electricity and heat, commonly used in the manufacture of wires and cables.

copper-clad wire Iron or steel wire plated with copper.

copper-constantan thermocouple A thermocouple consisting of a junction between wires or strips of copper and constantan. Typical output is 4.24 mV at 100°C.

copper loss Power (I^2R) loss in copper wires, cables, and/or coils.

copper-oxide diode A small diode in which the semiconductor material is copper oxide. Such diodes, widely used before the ready availability of selenium and silicon, are still occasionally found in meter-rectifier service.

copper-oxide modulator An amplitude modulator whose action is derived from the nonlinear conduction characteristic of copper-oxide diodes.

copper-oxide photocell A photoelectric cell in which the light-sensitive material is copper oxide.

copper-oxide rectifier A rectifier in which the semiconductor material is copper oxide. Rectifiers of this type are suitable for low-voltage service; they were widely used before the advent of germanium, silicon, and selenium rectifiers.

copper pyrites See CHALCOPYRITE.

copper-sulfide rectifier A rectifier in which the unilateral junction is between copper-sulfide and magnesium elements. Like the copper-oxide rectifier, the copper-sulfide unit was once widely used in low-voltage applications.

copy **1.** Also called *hard copy.* Printed or written text. **2.** In communications, a qualitative expression of the extent to which received data is intelligible (e.g., a radio operator's signal report, "You are solid (perfect) copy."). **3.** To duplicate data in a storage system, the original being in another system, or in a different location in the same system. **4.** An exact duplicate of data in any form.

copying telegraph A descriptive term for a *facsimile* system.

Corbino disk A variable resistor consisting of a semiconductor disk capable of exhibiting the CORBINO EFFECT. The disk is inserted into an adjustable magnetic field, which serves as the control medium.

Corbino effect A phenomenon similar to the HALL EFFECT, in which a current flows around a disk carrying a radial current when the disk is inserted into a magnetic field whose lines of flux are perpendicular to the disk. Compare HALL EFFECT.

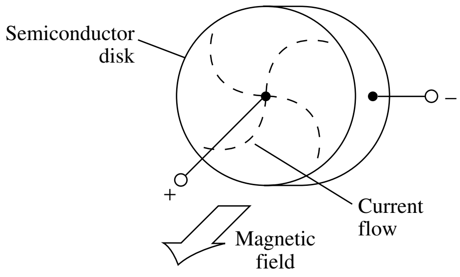

Corbino effect

cord **1.** A length of flexible, insulated cable, usually having two or three conductors. **2.** Tough, insulating string (e.g., *dial cord* or *lacing cord*).

cordless **1.** Descriptive of a plug without a flexible cord. **2.** Pertaining to radio-frequency (RF) or infrared short-range links for communications and control (e.g., a *cordless telephone* set).

cordless keyboard A computer keyboard that employs an infrared (IR), very-high frequency (VHF), or ultra-high-frequency (UHF) transmitter and receiver. Commonly used with so-called Web TV systems and in presentations using a display projection system. Operates according to the same electronic scheme as a CORDLESS MOUSE.

cordless modem See WIRELESS MODEM, 3.

cordless mouse A hand-controlled computer mouse that employs an infrared (IR), a very-high frequency (VHF), or an ultra-high-frequency (UHF) transmitter and receiver. The transmitter is inside the device, and the receiver is contained either inside the computer main unit, or in a small box attached to the computer main unit by a cord. The box can be placed somewhere out of the way; for example, at the back of the desk. Then the mouse can be moved around freely. This link is effective at distances of up to 20 or 30 feet.

cordwood A type of construction in which electronic components are sandwiched perpendicularly between layers of components. So called because it looks somewhat like stacked cordwood.

cordwood module A module containing discrete components mounted perpendicularly between two parallel printed circuits.

core **1.** The body or form on which a coil or transformer is wound. Can be made of ferromagnetic or dielectric material. The properties depend on the application. **2.** CORE MEMORY.

core dump Dumping core memory content to an output peripheral. Also see DUMP.

coreless induction heater An induction heater in which the body to be heated receives energy directly from the field of the energizing coil (there is no intervening core). Compare CORE-TYPE INDUCTION HEATER.

core loss Loss of energy in a magnetic core, caused by eddy currents and hysteresis in the core material.

core memory An older memory technology, consisting of a series of small ringshaped magnetic cores, into or out of which data can be written or read by changing the magnetization of the cores.

core plane A usually flat assembly of special magnetic cores, through which pass associated current-conducting wires to provide a CORE MEMORY.

core saturation The condition in which a core of magnetic material accommodates the maximum number of magnetic lines characteristic of that material. Increasing the magnetizing force produces no additional magnetization.

core shift register A shift register that uses special magnetic cores as bistable components. See CORE MEMORY.

core storage A high-speed magnetic core storage unit. Also see CORE MEMORY and CORE PLANE.

core transformer A transformer whose coils are wound around a ferromagnetic core.

core wrapping The placing of an insulating layer over an inductor or transformer core. This minimizes the chances of short-circuiting between the windings and the core material.

core-type induction heater An induction heater in which the body to be heated is magnetically linked, by a core, to the energizing coil. Compare CORELESS INDUCTION HEATER.

corner **1.** An abrupt turn in the axis of a waveguide. **2.** The line, and the region in the vicinity thereof, at which two intersecting plane surfaces meet (e.g., the reflector screen of a CORNER-REFLECTION ANTENNA). The plane surfaces are usually perpendicular to each other. **3.** The point, and the region in the vicinity thereof, at which three intersecting plane surfaces meet. Generally, the plane surfaces are mutually perpendicular. **4.** The passband frequency limit(s) of a bandpass, band-rejection, high-pass, or low-pass filter. **5.** A sharp bend in the attenuation-versus-frequency curve of a bandpass, band-rejection, high-pass, or low-pass filter, depicting the limit(s) of the passband.

corner diffraction **1.** The bending of sound waves around a corner. **2.** The bending of radio-frequency (RF) energy around an object, when the wavelength is great, compared with the size of the object.

corner effect A rounding off of the frequency response of a filter at the corner(s) [i.e., at the limit(s) of the passband].

corner frequency See CORNER, **4**.

corner reflection The reflection of a beam of light (or of microwave energy or other short-wavelength energy) from a corner reflector, so the beam leaves the reflector in exactly the opposite direction from which it approaches. See CORNER REFLECTOR, **2**.

corner-reflection antenna A directional antenna consisting of a dipole radiator situated at the apex formed by two nonparallel, flat reflecting sheets or a single folded sheet. See CORNER REFLECTOR, **1**.

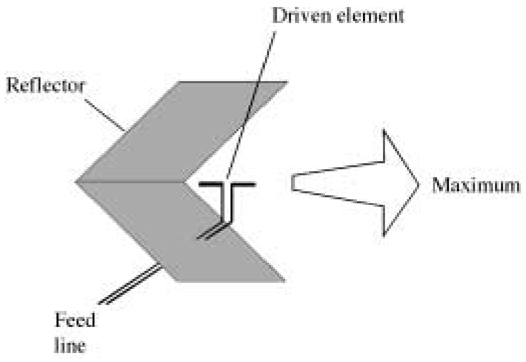

corner-reflection antenna

corner reflector **1.** An antenna with a half-wave driven element and a reflector made of wire mesh, screen, or sheet metal that resembles an open folder. The flare angle of the reflecting element is about 90 degrees. The antenna is used at ultra-

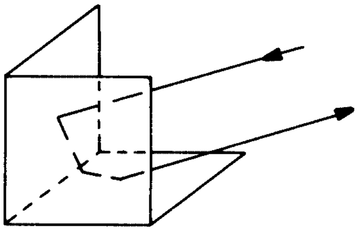

corner reflector, 2

high and microwave frequencies for television reception and satellite communications. Sometimes several half-wave dipoles are fed in phase and arranged along a common line with a single, elongated reflector. **2.** Also called *tricorner reflector*. A set of three flat metal surfaces or screens, attached together in a manner identical to the way two walls meet the floor or ceiling in a room. Such a device, if it is at least several wavelengths across, returns electromagnetic energy in exactly the same direction from which it arrives. Devices of this type are used as radar dummy targets and in optical and infrared (IR) wireless ranging systems.

corona A luminous discharge in the space surrounding a high-voltage conductor; caused by ionization of the air. The discharge constitutes a loss of energy.

corona effect The production of a luminous discharge, especially at the end of a pointed terminal, when the voltage gradient reaches a critical value.

corona failure A form of high-voltage failure, resulting from the erosion of an object (such as an electrical insulator) by corona.

corona loss Loss caused by energy dissipation through a corona. It occurs as a result of the emission of electrons from the surface of electrical conductors at high potentials, and depends on the curvature of the conductor surface, with most emission occurring from sharp points and the least from surfaces with a large radius of curvature. It is often accompanied by a blue glow and a crackling or hissing sound.

corona resistance The length of time that an insulating material can withstand a specified level of field-intensified ionization before completely breaking down.

corona shield A shield surrounding a high-voltage point to prevent corona by redistributing the electric flux.

corona starting voltage The minimum voltage between two electrodes, or on a single electrode in free space, at which corona occurs.

corona voltmeter A voltmeter used to measure the peak value of a voltage in terms of corona discharge. It consists of a metal tube in which a central wire is mounted, the parts being connected to the voltage source. The air density in the tube is varied until corona occurs.

corpuscle A tiny particle. It was the name given to the ELECTRON by some early experimenters and theorists.

correction **1.** The addition of a factor that provides greater accuracy in a measurement. **2.** A change in the calibration of an instrument to increase the accuracy.

correction factor A percentage, or numerical factor, added to or subtracted from a reading to provide a greater degree of accuracy. Often used with instruments known to be inaccurate by a certain amount.

corrective feedback Feedback that is used to correct (bring to a prescribed level) a quantity constituting the input to a system.

corrective maintenance The repair of a circuit or system after it has malfunctioned or broken down.

corrective network A network that improves the performance of the circuit into which it is inserted.

corrective stub A combination tuning-matching stub used in some antenna systems. It matches the resistive component of the antenna impedance to the characteristic impedance of a feed line, and also eliminates any reactance that might be present at the antenna feed point.

correed relay A sealed reed relay used as a high-speed switching device in communications equipment.

correlation A statistical expression or measure of the degree to which two sets of data are related. Can be given qualitatively (high-positive, low-positive, zero, low-negative, or high-negative) or quantitatively (as a number between –1 and 1). Does not necessarily imply causation.

correlation detector A detector that compares a signal of interest with a standard signal at every point, delivering an output that is proportional to the correspondence between the two signals.

correlation distance The smallest distance between two antennas that results in fading of signals under conditions of tropospheric propagation. It is used at very-high frequencies (VHF) and above, to determine the maximum range over which communications can be carried out reliably.

correlation tracking A method of target tracking in which phase relationships are used to determine positions.

correspondence The ability of a binocular machine vision system to tell when both of its optical sensors are processing an image from the same object; also, the ability of the system to keep both sensors tracking the same object.

corrosion-resistant Pertaining to materials that are treated to be immune to corrosion by the elements. Such substances are preferable for use in marine or tropical environments, where corrosion is especially severe.

corruption The altering of data or a code as a result of a program error or machine fault.

COS Abbreviation of COMPLEMENTARY-SYMMETRY CIRCUIT.

cosecant Abbreviation, csc. A trigonometric function; csc $q = c/a$, where c is the hypotenuse of a right triangle and a is the side opposite q. The cosecant is the reciprocal of sine: csc $q = 1/\sin q$.

cosecant-squared antenna A radar antenna that radiates a COSECANT-SQUARED BEAM.

cosecant-squared beam A radar beam whose intensity varies directly with the square of the cosecant of the angle of elevation.

cosech Abbreviation of HYPERBOLIC COSECANT. Also abbreviated as csch.

cosh Abbreviation of HYPERBOLIC COSINE.

cosine Abbreviation, cos. A trigonometric function; cos $q = b/c$, where b is the side adjacent to q and c is the hypotenuse of the right triangle.

cosine law The brightness in any direction from a perfectly diffusing surface is proportional to the cosine of the angle between the direction vector and a vector perpendicular to the surface.

cosine wave A periodic wave that follows the cosine of the phase angle. It has a shape identical with a SINE WAVE, but differs by 90 degrees of phase.

cosine yoke A magnetic-deflection yoke that has nonuniform windings for improved focus at the edges of a television picture. Also called *anastigmatic yoke* and *full-focus yoke.*

cosmic noise Radio noise produced by signals from extraterrestrial space.

cosmic rays Extremely penetrating rays consisting of streams of atomic nuclei entering the earth's atmosphere from outer space.

COS/MOS IC An integrated circuit (IC), such as an operational amplifier, utilizing metal-oxide-semiconductor (MOS) field-effect transistors in a complementary-symmetry (COS) arrangement.

cost analysis In a commercial or industrial organization, ascertaining the expense associated with a service, process, or job.

cot Abbreviation of COTANGENT.

cotangent Abbreviation, cot. A trigonometric function; cot $q = b/a$, where a is the side adjacent to q and b is the side opposite q (in a right triangle). Cotangent is the reciprocal of tangent: cot $q = 1/\tan q$.

coth Abbreviation of HYPERBOLIC COTANGENT.

Cotton-Mouton effect See KERR MAGNETO-OPTICAL EFFECT.

Cottrell process Dust precipitated by high voltage. Dust in the air is made to flow through a grounded metal chamber that contains a wire maintained at high voltage. The dust particles become charged and adhere to the chamber walls, from which they are later collected.

coul-cell A coulometer of the electrolytic-cell type.

coulomb (Charles Augustin Coulomb, 1736–1806). Abbreviation, C. The unit of electrical charge quantity, equal to the charge contained in 6.24×10^{18} electrons. A current of one ampere (1 A) represents 1 coulomb per second (C/s).

Coulomb's law The force between two electrically charged objects is directly proportional to the product of the charge quantities in coulombs, and inversely proportional to the square of the distance between the charge centers. This force is an attraction for opposite charges, and a repulsion for similar charges.

coulometer An instrument that measures electrical charge quantity in coulombs. A typical version keeps a cumulative count of coulombs (ampere-seconds) by integrating current, with respect to time. Also called *coulombmeter.*

Coulter counter See CELL COUNTER.

count **1.** The number of pulses tallied by a counting system in a given period of time. **2.** A single response by a radioactivity counter. **3.** A record of the number of times an instruction or subroutine in a computer program is executed (by increasing the value of a variable by one, as stated in a FOR-NEXT loop, for example).

countdown A decreasing count of time units remaining before an event or operation occurs showing time elapsed and time remaining.

counter **1.** A circuit, such as a cascade of flip-flops, that tracks the number of pulses applied to it and usually displays the total number of pulses. **2.** A mechanism, such as an electromechanical indicator, that tracks the number of impulses applied to it and displays the total. **3.** An electronic switching circuit, such as a flip-flop or stepping circuit, that responds to sequential input pulses applied to it, giving one output pulse after receiving a certain number of input pulses.

counter- Prefix meaning "opposite to" or "contrary to." Examples: *counter EMF, counterclockwise.*

counterclockwise Abbreviation, ccw. Pertaining to rotational motion in a sense opposite that of a typical analog clock. Movement is to the left at the top of the rotational circle, and to the right at the bottom of the circle. Compare CLOCKWISE.

counterclockwise-polarized wave An elliptically polarized electromagnetic wave whose electric-intensity vector rotates counterclockwise as observed from the point of propagation. Compare CLOCKWISE-POLARIZED WAVE.

counter efficiency The sensitivity of a radiation counter or scintillation counter to incident X-rays or gamma rays.

counterelectromotive cell A cell used to counteract a direct-current voltage.

counter EMF See BACK VOLTAGE and KICKBACK.

counter-meter A radioactivity instrument, such as a Geiger counter, that indicates the number of radioactive particles per unit time.

counterpoise A means of obtaining a radio-frequency (RF) ground by using a grid of wires or tubing in a plane parallel to the earth's surface or to average terrain. The radius of the grid is usually at least 0.25 wavelength, but might be smaller if the feed-point impedance of the antenna is very high.

counterpoise ground system A counterpoise with a radius such that resonance is obtained with a quarter-wavelength antenna operated at a height of more than 0.25 wavelength above actual ground. Usually such a system consists of three or four radials measuring 0.25 wavelength each, and extending outward from the base of the antenna nearly parallel to the average terrain.

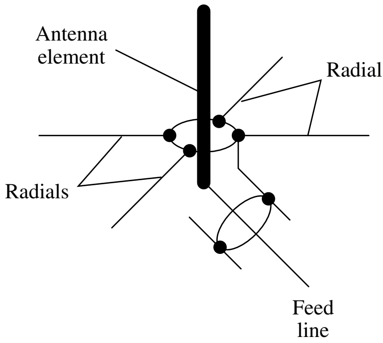

counterpoise ground system

counter tube A tube, such as the Geiger-Meuller tube, in which a penetrating radioactive particle ionizes a gas and produces an output pulse.

counter voltage See BACK VOLTAGE and KICKBACK.

counting-type frequency meter A direct-reading analog or digital frequency meter that indicates the number of pulses (or cycles) per second applied to it.

count-remaining technique See COMPLEMENT-SETTING TECHNIQUE.

couple Two dissimilar metals in contact with each other or immersed in an electrolyte.

coupled circuits Circuits between which energy is transferred electrostatically, electromagnetically, by some combination of the two, or by direct connection.

coupled impedance The impedance that a circuit "sees" when it is coupled to another circuit. Thus, when the secondary of a transformer is terminated with an impedance, the primary "sees" a combination of that impedance and its own.

coupler A device for transferring energy between two circuits and using capacitive coupling, direct coupling, inductive coupling, or some combination of these.

coupling **1.** Also called *electrostatic coupling* or *capacitive coupling.* The linking of two circuits or devices by electric flux. **2.** Also called *magnetic coupling* or *inductive coupling.* The linking of two circuits or devices by magnetic flux. **3.** Also called *direct coupling.* The linking of two circuits or devices by direct connection. **4.** Also called *resistive coupling.* The linking of two circuits or devices through a resistance. **5.** Also called *optical coupling.* The linking of two circuits or devices through an optoisolator.

coupling aperture A hole in a waveguide that is used to transmit energy to the waveguide, or receiving energy from outside the waveguide.

coupling capacitor A capacitor used to conduct ac energy from one circuit to another. Also see CAPACITIVE COUPLING.

coupling coefficient See COEFFICIENT OF COUPLING.

coupling diode A semiconductor diode connected between the stages of a direct-coupled amplifier. When the diode is connected in the correct polarity, it acts as a high resistance between stages when there is no signal, and does not pass the high dc operating voltage from one stage to the next. When a signal is present, the diode resistance decreases, and the signal gets through.

coupling efficiency A measure of the effectiveness of a coupling system (i.e., the degree to which it delivers an undistorted signal of correct amplitude and phase).

coupling loop **1.** A single turn of a coupling transformer. **2.** A small loop inserted into a waveguide to introduce microwave energy.

coupling probe A usually short, straight wire or pin protruding into a waveguide to electrostatically introduce microwave energy into the waveguide. It acts like a miniature whip antenna.

coupling transformer A transformer used primarily to transfer alternating-current (ac) energy electromagnetically into or out of a circuit.

covalent binding forces In a crystal, the binding forces resulting from the sharing of valence electrons by neighboring atoms.

covalent bonding The binding together of the atoms of a material as a result of shared electrons or holes.

coverage **1.** The area within which a broadcast or communication station can be reliably heard. **2.** The shielding effectiveness of a coaxial cable.

coversed sine Abbreviation, covers. The trigonometric functional equivalent of the *versed sine* of the complement of an angle [i.e., the difference between the sine of an angle and unity (1)]. Thus, covers $q = 1 - \sin q$.

CP Abbreviation of *chemically pure.*

cp **1.** Abbreviation of CANDLE POWER. **2.** Abbreviation of *central processor.*

cps **1.** Abbreviation of CYCLES PER SECOND. Cycles per second, to denote ac frequency, has been supplanted by HERTZ. **2.** Abbreviation of *characters per second.*

CPU Abbreviation of CENTRAL PROCESSING UNIT.

CQ A general call signal used in radio communication, especially by amateur stations, to invite a response from any station that hears it.

Cr Symbol for CHROMIUM.

cracked-carbon resistor A high-stability resistor in which the resistance material is particulate carbon.

cracker A hacker with malicious intent (also see HACKER). Such a person attempts to gain access to computer systems or databases in order to steal something or inflict damage. Examples include theft, erasure, or mutilation of data; fraudulent debiting of bank accounts; alteration of credit information; and identity theft.

cradle guard See GUARD WIRE.

cradlephone A telephone set in which the microphone and earphone are mounted on opposite ends of a handle. This handle, called the *receiver*, rests on the crossmember of a stand connected to a base containing the dial and ringing circuits. Also called *French phone, French telephone*, and *handset.*

crash **1.** A condition in which a computer or network server becomes inoperative because of a software or memory-management problem. **2.** In a computer hard disk or diskette drive, contact of the read/write head with the surface of a disk or platter. Usually, it is the result of excessive physical vibration or shock.

crate A foundation unit into which modules are plugged to establish a circuit.

crawl **1.** See CREEPING COMPONENT. **2.** The credits (names of staff and their contribution to content) superimposed and moving on a television picture at the end of a program.

crazing The formation of tiny cracks in materials, particularly in such dielectrics as plastic and ceramic.

creep See COLD FLOW.

creepage Current leakage across the surface of a dielectric.

creeping component A quantity, such as current, voltage, or frequency, that slowly changes in value with time.

crest factor See AMPLITUDE FACTOR.

crest value The maximum amplitude of a composite current or voltage.

crest voltmeter A peak-reading (or sometimes peak-responsive) voltmeter.

crippled mode The mode of operation for a computer or other hardware in which some of the components are inoperable. Compare GRACEFUL DEGRADATION.

crisscross neutralization See CROSS-CONNECTED NEUTRALIZATION.

crisscross rectifier circuit A conventional bridge rectifier circuit configured in such a way that two of the diodes are connected in crisscross fashion between the input and output terminals.

critical angle **1.** In radio communications, an angle of departure that a transmitted electromagnetic field subtends, with respect to the horizon at the transmitting (TX) point, below which the ionosphere will reliably return the signal to the earth, and above which the ionosphere will not reliably return the signal. This angle (shown by the double arc marked *X* in the drawing) depends on the frequency of the transmitted electromagnetic wave, and also on ionospheric conditions. **2.** For an electromagnetic wave or ray approaching a boundary at which the index of refraction abruptly decreases, the minimum angle of incidence (relative to a line perpendicular to a plane tangent to the boundary) at which the energy is totally reflected.

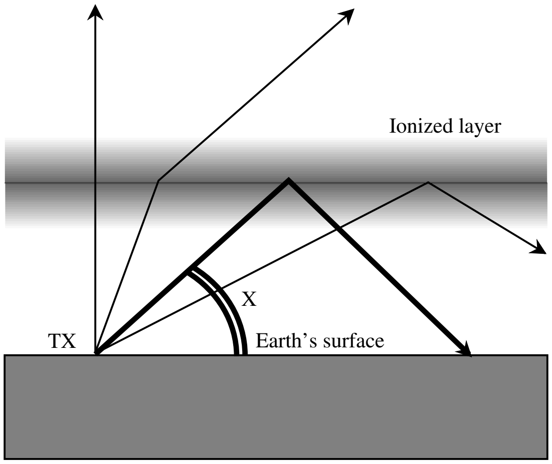

critical angle, 1

critical characteristic A parameter that has a disproportionate effect on other variables. A small change in this characteristic can result in a large change in the operating conditions of a circuit or system.

critical component A component or part that is especially important in the operation of a circuit or system.

critical coupling The value of coupling at which maximum power transfer occurs. Increasing the extent of coupling beyond the critical value decreases power transfer.

critical damping The value of damping that yields the fastest transient response without overshoot.

critical dimension The cross-sectional size of a waveguide that determines its minimum usable frequency.

critical failure A component or circuit failure that results in shutdown of a system, or a malfunction that results in improper operation.

critical field The smallest magnetic-field intensity in a magnetron that keeps an electron, emitted from the cathode, from reaching the anode.

critical frequency For a particular layer of the ionosphere, the high frequency at which a vertically propagated wave is no longer reflected back to the earth.

critical inductance In a choke-input power-supply filter, the minimum inductance that will maintain a steady value of average load current.

critical potential The potential difference required for an electron to excite or ionize an atom with which it collides.

critical voltage The voltage at which a gas ionizes.

critical wavelength The wavelength that corresponds to CRITICAL FREQUENCY.

CRO Abbreviation of *cathode-ray oscilloscope.*

Crookes dark space In a glow-discharge tube, the narrow dark space next to the cathode. Also see CROOKES TUBE.

Crookes tube A glow-discharge tube containing an anode, cathode, and a small amount of gas under low pressure.

cross antenna An antenna in which two (usually equal-length) horizontal radiators cross each other at right angles and are connected together to a feeder at their point of intersection. It takes its name from its horizontal-cross shape.

cross assembler A program used with one computer to translate instructions for another computer.

crossband operation **1.** Communications in which two frequency bands are used. Station X, for example, might transmit on frequency f_A in band A and receive on frequency f_B in band B; station Y would then transmit on f_B and receive on f_A. **2.** In satellite communications, the use of two frequency bands to facilitate full-duplex operation and to allow the satellite transponder to effectively function. The transponder receives signals from the earth within a specific frequency band, and converts this entire band of signals to a set of signals that occupies an equal amount of spectrum space on another frequency band. The converted signals are then retransmitted back to earth.

crossbar switch A three-dimensional array of switch contacts in which a magnetic selector chooses individual contacts, according to their coordinates in the matrix.

cross bearings A method of radionavigation, in which directional readings are taken from a receiving station (such as a ship or aircraft) for two fixed transmitting stations whose locations are known. Lines are drawn on a map from the transmitting stations, in directions 180 degrees opposite the bearings obtained from the receiving station. The intersection point of these lines is the location of the receiving station.

cross beat A spurious frequency arising from CROSS MODULATION.

cross-check To compare the result of a calculation or computer routine with the result obtained by a different method.

cross color In the chrominance channel of a color television receiver, crosstalk interference caused by monochrome signals.

cross-connected neutralization Neutralization of a push-pull amplifier by feedback through two capacitors—each connected from the output circuit of one transistor to the input circuit of the other.

cross-coupled multivibrator A multivibrator circuit in which feedback is provided by a coupling capacitor between the output of the second stage and the input of the first stage; the stages are forward-coupled by a capacitor of the same value.

cross coupling **1.** The state of being cross-coupled (see, for example, CROSS-COUPLED MULTIVIBRATOR). **2.** Undesired coupling between two circuits.

cross current A current that flows in the opposite direction from some other current.

crossed-pointer indicator **1.** Also called *crossed-needle meter.* A combination of two analog metering instruments in one case. Each needle has its own independently calibrated scale. A third scale corresponds to the intersection point of the needles. Commonly used in directional wattmeters that simultaneously show forward power, reflected power, and standing-wave ratio (SWR). **2.** A two-pointer meter used in aircraft to show the position of the aircraft, relative to the glide path.

crossed-wire thermoelement Two wires or strips of dissimilar metals joined or twisted at a point that constitutes a thermoelectric junction. In usual operation, a high-frequency current is passed through one wire, and a proportional direct-current (dc) voltage, generated by thermoelectric action, appears at the other wire.

cross flux The magnetic flux component that is perpendicular to the flux produced by field magnets.

cross-hair pattern A television test pattern consisting of a single vertical line and a single horizontal line, which form a simple cross. The pattern resembles the cross hairs of an optical instrument.

crosshatch generator A modulated radio-frequency (RF) signal generator that produces a crosshatch pattern on a picture-tube screen.

crosshatch pattern A grid of horizontal and vertical lines produced on a picture-tube screen by a cross-hatch generator. It is used in checking horizontal and vertical linearity.

cross modulation **1.** A type of radio-frequency interference (RFI) between two strong stations that are close in frequency. The desired carrier is modulated by the interfering signal. **2.** The production of signals by rectifier junctions in pipes and wiring near a radio receiver. These objects pick up waves and deliver energy at a different frequency, which finds its way into the receiver. Also called *external cross modulation.* **3.** The interaction between signals of different frequency when they magnetize a core of nonlinear magnetic material. Also see CROSSTALK.

cross-modulation factor An expression of the amount of cross modulation (or crosstalk) present in a particular instance. It is equal to M_1/M_2, where M_1 is the modulation percentage that a modulated wave produces in a superimposed unmodulated wave, and M_2 is the modulation percentage of the modulated wave.

cross-neutralized circuit See CROSS-CONNECTED NEUTRALIZATION.

crossover **1.** In a circuit diagram, a point at which lines representing wires intersect, but are not connected. **2.** In a characteristic curve, point at which the plot crosses an axis or operating point. **3.** See CROSSOVER NETWORK.

crossover distortion Distortion of a characteristic at a crossover point (see CROSSOVER, **2**); for example, a bend in the curve where the plot of a waveform passes through zero.

crossover frequency The frequency at which a crossover network delivers equal power to the two circuits it supplies.

crossover network Following final amplification in a sound-reproduction system, an outboard filter circuit that facilitates delivery of the low and high audio frequency (AF) components to the correct speakers.

crossover point See CROSSOVER, **2**.

crossover S-curve The S-shaped image obtained on an oscilloscope screen during sweep-generator alignment of a frequency-modulation (FM) detector. In correct alignment, the exact center of the S-curve (the crossover point) coincides with the zero point on the screen.

cross product Also called *vector product.* For vectors **A** and **B** having lengths A and B, respectively, and subtending an angle θ relative to each other, the cross product $\mathbf{A} \times \mathbf{B}$ points in a direction perpendicular to the plane containing both **A** and **B**. The length of $\mathbf{A} \times \mathbf{B}$ is equal to AB sin θ.

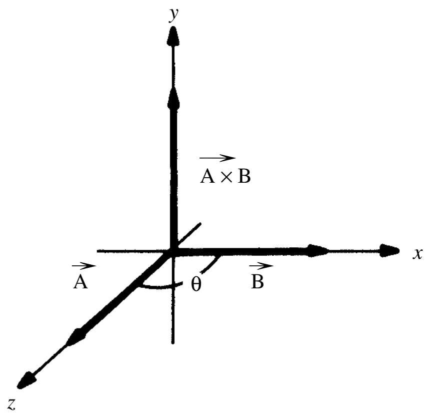

cross product

cross-sectional area **1.** The surface area of a face of a conductor after cutting through it at a right angle. Specified in square inches, square millimeters, or circular mils. **2.** The total of the cross-sectional areas of all the wires in a stranded conductor.

cross-sectional testing In quality assurance and quality control (QA/QC), a method of checking a large lot of units or components. Instead of testing every device, a fraction of the devices is tested. The sampling is taken uniformly from the group (e.g., every fifth unit).

crosstalk Undesired transfer of signals between or among telephone lines, data lines, or system components. In computer operations, this effect places a practical limit on the lengths of parallel data cables.

crosstalk coupling Undesired coupling between circuits, caused by crosstalk.

crosstalk factor See CROSS-MODULATION FACTOR.

crosstalk level The amplitude of crosstalk, usually expressed in decibels above a reference level.

crosstalk loss Loss of energy caused by crosstalk.

crowbar An action producing a high overload on a circuit protection device.

crowfoot **1.** A pattern formed by the cracking or crazing of solid plastics of solidified encapsulating compounds, so called from its resemblance to a bird's footprint. **2.** In a gravity battery cell, the zinc electrode, so called from its resemblance to a bird's foot.

CRT Abbreviation of CATHODE-RAY TUBE.

crud **1.** Broadband electrical noise, originating inside and/or outside a system. **2.** Undesired signals that interfere with a desired signal.

cryogenic device A device that exhibits unique electrical characteristics (such as superconductivity) at extremely low temperatures.

cryogenic motor A motor designed for operation at extremely low temperatures.

cryoelectronics The study of the behavior of electronic devices, circuits, and systems at extremely low temperatures.

cryogenics The branch of physics dealing with the behavior of matter at temperatures approaching absolute zero. Also concerned with methods of obtaining such temperatures in controlled environments.

cryosar A semiconductor switch utilizing low-temperature avalanche breakdown.

cryoscope An instrument used to determine freezing point.

cryostat A chamber for maintaining a very low temperature for cryogenic operations. Also see CRYOGENICS.

cryotron A switching device consisting essentially of a straight tantalum wire, around which a single-layer control coil is wound. The magnetic field generated by control current flowing through the coil causes the tantalum wire to become superconductive at a temperature of approximately 4.4 degrees K.

cryotronics Low-temperature electronics, concerned with such phenomena as superconductivity. The term is an acronym from cryogenics and electronics. Also see CRYOGENICS.

cryptanalysis The breaking of ciphers.

crypto- A prefix added to words, that implies encoding for the purpose of changing or hiding the meaning of a message or signal.

cryptography The creating and writing of ciphers.

cryptology The art and science of creating, writing, unscrambling, and breaking ciphers.

crystal **1.** A material distinguished by the arrangement of its atoms into a redundant pattern called a *lattice* that presents, in fragments of various sizes, a characteristic polyhedral shape. Common shapes include cubes, parallelepipeds, and hexagonal prisms. **2.** A fragment of material as defined in (1). **3.** A plate or bar cut from a piece of piezoelectric material.

crystal amplifier **1.** A semiconductor diode circuit using carrier storage. Transistor action and, accordingly, pulse amplification is obtained by alternately making one electrode of the diode an emitter or collector. **2.** Archaic term for TRANSISTOR.

crystal audio receiver An audio radar receiver, consisting of a crystal detector and audio-amplifier stages.

crystal axes The imaginary lines traversing a piezoelectric crystal, along which (or perpendicular to which) plates are cut for oscillators, resonators, or transducers.

crystal calibrator A crystal oscillator used to generate harmonic checkpoints for frequency calibration. Common fundamental calibrator frequencies are 100 kHz and 1 MHz.

crystal capacitor See VARACTOR.

crystal control The control of the operating frequency of a circuit by means of a piezoelectric crystal.

crystal-controlled receiver A superheterodyne radio receiver whose local oscillator is crystal controlled.

crystal-controlled transmitter A radio transmitter whose master oscillator is crystal controlled.

crystal counter A device for counting the frequency of subatomic particles, based on their ability to change the conductivity of a crystal. The particles can be photons, electrons, protons, neutrons, or the nuclei of atoms.

crystal current Current flowing through a crystal; specifically, the radio-frequency (RF) current flowing through a quartz plate in a crystal-controlled oscillator.

crystal cuts The classification of piezoelectric plates according to the angle at which they were cut from a quartz crystal. Common cut designations are AT, BT, CT, DT, X, Y, and Z. Various cuts afford such complementary factors as frequency, temperature, and thickness. Also see CRYSTAL AXES.

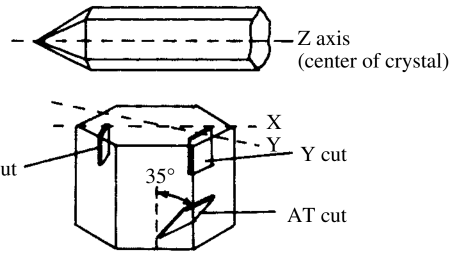

crystal cuts

crystal detector A rudimentary form of semiconductor diode consisting of a mounted lump of mineral (the crystal) in contact with a springy wire ("cat's whisker"). The point of the wire is moved to various points of contact on the crystal surface until the most-sensitive rectifying spot is found.

crystal diffraction The tendency of electromagnetic waves to be scattered when passing through a crystal material.

crystal diode Archaic term for SEMICONDUCTOR DIODE. Also see GALLIUM-ARSENIDE DIODE; GERMANIUM DIODE; JUNCTION DIODE; LASER DIODE; POINT-CONTACT DIODE; SELENIUM DIODE; SIGNAL DIODE; SILICON DIODE.

crystal earphone An earphone in which the transducer is a piezoelectric crystal. Electrical impulses applied to the crystal vary its shape and cause a vibration that is transmitted to a diaphragm; this in turn produces corresponding sound waves.

crystal filter See CRYSTAL RESONATOR.

crystal headphone See CRYSTAL EARPHONE.

crystal holder A fixture specially designed to hold a piezoelectric crystal; it ensures minimum distortion of crystal dimensions and minimum residual capacitance, inductance, and resistance.

crystal imperfection A flaw in the lattice structure of a crystal.

crystal lattice The orderly, redundant pattern of atoms and molecules within a crystalline material; it is a characteristic of a given material.

crystal-lattice filter A crystal resonator in which piezoelectric crystals are used to give a desired shape to the filter response curve.

crystalline material A material exhibiting the characteristic properties of a crystal (see CRYSTAL, **1**).

crystallogram An X-ray photograph or other record of crystal structure.

crystallography The science dealing with crystals and their properties (see CRYSTAL, **1**).

crystal loudspeaker A loudspeaker whose transducer is a piezoelectric crystal. Electrical impulses applied to the crystal vary its shape and cause vibrations that are transmitted to a diaphragm or cone, which produces corresponding sound waves.

crystal meter A rectifier-type ac meter using a semiconductor diode in series with a dc milliammeter or microammeter.

crystal microphone A microphone whose transducer is a natural or synthetic piezoelectric crystal. Sound waves striking the crystal (directly or via a diaphragm) vary its shape, making it produce an audio-frequency (AF) output voltage.

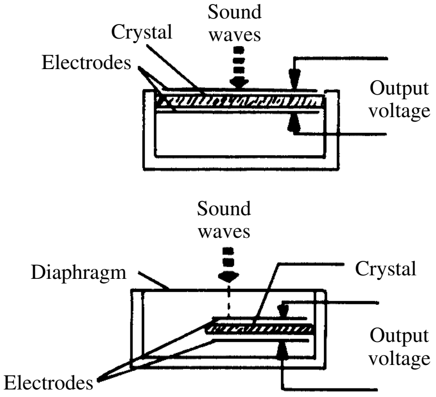

crystal microphones

crystal mixer A mixer (converter) circuit utilizing the nonlinearity of a semiconductor diode to mix signals.

crystal operation **1.** The characteristics of a piezoelectric crystal in a particular circuit. **2.** Crystal frequency control.

crystal oscillator An oscillator whose operating frequency is determined by the dimensions of an oscillating piezoelectric quartz-crystal plate. Compare SELF-EXCITED OSCILLATOR.

crystal oven A constant-temperature chamber for stabilizing the frequency of a quartz crystal by maintaining its operating temperature at a fixed point.

crystal photocell A photoelectric cell in which the light-sensitive material is a crystalline substance, such as germanium, selenium, silicon, etc.

crystal pickup A phonograph pickup whose transducer is a natural or synthetic piezoelectric crystal. The crystal is attached (either directly or through a mechanical linkage) to a stylus, whose movement in the disk groove varies the shape of the crystal. The resultant vibration generates a corresponding audio-frequency (AF) output voltage across the crystal.

crystal probe A radio-frequency (RF) probe, whose rectifying element is a semiconductor diode.

crystal pulling **1.** The extraction of a single crystal from a molten mass of crystalline material. Single crystals are used for high-quality semiconductor devices. Also see CZOCHRALSKI METHOD, SINGLE CRYSTAL, and SINGLE-CRYSTAL MATERIAL. **2.** The use of an inductor or capacitor in a crystal-controlled radio-frequency (RF) oscillator circuit to allow adjustment of the frequency over a small range.

crystal receiver See CRYSTAL SET.

crystal rectifier **1.** A semiconductor diode used for the purpose of rectifying alternating current (dc), usually in a power supply.

crystal resistor A temperature-sensitive resistor made from silicon, and exhibiting a positive temperature coefficient of resistance.

crystal resonator A highly selective resonant circuit in which the center frequency is the resonant frequency of a piezoelectric quartz-crystal plate.

crystal sensor See CRYSTAL TRANSDUCER.

crystal set A simple radio receiver that uses a tuned circuit, semiconductor-diode detector, and earphones.

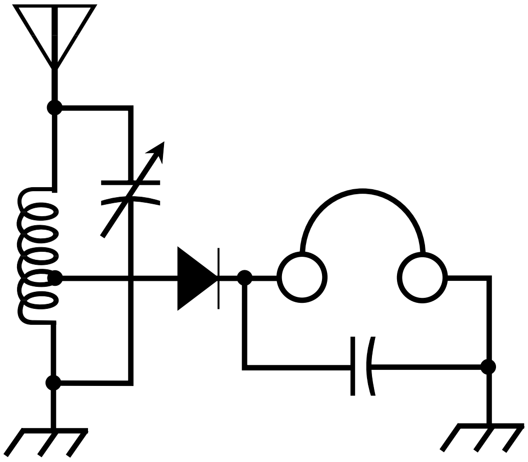

crystal set

crystal slab See QUARTZ BAR.

crystal socket **1.** A low-capacitance, low-loss socket for a piezoelectric crystal. **2.** A socket for a semiconductor diode.

crystal tester **1.** An oscillator used to check quartz crystals. Most such units check only the crystal's ability to oscillate; more elaborate ones also check crystal current, frequency, temperature coefficient, activity, filter action, etc. **2.** An instrument for checking the electrical characteristics of semiconductor diodes. **3.** An instrument for checking the performance of piezoelectric ceramics.

crystal tetrode A transistor having four elements: emitter, collector, and two bases.

crystal transducer A transducer using a piezoelectric crystal as the sensitive element. Examples: crystal earphone, crystal loudspeaker, crystal microphone, and crystal pickup.

crystal triode See TRANSISTOR.

Cs Symbol for CESIUM.

CS Abbreviation of COMPLEMENTARY SYMMETRY. Also COS.

C_S **1.** Symbol for *standard capacitance.* **2.** Symbol for *source capacitance.*

csc Abbreviation of COSECANT.

C scan See C DISPLAY.

csch Abbreviation of HYPERBOLIC COSECANT.

C scope A cathode-ray tube used in radar to provide a C DISPLAY.

CT-cut crystal A piezoelectric plate cut from a quartz crystal at an angle of rotation around the X-axis of +38°. Such a plate has a zero temperature coefficient of frequency at 25°C. Also see CRYSTAL AXES and CRYSTAL CUTS.

CTL Abbreviation of *complementary-transistor logic.*

Cu Symbol for COPPER.

cube **1.** A regular polyhedron with six identical square faces and eight vertices. At each vertex, three edges converge at mutual right angles. **2.** The third power of a number; thus the cube of n is written n^3.

cube tap An electrical adapter, in which a set of male prongs and three sets of female contacts are on the sides of a molded cube. Allows three appliances to be used with a single electrical socket.

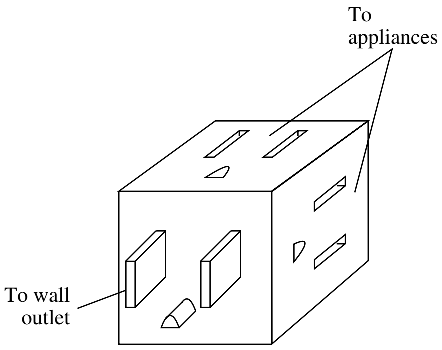

cube tap

cubical antenna An antenna in which the elements form the outline of a geometric cube or rectangular prism. The most common example is the QUAD ANTENNA.

cubical quad antenna See QUAD ANTENNA.

cubic equation A polynomial equation of the third degree. Its general form is $ax^3 + bx^2 + cx + d = 0$.

cue A condition or signal that alerts an operator, circuit or system to act in a specific manner.

cue circuit A device for transmitting cues used in program control.

cueing receiver **1.** A (usually miniature) radio receiver used to pick up cues. Example: a receiver carried by a technician, actor, or lecturer. **2.** A receiver or other pickup circuit that receives a cuing pulse, which it uses to set another circuit.

cu ft Abbreviation of *cubic foot* or *cubic feet.*

cu in Abbreviation of *cubic inch* or *cubic inches.*

cumulative error In a sum or other final value, the total error that has accumulated from the individual errors in the terms. Also called *systematic error.*

cup core A coil core that also forms a magnetic shield around the coil.

cuprous-oxide rectifier See COPPER-OXIDE RECTIFIER.

cur Abbreviation of CURRENT.

curie Abbreviation Ci. A unit of radioactivity; 1 curie is the amount of radiation from (or in equilibrium with) 1 gram of radium. Also equivalent to 3.7×10^{10} atomic breakdowns per second.

Curie point **1.** The temperature above which a ferromagnetic material loses its magnetism or becomes paramagnetic. **2.** The temperature at which the ferroelectric properties of a substance disappear.

curie temperature As a magnetized substance is heated, the lowest temperature at which magnetization is lost. It is generally measured in degrees Celsius or degrees Kelvin. For iron, this temperature is 760 degrees Celsius; for nickel, it is 356 degrees Celsius.

Curie's law For a paramagnetic substance, the ratio of the magnetization to the magnetizing force is inversely proportional to the absolute temperature.

Curie-Weiss law Above the Curie point, the susceptibility of a paramagnetic material varies inversely as the excess of temperature above the Curie point increases. This law is invalid for applications at or below the Curie point.

curium Symbol, Cm. A radioactive metallic element produced artificially. Atomic number, 96. Atomic weight, 247.

current Symbol, I or i. The movement of charge carriers, such as electrons, holes, or ions. Also see AMPERE.

current amplification **1.** An electronic process in which the instantaneous, average, or peak magnitude of a current is increased. **2.** The extent to which a current increases in a circuit; the ratio (always greater than one) of output current to input current, I_{out}/I_{in}. Also called *current gain.*

current amplifier An amplifier operated primarily to increase a signal current. Compare POWER AMPLIFIER and VOLTAGE AMPLIFIER.

current antinode See CURRENT LOOP.

current attenuation **1.** The reduction of current amplitude along a line. **2.** The extent to which a current decreases in a line or circuit; the ratio (always less than one) of output current to input current, I_{out}/I_{in}.

current balance An instrument for determining the size of the ampere. This is done by measuring the force between two current-carrying conductors.

current-balance switch A switch or relay, operated by the existence of a difference between two currents.

current-carrying capacity The maximum current (usually expressed in amperes) that a conductor or device can safely conduct.

current coil The series coil in a nonelectronic wattmeter. Compare POTENTIAL COIL.

current-controlled amplifier Abbreviation, CCA. An amplifier in which gain is controlled by means of a current applied to a control-input terminal.

current density The current (usually expressed in amperes per square centimeter) passing through a cross-sectional area of a conductor.

current drain **1.** The current supplied to a load by a generator or generator-equivalent. **2.** The current required by a device for its operation; also, the current taken by the device during standby periods.

current echo Reflected current in a transmission line that is not terminated in an impedance exactly matching its characteristic impedance.

current-fed antenna An antenna in which the transmission line is attached to the radiator at a current loop (voltage node). Compare VOLTAGE-FED ANTENNA.

current feed **1.** The delivery of power to a device or circuit at a point where current dominates. Compare VOLTAGE FEED. **2.** In an antenna, feeding it at a current maximum.

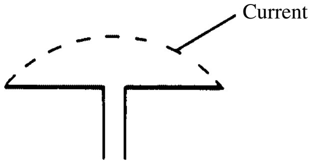

Current feed

current feedback **1.** A feedback signal consisting of current fed from the output to the input circuit of an amplifier. **2.** A system or circuit for obtaining current feedback.

current-feedback pair A two-stage, direct-coupled transistor amplifier having direct-current shunt-series feedback.

current flow Charge carriers passing through a solid, liquid, gas, or vacuum. Also see CURRENT and CURRENT DENSITY.

current gain See CURRENT AMPLIFICATION.

current hogging **1.** An undesirable condition that sometimes takes place when two or more transistors are operated in parallel. One device tends to do all the work, taking all the current. The result can be destruction of that device. **2.** The tendency of one component in a group of identical parallel-connected components to dissipate most of the power.

current-hogging injection logic Acronym, CHIL. A form of bipolar digital logic, similar to current-hogging logic but having the greater density characteristic of injection logic.

current instruction register A register in which are held instructions ready for execution by a program controller.

current lag A circuit condition in which current variations are delayed by up to 180 degrees of phase relative to voltage variations. Compare CURRENT LEAD.

current lead A circuit condition in which current variations occur earlier than voltage variations by up to 180 degrees of phase. Compare CURRENT LAG.

current limiting The controlling of current so that it does not exceed a desired value.

current-limiting resistor A series resistor inserted into a circuit to limit the current to a prescribed value.

current loop A point on a transmission line or antenna radiator at which the current reaches a local maximum. Compare CURRENT NODE.

current meter A usually direct-reading instrument, such as an ammeter, milliammeter, or microammeter, used to measure current strength. Also see ELECTRONIC CURRENT METER.

current-meter operation The operation of a voltmeter as a current meter by connecting it to respond to the voltage drop across a resistor that carries the current of interest.

current-mode logic In computer operations, transistor logic in which the transistors operate in the unsaturated mode.

current node A point on a transmission line or antenna radiator at which the current reaches a local minimum. Compare CURRENT LOOP.

current noise Electrical noise produced by current flowing through a resistor.

current probe A transformer usually having a snap-around, one-turn coil that picks up energy from a conductor and couples it into an alternating-current ammeter.

current rating **1.** A specified value of operating current. **2.** See CURRENT-CARRYING CAPACITY.

current-regulated supply See CONSTANT-CURRENT SOURCE.

current regulation The stabilization of current at a predetermined level or value.

current regulator See BARRETTER.

current relay A relay actuated by specific values of pickup and dropout current.

current saturation In the operation of a device (such as a transistor, saturable reactor, or magnetic amplifier), the leveling off of current at a value beyond which no further increase occurs—even though an input parameter is further increased.

current sense amplifier An amplifier used to increase the sensitivity of, or to decrease the loading of, a current-sensing component.

current sensing Sampling a current (e.g., when the voltage drop across a series resistor is used as a proportional indication of the current flowing through the resistor).

current-sensing resistor A low-value resistor inserted into a circuit primarily for current sensing.

current sensitivity In a current meter or galvanometer, current (in amperes or fractions thereof) per scale division.

current-sheet inductance Symbol, L_S. The low-frequency inductance of a single-layer coil, calculated with the formula $L_S = (0.10028\ a^2N^2)/s$, where L_S is in microhenrys, a is the coil radius in inches, N is the total number of turns, and s is the coil length in inches.

current shunt **1.** A resistor connected in parallel with a voltmeter to convert it into an ammeter. **2.** A resistor connected in parallel with the input of a voltage amplifier to make the response of the amplifier proportional to input-signal current.

current sink A circuit or device through which a constant current can be maintained.

current-sinking logic A form of bipolar digital logic. Current flows from one stage to the input of the stage immediately before.

current-squared meter An ammeter or milliammeter whose deflection is proportional to the square of the current.

current-stability factor In a common-base connected bipolar transistor, the ratio dI_E/dI_C, where I_E is the emitter current and I_C is the collector current.

current strength The magnitude of electric current (see CURRENT) (i.e., the number of carriers flowing past a given point per unit time, expressed in coulombs per second or in amperes).

current transformer **1.** A transformer used to increase or decrease current flow. A primary-to-secondary step-up turns ratio reduces the current; a primary-to-secondary step-down turns ratio increases the current. **2.** A particular transformer (as in 1) used to change the range of an alternating-current milliammeter or ammeter.

current vector In a vector diagram, a line with an arrowhead (vector) showing the magnitude and phase of a current. Compare VOLTAGE VECTOR.

current-voltage feedback In an amplifier or oscillator, the process of applying some of the output current and voltage to the input. This feedback might be in phase (positive) or out of phase (negative), with respect to the input.

cursor **1.** A marker that indicates the position where a character can be entered in a video alphanumeric display. Commonly used in computers and word processors. **2.** The sweeping line on a radar display. **3.** The movable marker on a slide rule.

curve trace **1.** A device that supplies a special variable test voltage to a component or circuit under test, at the same time supplying a sweep voltage to an oscilloscope. The component's output voltage is also presented to the oscilloscope. As a result, the response curve of the component appears on the oscilloscope screen. **2.** A device that produces a permanent record (photographic or graphic) of an electrical phenomenon. Also called OSCILLOGRAPH or RECORDER.

curvilinear trace A trace made on paper with curved vertical lines. The lines are curved to match the arc through which the recording pen swings.

cut-in angle In a semiconductor rectifier circuit, a phase angle slightly greater than zero degrees, at which current conduction begins. Compare CUTOUT ANGLE.

Cutler antenna A parabolic-dish antenna, in which the driven element consists of a waveguide that has two apertures on opposite sides of a resonant cavity.

Cutler feed An aircraft antenna feed system in which radio-frequency (RF) energy is fed to the reflector by a resonant cavity at the end of a waveguide.

Cutler tone control A dual resistance-capacitance (RC) filter circuit of the general bridged-tee variety. Variation of the series leg provides adjustable treble boost; variation of the shunt leg provides adjustable bass boost.

cutoff **1.** The process of reducing some operating parameter, such as collector current, to zero by adjusting the bias at the input electrode. **2.** The point on the characteristic curve of an amplifying device, at which the output current drops to zero under no-signal conditions. **3.** The lowest frequency at which a waveguide will efficiently function. **4.** The frequency or frequencies corresponding to the point or points in a filter response, at which the attenuation is three decibels greater than the lowest attenuation within the passband. See also CUTOFF FREQUENCY.

cutoff attenuator A variable, nondissipating attenuator consisting of a variable length of waveguide used at a frequency below cutoff.

cutoff bias In a transistor or vacuum-tube circuit, the value of control-electrode bias that produces output current cutoff.

cutoff current Symbol, I_{co}. In a transistor, the small collector current that flows when the emitter current is zero (common-base circuit) or when the base current is zero (common-emitter circuit).

cutoff frequency **1.** Symbol, f_{co}. The high frequency at which the current-amplification factor of a transistor drops to 70.7% of its 1-kHz value. **2.** In a filter, amplifier, or transmission line, the frequency point(s) at which transmission loss or filter rejection begins. It is generally specified as the half-power point(s), or the point(s) at which the attenuation is three decibels, relative to the lowest attenuation. Examples: the high-frequency cutoff of an amplifier and the upper and lower cutoff points of a bandpass filter.

cutoff limiting Output-peak clipping that results from overdrive in an amplifying device. Compare SATURATION LIMITING.

cutoff potential See CUTOFF BIAS.

cutoff voltage See CUTOFF BIAS.

cutoff wavelength **1.** The wavelength corresponding to cutoff frequency. **2.** For a waveguide, the ratio of the velocity of electromagnetic waves in free space (3×10^8 meters per second) to the cutoff frequency of the waveguide in Hz. The result is thus expressed in meters.

cutout **1.** A device, such as a circuit breaker, that automatically disconnects a circuit, usually to prevent overload, but occasionally to prevent underload. **2.** Emergency switch. **3.** Fuse.

cut-out angle In a semiconductor rectifier circuit, a phase angle slightly less than 180 degrees at which current conduction ceases. Compare CUT-IN ANGLE.

cutout base A fuse block.

cut rate **1.** The speed at which a cutter moves across the surface of a blank vinyl disk during the recording process. **2.** The number of cut lines per inch in a vinyl disk recording.

CW **1.** Abbreviation of CONTINUOUS WAVE. **2.** Abbreviation of CLOCKWISE.

CW filter In a communications receiver, a highly selective filter in the intermediate-frequency (IF) or audio-frequency (AF) stage. The bandwidth is typically 200 Hz to 500 Hz; some audio filters can be set for bandwidths as low as about 50 Hz.

CW laser A laser that emits energy in an uninterrupted stream, rather than in pulses.

CW monitor See KEYING MONITOR.

CW oscillator **1.** In a radio receiver, a variable-frequency oscillator that heterodynes a radiotelegraph signal in the intermediate-frequency (IF) amplifier chain, to make audible the continuous-wave dits and dahs. **2.** Sometimes, an external variable-frequency radio-frequency (RF) oscillator, whose output beats against the actual carrier of a continuous-wave radiotelegraph signal, making it audible as dits and dahs. **3.** An unmodulated, unkeyed oscillator.

CW radar A radar system in which radio-frequency (RF) energy is transmitted continuously.

CW reference signal A sinusoidal radio-frequency (RF) signal, used to control the conduction time of a synchronous demodulator in color television.

C_x Symbol for UNKNOWN CAPACITANCE.

cyan Blue-green, one of the three primary pigments.

cyber- A prefix that indicates relevance to, or involvement with, computers, computer systems, and electronic control systems.

cybernetics The study of control system theory in terms of the relationship between animal and machine behavior.

Cyber Sapiens An expression for a computer or robot with artificial intelligence (AI) on the forefront of current technology.

cyberspace **1.** Alternative expression for INFORMATION SUPERHIGHWAY. **2.** Alternative expression for VIRTUAL REALITY.

cyborg Acronym of the words *cybernetic* and *organism.* **1.** A human being with at least one artificial body part, such as a *prosthesis* (artificial limb). **2.** A human being who is largely composed of robotic body parts.

cycle **1.** Abbreviation, c. One complete, 360-degree revolution of the current or voltage vector in an alternating-current (ac) wave. An ac frequency of 1 cycle per second is 1 Hz (see HERTZ). **2.** A complete sequence of operations.

cycle counter A device that totals the number of cycles of a phenomenon repeated during a given period.

cycle index The number of times that a particular cycle has been, or must be, iterated in a computer program.

cycle index counter A variable that indicates how often a cycle of computer program instructions has been executed. In a program, for example, this can be accomplished by increasing, through instruction, the value of a location's content every time a loop operation is performed.

cycle life The total number of charge-discharge cycles a rechargeable cell or battery can tolerate before becoming useless.

cycle reset To change the value of a cycle count (making it zero or some other value).

cycle shift See CYCLIC SHIFT.

cycles per second Abbreviation, cps. Archaic term for HERTZ.

cycle time Pertaining to an operation, the duration of a complete cycle.

cycle timer A timer that switches a circuit or device on and off, according to a predetermined cycle. Also called *programmed timer.*

cyclic code See GRAY CODE.

cyclic memory In computer operations, a memory whose locations can only be accessed points in a cycle, as of a magnetic diskette.

cyclic shift The moving of data out of one end of a storage register and reentering it character-by-character or bit-by-bit at the other end in a closed loop (e.g., 87654 cyclically shifted one place to the right becomes 48765).

cyclic variations Periodic changes in the features of the ionosphere, occurring on a daily, seasonal, or sunspot-related basis. These changes are fairly predictable.

cycling The tendency of a parameter to oscillate back and forth between two different values.

cyclogram A method of showing the relationship between two signals on an oscilloscope. The two signals must have a fixed phase relationship.

cyclotron A type of particle accelerator. An applied electromagnetic field, acting together with an intense applied magnetic field, cause charged subatomic particles to travel with increasing velocity in a spiral path between two semicircular metal boxes called *dees.* When the particles go fast enough in the correct path, they are expelled and strike a target in their path.

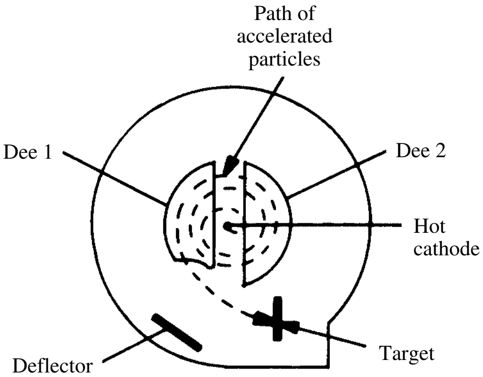

cyclotron

cyclotron frequency The angular frequency of a charged particle in a cyclotron. The cyclotron frequency depends on the number of times per second the magnetic field of the device is reversed.

cyclotron radiation An electromagnetic field produced by the circular movement of charged particles in a fluctuating magnetic field.

cylinder In computer operations, the combination of equal-radius tracks on the platters of a hard disk.

cylinder magnet A permanent magnet in the shape of a cylinder.

cylindrical capacitor See CONCENTRIC CAPACITOR.

cylindrical contour The most common curvature of the face of a magnetic tape recording head; it is a section of a cylinder having a constant radius of 0.5 inch to 1 inch.

cylindrical coordinate geometry A scheme for robot-arm movement. There are three coordinates, called *reach, angle,* and *elevation.* It allows precise positioning of a robot end effector within a region consisting of two concentric cylinders and all the volume in between.

cylindrical coordinates A method of locating a point in three-space in which height, distance, and angle are used to uniquely define points.

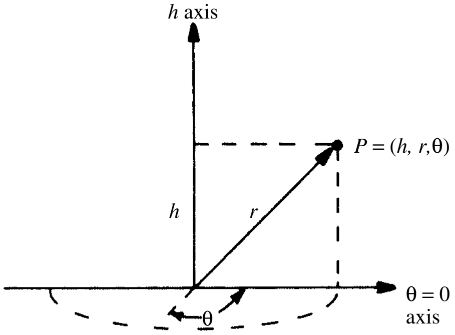

cylindrical coordinates

cylindrical magnet See CYLINDER MAGNET.

cylindrical wave An electromagnetic wave whose field surfaces are nearly perfect cylinders.

cylindrical waveguide A waveguide resembling a round pipe.

cylindrical winding A method of coil winding in which the wire is formed into a helix. There might be only one layer, or there might be several layers. The length of the coil is greater than the diameter. Also called a *linear winding.*

Czochralski method A technique for obtaining a relatively large single crystal from a substance, such as the semiconductors germanium and silicon. The method consists essentially of dipping a seed crystal into a molten mass of the same substance, then slowly withdrawing it while rotating it.

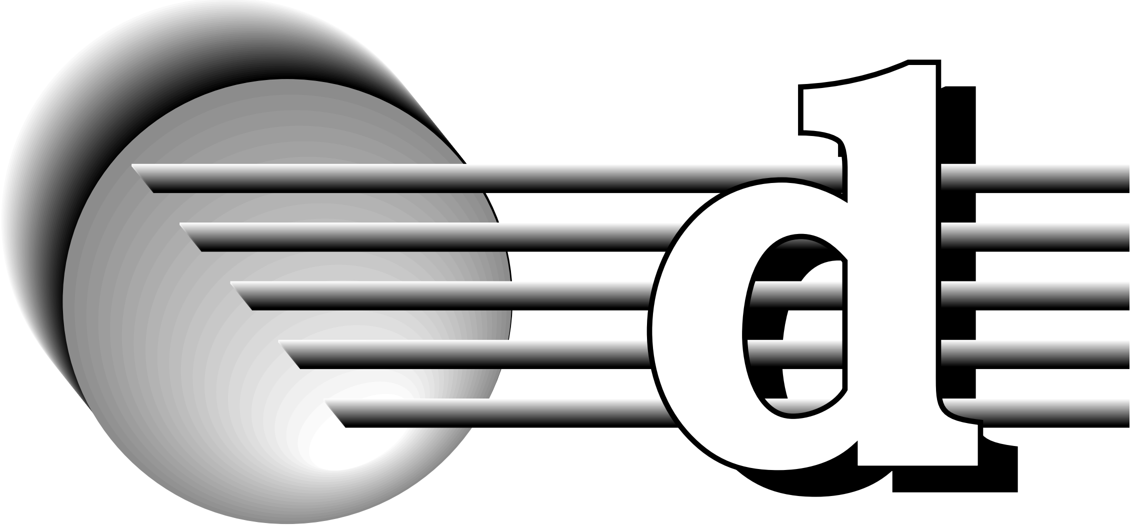

D **1.** Symbol for DEUTERIUM. **2.** Symbol for ELECTRIC DISPLACEMENT. **3.** Symbol for ELECTRIC FLUX DENSITY. **4.** Symbol for DISSIPATION FACTOR. **5.** Symbol for *drain* (see DRAIN, **3**). **6.** Abbreviation of DISSIPATION. **7.** Symbol for *determinant.* **8.** Symbol for DIFFUSION CONSTANT.

d **1.** Abbreviation of DECI. **2.** Symbol for DIFFERENTIAL. **3.** Symbol for distance. **4.** Symbol for DENSITY. **5.** Symbol for *drain* (see DRAIN, **3**). **6.** Abbreviation of DISSIPATION. **7.** Abbreviation of day. **8.** Abbreviation of DEGREE. **9.** Abbreviation of diameter. **10.** Abbreviation of DRIVE.

D/A Abbreviation of DIGITAL-TO-ANALOG. See DIGITAL-TO-ANALOG CONVERSION.

dA **1.** Symbol for DIFFERENTIAL OF AREA. **2.** Symbol for *differential of amplification.* **3.** Seldom-used abbreviation of *deciampere.*

da Abbreviation of DEKA.

DAC Abbreviation of DIGITAL-TO-ANALOG CONVERTER.

DACI Abbreviation of *direct adjacent-channel interference.*

DAGC Abbreviation of DELAYED AUTOMATIC GAIN CONTROL.

daisy chain A method of transferring a signal in a computer from one stage to the next.

daisy wheel A form of printing device consisting of a disk having several dozen radial spokes, each of which has a character molded on its face. The disk rotates to the proper position in the printing process, and a hammer strikes the spoke to press the molding against the ribbon and paper.

DAM Abbreviation of *data-addressed memory.*

Damon effect The change that the susceptibility of ferrite undergoes under the influence of high RF power.

damped galvanometer A galvanometer with a provision for overswing limiting or oscillation prevention.

damped loudspeaker A loudspeaker in which undesirable excursions are prevented by damping in the associated amplifier or speaker circuit.

damped meter **1.** A meter with a provision for overswing limiting or oscillation prevention. **2.** A meter that is protected during transport by a shorting bus between the two meter terminals.

damped natural frequency **1.** The frequency at which a damped system having one degree of freedom will oscillate after momentary application of a transient force. **2.** In the presence of damping, the rate at which a sensing element oscillates freely.

damped oscillations Oscillations in which the amplitude of each peak is lower than that of the preceding one; the oscillation eventually dies out (the amplitude becomes zero). Compare CONTINUOUS WAVE.

damped speaker See DAMPED LOUDSPEAKER.

damped wave A wave whose successive peaks decrease in amplitude (i.e., it decays), eventually reaching an amplitude of zero. Compare CONTINUOUS WAVE and UNDAMPED WAVE.

damped-wave decay See DECREMENT, **1**.

dampen To cause the amplitude of a signal to decay.

damper See DAMPING DIODE.

damper diode See DAMPING DIODE.

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

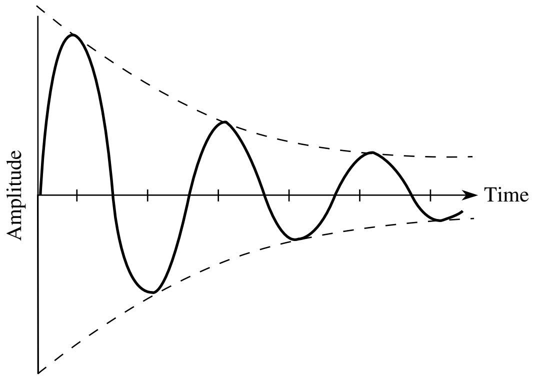

damped wave

damper winding A special short-circuited motor winding that opposes pulsation or rotation of the magnetic field.

damping **1.** See DAMPING ACTION. **2.** In a loudspeaker, sound-absorbent material used to minimize resonant effects within the enclosure.

damping action **1.** Quenching action. **2.** The prevention of overswing, dither, or flutter in a meter or loudspeaker (see DAMPED GALVANOMETER, DAMPED LOUDSPEAKER, DAMPED METER). **3.** The prevention of oscillation or ringing in a circuit. **4.** Inhibition of the vibration of an acoustic transducer to prevent ringing and other unwanted effects.

damping coefficient A figure expressing the ratio of the damping in a system to critical damping.

damping diode A diode used to prevent oscillation in an electric circuit (e.g., the diode that prevents ringing in the power supply of a television receiver). Also called *damper.*

damping factor **1.** Symbol, a. For a coil of inductance L and RF resistance R in a damped-wave circuit, the value $R/2L$, where L is in henrys and R in ohms. **2.** Abbreviation, F_o. For a torque motor, the ratio of the stall torque to the no-load rotational speed.

damping magnet A permanent magnet so situated, with respect to a moving conductor, disk, or plate, that the resulting field opposes the movement.

damping ratio See DAMPING COEFFICIENT.

damping resistance **1.** The value of shunt resistance required to prevent ringing in a coil. **2.** The value of resistance required for critical damping of a galvanometer.

damping resistor **1.** A shunt across a coil to prevent ringing. **2.** A resistor used to provide critical damping of a galvanometer.

Daniell cell A nonpolarizing primary wet cell with zinc (negative) and copper (positive) electrodes. The zinc plate is in a porous cup containing a weak zinc-sulfate solution with a little sulfuric acid; the cup is in a jar filled with a saturated copper-sulfate solution in which the copper electrode is immersed. Typical voltage for the cell is 1.1 V.

daraf The unit of ELASTANCE. Elastance in darafs is the reciprocal of capacitance in farads.

dark conduction The flow of dark current in a photoconductive or glow-discharge device.

dark current The usually tiny current flowing through a darkened photoconductive cell, phototransistor, or glow-discharge device.

dark discharge The occurrence of a discharge in a gas, without the production of visible light.

dark-spot signal A spurious signal generated by some camera tubes, arising from secondary-emission effects.

dark-trace tube An oscilloscope tube on whose white screen a long-persistence magenta image is traced by the electron beam. Illuminating the screen with bright light intensifies the image.

Darlington amplifier A high-gain amplifier that uses a COMPOUND CONNECTION of two bipolar transistors.

Darlington pair See COMPOUND CONNECTION.

D'Arsonval current A large, low-voltage, high-frequency current at one time thought to be therapeutic.

D'Arsonval meter A electromechanical analog meter, in which a coil turns on jeweled pivots between the poles of a strong magnet and against the force of spiral springs. A pointer is attached to the coil. The pointer moves over a calibrated scale.

D'Arsonval movement The mechanism of a D'Arsonval meter.

DART Abbreviation of *data analysis recording tape.*

dart leader A flow of electrons along a path traveled by a lightning stroke, preceding a second stroke. The dart leader, if any, occurs a few milliseconds after the first stroke. Several strokes could occur, each preceded by a dart leader, within less than 1 second.

dash The longer of the two characters (DOT and DASH) of the telegraph code. The duration of the dash is three times longer than that of a dot.

dashpot A delayed-action device in which the movement of a piston is slowed by air or a liquid in a closed cylinder.

dashpot relay A time-delay relay assembly in which the delay is obtained with a DASHPOT.

DAT **1.** Abbreviation of DIGITAL AUDIO TAPE. **2.** Abbreviation of *diffused-alloy transistor.*

data **1.** A collection of digital bits (binary digits) with informational content (e.g., a computer file, a digital image, or a digital sound recording). **2.** General expression for information, especially in encoded or written form.

data acquisition The reception and gathering of data (see DATA COLLECTION and DATA SYSTEM, **1**).

data-acquisition system A computer or dumb terminal used to gather data from one or more external points.

data analysis display unit A video display peripheral for online data analysis.

data area A computer memory area that holds data only (i.e., one that does not contain program instructions).

data bank A data file stored in a direct-access storage device, which can be drawn from by many system users through remote terminals.

database **1.** A computer file containing often-used information (e.g., names and addresses, or electronic part numbers). **2.** A popular form of computer software that allows users to create, maintain, and modify information.

data block A set of data bits, comprising an identifiable item.

data bus A conductor or medium over which digital data is transmitted from one place to another within a computer.

data carrier storage A medium of data storage outside of a computer (e.g., a magnetic disk).

data code A set of abbreviations or codes for data characters or words.

data collection The pickup of signals representing test data and their transmission to a computer, data processor, or recorder. Also see DATA SYSTEM, **1**.

datacom Acronym for DATA COMMUNICATION.

data communication The transmission and reception of data signals between or among points in a system.

data communication terminal A computer peripheral providing an input and output link to a central computer system, and that can be used offline for other functions.

data compression **1.** The process of reducing the size of a data file by eliminating redundancies. **2.** The process of minimizing the length of a data transmission by eliminating redundancies. **3.** The process of reducing the bandwidth of a data transmission. **4.** The process of reducing the dynamic amplitude range of a data transmission.

data control The automatic control of incoming and outgoing data in a data processing system.

data conversion The process of changing data from one form to another, e.g., from analog to digital (A/D), digital to analog (D/A), parallel to serial, or serial to parallel.

data converter **1.** A circuit or device for performing DATA CONVERSION. **2.** An analog-to-digital (A/D) converter. **3.** A digital-to-analog (D/A) converter. **4.** A parallel-to-serial converter. **5.** A serial-to-parallel converter.

data description The description of a unit of data, as included in a computer source program.

data display A device, such as a cathode-ray tube (CRT) or liquid-crystal display (LCD), that presents data for visual examination. Compare DATA PRINTOUT.

data element **1.** A component of a data signal (e.g., a number, letter, symbol, or the equivalent electrical pulses). **2.** A device or circuit for acquiring or processing data. **3.** A unit of data (e.g., a field in a file).

data-flow diagram A block diagram showing the movement of data through a data-processing system.

data format The form of data in a record or file (e.g., character format or numerical format).

data gathering See DATA COLLECTION.

data-handling capacity **1.** The amount of data that can be stored in a memory circuit. **2.** The amount of data that can be transmitted over a certain medium. **3.** The rate at which data can be transferred under certain conditions.

data-handling system A system that gathers, routes, transmits, or receives data, but does not necessarily process it.

data item A logical element (character, byte, or bit) describing a characteristic of a record used by a system for which there is a specific application.

data level Descriptive, through a programming language, of the relative weight of logical elements (data items) in a computer record. Also called *data hierarchy.*

data link The portion of a computer system that gathers data and, if necessary, converts it to a form acceptable by a computer.

data matrix Variables and their possible values stored as a series of columns and rows of values in a computer memory.

data name An operand specified in a computer source program.

data pickup **1.** A transducer that collects data signals from a source; it converts nonelectrical data into corresponding electrical signals and delivers its output to a data processing system. **2.** Data acquisition.

data playback The reproduction of data signals stored by some method of data recording.

data plotter See X-Y PLOTTER.

data printout **1.** A device that prints a record of data or the results of a computation. **2.** A permanent printed record, usually of a calculation or computation—especially the printed output of a computer peripheral device.

data processing Work performed on acquired data, as in solving problems, making comparisons, classifying material, organizing files. Usually done by a computer.

data-processing equipment A digital computer and the peripheral equipment needed to collate, store, analyze, and reduce data.

data-processing machine A computer or system used to collate, store, analyze, and reduce data, as opposed to a computer or system used primarily to solve problems or perform routine tasks. Also called *data processor.*

data-processing system An electronic system for automatic data processing. It can be based on analog and/or digital techniques.

data processor See DATA-PROCESSING MACHINE.

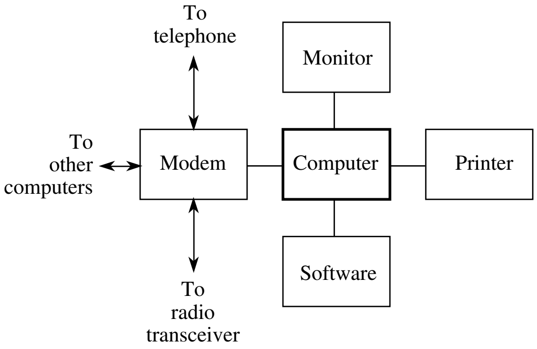

data-processing system

data receiver At a particular point in a data-processing system, a circuit or device for receiving data from a data transmitter.

data reception Receiving data signals from some point within or outside a data-processing system.

data-reception system A data receiver and its associated equipment.

data record A computer-processed record containing a data unit.

data recorder A machine for storing data acquired in the form of electrical signals (see DATA RECORDING).

data recording **1.** The preservation of data signals by some process, such as magnetic-disk encoding, optical-disk encoding, or tape recording, for future use or as a backup. **2.** A record of data signals, as on magnetic tape.

data reduction The summarization of a mass of electronically gathered data.

data-reduction system A system used to minimize the amount of data necessary to convey given information.

data representation Values and data as described by numerals, symbols, and letters (e.g., computer program instructions).

data segment As related to a particular computer process, a subunit of allocated storage containing data only.

data selector/multiplexer A digital circuit that has several or many input signals, and feeds one of them onto a common line.

data set A device that connects a data processor to a telegraph or telephone line.

data signal **1.** A signal (such as one of binary bit combinations) that can represent data as numbers, letters, or symbols. **2.** A signal current or voltage proportional to some sampled quantity, and that can be used to actuate indicating instruments during tests or measurements.

data statement A computer source program statement identifying a data item and specifying its format.

data storage The preservation of data, particularly computer files, for long periods of time in nonvolatile form (no source of power is required to ensure that the data remains intact).

data storage media Hardware that preserves data, particularly computer files, for long periods of time in nonvolatile form (no source of power is required to ensure that the data remains intact). Common media include magnetic disks, magnetic tape, and optical disks.

data synchronizer A device used to synchronize data transmission within a computing or processing system.

data system **1.** An arrangement for collecting, recording, and routing data in the form of electrical signals. **2.** An arrangement for processing data (i.e., for correlating, computing, routing, storing, etc.).

data terminal A remote input/output device connected to a central computer.

data throughput In a computer system, the amount of data per unit time (bytes, kilobytes, megabytes, gigabytes, or terabytes per second or minute) that can be transferred from one place to another.

data transducer In tests and measurements, a transducer that converts a monitored phenomenon into electrical quantities that can be used for computer analysis or calculations.

data transmission Sending data signals from a pickup point or processing stage to another point within a data-processing system; also, sending such signals to points outside the system.

data-transmission system A data transmitter and its associated equipment.

data transmission utilization measure The ratio of the useful data output of a data-transmission system to the total data input.

data transmitter A circuit or device for sending data from point to point within or outside of a data-processing system.

data unit Characters in a group that are related in a way that makes them a meaningful whole (e.g., a text word, or an object such as a circle in vector graphics).

data value A measure of the amount of information contained in a certain number of data bits. The greater the ratio of the actual information to the number of bits, the higher the data value.

data words In digital computer operations, words (bit groups) representing data, rather than program instructions.

DAVC Abbreviation of DELAYED AUTOMATIC VOLUME CONTROL.

David Phonetic alphabet code word for letter D.

daylight effect The modification of transmission paths during the day because of ionization of the upper atmosphere by solar radiation.

daylight lamp An incandescent lamp whose filament is housed in a blue glass bulb, which absorbs some red radiation and transmits most of the green, blue, and violet. So called because the spectral output resembles that of typical daylight.

daylight range The distance over which signals from a given transmitter are consistently received during the day.

DB **1.** Abbreviation of DIFFUSED BASE of a transistor. **2.** Abbreviation of DOUBLE BREAK (relay).

dB **1.** Abbreviation of DECIBEL or *decibels.* **2.** Symbol for *differential of susceptance.*

dBa Abbreviation of ADJUSTED DECIBELS.

dBc Abbreviation of *decibels referred to the carrier.*

DBD Abbreviation of *double-base diode.*

dBd The power gain of an antenna in the direction of maximum radiation, compared to the radiation in the favored direction of a half-wave dipole in free space receiving the same amount of power. Expressed in decibels.

dBi The power gain of an antenna in the direction of maximum radiation, compared to the radiation from a theoretical isotropic antenna in free space receiving the same amount of power. Expressed in decibels.

dBj The level of an RF signal, in decibels, relative to 1 millivolt.

dBk Abbreviation of DECIBELS REFERRED TO 1 KILOWATT.

DBM Abbreviation of *database management.*

dBm Abbreviation of DECIBELS REFERRED TO 1 MILLIWATT.

dBm0 Signal level in dBm, referred to a zero-transmission level.

dBm0p Noise in dBm0, measured with set psophometric weighting.

dB meter A usually high-impedance ac voltmeter with a scale reading directly in decibels.

dBmp The level in dBm, measured with psophometric weighting. Generally equal to dBm –2.5, for a noise level that is flat within the communications audio range.

dBmr Decibels measured with respect to zero transmission level.

dBmV Abbreviation of DECIBELS REFERRED TO 1 MILLIVOLT.

dBrap Abbreviation of DECIBELS ABOVE REFERENCE ACOUSTIC POWER (10^{-6} W).

dBrn Abbreviation for *decibels above reference noise.* A level of 0 dBrn is defined as noise power of 10^{-9} W (1 nanowatt).

dBrnc Noise power in dBrn for a circuit with message weighting c.

dBrnc0 Noise in dBrnc measured with respect to zero transmission level.

dBV Abbreviation of DECIBELS REFERRED TO 1 VOLT.

dBW Abbreviation of DECIBELS REFERRED TO 1 WATT.

dBx Abbreviation of DECIBELS ABOVE REFERENCE COUPLING.

dC Symbol for *differential of capacitance.*

dc **1.** Abbreviation of DIRECT CURRENT. **2.** Abbreviation of *direct-coupled.*

dc-ac converter A circuit that converts a dc input voltage into an ac output voltage, with or without step-up or step-down. Also called INVERTER.

dc alpha The current amplification factor (ALPHA) of a common-base transistor stage for a dc input (emitter) signal. Compare DC BETA.

dc amplifier **1.** A *direct-coupled amplifier.* 2. An amplifier for boosting direct-current signals.

dc balance **1.** Adjustment of a circuit or device for dc stability or dc null. **2.** Adjustment of a circuit for dc stability during gain changes. **3.** A potentiometer or other variable component used to stabilize or null a dc circuit.

dc bar See DC BUS.

dc base current Symbol, $I_{B(dc)}$. The static direct current in the base element of a bipolar transistor.

dc base resistance Symbol, $R_{B(dc)}$. The static dc resistance of a bipolar transistor's base element; $R_{B(dc)} = V_B/I_B$.

dc base voltage Symbol, $V_{B(dc)}$. The static dc voltage at the base element of a bipolar transistor.

dc beta The current amplification factor (BETA) of a common-emitter-connected transistor for a dc input (base) signal. Compare DC ALPHA.

dc block A coaxial section that has a capacitance in series with the inner or outer conductor, or both, to block dc while passing RF. Compare DC SHORT.

dc bus A supply conductor carrying direct current only.

dcc Abbreviation of double cotton covered (wire).

dc cathode current Symbol, $I_{K(dc)}$. The static direct current in the cathode element of an electron tube.

dc cathode resistance Symbol, $R_{K(dc)}$. The static dc resistance of the cathode path of an electron tube.

dc cathode voltage Symbol, $V_{K(dc)}$. The static dc voltage at the cathode of an electron tube.

dc circuit breaker A circuit breaker operated by direct-current overload or underload, depending on its design and application.

dc collector current Symbol, $I_{C(dc)}$. The static direct current in the collector element of a bipolar transistor.

dc collector resistance Symbol, $R_{C(dc)}$. The static dc resistance of a bipolar transistor's collector element; $R_{C(dc)} = V_C/I_C$.

dc collector voltage Symbol, $V_{C(dc)}$. The static dc voltage at the collector element of a bipolar transistor.

dc component In a complex wave (i.e., one containing both ac and dc), the current component having an unchanging polarity. The dc component constitutes the mean (average) value around which the ac component alternates, pulsates, or fluctuates.

dc converter A dynamoelectric machine for converting low-voltage dc into higher-voltage dc. It is essentially a low-voltage dc motor coupled me-

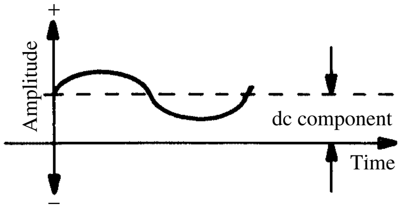

dc component

chanically to a higher-voltage dc generator. Compare DC INVERTER.

dc coupling See DIRECT COUPLING.

dc drain current Symbol, $I_{D(dc)}$. The static direct current in the drain element of a field-effect transistor.

dc drain resistance Symbol, $R_{D(dc)}$. The static dc resistance of an FET's drain element; $R_{D(dc)} = V_D/I_D$.

dc drain voltage Symbol, $V_{D(dc)}$. The static dc voltage at the drain element of a field-effect transistor.

dc dump In digital computer operation, removing dc power from a computer, which would eradicate material stored in a volatile memory.

dc emitter current Symbol, $I_{E(dc)}$. The static direct current in the emitter element of a bipolar transistor.

dc emitter resistance Symbol, $R_{E(dc)}$. The static dc resistance of a bipolar transistor's emitter element; $R_{E(dc)} = V_E/I_E$.

dc emitter voltage Symbol, $V_{E(dc)}$. The static dc voltage at the emitter element of a bipolar transistor.

dc equipment Apparatus designed expressly for operation from a dc power supply. Compare AC EQUIPMENT and AC/DC.

dc erase head In a magnetic recorder, a head supplied with a dc current for the purpose of removing data.

dc error voltage In a television receiver, the dc output of the phase detector, which is used to control the frequency of the horizontal oscillator.

dc gate current Symbol, $I_{G(dc)}$. The very small static direct current in the gate element of a field-effect transistor.

dc gate resistance Symbol, $R_{G(dc)}$. The very high, static dc resistance of an FET's gate element; $R_{G(dc)} = V_G/I_G$.

dc gate voltage Symbol, $V_{G(dc)}$. The static dc voltage at the gate element of a field-effect transistor.

dc generator **1.** A rotating machine (dynamo) for producing direct current. Also see DYNAMOELECTRIC MACHINERY. **2.** Generically, a device that produces direct current: batteries, photocells, thermocouples, etc.

dc generator amplifier A special type of generator that provides power amplification. The input signal energizes the field winding of a constant-speed machine; because the output voltage is proportional to field flux and armature speed, a high output voltage is obtained. Also see AMPLIDYNE.

dc grid bias Steady dc control-grid voltage used to set the operating point of an electron tube.

dc grid current Symbol, $I_{G(dc)}$. The static direct current in the control-grid element of an electron tube.

dc grid resistance Symbol, $R_{G(dc)}$. The static dc resistance in the control-grid element of an electron tube; $R_{G(dc)} = V_G/I_G$.

dc grid voltage Symbol, $V_{G(dc)}$. The static dc voltage at the control grid of an electron tube.

dc inserter In a television transmitter, a stage that adds the dc pedestal (blanking) level to the video signal.

dc inverter An electrical, electronic, or mechanical device that converts dc to ac. Also called INVERTER.

dcl Abbreviation of *dynamic load characteristic.*

dc leakage The unintended flow of direct current.

dc leakage current **1.** The direct current that normally passes through a correctly polarized electrolytic capacitor operated at its rated dc working voltage. **2.** The zero-signal reverse current in a semiconductor pn junction.

DCM Abbreviation of DIGITAL CAPACITANCE METER.

D/CMOS Combination of DMOS and CMOS on a monolithic chip.

dc motor A motor that operates from direct current only.

dc noise Noise heard during the playback of magnetic tape that was recorded while direct current was in the record head.

dc noise margin In a digital or switching circuit, the difference $V_o - V_i$, where V_o is the output-voltage level of a driver gate and V_i is the input threshold voltage of a driven gate.

dc operating point For a bipolar transistor, field-effect transistor, or vacuum tube, the static, zero-signal dc voltage and current levels.

dc overcurrent relay A relay or relay circuit actuated by dc coil current rising above a specified level. Compare DC UNDERCURRENT RELAY.

dc overvoltage relay A relay or relay circuit actuated as a result of the dc coil voltage rising above a specified level. Compare DC UNDERVOLTAGE RELAY.

dc patch bay A patch bay in which the dc circuits of a system are terminated.

dc picture transmission In television, transmission of the dc component of the video signal; this component corresponds to the average illumination of the scene.

dc plate current Symbol, $I_{P(dc)}$. The static direct current in the plate element of an electron tube.

dc plate resistance Symbol, $R_{P(dc)}$. The static dc resistance of the internal plate-cathode path of an electron tube; $R_{P(dc)} = V_P/I_P$.

dc plate voltage Symbol, $V_{P(dc)}$. The static dc voltage at the plate electrode of an electron tube.

dc positioning Alignment of the spot on the screen of an oscilloscope tube, by means of adjustable dc voltages applied to the horizontal and vertical deflecting plates.

dc power Symbol, P_{dc}. Unit, watt. The power in a dc circuit; $P_{dc} = EI$, where E is in volts and I is in amperes. Compare AC POWER. Also see POWER.

dc power supply A power unit that supplies direct current only. Examples: battery, transformer/rectifier/filter circuit, dc generator, and photovoltaic cell. Compare AC POWER SUPPLY.

dc relay A relay having a simple coil and core system for closure by direct current, which can be rectified ac.

dc resistance Resistance offered to direct current, as opposed to in-phase ac resistance.

dc resistivity The resistivity of a sample of material measured using a pure dc voltage under specified conditions (physical dimensions, temperature, etc.).

dc restoration The reinsertion of the dc component into a signal from which the component has been extracted through a capacitor or transformer.

dc restorer A circuit that reinserts the average dc component of a signal after the component has been lost because the signal passed through a capacitor or transformer.

DCS Abbreviation of DORSAL COLUMN STIMULATOR.

dc shift A shift in the DC OPERATING POINT.

dc short A coaxial fitting providing a dc path between the center and outer conductors, while permitting radio-frequency (RF) current to flow easily through the coaxial section. Compare DC BLOCK.

dc signaling A signaling procedure that uses direct current as the medium (e.g., simple wire telegraphy or telephony).

dc source **1.** DC GENERATOR. **2.** A live circuit point from which one or more direct currents can be taken.

dc source current Symbol, $I_{S(dc)}$. The static direct current in the source element of a field-effect transistor.

dc source resistance Symbol, $R_{S(dc)}$. The static dc resistance of an FET's source element.

dc source voltage Symbol, $V_{S(dc)}$. The static dc voltage at the source element of a field-effect transistor.

DCTL Abbreviation of DIRECT-COUPLED TRANSISTOR LOGIC.

dc-to-dc inverter See DC INVERTER.

dc transducer **1.** A transducer that depends on direct current for its operation (i.e., it has a dc power supply whose output is modulated by the sensed phenomenon). **2.** A transducer that converts a direct current into some other form of energy, such as heat, pressure, or sound.

dc transformer A dc-to-dc converter providing voltage step-up. The applied dc is usually first converted to ac, which is then stepped up by a transformer. The higher-voltage ac is then rectified to produce a high dc output voltage.

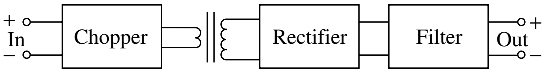

dc transformer

dc transmission **1.** Sending dc power from a generating point to a point of use. **2.** In television transmission, the retention of the dc component in the video signal.

dc tuning voltage The capacitance-varying dc voltage applied to a varactor in an inductance-capacitance (LC) tuned circuit.

dcu Abbreviation of decimal counting unit.

dc undercurrent relay A relay or relay circuit that is actuated as a result of the dc coil current dropping below a specified level. Compare DC OVERCURRENT RELAY.

dc undervoltage relay A relay or relay circuit that is actuated as a result of the dc voltage dropping below a specified level. Compare DC OVERVOLTAGE RELAY.

dcv Abbreviation of DC VOLTS or DC VOLTAGE.

dc voltage Abbreviation, dcv. A voltage that does not change in polarity, an example being the voltage delivered by a battery or dc generator. Also see VOLTAGE.

dc working voltage Abbreviation, dcwv. The rated dc voltage at which a component can be operated continuously with safety and reliability.

dc working volts Abbreviation, dcwv. The actual value, expressed in volts, of a DC WORKING VOLTAGE.

dcwv Abbreviation of DC WORKING VOLTAGE.

dD Symbol for *differential of electric displacement.*

DDA Abbreviation of *digital differential analyzer.*

DDD Abbreviation of DIRECT DISTANCE DIALING (telephone).

D display See D SCOPE.

DE Abbreviation of *decision element.*

dE Symbol for *differential of voltage.*

deac In frequency-modulation (FM) receivers, a device used for deemphasis. The name is short for *deaccentuator.*

deactuating pressure For an electrical contact, the pressure at which contact is made or broken as the pressure reaches the level of activation.

dead **1.** Unelectrified. **2.** Lacking electromagnetic signals or fields. **3.** Electrically or mechanically inoperative.

dead band **1.** A radio-frequency band on which no signals are heard. **2.** A range of values for which an applied control quantity (e.g., current or voltage) has no effect on the response of a circuit.

deadbeat The state wherein a moving body (such as the pointer of a meter or the voice coil of a loudspeaker) comes to rest without overswing or oscillation.

deadbeat galvanometer See DEADBEAT INSTRUMENT.

deadbeat instrument A meter or recorder that is highly damped to ensure that overswing or oscillation does not occur.

deadbeat meter See DEADBEAT INSTRUMENT.

dead break An unreliable contact of a relay, caused by insufficient pressure.

dead circuit A circuit that is electrically disabled.

dead end The unused end of a tapped coil (i.e., the turns between the end of the coil and the last turn used).

dead-end tower A supporting tower for an antenna or transmission line that can withstand stresses caused by loading or pulling.

dead file A computer file that is not in use, but is being kept in a record.

dead front panel A metal panel that, for safety and desensitization, is completely insulated from voltage-bearing components mounted on it; it is often grounded.

dead interval See DEAD TIME.

dead line A deenergized line or conductor.

dead period See DEAD TIME.

dead room An anechoic room in which acoustic tests and studies are made.

dead short A short circuit with extremely low (virtually no) resistance from dc into the radio-frequency spectrum.

dead space See DEAD BAND.

dead spot **1.** An area in which radio waves from a particular station are not received. **2.** On a vacuum-tube cathode (directly or indirectly heated), a spot from which no electrons are emitted.

dead stretch The tendency of insulating materials to permanently retain their approximate dimensions after having been stretched.

dead time **1.** DOWN TIME. **2.** An interval during which there is no response to an actuating signal. **3.** In a computer system, an interval between related events that is allocated to prevent interference between the events.

dead volume In a pressure transducer, the zero-stimulus volume of the pressure port cavity.

dead zone See ZONE OF SILENCE.

debatable time Computer time that cannot be placed in any other category.

debounced switch A switch in sensitive computer or control systems that has circuitry for eliminating the electrical effects of bounce (see BOUNCE, **1**).

de Broglie waves Electromagnetic waves that are believed to be associated with moving particles (such as electrons, protons, and neutrons).

debug **1.** To eliminate errors in, and maximize the efficiency of, a computer program or group of programs. **2.** To optimize the design and construction of electronic equipment.

debugging A process by which engineers eliminate the flaws in a circuit, machine, or computer program.

debugging aid routine A computer program used to test other programs.

debugging period The time interval following completion of a software design, a hardware interconnection, or the manufacture of a piece of electronic equipment, during which errors and imperfections are sought and corrected.

debunching In a velocity-modulated tube, such as a Klystron, a beamspreading space-charge effect that destroys electron bunching.

Debye length The maximum distance between an electron and a positive ion over which the electron is influenced by the field of the ion.

Debye shielding distance See DEBYE LENGTH.

deca- A prefix that indicates multiplication by 10.

decade **1.** A frequency band whose upper limit is 10 times the lower limit. Example: 20 Hz to 200 Hz. **2.** A set of 10 switched or selectable components in which the total value is 10 times that of individual values. Example: a *decade capacitor.* Also called DECADE BOX. **3.** A group, sometimes a unit of access, of 10 computer storage locations.

decade amplifier An amplifier or preamplifier whose gain can be adjusted in increments of 10 (×1, ×10, ×100, etc.).

decade box A group of components that provides values in 10 equal steps selected by a switch or jacks. For compactness, the components and the associated hardware are enclosed in a box or can. See, for example, DECADE CAPACITOR.

decade capacitor A composite capacitor whose value is variable in 10 equal steps. For example, the values might be set at 100 picofarads (pF), 200 pF, 300 pF, etc., up to 1000 pF. Compare DECADE INDUCTOR and DECADE RESISTOR.

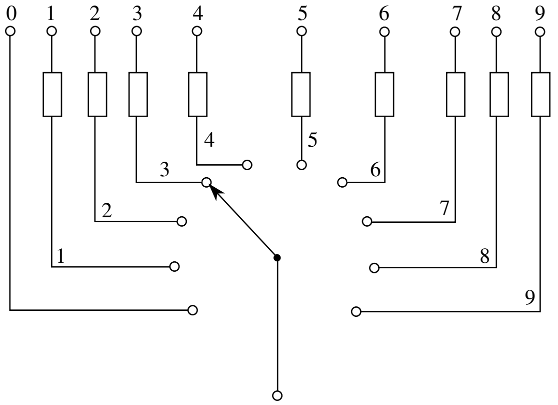

decade box

decade counter A counter (see COUNTER, **1**, **2**) in which the numeric display is divided into sections, each having a value 10 times that of the next and displaying a digit from zero to nine.

decade inductor An inductor whose value is variable in 10 equal steps. Compare DECADE CAPACITOR and DECADE RESISTOR.

decade resistor A resistor whose value is variable in 10 equal increments. Compare DECADE CAPACITOR and DECADE INDUCTOR.

decade scaler A scale-of-10 electronic counter (i.e., a circuit delivering one output pulse for each group of 10 input pulses).

decametric waves Waves in the 10- to 100-meter band (30 to 3 MHz).

decay **1.** The decrease in the value of a quantity, e.g., current decay in a resistance-capacitance circuit. **2.** The gradual, natural loss of radioactivity by a substance.

decay characteristics **1.** The decay of a parameter; usually an exponential function. **2.** The persistence time in a storage oscilloscope.

decay curve A curve, usually logarithmic, representing the function of quantity versus time for a signal decrement, the decrement of radioactivity, or other natural process.

decay rate A quantitative expression for the rapidity with which a quantity decreases. Generally listed in decibels per second (dB/s) or decibels per millisecond (dB/ms).

decay time The time required for pulse amplitude to fall from 90% to 10% of the peak value. Also called FALL TIME.

Decca A 70- to 130-kHz CW radio navigation system (British).

decelerated electron A high-speed electron that is abruptly decelerated upon striking a target, causing X-rays to be emitted.

decelerating electrode A charged electrode that slows the electrons in an electron beam.

deceleration Acceleration that results in a decrease in speed.

deceleration time **1.** The time taken by magnetic tape to stop moving after the last recording or playback has finished. **2.** The time taken by a mechanical data storage medium, such as a hard disk, to come to rest after completion of a read or write operation, or on powering-down.

decentralized data processing Data processing in which the computing equipment is distributed among managerial subgroups.

deception A method of producing misleading echoes in enemy radar.

deception device A radar device, or radar-associated device, for deception.

deci- Abbreviation, d. A prefix meaning one-tenth (10^{-1}). Examples: DECIBEL, DECIMETER.

decibel Abbreviation, dB. A practical unit of relative gain. In terms of power, the relative gain in decibels is equal to:

$$\text{Gain (dB)} = 10 \log_{10}(P_{out}/P_{in}),$$

where P_{out} is the output power and P_{in} is the input power. For voltage, if the input and output impedances are the same, the gain in decibels is given by:

$$\text{Gain (dB)} = 20 \log_{10}(V_{out}/V_{in}),$$

where V_{out} is the output voltage and V_{in} is the input voltage. For current, if the input and output impedances are the same, the gain in decibels is given by:

$$\text{Gain (dB)} = 20 \log_{10}(I_{out}/I_{in}),$$

where I_{out} is the output current and I_{in} is the input current. Losses are indicated by negative dB gain values.

decibels above reference acoustic power Abbreviation, dBrap. The ratio of a given acoustic power level to a lower reference acoustic power level, specified in decibels.

decibels above reference noise Abbreviation, dBrn. The ratio of the noise level at a selected point in a circuit to a lower reference noise level, in decibels.

decibels referred to 1 millivolt Abbreviation, dBmV. The relative voltage level of a signal when compared with a 1-mV signal measured at the same terminals.

decibels referred to 1 milliwatt Abbreviation, dBm. The ratio, in decibels, of an applied power level to the power level of 1 mW.

decibels referred to 1 volt Abbreviation, dBV. The ratio, in decibels, of a given voltage to 1 V, expressed in decibels.

decibels referred to 1 watt Abbreviation, dBW. The ratio of a given power level to the power level of 1 W, expressed in decibels.

decider See DECISION ELEMENT.

decigram A unit of mass equal to 0.1 gram.

deciliter A unit of volume equal to 0.1 liter, or 10^{-4} cubic meter.

decilog A unit equal to 0.1 times the common logarithm of a ratio.

decimal **1.** Pertaining to the base-10 number system (see DECIMAL NUMBER SYSTEM). **2.** A base-10 numerical fraction, represented by figures to the right of the radix point (decimal point), and arranged serially according to negative powers of 10. Examples: $0.12 = 1.2 \times 10^{-1}$, $0.00135 = 1.35 \times 10^{-3}$.

decimal attenuator An attenuator circuit whose resistances are chosen for attenuation in decimal steps. Thus, one section provides attenuation in steps of 0.1 times the applied voltage, another in steps of 0.01 times the applied voltage, another in steps of 0.001 times the applied voltage, etc.

decimal code A method of defining numbers, in which each place has a value of ten times that immediately to the right.

decimal-coded digit **1.** A numeral from 0 to 9. **2.** A numeral in the DECIMAL NUMBER SYSTEM. **3.** A binary representation of a decimal value from 0 to 9.

decimal digit A numeral from 0 to 9.

decimal equivalent The decimal number equal to a given fraction (e.g., the decimal equivalent of $^{21}/_{64}$ is 0.3281).

decimal fraction See DECIMAL, 2.

decimal notation See DECIMAL NUMBER SYSTEM.

decimal number system The familiar base-10 or radix-10 number system, in which the digits 0 through 9 represent values according to their position, relative to the decimal point (also called the *radix point*). Positions to the left of the point represent successive positive powers of 10, and those to the right represent successive negative powers of 10.

decimal point The radix point in a decimal number. It serves to separate the integral part from the fractional part of the number.

decimeter waves See MICROWAVES.

decimetric waves Electromagnetic waves having lengths ranging from 0.1 meter to 1 meter (3000 MHz to 300 MHz). Also known as *ultrahigh frequency (UHF).*

decineper A natural-logarithmic unit equal to 0.1 *neper.*

decipher See DECODING, **3**.

decision **1.** A choice based on the evaluation and comparison of data, and the identification of a specified objective. **2.** In digital computer operations, the automatic selection of the next step in a sequence, on the basis of data being compared by a relational test.

decision box A block on a computer flowchart indicating the point at which a decision (see DECISION, **2**) must be made as to which of several branches the program will take.

decision elements See LOGIC CIRCUITS.

decision instruction A computer program instruction to compare the values of operands and take an appropriate action, as per the BASIC instruction “IF A = B THEN GO TO (line number).”

decision procedure In decision theory, a series of calculations made to optimize the speed or efficiency of a process, or to minimize risk, failure, cost, etc.

decision theory A statistical discipline concerned with identifying and evaluating choices and alternatives, and determining the best sequence of steps to take in reaching an objective.

decision tree In decision theory, a diagram showing alternative choices, so called from its resemblance to a tree with branches.

decision value A value that defines the boundary between two intervals in the encoding process.

deck **1.** See TAPE DECK. **2.** A pack of punched cards in a computer file.

declarative macroinstruction As part of an assembly language, instructions to the compiler to do something or record a condition without affecting the object program.

declarative statement A computer source program instruction specifying the size, format, and kind of data elements and variables in a program for a compiler.

declination **1.** The angle representing the deviation of magnetic north from true north; it is the angle subtended by a freely turning magnetic needle and the meridian. Compare INCLINATION. **2.** Celestial latitude.

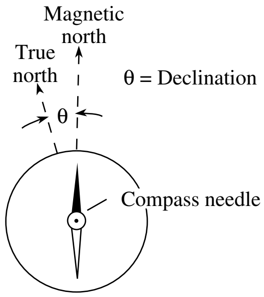

declination, 1

declinometer An instrument for measuring declination.

decode **1.** To unscramble a coded message. **2.** In digital computer operations, to deliver a specific output from character-coded inputs. **3.** In a multiplex system, the separation of the subcarrier from the main carrier.

decoder A circuit or device for performing DECODING.

decoder/demultiplexer A circuit that places an input signal on a selected output line.

decoder/driver An integrated circuit containing a decoder and driver.

decoding **1.** In computer and data-processing operations, DIGITAL-TO-ANALOG CONVERSION. **2.** The conversion to English of a message received in a code. **3.** Translating a message from a secret code (i.e., deciphering a message). **4.** The automatic conversion of a signal into the appropriate switching action (as the enabling of a transmitter or receiver by a tone in a selective calling system).

decoding circuit A circuit intended for the purpose of translating a code into ordinary language.

decollator An offline computer device for separating the parts of output continuous stationery sets. Also see CONTINUOUS STATIONERY.

decommutation The extraction of a signal component from the composite signal, resulting from commutation.

decommutator A circuit or device for performing decommutation, including demodulators, demultiplexers, and signal separators.

decoupler A device that isolates two circuits so that a minimal amount of coupling exists between them.

decoupling The elimination or effective minimization of coupling effects, as in decoupling amplifier stages to prevent interaction through a common power-supply lead.

decoupling capacitor **1.** A capacitor that provides a low-impedance path to ground to prevent undesired stray coupling among the circuits in a system. **2.** The capacitive member of a resistance-capacitance (RC) decoupling filter.

decoupling filter A resistance-capacitance (RC) filter, usually inserted into a common dc line in a multistage amplifier to prevent interstage feedback coupling through the common impedance of the line.

decoupling network One or more decoupling filters.

decoupling resistor The resistive member of a resistance-capacitance (RC) decoupling filter.

decoy In radar, an object that provides misleading reflections. Also see CHAFF.

decreasing function A function whose curve has a negative slope at all points in the domain.

decrement **1.** Also called *logarithmic decrement.* The rate at which a damped wave dies down. The decrement value is the natural (base-*e*) logarithm of the ratio of two successive peaks of the same polarity. **2.** A quantity used to lessen the value of a variable. **3.** To lower the value (of a register, for example) by a single increment.

decremeter An instrument for measuring the decrement of a radio wave.

decremeter capacitor A variable capacitor for use in a decremeter. The rotor plates are shaped so that equal angular rotations correspond to the same decrement at all settings. Thus, the percentage of capacitance change for a given angle of rotation is constant throughout the capacitance range.

decryption The conversion of an encrypted signal from a cipher into plain text, graphics, or other commonly recognizable form. Also see CIPHER. Compare ENCRYPTION.

decryption key An algorithm, or a set of algorithms, that converts an encrypted signal from a cipher into plain text, graphics, or other commonly recognizable form. Each cipher has its own unique algorithm or set of algorithms for this purpose. The signal cannot be decrypted unless all the components of the key are present.

dedicated Assigned exclusively to a certain purpose [e.g., a dedicated facsimile (fax) line].

deductive logic A form of symbolic logic used to demonstrate that a certain conclusion will always follow, given a certain set of circumstances. The logic of digital circuits is deductive. Compare INDUCTIVE LOGIC.

dee In a cyclotron, one of the D-shaped chambers in and between which particles accelerate in a spiral path to high velocity.

dee line In a cyclotron, a support for the dee, with which it forms a resonant circuit.

deemphasis In frequency modulation, the introduction of a low-pass characteristic (response falls as modulating frequency increases) to complement the rising response of preemphasis. Also called *postemphasis* or *postequalization.* Compare PREEMPHASIS.

deemphasis amplifier An amplifier used to remove the high-frequency preemphasis applied to signals prior to broadcasting, multiplexing, tape recording, or telemetering. Also see DEEMPHASIS and PREEMPHASIS.

deemphasis circuit A low-pass filter that provides deemphasis in an FM receiver.

deemphasis network See DEEMPHASIS CIRCUIT.

deenergize To take a circuit or device out of operation (i.e., to remove its power or signal excitation).

deep cycle Pertaining to a rechargeable cell or battery that can operate until it is almost completely discharged. It generally has a high ampere-hour capacity.

deep-diffused junction A pn junction made by diffusing the impurity material deep in the semiconductor wafer. Compare SHALLOW-DIFFUSED JUNCTION.

deep discharge The nearly complete discharge of a cell or battery; usually done prior to recharging.

deep-space net A radar system intended for constant monitoring of spacecraft.

defeating **1.** The disabling or circumvention of an alarm or security system, leaving the protected property vulnerable to intrusion. **2.** The dangerous, and potentially lethal, disabling of a safety device in an electrical or electronic system.

defect **1.** Absence of an electron (hence, presence of a hole) in the lattice of a semiconductor crystal. **2.** An abnormality of design, construction, or performance of an electronic circuit or device. **3.** In a computer system, a hardware or software fault that could be the eventual cause of a failure. **4.** A flaw in a crystalline substance.

defect conduction In a semiconductor material, conduction via holes.

deferred addressing Indirect addressing in which a preset counter makes several references to find a desired address.

deferred entry An entry into a computer subroutine, delayed because of a delay in the exit from a control program.

deferred exit An exit from a computer subroutine, delayed because of a particular command.

defibrillation Use of a CARDIAC STIMULATOR to halt fibrillation of the heart, as caused by electric shock.

defibrillator See CARDIAC STIMULATOR.

definite-purpose component A component designed for a specific use, rather than for a wide range of possible applications (e.g., a video detector diode, as opposed to a general-purpose diode). Compare GENERAL-PURPOSE COMPONENT.

definition **1.** Clarity of a video image (i.e., one having good contrast and faithful tones). **2.** Good intelligibility of reproduced sounds.

deflecting coil One of a set of external coils carrying sawtooth currents, which provide electromagnetic deflection of the cathode-ray beam in picture tubes, camera tubes, radar display tubes, sonar display tubes, and some oscilloscopes. Also called *deflection coil.*

deflecting electrode An electrode, such as a deflecting plate, used to alter the direction an electron beam. Also called *deflection electrode.*

deflecting plate In a cathode-ray tube, a plate that attracts or repels the electron beam, causing the spot to move horizontally or vertically on the screen. Also called *deflection plate.*

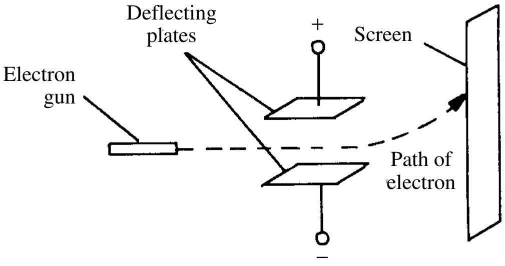

deflecting plate

deflecting torque The torque required to move the pointer of a meter, or the pen or mirror of a recorder.

deflection **1.** In a cathode-ray tube, movement of the electron beam by electric or magnetic fields. **2.** Movement of the pointer of a meter or the pen or mirror of a recorder by an applied current or voltage.

deflection factor Symbol, G. The reciprocal of DEFLECTION SENSITIVITY.

deflection plane In a cathode-ray tube, the plane perpendicular to the axis of the tube. This plane contains the electromagnetic and/or electrostatic lines of flux that result in deflection of the electron beam.

deflection coil See DEFLECTING COIL.

deflection electrode See DEFLECTING ELECTRODE.

deflection plate See DEFLECTING PLATE.

deflection polarity In a cathode-ray tube, the polarity of the voltage applied to a particular deflecting plate to move the electron beam in a particular direction.

deflection sensitivity Symbol, S. A quantitative measure of the extent to which the input voltage will displace the electron beam on the screen of an electrostatic cathode-ray tube. Expressed in volts per centimeter (V/cm) or volts per inch (V/in).

deflection voltage The potential difference between the deflection plates of an electrostatic cathode-ray tube. It is used to control the direction of the electron beam striking the phosphor screen.

deflection yoke An assembly of deflection coils in picture and camera tubes, and in some magnetically deflected oscilloscope tubes. The usual combination is two series-connected horizontal deflection coils and two series-connected vertical deflection coils.

deflector **1.** A beam-forming plate in a beam-power tube. **2.** A deflection plate in a cathode-ray tube. **3.** A deflection coil or yoke in a picture tube, camera tube, or magnetic-deflection oscilloscope tube. **4.** A mechanical attachment for improving the angle of radiation of a loudspeaker by spreading the higher-frequency waves.

defocusing Blurring of the image on the screen of a cathode-ray tube, caused by spreading of the electron beam.

deformation potential The voltage generated when a crystal lattice is subjected to pressure. An example is the voltage produced by a crystal microphone when acoustic waves strike the crystal.

defruiting The elimination of non-synchronized echoes in a radar system.

deg Abbreviation of DEGREE.

degassing During the evacuation of a vacuum tube or similar device, the removal of gas, including that which has bonded to the glass and metal parts.

degauss See DEMAGNETIZE.

degausser **1.** A circuit that performs DEGAUSSING. **2.** A device for bulk erasing magnetic tape; also called a *bulk tape eraser.*

degaussing **1.** The demagnetization of an object; in particular, the removal of all residual magnetism. **2.** The erasure of data from a magnetic or magneto-optical data-storage medium.

degaussing circuit In a color television receiver, a circuit including a thermistor, voltage-dependent resistor, and coil for automatically demagnetizing the picture tube when the receiver is switched on.

degaussing coil A coil carrying an alternating current; the resulting magnetic field demagnetizes objects that have become accidentally magnetized.

degeneracy In microwave practice, the appearance of a single resonant frequency for two or more modes in a resonator.

degenerate modes In microwave operations, a set of modes with the same resonant frequency or propagation constant.

degenerate parametric amplifier An inverting parametric amplifier, in which the two signals are of the same frequency, which is half the pump frequency.

degenerate semiconductor A semiconductor that behaves like a metal over a wide range of temperatures.

degeneration In an amplifier, the technique of feeding a portion of the output back to the input out of phase with the input signal, to improve fidelity at the expense of gain. Also called *negative feedback* or *inverse feedback.* Compare REGENERATION.

degenerative resistor An unbypassed emitter resistor in a common-emitter bipolar-transistor circuit, or an unbypassed source resistor in a common-source field-effect transistor circuit. Signal current flowing through the resistor produces negative feedback current (degeneration), which reduces the gain of the stage, but increases the linearity of the transfer characteristic.

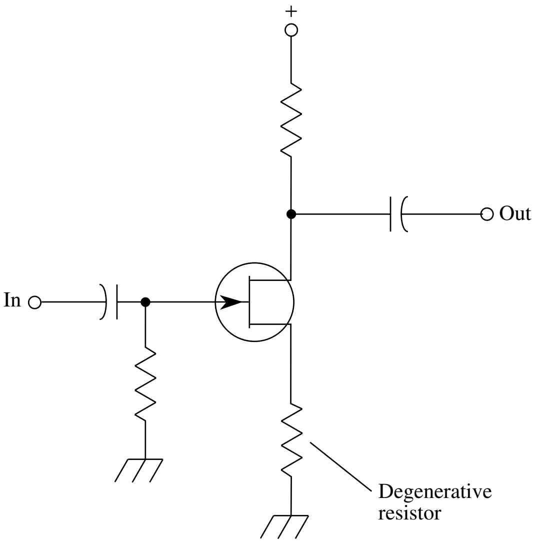

degenerative resistor

degradation **1.** Gradual deterioration in the condition or performance of a circuit or device. **2.** In a computer system, compromised performance caused by component failure.

degradation failure Failure occurring at the terminal point of degradation.

degraded operation See DEGRADATION.

degreaser See ULTRASONIC CLEANING TANK.

degree **1.** A unit of circular angular measurement equal to $1/360$ of the circumference of a circle. Also called GEOMETRIC DEGREE. **2.** A unit of temperature measurement. See DEGREE ABSOLUTE, DEGREE CELSIUS, DEGREE CENTIGRADE, DEGREE FAHRENHEIT, and DEGREE REAUMUR.

degree absolute Symbol, K. The unit of temperature on the absolute scale. Also see ABSOLUTE SCALE.

degree Celsius Symbol, °C. The unit of temperature on the CELSIUS SCALE.

degree centigrade Symbol, °C. The unit of temperature on the *centigrade scale* (now called CELSIUS SCALE).

degree Fahrenheit Symbol, °F. The unit of temperature on the FAHRENHEIT SCALE.

degree of current rectification For a rectifier, the ratio of the average direct output current to the root-mean-square (rms) alternating input current.

degree of voltage rectification For a rectifier, the ratio of the average direct-current (dc) output voltage to the root-mean-square (rms) alternating-current (ac) input voltage.

degree Reaumur Symbol, °R. The unit of temperature on the REAUMUR SCALE.

degrees of freedom **1.** The ways in which a point can move or a system can change. In three-dimensional space, a rigid body has six degrees of freedom: motion in three linear directions, and rotation around three linear axes extending through its center. **2.** The ways in which a robot arm can move, including linear motion and rotational motion.

degrees of rotation A measure of the extent to which a robot joint, or a set of robot joints, can be turned. Some reference axis is always used; angles are specified in degrees, relative to that axis.

degrees-to-radians conversion The conversion of angles in degrees to angles in radians. To change degrees to radians, multiply degrees by 0.01745. Compare RADIANS-TO-DEGREES CONVERSION.

deion circuit breaker A circuit breaker in which the arc occurring when the contacts open is quickly extinguished by an external magnetic device.

deionization The conversion of an ionized substance, such as a gas, to a neutral (non-ionized) state. The process changes the ions into uncharged atoms.

deionization potential The voltage at which an ionized substance becomes deionized; for example, the voltage at which a glow discharge is extinguished when the gas ions become neutral atoms at that voltage. Also called *extinction potential.*

deionization time The time required for an ionized gas to become neutral after the removal of the ionizing voltage.

deionization voltage See DEIONIZATION POTENTIAL.

deionize To restore to an electrically neutral condition (i.e., to convert ions to neutral atoms, as in the deionization of the gas when the discharge in a glow tube is extinguished).

deka- A prefix meaning ten(s) (e.g., DEKAMETER).

dekahexadecimal number system See HEXADECIMAL NUMBER SYSTEM.

delamination The splitting apart, in layers, of an insulating material, such as mica or bonded plastic film.

delay **1.** The interval between the instant at which a signal or force is applied or removed and the instant at which a circuit or device subsequently responds in a specified manner. **2.** The time required for a signal to traverse a given medium, such as air, mercury, or quartz.

delay action Response occurring some time after a stimulus has been applied or removed (e.g., the retarded opening of a delayed-dropout relay).

delay circuit **1.** A circuit, such as a resistance-capacitance (RC) or resistance-inductance (RL) combination, that introduces a time delay. **2.** See DELAY LINE.

delay coincidence circuit A coincidence circuit (see AND CIRCUIT) triggered by two pulses, one of which lags behind the other.

delay counter In a digital computer, a device that halts a program run long enough for an operation to be completed.

delay distortion **1.** Distortion resulting from variations in the phase delay of a circuit or device at different points in its frequency range. **2.** In a facsimile (fax) signal, variations in the delay of different frequency components of the signal.

delayed AGC See DELAYED AUTOMATIC GAIN CONTROL.

delayed automatic gain control An automatic gain control circuit that operates only when the signal amplitude exceeds a predetermined threshold level, thus providing maximum amplification of weaker signals.

delayed automatic volume control See DELAYED AUTOMATIC GAIN CONTROL.

delayed break In relay or switch operation, contacts separating some time after the switch has been thrown or the relay deenergized. Compare DELAYED MAKE.

delayed close See DELAYED MAKE.

delayed closure See DELAYED MAKE.

delayed contacts Contacts that open or close at a predetermined instant after their activating signal is applied or removed.

delayed drop-in See DELAYED MAKE.

delayed dropout See DELAYED BREAK.

delayed loop In security applications, a circuit or system that registers an alarm some time after intrusion is first detected. The delay can usually be selected or preadjusted.

delayed make In relay or switch operation, contacts closing some time after the switch has been thrown or the relay has been energized. Compare DELAYED BREAK.

delayed open See DELAYED BREAK.

delayed PPI Plan-position indicating radar having a delayed time base.

delayed pull-in See DELAYED MAKE.

delayed repeater A repeater that receives and stores information, and retransmits the information later, in response to a switching or interrogation signal.

delayed repeater satellite An active communications satellite that acts as a delayed repeater (i.e., it receives and records information at one time and retransmits it at a later time).

delayed sweep **1.** In an oscilloscope or radar, a sweep that starts at a selected instant after the signal under observation has started. **2.** The (usually calibrated) circuit for producing a sweep, as defined in (**1**).

delayed updating Updating a computer record or record set so that the record fields are left unchanged until all other changes attendant to the pertinent event are processed.

delay equalizer A network that corrects DELAY DISTORTION.

delay-frequency distortion Distortion caused by variation of envelope delay within a frequency band.

delay line A device (not always a line) that introduces a time lag in a signal. The lag is the time required for the signal to pass through the device, minus the time necessary for the signal to traverse the same distance through a wire, cable, optical fiber, or free space.

delay-line memory In a digital computer, a memory that uses a delay line, associated input- and output-coupling devices, and an external regenerative-feedback path. Information is kept stored by causing it to recirculate in the line by regeneration.

delay-line register In a digital computer, a register that operates in the manner of a DELAY-LINE MEMORY and has a register length (capacity) of an integral number of words.

delay-line storage See DELAY-LINE MEMORY and DELAY-LINE REGISTER.

delay multivibrator See MONOSTABLE MULTIVIBRATOR.

delay-power product Unit, watt-second. The figure of merit for an integrated circuit (IC) gate. Increasing gate power reduces propagation delay. Also called PROPAGATION DELAY-POWER PRODUCT.

delay relay A relay that opens or closes at the end of a predetermined time interval.

delay switch A switch having delayed make, delayed break, or both.

delay time **1.** The interval between the instant a voltage or current is applied and the instant a circuit or device operates. **2.** In an output pulse, the interval between the instant an ideal pulse is applied to the input of a system and the instant the output pulse reaches 10% of its maximum amplitude. **3.** The time elapsed between the presentation of a pulse to the input of a delay line and the appearance of the pulse at the output.

delay timer **1.** A timer that starts or stops an operation after a prescribed length of time. **2.** A delay relay or switch.

delay unit In a radar system, a circuit for delaying pulses.

delete **1.** To erase or blank out a signal. **2.** The elimination from a computer file of a record or record group. **3.** To remove a computer program from memory or storage.

deletion record In the master file of a digital computer, a new record that causes existing ones to be deleted.

delimiter In digital computer operations, a character limiting a sequence of characters of which it is not itself a member.

Dellinger effect The sudden disappearance of a radio signal as a result of an abrupt increase in atmospheric ionization caused by a solar eruption.

deliquescent material A material that absorbs enough moisture from the air to get wet. For example; calcium chloride, a deliquescent material, is often used to keep electronic equipment dry. Compare HYGROSCOPIC MATERIAL.

delta circuit A three-phase electrical circuit with no common ground.

delta connection A triangular connection of coils or load devices in a three-phase system, so called from its resemblance to the Greek letter delta. Compare WYE-CONNECTION.

delta-matched antenna See WYE-MATCHED IMPEDANCE ANTENNA.

delta-matched impedance antenna See WYE-MATCHED IMPEDANCE ANTENNA.

delta matching transformer In a WYE-MATCHED IMPEDANCE ANTENNA, the fanned-out (roughly delta-shaped) portion of the two-wire feeder at its point of connection to the radiator. It matches the impedance of the feeder to that of the radiator.

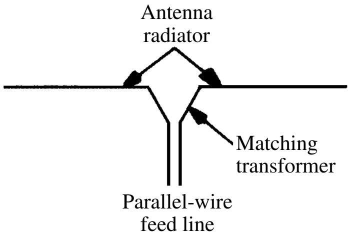

delta-matching transformer

delta modulation The conversion of an analog signal into a digital pulse train that can be decoded to yield the original analog signal.

delta network See DELTA CONNECTION.

delta pulse-code modulation In wire or radio communications, the conversion of an audio signal into a digital pulse train.

delta quantity An increment (i.e., the difference between two values of a variable).

delta rays The emission of secondary electrons as a result of radioactivity.

delta-sigma modulation A method of analog-to-digital conversion. The output is a pulse density function of the input. The input can be obtained by low-pass filtering of the output.

delta tune Also called *receiver incremental tuning (RIT).* In high-frequency (HF) communications transceivers, a control that allows the receiver frequency to be adjusted up to several kilohertz higher or lower than the transmitter frequency.

delta waves Brain waves having a frequency less than 9 Hz. Also see ELECTROENCEPHALOGRAPH and ELECTROENCEPHALOGRAM.

Deluc's pile See DRY PILE.

dem Abbreviation of DEMODULATOR.

demagnetization curve The portion of a magnetic hysteresis curve, showing reduction of demagnetization.

demagnetization effect The phenomenon in which uncompensated magnetic poles at the surface cause a reduction of the magnetic field inside a sample of a material.

demagnetize To remove magnetism from an object, either temporarily or permanently.

demagnetizer See DEGAUSSER.

demagnetizing current The half-cycle of an alternating current (or polarity of a direct current) flowing through a coil wound on a permanent magnet (as in a headphone, permanent-magnet loudspeaker, or polarized relay), that reduces the magnetic field.

demagnetizing force **1.** A magnetic force whose direction reduces the residual induction of a magnetized material. **2.** An effect that reduces the magnetism of a permanent magnet, such as high temperature or a physical blow.

demand factor In the use of electric power, the ratio of the consumer's maximum demand to the actual power consumed.

demand processing Descriptive of a system that processes data as it is available, without storing it.

demarcation strip An interface between a terminal unit and a carrier line.

Dember effect The appearance of a voltage between regions in a semiconductor when one of the regions is illuminated.

demodulation The process of retrieving the information (modulation) from a modulated carrier. In receivers and certain test instruments, this process is called DETECTION.

demodulator **1.** A circuit that recovers the information from a modulated analog or digital signal. In radio communications, such a device is usually called a DETECTOR. **2.** In computer communications, a device that performs ANALOG-TO-DIGITAL CONVERSION of incoming online signals.

demand read (write) Inputting or outputting data blocks to or from a central processor, as needed for processing.

demodulator probe A diode probe that removes the modulation envelope from an applied amplitude-modulated signal, and presents the envelope to a voltmeter or oscilloscope.

demonstrator A device used to show and teach the way in which a component, circuit, or system operates.

DeMorgan's theorem A rule of sequential or digital logic. It states that the negation of (A AND B), for any two statements A AND B, is equivalent to NOT A OR NOT B. Also, the negation of (A OR B) is equivalent, logically, to NOT A AND NOT B.

demultiplexer A circuit or device that separates the components of a multiplexed signal transmitted over a channel.

demultiplexing circuit See DEMULTIPLEXER.

denary band A band in which the highest frequency is 10 times the lowest frequency.

dendrite **1.** The branching (tree-like) structure formed by some materials, such as semiconductors, as they crystallize. **2.** The branching portion of a nerve cell; hence, the corresponding circuit element in the electronic model of such a cell.

dendritic growth **1.** Dendrite (see DENDRITE, **1**). **2.** The process of growing long, flat semiconductor crystals.

dendron See DENDRITE, **2**.

dens Abbreviation of DENSITY.

dense binary code A binary representation system, in which any possible combination of characters is assigned some correspondent.

densitometer An instrument for measuring the density of a body.

density **1.** Mass per unit volume of a material. **2.** Concentration of charge carriers or of lines of flux. **3.** The number of items per unit volume, area, distance, or time.

density modulation Modulation of the density, with respect to time, of electrons in an electron beam.

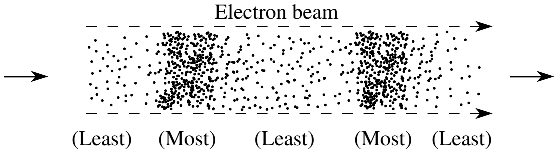

density modulation

density of electrons The concentration of electrons (i.e., the number per unit volume, area, distance, or time).

density packing A figure indicating the quantity of bits per inch or per centimeter, stored on a magnetic tape.

dependent equations Equations that are alike and have an infinite number of solutions. Compare INDEPENDENT EQUATIONS and INCONSISTENT EQUATIONS.

dependent linearity Linearity (especially in its deviation from an ideal slope) as a dependent variable.

dependent variable A changing quantity whose value at any instant is governed by the value at that instant of another changing quantity (the independent variable). Compare INDEPENDENT VARIABLE.

depletion-enhancement-mode MOSFET A metal-oxide-semiconductor field-effect transistor (MOSFET) designed for zero gate-bias voltage. An ac gate signal voltage drives the MOSFET alternately into the depletion mode (negative signal half-cycle) and enhancement mode (positive signal half-cycle). Compare DEPLETION-TYPE MOSFET and ENHANCEMENT-TYPE MOSFET.

depletion field-effect transistor A field-effect transistor whose operation is based on the control of depletion layer width.

depletion layer See BARRIER, **1**.

depletion-layer capacitance See JUNCTION CAPACITANCE.

depletion-layer rectification Rectification provided by a semiconductor junction.

depletion-layer transistor A transistor whose action depends on modulation of current carriers in a space-charge region (depletion layer).

depletion mode Operation characteristic of the DEPLETION-TYPE MOSFET.

depletion region See BARRIER, **1**.

depletion-type MOSFET A metal-oxide-semiconductor field-effect transistor (MOSFET) in which the channel directly under the gate electrode is narrowed by a negative gate voltage (in an n-channel device) or by a positive gate voltage (in a p-channel device).

depolarization **1.** In a primary cell, the removal of the agents that have caused polarization. **2.** The addition of a polarization-inhibiting substance to the electrolyte of a primary cell.

depolarizer A substance that retards polarization in an electrochemical cell. An example is the manganese dioxide used in dry cells.

depolarizing agent See DEPOLARIZER.

deposition The application of a layer of one substance (usually a metal) to the surface of another (the substrate), as in evaporation, sputtering, electroplating, silk-screening, etc.

depth finder See ACOUSTIC DEPTH FINDER.

depth indicator **1.** A sounding instrument for determining the depth of a body of water. **2.** On an ACOUSTIC DEPTH FINDER, the meter that indicates the depth of water.

depth of cut On a phonograph disk, the depth of the recorded groove.

depth of discharge Abbreviation, DOD. In a rechargeable cell or battery, a measure of the extent to which discharging has occurred. It is generally specified as a percentage. For example, if the DOD of a 10-ampere-hour (10-AH) battery is 80 percent, then 8 AH have been used up, and 2 AH remain before recharging will be necessary.

depth of heating In dielectric heating, the depth of heat penetration in the sample when both electrodes are applied to one of its faces.

depth of modulation The degree to which a carrier wave is modulated.

depth of penetration The extent to which a skin-effect current penetrates the surface of a conductor.

depth sounder See ACOUSTIC DEPTH FINDER.

de-Q **1.** To reduce the *Q* of a component or tuned circuit. **2.** To inhibit laser action during an interval when an ion population excess is pumped up.

derating To reduce an operating parameter (e.g., current, voltage, power) as another factor (such as temperature) increases, to ensure efficient, reliable, and safe operation.

derating curve A graph that shows the extent to which a quantity (such as allowable power dissipation) must be reduced as another quantity (such as temperature) increases.

derating factor The amount by which a current, power, or voltage must be decreased to ensure safe and efficient operation of a circuit or device in a given environment (temperature, altitude, humidity, etc.). Also see DERATING and DERATING CURVE.

derivative **1.** A mathematical expression indicating the rate at which a function changes, with respect to the independent variable. See DERIVATIVE FUNCTION. **2.** The slope of a line tangent to a curve at a given point. **3.** The output signal of a DIFFERENTIATOR, relative to the input signal.

derivative action In a control system, an action producing a corrective signal proportional to the rate of change (derivative) of the controlled variable.

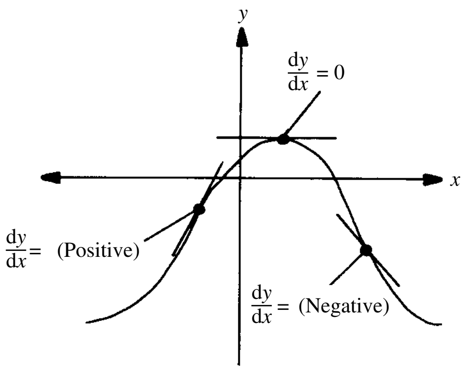

derivative

derivative control A method of automatic control, actuated according to the number of errors per second.

derivative function For a mathematical function $f(x)$, the function $f'(x) = df(x)/dx$, over the domain of f. For any specific point x_0 in the domain of f, the value of $f'(x_0)$ is equal to the slope of a line tangent to f at the point $(x_0, f(x_0))$.

derived center channel The sum or difference of the left and right channels in a stereophonic system.

Dershem electrometer A variation of the *quadrant electrometer.* In the Dershem instrument, the needle (to which a small mirror is attached) rotates within slots cut in the quadrant plates and, therefore, can never accidentally touch the plates.

descending node For a satellite orbiting the earth or another planet, any point at which the groundtrack crosses the equator as it moves from the northern hemisphere into the southern hemisphere. This node generally changes for each succeeding orbit, because the earth or planet rotates underneath the orbit of the satellite. Compare ASCENDING NODE.

descending pass For a specific point on the earth's surface, the time during which an artificial communications satellite is accessible when its latitude is moving southward. The duration of accessibility depends on the altitude of the satellite and on how close its groundtrack comes to the earth-based point. Compare ASCENDING PASS.

description A data element that is part of a record and is used to identify it.

desensitization **1.** The process of making a circuit or device less responsive to small values of a quantity. **2.** Also called *desensing.* In a communications receiver, an unwanted, often intermittent reduction in front-end gain, caused by an extremely strong local signal.

desensitize **1.** To reduce the sensitivity of a receiver. **2.** To reduce the gain of an amplifier. **3.** To reduce the small-quantity response of an instrument.

desiccant A compound, such as cobalt chloride, used for the purpose of keeping enclosed items dry.

design **1.** A unique, planned arrangement of electronic components in a circuit, in accordance with good engineering practice, to achieve a desired end result. **2.** A unique layout of components or controls, in accordance with good engineering practice, esthetics, and (often) ergonomics. **3.** Invention. **4.** Plan. **5.** To produce a design, as defined in **1**, **2**, **3**, or **4**.

designation Within a computer record, coded information identifying the record so that it can be handled accordingly.

design-center rating A specified parameter that, if not exceeded, should provide acceptable average performance for the greatest number of the components so rated.

design compatibility The degree to which a transmitter and receiver are designed for the rejection of unwanted electromagnetic noise.

design engineer An engineer who is skilled in the creation of new designs and in the comparative analysis of designs.

design-maximum rating See MAXIMUM RATING.

design-proof test A performance test made on a newly completed circuit or device to determine the suitability of the design.

Desk-Fax A facsimile transceiver that can be placed on a desktop, used for wire or radio transmission and reception of still images.

desk microphone A microphone equipped with a stand that sits on a table or desktop. It allows the operator to use both hands for equipment adjustment, taking notes, etc.

desktop computer A personal computer designed for nonportable use, usually equipped with a built-in hard disk, one or more diskette drives, a CD-ROM drive, and a fax/modem. It generally uses an external cathode-ray-tube display, printer, and keyboard. The power supply is intended for use with 117-volt utility circuits.

desolder To unsolder joints, usually with a special tool that protects delicate parts and removes melted solder by suction.

destaticization A chemical process used to minimize the retention of electrostatic charges by certain substances.

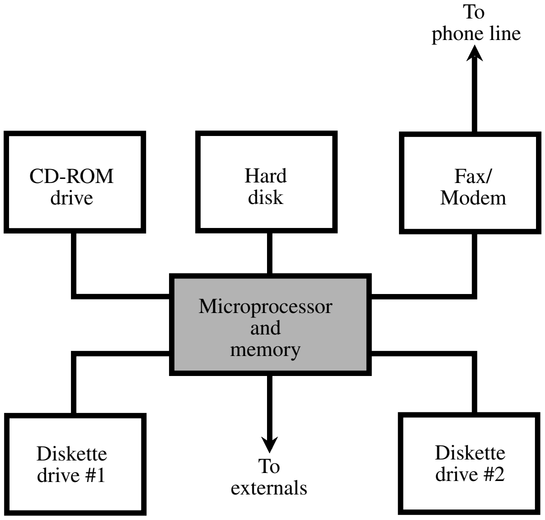

desktop computer (main unit only)

destination **1.** The point in a system to which a signal of any sort is directed. **2.** In communications, a receiving station.

destination file A computer file that receives data output during a specific program run.

destination register In a digital computer, a register into which data is entered.

Destriau effect Light emission resulting from the action of an alternating electric field on phosphors embedded in a dielectric.

destructive addition A computer logic operation in which the sum of two operands appears in the memory location occupied by one of the operands.

destructive breakdown A breakdown in which the effects are irreversible (e.g., permanent damage to a dielectric by excessive applied voltage).

destructive interference Interference resulting from the addition of two waves that have the same frequency, but opposite phase.

destructive read In a computer or calculator, the condition in which reading the answer erases the data (as from a location) used in the calculation.

destructive test A test that unavoidably destroys the test sample. Compare NONDESTRUCTIVE TEST.

DETAB A COBOL-based computer programming language permitting the programmer to present problems as decision tables.

detail constant Pertaining to a video signal, the ratio V_H/V_L, where V_H is the amplitude of high-frequency components, and V_L is the amplitude of the low-frequency reference component.

detected error In a computer system, an error that is identified, but remains uncorrected until final output is available.

detection **1.** See DEMODULATION. **2.** The sensing of a change in the operating parameters of a circuit or system.

detection range In security applications, the radius within which transducers or sensors can be expected to reliably operate. This radius varies, depending on the environment, the sensitivity of the receiving circuits and transducers, and the strength of the transmitted signal (if any).

detectophone A device for eavesdropping on a conversation. The device can use a tape recorder or a tiny radio transmitter.

detector **1.** In radio communications, a device or circuit that extracts the information from a modulated carrier. Also sometimes called a *demodulator.* **2.** A device that senses a signal or condition and indicates its presence.

detector balanced bias In a radar system, bias obtained from a controlling circuit and used to reduce or eliminate clutter.

detector bias Steady dc voltage applied to a detector to set its operating point.

detector blocking In a regenerative receiver, a phenomenon in which a strong signal tends to pull the detector oscillator into phase with itself, thereby causing the detector to oscillate at the signal frequency.

detector circuit A demodulator circuit (i.e., one used to recover the intelligence from a modulated carrier).

detector probe See DEMODULATOR PROBE.

detector pull-in See DETECTOR BLOCKING.

detector stage In a receiver or instrument, the separate stage that contains the detector circuit. Some systems, such as a superheterodyne receiver, have more than one detector. Also see FIRST DETECTOR and SECOND DETECTOR.

detent A mechanical stop used on a rotary switch to hold the switch pole securely in each selected position.

detune **1.** To adjust a circuit to some frequency other than its resonant frequency. **2.** To set the frequency of a receiver or transmitter to some point other than the frequency normally used. **3.** To stagger-tune a receiver intermediate-frequency system.

detuning Tuning to a point above or below the frequency to which a device or system is normally (or initially) adjusted (usually the resonant frequency of the device).

detuning stub A device used for the purpose of coupling a feed line to an antenna, while choking off currents induced on the feed line as a result of the near-field radiation of the antenna.

deupdating Producing an earlier form of a computer file by substituting older records for current ones.

deuterium Symbol, D, d, H^2, or ^{2}H. Also called *heavy hydrogen.* The hydrogen isotope having a nucleus consisting of one proton and one neutron.

deuterium oxide Symbol, D_2O. Also called *heavy water.* This compound has wide use in nuclear reactors.

deuteron The nucleus of a deuterium atom.

deuton See DEUTERON.

deutron See DEUTERON.

deviation **1.** In a frequency-modulated (FM) radio signal, the instantaneous amount of carrier frequency shift away from the unmodulated frequency. It is usually expressed in kilohertz; directly proportional to the amplitude of the modulating signal, up to a certain maximum that depends on the bandwidth allowed. **2.** The maximum instantaneous carrier frequency shift in a FM signal. **3.** The extent or amount by which a quantity drifts from its proper value.

deviation distortion In a frequency-modulation (FM) receiver, distortion resulting chiefly from discriminator nonlinearity and restricted bandwidth.

deviation ratio In a frequency-modulated (FM) signal, the ratio between the highest modulating frequency and the maximum carrier deviation.

deviation sensitivity For a frequency-modulation (FM) receiver, the smallest deviation that will produce a specified audio output power. Expressed in kilohertz, or as a percentage of rated deviation of the receiver, measured with the receiver set for maximum gain.

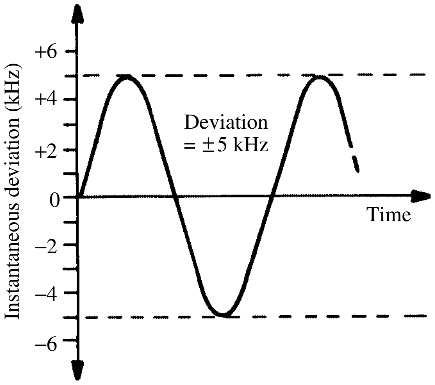

deviation

device **1.** A simple or complex discrete electronic component. **2.** A subsystem used as a unit, and regarded as a single component.

device complexity The number of components in an integrated circuit.

device independence A characteristic of a computer, that allows operation independent of the types of input/output devices used.

dew point For a gas containing water vapor (typically air), the highest temperature at which the vapor condenses as the gas is cooled. The dew point depends on the amount of vapor in the gas.

dew-point recorder An instrument for determining and recording the temperature at which water vapor in the air condenses to a liquid.

DF Abbreviation of DIRECTION FINDER.

DF antenna An antenna that is mechanically rotatable or has an electrically rotatable response pattern for use with a direction finder.

DF antenna system Two or more DF antennas arranged for maximum directivity and maneuverability, together with associated feeders and couplers.

D flip-flop A delayed flip-flop. The state of the input determines the state of the output during the following pulse, rather than during the current pulse.

dg Abbreviation of *decigram.*

dia Abbreviation of *diameter.*

diac A two-terminal, bilateral, three-layer semiconductor device that exhibits negative resistance. When the applied voltage exceeds a critical value, the device conducts.

diagnosis **1.** Determination of the cause and location of a hardware malfunction. **2.** In computer operations, determination of the cause of a system operation error.

diagnostic routine **1.** An efficient sequence of diagnostic tests for rapid, foolproof trouble-shooting of electronic hardware. **2.** A computer software package intended for debugging programs, or for finding the cause of a hardware or operating-system malfunction. Also called *diagnostic*, *diagnostic program*, or *diagnostic utility.*

diagnostic test **1.** A test made primarily to ascertain the cause of dysfunction in electronic equipment. Compare PERFORMANCE TEST. **2.** To apply a diagnostic routine to hardware faults, or to implement one to prevent such a fault.

diagnotor In digital computer operations, a troubleshooting routine combining both diagnosis and editing.

diagram A (usually line) drawing depicting a circuit, assembly, or organization. See, for example, BLOCK DIAGRAM and CIRCUIT DIAGRAM.

dial **1.** A graduated scale, arranged horizontally, vertically, in a circle, or over an arc. Used to show the distance through which a variable component (such as a potentiometer, variable capacitor, or switch) has been adjusted. A pointer can move over the scale, or the scale can be moved past a stationary pointer. **2.** The graduated face of a meter. **3.** In a telephone system, to press the keys or actuate the tones that establish contact with another subscriber.

dial cable A flexible cable or belt conveying motion on the shaft of an adjustable component (such as a potentiometer or variable capacitor) to a dial.

dial-calibrated attenuator A variable attenuator with a dial reading directly in decibels.

dial-calibrated capacitor A variable capacitor with a dial reading directly in picofarads.

dial-calibrated inductor A variable inductor with a dial reading directly in microhenrys.

dial-calibrated potentiometer A potentiometer with a dial reading directly in output volts, percentage of input voltage, number of turns (when resistance is a linear function), or other quantity.

dial-calibrated resistor A variable resistor with a dial reading directly in ohms, kilohms, or megohms.

dial-calibrated rheostat See DIAL-CALIBRATED RESISTOR.

dial cord A form of dial cable. Cord usually designates a fabric string, whereas a cable is a flexible, braided wire.

dial knob The knob used to turn a dial under a pointer, or to turn a pointer over a dial scale.

dial lamp See DIAL LIGHT.

dial light A small lamp sometimes used to illuminate a dial. Can also serve as a pilot light.

dial lock A small mechanism used to lock a dial at a particular setting to prevent further turning.

dialer See AUTOMATIC DIALING UNIT.

dialing key In a telephone system, a dial that uses keys, rather than a rotary dial.

dial jack In a telephone system, a set of jacks that facilitates interconnections between dial cords and external lines.

dial light A lamp or light-emitting diode placed in the dial mechanism of a radio receiver, transmitter, or transceiver. Allows the dial to be read in dim light or in darkness.

dialog equalizer In sound transmission and recording, a high-pass filter that reduces low-frequency response during dialog and extreme closeups.

dial pulse An interruption of the direct current in a telephone system when the dial contacts of the calling telephone open. The number of such interruptions corresponds to the digit dialed.

dial scale The graduated portion of a dial.

dial system **1.** See DIAL TELEPHONE SYSTEM. **2.** The arrangement of dials and knobs that facilitates adjustment of electronic equipment.

dial telephone A telephone set in which a numbered rotatable disk is used to produce the switch interruptions that cause generation of the transmitted multidigit telephone numbers.

dial telephone system The complete automatic circuit, including central-office facilities, for dial telephone operation.

dial tone In a telephone system, a constant hum or whine heard before dialing, indicating that the system is operational.

dial-up In a telephone system, the calling of one subscriber by another, using a dial system.

diam Abbreviation of *diameter.*

diamagnetic Pertaining to a material having magnetic permeability less than unity.

diamagnetism The state of having magnetic permeability less than unity. A material with this property reduces the flux density of a magnetic field, relative to the flux density in air or in free space.

diamond antenna Also called *rhombic antenna.* A nonresonant wideband directional antenna whose horizontal wire elements are arranged in the shape of a diamond (rhombus). The arrangement is fed at one corner, the opposite corner being terminated with a noninductive resistor.

diamond lattice The orderly internal arrangement of atoms in a redundant pattern in crystalline materials, such as germanium or silicon.

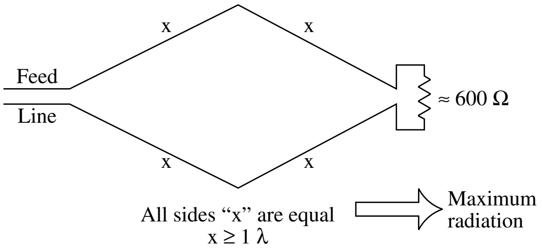

diamond antenna
]

diamond stylus A phonograph "needle" having as its point a small, ground diamond.

diapason **1.** Either of the two principal stops (open and closed) of an electronic organ that cover the entire range of the instrument. When one is used, a note played is automatically sounded in several octaves. **2.** Tuning fork.

diaphony See DISSONANCE.

diaphragm A usually thin metal or dielectric disk used as the vibrating member in headphones, loudspeakers, and microphones, and as the pressure-sensitive element in some sensors and barometers.

diaphragm gauge A sensitive gas pressure gauge using a thin metal diaphragm stretched flat. Increments of pressure move the diaphragm, relative to a nearby electrode, varying the capacitance between the two.

diathermic Pertaining to a substance that efficiently transfers heat or infrared energy.

diathermotherapy The use of diathermy in the treatment of various physiological disorders.

diathermy **1.** In medicine and physical therapy, the production of heat in subcutaneous (below the skin) tissues by means of high-frequency radio waves. **2.** A radio-frequency (RF) power oscillator and associated equipment used to produce heat in subcutaneous tissues.

diathermy interference Radio-frequency interference (RFI) resulting from the operation of unshielded and/or unfiltered diathermy equipment.

diathermy machine See DIATHERMY, **2**.

diatomic Having two atoms (e.g., a DIATOMIC MOLECULE).

diatomic molecule A molecule (such as that of oxygen) composed of two atoms. Compare MONATOMIC MOLECULE.

dibble A mathematical function in which a number (usually an integer) is doubled, and then one is added to the result. Thus, dibble $n = 2n + 1$.

dibit A combination of two binary digits (bits). The four possible dibits are 00, 01, 10, and 11.

dice Plural of DIE, **1**, **3**.

dichotomizing search Also called *binary search.* In digital computer operations, locating an item in a table of items that are arranged by key values in serial order. The required key is compared with a key halfway through the table; according to this relational test, half of the table is accepted and again divided for comparison, etc. until the keys match and the item is found.

dichotomy Characterized by the usually repetitive branching into two sets, groups, or factions.

dichroism Also called *dichromatism.* **1.** The property of a crystal showing different colors, depending on which axis corresponds to the line of sight. **2.** The property of a solid taking on different colors as the thickness of the transmitting layer changes. **3.** The property of a liquid changing color, according to solution concentration.

dichromate cell An electrolytic cell consisting of electrodes of carbon and zinc. The zinc electrode is immersed in a diluted solution of sulfuric acid, and the carbon electrode in a solution of potassium dichromate.

dicing The cutting of a semiconductor melt, crystal wafer, or other material into dice (see DIE).

dictionary A table of specifications for the size and format of computer file operands, and data names for field and file types.

die **1.** A small wafer of useful electrical material, such as a semiconductor or a precision resistor chip. **2.** A casting designed to mold molten metal into a specific configuration until the metal hardens. **3.** Any small object of roughly cubical proportions. **4.** To lose power or energy completely, usually unintentionally. **5.** In a computer program, to produce unpredicted and useless results following an initial run.

die bonding The bonding of dice or chips to a substrate.

die casting Making a casting by forcing molten metal (such as an aluminum alloy, lead, tin, or zinc) under high pressure into a die or mold.

dielectric A material that is a nonconductor of electricity; especially, a substance that facilitates the storage of energy in the form of an electric field. Such materials are commonly used in capacitors and transmission lines.

dielectric absorption The ability of certain dielectric materials to retain some of their electric charge—even after being momentarily short-circuited. Capacitors with this property must be shorted out continuously for a certain length of time before the dielectric has completely discharged.

dielectric amplifier A voltage amplifier circuit in which the active component is a capacitor having a nonlinear dielectric. A signal voltage applied to the capacitor varies the capacitance, thus varying the current. The modulated current flows through a load resistor, developing an output-signal voltage higher than the input-signal voltage.

dielectric antenna An antenna in which some or all of the radiating element is made of a dielectric material, such as polystyrene. Primarily used at microwave frequencies.

dielectric breakdown Sudden, destructive conduction through a dielectric when the applied voltage exceeds a critical value.

dielectric breakdown voltage The voltage at which DIELECTRIC BREAKDOWN occurs in an insulating material. Varies, depending on the particular dielectric substance.

dielectric capacity See DIELECTRIC CONSTANT.

dielectric constant Symbol, k. For a dielectric material, the ratio of the capacitance of a two-plate capacitor using the dielectric material, to the capacitance of the equivalent capacitor with dry air as a dielectric. Also called *inductivity* and *specific inductive capacity.*

dielectric current **1.** Current flowing over the surface of a dielectric material in response to a varying electric field. **2.** Current flowing through a dielectric as a result of its finite insulation resistance.

dielectric dissipation For a dielectric material in which an electric field exists, the ratio of the lost (dissipated) electrical energy to the recoverable electrical energy.

dielectric dissipation factor The cotangent of the dielectric phase angle, also equal to the reciprocal of the Q factor.

dielectric fatigue In some dielectric materials subjected to a constant voltage, the deterioration of dielectric properties with time.

dielectric guide A waveguide made from a solid dielectric, such as polystyrene.

dielectric heater A high-frequency power generator used for DIELECTRIC HEATING.

dielectric heating The heating and forming of a dielectric material, such as a plastic, by temporarily making the material the dielectric of a two-plate capacitor. This capacitor is connected to the output of a high-power radio-frequency (RF) generator. Losses in the dielectric cause its heating. Compare INDUCTION HEATING.

dielectric hysteresis See DIELECTRIC ABSORPTION.

dielectric isolation In a monolithic integrated circuit (IC), the isolation of circuit elements from each other by a dielectric film, as opposed to isolation by reverse-biased pn junctions.

dielectric lens A molded piece of dielectric material used to focus microwaves. Its operation is analogous to that of an optical lens.

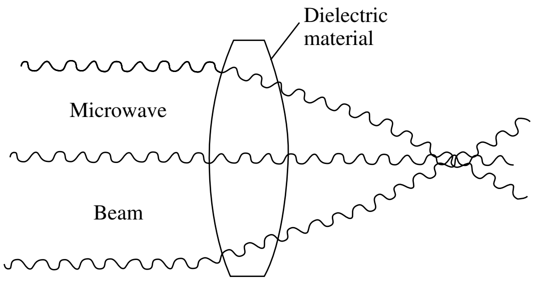

dielectric lens

dielectric loss For a dielectric material subjected to a changing electric field, the rate of transformation of electric energy into heat.

dielectric loss angle Ninety degrees minus the DIELECTRIC PHASE ANGLE.

dielectric loss factor For a dielectric material, the product of the dielectric constant and the tangent of the dielectric loss angle.

dielectric loss index See DIELECTRIC LOSS FACTOR.

dielectric matching plate A dielectric plate used in some waveguides for impedance matching.

dielectric mirror A reflector containing a number of layers of dielectric material. Its action depends on electromagnetic energy being partially reflected from the interfaces between materials having unequal indexes of refraction.

dielectric phase angle For a dielectric material, the angular phase difference between a sinusoidal voltage applied to the material and the component of the resultant current having the same period as that of the voltage.

dielectric phase difference See DIELECTRIC LOSS ANGLE.

dielectric polarization The effect characterized by the slight displacement of the positive charge in each atom of a dielectric material, with respect to the negative charge, under the influence of an electric field.

dielectric power factor The cosine of the dielectric phase angle, or the sine of the dielectric loss angle.

dielectric puncture voltage See DIELECTRIC BREAKDOWN VOLTAGE.

dielectric rating The breakdown voltage, and sometimes the power factor, of the dielectric material used in a device, such as a relay, motor, or switch.

dielectric ratings Electrical characteristics of a dielectric material: breakdown voltage, power factor, dielectric constant, etc.

dielectric resistance See INSULATION RESISTANCE.

dielectric rigidity See DIELECTRIC STRENGTH.

dielectric-rod antenna A unidirectional antenna that uses a dielectric substance to obtain power gain.

dielectric soak See DIELECTRIC ABSORPTION.

dielectric strain The distorted internal state of a dielectric, caused by the influence of an electric field. Also called DIELECTRIC STRESS.

dielectric strength The highest voltage a dielectric can withstand before DIELECTRIC BREAKDOWN occurs. Usually expressed in volts or kilovolts per mil of material thickness.

dielectric stress The distortion of electron orbits in the atoms of a dielectric material subjected to an electric field.

dielectric susceptibility For a polarized dielectric, the ratio of polarization to electric intensity.

dielectric tests Laboratory experiments performed to determine the dielectric characteristics of a substance—especially the dielectric constant and the dielectric breakdown voltage.

dielectric waveguide See DIELECTRIC GUIDE.

dielectric wedge A wedge-shaped dielectric slug placed inside a waveguide for impedance matching.

dielectric wire A small dielectric waveguide that acts as a wire to carry signals between points in a circuit.

Dietzhold network A four-terminal, shunt m-derived circuit used in some wideband amplifiers.

Dietzhold peaking In some wideband amplifiers, frequency compensation obtained with a shunt m-derived network (see DIETZHOLD NETWORK).

difference amplifier See DIFFERENTIAL AMPLIFIER.

difference channel In a stereophonic amplifier, an audio channel that handles the difference between signals in the right channel and those in the left channel.

difference detector A detector whose output is the difference between two simultaneous input signals.

difference frequency A signal frequency produced by mixing or heterodyning of signals at two other frequencies. If the lower input signal frequency is f_1 and the higher input signal frequency is f_2, then the difference frequency f_d is equal to $f_2 - f_1$.

difference of potential The absolute value of the algebraic difference of voltages at two points of different electrical potential. Thus, the difference of potential between a +5-V point and a –5-V point is +5 – (–5) V = 10 V. Also see POTENTIAL DIFFERENCE.

difference quantity See INCREMENT.

difference signal **1.** The resultant signal obtained by subtracting, at every instant for at least one full cycle, the amplitudes of two signals. **2.** The difference of the left- and right-channel outputs in a stereo system.

differential **1.** A device, consisting of a gear system, that adds or subtracts angular motions and delivers the result. **2.** A gear system in which the motion of a shaft is transferred to two other shafts aligned with each other and perpendicular to the first shaft. **3.** One of two coils arranged to produce opposite polarities at a point in a circuit. **4.** Pertaining to a difference between two signals or quantities.

differential amplifier A circuit, usually an operational amplifier, that amplifies the voltage difference between two input signals. The instantaneous output voltage is equal to some constant multiple of the difference between the instantaneous input voltages.

differential analyzer An analog computer that solves differential equations using integrators.

differential angle For a mercury switch, the angle between operation and release positions.

differential capacitor A dual variable capacitor with two identical stator sections, and a single rotor section that turns into one stator section and out of the other. The capacitance of one section decreases while that of the other increases.

differential coil See DIFFERENTIAL, **3**.

differential comparator A linear integrated circuit (IC) that delivers an output proportional to the difference between two input signals.

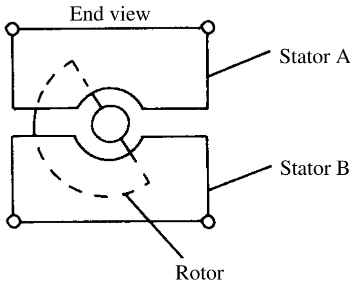

differential capacitor

differential compound dc generator A compound-wound dc generator in which the magnetomotive force of the series field opposes that of the shunt (main) field.

differential compound dc motor A compound-wound dc motor in which the magnetomotive force of the series field coil opposes that of the shunt (main) field coil.

differential cooling Reducing temperature at different points on a surface at different rates.

differential delay The difference $d_{\max} - d_{\min}$ across a frequency band, where $d_{\max}$ is the maximum frequency delay and $d_{\min}$ is the minimum frequency delay.

differential discriminator A device that passes pulses, whose amplitudes are between two predetermined values above or below zero.

differential distortion In an automatic-gain-control circuit, distortion from effects that cause shunting of the diode load resistor.

differential flutter Fluctuations in the speed of a magnetic tape that are nonuniform in different parts of the tape.

differential gain In a differential amplifier, the average gain of the two sections of the amplifier. Compare DIFFERENTIAL UNBALANCE.

differential gain control A circuit or device for setting the gain of a radio receiver in terms of an anticipated change in signal strength, to reduce the receiver output signal differential.

differential galvanometer A galvanometer in which currents in two similar coils neutralize each other; thus, there is zero deflection when the currents are equal.

differential gap The smallest range of values that a controlled variable must take to change a three-position controller's output from on to off, or vice versa.

differential heating Increase of temperature at different points on a surface at different rates.

differential impedance See DIFFERENTIAL-INPUT IMPEDANCE.

differential induction coil An induction coil having two differentially wound primary coils.

differential input In a differential amplifier, the circuit between input terminals 1 and 2, as opposed to the circuit between input 1 or input 2 and ground.

differential-input amplifier A differential amplifier whose output is proportional to the difference between two input signals—each applied between an input terminal and common ground.

differential-input capacitance In a differential amplifier, the capacitance between the input terminals.

differential-input impedance In a differential amplifier, the impedance between the input terminals.

differential-input measurement For a differential amplifier, a floating measurement made between the input terminals.

differential-input rating In an operational amplifier, the greatest difference signal that can be placed between the inputs while allowing proper operation.

differential-input resistance In a differential amplifier, the resistance between the input terminals.

differential-input voltage In a differential amplifier, the signal voltage presented to the floating input terminals.

differential-input voltage range In a differential amplifier, the range of signal voltages that can be applied between the differential input terminals without overdriving the amplifier.

differential input-voltage rating The maximum differential-input voltage that can be applied safely to a differential amplifier.

differential instrument A galvanometer or other meter in which deflection results from the differential effect of currents flowing in opposite directions through two identical coils. Also see DIFFERENTIAL GALVANOMETER.

differential keying A system of break-in keying, in which the oscillator stage of a transmitter containing a keyed amplifier is disabled when the key is open to prevent interference with the receiver at the keying station, and is enabled when the key is closed.

differential-mode gain In an operational amplifier, the ratio, in decibels, between the output voltage and the differential input voltage.

differential-mode input In an operational amplifier in differential mode, the difference between the two input signal voltages.

differential-mode signal In a balanced three-terminal circuit, such as the input of a differential amplifier, a signal applied between the floating (ungrounded) input terminals.

differential multiplexer An analog multiplexer that selects both the high and low portion of the input signal.

differential nonlinearity Incremental error from an ideal analog output difference when the input is changed by a certain value. Generally expressed as a fraction of full-scale output.

differential permeability The derivative of normal induction, with respect to magnetizing force.

differential phase In a television system tested with a low-level, high-frequency sine-wave signal (f_1) superimposed on a low-frequency, sine-wave signal (f_2), the difference in phase shift of f_1 throughout the system for two specified levels of f_2.

differential phase-shift keying Keying of a carrier by varying the carrier phase.

differential pressure The difference in pressure between two points.

differential-pressure transducer A transducer that delivers an output proportional to the difference between two sensed actuating pressures.

differential protective relay A differential relay that operates to protect equipment or personnel when the difference between the two actuating quantities reaches a prescribed level.

differential receiver A synchro differential that receives the electrical output of two synchro transmitters. The receiver can subtract one input voltage from the other.

differential relay A relay actuated by the difference between two currents or voltages.

differential selsyn A selsyn in which the position assumed by the rotor is proportional to the sum of rotor and stator field values.

differential stage See DIFFERENTIAL AMPLIFIER.

differential synchro See DIFFERENTIAL RECEIVER and DIFFERENTIAL TRANSMITTER.

differential transducer A dual-input, single-output sensor, such as a pressure transducer, that is actuated by two sensed quantities and delivers an output proportional to their difference.

differential transformer A variable inductance transformer having a (usually cylindrical) core that is moved in and out to provide adjustable coupling between the interwound primary and secondary windings. This permits adjustment of the amplitude and phase of the transformer output voltage, with respect to the input voltage.

differential transmitter A synchro differential connected to a synchro transmitter. In a synchro receiver supplied by this combination, the change in rotor position is the algebraic difference between the transmitter-rotor position and the differential-rotor position.

differential unbalance For a differential amplifier, the average difference in gain between the two amplifier sections. Compare DIFFERENTIAL GAIN.

differential voltage **1.** The voltage difference between the input signals to a differential device. **2.** The breakdown voltage minus the operating voltage for a lamp.

differential voltage gain **1.** The ratio, in decibels, between the differential output and differential input voltages of an amplifier. **2.** The instantaneous ratio, in decibels, between the rate of change of the output signal voltage and the rate of change of the input signal voltage in an amplifier.

differential-wound field In a motor or generator, a field winding having series and shunt coils whose fields are opposing.

differentiate **1.** To produce an output signal, the instantaneous amplitude of which is proportional to the instantaneous rate of change of the input amplitude. **2.** To determine the derivative of a mathematical function.

differentiating circuit See DIFFERENTIATING NETWORK.

differentiating network A four-terminal resistance-capacitance (RC) network whose output voltage is the derivative of the input voltage, with respect to time. Compare INTEGRATING NETWORK.

differentiation **1.** The processing of an input signal to create an output signal whose voltage waveform represents the derivative, with respect to time, of the input voltage waveform. **2.** The process of computing a mathematical derivative.

differentiator **1.** See DIFFERENTIATING NETWORK. **2.** An operational amplifier whose output waveform is the mathematical derivative of the input waveform.

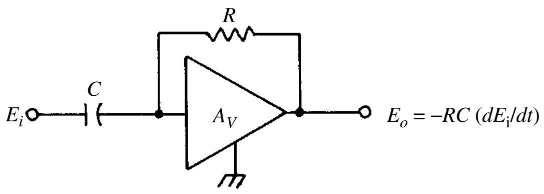

differentiator, 2

diffracted wave A wave or ray of energy undergoing DIFFRACTION.

diffraction **1.** Interference of one part of an energy beam with another part when the beam is deflected along two or more paths having different lengths. When this happens with visible light, dark and light bands or colored bands appear. This effect is responsible for the rainbow-like appearance of light reflected from the surface of a compact disc. **2.** The bending of electromagnetic waves around an object. This effect explains why radio signals can propagate around large obstructions, such as buildings and hills. The effect becomes more pronounced as the wavelength increases (the frequency decreases). **3.** The bending of acoustic waves around an object. This effect explains why sound propagates around large obstructions, such as buildings. The effect becomes more pronounced as the wavelength increases (the frequency decreases).

diffraction grating A transparent plate containing thousands of parallel lines or grooves spaced extremely close together. Light passing through the slits between the lines produces a rainbow spectrum as a result of DIFFRACTION.

diffraction spectrum **1.** The spectrum produced in visible light by a diffraction grating. **2.** The distribution of energy at various frequencies, produced by diffraction of electromagnetic waves. **3.** The distribution of energy at various frequencies, produced by diffraction of acoustic waves.

diffractometer An instrument for measuring the diffraction of radiation, such as light or X-rays.

diffuse **1.** To produce or cause DIFFUSION. **2.** Energy that is diffused.

diffused-alloy transistor See DRIFT-FIELD TRANSISTOR.

diffused-base transistor A bipolar transistor in which the base region has been diffused into the semiconductor wafer. Also see DIFFUSED JUNCTION.

diffused device A semiconductor device in which the junction is produced by diffusion (see DIFFUSION, **1**). Examples: DIFFUSED-BASE TRANSISTOR, DIFFUSED DIODE, DIFFUSED-JUNCTION RECTIFIER, and DIFFUSED-MESA TRANSISTOR.

diffused diode A semiconductor diode having a diffused junction.

diffused-emitter-and-base transistor A transistor in which n and p materials both have been diffused into the semiconductor wafer to provide emitter and base junctions. Also see DIFFUSION, **1** and DIFFUSED TRANSISTOR.

diffused junction In a semiconductor device, a pn junction formed by diffusing a gas into a semiconductor at a high temperature that is below the melting point of the semiconductor. Typically, a gas containing an n-type impurity is diffused into p-type semiconductor material. Compare ALLOY JUNCTION.

diffused-junction rectifier A semiconductor rectifier using a diffused junction.

diffused-junction transistor See DIFFUSED-BASE TRANSISTOR, DIFFUSED-MESA TRANSISTOR, and DIFFUSED TRANSISTOR.

diffused-layer resistor In an integrated circuit, a resistor produced by diffusing a suitable material into the substrate.

diffused-mesa transistor A transistor whose base is a n-type layer diffused into a p-type wafer (the remaining p-type material serving as the collector); its emitter is a small p-type area diffused into or alloyed with the n-layer. Unwanted diffused portions are etched away, leaving the transistor in a mesa shape.

diffused planar transistor A diffused transistor in which emitter, base, and collector electrodes are exposed at the face of the wafer, which has an oxide layer to forestall leakage between surface electrodes.

diffused resistor See DIFFUSED-LAYER RESISTOR.

diffused sound **1.** Sound distributed so that its energy flux is the same at all points. **2.** Sound whose source is difficult to locate or seems to shift, as that heard from out-of-phase stereo channels.

diffused transistor A transistor in which one or both electrodes are created by diffusion. See DIFFUSED JUNCTION.

diffused-junction transistor See DIFFUSED-BASE TRANSISTOR, DIFFUSED-MESA TRANSISTOR, and DIFFUSED TRANSISTOR.

diffusion **1.** In the fabrication of semiconductor devices, the slow, controlled introduction of a material into the semiconductor, for example, the high-temperature diffusion of a n-type impurity (from a gas containing it) into a p-type wafer to form a diode. **2.** The random velocity and movement of current carriers in a semiconductor, resulting from a high-density gradient. **3.** The characteristic spreading of light reflected from a rough surface or transmitted through a translucent material. **4.** The spreading-out of sound waves, for example when reflected from acoustic baffles. **5.** The migration of atoms from one substance to another, as in the spreading of one gas throughout another.

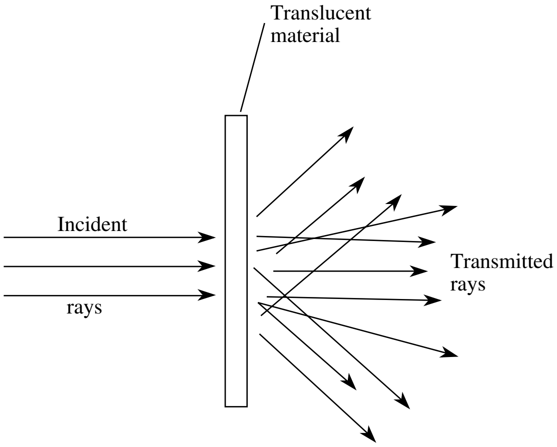

diffusion, 3

diffusion bonding A method of joining different substances by diffusing atoms of one into the other. This technique is employed in the manufacture of certain semiconductor diodes, transistors, and other devices.

diffusion capacitance The current-dependent capacitance of a forward-biased semiconductor junction.

diffusion current Current resulting from the diffusion of carriers within a substance (see DIFFUSION, **2**).

diffusion length In a semiconductor junction, the distance a current carrier travels to the junction during carrier life.

diffusion process **1.** The technique of processing semiconductor devices by diffusion (see DIFFUSION, **1**). **2**. Producing a high vacuum by means of diffusion (see DIFFUSION PUMP).

diffusion pump A pump for fast, efficient creation of a high vacuum in electron tubes and similar devices. In one form, the pump, in conjunction with a force pump, uses mercury vapor as the pumped medium. Gas molecules evacuated from the device diffuse into a chamber, where condensing mercury vapor traps and carries them off.

diffusion theory The notion that, in a homogeneous medium, current density is directly proportional to the gradient of particle flux density.

diffusion transistor A transistor whose operation is based principally on the diffusion of current carriers (see DIFFUSION, **2**).

diffusor In acoustics, a device or structure deliberately installed to spread sound waves throughout a region.

dig-in angle A stylus angle of 90 degrees, relative to the surface of a phonograph disc. Compare DRAG ANGLE.

DIGIRALT Acronym for *digital radar altimetry.* A system that utilizes digital techniques to enhance the accuracy of an altimeter using radar.

digit A single symbol in a numbering system (e.g., 0 through 9 in the decimal system, or 0 or 1 in the binary system), whose value depends on its position in a group and on the radix of the particular system used.

digital **1.** Pertaining to components, circuits, or systems that use signals having an integral number of discrete levels or values, rather than signals, whose levels or values vary over a continuous range. **2.** Pertaining to a numeric readout or display. **3.** See BINARY, **1**.

digital annunciator An annunciator that gives an alphanumeric digital display of information, as well as sounding an alarm.

digital audio tape Abbreviation, DAT. A magnetic tape intended for recording digitally encoded audio data. Used in some high-fidelity applications, and also for computer data storage.

digital barometer An electronic barometer providing a digital readout.

digital capacitance meter Abbreviation DCM. A meter with a digital readout for measuring capacitance values.

digital cellular See PERSONAL COMMUNICATIONS SERVICE.

digital circuit A circuit affording a dual-state switching operation (i.e., on or off, high or low, etc.). Also called *binary circuit.*

digital communications Radio or wire communications using a dual-state mechanism (on/off, positive/negative, or modulated/unmodulated) to represent information.

digital comparator A comparator that presents two digital values, one for each of the quantities being compared.

digital computer A high-speed, electronic machine for performing mathematical operations, file management, machine control, or other "intelligent" functions, and whose basic internal operations (data storage, comparing, and computation) are based on semiconductor devices assuming one of two states (on or off, high or low). Compare ANALOG COMPUTER.

digital data Information represented and processed in the form of combinations of digits (0 and 1, in the binary system).

digital-data cable A cable designed to conduct high-speed digital pulses with minimal distortion and loss.

digital data-handling system A system that accepts, sorts, modifies, classifies, or records digital data, displaying the final result or passing the data to a computer.

digital delay circuit A device that stores digitized audio data, and releases it after a specified delay.

digital device **1.** A digital integrated circuit (IC). **2.** Any circuit or system that operates by digital means.

digital differential analyzer Abbreviation, DDA. A digital computer that can perform integration using specialized circuitry.

digital display A presentation of information (such as the answer to a problem) in the form of actual digits, as opposed to one in the form of, for example, a meter deflection. See, for example, DIGITAL-TYPE METER.

digital divider In a computer, a device that can divide (i.e., provide a quotient and remainder using dividend and divisor signals).

digital electrometer An electrometer that has a digital current or voltage indicator.

digital electronics The branch of electronics concerned with components, circuits, and systems that use signals having an integral number of discrete levels or values, as opposed to signals whose levels or values vary over a continuous range. Compare ANALOG ELECTRONICS.

digital frequency meter A direct-reading frequency meter using high-speed electronic switching circuits and a digital readout. Such instruments read frequency from less than 1 Hz to many gigahertz.

digital HIC A hybrid integrated circuit (HIC) designed for digital applications. Also see DIGITAL INTEGRATED CIRCUIT.

digital IC See DIGITAL INTEGRATED CIRCUIT.

digital incremental plotter A device that can draw, according to signals received from a computer, graphs depicting solutions to problems.

digital information See DIGITAL DATA.

digital information display See DIGITAL DISPLAY.

digital integrated circuit An integrated circuit (IC) intended for binary operations, such as switching, gating, etc. Compare LINEAR INTEGRATED CIRCUIT.

digital integrator A device that can perform integration, in which increments in input variables, and an output variable, are represented by digital signals.

digital logic A form of Boolean algebra, consisting of negation, conjunction, and disjunction, in which the binary digit 1 has the value "true" and 0 the value "false" (in positive logic) or vice versa (in negative logic). Digital logic is the basis by which all digital devices function.

digital-logic module **1.** A circuit that performs digital operations. **2.** A logic gate.

digital meter A meter that produces a readout in discrete blocks or directly as numerals. The first, more primitive and less precise type, is known as a BAR METER. The second, more sophisticated type can resolve to several significant digits and often includes a fixed or floating radix point. This scheme eliminates the need for personnel to interpolate the reading on a scale. There is little chance for error on the part of the technician or engineer, because the readout is straightforward. Another advantage is the fact that there are no moving parts to wear out or be damaged by physical shock. Compare ANALOG METER.

digital multimeter Abbreviation, DMM. A voltohm-milliammeter producing a digital readout of measured values.

digital multiplex **1.** The combination of several or many digital signals into a single digital signal. **2.** Also called *digital demultiplex.* The reverse process from that defined in 1, in which the original signals are obtained from the combination signal. **3.** Communication using the techniques defined in 1 and 2.

digital multiplex equipment Equipment that accomplishes digital multiplexing or the reverse process, digital demultiplexing.

digital multiplier In a digital computer, a device that produces a product signal from multiplier and multiplicand signals.

digital output An output signal of digital pulses representing a number equal or proportional to the value of a corresponding input signal.

digital panel meter A numeric-readout meter whose relatively small size allows mounting on a panel.

digital phase shifter A phase shifter actuated by a digital control signal.

digital photometer An electronic photometer providing a digital readout of illumination values.

digital power meter An electronic wattmeter providing a digital readout of measured power.

digital readout An indicating device that displays a sequence of numerals that represent a measured value.

digital recording A system for tape-recording high-fidelity sound. The audio is converted from analog to binary digital form, and the binary digits (bits) are recorded on magnetic tape.

digital representation The use of digital signals to represent information as characters or numbers.

digital rotary transducer A device that delivers a digital output signal proportional to the rotation of a shaft.

Digital Satellite System Abbreviation, DSS. Trade name for a satellite television (TV) system developed by RCA. The analog signal is changed into digital pulses at the transmitting station via analog-to-digital (A/D) conversion. The digital signal is amplified and uplinked to a geostationary satellite. The satellite has a transponder that receives the signal, converts it to a different frequency, and downlinks it back to the earth. The downlink is picked up by a portable dish that can be placed on a balcony or patio, on a rooftop, or in a window. A tuner selects the channel that the subscriber wants to watch. The digital signal is amplified. If necessary, digital signal processing (DSP) can be used to improve the quality of reception under marginal conditions. The digital signal is changed back into analog form, suitable for viewing on a conventional TV set, via digital-to-analog (D/A) conversion.

digital signal A signal having an integral number of discrete levels or values, as opposed to a signal whose levels or values vary over a continuous range.

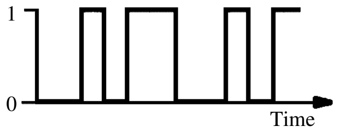

digital signal

digital signal processing Abbreviation, DSP. A method of signal enhancement that operates by eliminating confusion between digital states. This improves dynamic range and frequency response, reduces the number of errors, and virtually eliminates noise. It is used extensively in digital communication and recording, often in conjunction with analog-to-digital (A/D) and digital-to-analog (D/A) conversion to enhance the quality of analog signals and recordings.

digital sound Sound recording and reproduction accomplished with digital, rather than analog, signals. Advantages include wideband frequency response, superior dynamic range, and relative immunity to noise.

digital speech communications A system of voice communications, in which the analog voice signal is encoded into digital pulses at the transmitter, and decoded at the receiver.

digital subtractor In a computer, a device that produces an output signal whose value is equal to the difference of the values of two input signals.

digital switching Routing operations carried out on digital signals to establish communications links between specified system users.

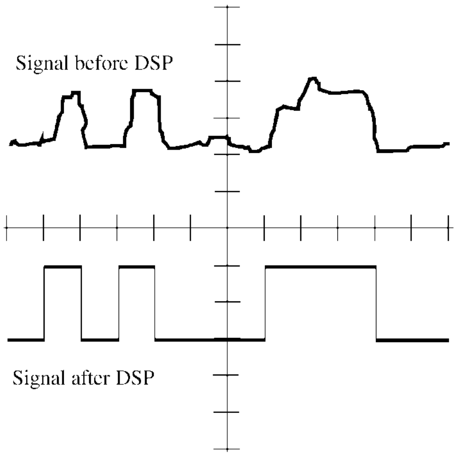

digital signal processing

digital television **1.** A television system in which the picture information is encoded into digital form at the transmitter, and decoded at the receiver. **2.** A form of television picture transmission that functions according to picture motion, rather than absolute brightness.

digital temperature indicator See DIGITAL THERMOMETER.

digital thermometer An electronic thermometer that provides a digital readout of temperature.

digital-to-analog conversion The conversion of a digital quantity into an analog representation, such as shown by a performance curve. Compare ANALOG-TO-DIGITAL CONVERSION.

digital-to-analog converter A circuit or device that performs DIGITAL-TO-ANALOG CONVERSION.

digital transmission **1.** A method of signal transmission in which the modulation occurs in defined increments, rather than over a continuous range. **2.** A message that is sent in digital form.

digital-type meter An indicating instrument in which a row of numeral indicators displays a value. Compare ANALOG-TYPE METER.

digital voltmeter Abbreviation, DVM. An electronic voltmeter having a direct numerical readout, rather than an analog display.

digital wattmeter See DIGITAL POWER METER.

digital compression In digital computer operation, the process of representing data with an economy of characters to reduce file size.

digit current In digital computer operations, the current associated with writing or reading a digit into or out of a memory cell.

digit delay element A logic element (gate) whose output signal lags the input signal by one digit period.

digit filter A device for detecting designations. See DESIGNATION.

digitize **1.** To express the results of an analog measurement in digital units. **2.** To convert an analog signal into corresponding digital pulses.

digitizer See ANALOG-TO-DIGITAL CONVERTER.

digit period In a digital circuit or system, the time interval between the start of one digital pulse and the start of the next pulse.

digit place See DIGIT POSITION.

digit plane In a matrix-type computer memory, the plane within a three-dimensional array of memory storage elements representing a DIGIT POSITION.

digit position The ordinal position of a digit in a numeral, the first position being occupied by the least-significant digit (e.g., 7 is in the third position in the numeral 756).

digit pulse A pulse that energizes magnetic core memory elements representing a digit position in several words.

digitron A display in which all of the characters lie in a single, flat plane.

digit time The duration of a digit signal in a series of signals.

digit time slot In digital communications, the interval of time assigned to one bit or one digit.

digit-transfer bus In a digital computer, a main line (of conductors) that transfers information among various registers; it does not handle control signals.

diheptal CRT base The 14-pin base of a cathode-ray tube. Also see BIDECAL, DUODECAL, and MAGNAL.

DIIC Abbreviation for *dielectric-isolated integrated circuit.* Several separate integrated-circuit wafers are contained in a single package, and kept electrically insulated by layers of dielectric.

dilatometer An instrument used to measure expansion.

dimension **1.** Any measurable quantity, such as distance, time, temperature, humidity, etc. **2.** An axis in the three-dimensional Cartesian coordinate system. **3.** An independent variable in a function of one or more variables.

dimensional analysis A mathematical procedure whereby an equation involving quantities with different units is verified as being *dimensionally correct.* The original variables are replaced with fundamental quantities, such as resistance (R), current (I), length or displacement (d), and time (t), applicable to electrical systems. The equation is dimensionally correct if it can be shown that the left and right sides of the equation are identical.

dimensional ratio In magnetism, the ratio of the longest diameter of an elongated ellipsoid of revolution to the shortest.

dimensional stability Nonvariance or little variance in the shape and size of a medium (such as film) during the processing of that material.

dimensionless quantity A quantity that is merely a real number. Example: logarithm, exponent, numerical ratio, etc. In contrast are physical quantities: 3 volts, 5000 hertz, 10 amperes, etc.

diminished radix complement See COMPLEMENT.

dimmer An electronic device used for controlling the brightness of incandescent lamps. Using amplified control, the device enables high-wattage lamp loads to be smoothly adjusted via a small rheostat or potentiometer. A photoelectric-type dimmer automatically controls lamps in accordance with the amount of daylight.

dimmer curve The function of a light-dimmer voltage output as a function of setting on a linear scale.

DIN Abbreviation for *Deutsche Industrie Normenausschuss.* A German association that sets standards for the manufacture and performance of electrical and electronic equipment, as well as other devices.

D indicator In radar operations, an indicator combining type B and C indicators (see B DISPLAY and C DISPLAY).

Dingley induction-type landing system An aircraft landing system that provides lateral and vertical guidance; instead of radio, it uses the magnetic field surrounding two horizontal cables laid on or under either side of the runway.

diode A two-element device containing an anode and a cathode, and providing unidirectional conduction. The many types are used in such devices as rectifiers, detectors, peak clippers, mixers, modulators, amplifiers, oscillators, and test instruments.

diode action **1.** The characteristic behavior of a diode (i.e., rectification and unidirectional conduction). **2.** Two-electrode rectification or unidirectional conductivity in any device other than a diode (e.g., asymmetrical conductivity between the collector and base of a transistor).

diode amplifier **1.** A parametric amplifier employing a varactor. **2.** An amplifier utilizing hole-storage effects in a semiconductor diode. **3.** A negative-resistance amplifier using a tunnel diode.

diode array A combination of several diodes in a single housing.

diode assembly See DIODE ARRAY.

diode bias A steady direct-current (dc) voltage applied to a diode to establish its operating point.

diode capacitance The capacitance existing at the p-n junction of a semiconductor diode when the junction is reverse-biased. The capacitance generally varies, depending on the reverse-bias voltage.

diode capacitor **1.** A capacitor normally operated with a diode. **2.** A voltage-variable capacitor utilizing the junction capacitance of a semiconductor diode (e.g., a varactor).

diode-capacitor memory cell A high-value capacitor in series with a high-back-resistance semiconductor diode. A data pulse forward-biases the diode and charges the capacitor, which remains charged, thus holding the data bit, because of the long time constant of the high capacitance and the high back resistance of the diode.

diode characteristic The current-versus-voltage curve for a diode.

diode checker An instrument for testing semiconductor diodes. There are two forms: A static checker, which measures forward and reverse current; and a dynamic checker (see DYNAMIC DIODE TESTER), which displays the entire diode response curve on an oscilloscope screen.

diode chopper A chopper using an alternately biased diode as the switching element.

diode clipper A clipper using one or more diodes. A single biased diode will limit the positive or negative peak of an applied alternating-current (ac) voltage, depending on diode polarity and bias. Two biased diodes with opposing polarity will clip both peaks. Also see LIMITER.

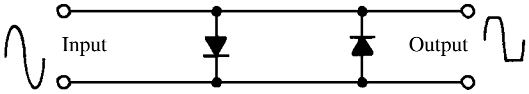

diode clipper

diode converter See DIODE MIXER.

diode current The forward or reverse current flowing through a diode.

diode current meter A direct-current (dc) milliammeter or microammeter with a semiconductor-diode rectifier that allows the measurement of alternating current (ac).

diode curve changer A diode or network of diodes used to make a linear current-voltage curve acquire some nonlinear shape.

diode demodulator See DEMODULATOR PROBE and DIODE DETECTOR.

diode detector A detector circuit in which a diode demodulates a signal. The diode, a simple device, provides linear response at high signal amplitudes, but affords no amplification.

diode feedback rectifier **1.** In a rectified-carrier, negative-feedback system for an amplitude-modulated (AM) transmitter, the diode that rectifies the modulated carrier and provides the audio envelope for use as negative-feedback voltage. This voltage is applied to the speech amplifier/modulator channel to reduce distortion, noise, and hum, at the same time providing automatic modulation control. **2.** The diode that rectifies a part of the signal at the output of an audio amplifier and provides a proportional direct-current (dc) voltage for use as bias in an automatic-gain-control (AGC) circuit.

diode field-strength meter A simple meter for measuring the intensity of a radio-frequency (RF) electromagnetic field. It consists of a short whip antenna, an inductance-capacitance (LC) tuned circuit, a diode detector, and a direct-current (dc) microammeter. The deflection of the meter is roughly proportional to the RF signal voltage.

diode gate A passive switching circuit of biased diodes. Also see AND CIRCUIT and OR CIRCUIT.

diode impedance The vector sum (resultant) of the resistive and reactive components of a diode. In a semiconductor diode, the inductive component of reactance is almost entirely the inductance of leads and electrodes, whereas the capacitive component of reactance is the shunting capacitance between leads and electrodes, plus the voltage-variable capacitance of the pn junction. The resistive component is almost entirely the voltage-variable resistance of the pn junction.

diode isolation A means of insulating an integrated-circuit chip from its substrate. The chip is surrounded by a pn junction that is reverse-biased. This prevents conduction between the chip and the substrate.

diode lamp See LASER DIODE.

diode laser See LASER DIODE.

diode light source See LASER DIODE.

diode limiter See DIODE CLIPPER.

diode load **1.** The current drawn from a diode acting as a rectifier or demodulator. **2.** The output (load) resistor into which a diode operates.

diode load resistance The required value for a diode load resistor.

diode load resistor A resistor usually connected to the output of a diode rectifier or diode detector.

diode logic Digital circuitry, such as AND and OR circuits, using diodes as the principal components.

diode matrix In some digital devices, a grid of wires, the intersections of some being interconnected through diodes, whose polarities determine circuit operation. A series of AND circuits is provided by this arrangement, which acts as a high-speed rotary switch when it is supplied with input pulses.

diode mixer A frequency converter that operates via the nonlinearity of semiconductor diodes.

diode noise limiter A noise limiter circuit having one or more biased diodes.

diode oscillator An oscillator based on the negative resistance or breakdown characteristics of certain diodes, such as high-reverse-biased germanium diodes, tunnel diodes, Gunn diodes, and four-layer diodes. It is generally used at microwave frequencies.

diode pack A device containing more than one diode. An example is the full-wave bridge-rectifier integrated circuit.

diode peak detector A diode detector whose load resistance is high at modulation frequencies; the voltage across the resistance is proportional to the peak amplitude of the modulated signal.

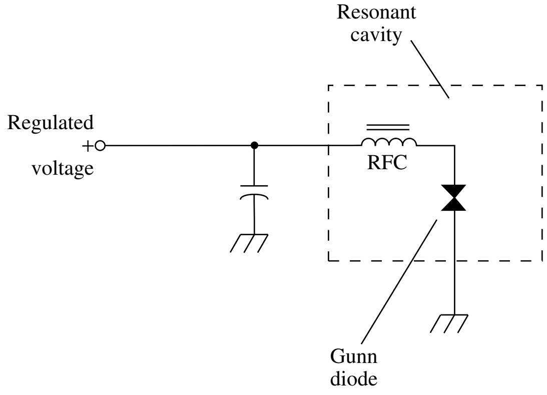

diode oscillator, 1

diode peak voltmeter A diode-type alternating-current (ac) voltmeter, in which the deflection of the direct-current (dc) milliammeter or microammeter is proportional to the peak value of the applied ac voltage.

diode probe A test probe containing a diode used as either a rectifier or demodulator.

diode recovery time The interval during which relatively high current continues to flow after the voltage across a semiconductor junction has been abruptly switched from forward to reverse. Recovery time is attributable to DIODE STORAGE.

diode rectification Conversion of alternating current (ac) to pulsating direct current (dc) by diode action.

diode rectifier **1.** A diode device that converts alternating current (ac) to pulsating direct current (dc) in a power supply. **2.** A small-signal diode device that converts ac to dc in the automatic-gain-control (AGC) circuit of a superheterodyne receiver. Also called *AGC rectifier.*

diode resistor **1.** A resistor usually operated with a diode. **2.** A voltage-variable resistor utilizing the (usually forward) resistance of a semiconductor diode.

diode storage The charge carriers (electrons and holes) remaining within a pn junction for a short time after forward bias has been either removed or switched to reverse polarity.

diode storage time See DIODE RECOVERY TIME.

diode switch See DIODE GATE.

diode sync separator A diode used in a television receiver circuit to separate and deliver the sync pulses from the composite video signal.

diode temperature stabilization **1.** Keeping the temperature of a diode at a constant level. **2.** Using the temperature-resistance characteristic of a forward-biased semiconductor diode to stabilize a circuit (such as a transistor amplifier stage) (i.e., to prevent variations caused by temperature changes).

diode tester See DIODE CHECKER.

diode transistor **1.** See UNIJUNCTION TRANSISTOR. **2.** A semiconductor diode whose operation simulates that of a transistor by means of pulsed operation that alternately makes the single junction an emitter or collector. **3.** A transistor connected to operate solely as a diode.

diode-transistor logic Abbreviation, DTL. Logic circuitry in which a diode is the logic element and a transistor acts as an inverting amplifier.

diode-type meter A rectifier-type alternating-current (ac) meter consisting of a semiconductor diode(s) and a direct-current (dc) milliammeter or microammeter. The diode rectifies the ac input, the resulting dc deflecting the meter.

diode varactor A conventional semiconductor diode or rectifier used as a makeshift varactor (voltage-variable capacitor).

diode variable resistor See DIODE VARISTOR.

diode varistor A conventional diode used as a makeshift varistor (voltage-variable resistor).

diode voltage reference See ZENER VOLTAGE REFERENCE.

diode voltage regulator See ZENER VOLTAGE REGULATOR.

DIP Abbreviation of DUAL IN-LINE PACKAGE.

dip **1.** A distinct decrease in the value of a varying quantity, followed by an increase [e.g., the sudden drop, followed by a rise, in collector current when a bipolar-transistor radio-frequency (RF) power amplifier is tuned through resonance]. **2.** Also called *magnetic inclination.* The slanting of a compass needle, resulting from the orientation of the geomagnetic lines of flux, with respect to the earth's surface. It varies, depending on magnetic latitude.

dip adapter An external accessory that allows a radio-frequency (RF) signal generator to be used as a DIP METER.

dip coating **1.** Applying a protective coat of insulating material to a conductor or component by dipping it into the liquid material, then draining and drying it. Compare SPRAY COATING. **2.** The coat applied in this way.

dip encapsulation Embedding a component or circuit in a protective block of insulating material (such as a plastic) while the material is in a liquid state, and then allowing the material to harden in ambient air or in an oven.

dip impregnation Saturating a component or material (such as absorbent film) with a substance (such as oil or wax) by dipping or vacuum forcing.

diplexer A coupler that permits two or more transmitters to operate simultaneously into a single antenna.

diplex operation **1.** Simultaneous transmission or reception of two signals using a single antenna. **2.** Simultaneous transmission or reception of two signals on a single carrier.

diplex reception The reception of signals while transmitting with the same antenna.

diplex transmission The transmission of signals while receiving with the same antenna.

dip meter A tunable radio-frequency (RF) instrument that, by means of a sharp dip of an indicating meter, indicates resonance with an external circuit under test. Specific names are derived from the active component used: *grid-dip meter*, *gate-dip meter*, etc.

dip needle See INCLINOMETER.

dipolar Also, *bipolar*. Possessing two poles (usually electric or magnetic).

dipolarization See DEPOLARIZATION.

dipole **1.** A pair of electrically opposite charge poles separated by a specific distance. **2.** A pair of magnetically opposite poles separated by a specific distance. **3.** See DIPOLE ANTENNA. **4.** See FOLDED DIPOLE.

dipole antenna Also called *dipole* and *doublet*. A half-wavelength radiator fed at the center with a two-wire or coaxial transmission line. Each "leg" of the antenna is one-quarter wavelength long. Such an antenna can be oriented horizontally or vertically, or at a slant. The radiating element is usually straight. For a straight wire radiator, properly insulated at the ends and placed well away from obstructions, the length L_{ft} (in feet) at a design frequency f (in megahertz) is approximately

$$L_{ft} = 467/f$$

The length L_m (in meters) is close to

$$L_m = 143/f$$

Because of its simplicity, this antenna is popular among shortwave listeners and radio amateurs, especially at frequencies below 10 MHz. A full-size antenna of this type has a feed-point impedance of approximately 73 ohms, purely resistive. Compare FOLDED DIPOLE.

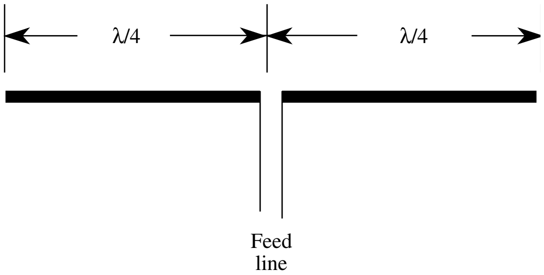

dipole antenna

dipole disk feed A method of coupling radio-frequency energy to a disk-shaped antenna. The energy is applied to a dipole located adjacent to the disk.

dipole feed A method of coupling radio-frequency energy to an antenna by means of a half-wave dipole. The dipole is directly fed by the transmission line, and the dipole radiates energy to the rest of the system.

dip oscillator The oscillator that provides the signal for a DIP METER.

dipotassium tartrate Abbreviation, DKT. An organic piezoelectric material.

dipped component A discrete electronic component that has been given a protective coating by dipping into a suitable material (such as oil, varnish, or wax) and draining off the surplus.

dipper Collective term for resonance-type instruments, such as a DIP METER or DIP ADAPTER.

dipper interrupter A cyclic switching device in which a contact pin is part of a revolving wheel partially immersed in mercury.

dipping **1.** The application of a protective coating or impregnant to a component by immersing it in a suitable material. Also see DIP COATING, DIP ENCAPSULATION, and DIP IMPREGNATION. **2.** In a resonant (tuned) amplifier circuit, the adjustment of the resonant circuit for minimum current through the amplifying device.

dipping needle See INCLINOMETER.

dip soldering **1.** Soldering leads or terminals by dipping them into molten solder and then removing excess solder. **2.** Tinning printed-circuit patterns by dipping the boards into molten solder or placing them in contact with the surface of a solder bath. **3.** Soldering leads in printed circuits by the methods defined in (2).

DIP switch A switch (or group of miniature switches) mounted in a dual-inline package (DIP) for easy insertion into an integrated-circuit socket or printed-circuit board.

direct-access storage device A computer memory in which data access time is unaffected by the data location. Also called *random-access memory device.*

direct-acting recorder See GRAPHIC RECORDER.

direct-acting recording instrument See GRAPHIC RECORDER.

direct address The actual address of a computer storage location (i.e., the one designated by machine code 0. Also called *absolute address* or *real address.*

direct capacitance The capacitance between two points in a circuit, as opposed to the capacitance between either point and other objects (including ground).

direct allocation In digital computer operations, to specify the necessary memory locations and peripherals for a particular program when it is written.

direct coding Computer programming in machine language.

direct control Control of one machine by another, for example, the control of a computerized mobile robot by a central computer system.

direct-conversion receiver A heterodyne receiver in which the incoming radio-frequency (RF) signal

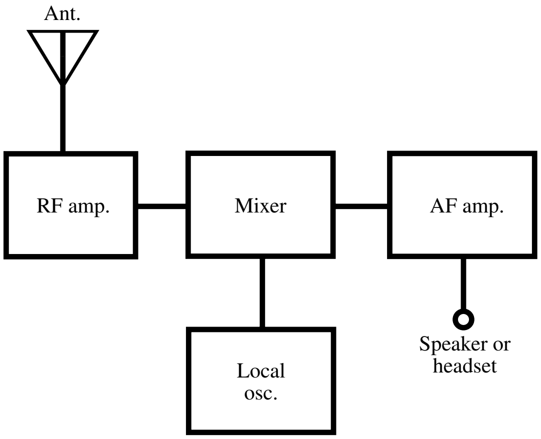

direct-conversion receiver

is amplified, then mixed with the RF output of a tunable local oscillator, producing an audio-frequency (AF) beat note. The AF is amplified; audio filtering can be added. Although the direct-conversion receiver somewhat resembles the superheterodyne type, it has no intermediate-frequency (IF) chain, and does not normally provide single-signal reception. Also see ZERO-BEAT RECEPTION.

direct-coupled amplifier An amplifier in which the output circuit of one stage is wired directly to the input circuit of the following stage (i.e., there is no intervening capacitor or transformer). Such an amplifier can handle alternating-current (ac) or direct-current (dc) signals, and has wide frequency response.

direct-coupled transistor logic Abbreviation, DCTL. In digital computer and switching circuits, a logic system using only direct-coupled transistor stages.

direct coupling Direct connection of one circuit point to another for signal transmission (i.e., without intermediate capacitors or transformers). Because coupling devices aren't used, direct coupling provides transmission of direct current (dc), as well as alternating current (ac).

direct current **1.** Abbreviation, dc. A current that always flows in the same direction (i.e., the polarity never reverses). The current might be constant, as from a battery or a regulated power supply; it might be pulsating, as from an unfiltered rectifier. **2.** Pertaining to current that always flows in the same direction. **3.** Descriptive of a voltage, resistance, or other parameter under conditions in which there is a usually constant current that always flows in the same direction.

direct-current amplifier An amplifier for boosting direct-current (dc) signals, as opposed to dc voltage signals.

direct-current bar See DC BAR.

direct-current beta See DC BETA.

direct-current block See DC BLOCK.

direct-current bus See DC BUS.

direct-current circuit breaker See DC CIRCUIT BREAKER.

direct-current component See DC COMPONENT.

direct-current converter See DC CONVERTER.

direct-current coupling See DC COUPLING.

direct-current dump See DC DUMP.

direct-current equipment See DC EQUIPMENT.

direct-current erase head See DC ERASE HEAD.

direct-current generator See DC GENERATOR.

direct-current inverter See DC INVERTER.

direct-current leakage See DC LEAKAGE.

direct-current motor See DC MOTOR.

direct-current noise See DC NOISE.

direct-current power See DC POWER.

direct-current relay See DC RELAY.

direct-current resistance See DC RESISTANCE.

direct-current shift See DC SHIFT.

direct-current short See DC SHORT.

direct-current signaling See DC SIGNALING.

direct-current source See DC SOURCE.

direct-current transducer See DC TRANSDUCER.

direct-current transformer See DC TRANSFORMER.

direct-current transmission See DC TRANSMISSION.

direct digital control In a digital computer, multiplexing or time sharing among a number of controlled loops.

direct display unit A cathode-ray-tube (CRT) peripheral that displays data recalled from memory.

direct-distance dialing A form of telephone service that allows dialing of long-distance numbers without involving a human operator.

direct drive **1.** Pertaining to electromechanical accessories for electronic equipment. **2.** The transmission of power directly from a source (such as a motor) to a driven device without intermediate gears, belts, or clutches.

direct-drive robot A robot that uses the minimum possible number of gears and other drive systems.

direct-drive torque motor In a positioning or speed-control system, a servoactuator connected directly to the driven load.

direct-drive tuning A tuning or adjusting mechanism in which the shaft of the variable component (such as a potentiometer or variable capacitor) is turned directly by a knob (i.e., without gearing, dial cable, or similar linkage).

directed number A number having direction as well as magnitude; a vector quantity.

direct electromotive force A direct-current (dc) voltage that does not fluctuate or pulsate.

direct emf See DIRECT ELECTROMOTIVE FORCE.

direct ground **1.** A ground connection made by the shortest practicable route. Compare INDIRECT GROUND. **2.** An earth ground.

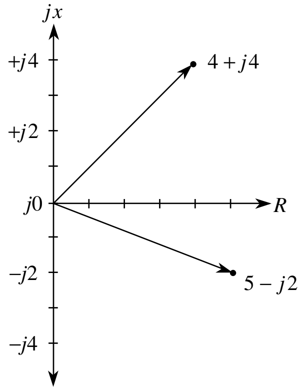

directed number

direct induced current A transient current induced in the same direction as the induction current when it is interrupted.

directing antenna See DIRECTIONAL ANTENNA.

direct-input circuit A circuit, especially an amplifier, whose input is wired directly to the input electrode of the active device (i.e., without a coupling capacitor or transformer).

direct-insert subroutine In digital computer operations, a subroutine directly inserted into a larger instruction sequence. It must be rewritten at every point it is needed.

direct instruction A computer program instruction that indicates the location of an operand in memory.

directional **1.** Depending on direction or orientation. **2.** Having a concentration in an identifiable direction. **3.** Pertaining to a transducer in which radiation, or sensitivity, is concentrated in certain directions at the expense of radiation or sensitivity in other directions.

directional antenna An antenna that transmits and receives signals more effectively in some directions than in others. Also called *beam*, *beam antenna*, and *directive antenna.*

directional array **1.** A directional antenna having a set of elements assembled in such a way that their combined action shapes the radiation into a unidirectional pattern. **2.** A group of antennas spaced and phased to produce unidirectional radiation and reception patterns.

directional beam **1.** An antenna whose radiation or reception pattern strongly favors a specific direction. **2.** The radiation or reception pattern of such an antenna.

directional characteristic The precise directional properties of an antenna or transducer.

directional CQ In amateur radio, a transmission that invites replies only from stations in a certain direction or in a particular city, state, or country.

directional coupler A microwave device that couples an external system to waves traveling through the coupler in one direction.

directional diode A high-back-resistance semiconductor diode inserted into a direct-current (dc) signal circuit or control circuit. Permits unidirectional current flow.

directional filter In carrier-current transmission, a filter that halves the frequency band, one half being for transmission in one direction, and the other half being for transmission in the opposite direction.

directional gain Symbol, k_S. The ratio of the power that would be radiated by a loudspeaker if the free-space axial sound pressure were constant over a sphere, to the actual radiated power. Usually expressed in decibels.

directional homing A scheme for locating the source of a radio signal. An effort is made to keep the bearing of the target or guiding station constant. Therefore, the search path is as direct (as nearly a straight line) as practicable.

directional horn See DIRECTIVE HORN.

directional hydrophone A hydrophone whose response pattern strongly favors one direction.

directional lobe In the spatial response pattern of a device, such as an antenna or loudspeaker, a portion showing emphasized response in a given direction.

directional microphone A microphone that strongly favors sound emanating from in front of it.

directional pattern See DIRECTIVITY PATTERN.

directional phase shifter A phase-shifting circuit in which the characteristics are different in one direction, as compared with the other direction.

directional power relay A relay that is actuated when the monitored power reaches a prescribed level in a given direction.

directional relay See POLARIZED RELAY.

directional response For any form of transducer, a radiation or sensitivity pattern that is concentrated in certain directions.

directional separation filter See DIRECTIONAL FILTER.

directional transducer A device that senses or emits some effect to an extent that depends on the direction from which the effect comes. Directional effects are often, but not always, accompanied by gain in the favored direction(s). Examples: *directional microphone*, *directional speaker*, and *directional antenna.*

directional variation of radio waves Changes in the field strength of radio waves, depending on the direction. There are various causes, including antenna directivity, ground characteristics, ionospheric factors, weather conditions, and the presence of obstructing objects.

directional wattmeter A device that can measure radio transmitter output power and can also give an indication of how well an antenna is matched to a transmission line. Such meters fall into two categories. One type has a single scale, calibrated in watts, and sometimes also in milliwatts or kilowatts (switch selectable). The meter reads either forward power or reflected power, depending on the position of a switch or rotatable internal element. Another type has two needles in a single enclosure, with a different calibrated scale for each needle. Both of these scales are graduated in watts, and sometimes also in milliwatts or kilowatts. One needle/scale indicates forward power and the other needle/scale indicates reflected power. There is a third scale, calibrated for the point where the two needles cross. This scale indicates the standing-wave ratio (SWR). See also CROSSED-POINTER INDICATOR.

direction angle In radar operations, the angle between the center of the antenna baseline and a line going to the target.

direction finder A receiver specially adapted to show the direction from which a signal is received, thus revealing the direction of the receiver with respect to the transmitting station, and vice versa. In its simplest form, it is a receiver with a loop antenna that is rotatable over a map or compass card. For increased accuracy, checks are made with signals from two transmitting stations; the exact location of the receiver is pinpointed by triangulation.

direction finding The taking of bearings by means of a direction finder.

direction of lay In a multiconductor cable, the lateral direction of winding of the topmost conductors as they recede from the observer; called *left-hand lay* or *right-hand lay.* If the cable is viewed from either end, left-hand lay is equivalent to conductors that rotate clockwise as they recede from the viewer; right-hand lay is equivalent to conductors that rotate counterclockwise as they recede from the viewer.

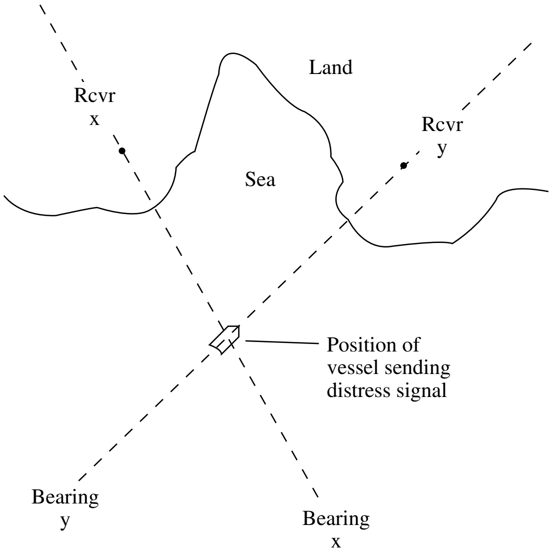

direction finding

direction of polarization The direction of the electrostatic field in a linearly polarized wave.

direction of propagation The direction in which energy moves from a transmitter, or between equivalent points in a sector of space under consideration.

direction rectifier In a control system, a rectifier whose direct-current (dc) output voltage has a magnitude and polarity dependent on the magnitude and phase of an alternating-current (ac) selsyn error voltage.

direction resolution **1.** The smallest difference in azimuth that a direction-finding device can detect. **2.** The smallest angular separation between two targets that allows a radar set to show two separate echoes rather than a single echo.

directive In a computer source program, a statement directing the compiler in translating the program into machine language without being translated itself. Also called *control statement.*

directive antenna An antenna designed for best propagation or reception in one (often steerable) horizontal direction. Also called *beam antenna* and *directional antenna.*

directive gain For a directional antenna, a rating equal to $12.566(Pr/Pt)$, where Pr is the radiated power per steradian in a given direction and Pt is the total radiated power.

directive horn A microwave antenna having the shape of a (usually rectangular) horn.

directivity **1.** In an antenna, a directional response. **2.** The degree to which the radiation or sensitivity of a transducer is concentrated in certain directions. **3.** The angle between the half-power points of a directive antenna in the azimuth plane. **4.** In an antenna system, the ratio, in decibels, between the power in the favored direction and the power in the exact opposite direction; also called front-to-back ratio. **5.** The forward power gain of an antenna, with respect to a dipole in free space. **6.** The forward power gain of an antenna, with respect to an isotropic radiator in free space.

directivity diagram A graph of the radiation/response pattern of a beam antenna or other directional device, usually in a horizontal or vertical plane. Also see DIRECTIVITY PATTERN.

directivity factor **1.** A measure of the directivity of an antenna or transducer. **2.** In acoustics, the ratio, in decibels, between the gain in the maximum direction and the gain in the minimum direction, for a transducer, such as a speaker or microphone.

directivity index **1.** For an acoustic-emitting transducer, the ratio, in decibels, of E_1 to E_2, where E_1 is the average intensity over an entire sphere surrounding the transducer, and E_2 is the intensity on the acoustic axis. **2.** For an acoustic pickup transducer, the ratio, in decibels, of E_1 to E_2, where E_1 is the average response over an entire sphere surrounding the transducer, and E_2 is the response on the acoustic axis.

directivity of antenna For a beam antenna, the ratio E_{max}/E_{avg}, where E_{max} is the maximum field intensity at a selected distance from the antenna and E_{avg} is the average field intensity at the same distance.

directivity of directional coupler The ratio, in decibels, of P_1 to P_2, where P_1 is the power at the forward wave-sampling terminals (measured with a forward wave in the transmission line) and P_2 is the power at the terminals when the wave is reversed in direction.

directivity pattern The calculated or measured radiation or response pattern (transmission or reception) of an antenna, microphone, loudspeaker, or similar device, with particular attention to the directional features of the pattern.

directivity signal A spurious output signal resulting from finite directivity in a coupler.

direct light Light rays traveling directly from a source to a receptor or target without reflection.

directly grounded Connected to earth or to the lowest-potential point in a circuit, without any intervening resistance or reactance.

directly heated cathode A vacuum-tube filament. It is so called because, when heated, it becomes the cathode of the tube (i.e., the emitter of electrons).

directly heated thermistor A thermistor whose temperature changes with the surrounding temperature, and also as a result of power dissipation in the device itself. Compare INDIRECTLY HEATED THERMISTOR.

directly heated thermocouple A meter thermocouple heated directly by signal currents passing through it. Compare INDIRECTLY HEATED THERMOCOUPLE.

direct measurement Immediate measurement of a quantity, rather than determining the value of the quantity through adjustments of a measuring device (e.g., measuring capacitance with a capacitance meter, rather than with a bridge). Compare INDIRECT MEASUREMENT.

direct memory access Abbreviation, DMA. The transfer of data from a computer memory to some other location, without the intervention of the central processing unit (CPU).

direct numerical control In a computer or data system, the capability for distributing information among numerically controlled machines whenever desired.

director In a multielement directional antenna, an element that is usually mounted in front of the radiator element, and that is phased and spaced to direct the radiation forward. The director functions in conjunction with the reflector element, which is usually mounted behind the radiator.

directory See DICTIONARY.

direct pickup The broadcasting, especially in television, of events at the same time as they occur (e.g., without recording/reproduction).

direct piezoelectricity The production of a piezoelectric voltage by mechanically stressing a suitable crystal.

direct playback In audio or video recording, the reproduction of a recording without additional processing (e.g., the playing of an original recorded tape, rather than a tape that has been mass produced).

direct-point repeater A relay-operated telegraph repeater. The received signals actuate the relay, which switches the second line.

direct-radiator loudspeaker A loudspeaker whose cone or diaphragm is directly coupled to the air.

direct ray An electromagnetic ray (wave) that reaches a receiver without reflection or refraction, and without encountering obstructions.

direct recording **1.** A record produced by a graphic recorder. **2.** The technique of producing such a record.

direct-recording instrument A device, such as a graphic recorder, that directly produces a permanent record (such as an inked trace) of the variations of a quantity.

direct resistance coupling A form of coupling in which the output of the first amplifying device is connected through a resistor directly to the input of the second device. The resistance value can vary; sometimes the connection is a short circuit.

directrix A fixed line to which a curve is referred (e.g., the axis of a parabola).

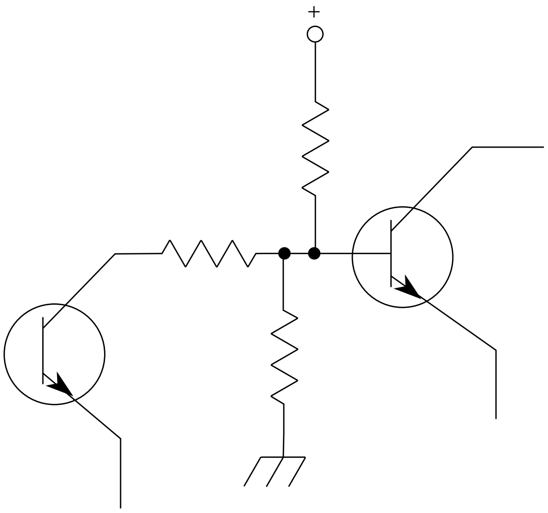

direct resistance coupling

direct scanning In television, the sequential viewing of parts of a scene by the camera—even though the entire scene is continuously illuminated.

direct serial file organization A technique of organizing files stored in a direct access device, in which a record can be chosen by number and amended where it is without altering other members of the file.

direct sound wave A sound wave arriving directly from its source—especially a wave within an enclosure that is not affected by reflection.

direct substitution **1.** An exact component replacement. **2.** Installing an exact component replacement.

direct synthesizer A device for producing random, rapidly changing frequencies for security purposes. A reference oscillator provides a comparison frequency; the output frequency is a rational-number multiple of this reference frequency.

direct voltage See DC VOLTAGE.

direct wave A wave that travels from a transmitter to a receiver without being reflected by the ionosphere or the ground. Compare SKYWAVE.

direct Wiedemann effect Twisting force (torque) in a wire carrying current in a longitudinal magnetic field. Occurs because of interaction between the longitudinal field and the circular magnetic field around the wire.

direct-wire circuit A communications or control line of wires connecting a transmitter (or control point) and a receiver (or controlled point) without an intermediary, such as a switchboard.

direct-writing recorder See GRAPHIC RECORDER.

direct-writing telegraph **1.** See PRINTING TELEGRAPH. **2.** See TELAUTOGRAPH.

dis- A prefix meaning "deprived of." For the formation of electronic terms, the prefix must be distinguished from *un-*, meaning "not." For example, a *discharged* body is one that was charged, but has been emptied of its charge. An *uncharged* body is one that ordinarily or presently is not charged.

disable **1.** To deliberately render a circuit, device, or system inoperative. **2.** In digital computer operations, to defeat a software or hardware function.

disc See DISK.

disc engraving **1.** Recording sound by cutting a groove in a record disc. **2.** The groove resulting from such a process.

discharge The emptying or draining of electricity from a source, such as a battery or capacitor. The term also denotes a sudden, heavy flow of current, as in DISRUPTIVE DISCHARGE. Compare CHARGE.

discharge current **1.** Current flowing out of a capacitor. **2.** Current flowing out of a cell—especially a storage cell. Compare CHARGING CURRENT.

discharge key See DISCHARGE SWITCH.

discharge lamp A gas-filled tube or globe in which light is produced by ionization of the gas between electrodes. Familiar examples are the neon bulb and fluorescent tube.

discharge phenomena The effects associated with electrical discharges in gases, such as luminous glow.

discharge potential See IONIZATION POTENTIAL.

discharger **1.** A short-circuiting tool for discharging capacitors. **2.** A spark gap or other device for automatically discharging an overcharged capacitor.

discharge rate **1.** The current that can be supplied by an electrochemical cell or battery reliably during its discharging cycle. Usually expressed in milliamperes or amperes. **2.** An expression of the speed with which a battery is being discharged at a specific point in time. It is usually specified in amperes or milliamperes.

discharge switch A switch for connecting a charged capacitor to a resistor or other load, through which the capacitor discharges. In some circuits, when the switch is in its resting position, it connects the capacitor to the charging source.

discharge voltage See IONIZATION POTENTIAL.

discharging **1.** The conversion of chemical energy to electrical energy by an electrochemical cell or battery. **2.** The release of stored electrical energy from a capacitor, or from a network containing capacitors.

discharging tongs See DISCHARGER, **1**.

discone antenna An antenna consisting of a horizontal metal or wire-mesh disk above a metal or wire-mesh cone. The antenna has an omnidirectional radiation pattern in the horizontal plane, and provides a good match to a coaxial transmission line over a frequency range of several octaves. Commonly used at very-high frequencies (VHF) and ultra-high frequencies (UHF).

disconnect **1.** To separate leads or connections, thereby interrupting a circuit. **2.** A type of connector whose halves can be pulled apart to open a cable or other circuit quickly. **3.** To open a switch or relay.

disconnector See DISCONNECT, **2** and DISCONNECT SWITCH.

disconnect signal A signal sent over a telephone line, ending the connection.

disconnect switch A switch whose main function is to open a circuit quickly (either manually or automatically) in the event of an overload.

discontinuity **1.** A break in a conductor. **2.** A point at which the impedance in a transmission line abruptly changes.

discontinuous wave trains See DAMPED WAVES.

discrete **1.** Complete and self-contained, as opposed to a part of something else. **2.** Composed of individual, separate members.

discrete capacitor Capacitance that is entirely self-contained, rather than being electrically distributed. Also called LUMPED CAPACITOR. Compare DISTRIBUTED CAPACITANCE.

discrete circuit A circuit comprised of discrete components, such as resistors, capacitors, diodes, and transistors, not fabricated into an integrated circuit.

discrete component A self-contained device that offers one particular electrical property in lumped form (i.e., concentrated at one place in a circuit, rather than being distributed). A discrete component is built especially to have a specific electrical property, and exists independently, not in combination with other components. Examples: disk capacitor, toroidal inductor, and carbon-composition resistor. Compare DISTRIBUTED COMPONENT.

discrete device Any component or device that operates as a self-contained unit.

discrete element A discrete device that forms part of a larger system.

discrete inductor An inductive component that is entirely self-contained, rather than being electrically spread out. Also called *lumped inductor.* Compare DISTRIBUTED INDUCTANCE.

discrete information source A source of data containing a finite number of individual elements, rather than a continuously variable parameter.

discrete part See DISCRETE COMPONENT.

discrete resistor A resistive component that is entirely self-contained, rather than being electrically spread out. Also called *lumped resistor.* Compare DISTRIBUTED RESISTANCE.

discrete sampling Sampling of individual bits or characters, one or more at a time.

discrete thin-film component A discrete component produced by the thin-film process (e.g., thin-film capacitor, thin-film potentiometer, etc.).

discretionary wiring A method of interconnecting the components and circuits on a semiconductor wafer for optimum performance. This requires a separate analysis and wiring pattern for every chip.

discrimination **1.** Sharp distinction between electrical quantities of different value. **2.** The detection of a frequency-modulated (FM) signal (i.e., the delivery of an audio signal corresponding to the frequency or phase variations in the FM carrier).

discriminator A second detector for frequency-modulated (FM) signals, in which two diodes are operated from the center-tapped secondary of a special intermediate-frequency (IF) transformer. The circuit is balanced for zero output when the instantaneous received signal frequency is at the unmodulated carrier frequency; the circuit delivers output when the instantaneous received signal frequency swings above or below the unmodulated carrier frequency. Also see FOSTER-SEELEY DISCRIMINATOR and TRAVIS DISCRIMINATOR.

discriminator transformer The special input transformer in a DISCRIMINATOR.

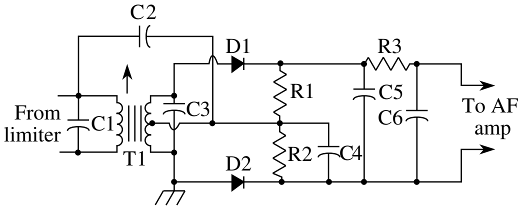

discriminator

discriminator tuner A device that tunes a discriminator to a selected subcarrier.

discriminator tuning device See DISCRIMINATOR TUNER.

dish See DISH ANTENNA.

dish antenna A transmitting and/or receiving antenna consisting of a driven element and a large reflector. The reflector has the shape of a shallow, circular section of a paraboloid or sphere. The feed point is at the focus of the reflector. This antenna, noted for its high directivity and gain, is used mainly at ultra-high and microwave frequencies for communications and satellite television. Large antennas of this type are used in some radio telescopes.

dish-type construction A type of panel-and-chassis construction in which the chassis is fastened vertically to the back of the panel.

disintegration **1.** The destructive breakdown of a material. **2.** The stripping of a vacuum-tube cathode of its emissive coating (see DISINTEGRATION VOLTAGE). **3.** The decay of a radioactive material.

disintegration voltage The anode voltage at which the cathode of a gas tube begins to be stripped of its electron-emitting material. For safety and reasonable tube life, the anode working voltage must be between the ionization and disintegration values.

disintegrator An ultrasonic device for reducing crystals or particles to fine suspensions.

disjunction The logical inclusive-OR operation.

disk **1.** A flat, circular plate (e.g., *rectifier disk*). **2.** See DISKETTE. **3.** See HARD DISK. **4.** See COMPACT DISC. **5.** See CD-ROM. **6.** A phonograph record or the equivalent unrecorded blank.

disk capacitor A fixed (usually two-plate) capacitor consisting of a disk of dielectric material on whose faces are deposited metal-film plates.

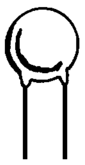

disk capacitor

disk coil See DISK WINDING.

disk dynamo A rudimentary direct-current (dc) generator, in which a copper disk rotates between the poles of a permanent magnet. The outer edge of the disk becomes positively charged; the center of the disk becomes negatively charged.

diskette A magnetic recording disk used for microcomputer data storage. Housed in a square, flat case. In personal computing, there are two sizes: 5.25 inches square and 3.5 inches square. The 5.25-inch version is flexible and is sometimes called a *floppy disk*.

disk files An information-storage system in which data are recorded on rotating magnetic disks.

disk generator **1.** See DISK DYNAMO. **2.** A disk-type electrostatic generator.

disk memory A common misnomer for DISK STORAGE.

Disk Operating System Abbreviation, DOS. Any of several command-driven operating systems commonly used in IBM-compatible personal computers.

disk pack In disk files, a set of disks that can be handled as a single unit.

disk recorder A device for recording (and usually also playing back) sound or other signals on record disks.

disk recording **1.** Recording sound or other signals on disks. **2.** A disk resulting from such a recording. See DISK.

disk rectifier A semiconductor rectifier (such as copper-oxide, selenium, magnesium-copper-sulfide, or germanium type) in which the active material is deposited on a metal disk.

disk resistor A resistor consisting of a resistive material deposited on a metal disk; or a disk of resistive material. In the latter, electrodes are plated on the faces of the disk, one or more of which are held between clips or screws for connections.

disk storage In digital computer systems, an online or offline data storage scheme, in which information is recorded on the magnetic coating of a rotating disk or set of disks. See DISKETTE, and HARD DISK.

disk system A sound-motion-picture system using audio disks synchronized with the film.

disk thermistor A thermistor having the general shape of a disk.

disk-type storage See DISK STORAGE.

disk varistor A varistor having the general shape of a disk.

disk winding An armature or coil winding that is flat, rather than cylindrical. Also called *disk coil*, *pancake coil*, and *spiral coil.*

dislocation A crystal region in which the arrangement of atoms does not have the perfect lattice structure of the crystal.

dispersion **1.** The property of a material that causes energy at different wavelengths to pass through it at different speeds. **2.** The separation of a wave into its various component frequencies (as when white light is broken up into the color spectrum by a prism). **3.** The scattering of a microwave beam when it strikes an obstruction. **4.** The scattering of sound or ultrasound as it emanates from an acoustic transducer. **5.** A suspension of finely divided particles within another substance.

dispersive medium A medium that disperses a wave passing through it.

displacement **1.** A change in the position of a point, particle, figure, or body. **2.** The vector representing a change in the position of a point, particle, figure, or body. **3.** Movement of a member through a specified distance.

displacement current **1.** An alternating current proportional to the rate of change of an electric field, and existing in addition to usual conduction current. **2.** The current flowing into a capacitor immediately after application of a voltage. This current continues to flow, although continually diminishing in value, until the capacitor becomes fully charged.

displacement of porches In a television signal, the amplitude difference between the front porch and back porch of a horizontal sync pulse.

displacement of vectors Vector rotation through a specified number of angular degrees or radians.

displacement transducer A transducer in which movement (displacement) of a rod, armature, core, reed, or other object converts mechanical energy into proportionate electrical energy.

display **1.** Visually observable presentation of information, such as data entered into a computer, an answer to a problem solved by a computer, the value of a measured quantity, or a graph of a function. **2.** The screen in a computer system or terminal that visually portrays text and graphical data. In laptop, notebook, and portable computers, this is usually a *liquid-crystal display (LCD)*; in desktop computers and terminals, it is usually a *cathode-ray tube (CRT).* **3.** To portray information in a visual manner (e.g., as text, numerals, symbols, or graphic images).

display blanking See DISPLAY INHIBIT.

display console In a computer system, a peripheral that is used to access and display data being processed or stored; often, it is a unit with a cathode-ray tube (CRT), keyboard, and light pen.

display control An interface device between a central processor and several visual display units (terminals).

display dimming See DISPLAY INHIBIT.

display inhibit In a digital meter, the blanking or dimming of the display when the instrument is not being used. It is used to conserve battery energy.

display loss The ratio P_1/P_2, where P_1 is the minimum input-signal power that can be detected by an ideal output device at the output of a receiver, and P_2 is the minimum input-signal power value seen by an operator using an output device with the same receiver. Also called *visibility factor.*

display mode **1.** A particular method of presenting a display. For example, a character display on a video unit might consist of bright characters on a dark background, or dark characters on a light background. **2.** An operating mode for a particular device, in which a display is used.

display module A self-contained unit with circuitry and readouts for indicating a numerical count.

display primaries Also called *primary colors.* In a color television receiver, the colors red, green, and blue. When mixed correctly, these three colors can produce any visible hue.

display-storage tube A special cathode-ray tube in which patterns and other information can be stored for later viewing. The tube has two electron guns: a writing gun and a reading (viewing) gun.

display unit A device that presents information for visual reading. Included are analog and digital meters, cathode-ray tubes, data printers, graphic recorders, etc. Also see DISPLAY CONSOLE.

display visibility The ease with which a display can be read by an operator.

display window **1.** In a panoramic display, the width of the presented frequency band in hertz. **2.** The panel opening through which the indication of a display unit appears.

displayed part That portion of a number displayed in the readout of a calculator or computer. There might be digits that are not displayed, but which the machine might take into consideration when making calculations. For example, in a 10-digit calculator display, the number 245.789378214895 would be displayed as 245.7893782. Depending on the calculator design, the machine might truncate (disregard) the undisplayed digits (14895), or take the undisplayed digits into account when making calculations.

disposable component A circuit component or machine part that is so inexpensive that it is more cost-effective to discard it than to repair it when it fails.

disruptive discharge Sudden, heavy current flow through a dielectric material when it fails completely under electric stress.

dissector A transducer that samples an illuminated image point by point.

dissector tube A camera tube using a flat photocathode, upon which the image is focused by the lens system. Electromagnetic deflection from external coils provides scanning. Electrons pass sequentially from the image cathode to a scanning tube at the opposite end of the camera tube. Also called *Farnsworth dissector tube* and *orthiconoscope.*

dissipation **1.** The consumption of power, often without contributing to a useful end, and usually accompanied by the generation of heat. **2.** In an amplifier, the difference between the collector, drain, or plate input power and the usable output power.

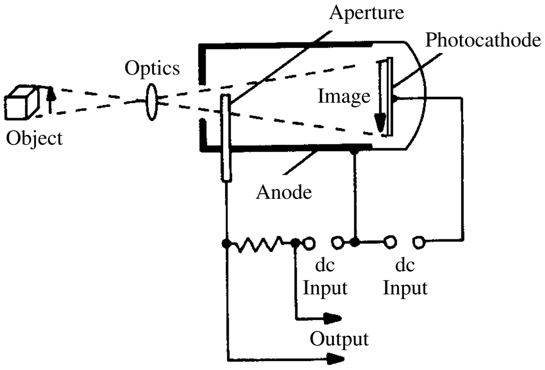

dissector tube

dissipation constant For a thermistor, the ratio of the change in power dissipation to a corresponding change in body temperature.

dissipation factor **1.** For a dielectric material, the tangent of the dielectric loss angle. Also called *loss tangent.* **2.** Symbol, *D.* For an impedance (such as a capacitor), the ratio of resistance to reactance; $D = R/X$. It is the reciprocal of the *figure of merit (Q).*

dissipation line A resistive section of transmission line, used for dissipating power at a certain impedance. Two parallel lengths of resistance wire are terminated by a large, noninductive resistor that has a value equal to the characteristic impedance of the line.

dissipator **1.** A device used primarily to consume power (i.e., a power sink). **2.** A device for removing heat generated by a device's operation (e.g., a heatsink attached to a power transistor).

dissociation The condition that characterizes electrolytes (certain acids, bases, or salts in water solution) in which the molecules of the material break up into positive and negative ions.

dissonance The unpleasant effect (especially in music) produced by nonharmonious combinations of sounds.

dissymmetrical network A network having unequal input and output image impedances.

dissymmetrical transducer A transducer having unequal input and output image impedances.

distance-double law A theoretical rule for determining the rate at which sound intensity decreases as distance increases. Under ideal conditions, when the distance from a sound source is doubled, the sound pressure is reduced to one-fourth of its original level, a reduction of 6 decibels. This is analogous to the *inverse-square law* for visible light and other radiant energy.

distance mark On a radar screen, a mark indicating the distance from the radar set to the target.

distance measurement Also called *ranging.* A method or system that allows a robot to navigate in its environment. It also allows a central computer to track the locations of robots under its control. Can use radar, sonar, visible light, or infrared.

distance-measuring equipment In radionavigation, a system that measures the distance of the interrogator to a transponder beacon in terms of the transmission time to and from the beacon.

distance protection The use of a protective device within a specified electrical distance along a circuit.

distance relay In circuit protection, a relay that operates to remove power when a fault occurs within a predetermined distance along the circuit.

distance resolution **1.** Qualitatively, the ability of a ranging system to differentiate between two objects or beacons that are almost, but not quite, the same distance away. See RANGING. **2.** Quantitatively, the minimum radial separation of objects or beacons necessary for a ranging system to tell them apart. **3.** For two targets having the same azimuth bearing, the minimum difference in range for which a radar display renders them as distinct blips.

distant control See REMOTE CONTROL.

distorted-drive multiplier A frequency multiplier whose excitation signal is a peaked wave that has been predistorted to decrease the angle of flow in the device, thus increasing its efficiency.

distorted nonsinusoidal wave A nonsinusoidal wave whose ideal shape (square, rectangular, sawtooth, etc.) has been altered.

distorted sine wave A wave that is approximately of sinusoidal shape (i.e., it is not an exact plot of a sine wave because of the presence of harmonics).

distortion **1.** Deformation of a signal waveform. **2.** The additional deformation of a signal exhibiting a less-than-ideal waveshape when it passes through a circuit. Some distortion originates within the signal generator itself; other forms result from circuits and devices transmitting the signal. **3.** Any degradation in the quality of a high-fidelity audio signal. **4.** See TOTAL HARMONIC DISTORTION.

distortionless **1.** Having no distortion. **2.** Having a propagation velocity that does not depend on frequency.

distortion meter An instrument for measuring harmonic distortion. It consists of a highly selective band-rejection filter (notch filter) that removes the fundamental frequency of the signal under test, and a sensitive voltmeter that can be switched between the filter input and the filter output. The distortion percentage is determined from the ratio between filter-output and filter-input voltages.

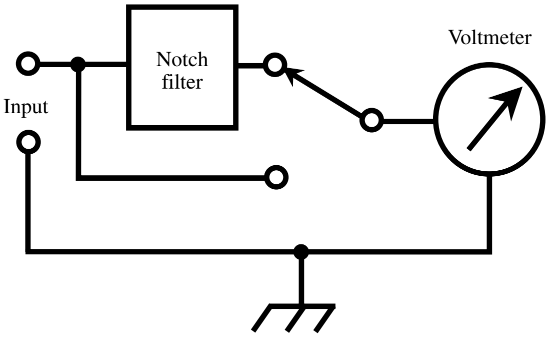

distortion meter

distortion tolerance The maximum amount of distortion that can be present in a signal without making it useless. This varies over wide limits. The maximum harmonic distortion that might be acceptable in a high-fidelity sound system could be less than 0.1% total, whereas in some applications of ac power, 10% would be acceptable.

distress frequency A radio frequency on which an emergency signal is transmitted. Ships at sea and aircraft over the sea use 500 kHz (by international agreement). In Citizen Band communications, channel 9 has been set aside for emergency use.

distress signal A signal indicating that trouble exists at the transmitting station and imploring aid from the recipient. The international radiotelegraph distress signal is the three-letter combination *SOS*; the international radiotelephone distress signal is the word *mayday*, the phonetic equivalent of the French *m'aidez* (help me).

distributed Existing over a measurable interval, area, or volume; not concentrated in a single place or places.

distributed amplifier A wideband, untuned amplifier whose active devices are spaced (distributed) along parallel, artificial delay lines consisting of coils that act in combination with the input and output capacitances of the active devices. Adding active devices to the lineup increases the gain. Commonly used as preamplifiers for television receivers.

distributed capacitance Symbol, C_d. Capacitance that is dispersed throughout a component or system, rather than being lumped in one place. An example is the distributed capacitance of a coil.

distributed component An electrical property that is spread throughout a circuit or device, rather than being concentrated at one point, as in a discrete component. For example, DISTRIBUTED CAPACITANCE and DISTRUBUTED INDUCTANCE are spread along the length of a transmission line. Another example is the DISTRIBUTED RESISTANCE of a wire coil. Distributed components are often unintended, but they can be useful. Compare DISCRETE COMPONENT and LUMPED COMPONENT.

distributed constant See DISTRIBUTED COMPONENT.

distributed-constant delay line A delay line whose capacitance and inductance are distributed throughout the line. Compare LUMPED-CONSTANT DELAY LINE.

distributed inductance Symbol, L_d. Inductance that is dispersed throughout a system or component, rather than being lumped in one place, such as in a coil (e.g., the inductance of an antenna or capacitor).

distributed network **1.** A network in which electrical properties (such as resistance, inductance, and capacitance) are distributed over a measurable interval, area, or volume. **2.** A network whose characteristics do not depend on frequency within a given range.

distributed-parameter network A network composed of distributed components, rather than lumped components.

distributed pole In a motor or generator, a pole having a DISTRIBUTED WINDING.

distributed resistance Symbol, R_d. Resistance that is dispersed throughout a component or circuit, rather than being lumped in one place, such as in a resistor. An example is the high-frequency resistance of an antenna system.

distributed-shell transformer A transformer having two complete closed cores that are perpendicular to each other.

distributed winding In a motor or generator, a winding that is placed in several slots (rather than in one slot) under a pole piece.

distributing amplifier An amplifier having a single input and two or more outputs that are isolated from each other; it distributes signals to various points.

distributing cable **1.** In cable television, the cable connecting the receiver to the transmission cable. **2.** In power service, the cable running between a feeder and a consumer's house.

distribution **1.** The selective delivery of a quantity (e.g., *power distribution*). **2.** In statistical analysis, the number of times particular values of a variable appear. Also called *frequency distribution.*

distribution amplifier A low-output-impedance power amplifier that distributes a radio, television, or audio signal to a number of receivers or speakers.

distribution cable See DISTRIBUTING CABLE.

distribution center **1.** The central point from which a signal is routed to various points of use. **2.** In electric power operations, the point at which generation, conversion, and control equipment is operated to route power to points of use.

distribution factor For a polyphase alternator, the factor by which the total voltage V_T can be determined in terms of the coil voltage V_C and the number of coils n: $V_T = nV_C$. Distribution factor $k_d = (\sin(sd/2))/(s \times \sin(d/2))$, where s is the number of slots per phase per pole, and d is the angle between adjacent slots.

distribution function In statistical analysis, the function $F(x)$ expressing the probability that F takes on a value equal to or less than x.

distribution switchboard **1.** A switchboard through which signals can be routed to or among various points. **2.** A switchboard for routing electric power to points of use.

distribution transformer A step-down transformer used to supply low-voltage alternating-current (ac) utility power to one or more consumers from a high-voltage line.

distributor **1.** See COMMUTATOR. **2.** A switching device consisting of a rotating blade and a number of contacts arranged in a circle. Accomplishes sequential switching of a voltage to a number of points in a circuit. A common example is the distributor in the ignition system of an automotive engine.

disturbance An undesired variation in, or interference with, an electrical or physical quantity.

disturbed-one output In digital computers, the one output of a magnetic cell that has received only a partial write pulse train because it was last written into. Compare UNDISTURBED-ONE OUTPUT.

disturbed-zero output In digital computers, the zero output of a magnetic cell that has received only a partial write pulse train since it was last read from. Compare UNDISTURBED-ZERO OUTPUT.

dither **1.** Vibrate; quiver. **2.** The condition of vibration or quivering (e.g., the dither of a meter pointer). **3.** To blend pixels in a digitized image to obtain various shades and colors.

divergence **1.** The tendency of a collimated beam of energy to spread out. **2.** The extent to which a collimated beam of energy spreads out, generally measured in seconds of arc, minutes of arc, angular degrees, or angular radians.

divergence loss Loss of transmitted sound energy, resulting from spreading.

diverging lens A lens having a virtual focus for parallel rays; generally a *concave lens.*

diversity **1.** The property of consisting of two or more independent components or media. **2.** See DIVERSITY RECEPTION. **3.** See DIVERSITY TRANSMISSION.

diversity factor **1.** A measure of the degree to which a system exhibits unity among its constituents. **2.** The sum of the requirements of each constituent of a system, divided by the total requirement of the system.

diversity gain **1.** Signal gain achieved by using two or more receiving antennas. **2.** Signal gain achieved by using two or more transmitting antennas.

diversity reception Also called *dual-diversity reception.* A method of minimizing the effects of fading in ionospheric communication at high frequencies (HF). Accomplished using two receivers whose antennas are 5 to 10 wavelengths apart. Each receiver, tuned to the same signal, feeds a common audio amplifier. The timing of the fading is different at the two antennas because of phasing effects. The composite signal, therefore, fades less than either of the component signals. Some diversity systems use three or more antennas and receivers to reduce the effects of fading even further; this is sometimes called *multiple-diversity reception.*

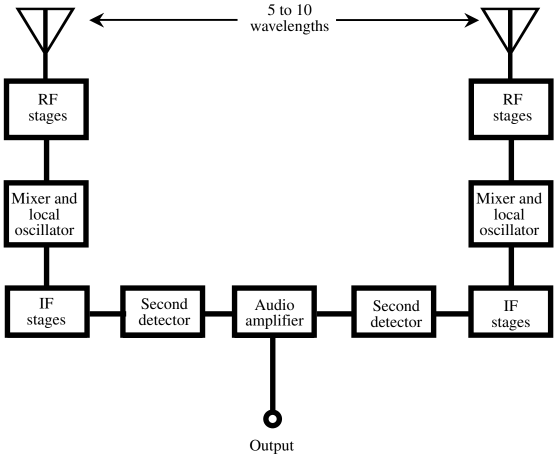

diversity reception

diversity transmission Also called *dual-diversity transmission* or *multiple-diversity transmission.* A scheme similar to DIVERSITY RECEPTION, except applied at the transmitting end of a communication circuit. The signals from two or more transmitters, at identical frequencies, are fed to antennas spaced several wavelengths apart.

diverter-pole generator A well-regulated direct-current (dc) generator, whose shunt winding is on the main field pole, the series winding being on a diverter pole whose flux opposes that of the main pole.

divide-by-seven circuit A three-stage binary circuit having feedback from stage three to stage one. Stage three is switched on by the fourth input pulse; at that time, the feedback pulse switches on stage one, simulating one input pulse and reducing the usual counting capacity from eight to seven.

divide-by-two circuit A circuit that delivers one output pulse for each two successive input pulses (i.e., a *flip-flop*).

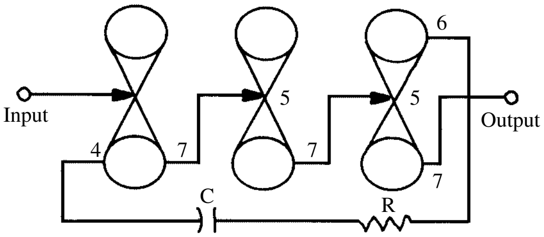

divide-by-seven circuit

divided-carrier modulation Modulation obtained by adding two identical frequency carriers that are 90 degrees out of phase.

divided circuit A *parallel circuit.*

divided equipment A system of modular electronic components interconnected with cables. A simple example is a radio receiver having an external power supply and external loudspeaker.

divider **1.** See VOLTAGE DIVIDER. **2.** See FREQUENCY DIVIDER. **3.** See PULSE-COUNT DIVIDER. **4.** A computing circuit or device for performing mathematical division.

divider probe A test probe that divides an applied signal voltage by some factor (such as 2, 5, or 10) to place it within the range of the instrument with which the probe is used.

dividing network See CROSSOVER NETWORK.

division **1.** Separating a quantity into a number of equal parts, as indicated by the divisor. **2.** Voltage division (see VOLTAGE DIVIDER). **3.** Frequency division (see FREQUENCY DIVIDER). **4.** Pulse-count division (see PULSE-COUNT DIVIDER).

division of vectors **1.** The quotient of two rectangular vectors determined by the principle of rationalization in algebra (i.e., by multiplying the numerator and denominator of the indicated division by the conjugate of the denominator, simplifying, and performing the division). **2.** To find the quotient of two polar vectors: the quotient of their moduli and the difference of their arguments.

dj Abbreviation of *diffused junction.*

DKT Abbreviation of *dipotassium tartrate.*

D layer A layer of the ionosphere that is below the E layer; its altitude is approximately 60 kilometers.

dm Abbreviation of *decimeter.*

DMA Abbreviation of DIRECT MEMORY ACCESS and *direct memory addressing.*

DME Abbreviation of DISTANCE-MEASURING EQUIPMENT.

DMM Abbreviation of DIGITAL MULTIMETER.

DMOS Abbreviation for *double-diffused metal-oxide semiconductor*, a type of field-effect transistor that exhibits extremely low capacitance and low source-drain resistance when conducting.

DNL See DIFFERENTIAL NONLINEARITY.

DNS Abbreviation of *Doppler navigation system.*

doctor To use unconventional (sometimes substandard) methods in fixing a circuit or device or in correcting a bad design.

document **1.** In digital computer operations, especially in file maintenance, a form that provides information pertinent to a transaction. Also see TRANSACTION. **2.** To perform documentation (see DOCUMENTATION, **2**). **3.** A computer text file.

documentation **1.** Paperwork explaining the scope of programs and how they can be optimized. **2.** Annotating a computer program at critical points during its writing (e.g., so that the purpose of various segments are understood). A measure of good programming, documentation becomes especially valuable for program modification or debugging.

document reader An electronic device that reads printed cards, usually for data entry into a computer.

dog **1.** A malfunctioning circuit or device. **2.** The cause of a circuit or device malfunction.

doghouse An enclosure for antenna loading inductors and other resonating components, placed at the base of a vertical broadcasting tower.

dog whistle See ULTRASONIC WHISTLE.

Doherty amplifier A highly efficient linear radio-frequency (RF) amplifier in which a *carrier tube* and a *peak tube* operate jointly, both receiving amplitude-modulated RF excitation. During unmodulated intervals, the carrier tube supplies carrier power to the load, while the peak tube, biased to cutoff, idles. On positive modulation peaks, the peak tube supplies output power that combines with that of the carrier tube, the increase in power corresponding to the condition of full modulation of the carrier. On negative modulation peaks, the peak tube does not supply power, and the output of the carrier tube is reduced to zero.

Dolby An electronic method of improving the audio reproduction quality of magnetic-tape systems. The gain is increased for low-level sounds during the recording process. During playback, the gain of the low-level sounds is reduced back to its original level.

Dolby A A Dolby system with four frequency ranges, operated independently. It is used mostly by recording professionals.

Dolby B A modified form of Dolby A, with only one band of noise-reducing circuitry. It is used primarily by hobbyists.

Dolezalek electrometer See QUADRANT ELECTROMETER.

dolly **1.** A low, wheeled frame or platform for transporting electronic equipment. **2.** A tool with which one end of a rivet is held while the head is hammered out of the other end.

DOM Abbreviation of DIGITAL OHMMETER.

domain **1.** A region of unidirectional magnetization in a magnetic material. **2.** A region of unidirectional polarization in a ferroelectric material. **3.** A region in which a variable is confined.

domestic electronics Also called *consumer electronics.* The branch of electronics concerned with appliances, automatic controls, protective devices, entertainment systems, communications devices, and other equipment for the home.

domestic induction heater A household cooking utensil heated by currents induced in it. A primary coil (connected to the power line) is imbedded in the utensil, which acts as a short-circuited secondary coil.

dome tweeter A speaker designed for high-frequency (treble), high-fidelity audio, and often functional at frequencies considerably above the limit of the human hearing. Characterized by a convex diaphragm. Usually part of an assembly including a woofer and midrange speaker.

dominant In statistical analysis, the nature of any quantity that imposes its effects even in the presence of other quantities.

dominant mode In a waveguide, the propagation mode exhibiting the lowest cutoff frequency.

dominant wave In a waveguide, the wave having the lowest cutoff frequency.

dominant wavelength For visible light of a given hue, the wavelength at which the emitted energy is the greatest.

donor An electron-rich impurity added to a semiconductor to make it into an n-type material. So called because it donates its excess electrons. Compare ACCEPTOR.

donor atom An atom having an excess electron. When a substance having such atoms is added to an intrinsic semiconductor, the extra electron is donated, making the semiconductor into an n-type material.

donor impurity A substance whose atoms have excess electrons, and that donates electrons to the atomic structure of the semiconductor crystal to which it is added. Donor elements make semiconductors into n-type materials. Also see DONOR ATOM. Compare ACCEPTOR IMPURITY.

do-nothing instruction A computer program instruction that causes no action to be taken. Can be used to provide space for future program updating, or to fill out a block of instructions, as needed by a compiler. Also called *dummy instruction.*

don't-care state In a logic function or gate, an input digit whose state (high or low) does not affect the output.

donut capacitor A flat, ring-shaped capacitor.

donut coil See TOROIDAL COIL.

donut crystal A relatively large, zero-temperature-coefficient piezoelectric quartz crystal cut in the form of a torus with the y-axis passing through the center of rotation.

donut magnet See RING MAGNET.

donut pattern The three-dimensional radio-frequency (RF) radiation/response pattern of a free-space straight antenna element measuring ½ wavelength, neglecting the effects of ground and nearby objects.

doohickey A usually unnamed device—especially one used to achieve some significant modification of circuit performance.

doorknob capacitor A high-voltage fixed capacitor, so called from its round package, which somewhat resembles a doorknob.

doorknob tube A special UHF vacuum tube, so called from its characteristic shape. The unique design provides short electron-transit time and low interelectrode capacitance. Largely replaced in recent years by semiconductor devices.

dopant An impurity added in controlled amounts to a semiconductor to make it an n-type or p-type material. Also see ACCEPTOR and DONOR.

dope To add impurities to a semiconductor material. Doping allows the manufacture of n-type or p-type semiconductors with varying degrees of conductivity. In general, the greater the extent of doping, the higher the conductivity.

doped junction In a semiconductor device, a junction produced by adding a dopant to the semiconductor melt.

doping Adding a dopant to a semiconductor to alter the way it conducts current.

doping agent See DOPANT.

doping compensation Opposite doping (i.e., adding a donor impurity to p-type semiconductor material or adding an acceptor impurity to n-type semiconductor material).

doping gas A gas diffused into a semiconductor material to dope it. For example, phosphorus pentoxide gas can be used to create an n-type region in a p-type silicon chip.

doping level The relative concentration of impurity added to a semiconductor material to obtain a certain resistivity and polarity. The greater the doping level, the lower the resistivity.

Doppler broadening In a spectrum, the spreading out or blurring of a spectral line caused by DOPPLER EFFECT, in turn resulting from motion of molecules, atoms, or other particles in the medium.

Doppler cabinet A loudspeaker enclosure with which a vibrato effect is achieved by rotating or reciprocating either the loudspeaker or a baffle board; the length of the sound path is altered cyclically.

Doppler effect A change in the frequency of a wave that occurs when the source and observer are in relative motion. The frequency of the wave increases (the wavelength shortens) as the source and observer approach each other; the frequency decreases (the wavelength becomes greater) as the source and observer recede from each other. This effect is often observed with sound waves, as when the pitch of an automobile horn seems to rise as the car approaches and to fall as the car passes. The effect is also observable in electromagnetic radiation at all wavelengths. It affects satellite communication and space communication.

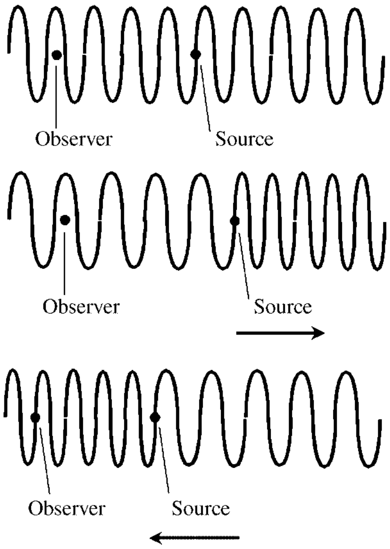

Doppler effect

Doppler enclosure See DOPPLER CABINET.

Doppler radar A radar that uses the change in carrier frequency of the signal returned by a moving target (approaching or receding) to measure its velocity. Used by law enforcement officers to determine the speed of moving vehicles. Also used by meteorologists to evaluate air circulation patterns in thunderstorms, and to determine wind speeds in hurricanes and tornadoes.

Doppler ranging See DORAN.

Doppler shift The extent to which the frequency or wavelength of a signal changes because of DOPPLER EFFECT. Can be measured in Hertz (for frequency) or in meters (for wavelength). In astronomy, the shift is also measured as displacement of absorption or emission lines in an infrared, visible, or ultraviolet spectrum.

Doppler's principle See DOPPLER EFFECT.

doran A continuous-wave trajectory-measuring system utilizing Doppler shift (see DOPPLER EFFECT). The name is a contraction of *doppler ranging.*

dorsal column stimulator Abbreviation, DCS. A machine that generates radio-frequency energy that is applied to human tissues for the temporary relief of pain.

dosage meter See DOSIMETER.

dose The total quantity of radiation received upon exposure to nuclear radiation or X-rays.

dosimeter An instrument for measuring the amount of exposure to nuclear radiation or X-rays.

dot **1.** The shorter of the two characters (dot and dash) of the telegraph code. The dot, a short sound, mark, or perforation, is one-third the length (duration) of a dash. Compare DASH. **2.** One of the small spots of red, green, or blue phosphor on the screen of a color-television picture tube or cathode-ray-tube (CRT) computer display. **3.** A small spot of material alloyed with a semiconductor to form an alloy junction. **4.** The junction of two lines on a schematic diagram, representing a wired connection; also called *solder dot.*

dot AND Externally connected circuits or functions whose combined outputs result in an AND function. Compare DOT OR.

dot-and-dash telegraphy Telegraphy (wire or radio) by means of dot and dash characters.

dot cycle One period of an alternation between two signaling conditions, each of which is of unit duration (e.g., a unit mark followed by a unit space).

dot encapsulation A method of packaging cylindrical components by pressing them into the holes of perforated disks; interconnections are made, to complete a circuit, on each face of the disks.

dot generator A special radio-frequency (RF) signal generator used to produce a pattern of red, green, and blue dots on the screen of a color television receiver.

dot matrix A rectangular array of spaces, some of which are filled in to form alphanumeric and punctuation characters.

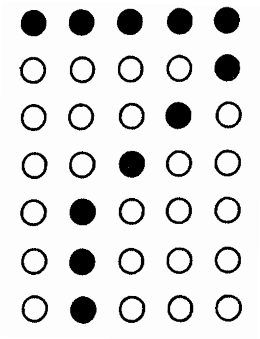

dot matrix

dot-matrix display A display that shows characters in dot-matrix form.

dot-matrix printer A computer output peripheral that prints characters and images on paper as a fine grid of dots. A print head, containing several pins, presses the ribbon against the paper as it moves laterally across each line. Can be used to print text and/or graphics.

dot movement pattern The movement of the red, green, and blue dots on the screen of a color television picture tube as the red, green, and blue magnets and the lateral magnet are adjusted for convergence of the dots at the center. The blue dots move horizontally or vertically; the red and green dots, diagonally.

dot OR Externally connected circuits or functions whose combined outputs result in an OR function. Compare DOT AND.

dot pattern In color television testing with a dot generator, dots of color (a red group, green group, and blue group) produced on the screen. With overall beam convergence, the three groups blend to produce white.

dot-sequential system The color television system in which the image is reproduced by means of primary-color dots (red, green, blue) sequentially activated on the screen of the picture tube. Compare FIELD-SEQUENTIAL SYSTEM and LINE-SEQUENTIAL SYSTEM.

double-amplitude-modulation multiplier A modulating system in which a carrier is amplitude-modulated first by one signal and then by a second signal. The resulting signal is fed to a detector, the output of which contains the product of the two modulating signals.

double-anode diode A semiconductor diode having two anodes and a common cathode.

double armature An armature (such as that of a dynamotor or a two-voltage generator) that has two separate windings on a single core, and has two separate commutators.

double-balanced mixer See BALANCED MIXER.

double-balanced modulator See BALANCED MODULATOR.

double-base diode See UNIJUNCTION TRANSISTOR.

double-base junction transistor A junction transistor having the usual emitter, base, and collector electrodes, plus two base connections, one on either side of the base region. The additional base connection acts as a fourth electrode to which a control voltage is applied. Also called *tetrode transistor.*

double-beam CRT See DUAL-BEAM OSCILLOSCOPE.

double-beam oscilloscope See DUAL-BEAM OSCILLOSCOPE.

double-bounce calibration In radar operations, a calibration technique for determining zero-beat error. Round-trip echoes are observed, the correct range being the difference between the two echoes.

double-bounce signal A signal that is received after having been reflected twice.

double-break contacts The member of a set of contacts that is normally closed on two others. Compare DOUBLE-MAKE CONTACTS.

double-break switch A switch that opens a previously closed circuit at two points simultaneously on closing. Compare DOUBLE-MAKE SWITCH.

double bridge See KELVIN DOUBLE BRIDGE.

double buffering In the input/output operation of a computer peripheral, the use of two memory areas for temporary storage.

double-button microphone A carbon microphone having two buttons mounted on each side of the center of a stretched diaphragm, and connected in push-pull. Also see BUTTON MICROPHONE.

double-channel duplex Two-way communication over two independent channels. One station transmits on one channel, and the other station transmits on the other channel. The result is conversation-mode communications, in which one operator can interrupt the other at any time; both receivers are always operational.

double-channel simplex A system of communication in which two channels are used. One station transmits on one channel, and the other station transmits on the other channel. Interruption is not possible because whenever either operator transmits, the station receiver is muted.

double-checkerboard pattern In a magnetic core memory, the maximum noise that appears when half of the half-selected cores are in the one state and the others are in the zero state. Also called *worst-case noise pattern.*

double circuit tuning A circuit whose output and input are tuned separately. Such tuning provides increased selectivity when the input and output are resonant at the same frequency, and decreased selectivity when they are tuned to different frequencies. Also see DOUBLE-TUNED AMPLIFIER.

double clocking A phenomenon that occurs in some digital circuits when the input pulse is nonuniform, and appears as two pulses to the device. The device is thus actuated at twice the desired frequency.

double-coil direction finder A radio direction finder (RDF) using an antenna that consists of two identical, perpendicular coils. The directivity of the antenna is the resultant of the directivity of the individual coils.

double conversion **1.** Two complete frequency conversions in a superheterodyne system. For example, the incoming signal might be converted to a 9-MHz first intermediate frequency (first IF); at a later stage, this signal might be converted to a 455-kHz second IF. The high first IF widely separates the signal from the image; the low second IF allows superior selectivity to be achieved at a reasonable cost. Also called *dual conversion.* **2.** Pertaining to a superheterodyne receiver with two intermediate frequencies.

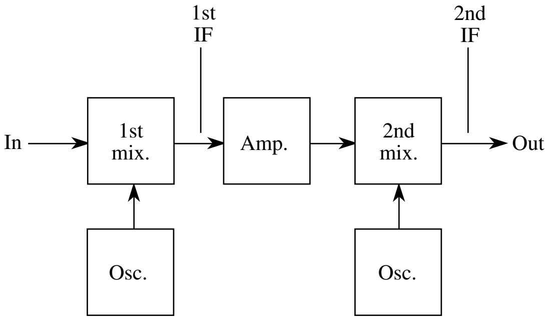

double conversion

double-conversion receiver Also called *double-conversion superheterodyne.* A superheterodyne receiver using DOUBLE CONVERSION to achieve optimum selectivity and image rejection.

double-current generator **1.** A dynamo-type generator supplying both alternating current (ac) and direct current (dc) from one armature winding. **2.** A rotary converter operating on dc and delivering ac.

double-diamond antenna A broadband antenna consisting of two rhomboid plates, one attached to each side of the feeder.

double-diffused epitaxial mesa transistor A transistor in which a thin mesa crystal is overlaid on another mesa crystal. Also called *epitaxial-growth mesa transistor.*

double-diffused transistor See DIFFUSED-EMITTER-AND-BASE TRANSISTOR.

double diode See DUODIODE.

double-diode limiter A limiter in which two diodes are connected back to back in parallel, to limit both peaks of an alternating-current (ac) signal.

double-doped transistor See GROWN-JUNCTION TRANSISTOR.

double edit In audio tape recording, to make two changes in a given span of the tape. For example, a producer might dislike the wording of a certain sentence, and re-record the sentence. Then, changing his or her mind, the producer might record the original sentence back over the re-recorded sentence. These changes increase the risk of audible irregularities appearing in the final recording.

double emitter follower See COMPOUND CONNECTION.

double-ended amplifier See PUSH-PULL AMPLIFIER and DOUBLE-ENDED CIRCUIT.

double-ended circuit A symmetrical circuit (i.e., one having identical halves, each operating on a half-cycle of the input signal). Example: a push-pull amplifier.

double-extended Zepp antenna A horizontal, collinear, center-fed antenna, in which each section measures 0.65 wavelength. This antenna gives increased gain over that of the Zepp and double Zepp (see DOUBLE ZEPP ANTENNA).

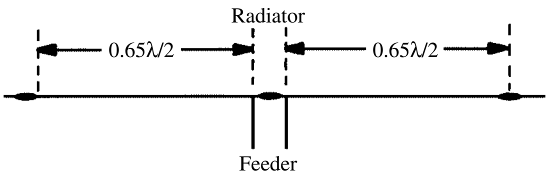

double-extended Zepp antenna

double-hump resonance curve A resonant response that is flattened by double tuning; it exhibits two resonance peaks. Also see DOUBLE-TUNED AMPLIFIER.

double-hump wave See DOUBLE-PULSE WAVE.

double image Two overlapping television pictures, one usually fainter than the other. Caused by the signal arriving over two different paths (one possibly attributable to reflection of the wave) and, hence, at different instants in time. The fainter image is called a GHOST.

double insulation The use of two layers of insulation on a conductor, made of different materials.

double ionization Ionization resulting from an electron colliding with an ion. In a gas, for example, a neutral atom might collide with an electron, which can knock an electron out of the atom. The atom then becomes a positive ion; it might in turn be bombarded by an electron, releasing still another electron.

double-junction photosensitive semiconductor See PHOTOTRANSISTOR.

double layer See HELMHOLTZ DOUBLE LAYER.

double local oscillator A mixer system in which a local oscillator generates two accurate radio-frequency (RF) signals separated by a few hundred hertz. The difference frequency is used as a reference in some applications.

double-make contacts A set of normally open contacts of which one closes against two others simultaneously. Compare DOUBLE-BREAK CONTACTS.

double-make switch A switch that closes a previously open circuit at two points simultaneously. Compare DOUBLE-BREAK SWITCH.

double moding In microwave operations, the abrupt changing of frequency at irregular intervals.

double modulation Using a modulated carrier to modulate another carrier of a different frequency.

double-play tape A thin magnetic recording tape that has approximately twice the playing time of the usual tape. Although the playing time is longer, double-play tape is more subject to jamming and stretching than standard-thickness recording tape.

double-pole Having two poles or switchable circuits (e.g., a double-pole switch).

double-pole, double-throw switch or relay Abbreviation, DPDT. A switch or relay having two contacts that can be closed simultaneously in one of two directions, to close or open two circuits.

double precision The use of two computer words to represent a single number to gain precision.

double-pole, single-throw switch or relay Abbreviation, DPST. A switch or relay having two contacts that can be closed in only one direction, to simultaneously close or open two circuits.

double precision hardware Within a computer, arithmetic units permitting the use of double-precision operands, sometimes also accommodating floating-point arithmetic.

double-precision number In digital computer operations, a number represented by two words for greater precision.

double pulse reading Pertaining to a magnetic core in a computer memory, recording bits as two states held simultaneously by one core having two areas that can be magnetized with alternate polarities. For example, positive-negative could represent zero, and negative-positive could represent one.

double-pulse wave An alternating-current (ac) wave having two successive positive peaks followed by two successive negative peaks within each cycle. The output voltage of a varistor bridge has such a waveshape for an ac input.

double-pulsing station A loran station that transmits at two pulse rates upon receiving two pairs of pulses.

double pumping A method of obtaining increased peak output power from a laser by pumping it for a comparatively long interval and then immediately pumping it for a short interval.

doubler **1.** A circuit or device for multiplying a frequency by two (see FREQUENCY DOUBLER). **2.** A circuit or device for multiplying a voltage by two (see VOLTAGE DOUBLER).

double probe A test probe that multiplies an applied signal voltage by two, so it can be handled more effectively by the instrument with which the probe is used.

double punching In perforating a punched card, putting two holes in one column; it is an error if it occurs in a field of a card that is part of a record.

double rail A form of logic system in which two lines are used, with three possible states. The output can be high, low, or undecided.

double response **1.** Two-point response, as that associated with tuning a receiver to a signal and then to its image. **2.** See DOUBLE-HUMP RESONANCE CURVE.

double screen A cathode-ray tube having a two-layer screen on which there is an additional, long-persistence coating of a different color.

double shield Two independent electromagnetic shields for a circuit enclosure or cable. The shielding structures are concentric, and can be connected together at a single point (the common point).

double sideband Abbreviation, DSB. In a modulated signal, the presence of both sidebands.

double-sideband suppressed carrier Abbreviation, DSSC. A transmission technique in which both sideband products of modulation are transmitted and the carrier is suppressed. Compare LOWER SIDEBAND, SUPPRESSED CARRIER, UPPER SIDEBAND, and SUPPRESSED CARRIER.

double-sideband system A modulation or demodulation system utilizing both sidebands, with or without the carrier.

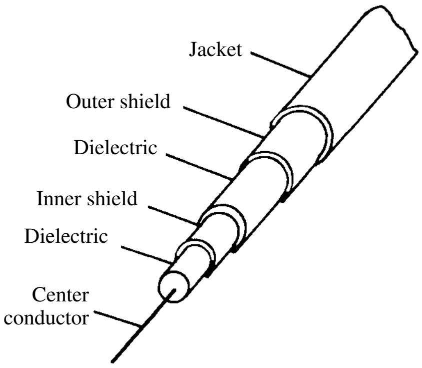

double shield

double-sideband transmitter A modulated transmitter using a double-sideband system.

double signal In reception, the property of having a signal on either side of the carrier frequency, as in a direct-conversion receiver. The two signals represent the sum and difference of the local oscillator signal and the actual received signal. Because the two signals convey identical information, the phenomenon is wasteful of received spectrum, and degrading to receiver selectivity and sensitivity.

double-signal receiver A receiver, such as a direct-conversion type, in which the signals occur in replicated form on either side of the local oscillator signal. Compare SINGLE SIGNAL.

double-spot tuning In a superheterodyne receiver, tuning in the same signal at two different places on the dial, a condition caused by image response.

double squirrel-cage induction motor A polyphase induction motor having a double squirrel-cage rotor. The rotor slots contain two bars, an upper bar having low reactance (being near the air gap) and high resistance, and a lower bar having high reactance and low resistance. This motor has low starting current, high starting torque, and a full-load slip of less than 5%.

double-stream amplifier A traveling-wave tube in which microwave amplification results from the interaction of two electron beams of different average velocity.

double-stub tuner Two stubs (see STUB) connected in parallel with a transmission line and usually spaced 0.375 wavelength (135 electrical degrees) apart; it is used as an impedance matcher.

double superheterodyne See DOUBLE-CONVERSION SUPERHETERODYNE.

double superheterodyne reception See DOUBLE-CONVERSION SUPERHETERODYNE.

double-surface transistor See COAXIAL TRANSISTOR.

doublet See DIPOLE ANTENNA.

double-throw Operating in opposite directions as selected (e.g., a double-throw relay or switch).

double-throw circuit breaker A circuit breaker that closes in both its pull-in and dropout positions.

double-throw switch or relay A switch or relay having two ganged poles.

double-trace recorder See DOUBLE-TRACK RECORDER, **2**.

double tracing Displaying two signals simultaneously on the screen of an oscilloscope through the use of an electronic switch.

double-track recorder **1.** A tape recorder whose head is positioned so that separate recordings can be made as two tracks on the tape. **2.** A graphic recorder that produces two separate parallel tracings.

doublet trigger A two-pulse, constant-spaced trigger signal used for coding.

double-tuned amplifier An amplifier whose input and output circuits are both tuned.

double-tuned circuit A circuit, such as an amplifier or filter, using separate input and output tuning. Also see DOUBLE CIRCUIT TUNING and DOUBLE-TUNED AMPLIFIER.

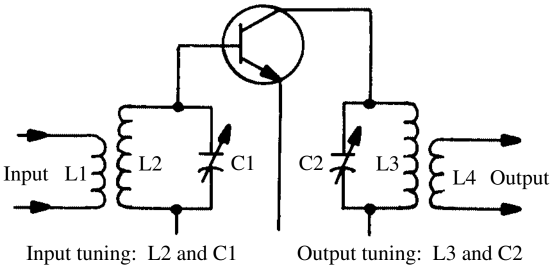

double-tuned circuit

double-tuned detector A form of frequency-modulation (FM) discriminator with two resonant circuits. One is tuned slightly higher than the channel center frequency, and the other is tuned an equal amount below the center.

double-vee antenna A broadband, modified dipole antenna resembling two vees in line. Also see VEE ANTENNA.

double-winding generator A dynamo-type generator having separate armature windings for supplying two voltages, either of which can be direct (dc voltage) or alternating (ac voltage).

double-wye rectifier A heavy-load circuit using six rectifier diodes, each conducting for 120 degrees of the cycle. An interphase winding is used. The circuit is equivalent to two three-phase, half-wave rectifiers connected in parallel.

double-Y rectifier See DOUBLE-WYE RECTIFIER.

double Zepp antenna A usually horizontal, straight, center-fed, full-wavelength antenna.

Also called *two half waves in phase.* Its name was derived because it is, in fact, two Zepp antennas forming a collinear array.

doubling **1.** Producing the second harmonic of a signal. **2.** In communication, unintentional simultaneous transmission by both operators, resulting in missed information. **3.** In a speaker, distortion resulting in large amounts of second-harmonic output.

doubly balanced modulator See BALANCED MODULATOR.

doughnut capacitor See DONUT CAPACITOR.

doughnut coil See TOROIDAL COIL.

doughnut crystal See DONUT CRYSTAL.

doughnut magnet See RING MAGNET.

down convert In superheterodyne conversion, to heterodyne a signal to an intermediate frequency lower than the signal frequency. Compare UP CONVERT.

down lead See LEAD-IN.

downlink The signal sent down from an active communications satellite to the earth, usually on a different frequency than the signal sent up. See UPLINK.

downlink beamwidth The angle subtended between the half-power points of the downlink signal from an active communications satellite.

downlink frequency The frequency of the downlink signal from an active communications satellite. Usually, the downlink signals occupy a certain band of frequencies, anywhere from several kilohertz to several megahertz wide.

downlink power **1.** The output power of the downlink transmitter in an active communications satellite. **2.** The effective radiated power (ERP) of the downlink signal from an active communications satellite.

down time A period of time during which electronic equipment is completely inoperative (for any reason).

downturn A usually sudden dip in a performance curve. Compare UPTURN.

downward modulation Modulation in which the average carrier component decreases during modulation. Example: amplitude modulation of a transmitter in which the antenna current decreases during modulation. Compare UPWARD MODULATION.

DP Abbreviation of DATA PROCESSING.

DPDT Abbreviation of *double-pole, double-throw* (switch or relay).

DPM **1.** Abbreviation of *digital power meter.* **2.** Abbreviation of DIGITAL PANEL METER. **3.** Abbreviation of *disintegrations per minute.*

DPS Abbreviaton of *disintegrations per second.*

DPST Abbreviation of *double-pole, single-throw* (switch or relay).

dr Abbreviation of *dram.*

drag **1.** A retarding force, caused by friction, acting on a moving body in contact with another moving or stationary body or medium. **2.** A retarding force introduced by an applied magnetic or electric field.

drag angle In disk recording, an angle of less than 90° between the stylus and the disk. The acute angle causes the stylus to drag instead of digging in.

drag cup A cup of nonmagnetic metal (usually copper or aluminum) that, when rotated in a magnetic field, acquires a voltage proportional to the speed of rotation. The device is often used as a brake.

drag-cup motor A servomotor whose shaft has a copper or aluminum drag cup that rotates in the field of a two-phase stator. Eddy currents set up in the cup by the field winding produce torque; braking action, direction control, and speed control are obtainable by means of associated electronics.

drag magnet In a motor-type meter, a braking magnet (i.e., one used to reduce speed through eddy-current effects). Also called *retarding magnet.*

drain **1.** The current or power drawn from a signal or power source. **2.** A load that absorbs current or power. **3.** The electrode in a field-effect transistor (FET) from which the output is usually taken; equivalent to the collector of a bipolar transistor.

drainage equipment Devices and systems for protecting circuits against transients generated by circuit breakers and similar safety devices.

drain-coupled multivibrator An oscillator that uses two field-effect transistors (FETs) in the circuit equivalent of a collector-coupled bipolar-transistor multivibrator. The drain of one stage is capacitance-coupled to the gate of the other stage.

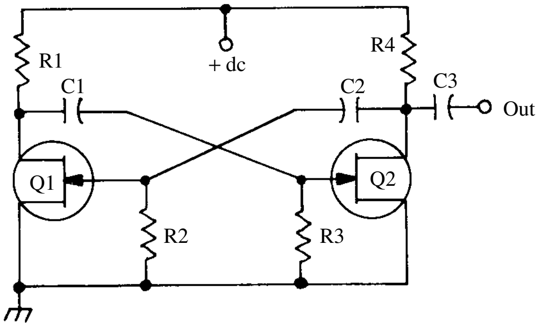

drain-coupled multivibrator

D-region A low region of the ionosphere beneath the E-region, whose ionization varies with the inclination of the sun. The greatest ionization is at midday; the layer disappears at night.

dress The (usually experimental) arrangement of leads for optimum circuit operation (minimum capacitance, best suppression of oscillation, minimum pickup, etc.).

dressed contact A contact having a permanently attached locking spring member.

drift **1.** Within a conductor or semiconductor, the controlled, directed movement of charge carriers resulting from an applied electric field. **2.** A usually gradual and undesirable change in a quantity, such as current, as a result of a disturbing factor, such as temperature or age.

drift current In a semiconductor, the current resulting from a flow of charge carriers in the presence of an electric field. The charge carriers are electrons in n-type material and holes in p-type material.

drift field The inherent internal electric field of a DRIFT-FIELD TRANSISTOR.

drift-field transistor An alloy-junction, bipolar, radio-frequency (RF) transistor for which the impurity concentration is graded from high on the emitter side of the base wafer to low on the collector side. This creates an internal drift field that accelerates current carriers and raises the upper frequency limit of the transistor.

drift-matched components Active or passive components that have been closely matched in terms of the drift of one or more parameters, with respect to time, temperature, etc.

drift mobility For current carriers in a semiconductor, the average drift velocity per unit electric field.

drift space **1.** In a vacuum tube, a space that is nearly free of alternating-current (ac) fields from the outside, and in which the repositioning of electrons is governed by the space-charge forces and the velocity distribution of the electrons. **2**. In a Klystron, the space between buncher and catcher cavities in which there is no field.

drift speed The average velocity of charge carriers moving through a medium.

drift transistor See DRIFT-FIELD TRANSISTOR.

drift velocity The net velocity of a charged particle (electron, hole, or ion) in the direction of the field applied to the conducting medium.

drift voltage The usually gradual change in voltage resulting from such causes as internal heating. Also called voltage drift.

drip loop In a transmission line for an antenna or power service, a loop near the point of entry to the building for the purpose of allowing condensation or rain water to drip off.

drip-proof motor A motor with ventilating apertures arranged so that moisture and particles cannot enter the machine.

drip-tight enclosure A housing designed to prevent entry of rain, snow, and dust; it also prevents accidental contact with the enclosed apparatus or machinery.

drive **1.** To excite (i.e., to supply with input-signal current, power, or voltage) (see DRIVING CURRENT, DRIVING POWER, and DRIVING VOLTAGE). **2.** Input-signal excitation (see DRIVING CURRENT, DRIVING POWER, and DRIVING VOLTAGE). **3.** A device that moves a recording medium (e.g., *tape drive* and *diskette drive*). **4.** The transmission of mechanical energy from one place to another (e.g., *motor drive*).

drive array A set of two or more hard-disk drives in a computer system. They function together to minimize the possibility of data loss. Such a system can also increase the amount of fast-access data storage.

drive belt A continuous belt used to transmit mechanical energy from a driving pulley to a driven pulley.

drive circuit **1.** A circuit used to provide the excitation to a motor. **2.** An amplifier that supplies drive to a more powerful amplifier.

drive control In a television receiver, the potentiometer used to adjust the ratio of horizontal pulse amplitude to the level of the linear portion of the sawtooth scanning-current wave.

driven element In a multielement antenna, an element to which electromagnetic energy is fed directly, as opposed to a PARASITIC ELEMENT, which is excited by a nearby radiator element.

driven-element directive antenna A multielement directional antenna whose elements are driven from the feed line (i.e., no element is parasitic). Compare PARASITIC-ELEMENT DIRECTIVE ANTENNA.

driven multivibrator A multivibrator whose operation or frequency is controlled by an external synchronizing or triggering voltage. Compare FREE-RUNNING MULTIVIBRATOR.

driven single sweep A single oscilloscope sweep that is initiated by the signal under observation.

drive pattern A pattern of interference in a facsimile system that is caused by improper synchronization of the recording spot.

driven sweep An oscilloscope sweep that is initiated by the signal under observation.

drive pin A pin used to prevent a record from slipping on the rotating turntable of a recorder or reproducer. It is similar to, and located near, the center pin of the turntable.

drive pulse In digital computer operations, a pulse that magnetizes a cell in a memory bank.

driver **1.** A device that supplies a useful amount of signal energy to another device to ensure its proper operation (e.g., a current driver for a magnetic-core memory, an oscillator driving a loudspeaker). **2.** A power amplifier stage that supplies signal power to a higher-powered amplifier stage. **3.** In a digital computer, a stage that increases the output current or power of another stage (e.g., a clock driver). **4.** The cone and magnet of a dynamic speaker.

driver element In a multielement directive antenna, the element excited directly by the feeder, the other elements (directors and reflectors) being parasitic.

driver impedance **1.** The output impedance of a driver stage. **2.** The impedance "seen" from the driven stage of an amplifier, through the driver transformer, to the driver stage. It is the vector sum of driver reactance and resistance.

driver inductance In an amplifier's driver transformer, the inductance, as "seen" looking through the transformer from the driven stage into the driver stage.

driver resistance In an amplifier's driver transformer, the resistance "seen" looking through the transformer from the driven stage into the driver stage.

driver stage An amplifier stage whose chief purpose is to supply excitation (input-signal current, power, or voltage) to the next stage. Also see DRIVER.

driver transformer The transformer that couples a driver stage to a driven stage. Example: the interstage transformer inserted between the collector of a single-ended driver transistor and the two bases of a push-pull power-output stage in an audio amplifier.

driving current In a power amplifier, the input signal current required to produce a given amount of output power.

driving-point admittance The reciprocal of DRIVING-POINT IMPEDANCE.

driving-point impedance The input impedance of a network.

driving power In a power amplifier, the input signal power required to produce a given amount of output power.

drive wire The wire forming the coil around the toroidal cell in a magnetic core memory; supplies pulses that magnetize the cell.

driving-range potential In cathodic protection, the difference of potential between the anode and (protected) cathode.

driving signal **1.** Drive (see DRIVE, **2**). **2.** In television, time-scanning signals (line-frequency pulses and field-frequency pulses) at the pickup location.

driving spring In a stepping relay, the spring that moves the wiper blades.

driving voltage In a power amplifier, the input signal voltage required to produce a given amount of output power.

DRO Abbreviation of DIGITAL READOUT.

drone A pilotless radio-controlled aircraft without a human pilot.

drone cone An undriven loudspeaker cone that is mounted in a bass-reflex enclosure with other speakers. Also called PASSIVE RADIATOR.

droop **1.** A dip in the graph of a function. **2.** In a pulse train, the decrease in mean amplitude (in percent of maximum amplitude) at a given instant after attainment of maximum amplitude.

drooping radials In a ground-plane antenna, radials that slope downward to provide a transmission-line impedance match. The slope angle depends on the characteristic impedance of the line; typically, the angle is between 45 degrees and 70 degrees, relative to horizontal.

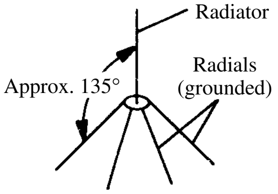

drooping radials

drop **1.** In wire communications, the line connecting a telephone cable to a subscriber's building. **2.** See VOLTAGE DROP.

drop bar A device that automatically grounds or short-circuits a capacitor when the door of a protective enclosure is opened.

drop cable See DISTRIBUTING CABLE, **1**.

drop channel In a communications system utilizing several channels, a channel that is not used.

drop-in The unintentional creation of bits when a magnetic storage device is being read from or written into. Compare DROP-OUT, **4**.

drop indicator In a signaling system, such as an annunciator, a hinged flap that drops into view when the signaling device is actuated.

drop-out **1.** The opening of a relay or circuit breaker. **2.** In digital computer operations, variation in signal level of the reproduced tape-recorded data. Such variation can result in errors in data reproduction. **3.** In the production of monolithic circuits, a special image placed at a desired point on the photomask. **4.** Digit loss during a read or write operation involving a magnetic storage device.

dropout current See DROPOUT VALUE.

dropout power See DROPOUT VALUE.

dropout value The level of current, power, or voltage at which a device, such as a circuit breaker or relay, is released.

dropout voltage See DROPOUT VALUE.

dropping resistor A series resistor providing a voltage reduction equal to the voltage drop across itself. For example, a 1000-ohm resistor in series with a 45-V battery, and carrying a current of 10 mA, will provide a voltage reduction equal to 10 V ($IR = 0.01 \times 1000 = 10$ V), thus dropping the 45 V to 35 V.

drop relay In a telephone system, a relay that is activated by the ringing signal. The relay is used to switch on a buzzer, light, or other device.

drop repeater A repeater intended for a termination of a communications circuit in a telephone system.

dropsonde A parachute-supported radiosonde dropped from a high-flying aircraft.

drop-tracks The tracks of radioactive particles made visible by moisture in an ionization chamber.

drop wire A wire that runs from a building to a pole (for line extension) or to a cable terminal (for cable extension).

drum **1.** A rotating cylinder coated with a magnetic material on which digital information can be recorded in the form of tiny magnetized spots. These spots are read as the drum rotates under pickup heads, or erased when the stored information is no longer needed. **2.** In some graphic recorders, facsimile receivers, etc., a rotating cylinder carrying the recording sheet.

drum capacitor See CONCENTRIC CAPACITOR.

drum controller The device that regulates the recording process on a drum memory.

drum mark On a track of a magnetic drum, a character that signifies the end of a character group.

drum memory In digital computers, a memory based on a magnetic drum (see DRUM, **1**). They have been largely replaced in recent years by electronic random-access memory, in the form of integrated circuits (ICs) and/or PCMCIA standard adapter cards.

drum parity The degree of accuracy in a drum recording/reproducing system.

drum programmer A device for sequencing operations. Its heart is a rotating drum, around whose surface contacts or points can be placed to actuate or terminate operations at selected times.

drum receiver A facsimile receiver using recording paper or photographic film wound around a revolving drum.

drum recorder A graphic recorder in which the record sheet is wound around a rotating drum.

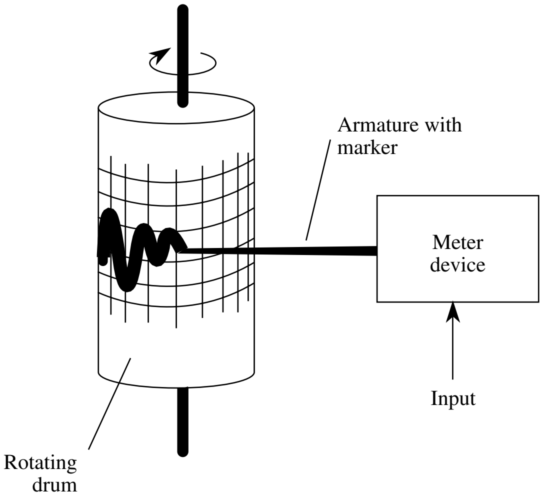

drum recorder

drum resistor A resistor consisting of a hollow cylinder of resistive material. Such a resistor can be cooled by circulating air or liquid through the cylinder.

drum speed The speed, usually measured in revolutions per minute (rpm), of the rotating drum in a graphic recorder, facsimile transmitter, or facsimile receiver.

drum storage The storage of data as magnetic impulses on a cylindrical, or drum-shaped, medium. Largely supplanted in recent years by magnetic disks, optical disks, and magnetic tapes.

drum switch A sequential switch whose contacts are pins or teeth placed at points around the outside of a revolving drum.

drum transmitter A facsimile transmitter in which the sheet bearing the material to be transmitted is wound around a revolving drum.

drum-type controller A motor-driven drum switch arranged to time various operations through sequential switching.

drum varistor A varistor that is a hollow cylinder of nonlinear resistance material. This varistor can be cooled by circulating air or liquid through it.

drum winding In a motor or generator, an armature whose conductors are on the outer face of the core, the two branches of a turn lying under adjacent poles of opposite polarity.

drunkometer An instrument for testing the extent of alcoholic intoxication. It electronically measures blood alcohol content through analysis of the subject's breath.

dry In an electric cell, a term used to describe an electrolyte that is semiliquid or solid.

dry battery A battery of dry cells.

dry cell **1.** A Leclanche primary cell in which the positive electrode is carbon, the negative electrode is zinc, and the electrolyte is a gel of ammonium chloride and additives. Also see CELL and PRIMARY CELL. **2.** A cell whose electrolyte is a gel or paste.

dry circuit A circuit in which the maximum voltage is 50 mV and the maximum current 200 mA.

dry-contact rectifier See DRY-DISK RECTIFIER.

dry contacts Contacts that neither make nor break a circuit.

dry-disk rectifier A solid-state rectifier, such as a copper-oxide, magnesium-copper-sulfide, or selenium type, that consists of a metal disk coated with a semiconductor material. The name was originally used to distinguish this rectifier from the *wet electrolytic rectifier.*

dry electrolytic capacitor An electrolytic capacitor whose electrolyte is a paste or solid. Compare WET ELECTROLYTIC CAPACITOR.

dry flashover voltage The breakdown voltage between electrodes in dry air when all insulation is clean and dry.

dry pile A voltaic pile containing numerous disks silvered or tinned on one face and covered with manganese dioxide on the other.

dry reed A metal contact, generally used as a relay or switch, that moves toward or away from another fixed contact under the influence of a magnetic field.

dry-reed relay See DRY-REED SWITCH.

dry-reed switch A switch consisting of two thin, metallic strips (reeds) hermetically sealed in a glass tube. The tube is surrounded by a coil of wire. When a current flows in the coil, a magnetic field affects the reeds. In the *normally open dry-reed switch*, the magnetic field causes the reeds to come together and close the circuit. In the *normally closed dry-reed switch*, the magnetic field causes the reeds to separate, opening the circuit. Compare MERCURY-WETTED REED RELAY.

dry run **1.** The preliminary operation of equipment for testing and appraisal. Such a procedure precedes putting the equipment into regular service. **2.** A step-by-step, paper-and-pencil "run" of a computer program before it is machine-implemented.

dry shelf life The life of a battery cell stored without its electrolyte.

dry-transfer process A method of transferring printed-circuit patterns and panel labels from sheets by rubbing them onto the substrate or panel.

dry-type forced-air-cooled transformer A DRY-TYPE TRANSFORMER that is cooled by convection of air circulated by a blower or fan. This increases the amount of power that the transformer can safely handle.

dry-type self-cooled transformer A DRY-TYPE TRANSFORMER that is cooled by natural air circulation (convection), without the use of a blower or fan.

dry-type transformer A transformer that, rather than being immersed in oil, is cooled entirely by the circulation of air.

DSB Abbreviation of DOUBLE SIDEBAND.

DSBSC Abbreviation of DOUBLE-SIDEBAND SUPPRESSED CARRIER. Also abbreviated DSSC.

dsc Abbreviation of *double silk covered* (wire).

D scope A radar whose display resembles that of a C scope, the difference being that blip height gives an approximation of the distance.

D service A Federal Aviation Agency service providing radio broadcasts of weather data, notices to aircraft personnel, and other advisory messages.

D-shell connector A multi-pin connector (either male or female) with a characteristic shape that ensures proper pin placement. Commonly used for computer ports, and for connections of peripherals to electronic equipment.

dsp Abbreviation of *double silver plated.*

DSR Abbreviation of DYNAMIC SPATIAL RECONSTRUCTOR.

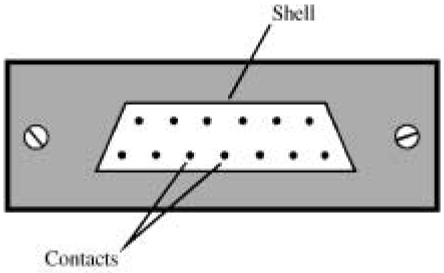

D-shell connector

DSS Abbreviation of *direct station selection* (telephone).

DSSC Abbreviation of DOUBLE-SIDEBAND SUPPRESSED CARRIER. Also abbreviated DSBSC.

DT Abbreviation of DATA TRANSMISSION.

DTA Abbreviation of *differential thermoanalysis.*

DT-cut crystal A piezoelectric plate cut from a quartz crystal at an angle of rotation about the z-axis of –53 degrees. It has a zero temperature coefficient of frequency at approximately 30 degrees Celsius. Also see CRYSTAL AXES and CRYSTAL CUTS.

DTL Abbreviation of DIODE-TRANSISTOR LOGIC.

DTn Abbreviation of DOUBLE TINNED.

DTS **1.** Abbreviation of DATA-TRANSMISSION SYSTEM. 2. Abbreviation of *digital telemetry system.*

DU Abbreviation of DUTY CYCLE.

dual **1.** Pertaining to a combination of two components such as diodes, transistors, etc., in a single housing. The components are often carefully matched. Compare QUAD. **2.** Pertaining to a device or circuit that behaves in a manner analogous to that of another operating with component and parameter counterparts. Thus, a current amplifier can be the dual of a voltage amplifier; a series-resonant circuit, the dual of a parallel-resonant circuit; or a field-effect transistor, the dual of a bipolar transistor.

dual-beam CRT A cathode-ray tube having two separate electron guns, for use in a dual-beam oscilloscope.

dual-beam oscilloscope Also called *dual-trace oscilloscope.* An oscilloscope having two electron guns and deflection systems; it can display two phenomena on the screen simultaneously for comparison.

dual capacitor **1.** Two fixed capacitors combined in a single housing, sometimes sharing a common capacitor plate. **2.** A two-section, ganged variable capacitor.

dual-channel amplifier An amplifier having two separate, independent channels (e.g., a stereo high-fidelity audio amplifier).

dual-cone speaker A speaker designed for a wide range of audio frequencies. One cone responds to the bass (low) and midrange audio frequencies, and a smaller cone responds to the treble (high) audio frequencies.

dual diode A discrete component consisting of two diodes in one package.

dual-diversity receiver A receiver or receiver system for DIVERSITY RECEPTION.

dual-diversity reception See DIVERSITY RECEPTION.

dual-emitter transistor A low-level silicon pnp chopper transistor of the planar passivated epitaxial type; it has two emitter electrodes.

dual-frequency calibrator A secondary frequency standard providing two fundamental test frequencies (e.g., 100 kHz and 1 MHz).

dual-frequency induction heater An induction heater whose work coils carry energy of two different frequencies. The coils heat the work either simultaneously or successively.

dual gate **1.** A digital integrated circuit (IC) consisting of two logic gate units. **2.** Pertaining to a field-effect transistor (FET) with two gates or gate electrode connections.

dual-gate FET A field-effect transistor with two gate (input) electrodes.

dual-gate MOSFET A metal-oxide-semiconductor field-effect transistor (MOSFET) with two gate (input) electrodes.

dual-inline package Abbreviation, DIP. A flat, molded integrated-circuit (IC) package having terminal pins along both long edges.

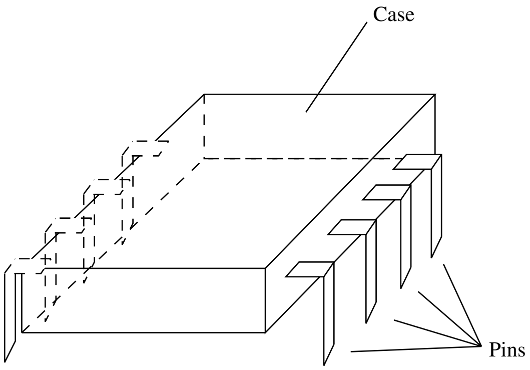

dual-inline package

duality **1.** The condition of being dual (see DUAL). This can be an aid in the design of certain circuits requiring complementary parameters, e.g., current-operated circuit analogs of voltage-operated circuits. **2.** See DUALITY OF NATURE.

duality of nature **1.** Any of various situations in which a phenomenon exhibits two distinct and different natures. A commonly cited example is the dual model of light. In some instances, visible light behaves like a barrage of particles, but in other environments it appears to be a wave effect. Another example is the dual model of electrostatic energy, behaving as point charges in some scenarios and as force fields in other situations. **2.** The tendency of a set of principles to be duplicated in sense by predictable analogies, as between inductance and capacitance, electrostatics and magnetics, etc.

dual local oscillator See DOUBLE LOCAL OSCILLATOR.

dual meter A meter having two meter movements and scales in a single case; the arrangement permits simultaneous monitoring of two quantities.

dual modulation The modulation of a single carrier or subcarrier by two different types of modulation—each carrying different information.

dual network A network that is the dual of another network having complementary parameters. For example, a common-emitter, current-sensitive, bipolar-transistor circuit is the dual of a common-source, voltage-sensitive, field-effect-transistor (FET) circuit. Also see DUALITY.

dual operation In digital logic, the operation resulting from inverting all of the digits. Every 1 is replaced with a 0, and vice versa.

dual-output power supply A power supply with two outputs. Often, one output is positive and the other is negative. In some cases, one output consists of alternating current (ac) and the other consists of direct current (dc).

dual pickup In disk reproduction, a pickup having two styli, one for large-groove records and one for fine-groove records.

dual potentiometer A ganged assembly of two potentiometers. The resistance values might or might not be the same.

dual preset counter A preset counter that will set alternately to two different numbers.

dual rail See DOUBLE RAIL.

dual resistor See DUAL POTENTIOMETER and DUAL RHEOSTAT.

dual rheostat A ganged assembly of two rheostats. The resistance values might or might not be the same.

dual stereo amplifier **1.** A two-channel audio amplifier for stereophonic audio applications. **2.** A two-channel linear integrated circuit (IC) for stereophonic audio applications.

dual-system loudspeaker See TWO-WAY SPEAKER.

dual trace In a cathode-ray oscilloscope, the use of two separate electron beams, which can show two different signals simultaneously on a single screen.

dual-trace recorder See DOUBLE-TRACK RECORDER, **2**.

dual-track recorder See DOUBLE-TRACK RECORDER.

dual recording In digital computer operations, updating two sets of master files simultaneously.

dual use The use of a communications system for two modes of data transfer at the same time.

Duant electrometer See BINANT ELECTROMETER.

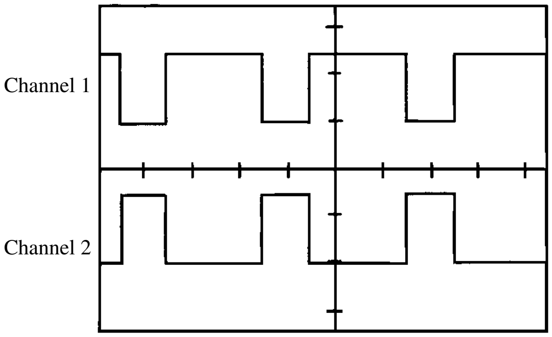

dual trace

duct **1.** A narrow propagation path, sometimes traveled by microwaves, created by unusual atmospheric conditions. **2.** A pipe or channel for cables and wires.

dubbing The adding of sound to a recorded magnetic tape, record disk, or film (e.g., replacing the sound track of a film in one language with that of another language).

dubnium Symbol, Db. Also clalled *unnilpentium* (Unp) and *hahnium* (Ha). Atomic number, 105. The most common isotope has atomic weight 262. Classified as a transition metal. It has a half-life on the order of a few seconds to a few tenths of a second (depending on the isotope), is human-made, and is not known to occur in nature.

duct **1.** A narrow propagation path, sometimes traveled by microwaves, created by unusual atmospheric conditions. **2.** A pipe or channel for cables and wires. **3.** An opening, vent, or other airway used for various purposes, such as cooling and acoustic wave transmission.

ducted port An opening in a speaker cabinet that has an airway (duct) extending several inches into the cabinet. It improves the quality of sound from a speaker system by equalizing the air pressure inside and outside the cabinet. Also provides resonant audio effects at frequencies that depend on the dimensions of the duct.

duct effect see TROPOSPHERIC DUCTING.

ductilimeter An instrument used for measuring the ductility of metals.

ducting The confinement of a radio wave to a duct (see DUCT, **1**) between two layers of the atmosphere or between an atmospheric layer and the earth.

Duddell arc A carbon copper arc circuit that produces audible continuous waves. Consists of a series inductance-capacitance (LC) circuit shunting an electric arc.

Duerdoth's multiple feedback system In an amplifier, feedback through several paths to improve response over that afforded by single-path feedback. In a simple application of multiple feedback, a single external loop is augmented with unbypassed emitter resistors in the amplifier stages.

Duerdoth's stability margin A feedback-amplifier stability margin equal to a 6-dB increase in gain at low and high frequencies over beta values between 0.3 and somewhat less than 2. For higher beta values, Duerdoth adopts an angular margin (for example, 15°); below $\beta = 0.3$, no danger of instability is present.

dummy **1.** A nonoperative model of a piece of equipment, usually assembled with dummy components (see DUMMY COMPONENT, **1**) for the purpose of developing a layout. **2.** DUMMY ANTENNA, DUMMY COMPONENT, or DUMMY LOAD. **3.** Part of a computer program that, rather than being useful for the problem at hand, only serves to satisfy some other format or logic requirement.

dummy antenna **1.** A nonradiating device that serves as a load for a transmitter (i.e., it takes the place of the regular antenna during tests and adjustments of the transmitter). **2.** A device containing a network of discrete inductive, capacitive, and resistive elements, inserted between a radio-frequency signal generator and receiver to simulate a standard antenna.

dummy component **1.** A nonoperative component used in developing a layout or package. **2.** A nonoperative component fraudulently included in a piece of equipment (e.g., an unwired transistor in a receiver circuit, a common occurrence during the early days of the transistor, when a 10-transistor radio brought more money than an 8-transistor radio, without regard to the circuit itself).

dummy instruction In a computer, a command that serves no operational purpose, other than to fill a format requirement.

dummy load **1.** A load device, usually consisting of resistance without reactance, used to terminate a power generator or power amplifier during adjustments and tests. The load resistance is equal to the output impedance of the generator or amplifier. **2.** See DUMMY ANTENNA.

dummy resistor A power-type resistor used as a dummy load.

dump **1.** In digital-computer operations, to transfer, completely or partially, the contents of memory into a peripheral. **2.** To switch off all power to a computer, deliberately or accidentally, thereby losing what is in the volatile memory.

dump and restart During a halt in a computer program run, to backtrack to the last dump point and use the data there to resume the run. Also see DUMP POINT.

dump check In digital-computer operations, the checking of all digits being transferred (see DUMP, **1**) to prevent errors when they are retransferred.

dumping To transfer the output at various stages in a computer program run to an external storage medium, so it will be available (in case of a failure) for the program's resumption from a point other than at the beginning.

dump point In writing a computer program, a point at which instructions are given to transfer data processed thus far to a storage medium that would be unaffected by a software or hardware failure. Also see DUMP AND RESTART and DUMPING.

dumping resistor **1.** See BLEEDER. **2.** A resistor having the minimum resistance permissible in a given situation. Used to discharge a capacitor, it acts to provide an alternative path to a potentially destructive short circuit.

duo Any pair of matched components, usually in a single package.

duodecal CRT base The 12-pin base of a cathode-ray tube. Also see BIDECAL, DIHEPTAL, and MAGNAL.

duodecal socket A 12-pin tube socket. Also see DUODECAL CRT BASE.

duodecimal **1.** Having 12 possibilities, states, choices, etc. **2.** Pertaining to the DUODECIMAL NUMBER SYSTEM. **3.** A number or numeral in the DUODECIMAL NUMBER SYSTEM.

duodecimal number system A system of numbering in which the radix, also called the base or modulus, is 12. The system uses the digits 0 through 9, plus two other characters (usually A and B) to represent 10 and 11. Thus, counting proceeds as 0, 1, 2, . . . , 9, A, B, 10, 11, 12, . . . , 19, 1A, 1B, 20, 21, 22, etc. At one time, some people seriously proposed that this system replace the DECIMAL NUMBER SYSTEM for general use.

duodiode See DUAL DIODE.

duolateral coil A multilayer, lattice-wound coil (see UNIVERSAL WINDING) in which the turns in successive layers are staggered slightly. Also called *honeycomb coil.*

duopole A two-pole all-pass device.

duplex **1.** A mode of communication in which two channels are used so that either operator in a conversation can interrupt the other at any time. **2.** The transmission of two messages over a single circuit, at the same time.

duplex artificial line In wire telephony, a balancing network that simulates the impedance of the actual line and the remote terminal equipment; it prevents an outgoing transmission from interfering with the local receiver.

duplex cable A cable consisting of a twisted pair of insulated stranded-wire conductors.

duplex channel A channel used for wire or radio DUPLEX OPERATION.

duplex communication See DUPLEX OPERATION.

duplex diode See DUAL DIODE.

duplexer In radar operations, a device operated by the transmitted pulse to automatically switch the antenna from the receiver to the transmitter.

duplexing assembly In a radar system, a device that automatically makes the receiver unresponsive to the outgoing transmitted signal while allowing incoming signals to reach the receiver easily. Also see TRANSMIT-RECEIVE SWITCH.

duplex computer system An installation of two computer systems, one standing by to take over in case the other fails.

duplex operation The simultaneous operation of a transmitter and receiver at a single location. This becomes possible (without mutual interference) through the use of two sufficiently separated carrier frequencies.

duplex system A system composed of two identical equipment sets—either of which will perform the intended function while the other stands by.

duplication check In digital-computer operations, the checking of an operation by doing it twice, using different methods in each case, to ensure the accuracy of results.

duplication house A professional person or company who makes high-quality copies of tape recordings (either audio or video). Charges vary, depending on the type and length of the recording.

duplicate To transfer data from one storage location to another. Compare DUMP, **1**.

dural See DURALUMIN.

duralumin An alloy of aluminum, copper, magnesium, manganese, and silicon. It offers strength with minimal weight.

duration control A potentiometer or variable capacitor for adjusting the duration of a pulse.

duration time The period during which a pulse is sustained (i.e., the interval between turn-on and turn-off time).

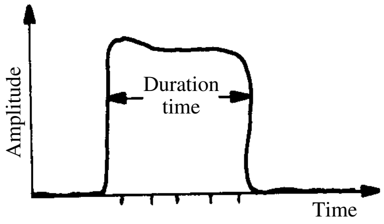

duration time

during cycle The interval during which a timer is in operation.

durometer An instrument for measuring the hardness of a material.

dust collector See DUST PRECIPITATOR.

dust core A magnetic core for radio-frequency coils consisting of very minute particles of iron or an alloy, such as Permalloy.

dust cover A removable, usually plastic cover for electronic and computer equipment, to protect hardware during periods of nonuse.

dust-ignition-proof motor A motor whose housing completely prevents the entry of dust, virtually eliminating the danger of fine dust sparking inside the machine.

dust precipitator An electrostatic device for removing dust, lint, and other particles from the air. It consists essentially of a pair of screens or wires through which the air passes; a potential of several thousand volts is maintained between them. The particles acquire a charge, then stick to the oppositely charged screen.

Dutch metal A copper-zinc alloy.

duty cycle **1.** The proportion or percentage of time during which a device, circuit, or system is operating or handling power. For example, when a radiotelegraph transmitter is keyed on and off to send Morse code, the duty cycle is approximately 50 percent; with frequency-shift keying, the duty cycle is 100 percent. **2.** The conditions under which an electrochemical cell or battery is used. In particular, the proportion or percentage of time during which current is drawn from the cell or battery for the purpose of operating an electrical or electronic circuit, device, or system.

duty cyclometer A direct-reading instrument for measuring duty cycle.

duty factor **1.** The ratio P_{avg}/P_{pk}, where P_{avg} is the average power in a system and P_{pk} is the peak power. **2.** The product of the duration and the repetition rate of regularly recurring pulses comprising a carrier.

duty ratio See DUTY FACTOR, 1.

DVM Abbreviation of DIGITAL VOLTMETER.

DVOM Abbreviation of DIGITAL VOLT-OHM-MILLIAMMETER.

dwell meter An instrument that shows the period (or angle) during which contacts remain closed.

dwell switching Switching action in which the contacts are held closed (or a circuit kept on) for specified periods, as opposed to MOMENTARY SWITCHING.

dwell tachometer A combination dwell meter/tachometer for automobile engine testing and adjustment. The dwell meter allows observation and adjustment of the ignition point cam angle; the tachometer shows the motor speed in revolutions per minute (rpm).

DX **1.** Radiotelegraph abbreviation meaning *long distance* or *foreign country.* **2.** A communication or broadcast station located far away and/or in a foreign country. **3.** Abbreviation of DUPLEX.

DXer An amateur radio operator who prefers to communicate with stations far away and/or in foreign countries.

Dy Symbol for DYSPROSIUM.

dyadic operation A binary operation (i.e., one using two operands).

dyn Abbreviation of DYNE.

dyna- A prefix (combined form) meaning *power* (e.g., dynamometer and dynatron).

dynamic acceleration Acceleration whose magnitude and direction are constantly changing.

dynamic allocation In multiprogramming, a system in which a monitor program assigns peripherals and areas of memory to a program.

dynamic analogy A mathematical similarity between or among various phenomena involving the motion of particles.

dynamic base current See AC BASE CURRENT.

dynamic base resistance See AC BASE RESISTANCE.

dynamic base voltage See AC BASE VOLTAGE.

dynamic behavior **1.** The behavior of a component, device, or system when signals are applied, as opposed to *static behavior* under no-signal conditions. **2.** The behavior of a device or system involving the motion of particles over a period of time.

dynamic braking A technique for stopping a motor quickly using a resistor (the dynamic braking resistance) connected across the spinning armature. The resistor dissipates the energy generated by the motor, producing a damping action that results in braking.

dynamic characteristic The performance characteristic of a device or circuit under alternating-current (ac) operating conditions, as opposed to the *static characteristic*, when only direct current (dc) flows.

dynamic check **1.** A test made under actual operating conditions of a device or circuit. **2.** A test made with an alternating-current (ac) applied signal, rather than with direct-current (dc) quantities.

dynamic collector current See AC COLLECTOR CURRENT.

dynamic collector resistance See AC COLLECTOR RESISTANCE.

dynamic collector voltage See AC COLLECTOR VOLTAGE.

dynamic contact resistance In relay or switch contacts, variation in the electrical resistance of the closed contacts because of variations in contact pressure.

dynamic convergence In a color picture tube, the meeting of the three beams at the aperture mask during scanning.

dynamic current amplification Abbreviation, DCA. An expression of gain in a bipolar transistor. Specified as the ratio of the difference in collector current I_C to the difference in base current I_B

$$\mathrm{DCA} = dI_C/dI_B$$

Geometrically, the dynamic current amplification at a given bias point is the slope of a line tangent to the I_C-versus-I_B curve at that point.

dynamic curve A characteristic curve that accounts for the presence of resistance in series with the device to which the curve applies.

dynamic debugging Any debugging operation performed on a computer system during a normal-speed program run.

dynamic decay Decay resulting from such factors as ion charging in a storage tube.

dynamic demonstrator A teaching aid consisting of a board displaying an electronic circuit, behind which is mounted the actual circuit. Various circuit components (especially adjustable ones) are mounted on the front of the board, in clear view at places where their circuit symbols appear. Pin jacks at important test points in the circuit allow connection of a meter, signal generator, and oscilloscope leads for testing or demonstrating the circuit.

dynamic deviation The difference between ideal output and actual output of a circuit or device operating with a reference input that changes at a constant rate and is free of transients.

dynamic diode tester An instrument that displays the response curve (or family of curves) of a diode on a calibrated oscilloscope screen. The horizontal axis of the screen indicates voltage, the vertical axis shows current, and zeros for both quantities are at center screen. Also see DYNAMIC RECTIFIER TESTER.

dynamic drain current See AC DRAIN CURRENT.

dynamic drain resistance See AC DRAIN RESISTANCE.

dynamic drain voltage See AC DRAIN VOLTAGE.

dynamic dump A dump that occurs during a program run. See DUMPING.

dynamic electric field An electric field whose intensity is constantly changing, either periodically or in a complex way.

dynamic emitter current See AC EMITTER CURRENT.

dynamic emitter resistance See AC EMITTER RESISTANCE.

dynamic emitter voltage See AC EMITTER VOLTAGE.

dynamic equilibrium **1.** The state of balance between constantly varying quantities. **2.** The tendency of two current-carrying circuits to maintain at a maximum the magnetic flux linking them.

dynamic error In a periodic signal delivered by a transducer, an error resulting from the restricted dynamic response of the device.

dynamic flip-flop A flip-flop (bistable multivibrator) that is kept on by recirculating an alternating-current (ac) signal. The device can be switched on or off by a single pulse. Compare STATIC FLIP-FLOP.

dynamic focus Compensation for defocusing, caused by the electron beam sweeping in an arc across a flat color picture-tube screen; the method uses an alternating-current (ac) focusing-electrode voltage.

dynamic gate voltage See AC GATE VOLTAGE.

dynamic impedance The impedance of a device (such as a transistor or diode) when it is operating with an applied alternating-current (ac) signal, as opposed to its *static resistance* with only direct current (dc) applied.

dynamic limiter A limiter, such as is used in frequency-modulation (FM) receivers, that maintains the output-signal level, despite appreciable excursions of input-signal amplitude.

dynamic loudspeaker See DYNAMIC SPEAKER.

dynamic magnetic field A magnetic field whose intensity is constantly changing, either periodically or in a complex way.

dynamic memory A usually random-access data storage method in which the memory cells must be electrically refreshed periodically to avoid the loss of held data.

dynamic microphone A microphone in which a small coil attached to a vibrating diaphragm or cone moves in a uniform magnetic field to generate the output signal.

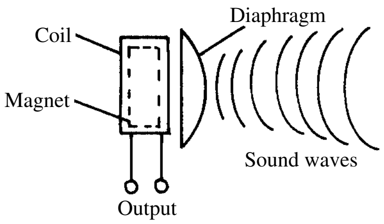

dynamic microphone

dynamic mutual conductance See DYNAMIC TRANSCONDUCTANCE.

dynamic noise suppressor A noise limiter consisting of an audio filter whose bandwidth is directly proportional to signal strength (i.e., it is varied automatically by signal amplitude).

dynamic operating line A curve displaying the control function of a device. For example, the collector-current-versus-base-current curve of a bipolar transistor is drawn between the limits of saturation and cutoff.

dynamic output impedance The output impedance of a power supply, as "seen" by the load.

dynamic pickup A phonograph pickup whose stylus causes a small coil to vibrate in the field of a permanent magnet. Works on the same principle as the DYNAMIC MICROPHONE.

dynamic printout A printout that occurs as a single function, actuated by one command, and completing itself in one operation.

dynamic problem checking A method of checking the solution obtained by an analog computer, to see that it makes sense (is not absurd).

dynamic programming A method of problem solving in which continual checks are made to ensure accuracy or conformance to a certain set of rules.

dynamic range **1.** In high-fidelity audio, the ratio of the loudest sound to the faintest sound that can be reproduced without significant distortion or noise. It is usually expressed in decibels. **2.** In a communications receiver, a measure of the ability to receive both weak and strong signals without excessive noise, distortion, desensitization or other undesirable effects. It is expressed in various ways, typically in decibels. **3.** The ratio between the loudest and faintest sounds, or between the strongest and weakest signals, encountered in a given environment or situation. It is usually expressed in decibels.

dynamic rectifier tester An instrument that displays the response curve of a rectifier on a calibrated oscilloscope screen. During the test, the rectifier receives an alternating-current (ac) voltage with a low positive peak and high negative peak, both corresponding to the rated forward and reverse voltages (respectively) of the rectifier. The horizontal axis of the screen indicates voltage, the vertical axis indicates current, and zeros for both quantities are at center screen.

dynamic regulation In an automatically regulated system, such as a voltage-regulated power supply, the transient response of the system. Dynamic regulation is determined from maximum overshoot and recovery time when the load or line value is suddenly changed.

dynamic regulator A circuit or device providing dynamic regulation.

dynamic reproducer **1.** See DYNAMIC MICROPHONE. **2.** See DYNAMIC PICKUP. **3.** See DYNAMIC SPEAKER.

dynamic resistance See DYNAMIC IMPEDANCE.

dynamic run See DYNAMIC CHECK, **1**. See also DYNAMIC DEBUGGING.

dynamics The study of bodies, charges, fields, forces, or pulses in motion. Compare STATICS.

dynamic sequential control In digital computer operation, the computer's changing the sequence of instructions during a run.

dynamic source current See AC SOURCE CURRENT.

dynamic source resistance See AC SOURCE RESISTANCE.

dynamic source voltage See AC SOURCE VOLTAGE.

dynamic spatial reconstructor Abbreviation, DSR. An advanced x-ray machine, developed at the Mayo Clinic, that displays organs in three-dimensional views in motion, and allows them to be electronically dissected without actually operating on the patient.

dynamic speaker A loudspeaker in which a small coil (voice coil), attached to a diaphragm or cone and carrying an audio-frequency signal current, moves back and forth in a permanent magnetic field and, accordingly, causes the diaphragm or cone to vibrate (emit sound). Compare MAGNETIC SPEAKER.

dynamic stability A measure of the ability of a robot to maintain its balance while in motion.

dynamic stop As caused by a computer program instruction, a loop indicating the presence of an error.

dynamic storage See DYNAMIC MEMORY.

dynamic subroutine A form of computer subroutine that allows the derivation of other subroutines in various forms.

dynamic test See DYNAMIC CHECK.

dynamic transconductance Transconductance determined from alternating-current (ac) signal parameters, rather than from direct-current (dc) parameters.

dynamic transducer A coil-and-magnet device that converts mechanical vibration into electric currents, or vice versa. Common examples include most microphones, headphones, and loudspeakers.

dynamic transfer characteristic An input-output characteristic determined, with respect to the load of a transfer device. Also see DYNAMIC CHARACTERISTIC.

dynamic transistor tester **1.** An instrument for checking the alternating-current (ac) gain of a transistor, rather than its direct-current (dc) beta. **2.** An instrument for determining the condition of a transistor from its performance in a simple oscillator circuit. **3.** An instrument that displays a transistor response curve, or a family of such curves, on a calibrated oscilloscope screen. Also see DYNAMIC DIODE TESTER and DYNAMIC RECTIFIER TESTER.

dynamo A mechanical generator of electricity; typically a rotating machine.

dynamoelectric machinery Rotating electric machinery. Examples: amplidynes, generators, dynamotors, rotary converters.

dynamometer **1.** See ELECTRODYNAMOMETER. **2.** A device for mechanically measuring the output power of a motor.

dynamometer ammeter See ELECTRODYNAMOMETER.

dynamometer voltmeter See ELECTRODYNAMOMETER.

dynamophone A dynamometer (see DYNAMOMETER, **2**) that uses two telephone circuits to measure the twist of a shaft.

dynamostatic machine A machine driven by alternating-current (ac) or direct-current (dc) power for the generation of static electricity.

dynamotor A (usually small) self-contained motor-generator. The motor and generator portions are enclosed in a common housing, giving the machine the appearance of a simple motor.

dynaquad A pnpn four-layer semiconductor device with three terminals, similar to the silicon-controlled rectifier or thyristor.

dynatron A form of vacuum tube that displays a negative-resistance characteristic, resulting in oscillation at ultra-high and microwave frequencies.

dynatron frequency meter A heterodyne-type frequency meter using a dynatron oscillator.

dyne Abbreviation, d. A unit of force. One dyne (10^{-5} newton) is the force that will give a mass of 1 gram an acceleration of 1 centimeter per second per second. Compare NEWTON.

dyne-centimeter See ERG.

dyne-five In the Giorgi mks system, a unit of force equal to 1 newton.

dyne per square centimeter Abbreviation, d/cm^2. A unit of pressure equal to 0.1 pascal (9.869 × 10^{-7} atmosphere).

dyne-seven A unit of force equal to 10^7 dynes.

dynistor A semiconductor diode that continues to conduct after the forward voltage is reduced below the normal threshold point. To stop the conduction, a reverse voltage must be applied, or voltage must be entirely removed from the device. It is used in switching applications.

dynode In a photomultiplier tube, any of several slanting electrodes that receives a beam of electrons generated by the light-sensitive cathode and reflects it, along with secondary electrons. This amplifies the beam; the process is repeated several times. Thus, the emission from the cathode is greatly amplified when it reaches the plate.

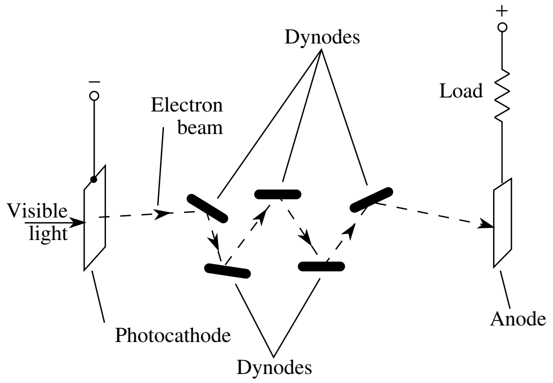

dynode

dysprosium Symbol, Dy. An element of the rare-earth group. Atomic number, 66. Atomic weight, 162.50. Dysprosium is a highly magnetic substance.

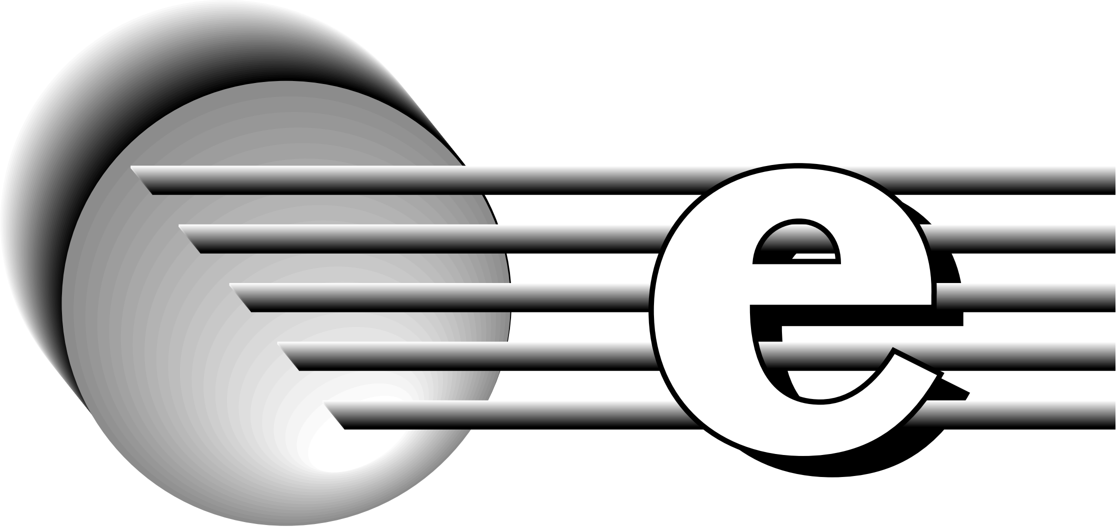

E **1.** Symbol for VOLTAGE. **2.** Symbol for ELECTRIC FIELD STRENGTH. **3.** Abbreviation of EMITTER. **4.** Symbol for prefix EXA. **5.** Symbol for ENERGY.

e **1.** Symbol for VOLTAGE. **2.** Abbreviation of EMITTER. **3.** Symbol for ELECTRON CHARGE. **4.** Symbol for the natural logarithm base (approximately equal to 2.71828). **5.** Symbol for ECCENTRICITY. **6.** Abbreviation of ERG.

EAM Abbreviation of *electronic accounting machine.*

E and M terminals The output and input leads in some signaling systems. Also called *E and M leads.*

early early sound Sound propagated through solids and/or liquids that reaches a pickup device (such as a microphone) before the sound propagated through the air. In general, sound waves travel faster as the medium becomes more dense.

early-failure period The period immediately after manufacture of a device, during which the failure rate (caused by defects in equipment or workmanship) is high.

early-warning radar Abbreviation, EWR. A radar system that produces immediate warning when enemy aircraft enter the monitored area.

earphone **1.** *Headphone* (usually a single unit). **2.** Telephone receiver. **3.** A miniature acoustic transducer that is small enough to be inserted into the ear.

earpiece See EARPHONE, **3**.

earth **1.** The ground. **2.** An electrical connection to the earth (see GROUND CONNECTION, **2**). **3.** In space communications, the planet Earth.

earth connection See GROUND CONNECTION, **2**.

earth currents **1.** Electric currents induced in the earth by current flowing through underground or underwater cables. **2.** Electric currents flowing through the earth between ground connections of electrical equipment.

earth ground **1.** A common connection to an electrode buried in the earth so that good conductivity is maintained between the common circuit point and the earth itself. **2.** A rod driven into the surface of the earth for use as a common circuit connection.

earth inductor A magnetometer consisting of a coil that is rotated in the earth's magnetic field. It delivers an alternating-current (ac) voltage proportional to the field strength. Also called *generating magnetometer.*

earth-moon-earth See MOONBOUNCE.

earth resonance A resonant effect at extremely low frequencies, caused by reflection of currents within the earth. Resonant currents have been tested as a means of communicating with submarines worldwide.

earth's magnetic field Also called *geomagnetic field.* The natural magnetic field whose lines of flux extend from north to south. The earth's magnetic poles, also called the *geomagnetic poles*, do not exactly coincide with the geographic poles. The field somewhat resembles that of a bar magnet.

Eastern Standard Time Abbreviation, EST. Local mean time at the 75th meridian west of Greenwich.

east-west effect The phenomenon in which the number of cosmic rays approaching earth near the equator from the west is greater than that from the east by 10 percent.

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

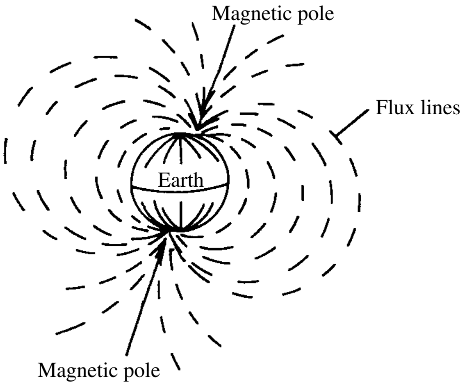

earth's magnetic field

E_{avg} Symbol for AVERAGE VOLTAGE.

E_b Symbol for BATTERY VOLTAGE.

E bend In a waveguide, a smooth change in the direction of the axis, which remains parallel to the direction of polarization.

EBI Abbreviation of EQUIVALENT BACKGROUND INPUT.

ebiconductivity Conductivity resulting from electron bombardment.

ebonite Hard rubber used as an insulant. Dielectric constant: 2.8. Dielectric strength: 30 to 110 kV/mm.

EBR Abbreviation of ELECTRON-BEAM RECORDING.

EBS Abbreviation of ELECTRON-BOMBARDED SEMICONDUCTOR.

EBS amplifier An amplifying device using an electron-bombarded semiconductor. The electron beam is modulated by the input signal, and the modulated resistance of the semiconductor target modulates a relatively heavy current to provide an amplified output. Current gains on the order of 2000 are possible.

ec Abbreviation of ENAMEL-COVERED (in reference to wire).

eccentric circle See ECCENTRIC GROOVE.

eccentric groove On a phonograph record, an off-center groove in which the stylus rides at the end of the recording, where it causes the tone arm to trip the record-changing mechanism.

eccentricity **1.** The condition of being off center, intentionally or not. It is often a consideration in the behavior of dials, potentiometers, and servomechanisms. **2.** On a phonograph record, the condition in which the spiral recording groove and the center hole of the disk are not concentric. **3.** A quantitative expression for the extent to which an ellipse is elongated.

eccentricity of orbit The extent to which the orbit or path of a satellite differs from a circle. A circular orbit has eccentricity zero. As the orbit becomes elliptical and deviates more and more from a perfect circle, the eccentricity increases. When the eccentricity reaches 1, the object takes a parabolic path through space. If the eccentricity exceeds 1, the path is a hyperbola.

ECCM Abbreviation of ELECTRONIC COUNTER-COUNTERMEASURES.

ECDC Abbreviation of *electrochemical diffused collector.*

ECG Abbreviation of electrocardiogram. (Also, EKG.)

ECG telemetry Use of a radio telemetering system to monitor the heart function of a person from a distance.

echelon **1.** A level of calibration accuracy, the highest echelon being the national standard for the particular measurement involved. **2.** A level of maintenance in which lower ordinal numbers refer to less-critical tasks, and higher ordinal numbers refer to tasks requiring progressively higher skills and technological expertise.

echelon grating A diffraction grating with extremely high resolution. Generally useful only over a small range of wavelengths.

echo **1.** A signal that is reflected back to the point of origin. **2.** A reflected or delayed signal component that arrives at a given point behind the main component. **3.** A radar blip, indicating an object or thundershower. **4.** Reflection of the signal on a telephone line, caused by improper impedance matching, or by overload of the system by too many subscribers attempting to use the system at the same time. **5.** In audio systems, a circuit that causes sounds to repeat one or more times, at intervals ranging from a fraction of a second to several seconds. **6.** The effect produced by a circuit, as defined in 5.

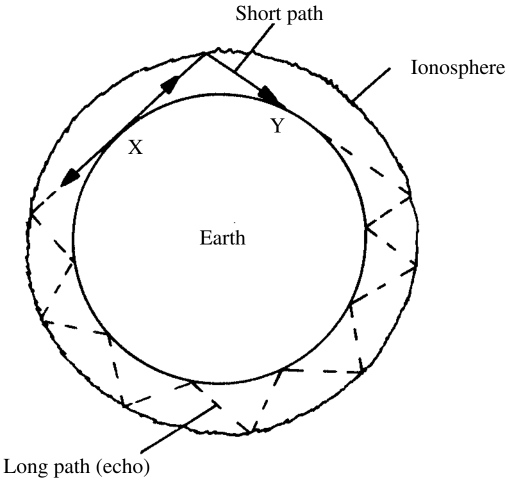

echo

echo area The area of a target that will return a radar signal as an echo.

echo attenuation In a bidirectional wire-communication circuit equipped with repeaters or multiplexers, the attenuation of echo currents set up by conventional operation.

echo box A resonant-cavity device used to test a radar set. Part of the transmitted energy enters the box, which retransmits it to the receiver. The signal reaching the receiver is a slowly decaying transient whose intensity eventually falls below the level that can be displayed on the screen; the time required to reach this level indicates radar performance.

echo chamber A reverberation chamber, electronic recording device, or room for acoustic tests or for simulating sonic delays.

echo check In data communication, a means of checking the accuracy of received data by sending it back to the transmitting station for comparison with the original data.

echo depth sounder See ACOUSTIC DEPTH SOUNDER.

echo eliminator **1.** A device that quiets a navigational instrument after receipt of a pulse, to prevent reception of a subsequent, delayed pulse. **2.** In a two-way telephone circuit, a voice-operated device that suppresses echo currents caused by conversation currents going in the opposite direction.

echoencephalograph An ultrasonic medical instrument that allows viewing of internal organs. Used for diagnostic purposes in certain situations, instead of the X-ray machine.

echogram In acoustics, a graph of the sound decrement in an enclosure. Time is plotted on the horizontal axis; signal amplitude is plotted on the vertical axis. An intense pulse is transmitted from a speaker; a microphone picks up the echoes and sends them to a pen recorder or microprocessor.

echo intensifier A device used at a radar target to boost the intensity of reflected energy.

echo interference Radio interference resulting from a reflected signal arriving slightly later than the direct signal.

echo matching In an echo-splitting radar system, the trial-and-error orientation of the antenna to find the direction from which the pulse indications are identical.

echo ranging An ultrasonic method of determining the bearing and distance of an underwater object.

echo send In an audio mixer, an output for delivering signals to external systems, such as an echo box (see ECHO, **5**). It can also provide an auxiliary output for a second set of speakers, a tape recorder, etc.

echo sounder See ACOUSTIC DEPTH SOUNDER.

echo splitting Separating a radar echo into two parts so that a double indication appears on the radar screen.

echo suppression In a telephone circuit, a device that chokes off reflected waves, thereby minimizing audible echo.

echo suppressor See ECHO ELIMINATOR.

echo talk Echo in a telephone system that results in distracting interference.

echo wave A reflected wave, such as a radio wave reflected alternately between earth's surface and the ionosphere.

ECL Abbreviation of EMITTER-COUPLED LOGIC.

eclipse effect A decrease in the critical frequency of the E and F1 layers of the ionosphere during a solar eclipse.

ecliptic orbit Any orbit that lies in the same plane as the orbit of the earth around the sun (the *ecliptic plane*). The ecliptic plane is slanted about 23.5° from the plane of the earth's equator.

ECM Abbreviation of ELECTRONIC COUNTERMEASURES.

ECO Abbreviation of ELECTRON-COUPLED OSCILLATOR.

econometer An instrument for continuously monitoring the amount of carbon dioxide in (factory) flue gases.

E core A transformer or transducer core having the shape of an E. Coils can be wound on one, two, or all three of the crosspieces.

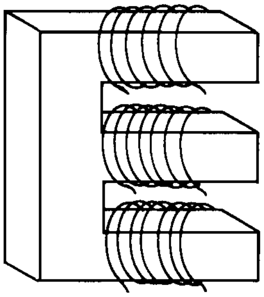

E core

ECPD Abbreviation of *Engineer's Council for Professional Development.*

ECTL Abbreviation of *emitter-coupled transistor logic.*

EDD Abbreviation of ENVELOPE-DELAY DISTORTION.

eddy current A circulating current induced in a conducting material by a varying magnetic field, often parasitic in nature. Such a current can, for example, flow in the iron core of a transformer.

eddy-current device A brake, coupling, clutch, drag cup, drive unit, or similar device whose operation is based on the generation of torque, pull, or opposition by the action of eddy currents.

eddy-current heating Heating caused by eddy-current loss in a material.

eddy-current loss Power loss resulting from eddy currents induced in nearby structures by an electromagnetic field. Eddy currents in the core of a transformer give rise to such loss.

edge connector A terminal block with a number of contacts, attached to the edge of a printed-circuit board for easy plugging into a foundation circuit.

edge control In the manufacture of paper, a robotic system for maintaining the width of a sheet by sensing the edges and correcting the machine accordingly. Transducers that sense the passing edges deliver output signals proportional to variations from standard width.

edge detection The ability of a machine vision system to locate and follow boundaries. Used extensively in mobile robots.

edge effect The extension of electric lines of flux between the outer edges of capacitor plates. This portion of the interplate field contributes a small amount of capacitance. Because the lines of flux are not confined to the space between plates, they can cause capacitive coupling with external bodies.

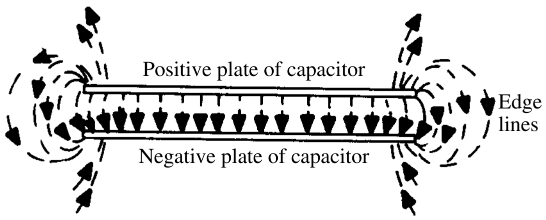

edge effect

edge-punched card In digital computer operations, a punched card whose edge is perforated in a narrow column, the center being used for written annotation.

edgewise meter A meter having a curved horizontal scale; this arrangement allows mounting the instrument edgewise in a panel.

edgewise-wound coil A coil made of a flat metal strip cut in the shape of a coil spring. The design allows the use of clips to vary the inductance, but this advantage is often offset by the coil's high distributed capacitance.

edging In a color television picture, extraneous color of a different hue than the objects around whose edges it appears.

Edison base A threaded base on light bulbs, cone-type heaters, and some pilot lamps.

Edison battery A group of Edison cells connected in series, parallel, or both, and contained in a single package with two electrodes.

Edison cell A secondary (storage) cell in which the active positive plate material consists of nickel hydroxide held in steel tubes assembled into a steel grid; the active negative plate material is powdered iron oxide mixed with cadmium; the electrolyte is potassium hydroxide. The open-circuit voltage of the cell is typically 1.2 V at full charge.

Edison distribution system A three-wire, 110- to 250-volt direct-current (dc) power-distribution system.

Edison effect Thermionic emission of negatively charged particles (electrons) from a hot filament sealed in an evacuated bulb; they are attracted by a cold, positively charged metal plate in the bulb.

E display A radar display in which the horizontal coordinate indicates distance, and the vertical coordinate indicates elevation.

edit **1.** In tape recording, the modifying of the recorded material by deleting (cutting out or erasing), adding (splicing or overrecording), or changing the sequence of the material by physically or magnetically altering the tape. **2.** In digital computer operations, to make data ready for processing.

edit decision list In the editing of a digital audio/video presentation, a record of every change (cut and paste). The list is automatically made by the computer and stored on disk for later reference if needed.

editing **1.** Alteration of a magnetic-tape recording by means of splicing. **2.** Alteration of data stored in memory, either by adding information, removing information, changing information, or (usually) a combination of these operations.

EDL Abbreviation of EDIT DECISION LIST.

EDM Abbreviation of ELECTRICAL-DISCHARGE MACHINING.

EDP Abbreviation of *electronic data processing.*

EDPC Abbreviation of ELECTRONIC DATA-PROCESSING CENTER.

EDPM Abbreviation of ELECTRONIC DATA-PROCESSING MACHINE.

EDPS Abbreviation of ELECTRONIC DATA-PROCESSING SYSTEM.

EDT **1.** Abbreviation of *ethylene diamine tartrate* (a synthetic piezoelectric material). **2.** Abbreviation of *Eastern Daylight Time.*

EDU Abbreviation of *electronic display unit.*

educational robot A robot that can be programmed for the purpose of teaching its users something. Popular as an educational toy for children.

EDVAC Acronym for *Electronic Discrete Variable Automatic Computer*, a development of the University of Pennsylvania.

EE Abbreviation of ELECTRICAL ENGINEER or ELECTRONICS ENGINEER.

E_{eff} Symbol for EFFECTIVE VOLTAGE.

EEG Abbreviation of ELECTROENCEPHALOGRAM.

EEPROM Abbreviation of ELECTRICALLY ERASABLE PROGRAMMABLE READ-ONLY MEMORY. See PROM and ROM.

EES Abbreviation of EARLY EARLY SOUND.

effective acoustic center The apparent point of propagation of spherically divergent sound waves radiated by an acoustic generator.

effective actuation time The total actuation time of a relay (i.e., the sum of the initial actuation time and subsequent intervals of contact chatter).

effective address The address a computer uses in implementing an instruction (i.e., one not necessarily coinciding with the address given in the instruction).

effective ampere An effective current of 1 ampere. Also see EFFECTIVE CURRENT.

effective antenna length See ELECTRICAL LENGTH.

effective antenna resistance The radiation resistance of an antenna, as measured at the input point.

effective bandwidth The bandwidth of an ideal bandpass filter, which, at a reference frequency, has the same transfer ratio as an actual bandpass filter under consideration; it also has the same current and voltage characteristics.

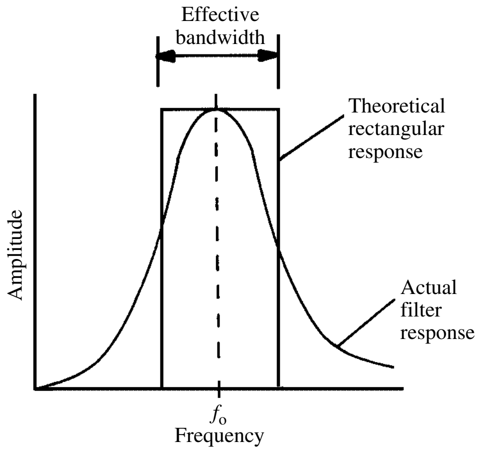

effective bandwidth

effective capacitance The actual capacitance between two points in a circuit resulting from the combination of inherent, lumped, and stray capacitances.

effective conductivity Conductivity measured between the parallel faces of a unit cube of a material.

effective confusion area In a radar system, an area in which interference makes it impossible to see whether a target is present.

effective current Symbol, I_{eff}. The root-mean-square (rms) value of alternating current (see EFFECTIVE VALUE). For a sinusoidal current, $I_{eff} = 0.707 I_{max}$, where I_{max} is the maximum value of the current. Also called *rms current.*

effective cutoff See EFFECTIVE CUTOFF FREQUENCY.

effective cutoff frequency For a filter or similar device operated between specified impedances, the frequency at which insertion loss is higher than the loss at a specified reference frequency in the passband.

effective field intensity The root-mean-square (rms) value of the field-strength voltage, averaged for all points at a horizontal distance of one mile from a transmitting antenna.

effective height The height of an antenna in terms of its performance as a transmitter or receiver of electromagnetic energy.

effective input capacitance The actual operative capacitance present at the input terminals of a circuit or device, caused by the shunt capacitance of the terminals themselves and the net capacitance of the circuit connected to the terminals.

effective internal resistance In an electrochemical cell or battery, a resistance that originates within the electrolyte and electrodes. This resistance is low when the current drain is low; it rises as the current drain increases. It limits the maximum current that the cell or battery can deliver.

effective isolation The condition of components or circuits being so well isolated or shielded that no significant direct coupling, capacitive coupling, or inductive coupling exists between them.

effective instruction The machine-language version of an instruction given in a computer program, as produced by resident software.

effectively bonded The condition afforded by an extremely low-resistance union between two conducting surfaces that are solidly fastened together.

effectively grounded The condition of being connected to earth or to the low-potential end of a circuit by means of an extremely low-resistance connection.

effective parallel capacitance Inherent capacitance that manifests itself in parallel with two circuit points in combination with any lumped capacitance.

effective parallel resistance **1.** The leakage resistance that manifests itself in parallel with a dielectric (e.g., the leakage resistance of a capacitor). **2.** Parallel-resistance effects caused by stray shunt-resistance components.

effective percentage of modulation For a complex waveform, an expression of the equivalent percentage of modulation by a pure sine wave. Given a certain proportion of power in the sidebands with modulation by a complex signal, the effective percentage of modulation is that percentage which, when the modulating signal is sinusoidal, results in the same proportion of power in the sidebands.

effective phase angle In alternating-current (ac) circuits, the phase angle, with respect to waveforms for current and voltage. When both waveforms are sinusoidal, the effective phase angle is the actual phase angle. But when harmonics are present in current or voltage, the angles differ, the difference being greater in capacitive circuits than in inductive circuits.

effective radiated power Abbreviation, ERP or P_{ER}. **1.** A measure of the performance of a wireless transmitting antenna. Suppose a test antenna, A_T, is set up and the field strength in its favored direction at a frequency f is measured at a distance d in free space. Let the field strength thus measured be F watts per square meter. Suppose A_T is replaced with an isotropic radiator, A_I, and the field strength at the same frequency f is measured at the same distance d in free space. Let the radio-frequency (RF) power at the feed point of the isotropic radiator A_I be varied until the field strength is F, the same as it was with the test antenna A_T. Let this RF power be symbolized P. Then P is defined as the effective radiated power (ERP or P_{ER}) of the test antenna, A_T. **2.** The figure defined as in (1), measured in some direction other than the favored direction of a test antenna.

effective resistance **1.** In a coupled circuit, the sum of the actual resistance of the circuit and the reflected resistance of the load. **2.** See EFFECTIVE ANTENNA RESISTANCE.

effective series inductance Inherent (distributed) inductance acting in series with other components in a circuit. The inherent inductance of the wire in a wirewound resistor, for example, manifests itself in series with the resistance of the device.

effective series resistance Inherent (distributed) resistance acting in series with other components in a circuit. Thus, the inherent resistance of the wire in a coil appears in series with the inductance of the coil. Likewise, a capacitor has an effective series resistance because of the resistance of the leads, plates, and connections.

effective shunt capacitance See EFFECTIVE PARALLEL CAPACITANCE.

effective shunt resistance See EFFECTIVE PARALLEL RESISTANCE.

effective sound pressure The root-mean-square (rms) value of instantaneous sound pressure at one point in a sound cycle.

effective speed of transmission In telegraphy (wire or radio) and in electronic data transmission, the transmission speed (characters per minute, bits per second, etc.) that can be reliably maintained for a given period.

effective thermal resistance The effective temperature rise (in degrees per watt of dissipation) of a semiconductor junction above an external reference temperature that is at equilibrium.

effective time For a computer, the time during which useful work is performed.

effective transmission speed See EFFECTIVE SPEED OF TRANSMISSION.

effective value The root-mean-square (rms) value of an alternating-current (ac) quantity. The effective value an alternating current produces in a pure resistance has the same heating effect as the equivalent direct current. See also ROOT MEAN SQUARE.

effective volt An effective potential of one root-mean-square (rms) volt. Also see EFFECTIVE VOLTAGE.

effective voltage Symbol, E_{eff}. The root-mean-square value of alternating-current (ac) voltage (see EFFECTIVE VALUE). For a sinusoidal voltage, $E_{eff} = 0.707E_{max}$, where E_{max} is the maximum value of the voltage. Also called *rms voltage.*

effective wavelength Wavelength in terms of measured frequency and effective propagation velocity.

effects processor In audio systems, a circuit that produces various sound effects via digital signal processing.

efficiency **1.** The ratio of useful power or energy output to total power or energy input to a device or system. **2.** The proportion of applied audio-frequency (AF) power that a loudspeaker converts into acoustic energy. **3.** See ELECTRICAL EFFICIENCY.

efficiency modulation A system of amplitude modulation in which the efficiency of a radio-frequency (RF) power amplifier is varied at an audio-frequency (AF) rate.

efficiency of rectification For a rectifier, the ratio of the direct-current (dc) output voltage to the peak value of alternating-current (ac) input voltage. For percent efficiency, the ratio is multiplied by 100.

efflorescence The giving up of water by a substance upon exposure to air. Some materials used in electronics exhibit this property. Common efflorescent compounds are hydrated ferrous carbonate, ferrous sulfate, and sodium carbonate.

efflorescent material A material exhibiting efflorescence. Compare DELIQUESCENT MATERIAL.

E field **1.** An electric field. **2.** the electric-field component of an electromagnetic wave.

EFL Abbreviation of *emitter-follower logic.*

E_g Symbol for GENERATOR VOLTAGE.

EHD Abbreviation of *electrohydrodynamic(s).*

EHF Abbreviation of EXTREMELY HIGH FREQUENCY.

EHP Abbreviation of *effective horsepower.*

E-H tee A waveguide junction in which E- and H-plane tee junctions intersect the main waveguide at the same point. Also see WAVEGUIDE TEE.

E-H tuner An impedance-transforming E-H tee with two arms that are terminated in tunable plungers for critical adjustments. See WAVEGUIDE PLUNGER.

EHV Abbreviation of *extra-high voltage.*

E_i Symbol for INPUT VOLTAGE.

EIA Abbreviation of *Electronics Industries Association.*

eight-level code A code, such as the *American Standard Code for Information Interchange (ASCII)*, in which each character is represented by eight bits.

E indicator A radar elevation display, in which the horizontal scale shows range and the vertical scale shows elevation.

Einstein equation The equation depicting the interconversion of mass and energy; $E = mc^2$, where E is energy (ergs), m is mass (grams), and c is the speed of light in a vacuum (centimeters per second).

einsteinium Symbol, Es. A radioactive element produced artificially. Atomic number, 99. Atomic weight, 252.

Einstein/Dehaas effect The tendency for an iron or steel cylinder to become magnetized as it rotates.

Einstein shift The decrease in frequency and loss of energy experienced by quanta acted upon by gravitation.

Einthoven string galvanometer A simple galvanometer in which a silvered glass filament carrying current is mounted in a magnetic field set up by either a permanent magnet or an electromagnet. The current causes the filament to be deflected through a distance proportional to current strength, the deflection being observed through a microscope.

EIT Abbreviation of *engineer-in-training.*

either-or operation The logical inclusive-OR operation.

EJC Abbreviation of *Engineers' Joint Council.*

eka-aluminum See GALLIUM.

eka-silicon See GERMANIUM.

EKG Abbreviation of ELECTROCARDIOGRAM. (Also, ECG.)

EKG telemetry See ECG TELEMETRY.

EL Abbreviation of ELECTROLUMINESCENCE.

elapsed time **1.** In data-processing and computer operations, what appears to be the duration of a process, compared with actual processing time, as measured by internal clocks, for example. **2.** The accumulated time, usually expressed in hours, minutes, and seconds, that an operation takes or a machine runs.

elapsed-time meter An instrument that indicates the time an electronic device or system has been in operation. Most such meters are based on electric clockwork that runs only while the system is in operation, holding the count during shutdown periods. Also see ELECTROLYTIC ELAPSED-TIME METER.

elastance Unit, daraf. The opposition of a capacitor to being charged. It is the reciprocal of capacitance.

elastic collision A collision between two charged particles in which neither loses energy—even though they are deflected from their normal paths.

elasticity **1.** The ability of a body to return to its original shape after being deformed. See YOUNG'S MODULUS. **2.** See ELASTIVITY.

elastic limit The maximum stress that can be tolerated by a material without being permanently deformed.

elastic wave A wave in an elastic medium, such as air or water; thus, a wave that is mechanically produced.

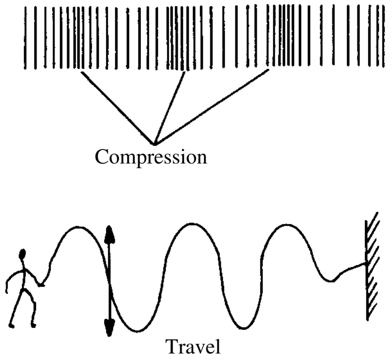

elastic waves

elastivity **1.** Specific elastance (i.e., the elastance in darafs per cubic unit of a dielectric). **2.** The ratio of electric stress to displacement.

elastomer A compressible, conducting substance used in pressure sensing. In one arrangement, an array of electrodes is connected to each side of a pad of the material. When pressure appears at some point, the material compresses, lowering the resistance in the pressure zone. A microprocessor determines the location and extent of the pressure.

E layer A layer in the ionosphere that is 50 to 90 miles above the surface of the earth, and is noted for returning radio waves at high and very-high frequencies.

elbow bend A 90-degree bend in a waveguide.

ELD Abbreviation of *edge-lighted display.*

electra A radio-navigational system in which equal-intensity signal zones (usually 24) are provided.

electralloy A nonmagnetic alloy used in the manufacture of radio hardware, such as chassis.

electre See ELECTRUM.

electrepeter A device used to change the direction of an electric current.

electret A device whose heart is a dielectric disk or slab that is permanently polarized electrically, and so possesses a permanent electric field. The electrical equivalent of the permanent magnet. Certain waxes, ceramics, and plastics acquire permanent polarization after they have been heated, then cooled slowly in an intense electric field.

electret microphone Also called *electret capacitor microphone.* A microphone in which sound waves cause an electret to vibrate and generate audio-frequency (AF) output current. An internal power supply (such as a battery) supplies the necessary voltage.

electric **1.** Pertaining to electricity and its various manifestations. **2.** See ELECTROSTATIC.

electric absorption See DIELECTRIC ABSORPTION.

electric accounting machine A self-contained data-processing machine that is neither a computer nor a computer peripheral.

electrical-acoustical transducer A transducer, such as a headphone, sonic applicator, or buzzer, that converts electrical energy into sound energy. Compare ACOUSTICAL-ELECTRICAL TRANSDUCER.

electrical angle The angle assumed at any instant by the rotating vector representing an alternating current or voltage. A complete cycle is divided into 360 electrical degrees. Thus, for an alternating-current (ac) sine wave, the angle is 0 degrees for zero and positive-going, 90 degrees for positive maximum, 180 degrees for zero and negative-going, and 270 degrees for negative maximum.

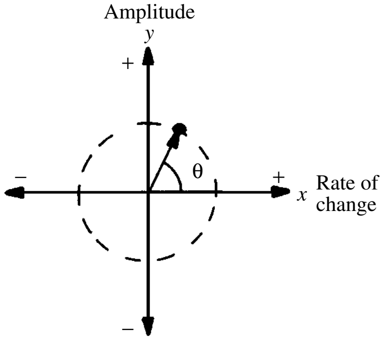

electrical angle

electrical attraction The attraction between two oppositely charged bodies or particles. Compare ELECTRICAL REPULSION.

electrical axis In a quartz crystal, the axis through opposite corners of the hexagonal cross section. The various electrical axes are *x*, *x*’, and *x*” (or *x*-*x*, *x*’-*x*’, and *x*”-*x*”). The electrical axis is perpendicular to the mechanical axis, which runs through the crystal’s length. Also see CRYSTAL AXES and X-AXIS, **2**.

electrical bail An action in which a special switch changes contact position and locks itself in that position after a station has been actuated, at the same time releasing a previously actuated station.

electrical bandspread In a tuned circuit, bandspread obtained by changing values of inductance or capacitance, rather than by mechanical gearing.

electrical bias A current maintained in a relay coil (sometimes an auxiliary coil) to keep the relay partially closed, thus sensitizing it. Compare MECHANICAL BIAS.

electrical boresight In radar operations, the tracking axis, as determined by an electrical test, such as one involving a sharp null or sharp peak response.

electrical center The point at which an adjustable component (variable resistor, variable inductor, etc.) has exactly half its total value. This point does not always coincide with the physical center.

electrical conductance See CONDUCTANCE.

electrical conduction The flow of charge carriers through a material. The degree of conduction is indicated by the material’s value of conductance.

electrical conductivity See CONDUCTIVITY.

electrical coupling The coupling of two or more circuits or elements by means of electric-field effects.

electrical degree **1.** In a periodic waveform, the length of time corresponding to $\frac{1}{360}$ of the time for completion of one cycle. **2.** In space, that distance representing $\frac{1}{360}$ of the wavelength in the medium through which electromagnetic energy travels.

electrical discharge The flow of current out of a voltage reservoir, such as a battery or capacitor.

electrical discharge in gases The phenomenon of electric conduction (current) by a gas, caused by sudden breakdown as a result of gas ionization. The discharge is often accompanied by light, as in the red glow of a neon bulb.

electrical-discharge machining A method of machining metals in which the metal is vaporized by an arc formed between an electrode and the metal workpiece (anode). In this way, metal is removed in tiny bits from the surface of the workpiece.

electrical distance Distance in terms of the time required for an electromagnetic wave to travel between two points in a particular medium.

electrical drainage The diversion of electric currents away from underground pipes to prevent corrosion by electrolysis.

electrical efficiency The ratio of the output of an electrical or electronic device to the total input. It can be expressed as a decimal or percentage. For example, for a bipolar transistor amplifier, the percent efficiency is equal to $100(P_{\text{out}}/P_{\text{in}})$, where P_{in} is the collector input in volt-amperes, and P_{out} is the output power in watts.

electrical elasticity See CAPACITANCE.

electrical element See ELEMENT, **2**.

electrical energy Energy in the form of electricity (see ELECTRICITY, **1**). The term is often used in place of ELECTRICITY.

electrical engineer Abbreviation, EE. A trained professional skilled in applying physics and mathematics to electricity, and in the theory and application of basic engineering and related subjects. Of particular interest to the EE are the generation and distribution of electrical energy and the design and application of electromechanical devices. Compare ELECTRONICS ENGINEER.

electrical erosion In electrical contacts, the loss of metal as a result of the evaporation or transfer of metal during switching.

electrical filter A bandpass, band-rejection, high-pass, or low-pass filter that operates by electrical means. Examples: Butterworth filter and Chebyshev filter.

electrical forming See ELECTROFORM, **1**.

electrical gearing In an electromechanical system, such as a servo, the condition in which an output shaft is electrically rotated at a speed different from that of an input shaft.

electrical glass High-temperature insulating materials made from glass fibers.

electrical inertia See INDUCTANCE.

electrical initiation **1.** Starting an action (electrical or nonelectrical) by means of an electrical signal. **2.** Using an enabling pulse.

electrical instrument A device for measuring an electrical quantity (such as voltage, current, or power).

electrical interlock Also called an *interlock switch* or *door-interlock switch.* A door- or lid-operated switch connected in series with the power switch of a piece of equipment. The interlock causes power to be removed from internal circuits whenever the door is opened, the lid lifted, or the case removed. This minimizes the chance for electric shock to occur to service personnel.

electrical length The length, in wavelengths, of an antenna or transmission line. The electrical length usually differs from the actual length because of ground-capacitance effects, end effects, and the speed of electromagnetic waves in conductors and/or dielectrics.

electrical load A device connected to a source of electricity (generator, amplifier, network, etc.) for a useful purpose (heat, work, etc.).

electrically connected Connected via direct path, such as through a wire, resistance, inductance, or capacitance.

electrically erasable PROM A programmable read-only memory (PROM) that can be erased by an electrical signal, rather than by exposure to ultraviolet light. Also see PROM and ROM.

electrically variable capacitor See VOLTAGE-VARIABLE CAPACITOR.

electrically variable inductor An inductor whose value varies inversely with the amount of direct current that is caused to flow through it or through an auxiliary winding on the same core.

electrically variable resistor See VOLTAGE-DEPENDENT RESISTOR.

electrical nature of matter The general behavior of matter as a complex interplay of waves and particles. Also see ELECTRON THEORY OF MATTER, WAVE MECHANICS, and WAVE THEORY OF MATTER.

electrical network A circuit containing two or more components (including generators and loads), usually arranged in some pattern.

electrical noise Extraneous currents and/or voltages that interfere with desired electrical quantities. Compare ACOUSTIC NOISE.

electrical polarity The distinct difference observable in electrification, designated positive (or plus) and negative (or minus). Negative electrification is generally characterized by a surplus of electrons; positive electrification is characterized by a deficiency of electrons.

electrical quantity **1.** See COULOMB and QUANTITY, **3**. **2.** An electrical unit (e.g., AMPERE, OHM, VOLT, and WATT).

electrical repulsion The mutual repulsion of bodies or particles having similar electric charges. Two positively charged objects will repel each other, as will two negatively charged objects.

electrical reset An electromechanical device for resetting a relay that ordinarily remains in the position resulting from actuation.

electrical resistance The in-phase current-retarding effect that all conductors exhibit to some extent. Also see RESISTANCE.

electrical resistivity See RESISTIVITY.

electrical resolver A synchro whose rotor has two perpendicular windings in addition to another winding.

electrical scan A method of changing the orientation of the major lobe of an antenna. The antenna is kept physically stationary, but the phase/amplitude relationships of the signals applied to different driven elements are varied.

electrical sheet Sheet iron or steel used for motor laminations.

electrical system **1.** The overall configuration of electrical elements for a set of apparatus. **2.** The wiring system that supplies power to a set of devices. **3.** One of several methods of quantizing electrical properties, such as METER-KILOGRAM-SECOND (mks), CENTIMETER-GRAM-SECOND (cgs) electromagnetic, cgs electrostatic, and the INTERNATIONAL SYSTEM OF UNITS (SI).

electrical taste See GALVANIC TASTE.

electrical technology The theory and practical application of electricity. Taught as a subengineering major, usually in two-year colleges that award the degree of Associate in Arts (AA) or Associate in Science (AS).

electrical time constant For a torque motor, the ratio of armature inductance to effective armature resistance. Compare MECHANICAL TIME CONSTANT.

electrical transcription **1.** A phonograph record made electrically, as opposed to one made mechanically. **2.** A radio program in which such a record is played. **3.** Any direct mechanical or electrical recording of an audio signal.

electrical transducer **1.** A transducer that converts a nonelectrical phenomenon into a proportional current, voltage, or frequency. **2.** A transducer that converts electricity in one form to electricity in another [e.g., a transducer actuated by direct-current (dc) voltage, delivering an alternating-current (ac) output, whose frequency is proportional to the dc input voltage].

electrical twinning A defect in which two quartz crystals intergrow in such a way that the electrical sense of their axes becomes reversed. Compare OPTICAL TWINNING.

electrical unit A standard for measuring an electrical quantity (e.g., ampere, ohm, volt, watt, siemens, etc.).

electrical wavelength The distance between one point in an electromagnetic wave cycle and the next identical point. This is usually expressed as the separation between points where the instantaneous amplitude of the electric field is zero and increasing positively. This quantity depends on the velocity factor, ν, of the medium through, or along, which the field propagates. The electrical wavelength also depends on the frequency, f, of the energy. In a transmission line with a velocity factor ν (given as a fraction rather than as a percentage), the electrical wavelength λ is given by

$$\lambda = 3.00 \times 10^8 \nu \;/\; f$$

where s is in meters, ν is in meters per second, and f is in hertz. Often, this formula is modified for values of f expressed in megahertz (Mhz) rather than in hertz. The equation in this case becomes

$$\lambda = 300 \nu \;/\; f$$

The wavelength of a signal in a radio-frequency (RF) transmission line is always less than the wavelength of the same signal in free space.

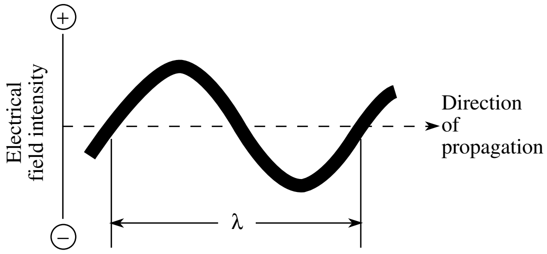

electrical wavelength

electrical zero **1.** A zero-output or minimum-output point resulting from the adjustment of a bridge or other zero-set circuit. **2.** In a meter whose pointer is mechanically set to some point above or below the zero on the scale, the zero setting obtained when the meter is deflected to scale zero by a current or voltage. **3.** For a synchro, the position at which the amplitudes and time phase of the outputs are defined.

electric and magnetic double refraction See KERR ELECTRO-OPTICAL EFFECT and KERR MAGNETO-OPTICAL EFFECT.

electric arc A sustained luminous discharge in the space between two electrodes. Compare ELECTRIC SPARK.

electric aura See ELECTRIC WIND.

electric balance See BRIDGE, **1** and **2**.

electric bell See BELL.

electric brazing A method of brazing in which electric current generates the required heat.

electric breakdown **1.** The usually sudden ionization of a gas by an electric field and the accompanying heavy current flow through the gas. **2.** The (destructive) puncture of a dielectric by the strain produced by high voltage. Also see DIELECTRIC STRENGTH. **3.** The usually nondestructive, abrupt increase in semiconductor junction current at a high reverse voltage. See, for example, AVALANCHE BREAKDOWN.

electric breakdown voltage **1.** The voltage at which avalanche effect occurs. **2.** Dielectric strength.

electric breeze See ELECTRIC WIND.

electric buzzer See BUZZER. Compare ELECTRONIC BUZZER.

electric calculator An electrically driven machine for performing mathematical operations. Its electromechanical nature distinguishes it from the electronic calculator, which features no moving parts, other than keys. Also see CALCULATOR.

electric catfish A fish native to tropical and northern Africa, capable of delivering a strong electric shock.

electric cell See CELL, **1**.

electric chair An electrode-bearing chair used in some states for administering the death penalty via high-voltage electricity. See ELECTROCUTION.

electric charge Potential energy as the electrification of a body or component. For a capacitance of C farads charged to a potential of E volts, the charge Q, in coulombs, is equal to the product CE. Also see ENERGY STORED IN CAPACITOR.

electric chronograph An instrument for accurately recording time intervals.

electric chronometer A precision electric or electronic timepiece. Also see ELECTRIC CLOCK and ELECTRONIC CLOCK.

electric circuit A network of interconnected components and devices, often including a source of electric power. Current flowing through a circuit is acted on by components, which produce specific desired effects.

electric clock A clock driven by electric current. Electric clocks fall into two categories: those driven by synchronous alternating-current (ac) motors, and those driven by stepping mechanisms that usually operate from direct current (dc).

electric column See VOLTAIC PILE.

electric conduction The flow of current carriers through a conductor.

electric constant Also called *permitivity of vacuum.* The fixed electrical permitivity of free space, the value of which is 8.8542×10^{-12} farad per meter.

electric contact See CONTACT, **1**, **2**.

electric controller An adjustable device for modifying the operating voltage or power of a component or system. Compare ELECTRONIC CONTROLLER.

electric cooling **1.** Cooling via PELTIER EFFECT. **2.** See ELECTROSTATIC COOLING. **3.** Forced-air cooling (of equipment) by electric blowers or fans.

electric current The phenomenon wherein charge carriers move in a directed manner through a material or vacuum. In most electrical conductors, current results from movement of electrons. In a semiconductor material, electric current can result from the movement of holes, as well as electrons; the proportion of holes to electrons depends on the nature of the semiconductor. In a gas or electrolyte, current consists of a flow of ions. In certain situations, electric currents can result from the movement of positrons, protons, anti-protons, alpha particles, and various atomic nuclei.

electric current density See CURRENT DENSITY.

electric delay line See DELAY LINE.

electric density See ELECTRIC SPACE DENSITY and ELECTRIC SURFACE DENSITY.

electric dipole A pair of equal charges having opposite polarity and separated by a fixed distance.

electric discharge See ELECTRICAL DISCHARGE.

electric-discharge lamp See DISCHARGE LAMP.

electric disintegration See ELECTRIC DISPERSION.

electric dispersion In a colloidal suspension, dispersion accomplished by passing an electric current through the material.

electric displacement The movement of a body or particle in response to an electric current or field.

electric double refraction See KERR ELECTRO-OPTICAL EFFECT.

electric dust precipitator See DUST PRECIPITATOR.

electric eel An eel (fish) capable of delivering a disabling shock on contact.

electric elasticity See ELASTIVITY, **1**, **2**.

electric endosmosis See ELECTRO-OSMOSIS.

electric eye A sensing device that uses a radiant energy beam to detect objects. It generally uses a laser diode as the beam source, and a photoelectric cell, phototransistor, or photovoltaic cell as the beam detector. The output is used to control some external machine or system.

electric fence A wire fence through which an electric current is passed. Anyone touching the fence will receive a shock. It is used in some prisons, and also by cattle ranchers to keep people or animals contained.

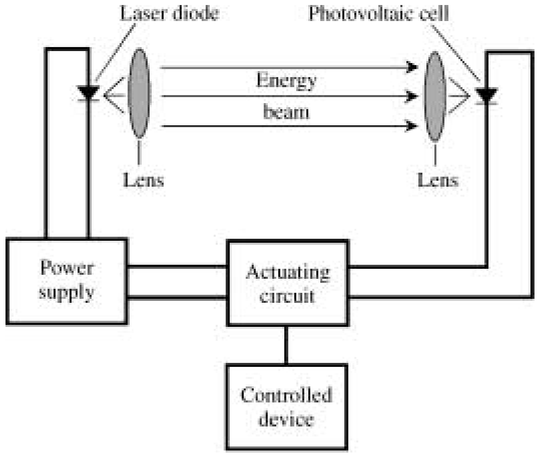

electric eye

electric fidelity The frequency response of a circuit or device.

electric field The space surrounding an electric charge or charged body, in which electric energy acts (electric lines of flux fill the space).

electric field intensity See ELECTROSTATIC FIELD INTENSITY.

electric field strength **1.** Symbol, *E.* In an electromagnetic wave, the amplitude of the electric component of the field, expressed in volts per meter. **2.** Dielectric strength.

electric-field vector See ELECTRIC-FIELD STRENGTH, **1**.

electric filter See ELECTRIC WAVE FILTER.

electric fish Fish capable of generating intense electric shocks (e.g., *electric catfish* and *electric eel*).

electric flux See ELECTROSTATIC FLUX.

electric flux density Symbol, *D.* In an electric field, the number of lines of flux per unit area, usually expressed in coulombs per square meter.

electric flux lines The direction of the electric field in the vicinity of a charged object. The field is denoted by means of "lines of flux" or "lines of force," with each line representing a designated electric field intensity. The closer together the lines, the more intense the field in a given region.

electric focusing See ELECTROSTATIC FOCUSING.

electric force The force exerted by an electrically charged particle or an electric field.

electric forces The forces exerted by electric charges or electric fields. Also see UNIT ELECTROSTATIC CHARGE.

electric furnace An electrically heated chamber (sometimes heated by an electric arc) used in ore reduction, carbide manufacture, and other high-temperature processes.

electric generator A device for producing electricity. Thus, many different devices, such as bat-

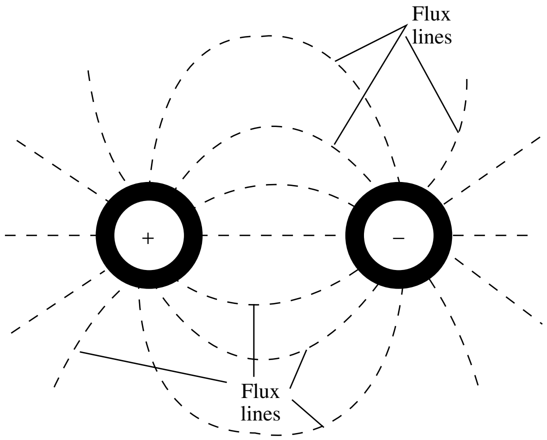

electric flux lines

teries, dynamos, oscillators, solar cells, and thermocouples, are classed as generators.

electric glow The light (usually pale blue) that occasionally accompanies an electric discharge in air.

electric guitar A guitar whose acoustic vibrations are converted by a transducer to an electrical signal for amplification.

electric hygrometer An instrument for measuring humidity in terms of the moisture-sensitive resistance of a sensor. A relatively simple direct-current (dc) circuit is used. Compare ELECTRONIC HYGROMETER.

electric hysteresis See ELECTROSTATIC HYSTERESIS.

electrician A professional person who installs and services electrical equipment and wiring.

electric image For solving certain problems involving electricity, an array of electrical points forming an image of certain other electrical points.

electricity **1.** Phenomena resulting from the existence of stationary or moving electric charge carriers, such as electrons, holes, and ions. **2.** A branch of physics concerned with phenomena resulting from the existence of stationary or moving charge carriers. **3.** Electrical energy or power. **4.** The voltage existing at utility outlets. In the North America, this is nominally an alternating-current (ac) voltage of 117 or 234 V at 60 Hz.

electric lamp An electric-powered device used primarily as a source of light. Common types are arc, incandescent, and fluorescent.

electric light **1.** Visible light produced by means of electricity. **2.** Electric lamp.

electric lines of flux Lines of flux associated with an electric charge and constituting the charge's electric field.

electric machine A mechanical device for generating static electricity. See, for example, ELECTROSTATIC GENERATOR, VAN DE GRAAFF GENERATOR, and WIMSHURST MACHINE.

electric meter **1.** An instrument such as an ammeter, voltmeter, or wattmeter, used to indicate an electrical quantity (usually directly). **2.** See KILOWATT-HOUR METER.

electric mirror See ELECTRON MIRROR.

electric moment In an electric field of unit intensity, the maximum torque exerted on an electric dipole.

electric motor A machine that converts electrical energy into mechanical work. The familiar form is a machine in which an armature rotates between the poles of a field magnet, mechanical energy being produced at the armature's revolving shaft.

electric needle A needle electrode carrying high-frequency current; it is used in surgery to cut tissue and sear it immediately to prevent bleeding.

electric network See ELECTRICAL NETWORK.

electric organ See ELECTRONIC ORGAN.

electric oscillations The alternate flow of electric charges in opposite directions, occurring at a defined frequency or frequencies.

electric osmosis See ELECTRO-OSMOSIS.

electric piano See ELECTRONIC PIANO.

electric polarization **1.** The orientation of flux lines in an electric field. **2.** The orientation of the electric field component in an electromagnetic field.

electric potential See ELECTROSTATIC POTENTIAL.

electric power Symbol, *P*. Unit, watt. The rate at which electrical energy is used. Power is energy per unit time; in the context of electricity, it is expressed as the product of current and voltage. In terms of heat losses, it is often expressed as I^2R (current in amperes squared, multiplied by resistance in ohms).

electric precipitator See DUST PRECIPITATOR.

electric probe A pin or rod inserted into an electrostatic field to sample it, or into an electromagnetic field to sample its electric component. See, for example, WAVEGUIDE PROBE. Compare MAGNETIC PROBE.

electric radiation **1.** The radiation of energy by means of electric waves. **2.** The energy so radiated.

electric recording Inkless recording on paper by direct use of an electric current. The two principal types are: (1) A current-carrying stylus burns away (in a fine line) the metallic coating of the recording paper, exposing the dark underlying layer. (2) A stylus delivers current that produces a line by means of electrolysis in a special paper (see ELECTROLYTIC RECORDER). Compare ELECTROSTATIC PRINTER.

electric reset See ELECTRICAL RESET.

electric residue A residual electric charge, such as might remain on a capacitor after it has been incompletely discharged.

electric rings Colored rings formed on a plate by the electrolytic deposition of substances, such as copper and some peroxides.

electric screen See ELECTROSTATIC SCREEN.

electric shield See ELECTROSTATIC SCREEN.

electric shock A physiological reaction caused by the passage of electric current through living tissue. When slight, it is characterized by tingling sensations and involuntary contractions of the muscles; a severe shock can cause paralysis, unconsciousness, heart fibrillation, and/or burns. If heart fibrillation occurs or if burns are severe, death can result.

electric spark A momentary, luminous discharge of electricity in the space between two electrodes. Compare ARC, **1**.

electric steel Steel that has been processed in an electric furnace.

electric strain See DIELECTRIC STRAIN.

electric strain gauge A device for detecting the strain that a certain stress produces in a body. Typically, such a gauge consists of one or more fine insulated wires cemented to the surface under test. As the surface becomes strained, the wire stretches, undergoing a change in electrical resistance that is proportional to the change in strain.

electric strength See DIELECTRIC STRENGTH.

electric stress See DIELECTRIC STRESS.

electric stroboscope See ELECTRONIC STROBOSCOPE.

electric surface density The ratio of the electric charge on a surface to the area of the surface. Compare ELECTRIC VOLUME DENSITY.

electric tachometer See ELECTRONIC TACHOMETER.

electric telemeter A device that sends metered data over a wire transmission line to a remote point for monitoring.

electric thermometer See ELECTRONIC THERMOMETER.

electric transcription See ELECTRICAL TRANSCRIPTION.

electric transducer A transducer that responds to non-electric energy, then delivers a proportional electric current or voltage to another circuit.

electric tuning A means of adjusting the frequency of a receiver, transmitter, transceiver, or oscillator, without the use of mechanical devices. An example is the use of a varactor diode, whose capacitance varies with voltage.

electric utilities The power companies that supply electric current for industrial and consumer use.

electric vane A small demonstration device consisting of a rotor having several spokes terminating in points. The rotor is mounted on a pivot bearing and, when it is connected to a source of high voltage, spins from the force of electricity escaping from the points into the surrounding air.

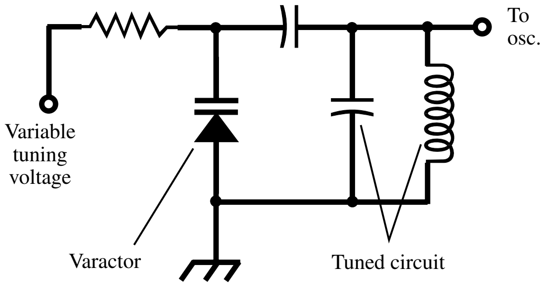

electric tuning

electric vector In an electromagnetic field, the vector representing the electric component. It is perpendicular to the magnetic vector.

electric volume density The ratio of the electric charge in a space to the volume of the space. Compare ELECTRIC SURFACE DENSITY.

electric watch A small timepiece driven by a tiny, self-contained electrochemical cell that drives an electrical escapement or other stepping mechanism. Compare ELECTRONIC WATCH.

electric wave See ELECTROMAGNETIC WAVES.

electric-wave filter A circuit or device for separating signals of one frequency from those of other frequencies.

electric whirl See ELECTRIC VANE.

electric wind **1.** Air currents set up by electrons escaping from the sharp point of a high-voltage electrode. **2.** The outward-rushing plasma (solar wind) ejected by the sun and traveling through space.

electrification **1.** Generating an electric charge in a body, as in charging a glass rod by rubbing it with a silk cloth. **2.** Providing electric service (e.g., RURAL ELECTRIFICATION). **3.** The conversion of a system from purely mechanical to electrical or electromechanical.

electroacoustic Pertaining to devices and systems that exhibit both electrical and acoustic effects. For example, speakers and microphones are electroacoustic devices.

electroacoustic device A device that transfers energy by converting it from electrical to acoustic form or vice-versa.

electroacoustic transducer A transducer that converts sound vibrations into electrical pulsations or, conversely, one that converts electricity into sound.

electroanalysis Chemical analysis performed by electrolytic methods.

electroanesthesia Anesthesia produced by an electric current going through some part of the body.

electroballistics A branch of physics concerned with the electrical or electronic measurement of the velocities and trajectories of projectiles.

electrobath A solution in which ELECTROPHORESIS or ELECTROPLATING is done.

electrobiology Biology concerned with electrical phenomena in living organisms.

electrobioscopy The examination of a body for viability by inducing muscular contractions with an electrical impulse.

electrocapillarity The production of capillary effects by means of electricity. See, for example, CAPILLARY ELECTROMETER.

electrocardiogram Abbreviation, ECG or EKG. A record made by an electrocardiograph of the changes in potential caused by the heartbeat; used as a diagnostic aid.

electrocardiograph An instrument that records changes in electrical potential caused by the heartbeat.

electrocardiophonograph A medical instrument that detects and records the impulses of the heart.

electrocatalysis Catalytic action produced by electricity (see CATALYSIS).

electrocautery **1.** In medicine, a cauterizing instrument consisting essentially of a platinum wire (at the tip of an insulated probe) that is heated by an electric current. Also see ELECTRIC NEEDLE. **2.** Cauterizing with an electrocautery.

electrochemical deterioration An electrochemical reaction that results in the permanent or temporary failure of a device.

electrochemical diffused-collector transistor A high-current pnp transistor in which metal has replaced the etched-away mass of p material, providing a built-in heatsink.

electrochemical equivalent In electrolysis or electroplating, a constant (Z) for the metal in plates. For a given metal, Z is the mass (in grams) of the metal deposited by 1 coulomb of electricity.

electrochemical junction transistor See SURFACE-BARRIER TRANSISTOR.

electrochemical measurements **1.** Measurements made on chemical substances with electrical instruments to determine such factors as conductivity, pH, dielectric strength, dielectric constant, etc. **2.** Measurements of electrical or electronic phenomena in terms of electrochemical response (e.g., current drain in terms of weight of plated metal, or voltage in terms of gas breakdown).

electrochemical polarization The disabling of a primary cell caused by gas products deposited around or on one of the electrodes.

electrochemical recording See ELECTRIC RECORDING.

electrochemical reduction Extracting a metal from a compound by means of electrolysis.

electrochemical series See ELECTROMOTIVE SERIES.

electrochemical switch A static, ionic alternating-current (ac) switch consisting of an anode, cathode, and a control electrode, all immersed in an electrolyte. A positive control-electrode voltage switches the device on, initiating ion current from anode to cathode through the liquid.

electrochemical transducer A transducer that converts chemical changes into electrical quantities, or vice versa. Examples: soil-acidity probe and electrolytic elapsed-time meter.

electrochemistry The branch of chemistry, overlapping with physics, concerned with chemical action arising from the effect of electricity on substances, and electrical effects produced by chemical action.

electrochromic display A display that operates by means of electric fields, which control the light-transmission and light-reflection characteristics in different regions of the material.

electrochronometer A precision electric or electronic clock. Also see ELECTRIC CLOCK and ELECTRONIC CLOCK.

electrocoagulation Use of a high-frequency alternating current to solidify tissue, as in arresting bleeding.

electrocorticogram See ELECTROENCEPHALOGRAM.

electroculture Acceleration or modification of plant growth through the application of electricity to plants, seeds, or soil.

electrocution **1.** Death resulting from electric shock. **2.** An electric shock inflicted for the purpose of causing death (e.g., in an ELECTRIC CHAIR).

electrode A body, point, or terminal in a device or circuit that delivers electricity, or to which electricity is applied. A positive electrode is usually an ANODE; a negative electrode is usually a CATHODE.

electrode admittance The admittance encountered by current flowing through an electrode; the property is entirely that of the electrode and is the reciprocal of ELECTRODE IMPEDANCE. Consists of a real-number component (ELECTRODE CONDUCTANCE) and an imaginary-number component that is 90 degrees out of phase (ELECTRODE SUSCEPTANCE).

electrode capacitance The capacitance between an electrode and a reference body, such as ground or another electrode.

electrode characteristic The mathematical function or graph of electrode current versus electrode voltage.

electrode conductance The conductance encountered by current flowing through an electrode; the property is entirely that of the electrode and is the reciprocal of ELECTRODE RESISTANCE.

electrode current Current entering or leaving an electrode.

electrode dark current See DARK CURRENT.

electrode dissipation The power lost in the form of heat in an electrode.

electrode drop Voltage drop resulting from ELECTRODE RESISTANCE.

electrode efficiency In an electrolytic cell, the ratio of the yield of metal deposited to the maximum possible theoretical yield. This value ranges between 0 and 1 (0 percent and 100 percent).

electrode impedance The impedance encountered by alternating current flowing through an electrode; the property is entirely that of the electrode and is the reciprocal of ELECTRODE ADMITTANCE. Consists of a real-number component (ELECTRODE RESISTANCE) and an imaginary-number component that is 90 degrees out of phase (ELECTRODE REACTANCE).

electrodeless discharge Discharge in a gas tube that is not directly connected to a power source. A familiar example is the glow of a neon lamp held in a strong radio-frequency (RF) electromagnetic field.

electrodeposit **1.** To deposit a substance by electrical action. Also see ELECTROPHORESIS and ELECTROPLATING. **2.** A deposit that is formed on an electrode by electrophoresis or electroplating.

electrodeposition The electrical application of a layer of one material (such as a metal) on the surface of another (the substrate) (e.g., electroplating, evaporation, and sputtering).

electrode potential See ELECTRODE VOLTAGE.

electrode reactance The imaginary-number component of ELECTRODE IMPEDANCE.

electrode resistance The resistance encountered by current flowing through an electrode; the property is entirely that of the electrode and is the reciprocal of ELECTRODE CONDUCTANCE.

electrodermography A method of monitoring the functions of the human body by measuring the resistance between two electrodes placed on the surface of the skin.

electrode susceptance The imaginary-number component of ELECTRODE ADMITTANCE.

electrode voltage The voltage between an electrode and a reference point, such as ground or another electrode.

electrodiagnosis **1.** The diagnosis of a disease or disorder through the use of electromedical instruments. **2.** Troubleshooting the electrical portion of electromechanical equipment.

electrodialysis See ELECTRO-OSMOSIS.

electrodissolution Dissolving a constituent substance of an immersed electrode by means of electrolysis.

electrodynamic Pertaining to electricity in motion (i.e., current flow and its accompanying electric and magnetic fields).

electrodynamic braking Stopping a tape-deck motor quickly by applying a braking voltage. In this method, direct braking current flows through the shaded-pole alternating-current (ac) reel motor.

electrodynamic instrument See ELECTRODYNAMOMETER.

electrodynamic loudspeaker See DYNAMIC SPEAKER.

electrodynamics The branch of physics concerned with moving electric charge carriers, such as electrons, holes, and ions. Also concerned with the interaction of electrical and mechanical phenomena. Important in the design and manufacture of devices, such as motors and generators.

electrodynamic speaker See DYNAMIC SPEAKER.

electrodynamism See ELECTRODYNAMICS.

electrodynamometer An indicating meter whose movable coil rotates between two stationary coils. All three coils are connected in series, and the magnetic fields of the two stationary coils are additive. This meter produces a positive indication for alternating current (ac), as well as for direct current (dc). Can be adapted for use as an ammeter, voltmeter, or wattmeter.

electroencephalogram Abbreviation, EEG. A record made by an ELECTROENCEPHALOGRAPH, showing changes in electric potential resulting from bioelectric action in the brain. The record is used as a diagnostic aid.

electroencephalograph An instrument that produces a record of voltage changes resulting from the brain's bioelectricity.

electroencephaloscope A type of oscilloscope used to pick up, amplify, and display changes in potential caused by the brain's bioelectric action.

electroextraction Extracting a substance from a mixture (e.g., a metal from an ore) by an electrical process, such as electrolysis.

electroform **1.** To precondition a material or device (e.g., a semiconductor junction) by passing a current through it for a specified period. **2.** To form articles by electrodepositing material on a mold or core.

electrogalvanized Electroplated with zinc.

electrogastrogram A recording of the electrical impulses and other functions of the stomach, for medical diagnostic purposes.

electrograph **1.** A picture transmitting or receiving device (see FACSIMILE RECEIVER and FACSIMILE TRANSMITTER). **2.** A device used for the electrolytic etching or transfer of designs.

electrographic recording A method of producing a visible pattern or record, using electrodes to create discharge through an insulating material.

electrographite Synthetic graphite prepared by heating carbon in an electric furnace.

electrojet A region of high current concentration in the sky near bright auroral displays or along the magnetic equator.

electrokinetic energy The energy of electric charge carriers in motion; a form of kinetic energy. The current in a wire (a flow of electrons) is a common manifestation of this form of energy. Compare ELECTROSTATIC ENERGY.

electrokinetics A branch of electricity concerned with (1) the behavior of moving charged particles (such as ions and molecules) and bodies in motion, and (2) the generation of static charges by moving liquids or solids in contact with each other.

electroless process Plating a metal from a solution of one of its salts without using electricity.

electroluminescence The ability of certain phosphors to emit light continuously when an alternating-current (ac) voltage is applied to them.

electroluminescent cell A device for generating light by electroluminescence. It consists of a luminescent-phosphor layer and two transparent metal films. An alternating-current (ac) voltage applied between the films causes the phosphor to glow through the transparent metal.

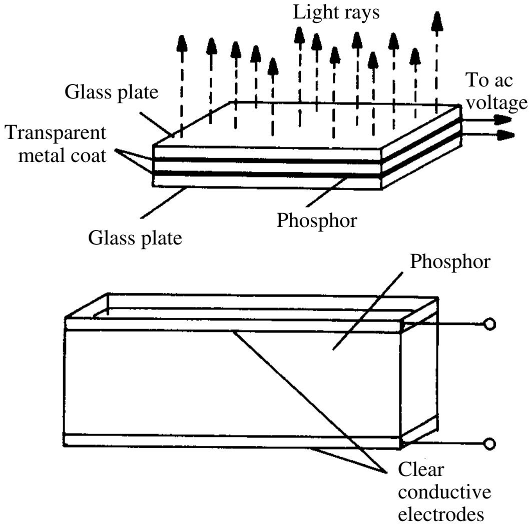

electroluminescent cell

electroluminescent lamp A lamp using one or more electroluminescent panels.

electroluminescent panel A panel that forms a complete electroluminescent cell. It delivers low-intensity visible light when an alternating-current (ac) voltage is applied to it. Available in various sizes and shapes.

electrolysis **1.** The action whereby a current passing through a conductive solution (electrolyte) produces a chemical change in the solution and the electrodes. **2.** An electrical method of destroying hair roots.

electrolyte A substance that ionizes in solution. Electrolytes conduct electricity; in batteries, they are instrumental in producing electricity by chemical action.

electrolytic **1.** Containing an electrolyte substance. **2.** See ELECTROLYTIC CAPACITOR.

electrolytic capacitor A component that is manufactured by rolling up aluminum foil strips separated by paper saturated with electrolyte. When a polarized voltage is applied, the aluminum oxidizes. The oxide layer forms the dielectric for the capacitor. The layer is extremely thin, producing high capacitance per unit volume. This type of capacitor must be connected with the proper polarity in a circuit. The component can have values up to thousands of microfarads, and some units can handle thousands of volts. They are used in audio circuits and in power supplies. Compare CERAMIC CAPACITOR, MICA CAPACITOR, PAPER CAPACITOR, PLASTIC-FILM CAPACITOR, TANTALUM CAPACITOR.

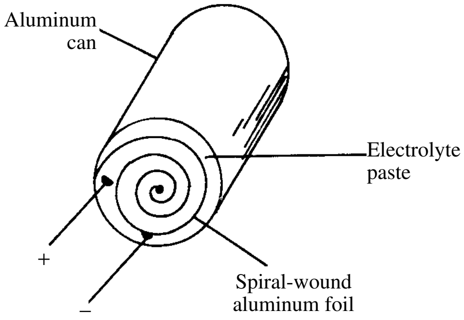

electrolytic capacitor

electrolytic cell An electrochemical device containing an electrolyte and at least two electrodes. Included in this category are voltaic cells, electrolytic capacitors, and electrolytic resistors.

electrolytic conduction Electric current flowing through an electrolyte, an action characterized by (1) positive electrolyte ions migrating to the negative electrode, where they acquire electrons; (2) negative ions migrating to the positive electrode, where they lose electrons; and (3) current flow in the external circuit, which consists of conventional electron flow (current in the electrolyte is a movement of ions).

electrolytic conductivity Conductance of an electrolyte. It is the conductance of a cube of the electrolyte, measuring one centimeter along each edge, at a specified temperature.

electrolytic corrosion Corrosion caused by an applied voltage, or that is accelerated by the voltage. Compare GALVANIC CORROSION.

electrolytic current meter See VOLTAMMETER.

electrolytic dissociation See DISSOCIATION.

electrolytic elapsed-time meter An instrument that indicates the time that equipment has been in operation in terms of the amount of metal electroplated on the cathode of an electrolytic cell by energy consumed during the period.

electrolytic gas A gas produced by electrolysis. Examples are hydrogen (H) and oxygen (O), generated in a ratio of two to one, by the electrolysis of water (H_2O).

electrolytic iron Very pure iron obtained by electrolytic refining.

electrolytic polarization In electrolysis, the tendency for the products to recombine. In an electrolytic cell, this effect interferes with the performance of the cell, reducing the voltage.

electrolytic potential The difference of potential that appears between a metal electrode in an electrolyte and the electrolyte immediately surrounding it. Also see ELECTROMOTIVE SERIES.

electrolytic recorder A data recorder that uses a paper impregnated with a chemical that turns dark when an electric current passes through the paper from the point of a stylus.

electrolytic rectifier A rectifier consisting of an aluminum electrode and a lead or carbon electrode in a solution of borax or sodium bicarbonate, or in a solution of ammonium citrate, ammonium phosphate, and potassium citrate. Also called *chemical rectifier.*

electrolytic refining Extracting or purifying metals by electrolysis.

electrolytic resistor An emergency resistor made by immersing two wire leads in an electrolyte; the weaker the solution, the higher the resistance.

electrolytic switch See ELECTROCHEMICAL SWITCH.

electrolyze To subject something to electrolytic action.

electrolyzer A cell used in the production of various materials by electrolysis. See, for example, ELECTROCHEMICAL REDUCTION and ELECTROLYTIC REFINING.

electromagnet **1.** A coil of insulated wire wound around an iron or steel cylinder, intended for use as a magnet. When current flows through the coil, a magnetic field develops, in effect rendering the cylinder a strong bar magnet. **2.** Any device that exhibits magnetism only while an electric current flows through it.

electromagnetic Exhibiting both electric and magnetic properties (e.g., an *electromagnetic wave*).

electromagnetic attraction **1.** The attraction of iron or steel to an electromagnet. **2.** The attraction of an electromagnetic pole to the opposite pole of another electromagnet. A unit pole attracts another unit pole 1 centimeter away with a force of 1 dyne or 10^{-5} newton. Compare ELECTROMAGNETIC REPULSION.

electromagnetic communication **1.** Any form of communication using a combination of electric and magnetic phenomena. Examples include wire telegraphy, wire telephony, radiotelegraphy, radiotelephony, facsimile, and television. **2.** Electronic communication via electromagnetic fields (i.e., radio communication).

electromagnetic compatibility **1.** The ability of a set of electronic devices to work together without being adversely affected by each other's electromagnetic fields. **2.** In radio communication, the relative immunity of a device or devices to the effects of electromagnetic fields.

electromagnetic complex A system that produces electromagnetic radiation.

electromagnetic component **1.** The magnetic component of an electromagnetic wave, which is perpendicular to the electrostatic component, and can be thought of as the wave's current component. **2.** A device operated by electromagnetism, such as a coil-type relay or a current-operated field magnet.

electromagnetic consonance An effect that takes place when two antenna elements, both having identical or nearly identical resonant frequencies, are in close proximity. If one antenna is fed with energy at its resonant frequency, currents will be induced in the other antenna, and it, too, will radiate. Parasitic arrays, such as the Yagi antenna and the quad antenna, operate on this principle. See RESONANCE.

electromagnetic constant **1.** Symbol, *c*. The propagation speed of electromagnetic waves in a vacuum, approximately equal to 299,792 kilometers per second or 186,282 miles per second. Also called SPEED OF LIGHT and VELOCITY OF LIGHT. **2.** The propagation speed of electromagnetic waves in a particular medium. Equal to vc, where v is the VELOCITY FACTOR of the medium, and c is the speed of electromagnetic waves in a vacuum (299,792 kilometers per second or 186,282 miles per second).

electromagnetic coupling See INDUCTIVE COUPLING.

electromagnetic crack detector An instrument that uses electromagnetic fields to find cracks in iron or steel.

electromagnetic CRT A cathode-ray tube using electromagnetic deflection.

electromagnetic cylinder See SOLENOID.

electromagnetic deflection In a television picture tube and some oscilloscopes, deflection of the electron beam by the magnetic fields of external horizontal- and vertical-deflection coils. Compare ELECTROSTATIC DEFLECTION.

electromagnetic deflection coil See DEFLECTION COILS.

electromagnetic delay line See DELAY LINE.

electromagnetic energy Energy in the form of electric and magnetic fields. A radio wave traveling through space, for example, has electric and magnetic components, between which energy oscillates.

electromagnetic energy conversion The conversion of electrical energy into mechanical work and vice versa, through the intermediary of an electromagnetic field.

electromagnetic environment A region in which electric and magnetic fields are present.

electromagnetic field A combination of alternating electric and magnetic fields. The electric lines of flux are perpendicular to the magnetic lines of flux at every point in space. The field propagates in a direction perpendicular to both the electric and magnetic lines of flux. The frequency of oscillation can range from a fraction of one Hz to many quadrillions of Hz. In order from longest wavelength (lowest frequency) to shortest wavelength (highest frequency), effects of this type include radio waves, infrared, visible light, ultraviolet, X rays, and gamma rays.

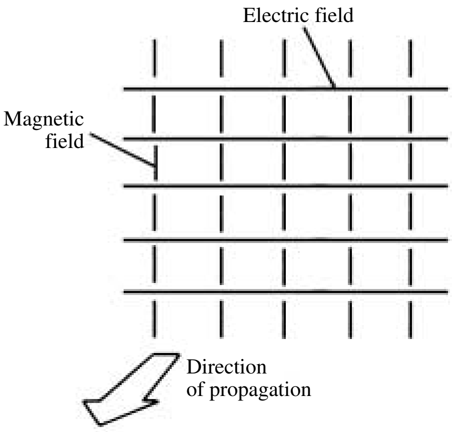

electromagnetic field

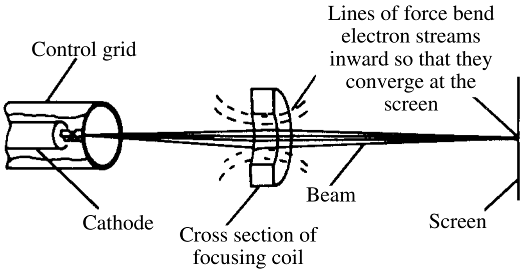
Control grid
Lines of force bend electron streams inward so that they converge at the screen
Cathode
Cross section of focusing coil
Beam
Screen

electromagnetic focusing

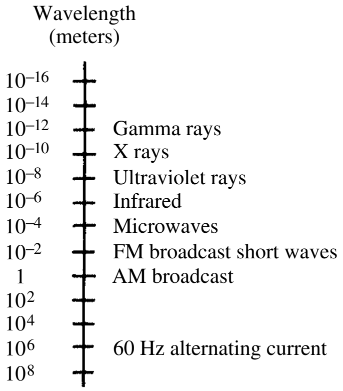

electromagnetic frequency spectrum

electromagnetic flux The magnetic field surrounding a coil or conductor carrying an electric current.

electromagnetic focusing In a television picture tube, electron-beam focusing obtained by varying the direct current flowing through an external focusing coil.

electromagnetic force The force that causes a conductor to be displaced from its position in a magnetic field when it conducts current.

electromagnetic frequency spectrum The frequency range of electromagnetic fields including radio waves, infrared, visible light, ultraviolet, X rays, and gamma rays. Ranges from a fraction of one Hz to many quadrillions of Hz.

electromagnetic horn radiator A horn used to radiate microwave energy. Also see HORN ANTENNA.

electromagnetic induction Inducing a voltage in a circuit or conductor by causing alternating current to flow in another nearby circuit or conductor. Compare ELECTROSTATIC INDUCTION.

electromagnetic inertia The tendency for the current in a circuit to lag the voltage at high frequencies.

electromagnetic interference Abbreviation, EMI. A phenomenon in which electronic devices upset each other's operation. Computers, television receivers, telephone sets, high-fidelity sound equipment, and certain medical devices can malfunction because of strong radio-frequency fields such as those from a nearby broadcast transmitter. The EMI is usually the result of improper or ineffective shielding in the affected device or system.

electromagnetic lens A coil or coil system whose magnetic field causes an electron beam passing through it to converge or diverge as a light beam does in passing through an optical lens. Compare ELECTROSTATIC LENS. Also see ELECTROMAGNETIC FOCUSING.

electromagnetic mass The mass that a moving electric charge is thought to possess.

electromagnetic microphone A microphone in which sound energy is converted into proportionate electrical energy by electromagnetism. Common examples are the *dynamic microphone* and *velocity microphone.*

electromagnetic mirror A reflector of electromagnetic waves (e.g., antenna elements, ionospheric layers, buildings, and hills).

electromagnetic momentum The momentum of a moving electric charge, comparable to that of matter in motion. Electromagnetic momentum is the product of electromagnetic mass and charge velocity.

electromagnetic oscillograph **1.** An electromechanical data recorder for tracing the waveform or variations of a signal. **2.** See ELECTROMECHANICAL OSCILLOSCOPE.

electromagnetic oscilloscope **1.** An oscilloscope using electromagnetic deflection. **2.** See ELECTROMECHANICAL OSCILLOSCOPE.

electromagnetic pump A device used for moving conducting or semiconducting fluids. When a current is passed through the fluid, a force is exerted on the molecules of the fluid because of the magnetic field set up by the current.

electromagnetic pulse Abbreviation, EMP. **1.** A broadband electromagnetic field emitted in a short, intense burst from a lightning stroke or nuclear explosion. This field can disrupt the operation of, and (in some cases) cause damage to, electrical and electronic apparatus. **2.** In electromagnetic induction, the displacement of an electron in a conductor by the magnetic field.

electromagnetic radiation The propagation of electromagnetic fields through space; it normally occurs at approximately 299,792 kilometers per second or 186,282 miles per second.

electromagnetic reaction The reaction between magnetic fields. Also see ELECTROMAGNETIC ATTRACTION and ELECTROMAGNETIC REPULSION.

electromagnetic reconnaissance In military applications, the use of electromagnetic apparatus to detect potential enemy activity in a certain geographic region.

electromagnetic relay See ELECTROMECHANICAL RELAY.

electromagnetic repulsion The repulsion of a pole of an electromagnet by the pole of another electromagnet (north pole opposing north pole, south opposing south). Compare ELECTROMAGNETIC ATTRACTION.

electromagnetics A branch of physics concerned with the theory and application of electromagnetism.

electromagnetic screen See ELECTROMAGNETIC SHIELD.

electromagnetic shield **1.** A partition, can, or box made of magnetic material (iron, steel, or special alloy) enclosing a magnetic component. The magnetic flux generated by the component is confined by the shield, thus preventing interference with external components. Likewise, external magnetic fields are prevented from reaching the component. **2.** A grounded partition, metal sheet, wire braid, or other barrier that prevents electromagnetic fields from passing through. Commonly used in electronic equipment to prevent ELECTROMAGNETIC INTERFERENCE. Also used in COAXIAL CABLE to confine electromagnetic fields to the transmission line.

electromagnetic shielding The use of an ELECTROMAGNETIC SHIELD to prevent undesired interaction among electrical and electronic devices and systems.

electromagnetic spectrum See ELECTROMAGNETIC FREQUENCY SPECTRUM and ELECTROMAGNETIC WAVELENGTH SPECTRUM.

electromagnetic switch **1.** A switch actuated by magnetism produced by control current flowing through a coil wound on an iron core. **2.** See ELECTROMECHANICAL RELAY.

electromagnetic theory of light The theory that light consists of electromagnetic waves that are similar to radio waves, but of shorter wavelength.

electromagnetic tube A cathode-ray tube using electromagnetic deflection (e.g., a television picture tube).

electromagnetic unit Abbreviation, emu. A unit of measure in the electromagnetic system of CENTIMETER-GRAM-SECOND (cgs) units.

electromagnetic vibrator See INTERRUPTER.

electromagnetic wavelength spectrum The wavelength range of electromagnetic fields, including radio waves, infrared, visible light, ultraviolet, X rays, and gamma rays. It ranges from many kilometers to a tiny fraction of one millimeter.

electromagnetic-wave polarization The orientation of the electric flux lines in an electromagnetic (EM) field, especially a field propagating through space. The polarization is generally parallel with the active element of a radio transmitting or receiving antenna. Thus, a vertical antenna radiates and receives fields with *vertical polarization*, and a horizontal antenna radiates and receives fields having *horizontal polarization.* Some antennas radiate and receive an EM field whose polarization continually and rapidly rotates. This is *elliptical polarization.* If the rate of rotation is constant, it is *circular polarization.*

electromagnetic waves Waves produced in a conductor or in space by the acceleration or oscillation of electric charge carriers. Such waves have an electric and a magnetic component acting at right angles to each other. The waves propagate at right angles to both the electric and magnetic flux lines.

electromagnetism **1.** Magnetism resulting from the movement of electric charge carriers (e.g., the magnetic field surrounding a coil of wire carrying an electric current). **2.** See ELECTROMAGNETICS.

electromagnetizer A magnetizer using continuous direct current (dc) as the magnetic-field source.

electromechanical Descriptive of any device that converts energy from electrical to mechanical form or vice-versa. Examples are the motor and the generator.

electromechanical amplifier An amplifier that converts an electrical input signal into mechanical motion (vibratory or rotary), which it then converts back into an electrical output signal of higher current, voltage, or power.

electromechanical chopper A vibrator-type interrupter used primarily to chop direct current, converting it into a square-wave signal, whose amplitude is proportional to current strength. Also see CHOPPER.

electromechanical counter A device that indicates the number of pulses that have been applied to it. Typically, it has a series of dials—each capable of displaying the numerals 0 to 9 in sequence, one for each decade in the count. The dials are geared together, the train being operated by the stepping action of an electromagnetic escapement. Compare ELECTRONIC COUNTER.

electromechanical energy The energy stored by an inductor or capacitor in an electromechanical device.

electromechanical filter See ULTRASONIC FILTER, **1**.

electromechanical flip-flop See BISTABLE RELAY.

electromechanical frequency meter A usually direct-reading instrument for measuring frequency in the lower and middle portions of the audio spectrum. It works via the mechanical motion resulting from the applied signal. The two varieties are the movable-iron type and the reed type.

electromechanical modulator See CHOPPER.

electromechanical oscillator An oscillator consisting of an electromechanical amplifier provided with regenerative feedback.

electromechanical oscilloscope A galvanometer-type instrument for displaying a varying or alternating current or voltage. The signal is applied to a meter movement having a movable coil, which swings or vibrates in response to the signal. A tiny mirror cemented to the coil reflects a beam of light to a rotating mirror that sweeps the beam across a translucent screen on which the image is produced.

electromechanical recorder An instrument in which a pen or stylus is moved on a sheet of paper by the varying signal current or voltage being recorded.

electromechanical rectifier A rectifier in which a moving part, such as a vibrating reed or rotating commutator-slip-ring unit, is driven by alternating current (ac) to close the circuit during positive or negative ac half-cycles, thus rectifying the ac. Compare ELECTROLYTIC RECTIFIER and ELECTRONIC RECTIFIER.

electromechanical relay An electromagnetic switch consisting of a multiturn coil wound on an iron core near an armature with a movable end contact. When control current flows through the coil, it becomes magnetized and attracts the armature, closing the movable contact against a stationary one.

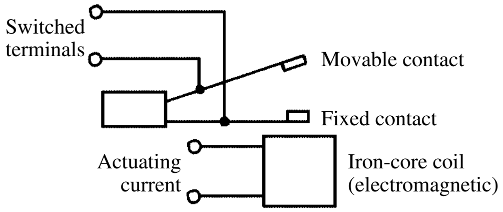

electromechanical relay

electromechanical timer A device for automatically timing a process or an observed event. Most such timers are based on an accurate clock (electric or spring driven) that opens or closes contacts at predetermined instants. Compare ELECTRONIC TIMER.

electromechanical transducer A transducer that translates mechanical signals directly into electrical ones or vice versa, without the intermediary of active devices, such as transistors or integrated circuits.

electromechanical valve A usually poppet-type valve for gases or liquids. The valve is operated by electromagnetic action and is often aided by an electronic (servo) circuit.

electromechanics The theory and application of electromechanical devices.

electromedical engineering The branch of electronics engineering concerned with the theory, design, and application of electronic equipment for medical diagnosis or treatment.

electromedical equipment Electrical or electronic equipment used in medical diagnosis or treatment.

electromerism Ionization in gases.

electrometallurgy The branch of metallurgy concerned with the use of electricity (especially in electrolysis) to separate or purify metals or to furnish heat for metallurgical processes.

electrometer A specially designed, highly sensitive electronic voltmeter used to measure extremely low voltages and, indirectly, extremely low currents. It is sometimes used as a galvanometer.

electrometer amplifier A stable low-noise amplifier for increasing the sensitivity of an electrometer.

electrometry The science of electrical measurements.

electromigration The movement of atoms in a substance from one place to another, because of interaction between electrons and ions in the presence of electric currents. This effect can cause the eventual deterioration of certain semiconductor devices.

electromotion Motion produced by electric charges or electrons.

electromotive force Abbreviation, emf. Electrical pressure, the potential that causes charge carriers to move through a substance or circuit. See VOLTAGE.

electromotive series A list of metals arranged according to the potential between the surface of the metals and an electrolyte into which they are immersed. Some metals acquire a positive potential (with respect to hydrogen, for which the potential is zero) and others, a negative potential.

electromotor A generator or motor, depending upon the context in which the term is used.

electromyogram The record produced by an ELECTROMYOGRAPH.

electromyograph An instrument for detecting, measuring, and analyzing the weak electrical currents generated by muscular activity.

electromyography The monitoring and analysis of the electrical activity of human muscles.

electron The subatomic particle that carries the unit negative charge of electricity. The electron has a mass of 9.109×10^{-31} kilogram and carries a charge of 1.602×10^{-19} coulomb.

electron acceleration See ELECTRON MOTION, **2**.

electronarcosis Loss of consciousness caused by passing a weak current through the brain. Useful in treating certain mental disorders, the process is somewhat similar to ELECTROSHOCK.

electron attachment The bonding of an electron to a neutral atom to form a negative ion. Also see ANION and ION.

electron avalanche See AVALANCHE and ELECTRON MULTIPLICATION.

electron band **1.** An emission line in the spectrum of an element or compound, caused by the movement of electrons from higher to lower energy levels within the atoms. **2.** An absorption line in the spectrum of an element or compound, caused by the movement of electrons from lower to higher energy levels within the atoms.

electron beam See ELECTRON STREAM, **1**.

electron-beam bender Any element that causes intentional deflection of the electron stream in a cathode-ray tube.

electron-beam focusing Reducing the size of the spot produced by the electron beam in a cathode-ray tube or television picture tube. This is accomplished by adjusting the direct-current (dc) bias voltage on a focusing electrode.

electron-beam generator **1.** See ELECTRON GUN. **2.** A tube, such as a Klystron, in which velocity modulation of the electron beam generates extremely high radio frequencies.

electron-beam instrument An instrument, such as an oscilloscope, based on a cathode-ray tube.

electron-beam machining Welding or shaping materials by controlled electron beams.

electron-beam magnetometer A magnetometer in which the magnetic field under measurement deflects the electron beam in a cathode-ray tube over a distance that is proportional to field intensity.

electron-beam recording In digital computer operations, a technique whereby the output of a computer is recorded on microfilm by an electron beam.

electron-beam scanning tube A tube in which an electron beam strikes a sensitized screen to produce a spot of light, which is deflected electrically or magnetically across a screen. Examples are oscilloscope tubes, storage tubes, and television picture tubes.

electron-beam tube Electron tubes (such as beam-power tubes, klystrons, oscilloscope tubes, and television picture tubes), in which an electron beam is generated and controlled.

electron-beam welding A method of welding in which an electron beam is focused on the workpiece to heat it.

electron-bombarded semiconductor A semiconductor wafer, plate, or junction that is acted on by an electron beam; it alters the resistance of the semiconductor to control the current in an external circuit.

electron-bombarded semiconductor amplifier See EBS AMPLIFIER.

electron bunching See BUNCHING.

electron charge See ELEMENTARY CHARGE.

electron cloud A mass of free electrons.

electron diffraction Diffraction that occurs when a beam of electrons passes through a crystal material. Fast-moving electrons have wavelike properties; the wavelength depends on the speed of the particles. This effect can also occur with beams of other particles, such as neutrons, protons, or alpha particles.

electron drift **1.** The movement of an electron from atom to atom in a conductor, as caused by the influence of an applied voltage. **2.** In a semiconductor, directed electron movement. Also see DRIFT, **1**.

electronegative Having negative electrification or polarity (see ELECTRICAL POLARITY). Compare ELECTROPOSITIVE.

electron emission The emission of electrons into surrounding space by a material. Depending on the material, this effect can be initiated by application of heat, light, torsion, electron impact, a high-voltage field, and other actions.

electron flow See ELECTRON DRIFT.

electron g-factor A physical constant that expresses the ratio of electron magnetic moment to the Bohr magneton, and equal to approximately 1.00116. Also called *free-electron g-factor.*

electron-gas binding forces See METALLIC BINDING FORCES.

electron-gas bonding See BONDING, **1** and METALLIC BINDING FORCES.

electron gun A composite electrode for generating an electron beam (see ELECTRON STREAM, **1**) in a vacuum. In a cathode-ray tube, the gun comprises a heated cathode, control electrode, accelerating electrodes, and a focusing electrode.

electron-hole pair In a semiconductor, an electron and a related hole. Each electron in the conduction band has a counterpart in the valence band, a vacancy (hole) left by the electron's moving to the conduction band.

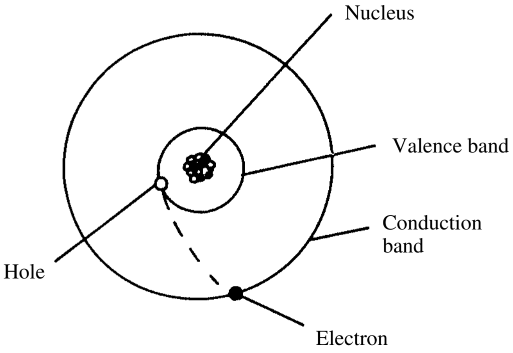

electron-hole pair

electronic **1.** Descriptive of any component, device, or system that functions, according to the principles of ELECTRONICS. **2.** Pertaining to electrons.

electronic adder A circuit (such as an operational amplifier) for performing arithmetic addition. In such a circuit, the output-signal amplitude is the sum of the input-signal amplitudes. Also see ADDER and ANALOG ADDER.

electronic aid An electronic device or circuit that contributes to the operation of a nonelectronic device or system; a pH meter, for example, is an electronic aid to chemistry.

electronics aide See ELECTRONICS TECHNICIAN.

electronic attenuator An attenuator in which the variable resistor is the output section of an active device, such as a transistor. Varying the direct-current (dc) bias of the input section varies the resistance of the output section.

electronic autopilot A servo system for detecting and automatically correcting an aircraft's flight path.

electronic balance An electronic scale, which uses a sensitive current-measuring device in conjunction with a movable tension device.

electronic brain **1.** A high-end computer. **2.** A robot controller. **3.** A system that possesses or uses artificial intelligence (AI).

electronic breadboard **1.** A thin, usually nonmetallic board or card having prepunched holes for the quick assembly of electronic circuits for test and evaluation. **2.** Any circuit prototype that is manually wired during the experimental phase of product development.

electronic bug **1.** A semiautomatic telegraph key that produces dots via an oscillating mechanical arm. The operator produces dashes manually. **2.** A telegraph keying device that automatically generates dots and dashes. Also see ELECTRONIC KEY, **1**. **3.** An undetermined source of problems or improper operation in an electronic circuit.

electronic buzzer **1.** A mechanical buzzer driven by a direct-current (dc) amplifier. **2.** An oscillator circuit that produces a sound similar to that of a mechanical buzzer.

electronic calculating punch A machine that punches on a card the result of calculations it has performed on data it has read from another punched card.

electronic calculator A fully electronic machine for performing mathematical calculations. The simplest machines perform basic arithmetic; more sophisticated ones can do operations with trigonometric, logarithmic, hyperbolic, and other scientific functions. Basic machines are available in department stores for about five dollars. It is usually powered via a small cell or battery, or by a small photovoltaic panel.

electronic camouflage The use of electronics by a target craft to minimize or prevent the reflection of radar echoes.

electronic carillon An electronic system that produces sounds resembling those of a bell carillon.

electronic chime See ELECTRONIC CARILLON.

electronic circuit An electric circuit containing active electronic components, such as transistors and integrated circuits, as opposed to a circuit containing only passive electrical components (such as resistors, switches, heating elements, etc.).

electronic clock **1.** An electric clock whose motor is driven by a constant-frequency oscillator (crystal or tuning fork type), followed by multivibrators and amplifiers. **2.** Any electronic timing circuit that produces pulses at predetermined intervals for the purpose of regulating the operation of other circuits, subsystems, or assemblies.

electronic commutator See COMMUTATOR, **2**.

electronic conduction A flow of electric current resulting from the movement of electrons among atoms in a conductor.

electronic control **1.** The science of automatically controlling machines and devices by means of electronic circuits. **2.** A circuit or device that provides automatic electronic control.

electronic controller A controller (see CONTROLLER, **2**) having no moving parts. For automatic operation, such a device often contains a circuit that senses control signals, compares them with a signal standard, and automatically adjusts the output control power accordingly.

electronic counter A fully electronic circuit that indicates the number of pulses that have been applied to it (see COUNTER, **1**). Unlike the electromechanical counter, the electronic counter has no moving parts and is therefore capable of extremely high-speed, noiseless operation.

electronic counter-countermeasures Abbreviation, ECCM. Military procedures for interfering with a foe's electronic countermeasures.

electronic countermeasures Abbreviation, ECM. Interference with enemy radio and radar emissions by electronic means. Also see JAMMING.

electronic coupling Coupling via electronic effects or devices.

electronic crowbar A switch that prevents destructive currents from flowing through the components of a circuit.

electronic current meter A current meter that uses an amplifier ahead of an analog or digital indicator to provide increased sensitivity.

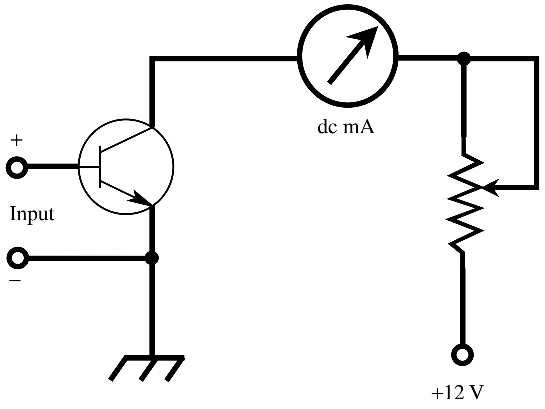

electronic current meter

electronic data processing Abbreviation, EDP. See DATA PROCESSING and ELECTRONIC INFORMATION PROCESSING.

electronic data-processing center Abbreviation, EDPC. An installation of electronic equipment and accessories for processing and storing data, usually in digital form. Also see DATA PROCESSING and ELECTRONIC INFORMATION PROCESSING.

electronic data-processing machine Abbreviation, EDPM. A device, such as an electronic computer, used in the automatic processing of data, usually in digital form.

electronic data-processing system **1.** A unique arrangement of machines for processing data. **2.** The sequence of steps in, and the underlying rationale for, the processing of data by automated equipment.

electronic deception See DECEPTION and DECEPTION DEVICE.

electronic differential analyzer See ANALOG DIFFERENTIATOR, DIGITAL DIFFERENTIAL ANALYZER, and DIFFERENTIAL ANALYZER.

electronic differentiator A circuit that performs mathematical differentiation. Also see DIFFERENTIATOR, **2**.

electronic divider **1.** An electronic device for performing arithmetic division. In a digital computer, such a divider can be a sequence of flip-flops, each of which produces a single output for every two input pulses. In an analog computer, the output signal amplitude is equal to the quotient of two input-signal amplitudes. **2.** See FREQUENCY DIVIDER. **3.** A voltage divider using active components, rather than resistors.

electronic dust precipitator See DUST PRECIPITATOR.

electronic efficiency A quantitative expression for the effectiveness of an electron beam as a medium of power transmission. The electronic efficiency, in percent, is equal to $100P_{out}/P_{in}$, where P_{out} is the (output) power delivered by the beam, and P_{in} is the (input) power supplied to the beam.

electronic equivalent of gravity In equations for the acceleration, velocity, and distance traveled for an electron, the factor equal to eF/m, where e is the electron charge, F is the potential gradient of field, and m is the electron mass. Also see ELECTRON MOTION, **2**.

electronic flash **1.** A device containing a circuit that uses an electronic flash tube as a light source for photography or other purposes. Also called *photoflash.* **2.** A bright momentary light burst produced by the equipment described in **1**, above.

electronic flash tube A tube used to produce brilliant bursts of light in photoflash units, stroboscopes, and laser exciters. A flash tube usually contains xenon gas, which is fired by a high-voltage pulse.

electronic frequency meter **1.** An instrument that gives direct readings of frequency in hertz, kilohertz, or megahertz on an analog scale or as a digital readout. **2.** Any device that indicates the operating frequency of another device, directly or indirectly, when used for such purpose.

electronic frequency synthesizer An instrument that supplies a number of selectable frequencies derived from one or more internally generated fixed frequencies.

electronic gas A collection of free electrons whose behavior resembles that of a gas.

electronic gate **1.** A logic gate that operates by electronic means. **2.** A security system, consisting of a mechanical gate controlled electronically. Similar to an electronic garage door.

electronic guitar See ELECTRIC GUITAR.

electronic heating The production of heat in an object via high-frequency energy. The two principal methods are *dielectric heating* and *induction heating.*

electronic hygrometer An electrical device for measuring relative humidity, whose sensitivity and stability have been increased by the addition of active amplifying devices.

electronician See ELECTRONICS TECHNICIAN.

electronic induction See ELECTROSTATIC INDUCTION.

Electronic Industries Association Abbreviation, EIA. An American association of electronic manufacturers and engineers. It sets standards, disseminates information, provides industry-government liaison, and maintains public relations for the industry.

electronic information processing The use of electronic equipment (especially digital computers and attendant devices) to perform mathematical operations on data entered into the system in the form of electrical signals. Also see DATA PROCESSING.

electronic instrument An instrument whose circuit uses active devices for increased sensitivity over that of the electrical counterpart, and for minimum loading of a device under test. Compare ELECTRICAL INSTRUMENT.

electronic integrator A active device (such as an operational amplifier) for performing mathematical integration. Also see INTEGRATOR, **2**.

electronic intelligence **1.** Information exchanged by electronic means. Examples: radio messages, radar information, and computer data. **2.** The faculties of reasoning and decision making, as apparently simulated by a high-level computer.

electronic interference The malfunctioning of a device because of nearby currents, voltages, or electromagnetic fields.

electronic inverter An electronic device for converting direct current (dc) to alternating current (ac). Typically, an inverter is a transistorized square-wave oscillator inductively coupled to ac output terminals. The dc to be inverted energizes the transistors, which perform the switching function at the rate determined by the components of the circuit. Also see INVERTER, **1**.

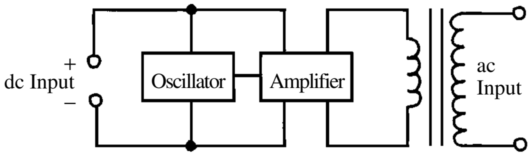

electronic inverter

electronic jamming The deliberate transmission of electromagnetic energy for the purpose of interfering with the operation of a device or devices.

electronic key **1.** For telegraphy (radio or wire), an electronic circuit that generates a continuous string of accurately spaced and timed Morse code dots or dashes, depending on its lever's position (right or left). **2.** For telegraphy (radio or wire), a keyboard that produces perfectly timed Morse code output that corresponds to the operator's typed input. **3.** An electronic device for opening an electronic lock.

electronic keyer See ELECTRONIC KEY, **1**.

electronic lock A lock that will open only after application of a special coded sequence of signals.

electronic mail Also called *e-mail.* A communications system that allows people to leave digital text messages for each other. It is popular among users of personal computers. Operates through the telephone lines using terminal emulation software and a modem. It can also be used via amateur radio packet communications.

electronic microammeter See FET CURRENT METER and TRANSISTOR CURRENT METER.

electronic microvoltmeter See MICROVOLTMETER.

electronic milliammeter See FET CURRENT METER and TRANSISTOR CURRENT METER.

electronic millivoltmeter A millivoltmeter that uses an amplifier ahead of an analog or digital indicator to provide high input impedance and increased sensitivity.

electronic multimeter A *voltohm-milliammeter* that uses active amplifying devices. Also see ELECTRONIC INSTRUMENT.

electronic multiplier A device, such as a Hall generator, whose output is equal (or proportional) to the product of two inputs (i.e., it can perform arithmetic multiplication).

electronic music **1.** Music produced by a combination of electronic oscillator, amplifier, and loudspeaker. A number of successful instruments have been developed. See, for example, ELECTRONIC CARILLON, ELECTRONIC ORGAN, ELECTRONIC PIANO, and THEREMIN. **2.** The electronically amplified sounds of conventional musical instruments.

electronic organ A musical instrument with a keyboard similar to that of a conventional organ, in which tones produced by oscillators or electrically driven reeds are processed and amplified for delivery to a system of loudspeakers.

electronic packaging See ENCAPSULATION.

electronic part A lowest replaceable unit, or component, in an electronic circuit.

electronic phase meter An electronic instrument for measuring phase difference. Direct readings, in degrees of lead or lag, are visible on an analog scale or as a digital readout.

electronic photoflash A transistorized light-intensity meter. Also see ELECTRONIC INSTRUMENT.

electronic piano A musical instrument having the keyboard of a conventional piano and provided with electronic amplification.

electronic power supply A power supply using transistors or integrated circuits for stabilization and output control.

electronic precipitator See DUST PRECIPITATOR.

electronic product Any commercially manufactured electronic device, intended for purchase by the public, by industry, or by government.

electronic profilometer An electronic instrument for measuring surface roughness.

electronic ratchet A stair-step circuit or other arrangement functioning in the manner of an equivalent electromechanical stepping switch. Also see COMMUTATOR.

electronic reconnaissance In military applications, the use of electronic systems to locate enemy installations (such as radio stations, guided-missile sites, and radar bases).

electronic rectifier A rectifier that uses active devices to change alternating current (ac) to direct current (dc).

electronic regulator A voltage regulator that uses active electronic circuits, as opposed to a reactor-type or electromechanical device. See, for example, VOLTAGE REGULATOR.

electronic relay **1.** A switching circuit that uses one or more transistors, and performs the relay function without moving parts. **2.** An electronic component designed to switch when gating signals are applied (e.g., triac, diac, or silicon-controlled rectifier).

electronic resistor **1.** The effective internal collector-emitter resistance of a common-emitter bipolar-transistor stage. **2.** The effective internal drain-source resistance of a field-effect transistor (FET) stage.

electronics The branch of physics concerned with the behavior and application of electric charge carriers in components, devices and systems that accomplish amplification, oscillation, signal processing, and/or switching.

electronics engineer A trained professional skilled in the physics and mathematics of electronics, and in the theory and application of basic engineering and related subjects. Compare ELECTRICAL ENGINEER.

electronics service person An electronics technician skilled in repairing and maintaining electronic equipment. Also called *electronics service technician.*

electronic shutter See KERR CELL.

electronics technician A professional skilled in building, testing, repairing, or maintaining electronic equipment.

electronics technology The theory and practical application of electronics. Taught as a subengineering major, usually in two-year junior colleges or technical institutes, awarding the degree of associate in arts (AA) or associate in science (AS).

electronic stethoscope A stethoscope employing a miniature microphone, amplifier, and earphones. The amplifier gain is continuously controllable, and its bandwidth often selectable for emphasizing particular heart sounds and other body noises.

electronic stimulator A device for applying controlled electrical pulses to the body to stimulate muscles or nerves during diagnosis or therapy.

electronic stroboscope A stroboscope that uses a rate-calibrated oscillator, rather than a mechanical contactor to generate pulses that strobe the lamp.

electronic subtracter An electronic circuit for performing arithmetic subtraction.

electronic surge A sudden, large increase in the current in a conductor. Can be caused by an electromagnetic pulse; can also occur when utility power is restored following a blackout.

electronic switch **1.** A nonmechanical device, such as a flip-flop or gate, whose characteristic on-off operation can be used to make and break an electric circuit. Compare CONTACT SWITCH. **2.** A device using electronic gating and sequencing circuits to present several signals alternately to the single input of an oscilloscope, allowing simultaneous viewing of the signals.

electronic tachometer An instrument for measuring angular velocity, usually in revolutions per minute (rpm). Ideally, the response is independent of sensor voltage amplitude, showing only the number of pulses per unit time reaching the meter circuit.

electronic thermal conductivity The thermal-conductivity component, resulting from the transfer of heat by electrons and holes.

electronic thermometer An instrument for measuring temperature as a result of variations in a temperature-sensitive component, such as a resistor, thermocouple, or thermistor.

electronic timer An electronic circuit or device for automatically timing a process or observed event. Most are based on the time constant of a stable resistance-capacitance (RC) circuit. Compare ELECTROMECHANICAL TIMER.

electronic tube See ELECTRON TUBE.

electronic tuning Variation of the resonant frequency of a device or circuit by changing the bias voltage or current of a controlling electronic component.

electronic voltmeter A voltmeter that uses electronic amplification ahead of the indicating meter to provide high input impedance and increased sensitivity. Also see FET VOLTMETER and TRANSISTOR VOLTMETER.

electronic voltohmmeter A voltohmmeter that uses electronic amplification ahead of the indicating meter to provide high input impedance and increased sensitivity.

electronic typewriter A typewriter with a microcomputer that provides features, such as buffering, automatic repeat, erase memory, and paragraph/page memory.

electronic warfare The use of electronic systems for military purposes, including interfering with an enemy's use of similar systems.

electronic watch **1.** A watch whose movement is a tiny, high-frequency, alternating-current (ac) motor driven by a stable oscillator. **2.** Any miniature timepiece incorporating solid-state circuitry, but especially one using a digital readout.

electronic wattmeter A wattmeter in which an amplifier is used for increased sensitivity.

electronic waveform synthesizer A signal generator that delivers an alternating or pulsating signal whose waveform can be tailored by means of adjustable circuit components.

electron lens A device that focuses an electron beam in a manner similar to the focusing of light rays by a glass lens. Also see ELECTROSTATIC LENS, ELECTROMAGNETIC LENS, and WAVEGUIDE LENS.

electron magnetic moment The energy per unit flux density available in an electron. Approximately equal to 9.2848×10^{-24} joule per tesla.

electron mass See MASS OF ELECTRON AT REST.

electron microscope A microscope in which the source of illumination is an electron beam focused by electromagnetic lenses. It allows much greater magnification than is possible with optical microscopes.

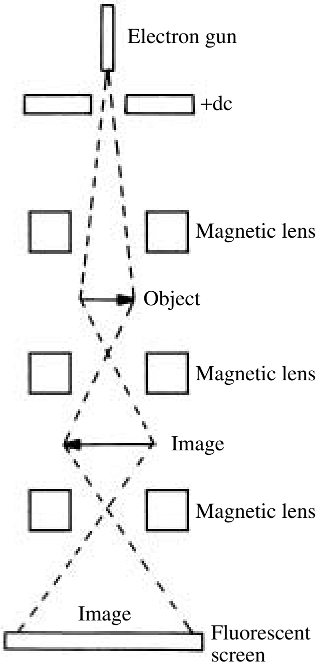

electron microscope

electron mirror A reflector of electrons—especially a dynode element in a photomultiplier tube or electron-multiplier tube.

electron motion **1.** The movement of electrons in a conductor, semiconductor, or space, as the result of electric or magnetic attraction or repulsion. **2.** The movement of an electron as a charged mass. In an electric field, this movement simulates that of a free-falling body in a gravitational field.

electron multiplication **1.** In a gas discharge, the production of additional electrons as a result of collisions between electrons, atoms, and molecules. **2.** The increased production of electrons in a semiconductor when avalanche occurs.

electron-multiplier tube **1.** A vacuum tube utilizing a sequence of secondary emissions for increased current amplification. Electrons from the cathode strike a positively biased dynode with a force that dislodges secondary electrons, which, upon joining those first emitted, are reflected to a second positive dynode that contributes more secondary electrons, reflecting the total to a third positive dynode, etc. The last dynode in the chain reflects the enhanced beam to an anode collector that passes the high current to an external circuit. **2.** See PHOTOMULTIPLIER TUBE.

electronography Printing by means of the electrostatic transfer of ink from a printing plate across a gap to an impression cylinder.

electron optics See ELECTRO-OPTICS.

electron orbits See ELECTRON SHELLS.

electron oscillator A device in which oscillation is obtained by causing electrons to move in an oscillatory path, to travel in bunches, etc. Examples: klystron, magnetron, and traveling-wave tube.

electron pair Two electrons from adjacent atoms, which sometimes share the same orbits, but always produce a bond between two adjacent atoms.

electron-pair bond The bond between an electron pair.

electron physics The physics of electronics, usually from a highly theoretical viewpoint.

electron-proton magnetic moment ratio A physical constant whose value is approximately equal to 658.211, derived from the division of the magnetic moment of the electron by that of the proton.

electron recoil The recoil of an electron from a photon it has collided with.

electron rest mass See MASS OF ELECTRON AT REST.

electron scanning Deflection of an electron beam. See, for example, ELECTROSTATIC DEFLECTION and ELECTROMAGNETIC DEFLECTION.

electron shells The spheres, concentric with the nucleus of an atom, that represent the median distances from the nucleus around which electrons migrate.

electron spin The rotation of an electron (i.e., around its axis). This motion is independent of the electron's movement around the nucleus of an atom.

electron stream **1.** The beam of electrons generated by the electron gun in a cathode-ray tube.

2. The electrons moving between the cathode and plate in an electron tube.

electron-stream instrument See ELECTRON-BEAM INSTRUMENT.

electron-stream meter An oscilloscope (or cathode-ray tube alone) used as a device for making measurements.

electron-stream transmission efficiency The ratio of the current through a positive electrode to the current impinging on it. In a tube, for example, some electrons are absorbed by the plate, and others are reflected.

electron telescope A telescope using a combination of a glass lens, photocathode, and electrostatic focusing. Light from the object is focused on the photocathode by the lens, the electrons emitted being focused electrostatically on a phosphorescent viewing screen.

electron transit time The time required for an electron to travel a given distance. For a vacuum tube, the upper frequency limit of operation is governed by the time required for an electron to reach the plate after leaving the cathode. Transit time is usually stated in fractions of a microsecond.

electron tube An evacuated or gas-filled chamber in which electrons are emitted (usually by a hot cathode) and controlled (usually by a voltage applied to a grid electrode).

electron unit See ELEMENTARY CHARGE.

electron velocity The velocity acquired by an electron that moves between two points having a given potential difference. Also see ELECTRON MOTION, **2**.

electronvolt Abbreviation, eV. The energy acquired by a unit charge moving through a potential difference of one volt; it is equal to approximately 1.6022×10^{-19} joule.

electron-wave tube A tube, such as a klystron or traveling-wave tube, in which electrons traveling at different velocities interact with each other, modulating the electron stream.

electro-oculogram A recording of the voltage that is found between the anterior and posterior parts of the eyeball.

electro-optical transistor A phototransistor or pair of phototransistors in a single package. Used in electronic circuits to sense changes in light levels.

electro-optical valve See KERR CELL.

electro-optic radar A form of radar that makes use of visual apparatus for locating a target.

electro-optics The branch of electronics dealing with related electrical and optical phenomena: photoelectricity, light generation, laser technology, light amplification, etc. It is also concerned with electronic phenomena that are analogous to optical phenomena, such as electronic focusing, reflection, refraction, diffraction, etc. Also called *electron optics* and *optoelectronics.*

electro-osmosis Causing liquids to flow by applying an electric field across the walls of a porous plug. The force exerted by the field on ions in the liquid causes it to flow.

electropad The skin-contacting electrode of an electrocardiograph.

electropathy See ELECTROTHERAPY.

electrophilic Pertaining to the tendency to seek electrons.

electrophobia An irrational fear of electricity, a psychological condition sometimes exhibited by victims of serious electric shock.

electrophonic effect Sound heard by a person when an alternating current is passed through some part of the body.

electrophoresis The movement of dielectric particles through a liquid in which they are suspended, produced by the electric field between electrodes immersed in the suspension.

electrophoresis equipment 1. Any device intended for the purpose of depositing a dielectric material onto a metal by means of electrophoresis. **2.** Any equipment in which electrophoresis occurs.

electrophoresis scanner A device that senses the movement of charged particles caused by electrophoresis effects.

electrophoretic deposition A type of deposition in which a low-voltage direct current passing through a colloidal suspension of dielectric polymer particles deposits them as a coating on a metallic body (the anode in the process). It can provide a better coating than one obtained with spray painting or dipping.

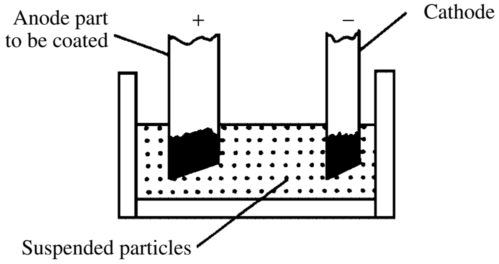

electrophoretic deposition

electrophorus A simple device used to demonstrate electrostatic generation and induction. It consists of a smooth metal plate at the end of an insulating handle and an accompanying cake of resin or hard rubber. The cake is rubbed with cat's fur, making it negatively charged. The metal plate is touched to the charged cake; by induction, it acquires a bound positive charge on the face that touched the cake and a free negative charge on the opposite face. When the plate is lifted and its top face touched momentarily with the finger or grounded, the negative charge leaks off, often with a sharp spark, but the positive charge remains.

electrophotographic process See XEROGRAPHY.

electrophotography The production of photographs by means of electricity. See XEROGRAPHY.

electrophotometer A light-intensity meter using a photoelectric sensor and a meter, but usually not incorporating an amplifier. Compare ELECTRONIC PHOTOMETER.

electrophrenic respiration A system of inducing respiration in which one or both of the phrenic nerves (i.e., of the diaphragm) are stimulated electrically to produce contractions of the diaphragm muscles.

electrophysiology **1.** The study of electrical processes in the human body. **2.** The study of how electrical impulses affect, and are produced by, body organs.

electroplaques In electric fish, small voltage-generating cells that are connected in series-parallel networks.

electroplate **1.** To cause one metallic substance to adhere to the surface of another through the effects of electrolysis. **2.** A metal plating deposited via electrolysis.

electroplating Depositing one metal on the surface of another by electrolytic action.

electropolar Having electrical polarity (either positive or negative).

electropolishing An electrolytic method of smoothing a rough metal surface. The workpiece to be polished becomes the anode of an electrolytic cell in which electrolytic action dissolves tiny surface irregularities.

electropositive Having positive electrification or polarity. Compare ELECTRONEGATIVE.

electropotential series See ELECTROMOTIVE SERIES.

electropsychrometer An electronic instrument for humidity measurements.

electroreduction In electrolysis, reduction of the cathode electrode.

electrorefining The refining of metals by means of electrolysis.

electroretinograph An instrument used to measure the electrical response of the human retina to light.

electroretinography The process of detecting and measuring electrical impulses from the retina.

electroscope An instrument for detecting electric charges and fields. The common type uses a pair of gold-leaf strips hung from the end of a metal rod in a glass tube or jar. When the exposed end of the rod is brought near a charged object, the leaves repel each other and spread apart.

electrosection The use of an arc-generating device for making surgical incisions.

electrosensitive recording See ELECTRIC RECORDING.

electroshock **1.** The system of creating a controlled electric shock in the brain as a treatment for certain mental disorders. **2.** The electric shock used in the therapy described in **1**.

electrospinograph An instrument that senses and records electrical impulses from the spinal cord.

electrostatic Pertaining to stationary electric charges and fields, and their application.

electrostatic actuator A device for measuring the sensitivity of a microphone. Electrostatic charges produce forces on the diaphragm of the microphone, and the resulting output is recorded.

electrostatic amplifier See DIELECTRIC AMPLIFIER.

electrostatic charge See ELECTRIC CHARGE.

electrostatic component The electric component of an electromagnetic wave. It is perpendicular to the magnetic component and can be thought of as the wave's voltage component.

electrostatic constant See ELECTRIC CONSTANT.

electrostatic convergence See ELECTROSTATIC FOCUSING.

electrostatic cooling Accelerated cooling of a body through the application of an intense electrostatic field. The body must be in a free convection state, a corona discharge must be present, and the field must not be uniform.

electrostatic copier A document-copying apparatus that uses electrostatic effects to reproduce printed material.

electrostatic coupling See CAPACITANCE COUPLING.

electrostatic deflection In a cathode-ray tube, deflection of the electron beam by the electrostatic fields between pairs of internal horizontal and vertical deflecting plates. It is primarily used in laboratory oscilloscopes. Compare ELECTROMAGNETIC DEFLECTION.

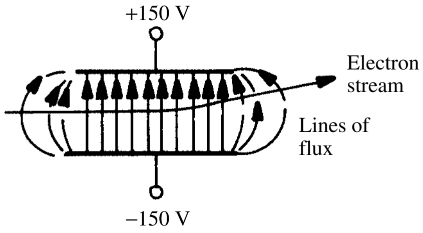

electrostatic deflection

electrostatic electrometer See KELVIN ABSOLUTE ELECTROMETER.

electrostatic electrophotography See XEROGRAPHY.

electrostatic energy The potential energy contained in an electrostatic field (e.g., the energy in a charged capacitor). Compare ELECTROKINETIC ENERGY.

electrostatic field A stationary electric field.

electrostatic field intensity An expression of the strength of an electrostatic field. Usually expressed in volts per meter, millivolts per meter, or microvolts per meter.

electrostatic flux The flux existing around an electric charge or a charged body.

electrostatic focusing In a cathode-ray tube, electron-beam focusing achieved by varying the dc bias voltage on a focusing electrode. Compare ELECTROMAGNETIC FOCUSING.

electrostatic galvanometer A galvanometer operating on the principle of the electrostatic voltmeter.

electrostatic generator A device for producing high-voltage electric charges; e.g., a *Van de Graaff generator.*

electrostatic headphone A device similar to an electrostatic speaker, but held against the head for private listening. Incoming audio signals cause attraction and repulsion among charged plates, resulting in acoustic vibration.

electrostatic hysteresis The tendency of some dielectrics (especially ferroelectric materials) to saturate and retain a portion of their polarization when an alternating electric field to which they are exposed reverses polarity. This causes the charge to lag behind the charging force.

electrostatic induction The charge acquired by a body inserted into an electric field. Compare ELECTROMAGNETIC INDUCTION.

electrostatic instrument See ELECTROSTATIC VOLTMETER.

electrostatic lens An assembly of deflecting plates or cylinders, whose electric field causes an electron beam to converge or diverge in much the same way as a visible light beam passing through an optical lens. Compare ELECTROMAGNETIC LENS.

electrostatic loudspeaker See ELECTROSTATIC SPEAKER.

electrostatic memory A memory unit in which an information bit is stored as an electric charge.

electrostatic memory tube A cathode-ray tube in which information bits are stored in capacitive cells swept by the scanning electron beam.

electrostatic microphone See CAPACITOR MICROPHONE.

electrostatic phase shifter See PHASE-SHIFTING CAPACITOR.

electrostatic potential In an electric field, the potential energy represented by the voltage between the two elements creating the field, or between any two points within the field.

electrostatic precipitator See DUST PRECIPITATOR.

electrostatic printer A computer output peripheral in which the printing medium, a fine dust, is fused by heat onto paper that has been charged according to the data being represented.

electrostatic process **1.** Any process that uses electrostatic action. **2.** A method of photography in which visual images are converted to electrostatic images.

electrostatic recording A method of recording that employs a signal-controlled electric field.

electrostatic relay A high-input-impedance relay consisting of two polarity-controlled contacts; opposite charges on the contacts close the relay, and like charges open it.

electrostatics The branch of electricity concerned with electrical charges at rest. Compare ELECTRODYNAMICS and ELECTROKINETICS.

electrostatic screen A shield against electric flux consisting of a number of straight, narrowly separated rods or wires joined at only one end. The shield has little effect on magnetic flux. Also called *Faraday shield.*

electrostatic separator A device for separating fine particles from a mixture by exposing the mixture to an intense electrostatic field.

electrostatic series A list of materials arranged in this sequence: any one of them becomes positively electrified when rubbed with another lower in the list, or negatively electrified when rubbed with another higher in the list. Compare ELECTROMOTIVE SERIES.

electrostatic shield Any metallic enclosure designed to confine an electric field.

electrostatic speaker A loudspeaker whose vibrating diaphragm is one of two plates in a large air-dielectric capacitor, the other being a closely situated metal plate (or plug). An audio voltage applied to the plates causes them to vibrate. Also called *capacitive loudspeaker* and *capacitor loudspeaker.*

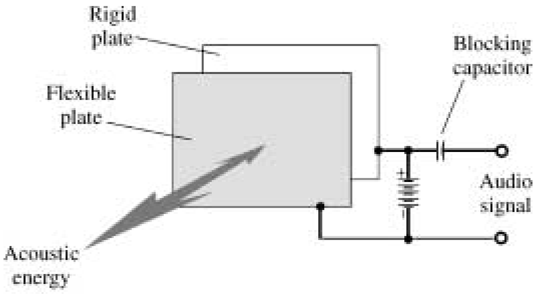

electrostatic speaker

electrostatic sprayer An equipment for spray painting in which fine droplets of paint are attracted by an electrostatic field to the surface to be coated.

electrostatic storage See ELECTROSTATIC MEMORY.

electrostatic stress **1.** Stress in the vicinity of a charged body or particle. **2.** See DIELECTRIC STRESS.

electrostatic transducer See CAPACITANCE TRANSDUCER.

electrostatic tube of flux The space between electric lines of flux going through adjacent points on the boundary of a given area in an electric field.

electrostatic tweeter A small electrostatic speaker for reproducing high-frequency sounds. Compare WOOFER.

electrostatic unit Abbreviation, esu. A unit of measure in the electrostatic system of cgs units. Also see CENTIMETER-GRAM-SECOND. Compare ELECTROMAGNETIC UNIT.

electrostatic vector See ELECTRIC VECTOR.

electrostatic voltmeter An indicating meter whose movement consists of a stationary metal plate near a rotating metal plate. A voltage applied to the plates charges them, and the attraction between them causes the movable member to rotate against the torque of a returning spring over an arc proportional to the voltage.

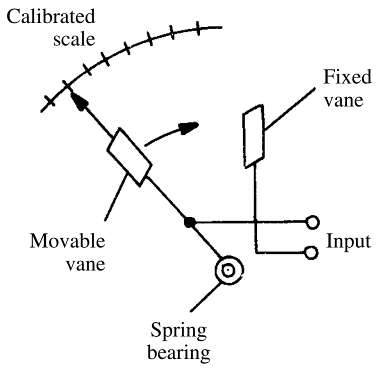

electrostatic voltmeter

electrostatography See XEROGRAPHY.

electrosteel See ELECTRIC STEEL.

electrostenolysis The deposition of certain metals from a solution in capillary tubes when an electric current passes through the solution.

electrostimulation Electrical excitation of nerves for the relief of pain.

electrostriction In certain materials, the physical contraction that occurs when a voltage is applied. Compare MAGNETOSTRICTION.

electrostrictive ceramic A ceramic exhibiting ELECTROSTRICTION when voltage is applied.

electrostrictive relay A relay in which the movable contact is carried by a bar of electrostrictive material, such as barium titanate. A control voltage deforms the material, causing the contacts to close.

electrosurgery Surgery, sometimes bloodless, achieved with diathermy-like equipment. See DIATHERMY, **2**.

electrosynthesis Chemical synthesis produced by means of electric currents or fields.

electrotape Also called *electronic tape measure.* Any device that measures distance by electronic means, such as radar or sonar.

electrotechnology See ELECTRICAL TECHNOLOGY and ELECTRONICS TECHNOLOGY.

electrotellurograph An instrument for measuring ground currents.

electrotherapeutics See ELECTROTHERAPY.

electrotherapy The treatment of disorders or diseases by electrically induced heat—especially by DIATHERMY.

electrothermal **1.** Pertaining to electrically generated heat. **2.** Pertaining to a combination of electricity and heat.

electrothermal device A device whose operation depends on the heat generated by an electric current (e.g., a bolometer, hot-wire ammeter, thermocouple, or varistor).

electrothermal expansion element A thermostatic element, such as a bimetallic strip, whose expansion is used in heat-sensitive switches.

electrothermal instrument A hot-wire or thermocouple-type meter.

electrothermal recorder See ELECTRIC RECORDING, **1** and THERMAL RECORDER.

electrothermic See ELECTROTHERMAL.

electrothermics The study and application of the heating effects of electricity in conductors and junctions.

electrotitration In chemistry, the completion of titration, as indicated by an electrical measurement, such as of the resistance of the solution being titrated.

electrotonic Pertaining to ELECTROTONUS.

electrotonus Modification of a nerve's sensitivity by passing a constant current through it.

electro-ultrafiltration In physical chemistry, filtering a colloidal suspension by electro-osmosis.

electrovalence **1.** The number of charges acquired by an atom gaining electrons. **2.** The number of charges forfeited by an atom losing electrons. **3.** Valence resulting from electron transfer between atoms and the resulting creation of ions.

electrovalency See ELECTROVALENCE.

electrovalent bond See IONIC BOND.

electrowin To recover (win) a metal from a solution of its salts by means of electrolysis.

electrum A natural alloy of gold and silver.

element **1.** See ELECTRODE. **2.** A circuit component intended for a specific purpose. **3.** A specific part of an antenna array (e.g., *driven element* or *parasitic element*). **4.** A fundamental, unique substance whose atoms are of only one kind (examples: aluminum, carbon, silicon, and sulfur). There are more than 100 elements, some man-made. Elements combine to form compounds. **5.** A circuit, such as an AND gate, that can be taken as a unit because it performs a special function. **6.** In digital computer operations, a subunit of a category that cannot be further categorized [e.g., a *bit* (word element) or a *record* (file element)].

elemental area In a facsimile or television picture, a scanning line segment as long as the line's width.

elemental charge See ELEMENTARY CHARGE.

elemental semiconductor A semiconductor containing one undoped chemical element.

elementary charge Symbol, *e*. Also called *unit electric charge.* The electric charge of a single electron or proton. This charge is approximately equal to 1.6022×10^{-19} coulomb.

elementary particle **1.** A minute charged or uncharged particle within the atom (i.e., electron, proton, neutron, quark, etc.). **2.** In theory, a subatomic particle that cannot be broken down into smaller particles.

element error rate In communications or data transfer, the ratio n_i/n_t, where n_i is the number of elements received incorrectly and n_t is the number of elements transmitted.

element spacing **1.** The spacing between radiator, director, and reflector elements in a directional antenna. **2.** The spacing between the internal electrodes of a vacuum tube.

elevation Angular position (in degrees) of a point above the horizontal.

elevation-position indicator A type of radar display simultaneously indicating the elevation of, and the line-of-sight distance to, the target.

elevator control **1.** An electronic system for automatically stopping an elevator and opening the doors. Various safety functions are included, an example being the reopening of a closing door when a passenger steps into the car. **2.** In an aircraft, the mechanical, electronic, or electromechanical devices or circuits involved in actuation of the elevators.

ELF Abbreviation for EXTREMELY LOW FREQUENCY.

eliminator **1.** A device or circuit acting as a surrogate for an inconvenient or undesirable component (e.g., battery eliminator). **2.** A device for removing or minimizing an undesirable signal or quantity (e.g., harmonic eliminator); interference eliminator.

ell A coaxial fitting that is a right-angle line section with a coaxial connector at each end. It takes its name from its L shape.

ellipse A geometric figure having the Cartesian-plane formula $(x - x_0)^2/a^2 + (y - y_0)^2/b^2 = 1$, where a and b are constants, and x_o and y_0 represent the center point.

elliptical filter See ELLIPTIC FILTER.

elliptical function See ELLIPTIC FUNCTION.

elliptical load line For any amplifier with an output transformer, a load line in the shape of an ellipse, obtained when the load connected to the output element is reactive, rather than purely resistive.

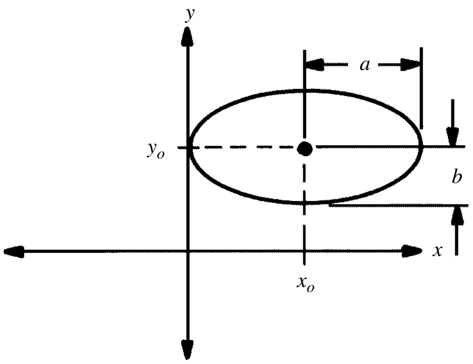

ellipse

elliptically polarized wave An electromagnetic wave in which the rotation of the electric-intensity vector at one point describes an ellipse.

elliptical orbit A satellite orbit that is not a perfect circle. In theory, all satellites deviate slightly from perfectly circular orbits. Sometimes a satellite is deliberately put into an orbit that is greatly elongated. The closer the satellite is to the earth, the faster it moves.

elliptical polarization Polarization characterized by elliptical rotation of the wave vector at a given point.

elliptical stylus In a phonograph (turntable) system, a stylus with a characteristic ellipsoidal shape.

elliptic filter A band-pass, band-stop, high-pass, or low-pass inductance-capacitance (LC) filter, designed according to an ELLIPTIC FUNCTION. Characterized by a steep attentuation-versus-frequency cutoff response with ripple in both the passband and the stopband.

elliptic function A function, similar to the Chebyshev and Butterworth functions, used in the design of certain selective filters. The elliptic function results in a better filter magnitude response than the Chebyshev or Butterworth functions in some applications. See ELLIPTIC FILTER.

elongation A form of modulation distortion resulting from multipath propagation. Some of the paths result in greater propagation delay than other paths; this causes the modulation envelope to spread out. The higher the modulating frequency, the greater the effect.

ELSE A word used in a BASIC computer program that provides an instruction based on a relational test and, in this respect, is related to IF-THEN, ON-GOTO, etc. It specifies the operation to be done if the conditions given in the same program line don't occur.

ELSIE Abbreviation of *electronic letter-sorting and indicator equipment.*

EM **1.** Abbreviation of EFFICIENCY MODULATION. **2.** Abbreviation of ELECTROMAGNETIC(S). **3.** Abbreviation of *electromagnetic iron.* **4.** Abbreviation of ELECTROMAGNETIZER. **5.** Abbreviation of ELECTRON MICROSCOPE. **6.** Abbreviation of EXPOSURE METER. **7.** Abbreviation of *electromotive.*

E_m **1.** Symbol for MAXIMUM VOLTAGE. **2.** Symbol for MAXIMUM JUNCTION FIELD.

emanation **1.** Emission of electrons. **2.** Emission of radioactive particles or ionizing radiation. **3.** Emission of electromagnetic energy.

E_{max} Symbol for MAXIMUM VOLTAGE.

embedded path A means of guiding a mobile robot along a specific route. One common scheme uses a buried, current-carrying wire that produces a magnetic field. The robot can sense and follow this field. Colored paints and tapes can also be used in conjunction with machine vision systems. Compare EDGE DETECTION.

embedded training The inclusion of training/tutorial programs in computerized equipment that assist users in the operation of the equipment.

embedding See ENCAPSULATION.

embossed-foil printed circuit A printed circuit made by pressing the pattern from metal foil into the insulating substrate and then removing the surplus foil.

embossed-groove recording **1.** A phonograph record into which grooves are embossed, rather than scribed. **2.** Recording sound by embossing grooves on record disks.

embossing stylus The rounded-tip stylus used to make an embossed-groove recording.

EMC Abbreviation of ELECTROMAGNETIC COMPATIBILITY.

EME Abbreviation of *earth-moon-earth.* See MOONBOUNCE.

e/m_e The ratio of the elementary electron charge to its mass: 1.7588×10^{11} coulombs per kilogram. Also see CHARGE-MASS RATIO.

Emergency Broadcast System In the United States, a general plan for dissemanating information via broadcast stations in the event of a national emergency.

emergency channel A communication channel allocated for emergency service.

emergency communication Radio or other electronic transmission and reception of urgent messages (e.g., distress signals, storm warnings, etc.).

emergency equipment **1.** Apparatus kept in standby status for immediate operation when regularly used equipment fails. **2.** Equipment, especially vehicular, for use in emergency situations. Examples are ambulances, fire-fighting trucks and equipment, etc.

emergency power supply An alternating-current (ac) or direct-current (dc) power unit kept in standby status for immediate use when the regular power supply fails.

emergency service A communications service devoted exclusively to emergency communication.

EMF, emf Abbreviation of ELECTROMOTIVE FORCE.

emf standard See STANDARD CELL.

EMG **1.** Abbreviation of ELECTROMYOGRAM. **2.** Abbreviation of ELECTROMYOGRAPH.

EMI Abbreviation of ELECTROMAGNETIC INTERFERENCE.

E microscope See ELECTRON MICROSCOPE.

emission **1.** The ejection of particles, especially electrons, from a material. **2.** Waves radiated from any source (as from a transmitting antenna or from an amplifier stage). **3.** The emanation of radiant, electromagnetic, acoustical, electrical, or magnetic energy.

emission code A system of abbreviating the various types of radio emission. See EMISSION MODE.

emission frequency **1.** In communications, the carrier frequency of the transmitted signal as it is radiated from the antenna or fed into a transmission line. **2.** The actual frequency or frequency range of a signal as it is transmitted or radiated. This might be the carrier frequency. **3.** The frequency of energy in an emission band or bands in a spectrum.

emission lines In a spectrum, radiation intensity peaks that appear as bright lines in a visible display. In a radio-frequency spectrum, the emission lines occur as sharp peaks in radiated energy at specific wavelengths.

emission mode Any of various official classifications of radio communication emission types. Emissions are designated according to the modulation method used (e.g., continuous waves, amplitude modulation, single-sideband with suppressed carrier, frequency modulation, pulse modulation, etc.).

emission power **1.** The rate at which energy is radiated from an object. **2.** In radio communication, the transmitter output power.

emission spectrum The radiation spectrum of a substance that emits energy (e.g., the light spectrum of an incandescent metal).

emission types See EMISSION MODE.

emission velocity The initial velocity of an electron as it leaves an emitting surface.

emission wavelength **1.** In communications, the carrier wavelength of the transmitted signal as it is radiated from the antenna or fed into a transmission line. **2.** The actual wavelength or wavelength range of a signal as it is transmitted or radiated. This might or might not be the carrier wavelength. **3.** The wavelength of energy in an emission band or bands in a spectrum.

emissive power The rate at which a surface emits energy of all wavelengths in all directions, per unit area of radiating surface, regardless of temperature.

emissivity For a radiating source, the ratio W_1/W_2, where W_1 is the energy emitted by the source at a particular temperature, and W_2 is the energy emitted by a blackbody (i.e., a theoretically perfect radiator) at the same temperature.

emittance For an energy-radiating source, the radiated power per unit area of radiating surface.

emitted electron An electron that has left an atom of a material and has escaped into surrounding space or entered a neighboring material.

emitter **1.** A body that discharges particles or waves (see EMISSION). **2.** In a semiconductor device, the area, region, or element from which current carriers are injected into the device. In a transistor symbol, the emitter is that electrode shown with an arrowhead.

emitter-base junction In a bipolar transistor, the boundary between base and emitter regions.

emitter bias Emitter current or voltage maintained to set the operating point of a bipolar transistor.

emitter bulk resistance The portion of the resistance of the semiconductor material in a transistor that affects emitter resistance.

emitter-coupled logic A bipolar form of digital logic, abbreviated ECL.

emitter-coupled multivibrator A two-transistor multivibrator circuit in which the emitters share a common resistor.

emitter-coupled phase inverter A transistor phase inverter in which the out-of-phase component is taken from the collector and the in-phase component from the emitter resistor (of the same transistor). Another transistor is often used to amplify the in-phase component so that both outputs are equal in magnitude.

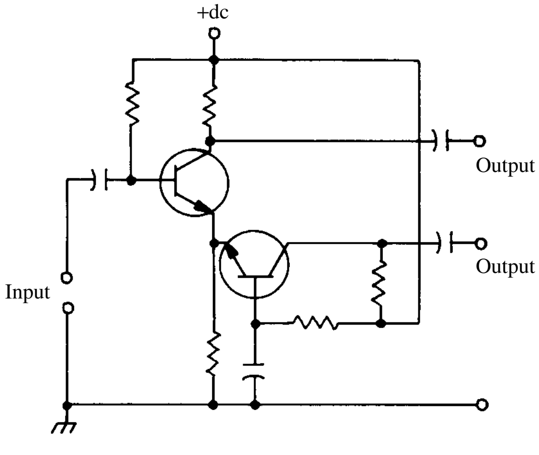

emitter-coupled phase inverter

emitter current Symbol, I_E. The current in the emitter electrode of a bipolar transistor.

emitter degeneration In a transistor amplifier, current degeneration obtained by use of an unbypassed emitter resistor. The arrangement results in virtually distortion-free amplification at a sacrifice in voltage gain.

emitter follower A transistor circuit in which the input signal is applied to the base, and the output signal is taken from the emitter resistor. Gain is always less than unity; output impedance is low.

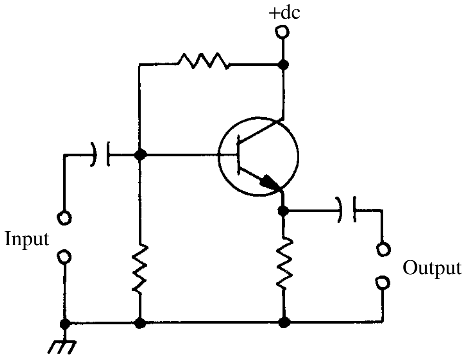

emitter follower

emitter-input circuit See COMMON-BASE CIRCUIT.

emitter junction See EMITTER-BASE JUNCTION.

emitter resistance Symbol, R_E. **1.** The resistance of the emitter electrode in a bipolar transistor. **2.** External resistance connected to a transistor's emitter terminal.

emitter stabilization In a common-emitter transistor stage, an emitter resistor that stabilizes the circuit against temperature variations.

emitter-to-base junction See EMITTER-BASE JUNCTION.

emitter voltage Symbol, V_E. The voltage at the emitter electrode of a bipolar transistor.

EMP, emp **1.** Abbreviation of ELECTROMAGNETIC PULSE. **2.** Abbreviation of *electromagnetic power.*

emphasis Modification of the amplitude-versus-frequency output or response of an audio circuit, for the purpose of optimizing signal intelligibility.

emphasizer An audio-frequency device with a specially tailored response, intended to maximize intelligibility of a voice.

Empire cloth Varnished cambric used as an insulating sheet or tape.

empirical Observable; derived from experimentation.

empirical curve A curve plotted from data acquired from observations, tests, and calculations, rather than from mathematical laws or other theory.

empirical design The design of electronic circuits by cut-and-try methods and, to some extent, through intuition arising from experience (i.e., practical as opposed to theoretical design).

empirical probability Probability estimated from experience and observations. This method is often used in quality-control and reliability procedures.

empty medium A computer storage medium, such as a magnetic tape or disk, that is ready to accept data (i.e., rather than being completely blank, it contains the signals necessary for processing the to-be-added data).

EMU, emu Abbreviation of ELECTROMAGNETIC UNIT(S).

emulator In computer engineering, a sophisticated device that substitutes for a similar device or stage in the computer, and thereby provides a basis for experimenting and troubleshooting without disturbing the equivalent part of the computer.

E_n Symbol for *voltage remaining at null.*

enable To initiate the operation of a circuit or device by applying a pulse or trigger signal.

enable pulse **1.** A pulse that initiates the operation of a circuit or device. **2.** A binary pulse that augments a write pulse to make a magnetic core change state.

enabling gate A digital device that regulates the length of a pulse for specialized use.

enameled wire Wire that is insulated by a thin coat of baked enamel. Commonly used in coil winding because the thin enamel allows for a maximum number of turns in a given volume for a given wire gauge.

encapsulant A material, such as potting resin, used to embed (encapsulate) a component, circuit, or device.

encapsulated circuit A component, circuit, or device embedded in plastic or wax (see ENCAPSULATION).

encapsulated component An electronic part that is embedded in plastic or wax (see ENCAPSULATION).

encapsulating material See ENCAPSULANT.

encapsulation The embedding of a circuit or component in a solid mass of plastic or wax. The mold or container remains as part of the assembly after the plastic or wax has solidified. Protects against the environment, and/or against the effects of physical vibration. Compare POTTING.

encephalogram See ELECTROENCEPHALOGRAM.

encephalograph See ELECTROENCEPHALOGRAPH.

enciphered facsimile Facsimile communications that have been rearranged or scrambled at the transmitting location so that it cannot be intercepted by a third party. A deciphering device is needed at the receiver end of the circuit.

enclosure **1.** A cabinet, case, or other housing for electronic equipment, such as a receiver, transmitter, or test instrument. **2.** A specially designed housing for a loudspeaker.

encode **1.** To convert signals or data into a desired (usually digital) form. Also called CODE. **2.** To equip a transmitter with a tone-producing device (encoder). **3.** To develop and apply an encoding system to a group of transceivers or transmitters of a communications network.

encoder **1.** An analog-to-digital or digital-to-analog converter. **2.** An electromechanical device for translating the angular position of a rotating shaft into a corresponding series of digital pulses. Also see SHAFT-ANGLE ENCODER. **3.** A device for encoding data (see ENCODE). **4.** A machine with a keyboard for printing characters that can be read by optical character recognition (OCR) equipment. **5.** A tone generator used as a receiver enabler in the transmitters of a communications network.

encoding **1.** The translation, either by a machine or by a human operator, of a spoken or written language into digital code. **2.** Any function performed by an ENCODER.

encryption The conversion of a signal from plain text, graphics, or other commonly recognizable form into a cipher. See also CIPHER. Compare DECRYPTION.

end-around carry In a computer, a carry produced in the most significant position, causing a carry into the least-significant position.

end-around shift In digital-computer operations, the transfer of characters from one end of a register to the other end. Also called LOGICAL SHIFT.

end bell **1.** The part of a motor housing that supports the bearing and protects internal rotating parts. **2.** A clamping part fastened to the back of a plug or receptacle. **3.** Either of the two frames of a transformer that contains the mounting lugs.

end bracket See END BELL, **2**.

end cell A cell intended for series operation in conjunction with a storage battery. As the voltage of the battery drops, the end cell can be added into the circuit.

end effect **1.** In a tapped coil, losses because of induced currents flowing in the inductance and distributed capacitance of the unused end of the coil. **2.** EDGE EFFECT in a capacitor. **3.** An effective capacitance at the ends of an antenna, resulting from air discharge. This lowers the resonant frequency slightly below that predicted by theory. The effect is exaggerated by the proximity of objects, such as trees and buildings, or when an antenna is placed close to the earth.

end effector The device or tool connected to the end of a robot arm (e.g., a gripper, screwdriver, drill, or soldering iron).

end-fed antenna An antenna whose lead-in or feeders are attached to an end of the radiator.

end feed A method of feeding electromagnetic fields to an antenna by connecting the transmission line to the end. Ordinarily, the antenna must be a multiple of 0.5 wavelength long for end feed to be effective.

end-fire antenna See END-FIRE ARRAY.

end-fire array Also called *end-fire antenna.* A phased antenna in which the greatest radiation/response takes place off one or both ends. The array consists of two or more parallel driven elements, all of which lie in a single plane. A typical system might consist of two half-wave dipoles, fed 90 degrees out of phase and spaced one-quarter wavelength apart in free space. This produces a unidirectional directivity pattern. Two elements might be driven in phase and spaced 1 wavelength apart, producing a bidirectional pattern. These systems show some power gain, in their favored directions, compared to a single half-wave dipole. The larger the number of elements, with optimum phasing and spacing, the greater the gain.

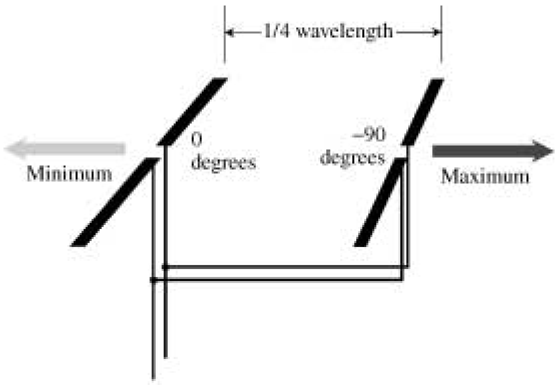

end-fire array

end-fire directivity In a directive antenna, beaming a signal along the plane of the antenna (i.e., off its ends).

end instrument A device capable of converting intelligence into electrical signals or vice versa, and that needs to be connected to only one terminal of a loop.

end item A final, completed product or component.

endless loop Also called *infinite loop.* A computer programming bug that causes the machine to go in an indefinite, and often useless, logical circle. For example, suppose that at line 180, the computer encounters the command GOTO 250, meaning "Go to line 250," but line 250 gives the command GOTO 180. Once the computer gets to line 180, it enters a loop in which nothing is accomplished, and from which the only escape is intervention by the operator (e.g., terminating the program).

end mark In digital-computer operations, a signal or code indicating the close of an information unit.

endodyne reception See ZERO-BEAT RECEPTION.

end-of-charge voltage For a rechargeable cell or battery, the voltage at full charge (i.e., just after disconnection of the charging apparatus and before use).

end-of-data mark A code or character signaling that all the data in a computer storage medium has been read or used.

end-of-discharge voltage For a rechargeable cell or battery, the voltage at the termination of a discharging cycle, immediately before the unit is taken out of use and the charging apparatus is connected.

end-of-field mark In computer operations, a "flag" code that signals when the end of a field has been reached.

end-of-file mark In computer operations, a code instruction that signals when the last record in a file has been read.

end-of-line unit The last device or circuit in a chain.

end-of-message character A character or code signaling the end of a message.

end-of-run The end of a computer program or program run, as indicated by the program.

end-of-tape mark A physical marker at the end of a magnetic tape (e.g., something that can be sensed by methods other than that used to read the tape).

end-of-tape routine A computer program that handles the processing needed after the last record on a reel of magnetic tape has been reached.

end-on armature A relay armature that moves in the direction of the core's axis.

end-on directional antenna See END-FIRE ANTENNA and END-FIRE ARRAY.

endoradiograph An X-ray picture, derived or enhanced by the introduction of substances into the body.

endoradiosonde A tiny pill-enclosed transducer and radio transmitter for sensing physiological conditions in the stomach and intestines; it transmits corresponding signals to instruments outside.

endothermic reaction A chemical reaction producing cold (i.e., one in which kinetic energy is lost). Compare EXOTHERMIC REACTION.

end-plate magnetron A magnetron whose oscillation intensity is increased by a positive and a negative end plate, the electric field between them causes the electrons to move axially while spinning.

end point **1.** For a precision potentiometer, the shaft position between the last and first positions of measurement. **2.** The point at which the useful life of a device can be considered spent. **3.** The point at which a time interval or operational sequence ends. **4.** The end-point voltage of a primary or Edison storage cell. **5.** For a

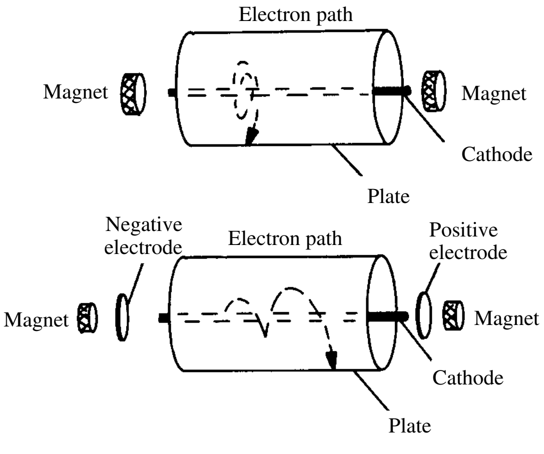

end-plate magnetron

lead-acid storage cell, the specific-gravity value of the electrolyte at which the cell is considered in need of recharging (nominally 1.150 to 1.175).

end-point control A form of quality control in which the end item is checked for defects.

end-point sensitivity A means of expressing the sensitivity of a meter or other indicating device: the ratio, in decibels, between the input signal required to produce a full-scale or maximum reading and the smallest detectable input signal.

end-point voltage The voltage of a battery or cell terminal when the device is no longer useful.

end resistance In a rheostat or potentiometer, the resistance between the wiper and the end terminal when the wiper is set to the end point of the device.

end-resistance offset In a potentiometer, the resistance between the wiper and an end terminal when they are in contact.

end-scale deflection See END-SCALE VALUE.

end-scale value For an indicating meter, the electrical quantity indicated at the last graduation on the scale.

end section Either the input or output section of a multisection filter.

end setting **1.** The fully clockwise or fully counterclockwise setting of a rotatable control. **2.** The minimum or maximum setting of a control.

end shield In a magnetron, a shield that confines the space charge to the interaction space.

end spaces The cavities at either end of the anode block in a multicavity magnetron tube; they terminate all the anode-block cavity resonators.

end use The intended application of a circuit or device.

energize To apply operating power and input signals to a circuit or device.

energized The condition of a circuit or device that is powered or excited.

energy Symbol, W. Common units: joule, watt-hour, and kilowatt-hour. **1.** The capacity for doing work. Some common forms of energy are electrical, mechanical, and chemical. Also see CONSERVATION OF ENERGY, KINETIC ENERGY, and POTENTIAL ENERGY. **2.** The work performed by electric power. The unit used by utility companies is the kilowatt-hour (kWh), equal to the product *Pt*, where *P* is power in kilowatts and *t* is the period (hours) during which the power is used.

energy-band diagram A diagram depicting the various energy levels within the atom of a conductor, semiconductor, or insulator.

energy barrier The natural potential gradient across a semiconductor junction. In the absence of an applied voltage, the gradient, not measurable from the outside, prevents total interaction between the n- and p-type materials.

energy cell **1.** A usually small primary or secondary cell—especially the kind used in hearing aids and electronic watches. **2.** A capacitive-type direct-current (dc) source (see ENERGY-STORAGE DEVICE, **2**).

energy consumption **1.** The conversion of energy from one form to another by a component, circuit, system, or machine, in the process of performing some useful task. **2.** The amount of energy involved in the process defined in **1**.

energy conversion The transformation of energy from one form to another. See also CONSERVATION OF ENERGY and ENERGY TRANSFORMATION.

energy-conversion device A component, circuit, system, or machine that changes energy from one form to another. See also CONSERVATION OF ENERGY.

energy density **1.** For an energy-producing cell, such as an electrochemical cell, the ratio of available energy to cell mass. It is expressed in joules per gram or in watt-hours per kilogram. **2.** For an energy-producing cell, the ratio of available energy to cell volume. Expressed in joules per cubic centimeter or in watt-hours per cubic centimeter.

energy gap In the energy-level diagram for a semiconductor or insulator, the region between valence and conduction bands representing the minimum energy required to make the electron pass from the valence to the conduction band (i.e., to become a current carrier). Also called *forbidden energy band.*

energy level A constant-energy state, such as one of the energy levels of an electron in an atom.

energy-level diagram **1.** A diagram showing the energy levels (in electronvolts) of electrons in the various shells of an atom. **2.** A diagram showing variations in power that correspond to variations in current in a channel.

energy loss In any system, the energy that is unavoidably lost (i.e., it is not converted into useful work). Also see ENTROPY and POWER LOSS.

energy of a charge The energy level of an electrostatic charge. It is $QV/2$ ergs, where Q is the quantity of electricity in coulombs, and V is the potential in volts.

energy product An expression of the effectiveness of a permanent magnet. The magnetic flux density is multiplied by the magnetic field strength to obtain the energy product, specified in gauss-oersteds.

energy redistribution A mathematical process for determining the effective duration of an irregular pulse. The instantaneous power output of the irregular pulse is integrated from the start to the end of the pulse. Then, a rectangular pulse is constructed having the same peak power and the same total energy content (area under the power curve). The length of this rectangular pulse is considered to be the effective duration of the irregular pulse.

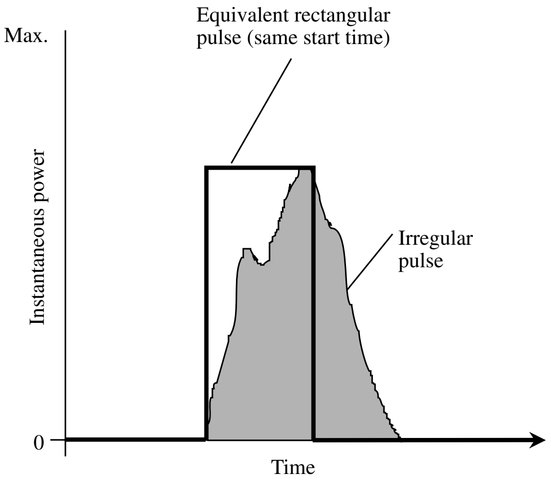

energy redistribution

energy state The condition of an electron, as expressed by its position and velocity, with respect to the position and velocity of other electrons.

energy-storage capacitor A usually high-value capacitor used primarily to store the charge used to fire a lamp (as in a photoflash unit), create a spark discharge (as in electronic ignition), or perform some similar function.

energy-storage device **1.** See CAPACITOR. **2.** A small, electrochemical component offering very high capacitance (e.g., several farads) and low leakage current (less than 1 pA). It has a number of applications, including long-interval timing, power-supply filtering, and energy-cell service. Its active ingredients are compressed powders.

energy stored in capacitor The electrical energy in the field between the plates of a charged capacitor. In this instance, energy $W = CE^2/2$, where W is the energy in joules, C is the capacitance in farads, and E is the voltage in volts.

energy stored in inductor The magnetic energy in the field surrounding an inductor carrying current. In this instance, energy $W = LI^2/2$, where W is the energy in joules, L is the inductance in henrys, and I is the current in amperes.

energy transformation The conversion of one form of energy into another, as with a transducer.

engine analyzer An instrument for checking the performance of an automobile engine. In addition to measuring voltage and resistance throughout a car's electrical system, the instrument measures engine speed, cam dwell angle, and other factors.

engineer **1.** A person who designs machines, circuits, and other devices. **2.** A person who develops methods of utilizing machines, circuits, or other devices more efficiently, or for new applications. **3.** To design or implement an apparatus.

engineering The science of applying scientific laws to technical problems and designing practical devices. Also see ELECTRICAL ENGINEER and ELECTRONICS ENGINEER.

enhanced-carrier demodulation A method of reducing distortion in the demodulation of amplitude-modulated (AM) signals. A properly phased and synchronized local carrier is added to the signal in the demodulator.

enhancement mode Operation characteristic of an ENHANCEMENT-TYPE MOSFET.

enhancement-type MOSFET A metal-oxide semiconductor field-effect transistor (MOSFET) in which the channel directly under the gate electrode is widened (enhanced) by a negative gate voltage in the n-channel unit or by a positive gate voltage in the p-channel unit. Compare DEPLETION-TYPE MOSFET.

ENIAC An electronic computer developed at the University of Pennsylvania. The name is an acronym for *Electronic Numerical Integrator And Calculator.*

ENIC Abbreviation of *voltage negative-impedance converter.*

enrichment In a mixture of different isotopes of a given element, the increase in the relative concentration of one particular isotope.

ensemble **1.** A collection of devices that functions together as a complete unit. **2.** In music recording, the ability of all the musicians to hear each other during the session. **3.** A set of random mathematical functions, all starting at the same point.

ENSI Abbreviation of EQUIVALENT-NOISE-SIDEBAND INPUT.

entladungsstrahlen Ultraviolet radiation emitted by electric arcs. At atmospheric pressure, the wavelength is approximately 40 to 90 nanometers, depending on the arc length. The term is derived from the German word for *discharge rays.*

entrainment Providing a path for gases to escape from an electrochemical cell or battery.

entrance delay In security applications, a delay that allows authorized people time to leave the protected area after activating the alarm system, or to deactivate the system after entering the protected area. The delay is approximately 30 to 45 seconds.

entropy **1.** In all closed physical systems, the measure of energy wasted. According to the second law of thermodynamics, for example, supplied heat can never be converted entirely into work. **2.** In communications, the amount of information in a message, defined as the base-10 logarithm of the number of equivalent messages that can exist. **3.** A natural process in which the energy in the universe tends to become more uniformly distributed with the passage of time.

entropy coding A form of digital encoding that minimizes redundancy, thereby increasing the amount of data in a given amount of memory or storage space.

entry **1.** A unit of computer input or output information. **2.** A data item in a table or list. **3.** A computer source program statement. **4.** In a computer program, the address of the first instruction.

entry condition A condition that must be specified before a computer program is run (e.g., establishing operand values).

entry-level system **1.** The least-sophisticated computer that will perform the things that a user requires. **2.** A simple electronic or computer system (e.g., an amateur radio transciever or personal computer, intended for ease of operation, and from which the user expects to upgrade to a more powerful system at a later date).

entry point In a computer program, the first instruction to be implemented, or a point during the run when data can be entered.

envelope **1.** On a graph, the imaginary line joining successive signal peaks. In the graph for an amplitude-modulated signal, the line reproduces the modulating wave. **2.** The enclosure of a transistor or integrated circuit. **3.** The glass shell of a vacuum tube.

envelope delay In a tuned amplifier, time delay introduced in the envelope of a modulated signal by varying the phase of the envelope with the modulating frequency. This delay varies directly with the amount by which the sidebands shift, with respect to the carrier frequency.

enveloped file A computer file with labels permitting it to be handled by a computer of a type different from that used to make the file.

environmental conditions See ENVIRONMENTAL FACTORS.

environmental factors Aspects of the space immediately surrounding and sometimes influencing electronic equipment. Examples: altitude, dust, light, moisture, noise, pressure, shock, temperature, and vibration.

environmentally sealed Sealed against the effects of adverse environmental factors.

environmental test chamber See CLIMATE CHAMBER.

E_o **1.** Symbol for OUTPUT VOLTAGE. **2.** Symbol for *zero reference voltage.*

EOF Abbreviation of *end of file.*

EOL Abbreviation of *end of line.*

EOLM Abbreviation of *electro-optical light modulator.*

EOR Abbreviation of END OF (program) RUN.

EOS Abbreviation of *electro-optical system(s).*

EOT Abbreviation of *end of tape.*

EOTS Abbreviation of *electro-optical tracking system.*

E_P **1.** Symbol for PLATE VOLTAGE. **2.** Symbol for PEAK VOLTAGE.

EP Abbreviation for EXTENDED PLAY.

ephemeris time Time measured with respect to the orbit of the earth around the sun. Initiated in the year 1900 AD.

epipolar navigation A scheme for position sensing and navigation that uses an artificially intelligent vision system. Allows calculation of position and velocity, based on changes in the visualized direction, size, and shape of an object whose actual location, size, and shape are precisely known. It is used in some mobile robots.

episcotister A mechanical light beam modulator. The device consists of a series of rotating disks having transparent and opaque sections that alternately interrupt and pass the light beam at an audio-frequency rate.

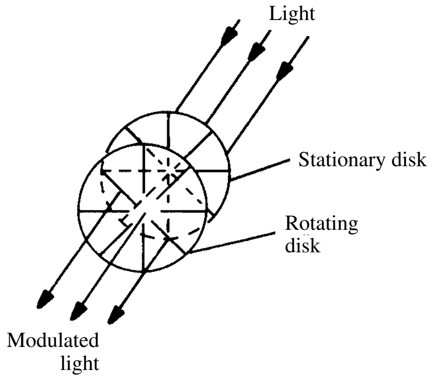

episcotister

epitaxial Pertaining to, or having the property of, EPITAXY.

epitaxial deposition The tendency of certain materials to grow on a semiconductor substrate under certain conditions.

epitaxial device A semiconductor device built by means of EPITAXIAL GROWTH.

epitaxial film A film of single-crystal semiconductor material deposited onto a single-crystal semiconductor substrate.

epitaxial growth Growing monocrystalline silicon on a silicon wafer by precipitating silicon from a gas in which the wafer is placed. Epitaxy is secured between the precipitate and the wafer.

epitaxial layer A semiconductor layer exhibiting epitaxy. Also see EPITAXIAL GROWTH.

epitaxial mesa transistor See DOUBLE-DIFFUSED EPITAXIAL MESA TRANSISTOR.

epitaxial planar transistor A planar transistor having an epitaxially grown collector on a low-resistivity substrate, and a diffused base and emitter.

epitaxial process See EPITAXIAL GROWTH PROCESS.

epitaxial transistor A transistor in which an epitaxial layer (into which a base region later is diffused and an emitter region alloyed) is grown on the face of a semiconductor wafer, which serves as the collector. Also see DOUBLE-DIFFUSED EPITAXIAL MESA TRANSISTOR.

epitaxy The condition in which atoms in a thin film of single-crystal semiconductor material grown on the surface of the same kind of wafer continue their characteristic alignment. Also see EPITAXIAL GROWTH.

E plane The plane of an antenna containing the electric field.

E plane bend See E BEND.

E-plane tee junction A waveguide junction whose structure changes in the plane of the electric field.

epoxy resin A synthetic resin used to encapsulate electronic equipment, or as a cement. Epoxy resins are based on ethylene oxide or its derivatives.

EPROM Abbreviation of *erasable programmable read-only memory.*

EPU **1.** Abbreviation of *electronic power unit.* **2.** Abbreviation of *emergency power unit.*

Eq **1.** Abbreviation of *equation.* **2.** Abbreviation of EQUALIZER or EQUALIZATION.

equal alternations Positive and negative half-cycles of a wave that have identical shape and amplitude.

equal-energy source A light source that has a constant emission rate (energy per unit wavelength).

equal-energy white The color of light emitted by a source radiating equally the wavelengths of the visible-light spectrum.

equal heterodyne In a beat-frequency system, the condition in which the outputs of the two heterodyning oscillators are identical.

equality circuit A logic circuit that, when two numbers are put into it, outputs logic 1 if the numbers are equal, and logic 0 if the numbers are not equal.

equalization **1.** The use of an EQUALIZER to make the frequency response of a line, amplifier, or other device uniform over a given frequency range. **2.** The use of an EQUALIZER to modify the frequency response of a line, amplifier, or other device.

equalizer A circuit or device, such as a compensated attenuator, that allows the user to tailor the frequency response of a line, amplifier, or other device. Sometimes used in sophisticated high-fidelity stereo amplifier systems, to obtain a desired bass/midrange/treble frequency output.

equalizer circuit breaker A form of circuit breaker that trips in the event of unbalance in an electrical system.

equalizing current A current that flows in the circuit of two compound generators connected in parallel.

equalizing network A circuit used to equalize a line.

equalizing pulses In a television signal waveform, several pulses (preceding and following the vertical sync pulse and having a repetition rate of twice the power-line frequency) that start the vertical retrace at the correct instant for good interlace.

equal-loudness curves See AUDIBILITY CURVES.

equal vectors Vectors having the same magnitude and the same direction. They do not necessarily originate at the same point. Compare IDENTICAL VECTORS.

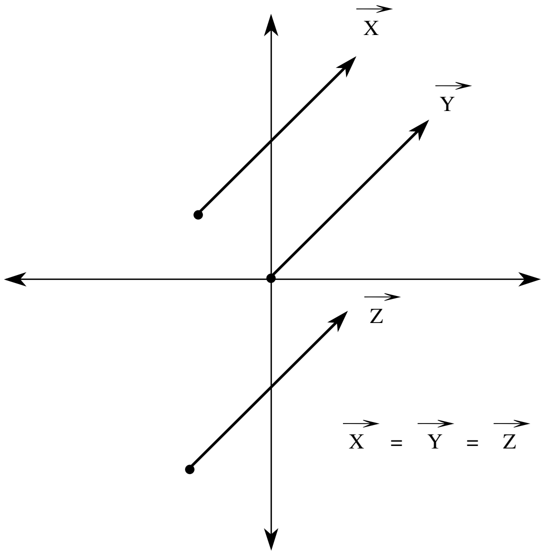

equal vectors

equation solver A (usually analog) computer for solving linear simultaneous equations or for determining the roots of polynomials.

equatorial orbit A satellite orbit that lies in the plane of the earth's equator.

equiphase surface Any surface in a wave, over which the field vectors at a particular instant have either 0° or 180° phase difference.

equiphase zone The space region in which two radionavigation signals show no phase difference.

equipment **1.** Collectively, apparatus or components designated for a specific purpose (e.g., *radio equipment*). **2.** A functional electronic unit, such as a test instrument, receiver, or memory unit.

equipment chain A system consisting of series-connected circuits or devices.

equipment ground An electrical ground connection intended to reduce the chances of electric shock. An equipment ground does not necessarily constitute a good radio-frequency ground; it serves only to eliminate potential differences among the individual units in a system.

equipment life The period during which electronic equipment functions according to specifications; it is terminated at an END POINT.

equipment test A usually preliminary, qualifying test of electronic equipment.

equipotential Having a potential difference of zero; being at the same voltage level.

equipotential line Between two charged plates, the locus (an imaginary line) of points having the same potential, with respect to the plates.

equipotential surface A surface on which all points have the same electrical potential.

equisignal Pertaining to signals having equal intensity.

equisignal localizer See TONE LOCALIZER.

equisignal radio-range beacon For aircraft guidance, a radio-range beacon that transmits two distinct signals that are received by aircraft with equal intensity only in certain directions.

equisignal surface The "surface" around a transmitting antenna formed by points of equal field intensity.

equisignal zone The region in which two radionavigation signals have identical amplitude.

equivalence The condition existing when one network can be substituted for another without disturbing currents, impedances, and voltages at the terminals.

equivalent absorbing power See EQUIVALENT STOPPING POWER.

equivalent absorption Unit, sabin. The rate at which a surface absorbs sound energy.

equivalent binary digits For a given decimal number or specific character, the corresponding binary digits (bits).

equivalent bit rate The number of binary digits (bits) that can be sent in a given unit of time, such as one second, in a digital communications system.

equivalent capacitance The value of a single lumped capacitance, that would cause the same action as the capacitance distributed throughout a circuit.

equivalent circuit A circuit that has the same overall current, impedance, phase, and voltage relationships as a more-complicated counterpart that it usually replaces for analysis.

equivalent component density For a circuit in which discrete components are not used or are not evident, the volume of that circuit divided by the number of discrete components that would be required if the circuit used them.

equivalent conductivity The conductivity of a solution that contains 1 gram equivalent of the solute in the space between electrodes 1 centimeter apart.

equivalent dark-input current For a photoelectric device, the illumination required for an output current equal to the DARK CURRENT of the device.

equivalent decrement The value of decrement in a damped wave that would result in the same amount of interference at a receiver as the interference caused by the sidebands of an amplitude-modulated signal.

equivalent delay line A comparatively simple network, such as a resistance-capacitance (RC) circuit, that will provide the attenuation and phase characteristics of an ideal delay line.

equivalent delta In a three-phase system, a delta-connected circuit that is equivalent to a given wye-connected circuit, from the standpoint of impedance and phase. Also see DELTA CONNECTION and WYE CONNECTION. Compare EQUIVALENT WYE and WYE-EQUIVALENT CIRCUIT.

equivalent differential input capacitance For a differential amplifier, the equivalent input capacitance (see EQUIVALENT CAPACITANCE) at one input (inverting or noninverting) when the opposite input is grounded.

equivalent differential input impedance For a differential amplifier, the equivalent input impedance at one input (inverting or noninverting) when the other input is grounded.

equivalent differential input resistance For a differential amplifier, the equivalent input resistance at one input (inverting or noninverting) when the other input is grounded.

equivalent equations Two equations for an unknown that have the same root.

equivalent four-wire system A two-wire line over which full-duplex operation is obtained by use of frequency division.

equivalent height See VIRTUAL HEIGHT.

equivalent impedance **1.** The value of a single lumped impedance that would cause the same action as the impedance distributed throughout a circuit. **2.** An impedance that draws current of the same strength and phase as that drawn by an impedance it replaces.

equivalent inductance The value of a single lumped inductance that would cause the same action as the inductance distributed throughout a circuit.

equivalent input offset current For a differential amplifier, the difference between currents flowing into the inverting and noninverting inputs when the output voltage is zero.

equivalent input offset voltage For a differential amplifier, the input voltage required to reduce the output voltage to zero.

equivalent input wideband noise voltage For a differential amplifier, the ratio V_o/G_v, where V_o is the root-mean-square (rms) output-noise voltage, and G_v is the direct-current (dc) voltage gain.

equivalent length of antenna **1.** The electrical length of an antenna, as measured in degrees or wavelengths. **2.** The free-space length of an antenna. **3.** The length d (in feet) of a quarter-wave resonant antenna at a specific frequency f (in megahertz), given by the formula $d = 234/f$. **4.** The length d (in feet) of a half-wave resonant antenna at a specific frequency f (in megahertz), given by the formula $d = 468/f$.

equivalent length of electric dipole The distance, measured in a straight line, separating the points that represent the charge centers of an electric dipole.

equivalent length of feed line The electrical length of a feed line as measured in degrees or wavelengths. Generally, this is equal to $1/v$ times the line length in free-space wavelengths, where v is the VELOCITY FACTOR of the line, expressed as a fraction between 0 and 1.

equivalent length of magnet The distance separating the poles of a magnet. In a bar magnet, these poles are not exactly at the ends. The actual equivalent length is about 83% of the length of the bar magnet.

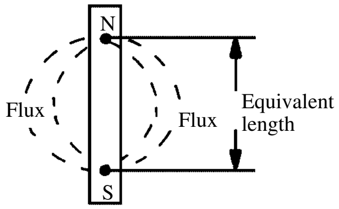

equivalent length of magnet

equivalent loudness The actual intensity, in decibels, of a given sound whose apparent loudness changes with frequency.

equivalent network A network that can replace a more-complex network for analysis purposes.

equivalent noise input The value of modulated luminous flux that, when applied to a photoelectric device, produces a root-mean-square (rms) output current equal to the device's rms noise current.

equivalent-noise-sideband input Abbreviation, **ENSI.** A specification for receiver noise characteristics. Numerically, $ENSI = 0.3E_s(P_n/P_s)^{1/2}$, where E_s is the voltage of an unmodulated radio-frequency (RF) carrier applied to the receiver, P_n is the resulting noise-output power of the receiver (measured with an rms meter), and P_s is the noise-output power measured with the RF signal 30% amplitude modulated at 400 Hz with a 400-Hz bandpass filter inserted between the receiver output terminals and the meter.

equivalent noise temperature For a component having resistance, the temperature (degrees absolute) at which a theoretically perfect resistor having the resistance of the component would generate the same noise the component generates at room temperature.

equivalent optics The analogy between certain optical lenses and prisms and the electrostatic deflection of an electron beam. Thus, when the upper deflecting plate in an electrostatic deflection system is made negative and the lower plate is positive, the beam is deflected downward, like horizontal light rays bent by a prism. When both plates are made equally negative, the beam converges to a point, as light rays do when they pass through a double convex lens. When both plates are made equally positive, the beam spreads out, as do light rays passing through a double concave lens.

equivalent permeability The permeability of a component made of certain materials, compared with that of a component having the same reluctance, shape, and size, but made of different materials.

equivalent reactance The value of a single lumped reactance that would cause the same action as the reactance distributed throughout a circuit.

equivalent resistance The value of a single lumped resistance that would cause the same action as the resistance distributed throughout a circuit.

equivalent series and parallel circuits Series and parallel circuits in which current, voltage, phase, and frequency relationships are identical. Any series circuit can be transformed into an equivalent parallel circuit.

equivalent series resistance The equivalent resistance acting in series with circuit components.

equivalent sine wave A sine wave of the same frequency and effective voltage as a given wave.

equivalent stopping power For a material in the path of radioactive particles, the thickness of the material that produces the same energy loss as that produced by one centimeter of air.

equivalent time The effective duration of some phenomenon, such as a pulse.

equivalent volt See ELECTRONVOLT.

equivalent wye In a three-phase system, a wye-connected circuit that is equivalent to a given delta-connected circuit from a standpoint of impedance and phase. Also see DELTA

CONNECTION, EQUIVALENT DELTA, WYE CONNECTION, and WYE-EQUIVALENT CIRCUIT.

equivalent Y See EQUIVALENT WYE.

equivocation A condition in which the meaning of data depends on certain parameters.

ER Abbreviation of ECHO RANGING.

Er Symbol for ERBIUM.

E_r Symbol for voltage drop across a resistance.

erasable storage In computer operations, any storage medium holding information that can be erased.

erasable PROM A programmable read-only memory (PROM) from which the data can be removed, usually by exposure to ultraviolet light. Also see PROM.

erase To obliterate or remove a signal, especially a recorded one, as in the erasure of recorded material from a magnetic tape or the data from a computer disk.

erase button A pushbutton that actuates the circuit supplying a signal that erases stored material (as the display on a storage oscilloscope).

erase current In an electromagnetic erase head, the current flowing through the coil of the head. In most instances, it is a high-frequency current (usually the regular bias current), but it can be as low as 60 Hz, as long as the speaker does not respond to what remains of it on the tape after erasure.

erase head In a tape recorder, a head used to erase recorded material from tape. It can contain a permanent magnet (see ERASE MAGNET) or an electromagnet whose coil carries erase current.

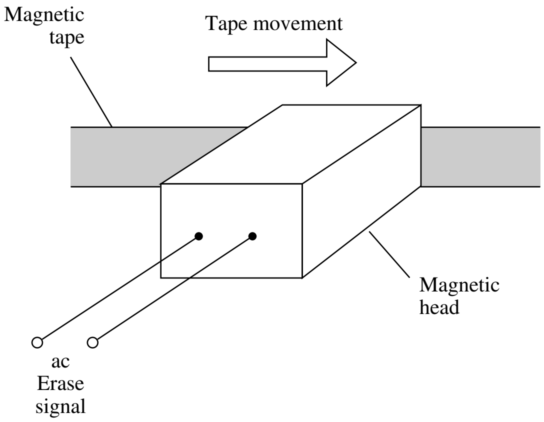

erase head

erase magnet In a tape recorder, a magnet used to erase recorded material from tape. Because the strength of the magnet is greater than that of the magnetized areas on the tape, erasure is complete (the tape left demagnetized).

erase oscillator In a tape recorder, a high-frequency (typically 30 to 80 kHz) oscillator that supplies erase current.

eraser See BULK ERASER.

erase signal A signal that causes recorded material to be erased (see ERASE and ERASE CURRENT).

erasing speed The rate at which successive storage elements are erased, as in a charge-storage tube.

erasure **1.** In tape-recording and digital-computer operations, the process of erasing a recorded signal (see ERASE). **2.** An erasure accomplished by the process described in 1, above.

erbium Symbol, Er. A metallic element of the rare-earth group. Atomic number, 68. Atomic weight, 167.26.

E region See E LAYER.

e register In a computer, a register used in double-precision calculations.

E_r-E_y signal In color television, the resultant signal that is the difference between the original full-red and E_y signals.

ERG Abbreviation of ELECTRORETINOGRAPHY.

erg Abbreviation, *e*. A unit of work. It is the work done by a force of one dyne (10^{-5} newton) acting through a distance of one centimeter.

ergograph An instrument used to measure and record work done by muscles.

ergometer An instrument for measuring energy consumed or work accomplished.

ergon See ERG.

E_{rms} Symbol for ROOT-MEAN-SQUARE VOLTAGE.

ERP Abbreviation of EFFECTIVE RADIATED POWER.

error **1.** In calculations and measurements, the difference between a true value and an observed or calculated value. **2.** In electronic circuits, especially those of automatic control systems, the difference between a required (or reference signal) level and the actual signal level. **3.** In communications, a discrepancy between the transmitted data and the received data.

error accumulation The adding-up of maximum possible error when measurements are repeatedly made. Generally, the maximum plus-or-minus error per measurement is multiplied by the number of measurements.

error amplifier An amplifier for boosting error current or voltage.

error-checking code An error-correcting or error-detecting code.

error-correcting code An error-detecting code that, in addition to the function indicated by its name, indicates the correct code.

error-correcting telegraph A digital communications system in which an improbable or incorrect character is not accepted. In the event that such a character is received, the receiver instructs the transmitter to send that character again.

error correction **1.** The restoration of mutilated, corrupted, or missing data in a digital system, especially in magnetic data storage media, such as tapes and disks. **2.** In digital communications, any scheme in which the receiver (destination) automatically eliminates (to the greatest possible extent) errors in data from a transmitter (source). For example, the destination can instruct the source to repeat questionable characters or words.

error-correction routine In computer operations, a series of programmed instructions to detect and correct errors in files. A common example is a spell-checking program for word-processed document files.

error current An error signal that is a feedback current for automatically correcting a system.

error curve A bell-shaped curve that describes the distribution of errors in measurement around a true value.

error-detecting code In computer operations, a character-coding system that ensures that an impossible combination (forbidden characters) will be generated by an error (for error detection).

error-detecting routine A computer program that detects errors by checking the validity of data.

error detection and feedback In computer operations, a system in which an error (sensed by an error-detecting code) automatically generates a request to repeat the suspect signal.

error detector A sensor that responds to an error signal by delivering a signal proportional to the error.

error diagnostics As performed by a compiler, detecting and indicating the presence of errors in source language statements.

error interrupt A computer program halt caused by a software or hardware error and accompanied by a display of what has happened.

error list As produced by a compiler, a list of source language statement faults.

error message During a computer program run, a statement (displayed on a peripheral) of what is in error.

error of measurement The positive or negative difference between the value of an actual measurement and the true (or most probable) value.

error range For a data item, the range of values over which it will cause an error.

error rate In data transmission, the ratio of errors transmitted to the data transmitted.

error-rate damping Damping that involves adding to an error signal another signal that is proportional in rate of change.

error ratio **1.** In a received message, the number of incorrect characters divided by the total number of characters. Can be represented as a fraction between 0 and 1 or as a percentage by multiplying the fraction by 100. **2.** A measure of distortion for digital signal communications. The number of inaccurately received bits divided by the total number of received bits.

error routine A computer program segment that is input when an error is detected so that an appropriate action is taken (correct the error, repeat the process, etc.).

error-sensing circuit A circuit that samples the output current or voltage of a power supply, amplifier, or control system, compares this output with a standard value, and delivers a feedback (correction) signal whose amplitude is proportional to the difference (error).

error signal In a servo system, an output signal whose value is proportional to the difference between the actual operating quantity of the system and a standard reference quantity. The signal is fed back to the input of the system for automatic correction.

error tape In data processing, a record tape designed and used for storing errors for subsequent study.

error voltage An error signal that is a feedback voltage for automatically correcting a system.

Es Symbol for EINSTEINIUM.

Esaki diode See TUNNEL DIODE.

E scope See E DISPLAY.

escape character In computer operations, a character indicating that the next character belongs in a new group.

escapement A (usually oscillating) mechanical or electromechanical device that stores energy (often in a spiral spring) on one swing, and returns that energy on the next swing. Such a mechanism advances a shaft progressively in a clock or watch, and in some control equipment.

escape velocity **1.** The minimum velocity (about 25,000 miles per hour or seven miles per second) required for a space vehicle to completely escape the gravitational field of the earth. **2.** The minimum velocity required for a space vehicle to completely escape the gravitational field of a planet or star. **3.** The minimum velocity required for an electron to escape the electrical influence of an atomic nucleus.

escutcheon A usually decorative plate that frames an opening or covers a panel in a piece of equipment (e.g., the escutcheon of a radio tuning dial).

ESD Abbreviation of ENERGY-STORAGE DEVICE.

ESG Abbreviation of *electronic sweep generator.*

Esnault-Pelterie formula A formula for approximately calculating the inductance of a single-layer solenoidal coil:

$$L = 0.1008(a^2n^2)/(s + 0.92a)$$

where L is the coil inductance in microhenrys, a is the radius of the coil in inches, s is the length of the coil in inches, and n is the number of turns in the winding. The formula is accurate to 0.1 percent for all values of $2a/s$ between 0.2 and 1.5.

ESS Abbreviation of *electronic switching system.*

EST Abbreviation of EASTERN STANDARD TIME.

esu Abbreviation of ELECTROSTATIC UNIT(S).

ET Abbreviation of EPHEMERIS TIME.

ETC Abbreviation of *electronic temperature control.*

etchant Any substance such as cupric chloride, ferrous chloride, or hydrochloric acid, used in etching.

etched circuit A circuit produced by etching the metallic coating of a substrate to provide the required pattern of conductors and terminals to which discrete components are soldered.

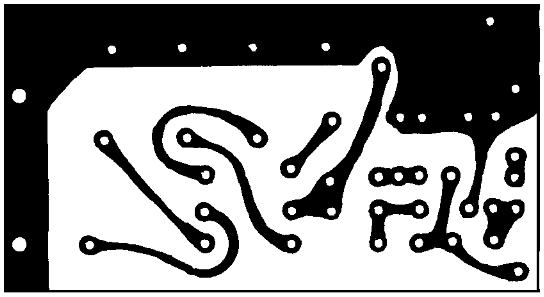

etched circuit

etch factor The ratio of the depth to the width of an etched track in an etched circuit.

etching **1.** Chemically eating away a metal to form a desired pattern, such as an etched circuit. **2.** Thinning a quartz-crystal plate by slowly eroding one or both of its faces with hydrofluoric acid to fine-tune the resonant frequency.

ET-cut crystal A piezoelectric plate cut from a quartz crystal at an angle of +66°, with respect to the z-axis. Also see CRYSTAL AXES and CRYSTAL CUTS.

ether **1.** Also called *luminiferous ether.* A nonviscous fluid once thought to fill space, convey waves (radio, light, etc.), and sustain fields. **2.** A volatile liquid occasionally used in electronics as a solvent [e.g., ethyl oxide $(C_2H_5)_2O$].

ether drift The postulated motion between a material body and the ether (see ETHER, **1**). The concept was checked by Michelson and Morley, who failed to find that the earth moves relative to the ether. This eventually led to scientific rejection of the so-called *ether theory* of the propagation of light.

ethical slave A machine, especially a smart robot, that is treated in the manner of a slave, based on the notion that a machine cannot have "feelings." Some researchers fear that the use of ethical slaves could lead to technological nightmares. For example, robots might be used as soldiers in a marauding offensive army; the commanders could rationalize that there is nothing immoral about the war because there is no loss of life on their side.

E transformer A differential transformer whose primary is wound on the center leg of an E core, the secondaries being wound on the outer legs.

Ettinghausen effect A phenomenon somewhat like the HALL EFFECT. It occurs when a metal strip, carrying current longitudinally, is placed into a magnetic field perpendicular to the plane of the strip: corresponding points on opposite edges of the strip exhibit different temperatures.

Eu Symbol for EUROPIUM.

eudiometer **1.** An instrument for measuring the amount of oxygen in the air. **2.** An instrument for analyzing gases.

eureka **1.** See CONSTANTAN. **2.** The ground transponder beacon in the British rebecca-eureka radar navigational system (see REBECCA-EUREKA SYSTEM).

europium Symbol, Eu. An element of the rare-earth group. Atomic number, 63. Atomic weight, 151.96.

eutectic **1.** A form of reaction in which mixed liquids solidify when cooled. **2.** The solid substance resulting from a reaction as defined in **1**.

eutectic alloy A metallic alloy with a specific melting point, made via eutectic process.

eutectic bond A connection between two dissimilar metals, facilitated by a third metal alloyed, via eutectic process, to the adjoining faces.

eV Symbol for ELECTRONVOLT.

evacuation The removal of air or other gases from a tube or chamber, specifically, the envelope of a vacuum tube that houses the internal elements.

evaporation **1.** A technique for electrically depositing a film of a selected metal on a metallic or nonmetallic surface. A filament of the metal to be deposited is heated by an electric current in a vacuum chamber, which makes filament particles travel to the (nearby) object to be coated, where they condense as a film. In an alternate method, a piece of the metal to be deposited is laid on or wrapped around a filament of some other metal. **2.** Electron emission by a hot cathode.

evaporation theory The theory that electrons will acquire sufficient escape velocity to leave a material when the energy acquired by (or imparted to) the electron exceeds the work function of the material. Also see WORK FUNCTION.

E vector The vector that represents the electric component of an electromagnetic wave.

even-even nucleus An atomic nucleus containing an even number of protons and an even number of neutrons. An example is the alpha particle, or helium nucleus, which contains two protons and two neutrons.

even harmonic In a complex waveform, an even-numbered multiple of the fundamental frequency. Compare ODD HARMONIC.

even line In a television picture, an even-numbered member of the 262.5 horizontal lines scanned by the spot in developing the even-line field. Compare ODD LINE.

even-line field On a television screen, the complete field obtained when the spot has traced all the even-numbered lines. Compare ODD-LINE FIELD.

even parity check A check to verify the presence of an even number of ones or zeros in a group of bits.

event An occurrence that affects the state of a computer file.

event counter Any device that measures the number of specified events taking place within a certain interval of time.

evolution Extracting a root of a number (e.g., square root, cube root, etc.).

E wave In microwave operations, the transverse magnetic (TM) wave. Also see WAVEGUIDE MODES.

EWR Abbreviation of EARLY-WARNING RADAR.

EWS Abbreviation of *early-warning system.*

E_X **1.** Symbol for voltage drop across a reactance. **2.** Symbol for EXCITATION ENERGY.

exa- Symbol, E. A prefix meaning 10^{18} (International System of Units).

exalted-carrier reception In radio reception, overcoming the effects of selective fading by maintaining the carrier at a high amplitude. This is accomplished before demodulation by removing the carrier from an amplitude-modulated or phase-modulated signal, amplifying it, and reinserting it at a higher amplitude with the sidebands.

exc **1.** Abbreviation of EXCITER. **2.** Abbreviation of EXCITATION.

except gate A logic gate that delivers an output pulse when an input pulse is present at one or more of a set of input terminals, and absent from one or more of another set of input terminals. Also called *exclusive-OR element.*

excess charge The amount of overcharge for a storage battery.

excess conduction In a semiconductor, current conduction by excess electrons.

excess electron **1.** An electron that, when introduced into an atom, results in a negative ion. **2.** An electron resulting from the addition of a donor impurity to a semiconductor substance.

excess meter A meter that integrates the amount of power in excess of some predetermined level.

excess minority carriers The number of minority carriers in excess of the normal equilibrium number in a semiconductor material.

excess modified index of refraction Symbol, M. For waves transmitted through a refracting medium, a modified index of refraction greater than unity.

excess noise Electrical noise caused by current in a semiconductor material.

excess sound pressure Unit, dyne/cm^2. In a medium conducting sound waves, the quantity $P_i - P_s$, where P_i is total instantaneous pressure at a given point in the medium, and P_s is static pressure in the absence of the sound waves.

excess-three code A computer code derived from binary notation by adding binary three (i.e., 0011) to each four-bit group. Thus, decimal seven is 1010 in the code (it is 0111 in binary). Unlike the binary representation for zero, the excess-three representation (0011) contains two ones, a feature that distinguishes actual zero from a machine fault.

exchange **1.** To reverse the contents of two memory banks. For example, if the memory banks are called A and B, an exchange is the placing of the contents of memory A into memory B, and the placing of the contents of memory B into memory A. The original contents are removed. **2.** A two-way sequence of data transmissions. **3.** A designated location in a telephone circuit.

exchange line A telephone line.

exciplex In a laser, a method of adjusting the color by means of chemical reactions in organic dyes.

excitant The electrolyte in a voltaic cell.

excitation **1.** Supplying input-signal driving current, driving power, or driving voltage. **2.** Input-signal driving current, driving power, or driving voltage.

excitation anode In a mercury-pool tube, an auxiliary anode whose operation maintains the cathode spot when no output current is being drawn from the tube.

excitation current **1.** Input-electrode current in an excited transistor amplifier. **2.** Grid current in an excited vacuum-tube amplifier. **3.** Current flowing in the circuit of the excitation anode of a mercury-pool tube. **4.** Current flowing in the exciter circuit of an alternator. **5.** Shunt-field current in a motor.

excitation energy **1.** Symbol, E_X. In artificial transmutation, the energy of a nucleus when protons of less than maximum energy have been emitted from the atom. **2.** Electrical energy required by a transducer.

excitation purity In color television, complete saturation of a hue (i.e., there is no contamination by other colors, and the saturated hue is distributed uniformly).

excitation voltage **1.** The signal voltage that achieves, or is required for, excitation (see EXCITATION, **1**). **2.** The value of driving voltage.

excitator An electrical discharger.

excited atom An atom in which one or more electrons have been pushed out of their normal orbits into higher ones by energy applied from the outside.

excited-field speaker A dynamic speaker in which the magnetic field is provided, not by a permanent magnet, but by direct current flowing through a large coil of wire wound around the speaker core. The coil usually acts as a filter choke in the power supply of the attendant amplifier or receiver.

excited state In artificial transmutation, the state of the nucleus when protons of less-than-maximum energy have been emitted from the atom. The energy of the protons, in this instance, is greater than the ground state.

exciter **1.** An amplifier or oscillator (or a system of such units) that supplies the input (driving) signal to the output amplifier in a radio transmitter or similar device. **2.** A small direct-current (dc) generator that supplies direct current to the field winding of an alternating-current (ac) generator. **3.** See INDUCTION COIL.

exciter lamp **1.** A concentrated-filament, high-intensity incandescent lamp used in sound-on-film recording and reproduction and in some types of electromechanical television. **2.** In a facsimile transmitter, the lamp illuminating what is being scanned.

exciter relay In an electromechanical generator, the relay that activates the direct-current (dc) field excitation during machine startup.

exciter response **1.** A change in the exciter voltage of a motor when the field-circuit resistance changes. **2.** A change in the operating conditions of a radio frequency exciter, as a result of a change in the impedance at the input of the final amplifier.

exciter unit See EXCITER.

exciting current **1.** The output current produced by the exciter of a generator (see EXCITER, **2**). **2.** The field current of a dynamo-type generator. **3.** Primary current in an unloaded transformer.

exciting power **1.** The output power produced by an exciter. **2.** The input-signal power required for full output from a power amplifier. Also called DRIVING POWER.

exciting voltage **1.** Input-signal voltage. **2.** The input-signal-voltage amplitude required for full rated output from a power amplifier. Also called DRIVING VOLTAGE. **3.** The output voltage produced by an exciter.

exciton In a semiconductor or dielectric, a bound electron-hole pair.

excitron A mercury-pool rectifier whose arc is initiated mechanically (e.g., by means of a magnetic plunger in the tube).

exclusion principle The rule that only one particle of a particular kind can occupy a given quantum state at one time.

exclusive-NOR A logic function where the output is 1 if both inputs are 1 or both are 0 (same). The output is 0 if one input is 0 and the other is 1 (different). Compare EXCLUSIVE OR.

exclusive-OR A logic function in which the output is 1 when the two inputs are different, and is 0 when the two inputs are the same.

excursion **1.** A change in the value of a quantity in a given direction. **2.** In an oscillatory system, a body's moving away from the point of equilibrium or mean position.

execution A computer's performance of the operations required by an instruction.

execution time The length of time required for a computer to complete a designated operation.

executive routine In computer operations, a program that controls and processes other routines. Also called *monitor program.* Compare MONITOR SYSTEM.

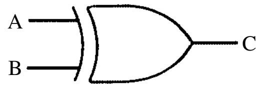

A	B	C
0	0	0
0	1	1
1	0	1
1	1	0

exclusive-OR

exhaust analyzer An instrument for examining the exhaust fumes of an internal combustion engine, to measure the presence of noxious materials and to evaluate air-to-fuel ratio and combustion efficiency.

exhaustion See EVACUATION.

exit **1.** In computer operations, the last instruction in a program or program segment, often taking a subroutine back to the main program. **2.** To leave a computer application, routine or subroutine.

exoskeleton A robot that resembles a suit of armor, and that greatly magnifies the force of physical movements. A human operator occupies the interior. Thus, for example, the operator might use the machine to throw a football 2500 yards, or to run 50 miles an hour, or to smash through walls. Primarily a tool of science fiction writers, this machine is within the scope of current robotic technology.

exosphere The extreme outer layer of the earth's atmosphere.

exothermic Pertaining to a chemical or electrochemical reaction in which heat is given off. Compare ENDOTHERMIC.

exothermic reaction In a chemical reaction, the production of positive reaction energy (i.e., kinetic energy is gained). Compare ENDOTHERMIC REACTION.

exp **1.** Symbol for EXPONENTIAL. **2.** Abbreviation of EXPERIMENT(AL).

expand **1.** In communications, to increase the bandwidth of a signal, restoring it to normal bandwidth after it has been compressed. **2.** In communications, to increase the dynamic range of a signal. **3.** In computer operations, to restore a file to full or normal size after it has been compressed. **4.** To widen the scale of a meter. **5.** To widen (or magnify a portion of) the trace of an oscilloscope beam. Compare COMPRESS.

expandable Capable of being built up into larger circuits or systems.

expandable gate In digital logic, a gate that can be provided with an unlimited number of input lines by electrical interconnection with other gates.

expanded memory In personal computer systems, memory beyond the basic 640 kilobytes (640 kb), up to one megabyte (1 MB). This memory resides in integrated circuits (ICs) in the computer, and is normally volatile (i.e., it is not retained when power is removed). Compare EXTENDED MEMORY.

expanded-scale meter A meter having a scale designed to display a narrow range of values. Such a meter used for monitoring the 117-V power line might have a scale reading 100 to 140 V, instead of a conventional scale beginning at zero.

expanded sweep **1.** In an oscilloscope, speeding up the deflection of the beam during a selected portion of the trace. **2.** The circuit for the action described in **1**.

expander A circuit for increasing the dynamic range over which a signal or quantity can vary. A typical example is the volume expander, a device that greatly increases the amplitude of strong signals while weakening, or having no effect on, signals of low amplitude.

expansion **1.** In communications, the process of increasing the bandwidth of a signal, restoring it to normal bandwidth after it has been compressed. **2.** In communications, a process in which stronger components are amplified more than weak ones, restoring a signal to its normal dynamic range after it has been compressed. **3.** In computer operations, the restoration of a file to full or normal size after it has been compressed. **4.** The widening of a meter scale. **5.** The widening or magnification of an oscilloscope trace. Compare COMPRESS.

expansion chamber A cloud chamber for viewing the paths of radioactive particles. It consists of a closed glass cylinder containing humid air and a piston. An electrostatic field is applied through the cylinder, the piston is pulled quickly, and the volume of the chamber expands. The temperature inside falls below the dew point, a cloud is formed, and droplets of water condense on ions, making their paths visible for observation or photography through the cylinder walls.

expansion ratio In communications, the inverse of COMPRESSION RATIO.

expansion time For an expansion chamber, the interval during which expansion occurs. The interval is kept short to ensure that the temperature will drop low enough for vapor condensation, and to minimize the possibility of continuing gas motion distorting the track of a particle.

expectation In probability theory, the middle value (average or mean) of a random variable.

expendable A component or system that, for economy, is best discarded instead of repaired when it fails. Also called DISPOSABLE COMPONENT.

experiment One or a series of carefully planned tests carried out under controlled conditions to obtain data or to check performance.

experimental chassis See ELECTRONIC CHASSIS.

experimental model A prototype of an electronic circuit or device, produced solely for operational tests or as a model against which theory and design can be checked.

experimental service A special, nonamateur radio service intended for on-the-air testing of new methods and equipment.

experimental station A station specially licensed to operate on specific frequencies in the experimental service.

expert system Also called *rule-based system.* A form of artificial intelligence (AI) that allows a computer or smart robot to act as a highly talented specialist in a specific field. An example is the use of a computer to help a physician diagnose a complex disease. A smart robot might be used as a surgical assistant.

exploring coil A pickup coil for sensing a signal or magnetic field. Sometimes called a *sniffer.*

exploring electrode **1.** A sampling electrode sealed in a discharge tube for measuring ionization at the point of insertion. **2.** Broadly, a test probe.

explosion-proof device A device that is housed and operated so that its sparking, heating, or production of radiant energy will not cause materials in the environment to explode.

exponent A number written as a superscript indicating the power to which another number (called the *base*) is to be raised. For example, 2^2 is the square (second power) of 2; x^3 is the cube (third power) of x.

exponential **1.** A base (such as the natural number e) modified by an exponent. **2.** Related to a change in value as determined by an exponent. Thus, using increments for x in the equation $y = e^x$ produces an exponential curve.

exponential curve A curve based on powers of a number (such as for $y = e^x$). Also see EXPONENTIAL, EXPONENTIAL DECREASE, and EXPONENTIAL INCREASE.

exponential damping Damping action described by an exponential curve.

exponential decay See EXPONENTIAL DECREASE.

exponential decrease The continuous reduction in the value of a quantity, according to the equation $y = e^{-x}$, which depicts the natural decay curve.

exponential function A function, such as $f(x) = e^x$, that varies exponentially. See, for example, EXPONENTIAL DECAY, EXPONENTIAL GROWTH, and EXPONENTIAL SERIES.

exponential horn A horn of circular or rectangular cross section, whose cross-sectional area S at any point x feet along its axis is given by the formula $S = S_0e^{mx}$, where S_0 is the cross-sectional area at the throat, e is the natural logarithm base (approximately 2.71828), and *m* is the horn flaring constant.

exponential increase The continuous increase of a quantity, according to the equation $y = e^x$, which depicts the natural growth curve.

exponential line A transmission line whose characteristic impedance varies exponentially with its electrical length.

exponential quantity A quantity involving an exponential (e.g., $3e^x$).

exponential series A mathematical series based on exponential expressions. Example: $e^x = 1 + x + (x^2/2!) + (x^3/3!) + (x^4/4!) + \ldots$

exponential sweep In cathode-ray-tube (CRT) devices, such as oscilloscopes, a beam sweep that starts fast and slows exponentially.

exponential transmission line See EXPONENTIAL LINE.

exponential waveform Any waveform in which the rate of change in the amplitude is directly or inversely proportional to the instantaneous amplitude. The absolute value of the derivative of such a waveform is equal to the absolute value of the instantaneous amplitude, multiplied by a constant that depends on the amplitude units.

exposure **1.** The total amount of radiation received in a given area, or by a given sample, or by a person, over a specified length of time. **2.** The extent to which a photographic film has been darkened or otherwise modified by visible light, infrared, ultraviolet, or X rays.

exposure meter **1.** A usually simple instrument for measuring light intensity—especially for photographic purposes. A common form consists of a self-generating photocell connected to a direct-current microammeter. **2.** A device that indicates the amount of ionizing radiation that has been received over a given period of time.

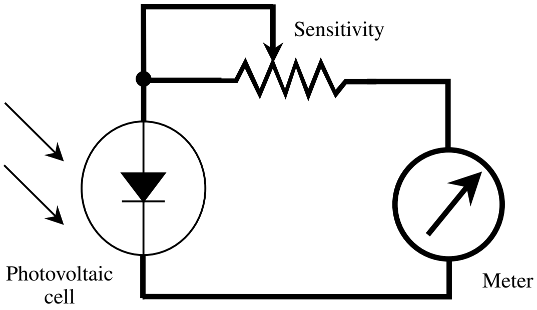

exposure meter

expression control A volume control in an electronic organ.

extended double Zepp antenna See DOUBLE-EXTENDED ZEPP ANTENNA.

extended memory In personal computer systems, memory beyond the first megabyte (1 MB). There is no limit in theory to the extent of this memory, although it is limited by current technology. Used in high-level programs—especially those involving graphical applications, artificial intelligence, or intensive calculations. Resides in integrated circuits (ICs) in the computer, and is normally volatile (i.e., it is not retained when power is removed). Compare EXPANDED MEMORY.

extended octaves Audio-frequency tones above or below the normal range of an electronic musical instrument. Special circuits must be added to make the extended octaves available.

extended play Pertaining to a recorded phonograph disc that provides a longer playing time than conventional discs of the same size and recording speed.

extender A substance added to an encapsulant to make it go further.

extensimeter See EXTENSOMETER.

extension cable A flexible, low-capacitance (usually concentric) cable for connecting part of one circuit to part of another. Extension cables are available with a variety of end connectors.

extension cord A flexible power cord having a male plug on one end and female receptacle on the other.

extension loudspeaker An auxiliary loudspeaker serving areas in which the main speakers can't be adequately heard.

extensometer An instrument used to measure small amounts of expansion, contraction, or deformation.

exterior label On a diskette or tape cartridge used for computer data storage, a written identification on the housing or cartridge, as opposed to the label, which is recorded on the diskette or tape itself.

external armature In a dynamo-type machine, an armature that rotates around the outside of the field magnets, as opposed to the usual (inside) arrangement.

external capacitor A high-value capacitor connected externally to an oscillator or sweep generator to lower its frequency.

external circuit A circuit or subcircuit connected and external to a main equipment.

external controls **1.** Control devices that are connected to, but operated away from, a main circuit. **2.** Manual or screwdriver-adjusted controls that are mounted on the panel of an equipment, as opposed to those mounted in the case or behind the panel.

external critical damping resistance The value of external resistance that must be connected to a galvanometer or other meter to produce critical damping.

external damping device A resistor or short-circuiting bar connected temporarily between the terminals of a meter to keep its movement immobile during transportation.

external feedback Negative or positive feedback through a separate path outside of and around the main circuit. Example: negative feedback through a resistance-capacitance (RC) path between the output terminals and the input terminals of an amplifier.

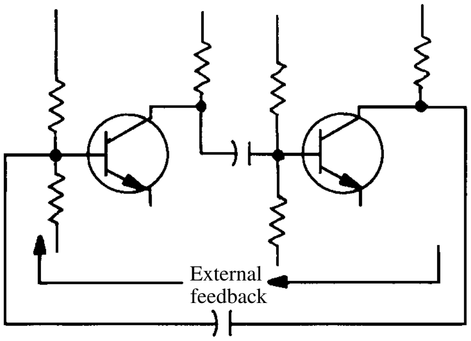

external feedback

external impedance Load impedance (i.e., an impedance connected to the output terminals of a generator or amplifier).

external load See EXTERNAL IMPEDANCE.

external loudspeaker See EXTERNAL SPEAKER.

externally caused chatter In a relay, contact chatter caused by mechanical vibration outside of the relay.

externally caused failure Failure of a circuit or component resulting from unfavorable environmental factors.

external memory In computer operations, a memory unit outside of the computer mainframe.

external power supply A power supply unit situated apart from the powered equipment. Such separation is helpful in eliminating the disturbing effects of heat, hum, and vibration associated with internal power units.

external *Q* For a microwave tube, the quantity $1/(1/Q_1 + 1/Q_2)$, where Q_1 is the loaded Q and Q_2 is the unloaded Q.

external S-meter A signal-strength meter connected to a receiver, but not installed in its panel.

external speaker A loudspeaker that doesn't share an enclosure with an amplifier, receiver, or other device that drives it. Such isolation is helpful in eliminating the undesirable effects of vibration and acoustic feedback.

external storage In computer operations, storage media (such as magnetic diskettes or tapes) that are outside of the computer.

extinction potential See DEIONIZATION POTENTIAL.

extinction voltage See DEIONIZATION POTENTIAL.

Extra-class license An amateur-radio license that conveys all available amateur operating privileges in the United States. The highest class of amateur license.

extract **1.** To remove a signal or quantity from some product containing it, or from its source. Examples: extracting a fifth harmonic from a complex signal, extracting the direct-current (dc) component from a signal containing both alternating current (ac) and dc. **2.** To derive a factor (e.g., to extract a root). **3.** To separate certain classes of information from an aggregate of information.

extract instruction In computer operations, the instruction to generate a new word by the serial arrangement of designated segments of specified words.

extractor **1.** A circuit or device for removing a signal (or a signal component) from another circuit or device. A demodulator probe, for example, extracts the modulation from a modulated signal. **2.** A device for removing used active devices from a circuit board. Such extractors can also employ heat to desolder, as well as remove the devices.

extraneous component A usually undesired inherent effect that results from the physical nature of a component or device. Examples: distributed capacitance of a coil, internal inductance of a capacitor.

extraneous emission Undesired emission from a transmitter (e.g., excessive harmonics).

extraneous response The unintended response of a circuit or device (e.g., image response in a superheterodyne communications receiver).

extraneous root In the solution of an equation derived from another equation, one or more roots that satisfy the derived equation but not the original one.

extraneous signal A superfluous and potentially interference-causing signal.

extranuclear Outside the nucleus of an atom.

extraordinary ray Of the two rays resulting from the double refraction of electromagnetic waves, the one that does not follow the usual laws of refraction. Also see X WAVE.

extraordinary wave See X WAVE.

extrapolar **1.** Outside of electrical or magnetic poles. **2.** Not between electrical or magnetic poles.

extrapolation Estimation of values beyond the range of available data. An example is the extension of a curve beyond its final plotted point to determine a value for a variable. There is always some margin for error, which increases as the process is extended further and further beyond the range of actual data values. Commonly done by computers (e.g., prediction of the probable path of a hurricane 12, 24, 36, and 48 hours in advance).

extrared See INFRARED.

extraviolet See ULTRAVIOLET.

extreme **1.** The lowest or highest value of a quantity. **2.** The lowest or highest point of the dependent variable in the range of a function. **3.** An unusual value for a parameter (e.g., temperature extreme and pressure extreme).

extremely high frequency Abbreviation, EHF. A frequency near the upper limit of the radio-frequency spectrum—especially one in the 30- to 300-GHz band.

extremely low frequency **1.** Abbreviation, ELF. Pertaining to a signal or current within the audio-frequency (AF) range, but not used for audio applications. **2.** Electromagnetic emissions from a cathode-ray tube (CRT), resulting from the currents in the deflecting coils and/or electron beam. **3.** The 60-Hz electromagnetic field generated by utility power lines and wiring.

extrinsic base-resistance/collector-capacitance product Units: seconds, milliseconds, and microseconds. For a bipolar transistor, the product R_BC_C, where R_B is the base resistance and C_C is the collector capacitance. This product is a time constant that determines the high-frequency operating limit of the transistor.

extrinsic conductance For a material, the conductance resulting from impurities or such external factors as environmental conditions.

extrinsic properties For a semiconductor material, properties resulting from doping (e.g., altered resistivity or majority/minority carrier differentiation). Also see EXTRINSIC SEMICONDUCTOR.

extrinsic semiconductor A semiconductor material, such as germanium or silicon, to which a controlled amount of a suitable impurity material has been added to give the semiconductor a desired resistivity and polarity. Compare INTRINSIC SEMICONDUCTOR.

extrinsic transconductance Symbol, g_m. For a bipolar transistor, the first derivative of collector current, with respect to base-emitter voltage. It is the ratio of a small change in collector current (dI_C) to the small change in base-emitter voltage (dV_{BE}) that produced it, collector voltage being constant; $g_m = dI_C/dV_{BE}$.

extrusion The process of forming a material such as metal or plastic, by forcing it through dies. Many pieces of electronic hardware are mass produced in this manner. Examples are insulating rods and tubes, metal cans, and metal tubing.

eye-in-hand system In robotics, a scheme that uses a vision system, servo, and microprocessor to precisely manipulate an END EFFECTOR, such as a robot gripper ("hand"). A camera (the "eye") is contained in the end effector itself, and is designed to work at close range (i.e., approximately one meter down to a fraction of a millimeter).

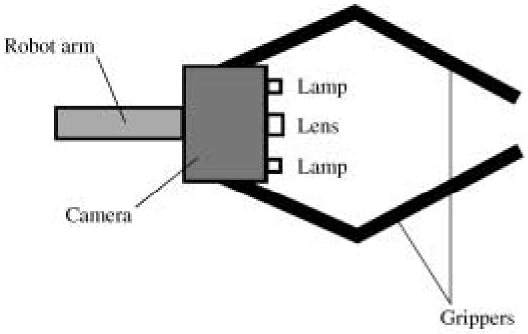

eye-in-hand system

eyelet connection A connection made by fastening conductors together with an eyelet or by soldering leads or pigtails to an eyelet.

eyepiece A small lens system for viewing an oscilloscope screen through a camera setup.

E_Z Symbol for *voltage drop across an impedance.*

E zone A portion of the earth including most of the eastern hemisphere. When propagation forecasts are made, this region is one of three longitude zones specifie

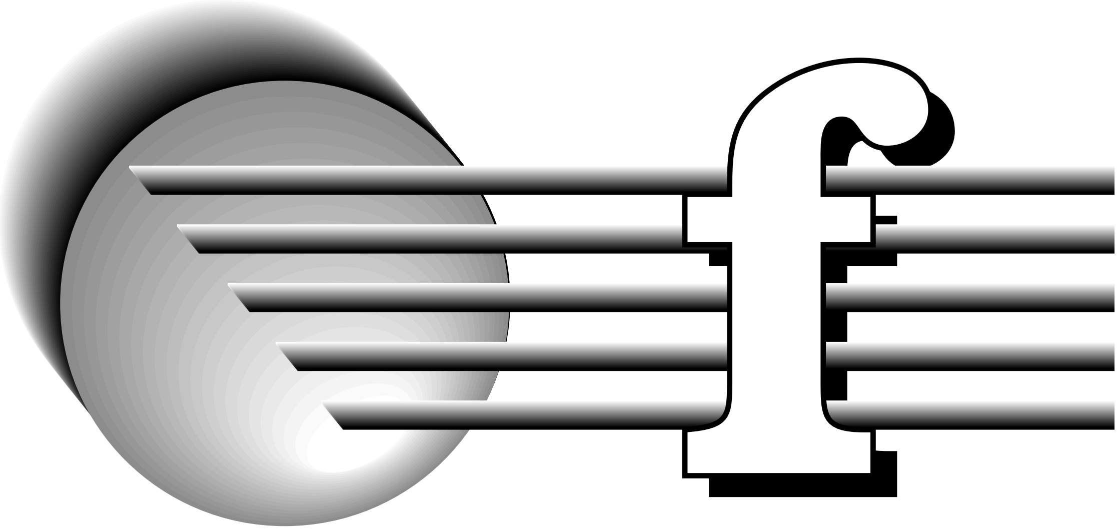

F **1.** Symbol for FORCE. **2.** Symbol for FLUORINE. **3.** Abbreviation of FAHRENHEIT. **4.** Abbreviation of FARAD. **5.** Abbreviation of FERMI. **6.** Symbol for FOCAL LENGTH. **7.** Symbol for FUSE. **8.** Symbol for FARADAY CONSTANT.

f **1.** Abbreviation of FEMTO. **2.** Symbol for FREQUENCY. **3.** Symbol for FUNCTION.

F_0 Symbol for DAMPING FACTOR.

F1 layer The lower part of the ionosphere's F REGION. Also called *F1 region.*

F2 layer The upper part of the ionosphere's F REGION. Also called *F2 region.*

fA Abbreviation of FEMTOAMPERE.

fabrication tolerance The amount of variation that can be tolerated in the manufacture of components.

Fabry-Perot interferometer A resonant cavity, often used with lasers, that has mirrors at each end; the interferometer produces the optical equivalent of standing waves.

face **1.** A flat crystal surface whose orientation can be expressed as its position relative to other faces. **2.** The viewer's side of a screen. **3.** The scale part of a meter.

face material In a tape recorder, the plastic used to coat the face of a head.

face-parallel cut See Y-CUT CRYSTAL.

face-perpendicular cut See X-CUT CRYSTAL.

face side The side of pressure-sensitive insulating tape that is coated with adhesive.

facom A radionavigation system that operates by means of phase comparison at low frequencies. Effective over long distances and under poor conditions.

facsimile Also called *fax.* The transmission and reception, through the medium of radio or by wire, of permanent pictures, writing, and other graphic material.

facsimile receiver The complete device or system that selects, amplifies, and demodulates a picture signal received up from the air, wires, or cable, and uses the elements of this signal to reproduce the picture. Also see FACSIMILE.

facsimile recorder The machine that puts a transmitted facsimile image on paper.

facsimile transmitter The complete device or system that generates signals depicting graphic material (pictures, writing, printing, etc.) and sends them to a distant point via cable, wire lines, or radio for subsequent reproduction. Also see FACSIMILE.

factor **1.** A data element that is an operand in an arithmetic operation. **2.** To find the two or more numbers whose product is the number being factored. **3.** One of two or more numbers whose product is the number being factored.

factorial Symbol, !. For an integer n, the product of all positive integers up to and including n. Thus, $5! = 1 \times 2 \times 3 \times 4 \times 5 = 120$. The term $n!$ is read "*n* factorial."

factor of merit See FIGURE OF MERIT.

factor of safety See SAFETY FACTOR.

fade in To gradually increase an audio or video signal—especially for recording.

fade out **1.** To gradually decrease an audio or video signal—especially for recording. **2.** The complete disappearance of a radio communications signal. See FADING.

fader **1.** In sound amplification systems, an attenuator circuit that enables the operator to fade out one signal and fade in another. Ordinarily, a fader does not provide mixing action. **2.** In an au-

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

tomotive high-fidelity sound system, a control that adjusts the volume of the front speaker set and rear speaker set. Usually, this is a single knob; increasing the front volume decreases the rear volume, and vice-versa.

fading **1.** Repeated increases and decreases of radio communications signal amplitude at the reception point. Fading of a radio wave results from multiple propagation paths from transmitter to receiver, causing the received waves to arrive in constantly varying phase relationships. **2.** In a cell or battery power supply, a gradual loss of power-delivering capability that occurs during the discharge cycle.

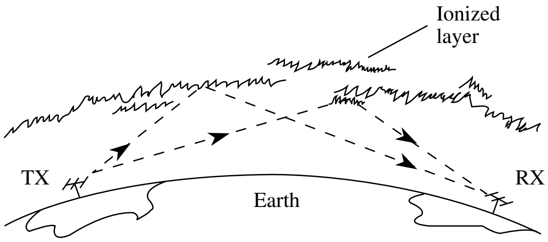

fading

factory automation The use of robotic devices, to the greatest extent possible, as laborers in manufacturing plants.

Fahnestock clip A flat, sheet-metal spring clip for holding a wire (usually in a temporary breadboard setup).

Fahrenheit scale A temperature scale on which the freezing point of water is 32 degrees, and the boiling point of water is 212 degrees. Compare ABSOLUTE SCALE and CELSIUS SCALE.

failsafe Pertaining to devices or circuits that, upon failure, cause no damage or serious malfunction.

failsoft In a computer, a system in which operation is maintained—even in the event of partial failure. Efficiency is reduced but the computer does not completely shut down.

failure The condition wherein a circuit, system, or device is not operating correctly.

failure analysis **1.** The process of determining the failure rate for a component, system, or device. **2.** The process of determining the cause of a failure.

failure mode The particular way in which a failure of equipment or a method occurs.

failure unit A unit of machine or device failure: one failure per billion (10^9) hours of operation.

fall-in The time when synchronous speed is attained in a synchronous motor.

falling characteristic A NEGATIVE RESISTANCE characteristic.

fall time **1.** Decay time. **2.** The time required for the amplitude of a pulse to decrease from 90 percent to 10 percent of its peak amplitude. Compare RISE TIME.

false add A logic add (i.e., addition without carries).

false alarm **1.** Improper operation of an electronic security system, resulting in actuation of the device when no breach of security has occurred. **2.** In radar, the presence of a false echo that causes the attendant circuits or personnel to act as though an enemy target is present.

false error A condition in which a computer system erroneously signals the existence of an error.

false precision See MISLEADING PRECISION.

false retrieval The incorrect specification of criteria for information to be selected for retrieval so that an unwanted item of data is selected. Also called *false drop.*

family Any group of components, circuits, ratings, or characteristics, classed together because of some common or analogous feature or application. Examples: IC family, family of curves, and family of equations.

family of curves A group of curves plotted on the same axes, that depict the performance of a circuit or device at several levels of a third parameter (e.g., curves showing transistor collector current vs. collector voltage for several levels of base current).

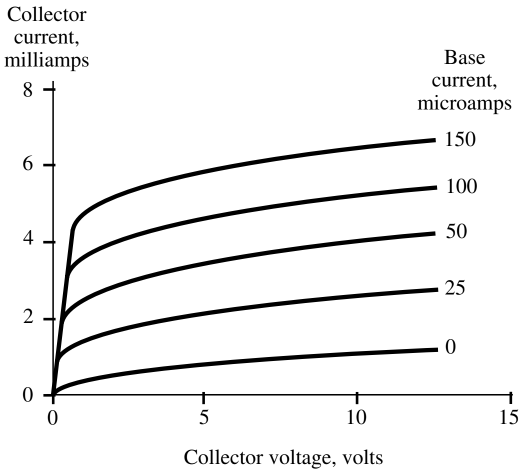

family of curves

fan antenna See DOUBLE-V ANTENNA.

fan-in **1.** A number of inputs entering a common input terminal. **2.** In digital computer operations, the number of inputs that can be accommodated by a logic circuit. Compare FAN-OUT.

fan-in circuit A circuit having a number of input lines entering a common input point. Also see FAN-IN, **1**.

fan-out **1.** One common output terminal feeding a number of output lines. **2.** In digital computer operations, the number of outputs that can be fed by a logic circuit. Compare FAN-IN.

fan-out circuit A circuit in which a number of output lines leave a common output terminal. Also see FANOUT, **1**.

fantasy robot A robot as portrayed in science fiction or fantasy. Such robots are usually, but not always, *androids* (that is, they have humanoid form). Some fictional robot characteristics are later realized. Science fiction can even give roboticists ideas for future designs.

farad (Michael Faraday, 1791–1867) Abbreviation, F. The basic unit of capacitance. A capacitor has a capacitance of 1 F when a charge of 1 volt per second across the capacitor produces a current of 1 ampere through it.

faraday An electrical quantity approximately equal to 9.65×10^4 coulombs; it is the quantity of electricity required in electrolysis to free 1 gram atomic weight of a univalent element. The equivalent, and preferred, unit is the COULOMB. Also called *Faraday constant.*

Faraday cage See ELECTROSTATIC SCREEN.

Faraday constant See FARADAY.

Faraday cylinder A cylindrical metal shield placed around electrical equipment or circuits to prevent electromagnetic fields from affecting the equipment or circuits.

Faraday rotation **1.** A change in the polarization of an electromagnetic wave as it passes through a medium. **2.** Constant and random fluctuations in the polarization of radio signals as they pass through the earth's ionosphere. This phenomenon affects signals having linear polarization at frequencies well below 10 GHz. It has little or no effect on circularly polarized signals, or on signals at frequencies above 10 GHz. It can be a problem with some satellite communications links.

Faraday effect See MAGNETO-OPTICAL ROTATION.

Faraday's disk dynamo See DISK DYNAMO.

Faraday shield See ELECTROSTATIC SCREEN.

Faraday's law The voltage induced in a conductor moving in a magnetic field is proportional to the rate at which the conductor cuts the magnetic lines of flux.

Faraday's laws of electrolysis **1.** In electrolysis, the mass of a substance liberated from solution is proportional to the strength and duration of the current. **2.** For different substances liberated by the same current in a certain time, the masses are proportional to the electrochemical equivalents of the substances. Also see ELECTROCHEMICAL EQUIVALENT; ELECTROLYSIS, **1**; and ELECTROLYTE.

faradic current The lopsided alternating current produced by an induction coil.

faradmeter An alternate term for MICROFARAD METER.

far field **1.** The region beyond the near field of an antenna (see NEAR FIELD, **1**). **2.** The region beyond the near field of a loudspeaker (see NEAR FIELD, **2**).

far infrared Also called far IR. The lower-frequency portion of the infrared (IR) spectrum.

farming robot A robot that performs labor in an agricultural setting. Examples: fruit-picking robot, field-watering robot, cow-milking robot. Such machines are usually overseen by a central computer, via which the farmer can monitor and control the robots' actions.

far zone See FAR FIELD, **1**, **2**.

fast access storage In a computer memory, the section from which information can be most quickly accessed, depending on the relative speed of other system devices.

fast-break, fast-make relay A relay that opens and closes rapidly.

fast-break, slow-make relay A relay that opens rapidly and closes slowly.

fast charge The rapid charging of a rechargeable cell or battery, particularly of a nickel-cadmium (NICAD) or nickel-metal-hydride (NiMH) device.

fast diode See COMPUTER DIODE.

fast drift The rapid change of a quantity or setting, usually in one direction. Compare SLOW DRIFT.

fast-food robot A robot that prepares and/or serves items in a fast-food establishment. Robots have not yet gained wide acceptance in this setting, mainly because of technical difficulties and prohibitive cost. Another problem is that some customers are put off or intimidated by robots.

fast-forward Abbreviation, FF or FFWD. In a tape recorder, a mechanism for running the tape through the machine rapidly.

fast groove The informationless groove between tracks on a disk recording.

fast-make/fast-break relay A relay that closes and opens rapidly.

fast-make, slow-break relay A relay that closes rapidly and opens slowly.

fast-forward playback In a videotape recorder, the playing back of the tape at faster than real-life speed. It allows the viewer to move quickly ahead in a program, and also to watch the images so as to know when to resume normal replay.

fast-reverse playback In a videotape recorder, the playing back of the tape rapidly backwards. It allows the viewer to move quickly back in a program, and also to watch the images so as to know when to resume normal replay.

fast time constant **1.** The property of responding quickly to changes in input parameters. **2.** In radar, a method of defeating attempts at jamming by modification of the receiving circuitry.

fathometer See ACOUSTIC DEPTH FINDER.

fathom **1.** To measure the depth of a body of water, as in the use of sonar for this purpose. **2.** A unit of length (distance) equal to six feet.

fatigue **1.** The degradation of the performance of circuits or materials with time. **2.** The tendency of bodies and materials to weaken, deform, or fracture under repeated strain.

fault **1.** A defective point or region in a circuit or device. **2.** A failure in a circuit or device.

fault current **1.** A momentary current surge. **2.** A leakage current.

fault finder A troubleshooting instrument or device (e.g., a multimeter).

fault resilience **1.** A design scheme for an electronic or computer device or system so that if a component or circuit fails, the system will continue to operate, although perhaps at reduced efficiency. The operator is notified of the problem so that it can be repaired with minimal downtime. **2.** In a computer system, the property of being as nearly sabotage-proof as possible.

fault tolerance Total redundancy in an electronic or computer system so that if a component or circuit fails, the system will continue to function at full efficiency. Every component has a backup that automatically takes over in case of failure. The operator is notified of the problem, so the defective part or circuit can be replaced while its backup keeps the circuit working continuously at 100-percent capacity.

Faure plate A storage battery plate consisting of a lead grid containing a chemical electrolytic paste.

fax Abbreviation of FACSIMILE.

fc Abbreviation of FOOT-CANDLE.

f_c Abbreviation of CARRIER FREQUENCY.

FCC See FEDERAL COMMUNICATIONS COMMISSION.

f_{co} Abbreviation of CUTOFF FREQUENCY.

F connector A type of antenna feedline connection especially common on television receivers and videocassette recorders.

F display See F SCAN.

FDM Abbreviation of *frequency-division multiplex.*

FE Abbreviation of FERROELECTRIC. See FERROELECTRICITY.

Fe Symbol for IRON.

feasibility study The procedures for evaluating the potential gains in applying a computer system to a job or to an organization's process, or in modifying or replacing an existing system.

FEB Abbreviation of FUNCTIONAL ELECTRONIC BLOCK.

Federal Communications Commission Abbreviation, FCC. Established in 1934, the U.S. Government agency that regulates electronic communications. The FCC succeeded the Federal Radio Commission (FRC), which was established in 1927; the FRC succeeded the Radio Division of the Bureau of Navigation in the Department of Commerce, whose jurisdiction over radio began in 1912.

feed **1.** To supply power or a signal to a circuit or device. **2.** The method of supplying such a signal or power. See, for example, PARALLEL FEED and SERIES FEED. **3.** To cause data to be entered into a computer for processing.

feedback **1.** The transmission of current or voltage from the output of a circuit or device back to the input, where it interacts with the input signal to modify operation of the device. Feedback is positive when it is in phase with the input, and is negative when it is out of phase. **2.** To input the result at one point in a series of operations to another point; the method allows a system to monitor its actions and make necessary corrections.

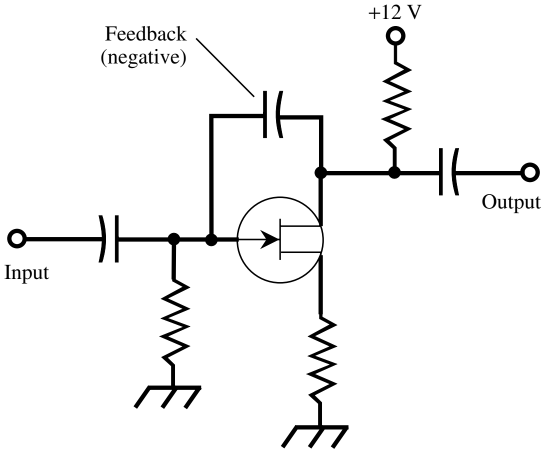

feedback, 1.

feedback amplifier **1.** An amplifier whose performance (especially frequency response) is modified by means of positive, negative or both positive and negative feedback. **2.** An amplifier placed in the feedback path of another circuit to increase the amplitude of feedback.

feedback attenuation **1.** In an operational-amplifier circuit, the attenuation in the voltage from output to input. **2.** In an audio-frequency or radio-frequency amplifier circuit, the reduction of feedback by electronic means.

feedback bridge A bridge circuit in the feedback channel of an amplifier or oscillator.

feedback capacitance **1.** A capacitance through which feedback current is coupled from the output to the input of a circuit or system. **2.** The interelectrode capacitance of a vacuum tube.

feedback control **1.** The variable component (potentiometer or variable capacitor) used to adjust the level of feedback current or voltage. **2.** The control of circuit performance by means of feedback.

feedback cutter A device used for the purpose of cutting grooves in phonograph disks. Feedback is used to provide a flat frequency response.

feedback factor For a feedback amplifier, the quantity $1 - bA$, where A is the open-loop gain of the amplifier and b is the FEEDBACK RATIO.

feedback input current In a feedback network, the current drawn by the feedback input. This current affects the design parameters of a network.

feedback loop The part of a circuit that provides controlled feedback in an operational-amplifier circuit.

feedback oscillator A circuit in which oscillation is obtained by feeding a portion of the output of an active amplifying device back to the input circuit by inductive coupling. Also called *tickler oscillator.*

feedback path A path over which feedback, either positive or negative, can occur in a circuit. The feedback can be intentionally produced, or it can be undesirable.

feedback percentage Symbol, n. In a feedback circuit, the percentage of output voltage that is fed back; $n = 100V_f/V_o$, where V_f is the feedback voltage and V_o is the open-loop output voltage. Compare FEEDBACK RATIO.

feedback ratio For a feedback system, the ratio V_f/V_o, where V_f is the voltage that is fed back and V_o is the open-loop output voltage of the system.

feedback rectifier See DIODE FEEDBACK RECTIFIER.

feedback regulator In a controlled-feedback circuit, the device that determines the amount of feedback.

feedback resistance **1.** The internal base resistance of a point-contact transistor. **2.** The resistance in a feedback loop.

feedback transfer function The transfer function of a feedback loop exclusively.

feedback winding A special winding on a magnetic amplifier or saturable reactor, for the introduction of feedback currents.

feeder **1.** A conductor or set of conductors that carries electric power from one point to another. **2.** The transmission line connecting a transmitter to an antenna.

feeder cable **1.** A communication cable running in a primary route from a central station (or in a secondary route from a main feeder cable) as a means of making connections to distribution cables. **2.** In a cable television system, the cable carrying transmission from the head end to the trunk amplifier. Also called TRUNK CABLE.

feeder loss Loss of energy resulting from resistance in, or radiation from, feeder lines.

feeding In character recognition, a system in which documents go into the transport of a character reader at a steady, specified rate.

feed pitch The distance between feed holes.

feed reel The tape supply reel of a tape recorder.

feedthrough **1.** The usually undesirable transmission of a signal through a circuit without being processed by the circuit, because of unavoidable capacitive coupling, for example. **2.** Contraction of FEEDTHROUGH COMPONENT.

feedthrough capacitor A capacitor whose design is like that of a feedthrough terminal; it is mounted in a hole in a chassis. The center screw or wire is the "high" terminal of the capacitor, to which connections can be made above or below the chassis. The body of the device is the "low" terminal of the capacitor; it is soldered to the chassis or secured with a nut.

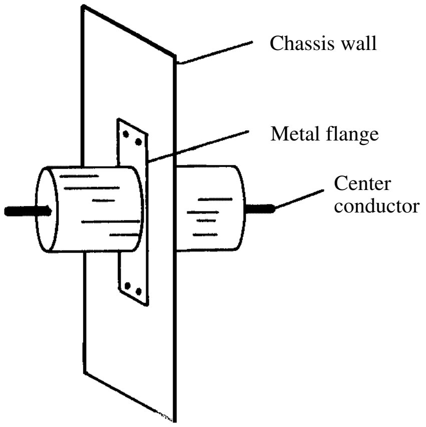

feedthrough capacitor

feedthrough component A passive device permanently installed in a panel or plate (e.g., a FEEDTHROUGH CAPACITOR or FEEDTHROUGH INSULATOR).

feedthrough insulator An insulator mounted tightly in a hole in a wall or chassis, and provided with a center hole for a lead.

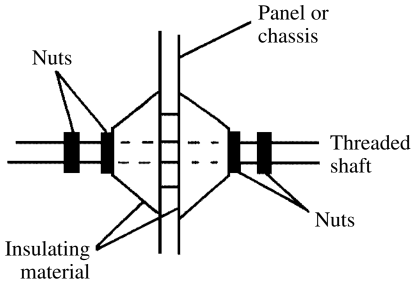

feedthrough insulator

feedthrough terminal A terminal mounted tightly in a hole in a chassis or wall; it consists of a screw going through a feedthrough insulator. Connections can be made to either end of the screw.

FE-EL Abbreviation of *ferroelectric-electroluminescent.*

feeler **1.** A wire or blade contact (e.g., a finger that senses holes in a punched card). **2.** The wire portion of a point-contact diode.

Felici mutual-inductance balance An inductive null circuit for determining mutual inductance (M_x) in terms of a standard mutual inductance (M_s). The secondary coils of two mutual-inductance circuits are connected in phase opposition. The standard mutual inductor, which is variable, is adjusted for null. At null, $M_x = M_s$.

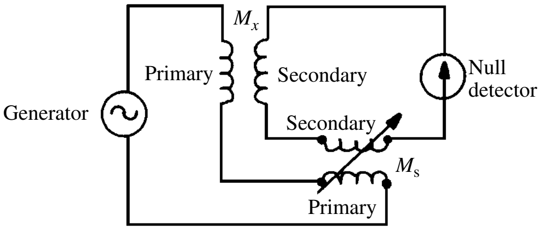

Felici mutual-inductance balance

female plug A plug whose contacts are separated by a recess into which the prongs of a mating male plug are inserted. Compare HERMAPHRODITE PLUG and MALE PLUG.

femto- Abbreviation, f. A prefix meaning quadrillionth (10^{-15}).

femtoampere Abbreviation, fA. A unit of extremely low current; 1 fA equals 10^{-15} A.

femtofarad Abbreviation, fF. A unit of extremely low capacitance; 1 fF equals 10^{-15} F.

femtovolt Abbreviation, fV. A unit of extremely low voltage; 1 fV equals 10^{-15} V.

fence A system or string of early warning radar stations.

Fermat's principle The principle that the path of a ray of radiant energy between two points in any medium is the shortest distance between those points in that medium. This path is also the path of least propagation time. This might not necessarily be a straight line. On a flat plane, the path of least time between any two points is always a straight line, assuming constant velocity of propagation. However, on a sphere, the path of least time between two points is represented by a great circle.

fermi Abbreviation, F. An extremely small unit of length and wavelength, equal to one quadrillionth (10^{-15}) of a meter.

Fermi-Dirac distribution function The probability that an electron will be in a certain quantum state under conditions of thermal equilbrium. This is determined as the probability that the electron will be in a given energy band at a particular instant of time.

fermium Symbol, Fm. A radioactive metallic element that is artificially produced. Atomic number, 100. Atomic weight, 257.

ferpic Acronym for *ferroelectric ceramic picture* device. An image-storing device containing a photoconductive film, transparent electrodes, and a ferroelectric ceramic, in layers.

ferreed A form of magnetic switching device, similar to a reed relay, that maintains its position indefinitely without the need for a continuous current.

ferret A vehicle or craft equipped for determining the locations of enemy radar transmitters.

ferri- A prefix used to denote magnetic properties.

ferric oxide Formula, Fe_2O_3. A red oxide of iron used to coat magnetic recording tape.

ferristor A high-frequency magnetic amplifier using a ferroresonant circuit.

ferrite A high-resistance magnetic material consisting principally of ferric oxide and one or more other metals. After being powdered and sintered, ferrites exhibit low eddy-current loss at high frequencies and make ideal core material for inductors and switching elements. Also used in television deflection yokes and in miniature antennas. Also see FERROSPINELS.

ferrite antenna See FERRITE-ROD ANTENNA.

ferrite bead **1.** A magnetic storage device in the form of a bead of ferrite powder fused onto the signal conductors of a memory matrix. **2.** A tiny ring of ferrite that can be slipped over a wire or cable to choke off radio-frequency (RF) currents.

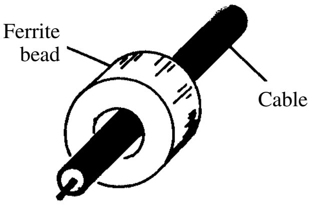

ferrite bead

ferrite core A coil or switching-element core made from a ferrite; specifically, in a core memory, a small magnetic toroid that can retain its polarity when charged by a pulse.

ferrite core memory A magnetic memory in which ferrite cores are interconnected by a network of input and output wires.

ferrite isolator A microwave device that permits energy to pass with negligible loss in one direction through a waveguide or coaxial line, while absorbing energy passing in the opposite direction.

ferrite limiter A device used in the antenna circuit or front end of a receiver, to prevent overload while maintaining a linear response. Used mostly at ultra-high and microwave frequencies.

ferrite loopstick See FERRITE-ROD ANTENNA.

ferrite memory A static memory using ferrite cores. See CORE MEMORY.

ferrite-rod antenna Also called *loopstick antenna*. A small antenna that can be used for wireless reception at frequencies below approximately 20 MHz. This antenna consists of a coil wound on a solenoidal, high-permeability, powdered-iron core, usually less than 20 centimeters (cm) long and 1 cm in diameter. A series or parallel capacitor, in conjunction with the coil forms a tuned circuit. The operating frequency is determined by the resonant frequency of the inductance-capacitance (LC) combination. Response is maximum off the sides of the coil, and a sharp null occurs off the ends. This antenna has narrow bandwidth. The null can be oriented to minimize system response to undesired local signals or humanmade noise. Compare SMALL LOOP ANTENNA.

ferrite switch A device that regulates the flow of power through a waveguide. The electric-field vector is rotated, resulting in a high degree of attenuation when actuated, but little or no attenuation when not activated.

ferroelectric **1.** Producing ferroelectricity. **2.** A ferroelectric material.

ferroelectric amplifier See DIELECTRIC AMPLIFIER.

ferroelectric capacitor A capacitor in which a ferroelectric material is the dielectric.

ferroelectric cell See FERROELECTRIC CAPACITOR.

ferroelectric crystal A crystal of ferroelectric material.

ferroelectric flip-flop A flip-flop based on the hysteresis of a ferroelectric capacitor. Compare FERRORESONANT FLIP-FLOP.

ferroelectricity Electric polarization in certain crystalline materials. The effect is analogous to the magnetization of a ferromagnetic material by a magnetic field.

ferroelectric-luminescent Pertaining to a ferroelectric cell that emits light.

ferroelectric material A nonlinear dielectric material capable of producing ferroelectricity. Examples: barium titanate, barium strontium titanate, potassium dihydrogen phosphate, guanadine aluminum sulfate hexahydrate (GASH), Rochelle salt, and triglycene sulfate.

ferromagnetic **1.** Pertaining to a substance that conducts a magnetic field with relative ease. **2.** Pertaining to a material in which a magnetic-field change causes a voltage, which in turn results in a measurable current flow.

ferromagnetic-core inductor A coil of wire designed to introduce inductive reactance into a circuit or system, wound around a core consisting of ferromagnetic material that greatly increases the inductance for a given number of turns. With proper design and choice of core material, the Q factor (selectivity attainable) can be extremely high. The core confines much of the magnetic flux within itself. When the coil current exceeds a certain level, *core saturation* occurs, and further increases in the current will not produce a corresponding increase in magnetic flux. This decreases the effective inductance, and reduces the efficiency because power is dissipated as heat in the core.

ferromagnetic material A substance that concentrates magnetic lines of flux relative to their concentration in free space. Iron, powdered iron, and ferrite are common examples.

ferromagnetic resonance The point at which the permeability of a magnetic material peaks at a microwave frequency.

ferromagnetic spinels Highly permeable and resistive ceramic-like materials. The low eddy-current losses and high permeability of these materials suit them for use as cores in radio-frequency (RF) transformers and inductors. Also see FERRITE.

ferromagnetic tape Magnetic tape used for winding closed transformer cores.

ferrometer An instrument for testing hysteresis and permeability in steel and iron.

Ferron detector See IRON-PYRITES DETECTOR.

ferroresonant circuit An inductance-capacitance (LC) circuit in which the coil is a saturable reactor. Because of coil nonlinearity, the circuit is resonant at only one value of alternating-current (ac) voltage, and exhibits both negative resistance and bistable operation.

ferroresonant counter A digital counter using ferroresonant flip-flops, rather than semiconductor devices.

ferroresonant flip-flop A flip-flop using one or two ferroresonant circuits instead of semiconductor devices. See BISTABLE MULTIVIBRATOR. Compare FERROELECTRIC FLIP-FLOP.

ferroresonant shift register A shift register using ferroresonant circuits instead of semiconductor devices.

ferrosoferric oxide See MAGNETITE.

ferrospinels See FERROMAGNETIC SPINELS.

ferrous Pertaining to a substance that contains iron and is magnetizable.

Ferroxcube A nonmetallic ferromagnetic material having high permeability and resistivity, and a Curie point near room temperature. These characteristics make the material suitable for the cores of radio-frequency (RF) inductors and transformers, and for high-frequency magnetic shields.

Fessenden oscillator In underwater communications, a transmitter of acoustic waves.

FET Abbreviation of FIELD-EFFECT TRANSISTOR.

fetch An operation in a computer run in which the location of the next instruction is taken from memory and changed if necessary; it then goes to the control register.

FET current meter An ammeter, milliammeter, or microammeter having a self-contained amplifier that uses field-effect transistors. Also see ELECTRONIC CURRENT METER.

FET op-amp **1.** An operational amplifier composed of field-effect transistors and associated components. **2.** An operational amplifier having a field-effect transistor in its input stage.

FET voltmeter A voltmeter using a field-effect transistor amplifier for high-impedance input. Also see ELECTRONIC VOLTMETER.

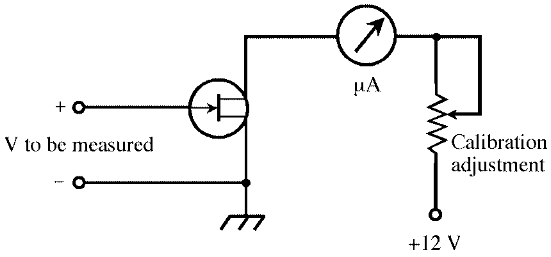

FET voltmeter

FET VOM A volt-ohm-milliammeter (VOM) using a field-effect transistor amplifier for increased sensitivity and high input impedance.

FF Abbreviation of FLIP-FLOP.

fF Abbreviation of FEMTOFARAD.

FFI Abbreviation of FUEL-FLOW INDICATOR.

fhp Abbreviation of FRACTIONAL HORSEPOWER.

fiber **1.** A tough, vulcanized insulating material. Dielectric constant, 2.5 to 5. Dielectric strength, 2 kV/mm. **2.** A thin thread of a material. **3.** Also called *optical fiber.* A light-conductive transparent filament; see FIBEROPTICS, **1**.

fiber electrometer An instrument for measuring small quantities of electricity. It consists of a thin thread, such as one of plasticized quartz, hanging freely between two knife-edged metal pieces that are charged by the electricity being measured. The charge draws the fiber away from its position of rest. The movement can be observed with a microscope. A special form of this instrument, using two fibers, is the *bifilar electrometer.*

fiber needle A soft phonograph needle made from a fiber. It produces less disk wear than other styli, but is short-lived.

fiber metallurgy A process in which metallic fibers or filaments are grown.

fiberoptic bundle A cable of optical fibers. See FIBEROPTICS, **1**.

fiberoptic coupling Also called *optical coupling.* A method of coupling in which a light-conducting fiber is placed between a light (signal) source and a photoreceptor. Also see FIBEROPTICS, **1**, **2**.

fiberoptics **1.** Also called *optical fibers.* Extruded materials, such as certain plastic filaments, that provide paths for light. **2.** The science of developing and using communications systems that use optical fibers to transfer data.

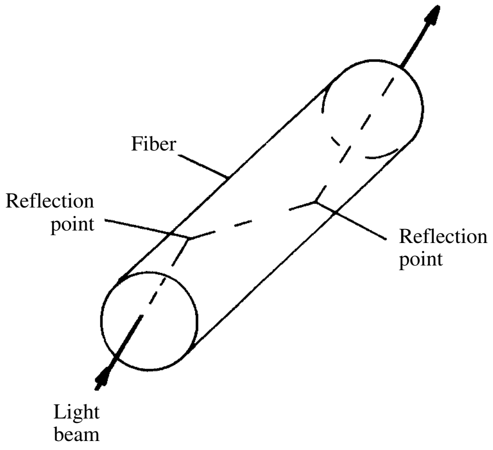

fiberoptics

fiberoptic scrambler A fiberscope in which a center section of fibers in the core is deliberately disoriented before the bundle is encapsulated; when cut, one half can decode the image encoded by the other half. See FIBERSCOPE.

fiberscope A flexible bundle of optical fibers having a lens at each end; it is used to view areas that are otherwise inaccessible to view.

fiber stylus See FIBER NEEDLE.

fibre Alternate (Brit.) spelling of FIBER.

fibrillation Dangerous, irregular beating of the heart that often follows electric shock. Also see CARDIAC STIMULATOR and DEFIBRILLATION.

fidelity The degree to which a circuit or device transmits a signal without distorting it. Pertains especially to acoustic devices and stereo audio equipment.

field **1.** A volume of space in which a force is operative. See, for example, ELECTRIC FIELD and MAGNETIC FIELD. **2.** Half of a video image, consisting either of all the even-numbered lines, or all the odd-numbered lines. **3.** A computer record subdivision containing an information unit (e.g., a bank account record might have deposits as a field).

field circuit breaker A circuit breaker designed to control the field excitation of a motor or other device.

field coil **1.** The winding on the field pole of a motor or generator. **2.** The winding on the pole of an electrodynamic speaker. **3.** The main coil of a relay. **4.** The fixed coil in an electrodynamometer.

Field Day In amateur radio, an annual contest sponsored by the American Radio Relay League (ARRL) to simulate emergency communication conditions. It occurs during the last full weekend in June.

field direction The direction in which an electric field or magnetic field exerts its force.

field effect The phenomenon in which the flow of current carriers in a solid substance is controlled by an external electric field. A useful application is the FIELD-EFFECT TRANSISTOR (FET).

field-effect tetrode Also called *dual-gate FET.* A field-effect transistor (FET) in which the gate electrode is split into two parts, each connected to a separate external lead. The reverse bias between the channel and either gate lead affects the conductivity through the device.

field-effect transistor Abbreviation, FET. A monolithic semiconductor amplifying device in which a high-impedance GATE electrode controls the flow of current carriers through a thin bar of silicon (rarely, germanium) called the CHANNEL. Ohmic connections made to the ends of the channel constitute SOURCE and DRAIN electrodes. Also see JUNCTION FIELD-EFFECT TRANSISTOR, METAL-OXIDE SEMICONDUCTOR FIELD-EFFECT TRANSISTOR, and CHANNEL JUNCTION FIELD-EFFECT TRANSISTOR.

field-effect varistor A nonlinear dual-terminal semiconductor device capable of maintaining a value of current for a range of voltages.

field-emission microscope An instrument for examining the atomic structure of high-melting-point metals; it magnifies more than 2 million times. The metal to be examined is made into a needle that is subjected to 5 to 30 kV; electrons emitted by the tip of the needle form an image on a fluorescent screen.

field forcing A method of controlling a motor by changing the magnetic field in the windings.

field frequency In television, the product of frame frequency and fields per frame (in the United States, 60 per second).

field intensity **1.** The strength of an electric or magnetic field. **2.** The strength of an electromagnetic field, usually expressed in microvolts per meter or millivolts per meter.

field-intensity meter See FIELD-STRENGTH METER.

field ionization The tendency for atoms to be ionized in a gas by a high-intensity electric field. The ionization occurs mostly near the poles of the electric field.

field-ion microscope A high-resolution field-emission microscope that uses helium ions instead of electrons. The ions are repelled by the tip of the metal needle under observation, forming an image on a fluorescent screen. Also see FIELD-EMISSION MICROSCOPE.

field length Record field size in applicable units, usually bytes, characters, or words.

field magnet **1.** The permanent magnet in a dynamic speaker. **2.** A similar magnet in an earphone, generator, microphone, motor, phono pickup, transducer, etc.

field pickup **1.** A probe or sensor for insertion into an electric or magnetic field. **2.** An on-location radio or television program (i.e., one coming from outside the studio). Also called remote or NEMO.

field resistor A resistive component consisting of an insulated form with a thin layer of conductive material.

field rheostat The rheostat whose setting determines the amount of current flow through the field coil of a motor or generator.

field scan A form of television scanning in which the lines are scanned alternately.

field-sequential system A color-television system in which the image is reproduced by means of primary color fields (red, green, and blue) flashed sequentially on the screen of the picture tube. Compare DOT-SEQUENTIAL SYSTEM and LINE-SEQUENTIAL SYSTEM.

field strength See FIELD INTENSITY.

field-strength meter An instrument for measuring the radio-frequency (RF) voltage of a signal reaching a chosen location. The instrument consists essentially of a radio detector equipped with a portable antenna and an output meter.

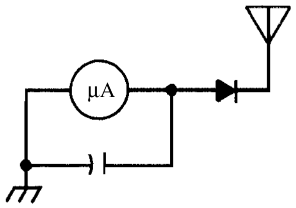

field-strength meter

field telephone A rugged, portable telephone system for outdoor use.

field test A test of equipment under actual operating conditions (i.e., outside the laboratory or factory).

field winding See FIELD COIL, **1**, **2**.

fig. Abbreviation of *figure* (usually in reference to an illustration).

figure-8 pattern **1.** A bidirectional antenna pattern whose cross section, in a specified plane, resembles the numeral 8. **2.** A Lissajous figure resembling the numeral 8.

figure of merit **1.** For a capacitor, inductor, or tuned circuit, the ratio of reactance to resistance. Also called *Q factor* or *Q.* **2.** For a magnetic amplifier, the ratio of power amplification to control time constant. **3.** For a transistor, the *gain-bandwidth factor.*

fil Abbreviation of FILAMENT.

filament In a vacuum tube or incandescent lamp, the thin wire heated by electric current; it emits electrons, light, and heat. The filament is the cathode in a filament-type tube, but serves only to heat indirectly the cathode sleeve in an indirectly heated tube.

filament battery See A BATTERY.

filament choke A radio-frequency (RF) choke operated in the filament lead of an electron tube. Such chokes are necessary in filament-type tubes in RF power amplifiers.

filament circuit The circuit carrying filament current.

filament coil See FILAMENT WINDING.

filament current Symbol, I_F. The current flowing through the filament of a vacuum tube.

filament emission Electrons emitted directly by the filament in an electron tube (thermionic emission) or the amount of such emission.

filament hum A hum signal caused by voltage induced in a circuit by the alternating-current (ac)-operated filaments (heaters) of tubes or by the filament wiring.

filament lag The time delay in the heating and cooling of an ac tube or lamp filament as filament-current changes polarity.

filament power supply A source of power, usually alternating-current, for heating the filament of a vacuum tube or tubes.

filament resistance **1.** The resistance of the filament in an electron tube or incandescent lamp. **2.** The resistance of an external dropping resistor in the filament circuit of a tube or lamp.

filament transformer A step-down transformer that supplies power exclusively to the filament (heater) of an electron tube.

filament-type bolometer A BOLOMETER in which the sensitive element is a wire filament. Examples: barretter, incandescent lamp, and wire fuse.

filament voltage Symbol, V_F. The voltage across the filament of an electron tube.

filament winding In a power transformer, the coil that supplies heating power to the filament of a vacuum tube. Also called FILAMENT COIL.

file An organized collection of records related by a common format, data source, or application.

file conversion Converting data files from one form to another, often for the purpose of making them compatible with other computers.

file gap An area of a data medium that signifies the end of a file; it can also mark the start of another file.

file identification A code that identifies a file.

file label File identification in which the first record in the file is a set of characters unique to the file; it conveys such information about (for example, a tape file as a description of content, generation number, reel number, date of writing, etc.). Also called HEADER LABEL.

file layout How the contents of a file are organized; usually defined by the system or specified by a program.

file maintenance To delete, add, or correct records in a file. Unlike updating, which is done to reflect changes in events recorded in the file, maintenance ensures that the contents of the file are accurate records of the necessary data.

file management A method of storing and recalling data from computer storage media, such as magnetic disks.

filename In a computer file label, the alphanumeric character set that identifies and describes the file. It generally consists of one to eight characters, often including an extension of one to three characters. In some computer operating systems, filenames can be much longer (e.g., up to 32 characters). See FILENAME EXTENSION.

filename extension A group of one, two, or three characters following the main body of a FILENAME, and separated from it by a period (.). It generally denotes the application or purpose of the file. The extension .BAK, for example, denotes a backup file; .TXT denotes a text file; .BMP denotes a bit-mapped graphics file.

file organization The way words, bits, or records are physically arranged in the storage medium for a file, possibly including the method of access (serial, alphabetical, random, etc.).

file-oriented programming Computer programming that uses a general file and record-control program to simplify I/O coding.

file-oriented system A system having file storage as its basis.

file print A hard copy (printout) of the contents of a file.

file processing The operations associated with making and using files.

file protection Preventing the possibility of writing over data files before they are made available for use. It is usually done by having a program check file labels.

file reconstitution Restoring a partially or completely damaged file by updating a previous generation of the file using a file of interim transactions.

file recovery Following the interruption of file processing because of system failure, the procedure for reestablishing the file's condition as necessary for the resumption of processing without losing accuracy.

file section Part of a file in certain consecutive locations on a storage medium.

file security Protective and security measures (e.g., the issuance of clearances, status markers, etc.) as they relate to computer files.

file set A collection of interrelated files stored consecutively in a magnetic disk volume (package).

fill The percentage of lines in a cable that are actually in use at a given time.

filler A nonessential data part used, for example, to bring a record to a standard size.

film See THIN-FILM MEMORY.

film capacitor A capacitor in which the electrodes are plated or deposited on the faces of a thin film of plastic or other dielectric material.

film chain A system designed for the transmission of movies over a television system. This requires synchronization of the movie frame rate with the television scanning rate.

film frame A single picture on a strip of motion picture film.

film-frame blanking interval The interval during which a film frame is blanked out as motion-picture film moves through a camera, projector, or pickup. The blanking action allows a frame to move into position without blurring the image as seen by viewers.

film integrated circuit A monolithic circuit whose elements are films formed on an insulating substrate.

film pickup A photocell, photodiode, or phototransistor circuit used to pick up recordings from the sound track of motion-picture film.

film reader A device for converting data on film into digital form for a computer.

film recorder An apparatus that records data as a sound pattern on film. Compare FILM REPRODUCER.

film reproducer An apparatus that plays back data recorded on photographic film. Compare FILM RECORDER.

film resistor Also called *carbon-film resistor* or *metal-film resistor.* An electronic component consisting of a mixture of ceramic, carbon, and/or metal, applied to a cylindrical form in a thin layer to obtain a desired value of resistance. The cylindrical form is made of an insulating substance, such as porcelain. The film can be deposited on this form by various methods, and the value tailored as desired. Metal-film units can be made to have nearly exact values. Film type resistors usually have low to medium-high resistance. A major advantage of film-type resistors is that they, like carbon-composition units, do not have much inductance or capacitance. A disadvantage, in some applications, is that they cannot handle as much power as the more massive carbon-composition units, or as wirewound types. Compare CARBON-COMPOSITION RESISTOR, WIREWOUND RESISTOR.

film scanning The conversion of a movie into a form suitable for transmission by television.

film speed **1.** The speed at which motion-picture film moves intermittently through a camera, projector, or pickup, measured in feet or frames per second. **2.** A measure of film's light sensitivity, given as an ASA (American Standards Association) or DIN (European) number; in either system, the higher the number, the greater the light sensitivity.

filter **1.** A circuit or device that passes alternating currents at some frequencies while attenuating or blocking currents at other frequencies. Examples: BANDPASS FILTER, BAND-REJECTION FILTER, HIGH-PASS FILTER, and LOW-PASS FILTER. **2.** An inductance-capacitance (LC) or resistance-capacitance (RC) circuit for removing the ripple from the output of a power-supply rectifier. **3.** A transparent disk with special optical properties placed in front of a camera lens for a special photographic effect. **4.** A character patterned to control the elimination or selection of characters in another pattern. **5.** A device or program that separates information, according to certain specifications or characteristics. **6.** A machine word that specifies the elements to be treated in another machine word; also called MASK.

filter attenuation In a selective filter, the power, current, or voltage loss, in decibels, that occurs within the passband.

filter attenuation band In a selective filter, the frequency band(s) outside the passband; that is, the frequency range over which signals are significantly attenuated.

filter bank In audio applications, a set of bandpass filters, each of which covers a specific portion of the audio-frequency (AF) spectrum. There is some (minimal) overlap between the passbands of the filters, so an AF signal of a specific frequency will always pass through at least one of the filters, but will never pass through more than two of the filters.

filter capacitor A capacitor that provides capacitive reactance in a wave filter or power-supply filter while also blocking direct current.

filter center A place where information is modified for transmission to aircraft pilots. Such information can include weather data, course changes, or other instructions.

filter choke An inductor that provides inductive reactance in a wave filter or power-supply filter while affording relatively easy conduction of direct current.

filter crystal A piezoelectric crystal used in a CRYSTAL RESONATOR.

filter cutoff The frequency or frequencies at which the transmission figure of a filter is below its maximum value by a prescribed amount, usually 3 dB, representing the half-power point(s).

filter discrimination The amount of fluctuation in the insertion loss of a bandpass, band-rejection, high-pass, or low-pass selective filter. The fluctuation is measured at various points in the filter passband.

filter inductor See FILTER CHOKE.

filter passband The frequency range over which a selective filter passes signals with minimum attenuation.

filter reactor See FILTER CHOKE.

filter slot In a waveguide, a slot that acts as a choke to suppress undesirable modes.

filter stopband See FILTER ATTENUATION BAND.

filter transmission band See FILTER PASSBAND.

fin **1.** A metal disk or plate attached to a component for the purpose of radiating heat. **2.** A projection in an irregular heatsink.

final amplifier Also called *output amplifier* or *power amplifier.* In a wireless transmitter, the amplifier that boosts the signal power to the level to be emitted over the airwaves. The input is provided by the driver; the output is connected to the system, to a transverter, or to a network cable power amplifier. In some transmitters, the output amplifier requires no tuning; in others, adjustment of the output circuitry is necessary to obtain proper coupling to the antenna system. In a single-sideband (SSB) or amplitude-modulated (AM) transmitter, all amplifiers following the modulator must be linear. In transmitters intended for use only in digital modes or for frequency modulation (FM), power amplifiers need not be linear.

final result A result displayed at the end of a data processing operation. Compare INTERMEDIATE RESULT.

finder The switch or group of relays that selects the path for a call going through a telephone switching system. Also called LINE FINDER.

fine adjustment Adjustment of a quantity in small increments or as a smooth, continuous variation. Compare COARSE ADJUSTMENT.

fine-chrominance primary See I SIGNAL.

fine frequency control A variable component, such as a potentiometer or variable capacitor, that permits a signal or response frequency to be varied over a small increment; it is often used in conjunction with a coarse frequency control.

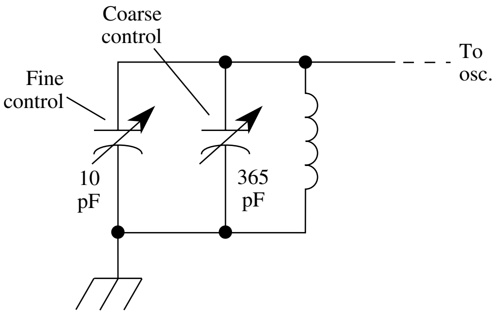

fine frequency control

fine-groove record See MICROGROOVE RECORD.

fine index In computer operations, a secondary, supplemental index used with a main, or gross, index when the latter does not adequately detail the differences between the items being indexed.

fine-motion planning The scheme used by a robotic end effector (“hand”) to get in exactly the correct position. It uses machine vision, tactile sensors, pressure sensors, or high-precision displacement transducers.

finger See FEELER, **1**.

finger plethysmograph A device that senses and records the resistance through the human finger during various parts of the heart cycle.

finger rules See FLEMING’S LEFT-HAND RULE, FLEMING’S RIGHT-HAND RULE, and RIGHT-HAND RULE FOR WIRE.

finish lead The lead attached to the last turn of a coil. Also called *outside lead.* Compare START LEAD.

finished blank The end product in the crystal manufacturing process, often including electrodes.

finishing The careful handwork and testing involved in bringing a crystal blank to a condition that is acceptable as finished, according to specifications.

finishing rate The rate of charging a battery, as the battery approaches a full charge. Generally, the finishing rate is less than the normal charging rate.

finite Pertaining to that which has defined limits.

finite sample space In statistics, a sample space having definite limits.

finite series A mathematical series having a limited number of terms. Example: 1 + 0.1 + 0.01 + 0.001, a finite series containing four terms and whose sum is equal to 1.111.

finned surface The irregular surface of a heatsink. The ratio of surface area to volume is greater than with a flat surface; this increases the cooling effectiveness of the heatsink.

fins Metal vanes radiating from components that dissipate large amounts of power as heat.

FIR Abbreviation of FAR INFRARED.

fire A transition from non-conduction to conduction in an ionizing switching device.

fire control The aiming and firing of guns automatically via radar and associated electronic systems.

firefighter robot A robotic device or system intended to protect lives and property from fire. Such a robot can be simple, such as an automatic sprinkler system actuated via heat sensors and/or smoke detectors. More sophisticated systems might use autonomous or remotely controlled androids that perform the same functions as human firefighters.

Fire Underwriter’s regulations See NATIONAL ELECTRIC CODE.

firing The pulse that initiates conduction in an ionization switching device.

firing angle **1.** For a magnetic amplifier, the angular distance through which the input-voltage vector rotates before the core is driven into saturation. **2.** For a silicon-controlled rectifier (SCR), the point, as an angle (in degrees or radians), along the control voltage half-cycle at which the SCR fires.

firing circuit Any circuit, such as a phase shifter, that permits adjustment of the firing angle of a silicon-controlled rectifier (SCR) or similar device, or which delivers the required pulse or other signal to initiate firing.

firmware **1.** Programs (software) in nonvolatile computer memory [e.g., in a read-only memory (ROM), that can only be changed by replacement with an alternate unit]. **2.** Unalterable internal interconnections that determine what a computing device or system can do. Also called MICROPROGRAM.

first detector In a superheterodyne circuit, the signal frequency detector. Compare SECOND DETECTOR.

first filter capacitor The input capacitor in a capacitor-input power-supply filter.

first Fresnel region A portion of a directional transmitted electromagnetic ray, shaped generally like a paraboloid with the apex at the transmitter and the axis in the direction of transmission. Any point in the first Fresnel zone is in such a position that the sum of the lengths of the paths from the point to the receiver, and the point to the transmitter, is no greater than 0.5 wavelength more than the distance from the transmitter to receiver.

first harmonic The fundamental frequency in a complex waveform from which multiples are generated.

first-in/first-out Acronym, FIFO. A read-write memory commonly used as a buffer to smooth out the flow of data bits in a digital system. The output bits are in the same order (sequence) as the input bits. If a data bit *x* goes into the FIFO before data bit *y*, then *x* will always emerge from the FIFO before *y*. Compare PUSHDOWN STACK.

first-in/last-out See PUSHDOWN STACK.

first law of thermodynamics Quantities of heat can be converted into mechanical work, and vice versa. Also see MECHANICAL EQUIVALENT OF HEAT.

first level address See ABSOLUTE ADDRESS.

first selector The selector that responds to the first-digit dial pulses when a telephone number is called.

fishbone antenna An untuned, wideband directional antenna of the general end-fire type. Consists of a number of collector antennas, each capacitively coupled to the resistor-terminated transmission line in collinear pairs. It is so called from its resemblance to the skeleton of a fish.

fishpaper A chemically treated, vulcanized-fiber paper used for electrical insulation.

fish pole A microphone boom that can be held in the hand.

fist In radiotelegraphy and wire telegraphy, an operator's manual sending style.

fitting A device intended to mechanically fasten a wire or cable in place.

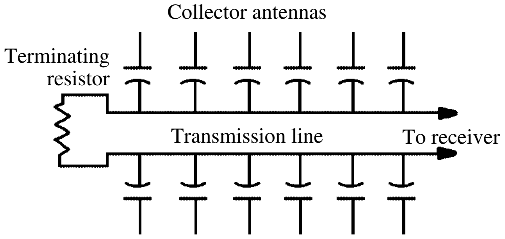

fishbone antenna

five-element code A five-impulse telegraph code that describes a character (e.g., Baudot, which also includes start and stop elements).

five-layer device A semiconductor device containing four pn junctions. Examples: diac and triac.

five-level code A teletype code that utilizes five binary elements to define a character.

fix **1.** In direction finding, the point at which two lines of direction intersect. **2.** In electronics maintenance, to repair successfully. **3.** To subject an in-process circuit board to a solution or other medium to stop a photographic action permanently. Also, the solution in which such photosensitive materials can be immersed to halt development.

fixed bias Bias voltage or current supplied from a fixed external source (such as a battery or power supply) independent of the operation of the biased device. Compare AUTOMATIC BIAS.

fixed block length Blocks of data having a fixed number of words or characters, as required because of hardware limitations or a program instruction. Compare VARIABLE BLOCK LENGTH.

fixed capacitor A nonadjustable capacitor (i.e., one having a single unalterable value).

fixed component Any component (e.g., a capacitor, inductor, or resistor that has one unalterable value).

fixed contact The stationary contact in a relay or switch. Compare MOVABLE CONTACT.

fixed-crystal detector A simple crystal detector in which the point of the contact wire is permanently placed in contact with a sensitive spot on the surface of the crystal.

fixed field Fields in records organized so that those containing similar information in each record are the same length and in the same relative position in the record. Compare VARIABLE FIELD.

fixed form coding The coding of source languages so that each part of the instruction is in a fixed field.

fixed-frequency amplifier An amplifier that is pretuned to operate on one frequency, or in a relatively narrow band of frequencies.

fixed-frequency oscillator An oscillator that is preset to operate on one frequency. Such an oscillator can be self-excited or controlled (crystal, fork, magnetostriction, etc.).

fixed-frequency receiver A receiver that is pretuned to receive signals of one frequency.

fixed-frequency transmitter A transmitter that is pretuned to radiate signals of one frequency. Such a transmitter can contain a self-excited or crystal-controlled oscillator.

fixed inductor A nonadjustable inductor (i.e., one having an unalterable value of inductance).

fixed-length record A record in which word or character size is constant. Compare VARIABLE LENGTH RECORD.

fixed logic Applicable to computers or peripherals whose logic can only be altered internally by changing connections.

fixed memory A nonvolatile readout computer memory that can only be altered mechanically.

fixed placement file A file that has been allocated a fixed location in storage.

fixed-point system A notation system in which a single set of digits represents a number, and the radix point (in the decimal system, the decimal point) can only be placed in one position for the value being expressed. Also see FLOATING-POINT CALCULATION.

fixed resistor A nonadjustable resistor (i.e., one having an unalterable value of resistance).

fixed-sequence robot A robot that performs one task or set of tasks, making exactly the same movements every time. The sequence is programmed in the robot controller via firmware. Many assembly robots and toy robots are of this type. Compare FLEXIBLE AUTOMATION.

fixed station A radio station operating from a stationary point; one that is not mobile.

fixed-step potentiometer A potentiometer whose output is varied in one or more discrete steps by fixed-resistor sections. Also see POTENTIOMETER.

fixed wireless 1. A cellular communications network in which some, or all, of the end users are at fixed locations. This is increasingly popular as an alternative to conventional hard-wired telephone and Internet access. 2. The use of wireless equipment, such as cell phones and wireless-modem–equipped computers, at fixed (rather than mobile or portable) locations.

fixed word length Applicable to the organization of information in storage in which each computer word stored has a fixed number of characters or bits.

fixture A piece of hardware used in equipment setups (e.g., microwave couplers, joints, sections, etc.).

fL Abbreviation of FOOT-LAMBERT.

flag **1.** A piece of information added to a data item that gives information about the data item. **2.** A bit added to a character or word to delineate a field boundary. **3.** An indication that an operation is complete and need not be done by the program. **4.** An indicator identifying the members of mixed sets. **5.** A character that signals the presence of some condition (e.g., an *error flag* indicates that a data item caused an error).

flag event A program condition that causes a flag to be set.

flag line An input pulse to a microprocessor that depends on specific external instructions. Indicates a certain condition or change of state.

flagpole antenna **1.** Any of several vertical UHF or VHF antennas consisting of a radiator mounted atop a coaxial pipe or cable (see, for example, COAXIAL ANTENNA). It takes its name from its resemblance to a flagpole. **2.** A vertical antenna formed by shunt-feeding a pole already in existence, such as a flagpole. This can be done to conceal the antenna.

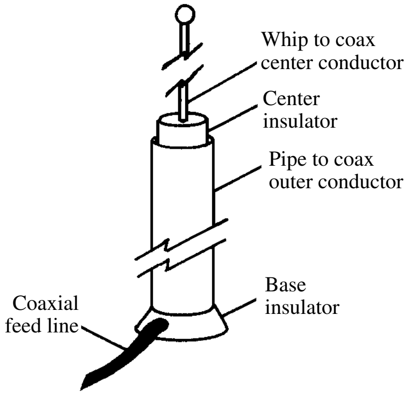

flagpole antenna, 1

flag terminal A form of terminal that does not require soldering for electrical contact. A protruding “flag” is crimped around the conductor.

flame alarm A (usually photoelectric or thermoelectric) device or circuit for detecting a flame and actuating an alarm.

flame control A (usually photoelectric or thermoelectric) device or circuit for sensing and automatically controlling the height of a flame, such as a gas pilot.

flame-failure control A FLAME CONTROL that automatically cuts off the fuel if the flame goes out.

flame microphone A microphone in which two electrodes in a flame undergo a change in electrical resistance when the flame is influenced by sound waves, thus modulating current passing between the electrodes.

flange **1.** A flat, protruding edge used for fastening a connector or plug to the chassis of a piece of equipment. **2.** In a waveguide, a coupling used for connection to another section of waveguide, or to a horn or other external device.

flange focus The focal length of a lens, based on the distance from the mounting flange to the focal plane.

flanging In sound systems, a means of obtaining special effects using a COMB FILTER. Creates an "extraterrestrial," digitized sound. Commonly used by rock music bands.

flanking Modification of the response of a selective filter, resulting from the parallel connection of two or more similar filters.

flap See DROP INDICATOR.

flap attenuator A waveguide attenuator consisting of a sheet of resistance material inserted transversely into the waveguide through a slot.

flare **1.** The hyperbolic cross section of a horn antenna or loudspeaker. **2.** A transient or stationary bright area with (usually) a central pip on the screen of an otherwise blank oscilloscope or television picture tube.

flare angle The gradual change in a waveguide's diameter over its length.

flare factor The angle at which the faces of a horn speaker are curved or turned outward.

flaring constant Symbol, m. A number expressing the degree of flare (see FLARE, **1**) in a horn antenna or loudspeaker. The value of m is equal to $0.6931/d$, where d is the distance along the axis required for the cross-sectional area A of the horn to double. If d is given in feet (ft), for example, then A is in units of square feet (ft^2) and m is in units of reciprocal feet (1/ft). If d is in centimeters (cm), then A is in units of square centimeters (cm^2) and m is in units of reciprocal centimeters (1/cm).

flash **1.** A photographic camera flash. **2.** To vaporize a metal (such as magnesium) in an electron tube being evacuated, to absorb gases. **3.** Flashover.

flash arc In a vacuum tube, a sudden high-current arc between cathode and plate at high plate voltages; it can short-circuit the plate power supply.

flashback voltage The maximum inverse voltage that causes the gas in a tube to ionize.

flash delay A device that automatically postpones the operation of a FLASHTUBE until a predetermined instant, such as the moment when a moving object arrives at a particular point before a camera.

flasher An electrical or electronic device or circuit that flashes a light or a series of lights sequentially.

flasher LED A light-emitting diode (LED) that, when connected to a low-voltage direct-current (dc) source, emits light that flashes at a basic rate of a few pulses per second.

flashlamp **1.** See FLASHTUBE. **2.** A small portable light operated from self-contained cells; a flashlight or lantern.

flashlight See FLASHLAMP, **2**.

flash memory High-capacity, nonvolatile random-access memory, used in some computers in place of an electromechanical hard drive. This type of memory allows a computer to store, access, and run large data files or programs much faster than is possible with a conventional hard drive. See also NONVOLATILE MEMORY, RANDOM-ACCESS MEMORY.

flashover The sudden discharge of electrical energy between electrodes or conductors, often accompanied by light; it is usually the result of excessive voltage.

flashover voltage **1.** The peak voltage at which FLASHOVER occurs. **2.** The voltage at which disruptive discharge occurs between electrodes and across the surface of an insulating material.

flash plating Electroplating in which a thin layer is deposited quickly.

flash test Insulation testing by applying a higher-than-normal voltage for a short time.

flashtube A straight or coiled glass tube filled with gas and provided with electrodes. When a high voltage is applied to the electrodes, the tube emits a brilliant flash of light.

flat cable Also called *ribbon cable.* A cable whose flexible conductors are molded side by side in a flexible, flat ribbon of plastic (such as polyethylene).

flat-compounded generator A compound-wound generator whose windings are proportioned so that the full-load and no-load voltages are identical.

flat fading Fading of a radio signal that occurs independently of frequency; all frequency components of the signal fade to the same extent at the same time.

flat file A computer file containing unfolded documents.

flat frequency response Relatively equal response to all fixed-point frequencies within a given spectrum, exhibited by an amplifier or other circuit that must transmit a band of frequencies.

flat line A transmission line in which there are no standing waves, or for which the standing-wave ratio is very low.

flat pack An integrated circuit package consisting of a square or rectangular flat housing, with pins projecting straight outward from the edges.

flat response A response characteristic in which the dependent variable is substantially constant over a specified range of values of the independent variable. For example, in amplifier operation, an output signal whose component fundamental frequencies and their harmonics are in the same proportion as those of the input signal being amplified.

flat-ribbon line A transmission line (feeder) consisting of two flexible conductors molded in a flexible, flat ribbon of plastic, such as polyethylene. Also called *twinlead.*

flattening The leveling-off or blunting of a normally peaked or curved response, often caused by signal saturation within a circuit. Sine-wave clipping is an example.

flat top **1.** The horizontal radiating portion of an antenna. **2.** See FLAT-TOP ANTENNA. **3.** Of an amplifier, to distort by clipping of the positive half-cycles.

flat-top antenna An antenna having a horizontal wire or wires at the top to lower the resonant frequency and increase the bandwidth.

flat-top beam A bidirectional, end-fire antenna consisting of two close-spaced dipoles center-fed out of phase. Also see KRAUS ANTENNA.

flat-topping The positive-peak clipping of a modulation envelope that occurs when an amplifier is overdriven or when a signal is overmodulated. This is in contrast to deliberate clipping. Flattopping results in "splatter" and is, therefore, an undesirable characteristic in an amplitude-modulated signal.

flat-top response The ability to uniformly transmit frequencies in a given band.

flat transmission line **1.** A transmission line that is free of standing waves. Also see MATCHED TRANSMISSION LINE. **2.** See FLAT-RIBBON LINE.

flaw An irregularity in a substance that can result in problems, such as mechanical failure or poor equipment performance.

flaw detector An instrument that uses ultrasonic waves to detect internal flaws in metal. The waves are reflected by flaws.

F layer See F REGION.

fLb Abbreviation of FOOT-LAMBERT.

Fleming/Kennelly law Under conditions at or near magnetic saturation in a ferromagnetic substance, the reluctivity is directly proportional to the intensity of the magnetic field.

Fleming's generator rule See FLEMING'S RIGHT-HAND RULE.

Fleming's left-hand rule A simple way of indicating certain relationships in the behavior of electric generators and motors. If the thumb, index finger, and middle finger of the left hand are positioned so that they are at right angles to each other, the thumb will point in the direction of force or motion when the index finger is pointed in the direction of flux; the middle finger points in the direction of current flow. Compare FLEMING'S RIGHT-HAND RULE.

Fleming's motor rule See FLEMING'S LEFT-HAND RULE.

Fleming's right-hand rule A simple way of indicating certain relationships in the behavior of electric generators and motors. If the thumb, index finger, and middle finger of the right hand are positioned so that they are at right angles to each other, the middle finger points in the direction of an induced voltage, the thumb in the direction of the motion of a conductor, and the index finger in the direction of the magnetic field. Compare FLEMING'S LEFT-HAND RULE.

Fletcher-Munson curves A set of curves depicting the uneven frequency response of human hearing. Also called AUDIBILITY CURVES.

flexible automation The ability of a robot or system to do various different tasks. Changing from one task to another is simply a matter of changing software. Compare FIXED-SEQUENCE ROBOT.

flexible collodion A viscous solution of pyroxylin (cellulose nitrates) used sometimes as a binder for coils.

flexible contact A contact made from flat, metal spring stock; it is usually bent or curved. Also called *spring contact.*

flexible coupling A device for joining two shafts and conveying rotary motion from one to the other; it is elastic, so the shafts need not be exactly aligned with each other.

flexible flat cable See FLAT CABLE.

flexible manufacturing system A roboticized manufacturing plant that can turn out a variety of different products. One or more central computers oversee the operation of the facility. Such factories are commonly used in the production of electronic devices (such as printed circuits, calculators, and portable radios).

flexible resistor An insulated, wirewound resistor that can be bent, coiled, or knotted.

flexible shaft A control shaft that can be bent somewhat while still allowing easy adjustment.

flex life A measure of how much bending a conductor or other flexible object can take without breaking.

flexode A diode that is flexible in that its junction can be changed (i.e., reversed without reversing its leads, its resistance being variable from the forward- to backward-resistance value).

flicker **1.** A tendency for a video image to appear, disappear, and reappear, or to increase and decrease in intensity frequently. **2.** The effect created by such action (as in a flickering light).

flicker frequency The number of times the screen illumination flashes on and off in the projection of a motion picture. It is 48 per second (twice the frame rate) in conventional movie projectors; for each frame, the screen is blanked once when the frame is pulled into position and once again during projection of the frame.

flight control Electronic monitoring and control of an aircraft in flight.

flight path The course planned for an aircraft's flight.

flight-path computer A computer that controls the course of an aircraft in flight, from takeoff to landing.

flight-path deviation The departure of an aircraft in flight from the course in the flight plan. Also see FLIGHT PATH.

flight-path-deviation meter An instrument that provides a visual indication of the departure of an aircraft in flight from the course in the flight plan.

flight telerobotic servicer A remotely controlled robot used to maintain and repair space vehicles and satellites. The machine can be controlled by a computer program or by a human operator. The most-sophisticated machines can use TELEOPERATION and/or TELEPRESENCE so that a human can perform dangerous work without being placed at personal risk.

flight test **1.** To test airborne electronic equipment in actual flight. **2.** Any test made as in **1**.

Flinders bar In a magnetic compass, a metal bar that corrects for the vertical component (inclination) of the earth's magnetic field. The bar must be designed differently in different geographic locations, because the inclination varies from place to place. Inclination is greatest near the geomagnetic poles, and is zero at the geomagnetic equator.

flint glass A hard, bright, lead glass. Dielectric constant, 7 to 9.9. Dielectric strength, 30 to 150 kV/mm. Also see GLASS.

flip chip A monolithic semiconductor device (such as a diode, transistor, or integrated circuit), in which bead-like terminals are provided on one face of the chip for bonding.

flip-chip bonding A scheme for making connections between a semiconductor chip and a header, in which leads are not run between chip and header. Instead, bead-like projections are electrodeposited as terminals around one face of the chip, which is then registered with the header terminals and bonded to them.

flip-flop **1.** See BISTABLE MULTIVIBRATOR. **2.** A two-position relay that locks in alternate positions upon receiving successive actuating pulses.

flip-flop key In a video display, a key that, when pressed, allows viewing of one half of the screen and then the other.

flip-flop memory A bistable computer memory that stores bits of data as flip-flop states.

flip-flop relay See BISTABLE RELAY.

float charging The constant charging of a storage battery, keeping the battery at or near the fully charged state at all times.

floated battery A storage battery connected in parallel with a generator, which supplies the load; the battery, always completely charged, helps during high-current demands.

floating **1.** To float a storage battery; see FLOAT. **2.** An ungrounded device or circuit that is not connected to a source of voltage. **3.** Not loaded or driven. **4.** Not fixed in position. **5.** A dedicated ground connection that remains isolated from the common circuit ground.

floating address See RELATIVE ADDRESS.

floating charge See TRICKLE CHARGE.

floating control **1.** A potentiometer, such as a gain control, installed with its shaft insulated from ground and, accordingly, subject to body-capacitance effects. **2.** A type of automatic control in which the rate of final control element movement depends on the amount that the controlled variable deviates from a prescribed value.

floating ground See FLOATING, **5**.

floating input An ungrounded input circuit.

floating-input measurement See DIFFERENTIAL-INPUT MEASUREMENT.

floating instrument An instrument whose signal terminals are above ground.

floating I/O port An input/output (I/O) terminal that is not loaded or being driven.

floating junction A junction (in a semiconductor device, for example), that has no net current flowing through it.

floating neutral A circuit with a variable common voltage reference.

floating paraphase inverter A dual-transistor adaptation of the paraphase inverter. The second stage receives its input signal from a tap on the load resistor of the first stage and provides the additional phase shift that is required.

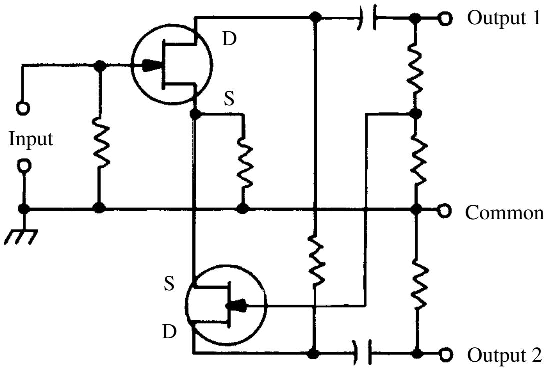

floating paraphase inverter

floating-point calculation An electronic calculation using a floating point number [i.e., a number whose value is represented by two sets of digits, a fixed-point part (see FIXED-POINT SYSTEM) and a radix (base number) with an exponent]. Important in the operation of microprocessors, especially in computer systems.

floating-point number A number expressed in the floating-point system.

floating-point package Computer-vendor software that enables that computer to perform floating-point calculations.

floating-point system A system of notation in which a number n is represented by two sets of numbers: a fixed-point part (see FIXED-POINT SYSTEM) a, the radix (base number) r, and an exponent b as follows: $n = a \times r^b$. For example, in the floating-point system, 623 can be written 6.23×10^2. Floating-point numbers can be stored economically (in terms of memory) and in magnitudes that might otherwise be beyond the capacity of the computer to operate upon with relatively consistent accuracy. Compare FIXED-POINT SYSTEM.

floating probe A test electrode (wire or plate) inserted in a discharge tube at a desired point to sample the potential gradient, but which acquires a misleading negative charge, with respect to the gas cloud, because electrons (traveling faster than the positive ions) tend to accumulate on the probe.

floating zero A control system in which the reference point is easily moved.

floating zone In a semiconductor ingot undergoing purification, a molten zone in which impurities float. The material in the zone is melted by the radio-frequency (RF) field of an external heating coil, which is passed along the ingot to move the molten zone to one end, picking up impurities along the way and concentrating them at the end that is later sawed off.

float switch A switch operated by a float, such as in a sump pump.

flocking **1.** Particulate felt used on phonograph turntables to protect disks from being scratched. **2.** To coat with flocking.

flood gun In a storage (image-holding) oscilloscope, the electron gun that sprays the storage target with low-velocity electrons and makes the image visible on the viewing screen. The gun is mounted next to one pair of deflection plates. Compare WRITING GUN.

floor stand A support for a microphone, consisting of a heavy base that rests on the floor, and an adjustable, vertical boom that allows the microphone to be set at various heights.

floppy disk A flexible magnetic disk used in recording, as in computer and data system storage. It usually refers to a 5.25-inch diskette.

flow **1.** The movement of current carriers under the influence of an electric field. **2.** See ANGLE OF CONDUCTION. **3.** A series of interrelated events in a time sequence.

flow angle See ANGLE OF CONDUCTION.

flowchart **1.** A diagram depicting the logic steps in a digital-computer program. **2.** A diagram showing the flow of material through a sequence of processes.

flow direction The method of delineating antecedent and successor events on a flowchart; usually arrows and flowlines connecting the events in the way a page is read (top to bottom, left to right).

flowed-wax disk A form of recording disk, in which wax is melted onto a plastic or metal base. The grooves are cut in the wax layer.

flowline A line showing flow direction on a flowchart.

flowmeter An instrument for measuring liquid flow rate.

flow relay A relay that is actuated by a predetermined rate of fluid flow.

fluctuating current See COMPOSITE CURRENT.

fluctuating voltage See COMPOSITE VOLTAGE.

fluid absorption See LIQUID ABSORPTION.

fluid analogy The comparison of electric current flow to the movement of a simple fluid. Also see WATER ANALOGY.

fluid capacitor See WATER CAPACITOR.

fluid computer A digital computer that uses fluid logic elements (i.e., one that contains no electronic circuits or moving parts).

fluid damping Use of a viscous fluid to damp a mechanical member's movement.

fluid-flow alarm An electronic circuit that actuates an alarm when fluid flowing through pipes or other channels changes from a predetermined rate.

fluid-flow control A servo system that automatically maintains or adjusts liquid flow through pipes or other channels.

fluid-flow gauge See FLUID-FLOW METER.

fluid-flow indicator See FLUID-FLOW METER.

fluid-flow meter An instrument that indicates fluid flow rate through pipes or other channels.

fluid-flow switch In a fluid-cooled system, a switch that actuates an alarm when the fluid slows or stops.

fluidics **1.** A form of digital logic in which circuits operate by means of fluid flow. **2.** A branch of physics concerned with the behavior of fluids; more commonly called *fluid dynamics.*

fluid-level control A servo system that automatically maintains the level of a fluid in a tank.

fluid-level gauge An electronic system that provides direct readings of the level of a fluid in a tank.

fluid-level indicator See FLUID-LEVEL GAUGE.

fluid logic Logic operations carried out by varying the flow and pressure of a gas or liquid in a circuit of channels. Also see FLUID COMPUTER.

fluid ounce (U.S.) Abbreviation, fl. oz. A unit of volume equal to 2.957×10^{-5} cubic meters, or 0.02957 liter. A quart is 32 ounces; a gallon is 128 ounces.

fluid-pressure alarm An electronic circuit that actuates an alarm when fluid pressure rises or falls beyond set limits.

fluid-pressure control A servo system that automatically maintains or adjusts fluid pressure in pipes or other channels.

fluid-pressure gauge See FLUID-PRESSURE METER.

fluid-pressure indicator See FLUID-PRESSURE METER.

fluid-pressure meter An instrument that indicates the pressure of a fluid in a pipe or other channel.

fluid valve See ELECTROMECHANICAL VALVE.

fluorescence The property of some materials to glow when excited by a stimulus, such as ultraviolet, X rays, or an electron beam. Compare PHOSPHORESCENCE.

fluorescent lamp See FLUORESCENT TUBE.

fluorescent materials Materials that glow when irradiated, but cease to glow when the source of excitation is removed. An example is the phosphor coating on the screen of a cathode-ray tube (CRT).

fluorescent screen A transparent or translucent plate (such as the end of a cathode-ray tube or fluoroscope) coated with phosphors that glow when struck by an electron beam, or by high-energy electromagnetic radiation, such as ultraviolet or X rays.

fluorescent tube A mercury-vapor glow lamp distinguished by having a glass tube whose inner wall is coated with a phosphor that emits light when excited by the ultraviolet glow discharge in the vapor.

fluorescent X rays X rays reradiated by the atoms of a material that has absorbed X radiation. During initial exposure, energy absorbed from the radiation raises the energy level of electrons in the atoms; when the electrons return to their normal energy levels, they reradiate some of the absorbed energy.

fluorine Symbol, F. A gaseous element of the halogen family. Atomic number, 9. Atomic weight, 18.998.

fluoroscope A device used for viewing the internal structures of objects. A screen coated with material that fluoresces when exposed to X rays is mounted in one end of a light-tight viewing hood. When an object is placed between the screen and an X-ray tube, an image is produced on the screen. In medical applications, this device has been supplanted by methods that do not use ionizing radiation; *nuclear magnetic resonance imaging (NMRI)* is one example.

fluoroscopy The art of using a fluoroscope in the inspection of materials and parts or in medical examinations.

flush A form of mounting in which there is little or no protrusion from the panel surface.

flutter **1.** In a high-frequency superheterodyne receiver, a rapid fluctuation in signal strength, caused by tuning and detuning of the oscillator stage. This usually results from poor direct-current (dc) power-supply regulation. **2.** Repetitive, rapid fluctuations in the output of a sound reproducer. Also see WOW. **3.** An echo effect sometimes observed in rooms or auditoriums of poor acoustic design.

flutter bridge A bridge-type instrument for measuring flutter in constant-speed machines, such as sound recording and reproducing devices.

flutter rate The frequency of flutter, in cycles per second (Hertz).

flux **1.** Theoretical lines of force that extend in all directions from an electric charge (electric flux) or from a magnetic pole (magnetic flux). **2.** A material that makes metals more amenable to being joined by soldering. **3.** The number of photons that pass through a surface for a given time.

flux density Symbol, *B*. Unit, tesla. The degree of concentration of magnetic lines of force. One tesla represents a flux density of one volt-second per square meter (V•s/m^2).

fluxgate A device that controls the azimuth bearing of a directional system by means of interaction with the geomagnetic field.

fluxgate magnetometer A magnetic compass for robot guidance. Uses coils to sense changes in artificially generated reference fields. Output from the sensors is sent to a computer that calculates the robot's position, based on the orientation and intensity of the lines of flux in the reference fields.

flux graph A device that graphically records the intensity of a magnetic field around a permanent magnet or electromagnet, or around an inductor carrying a current.

flux leakage See MAGNETIC LEAKAGE.

flux lines The theoretical lines of force in an electric or magnetic field.

flux linkage The passage of lines of force set up by one component through another component, so as to enclose most of the penetrated component's volume.

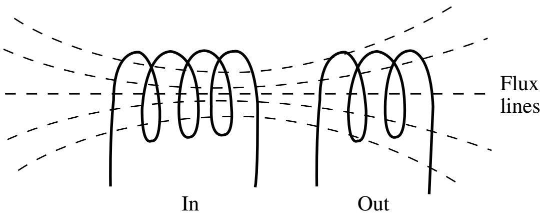

flux linkage

fluxmeter An instrument for measuring magnetic flux density. Also called *gaussmeter.*

flux refraction The tendency for magnetic lines of flux to change direction at the boundary between substances having different permeability. Flux refraction resembles refraction of electromagnetic radiation in or at a boundary between substances having different indices of refraction.

flyback **1.** The abrupt drop or reversal of a current or voltage that was previously increasing (e.g., the rapid fall of a sawtooth wave). Also see KICKBACK. **2.** The duration of the drop of a current or voltage that was previously increasing, for a sawtooth or similar wave. **3.** In an oscilloscope or picture tube, the rapid return of the beam to its starting position.

flyback checker An apparatus that senses the presence of short or open circuits in motors, transformers, and generators, by measuring the amount of flyback (kickback).

flyback power supply See KICKBACK POWER SUPPLY.

flyback time The time taken for the electron beam in an oscilloscope tube, picture tube, or camera tube to return to its starting point after it has reached the point of maximum deflection.

flyback transformer In a television receiver circuit, the horizontal output transformer. The unit supplies horizontal scanning voltage and kickback voltage, which is rectified to produce the high-voltage direct-current (dc) anode potential. Also see FLYBACK and KICKBACK POWER SUPPLY.

flying eyeball An undersea exploration robot consisting of a television camera, illumination lamps, and thrusters (such as jets or propellers). A cable, which also serves as a tether, sends data to a human operator, and allows the operator to control the movements of the robot. In some cases, the tether/cable can be replaced by an infrared or visible-light laser data link.

flying-spot tube A tube, such as a camera tube, in which a rapidly deflected spot of light scans an image on a transparent screen; the spot is projected through the picture to a photomultiplier.

fly's-eye lens A lens consisting of hundreds of much smaller lenses. Used in microelectronic circuit fabrication to produce many images of the same circuit.

flywheel effect **1.** In an inductance-capacitance (LC) tank circuit, the completion of a partial input wave cycle at the resonant frequency, resulting from the storage and release of energy. This provides a nearly perfect sine-wave output for class-AB, class-B, and class-C radio-frequency (RF) power amplifiers. **2.** In an LC tank circuit, the action in which energy continues to oscillate between the capacitor and inductor after an input signal has been removed. The oscillation stops when the tank-circuit finally loses the energy absorbed. The lower the inherent resistance of the circuit, the longer the decrement (decay time).

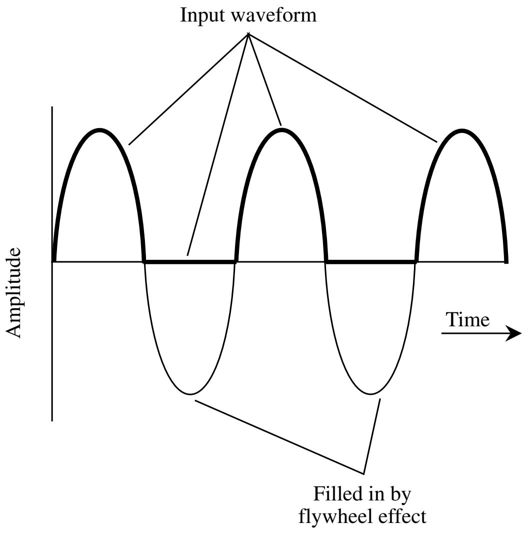

flywheel effect, 1.

flywheel synchronization A form of television scanning synchronization used when the received signal is very weak. The synchronization signals from the transmitter are sensed by the receiver, which then produces its own local pulses based on the rate of received pulses.

flywheel tuning A tuning dial mechanism in which the control shaft has a flywheel for the smoother tuning action afforded by the added momentum.

Fm Symbol for FERMIUM.

FM Abbreviation of FREQUENCY MODULATION.

f_m Abbreviation of MODULATION FREQUENCY.

FM-AM Pertaining to equipment that will operate with either amplitude-modulated or frequency-modulated signals.

FM-AM multiplier A method of frequency multiplication using both amplitude and frequency modulation of a carrier wave.

FM broadcast band The 88- to 108-MHz frequency band, within which channels spaced 200 kHz apart occupy positions from 88.1 to 107.9 MHz.

FM detector See DISCRIMINATOR, RATIO DETECTOR, and SLOPE DETECTOR.

FM-FM Frequency modulation by one or more FM subcarriers.

FM limiter In a frequency-modulation circuit, a stage which holds the amplitude of the FM signal to a constant value. The limiter can be active (e.g., an amplifier-limiter transistor) or passive (e.g., a diode clipper).

FM multiplex See MULTIPLEX ADAPTER.

FM noise Unintentional modulation of a frequency-modulated transmitter, resulting from noise in the audio-input stages.

FM-PM A system of modulation in which a carrier is phase modulated by frequency-modulated subcarriers.

FM radar A radar system in which the signal is frequency modulated; the distance to the target is measured in terms of the beat note between transmitted and reflected waves.

FM repeater A two-way radio system composed of a simultaneously operating receiver and transmitter, the latter of which retransmits (usually on a different frequency) all signals picked up by the receiver. The system is usually tower- or hilltop-mounted, and is used to extend the range of two-way units in a communications network.

FM stereo The use of multiplex methods to transmit and receive stereophonic programs in an FM channel. Also see MULTIPLEX ADAPTER.

FM tuner A compact radio receiver that handles frequency-modulated (FM) signals, and delivers its low-amplitude audio output to a high-fidelity system. Compare AM TUNER and AM-FM TUNER.

focal length Symbol, F. The distance from the center of a lens or dish antenna to the principal focus. Also see PRINCIPAL FOCUS.

focus **1.** The point at which rays converge. Also see PRINCIPAL FOCUS. **2.** To bring rays to a point of convergence.

focus coil See FOCUSING COIL.

focus control In an oscilloscope or television circuit, the potentiometer that controls the voltage on the focusing electrode of the cathode-ray tube and, accordingly, the sharpness of the image.

focus grid **1.** The focusing electrode in an electrostatic cathode-ray tube. **2.** The focusing electrode in an oscilloscope tube.

focusing Bringing a ray of particles or energy to a common point. This can be done using lenses, deflecting coils, deflecting plates, or reflecting devices. Focusing can be done with acoustic waves, electromagnetic waves, and theoretically with any kind of disturbance propagated through any medium.

focusing anode See FOCUSING ELECTRODE.

focusing coil An external coil used to focus an electron beam in a cathode-ray tube. Also see ELECTROMAGNETIC FOCUSING.

focusing electrode The internal electrode (grid or ring) used to focus the electron beam in a cathode-ray tube. Also called focus electrode. Also see ELECTROSTATIC FOCUSING and FOCUS GRID, **1**, **2**.

focusing magnet A permanent magnet assembly for focusing the electron beam in a cathode-ray tube.

foil **1.** The thin conductive strips on a printed-circuit board. **2.** Also known as *tape.* Thin metal supplied in strips, intended for use in certain security systems. It can be installed in closed loops at potential points of entry.

foil capacitor A capacitor whose plates are sheets or strips of metal foil separated by a dielectric film.

foil coil See FOIL-WOUND COIL.

foil conductor A conductor that is a strip of metal foil, rather than wire. Also see FOIL PATTERN.

foil electroscope See LEAF ELECTROSCOPE.

foil pattern The pattern of thin metal circuit paths that constitute the "wiring" of a printed circuit. Also see ETCHED CIRCUIT and PRINTED CIRCUIT.

foil-wound coil A coil wound with metal foil (usually aluminum or copper) instead of wire. Such coils substantially reduce the weight of large transformers and filter chokes.

foldback **1.** See FOLDBACK CURRENT LIMITING. **2.** In audio recording, the routing of sound (via an audio mixer) to some other destination in addition to the recording medium. Example: playing an electronic organ while singing, recording the arrangement on tape, and also listening to it (organ and voice) in a headset.

foldback current limiting In a power supply, a method of automatically reducing the output current to a safe level when the load current exceeds the maximum recommended value. This action protects both the power supply and the powered equipment.

folded dipole A half-wavelength, center-fed antenna constructed of two parallel wires with their ends connected together. It has the same length as a simple dipole antenna, but the feed-point impedance is four times that of the ordinary dipole. Instead of approximately 73 ohms, the folded dipole presents a resistive impedance of about 300 ohms. This makes the folded antenna desirable for use with high-impedance, parallel-wire transmission lines. It also can be used to obtain a good match with 75-ohm coaxial cable when four antennas are connected in phase, or with 50-ohm coaxial cable when six antennas are connected in phase. Compare DIPOLE ANTENNA.

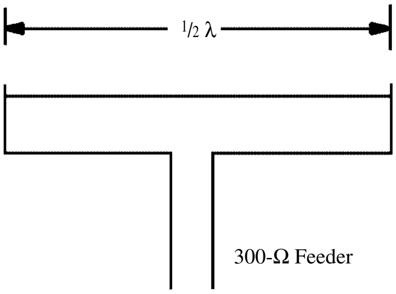

folded dipole

folded horn A loudspeaker having a horn whose flare is divided into several zigzagging chambers; that is, the horn is, in effect, folded to squeeze a required length into a small cabinet.

folded-horn enclosure See LABYRINTH SPEAKER.

folded pattern An oscilloscope image having an elongated time axis obtained by successive horizontal sweeps—each placed slightly lower on the

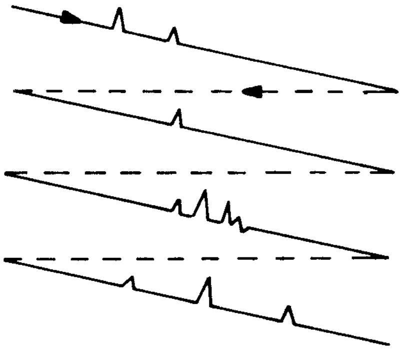

folded pattern

screen than the preceding one. The folded-pattern technique provides a time axis several times longer than the screen width.

folding frequency In a system where sampling is made at uniform frequency increments, the frequency corresponding to half the sampling rate in hertz.

foldover Distortion characterized by the horizontal or vertical overlapping of a television picture.

follower A single-stage, active circuit characterized by zero phase reversal, and voltage gain less than unity. The *emitter follower* is also called a *common-collector circuit*; the *source follower* is also known as a *common-drain circuit.* Characterized by moderate to high input impedance, and low output impedance over a wide band of frequencies.

follower drive In a servo system, the drive that mechanically follows the master drive.

following blacks In a television picture, the effect in which a moving white object has a black border following it.

following whites In a TV picture, the effect in which a moving dark object has a white border following it.

follow-up motor See SERVOMOTOR.

font The physical shape and size of the letters and numbers in an alphanumeric system.

font reticle In optical character recognition, an overlay reference pattern of lines used to check the size and configuration of an input character, the size of punctuation marks, and spacing between lines and characters.

food-service robot Any robot that is used for the purpose of packaging, preparing, and/or serving food.

foot Abbreviation, ft. A unit of linear measure in the English system equal to 0.3048 meter.

foot-candle Abbreviation, fc. A unit of illuminance; 1 fc is the amount of direct light emitted by 1 candela (see CANDLEPOWER) that falls on 1 square foot of a surface on which every point is 1 foot away from the source. In the International System of Units, the unit is *lux* (lumens per square meter). Compare METER-CANDLE.

foot-candle meter A light meter whose scale reads directly in foot-candles.

foot-lambert Abbreviation, fL. A unit of luminance; the average brightness of a surface that emits or reflects 1 lumen per square foot. The Standard International (S.I.) unit is the *candela per square meter* (cd/m^2); 1 fL = 3.426 cd/m^2.

foot-pound Abbreviation, ft-lb. In the English system, a unit of energy equal to 1 pound displaced through 1 foot in the direction of the exerting force. The Standard International (S.I.) unit is the *joule* (j); 1 ft-lb = 1.356 j.

foot-pound-second system See FPS SYSTEM OF UNITS.

foot switch A switch operated by the foot, generally used for the purpose of turning a playback or recording system on and off. Often used for taking dictation.

forbidden band See ENERGY GAP.

forbidden character code An error-finding code using forbidden characters: combinations of prohibited bits. Also called *forbidden combination.*

forbidden energy band See ENERGY GAP.

force **1.** Symbol, F. Units: newton, dyne, poundal. The agency or influence that accomplishes work. **2.** An operator interjection made during a program run that causes the computer to execute a branch instruction; it is usually necessary when a condition responsible for halting a program must be bypassed.

forced coding Programming that minimizes the time required to retrieve information from storage. Also called *minimum latency programming* or *minimum access programming.*

forced oscillations Oscillations in a circuit, such as in an inductance-capacitance (LC) tank, that result from continuously applied alternating-current (ac) excitation. Compare FREE OSCILLATIONS.

foreground job A relatively high-priority, short-running program that is carried out by interrupting a low priority, long-running program. Compare BACKGROUND JOB.

force pump In a multistage vacuum system, the first pump that reduces the pressure considerably below atmospheric pressure. Also see DIFFUSION PUMP and VACUUM PUMP.

force summing device A transducer element that is physically moved by a force being transduced.

foreshortened addressing In control computers, the mixing of available storage by using simplified addressing instructions.

fork oscillator An audio-frequency oscillator controlled by a tuning fork. The dimensions of the fork determine its vibration frequency and, accordingly, the frequency of the oscillator.

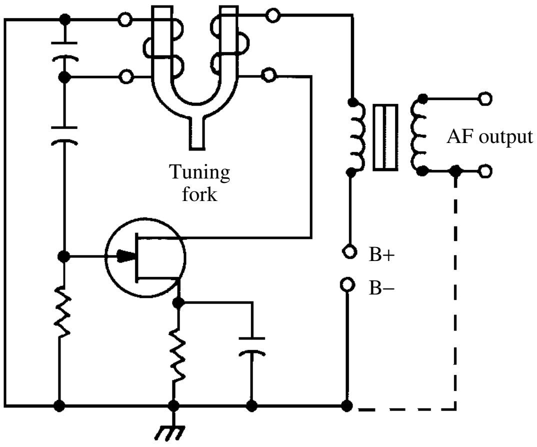

fork oscillator

form **1.** The core or frame upon which an inductor is wound. **2.** A vessel, such as a mold, used in the shaping stage of a manufacturing process.

formaldehyde Formula, HCHO. A colorless, pungent gas that is a constituent of many well-known plastic insulating materials (see PHENOLFORMALDEHYDE PLASTICS).

formant **1.** The audio-frequency range in which the sound of a spoken syllable is concentrated. **2.** Any general group of audio frequencies.

formant filter In an electronic organ, an audio filter that changes the waveshape of a tone so that the tone will have the desired characteristics.

format **1.** The form in which data is presented (e.g., the arrangement of characters, fields, words, totals, etc.). **2.** To prepare a computer disk or tape so that it will accept data.

form factor **1.** The SHAPE FACTOR for a filter or tuned circuit. **2.** For a half-cycle of an alternating-current (ac) quantity, the ratio of the root-mean-square (rms) value to the average value.

form feed **1.** A mechanical system that positions paper being supplied to a line printer. **2.** The FF character that initiates advancement of printout paper in a printer. **3.** The advancement of printout paper in a printer.

form feed character In a control loop, a character (symbol, FF) used on printing devices for controlling form feed.

forming See ELECTROFORM, **1**.

form stop An automatic device that stops a printer when the paper runs out.

FORTH A high-level computer programming language used in certain robots, automated factories, medical electronic devices, and electronic games. It was originally developed in the 1970s to facilitate computer control of equipment in astronomical observatories.

FORTRAN A high-level computer programming language developed in the 1950s, and still used in some scientific and mathematical applications. It is not especially useful for the control of electronic or mechanical devices.

fortuitous conductor A medium that creates an unwanted electrical path.

fortuitous distortion Waveform distortion that results from causes other than characteristic effects or bias effects.

forward AGC Automatic gain control provided by special transistors whose transconductance decreases with increasing emitter current, and vice versa. Compare REVERSE AGC.

forward-backward counter A counter that runs forward to perform addition and backward to perform subtraction.

forward bias Forward voltage or current in a transistor or semiconductor diode.

forward-blocking state For a silicon-controlled rectifier, the off state, during which the forward bias is so much less than the forward breakover voltage that only small off-state current flows.

forward breakover voltage **1.** For a semiconductor pn junction, the smallest forward voltage at which appreciable conduction occurs. This is about 0.3 V for germanium and 0.6 V for silicon. **2.** For a silicon-controlled rectifier, the forward voltage value at which the device abruptly switches on.

forward characteristic The current-voltage response of a semiconductor junction that is biased in the forward (high-conduction) direction. Compare REVERSE CHARACTERISTIC.

forward compatibility standards Standards developed to make programs for one computer system usable for additional or replacement equipment.

forward conduction The increased current conduction through a pn junction that is forward biased. Compare REVERSE CONDUCTION.

forward current Symbol, I_f. The increase in current flow through a pn junction that is forward biased. Compare REVERSE CURRENT.

forward current-transfer ratio The current gain of a bipolar transistor (alpha for the common-base connection and beta for the common-emitter connection).

forward power **1.** In a transmission line, the power leaving the generating source, as measured by a directional wattmeter at that location. **2.** The power arriving at the load at the terminating end of a transmission line.

forward propagation by ionospheric scatter See FORWARD SCATTER.

forward propagation by tropospheric scatter Abbreviation, FPTS. A method of transmitting part of a radio signal beyond the horizon using the scattering effect of the troposphere. Also see FORWARD SCATTER and TROPOSPHERE.

forward resistance Symbol, R_f. The resistance of a forward-biased pn junction. Also see FORWARD BIAS. Compare REVERSE RESISTANCE.

forward-reverse ratio See FRONT-TO-BACK RATIO.

forward scatter Also called *forward propagation by ionospheric scatter.* The scattering of a radio wave in the normal direction of propagation to points beyond the skip zone. The phenomenon occurs because of waves returned from regions in the ionosphere. Compare BACK SCATTER.

forward transconductance Symbol, g_{fs}. For a common-source-connected FET, the ratio of a drain-current differential to the differential of gate-to-source voltage that produces it; $g_{fs} = 1000(dI_D/dV_{GS})$, where g_{fs} is in microsiemens, I_D is the drain current in milliamperes, and V_{GS} is the gate-to-source voltage in volts.

forward voltage Symbol, E_f or V_f. Voltage whose polarity causes maximum current to flow through a pn junction. Compare REVERSE VOLTAGE.

forward voltage drop The voltage across a semiconductor junction that is biased in the forward (high-conduction) direction. Compare REVERSE VOLTAGE DROP.

FOSDIC Acronym for *film optical scanning devices for input to computer.*

Foster-Seeley discriminator A discriminator circuit in which the diodes are operated from a single-tuned, center-tapped secondary of the input transformer. The center tap is also capacitively coupled to the top of the transformer's primary coil. Compare TRAVIS DISCRIMINATOR.

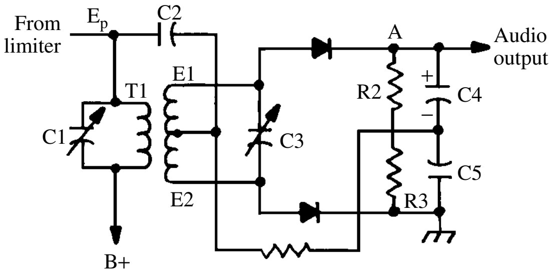

Foster-Seeley discriminator

Foucault currents See EDDY CURRENTS.

four-address instruction A computer instruction in which the address is comprised of four addresses: two for operands, one for the result of the operation, and one for the upcoming instruction.

four-channel sound system Also called *quadraphonic sound system.* A high-fidelity, stereophonic sound reproduction system, in which there are four channels, rather than the usual two. The channels are generally designated left front, left rear, right front, and right rear. The four-channel system is an enhancement of two-channel stereo.

four-dimensional continuum 1. In relativistic theory, the space-time continuum. There are three spatial dimensions and one time dimension. A point in the continuum can be uniquely defined by three space coordinates and one time coordinate. **2.** Any continuum that requires four and only four coordinates to uniquely determine the position of a point.

Fourier analysis Use of the FOURIER SERIES to evaluate the components of a complex wave, or to define a complex wave in terms of its components.

Fourier series An infinite mathematical series that shows any periodic function to be a combination of sine terms and cosine terms. Any complex wave (e.g., a square wave) consists of fundamental and harmonic sine-wave components. In simplified form, the series is:

$$f(x) = A_0/2 + A_1 \cos x + B_1 \sin x + A_2 \cos 2x + B_2 \sin 2x + A_3 \cos 3x + B_3 \sin 3x + \ldots$$

In general, the more terms to which the series is calculated, the better the approximation of a complex waveform.

four-layer diode A dual-terminal npnp device that is usable as a bistable switch, sawtooth or pulse generator, memory device, etc.

four-layer transistor A transistor in which the wafer or block has four processed regions; however, the device might have only three terminals. Some examples are the silicon-controlled rectifier, silicon-controlled switch, and thyristor.

four-level laser A laser identical to the three-level laser, except for the addition of one excited state.

four-phase system A two-phase system in which the center taps of the coils are interconnected. Also called QUARTER-PHASE SYSTEM.

four-terminal network A network having two input terminals and two output terminals. One input terminal can be internally connected to one output terminal (as when a common ground is present), but this is not mandatory.

four-space 1. A mathematical space in which four coordinates (w,x,y,z) are necessary to uniquely define a point. **2.** Three-dimensional space with the addition of time as a fourth dimension; coordinates are, for example, (x,y,z) for space and t for time.

four-space coordinates 1. A system of coordinates for uniquely determining points in four-space. Such a system can be Euclidean or non-Euclidean, as with space of any number of dimensions. **2.** The set of numbers that defines a particular point uniquely in four-space; for example, $P = (3,-15,0,-7)$.

four-sphere The set of all points equidistant from a given point P in four-space. Formula is $w^2 + x^2 + y^2 + z^2 = r^2$, where r is the radius and the coordinates are (w,x,y,z) in the Euclidean, Cartesian system.

four-track recording A tape recording in which four channels are recorded in two adjacent tracks on the tape. Usually, tracks number 1 and 3 are in the forward direction, and tracks number 2 and 4 are in the reverse direction.

four-track tape A magnetic tape with four parallel sound paths.

four-wire wye system A three-phase system in which three wires supply the respective phases, a fourth being the neutral conductor.

F_p Symbol for POWER-LOSS FACTOR.

fp Abbreviation of FREEZING POINT.

FPC Abbreviation of *Federal Power Commission.*

FPI Abbreviation of FUEL-PRESSURE INDICATOR.

FPIS Abbreviation of *forward propagation by ionospheric scatter.* See FORWARD SCATTER.

fpm Abbreviation of *feet per minute.*

fps 1. Abbreviation of *feet per second.* **2.** Abbreviation of *frames per second.* **3.** Abbreviation of *foot-pound-second (fps),* a chiefly British system of units.

fps system of units The British system of units of measurement that uses the foot for length, the pound for mass, and the second for time. Compare CENTIMETER-GRAM-SECOND and INTERNATIONAL SYSTEM OF UNITS.

FPTS Abbreviation of FORWARD PROPAGATION BY TROPOSPHERIC SCATTER.

Fr Symbol for FRANCIUM.

fr Abbreviation of FRANKLINE.

fractional exponent An exponent indicating that a number is to be raised to a fractional power (e.g., $10^{4/3}$). The numerator of the exponent indicates the power to which the base number must be raised; the denominator of the exponent indicates the root that must be taken of the result. Thus, *na/b* is equal to the *b*th root of *na*.

fractional gain Amplification less than unity. A notable example is the transfer function of a source follower or emitter follower.

fractional horsepower Any power rating lower than one horsepower (1 hp). Also see HORSEPOWER.

fractional uncertainty See RELATIVE UNCERTAINTY.

frame **1.** A single, complete video image, scanned in 1/30 second in conventional television receivers. **2.** A single motion-picture (film) image. **3.** In packet communications, a fundamental unit of data. The three types are called *information (I) frame*, *supervisory (S) frame*, and *unnumbered (U) frame*. **4.** One of a recurring cycle of pulses. **5.** In pulse-code modulation (PCM), a cyclic word group including a sync signal. **6.** A complete commutator cycle. **7.** A digital representation of a set of objects, useful in robotics and artificial intelligence (AI).

frame alignment The condition in which the receiver, or receiving apparatus, is in correct alignment with the signal to be received. In television, for example, this results in true rendition of the picture. Incorrect frame alignment (misalignment) might result in the picture being split with the top and bottom interposed. For other types of signals, misalignment would result in garbled reception.

frame-alignment signal In television, a transmitted signal that is used to ensure that frame alignment occurs in the receiver. It is a form of synchronizing pulse.

frame-alignment time slot In a transmitted television frame, an interval of time that is used for the purpose of transmitting a frame-alignment signal. There might or might not be other signal information transmitted during this time interval.

frame frequency The number of frames of a motion-picture film that come into position per unit of time in a camera, projector, or pickup.

frame of reference Geometric relationships used to describe the location of a body in space.

frame rate See FRAME FREQUENCY.

frame-repetition rate See FRAME FREQUENCY.

frame roll Momentary vertical roll in a television picture.

frame synchronizing signal **1.** In pulse amplitude modulation (PAM), a coded pulse indicating initiation of a commutation frame. **2.** In pulse-code modulation (PCM), a signal used to identify an information frame.

framing **1.** Synchronization of the vertical component of a video signal so that the top and bottom of the transmitted and received pictures line up. **2.** The process of lining up the top and bottom of a movie picture. **3.** Alignment of the characters in a digital alphanumeric transmission.

francium Symbol, Fr. A radioactive metal element of the alkali-metal group. Produced artificially through radioactive disintegration. Atomic number, 87. Atomic weight, 223.

Frankenstein scenario A theme often depicted in science fiction, in which intelligent machines seize power from their human creators. With the rapid advance of technology, especially in robotics and artificial intelligence (AI), some people, including a few educated researchers, believe this scenario is within the realm of possibility. Most scientists think it is highly improbable.

Franklin antenna A vertical collinear array that produces omnidirectional gain because of phasing among the individual components.

frankline (Benjamin Franklin, 1706–1790) Abbreviation, fr. A name that has been suggested for the unit of electric charge; 1 fr is the charge that exerts a force of 1 dyne on an equal charge at a distance of 1 centimeter in a vacuum.

Franklin oscillator A dual-terminal, audio/radio-frequency (AF/RF) oscillator circuit. Consists of a two-stage, resistance-capacitance (RC) coupled amplifier, with a tuned inductance-capacitance (LC) tank in the input gate circuit, and with capacitive feedback from the second drain to the tank.

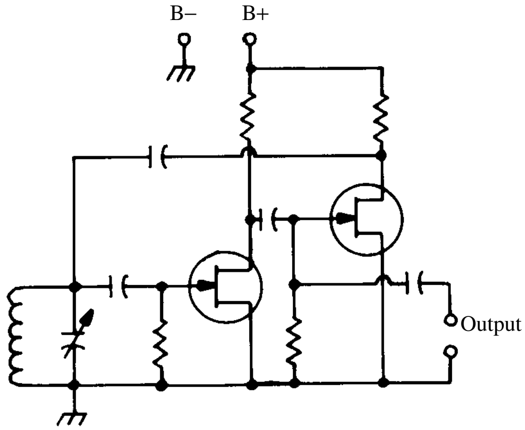

Franklin oscillator

Fraunhofer region The area surrounding a radiating antenna, throughout which the energy appears to come from a single point located near the actual antenna.

free air resonance For a speaker, the resonant frequency or frequencies exhibited when the device is not mounted in a cabinet.

free carrier A free electron or, in a semiconductor material, the equivalent hole. Also see ELECTRON and HOLE.

free charge The portion of a charge on a conductor that, being unaffected by a neighboring charge, will escape to ground when the conductor is grounded. Compare BOUND CHARGE.

free electron **1.** An electron situated in one of the outer orbits of an atom, held loosely by the nucleus. Because free electrons can easily escape the attraction of atomic nuclei, they will drift among atoms if the material is subjected to an electric potential. The result is electric current. Also see ELECTRON and BOUND ELECTRON. **2.** An electron that is not associated with any atomic nucleus.

free field Data organized in a storage medium in such a way that a data item or field can be anywhere in the medium. Compare FIXED FIELD.

free impedance For a transducer, the input impedance produced by a perfectly short-circuited load.

free magnetic pole A magnetic pole that is so well isolated from its opposing pole that it experiences little or no influence from the latter.

free magnetism A theoretical medium or fluid to which magnetic effects are conventionally given. The sum of free magnetism in any given object is always zero. Within any small part of the field, the free magnetism is thought of as flux lines. This theoretical medium can be any nonmagnetic material.

free net In radio communications, a network in which stations are free to communicate with other stations in the net without constant supervision by the net control station. Such communication is carried out on a frequency slightly above or below that of the net's formal operation.

free oscillations Oscillations in a circuit, such as an inductance-capacitance (LC) tank, that continue after excitation has been removed. Also see FLYWHEEL EFFECT. Compare FORCED OSCILLATIONS.

free path In a gas tube, the path taken by an electron as it collides with atoms. Also see MEAN FREE PATH.

free-power supply **1.** A simple tuned radio-frequency (RF) detector diode, used to rectify a radio signal and supply small amounts of direct current for the operation of low-powered transistor circuits. **2.** See SOLAR BATTERY.

free reel The supply reel of a magnetic-tape recorder.

free-running frequency The frequency at which a synchronized generator, such as a multivibrator or self-excited oscillator, will operate when the synchronizing voltage is removed.

free-running multivibrator See ASTABLE MULTIVIBRATOR and UNCONTROLLED MULTIVIBRATOR.

free space Empty space; a theoretical ideal.

free-space loss Radio transmission loss disregarding variable factors (a theoretical condition).

free-space pattern The ideal directivity pattern of an antenna that is situated many wavelengths above ground. In use, this pattern is modified by reflections from ground.

free speed The angular velocity of an unloaded motor.

free-standing display In a computer system, a remote display unit for prompting peripheral operators.

freezing point Abbreviation, fp. The temperature at which a liquid starts becoming a solid at normal pressure. Compare MELTING POINT.

F region Also called *F layer.* A region of the ionosphere with an altitude at night of approximately 175 miles. In daytime, the region splits into the lower *F1 region* and the higher *F2 region.* This layer is primarily responsible for long-distance propagation of radio waves at high frequencies (3 MHz to 30 MHz). At times it returns waves at frequencies as high as about 70 MHz.

Fremodyne detector A frequency-modulation (FM) detector that is essentially a conventional amplitude-modulation (AM) circuit detuned to one side of resonance (slope-tuned) to demodulate a frequency-modulated signal. Also see SLOPE DETECTOR.

French phone See CRADLEPHONE.

freqmeter Contraction of FREQUENCY METER.

frequency Symbol, f. The rate at which a phenomenon is repeated. The basic unit of frequency is the *Hertz (Hz)*, which represents one complete cycle per second. Common units encountered in electronics are the *kilohertz (kHz)*, *megahertz (MHz)*, and *gigahertz (GHz)*, where 1 kHz = 10^3 Hz, 1 MHz = 10^6 Hz, and 1 GHz = 10^9 Hz. Occasionally, the *terahertz (THz)* is used; 1 THz = 10^{12} Hz.

frequency-agile radar A radar system in which the transmitter frequency is shifted in a predetermined pattern for the purpose of avoiding detection. A frequency-agile radar system, with a complex frequency control program, is very difficult to jam.

frequency allocation **1.** The assignment of frequencies to radio and allied services by the licensing authority (in the United States, the Federal Communications Commission). **2.** A specific assignment of a frequency or a band of frequencies. Also see RADIO SPECTRUM.

frequency band A given range of frequencies, usually specified for some application (e.g., the band allocated for standard radio broadcast service). Also see BAND.

frequency bias An intentional change in the frequency of a transmitted signal.

frequency bridge **1.** Any alternating-current bridge, such as the *Wien bridge* or *resonance bridge*, that can be nulled at only one frequency for a given set of bridge-arm values. **2.** Any alternating-current bridge that is used to measure unknown frequencies.

frequency calibrator A device, such as a crystal oscillator, that provides a signal of precise frequency with which other signals can be compared. Also see SECONDARY FREQUENCY STANDARD.

frequency changer **1.** A superheterodyne converter (see CONVERTER). **2.** A motor-generator in which the output voltage has the same value as the input voltage, but is of a different frequency. **3.** See FREQUENCY-MULTIPLYING TRANSFORMER. **4.** See FREQUENCY MULTIPLIER.

frequency-change signaling See FREQUENCY-SHIFT KEYING.

frequency channel A relatively narrow segment of a frequency band allocated to a station in a particular service. The bandwidth of the channel depends on the type of modulation used, the type of data to be transmitted, and the speed or fidelity of the data to be transmitted.

frequency comparator A device, such as an oscilloscope or zero-beat indicator, used to check one frequency against another. Also see FREQUENCY COMPARISON.

frequency comparison The observation of a current or voltage of one frequency for similarities in that of another frequency. Comparisons (as in frequency matching) can be made by audio means, by visual means, or both. Common instruments used are oscilloscopes, beat-note detectors, and beat-note meters.

frequency-compensated attenuator An attenuator, such as one in an electronic voltmeter or wideband oscilloscope, that has been modified by the addition of capacitors or inductors to achieve reasonably flat response over a wide range of frequencies.

frequency compensation The modification of a circuit, such as an amplifier or attenuator, by the addition of capacitors or inductors to tailor its response at specified frequencies.

frequency control **1.** An adjustable component (potentiometer, variable capacitor, or variable inductor) with which the frequency or frequency response of a circuit is adjusted. **2.** A device, such as a quartz crystal or tuning fork, that automatically sets the frequency of an oscillator.

frequency conversion The process of changing a signal from one frequency to another, usually without altering the signal bandwidth. In some cases, a signal is turned “upside down” by this process [e.g., an upper-sideband (USB) signal might be changed to a lower-sideband (LSB) signal]. Generally accomplished by means of a MIXER.

frequency converter **1.** An active or passive device for changing the frequency of a signal. **2.** The mixer in a superheterodyne circuit.

frequency correction Manual or automatic resetting of a deviated frequency to its original value.

frequency counter An instrument that counts signal cycles or pulses over a standard time base (a frequency measurement). Often used to accurately measure the frequencies of radio or television signals; in this application, it is a precision FREQUENCY METER.

frequency cutoff See CUTOFF FREQUENCY.

frequency detector See FREMODYNE DETECTOR.

frequency deviation **1.** The degree to which a frequency changes from a prescribed value. Thus, if the frequency of a 1000-Hz oscillator drifts between 990 and 1010 Hz, the deviation is ±10 Hz. **2.** In a frequency-modulated (FM) signal, the amount of instantaneous frequency shift above and below the unmodulated carrier frequency.

frequency-deviation meter In frequency-modulation (FM) communications operations, a meter that gives a direct reading of frequency deviation resulting from a modulating signal. It uses either a tuned circuit or a frequency comparator.

frequency difference **1.** In a superheterodyne circuit, the difference between the signal frequency and the oscillator frequency. **2.** In any beat-frequency operation, the quantity $f_2 - f_1$, where f_2 is the higher frequency and f_1 is the lower frequency. Compare FREQUENCY SUM.

frequency discriminator See DISCRIMINATOR.

frequency distortion A form of distortion in which the amplification of some frequencies is different from that of others.

frequency distribution See DISTRIBUTION, **2**.

frequency diversity The transmission and reception of signals at two or more frequencies for the purpose of reducing the effects of fading. It is generally used in long-distance, high-frequency circuits.

frequency divider A circuit or device whose output frequency is a fraction of the input frequency. Compare FREQUENCY MULTIPLIER.

frequency-dividing network See CROSSOVER NETWORK.

frequency-division multiplex A form of multiple-signal parallel transmission in which a single carrier is modulated by two or more signals simultaneously.

frequency doubler A circuit that multiplies an input frequency by two. If a doubler’s input circuit is tuned to frequency f, then its output circuit is generally tuned to $2f$. Frequency doubling is performed by various nonlinear devices, including transistors, varactors, and biased diodes.

frequency drift An undesired, usually gradual, change in the frequency of a signal from its intended frequency or channel; expressed in hertz or kilohertz.

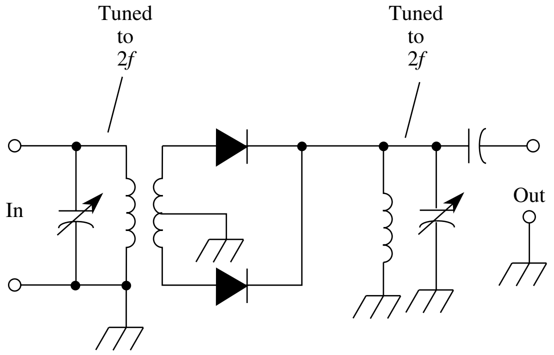

frequency doubler

frequency function See PROBABILITY DENSITY FUNCTION.

frequency indicator **1.** A device that indicates when a phase or frequency is common to two alternating currents. **2.** The display or dial that shows the operating frequency of a radio receiver or transmitter. **3.** See FREQUENCY METER.

frequency keying See FREQUENCY-SHIFT KEYING.

frequency meter An instrument for measuring the frequency of an alternating current. The several different types are used in different applications. Also see AUDIO FREQUENCY METER, FREQUENCY COUNTER, POWER-FREQUENCY METER, and WAVEMETER.

frequency-modulated radar See FM RADAR.

frequency modulation Abbreviation, FM. A method of conveying intelligence in wireless communications and broadcasting. The amplitude of the carrier remains constant, and the instantaneous frequency varies. One scheme for obtaining this type of modulation is to apply the modulating signal to a varactor in an oscillator. Another method is to modulate the phase of the oscillator signal. This causes small fluctuations in the frequency as well, because any instantaneous phase change shows up as an instantaneous frequency change (and vice versa). Also see AMPLITUDE MODULATION, PHASE MODULATION, SINGLE SIDEBAND.

frequency modulation deviation **1.** In frequency modulation (FM), the largest difference between the instantaneous signal frequency and the unmodulated carrier frequency. **2.** The maximum bandwidth of an FM signal at its audio modulation amplitude peak.

frequency modulator **1.** A circuit or device that modulates the frequency of an oscillator. **2.** The modulator section of an FM transmitter.

frequency monitor A device used (often continuously) to check the frequency of a signal (e.g., a frequency-deviation meter used in radio broadcast stations or a frequency meter used in electric-generating stations).

frequency multiplier A circuit or device whose output frequency is a multiple of the input frequency. See, for example, FREQUENCY DOUBLER.

frequency-multiplying amplifier See MULTIPLIER AMPLIFIER.

frequency-multiplying transformer A magnetic amplifier that generates harmonics of the supply frequency. The effect results from the nonlinearity of the transformer core material.

frequency offset **1.** The difference between an actual frequency and the desired frequency. **2.** In a communications transceiver, the difference between the receiver frequency and the transmitter frequency. In some modes, such as single-sideband (SSB), the offset is normally zero. In other modes, notably continuous-wave (CW) Morse code, the offset is normally several hundred Hz. **3.** See FREQUENCY SPLIT, **1**.

frequency overlap **1.** A common band of frequencies between two adjacent channels in a communications system. **2.** A common frequency region between two assigned bands. **3.** A condition in which parts of the sidebands of two signals occupy the same range of frequencies.

frequency pulling A change in the frequency of a circuit, especially of a self-excited oscillator, resulting from the detuning effects of an external circuit, device, or condition (such as body capacitance or a change in the temperature).

frequency pushing An effect in which a current change in a source oscillator causes a shift in source frequency.

frequency quadrupler See QUADRUPLER, **2**.

frequency quintupler See QUINTUPLER, **2**.

frequency range **1.** A communication system's frequency transmission limits, beyond which the power output is attenuated below a specified amount. **2.** The frequency band or bands within which a radio transmitter, receiver, or transceiver is designed to operate.

frequency ratio counter See FREQUENCY RATIO METER.

frequency ratio meter A meter that indicates the ratio between two frequencies, and is useful in the quick identification of harmonics.

frequency record A phonograph test disk containing recordings of various frequencies at specified amplitudes.

frequency rejection The elimination, usually by a filter, of a single frequency (or narrow band of frequencies) from a mixture of frequencies. Compare FREQUENCY TRANSMISSION.

frequency relay A frequency-sensitive relay (see SELECTIVE RELAY, **1**).

frequency response A performance characteristic that describes the operation of a device or circuit over a specified range of signal frequencies (e.g., the gain-versus-frequency characteristic of an amplifier).

frequency-response recorder A graphic recorder that automatically plots a frequency-response curve for a device under test.

frequency run A test, or test sequence, that determines the loss characteristics of a circuit as a function of the operating frequency.

frequency scanning **1.** A controlled fluctuation of the transmitter frequency in a frequency-agile radar or communications system. **2.** In a programmable, digital communications receiver or transceiver, a form of simultaneous digital monitoring of two or more channels. **3.** The frequency-response change in a spectrum analyzer.

frequency scaler See SCALER.

frequency-selection sensor A sensor that passes or rejects phenomena at certain frequencies while ignoring those at other frequencies.

frequency-selective relay See SELECTIVE RELAY, **1**.

frequency-sensitive bridge A bridge, such as the *Wien bridge* or *resonance bridge*, that can be balanced at only one frequency for a given set of bridge-arm values.

frequency separator In a television receiver, the circuit that separates horizontal- and vertical-scanning sync pulses.

frequency-shift keying Abbreviation, FSK. A method of digital signal transmission. The logic 1 (*high* or *mark*) pulses are sent at a specific carrier frequency, and the logic 0 (*low* or *space*) pulses are transmitted at another frequency slightly higher or lower than the logic 1 pulses. This is the most primitive form of frequency modulation (FM). The difference between the mark and space frequencies is called the *shift*, and is usually between 100 and 1000 Hz. Compare AUDIOFREQUENCY-SHIFT KEYING.

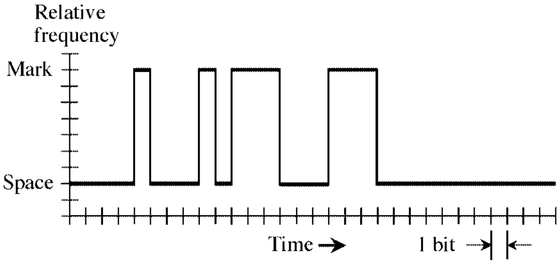

frequency-shift keying

frequency-shift radar See DOPPLER RADAR.

frequency span The difference $f_2 - f_1$, where f_1 is the lowest frequency in a given range of frequencies and f_2 is the highest frequency. Compare FREQUENCY SPREAD.

frequency spectrum All electromagnetic radiation, from longest to shortest wavelengths, within a set of specified limits.

frequency split **1.** The difference between the receiver frequency and the transmitter frequency in a communications repeater. **2.** See FREQUENCY OFFSET, **2**.

frequency spotting The setting-up of signals at reference frequencies (usually harmonics of a standard-frequency oscillator), and their use in identifying unknown frequencies. Also see FREQUENCY CALIBRATOR.

frequency spread The ratio f_2/f_1, where f_1 is the lowest frequency in a given range of frequencies and f_2 is the highest frequency. Compare FREQUENCY SPAN.

frequency stability The degree to which a frequency remains constant during variations in temperature, current, voltage, and similar factors. It is specified in Hertz (Hz), kilohertz (kHz), or megahertz (MHz), or in parts per million per unit of the variable parameter.

frequency standard A signal source of a precise frequency, against which other signal sources can be calibrated. See specifically PRIMARY FREQUENCY STANDARD and SECONDARY FREQUENCY STANDARD.

frequency sum In a beat-frequency system, the quantity $f_1 + f_2$, where f_1 is the lower frequency and f_2 is the higher frequency. Compare FREQUENCY DIFFERENCE, **2**.

frequency swing See FREQUENCY DEVIATION, **1**, **2**.

frequency synthesizer A generator of signals at a precise frequency or set of frequencies, generally adjustable in discrete frequency steps. It is used for test or communications purposes. The signals are derived from a single-frequency source, usually a crystal oscillator. Also see SIGNAL SYNTHESIZER.

frequency tolerance The acceptable amount by which a frequency can vary from its intended value. The tolerance can be specified as a percentage of the stated frequency, a certain number of parts per million, or a certain number of hertz (Hz), kilohertz (kHz), or megahertz (MHz). Example: 3.675000 MHz ±10 Hz.

frequency-to-voltage converter A device or circuit that delivers an output voltage that is proportional to the input frequency.

frequency translation **1.** The conversion of a given frequency band from one part of the electromagnetic spectrum to another, without changing the actual separation of channels or the overall width of the band. **2.** See FREQUENCY CONVERSION.

frequency transmission The passage of a frequency or band of frequencies from a mixture of frequencies through a filter or other circuit. Compare FREQUENCY REJECTION.

frequency tripler See TRIPLER, **2**.

frequency-variation method A method of determining the figure of merit (Q) of a tuned circuit by varying the frequency of the applied test voltage from resonance (f_r) to a high point (f_2) and a low point (f_1). At the high and low points, the circuit voltage is 0.707 times the voltage at resonance. The figure of merit then is calculated from the formula $Q = f_r/(f_2 - f_1)$.

frequency-voltage converter See FREQUENCY-TO-VOLTAGE CONVERTER.

frequency-wavelength conversion See WAVELENGTH-PERIOD-FREQUENCY RELATIONSHIPS.

fresnel (A.J. Fresnel, 1788–1827) A unit of frequency equal to 10^{12} Hz. Also called *terahertz* and abbreviated THz.

fresnel lens A usually square plastic sheet with progressively thicker concentric areas; its effect is similar to that of an automotive headlight lens.

Fresnel number A measure of the relative effects of diffraction in an optical lens. The Fresnel number is equal to the radius of the lens divided by the product of the light wavelength and the lens focal length, all measured in the same units.

Fresnel region For a radio-frequency transmitting antenna, the zone between the antenna and the FRAUNHOFER REGION. The size of the Fresnel region depends on the wavelength of the radiated energy.

friction The resistance to mechanical motion when one material is rubbed against another. Friction was one of the earliest sources of human-made electricity (see FRICTIONAL ELECTRICITY and ELECTRIC MACHINE). Electrical resistance, opposing the flow of current, is analogous to friction.

frictional electricity Static electricity generated by rubbing one material against another.

frictional electric machine See ELECTRIC MACHINE.

frictional error The change in parameters of a phonograph pickup, resulting from friction with the disk surface.

frictional loss A decrease or impairment in the efficiency with which energy is converted into useful work, caused by friction between moving parts.

fringe area The region in which a signal falls to the minimum field strength necessary for satisfactory communication.

fringe howl In a regenerative detector, a howling sound that occurs when the transistor first begins to oscillate, obscuring the signal. The term is used because the circuit is operated at the fringe of oscillation.

fringing See EDGE EFFECT.

Fritch Trade name (American Telephone & Telegraph Co.) for *frequency-selective switch.*

fritting A condition in which electrical contact corrosion creates a small hole, through which molten contact material passes to form a conductive bridge.

front contact The movable contact of a relay.

front end 1. The first radio-frequency (RF) amplifier stage in a radio or television receiver. **2.** The converter portion of a superheterodyne communications receiver (i.e., the RF amplifier, first detector, and local oscillator). Compare REAR END.

front layer photocell See RECTIFIER PHOTOCELL.

front porch In a television horizontal sync pulse, the interval between the end of the sync pulse and the fall of the blanking pedestal. Compare BACK PORCH.

front projection In big-screen video, a scheme in which the images from a set of bright cathode-ray tubes (CRTs) are projected onto a reflective screen, in a manner similar to the way the film is projected in a movie theater.

front-surface mirror Also called *first-surface mirror.* A mirror that has its reflective material on the front, instead of on the back.

front-to-back ratio Abbreviation, f/b. An expression of the ability of a unidirectional antenna to concentrate its radiation or response in its favored direction. This specification is nearly always given in decibels (dB). The field strength in the favored direction is compared with the field strength exactly opposite the favored direction at the same distance from the antenna in free space, at the same frequency, and with the same power applied to the antenna feed point. Measurements can be made with a calibrated field-strength meter. Compare FRONT-TO-SIDE RATIO.

front-to-side ratio Abbreviation, f/s. An expression of the directivity of a unidirectional or bidirectional antenna system. This specification is nearly always given in decibels (dB). The field strength in the favored direction(s) is compared with the field strength at right angles to the favored direction(s) at the same distance from the antenna in free space, at the same frequency, and with the same power applied to the antenna feed point. Measurements can be made with a calibrated field-strength meter. Compare FRONT-TO-BACK RATIO.

frost alarm A device or circuit that responds to the presence of frost and actuates an alarm. Such alarms are sensitive to temperature, moisture, or both.

FRUGAL Acronym for *FORTRAN rules used as a general applications language.*

FRUSA Acronym for *flexible rolled-up solar array*, such as the type used in spacecraft and communications satellites.

F scan In radar operations, a display in which a central blip represents the target at which the antenna is pointed; horizontal and vertical displacement of the blip indicate corresponding horizontal and vertical aiming errors.

FSK Abbreviation of FREQUENCY-SHIFT KEYING.

FSM Abbreviation of FIELD-STRENGTH METER.

FS meter See FIELD-STRENGTH METER.

FSR Abbreviation of *feedback shift register.*

ft Abbreviation of *foot* or *feet.*

FT-cut crystal A piezoelectric plate cut from a quartz crystal at an angle of +57°, with respect to the z-axis. Also see CRYSTAL AXES and CRYSTAL CUTS.

ft-Lb Abbreviation of FOOT-LAMBERT.

ft-lb Abbreviation of FOOT-POUND.

Fuchs antenna A simple antenna consisting of a single-wire radiator without feeder or transmission line, connected directly to the transmitter. It is usually an odd multiple of 0.25 wavelength long. When a good radio-frequency (RF) ground is used, this antenna can be effective at high frequencies, although part of its radiated field is often inside the transmitter building.

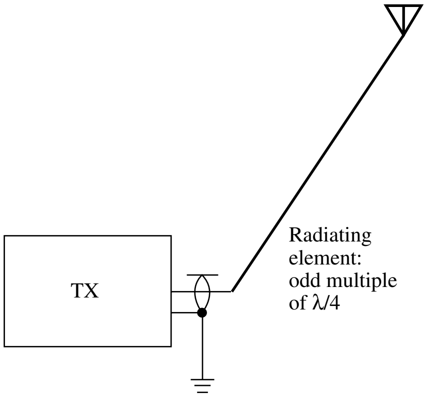

Fuchs antenna

fuel alarm A sensing circuit that actuates an alarm when the fuel in a tank or reservoir falls to a prescribed level.

fuel cell A generator that produces electricity directly from a reaction between fuel substances, such as hydrogen and oxygen.

fuel-flow alarm An electronic circuit that actuates an alarm when fuel flow changes from a prescribed value.

fuel-flow control A servo system that automatically maintains or corrects the flow rate of a fuel.

fuel-flow gauge See FUEL-FLOW METER.

fuel-flow indicator See FUEL-FLOW METER.

fuel-flow meter An instrument for measuring fuel flow rate.

fuel-flow switch A switch that is actuated by fuel flowing in pipes or other channels.

fuel gauge An instrument consisting of a transducer that senses the level of liquid fuel in a tank and delivers a proportional output current or voltage, and an electric meter whose needle is deflected in proportion to the current or voltage and, therefore, to the fuel level. Alternatively, the meter can be a direct-readout digital device, showing the number of gallons remaining in the tank, and/or the extent to which the tank is full (fraction or percentage).

fuel meter See FUEL GAUGE.

fuel-pressure indicator An instrument for measuring fuel pressure in pipes or other channels.

fuel-pressure meter See FUEL-PRESSURE INDICATOR.

full adder In a digital computer, an adder circuit that can handle the carry signal, as well as the binary elements that are to be added. Also see ADDER and CARRY. Compare HALF ADDER.

full bridge A bridge-rectifier circuit in which each of the four arms contains a diode. By comparison, the three-quarter bridge contains a resistor in one arm; the half bridge, resistors in two arms; and the quarter bridge, resistors in three arms.

full-duplex system In data communications, a system that transmits data in both directions simultaneously and continuously. Compare HALF-DUPLEX SYSTEM.

full-focus yoke See COSINE YOKE.

fullhouse A multichannel radio-control model plane system that allows the use of a realistic complement of working control surfaces.

full-load current The output current from a source when the load is maximum (that is, the load resistance is minimum).

full-load power The power drawn from a source when the load is maximum (that is, the load resistance is minimum).

full-load voltage The output voltage of a source when full power is drawn [i.e., when the load is maximum (that is, the load resistance is minimum)].

full-load wattage See FULL-LOAD POWER.

full-power frequency response The highest frequency at which a signal can fluctuate at full voltage (peak-to-peak) without causing distortion of more than a certain specified amount.

full-range speaker See MONORANGE SPEAKER.

full scale **1.** The operating range of an instrument. **2.** In an analog meter, the quantity indicated by maximum deflection of the needle (usually at the extreme right-hand end of the calibrated scale). **3.** Transducer output as a function of highest allowable input stimulus.

full-scale current Symbol, I_{FS}. In a digital-to-analog converter, the maximum current that can occur at the output.

full-scale error For an electrical indicating instrument, the rated full-scale input signal minus the actual input signal that causes a full-scale deflection. Thus, the predictable error in an instrument, expressed as a percentage of the full-scale reading.

full-scale frequency Generally expressed in Hertz (Hz) or kilohertz (kHz). The maximum frequency at which a voltage-to-frequency converter can operate while remaining within its specifications.

full-scale sensitivity The current, voltage, or power required to deflect a meter mechanism to full scale.

full-scale symmetry Expressed in microamperes (mA). The mathematical difference between the full-scale current outputs in a complementary-output digital-to-analog converter.

full track A recording track covering the full width of a magnetic tape.

full-track head A tape-recorder head having a gap that covers the full width of the tape.

full-track recording Usually applicable to quarter-inch or narrower magnetic recording tape, a one-track recording made by a head that magnetizes essentially the entire width of the tape.

full-wave bridge rectifier See BRIDGE RECTIFIER.

full-wave, center-tap rectifier A circuit in which the center-tapped secondary winding of a transformer operates two rectifier diodes, each on an alternate half-cycle of secondary voltage. The frequency of the ripple in the direct-current (dc) output is equal to twice the alternating-current (ac) input frequency. Compare BRIDGE RECTIFIER.

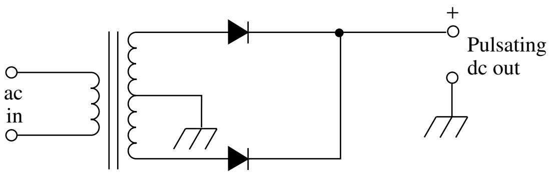

full-wave, center-tap rectifier

full-wave detector A detector circuit using two diodes in a full-wave, center-tap rectifier configuration.

full-wave doubler See FULL-WAVE VOLTAGE DOUBLER.

full wavelength Symbol, λ. The distance that corresponds to 360 degrees of phase as an electromagnetic (EM) field is propagated. In free space, it is related to the frequency by a simple equation:

$$L_{\text{ft}} = 984/f$$

where L_{ft} represents λ in feet, and f represents the frequency in megahertz. If λ is expressed in meters, then the formula is:

$$L_{\text{m}} = 300/f$$

Where L_{m} represents the displacement in meters. In general, if ν is the velocity factor (expressed as a ratio) in a given medium, then:

$$L_{\text{ft}} = 984\nu/f$$

and

$$L_{\text{m}} = 300\nu/f$$

Compare HALF WAVELENGTH, QUARTER WAVELENGTH.

full-wave loop antenna A loop antenna with a circumference of one wavelength, fed at a break. The loop can be circular, square, triangular, or occasionally some odd shape. The most efficient configuration is the circular loop. Maximum radiation occurs perpendicular to the plane of the loop.

full-wave rectifier A rectifier that delivers a half-cycle of pulsating direct-current (dc) output voltage for each half-cycle of applied alternating-current (ac) voltage. The successive output half-cycles have the same polarity. See, specifically, BRIDGE RECTIFIER and FULL-WAVE, CENTER-TAP RECTIFIER. Compare HALF-WAVE RECTIFIER.

full-wave vibrator **1.** In a vibrator-type power supply, an interrupter that closes contacts on both ends of its swing, thus causing direct current (dc) to flow through the transformer in alternate directions. **2.** A vibrator-type rectifier that closes in both directions.

full-wave voltage doubler A voltage-doubler circuit whose direct-current (dc) output has a ripple of twice the alternating-current (ac) supply frequency. Compare HALF-WAVE VOLTAGE DOUBLER.

function **1.** A mapping between two sets of quantities or points A and B, such that: (1) For each y in B, there exists at least one corresponding x in A; and (2) For each x in A, there exists exactly one y in B. In this case y is said to be a function of x; this can be written as $y = f(x)$. The set A is called the *domain* of f; the set B is called the *range* of f. **2.** A mathematical expression, using symbols, relating variables (e.g., the expression $x - y = z$ is a function of variables x, y, and z). **3.** The behavior and application for which a device or system is designed. **4.** Part of a computer instruction specifying the operation to be done.

functional blocks Combinations of substances or components that perform specific tasks in an electronic circuit. An example is a tuned circuit, containing inductive reactance, capacitive reactance, and resistance.

functional character See CONTROL CHARACTER.

functional design Design specifications encompassing a description of how system elements will interrelate, and what their logic design will be.

functional diagram FUNCTIONAL DESIGN represented in graphic form; that is, as an illustration or set of illustrations.

functional electronic block Abbreviation, FEB. A complete integrated circuit. See INTEGRATED CIRCUIT.

functional end (FE) point In a system operating from a battery power supply, the lowest voltage at which the equipment will properly operate. As a battery discharges, the voltage decreases; when the voltage drops to the FE point, the battery must be replaced or recharged.

functional test A performance test of a device or circuit, to see that it behaves as intended in the environment in which it is to be used.

function generator **1.** A signal generator whose output is any of several selectable waveforms (e.g., sine, square, triangular, step-pulse) and frequencies (or repetition rates). **2.** An analog computer circuit that produces a variable based on a mathematical function and one or more input variables.

function key **1.** In digital communications, a keyboard key used to control the form in which a message will be received. **2.** On a computer keyboard, any of 12 keys (usually designated F1 through F12) that activates special functions. The precise action of a given key depends on the program being run.

function polling A polling technique in which a disabled device signals its condition and specifies the remedy.

function switch In a multifunction instrument, such as a voltohm-milliammeter, the switch that permits selection of the various functions.

function table **1.** A table of mathematical function values. **2.** Hardware or software that translates one representation of information into another. **3.** A routine that allows a computer to use the values of independent variables to determine the value of a dependent variable.

fundamental Contraction of FUNDAMENTAL FREQUENCY.

fundamental component The FUNDAMENTAL FREQUENCY of a complex wave.

fundamental frequency **1.** The lowest frequency in a complex wave containing harmonic energy. **2.** In a radio or television transmitter, the intended frequency of operation. **3.** In acoustics and audio applications, the predominant pitch of a musical tone.

fundamental group A set of trunk lines in a telephone system, through which zone centers are interconnected.

fundamental mode See DOMINANT MODE.

fundamental suppression Removal of the fundamental frequency from a complex wave, leaving only the harmonics, as in the operation of a null network adjusted to the fundamental frequency.

fundamental units Base units of an absolute system of units. Example: the meter (m), the kilogram (kg), and the second (s) in the mks system.

fundamental wavelength The wavelength that corresponds to the FUNDAMENTAL FREQUENCY of a wave or signal.

fuse A safety device consisting of a wire of low-melting-point metal. When current passing through the wire exceeds a prescribed (safe) level, the resulting heat melts the wire and opens the circuit, protecting equipment from damage. See PROXIMITY FUSE.

fuse box A set of electrical fuses, usually enclosed in a metal box.

fused junction In a semiconductor, a junction produced by alloying metals to the semiconductor material.

fused junction See ALLOY JUNCTION.

fuse resistor See FUSIBLE RESISTOR.

fuse wire The low-melting-point wire used in fuses. See FUSE.

fusible resistor A low-value resistor that also serves as a fuse in certain appliances, such as television receivers.

fusing current The specified current level at which a wire of a given diameter and material composition will melt.

fusion **1.** In acoustics, pertaining to delayed or reflected waves that arrive within approximately $\frac{1}{25}$ of a second of the direct wave. So called because the human ear/brain “fuses” (blends) sounds together when they are separated by less than about $\frac{1}{25}$ second. If the delay is longer, the ear/brain usually perceives an echo instead. **2.** In a nuclear reaction, the uniting of two atomic nuclei, accompanied by the release of energy.

future labels In a computer system, program instruction labels that refer to locations not designated as absolute addresses by a compiler or assembler.

futurist A person who tries to anticipate or predict, based on current technology and trends, what will be accomplished in a given field in the next several years or decades.

fuzz A form of deliberate distortion in the waveform produced by an electric guitar.

fuzzbox A circuit that distorts the waveform produced by an electric guitar, for the purpose of creating various musical sound effects.

fuzz buster Slang for a specialized mobile radio receiver, used by drivers of vehicles to signal the presence of law-enforcement radar equipment.

fV Abbreviation of *femtovolt.*

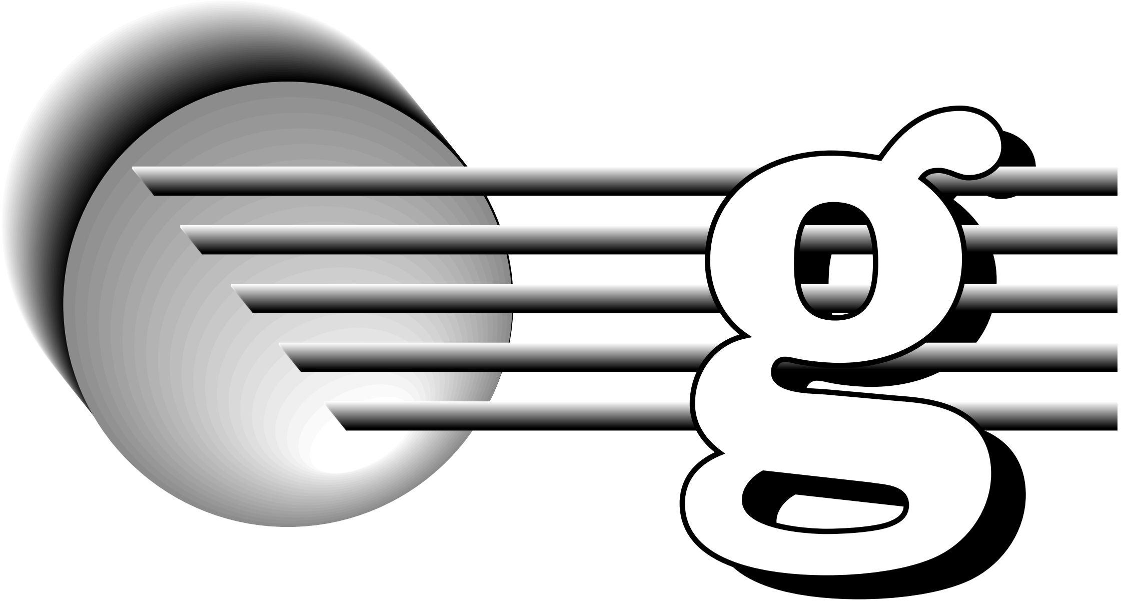

G **1.** Symbol for CONDUCTANCE. **2.** Abbreviation of GIGA-. **3.** Symbol for DEFLECTION FACTOR. **4.** Symbol for PERVEANCE. **5.** Symbol for GRAVITATIONAL CONSTANT. **6.** Symbol for GENERATOR. **7.** Symbol for GATE. **8.** Abbreviation for GAUSS.

g **1.** Symbol for CONDUCTANCE. **2.** Abbreviation of GRAM. **3.** Subscript for GATE. **4.** Subscript for GENERATOR. **5.** Symbol for GRAVITY.

GA Radiotelegraph abbreviation of "Go ahead."

G/A Abbreviation of *ground-to-air.*

Ga Symbol for GALLIUM.

GaAs **1.** Formula for *gallium arsenide.* **2.** Pertaining to semiconductor devices based on gallium arsenide.

GA coil A special form of coil, wound with extra space among the turns to reduce the distributed capacitance.

gadget **1.** A device or component. **2.** A superfluous or makeshift device.

gadolinium Symbol, Gd. A metallic element of the rare-earth group. Atomic number, 64. Atomic weight, 157.25.

gage See GAUGE.

gain The extent to which a component, circuit, device, or system increases current, voltage, or power. Applicable especially to active devices, such as transistors and integrated circuits (ICs), and to amplifiers and filters that use them. Also used to express the directional properties of some antenna systems. Usually specified in decibels (dB). See AMPLIFICATION, CURRENT AMPLIFICATION, DECIBEL, VOLTAGE AMPLIFICATION, and POWER AMPLIFICATION.

gain bandwidth product Symbol, f_T. The frequency at which the gain of a bipolar transistor is equal to 1 (no amplification or loss) in the common-emitter configuration. The f_T represents an absolute upper limit for the frequency at which a bipolar transistor will works as an amplifier. Any attempt to design an amplifier using a bipolar transistor at a frequency higher than its f_T will inevitably fail. Compare ALPHA CUTOFF FREQUENCY.

gain control **1.** To adjust the gain of an amplifier. **2.** A potentiometer used to adjust amplifier gain.

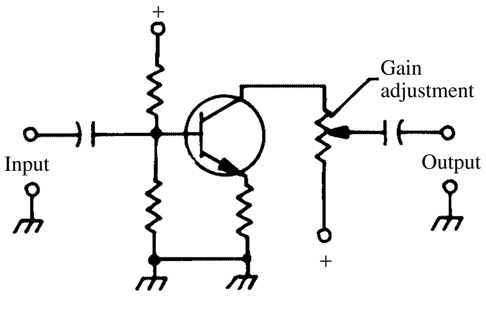

gain control

gain function A function between two currents or voltages in a circuit with gain.

gain-level linearity The quantitative measure of the extent to which the gain of a device depends on the signal level. The level is found by compar-

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

ing the output to the input level over a range of input signal levels.

gain reduction The drop in gain of an amplifier at high- and low-frequency extremes.

gain sensitivity control See DIFFERENTIAL GAIN CONTROL.

gain stability The degree to which the gain of a system remains constant during changes in related factors, such as temperature, supply power, and loading.

gain temperature coefficient The extent to which the full-scale current varies over a certain temperature range, expressed in parts per million per degree Celsius (ppm/°C).

galactic noise Radio noise propagated from the plane of our galaxy, and especially from the center, located in the direction of the constellation Sagittarius. It is of significance in space communications and radio astronomy.

galena Formula, PbS. Natural lead sulfide, which in nature takes the form of bluish-gray, cubical crystals.

gallium Symbol, Ga. Atomic number, 31. Atomic weight, 69.72. One of the constituents of the semiconductor compound GALLIUM ARSENIDE.

gallium arsenide Formula, GaAs. A compound of gallium and arsenic, used as a semiconductor material. It is noted for its low-noise characteristics.

gallium-arsenide diode A diode in which the semiconductor material is processed gallium arsenide.

gallium-arsenide varactor A low-noise, microwave varactor in which the semiconductor material is gallium arsenide.

gallium-phosphide diode A light-emitting diode in which the semiconductor material is processed gallium phosphide.

galloping ghost A form of radio-control system in which the elevation and rudder can be moved to the desired extent.

Galton whistle A device for producing high-frequency acoustic waves (ultrasound), similar to a common dog whistle.

galvanic cell Generic term for any electrochemical primary voltaic cell.

galvanic corrosion Corrosion that occurs on one of two dissimilar metals when they are immersed in an electrolyte. Caused by battery action between them. Compare ELECTROLYTIC CORROSION.

galvanic couple See VOLTAIC COUPLE.

galvanic current A very small direct current such as that produced by dissimilar metals in acid or by nervous reaction in living tissue.

galvanic pile See VOLTAIC PILE.

galvanic series A list of metals and alloys arranged in order of the most to least likely to oxidize in a given environment.

galvanic skin response Abbreviation, GSR. The variations in electrical resistance of the (usually human) skin. This phenomenon is a useful indicator in physiology, psychology, and criminology.

galvanic taste A sharp, metallic taste experienced when a small electric current is passed through the tip of the tongue.

galvanism (After Luigi Galvani, 1737–1798) The production of an electric current by chemical action, as in a battery.

galvanize To coat steel with zinc to forestall corrosion.

galvanometer A sensitive, bi-directional current meter. Used in various electrical tests—especially as a null indicator in bridge operation. Also see MICROAMMETER.

galvanometer constant The number by which a galvanometer reading must be multiplied in order to obtain the current in microamperes, milliamperes, or amperes.

galvanometer recorder A graphic recorder in which a mirror in a movable-coil galvanometer reflects a beam of light to a passing strip of photographic film.

galvanometer shunt A resistor placed in parallel with a galvanometer to decrease its sensitivity. Also see SHUNT RESISTOR.

galvanometry The use of galvanometers to determine the intensity and direction of electric currents.

galvanoplastics The science of ELECTROPHORESIS and ELECTROPLATING.

galvanoscope An instrument for detecting and showing the direction of very weak electric currents.

galvanotherapy The use of electric currents to produce heat in the body of a human or animal.

gamma ferric oxide A form of coating used in formulation of magnetic recording tape.

gamma match A linear transformer for matching an unbalanced (usually coaxial) feed line to a balanced (usually half-wave) antenna. The outer conductor of the cable is connected to the center of the radiator, and an extension of the center conductor runs for a short distance parallel to the radiator, making a right-angle bend before connecting to the radiator.

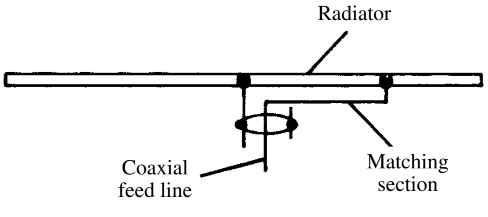

gamma match

gamma rays High-energy, ionizing radiation emitted by radioactive substances; similar to X rays, but of a shorter wavelength.

gamma section See GAMMA MATCH.

gang To mechanically couple components (pots, switches, etc.) for operation by a single knob.

gang capacitor A variable capacitor consisting of sections mounted on the same shaft for simultaneous variation. It is usually specified by the number of sections (e.g., four-gang capacitor). Compare GANGED CAPACITORS.

ganged capacitors Separate variable capacitors mechanically connected together (e.g., by belt or gear drive) for simultaneous variation. Compare GANG CAPACITOR.

ganged potentiometers Separate potentiometers mechanically connected together (e.g., by belt or gear drive) for simultaneous variation. Compare GANG POTENTIOMETER.

ganged rheostats See GANGED POTENTIOMETERS.

ganged switches Separate switches mechanically connected together for simultaneous operation. Compare MULTISWITCH.

ganged tuning Simultaneous tuning of separate circuits by means of ganged capacitors or ganged potentiometers.

gang potentiometer A potentiometer consisting of sections mounted on the same shaft for simultaneous variation. Usually specified according to the number of sections (e.g., *dual potentiometer*).

gang printer In digital computer and data processing operations, an electromechanical printer capable of printing an entire line at one time.

gang punch **1.** To punch identical or nonvarying information into the cards of a group. **2.** A machine for this operation.

gang rheostat See GANG POTENTIOMETER.

gang switch See MULTISWITCH.

Gantt chart A chart of activity versus time used in industry as an aid in making decisions regarding the allocation of resources for specific activities [e.g., as applied to PERT (project evaluation and review techniques)].

gap **1.** A space between electrodes or magnetic poles. **2.** A device consisting essentially of separated electrodes (e.g., *spark gap*). **3.** A relatively narrow space cut in iron cores to provide a break in a magnetic circuit. Also see SLOT, **1**. **4.** The opening between the opposite poles of a tape recorder or playback head.

gap arrester A lightning arrester consisting of a number of metal cylinders separated by air gaps.

gap coding A system in which silent periods are inserted, according to a specific timing code, into a transmission.

gap depth In a magnetic recording head, the depth of the gap (taken perpendicular to the face). Compare GAP WIDTH.

gap digit A digit that contributes no intelligence to the word in which it appears (e.g., a *parity bit*).

gap energy The energy represented by the forbidden gap in electron energy levels (e.g., between the M-valence band and the N-conduction band in a material).

gap filling Modification of an antenna for the purpose of eliminating nulls in the directional pattern.

gap insulation See SLOT INSULATION, **1**, **2**.

gap loss In a reproducing head, the loss that occurs because of the GAP DEPTH.

gap-type protector A spark gap used to protect equipment from high-voltage transients.

gap voltmeter See NEEDLE GAP and SPHERE GAP.

gap width In a magnetic recording head, the width of the gap (taken parallel with the face). Compare GAP DEPTH.

garbage **1.** In digital computer operations, a colloquialism for useless or incorrect data. **2.** Colloquialism for unreadable signals or severe intermodulation in a radio communications circuit. **3.** Colloquialism for an unsound theory.

garble **1.** Garbled matter. **2.** Also called *scramble.* To purposely render communications or data unintelligible to everyone, except the intended recipient(s). See SCRAMBLER CIRCUIT.

garbled matter Confused communications or data, usually resulting from distortion in a circuit or system. Also called GARBLE.

garbler See SCRAMBLER CIRCUIT.

garnet maser A maser that uses natural or synthetic garnet as the stimulated material. Also see YTTRIUM-IRON-GARNET.

gas One of the states of matter, characterized by molecules that are widely separated and are in continual, relatively rapid motion. Because it is a fluid, a gas will readily conform to a container of any shape. Gases can readily be compressed and liquefied. Compare LIQUID, PLASMA, and SOLID.

gas amplification In a radiation-counting device, the ratio, in decibels, of the charge collected to the charge produced in the gas.

gas breakdown The ionization of a gas by means of high voltage. The intensity of the electric field prevents recombination of ions. Collisions among atoms cause further ionization. Thus, the gas becomes a good conductor of current.

gas cell A cell whose operation depends on gas absorption by the electrodes.

gas cleanup Loss of pressure in a gas-filled tube, eventually leading to failure. Caused by gas ions forming compounds with metal parts or with the glass envelope.

gas detector A device for sensing presence of various gases in the air—especially toxic or explosive gases (such as chlorine, hydrogen, or methane).

gaseous conduction The conduction of an electric current through an ionized gas.

gaseous phototube A PHOTOTUBE containing a small amount of a gas suitable for ionic conduction.

gaseous voltage regulator A gas-filled diode across which the voltage drop is substantially constant during the gas discharge and which accordingly delivers a constant output voltage.

gas-filled cable A sealed cable filled with an inert gas that serves as a low-loss dielectric and moisture barrier.

gas-filled counter tube A radiation counter tube containing a gas that ionizes when irradiated. See GEIGER-MUELLER TUBE.

gas-filled lamp **1.** An incandescent lamp filled with a gas, such as nitrogen, for improved performance. **2.** Discharge lamp.

gas-filled tube An enclosure filled with gas that is subjected to an electric potential, intended to produce specific effects, such as light emission or voltage regulation.

gas-flow alarm An electronic circuit that actuates an alarm when the flow of gas through a pipe changes from a predetermined rate.

gas-flow control A servo system for automatically maintaining or adjusting the flow of gas through pipes.

gas-flow gauge See GAS-FLOW METER.

gas-flow indicator See GAS-FLOW METER.

gas-flow meter An instrument that indicates the rate of gas flow through a pipe, in terms of volume per unit time (e.g., cubic feet per minute).

gas-flow switch In a gas-circulating system, a switch that actuates an alarm when the gas flow rate changes.

gas focusing In a cathode-ray tube, a technique by which a gas is used for the purpose of focusing an electron beam. The ionization of the gas causes the electron beam to be made more narrow.

GASH Acronym for *guanidine aluminum sulfate hexahydrate*, an organic crystalline material used as the dielectric in certain ferroelectric capacitors and ferroelectric memory elements.

gas laser A LASER that uses a gas or mixture of gases (instead of a solid rod) as the stimulated medium. Some of the gases used are argon, carbon dioxide, helium, krypton, and neon.

gas maser A MASER in which the stimulated material is a gas, such as ammonia.

gas multiplication See GAS AMPLIFICATION.

gas noise Electrical noise resulting from the undirected motion of gas molecules in a gas-filled tube.

gas-pressure alarm An electronic circuit that actuates an alarm when gas pressure rises or falls.

gas-pressure control A servo system for automatically maintaining or adjusting gas pressure in pipes or other channels.

gas-pressure gauge See GAS-PRESSURE METER.

gas-pressure indicator See GAS-PRESSURE METER.

gas-pressure meter An instrument that indicates gas pressure in a pipe or container, but provides no means for automatically correcting the pressure.

gas ratio For a gas-filled tube, the ratio I_i/E_i, where I_i is the ion current and E_i is the ionization potential.

gas sensor Any element, such as the filament in a hot-filament gas detector, that responds to the presence of a gas in the environment and activates the detector or alarm circuit.

gassing **1.** The generation of gas by a storage battery, especially while it is being charged. **2.** The generation of gas during electrolysis.

gas sniffer See GAS DETECTOR.

gaston A device intended for the purpose of modulating an aircraft signal, making the signal difficult to jam. The signal is randomly modulated by noise from the device.

gas tube See GAS-FILLED TUBE.

gas-tube lightning arrester A lightning arrester, consisting of a special gas diode. The tube has virtually infinite resistance at low voltages, but provides a low-resistance path to ground when the high voltage of a lightning stroke ionizes the gas.

gas-tube oscillator A relaxation oscillator using a two-element gas tube, such as a neon lamp, as the breakdown device.

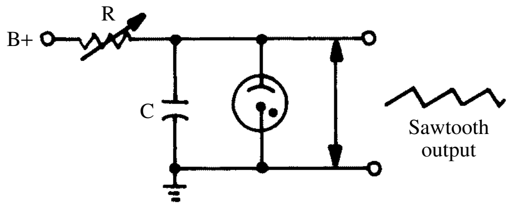

gas-tube oscillator

gas valve See ELECTROMECHANICAL VALVE.

gas X-ray tube An X-ray tube in which the positive ions of a gas bombard the cathode, which emits electrons.

gate **1.** A device or circuit that has no output until it is triggered into operation by one or more enabling signals, or until an input signal exceeds a predetermined threshold amplitude. **2.** The input (control) electrode of a field-effect transistor or thyristor device (e.g., silicon-controlled rectifier). **3.** A signal that triggers the passage of other signals through a circuit.

gate array Basic gates arranged in a pattern on a chip; the gates can be interconnected during manufacture to form a unit that performs whatever function is needed.

gate circuit **1.** An electronic switching circuit (see GATE, **1**). **2.** The circuit associated with the gate electrode of a field-effect transistor.

gate-controlled switch A device similar to a silicon-controlled rectifier or thyristor. A negative current, applied to the gate, switches the device off.

gate current Symbol, I_G. Current flowing in the gate (control) circuit of a semiconductor device. The current is finite in thyristors, but is almost zero in some types of field-effect transistors.

gated amplifier An amplifier whose input is effectively switched on and off by gating signals.

gated buffer A low-voltage, high-current driver, used for differentiation in a multivibrator circuit.

gated flip-flop A flip-flop in which it is impossible for both outputs to be low simultaneously.

gate-dip meter A dip meter using a field-effect-transistor oscillator with the indicating microammeter in the gate circuit.

gate-dip oscillator See GATE-DIP METER.

gated multivibrator A rectangular-wave generator that produces a gate voltage when triggered.

gate-drain voltage Symbol, V_{GD}. In a field-effect transistor, the maximum voltage permitted between the gate and drain electrodes.

gated sweep **1.** In radar, a sweep whose initiation and duration are closely controlled to eliminate echoes in the image. **2.** A circuit providing the action described in **1**.

gate electrode See GATE, **2**.

gate impedance **1.** The impedance of the gate electrode of a field-effect transistor with respect to the other electrode, which serves as the return. **2.** The impedance of the gate winding of a magnetic amplifier.

gate leakage current See GATE REVERSE CURRENT.

gate nontrigger voltage Symbol, V_{Gnt}. For a thyristor, the direct-current (dc) voltage applied between the gate and the cathode, above which the device fails to maintain its rated blocking voltage.

gate power dissipation Symbol, P_G. In a silicon controlled rectifier, the power consumed by the gate-cathode path.

gate-protected MOSFET A metal-oxide-semiconductor field-effect transistor in which the gate electrode is protected from accidental burnout via built-in Zener diodes, connected back-to-back.

gate pulse **1.** A pulse applied to the gate electrode to actuate a gate-controlled semiconductor device. **2.** An actuating pulse in a gate circuit.

gate recovery time Symbol, t_{Gr}. For a silicon-controlled rectifier, an extension of the reverse recovery time: the interval following application of the reverse voltage required before the forward blocking voltage can be reapplied and then blocked by the device.

gate reverse current Symbol, I_{GSS}. In a field-effect transistor, reverse current in the gate-source circuit. Also called *gate leakage current.*

gate signal **1.** The input or control signal applied to the gate electrode of a semiconductor device. **2.** An actuating signal in a gate circuit.

gate-source breakdown voltage Symbol, BV_{GSS}. The voltage at which the gate junction of a junction field-effect transistor (JFET) enters avalanche.

gate-source pinchoff voltage Symbol, V_P. In a field-effect transistor, the gate-source voltage at which the conduction channel just closes.

gate-source voltage Symbol, V_{GS}. In a field-effect transistor, the direct-current (dc) voltage between the gate and source electrodes.

gate terminal **1.** The terminal connected to the gate semiconductor in a field-effect transistor. **2.** The terminal, or terminals, connected to the input or inputs of a digital-logic network.

gate trigger current In a gate-controlled semiconductor switch, the current flowing in the gate circuit when the device is being switched on by a gate trigger voltage.

gate trigger voltage In a gate-controlled semiconductor switch, the trigger voltage required to actuate the device.

gate turn-off current In a gate-controlled semiconductor switch, the low value of gate current that flows when the device is being switched off. Turn-off current varies with collector (anode) current.

gate turn-off voltage In a gate-controlled semiconductor switch, the low value of gate voltage that causes the device to switch off.

gate voltage **1.** The voltage applied to the gate electrode of a field-effect transistor. See GATE-SOURCE VOLTAGE. **2.** The instantaneous gate-cathode voltage in a silicon-controlled rectifier. **3.** The voltage across the gate winding of a magnetic amplifier.

gate winding In a magnetic amplifier, a winding that produces gating action.

gating **1.** The process of using one signal to switch another (or part of another) on or off for a desired interval. **2.** Selecting a part of a wave for observation or for control purposes.

gauge **1.** Any device, such as a METER, used for the purpose of measuring the magnitude of a quantity. **2.** Wire data and measurements (see WIRE GAUGE, **1**, **2**, **3**). **3.** Sheet metal thickness (e.g., 10 gauge).

Gauss (Karl F. Gauss, 1777–1855) Unit of magnetic flux density, equivalent to one line of flux per square centimeter. The SI (preferred) unit of magnetic flux density is the TESLA (webers per square meter); 1 gauss equals 10^{-4} teslas (symbol, T). Also see FLUX DENSITY.

Gaussian curve See BELL-SHAPED CURVE.

Gaussian distribution In statistics, the symmetrical distribution described by a bell-shaped curve. Also called NORMAL DISTRIBUTION.

Gaussian function A mathematical function used in the design of lowpass filters. The Gaussian response is characterized by low band-edge selectivity, high transition bandwidth, negligible step response overshoot, and low step-response delay time.

Gaussian noise Electrical noise whose amplitude-versus-frequency characteristic is described by the GAUSSIAN DISTRIBUTION.

Gaussian waveform A waveform that results in minimal side lobes in a pulse-compression system.

gaussmeter See FLUXMETER.

Gauss' theorem Across any closed surface within an electric field, the total flux is approximately equal to 12.566 times the enclosed quantity of electric charge.

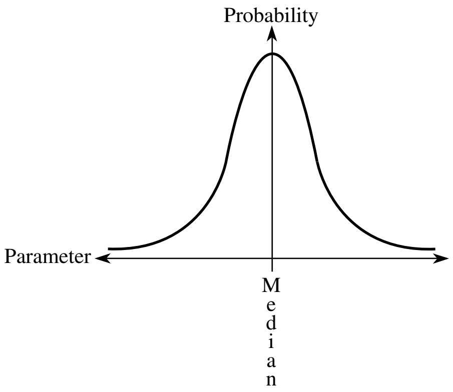

Gaussian distribution

gauze resistor See WOVEN RESISTOR.

GAVRS Abbreviation of *gyrocompass attitude vertical reference system.*

GCA Abbreviation of GROUND-CONTROLLED APPROACH.

GCI Abbreviation of GROUND-CONTROLLED INTERCEPTION.

GCM Abbreviation of *gyrocompass module.* See GYROCOMPASS.

GCT Abbreviation of GREENWICH CIVIL TIME.

Gd Symbol for GADOLINIUM.

G display See G SCAN.

GDO **1.** Abbreviation of GRID-DIP OSCILLATOR. **2.** Abbreviation of GATE-DIP OSCILLATOR.

Ge Symbol for GERMANIUM.

gear **1.** Collectively, electronic equipment. **2.** A toothed wheel commonly used in mechanical devices.

gearmotor An electric motor with a gear train for speed changing.

gear-wheel pattern A pattern produced on an oscilloscope by intensity-modulating a circular trace. A signal of unknown frequency f_x is applied to the intensity-modulation (z-axis) input. The signal produces corrugations in the trace. If there are n corrugations around the trace, and if the trace itself completes f_s revolutions per second, then the unknown frequency f_x, in Hz, is equal to nf_s. Compare SPOT-WHEEL PATTERN.

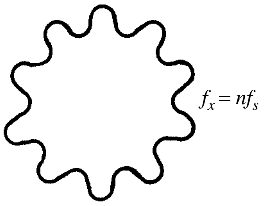

gear-wheel pattern

Geiger counter A radioactivity rate-counting instrument based on the GEIGER-MUELLER TUBE. Pulses from the tube drive a transistor, which, in turn, drives a meter or digital counter to indicate the count.

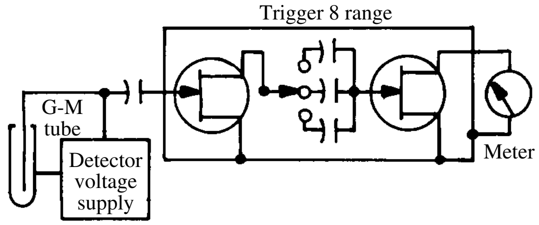

Geiger counter

Geiger-Mueller counter See GEIGER COUNTER.

Geiger-Mueller region For a GEIGER-MUELLER TUBE, the range of voltages within which the output pulse amplitude is constant, regardless of ionizing radiation intensity.

Geiger-Mueller threshold The lowest voltage in the Geiger-Mueller region.

Geiger-Mueller tube A gas-filled radiation detector/counter tube consisting of a straight wire as an anode, surrounded by a cylindrical cathode. The tube is biased by high voltage. Radiation penetrating the tube ionizes the gas; each ionizing event causes an output pulse.

Geiger point counter See POINT COUNTER.

Geiger region See GEIGER-MUELLER REGION.

Geiger threshold See GEIGER-MUELLER THRESHOLD.

Geissler tube A simple gas-filled glow-discharge tube with metal electrodes sealed in each end. When a sufficiently high voltage is applied between the electrodes, the highly rarefied gas ionizes and glows with the color associated with the particular gas used.

gel A substance equivalent to colloidal solution in the solid phase (e.g., silica gel).

gel battery A rechargeable electrochemical BATTERY designed for use with portable electronic and computer equipment. Noted for its ability to deliver high current for short periods, when necessary, and also to deliver moderate current continuously throughout its discharge cycle.

gen Abbreviation of GENERATOR.

genemotor Contraction of *generator/motor*, a (usually battery-driven) dynamotor that has separate motor and generator windings on the same armature core.

general class license An amateur-radio license that conveys some privileges in the high-frequency bands, and all operating privileges in the very-high-frequency region and above. An examination of moderate difficulty is required.

General Packet Radio Service A form of packet radio especially designed for mobile use. Subscribers pay for the volume of data they send and receive, rather than for the time they spend accessing the network. This optimizes the efficiency of the network, both from an engineering standpoint and a financial standpoint. See PACKET COMMUNICATIONS.

general-purpose bridge See UNIVERSAL BRIDGE.

general-purpose component A component designed or used for a wide range of applications. For example, a general-purpose germanium diode is useful as a detector, mixer, limiter, clipper, meter rectifier, automatic-gain-control (AGC) rectifier, and curve changer.

general-purpose computer A computer that can be used in a number of applications for which it was not specifically designed.

general-purpose diode A small-signal semiconductor diode that is useful for a variety of applications, such as detection, light-duty rectification, limiting, logic switching, etc.

general-purpose function generator A nonspecialized function generator that is capable of generating a variety of different waveforms.

general-purpose program A program for the solution of a class of problems or for a specific problem, according to certain parametric values. Also called *general routine.*

general-purpose relay Any relay that can be used in various situations, such as for switching alternating or direct currents.

general-purpose tester An instrument, such as a voltohm-milliammeter, that offers several test capabilities.

general-purpose transistor A transistor that can be used in several applications, such as audio amplification, detection, and oscillation.

general service code See CONTINENTAL CODE.

generate **1.** To produce a signal or carrier wave. **2.** To convert some non-electrical form of energy (usually mechanical) into electrical energy. **3.** To develop subroutines from parameters applied to skeletal coding. **4.** To use a program generator to produce a specialized version of a general-purpose program.

generated address An address developed by program instructions for later use by that program.

generated noise **1.** Electrical noise caused by battery action (i.e., between dissimilar metals) in a component, such as in a potentiometer. **2.** Electrical noise caused by small output variations of generating devices (rotating machines, vibrators, etc.). Also called *generator noise.*

generating magnetometer See EARTH INDUCTOR.

generating station An electric power station.

generating voltmeter An instrument based on a rapidly spinning variable capacitor. A direct-current (dc) voltage applied to the capacitor is converted into an alternating current (ac) by the varying capacitance; the ac is proportional to the voltage.

generation **1.** The production of a signal or carrier wave. **2.** The conversion of some non-electrical form of energy (usually mechanical) into electrical energy. **3.** The number of recording steps between a master recording and a copy. **4.** A copy of data in any form (e.g., tape recording, disk file, and photocopy).

generation number A number that identifies the age of a file; it is included in the file label on the disk or tape containing the file.

generator **1.** Symbol, G. Any signal source. **2.** A rotating machine for producing electricity. **3.** An electronic device for converting direct current into alternating current of a specific frequency and waveshape. **4.** In computer operation, a routine (akin to a compiler) that will produce a program to perform a specific version of some general operation by implementing skeletal coding, according to specific parameters (e.g., *sort generator*).

generator efficiency The ratio of consumed power to delivered power in a generator. It is usually expressed as a percentage.

generator noise Electrical noise caused by a rotating generator. Also see GENERATED NOISE, **2**.

generator-type microphone A microphone that produces an output voltage without the need for a supply voltage. Examples: ceramic, crystal, dynamic, electret, and velocity types.

generator-type transducer A transducer that converts mechanical motion into an electrical signal of a proportional voltage. In such a transducer, an armature or conductor moves in a magnetic field.

generic A form of software collection. Several specialized software packages can be derived from the generic collection, for use in different systems.

geodesic **1.** On a surface, the shortest path between two points. **2.** The shortest path between two geographical locations, measured over the surface of the earth. Also called *geodetic* and *great circle.*

geodesy The branch of applied mathematics concerned with the precise dimensions of the earth.

geodetic system The application of a computer to seismographic studies for the purpose of reducing drilling and mining costs.

geomagnetic field See EARTH'S MAGNETIC FIELD.

geomagnetism The earth's magnetism. Also see EARTH'S MAGNETIC FIELD.

geometric capacitance The ratio of the free charge of a capacitor to the voltage across its terminals.

geometric mean The *n*th root of the product of *n* quantities.

geometric progression A mathematical sequence in which each term after the first is obtained by multiplying the preceding one by a constant quantity (e.g., 1, 2, 4, 8, 16, 32, 64, . . .). Also called *geometric sequence.*

geometric symmetry In a bandpass or band-rejection filter, a condition in which the response is identical on either side of the center frequency. Also called *mirror-image symmetry.*

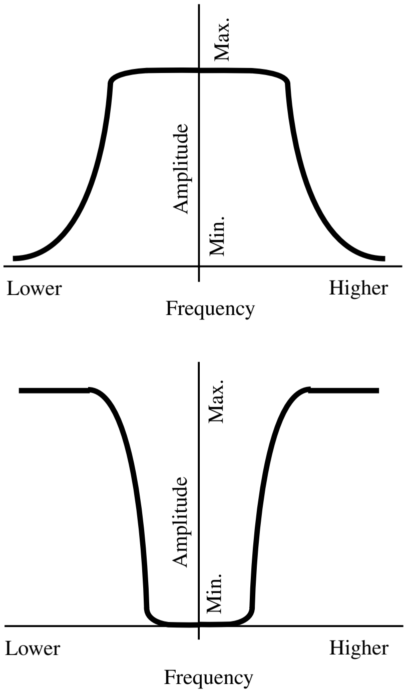

geometric symmetry

george box In an intermediate-frequency amplifier, a device used to reject jamming signals. Any jamming signal with an amplitude lower than a certain minimum is rejected.

geostationary orbit An orbit in which a satellite revolves around the earth exactly once a day, so it remains over the same place on the earth all the time. The altitude must be 22,300 miles, and the orbit must lie in the plane of the earth's equator.

geostationary satellite A satellite in GEOSTATIONARY ORBIT. It is always in the same spot in the sky from any given observing point. A geostationary satellite provides coverage over about 40 percent of the earth, and is commonly used for communications and weather-observation purposes.

germanium Symbol, Ge. A metalloidal element. Atomic number, 32. Atomic weight, 72.59. Used in semiconductor diodes, photocells, rectifiers, and transistors.

germanium diode A diode in which the semiconductor material is specially processed germanium.

germanium dioxide Formula, GeO_2. A gray or white powder obtainable from various sources; it is reduced in an atmosphere of hydrogen or helium to yield germanium, a semiconductor material.

germanium junction diode A germanium diode that contains a pn junction.

germanium photocell A photoconductive cell consisting of a reverse-biased germanium point-contact diode or germanium junction diode.

germanium point contact The contact between a pointed metal wire and a germanium wafer, as in a point-contact diode or point-contact transistor.

germanium rectifier A power rectifier that contains a germanium pn junction.

germanium transistor A transistor in which germanium is the semiconductor material. Such a transistor has lower internal resistance and greater temperature drift than a silicon transistor.

German silver A copper-nickel-zinc alloy used in some resistance wires. Also called NICKEL SILVER.

GEV Abbreviation of *ground effect vehicle.*

GeV Abbreviation of GIGAELECTRONVOLT.

gfi Abbreviation of GROUND-FAULT INTERRUPTER.

g-force See GRAVITY, **2**.

g_{FS} Symbol for FORWARD TRANSCONDUCTANCE.

G/G Abbreviation of GROUND-TO-GROUND.

ghost In television reception, a slightly displaced image appearing on the screen simultaneously with its twin (the false member of a double image).

ghost signal Any signal (such as an undesired reflection) that produces a ghost.

GHz Abbreviation of GIGAHERTZ.

G_i Symbol for INPUT CONDUCTANCE.

Gibson girl A portable radio transmitter, powered by an integral crank-operated generator, introduced during World War II for pilots forced down at sea.

giga- Abbreviation, G. **1.** A prefix meaning billion (10^9). **2.** In digital data applications, a prefix meaning 2^{30}.

gigabit A unit of digital data, equal to 2^{30} bits or 1024 megabits. Also see BIT and MEGABIT.

gigabyte A unit of digital data, equal to 2^{30} bytes or 1024 megabytes. Also see BYTE and MEGABYTE.

gigacycle See GIGAHERTZ.

gigaelectronvolt Abbreviation, GeV. A large unit of voltage; 1 GeV equals 10^9 eV. Also see BEV, ELECTRONVOLT, MEV, and MILLION ELECTRON VOLTS.

gigahertz Abbreviation, GHz. A unit of ultra-high frequency; 1 GHz equals 1,000,000,000 Hz = 10^9 Hz.

gigaohm A unit of extremely high resistance, reactance, or impedance, equal to 1,000,000,000 ohms = 10^9 ohms.

GIGO Abbreviation of *garbage in = garbage out,* an expression signifying that incorrect or improper input to a computer will produce meaningless output.

gilbert (William Gilbert, 1540–1603) A unit of magnetomotive force, equal to 1.26 times the number of ampere-turns. The SI (preferred) unit of magnetomotive force is the ampere (symbol, A); 1 gilbert = 0.796 A.

gilbert per centimeter See OERSTED.

gimbal A suspension device whose orientation can be changed without affecting the attitude of the body being suspended.

gimmick **1.** Colloquialism for any unnamed device. Also see GADGET. **2.** Colloquialism for any tricky manipulation or design. **3.** A low-value capacitor made by twisting two short pieces of insulated wire together.

gimp Colloquialism for the tinsel and cloth conductor used in some earphone cords.

Giorgi system The *meter-kilogram-second (mks)* system of units.

GJD Abbreviation of GERMANIUM JUNCTION DIODE.

glass A hard, brittle, amorphous, and usually transparent substance that is largely silicon dioxide. Glass has a multitude of uses in electronics, and there are several kinds, each having different electrical properties. The dielectric constant ranges from about 4 to 10; the dielectric strength ranges from about 20 to 300 kilovolts per millimeter.

glass arm A stiffness of the wrist or forearm, somewhat resembling writer's cramp, sometimes experienced by radiotelegraph operators or wire telegraph operators after prolonged use of a hand key.

glass bulb The glass enclosure of electron tubes and incandescent lamps.

glass capacitor A capacitor that uses thin glass as the dielectric, and usually has plates consisting of metal electroplated or electrodeposited on opposite faces of the glass. Also see MOLDED GLASS CAPACITOR.

glass diode A semiconductor diode molded in glass.

glass electrode A probe used with a pH meter; it consists of a thin-walled glass tube containing potassium chloride and mercurous chloride. Also see CALOMEL ELECTRODE.

glass envelope See GLASS BULB.

glassivation A procedure for encapsulating semiconductor devices in glass or other dielectric material.

glass-metal seal See GLASS-TO-METAL SEAL.

glass plate capacitor See GLASS CAPACITOR.

glass shell See GLASS BULB.

glass-to-metal seal A bond between glass and metal in electronic devices, such as vacuum tubes, feedthrough terminals, and glass capacitors.

glass tube A vacuum tube whose elements are housed in an evacuated glass envelope.

glide path The guidance beam used by aircraft making instrument landings.

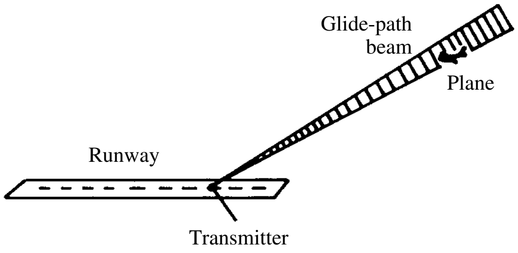

glide-path transmitter

glide-path transmitter A radio-frequency transmitter that produces a guidance beam for aircraft landing purposes. The aircraft follows the beam toward the runway.

glide slope See GLIDE PATH.

G line A microwave conductor consisting of a round wire coated with a dielectric.

glitch **1.** In a television image, a narrow, horizontal interference bar that moves vertically. **2.** A very short and unwanted high-amplitude transient that recurs irregularly in an electronic system.

glitter **1.** In radar, an echo or set of echoes that fluctuates rapidly in intensity because of motion in the target. **2.** A system in which moving devices are used to confuse enemy radar systems.

gloss factor For a reflecting surface, the ratio of reflected light in a selected direction to reflected light in all directions.

glossmeter An instrument for determining GLOSS FACTOR.

glow discharge The luminous electrical discharge resulting from the passage of current through ionized gas in a partially evacuated tube. The color of the glow is characteristic of the particular gas used.

glow-discharge microphone A device that produces audio-frequency currents from the action of sound waves in a glow-discharge tube.

glow-discharge tube A partially evacuated tube that contains two or more electrodes. The rarefied gas in the tube glows when a sufficient voltage is applied to the electrodes. See DISCHARGE LAMP, FLUORESCENT TUBE, and NEON BULB.

glow lamp See DISCHARGE LAMP.

glow modulator tube A gas tube whose luminous output can be modulated by an audio input signal.

glow potential The voltage at which glow discharge just begins in a gas-filled tube.

glow switch In fluorescent light circuits, an electron tube containing two bimetal strips that make mutual contact when heated by the glow discharge.

glow tube See DISCHARGE LAMP, FLUORESCENT TUBE, GLOW MODULATOR TUBE, NEON BULB, and STROBOTRON.

glow voltage See BREAKDOWN VOLTAGE, **2**.

glucinium See BERYLLIUM.

gluon A subatomic particle that is believed to bind quarks together (coined by Prof. Murray Gell-Mann of California Institute of Technology).

GM Abbreviation of *Geiger-Mueller* (see, for example, GEIGER-MUELLER TUBE).

gm Abbreviation of *gram.* Also abbreviated g.

g_m Symbol for TRANSCONDUCTANCE.

g-m Abbreviation of GRAM-METER.

gm-cal Abbreviation of GRAM-CALORIE.

gm-cm Abbreviation of GRAM-CENTIMETER.

G-M counter See GEIGER COUNTER.

GMT Abbreviation of GREENWICH MEAN TIME.

G/M tube See GEIGER-MUELLER TUBE.

gnd Abbreviation of GROUND.

g_o Symbol for OUTPUT CONDUCTANCE.

gold Symbol, Au. A precious metallic element. Atomic number, 79. Atomic weight, 196.967. Electrical contacts that must have low radio-frequency resistance, and that must be relatively immune to corrosion, are often plated with gold.

gold-bonded diode A germanium point-contact diode having a fine gold wire whose point is bonded to the germanium wafer. Its principal features are high forward current and almost constant, low reverse current.

gold doping The diffusion of gold into the base and collector regions of a diffused-mesa transistor; it shortens carrier storage time.

golden ratio A set of proportions used in the design of some speaker cabinets. The width (W) is ⅝ of the height (H); the depth (D) is ⅝ of the width. These correspond to a ratio $H:W:D = 1.000:0.625:0.391$. These proportions are thought by some acoustics engineers to result in the best possible sound quality.

gold-leaf electroscope See ELECTROSCOPE.

Goldschmidt alternator An early dynamo for generating radio-frequency power. The high-frequency energy was not generated directly by the machine, but by resonant circuits and frequency-multiplying interaction between components.

Golf Standard phonetic alphabet code word for the letter G.

goniometer **1.** Generically, any radio direction finder. **2.** An inductive coupler having a secondary coil rotated by a dial calibrated to read azimuth. The coupler, when used with a suitable antenna system, comprises a direction finder. **3.** A device for electrically varying the directional pattern of an antenna.

go-no test A test that indicates only acceptance or rejection of a device. No diagnosis is made.

GOTO In computers and programmable calculators, an instruction that, followed by a suitable label, directs the program to that label.

goto circuit In a digital-logic circuit, a device that senses the direction of electric current.

goto pair A pair of diodes connected in reverse series used in digital-logic circuits.

governor **1.** A device that prevents a motor or engine from running faster than a certain speed. **2.** Any device that limits a circuit parameter.

g parameters Conductance parameters obtained for the equivalent-pi model of a transistor: g_{BE}, g_{GC}, g_{CE}, and g_m.

gpc Abbreviation of GERMANIUM POINT CONTACT.

GPI Abbreviation of GROUND-POSITION INDICATOR.

GPRS Abbreviation for GENERAL PACKET RADIO SERVICE.

gr Abbreviation of *grain(s).*

graceful degradation A computer programming technique used to prevent debilitating breakdown by operating the system—even though several subsystems have malfunctioned; also known as *crippled mode.*

grad A unit of angular measurement equal to 0.9 degree.

graded-base transistor See DIFFUSED-BASE TRANSISTOR.

graded filter A power-supply filter that supplies direct-current output at various points in the filter sequence. Thus, the points in the powered equipment that can tolerate the least ripple are connected to the filter output, and those that can tolerate appreciable ripple are connected to the filter input; fairly critical points are connected to an intermediate position in the filter, such as at the junction of two chokes.

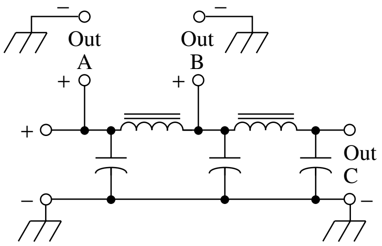

graded filter

graded-junction transistor A grown-junction transistor in which the temperature of the melt and the rate at which the crystal is pulled from it are closely controlled as the n and p layers are formed.

gradient The rate at which a variable quantity increases or decreases. See, for example, VOLTAGE GRADIENT.

gradient microphone A microphone whose output varies with sound pressure. Also see PRESSURE MICROPHONE.

Graetz bridge A full bridge rectifier (i.e., one having a diode in each arm).

Graffiti **1.** Trade name for a specialized written alphabet used for entering data into a handheld computer using a stylus. **2.** The entry of data into a handheld computer using a stylus. See HANDHELD COMPUTER.

grain boundary In a polycrystalline solid, a boundary between single crystalline regions.

gram **1.** Abbreviation, g. A unit of mass and weight, equal to 0.001 kilogram or 0.0353 ounce. **2.** A suffix meaning something drawn (written), or recorded, as in *radiogram* or *electrocardiogram.*

gram atom See GRAM ATOMIC WEIGHT.

gram atomic weight Also called *gram-equivalent.* The quantity of an element with a mass in grams that is equal to the atomic weight of the element. For example, copper has an atomic weight of 63.546; therefore, one gram atomic weight of copper has a mass of 63.546 grams.

gram-calorie Abbreviation, gm-cal. The amount of heat required to raise the temperature of one gram of water by one degree Celsius.

gram-centimeter Abbreviation, gm-cm. The work done by a force of one gram exerted over a distance of one centimeter. Also see JOULE.

gram-equivalent See GRAM ATOMIC WEIGHT.

grammar **1.** The sequence of words and/or abbreviations in a communication or part of a communication. **2.** The sequence of codes and/or commands in a high-level computer programming language.

gramme armature See GRAMME RING.

Gramme ring A type of armature for a motor or generator, consisting of an iron ring onto which is wound a coil of wire, each turn being connected to a commutator bar.

gram-meter Abbreviation, g-m. A unit of work equal to a force of one gram exerted over a distance of one meter. Compare JOULE.

gram-molecular weight See MOL.

gram molecule See MOL.

gramophone Archaic term for *phonograph.*

grandfather cycle A backup scheme in a magnetic reproduction system. The original records are retained for a period of time so that new copies of high precision can be made in case of loss.

grandfather file An original copy of a file on a magnetic disk or tape, retained as a source for reconstruction as needed. Usually, three generations of a file (grandfather, father, and son) are kept, each identified by a generation number. Sometimes the terms *grandparent*, *parent*, and *child* are used instead. See GENERATION NUMBER.

grand synthesizer A hypothetical child with a mind perfectly attuned to artificial intelligence (AI), who grows up to revolutionize the whole industry; a "grand master of AI."

granular carbon Carbon in the form of fine granules, used in the button of a carbon microphone.

granularity **1.** In a digital device, the smallest increment that can be differentiated. **2.** The limit of detail in a reproduction system.

granule One of many narrow frequency subbands, the combination of which composes a complete signal. Each subband carries its own specific data.

graph **1.** A presentation of data, particularly a depiction of the manner in which one variable or set of variables changes, with respect to another. Can be in the form of discrete points, curves, bars, columns, pie-shaped slices, etc. **2.** A curve or set of curves in a coordinate system.

graphical analysis The solution of problems through the use of graphic devices, such as vector diagrams, load lines, Nyquist plots, topological flow diagrams, etc.

graphical harmonic analysis See SCHEDULE METHOD.

graphical user interface Acronym, GUI (pronounced "gooey"). An operating system or software that makes it easy for lay people to use a computer. Commands are given by making choices from among items displayed on the screen. Popular versions use symbols, called *icons*, along with a pointing device, such as a *mouse* or *trackball.*

graphic documentation Records of data in the form of graphs, charts, tables, diagrams, etc.

graphic equalizer A device for tailoring the amplitude-versus-frequency response of a high-fidelity audio system. Consists of a splitter, several audio filters, and a mixer. The gain of each filter is adjustable via a slide potentiometer. The potentiometers are arranged on the front panel in such a way that their relative positions show the approximate shape of the response curve.

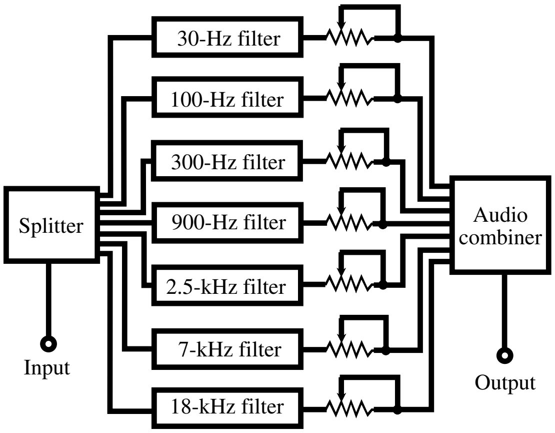

graphic equalizer

graphic instrument See GRAPHIC RECORDER.

graphic-level recorder An instrument that makes a permanent recording of signal amplitude as a function of some independent variable (such as time, frequency, or direction).

graphic panel In process control, a panel of illuminated lights or dials that display the status of a process.

graphic recorder An instrument in which a signal-driven pen or stylus makes a permanent record of a quantity on graph paper. The paper passes at a controlled speed beneath the pen.

graphics **1.** Diagrams, charts, photos, tables, or similar, often symbolic, artwork used to convey information. **2.** The video display in a computer system. **3.** Computer software designed specifically to create and edit illustrations.

graphic solution **1.** A method of depicting the solution(s) to a problem or equation by means of graphs. **2.** The process of solving problems or equations with the aid of graphs.

graphic terminal A display or plotter that provides visual output of a computer run.

graphite A soft form of carbon used in resistors, attenuators, contacts, brushes, vacuum-tube plates, cathode-ray tube coatings, etc.

graphite-line resistor An emergency, makeshift resistor consisting of a pencil line drawn on a piece of paper. The heavier the line for a given width and length, the lower its resistance.

graphophone Archaic term for *phonograph.*

grasping planning A scheme that a robot arm and gripper use to get hold of a particular object. It can use a vision system, a bar-code reader, tactile sensing, and/or proximity sensing. The robot controller (computer) must be programmed to recognize the input from these devices, and to seek out the object.

grass The background noise (noise floor) on the display of a spectrum analyzer; it also appears on certain types of radar displays.

grasshopper fuse A special type of spring-operated fuse. When it burns out, it actuates an alarm that alerts personnel of a possible circuit or system malfunction.

graticule Calibrated gridwork, as on the face of an oscilloscope or spectrum-analyzer screen.

grating A set of parallel, closely spaced, equidistant conductors or bars. When an electromagnetic field that has a wavelength comparable to the conductor spacing passes through the plane containing the conductors, diffraction occurs, producing an interference pattern. An adaptation of this device, called a DIFFRACTION GRATING, is used to separate visible light into its constituent spectral colors.

grating reflector A metal antenna reflector consisting of numerous parallel, straight, closely spaced conductors. When the conductor spacing is much smaller than the wavelength, the set behaves like a solid sheet of metal.

Gratz rectifier A form of full-wave rectifier circuit in a three-phase, alternating-current system.

gravitational constant Symbol, g. The acceleration produced by the attraction of a unit mass at unit distance; $g = 6.673 \times 10^{-11}$ N•m^2/kg^2.

gravitational wave See GRAVITY WAVE.

gravity **1.** The universal force of attraction between material bodies—especially that force evidenced by the earth's drawing of bodies toward its center, causing them to have weight. **2.** Abbreviation, g. The rate at which a free-falling mass accelerates in a vacuum at the earth's surface; equal to 9.802 meters per second per second (m/s^2). **3.** The rate at which a free-falling mass accelerates in a vacuum in the vicinity, or at the surface, of an astronomical object, such as a planet or star.

gravity cell An electrochemical cell in which the positive electrode is made of copper and the negative electrode is made of zinc. The copper electrode is placed at the bottom of a jar, and the zinc electrode is placed at the top. The jar is half filled with copper sulfate solution, and then filled with zinc sulfate solution. The solutions remain separate because copper sulfate has a higher specific gravity than zinc sulfate.

gravity wave A disturbance in a gravitational field, such as might be caused by a collapsing star. These waves might emanate from black holes or rapidly spinning neutron stars.

gray body A radiating body exhibiting constant spectral emissivity at all wavelengths. That is, the emitted energy is the same at all wavelengths and all frequencies.

Gray code A computer code in which the expressions representing sequential numbers differ in only one bit.

gray scale A reference scale for use in black-and-white television and video display images, consisting of several defined levels of brightness with neutral color.

Gray scale:
hypothetical 16-shade binary codes.

Code	Relative shade	Percent brightness
0000	Black	0.00
0001		6.67
0010	Very dark gray	13.33
0011		20.00
0100	Dark gray	26.67
0101		33.33
0110	Medium-dark gray	40.00
0111		46.67
1000	Medium gray	53.33
1001		60.00
1010	Medium-light gray	66.67
1011		73.33
1100	Light gray	80.00
1101		86.67
1110	Off-white	93.33
1111	White	100.00

gray tin A form of tin that exhibits some properties of a semiconductor at temperatures below 18 degrees Celsius.

greatest lower bound The largest value of a parameter that can be obtained without changing some characteristic of a circuit, program, or system.

Greek alphabet The 24-letter alphabet of the Greek language. Virtually all of letters are used as symbols in electronics and related sciences.

Green Book A specialized format for *compact-disk read-only memory (CD-ROM)* computer data storage media, developed by Sony and Philips. Requires the use of a *Compact Disk-Interactive (CD-I)* player for data retrieval. See also CD-ROM, ORANGE BOOK, RED BOOK, and YELLOW BOOK.

green gun The electron gun in a color cathode-ray tube whose correctly adjusted beam strikes only the green phosphors on the screen.

green video voltage In a three-gun color cathode-ray-tube circuit, the green-signal voltage, which actuates the green gun.

Greenwich Civil Time Abbreviation, GCT. Mean time counted from mean midnight at Greenwich, England, the location of zero meridian.

Greenwich Mean Time Abbreviation, GMT. Mean solar time at zero degrees longitude, also called the *Greenwich meridian* because it passes through Greenwich, England. In recent years, GMT has been supplanted by COORDINATED UNIVERSAL TIME (UTC) as the basis of standard time throughout the world.

grid **1.** The prime control electrode in a vacuum tube. Usually, it is a coil or mesh, but it can have other forms. Also called *control grid.* **2.** Any electrode in a vacuum tube placed between the cathode and the anode (plate) (e.g., *screen grid* and *suppressor grid*). **3.** Two sets of straight, uniformly spaced, parallel conducting wires or rods, one set perpendicular to the other. The conductors are electrically connected at all crossing points. Used as an electrostatic or electromagnetic shield, or as a reflector of electromagnetic waves. **4.** Two sets of uniformly spaced parallel lines, one set perpendicular to the other, used as a system of coordinates or as a basis for physical measurements.

grid capacitor **1.** A capacitor in series with the grid of a vacuum tube, used for blocking purposes. **2.** A bypass capacitor in a grounded-grid tube type amplifier. **3.** The capacitor in the grid tank circuit of a tube type oscillator or amplifier.

grid-cathode capacitance Symbol, *CGK*. The internal capacitance between the control grid and cathode of an electron tube. Also called INPUT CAPACITANCE.

grid characteristic The grid-current-versus-grid-voltage performance curve for a vacuum tube.

grid circuit The external circuit associated with the control grid of a vacuum tube.

grid current Symbol, I_G. Current flowing between the control grid and cathode in a vacuum tube.

grid cylinder The metal cylinder that acts as a control grid in a cathode-ray tube.

grid-dip meter **1.** A dip meter that contains a vacuum-tube oscillator; the indicating microammeter is in the grid circuit. **2.** Loosely, any frequency-sensitive wavemeter that indicates resonance by a marked dip in input (base, grid, and gate) current.

grid-dip oscillator See GRID-DIP METER.

grid dissipation **1.** The amount of power given up as heat in the grid circuit of a vacuum-tube amplifier. **2.** The maximum amount of power that a tube can safely dissipate as heat in the grid.

grid drive See GRID EXCITATION.

grid-driving power The signal power required by the control grid of a power tube.

grid emission Electron or ion emission by the control grid of a vacuum tube.

grid excitation Signal voltage or power applied to the control grid in a vacuum-tube amplifier circuit.

grid impedance Symbol, Z_G. The internal impedance of the grid-cathode path in a vacuum tube.

grid input impedance The impedance of the grid input section of a vacuum-tube circuit. It is a complex combination of grid impedance and the impedance of input-circuit components.

gridistor A special form of field-effect transistor with several channels.

grid-limiter resistor A resistor connected in series with the grid of a tube to limit grid current during the positive half-cycle of grid-signal voltage.

grid limiting The cutting off of plate current in a vacuum tube, with consequent limiting action, by means of a high, negative grid voltage developed by overdriving the grid.

grid loading effect The tendency of the internal grid-cathode path of a vacuum tube to load a tuned circuit—especially when the grid draws current.

grid locking A vacuum-tube fault in which the grid potential has become permanently positive because of excessive grid electron emission.

grid mesh The mechanical structure of a grid (e.g., gauze or a metal screen).

grid neutralization See GRID-NEUTRALIZED AMPLIFIER.

grid-neutralized amplifier A neutralized radio-frequency power amplifier in which the neutralizing capacitor is connected from the plate of the tube to the free end of a center-tapped grid-tank coil.

grid north In the grid system of navigation, the direction most nearly corresponding to geographic north.

grid-plate capacitance See PLATE-GRID CAPACITANCE.

grid pool tube A gas-discharge tube in which the cathode is a pool of mercury.

grid power loss Driving-power loss in the grid-input circuit of a power amplifier.

grid resistor A high-value resistor connected between the control grid and ground in a vacuum-tube amplifier circuit.

grid return The circuit path through which the control grid of a vacuum tube is returned to ground or to the negative grid bias supply.

grid-separation circuit A vacuum-tube circuit in which the control grid is grounded. See COMMON-GRID CIRCUIT.

grid swing The peak-to-peak variation of a grid excitation signal.

grid tank A resonant inductance-capacitance circuit operating in the control-grid circuit of a vacuum tube. Compare PLATE TANK.

grid tank capacitance The capacitance required to tune a GRID TANK to resonance.

grid tank inductance The inductance of the coil in a GRID TANK.

grid tank voltage The alternating-current (ac) voltage developed across the grid tank of a vacuum-tube circuit.

grid tuning Tuning of a vacuum-tube circuit by varying the capacitance, inductance, or both in the GRID TANK.

grid voltage **1.** Symbol, V_G. The direct-current (dc) bias voltage applied to the control grid of a vacuum tube. **2.** Symbol, $V_{G(ac)}$. The voltage of the radio-frequency (RF) or audio-frequency (AF) signal in the grid circuit.

grille A covering for an acoustic speaker, used primarily to protect the speaker cone, but also for esthetic appeal.

grille cloth A durable fabric often used for speaker grilles in high-fidelity sound systems. It transmits sound at all audio frequencies, but protects the speaker(s) and provides an attractive physical appearance.

gripper See ROBOT GRIPPER.

grommet An elastic washer inserted through a hole in a chassis to prevent accidental grounding of a conductor or to reduce wear on a cord or cable exiting the chassis.

groove **1.** See KEYWAY. **2.** The fine, spiral line cut into a phonograph disc when it is manufactured.

groove angle On a phonograph disk, the angle between the walls of the unmodulated groove.

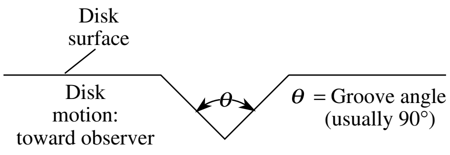

groove angle

groove speed In a phonograph recording or reproducing system, the speed of the cutter or needle, with respect to the disc. The speed is greatest near the outer edge of the disk, and least near the center.

gross content The overall amount of data contained in a message. It can be expressed in bits, bytes, words, or other units.

gross index One of a pair of indexes, used to give a reference in the *fine index*, a supplement; both indexes are used to locate computer records in storage.

gross-motion planning The method(s) that a robot uses to navigate in a general area without running into objects, knocking things over, falling down stairs, or losing its balance. It is often performed using a COMPUTER MAP of the work environment.

ground **1.** The earth in relation to electricity and magnetism. **2.** An electrical connection to the earth. **3.** The return point in a circuit. **4.** A short-circuit to the earth or to a circuit return point. **5.** A short-circuit to the metal chassis, case, or panel of a piece of equipment.

ground absorption The absorption (and resulting loss) of radio-frequency electromagnetic energy by the earth.

ground bus A conductor connected to an earth ground, and to which devices in a system are individually connected. The common ground points (e.g., chassis) of the individual devices are not directly connected to each other, so *ground loops* are avoided. This scheme minimizes the probability of ELECTROMAGNETIC INTERFERENCE to or from the system. Compare GROUND LOOP.

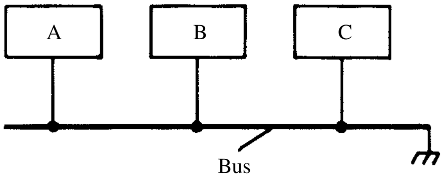

ground bus

ground clamp A device that provides a mechanical and electrical bond between a conductor and a ground rod or pipe. It is generally capable of passing a large amount of current.

ground clutter **1.** In a surface-based radar system, echoes from nearby buildings, hills, and other objects, producing blips or blobs near the center of the screen. **2.** In radar operations, an interference pattern on the screen, caused by accidental grounding in the system.

ground conductivity The ease with which the earth will carry electrical current. Saltwater has the best conductivity; dark, moist soil and fresh water (because of the mineral content) rate from fair to good. Sandy soil has the poorest ground conductivity.

ground conduit A pipe housing one or more ground leads.

ground connection **1.** A low-resistance connection to the earth. **2.** The common point, such as a chassis, to which zero-potential terminals of circuit components are connected.

ground constants The conductivity and dielectric constant of the earth for a particular kind of terrain and soil at a given location. Affects the behavior of radio antenna systems and electrical utility systems. Usually, high conductivity (low resistance) and low dielectric constant are desirable characteristics.

ground-controlled approach Abbreviation, GCA. In air navigation, a ground radar system that provides information for radio-directed aircraft approaches.

ground-controlled interception Abbreviation, GCI. A ground radar system by which an aircraft can be directed to intercept enemy aircraft.

ground current **1.** A direct electric current flowing into the earth from an electrical or electronic device, or into a device from the earth. **2.** An alternating current flowing between the earth and a device. **3.** An electric current flowing through the earth between two points. **4.** A current flowing in the normal ground (low-potential) line of a circuit.

ground detector A device that indicates whether or not a given circuit point is at direct-current ground.

grounded antenna See MARCONI ANTENNA.

grounded-base circuit See COMMON-BASE CIRCUIT.

grounded-cathode circuit See COMMON-CATHODE CIRCUIT.

grounded-collector circuit See COMMON-COLLECTOR CIRCUIT.

grounded-drain circuit See COMMON-DRAIN CIRCUIT.

grounded-emitter circuit See COMMON-EMITTER CIRCUIT.

grounded-gate circuit See COMMON-GATE CIRCUIT.

grounded-grid circuit See COMMON-GRID CIRCUIT.

grounded outlet An outlet with a receptacle having a ground contact that can be connected to equipment-grounding conductors.

grounded-source circuit See COMMON-SOURCE CIRCUIT.

grounded system A set of electrical conductors or a transmission line in which one conductor is deliberately grounded.

ground effect **1.** Modification or distortion of the ideal free-space directivity pattern of an antenna by reflections from, and absorption by, the earth. **2.** Effects of the earth on radio-wave propagation (e.g., the production of a *reflected wave* and a *surface wave*, neither of which can exist if the earth is not part of the signal path). **3.** Effects on antenna behavior, such as modification of the impedance and resonant frequency, caused by the proximity of the earth.

ground efficiency In an antenna system, the quality of the ground circuit. For some antenna systems, such as a balanced dipole at great height, this is not a consideration. Ideal ground efficiency (100%) results in zero ground loss.

ground environment **1.** See GROUND CONSTANTS. **2.** The ground characteristics in the vicinity of an unbalanced antenna working against ground. **3.** In aviation, the set of ground-based installations.

ground fault **1.** Loss of a ground connection. **2.** A short-circuit to ground.

ground-fault interrupter Abbreviation, GFI. A fast-acting electronic circuit breaker that opens the power-line circuit breaker to prevent electric shock or equipment damage when the path of current flow is through the earth.

grounding electrode A device, such as a ground plate or ground rod, that facilitates low-resistance connections to the earth.

grounding plate A metal plate connected to the earth, on which a person stands to discharge static electricity from the body.

grounding rod See GROUND ROD.

ground insulation Electrical insulation used between adjacent energized and grounded parts, such as transformer windings and metal cores.

ground level See GROUND STATE.

ground loop A closed current path resulting from improper grounding of the components in a system. A loop is formed when two devices are connected to each other, and also to separate earth grounds or to a single earth ground via conductors of appreciable length. The loops can act as antennas, increasing the likelihood of ELECTROMAGNETIC INTERFERENCE to or from the system. The use of a single GROUND BUS is preferred.

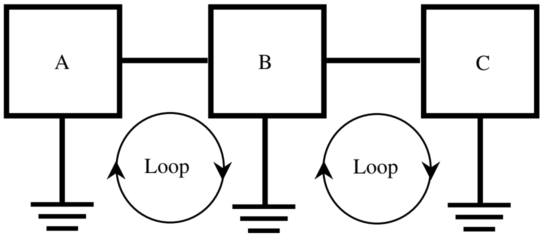

ground loop

ground mat A grid or network of conductors, connected to earth ground, for the purpose of improving the earth conductivity.

ground-mounted vertical antenna A vertical radiator mounted at the earth's surface or at the surface of a body of water, and fed with coaxial cable. The radiator can be any physical length, but it is tuned to resonance at the operating frequency.

The center conductor of the cable is connected to the base of the radiator, and the shield is connected to a ground system. Radial wires, usually buried just below the surface, minimize losses. The antenna can be made resonant on several frequencies by inserting multiple loading coils or traps in the radiator. The chief advantages of this antenna are unobtrusiveness and ease of installation. The main problems are marginal efficiency and susceptibility to human-made electromagnetic noise. See also RADIAL, 1. Compare GROUND-PLANE ANTENNA.

ground noise **1.** Electrical noise that results from a faulty ground connection. **2.** Background noise. **3.** In wire circuits, such as a telephone system, electrical noise that results from fluctuations in ground current.

ground plane **1.** A metal plate or a system of horizontal rods or wires mounted high on a mast, at the base of a vertical antenna, to provide a radio-frequency ground at a point several wavelengths above the surface of the earth. Also see GROUND-PLANE ANTENNA. **2.** In noise and interference tests, a sheet metal structure used to simulate the skin of an aircraft or missile. **3.** On a circuit board, a thin metallic sheet, usually bound to the underside, that serves as a common ground and RF shield.

ground-plane antenna A vertical radiator operated against a system of quarter-wave radials, elevated at least a quarter wavelength above the earth's surface and fed with coaxial cable. The center conductor of the cable is connected to the base of the radiator, and the shield is connected to three or four quarter-wave radials that run outward from the base of the radiator. At shorter wavelengths, a disk or cone can replace the radials. The chief advantages of this antenna are simplicity and high efficiency. The main problem is susceptibility to human-made electromagnetic noise. See also RADIAL, 1. Compare GROUND-MOUNTED VERTICAL ANTENNA.

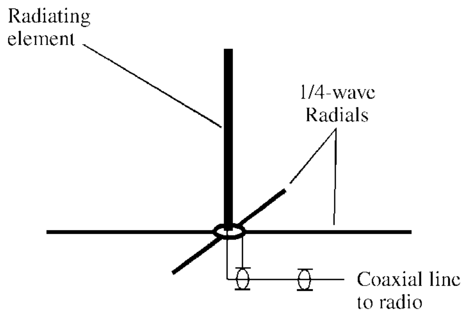

ground-plane antenna

ground plate A metal plate buried in the earth to provide a low-resistance ground connection.

ground-position indicator Abbreviation, GPI. A computer system that gives a continuous indication of an aircraft's position in terms of heading, elapsed time, and speed, with respect to the surface. This provides a more useful indication of the aircraft position than an *air-speed indicator* because it is not affected by high-altitude winds.

ground potential See ZERO POTENTIAL, **3**.

ground protection The use of a GROUND-FAULT INTERRUPTER.

ground-reflected wave A radio wave component that results from ground reflection.

ground reflection The reflection of a radio wave by the earth.

ground resistance The direct-current resistance of a connection to the earth, or the resistance between two points through the earth. The magnitude of the resistance depends on several factors: composition of the soil, amount of moisture, soil electrolytic action, and the area of contact with the earth.

ground return **1.** The point or path used to return a circuit to ground for completion. **2.** In radar, echoes returned from the earth's surface (including reflections from objects on it).

ground-return circuit A circuit, such as a single-wire telephone line, in which earth ground forms one leg of the circuit. Compare METALLIC CIRCUIT.

ground rod A strong metal rod driven deep into the earth as a point of ground connection.

ground speed The speed of an aircraft or missile, relative to the surface of the earth.

ground state The least-energy level of all possible states in a system.

ground support equipment Electronic surface-based apparatus upon which the functioning of a weapons system is dependent.

ground switch A switch for grounding an outside antenna during idle periods. Also called *lightning switch.* Ideally, antennas should be disconnected from equipment, as well as grounded, when not in use.

ground-to-air communication Radio or radar transmission from a land station to an aircraft in flight.

ground-to-ground **1.** Pertaining to communications between land-based stations. **2.** Pertaining to missiles intended for use between points on the surface of the earth.

ground-to-ground communication Communications between land-based stations.

groundtrack For an earth-orbiting satellite, the path followed by the point on the earth's surface from which the spacecraft appears directly overhead. For most satellites, this path shifts toward the west for each succeeding orbit, because the earth rotates eastward underneath the satellite. For geostationary satellites, the point never changes position on the surface. For satellites in equatorial orbits, the track follows the equator.

ground wave In wireless communications and broadcasting, an electromagnetic (EM) wave that consists of three distinct components: the *direct wave* (also called the *line-of-sight wave*), the *reflected wave,* and the *surface wave.* The direct wave is significant only when the transmitting and receiving antennas are connected by a line through free space without obstructions. The reflected wave, after returning from the earth's surface or a human made structure, combines with the direct wave (if any) at the receiving antenna. The surface wave travels in electrical contact with the earth. This occurs only with vertically polarized EM fields at frequencies below about 15 MHz. Below about 300 kHz, the surface wave propagates for hundreds or even thousands of miles. Sometimes the surface wave is called the *ground wave.*

ground wire A conductor between an equipment and a ground connection, either for circuit completion or for safety.

group **1.** A series of computer storage locations containing a specific record or records. **2.** The data in these locations. **3.** A record set having a common key value in a sorted file.

group busy In a telephone system, an audio signal indicating that all of the lines in a group are in use.

group code In digital communications, an error-detecting code used to verify a character group transferred between terminals.

group delay In a modulated signal, a delay in the transmission of data.

grouped-frequency operation In a two-wire communications system, the grouping of directional signals into certain frequency bands.

grouped records A set of data records in which the key of one record identifies the entire set.

grouping **1.** The arrangement of data into blocks or sets. **2.** On a phonograph disc, the insertion of gaps in the arrangement of grooves. **3.** Any periodic irregularity in the spacing of a data transmission. **4.** The bunching of grooves on a disc recording. **5.** In a facsimile system, occasional spacing errors between recorded lines. **6.** A mass of data arranged into groups, according to common characteristics.

group mark **1.** In telegraphy, an indicator that signals the end of a data unit. **2.** A character indicating the end of a character group; usually, it is a logical record that is addressed and processed as a unit.

group velocity The velocity at which a group of waves or a pulse is propagated.

Grove cell A closed-circuit primary cell in which the positive electrode, platinum, is immersed in nitric acid; the negative electrode, zinc, is immersed in sulfuric acid. The nitric acid is held in a porous cup, surrounded by a larger jar of sulfuric acid.

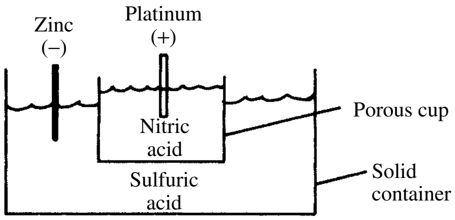

Grove cell

growler **1.** An electromechanical troubleshooting device that indicates the location of short circuits and grounds (especially in electric motors) by emitting a growling or rumbling sound. **2.** Any tester that provides an audible signal, which indicates electrical continuity.

grown-diffused transistor A transistor that is made by first growing the emitter and collector regions as a crystal, into which the base region is later diffused while the crystal is being pulled.

grown diode A semiconductor diode created by growing a layer of p-type material into n-type material (or vice versa) as the single-crystal material is being pulled from the melt.

grown junction A pn junction produced by adding impurities in various amounts to a crystal while it is being pulled from molten semiconductor material.

grown-junction diode See GROWN DIODE.

grown-junction photocell A grown-junction diode used as a photoconductive cell.

grown-junction transistor A transistor made by adding n-type and p-type impurities successively to a crystal in its molten state, then slicing the resulting npn formations from the finished crystal.

G-scan A rectangular radar display consisting of a laterally centered blip that "grows wings" as a target approaches. Horizontal and vertical displacement of the blip indicate horizontal and vertical aiming errors.

G scope See G SCAN.

GSR Abbreviation of GALVANIC SKIN RESPONSE.

G-string antenna In microwave operations, a communications path provided by a dielectric-coated wire that behaves like an extremely low-loss coaxial line with its outer conductor removed to infinity. A horn at each end matches the line to the transmitter and receiver. The term comes

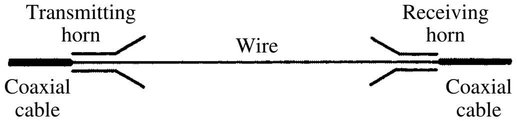

G-string antenna

from the first initial of Dr. George Groubau, inventor of the device, and the stringy appearance of the wire.

GTO Abbreviation of GOTO.

guard band A narrow unoccupied band of frequencies at the upper and lower limits of an assigned channel; its purpose is to prevent adjacent-channel interference by ensuring adequate separation between channels.

guard circle On a phonograph disk, an inner groove that prevents collision of the pickup with the spindle at the center of the disc.

guard circuit An auxiliary circuit added to an alternating-current bridge to compensate for the effects of stray capacitance in the bridge arms. One of its several forms is the WAGNER GROUND.

guarded input An input-terminal arrangement in which one terminal, maintained at the proper potential, shields the entire input-terminal combination.

guarding A method of short-circuiting a leakage current to ground. On a printed-circuit board, guarding is usually accomplished by the use of a large conducting foil surface near critical components.

guard relay A relay that ensures that only one linefinder will be connected to a line circuit when other line relays are in operation.

guard ring A metal ring (or other configuration) surrounding, but separate from, a charged body or terminal, for the purpose of evenly distributing the electric charge over the latter's surface.

guard shield A shield that encloses the input circuit of an amplifier or instrument.

guard terminal In a GUARDED INPUT, the terminal that shields the combination.

guard wire A grounded wire that is intended to catch and ground a broken high-voltage line.

Gudden-Pohl effect The tendency of an ultraviolet irradiated phosphor to glow momentarily when subjected to an electric field.

GUI Acronym for GRAPHICAL USER INTERFACE. Can be spelled out or pronounced "gooey."

guidance Electronic control of the path or course of a robot, missile, or other vehicle.

guidance system The complete electromechanical system for control of a robot, missile, or other vehicle. It consists of hardware and software. The hardware includes beacons, sensors, drive systems, rockets, etc. The software interprets data from, and transmits commands to, the hardware. The nature of the hardware and software depend on the application.

guidance tape In a guided missile, a magnetic tape containing computer instructions for steering the missile in a designated course.

guide See WAVEGUIDE.

guide connector See WAVEGUIDE CONNECTOR.

guided missile A missile whose progress to a target is controlled electronically by signals from a control station or by sensing equipment aboard the missile.

guided propagation A form of radio-wave propagation in which air masses of different temperatures or humidity levels cause refraction and/or reflection of electromagnetic waves, guiding signals over long distances with very little attenuation. Commonly observed at very-high and ultra-high frequencies.

guide elbow See WAVEGUIDE ELBOW.

guide flange See WAVEGUIDE FLANGE.

guide gasket See WAVEGUIDE GASKET.

guide junction See WAVEGUIDE JUNCTION.

guide load See WAVEGUIDE LOAD.

guide slot See KEYWAY.

guide wavelength See WAVEGUIDE WAVELENGTH.

Guillemin effect The tendency for a strip of ferromagnetic material to become straight in a strong magnetic field. This is a form of MAGNETOSTRICTION.

Guillemin line In radar operations, a special pulse-forming network for controlling modulation pulse duration.

guillotine capacitor A variable capacitor in which a sliding (instead of rotary) plate moves between two stator plates. Its name results from its resemblance to the infamous beheading apparatus.

gulp Several bytes of digital information.

gun See ELECTRON GUN.

Gunn diode A semiconductor diode that can operate as an oscillator in the ultra-high-frequency (UHF) and microwave parts of the radio spectrum. Oscillation takes place as a result of a negative-resistance effect in which, within a certain range of applied voltages, the current decreases as the voltage increases. The device is not particularly efficient. Only a small fraction of the input power results in useful signal output. The frequency and oscillation stability are sensitive to changes in temperature and bias voltage. See GUNN EFFECT, NEGATIVE RESISTANCE.

Gunn effect A semiconductor phenomenon named after J. Gunn of International Business Machines (IBM) who first observed and studied it in the 1960s. In certain types of diodes having a negative-resistance characteristic, oscillation occurs when the applied voltage is within a certain range. See GUNN DIODE, NEGATIVE RESISTANCE.

Gunn-effect circuit Any circuit exploiting the Gunn effect, especially a GUNN OSCILLATOR.

Gunn oscillator A discrete semiconductor microwave oscillator using a GUNN DIODE.

Gunnplexer A microwave transmitter consisting of a Gunn-diode oscillator connected to a horn antenna. Usually, the entire device is a compact, self-contained unit. The output power is low. When the device is placed at the focal point of a large dish antenna with the horn pointing toward the dish reflector (conventional dish feed), considerable transmission range is possible. Used primarily by experimenters, and in wireless communications links. See GUNN DIODE.

gutta percha A hard, rubberlike, organic insulating material. Dielectric constant, 3.3 to 4.9. Dielectric strength, 203 to 508 kV/in.

guyed tower In radio communications or in microwave links, a structure that is supported by one or more sets of guy wires to add strength and to prevent collapse.

guying The support of a radio communications or microwave-link tower by the use of one or more sets of guy wires.

guy insulator Also called *egg insulator.* An insulator designed to electrically break a guy wire while maintaining its ability to support a structure. Such an insulator has two slots with holes placed at right angles, in such a position that the wire will not separate even if the insulator breaks. The stress exerted on the insulator is compression, and the insulating material is stronger under this type of stress than under tension (pulling).

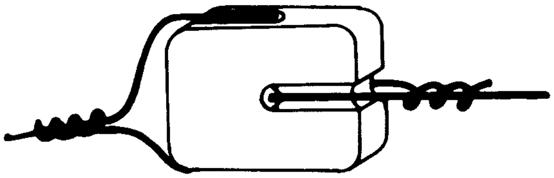

guy insulator

guy wire A bracing wire for antenna masts or towers.

gyrator An active (usually cascaded-transistor) device exhibiting nonreciprocal phase shift. It provides, among other functions, the simulation of inductance using capacitors.

gyro Contraction of GYROSCOPE.

gyro- A prefix meaning “pertaining to gyroscopes,” “containing a gyroscope,” or “behaving like a gyroscope.”

gyrocompass A type of compass in which a spinning gyroscope, acted upon by the earth’s rotation, causes the device to point to true north. Compare MAGNETIC COMPASS.

gyrofrequency The natural frequency of rotation of charged particles around the earth’s magnetic lines of flux.

gyromagnetic Pertaining to the magnetic properties of rotating electric charges (e.g., the effect of electrons spinning inside an atom).

gyromagnetic effect The tendency of a rotating body to become magnetized because of the magnetic field of the earth.

gyropilot See AUTOPILOT.

gyroscope A device that consists of a spinning wheel mounted in a gimbal. The shaft of the wheel will point in one direction, despite the movement of the earth beneath it.

gyrostat See GYROSCOPE.

G-Y signal In a color-television circuit, the signal representing primary green (G) minus luminance (Y). A primary green signal is obtained when the G-Y signal is combined with the luminance (Y) signal. Compare B-Y SIGNAL and R-Y SIGNAL.

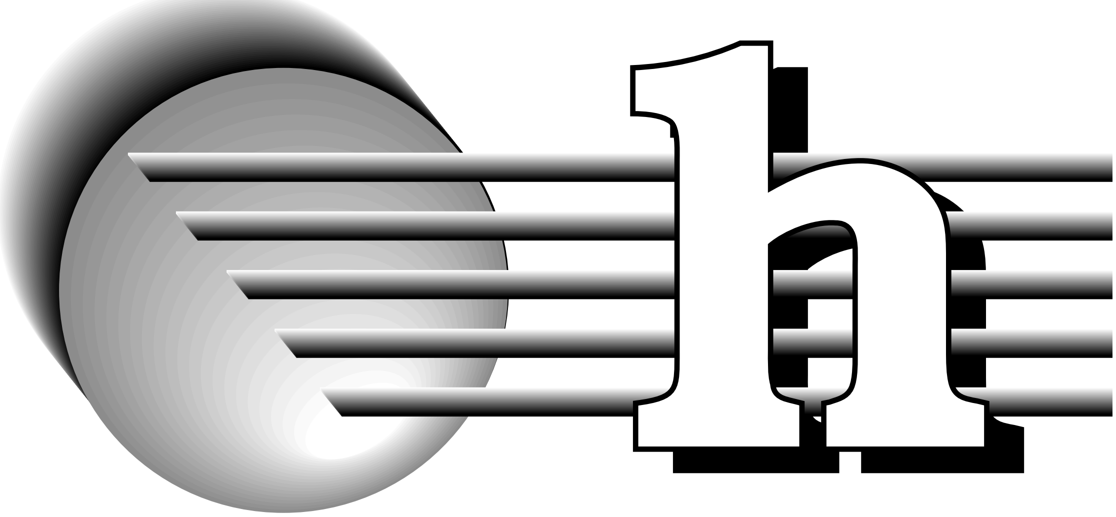

H **1.** Symbol for *magnetic field strength.* **2.** Symbol for MAGNETIZING FORCE. **3.** Symbol for HYDROGEN. **4.** Symbol for UNIT FUNCTION. **5.** Abbreviation of HORIZONTAL. (Also, hor and horiz.) **6.** Symbol for HENRY. **7.** Symbol for HARMONIC.

h **1.** Abbreviation of prefix HECTO-. **2.** Symbol for the PLANCK CONSTANT. **3.** Abbreviation of HOUR.

Haas effect See FUSION, **1**.

hack In computer networking, to access, and sometimes to change, sensitive data without authorization.

hacker A person knowledgeable in computer networking who uses his or her expertise to access, and sometimes alter, sensitive data. This is illegal and can sometimes be destructive. The intent, is rarely malicious, however, the activity is viewed as a challenge. Compare CRACKER.

HACKER program A computer program developed by Gerry Sussman as an early experiment with artificial intelligence (AI), to see how the machine would deal with complex decision-making problems.

hadron A subatomic particle consisting of quarks.

hafnium Symbol, Hf. A metallic element. Atomic number, 72. Atomic weight, 178.49. Readily emits electrons.

hahnium See DUBNIUM.

hailer **1.** A marine microphone-amplifier-speaker system for calling to other boats or persons ashore. **2.** A comparable system for land vehicles, such as police cars. Also see MEGAPHONE, **1**.

hair See HAIRLINE.

hair hygrometer A device for measuring relative humidity, in which a stretched hair is the moisture-sensitive element.

hairline A fine line used as an index or a graticule marker in a precision instrument.

hairpin coil A quarter-turn coil, so called from its resemblance to a hairpin.

hairpin coupling coil A hairpin coil used as a low-impedance primary or secondary coil for input or output coupling.

hairpin match A form of impedance-matching network used at the feed point of a half-wave dipole antenna. A short length of open-wire transmission line, short-circuited at the far end, is connected in parallel with the antenna at the feed point.

hairpin pickup A short, doubled length of wire that acts as a pickup coil at very-high and ultra-high frequencies.

hairspring A fine, usually spiral spring—especially the one in a movable-coil meter or the one connected to the balance wheel of a watch or clock.

hair-trigger Pertaining to extreme sensitivity of response, such as the tendency of a switching device to change state when excited by a weak pulse.

hair wire **1.** An extremely thin wire filament in a lamp or bolometer. **2.** Very small gauge wire (e.g., #44).

hal Abbreviation of HALOGEN.

half-add The sum of two binary digits, in which the carry operation is omitted. Thus 0 + 0 = 0, 0 + 1 = 1, 1 + 0 = 1, and 1 + 1 = 0.

half-adder In digital systems, an adder circuit that can handle the two binary bits that are to be added, but that cannot accommodate a carry signal. Compare FULL ADDER.

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

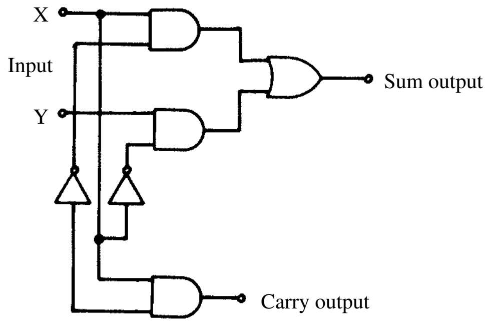

half-adder

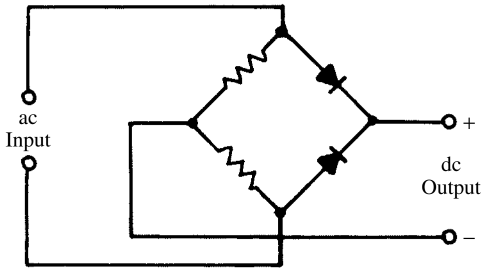

half-bridge

half-bridge A bridge rectifier that has diodes in two arms and resistors in the other two.

half-cell A voltaic cell consisting of a single electrode immersed in an electrolyte and having a definite difference of potential; it is, in effect, half of a primary cell. Also see HELMHOLTZ DOUBLE LAYER.

half-cycle Half of a complete alternation (i.e., 180 degrees of phase).

half-cycle magnetizer A magnetizer using half-cycles of rectified alternating current as the magnetic-field source.

half-duplex channel A communications channel in a HALF-DUPLEX SYSTEM.

half-duplex system In data communications, a system that transmits data in both directions, but not simultaneously. Compare FULL-DUPLEX SYSTEM.

half-lattice crystal filter A band-pass crystal-filter circuit using two piezoelectric crystals in a four-arm bridge. Also see CRYSTAL RESONATOR.

half-nut In a facsimile receiver, a device that guides the lead screw.

half-power point In a response curve or directional pattern, such as for a selective filter or a unidirectional antenna, the points on each side of maximum at which the power is 3 dB below the peak value.

half-power width In a directional antenna system, an expression of beamwidth. It is usually listed as the horizontal-plane angle, in degrees, between the half-power points in the main lobe of the directive pattern.

half-step **1.** In audio engineering, the frequency interval between two sounds, whose ratio is 1.06:1. **2.** The difference in pitch between the notes produced by two adjacent keys on a piano.

half tap A bridging circuit or device that can shunt another circuit with the least electrical disturbance.

half-track recorder A magnetic tape recorder that applies signals to both halves of a tape with a head that covers only half the tape's width in each of two directions. Also called *dual-track recorder.*

half-track tape Magnetic tape recorded by a HALF-TRACK RECORDER.

half-wave Half of a complete wave (i.e., a complete rise and fall in one direction). Its graphic representation is similar in appearance to that for a half-cycle.

half-wave antenna An antenna whose radiator measures an electrical half wavelength from end to end. Such a radiator is about 5 percent less than a free-space half-wavelength long, because of capacitive effects and the velocity factor of the conductor.

half-wave chopper A chopper that closes a circuit during only half the switching signal cycle.

half-wave dipole A center-fed antenna whose radiator measures 0.5 electrical wavelength from end to end. Also see DIPOLE ANTENNA.

half-wave doubler See HALF-WAVE VOLTAGE DOUBLER.

half-wave feeder See HALF-WAVE TRANSMISSION LINE.

half-wavelength Symbol $\lambda/2$. The distance that corresponds to 180 degrees of phase as an electromagnetic (EM) field is propagated in free space, it is related to the frequency by a simple equation:

$$L_{\mathrm{ft}} = 492/f$$

where L_{ft} represents $\lambda/2$ in feet, and f represents the frequency in megahertz. If $\lambda/2$ is expressed in meters, then the formula is:

$$L_{\mathrm{m}} = 150/f$$

where L_{m} represents the displacement in meters. In general, if v is the velocity factor (expressed as a ratio) in a given medium, then:

$$L_{\mathrm{ft}} = 492\mathrm{v}/f$$

and

$$L_{\mathrm{m}} = 150\mathrm{v}/f$$

Compare FULL WAVELENGTH, QUARTER WAVELENGTH.

half-wave loop antenna A loop antenna having a circumference of 0.5 wavelength with a break opposite the feed point. It is, in effect, a HALF-WAVE DIPOLE bent into a circle or square (although any symmetrical configuration can be used). The circle is the most efficient configuration.

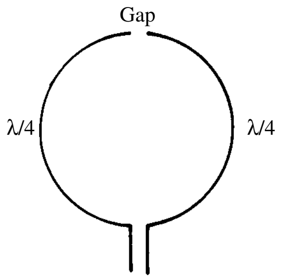

half-wave loop antenna

half-wave radiator An antenna consisting of a single, usually straight, active element that measures an electrical half wavelength from end to end. It is therefore a resonant element. A simple half-wavelength (λ/2) conductor with a high length-to-diameter ratio measures approximately 95 percent of λ/2 in free space. The element can be much shorter than free-space λ/2 yet remain λ/2-resonant if inductance is inserted in series with the radiator. The element can be much longer than free-space λ/2 yet remain λ/2-resonant if distributed capacitances are inserted in series with the radiator.

half-wave rectification The conversion of alternating current (ac) to direct current (dc) during half of each ac cycle. Also see HALF-WAVE RECTIFIER.

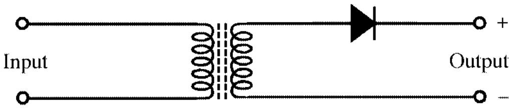

half-wave rectifier

half-wave rectifier A rectifier that delivers a half-cycle of direct-current (dc) output for every other half-cycle of applied alternating-current (ac) voltage. Because the successive dc half-cycles are 180 degrees apart, they have the same polarity. Compare FULL-WAVE RECTIFIER.

half-wave transmission line A transmission line measuring 0.5 electrical wavelength at the transmission frequency. The physical length is somewhat less than a free-space half wavelength because of the VELOCITY FACTOR of the line.

half-wave vibrator A vibrator (see INTERRUPTER) whose reed operates against only one stationary contact. Compare FULL-WAVE VIBRATOR.

half-wave voltage doubler A voltage-doubler circuit whose direct-current (dc) output has a ripple frequency equal to that of the alternating-current (ac) supply. Although its output is harder to filter than that of a full-wave doubler, this circuit has the advantage of a common ground. Compare FULL-WAVE VOLTAGE DOUBLER.

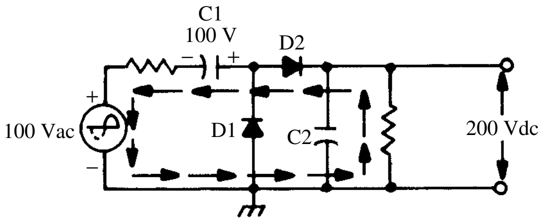

half-wave voltage doubler

halide A compound of a HALOGEN. Examples: sodium iodide, used as a scintillating crystal; ammonium chloride, used as the electrolyte in a dry cell.

halide crystal A halogen-compound crystal, such as mercuric iodide and sodium iodide, useful in detecting radioactivity.

Hall coefficient For a current-carrying conductor, the constant relationship between the Hall (transverse electric) field and the magnetic flux density.

Hall constant For a current-carrying conductor, the constant of proportionality k given by the equation $k = e/(im)$, where e is the transverse electric field (Hall field), i is the current density, and m is the magnetic field strength.

Hall effect A phenomenon observed in thin strips of metal and in some semiconductors. When a strip carrying current longitudinally is placed in a magnetic field that is perpendicular to the strip's plane, a voltage appears between opposite edges of the strip that, although feeble, will force a current through an external circuit. The voltage is positive in some metals (such as zinc) and negative in others (such as gold). Also see ETTINGHAUSEN EFFECT, NERNST EFFECT, and RIGHT-LEDUC EFFECT.

Hall-effect modulator A device that uses the HALL EFFECT to modulate a signal, or to mix two signals.

Hall-effect multiplier A device based upon the Hall generator and used in analog mathematical operations, such as multiplication and the extraction of roots.

Hall field The transverse electric field of a conductor carrying current in a magnetic field.

Hall generator A semiconductor device exhibiting the HALL EFFECT. It is a thin wafer or film of indium antimonide or indium arsenide with leads on opposite edges.

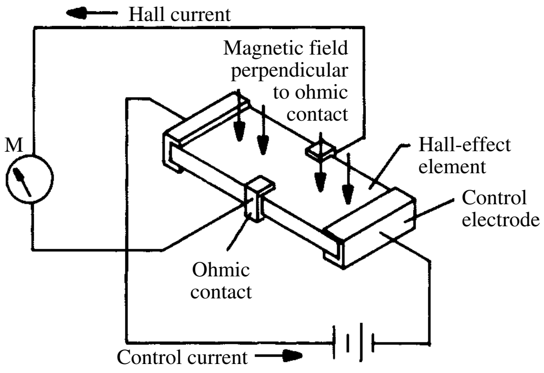

Hall generator

Hall mobility For a conductor or semiconductor, the product of conductivity and the HALL CONSTANT.

Hall network A resistance-capacitance null circuit whose general configuration is two cascaded high-pass tee-sections bridged by a high resistance. The circuit can be tuned with one potentiometer.

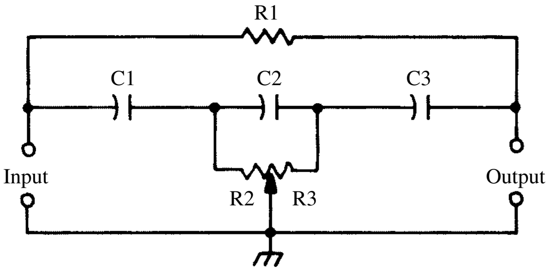

Hall network

Hallwacks effect The phenomenon (observed by Hallwacks in 1888) in which ultraviolet light falling on a polished zinc plate causes a negatively charged electroscope to which it is connected to discharge.

hallucination In complex computers and artificially intelligent systems, the generation or appearance of data for no apparent reason.

halo See AFTERGLOW and PERSISTENCE.

halo antenna A horizontally polarized antenna, consisting of a circular half-wave dipole whose ends are capacitance loaded. Commonly used at very high frequencies (VHF).

halogen Abbreviation, hal. A group of five very active nonmetallic elements whose similar chemical properties put them in group VIIA of the periodic table; they are *astatine*, *bromine*, *chlorine*, *fluorine*, and *iodine*.

Halowax A chlorinated naphthalene wax used as an impregnant for paper capacitors. Dielectric constant, 3.4 to 5.5. Resistivity, 10^{13} to 10^{14} ohm-cm.

halt A stop during the execution of a computer program run, often resulting from a HALT INSTRUCTION.

halt instruction An instruction in a computer program that causes a break in the program's execution, as by BASIC's STOP command, for example.

ham Colloquialism for AMATEUR RADIO operator.

ham radio See AMATEUR RADIO.

Hamilton's principle Also called the *principle of least action.* Motion tends to occur in such a way that the integral of the product of kinetic energy and elapsed time is minimal.

hammer **1.** The striking member in a WHEEL PRINTER. **2.** The clapper in an electric bell or gong.

hammer-and-wheel See WHEEL PRINTER.

Hamming code An error-correction code used in some digital communications circuits.

hand capacitance Also called *body capacitance.* Capacitive coupling effects between a circuit and the human body (e.g., as evidenced between an operator's hand and a device having extremely high impedance and poor grounding and/or shielding).

hand generator An electric generator operated by turning a hand crank.

handheld computer Also called *personal digital assistant (PDA)* or *palmtop computer.* The names PalmPilot and Palm are proprietary (Palm Computing, Inc.) and refer to specific families of handheld computers, although they might someday become generic and refer to handheld computers in general. A battery-powered portable computer, smaller than a notebook computer, and used for simple tasks such as note-taking and record keeping. Some units incorporate wireless modems for connection to the Internet. Others include paging, wireless fax, videoconferencing capability, remote-control capability, and other features. Many units can recognize a specialized form of handwriting so users can enter data with a penlike device called a *stylus.*

Handie-Talkie Abbreviation, HT. Tradename for a portable transceiver small enough to be held in the hand during operation.

hand key Also called *brass pounder.* An old-fashioned, hand-operated telegraph key, operated by manual downward pressure.

handoff In cellular communications networks, the changeover of reception from, and transmission to, a mobile or portable set from one repeater to another as the subscriber moves from one cell into another. When the subscriber is moving rapidly—for example, driving along a freeway, such transfers occur relatively often. When a subscriber is moving slowly, for example, walking along a trail, such transfers occur rarely. When a subscriber is in a fixed location, such transfers do not normally occur.

hand-operated device A device manipulated directly and manually by the operator's hand(s). Also called *manual device.*

hand receiver **1.** A single earphone that must be held against the ear. **2.** A telephone receiver.

hand rules See FLEMING'S LEFT-HAND RULE, FLEMING'S RIGHT-HAND RULE, and RIGHT-HAND RULE FOR WIRE.

handset See CRADLEPHONE.

handshaking **1.** A controlled, periodic exchange of synchronizing pulses between a digital transmitter and receiver. **2.** In a digital communications system, a method of error correction. The receiver detects nonstandard or improbable character sequences, and instructs the transmitter to repeat them for double-checking.

hand-type pointer In an electric meter, a spear-like pointer (resembling the hand of a clock), as opposed to a knife-edged pointer.

hand-wired Pertaining to electronic equipment wired by hand, rather than being assembled on printed-circuit boards. This form of construction is rarely seen nowadays, except in some radio-frequency power amplifiers.

hang AGC An automatic-gain-control (AGC) circuit whose action is sustained for a brief interval after an actuating signal has passed, an advantage in some applications. Also called *fast-attack/slow-release AGC.*

hangover In sound operations, the blurring or smearing of low-frequency (bass) notes by a poorly damped or poorly mounted loudspeaker.

hangup **1.** In phonograph operation, the state in which the same material is played repetitiously (i.e., the stylus does not move toward the spindle). **2.** In digital-computer operations, an unexpected break during a program run as a result of software or hardware failure. Sometimes called UNEXPECTED HALT.

H antenna See LAZY-H ANTENNA.

hard copy **1.** In digital computer operations, a readable document (printout) of material being translated to a form understood by a computer. **2.** Generally, written or typed documents, as opposed to data on other media, such as diskettes, tapes, CD-ROM, etc.

hard disk An electromechanical data storage medium commonly used in personal computers. Consists of several rigid disks, called *platters*, coated with ferromagnetic material.

hard-drawn wire High-tensile-strength unannealed wire.

hard dump See HARDWARE DUMP.

hard magnetic material High-retentivity magnetic material. Also see RETENTIVITY.

hardness **1.** The property that causes a material to resist penetration, deformation, scratches, etc. **2.** The penetrative ability of ultraviolet rays, X rays, or other ionizing radiation. Generally, the radiation hardness increases as the wavelength decreases, and as the photon or particle energy increases.

hardness tester A device for measuring the hardness of a solid in terms of the force required to penetrate its surface. Also see HARDNESS, **1**.

hard radiation In general, any radiation with high penetrating power. Usually, this term is used in reference to short-wavelength (high-energy) ultraviolet rays or X rays.

hard solder Solder that melts at a comparatively high temperature. Compare SOFT SOLDER.

hard vacuum A nearly perfect vacuum, that is, a medium essentially devoid of atomic or subatomic particles.

hardware **1.** Collectively, electronic circuit components and associated fittings and attachments. **2.** In a computer system, the electronic and electromechanical components (e.g., integrated circuits, keyboards, and disk drives) associated with operation. Compare SOFTWARE.

hardware availability ratio A figure depicting the availability of a computer system to do productive work; as a percentage, it is given by the formula:

$$A = 100(ta - td)/ta,$$

where A is the availability ratio, ta is the operational time, and td is the downtime over a specified time period.

hardware check A check on data being transferred within a computer, as done by hardware (e.g., a parity check).

hardware cloth A finely woven wire screen sometimes used in place of a metal plate for an antenna element, an antenna reflector, or a shielded enclosure. Especially useful when free-air circulation is required.

hardware dump During a computer program run, data sent to a storage device for later evaluation; it occurs at the time of a failure. Also called AUTOMATIC HARDWARE DUMP.

hardware engineer A person who designs and perfects the actual electronic circuitry in a system. The hardware engineer is not involved with the programming of the system.

hardware recovery A computer system's ability (through software or hardware) to recover from a failure (i.e., to proceed from the point of failure).

hardware serviceability ratio See HARDWARE AVAILABILITY RATIO.

hardwire **1.** To construct a circuit for direct-current conductivity. **2.** A circuit exhibiting direct-current conductivity over a complete, closed path.

hard-wire telemetry See WIRE-LINK TELEMETRY.

hard wiring **1.** In computer systems, functions or programs built directly into the machine hardware. In order to alter such functions or programs, the system wiring and/or components must be physically changed. **2.** A system interconnected entirely by wires and cables, and using no free-space links, such as radio or infrared.

hard X rays High-frequency (shortwave) X rays. Such radiation has high penetrating power. Compare SOFT X RAYS.

harmonic **1.** Symbol, H. In a complex sound or signal wave, a component whose frequency is a multiple of the FUNDAMENTAL FREQUENCY by a whole-number factor of 2 or more. **2.** Pertaining to whole-number multiples of the FUNDAMENTAL FREQUENCY of a sound or signal, as defined in **1**.

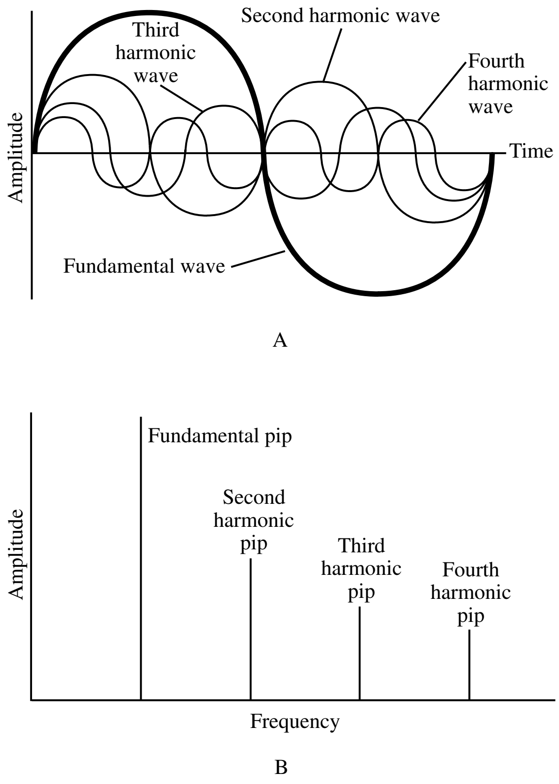

harmonic: waves (A) and spectrum-analyzer display (B)

harmonic accentuation Increasing the amplitude of harmonic components in a complex wave using filters, amplifiers, or special modes of operation.

harmonic accentuator A circuit or device, such as a harmonic amplifier or bandpass filter, for emphasizing signal harmonics.

harmonically related bands In communications, frequency bands arranged so that the frequencies in one band are harmonics of the frequencies in another band. An example of bands that are precisely related in this way are 4.1 to 4.3 MHz and 8.2 to 8.6 MHz. Various amateur radio bands are harmonically related to some extent, such as 40 meters (7.0 to 7.3 MHz) and 20 meters (14.0 to 14.35 MHz).

harmonic amplifier An amplifier, such as one used with a frequency standard, used to increase the amplitude of weak harmonics. Also see HARMONIC ACCENTUATION.

harmonic analysis **1.** The evaluation of the harmonic content of a complex wave. See, for example, HARMONIC WAVE ANALYZER; SCHEDULE METHOD, **2**; SPECTRUM ANALYZER; and WAVE ANALYZER. **2.** See FOURIER ANALYSIS.

harmonic analyzer See HARMONIC WAVE ANALYZER, SPECTRUM ANALYZER, and WAVE ANALYZER.

harmonic antenna An antenna operated at a harmonic of the lowest frequency at which it is resonant. For example, a half-wave dipole cut for 7.0 MHz, but used for transmitting and receiving at 21.0 MHz, is functioning at the third harmonic.

harmonic attenuation Reduction of the amplitude of harmonic components in a complex wave using filters, tuned amplifiers, or special modes of operation.

harmonic attenuator A circuit, device, or method of operation (such as a filter, tuned amplifier, special biasing, or special bypassing) for reducing the amplitude of harmonics.

harmonic component See HARMONIC.

harmonic composition See HARMONIC DISTRIBUTION.

harmonic content The amount of harmonic energy present in a complex wave. Also see HARMONIC-DISTORTION PERCENTAGE and HARMONIC RATIO.

harmonic-cut crystal Also called *overtone crystal.* A quartz crystal that, when operated in the proper circuit, oscillates at a harmonic of the (fundamental) frequency dictated by its thickness.

harmonic detector A detector tuned to respond to a harmonic of a signal.

harmonic distortion **1.** The generation of harmonics by the circuit or device by which the signal is processed. **2.** The deformation of the original signal that results from the action described in **1**. **3.** The disproportionate reproduction of a signal's harmonic components.

harmonic distortion meter See DISTORTION METER.

harmonic-distortion percentage In a signal containing harmonics, the harmonic energy as a percentage of the total signal energy (fundamental plus all harmonics). Also called *total harmonic distortion (THD).*

harmonic distribution For a given signal, the various frequencies and amplitudes of its harmonics, specified within a certain range of frequencies.

harmonic elimination The complete removal of one or more harmonics from a complex wave using a filter or special mode of operation.

harmonic eliminator A circuit or device, such as a band-suppression filter, for removing harmonics.

harmonic filter **1.** A bandpass filter for transmitting one or more harmonics of a complex input wave. **2.** A band-suppression filter for removing one or more harmonics of a complex input wave.

harmonic frequency **1.** In a complex wave, the frequency of a component that is a multiple of the FUNDAMENTAL FREQUENCY by a whole-number factor of two or more. **2.** A frequency that is a whole-number (two or more) multiple of another frequency to which it is referred. Compare NONHARMONIC FREQUENCY.

harmonic generator **1.** An oscillator operated so that it generates strong harmonics of the fundamental frequency. **2.** See FREQUENCY MULTIPLIER. **3.** See HARMONIC AMPLIFIER.

harmonic intensification See HARMONIC ACCENTUATION.

harmonic intensifier See HARMONIC ACCENTUATOR.

harmonic interference Interference resulting from the harmonics of radio or test signals.

harmonic motion Periodic motion typified by a swinging pendulum and illustrated by the plot of a sine wave.

harmonic oscillator A crystal oscillator whose output frequency is a harmonic of the crystal frequency.

harmonic percentage See HARMONIC-DISTORTION PERCENTAGE.

harmonic producer **1.** An oscillator that uses a tuning fork to establish the fundamental frequency. The output can be an odd or even harmonic of this frequency. **2.** See FREQUENCY MULTIPLIER. **3.** A nonlinear circuit used in a calibrator to generate markers at integral multiples of the fundamental frequency.

harmonic ratio **1.** In a complex wave, the ratio of harmonic energy to total signal energy (fundamental plus all harmonics). **2.** In a complex wave, the ratio of harmonic energy to fundamental-frequency energy.

harmonic reducer See HARMONIC ATTENUATOR.

harmonic reduction See HARMONIC ATTENUATION.

harmonic resonance Resonance of an antenna or a circuit at a whole multiple of the applied signal frequency.

harmonic ringing In wire telephony, the use of alternating-current signal harmonics for selective ringing.

harmonic series of tones A set of audio-frequency tones in which the frequencies can be specified by *f*, 2*f*, 3*f*, 4*f*, and so on.

harmonic suppression See HARMONIC ELIMINATION.

harmonic suppressor See HARMONIC ELIMINATOR.

harmonic tolerance The harmonic content permissible in a given system.

harmonic totalizer An instrument for measuring *total harmonic distortion.* See, for example, DISTORTION METER.

harmonizer A circuit that changes the frequency of an audio signal, or produces an output at several audio frequencies from an input having only one audio frequency. Used in sound recording for special effects.

harness A tied bundle of wires or cables for wiring electronic equipment.

harp antenna A vertical antenna consisting of a number of wires that fan out from point to point along a horizontal supporting wire.

hartley A unit of digital information equivalent to 3.32 bits. Used in certain computer applications.

Hartley oscillator A radio-frequency (RF) oscillator that uses a single inductor with a tap on the windings to provide the feedback. The amount of feedback is controlled by the position of the coil tap. A variable capacitor in parallel with the inductor determines the oscillating frequency and allows for frequency adjustment. The circuit uses about 25 percent of its output power to produce feedback. The other 75 percent of the power can be delivered to external circuits or devices. Compare COLPITTS OSCILLATOR.

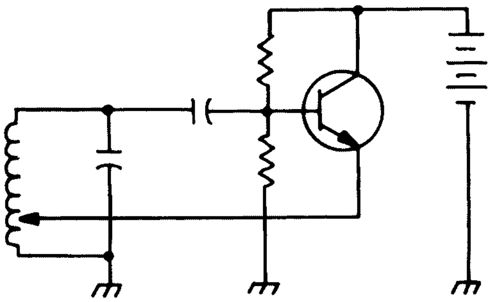

Hartley oscillator

hash **1.** Electrical noise, especially wideband noise with a characteristic hissing sound in a radio receiver. **2.** Undesirable or purposefully meaningless information, as used in a hash total (checksum).

hash filter A radio-frequency filter for eliminating HASH noise in a radio receiver.

hassium Symbol, Hs. Also called *unniloctium* (Uno). Atomic number, 108. The most common isotope has atomic weight 265. Classified as a transition metal. It is human-made and not known to occur in nature.

hat **1.** Also called *capacitance hat.* A small disk or set of wires attached to the end(s) of an antenna radiator, lowering the resonant frequency and increasing the usable bandwidth. **2.** A procedure for randomizing data.

hash total See CHECKSUM.

hatchdot pattern A television test pattern consisting of a crosshatch pattern with dots around its outer edges and one dot at its center.

hatted code A form of code in which randomization is used to maximize the difficulty of breaking the code.

Hay bridge An alternating-current bridge for measuring the inductance and Q of an inductor in terms of resistance, frequency, and a standard capacitance.

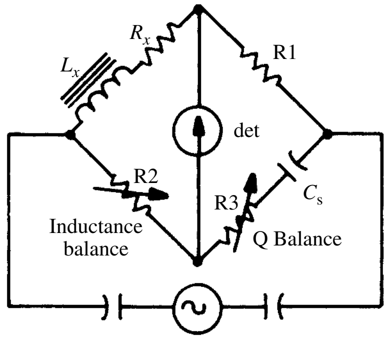

Hay bridge

haywire Loose, disorderly, or apparently careless wiring.

haz Abbreviation of HAZARD.

hazard Abbreviation, haz. A dangerous or potentially dangerous circuit, device, material, method, situation, or system (e.g., *electric-shock hazard*).

H beacon A form of homing beacon with an omnidirectional radiation pattern and a radio-frequency output of between 50 W and 2 kW.

H bend See H-PLANE BEND.

HCD Abbreviation of *hard-copy device.*

HCM Abbreviation of HALF-CYCLE MAGNETIZER.

HDB-3 code Abbreviation of HIGH-DENSITY BIPOLAR-3 CODE.

HDF Abbreviation of HIGH-FREQUENCY DIRECTION FINDER.

HDTV Abbreviation of HIGH-DEFINITION TELEVISION.

H-display See H-SCAN.

He Symbol for HELIUM.

head **1.** The top or operating portion of a device (e.g., microphone head or dynamic-speaker head). **2.** In magnetic recording and reproduction, the magnetic device (transducer) that delivers or picks up recorded impulses. **3.** In a hard disk or diskette drive, a transducer that delivers and picks up recorded data.

head alignment **1.** Positioning the cone of a dynamic speaker so that the voice coil moves freely (i.e., without rubbing against the core). **2.** Positioning a magnetic-recorder head so that a proper relationship to the moving tape is maintained.

head amplifier A self-contained amplifier or preamplifier in the head of a microphone or sound-on-film pickup.

head degausser A device used for the purpose of demagnetizing the head of a tape recorder. Unwanted magnetization can build up because of direct-current components in the driving signal.

head demagnetizer See HEAD DEGAUSSER.

head end In a television network or system, the location from which signals are sent to subscribers.

header **1.** A (usually glass) disk or wafer through which one or more leads pass and to which they are fully sealed. Can be used as the terminal base of an enclosed plug-in unit, such as a miniature coil, filter, or similar components. Also see GLASS-TO-METAL SEAL. **2.** A data set placed before other sets as a means of identifying them and, possibly, including control data pertinent to the sets so identified.

header capacitance Capacitance between or among the leads in a header (see HEADER, **1**).

header label A header recorded on a magnetic tape file (see HEADER, **2**).

head gap **1.** In computer disk or tape drives, the distance between the head and the magnetic medium. **2.** In audio operations, the spacing between tape-unit head electrodes; also called *gap width.*

heading The direction taken by a vehicle with reference to some point (such as a radio beacon, true north, or magnetic north).

headlight In radar operations, a small rotating antenna.

headphone A small acoustic transducer worn against the ear for listening to music without disturbing others, or for monitoring live or recorded material without being disturbed by noise in the environment. Also see RECEIVER, **2**.

headphone amplifier An audio-frequency amplifier designed and operated primarily to supply a signal to headphones.

headphone receiver A portable radio receiver, usually for AM and/or FM broadcast, consisting of a pair of headphones or a headset with the radio built into it.

head room **1.** In a high-fidelity sound system, the extent, measured in decibels, to which an amplifier can be operated beyond the zero point on its volume-unit (VU) meter without causing objectionable distortion on sound peaks. **2.** In tape recording, the region between the maximum recording level specified by the manufacturer of the equipment, and the amplitude at which tape overload occurs. It is specified in decibels.

headset An assembly consisting of one or two earphones, a headband, and a flexible cord. Also see HEADPHONE and RECEIVER, **2**.

head stack In magnetic recording, an assembly of two or more heads for multitrack service. Also see HEAD, **2**.

head station See BASE STATION.

head-to-tape contact In magnetic-tape recording or playback, physical contact between the tape and the head.

hearing aid A miniature audio-frequency device that amplifies sound for people with impaired hearing. It consists of a microphone, a high-gain amplifier, and an earphone or bone-conduction transducer.

hearing-aid battery A physically small battery designed for use with hearing aids. Such a battery is usually of the lithium type, or some other type that has long life under conditions of low current drain.

hearing loss A measure of hearing impairment. Generally expressed as the ratio, in decibels, of an individual's threshold of hearing to the normal threshold of hearing. Also see AUDIOLOGIST, AUDIOMETER, and AUDIOMETRIST.

heart fibrillation A condition in which the heart muscle twitches at random, rather than pumping blood normally. This can be caused by an electric shock through the heart of 100 mA to 300 mA. If normal heart function is not restored, death will follow.

heart pattern See CARDIOID PATTERN.

heart telemetry See ECG TELEMETRY.

heat A form of energy transferred by conduction, convection, or radiation between two bodies having different temperatures. The amount of heat is expressed in degrees, British thermal units, calories, joules, or kelvins.

heat aging **1.** The degeneration of a substance, aggravated by high temperatures. **2.** A test that indicates the immunity of a substance to degeneration because of high temperatures.

heat coil A device that disconnects a circuit when the temperature reaches a certain minimum level.

heat detector A sensor of heat. See, for example, BOLOMETER, INFRARED DETECTOR, MICRORADIOMETER, RADIOMETER, THERMISTOR, THERMOCOUPLE, and THERMOPILE.

heated-pen recorder See THERMAL RECORDER.

heated-stylus recorder See THERMAL RECORDER.

heated-wire flowmeter See HOT-WIRE FLOWMETER.

heated-wire sensor A hot wire used to discriminate between substances, according to how they affect its heating. See, for example, GAS DETECTOR (also usable as a vacuum gauge), HOT-WIRE ANEMOMETER, and HOT-WIRE MICROPHONE.

heat engine A machine that converts heat energy into mechanical energy.

heater **1.** The filament of an indirectly heated vacuum tube. **2.** The filament in an indirectly heated thermistor.

heater-voltage coefficient The amount of frequency change per volt of fluctuation in the filament voltage of a Klystron.

heat exchanger A device or system that removes heat from a hot body and transfers it to another body or to the surrounding air.

heat-eye tube An infrared-sensitive device used for the purpose of locating objects in visible darkness. The tube consists of a cathode-ray device that is sensitive to infrared radiation.

heat gradient The temperature difference between two points on a body, divided by the distance between the two points.

heating depth See DEPTH OF HEATING.

heating effect The production of heat (power loss) by electric current flowing in a conductor.

heating element **1.** See HEATER. **2.** The resistance element (such as a strip or coil) that generates heat in an electric-heating device.

heat loss **1.** Heat emitted by conduction, convection, or radiation from a body at a relatively high temperature. **2.** Power loss as a result of the heating effect of an electric current.

heat of fusion The amount of heat required to melt a unit mass of a solid that has reached its melting point.

heat of radioactivity Heat generated during the process of radioactive disintegration.

heat of reaction In a chemical or electrochemical reaction, the heat (in calories) absorbed or released.

heat of vaporization The amount of heat required to convert 1 gram of a liquid to a vapor without raising its temperature.

heat radiator See HEATSINK.

heat rays See INFRARED RAYS.

heat remover **1.** See HEATSINK. **2.** A forced-air or forced-liquid cooling system.

heat-resistant glass See PYREX.

heat-sensitive resistor See THERMISTOR.

heat-sensitive switch A make-and-break device, such as a thermostat, that is actuated by a change in temperature.

heat-shrink tubing An insulated flexible sleeving made from a plastic that shrinks permanently for a tight fit when heated; it is commonly used at the joint between a cable and connector.

heatsink A heat exchanger in the form of a heavy, metallic mounting base or a set of radiating fins. It conducts heat away from such devices as power transistors, heavy-duty resistors, or power tubes, and dissipates the heat into the surrounding environment via convection and radiation.

heatsink resistance The opposition offered by a heatsink to the flow of heat.

heat therapy **1.** The use of radio-frequency heating for therapeutic purposes. Also see DIATHERMY. **2.** The use of infrared rays for therapeutic purposes.

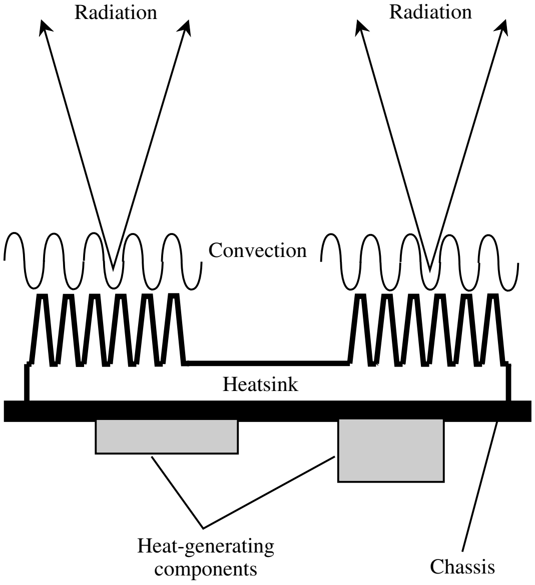

heatsink

heat transfer The movement of heat from one point to another by absorption, conduction, convection, or radiation.

heatronic Pertaining to the heating of a dielectric material subjected to a high voltage.

heat unit **1.** See BRITISH THERMAL UNIT. **2.** See CALORIE. **3.** See KELVIN.

heat waves See INFRARED RAYS.

heat writer See THERMAL RECORDER.

Heaviside-Campbell bridge A form of mutual-inductance bridge. Mutual inductance is determined without regard to the operating frequency.

Heaviside layer See KENNELLY-HEAVISIDE LAYER.

heavy hydrogen An isotope of hydrogen. The term is applied to *deuterium*, whose nucleus consists of one proton and one neutron, and also to *tritium*, whose nucleus consists of one proton and two neutrons.

heavy metal A metal having a specific gravity of 5.0 or higher. Examples: iron (7.85 to 7.88), lead (11.3), nickel (8.6 to 8.9), mercury (13.6), platinum (21.4).

heavy water Formula, D_2O. Water in which deuterium (HEAVY HYDROGEN), rather than ordinary hydrogen has combined with oxygen.

hecto- Abbreviation, h. A prefix meaning hundred(s), (i.e., 10^2).

hectometric wave An electromagnetic field whose wavelength is on the order of hundreds of meters (i.e., at least 100 meters, but less than 1000 meters). The frequency ranges from 300 kHz to 3 MHz.

hectowatt Abbreviation, hW. A unit of power equal to 100 watts. Seldom used; power in this range is usually expressed in terms of the WATT or the KILOWATT.

heelpiece A part of an electronic relay that provides mechanical support for the armature.

Hefner candle A unit of luminous intensity equal to 0.9 candela; the standard (German) is the HEFNER LAMP.

Hefner lamp A standard light source whose luminous intensity is 0.9 candela. It burns amyl acetate (banana oil) and its flame has been the standard of the HEFNER CANDLE, a unit of luminous intensity devised in Germany. Also see CANDLE POWER and LUMINOUS INTENSITY.

height control In a television receiver circuit, the potentiometer or rheostat that controls the vertical dimension of the picture by varying the amplitude of vertical scanning pulses.

height finder An altitude-measuring radar system.

height-position indicator Abbreviation, HPI. A radar displaying the height of a target, its angular elevation, and the slant range.

Heil oscillator An oscillator based on a special tube consisting of a heated cathode, first anode, metal cylinder, and second anode. Electrons emitted by the cathode pass through a hole in the first anode and become a beam, which passes through the cylinder and strikes the second anode (collector). Electron bunching in the cylinder causes energy to be transferred to a tank circuit between the cylinder and anodes.

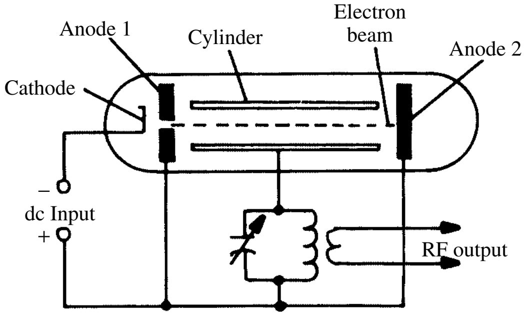

Heil oscillator

Heisenberg uncertainty principle See UNCERTAINTY PRINCIPLE.

hekto- See HECTO-.

heliacal cycle See SUNSPOT CYCLE.

helical antenna A spring-shaped antenna mounted perpendicular to a flat metal-plate reflector, an arrangement that produces circularly polarized waves in a narrow beam. It is used primarily at ultra-high and microwave radio frequencies.

helical-beam antenna See HELICAL ANTENNA.

helical line The helix in a backward-wave oscillator or traveling-wave tube.

helical potentiometer A potentiometer whose resistance element is a wire wound into a coil of several turns. The slider moves over the wire (or the larger coil) from one end to the other as the slider or coil is turned through several complete revolutions. Also called MULTITURN POTENTIOMETER.

helical scanning Radar scanning by an antenna that moves vertically as it moves horizontally, producing a spiral motion to the radiated beam.

helical sweep See SPIRAL SWEEP, **1**, **2**.

helical transmission line See HELICAL LINE.

helicone An antenna used at ultra-high and microwave frequencies, consisting of a helical radiator within a cone-shaped reflector.

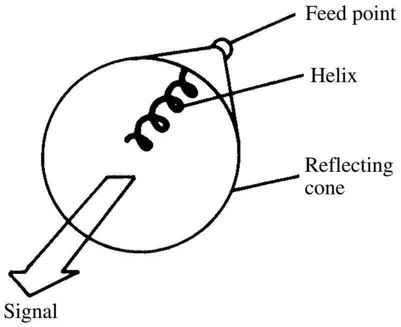

helicone

helionics The science of converting solar heat into electrical energy. The term is an acronym from *helio-* (pertaining to the sun) and *electronics.*

heliostat **1.** A servo-controlled motor-driven device that drives a mirror to keep sunlight trained upon a specific target. **2.** By extension, any similar device to keep a solar cell pointed to the sun.

helitron A form of oscillator used at ultra-high and microwave frequencies. The output frequency is variable over a wide range.

helium Symbol, He. A gaseous element. Atomic number, **2**. Atomic weight, 4.0026.

helium group The six inert gases in group 0 of the periodic table: *argon*, *helium*, *krypton*, *neon*, *xenon*, and *radon.*

helium-neon laser A laser in which the lasing substance is a mixture of helium and neon. Produces a characteristic brilliant red visible output. Also see HELIUM and NEON.

helix **1.** A single-layer coil. **2.** That which is coil-shaped (i.e., spiral in configuration). **3.** See HELICAL ANTENNA. **4.** See HELICAL LINE.

helix line See HELICAL LINE.

helix recorder An information recorder using a spiral method of scanning. The recording medium is usually drum-shaped.

Helmholtz coil A device consisting of two crossed-field primary windings in which an inductively coupled secondary winding rotates. The primary windings carry currents that differ in phase by 90 degrees. Rotating the secondary coil provides 360 degrees of continuously variable phase shift.

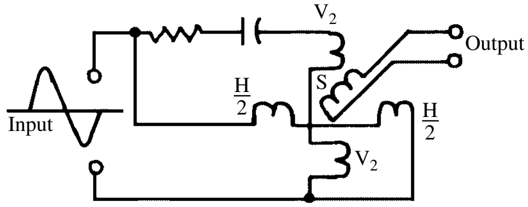

Helmholtz coil

Helmholtz double layer An intermolecular layer between a metal and an electrolyte in which it is immersed. It is formed when the adhesive force between the metal and electrolyte decreases the surface tension of the metal, causing positive ions to migrate from the metal into the liquid. The metal, charged negatively, and the electrolyte, charged positively, form a capacitor whose dielectric is the Helmholtz layer.

Helmholtz resonator An acoustic (sound) chamber whose geometry, in combination with the size of a small opening, results in resonance at a specific frequency.

HEM Abbreviation of *hybrid electromagnetic* (see, for example, HYBRID ELECTROMAGNETIC WAVE).

hemimorphic Pertaining to an object with ends that have unlike faces.

He-Ne laser See HELIUM-NEON LASER.

henry Symbol, H. The standard unit of inductance. It is the inductance exhibited by a closed circuit in which one volt is produced by a current changing uniformly at one ampere per second. This is a large unit of inductance; more common units are the MILLIHENRY and the MICROHENRY.

hermaphroditic plug A plug that has the prongs of a male plug and the recessed contacts of a female plug. Compare FEMALE PLUG and MALE PLUG.

hermetically sealed Constructed in manufacture so as to be permanently closed against the entry of air or other gases, dust, and moisture.

hermetic seal A permanent, air-tight seal that effectively prevents corrosion from elements in the outside environment.

herringbone pattern A pattern of interference in a television picture, so named because of its resemblance to the skeleton of a fish.

Herschel-Quincke tube An acoustic device that demonstrates sound interference. The device has two hollow cylinders, one of which can be adjusted in length. At the far end of the apparatus, the cylinders are joined together. The resultant amplitude depends on the difference in length between the cylinders. Sound wavelengths can be measured using this apparatus.

hertz Abbreviation, Hz. The standard unit of frequency (of periodic phenomena, such as alternating or pulsating currents); 1 Hz = 1 cycle per second. One Hz is an extremely small unit of frequency; more common units are the KILOHERTZ, the MEGAHERTZ, and the GIGAHERTZ.

Hertz antenna An ungrounded halfwave antenna fed by a transmission line attached to one end or to the center of the radiator. See, for example, CENTER-FED ANTENNA and END-FED ANTENNA. Compare MARCONI ANTENNA.

Hertz effect Ionization of a gas, produced by intense ultraviolet radiation.

Hertzian antenna See HERTZ ANTENNA.

Hertzian oscillator See HERTZ OSCILLATOR.

Hertzian radiation Radiation of electromagnetic (radio) waves.

Hertzian waves Electromagnetic waves in the radio spectrum, with wavelengths longer than those of infrared, visible light, ultraviolet, X rays, or gamma rays.

Hertz oscillator A damped-wave generator of oscillations, used by Hertz in his demonstration of radio waves in 1888 (verifying the earlier prediction by James Clerk Maxwell). The oscillator contains a spark gap supplied by an induction coil, attendant coils, capacitors (in the prototype, Leyden jars), and two large metal plates.

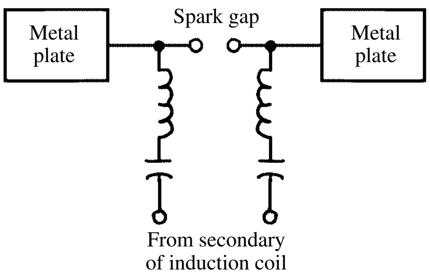

Hertz oscillator

Hertz vector A single vector that specifies the electromagnetic field (electric and magnetic components) of a radio wave.

hesitation As distinct from a halt, a brief break in a computer program run during which internal operations are occurring, such as data transfer to a peripheral.

heterochromatic Consisting of different frequencies, wavelengths, or colors. Compare MONOCHROMATIC.

heterodyne **1.** To beat one alternating-current signal against another to produce one or more beat-frequency signals. Also see BEAT FREQUENCY and BEAT NOTE. **2.** The whistle produced when two signals very close in frequency are mixed in a radio receiver. **3.** To combine radio signals in a mixer, the purpose of which is to produce a sum or difference signal for further processing.

heterodyne detection **1.** Signal detection by beating the incoming signal against one produced by a local oscillator. In this way, an unmodulated signal is made audible (the beat note is an audio frequency). **2.** Signal detection by a superheterodyne circuit.

heterodyne detector **1.** A detector that makes a radio-frequency (RF) signal audible by beating it against the RF signal of a local oscillator, the product being an audio-frequency (AF) beat note. **2.** The FIRST DETECTOR or SECOND DETECTOR in a SUPERHETERODYNE CIRCUIT. **3.** A combination linear detector and local RF oscillator used to detect and measure the frequency of test signals. Also see HETERODYNE FREQUENCY METER.

heterodyne eliminator See WHISTLE FILTER.

heterodyne filter See WHISTLE FILTER.

heterodyne frequency The frequency of the signal obtained by beating one signal against another.

heterodyne frequency meter A frequency-measuring device that contains a variable-frequency oscillator (VFO), a mixer, and an indicator such as an analog meter. The oscillator frequency is adjusted until zero beat is reached with the signal source. This condition is shown by a dip in the meter indication. An audio amplifier can be coupled to the output of the device instead of a meter; in this case the heterodyne appears as an audible tone whose frequency drops to zero when the oscillator frequency is equal to the signal frequency being measured.

heterodyne method See HETERODYNE, **1**.

heterodyne oscillator A signal generator whose output is the beat product of outputs from two internal oscillators. The output frequency can be either the sum or the difference of the oscillator frequencies, as selected by output filtering or tuning. See BEAT-FREQUENCY OSCILLATOR.

heterodyne reception Radio reception (especially in telegraphy) by means of the beat-note process. Also see HETERODYNE DETECTION, **1** and HETERODYNE DETECTOR, **1**.

heterodyne repeater A REPEATER in which the received signals are converted to another frequency before transmission.

heterodyne-type frequency meter See HETERODYNE FREQUENCY METER.

heterodyne wave analyzer A type of audio-frequency (AF) wave analyzer. The input signal is heterodyned in a balanced modulator with the signal from an internal tunable oscillator. One of the resulting sidebands is passed through a sharp bandpass filter, whose output actuates an alternating-current (ac) voltmeter. The internal oscillator is tuned slowly so that different components of the balanced-modulator output sideband pass successively through the filter. The amplitude-versus-frequency function of the input signal is determined by noting the meter readings as the internal oscillator is tuned.

heterodyne wavemeter See HETERODYNE FREQUENCY METER.

heterodyne whistle See HETERODYNE, **2**.

heterogeneous Pertaining to a group of objects or devices that have differing characteristics.

heterogeneous radiation Any broadband form of radiation. In particular, broadbanded radio waves, infrared radiation, visible light, ultraviolet, X rays, or gamma rays.

heterolysis The HYDROLYSIS of a compound into two oppositely charged ions.

heteropolar generator An electric generator in which the active conductors move through magnetic fields, first in one direction, then in the other direction. This is done by means of rotation in a nonuniform magnetic field. Most generators in common use are of this kind.

heuristic knowledge In artificial intelligence (AI), a form of machine knowledge in which a computer learns from its mistakes. As a complex program is repeatedly run over a period of time, the number of errors per run approaches zero.

heuristic program In artificial intelligence (AI), a program with which the computer solves a problem by trial and error, often learning in the process so that mistakes will not be repeated on subsequent runs.

Heusler's alloys Ferromagnetic alloys that contain one or more non-magnetic metals (such as aluminum, copper, or manganese).

hexadecimal number system An alphanumeric, base-16 system of number notation used in some computers. The system uses the usual digits plus the letters A through F to represent the numbers 10 through 15 (each place can only hold one symbol).

hex inverter A collection of six digital inverters, or NOT gates, contained within one package, usually an integrated circuit.

Hf Symbol for HAFNIUM.

HF Abbreviation of HIGH FREQUENCY.

Hg Symbol for MERCURY.

HH beacon In radionavigation, a nondirectional homing beacon.

hi **1.** Contraction of HIGH. **2.** Radiotelegraph symbol for a laugh, often verbalized by radio amateurs.

HIC Abbreviation of HYBRID INTEGRATED CIRCUIT.

HIDM Abbreviation of HIGH-INFORMATION DELTA MODULATION.

hierarchical password protection Also called *multilevel password protection.* A security feature that prevents unauthorized use of a computer, network, or database. The password levels allow users various degrees of control over the host machine.

hi-fi **1.** Contraction of HIGH-FIDELITY. **2.** In video recording, the addition of sound having high fidelity.

high **1.** Pertaining to a circuit point or condition at some potential above ground. **2.** The logical digit 1. **3.** The condition of having relatively large magnitude (e.g., HIGH FREQUENCY and HIGH VOLTAGE). **4.** Pertaining to the upper portion of a range, as in HIGH BAND or HIGH FREQUENCY. **5.** Characterized by greater-than-normal response or performance, as in HIGH *Q* or HIGH FIDELITY.

high band **1.** The higher or highest frequency band used in communications, testing, or processing, when several bands are available. **2.** The very-high-frequency (VHF) television channels 7 through 13. **3.** The communications frequency range from about 144 MHz to about 170 MHz.

high boost In sound recording and reproduction, the emphasis of high frequencies in an operating spectrum. Also called HIGH-FREQUENCY COMPENSATION.

high-C circuit A tuned circuit having high capacitance and low inductance at a given frequency. Such a circuit is characterized by high selectivity and low voltage. Compare HIGH-L CIRCUIT. Also see LC RATIO.

high contrast In an image, a limited range of gray values between black and white, or a similar condition in a color image (overbright whites, little shadow detail). Also see CONTRAST.

high definition In facsimile or television, a condition of minute detail so that the original scene is faithfully reproduced.

high-definition television Abbreviation, HDTV. A method of getting enhanced detail into a television (TV) picture and for obtaining better audio quality, compared with standard analog TV. There are several different schemes. The most noticeable feature is the crispness of the picture. This is vividly apparent in big-screen installations, which have traditionally suffered from image blurring. A standard TV picture has 525 lines per frame, but HDTV systems can have more than twice that number. Another important difference is that HDTV is digital; this minimizes susceptibility to interference. Interlacing is used in some systems.

high-density bipolar-3 code A communications or digital code in which two logic highs (ones) can occur in sequence, without the need for an intervening logic low (zero) to separate them.

high-efficiency linear amplifier A LINEAR AMPLIFIER with higher operating efficiency than is obtainable with conventional class-B linear amplifiers. Efficiencies on the order of 60% at 100% modulation are possible.

high-energy materials See HARD MAGNETIC MATERIALS.

high-energy particle **1.** A SUBATOMIC PARTICLE that has been given high velocity by a particle accelerator. **2.** High-speed subatomic particles emitted by the sun during a solar flare, or arriving from outer space.

high-energy physics The discipline dealing with the characteristics, properties, and applications of HIGH-ENERGY PARTICLES.

higher-order language See HIGH-LEVEL LANGUAGE.

high fidelity Abbreviation, hi-fi. Pertaining to an audio-frequency system that is very faithful to the signal it is processing (i.e., one characterized by extremely low distortion and wide frequency response).

high frequency Abbreviation, HF. Pertaining to frequencies in the 3- to 30-MHz band (wavelengths from 10 to 100 meters). Also see RADIO SPECTRUM.

high-frequency alternator A dynamo for generating radio-frequency energy.

high-frequency bias In a tape recorder, a high-frequency sinusoidal signal superimposed on the signal being recorded, for improving linearity and dynamic range.

high-frequency compensation See HIGH BOOST.

high-frequency converter See SHORTWAVE CONVERTER.

high-frequency crystal See HARMONIC CRYSTAL.

high-frequency direction finder Abbreviation, HDF. A direction finder operated at high radio frequencies (i.e., between about 3 MHz and 30 MHz).

high-frequency heating Electronic heating of materials by high-frequency energy. See, for example, DIELECTRIC HEATING and INDUCTION HEATING.

high-frequency resistance See RADIO-FREQUENCY RESISTANCE.

high-frequency speaker See TWEETER.

high-frequency trimmer **1.** In older high-frequency communications receivers, a low-value variable capacitor operated in parallel with a usually front-panel tuning capacitor to set the high-frequency end of the tuning range. See, for example, OSCILLATOR TRIMMER. **2.** A small variable capacitor used in conjunction with a larger tuning capacitor, the function of which is to permit precision tuning of the larger device.

high-impedance-state output current Pertaining to tests that ensure that an integrated circuit will not overload a bus line.

high-impedance voltmeter A voltmeter having an input impedance of at least several megohms.

high-information delta modulation A companded form of delta modulation, operating at comparatively low sample rate.

high-L circuit A tuned circuit having high inductance and low capacitance at a given frequency. Such a circuit is characterized by low selectivity and high voltage. Compare HIGH-C CIRCUIT. Also see LC RATIO.

high-level audio signal An audio-frequency signal that has been preamplified (e.g., the output of a compact-disc player). Compare LOW-LEVEL AUDIO SIGNAL.

high-level input current **1.** Pertaining to the testing of intertransistor leakage in an integrated circuit (IC) having multiple emitter inputs. **2.** The current into an IC input at minimum high-level voltage.

high-level language Also called *higher-order language.* A computer programming language in which the operator is easily able to communicate with the machine. It generally serves as an interface between a human programmer and the MACHINE LANGUAGE. Examples are BASIC, C, C++, COBOL, and FORTRAN.

high-level modulation In an amplitude-modulated transmitter, introduction of the audio at the final stage of radio-frequency amplification, permitting 100% modulation of the full-power signal.

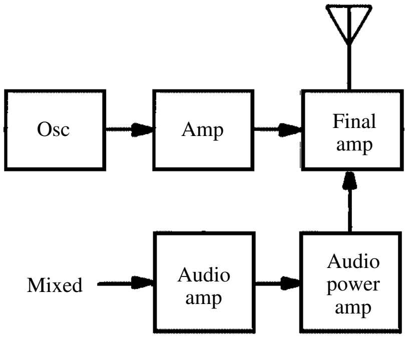

high-level modulation

high-level output current **1.** Pertaining to the testing of drive capability and fanout of an integrated circuit (IC). **2.** The current flowing from an IC output when in the high state.

high-level recovery Hardware recovery using data not involved in the failure, such as that on a magnetic storage medium. Also see HARDWARE RECOVERY.

highlight **1.** A bright area in a television picture. **2.** In computer data processing, the defining or setting-off of a block of data (such as text), with the intention of relocating, editing, or deleting it.

high-noise-immunity logic Abbreviation, HNIL. A form of bipolar digital logic designed for minimal sensitivity to noise. Also known as *high-threshold logic (HTL).*

high order Descriptive of the relationship between bits or digits in a word or number. Of two digits, the one holding the higher place value is the high-order digit (e.g., 2 is the high-order digit in 25).

high-pass filter A combination of capacitance, inductance, and/or resistance, intended to produce large amounts of attenuation below a certain frequency and little or no attenuation above that frequency. The frequency at which the transition occurs is called *cutoff*. At cutoff, the power attenuation is 3 dB with respect to the minimum attenuation. At frequencies above cutoff, the power attenuation is less than 3 dB. At frequencies below cutoff, the power attenuation is more than 3 dB. The simplest circuit consists of a parallel inductor or a series capacitor. The inductance-capacitance (LC) circuit has a combination of parallel inductors and series capacitors. In the resistance-capacitance (RC) circuit, resistors are substituted for the inductors. Compare BAND-PASS FILTER, BAND-REJECTION FILTER, LOW-PASS FILTER.

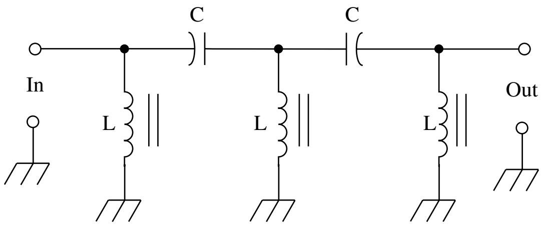

high-pass filter

high-pass-filter method A method of measuring the total harmonic distortion (THD) percentage using a high-pass filter to separate the harmonics from the fundamental. The output voltage V_o of the filter is compared with the input voltage V_i; then THD = $100V_o/V_i$.

high-performance Pertaining to apparatus designed for continuous operation with maximum reliability.

high-performance navigation system Acronym HIPERNAS. An electromechanical guidance system that is purely inertial and self-compensating.

high-potential test A high-voltage test of insulation, in which the applied voltage is continuously increased until the breakdown point of the dielectric is reached.

high-power rectifier A rectifier designed for high-voltage, high-current operation.

high Q For a component or circuit, a high value for the ratio X/R (reactance to resistance). This is a relative term because a particular numerical value of Q, considered high in one situation, might be regarded as low under other circumstances. Also see FIGURE OF MERIT.

high-resistance joint In the wiring of electronic equipment, a joint or connection between conductors that is poorly made, thereby introducing a high resistance between the parts.

high-resistance voltmeter A voltmeter having an input resistance of at least several megohms.

high-speed carry In computer operation, a carry into a column causing a carry out, circumventing the usual intermediate adding circuit.

high-speed diode See COMPUTER DIODE.

high-speed flip-flop A flip-flop having short switch-on and switch-off time.

high-speed oscilloscope An oscilloscope with excellent high-frequency and unit-function response. It can reproduce high-speed pulses faithfully.

high-speed relay A relay with a short make or short break interval.

high-speed transistor See SWITCHING TRANSISTOR.

high tension Pertaining to utility power-transmission lines on which there are very high voltages, typically 100 kilovolts (100 kV) or more.

high-tension line A power-transmission line carrying a very high voltage. It is generally used for the transfer of electric power over long distances.

high-threshold logic See HIGH-NOISE-IMMUNITY LOGIC.

high voltage **1.** A voltage considerably higher than those ordinarily encountered in a particular application. The term is comparative; a few hundred volts might be considered high in one situation, but low in another. **2.** In a cathode-ray tube, the voltage that accelerates the beam electrons. **3.** In a television receiver, the picture-tube anode voltage. **4.** In a vacuum-tube power amplifier, the plate supply voltage.

high-voltage probe A very-high-resistance probe for measuring high voltages with a low-range voltmeter.

highway A path over which multiple signals are propagated using time-division multiplexing.

HILAC Acronym for *heavy-ion linear accelerator.*

hill-and-dale recording See VERTICAL RECORDING.

hinged-iron instrument An alternating-current meter whose input transformer core is hinged in two parts. By means of a thumb trigger, the core can be opened, then closed around the current-carrying conductor that induces magnetism in the core; a secondary coil delivers current to the meter. Also called *clamp ammeter* or *clamp voltmeter.*

HIPERNAS Acronym for HIGH-PERFORMANCE NAVIGATION SYSTEM.

hipernick A high-permeability alloy of iron and nickel.

hipot Contraction of *high potential.* See HIGH-POTENTIAL TEST.

hiss **1.** A high-pitched sound rich in sibilants (*s*, *sh*, and *z* sounds) produced by random high-frequency fluctuations in current. **2.** The characteristic, high-pitched background noise (as in 1) accompanying super-regeneration. **3.** Internally generated noise in a communications receiver, amplified by the audio-frequency stages and appearing at the speaker or headphones.

hiss filter See HASH FILTER.

hit **1.** The occurrence of a lightning stroke at a specific point on the ground. Also called *direct hit.* **2.** The coincidence of two pulses.

H lines Magnetic lines of flux.

HLL Abbreviation of HIGH-LEVEL LANGUAGE.

H network A network of five impedances: two connected in series between the upper input and output terminals, two between the lower input and output terminals, and one shunted between the junctions of the series-connected impedances. Also called *H pad*, *balanced tee network*, and *balanced tee pad.*

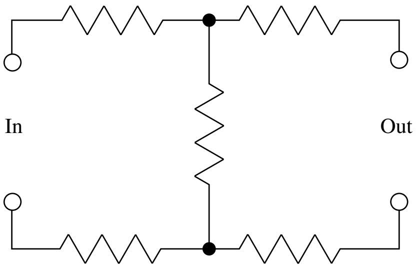

H network

Ho Symbol for HOLMIUM.

hobby robot A robot intended for amusement and, sometimes, for education. Such robots often take humanoid form. Some are programmable, and can give demonstrations, play musical instruments, and do other complex routines.

hockey-stick lead On a capacitor, resistor, or other component, a pigtail lead that is given a single crimp for easy insertion into a printed-circuit board.

hodoscope An instrument consisting essentially of closely spaced ion counters, for studying the path of an ionizing particle.

Hoffmann electrometer See BINANT ELECTROMETER.

hog horn A form of horn antenna used in microwave applications. It is generally used in the feed system of a dish antenna. The horn opening points in the direction of the feed waveguide.

hold **1.** To retain data in a storage device after the data has been duplicated in another location or device. **2.** A momentary halt of an operation or process. **3.** In a television receiver, a control that stabilizes the vertical or horizontal synchronization.

hold circuit **1.** See HOLDING CIRCUIT. **2.** In a television receiver, the circuit associated with the hold control(s). Also see HORIZONTAL-HOLD CONTROL and VERTICAL-HOLD CONTROL.

hold control See HORIZONTAL-HOLD CONTROL and VERTICAL-HOLD CONTROL.

hold current Symbol, I_h. The minimum current that will keep a normally open relay closed or a normally closed relay open.

hold electrode In a mercury switch, the electrode that is in permanent contact with the mercury.

holding beam In an electrostatic cathode-ray storage tube, the electron beam that generates replacement charges for those that were stored on the dielectric surface and then lost.

holding circuit In an electromechanical relay, a separate circuit that, when energized, keeps the relay actuated.

holding coil In an electromechanical relay, the extra coil that is associated with the holding circuit.

holding current **1.** Current in the holding coil of a relay. **2.** In a gas tube, the minimum current required to maintain ionization.

holding gun In an electrostatic cathode-ray storage tube, the electron gun that generates the holding beam.

hold mode A condition in which the output state of a digital-logic circuit remains unchanged while the input signals are removed.

hold-off voltage The highest voltage that can be applied to a flashtube without causing it to fire.

holdover The flow of current through the ionized path created by an electric arc.

hold time **1.** The time permitted for a weld to harden in resistance welding. **2.** In digital communications, the time for which a signal is maintained at a certain input after changing state at another specified input.

hole **1.** In a semiconductor atom, the vacancy resulting from the loss of an electron. When an electron is lost, so is its negative charge, leaving an equivalent net positive charge. This charge, like that of an electron, can move as a current carrier. **2.** The punched-out portion of a chassis or panel, through which wires can be passed or components mounted.

hole conduction In a semiconductor material, electrical conduction as a result of HOLE CURRENT.

hole current In a semiconductor material, the electrical current resulting from the movement of positive charge carriers (holes). Also see HOLE.

hole density The degree of concentration of holes in a semiconductor. Also see HOLE.

hole-electron pair In a semiconductor, a hole and a related electron. Each electron in the conduction band has a counterpart in the valence band, a vacancy (HOLE) left by the movement of the electron to the conduction band.

hole injection The creation of mobile holes in a semiconductor by applying an electric charge. Also see HOLE.

hole injector **1.** The emitter electrode of a bipolar pnp transistor. **2.** The metal whisker of a point-contact diode having an n-type wafer. **3.** The player of a forward-biased junction diode.

hole mobility The ease with which a hole moves within a semiconductor. Also see CARRIER MOBILITY.

hole storage See CARRIER STORAGE.

hole storage factor In a bipolar transistor biased to saturation, the amount of storage charge caused by excess base current.

hole trap In a semiconductor, an impurity that can cancel holes by releasing electrons to fill them.

hollow coil A coreless inductor.

hollow conductor Tubing used as a low-loss conductor at radio frequencies.

hollow core A core that is not solid throughout—especially one that has a central mounting hole.

holmium Symbol, Ho. A metallic element of the rare-earth group. Atomic number, 67. Atomic weight, 164.93. Forms highly magnetic compounds.

holocamera A camera for making holograms. Also see HOLOGRAM, **1** and HOLOGRAPHY.

hologram **1.** A wavefront recording made on photographic film by the process of HOLOGRAPHY. By changing the frequency of the light transmitted, various magnifications of the image can be obtained. Produces a true three-dimensional image. **2.** A visible, three-dimensional display projected in the air or underwater by means of lasers. They are often used at outdoor music concerts and other events.

holography A method of producing a wavefront recording of an object illuminated by laser light. The result, an interference pattern, appears meaningless when viewed in ordinary diffuse light. But when a point source of illumination is used, especially a laser, an image appears that is convincingly three-dimensional.

homeostasis The condition of being in static equilibrium.

home station See BASE STATION.

homing **1.** Guidance by means of an electronic beacon. The vehicle maintains a course toward the beacon. **2.** Guidance by means of some form of emission from a target object. The emission can be acoustic or electromagnetic energy.

homing antenna A direction-finding antenna—especially one on a mobile vehicle.

homing beacon A station radiating a beam for use in direction finding by mobile vehicles.

homing device A receiving device mounted on a mobile vehicle, and that continuously indicates the direction of a selected transmitting station that is the vehicle's destination.

homing relay A stepping relay that returns to its starting position after each switching sequence. Also see STEPPING SWITCH.

homing station See HOMING BEACON.

homodyne reception See ZERO-BEAT RECEPTION.

homogeneous **1.** Uniform in structure; similar at all points or locations. **2.** Consisting of many identical elements.

homogous field A field whose lines of flux in one plane pass through a single point.

homolysis The decomposition of a compound into a pair of neutral atoms or radicals.

homomorphism A one-to-one correspondence between the elements of two sets.

homopolar Pertaining to the union of atoms of the same polarity; nonionic.

homopolar generator A direct-current (dc) generator whose poles have the same polarity, with respect to the armature. Thus, no commutator is necessary.

homopolar magnet A magnet whose pole pieces are concentric.

homunculus In artificial intelligence (AI), a computer or robot that exhibits characteristics of a living being; especially, an ANDROID.

honeycomb coil A multilayer coil having a UNIVERSAL WINDING.

honeycomb winding See UNIVERSAL WINDING.

honker Also called *midrange speaker.* A loudspeaker that favors the middle audio frequencies. Compare TWEETER and WOOFER.

hood A light shield for a cathode-ray tube; it allows the screen to be viewed with a minimum of interference from room light.

Hooke's law Strain is proportional to the stress that produces it, as long as the ELASTIC LIMIT is not exceeded.

hook switch A switch that closes a circuit when a headset or handset is lifted from the resting position. The common telephone receiver uses such a switch.

hook transistor A four-layer pnpn semiconductor device, in which the outer p and n layers serve as emitter and collector, the inner n layer being the base. This places a p layer between the base and collector, resulting in a transistor that provides high alpha as a result of carrier multiplication by the additional junction in the collector layer.

hookup See SCHEMATIC DIAGRAM.

hookup wire Flexible, insulated wire used in the wiring of some electrical and electronic devices.

hoop antenna See CAGE ANTENNA.

hoot stop During a computer program run, a loop made evident by a sound signal.

hop In long-distance radio communications, the transmission of a wave and its subsequent return to the earth from the ionosphere; it is of importance mainly at low, medium, and high frequencies.

hor Abbreviation of HORIZONTAL. (Also, *H* and *horiz.*)

horiz Occasional abbreviation of HORIZONTAL. The usual form is *hor*; another alternate is *H.*

horizon **1.** For a specific location, the circle on the celestial sphere midway between the zenith (the point directly overhead) and the nadir (the point directly underfoot). **2.** Also called *visual horizon.* The set of points, as viewed from a particular location, where the sky and the earth appear to meet (i.e., the last visible part of the earth's surface from a given observation point). **3.** Also called *radio horizon.* The set of points, from a particular location at a given radio frequency, representing the maximum communications range via the ground wave under normal conditions. Also see ARTIFICIAL HORIZON.

horizontal **1.** Pertaining to objects or effects in a plane perpendicular to a line connecting the zenith (the point directly overhead) and the nadir (the point directly underfoot). **2.** Pertaining to that which is parallel to an assumed flat surface. **3.** Pertaining to width deflection on a cathode-ray tube.

horizontal AFC In a television receiver circuit, *automatic frequency control (AFC)* of the horizontal sweep. It keeps the receiver's horizontal scanning in step with that of the camera at the transmitting station.

horizontal amplification Gain provided by the horizontal channel of a device, such as an oscilloscope, cathode-ray electrocardiograph, or television receiver. Compare VERTICAL AMPLIFICATION.

horizontal amplifier A circuit or device that provides HORIZONTAL AMPLIFICATION. Compare VERTICAL AMPLIFIER.

horizontal angle of radiation For an antenna, the direction of maximum radiation in the horizontal plane (see HORIZONTAL, **1**), provided as an azimuth angle measured clockwise from geographic north.

horizontal angle of deviation In a communications circuit, the angular difference, in degrees, between the compass direction from which a received signal arrives, and the great-circle path connecting the receiving station with the transmitting station.

horizontal axis The axis that is parallel to an assumed horizontal surface (of the earth, for example) or the one so represented in a diagram. Also see X-AXIS.

horizontal beamwidth In a directional antenna system, the angle, measured in the horizontal plane, between the half-power points in the major lobe.

horizontal blanking See HORIZONTAL RETRACE BLANKING.

horizontal-blanking pulse In a television signal, the rectangular pedestal-shaped pulse that occurs between the active horizontal lines to achieve horizontal retrace blanking. Compare VERTICAL-BLANKING PULSE.

horizontal centering control See CENTERING CONTROL.

horizontal channel The system of amplifiers, controls, and terminations that constitute the path of the horizontal signal in an equipment, such as an oscilloscope or graphic recorder. Compare VERTICAL CHANNEL.

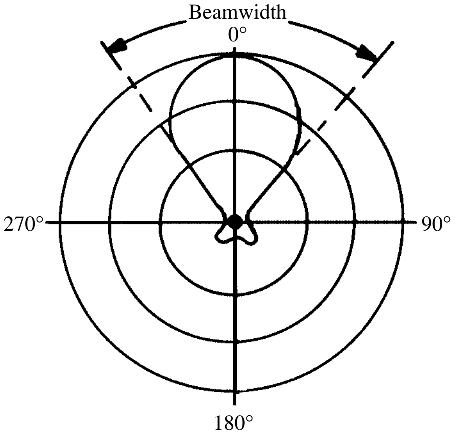

horizontal beamwidth

horizontal-convergence control In a color television receiver, the variable component for adjusting the horizontal dynamic convergence voltage.

horizontal coordinates See CARTESIAN COORDINATES.

horizontal deflection In a cathode-ray tube, the lateral movement of the electron beam on the screen. Compare VERTICAL DEFLECTION.

horizontal deflection coils The pair of coils that produces the magnetic field to horizontally deflect the electron beam in an electromagnetic cathode-ray tube. Also see DEFLECTION COIL.

horizontal-deflection electrodes See HORIZONTAL DEFLECTION COILS and HORIZONTAL DEFLECTION PLATES.

horizontal deflection plates In an electrostatic cathode-ray tube (typical of laboratory oscilloscopes and some early television picture tubes), a pair of plates that produces an electric field to horizontally deflect the electron beam. Compare VERTICAL DEFLECTION PLATES.

horizontal directivity The radiation or reception pattern of a directional antenna in the horizontal plane.

horizontal-drive control See DRIVE CONTROL.

horizontal dynamic convergence During the scanning of a horizontal line in a color picture tube, convergence of the electron beams at the aperture mask. Compare VERTICAL DYNAMIC CONVERGENCE.

horizontal field strength The field strength of signals passing through an antenna in a horizontal plane. Compare VERTICAL FIELD STRENGTH.

horizontal-field-strength diagram A plot of horizontal field strength, usually in polar form. Compare VERTICAL-FIELD-STRENGTH DIAGRAM.

horizontal flowcharting Flowcharting the movement of documents or files, rather than the data bits themselves, through a digital system.

horizontal frequency In television circuits, the horizontal scanning frequency [i.e., the frequency at which the horizontal lines are traced (generally 15.750 kHz)].

horizontal frequency response The gain-vs-frequency characteristic of the horizontal channel of an oscilloscope or graphic recorder. Compare VERTICAL FREQUENCY RESPONSE.

horizontal gain At a specified frequency, the overall amplification of the horizontal channel of an oscilloscope or graphic recorder. Compare VERTICAL GAIN.

horizontal-gain control A control, such as a potentiometer, for adjusting horizontal gain. Compare VERTICAL-GAIN CONTROL.

horizontal-hold control In a television receiver, the control for adjusting the horizontal oscillator frequency to prevent horizontal tearing of the picture. Compare VERTICAL-HOLD CONTROL.

horizontal hum bars Dark, horizontal interferential bars in a television picture, caused by HUM interference.

horizontal linearity The precision of gain and deflection in the horizontal channel of an oscilloscope, graphic recorder, or television receiver. A linear picture is a faithful (undistorted) reproduction of the original image. Compare VERTICAL LINEARITY.

horizontal-linearity control In an oscilloscope or television receiver, the control with which horizontal linearity is adjusted. Compare VERTICAL-LINEARITY CONTROL.

horizontal line frequency See HORIZONTAL FREQUENCY.

horizontal lock See HORIZONTAL-HOLD CONTROL.

horizontally polarized wave An electromagnetic wave whose electric lines of flux are parallel to the plane of the horizon. Compare VERTICALLY POLARIZED WAVE.

horizontal multivibrator In a television receiver, a 15.750-kHz multivibrator that generates the horizontal sweep signal.

horizontal oscillator In a TV receiver, the oscillator (usually a multivibrator) that generates the horizontal sweep signal. Compare VERTICAL OSCILLATOR.

horizontal output stage In a television receiver, an output amplifier following the horizontal oscillator. Compare VERTICAL OUTPUT STAGE.

horizontal output transformer In a television receiver, the output transformer in the horizontal-oscillator-output amplifier section. Also called FLYBACK TRANSFORMER.

horizontal polarization Pertaining to an electromagnetic wave whose electric lines of flux are horizontal. In general, when the radiating element of an antenna is horizontal, the electric lines of flux in the transmitted waves are horizontal, and the antenna is most sensitive to incoming signals whose electric lines of flux are horizontal. Compare VERTICAL POLARIZATION.

horizontal positioning control See CENTERING CONTROL.

horizontal quantity The quantity measured along the X-axis of a graph represented by the horizontal deflection of an oscilloscope beam. Compare VERTICAL QUANTITY.

horizontal recording See LATERAL RECORDING.

horizontal repetition rate See HORIZONTAL FREQUENCY.

horizontal resolution In a television picture, the number of picture elements (*pixels*) that can be discerned in a horizontal scanning line. Compare VERTICAL RESOLUTION.

horizontal retrace In a cathode-ray device, such as an oscilloscope or television receiver, the rapid return of the electron beam to its starting point after completing a horizontal sweep of the screen. Compare VERTICAL RETRACE.

horizontal retrace blanking In oscilloscopes and television receivers, the automatic cutoff of the electron beam during a horizontal retrace period, preventing an extraneous line on the screen during the period. Compare VERTICAL RETRACE BLANKING.

horizontal scanning **1.** The lateral sweeping of the electron beam in a cathode-ray tube. **2.** The sampling of x-axis values in a repetitive or nonrepetitive sweep of that axis.

horizontal scanning frequency See HORIZONTAL FREQUENCY.

horizontal sensitivity The signal voltage required at the input of a horizontal channel for full horizontal deflection. Also see HORIZONTAL GAIN. Compare VERTICAL SENSITIVITY.

horizontal signal A signal serving as a horizontal quantity. Compare VERTICAL SIGNAL.

horizontal sweep **1.** In a cathode-ray tube, the horizontal movement of the spot on the screen; in particular, the movement from left to right, during which a line of the image is formed on the screen. **2.** The circuit that produces horizontal sweep.

horizontal sweep frequency The frequency at which horizontal sweep occurs; in a television receiver, it is generally 15.750 kHz. Also called *horizontal sweep rate.* Compare VERTICAL SWEEP FREQUENCY.

horizontal sweep rate See HORIZONTAL SWEEP FREQUENCY.

horizontal-sync discriminator In a television receiver, a circuit that compares horizontal sync-pulse phase with the phase of the signal from the horizontal sweep oscillator.

horizontal synchronization In a television receiver, synchronization of the horizontal component of scanning with that of the transmitting camera. Also see HORIZONTAL SYNC PULSE. Compare VERTICAL SYNCHRONIZATION.

horizontal sync pulse In a video signal, the pulse that synchronizes the horizontal scanning component in a television receiver with that of the camera; it also triggers horizontal retrace and blanking. Also see BACK PORCH. Compare VERTICAL SYNC PULSE.

horizontal wave See HORIZONTALLY POLARIZED WAVE.

horizontal-width control See WIDTH CONTROL, **1**, **2**.

horn A radiating device that is essentially a cylindrical or rectangular pipe, whose surface flares from a narrow entry to a wide exit. See, for example, HORN ANTENNA, HORN SPEAKER, and MEGAPHONE.

horn antenna An antenna used for transmission and reception of signals at microwave frequencies. There are several configurations; most have a square or rectangular cross section. The antenna provides a unidirectional radiation and response pattern, with the favored direction coincident with the opening of the horn. The feed line consists of a waveguide that joins the antenna at the narrowest point (throat) of the horn. This type of device is used alone in microwave transmitters known as *Gunnplexers*. The device works well in the feed systems of large dish antennas; the horn is pointed toward the center of the dish, in the opposite direction from the favored direction of the dish. See also DISH ANTENNA, GUNNPLEXER.

horn cutoff frequency The lowest frequency at which an exponential horn will function properly.

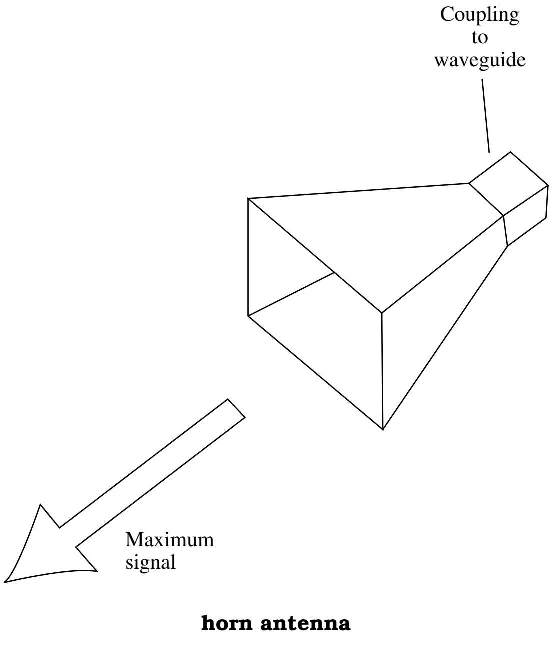

horn antenna

horn loading In a sound-transmission system, a form of propagation that makes use of a horn-shaped speaker.

horn mouth The wider (radiating) end of a horn antenna or speaker. Compare HORN THROAT.

horn radiator See HORN ANTENNA.

horn speaker A loudspeaker using a horn in conjunction with an acoustic transducer. It generally produces sound of low fidelity, and is intended for use in marginal environments, such as outdoor public-address (PA) systems.

horn throat The narrower (input) end of a horn antenna or speaker. Compare HORN MOUTH.

horsepower Abbreviation, hp. A unit of power equal to 746 watts. It is generally used to express mechanical power delivered by engines and electric motors.

horsepower-hour Abbreviation, hp-hr. A unit of energy or work represented by the equivalent of one horsepower delivered for a period of one hour; equal to 746 watt-hours.

horseshoe coil See HAIRPIN COIL and HAIRPIN COUPLING COIL.

horseshoe magnet A (usually permanent) magnet having the shape of a horseshoe or a U and a rectangular cross section.

host A programmable computer that gathers and stores the information from all of the data-entry terminals in a system.

HOT **1.** Abbreviation of HORIZONTAL OUTPUT TRANSFORMER. **2.** Abbreviation of HORIZONTAL OUTPUT TUBE.

hot **1.** Pertaining to an object from which heat energy is transferred to the environment by radiation, convection, and/or conduction. **2.** Charged to an electrical potential, either positive or negative, with respect to common ground. **3.** Pertaining to a conductor that carries a high radio-frequency (RF) current or voltage.

hot carrier In a semiconductor, a charge carrier (electron or hole) whose energy is higher than that of majority carriers normally encountered in the same material.

hot-carrier diode A semiconductor diode having a metal base that receives HOT CARRIERS from a semiconductor layer. The unit has a fast switching speed because there are virtually no minority carriers—either injected or stored.

Hotel Phonetic alphabet code word for the letter H.

hot-filament gas detector A gas detector in which the sensor is a heated filament acting as one arm of a Wheatstone bridge circuit. The bridge, previously balanced, becomes unbalanced when gas impinges upon the filament and changes its resistance.

hot junction The heated junction in a two-junction thermocouple circuit. Compare COLD JUNCTION.

hot line **1.** An energized wire, transmission line, or busbar. **2.** A private communications channel (wire or radio) kept in constant readiness for instant use between persons of high authority.

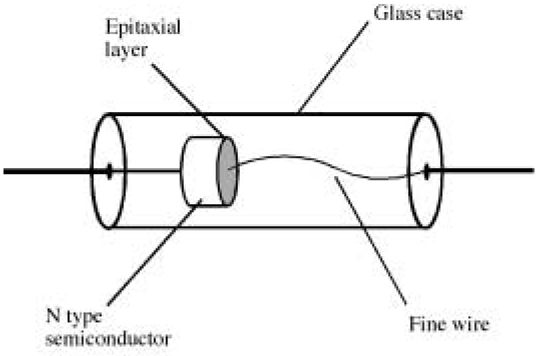

hot carrier diode

hot-pen recorder See THERMAL RECORDER.

hot plate A metal device, usually heated by means of electricity, used for the purpose of conducting certain experiments.

hot resistance The resistance of a component during its operation (i.e., after it has been heated by ambient temperature or internal power dissipation). Compare COLD RESISTANCE.

hot spark A brilliant flash seen when a capacitor discharges through a spark gap in a vacuum.

hot spot **1.** In a circuit or component, an area whose temperature is ordinarily higher than that of the surrounding area. **2.** A local current or voltage maximum on an antenna or transmission line. **3.** In communications operations, a geographic location in which reception is markedly better than in other nearby places or from which the transmitted signal appears to be stronger.

hot-strip ammeter A current meter similar to the hot-wire meter, except that it has a heated metallic strip instead of a heated wire.

hot-stylus recorder See THERMAL RECORDER.

hot-tip writing The use of a heated-tip stylus in graphic recording. Also see THERMAL RECORDER.

hot-wire ammeter See HOT-WIRE METER.

hot-wire anemometer An electrical anemometer whose indication is based on the cooling effect of the wind on a heated filament.

hot-wire flowmeter An instrument for determining the rate of flow of a gas in a pipe or other channel. The circuit is similar to that of the gas detector and hot-wire anemometer.

hot-wire gasmeter See GAS DETECTOR.

hot-wire instrument See HOT-WIRE METER.

hot-wire meter A meter in which current heats a wire, stretching it so that it moves a pointer across a scale over a distance proportional to the magnitude of the current.

hot-wire microphone A microphone in which sound waves vary the temperature of a heated wire and, accordingly, its electrical resistance.

hot-wire relay A time-delay relay in which actuating current heats a wire, causing it to expand, eventually opening or closing the contacts. Also see DELAY RELAY.

hot-wire sensor See HEATED-WIRE SENSOR.

hot-wire transducer See HOT-WIRE MICROPHONE.

hour Abbreviation, h; (sometimes, hr). A unit of time measure equal to 60 minutes or 3600 seconds. Compare MINUTE, **1** and SECOND, **1**. Also see TIME.

housekeeping In digital-computer operations, the part of a program that attends to chores (e.g., setting variables to zero), rather than being involved in making computations for a solution.

howl A discordant sound produced in headphones or a loudspeaker, usually as a result of acoustic or electrical feedback.

howler **1.** An audio-frequency alarm device. **2.** A sound-emitting test device (see GROWLER, **1**, **2**).

howl repeater A form of electric feedback in which a hum or howl occurs because of oscillation. The term is used to describe an oscillating conduction in a wire-communications-system repeater.

hp Abbreviation of HORSEPOWER.

h-p Abbreviation of HIGH PRESSURE.

H pad See H NETWORK.

***h* parameters** Parameters of the four-terminal network equivalent of a transistor. They are *hybrid parameters* (thus, *h*) because of their appearance in mesh and nodal equations. The basic *h* parameters are h_{11}, input resistance with output short-circuited; h_{12}, reverse voltage ratio with input open-circuited; h_{21}, forward current gain with output short-circuited; and h_{22}, output conductance with input open-circuited.

h particle A positive hydrogen ion or proton obtained by bombarding a hydrogen atom with alpha particles or high-velocity positive ions.

HPF Abbreviation of *highest probable frequency.*

hp-hr Abbreviation of HORSEPOWER-HOUR.

HPI Abbreviation of HEIGHT-POSITION INDICATOR.

H plane The plane of the magnetic field of an antenna, or the magnetic component of an electromagnetic field. Compare E PLANE.

H-plane bend In a waveguide, a smooth change in the direction of the axis perpendicular to the direction of polarization.

H-plane tee junction A waveguide tee junction, whose structure changes in the magnetic-field plane. Also see WAVEGUIDE JUNCTION and WAVEGUIDE TEE.

hr Abbreviation of HOUR. (Also, h.)

H scan A radar display in which the target is represented by two close blips, approximating a line, whose shape is proportional to the sine of the target's angle of elevation.

H scope See H SCAN.

HSM Abbreviation of *high-speed memory.*

HTL Abbreviation of HIGH-THRESHOLD LOGIC.

hub The hole in the center of a magnetic tape reel.

hue The quality of having a particular visible color; an attribute that depends on the wavelength(s) of light emitted or reflected. Thus, electromagnetic energy at a wavelength of 700 nanometers (nm) has a red hue; energy at a wavelength of 400 nm has a violet hue. Compare SATURATION, **2**.

hue control Also called *tint control.* In a color television receiver, a control that allows adjustment of the color wavelength, but does not affect the saturation (intensity).

hum **1.** Alternating current having a frequency of 60 Hz. **2.** Residual ripple in the output of a power supply, having a frequency of 60 Hz or 120 Hz, depending on the type of rectifier circuit used. **3.** An electromagnetic field of long wavelength, usually originating from utility lines and having a primary frequency of 60 Hz. **4.** The effects of low-frequency electromagnetic fields or currents, such as moving horizontal bars on a television screen. **5.** An acoustic disturbance of long wavelength (low pitch).

human engineering **1.** Also called *user-friendliness.* The extent to which an electronic or electromechanical device or system is easy to operate. **2.** The branch of engineering devoted to interfacing human beings with the machines and instruments they operate. Both a science and an art, the discipline is concerned with the safest and most efficient design, arrangement, and operation of equipment.

human interface The interface between a sophisticated electronic device and a human operator.

human-made interference Also called *artificial interference.* Electromagnetic interference to radio and television receiving systems or to data terminals, originating from artificial sources (such as radio transmitters, certain electrical appliances, and internal-combustion engines). Of the numerous different forms, some affect only a narrow band of frequencies or set of frequencies, and others affect a wide band of frequencies.

humanoid robot A robot that bears structural resemblance to a human being (e.g., has arms, a head, and perhaps legs). In its most advanced form, such a robot is an ANDROID.

hum-balance potentiometer A potentiometer connected across an alternating-current power supply, with its slider grounded. At a certain setting, hum interference is nulled.

hum bars See HORIZONTAL HUM BARS.

hum bucking The reduction of hum interference by introducing an alternating-current voltage of the same frequency and amplitude as the hum, but opposite in phase.

hum-bucking coil An auxiliary coil used in conjunction with the field and voice coils of an electrodynamic speaker. Reduces hum interference via HUM BUCKING.

hum field The magnetic field surrounding a conductor carrying hum-frequency alternating current.

humidity The amount of moisture in the air. Also see ABSOLUTE HUMIDITY and RELATIVE HUMIDITY.

humidity meter See ELECTRIC HYGROMETER and ELECTRONIC HYGROMETER.

humidity sensor A pickup whose resistance or capacitance varies proportionally with ambient humidity.

hum interference Electrical interference resulting from HUM in any of its various electrical forms (see HUM, **1–4**).

hum loop A ground loop that results in undesired hum in the output of an amplifier.

hummer A nonelectronic audio oscillator similar to the fork oscillator, but using a thick, metal reed, instead of a tuning fork. A carbon microphone button attached to the reed provides the feedback path necessary for sustained oscillation.

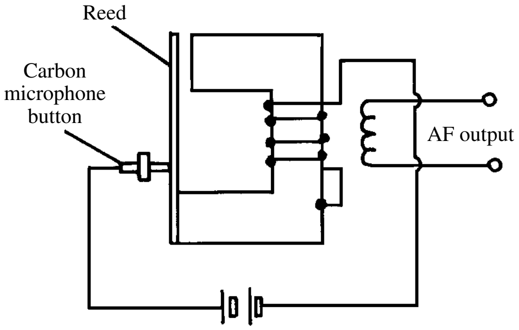

hummer

hum modulation Undesirable modulation of a radio signal or audio amplifier output signal by HUM interference.

hump **1.** Either the positive or the negative half-cycle of a sine wave; that is, either the portion from 0 degrees to 180 degrees, or the portion from 180 degrees to 360 degrees. **2.** A curve whose graph is a multiple of the sine function, and whose domain is restricted to a half-cycle (0 to 180 degrees or 180 to 360 degrees). See, for example, DOUBLE-HUMP RESONANCE CURVE and DOUBLE-HUMP WAVE.

hunting An undesirable condition in which an electrical or electromechanical system oscillates back and forth, relative to some mean mode of operation ("hunts" for the mode), sometimes eventually settling down at the mode. It can occur in a servo system because of overcompensation, lack of hysteresis, or both.

Huygens' principle The observation that an advancing wave is the resultant of secondary waves that arise from points in the medium that have already been passed.

HV Abbreviation of HIGH VOLTAGE.

H vector A vector representing the magnetic field of an electromagnetic wave. Compare E VECTOR.

hW Abbreviation of HECTOWATT.

H-wave mode In a waveguide, a mode of transmission in which the electric lines of flux are at right angles to the direction of the waveguide. Also called *transverse-electric (TE) mode.*

hy Occasional abbreviation of HENRY. The SI abbreviation and symbol, H, is preferred.

hybrid Descriptive of a device that is an offspring of other devices or a product of dissimilar technologies (but using elements of each). See, for example, HYBRID JUNCTION and HYBRID COIL.

hybrid active circuit An active circuit (such as an amplifier, oscillator, or switch) using a combination of two dissimilar active devices (e.g., transistors and vacuum tubes).

hybrid coil A special type of bridging transformer used in wire telephony to prevent self-oscillation in a repeater amplifier that operates in both directions.

hybrid computer A computer system incorporating more than one major computer technology. Examples: analog/digital and digital/neural-network.

hybrid electromagnetic wave Abbreviation, HEM wave. An electromagnetic wave whose electric-field and magnetic-field vectors are both in the direction of propagation.

hybrid IC See HYBRID INTEGRATED CIRCUIT.

hybrid integrated circuit Abbreviation, HIC. An integrated circuit embodying both integrated and microminiature discrete components (i.e., one combining both monolithic and thin-film construction).

hybrid junction **1.** See MAGIC TEE. **2.** A four-terminal device, such as a resistor circuit, special transformer, or waveguide assembly, in which a signal applied to one pair of terminals divides and appears at only the two adjacent terminals.

hybrid microcircuit A microcircuit containing diffused or thin-film elements interconnected with separate chip elements.

hybrid parameters See H PARAMETERS.

hybrid ring A hybrid waveguide junction (see HYBRID JUNCTION, **2**) consisting essentially of a reentrant line with four side arms. Used as an equal power divider.

hybrid-tee See HYBRID JUNCTION, **1**.

hybrid thin-film circuit A microcircuit in which semiconductor devices and discrete components are attached to passive components and conductors that have been electrodeposited on a substrate.

hybrid transformer See HYBRID COIL.

hydroacoustic Pertaining to the sound of fluids, especially water, under pressure.

hydroacoustic transducer A transducer that converts energy from the high-pressure flow of a fluid into acoustic energy.

hydrodynamic pressure The pressure of a fluid in motion. Compare HYDROSTATIC PRESSURE.

hydroelectric Pertaining to the production of electricity by water power, as by a generator turned by water turbines.

hydroelectric machine A device for generating electricity from high-pressure steam escaping from a series of jets.

hydroelectric power See WATER POWER.

hydrogen Symbol, H. A gaseous element. Atomic number, **1**. Atomic weight, 1.00794. Used in making semiconductor materials, it is the lightest and most abundant element in the universe. Compare DEUTERIUM and TRITIUM.

hydrogen atmosphere The nonoxidizing atmosphere in which semiconductor materials are melted and processed, and in which semiconductor crystals are grown. Occasionally, helium is used instead of hydrogen.

hydrogen atom A single atom of the element hydrogen, consisting of one electron and one proton.

hydrogen-ion concentration See pH.

hydrogen lamp A glow-discharge lamp that produces light by means of the ionization of rarefied hydrogen gas. Visible light is emitted at discrete wavelengths.

hydrokinetic Pertaining to fluids in motion or the forces behind such motion.

hydrolysis The process whereby chemical substances become ionized in water solution, producing electrolytes.

hydromagnetics See MAGNETOHYDRODYNAMICS.

hydromagnetic wave In a fluid, a wave in which the energy is propagated via magnetic and dynamic modes.

hydrometer An instrument for measuring the specific gravity of liquids.

hydrophone An underwater sound-to-electricity transducer (microphone).

hydropower See WATER POWER.

hydrostatic pressure The pressure of a fluid at rest. Compare HYDRODYNAMIC PRESSURE.

hygristor A resistor, the resistance of which depends on humidity. A common application is in the measurement of relative atmospheric humidity.

hygrogram A graphic rendition of atmospheric humidity versus time, made by a HYGROGRAPH.

hygrograph A device that produces a graphic recording of atmospheric humidity as a function of time. Consists essentially of a HYGROMETER connected to a permanent data-storage device, such as a computer or pen recorder.

hygrometer An instrument for measuring humidity. Also see ELECTRIC HYGROMETER, ELECTRONIC HYGROMETER, and HAIR HYGROMETER.

hygroscopic material A material that absorbs moisture from the air, but not enough to get wet (e.g., lime and silk). Compare DELIQUESCENT MATERIAL.

hygrostat A humidity-sensitive relay or switching circuit.

hygrothermograph A graphic recorder indicating humidity and temperature on the same chart, both as functions of time.

hyperacoustical zone In the upper atmosphere, a region in which the distance between air molecules is comparable to the wavelengths of audible sound. As the altitude increases within this zone, the upper-frequency limit of effective sound propagation decreases. At altitudes above this zone, no audible sound can be propagated.

hyperbola A conic-section curve satisfying the equation:

$$(x - x_0)^2/a^2 - (y - y_0)^2/b^2 = 1$$

where x and y are the independent and dependent variables, x_0 and y_0 are the coordinates of the center, a is half the length of the major axis, and b is half the length of the minor axis.

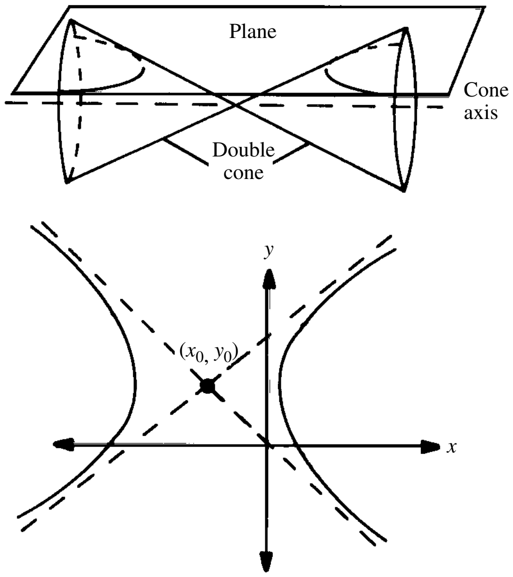

hyperbola

hyperbolic angle An angle subtended by a sector of a hyperbola in a manner analogous to that in which a circular angle is subtended by an arc of a circle.

hyperbolic-cosine horn See CATENOIDAL HORN.

hyperbolic error **1.** In an interferometer, a miscalculation in the direction of arrival of a signal. The signal from one antenna in the system can be assumed to be in phase with the signal from another antenna, when actually the two components differ by an integral number of whole wavelengths. **2.** The angular error, in degrees, minutes, or seconds of arc, resulting from a miscalculation of phase in an interferometer.

hyperbolic face contour See HYPERBOLIC GRIND.

hyperbolic functions The nonperiodic functions of a hyperbolic angle. Hyperbolic functions are related to the hyperbola in the same manner that common trigonometric functions are related to the circle. Hyperbolic functions are useful in calculations involving traveling waves on transmission lines, attenuator design, and wave-filter design.

hyperbolic grind The shape (approximately hyperbolic) to which the face of a magnetic recording head is ground. It provides optimum contact with the tape and ensures good high-frequency response.

hyperbolic horn A horn antenna whose cross-sectional area is a hyperbolic function of the distance along the axis.

hyperbolic logarithm See NAPIERIAN LOGARITHM.

hyperbolic navigation A radionavigation system in which the operator of an aircraft or boat determines position by comparison of two received signals. The two transmitters radiate signals from known positions and with known timing characteristics. The time delay from each transmitter is determined, resulting in two hyperbolic curves on a map. The point of intersection of the curves is the location of the aircraft or ship.

hyperbolic radian A unit of measure derived from a hyperbolic angle. A hyperbolic radian is the hyperbolic angle that encloses an area of 0.5 when the distance along the x-axis to the hyperbola is unity. Also see HYPERBOLIC ANGLE. Compare CIRCULAR RADIAN.

hyperbolic trigonometry The branch of mathematics dealing with the theory and application of hyperbolic angles and their functions.

hypercardioid microphone A unidirectional microphone with exceptional sensitivity in front, and minimal responsiveness from the sides and rear.

hypercardioid pattern A directional CARDIOID PATTERN with accentuated responsiveness or emission in the favored direction (front), and greatly suppressed responsiveness or emission in other directions (sides and rear).

hyperfocal distance The shortest distance to which a lens can be focused without degrading definition at infinity.

hyperfrequency waves See MICROWAVES.

hypernik See HIPERNICK.

hyperon Any one of various particles having a mass greater than that of a neutron or proton.

hyperpolarization The production of an increased voltage across a biological membrane.

hypersonic Pertaining to speeds of at least five times that of sound. In air at normal atmospheric pressure, such speeds are more than approximately 1700 meters, or 5600 feet, per second.

hypersonic speed Any speed greater than five times the speed of sound in a given medium.

hypersyn motor A high-efficiency, high-power-factor synchronous motor combining the advantages of the *direct-current-excited synchronous motor* (stiffness), the *hysteresis motor* (synchronizing torque), and the *induction motor* (high starting torque).

hypervelocity Velocity in excess of 3 kilometers per second.

hypotenuse The side of a right triangle opposite to the right angle.

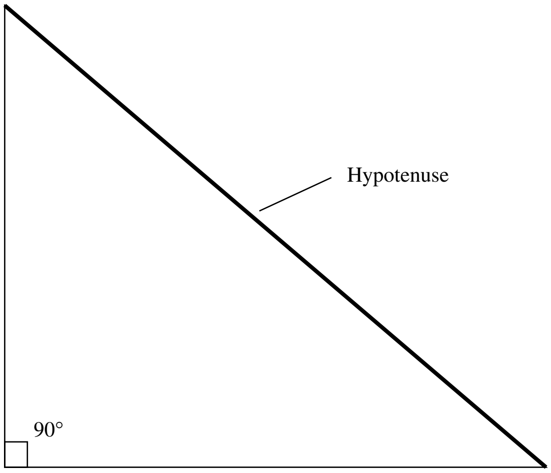

Hypotenuse

hypothesis An idea, concept, or system that seems true or workable, but must be subjected to logical analysis and/or practical testing to prove its validity.

hypsometer An altimeter in which a thermistor (connected to a battery and current meter) is immersed in a boiling liquid. Because the liquid's boiling point is proportional to altitude, it affects the resistance of the thermistor and, hence, the deflection of the meter.

hysteresigram The hysteresis-curve record produced by a HYSTERESIGRAPH.

hysteresigraph A graphic recorder that displays or records the hysteresis curve for a material. Also see HYSTERESISCOPE.

hysteresimeter See HYSTERESIS METER.

hysteresis **1.** The tendency of a magnetic material to saturate and retain some of its magnetism after the alternating magnetic field to which it is subjected reverses polarity, thus causing magnetization to lag behind the magnetizing force. **2.** A similar electrostatic action in a ferroelectric dielectric material. **3.** In a servo system, the condition in which a variable quantity lags the effect or stimulus that causes it to change. The plot for this is a double-line HYSTERESIS CURVE.

hysteresis brake A brake whose retarding action comes from hysteresis in a permanent-magnet motor.

hysteresis clutch A magnetic clutch whose output torque (for synchronous drive or continuous slip) comes from hysteresis in a permanent-magnet motor.

hysteresis coefficient In a sample of iron whose volume is one cubic centimeter, the energy in ergs dissipated during one cycle of magnetization. Also called *coefficient of hysteresis.*

hysteresiscope An oscilloscope that is specially designed to display the hysteresis curve of a material. Compare HYSTERESIGRAPH.

hysteresis curve A response curve depicting hysteresis in a magnetic material, a dielectric, or a servo system. A graph of the extent to which a variable quantity lags the effect or stimulus that causes it to change (e.g., a curve showing the effect of response delay in a thermostatically controlled heating/cooling system).

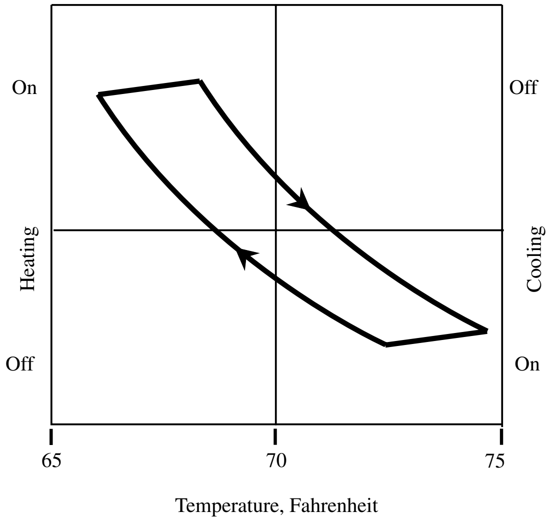

hysteresis curve

hysteresis cycle A complete hysteresis curve.

hysteresis distortion Signal distortion in iron-core components, such as coupling transformers, resulting from hysteresis in the iron.

hysteresis error In a meter, a difference in indications for increasing and decreasing current, an effect caused by hysteresis in iron meter parts.

hysteresis heater An induction heater in which heating results from hysteresis loss in the load.

hysteresis loop See HYSTERESIS CURVE.

hysteresis loss Power loss caused by hysteresis in a magnetic material exposed to an alternating magnetic field, or in a dielectric material exposed to an alternating electric field. It is characterized by the generation of heat.

hysteresis meter An instrument that determines the hysteresis loss in a ferromagnetic material in terms of the torque produced when the material is rotated in a magnetic field, or vice versa.

hysteresis motor A synchronous motor that does not require direct-current excitation, and does not have salient poles. It is started by means of hysteresis losses that the rotating magnetic field causes in the secondary.

hysteretic constant For a ferromagnetic material, hysteresis loss in ergs per cubic centimeter of material per cycle of magnetization.

hysteretic loss See HYSTERESIS LOSS.

hystoroscope A device used to determine the magnetic characteristics of a material.

Hz Abbreviation of HERTZ.

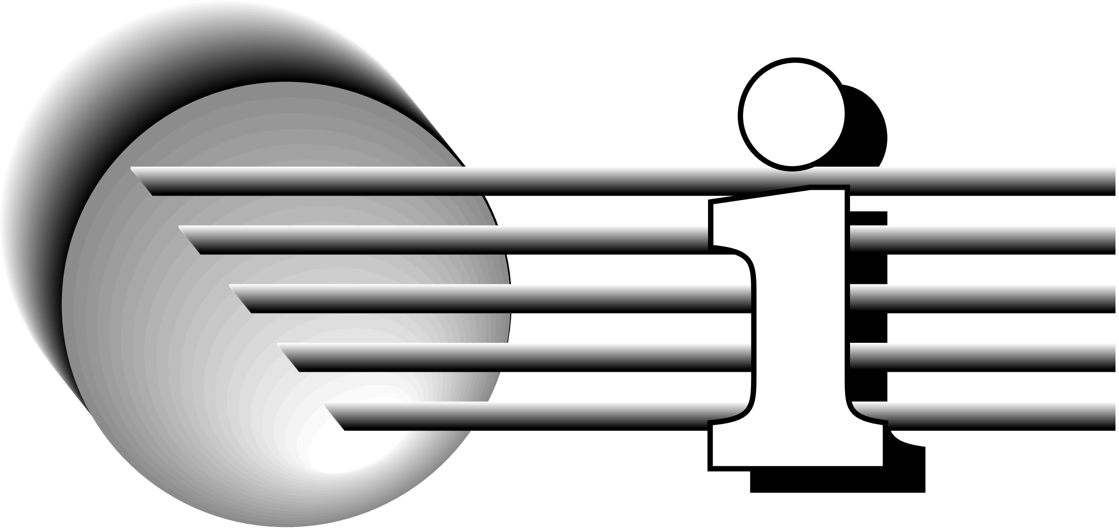

I **1.** Symbol for CURRENT. **2.** Symbol for INTRINSIC SEMICONDUCTOR. **3.** Symbol for LUMINOUS INTENSITY. **4.** Symbol for IODINE.

i **1.** Symbol for the unit imaginary number, the square root of -1. (Also, *j*.) **2.** Subscript for INSTANTANEOUS VALUE. **3.** Symbol for INTRINSIC SEMICONDUCTOR. **4.** Symbol for ANGLE OF INCIDENCE. **5.** Symbol for INSTANTANEOUS CURRENT. (Also, I_i.) **6.** Symbol for a unit vector parallel to the x-axis. **7.** Symbol for INCIDENT RAY.

I_{ac} Symbol for the ac component of a COMPOSITE CURRENT.

I_{AF} Symbol for AUDIO-FREQUENCY CURRENT.

IAGC Abbreviation of INSTANTANEOUS AUTOMATIC GAIN CONTROL.

IAVC Abbreviation of INSTANTANEOUS AUTOMATIC VOLUME CONTROL.

I_B **1.** Symbol for *plate power-supply current.* **2.** Occasional symbol for *plate current* (usually, I_P).

IBM Abbreviation for *International Business Machines* Corporation.

IC **1.** Abbreviation of INTEGRATED CIRCUIT. **2.** Abbreviation of INTERNAL CONNECTION.

I_C **1.** Symbol for transistor COLLECTOR CURRENT. **2.** Occasional symbol for *grid current* (usually, I_G).

ICAD Abbreviation of *integrated control and display.*

ICAS Abbreviation of INTERMITTENT COMMERCIAL AND AMATEUR SERVICE.

I_{CBO} Symbol for the static reverse collector (leakage) current in a common-base connected transistor with an open-circuited emitter.

ICBS Abbreviation of *interconnected business system.*

I_{CEO} Symbol for the static reverse collector (leakage) current in a common-emitter-connected transistor with an open-circuited base.

ice loading **1.** In an antenna, power-line system, or other structure, the additional stress caused by accumulation of ice. **2.** The weight or thickness of ice a structure can safely withstand.

ice-removal circuit A high-voltage, low-frequency power supply used to heat certain antennas to melt ice that accumulates on them.

ICET Abbreviation of *Institute for the Certification of Engineering Technicians* (National Society of Professional Engineers).

ICME Abbreviation of *International Conference on Medical Electronics.*

I_{co} Symbol for the collector cutoff current (static leakage current) of a bipolar transistor (see CUTOFF CURRENT).

icon In a graphical computer interface, a symbol that aids the user in recognizing a selection that can be made.

iconoscope A camera tube in which an electron beam scans a photomosaic on which the image is focused. The light-sensitive droplets of the mosaic form tiny capacitors with the insulated, metallic backplate of the mosaic, each capacitor becoming charged by the light of the picture. As the electron beam scans the mosaic, each capacitor discharges as the beam strikes it, delivering an output pulse proportional to the light intensity at that spot in the picture.

IC tester An instrument for checking the operation of an integrated circuit.

ICW Abbreviation of INTERRUPTED CONTINUOUS WAVE.

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

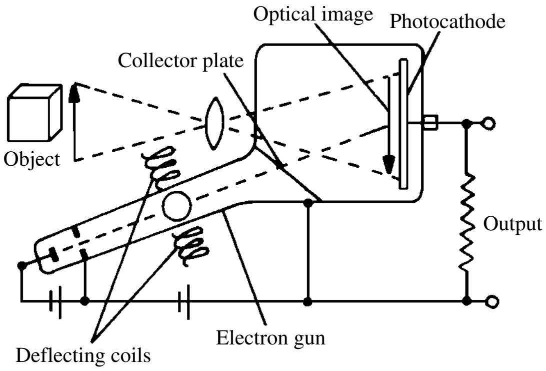

iconoscope

ID Abbreviation of INSIDE DIAMETER.

IDA Abbreviation of *integrodifferential analyzer.*

I_{dc} Symbol for the dc component of a COMPOSITE CURRENT.

ideal Pertaining to a circuit, device, material, or manner of operation that is assumed to conform to the theoretical best-case example; it is not usually realized in real life. Thus, an ideal reactance has no inherent resistance.

ideal capacitor A capacitor having zero dielectric loss and a constant value of capacitance at all alternating-current frequencies.

ideal component A theoretical component that is assumed to be 100-percent efficient and to have no extraneous properties. Compare PRACTICAL COMPONENT.

ideal crystal A piezoelectric crystal that acts as a theoretically perfect tuned circuit, that is, an ideal capacitor and inductor.

ideal inductor An inductor having zero loss and a constant value of inductance at all alternating-current frequencies.

I demodulator In a color-television receiver circuit, a demodulator that receives the chrominance and 3.58-MHz oscillator signals and delivers a video output corresponding to color in the picture.

identical vectors Equal vectors that have the same initial point or point of application.

identification **1.** In radar operations, the (often automatic) determination of the target's identity. See, for example, IDENTIFICATION, FRIEND OR FOE. **2.** In digital computer operations, a symbol or set of symbols within a label identifying a unit of data or its location.

identification beacon **1.** A beacon used for the determination of a particular geographic location. **2.** An automatically transmitted station-identification signal or code, usually superimposed on the regular transmission in the form of a subcarrier or subaudible signal.

identification division The division (one of four) in a COBOL program that describes and identifies the program being compiled.

identification, friend or foe Abbreviation, IFF. A technique in which a radar station transmits an interrogating signal and the station questioned replies automatically with a suitable pulse or other signal if it is aboard a friendly aircraft or vessel. If it is aboard an enemy vehicle, the station gives no reply or sends an unsatisfactory one.

identifier A data file identification label in an input/output device, or a label that identifies a specific storage location.

identity element A logic element that, upon receipt of two input signals, provides an output that is logic 1 (high) only if the input signals are both logic 1 (high) or logic 0 (low).

idiochromatic Possessing the photoelectric properties of a true crystal.

I display See I SCAN.

idle channel noise A measurement of wideband noise. The total energy at an output of an integrated circuit or other communications stage with that output grounded (idle).

idle character A digital character that conveys no information, but helps maintain synchronization between the transmitter and receiver. Sometimes called a *blank.*

idler wheel In a phonograph or magnetic tape recorder, an auxiliary, rubber-coated wheel that transfers rotary motion from the motor pulley to the platter or the rim of the capstan flywheel.

idle time The period during which data processing equipment, although operable, is not in use.

idling Standby equipment operation, as when vacuum-tube filaments in a radio-frequency power amplifier are kept hot—even when the amplifier is not actually being used to transmit a signal.

idling current The current flowing in a device during a standby period, as opposed to operating current. Also called STANDBY CURRENT.

idling frequency In a parametric amplifier, the difference between the signal frequency and pump frequency.

idling power See STANDBY POWER.

idling voltage The voltage required by or measured in a device that is in standby mode.

$I_{D(off)}$ Symbol for drain cutoff current in a field-effect transistor.

IDOT Abbreviation of *instrumentation online transcriber.*

IDP **1.** Abbreviation of INDUSTRIAL DATA PROCESSING. **2.** Abbreviation of INTEGRATED DATA PROCESSING. **3.** Abbreviation of INTERMODULATION-DISTORTION PERCENTAGE.

I_{DSS} Symbol for DRAIN CURRENT AT ZERO GATE VOLTAGE in a field-effect transistor.

I_E Symbol for EMITTER CURRENT.

IEC Abbreviation of *integrated electronic component.*

IEE Abbreviation of *Institution of Electrical Engineers* (British).

IEEE Abbreviation of *Institute of Electrical and Electronics Engineers.*

IES Abbreviation of *Illuminating Engineering Society.*

IF Abbreviation of INTERMEDIATE FREQUENCY.

IF amplifier See INTERMEDIATE-FREQUENCY AMPLIFIER.

I_{FB} In an integrated circuit, current limit feedback or feedback current.

IF channel See INTERMEDIATE-FREQUENCY CHANNEL.

IF converter The converter (first detector-oscillator) section of a superheterodyne radio receiver.

IFF Abbreviation of IDENTIFICATION, FRIEND OR FOE.

IF gain **1.** The amplification provided by the intermediate-frequency (IF) channel of a superheterodyne receiver, usually specified in decibels. **2.** A control that allows adjustment of the amplification in the IF channel of a superheterodyne radio receiver.

IF interference See INTERMEDIATE-FREQUENCY INTERFERENCE.

IFIPS Abbreviation of *International Federation of Information Processing Societies.*

IF selectivity See INTERMEDIATE-FREQUENCY SELECTIVITY.

IF strip A (sometimes removable) circuit section containing a complete intermediate-frequency channel.

IF/THEN/ELSE In computer systems and artificial intelligence, a two-way choice that is made during the execution of a program. For example, in finding the absolute value of a number, the logic proceeds as follows: “IF $x < 0$, THEN multiply x by -1; otherwise (ELSE) leave it alone.”

IF transformer See INTERMEDIATE-FREQUENCY TRANSFORMER.

I_G Symbol for GATE CURRENT.

I_G Symbol for GRID CURRENT.

IGFET Abbreviation of INSULATED-GATE FIELD-EFFECT TRANSISTOR.

ignition coil A small open-core transformer having a high step-up turns ratio for converting 6- or 12-volt battery potential to the high voltage needed in an automotive ignition system.

ignition interference Electrical noise generated by the ignition system of an internal combustion engine.

ignition potential **1.** The minimum voltage needed to cause electrical discharge in an internal combustion engine. **2.** See BREAKDOWN VOLTAGE, **1**, **2**.

ignition reserve The extra voltage provided by the starter, as compared with the voltage actually needed for ignition of an internal-combustion engine.

ignition system An electrical or electronic system that supplies the high voltage in an automotive engine. See, for example, CAPACITOR-DISCHARGE IGNITION SYSTEM.

ignition voltage See IGNITION POTENTIAL.

ignitron A form of rectifier tube that contains ionized mercury. It is used with some extremely high-voltage power supplies.

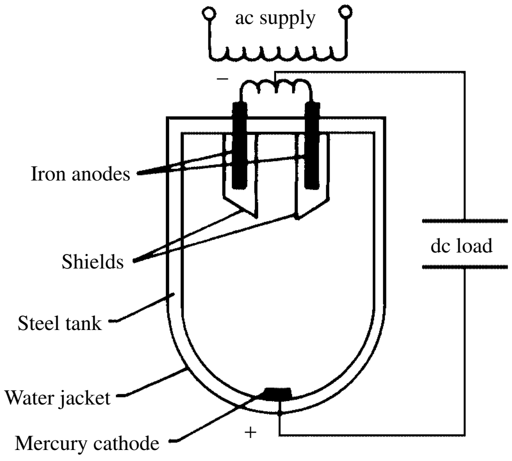

ignitron

ignore character **1.** A character used as a signal to inhibit an action. **2.** A digital character that is ignored by the system.

I_{GSS} Symbol for GATE REVERSE CURRENT in a field-effect transistor.

IGY Abbreviation for INTERNATIONAL GEOPHYSICAL YEAR.

I_H **1.** Symbol for HOLD CURRENT. **2.** Symbol for HOLDING CURRENT.

IHF Abbreviation of INHIBIT FLIP-FLOP.

IHFM Abbreviation for *Institute of High Fidelity Manufacturers.*

ihp Abbreviation of INDICATED HORSEPOWER.

ihp-hr Abbreviation of INDICATED HORSE-POWER-HOURS.

I_i **1.** Symbol for INPUT CURRENT. **2.** Symbol for INSTANTANEOUS CURRENT (also, i).

I_K Symbol for CATHODE CURRENT.

I_L Symbol for current in an inductor.

I^2L Abbreviation of INTEGRATED INJECTION LOGIC. (Also IIL.)

illegal character **1.** A character or bit group that is, according to some standard, invalid. **2.** A bit group that represents a symbol in a character set.

illuminance The amount of luminous flux received per unit surface area, measured in lux (lumens per square meter).

illuminant-C In color television, the reference white that closely resembles average daylight.

illuminated pushbutton See LIGHTED PUSHBUTTON.

illuminated switch See LIGHTED SWITCH.

illumination **1.** Visible light. **2.** The condition of being irradiated by visible light. **3.** The intensity of visible light irradiating an object or region.

illumination control A photoelectric circuit that automatically regulates electric lights, according to the amount of daylight.

illuminometer A device for measuring the intensity of visible light, or, less often, infrared or ultraviolet. In its simplest form, the meter consists of a photovoltaic cell, a potentiometer for adjusting the sensitivity, and a microammeter. More sophisticated devices include direct-current amplification.

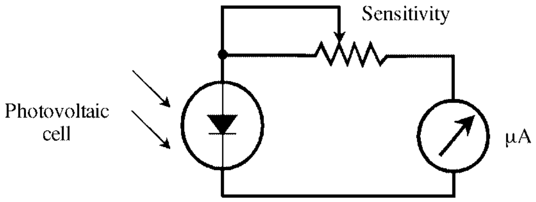

illuminometer

ILS Abbreviation of INSTRUMENT LANDING SYSTEM.

IM Abbreviation of INTERMODULATION.

I_m **1.** Abbreviation of *maximum current.* (Also, I_{max}). **2.** Abbreviation of *meter current.*

image **1.** In a superheterodyne circuit, a spurious response whose frequency differs from that of the desired signal by twice the intermediate frequency. **2.** A picture on the screen of a television receiver or computer display. **3.** A pattern on the screen of an oscilloscope tube. **4.** A picture on the mosaic of a television camera tube. **5.** A duplicate of a computer storage area that is in another part of storage or on another medium.

image admittance The reciprocal of IMAGE IMPEDANCE.

image antenna An imaginary "mirror" antenna below the effective radio-frequency (RF) ground surface, at a depth equal to the height of the true radiating antenna above effective RF ground.

image attenuation constant The real-number part of the IMAGE TRANSFER CONSTANT. Also see IMAGE PHASE CONSTANT.

image compression Any of various methods by which the number of bytes in a digital image can be reduced, without significant degradation of image quality. The two most common methods are *JPEG (Joint Photographic Experts Group) image compression* and *fractal image compression.* Other schemes include *MPEG (Moving Picture Experts Group) image compression* and *Indeo*, developed by Intel Corporation. Fractal image compression can provide a compression factor of as much as 100:1 without significant degradation of image quality.

image converter **1.** A device that changes an invisible image into a visible image. Examples include the *snooperscope*, an infrared-to-visible converter, and photographic apparatus for infrared, ultraviolet, and X-ray wavelengths. **2.** A tube that operates as an image converter.

image dissector See DISSECTOR TUBE.

image effect The effect of reflection of electromagnetic waves from the ground. An IMAGE ANTENNA appears to radiate from a point beneath the effective RF ground plane. The depth of the image antenna below the effective RF ground plane is equal to the height of the actual antenna above the effective RF ground plane.

image frequency The frequency of the image response in a superheterodyne radio receiver. See IMAGE, **1**.

image impedance The property of a network in which the load impedance is "seen," looking into the output terminals with the generator connected to the opposite end, and the generator impedance is "seen," looking into the input terminals with a load connected to the opposite end.

image intensification An increase of the brightness of the display on a cathode-ray tube.

image intensifier A device that increases the brightness of the spot produced when a beam of electrons or X rays hits a fluorescent screen.

image interference A type of interference that can occur in superheterodyne circuits. It occurs when there is a sufficiently strong signal on the IMAGE FREQUENCY.

image orthicon See ORTHICON.

image phase constant The imaginary-number part of the IMAGE TRANSFER CONSTANT. Also see IMAGE ATTENUATION CONSTANT.

image potential The potential energy of a charged particle at a given distance from a metal surface. The metal surface acts in a way similar to a mirror; a reflected image of the object, having equal charge but opposite polarity, is formed on the other side of the surface.

image ratio See SIGNAL-TO-IMAGE RATIO.

image rejection In a superheterodyne radio receiver, the suppression or elimination of IMAGE INTERFERENCE by means of a selective circuit, such as a radio-frequency preamplifier.

image response In a superheterodyne receiver, an undesired response to signals removed from the desired frequency by twice the intermediate frequency.

image transfer constant A number depicting the transfer of power by an impedance network. It has the same value, regardless of the direction of transmission through the network. Also see IMAGE ATTENUATION CONSTANT and IMAGE PHASE CONSTANT.

imaginary axis In a vector diagram of complex impedance, the axis of the imaginary-number component (*jX*).

imaginary number A real-number multiple of the positive square root of -1. Engineers depict the positive square root of -1 by the lowercase letter *j*; mathematicians use the lowercase letter *i*. Engineers write imaginary numbers in the form *jX*, and mathematicians write *Xi*, where *X* is a real number.

imaginary-number component The imaginary-number part of a COMPLEX NUMBER.

I_{max} Abbreviation of *maximum current.*

IM distortion meter INTERMODULATION METER.

IM distortion percentage See INTERMODULATION-DISTORTION PERCENTAGE.

I_{min} Abbreviation of *minimum current.*

imitation The transmission of false signals for purposes of deception. For example, during wartime, the signals from an enemy station might be recorded and retransmitted.

immediate access **1.** The ability of a computer to store and retrieve data in a minimal amount of time. **2.** Computer storage that can be accessed in a minimal amount of time.

immediate address An instruction address that is used as data by that instruction.

IM meter See INTERMODULATION METER.

immitance Impedance or admittance; a contraction of IMPEDANCE and ADMITTANCE. Example: a *negative-immitance circuit.*

immortal knowledge Knowledge that can be kept indefinitely in the form of detailed computer data files. Some scientists and sociologists believe that this will eventually have a profound effect on the way that future generations view history.

impact excitation See SHOCK EXCITATION.

impact strength **1.** The ability of a component or material to withstand mechanical shock loading. **2.** The work required to fracture the material under shock loading.

IMPATT diode Acronym for *impact avalanche transit time* diode. A microwave semiconductor (silicon or gallium arsenide) diode exhibiting negative resistance resulting from the combined effects of charge-carrier transit time and impact avalanche breakdown. It is used as an oscillator or amplifier.

IMPATT oscillator A microwave oscillator that uses an IMPATT DIODE.

impedance Symbol, *Z*. Unit, ohm. The total opposition offered by a circuit or device to the flow of alternating current. It is the vector sum of RESISTANCE and REACTANCE. This is a COMPLEX NUMBER whose real-number component is resistance *R*, and whose imaginary-number component is reactance *jX*; mathematically, $Z = R + jX$.

impedance angle The angle between the resistance and impedance vectors in an IMPEDANCE TRIANGLE.

impedance arm The network branch that contains one or more impedances, as opposed to an arm that contains only resistance or (predominantly, reactance). Also called *impedance leg.*

impedance branch See IMPEDANCE ARM.

impedance bridge **1.** An alternating-current (ac) bridge (commonly operated at 1 kHz) used to measure resistance, inductance, capacitance, and resistive components associated with inductors and capacitors, from which impedance can be calculated. **2.** Sometimes, an ac half-bridge circuit in which an unknown impedance is compared with a known resistance. **3.** A radio-frequency bridge circuit whose balancing element reads impedance directly in ohms.

impedance bump A discontinuity in the characteristic impedance of a radio-frequency transmission line. It is often caused by the use of improper splicing techniques.

impedance coil See CHOKE COIL.

impedance converter See IMPEDANCE TRANSFORMER.

impedance-coupled amplifier An amplifier using capacitor/coil combinations for interstage and output load coupling.

impedance drop In an alternating-current circuit, the complex sum of the resistance drop and reactance drop.

impedance ground A ground connection in which the impedance at the operating frequency is determined by a network of resistors, capacitors, and/or inductors.

impedance leg See IMPEDANCE ARM.

impedance magnetometer A device that is used for measuring small local variations in the intensity of the earth's magnetic field. A small change in the intensity of the magnetic field will cause a change in impedance of a nickel-iron wire having high permeability.

impedance match The condition (for maximum power transfer) when the transmitting impedance equals the receiving impedance, or when a suitable transformer is inserted between different impedances for matching purposes. Also see IMPEDANCE MATCHING.

impedance matching **1.** The adjustment or modification of two impedances so that they are identical, that is, the two resistive components are equal, and the two reactive components are equal. **2.** The insertion of a suitable transformer or network between circuits having different impedances, for the purpose of optimizing power transfer.

impedance-matching network A network of discrete components, often adjustable, that is used to match a circuit having a certain impedance to a circuit having a different impedance. An example is the inductive-capacitive (LC) coupler, also called a *transmatch*, commonly used to match a radio transmitter to an antenna system.

impedance-matching transformer See IMPEDANCE TRANSFORMER.

impedance meter See Z METER.

impedance plethysmograph An electronic device used to measure changes in the chemical content of body cells.

impedance poles See POLES OF IMPEDANCE.

impedance ratio The quotient of two impedances that are related in some situation, such as impedance match or impedance mismatch. The impedance ratio of a transformer is equal to the square of the turns ratio.

impedance transformer **1.** A transformer for converting an impedance to a different value. The turns ratio is equal to the square root of the impedance ratio. **2.** An emitter follower or source follower circuit, used primarily to match a high impedance to a lower impedance. **3.** A short-circuited transmission-line section used to match or convert impedances at radio frequencies.

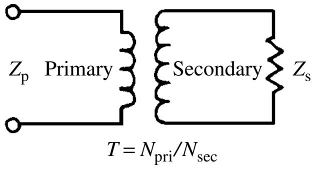

impedance transformer

impedance triangle A triangular vector diagram in which the impedance vector is the hypotenuse, and the reactance and resistance vectors are the perpendicular sides.

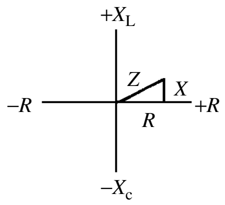

impedance triangle

impedance vector In a vector diagram, the resultant (vector) representing the combined reactance and resistance vectors. Also see IMPEDANCE TRIANGLE.

impedance zeros See ZEROS OF IMPEDANCE.

imperative macroinstruction A MACROINSTRUCTION used to create object (machine language) program instructions.

imperative statement A source language program instruction that is converted into object program (machine language) instructions.

imperfection A fault in the lattice of a crystal. Also see CRYSTAL LATTICE.

implementation **1.** Putting a system into operation and evaluating its performance. **2.** Doing all that is necessary to install a device or system.

implosion The inward collapse of an evacuated chamber, such as the envelope of a cathode-ray tube. It is the opposite of explosion, an outward bursting.

impregnant A substance for the IMPREGNATION of electronic components. Examples: waxes, oils, liquid plastics, and varnish.

impregnation The process of filling the spaces within a material or structure with an insulating compound. Various electronic components, such as capacitors, inductors, transformers, transistors, and diodes, are impregnated for protection and stability.

impressed voltage The voltage applied or presented to a circuit or device.

impulse **1.** A signal of extremely short duration (e.g., a noise burst). **2.** A brief, unidirectional surge in voltage (usually) or current.

impulse excitation Driving a tank circuit with a single pulse, then allowing it to oscillate at its own frequency until another driving pulse arrives.

impulse frequency In a digital telephone, the number of pulse periods per second, corresponding to a dialed digit, generated by the pulse springs.

impulse generator See MARX GENERATOR.

impulse noise Electrical noise from short-duration pulses, such as those produced by an automotive ignition system.

impulse ratio The ratio V_p/V_s, where V_p is the breakdown (or flashover or sparkover) voltage caused by an impulse, and V_s is the corresponding voltage at the crest of the power-frequency cycle.

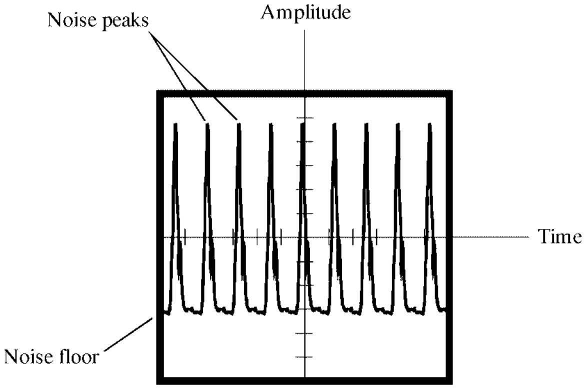

impulse noise

impulse relay A relay that is able to close or open completely when driven by a short pulse.

impulse speed The switching rate of a telephone dialing device as it transmits pulses.

impulse timer A synchronous-motor-driven timer whose cams can control many circuits; it can advance by a number of specified increments, as controlled by an integral stepping mechanism.

impulse transmission A method of transmission in which defined impulses are used to denote changes in signal content or format.

impurity A substance added to an intrinsic semiconductor to alter its electrical properties.

impurity atom In a processed semiconductor material, an atom of an IMPURITY material that produces either n- or p-type properties to the intrinsic semiconductor.

impurity density In the manufacturing process of a semiconductor material, the amount of impurity added to the original semiconductor.

impurity ion In a crystal, an ion in a space between atoms, or one taking the place of an atom.

impurity level **1.** The energy existing in a semiconductor material as a result of doping (the addition of an impurity). **2.** See IMPURITY DENSITY.

impurity material See IMPURITY.

IN Abbreviation for INTELLIGENT NETWORK.

In Symbol for INDIUM.

I_n Symbol for the *n*th value of current in a series of values.

in **1.** Abbreviation of INPUT. **2.** Abbreviation of INCH.

inaccuracy **1.** The state or condition of instrument error. **2.** The difference between the actual value of a parameter and the value indicated by an instrument. **3.** The percentage of instrument error.

inactive leg Within a transducer, an electrical component whose characteristic remains unchanged when the stimulus (quantity being transduced) is applied; specifically, a Wheatstone bridge element in a transducer.

inactive lines In a conventional television picture, blanked lines that do not contribute to the visible part of the image. Approximately half of these lines are at the top of the screen; the other half are at the bottom.

inactive time The period during which a radioactivity counter is insensitive to ionizing agents.

incandescence The state of glowing from intense heat, as when a metal becomes white hot from an electric current flowing through it.

incandescent lamp A filament-type lamp. The filament becomes so hot when an electric current passes through it that it glows brilliantly.

inch Abbreviation, in. A unit of linear measure in the English system; 1 in. = 2.54 centimeters = 0.0254 meter.

inching See JOGGING.

inch-pound Abbreviation, in-lb. A unit of work equal to a force of 1 pound exerted over a distance of 1 inch. Compare FOOT-POUND.

incident **1.** A failure in a computer system requiring the intervention of an operator in removing or revising the job involved. **2.** Pertaining to a ray or rays of energy striking a surface or boundary, or encountering a device (such as an antenna or photovoltaic cell).

incidental AM Undesired amplitude modulation in a frequency-modulated signal. Compare INCIDENTAL FM.

incident field intensity The field strength of an electromagnetic field as it arrives at a receiving antenna.

incidental FM Undesired frequency modulation in an amplitude-modulated signal. Compare INCIDENTAL AM.

incidental time Computer time devoted to other than program runs or program development.

incident light The light that strikes or enters an altering device or medium. See INCIDENT RAY.

incident power In a transmission line, the power that reaches the end of the line. Compare REFLECTED POWER.

incident ray The ray that strikes the surface of a reflecting, refracting, or absorbing body. Compare REFLECTED RAY and REFRACTED RAY.

incident wave **1.** A wave propagated to the ionosphere. Compare REFLECTED WAVE and REFRACTED WAVE. Also see IONOSPHERE and IONOSPHERIC PROPAGATION. **2.** A wave that encounters a change (in density, for example) in a propagation medium or the transition point between media.

in-circuit tester An instrument that permits the checking of components (especially transistors) without removing them from the circuit in which they are wired.

inclination **1.** At a given geographic location, the angle between the earth's magnetic lines of flux and the horizon. This angle ranges from 0 degrees at the geomagnetic equator to 90 degrees at the geomagnetic poles. Compare DECLINATION, 1. **2.** The angle between the plane of a satellite orbit and the plane containing the equator of the earth, planet, or star around which the satellite revolves. Generally measured as the angle at which the satellite crosses the equator while passing from the southern hemisphere to the northern hemisphere. This angle is between 0 and 90 degrees for satellites in non-retrograde orbits, and between 90 and 180 degrees for satellites in retrograde orbits.

inclinometer An instrument for measuring INCLINATION. One form consists of a magnetic needle mounted so that it can swing inside a vertically mounted circular scale.

inclusive-OR Also called *OR operation.* A logical operation between two operands, the result of which depends on rules for combining bits in each position within the operands: an output of 1 results if one or both of the bits have a value of 1; zero only if both are zero. Compare EXCLUSIVE-OR.

incoherent light Electromagnetic radiation in the visible spectrum in which the waves are not aligned (not in phase) and can be of more than one wavelength.

incoherent radiation Electromagnetic radiation in which the waves are not aligned (not in phase) and can be of more than one frequency or wavelength.

incoming inspection The examination of equipment and materials as they enter a factory or laboratory, for the purpose of identifying damaged or faulty units.

incoming line A line that enters a device, facility, or stage. Compare OUTGOING LINE.

Incompleteness Theorem A mathematical theorem of significance in artificial intelligence. For any consistent set of axioms (postulates), there are always more true statements than provable statements.

incomplete program Also called *incomplete routine.* A computer program of generalized steps that must be augmented with specific requirements to be implemented for a given operation.

inconsistency Contradictory computer statements, as detected by the program.

inconsistent equations A set of equations that have no common solution.

Increductor A specialized radio-frequency magnetic amplifier or saturable reactor.

increment **1.** The difference between two successive values of a variable. **2.** A small change in a quantity, such as mass, distance, or time. **3.** A quantity to be added to another quantity. **4.** The difference in value between adjacent hash marks on a meter scale.

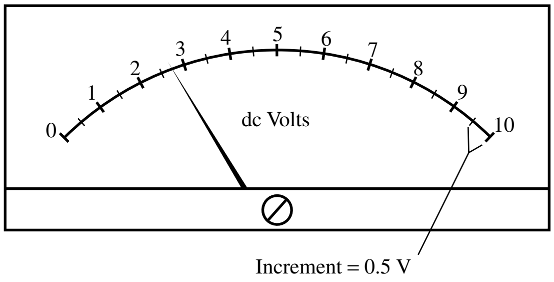

increment, 4

incremental computer A computer that operates on changes in variables. Example: differential analyzer.

incremental digital recorder A magnetic tape recorder that moves the tape across the record head in increments.

incremental display A device that converts digital data into a form for display (characters; graphs).

incremental inductance The inductance exhibited by an inductor, such as an iron-core choke, carrying a direct current.

incremental permeability The permeability exhibited by a material when an alternating-current magnetizing force is superimposed upon a direct current.

incremental plotter A device that, by direction of a computer program, provides the results of a program run in the form of curves or points on a curve, along with annotational characters.

incremental representation For incremental computers, a method of representing variables in terms of changes in the variables.

incremental sensitivity The smallest change in a quantity under measurement that can be detected by the instrument used.

ind **1.** Abbreviation of *indicator.* **2.** Abbreviation of INDUCTANCE (more commonly, *L*). **3.** Abbreviation of INDUCTOR (More commonly, *L*).

indefinite integral See INTEGRAL, **1**.

Indeo Also called *Digital Video Interactive (DVI).* A method of IMAGE COMPRESSION developed by Intel Corporation.

independent equations A set of equations having one common solution.

independent events In probability and statistics, the case where the occurrence of one event has no effect on the occurrence of another.

independent failure A circuit failure whose direct cause is not related to malfunctions elsewhere in the system.

independent mode In tracking supplies, an optional method of operation in which the separate units are adjustable independently of each other. Compare TRACKING MODE.

independent variable A changing quantity whose value at any instant is not governed by the value of any other quantity. Compare DEPENDENT VARIABLE.

index **1.** A reference line, hair, or point (e.g., a *file index*). **2.** In mathematics, an exponent. **3.** A ratio of one quantity to another, as *index of refraction.* **4.** In a computer memory, a table of references in a key sequence; it can be addressed to find the addresses of other data items. **5.** A number that is used to select a specific item within an array of items in memory.

index counter Also called *tape counter.* In a magnetic tape recorder, a (usually electromechanical) counting device that the operator can refer to in order to find material on the tape.

indexed address During or preceding the execution of a computer program instruction, an address that is modified by the content of an index register.

indexing **1.** An information retrieval technique used with files on a direct-access storage medium or on tables in memory. **2.** To modify an instruction using an index word.

index of modulation In frequency modulation, the ratio of carrier frequency deviation to modulating frequency.

index of refraction Symbol, *n*. The ratio v_1/v_2, where v_1 is the speed of energy propagation in the first medium through which the energy passes, and v_2 is the speed in the second medium.

index register Abbreviation, XR. In digital computer operations, a register holding a modifier that allows data to be directly addressed (each program refers to an index register when addressing storage locations). Also called MODIFIER REGISTER.

index word A word (bit group) containing a modifier that will be added to a basic instruction when it is executed during a program run.

India mica High-grade mica mined in India. Its excellent dielectric properties make it useful for capacitor stacks, high-*Q* radio-frequency circuits, and other critical applications.

indicated horsepower Abbreviation, ihp. Horsepower calculated from data or ratings, as opposed to measured horsepower.

indicated horsepower-hours Abbreviation, ihp-hr. Horsepower-hours based on calculation of indicated horsepower.

indicating fuse A fuse that provides some signal (such as a protruding pin) to show that it has blown.

indicating instrument An instrument, such as a meter, that provides direct readings of a measured quantity, as opposed to an instrument, such as a bridge, that must be manipulated and whose operation must often be followed by calculations.

indicating lamp A lamp that is marked or coded so that when it is on or off it conveys information.

indicator **1.** Meter (see METER, **1**). **2.** See MONITOR. **3.** See ANNUNCIATOR. **4.** In a computer, a device that can be set by a specific condition (e.g., by a negative result or error indicator).

indicator probe A test probe having a built-in meter.

indicial response Symbol, I(T). The sum of the transient and steady-state responses to a unit function.

indirect addressing In computer programming, a technique in which the address in an instruction refers to a different location containing another address, that can specify yet another address or an operand. Also called *multilevel addressing.*

indirect coupling Collectively, capacitive and inductive coupling, as opposed to direct coupling.

indirect ground An unintentional ground connection (e.g., accidental grounding of part of a circuit) or one obtained through a roundabout path. Compare DIRECT GROUND.

indirect light Light that has been reflected from one or more surfaces. Compare DIRECT LIGHT.

indirectly controlled Influenced by a directly controlled parameter, but not itself directly controlled.

indirectly grounded Connected to earth or to the lowest-potential point in a system inadvertently or through a roundabout path (e.g., by means of an indirect ground). Compare DIRECTLY GROUNDED.

indirectly heated cathode An electron-tube cathode consisting of a cylindrical or rectangular sleeve coated with a substance that is a rich emitter of electrons; it is heated by a filament inside the cylinder.

indirectly heated thermistor A thermistor whose temperature is changed by a built-in heater (filament) operated by the control current.

indirectly heated thermocouple A meter thermocouple heated by a small heater (filament) through which the signal current passes.

indirect material A semiconductor substance in which electrons move from the conduction band to the valence band in discrete jumps or steps.

indirect measurement The measurement of a quantity by comparing it with a similar quantity, using an instrument that requires adjustment or manipulation (rather than a simple meter). For example, resistance can be measured with a bridge, instead of an ohmmeter. Compare DIRECT MEASUREMENT.

indirect piezoelectricity In a piezoelectric crystal, the application of a voltage for the purpose of producing a strain on the crystal. A piezoelectric buzzer operates on this principle.

indirect scanning A method of video scanning, in which a fast-moving spot of light scans the film or an object and is passed through the film (or reflected by the object) to a photocell.

indirect wave **1.** In communications, a wave that arrives at a receiver after having traveled via refraction, reflection, or both. **2.** A radio wave propagated via the ionosphere. **3.** A wave reflected from some object, such as the moon or a meteor trail. **4.** A wave received from a satellite, originating from a distant earth station.

indium Symbol, In. A metallic element. Atomic number, 49. Atomic weight, 114.82. Used as a dopant in semiconductor processing.

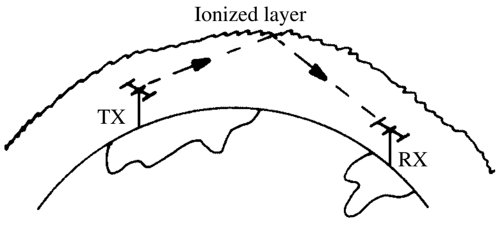

indirect wave

indoor antenna An antenna erected and operated in a building, but kept away from other objects as much as possible.

indoor radiation Electromagnetic radiation from the part of an antenna feeder or lead-in that is inside the transmitter building.

indoor transformer A power service transformer that, for protection against the elements, is installed inside the building it serves.

induced Brought about by the influence of a magnetic or electric field.

induced charge An electric charge produced in a body by the electric field surrounding another charge.

induced current An alternating current established in one circuit by the alternating magnetic field of another circuit. Also see INDUCTION.

induced EMF See INDUCED VOLTAGE.

induced failure A form of component failure that occurs because of operation beyond the normal specifications.

induced voltage An alternating voltage set up across one circuit (especially a coil) by the alternating magnetic field of another circuit. Also see INDUCTION.

inductance Symbol, *L*. Unit, henry. In a conductor, device, or circuit, the inertial property (caused by an induced reverse voltage) that opposes the flow of current when a voltage is applied; it opposes a change in current that has been established. Also see HENRY, INDUCTION, and MUTUAL INDUCTANCE.

inductance bridge An alternating-current bridge for measuring inductance in terms of a standard inductance or a standard capacitance. See, for example, HAY BRIDGE, MAXWELL BRIDGE, and OWEN BRIDGE.

inductance-capacitance Abbreviation, *LC*. **1.** A combination of inductance and capacitance in a circuit, such as a filter, a parallel-resonant circuit, or a series-resonant circuit. **2.** Pertaining to a device for measuring inductance and capacitance (e.g., *LC bridge* and *LC meter*).

inductance-capacitance bridge An alternating-current bridge for measuring inductance and capacitance only.

inductance-capacitance filter A filter composed of inductors and capacitors. Also called *LC filter.*

inductance-capacitance meter A direct-reading meter for measuring inductance and capacitance.

inductance-capacitance-resistance Abbreviation, *LCR*. **1.** A combination of inductance, capacitance, and resistance in a circuit, such as a basic tuned circuit. The resistive component represents loss in the inductor and capacitor. **2.** Pertaining to a device for measuring inductance, capacitance, and resistance (e.g., *LCR bridge* and *LCR meter*).

inductance-capacitance-resistance bridge See IMPEDANCE BRIDGE.

inductance coil See INDUCTOR.

inductance filter A filter using only an inductor, usually a coil of wire.

inductance-resistance time constant The time constant *t* (see ELECTRICAL TIME CONSTANT) of a circuit containing, ideally, only inductance and resistance. Mathematically, $t = LR$, where t is in seconds, L is the inductance in henrys, and R is the resistance in ohms. Also called *LR time constant.*

inductance standard A highly accurate, stable inductor used in precision measurements. Also see PRIMARY STANDARD and SECONDARY STANDARD.

induction **1.** The ability of an alternating, pulsating, or otherwise changing current flowing in one circuit to set up a current in a nearby circuit. The circuits need not be physically connected, but need only be linked by magnetic lines of flux. Also see SELF-INDUCTION. **2.** The phenomenon whereby a body becomes electrically charged by the field surrounding a nearby charged body. Also see ELECTRIC CHARGE.

induction coil A special high-voltage step-up transformer having an open core and a vibrator-interrupter in series with the primary winding, which carries direct current from a battery. The current is broken up into short pulses by the interrupter, and a high alternating-current voltage is generated in the secondary winding.

induction compass A compass whose indications depend on current induced in a coil revolving in the earth's magnetic field. Compare GYROCOMPASS and MAGNETIC COMPASS.

induction factor The ratio of total current to nonproductive current in an alternating-current circuit.

induction field The portion of an electromagnetic field that returns to a radiator, such as a coil, as opposed to the RADIATION FIELD.

induction frequency converter A mechanical device used for converting a signal at a fixed frequency to a signal at another fixed frequency.

induction furnace A furnace in which high-frequency magnetic fields induce currents in metal ores, causing the ore to become hot enough to melt.

induction heater A high-power, radio-frequency generator designed especially for induction heating.

induction heating The heating of metallic work samples by placing them in (but insulated from) a WORK COIL carrying current from a high-power radio-frequency generator. The workpiece is heated by radio-frequency currents induced in it as a result of the intense alternating magnetic field within the coil. Compare DIELECTRIC HEATING.

induction loss Loss of energy from a current-carrying conductor because of inductive coupling to a nearby conductor.

induction modulator See ELECTROMECHANICAL MODULATOR.

induction motor An electric motor in which the stator's rotating magnetic field makes the rotor revolve.

induction speaker An acoustic loudspeaker in which an audio-frequency current is passed through a diaphragm or coil located in a constant magnetic field. This results in movement of the diaphragm or coil.

induction transducer See INDUCTIVE TRANSDUCER.

induction-type landing system See DINGLEY INDUCTION-TYPE LANDING SYSTEM.

induction welding Welding in which the heating current flowing in the workpieces is induced by an electromagnetic field.

inductive capacitor A wound capacitor in which the inductance of the roll is controlled and specified. Such a capacitor is useful in compact filters and in single-frequency bypassing, where the reactive components are supplied by the capacitor. Compare NONINDUCTIVE CAPACITOR.

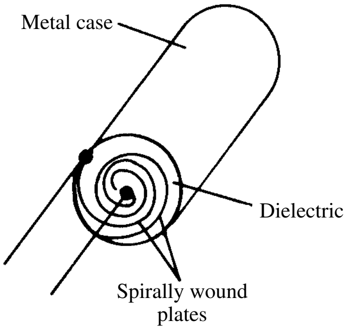

inductive capacitor

inductive circuit **1.** A circuit in which inductance predominates. **2.** A (theoretical) circuit containing inductance only.

inductive coupling The transfer of energy between two inductors (or inductive devices) by a linking electromagnetic field. Also see COEFFICIENT OF COUPLING, COUPLING, INDUCTION, and MUTUAL INDUCTANCE.

inductive feedback See MAGNETIC FEEDBACK.

inductive heater See INDUCTION HEATER.

inductive heating See INDUCTION HEATING.

inductive kick See BACK VOLTAGE and KICKBACK.

inductive load A load device that approaches a pure inductive reactance (e.g., loudspeaker and electric motor).

inductive loading In an antenna, the addition of inductance in series with the element(s). This reduces the resonant frequency for a radiator having a given physical length. It can also serve to reduce the physical length required for a radiator having a specified resonant frequency. Compare CAPACITIVE LOADING.

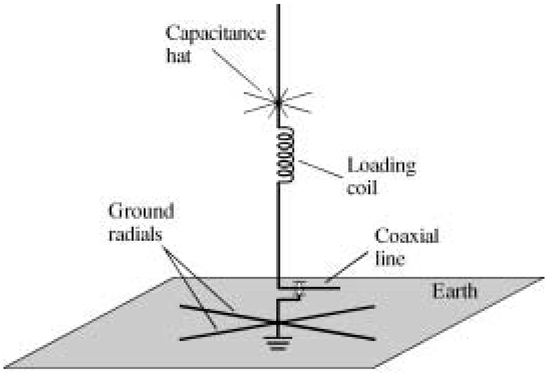

inductive loading

inductive logic A form of reasoning that demonstrates that a certain conclusion is highly probable, given a certain set of circumstances. This is of interest to researchers in artificial intelligence (AI). Compare DEDUCTIVE LOGIC.

inductive microphone A microphone in which sound waves vibrate a conductor or coil in a strong magnetic field, producing a corresponding alternating-current output by the resulting induction. Example: *dynamic microphone.*

induction neutralization Neutralization of a vacuum-tube radio-frequency power amplifier, via negative feedback from the output to the input through coupling coils.

inductive reactance Symbol, X_L. Unit, ohm. The reactance exhibited by an ideal inductor, considered as a positive imaginary-number quantity; $X_L = j6.28f_L$, where X_L is in ohms, f is the frequency in Hertz, L is the inductance in henrys, and j is the unit imaginary number (the square root of -1). Alternatively, f can be specified in megahertz, and L in microhenries. In a pure inductive reactance, current lags 90 degrees behind voltage. Also see INDUCTANCE, INDUCTION, INDUCTOR, and REACTANCE.

inductive switching Switching operations in a circuit containing an inductor. Switching time is influenced by the INDUCTANCE-RESISTANCE TIME CONSTANT of the inductor; overall operation is affected by the back voltage generated by the inductor.

inductive transducer A transducer in which the sensed phenomenon causes a change in inductance (or reluctance), which, in turn, causes a proportional change in output current, voltage, frequency, or bridge balance. Compare CAPACITIVE TRANSDUCER, CRYSTAL TRANSDUCER, MAGNETIC TRANSDUCER, and RESISTIVE TRANSDUCER.

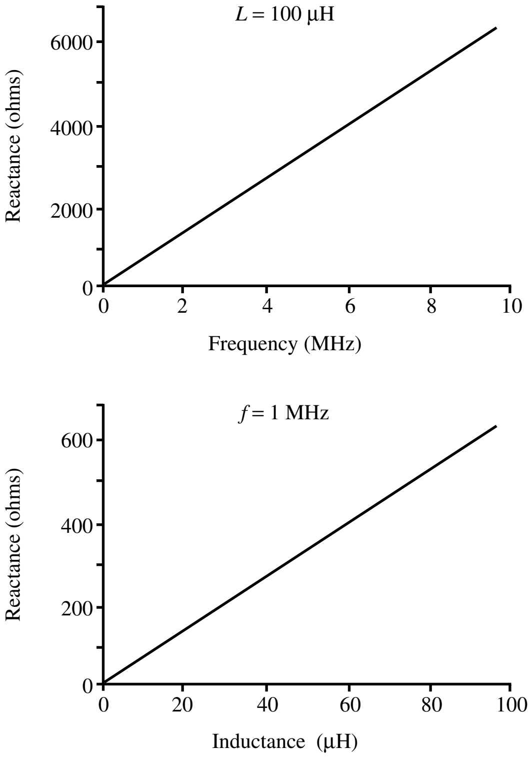

inductive reactance

inductive trimmer See TRIMMER INDUCTOR.

inductive tuning Also called *permeability tuning.* In a radio receiver, transmitter or transceiver, the adjustment of frequency by changing the inductance of a coil having a movable core.

inductivity See DIELECTRIC CONSTANT.

inductometer An instrument for measuring inductance in terms of the resonant frequency of an INDUCTANCE-CAPACITANCE (*LC*) circuit, in which *L* is the unknown inductance and *C* is calibration capacitance.

inductor A coil of wire wound according to various designs, with or without a core of ferromagnetic material, to concentrate the magnetic flux resulting from current flowing in the wire. The coiling of the wire and/or the addition of a ferromagnetic core increases the self-inductance compared with that of a straight wire having the same length. Also see INDUCTANCE; INDUCTION, **1**; and SELF-INDUCTANCE, **1**.

inductor alternator See ALTERNATOR.

inductor amplifier See MAGNETIC AMPLIFIER.

inductor decade See DECADE INDUCTOR.

inductor microphone See INDUCTIVE MICROPHONE.

inductors in parallel See PARALLEL INDUCTORS.

inductors in parallel-series See PARALLEL-SERIES INDUCTORS.

inductors in series See SERIES INDUCTORS.

inductors in series-parallel See SERIES-PARALLEL INDUCTORS.

inductor substitution box An enclosed assortment of common-value inductors that can be switched, one at a time, to a pair of terminals. In troubleshooting and circuit development, any of several useful fixed inductances can be thus obtained.

industrial data processing Abbreviation, IDP. The application of digital computers and associated equipment to industrial problems, through the classification, sorting, storing, and manipulation of information.

industrial electronics The branch of electronics concerned with manufacturing processes and their control, and with the operation and safeguarding of factories.

industrial instrumentation **1.** Supplementing an industrial process with electrical and electronic measuring instruments. **2.** The instruments used for the purpose defined in **1**.

industrial television Abbreviation, ITV. A usually closed-circuit television (CCTV) system, used as an adjunct to a manufacturing process, or as a means of communication or surveillance within an industrial plant.

industrial robot A robotic device used in industrial applications (e.g., mining, construction, manufacturing, or laboratory work).

industrial tube An (often heavy-duty) highly reliable vacuum tube designed expressly for industrial service, such as high-power radio or television broadcasting.

ineffective time The period during which an otherwise operational computer is not being used effectively because of delays or idle time.

inelastic collision A collision between charged particles in which one gains energy and the other loses energy.

inert gas A gas that does not readily react with other elements. Inert gases include *argon*, *helium*, *krypton*, *neon*, and *xenon.* Such gases are often used in hermetically sealed devices to retard corrosion.

inertance See ACOUSTIC INDUCTANCE.

inertia The tendency of a body at rest to remain at rest unless acted on by an outside force. Also, the tendency for a body in motion to maintain that motion unless acted on by an outside force. Compare MOMENTUM.

inertia in electric circuit The condition in a circuit containing inductance, in which a current change lags behind a voltage change (analogous to mechanical inertia; see INERTIA).

inertial guidance A system that automatically guides missiles and satellites in a desired trajectory without the need for continuous control by signals from a station.

inertia relay A time-delay relay whose operation is slowed by the addition of weights or other attachments.

inertia switch A switch that can sense a disturbance of its inertia.

inference engine A circuit that gives instructions to a computer or robot, by applying programmed rules to commands issued by a human operator. Comprises the functional portion of an EXPERT SYSTEM.

infinite Pertaining to a quantity or region that has no defined limits.

infinite baffle A loudspeaker baffle having no openings for the passage of sound from the front to the back of the speaker cone.

infinite-impedance detector A detector that offers the very high input impedance of a gate-source circuit and the large-signal capabilities of a diode detector. Audio-frequency output is taken across the source resistor, which is bypassed for radio-frequency signals. There is no drain resistor. Drain current increases with the input signal from a very low value at zero signal level.

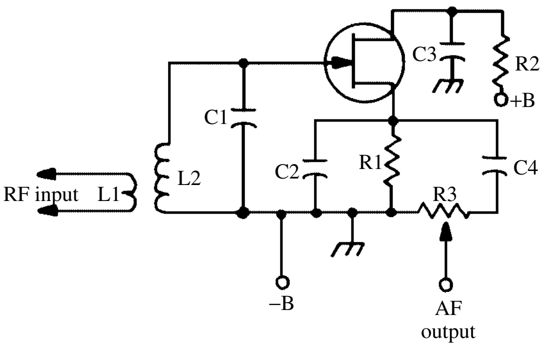

infinite-impedance detector

infinite line See INFINITE TRANSMISSION LINE.

infinite regress A reasoning pattern (either human or machine-based) that is fallacious because it defines or explains something in terms of itself.

infinite sample space In statistics, a sample space having no definite limits.

infinite series A mathematical series in which the number of terms is limitless. For example, $\frac{1}{6} = 0.1 + 0.06 + 0.006 + 0.0006 + \ldots$

infinitesimal **1.** A quantity, such as a differential, that approaches zero as the limit. **2.** Pertaining to a quantity whose magnitude is extremely small or negligible. **3.** Pertaining to an extremely small change in a quantity or measured value.

infinite transmission line A theoretical transmission line with normal characteristics, but extending away from the signal generator or receiver for a limitless distance.

infinity Symbol, ∞. A quantity that is unlimited in duration or dimension. A quantity that increases without limit is sometimes said to "approach infinity."

infix notation A system of logical operation notation wherein operands are separated by operators, thus, A & B, where the ampersand means AND. Compare PREFIX NOTATION.

Infobahn See INFORMATION SUPERHIGHWAY.

infobond On a printed circuit board, a form of wiring on the side opposite the components. The wiring is used in place of the foil normally on such a circuit board.

information **1.** Collectively, data or communications, excluding the symbols or signals used to describe, present, or store them. **2.** The result of data processing (i.e., that which is derived from the compilation, analysis, and distillation of data).

information bits In an encoded signal, data characters or digits that can be treated to give information (excluding control characters).

information center A storage bank designed for use by many different subscribers, via computer.

information channel A channel through which data and associated signals are transmitted and received.

information feedback system In message transmission, a control system in which intelligence received at a terminal is returned to the sending unit for automatic verification.

information gate A device or circuit that opens and closes an information channel.

"information processor" species Anything that uses data to derive conclusions, to produce other data, or to take specific actions, and whose functioning can be explained entirely on the basis of data-processing operations. This includes computers and smart robots. Many (but not all) scientists believe that animals are also included; some believe that human beings qualify as well.

information retrieval In digital computer and data-processing operations, the categorizing and storage of information and the automatic recall of specific file items. Also see ACCESS TIME.

information separator An indicator that separates items of information or fields in a (usually variable-length) record.

information storage In digital computer and data processing operations, holding information in memory pending retrieval.

information superhighway **1.** General expression for a worldwide network consisting of computers (personal, educational, industrial, and government) interconnected by telephone lines. **2.** See INTERNET. **3.** A massive, evolving, somewhat controversial data communication network linking computers, television, and telephone systems. It uses high-speed, high-volume data links. Communication technologies include fiberoptics, radio-frequency repeaters, microwaves, geostationary satellites, and low-earth-orbit (LEO) satellite systems.

information word A character group representing stored information and managed, as a unit, by hardware or software.

infra- Prefix meaning *below* or *lower than* (e.g., INFRARED).

infrablack region In a composite video signal, the blacker-than-black region (see BLACKER THAN BLACK).

infradyne receiver A superheterodyne receiver in which the intermediate frequency is the sum of the signal and oscillator frequencies, rather than their (usual) difference.

infrared Pertaining to electromagnetic energy in a band whose wavelength is longer than that of visible light, but shorter than that of microwave energy.

infrared communication Communication by keying or modulating infrared rays.

infrared counter-countermeasure A military tactic in which action is taken against an enemy infrared countermeasure.

infrared countermeasure A military tactic using countermeasure methods to cripple enemy infrared equipment.

infrared detector A device that senses the presence of infrared energy. Some such detectors are bolometers, radiometers, radiomicrometers, and photocells.

infrared-emitting diode Abbreviation, IRED. A semiconductor diode, such as the gallium-arsenide type, that emits infrared rays when a current passes through the p-n junction in the forward direction.

infrared guidance A navigation and reconnaissance system using infrared rays.

infrared homing The method whereby a guided missile uses infrared rays to guide it to its target.

infrared light See INFRARED RAYS.

infrared motion detector See MOTION DETECTOR and INFRARED.

infrared photography Photography in which the scene is illuminated with infrared light or emits infrared rays, and the film is infrared sensitive.

infrared radiation See INFRARED RAYS.

infrared rays Radiation at frequencies in the INFRARED region. Also (somewhat mistakenly) called *heat rays.*

infrared remote control **1.** The use of an infrared link, usually over short line-of-sight distances, for the purpose of controlling the operation of electronic equipment. A common example is the local remote control of a television receiver or high-fidelity system. **2.** A small box, containing buttons, a transmitter and an infrared-emitting diode (IRED), used for local remote control of devices, such as television receivers and high-fidelity sound systems.

infrared spectrum The region of the electromagnetic spectrum in which INFRARED radiation is found. This band lies between the microwave radio spectrum and the visible-light spectrum.

infrared therapy The use of infrared rays by physicians and other practitioners to treat certain disorders.

infrared waves Electromagnetic waves whose lengths are greater than those of visible light waves, but less than those of microwaves.

infrared window Any portion of the infrared spectrum in which energy is easily transmitted through the lower atmosphere of the earth.

infrasonic Pertaining to acoustic disturbances whose frequencies are below the range of human hearing (less than about 20 Hz).

infrasonic intrusion detector A system that detects the presence of extremely low-frequency acoustic disturbances, and sends a signal to an alarm. Such INFRASONIC waves can be caused by various actions such as walking on a wooden floor, opening or closing a door, etc.

infrasonics The branch of physics dealing with INFRASONIC phenomena.

infrasound Acoustic disturbances in the air, whose frequencies are lower than about 20 Hz, and whose wavelengths are longer than about 55 feet (17 meters).

inharmonic distortion Distortion in which the frequencies of extraneous components are not harmonically related to the fundamental frequency. It is sometimes experienced when a tone-burst signal is applied to a loudspeaker.

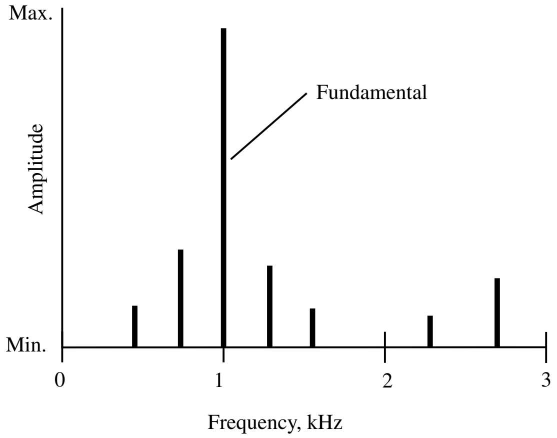

inharmonic distortion

inherent component A (usually extraneous) property possessed by a device because of its internal peculiarities. Thus, an inductor has inherent capacitance; a capacitor has inherent inductance.

inherited error In an extended calculation, an error carried through from one of the earlier steps.

inhibit **1.** In digital computer and logic operations, to prevent an action or block the input of data by means of a pulse. **2.** To delay an action or process.

inhibit gate A pulse-actuated gate circuit that acts as an INHIBITOR.

inhibitor **1.** A device or circuit that produces a pulse or signal that prevents an action, or that blocks data input. **2.** An additive, such as an organic liquid, that delays the hardening of a mixture, such as an encapsulating compound.

inhibit pulse In a computer, a drive pulse that prevents other pulses from changing the direction of magnetization in the cells of a magnetic core memory.

inhibit signal In digital computer and logic operations, the signal that causes an INHIBIT action.

initial drain **1.** The current supplied by a battery or cell at its rated voltage. **2.** The current delivered by a rechargeable battery or cell when it is put to use immediately after receiving a full charge.

initial failure The first failure occurring in the operation of a circuit or device.

initial instructions A resident computer routine used to aid program loading. Also called *initial order.*

initial ionizing event In the operation of a radioactivity counter, the first event that starts the chain of similar events constituting the count.

initialization A computer program instruction that sets the value of a variable to zero.

initial permeability Permeability in the low magnetization region of a material.

initial time delay Abbreviation, ITD. In acoustics, the elapsed time between the instant the direct sound wave is first heard, and the instant the first echoes (reflected sound waves) arrive.

initiate See TRIGGER.

injection **1.** Introducing a signal into a circuit or device. **2.** Introducing charge carriers (electrons or holes) into a semiconductor.

injector **1.** An element or electrode for INJECTION. **2.** A device or circuit that injects a signal into another device or circuit.

injector electrode See INJECTOR, **1**.

ink bleed In the printing of matter for optical character recognition, ink flow around the characters, often making them unrecognizable to the reader.

inkjet galvanometer A galvanometer whose movement controls the pressure of a jet of ink for making a recording on a paper chart. Also see LIQUID-JET OSCILLOGRAPH.

inkjet printer A printer commonly used with personal computers, in which images are created by jets of ink sprayed directly onto the paper. Noted for low operating noise level, high image resolution, and excellent color-reproduction capability.

ink-mist recorder A graphic recorder in which the line is traced by a mist of ink.

ink recorder A graphic recorder using a pen-and-ink stylus.

ink squeeze-out In the printing of matter for optical character recognition, the squeezing of ink from a character's center.

ink-vapor recorder See INK-MIST RECORDER.

in-lb Abbreviation of INCH-POUND.

inlead The part of an electrode that passes through the external shell or case of a component.

inline procedure The main portion of a COBOL computer program, responsible for the primary operations.

inline processing The action peculiar to a system that processes data almost immediately upon receipt (i.e., one that need not be capable of storing a lot of unprocessed data).

inline readout In digital computer operations, a readout device that displays digits side-by-side horizontally.

inline subroutine A subroutine that must be written each time it is needed, as compared with one that can be accessed by a program branch.

inline tuning Tuning of all the stages of a channel, such as an intermediate-frequency amplifier, to the same frequency.

inner conductor The inner wire or rod of a coaxial cable or coaxial tank. It generally carries the signal, and is isolated from the surrounding environment by the grounded OUTER CONDUCTOR.

inorganic Consisting of materials other than carbon compounds; therefore, it is not related to living things.

inorganic electrolyte Any electrolyte that is completely inorganic: containing no compounds of carbon.

in phase The condition in which alternating or pulsating waves or wave phenomena are in step with each other at all points. Compare OUT-OF-PHASE.

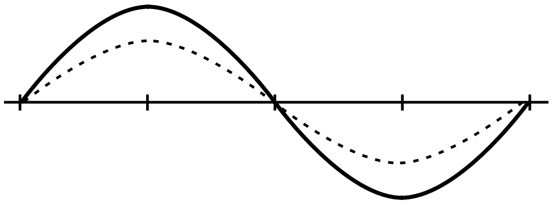

in phase

in-phase carrier See I-PHASE CARRIER.

in-phase current Resistive current in an ac circuit (i.e., current in phase with voltage). Compare QUADRATURE CURRENT.

in-phase feedback Feedback in phase with a main signal. Also called POSITIVE FEEDBACK and REGENERATION.

in-phase voltage A voltage that is in phase with another (reference) voltage.

in-plant system An automatic data communications system within a specific building or complex.

input **1.** Energy or information delivered or transferred to a circuit or device. **2.** The terminals of a device or circuit to which energy or information is applied. **3.** To deliver or transfer energy or information to a circuit or device (as to input data from a computer peripheral to memory).

input admittance Symbol, Y_i. The internal admittance of a circuit or device, as "seen" from the input terminals; the reciprocal of input impedance. Compare OUTPUT ADMITTANCE.

input area In a computer memory, an area set aside for data input from a source other than a program.

input bias current The input bias required by an operational amplifier.

input capacitance Symbol, C_i. **1.** The internal capacitance of a circuit or device, as "seen" from the input terminals. Compare OUTPUT CAPACITANCE. **2.** The grid-cathode capacitance of a vacuum tube.

input capacitor **1.** In a capacitance-coupled circuit, the input coupling capacitor. Compare OUTPUT CAPACITOR. **2.** The first capacitor in a capacitor-input filter (i.e., that capacitor electrically nearest the rectifier output electrode).

input choke The first choke in a choke-input filter (i.e., that choke electrically nearest the rectifier output electrode, when no preceding capacitor is used).

input circuit The circuit or subcircuit constituting the input section of a network or device. Compare OUTPUT CIRCUIT.

input clamp current The current from an input when the input is in a state below ground potential. A test for the input clamp diode.

input conductance Symbol, G_i. The internal conductance of a circuit or device, as "seen" from the input terminals; it is the reciprocal of INPUT RESISTANCE. Compare OUTPUT CONDUCTANCE.

input coupling capacitor See INPUT CAPACITOR, **1**.

input coupling transformer See INPUT TRANSFORMER.

input current Symbol, I_i. **1.** The current delivered to a circuit or device. **2.** Current flowing in the input leg or electrode of a circuit or device.

input device **1.** A device, such as an input transformer, that couples energy or information to a circuit or device. Compare OUTPUT DEVICE. **2.** A device through which another device receives data.

input equipment Collectively, input devices used with a computer.

input error voltage In an operational amplifier, the error voltage at the input terminals when a feedback loop operates around the amplifier.

input extender A diode network that provides increased fan-in for a logic circuit. Also see FAN-IN, **1**.

input gap In a velocity-modulated tube, the gap in which the electron stream is initially modulated.

input guarding A method of eliminating stray coupling among inputs in an integrated circuit. A shield is provided at the input; it is driven to follow along with the input voltage. This ensures low loss and minimum errors resulting from unwanted coupling.

input impedance Symbol, Z_i. The internal impedance of a circuit or device, as "seen" from the input terminals. Compare OUTPUT IMPEDANCE.

input limited The processing time limitation imposed by an input unit on the speed of a program run.

input noise current At the input of an integrated circuit, the root-mean-square (rms) or peak-to-peak (pk-pk) noise current existing within a specified range of frequencies.

input noise current density The noise current, usually expressed as a root-mean-square (rms) value, in a band 1 Hz wide around a given frequency.

input noise voltage At the input of an integrated circuit, the root-mean-square (rms) or peak-to-peak (pk-pk) noise voltage existing within a specified range of frequencies.

input noise voltage density The noise voltage, usually expressed as a root-mean-square (rms) value, in a band 1 Hz wide around a given frequency.

input offset current In an operational amplifier, the difference between the currents going to the input terminals when the output is zero.

input offset voltage In an operational amplifier, the potential that has to be applied between the input terminals for a zero output voltage.

input/output Abbreviation, I/O. **1.** Data transmitted to, or received from, a computer. **2.** A terminal through which data is transmitted to, or received from, a device.

input/output bound A condition affecting a system in which the time consumed by input and output operations is greater than that required for other processes.

input/output buffer A computer memory area specifically reserved for the receipt of data coming from or going to a peripheral.

input/output control The part of a computer system that coordinates activity between a central processor and peripherals.

input/output equipment **1.** In digital computer operations, devices for entering information into the computer or for reading information from it. Examples: keyboard, mouse, display, and optical scanner. **2.** In robotics and artificial intelligence, a data link between a controller and one or more robots, and/or between or among two or more controllers.

input/output isolation Arrangement or operation of a circuit or device so that there is no direct path between input and output terminals around the circuit or device. Also see ISOLATION.

input/output module See INPUT/OUTPUT EQUIPMENT.

input/output routine A routine for simplifying the programming of standard input/output equipment operations.

input/output switching The allocation of more than one channel to peripherals for communications with a central processor.

input/output voltage differential At a given load current, the potential difference that is necessary for an integrated circuit to operate according to its output voltage specifications.

input power Symbol, P_i. **1.** The power presented to the input terminals of a circuit or device. Also called POWER INPUT. Compare OUTPUT POWER. **2.** The operating power of a circuit or device (i.e., the power-supply requirement).

input protection In an integrated circuit, a means of preventing damage to the device from excessive voltage at the input, such as transient spikes or the result of malfunctioning of some other circuit.

input record **1.** A computer record of immediate interest that is ready for processing. **2.** During a computer program run, a record read into memory from an input device.

input recorder A device that makes a permanent record of the signals or data input to a circuit or system.

input register In a computer, a register that receives data from a peripheral relatively slowly and then passes it on to a central processor at a faster speed as a sequence of informational units. Also see REGISTER.

input resistance Symbol, R_i. The internal resistance of a circuit or device, as "seen" from the input terminals. Compare OUTPUT RESISTANCE.

input resonator In a velocity/modulated tube, the resonator in which electron bunching occurs.

input routine A computer program section that manages data transferal between an external storage medium and a memory input area.

input section **1.** See INPUT ROUTINE. **2.** See INPUT AREA.

input sensitivity **1.** The level of input-signal amplitude that will result in a certain signal-to-noise ratio at the output of a device. The specified signal-to-noise ratio is usually 10 or 20 dB. **2.** The level of input signal in a frequency-modulated device, required to produce a specified amount of noise quieting. The specified level of noise quieting is usually 20 dB. Alternatively, 12-dB SINAD (ratio of signal to the level of noise and distortion) can be specified. **3.** The minimum level of input voltage required to actuate a logic gate.

input signal The signal (current, voltage, and power) presented to the input terminals of a circuit or device for processing.

input tank In a double-tuned stage of a transmitter or power generator, the tank circuit in which the input signal is resonated. This is generally the base or gate circuit. Compare OUTPUT TANK.

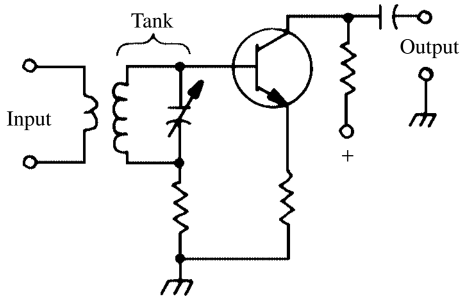

input tank

input terminals Terminals (usually a pair) associated with the input section of a circuit or device. Compare OUTPUT TERMINALS.

input transformer The transformer that delivers signal voltage or power to the input circuit of a network or device. Compare OUTPUT TRANSFORMER.

input uncertainty The combination of all parameters that result in adverse behavior in an operational amplifier.

input unit In a digital computer, the device or circuit that receives information from peripherals.

input voltage **1.** Symbol, E_i or V_i. The voltage presented to a circuit or device. Compare OUTPUT VOLTAGE, **1**. **2.** The voltage across the input leg or electrode of a circuit or device. Compare OUTPUT VOLTAGE, **2**.

input-voltage drift For an integrated circuit (IC), the time- and temperature-dependent change in output voltage divided by the IC's open-loop voltage gain.

input-voltage offset For a differential amplifier, the input signal voltage at the differential input that results in zero output voltage.

input-voltage range The range, in volts, over which the input voltage can fluctuate in an integrated circuit so that the common-mode rejection ratio (CMRR) specifications are not exceeded.

input winding The signal winding of a magnetic amplifier.

inquiry A programmed request for information from storage in a computer.

inquiry display terminal A video display/keyboard terminal used to make an inquiry to a computer system, and display the response.

inquiry station A terminal from which an inquiry can be sent to a central computer.

inrush The initial surge of current that occurs when voltage is first applied to the primary winding of a transformer with no load connected.

inscribe To convert data to a form on a document that is readable by a character-recognition device, as through the use of magnetic ink, for example.

insect robot A member of a fleet of robots, all of which are under the control of a single computer.

The term arises because the system functions like an anthill or beehive, in which the individual machines are "stupid," but the system as a whole is "smart." Such robots often have six legs, like insects. Compare AUTONOMOUS ROBOT.

insert A (usually metallic) bushing that can be molded into a plastic part (or pressed into it after molding is completed) to provide a bearing sleeve or threaded hole.

insert core A ferromagnetic core whose position can be adjusted to vary the inductance of the coil surrounding it.

insert edit **1.** In magnetic tape recording, a section of tape on which new audio is recorded over existing audio. **2.** The process of recording new audio over existing audio in a defined interval on a magnetic tape.

insertion gain In a circuit or system, the gain resulting from the amplifier inserted into the system; it is usually expressed in decibels. Compare INSERTION LOSS.

insertion loss Loss of energy or gain by placing certain devices or subcircuits (filters, impedance matchers, etc.) in a circuit. It is usually expressed in decibels. Also see INSERTION RESISTANCE.

insertion phase shift The difference in phase produced by a circuit installed in an electrical transmission line.

insertion resistance The resistance of a component or instrument that is introduced into a circuit. Thus, the internal resistance of a microammeter becomes an insertion resistance in the circuit in which the meter is connected for current measurement.

inside antenna See INDOOR ANTENNA.

inside diameter Abbreviation, ID. The innermost diameter of a body or figure having two concentric diameters. Compare OUTSIDE DIAMETER.

inside lead See START LEAD.

inside radiation See INDOOR RADIATION.

inside spider A voice-coil centering device within a loudspeaker.

inst **1.** Abbreviation of INSTRUMENT or INSTRUMENTATION. **2.** Abbreviation of INSTANT.

instability Inconsistency in the operation of a circuit or device, in the parameters of a device, or in an electrical quantity. It can be attributed to a number of causes, including temperature, loading, age, humidity, negative resistance, and radioactivity.

installation tape number An identification number given to a reel of magnetic tape by the processing facility.

instant Abbreviation, inst. The point in time at which an event occurs, or at which a quantity reaches a particular value.

instantaneous Occurring at a specified moment, or instant, of time.

instantaneous amplitude The amplitude, specified in amperes, volts, or watts, of a signal, specified at a particular moment in time.

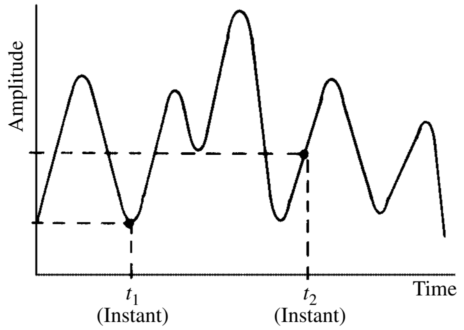

instantaneous amplitude

instantaneous automatic gain control Abbreviation, IAGC. An automatic gain control whose operation almost immediately follows a change in signal amplitude.

instantaneous automatic volume control Abbreviation, IAVC. An instantaneous automatic gain control system for the immediate control of volume in receivers and audio-frequency amplifiers.

instantaneous companding A form of companding that operates according to the instantaneous amplitude of the input signal.

instantaneous contacts Timer contacts that open or close almost immediately upon application of the control signal.

instantaneous current Symbol, *i* or I_i. The value of an alternating or fluctuating current at a particular instant in the cycle.

instantaneous disc A phonograph disc that can be played back immediately after being recorded.

instantaneous frequency The frequency of a signal at a particular moment in time. The instantaneous frequency changes in frequency-modulated or phase-modulated signals.

instantaneous power **1.** In a single-sideband, suppressed-carrier signal, the power at a specified instant in time. It varies between zero and the peak envelope power (PEP). **2.** The output power of an audio amplifier at a specified instant in time.

instantaneous power output The rate of power delivery to a load at a given instant.

instantaneous relay A relay, such as a fully electronic type (having no moving parts), that shows virtually no delay in its operation.

instantaneous sample A measurement obtained by INSTANTANEOUS SAMPLING.

instantaneous sampling The measurement of wave or signal amplitude at a specific moment in time. See, for example, INSTANTANEOUS CURRENT and INSTANTANEOUS VOLTAGE.

instantaneous speech power In the output of an audio amplifier, the instantaneous value of power in a speech wave, as opposed to that in a sine wave. Also see INSTANTANEOUS VALUE and SPEECH POWER, **1**.

instantaneous value The magnitude of a fluctuating value at a selected instant in time. See, for example, INSTANTANEOUS CURRENT, INSTANTANEOUS POWER, INSTANTANEOUS SPEECH POWER, and INSTANTANEOUS VOLTAGE. Compare AVERAGE VALUE, and EFFECTIVE VALUE.

instantaneous voltage Symbol, e or E_i. The value of an alternating or fluctuating voltage at a particular instant in the cycle.

instant loop In electronic security applications, a circuit that actuates an alarm without delay when an intrusion is detected.

instruction In digital computer operations, a set of bits defining an operation. Consists of an operation code specifying the operation to be performed, one or more operands or their addresses, and one or more modifiers or their addresses (to modify the operand or its address).

instruction address In a computer memory, the address of a location containing an instruction.

instruction address register Also called *program counter.* A register that holds instruction addresses so that the retrieval of the instructions from memory can be controlled during a program run.

instruction code Also called INSTRUCTION SET. The symbols and characters that compose the syntax of a computer programming language.

instruction format In a computer's basic machine code, the part that specifies how characters or digits are used to represent the codes within the machine's instruction set.

instruction modification In a computer instruction, a change in the instruction code that makes the computer do a different operation when the routine containing the code is encountered again.

instruction register A register in a computer containing the address of the current instruction. Also called CONTROL REGISTER (abbreviation, CR).

instruction set **1.** The range of commands that form a programming language. **2.** See INSTRUCTION CODE.

instruction storage A memory circuit that stores computer instructions or programs.

instruction time The time required for a control unit to analyze and implement a computer program instruction.

instruction word In digital computer programming, a word containing the instruction code (type of operation to be performed) and the address part (location of the associated data in storage).

instrument A device for measuring electrical quantities or the performance of electronic equipment. A *meter* provides a direct indication; other devices, such as a *bridge*, must be adjusted, the measured quantities being determined from one or more adjustments (sometimes augmented with calculations).

instrumental error See INSTRUMENT ERROR.

instrument amplifier Also called INSTRUMENTATION AMPLIFIER. A high-gain, wideband amplifier that increases the sensitivity of an instrument (such as an oscilloscope, meter, or graphic recorder).

instrument-approach system See INSTRUMENT LANDING SYSTEM.

instrumentation Planning and providing instruments and instrument systems for the collection and, sometimes, storage and analysis of data.

instrumentation amplifier **1.** A form of integrated-circuit voltage amplifier designed for high linearity, high input impedance, and high common-mode rejection. It is intended for use with electronic instruments. **2.** See INSTRUMENT AMPLIFIER.

instrument chopper A refined chopper for converting a direct-current (dc) signal to alternating current (ac) for an ac instrument, such as a voltmeter or recorder.

instrument error Discrepancy in measured quantities resulting from inaccuracy of the instrument used, insertion resistance, environmental factors, operator error, etc.

instrument flight Also called *blind flight.* Aircraft flight guided by navigational instruments and signals alone. Required when visibility is extremely poor.

instrument fuse A fast-acting, low-current fuse used to protect a sensitive instrument, such as a galvanometer, milliammeter, and/or microammeter.

instrument lamp A light or lamp that illuminates the face of an instrument to facilitate viewing in the dark.

instrument landing Also called *blind landing.* Aircraft landing guided entirely by instruments. Required when visibility is poor and when landing is imperative at a given location at a given time.

instrument landing station The radio or radar station in a blind-landing system (see INSTRUMENT LANDING SYSTEM).

instrument landing system Abbreviation, ILS. The complete instrument and signal system (on the ground or in aircraft) required for an INSTRUMENT LANDING.

instrument multiplier See MULTIPLIER PROBE, **1**.

instrument preamplifier An external, sensitive amplifier for an instrument that has an internal input amplifier. Also see INSTRUMENT AMPLIFIER.

instrument relay See METER RELAY.

instrument resistance See METER RESISTANCE.

instrument shunt A resistance connected in parallel with a current-measuring instrument, used to increase the range of currents that can be measured.

instrument transformer A transformer used to change the range of an alternating-current meter.

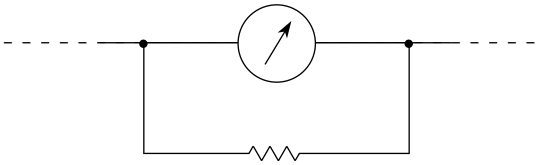

instrument shunt

For ammeters, it is called a *current transformer*; for voltmeters, it is called a *potential transformer.*

insulant A nonconducting material, used to prevent the flow of electric current between or among points. See INSULATOR, **1**.

insulated Isolated from conductors by an INSULANT.

insulated-gate field-effect transistor Abbreviation, IGFET. See METAL-OXIDE SILICON FET.

insulated resistor A resistor around which is molded a nonconducting material, such as vitreous enamel or a plastic.

insulating tape Electrical insulation in the form of a thin, usually adhesive, strip of fabric, paper, or plastic.

insulation **1.** A coating of dielectric material that prevents a short circuit between a conductor and the surrounding environment. **2.** The application of a dielectric coating to an electrical conductor. **3.** Electrical separation between or among different components, circuits, or systems.

insulation breakdown Current leakage through, and rupture of, an insulating material because of high-voltage stress.

insulation ratings Collectively, the dielectric constant, dielectric strength, power factor, and resistivity of an insulating material. Sometimes included are such physical properties as rupture strength, melting point, etc.

insulation resistance The very high resistance exhibited by a good insulating material. It is expressed in megohms (or higher units of resistance) for a sample of material of stated volume or area.

insulation system Collectively, the materials needed to insulate a given electronic device.

insulator **1.** A material that, ideally, conducts no electricity; it can, therefore, be used for isolation and protection of energized circuits and components (also see DIELECTRIC). Actually, no insulator is perfectly nonconductive (see, for example, INSULATION RESISTANCE). **2.** A molded piece of solid insulating material, used to electrically isolate conductors—especially in antenna systems and power transmission lines. **3.** Any body made from an insulating material.

insulator arcover A sudden arc, or flow of current, over the surface of an insulator, because of excessive voltage.

integer A positive or negative whole number, as opposed to a fraction or mixed number.

integral **1.** Also called *indefinite integral* and *antiderivative.* For given mathematical function *f*, function *g*, whose derivative is equal to *f*. **2.** Also called *definite integral.* The area under a curve of a function, between two vertical lines defined by two specific points in the domain of the function. **3.** The part of a number to the left of the radix point. **4.** Pertaining to integers (positive or negative whole numbers) or quantities that can be represented by integers.

integral action In automatic control operations, a control action delivering a corrective signal proportional to the time that the controlled quantity has differed from a desired value.

integral contact In a relay or switch, a contact that carries current to be switched.

integral-horsepower motor A motor rated at one horsepower.

integral multiple A whole multiple of a number. Thus, a harmonic is an integral multiple of fundamental frequency *f*: 2*f*, 3*f*, 4*f*, etc.

integral number See INTEGER.

integrate **1.** To perform the function of mathematical or electrical INTEGRATION. **2.** To construct a circuit on a piece of semiconductor material.

integrated Constructed on a single piece of material, such as a semiconductor wafer.

integrated amplifier An audio-frequency (AF) amplifier having a preamplifier, intermediate amplifier, and output amplifier on a single chassis.

integrated capacitor In an integrated circuit, a fixed capacitor in which one plate is a layer of material diffused into the substrate, the dielectric is a thin-oxide film grown on top of the first layer, and the other plate is a metal layer deposited on top of the oxide film.

integrated circuit Abbreviation, IC. A circuit whose components and connecting "wires" are made by processing distinct areas of a chip of semiconductor material, such as silicon. Classified according to construction (e.g., *monolithic IC*, *thin-film IC*, *hybrid IC*).

integrated data processing Abbreviation, IDP. The detailed electronic classification, sorting, storage, and mathematical processing of data within a coordinated system of equipment, usually at one location.

integrated electronics The branch of electronics that is concerned with the design and fabrication of integrated circuits.

integrated resistor See DIFFUSED-LAYER RESISTOR.

Integrated Services Digital Network Abbreviation, ISDN. A communications network or connection intended primarily for Internet access via telephone lines. Allows significantly higher data speed than is possible with a conventional analog connection. In addition, it is possible to use a digital system, such as a computer, online simultaneously with an analog voice conversation.

integrating circuit See INTEGRATING NETWORK.

integrating galvanometer A device for measuring the change in electric flux produced in a coil in an electric field. Even very slow changes can be measured.

integrating meter An instrument whose indication is a summation (usually) of an electrical quantity that is time-dependent (e.g., *ampere-hour meter* and *watt-hour meter*).

integrating motor An electric motor that follows the integral of the input signal. The angle of rotation of the motor shaft is equal to the integral of the input waveform.

integrating network A four-terminal network whose output voltage is proportional to the time integral of the input voltage. It can be a passive resistance-capacitance (RC) circuit or it can use an operational amplifier. Compare DIFFERENTIATING NETWORK.

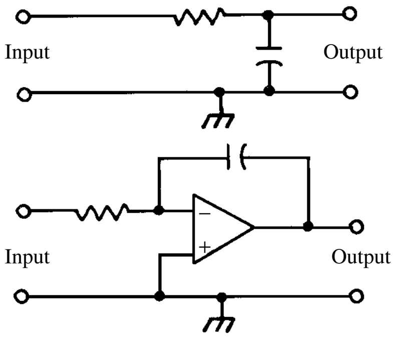

integrating networks

integrating photometer A photometer whose reading is the average candlepower at all angles in one plane.

integration **1.** The process of determining a mathematical function when its derivative is given. **2.** The processing of a signal by an INTEGRATOR circuit. **3.** Collectively, the processes by which an INTEGRATED CIRCUIT is manufactured.

integrator **1.** See INTEGRATING NETWORK. **2.** A device having an output variable whose value is proportional to the integral of one variable, with respect to another, or is proportional to the integral of an input variable, with respect to elapsed time.

intelligence **1.** Meaningful data that modulates a carrier [e.g., the voice or music in a frequency-modulated (FM) radio signal, or the image in a television signal]. **2.** Also called *machine intelligence.* The quality of a system or device, especially a computer, that allows it to “learn” (i.e., to better its capability by repeatedly operating on a given problem).

intelligence bandwidth **1.** The bandwidth necessary to convey a specified amount of data within a certain period of time. **2.** The total bandwidth of one complete signal channel in a communications or broadcast system.

intelligence signal **1.** A signal that conveys data or information. **2.** The modulating waveform in a communications or broadcast transmission.

intelligent network Abbreviation, IN. **1.** In general, any advanced and sophisticated communications network, particularly a broadband digital network. **2.** A network designed to readily accommodate new technologies and services, such as videoconferencing, interactive television, or remote control.

intelligent terminal A computer terminal (e.g., an input/output video display/keyboard unit) that through its circuitry (i.e., by use of a microprocessor) has some data-processing ability.

intelligibility tests Tests that measure the coherence of electronically reproduced speech.

intensification of image See IMAGE INTENSIFICATION.

intensifying ring In some electrostatic cathode-ray tubes, an internal metal ring serving as an extra anode to accelerate the beam and, thus, brighten the image.

intensity The degree or extent of a phenomenon (such as amplitude, brightness, loudness, power, force, etc.).

intensity control In an oscilloscope circuit, the potentiometer that adjusts the direct-current voltage on the control electrode of the cathode-ray tube and, accordingly, the brightness of the image. Also called BRIGHTNESS CONTROL and BRILLIANCE CONTROL.

intensity level **1.** A measure of sound magnitude, expressed in decibels, with respect to a value of one microwatt per square centimeter (10^{-6} W/cm^2) at sea level in the atmosphere. **2.** The setting of the brightness control in a cathode-ray-tube device.

intensity modulation **1.** Modulation of electron-beam intensity in a cathode-ray tube. Also called *z-axis modulation.* **2.** Sometimes, the video-signal modulation in a television image.

intensity-modulation amplifier The z-axis amplifier in an oscilloscope. Also see INTENSITY MODULATION.

interaction The (sometimes mutual) influence of one circuit or device on the behavior of another, as in induction.

interactive display A computer display device with which its operator can supply data to the computer in response to what is displayed. Example: *touch screen.*

interactive graphics A computer graphics system using a cathode-ray tube to draw or modify three-dimensional representations.

interactive photovoltaic system A solar-power plant that operates in conjunction with the utility companies. Energy is sold to the companies during times of daylight and minimum usage, and is bought back from the companies at night or during times of heavy usage. The principal advantage of this system is that the user can keep using electricity (by buying it all from the utilities) if the solar-energy system breaks down. But such a system does not provide the independence from utility companies that some users desire. Compare STAND-ALONE PHOTOVOLTAIC SYSTEM.

interactive program A computer program in which the machine and its operator engage in two-way communication. Most personal computing software is of this type, in contrast to programs that carry out all their functions without operator intervention (other than initialization).

interactive television Television provided to consumers, in which viewers can transmit data, as well as receive it. For example, a survey might be conducted in which viewers are polled and send in their responses. Another example: products might be ordered while viewing an advertisement.

interactive mode See CONVERSATIONAL MODE.

interbase resistance The internal resistance between the bases of a unijunction transistor.

interblock A part of a computer program or a hardware device that will prevent interference between parts of a computer system.

interblock space **1.** On a magnetic tape, the space between recordings, caused by starting and stopping the tape. **2.** On magnetic tape used as a computer storage medium, the interval between recorded blocks.

intercarrier receiver A television (TV) receiver circuit in which video, sound, and sync components of the composite TV signal are amplified together in the radio-frequency (RF), intermediate-frequency (IF), and video IF stages; then they are separated in the video detector and video amplifier stages. Compare SPLIT-SOUND RECEIVER.

intercept receiver In military service, a search receiver tuned over a wide band of frequencies to locate and evaluate enemy signals.

interchangeability The ability of one component to substitute directly for another component of the same kind. Example: capacitor interchangeability, transistor interchangeability. Also see REPLACEMENT.

intercharacter space The three-unit interval between letter symbols in telegraphy. Compare INTERWORD SPACE.

intercom A comparatively simple two-way telephone or low-power radio system for use on the premises of a home or business.

intercommunicator See INTERCOM.

interconnection **1.** A mutual connection of separate circuits. **2.** The interconnection of two or more separate power-generating systems.

interdigital contacts A pair of contacts with “fingers” that are plated, printed, or deposited on the surface of a resistor material or semiconductor substrate. The fingers of each contact are interconnected at one end, the fingers of one contact being interleaved with those of the other.

interdigital tube A magnetron having a cathode surrounded by anode segments that are alternately interconnected at opposite ends in the manner of INTERDIGITAL CONTACTS.

interelectrode capacitance Capacitance between or among electrodes—especially between the plate and control grid of a vacuum tube.

interelement capacitance Internal pn-junction capacitance in a semiconductor device, such as a diode or transistor.

interface **1.** The circuitry that interconnects and provides compatibility between a central processor and peripherals in a computer system. **2.** Collectively, the hardware and software that allows a computer to interact with its operator. **3.** To provide an efficient pathway for data between two devices or systems. **4.** The meeting of surfaces or regions in a material. **5.** The surface of a body that mates with another body similar or identical to it.

interface resistance See CATHODE INTERFACE.

interface routine A computer program routine that links one system to another.

interfacial connection A connection that runs through a printed-circuit board and joins circuit joints on opposite faces of the board.

interference **1.** In communications, degradation of reception caused by noise or undesired signals. **2.** The interaction of acoustic or electromagnetic waves from more than one source, especially when they are of the same frequency, producing a characteristic INTERFERENCE PATTERN of high-amplitude and low-amplitude regions.

interference attenuator A device or mode of operation that reduces the amplitude of interference.

interference eliminator A filter, wavetrap, or similar device that removes interfering signals or noise. Also see INTERFERENCE.

interference filter See INTERFERENCE ELIMINATOR.

interference pattern A regular pattern of high-amplitude and low-amplitude regions, lobes, or bands, produced when waves of identical frequency from two or more sources combine in varying phase. Such patterns can be observed with sound, radio waves, infrared, visible light, ultraviolet, X rays, and gamma rays. The phenomenon is of interest in acoustic engineering, the design of radio antenna systems, and in physics (particularly optics).

interference stub A length of twin-lead feeder cut to appropriate length, connected to the antenna-input terminals of a television receiver, and short-circuited at the opposite end. A stub of the correct length resonates at the frequency of an interfering signal and, acting as a wavetrap, keeps it out of the receiver. Also see STUB.

interference trap A wavetrap that suppresses interfering signals at the rejection frequency of the trap.

interferometer **1.** A radio telescope having two antennas spaced at a distance of many wavelengths, providing much greater resolution than a single antenna. Pioneered by M. Ryle of England and J.L. Pawsey of Australia. **2.** Any device that displays an INTERFERENCE PATTERN for testing or experimental purposes.

interfix A method used in information-retrieval systems that eliminates ambiguity in the responses to inquiries by describing the relationship between keywords in a record.

Interframe A method of digital IMAGE COMPRESSION developed by *MPEG (Moving Picture Experts Group).* It operates by eliminating redundant data from between image frames. Compare INTRAFRAME.

interharmonic beats Beat notes produced by various combinations of the harmonics of a signal.

interim storage See TEMPORARY STORAGE.

interior label On a magnetic tape used as a computer-storage medium, a label recorded at the beginning of the tape. Compare EXTERIOR LABEL.

interior protection **1.** In electronic security applications, a set of sensors contained entirely within the region to be protected. **2.** The installation and operation of a security system whose sensors are all within the region to be protected.

interlace A form of data storage in which portions of the data are stored in alternate locations in the tape or disk.

interlaced field A video image field produced by INTERLACED SCANNING.

interlaced scanning In the display of a video image, the alternate presentation of the even- and odd-line fields. This process increases the obtainable image resolution for a given refresh rate, but can result in "jerkiness" of the image when rapid motion is portrayed.

interlace factor A number expressing the extent to which two fields are interlaced. Also see INTERLACED SCANNING.

interleaving In multiprogramming, the inclusion in a program of segments of another program so that both can be effectively executed simultaneously.

interlock switch See ELECTRICAL INTERLOCK.

intermediate amplifier See BUFFER.

intermediate frequency Abbreviation, IF. In a superheterodyne circuit, the frequency of the signal that results from beating the incoming signal with the signal produced by the local oscillator.

intermediate-frequency amplifier In a superheterodyne circuit, the fixed-frequency amplifier that boosts the intermediate-frequency signal. Also see INTERMEDIATE FREQUENCY.

intermediate-frequency channel Usually, the intermediate-frequency amplifier in a superheterodyne circuit, but sometimes including the second detector, automatic gain control (AGC), and oscillator stages.

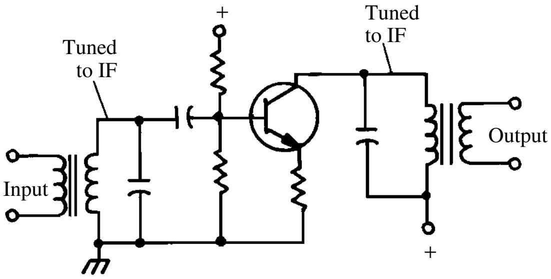

intermediate-frequency amplifier

intermediate-frequency converter See IF CONVERTER.

intermediate-frequency interference Interference from signals at the intermediate frequency of a receiver or instrument.

intermediate-frequency selectivity The selectivity of an intermediate-frequency (IF) channel alone, usually determined by the characteristics of the bandpass filter(s) in the IF channel.

intermediate-frequency transformer A coupling transformer designed for use in an intermediate-frequency amplifier.

intermediate-puck drive In a tape recorder, a speed-reducing drive system in which an intermediate wheel conveys motion from the motor shaft to the rim of the flywheel.

intermediate puck wheel See IDLER WHEEL.

intermediate repeater In wire telephony, a repeater inserted into a line or trunk at some point other than the end.

intermediate result Obtained during a program run or the execution of a subroutine, a result that is used again as an operand in deriving the final result.

intermediate section Any of the internal sections of a multisection filter. Thus, the middle section of a three-section filter.

intermediate storage In a computer system, a storage medium for temporarily holding totals or working figures. Also called WORK AREA.

intermediate subcarrier A modulated or unmodulated subcarrier that modulates either a carrier or another intermediate subcarrier.

intermittent **1.** Pertaining to a circuit fault, such as an open or short circuit, that occurs some of the time, but not all the time. **2.** See INTERMITTENT CONDITION. **3.** Pertaining to a phenomenon that is observed some of the time, but not all the time; sporadic. **4.** Pertaining to a DUTY CYCLE greater than zero, but less than 100 percent; usually between 25 and 50 percent.

intermittent commercial and amateur service Abbreviation, ICAS. Operation of equipment, such as radio transmitters, for short, irregular periods, as in amateur (hobbyist) activity or infrequent commercial service. ICAS ratings are higher than continuous commercial service (CCS) ratings. Compare CONTINUOUS COMMERCIAL SERVICE.

intermittent condition A defect in a circuit or device that causes erratic and unreliable operation. The cause of such a problem is often difficult to determine.

intermittent dc See INTERMITTENT DIRECT CURRENT.

intermittent direct current A regularly pulsed unidirectional current. Also called PULSATING DIRECT CURRENT.

intermittent duty A DUTY CYCLE of less than 100 percent, but greater than zero. Generally, an operating duty cycle of 25 to 50 percent.

intermittent-duty rating The dissipation or power rating of a component, circuit, or system, under conditions of intermittent use, usually a 25-percent to 50-percent DUTY CYCLE.

intermittent operation Operation characterized by often long nonoperating intervals. Intermittent operation is often random, whereas on-off operation tends to be regular.

intermittent signal An interrupted signal resulting from the intermittent operation of a circuit or device.

intermodulation Abbreviation, IM. **1.** The (usually undesired) modulation of one signal by another, caused by nonlinear processing of the signals. **2.** The heterodyning of components in the sidebands produced by an amplitude-modulated (AM) or single-sideband (SSB) transmitter.

intermodulation distortion Abbreviation, IMD. **1.** Distortion products in the output of an amplitude-modulated (AM) or single-sideband (SSB) transmitter, caused by heterodyning of components in the sidebands. **2.** Distortion products in the output of an audio amplifier, caused by heterodyning of the fundamental components. **3.** The extent to which distortion as defined in 1 occurs. See INTERMODULATION-DISTORTION PERCENTAGE.

intermodulation-distortion meter See INTERMODULATION METER.

intermodulation-distortion percentage Abbreviation, IDP. The degree to which a low-frequency test signal modulates a higher-frequency test signal when both are applied simultaneously (in a prescribed amplitude ratio) to a device under test; $IDP = 100(b - a)/a$, where a is the peak-to-peak amplitude of the unmodulated high-frequency wave and b is the peak-to-peak amplitude of the modulated high-frequency wave.

intermodulation meter An instrument for measuring percentage of intermodulation distortion (IMD). The instrument combines a dual-frequency signal generator, filter circuits, and percent-of-modulation meter. Also see INTERMODULATION-DISTORTION PERCENTAGE.

intermodulation noise Electrical noise produced in one channel by signals in another; it is caused by INTERMODULATION.

internal absorptance The ratio of flux absorbed in a substance to the flux leaving at the entry surface of the substance. It does not include energy reflected at the entry surface.

internal amplification In a radioactivity counter tube, current enhancement resulting from cumulative ionization initiated by an ionizing particle.

internal arithmetic In a digital computer, arithmetic operations performed in the computer, as opposed to those performed by peripherals.

internal impedance The impedance in a device, as opposed to that added from the outside. Compare INTERNAL RESISTANCE.

internal input impedance The impedance in a circuit or device, as “seen” from the input terminals. Compare INTERNAL OUTPUT IMPEDANCE.

internal input resistance The resistance in a circuit or device, as “seen” from the input terminals. Compare INTERNAL OUTPUT RESISTANCE.

internal noise Electrical noise generated within a circuit, as opposed to that picked up from outside. Such noise comes from transistors, diodes, integrated circuits, resistors, and any other component through which current flows.

internal output impedance The impedance in a circuit or device, as “seen” from the output terminals. Compare INTERNAL INPUT IMPEDANCE.

internal output resistance The resistance in a circuit or device, as “seen” from the output terminals. Compare INTERNAL INPUT RESISTANCE.

internal resistance **1.** The resistance of a device, as opposed to added resistance. See, for example, METER RESISTANCE. **2.** In a cell or battery, the equivalent resistance, resulting from imperfect conductivity of the electrolyte and electrodes, which limits the maximum deliverable current.

internal thermal shutdown In an integrated circuit, the junction temperature at which thermal shutdown occurs. It is generally indicated in degrees Celsius (°C).

internal transmittance The ratio of flux reaching the exit surface of a material to the flux leaving the entry surface. Reflection is not taken into account. The sum of internal transmittance and INTERNAL ABSORPTANCE is always equal to 1.

international broadcast station A shortwave broadcast station transmitting programs for international reception between 6 and 26.6 MHz.

international callsign The call letters of a station, assigned within a country, according to the method of arrangement (identifying letters, or letters and numerals) prescribed by the International Telecommunication Union.

international candle See CANDLE.

international coulomb A unit of electrical quantity, equal to 0.99985 absolute coulomb.

international farad A unit of capacitance, equal to 0.99952 absolute farad.

international henry A unit of inductance, equal to 1.00018 absolute henry.

international joule A unit of energy, equal to 1.00018 absolute joule.

International Morse Code See CONTINENTAL CODE.

international ohm A unit of electrical resistance, equal to 1.000495 absolute ohm. The other international units are derived from this value.

International Radio Consultative Committee Abbreviation, CCIR (*Comite Consultatif International Radiodiffusion*). An international organization reporting to the International Telecommunication Union, and studying technical operations and tariffs of radio and television.

International Steam Table calorie A unit of heat energy equal to 4.1868 joules.

International System of Units Abbreviation, SI (*Systeme International d'Unites*). The system of units of measurement established in 1960 under the Treaty of the Meter. The base units are as follows.

–METER (m), length: 1,650,763.73 times the wavelength of the light emitted in a vacuum of krypton 86

–KILOGRAM (kg), mass: the mass of the protype kilogram kept at Sevres, France

–SECOND (s), time: the duration of 9,192,631,770 periods of the radiation that corresponds to the transition between the two hyperfine levels of the ground state of cesium 133

–KELVIN (K), thermodynamic temperature: $\frac{1}{273.16}$ the thermodynamic temperature of the triple point of water

–AMPERE (A), electric current: the current that, flowing through two infinitely long parallel wires in a vacuum and separated by 1 meter, produces a force of 2×10^{-7} newton per meter of length between the wires

–CANDELA (cd), luminous intensity: the luminous intensity of $\frac{1}{600{,}000}$ square meter of perfectly radiating surface at the temperature of freezing platinum

International Telecommunication Union Abbreviation, ITU. An international, nongovernmental organization devoted to standardizing worldwide communications practices and procedures.

International Telegraph and Telephone Consultative Committee See CCITT.

International Telegraph Consultative Committee See CCIT.

international units A system of electrical units, based on the resistance through a specified quantity and configuration of the element mercury. The INTERNATIONAL OHM forms the basis for the international system of units.

international volt A unit of electrical potential, equal to 1.00033 absolute volt.

international watt A unit of power, equal to 1.00018 absolute watt.

Internet A worldwide, interconnected system of computer networks. Originated in the late 1960s as *ARPAnet* (Advanced Research Projects Agency network). Still used extensively by educators and scientists, but gaining popularity among personal-computer users. Sometimes called the INFORMATION SUPERHIGHWAY.

interphone An intercom aboard a mobile vehicle.

interpolation Finding a value that falls between two values listed in a table, indicated by a dial, plotted on a graph, derived by estimate, or given by intermediate calculation. For example, if a linear variable capacitor has a value of 100 pF when its dial is set to 10, and 140 pF when the dial is set to 20, then the capacitance when the dial reads 15 (midway between 10 and 20) can be assumed to be 120 pF (midway between 100 pF and 140 pF). When functions are not linear, interpolation is usually not exact.

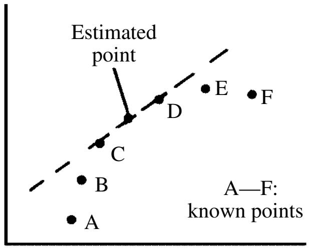

interpolation

interpolation meter See INTERPOLATION-TYPE INSTRUMENT.

interpolation oscillator A frequency-measuring signal generator with a built-in crystal oscillator, whose harmonics provide calibration points on the tuning dial. The dial provides a continuous indication of generator frequencies between crystal-harmonic points. Also see INTERPOLATION-TYPE INSTRUMENT.

interpolation-type instrument An instrument, such as a meter or signal generator, that is used to transfer an accurate quantity point from a standard to another instrument and to provide a range of values between such points. A secondary standard is sometimes used as an interpolation-type instrument (see, for example, INTERPOLATION OSCILLATOR).

interpole motor A direct-current motor with small auxiliary poles (*interpoles*) between its main field poles. The interpoles reduce sparking at the commutator.

interpreter A computer program that can convert instructions given in a high-level language (BASIC, for example) into the machine language that a computer uses; if it is not resident in the computer's nonvolatile memory, it must be loaded each time the machine is activated.

interrecord gap See INTERBLOCK SPACE.

interrupt A break in a computer program, as when a background job is interrupted so that a foreground job can be run. Also see BACKGROUND JOB and FOREGROUND JOB.

interrupted commercial and amateur service See INTERMITTENT COMMERCIAL AND AMATEUR SERVICE.

interrupted continuous wave Abbreviation, ICW. A continuous wave that is interrupted at regular intervals, as in the chopping of a wave at a regular rate. Compare CONTINUOUS WAVE (CW) and MODULATED CONTINUOUS WAVE (MCW).

interrupted dc A direct current or voltage that is periodically started and stopped by switching or chopping.

interrupter contacts Auxiliary contacts operated directly by the armature of a stepping switch.

interruption frequency See QUENCHING FREQUENCY.

interruption-frequency oscillator See QUENCH OSCILLATOR.

interrupt signal The signal that causes a break (INTERRUPT) in a computer program.

intersatellite communication **1.** Communication between or among satellites. **2.** Communication between two earth-based stations, using two or more satellites. **3.** Communication between an earth-based station and a satellite-based station, using at least one intermediate relaying satellite.

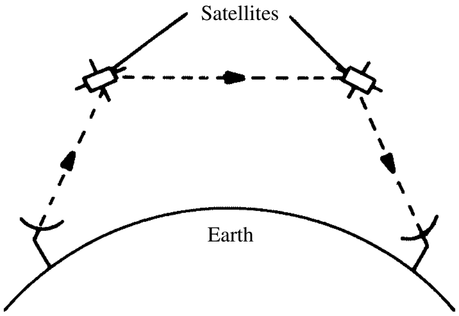

intersatellite communication

intersecting storage ring A device for producing great amounts of energy. It is similar to a vacuum tube. High-speed subatomic particles are fed to a ring-shaped evacuated structure in opposite directions. The two particle beams collide at various points, yielding high energy.

intersection The logical AND operation.

interstage capacitor A coupling capacitor used between two circuit stages.

interstage coupling The transfer of a signal between two circuit stages, such as those of an amplifier. Common forms of interstage coupling include *direct coupling*, *capacitive coupling*, *transformer coupling*, *diode coupling*, and *optoisolator coupling.*

interstage diode A semiconductor coupling diode used between two circuit stages.

interstage transformer A coupling transformer used between two circuit stages. It provides direct-current isolation, and also can match purely resistive impedances.

intersymbol interference In a digital communications signal, a condition in which a given symbol overlaps with one or more other symbols (either immediately preceding it or immediately following it), upsetting the ability of the receiver to decipher signals in certain time intervals. The phenomenon is sometimes a problem in time-division multiplexing—especially at data speeds near the maximum for the system.

intersystem A power-generating network of interconnected separate systems.

intersystem communications Communications between computer systems, either through direct linking of central processors, or by mutual use of peripherals and input/output channels.

intertie See INTERCONNECTION, **2**.

interval **1.** The amount of separation between successive points, events, or quantities. **2.** The continuous range of values between two defined points. **3.** A specific period of time, with defined beginning and ending points.

intervalometer A timing device for operating equipment over a precisely defined time interval.

interval timer A device that provides power to an equipment for a precise interval upon application of a simple initiating signal or action. See also INTERVALOMETER.

interword space The seven-unit interval between words or code groups in telegraphy. Compare INTERCHARACTER SPACE.

intoxication tester See DRUNKOMETER.

intracoding The coding of data using only data that it contains.

Intrafax Western Union's private facsimile system.

Intraframe A method of digital IMAGE COMPRESSION developed by *MPEG (Moving Picture Experts Group)* and *JPEG (Joint Photographic Experts Group).* It operates by eliminating redundant data within image frames. Compare INTERFRAME.

intrinsic-barrier diode See PIN DIODE.

intrinsic-barrier transistor A bipolar transistor with a layer of intrinsic semiconductor between one of its pn junctions.

intrinsic concentration The number of minority carriers exceeding the normal equilibrium number in a semiconductor.

intrinsic conduction The flow of electron/hole pairs in an intrinsic semiconductor subjected to an electric field.

intrinsic flux A quantity equal to the product of the intrinsic flux density and the cross-sectional area in a magnet.

intrinsic flux density The increased flux density of a magnet in its actual environment, as compared with the flux density resulting from the same magnetizing force in a perfect vacuum.

intrinsic mobility Electron mobility in an intrinsic semiconductor. Also see CARRIER MOBILITY and MOBILITY.

intrinsic *Q* The value of the Q, also known as the FIGURE OF MERIT, for an unloaded circuit. This value is generally higher than the value when a load is connected to the circuit.

intrinsic semiconductor A semiconductor whose characteristics are identical to those of a pure crystal of the material. In this condition, the semiconductor is nearly an insulator. Example: highly purified germanium or silicon before n- or p-type impurities have been added. Compare EXTRINSIC SEMICONDUCTOR.

intrusion alarm A set of electronic sensors and associated circuitry composing a system that detects and warns of the presence of unauthorized personnel within a specific region.

intrusion sensor A sensitive pickup (such as a photocell, ultrasonic detector, or capacitive transducer) that responds to a nearby body by delivering an actuating signal to an intrusion alarm.

INV Abbreviation of INVERTER.

inv Abbreviation of INVERSE.

Invar A nickel-steel alloy (36% nickel) having a low temperature coefficient of linear expansion (1 ppm/°C). Invar is used in electronic equipment where mechanical distortion resulting from temperature changes must be negligible, and in magnetostrictive circuits (see MAGNETOSTRICTION).

inverse **1.** Opposite in nature (e.g., an INVERSE CHARACTERISTIC). **2.** Of opposite sign (e.g., a negative current or voltage). **3.** An operation of opposite kind; thus, subtraction is the inverse of addition, and division is the inverse of multiplication.

inverse beta The beta of a transistor operated with the emitter and collector interchanged.

inverse bias See REVERSE BIAS.

inverse characteristics The characteristics of a bipolar transistor when operated with the emitter and collector reversed.

inverse conduction See REVERSE CONDUCTION.

inverse cube law A principle relating the intensity of an effect to the reciprocal of the cube of the distance from the source. The magnetic field around a solenoidal coil of wire obeys this principle.

inverse current See REVERSE CURRENT.

inverse-distance law The inverse-square law applied to the propagation of radio waves, assuming that the waves do not encounter obstacles.

inverse feedback See DEGENERATION.

inverse fourth-power law A rule of propagation for certain complex forms of energy: $I = k/d^4$, where I is the intensity of the field, d is the distance from the source, and k is a constant.

inverse impedances Impedances (Z_1 and Z_2) that are the reciprocal of another impedance (Z_3), satisfying the relationship $Z_1Z_2 = (Z_3)^2$. Also called RECIPROCAL IMPEDANCES.

inverse leakage The flow of a small static reverse current in semiconductor devices.

inverse-parallel circuit See BACK-TO-BACK CIRCUIT and BACK-TO-BACK CONNECTION.

inverse peak voltage See PEAK INVERSE VOLTAGE.

inverse piezoelectric effect Mechanical movement in a piezoelectric material, caused by application of voltage.

inverse resistance See REVERSE RESISTANCE.

inverse resonance See PARALLEL RESONANCE.

inverse-square law The energy or power intensity of a phenomenon is inversely proportional to the square of the distance from the source. This is often applied to quantitative reasoning about radiant energy, electromagnetic energy, and acoustic energy. Thus, if the distance doubles, the energy or power drops to ¼ its previous value.

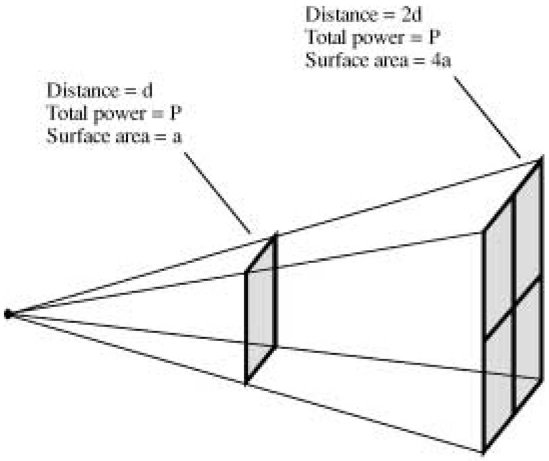

inverse-square law

inverse trigonometric function An angle expressed in terms of a given trigonometric function, followed by the exponent -1 or preceded by "arc" (sin^{-1} is the equivalent of *arcsin*).

inverse voltage **1.** The negative voltage at the anode of a rectifier during the negative half-cycle of alternating-current (ac) input. **2.** The voltage across a power-supply filter capacitor during the negative half-cycle of ac input. **3.** Semiconductor-junction reverse voltage.

inverse Wiedeman effect See DIRECT WIEDEMAN EFFECT.

inversion **1.** A reversal of the normal vertical temperature gradient of the atmosphere, often resulting in long-distance tropospheric radio-wave propagation. **2.** Speech scrambling (see SCRAMBLER CIRCUIT). **3.** Phase inversion (see PHASE INVERTER). **4.** Changing direct current into alternating current, often increasing the voltage (see INVERTER, **1**).

inverted amplifier A push-pull, grounded-gate, field-effect-transistor (FET) amplifier.

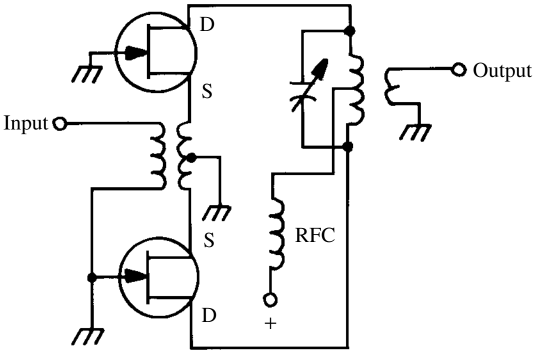

inverted amplifier

inverted-L antenna An antenna having a horizontal radiator and a vertical feeder or lead-in attached to one end of the radiator. The entire arrangement resembles an upside-down L. The overall length is generally ¼ to ½ wavelength.

inverted speech See SCRAMBLED SPEECH.

inverter **1.** Also called *power inverter.* A device that converts direct current (dc) into alternating current (ac), often of a much higher voltage (e.g., 12 Vdc into 117 Vac). **2.** A logic circuit that provides an output pulse that is a negation of the input pulse. Also called a COMPLEMENTER or a NOT CIRCUIT. **3.** See PHASE INVERTER.

inverting adder An analog adder circuit that is provided with an amplifier for a 180° phase shift.

inverting amplifier An amplifier providing a 180° phase shift between input and output.

inverting connection Connection to the inverting input terminals of a differential amplifier or operational amplifier. Also see INVERTING INPUT. Compare NONINVERTING INPUT.

inverting input In a differential amplifier or operational amplifier, the input circuit that produces a phase reversal between the input and output. Compare NONINVERTING INPUT.

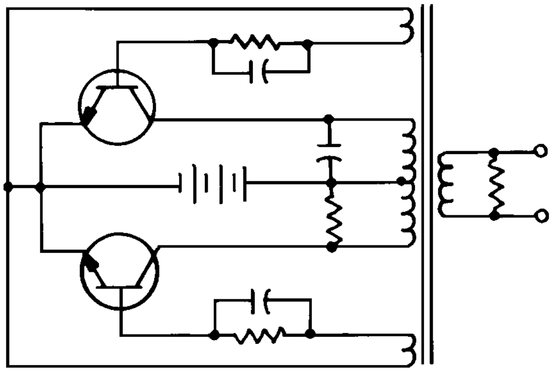

inverter, 1

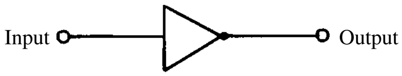

In	Out
0	1
1	0

inverter, 2

inverting transponder In a communications satellite, a transponder in which the downlink band is "upside-down" in frequency relative to the uplink band. That is, the highest downlink frequency corresponds to the lowest uplink frequency, and the lowest downlink frequency corresponds to the highest uplink frequency. Compare NONINVERTING TRANSPONDER. See also DOWNLINK, TRANSPONDER, UPLINK.

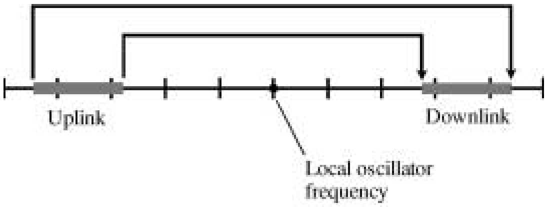

inverting transponder

invisible failure In a computer system, a hardware or software failure whose effect on the system is unnoticeable in a given application. A failure that is invisible in one application might be vividly apparent in some other application.

invister A unipolar semiconductor material, capable of operation at very high frequencies.

involution Raising a number to a power: squaring, cubing, etc. Compare EVOLUTION.

inward-outward dialing Also called *direct dialing.* In a telephone system, a method of dialing in which calls can be made to and from branch exchanges, without operator assistance.

I/O Abbreviation of *input/output* (see INPUT/OUTPUT EQUIPMENT).

Io Symbol for OUTPUT CURRENT.

iodine Symbol, I. A nonmetallic element of the halogen family. Atomic number, 53. Atomic weight, 126.905. Also see HALOGEN.

ion A charged atom [i.e., one that has gained one or more electrons (a negative ion, or *anion*) or lost one or more electrons (a positive ion, or *cation*)].

ion burn A spot burned on the screen of a cathode-ray tube by negative ions from the cathode striking a single point on the faceplate with high intensity for long periods.

ion concentration **1.** The number of ions, expressed as a percentage or as a number per unit volume, in a substance. **2.** Ionization density in the atmosphere.

ion exchange resins Granular resins that contain acid or base groups, and that trade ions with salts in solutions. The resins play a part in the purification of water for various industrial processes.

ionic binding forces In a crystal, the binding forces that occur when valence electrons of one atom are joined to those of a neighboring atom whose outer shell they fill.

ionic bond In a solid, a bond between atoms, formed as a result of the attraction between oppositely charged atoms (ions).

ionic conduction Conduction, as in a gas or electrolyte, by ion migration (positive to the cathode, negative to the anode).

ionic crystal A crystal whose lattice is held together by the electric forces between ions. Also see IONIC BINDING FORCES and IONIC BOND.

ionic current Current caused by ion movement in a gas or liquid. Also see ION, IONIZATION, and IONIC CONDUCTION.

ionic semiconductor A semiconductor in which the carrier is an ion, as opposed to an electron or hole.

ionic switch See ELECTROCHEMICAL SWITCH.

ionization **1.** The loss or gain of one or more electrons by an atom. Also see ANION, CATION, and ION. **2.** The formation or existence of significant numbers of ions in a gas, liquid, or solid (e.g., ionization of the upper atmosphere).

ionization arc The electrical discharge resulting from the ionization of a material because of high voltage.

ionization chamber An enclosure containing a gas and a pair of electrodes between which a high voltage is applied. Radiation, such as X rays or radioactive particles, passing through the walls of the chamber ionize the gas, creating an ionization current that is proportional to the intensity of the radiation.

ionization current **1.** Current in an ionized gas (such as air). **2.** Current flowing in an electrolyte. **3.** Current in an ionization chamber, Geiger-Mueller tube, or similar gaseous device. **4.** In a gas tube, current flowing after the ignition potential has been reached. **5.** Negative grid current resulting from gassiness in a vacuum tube.

ionization density The extent to which ionization exists in an ionized layer of the atmosphere. The higher the ionization density, the greater the effect on radio waves—especially at frequencies below about 150 MHz.

ionization gauge A form of vacuum tube that can be used to measure the hardness of a vacuum. It consists of a cathode, an anode (plate), and a positively charged grid. Plate current flows as a result of ionization of the atoms within the tube. The more nearly perfect the vacuum, the lower the plate current.

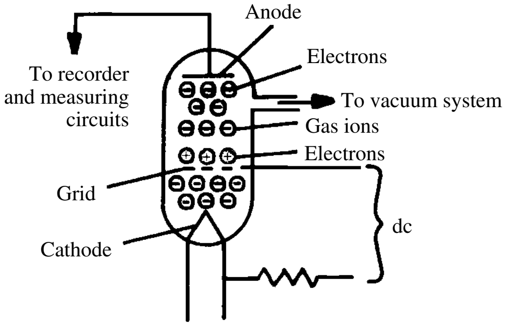

ionization gauge

ionization potential The voltage at which a substance (especially a gas) ionizes. Also called (for a gas) *ignition potential* (see BREAKDOWN VOLTAGE).

ionization pressure In an ionized gas, the pressure increase resulting from the ionization, as compared with the same volume and mass of gas when not ionized.

ionization resistance See CORONA RESISTANCE.

ionization smoke detector A device that senses the presence of smoke or other particles as a result of changes in the IONIZATION CURRENT through the air. When the ionization current suddenly changes, a signal is sent to an alarm circuit.

ionization time The interval, usually measured in microseconds or milliseconds, between the instant that an ionizing potential is applied to a gas, and the instant at which the gas begins to ionize. Compare DEIONIZATION TIME.

ionization trail The ionized path of a meteor as it passes through the upper atmosphere.

ionize To cause the electrons in a substance, particularly a gas, to move freely from atom to atom.

ionized gas A gas whose atoms, under the influence of a strong electric field or IONIZING RADIATION, have become positive or negative ions.

ionized layer See KENNELLY-HEAVISIDE LAYER.

ionized liquid See ELECTROLYTE.

ionizing radiation **1.** Any high-energy electromagnetic radiation that causes ionization in a gas through which the field passes. Examples: ultraviolet, X rays, and gamma rays. **2.** High-speed atomic nuclei (e.g., protons or alpha particles).

ion migration The movement of ions through a solid, liquid, or gas because of the influence of an electric field.

ionosphere **1.** Any of several ionized regions at specific altitudes above the earth's surface. These layers cause absorption and refraction of electromagnetic (EM) fields at some radio frequencies. The D layer exists at an altitude of about 30 miles (50 km) and is ordinarily present only on the daylight side of the planet. This layer does not contribute to, and in fact sometimes hinders, wireless communications. The E layer, about 50 miles (80 km) above the surface, also exists mainly during the day, but nighttime ionization is sometimes observed. The E layer can facilitate medium-range radio communication at certain frequencies. The uppermost regions are the F1 and F2 layers. The F1 layer, normally present only on the daylight side of the earth, forms at about 125 miles (200 km) altitude; the F2 layer exists at about 180 miles (300 km) more or less around the clock. Sometimes the distinction between the F1 and F2 layers is ignored, and they are spoken of together as the F layer.

ionospheric disturbance See IONOSPHERIC STORM.

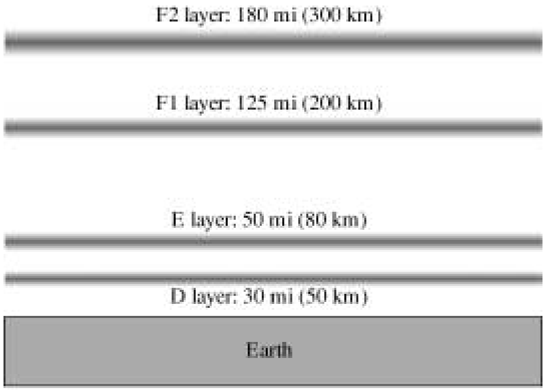

ionosphere

ionospheric forecasting Predicting ionospheric conditions. Radio propagation data is derived from such predictions.

ionospheric layers The respective layers of the ionosphere: *D layer* (at an altitude of about 30 mi or 50 km), *E layer* (at an altitude of about 50 mi or 80 km), the *F1 layer* (at an altitude of about 125 mi or 200 km), and the *F2 layer* (at an altitude of about 180 mi or 300 km). The F1 layer is generally present only in the daytime.

ionospheric propagation Propagation of radio waves by means of reflection or refraction by the ionosphere. Also see HOP; INCIDENT WAVE; IONOSPHERE; MULTIHOP PROPAGATION; REFLECTED WAVE, **1**; REFRACTED WAVE; and SKYWAVE.

ionospheric storm Turbulence in the ionosphere, usually accompanied by a magnetic storm and caused by high-speed particles emitted from an eruption on the sun.

ion sensor A device whose operation is based on the detection of ions and the delivery of a proportionate voltage. Examples are the *Geiger counter*, *halogen gas leak detector*, *mass spectrometer*, and *vacuum gauge*.

ion spot See ION BURN.

ion trap See BENT-GUN CRT.

ion-trap magnet An external (usually double) magnet used with a television picture tube to deflect the ion beam away from the screen. This prevents ION BURN.

I/O port That part of a computer providing, via a connector, a point through which data can enter from, or exit to, peripheral equipment.

I_P **1.** Abbreviation of PLATE CURRENT. **2.** Abbreviation of PEAK CURRENT.

I-phase carrier In color television, a carrier separated by 57 degrees from the color subcarrier.

I-picture A video image that is coded using only data that it contains.

ipm Abbreviation of INCHES PER MINUTE.

ips Abbreviation of INCHES PER SECOND.

IR **1.** The product of current and resistance (see, for example, IR DROP). **2.** Abbreviation of INSULATION RESISTANCE. **3.** Abbreviation of INFRARED.

$\mathrm{I_r}$ Symbol for IRIDIUM.

I_r Symbol for CURRENT in a resistor.

IRAC Abbreviation for *Interdepartment Radio Advisory Committee* (a federal government group in the United States).

IR diode See INFRARED DIODE.

IR drop The voltage drop (E) across a resistance (R) when there is a current (I) through the resistor; according to Ohm's law, $E = IR$.

IRE Abbreviation for *Institute of Radio Engineers*, the predecessor of the IEEE (*Institute of Electrical and Electronics Engineers*).

IRED Abbreviation for INFRARED-EMITTING DIODE.

I_{RF} Symbol for RADIO-FREQUENCY CURRENT.

iridescence A sparkling, colorful appearance in a material, resulting from refraction, internal reflection, and interference in light waves passing through the substance. It is especially noticeable in quartz and certain gems.

iridium Symbol, Ir. A metallic element of the platinum group. Atomic number, 77. Atomic weight, 192.22.

iron Symbol, Fe. A ferromagnetic, metallic element. Atomic number, 26. Atomic weight, 55.847. Iron (and its special form, steel) is widely used in magnetic circuits.

iron/constantan thermocouple A thermocouple consisting of a junction between wires or strips of IRON and CONSTANTAN.

iron core A transformer or choke core made from iron or steel. The core is usually laminated to reduce eddy-current loss.

iron-core coil An inductor having an iron or steel core, usually laminated to reduce eddy-current loss.

iron-core IF transformer An intermediate-frequency transformer having a core of powdered iron, a form of iron that has the advantage of high permeability while greatly minimizing eddy currents.

iron-core transformer A transformer whose coils are wound on a core of laminated iron or steel.

iron loss Power lost in the iron cores of transformers, inductors, and electrical machinery as a result of eddy currents and hysteresis.

iron magnet A permanent magnet consisting of magnetized iron or a mixture of iron and nickel.

iron oxide A compound of iron and oxygen, whose most familiar form is common rust. The several variants have characteristics that depend on the number of iron and oxygen atoms in the iron-oxide molecule. See, for example, MAGNETITE and RED OXIDE OF IRON.

iron pyrites Formula, FeS_2. Natural iron sulfide that occurs as bright yellow crystals in its natural state.

iron-vane meter An alternating-current meter whose movable element, a soft iron vane, carries the pointer and pivots near a similar, stationary vane. The vanes are mounted in a multiturn coil of wire. The current flows through the coil, the resulting magnetic field magnetizing the vanes. Because the magnetic poles of the vanes are identical, they repel each other; the movable vane is deflected (against the torque of returning springs) over an arc proportional to the current, carrying the pointer over the scale.

irradiance The amount of radiant flux impinging on a unit surface area; it is generally specified in watts per square meter (W/m^2).

irradiation **1.** Exposure of a device to radioactivity or X rays. **2.** The total radiant power density that is incident upon a receiving surface.

irrational number A number that cannot be expressed as the quotient of two integers. Its decimal expansion is nonterminating and nonrepeating.

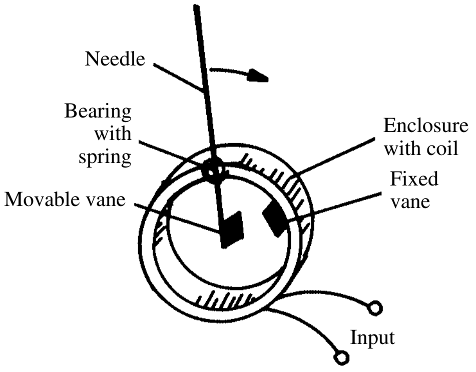

iron-vane meter

irregularity **1.** The condition of being nonuniform, or rapidly fluctuating, rather than constant. **2.** A departure from normal operating conditions. **3.** Nonuniformity in a surface. **4.** Nonuniform distribution of matter. **5.** Nonuniform distribution of data.

irregular wave A wave disturbance, or the plot of such a disturbance versus time, that has a complex, periodic, repeating nature.

IR viewer A device that allows observation of images at infrared wavelengths. See SNIPERSCOPE and SNOOPERSCOPE.

I_s **1.** Symbol for source current in a field-effect transistor. **2.** Symbol for *screen current* in a vacuum tube.

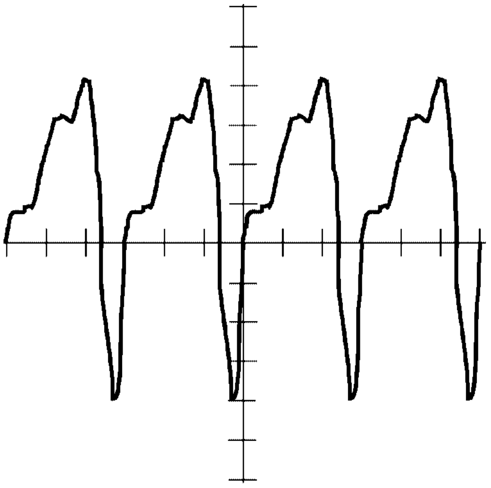

irregular wave

ISDN Abbreviation for INTEGRATED SERVICES DIGITAL NETWORK.

ISCAN Abbreviation of *inertialess steerable communications antenna.*

I-scan A radar display in which the target is shown as a complete circle, whose radius is proportional to the distance to the target.

I-signal With the Z-signal, one of the two signals that modulates the chrominance subcarrier in color television. The I-signal results from mixing a B-Y signal (with -0.27 polarity) and an R-Y signal (with +0.74 polarity).

Isinglass Thinly laminated mica.

ISO Abbreviation for *International Standards Organization.*

ISO 9660 A standard format for producing CD-ROM (COMPACT DISK READ-ONLY MEMORY) mass storage media for use with computers. It is a part of the YELLOW BOOK scheme.

isobar **1.** An atom whose nucleus has the same weight as that of another atom but differs in atomic number. **2.** On a weather map, a line connecting points of equal pressure. Also see BAR, **1**.

isochromal phenomena **1.** Effects occurring at regular time intervals. **2.** Effects of equal duration.

isochromatic Also *orthochromatic.* **1.** The quality of having or producing natural visible-light hues. **2.** Color sensitivity excluding a response to red.

isochronal See ISOCHRONE.

isochrone On a map, a line connecting points of constant time difference in radio-signal reception. It is useful in radiolocation and radionavigation.

isochronous Having identical resonant frequencies or wavelengths.

isoclinic line See ACLINIC LINE.

isodose Pertaining to points receiving identical dosage of radiation.

isodynamic line On a map of the geomagnetic field (the earth's magnetic field), a line connecting points of equal flux density.

isoelectric Having a potential difference of zero.

isoelectronic Having the same number of electrons.

isogonal **1.** See ISOGONIC LINE. **2.** Having uniform magnetic declination at all points.

isogonic line On a map of the geomagnetic field (the earth's magnetic field), a line connecting points of equal magnetic declination.

isolantite An insulating ceramic. Dielectric constant, 6.1.

isolated **1.** Electrically insulated. **2.** Separated in such a way that interaction does not take place.

isolated input **1.** An ungrounded input. **2.** An input circuit with a blocking capacitor to prevent the passage of direct current.

isolated location In a computer, a storage location that is hardware-protected from being addressed by a user's program.

isolating amplifier See BUFFER, **1**.

isolating capacitor A series capacitor inserted in a circuit to pass an alternating-current signal while blocking direct current. Also called a BLOCKING CAPACITOR.

isolating diode A diode used (because of its unidirectional conduction) to pass signals in one direction, but block them in the other direction.

isolating resistor A high-value resistor connected in series with the input circuit of a voltmeter or oscilloscope to protect the instrument from stray pickup. In most voltmeters, this resistor is built into the probe.

isolating transformer A power transformer, usually having a 1:1 turns ratio, for isolating equipment from direct connection to the power line.

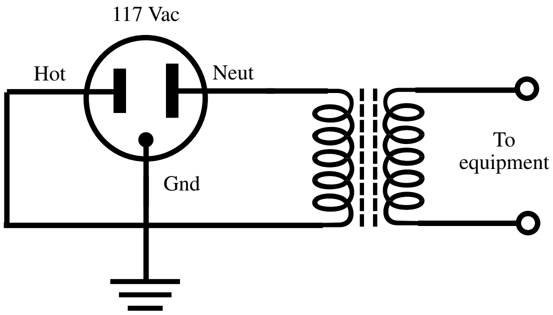

isolating transformer

isolation The arrangement or operation of a circuit so that signals in one portion are not transferred to (nor affect) another portion.

isolation amplifier See ISOLATING AMPLIFIER.

isolation capacitor See ISOLATING CAPACITOR.

isolation diode **1.** In an integrated circuit, a reverse-biased diode that is formed in the substrate to prevent cross-coupling and grounds. **2.** See ISOLATING DIODE.

isolation resistor See ISOLATING RESISTOR.

isolation transformer See ISOLATING TRANSFORMER.

isolator See OPTOELECTRONIC COUPLER.

isolith A form of monolithic integrated circuit, in which the semiconductor is removed in certain places for the purpose of isolating different parts of the circuit.

isomagnetic Having equal magnetic intensity.

isomer A material that has the same atomic number or chemical formula as some other substance, but, because of a difference in the atomic structure, is an entirely different substance. An example is carbon; it can be either *graphite* (by far the more common form) or *diamond.*

isophote On a graph of visible-light intensity, a curve joining points of equal brightness.

isoplanar An integrated-circuit configuration in which insulating barriers or metal oxides are fabricated among the bipolar elements.

isothermal process A physical or chemical process in which there is no temperature change as other factors vary. Compare ENDOTHERMIC REACTION and EXOTHERMIC REACTION.

isotope An atom having the same number of protons as another atom, thereby composing the same chemical element, but having a different number of neutrons. Thus, *deuterium* is an isotope of *hydrogen.* Some isotopes are radioactive [e.g., *carbon 14* (an isotope of the more-common *carbon 12*)]. The two extra neutrons in carbon 14 make it less stable than carbon 12.

isotropic antenna A theoretically ideal antenna that transmits and/or receives electromagnetic fields equally well, and with 100-percent efficiency, in all directions in three-dimensional free space.

isotropic radiator A theoretically ideal radiating element that transmits electromagnetic fields equally well, and with 100-percent efficiency, in all directions in three-dimensional free space.

I_{sup} Symbol for *suppressor current.*

I(t) Symbol for INDICIAL RESPONSE.

ITD Abbreviation for INITIAL TIME DELAY.

item **1.** Component. **2.** Any one of a number of similar or identical components, circuits, or systems.

iteration Repeating a series of arithmetic operations to arrive at a solution to a problem. Computers are commonly programmed to do this thousands, millions, or billions of times. Such a program must include a statement of acceptable accuracy so that it knows when to leave the iteration loop.

iterative impedance In a network consisting of identical, cascaded sections, the input impedance of a section to which the output impedance of the preceding section is made equal.

iterative routine A program or subroutine that provides a solution to a problem by iteration.

iterative transfer constant Symbol, *P.* A property of ITERATIVE IMPEDANCE networks. If I_i is the network input current and I_o is the network output current, then $P = \log_e(I_o/I_i)$.

ITU Abbreviation for *International Telecommunication Union.*

ITV **1.** Abbreviation of INDUSTRIAL TELEVISION. **2.** Abbreviation for INTERACTIVE TELEVISION.

I-type semiconductor See INTRINSIC SEMICONDUCTOR.

I_X Symbol for *current in a reactance.*

I_Y Symbol for *current in an admittance.*

I_Z Symbol for *current in an impedance.*

-ize A suffix used, with some liberty, to form verbs from nouns. In electronics, this commonly refers to procedures or processes (e.g., to ANODIZE, ELECTROCIZE, PLASTICIZE, or TRANSISTORIZE).

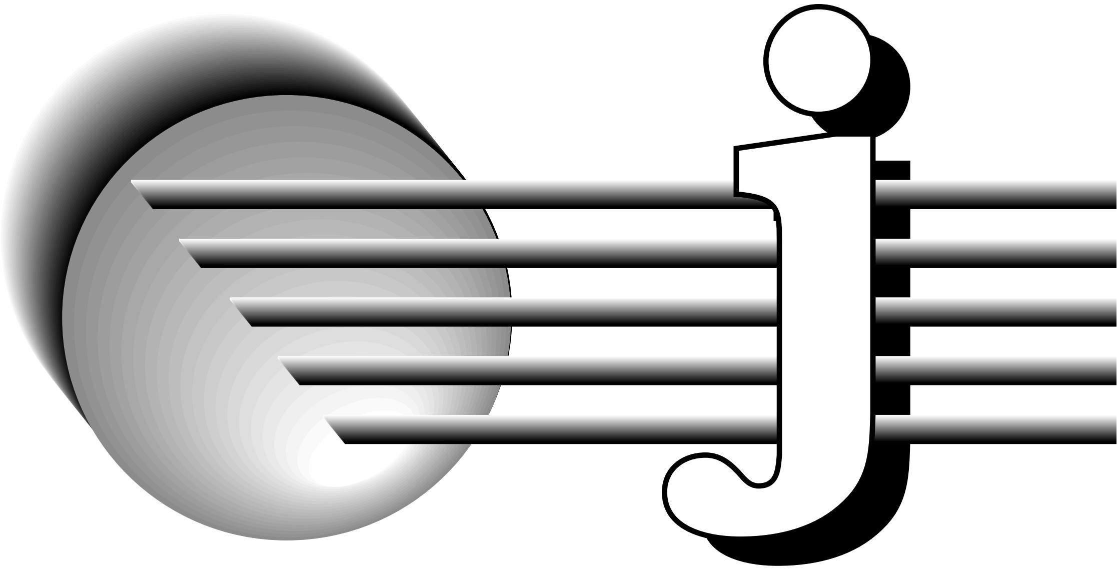

J **1.** Abbreviation for JOULE. **2.** Symbol for JACK or CONNECTOR. **3.** Symbol for EMISSIVE POWER.

***j* (*j* operator)** The square root of –1; an imaginary number (usually denoted *i* in mathematics). Assigned to reactance values depicted on the vertical axis of the resistance-reactance (RX) plane in impedance vector diagrams, and whose currents are 90 degrees out of phase with the current in the resistive part of an alternating-current circuit.

jack A receptacle for a plug. A plug (a male connector) is inserted into a jack (a female connector) to complete a circuit or removed from it to break a circuit.

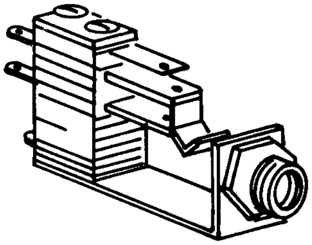

jack

jack box A (usually metallic) box or can used to hold, shield, or protect a jack or group of jacks.

jacket **1.** The outer covering on a cable, as opposed to the insulation or dielectric separating the individual conductors within the cable. **2.** An insulating outer case or wrapper on a component, such as a capacitor. **3.** A shield can or shield box. **4.** A heat-radiating or water-conducting enclosure used in cooling a power vacuum tube.

jack panel A (usually metallic) panel in which a number of jacks are mounted, usually in some order or sequence as denoted by labels.

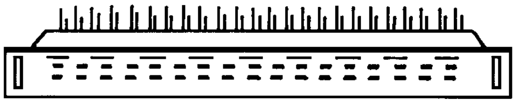

jack panel

jackscrew In a two-piece connector, a screw for mating or separating the halves of the connector.

Jacob's law A principle concerning the behavior of motors. An electric motor develops maximum power when $E_i = 2E_{bk}$, where E_i is the applied voltage and E_{bk} is the back voltage.

JAES Abbreviation for *Journal of the Audio Engineering Society.*

jaff Colloquial term for radar jamming that combines electronic and chaff techniques.

jag Distortion caused by temporary loss of synchronization between the scanner and recorder in a facsimile system.

jam input **1.** A means of setting a logic line to the desired condition by directly applying the desired high or low voltage. **2.** A voltage applied to a logic line to force it high or low.

jammer **1.** A radio transmitter or station used for the purpose of JAMMING communications be-

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

tween or among other stations. **2.** A radio operator who engages in the practice of deliberately JAMMING communications between or among other stations.

jamming The deliberate use of countermeasures, such as malicious transmission of interfering signals, to obstruct communications.

jamming effectiveness The extent to which JAMMING is able to disrupt a service. It can be expressed quantitatively as the ratio of jamming signal voltage to jammed signal voltage. It can also be determined according to the percentage of data that is effectively obliterated.

JAN Abbreviation of JOINT ARMY-NAVY.

janet A system for point-to-point communication via meteor-trail forward scatter. It is generally used at very high frequencies (VHF).

Jansky noise Wideband, high-frequency electromagnetic noise generated by objects in interstellar and intergalactic space.

J antenna An end-fed half-wave antenna having a quarter-wave, parallel-wire matching section. The entire antenna, when oriented vertically, resembles the letter J.

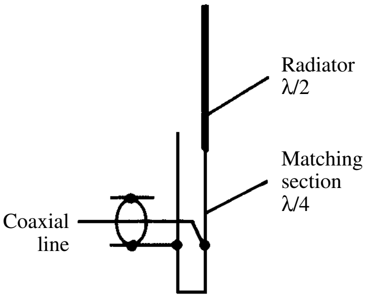

J antenna

Janus antenna array (from *Janus*, an ancient Roman god.) A Doppler-navigation antenna array radiating forward and backward beams.

jar **1.** (From *Leyden jar*) An obsolete unit of capacitance equal to 1/900 microfarad. **2.** The container for the elements of a storage cell.

JASA Abbreviation for *Journal of the Acoustical Society of America.*

J-carrier system In carrier-current (wired/wireless) telephony, a broadband system that provides 12 telephone channels at frequencies up to 140 kHz.

JCET Abbreviation of *Joint Council on Educational Television.*

JCL Abbreviation of *job control language.*

J-display A radar display having a circular time base. The transmitted pulse and reflected (target) pulse are spaced around the circumference; distances can be measured circumferentially between them.

JEDEC Acronym for *Joint Electron Device Engineering Council.*

jerk The rate of change of acceleration; the third derivative of displacement.

JETEC Acronym for *Joint Electron Tube Engineering Council.*

jewel bearing A low-friction bearing used in electric meters and other sensitive devices. It takes its name from a jewel pivot (such as a sapphire) in the groove of which rides the pointed end of a rotating shaft. Also called *jeweled bearing* or *jewel.*

jezebel A passive sonobuoy used in military applications. It detects enemy submarine noises, and transmits them by radio to a monitoring station.

JFET Abbreviation of *junction field-effect transistor.*

JHG Abbreviation of JOULE HEAT GRADIENT.

jig A device constructed especially for the purpose of holding an equipment or circuit board during its repair.

jitter A (usually small and rapid) fluctuation in a phenomenon, such as a quantity or wave, because of noise, mechanical vibration, interfering signals, or similar internal or external disturbances. It is used especially in reference to cathode-ray-tube (CRT) displays.

J/K Abbreviation for *joule(s) per Kelvin*, the SI unit of entropy; also the unit for the Boltzmann constant.

J-K flip-flop A transistor-resistor flip-flop stage producing an output signal even when both inputs are in the logic 1 state (high). It is so called because its input terminals are labeled J and K.

J/(kg·K) Abbreviation for *joule(s) per kilogram Kelvin*, the SI unit of specific heat capacity.

job A unit of computer work, usually consisting of several program runs.

job control language An operating system language used to describe the control requirements for jobs within the system.

job control program A program that uses control language statements and implements them as instructions controlling a job in an operating system.

job control, stacked See SEQUENTIAL-STACKED JOB CONTROL.

job flow control To control the order of jobs being processed by a computer to make the most efficient use of peripherals and central processor time—either manually or by an operating system.

job library A series of related sets of data that will be loaded for a given job.

job-oriented terminal A terminal that produces data in computer-ready form.

job statement A control statement identifying the beginning of a series of job control statements for a job.

job step The execution of a computer program according to a job control statement; several job steps can be specified by a job.

job stream In a processing system, a group of consecutively run jobs.

jogging Rapid, repetitive switching of power to a motor to advance its shaft by small amounts. Also called *inching.*

Johnson counter See RING COUNTER.

Johnson curve A spectral curve (important in appraising solar cell performance) for air mass zero (i.e., for conditions beyond the earth's atmosphere).

Johnson-Lark-Horowitz effect The resistivity gained by a metal or degenerate semiconductor (one in which conduction is nearly equal to that of a simple metal) because of electron scattering by impurity atoms.

Johnson noise See THERMAL NOISE.

Johnson-*Q* feed system See *Q*-ANTENNA.

join Also called *disjunction.* The logical inclusive-OR operation.

joined actuator A form of multiple circuit breaker in which the opening of one circuit results in the opening of all circuits.

joint See JUNCTION, **1**.

joint circuit A communications circuit shared by two or more services.

joint communications Communication facilities being used by more than one service of the same country.

joint denial The logical NOR (NOT-OR) operation.

joint-force sensor A feedback servo device that prevents a robot joint from exerting excessive force. It works by detecting the mechanical resistance the robot arm encounters. If the resistance becomes too great, the joint force is reduced or removed.

jolt **1.** Colloquialism for KILOVOLT. **2.** Colloquialism for ELECTRIC SHOCK. **3.** Colloquialism for TRANSIENT. **4.** Colloquialism for a lightning discharge.

Joly transformer A frequency-tripling transformer whose frequency-multiplying action depends on the nonlinearity of the magnetic induction curve of the core material.

Jones plug A special form of polarized receptacle having numerous contacts.

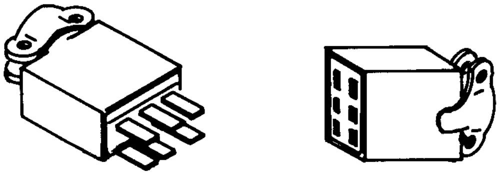

Jones plug

***j* operator** See *J.*

Josephson effect The phenomenon, predicted by Brian Josephson, wherein a current flows across the gap between the tips of two superconductors brought close together; a high-frequency wave is generated.

Joshi effect The phenomenon whereby current in a gas changes as the result of irradiation by light.

joule Abbreviation, J. The SI unit of work. One joule is the work performed when the point of application of one newton is moved one meter in the direction of the applied force. Also see NEWTON.

Joule calorimeter A heat-measuring device that operates electrically.

Joule constant See MECHANICAL EQUIVALENT OF HEAT.

Joule effect **1.** The heat resulting from current flowing through a resistance. **2.** See MAGNETOSTRICTION.

joule heat See JOULE EFFECT, **1**.

joule heat gradient The rate of change in the temperature of a resistive object through which a current flows.

joule meter An integrating wattmeter producing readings in joules.

Joule's law The rate at which heat is produced by current flowing in a constant-resistance circuit is proportional to the square of the current.

journal A file of messages within an operating system, providing information for restarts and historical analysis of system functioning.

joystick A two- or three-dimensional potentiometer with a movable lever, allowing control of a parameter, according to the position of the lever in the up/down and left/right directions. It is often used in computer games for the purpose of manipulating images on a screen. In some such devices, the lever can be rotated clockwise or counterclockwise to obtain additional functions.

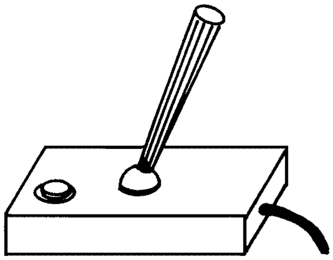

joystick

JPEG Abbreviation of *Joint Photographic Experts Group.*

JPEG image compression An image-compression scheme that eliminates redundancies, greatly reducing the necessary digital storage space. The process is fast enough to allow animated graphics at moderate speed.

JPL Abbreviation of *Jet Propulsion Laboratory.*

J-pole antenna See J ANTENNA.

J rule During transitions of orbital electrons from higher to lower energy states (accompanied by the emission of photons), changes in the inner quantum number can only be by a factor of 0 or +1, or −1.

J scan See J DISPLAY.

J scope A radar device that produces a J DISPLAY.

JSR Abbreviation of *jump to subroutine.*

JTAC Abbreviation of *Joint Technical Advisory Committee.*

judder In facsimile transmission, distortion caused by movements of the transmission or reception equipment.

juice Colloquialism for electric voltage or current.

jukebox An automatic phonograph (usually found in public places) that contains a large assortment of records.

Juliet Phonetic representation for the letter J, used in voice communications.

jump **1.** To purposely provide a short circuit around some component or circuit. **2.** Also called *branch.* In digital computer operations, a programming instruction specifying the memory location of the next instruction and directing the computer to it.

jump, conditional See CONDITIONAL BRANCH.

jump, unconditional See UNCONDITIONAL BRANCH INSTRUCTION.

jumper A short piece of wire (usually flexible, insulated, and equipped with clips) for jumping a component or circuit. See JUMP, **1**.

jump instruction See BRANCH INSTRUCTION.

junction **1.** A joint (connection) between two conductors. **2.** The region of contact between semiconductor materials of opposite type (e.g., *pn junction*). **3.** A waveguide fitting used to attach a branch waveguide to a main waveguide at an angle.

junction barrier See DEPLETION REGION.

junction battery A nuclear battery in which a silicon pn junction is irradiated by strontium 90.

junction box A (usually metal) protective box or can into which several conductors are brought together and connected.

junction capacitance In a semiconductor pn junction, the internal capacitance across the junction; it is of special interest when the junction is reverse-biased. Also called *barrier capacitance.*

junction capacitor See VOLTAGE-VARIABLE CAPACITOR.

junction diode A semiconductor diode created by joining an n-type region and p-type region as a wafer of semiconductor material, such as germanium or silicon.

junction field-effect transistor Abbreviation, JFET. A field-effect transistor in which the gate electrode consists of a pn junction.

junction filter A combination of separate low- and high-pass filters having a common input, but separate outputs. The filter is used to separate two frequency bands and transmit them to different circuits.

junction laser See LASER DIODE.

junction light source See LIGHT-EMITTING DIODE.

junction loss **1.** The loss that occurs in a telephone circuit at connecting points. **2.** Loss in a semiconductor pn junction. **3.** Loss occurring at an electrical connection because of poor bonding.

junction photocell A photoconductive or photovoltaic cell that is essentially a light-sensitive JUNCTION DIODE.

junction point **1.** A point at which two or more conductors, components, or circuits join in an electrical connection. **2.** The point in a computer routine at which one of several choices is made.

junction rectifier A semiconductor rectifier that is, in effect, a heavy-duty junction diode, or the equivalent of several heavy-duty junction diodes in combination.

junction station A microwave relay station joining one or more microwave radio legs to the main route or through route.

junction transistor A transistor in which the emitter and collector consists of junctions (see JUNCTION, **2**) between p and n semiconductor regions. Compare POINT-CONTACT TRANSISTOR.

junctor In a crossbar system, a circuit that bridges frames of a switching unit and terminates in a switching device on each frame.

Jungian-world theory The theory that societies keep repeating the same mistakes, generation after generation. Of interest to some researchers in artificial intelligence. It has been suggested that supercomputers might help find solutions to recurring social problems.

justification **1.** The alignment of text along the left margin, the right margin, or both margins. **2.** A method of altering the speed of a digital signal so that it can be received by equipment designed for a different data speed. The rate of speed is lower after this process is applied. Also called *pulse stuffing.*

justify **1.** To adjust the printing of words for aligned left and/or right margins. **2.** In computer operations, to shift an item in a register so that the most- or least-significant digit is at the corresponding end of the register.

just-operate value The current or voltage level at which a relay or similar device just closes. Also called *just-close value.*

just-release value The current or voltage level at which a relay or similar device just opens. Also called *just-open value.*

just scale A musical scale of three consecutive triads, the highest note of each being the lowest note of the next. Each triad has the ratio 4:5:6 or 10:12:15.

jute Tar-saturated fiber, such as hemp, used as a protective covering for cable.

jute-protected cable A cable whose outer covering is a wrapping of JUTE or similar material.

juxtaposed elements Components placed or mounted side by side.

juxtaposed images Images (e.g., those on the screen of a dual-beam cathode-ray tube) that are close to each other for simultaneous viewing, but do not overlap at any point.

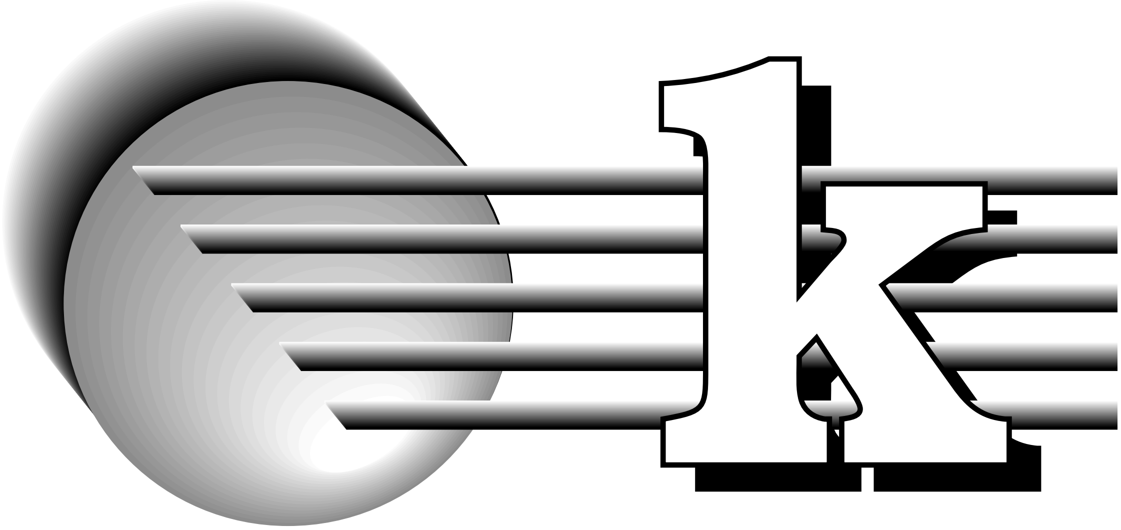

K **1.** General symbol for CONSTANT. **2.** Symbol for POTASSIUM. **3.** Symbol for KELVIN. **4.** Radiotelegraph symbol for *go ahead* or *over.* **5.** Symbol for CATHODE. **6.** Symbol for KILOBYTE. **7.** Abbreviation of KILOHM (k is preferred). **8.** Abbreviation of KILO-.

k **1.** Abbreviation of KILO-. **2.** Symbol for the BOLTZMANN CONSTANT. **3.** General symbol for a mathematical constant. **4.** Symbol for DIELECTRIC CONSTANT. **5.** Abbreviation for KILOHM.

kA Abbreviation of KILOAMPERE.

Karnaugh map A logic chart showing switching-function relationships and used in computer logic analysis to determine quickly the simplest form of logic circuit for a given function. The Karnaugh map is sometimes regarded as a tabular form of the Venn diagram.

Kansas City standard A frequency-shift modulation standard for a computer/tape-recorder interface. Also called *byte standard.*

KB **1.** Abbreviation of KEYBOARD. **2.** Abbreviation of KILOBYTE.

K band The 18- to 27-GHz band of frequencies, pertaining to their use in radar applications.

kcal Abbreviation of KILOCALORIE.

K carrier system A four-wire carrier-current telephone system using frequencies up to 60 kHz and providing 12 channels.

kCi Abbreviation of KILOCURIE.

kcs Abbreviation of *kilocharacter per second* (1000 characters per second).

K display See K SCAN.

KDP Abbreviation of POTASSIUM DIHYDROGEN PHOSPHATE, a ferrolelectric material.

keeper A small iron bar placed across the poles of a permanent magnet to forestall demagnetization.

K electron In certain atoms, one of the electrons whose orbit is nearest the nucleus.

Kel-f Abbreviation of *polymonochlorotrifluorethylene,* a high-temperature insulating material.

Kellie bond **1.** The junction of two mated electrically conductive surfaces that are held together by an adhesive that exhibits negligible resistance thermally and electrically when set. **2.** To make a thermally conductive joint, as from a heatsink to a chassis or component.

kelvin (Lord Kelvin, 1824–1907) Symbol, K. The SI unit of thermodynamic temperature; 1 K = 1/273.16, the thermodynamic temperature of the triple point of water.

Kelvin absolute electrometer An electrostatic voltmeter consisting of a movable metal plate (or plates) and a stationary metal plate (or plates)

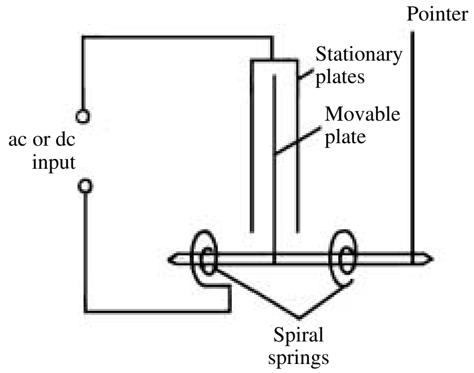

Kelvin absolute electrometer

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

between which voltage is applied. The movable plate(s) is displaced over a distance that is proportional to the potential, against the torque of a return spring.

Kelvin balance An apparatus for measuring current in terms of magnetic pull. A coil is attached to each end of the beam of a balance, and coils ride directly above two stationary coils. Current flows through all coils, making one pair attractive and the other repellent, thus unbalancing the beam. Balance is restored by sliding a weight whose position along a graduated scale indicates the current strength.

Kelvin contacts Electrical contacts designed to eliminate the effect of lead resistance on the accuracy of measurement. Two leads run to each test point, one lead carrying the test signal and the other leading directly to the measuring instrument.

Kelvin double bridge A special bridge for measuring very low resistance (0.1 ohm or less). The arrangement of the bridge reduces the effects of contact resistance that causes significant error when such low resistances are connected to conventional resistance bridges.

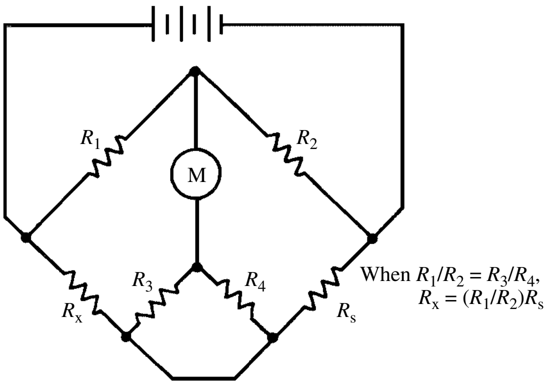

When $R_1/R_2 = R_3/R_4$,
$R_x = (R_1/R_2)R_s$

Kelvin double bridge

Kelvin replenisher A static generator consisting of two pairs of concentric, rotating semicircular conductors connected to brushes. The machine can be regarded as a rotating electric doubler.

Kelvin scale See ABSOLUTE SCALE.

Kelvin temperature scale See ABSOLUTE SCALE.

Kelvin voltmeter An electrostatic voltmeter in which an assembly of figure-8-shaped metal plates rotates between the plates of a stationary assembly when a voltage is applied between the assemblies. The length of the arc of rotation is proportional to the electrostatic attraction and, thus, to the applied voltage.

Kendall effect Distortion in a facsimile record caused by unwanted modulation produced by a carrier signal.

Kennelly-Heaviside layer (A.E. Kennelly, 1861–1939; Oliver Heaviside, 1850–1925) An early name for ionized regions in the upper atmosphere. These regions reflect and refract radio waves at certain frequencies. There are several layers at various altitudes. Also see IONOSPHERE and IONOSPHERIC LAYERS.

keraunograph A meteorological instrument for detecting distant electrical storms. In its simplest form, it consists of a galvanometer connected in series with an antenna and ground.

keraunophone A radio-receiving KERAUNOGRAPH.

kernel Inside an electrical conductor, a line along which the magnetic field strength is zero. Generally, this line is near the center of the conductor.

Kerr cell A nitrobenzene-filled cell that makes use of the KERR ELECTRO-OPTICAL EFFECT. It can function as an electric light shutter or control.

Kerr electro-optical effect The tendency of certain dielectric materials to become double-refracting in an electric field.

Kerr magneto-optical effect The tendency of glass and some other solids and liquids to become double-refracting in a magnetic field.

Kerst induction accelerator See BETATRON.

keV Abbreviation of KILOELECTRONVOLT.

Kew magnetometer A special magnetometer used to measure the intensity of the earth's magnetic field, and also the magnetic declination at a given point on the earth's surface. The device is designed for very high accuracy, using magnifying lenses.

key **1.** See DECRYPTION KEY. **2.** A projection or pin that guides the insertion of a plug-in component into a holder or socket. **3.** A digit or digits used to locate or identify a computer record (but not necessarily part of the record). **4.** A specialized hand-operated switch, used to make and break a circuit repetitively to form the dot and dash signals of Morse code telegraphy. Primarily of historical interest. Used by some amateur radio operators for hobby purposes. **5.** Slang for *principal* or *main.*

keyboard An array of lettered or numbered, low-torque push buttons, used to enter information into a computer, telegraph, teletypewriter, or automatic control system.

keyboard computer A digital computer in which the input device is an electrical keyboard of the typewriter or calculator type.

keyboard entry The operation of a keyboard to enter information into a computer for processing.

keyboard keyer A device for automatically sending Morse code using a typewriter-like keyboard, rather than a paddle or straight key. Each key on the keyboard, when pressed, produces the complete character and a space following it. Most keyboard keyers have buffers to allow typing well ahead of the code being sent, with insertion of all the correct spaces. The speed range is usually from about 5 words per minute (wpm) to 60 or 70 wpm, although some keyboard keyers are programmed for speeds over 100 wpm.

keyboard lockout A keyboard interlock in a data transmission circuit that prevents data from being transmitted while the transmitter of another station on the same circuit is operating.

keyboard send-receive unit A teletypewriter lacking an automatic input device.

key cabinet In a telephone system, a facility that shows a subscriber which lines are busy and which lines are open.

key chirp A chirping sound in a received signal, resulting from slight frequency shift when a radiotelegraph transmitter is keyed. It does not occur with a well-designed transmitter.

key-click filter An inductance-capacitance (LC) or resistance-capacitance (RC) filter for smoothing a keying wave to eliminate KEY CLICKS. It functions by optimizing the rise and decay times of the keyed waveform.

key clicks Excessive bandwidth of a radiotelegraph signal that can result when a keyed signal has rise and decay times that are too rapid. Produces characteristic clicking or popping sounds, with resulting interference, in receivers tuned to frequencies near that of the transmitted signal. A KEY-CLICK FILTER can eliminate this.

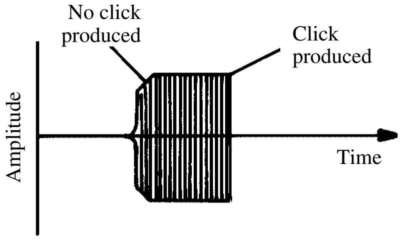

key clicks

keyed AGC A controlled automatic gain control (AGC) system in a television receiver circuit. The AGC acts when the horizontal sync pulse appears; it is inactive between pulses. This prevents unwanted actuation of the AGC by noise transients and picture-signal elements.

keyed clamp A clamping circuit that uses a control signal to determine the clamping time.

keyed interval In a transmission system that is keyed periodically, an interval beginning with a change in state and having a duration of the shortest time between changes in state.

keyed rainbow generator For testing of color-television receivers, a signal generator that produces a rainbow color pattern on the screen (i.e., a set of 10 vertical color bars representing the spectrum, with blank bars in between). The pattern results from gating the 3.56-MHz oscillator in the receiver at a frequency of 189 kHz.

keyer An automatic device for keying a radiotelegraph transmitter or wire telegraph circuit. The keyer can operate from perforated tape, an embossed disk, magnetic tape, or other similar recording.

keyer adaptor A modulated-signal detector that produces a direct-current signal having an amplitude sympathetic with the modulation; it provides the keying signal for a frequency-shift exciter in radio facsimile transmission.

key escrow A controversial system in which the government is provided with certain components of decryption keys to all communications ciphers, according to laws that would allow the government to eavesdrop on private communications or transactions after getting a court order. The key components are held in a secure place, that is, in “escrow,” unless and until the necessary court order is obtained. See DECRYPTION KEY.

keying **1.** The modulation of a carrier by switching it on and off. It is commonly used in radiotelegraphy. **2.** The modulation of a carrier by switching its frequency between two defined values. It is also called FREQUENCY-SHIFT KEYING (FSK) and it is used in data transmission. **3.** The modulation of a carrier with an audio tone that is switched on and off. The carrier can be modulated via amplitude modulation (AM), frequency modulation (FM), pulse modulation, or any other form that will convey the audio tone. It is occasionally used in radiotelegraphy at very-high frequencies (VHF) and above. Also called *audio keying.* **4.** The modulation of a carrier with an audio tone whose frequency is switched between two defined values. The carrier can be modulated via amplitude modulation (AM), frequency modulation (FM), pulse modulation, or any other form that will convey the audio tone. It is commonly used in data transmission at VHF and above. Also called AUDIO FREQUENCY-SHIFT KEYING (AFSK).

keying chirp A rapid change in the frequency of a continuous-wave signal, occurring at the beginning of each code element. In the receiver, the resulting sound is a chirp.

keying error rate In data transmission, the ratio of incorrectly keyed signals to the total number of signals keyed.

keying filter See KEY-CLICK FILTER.

keying frequency **1.** In audio-keyed radiotelegraphy, the audio frequency (tone) of the dot and dash signals (as opposed to the carrier frequency). **2.** In radiotelegraphy, the transmission speed (see KEYING SPEED). **3.** The number of times per second that a black-line signal occurs while an object is scanned in a facsimile system.

keying monitor A simple detector used by an operator to listen to the keying of a radiotelegraph transmitter.

keying speed The speed (in words per minute) of a telegraph or radio-telegraph transmission.

keying transients **1.** Transients arising from the keying of a radiotelegraph transmitter or wire telegraph circuit. **2.** Transients that arise from the repetitive making and breaking of any circuit.

keyless ringing In a telephone system, ringing that begins as soon as the calling plug is put in the appropriate jack on the jack panel.

key pulse In telephone operations, a signaling system in which the desired numbers are entered by pressing corresponding pushbuttons or keys.

key punch A keyboard-operated machine for recording information by perforating a tape or cards.

keyshelf A shelf that supports manually operated telephone switchboard keys.

key station The master (control) station in a communications or control network.

keystoning A form of video image distortion in which the top of the picture is wider than the bottom, or vice versa. Thus, the image area is shaped like a trapezoid, rather than a rectangle.

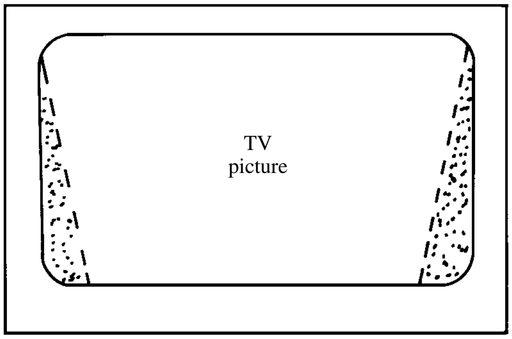

keystoning

key switch **1.** A lockable switch that is operated by inserting and turning a key in it. **2.** A switch having a long handle that transmits motion to the mechanism through a cam. **3.** See KEY, **1**. **4.** The separate short-circuiting switch sometimes mounted on the base of a telegraph key (see KEY, **1**). **5.** A switch that actuates the keys of an electronic organ.

key-to-disk unit A keyboard-to-magnetic-disk data-processing unit.

keyway A groove or slot into which a mating key slides to position a plug-in component (see KEY, **2**).

keyword In information retrieval systems, the significant word in the title describing a document (e.g., in the title "A Primer on French Cuisine," the word *cuisine* would be the keyword, the others having no singular significance).

kg Abbreviation of KILOGRAM.

kgC Abbreviation of KILOGRAM-CALORIE.

kgm Abbreviation of KILOGRAM-METER.

kg/m^3 Symbol for KILOGRAMS PER CUBIC METER, the SI unit of density.

K-H layer See KENNELLY-HEAVISIDE LAYER.

kHz Abbreviation of KILOHERTZ.

kick **1.** To place into sudden operation, as by the quick, forcible closure of a switch or the rapid application of an enabling pulse. **2.** See TRIGGER. **3.** Colloquialism for an abrupt, momentary electric shock.

kickback **1.** The counter EMF that appears across an inductor when current is interrupted. **2.** See FLYBACK.

kickback power supply A high-voltage power supply using the flyback principle. See FLYBACK.

kidney joint A waveguide coupling used in radar. The joint is flexible, or can consist of an air gap, to allow rotation of the antenna.

Kikuchi lines A characteristic spectral pattern produced by the electrons scattered when an electron beam strikes a crystal.

killer **1.** A pulse or other signal used to disable a circuit temporarily (e.g., a blanking pulse). **2.** In personal computing, an application of such importance that it alone serves as the motivation for someone to purchase or upgrade a system. Common examples are word processing and online communications.

killer circuit **1.** A circuit that disables some function of a system, such as the audio in a television receiver. **2.** The blanking circuit in a radar receiver. **3.** A circuit that prevents responses to side-lobe signals in a repeater or transponder.

Kilo Pronunciation, *KEY-low.* Phonetic alphabet code word for the letter K.

kilo- Abbreviation, K or k. **1.** A prefix meaning thousand(s). **2.** In digital data applications, a prefix meaning 2^{10} (1024).

kiloampere Abbreviation, kA. A unit of current equal to 1000 amperes.

kilobit A unit of digital data equal to 2^{10} (1024) bits. Also see BIT.

kilobyte Abbreviation, K or KB. A unit of digital data equal to 2^{10} (1024) bytes. Also see BYTE.

kilocalorie Abbreviation, kcal. A large unit of heat; 1 kcal equals 1000 calories. See CALORIE.

kilocurie Abbreviation, kCi. A large unit of radioactivity equal to 3.71×10^{13} disintegrations per second; 1 kCi equals 1000 curies. Also see CURIE, MEGACURIE, MICROCURIE, and MILLICURIE.

kilocycle See KILOHERTZ.

kiloelectronvolt Abbreviation, keV. A large unit of electrical energy equal to 1000 electronvolts. See ELECTRONVOLT.

kilogauss A large unit of magnetic flux density; 1 kilogauss equals 1000 gauss or 0.1 tesla.

kilogram Abbreviation, kg. The SI base unit of mass; it is equal to 1000 grams.

kilogram-calorie The heat required to raise 1 kilogram of water 1°C.

kilogram-meter Abbreviation, kgm. A unit of mechanical energy (work); 1 kgm is the energy required to raise a mass of 1 kilogram vertically by a distance of 1 meter (equal to 7.2334 foot-pounds). Also see JOULE.

kilohertz Abbreviation, kHz. A unit of frequency; 1 kHz equals 1000 Hz.

kilohm Symbol, kΩ. A unit of high resistance, reactance, or impedance; 1 kΩ equals 1000 ohms.

kilojoule Abbreviation, kJ. A unit of energy or work; 1 kJ equals 1000 joules. See JOULE.

kilolumen Abbreviation, klm. A unit of luminous flux equal to 1000 lumens. See LUMEN.

kilomega- See GIGA-.

kilomegahertz See GIGAHERTZ.

kilometer Abbreviation, km. A large metric unit of linear measure; 1 km equals 1000 meters (3280.8 feet).

kilo-oersted Abbreviation, kOe. A unit of magnetic field strength; 1 kOe equals 1000 oersteds. See OERSTED.

kiloroentgen Abbreviation, kr. A large unit of radioactive radiation; 1 kr equals 1000 roentgens. See ROENTGEN.

kilorutherford Abbreviation, krd. A large unit of radioactivity equal to 10^9 disintegrations per second.

kilovar A compound term coined from *kilo-* and *VAR* (the abbreviation of *volt-amperes reactive*). It is equal to a reactive power of 1000 watts.

kilovar-hour A large unit of reactive electrical energy, equivalent to 1000 reactive watts manifested for a period of one hour.

kilovolt Abbreviation, kV. A unit of high voltage; 1 kV equals 1000 V.

kilovolt-ampere Abbreviation, kVA. A unit of high power that gives the TRUE POWER in a direct-current circuit and the APPARENT POWER in an alternating-current circuit; 1 kVA equals 1000 W. Also see DC POWER.

kilovolt-ampere reactive See KILOVAR.

kilovoltmeter A voltmeter designed to measure thousands of volts (kilovolts).

kilowatt Abbreviation, kW. A unit of high power; 1 kW equals 1000 watts. Also see WATT.

kilowatt-hour Abbreviation, kWh. A common unit of electrical energy; 1 kWh equals 1000 watt hours, or the equivalent of 1000 watts dissipated for a period of one hour. Also see ENERGY, KILOWATT-HOUR, POWER, WATT-HOUR, and WATT-SECOND.

kilowatt-hour meter A motorized meter for recording (electrical) power consumption in kilowatt-hours. Also see KILOWATT-HOUR.

kinematograph A motion picture camera. Also called CINEMATOGRAPH and KINETOGRAPH.

kine **1.** See KINESCOPE, **1**. **2.** See KINESCOPE RECORDING.

kinescope **1.** The picture tube in a television receiver. **2.** See KINESCOPE RECORDING.

kinescope recorder A film or tape apparatus for recording television pictures.

kinescope recording A motion-picture or videotape made from the screen (or taken from the circuit) of a television picture tube.

kinetic energy The energy associated with particles, bodies, or electric charge carriers in motion.

kinetograph See KINEMATOGRAPH.

kinetoscope A motion-picture projector.

kiosk A computer and peripherals set up for the purpose of multimedia use by the general public. It generally uses a touch screen for inputting data and must be ruggedly constructed to tolerate rough treatment.

Kirchhoff's first law The sum of the currents flowing out of a point in a direct-current circuit equals the sum of the currents flowing into that point.

Kirchhoff's laws (Gustav Robert Kirchhoff, 1824–1887) Two laws of electric circuits that account for the behavior of certain networks. See KIRCHHOFF'S FIRST LAW and KIRCHHOFF'S SECOND LAW.

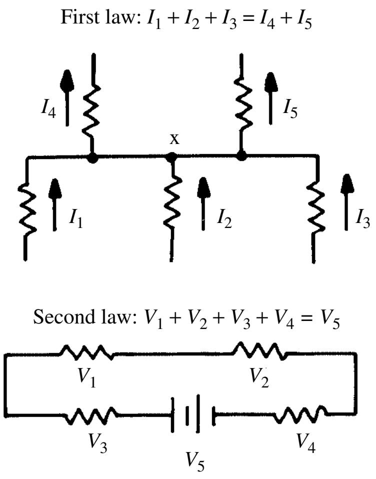
First law: $I_1 + I_2 + I_3 = I_4 + I_5$
I_4
I_5
x
I_1
I_2
I_3
Second law: $V_1 + V_2 + V_3 + V_4 = V_5$
V_1
V_2
V_3
V_5
V_4

Kirchhoff's laws

Kirchhoff's second law The algebraic sum of all the voltage drops around a direct-current circuit (including supply voltages) is always equal to zero.

kit A selection of components, associated equipment, supplies (such as wire and hardware), and instructions for constructing a piece of electronic equipment.

kite-supported antenna A longwire antenna that uses a kite as a support for the far (nonstation) end. A tether is used to reduce the chance that the kite will fly away with wire attached. It was used by Marconi in early experiments with radio. Radio amateurs and shortwave listeners sometimes use this scheme at low and medium frequencies. It is a dangerous antenna because of electrostatic buildup, a tendency to attract lightning, the possibility of its breaking loose, and the risk that it might contact utility lines.

kJ Abbreviation of KILOJOULE.

k-line programming A method by which an artificially intelligent robot can learn as it does a job, so it will have an easier time doing the same job in the future. The robot controller actually learns from the robot's mistakes.

Klipsch horn A loudspeaker that includes a folded low-frequency horn housed in a corner enclosure.

klm Abbreviation of KILOLUMEN.

kludge **1.** A crude, useless, or grossly inefficient machine or process. **2.** A hastily contrived prototype of a circuit or device, put together for the purpose of testing a concept, but not intended as a representative of a production unit.

klydonograph A device that photographically records the voltage gradient in the presence of an electric field.

klystron A microwave tube whose operation is based on the velocity modulation of an electron beam by buncher and cavity reentrant cavities.

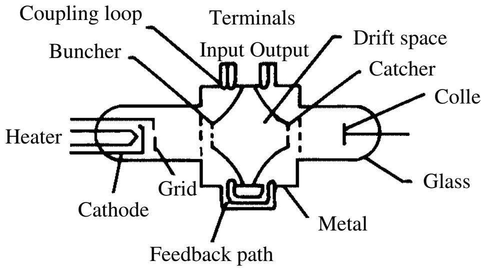

klystron

klystron amplifier A microwave amplifier using a klystron.

klystron harmonic generator A frequency-multiplying power amplifier using a klystron. It is used at microwave radio frequencies.

klystron oscillator A klystron operated as a self-excited microwave oscillator.

klystron repeater A microwave amplifier in which a klystron inserted in a waveguide boosts the amplitude of an incoming signal.

km Abbreviation of KILOMETER.

knee **1.** A sharp bend in a response curve for a device, usually indicating the onset of conduction, saturation, cutoff, pinchoff, or limiting action. It applies especially to semiconductor diodes and transistors. **2.** The characteristics of a device when it is operated at a point in the vicinity of a sharp bend in its response curve.

knee noise Electrical noise generated by rapidly repeating current fluctuations at the knee in a Zener diode.

knife-edge diffraction The lessening of atmospheric signal attenuation when the signal passes over a sharp obstacle and is diffracted.

knife switch A switch composed of one or more flat blades roughly resembling knife blades, which are slid firmly between the jaws of pinching contacts to close a circuit.

knob **1.** A (usually round and insulated) finger dial for adjusting a variable electronic component, such as a potentiometer, variable capacitor, or rotary switch. **2.** A solid round insulator usually having a low diameter to height ratio. **3.** A small ball- or rod-shaped electrode or protuberance.

knocker A fire-control radar subassembly of synchronizing and triggering circuits.

knockout An area in a metal box or chassis that is easily removed by tapping or knocking to provide an opening.

knot A unit of speed, corresponding to 1 nautical mile per hour. A speed of 1 knot is about 1.15 statute miles per hour; a speed of 1 statute mile per hour is about 0.868 knots. It is used by mariners for specifying speeds at sea, and also occasionally by meteorologists in specifying wind speeds.

knowledge The data in a computer and in mass-storage media, accumulated over time and capable of being put to practical use.

kOe Abbreviation of KILO-OERSTED.

Kolster decremeter An absorption wavemeter with a movable scale; it permits measurement of the decrement of a radio wave.

Kooman antenna A unidirectional antenna consisting of stacked, full-wave, center-fed driven elements, and a reflecting screen.

Korner Killer Trade name for an acoustically absorbent object that reduces sound echoes that can occur in enclosed rooms. The name results because the device works best when placed in a corner (where two walls meet).

Kovar An alloy of cobalt, iron, and nickel. It is used mostly in glass-to-metal seals because it has characteristics of both kinds of material.

Kozanowski oscillator A positive-grid vacuum-tube UHF oscillator circuit using two tubes having cylindrical elements, and a pair of parallel-wire tanks.

Kr Symbol for KRYPTON.

kr Abbreviation of KILOROENTGEN.

K radiation X rays emitted from an atom when an electron becomes a K electron.

kraft paper Strong brown paper used for insulation and as the dielectric of paper capacitors.

Kramer system A system of three-phase motor control providing constant horsepower, and having a direct-current (dc) motor coupled to the shaft of a wound-rotor three-phase induction motor. The dc supply for the motor also supplies a rotary converter. The speed-control rheostat is connected in series with the field of the motor and the dc power supply.

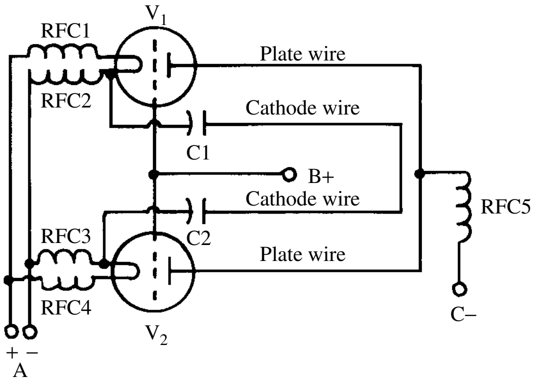

Kozanowski oscillator

Kraus antenna A bidirectional, flat-top beam antenna consisting of a pair of closely spaced dipoles. Several such sections can be connected in series by crisscrossing the wires at voltage loops.

krd Abbreviation of KILORUTHERFORD.

Kryptol A mixture of clay, graphite, and silicon carbide. It is used in electric heater elements because of its low resistance and high melting point.

krypton Symbol, Kr. An inert, gaseous element. Atomic number, 36. Atomic weight, 83.80. Krypton is present in trace amounts in the earth's atmosphere.

K scan In radar operations, a modified A-scan used in aiming antennas in which two pips are displayed; their relative amplitudes indicate the antenna-aiming error.

K series A series of spectral lines for the shortest wavelengths of radiation from the innermost electron shell of a radiating atom.

KSR Abbreviation of *keyboard send-receive* unit.

Ku band A band of microwave frequencies between approximately 12 and 18 GHz.

Kundt's law The index of refraction of a medium does not change continuously with wavelength in the absorption bands.

Kundt tube A device used to measure the speed of sound in gases under various conditions. Suspended particles in the gas form standing waves that can be easily seen. Knowing the frequency of the disturbance and the distance between nodes of the standing waves, the speed can be determined. The pressure and density of the gas, as well as temperature and humidity, affect the speed.

kurchatovium See RUTHERFORDIUM.

kV Abbreviation of KILOVOLT.

kVA Abbreviation of KILOVOLT-AMPERE.

kVAR Abbreviation of REACTIVE KILOVOLT-AMPERE.

kVARh Abbreviation of KILOVAR-HOUR.

kW Abbreviation of KILOWATT.

kWh Abbreviation of KILOWATT-HOUR.

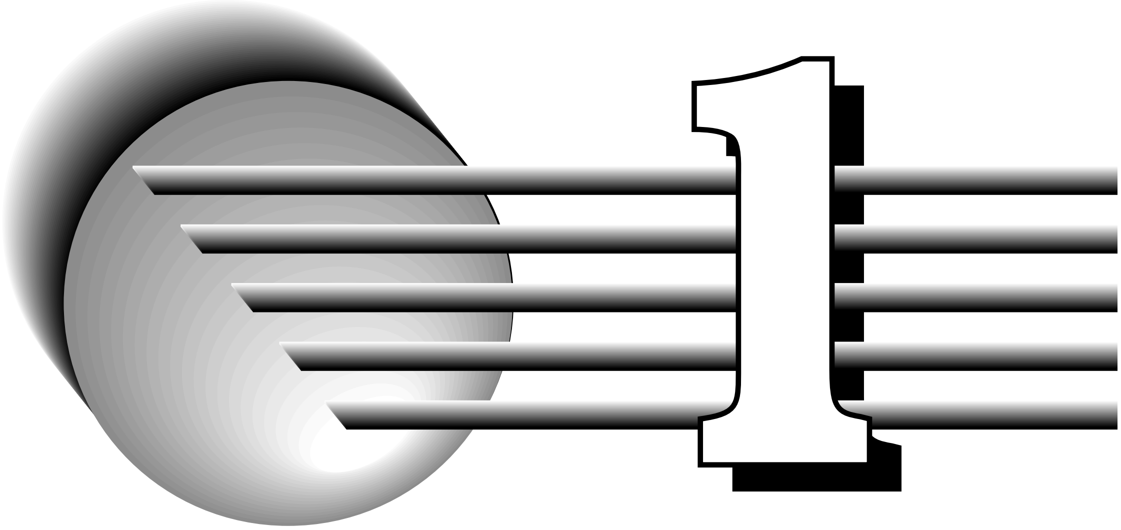

L **1.** Symbol for INDUCTANCE. **2.** Symbol for MEAN LIFE. **3.** Abbreviation of LOW. **4.** Resembling the capital letter L in physical shape. **5.** Symbol for LAPLACE TRANSFORM.

l **1.** Symbol for LENGTH. **2.** Abbreviation of LITER. **3.** Subscript for LOW. **4.** Abbreviation of LUMEN; also abbreviated *lm* (preferred) and *lum.*

La Symbol for LANTHANUM.

label **1.** A symbolic group of characters that identifies an area of memory, an item of data, a file, or a record. **2.** A name assigned to a source program instruction step to identify the step as a coding entry point, or to make the step usable as a reference point for entry to the routine or subroutine in which it appears.

label group A collection of labels, usually of the same type, held in an operating system.

label identifier Within a label, a character set used to name the kind of item labeled.

label record A record identifying a file recorded on a magnetic storage medium (e.g., magnetic tape).

label set A collection of labels having a common label identifier.

labile oscillator A frequency-controlled local oscillator.

laboratory conditions The environmental, mechanical, and electrical parameters characteristic of controlled conditions. Actual operating conditions can be much different.

laboratory-grade instrument An instrument having the high accuracy and stability that suit it to precision measurements in a laboratory. Also called PRECISION INSTRUMENT. Compare SERVICE-TYPE INSTRUMENT.

laboratory power supply A regulated direct-current source whose adjustable output is less than 10 kV at no more than 500 W.

laboratory standard See PRIMARY STANDARD and SECONDARY STANDARD.

labyrinth speaker A loudspeaker whose enclosure (a wooden cubicle) includes a folded pipe or acoustic transmission line (behind the speaker); the inner walls are lined with a sound-absorbent material. When the pipe, which is open-ended, is half as long as the wavelength of the frequency being reproduced, the sound emerging from the open end is in phase with that radiated by the front of the speaker and, therefore, reinforces it.

laced wiring Circuit wiring in which wires or cables run parallel in bundles that are tied together with LACING CORD.

lacing cord Strong, sometimes waxed cord used to tie together wires running parallel in a bundle. Also see LACED WIRING.

lacquer disk See CELLULOSE-NITRATE DISK.

lacquer-film capacitor A fixed capacitor with a plastic film dielectric; the film is applied as liquid lacquer to the metal foil.

lacquer master A master recording made on a CELLULOSE-NITRATE DISK.

lacquer original See LACQUER MASTER.

ladar Abbreviation of LASER DOPPLER RADAR. Also abbreviated *lopplar.*

ladder attenuator See LADDER-TYPE ATTENUATOR.

ladder filter A form of delay line or filter. It generally consists of series and parallel impedances, either in a balanced or unbalanced form.

ladder line See AIR-INSULATED LINE.

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

ladder network A network consisting of several L sections in cascade. See L SECTION, **1**, **2**, **3**.

ladder-type attenuator An attenuator consisting of a ladder network equipped with a switching circuit for selecting the output at various sections.

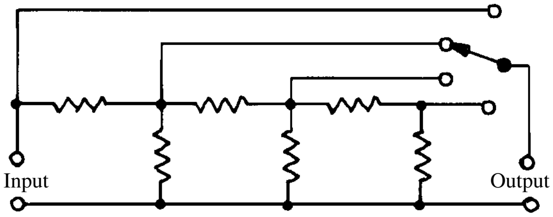

ladder-type attenuator

LAFOT Coded weather broadcasts aired every six hours by the U.S. Weather Bureau through marine radiotelephone broadcasting stations for the Great Lakes region.

lag In computations relating the phase of alternating-current signals, the extent to which one quantity follows another in time (e.g., the current lags the voltage by 90 degrees in a pure inductive reactance). Compare LEAD.

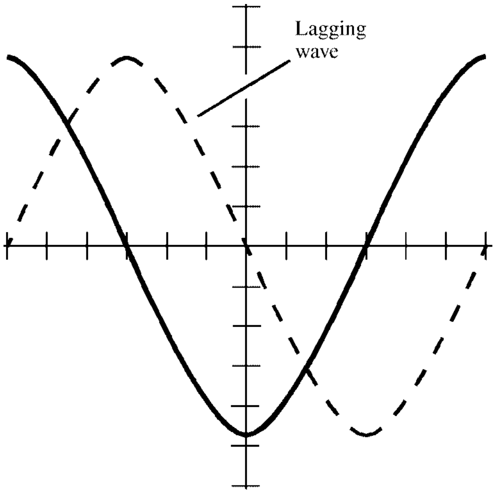

lag

lagged-demand meter A meter with a built-in time delay.

lagging current Current that follows voltage (in time). Also see LAG.

lagging load A load in which current lags behind voltage (i.e., an inductive load). Compare LEADING LOAD.

lag network A phase-shifting circuit containing series-resistance and shunt-capacitance arms. It produces a lagging phase shift. Compare LEAD NETWORK.

lambda wave An electromagnetic disturbance that travels along the surface of an object. An example is the surface wave characteristic of low-frequency propagation.

lambert Symbol, L. The centimeter-gram-second (cgs) unit of luminance, equal to the brightness of an ideal diffusing surface that radiates or reflects light at 1 lumen per square centimeter. The SI (preferred) unit of luminance is the candela per square meter (cd/m^2); 1 lambert equals 10^4 cd/m^2. Also see CANDELA.

Lambert's law of illumination The illumination of a surface by a point light source is inversely proportional to the square of the distance between the surface and the source. If the surface is not perpendicular to the rays, the illumination is proportional to the cosine of the angle of incidence.

laminated armature An armature for a motor or generator, made of stacked laminations.

laminated contact A switch contact consisting of a number of laminations—each contacting a conducting counterpart.

laminated core A core for a transformer, choke, relay, or similar device, made of stacked laminations.

laminated disk A layered recording disk.

laminated pole A pole within a motor, generator, relay, electromagnet, or similar device, made of stacked laminations.

lamination **1.** A relatively thin sheet of metal cut to a required shape to be stacked with other similar sheets to form a laminated core or pole. **2.** A relatively thin sheet of plastic that is bonded together and heat-formed with other similar sheets to produce a sheet or piece of desired thickness and strength.

lamp A device for converting electrical energy into visible light. The term includes a number of devices (e.g., arc lamp, fluorescent tube, incandescent lamp, mercury-vapor lamp, and neon bulb).

lamp-bank resistor A makeshift heavy-duty resistor consisting of several incandescent lamps arranged so that they can be switched in various series, parallel, and series-parallel combinations to vary the resistance provided by the filaments.

lampblack Carbon obtained from soot deposited by a smoky flame. The substance is used as the basic material for some resistors.

lamp cord A two-wire insulated cord, used with low-wattage alternating-current appliances at 117 volts. The wire is usually stranded copper equivalent to American Wire Gauge (AWG) #16.

lamp dimmer See DIMMER.

lamp driver A usually single-stage circuit for amplifying a small pulse to drive an indicator lamp.

lamp extractor A special tool used to insert or extract miniature lamps for electronic equipment.

lamp jack A receptacle with a spring release that holds a small incandescent bulb. The bulb is removed and replaced by pushing and twisting.

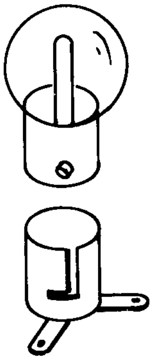

lamp jack

lamp-type expander A volume expander in which the tungsten filament in an incandescent lamp serves as the nonlinear resistor.

lamp-type readout For counters, calculators, and digital meters, a readout device in which each digit is indicated by a lamp.

LAN Abbreviation of LOCAL AREA NETWORK.

land **1.** The flat, reflective surface between pits on a compact disc (CD). Compare PIT, **1**. **2.** The thin vinyl wall between grooves on a phonograph record. **3.** A bonding point in a microcircuit. **4.** Pertaining to earthbound communications stations.

Land camera See POLAROID CAMERA.

landing beacon The aircraft landing-beam transmitter. Also see LANDING BEAM.

landing beam A highly directional airport radio signal beamed upward to guide aircraft landing during conditions of poor visibility.

landline A telephone or telegraph circuit completed with wires.

landmark beacon Any beacon that is not an airway or airport beacon.

land mobile service Two-way radio service between a base station and mobile land vehicles, or among mobile land vehicles.

land mobile station A radio station aboard a mobile, earthbound vehicle.

land return Ground reflection of radar signals back to the transmitter.

land station A fixed ground station.

Langevin ion An electrically charged particle, such as a grain of dust or droplet of water, resulting from the accumulation of ions.

Langmuir dark space In a luminous gas discharge, the dark region around a negatively charged probe inserted into the positive column. Compare CROOKES' DARK SPACE.

Langmuir's law See CHILD'S LAW.

language In digital-computer operations, any one of the detailed systems for representing data, instructions, and procedures through the use of symbols and symbol sequences. See MACHINE LANGUAGE, ASSEMBLY LANGUAGE, COBOL, FORTRAN, and BASIC.

language laboratory An electronic contribution to the teaching and learning of languages. It consists of recordings in a language being studied and all the equipment associated with recording, playback, and monitoring. Students listen to the speech of experts in the language record, listen to, and later erase their own utterances in the language.

language translation **1.** The conversion of statements in one computer language to equivalent statements in another. **2.** The conversion of one written natural language into another (e.g., French to Russian) by means of a computer program.

language translator **1.** An assembly program, compiler, or other routine used for translation between computer languages. **2.** A high-level program that allows a computer to translate one written natural language into another (e.g., Chinese to Italian).

L antenna See INVERTED-L ANTENNA.

lantern battery A moderate-sized electrochemical battery usually rated at 6 volts. Derives its name from its original use as a power source for portable lamps. One type has spring contacts on the top. Another type has thumbscrew terminals. The nonrechargeable battery consists of 4 zinc–carbon or alkaline cells in series. Some varieties are rechargeable, consisting of nickel–cadmium (NiCd) or nickel–metal-hydride (NiMH) cells. This type of battery can provide enough energy to operate a low-power radio transceiver. Two units connected in series, or four units in series-parallel, make a 12-volt battery that can power a small portable Citizen Band (CB) or amateur radio station.

lanthanum Symbol, La. An elemental metal of the rare-earth group. Atomic number, 57. Atomic weight, 138.906.

lanyard A wire or cable used to quickly pull apart the halves of a quick-disconnect connector.

lap A device used for grinding piezoelectric crystals for resonance at a desired frequency.

lap dissolve The simultaneous fading out of one televised scene while another is fading in so that one is apparently dissolving into the next. It is also applicable to motion pictures.

lapel microphone A small microphone that is clipped to a lapel of a user's jacket or coat.

lap joint An overlapping splice of two conductors.

Laplace transform Symbol, *L*. An operator that reduces the work of solving certain differential equations by permitting them to be handled by simpler algebraic methods.

lapping Fine-tuning quartz crystal plates by moving them over a flat plate coated with a liquid abrasive.

laptop computer See NOTEBOOK COMPUTER.

lap winding In a motor or generator armature, a winding in which the opposite ends of each coil are connected to the adjoining segments of the commutator.

lap wrap **1.** A form of asbestos cloth wire insulation. **2.** A method of wrapping with electrical tape, in which there is considerable overlap among the turns of the tape.

large calorie See KILOGRAM-CALORIE.

large loop antenna A single-turn open or closed loop, usually having a circumference of 0.5 wavelength or 1 wavelength. With a half-wavelength open or closed loop, maximum radiation occurs in the plane of the loop. A full-wavelength closed loop exhibits maximum radiation and response along the axis. This type of antenna can be used for wireless transmitting and receiving applications. Either the open or the closed half-wavelength loop exhibits a slight power loss relative to a dipole, but the full-wavelength loop shows a small gain over a dipole in its favored directions. Compare SMALL LOOP ANTENNA.

large-scale integration Abbreviation, LSI. The inclusion of more than 100 transistors, performing various individual, but interrelated circuit functions, on a single integrated-circuit chip.

large signal A relatively high-amplitude signal that traverses so large a part of the operating characteristic of a device that nonlinear portions of the characteristic are usually encountered. Compare SMALL SIGNAL.

large-signal analysis The rigorous study of circuits and devices that process large signals.

large-signal component **1.** A coefficient or parameter such as amplification, transconductance, or dynamic resistance, measured under conditions of large-signal operation. Also see LARGE SIGNAL and LARGE-SIGNAL EQUIVALENT CIRCUIT. **2.** A device designed for operation at high signal levels.

large-signal equivalent circuit For a given transistor circuit, the equivalent circuit at high signal levels (i.e., at amplitudes approaching saturation and cutoff levels). Also see EQUIVALENT CIRCUIT.

large-signal operation The use of a circuit or device at signal levels sufficiently high so that nonlinear portions of the characteristic are usually encountered. Compare SMALL-SIGNAL OPERATION.

large-signal transistor See POWER TRANSISTOR.

large-signal voltage gain In an integrated-circuit amplifier, the voltage gain under open-loop conditions, determined as the difference in the output voltage divided by the difference in the input voltage. It is usually specified in volts per millivolt or volts per microvolt.

Larmor orbit The path followed by a charged particle in a constant magnetic field. Because of interaction between the external field and the field generated by the particle, the charged particle travels in a circular path.

laryngaphone See THROAT MICROPHONE.

LASCR Abbreviation of LIGHT-ACTIVATED SILICON-CONTROLLED RECTIFIER.

LASCS Abbreviation of LIGHT-ACTIVATED SILICON-CONTROLLED SWITCH.

lase To emit coherent electromagnetic energy in the visible-light spectrum. See LASER.

laser Acronym for *light amplification by stimulated emission of radiation.* A device that produces coherent radiation in the visible-light range, that is, between 750 and 390 nanometers (one nanometer is 10^{-9} meter). Some devices that produce coherent radiation in the infrared, ultraviolet, or X-ray parts of the spectrum are also referred to as lasers. Lasers can be either continuous or pulsed, and are characterized by coherent, monochromatic emissions. The peak intensity ranges from a few microwatts to many megawatts.

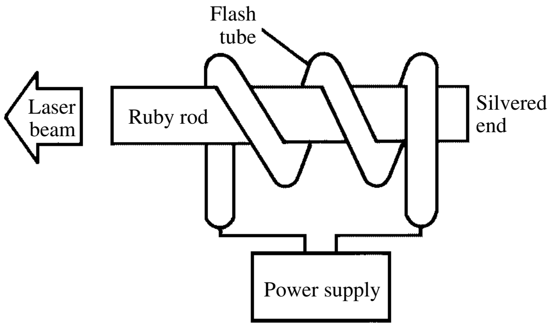

laser

laser beam The radiation from a laser—especially if the divergence is very low, that is, the rays are almost perfectly parallel, resulting in minimal divergence.

laser-beam communication A form of coherent infrared or optical communication in which a laser beam is the link between transmitting and receiving stations. Also see LASER, LASER DIODE, and LIGHT-BEAM COMMUNICATION.

laser capacitor An energy-storage capacitor used to discharge-fire the exciter lamp of a laser. Also see LASER.

laser cavity An optical-resonant cavity that results in the emission of coherent light.

laser cutting A method of using a laser for severing materials.

laser diode A special form of semiconductor light-emitting diode (LED), usually of the gallium-arsenide type, that emits coherent light when a voltage is applied to its terminals. Also see LASER.

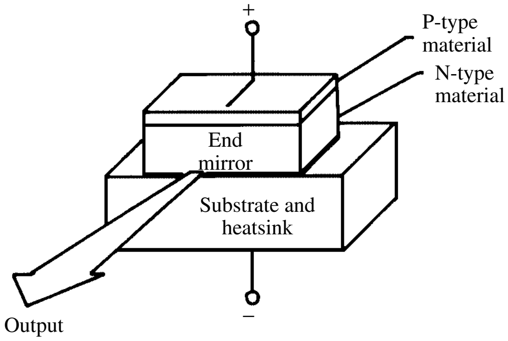

laser diode

laser disk A method of reproducing sound in which a laser is used to recover the sound from a compact disk.

laser Doppler radar Acronyms, *ladar* or *lopplar.* A form of Doppler radar using the light beam of a laser instead of radio waves.

laser eye surgery A method of repairing the retina of the eye without cutting the eyeball, using laser beams to push loose retina tissue back into place.

laser gun A colloquial term for a weapon that makes use of a laser as a device of destruction.

laser optical videodisc system A system in which a low-powered laser reads audio and video information from a videodisc and delivers it to a television receiver.

laser ranger A radar-like device using intense light (instead of microwaves).

laser show A three-dimensional, midair display having motion, made by using lasers in various combinations.

laser surgery The application of a laser in medicine for the purpose of assisting in, or actually performing, operations on human subjects.

laser welding Welding (especially of tiny pieces) with the heat produced by a laser beam.

lasing The emission of coherent electromagnetic energy in the visible-light spectrum. See LASER.

lat Abbreviation of LATITUDE.

latch **1.** A feedback loop in a symmetrical digital circuit, such as a flip-flop, used to maintain a given state. **2.** A simple logic-circuit storage element that consists of two gates as a unit. **3.** To maintain a closed (energized) state in a pair of relay contacts after initial energization from a single electrical pulse. See LATCHING RELAY.

latching current In a thyristor, the minimum value of anode current (slightly higher than the holding current) that will sustain conduction immediately after switch-on.

latching relay An electromechanical or fully electronic relay that locks into whichever mode it is energized for (on or off).

latch-on relay See LOCKING RELAY.

latchup In a transistor switching circuit, the abnormal condition in which the collector voltage remains at its switched-on level after the transistor is switched to cutoff from saturation.

latch voltage The input voltage at which a flip-flop changes states.

late contacts Relay contacts that are operated following the movement of other contacts during the relay's operation.

latency **1.** The time taken by a digital computer to deliver information from its memory. **2.** In a serial storage system, the access time less the word time.

latent image **1.** In a storage tube, a stored image that is not yet visible. **2.** An image stored in the mosaic of an iconoscope. **3.** The image that will appear when photographic film or paper is developed.

lateral chromatic aberration An aberration affecting the sharpness of off-axis color television images.

lateral compliance In phonograph reproduction, the ease with which the stylus can move laterally as it follows the groove. Also see COMPLIANCE and LATERAL RECORDING.

lateral-correction magnet In a color picture tube, a magnet operated with a set of pole pieces attached to the focus element of the blue gun; it controls horizontal positioning of the blue beam for convergence.

lateral magnet See LATERAL-CORRECTION MAGNET.

lateral recording A disc recording in which the groove undulates from side to side. Compare VERTICAL RECORDING.

latitude Abbreviation, lat. Angular distance measured around the earth's circumference to the north and south from the equator. Compare LONGITUDE.

latitude effect The tendency of the earth's magnetic field to decrease the number of charged subatomic particles that reach the surface of the earth near the equator, as compared with the number reaching the surface at other latitudes.

Latour alternator See BETHENOD ALTERNATOR.

lattice **1.** The orderly internal pattern (matrix) of atoms in a crystal. Also see CRYSTAL LATTICE. **2.** A symmetrical arrangement of components in a network (such as an attenuator, a filter, or a bridge circuit).

lattice filter A lattice network having reactance in its arms that makes it a selective circuit.

lattice network See LATTICE, **2**.

lattice section See LATTICE, **2**.

lattice structure See LATTICE, **1**.

lattice winding See UNIVERSAL WINDING.

laue diagram A pattern of spots on a photographic plate produced by the scattering of high energy radiation as it falls on a thin crystal. This pattern is used to determine the nature of the crystal material.

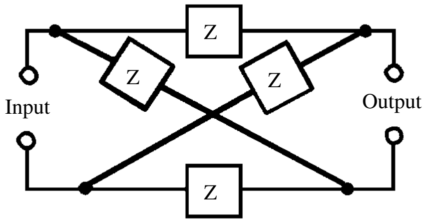

lattice filter

launching The energy transference from a cable into a waveguide.

lavalier microphone A small microphone that can be hung from the user's neck on a cord or chain.

law **1.** A general, verifiable statement that describes the behavior of entities or the relationships between phenomena or concepts. The product of inductive reasoning that follows many observations and controlled experiments (e.g., *first law of thermodynamics*, *inverse-square law*, *Kirchhoff's laws*, and *Ohm's law*. **2.** The nature of the change of a dependent variable, particularly as depicted by a response curve (e.g., *square law*).

LAWEB Civilian weather bulletins issued every six hours from ship and shore positions along the Great Lakes during the sailing season.

lawn mower **1.** A facsimile term for a helix recording mechanism. **2.** A radar receiver preamplifier.

law of a curve See LAW, **2**.

law of averages In probability and statistics, a principle stating that for a large sampling of events, the numerical probability value will be more closely approached than when the sampling is small. Compare LAW OF LARGE NUMBERS.

law of charges Different electric charges attract each other, and similar charges repel each other.

law of electric charges See LAW OF CHARGES.

law of electromagnetic induction See LENZ'S LAW.

law of electrostatic attraction See COULOMB'S LAWS.

law of electrostatic repulsion See COULOMB'S LAWS.

law of first wavefront In acoustics, a rule stating that the perceived direction from which a sound arrives is determined by the first wavefront that reaches the listener's ears.

law of induction See FARADAY'S LAW.

law of inverse squares See INVERSE-SQUARE LAW.

law of large numbers In probability and statistics, a principle stating that with a large sample, the sample average will approximate the population average. It is often erroneously called LAW OF AVERAGES.

law of magnetism Different magnetic poles attract each other, and similar magnetic poles repel each other.

law of natural decay See EXPONENTIAL DECREASE.

law of natural growth See EXPONENTIAL INCREASE.

law of normal distribution Gauss' law of the frequency distribution of a repetitive function, describing the probability of deviations from the mean.

law of octals Chemical activity occurs between two atoms lacking eight valence electrons, and continues until the requirement of eight electrons is satisfied for all but the first orbit, where only two electrons are required. Of interest in the study of semiconductors.

law of radiation See QUANTUM THEORY.

law of reflection For a ray of energy striking a smooth reflective surface, the angle of reflection is equal to the angle of incidence, with respect to a plane tangent to the surface at the point of incidence.

Lawrence accelerator See CYCLOTRON.

lawrencium Symbol, Lr (occasionally Lw). A short-lived radioactive element produced artificially from californium. Atomic number, 103. Atomic weight, about 260.

lay See DIRECTION OF LAY.

layer **1.** A complete coil winding consisting of turns laid side by side (not on top of each other). **2.** In a semiconductor device, a region having unique electrical properties (e.g., *n layer*). **3.** A region of the ionosphere. See IONOSPHERIC LAYERS. **4.** The tape on a reel or in a cassette, encompassing one complete turn (rotation). **5.** In general, a single stratum of a stratified medium.

layer-to-layer transfer In a roll of magnetic tape, unwanted transfer of data between adjacent turns on the reel. If severe, this transfer can cause drop-in or drop-out in a computer. In audio applications, it can sometimes be heard as a delayed echo or a faint sound occurring just prior to the actual recorded sound.

layer winding A coil winding in which the turns are arranged in two or more concentric layers.

layerwound coil An inductor wound in layers, one on top of the other. Also see LAYER, **1**. Compare SINGLE-LAYER COIL.

layout The arrangement of components on a chassis, printed circuit board, or panel.

lazy-H antenna An antenna consisting of two vertically stacked collinear elements, producing both horizontal and vertical directivity.

Lb Abbreviation of LAMBERT; also, *L* (preferred).

lb Abbreviation of POUND.

L band A radio-frequency band extending from 390 MHz to 1.55 GHz. For subdivisions of this band, see L_C BAND, L_F BAND, L_K BAND, L_L BAND, L_P BAND, L_S BAND, L_T BAND, L_X BAND, L_Y BAND, and L_Z BAND.

LC **1.** Abbreviation of LIQUID CRYSTAL; also abbreviated *lix.* **2.** Abbreviation of INDUCTANCE-CAPACITANCE. **3.** Symbol for LC CONSTANT.

L carrier In a telephone system, a carrier having a frequency between approximately 68 kHz and 10 MHz. It can be used in wire-transmission or radio links.

L_C band A section of the L BAND extending from 465 MHz to 510 MHz.

LC bridge See INDUCTANCE-CAPACITANCE BRIDGE.

LC constant Abbreviation, LC. The product of the inductance and capacitance required for resonance at a given frequency.

LCD Abbreviation of LIQUID-CRYSTAL DISPLAY.

LC filter See INDUCTANCE-CAPACITANCE FILTER.

L circuit See L NETWORK.

LC meter See INDUCTANCE-CAPACITANCE METER.

LCR See INDUCTANCE-CAPACITANCE-RESISTANCE.

LC ratio In a tuned circuit, the ratio of inductance to capacitance.

LCR bridge See IMPEDANCE BRIDGE, **1**.

L_d Symbol for DISTRIBUTED INDUCTANCE.

LDF See LOW-FREQUENCY DIRECTION FINDER.

L display Also called *L scan.* A radar display in which the target appears as two horizontal traces, one extending from a vertical timebase to the right, the other to the left.

lead Pronunciation, *leed.* **1.** A conductor (usually a wire) leading to or emerging from a terminal or electrode. **2.** In computations relating phase, the extent to which one quantity precedes another (e.g., current leads voltage by 90 degrees in a pure capacitance). Compare LAG.

lead Pronunciation, *led.* Symbol, Pb. A heavy metallic element. Atomic number, 82. Atomic weight, 207.2. It can be used as a shield against atomic radiation, and has various applications in electronics (e.g., as electrodes in batteries and as a component of solder). See LEAD-ACID BATTERY, FUSE, and SOLDER.

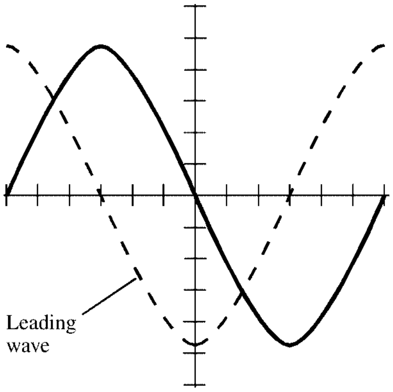

lead, 2.

lead–acid battery A set of two or more lead–acid cells connected in series, usually housed in a common enclosure. Some batteries of this type, notably automotive batteries, are made from sets of lead–acid cells having a free-flowing liquid acid. Other cells have a semisolid "paste" electrolyte. These batteries are popular in consumer electronic devices that require a moderate amount of current. They are also used in uninterruptible power supplies (UPSs) for personal computers. See LEAD–ACID CELL.

lead–acid cell A rechargeable electrochemical cell having an electrolyte of sulfuric acid. The electrodes are lead (negative) and lead dioxide (positive). Produces about 1.5 volts under no-load conditions when fully charged. A large cell of this type can store several tens of ampere-hours. Smaller units have less capacity but more versatility. Their main advantage is reasonable cost, considering that they can be charged and discharged many times.

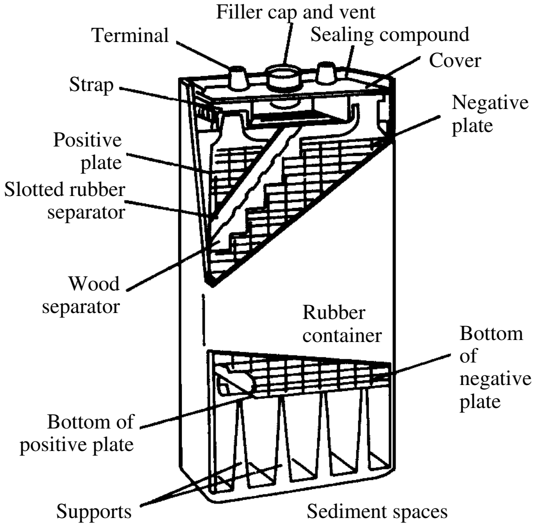

lead-acid cell

lead cell **1.** A lead-acid cell. **2.** A lead-sulfide photocell; see LEAD SULFIDE, **1**.

lead dress See DRESS.

leader **1.** The blank section at the beginning of a magnetic tape. It is usually made of plastic. **2.** A record, preceding a group of records, that identifies the group and provides other data pertinent to the group. **3.** In a lightning stroke, the initial movement of electrons or positive ions, creating the ionized path that allows discharge.

lead frame The metal frame holding the leads of a circuit package (DIP) in place before encapsulation.

lead-in The wire connecting an antenna to a receiver or transmitter.

leading current Current that precedes voltage in time. Also see LEAD.

leading edge The rising edge of a pulse; compare TRAILING EDGE.

leading ghost A twin television image to the left of the original on the screen.

leading load A load in which current leads voltage (i.e., a capacitive load).

lead-in groove Around the outer edge of a phonograph record, a blank spiral groove that leads the stylus into the first groove of the recording. Compare LEAD-OUT GROOVE.

leading whites In a television picture, an abnormal condition in which the leading (left) edge of a black object has a white border.

leading-zero suppression In a digital meter, the blanking out of all zeros to the left of the decimal point.

lead-in spiral See LEAD-IN GROOVE.

lead-in tube A tube of insulating material, such as plastic or ceramic, used to conduct an antenna lead-in through a wall.

lead-in wire **1.** A single wire, used as a feed line for a shortwave receiving antenna. **2.** The feed line for a television receiving antenna. **3.** A feed line for a transmitting antenna.

lead network A phase-shift circuit containing series capacitance and shunt resistance; it produces a leading phase shift. Compare LAG NETWORK.

lead-out groove Around the inner edge of a phonograph record, a blank spiral groove leading into the eccentric or locked groove.

lead-over groove On a phonograph record containing several recorded tracks, a blank groove that conducts the stylus from the end of one recording to the beginning of the next.

lead screw **1.** A threaded rod that guides the cutter across the surface of a disc during its recording. **2.** In facsimile transmission, a threaded shaft that moves the drum or scanning mechanism lengthwise.

lead sulfate Formula, $PbSO_4$. An insulating compound formed in a LEAD-ACID CELL by the chemical action between the lead in the plates and the sulfuric-acid electrolyte. If the sulfate is not broken down during charging of the cell, it will eventually ruin the cell.

lead sulfide Formula, PbS. A compound of lead and sulfur used as the light-sensitive material in some photoconductive cells.

lead zirconate titanate A synthetic piezoelectric material.

leaf electroscope An electroscope using a pair of gold leaves or a single gold leaf and a solid strip of metal.

leak **1.** The loss of energy through a stray path not intended for conduction. **2.** A point from which energy is lost through a stray path not intended for conduction.

leakage **1.** The small current that flows through an electrical insulator. **2.** The electromagnetic field that is radiated or received by a feed line that should theoretically have 100-percent shielding.

leakage current The zero-signal current flowing across a reverse-biased semiconductor junction.

leakage flux Collectively, magnetic lines of flux around a transformer that do not link the primary and secondary coils.

leakage inductance Self-inductance caused by LEAKAGE FLUX. Leakage inductance is effectively in series with the primary or secondary winding of a transformer.

leakage resistance **1.** In an imperfect insulator, the ohmic resistance, calculated by dividing the voltage across the insulator by the current flowing through the insulator. **2.** The quotient of voltage and current in a reverse-biased semiconductor junction.

leakage radiation Radiation from parts of a system, as compared with that from the true radiator.

leakage reactance Inductive reactance caused by leakage inductance in the primary or secondary circuit of a transformer.

leakance The conductance of an insulating material, measured in siemens. It is equal to the reciprocal of the leakage resistance in ohms.

leaky **1.** Descriptive of a capacitor in which the dielectric material is not a perfect insulator. **2.** Descriptive of imperfect shielding in a coaxial transmission line. **3.** Descriptive of a waveguide with imperfect shielding.

leaky dielectric See LEAKY INSULATOR.

leaky insulator An insulator that conducts significant current at a specified (test) voltage.

leaky waveguide A waveguide that has imperfect shielding, allowing some electromagnetic field to escape.

leapfrogging In radar, a phasing process that eliminates false echoes resulting from the signals of other radar sets.

leapfrog test A test performed on different locations by a computer program in memory; it moves to another memory area to continue tests on other locations.

learning The ability of an artificially intelligent machine to improve or expand its knowledge with time. This can occur as a result of accumulation of information; it can also occur in systems that track their errors to avoid repeating them.

leased line A communications circuit reserved exclusively for a specific user.

least-significant bit Abbreviation, LSB. The digit with the lowest place value in a binary number.

least-significant character Abbreviation, LSC. In positional notation, the extreme right-hand character in a group of significant characters. See POSITIONAL NOTATION.

least-significant digit Abbreviation, LSD. The digit in a number that is at the extreme right (i.e., the one having the lowest place value).

least upper bound The smallest value of a parameter that can be obtained without changing some characteristic of a circuit, program or system.

Lecher frame A sturdy assemblage of LECHER WIRES.

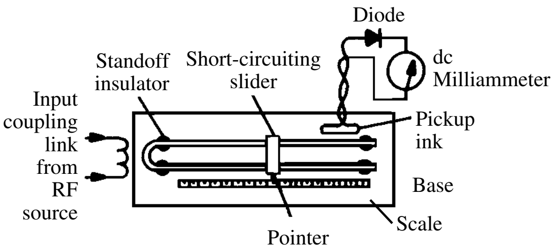

Lecher frame

Lecher lines See LECHER WIRES.

Lecher oscillator A self-excited radio-frequency oscillator using Lecher wires in place of an inductance-capacitance (LC) tank circuit. Also see LINE-TYPE OSCILLATOR.

Lecher wires A circuit segment consisting of two parallel wires or rods joined by a coupling loop on one end, the other end being open. A short-circuiting bar is moved along the wires to vary the effective length of the circuit. Radio-frequency energy is inductively coupled into the system through the loop, and the bar is slid along to various response points, as indicated by a meter or lamp coupled to the wires. The frequency can be determined by measuring the distance between adjacent response points. It is used at very-high frequencies (VHF), ultra-high frequencies (UHF), and microwave frequencies. Also called *Lecher frame* and *Lecher lines.*

Leclanche cell See DRY CELL, **1**.

LED See LIGHT-EMITTING DIODE.

LED phototransistor isolator An optoelectronic isolator in which the light source is a light-emitting diode and the light-sensitive component is a phototransistor.

LEDE Abbreviation for LIVE END/DEAD END.

LEF See LIGHT-EMITTING FILM.

left-hand lay See DIRECTION OF LAY.

left-hand motor rule See FLEMING'S LEFT-HAND RULE.

left-hand polarized wave See COUNTERCLOCKWISE-POLARIZED WAVE.

left-hand taper Potentiometer or rheostat taper in which most of the resistance is in the counterclockwise half of rotation as viewed from the front. Compare RIGHT-HAND TAPER.

left justified An item of data that occupies consecutive locations in storage, starting at the left-hand end of its area; empty locations might appear consecutively at the right-hand end if the item needs fewer positions than have been provided.

left shift A shift operation in which the digits of a word are displaced to the left; the effect is multiplication in an arithmetic shift.

leg **1.** Any one of the distinct branches of a circuit or network; also called ARM or BRANCH. **2.** In a computer program, a path in a routine or subroutine.

L electron In certain atoms, an electron whose orbit is outside of, and nearest to, the orbit of the K electrons.

Lenard rays See CATHODE RAYS.

length **1.** The number of bits or characters in a record, word, or other data unit. **2.** The end-to-end dimension of a device, circuit, line, etc. **3.** The start-to-finish duration of a time interval.

length to fault In cable or line measurements from the home station, the distance (i.e., the cable or line length) to the point at which a fault, such as a short circuit or ground, is located.

lens **1.** A usually circular piece of transparent material with one or both surfaces curved in cross section, used (through its refractive properties) to focus or spread rays that pass through it. Lenses for visible light must be transparent, but those for other radiation, such as radio waves, need not transmit light. Also see ANTENNA LENS. **2.** In a cathode-ray tube, one or more high-voltage electrodes for focusing an electron beam to a fine point on the screen.

lens antenna See ANTENNA LENS.

lens disk A NIPKOW DISK having a lens in each hole.

lens speed The light-transmitting ability of a lens, given as an f-stop number: the lens' focal length divided by its diameter.

lens turret A multiple-lens mount on a camera that can be rotated for quick lens interchange.

LEO Acronym for LOW EARTH ORBIT.

LEO satellite system A set of several dozen low earth orbit (LEO) satellites in polar orbits spaced strategically around the globe such that, for any point on the earth, there is always at least one satellite in range. The satellites can relay messages throughout the fleet. Any two points on the surface are always able to make contact through the satellites.

Lepel discharger See QUENCHED SPARK GAP.

letter-identification word See PHONETIC ALPHABET CODE WORD.

letters patent See PATENT.

letters shift In a text communications terminal, a control character that causes all of the following characters to occur in the lower case. This can be an automatic or a manual control character.

let-through current The current conducted by a circuit breaker during a short-circuit.

level-0 security In communications, the complete lack of security measures. Anyone can listen in on a conversation at any time, provided they are willing to spend the money and/or time to obtain the necessary equipment. Examples are amateur ("ham") radio and Citizens' Band (CB) voice communications.

level-1 security Also called *wire-equivalent security.* In communications, the implementation of security measures such that the circuit is protected to the extent of a typical hard-wired link. The encryption is anticipated to be unbreakable for at least 12 months, and preferably for 24 months or more. The technology is updated at least every 12 months, and preferably every 6 months.

level-2 security Also called *commercial-level security.* In communications, the implementation of security measures such that the circuit is deemed safe for ordinary commercial transactions. The encryption is anticipated to be of such a nature that engineers believe it would take a hacker at least 10 years, and preferably 20 years or more, to break the cipher. The technology should be updated at least every 10 years, but preferably every 3 to 5 years, and more often if possible.

level-3 security Also called *mil-spec security.* In communications, the implementation of security involving the most sophisticated forms of encryption and personnel restriction that a government can muster.

level **1.** The amplitude at which a device is functioning or at which a phenomenon occurs (e.g., collector-current level, or received signal level). **2.** The minimum amplitude at which a phenomenon occurs; also called *threshold amplitude.* **3.** A functional plateau or echelon.

level clipper See CLIPPER.

level compensator **1.** An automatic gain control (AGC) that effectively reduces amplitude variations in a received signal. **2.** An automatic gain control in telegraph receiving equipment.

level control **1.** The adjustment of amplitude or threshold. **2.** A potentiometer or other variable component for adjusting the amplitude or threshold of a quantity.

level indicator See VOLUME INDICATOR.

level translator Any circuit or device that alters the voltage levels of input signals. An example is a converter that changes positive-logic signals to negative-logic signals.

level-triggered flip-flop A flip-flop that responds to voltage level, rather than to the frequency of an input signal.

lever switch **1.** A switch designed for rapid making and breaking of a circuit. **2.** A radiotelegraph key.

Lewis antenna A form of antenna used at ultrahigh and microwave frequencies. It resembles a horn antenna.

Leyden bottle See LEYDEN JAR.

Leyden jar [Leyden, Holland (also Leiden), site of the invention in 1745 by Peiter van Musschenbroek, 1692–1761.] The first practical capacitor. In modern form, it is a glass jar covered inside and out with metal foil and has a rod topped by a metal ball that touches the inner foil. It is still used occasionally in classrooms for demonstrating static electricity. The Leyden jar was co-invented by van Musschenbroek and invented independently by E. G. von Kleist of Pomerania, among others.

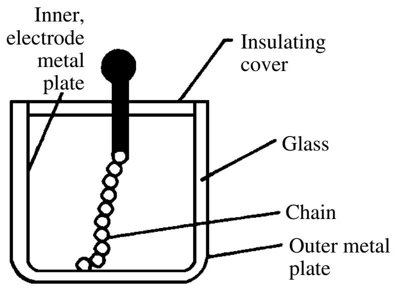

Leyden jar

Leyden phial See LEYDEN JAR.

Leyden vial See LEYDEN JAR.

LF Abbreviation of LOW FREQUENCY.

L_F band A section of the L BAND extending from 1.35 to 1.45 GHz.

L fitting See ELL.

LHD Abbreviation of LOAD/HAUL/DUMP.

Li Symbol for LITHIUM.

librarian program A computer program controlling a LIBRARY.

library In digital-computer and data-processing operations, the permanent storage of data or instructions. Also called *permanent mass storage.*

libration fading In earth-moon-earth (EME) communications, also known as *moonbounce,* rapid and deep fading, accompanied by phase modulation, that takes place because the moon does not always show the earth exactly the same portion of its surface. The moon "wobbles" slightly back and forth on its axis. This causes changes in the relative phases of signals reflected from the topographic features (mountains, in particular) on the lunar surface.

licrystal An acronym from *liquid* and *crystal.* See LIQUID CRYSTAL.

lidar See LIGHT DETECTION AND RANGING.

lie detector See POLYGRAPH.

life **1.** The duration of useful service (or of operation before failure) of electronic equipment. **2.** For a non-rechargeable cell or battery, the length of time it will last in a given application before it must be discarded and replaced. **3.** In robotics and artificial intelligence (AI), a general term that refers to qualitative similarities between machines and animate creatures, including human beings.

life test An assessment of the life of electronic equipment—either by means of full-time test runs or accelerated time tests.

lifetime See CARRIER LIFETIME.

lifter A device in a magnetic tape recorder that removes the tape from the recording and playback heads under fast-forward and rewind conditions.

light Visible electromagnetic radiation occurring in the wavelength band of about 750 nanometers (red light) to 390 nanometers (violet light). Included sometimes in the category of light are infrared and ultraviolet rays.

light-activated silicon-controlled rectifier Abbreviation, LASCR. A silicon-controlled rectifier that functions both as a photosensor and a heavy-duty bistable electronic switch, allowing high currents to be switched by means of a light beam.

light-activated silicon-controlled switch Abbreviation, LASCS. A pnpn device that acts simultaneously as photocell and electronic switch.

light adaption **1.** The process whereby the eye adjusts itself to an increase or decrease in illumination. **2.** Similar action in photoelectric devices. **3.** The length of time required for the eye to adjust itself to an increase or decrease in illumination.

light amplifier A solid-state amplifier using an input electroluminescent cell and an output photocell, or some similar pair of components. The device is essentially an optoelectronic coupler with gain.

light-beam communication A system of communication in which a beam of light between transmitting and receiving stations is modulated or interrupted to convey intelligence. A laser is commonly used because it has minimal beam divergence, allowing maximum communication range.

light-beam meter An electric meter using a LIGHT-BEAM POINTER.

light-beam pointer A slender beam of light that replaces the pointer in a moving-coil meter. The light comes from a small incandescent lamp and is reflected by a mirror attached to the coil; when the coil moves, a spot of light moves over the scale of the meter.

light-beam receiver The receiver in a LIGHT-BEAM COMMUNICATION system.

light-beam recorder A graphic recorder using a light-beam pointer. In this device, a small spot of light traces a pattern on moving photographic film, which is subsequently developed to produce a permanent record.

light-beam transmitter The transmitter in a LIGHT-BEAM COMMUNICATION system.

light cable A cable, consisting of numerous thin optical fibers, through which light can be transmitted for communication or control purposes. See, for example, LIGHT-WAVE TELEPHONY.

light chopper A device that modulates a light beam by interrupting it repetitively.

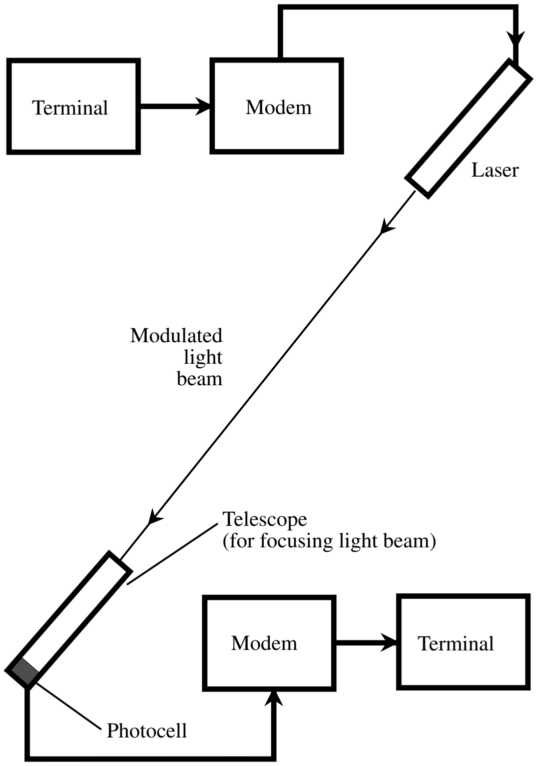

light-beam communication

light detection and ranging Acronym, lidar. A navigation and surveillance system in which laser light scans in a manner similar to that of RADAR.

light dimmer See DIMMER.

lighted pushbutton See LIGHTED SWITCH.

lighted switch A pushbutton switch containing a pilot light that glows to show when the switch is on. Also called *illuminated switch.*

light-emitting diode Abbreviation, LED. A semiconductor device that emits visible light when forward biased. Also see LASER DIODE.

light-emitting film A thin phosphor film that becomes luminescent when a high-frequency voltage is applied across its surface. Also see ELECTROLUMINESCENCE and ELECTROLUMINESCENT CELL.

light flasher An electronic circuit or simple automatic flasher switch for flashing a lamp at regular intervals.

light flicker See LOAD FLICKER.

light flux See LUMINOUS FLUX.

light hood See HOOD.

light-induced electricity See PHOTOELECTRICITY.

light load A load that is a fraction of the usual value for a given application. That is, its resistance or impedance is several times higher than normal.

light meter An electronic instrument for measuring the intensity of light. It generally consists of a photodiode, a battery, and a direct-current microammeter connected in series. A direct-current amplifier can be used to increase the sensitivity.

light microsecond A unit of electrical distance; the distance that light, or any electromagnetic disturbance, travels in free space in 1 microsecond. Approximately equal to 300 meters.

light modulation Variation of the instantaneous brightness of a visible light beam in synchronization with the instantaneous amplitude of a modulating signal. Also see LIGHT MODULATOR.

light modulator A device with which a beam of light can be modulated by an electrical signal.

light negative Pertaining to negative photoconductivity, the decrease in conductivity of a photosensitive material under illumination. Compare LIGHT POSITIVE.

lightning The discharge that occurs between positive and negative poles in the atmosphere. Common in and near areas where heavy rainfall is occurring. It also can occur in snow storms, in sand storms, and over erupting volcanoes. Generally, the negative pole is in a cloud and the positive pole is at the surface of the earth, resulting in a flow of electrons from cloud to ground. Some lightning occurs as a flow of electrons from ground to cloud, or between two clouds. Such discharges sometimes attain current levels of more than 1,000,000 amperes.

lightning arrester A device that bypasses high-voltage pulses from most nearby lightning discharges to the earth, helping to protect electronic equipment connected to an outdoor antenna or power line. Is not a total guarantee of protection, however.

lightning conductor **1.** A system for protecting buildings from lightning strikes. A common system includes a lightning rod, heavy conductor, and ground rod. The ground rod is placed at least six feet from the base of a building and is at least eight feet long. **2.** The conductor between the lightning rod and ground rod in a system, as defined in **1**.

lightning detector See KERAUNOGRAPH and KERAUNOPHONE.

lightning rod A protective device mounted on the outside of structures, consisting of a pointed, grounded metal rod that will conduct a lightning discharge to earth.

lightning strike A direct hit of an object by a lightning stroke. It usually causes extensive damage to electrical and electronic equipment through which the discharge passes.

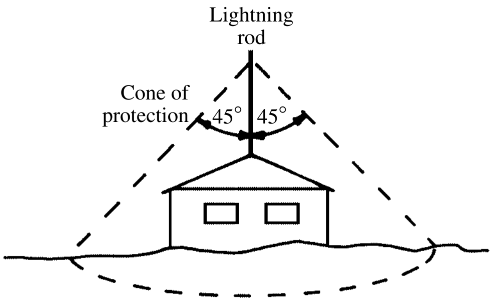

lightning rod

lightning stroke The discharge that occurs with lightning. The peak current is typically several tens of thousands of amperes, but in some cases can exceed 1,000,000 amperes. A stroke can consist of one discharge or several individual discharges in rapid succession.

lightning switch See GROUND SWITCH.

light-operated relay See PHOTOELECTRIC RELAY.

light-operated switch A PHOTOELECTRIC RELAY, or a switch operated by such a relay.

light pen A probe containing a tiny photosensor in its tip. The tip of the light pen is touched to the screen of a cathode-ray tube to sense the beam when it passes the spot of contact. It is used as an input device in some computers and terminals.

light pipe **1.** An OPTICAL FIBER. **2.** A cable consisting of numerous optical fibers in a bundle. See FIBEROPTICS, **1**.

light positive Pertaining to positive photoconductivity, when the conductivity of a photosensitive material increases under illumination. Compare LIGHT NEGATIVE.

light quantum See PHOTON.

light ray A thin beam of light. Theoretically, a ray emerges from a point source (i.e., it has no width).

light receiver See LIGHT-BEAM RECEIVER.

light relay A photoelectric device that operates a relay, according to fluctuations in the intensity of a light beam.

light-sensitive cathode Also called *photocathode.* A cathode that emits electrons when exposed to light.

light-sensitive diode A semiconductor diode usable as a photoconductive cell. Such diodes are available as both junction and point-contact types.

light-sensitive material A photoconductive or photoemissive substance.

light-sensitive resistor See PHOTOCONDUCTIVE CELL.

light sensor **1.** A light-sensitive device, such as a photocell, photodiode, phototransistor, or phototube. **2.** A light-sensitive substance, such as cesium, selenium, silicon, cadmium selenide, or lead sulfide.

light source Any generator of light. Under some conditions, the source is regarded as a point.

light spectrum See ELECTROMAGNETIC THEORY OF LIGHT.

light-spot scanner Also called *flying-spot scanner.* A television camera using (as a source of illumination) a spot of light that scans what is to be televised.

light transmitter See LIGHT-BEAM TRANSMITTER.

light valve **1.** An electromechanical device for varying the intensity of light passing through its adjustable aperture. **2.** See KERR CELL.

light-wave telephony Telephone communication by means of modulated-light transmission, usually through an OPTICAL FIBER.

light-year Abbreviation, lt-yr. Pertaining to astronomy, a unit of distance equal to the distance traveled by light in one year in a vacuum: $9.460\ 55 \times 10^{15}$ meters (5.878×10^{12} miles).

likelihood In probability and statistics, the chance that an event will occur or that an outcome will be realized. Also see PROBABILITY, **1**, **2**.

lim Abbreviation of LIMIT.

Lima Pronunciation, *LEE-ma.* Phonetic alphabet word for the letter L.

limen A unit that has been proposed as the minimum audible change in frequency that can be detected by at least half of a group of listeners.

limit **1.** The lowest or highest frequency in a band. **2.** In mathematics, a fixed value that a variable approaches. **3.** The upper and lower extremes in any performance range or value range.

limit bridge A bridge used to check a component (e.g., resistance, capacitance, or inductance) in terms of the tolerance limits, rather than the nominal (named) value, of that component. Also see BRIDGE, **2**.

limited integrator A circuit that integrates two input signals until the corresponding output signal exceeds a certain limit.

limited stability A characteristic of a circuit or system, allowing proper operation only if the input signal and applied voltages are within certain maximum and minimum limits.

limiter A device or circuit whose output-signal amplitude remains at some predetermined level, despite wide variations in input-signal amplitude.

limiting The restriction of the maximum peak amplitude of a signal to a designated level.

limiting amplifier An amplifier that automatically holds the output-signal level to a prescribed value.

limiting current In electrolysis, the highest current that conducts under certain conditions of ion concentration. This current depends on the electrolyte material, the concentration of the electrolyte in solution, the electrode substance, and the size of the electrolytic cell.

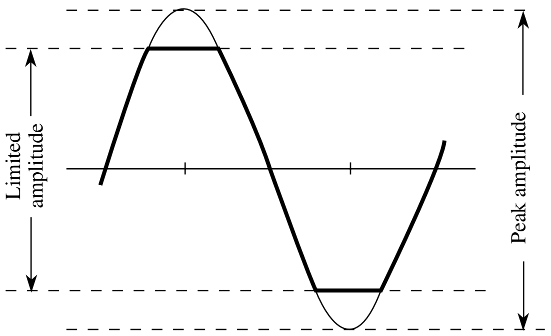

limiting

limiting error The anticipated maximum value of the absolute error in a computation.

limiting resistor See CURRENT-LIMITING RESISTOR.

limiting resolution As a measure of video image resolution, the maximum number of lines for picture height that can be discriminated on a test chart.

limit switch A switch that is actuated when a monitored quantity (e.g., current, voltage, or illumination) reaches the limit of its range.

line **1.** A wire, cable, or waveguide, along which electrical or electromagnetic energy travels from one defined place to another. **2.** One lengthwise path in which a force, such as electricity or magnetism, is evidenced. Such a *line of flux* has theoretically zero width.

line advance **1.** The physical separation between the centers of adjacent scanning lines in a television system. **2.** Line feed in a text data transmission system.

line amplifier An amplifier in a telephone line or similar channel, or one feeding such a line from the input end.

linear **1.** In a straight line. **2.** In the manner of a straight line. Thus, linear response is indicated when one quantity varies directly with another; the graph of this response is a straight line (i.e., one of constant slope). **3.** The characteristic of a signal that is a replica of another (e.g., an amplifier output signal of the same waveform as that of the input signal).

linear absorption coefficient A number expressing the extent to which the intensity of an X-ray beam is reduced per centimeter of the material through which it passes.

linear accelerator A device in which subatomic particles are accelerated in a straight line through a long tube. This action is in contrast with that occurring in a circular accelerator, such as a CYCLOTRON.

linear algebra A branch of mathematics that deals with the solving of linear equations or sets of linear equations.

linear amplifier **1.** An amplifier for which a linear relationship exists between input and output parameters (e.g., a high-fidelity audio amplifier). **2.** A class-AB radio-frequency power amplifier that does not distort the envelope of an amplitude-modulated (AM) or single-sideband (SSB) signal. It is commonly used by amateur radio operators.

linear array A directional antenna having equally spaced, in-line elements.

linear circuit **1.** A circuit whose output is a faithful reproduction of the input. See LINEAR AMPLIFIER, **1** and LINEAR DETECTOR. **2.** A circuit whose performance is linear. See LINEAR RESPONSE, **1**.

linear decrement In a damped wave, a constant decrease in amplitude with time, as opposed to a decrease that is a logarithmic or otherwise nonlinear function of time. Compare DECREMENT.

linear detector A detector whose output/input relationship is linear. Also see LINEAR, **2**, **3**; LINEAR CIRCUIT, **1**; and LINEAR RESPONSE, **1**, **2**, **3**.

linear differential transformer A device that converts the physical position of an object into an output voltage or current. The voltage or current is directly proportional to the displacement.

linear distortion Amplitude distortion in which the output and input signal envelopes are disproportionate (in the absence of spurious frequencies).

linear equation See FIRST-DEGREE EQUATION.

linear function **1.** In Cartesian two-space, a function of the form $y = mx + b$, where the coordinates are (x,y), the slope is m (any real number) and b is the point on the y axis at which the graph crosses the axis. **2.** Any function in any number of dimensions whose graph appears as a straight line in the Cartesian system of coordinates.

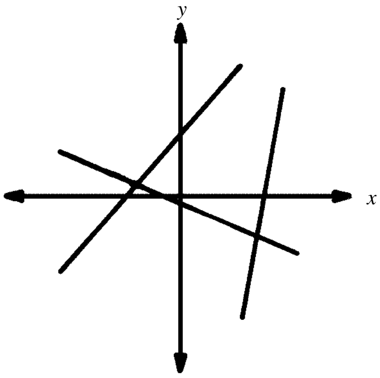

linear function

linear IC See LINEAR INTEGRATED CIRCUIT.

linear integrated circuit An integrated circuit designed for analog operations (such as signal amplification, oscillation, nondigital regulation, analog instrumentation, and similar applications). Compare DIGITAL INTEGRATED CIRCUIT.

linearity **1.** The degree to which performance or response approaches the condition of being linear, expressed in percent or parts per million. Also see LINEAR AMPLIFIER, **1**; LINEAR CIRCUIT, **1**, **2**; LINEAR OSCILLATOR, **1**; LINEAR RESPONSE, **1**, **2**, **3**; and LINEAR TAPER. **2.** In a cathode-ray-tube image, absence of compression or stretching of any portion of the image; that is, an undistorted reproduction.

linearity control In a cathode-ray-tube display, the potentiometer used to correct image linearity. See LINEARITY, **2**.

linearity error **1.** The difference between a theoretically linear function and the actual function, as observed under experimental conditions. **2.** The degree of nonlinearity in an amplifier that is supposed to be linear.

linear modulation **1.** Modulation in which the instantaneous amplitude of the input signal is directly proportional to the instantaneous amplitude of the output signal. **2.** Modulation in which the instantaneous amplitude of the input signal is inversely proportional to the instantaneous amplitude of the output signal. **3.** Modulation in which the instantaneous amplitude of the input signal is directly proportional to the frequency or phase deviation of the output signal.

linear motor A motor in which the stator and rotor are parallel and straight.

linear oscillator **1.** An oscillator whose alternating-current output amplitude varies linearly with its direct-current input. **2.** A line-type oscillator.

linear programming A method of determining the optimum value for a certain set of linear equations. Generally, this is done by finding the point on a plane in space closest to some point not on the plane.

linear quantizing A method of quantizing in which all of the intervals are of equal size or duration. An example is a linear analog-to-digital converter circuit.

linear reflex detector See INFINITE-IMPEDANCE DETECTOR.

linear response **1.** A response in which the value of the dependent variable is equal or directly proportional to the value of the independent variable. Thus, the graph of the response function is a straight line over the range of normal operating values. Compare LOGARITHMIC RESPONSE, **1** and SQUARE-LAW RESPONSE. **2.** A type of response in which a quantity (such as current) varies directly with another quantity (such as voltage). Compare LOGARITHMIC RESPONSE, **2**. **3.** Low-distortion response. Also see HIGH FIDELITY.

linear scale A scale in which all of the divisions represent the same differential and are equally spaced. On a linear scale, a given difference

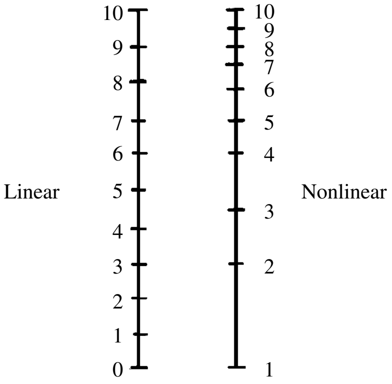

linear scale

always has the same physical length, no matter where on the scale it appears. For example, the interval between 3 and 4 on a linear scale is the same as any interval between x and $x + 1$, where x is any real value on the scale.

linear sweep In a television or oscilloscope circuit, the scanning of the electron beam across the screen at a constant speed. Also see LINEAR, **1**; LINEAR RESPONSE, **1**; and LINEARITY, **2**.

linear taper In a potentiometer or rheostat, resistance variation that is directly proportional to shaft rotation. Thus, half the total resistance corresponds to movement of the shaft over half the arc of full rotation. Compare LOG TAPER. Also see TAPER.

linear time base For an oscilloscope, the base provided by sweeping the electron beam horizontally at a uniform rate. Also see LINEAR SWEEP.

linear track On a video tape, the track that contains audio information that accompanies the video data. It was so named because it is a single, straight track, in contrast to the video tracks, which are angled.

linear tracking In a turntable system, a design scheme in which the lateral movement of the stylus, as the disc is played, occurs in a straight line, rather than in an arc.

linear transformer A radio-frequency transformer consisting of a section of transmission line.

linear variable differential transformer Abbreviation, LVDT. A differential transformer exhibiting linear response. Also see LINEAR RESPONSE, **1**, **2**.

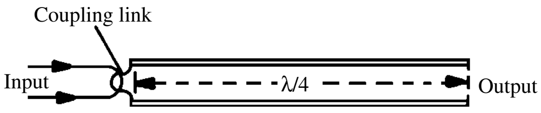

linear transformer

line balance The degree of electrical similarity between transmission line conductors, or between a conductor and ground.

line-balance converter A device used to isolate the outer conductor at the end of a coaxial line from ground.

line characteristic distortion Fluctuations in the duration of received signal impulses in text data communications, caused by changing current transitions in the wire circuit.

line circuit The telephone system relay equipment associated with stations connected to a switchboard.

line code A code between the digits in processing equipment and the pulses representing the digits in a line transmission.

line conditioning In data communications, the modification of private or leased lines by adding compensating reactances to reduce amplitude variations or phase delays over a band of frequencies.

line contact stylus A needle used to reproduce stereo high-fidelity sound from vinyl discs. It has a characteristic oblate ellipsoidal shape.

line coordinate A symbol identifying a specific row of cells in a matrix; a specific cell can be located with an additional column coordinate.

line cord A flexible two- or three-wire insulated cable connecting equipment to the power line by means of a plug that mates to a standard electrical outlet.

line current **1.** Current flowing from a power line into equipment. **2.** Current flowing in a transmission line. **3.** Current flowing into a parallel-resonant circuit.

line diffuser A circuit that creates minor vertical oscillations of the spot on a television screen, making the individual scanning lines less noticeable.

line driver An integrated circuit capable of transmitting logic signals through long lines.

line drop The voltage drop along a line supplying power to a device.

line equalizer See EQUALIZER.

line fault A discontinuity in a transmission line, resulting in signal loss at the receiving end of a circuit.

line feed In a text data transmission system, the movement of the paper, platen, or cursor to allow for printing or displaying an additional line of text.

line filter **1.** A circuit that can be inserted in the utility power cord of an appliance, device, or system for transient suppression and the minimization of electromagnetic interference (EMI). The filter employs series inductors, parallel capacitors, and in some cases, a clamping device. Installed between a radio transmitter and the utility lines, such a filter can choke off radio-frequency (RF) current and help keep utility wiring from acting as an antenna. Installed in the

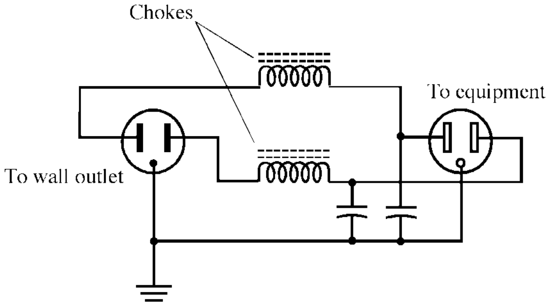

line filter

power cords of home entertainment equipment, such as stereo audio amplifiers, line filters can keep RF from entering the apparatus through the power supply. **2.** A circuit containing parallel capacitors and/or series inductors, installed in a twisted-pair telephone line to reduce or eliminate the effects of EMI and transients in a telephone set.

linefinder A switching device that finds one of a group of calling telephone lines and connects it to a trunk, connector, or selector.

line frequency **1.** The frequency of power-line voltage in the United States, 60 Hz. **2.** The rate at which the horizontal lines are traced in a video image. Also see HORIZONTAL FREQUENCY.

line group **1.** A group of signals sent by wire transmission. **2.** The frequency spectrum occupied by a group of signals sent by wireless transmission.

line leakage Resistance between insulators of two wires in a telephone line loop.

line loss The sum of energy losses in a transmission line.

lineman A technician who works mainly with telephone or telegraph lines.

line matching transformer A device that provides an impedance match in an audio circuit, and also can adapt balanced to unbalanced audio transmission lines (or vice versa).

line noise **1.** Electrical noise (as received by a radio) arising from fluctuations of current or voltage in a power line. **2.** Noise in a data transmission line.

line of flux **1.** A line (usually curved) depicting the points of equal-intensity field strength in the vicinity of an electric charge or a charged body. Also see FLUX. **2.** A line depicting the points of equal-intensity field strength in the vicinity of a magnetic pole or a magnetized body. Also see FLUX.

line-of-sight communication Radio communications between points located so that a straight line between them does not pass through the earth, or through any major obstructions. Also see LINE-OF-SIGHT DISTANCE.

line-of-sight distance The maximum distance over which an ultra-high-frequency (UHF) or microwave signal can be directly transmitted along the surface of the earth. It is slightly more than the maximum optical line-of-sight distance.

line oscillator See LINE-TYPE OSCILLATOR.

line plug The plug terminating a line cord. Also see MALE PLUG.

line printer A machine that prints the results of a computer run, line by line.

line radio See WIRED RADIO.

line regulation Automatic stabilization of power-line voltage.

line-sequential system The color television system in which the image is reproduced by means of primary color lines (red, green, and blue) sequentially beamed across the screen of the picture tube. Compare DOT-SEQUENTIAL SYSTEM and FIELD-SEQUENTIAL SYSTEM.

lines of cleavage See CLEAVAGE.

lines oscillator See LINE-TYPE OSCILLATOR.

line supervision In electronic security systems, a method of monitoring circuit characteristics to detect possible tampering.

line switch **1.** The main power-line switch to a system. **2.** Within a piece of electronic equipment, the switch that opens and closes the circuit to the incoming power line.

line-type amplifier A radio-frequency amplifier in which the tuned circuits are transmission lines consisting of parallel wires, rods, or tubing, or of coaxial cable sections.

line-type oscillator A radio-frequency oscillator in which the tuned circuits are transmission lines consisting of parallel wires, rods, or tubing, or of coaxial cable sections.

line unit In a wire data transmission system, the terminal unit, or device that converts the text signals into electrical impulses and vice versa.

line voltage **1.** The voltage between the conductors in a power line. **2.** The voltage between the conductors of a transmission line.

line-voltage drop See LINE DROP.

line-voltage monitor See POWER-LINE MONITOR.

linguistics The study of languages, including structure, symbology, and phonetics.

link **1.** The small coupling coil used in link coupling. **2.** A communication path between two radio facilities for the purpose of extending the range of one, as between a remote pickup point and a broadcast transmitter. **3.** A data connection between two different computers. **4.** The act or process of creating a signal path or data connection, as defined in **1**, **2**, or **3**. **5.** In a digital computer, a branch instruction, or an address in such an instruction, used to leave a subroutine to return to some point in the main program.

linkage Coupling between separated conductors or devices through the medium of electric or magnetic lines of flux.

link circuit A closed-loop coupling circuit having two coils of a few turns of wire; each coil is placed near one of the circuits to be coupled.

link coupling Low-impedance coupling via a small (usually one-turn) input or output coil fed by a twisted pair or a coaxial line.

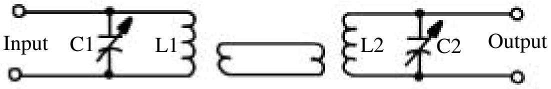

link coupling

linked subroutine A subroutine, entered by a branch instruction from a main routine, that executes a branch instruction returning control to the main routine.

link fuse A fuse consisting of an exposed length of fuse wire.

link neutralization Neutralization achieved by out-of-phase current fed back via link coupling from the output to the input of an amplifier. Also called INDUCTIVE NEUTRALIZATION.

lin-log receiver A radar receiver whose amplitude response is linear for small signals, but logarithmic for large ones.

lip microphone A small microphone operated close to or in contact with the lips.

liquid A state of matter characterized by a level of molecular motion intermediate between that of gases and solids; liquids have the ability (like gases) to take the shape of a container and are only slightly compressible. Compare GAS, PLASMA, and SOLID. Also see STATE OF MATTER.

liquid absorption For a solid material, such as dielectric, the ratio of the weight of liquid absorbed by the material to the weight of the material.

liquid capacitor See WATER CAPACITOR.

liquid cell See ELECTROLYTIC CELL.

liquid conductor See ELECTROLYTE.

liquid cooling Use of circulating water, oil, or other fluid to remove heat from components or equipment, such as microprocessors or power amplifiers.

liquid crystal A liquid exhibiting some of the characteristics of a crystal. Also see NEMATIC CRYSTAL and SMECTIC CRYSTAL.

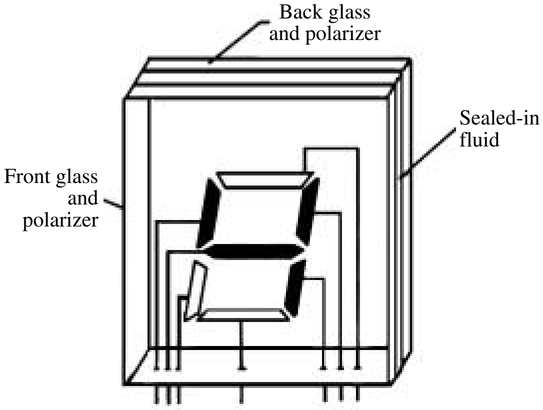

liquid-crystal display

liquid-crystal display Abbreviation, LCD. A flat-panel display noted for its thin profile, light weight, and low power consumption. The simplest devices are used in calculators, meters, wristwatches, and radios. More sophisticated displays are used in computers and portable video units. This type of display can operate at a much lower voltage than a cathode-ray tube (CRT). This makes it ideal for portable electronic systems in which batteries are used. Older displays of this type can be difficult to read from certain viewing angles. This situation has improved in recent years with the advent of *active-matrix*, also known as *thin-film-transistor* (TFT), displays.

liquid-filled transformer A transformer filled with a protective liquid insulator, such as oil.

liquid-flow alarm An electronic circuit that actuates an alarm when the flow of a liquid through pipes or other channels changes from a desired rate.

liquid-flow control A servo system that automatically maintains or corrects the rate of liquid flow through pipes or other channels.

liquid-flow gauge See LIQUID-FLOW METER.

liquid-flow indicator See LIQUID-FLOW METER.

liquid-flow meter An instrument that indicates the rate at which a liquid moves through pipes or other channels.

liquid-flow switch In a liquid-cooled system, a switch that actuates an alarm when the liquid slows or stops.

liquid-jet oscillograph A graphic recorder using an ink-jet galvanometer to trace the pattern on a paper chart.

liquid laser A laser in which the active material is a liquid.

liquid-level alarm An electronic device that actuates visual or audio signal devices when the surface of a liquid inside a tank rises or falls to a predetermined level.

liquid-level control A servo system that automatically maintains the liquid in a tank at a predetermined level.

liquid-level gauge An electronic system that provides direct readings of the level of a liquid in a tank.

liquid-level indicator See LIQUID-LEVEL GAUGE.

liquid-level meter See LIQUID-LEVEL GAUGE.

liquid load See WATER LOAD.

liquid-pressure alarm An electronic circuit that actuates an alarm when the pressure of a liquid changes.

liquid-pressure control A servo system that automatically maintains or corrects liquid pressure in pipes or other channels.

liquid-pressure gauge See LIQUID-PRESSURE METER.

liquid-pressure indicator See LIQUID-PRESSURE METER.

liquid-pressure meter An instrument that provides direct readings of liquid pressure in a pipe or other channel.

liquid-pressure switch A switch that actuates an external circuit or device when the pressure of a liquid changes.

liquid rheostat See WATER RHEOSTAT.

liquid valve See ELECTROMECHANICAL VALVE.

liser An oscillator that produces an extremely pure microwave carrier signal.

LISP A digital-computer language written in the form of lists. A program can be directly interpreted as data, and vice versa. The entire language is derived from a few basic functions. Programs are easy to debug. It is used in artificial intelligence (AI) research. Also see LANGUAGE.

Lissajous figure Any one of several curves resulting from the combination of two harmonically related sine waves. These figures are familiar in electronics; they are obtained when signals are applied simultaneously to both axes of an oscilloscope. It is also called *Lissajous pattern.*

list **1.** To print serially the records in a file or in memory. **2.** To print (instruct a computer to display) every item of input data in a program. **3.** A one-dimensional array of numbers.

listener fatigue Physiological symptoms, such as headaches, caused by prolonged listening to certain sounds (e.g., a pure sine wave or poorly reproduced music).

listening angle In stereo sound reproduction, the angle between the speakers, with respect to the listener. This angle can vary from zero to 180 degrees. Larger angles result in increased apparent channel separation.

listening test The subjective evaluation of audio equipment by listeners.

liter Abbreviation, l. A metric unit of volume equal to 1.0567 U.S. liquid quarts or 0.908 U.S. dry quart. A liter is the volume of 1 kilogram of water at 4¯C and under a pressure of 1 pascal.

literal operands It is usually applicable to source language instructions, operands that specify the value of a constant, rather than an address of a location in which the constant is stored.

lithium Symbol, Li. An element of the alkali-metal group, and the lightest elemental metal. Atomic number, 3. Atomic weight, 6.941.

lithium battery See LITHIUM CELL.

lithium cell A type of electrochemical cell with exceptional energy-to-mass ratio and long shelf life. There are several variations in the chemical makeup; they all contain lithium, a light, highly reactive metal. These units can be manufactured to supply 1.5 to 3.5 volts, depending on the particular chemistry used. These cells can last for years in very-low-current applications such as memory backup.

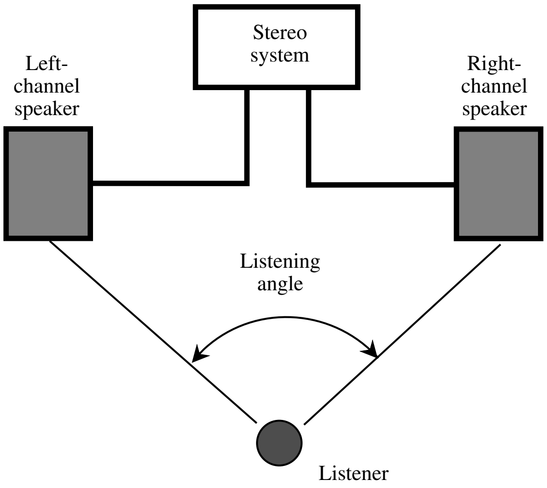

listening angle

Litzendraht wire See LITZ WIRE.

Litz wire A woven wire having a number of copper strands, each separately enameled to insulate it from the others. The wire is woven so that inner strands come to the surface at regular intervals. It is noted for its low losses at radio frequencies.

live **1.** Electrically activated (i.e., sustaining voltage or current). **2.** Being broadcast as it occurs. **3.** Acoustically reflective, as in a LIVE ROOM (contrasted with one that is acoustically absorbent).

live end **1.** In a recording or broadcasting studio, the part of the room in which the acoustic concentration is greatest. **2.** In a utility circuit, the wire or terminal that carries 117 volts alternating current (the ungrounded end).

live end/dead end Pertaining to a room that is acoustically reflective (live) at one end, and acoustically absorbent (dead) at the other end.

live room A room with little or no acoustically absorbent material in its ceiling, walls, and floor, with the result that echoes and reverberation are pronounced. Compare DEAD ROOM.

lix Abbreviation of LIQUID CRYSTAL.

L_K band A section of the L BAND that extends from 1.150 to 1.350 GHz.

L_L band A section of the L BAND that extends from 510 to 725 MHz.

LLL Abbreviation of LOW-LEVEL LOGIC.

lm Preferred abbreviation of LUMEN.

lm/ft^2 Abbreviation of *lumens per square foot.* Also see LUMEN.

lm-hr Abbreviation of LUMEN-HOUR.

lm/m^2 Abbreviation of *lumens per square meter.* Also see LUMEN and LUX.

lm/W Abbreviation of *lumens per watt*, a unit of luminosity. Also see LUMEN.

ln Representation of the natural (base-*e*) logarithm function (see NAPIERIAN LOGARITHM). Also written *loge*.

L network An impedance-matching circuit, filter, or attenuator whose schematic representation resembles an inverted letter L.

LO Abbreviation of LOCAL OSCILLATOR.

lo Abbreviation of *low*, usually as a prefix or subscript. Also abbreviated L.

load **1.** Also called *electrical load.* A device or circuit that is operated by the electrical power output of another device or circuit. Examples: speaker, light bulb, power amplifier, antenna system. **2.** Also called *mechanical load.* The mechanical power output demanded of a machine—especially a motor. **3.** To fill an internal computer storage with information from an external storage [e.g., from a magnetic disk to a computer's random access memory (RAM)]. **4.** To add inductors and/or capacitors to an antenna system to alter the characteristics of the system—especially the resonant frequency. **5.** To adjust the output circuit of a radio-frequency power amplifier for optimum energy transfer to the antenna system.

load-and-go Automatic coding in which a user's (source) program is translated automatically into machine language and stored.

load capacitance **1.** The capacitance present in an electrical load (see LOAD, **1**). **2.** A capacitance used as an electrical load.

load capacity **1.** In pulse-code modulation, the level at which a sine-wave signal has peaks coinciding with the plus/minus virtual decision values of the encoder. **2.** The maximum number of signals that a medium or line can carry without serious degradation of reception.

load circuit **1.** The circuit that forms the load, or power-consuming portion, of a system. **2.** A circuit that facilitates transfer of power to a load.

load coil See WORK COIL.

load current The current flowing in a load. See LOAD.

load division A method of connecting two or more power sources to a single load, for optimum power transfer.

loaded antenna An antenna incorporating a LOADING COIL and/or LOADING DISK to increase its electrical (effective) length. See LOADING COIL and LOADING DISK.

loaded line A transmission line in which inductors or capacitors are inserted at appropriate points to alter the characteristics of the line.

load end The end of a transmission line to which a radiator or receiver is connected.

load flicker Fluctuations in the brightness of lamps, caused by intermittent loading of the power line by other devices.

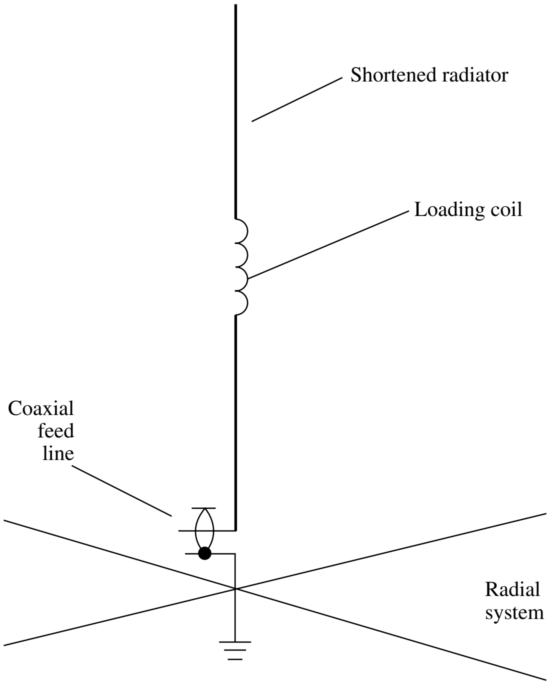

loaded antenna

load/haul/dump Abbreviation, LHD. A remote-controlled or computer-controlled robot used in mining and construction work. Under the direction of a human operator or a computer, it loads, hauls, and dumps materials (hence its name). It can use various navigation methods, including beacons, computer maps, position sensors, and vision systems.

load impedance Symbol, Z_L. The impedance presented by a load connected to a generator or other source.

loading **1.** The matching of source impedance to load impedance, usually by means of the introduction of an inductance or capacitance into the load itself. **2.** Any form of impedance matching. **3.** The addition of inductance and/or capacitance to an antenna system to alter the characteristics of the system—especially the resonant frequency. **4.** The modification of the acoustic impedance of a loudspeaker.

loading coil An inductor inserted in a circuit to increase its total inductance or to provide some special effect, such as canceling capacitive reactance. See LOADED ANTENNA and PUPIN COIL.

loading disk Also called *capacitance hat.* A metal disk mounted atop a vertical antenna to increase its effective length and thereby lower its resonant frequency. It also increases the bandwidth of the system.

loading factor The ratio of source impedance to load impedance before the introduction of loading circuits.

loading inductance **1.** The inductance present in a load. **2.** An inductance used as a load.

loading routine Also called *loading program.* A routine permanently in memory; it allows a program to be loaded into memory from an external storage medium.

load life The longevity of a device in terms of the number of hours it can withstand its full power rating.

load line In a group of voltage-current (*EI*) curves, a line connecting points of equal resistance (*E/I*) that are equal to a particular value of load resistance (impedance).

load power The power dissipated in a load.

load regulation Automatic stabilization of load resistance (impedance) at a constant value.

load resistance **1.** The resistance present in a load. **2.** A resistance used as load.

load stabilizer A device for holding load current or load voltage to a constant value.

loadstone Alternate spelling of LODESTONE.

load termination The load connected to the output of a circuit or device as the terminal element in a circuit or system.

load voltage The voltage developed across a load.

load-voltage stabilization Automatic regulation of load voltage.

load wattage See LOAD POWER.

lobe In the directivity pattern of a transducer, a figure, such as a circle or ellipse enclosing an area of intensified response. Applicable especially to antennas.

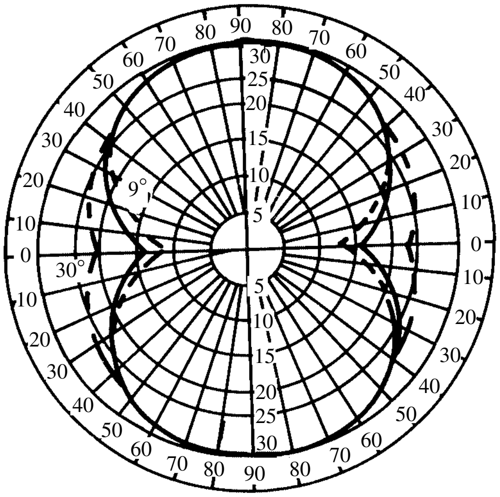

lobe (pattern of horizontal half-wave antenna)

lobing **1.** In a transmitting or receiving antenna, the effect of ground reflection, resulting in phase reinforcement (lobes) at some elevation angles and phase opposition (nulls) at other angles. **2.** The pattern of secondary maxima in the radiation response of a directional antenna.

local action Electrolysis between separate areas of a single electrode immersed in an electrolyte. The action is caused by impurities at different spots in the electrode metal, causing one spot to act as an anode and the other as a cathode, thereby creating a small battery cell.

local alarm In security applications, a visible and/or audible alarm that can be seen and/or heard easily throughout the protected zone.

local area network Abbreviation, LAN. A group of computers and/or terminals that are linked together within a relatively small geographic area, such as a college campus. Interconnections are usually made via cable. There are several different configurations, called *topologies.*

local battery In wire telephony, a battery installed on the subscriber's premises.

local broadcast station A standard broadcast station licensed in the local service. See LOCAL CHANNEL and LOCAL STATION.

local channel A channel in the standard amplitude-modulation (AM) broadcast band, intended to serve only the area near the station. Transmitter power and operating time are restricted to prevent long-distance propagation.

local control The control of a radio transmitter from the site (in contrast to remote control).

local feature focus In robotic systems, the use of only a small portion of the available image data to perform a function. The robot computer recognizes and acts on the data it needs, disregarding the rest.

local feedback Feedback within a circuit stage.

localizer A radionavigation transmitter whose signal guides aircraft to the centerline of a runway.

local oscillator Abbreviation, LO. In a wireless transmitter, one of the oscillators that contributes to the signal that is ultimately modulated and transmitted. The LO output is mixed with the output of a variable-frequency oscillator (VFO). The LO frequency can be switchable at increments of several hundred kilohertz (500 kHz and 1000 kHz are typical), facilitating band changes when mixed with the VFO output. In some transmitters, the LO can operate at a single, fixed frequency if the VFO tunes over an exceptionally wide range.

local program A program that originates at the same single broadcast station from which it is transmitted.

local reception The reception of signals from local stations. Compare LONG-DISTANCE COMMUNICATIONS.

local side The group of circuits and components associated with a communications terminal at a given location.

local station A station situated within the same general area as the receiver, as opposed to a distant station.

local system library A computer program library containing standard software associated with a specific system.

local transmission The sending of signals to receivers in the same general locality as the transmitter, as opposed to long-distance transmission.

local trunk In a telephone system, the interconnecting line between local and long-distance lines.

location In digital computer operation, a memory position (often a register) specified by an address and usually described in terms of the basic storage unit a particular system uses (e.g., a character is a location in a character-oriented machine).

location counter A register in the control section of a computer containing the address of the instruction being executed.

locked groove A continuous blank groove around the inside of a phonograph record. When the disc is done playing, this groove keeps the stylus from running into the label or sliding across the disc.

locked oscillator **1.** A fixed-frequency oscillator, such as a crystal-controlled oscillator. **2.** See BRADLEY DETECTOR.

lock-in A state of synchronism, as when a self-excited oscillator is synchronized (locked-in) with a standard-frequency generator.

lock-in amplifier A detector that makes use of a balanced amplifier. The output is the difference between the collector or drain currents of the two devices.

locking circuit See HOLDING CIRCUIT.

locking relay See LATCHING RELAY.

lock-in relay See LATCHING RELAY.

lock-out **1.** To prevent a hardware unit or routine from being activated (e.g., when there would be a conflict between operations using the same areas of memory). **2.** A safeguard against an attempt to refer to a routine in use.

lock-up relay An electromagnetic relay that can be locked in the actuated state nonmechanically (i.e., by means of an electromagnet or permanent magnet).

locus The set of all points located by stated conditions (e.g., the locus of secondary points that are all equidistant from a primary point is a sphere).

lodestone A natural magnet; a form of the mineral magnetite. Also spelled *loadstone.*

log **1.** Abbreviation of LOGARITHM. **2.** A continuous record of communications kept by a station, or a record of the operation of an equipment.

log$_{10}$ Abbreviation of *common logarithm* (base-10 logarithm). Also called *Briggsian logarithm.*

logarithm Abbreviation, log. The power y to which a fixed number a, called the *base,* must be raised to equal a given number x. Suppose $x = a^y$, where a, x, and y are real numbers. Then, $\log_a x = y$. The most common logarithmic bases are 10 and the transcendental number e, approximately equal to 2.71828. See COMMON LOGARITHM and NAPIERIAN LOGARITHM.

logarithmic amplifier An amplifier whose output-signal amplitude is proportional to the logarithm of the input-signal amplitude.

logarithmic curve A graphical representation of a logarithmic function, having the form $y = k \log_a x$, where k is a nonzero real-number constant, and a is a positive real number (the logarithmic base).

logarithmic decrement See DECREMENT.

logarithmic graph Also called *log-log graph.* A graph in which the x and y axes are both incremented logarithmically. Compare SEMILOGARITHMIC GRAPH.

logarithmic horn A horn whose diameter varies directly, according to the logarithm of the displacement along the axis. See HORN.

logarithmic mean See GEOMETRIC MEAN.

logarithmic meter A current meter or voltmeter whose deflection is proportional to the logarithm of the quantity under measurement. The increments on the scale of such an instrument are closer together in the upper portion.

logarithmic rate of decay See EXPONENTIAL DECREASE.

logarithmic rate of growth See EXPONENTIAL INCREASE.

logarithmic response **1.** Response in which the value of a dependent variable is at every point proportional to the logarithm of the independent variable. **2.** A type of response in which a quantity (such as current) varies directly with the logarithm of another quantity (such as voltage).

logarithmic scale A graduated scale in which the coordinates are positioned, according to the logarithm of the actual distance from the origin.

logarithmic voltmeter See LOGARITHMIC METER.

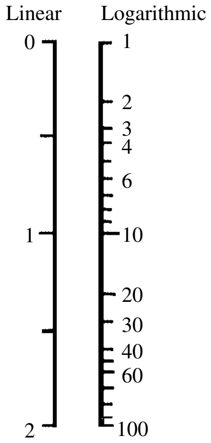

logarithmic scale

log$_e$ Abbreviation of *logarithm to the base e* (the NAPIERIAN LOGARITHM). Also written *ln.*

logic **1.** In digital circuits, the mathematics dealing with the truth or falsity of statements (represented by variables) and their combinations. Also see SYMBOLIC LOGIC. Generally, "truth" is represented by the binary digit 1, and "falsity" is represented by the binary digit 0. **2.** Collectively, the switching circuits and associated hardware for implementing digital functions (see 1), such as AND, NAND, NOR, OR, etc.

logical decision During a computer program run, a choice between alternatives based on specified conditions. For example, one alternative path in a routine might be selected because an intermediate result was negative.

logical diagram A schematic diagram showing the interconnection between gates of a logic circuit.

logical equivalence The condition in which two logical statements have identical truth value for all possible combinations of truth value of their constituents.

logical file A data set composing one or several logical records.

logical implication For logical statements *x* and *y*, the condition that *y* is true whenever *x* is true: If *x*, then *y*.

logical operation **1.** An operation using logical operators: AND, NOR, OR, and NAND. **2.** A processing operation in which arithmetic is not involved (e.g., a shift).

logical operator A word or symbol representing a logic function operating on one or more operands. Examples: NOT, OR, AND, NOR, and NAND.

logical shift A shift operation in which digits in a word are moved to the left or right in circular fashion; digits displaced at one end of the word are returned at the other. Also called CYCLIC SHIFT.

logic array In logic circuits, a redundant arrangement of identical components in a single package.

logic circuit In digital systems, a gating or switching circuit that performs such logical operations as AND, NAND, NOR, OR, and XOR. It can use diodes, transistors, charge-coupled devices, tunnel diodes, thyristors, ferroelectric elements, magnetic-core elements, or a combination of these. It usually consists of diodes and transistors fabricated onto an integrated-circuit (IC) chip.

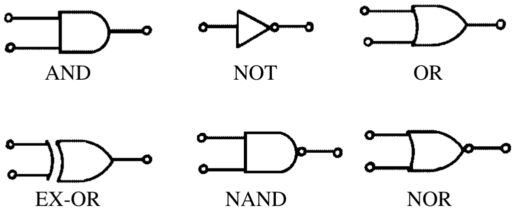

logic circuits (or gates)

logic comparison An operation in which two operands are compared for equal value.

logic connectives Words connecting operands in a logic statement; the truth or falsity of the statements can be determined from their content and the connectives' meanings.

logic diagram **1.** A graphic representation of a logic function (e.g., AND, NAND, NOR, OR, and XOR). **2.** The design of a device or system represented by graphic symbols for logic elements and their relationships.

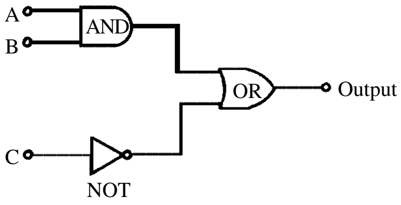

A	B	C	$AB+\bar{C}$
0	0	0	1
0	0	1	0
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	1
1	1	1	1

logic diagram and truth table

logic diode See COMPUTER DIODE.

logic flowchart The logical steps in a program or subroutine represented by a set of symbols.

logic function An expression for an operation involving one or a combination of logic operators.

logic gate See LOGIC CIRCUIT.

logic input current In an integrated circuit, the input current to the logic gate at a certain voltage value.

logic input levels The range of voltages over which the logic trip level occurs, from low to high or high to low. It is usually expressed in volts for the low state and the high state; for example, low is –1 to +2 volts, high is +4 to +6 volts.

logic instruction A command to execute a logical function.

logic level **1.** One of the two logic states 0 or 1 (on or off, high or low). **2.** Of the two logic states, that which represents the "true" condition. **3.** The voltage amplitude of digital signals in a logic system.

logic probe A test probe with a built-in amplifier and (usually) indicating LEDs; its tip is touched to various test nodes in a digital logic circuit to trace logic levels and pulses.

logic relay See BISTABLE RELAY.

logic swing In a logic circuit, the difference between the voltage corresponding to the high state and the voltage corresponding to the low state.

logic symbol **1.** A symbol used to represent a logic element in a circuit diagram. **2.** A symbol used to represent a logic connective.

log-log graph See LOGARITHMIC GRAPH.

LOGO A high-level computer programming language used for robot control and as an education aid. It is especially useful for teaching children how to operate computers and computerized robots. Statements are simple enough so that elementary-school children can learn to write short programs.

log-periodic antenna Also called *log-periodic dipole array (LPDA).* A broad-spectrum, multiband directional antenna in which the lengths and spacing of the radiator and elements increase logarithmically from one end of the antenna to the other.

log polar navigation A computerized navigation system in which polar-coordinate data is converted into rectangular-coordinate data. In the transformation process, the logarithm of the radius (range) is taken. This results in improved image resolution at close range, although it sacrifices resolution at greater ranges.

log taper In a potentiometer or rheostat, resistance variations that correspond to the logarithm of shaft rotation, or vice versa. Compare LINEAR TAPER. Also see TAPER.

long Abbreviation of LONGITUDE.

long-distance communication **1.** Radio communication between stations separated by distances too great for ground-wave propagation to be effective. **2.** In telephone service, communications that require the dialing of an area code in addition to the local exchange number.

long-distance loop A direct telephone line connecting a subscriber's station to a long distance switchboard.

long-distance reception Reception of radio signals from stations beyond the range of ground-wave propagation. See also LONG-DISTANCE COMMUNICATION, **1**.

long-distance transmission Transmission of radio signals to points beyond the range of ground-wave propagation. See also LONG-DISTANCE COMMUNICATION, **1**.

longitude Abbreviation, long. Angular displacement, measured in degrees around the earth's circumference, to the east and west of the prime meridian that passes through Greenwich, England. Compare LATITUDE. Also see MERIDIAN.

longitude effect The variation (caused by the earth's rotation and magnetic field) of the strength of cosmic rays arriving at different longitudes on the surface of the earth.

longitudinal current Current flowing in the same direction in the parallel wires of a pair (the return circuit is via ground).

longitudinal parity Parity associated with bits recorded on one track of a magnetic storage medium to indicate whether the number of bits is even or odd.

longitudinal redundancy A computer condition, generally affecting magnetic tape records, in which the bits in each track of a record do not meet the required parity, as determined by a LONGITUDINAL REDUNDANCY CHECK.

longitudinal redundancy check A parity check performed on a block of characters or bits (for example, on a track of a magnetic disk). A parity character is generated and transmitted as the last character of the block; thus, each longitudinal block has either even or odd parity.

longitudinal wave A wave in which the movement of particles in a medium is parallel with the direction of propagation.

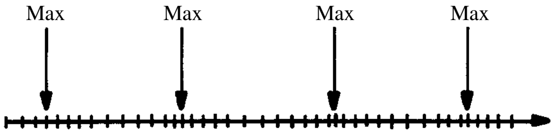

longitudinal wave

long line **1.** A single-wire antenna whose length is greater than the length of the wave fed to it for propagation. Also see LONG-WIRE ANTENNA. **2.** In wire telegraphy, an electrical line that has great physical length. **3.** In electronics theory, a transmission line of indeterminate length, but whose characteristics remain stable and predictable to infinity.

long-persistence screen A cathode-ray-tube screen on which the image remains for a time after the electron beam has passed.

long-play Abbreviation, LP. Descriptive of phonograph discs designed to play at 33.3 revolutions per minute (rpm). Also see MICROGROOVE RECORD.

long-play record See MICROGROOVE RECORD.

long-range navigation See LORAN.

long-range radar Radar that can detect targets at distances of 200 miles or more.

long skip Ionospheric radio-wave propagation, usually via the F layer, between or among stations separated by large geographic distances. The wave angles of departure (from transmitting stations) and arrival (at receiving stations) are very small relative to the horizon. Compare SHORT SKIP.

long-term drift Gradual change in the value of a quantity, such as voltage or frequency, observed over a long period, in contrast to that noted for a brief interval. Compare SHORT-TERM DRIFT.

long-term input offset voltage stability Expressed in microvolts per month. The extent to which the input offset voltage in an integrated circuit stays stable over long periods of time.

long-term stability Stability reckoned over a period of weeks, months or years, as contrasted to that noted for brief intervals of time (minutes or hours).

long throw A speaker design term that describes a woofer moving through long excursions; the objective is to provide good low-frequency response with low distortion.

long waves Low-frequency radio waves, particularly those in the frequency range of 30 kHz to 300 kHz (10 km to 1 km).

long-wire antenna A horizontal or sloping wire antenna measuring a full wavelength or more and fed at a high-current point or at one end. As the wire is made longer, the main radiation/response lobes get more nearly in line with the antenna, and their amplitudes increase. As the wire is made shorter, the main lobes get farther from the axis of the antenna, and their amplitudes decrease. This antenna can produce high gain and excellent low-angle radiation, provided it is straight, is at least several wavelengths long, and is clear of obstructions. However, it cannot be rotated conveniently to change the direction in which maximum gain occurs. Also, a great deal of real estate is needed at medium and high frequencies.

lookup A computer programming technique in which a data item identified by a key is selected from an array.

loop **1.** An electrical circuit consisting of elements connected in series. **2.** In a standing-wave system, a maximum-response point (e.g., *current loop* and *voltage loop*). Compare NODE, **2**. **3.** See LOOP ANTENNA. **4.** A signal path (e.g., *feedback loop*). **5.** A one- or two-turn coil for low-impedance coupling. Also see LINK, **1**. **6.** In a computer program run, the repetitious execution of a series of instructions that terminates when some specified condition is satisfied by a relational test, at which point the next instruction in the main program is obeyed.

loop antenna **1.** A small portable receiving antenna in the form of a wire coil. **2.** A half-wave conductor bent into a circle or square. The conductor is broken at the point opposite the feed point. It can be used for transmitting and receiving. **3.** A full-wavelength, continuous conductor bent into a circle or square. It can be used for transmitting and receiving.

loop checking A method of checking the accuracy of data transmitted over a data link by returning signals received at one terminal to the transmitting terminal for comparison with the original data.

looped amplification See FEEDBACK FACTOR.

looping plug A double phone-plug unit for simultaneously plugging into two phone jacks. Completes (loops) the circuit between the two jacks.

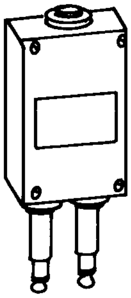

looping plug

loop-input signal A signal introduced into a feedback control loop.

loop pulsing The regular, intermittent breaking of the direct-current path at the transmitting end of a transmission line; also called DIAL PULSING.

loop resistance The resistance of the electrical path around a complete loop (see LOOP, **1**).

loop response time In a security system, the length of time between the first sensing of an abnormal condition (e.g., an intrusion) and the recognition of that condition by the controller.

loopstick antenna See FERRITE-ROD ANTENNA.

loop test A means of locating a discontinuity in a circuit by creating a closed loop, including the suspected fault point.

loose coupling Coupling that transfers only small amounts of energy, as when a primary and secondary coil are spaced so far apart that the coefficient of coupling is small. Compare CLOSE COUPLING.

loosely coupled twin In computer operations, a system in which two processors, each having its own operating system, are used with switches so that they can use common peripherals. Also see SWITCH.

lopplar Acronym for *laser Doppler radar.* Also abbreviated as ladar.

lorac A radio-navigation system that operates by means of phase comparison. Similar to LORAN. Trade name of Seismograph Service Corporation.

loran A long-range radionavigation system in which two pairs of ground stations transmit pulsed signals that are used by aircraft and ships to determine their positions. The name is an acronym for *long-range navigation.*

loran C A radionavigation system that operates at a frequency of 100 kHz. It operates on the hyperbolic principle.

loran D A radionavigation system similar to loran C. It is used by aircraft, operates independent of ground stations, and prevents unwanted enemy detection of aircraft position.

Lorentz force For a charge Q moving with velocity v in a magnetic field B and an electric field E, the force $F = Q(E + (v \times B))$.

loss Energy that is dissipated without doing useful work in a circuit or system. See POWER LOSS.

loss angle For an insulating material, 90 degrees minus the PHASE ANGLE.

loss index For an insulating material, the product of the POWER FACTOR and the DIELECTRIC CONSTANT.

lossless data compression A process in which the number of bits in a data file is reduced by eliminating redundancies, without sacrificing any of the precision or detail in the file. Compression of text and programs must usually be lossless. Image compression need not generally be lossless. Compare LOSSY DATA COMPRESSION.

lossless line A perfect transmission line (i.e., one having no resistance loss, no dielectric loss, and no radiation loss). It is not realizable in practice, but useful in some theoretical calculations.

loss tangent See DISSIPATION FACTOR, **1**.

lossy data compression A process in which the number of bits in a data file (especially a digital image) is reduced by eliminating redundancies, with some sacrifice of precision or detail. In image compression, some loss can usually be tolerated, allowing larger compression ratios than would be possible if zero loss were mandatory. Lossy compression is not generally acceptable for text files and programs. Compare LOSSLESS DATA COMPRESSION.

lossy line A line or cable having comparatively high or excessive attenuation per unit length.

loudness The amplitude of sound, especially in audio reproduction equipment, such as a high-fidelity stereo amplifier. Also called VOLUME.

loudness control See COMPENSATED VOLUME CONTROL.

loudness curves See AUDIBILITY CURVES.

loudness switch/button In a high-fidelity audio amplifier, a switch or button that can be actuated when music is played at low loudness. Increases the volume of the bass relative to the midrange and treble.

loudspeaker A transducer that converts electrical impulses into sound waves of sufficient volume to be heard easily by a number of listeners situated at some distance from the device. Also called *speaker.*

loudspeaker damping See DAMPED LOUDSPEAKER.

loudspeaker dividing network See CROSSOVER NETWORK.

low **1.** The logical digit 0. **2.** Of relatively small magnitude (e.g., LOW VOLTAGE and LOW-FREQUENCY). **3.** In a rechargeable cell or battery, the condition of being near the end of the discharge cycle.

low band **1.** The low or lowest frequency band used in communications, testing, or processing in a given situation. **2.** Television channels 2 to 6 (54 to 88 MHz). **3.** In two-way radio operations, those radio channels between 30 MHz and about 70 MHz.

low battery **1.** The condition of a battery needing replacement or recharging. **2.** An indicating device that shows when a battery needs to be replaced or recharged.

low-battery bias current The current into a designated pin of an integrated circuit required for proper operation of the LOW-BATTERY INDICATOR.

low-battery indicator A device, such as a light-emitting diode (LED) and associated circuitry, that gives a visible indication of the condition of LOW BATTERY.

low-capacitance probe A test probe in which capacitance has been minimized to reduce loading and detuning of the circuit under test.

low earth orbit Acronym, LEO. An artificial satellite orbit that is comparatively low in altitude, resulting in a short orbital period (in some cases less than 2 hours). See also LEO SATELLITE SYSTEM.

low-energy criterion See VON HIPPEL BREAKDOWN THEORY.

lower sideband Abbreviation, LSB. In an amplitude-modulated wave, the lower band of frequencies equal to the difference between the carrier frequency and the modulating frequency. Compare UPPER SIDEBAND.

lower sideband suppressed carrier Abbreviation, LSSC. A single-sideband transmission technique in which the lower sideband is transmitted, but the upper sideband and carrier are suppressed. Compare DOUBLE SIDEBAND SUPPRESSED CARRIER and UPPER SIDEBAND SUPPRESSED CARRIER.

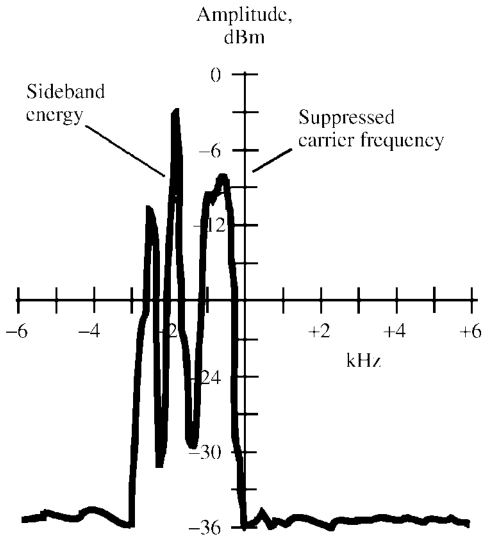

lower sideband

lowest usable frequency Abbreviation, LUF. The lowest frequency that can be used successfully at a given time for communication via the ionosphere. Compare MAXIMUM USABLE FREQUENCY.

low filter A highpass filter that removes low-frequency audio noise from the modulating waveform of a broadcast station. The result is a lower level of transmitted hum and rumble.

low-frequency Abbreviation, LF. **1.** Pertaining to radio frequencies in the band from 30 kHz to 300 kHz (wavelengths from 10 kilometers to 1 kilometer). Also see RADIO SPECTRUM. **2.** Pertaining to audio frequencies below 500 Hz.

low-frequency compensation **1.** In video-amplifier design, special measures, such as use of high coupling and bypass capacitances, to boost low-frequency gain. **2.** Use of special circuits to increase the low-frequency response of an audio amplifier. Also see BASS BOOST, **1**, **2**.

low-frequency direction finder Abbreviation, LDF. A direction finder operated in or below the standard amplitude-modulation (AM) broadcast band, that is, below 1.605 MHz.

low-frequency padder See OSCILLATOR PADDER.

low-frequency parasitics Parasitic oscillations of a frequency lower than that being processed by the amplifier or generated by the oscillator in which they occur.

low-level **1.** A logic term for the more negative of the two (binary) logic levels. **2.** Having an amplitude that is below that normally available in comparable circuits or systems. **3.** In computer operations, pertaining to programming languages (such as assembly language or machine language) that control the machine, but do not directly interface with the operator.

low-level audio signal In audio operations, a signal that has not been amplified by any means (e.g., the output of a dynamic microphone). Compare HIGH-LEVEL AUDIO SIGNAL.

low-level contact A switch or relay contact intended for use with low values of current and voltage.

low-level input current **1.** A test used to check an input pull-up resistor in an integrated circuit to ensure that the fan-in is as specified. **2.** The current flowing from an input when the highest low-level output voltage specified is applied to the input of the device.

low-level language A computer programming language in which each instruction has only one equivalent machine code. Examples are *machine language* and *assembly language.* Compare HIGH-LEVEL LANGUAGE.

low-level logic Abbreviation, LLL. In digital-computer operations, any logic system that operates at low voltage or current levels.

low-level modulation Modulation of a radio or television transmitter at a stage preceding the final radio-frequency (RF) power amplifier.

low-level output current **1.** A test to ascertain that the fan-out and current-sinking capability of an integrated circuit are as specified. **2.** The current flowing into an output with input conditions that cause the output to be at logic low.

low-level signal **1.** A signal with small amplitude. **2.** A signal with peak-to-peak voltage so low that it does not drive an amplifier circuit out of the linear range of operation.

low-loss material A material, particularly a dielectric, having low electrical loss at a given frequency. Also see LOSS.

low-noise Pertaining to circuits, especially weak-signal communications receiving amplifiers and converters, designed to generate the smallest possible amount of internal noise.

low-noise down converter In a satellite television receiving system, a circuit that converts the signals from the dish antenna to frequencies that correspond to the channels on a conventional television set.

low order The lesser-value place(s) of characters or digits in the hierarchy of a group (number or word). For example, 5 and 6 are low-order digits in the number 123,456.

low-order position The extreme right-hand (least significant) position in a number or word.

low-pass filter A combination of capacitance, inductance, and/or resistance, intended to produce large amounts of attenuation above a certain frequency and little or no attenuation below that frequency. The frequency at which the transition occurs is called *cutoff.* At cutoff, the power attenuation is 3 dB with respect to the minimum attenuation. At frequencies below cutoff, the power attenuation is less than 3 dB. At frequencies above cutoff, the power attenuation is more than 3 dB. The simplest circuit consists of a series inductor or a parallel capacitor. The inductance-capacitance (LC) circuit has a combination of series inductors and parallel capacitors. In the resistance-capacitance (RC) circuit, resistors are substituted for the inductors. Compare BANDPASS FILTER, BAND-REJECTION FILTER, HIGH-PASS FILTER.

low power Abbreviation, LP. Power considerably lower than that ordinarily encountered in a

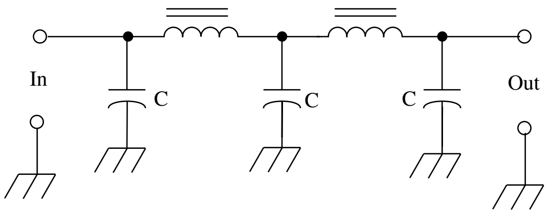

low-pass filter

particular application. The term is arbitrary; several hundred watts might be regarded as low power in one situation, whereas a fraction of a watt would be implied in another.

low-print recording tape Magnetic tape that is less susceptible to print-through than conventional tape.

low *Q* For a component or circuit, a low quotient for the ratio of reactance to resistance (X/R). This is a relative term because a particular Q value considered low in one situation might be high in other circumstances. Also see FIGURE OF MERIT, **1**.

low tension See LOW VOLTAGE.

low voltage **1.** A voltage considerably lower than that ordinarily encountered in a particular application. The term is arbitrary; several hundred volts might be regarded as low in one situation, and a fraction of a volt would be implied in another. **2.** In a television receiver, the supply voltage applied to all points other than the high-voltage circuit or the picture tube.

low-voltage rectifier In a television receiver, the rectifier that supplies power for the low-voltage stages. See LOW VOLTAGE, **2**.

LP **1.** Abbreviation of LOW POWER. **2.** Abbreviation of LONG PLAY. **3.** Abbreviation of *low pressure.*

L pad An attenuator consisting of one series arm and one shunt arm, arranged in such a way that the schematic representation of the circuit resembles an inverted capital letter L. It is noted for its constant input resistance or impedance as the amount of attenuation is varied.

LPB Abbreviation of LIGHTED PUSHBUTTON.

L_P band A section of the L BAND extending from 390 to 465 MHz.

lpm Abbreviation of *lines per minute*: the output speed of a line printer.

lpW Abbreviation of *lumens per watt*; lm/W is preferred.

Lr Symbol for LAWRENCIUM.

L + R, L – R The sum and difference of the left (L) and right (R) channel signals in a stereo high-fidelity sound system. The L + R signal is the in-phase combination of the two channels; the L – R signal is the out-of-phase combination.

L regulator See L-TYPE VOLTAGE REGULATOR.

LRR Abbreviation of LONG-RANGE RADAR.

LR time constant See INDUCTANCE-RESISTANCE TIME CONSTANT.

LSA diode Abbreviation for *limited-space-charge-accumulation diode.* A solid-state diode that acts as a microwave oscillator.

LSB **1.** Abbreviation of LOWER SIDEBAND. **2.** Abbreviation of LEAST-SIGNIFICANT BIT.

L_S band A section of the L band that extends from 900 to 950 MHz.

LSC Abbreviation of LEAST-SIGNIFICANT CHARACTER.

LSD Abbreviation of LEAST-SIGNIFICANT DIGIT.

L section **1.** A filter section whose schematic representation has the general shape of an inverted capital letter L. **2.** An attenuator circuit whose schematic representation has the general shape of an inverted capital letter L. **3.** A network section consisting of a series (input) impedance arm and a shunt (output) impedance arm. See L PAD. **4.** A right-angle bend in coaxial cable (see ELL).

LSI See LARGE-SCALE INTEGRATION.

LSSC Abbreviation of LOWER SIDEBAND SUPPRESSED CARRIER.

L_T band A section of the L BAND extending from 780 to 900 MHz.

LTROM Abbreviation of LINEAR-TRANSFORMER READ-ONLY MEMORY.

L-type antenna See INVERTED-L ANTENNA.

L-type voltage regulator A simple voltage regulator containing a series current-limiting resistor and shunt regulator (zener diode, VR tube, voltage-dependent resistor, etc.). The schematic representation resembles an inverted capital letter L.

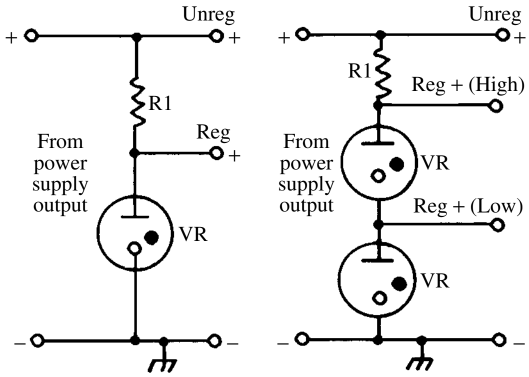

L-type voltage regulator

lt-yr Abbreviation of LIGHT-YEAR.

Lu Symbol for LUTETIUM.

Lucalox General Electric's translucent ceramic; its chief constituent is polycrystalline alumina. The material has many applications in electro-optics.

Lucite Trade name for METHYL METHACRYLATE RESIN.

Luddites During the Industrial Revolution, people who sabotaged automated factory equipment because they believed the machines would put them out of work. The word comes from the name of Ned Ludd, their supposed leader. The term is sometimes applied nowadays to people who fear computers, robots, and other technological innovations because of real or imagined threats to personal job security.

LUF Abbreviation of LOWEST USABLE FREQUENCY.

lug **1.** A contact attached to the end of a wire lead to facilitate connection to a binding post. **2.** A contact attached to a terminal strip, to which wire leads are soldered.

lum Abbreviation of LUMEN. The preferred (SI) form is lm.

lumen Abbreviation, lm, and sometimes l or lum. The SI unit of luminous flux; it is equal to the light that is emitted in one steradian (the unit solid angle) by a uniform point source of one candela. Also see CANDLE POWER, ILLUMINANCE, LUMINOUS INTENSITY, SOLID ANGLE, and STERADIAN.

lumen-hour Abbreviation, lm-hr. The amount of light that a source having a luminous flux of one LUMEN delivers in a time period of one hour.

luminaire A complete and self-contained lighting system, for television-studio use or photographic use. The kit includes all of the needed parts and accessories.

luminance The amount of light emitted or scattered by a surface. This property is expressed in *candelas per square meter* (cd/m^2).

luminance channel In a color television circuit, the channel that processes the Y SIGNAL.

luminance signal See Y SIGNAL.

luminescence The production of visible light, but not heat, by a material stimulated by radiation or electron bombardment. See ELECTROLUMINESCENT CELL and LUMINESCENT SCREEN.

luminescent cell See ELECTROLUMINESCENT CELL.

luminescent screen A cathode-ray tube whose screen is coated with a material that glows under the influence of ionizing radiation, X rays, or electron beams.

luminiferous ether See ETHER, **1**.

luminosity The luminous efficiency of radiant energy, as given by the ratio of luminous flux to radiant flux (lumens per watt) for a specific wavelength.

luminosity factor Abbreviated K, and expressed in lumens per watt. The luminous intensity divided by the actual radiant intensity at a given wavelength of visible light.

luminous energy The energy in visible electromagnetic radiation.

luminous flux The rate of transfer or flow of luminous energy.

luminous intensity Luminous flux through a unit solid angle, expressed in candelas. Also see CANDELA.

lumistor An amplifier or coupling device in which the input signal varies the brilliance of a lamp, electroluminescent cell, or light-emitting diode, and a photocell (or other light-sensitive device) picks up the fluctuating light and uses it to modulate an output current. In a compact lumistor, the light-emitting and light-sensing components are separate layers in a wafer or block of material. Compare LIGHT AMPLIFIER.

lumped Pertaining to a property that is concentrated at or around a single point, rather than being distributed through a circuit (e.g., *lumped capacitance* and *lumped inductance*).

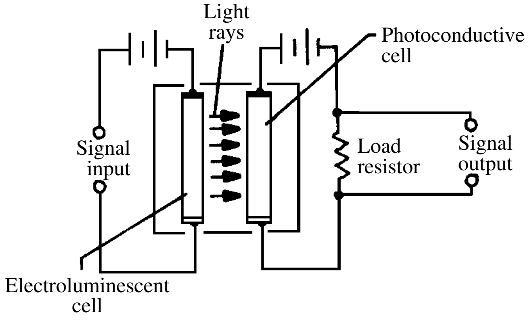

lumistor

lumped capacitor See DISCRETE CAPACITOR.

lumped component A discrete component (i.e., one that is self-contained). Compare DISTRIBUTED COMPONENT.

lumped constant The total value of any single electrical property in an electrical or electronic component.

lumped-constant delay line A delay line having discrete capacitance and inductance components. Compare DISTRIBUTED-CONSTANT DELAY LINE.

lumped impedance A reactance and/or resistance manifested in a definite location. Examples are ordinary components, such as capacitors, inductors, and resistors.

lumped inductor See DISCRETE INDUCTOR.

lumped parameter Any circuit parameter that can be considered as a discrete parameter—even if it is not made up of a single component.

lumped resistor See DISCRETE RESISTOR.

lutetium Symbol, Lu. A metallic element of the rare-earth group. Atomic number, 71. Atomic weight, 174.967.

lux The unit of illuminance, equivalent to one LUMEN per square meter.

Luxemberg effect The generation of interference by cross-modulation of two or more signals whose paths intersect in the same region of the ionosphere.

luxmeter A device for measuring visible illuminance.

LV Abbreviation of LOW VOLTAGE.

LVDT Abbreviation of LINEAR VARIABLE DIFFERENTIAL TRANSFORMER.

Lw Symbol for LAWRENCIUM; more commonly *Lr.*

lx Abbreviation of LUX.

L_X band A section of the L BAND that extends from 950 MHz to 1.15 GHz.

L_Y band A section of the L BAND that extends from 725 to 780 MHz.

L_Z band A section of the L BAND that extends from 1.450 to 1.550 GHz.

LZT Abbreviation of LEAD ZIRCONATE-TITANATE, a ceramic used in electronics.

m

M **1.** Abbreviation of prefix MEGA-. **2.** Symbol for MUTUAL INDUCTANCE. **3.** Symbol for MODIFIED INDEX OF REFRACTION.

m **1.** Abbreviation of prefix MILLI-. **2.** Symbol for MASS. **3.** Abbreviation of METER. **4.** Abbreviation of MILE. (Also, mi.) **5.** Symbol for MODULATION COEFFICIENT.

m² Abbreviation of *square meter*, the SI unit of area.

m³ Abbreviation of *cubic meter*, the SI unit of volume.

MA **1.** Abbreviation of MAGNETIC AMPLIFIER. (Also, *magamp.*) **2.** Abbreviation of MEGAMPERE.

mA Abbreviation of MILLIAMPERE.

Mache unit A unit of radioactivity equivalent to 13.47 disintegrations per second (3.64×10^{-10} curie) per liter. It represents the concentration of radon gas per liter (when all radiation is absorbed) that will result in a saturation current of 10^{-3} esu (not to be confused with MACH NUMBER).

machine address See ABSOLUTE ADDRESS.

machine code See MACHINE LANGUAGE.

machine cycle In a machine whose operation is periodic, a complete sequence constituting a period of operation.

machine error In a computer or data-processing system, an error attributable to a hardware failure, rather than to a software fault.

machine instruction A computer program instruction written in MACHINE LANGUAGE.

machine knowledge General term for data stored in an artificially intelligent computer system, and the ability of the computer to use that data in meaningful ways.

machine language Computer program instructions and data represented in binary form. In the hierarchy of programming languages, it is the lowest; the computer works directly with it. All high-level languages are translated to machine language by an assembler, compiler, interpreter, or monitor system.

machine learning In artificial intelligence, a computer's ability to learn through repeated calculations for particular problems.

machine logic **1.** The way that a computer's functional parts are interrelated. **2.** The facility whereby a computer solves problems.

machine operation The performance by a computer of a built-in function (e.g., subtraction).

machine operator A person participating in implementing and overseeing the processing of computer programs.

machine word In computer operations, the address of a memory location composed of the full number of bits normally handled by each register of the machine.

machining In industrial robotics, the mechanical modification of parts during assembly. Examples: drilling, welding, sanding, polishing, and painting.

Mach number For a medium such as air, the ratio of the speed of a body in motion to the speed of sound in the medium. (Not to be confused with MACHE UNIT.)

macro **1.** A control shortcut, in which a function requiring the actuation of several switches is abbreviated, via a microcomputer, so that it can be executed by actuating only one or two switches. **2.** Abbreviation of MACROINSTRUCTION.

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

macro- Prefix denoting extremely large. Compare MICRO-.

macro assembly program An assembly program whose source statements are translated to several machine-language instructions.

macroinstruction A source program instruction that becomes several machine-language instructions when operated on by a compiler.

macroknowledge In artificial intelligence, knowledge in the large sense (i.e., knowledge about information). Example: a set of definitions in an expert system. Compare MICROKNOWLEDGE.

macroprogram A computer program consisting of macroinstructions.

macrosonics The theory and applications of high-amplitude sound waves.

madistor A component that produces changes in current by means of magnetic-field effects. It is used as an oscillator or amplifier.

MADT Abbreviation of MICROALLOY DIFFUSED TRANSISTOR.

MAG Abbreviation of MAXIMUM AVAILABLE GAIN.

magamp Acronym for MAGNETIC AMPLIFIER.

magazine A tape or film cartridge.

magenta One of the primary pigments used in color printers. It has a pinkish-red hue.

magnal CRT base An 11-pin base typical of many cathode-ray tubes.

magnesium Symbol, Mg. A metallic element. Atomic number, 12. Atomic weight, 24.305.

magnesium fluoride phosphor A substance used as a phosphor coating on the screen of a very-long-persistence cathode-ray tube. The fluorescence and phosphorescence are orange.

magnesium silicate phosphor A substance used as a phosphor coating on the screen of a cathode-ray tube. The fluorescence is orange-red.

magnesium tungstate phosphor A substance used as a phosphor coating on the screen of a cathode-ray tube. The fluorescence is very light blue.

magnet A device or body of material that has the ability to attract to itself pieces of iron and other magnetic metals, and the ability to attract or repel other magnets. Also see ELECTROMAGNET, PERMANENT MAGNET, and TEMPORARY MAGNET.

magnet armature See KEEPER.

magnet battery A group of several magnets placed together in parallel (i.e., with similar poles touching or resting nearby) to act as a single magnet.

magnet charger A device that produces an intense magnetic field for restoring weakened magnets or for making new magnets.

magnetic **1.** Pertaining to MAGNETISM. **2.** Possessing MAGNETISM. **3.** Capable of being magnetized. **2.** See MAGNETIC MATERIAL.

magnetic air-gap A space between two magnetic poles, either the same (in which case the force is repulsive) or opposite (in which case the force is attractive).

magnetic amplifier An iron-core device that uses the principle of the saturable reactor to obtain amplification. In its simplest form, it consists of input and output coils wound on a core of square-loop magnetic metal. The input coil consists of two identical windings connected in series-opposition so that currents in the output winding cannot induce voltage in the input winding. The output coil is connected in series with a load and an alternating-current (ac) supply. A small ac signal applied to the input winding causes a large change in the impedance of the output winding and, therefore, a large change in the voltage across the load.

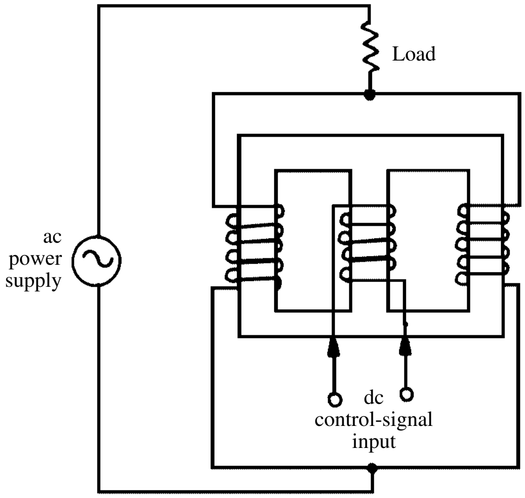

magnetic amplifier

magnetic analysis See MASS SPECTROMETER.

magnetic attraction **1.** The force that causes a magnetic pole to draw to itself an opposite magnetic pole. Thus, a north pole attracts a south pole, and a south pole attracts a north pole. Compare MAGNETIC REPULSION. **2.** The force that causes a magnetic pole to draw to itself a piece of magnetic material, such as iron or steel.

magnetic axis A straight line joining the poles of a magnet.

magnetic bearing The azimuth, or compass direction, measured with respect to magnetic north (the direction of the north geomagnetic pole). It is usually expressed in degrees and can be read directly from a compass.

magnetic azimuth An azimuth bearing relative to magnetic north (the direction of the north geomagnetic pole).

magnetic balance An instrument for measuring the force, either attractive or repulsive, between two objects that are magnetized, or between a magnet and a magnetic substance. The device can also be used for measuring the intensity of a magnetic field, either from a permanent magnet, an electromagnet, or from the earth.

magnetic bias A steady magnetic force applied to another magnetic field to set the latter's quiescent point (e.g., sensitizing a relay by using a permanent magnet to lower the relay draw-in point).

magnetic blowout **1.** The extinction of an electric arc by a strong magnetic field. **2.** The apparatus for accomplishing the action described in **1**.

magnetic bottle A container envisioned for atomic fusion reactions, and that would consist of a magnetic field. Conventional containers cannot withstand the extremely high temperatures involved in atomic fusion.

magnetic braking See ELECTROMAGNETIC BRAKING.

magnetic bridge An instrument comparable to the WHEATSTONE BRIDGE, used to measure magnetic permeability.

magnetic bubble memory See BUBBLE MEMORY.

magnetic capacity The maximum magnetization a given material can receive.

magnetic card A computer storage medium in the form of a card that can be selectively magnetized or imprinted with magnetic ink to represent data.

magnetic cartridge A variable-reluctance phonograph pickup. As the stylus moves in the groove, the vibrations are translated into electric currents by a magnet and coil.

magnetic centering Centering the beam in a television picture tube by means of an electromagnetic focusing coil, a permanent magnet, or both.

magnetic character A letter, numeral, or other symbol written or printed in (visible) magnetic ink for its automatic sensing or reading in computing and signaling operations.

magnetic circuit The closed path determined by a line of magnetic flux or by a set of lines of flux.

magnetic clutch A clutch in which the magnetism of one rotating member causes a second member to lock in and rotate. There need not be physical contact between the two.

magnetic coil The winding in an electromagnet or similar device.

magnetic compass A direction-indicating device using a horizontally suspended magnetic needle as the indicator. The needle tends to point in the direction of the north geomagnetic pole. Compare GYROCOMPASS.

magnetic component See ELECTROMAGNETIC COMPONENT.

magnetic conductivity See PERMEABILITY.

magnetic constant The absolute permeability of free space. It is approximately 1.26×10^{-6} H/m.

magnetic controller A controller that uses electromagnets for some of its functions.

magnetic core The iron core of an electromagnet, choke, transformer, relay, or similar device.

magnetic coupling See INDUCTIVE COUPLING.

magnetic course In navigation, a course referenced to geomagnetic north, rather than geographic north.

magnetic crack detector See ELECTROMAGNETIC CRACK DETECTOR.

magnetic creeping A gradual increase in the magnetization of a material under the influence of a steady magnetizing force.

magnetic cycle **1.** For a material in an alternating magnetic field, the change in magnetic flux as a function of time. **2.** The change in the magnetic-field polarity of the earth. This polarity reverses every few thousand years.

magnetic damping The production of a damping effect or drag in a machine or meter by means of magnetic action on a moving member, in accordance with LENZ'S LAW.

magnetic declination See DECLINATION.

magnetic deflection See ELECTROMAGNETIC DEFLECTION.

magnetic density The concentration of magnetic flux in a region, expressed as the number of lines per unit area of cross section.

magnetic dip At a particular location on the earth's surface, the angle between the terrestrial magnetic field and a horizontal line.

magnetic dipole **1.** A molecule or particle with a north and south magnetic pole. **2.** Any pair of adjacent north and south magnetic poles.

magnetic direction finder Abbreviation, MDF. A type of compass operated by an electric signal delivered by a gyrostabilized magnetic-compass movement.

magnetic disk A rotating disk coated with a layer of magnetic material for the recording, storage and retrieval of information. They are available in various sizes, configurations, and storage capacities. Commonly used with personal computers. Also see DISKETTE and HARD DISK.

magnetic doublet See DOUBLET, **2**.

magnetic drive A device in which mechanical movement is conveyed from one moving part to another by means of a magnetic clutch.

magnetic drum See DRUM.

magnetic effect of electric current The presence of a magnetic field around a conductor carrying electric current.

magnetic equator Also called *geomagnetic equator.* An imaginary circle around the earth, along which a magnetic needle shows no dip. It is near, but slightly displaced from, the geographic equator, and is midway between the geomagnetic poles.

magnetic feedback Feedback by means of inductive coupling between the output and input circuits of a system. It can be positive or negative.

magnetic field The space around a magnetic pole or magnetized body in which magnetic energy acts.

magnetic field intensity See MAGNETIC INTENSITY.

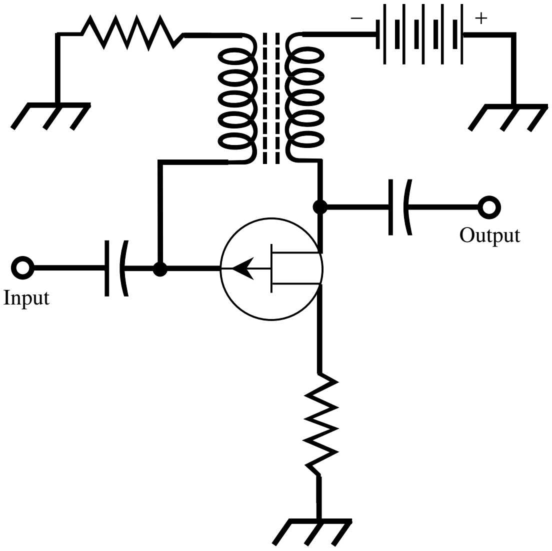

magnetic feedback

magnetic field strength See MAGNETIC INTENSITY.

magnetic-field viewer A device for visually examining a magnetic field. It consists of a clear plastic watchcase filled with iron-oxide particles in liquid suspension. When it is placed within a magnetic field, the particles align themselves in the direction of the magnetic lines of flux.

magnetic-film memory A magnetic memory in which memory cells consist of a thin film (thick film in some instances) of a magnetic material deposited on a substrate. Information is written into and read out of the cell through coils. Also called *thin-film memory.*

magnetic flip-flop A bistable multivibrator using magnetic amplifiers or square-loop cores in place of transistors.

magnetic flux The intensity of a magnetic field through a given area. The unit of magnetic flux is the weber, and the symbol is F. It can loosely be expressed as the number of lines passing through a region of a certain area or of a unit area, such as one square meter. See FLUX.

magnetic flux density See FLUX DENSITY.

magnetic flux linkage The passage of magnetic lines of flux through separate materials or circuits, thereby coupling them magnetically.

magnetic focusing See ELECTROMAGNETIC FOCUSING.

magnetic force The force exerted by a magnet on a body of magnetic material, or on another magnet, within its field.

magnetic friction **1.** See HYSTERESIS, **1**. **2.** The resistance experienced by a magnetic material moving in a magnetic field.

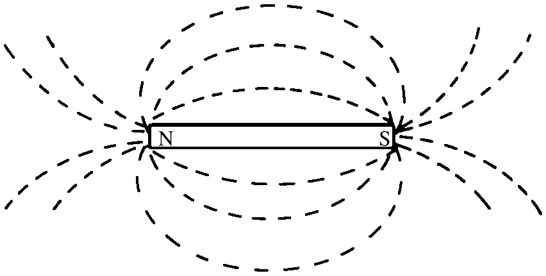

magnetic flux

magnetic gap A space separating the materials in a magnetic circuit. This break is either an air space or one filled with a comparatively thin piece of nonmagnetic material (e.g., the gap in a choke-coil core).

magnetic head See MAGNETIC PICKUP HEAD and MAGNETIC RECORDING HEAD.

magnetic hysteresis See HYSTERESIS, **1**.

magnetic inclination See MAGNETIC DIP.

magnetic induction **1.** The magnetization of a magnetic material, such as iron or steel, when it is placed in a magnetic field. **2.** The induction of an alternating voltage in a conductor by a nearby alternating magnetic field. Also see ELECTROMAGNETIC INDUCTION.

magnetic ink Writing or printing ink that is a suspension of finely divided particles of magnetic material. Also see MAGNETIC CHARACTER.

magnetic instability **1.** The tendency of a magnetic recording medium to deteriorate with time. **2.** Any fluctuation in the intensity of a magnetic field.

magnetic intensity The free-space strength of a magnetic field at a particular point. Specifically, the force (in dynes) that the magnetic field would exert on a unit magnetic pole placed at that point.

magnetic iron oxide See MAGNETITE.

magnetic leakage The usually undesired extension of magnetic flux beyond the confines of a magnetic body, such as the core of a choke.

magnetic lens See ELECTROMAGNETIC LENS.

magnetic line of flux See LINE OF FLUX, **2**.

magnetic load An electromagnetic device operating on the output of an electrical source. Such devices include actuators, alarms, electromagnets, magnetic tapes and disks, relays, and loudspeakers.

magnetic loudspeaker See MAGNETIC SPEAKER, **1**, **2**.

magnetic material **1.** A material, such as magnetite, that exhibits natural magnetism. **2.** A material, such as iron or steel, that is capable of being magnetized.

magnetic media Any medium that stores data as tiny magnetic fields; in particular, MAGNETIC DISK or MAGNETIC TAPE.

magnetic memory **1.** See RETENTIVITY. **2.** A digital memory circuit using magnetic fields to store data bits. Example: BUBBLE MEMORY.

magnetic meridian The circle of the celestial sphere that passes through the zenith and earth's magnetic poles.

magnetic mine A naval mine detonated by a magnetic switch that is closed by the proximity of the steel hull of a ship.

magnetic modulator A core-type device that is somewhat similar to a magnetic amplifier used for amplitude modulation. Modulating current passes through the control winding, and the carrier current through the output winding.

magnetic moment Unit, joule per tesla. For a magnet, the product of pole strength and the distance between poles.

magnetic needle The pivoted magnetic pointer in a magnetic compass.

magnetic north See NORTH MAGNETIC POLE.

magnetic oxide Iron oxide used as the sensitive coating of magnetic recording tape.

magnetic pickup **1.** A phonograph pickup of the variable-reluctance type (see VARIABLE-RELUCTANCE PICKUP). **2.** A magnetic transducer (such as a phono cartridge, tape recording head, or similar input element).

magnetic-pickup head In a tape recorder, the transducer that receives up magnetic impulses from the passing tape and converts them into alternating currents. These currents are amplified to obtain the original sound. Compare MAGNETIC RECORDING HEAD.

magnetic-plate wire Wire in which a magnetic metal has been plated on top of a nonmagnetic metal.

magnetic poles **1.** The points in a MAGNET at which the magnetic lines of flux converge. **2.** The points on the earth at which the geomagnetic lines of flux converge. See NORTH MAGNETIC POLE and SOUTH MAGNETIC POLE.

magnetic pressure See MAGNETOMOTIVE FORCE.

magnetic printing **1.** Also called *magnetic print-through.* In a recording material, such as magnetic tape, the transfer of information from one part of the material to another part (or from one medium to another) by the magnetic field of the recorded material. This phenomenon, which is also called PRINT-THROUGH, sometimes occurs in recording tape on a reel. **2.** Conventional lithography, letterpress, or other reproduction process in which MAGNETIC INK is used.

magnetic print-through See MAGNETIC PRINTING, **1**.

magnetic probe A loop or coil inserted in an electromagnetic field to sample the magnetic component. See, for example, WAVEGUIDE PROBE. Compare ELECTRIC PROBE.

magnetic recording **1.** The recording of sounds or data by varying the magnetization of a medium, such as a magnetic disk or tape. **2.** A magnetic medium on which data has been recorded.

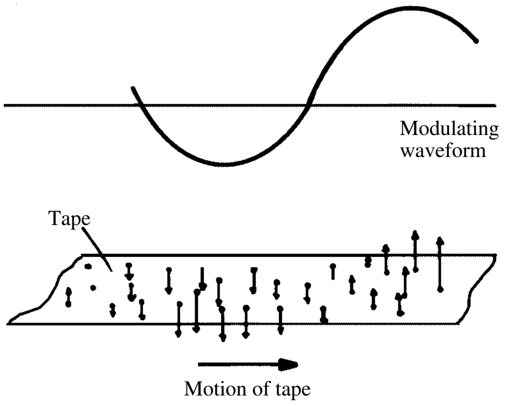

magnetic recording

magnetic recording head In a tape recorder, the transducer that receives current impulses (analogs of the original sound vibrations) from an amplifier and converts them into magnetic impulses that magnetize spaces on the passing tape. Compare MAGNETIC PICKUP HEAD.

magnetic recording medium **1.** A magnetic cylinder, disk, drum, tape, or wire used in the recording of sound or data. **2.** The sensitive material with which any of these is coated.

magnetic relay A relay having a permanent magnet in whose field a coil, bar, or reed moves to open or close a pair of contacts.

magnetic remanence See RESIDUAL MAGNETISM.

magnetic repulsion The force that causes a magnetic pole to push away a similar magnetic pole, although they are not in mutual contact. Thus, two north poles repel each other, and so do two south poles. Compare MAGNETIC ATTRACTION.

magnetic-resonance accelerator See CYCLOTRON.

magnetics **1.** Collectively, magnetic components and equipment. **2.** Collectively, magnetic materials. **3.** A branch of physics dealing with magnets and magnetism.

magnetic saturation The condition in which a magnetic material passes all of the magnetic lines of flux that its permeability allows. Increasing the intensity of the magnetizing force will produce no increase in magnetization.

magnetic scan See ELECTROMAGNETIC DEFLECTION.

magnetic screen See ELECTROMAGNETIC SHIELD.

magnetic shield See ELECTROMAGNETIC SHIELD and MAGNETIC SHIELDING.

magnetic shielding **1.** Enclosing a magnetic field to confine its flux, thus preventing interaction with outside bodies. **2.** Devices (such as boxes, cans, or shells of iron, steel, or a magnetic alloy) used for the purpose described in **1**.

magnetic shift register A shift register using magnetic flip-flops.

magnetic shunt A device that allows the useful magnetic flux of an instrument's magnet to be controlled. The device consists of a piece of magnetic material near the magnet in an electrical measuring instrument.

magnetic south See MAGNETIC SOUTH POLE.

magnetic speaker **1.** A loudspeaker that is essentially an enlarged earphone with a horn that conveys and intensifies the sound from the vibrating diaphragm. **2.** A loudspeaker in which the vibration of a diaphragm or reed in the field of a permanent magnet is conveyed by a pin to a paper or composition cone. Compare DYNAMIC SPEAKER.

magnetic storage **1.** A data bank or memory that stores information in the form of magnetic fields. **2.** The data on a magnetic tape or disk.

magnetic storm A disturbance in the earth's magnetic field that typically follows a solar flare. Often causes interference to radio communications at low, medium and high frequencies.

magnetic strip A strip of powdered iron or ferrite on the back of an identification card, bank cash card, or credit card, that carries a code to identify the account number and verify that the secret entry code (if any) is correct, or that the credit is good.

magnetic susceptibility See SUSCEPTIBILITY.

magnetic switch **1.** In security systems, a switch kept open by the presence of a magnet attached to a door, window, or other movable object. When the object is moved, the magnet moves away from the switch, closing the switch and actuating an alarm. **2.** A REED SWITCH operated by a magnetic field.

magnetic tape Plastic tape coated with a film of magnetic material; it can be magnetized along its length to record sounds, video signals, and computer information.

magnetic-tape core A strip of magnetic metal wound spirally to create a toroid (donut) shape. Such construction is sometimes used in choke or transformer cores. Also see TOROID.

magnetic tape deck See TAPE DECK.

magnetic tape drive See TAPE TRANSPORT.

magnetic tape head See MAGNETIC PICKUP HEAD and MAGNETIC RECORDING HEAD.

magnetic tape library In a computer installation, the place where magnetic tape files are kept, or magnetic tape files and the records needed to utilize them.

magnetic tape parity As a safeguard against losing information bits during the transfer of information between magnetic tape and a memory device, a technique in which an extra bit is generated and added to characters under certain conditions, to make the output uniform temporarily. Lack of uniformity in output then serves as an error indicator. The original quantity is recovered by dropping the extra bit following a parity check.

magnetic tape reader A tape deck for playing back data on magnetic tape.

magnetic tape recorder A recorder-reproducer using magnetic tape.

magnetic test coil See SEARCH COIL.

magnetic thick film A film of magnetic material at least 10^{-6} meter in thickness, deposited on a substrate. Compare MAGNETIC THIN FILM.

magnetic thin film A film of magnetic material, less than 10^{-6} meter in thickness, deposited on a substrate. Compare MAGNETIC THICK FILM.

magnetic transducer A transducer that uses a coil, magnet, or both, to convert displacement into variable magnetic fields or electric currents. Common varieties are the inductance type, transformer type, and generator type. Compare CAPACITIVE TRANSDUCER, CRYSTAL TRANSDUCER, and INDUCTIVE TRANSDUCER.

magnetic tuning In a microwave oscillator, a means of tuning in which a ferrite rod in the cavity resonator is made to have adjustable magnetization so that the resonant frequency of the cavity varies and the frequency of the oscillator is thus adjustable. It is also used at ultra-high frequencies (UHF) and occasionally at very-high frequencies (VHF).

magnetic-vane meter See IRON-VANE METER.

magnetic vector In an electromagnetic field, the vector representing the magnetic component. It is perpendicular to the electric vector.

magnetic viscosity A property of certain materials, described in terms of the time required to magnetize a given substance to a specified level.

magnetic whirl One of the circular magnetic lines of flux around a straight conductor that carries electric current.

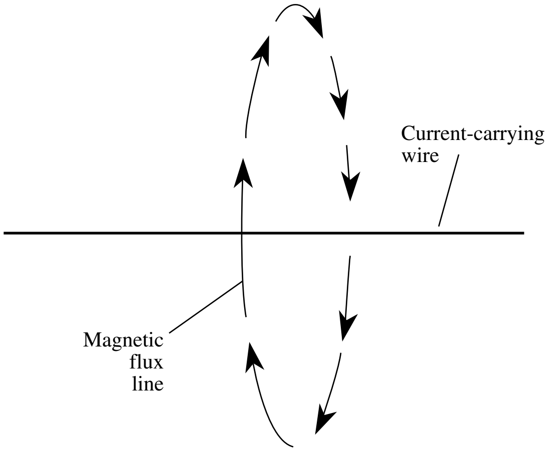

magnetic whirl

magnetic wire The thin wire used in wire recording and playback. See WIRE RECORDER.

magnetism The property of having or causing a magnetic field. It occurs when magnetic dipoles are aligned and when electric charge carriers are in motion.

magnetite A natural magnetic oxide of iron. Also called LODESTONE.

magnetization curve A curve depicting the magnetization of a material versus the applied magnetizing force. See, for example, HYSTERESIS CURVE.

magnetizer A device for magnetizing magnetic materials, as in the making of permanent magnets. Also see MAGNET CHARGER. Compare DEMAGNETIZER.

magnetizing current **1.** A current that sets up a magnetic field of useful intensity. **2.** The half-cycle of an alternating current or the polarity of a direct current flowing through a coil wound on a permanent magnet (as in a headphone, permanent-magnet loudspeaker, or polarized relay) that increases magnetic field strength. Compare DEMAGNETIZING CURRENT. **3.** The field current of a dynamo.

magnetizing force **1.** Magnetomotive force (in gilberts) divided by spatial distance (in meters). **2.** The intensity of a magnetic field that causes a material to become magnetized.

magnet keeper See KEEPER.

magnet meter See MAGNET TESTER.

magnet motor See PERMANENT-MAGNET MOTOR.

magneto See PERMANENT-MAGNET GENERATOR.

magnetocardiogram Abbreviation, MCG. A record, made by a MAGNETOCARDIOGRAPH, of the pulsating magnetic field of the heart. It is used as a diagnostic aid.

magnetocardiograph An instrument that produces a record of the pulsating magnetic field generated around the torso by natural ion currents in the heart.

magnetoelectric generator See MAGNETOGENERATOR.

magnetofluid mechanics See MAGNETOHYDRODYNAMICS.

magnetofluidynamics See MAGNETOHYDRODYNAMICS.

magnetogasdynamics See MAGNETOHYDRODYNAMICS.

magnetoionic duct A propagation path for radio waves between two points that have the same geomagnetic longitude on the surface of the earth. The radio waves tend to travel with the geomagnetic lines of flux at some frequencies under certain conditions.

magnetoionics The study of the effects of the geomagnetic field on the propagation of radio waves.

magnetogenerator See PERMANENT-MAGNET GENERATOR.

magnetograph An instrument for automatically recording a magnetic field.

magnetohydrodynamic generator A device using magnetohydrodynamic principles to generate electric power directly from gases. In the generator, a hot gas is passed through an intense magnetic field; a pair of collector plates picks up electrons from the ionized gas.

magnetohydrodynamic gyroscope A gyroscope whose spin is obtained by a rotating magnetic field circulating a conducting fluid, such as mercury, around a closed loop. Also see MAGNETOHYDRODYNAMICS.

magnetohydrodynamic power generator See MAGNETOHYDRODYNAMIC GENERATOR.

magnetohydrodynamics Abbreviation, MHD. The theory and application of phenomena produced by electrically conductive fluids and gases in electric and magnetic fields.

magnetometer An instrument for measuring the strength and direction of magnetic fields.

magnetomotive force Abbreviation, mmf. Unit, ampere. The phenomenon that is sometimes descriptively called *magnetic pressure.* It is analogous to electromotive force (and to water pressure) and is the agent that produces a magnetic field.

magneton See BOHR MAGNETON.

magneto-optical rotation The tendency of a magnetic field to rotate the plane of polarization of light passing through a substance. Also see KERR MAGNETO-OPTICAL EFFECT.

magneto-optical technology A computer data-storage technology that uses lasers to guide the read/write head in a magnetic disk drive. This greatly increases the amount of data that can be effectively stored on, and retrieved from, a magnetic disk.

magneto-optical valve See KERR MAGNETO-OPTICAL EFFECT.

magnetopause The high-altitude limit of the MAGNETOSPHERE.

magnetoplasmadynamics See MAGNETOHYDRODYNAMICS.

magnetoresistance The phenomenon whereby the resistance of a material, such as a semiconductor, changes when it is exposed to a magnetic field. Also see MAGNETORESISTOR.

magnetoresistor A material (such as bismuth wire, indium antimonide, or indium arsenide) whose resistance varies with the strength of a magnetic field in which it is placed.

magnetosphere In the upper atmosphere, a region extending thousands of kilometers from the earth, in which charged particles are trapped by the earth's magnetic field.

magnetostatic field A stationary magnetic field, such as that produced by a permanent magnet.

magnetostriction The expansion or contraction of a bar or rod of magnetic material (such as Invar, Monel metal, Nichrome, nickel, or Stoic metal) in proportion to the strength of an applied magnetic field. Magnetostrictive vibration in such a rod is

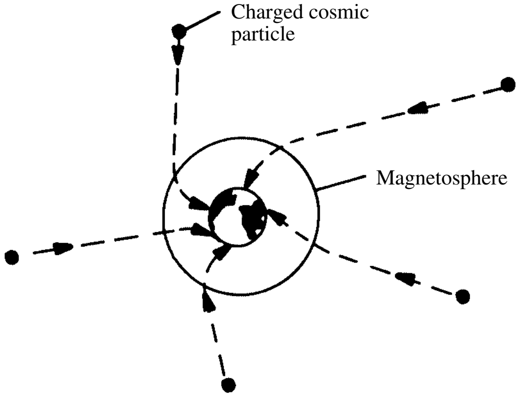

magnetosphere

comparable to piezoelectric vibration in a quartz crystal.

magnetostriction filter See ULTRASONIC FILTER, **1**.

magnetostriction oscillator An oscillator whose frequency is controlled by a magnetostrictive rod (see MAGNETOSTRICTION). The dimensions of the rod and the type of metal it contains determine its vibration frequency and, accordingly, the operating frequency of the oscillator.

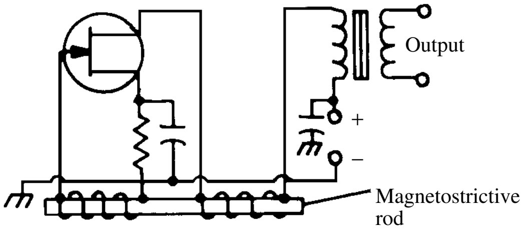

magnetostriction oscillator

magnetostrictive delay line A delay line in which the signal is propagated through a magnetostrictive rod. Also see MAGNETOSTRICTION.

magnetostrictive microphone A microphone in which sound vibrations produce changes in a magnetostrictive element, which, in turn, are converted into output-voltage changes. Also see MAGNETOSTRICTION.

magnetostrictive transducer A transducer in which some phenomenon, such as vibration or pressure, produces changes in a magnetostriction element, which in turn are converted into output-voltage changes. Also see MAGNETOSTRICTION.

magnet protector See KEEPER.

magnetrode The trademark of a radio-frequency device for externally producing hyperthermia (elevated temperature) inside the body, as in heating a cancerous tumor for therapeutic purposes.

magnetron A microwave vacuum tube consisting of a diode (with a cylindrical anode) through which the field of a powerful external permanent magnet passes. The magnetic field causes electrons leaving the cathode to travel in spiral paths between the electrodes. This action gives the tube a negative-resistance characteristic, resulting in oscillation when the tube is connected in an appropriate circuit. Some magnetrons have a built-in resonant cavity.

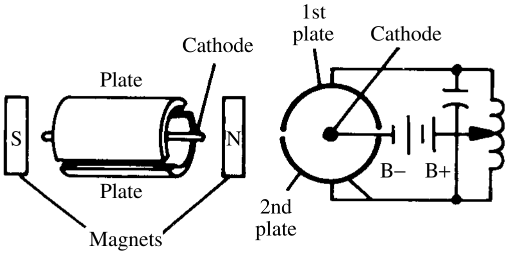

magnetron

magnet steel A high-retentivity alloy of chromium, cobalt, manganese, steel, and tungsten, used in the manufacture of permanent magnets.

magnet tester An instrument used to measure the flux of a magnet. Also see FLUXMETER.

magnet wire Insulated wire (usually solid copper) of 14 to 40 gauge, so called because of its original major use in winding the coils of electromagnets.

magnitude **1.** General expression for degree, size, or extent. **2.** Signal strength (amplitude). **3.** For a number or vector quantity, the absolute value or length. **4.** A measure of the relative or absolute brightness of celestial objects.

mAh Abbreviation of *milliampere-hour.*

main British expression for the alternating-current (ac) utility power available in a house or building.

main bang **1.** In a radar display, the pip or pulse resulting from the actual transmitted signal. This pulse is blanked out. **2.** In a spectrum analyzer, the pip corresponding to a frequency of zero, and caused by the local oscillator.

mainframe **1.** The chassis containing the central processor and arithmetic and logic circuits for a large computer. **2.** The general term for a large, powerful computer.

main lobe Also called *major lobe.* In a directional antenna system, the portion of the directivity pattern representing the greatest transmitted signal gain and/or the greatest received signal response. Also see MINOR LOBE, SIDE LOBE.

main memory The principal (immediate process) memory unit in a digital computer or data-processing system.

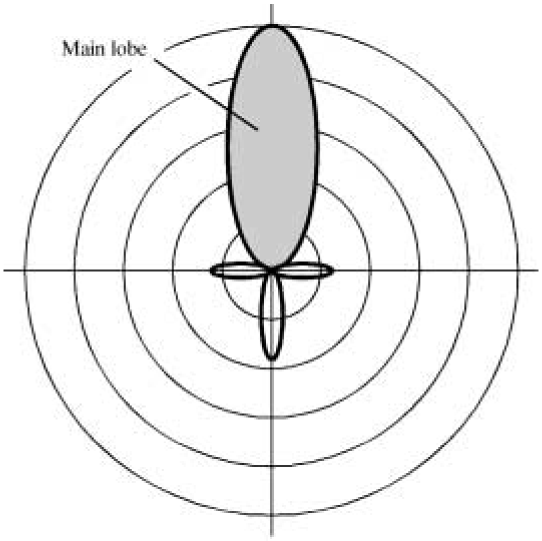

main lobe

main path In a computer program, the sequence of instruction execution disregarding the execution of subroutines.

main program The part of a computer program other than a subroutine.

main routine See MAIN PROGRAM.

mains **1.** In a power-distribution center, the lines that supply the entire system. An example is the set of lines leading into a house. **2.** The utility wires and associated outlets in a house or building.

maintenance The process of keeping a system, circuit, or component in operating condition, with minimal down time.

maintenance routine A computer program used by computer service personnel for diagnosis during a regular service interval.

major beats The principal beats produced in a beat-note system; they are usually the sum and/or difference of two fundamental frequencies. Compare MINOR BEATS.

major face In a hexagonal quartz crystal, one of the three larger faces. Compare MINOR FACE.

majority carrier The predominant charge carrier in processed semiconductor material. Electrons are the majority carriers in n-type material; holes are the majority carriers in p-type material. Compare MINORITY CARRIER.

majority logic A logic gate in which the output is high whenever the majority of its inputs is high, regardless of which inputs are high. Thus, in a five-input gate of this type, the output is high when any three or more of the inputs are high.

major lobe See MAIN LOBE.

major loop The principal path for the circulation of information or control signals in an electronic system (e.g., *major feedback loop*). Compare MINOR LOOP.

make **1.** The closing of a pair or set of contacts. **2.** To close a pair or set of contacts.

make-before-break contacts A pair of contacts in which the movable arm closes with the next contact before breaking with the previous one. Compare BREAK-BEFORE-MAKE CONTACTS.

make time The time required for a relay to latch completely, or for a switch (either mechanical or electronic) to close completely. Compare BREAK TIME.

male plug A plug having one or more protruding contacts in the form of pins, blades, or prongs. Compare FEMALE PLUG and HERMAPHRODITIC PLUG.

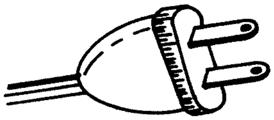

male plug

Malter effect The tendency for a layer of semiconductor having a high secondary emission ratio to become positively charged when bombarded by electrons. This occurs when a thin insulator separates the semiconductor from a metal plate. The insulator must be very thin (on the order of 10^{-7} meters). This results in a potential difference of up to about 100 volts.

manganese Symbol, Mn. A metallic element. Atomic number, 25. Atomic weight, 54.938.

manganese-dioxide depolarizer In a dry cell, manganese dioxide mixed with powdered carbon, the mixture being a depolarizing agent. Also see DEPOLARIZER.

manganin A low-temperature-coefficient alloy used in making wire for precision resistors. A typical composition is: copper (84 percent), manganese (12 percent), and nickel (4 percent).

manipulator A robot arm and end effector, as used in mechanical processes.

man-made interference See HUMAN-MADE INTERFERENCE.

man-made static See HUMAN-MADE INTERFERENCE.

manometer An instrument for measuring gas or vapor pressure—especially at low levels.

manpack A portable radio transceiver that can be used while walking.

mantissa **1.** The portion of a logarithm to the right of the decimal point. Thus, in 3.952502 ($\log_{10}$ 8964), the mantissa is 0.952502. **2.** The fixed point part of a number in scientific notation; thus, in 4×10^3, the mantissa is 4.

manual **1.** Actuated or operated directly by mechanical means, rather than automatically. **2.** A book, or a set of online information files, detailing the operation and maintenance procedures for a device or system.

manual input Use of a keyboard, mouse, trackball, or other electromechanical input device to enter data into a computer program or system.

manual operation In data processing, an operation in which automatic machines are not involved.

manual telegraphy Telegraphy that consists signals transmitted by a hand-operated key and recorded by hand (pen, pencil, or typewriter).

manual tuning Tuning performed entirely by adjusting variable circuit components by hand.

manual word generator A device by which an operator can originate information words for input into computer memory.

manufacturing automation protocol In a factory using computer-controlled robots, the set of standards for data communication between the robots and the controller and/or between individual robots. It keeps the factory operating smoothly.

MAR Abbreviation of MEMORY-ADDRESS REGISTER.

Marconi antenna A quarter-wave radio transmitting or receiving antenna operated against an earth ground.

Marconi effect The undesired tendency of an entire receiving antenna system, including lead-in or feeders, to act as a MARCONI ANTENNA.

margin **1.** A gap or space between two objects, such as adjacent plates of a capacitor. **2.** Clearance. **3.** The maximum error that can be tolerated without risk of improper or abnormal operation. **4.** In a teletypewriter, the range of adjustments in which the error frequency is acceptable.

marginal relay A relay having a small difference between its on and off currents or voltages.

marginal test As performed on equipment in a computer installation, a test to either determine the cause of an intermittent malfunction, or verify an equipment's operating tolerances.

marine broadcast station A coastal station that broadcasts information of interest to shipping: time, weather, ocean currents, etc.

marine radio Radio communications between seagoing vessels or between vessels and shore stations.

marine radiobeacon station A land-based radionavigation station whose transmitted signals are used for taking bearings.

mariner's compass See MAGNETIC COMPASS.

mark **1.** In telegraphy, the dot or dash portion of a character, as opposed to the dead space between such portions. **2.** The intelligence part of a similar signal (such as sound, light, etc.). **3.** The high (logic 1) state represented by a binary bit, as opposed to the low (logic 0) state. **4.** A character identifying the end of a data set. Also called MARKER (see MARKER, **2**).

marker **1.** A pip that indicates a particular frequency on a response curve displayed on an oscilloscope screen. **2.** A character that identifies the end of a data set. Also called MARK (see MARK, **4**).

marker beacons Individual coded-signal transmitters placed along a radio range and indicating features of the course marked by them.

marker frequency **1.** A known frequency that can be used to identify a spot-frequency harmonic of a frequency-standard signal. **2.** A known accurate signal used to identify the limit of a radio band. **3.** The frequency at some point on a response curve as identified by a marker pip (see MARKER).

marker generator An oscillator that supplies a marker pip (see MARKER).

mark hold In telegraphy, an unmodulated signal meaning information is not being sent.

mark reading The reading by an optical scanning device of marks made in specific areas of a document; the process also includes the marks' conversion to digital signals for input to a computer.

mark scanning See MARK READING.

mark sensing A process similar to MARK READING, except that the marks are sensed electrically.

marker trap A wave trap that supplies a dip-type marker pip when used in conjunction with a radio-frequency test oscillator (see MARKER).

market scanner Also called *bar-code reader.* A device that scans a black-bar binary label printed on a carton or other package (or magazine), and indicates the price of the merchandise on the readout of the checkout register.

mark-to-space ratio In radiotelegraphy, the ratio of the duration (mark) of a dot to the interval (space) between successive dots.

Marx generator An impulse-type high-voltage direct-current generator circuit in which several capacitors are charged in parallel through a high-resistance network. When the capacitor voltage reaches a critical high value, discharge occurs in series through spark gaps, producing a high-voltage pulse for each discharge.

maser A low-noise microwave amplifying device in which a microwave input signal causes high-energy-state molecules of ammonia or ruby to fall to the low-energy state and, as a result, to emit large amounts of energy as an output signal. The name is an acronym for *microwave amplification by stimulated emission of radiation.*

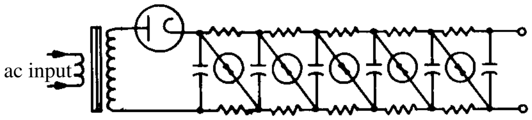

Marx generator

mask **1.** A kind of stencil through which plating, electrodeposition, or diffusion can be done. **2.** The viewing screen, or GRATICULE, of an oscilloscope. **3.** To obliterate a signal with a stronger one. **4.** A bit or character pattern used to change or extract bit positions in another pattern.

masking **1.** The use of a MASK of any type. **2.** The tendency of one effect or phenomenon to obscure another. It applies especially in audio systems, where certain sounds impair the ability of a listener to hear other sounds that occur at the same time. **3.** The extent to which one effect or phenomenon obscures another.

Masonite Masonite Corporation's tough fiberboard used for panels and bases of some electronic equipment.

mass The quantity of matter in a body. Like weight, mass is expressed in kilograms in the metric (SI) system and in pounds in the English system. For a given piece of material, mass can be determined by dividing the weight by the acceleration of gravity.

mass data Data in excess of the maximum amount that can be stored in the main (internal) storage unit of a digital computer (i.e., that which can only be accommodated by external media such as magnetic disks or tapes).

mass-energy equation Energy (E) is the product of a given mass (m) and the square of the speed of light (c^2): $E = mc^2$. It is also called the *Einstein equation.*

mass number **1.** Symbol, A. A number representing the total of neutrons and protons in the nucleus of an atom. The approximate mass of an atom is equal to $A \times m_p$, where m_p is the total proton (rest) mass. **2.** The number indicating the sum of nuclear protons and neutrons in an atom. It is usually written following the symbol for the atom: thus, U238 is uranium having 238 nucleons. An isotope of an element will have a different mass number than that of the normal atom.

mass of electron at rest Symbol, m_e. The amount of matter in an electron; $m_e = 9.1093897 \times 10^{-31}$ kg.

mass of neutron at rest Symbol, m_n. The mass of a neutron in the nucleus of an atom; $m_n = 1.6749286 \times 10^{-27}$ kg.

mass of proton at rest Symbol, m_p. The mass of a proton in the nucleus of an atom; $m_p = 1.6726231 \times 10^{-27}$ kg.

mass resistivity **1.** The resistance of a wire one meter long having a mass of one gram. It varies, depending on the composition of the wire. **2.** The resistance of a wire one mile long having a weight of one pound. It varies, depending on the composition of the wire.

mass spectograph An instrument used to analyze chemical compounds and mixtures in terms of their distinctive mass spectra, exhibited by ionized samples of the materials in a magnetic field.

mass spectrometer Abbreviation, MS. An instrument that permits rapid analysis of chemical compounds via the MASS SPECTRUM.

mass spectrum An electron spectrum that can be used to identify a chemical element. Different elements have nuclei with different charge-to-mass ratios. This results in each element having a unique mass spectrum.

mass storage In a computer system, a magnetic or optical storage medium capable of holding large amounts of data. Examples: magnetic diskette, magneto-optical diskette, external hard disk, compact-disk read-only memory (CD-ROM), and magnetic tape.

mass unit See ATOMIC MASS UNIT.

master **1.** The primary or main element or device in a system. **2.** A primary data medium or recording from which copies are made. **3.** A primary reference standard. See the following several definitions.

master clock **1.** In a digital computer, the primary generator of timing pulses. **2.** A standard time clock that drives other (slave) clocks, or to which clocks of lesser accuracy can be referred.

master console In a computer system, an equipment with panel instruments and controls, which permits operations to be governed, monitored, and controlled by a human operator.

master control **1.** The main control circuit in a system. **2.** A point from which signals or programs are distributed in a communications or broadcast system.

master data Also called *archives.* In a computer record, data elements that remain unaltered for a long time, and from which copies are made.

master file A computer file of data used routinely and remaining unchanged for a long time.

master gain control The principal gain control in an audio amplifier or mixer [i.e., the one used to adjust the gain (volume) of the entire system].

master instruction tape Magnetic tape on which various related computer programs are recorded.

master library tape See MASTER PROGRAM FILE.

master oscillator Abbreviation, MO. The main oscillator in an electronic system (e.g., the oscillator stage in an oscillator-amplifier type of radio transmitter). This oscillator can be either self-excited or crystal-controlled.

master oscillator-power amplifier Abbreviation, MOPA. A type of transmitter or signal generator in which a frequency-determining oscillator drives a power amplifier, which in turn delivers an output signal. Because the oscillator is isolated from the output load, this arrangement has greater stability than one in which the oscillator alone supplies power to the load.

master pattern The etching pattern used for manufacture of a batch of identical printed-circuit boards.

master program file A reel of magnetic tape on which is recorded the programs regularly used in a data-processing installation. It is also called *master library tape.*

master record In a data-processing system, the current record (usually stored on a disk or tape) that will be used for the next computer run.

master relay A relay that operates other (slave) relays. Compare SLAVE RELAY.

master station See KEY STATION.

master switch A switch that can actuate or deactuate an entire installation or system.

master tape **1.** In sound recording and reproduction, a magnetic tape that contains material from which other tapes and discs can be made. **2.** In automation, a magnetic tape on which is recorded the basic signal sequence for controlling a process and other recorders. **3.** In data processing, a magnetic tape that must not be erased.

master volume control See MASTER GAIN CONTROL.

masurium See TECHNETIUM.

MAT Abbreviation of MICROALLOY TRANSISTOR.

match **1.** To mate devices, signals, impedances, etc. for optimum compatibility in terms of signal transfer, equipment interfacing, and other optimizing qualities. **2.** The condition of being compatibly mated, physically or electrically.

matched components Circuit components (capacitors, coils, diodes, resistors, transistors, etc.) that are carefully selected for similar or particularly compatible operating characteristics.

matched filter **1.** A filter with input and output impedances matched to the input line and output load, respectively. **2.** A filter designed for separating a signal with a particular waveform from other signals and noise.

matched impedance A usually non-reactive impedance that has the same value as that of another impedance with which it is operated. Maximum power is transferred between impedances that are matched.

matched load A purely resistive load, the impedance of which is the same as the characteristic impedance of the feed line. This results in optimum power transfer from the line to the load.

matched pair A pair of matched components offered in a single package.

matched transmission line A transmission line terminated in a purely resistive impedance whose value is identical to the characteristic impedance of the line. Such a line transfers all of its energy to its load without reflection; no standing waves are on the line.

matching pad An inductance-capacitance (LC) network for matching the impedance of a load to the output impedance of a signal generator.

matching stub See STUB.

matching transformer An audio-frequency (AF) or radio-frequency (RF) transformer used to match one purely resistive impedance to another.

matchtone A transistorized, single-frequency audio oscillator that can be used to monitor transmitted radiotelegraph signals. The carrier wave from the transmitter is rectified by a small semiconductor diode, whose direct-current output powers the oscillator.

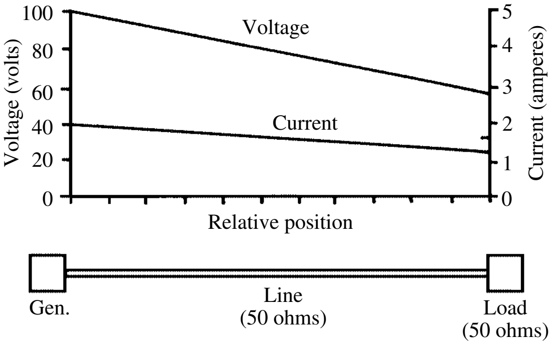

matched transmission line

Mateucci effect The generation of a potential difference in a helically wound, ferromagnetic wire when its magnetization fluctuates.

mathematical check A test of the validity of the result of an arithmetic process (by using alternate methods, for example).

mathematical logic **1.** A branch of mathematics that involves the theoretical behavior of various systems of reasoning. **2.** See BOOLEAN ALGEBRA. **3.** See DIGITAL LOGIC.

mathematical model See MODEL, **2**.

mathematical subroutine Within a computer program, a subroutine serving as an arithmetic function (i.e., one for performing an operation not integral to the monitor program).

matrix **1.** A high-speed switching or memory array used in counters and computers. **2.** Generally, any two-dimensional array of objects. **3.** A device for solving linear simultaneous equations, consisting of a rectangular array of coefficients.

matrix printer See WIRE PRINTER.

mat switch A form of PRESSURE SENSOR used in some security systems. When weight appears on the mat, switches close, actuating an alarm.

matter The building material of the universe that occupies space and has mass that can be measured. See, for illustration, ATOMIC THEORY and STATES OF MATTER.

matter waves See DE BROGLIE WAVES.

max Abbreviation of MAXIMUM.

maxima Points along a curve at which a function reaches a local maximum value. Also see MAXIMA AND MINIMA.

maxima and minima **1.** The loops and nodes of current or voltage on an antenna or transmission line. **2.** The lobes and nulls in a directivity pattern. **3.** The bright and dark bands in a visible-light interference pattern. **4.** In radar reflections, regions of localized maximum and minimum

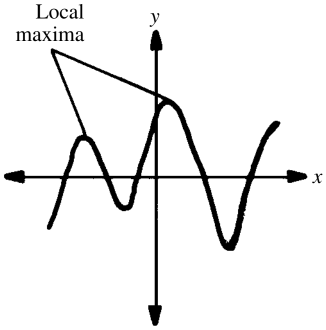

maxima

intensity. **5.** The study and solution of maximum, minimum, and inflection points on the curve of a function.

maximal flatness For an amplifier or network, the condition in which peaks are not present in the normal passband response.

maximum Abbreviation, max. The highest value in a range or set. Also see MAXIMA AND MINIMA and PEAK.

maximum available gain Abbreviation, MAG. The amplification provided by a circuit or device whose input and output impedances are correctly matched to source and load.

maximum current **1.** Symbol, I_m or I_{max}. The highest value reached by an alternating-current half-cycle or by a pulse current. Also called PEAK CURRENT. **2.** The highest value of current in a series of current values.

maximum power **1.** Symbol, P_m or P_{max}. The highest value of power that an equipment can be called upon to supply. **2.** The highest value of power in a series of measurements or calculations.

maximum-power discharge current For a cell or battery, the current at which the greatest amount of power is delivered.

maximum power output See MAXIMUM POWER, **1**.

maximum power transfer The condition in which the largest amount of power is delivered by a source to a load.

maximum power transfer theorem Maximum power is transferred from a generator to a load when the impedance of the load equals the internal impedance of the generator. Compare COMPENSATION THEOREM, NORTON'S THEOREM, RECIPROCITY THEOREM, SUPERPOSITION THEOREM, and THEVENIN'S THEOREM.

maximum rating **1.** The highest value of a quantity (e.g., current, voltage, or power) that can safely be used with a given device. **2.** The highest value of a quantity afforded by a given device (e.g., *maximum capacitance* of a variable capacitor).

maximum record level **1.** In a magnetic tape, magnetic disk, or phonograph disc, the highest amplitude of input signal that can be recorded with an acceptable amount of distortion. **2.** The recording-head current or power that results in third-harmonic distortion of three percent.

maximum signal level **1.** In an amplitude-modulated signal, the peak power. **2.** In an amplitude-modulated facsimile or television system, the amplitude that results in a black or white picture (depending on whether the highest amplitude produces black or white).

maximum undistorted power output Abbreviation, MUPO. The highest power that an active amplifying device will deliver before significant distortion occurs.

maximum usable frequency Abbreviation, MUF. The highest frequency that can be used successfully at a given time, between two specific geographic locations, for communication via the ionosphere.

maximum voltage **1.** Abbreviation, E_m, E_{max}, V_m, or V_{max}. The peak value reached by an alternating-current voltage half-cycle, or by a voltage pulse. **2.** The highest value of voltage in a series of voltage measurements or calculations.

maximum wattage See MAXIMUM POWER.

maxterm form In mathematical calculations, the factored form of a function, expressed as a product of sums. For example, the maxterm form of $f(x) = x^2 + 5x + 6$ is $f(x) = (x + 2)(x + 3)$.

maxwell Symbol, Mx. The cgs unit of magnetic flux, equivalent to one line of flux or 10^{-8} weber.

Maxwell bridge A four-arm alternating-current bridge for measuring inductance against a standard capacitance.

Maxwell's equations A set of four equations developed by James Clerk Maxwell in 1864 and 1873, describing vector quantities pertaining to points in space subjected to varying electric and magnetic forces. Through his classic presentation, Maxwell

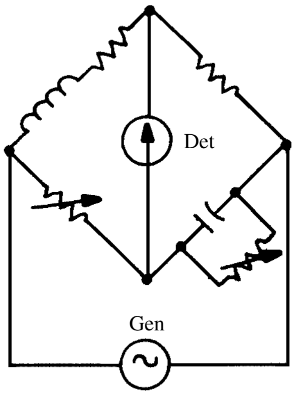

Maxwell bridge

predicted the existence of electromagnetic waves, whose later discovery made radio possible.

Maxwell's law Also called *Maxwell's rule.* Every part of an electric circuit is acted upon by a force tending to move it in the direction that results in the maximum magnetic flux being enclosed.

maxwell-turn A unit of magnetic coupling (linkage) equal to 1 maxwell per turn of wire in a coil linked by magnetic flux. Also see MAXWELL.

mayday In radiotelephony, a word spoken as an international distress signal equivalent to SOS in radiotelegraphy. The word is the phonetic equivalent of the French *m'aidez* (help me).

MB Abbreviation of MIDBAND.

Mb Abbreviation of MEGABAR.

MBB Abbreviation of MAKE BEFORE BREAK.

MBM Abbreviation of *magnetic bubble memory.*

MBO Abbreviation of MONOSTABLE BLOCKING OSCILLATOR.

MBS Abbreviation of *magnetron beam switching.*

mc **1.** Symbol for MILLICURIE (mCi is preferred). **2.** Symbol for METER-CANDLE.

Mc **1.** Symbol for MEGACURIE (MCi is preferred). **2.** Obsolete abbreviation of *megacycle(s)*, a term superseded by MEGAHERTZ.

MCG Abbreviation of MAGNETOCARDIOGRAM.

McLeod gauge An instrument for measuring gas under low pressure. A measured volume of the gas under test is first compressed (to a lower known volume) to a pressure more easily measured via a mercury manometer and the application of Boyle's law.

MCM Abbreviation of MONTE CARLO METHOD.

McProud test A simple test for checking the tracking efficiency of a phonograph pickup and arm for microgroove discs.

Mc/s Obsolete abbreviation of *megacycle(s) per second*, a term superseded by MEGAHERTZ.

MCS **1.** Abbreviation of *Master of Computer Science.* **2.** Abbreviation of *missile control system.*

MCW Abbreviation of MODULATED CONTINUOUS WAVE.

Md Symbol for MENDELEVIUM.

MDAS Abbreviation of *medical data acquisition system.*

m-derived filter A filter whose inductance (L) and capacitance (C) values are derived by multiplying those of a constant-k filter by a factor m between zero and 1. This factor is a function of the ratio f_i/f_c, where f_i is the frequency of infinite attenuation, and f_c is the cutoff frequency. This type of filter exhibits sharper response than the equivalent constant-k filter.

MDI Abbreviation of MAGNETIC DIRECTION INDICATOR.

M-display See M-SCAN.

MDS Abbreviation of MINIMUM DISCERNIBLE SIGNAL.

m_e Symbol for MASS OF ELECTRON AT REST.

Meacham oscillator A highly stable radio-frequency oscillator consisting of an amplifier provided with a feedback circuit containing a four-arm bridge, one arm of which is a quartz crystal, and another, a tungsten-filament lamp acting as a nonlinear resistor. Also called *bridge-stabilized oscillator.*

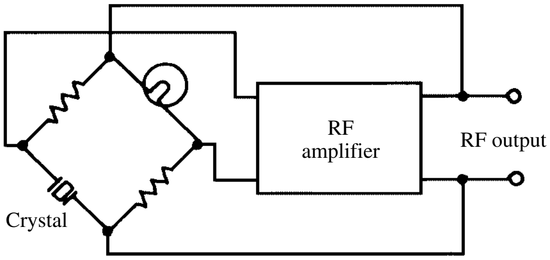

Meacham oscillator

mean **1.** A general term meaning *average.* **2.** See ARITHMETIC MEAN. **3.** See GEOMETRIC MEAN.

mean charge **1.** In an object that is nonuniformly charged, the average charge per unit distance, area, or volume. **2.** In a capacitor carrying a fluctuating current, the average amount of charge held by the plates.

mean free path **1.** In acoustics, the average distance that sound waves travel before striking a barrier or reflecting surface. **2.** The average distance that sound waves travel between reflections (echoes) in a chamber. **3.** In a gas tube, the average of all the free paths of electrons at a specified temperature.

mean life **1.** Symbol, L. The average life of a radioactive substance [i.e., the time taken for $1/e$ (e = base of natural logarithms) of the substance to disintegrate]. **2.** The time required for excess carriers injected into a semiconductor to recombine with carriers of opposite sign. Also called *average life.*

mean proportional See GEOMETRIC MEAN.

mean time before failure Abbreviation, MTBF. The average length of time that a component or system will perform before the first failure occurs. It is generally specified in hours.

mean time between failures Abbreviation, MTBF. The average length of time that a component or system will perform before failure occurs—either initially or after repair or replacement. It is generally specified in hours.

measured A quantity that is presented to an instrument for measurement.

measured service Any service in which charges are assessed per unit-time usage block. Online computer services are a common example. In some cases, other factors, such as distance, affect the cost per unit time; most long-distance telephone services fall into this category.

measurand Any quantity that is measured with an instrument.

measurement **1.** The process by which the magnitude, extent, or duration of a parameter is found. **2.** The value of a parameter, as obtained, according to **1**.

measurement error The difference between the measured value of a quantity and its true value. Also see NEGATIVE ERROR OF MEASUREMENT and POSITIVE ERROR OF MEASUREMENT.

measurement range In a measuring device, the range within which the error is smaller than a specified value.

mechanical analogs Familiar mechanical devices, systems, or effects with which certain electrical counterparts can be compared for ease in teaching or understanding (e.g., inductance compared with mass, capacitance with elasticity, voltage with pressure, and current with velocity).

mechanical axis In a quartz crystal, the axis perpendicular to the faces of the hexagon. Also see Y-AXIS, **2**.

mechanical bandspread Bandspread tuning obtained by reduction-ratio gearing of the tuning mechanism. Compare ELECTRICAL BANDSPREAD.

mechanical bias **1.** A steady pull applied by a spring to the armature of a relay to sensitize it by decreasing the distance that the armature must move to close the contacts. **2.** Bending of a relay frame to position the armature closer to the magnet for the purpose defined in **1**.

mechanical damping Damping action obtained entirely by mechanical devices (such as weights, dashpots, etc.).

mechanical equivalent of heat The amount of mechanical work required to produce a unit quantity of heat. For example, 4.183 joules can be converted into 1 calorie of heat.

mechanical equivalent of light The expression of luminous energy in equivalent power units. In practical measurements, this is taken as the total power output of a lamp minus the power absorbed by a transparent jacket used to remove the infrared and ultraviolet rays.

mechanical filter See ULTRASONIC FILTER, **1**.

mechanical joint A union of electrical conductors consisting exclusively of a junction or splice made without brazing, soldering, or welding.

mechanical load An electromechanical device that uses the output of an electrical source. Such devices include actuators, brakes, clutches, meters, motors, and relays.

mechanical rectifier A vibrator or commutator used to change an alternating current into a direct current by selecting and passing only positive or negative half-cycles. Also see ELECTROMECHANICAL RECTIFIER.

mechanical scanner **1.** A mechanical device for scanning an object or scene and breaking it into horizontal lines that are converted to signals. **2.** A device that scans the reproducer lamp in a mechanical television receiver. See, for example, NIPKOW DISK.

mechanical switch A switch actuated by moving or sliding a lever, pressing a button, or otherwise applying mechanical pressure.

mechanical time constant For a torque motor, the ratio of moment of inertia to damping factor. Compare ELECTRICAL TIME CONSTANT.

mechanical wave filter See ULTRASONIC FILTER, **1**.

mechanics The branch of physics concerned with forces and motion and the laws of gases and liquids. It is subdivided into *kinematics* and *kinetics*.

mechatronics Combination of the words *mechanics* and *electronics*, referring to the use of electromechanical devices (especially robots) in manufacturing. The term was originally coined in Japan.

median **1.** The middle value in a sequence of numbers. For example, in the series: 1, 2, 3, 4, 5, 6, 7, the median is 4. Compare ARITHMETIC MEAN and GEOMETRIC MEAN. **2.** In a statistical distribution, the value *s* in the domain so that the area under the curve for all values less than *s* is equal to the area under the curve for all values greater than *s*.

medical electronics See ELECTROMEDICAL ENGINEERING.

medical robot **1.** A robot used in a doctor's office, or in a hospital to assist doctors and nurses. There are various applications, some of which have provoked controversy (e.g., *robotic surgical assistant*). It generally performs simple, noncritical tasks. It has been suggested as a means of entertaining hospital patients—especially children. **2.** See BIOMECHANISM.

medium In a computer system, that storage device onto or into which data is recorded for input into memory (e.g., magnetic disk, magnetic tape, optical disk, etc.).

medium-frequency Abbreviation, MF. Pertaining to frequencies in the range 300 kHz to 3 MHz, representing wavelengths from 1000 meters to 100 meters.

medium of propagation The substance (or vacuum) through which electromagnetic energy is transmitted (e.g., outer space, the atmosphere, or a dielectric material).

medium-scale integration A method of manufacturing integrated circuits, in which there are at least 10, but less than 100, individual gates on each chip. Abbreviated MSI.

medium-scan television A television (TV) communications medium in which the scanning rate is slowed down compared to regular (fast-scan) TV, but is faster than the commonly used slow-scan TV. It provides some conception of motion, although not as realistic as fast-scan TV.

medium tension Medium voltage. A relative term, but generally referring to common alternating-current utility voltage (e.g., 117 volts or 234 volts).

medium wave Abbreviation, MW. Pertaining to wavelengths corresponding to medium frequencies (see MEDIUM-FREQUENCY) (i.e., those in the 100- to 1000-meter range).

meg **1.** Colloquialism for MEGOHM(s). **2.** Colloquialism for MEGABYTE(s).

mega- Abbreviation, M. **1.** A prefix meaning *million(s)*, (i.e., 10^6 or 1,000,000). **2.** In digital data applications, a prefix meaning 2^{20} or 1,048,576.

megabar Abbreviation, Mb. A cgs unit of high pressure. 1 Mb = 10^6 bars = 10^{11} pascals. Also see BAR, **1**.

megabit A unit of digital data equal to 2^{20} (1,048,576) bits. Also see BIT.

megabyte Abbreviation, M or MB. A unit of digital data equal to 2^{20} (1,048,576) bytes. Also see BYTE.

megacurie Abbreviation, MCi. A large unit of radioactivity equal to 3.71×10^6 disintegrations per second; 1 MCi = 10^6 curies. Also see CURIE.

megacycle See MEGAHERTZ.

megaelectronvolt Abbreviation, MeV. A large unit of electrical energy; 1 MeV = 10^6 eV. Also see ELECTRONVOLT.

megahertz Abbreviation, MHz. A unit of frequency; 1 MHz = 10^6 Hz = 1,000,000 Hz.

megampere Abbreviation, MA. A unit of high current; 1 MA = 10^6 A = 1,000,000 A.

megaphone **1.** A hand-held microphone/amplifier/loudspeaker used to amplify the voice of a person who must be heard over an appreciable area. **2.** A simple horn for amplifying the voice.

megarutherford Abbreviation, Mrd. A large unit of radioactivity equal to 1 trillion (10^{12}) disintegrations per second; 1 Mrd = 10^6 rd = 1,000,000 rd. Also see RUTHERFORD.

megavolt Abbreviation, MV. A unit of extremely high voltage; 1 MV = 10^6 V = 1,000,000 V.

megavolt-ampere Abbreviation, MVA. A unit of extremely high reactive power; 1 MVA = 10^6 VA = 1,000,000 VA. Also see VOLT-AMPERE.

megawatt Abbreviation, MW. A unit of high power; 1 MW = 10^6 W = 1,000,000 W. Also see WATT.

megawatt-hour Abbreviation, MWh. A large unit of electrical energy or of work; 1 MWh = 10^6 Wh = 1,000,000 Wh = 3.6×10^9 joules. Also see WATT-HOUR.

megger An instrument containing an internal high-voltage direct-current power supply, used for measuring high values of resistance. Compare MEGOHMMETER.

meg-mike **1.** Colloquialism for MEGOHM-MICROFARAD(s). **2.** Colloquialism for MEGOHM-FARAD(s).

megohm Symbol, M. A unit of high resistance, reactance, or impedance; 1 M = 10^6 ohms = 1,000,000 ohms.

megohm-farads For a large capacitor, the product of leakage resistance (megohms) and capacitance (farads). Also see MEGOHM-MICROFARADS.

megohmmeter A special ohmmeter for measuring resistances in the megohm range.

megohm-microfarads For a capacitor, the product of leakage resistance (megohms) and capacitance (microfarads). The figure is an expression for the relative insulation resistance of a capacitor.

Meissner circuit An oscillator tuned by means of LECHER WIRES (parallel-conductor resonant circuits). It is used primarily at ultra-high frequencies (UHF).

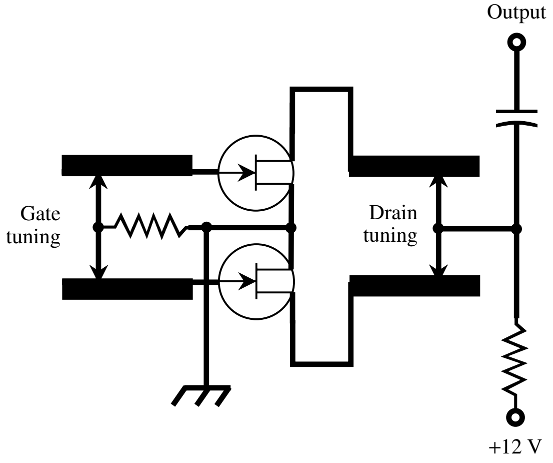

Meissner circuit

Meissner effect In a superconductive material, the abrupt loss of magnetism when the temperature of the material is reduced to a value below that required for superconductivity.

Meissner oscillator See MEISSNER CIRCUIT.

meitnerium Symbol, Mt. Also called *unnilenium* (Une). Atomic number, 109. The most common isotope has atomic weight 266. Classified as a transition metal. It is human-made and not known to occur in nature.

mel An expression of apparent or perceived sound pitch. A tone of 1 kHz, at a level of 40 dB, with respect to the threshold of hearing, represents 1 mel. The perceived pitch depends, to some extent, on the intensity of the sound, as well as on the actual frequency.

M electron In certain atoms, one of the electrons whose orbits are outside of and nearest to those of the L electrons.

meltback process The technique of remelting a doped semiconductor material and allowing it to refreeze to form a grown junction.

meltback transistor A grown-junction transistor produced by the MELTBACK PROCESS.

melting point Abbreviation, mp. The temperature at which a solid starts becoming liquid at a pressure of one atmosphere. Compare FREEZING POINT and MIXTURE MELTING POINT.

memory **1.** The section of a digital computer that records and holds data until it is necessary. In personal computers, the term generally refers to RANDOM-ACCESS MEMORY and READ-ONLY MEMORY, contained in integrated circuits (ICs). Compare STORAGE. **2.** See MEMORY DRAIN.

memory address register In computer storage, a register in which is stored the address of operands in other locations.

memory area A portion of computer memory reserved for a specific type of data. Also called *area.*

memory capacity As a function of the number of memory locations available, the number of bytes that can be stored. It is usually specified in kilobytes, megabytes, or gigabytes. Also see GIGABYTE, KILOBYTE, and MEGABYTE.

memory cycle **1.** The period of execution of a sequence of operations. **2.** The complete operational cycle for inputting data to memory or retrieving it.

memory dialing In a telephone set, a feature that allows rapid dialing of stored digits. The simplest version is the "redial" feature, in which the most recently dialed number is rapidly dialed at the touch of a button. Some sets can store several different numbers, usually including area codes, and sometimes country codes as well.

memory drain Also called *battery memory.* A phenomenon occasionally exhibited by nickel–cadmium cells and batteries, in which the useful ampere-hour capacity is reduced even though the unit is not physically damaged. The depth of the charge cycle decreases to a fraction of its rated value. The problem can usually be overcome by discharging the cell or battery fully, then recharging fully, and repeating the process several times. See also NICKEL–CADMIUM.

memory dump In computer operations, to either print out what is stored in some of or all of the memory locations or transfer the data from a bank of memory cells to some external storage medium.

memory effect See MEMORY DRAIN.

memory guard In a computer, hardware or software that keeps certain memory locations from being addressed by a program being run.

memory location In a computer memory, a place where an information unit (word or character) can be stored; the stored information can be retrieved by appropriate addressing instructions.

memory organization packets In artificial intelligence (AI) and expert systems, a method of arranging computer memory into general rules or statements. The statements are used by software to derive models, forecasts, diagnoses, etc.

memory power Computer memory efficiency in terms of data processing (cycle) speed.

memory protection A hardware device in a multiple programming computer that prevents programs from being altered by other operating programs in the installation.

memory register In a digital computer, a register used in all instruction and data transfers between the memory and other sections of the machine.

memory unit See MEMORY.

mendelevium Symbol, Md. A radioactive element produced artificially. Atomic number, 101. Atomic weight, 258 (approx.).

menu In computer operations, a list of commands for using various functions of the system.

MEP Abbreviation of *mean effective pressure.*

mercuric iodide Formula, HgI_2. A compound whose crystals are useful at room temperature as detectors in high-resolution gamma-ray spectroscopy.

mercury Symbol, Hg. A metallic element. Atomic number, 80. Atomic weight, 200.59. The only metal that is liquid at room temperature. It is used extensively in switches, certain high-voltage rectifiers, high-vacuum pumps, and thermometers.

mercury arc The arc discharge occurring in mercury vapor between solid or liquid (mercury) electrodes. The discharge emits ultraviolet radiation.

mercury-arc rectifier A heavy-duty rectifier tube utilizing ionized mercury vapor. The two general types are MERCURY-VAPOR RECTIFIER and MERCURY-POOL RECTIFIER.

mercury battery Also called *mercuric-oxide battery.* A set of two or more mercury cells stacked one atop the other, electrically connected in series. The resulting battery has a cylindrical shape. A set of four cells provides approximately 5.4 volts under no-load conditions; a battery of seven cells provides 9.5 volts; a battery of nine cells provides 12 volts. See MERCURY CELL.

mercury cadmium telluride Formula HgCdTe. An alloy used as a semiconductor in certain transistors, integrated circuits, and infrared detectors.

mercury cell Also called *mercuric-oxide cell.* An electrochemical cell having a button-like shape, small enough to fit inside a wristwatch. The unit is housed in a steel container and has a mercuric-oxide cathode, amalgamated-zinc anode, and potassium hydroxide and zinc-oxide electrolyte. The potential difference under no-load conditions is 1.35 volts, with a high ratio of stored energy per unit mass. The cell has a flat discharge curve; the voltage remains essentially constant until the charge is almost depleted, and then the voltage drops rapidly. There has been a decrease in the use of mercury cells and batteries in recent years, because mercury is toxic. Unless these cells are discarded in a special way, the mercury from them can cause dangerous contamination of soil and water.

mercury delay line A delay line in which delay is obtained by propagating the signal through a pipe of mercury.

mercury diffusion pump A vacuum diffusion pump using mercury vapor.

mercury displacement relay A form of switching relay in which the electrical contact is made by moving mercury.

mercury-jet switch A multipoint switch using a jet of mercury instead of the conventional wiper arm, for high-speed operation and reduced wear.

mercury memory A recirculating memory using a mercury delay line. Also see DELAY LINE and DELAY-LINE MEMORY.

mercury-pool cathode In certain industrial electron tubes, such as the ignitron, a cathode electrode consisting of a pool of mercury.

mercury-pool rectifier A type of mercury-arc rectifier whose cathode is a pool of mercury. In one type, the arc is initiated by tilting the tube momentarily to bring the mercury into contact with a third electrode, thus causing a starting current to flow through the pool. In another type, the *ignitron*, a starter electrode is in continual contact with the mercury.

mercury pump See MERCURY DIFFUSION PUMP.

mercury rectifier See MERCURY-POOL RECTIFIER and MERCURY-VAPOR RECTIFIER.

mercury relay A relay in which at least one of the contacts is mercury.

mercury storage See MERCURY MEMORY.

mercury switch A switch consisting essentially of two or more stiff wire electrodes and a drop of mercury hermetically sealed in a glass tube. Tilting the tube causes the mercury to flow toward one end, where it immerses the electrodes, providing a conductive path between them.

mercury-vapor lamp A glow lamp emitting blue-green light that causes ionization of mercury vapor by an electric current.

mercury-vapor rectifier A tube-type high-voltage diode rectifier containing a small amount of mercury that vaporizes and ionizes during tube operation.

mercury-vapor tube **1.** See MERCURY-VAPOR LAMP. **2.** See MERCURY-VAPOR RECTIFIER.

mercury-wetted reed relay A reed relay in which the reeds are wetted with mercury in a pool by capillary action. The film of mercury forms a tiny bridge when the reeds open; when this bridge separates, a clean, high-speed break occurs without contact bounce. Compare DRY-REED SWITCH.

merge **1.** In computer operations, to make a single set or file from two or more record sets. **2.** In word processing, to create a corrected master recording from two input media: the original master recording and the recording that contains the corrections.

meridian **1.** A great circle passing through earth's geographic poles and a given point on the surface of the earth. **2.** A line of longitude on a map or globe. Also see TIME ZONE, ZERO MERIDIAN, and ZONE TIME.

mesa A flat-topped, protruding region in a semiconductor wafer. The mesa is produced by etching the surrounding part of the material. Some bipolar transistors are manufactured in this way.

mesa diffusion A method of manufacturing bipolar transistors. The different semiconductor materials are first diffused together. Then part of the resulting wafer is etched away, resulting in a mesa shape.

mesa transistor A diffused planar transistor in which the silicon area around the base has been etched away to reduce collector-to-base capacitance; the base-emitter region remains elevated like a high plateau (mesa).

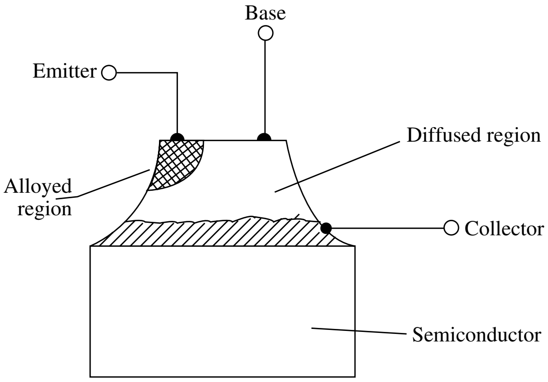

mesa transistor

MESFET A form of field-effect transistor combining depletion-mode and enhancement-mode properties. A Schottky barrier forms the gate electrode.

mesh **1.** A combination of the elements that form a closed path in a network. **2.** The closed figure (such as the delta or star) obtained by connecting polyphase windings together. **3.** A grid, screen, or similar structure in a vacuum tube.

mesh equations Equations describing fully the current and voltage relations in a network of meshes (see MESH, **1**).

Mesny circuit A push-pull ultra-high-frequency (UHF) oscillator whose gate or base tank is a pair of parallel wires short-circuited by a slider; the drain or collector tank is a similar pair of wires. The frequency is varied by moving the sliders along the wires.

mesochronous A condition for signals in which significant instants pass at identical average speeds, such as bits per second.

meson An unstable nuclear particle first observed in cosmic rays. A meson can be electrically positive, negative, or neutral. Its mass lies between that of the electron and proton.

mesotron See MESON.

message **1.** In communications, a body of information sent from a source (transmitter) to a destination (receiver). **2.** Data entered into a transaction-processing system.

message exchange In a digital communications channel, a hardware unit that carries out certain switching functions that would otherwise have to be done by a computer.

message switching system A data communications system having a central computer that receives messages from remote terminals, stores them, and transfers them to other terminals as needed.

metadyne See DC GENERATOR AMPLIFIER.

metal An elemental material that exhibits several familiar properties (such as luster, ductility, malleability, good electrical and heat conductivity, relatively high density, and the ability to emit electrons). Common examples are aluminum, copper, gold, lead, and silver. Compare METALLOID and NONMETAL.

metal-base transistor A bipolar transistor in which the base is a metal film, and the emitter and collector are films of n-type semiconductor material.

metal-ceramic construction The building of certain electronic components by bonding ceramic parts to metal parts. Also see CERMET.

metal-film resistor A fixed or variable resistor in which the resistance element is a film of a metal alloy deposited on a substrate such as a plastic or ceramic.

metal finder See METAL LOCATOR.

metallic binding forces In a crystal, the binding electrostatic force between cations and electrons. Also called *electron-gas binding forces.*

metallic bonding See BONDING, **1** and METALLIC BINDING FORCES.

metallic circuit A circuit, such as a two-wire telephone line, in which earth ground is not a part of the circuit. Compare GROUND-RETURN CIRCUIT.

metallic crystal A crystal substance in which positive ions and free electrons exist; it is, therefore, a good electrical conductor.

metallic insulator A short-circuited quarter-wave section of transmission line that acts as an insulator at the quarter-wavelength frequency.

metallicize To make a circuit fully metallic, as when two wires are used instead of one wire and a ground connection. (Not to be confused with METALLIZE.)

metallic rectifier A dry rectifier using a metal disk or plate coated with a material (such as selenium, an oxide, or a sulfide).

metallic tape Recording tape made from metal, rather than from plastic. Noted for its excellent audio-reproduction characteristics.

metallize To treat, coat, or plate with a metal. (Not to be confused with METALLICIZE.)

metallized capacitor A capacitor in which each face of a dielectric film is metallized to form conductive plates.

metallized-paper capacitor A paper-dielectric capacitor whose plates are metal areas electrodeposited on each side of a paper film.

metallized-polycarbonate capacitor A fixed capacitor in which the dielectric is a polycarbonate plastic film, and the plates are metal areas electrodeposited on each face of the film.

metallized resistor See METAL-FILM RESISTOR.

metal locator An electronic device for locating metal deposits, pipes, or wires underground, in walls, or under floors. It operates via the disturbance that these objects cause to a radio-frequency or magnetic field.

metalloid An element that has some of the properties of a metal. Examples of metalloidal elements widely used in electronics are antimony, arsenic, germanium, silicon, and tin.

metal master See ORIGINAL MASTER.

metal negative See ORIGINAL MASTER.

metal-oxide resistor A resistor in which the resistance material is a film of tin oxide deposited on a substrate.

metal-oxide semiconductor field-effect transistor Abbreviation, MOSFET. A field-effect transistor in which the gate electrode is not a pn junction (as in the junction field-effect transistor), but a thin metal film insulated from the semiconductor channel by a thin oxide film. Gate-control action is entirely electrostatic. Also called *insulated-gate field-effect transistor.* Also see DEPLETION-TYPE MOSFET, DEPLETION-ENHANCEMENT-TYPE MOSFET, and ENHANCEMENT-TYPE MOSFET.

metal-oxide varistor A VOLTAGE-DEPENDENT RESISTOR in which the resistance material is a metallic oxide, such as zinc oxide.

metal-plate rectifier See METALLIC RECTIFIER.

metal tube A vacuum tube housed in a metal envelope for self-shielding and mechanical ruggedness.

metamer A visible-light beam that is identical in color (hue), but different in concentration (saturation), with respect to a reference color.

meteor-burst signals Momentary signals, or increases in signal strength, resulting from reflection of electromagnetic energy from meteor ionization trails. See METEOR SCATTER, **1**.

meteor ionization trail A cloud of ions left in the upper atmosphere as a meteor passes. This cloud tends to reflect radio signals at certain frequencies for a short period of time. During a meteor shower, there could be a sufficient number of such trails to allow continuous over-the-horizon communication when other over-the-horizon modes are unusable.

meteorograph An instrument for the simultaneous measurement of various meteorological phenomena such as temperature, humidity, etc.

meteorology The science of the atmosphere, especially the study of weather and climate. (Not to be confused with METROLOGY.)

meteor-scatter propagation The reflection of radio signals from the ionized trails produced by meteors as they pass through the upper atmosphere. This can result in over-the-horizon radio communication or reception. A meteor produces a trail that persists for a few tenths of a second up to several seconds, depending on the size of the meteor, its speed, and the angle at which it enters the atmosphere. This is not sufficient time for the transmission of very much information, but during a meteor shower, ionization can be almost continuous. Meteor-scatter propagation has been observed at frequencies considerably above 30 MHz.

meteor-trail reflections Momentary reflection of signals by the ionized trails of meteors passing through a signal path.

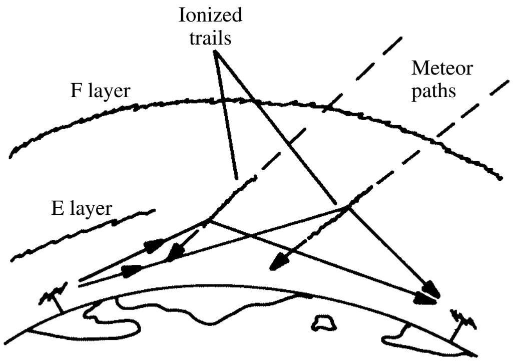

meteor-trail reflections

meter **1.** An instrument for measuring and indicating the value of a particular quantity. See, for example, CURRENT METER and VOLTMETER. **2.** Abbreviation, m. A unit of linear measure and of electrical wavelength, equivalent to 1.65076373×10^6 wavelengths (in a vacuum) of the radiation corresponding to the transition between the two levels of the krypton-86 atom, and approximately equal to 39.37 inches. **3.** To supply in specific increments or by a governed amount.

meter alignment See VISUAL ALIGNMENT.

meter-ampere A unit of transmitted radio signal intensity. Determined by multiplying the antenna current (in amperes) by the height (in meters) of the antenna above ground.

meter-candle Abbreviation, mc. A metric unit of illuminance, equivalent to the illumination on a surface 1 meter from a light source of 1 candle power. Compare FOOT-CANDLE and LUX.

meter equivalent The number of meters equal to a given English measure of length (e.g., the meter equivalent of 3 feet is approximately 0.9144).

meter-kilogram-second Abbreviation, mks. The system of units in which the meter is the standard unit of length, the kilogram is the standard unit of mass, and the second is the standard unit of time. Compare CENTIMETER-GRAM-SECOND and INTERNATIONAL SYSTEM OF UNITS.

meter multiplier See MULTIPLIER RESISTOR.

meter protector A nonlinear resistor, such as a varistor or semiconductor diode, used to prevent overswing in an electric meter by limiting the current flowing through it.

meter rating The maximum reading on a meter, at or below which the accuracy is within a specified limit, but above which the error might exceed that limit.

meter rectifier A light-duty semiconductor diode or bridge circuit, used to change alternating current (ac) to direct current (dc) for deflection of a D'Arsonval-type dc milliammeter or microammeter.

meter relay A sensitive relay that is essentially a moving-coil meter, whose pointer closes against a stationary contact mounted at some point along the scale.

meter resistance Symbol, R_m. The internal resistance of an electric meter. In a simple D'Arsonval meter, it is the resistance of the movable coil. In more-complicated meter circuits, it is the resistance of the parallel combination of the coil and METER SHUNT.

meter scale factor See SCALE FACTOR, **1**.

meter sensitivity See VOLTMETER SENSITIVITY.

meter shunt A resistor connected in parallel with an ammeter, milliammeter, or microammeter to increase the range of currents that the device can measure.

meter torque See DEFLECTING TORQUE.

meter-type relay See METER RELAY.

methyl methacrylate resin Also known by the trade name *Lucite.* A plastic insulating material. Dielectric constant, 2.8 to 3.3. Dielectric strength, 20 kV/mm.

metre Abbreviation, m. Alternate spelling of *meter* when used to specify displacement or wavelength. See METER, **2**.

metric system The decimal system of weights and measures based on the meter, kilogram, and second. Also see CENTIMETER-GRAM-SECOND, METER-KILOGRAM-SECOND, and INTERNATIONAL SYSTEM OF UNITS.

metric ton Abbreviation, MT. A metric unit of weight equal to 1000 kilograms or 1.1023 English tons.

metric waves British designation for electromagnetic energy having wavelengths ranging from 10 meters down to 1 meter, corresponding to frequencies from 30 MHz up to 300 MHz.

metrology The science of weights and measures, including electrical standards and electronic instruments and measurements. (Not to be confused with METEOROLOGY.)

metronome A mechanical or electronic device that produces audible beats (ticks). It is commonly used in setting the tempo for music, and for audibly timing certain processes.

MeV Abbreviation of MEGAELECTRONVOLT(s).

MEW Abbreviation of MICROWAVE EARLY WARNING.

mF Abbreviation of MILLIFARAD.

MF **1.** Abbreviation of MEDIUM FREQUENCY. **2.** Abbreviation of MIDFREQUENCY.

MFD Abbreviation of MAGNETOFLUID DYNAMICS (see MAGNETOHYDRODYNAMICS).

MFSK Abbreviation of MULTIPLE FREQUENCY-SHIFT KEYING.

Mg Symbol for MAGNESIUM.

mg Abbreviation of MILLIGRAM.

MGD Abbreviation of MAGNETOGASDYNAMICS (see MAGNETOHYDRODYNAMICS).

MHD Abbreviation of MAGNETOHYDRODYNAMICS.

MHD generator See MAGNETOHYDRODYNAMIC GENERATOR.

MHD gyroscope See MAGNETOHYDRODYNAMIC GYROSCOPE.

MHD power generation See MAGNETOHYDRODYNAMIC GENERATOR.

mho Obsolete term for the standard unit of electrical conductance. See SIEMENS.

mhp Abbreviation of MILLIHORSEPOWER.

MHz Abbreviation of MEGAHERTZ.

mi Abbreviation of MILE. (Also, m.)

MIC Abbreviation of MICROWAVE INTEGRATED CIRCUIT.

mic Abbreviation of MICROPHONE.

mica A dielectric mineral of complex silicate composition, easily separated into numerous thin, transparent sheets. It is widely used as a capacitor dielectric and high-temperature electrical insulator. Dielectric constant, 2.5 to 7. Dielectric strength, 50 to 220 kV/mm. Also see MUSCOVITE.

mica capacitor A component that is made by alternately stacking metal sheets and layers of mica, or by applying silver ink to sheets of mica. The metal sheets are wired together into two meshed sets, forming the two terminals of the capacitor. This type of capacitor is noted for low loss. Voltage ratings can be up to several thousand volts if thick sheets of mica are used. But these capacitors are physically bulky in proportion to their capacitance. The main application is in radio receivers and transmitters. Compare CERAMIC CAPACITOR, ELECTROLYTIC CAPACITOR, PAPER CAPACITOR, PLASTIC-FILM CAPACITOR, TANTALUM CAPACITOR.

MICR Abbreviation of MAGNETIC INK CHARACTER RECOGNITION.

micro- **1.** A prefix meaning MILLIONTH(S) (i.e., 10^{-6}). **2.** A prefix meaning extremely small (as in microstructure). Compare MACRO-.

microalloy diffused transistor Abbreviation, MADT. A MICROALLOY TRANSISTOR having a uniform base region that is diffused into the wafer before the emitter and collector electrodes are produced by alloying.

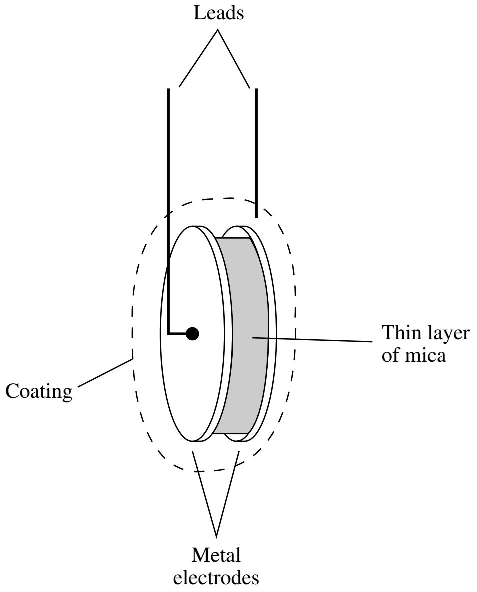

mica capacitor

microalloy transistor Abbreviation, MAT. A transistor having tiny emitter and collector electrodes that are formed by alloying a thin film of impurity material with a collector pit and emitter pit facing each other on opposite surfaces of the semiconductor wafer. Also see SURFACE-BARRIER TRANSISTOR.

microammeter A usually direct-reading instrument used to measure current in the microampere range. Also see CURRENT METER.

microampere A small unit of current, equal to 10^{-6} (0.000001) A.

microbalance A sensitive electronic weighing device. One type uses one or more servo amplifiers for the balancing operation.

microbar A cgs unit of low pressure, equal to 10^{-6} b or 0.1 pascal. Also see BAR, **1** and MILLIBAR.

microbarograph A barograph that is sensitive to small changes in pressure.

microbeam An energy beam (ray) having extremely small cross section.

microcircuit An extremely small circuit fabricated upon and within a substrate, such as a semiconductor chip. Also see INTEGRATED CIRCUIT.

microcode **1.** See MICROINSTRUCTION. **2.** A code for MICROPROGRAMMING.

microcomponent A tiny component in an electronic circuit. Examples are the resistors, capacitors, diodes, and transistors fabricated onto an integrated-circuit chip.

microcomputer **1.** A computer contained within a single integrated circuit (IC). The simplest such devices perform elementary functions and are available for a few dollars. More sophisticated devices control radio receivers and transmitters, television sets, automobiles, aircraft, and robots. The most advanced devices can be programmed to provide electrical impulses to control erratically functioning body organs, move the muscles of paralyzed persons, and transcribe speech to writing or vice versa. Compare MICROPROCESSOR. **2.** In general, any small computer.

microcrystal A crystal that is invisible to the naked eye.

microcurie A small unit of radioactivity equal to 3.71×10^4 disintegrations per second or 10^{-6} curie. Also see CURIE.

microelectrode **1.** An electrode used in MICROELECTROLYSIS. **2.** A tiny electrode, especially one of those used in integrated circuits and in certain biological applications.

microelectrolysis Electrolysis of tiny amounts of material. Also see ELECTROLYSIS, **1**.

microelectronic circuit **1.** A tiny electronic circuit other than an INTEGRATED CIRCUIT (i.e., one assembled in a small space with small discrete or integrated components). **2.** See MICROCIRCUIT.

microelectronic device See MICROELECTRONIC CIRCUIT.

microelectronics The branch of electronics dealing with extremely small components and circuits fabricated on substrates. Also see INTEGRATED CIRCUIT.

microelectrophoresis Electrophoresis of single particles.

microelectroscope A very sensitive electroscope used to detect minute quantities of electricity.

microelement A tiny component (capacitor, resistor, coil, semiconductor device, or transformer) mounted on a wafer and used in a MICROCIRCUIT.

microelement wafer A microwafer on which a microelement is mounted or deposited.

microfarad Abbreviation, μF. A unit of capacitance equal to 10^{-6} (0.000001) F.

microfarad meter **1.** A dynamometer-type meter that indicates the value of a capacitor directly in microfarads. Such instruments operate from an alternating-current power line. **2.** A direct-reading capacitance meter.

microfiche A method of storing printed information on small film cards. The pages are reduced and arranged in order from left to right and top to bottom. The card is inserted into a projecting machine to allow retrieval of the information. The photographic method is similar to that used in MICROFILM.

microfilm A method of storing printed or photographic information. The pages are reduced and arranged sequentially on a strip of film, usually 35-mm size. The film is inserted into a projecting device for retrieval of the information.

microgalvanometer A highly sensitive GALVANOMETER.

microgauss A magnetic unit equal to 10^{-6} (0.000001) gauss.

microgram A metric unit of weight or mass equal to 10^{-6} (0.000001) gram.

microgroove record A phonograph disc with a very fine groove (200 to 300 per inch), designed for playback at 33⅓ revolutions per minute (rpm).

microhenry Symbol, μH. A unit of inductance, equal to 10^{-6} (0.000001) H.

microhm Symbol, μΩ. A unit of low resistance, reactance, or impedance, equivalent to 10^{-6} (0.000001) ohm.

microhm-centimeter A unit of low resistivity, equal to 10^{-6} (0.000001) ohm-cm. See OHM-CENTIMETER and RESISTIVITY.

microhmmeter An instrument for measuring ultra-low resistance. Such an instrument must have a special provision for canceling the effects of contact and lead resistance.

microinch A unit of linear measure equal to 10^{-6} (0.000001) inch.

microinstruction A machine-code instruction that controls the operation of a computer directly (i.e., it is a "wired-in" instruction, or one set by DIP switches, independent of programs loaded into the machine).

microknowledge In artificial intelligence (AI), detailed machine knowledge. It includes logic rules, computer programs, and data in memory. Compare MACROKNOWLEDGE.

microliter A unit of volume, equal to 10^{-6} (0.000001) liter.

microlock A special form of phase-locked-loop system, used especially with radar to improve the signal-to-noise ratio.

micromanipulator A machine that permits handling tiny parts in very small areas. An example of its use is in placing connections close together in microcircuits.

micrometer **1.** An instrument for measuring very small thicknesses, diameters, etc. **2.** Also called *micron.* The SI unit of length, equal to 10^{-6} (0.000001) meter, or 10^{-3} (0.001) millimeter.

micromho See MICROSIEMENS.

micromicro- See PICO-.

micromicrofarad See PICOFARAD.

micromicrohenry See PICOHENRY.

micromicron A unit of linear measure equal to 10^{-12} meter, or 10^{-6} (0.000001) micrometer.

micromillimeter See NANOMETER.

microminiature Pertaining to an extremely small body, component, or circuit; the last adjective in the sequence of those describing size: *standard*, *small*, *midget*, *miniature*, *subminiature*, and *microminiature.*

micromodule A small, encapsulated circuit, consisting of smaller components. The components can be discrete, can consist of integrated circuits, or can be a combination of both. The module is easily removed and replaced by means of a plug-in socket.

micron See MICROMETER.

microphone A transducer that converts sound waves, especially speech and music, into electrical voltage analogs.

microphone amplifier A sensitive, low-distortion, low-noise amplifier used in voice wireless transmitters and public address systems. Most amplifiers of this type have a tailored frequency response, passing audio between about 300 Hz nd 3000 Hz, and attenuating audio outside this a range. The range 300 Hz to 3000 Hz is sufficient to convey intelligible voice signals, and also allows for audio-frequency-shift keying (AFSK) and slow-scan television (SSTV) audio input.

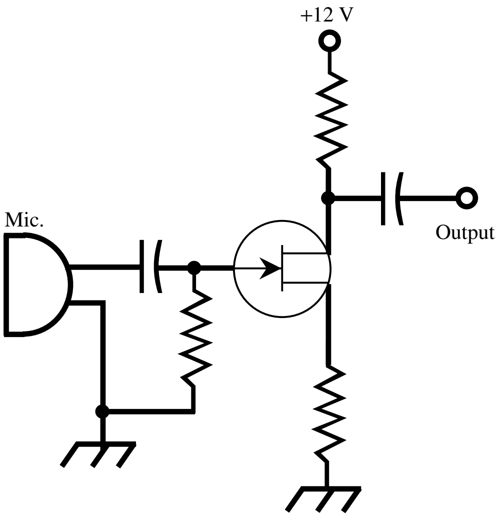

microphone amplifier

microphone boom A device used to hang a microphone, with the base out of the way. It is often used in radio broadcasting.

microphone hummer See HUMMER.

microphone input In an audio amplifier, a jack or other receptacle provided for connection to an external microphone. It can also be used with other low-level audio apparatus. The jack is connected to a MICROPHONE AMPLIFIER that provides high gain with minimum internal noise.

microphone oscillator See HUMMER.

microphonics Ringing (electrical noises) set up by the vibration of a component having loose or movable elements. For example, ringing noises are generated in some circuit boards when they receive a physical blow.

microphonograph A recorder of very low-intensity sound.

microphonoscope An electronic stethoscope, using amplification to enhance the response.

microphotograph An extremely small photograph, often of a pattern or mask used in producing transistors and integrated circuits. Not to be confused with *photomicrograph*, a photograph taken through a microscope.

microphotometer A sensitive instrument for measuring small-area light intensity.

microphysics The branch of physics concerned with atoms, molecules, and subatomic particles.

micropower Extremely small amounts of power—especially the very low direct-current supply power required by some transistors.

microprocessor The integrated circuit (IC), also known as a *chip*, that coordinates the actions of a computer and does the calculations. It is located on the motherboard (sometimes called the *logic board*). These devices get more powerful every year. Physically, this translates to an increasing number of digital switching transistors per chip. The number of digital switches that can be fabricated onto a semiconductor chip of a particular size is ultimately limited by the structure of matter. Compare MICROCOMPUTER.

microprogram **1.** In computer operations, a routine of microinstructions that provides a computer a specific function, independent of those established by programs being run or by the monitor program. Also see MICROINSTRUCTION. **2.** In the direction of a computer, use of a routine that is stored specifically in the memory, instead of elsewhere.

microprogramming In the direction of a computer, the use of a routine that is stored specifically in the memory, instead of elsewhere.

micropulsation Also called *micropulse.* A pulse of extremely short duration.

microradiometer A sensitive detector of heat and infrared radiation, consisting essentially of a thermopile carried by the moving coil of a galvanometer.

microrutherford A unit of radioactivity equal to one disintegration per second or 10^{-6} (0.000001) rutherford. Also see RUTHERFORD.

microsecond A unit of time measure equal to 10^{-6} (0.000001) second.

microsiemens A unit of conductance equal to 10^{-6} (0.000001) siemens.

microspectrophotometer An extremely sensitive spectrophotometer for examining light from tiny areas.

microstrip A microwave component that is, in effect, a single-wire transmission line operating above ground.

microsyn A device that translates rotational position into an electrical signal. Similar to a SELSYN. It is used for such purposes as rotator-direction reading.

microsystems electronics The technology of electronic systems using tiny electronic components. Also see INTEGRATED CIRCUIT, MICROELECTRONIC CIRCUIT, MICROELEMENT, MICROELEMENT WAFER, and MICROWAFER.

microvolt A unit of low voltage, equal to 10^{-6} (0.000001) volt.

microvolter An accurate, external attenuator (usually for an audio signal generator) providing stepped and continuously variable output in microvolts and millivolts.

microvoltmeter A usually direct-reading instrument used to measure voltages in the microvolt range. An input amplifier boosts the test voltage sufficiently to deflect the indicating meter.

microvolts per meter A unit of radio-frequency (RF) field strength. It refers to the RF voltage (in microvolts) between an antenna and ground, divided by the height of the antenna (in meters) above ground. Compare MILLIVOLTS PER METER.

microvolts per meter per mile A means of expressing absolute radio-frequency (RF) field strength. Generally, the numerical value is based on the field strength, in MICROVOLTS PER METER, at a distance of 1 statute mile (5280 feet) from the source.

microwafer A wafer of insulating material, such as a ceramic, on which one or more microelements are mounted and terminals deposited or plated.

microwatt A unit of low power, especially electrical power, equal to 10^{-6} (0.000001) watt.

microwattage See MICROPOWER.

microwattmeter An instrument for measuring power in the microwatt range. Such an instrument obtains its sensitivity from a built-in input amplifier.

microwave See MICROWAVES.

microwave security system A circuit using microwave radio-frequency energy to detect intruders. When an object moves within the field, the intensity of the field changes at one or more pickup sensors, triggering an alarm.

microwave acoustics See ACOUSTOELECTRONICS and ACOUSTIC DELAY LINE.

microwave dish A dish antenna for use at microwave frequencies.

microwave early warning Abbreviation, MEW. A high-power early warning radar system that affords large traffic-handling capacity and long range.

microwave filter A bandpass filter built into a waveguide for use at microwave frequencies.

microwave frequencies The general expression for radio frequencies above the ultra-high range, that is, 3 GHz or more, but below the frequencies of infrared energy. This corresponds to radio wavelengths of 10 centimeters or less.

microwave integrated circuit Abbreviation, MIC. An integrated circuit designed for use at microwave frequencies.

microwave lens See WAVEGUIDE LENS.

microwave mirror A reflector of microwaves.

microwave oven A device consisting essentially of a radio-frequency heater using a magnetron oscillator. It produces microwave energy that causes heating of certain substances via excitation of the molecules.

microwave plumbing Collectively, the waveguides, tees, elbows, and similar fixtures and connections used in microwave setups.

microwave radio relay The use of microwaves to relay radio, television, and control signals from point to point.

microwave refractometer An instrument using microwaves (around 10 GHz) to measure the refractive index of the atmosphere.

microwave region See MICROWAVE FREQUENCIES and MICROWAVES.

microwave relay See MICROWAVE RADIO RELAY.

microwave relay system A series of microwave transmitter-receiver stations for relaying communications in several line-of-sight hops.

microwaves Radio-frequency electromagnetic energy at wavelengths shorter than about 10 centimeters, but longer than the wavelengths of infrared energy. See also MICROWAVE FREQUENCIES.

microwave spectrum See MICROWAVE FREQUENCIES and MICROWAVES.

microwave transistor A transistor whose semiconductor properties and special fabrication enable it to operate at microwave frequencies.

microwave tube A KLYSTRON, MAGNETRON, or similar tube, used to generate or amplify microwave radio-frequency signals.

midband Abbreviation, MB. The region whose limits are immediately above and below a MIDFREQUENCY; the limits are usually specified for a particular case.

midband frequency See MIDFREQUENCY.

midfrequency The center frequency in a specified band of frequencies.

midget Of reduced size (smaller than *small* and larger than *miniature*).

MIDI Acronym for MUSICAL INSTRUMENT DIGITAL INTERFACE.

midpoint voltage The voltage at the terminals of a cell or battery when it has been discharged halfway (i.e., when the amount of energy used up is equal to the amount of energy remaining).

midrange Pertaining to audio frequencies in the middle of the human hearing range, where the ear is the most sensitive. These frequencies lie between the BASS and TREBLE.

midrange horn A MIDRANGE SPEAKER equipped with a flared horn to give the device a unidirectional sound-emission pattern. It is used primarily in high-power systems and by popular music bands or high-end audio enthusiasts.

midrange speaker A loudspeaker operating most efficiently at frequencies in the middle of the audio spectrum. Such a speaker is intermediate in performance between a WOOFER and a TWEETER.

midsection The center section of a multisection filter having an odd number of sections; thus, the second section of a three-section filter.

migration See ION MIGRATION.

mike **1.** See MICROPHONE. **2.** See MICROFARAD. **3.** See MICROMETER.

MIL Abbreviation of *military.*

mil **1.** A small unit of linear measure; 1 mil = 10^{-3} (0.001) inch = 0.0254 mm. **2.** Thousand, as in *n* parts per mil.

mile Abbreviation, m or mi. A large unit of linear measure, 1 mi = 1.609 km = 5280 feet.

military robot A robot designed and used for the purpose of executing some task in warfare. The two general types are: human-operated and computer-controlled. An example of a human-operated military robot is an aircraft that is flown by remote control by a ground-based pilot. The same robot, or a whole fleet of them, might be flown by a computer using sophisticated EXPERT SYSTEMS.

mill A telegraph operator's typewriter.

Miller oscillator A crystal oscillator circuit in which the crystal is connected between the control electrode and ground. The tuned tank is connected in the output circuit. The internal capacitance of the active device provides feedback coupling. Sometimes called *conventional crystal oscillator.*

milli- Abbreviation, m. A prefix meaning thousandth(s): 10^{-3} (0.001).

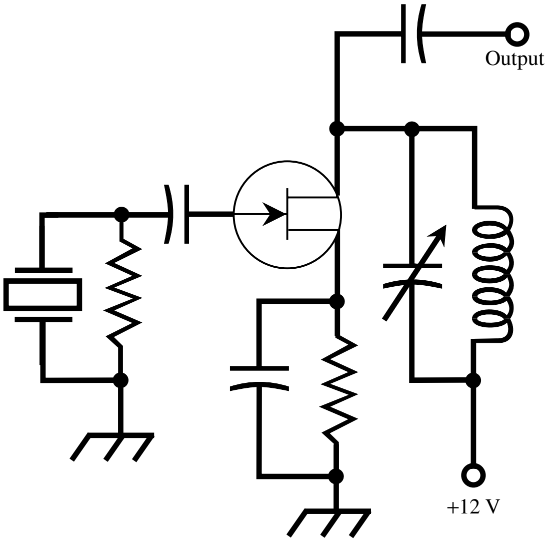

Miller oscillator

milliammeter A usually direct-reading instrument for measuring current in the milliampere range. Also see CURRENT METER.

milliampere Abbreviation, mA. A unit of current equal to 10^{-3} (0.001) ampere.

milliampere-hour Abbreviation, mAh. A unit of low current drain or charging rate, equal to 10^{-3} (0.001) ampere-hour. Also see AMPERE-HOUR and BATTERY CAPACITY.

millibar Abbreviation, mb. A unit of low pressure equal to 10^{-3} (0.001) bar or = 100 pascals.

millicurie Abbreviation, mCi. A small unit of radioactivity equal to 3.71×10^7 disintegrations per second, or 10^{-3} (0.001) curie. Also see CURIE.

millifarad Abbreviation, mF. A seldom-used unit of capacitance, equal to 10^{-3} (0.001) farad or 1000 microfarads.

milligram Abbreviation, mg. A metric unit of weight equal to 10^{-3} (0.001) gram.

millihenry Abbreviation, mH. A unit of inductance, equal to 10^{-3} (0.001) henry.

millihorsepower Abbreviation, mhp. A unit of power equal to 10^{-3} (0.001) horsepower or 0.746 watt. Also see HORSEPOWER.

millilambert Abbreviation, mL. A small unit of brightness equal to 10^{-3} (0.001) lambert.

milliliter Abbreviation, ml. A metric unit of volume equal to 10^{-3} (0.001) liter.

millimaxwell Abbreviation, mMx. A small unit of magnetic flux equal to 10^{-3} (0.001) maxwell or 10^{-11} weber.

millimeter Abbreviation, mm. A metric unit of linear measure equal to 10^{-3} (0.001) meter or 0.03937 inch.

millimeter equivalent The number of millimeters equal to a given English measure fraction (e.g., the millimeter equivalent of $\frac{5}{16}$ inch is 7.937).

millimeter waves Wavelengths between 0.6 and 10 mm (frequencies from 30 to 500 GHz).

millimicro See NANO-.

millimicrofarad See NANOFARAD.

millimicrohenry See NANOHENRY.

millimicron Abbreviation, mm. A unit of wavelength equal to 10^{-3} micron or one nanometer (10^{-9} meter).

millimilliampere See MICROAMPERE.

millimole Abbreviation, mmol. A unit in chemistry equal to 10^{-3} (0.001) mole.

milliohm A small unit of resistance, reactance, or impedance, equal to 10^{-3} (0.001) ohm.

milliohmmeter An ohmmeter for measuring resistances in the milliohm range.

million electronvolt(s) See MEGAELECTRONVOLT.

milliphot A unit of illumination equal to 10^{-3} (0.001) phot.

millipuffer See PUFFER.

milliradian Abbreviation, mrad. A unit of angular measure equal to 10^{-3} (0.001) radian.

milliroentgen Abbreviation, mr. A small unit of radioactive dosage; 1 mr = 10^{-3} (0.001) roentgen = 2.57976×10^{-7} Ci/kg.

millirutherford Abbreviation, mrd. A small unit of radioactivity equal to 1000 disintegrations per second; 1 mrd = 10^{-3} (0.001) rutherford.

millisecond Abbreviation, ms or msec. A small unit of time equal to 10^{-3} (0.001) second.

millitorr Abbreviation, mT. An obsolete unit of low pressure equal to 10^{-3} (0.001) torr, or 0.133322 pascal. Also see TORRICELLI.

millivolt Abbreviation, mV. A unit of voltage equal to 10^{-3} (0.001) volt.

millivoltmeter A usually direct-reading instrument for measuring low electric potential. Its sensitivity is provided by a high-gain amplifier operated ahead of the indicating meter.

millivolt potentiometer Abbreviation, MVP. A potentiometer-type null instrument for accurately measuring small direct-current voltages, such as those delivered by a thermocouple. Also see POTENTIOMETRIC VOLTMETER.

millivolts per meter Abbreviation, mV/m. A unit of radio-frequency (RF) field strength. It refers to the RF voltage (in millivolts) developed between an antenna and ground, divided by the height (in meters) of the antenna above ground. Compare MICROVOLTS PER METER.

milliwatt Abbreviation, mW. A unit of power equal to 10^{-3} (0.001) watt.

milliwattmeter An instrument for measuring power in milliwatts. Such instruments usually obtain their sensitivity from a built-in preamplifier.

Mills cross A radio-telescope antenna, consisting of two collinear or phased arrays with a common intersecting lobe. The result is high resolution.

mil-spec security See LEVEL-3 SECURITY.

min **1.** Abbreviation of MINIMUM. **2.** Abbreviation of MINUTE.

mineral An element or compound that occurs naturally in the earth's crust. Most minerals are crystalline and many of these have found use in electronics. Some have been produced artificially.

mineral oil A natural liquid insulant derived from petroleum. Dielectric constant, 2.7 to 8.0. Power factor, 0.08 to 0.2 percent at 1 kHz.

mineral-oil capacitor An oil capacitor whose paper dielectric has been impregnated with mineral oil, which is also the filler.

miniature Very small (smaller than midget and larger than subminiature).

miniature cell An electrochemical cell of very small size (e.g., a *button cell* of the kind used in cameras and watches).

miniaturization The technology of minimizing the physical size of a circuit or system, while maintaining its ability to accomplish a given task.

minicalculator A pocket-size electronic calculator.

minicomputer General term for a computer that is more sophisticated than a MICROCOMPUTER, but less powerful than a MAINFRAME.

mini connector A jack or plug having two or three conductors, and measuring ⅛ (0.125) inch in diameter. It is commonly used with audio equipment.

minifloppy A smaller than standard flexible magnetic disk (floppy).

minima Points along a curve at which a function reaches a local minimum value. Also see MAXIMA AND MINIMA.

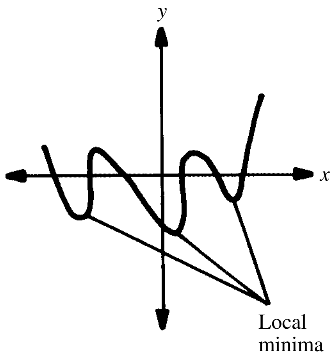

minima

minimum Abbreviation, min. The smallest value in a range or set. Also see MAXIMA AND MINIMA.

minimum detectable signal A signal whose intensity is just higher than the threshold of detection.

minimum discernible signal Abbreviation, MDS. The lowest input-signal amplitude that will produce a discernible output signal in a radio receiver.

miniscope A very-small-sized, lightweight oscilloscope.

minitrack A system used to track an earth satellite, using signals transmitted to the satellite by a line of ground radio stations.

minometer A radioactivity-measuring instrument composed of an ionization chamber and a string galvanometer.

minor beats Secondary or extraneous beats produced in a beat-note system, caused by various sum and difference frequency byproducts of the heterodyne process. Compare MAJOR BEATS.

minor bend A bend in a rectangular waveguide, made without twisting.

minor cycle See WORD TIME.

minor face In a hexagonal quartz crystal, one of the three smaller faces. Compare MAJOR FACE.

minority carrier The type of charge carrier present in relatively small numbers in a processed semiconductor material. Electrons are minority carriers in p-type material; holes are minority carriers in n-type material. Compare MAJORITY CARRIER.

minor lobe In a directional antenna system, any lobe other than the main lobe(s). Such a lobe represents reduced sensitivity and/or power gain relative to the main lobe. Also see MAIN LOBE, SIDE LOBE.

minor loop A subordinate path for the circulation of information or control signals in an electronic system (e.g., *minor feedback loop*). Compare MAJOR LOOP.

minute **1.** Abbreviation, min. A unit of measure of time equal to 60 seconds or $\frac{1}{60}$ hour. **2.** Also called *minute of arc.* Symbol (′). A unit of arc measure equal to $\frac{1}{60}$ angular degree or 60 seconds. **3.** General term meaning "extremely small."

MIR Abbreviation of *memory-information register.*

mirror **1.** A device consisting chiefly of a highly polished or silvered surface that reflects a large part of the radiation (such as light) striking it. **2.** Radar-interference material (see CHAFF). **3.** To reflect, as by a mirror.

mirror galvanometer A galvanometer in which a mirror is moved by the coil. The mirror either reflects a spot of light along an external scale, or it reflects the scale, which is then read through a small telescope.

mirror-galvanometer oscillograph See ELECTROMECHANICAL OSCILLOSCOPE.

mirror image **1.** An image or curve that is exactly reversed relative to a straight line or flat plane, compared to a reference image or curve. Compare BILATERAL SYMMETRY. **2.** For a quarter-wave Marconi antenna, the extra quarter-wave element supplied by the earth. **3.** For an antenna at a distance *d* above a ground plane, an effective antenna at an equal distance *d* below the ground plane.

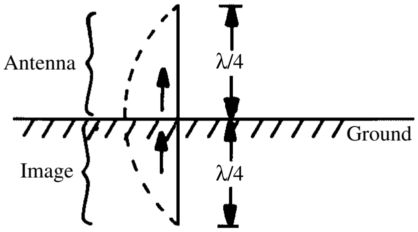

mirror image, 2

mirror-reflection echo A false radar echo or set of echoes, caused by reflection of the radar beam from a plane surface prior to its encountering the target or targets. The beam can also be reflected from one target to another.

mirror-type meter A meter whose movable coil carries a small mirror (rather than a pointer) that reflects a beam of light to produce a spot on a translucent scale.

misaligned head In a tape recorder, a record or a pickup head that is incorrectly oriented, with respect to the passing tape.

misfire Failure of a gas tube or mercury-arc tube to ignite at the correct instant.

misleading precision In electronic calculations and data recording, greater precision than the instruments or conditions justify. Also see SIGNIFICANT FIGURES.

mismatch The condition resulting from joining two circuits, or connecting a line to a circuit, in which the impedances are substantially different.

mismatched impedances Impedances that are unequal, and thus do not satisfy the conditions for maximum power transfer.

mismatch factor For a load not perfectly matched to a driving circuit, the ratio of current flowing in the load to the current that would flow in the load if its impedance were perfectly matched to the output impedance of the driving circuit.

mismatch loss For a load that is mismatched to a source, the ratio P_1/P_2, where P_1 is the power a matched load would absorb from the source, and P_2 is the power actually absorbed by the mismatched load.

mistor A variable-resistance device, used to detect the presence of a magnetic field and to measure magnetic-field strength.

MIT Abbreviation of MASTER INSTRUCTION TAPE.

mix **1.** To produce a beat signal (either the sum or the difference frequency) from two input signals. **2.** The proportion of powdered iron and other inert substances in a ferromagnetic transformer core. Different mixes result in different operating characteristics.

mixdown A method of combining recorded sound from two or more audio tracks, and recording the result onto an audio tape or disc. It is used to create special audio effects.

mixed-base notation A number system in which the base (radix) alternates between two digit positions. Also called *mixed radix notation.*

mixed calculation A mathematical calculation or expression in which more than one operation is used.

mixed modulation Modulation of several kinds coexisting in a system. Thus, a small amount of undesired frequency modulation might accompany amplitude modulation, or vice versa.

mixed number A number having integral (whole) and fractional parts (e.g., 3.14159).

mixer **1.** A device, such as a transistor or semiconductor diode, used to mix two input signals and deliver an output equal to their difference and/or sum (see MIXING). **2.** See AUDIO MIXER. **3.** Any device that combines two or more signals, yielding one output signal whose nature is determined by the characteristics of the circuit.

mixer noise Electrical noise that occurs in a MIXER.

mixing Combining several signals so that some desired mixture of the original signals is obtained. Compare MODULATION.

mixture **1.** A combination of two or more signals that retain their characteristics—even when they interact to produce beat-frequency products. **2.** A

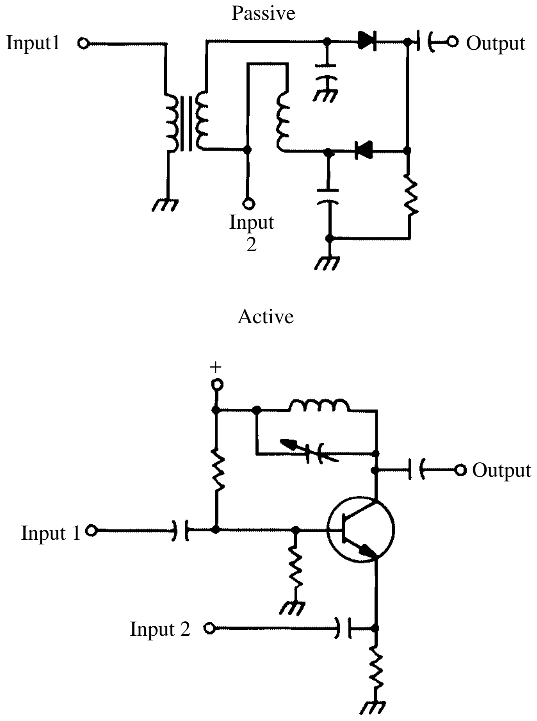

mixer

diffusion of one substance throughout another, without a solution or a chemical reaction resulting.

mixture melting point Abbreviation, mmp. The temperature at which a mixture of solid substances starts turning into a liquid at 1 atmosphere of pressure. This melting point depends upon the melting points of the substances and their relative concentration in the mixture. Also see MELTING POINT and MIXTURE.

mks Abbreviation of METER-KILOGRAM-SECOND.

mL Abbreviation of MILLILAMBERT.

ml Abbreviation of MILLILITER.

mm Abbreviation of MILLIMETER.

mmf Abbreviation of MAGNETOMOTIVE FORCE.

mmol Abbreviation of MILLIMOLE.

mmp Abbreviation of MIXTURE MELTING POINT.

mmv Abbreviation of MONOSTABLE MULTIVIBRATOR.

Mn Symbol for MANGANESE.

m_n Symbol for MASS OF NEUTRON AT REST.

mnemonic **1.** Pertaining to MEMORY or to memory systems. **2.** A memory code or device.

mnemonic code In computer operations, a programming code, such as assembly language, that, although easily remembered by the programmer, requires subsequent conversion to machine language.

mntr Abbreviation of MONITOR. (Also, mon.)

MO Abbreviation of MASTER OSCILLATOR.

Mo Symbol for MOLYBDENUM.

mobile communications **1.** Radio communications between or among stations on board moving or stationary land, waterborne, or airborne vehicles. **2.** Radio communications between at least one fixed station and one or more moving or stationary land, waterborne, or airborne stations.

mobile radio service See MOBILE COMMUNICATIONS.

mobile receiver A radio, television, or other receiver aboard a moving or stationary land, waterborne, or airborne vehicle.

mobile-relay station A fixed station that receives a signal from a MOBILE STATION and retransmits it to one or more other mobile stations.

mobile station A station installed and operated aboard a moving or stationary vehicle. The vehicle might be on land, under water, or in the air. Compare PORTABLE STATION.

mobile transmitter A radio, television, or other transmitter aboard a moving or stationary vehicle. The vehicle might be on land, under water, or in the air.

mobility See CARRIER MOBILITY.

mockup See DUMMY, **1**.

mod **1.** Abbreviation of MODULATOR. **2.** Abbreviation of MODULUS. **3.** Abbreviation of MODIFICATION.

mode **1.** One of the ways a given resonant system can oscillate. **2.** One of the ways that electromagnetic energy can be propagated through a device or system. See MODES OF PROPAGATION. **3.** The method via which intelligence is conveyed in a communications or broadcast signal. See EMISSION MODE. **4.** Resonance of sound waves within an acoustic chamber.

mode coupling The exchange or interaction of energy between identical modes (see MODE, **1**, **2**).

mode filter A waveguide filter that separates waves of different propagation mode, but of the same frequency (see MODES OF PROPAGATION).

model **1.** A working or mockup version of a circuit, system, or device, illustrative of the final version. **2.** A mathematical representation of a process, device, circuit, or system. Example: the Rutherford model of the atom.

modeling The creation of an object in three-dimensional computer graphics.

modem See MODULATOR-DEMODULATOR.

mode purity In a modulated radio-frequency signal, the condition in which no undesirable types of modulation exist. For example, a frequency-modulated signal in which there is zero amplitude modulation.

moderator A substance, such as graphite or heavy water, used to slow neutrons in an atomic reactor. Also see ACCELERATOR, **1** and REACTOR, **2**.

modes of propagation The configurations in which microwave energy can be transmitted through a WAVEGUIDE.

modes of resonance In a microwave cavity, the configurations in which resonant oscillation can exist, depending on the way the cavity is excited.

modification **1.** Changing the configuration of a circuit, device, or system, usually to a minor extent, to tailor its characteristics for a specific purpose. **2.** Changing some aspect of a signal for a specific purpose (e.g., reducing the emission bandwidth to allow more signals to fit within a given band of frequencies). **3.** In computer operations, changing program addresses and instructions by performing logic and arithmetic on them, as if they were data. Also see PROGRAM MODIFICATION.

modified alternate mark inversion A signal that is similar to alternate mark inversion (AMI), but contains certain differences that are specified by a rigorous set of standards for the particular signal.

modifier A data item used to change a computer program instruction so that it can be used to implement different successive operations. Also see PROGRAM MODIFICATION.

modifier register See INDEX REGISTER.

modify To perform a MODIFICATION to a circuit, device, system, signal, program address, etc.

moding A fault characterized by oscillation of a MAGNETRON in undesirable modes.

modular technique See BUILDING-BLOCK TECHNIQUE.

modulated amplifier A usually high-frequency amplifier whose output is varied in some way for the purpose of conveying intelligence. Compare MODULATED OSCILLATOR. In the amplitude modulation of an amplifier, there is little or no disturbance of the carrier frequency. Also see MODULATION.

modulated beam **1.** An electron beam (as in a cathode-ray tube), whose intensity is varied by a desired signal. **2.** A light beam whose intensity is varied for communications or control purposes.

modulated carrier A carrier wave whose amplitude, frequency, or phase is varied to convey intelligence.

modulated continuous wave Abbreviation, mcw. A high-frequency carrier wave modulated by a continuous, lower-frequency wave, as in MCW telegraphy.

modulated CW See MODULATED CONTINUOUS WAVE.

modulated electron beam See MODULATED BEAM, **1**.

modulated light beam See MODULATED BEAM, **2**.

modulated oscillator A usually high-frequency oscillator whose output is varied in some way to convey intelligence. Compare MODULATED AMPLIFIER. Also see MODULATION.

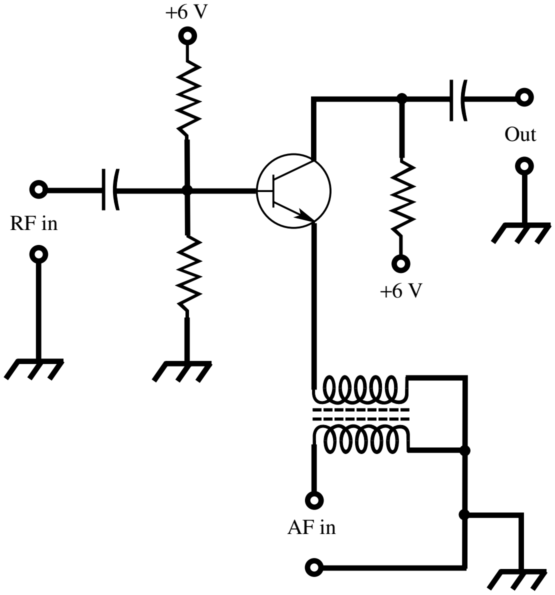

modulated amplifier

modulated-ring pattern See GEAR-WHEEL PATTERN and SPOT-WHEEL PATTERN.

modulated stage A transmitter, amplifier, or oscillator in which the signal information is impressed on the carrier.

modulated wave See MODULATED CARRIER.

modulatee A stage or circuit upon which modulation is impressed (e.g., a MODULATED AMPLIFIER or a MODULATED OSCILLATOR).

modulating electrode **1.** In an oscilloscope, an electrode (usually the intensity electrode) to which a signal can be applied to intensity-modulate the electron beam. **2.** In a cathode-ray tube, the electrode to which the video signal is applied.

modulating signal Intelligence that modulates a carrier (e.g., binary data in radioteletype, or music in broadcasting).

modulation Combining two signals with the result that some aspect of one signal (the carrier) is varied by and in sympathy with the other (the modulating signal). Usually, the carrier has a frequency considerably higher than that of the modulating signal.

modulation bars A form of television interference in which an amplitude-modulated signal causes horizontal bars, alternating light and dark, to appear on the picture screen. The higher the modulating frequency, the closer together the bars appear. In severe cases, the bars completely obliterate the picture.

modulation capability The maximum percentage of modulation a transmitter will permit before

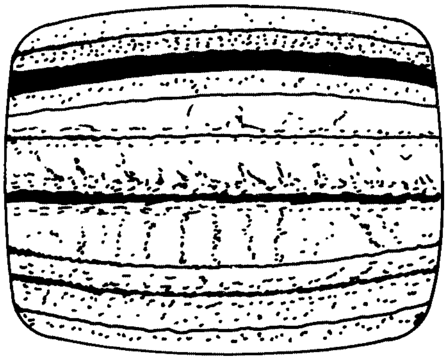

modulation bars

nonlinearity occurs. Also see MODULATION LINEARITY.

modulation characteristic For an amplitude-modulated wave, the ratio of the instantaneous amplitude of the modulated signal to the instantaneous modulating voltage.

modulation code In a modulated transmitter, a system of modulation in which certain signal variations or pulses represent particular characters. Examples are ASCII, BAUDOT CODE, and MORSE CODE.

modulation coefficient Symbol, m. A figure expressing the depth (extent) to which a signal is amplitude-modulated. For a signal in which the upward modulation is equal to the downward modulation, $m = (E_m - E_c)/E_c$, where E_c is the peak-to-peak voltage of the unmodulated signal, and E_m is the peak-to-peak voltage of the modulated signal. For full (100%) modulation, $m = 1$.

modulation depth See DEPTH OF MODULATION.

modulation distortion **1.** In a modulated signal, envelope distortion introduced by the modulation process or by the receiver circuit. **2.** External cross modulation (see CROSS MODULATION, **1**).

modulation envelope See ENVELOPE, **1**.

modulation-envelope distortion Undesirable distortion in the ENVELOPE of the modulating intelligence in an amplitude-modulated or single-sideband signal at the output of a radio transmitter.

modulation factor See MODULATION COEFFICIENT.

modulation frequency Abbreviation, f_m. The frequency of a modulating signal.

modulation linearity In a modulated signal, the degree to which the instantaneous signal amplitude or frequency follows the instantaneous amplitude of the modulating signal.

modulation meter See PERCENTAGE-MODULATION METER.

modulation monitor **1.** A linear detector with a pickup coil (or antenna) and headphones for listening to an amplitude-modulated signal. **2.** See PERCENTAGE-MODULATION METER. **3.** A device that displays the modulation characteristics of a signal (e.g., the envelope for amplitude modulation and single sideband or the frequency-vs. time function for frequency modulation).

modulation noise See NOISE BEHIND THE SIGNAL.

modulation percentage A measure of the extent to which a signal is amplitude modulated. It is expressed as the MODULATION COEFFICIENT multiplied by 100.

modulation ratio For a modulated signal, the quotient M_r/M_i, where M_r is the percentage of radiated-signal modulation, and M_i is the percentage of current modulation.

modulator A device or circuit for modulating a carrier.

modulator cell See KERR CELL.

modulator crystal A transparent piezoelectric crystal to which a signal voltage can be applied to modulate a beam of polarized light passing through it.

modulator-demodulator Also called *modem.* **1.** A device that converts binary data to audio-frequency-shift-keyed (AFSK) analog signals, and vice-versa. It is commonly used to interface computers and terminals with telephone lines and radio transceivers. **2.** A circuit or device, such as a

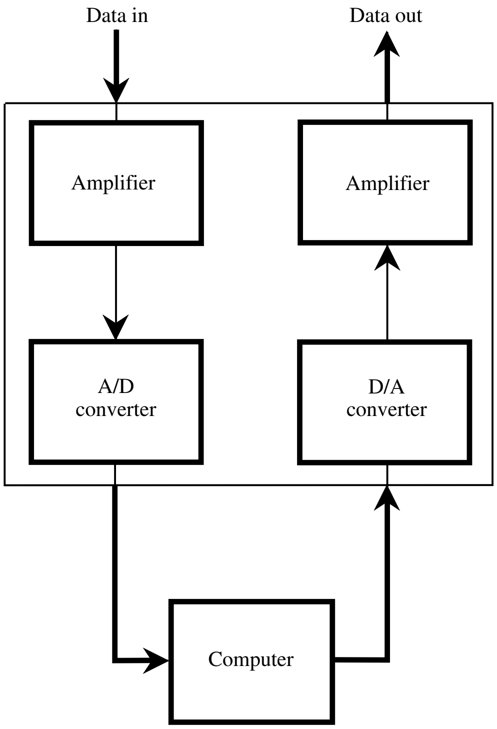

modulator-demodulator

biased diode or diode bridge, that can perform either modulation or demodulation.

modulator driver An amplifier stage that delivers excitation current, voltage, or power to a modulator stage.

module An assembly containing a complete self-contained circuit (or subcircuit), often miniaturized and made for plug-in operation.

modulometer Any instrument, such as a percentage-modulation meter, used to measure the degree of modulation of a signal. Often, it can also measure other signal characteristics (e.g., carrier shift, extraneous amplitude modulation, and extraneous frequency modulation).

modulo n check In computer operations, a technique for verifying the validity of a number used as an operand. The number being so checked is divided by another number to provide a remainder (check digit) that goes with the number. After the number is, for example, transmitted through some part of a computer system, it is again divided by the original divisor, and if the remainder is the check digit, the data has retained its integrity.

modulus **1.** Absolute magnitude. Also see ABSOLUTE VALUE and IMPEDANCE. **2.** Abbreviation, mod. In computer operations, a whole number that indicates the number of states a counter sequences through in each cycle. **3.** Abbreviation, mod. A number (constant or coefficient) expressing the degree to which some property is possessed by a material or body (e.g., *modulus of elasticity*, *shear modulus*, and *bulk modulus*). **4.** A constant by which a logarithm to one base must be multiplied to obtain a logarithm of the same number to another base.

modulus of elasticity The stress-to-strain ratio in a material under elastic deformation.

moire In a television or facsimile picture, an effect produced by the convergence of straight lines. When the lines are nearly parallel to the scanning lines, the converging lines appear irregular.

moisture meter See ELECTRIC HYGROMETER and ELECTRONIC HYGROMETER.

mol Abbreviation of MOLE.

molar conductance See MOLECULAR CONDUCTANCE.

molar polarization Any molecule in an electric field undergoes a small displacement of the positive and negative electrical centers. This results in an electric dipole.

molar solution A solution, such as an electrolyte, containing 1 mol of solute per liter of solvent. Compare NORMAL SOLUTION.

mold **1.** To form matter into a desired shape, as by pouring liquefied material into a container or liquefying the material in the container, then allowing the liquid material to solidify. In *hot molding*, the material is melted in the container and then cooled to hardness; in *cold molding*, the material is shaped without heat and it solidifies with time. **2.** The hollow container used to shape a material, as in **1**.

molded capacitor A capacitor that is molded into a protective body of insulating material. Also see MOLD, **1**, **2** and MOLDED COMPONENT.

molded ceramic capacitor A ceramic-dielectric capacitor enclosed in a molded housing. Also see MOLD, **1**, **2**; MOLDED CAPACITOR; and MOLDED COMPONENT.

molded coil See MOLDED INDUCTOR.

molded component A part (such as a capacitor, coil, or resistor) that is completely enclosed in a protective material (such as a plastic) that is molded around it. Also see MOLD, **1**, **2**.

molded electrolytic capacitor A solid-dielectric electrolytic capacitor enclosed in a molded housing. Also see MOLD, **1**, **2**; MOLDED CAPACITOR; MOLDED COMPONENT; and SOLID-ELECTROLYTIC CAPACITOR.

molded glass capacitor A glassplate-dielectric capacitor enclosed in a molded glass housing. Also see GLASS CAPACITOR; MOLD, **1**, **2**; MOLDED CAPACITOR; and MOLDED COMPONENT.

molded inductor An inductor that is molded into a protective housing of insulating material. Also see MOLD, **1**, **2** and MOLDED COMPONENT.

molded mica capacitor A mica-dielectric capacitor enclosed in a molded housing. Also see MICA CAPACITOR; MOLD, **1**, **2**; MOLDED CAPACITOR; and MOLDED COMPONENT.

molded mud A molding compound that has inferior electrical characteristics. Also see MOLD, **1**, **2** and MOLDED COMPONENT.

molded paper capacitor A paper-dielectric capacitor enclosed in a molded housing. Also see MOLD, **1**, **2**; MOLDED CAPACITOR; MOLDED COMPONENT; and PAPER CAPACITOR.

molded porcelain capacitor A capacitor enclosed in a body of molded porcelain. Also see MOLD, **1**, **2**; MOLDED CAPACITOR; MOLDED COMPONENT; and PORCELAIN.

molded resistor A resistor that is molded in a protective housing of insulating material. Also see MOLD, **1**, **2** and MOLDED COMPONENT.

molded transistor A transistor that is encapsulated in a protective molding compound, such as epoxy resin. Also see MOLD, **1**, **2** and MOLDED COMPONENT.

mole Abbreviation, mol. **1.** The amount of substance in a system containing as many specified entities (atoms, molecules, ions, subatomic particles, or groups of such particles) as there are atoms in 12 grams of carbon 12. **2.** It is also called the *Avogadro constant*. A unit of quantity in chemistry, equal to approximately 6.022×10^{23}.

molectronics See MOLECULAR ELECTRONICS.

molecular circuit See MONOLITHIC INTEGRATED CIRCUIT.

molecular conductance For a solution, such as an electrolyte, the product of specific conductivity and the volume (in liters) of a solution that contains 1 gram molecule of the solute. Also see SOLUTE; SOLUTION, **1**; and SOLVENT, **1**, **2**.

molecular conductivity See MOLECULAR CONDUCTANCE.

molecular electronics The technique of processing a single block of material so that separate areas perform the functions of different electronic components. The entire block constitutes a circuit (e.g., a MONOLITHIC INTEGRATED CIRCUIT).

molecular magnets According to the molecular theory of magnetism, the elemental magnets formed by individual molecules.

molecular theory of magnetism Each molecule in a piece of magnetic metal is itself a magnet (possessing a north and a south pole). These tiny magnets are thought to be normally oriented at random, but when the material is magnetized by an external force, they align themselves with each other.

molecular weight Abbreviation, mol wt. In a molecule of a substance, the sum of the atomic weights of the constituent atoms.

molecule The basic particle of a compound; each molecule usually contains two or more atoms. For example, the formula $AgNO_3$ represents silver nitrate, each molecule of which contains one atom of silver (Ag), one atom of nitrogen (N), and three atoms of oxygen (O).

moletronics See MOLECULAR ELECTRONICS.

mol wt Abbreviation of MOLECULAR WEIGHT.

molybdenum Symbol, Mo. A metallic element. Atomic number, 42. Atomic weight, 95.94. It is used in the grids and plates of certain vacuum tubes.

moment The tendency to produce motion around a point, as by torque, or the product of a quantity and the distance to a point. The moment of force is expressed as the product *Fd*, where *F* is force and *d* is distance.

momentary-contact switch A switch that maintains contact only while it is held down. Such a device is usually a *pushbutton switch*, although it might be a *toggle switch*, a *slide switch*, or a *lever switch.*

momentary switching Switching of short duration, often characterized by a quick make and break immediately following activation of the switch. Compare DWELL SWITCHING.

moment of inertia For a torque motor, the inertia of the armature around the axis of rotation. Also see MOMENT.

mon **1.** Abbreviation of MONITOR. **2.** Abbreviation of MONAURAL.

monatomic **1.** Pertaining to a molecule with only one atom. **2.** Pertaining to a molecule with only one replaceable atom or radical.

monatomic molecule A molecule having a single atom (e.g., argon, helium, and neon). Compare DIATOMIC MOLECULE.

monaural **1.** Pertaining to an audio system having one channel. **2.** Pertaining to hearing with one ear, as opposed to BINAURAL.

monaural recorder A single-track recorder, as opposed to a stereophonic recorder.

Monel metal An alloy of nickel (67%), copper (28%), iron, manganese, and other metals (5%). Its resistivity is approximately 42 microhm-centimeters at 20°C.

monimatch An amateur version of the reflected-power meter and SWR meter.

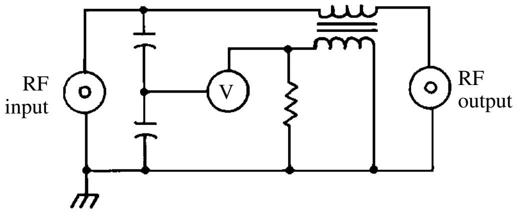

monimatch

moniscope A special cathode-ray tube that produces a stationary picture for testing television equipment. Its name is a contraction of monitor and scope.

monitor **1.** A device that allows the sampling of a signal or quantity. Examples: *line-voltage monitor*, *television monitor*, and *modulation monitor.* **2.** A cathode-ray-tube (CRT) computer display.

monitor head A separate playback head included in some tape recorders for listening to the tape as it is being recorded.

monitoring The act, process, or technique of observing an action while it is in progress or checking a quantity while it is varying. Examples: *carrier monitoring*, *modulation monitoring*, and *line-voltage monitoring.*

monitoring amplifier An auxiliary amplifier used in monitoring an audio-frequency system.

monitoring antenna A usually small pickup antenna used with a signal monitor or monitoring receiver.

monitoring key In a telephone system, a device used to listen to a two-way conversation.

monitoring receiver A radio or television receiver used specifically to monitor a transmission directly.

monitoring station In a security system, a central control location from which personnel can observe the input from sensors, cameras, and other devices at remote locations throughout the secured area.

monitor system A computer program usually stored in the read-only memory (ROM) supplied by the hardware vendor. It controls the implementation of programs written by the user, and the operation of peripherals associated with program runs and inputting or outputting data to or from memory. Also called *executive program.*

monkey chatter The characteristic sound of a single-sideband (SSB) signal when the receiver is mistuned, or when the receiver is set for the wrong sideband.

monk's cloth A coarse drapery fabric sometimes used to soundproof the walls and ceiling of a radio studio or recording booth.

mono See MONAURAL.

mono- Prefix meaning *single.*

monobrid circuit An integrated circuit in which either several monolithic IC chips are interconnected to form a larger, single-package circuit, or monolithic IC chips are interwired with thin-film components into a single-package circuit. The name is a contraction of *monolithic hybrid.*

monochromatic **1.** Being of one color (hue) in nature. **2.** Being of a single wavelength in nature (pertaining to radiation of any kind). **3.** Pertaining to black-and-white television.

monochromaticity Consisting of one color of visible light. The brightness can vary from black to maximum.

monochromatic power density At a given temperature, the energy radiated per square centimeter of blackbody surface per second per unit wavelength range. Also see BLACKBODY and BLACKBODY RADIATION.

monochromatic sensitivity Sensitivity to light of one color only.

monochrome television Black-and-white television.

monoclinic crystal A crystal having three axes of unequal length; two of them intersect obliquely and are perpendicular to the third [e.g., the type of crystal found in one form of sulfur (monoclinic sulfur)].

monocrystalline material See SINGLE-CRYSTAL MATERIAL.

monode A one-element device, such as a filament-type lamp, thermistor, voltage-dependent resistor, barretter, etc.

monogroove stereo A method of making a stereophonic phonograph disc in which both channels are recorded as a single groove.

monolayer A thin film having a thickness of one molecule.

monolithic integrated circuit An integrated circuit (IC) formed in a single block or wafer of semiconductor material. The name is derived from the Greek monolithos ("one stone"). Compare HYBRID INTEGRATED CIRCUIT and THIN-FILM INTEGRATED CIRCUIT.

monometallic Containing or using only one metal.

monomolecular film See MONOLAYER.

monophonic recorder See SINGLETRACK RECORDER.

monophonic system A single-channel sound system. Compare STEREO SYSTEM.

monopole antenna See QUARTER-WAVE MONOPOLE.

monopulse In radar and electronic-navigation operations, using one pulse to determine azimuth and elevation simultaneously.

monorange speaker A loudspeaker that reproduces most of the full audio range. Also called *extended range speaker.* Compare TWEETER and WOOFER.

monostable Having one stable state.

monostable blocking oscillator Abbreviation, MBO. A blocking oscillator that behaves somewhat like a one-shot multivibrator. The oscillator delivers a single output pulse each time it receives an input (trigger) pulse.

monostable multivibrator A multivibrator that delivers one output pulse for each input (trigger) pulse. Also called *one-shot circuit* and *single-shot multivibrator.* Compare ASTABLE MULTIVIBRATOR and BISTABLE MULTIVIBRATOR.

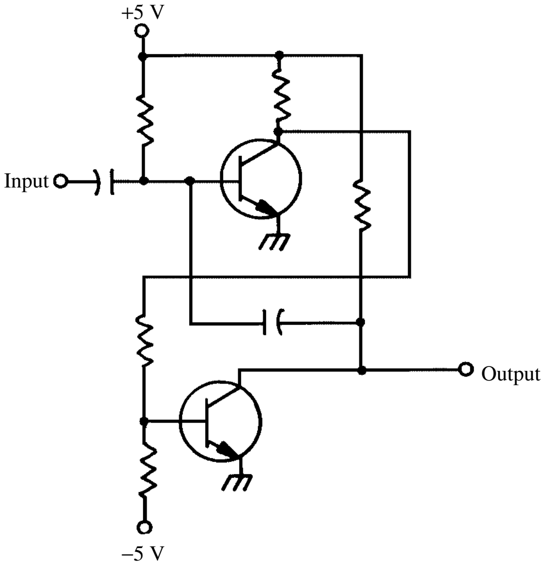

monostable multivibrator

monostatic reflectivity The property whereby, for certain reflectors (such as a tricorner reflector), all incident rays are reflected in exactly the opposite direction from which they arrive.

monotone A sound or series of sounds having a single, constant pitch (frequency).

monotonicity In a digital-to-analog converter, a condition where the output either remains the same or increases for any single increase in the input code.

monovalent See UNIVALENT.

Monte Carlo method **1.** The use of statistical sampling in the approximate solution of an engineering problem. **2.** In computer operations, the construction of mathematical models from randomly selected components taken from representative statistical populations.

Moog synthesizer An electronic device that can be made to simulate virtually any sound, including that of musical instruments and the human voice, through the use of several audio oscillators, whose output can be controlled to produce tones of various harmonic content, duration, attack, and decay periods.

moonbounce Also called *earth-moon-earth (EME).* Radio communication, usually at very-high frequencies (VHF) or ultra-high frequencies (UHF), in which the moon is used as a passive reflector. This is a popular mode among some amateur radio operators.

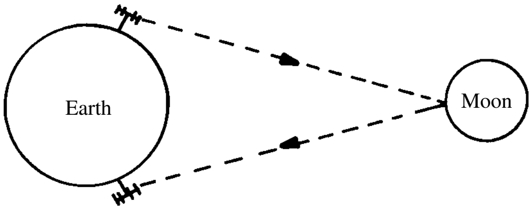

moon bounce

MOPA Abbreviation of MASTER OSCILLATOR/POWER AMPLIFIER.

morphological electronics See MOLECULAR ELECTRONICS.

Morse **1.** See MORSE CODE. **2.** Telegraphy (wire or radio). **3.** To signal by means of the Morse code.

Morse code Either of two similar binary codes used in radio and wire telegraphy. It uses short pulses (dots or dits) and long pulses (dashes or dahs) to represent letters of the alphabet, numerals, and punctuation marks. It usually refers to the CONTINENTAL CODE, but occasionally it refers to the AMERICAN MORSE CODE.

MOS Abbreviation of *metal-oxide-semiconductor.*

mosaic **1.** See PHOTOMOSAIC. **2.** The pattern of tiny photoelectric particles in a television camera tube that convert the image into electric charges.

mosaic crystal A form of imperfect crystal. The defects have certain properties, one of which is to cause additional energy levels in semiconductor materials manufactured from such crystals.

MOS capacitor A capacitor using metal-oxide-semiconductor (MOS) technology. It is used in MOS integrated circuits. A silicon substrate forms one electrode (generally p-type material). An oxide layer forms the dielectric. An electrode forms the other plate of the capacitor. The capacitance might be variable by changing the applied voltage at the metal gate electrode.

MOSFET Abbreviation of METAL-OXIDE-SEMICONDUCTOR FIELD-EFFECT TRANSISTOR.

MOSROM Abbreviation of *metal-oxide-semiconductor read-only memory* (see READ-ONLY MEMORY).

MOST Abbreviation of *metal-oxide-semiconductor transistor.*

most significant character Abbreviation, MSC. In positional number representation, the leftmost character in a significant group, such as a word.

most significant digit Abbreviation, MSD. In a number, the leftmost digit that is not zero (zero being insignificant in this context).

mother **1.** A term or prefix referring to a component that supports (or appears to support) other similar components (e.g., MOTHERBOARD). **2.** A term referring to the source from which samples or components are derived (e.g., MOTHER CRYSTAL). **3.** A mold that has been electroformed from a master phonograph disc.

motherboard In a computer or data-processing device, the circuit board on which most of the main circuitry is mounted.

mother crystal A natural quartz crystal from which is produced the piezoelectric plates and other components used in electronics.

motion detector A device for sensing the movement or stopping of a body, such as a rotating shaft. Various sensors are used in different detectors: magnetic, photoelectric, capacitive, etc.

motion frequency The natural frequency (especially that of oscillation) of a servo.

Motion Picture Experts Group Abbreviation, MPEG. A data-compression standard used for animated digital video. It is useful in computer videoconferencing and telecommuting, and for the development of multimedia presentations.

motion-picture pickup In television operations, a camera (and the technique for using it) for picking up scenes directly from motion-picture film.

motion sensor In security systems, a set of devices that produces an output signal whenever anything moves within a certain area. Such equipment might use infrared, ultrasound, microwaves, capacitive effects, thermal sensors, air-current detectors, sound detectors, video cameras, or a combination of these.

motor **1.** A machine for converting electrical energy into mechanical energy. **2.** The driving mechanism of a loudspeaker.

motorboard The basic mechanism of a tape recorder, embodying motor, flywheel, capstan, rollers, etc., assembled on a board or panel.

motorboating A repetitive, low-frequency popping or puffing noise sometimes observed in malfunctioning audio circuits. It was so named because of its resemblance to the sound of a motorboat.

motorboating filter In an audio amplifier circuit, a simple filter installed to prevent

MOTORBOATING caused by unwanted positive feedback. Also see DECOUPLING FILTER.

motor capacitor See MOTOR-RUN CAPACITOR and MOTOR-START CAPACITOR.

motor constant The ability, expressed numerically, of a torque motor to convert electric input power into torque.

motor converter An electromechanical device for converting alternating current to direct current.

motor-driven relay An electromechanical relay whose contacts are opened and closed by a rotating motor.

motor effect Magnetic force between adjacent current-carrying conductors.

motor-generator A combination of motor and generator in a single machine assembly. A common arrangement is a low-voltage motor turning a high-voltage generator. The machine shafts can be coupled, or the motor can turn the generator through a belt, chain, or gear train.

motor meter A meter in which the movable element is essentially a continuously rotating motor. See, for example, KILOWATT-HOUR METER.

motor-run capacitor A power-factor-boosting capacitor connected (together with its auxiliary winding) in parallel with the main winding in an induction motor. Also see CAPACITOR MOTOR.

motor-speed control A method of controlling the speed of a motor by varying the magnitude and/or phase of its current. Electronic devices usually use diode, diac, or triac circuits.

motor-start capacitor A capacitor that, with an auxiliary winding, is switched into a motor circuit during the starting period, and is automatically disconnected (with the winding) after the motor reaches normal running speed. Also see CAPACITOR MOTOR.

motor starter See STARTING BOX.

mould See MOLD.

mount **1.** A mechanical device with which a component is attached to a circuit board or chassis. **2.** The attachment of a circuit board to a chassis. **3.** The hardware with which an antenna is attached to a mast. **4.** The hardware with which a microphone is attached to a boom or other support. **5.** In general, any attaching hardware.

Mountain standard time Abbreviation, MST. Local mean time at the 105th meridian west of Greenwich. Also see GREENWICH MEAN TIME, STANDARD TIME, TIME ZONE, and COORDINATED UNIVERSAL TIME.

mounting flange The portion of a speaker frame, usually made of metal, that is equipped with holes for attachment to a cabinet or panel.

mouse A pointing device commonly used with personal computers. It is about the size of a candy bar. The computer operator moves the display pointer or cursor by sliding the mouse on a flat surface. Functions and commands are executed by pressing a button (clicking) on the device.

mouth The radiating end of a horn (antenna, loudspeaker, etc.).

MOV Abbreviation of METAL-OXIDE VARISTOR.

movable-coil meter See D'ARSONVAL METER.

movable-coil modulator See ELECTROMECHANICAL MODULATOR.

movable contact The traveling contact in a relay or switch. Compare FIXED CONTACT.

movable-iron meter See IRON-VANE METER.

movies on demand In television, a system that makes it possible for subscribers to choose the programs they want to see, and also to choose the viewing times for the programs.

moving-coil galvanometer A galvanometer whose movable element is a coil of fine wire suspended or pivoted between the poles of a magnet.

moving-coil microphone See DYNAMIC MICROPHONE.

moving-coil motor The driving mechanism of a moving-coil DYNAMIC SPEAKER.

moving-coil pickup See DYNAMIC PICKUP.

moving-coil speaker See DYNAMIC SPEAKER.

moving-conductor microphone See VELOCITY MICROPHONE.

moving-diaphragm meter A headphone used as a sensitive indicator in alternating-current bridge measurements.

moving element In an electromechanical device, the portion that moves physically under variable operating conditions.

moving-film camera An oscilloscope camera in which the film is drawn past the lens continuously at a constant speed, rather than being advanced frame by frame, as in a motion-picture camera.

moving-iron meter See IRON-VANE METER.

moving-vane meter See IRON-VANE METER.

mp Abbreviation of MELTING POINT.

m_p Symbol for MASS OF PROTON AT REST.

MPEG Abbreviation for MOTION PICTURE EXPERTS GROUP.

MPG Abbreviation of *microwave pulse generator.*

mph Abbreviation of *miles per hour.* Also, mi/h.

MPO Abbreviation of *maximum power output.*

mps **1.** Abbreviation of *meters per second.* Also, m/s. **2.** Abbreviation of *miles per second.* Also, mi/s.

MPT Abbreviation of MAXIMUM POWER TRANSFER.

MPX Abbreviation of MULTIPLEX.

MR Abbreviation of MEMORY REGISTER.

mrad Abbreviation of MILLIRADIAN.

MRIA Abbreviation of *Magnetic Recording Industry Association.*

mrm Abbreviation of *milliroentgens per minute.*

MS Abbreviation of MASS SPECTROMETER.

m^2/s Abbreviation of *meters squared per second.* The unit of kinematic viscosity.

M scan In radar operations, a modified A-scan display in which a pedestal signal is moved along the base line to a point where it coincides with the base line of the reflected signal to determine the distance to the target.

msec Abbreviation of MILLISECOND. Also, ms.

msg Abbreviation of MESSAGE.

MSI Abbreviation of MEDIUM SCALE INTEGRATION.

MST Abbreviation of MOUNTAIN STANDARD TIME.

MT Abbreviation of METRIC TON.

MTR Abbreviation of MAGNETIC TAPE RECORDER.

mtr Alternate abbreviation of METER.

MTS Abbreviation for MULTICHANNEL TELEVISION SOUND.

M-type backward-wave oscillator A broadband, voltage-tuned oscillator in which the electrons interact with a backward traveling radio-frequency wave. Compare O-TYPE BACKWARD-WAVE OSCILLATOR.

mu Symbol, μ. **1.** Abbreviation of the prefix MICRO-. **2.** Expression for AMPLIFICATION FACTOR. **3.** Expression for PERMEABILITY. **4.** Expression for MICRON. **5.** Expression for ELECTRIC MOMENT. **6.** Expression for INDUCTIVITY. **7.** Expression for MAGNETIC MOMENT. **8.** Expression for MOLECULAR CONDUCTIVITY.

μ **1.** Symbol for MU.

μ**A** Abbreviation of MICROAMPERE.

μ_B Symbol for BOHR MAGNETON.

μ**b** Abbreviation of MICROBAR.

μ**Ci** Abbreviation of MICROCURIE.

μ_e Symbol for ELECTRON MAGNETIC MOMENT.

μ**F** Abbreviation of MICROFARAD.

μ**g** Abbreviation of MICROGRAM.

μ**H** Abbreviation of MICROHENRY.

μ**in** Abbreviation of MICROINCH.

μ**l** Abbreviation of MICROLITER.

μ**mm** Abbreviation of *micromillimeter* (see NANOMETER).

μ_n Symbol for NUCLEAR MAGNETON.

μ_o Symbol for FREE-SPACE PERMEABILITY CONSTANT.

μ_p Symbol for PROTON MAGNETIC MOMENT.

μ**P** Abbreviation of MICROPROCESSOR.

μ**rd** Abbreviation of MICRORUTHERFORD.

μ**S** Abbreviation of MICROSIEMENS.

μ**s** Abbreviation of MICROSECOND. (Also, μsec.)

μ**V** Abbreviation of MICROVOLT.

μ**V/m** Abbreviation of MICROVOLTS PER METER.

μ**W** Abbreviation of MICROWATT.

MUF Abbreviation of MAXIMUM USABLE FREQUENCY.

mu factor See MU, **2**.

mu law In communications, a companding standard generally used in the United States and Canada.

Muller tube A tube that conducts by means of ionization of an internal gas. It is used for radiation detection.

multiaddress In a computer, a multiple memory location.

multiband amplifier A radio-frequency power amplifier capable of operation on more than one frequency band. The bands are often (but not always) harmonically related. Compare WIDEBAND AMPLIFIER.

multiband antenna A single antenna, such as one used in amateur radio and television reception, operated in several frequency bands or channels. The bands are often (but not always) harmonically related.

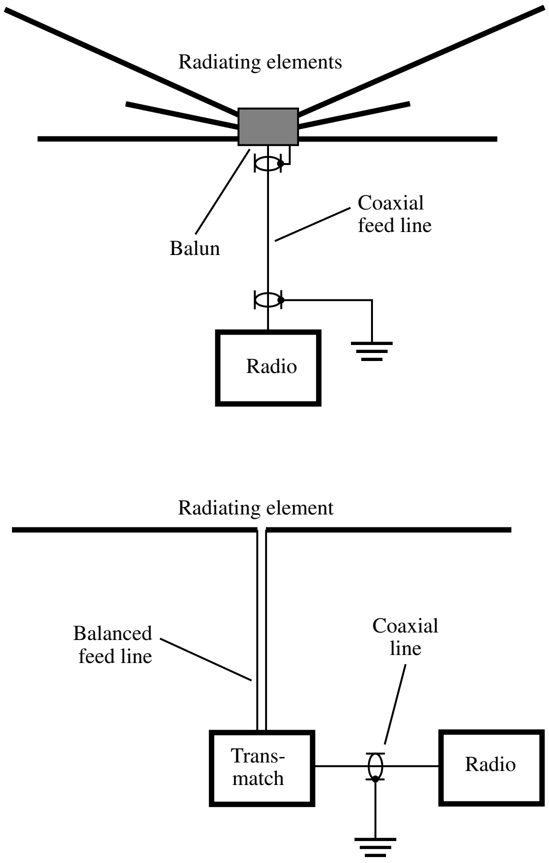

multiband antenna

multiband device A device (such as a tuner, receiver, transmitter, or test instrument) that operates in several selectable frequency bands.

multiband oscilloscope See WIDEBAND OSCILLOSCOPE.

multiband receiver A communications receiver capable of operation on more than one frequency band. Such a receiver might, but does not necessarily, provide continuous coverage over a wide range of frequencies. Compare WIDEBAND RECEIVER.

multicasting The use of two frequency-modulation (FM) broadcast stations, or one FM station and one television station, to broadcast separately the two channels of a stereo program. The program is picked up simultaneously with two receivers.

multicavity magnetron A magnetron whose anode block has two or more cavities.

multicellular horn A loudspeaker in front of which are placed rectangular cells through which the sound passes.

multichannel Pertaining to a radio-communication system that operates on more than one channel at the same time. The individual channels might contain identical information, or they might contain different signals.

multichannel analyzer A test instrument, such as a spectrum analyzer, that splits an input into several channels for testing.

multichannel television sound Abbreviation, MTS. In television broadcasting, audio transmitted on more than one channel to provide stereo sound to subscribers.

multichip circuit A MICROCIRCUIT composed of interconnected active and passive chip-type components.

multichip integrated circuit An INTEGRATED CIRCUIT composed of circuit elements on separate, interconnected chips.

multicontact switch A switch having more than two contacting positions.

multicoupler An impedance-matching device used to couple several receivers to a single antenna.

multielement antenna A directive antenna having more than one element. Such antennas include phased arrays and parasitic arrays. Common examples are the LOG-PERIODIC ANTENNA, the QUAD ANTENNA, and the YAGI ANTENNA.

multiemitter transistor 1. A bipolar transistor having more than one emitter. **2.** A bipolar power transistor having several emitters connected in parallel in the transistor structure.

multiframe A set of frames in a signal that are adjacent to each other in time (consecutive) and positioned according to an alignment signal.

multigun CRT A cathode-ray tube (CRT) having more than one electron gun.

multihop propagation Propagation of a radio wave by several successive reflections between the ionosphere and the surface of the earth.

multilayer circuit A circuit consisting of several sections printed or deposited on separate layers, which are subsequently stacked in a sandwich-like manner.

multilayer coil A coil in which the turns of wire are wound in several complete layers, one on top of the other. Compare SINGLE-LAYER COIL.

multilevel password protection See HIERARCHICAL PASSWORD PROTECTION.

multimedia 1. In computing, the use of video, voice, music, electromechanical control, and/or data transfer at the same time. Useful in education, entertainment, business, and gaming applications. **2.** The use of voice, images, data, and/or video in a communications system or network, and, in particular, in wireless applications. An example is videoconferencing between a corporate office and the user of a handheld computer equipped with a wireless modem.

multimedia computer A personal computer designed especially for multimedia use. It includes a sound board, speakers, a microphone, and a CD-ROM (compact-disk read-only memory) drive. It often has a large-screen, high-resolution monitor (17 inches or greater diagonal measure).

multimeter A meter that allows measurement of different quantities (e.g., current, voltage, and resistance); the functions are usually made available through a selector switch.

multimode operation 1. In radio communication, the use of two or more transmitters simultaneously, operating in different modes [e.g., one using single-sideband (SSB) and another using frequency modulation (FM)]. **2.** The transmission of visible light or infrared energy through an optical fiber in more than one mode at the same time. **3.** The operation of any device in more than one of its modes simultaneously.

multipactor A microwave switching tube capable of operating at high power levels. Characterized by high operating speed.

multipath cancellation A phenomenon sometimes observed in radio-wave propagation. Separate signal components arrive at the receiver in equal amplitude, but opposite phase. Also see MULTIPATH FADING and MULTIPATH RECEPTION.

multipath delay In MULTIPATH RECEPTION, the lag between signal components arriving over different paths.

multipath effect At a receiver, the difference in arrival time of multipath signals. Also see MULTIPATH DELAY.

multipath fading In radio communication, variations in the received signal that result from phasing various components of the transmitted signal that are propagated over different paths. At low, medium, and high frequencies, this effect is usually the result of ionospheric fluctuation. At very-high and ultra-high frequencies, it could occur as a result of changes in the state of the intervening atmosphere or, occasionally, reflection from objects, such as aircraft.

multipath reception Reception of a signal over more than one path. This often results in fading (see MULTIPATH FADING) and other undesirable effects, such as ghosting in television reception.

multipath transmission 1. Transmission of a signal over two or more paths. Also see MULTIPATH RECEPTION. **2.** See MULTIHOP PROPAGATION.

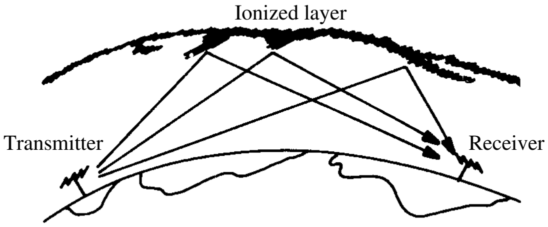

multipath reception

multiphase system See POLYPHASE SYSTEM.

multiple access system A timesharing data processing system that can be used by a number of people at remote locations, usually through peripherals other than conventional terminals (e.g., cash registers linked to a computer system for inventory control).

multiple-address code In computer operations, an instruction code that requires the user to specify the address of more than one operand per instruction.

multiple address instruction A computer program instruction specifying the address of more than one operand.

multiple break **1.** Interruption of a circuit at several points. **2.** Contact bounce.

multiple-chip circuit See HYBRID INTEGRATED CIRCUIT.

multiple connector In a flowchart, a symbol showing the merging of several flowlines.

multiple ionization Successive ionization, as when an ion repeatedly collides with electrons. Also see DOUBLE IONIZATION.

multiple-length number In computer operations, a number occupying more than one register.

multiple-loop feedback system A feedback system using more than one feedback loop. An example is a tunable audio amplifier (see PARALLEL-TEE AMPLIFIER) in which a negative-feedback path provides tuning, while a positive-feedback path sharpens selectivity.

multiple modulation See COMPOUND MODULATION.

multiple precision arithmetic An arithmetic process using several words (bit groups) for an operand to ensure accuracy.

multiple programming **1.** Computer operation in which several programs in memory share peripherals and processor time. **2.** Programming a computer so that several logic or arithmetic operations can be carried out at the same time.

multiple-purpose meter See MULTIMETER.

multiple-purpose tester A multimeter sometimes combined with some other instrument, such as a test oscillator.

multiple reel file In a data-processing system, a magnetic tape data file of more than one reel.

multiple speakers A group of more than two loudspeakers, usually operated from a single amplifier system.

multiple-stacked broadside array A stacked array in which a number of collinear elements are stacked above and below each other. Also see BROADSIDE ARRAY and COLLINEAR ANTENNA.

multiple station Descriptive of a communications network in which more than one terminal is used.

multiple-unit steerable antenna A shortwave antenna system intended to prevent or minimize the effects of fading in received signals. It consists principally of a number of rhombic antennas feeding the receiver. The system utilizes the various waves arriving at different angles. An electric steering system causes the antennas to be selected automatically in the best combination and with their outputs in the proper phase. Also see DIVERSITY RECEPTION.

multiple winding See DRUM WINDING.

multiplex Pertaining to a communications or broadcasting system in which *multiplexing* is used. See MULTIPLEXING, **1**, **2**, **3**.

multiplex adapter A special circuit (or auxiliary unit) used in frequency-modulated radio receivers for stereophonic reception from a station transmitting a multiplex broadcast. Also see MULTIPLEXING, **1** and MULTIPLEX STEREO.

multiplex code The transmission of multiple signals over a single medium, using a code (such as Morse, Baudot, or ASCII). See MULTIPLEXING, **2**, **3**.

multiplex data terminal A computer terminal acting as a modem by virtue of its accepting and transferring signals between its input/output devices and a data channel.

multiplexer A device that allows two or more signals to be transmitted simultaneously on a single carrier wave, communications channel, or data channel. See MULTIPLEXING, **2**, **3**.

multiplexing **1.** A process in which a comparatively low-frequency carrier is modulated, then mixed with a signal that has a higher frequency. Also see SUBCARRIER and SUBSIDIARY COMMUNICATION AUTHORIZATION. **2.** The simultaneous transmission of numerous relatively low-frequency signals on a single carrier having a higher frequency. Also called *frequency-division multiplexing.* **3.** The transmission of numerous signals on a single channel by breaking each signal into timed fragments (intervals) and transmitting the fragments in a rotating sequence. Also called *time-division multiplexing.* **4.** In certain digital light-emitting-diode (LED) displays, the lighting of various parts of the display in a rapidly rotating sequence.

multiplex stereo The use of *multiplexing* to broadcast both channels of a stereophonic program on a single carrier wave. See MULTIPLEXING, **1**, **2**.

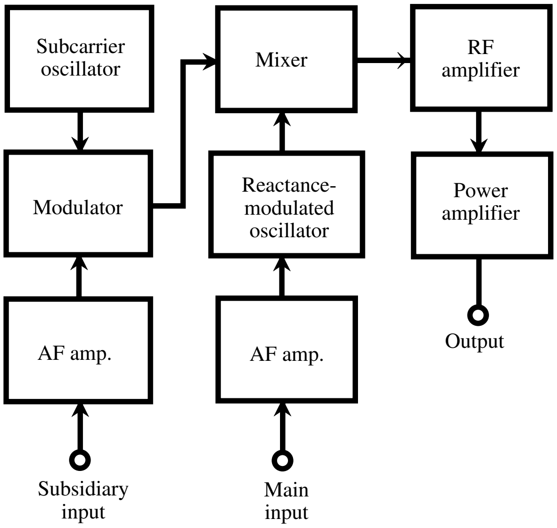

multiplexing, 1.

multiplex telegraphy **1.** A system of wire telegraphy in which two or more messages are sent simultaneously in one or both directions. **2.** A system of radiotelegraphy in which two or more messages can be sent simultaneously on the same carrier wave.

multiplex telephony **1.** A system of wire telephony in which two or more messages can be sent simultaneously in one or both directions over the same line. **2.** A system of radiotelephony in which two or more messages can be sent simultaneously on the same carrier wave.

multiplication **1.** The arithmetic process whereby a certain factor is added to itself the number of times indicated by another factor (the multiplier). **2.** A method of increasing a quantity, magnitude, or rate by some desired factor. See, for example, FREQUENCY MULTIPLIER and VOLTAGE MULTIPLIER.

multiplier **1.** See FREQUENCY MULTIPLIER. **2.** See VOLTAGE MULTIPLIER. **3.** A circuit or device for performing arithmetic multiplication. **4.** See VOLTMETER MULTIPLIER.

multiplier amplifier A frequency-multiplying amplifier (such as a doubler, tripler, or quadrupler), whose output circuit is tuned to an integral multiple of the input frequency. Compare STRAIGHT-THROUGH AMPLIFIER.

multiplier phototube See PHOTOMULTIPLIER TUBE.

multiplier prefix A prefix that, when affixed to the name of a quantity (e.g., Hertz and byte) indicates the amount, usually a power of 10 or a power of 2, by which that quantity is to be multiplied. Also called *prefix* and *prefix multiplier.* See PREFIXES table in Appendix B. Also see, for example, GIGA-, KILO-, MEGA-, MICRO-, MILLI-, NANO-, and PICO-.

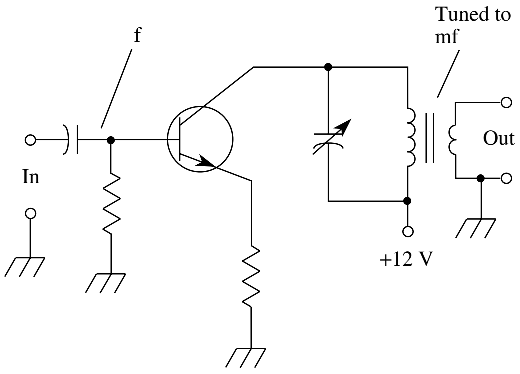

multiplier amplifier

multiplier probe **1.** A resistor-type test probe for a voltmeter, which increases the range of the meter and, therefore, acts as a voltmeter multiplier. **2.** A voltage-multiplier type of test probe for a voltmeter, which multiplies the test voltage before it is applied to the instrument (e.g., voltage-doubler probe).

multiplier register In a computer, a register that retains a number during a multiplication operation.

multiplier resistor A resistor connected in series with a current meter (usually a milliammeter or microammeter) to make it a voltmeter. Also called VOLTMETER MULTIPLER.

multiplier tube See PHOTOMULTIPLIER TUBE.

multipoint circuit A transmission circuit or system, in which information can be entered or retrieved at two or more locations.

multipoint distribution service In radio broadcasting, the transmission of programs via microwave links from a central facility to various local stations.

multipolar Pertaining to the presence of more than two magnetic poles.

multipolar machine A motor or generator having a number of poles in its armature and field.

multiposition relay A relay having more than two positions of closure. See, for example, SELECTOR RELAY.

multiposition switch A switch having more than two contacting positions. See, for example, SELECTOR SWITCH and STEPPING SWITCH.

multiprocessing **1.** In a computer system, the running of more than one program at once. **2.** The use of more than one microprocessor simultaneously in a single computer. Also called *parallel processing.* **3.** See TIME SHARING.

multiprogramming Also called *multitasking.* In a computer system, a technique that allows two or more programs to be executed at the same time.

multipurpose meter An electric meter that performs several functions, usually available through a function selector. Examples: VOLTAMMETER, VOLT-OHMMETER, and VOLT-OHM-MILLIAMMETER.

multirange instrument An instrument provided with several separate ranges, usually selectable by range switching (e.g., five-range voltmeter, two-scale ammeter, and four-band oscillator).

multisection filter A filter having two or more selective sections connected in cascade.

multisegment magnetron See MULTICAVITY MAGNETRON.

multiskip propagation See MULTIHOP PROPAGATION.

multistage device A device having several stages operating in cascade or otherwise coordinated with each other (e.g., a five-stage amplifier).

multistage feedback Feedback (positive or negative) between several stages in a system, as opposed to feedback between the output and input of a single stage.

multistage oscillation Oscillation resulting from positive feedback between or among two or more stages of an amplifier chain, as opposed to oscillation occurring between the output and input of a single amplifier stage.

multistage X-ray tube An X-ray tube providing electron acceleration by means of successive ring-shaped anodes—each biased to a higher voltage than the preceding one.

multiswitch A switch having a number of poles and contacts.

multitester An instrument, such as a multimeter or a combined signal generator and oscilloscope, that performs a number of different test functions.

multitrack recording **1.** A recording on two or more tracks (e.g., *multitrack disk* and *multitrack tape*). **2.** Making a recording on a tape or disc with two or more tracks.

multiturn loop antenna A small loop antenna (whose diameter measures less than approximately 10 degrees of phase in free space) that has two or more turns of wire or tubing. The loop has significant inductance. A capacitor, connected in parallel or in series with the loop, can produce a narrowband frequency response.

multiturn potentiometer A potentiometer whose shaft must be rotated through several complete revolutions to cover the full resistance range.

multivalent Having a valence greater than 2. Compare UNIVALENT.

multivibrator A circuit usually containing two transistors in a resistance-capacitance (RC) coupled amplifier, whose output is capacitance-coupled to the input. The two stages switch each other alternately in and out of conduction at a frequency determined by the *R* and *C* values. Also see ASTABLE MULTIVIBRATOR, BISTABLE MULTIVIBRATOR, and MONOSTA-BLE MULTIVIBRATOR.

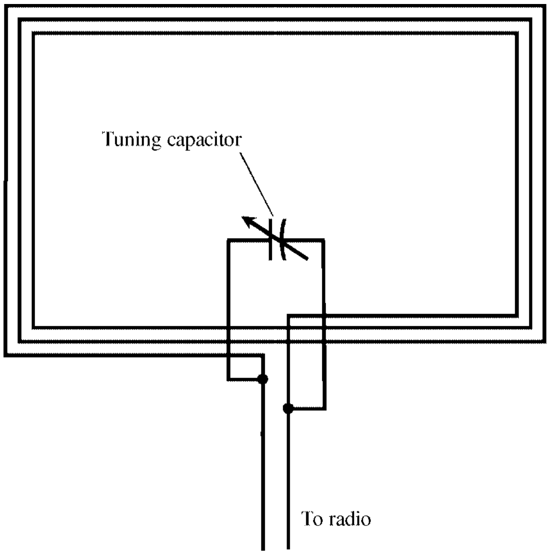

multiturn loop antenna

multiwire antenna An antenna having more than one wire in the radiating section. In early flat-top antennas, such wires were usually connected together at one end.

multiwire doublet antenna A doublet antenna having more than one wire in its radiator. A common form is the *folded dipole*, in effect two closely spaced dipole radiators connected together at both ends, one being center-fed.

Mumetal A high-permeability alloy of iron and nickel, valued especially for use as a magnetic shield for cathode-ray tubes.

Munsell color system A system for specifying colors in terms of hue, saturation, and brilliance, according to charts. Also see COLOR MATCHING.

Muntz metal See YELLOW METAL.

MUPO Abbreviation of MAXIMUM UNDISTORTED POWER OUTPUT.

Murray-loop bridge A specialized form of Wheatstone bridge in which two of the resistance arms are supplied by a two-wire line (such as a telephone line). By means of resistance measurements made with the bridge, the distance from an office or station to a ground fault on the line can be determined.

MUSA Abbreviation of *multiple-unit steerable antenna.*

muscovite Formula, $KH_2Al_2(SiO_4)_3$. A high-grade variety of mica having low dielectric loss.

Musical Instrument Digital Interface Acronym, MIDI. A computer language used in electronic music. It “tells” the computer when to play a note, how long to play it, and how loud to play it. It also sets the tempo of the music, based on how long a quarter-note lasts. It controls the operation of a music synthesizer, and allows two or more synthesizers to communicate.

musical quality See TIMBRE.

musical scale A series of tones between a given tone and its second harmonic, scaled in intervals of ⅛ octave. The most common example can be found on a piano. Variations exist (e.g., *13-note scale*, *six-note scale*, and *five-note scale*).

music chip An integrated circuit for producing various musical effects (such as tones, percussion, etc.).

music power For a power amplifier, the short-term output power obtained in the reproduction of music waveforms, in contrast to root-mean-square (rms) or effective power output.

music synthesizer A set of oscillators, usually operated with a computer, used to create or playing electronic music. Also see MOOG SYNTHESIZER and MUSIC INSTRUMENT DIGITAL INTERFACE.

music under Pertaining to low-volume, continuous, background music, often added to presentations, radio and television advertisements, educational programs, etc.

muting **1.** Disabling a receiver or amplifier under no-signal or weak-signal conditions. **2.** Softening or muffling a sound.

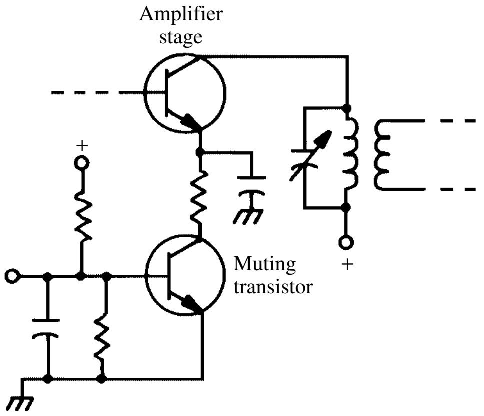

muting

muting circuit **1.** An interference-preventing device that automatically shuts off a radio receiver during the operation of a transmitter, and vice versa. **2.** A squelch circuit. **3.** In a stereo receiver, an electronic element that cuts off all audio when no signal is present or when the receiver is being tuned between carriers.

muting switch A switch or relay that cuts off a receiver during periods of transmission, or when reception is not desired. Generally, a cutoff voltage is applied to one of the intermediate-frequency stages to accomplish muting. In some cases, the audio-frequency circuit is disabled.

mutual antenna coupling Electromagnetic coupling between or among antennas when they are placed too close together. Usually, it is an undesirable phenomenon.

mutual capacitance Inherent capacitance between two conductors.

mutual-capacitance attenuator An attenuator that, in its simplest form, is essentially a shielded, two-plate variable capacitor.

mutual conductance See TRANSCONDUCTANCE.

mutual impedance An impedance shared by two or more branches of a circuit.

mutual inductance Symbol, *M.* Unit, henry. The property shared by neighboring inductors or inductive devices that enables induction to occur. A mutual inductance of 1 henry is present when a current change of 1 ampere per second in one inductor induces 1 volt across another inductor. Also see INDUCTANCE.

mutual-inductance attenuator An attenuator consisting essentially of two coupled coils (input and output), whose spacing can be gradually changed.

mutual-inductance bridge See CAREY-FOSTER MUTUAL-INDUCTANCE BRIDGE.

mutual induction The action whereby the magnetic field produced by alternating current in one conductor produces a voltage in another isolated conductor.

mutual interference **1.** See ADJACENT-CHANNEL INTERFERENCE. **2.** Any kind of interference between or among radio-frequency communications circuits.

mutually exclusive events Two or more events (or data points) so that the occurrence of one prevents the occurrence of the other(s). That is, two events cannot take place simultaneously.

MV **1.** Abbreviation of MEGAVOLT. **2.** Abbreviation of MULTIVIBRATOR.

mV Abbreviation of MILLIVOLT.

MV Abbreviation of MEDIUM VOLTAGE.

MVA Abbreviation of MEGAVOLT-AMPERE.

mV/m Abbreviation of millivolts per meter.

mVP Abbreviation of MILLIVOLT POTENTIOMETER.

MW Symbol for MEGAWATT.

mW Symbol for MILLIWATT.

mw Abbreviation of MEDIUM WAVE.

MWH Abbreviation of MEGAWATT-HOUR.

mW RTL Abbreviation of *milliwatt resistor-transistor logic* or *low-power resistor-transistor logic.*

Mx Abbreviation of MAXWELL.

Mycalex Trade name for an insulating material consisting of mica bonded with glass. Dielectric constant, 6 to 8. Resistivity, 10^{13} ohm-cm.

Mylar A DuPont registered trademark. A tough, plastic insulating material commonly used as a magnetic tape base. Dielectric constant, 2.8 to 3.7. Dielectric strength, 7000 V/mil.

Mylar capacitor A capacitor in which the dielectric film is Mylar.

Mylar tape Magnetic recording tape using a Mylar film as the substrate.

mym Abbreviation of MYRIAMETER.

myoelectricity **1.** Bioelectric pulses of 10- to 1000-mV amplitude produced by muscular activity and detectable by electrodes attached to the skin. Also see ELECTROMYOGRAM, ELECTROMYOGRAPH, and ELECTROMYOGRAPHY. **2.** The study of phenomena, as defined in **1**.

myogram See ELECTROMYOGRAM.

myograph See ELECTROMYOGRAPH.

myography See ELECTROMYOGRAPHY.

myriameter Abbreviation, mym. A metric unit of linear measure equal to 10 kilometers (10^4 meters).

myriametric waves British designation for electromagnetic energy having wavelengths from 100 kilometers down to 10 kilometers, corresponding to frequencies from 3 kHz up to 30 kHz.

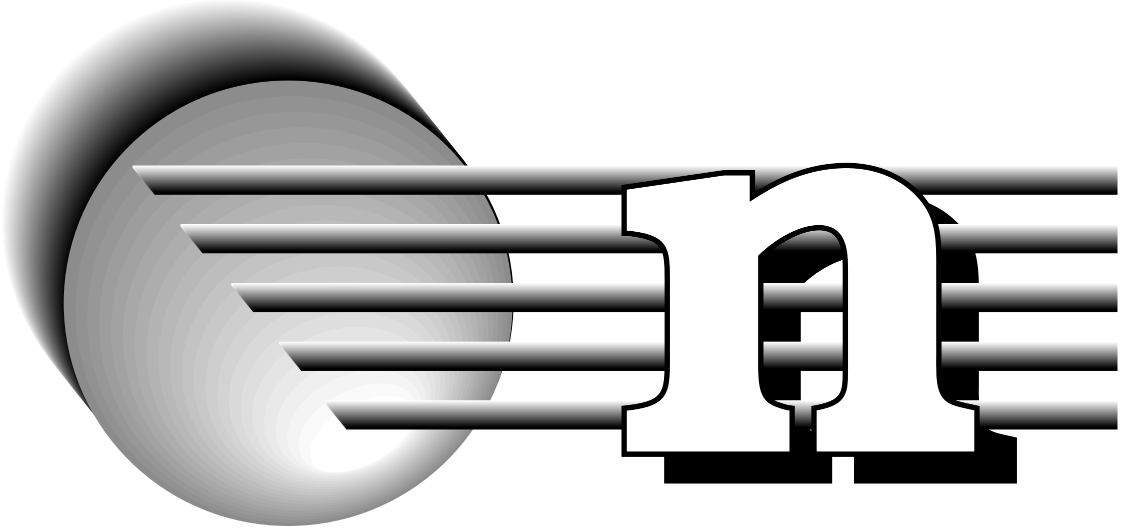

N **1.** Symbol for NUMBER. **2.** Symbol for NITROGEN. **3.** Abbreviation for NEWTON. **4.** Abbreviation of NUMBER (also, No.). **5.** Symbol for the set of natural (counting) numbers (0, 1, 2, 3, . . .).

n **1.** Abbreviation of prefix NANO-. **2.** Symbol for NUMBER. **3.** Symbol for a term (e.g., a multiplier) having an assigned value. **4.** Symbol for part of an expression or operator (as in $2n$, $n - 1$, n^2, etc.). **5.** Symbol for INDEX OF REFRACTION. **6.** Symbol for amount of substance (unit, mole).

NA **1.** Abbreviation of *not available* (as on a specifications sheet). **2.** Abbreviation of *not applicable.*

N_A Symbol for AVOGADRO'S CONSTANT.

Na Symbol for SODIUM.

nA Abbreviation of NANOAMPERE.

NAB Abbreviation of NATIONAL ASSOCIATION OF BROADCASTERS.

NAB curve In audio-frequency operations, the standard magnetic tape playback equalization curve developed by the National Association of Broadcasters (NAB).

NAE Abbreviation of *National Academy of Engineering.*

NAND circuit Also called *NOT-AND circuit.* A binary digital circuit whose output is high (logic 1) if any of the input signals are low (logic 0); the output is low only when all the inputs are high. The performance of the NAND circuit is the inverse of that of the AND circuit.

NAND gate A gate that performs the function of a NAND circuit.

nano- Abbreviation, n. A prefix meaning billionth(s) (i.e., 10^{-9}).

nanoampere Abbreviation, nA. A unit of current; 1 nA equals 10^{-9} ampere.

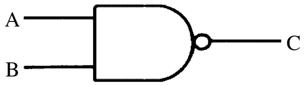

A	B	C
0	0	1
0	1	1
1	0	1
1	1	0

NAND circuit

nanofarad Abbreviation, nF. A unit of capacitance; 1 nF equals 10^{-9} farad.

nanohenry Abbreviation, nH. A unit of low inductance; 1 nH equals 10^{-9} henry.

nanometer Abbreviation, nm. A unit of short wavelength or displacement; 1 nm equals 10^{-9} meter.

nanosecond Abbreviation, ns. A time interval of 10^{-9} second.

nanotechnology The design and construction of superminiature electronic and electromechanical devices. For example, microscopic, computer-controlled robots might someday be used as programmable antibodies to fight infectious diseases, or to build integrated circuits, atom by atom.

nanovolt Abbreviation, nV. A unit of low voltage; 1 nV equals 10^{-9} volt.

nanovoltmeter A sensitive voltmeter for measuring potential in the nanovolt range.

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

nanowatt Abbreviation, nW. A unit of low power; 1 nW equals 10^{-9} watt.

nanowattmeter A meter for measuring power in the nanowatt range.

NAP Abbreviation of *nuclear auxiliary power.*

napier See NEPER.

Napierian base Symbol, *e.* An irrational number, equal to approximately 2.718282, and used as the NAPIERIAN LOGARITHM base.

Napierian logarithm Abbreviation, ln or $\log_e$. A logarithm to the base *e* (see NAPIERIAN BASE). Also called NATURAL LOGARITHM. Compare COMMON LOGARITHM.

NAPU Abbreviation of *nuclear auxiliary power unit.*

narrative A computer program statement that, rather then being an instruction, merely describes the purpose of what follows (usually the steps in a routine or a block of instructions) as a debugging or program modification aid. In a program written in BASIC, such a statement is preceded by the abbreviation REM, for remark. Also called COMMENT.

narrowband **1.** A frequency band in which the difference between upper and lower limits is small compared with bandwidths typical of the service specified. **2.** Pertaining to a radio-frequency emission whose bandwidth is limited or restricted. **3.** Pertaining to a circuit or device that operates over a comparatively small range of frequencies. **4.** In a digital network, a channel having a bandwidth of 64 kbps or less.

narrowband amplifier An amplifier whose passband is restricted to a fraction of the frequency spread common to the amplifier's application.

narrowband FM See NARROWBAND FREQUENCY MODULATION.

narrowband frequency modulation Abbreviation, NBFM or NFM. Frequency modulation in which the maximum instantaneous frequency deviation is no greater than the highest modulating frequency.

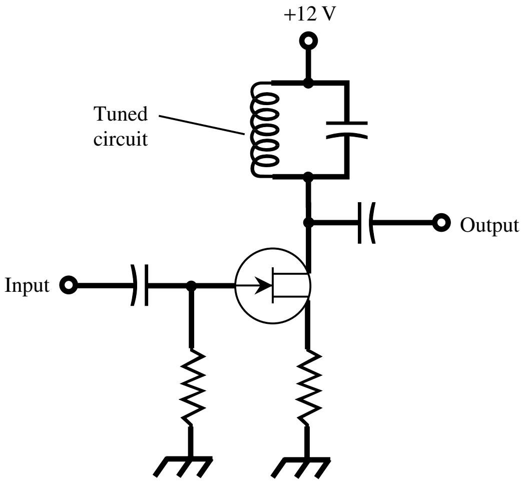

narrowband amplifier

narrowband interference Signal interference whose bandwidth is narrow, compared with that of the circuit affected.

narrowband voice modulation Abbreviation, NBVM. A scheme via which an analog voice signal is compressed into a band of frequencies narrower than that normally required for effective communication.

narrow bandwidth Pertaining to a radio-frequency emission whose bandwidth is comparatively narrow (e.g., a small fraction of the available spectrum space in the frequency band being used).

narrow-sector recorder A directional radio receiver for locating sources of atmospheric noise.

NARTB Abbreviation of *National Association of Radio and Television Broadcasters.*

NAS Abbreviation of *National Academy of Sciences.*

National Association of Broadcasters Abbreviation, NAB. A countrywide organization of radio and television broadcasters.

National Bureau of Standards Abbreviation, NBS. An agency in the U.S. that maintains values for physical constants in the Standard International (SI) System of Units. It also maintains radio broadcast stations that transmit standard time and frequency signals.

National Electrical Manufacturers Association Abbreviation, NEMA. A countrywide organization of manufacturers of electrical and electronic equipment and supplies.

National Electric Code Abbreviation, NEC. Safety regulations and procedures issued by the National Fire Protection Association for the installation of electrical wiring and equipment in the United States. Although the code is advisory from the Association's standpoint, it is enforced to various degrees by local authorities.

National Television Standards Committee Abbreviation, NTSC. A U.S. organization of television companies and other interested organizations. It developed the original black-and-white and color television standards that were approved by the Federal Communications Commission.

natural-decay curve See EXPONENTIAL DECREASE.

natural antenna frequency The fundamental resonant frequency of an electromagnetic antenna.

natural disintegration **1.** The decay of a radioactive substance as a result of the continuous emission of particles and rays. **2.** Also called *half life.* The time required for half of a quantity of a radioactive substance to decay into a different isotope or element.

natural electricity **1.** Atmospheric electricity. **2.** The electricity in living organisms, that is, BIOELECTRICITY.

natural frequency See NATURAL RESONANT FREQUENCY.

natural-growth curve See EXPONENTIAL INCREASE.

natural interference Interference from atmospheric and celestial sources, as opposed to human-made interference.

natural language A spoken or written human language (such as French or Japanese), as opposed to a computer language (such as machine language, C++, or LISP). Translation between natural language and machine language is important in optical character recognition, speech recognition, and speech synthesis.

natural logarithm See NAPIERIAN LOGARITHM.

natural magnet A material, such as magnetite (lodestone), found in nature and exhibiting permanent magnetism.

natural magnetism Magnetism found in some natural materials (see NATURAL MAGNET) and in the earth itself.

natural number **1.** Any nonnegative whole number, that is, a member of the set $N = (0, 1, 2, 3, \ldots)$. **2.** See NAPIERIAN BASE.

natural period The time required for one complete wave cycle to occur in a device at its NATURAL RESONANT FREQUENCY.

natural radiation Noise in the form of radiation emitted by natural radioactive substances, cosmic rays, etc. Also called BACKGROUND RADIATION.

natural resonance Resonance resulting from the unique physical constants of a body, circuit, or system. Also see NATURAL RESONANT FREQUENCY.

natural resonant frequency **1.** The frequency at which a circuit or device responds with maximum amplitude to applied signals. **2.** The frequency at which a circuit or device generates maximum energy. **3.** The frequency at which an object vibrates at maximum amplitude.

natural wavelength The wavelength corresponding to the NATURAL RESONANT FREQUENCY of a circuit, device, or object.

nautical mile A nautical unit of linear measure equal to 1.852 kilometers (1.1508 statute miles).

nav **1.** Abbreviation of *navigation.* **2.** Abbreviation of *navigational.*

Navaglobe A radionavigation system used at very-low or low frequencies over long distances.

NAVAIDS Abbreviation of NAVIGATIONAL AIDS.

navar A radar system in which a ground radar scans the immediate vicinity of an airport, observes the flight activity, and transmits such observations to aircraft in flight. The name is a contraction of the term *navigation and ranging.*

navigation aid An electronic device or system, such as radar or radio direction finding, that assists in the navigation of vehicles on land, at sea, or in the air.

navigation beacon A beam that provides aircraft and ships with navigational aid.

NAWAS Abbreviation of *National Attack Warning System.*

Nb Symbol for NIOBIUM.

nb Abbreviation of NARROWBAND.

NBFM Abbreviation of NARROWBAND FREQUENCY MODULATION. Also abbreviated NFM.

NBS Abbreviation for NATIONAL BUREAU OF STANDARDS.

NBTDR Abbreviation of *narrowband time-domain reflectometry.*

nbw Abbreviation of NOISE BANDWIDTH.

NC **1.** Abbreviation of NORMALLY CLOSED. **2.** On drawings, abbreviation of *no connection.* **3.** Abbreviation of NUMERICAL CONTROL.

N/C Abbreviation of NUMERICAL CONTROL.

nc Abbreviation of *no connection.*

n-channel junction field-effect transistor Abbreviation, NFET. A junction field-effect transistor in which the gate junction is formed on a bar or die of n-type semiconductor material.

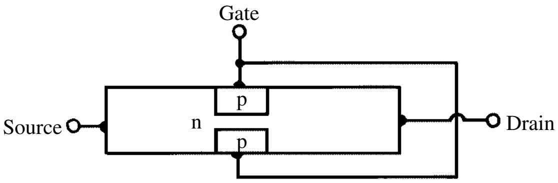

n-channel junction field-effect transistor

n-channel MOSFET A metal-oxide-semiconductor (MOS) field-effect transistor in which the channel is n-type material.

NCMT Abbreviation of NUMERICALLY CONTROLLED MACHINE TOOL.

NCS Abbreviation of *net control station(s).*

N curve A negative-resistance voltage-current curve having a shape roughly resembling an uppercase letter N. Compare S CURVE.

Nd Symbol for NEODYMIUM.

N display A radar display in which the target is represented by a pair of vertical blips, whose amplitude indicates target direction and whose position along the horizontal base line (as determined by lining up a pedestal signal with the blips) is read from the pedestal-adjustment calibration.

n-doped zinc-oxide ceramic A nonlinear-resistance material used in the manufacture of some voltage-dependent resistors.

NDR Abbreviation of NONDESTRUCTIVE READ.

NDT Abbreviation of NONDESTRUCTIVE TEST.

Ne Symbol for NEON.

near-end crosstalk Crosstalk originating at, or near, the telephone subscriber line in which the interference is noticed.

near field **1.** The radiation field within a radius of 1 wavelength from a transmitting antenna. **2.** The sound field near a loudspeaker or other reproducer.

near infrared Pertaining to electromagnetic radiation of 3 to 30 micrometers in wavelength.

near ultraviolet That part of the ultraviolet spectrum nearest the visible light wavelengths. These are the least-penetrating ultraviolet rays. They are also called *soft ultraviolet.*

near video on demand In television, a local service in which various programs/videos are broadcast at specified times.

near zone See NEAR FIELD, **1**.

NEB Abbreviation of *noise equivalent bandwidth.*

NEC Abbreviation of NATIONAL ELECTRIC CODE.

necessary bandwidth The minimum bandwidth needed (with a given emission) to transmit information at a required rate and of a required quality. In general, as data speed increases (as measured in bits per second, baud, or words per minute), the minimum required bandwidth increases. For slow-speed Morse code signals, it can be as small as a few Hertz; for fast-scan television, it is several megahertz.

neck The straight portion of a cathode-ray-tube (CRT) envelope (i.e., the part between the base and the flared portion).

NEDA Abbreviation of *National Electronic Distributors' Association.*

needle **1.** The stylus of a phonograph cartridge. **2.** The pointer of an indicating meter. **3.** One of the electrodes in a voltage-measuring spark gap (see NEEDLE GAP). **4.** The slender, pointed metal tip of a test probe.

needle chatter See NEEDLE TALK.

needle drag In disc recording and reproduction, friction between the needle (stylus) and disc. Also called *stylus drag.*

needle electrode See NEEDLE, **3**, **4**.

needle gap A spark gap composed of two needles having an adjustable air gap between their points. An unknown high voltage is measured in terms of the gap width necessary for sparking.

needle memory A computer memory in which the dual-state elements are thin magnetic needles.

needle pointer The pointer of a meter or compass.

needle pressure See VERTICAL STYLUS FORCE.

needle probe See NEEDLE-TIP PROBE.

needle scratch In disc recording and reproduction, noise resulting from vibration of the needle (stylus) because of an irregular groove surface. Also called *surface noise.*

needle talk Direct radiation of sound by the stylus of a phonograph pickup. Also called *needle chatter.*

needle telegraph A telegraph in which the Morse code characters are converted into magnetic needle deflections.

needle test point See NEEDLE, **4**.

needle-tip probe A test probe that terminates in a sharp point. Also see NEEDLE, **4**.

needle voltmeter See NEEDLE GAP.

neg Abbreviation of NEGATIVE.

negate **1.** To insert the NOT operation in front of a digital expression. **2.** To change logic 1 to logic 0 (high to low) or vice versa.

negation The logical NOT operation (see NOT) in digital systems. Also see NAND CIRCUIT, NOR CIRCUIT, NOR GATE, NOT CIRCUIT, and NOT-OR CIRCUIT.

negation element In a computer system, a device that can give the reverse of a condition, event, or signal.

negative **1.** Possessing NEGATIVE ELECTRIFICATION. **2.** In voice communications, a word often used for "no," especially when interference is present or signals are weak. **3.** Pertaining to a real number less than zero. **4.** An image whose shadings are opposite to those in the scene.

negative acceleration A decrease in speed or velocity; also called *deceleration.*

negative acknowledgment character In a handshaking or forward-error-correction (FEC) system, a response by the receiving station that indicates a missed bit or bits.

negative angle **1.** An angle in the third or fourth quadrant in a system of rectangular coordinates. **2.** An angle measured clockwise from the positive *x*-axis in a rectangular coordinate system. Compare POSITIVE ANGLE.

negative bias A steady, negative direct-current voltage or current applied continuously to an electrode of a device, such as a transistor to establish the operating point. Compare POSITIVE BIAS.

negative bus See NEGATIVE CONDUCTOR.

negative charge An electric charge consisting of a quantity of NEGATIVE ELECTRIFICATION. Also see CHARGE, **1**; ELECTRIC CHARGE; and UNIT ELECTROSTATIC CHARGE. Compare POSITIVE CHARGE.

negative conductor The conductor or wire connected to the negative terminal of a current, power, or voltage source. Compare POSITIVE CONDUCTOR.

negative electricity See NEGATIVE CHARGE and NEGATIVE ELECTRIFICATION.

negative electrification Electrification characterized by an excess of electrons. For example, when a glass rod is rubbed with a silk cloth, the cloth becomes negatively charged because electrons are transferred from the glass to the cloth. Similarly, when a neutral atom acquires an extra electron, the atom becomes negatively charged because it has an excess of electrons. Compare POSITIVE ELECTRIFICATION.

negative electrode **1.** An electrode connected to the negative terminal of a current, power, or voltage source. **2.** The negative output terminal of a current, power, or voltage source, such as a battery or generator.

negative error of measurement An error of measurement in which the difference between a measured value and the true or most probable value is negative. Compare POSITIVE ERROR OF MEASUREMENT.

negative exponent In mathematical notation, an exponent indicating that a number is to be raised to a negative power. Raising a number *x* to a negative power –*n* means taking the reciprocal of the number raised to the power *n*. That is, $x^{-n} = 1/x^n$; for example, $10^{-2} = 1/10^2 = 1/100 = 0.01$.

negative feedback Feedback that is out of phase with the input signal. Also called INVERSE FEEDBACK, DEGENERATION, and DEGENERATIVE FEEDBACK. Compare POSITIVE FEEDBACK.

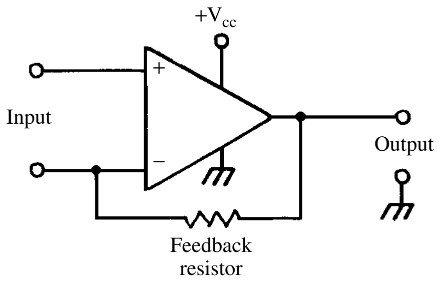

negative feedback

negative-feedback amplifier An amplifier in which negative feedback is used to improve performance or modify response.

negative function A trigonometric function having the negative sign. In a rectangular coordinate system, the sine function is negative in the third and fourth quadrants, the cosine in the second and third, and the tangent in the second and fourth. Compare POSITIVE FUNCTION.

negative gain A misnomer for FRACTIONAL GAIN, arising because fractional gain can be expressed in negative decibels.

negative ghost In a television picture, a ghost with negative (see NEGATIVE, **4**) shading.

negative ground In a direct-current power system, the connection of the negative pole to common ground.

negative image **1.** A picture in which the blacks, whites, and shades in between are the reverse of those in the scene (see NEGATIVE, **4**). **2.** An abnormal image that has the reverse shading described in **1**.

negative impedance An impedance that displays the same behavior as that of NEGATIVE RESISTANCE.

negative ion An atom with an excess of electrons and, consequently, a net negative charge. Also called ANION. Compare POSITIVE ION.

negative-ion generator A device for generating negative ions and circulating them into the surrounding air. This is believed by some people to be beneficial to health.

negative lead See NEGATIVE CONDUCTOR.

negative-lead filtering Power-supply filtering in which the choke coils and capacitors are in the negative direct-current lead, rather than in the positive lead (the usual position). One advantage of this arrangement is the lower insulation requirement of the choke.

negative light modulation In television image transmission, the condition in which transmitted power is increased by a decrease in the initial intensity of light. Compare POSITIVE LIGHT MODULATION.

negative line See NEGATIVE CONDUCTOR.

negative logic **1.** Binary logic in which a high negative state represents logic 1, and a low negative state represents logic 0. **2.** Binary logic in which a low positive state represents logic 1, and a high positive state represents logic 0. Compare POSITIVE LOGIC.

negative measurement error See NEGATIVE ERROR OF MEASUREMENT.

negative modulation Amplitude-modulated television transmission in which the transmitted power decreases as image brightness increases. Compare POSITIVE MODULATION.

negative modulation factor For an amplitude-modulated wave having unequal positive and negative modulation peaks, a ratio expressing the maximum negative deviation from the average for the envelope. Compare POSITIVE MODULATION FACTOR.

negative number A real number less than zero (i.e., one to which the minus sign is assigned).

negative peak The maximum negative instantaneous current or voltage in an alternating-current waveform.

negative-peak voltmeter An electronic meter for measuring the NEGATIVE PEAK voltage of an alternating-current waveform.

negative phase-sequence relay A relay that responds to the negative phase sequence in a polyphase circuit. Compare POSITIVE PHASE-SEQUENCE RELAY.

negative picture modulation See NEGATIVE MODULATION.

negative picture phase In a television signal, the picture-signal voltage swing from zero to negative, in response to an increase in brightness. Compare POSITIVE PICTURE PHASE.

negative plate The negative member of a battery cell; electron flow is from the plate through the external circuit.

negative pole See NEGATIVE ELECTRODE, **1**, **2**.

negative positive zero Abbreviation, NP0. Pertaining to temperature-compensating capacitors having a temperature coefficient of capacitance that changes sign within a specified temperature range.

negative potential **1.** The potential measured at a negative electrode, with respect to the positive electrode or to ground. **2.** Potential less than that of the earth as a reference.

negative resistance **1.** A decrease in voltage across a device as the current through it increases, and an increase in voltage as the current decreases. **2.** A decrease in current through a device as the voltage across it increases, and an increase in current as the voltage decreases. This is opposite to the behavior of an ohmic (positive) resistance. Also see N-CURVE, NEGATIVE-RESISTANCE REGION, NEGATIVE RESISTOR, and S CURVE.

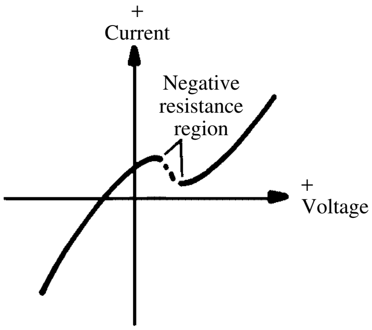

negative resistance

negative-resistance amplifier A simple circuit in which a negative-resistance device, such as a TUNNEL DIODE, cancels the positive resistance of the circuit, causing amplification or oscillation.

negative-resistance device See NEGATIVE RESISTOR.

negative-resistance diode **1.** See TUNNEL DIODE. **2.** A reverse-biased germanium diode (and occasionally a silicon diode) that exhibits NEGATIVE RESISTANCE. **3.** A special diode tube that, when operated at ultrahigh frequencies, exhibits negative resistance because of transit-time effects.

negative-resistance magnetron A split-anode magnetron operated at a combination of anode voltage and magnetic field strength corresponding to cutoff; it exhibits negative resistance to voltage applied symmetrically between the anode halves. The frequency of oscillation is determined by an external tank circuit.

negative-resistance oscillator An oscillator that consists of a negative-resistance device connected across a tuned circuit. The arrangement oscillates because the negative resistance cancels the positive resistance (losses) of the tuned circuit. See, for example, NEGATIVE-RESISTANCE MAGNETRON and TUNNEL-DIODE OSCILLATOR.

negative-resistance region In certain devices, a portion of the voltage-versus-current curve having negative slope (i.e., the current decreases as the voltage increases, or vice versa). Also see N CURVE, NEGATIVE RESISTANCE, and S CURVE.

negative-resistance repeater A repeater that produces gain by means of NEGATIVE RESISTANCE effects.

negative resistor Any device exhibiting NEGATIVE RESISTANCE.

negative space charge The cloud of electrons (negative particles) in the region surrounding an emitter, such as the hot cathode of a vacuum tube.

negative temperature coefficient Abbreviation, NTC. A number expressing the amount by which a quantity (such as the rating of a component) decreases when the temperature is raised. The coefficient is stated as a percentage, or as a certain number of parts per million (ppm) per degree temperature rise. Compare POSITIVE TEMPERATURE COEFFICIENT and ZERO TEMPERATURE COEFFICIENT.

negative terminal See NEGATIVE ELECTRODE, **2**.

negative torque In an electric motor, a torque that acts against the operating torque.

negative transmission In a television or facsimile system, the condition in which brighter light corresponds to lower transmitted power, and dimmer light corresponds to higher transmitted power.

negative valence The valence of a negative ion. Also see VALENCE.

negator A logical NOT element [i.e., one that outputs the complement of an input bit (1 for 0 and vice-versa)].

negatron The term that specifically differentiates the familiar ELECTRON from a POSITRON (positive electron).

NEI Abbreviation of *noise equivalent input.*

neighborhood node A local site at which a fiberoptic television (video) signal is converted to a cable signal.

NEL Abbreviation of *National Electronics Laboratory.*

NELA Abbreviation of *National Electric Light Association.*

N electron In certain atoms, an electron whose orbit is outside of, and nearest to, those of the M electrons.

NEMA Abbreviation of *National Electrical Manufacturers' Association.*

nematic crystal A normally transparent liquid crystal that becomes opaque when an electric field is applied to it, and becomes transparent again when the field is removed. The crystal material is cut in the form of a letter or numeral and provided with a reflecting backplate for display readouts in calculators, watches, and various electronic devices.

nematic-crystal display A device in which an electrically controlled film of nematic-crystal material is used to transmit and interrupt light from a lamp or from a reflecting mirror, in this way displaying characters in whose shape the film has been formed.

nematic liquid In a liquid-crystal display (LCD), a normally clear liquid that becomes opaque in the presence of an electric field.

nemo A radio or television program that is picked up from a location outside the studio. Also called *field pickup* or *remote.*

neodymium Symbol, Nd. A metallic element of the rare-earth group. Atomic number, 60. Atomic weight, 144.24.

neodymium-YAG laser A laser that uses neodymium and yttrium-aluminum-garnet (YAG) crystal. It is used in medical applications, such as surgery and various other jobs where high precision is required. Generally, it is a low-to-medium-power laser.

neomatachograph An instrument used in psychology to measure complex reaction time. Compare NEOMATACHOMETER.

neomatachometer An instrument used in psychology to measure simple reaction time. Compare NEOMATACHOGRAPH.

neon Symbol, Ne. An inert-gas element. Atomic number, 10. Atomic weight, 20.179. Neon, present in trace amounts in the earth's atmosphere, is used in some glow tubes, readout devices, and indicator lamps.

neon bulb A (usually small) neon-filled gas diode. It has a characteristic pink glow and is ignited by a firing voltage for the particular unit. Also called *neon glow lamp* and *neon tube.*

neon-bulb flip-flop A flip-flop circuit (bistable multivibrator) using two neon bulbs as the bistable components.

neon-bulb gate A gate circuit containing a neon bulb biased below the firing point. A trigger voltage added to the bias voltage raises the applied voltage and fires the bulb, producing an output pulse.

neon-bulb logic Logic circuits composed of neon-bulb gates.

neon-bulb memory See NEON-BULB STORAGE.

neon-bulb multivibrator A multivibrator using two neon bulbs as the switching components.

neon-bulb oscillator A simple relaxation oscillator consisting essentially of a neon bulb, capacitor, resistor, and direct-current supply. The frequency of the sawtooth-wave output depends principally on the capacitance and resistance values. The maximum operating frequency is limited to about 5 kHz by the deionization time of the gas.

neon-bulb overmodulation indicator The application of a neon-bulb overvoltage indicator as a monitor for amplitude-modulated radio-frequency signals. The bulb flashes each time the modulation percentage exceeds a predetermined value. Also called *neon-bulb modulation alarm.*

neon-bulb overvoltage indicator A relatively simple circuit in which a neon bulb flashes each time a voltage monitored by the circuit exceeds a predetermined value. The flash shows that the portion of the voltage presented to the bulb has exceeded the firing potential.

neon-bulb peak indicator See NEON-BULB OVERVOLTAGE INDICATOR.

neon-bulb ring counter A ring counter composed of neon-bulb flip-flops. The maximum counting speed is limited by the deionization time of the neon bulbs to approximately 300 events per second.

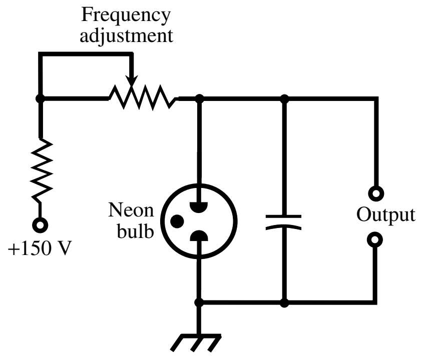

neon-bulb oscillator

neon-bulb sawtooth generator A relatively simple relaxation oscillator using a neon bulb, capacitor, and resistor. The output is a sawtooth wave, whose frequency is determined principally by the capacitance and resistance values.

neon-bulb scale-of-two circuit A scale-of-two circuit (frequency halver) using neon bulbs as the bistable elements.

neon-bulb storage A storage (memory) device composed of neon bulbs. A fired bulb (representing a bit of stored information) remains fired until turned off by an erase signal.

neon-bulb stroboscope A stroboscope in which a neon bulb supplies the light flashes. The circuit is essentially that of the neon-bulb oscillator, the flash rate being continuously variable by an adjustable frequency control.

neon-bulb voltage regulator A simple circuit utilizing the constant voltage drop across a fired neon bulb as a regulated voltage. The usual circuit configuration is a neon bulb and current-limiting resistor in series with a power supply.

neon-bulb volume indicator A neon-bulb overvoltage indicator used in some tape recorders to show when the volume exceeds a predetermined level—especially when the volume is high enough to cause an unacceptable amount of distortion.

neon glow lamp See NEON BULB.

neon lamp See NEON BULB.

neon pilot lamp A neon bulb used as a pilot lamp operated from the power-line circuit of an electronic equipment. Also called *neon pilot light.*

neon tube See NEON BULB.

NEP Abbreviation of NOISE EQUIVALENT POWER.

NEPD Abbreviation of *noise equivalent power density.*

neper Abbreviation, Np. A Napierian-logarithmic unit expressing a ratio of power levels; $Np = \log_e (P_1/P_2)^{1/2}$, where P_1 and P_2 are the power values being compared and e is the Napierian logarithm base, equal to approximately 2.71828. The neper is related to the decibel (dB), a similar unit based on common logarithms, in the following manner: 1 Np = 8.686 dB; 1 dB = 0.1151 Np.

neptunium Symbol, Np. A radioactive metallic element produced artificially. Atomic number, 93. Atomic weight, 237.05.

Nernst effect The appearance of a voltage between the opposite edges of a metal strip that is conducting heat longitudinally when the strip is placed in a magnetic field perpendicular to the plane of the strip.

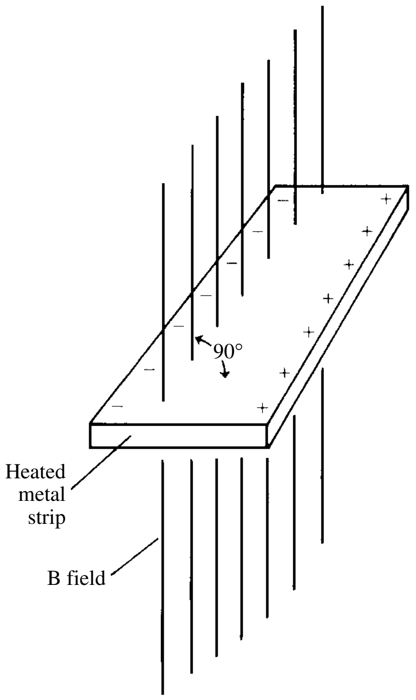

Nernst effect

Nernst-Ettinghausen effect In a piezoelectric crystal, the tendency for a temperature gradient to exist as a result of applied electromagnetic fields.

Nernst lamp An incandescent lamp in which the filament material is a mixture of magnesia and certain rare oxides (such as yttria). A surrounding current-carrying coil heats the filament, lowering its resistance and causing it to glow, after which it continues to operate on a low current.

nerve center The most essential part of a control system or communication network.

nerve current In an animal organism, the small electrical current that flows along nerve fibers.

nervous breakdown The upset of a function in a digital computer (which becomes disabled as a result).

NESC Abbreviation of *National Electrical Safety Code.*

nested loops See NESTING.

nesting In digital computers and data processing, the inclusion of one or more loops or subroutines inside a larger loop or subroutine. Nested loops, for example, can be used to assign values to double-subscripted variables.

NET Abbreviation of *noise equivalent temperature.*

net **1.** An effective, useful, or resultant quantity (e.g., NET CAPACITANCE). **2.** Colloquialism for *communications network.* See NETWORK, **2**.

net authentification A password or other special identification used in a radio network.

net capacitance The resultant capacitance in a circuit in which capacitances act in combination with each other.

net component The total value of two or more passive components of the same sort. See, for example, NET CAPACITANCE, NET CURRENT, NET IMPEDANCE, NET INDUCTANCE, NET POWER, NET REACTANCE, NET RESISTANCE, and NET VOLTAGE.

net current The current flowing in a circuit in which currents aid or oppose each other.

net gain For an amplifier, the amount of gain remaining after all losses in the device have been subtracted.

net impedance The impedance of a circuit in which impedances act in combination with each other.

net inductance The inductance of a circuit in which inductances act in combination with each other.

net loss For an amplifier or other system, the algebraic sum of gains and losses between two points in the system.

net power The resultant power observed when power signals aid or oppose each other in a single circuit or system.

net radiometer A device for measuring the difference in intensity between radiation entering the earth's surface and radiation leaving it. This is generally visible light or infrared radiation from the sun.

net reactance Symbol, X_t. The combined inductive reactance (X_L) and capacitive reactance (X_C) in a circuit or device.

net resistance The resistance of a circuit in which resistances act in combination with each other.

net voltage The resultant voltage at a point where voltages aid or oppose each other.

network **1.** A circuit arrangement of electronic components, sometimes redundant in its design [e.g., resistance-capacitance (RC) network]. **2.** A group of interconnected computers, communications stations, or other facilities, often organized for simultaneous operation and data transfer. **3.** To conduct research or gather information using a group of interconnected computers, communications stations, or other facilities.

network analog A circuit or circuits representing variables and used to express and solve a mathematical relationship between the variables.

network analysis The rigorous examination of a network to determine its properties and performance. Compare NETWORK SYNTHESIS.

network analyzer An analog or digital circuit for simulating and analyzing a network (see NETWORK, **1**) by means of measurements and calculations from the model so obtained.

network assurance The construction and operation of a network in such a manner that failures can be quickly and easily located and corrected.

network calculator An analog or digital device for determining the component values and performance of a given network (see NETWORK, **1**).

network constant The value of a passive component (capacitance, inductance, resistance, etc.) used in a network (see NETWORK, **1**).

network filter A transducer that passes or rejects signals, depending on their frequency.

network interface unit Abbreviation, NIU. A device, provided to each subscriber, that connects telephones, television sets, and personal computers to an electrical or fiberoptic cable. Envisioned as an integral component of the multimedia INFORMATION SUPERHIGHWAY.

network provisioning The control of a network by means of computer software, rather than via manual rearrangement of system devices.

network relay **1.** A relay that provides protection of the circuits in a network (see NETWORK, **1**); a circuit breaker. **2.** In a communications network (see NETWORK, **2**), the reception and retransmission of a message by a human operator.

network synthesis The design and fabrication of a network (see NETWORK, **1**) by rigorous engineering methods to achieve a prescribed performance.

network theorems See COMPENSATION THEOREM, MAXIMUM POWER TRANSFER THEOREM, NORTON'S THEOREM, RECIPROCITY THEOREM, SUPERPOSITION THEOREM, and THEVENIN'S THEOREM.

network topology The analysis of networks (see NETWORK, **1**) with signal-flow diagrams. Components of the diagram represent signal paths, open loops, closed loops, and node points.

network transfer function A function describing the overall processing of energy by a network (see NETWORK, **1**); it is equal to V_o/V_i, where V_i is the input voltage and V_o is the output voltage of the network.

Neuman's law A property of mutual inductances. For a given orientation and environment for two inductors, the value of the mutual inductance does not change, regardless of the magnitude, frequency, or phase of the currents in the coils. That is, mutual inductance is subject only to the physical environment surrounding the coils.

neuristor A two-terminal semiconductor device that simulates the behavior of a neuron (nerve cell) and allows machines to duplicate some of the neurological phenomena observed in the human body.

neuroelectricity Low-voltage electricity in the nervous system of a human being or animal.

neuron A nerve cell in a living organism.

neurosurgery assistance robot An electromechanical device, controlled by a human operator and assisted by computers, that helps to perform simple neurosurgical operations requiring extreme precision and steadiness. One procedure for which a robot has been used is the drilling of holes in the skull prior to brain surgery.

neutral **1.** Having no electric charge; thus, an atom is normally neutral because its internal positive charges neutralize its internal negative charges. **2.** Devoid of voltage (e.g., a neutral line). **3.** Pertaining to a chemical, especially a solution, that is neither acidic nor alkaline.

neutral bus See NEUTRAL WIRE.

neutral circuit **1.** A deenergized circuit (i.e., a "dead" one). **2.** In a teletypewriter system, a circuit in which current flows in one direction.

neutral conductor See NEUTRAL WIRE.

neutral ground A ground connection to a neutral wire or to the neutral point of a circuit.

neutralization The process of balancing out positive feedback in an amplifier to prevent self-oscillation. Also see NEUTRALIZING CAPACITOR and NEUTRALIZING COIL.

neutralization indicator A neon bulb or radio-frequency meter used during neutralization of an amplifier. When the amplifier is properly neutralized, the bulb or meter will show zero signal with no input to the amplifier, indicating that the circuit is not oscillating.

neutralize To eliminate positive feedback in a radio-frequency amplifier.

neutralized amplifier An amplifier in which neutralization has been performed to prevent self-oscillation.

neutralizing capacitor In a capacitively neutralized circuit, a small capacitor that serves as a coupler of signal energy from the output back to the input in reverse phase, to cancel self-oscillation of the circuit.

neutralizing circuit Any component or set of components that is used to neutralize a radio-frequency amplifier.

neutralizing coil In an inductively neutralized circuit, a small coupling coil that picks up signal energy from the output and applies it in reverse phase to the input, to cancel circuit self-oscillation.

neutralizing tool A nonconducting screwdriver-like device for adjusting a neutralizing capacitor or coil. It is usually made of fiber or plastic.

neutralizing voltage The feedback voltage that cancels self-oscillation in the process of neutralization.

neutralizing wand See NEUTRALIZING TOOL.

neutral line See NEUTRAL WIRE.

neutral relay See UNPOLARIZED RELAY.

neutral wire In a polyphase power-transmission system, the line (wire) that does not carry current until the system is unbalanced.

neutrodon See NEUTRODYNE.

neutrodyne A radio-frequency amplifier that is neutralized by feeding energy through a small capacitor from a tap on the secondary coil of the collector or drain output transformer back to the base or gate.

neutron An uncharged atomic particle having a mass approximately equal to that of the proton. The neutron is present in the nucleus of every atom, except that of hydrogen.

neutron age 16.7% of the mean square displacement of a neutron as it moves through a specified energy range.

neutron rest mass See MASS OF NEUTRON AT REST.

new candle See CANDELA.

newton (Sir Isaac Newton.) Symbol N. The SI unit of force. A force of 1 newton imparts an acceleration of 1 meter per second per second to a mass of 1 kilogram. The newton is equal to 10^5 dynes.

Newton's laws of motion Three natural laws discovered in 1686 by Sir Isaac Newton. First law: A body at rest or in motion tends to remain in that state, unless it is acted upon by some force. Second law: A body tends to accelerate or decelerate when it is acted upon by a force, the acceleration or deceleration being directly proportional to the force and inversely proportional to the mass of the body. Third law: For every action or acting force, there is an equal and opposite reaction or reacting force.

nexus An interconnection point in a system.

NF Abbreviation of NOISE FIGURE.

nF Abbreviation of NANOFARAD.

NFET Abbreviation of N-CHANNEL JUNCTION FIELD-EFFECT TRANSISTOR.

NFM Abbreviation of NARROWBAND FREQUENCY MODULATION. Also abbreviated NBFM.

NFM reception Reception of a signal having NARROWBAND FREQUENCY MODULATION. Standard discriminators and ratio detectors can be used. Slope detection can be achieved with an amplitude-modulation (AM) receiver by tuning slightly to one side of the center frequency of the FM signal.

NFM transmission Transmission of a signal via NARROWBAND FREQUENCY MODULATION. A simple method consists of frequency modulating the master oscillator of the transmitter at the modulation frequency. Phase modulation can also be used.

NFPA Abbreviation of *National Fire Protection Association* (see NATIONAL ELECTRIC CODE).

NFQ Abbreviation of *night frequency.*

NG **1.** Abbreviation of NEGATIVE GLOW. **2.** Abbreviation of *no good*, used for marking inoperative or malfunctioning components.

nH Abbreviation of NANOHENRY.

nhp Abbreviation of NOMINAL HORSEPOWER.

Ni Symbol for NICKEL.

nibble In computer operations, a four-bit word.

NICAD Acronym for NICKEL–CADMIUM.

NiCd Chemical symbol for NICKEL–CADMIUM.

Nichrome A nickel-chromium alloy used in the form of a wire or strip for resistors and heater elements.

nickel Symbol, Ni. A metallic element. Atomic number, 28. Atomic weight, 58.69. Nickel is familiar as an alloying metal in some resistance wires and as the material used in some electron-tube elements.

nickel–cadmium Acronym, NICAD. Chemical symbol, NiCd. A mixture used in certain rechargeable electrochemical cells and batteries. In recent years, these cells and batteries have been largely replaced by nickel–metal-hydride (NiMH) and lithium-based units. These batteries must be disposed of in a special way because of the toxic cadmium they contain. Compare NICKEL–METAL-HYDRIDE.

nickel–cadmium battery **1.** A battery of nickel–cadmium cells. **2.** Loosely, a nickel–cadmium cell. See also NICKEL–CADMIUM.

nickel-cadmium cell A small, rechargeable, electrochemical cell. The anode is cadmium, the cathode is nickel hydroxide, and the electrolyte is potassium hydroxide. See also NICKEL–CADMIUM.

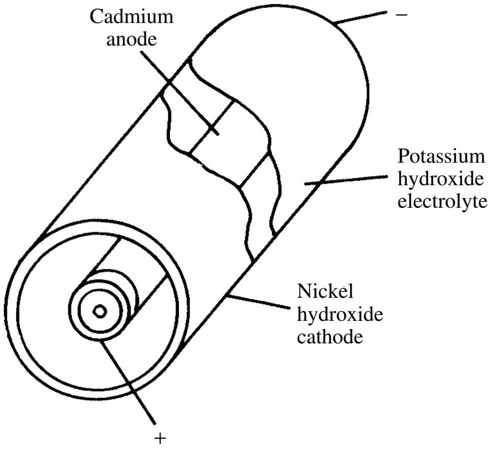

nickel-cadmium cell

nickel hydroxide Chemical symbol, NiOH. A compound used in certain rechargeable electrochemical power supplies. Examples are nickel–cadmium (NiCd or NICAD) and nickel–metal-hydride (NiMH) batteries, used in older notebook computers.

nickel–iron Chemical symbol, NiFe. A mixture used for a specialized rechargeable electrochemical cell in which the active positive plate material consists of nickel hydroxide. The active negative plate material is powdered iron oxide mixed with cadmium, and the electrolyte is potassium hydroxide.

nickel-iron battery See EDISON BATTERY.

nickel-metal–hydride Chemical symbol, NiMH. A mixture used in certain rechargeable electrochemical cells and batteries. This type of cell or battery is less likely than nickel–cadmium units to develop *battery memory* or *memory drain,* in which the ampere-hour capacity is reduced because of repeated partial discharging. In addition, the chemicals in NiMH cells and batteries are less toxic than those in NICADs. Compare NICKEL–CADMIUM.

nickel–metal-hydride battery **1.** A small, rechargeable, nickel-based electrochemical battery, similar to the nickel–cadmium (NICAD) type. **2.** Loosely, a nickel–metal-hydride cell. See NICKEL–METAL-HYDRIDE.

nickel–metal-hydride cell A small, rechargeable, nickel-based electrochemical cell, similar to the nickel–cadmium type. See also NICKEL–METAL-HYDRIDE.

nickel-oxide diode A diode fabricated from nickel-oxide semiconductor material.

nickel silver An alloy of copper, nickel, and zinc, sometimes used for making resistance wire. Also called GERMAN SILVER.

Nicol prism An optical component for producing or analyzing plane-polarized light. It consists of two prisms of Iceland spar cemented together. Light entering the device strikes the interface, where the ordinary ray is totally reflected and the extraordinary ray passes through, both rays being plane-polarized perpendicular to each other.

NIDA Abbreviation of *numerically integrating differential analyzer.*

NIF Abbreviation of NOISE IMPROVEMENT FACTOR.

NiFe Chemical symbol for NICKEL–IRON.

night effect A phenomenon sometimes observed at frequencies below approximately 500 kHz. Direction-finding signals received between sunset and the following sunrise appear to come from a transmitter that moves slowly back and forth.

night-effect errors In radio direction-finding, inaccurate or uncertain readings resulting from the NIGHT EFFECT.

night range The distance over which signals from a given transmitter are consistently received after sunset.

ni junction In a semiconductor device, the junction between an n-type layer and an intrinsic layer.

nil **1.** Colloquialism for *negligible.* **2.** British expression for *zero.* Also, *nought.* **3.** Colloquialism for *nothing.*

NiMH Chemical symbol for NICKEL–METAL-HYDRIDE.

niobium Symbol, Nb. A metallic element chemically resembling tantalum. Atomic number, 41. Atomic weight, 92.91.

NIPO Abbreviation of *negative input/positive output.*

NIR Abbreviation of NEAR INFRARED.

nit In digital-computer operations, a unit of data equal to 1.44 bits.

nitrocellulose See CELLULOSE NITRATE.

nitrogen Symbol, N. A gaseous element. Atomic number, 7. Atomic weight, 14.007. Nitrogen is the most-abundant component (about 78%) of the Earth's atmosphere.

NIU Abbreviation of NETWORK INTERFACE UNIT.

Nixonite A trade name for cellulose acetate, a plastic.

Nixonoid A trade name for cellulose nitrate, a plastic.

NJCC Abbreviation of *National Joint Computer Conference.*

n layer A semiconductor layer doped to provide current carriers that are predominantly electrons. Compare P LAYER.

n-level logic **1.** A multilevel form of logic, with *n* different possible states. **2.** In a computer, the connection of up to *n* logic gates.

NLR **1.** Abbreviation of *nonlinear resistance.* **2.** Abbreviation of NONLINEAR RESISTOR.

NLS Abbreviation of NO-LOAD SPEED.

N/m^2 Abbreviation of *newtons per meter squared* (pascals).

Nm^2/kg Newton meters squared per kilogram, the SI unit of the *gravitational constant.*

NMAA Abbreviation of *National Machine Accountants Association.*

NMOS A metal-oxide semiconductor device made on a p-type substrate whose active carriers, electrons, migrate between n-type source and drain contacts.

NMR **1.** Abbreviation of NUCLEAR MAGNETIC RESONANCE. **2.** Abbreviation of NORMAL-MODE REJECTION.

n-n junction In a semiconductor device, especially an integrated circuit, the junction between two n-type regions having somewhat different properties (sometimes designated n_1 and n_2).

NO Abbreviation of NORMALLY OPEN.

No Symbol for NOBELIUM.

No. Abbreviation of NUMBER.

no-address instruction In digital computer operations, an instruction requiring no reference to storage or memory for its execution.

nobelium Symbol, No. A radioactive element produced artificially. Atomic number, 102. Atomic weight, approximately 259 (varies with isotope).

Nobili's rings See ELECTRIC RINGS.

noble Chemically inert or inactive. For example, *noble metals* oxidize less rapidly than *base metals.*

noble gas An inert rare gas (such as argon, helium, krypton, neon, or xenon). It is used in electronic glow devices.

noble metal A comparatively nonreactive metal (such as gold, silver, or platinum).

noctovision A television transmission system using infrared rays instead of visible light to scan the object. This makes it possible to televise images in complete visual darkness (hence, the prefix *nocto-*, meaning "night").

no-charge machine fault time Unproductive computer time resulting from errors or a malfunction.

nodal point See NODE.

node **1.** The terminal point at which two or more branches of a circuit meet, or a point that is common to two circuits. **2.** In a standing-wave system, a zero point or minimum point, e.g., current node. Compare LOOP, **1. 3.** A database management system expression defining the location of information about a record, user, field, etc. **4.** A point at which a satellite crosses the equator. See also ASCENDING NODE, DESCENDING NODE.

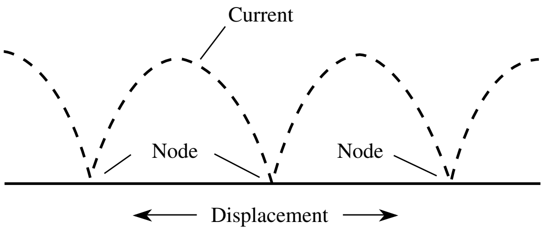

node, 2

nodules **1.** Oxide particles that protrude above the surface of magnetic tape. **2.** In a planar pattern describing radiation or pickup characteristics (as for antennas, microphones, loudspeakers), a small peak aligned in a direction other than that of the main lobe.

no-field release In the starting box for a shunt motor, the electromagnet that normally holds the arm in full-running position; it is connected in series with the field winding. When the field current is lost, the arm is released, disconnecting the armature for safety. Compare NO-VOLTAGE RELEASE.

noise **1.** A random-frequency current or voltage signal extending over a considerable frequency spectrum and having no useful purpose, unless it is intentionally generated for test purposes. **2.** Dissonant, interferential sound; unlike harmonious sound, it is disagreeable. **3.** In audio operations, unwanted hiss and/or hum. **4.** Extra bits or bytes that must be removed from digital data before it can be useful.

noise abatement The elimination or reduction of noise intensity—especially a measure in a program concerned with noise pollution in the environment.

noise analysis The measurement of the amplitude and spectral distribution of noise and the determination of its character.

noise analyzer An instrument for evaluating the nature of noise in a communications system. See, for example, NOISE METER. Noise analyzers are sometimes adapted for vibration analysis.

noise-balancing system A bridge circuit inserted between a receiver and antenna for balancing out interferential signals resulting from nearby power-line leaks or similar causes.

noise bandwidth Abbreviation, NBW. A figure obtained by dividing the area under the power-output-vs.-frequency curve of a device by the power amplitude at the noise frequency of interest.

noise behind the signal Noise caused by, but exclusive of, a signal.

noise blanker A device that cuts off one of the intermediate-frequency stages of a radio receiver during a noise pulse. The noise blanker is effective against high-amplitude impulses of short duration.

noise-canceling antenna A specialized receiving antenna system that uses two elements, one intended for receiving primarily the desired signal(s) and the other intended for receiving primarily local human-made noise. The outputs from the two elements are combined so that the noise impulses appear out of phase, but of equal amplitude. Because one element is located in a favorable position for reception of desired signals while the other element is not, the desired signals do not cancel at the receiver input. As a result, the signal-to-noise ratio is greatly improved, compared with the use of the signal antenna alone.

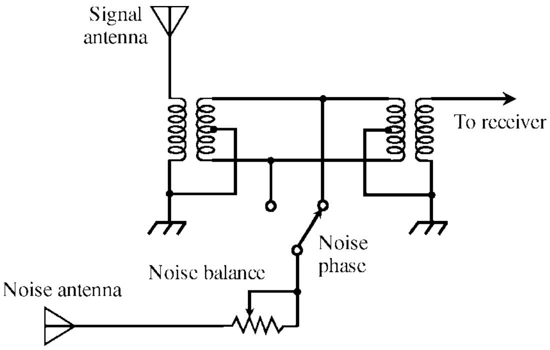

noise-canceling antenna

noise-canceling microphone A microphone that discriminates against background sounds. It is usually directional and relatively insensitive, requiring the user to talk directly into it at close range.

noise clipper A biased-diode circuit used as an automatic noise limiter. The device cuts off all signals above a predetermined amplitude on the theory that noise peaks are high-level transients in an otherwise uniform signal. Noise is reduced at the sacrifice of system reproduction fidelity.

noise criteria An expression for the level of ambient acoustic noise.

noise current Noise-generated current.

noise-current generator A noise generator that supplies a useful current. Compare NOISE-VOLTAGE GENERATOR.

noise digit A digit (usually zero) generated during normalization of a floating-point number. See NORMALIZE.

noise diode A reverse-biased semiconductor diode that produces a standard noise voltage.

noise elimination The nearly complete removal of noise effects from a system. Noise can never be eliminated altogether because the movement of electrons and atoms generates some electrical and thermal noise. However, in some digital systems, the effects of noise can be almost totally overcome. Compare NOISE SUPPRESSION.

noise equivalent power Abbreviation, NEP. The power that produces an rms signal-to-noise ratio of 1 in a detector.

noise factor For a circuit, especially a communications receiver or weak-signal amplifier, the ratio R_1/R_2, where R_1 is the signal-to-noise power ratio of an ideal circuit, and R_2 is the signal-to-noise ratio of the circuit under test. Compare NOISE FIGURE.

noise figure The NOISE FACTOR of a circuit, expressed in decibels. If N is the noise factor expressed as a ratio, then noise figure N_{dB} can be determined by $N_{dB} = 10 \log_{10} N$.

noise filter A filter designed to suppress noise that would otherwise enter an electronic circuit (e.g., a power-line noise filter).

noise floor **1.** In a receiver, the level of noise in microvolts that determines the weakest signal that can be heard or accurately received. **2.** In a spectrum analyzer, the level of noise that determines the weakest signal that will be visibly displayed.

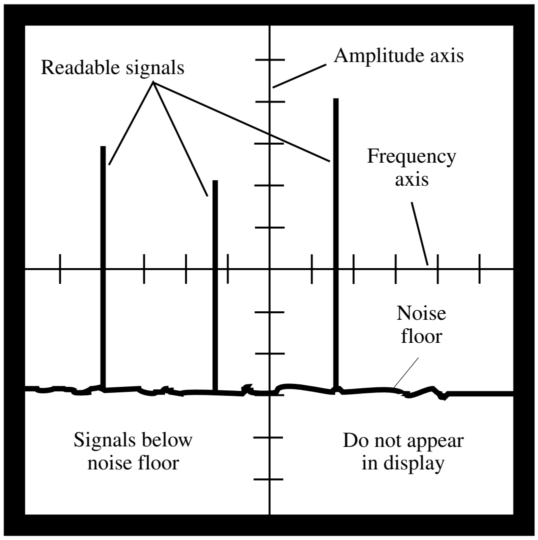

noise floor, 2.

noise generator A device for generating precise amounts of noise voltage for test purposes.

noise grade **1.** The relative level of radio-frequency background noise, over all electromagnetic frequencies, in a particular geographic location. The noise grade is generally lowest near the poles and highest near the equator. **2.** The mathematical function of relative electromagnetic noise intensity versus latitude and longitude.

noise immunity The degree to which a circuit or device is insensitive to extraneous energy—especially noise signals.

noise-improvement factor Abbreviation, NIF. For a radio receiver, the ratio SN_i/SN_o, where SN_i is the input signal-to-noise ratio and *SNo* is the output signal-to-noise ratio.

noise killer **1.** See AUTOMATIC NOISE LIMITER. **2.** See NOISE FILTER. **3.** See NOISE BLANKER.

noiseless alignment See VISUAL ALIGNMENT.

noise level **1.** The amplitude of ambient electrical noise generated outside an electronic system of interest. **2.** The amplitude of electrical noise generated in an electronic system of interest. **3.** The intensity of ambient acoustic noise.

noise limiter See AUTOMATIC NOISE LIMITER.

noise margin In a binary logic circuit, the difference between operating and threshold voltages.

noise-measuring set See NOISE METER.

noise meter An instrument for measuring acoustic noise level. It consists essentially of a sensitive, multirange voltmeter provided with a microphone, amplifier, and attenuators. The meter scale reads noise level directly in decibels.

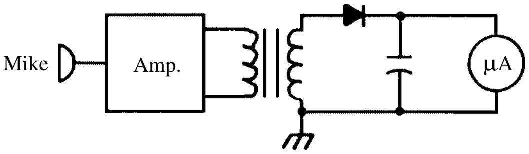

noise meter

noise power The power component of a noise signal.

noise power ratio The ratio of noise power at the output of a circuit (such as a receiver) to the noise power at the input.

noise pulse A random short-duration noise burst whose amplitude exceeds the average peak noise level.

noise quieting In a radio receiver, the reduction (in decibels) of background noise, with respect to a signal of interest.

noise ratio See NOISE POWER RATIO.

noise-reducing antenna A receiving antenna having a balanced transmission line and usually some form of noise-balancing system for reducing electrical noise picked up by the antenna.

noise reduction See NOISE SUPPRESSION.

noise residue The residual output (see NULL VOLTAGE, **1**) of a balanced bridge, caused entirely by noise.

noise silencer A noise-limiting circuit that removes noise-pulse transients with little or no effect on the signal from which the noise is removed. Compare NOISE CLIPPER.

noise source See NOISE GENERATOR.

noise spike See NOISE PULSE.

noise suppression **1.** In communications, the reduction of noise amplitude to a level that is noncompetitive with desired signals. **2.** In audio recording and reproduction, reduction of unwanted noise (e.g., hiss) to the greatest extent possible without degrading the fidelity of the desired audio. Compare NOISE ELIMINATION.

noise suppressor A device for eliminating electrical noise or reducing its amplitude. See, for example, AUTOMATIC NOISE LIMITER and NOISE BLANKER.

noise temperature At a given frequency, the temperature of a passive system that has the same noise power per unit bandwidth as that observed at the terminals of a device under test.

noise voltage The voltage component of an electrical noise signal.

noise-voltage generator A signal generator that supplies an alternating-current waveform containing random-frequency pulses of relatively uniform distribution over a given frequency spectrum. Compare NOISE-CURRENT GENERATOR.

noisy mode During normalization of a floating-point number, the generation of digits, excluding zero, as part of the fixed-point part (see NORMALIZE).

NOL **1.** Abbreviation of *National Ordnance Laboratory.* **2.** Abbreviation of *Naval Ordnance Laboratory.*

no-load current **1.** Output-electrode current (e.g., drain, plate, or collector current) when a device is not delivering output to an external load. **2.** Current flowing in the primary winding of an unloaded transformer.

no-load losses Losses in an unloaded transformer (see NO-LOAD CURRENT, **2**).

no-load speed The rotational speed of an unloaded motor.

no-load voltage The open-circuit output voltage of a power supply, amplifier, generator, or network.

nominal **1.** Named, rated, or specified. The nominal value of a speaker, for example, might be 8 ohms, even though the actual impedance value depends on the frequency of the applied signal. **2.** Approximate, and specified as a typical example only, for the purpose of identifying the operating or value range. For example, an automotive circuit might have a nominal rating of 12 volts—even though it can be operated at 10 volts to 14.6 volts.

nominal band In a facsimile signal, the waveband extending between zero and the maximum frequency of modulation.

nominal bandwidth **1.** For a filter, the difference $f_{c2} - f_{c1}$, where f_{c1} is the nominal lower cutoff frequency and f_{c2} is the nominal upper cutoff frequency. **2.** For an allocated communication channel, the total bandwidth, including upper and lower guard frequencies. **3.** The intended and specified bandwidth of a given channel, regardless of the bandwidth of the signal on that frequency at any given time.

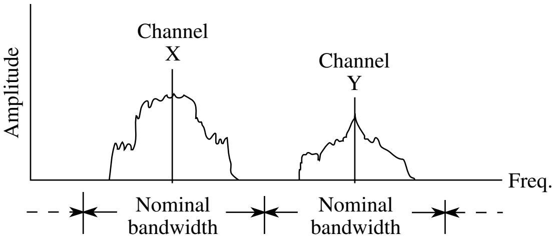

nominal bandwidth, 2

nominal capacitance The rated ("label") value of a capacitor. Also see NOMINAL VALUE.

nominal current The rated ("nameplate") value of required current or of current output. Also see NOMINAL VALUE.

nominal horsepower The rated ("nameplate") horsepower of a machine, such as a motor. Also see NOMINAL VALUE.

nominal impedance The rated impedance of a circuit or device. Also see NOMINAL VALUE.

nominal inductance The rated ("label") value of a coil's inductance. Also see NOMINAL VALUE.

nominal line pitch In a television raster, the average center-to-center separation between adjacent lines.

nominal line width **1.** For a television raster, the factor $1/n$, where n is the number of lines per unit width for the direction in which the lines progress. **2.** In facsimile, the average center-to-center separation of scanning or recording lines.

nominal power factor The rated ("nameplate") value of power factor of a device. Also see NOMINAL VALUE.

nominal power rating The rated ("nameplate") value of power output, power drain, or power dissipation. Also see NOMINAL VALUE.

nominal *Q* The rated ("nameplate") value of Q of a capacitor, inductor, transformer winding, or tank circuit. Also see NOMINAL VALUE.

nominal rating See NOMINAL, **1**, **2**.

nominal resistance The rated ("label") value of a resistor or similar device. Also see NOMINAL VALUE.

nominal speed The highest speed of a data-processing unit or system, disregarding slowdowns because of factors other than computational operations.

nominal value A named, specified, rated, or labeled value, given without reference to tolerance.

This can differ significantly from the actual value. For example, the nominal value of a capacitor might be 100 pF; but if the tolerance is ±10 percent, the actual capacitance might be any value between 90 pF and 110 pF.

nominal voltage **1.** The rated ("nameplate") value of required voltage or of voltage output. Also see NOMINAL VALUE. **2.** In a cell or battery, the MIDPOINT VOLTAGE.

nomogram See ALIGNMENT CHART.

nomograph See ALIGNMENT CHART.

nomography The geometric representation of a mathematical function or relation by means of alignment charts.

nonaccountable time The period during which a computer system is unavailable to the user (because of a power outage, for example).

nonarithmetic shift See LOGICAL SHIFT.

nonblinking meter A digital meter that does not alternate, or oscillate, between two different values when the measured parameter is between two discrete values. Instead, the display is rounded off to the nearest value, and the display remains at that value continuously.

nonblocking system In a telephone communications network, a system that ensures a circuit will be completed when necessary. That is, at no time is it impossible for a connection to be made. Under conditions of extremely heavy usage, the quality of communications might be degraded, but the connection will not be cut off.

nonbridging contact In a switch or relay, a movable contact that leaves one stationary contact before contacting another.

nonchargeable battery A primary battery (i.e., one that cannot ordinarily be recharged). An example is a battery of common zinc-carbon or alkaline cells.

noncoherent Pertaining to electromagnetic radiation in which the wave disturbances are not all precisely aligned in frequency and phase.

nonconductor See DIELECTRIC.

noncontact temperature measurement The use of infrared or optical electronic equipment to measure the temperature of bodies without touching them.

noncorrosive flux A solder flux that does not corrode the metals to which solder is applied.

noncrystalline Pertaining to materials that possess none of the characteristics of crystals. Compare CRYSTALLINE MATERIAL.

nondestructive read In digital computer and counter operation, the process of reading data without erasing it as a result. The name is also applied to the readout device.

nondestructive test Abbreviation, NDT. A test that does little or no irreversible harm to the test sample. Compare DESTRUCTIVE TEST.

nondeviated absorption Absorption that slows waves by a negligible amount; also, normal skywave absorption.

nondirectional antenna An antenna that displays equally intense radiation or equally sensitive reception in all directions within a specified plane. An example is a vertical dipole antenna, which is nondirectional in the horizontal plane.

nondirectional microphone A microphone that responds equally well to sound from any direction; an omnidirectional microphone.

nondissipative load A purely reactive load. In such a load, the only power consumed is that which is dissipated in the inherent resistance (losses) of the load.

nondissipative stub A stub that exhibits only slight losses; it consumes no power, except that dissipated in small, inherent losses. Also see STUB.

nonelectrical Not electrical in nature. The term is commonly used to designate the mechanical parts of electromechanical systems, such as robots and servomechanisms.

nonelectrolyte A substance that does not ionize in water solution. Compare ELECTROLYTE.

nonelectronic meter A meter that uses no electronic devices (such as transistors, liquid crystals, light-emitting diodes, or integrated circuits). Also called *conventional meter.*

nonequivalence operation See EXCLUSIVE-OR OPERATION.

nonerasable storage In digital-computer and data-processing operations, storage media that cannot be erased under ordinary circumstances. A common example is CD-ROM (compact-disk read-only memory).

nonferrous metal A metal or alloy that does not contain iron, and is not related to iron in the sense that it is not attracted to magnets.

nonflammable Pertaining to a material that is resistant to burning.

nonharmonic frequency A frequency that has no integral numerical relationship to another frequency of interest. Compare HARMONIC FREQUENCY.

nonharmonic oscillations Parasitic oscillations that do not occur at the fundamental frequency nor at any harmonic frequency of an oscillator or amplifier in which they appear.

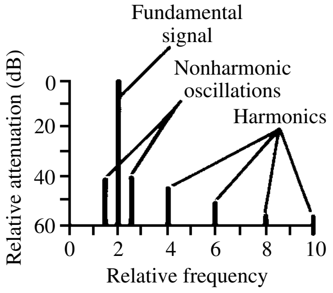

nonharmonic oscillations

nonillion The number 10^{30}, so called because when written out, the number contains nine groups of three zeros (following the first 1000).

noninductive capacitor A wound capacitor in which the edges of one of the spiral windings are connected together to minimize the inductance of the roll. Compare INDUCTIVE CAPACITOR.

noninductive resistor A wirewound resistor constructed so that the magnetic field of the coil is self-canceling. Therefore, the inductance is practically eliminated.

noninterlaced scanning In the display of a video image, the presentation of all the raster lines in a single scan. It is commonly used in cathode-ray-tube (CRT) computer monitors. This process minimizes "jerkiness" in rapidly moving images.

noninverting connection Connection to the noninverting input of a differential or operational amplifier. Also see NONINVERTING INPUT. Compare INVERTING CONNECTION.

noninverting input In a differential or operational amplifier, the input that provides an output signal in phase with the input. Compare INVERTING INPUT.

noninverting transponder In a communications satellite, a transponder in which the downlink band is "rightside-up" in frequency relative to the uplink band. That is, the highest downlink frequency corresponds to the highest uplink frequency, and the lowest downlink frequency corresponds to the lowest uplink frequency. Compare INVERTING TRANSPONDER. Also see DOWNLINK, TRANSPONDER, UPLINK.

nonionic **1.** Possessing none of the properties of ions. **2.** Electrically neutral.

nonionizing radiation Electromagnetic radiation that does not cause ionization of gases under a given set of conditions. Examples: radio signals, television signals, and visible light.

nonlinear **1.** Pertaining to components, circuits, or devices in which the instantaneous output signal amplitude is not directly proportional to the instantaneous input signal amplitude. The graph of instantaneous output versus instantaneous input is a curve, not a straight line. Example: CLASS-C AMPLIFIER. **2.** Pertaining to components, circuits, or devices in which a specified value is not directly proportional to some other specified value. Example: NONLINEAR CAPACITOR.

nonlinear bridge See VOLTAGE-SENSITIVE BRIDGE.

nonlinear capacitor A capacitor whose value varies nonlinearly with applied voltage. Also see VOLTAGE-VARIABLE CAPACITOR, **1**, **2**.

nonlinear coil See SATURABLE REACTOR.

nonlinear dielectric A material (such as processed barium-strontium titanate) whose dielectric constant varies with applied voltage.

nonlinear distortion Distortion caused by nonlinear response of an amplifier or component. This causes different parts of the signal to be amplified or transmitted by different amounts; therefore, the amplitude variations in the output signal differ from those in the input signal.

nonlinear inductor See SATURABLE REACTOR.

nonlinearity **1.** The condition of being NONLINEAR. In an amplifier, this means that the output signal is not a faithful reproduction of the input signal, and distortion occurs. **2.** A measure of the extent to which a circuit is nonlinear. Expressed as a percentage of peak-to-peak full-scale output, the maximum extent to which the output differs from a perfect reproduction of the input.

nonlinearity error An error in received signals resulting from nonlinearity in one or more of the stages in the communications circuit.

nonlinear mixing The mixing of signals as a result of the nonlinear response of a device (such as a semiconductor diode operated in its square-law region) through which they are passed simultaneously. Also see MIXER and MIXING.

nonlinear network A circuit that produces distortion in an input waveform; the output and input waves are not related by a linear function.

nonlinear quantizing A method of signal quantizing in which the intervals are not all the same size or duration.

nonlinear resistor A resistor whose value varies with applied voltage. Also see VOLTAGE-DEPENDENT RESISTOR.

nonlinear response Any response for which the corresponding plot is not a straight line; doubling the independent variable, for example, does not double the dependent variable.

nonloaded *Q* See UNLOADED *Q*.

nonmagnetic **1.** Possessing no magnetism. **2.** Incapable of being magnetized.

nonmathematical Pertaining to materials and methods that rely upon physical description and qualitative procedures instead of mathematical development, prediction, and quantitative procedures.

nonmetal An elemental material devoid of the properties exhibited by metals (e.g., luster, good ductility, electrical conductivity, heat conductivity, and malleability). Examples: carbon, phosphorus, sulfur. Compare METAL and METALLOID.

nonmetallic conduction Collectively, ionic conduction in liquids and gases, conduction in dielectrics by small leakage currents, and thermionic conduction in a vacuum.

nonmicrophonic Without microphonic properties, e.g., a nonmicrophonic integrated circuit does not produce electrical ringing when physically struck.

nonnumeric character A character that is not a numeral, i.e., a symbol or letter.

nonohmic response **1.** Nonlinear resistance or reactance. Compare OHMIC RESPONSE. **2.** See NEGATIVE RESISTANCE.

nonoscillating detector A detector devoid of positive feedback action and, therefore, unable to generate a signal on its own. Compare OSCILLATING DETECTOR.

nonplanar **1.** Existing in three spatial dimensions. **2.** Pertaining to a circuit that cannot be fabricated on a two-dimensional board without the use of jumper wires.

nonpolar **1.** Having no pole(s). **2.** Pertaining to atoms that share electrons to complete their outer shells. **3.** Not polarized nor requiring polarization. Example: a 100-pF disk ceramic capacitor is nonpolar because it can be used in circuits without consideration of voltage polarity.

nonpolar crystal A crystal in which lattice points are identical.

nonpolarized electrolytic capacitor An electrolytic capacitor that has no definite negative and positive terminals and, consequently, can be used in alternating-current circuits, as well as in direct-current circuits. See also NONPOLAR, **3**.

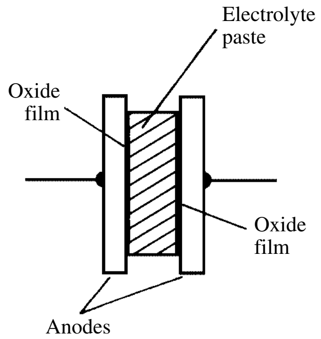

nonpolarized electrolytic capacitor

nonpolarized reactor A saturable reactor in which the lines of flux produced in the three-leg core by the coils on the two outer legs oppose each other in the center leg. When direct current (dc) is passed through the coil on the center leg to saturate the core, operation remains the same for either dc polarity. Compare POLARIZED REACTOR.

nonpolarized relay See UNPOLARIZED RELAY.

nonprint code In telegraphy, a code used to start teleprinter functions excluding printing.

nonreactive circuit A circuit containing pure resistance only.

nonrechargeable battery See NONCHARGEABLE BATTERY.

nonrecurrent Pertaining to phenomena that do not repeat periodically. Thus, a single sweep in an oscilloscope is nonrecurrent.

nonrecurrent sweep See NONREPETITIVE SWEEP.

nonregenerative detector A detector having no regenerative feedback. Such a detector is stable, but relatively insensitive. Compare REGENERATIVE DETECTOR.

nonregenerative receiver A radio receiver in which no local signal whatever is generated.

nonrepetitive phenomena See NONRECURRENT.

nonrepetitive sweep In an oscilloscope, a single horizontal sweep of the electron beam, initiated either by the operator or by the signal under observation. Also called SINGLE SWEEP. Compare RECURRENT SWEEP.

nonreset timer A timer that must be reset manually.

nonresident routine A computer routine not permanently stored in memory. Compare RESIDENT ROUTINE.

nonresonant **1.** Pertaining to a resonant circuit or device operated at some frequency other than one of its resonant frequencies. Thus, reactance is present at the operating frequency. See RESONANCE. **2.** Pertaining to a circuit or device that exhibits pure resistance (without reactance) over a wide range of frequencies.

nonresonant lines Transmission lines so dimensioned and operated that they do not resonate at the operating frequency.

nonresonant load An alternating-current load that is either purely resistive or is detuned from the fundamental and harmonic frequencies of the source from which it is operated.

nonreturn to zero In the magnetic recording of digital data, the system in which the current flowing in the write-head coil is sustained (i.e., does not return to zero) after the write pulse.

nonsalient pole A nonprojecting (often flush) pole. Compare SALIENT POLE.

nonsaturated color **1.** Visible light that consists of energy at more than one wavelength. **2.** A color that contains some white, in addition to the pure color.

nonsaturated logic A logic circuit in which transistors are prevented from saturating. This results in higher operating speed than SATURATED LOGIC using the same transistors.

nonshorting switch A multiple-throw switch that disconnects one circuit before completing another; that is, no two poles are ever connected simultaneously.

nonsinusoidal waveform A waveform whose curve cannot be represented by the equation $y = c \sin a(x + b)$, where a, b, and c are constants, y is the dependent variable (usually instantaneous amplitude or frequency), and x is the independent variable (usually time), for any real-number values of a, b, and c. Examples: BACK-TO-BACK SAWTOOTH, SAWTOOTH, and SQUARE WAVE. The COSINE WAVE is sinusoidal, being a SINE WAVE shifted in phase by 90 degrees.

nonsymmetrical wave See ASYMMETRIC WAVE.

nonsynchronous Unrelated in cyclic quality to other such qualities in the system.

nonsynchronous network A communications network in which the clocks are not all synchronized.

nonsynchronous vibrator A power-supply vibrator that is essentially a single-pole, double-throw switch providing no mechanical rectification. A separate rectifier must be used. Compare

VIBRATOR-TYPE RECTIFIER. Also see VIBRATOR-TYPE POWER SUPPLY.

nontechnical Pertaining to circuits, devices, systems, or phenomena described in lay terms, using concise graphics and little or no mathematics. An example is the simplified explanation of the operation of a spread-spectrum radio transmitter.

nontonal components See NOISE, **3**.

nontrigger voltage For a thyristor, the maximum gate-to-cathode voltage that can be applied without triggering the device. The amplitude of interferential signals, including noise, must be below this level to prevent accidental triggering.

nonuniform field An electric or magnetic field whose intensity is not the same at all points.

nonvolatile memory Memory whose data is retained even when power is removed for extended periods. This type of memory requires no backup power source. The main advantage of this type of memory is the fact that the data is not lost in case of a power interruption. Memory should not be confused with storage. Magnetic, magneto-optical, or optical disks (including hard drives) are storage media, not memory. Memory data can be stored and retrieved much faster than storage data, because memory uses no mechanical parts. Also see RANDOM-ACCESS MEMORY, READ-ONLY MEMORY. Compare VOLATILE MEMORY.

nonvolatile storage A computer storage medium in which the data does not require a source of power to be retained. Examples: MAGNETIC DISK, MAGNETIC TAPE, and COMPACT-DISK READ-ONLY MEMORY.

no-op instruction An instruction that commands a computer to perform no operation, other than to proceed to the following instruction.

NOR circuit Also called NOT-OR CIRCUIT. In computer and control operations, a circuit that delivers a zero output signal, except when two or more input signals are zero. The NOR circuit function is the inverse of that of the OR circuit.

NOR gate A gate that performs the functions of a NOR circuit.

norm The average or ambient condition.

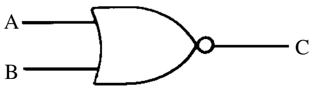

A	B	C
0	0	1
0	1	0
1	0	0
1	1	0

NOR circuit

normal **1.** Pertaining to the most commonly observed set of conditions or parameters. **2.** Standard. **3.** Perpendicular; oriented at right angles. **4.** Pertaining to a NORMAL DISTRIBUTION. **5.** Pertaining to a NORMAL SOLUTION. **6.** Pertaining to an atom at its lowest energy state. See NORMAL STATE OF ATOM.

normal curve See BELL-SHAPED CURVE.

normal distribution In a statistical evaluation, a probability distribution represented by the so-called bell-shaped curve. The maximum probability occurs at the 50-percent value.

normal-distribution curve See BELL-SHAPED CURVE.

normal electrode A standard electrode used in electrode-potential measurements.

normal fault An unintended path between the hot terminal of a load and ground.

normal fault plus grounded neutral fault A combination of NORMAL FAULT and GROUNDED NEUTRAL FAULT.

normal glow discharge In a glow-discharge tube, the discharge region between the Townsend discharge and the abnormal glow in which current increases sharply, but a constant voltage drop is maintained across the tube.

normal impedance A transducer's input impedance when the load impedance is zero.

normal induction curve A saturation curve for a magnetic material. Also see BOX-SHAPED LOOP and SATURABLE REACTOR.

normalize In computer programming, to use floating-point numbers to modify the fixed-point part of a number so that it is within a desired range.

normalized admittance The quantity $1/Z_n$, where Z_n is NORMALIZED IMPEDANCE.

normalized frequency The unitless number represented by the ratio f/f_r, where f_r is a reference frequency and f is a frequency of interest. Response plots are sometimes conveniently drawn on the basis of normalized frequency, the reference (or resonant) frequency being indicated as **1**, twice the reference frequency as **2**, etc.

normalized impedance A value of impedance divided by the characteristic impedance of a waveguide.

normally closed Abbreviation, NC. Pertaining to a switch or relay whose contacts are closed when the device is at rest. Compare NORMALLY OPEN.

normally open Abbreviation, NO. Pertaining to a switch or relay whose contacts are open when the device is at rest. Compare NORMALLY CLOSED.

normal mode Pertaining to a device or system operated in its usual or most common manner.

normal mode A state of acoustic resonance in an enclosure, such as a speaker cabinet or a room.

normal-mode rejection Abbreviation, NMR. In a digital direct-current voltmeter, the level of noise on the applied voltage that will be rejected by the instrument. Compare COMMON-MODE REJECTION.

normal position In a switch or relay, the state of the contacts when the device is at rest.

normal solution A solution, such as an electrolyte, in which the amount of dissolved material is chemically equivalent to 1 gram-atomic weight of hydrogen per liter of the solution. Compare MOLAR SOLUTION.

normal state of atom The condition in which an atom is at its lowest energy level. For the hydrogen atom, for example, the state in which the electron is in the lowest-energy orbit.

normal-through A feature in an audio PATCH BAY or PATCH PANEL that connects two sockets by default. The top socket and the one immediately below it are connected, even when a patch cord is not plugged into either of them.

northern lights See AURORA.

north magnetic pole The north pole of the equivalent bar magnet constituted by the EARTH'S MAGNETIC FIELD. The north magnetic pole lies close to the geographic north pole. Compare SOUTH MAGNETIC POLE.

north pole 1. See NORTH MAGNETIC POLE. **2.** The earth's geographic north pole. **3.** See NORTH-SEEKING POLE.

north-seeking pole Symbol, N. The so-called north pole of a magnet. When the magnet is suspended horizontally, this pole points in the direction of the earth's north magnetic pole. Compare SOUTH-SEEKING POLE.

Norton's equivalent An equivalent circuit based on NORTON'S THEOREM, replacing a Thevenin equivalent for a current-actuated device, such as a bipolar transistor. Also see THEVENIN'S THEOREM.

Norton's theorem With reference to a particular set of terminals, any network containing any number of generators and any number of constant impedances can be simplified to one constant-current generator and one impedance. The equivalent circuit will deliver to a given load the same current that would flow if the output terminals of the original circuit where short-circuited. Compare COMPENSATION THEOREM, MAXIMUM POWER TRANSFER THEOREM, RECIPROCITY THEOREM, SUPERPOSITION THEOREM, and THEVENIN'S THEOREM.

NOT In binary logic, an operation that changes high to low and vice versa. Also see NAND CIRCUIT, NOR CIRCUIT, NOR GATE, NOT CIRCUIT, and NOT-OR CIRCUIT.

NOT-AND circuit See NAND CIRCUIT.

notation The way that numbers, quantities, or formulas are represented (e.g., *binary notation*, *Polish notation*, and *scientific notation*).

notch A dip in frequency response, typical of a band-suppression (band-elimination) filter or other frequency-rejection circuit. Compare PEAK, **3**.

notch amplifier An amplifier containing a notch filter or other arrangement that permits it to reject one frequency or a given band of frequencies while passing all higher and lower frequencies.

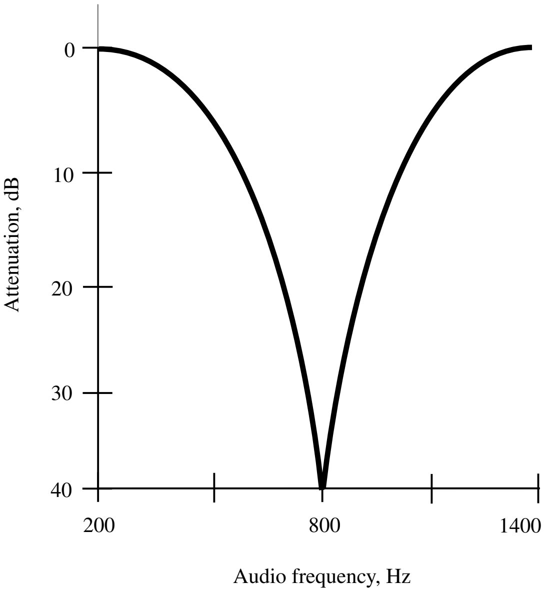

notch

notch antenna An antenna with a slot in the radiating surface, for the purpose of obtaining a directional response.

notcher See NOTCH FILTER.

notcher-peaker A circuit or device that can be set to perform either as a NOTCH FILTER or PEAK FILTER.

notch filter A circuit that exhibits high attenuation at and near a single frequency and little or no attenuation at all other frequencies. This type of device is used in some radio communications receivers, and can reduce interference caused by strong, unmodulated carriers within the passband. The notch frequency is adjustable, so that the deep null can be tuned to any frequency within the receiver passband. A properly designed circuit can produce attenuation in excess of 40 dB in the center of the notch. Some sophisticated types, especially audio designs, can provide more than 60 dB of attenuation at the notch frequency. Audio notch filters employ operational amplifiers with resistance-capacitance (RC) circuits. In some audio notch filters, the notch width (sharpness) and frequency are both adjustable. Compare BAND-REJECTION FILTER.

notch gate In radar, a gate that determines the minimum and maximum range.

notch sweep An oscilloscope sweep that expands only a small portion (notch) of the pattern on the screen, leaving the portions on either side of the notch untouched. Thus, the first dozen or so cycles might appear at the normal sweep speed, the next two cycles expanded, and the remaining two or three at normal sweep speed.

NOT circuit A logic circuit that provides an output pulse when there is no input pulse, and vice versa. Also called COMPLEMENTER, NEGATOR, and INVERTER.

note See BEAT NOTE.

notebook computer A portable personal computer, also called a *laptop computer.* It is about the size of a typical three-ring notebook, and generally contains a DISKETTE DRIVE, a HARD DISK, a MODEM, and attachments for peripherals, such as printers. It uses rechargeable batteries and can be operated for approximately two to six hours between battery charges.

NOT gate A digital circuit that inverts a logical condition—either from high (logic 1) to low (logic 0) or vice versa. Also called an *inverter.*

NOT-OR circuit A logical OR CIRCUIT combined with a NOT CIRCUIT.

novelty calculator See SPECIAL-PURPOSE CALCULATOR.

November Phonetic alphabet code word for the letter N.

novice **1.** A beginner class of amateur radio license. **2.** Any beginner or inexperienced practitioner.

no-voltage release In the starting box for a shunt motor, the electromagnet that normally holds the arm in full-running position. It is connected directly across the power line to disconnect the motor in the event of power failure. When the arm is released, it falls to its off position, thereby preventing burnout that would result if the motor were left connected to the line in the full-running position when power resumed. Compare NO-FIELD RELEASE.

noys scale A scale of apparent acoustic noise, based on a linear function instead of the more common logarithmic function.

Np **1.** Symbol for NEPTUNIUM. **2.** Abbreviation of NEPER.

N_p Symbol for number of primary turns in a transformer.

n-phase system A polyphase system having *n* phases.

npin transistor A junction transistor having an intrinsic layer between a p-type base and an n-type collector. The emitter is a second n-type layer on the other side of the base.

N plant See NUCLEAR POWER PLANT.

n-plus-one address instruction A computer program instruction containing two addresses, one of which specifies the location of an upcoming instruction to be executed.

NPM Symbol for *counts per minute.*

npnp device A semiconductor switching device having three junctions. Examples: FOUR-LAYER DIODE, and SILICON-CONTROLLED RECTIFIER. Also called *pnpn device.*

npn transistor A bipolar transistor in which the emitter and collector layers are n-type semiconductor material, and the base layer is p-type semiconductor material. Compare PNP TRANSISTOR.

NP0 Abbreviation of NEGATIVE POSITIVE ZERO.

NP0 capacitor A fixed capacitor exhibiting temperature-compensating ability over a wide temperature range, in which the coefficient has negative, positive, and zero values.

NPS Symbol for *counts per second.*

N radiation X rays emitted as a result of an electron becoming an N electron.

NRD Abbreviation of NEGATIVE-RESISTANCE DIODE.

N region See N LAYER.

NRZ Abbreviation of NONRETURN TO ZERO.

N_s Symbol for number of secondary turns in a transformer.

ns Abbreviation of NANOSECOND.

N scan See N DISPLAY.

N scope Colloquialism for a radar set using an N DISPLAY.

nsec Alternate abbreviation of NANOSECOND.

Ns/m^2 Newton-seconds per meter squared, the unit of dynamic viscosity.

n-space A coordinate system in *n* variables. It is generally of mathematical interest. The coordinates are written ($x_1, x_2, x_3, \ldots, x_n$) and are called *ordered n-tuples.*

NSPE Abbreviation of *National Society of Professional Engineers.*

NTC Abbreviation of NEGATIVE TEMPERATURE COEFFICIENT.

nth harmonic An unspecified harmonic, having a frequency of *n* times the fundamental frequency, where *n* is some positive integer. Also see HARMONIC and HARMONIC FREQUENCY.

nth term An unspecified term in a mathematical sequence or series.

NTP Abbreviation of NORMAL TEMPERATURE AND PRESSURE.

NTSC Abbreviation of *National Television Standards Committee.*

NTSC color signal The color-television signal specified by the National Television Systems Committee. In the signal, the phase of a 3.58-MHz signal varies with the instantaneous hue of the transmitted color, and the amplitude varies with the instantaneous saturation of the color.

NTSC triangle On a chromaticity diagram, a triangle whose sides encompass the range of colors obtainable from the additive primaries.

NTSC-type generator A special radio-frequency signal generator for color-television tests. It provides separate, individually selected color bars that are fully saturated. The signals are strictly in accordance with NTSC standards.

n-type conduction In a semiconductor, current flow consisting of electron movement. Compare P-TYPE CONDUCTION.

n-type material Semiconductor material that has been doped with a donor-type impurity and, consequently, conducts a current via electrons. Germanium, for example, when doped with arsenic, becomes n-type. Compare P-TYPE MATERIAL.

n-type semiconductor See N-TYPE MATERIAL.

nuclear battery See ATOMIC BATTERY.

nuclear bombardment In nucleonics, the bombarding of the nucleus of an atom with subatomic particles, usually neutrons.

nuclear charge The net positive charge of the nucleus of an atom.

nuclear clock A chronometer based on the rate of disintegration of a radioactive material.

nuclear energy Energy resulting from the splitting of the nucleus of an atom or from the fusion of nuclei. Also see ATOMIC ENERGY, ATOMIC POWER, NUCLEAR FISSION, NUCLEAR FUSION, NUCLEAR REACTOR, and NUCLEUS.

nuclear fission A nuclear reaction resulting from the bombardment of nuclei in the atoms of certain radioactive materials. The bombardment with neutrons creates two new nuclei (by splitting) and several new neutrons that split several other nuclei, producing still more nuclei and neutrons, etc. The result is a *chain reaction* that can lead to a violent explosion if not checked. Compare NUCLEAR FUSION.

nuclear force A strong attraction that holds together pairs of nucleons in an atomic nucleus. This prevents an electric charge of protons from driving the nucleus apart. Nuclear force acts only over very minute distances. At greater distances, electrostatic repulsion is stronger.

nuclear fusion A nuclear reaction resulting from the violent collision of the nuclei of the atoms of a hydrogen isotope (such as deuterium) at extremely high temperature. The process produces more energy than does NUCLEAR FISSION, and leaves no hazardous radioactive waste.

nuclear magnetic resonance An atomic phenomenon in which a particle, such as a proton, in a steady magnetic field “flips over” when an alternating magnetic field is applied perpendicular to the steady field.

nuclear magnetic resonance imaging Abbreviation, NMRI. The use of NUCLEAR MAGNETIC RESONANCE effects to produce a picture of internal body organs. Using computers, three-dimensional renditions can be generated. It is useful in medicine as a diagnostic aid.

nuclear medicine A branch of medicine involving the use of radioactive isotopes in diagnosing and treating disease. A radioactive isotope is put inside the body and it tends to accumulate in certain areas. Abnormal concentration of radioisotopes might indicate abnormal body activity in a certain area.

nuclear pile See NUCLEAR REACTOR.

nuclear power plant A power-generating plant using a NUCLEAR REACTOR.

nuclear reaction **1.** A reaction in which a heavy atomic nucleus is split into two or more lighter nuclei, with an accompanying release of radiant energy. Also called NUCLEAR FISSION. **2.** A reaction in which two or more light nuclei combine to form a heavier nucleus, accompanied by the release of radiant energy. Also called NUCLEAR FUSION.

nuclear reactor **1.** A device in which nuclear fission can be initiated and controlled. At the center of the reactor is a core of nuclear fuel, such as a fissionable isotope of uranium. The core is surrounded by a graphite moderator jacket that is, in turn, surrounded by a coolant jacket; the whole is surrounded by a thick concrete shield. Neutron-absorbing rods are inserted through various walls to different depths in the fuel to control the reaction. Also called *atomic pile.* **2.** A controlled nuclear fusion device, not yet perfected, but under development. It would provide all the benefits of an atomic pile (fission reactor), but would be more efficient and would not produce hazardous radioactive waste.

nuclear recoil An observable vibration of an atomic nucleus when it disintegrates.

nuclear resonance The condition wherein a nucleus absorbs a gamma ray emitted by an identical nucleus.

nuclear service robot A remotely controlled (teleoperated) robot used for general work in environments where the level of radioactivity is too high for humans (e.g., the maintenance of a nuclear reactor). It could also be used, if necessary, for such tasks as disarming nuclear warheads and cleaning up after a nuclear accident.

nucleon A proton or neutron in the nucleus of an atom.

nucleonics The branch of physics concerned with nucleons and nuclear phenomena. The name is an acronym for *nuclear electronics.*

nucleon number See MASS NUMBER.

nucleus The center or core of an atom. Contains neutrons, protons, and other particles. The net electric charge of the nucleus is positive, and is equal to the sum of the negative charges of the orbital electrons of the atom.

null **1.** The condition of zero output current or voltage resulting from adjusting or balancing a circuit, such as a bridge. **2.** A local minimum in an interference pattern or directivity pattern.

null balance In potentiometric-measuring circuits for comparing one voltage to another, the balance condition in which no current flows through the galvanometer.

null current In potentiometric-measuring circuits for comparing one voltage to another, the galvanometer current remaining at null when the null point is not fully zero.

null detection Direction finding by means of an antenna with a bidirectional or unidirectional null response.

null detector See BRIDGE DETECTOR.

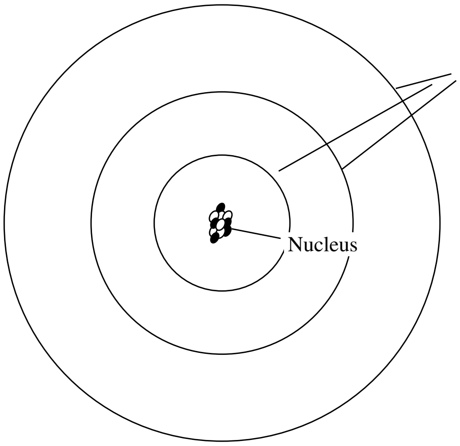

nucleus

null frequency The frequency at which a frequency-sensitive circuit, such as a Wien bridge or twin-tee network, can be balanced.

null meter See BRIDGE DETECTOR.

null method See ZERO METHOD.

null point In a balanced circuit, such as a bridge or potentiometer, the point of zero output voltage (or current) or minimum output voltage (or current).

null potentiometer **1.** The variable resistor that constitutes one arm of a four-arm bridge and is used to balance the bridge. **2.** A potentiometric circuit using the null method to compare one voltage with another. Also see POTENTIOMETER, **2**.

null setting **1.** The setting of a bridge circuit or other null device that balances the circuit. **2.** The electrical zero setting of an electronic voltmeter.

null voltage **1.** In a conventional bridge, the output voltage remaining when the bridge is set for its best null. **2.** For a voltage-sensitive bridge, the input voltage that will produce zero output voltage.

number The mathematical representation of a quantity. It is generally used in electronics to denote coefficients, magnitudes, component values, frequencies, etc.

number cruncher A computer with great computational power, but one not necessarily able to process large amounts of data (such as payroll information).

number system A systematic sequence of numbers based on a radix and a logical arrangement. See, for example, BINARY NUMBER SYSTEM and DECIMAL NUMBER SYSTEM.

numeral A member of a digit set in a number system.

numerical analysis A mathematical approach to solving problems numerically, including finding the limits of error in the results.

numerical code A code having a character set restricted to digits.

numerical control A method of programming computer-controlled mechanical devices, used in some early robots. An automated system in which number sequences fed to a digital computer cause it to control machines or processes in a manufacturing operation.

nV Abbreviation of NANOVOLT.

nW Abbreviation of NANOWATT.

nybble A piece of digital information that is larger than a bit and smaller than a byte. Compare GULP.

Nylon The trade name for a synthetic fiber-forming polyamide, useful for electrical insulation.

Nyquist criterion of stability With reference to a NYQUIST DIAGRAM for a feedback amplifier, the amplifier is stable if the polar plot of loop amplification for all frequencies from zero to infinity is a closed curve that neither passes through nor encloses the point $1 + j0$.

Nyquist diagram A graph of the performance of a reactive feedback system (such as a degenerative amplifier) that depicts the variation of amplitude and phase of the feedback factor with frequency. The plot is polar and accounts for the real and imaginary components.

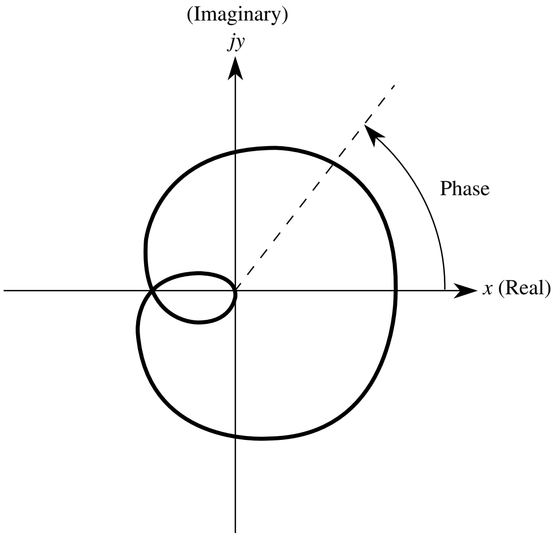

Nyquist diagram

Nyquist noise rule The power dissipated in a resistor because of thermal noise at a given frequency. The derivative of frequency, with respect to power, is equal to the absolute temperature times the Boltzmann constant.

N zone See N LAYER.

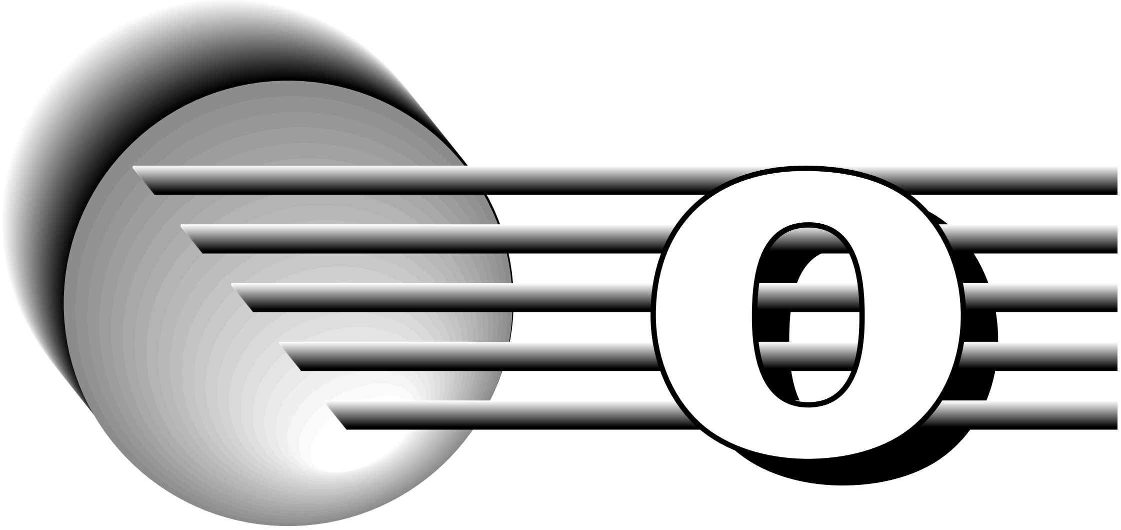

O **1.** Symbol for OXYGEN. **2.** Abbreviation of OUTPUT.

o **1.** Symbol for OUTPUT. (Also, OUT; both are used as subscripts.) **2.** Symbol for ORIGIN.

O_2 Symbol for OXYGEN.

O_3 Symbol for OZONE.

OAT Abbreviation of OPERATING AMBIENT TEMPERATURE.

object code **1.** In a computer system, the machine-language output of the compiler, directed to the computer. **2.** In a computer, the high-level output of the compiler, directed to the operator. **3.** The assembly language, directed to the compiler for translation between machine language and high-level language and vice versa.

object language The computer language that a compiler derives from a high-level (source) language (such as C++, LISP, etc.); it is usually machine language, but it can be an intermediate code that requires further conversion.

object-oriented graphics Also called *vector graphics.* In computer graphics, the use of equations to represent curves in a coordinate plane to define shapes, rather than defining the shapes pixel-by-pixel.

object-oriented language A computer program that uses on-screen objects, called icons, to represent commands. A movable device, such as a mouse or trackball, is used to move an arrow or other pointer to the icon; then a button is pressed to carry out the function indicated by the icon.

object-oriented programming Abbreviation, OOP. A computer programming language that builds sophisticated programs from basic programs called modules.

object program A machine or high-level language version of a user's computer program, as produced by a compiler.

object recognition In robotic systems, any method used to identify specific objects, according to characteristics, such as shape, texture, weight, etc. Common schemes use bar-code labels and machine vision. More-complex methods make use of pattern-recognition programs.

oblique-incidence transmission Transmission of radio signals via ionospheric reflection.

oblique mode In acoustics, a resonance within a room that involves all four walls, the floor, and the ceiling.

oboe A system of radar navigation in which a pair of ground stations measures the distance to an airborne transponder beacon and then transmits the information to the aircraft.

obsolescence-free Pertaining to a design or process that is not likely to become obsolete in the near future. Compare OBSOLESCENCE-PRONE.

obsolescence-prone Pertaining to a design or process subject to being soon outdated. Compare OBSOLESCENCE-FREE.

OBWO Abbreviation of O-TYPE BACKWARD-WAVE OSCILLATOR.

o/c Abbreviation of OPEN CIRCUIT.

occluded gas Gas that has been absorbed or adsorbed by solid material, such as glass or metals, and that must be eliminated during the evacuation of an electronic device, such as a vacuum tube. Also see OUTGASSING.

occultation **1.** The passage of the moon or other planetary body in front of a more distant celestial object, resulting in the cutting off of electromag-

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

netic radiation from that object. **2.** The eclipsing of one object by another.

occupied band A frequency band used by at least one transmitting station regularly.

occupied bandwidth For a given emission, the continuous band of frequencies ($f_2 - f_1$) for which the mean (average) radiated power above f_2 and below f_1 is 0.5 percent of the total mean radiated power.

occupied orbit In an atom, an orbit in which an electron is present.

Ocean Phonetic alphabet code word for the letter O.

OCR Abbreviation of OPTICAL CHARACTER RECOGNITION.

oct Abbreviation of OCTAL.

octal Abbreviation, oct. Based on the number eight. See, for example, OCTAL NUMBER SYSTEM.

octal digit One of the figures in the group 0 through 7 used in the OCTAL NUMBER SYSTEM.

octal notation See OCTAL NUMBER SYSTEM.

octal number system The base-eight system of number notation. It uses digits 0 through 7. Compare BINARY NUMBER SYSTEM and DECIMAL NUMBER SYSTEM. The octal system is often used as shorthand for otherwise-cumbersome binary numbers. The binary number is separated into groups of three digits from right to left; each such group is then converted into its decimal equivalent, with the result being the octal form of the binary number [e.g., binary 111 001 011 = octal 713 (111 = decimal 7; 001 = decimal 1; and 011 = decimal 3)].

Decimal	*Octal*
0	0
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	10
9	11
10	12
11	13
12	14
13	15
14	16
15	17
16	20

octal number system

octal-to-decimal conversion Conversion of numbers in the octal number system to numbers in the more-familiar decimal number system. This is done by expressing the octal number serially in powers of eight. Thus, octal 12 = $(1 \times 8^1) + (2 \times 8^0) = 8 + 2$ = decimal 10.

octant One-eighth of a circle; therefore, 45 degrees, or 0.5 quadrant.

octave **1.** The range of frequencies between a given frequency (f) and twice that frequency ($2f$). **2.** The range of frequencies between a given frequency (f) and half that frequency ($f/2$).

octave band A band of frequencies one octave wide (see OCTAVE).

octave-band noise analyzer A noise analyzer having bandpass-filter channels whose center frequencies are one octave apart.

octave-band pressure level Sound pressure level within an OCTAVE BAND.

octave pressure level See OCTAVE-BAND PRESSURE LEVEL.

OCTL Abbreviation of OPEN-CIRCUITED TRANSMISSION LINE.

octonary signal An eight-level signaling code.

octonary system See OCTAL NUMBER SYSTEM.

OD Abbreviation of OUTSIDE DIAMETER.

odd-even check A method of checking the integrity of data transferred in a computer system, in which each word carries an extra digit to show whether the sum of ones in the word is odd or even.

odd harmonic In a complex waveform, a harmonic that is an odd-numbered multiple of the fundamental frequency (e.g., third harmonic, fifth harmonic, etc.). Compare EVEN HARMONIC.

odd-harmonic intensification In a complex waveform, emphasis of the amplitude of odd harmonics, with respect to that of even harmonics, a property of some multivibrators and nonsinusoidal waves.

odd line In a conventional television picture, one of the 262.5 odd-numbered horizontal lines scanned by the electron beam in developing the odd-line field. Compare EVEN LINE.

odd-line field On a conventional television screen, the complete field obtained when the electron beam has traced all the odd-numbered lines. Compare EVEN-LINE FIELD.

odd-line interlace See ODD-LINE FIELD.

odd parity check A computer check for an odd number of ones or zeros in digital data.

Odex Trade name for a series of autonomous robots developed by Odetics, Inc. They use legs for locomotion. Noted for their ability to maneuver in places that most robots cannot reach.

odograph An electromechanical or electronic plotter that traces the path of a vehicle on a map, or on the image of a map as portrayed on a display screen.

odometer An electromechanical device that indicates the speed of, and distance covered by, a moving vehicle or robot. Some such devices give a constant position indication, via mathematical integration of the measured velocity (speed and direction), relative to time. It can function in one, two, or three dimensions.

odometry A method of speed, velocity, and/or position sensing. It is commonly used in mobile vehicles and mobile robots. The most-sophisticated systems give a constant indication of position on the basis of a starting point (origin) and mathematical integration of the velocity (speed and direction), with respect to time. It can function in one, two, or three dimensions.

Oe Symbol for OERSTED.

oersted Symbol, Oe. The cgs unit of magnetic field intensity; 1 Oe = 79.58 A/m. Formerly the unit of reluctance.

off-air alarm A device that gives a visible or audible indication when the carrier of a transmitter is lost. In its most rudimentary form, the device consists of a radio-frequency relay that actuates a bell, buzzer, horn, or lamp.

off-center display A radar display whose center point does not correspond to the location of the antenna.

off-center-fed antenna An antenna in which a feeder is attached to one side of the center point of the radiator. See, for example, WINDOM ANTENNA.

off-center feed The connection of a feed line to an antenna radiator at some point other than the physical center of the radiator.

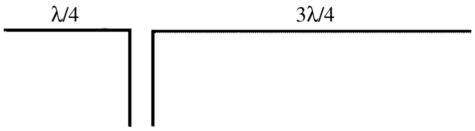

off-center feed

off delay The interval during which a circuit remains on after the control signal has been switched off. Compare ON DELAY.

off-ground **1.** An above- or below-ground operating voltage. **2.** The ground return for a voltage, as defined in **1**, not common with the system ground.

offhook In a telephone system, the condition in which the receiver is removed from its receptacle, or in which the line is otherwise engaged (e.g., a fax machine or modem is in use).

off isolation A measure of the extent to which an open switch isolates the output from the input, specified as the output voltage divided by the input voltage. For percentage, multiply this quantity by 100. The ideal value is zero (or zero percent).

off-limit In a stepping relay, a condition in which the armature has gone past the limit of its travel.

offline **1.** In a computer installation, equipment that is not controlled by the central processor. **2.** In computers and data processing, operations that are not under the direct real-time control of a central processor. **3.** A computer memory facility or device not connected to a central processor. **4.** In personal computing, the use of a computer by itself (i.e., not connected in a network). Compare ONLINE.

offline editing In video production, a process in which copies are made of the original recording; these copies are edited in a trial-and-error scheme to develop the final presentation. Compare ONLINE EDITING.

offloading The use of computers and/or robots to perform repetitive, mundane tasks, leaving people free to do more interesting things. Example: a personal robot that mows the lawn.

off-on operation See ON-OFF OPERATION.

off period **1.** The interval during which an on-off circuit or device is off. **2.** The time during which an equipment is shut down.

offset An imbalance between the halves of a normally symmetrical circuit, such as that of a differential amplifier. Also see OFFSET CURRENT and OFFSET VOLTAGE.

offset adjustment range The change in offset voltage, in millivolts or in microvolts, that can be affected by means of the external offset adjustment circuit(s).

offset current For an operational amplifier, the input current when the offset voltage is zero.

offset voltage For an operational amplifier, the particular value of direct-current bias voltage required at the input to produce zero output voltage.

off state **1.** The condition of an on-off circuit or device, such as a flip-flop, that is off. Compare ON STATE, **1**. **2.** The condition in which a circuit or device is shut down. Compare ON STATE, **2**.

off-state voltage The voltage drop across a semiconductor device, such as a diode, rectifier, or thyristor, when the device is in its normal off (nonconducting) state. Compare ON-STATE VOLTAGE.

off-target jamming In radio or radar jamming, the use of a remote jamming transmitter that will not betray the location of the base station.

off time The period during which no useful work is being performed, as of equipment when it is not functioning because of a circuit breakdown.

OGL Abbreviation of OUTGOING LINE.

OGO Abbreviation of *orbiting geophysical observatory.*

ohm Symbol, Ω. The basic unit of resistance, reactance, or impedance. A resistance of 1 ohm passes a current of 1 ampere in response to an applied emf of 1 volt.

ohmage Electrical resistance or resistivity expressed in ohms.

ohm-centimeter A unit of volume resistivity (see RESISTIVITY): the resistance of a centimeter cube of the material under measurement. Also see MICROHM-CENTIMETER.

ohmic component A resistive or reactive circuit or device exhibiting an OHMIC RESPONSE.

ohmic contact A usually very-low-resistance connection between two materials that provides bilateral linear conduction. It exhibits none of the properties of a rectifying junction or a nonlinear resistance.

ohmic heating **1.** Heating caused by current passing through a resistive material (i.e., I^{2R} losses in the material). **2.** In an electric field, heat generated by charged particles when they collide with other particles.

ohmic loss Loss resulting from the direct-current resistance in a circuit or transmission line.

ohmic region The portion of the response curve of a negative-resistance device that exhibits positive (ohmic) resistance. The *E-I* curve of a tunnel diode, for example, has two such positive-slope regions with a negative-slope (negative-resistance) region between them.

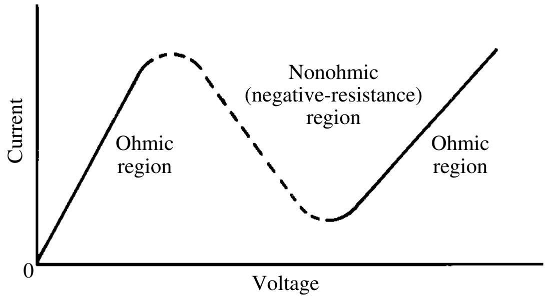

ohmic region

ohmic resistance A resistance exhibiting OHMIC RESPONSE.

ohmic response Response that follows OHM'S LAW: $I = E/R$. In strictly ohmic devices, neither resistance nor reactance changes with current or voltage. Compare NONOHMIC RESPONSE.

ohmic value Electrical resistance expressed in ohms, or in multiples or fractions of ohms (kilohms, megohms, milliohms, etc.)

ohmmeter An instrument for the direct measurement of electrical resistance. It usually consists of a milliammeter or microammeter, a battery, and several switchable resistors having very close tolerances. The scale is calibrated in ohms; the switch selects multiplier factors (e.g., ×100, ×10,000, ×1,000,000). The scale is usually reversed (i.e., 0 is at the extreme right and "infinity" is at the extreme left).

ohmmeter zero **1.** The condition of proper adjustment of an ohmmeter, indicating zero resistance for a direct short circuit. **2.** The potentiometer, or other control, used for adjusting an ohmmeter to obtain a reading of zero with a short circuit.

ohm-mile A rating meaning 1 mile of wire having a resistance of 1 ohm.

ohms adjust The rheostat or potentiometer used to set the pointer of an ohmmeter before it is used to take resistance readings.

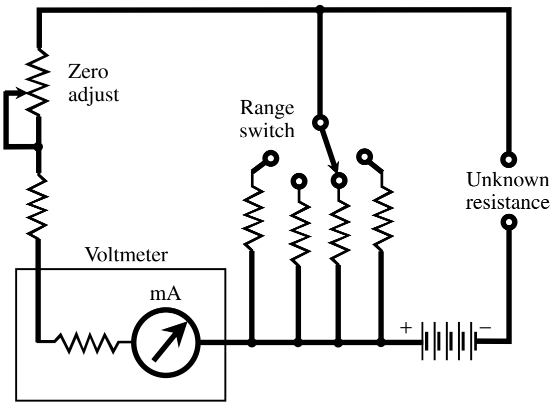

ohmmeter

Ohm's law The statement of the relationship among current, voltage, and resistance. In a direct-current circuit, current varies directly with voltage and inversely with resistance: $I = E/R$, where I is the current in amperes, E is the voltage in volts, and R is the resistance in ohms. For alternating current, Ohm's law states that $I = E/X = E/Z$, where X is reactance and Z is impedance.

ohms per square The resistance (in ohms) between two parallel edges of a square of thin-film resistance material.

ohms-per-volt A specification that indicates the sensitivity and impedance of a voltmeter. In general, the higher the rating, the better. When measuring voltage in high-impedance circuits, the rating should be as high as possible. Field-effect-transistor (FET) voltmeters and vacuum-tube voltmeters have the highest ratings.

oil-burner control An electronic system for starting and stopping the operation of an oil burner to prevent puffback and to interrupt the supply when the flame becomes erratic.

oil calorimeter A calorimeter used to measure power in terms of the rise in temperature of oil heated by the electrical energy of interest.

oil capacitor A capacitor impregnated or filled with oil, such as high-grade castor or mineral oil. Also see OIL DIELECTRIC.

oil circuit breaker A circuit breaker filled with a high-grade insulating oil for cooling and arc elimination.

oil-cooled transformer A heavy-duty transformer, through which oil is circulated for heat removal and arc prevention.

oil dielectric A highly refined oil used as an electrical insulating material (e.g., between the plates of a capacitor). Familiar examples are *castor oil*, *mineral oil*, and the synthetic oil *chlorinated diphenyl.*

oil diffusion pump See OIL PUMP.

oiled paper Insulating paper impregnated with oil for waterproofing and to increase its dielectric strength.

oil-filled cable A cable whose insulation is impregnated with oil that can be maintained at a constant pressure.

oil-filled capacitor See OIL CAPACITOR.

oil-filled circuit breaker See OIL CIRCUIT BREAKER.

oil-filled transformer A transformer whose case is filled with an insulating oil.

oil fuse cutout A fuse cutout that is filled with an insulating oil. Compare OPEN-FUSE CUTOUT.

oil-immersed transformer See OIL-FILLED TRANSFORMER.

oil-impregnated capacitor See OIL CAPACITOR.

oil pump A vacuum diffusion pump using oil instead of mercury. Also see DIFFUSION PUMP.

oil switch A switch enveloped by insulating oil.

OLRT Abbreviation of *online real time* (operation).

OM Amateur radio jargon for OLD MAN: chief (male) operator, or husband of female operator.

omega A phase-dependent radionavigation system using single-frequency, time-shared, ICW transmissions from two or more locations.

omnibearing A navigational bearing obtained by means of OMNIRANGE.

omnibearing converter An electromechanical device in which an OMNIRANGE signal and vehicle heading information are combined, the output being a signal that is fed to a meter.

omnibearing indicator Abbreviation, OBI. An omnibearing converter with a dial and pointer.

omnibearing line In an OMNIRANGE system, one of the imaginary lines extending from the geographic center of the omnirange.

omnibearing selector A device that can be set manually to a selected omnibearing.

omniconstant calculator A calculator that adds or multiplies numbers in succession in such a manner as to arithmetically or geometrically increase the exponent as a single button is repeatedly pressed.

omnidirectional **1.** Pertaining to a device that responds equally well to acoustic or electromagnetic energy from any direction in three dimensions. **2.** Pertaining to a device that radiates acoustic or electromagnetic energy equally well in any direction in three dimensions. **3.** Also, NONDIRECTIONAL. Pertaining to an antenna that intercepts or radiates equally well in any azimuth (horizontal) direction.

omnidirectional antenna See NONDIRECTIONAL ANTENNA.

omnidirectional hydrophone A hydrophone that picks up underwater sounds coming from any direction.

omnidirectional microphone A microphone that picks up sounds coming from any direction.

omnidirectional radio range See OMNIRANGE.

omnidirectional range See OMNIRANGE.

omnidirectional range station Abbreviation, ORS. A radionavigation station for OMNIRANGE service.

omnigraph A Morse-code generator that operates via marked or perforated paper tape.

omnirange A radionavigation system in which each station in a chain broadcasts a beam in all directions. It usually operates at very-high frequencies (VHF) or ultra-high frequencies (UHF). Pilots of aircraft home on a particular station by tuning it in and noting its bearings.

OMR Abbreviation for OPTICAL MARK RECOGNITION.

on air See ON THE AIR.

on-call channel An assigned radio channel of which exclusive, full time use is not demanded.

on-course curvature In navigation, the rate at which the course of a vehicle deviates with reference to the distance along the true course.

on-course signal A single-tone-modulated signal indicating to an aircraft pilot following a radio beam that the flight is substantially on course.

on current See ON-STATE CURRENT.

on delay An interval during which a circuit remains off after an actuating signal has been supplied. Compare OFF DELAY.

on-demand system A system, especially in computer and data-processing operations, that delivers information or service immediately upon request.

ondograph An electromechanical device that graphically draws alternating-current waveforms on paper.

ondoscope A radio-frequency (RF) energy detector that consists of a neon bulb attached to the end of an insulating rod. When a bulb is held in an intense RF field, the field energy ionizes the gas in the bulb, causing it to glow without direct connection to the RF circuit.

one-address code In computer programming, a code in which the address in an instruction refers to only one memory location.

one condition See ONE STATE.

one-digit adder See HALF ADDER.

one-element rotary antenna A directional antenna consisting of a radiator only (no directors or reflectors) that can be rotated. A straight, rigid, half-wave rotatable dipole is the most common configuration.

one-for-one compiler A compiler that generates one machine language instruction from one source language instruction.

one-input terminal In a flip-flop, the input terminal that must be energized to switch the circuit to its logic 1 output.

one-level address See ABSOLUTE ADDRESS.

one-level subroutine In a computer program, a subroutine in which no reference is made to other subroutines.

one-lunger Colloquialism for a radio transmitter consisting entirely of a one-transistor oscillator.

one output See ONE STATE.

one-output signal The signal that results from reading a computer memory unit that is in the logic 1 state.

one-output terminal In a flip-flop, the output terminal that is energized when the circuit is in its logic 1 state.

one-plus-one address A method of computer programming in which instructions contain two addresses and an operation, the addresses referring to the location of the next instruction and the location of the data to be used.

ones complement Binary notation in the *radix-minus-one-complement form.*

one shot See MONOSTABLE MULTIVIBRATOR.

one-shot circuit See MONOSTABLE MULTIVIBRATOR.

one-shot multivibrator See MONOSTABLE MULTIVIBRATOR.

one-sided wave A waveform consisting of only negative or positive half-cycles. Example: a rectified alternating-current signal.

one state Also, *logic 1 state.* The high, on, or true logic state of a bistable device, such as a flip-flop. Compare ZERO STATE. In binary notation, the one state is represented by the digit 1.

one-third-octave band A frequency band that is ⅓ octave wide; that is, the difference between the upper-frequency limit (f_2) and the lower-frequency limit (f_1) is one-third of an octave. Also see OCTAVE BAND.

one-to-one assembler An assembler computer program that produces a machine language instruction as a result of translating a source-language statement.

one-to-one correspondence A mapping between two sets *A* and *B* so that every element in set *A* has exactly one correspondent in *B*, and every element in *B* has exactly one correspondent in *A*.

O network A four-impedance network containing two series (upper and lower) arms and two shunt (input and output) arms.

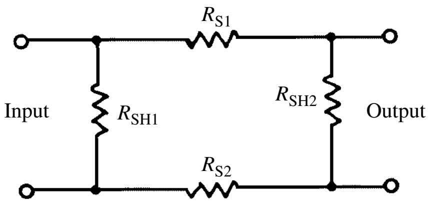

O network

one-way communication **1.** The transmission of a message to one or more stations that receive only. Compare TWO-WAY COMMUNICATION. **2.** See BROADCASTING.

one-way conduction See UNILATERAL CONDUCTIVITY.

one-way radio See ONE-WAY COMMUNICATION.

one-way repeater In wire telephony, a device that amplifies and retransmits a signal in the direction the signal was traveling when it arrived at the repeater. Compare TWO-WAY REPEATER.

one-way valve A diode or rectifier (British variation).

on interval See ON TIME.

online **1.** Descriptive of equipment under the control of a central processor. **2.** Pertaining to operations being controlled by a central processor. **3.** Relating to a computer storage device independent of a central processor. **4.** In personal computing, the use of a computer in a network (e.g., with a modem for operation on the Internet). Compare OFFLINE.

online data reduction Processing data as it enters a computer system.

online editing In video production, the final phase of editing. The original recording (not copies) is used to prepare the presentation, based on the trials that were performed using the copies. Compare OFFLINE EDITING.

online service A computer network accessible via telephone lines, television cable systems, and/or radio. It is generally available to owners of personal computers. It provides such services as electronic mail, special-interest forums, news, weather, sports, shopping, and general information.

on-off keying Keying, as in radiotelegraphy and wire telegraphy, by switching a signal source on and off to form dots and dashes or a binary code, rather than alternately changing the amplitude or frequency of the signal.

on-off operation A switching operation, especially that performed by nonmechanical (fully electronic) circuits.

on-off ratio **1.** For a circuit or device, the ratio of off time to on time. **2.** For a pulse, the ratio of pulse duration to the dead space between pulses.

on-off switch **1.** In electronic equipment, the switch by which the equipment can be started or stopped. It can be, but is not necessarily, the power-line or B-plus switch. **2.** An electronic circuit or stage that is designed to operate as a conventional switching element when triggered by an appropriate signal.

on period See ON TIME.

ONR Abbreviation of *Office of Naval Research.*

on resistance See ON-STATE RESISTANCE.

on state **1.** For a switch or switching device (such as a flip-flop), the condition of the device when it conducts current or delivers an output voltage. Compare OFF STATE, **1**. **2.** The condition of a circuit or device that is activated for operation. Compare OFF STATE, **2**.

on-state current The current flowing through a semiconductor device (such as a diode, rectifier, or thyristor) when it is conducting. Also see ON-STATE VOLTAGE.

on-state resistance The resistance of a voltage-dependent resistor that is conducting current.

on-state voltage The voltage drop across a semiconductor device (such as a diode, rectifier, or thyristor) when the device is conducting current. Also see ON-STATE CURRENT.

on the air **1.** The state of a radio station that is transmitting. **2.** In broadcasting, the condition in which the transmitter is operational and the sound and/or video from the studio is being disseminated.

on time The length of time a switch or switching device (such as a flip-flop) remains on.

on voltage See ON-STATE VOLTAGE.

OOP Abbreviation of OBJECT-ORIENTED PROGRAMMING.

op **1.** Abbreviation of OPERATE. **2.** Abbreviation of *operator* (see OPERATOR, **1**). **3.** Abbreviation of OPERATION. **4.** Abbreviation of *operational.*

opacimeter **1.** An instrument for measuring the extent to which a material blocks light. **2.** An instrument, such as a field-strength meter, used to measure the effectiveness of an electrical shielding material in blocking radio waves, X rays, or other radiation.

opacity The condition of being opaque. This applies to all forms of radiation. For example, a material can be opaque to light rays, but be transparent to radio waves, or it can be transparent to gamma rays while being opaque to alpha particles.

op amp See OPERATIONAL AMPLIFIER.

op code Abbreviation of OPERATION CODE.

open-air line See OPEN-WIRE LINE.

open-back cabinet A loudspeaker enclosure in which the space behind the speaker is open to the room.

open capacitance The value of a variable capacitor whose rotor plates have been rotated completely out of mesh with the stator plates. Compare CLOSED CAPACITANCE.

open-center display A radar display in which a ring around the center indicates zero range.

open-chassis construction A method of assembling electronic equipment by mounting components and wiring them on an unenclosed chassis, often without a front panel. Similar to breadboard construction.

open circuit **1.** A discontinuous circuit (i.e., one that is broken at one or more points and, consequently, cannot conduct current nor present a voltage at its extremities). Compare CLOSED CIRCUIT. **2.** Pertaining to no-load conditions, for example, the open-circuit voltage of a battery. **3.** For a bipolar transistor, the operating characteristics under independent input and output conditions.

open circuit breaker A circuit breaker whose contacts are open.

open-circuit current Current flowing in the primary winding of an unloaded transformer.

open-circuited line See OPEN-CIRCUITED TRANSMISSION LINE.

open-circuited transmission line Abbreviation, OCTL. An unterminated transmission line in which the conductors at the far end are not connected together.

open-circuit impedance For a transmission line or a four-terminal network, the input or driving-point impedance when the output end of the line or network is open-circuited.

open-circuit jack A telephone jack that introduces a break in a circuit until a plug connected to a closed external circuit is inserted.

open-circuit plug See OPEN PLUG.

open-circuit resistance For a four-terminal network, the input or driving-point resistance when the output end of the network is unterminated.

open-circuit signaling A system of signaling in which no current flows until the signal circuit is in active operation. In a simple telegraph circuit, for example, current flows only when the key is pressed (to form a dot or dash).

open-circuit alarm system A security system in which all the actuating sensors are normally open and connected in parallel. When one of the sensors is actuated, it closes, causing a short circuit that triggers the alarm.

open-circuit voltage See NO-LOAD VOLTAGE.

open-collector configuration In an integrated circuit, an output scheme utilizing no internal pull-up resistor. Wired-OR outputs can thus have opposite states without risk of damage to the device.

open component An open-circuit component (e.g., an open diode, coil, resistor, etc.).

open core A magnetic core having a cylindrical shape. A disadvantage of this core configuration, in some applications, is that much of the magnetic flux extends outside the core. Compare CLOSED CORE.

open-core choke A choke coil wound on an open core. Also called OPEN-CORE INDUCTOR. Compare CLOSED-CORE CHOKE.

open-core transformer A transformer wound on an open core. Compare CLOSED-CORE TRANSFORMER.

open-delta connection See VEE-CONNECTION OF TRANSFORMERS.

open-ended Pertaining to a circuit or device that can be built upon without modifying its original configuration.

open-end stub A stub that is neither short-circuited nor terminated at its far end.

open-end stub tuning Adjustment of an OPEN-END STUB, by pruning its length, for optimum operation at a given frequency.

open-entry contact In a connector, an unprotected, opening contact of the female type.

open feeder See OPEN-WIRE LINE.

open-frame machine See OPEN GENERATOR and OPEN MOTOR.

open-fuse cutout A type of enclosed-fuse cutout having an exposed fuse holder and support. Compare OIL FUSE CUTOUT.

open generator An unsealed generator (i.e., one that has openings in its housing for air circulation).

open line **1.** See OPEN-WIRE LINE. **2.** See OPEN-CIRCUITED TRANSMISSION LINE.

open loop **1.** In a control system, a feedthrough path having no feedback and that is not self-regulating. **2.** In an operational amplifier, the configuration in which there is no feedback, and in which the device operates at maximum gain. **3.** A loop within a program from which the computer automatically exits after a specific number of iterations.

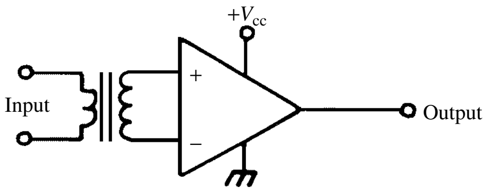

open loop, 2.

open-loop bandwidth The bandwidth of an open-loop device, such as an amplifier, without feedback. Also see OPEN LOOP.

open-loop control system A control system having no feedback and that is not self-regulating. An example is a fluid-level gauge that indicates the height of fluid in a tank, but that cannot correct the level automatically. Compare CLOSED-LOOP CONTROL SYSTEM.

open-loop differential voltage gain For a differential amplifier, the overall voltage gain (when either input is used) when the amplifier has no feedback.

open-loop gain The overall gain (ratio of output to input) of an open-loop device, such as an amplifier without feedback. Also see OPEN LOOP.

open-loop input impedance The input impedance of an open-loop device, such as an amplifier without feedback. Also see OPEN LOOP.

open-loop output impedance The output impedance of an open-loop device, such as an amplifier without feedback. Also see OPEN LOOP.

open-loop output resistance The output resistance of an open-loop device, such as an amplifier without feedback. Also see OPEN LOOP and OPEN-LOOP IMPEDANCE.

open-loop system **1.** A circuit in which the input and output currents are independent. **2.** A robot that does not use a servo system. It depends entirely on alignment and precision for positioning accuracy. **3.** An electromechanical device that does not use corrective feedback.

open-loop voltage gain The overall voltage gain of an open-loop amplifier (i.e., one having no feedback). Also see OPEN LOOP.

open magnetic circuit A magnetic circuit in which a complete path is not provided for magnetic flux. See, for example, OPEN CORE. Compare CLOSED MAGNETIC CIRCUIT.

open motor An unsealed motor (i.e., one that has openings in its housing for air circulation).

open-phase protection Use of an automatic device, such as an open-phase relay, to interrupt the power to a polyphase system when one or more phases are open-circulated.

open-phase relay In a polyphase system, a protective relay that opens when one or more phases are open-circuited. Also see OPEN-PHASE PROTECTION.

open plug A phone plug to which no external connections are made; it is used to hold the blades of a jack as if they were plugged in.

open-reel A tape-recording system in which the tape, during record or playback condition, is wound into a take-up reel that is physically separate from the tape reel. Also called *reel-to-reel arrangement.*

open relay **1.** A relay in its open-contact state. **2.** A relay having an open-circuited coil. **3.** An unenclosed relay.

open routine In computer operations, a routine that can be inserted directly into a larger routine and requires no link to the main program.

open subroutine In computer operations, a subroutine that can be inserted into a larger instructional sequence and must be recopied whenever it is required. Also called *direct-insert subroutine.*

Open Systems Interconnection Reference Model Abbreviation, OSI-RM. A standard set of protocols for computer network communication. It consists of seven levels, also called layers: *physical layer*, *link layer*, *network layer*, *transport layer*, *session layer*, *presentation layer*, and *application layer.*

open temperature pickup A temperature transducer that must be placed directly in contact with the monitored medium.

open volume Pertaining to the maximum-gain operation of a sound-reproducing system (i.e., operation at full volume).

open wire **1.** An unterminated wire. **2.** A wire supported above the surface of the earth and often ungrounded. **3.** See OPEN-WIRE LINE.

open-wire feeder See OPEN-WIRE LINE.

open-wire line A transmission line or feeder usually consisting of two straight, parallel wires held apart by bars of low-loss insulating material at regular intervals along the line.

open-wire loop A branch line connected to a main open-wire line.

open-wire transmission line See OPEN-WIRE LINE.

open-wire wavemeter See LECHER WIRES.

operand In computer operations, a quantity that enters into or results from an operation.

operate **1.** To manipulate according to an established procedure (e.g., to operate an instrument). **2.** To perform according to specifications, in the sense that an electronic circuit functions.

operate current A signal current or trigger current required to actuate a device. Compare OPERATE VOLTAGE.

operate delay See OPERATE TIME, **1**.

operate interval See OPERATE TIME, **2**.

operate time **1.** The interval starting after the application of an operate current or voltage to a device, and ending when the device operates. **2.** The period during which an electronic equipment is in operation. Also see OPERATING TIME, **1**.

operate voltage The signal voltage or trigger voltage required to actuate a device. Compare OPERATE CURRENT.

operating ambient temperature Abbreviation, OAT. The maximum or recommended temperature in the space immediately surrounding an equipment in operation.

operating angle In an amplifier circuit, the excitation-signal cycle, in degrees, during which drain, collector, or plate current flows. Class-A amplifiers operate for 360 degrees of the input signal cycle; class-AB amplifiers operate for more than 180 degrees, but less than 360 degrees of the input signal cycle; class-B amplifiers operate for 180 degrees of the input signal cycle; class-C amplifiers operate for less than 180 degrees of the input signal cycle.

operating bias In a circuit containing transistors, diodes, or vacuum tubes, the value(s) of direct-current bias required for normal operation.

operating code The code used by the operator in a computer or data-processing system.

operating conditions The environment in which a circuit or system functions in normal use.

operating current The current required by a device during its operation. Compare IDLING CURRENT.

operating cycle The sequence of events in the operation of a device. For example, the repetitive operation of a neon-bulb relaxation oscillator is a sequence of three events: (1) slow charge of capacitor, (2) firing of bulb, and (3) abrupt discharge of capacitor.

operating frequency **1.** The fundamental frequency at which a circuit or device is operated. **2.** The frequency of the current, voltage, or power delivered by a generator.

operating frequency range **1.** The range of operating frequencies, expressed as a minimum and maximum, for a communications receiver, transmitter, or transceiver. **2.** For an integrated-circuit oscillator, the minimum and maximum frequency, and all frequencies in between, at which the device is guaranteed to operate.

operating life The maximum period (from seconds to years) over which a device will operate before failure (from which it usually cannot recover). Compare SHELF LIFE.

operating line A line drawn across a family of curves depicting the performance of a device. It intersects each curve at a single point and graphically displays the performance of the device for a given condition. Thus, an operating line on a family of output curves for a transistor might depict operation with a given load resistor.

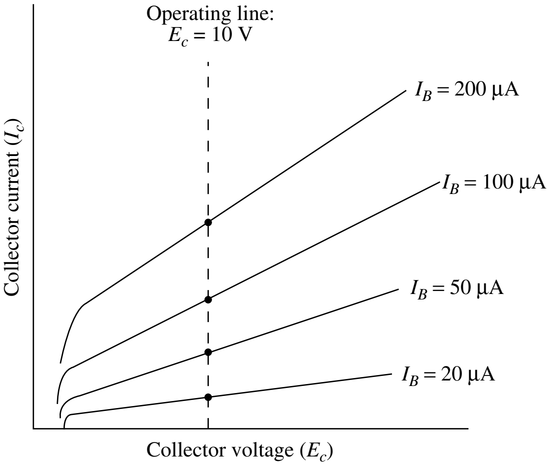

operating line

operating overload The extent and/or duration of overload to which an equipment can be exposed during customary operation, and still continue to properly function.

operating point On the response curve for a device, the point indicating the quiescent level of operation (such as determined by a fixed bias voltage or current). An alternating-current signal applied at this point oscillates above and below the point as a mean.

operating-point shift A movement of an operating point due to faulty operation of a circuit or device, or to a value change in some critical component.

operating position **1.** The control point in a system, [i.e., the place where an operator (see OPERATOR, **1**) normally functions]. **2.** The actual or recommended physical orientation of a device during its operation (e.g., a vertical operating position for a power vacuum tube).

operating power **1.** The power actually used by a device during its operation. **2.** The antenna power of a radio station.

operating ratio For a given period, the ratio t_o/t, where t_o is the time during which an equipment is operating correctly, and t is the duration of the period.

operating station In a computer installation, one or more consoles for the control of a data-processing system by an operator.

operating system Abbreviation, OS. In a computer, a set of programs that oversee the functioning of the hardware and application software. It works with the basic input/output system (BIOS).

operating temperature The actual or recommended temperature of a device during its operation.

operating-temperature range For a given device, the spread between maximum and minimum values of operating temperature.

operating time **1.** The interval during which an equipment is in operation. **2.** The period corresponding to OPERATING ANGLE.

operating-time characteristic For a coil-type relay, the relationship between operating time and operating power.

operating voltage The voltage required by a device, or measured at the device, during its operation. Compare IDLING VOLTAGE.

operation **1.** The working of a circuit or device (i.e., its performance). **2.** A process usually involving a sequence of steps (e.g., a mathematical operation).

operational amplifier Also called an *op amp.* A specialized linear integrated circuit (IC) that consists of several transistors, resistors, diodes, and capacitors, interconnected to produce gain over a wide range of frequencies. A single amplifier can comprise an entire IC, or an IC can contain several amplifiers. The *dual op amp* and the *quad op amp* are common variants. Some ICs contain one or more amplifiers in addition to other circuits. The devices can be used with resistance-capacitance (RC) network combinations to build active filters for use at audio frequencies.

operational differential amplifier An OPERATIONAL AMPLIFIER preceded by a DIFFERENTIAL AMPLIFIER.

operational readiness In statistical analysis, the probability that a system will, at a certain time, be correctly operating or ready to operate.

operational reliability Reliability determined empirically from a study of the actual operation of a device or system under controlled conditions. Also called ACHIEVED RELIABILITY.

operational transconductance amplifier Abbreviation, OTA. An integrated-circuit amplifier that differs from the conventional operational amplifier in that its output current is proportional to its input-signal voltage.

operation code In computer operations, the part of an instruction that specifies an operation.

operation decoder In a digital computer, the circuit that reads an OPERATION CODE and directs other circuitry in the execution of the code.

operation envelope See WORK ENVELOPE.

operation number In computer programming, a number that indicates the position in the program of a particular operation or subroutine.

operation part In a computer program, the part of an instruction containing the OPERATION CODE.

operation register In a digital computer, the register that stores the operation code of an instruction.

operations research A branch of computer engineering, devoted to the solution and/or optimization of functions of many variables.

operation time The interval between the instant of application of all voltages to a circuit, and the instant when the current reaches a specified percentage of its final steady value.

operator **1.** A person who performs an operation (see OPERATE, **1**). **2.** In mathematics, a symbol indicating an operation (e.g., *j*, +, −, ×, etc.).

operator j See J OPERATOR.

opposition **1.** The state of two quantities that are 180 degrees out of phase with each other. **2.** A state in which two effects or quantities operate against each other in some manner (e.g., physically, mathematically, electrically, etc.).

opt Abbreviation of OPTICAL.

Optacon An electronic aid for the blind. It has a camera that scans printed matter and a device that forms corresponding raised letters that can be read, as would Braille, with the fingertips. The name is a contraction of *optical to tactile converter.*

optical **1.** Generating or sensing visible light. **2.** Visible electromagnetic radiation in the range of approximately 390 to 750 nanometers wavelength.

optical ammeter A type of OPTICAL PYROMETER that measures the current flowing through the filament of an incandescent lamp.

optical antenna A photosensitive device intended for the reception of modulated-light or modulated-infrared signals. Usually consists of a convex lens or a concave reflector and a semiconductor photodetector.

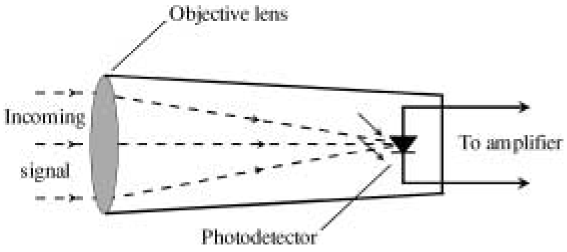

optical antenna

optical axis The z-axis of a quartz crystal (see Z-AXIS, **4**).

optical character reader A device that uses OPTICAL CHARACTER RECOGNITION to discern printed characters.

optical character recognition Abbreviation, OCR. In computer and data-processing operations, the reading of alphabetical, numerical, and other characters from hard copy (usually printed matter) by photoelectric methods. It converts the characters into digital data that can be stored in computer memory, on disks and tapes, and transmitted via digital communications networks. It can also allow a computer or robot to read signs, maps, etc.

optical communications One-way or two-way communications via modulated visible light. It can be conducted through optical fibers, the atmosphere, water, or any other transparent medium.

optical coupler A coupling device consisting essentially of a light source (actuated by an input signal) mounted in an opaque housing with a light-sensitive device (that delivers the output signal). In its simplest form, the arrangement consists of a light-emitting diode (LED) and a photodiode.

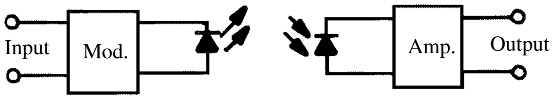

optical coupler

optical detector An integrated-circuit (IC) that provides light-to-voltage conversion. Its direct-current output voltage is proportional to the intensity of light impinging on its sensor.

optical encoder An electronic device that measures the extent to which a mechanical shaft has rotated. It can also measure the rate of rotation (angular speed). It consists of a light-emitting diode (LED), a photodetector, and a disk with alternate light and dark bands (chopping wheel). It is commonly used in digital controls and jointed robot arms.

optical fiber A glass or plastic medium through which light is propagated for optical-communications purposes. The refractive characteristics of the fiber keep the visible light inside.

optical lever A device for amplifying the effect of a small rotation. The rotating member carries a small mirror that reflects a light beam over a curved scale, the distance through which the light spot travels on the scale being proportional to the distance between the scale and the rotating mirror. In this way, the deflection on the scale is several times the length of the arc described by the mirror, the rotation being thus amplified.

optical link See OPTICAL COUPLER.

optical mark recognition A method of data transfer that involves the use of optical techniques.

optical maser See LASER.

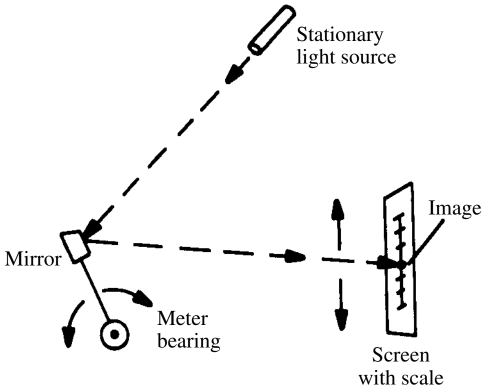

optical lever

optical mode The vibration mode that produces an oscillating dipole in a crystal lattice.

optical pattern A Christmas-tree-like pattern produced by parallel rays of light striking a phono disk. Also see DIFFRACTION.

optical playback See OPTICAL SOUND REPRODUCER.

optical pyrometer A pyrometer for measuring the temperature of a hot body in terms of the intensity and color of light it emits.

optical scanner An electronic device that converts hard-copy text and graphics into digital form suitable for processing and storage in a computer. There are several types; the most common resembles a small photocopy machine and is known as a *flatbed scanner*. See OPTICAL SCANNING.

optical scanning An electronic process that converts hard-copy text and graphics into digital form suitable for processing and storage in a computer. Text scanners work in conjunction with specialized software. The optical scanner and software together form the equivalent of audio speech recognition and sound interpretation. Usually, scanners are employed to convert printed matter to digital data. However, scanning technology can also be used at a distance, for example, by a robot for the purpose of reading road signs.

optical sound recorder A photoelectric machine for recording sound on photographic film. Also see OPTICAL SOUND RECORDING.

optical sound recording A system for recording sound on photographic film, as in sound motion pictures. The sound is picked up by a microphone and amplified to vary the intensity of a light source. The film passing this modulated beam becomes exposed to a variable-width or variable-density track corresponding to the modulation (see VARIABLE-DENSITY SOUND RECORD and VARIABLE-WIDTH SOUND RECORD). When developed film is played back, its sound track modulates a light beam in the reproducer by actuating a photocell or phototube to produce the audio signal, which is then amplified.

optical sound reproducer A photoelectric machine for reproducing sound on film. A light beam in the device is modulated by the passing sound track, and it, in turn, modulates photocell or phototube current, which is amplified to drive a loudspeaker. Also see OPTICAL SOUND RECORDING.

optical system Collectively, the functional arrangement of lenses, mirrors, prisms, and related devices in optoelectronic apparatus.

optical tachometer An optoelectronic instrument for measuring (by means of reflected-light variations) the speed of a body, such as a rotating shaft, without electrical or mechanical attachments to the latter.

optical thermometer See OPTICAL PYROMETER.

optical twinning A kind of defect in which two types of quartz occur in the same crystal. Compare ELECTRICAL TWINNING.

optical type font A special type (printing) style designed for use with OPTICAL CHARACTER RECOGNITION equipment.

optical wand A pencil-like optical probe, used to read bar codes from a printed page and translate the codes into information that is then loaded into a computer or calculator.

optic axis See OPTICAL AXIS.

optic flow See EPIPOLAR NAVIGATION.

optics **1.** The science of light, its measurement, application, and control. **2.** A system of lenses, prisms, filters, or mirrors used in electronics to direct, control, or otherwise modify light rays.

optimization The adjustment or manipulation of the elements of a process or system for the best operation or end result.

optimize **1.** To manipulate a set of variables or parameters for the best possible performance of a circuit or system. **2.** To maximize the value of a multivariable function.

optimum angle of radiation For a given altitude of the ionosphere, the angle, with respect to the horizon, at which a radio signal should be transmitted from a specific geographic location to provide optimum reception at some other specific location. This angle varies, depending on the altitude of the ionospheric layer from which signals are returned, the distance between the transmitting and receiving stations, and the number of hops taken by the signal between the earth and the ionosphere.

optimum bunching In a velocity-modulated tube, such as a klystron, the bunching associated with maximum output.

optimum collector load The ideal load impedance (see OPTIMUM LOAD) for a particular transistor operated in a specified manner.

optimum coupling The degree of coupling between two circuits tuned to the same frequency that results in maximum energy transfer. Also called CRITICAL COUPLING.

optimum current The value of current that produces the most-effective performance of a circuit or device.

optimum frequency See OPTIMUM WORKING FREQUENCY.

optimum load The value of load impedance that produces maximum transfer of power from a generator or amplifier.

optimum plate load The ideal load impedance (see OPTIMUM LOAD) for a particular electron tube operated in a specified manner.

optimum *Q* The most effective figure of merit for a capacitor, inductor, or tuned circuit at a specified frequency.

optimum reliability The value of reliability that ensures minimum project cost.

optimum voltage The value of voltage that results in the most effective performance of a circuit or device.

optimum working frequency In radio transmission involving reflection from the ionosphere, the frequency of use that results in the most reliable communication between two points.

optoelectronic coupler An assembly consisting of an LED and a phototransistor. An input signal causes the diode to glow, and the light activates the transistor, which in turn delivers an output signal of higher amplitude than that of the input signal.

optoisolator A light-emitting diode (LED) or infrared-emitting diode (IRED) and a photodiode combined in a single package with a transparent gap between them. The LED or IRED converts an electrical signal to visible light or IR; the photodiode changes the visible light or IR back into an electrical signal. This device eliminates interstage impedance reflections that can plague systems that employ electronic coupling. Even the most drastic changes in load impedance have no effect on the impedance "seen" by the signal source. The devices can also protect expensive equipment against electrical transients; for example, in telephone-line modems used with personal computers.

optoelectronics A branch of electronics that involves the use of visible light for communications or data-transfer purposes.

optoelectronic transistor A transistor having an electroluminescent emitter, transparent base, and photoelectric collector.

optophone A photoelectric device for converting light into sounds of proportionate pitch to enable blind persons to "see" by ear.

Orange Book Trade name (Sony and Philips) for a specialized format for *compact-disc recordable (CD-R)* computer data storage media. It allows recording of data onto, as well as retrieval of data from, an optical compact disc. See also GREEN BOOK, RED BOOK, and YELLOW BOOK.

orange peel On a phonograph disc, a mottled surface that produces high background noise; so called from its resemblance to an orange peel.

orbit **1.** Also called *shell.* The median path of an electron around the nucleus of an atom. **2.** The circular or elliptical path of an artificial satellite (e.g., a communications satellite) around the earth. **3.** The elliptical path of the moon around the earth. **4.** The elliptical path of the earth or another planet around the sun. **5.** Any closed, usually circular or elliptical path, that an object follows around another object.

orbital-beam multiplier tube An electron-multiplying UHF oscillator or amplifier tube in which a positively charged electrode focuses electrons in a circular path.

orbital electron An electron in orbit around the nucleus of an atom.

orbital period The length of time required by a satellite to go around a celestial body, usually the earth. It can vary from about 90 minutes for very low orbits to months or even years for orbits beyond that of the moon. The moon's period is about a month.

OR circuit Also called *inclusive-OR circuit.* In digital systems and other switching circuits, a logic gate whose output is high (logic 1) when any of the input signals is high. The output is low (logic 0) only when all of the inputs are low. Compare AND CIRCUIT.

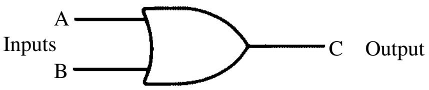

A	B	C
0	0	0
0	1	1
1	0	1
1	1	1

2-input OR circuit

order **1.** An instruction to a digital computer. **2.** Sequence.

OR element See OR CIRCUIT.

ordered pair A set of two variables or numbers that can be mapped one-to-one onto a set of points in a plane. The most familiar example is (x,y), used to depict the Cartesian (rectangular) coordinate system on which characteristic curves and other functions are plotted.

order of interference The position of an interference fringe that depends on whether the interference arises from one, two, three, or more wavelength differences. Generally, the first fringe results from a difference of one wavelength, the second fringe from a difference of two wavelengths, etc. The center fringe indicates that the two signals are exactly in phase; that is, the difference is zero wavelengths.

order of logic **1.** A mathematical expression for the complexity of a system of logic. **2.** The relative speed with which logical information is processed in a system.

order tone A warning signal to receiving operators in the form of a tone transmitted over a trunk preceding the transmission of an order.

ordinary ray As a result of the double refraction of electromagnetic waves, the member of a pair of rays that follows the usual laws of refraction. Also see O WAVE. Compare EXTRAORDINARY RAY.

ordinary wave See O WAVE.

ordinate In the rectangular coordinate system, a point located on the vertical axis.

organ **1.** A computer subsystem. **2.** An electronic device used for the purpose of generating music.

organic decay See EXPONENTIAL DECREASE.

organic electricity Electricity in the living tissues of animals and plants.

organic growth See EXPONENTIAL INCREASE.

organic semiconductor A semiconductor material consisting of, or combined with, some compound of carbon.

OR gate See OR CIRCUIT.

orientation **1.** The direction or position of an object in space, expressed in terms of coordinate values. **2.** In a teleprinter, the calibration or alignment that determines the speed of response to a received character.

orientation of quartz plates See CRYSTAL AXES and CRYSTAL CUTS.

orifice An opening or window, such as that in a loudspeaker enclosure or waveguide, through which energy is transmitted.

origin **1.** The starting point in a coordinate system. **2.** Relative to address modification in computer operations, an address to which a modifier is added to derive a variable operand address.

original lacquer In disc-recording operations, an original recording made on a lacquer-surfaced disc that is subsequently used to make a master.

original master In disk recording, the master disc produced from a wax or lacquer disc (see ORIGINAL LACQUER) by means of electroforming.

origin distortion The change in the shape of a wave as it swings through zero (polarity).

OR operation See OR CIRCUIT.

ORS Abbreviation of OMNIDIRECTIONAL RANGE STATION.

orthicon A television camera tube somewhat similar to the iconoscope, but which provides internal amplification of light and, accordingly, can be used in dimmer places than the iconoscope. Light amplification is provided by an arrangement similar to that of a photomultiplier tube.

orthiconoscope See ORTHICON.

orthoacoustic recording 1. A system of disc recording in which the inherent differences between high-frequency recording and low-frequency recording are compensated to provide reproduction that more closely resembles the actual sound. **2.** A disc made by the method defined in **1**.

orthogonal axes Perpendicular axes [e.g., those in a Cartesian (rectangular) coordinate system].

Os Symbol for OSMIUM.

OS Abbreviation of OPERATING SYSTEM.

osc Abbreviation of OSCILLATOR.

OSCAR Abbreviation for *Orbiting Satellite Carrying Amateur Radio.* A satellite with a transponder that has an uplink in one amateur band and a downlink in another amateur band.

osciducer See OSCILLATING TRANSDUCER.

oscillate 1. To fluctuate in amplitude in a uniform manner. **2.** To vary above and below a specified value at a constant rate.

oscillating arc A small arc, especially one produced by slow-opening relay contacts, that generates high-frequency oscillations.

oscillating circuit A closed circuit containing inductance, capacitance, and inherent resistance, in which energy passes back and forth between inductor and capacitor at a frequency determined by the inductance (L) and capacitance (C) values.

oscillating crystal 1. A piezoelectric plate maintained in a state of oscillation in a circuit. See, for example, CRYSTAL OSCILLATOR and QUARTZ CRYSTAL. **2.** An oscillating semiconductor diode (see NEGATIVE-RESISTANCE DIODE, **1**, **2**).

oscillating current See OSCILLATORY CURRENT.

oscillating detector A detector provided with positive feedback; therefore, it is capable of generating a signal of its own. Compare NON-OSCILLATING DETECTOR.

oscillating diode 1. A semiconductor diode biased into its negative-resistance region so that it oscillates in a suitable circuit. **2.** An oscillating tunnel diode. **3.** Any of several microwave diodes, such as the IMPATT diode, which will oscillate in a suitable system. **4.** See MAGNETRON. Also see DIODE OSCILLATOR.

oscillating field An alternating electric or magnetic field.

oscillating rod A rod of magnetostrictive metal maintained in a state of oscillation in a circuit. See, for example, MAGNETOSTRICTION and MAGNETOSTRICTION OSCILLATOR.

oscillating transducer A transducer in which an input quantity varies a frequency proportionately from its center value.

oscillating wire A wire of magnetostrictive metal maintained in a state of oscillation in a circuit. See, for example, MAGNETOSTRICTION and MAGNETOSTRICTION OSCILLATOR.

oscillation The periodic change of a body or quantity in amplitude or position (e.g., oscillation of a pendulum, voltage, or crystal plate).

oscillation constant For an oscillating inductance-capacitance (LC) circuit, the expression $(LC)^{1/2}$, where L is the inductance in henrys and C is the capacitance in farads. The reciprocal of the OSCILLATION NUMBER.

oscillation control A manual or automatic device for adjusting the frequency or amplitude of the signal generated by an oscillator.

oscillation efficiency The ratio, as a percentage, of the alternating-current (signal) power output of an oscillator (P_{out}) to the corresponding direct-current power input (P_{in}). $Efficiency = 100P_{out}/P_{in}$.

oscillation number For an oscillating circuit, the number of complete oscillation cycles that occur in 6.28 (2 pi) seconds.

oscillation test 1. A test of an oscillator to determine if a signal is being generated. **2.** A test for transistors wherein the transistor is used as an oscillator to give a rough indication of its condition in terms of oscillation amplitude.

oscillation transformer A tank coil of a radio transformer—especially one that includes an output coupling coil.

oscillator A device that produces an alternating or pulsating current or voltage electronically. The term is sometimes used to describe any alternating-current-producing device other than an electromechanical generator.

oscillator circuit The specific manner in which the components of an OSCILLATOR are interconnected. The three general types are: positive-feedback, negative-resistance, and relaxation.

oscillator coil A tapped coil that provides the input and output windings required for an oscillator circuit. Such coils are used in signal generators, oscillators, and superheterodyne receivers.

oscillator-doubler A circuit consisting of an oscillator and a frequency doubler (e.g., a crystal oscillator whose output frequency is twice the crystal frequency).

oscillator drift A usually gradual change in frequency of an oscillator caused by such factors as warmup time, voltage variations, capacitance change, inductance change, or change in transistor characteristics.

oscillator frequency The fundamental frequency at which an oscillator operates. It can be determined by a tuned circuit, crystal, cavity, section of waveguide or transmission line, or by a resistance-capacitance circuit.

oscillator harmonic interference In a superheterodyne receiver, interference that is the beat product of local oscillator harmonics and received signals.

oscillator interference Radio-frequency interference caused by signals from the high-frequency oscillator of a receiver.

oscillator keying Keying by making and breaking the signal output, direct-current (dc) power, or dc bias of the oscillator stage of a radiotelegraph transmitter.

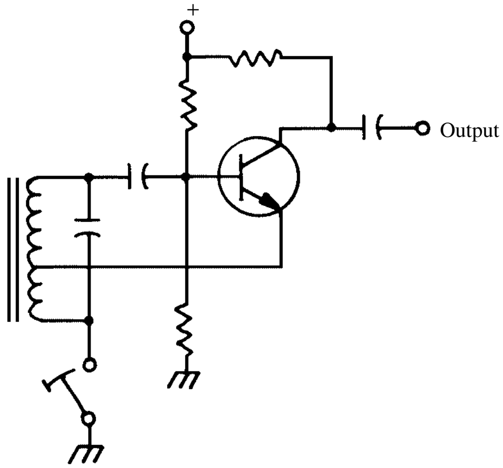

oscillator keying

oscillator-mixer **1.** A combination stage in which a transistor functions as a local oscillator and mixer in a receiver or test instrument. **2.** A device designed specifically to function as a local oscillator and mixer.

oscillator-mixer-detector **1.** In a superheterodyne receiver, a stage in which the functions of high-frequency oscillator, mixer, and first detector are performed by a single transistor. **2.** A device designed specifically to function as a local oscillator, mixer, and detector.

oscillator-multiplier A single circuit that serves simultaneously as an oscillator and frequency multiplier. See, for example, OSCILLATOR-DOUBLER.

oscillator padder In a superheterodyne receiver, a small, limited-range variable capacitor connected in series with the oscillator coil for tracking oscillator tuning at the low end of a band. Compare OSCILLATOR TRIMMER.

oscillator power supply **1.** The direct-current or alternating-current power supply for an oscillator. **2.** See OSCILLATOR-TYPE POWER SUPPLY.

oscillator radiation The emission of radio-frequency energy by the oscillator stage of a superheterodyne receiver. Also see OSCILLATOR INTERFERENCE.

oscillator-radiation voltage The radio-frequency voltage at the antenna terminals of a superheterodyne receiver that results from signal emission by the oscillator stage.

oscillator stabilization **1.** The automatic compensation of an oscillator circuit for the frequency drift resulting from changes in temperature, current, voltage, or component parameters. **2.** The automatic stabilization of the operating point of an oscillator circuit against variations resulting from changes in temperature, supply current or voltage, or component parameters.

oscillator synchronization The locking of an oscillator in step with another signal source, such as a frequency-standard generator.

oscillator tracking In a superheterodyne receiver, the constant separation of the oscillator frequency from the signal frequency by an amount equal to the intermediate frequency at all settings of the tuning control.

oscillator transmitter A radio transmitter consisting only of a radio-frequency oscillator and its power supply. The oscillator can be modulated in various ways [e.g., on-off keying, frequency-shift keying, voice amplitude modulation (AM), voice frequency modulation (FM)].

oscillator trimmer In a superheterodyne receiver, a small, limited-range capacitor connected in parallel with the oscillator coil for tracking oscillator tuning at the high end of a band. Compare OSCILLATOR PADDER.

oscillator tuning The separate, often ganged, tuning of the oscillator stage in a circuit.

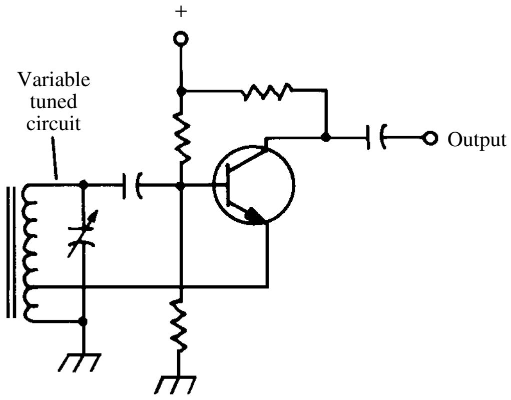

oscillator tuning

oscillator-type power supply A high-voltage, direct-current (dc) power supply in which a radio-frequency (RF) oscillator generates a low-voltage alternating current (ac). This ac voltage is stepped up by an RF transformer, and is finally rectified to obtain high-voltage dc.

oscillator-type transmitter See OSCILLATOR TRANSMITTER.

oscillator wavelength The fundamental wavelength at which an oscillator operates. It is usually expressed in meters, centimeters, or millimeters. It can be determined by a tuned circuit, crystal, cavity, section of waveguide or transmission line, or by a resistance-capacitance circuit.

oscillatory current A current that alternates periodically, particularly the current in an inductance-capacitance (LC) tank circuit that results from the oscillation of energy back and forth between the inductor and capacitor.

oscillatory discharge An electrical discharge, such as that of a capacitor, that sets up an OSCILLATORY CURRENT.

oscillatory surge A current or voltage surge that includes both positive and negative excursions.

oscillatory transient See OSCILLATORY SURGE.

oscillistor A device consisting essentially of a bar of semiconductor material positioned in a magnetic field; it will produce oscillations under certain conditions.

oscillogram **1.** The image produced on the screen of an oscilloscope. **2.** A permanent, usually photographic, record made from the screen of an oscilloscope.

oscillograph **1.** An instrument that makes a permanent record (photograph or pen recording) of a rapidly varying electrical quantity. Also called *recorder* (see RECORDER, **2**). Compare OSCILLOSCOPE. **2.** An obsolete term for OSCILLOSCOPE.

oscillograph recorder A direct-writing recorder (see RECORDER, **2**).

oscillography The use of a graphic oscillation recorder (OSCILLOGRAPH).

oscillometer A device used for determining the peak amplitude of an oscillation.

oscilloscope An instrument that presents for visual inspection the pattern representing variations in an electrical quantity. Also see CATHODE-RAY OSCILLOSCOPE. Compare OSCILLOGRAPH.

oscilloscope camera A special high-speed, short-focus camera with fixtures for attachment to an oscilloscope to record images from the screen. Standard and instant-film types are available.

oscilloscope differential amplifier An amplifier that processes the difference between two signals, for the purpose of displaying on an oscilloscope or oscillograph.

oscilloscope tube A cathode-ray tube for use in an oscilloscope. It contains an electron gun, accelerating electrode, horizontal and vertical deflecting plates, and a fluorescent screen.

Os-Ir Symbol for OSMIRIDIUM.

OSI-RM Abbreviation of OPEN SYSTEMS INTERCONNECTION REFERENCE MODEL.

OSL Abbreviation of *orbiting space laboratory.*

osmiridium Symbol, Os-Ir. A natural alloy of osmium and iridium.

osmium Symbol, Os. A metallic element of the platinum group. Atomic number, 76. Atomic weight, 190.2.

osmotic pressure The force that causes the positive ions to pass out of a solution toward a metal body immersed in an electrolyte. Also see HELMHOLTZ DOUBLE LAYER.

OSO Abbreviation of *orbiting solar observatory.*

osteophone A bone-conduction hearing aid.

OTA Abbreviation of OPERATIONAL TRANSCONDUCTANCE AMPLIFIER.

OTL Abbreviation of OUTPUT-TRANSFORMERLESS.

O-type backward-wave oscillator Abbreviation, OBWO. A backward-wave oscillator using harmonics having opposing phases.

ounce Abbreviation, oz. A unit of weight equal to 1⁄16 pound or 28.35 grams.

ounce-inch Abbreviation, oz-in. A unit of torque equal to the product of a force of 1 ounce and a moment arm of 1 inch. Compare POUND-FOOT.

outage **1.** Loss of power to a system. **2.** Loss of a received signal.

outboard components **1.** Discrete components (capacitors, coils, resistors, or transformers) connected externally to an integrated circuit. **2.** Discrete components connected externally to any existing electronic device.

outcome In statistical analysis, the result of an experiment or test. An outcome can be numerical or nonnumerical.

outdoor antenna An antenna erected outside, usually high above the surface of the earth clear of obstacles. It generally provides superior performance compared with an INDOOR ANTENNA. Also reduces the probability of radio-frequency interference (RFI) when used for transmitting.

outdoor booster A signal preamplifier mounted on an outdoor television receiving antenna for improved reception.

outdoor transformer A weatherproof distribution transformer installed outside the building it services.

outer conductor The outer metal cylinder or jacket of a coaxial cable or coaxial tank. Compare INNER CONDUCTOR.

outgassing **1.** In the evacuation of electronic devices, such as vacuum tubes, the removal of occluded gases from glass, ceramic, and metal by means of slow baking and by flashing an internal metal getter (such as one of magnesium). **2.** The production of gases in certain electrochemical cells and batteries during the final stage of charging.

outgoing line A power or signal line that leaves a device, facility, or stage. Compare INCOMING LINE.

outlet A female receptacle that delivers a signal or operating power to equipment plugged into it.

outline flowchart In computer operations, a preliminary flowchart showing how a program will be divided into routines and segments, input and output functions, program entry points, etc.

out-of-line coding Instructions for a computer program routine stored in an area of memory other than that in which the routine's program is stored.

out of phase Pertaining to the condition in which the alternations or pulsations of two or more separate waves or wave phenomena, having identical frequencies, are out of step with each other. Compare IN PHASE.

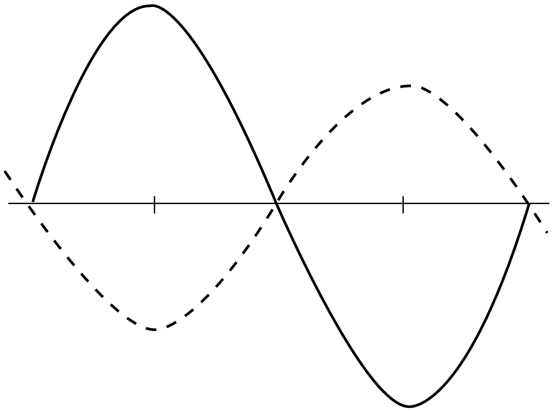

out of phase

out-of-phase current Reactive current in an alternating-current circuit (i.e., current that is out of phase with voltage. Also see QUADRATURE CURRENT).

out-of-phase voltage Voltage across a reactance; so called because it is out of phase with the current.

outphaser A device that converts a sawtooth wave to a square wave. It is used in electronic organs and synthesizers.

outphasing modulation A system of modulation in which the sideband frequencies are shifted 90 degrees from the phase position in an amplitude-modulated wave. The resulting constant-envelope wave is then amplified with high efficiency and low distortion by a class-C stage; then the signal is reconverted to an amplitude-modulated one by phase shifting the carrier, with respect to the sidebands.

out-plant system A data-processing system in which a central computer receives data from remote terminals.

output **1.** Energy or information delivered by a circuit, device, or system. Compare INPUT, **1**. **2.** The terminals at which energy or information is taken from a circuit, device, or system. Compare INPUT, **2**.

output admittance Symbol, Y_o. The internal admittance of a circuit or device, as “seen” at the output terminals; the reciprocal of OUTPUT IMPEDANCE. Compare INPUT ADMITTANCE.

output amplifier See FINAL AMPLIFIER.

output area In a computer system, the portion of storage holding information for delivery to an output device. Also called *output block.*

output axis For a gyroscope that has received an input signal, the axis around which the spinning wheel precesses.

output block See OUTPUT AREA.

output buffer **1.** A circuit that follows an oscillator and reduces the effects of variable load impedance on the oscillator frequency or signal amplitude. **2.** An amplifier, usually with a voltage gain of 6 dB, that follows a video multiplexer. The amplifier drives a coaxial transmission line.

output bus driver In a computer, a device that amplifies output signals sufficiently to provide signals to other devices without undue loading of the supply line (bus).

output capability The maximum power or voltage output that a circuit can deliver without distortion or other improper operating conditions.

output capacitance Symbol, C_o. The internal capacitance of a circuit or device, as seen at the output terminals. Compare INPUT CAPACITANCE.

output capacitive loading For an operational amplifier at unity gain, the maximum capacitance that can be connected to the output of the amplifier before phase shift increases to the point of oscillation.

output capacitor **1.** In a capacitance-coupled circuit, the output coupling capacitor. Compare INPUT CAPACITOR. **2.** The last capacitor in a power-supply filter circuit.

output capacity The maximum output capability of a device or system expressed in appropriate units, such as current, voltage, power, torque, horsepower, etc.

output choke The last choke (inductor) in a power-supply filter circuit.

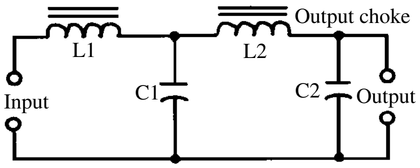

output choke

output circuit The circuit or subcircuit that constitutes the output portion of a network or device. Also see OUTPUT and OUTPUT TERMINALS. Compare INPUT CIRCUIT.

output-circuit distortion Distortion in the output portion of a circuit or device (such as a transistor or transformer), usually caused by an overload or nonlinear response.

output conductance Symbol, G_o. The internal conductance of a circuit or device, as “seen” at the output terminals. It is the reciprocal of OUTPUT RESISTANCE. Compare INPUT CONDUCTANCE.

output control **1.** The gain control of an amplifier. **2.** The level control of a variable power supply.

output coupling capacitor See OUTPUT CAPACITOR.

output coupling transformer See OUTPUT TRANSFORMER.

output current **1.** Symbol, I_o. The current delivered by a source, such as a battery, generator, or amplifier. Compare INPUT CURRENT, **1**. **2.** Symbol, I_o. Current flowing in the output leg or electrode of a circuit or device. Compare INPUT CURRENT, **2**.

output device **1.** A load device, such as a resistor, loudspeaker, lamp, relay, motor, etc., that utilizes the output energy delivered by a generator, amplifier, or network. **2.** A device, such as an output transformer, that serves to transfer energy or information from a circuit or device. Compare INPUT DEVICE. **3.** In computer operations, a device that presents the results of computer operation in a comprehensible form. Examples: printer, monitor, disk drive, tape drive, modem, etc.

output efficiency The efficiency of a device, such as a generator or amplifier, in delivering an output signal. For an amplifier, the efficiency ($Eff_{\%}$) is given as a percentage by the formula $Eff_{\%} = 100P_o/P_i$, where P_i is the direct-current power input, and P_o is the alternating-current (signal) power output.

output equipment See OUTPUT DEVICE, **3**.

output filter The direct-current filter of a power supply operating from alternating current. Also see CAPACITOR-INPUT FILTER and CHOKE-INPUT FILTER.

output gap A device via which current or power is intercepted from an electron beam in a beam-power tube.

output impedance Symbol, Z_o. The impedance "looking" into the output terminals of an amplifier, generator, or network. Compare INPUT IMPEDANCE.

output indicator A device, such as an analog meter, digital meter, or bar-graph meter, that provides a visual indication of the output-signal amplitude of an equipment.

output leakage current In an open-collector integrated circuit, the current from collector to emitter with the output in the "off" condition and a certain specified voltage applied to the device. It can be expressed in milliamperes or microamperes.

output limiting A process for automatically maintaining the amplitude of the signal delivered by a generator or amplifier. See, for example, AUTOMATIC GAIN CONTROL, AUTOMATIC MODULATION CONTROL, VOLUME COMPRESSION, and VOLUME LIMITER.

output load See OUTPUT DEVICE, **1**.

output load current **1.** The current through the output load of an amplifier. Generally, this current is expressed in root-mean-square (rms) form. **2.** The highest rms current that an amplifier can deliver to a load of a specified impedance.

output meter A meter that gives a quantitative or qualitative indication of the output of an amplifier or generator. See, for example, OUTPUT-POWER METER.

output offset In an integrated circuit, the voltage at the output when the inputs are grounded.

output port The output terminal of a logic device.

output power Symbol, P_o. The power deliverable by an amplifier, generator, or circuit. Also called *power output.* Compare INPUT POWER.

output-power meter A type of direct-reading wattmeter for measuring the power output of an amplifier or generator.

output regulator A circuit or device that automatically maintains the output of a power supply or signal source at a constant amplitude.

output resistance Symbol, R_o. The internal resistance of a circuit or device, as "seen" at the output terminals. Compare INPUT RESISTANCE.

output routine In computer operations, a routine (program segment) that performs the work involved in moving data to an output device, often including intermediate transferals and modifying the data as necessary.

output section See OUTPUT AREA.

output sink current In an integrated circuit, for a specified set of conditions at the input and output, current into the output as measured in milliamperes or microamperes.

output source current In an integrated circuit, for a specified set of conditions at the input and output, the current out of the output, as measured in milliamperes or microamperes.

output stage The last stage of an amplifier. Delivers the signal to the load.

output terminals Terminals (usually a pair) associated with the output of a circuit or device (see OUTPUT, **1**, **2**). Compare INPUT TERMINALS.

output tank In a transmitter or power generator, a parallel-tuned combination of inductance and capacitance in the collector, drain, or plate circuit, that is generally tuned to resonance at the operating frequency. IT optimizes efficiency and couples the signal to the load. Compare INPUT TANK.

output transformer The output-coupling transformer that delivers signal voltage or power from an amplifier, generator, or network to a load or to another circuit. Compare INPUT TRANSFORMER.

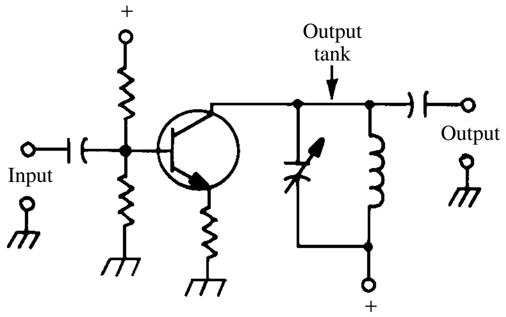

output tank

output-transformerless Abbreviation, OTL. Pertaining to an oscillator, amplifier, or generator that requires no output coupling transformer.

output transistor A transistor in the final stage of an amplifier or generator; usually, it is a power transistor.

output tube A vacuum tube in the final stage of an amplifier or generator; usually, it is a power tube.

output unit See OUTPUT DEVICE.

output voltage **1.** Symbol, E_o or V_o. The voltage delivered by a source, such as a battery, generator, or amplifier. Compare INPUT VOLTAGE, **2**. **2.** Symbol, E_o or V_o. The voltage across the output leg or electrode of a circuit or device. Compare INPUT VOLTAGE, **2**.

output voltage compliance In an integrated circuit, the voltage range over which the output can be made to swing, while keeping the operation of the circuit within a certain maximum allowable nonlinearity limit. It is measured in volts or millivolts.

output voltage noise In an integrated circuit, the output noise over a given range of frequencies, as measured in peak-to-peak millivolts or microvolts. It can also be measured as the root-mean-square (rms) value.

output voltage swing In an integrated circuit with a specified load, the output-voltage change measured as a difference between maximum and minimum in volts or millivolts.

output voltage tracking For an integrated-circuit dual regulator, the difference between the absolute values of the output voltages of a dual regulator. It can be expressed as a specific voltage or as a percentage of the specified output voltage of the device.

output winding The secondary coil of an output transformer.

outside antenna See OUTDOOR ANTENNA.

outside booster See OUTDOOR BOOSTER.

outside diameter Abbreviation, OD. The outermost diameter of a body or figure having two concentric diameters (e.g., tubing or conduit). Compare INSIDE DIAMETER.

outside lead See FINISH LEAD.

outside transformer **1.** See OUTDOOR TRANSFORMER. **2.** A transformer mounted outside of an equipment in whose circuit it is included. External mounting can eliminate hum in the equipment circuit, and can help to prevent overheating.

oven **1.** Also called *crystal oven.* A chamber providing a closely controlled operating temperature for an electronic component, such as a quartz crystal. **2.** An enclosure in which electronic equipment can be tested at selected, precise high temperatures. Compare COLD CHAMBER.

overall feedback Positive or negative feedback around an entire system (such as a public-address system), as opposed to feedback confined to one stage or a few stages within the system.

overall gain The total gain of an entire system (such as a multistage amplifier), as opposed to that of one or several stages.

overall loudness The apparent intensity of an acoustic disturbance, generally measured with respect to the threshold of hearing, and expressed in decibels, relative to the threshold level.

overbiased unit A component, such as a transistor or vacuum tube, whose bias current or voltage is higher than the correct value for a given mode of operation. Compare UNDERBIASED UNIT.

overbunching In a velocity-modulated tube, such as a Klystron, the condition in which the buncher voltage exceeds the value required for optimum bunching.

overcharging In a secondary cell or battery, the application of charging current longer than necessary to obtain full charge. This can sometimes cause problems, such as cell heating.

overcompounded generator A dynamo-type generator having a compound field winding in which the series-field winding increases the field intensity beyond the point needed to maintain the output voltage. Compare UNDERCOMPOUNDED GENERATOR.

overcompounding A characteristic of electromechanical motors, resulting in increased running speed with decreasing load resistance.

overcoupled transformer A transformer having greater than critical coupling between its primary and secondary windings. In tuned circuits, such as intermediate-frequency (IF) transformers, this produces a double-peak response.

overcoupling Extremely close coupling (see CLOSE COUPLING).

overcurrent A current greater than the specified, nominal, or desired level. Compare UNDERCURRENT.

overcurrent circuit breaker A circuit breaker that opens when current exceeds a predetermined value.

overcurrent protection The use of a circuit breaker, relay, or other device to protect a circuit or system from damage resulting from an excessive flow of current.

overcurrent relay A protective relay that opens a circuit when current exceeds a predetermined value. Compare UNDERCURRENT RELAY.

overcutting In disc recording, the condition in which an excessively high amplitude signal causes the stylus to cut through the wall between adjacent grooves. Compare UNDERCUTTING.

overdamping Damping greater than the critical value (see DAMPING ACTION, **2**). Compare UNDERDAMPING.

overdesign Also called *overengineering.* **1.** To use an unnecessarily high safety factor in the design of equipment. **2.** To design equipment for performance superior to that which is required in the intended application. **3.** A design that results from operations defined in **1** and **2**.

overdrive In an analog amplifier, an undesirable operating condition that occurs with the application of excessive input signal voltage or power. This results in increased distortion, nonlinearity, excessive harmonic generation, and excessive output signal bandwidth. In a high-fidelity audio system, such a condition grossly degrades the quality of the sound. The amplifying device (transistor or vacuum tube) is in or near saturation during part of the signal cycle. This reduces the efficiency of the circuit, can cause excessive collector, drain, or plate current, and can overheat the base–collector (B-C) junction of a bipolar transistor. In the worst case, it can destroy the component.

overdriven amplifier See OVERDRIVEN UNIT.

overdriven unit An amplifier, oscillator, or transducer whose driving signal (current, voltage, power, or other quantity) is higher than that which the device can properly or efficiently handle for correct or intended operational performance.

overdub In audio recording, a method of combining two or more signals onto a single tape track. For example, a live voice can be recorded on a tape containing pre-recorded music.

overexcited Receiving higher than normal excitation, as in radio-frequency amplifiers or alternating-current generators.

overflow **1.** In computer or calculator operation, the condition in which an arithmetic operation yields a result exceeding the capacity of the location or display for a result. **2.** The carry digit that results from the condition described in (1).

overflow indicator **1.** In a digital calculator, a display that indicates that a numerical value is too large or too small to be shown with the available number of decimal places. **2.** In data processing, a display that indicates the presence of too many bits or characters for the available storage capacity.

overflow position In a digital computer, an auxiliary register position for developing the overflow digit (see OVERFLOW, **1**, **2**).

overflow record In data processing, a record that will not fit the storage area allotted for it, and that must be kept where it can be retrieved, according to some reference stored in its place.

overflow storage In a calculator or computer, extra storage space, allowing a small amount of overflow without loss of accuracy.

overhanging turns The turns in the unused portion(s) of a tapped coil.

overhead line A power or transmission line suspended above the ground between poles or towers.

over-horizon radar A form of radar used at high frequencies, in which pulses are transmitted and received. The signals are returned to earth via the ionosphere, both in the forward and reflected directions, making it possible to detect such things as missile launchings from thousands of miles away.

over-horizon transmission See FORWARD SCATTER.

overinsulation Use of excessive insulation for a particular application. Compare UNDERINSULATION.

over insulation The insulation (usually a strip of tape) laid over a wire brought up from the center of a coil. Compare UNDER INSULATION.

overlap **1.** The time during which two successive operations are performed simultaneously. **2.** In a facsimile or television system, a condition in which the scanning line is wider than the center-to-center separation between adjacent scanning lines.

overlap radar A long-range radar situated in one sector and covering part of another sector.

overlay **1.** A sheet of transparent or translucent material laid over a schematic diagram for the purpose of tracing connections that have been made in wiring an equipment from the diagram. **2.** In computer operations, a method whereby the same internal storage locations are used for different parts of a program during a program run. It is used when the total storage requirements for instructions exceed the available main storage capacity.

overlay transistor A double-diffused epitaxial transistor having separate emitters connected together by means of diffusion and metallizing to increase the edge-to-area ratio of the emitters. This design raises the current-handling ability of the transistor. Also see DIFFUSED TRANSISTOR and EPITAXIAL TRANSISTOR.

overload **1.** Current or power drain in excess of the rated output of a circuit or device. **2.** An excessive driving signal.

overload circuit breaker See CIRCUIT BREAKER.

overloaded amplifier A power amplifier delivering excessive output power. Compare UNDERLOADED AMPLIFIER, **2**.

overloaded oscillator An oscillator from which excessive power is drawn, causing instability, frequency shift, lowered output voltage, and overheating.

overload indication Any attention-catching method, such as an audible or visual alarm, for warning that a prescribed signal or power level has been exceeded.

overload level The amount of overload that can safely be applied to an equipment (see OVERLOAD, **1**).

overload protection The use of circuit breakers, relays, automatic limiters, and similar devices to protect equipment from overload damage by reducing current or voltage, disconnecting the power supply, or both.

overload recovery time Following an overdrive at the input of an integrated-circuit device, the time required for the output to resume its normal characteristics.

overload relay A relay actuated when circuit current exceeds a predetermined value. Compare UNDERLOAD RELAY.

overload time The maximum length of time that an equipment can safely be subjected to an overload level of current.

overmodulation Modulation in excess of a prescribed level—especially amplitude modulation greater than 100%. Compare COMPLETE MODULATION and UNDERMODULATION.

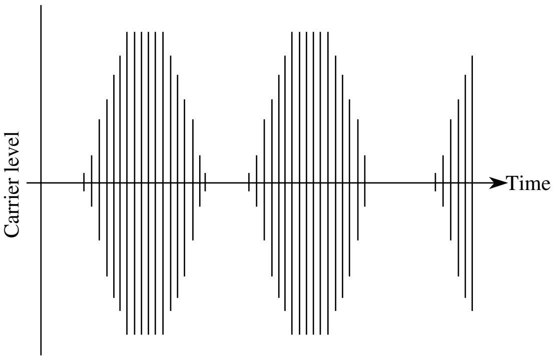

overmodulation

overmodulation alarm See OVERMODULATION INDICATOR.

overmodulation indicator A device, such as a neon bulb, incandescent lamp, light-emitting diode, analog meter, or digital meter, adapted to give an alarm when the modulation percentage of a signal exceeds a predetermined value.

overpotential See OVERVOLTAGE.

overpower relay A relay actuated by a rise in power above a predetermined level. Compare UNDERPOWER RELAY.

overpressure For a pressure transducer, pressure in excess of the maximum rating of the device.

override **1.** To intentionally circumvent an automatic control system. **2.** To bridge a functional stage of a system.

overscanning The deflection of the beam of a cathode-ray tube beyond the edges of the screen.

overshoot **1.** The momentary increase of a quantity beyond its normal maximum value (e.g., the spike sometimes seen on a square wave because of the overswing of a rising voltage). **2.** Momentary overtravel of the pointer of an analog meter.

overswing See OVERSHOOT, **2**.

overtemperature protection The use of an automatic device, such as a thermal relay or thermostat, to disconnect a device from the power supply when the device's temperature becomes excessive.

overthrow See OVERSHOOT, **2**.

overtone See HARMONIC.

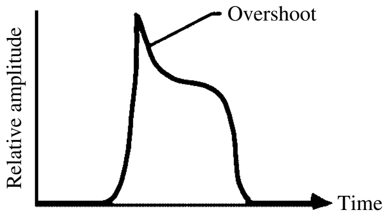

overshoot

overtone crystal A piezoelectric quartz crystal that oscillates at odd multiples of the frequency for which it was cut. This allows crystal operation at frequencies otherwise obtainable only from a fundamental-frequency crystal ground so thin as to be prohibitively fragile.

overtone oscillator A crystal oscillator using an OVERTONE CRYSTAL.

overtravel See OVERSHOOT, **2**.

overvoltage A voltage higher than a specified or rated value. Compare UNDERVOLTAGE.

overvoltage circuit breaker A circuit breaker that opens when voltage exceeds a predetermined value.

overvoltage protection The use of a special circuit or device to protect equipment from excessive voltage. When voltage increases beyond the overvoltage limit, the protective circuit causes shutdown.

overvoltage relay A relay actuated when voltage rises above a predetermined value. Compare UNDERVOLTAGE RELAY.

overwrite In computer operations, to record new data over existing data (e.g., to update the files on a magnetic disk or tape).

Ovshinsky effect In thin-film solid-state devices, the tendency for switches to have the same characteristics for currents in either direction.

O wave One (the ordinary) of the pair of components into which an ionospheric radio wave is divided by Earth's magnetic field. Compare X WAVE.

Owen bridge A wide-range four-arm bridge that measures inductance in terms of a standard capacitance and bridge-arm resistances.

own coding Additional program steps added to vendor-supplied software so that it can be modified to fit special needs.

ox Abbreviation of OXYGEN.

oxidation **1.** The combination of a substance with oxygen. Generally a slow process, such as the corrosion of iron or aluminum in the atmosphere. The process is accelerated by the presence of moisture and/or high temperatures. **2.** The loss of electrons from a cell or battery during discharge.

oxidation-reduction potential The potential at which oxidation occurs at the anode of an electrolytic cell, and at which reduction occurs at the cathode.

oxide-coated cathode See OXIDE-COATED EMITTER.

oxide-coated emitter An electron-tube cathode or filament coated with a material, such as thorium oxide, for increased electron emission at low emitter temperatures.

oxide-coated filament See OXIDE-COATED EMITTER.

oxide film **1.** The thin film of iron oxide that constitutes the recording surface of a magnetic disk or tape. **2.** The layer of copper oxide formed on the copper plate of a copper-oxide rectifier.

oxide-film capacitor An electrolytic capacitor, so called because the dielectric is a thin oxide film.

oxide rectifier A solid-state rectifier using a junction between copper and copper oxide. Also called COPPER-OXIDE RECTIFIER.

oximeter A photoelectric instrument for measuring the oxygen content of the blood. It operates by passing visible light through the earlobe, and analyzing the color and intensity of the emerging beam. Also called ANOXEMIA TOXIMETER.

oxygen Symbol, O. Abbreviation, O_2. A gaseous element. Atomic number, 8. Atomic weight, 15.999. Constitutes 21% of Earth's atmosphere. It readily combines with various elements to form compounds (see OXIDATION).

oxygen analyzer An electronic gas analyzer designed especially to measure oxygen content. The operation of this instrument is based on the paramagnetic properties of oxygen.

oxygen recombination In nickel-cadmium (NICAD) cells and batteries, a process in which oxygen is generated in the vicinity of the positive electrode, and is reduced with water in the vicinity of the negative electrode. This produces battery heating.

oz **1.** Abbreviation of OUNCE. **2.** Abbreviation of OZONE.

oz-in Abbreviation of OUNCE-INCH.

ozocerite An insulating mineral wax. Dielectric constant, 2.2. Dielectric strength, 4 to 6 kV/mm. Also spelled OZOKERITE.

ozone Symbol, O_3. An allotropic form of oxygen. Its formula indicates that each molecule has three atoms. Produced by the action of ultraviolet rays (or electrical discharge) on oxygen, its characteristic odor (somewhat like weak chlorine) can often be detected around sparking contacts or in the air after a thunderstorm.

ozone layer In the earth's atmosphere, a layer of ozone gas in the upper troposphere and lower stratosphere. It is produced by ultraviolet radiation from space, mainly from the sun. The ozone layer tends to block ultraviolet radiation, reducing the amount that reaches the surface of the earth.

ozone monitor An instrument for measuring the concentration of ozone in the atmosphere. One version measures the extent to which ultraviolet radiation is absorbed by a sample of air; the greater the absorption, the higher the ozone concentration in the sample.

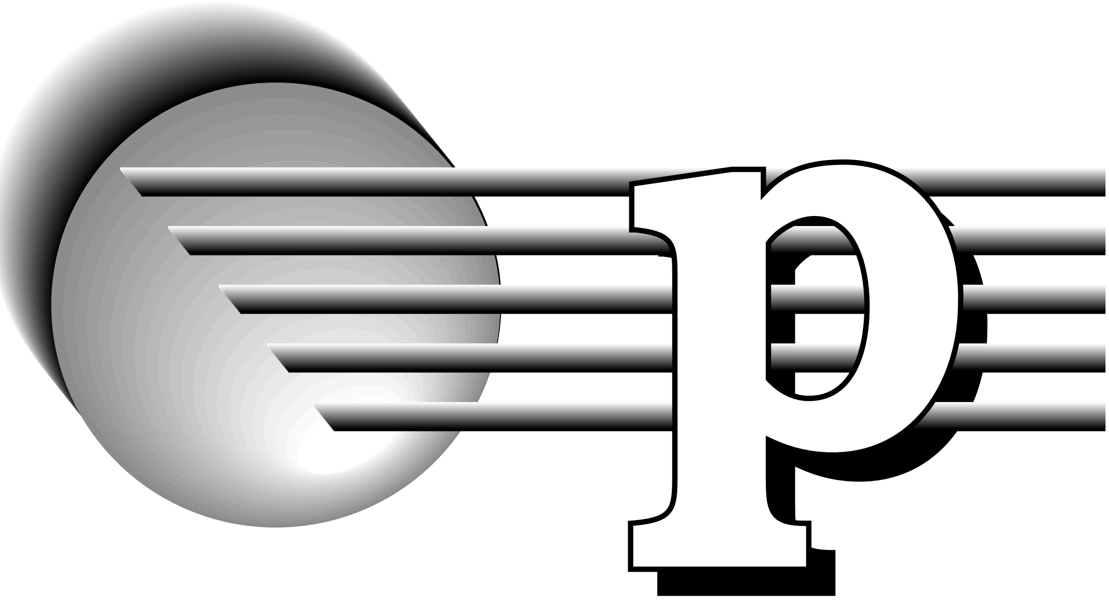

P **1.** Symbol for POWER. **2.** Symbol for PLATE (of a vacuum tube). **3.** Symbol for PHOSPHORUS. **4.** Abbreviation of PRESSURE. **5.** Symbol for PRIMARY. **6.** Abbreviation for prefix PETA-. **7.** Symbol for PERMEANCE. **8.** Abbreviation of POINT.

p **1.** Abbreviation of prefix PICO-. **2.** Subscript for PEAK. **3.** Abbreviation of POUND. **4.** Abbreviation of POINT (often capitalized). **5.** Subscript for PRIMARY. **6.** Subscript for PLATE (of a vacuum tube). **7.** Abbreviation of PITCH. **8.** Abbreviation of PER.

PA **1.** Abbreviation of POWER AMPLIFIER. **2.** Abbreviation of PULSE AMPLIFIER. **3.** Abbreviation of *particular average.* **4.** Abbreviation of *pilotless aircraft.* (Also, P/A.) **5.** Abbreviation of PUBLIC ADDRESS (as in PA system).

Pa **1.** Symbol for PROTACTINIUM. **2.** Symbol for PASCAL.

pA Abbreviation of PICOAMPERE.

pacemaker See CARDIAC STIMULATOR.

pacer See CARDIAC STIMULATOR.

Pacific Standard Time Abbreviation, PST. Local mean time at the 120th meridian west of Greenwich. Also see GREENWICH MEAN TIME, STANDARD TIME, TIME ZONE, and COORDINATED UNIVERSAL TIME.

pack A technique for maximizing a computer memory device's storage capacity, wherein more than one information item is stored in a single storage unit. Also called *crowd.*

package **1.** The enclosure for an electronic device or system. This includes a wide range of housings, from the simple encapsulation of miniature transistors to forced-air-cooled enclosures for heavy power units. **2.** To assemble and house an electronic equipment, or to design a housing for it, in accordance with good engineering techniques. **3.** A computer program of general use for an application (e.g., *payroll package*).

package count The number of discrete packaged circuits in a system.

packaging density **1.** See VOLUMETRIC EFFICIENCY. **2.** Computer storage capacity in terms of the number of information units that can be contained on a given segment of a magnetic medium. Also called PACKING DENSITY. **3.** Within a given integrated circuit, the capacity in terms of the number of active devices that can be contained on a single silicon chip.

packet **1.** A unit of digital information in PACKET COMMUNICATIONS. It consists of a header followed by a certain number of data bits or bytes. **2.** See WAVE PACKET. **3.** See PACKET COMMUNICATIONS. **4.** See PACKET RADIO.

packet communications A method via which data is exchanged through a network between or among people or computers. Information is sent and received in blocks of information called *packets.* Each packet is routed individually through the network according to the most efficient possible path at the time of its transit. At the destination, the packets are reassembled into the original signal. This scheme makes more efficient use of network resources than continuous-connection or single-path methods. However, when network usage is heavy, there can be a delay in the arrival of a sufficient number of packets to produce an intelligible received signal.

packet radio The transmission and reception of PACKET COMMUNICATIONS data via radio.

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

packet switching In telephony, a method of connection in which data is exchanged between subscribers by splitting the data into units (packets). Each packet is sent over the optimum path at the time of transmission. The signal path can, and usually does, vary from packet to packet. At the destination, the packets are reassembled into the original singal. The connection is in effect nonexistent during periods of silence (no data transmitted by either subscriber). Compare CIRCUIT SWITCHING.

packing In the button of a carbon microphone, bunching and cohesion between the carbon granules.

packing density The number of discrete package circuits within a given surface area or volume.

packing factor **1.** See VOLUMETRIC EFFICIENCY. **2.** In computer operations, the number of bits that can be recorded in a given length of magnetic memory surface. Also called PACKING DENSITY.

pack transmitter A portable transmitter that can be strapped to the operator's back.

pack unit A portable transceiver that can be strapped to the operator's back or carried on an animal's back.

PACM Abbreviation of *pulse-amplitude code modulation.*

pad **1.** An attenuator network (usually a combination of resistors) that reduces the amplitude of a signal by a desired amount while maintaining constant input and output impedance. **2.** In computer operations, to make a record a fixed size by adding blanks or dummy characters to it. **3.** To lower the frequency of an inductance-capacitance (LC) circuit by adding capacitance to an already capacitively tuned network.

padder See OSCILLATOR PADDER.

padding capacitor See OSCILLATOR PADDER.

padding character In a digital communications system, a character that is inserted solely for the purpose of consuming time while no meaningful characters are sent. The insertion of such characters maintains the synchronization of the system.

paddle-handle switch A toggle switch the lever of which is a flattened rod. Compare BAT-HANDLE SWITCH, ROCKER SWITCH, and SLIDE SWITCH.

PADT Abbreviation of POST-ALLOY-DIFFUSED TRANSISTOR.

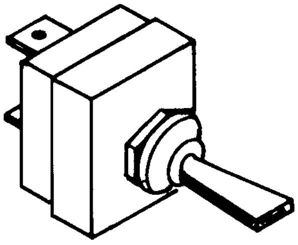

paddle-handle switch

page A display of text data on a computer display that completely fills the screen.

page printer A computer peripheral that prints a message in lines on a page, according to an established format, rather than in a single line.

pager **1.** A public-address system used for summoning purposes. **2.** See BEEPER, **2**.

page turning The successive display of pages (see PAGE).

pair **1.** Two wires, especially two insulated conductors in a cable. **2.** A set of two particles or charge carriers (e.g., *electron-hole pair*). **3.** A set of two transistors or vacuum tubes, operating together in push-pull or parallel in a power amplifier.

paired cable A cable consisting of separate twisted pairs of conducting wires.

paleomagnetism The study of certain rocks and minerals to determine the nature of the earth's magnetic field at the time the rocks were formed. When the age of the rock is determined by means of radioactive dating, and numerous rock samples are found covering many different eras, the nature of the earth's magnetic field can be graphed over time.

palladium Symbol, Pd. A metallic element of the platinum group. Atomic number, 46. Atomic weight, 106.42.

palletizing In industrial robots, the automatic placing of objects in a tray according to a computer program.

Palm See HANDHELD COMPUTER.

Palmer scan In radar, a method of simultaneously scanning the azimuth and the elevation.

PalmPilot See HANDHELD COMPUTER.

palmtop computer See HANDHELD COMPUTER.

PAM Abbreviation of PULSE-AMPLITUDE MODULATION.

Pan In radiotelephony, a spoken word indicating that an urgent message is to follow. It is equivalent to the XXX of radiotelegraphy.

pan **1.** To make a panoramic sweep [e.g., to sweep a wide area with a beam (as from an antenna), or to sweep a wide band of frequencies with a suitable tuning circuit]. **2.** A panoramic sweep made as defined in **1**. **3.** In audio engineering, to gradually shift from one audio channel to another or from one reproducer to another.

pan and tilt **1.** An azimuth-elevation mounting for a television camera. **2.** The simultaneous movement of a television camera in the vertical and horizontal directions.

pancake coil See DISK WINDING.

panel A flat surface on which are mounted the controls and indicators of an equipment, for easy access to the operator.

panel lamp **1.** See ELECTROLUMINESCENT PANEL. **2.** See PANEL LIGHT.

panel light A pilot light for illuminating the front panel of a piece of equipment.

panel meter A usually small meter for mounting on, or through an opening in, a panel.

panic button In a security system, a button or switch that immediately triggers an alarm when it is closed.

panoramic adapter An external device that can be connected to a receiver to sweep a frequency band and indicate carriers on the air as pips on a screen at the corresponding frequency points. Also called *pan adapter.*

panoramic display **1.** A wide-angle display. **2.** A spectrum-analyzer display that shows a wide range of frequencies, from zero to well above the maximum frequency in the monitored system.

panoramic radar An omnidirectional radar (i.e., one that transmits wide-beam signals in all directions without scanning).

panoramic receiver A receiver that displays pips on a screen to show carriers on the air in a given frequency band. All frequencies in the band are presented along the horizontal axis of the screen.

panpot A potentiometer with which panning can be achieved (see PAN, **3**).

pan-range A form of radar display in which target motion can be ascertained.

pantography The transmission of radar information to a distant location for observation or recording.

Papa Phonetic alphabet code word for the letter P.

paper advance mechanism In a data-processing system, the part of a printer that moves (sometimes by computer control) the paper through the printer.

paper capacitor A component that is made by placing paper, soaked with mineral oil, between two strips of foil. The assembly is rolled up, and wire leads are attached to the two pieces of foil. Finally, the rolled-up foil and paper are enclosed in a cylindrical case. These components are sometimes found in radio-frequency (RF) electronic equipment. They have values ranging from about 0.001 microfarads (µF) to 0.1 µF, and can handle low to moderate voltages, usually up to about 1000 volts. Compare CERAMIC CAPACITOR, ELECTROLYTIC CAPACITOR, MICA CAPACITOR, PLASTIC-FILM CAPACITOR, TANTALUM CAPACITOR.

PAR Abbreviation of PRECISION APPROACH RADAR.

par Abbreviation of PARALLEL.

parabola A plane curve that is the locus of points that are equidistant from a fixed point (the focus) and a fixed straight line (the directrix). In the Cartesian *xy*-plane, the general equation is $y = ax2 + b$, where a and b are constants.

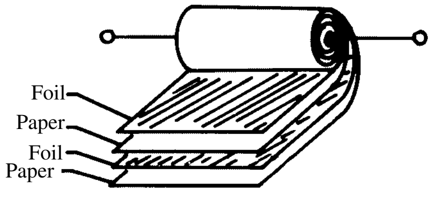

paper capacitor

parabola control See VERTICAL-AMPLITUDE CONTROL, **2**.

parabola generator A circuit for generating a parabolic-waveform signal.

parabolic microphone A directional microphone mounted at the principal focus of a parabolic sound reflector; the front of the microphone faces the inside of the parabola. It is useful for detecting sounds from great distances.

parabolic reflector Also called *paraboloidal reflector.* A reflector having the shape of a paraboloid. It is particularly useful for focusing or directing radiation. For example, if a radiator, such as an antenna rod, is placed at the focus of the paraboloid, a beam of parallel rays will be emitted by the reflector.

paraboloid The surface generated by a PARABOLA rotated about its axis of symmetry.

paraffin A relatively inexpensive, easily available, solid, white petroleum wax. At one time, it was used to impregnate capacitors and coils and to waterproof paper used for insulating purposes.

parallax The apparent shift in the position of a relatively nearby object when the observer moves or alternately blinks either eye. Thus, a pointer-type meter will seem to give different readings when viewed from different angles. Some meters have mirrored scales to eliminate this effect.

parallel **1.** Pertaining to the type of operation in a computer when all elements in an information item (e.g., bits in a word) are acted upon simultaneously, rather than serially (one at a time). **2.** The condition in which two comparably sized objects or figures are equidistant at all facing points. **3.** Pertaining to the shunt connection of components or circuits.

parallel access In computer operations, inputting or outputting data to or from storage in whole elements of information items (a word, rather than a bit at a time, for example).

parallel adder In a computer or calculator, an adder in which corresponding digits in multibit numbers are added simultaneously. Also see PARALLEL, **1**.

parallel antenna tuning Antenna-feeder tuning in which the tuning capacitor is connected in parallel with the two feeder wires. Compare SERIES ANTENNA TUNING.

parallel arithmetic unit See PARALLEL ADDER.

parallel capacitance **1.** A capacitance connected in parallel (shunt) with some other component. **2.** The capacitance between the turns of a coil. Also see DISTRIBUTED CAPACITANCE.

parallel capacitors Two or more capacitors connected in parallel (shunt) with each other. The total capacitance is equal to the sum of the individual capacitances. Also see PARALLEL CIRCUIT.

parallel circuit A circuit in which the components are connected across each other (i.e., so that the circuit segment could be drawn showing component leads bridging common conductors as rungs would across a ladder). Compare SERIES CIRCUIT.

parallel-component amplifier An amplifier stage in which the active devices (transistors or vacuum tubes) are connected in parallel with each other for increased power output. Also see PARALLEL CIRCUIT.

parallel-component oscillator An oscillator stage in which transistors are connected in parallel with each other for increased power output. Also see PARALLEL CIRCUIT.

parallel computer A computer equipped to handle more than one program at a time, but not through the use of multiple programming or time-sharing.

parallel-cut crystal See Y-CUT CRYSTAL.

parallel-diode half-wave rectifier See PARALLEL LIMITER.

parallel-fed amplifier An amplifier circuit in which the direct-current operating voltage is applied in parallel with the alternating-current output voltage. Also see PARALLEL FEED.

parallel-fed oscillator An oscillator circuit in which the direct-current operating voltage is applied in parallel with the alternating-current output voltage. Also see PARALLEL FEED.

parallel feed **1.** The presentation of parallel alternating-current (ac) and direct-current (dc) voltages to a device. **2.** The presentation of a dc operating voltage in parallel with the ac output voltage of a device (as in a parallel-fed amplifier or oscillator). Also see SHUNT FEED.

parallel gap welding A welding technique using two electrodes separated by a gap.

parallel gate circuit **1.** A gate circuit using two bipolar transistors with parallel-connected collectors and emitters, and a common collector resistor. The input signal is applied to one base, and the control signal to the other. **2.** A gate circuit using two field-effect transistors with parallel-connected drains and sources, and a common drain resistor. The input signal is applied to one gate, and the control signal to the other.

parallel inductance An inductance connected in parallel (shunt) with some other component.

parallel inductors Inductors connected in parallel and separated or oriented to minimize the effects of mutual inductance. Also see PARALLEL CIRCUIT.

parallel inverse feedback In a single-ended audio amplifier circuit, a simple system for obtaining negative feedback: A high resistance is connected from the output-transistor collector or drain to the driver-transistor collector or drain.

parallel limiter A limiter (clipper) circuit in which the diode is in parallel with the signal. Compare SERIES LIMITER.

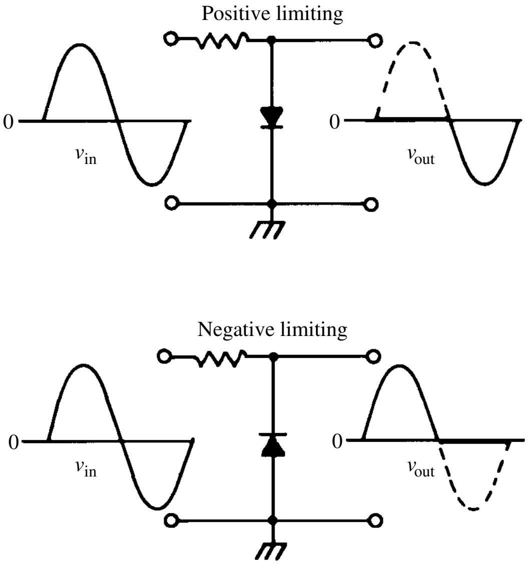

parallel limiter

parallel-line tuning At ultra-high frequencies (UHF) and microwave frequencies, the use of two parallel wires or rods for tuning. A straight short-circuiting bar is slid along the wires to accomplish tuning.

parallelogram A two-dimensional geometric figure that has four sides. Opposite pairs of sides are parallel. Opposite interior angles have equal measure.

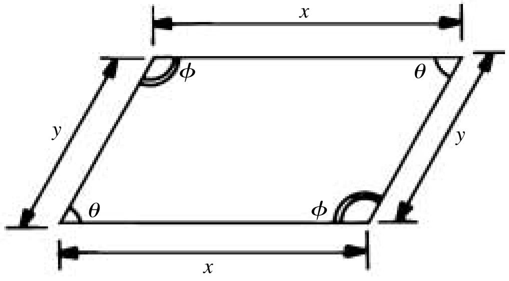

parallelogram

parallelogram of vectors A graphic device for finding the sum of two vectors. A parallelogram is constructed for which the two vectors are adjacent sides. The sum of the vectors is represented by the diagonal of the parallelogram.

parallel operation In computer operations, the simultaneous transmission of all bits in a multibit word over individual lines, as compared with the serial transmission of a word bit by bit.

parallel output A digital output consisting of two or more lines, all of which carry data at the same time.

parallel processing In computer operations, the simultaneous processing of several different programs through separate channels. Compare SERIAL PROCESSING.

parallel *Q* Symbol, Q_p. The figure of merit of a parallel circuit of inductance, capacitance, and resistance.

parallel resistance **1.** A resistance connected in parallel (shunt) with some other component. **2.** The resistance between the plates of a capacitor. **3.** The resistance across a coil.

parallel resistors Resistors connected in parallel. If the individual resistances are represented by R_1, R_2, R_3, . . ., Rn, then total resistance R_t is equal to $1/(1/R_1 + 1/R_2 + 1/R_3 + \ldots + 1/R_n)$. Also see PARALLEL CIRCUIT.

parallel resonance Resonance in a circuit consisting of a capacitor, inductor, and alternating-current source connected in parallel. At the resonant frequency, the inductive reactance is equal in magnitude, but opposite in effect, to the capacitive reactance. The capacitor current and inductor current are maximum, the line current is minimum, and the circuit impedance is maximum. Compare SERIES RESONANCE.

parallel-resonant circuit A resonant circuit in which the capacitor, inductor, and alternating-current source are connected in parallel. Compare SERIES-RESONANT CIRCUIT.

parallel-resonant trap A wavetrap consisting of a parallel-resonant inductance-capacitance (LC) circuit. Compare SERIES-RESONANT TRAP.

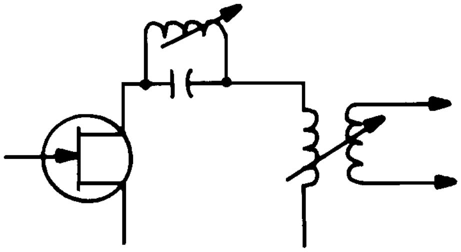

parallel-resonant trap

parallel-resonant wavetrap See PARALLEL-RESONANT TRAP.

parallel-rod oscillator An ultra-high-frequency (UHF) oscillator tuned by means of two straight, parallel quarter- or half-wave rods, one rod connected to the base or gate of a transistor, and the other rod connected to the collector or drain.

parallel-rod tuning Adjustment of the resonant frequency of a section of open-wire transmission line. A movable shorting bar allows quarter-wave resonance. The impedance at resonance is very high.

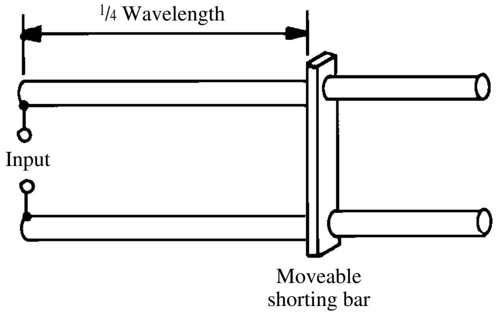

parallel-rod tuning

parallel-series Also called *series-parallel*. Pertaining to an arrangement of components, usually similar (e.g., resistances), consisting of parallel circuits connected in series with each other, or of series circuits connected in parallel with each other. Usually, the component values or ratings are all identical, so currents and/or voltages are equally shared among them. Also see PARALLEL CIRCUIT and SERIES CIRCUIT.

parallel-series capacitors Capacitors connected in PARALLEL-SERIES, usually to obtain higher voltage and/or current ratings than an individual capacitor can provide.

parallel-series inductors Inductors connected in PARALLEL-SERIES and separated or oriented to minimize the effects of mutual inductance.

parallel-series resistors Resistors connected in PARALLEL-SERIES, usually to obtain a higher power rating than an individual resistor can provide.

parallel storage In a computer, storage in which all information items can be made available in the same amount of time.

parallel-tee amplifier A bandpass amplifier having a parallel-tee network in its negative-feedback path. The null frequency of the network determines the pass frequency of the amplifier.

parallel-tee measuring circuit A parallel-tee network used for measuring circuit constants. Also called TWIN-TEE MEASURING CIRCUIT.

parallel-tee network A resistance-capacitance (RC) network containing two tee sections (with R and C elements opposite in the tees) connected in parallel. The network produces a null at one frequency. Also called TWIN-T NETWORK.

parallel-tee oscillator A resistance-capacitance tuned oscillator having a parallel-tee network in its negative-feedback path. The null frequency of the network determines the oscillator frequency.

parallel transfer A form of digital information transfer, consisting of two or more lines that carry data at the same time.

parallel-wire line A transmission line consisting of two parallel wires whose separation is kept constant by dielectric rods (open-wire line) or a solid dielectric web (ribbon line).

parallel-wire tank In an ultra-high-frequency (UHF) amplifier or oscillator, a resonant circuit consisting of two separate parallel wires connected to the transistor(s) or tube(s) at one end, and short-circuited or tuned at the other end.

paramagnet A paramagnetic substance (see PARAMAGNETISM). Compare DIAMAGNET.

paramagnetic Possessing PARAMAGNETISM. Compare DIAMAGNETISM.

paramagnetism The state of having a magnetic permeability slightly greater than **1**. Compare DIAMAGNETISM.

parameter **1.** An operating value, constant, or coefficient that can be either a dependent or an independent variable (e.g., a transistor-electrode current or voltage). **2.** The ratio of one coefficient to another, where both are either fixed or variable (e.g., transconductance of a vacuum tube).

parameter word In a computer memory, a place having a capacity of a word (bit group) in which is stored a parameter for a program.

parametric amplifier A radio-frequency power amplifier based on the action of a voltage-variable capacitor in a tuned circuit.

parametric amplifier diode See VARACTOR.

parametric converter A frequency converter in which a parametric device, such as a varactor, is used to change a signal of one frequency to a signal of another frequency. Also see PARAMETRIC DOWN-CONVERTER and PARAMETRIC UP-CONVERTER.

parametric diode A variable-capacitance diode (see VOLTAGE-VARIABLE CAPACITOR, **1**).

parametric down-converter A parametric converter in which the output signal is of a lower frequency than the input signal. Compare PARAMETRIC UP-CONVERTER.

parametric equalizer A set of audio filters similar to a GRAPHIC EQUALIZER, except that the center frequencies are adjustable, rather than fixed. The center frequencies are selected by the operator; then the attenuation level (in decibels) is set for each frequency. It is used in audio recording studios.

parametric modulation Modulation in which either the inductance or capacitance of a tank circuit or coupling device is varied at the modulation frequency.

parametric oscillator An oscillator that generates visible light energy by means of a parametric amplifier and a tunable cavity.

parametric up-converter A parametric converter in which the output signal is of a higher frequency than the input signal. Compare PARAMETRIC DOWN-CONVERTER.

parametron See PHASE-LOCKED OSCILLATOR.

paramistor A device consisting of several digital circuit elements that use parametric oscillators.

paramp Abbreviation of PARAMETRIC AMPLIFIER.

paraphase inverter A single-transistor phase inverter in which the two out-of-phase output signals are obtained by taking one output from the collector or drain, and the other output from the emitter or source. Thus, the 180-degree phase difference between collector/drain and emitter/source is exploited.

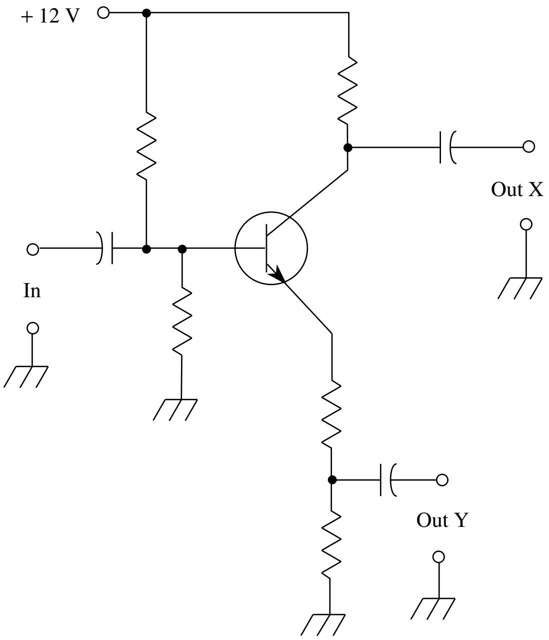

paraphase inverter

parasitic See PARASITIC OSCILLATION.

parasitic antenna See PARASITIC ARRAY, PARASITIC ELEMENT.

parasitic array Any of several types of directional antenna employing parasitic elements, sometimes in combination with phasing, to obtain directivity and gain. Common examples include the Yagi and quad. Arrays with numerous parasitic elements (usually directors) can produce upwards of 15 dBd forward gain. When several high-gain arrays are phased, the realizable gain becomes greater still. Two-element, three-element, and four-element arrays are common below 30 MHz. At very-high and ultra-high frequencies, especially above 100 MHz, bays of antennas of this type are used in satellite communications and radio astronomy. See PARASITIC ELEMENT, QUAD ANTENNA, YAGI ANTENNA.

parasitic capacitance Stray capacitance. It can be internal or external to a circuit and can introduce undesirable coupling or bypassing.

parasitic choke A small radio-frequency choke coil (with or without a shunting resistor) that suppresses or eliminates parasitic oscillation in a power amplifier.

parasitic director In a multielement directional antenna, a parasitic element acting as a director; usually, it is a few percent shorter than the driven element.

parasitic element An electrical conductor that comprises an important part of an antenna system, but that is not directly connected to the feed line. Such elements are used for the purpose of obtaining directivity and power gain. They operate via electromagnetic coupling to the driven element(s). The principle of operation was first discovered by the Japanese engineers Yagi and Uda, who observed that antenna elements parallel to a driven element but not connected to anything, at a specific distance from the driven element, and having a certain length, cause the radiation pattern to show gain in one direction and loss in the opposite direction. See DRIVEN ELEMENT, PARASITIC ARRAY, PARASITIC DIRECTOR, PARASITIC REFLECTOR.

parasitic-element directive antenna See PARASITIC ARRAY.

parasitic eliminator See PARASITIC SUPPRESSOR.

parasitic excitation Excitation of a beam-antenna element without a direct connection to the transmitter. Thus, a director or reflector element can be excited by the field of the radiator element.

parasitic inductance Stray inductance (e.g., the internal inductance of a wirewound resistor).

parasitic oscillation Extraneous, useless oscillation present as a fault in an electronic circuit, particularly a radio-frequency power amplifier.

parasitic reflector In a multielement beam antenna, a parasitic element acting as a reflector; usually, it is a few percent longer than the driven element.

parasitic resistance Stray resistance (e.g., the inherent, internal resistance of a multilayer coil).

parasitic suppressor A small resistor, coil, or parallel combination of the two, connected in series with the plate or collector of a vacuum tube or transistor to eliminate parasitic oscillations in a radio-frequency power amplifier.

PARD Abbreviation of PERIODIC AND RANDOM DEVIATION.

parity **1.** At par (with respect to the even-or-odd state of the characters in a group). **2.** Having the quality that the number of bits (or the number of similar bits) are even or odd, as intended.

parity bit **1.** In computer operations, a logic 1 added to a group of bits so that the number of 1s in the group is, according to specification, even or odd. **2.** In computer operations, a check bit that can be a logic 1 or 0, depending on the parity (see PARITY, **1**) of the total of 1s in the bit group being checked.

parity check A check of the integrity of data being transferred by adding the bits in, for example, a word, and then determining the parity bit needed and comparing that with the transmitted parity bit.

parity error An error disclosed by a parity check.

parity tree A digital device used to check parity.

parsec Abbreviation, pc. The distance at which the mean radius of the earth's orbit around the sun subtends an angle of 1 second of arc; 1 pc = 3.0857×10^{13} kilometers or 3.2616 light years.

part See CIRCUIT COMPONENT, **1**.

part failure The usually destructive breakdown of a circuit component.

partial One of the frequencies in a complex musical tone. It might be a harmonic of the fundamental frequency, although this is not always the case.

partial carry The temporary storage of some or all of the carry information in a digital calculation.

particle **1.** A tiny, discrete bit of matter. **2.** A unit of matter smaller and lighter than an atom. See, for example, ANTIPARTICLE, ELECTRON, MESON, NEUTRETTO, NEUTRINO, NEUTRON, NUCLEON, POSITRON, and PROTON.

particle accelerator See ACCELERATOR, **1**.

particle theory of radiation In physics, a model that explains the nature of electromagnetic radiation (radio waves, infrared, visible light, ultraviolet, X rays, and gamma rays) in terms of discrete particles. Each particle, called a PHOTON, carries a certain amount of energy that depends on the wavelength of the radiation.

particle velocity **1.** The speed and direction of the particles from a source of atomic radiation. **2.** The speed and direction of the molecules in the medium of an acoustic disturbance.

partitioning In computer operations, breaking down a large block of data into smaller blocks that can be better handled by the machine.

parton model A model for atomic nuclei, in which protons and neutrons are made up of smaller particles called *partons.* Subparticles have been found, commonly called *quarks.*

Pascal A high-level computer programming language, similar to BASIC or FORTRAN in structure. It is used in some schools to teach computer programming.

pascal Symbol, Pa. The SI (derived) unit of pressure; 1 $Pa = 1\ \mathrm{N/m^2} = 1.4503 \times 10^{-4}\ \mathrm{lb/in^2}$.

Paschen-Back effect See ZEEMAN EFFECT.

Paschen's law For a two-element, parallel-plate, gas-discharge tube, the plate-to-plate sparking potential is proportional to Pd, where P is the gas pressure, and d is the distance between plates.

pass amplifier A tuned amplifier having the response of a bandpass filter. Like the filter, the amplifier passes one frequency (or a narrow band of frequencies) readily while rejecting or attenuating others. Compare REJECT AMPLIFIER.

passband The continuous spectrum of frequencies transmitted by a filter, amplifier, or similar device. Compare STOPBAND.

passband ripple Multiple low-amplitude attenuation variations within the passband of a filter or tuner, resulting in a ripple pattern on the nose of the response curve.

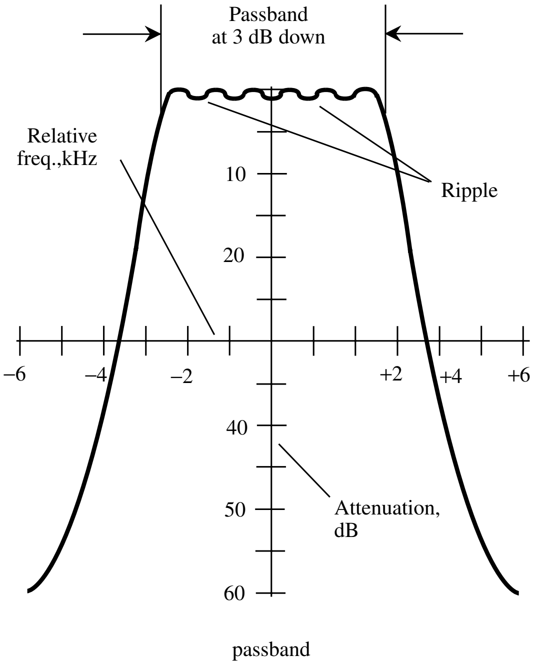

passband ripple

passivation The process of growing a thin oxide film on the surface of a planar semiconductor device to protect the exposed junction(s) from contamination and short circuits. See, for example, PLANAR EPITAXIAL PASSIVATED TRANSISTOR and PLANAR TRANSISTOR.

passive absorber A substance that reflects minimal sound energy. Examples include acoustical ceiling tile and thick carpeting.

passive circuit A circuit consisting entirely of nonamplifying components, such as capacitors, resistors, inductors, and diodes.

passive communications satellite A communications satellite that reflects electromagnetic waves, but does not contain a transponder; that is, it does not receive and retransmit the signals. Also called *passive comsat.* Compare ACTIVE COMMUNICATIONS SATELLITE.

passive component A device that is basically static in operation (i.e., it is ordinarily incapable of amplification or oscillation and usually requires no power for its characteristic operation). Examples: conventional resistor, capacitor, inductor, diode, rectifier, and fuse. Compare ACTIVE COMPONENT, **1**.

passive comsat See PASSIVE COMMUNICATIONS SATELLITE.

passive decoder A decoder that responds to only one signal code, rejecting all others.

passive detection In reconnaissance, detecting a target without betraying the location of the detector.

passive electric network See PASSIVE NETWORK.

passive frequency multiplier A frequency multiplier that does not require a power supply, but operates only from the input signal energy. Usually consists of one or more semiconductor diodes, sometimes in conjunction with inductors and/or capacitors. The output signals appear at integral multiples of the input frequency.

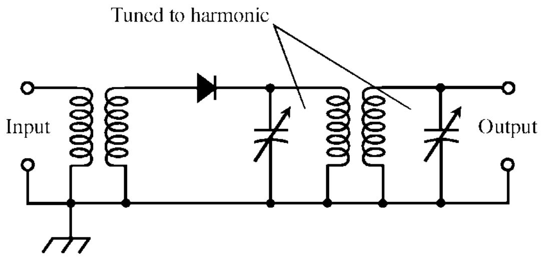

passive frequency multiplier

passive infrared sensor A device that detects infrared directly, such as that given off by humans because of their body heat. It does not generate energy of any kind. It is used in some intrusion detection systems.

passive mixer A signal mixer using only passive components (diodes, nonlinear resistors, and nonlinear reactances) (i.e., one without active components, such as transistors). Passive mixers introduce some loss. Compare ACTIVE MIXER.

passive modulator A modulator using only passive components (diodes, nonlinear resistors, and nonlinear reactances) (i.e., one without active components, such as transistors). Passive modulators introduce some loss. Compare ACTIVE MODULATOR.

passive network A network composed entirely of passive components (i.e., one containing no generators and providing no amplification).

passive radiator See DRONE CONE.

passive reflector A metal surface used to reflect electromagnetic energy at ultra-high and microwave frequencies.

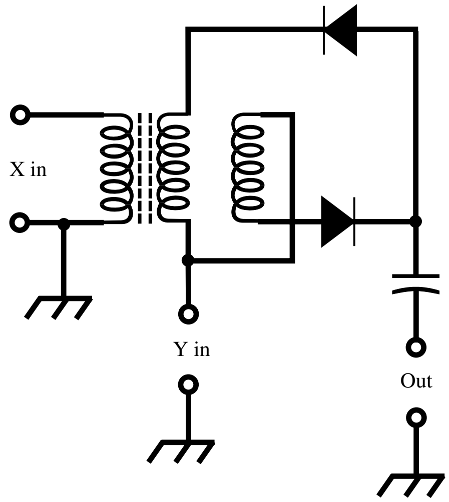

passive mixer

passive transponder A device that allows a machine, such as a computer or robot, to identify an object. A bar-code reader is a common example. Magnetic labels, such as those on credit cards and bank cash cards, are another example. It is so named because it does not transmit data; it requires a sensor for data detection.

password As a security device in computer operations, a group of characters upon whose presentation to the system via a terminal the user is allowed access to memory or control of information.

password retry limitation A security feature that prevents hackers from making repeated guesses at passwords in an attempt to break into a computer, network, or database. If more than three unsuccessful entries are made in succession, for example, the system will not accept further access attempts for a certain preprogrammed length of time, say 1 hour.

paste In "dry" batteries and electrolytic capacitors, a gelatinous electrolyte.

patch **1.** A temporary signal path, as between a radio receiver and a telephone or, conversely, between a telephone line and a radio transmitter. **2.** To make quick, usually temporary connections, as with a patch cord. **3.** Instructions entered by an unconditional branch to a computer program for the purpose of correction.

patch bay **1.** See PATCH PANEL. **2.** A set of patch panels.

patch cord A flexible line of one or more conductors with a jack or connector at each end, used to interconnect (patch) circuit points exposed for the purpose on a panel or breadboard.

Patchett tone control A dual tone-control circuit using a variable series resistance-capacitance (RC) filter for treble boost, and a variable shunt RC filter for bass boost. The input signal is applied in parallel to both filters. The outputs are combined in an audio mixer.

patching The interconnection of two or more signal media or lines.

patch panel A panel on which the terminals of a system are accessible for interconnection, tests, etc. It is used especially in high-fidelity audio recording systems. It was once commonly used in manual telephone-switching applications.

patch up **1.** To replace faulty or damaged parts in an electronic system with roughly appropriate surrogates to restore operation quickly (usually under emergency conditions). Also see DOCTOR. **2.** To wire a circuit quickly using patch cords for preliminary test and evaluation.

patent **1.** A document awarded by a government body, giving to an inventor the exclusive right to exploit an invention for a specified number of years. Formally called *letters patent.* **2.** The monopoly granted by a document, as defined in **1**.

path **1.** The route over which current flows. **2.** In radio and navigation, the imaginary line extending directly between transmitter and receiver (or target). **3.** In a computer program, the logical order of instructions.

pathometer A form of lie detector that indicates changes in the electrical resistance of the human body.

pattern **1.** An established sequence of steps in a process. **2.** An arrangement of terms in a matrix. **3.** The graphical representation of a varying quantity (e.g., an alternating-current wave pattern). **4.** The image on the screen of an oscilloscope, or the record traced by an oscillograph. **5.** The graphic polar representation of the radiation field of an antenna. **6.** The arrangement of bits in a word or field.

pattern recognition In machine-vision systems, a method of identifying an object or decoding data according to geometric shape. Optical character recognition (OCR) is an example. The machine recognizes combinations of shapes, and deduces their meanings via a computer program.

pause editing In the editing of audio tape recordings, the use of a "pause" switch to temporarily stop the tape when necessary.

PAV Abbreviation of PHASE-ANGLE VOLTMETER.

pawl In a mechanical stepping device, as in a nonelectric clock, a device made to engage the sloping sprockets on a wheel to ensure shaft rotation in one direction only.

PAX Abbreviation of PRIVATE AUTOMATIC EXCHANGE.

pay-per-view Abbreviation, PPV. Television service in which each subscriber pays only for individually selected programs.

pay TV See SUBSCRIPTION TV.

Pb Symbol for LEAD.

P band A radio-frequency band extending from 225 to 390 MHz.

PBX Abbreviation of PRIVATE BRANCH EXCHANGE.

PC **1.** Abbreviation of PERSONAL COMPUTER. **2.** Abbreviation of PRINTED CIRCUIT. **3.** Abbreviation of PHOTOCELL. **4.** Abbreviation of POSITIVE COLUMN. **5.** Abbreviation of POINT-CONTACT. **6.** Abbreviation of PERCENT (also, pct.). **7.** Abbreviation of PROGRAM COUNTER.

pc **1.** Abbreviation of PICOCOULOMB. Also, pC (preferred). **2.** Abbreviation of PICOCURIE. Also, pCi (preferred). **3.** Abbreviation of PARSEC.

pC Abbreviation for PICOCOULOMB.

PCB Abbreviation of PRINTED-CIRCUIT BOARD.

PC board See PRINTED-CIRCUIT BOARD.

PC diode See POINT-CONTACT DIODE.

p-channel JFET See P-CHANNEL JUNCTION FIELD-EFFECT TRANSISTOR.

p-channel junction field-effect transistor Abbreviation, PFET. A junction-type FET in which the gate junction has been formed on a bar or die of p-type semiconductor material. Compare N-CHANNEL JUNCTION FIELD-EFFECT TRANSISTOR.

p-channel MOSFET A metal-oxide semiconductor field-effect transistor in which the channel is composed of p-type silicon. Also see DEPLETION-TYPE MOSFET, DEPLETION-ENHANCEMENT-TYPE MOSFET, and ENHANCEMENT-TYPE MOSFET.

pCi Symbol for PICOCURIE.

PCL Abbreviation of PRINTED-CIRCUIT LAMP.

PCM Abbreviation of PULSE-CODE MODULATION.

PCM-FM Pertaining to a carrier that is frequency modulated by information that is pulse-code modulated. Also see FREQUENCY MODULATION and PULSE-CODE MODULATION.

PCM-FM-FM Pertaining to a carrier that is frequency modulated by one or more subcarriers that are frequency modulated by information that is pulse-code modulated. Also see FREQUENCY MODULATION and PULSE-CODE MODULATION.

PCM level In a pulse-code-modulated signal, one of several different possible signal conditions.

PCM-PM Pulse-code modulation that is accomplished by varying the phase of the carrier wave.

PC relay See PRINTED-CIRCUIT RELAY.

PCS Abbreviation for PERSONAL COMMUNICATION SERVICE.

PC transistor See POINT-CONTACT TRANSISTOR.

PD **1.** Abbreviation of PLATE DISSIPATION. **2.** Abbreviation of PULSE DURATION. **3.** Abbreviation of PROXIMITY DETECTOR. **4.** Abbreviation of POTENTIAL DIFFERENCE.

Pd Symbol for PALLADIUM.

PDA Abbreviation for *personal digital assistant.* See HANDHELD COMPUTER.

PDAS Abbreviation of *programmable data acquisition system.*

P display See PLAN POSITION INDICATOR.

PDM Abbreviation of PULSE-DURATION MODULATION.

PDM-FM Pertaining to a carrier that is frequency modulated by one or more subcarriers that are frequency modulated by pulses that are pulse-duration modulated. Also see FREQUENCY MODULATION and PULSE-DURATION MODULATION.

PDM-FM-FM Pertaining to a carrier that is frequency modulated by one or more subcarriers that are frequency modulated by pulses that are pulse-duration modulated. Also see FREQUENCY MODULATION and PULSE-DURATION MODULATION.

PDM-PM Pertaining to a carrier that is phase modulated by pulse-duration-modulated information. Also see PHASE MODULATION and PULSE-DURATION MODULATION.

PDT Abbreviation of PACIFIC DAYLIGHT TIME.

PDVM Abbreviation of PRINTING DIGITAL VOLTMETER.

PE **1.** Abbreviation of POTENTIAL ENERGY. **2.** Abbreviation of PROFESSIONAL ENGINEER. **3.** Abbreviation of PROBABLE ERROR.

peak **1.** The maximum value of a quantity. **2.** In an alternating-current cycle, the maximum positive or negative current or voltage point. **3.** The frequency at which the transmission by a bandpass circuit or device is maximum (attenuation is minimum), evidenced by a maximum in the frequency-response curve.

peak amplitude **1.** The maximum positive or negative current or voltage of a wave. **2.** The maximum instantaneous power of a signal.

peak anode (plate) current The maximum instantaneous current flowing in the anode (plate) circuit of a vacuum tube.

peak anode (plate) voltage The maximum instantaneous voltage applied to the anode (plate) of a vacuum tube.

peak chopper See PEAK CLIPPER.

peak current Abbreviation, I_p. The highest value reached by an alternating-current half-cycle or a current pulse. Also called MAXIMUM CURRENT.

peak detector See PEAK PROBE.

peak distortion **1.** The maximum instantaneous distortion in a signal, generally expressed as a percentage. **2.** Distortion of a modulated signal at envelope peaks.

peaked sawtooth A wave composed of a sawtooth and peaking-pulse components. The deflection voltage of a magnetic-deflection cathode-ray tube requires this waveform to produce a current sawtooth in the deflecting coils.

peaked waveform An alternating-current waveform having nearly pointed positive and negative half-cycles. Such a wave contains appreciable third-harmonic energy.

peak envelope power Abbreviation, PEP. For a linear radio-frequency (RF) power amplifier handling a modulated signal, the average RF output

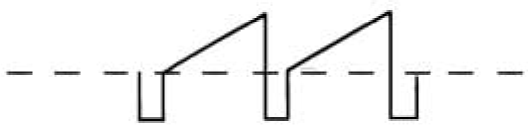

peaked sawtooth

power during a single RF cycle at the highest peak of the modulation envelope.

peaker **1.** See PEAK FILTER. **2.** See PEAKING TRANSFORMER.

peaker-notcher See NOTCHER-PEAKER.

peak factor For an alternating-current wave, the ratio E_m/E_{rms} or I_m/I_{rms}, where E_m is the maximum voltage, E_{rms} is the effective (root-mean-square) voltage, I_m is the maximum current, and I_{rms} is the effective current.

peak filter A frequency-selective circuit, such as a bandpass filter, for producing a peak response (see PEAK, **3**).

peak inductor current In a switching regulator, the maximum instantaneous current through the inductor when the device is switching at its fully rated duty cycle.

peaking The adjustment of a control or device for maximum indication on a meter or other display.

peaking coil A small inductor used to compensate the frequency response of a circuit, such as a video amplifier or video detector. Both series and shunt peaking coils are used.

peaking transformer A transformer whose output waveform is sharply peaked (of short duration, with respect to a cycle). The effect is obtained by means of a special core that, because it contains little iron, saturates easily.

peak inverse voltage Abbreviation, PIV. Often used interchangeably with the term PEAK REVERSE VOLTAGE. **1.** The peak value of the voltage applied to a rectifier diode in the reverse direction. **2.** The maximum value of reverse voltage that a rectifier diode will tolerate according to its specifications.

peak level lamp In audio recording and reproduction, a bulb or light-emitting diode (LED) that illuminates when sound peaks exceed a predetermined amplitude.

peak limiting **1.** A method of limiting the maximum amplitude of a signal. When the instantaneous peak amplitude, either positive or negative, exceeds a certain value, the output is clipped at that value. **2.** In pulse-code modulation, the effect resulting from the application of an input signal in excess of the virtual decision value.

peak modulated power In an amplitude-modulated wave, the maximum instantaneous signal power (including the carrier and sidebands). In 100-percent sinusoidal modulation, the peak modulated power is four times the unmodulated carrier power.

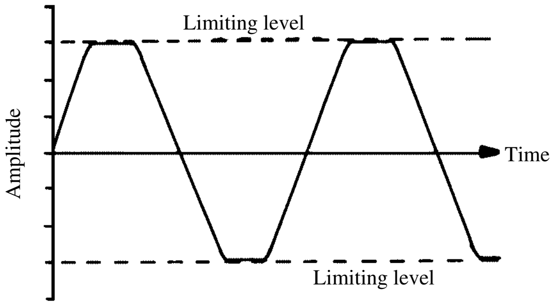

peak limiting

peak point The highest current point in the current-voltage response curve of a tunnel diode. Immediately beyond this point, the current decreases as the applied voltage is increased, indicating a negative-resistance region. Compare VALLEY POINT.

peak power **1.** Symbol, P_p. Unit, watt. Alternating-current power that is the product of the peak voltage (E_p) and the peak current (I_p). For E_p in volts and I_p in amperes, the peak power in watts is given by $P_p = E_pI_p$. **2.** The highest output power that an amplifier or device can produce without excessive distortion. **3.** The maximum instantaneous power that a speaker can handle without risk of damage.

peak probe A voltmeter test probe containing a diode circuit whose direct-current output voltage is close to the peak value of the applied alternating-current test voltage.

peak recurrent forward current For a semiconductor diode, the maximum repetitive instantaneous forward current as measured under specified conditions of operation.

peak reverse voltage Abbreviation, PRV. In semiconductor operations, the peak value of the voltage applied in reverse polarity across the junction. It is often used interchangeably with the term PEAK INVERSE VOLTAGE.

peak signal level The maximum instantaneous signal power or voltage specified for particular operating conditions.

peak-to-peak Abbreviations, p-p or pk-pk. For an alternating-current waveform, pertaining to the arithmetic difference between the positive peak and negative peak values of current or voltage.

peak-to-peak probe A voltmeter test probe containing a diode circuit whose direct-current output voltage is close to the peak-to-peak value of the applied alternating-current test voltage.

peak-to-peak voltage The arithmetic sum of positive and negative peak voltages in an alternating-current (ac) wave. Thus, a symmetrical sine-wave ac voltage of 115.0 V rms has a peak value of 162.63 V and a peak-to-peak value of 325.3 V. Also see PEAK VOLTAGE.

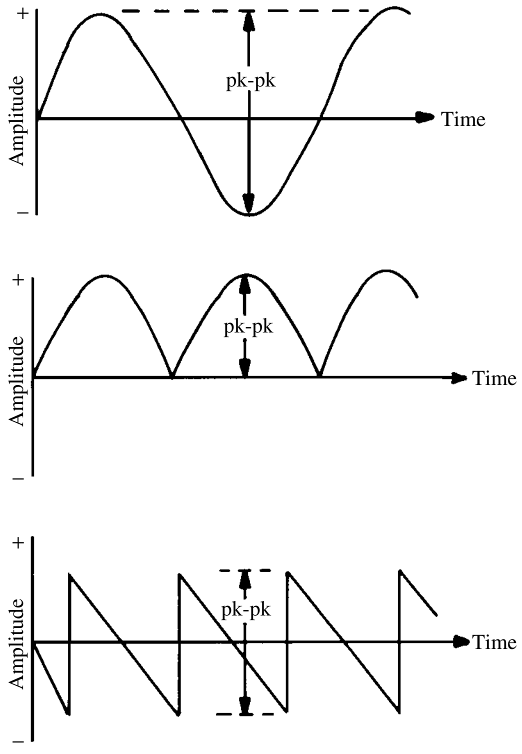

peak-to-peak voltage

peak torque Symbol, T_p. For a torque motor, the maximum useful torque at maximum recommended input current.

peak voltage Abbreviation, E_p. The highest value reached by an alternating-current voltage half cycle, or by a voltage pulse. Also called MAXIMUM VOLTAGE.

peak voltmeter **1.** An alternating-current (ac) voltmeter that responds to the peak value of the applied voltage. **2.** An ac voltmeter that responds to the average value of the applied voltage—even though its scale reads in peak volts.

pea lamp A miniature incandescent bulb, sometimes used as a control-panel or meter light.

PEC Abbreviation of PHOTOELECTRIC CELL.

pedestal See BLANKING PEDESTAL.

pedestal level See BLANKING LEVEL.

pel See PIXEL.

Peltier effect A drop below ambient temperature at the junction between two dissimilar metals when an electric current is passed through the junction.

PEM Abbreviation of *photoelectromagnetic.*

pen-and-ink recorder A graphic recorder in which a fountain-pen-type stylus inscribes an ink line on a paper chart. Also called *pen recorder.*

pencil **1.** A beam of electrons or other particles or rays that either converges to, or diverges from, a specific point. **2.** A pair of geometric entities sharing a property (e.g., lines intersecting at a single point).

pendulum switch A device that closes a circuit when subjected to physical shock. One type consists of a dangling element resembling a pendulum, with one or more nearby contacts.

penetrating frequency For a particular layer of the ionosphere, the lowest high frequency at which a vertically propagated wave penetrates the layer (i.e., it is not reflected back to earth). Also called CRITICAL FREQUENCY.

penetrating radiation Ionizing radiation that passes through otherwise opaque materials. A relative term; low-energy X rays are less penetrating than high-energy X rays, which, in turn, are less penetrating than gamma rays.

penetrating rays See COSMIC RAYS.

penetration depth See DEPTH OF PENETRATION.

pent Abbreviation of PENTODE.

pentavalent element An element whose atoms have five valence electrons (e.g., antimony or arsenic).

pentode A five-electrode vacuum tube in which the electrodes are an anode, cathode, control grid, screen grid, and suppressor grid.

pentode field-effect transistor A field-effect transistor with three separate gates.

pentode transistor A bipolar transistor with three emitters.

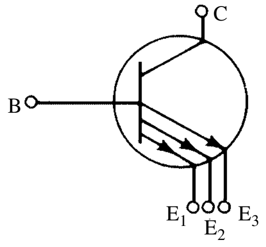

pentode transistor

penumbra **1.** That part of a shadow in which the light source is not fully obscured by the eclipsing object. **2.** In a sunspot, the outer part of the spot; it is less dark than the inner portion.

PEP **1.** Abbreviation of PLANAR EPITAXIAL PASSIVATED. **2.** Abbreviation of PEAK ENVELOPE POWER.

PEP diode See PLANAR EPITAXIAL PASSIVATED DIODE.

PEP reading wattmeter A wattmeter that shows the peak envelope power output of a transmitter.

PEP transistor See PLANAR EXPITAXIAL PASSIVATED TRANSISTOR.

perceived level The level of a disturbance, particularly sound, as sensed by a person. It is generally expressed in decibels, with respect to a certain threshold value. The threshold is assigned an intensity of 0 dB.

percent An expression of a fraction, in terms of hundredths. A quantity of *x* percent indicates a fraction of $x/100$. Percent is usually abbreviated by the symbol %.

percentage error The amount by which a measured value differs from the true value, expressed as a percentage (the number of parts per 100 that the measurement is in error).

percentage-modulation meter An instrument that provides direct readings of the modulation percentage of an amplitude-modulated signal. The meter scale or dial is graduated in increments from 0 to somewhat more than 100 percent.

percentage uncertainty The maximum possible error in a measurement, expressed as a percentage of the measured value. Also see UNCERTAINTY IN MEASUREMENT.

percent distortion Symbol, $\%D$. In the determination of harmonic distortion, the total harmonic voltage expressed as a percentage of the fundamental voltage, plus total harmonic voltage; $\%D = 100E_h/E_t$, where E_h is the total voltage of the harmonic components, and E_t is the total signal voltage (fundamental plus harmonics).

percent modulation See MODULATION PERCENTAGE.

percent modulation meter See PERCENTAGE-MODULATION METER.

percent ripple The amount of ripple voltage in the direct-current (dc) output of a rectifier or generator, expressed as a percentage of the nominal dc output voltage.

perfect crystal A crystal without defects or impurities. The atoms are arranged in a regular pattern with no faults.

perforated board A plastic panel provided with a number of small holes in orderly columns and rows for the insertion of the pigtails of components, or of push-in terminals to facilitate quick assembly of prototype circuits. Also called *perfboard.*

performance curve A curve depicting the behavior of a component or circuit under specified conditions of operation. Such a curve, for example, might display the variation of output power with input power, the variation of frequency with voltage, etc. Compare CHARACTERISTIC CURVE.

performance test A test made primarily to ascertain how a system behaves. The test is concerned with normal operation, whereas a diagnostic test is a troubleshooting procedure. Compare TROUBLESHOOTING TEST.

perigee **1.** The point at which an earth-orbiting satellite attains its lowest altitude. It occurs once for every complete orbit. At this point, the satellite travels faster than at any other point in the orbit. **2.** The altitude, measured from the earth's surface or the earth's center, of an earth-orbiting satellite at its closest approach.

perihelion **1.** The point at which a solar-orbiting satellite attains its lowest altitude. It occurs once for every complete orbit. At this point, the satellite travels faster than at any other point in the orbit. **2.** The altitude, measured from the sun's surface or the sun's center, of a solar-orbiting satellite at its closest approach.

perimeter protection The use of a security system to restrict or prevent access to a designated area, using sensors and/or barriers around the boundaries of the area.

period Symbol, *T.* Unit, second. The duration of a complete alternating-current cycle or of any cyclic event; $T = 1/f$, where f is the frequency in Hertz. Also see CYCLE, FREQUENCY, and HERTZ.

periodic and random deviation Abbreviation, PARD. In the direct-current output of a rectifier, the combined PERIODIC DEVIATION, including ripple, noise, hum, and transient spikes.

periodic curve A curve that repeats its shape in each period (e.g., a sine curve).

periodic deviation Repetitive deviation of a quantity from its normal value (e.g., ripple in the direct-current output of a rectifier).

periodic function A mathematical function that is represented by a periodic curve (e.g., the sine function $y = \sin x$).

periodicity In a transmission line, the tendency for power to be reflected at a point or points where the diameter of the line changes.

periodic law The observation that when the chemical elements (see ELEMENT, **3**) are arranged in increasing order of atomic number, their physical and chemical properties recur periodically. Also see PERIODIC TABLE.

periodic table A table in which the chemical elements (see ELEMENT, **3**) are arranged according to the periodic law. The vertical columns in the table, labeled groups, contain elements possessing related properties (e.g., silicon and germanium in group IV). The rows, labeled periods, depict the periodic shift in the properties of the elements.

peripheral **1.** Pertaining to equipment accessory to a central system (e.g., peripheral input/output devices online or offline to computers, data recorders, and indicators). Also see ANCILLARY EQUIPMENT. **2.** Peripheral equipment in a computer system (e.g., printers, modems, external disk drives, tape drives, etc.).

peripheral buffer As part of a peripheral in a computer system, a storage unit in which data temporarily resides on its way to or from the central processing unit. Also called INPUT/OUTPUT BUFFER.

peripheral electron See VALENCE ELECTRON.

peripheral equipment See PERIPHERAL, **1**, **2**.

peripheral interface adapter Abbreviation, PIA. An integrated circuit that acts as an input/output port to interface a microprocessor with peripheral devices.

peripheral transfer In a computer system, the transfer of a unit of data between peripherals, or between a peripheral and the central processing unit.

Permalloy A high-permeability alloy of iron and nickel.

permamagnetic speaker See PERMANENT-MAGNET SPEAKER.

permanent magnet A body that is always magnetized (i.e., without the application of electricity and without requiring the presence of another magnet). Compare TEMPORARY MAGNET.

permanent-magnet erase Erasure of magnetic tape by the field of a permanent magnet. Typically, it is a two-step process: a magnet erases what it can of the signal, leaving any residual magnetization for erasure by a second magnet.

permanent-magnet focusing In a cathode-ray tube, the focusing of the electron beam by means of permanent magnets.

permanent-magnet generator An electromechanical generator in which the field (either stationary or rotating) is provided by a multipole permanent magnet. Also called *magneto.*

permanent-magnet loudspeaker See PERMANENT-MAGNET SPEAKER.

permanent-magnet magnetizer A magnetizer using a permanent magnet as the magnetic-field source.

permanent-magnet meter An indicating meter in which a movable coil rotates between the poles of a permanent magnet. Compare ELECTRODYNAMOMETER and IRON-VANE METER.

permanent-magnet motor A motor having a permanent-magnet field.

permanent-magnet relay A polarized relay using a permanent magnet.

permanent-magnet speaker An acoustic loudspeaker in which the core is a strong permanent magnet (as opposed to a direct-current electromagnet). Also see MAGNETIC SPEAKER.

permanent storage See NONVOLATILE MEMORY.

permeability Unit, H/m. A quantitative indicator of the extent to which a material concentrates magnetic flux: for a given constant magnetic-field intensity, the ratio of magnetic flux density in the material to the magnetic flux density in air.

permeability curve See B-H CURVE.

permeability-tuned oscillator A radio-frequency oscillator in which the frequency is varied or adjusted by moving a ferromagnetic core in and out of the coil of an inductance-capacitance (LC) tuned circuit.

permeability tuning Variation of the resonant frequency of an inductance-capacitance LC circuit by changing the position of a magnetic core within the inductor. This type of tuning is used in amplifiers, oscillators, filters, and wavetraps.

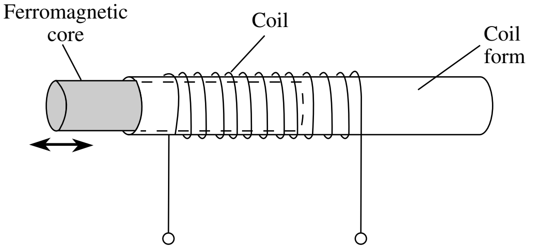

permeability tuning

permeameter An instrument for measuring permeability.

permeance Unit, Wb/A. In a magnetic circuit, the ease with which a magnetic field is established. The reciprocal of RELUCTANCE.

Permendur A high-permeability magnetic alloy containing equal parts of iron and cobalt. At saturation, the flux density of this material can be 2 teslas (20,000 gauss).

Perminvar A high-permeability magnetic alloy of cobalt, iron, and nickel. At saturation, the flux density of this material can approach 1.2 teslas (12,000 gauss).

permittivity See DIELECTRIC CONSTANT.

permutation A selection of several factors or objects from a group, in a specific ordered sequence. For example, one of the permutations of the elements of the set (1, 2, 3, 4, and 5) is the ordered sequence 4, 1, 3, 5, 2.

permutation modulation A method of modulation accomplished by varying the sequence of digital bits.

peroxide of lead In a lead-acid cell or battery, a compound of lead and oxygen that composes the positive electrode or electrodes.

persistence **1.** The effect whereby the retina of the eye continues to register a projected scene for approximately 0.05 second after the scene disappears. This allows perception of a sequence of video frames as a continuous moving image. **2.** The tendency of certain phosphors to glow after the excitation has been removed. Thus, after the electron beam in a cathode-ray tube has passed over the screen, the phosphor might continue to glow for a certain time along the path traced by the beam. Some phosphors, such as those used in high-speed oscilloscopes, have virtually no persistence, whereas others have long persistence.

persistent oscillations Successive oscillations of constant amplitude. Also called CONTINUOUS WAVE.

persistor A device used at low temperatures for temporary memory storage that operates between superconducting and normal conditions.

personal communications service Abbreviation, PCS. Also called *digital cellular.* An enhanced wireless network using digital modulation, cellular repeaters, and facilitating telephone and Internet connections. Emphasis is on maximizing user mobility and portability, and minimizing blind zones. Compare CELLULAR COMMUNICATIONS.

personal computer A small computer equipped with a keyboard, display, hard disk, diskette drive(s), a modem or fax/modem, one or more serial data ports, and one or more parallel data ports. They are used extensively by individuals and businesses for record keeping, data processing, communications, word processing, graphics, etc.; they are also used in schools as an educational aid.

personal digital assistant Abbreviation, PDA. See HANDHELD COMPUTER.

personal equation The value of systematic error for a person observing specific phenomena or making measurements.

personality Characteristics that make an intelligent computer or robot human-like. In general, the more powerful the computer, the more personality it can have, depending on the installed software. In some cases, certain malfunctions in a computer can produce personality quirks.

personal robot A usually autonomous robot intended for use by individuals. The most common examples are robot toys, programmable with a PERSONAL COMPUTER, intended for the education and entertainment of children. More sophisticated devices can perform domestic tasks, such as cleaning floors and mowing lawns.

peta- Abbreviation, P. A prefix meaning 10^{15}.

petagram Abbreviation, Pg. A large unit of mass or force, equal to 10^{15} grams or 10^{12} kilograms.

petameter Abbreviation, Pm. A large unit of (astronomical) distance, equal to 10^{15} meters or 10^{12} kilometers.

pF Abbreviation of PICOFARAD.

pf Symbol for POWER FACTOR.

PFET Abbreviation of P-CHANNEL JUNCTION FIELD-EFFECT TRANSISTOR.

PFM Abbreviation of PULSE-FREQUENCY MODULATION.

PG Abbreviation of POWER GAIN.

Pg Abbreviation of PETAGRAM.

pH **1.** Symbol for *hydrogen-ion concentration.* Numerically, pH is the negative logarithm of the effective hydrogen-ion concentration in gram equivalents per liter. The scale runs from zero to 14, on which 7 denotes neutrality relative to acidity vs. alkalinity; values between zero and 7 denote acidity, and values between 7 and 14 denote alkalinity. **2.** Abbreviation of PICOHENRY.

phantom Radio interference in the form of a beat note (heterodyne), resulting from interference between two strong carriers, often from local radio stations. When the phantom frequency lies within the tuning range of a receiver, the phantom can be tuned in as a separate signal. But when the phantom corresponds to the intermediate frequency (IF) of the receiver, it will ride into the IF amplifier and be present as an untunable interferential signal.

phantom channel In a properly phased high-fidelity stereo sound system, the apparent sound source centered between the left- and right-channel loudspeakers.

phantom circuit In wire telephony, a third circuit that has no wires; it results from a method (using repeating coils) of making two other circuits do the work of (this third) one.

phantom signal Also called *bogey.* In a radar system, a signal that does not correspond to an actual target. The origin of the phantom signal or echo cannot be readily determined.

phantom target See ECHO BOX.

phase angle Unit, degree or radian. In an alternating-current (ac) circuit, the lag or lead between the instant that one alternating quantity reaches its maximum value and the instant that another alternating quantity reaches its maximum value. It is usually given in degrees (a complete cycle being 360 degrees) along the horizontal axis of the time-versus-magnitude graph of the ac quantity.

phase-angle voltmeter An instrument that indicates both the magnitude and phase of a voltage.

phase coincidence For signals having the same frequency, the condition of their coinciding in terms of instantaneous amplitudes, so positive peaks of the first signal correspond to positive peaks of the second signal, and negative peaks of the first signal correspond to negative peaks of the second signal. For periodic waves that do not change their characteristics with time, this is the same thing as being shifted by an integral multiple of 360 degrees in phase. Compare PHASE OPPOSITION.

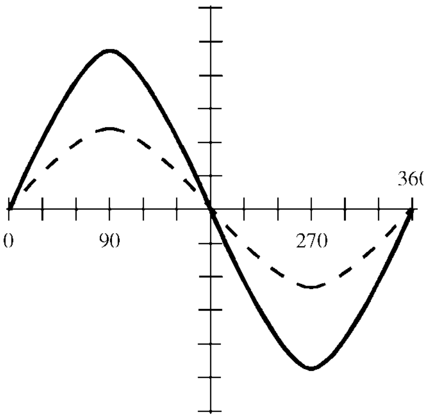

phase coincidence

phase compensation In an operational amplifier, compensation for excessive phase shift in the feedback.

phase compressor A push-pull phase-inverter circuit in which a capacitor is connected between each collector or drain and the opposite output terminal to attenuate in-phase components, such as even-numbered harmonics.

phase constant A figure providing the rate (in degrees of phase per unit length) at which the phase lag of the current or voltage field component in a traveling wave increases linearly in the propagation direction.

phase corrector A circuit that returns a signal to a certain phase after the signal has passed through a circuit or medium that has caused phase distortion.

phased antenna See PHASED ARRAY.

phased array Also called *phased antenna*. An antenna system having two or more driven elements fed with a certain relative phase, and spaced at a certain distance, resulting in a directivity pattern that exhibits gain in some directions and little or no radiation/response in other directions. Such an array can have two elements, producing a unidirectional cardioid or bidirectional figure-eight pattern. More complex arrays have several elements, usually vertical antennas, strategically positioned and fed with signals of specified phase to produce a highly tailored pattern. The most sophisticated systems have rotatable or steerable radiation/response patterns.

phase-delay equalizer See DELAY EQUALIZER.

phase detector See PHASE-SENSITIVE DETECTOR.

phase diagram A graphical representation of waves having equal frequency, but differing in phase. The phase difference for two identical waveforms is greater than or equal to zero degrees, but less than 360 degrees.

phase difference **1.** The difference (in time, angle, or fractional cycle) between the instants at which two alternating quantities reach a given value. **2.** For a dielectric, the complement of PHASE ANGLE; that is, 90 degrees minus the phase angle in degrees.

phase discriminator See DISCRIMINATOR, FOSTER-SEELEY DISCRIMINATOR, RATIO DETECTOR, and TRAVIS DISCRIMINATOR.

phase distortion Distortion characterized by input/output phase shift between various components of a signal passed by a circuit or device.

phase inverter A resistance-capacitance-coupled amplifier with a single-ended input and a push-pull output. This circuit enables a push-pull amplifier to be driven without an input transformer.

phase-locked loop Abbreviation, PLL. An oscillator that combines the flexibility of a conventional variable-frequency oscillator (VFO) with the stability of a crystal oscillator. The oscillator output is passed through a programmable divider that divides the frequency by a specific integral value n chosen by the operator. The output frequency of the divider is locked, by means of a phase comparator, to the signal from a crystal-controlled reference oscillator. As long as the output from the divider is exactly at the reference-oscillator frequency, the two signals are in phase, and the output of the phase comparator is zero volts dc. If the VCO frequency changes, the phase also changes, and the phase comparator produces a dc error voltage. The error voltage is applied to the VCO, causing the VCO frequency to correct itself. This maintains the VCO frequency at precisely n times the reference-oscillator frequency.

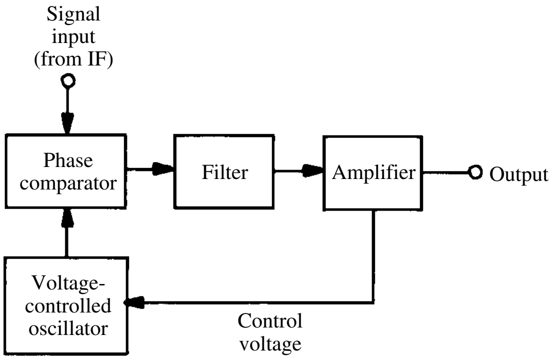

phase-locked loop

phase-locked oscillator An oscillator in which the inductance or the capacitance is varied periodically at half the driving frequency.

phase margin In an integrated-circuit amplifier, the extent to which the device shifts the phase of a signal more or less than one-half cycle (180 degrees) for a certain signal voltage.

phase modulation Abbreviation, PM. A method of modulation in which the phase of the carrier current is varied in accordance with the instantaneous modulating-signal voltage. The result is similar to FREQUENCY MODULATION.

phase modulator A circuit or stage that produces PHASE MODULATION.

phase multiplier A circuit used for the purpose of phase comparison between signals. The frequency of the measured signal is multiplied, resulting in multiplication of the phase difference. This improves the sensitivity of the measuring apparatus.

phase opposition For signals having the same frequency, the condition of their being inverted relative to each other in terms of instantaneous amplitudes, so positive peaks of the first signal correspond to negative peaks of the second signal, and negative peaks of the first signal correspond to positive peaks of the second signal. This is not the same thing as being shifted by an odd

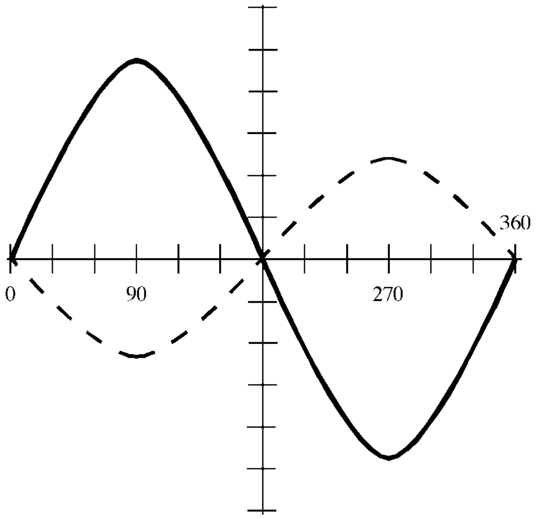

phase opposition

integral multiple of 180 degrees in phase, although in practice, with sine waves and square waves, the effect is the same. Compare PHASE COINCIDENCE.

phase resonance See VELOCITY RESONANCE.

phase reversal **1.** The inversion of an alternating-current (ac) signal. The instantaneous amplitude (current or voltage) is multiplied by a negative constant. Thus, the positive half-cycles become negative, and the negative half-cycles become positive. **2.** A phase shift of ±180 degrees (±½ cycle) in an ac signal.

phase-rotation relay See PHASE-SEQUENCE RELAY.

phase-rotation system A system for producing single-sideband signals without using selective filters. In one such system, two balanced modulators are used. One of these receives carrier and modulating voltages that are 90 degrees out of phase with voltages that are fed to the other balanced modulator.

phase-sensitive detector Abbreviation, PSD. A detector for frequency modulation (FM) and phase modulation (PM). It delivers a direct-current output voltage whose value is proportional to the difference in phase between a reference signal and the signal from a local oscillator.

phase-sequence relay In a polyphase system, a relay or relay circuit that is actuated by voltages reaching maximum positive amplitude in a predetermined phase sequence. Also called PHASE-ROTATION RELAY.

phase shift **1.** A change in the displacement, as a function of time, of a periodic disturbance having constant frequency. **2.** The magnitude of a change, as defined in **1**, measured in fractions of a wavelength or in electrical degrees.

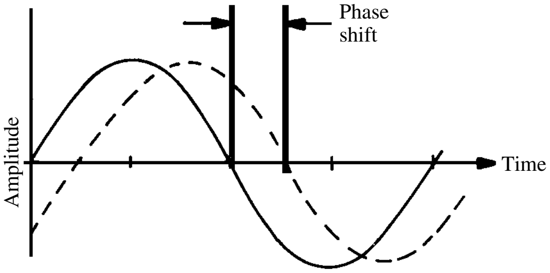

phase shift

phase-shift bridge A four-arm-bridge circuit for shifting the phase of an alternating-current signal. Such a circuit is often used (with one arm variable) to shift the phase of the firing voltage for a thyratron.

phase-shift discriminator See FOSTER-SEELEY DISCRIMINATOR.

phase shifter A circuit, such as an inductance-capacitance (LC) or resistance-capacitance (RC) network, or a device, such as a Helmholtz coil or phase-shifting capacitor, that introduces a phase shift between input and output signals.

phase-shifting capacitor A special four-stator, one-rotor variable capacitor that, with a transformer-coupled resistance-capacitance (RC) circuit, provides 360 degrees of continuously variable phase shift for one rotation of the rotor. The rotor plate turns like a cam under the stators because of the off-center insertion of the rotor shaft.

phase-shift oscillator A single-stage oscillator in which the required 180-degree phase shift in the signal (fed back from output to input) is obtained by passing the output through a phase-shifting network.

phase-shift-type distortion meter A distortion meter in which the output signal of a device under test is compared with a distortion-free input test signal. The output signal phase is shifted 180 degrees, with respect to the input, and the two amplitudes are made equal. If there is no distortion, the signals cancel each other, and the result is zero. Any remaining signal is proportional to the total harmonic distortion (THD).

phase-splitting circuit A circuit that produces, from a single input signal, two output signals differing in phase.

phase-splitting driver A PHASE INVERTER used as the driver of a push-pull amplifier.

phase velocity The velocity of a wave, provided by the product of the frequency and the wavelength.

phase windings In an alternating-current generator, windings that deliver voltages that differ in phase.

phasing capacitor In a crystal filter, a small variable capacitor that constitutes one arm of a four-arm bridge in which the crystal is another arm. Adjustment of this capacitor balances the bridge, thus preventing the undesirable passage of a signal through the capacitance of the crystal holder.

phenol-formaldehyde plastics A family of plastic insulating materials made with phenolic resin, and occasionally used as dielectrics and air-core coil forms. Some of the trade names for these materials include *Bakelite*, *Catalin*, *Durez*, *Durite*, *Formica*, and *Micarta*.

phenolic insulants See PHENOL-FORMALDEHYDE PLASTICS.

phenolic resin A synthetic resin made by condensing phenol (carbolic acid) with formaldehyde.

phenomenon An event or circumstance that can be verified by the senses, as opposed to one subject to theory or speculation (e.g., the phenomenon of magnetic attraction).

Phillips gate A device that allows measurement of the gas pressure in a confined chamber. A current is passed through the gas. The magnitude of the current, for a given gas, is a function of the gas pressure and temperature.

Phillips screw A screw with a pair of slots in its head. The slots are arranged like an x. Phillips screws are available in many different sizes, as are ordinary screws. The x-shaped pair of slots reduces the tendency for the screwdriver to slip out of the screw head as the screw is rotated.

Phi phenomenon The illusion of motion resulting from the rapid presentation to the eye of pictures showing objects in a succession of different positions. Television and motion pictures exploit this illusion. Also see PERSISTENCE.

pH meter An instrument used to measure the acidity or alkalinity of solutions. Also see PH, **1**.

phon A unit of apparent change in loudness discerned by a listener. Unlike the decibel, the phon includes compensation for the ear's nonlinear response to attendant frequency changes. At a frequency of 1 kHz, a change in loudness of 1 phon is the equivalent of 1 decibel.

phone **1.** Telephone (wire or radio). **2.** To establish communication via telephone. **3.** Colloquialism for voice communication (radiotelephone), particularly via amateur-radio single sideband on the high-frequency bands (160 through 10 meters). **4.** A minimal, unique speech sound. Also called SOUND UNIT.

phoneme An individual sound or syllable in the human voice, with a characteristic amplitude-vs. frequency spectral pattern. It is important in speech recognition and speech synthesis. Computers can be programmed to identify and transcribe these sounds; computers can also be programmed to generate the sounds from text data.

phone jack The female mating device for a PHONE PLUG.

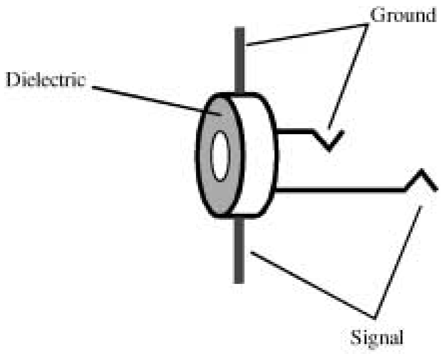

phone jack

phone monitor A simple device for listening to amplitude-modulated radio transmissions to test their quality. In its most rudimentary form, it consists of a pickup antenna, semiconductor-diode detector, and high-resistance headphones.

phone patch A device for establishing a connection (patch) between radio and wire-telephone facilities. Also see PATCH.

phone plug A type of plug originally designed for patching telephone circuits, now widely used in electronics and instrumentation. In its conventional form, it has a rod-shaped neck that serves as one contact, and a ball on the tip of the neck, but insulated from it, that serves as the other contact. Typical diameters are ⅛ inch and ¼ inch.

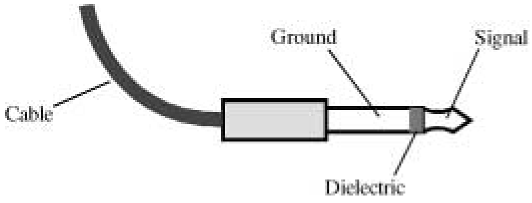

phone plug

phone test set An instrument for checking the performance of a radiotelephone transmitter. The set combines the functions of field-strength meter, modulation indicator, and aural monitor. Sometimes it includes a volt-ohm-milliammeter for troubleshooting the transmitter.

phonetic alphabet Words whose initial letters are used to identify the letters of the alphabet for which they stand. These words are spoken in radiotelephony to identify letters that, if spoken by themselves, might not be clearly heard.

Phonetic alphabet

Letter	Phonetic (Capitals indicate emphasis)
A	AL-fa
B	BRAH-vo
C	CHAR-lie
D	DEL-ta
E	ECK-o
F	FOX-trot
G	GOLF
H	ho-TEL
I	IN-dia
J	Ju-li-ETTE
K	KEE-low
L	LEE-ma
M	Mike
N	No-VEM-ber
O	OS-car
P	pa-PA
Q	Que-BECK
R	ROW-me-oh
S	see-AIR-ah
T	TANG-go
U	YOU-ni-form
V	VIC-tor
W	WHIS-key
X	X-ray
Y	YANK-key
Z	ZOO-loo

phonetic alphabet code word In radio and wire telephony, a word chosen for its easy recognition by ear to identify the letter of the alphabet with which it begins. For example: *Golf* for G, *Juliet* for J, *X-ray* for X.

phonics See ACOUSTICS, **1**.

phonocardiogram The record made by a PHONOCARDIOGRAPH.

phonocardiograph An instrument that makes a graphic record of heart sounds.

phono cartridge The vibration-to-electricity transducer (pickup) of a phonograph; it is actuated by the stylus (needle). Common types are ceramic, variable-inductance, and variable-reluctance. See PHONOGRAPH and PHONOGRAPH DISC.

phonocatheter A microphone that can be inserted into the body for the purpose of listening to the functions of internal organs.

phonograph A device for reproducing sound recorded on disc. It consists of a turntable, an amplifier, and one or more speakers.

phonograph cartridge See PHONO CARTRIDGE.

phonograph disc A thin, lightweight disc, usually made of vinyl or similar plastic, on which audio-frequency signals are recorded as irregularities in a spiral groove. In reproduction, these irregularities cause vibration in a PHONO CARTRIDGE as the turntable rotates.

phonograph oscillator See PHONO OSCILLATOR.

phono jack Also called *RCA jack.* A jack similar to a PHONE JACK, designed especially for the quick connection and disconnection of coaxial cables used with audio and low-frequency devices.

phonon A unit of energy resulting from vibration, as of a piezoelectric crystal.

phono oscillator A small radio-frequency (RF) oscillator modulated by the audio-frequency (AF) voltage from a phonograph. The modulated RF signal is picked up by a remote radio receiver (usually in the same room), and the sound is reproduced through a loudspeaker connected to the receiver.

phono plug Also called *RCA plug.* A plug similar to a PHONE PLUG, designed especially for the quick connection and disconnection of coaxial cables used with audio and low-frequency devices.

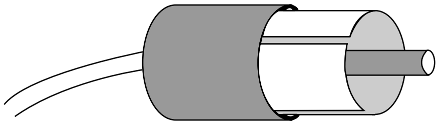

phono plug

phonoreception The hearing of high-frequency sounds.

phonorecord A PHONOGRAPH disc.

phonoselectroscope A special type of stethoscope, in which the main heartbeat is attenuated. This makes abnormal sounds more audible. The device can be adjusted in various ways to listen for abnormalities characteristic of various heart diseases.

phosphor A substance that glows when an electron beam strikes it. Such a substance is used as a coating on the screens of cathode-ray tubes. See also BEAT ZINC SILICATE; CADMIUM BORATE, SILICATE, and TUNGSTATE; CALCIUM PHOSPHATE, SILICATE, and TUNGSTATE; MAGNESIUM FLUORIDE, SILICATE, and TUNGSTATE; ZINC ALUMINATE; ZINC BERYLLIUM SILICATE; ZINC BERYLLIUM ZIRCONIUM SILICATE; ZINC BORATE; ZINC CADMIUM SULFIDE; ZINC GERMANATE; ZINC MAGNESIUM FLUORIDE; ZINC ORTHOSILICATE; ZINC OXIDE; ZINC SILICATE; and ZINC SULFIDE.

phosphor bronze A form of bronze whose elasticity, hardness, and toughness have been greatly improved by the addition of phosphorus. The metal is used for brushes, springs, switch blades, and contacts.

phosphor copper An alloy of copper and phosphorus used in the manufacture of PHOSPHOR BRONZE.

phosphorescence The property of some materials that ordinarily fluoresce to continue to glow after the stimulus (light or an electron beam) has been removed. Compare FLUORESCENCE.

phosphorescent screen A viewing screen coated with a phosphor (e.g., oscilloscope screen).

phosphorous Exhibiting the properties of phosphor (e.g., glowing after stimulation with light). Not to be confused with PHOSPHORUS.

phosphorus Symbol, P. A nonmetallic element of the nitrogen family. Atomic number, 15. Atomic weight, 30.974. It is used as a dopant in semiconductor processing.

phot The cgs unit of illumination: The direct illumination produced upon a one-centimeter-distant surface by a uniform point source of one international foot-candle. Equivalent to one lumen per square centimeter.

photocathode **1.** The photomosaic of a video camera tube. **2.** The light-sensitive cathode in a phototube.

PhotoCD Trade name for an image-recording system developed by Kodak, in which photographs can be stored on compact discs. Viewing is accomplished using personal computers.

photocell See PHOTOELECTRIC CELL.

photocell amplifier An amplifier used to boost the output of a photocell. With respect to the nature of the input signal, it can be an alternating-current (ac) or direct-current (dc) amplifier, depending on whether the output of the photocell is modulated dc or pure dc.

photochemical effect The phenomenon whereby certain substances undergo chemical change when exposed to light or other radiant energy. An example of such a substance is the silver bromide, silver chloride, or silver iodide on photographic film.

photoconductive cell A photoelectric cell, such as the cadmium-sulfide type, whose resistance is proportional to the intensity of light impinging upon it. The photoconductive cell acts as a light-sensitive variable resistor in a current path. Also see PHOTOCONDUCTIVE MATERIAL.

photoconductive effect The tendency for the electrical resistance of a substance to change when infrared radiation, visible light, or ultraviolet radiation strikes it. Different substances exhibit different degrees of this effect.

photoconductive material A substance that exhibits decreased electrical resistance when exposed to infrared rays, visible light, or ultraviolet. Some photoconductive substances are cadmium selenide, cadmium sulfide, germanium, lead sulfide, selenium, silicon, and thallous sulfide. Also see ACTINOELECTRIC EFFECT.

photoconductivity The phenomenon whereby the electrical resistance of certain materials (such as cadmium sulfide, cadmium selenide, germanium, selenium, and silicon) is lowered upon exposure to infrared rays, visible light, or ultraviolet. Also see PHOTOCONDUCTIVE MATERIAL.

photoconductor **1.** See PHOTOCONDUCTIVE MATERIAL. **2.** See PHOTOCONDUCTIVE CELL.

photocurrent See PHOTOELECTRIC CURRENT.

photo-Darlington Also, *photodarlington.* **1.** A phototransistor fabricated as a Darlington amplifier for high output current. **2.** A combination of photodiode (see LIGHT-SENSITIVE DIODE) and Darlington amplifier.

photodecomposition Chemical breakdown by the action of radiant energy. Also called *photolysis.*

photodetachment The removal of an electron from an atom or ion, resulting from the impact of a PHOTON.

photodetector **1.** An illumination meter that uses a PHOTOCELL. **2.** See OPTOELECTRONIC COUPLER.

photodielectric effect The tendency for the dielectric constant of a substance to change when infrared radiation, visible light, or ultraviolet radiation strikes it. Different substances exhibit different degrees of this effect.

photodiffusion effect See DEMBER EFFECT.

photodiode See LIGHT-SENSITIVE DIODE.

photodisintegration In the nucleus of an atom, disintegration resulting from PHOTON bombardment.

photoelasticity The tendency for the light-transmission characteristics of a substance to change with externally applied forces.

photoelectric alarm An alarm actuated when a light beam impinging on a photocell is interrupted.

photoelectric amplifier **1.** An amplifier for boosting the output of a photosensitive device. **2.** An OPTOELECTRONIC COUPLER possessing gain.

photoelectric cell A device that converts infrared, visible-light, or ultraviolet energy into electricity or electrical effects. It can function by producing a voltage (see PHOTOVOLTAIC CELL, SELENIUM CELL, SILICON CELL, SOLAR CELL, and SUN BATTERY) or by acting as a light-sensitive resistor (see LIGHT-SENSITIVE DIODE, PHOTOCONDUCTIVE CELL, and SELENIUM CELL).

photoelectric constant The quantity h/e, where h is Planck's constant and e is the unit electron charge.

photoelectric counter A counting device (electromechanical or fully electronic) that counts objects as they interrupt a light beam impinging upon a photocell.

photoelectric disintegration See PHOTODISINTEGRATION.

photoelectric effect The phenomenon whereby temporary changes occur in the atoms of certain substances under the influence of infrared, visible light, or ultraviolet radiation. Some of these materials undergo a change in their electrical resistance, whereas others generate electric current (see, for comparison, PHOTOCONDUCTIVE MATERIAL and PHOTOVOLTAIC MATERIAL).

photoelectric efficiency See QUANTUM YIELD.

photoelectric field-effect transistor See PHOTOFET.

photoelectricity Electricity produced by the action of light on certain materials, such as cesium, selenium, and silicon. Also see PHOTOEMISSION and PHOTOVOLTAIC CELL.

photoelectric material See PHOTOCONDUCTIVE MATERIAL and PHOTOEMISSIVE MATERIAL.

photoelectric multiplier A device that internally amplifies the current resulting from bombardment by infrared, visible light, or ultraviolet radiation. A PHOTOMULTIPLIER TUBE is an example of such a device.

photoelectric photometer An instrument that uses a photoelectric device for the purpose of measuring the intensity of infrared radiation, visible light, or ultraviolet radiation.

photoelectric proximity sensor A device that uses a light-beam generator, a photodetector, an amplifier, and a microprocessor to detect the presence of nearby objects. It is useful in robot guidance systems.

photoelectric pyrometer An optical pyrometer in which a photocell and appropriate filters act instead of the human eye.

photoelectric relay A relay actuated directly by a photocell or a photocell and amplifier. This type of relay is the basis of some PHOTOELECTRIC ALARM devices.

photoelectric sensor **1.** See ELECTRIC EYE. **2.** See PHOTOELECTRIC PROXIMITY SENSOR. **3.** See PHOTOELECTRIC CELL.

photoelectric smoke alarm An alarm that is tripped by a PHOTOELECTRIC SMOKE DETECTOR when the density of smoke exceeds a safe level.

photoelectric smoke control A system for making automatic adjustments to a burning process when the smoke density exceeds a prescribed level. The initial element in the system is a PHOTOELECTRIC SMOKE DETECTOR.

photoelectric smoke detector A smoke detector in which a photocell, photodiode, phototransistor, or phototube is excited by a light beam passing through the air. The cell output current decreases when smoke fills the air. This current change trips an alarm or deflects an indicating meter when the density of the smoke exceeds a prescribed level.

photoelectric tape reader A punched-tape reader using a photocell, photodiode, phototransistor, or phototube to sense light passing through the holes.

photoelectric transducer A photocell, photodiode, phototransistor, or phototube used as a sensor.

photoelectric tube See PHOTOTUBE.

photoelectric wattmeter A power-measuring instrument useful for the approximate measurement of radio-frequency power. It consists of an incandescent lamp sharing an opaque enclosure with a photovoltaic cell. The power to be measured is applied to the lamp, which glows proportionately. The light excites the cell, causing it to deliver a direct current proportional to the power. This current deflects a milliammeter or microammeter. The meter can be calibrated to read directly in watts.

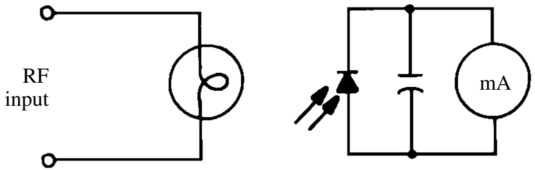

photoelectric wattmeter

photoelectromotive force The electromotive force (voltage) produced by a photovoltaic cell.

photoelectron An electron displaced within, or ejected from, an atom, as the result of infrared, visible light, or ultraviolet radiation striking the atom.

photo-emf See PHOTOELECTROMOTIVE FORCE.

photoemission The ejection of electrons from certain materials, such as cesium, when these materials are exposed to infrared, visible light, or ultraviolet radiation. Also see PHOTOEMISSIVE MATERIAL.

photoemissive material A substance that emits electrons when exposed to infrared, visible light, or ultraviolet radiation. A typical use of such a material is in the coating of the light-sensitive cathode of a phototube. The metals cesium, potassium, rubidium, and sodium are photoemissive.

photofabrication **1.** A method of circuit-board manufacturing. The etching pattern is placed over the circuit-board material, the board is placed in a special solution, then the assembly is exposed to visible light. The light interacts with the solution to dissolve the metal in areas exposed to the light, but not in areas covered by the etching pattern. **2.** The technique in **1**, applied to the manufacture of integrated circuits.

photoFET A FIELD-EFFECT TRANSISTOR that exhibits properties similar to those of a bipolar PHOTOTRANSISTOR.

photoflash See ELECTRONIC FLASH, **1**.

photoglow tube See DISCHARGE LAMP.

photogram The permanent shadow produced by an object placed between a light source and photographic paper.

photographic exposure meter See EXPOSURE METER, **1**.

photographic recorder A graphic recorder that uses a light beam, deflected by galvanometer movement, that moves across photographic film or paper to produce a trace representing a varying quantity.

photographic sound recording See OPTICAL SOUND RECORDING.

photograph reception **1.** The use of FACSIMILE to print photographs transmitted in analog form via wire or radio. **2.** The use of a computer, equipped with a modem and graphics software, to display and/or store photographs transmitted in digital form via wire or radio.

photograph transmission **1.** The use of FACSIMILE to scan and send photographs in analog form via wire or radio. **2.** The use of a computer, equipped with a modem and video camera or optical scanner, to digitize and send photographs via wire or radio.

photoionization The ejection of electrons from atoms or molecules by the action of infrared, visible light, or ultraviolet radiation.

photoisolator See OPTOELECTRONIC COUPLER.

photojunction cell A photocell consisting of a semiconductor pn junction. The cell is useful mainly for its photoconductivity, although infrared, visible light, or ultraviolet energy striking the junction produces a small amount of photovoltaic action.

photokinesis Light-induced motion, as in a RADIOMETER.

photolithographic process A method of producing integrated circuits and printed circuits by photographing (often at considerable reduction) an enlarged pattern of the circuit on a suitable light-sensitized surface of metal or semiconductor, and chemically etching away unwanted portions of the surface.

photolysis See PHOTODECOMPOSITION.

photomagnetic effect Light-sensitive magnetic susceptibility in some materials.

photomap A photo taken of terrain from a high altitude and usually overlaid with a reference grid.

photomask In PHOTOFABRICATION, the transparent film or template on which the etching pattern is drawn.

photometer An instrument used to compare the luminous intensity of two light sources.

photometric measurement of power See PHOTOELECTRIC WATTMETER.

photometry The science of visible-light measurement. The response of the human eye is used as the basis for preferred sensors (those used with photometric instruments, which have spectral sensitivity curves resembling those of the eye). Compare RADIOMETRY.

photomosaic In a television camera tube, the flat photocathode screen on which the image is projected by the lens system and scanning electron beam. The surface of the screen is covered with tiny light-sensitive droplets. Also see DISSECTOR TUBE, ICONOSCOPE, and ORTHICON.

photomultiplier tube A type of PHOTOTUBE that delivers high output current for a given light intensity by utilizing the secondary emission of electrons. The initial light-sensitive cathode emits electrons; these strike a specially placed metal plate with a force that dislodges more electrons. These electrons, together with the initial emission, are reflected to a second plate, where they dislodge still more electrons. This process continues from deflection plate to deflection plate through the tube. The final plate deflects the accumulated electrons to the anode (collector electrode).

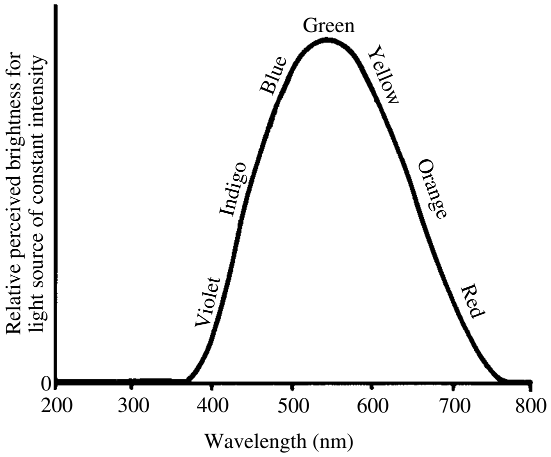

photometry

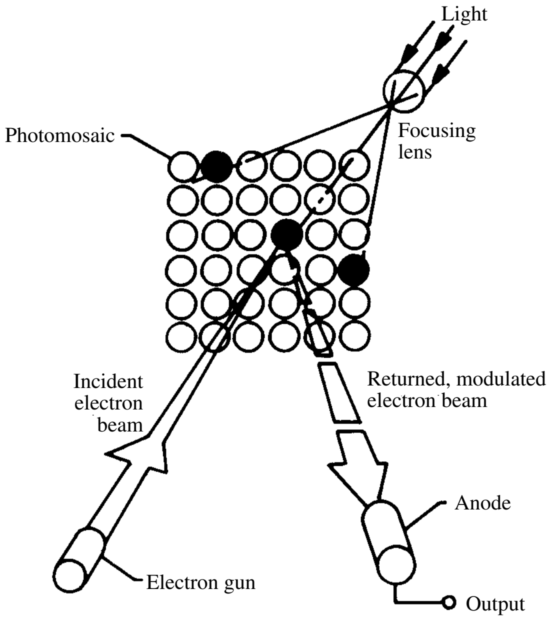

photomosaic

photon A quantum of radiant energy whose energy constant *W* (in joules) is equal to *hf*, where *h* is the PLANCK CONSTANT and *f* is the frequency in Hertz.

photoneutron A neutron released by PHOTODISINTEGRATION.

photophone **1.** A telephone-type communication system using a modulated light beam transmitted between stations. **2.** A process for recording sound on motion-picture film (see OPTICAL SOUND RECORDING).

photorelay See PHOTOELECTRIC RELAY.

photoresistive cell See PHOTOCONDUCTIVE CELL.

photoresistive material See PHOTOCONDUCTIVE MATERIAL.

photoresistivity See PHOTOCONDUCTIVITY.

photoresistor See PHOTOCONDUCTOR, **1**, **2**.

photosensitive device A light-sensitive electronic device. See, for example, PHOTOCONDUCTIVE CELL, PHOTODIODE, PHOTOFET, PHOTOMULTIPLIER TUBE, PHOTOTRANSISTOR, PHOTOTUBE, and PHOTOVOLTAIC CELL.

photosphere The luminous layer at the surface of a star.

photoswitch A light-activated switch. Some photoswitches contain an electromechanical relay; others, such as the light-activated silicon-controlled switch, have no moving parts.

phototimer An electronic timer for timing photographic processes.

phototransistor A transistor in which current carriers emitted as a result of illumination constitute an input-signal current. This current is amplified by the transistor. The output signal delivered by the transistor, accordingly, is larger than the output of an equivalent photodiode.

phototube An electron tube that converts light energy into electrical energy by acting as a light-sensitive resistor. Characteristically, the tube contains an illuminated cathode coated with a photoemissive material, and an anode wire situated nearby. Light energy causes electrons to be emitted from the cathode in amounts proportional to light intensity; the electrons are attracted by the anode, which is connected externally to a positive direct-current voltage.

photovoltaic cell Also called *solar cell.* A semiconductor diode, usually made from silicon, that converts visible light, infrared, and/or ultraviolet directly into electric current. The device consists of a flat P-N junction; the assembly is transparent so that radiant energy can fall directly on the P-type silicon. Metal ribbing, forming the positive electrode, is interconnected with tiny wires. The negative electrode is a metal backing, placed in contact with the N-type material. The component produces about 0.5 to 0.6 volts in direct sunlight under no-load conditions.

photovoltaic material A substance that generates a voltage when exposed to light. The principal substances exhibiting this effect are silicon, selenium, and germanium. Also see ACTINOELECTRIC EFFECT.

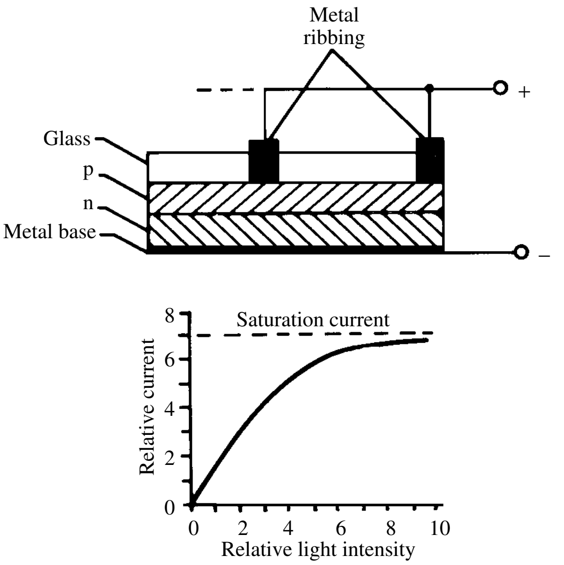

photovoltaic cell

photran A light-sensitive, four-layer semiconductor device, used for switching purposes.

physical properties The distinguishing characteristics of matter, apart from its chemical properties. Included are *boiling point*, *density*, *ductility*, *elasticity*, *electrical conductivity*, *hardness*, *heat conductivity*, *index of refraction*, *malleability*, *melting point*, *specific heat*, and *state* (solid, liquid, gaseous, or plasma).

physical quantity A quantity expressing the actual number of physical units under consideration, as compared with a dimensionless number. Examples: 50 volts, 39 kilometers, and 30 picofarads. Compare DIMENSIONLESS QUANTITY.

physics The science of energy and matter and their interactions. Physics is subdivided into several fields, including *mechanics*, *thermodynamics*, *acoustics*, *optics*, and *electricity/magnetism.* Many subdivisions are within the traditional fields.

P_i Symbol for INPUT POWER.

picket fencing An effect often observed at very-high frequencies (VHF) and ultra-high frequencies (UHF), in which movement of the transmitting station antenna, the receiving station antenna, or both antennas causes rapid fading. The fading is the result of phase effects between the direct wave and indirect wave(s). These effects are most pronounced with vertically polarized antennas.

PIA Abbreviation of PERIPHERAL INTERFACE ADAPTER.

pickoff **1.** To monitor a voltage, current, or other characteristic in an active circuit, without disturbing the operation of the circuit. **2.** A device for electronically monitoring linear or angular displacement.

pickup **1.** A device that serves as a sensor of a signal or quantity. This covers a wide variety of items, including *temperature sensors*, *vibration detectors*, *microphones*, *phonograph pickups*, etc. **2.** Collectively, energy or information that is received (e.g., *sound pickup*).

pickup arm The pivoted arm that holds the cartridge and stylus of a phonograph.

pickup cartridge See PHONO CARTRIDGE.

pickup current **1.** The current required to close a relay. **2.** Current flowing through, or generated by, a pickup.

pickup pattern The directional pattern of a microphone or other transducer that converts acoustic energy into electrical signals.

pickup voltage **1.** The voltage required to close a relay or circuit breaker. **2.** The voltage delivered by a pickup.

pico- **1.** Abbreviation, p. A prefix meaning 10^{-12}. **2.** A prefix meaning extremely small.

picoammeter A usually direct-reading instrument used to measure current in the picoampere range. Also see CURRENT METER.

picoampere Abbreviation, pA. A small unit of current equal to 10^{-12} ampere.

picocoulomb Abbreviation, pC. A small unit of electrical quantity equal to 10^{-12} coulomb.

picocurie Abbreviation, pCi. A small unit of radioactivity equal to 10^{-12} curie.

picofarad Abbreviation, pF. A small unit of capacitance equal to 10^{-12} farad.

picohenry Abbreviation, pH. A small unit of inductance equal to 10^{-12} henry.

picosecond Abbreviation, ps or psec. A small unit of time equal to 10^{-12} second.

pi coupler See COLLINS COUPLER.

picovolt Abbreviation, pV. A small unit of voltage equal to 10^{-12} volt.

picovoltmeter A usually direct-reading electronic instrument used to measure electromotive force in the picovolt range.

picowatt Abbreviation, pW. A small unit of power equal to 10^{-12} watt.

pictorial See PICTORIAL WIRING DIAGRAM.

pictorial diagram See PICTORIAL WIRING DIAGRAM.

pictorial wiring diagram A wiring diagram in the form of a drawing or photograph of the components, as opposed to one of circuit symbols. The components are shown in their positions in the finished equipment, and the wiring as lines running between them.

picture black In facsimile or television, the signal condition resulting from the scanning of a black portion of the image.

picture detector See VIDEO DETECTOR.

picture diagram See PICTORIAL WIRING DIAGRAM.

picture element See PIXEL.

picture information In a television signal, the variable-amplitude component (i.e., the one carrying energy corresponding to the picture elements) that fills the space between blanking pulses.

picture-in-picture Abbreviation, PIP. In some television receivers, a feature that allows simultaneous viewing of two programs. One program occupies the full screen, and another program appears in a small portion of the screen.

picture reception **1.** See PHOTOGRAPH RECEPTION. **2.** The reception of television signals.

picture transmission **1.** See PHOTOGRAPH TRANSMISSION. **2.** The transmission or broadcasting of television signals.

picture tube The cathode-ray tube used in a television receiver to display the image. Also called KINESCOPE.

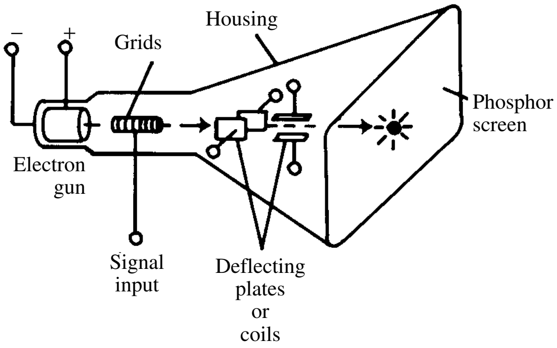

picture tube

pie chart See CIRCLE GRAPH.

Pierce oscillator A simple crystal oscillator in which the crystal is connected directly between the input and output terminals of the active device (usually a bipolar or field-effect transistor). A tuned inductance-capacitance (LC) circuit might be included, but is not required.

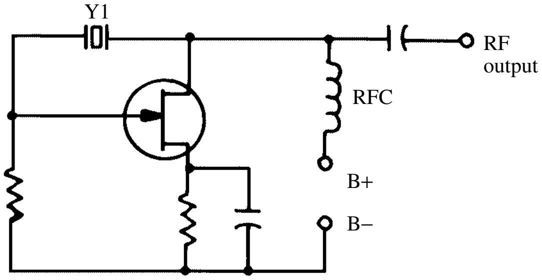

Pierce oscillator

pie winding A method of coil winding in which two or more separate, multilayer coils are connected in series and placed along a common axis. It is sometimes used in radio-frequency chokes to minimize capacitance among the windings.

piezo- A prefix meaning *pressure* (see PRESSURE, **2**).

piezodielectric A substance that, when stretched or compressed, exhibits a change in dielectric constant.

piezoelectric accelerometer An accelerometer using a piezoelectric crystal, whose voltage output is proportional to acceleration.

piezoelectric ceramic A ceramic material that delivers a voltage when deformed, or that changes in shape when a voltage is applied to it.

piezoelectric crystal A crystal (such as quartz, Rochelle salt, tourmaline, or various synthetics) that delivers a voltage when mechanical force is applied between its faces, or that changes its shape when a voltage is applied between its faces.

piezoelectric earphone See CRYSTAL EARPHONE.

piezoelectric filter See CRYSTAL FILTER and CRYSTAL RESONATOR.

piezoelectricity Electricity produced by deforming (squeezing, stretching, bending, or twisting) certain crystals, such as those of quartz, Rochelle salt, or tourmaline.

piezoelectric loudspeaker See CRYSTAL LOUDSPEAKER.

piezoelectric microphone See CERAMIC MICROPHONE and CRYSTAL MICROPHONE.

piezoelectric oscillator See CRYSTAL OSCILLATOR.

piezoelectric pickup See CRYSTAL PICKUP.

piezoelectric resonator See CRYSTAL FILTER and CRYSTAL RESONATOR.

piezoelectric sensor See CRYSTAL TRANSDUCER.

piezoelectric transducer See CRYSTAL TRANSDUCER.

piezoid A complete piezoelectric crystal device.

piezoresistance In certain substances, the tendency of the resistance to change with stretching or compression.

piezo tweeter A tweeter of the piezoelectric type (see CRYSTAL LOUDSPEAKER).

pi filter An unbalanced filter section having one series arm and two shunt arms; its schematic representation has the general shape of the uppercase Greek letter pi.

piggyback component See OUTBOARD COMPONENT.

piggyback control See CASCADE CONTROL.

piggyback tuner A separate ultra-high-frequency (UHF) television tuner operated in conjunction with the very-high-frequency (VHF) tuner of the receiver.

pigtail **1.** A usually long and sometimes flexible lead, such as the pigtail of a fixed capacitor. **2.** Descriptive of a device containing a long lead or leads, and usually mounted by such leads.

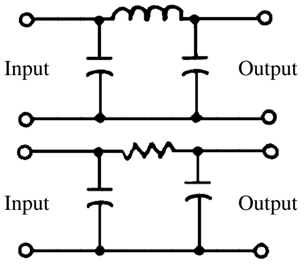

pi filters

pile **1.** See VOLTAIC PILE. **2.** See NUCLEAR REACTOR. **3.** A battery of electrochemical cells. **4.** Any packed group of particles or granules.

pillow speaker A small, flat loudspeaker intended for use under a pillow.

PILOT Acronym for *programmed inquiry learning or teaching.* A straightforward high-level computer programming language, used in computer-assisted instruction (CAI).

pilot lamp See PILOT LIGHT.

pilot light A usually small, incandescent or neon lamp. When glowing, it serves as a signal that a piece of equipment is in operation.

pilot model A preliminary model of a circuit or device constructed primarily to test the efficacy of a production process. The pilot model usually follows the PROTOTYPE.

pilot production The often small-scale production of a device in a special assembly line apart from the main line in a factory.

pilot regulator A variable-gain circuit that maintains a constant output—even if the input amplitude changes.

PIM Abbreviation of PULSE-INTERVAL MODULATION.

pi mode In a vane-anode magnetron, the mode of operation in which adjacent vanes have radio-frequency voltages of opposite polarity.

pin **1.** A semiconductor junction consisting of a layer of instrinsic semiconductor material situated between n and p layers. **2.** A slender, straight, stiff prong used as a terminal or locking device (see, for example, BASE PIN and BAYONET BASE).

pinchoff In a junction field-effect transistor, the condition in which the gate voltage causes the two depletion regions to meet and close the channel to obstruct drain-current flow.

pinchoff voltage In a junction field-effect transistor, the lowest value of gate voltage that will produce pinchoff.

pinch roller In a tape recorder, a rubber-tired, rotating cylinder that helps to pull the tape past the recording and/or playback heads.

pincushion A type of television picture distortion in which each side of the raster sags toward center screen. Also see ANTIPINCUSHIONING MAGNETS.

pincushion-correction generator A circuit for generating a deflection signal to correct pincushion distortion (see PINCUSHION). One form consists of a parabola generator and op-amp-type differentiator.

pin diode A silicon junction diode having a lightly doped intrinsic layer serving as a dielectric barrier between p and n layers.

pi network See COLLINS COUPLER.

ping An acoustic pulse; it can be audible sound or ultrasound.

pinhole **1.** A tiny hole present as a defect in a film of dielectric, semiconductor, or metal. **2.** A tiny aperture that acts as a universal lens by permitting the passage of a very small bundle of light rays. The smaller the aperture, the greater the depth of field.

pinhole detector An electronic device for finding pinholes in materials. Also see PINHOLE, **1**.

pin jack A jack into which a pin plug is inserted for quick connection.

pink noise Acoustical noise whose amplitude is inversely proportional to the frequency within a limited frequency spectrum. In the extreme, it creates a hissing sound. Compare WHITE NOISE.

pinout A diagram of an integrated circuit depicting the locations of the pins for various functions. It generally takes the form of a rectangle for the circuit itself, and short lines for the pins with designators printed next to the lines.

pin plug A plug consisting of a slender metal pin inserted between the blades of a PIN JACK for a quick connection. The plug usually has a small insulated back for convenient handling.

pin straightener A device for straightening the pins of a transistor, integrated circuit, or other electronic component.

pin switch A switch that changes state when a small pin is pushed or pulled.

pin-usage factor For an integrated circuit, the number of gate equivalents per package pin. Also see GATE EQUIVALENT.

PIO Abbreviation of *parallel input/output.*

pion A subatomic particle consisting of one quark and one antiquark.

PIP Abbreviation of PICTURE-IN-PICTURE.

pip See BLIP.

pi pad A resistance-type attenuator having a series arm, a shunt input arm, and shunt output arm; its name is derived from its resemblance to the Greek letter pi. Also see PAD.

pipe radiator A waveguide having an open end from which microwave energy is radiated.

Pirani gauge A type of vacuum gauge in which a heated filament, composing one arm of a four-arm resistance bridge, is sealed into the vacuum system. The bridge is balanced before evacuation starts. As evacuation progresses, the heat removed from the filament becomes proportional to the pressure in the system, and the resistance of the filament changes accordingly. The bridge is then rebalanced, and the difference between initial and subsequent null conditions indicate the extent of the vacuum when the bridge has been appropriately calibrated.

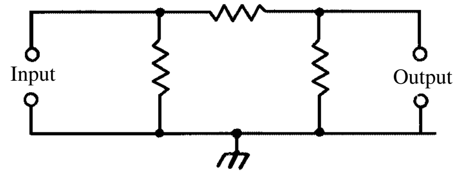

pi pad

pi section An unbalanced filter or tuner section whose schematic representation has the general shape of the uppercase Greek letter pi.

pi-section coupling Use of a PI SECTION for coupling a radio transmitter to an antenna. Also see COLLINS COUPLER and PI-SECTION TANK.

pi-section filter A pi section used as either a low-pass or high-pass filter, depending on the position of the capacitors in the circuit.

pi-section tank A pi section used as the collector, drain, or plate tank circuit of a radio-frequency power amplifier, and also serving as an antenna coupler.

piston **1.** The movable element (cone) of a loudspeaker. **2.** The movable, solid plunger of a trimmer capacitor that consists of a plug within a cylinder.

piston directivity Directivity of sound emitted by the piston of a loudspeaker (see PISTON, **1**). As the frequency of the audio signal increases, radiation from a loudspeaker tends to be concentrated along the axis of the piston.

pit **1.** A microscopic depression in a compact disc; scatters and/or absorbs light from the laser, rather than reflecting it. Compare LAND, **1**. **2.** In a printed-circuit board, a pockmark in a component or foil run. **3.** A pockmark in a metallic substance, resulting from corrosion.

pitch **1.** The frequency of a sound, either in general terms (e.g., low, midrange, and high) or as a specific quantity (e.g., 2450 Hz). **2.** The distance between the peaks of adjacent grooves on a phonograph disc. **3.** The distance between adjacent threads of a screw. **4.** The distance between centers of turns in a coil (see PITCH OF WINDING) **5.** The number of teeth or threads per unit length. **6.** The distance along its axis a propeller moves in a revolution. **7.** Up-and-down motion of a robot end effector or other electromechanical device. **8.** The extent or range of movement, as defined in **7**.

pitch of winding In a coil, the distance from the center of one turn to the center of the adjacent turn in a single layer of winding.

PIV Abbreviation of PEAK INVERSE VOLTAGE.

pivot The sometimes jeweled, stationary member of the bearing in an analog meter movement.

pi-wound choke A choke coil consisting of several series-connected sections, mounted on a single core and separated to reduce internal capacitance.

pix Abbreviation of PICTURE.

pixel Contraction of *picture element.* The smallest bit of data in a video image. Also called *pel.* The smaller the size of the pixels in an image, the greater the resolution for a given image area.

pixel aspect ratio In a video image, the ratio of PIXEL height to pixel width.

pix tube See PICTURE TUBE.

PL Abbreviation for PRIVATE LINE.

place effect An apparent change in the perceived pitch of a sound, caused by variations in the way the waves interact inside the human ear.

planar diffusion In the production of a semiconductor device, the diffusion of all the elements into one face of a wafer. Consequently, connections to the elements all lie in one plane. Also see EPITAXIAL PLANAR TRANSISTOR and PLANAR TRANSISTOR.

planar diode A semiconductor diode, having a pn junction that lies entirely within a single plane.

planar epitaxial passivated diode A junction diode that, like the planar epitaxial transistor, has been manufactured by planar diffusion, then passivated to protect the junction. Also see EPITAXIAL GROWTH, EPITAXY, PASSIVATION, and PLANAR DIFFUSION.

PL/1 A computer programming language that is a hybrid of scientific and commercial types (like ALGOL and COBOL), the combined features being powerful problem-solving and mass-data-handling abilities.

planar epitaxial passivated transistor A planar epitaxial transistor that has been passivated to protect the exposed junctions. Also see EPITAXIAL PLANAR TRANSISTOR, PLANAR TRANSISTOR, EPITAXIAL TRANSISTOR, PASSIVATION, and PLANAR TRANSISTOR.

planar transistor A transistor in which the emitter, base, and collector elements terminate on the same face (plane) of the silicon wafer. A thin film of silicon dioxide is grown on top of the wafer to insulate the exposed junctions after the leads have been attached (i.e., the transistor is passivated).

Planck constant Symbol, *h.* Unit, joule-second. The constant of proportionality in the fundamental law of the quantum theory, stating that radiant energy is composed of quanta proportional to the frequency of the radiation; $h = q/f = 6.62608 \times 10^{-34}$ J•s, where q is the value of the quantum and f is the frequency in Hz.

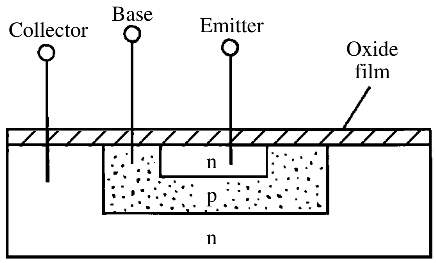

planar transistor

plane of polarization The plane containing the direction of propagation and the electric field vector of a plane-polarized wave (see POLARIZATION, **3** and POLARIZED LIGHT).

plane-polarized light See POLARIZATION, **3** and POLARIZED LIGHT.

plane-reflector antenna A directive antenna in which the reflector is a sheet of metal or a metal screen. In a corner-reflector antenna, the reflector is a folded sheet, or two sheets joined along one edge.

planetary electron See ORBITAL ELECTRON.

planimeter A mechanical instrument for measuring the area of a closed figure. The outline of the figure is traced with the pointer of the device, and the area is read from a pair of dials. In this application, the planimeter does the work of integral calculus.

PLANNER A high-level computer programming language sometimes used in artificial intelligence. It is a "goal-oriented" language in that it seeks a solution to a problem using various schemes, as necessary.

plan position indicator Abbreviation, PPI. A radar display on whose screen small spots of light reconstruct the scanned vicinity, revealing objects, such as buildings, boats, aircraft, etc. The distance from the center of the screen to a spot depicts the range of an object, and the radial angle reveals its bearing.

plant In computer operations, to put the result of an operation specified by a routine in a storage location from which it will be taken for implementation of an instruction further on in the program.

plaque resistor A flat, noninductive, power resistor, often used as a dummy load during high-frequency power measurements.

plasma A usually high-temperature gas that is so highly ionized that it is electrically conductive and susceptible to magnetic fields; it is recognized as one of the states of matter. Also see PHYSICAL PROPERTIES.

plasma diode A diode in which a plasma substance produces conduction in one direction, but not in the other.

plasma length See DEBYE LENGTH.

plasma oscillation In a plasma, a form of electric-field oscillation of the rapidly moving electrons.

plasma torch A torch, used for such high-heat applications as melting metal, in which a gas is heated by electricity to the high temperature at which it becomes a plasma.

plasmatron A form of amplifier tube sometimes used at ultra-high and microwave frequencies. It is similar to a thyratron. An inert gas is excited until it becomes a plasma, producing amplification under certain operating conditions.

Plastacele See CELLULOSE ACETATE.

plastic A synthetic material usually made from various organic compounds through polymerization (see POLYMERIZE). Plastics can be molded into solid shapes and are available as films. Examples: *celluloid*, *cellulose acetate*, *cellulose nitrate*, *polyethylene*, and *polystyrene.* Also see THERMOPLASTIC MATERIAL and THERMOSETTING MATERIAL.

plastic-film capacitor A capacitor made using polyester, polyethylene, or polystyrene. The method of manufacture is similar to that for paper capacitors when the plastic is flexible. Stacking methods can be used if the plastic is more rigid. The geometries can vary, and these capacitors are therefore found in several different shapes. Capacitance values for plastic-film units range from about 50 pF to several tens of microfarads. Most often they are in the range of 0.001 µF to 10 µF. Plastic-film capacitors are employed in audio equipment, and also in wireless transmitters and receivers. The efficiency is good, although not as high as that for mica-dielectric units. Compare CERAMIC CAPACITOR, ELECTROLYTIC CAPACITOR, MICA CAPACITOR, PAPER CAPACITOR, TANTALUM CAPACITOR.

plasticizer A substance added to a plastic to make it softer or more flexible.

plastic-leaded chip carrier Abbreviated PLCC. A surface-mounted package for an integrated circuit. It is small in size and has high electrical and mechanical reliability.

plate **1.** The anode of an electron tube. **2.** One of the electrodes of a primary or secondary battery cell. **3.** One of the electrodes of a capacitor.

plateau In a response curve, a region in which an increase in the independent variable produces no further change in the dependent variable. Example: the saturation region in a common-base transistor collector-current curve.

plate blocking capacitor A capacitor connected between the plate of an electron tube and the plate tank. It allows the direct-current supply voltage to be applied directly to the plate without it passing through the tank coil, while at the same time preventing the tank coil from short-circuiting the plate power supply. The capacitor freely transmits alternating-current signal energy to the tank.

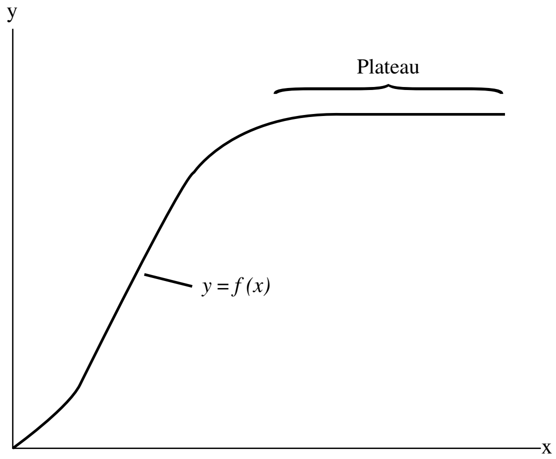

plateau

plate capacitance See PLATE-CATHODE CAPACITANCE.

plate-cathode capacitance Symbol, C_{PK}. Unit, pF. The internal capacitance between the plate and cathode of an electron tube. Also called OUTPUT CAPACITANCE.

plate characteristic For an electron tube, the family of plate current-vs-plate voltage curves for various grid-bias voltages.

plate circuit The external circuit associated with the plate of an electron tube.

plate-circuit relay A direct-current relay operated in series with the plate of an electron tube.

plate conductance Symbol, g_P. Unit, siemens. Conductance of the internal plate circuit of an electron tube. The value of static g_P is equal to the plate current divided by the plate voltage (I_P/E_P). The value of dynamic g_P is equal to the derivative of the static g_P: d_{IP}/d_{EP}. Plate conductance is the reciprocal of PLATE RESISTANCE.

plate current Symbol, I_P. Direct current flowing in the plate circuit of an electron tube.

plate-current shift A change in the plate current of a radio-frequency power amplifier during amplitude modulation. The action discloses faulty operation because the average plate current should remain constant during modulation.

plate dissipation Abbreviation, *PD.* Unit, watt. Power expended in the plate of an electron tube. For an unloaded tube, PD = $E_P I_P$, where E_P is the direct-current (dc) plate voltage in volts, and I_P is the dc plate current in amperes. For a loaded tube, $PD = P_o - P_i$, where P_o is the alternating-current (ac) power output of an amplifier or oscillator in which the tube operates, and P_i is the dc plate power input.

plated magnetic wire A wire with a ferromagnetic outer coating on a core that is not magnetic.

plated-wire memory See WIRE MEMORY.

plate-grid capacitance Symbol, C_{PG} or C_{GP}. Unit, pF. The internal capacitance between the plate and control grid of an electron tube. Also called INTERELECTRODE CAPACITANCE and FEEDBACK CAPACITANCE.

plate load The power-consuming load into which the plate circuit of an electron tube operates. In an intermediate stage of a multistage amplifier, this load is the grid circuit of the following tube.

plate load impedance Symbol, Z_{LP}. Unit, ohm. In a tube circuit, the (output) impedance that is ac-connected between the plate and ground, or dc-connected between the plate electrode and dc plate power supply.

plate meter A direct-current ammeter or milliammeter that indicates the plate current of an electron tube.

plate modulation A method of AMPLITUDE MODULATION in which a modulating voltage is superimposed on the direct-current plate voltage of a higher-frequency amplifier or oscillator.

platen The "roller" in a teletypewriter or impact printer. It supports the paper against impact by the print head; it also helps to move the paper through the machine.

plate-neutralized amplifier A vacuum-tube radio-frequency power amplifier in which a neutralizing capacitor is connected between the control grid and the free end of a center-tapped plate-tank coil.

plate power Symbol, P_P. Unit, watt. Power in the plate circuit of an electron tube; $P_P = E_P I_P$, where E_P is the plate voltage in volts and I_P is the plate current in amperes.

plate power input See PLATE POWER.

plate power output The output signal power delivered by the plate circuit of an electron tube. Compare PLATE POWER INPUT.

plate power supply The (usually direct current) power supply that furnishes energy to the plate of an electron tube.

plate pulse modulation A method of obtaining pulse modulation by injecting high-voltage pulses into the plate circuit of a vacuum-tube amplifier.

plate relay A relay operated in series with the plate of an electron tube.

plate resistance Symbol, r_P. Unit, ohm. Resistance of the internal plate circuit of an electron tube. The static value of r_P is equal to E_P/I_P, where E_P is the plate voltage in volts, and I_P is the plate current in amperes. The dynamic value is dE_P/dI_P.

plate saturation In an electron tube, the point at which, while plate voltage is increasing, the plate attracts all the electrons emitted by the cathode (i.e., the point beyond which no further significant increase in plate current results from a further increase in plate voltage).

plate-screen capacitance Symbol, C_{PS}. Unit, pF. The internal capacitance between the plate and screen grid of an electron tube.

plate-screen modulation A method of AMPLITUDE MODULATION in which the modulating voltage is superimposed simultaneously on the direct-current plate and screen voltages of a higher-frequency amplifier or oscillator.

plate series compensation In an audio amplifier, the use of a plate decoupling circuit to obtain a fixed amount of bass boost.

plate shunt compensation The addition of a network to the plate-output circuit of a tube to boost the bass response of an amplifier.

plate spacing **1.** The distance between plates in a fixed capacitor. This dimension is usually the same as dielectric thickness. **2.** The distance between plates in a variable capacitor. Also called *capacitor air gap.*

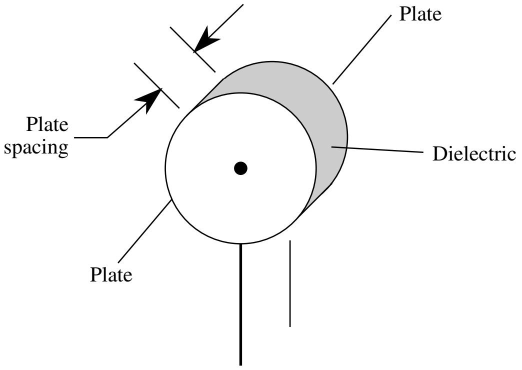

plate spacing, 1

plate supply voltage Symbol, E_{BB}. The output voltage of a plate power supply.

plate tank A resonant inductance-capacitance (LC) circuit operated from the plate of an electron tube.

plate tank capacitance The capacitance required to tune a PLATE TANK to resonance.

plate tank inductance The inductance of the coil in a PLATE TANK.

plate tank *Q* Figure of merit (see *Q*) of a plate tank, a function of the load resistance and the tank inductance-to-capacitance (L/C) ratio.

plate tank voltage The audio-frequency or radio-frequency voltage developed across the plate tank of an electron-tube circuit.

plate tuning Tuning an electron-tube circuit by varying the capacitance, inductance, or both in the plate tank.

plate tuning capacitance See PLATE TANK CAPACITANCE.

plate tuning inductance See PLATE TANK INDUCTANCE.

plate-type capacitor A capacitor having flat metal plates, rather than concentric cylinders, a cylinder and rod, etc.

plate voltage Symbol, E_P. The direct-current voltage applied to the plate of an electron tube.

plate winding **1.** An inductor connected in series between the plate of a vacuum tube and the positive power-supply voltage. **2.** The primary winding of a plate-circuit output transformer.

platiniridium A natural alloy of PLATINUM and IRIDIUM.

platinotron A form of traveling-wave vacuum tube used as an amplifier at ultra-high and microwave frequencies. There are two output connections.

platinum Symbol, Pt. A precious metallic element. Atomic number, 78. Atomic weight, 195.08. It is sometimes used for plating of relay and switch contacts, and for certain parts of vacuum tubes.

platinum metals The rare metals IRIDIUM, OSMIUM, PALLADIUM, PLATINUM, RHODIUM, and RUTHENIUM. They do not react readily with other elements.

platinum-tellurium thermocouple A thermocouple using the junction between platinum and tellurium wires; it is used in thermocouple-type meters.

platter **1.** One of the individual disks in a computer HARD DISK drive. **2.** The rotating turntable in a PHONOGRAPH.

playback The reproduction of recorded material in audio-tape, audio-disc, video-tape, or video-disc systems.

playback computer system A personal computer and associated peripherals, equipped for reproducing multimedia that has been recorded on CD-ROM.

playback head In a magnetic recorder/reproducer, the head that picks up the signal from the tape or disc for reproduction. Also called *read head* and *play head.*

playback loss In disc recording, the difference (at a particular point on the disc) between the recorded level and the reproduced level.

player A semiconductor layer that is doped to provide current carriers that are predominantly holes. Compare N LAYER.

player A device or system that reproduces (plays back) data from a tape or disc, but cannot be used to record data onto the tape or disc.

play head See PLAYBACK HEAD.

playthrough The condition in which an amplifier delivers a small output signal when the gain control is set to zero.

PLC Abbreviation of POWER-LINE COMMUNICATION.

PLCC Abbreviation of PLASTIC-LEADED CHIP CARRIER.

plethysmograph A medical-electronic device that allows the monitoring of the amount of blood in different parts of the body.

PLL Abbreviation of PHASE-LOCKED LOOP.

PLM Abbreviation of *pulse-length modulation* (see PULSE-DURATION MODULATION).

PLO Abbreviation of PHASE-LOCKED OSCILLATOR.

plot **1.** A curve depicting the variations of one quantity, with respect to another. **2.** To generate, print, or display a curve of the type defined in **1**.

plotter A machine that plots (see PLOT, **2**) automatically, often by the direction of a computer.

PL tone encoder An audio oscillator and modulator that cause a subaudible-tone modulation of a signal for use in restricted communications systems.

plug A usually male quick-connect device that can be inserted into a JACK to make a circuit connection, or be pulled out of the jack to break the connection. See, for example, MALE PLUG, PHONE PLUG, POWER PLUG, and POLARIZED POWER PLUG.

plug-and-jack connection A connection made by inserting a PLUG into a JACK.

plug fuse A fuse provided with an Edison base for screwing into a socket.

pluggable Capable of being completely removed from the rest of the system without the need for removing any wiring. Pluggable components and circuit boards simplify the servicing of electronic equipment.

plug-in See PLUG-IN COMPONENT and PLUG-IN UNIT.

plug-in capacitor A capacitor with pins or ferrules that can be quickly inserted into, or removed from, a socket.

plug-in coil A coil wound on a form having pins that can be quickly inserted into, or removed from, a socket.

plug-in coil form An insulating form with base pins that mate with socket terminals so that a coil wound on the form can be quickly inserted into, or removed from, a circuit.

plug-in component A component or module, such as a transistor, capacitor, coil, lamp, etc., provided with pins, clips, or contacts for easy insertion into, or removal from, a circuit. See, for example, PLUG-IN CAPACITOR, PLUG-IN COIL, PLUG-IN FUSE, PLUG-IN LAMP, PLUG-IN METER, PLUG-IN RESISTOR, PLUG-IN TRANSFORMER, and PLUG-IN UNIT.

plug-in fuse A cartridge fuse having a metal ferrule on each end for insertion into a matching clip for easy installation and removal.

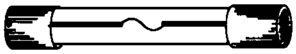

plug-in fuse

plug-in lamp A lamp with base pins for quick insertion into, or removal from, a socket.

plug-in meter A meter with pins or banana plugs for quick insertion into, or removal from, a circuit.

plug-in resistor A resistor with pins or ferrules for quick insertion into, or removal from, a socket.

plug-in transformer A small transformer with pins for quick insertion into, or removal from, a socket.

plug-in unit A unit, such as a tuned circuit, amplifier, or meter, that has pins or contacts for easy insertion into, or removal from, a larger piece of equipment.

plumber's delight An antenna whose construction, including that of the mast, is entirely of metal rods or tubing, with no insulating parts. Short circuits and grounds are prevented by making all attachments and joints at points that are at zero voltage, with respect to the standing-wave pattern.

Plumbicon A television camera tube, similar to the VIDICON, with a lead-oxide target. It is noted for high sensitivity. The image lag time is shorter than in the conventional vidicon.

plumbing Collectively, the waveguides, tees, elbows, and similar pipelike devices and fixtures used in microwave setups.

plunger-type meter A meter in which an iron or steel plunger is pulled into a coil by the magnetism produced by a current flowing in the coil. The plunger is attached to a pointer that moves over the scale.

plutonium Symbol, Pu. A radioactive metallic element that is artificially produced. Atomic number, 94. Atomic weight, approximately 244.

PM **1.** Abbreviation of PERMANENT MAGNET. **2.** Abbreviation of PULSE(D) MODULATOR. **3.** Abbreviation of *post meridian.* **4.** Abbreviation of PHASE MODULATION.

Pm **1.** Symbol for PROMETHIUM. **2.** Abbreviation of PETAMETER.

Pm Symbol for MAXIMUM POWER.

PME Abbreviation of *photomagnetoelectric.*

PMG Abbreviation of PERMANENT-MAGNET GENERATOR.

PMM Abbreviation of PERMANENT-MAGNET MAGNETIZER.

PMOS Abbreviation of P-CHANNEL METAL-OXIDE SEMICONDUCTOR.

PMU Abbreviation of *portable memory unit.*

PN **1.** Abbreviation of POLISH NOTATION. **2.** Abbreviation of POSITIVE-NEGATIVE (often lowercase).

pn Abbreviation of POSITIVE-NEGATIVE.

pn boundary See PN JUNCTION.

pnip transistor A junction transistor having an intrinsic layer between an n-type semiconductor base and one of the p-type semiconductor layers.

pneumatic computer A computer that uses fluid logic [i.e., one in which information is stored and transferred by the flow of a fluid (gas or liquid) and pressure variations therein].

pn junction The boundary between p-type and n-type semiconductor materials in a single block or wafer of the materials. The junction cannot be duplicated by merely touching two pieces of material (one n-type and one p-type) together, however smooth their mating faces.

pn-junction diode A diode consisting of the junction between p-type and n-type regions in the same wafer of semiconductor material.

PNM Abbreviation of PULSE-NUMBERS MODULATION.

pnpn device See NPNP DEVICE.

pnp transistor A bipolar junction transistor in which the emitter and collector layers are p-type semiconductor material, and the base layer is n-type semiconductor material. Compare NPN TRANSISTOR.

Po Symbol for POLONIUM.

Po Symbol for OUTPUT POWER or POWER OUTPUT.

POGO Abbreviation of *polar orbiting geophysical observatory.*

point **1.** A dot indicating the place of separation between the integral and fractional parts of a number (e.g., *decimal point*). Also called RADIX POINT. **2.** A precisely defined location in three-dimensional space that has theoretically zero length, zero width, and zero depth (e.g., *focal point*). **3.** The place on a graph in any number of dimensions, at which two or more curves or coordinates intersect. **4.** A set of operating conditions for a component, device, or system (e.g., *cutoff point, operating point*). **5.** A defined condition at which some specific physical phenomenon occurs (e.g., *melting point*).

point charge An electric charge imagined to occupy a single point in space; thus, it has neither area nor volume.

point contact The point at which the sharply pointed tip of a wire or rod conductor touches a second conductor (e.g., the contact between a "cat whisker" and a semiconductor wafer).

point-contact diode A semiconductor diode having a fine wire ("cat whisker"), whose point is in permanent contact with the surface of a wafer of semiconductor material, such as germanium or silicon.

point-contact junction The pn junction electroformed under the point at which the "cat whisker" touches the semiconductor wafer in a point-contact diode or transistor.

point-contact transistor A transistor having two fine wires ("cat whiskers") that serve as the emitter and collector electrodes. The pointed tips of the wires are nearly in contact with (a few mils apart from) the surface of a wafer of semiconductor material, such as germanium. The semiconductor serves as the base electrode. This device was a predecessor of the JUNCTION TRANSISTOR.

point counter A Geiger counter tube in which the central electrode is a pointed, fine wire. Also see PROPORTIONAL COUNTER.

point defect **1.** In a semiconductor substance or piezoelectric crystal, the absence of an atom from its place in the lattice structure. **2.** The presence of an extra atom in the lattice structure.

point effect The tendency of an electrical discharge to occur more readily at a sharp point than at a blunt surface (as of an electrode).

pointer A pointed blade, stiff wire, or inscribed line on a transparent blade; it moves over a scale to indicate a setting or the value of a quantity. Also called NEEDLE.

pointer-type meter An analog meter in which a pointer moves over a calibrated scale.

point impedance **1.** The impedance observed at a given point in a circuit. **2.** In a transmission line, the intensity of the electric field divided by the intensity of the magnetic field at a given point.

points of saturation For a magnetic core, saturation as evidenced by a leveling-off of the positive and negative halves of the magnetization curve.

point source A source from which electromagnetic radiation emanates, and that appears as a geometric point from a great distance.

point mode Descriptive of cathode-ray-tube display operation (in a computer system), in which data is portrayed as plotted dots.

point-to-point communication Communication between two stations whose location can be precisely specified.

point-to-point motion A method of robot arm movement in which the device can attain only certain positions. The coordinates of each stopping point are stored in the robot controller (computer) memory.

point-to-point station A radio station that provides POINT-TO-POINT COMMUNICATION.

point-to-point wiring A method of wiring an electronic circuit in which wires are run directly between the terminals or components, usually by the shortest practicable route. It is used mainly in high-voltage circuits, such as power amplifiers. Compare CABLED WIRING.

poise The cgs unit of absolute viscosity; 1 poise is the absolute viscosity of a fluid that requires a shearing force of 1 dyne to move a 1-sq-cm area of one of two parallel layers of the fluid (1 cm apart) with a velocity of 1 cm per second, with respect to the other layer. The comparable SI unit is the *newton-second per meter squared* ($N \cdot s/m^2$); 1 poise = $0.1\ N \cdot s/m^2$.

polar axis **1.** In a crystal, the axis of rotation not perpendicular to a reflection plane. **2.** The straight line connecting epicenters of electric, magnetic, or gravitational poles in a system. **3.** The axis about which the earth or another planet rotates.

polar coordinate conversion See POLAR COORDINATE TRANSFORMATION.

polar-coordinate geometry A two-dimensional system for movement of industrial-robot arms. It is based on a system of POLAR COORDINATES, in which a radius and a direction angle are assigned to each point in the working plane.

polar coordinates The magnitude and direction of a vector in a defined plane, listed as a radius (magnitude) in combination with an angle (direction) between the vector and the polar axis.

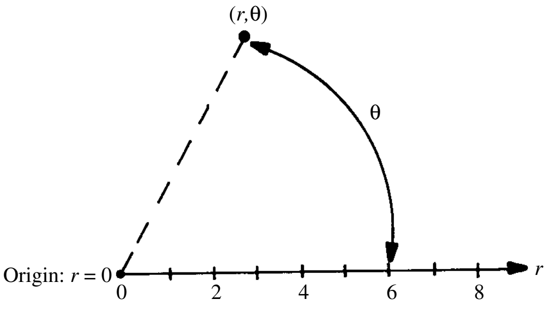

polar coordinates

polarimeter An instrument for measuring the amount of polarized light in a ray that is only partially polarized.

polariscope An instrument used in the observation or testing of materials under POLARIZED LIGHT.

polarity **1.** The condition of being electrically positive or negative. **2.** The condition of being magnetically north or south. **3.** The orientation of the positive and negative poles in a battery or power supply relative to a circuit. **4.** The orientation of a magnetic field, relative to the surrounding environment.

polarity blanking See POLARITY INHIBIT.

polarity inhibit In some instruments, especially those having automatic polarity, the automatic blanking of the polarity sign.

polarity-sensitive relay A direct-current relay actuated only when coil current flows in one direction. One of the simplest versions is a relay having a semiconductor diode connected in series with its coil.

polarity shifter A potentiometer connected to two direct-current sources so that a pair of output terminals has plus and minus polarities at one extreme of potentiometer adjustment, and minus and plus at the other extreme. At the center of the range, the output voltage is zero.

polarity switch A double-pole, double-throw switch connected between a pair of direct-current input terminals so that the polarity of a pair of output terminals can be interchanged.

polarization **1.** In a radio wave, the orientation of the electric lines of flux, with respect to the surrounding environment (e.g., *horizontal polarization* and *vertical polarization*). **2.** The disabling of a battery cell by the formation of insulating gas on one of the plates. **3.** The condition in which transverse waves of light are confined to a specific (e.g., horizontal or vertical) plane.

polarization diversity A form of reception in which two separate receivers, tuned to the same signal, are connected to independent antennas. One antenna is vertically polarized and the other is horizontally polarized. The result is a reduction in fading caused by ionospheric effects on the polarization of the incoming signal.

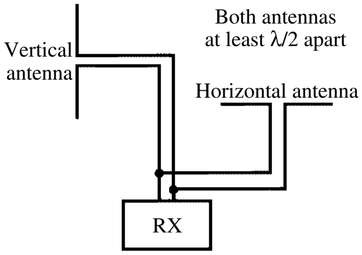

polarization diversity

polarization error In the operation of a loop antenna (e.g., that of a direction finder), null error caused by waves arriving with polarization opposite that of the loop (thus, vertically polarized waves at a horizontal-plane loop, and vice versa).

polarization fading In radio reception, a form of fading that results from changes in the polarization of the arriving signal with respect to the receiving antenna. When the polarization of the arriving signal coincides with that of the receiving antenna, the received signal strength is maximum. When the received-signal polarization is at right angles to the receiving antenna, the signal strength is minimum.

polarization modulation A method of impressing information on a signal by changing the polarization of the radiated electromagnetic field.

polarization selectivity For a photoemissive surface, the condition in which the ratio of photocurrents for two different angles of plane polarization of the light incident to the surface differs from the ratio of the corresponding amounts of light absorbed by the surface. Also see PHOTOEMISSION; PHOTOEMISSIVE MATERIAL; and POLARIZATION, **3**.

polarized capacitor A conventional electrolytic capacitor, so called because one particular terminal must be connected to the more positive of the two connection points. Compare NONPOLARIZED ELECTROLYTIC CAPACITOR.

polarized light Visible light waves whose electric lines of flux are confined more or less to a single plane. This effect can be obtained via filtering; it also occurs naturally under certain conditions. Scattered sunlight is polarized to some extent by the atmosphere. Light is polarized to some extent when it reflects from a plane surface. Also see POLARIZING FILTER.

polarized plug A plug that can be inserted into a socket or receptacle in only one way to ensure safe and foolproof operation.

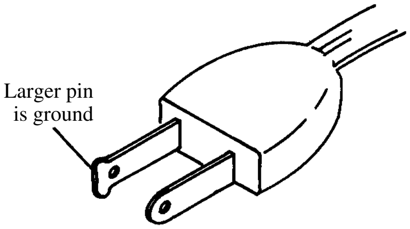

polarized plug

polarized power plug A polarized plug for connection of equipment to an alternating-current utility power source.

polarized reactor A saturable reactor in which the lines of flux produced in the three-leg core by the coils on the two outer legs are added in the center leg. Consequently, the flux can reinforce or oppose the controlling current in the coil on the center leg, depending on the direction of the current. Compare NONPOLARIZED REACTOR.

polarized receptacle A receptacle constructed so that it can receive a plug in only one way, thus preventing incorrect connections.

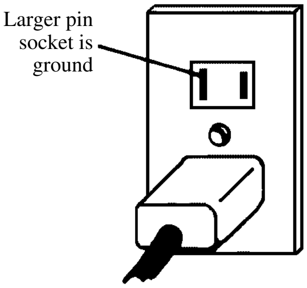
Larger pin socket is ground

polarized receptacle

polarized relay A relay actuated by one polarity of direct current, or by one particular phase of alternating current. Such a relay sometimes contains an armature-centering permanent magnet.

polarized socket See POLARIZED RECEPTACLE.

polarized X rays X-ray waves whose electric lines of flux are more or less oriented in a specific plane, as when they are scattered by carbon blocks. See, for illustration, POLARIZATION, **3** and POLARIZED LIGHT.

polarizing filter A light filter that consists essentially of microscopic heraphathite plastic plate. The plate can be rotated to cause light passing through it to be polarized in any particular plane between horizontal and vertical, depending on the degree of rotation. Also see POLARIZED LIGHT.

polarography In chemistry, a form of qualitative or quantitative analysis utilizing IE curves obtained when the voltage is gradually increased across a solution in ELECTROLYSIS.

polar orbit An orbit that carries a satellite over the geographic polar regions of the earth. This type of orbit is oriented at, or nearly at, 90 degrees with respect to the equator, and can have a period ranging from about 90 minutes to several weeks or even months. Low-earth-orbit (LEO) satellites generally have such orbits.

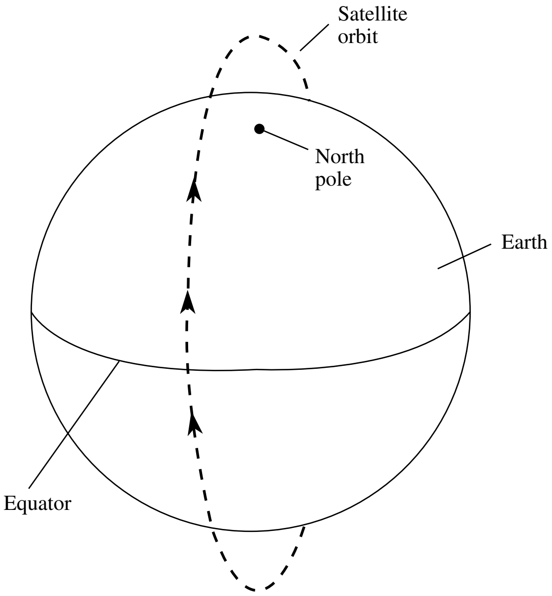

polar orbit

polar-orbiting satellite Any satellite in a POLAR ORBIT.

polar planimeter See PLANIMETER.

polar relay See POLARIZED RELAY.

polar response The horizontal-plane directional response of an antenna or other transducer.

pole **1.** An extremity or terminus that possesses POLARITY. Examples: *magnetic pole* and *electric pole.* **2.** The movable member of a switch. **3.** One of the frequencies at which a transfer function becomes infinite.

pole face The smooth end surface of a pole piece.

pole piece **1.** A section of specially shaped iron or steel that is attached to a magnetic core. **2.** Half of a two-piece magnetic core that terminates in a pole.

pole shoe In an electric motor, the section of the field pole nearest the armature.

poles of impedance For a reactive network, the frequencies at which the impedance is infinite. Compare ZEROS OF IMPEDANCE.

poles of network function The values at which a network function is infinite. Compare ZEROS OF NETWORK FUNCTION.

poles of transfer function The frequencies at which a transfer function becomes infinite. Compare ZEROS OF TRANSFER FUNCTION.

police robot A proposed robotic machine, either autonomous or remotely controlled (teleoperated), that would be used to assist in law enforcement, particularly in dangerous operations.

poling The deliberate adjustment of electromagnetic-field polarity.

Polish notation In Boolean algebra, a form of notation wherein the variables in a statement are preceded by the operators.

polling In data transmission, a technique in which channels being shared by more than one terminal are tested to find one over which data is coming in, or to ascertain which is free for transmission.

polonium Symbol, Po. A radioactive metallic element. Atomic number, 84. Atomic weight, approximately 209.

polychromatic radiation **1.** Visible-light radiation having more than one wavelength; in particular, covering a broad range of wavelengths. **2.** Electromagnetic radiation over a broad band of wavelengths.

polycrystalline material A substance, such as a semiconductor, of which even a very small sample consists of a number of separate crystals bound tightly together. Compare SINGLE-CRYSTAL MATERIAL.

polydirectional microphone See OMNIDIRECTIONAL MICROPHONE.

polyelectrolyte An ELECTROLYTE having high molecular weight.

polyester A resin made by reacting a dihydroxy alcohol with a dibasic acid.

polyester backing A polyester tape on the surface of which iron oxide is deposited to yield a magnetic recording tape.

polyethylene A plastic insulating material. Dielectric constant, 2.2. Dielectric strength, 585 V/mil.

polyethylene disc A phonograph disc made of polyethylene.

polygonal coil A coil wound on a form having a polygonal, rather than circular cross section. Some polygonal forms have as many as 12 sides in cross section.

polygraph An instrument for measuring and recording electrical signals proportional to blood pressure, skin resistance, breathing rate, and other reactions that vary under emotional stress. Also called LIE DETECTOR. Compare PATHOMETER.

polymer A compound that is the product of *polymerization*, resulting from the chemical union of monomers. Also see POLYMERIZE.

polymerize To unite monomers or polymers of the same kind to form a molecule having a higher molecular weight.

polyphase In alternating-current circuits, pertaining to the existence or generation of two or more specific electrical phases. Compare SINGLE PHASE.

polyphase antenna An antenna consisting of two dipole radiators mounted perpendicular to each other at their midpoint and excited 90 degrees out of phase. The radiation pattern is approximately circular in the plane of the elements. Also called TURNSTILE ANTENNA.

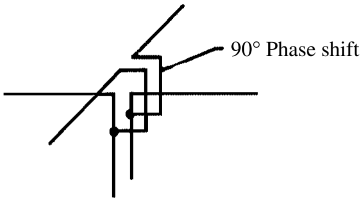

polyphase antenna

polyphase generator **1.** A dynamo-type generator of polyphase power (two-phase, three-phase, etc.). **2.** See POLYPHASE OSCILLATOR.

polyphase oscillator An oscillator circuit that generates polyphase alternating current. The circuit contains separate oscillators for each phase. A three-phase circuit, for example, has three symmetrical oscillators with matched inductance and capacitance values.

polyphase power **1.** The total dissipated power in a polyphase alternating-current circuit. **2.** Polyphase alternating current provided for utility purposes.

polyphase rectifier A rectifier of polyphase alternating current generally obtained from a three-phase power line through a transformer. The several common circuits usually contain a diode for each phase. Such rectifiers offer the advantage of higher ripple frequency than is obtainable by single-phase operation. For a three-phase rectifier, for example, the ripple frequency is three times the line frequency; for a six-phase rectifier, it is six times the line frequency.

polyphase system An alternating-current circuit in which voltages or currents are normally out of phase with each other by some fixed amount. Familiar types are two-phase and three-phase.

polyphase transformer An alternating-current transformer specifically designed for use in circuits that have two or more simultaneous current phases.

polypropylene A plastic material commonly used as an electrical insulator. Dielectric constant, 2.0. Dielectric strength, 600 V/mil.

polyrod antenna A tapered dielectric antenna, usually made of polystyrene, for directional microwave transmission.

polysilicon A polycrystalline form of silicon (see POLYCRYSTALLINE MATERIAL).

polystyrene A clear, colorless thermosetting-type plastic. It is widely used as an insulating material in radio-frequency circuits and, to some extent, as a dielectric film in fixed capacitors. Dielectric constant, 2.4 to 2.9. Dielectric strength, 20 to 28 kV/mm.

polystyrene capacitor A high-Q capacitor in which the dielectric film is polystyrene.

polyvinyl chloride Abbreviation, PVC. A plastic insulating material. Dielectric constant, 3.6 to 4.0. Dielectric strength, 800 V/mil.

pool cathode In an industrial electron tube, a cathode consisting of a pool of mercury.

pool-cathode tube An industrial electron tube using a pool cathode. Examples: excitron, ignitron, and mercury-arc rectifier.

popcorn noise A temperature-dependent, random-shot electrical noise. In a radio receiver or audio circuit, this noise resembles the sound of popping corn. It occurs in some operational amplifiers.

population In statistical analysis, the total group of items, quantities, or values under consideration. Sometimes called *universe.*

porcelain A hard, white, usually glazed ceramic used as a dielectric and insulant. Dielectric constant, 6 to 7.5. Dielectric strength, 40 to 100 V/mil. Also called *china.*

porcelain capacitor A ceramic-dielectric capacitor in which the dielectric is composed of porcelain or a related substance.

porcelain insulator An electric insulator fabricated from porcelain.

porch See BACK PORCH and FRONT PORCH.

port **1.** In a circuit, device, or system, a point at which energy or signals can be introduced or extracted in a particular manner (e.g., *two-port circulator* and *I/O port*). **2.** An aperture in a loudspeaker enclosure.

portable-mobile station See MOBILE STATION.

portable station A communications station that can be carried from one location to another. A portable station differs from a mobile station in that a portable station does not usually operate while in motion, whereas a mobile station does.

ported reflex enclosure A loudspeaker cabinet with openings that facilitate bass (low-frequency) sound reproduction.

pos **1.** Abbreviation of POSITIVE. **2.** Abbreviation of POSITION.

position **1.** The location of a point or object with respect to one or more (usually fixed) references. **2.** The setting of an adjustable device, such as a potentiometer, rotary switch, or variable capacitor.

positional notation A method of representing numbers in which the number is indicated by the positions and value of the component digits. The decimal number system belongs in this category (e.g., the decimal number 1284.67 is equal to $1 \times 10^3 + 2 \times 10^2 + 8 \times 10^1 + 4 \times 10^0 + 6 \times 10^{-1} + 7 \times 10^{-2}$).

positional number system See POSITIONAL NOTATION.

positional representation See POSITIONAL NOTATION.

position-control potentiometers In an oscilloscope, potentiometers used to control the voltage applied to the horizontal and vertical deflecting plates to position the spot on the screen. Also see CENTERING CONTROL.

position controls See POSITION-CONTROL POTENTIOMETERS.

position feedback In a servo or other control system, feedback current, or voltage that is proportional to the position assumed by a member.

position fixing The determination of a position from the intersection on a map of two lines derived from the direction-finding pickups of two transmitting stations. Also see DIRECTION FINDER.

position indicator In a tape recorder, a counter whose numbered wheels revolve when the reels do, thus aiding in locating a desired spot on the tape. Also called *tape counter.*

positioning circuit The circuit associated with a horizontal or vertical centering control (see CENTERING CONTROL).

position sensing **1.** In robotics and navigation systems, a method of determining location, relative to the surrounding environment. **2.** Any method via which a robot can accurately determine the location(s) of its end effector(s).

position sensor An electronic circuit that detects physical displacement, and transmits a signal proportional to the displacement.

positioning control See CENTERING CONTROL.

positive **1.** Possessing positive (plus) direct-current electrical polarity. **2.** Pertaining to real numbers greater than zero. **3.** A photographic image whose shadings are the same as those in the scene.

positive angle **1.** In a system of rectangular coordinates, an angle in the first or second quadrant. **2.** In rectangular coordinates, an angle measured counterclockwise from the positive x-axis. Compare NEGATIVE ANGLE.

positive bias A positive voltage or current applied continuously to an electrode of a device (as to a transistor base) to maintain the device's operating point. Compare NEGATIVE BIAS.

positive bus See POSITIVE CONDUCTOR.

positive charge An electrical charge characterized by having relatively fewer electrons than a negative charge. Also see CHARGE, **1**; ELECTRIC CHARGE; and UNIT ELECTROSTATIC CHARGE. Compare NEGATIVE CHARGE.

positive conductor The conductor or line connected to the positive terminal of a current, voltage, or power source. Compare NEGATIVE CONDUCTOR.

positive electricity See POSITIVE CHARGE and POSITIVE ELECTRIFICATION.

positive electrification Electrification characterized by a deficiency of electrons. For example, when a glass rod is rubbed with a silk cloth, the rod becomes positively charged because electrons are rubbed off the glass onto the silk. Similarly, when an atom loses an electron, it becomes electrified positively because it has a deficiency of electrons. Compare NEGATIVE ELECTRIFICATION.

positive electrode **1.** An electrode connected to the positive terminal of a current, voltage, or power source. **2.** The positive terminal of a current, voltage, or power source, such as a battery or generator.

positive electron See POSITRON.

positive element See POSITIVE ELECTRODE, **1**.

positive error of measurement An error of measurement in which the difference between a measured value and the true or most probable value is positive. Compare NEGATIVE ERROR OF MEASUREMENT.

positive exponent A positive superscript indicating that a number (x) is to be raised to the positive nth power. Thus, in the expression xn, the value of n is greater than zero. Compare NEGATIVE EXPONENT.

positive feedback Feedback that is in phase with an input signal. Also called REGENERATION and REGENERATIVE FEEDBACK. Compare NEGATIVE FEEDBACK.

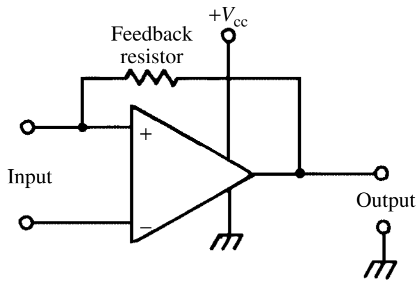

positive feedback

positive function A function having the positive sign. In the rectangular coordinate system, the trigonometric sine function is positive in the first and second quadrants, the cosine in the first and fourth, and the tangent in the first and third. Compare NEGATIVE FUNCTION.

positive ghost In a television picture, a ghost with positive shading (see POSITIVE, **3**). Also see GHOST.

positive-going Pertaining to a signal whose value is changing in a positive direction. This is not restricted to signals of actual positive polarity; a decreasing negative voltage, for example, is positive-going as it falls in the direction of zero—even if it never crosses the zero line.

positive grid In an electron tube, a control grid whose bias or signal voltage is positive, with respect to the cathode.

positive-grid oscillator A microwave oscillator circuit in which the control grid of a triode tube is operated at a positive direct-current potential, and the plate at a negative potential. Electrons move back and forth between cathode and plate, through the grid, and thus give rise to an oscillating current.

positive ground A direct-current electrical system in which the positive power-supply terminal is connected to the common ground. It is not generally used in North America.

positive half-alternation See POSITIVE HALF-CYCLE.

positive half-cycle That half of an alternating-current cycle in which the current or voltage increases from zero to maximum positive and returns to zero.

positive image **1.** A picture in which the blacks, whites, and grays correspond to those in the actual scene (see POSITIVE, **3**). **2.** A normal television picture (i.e., one that has the shading described in **1**).

positive ion An atom that has a deficiency of electrons and, consequently, exhibits a net positive charge. Also called CATION.

positive lead See POSITIVE CONDUCTOR.

positive light modulation In television transmission, the condition in which transmitted power increases as the light intensity increases. Compare NEGATIVE LIGHT MODULATION.

positive line See POSITIVE CONDUCTOR.

positive logic **1.** Binary logic in which a low positive state represents logic 0, and a high positive state represents logic 1. **2.** Binary logic in which a high negative state represents logic 0, and a low negative state represents logic 1. Compare NEGATIVE LOGIC.

positive magnetostriction A form of MAGNETOSTRICTION in which the physical size of a substance is directly proportional to the intensity of the surrounding magnetic field.

positive measurement error See POSITIVE ERROR OF MEASUREMENT.

positive modulation In amplitude-modulated television transmission, the increase in transmitted power when the brightness of the scene increases. Compare NEGATIVE MODULATION.

positive modulation factor For an amplitude-modulated wave having unequal positive and negative peaks, a ratio expressing the maximum positive deviation (increase) from the average value of the envelope. Compare NEGATIVE MODULATION FACTOR.

positive number A real number, whose value is greater than zero. Compare NEGATIVE NUMBER.

positive peak The maximum amplitude of a positive half-cycle or positive pulse.

positive-peak clipper A peak clipper that levels off the positive half-cycle of an alternating-current wave to a predetermined level.

positive-peak modulation Amplitude modulation of the positive peaks of a carrier wave.

positive-peak voltmeter An electronic voltmeter for measuring the amplitude of the positive peak of an alternating-current (ac) wave. In its simplest form, it consists essentially of a direct-current (dc) microammeter with a diode oriented to pass the positive half-cycle. A series capacitor in the circuit is charged to approximately the peak value of the applied ac voltage. Compare NEGATIVE-PEAK VOLTMETER.

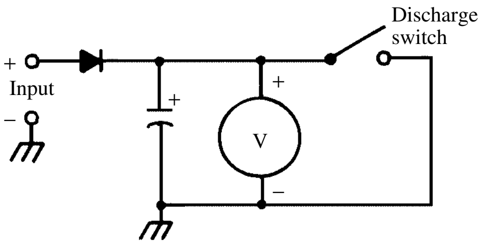

positive-peak voltmeter

positive phase-sequence relay A phase-sequence relay that responds to the positive phase sequence in a polyphase circuit. Compare NEGATIVE PHASE-SEQUENCE RELAY.

positive picture modulation See POSITIVE MODULATION.

positive picture phase In a television signal, the swinging of the picture-signal voltage from zero to positive, in response to an increase in brightness in the scene. Compare NEGATIVE PICTURE PHASE.

positive plate **1.** The positive member of an electrochemical cell or battery. Electrons flow to this plate from the negative plate, through the external circuit. **2.** A vacuum-tube plate that is biased positively, as in a conventional tube circuit.

positive pole See POSITIVE ELECTRODE, **1**, **2**.

positive potential **1.** The voltage at a positive electrode (with respect to the negative electrode). **2.** Voltage greater than that at ground as a reference.

positive power See POSITIVE EXPONENT.

positive resistance Ohmic resistance (see OHMIC RESPONSE). Compare NEGATIVE RESISTANCE.

positive resistor A resistor whose value does not change with current or voltage changes. Compare NEGATIVE RESISTOR.

positive temperature coefficient Abbreviation, PTC. A number expressing the amount by which a quantity (such as the value of a component) increases when temperature is increased. The coefficient is stated as a percentage of the rated value per degree, or in parts per million per degree. Compare NEGATIVE TEMPERATURE COEFFICIENT and ZERO TEMPERATURE COEFFICIENT.

positive transmission In facsimile or television, a form of amplitude modulation in which the picture brightness is directly proportional to the signal strength at any given instant of time.

positive valence The valence of a positive ion.

positron A positively charged particle having the same mass as that of the electron, and the same magnitude of electric charge, but positive (instead of negative). Sometimes called *positive electron.*

post See BINDING POST.

post- Prefix meaning “following,” “subsequent to,” or “behind.”

post-accelerating electrode In a cathode-ray tube, the high-voltage electrode that produces POST-DEFLECTION ACCELERATION of the electron beam. Also called INTENSIFIER ELECTRODE.

postacceleration See POST-DEFLECTION ACCELERATION.

post-alloy-diffused transistor Abbreviation, PADT. A transistor in which electrodes are diffused into the semiconductor wafer after other electrodes have been alloyed.

post-conversion bandwidth The bandwidth of a signal after it has been converted from one frequency to another.

post-deflection accelerating electrode See POST-ACCELERATING ELECTRODE.

post-deflection acceleration In a cathode-ray tube, the intensification of the electron beam following beam deflection. Also see POST-DEFLECTION CRT.

post-deflection CRT An oscilloscope tube provided with a high-voltage intensifier electrode in the form of a ring encircling the inside flare of the tube, between the deflecting plates and the screen. The deflected electron beam is accelerated by this electrode. This arrangement allows the beam to be deflected at low velocity and high sensitivity, then to be accelerated for a brighter image.

post edit The editing of data in a computer output.

postemphasis See DEEMPHASIS.

post-equalization **1.** In sound recording and reproduction, equalization during playback. Compare PREEQUALIZATION. **2.** See DEEMPHASIS.

postmortem An investigation into the cause of failure of a circuit, device, or system.

postmortem dump At the end of a computer program run, a dump to supply information for debugging purposes.

post office box A type of wheatstone bridge that contains resistance coils in a special box. The coils are connected so that they can be replaced by shorting connectors.

pot **1.** See POTENTIOMETER. **2.** See DASHPOT. **3.** Abbreviation of POTENTIAL. **4.** To encapsulate a circuit in a potting compound, such as epoxy resin.

potassium Symbol, K. A metallic element of the alkali-metal group. Atomic number, 19. Atomic weight, 39.098.

potassium chloride Formula, KCl. A compound used as a phosphor coating on the screen of a nearly permanent-persistence cathode-ray tube. The fluorescence is magenta or white, as is the phosphorescence.

potassium cyanide Formula, KCN. A highly toxic salt that is an electrolyte in some forms of electroplating.

potassium dihydrogen phosphate Abbreviation, KDP. An inorganic ferroelectric material.

pot core A magnetic core for a coil, made of ferrite or of powdered iron, consisting of a central rod, a surrounding potlike enclosure, and a lid. The rod passes through the center of the coil, and the pot and lid completely enclose the coil. This arrangement provides a completely closed magnetic circuit and coil shield.

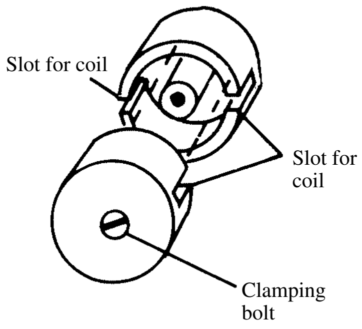

pot core

potential See ELECTROMOTIVE FORCE.

potential barrier The electric field produced on each side of a semiconductor junction by minority carriers (i.e., by holes in the n-layer and electrons in the p-layer) that face each other across the junction, but cannot diffuse across the junction and recombine.

potential coil The shunt coil in a conventional wattmeter.

potential difference See ELECTROMOTIVE FORCE and VOLTAGE.

potential divider See VOLTAGE DIVIDER.

potential drop **1.** A voltage difference between two points in a circuit. **2.** The voltage across a resistor in a direct-current circuit.

potential energy Energy resulting from the position of a body or particle (e.g., the energy stored in something lifted against gravity and held in its new position) or from the position of charges (e.g., the energy stored in a charged capacitor). Compare KINETIC ENERGY.

potential gradient See VOLTAGE GRADIENT.

potential profile A rectangular-coordinate display of the VOLTAGE GRADIENT across a body (e.g., the cross section of a transistor).

potential transformer A small step-up transformer for increasing the range of an alternating-current voltmeter.

potentiometer **1.** A variable resistor used as a voltage divider. The input voltage is applied across the entire resistance element and the output voltage is taken from the wiper, relative to one end of the element. One end is usually grounded (at zero potential). **2.** A null device whose operation is based on a variable resistor, and is used for precise voltage measurements. The unknown voltage is applied to the input of a variable resistor whose settings are known with great accuracy; the resistance is adjusted for an output voltage that exactly equals the voltage of a standard cell (as indicated by a null between the two voltages). The unknown voltage is then determined from the resistance and the standard-cell voltage.

potentiometer noise In a current-carrying potentiometer, electrical noise generated when the wiper blade rubs against the resistance element, or by contact between the blade and element.

potentiometric recorder A type of graphic recorder. It consists essentially of a resistance-calibrated potentiometer, a standard cell, and a galvanometer. When an unknown voltage (E_x) is applied to the input terminals of the potentiometer and the potentiometer is set for null, $E_x = E_s$ (R_2/R_1), where Es is the voltage of the standard cell, $R1$ is the input resistance of the potentiometer, and $R2$ is the output resistance of the potentiometer.

Potier diagram An illustration of the phase relationship between current and voltage in an alternating-current circuit that contains reactance.

POTS Acronym for *plain old telephone service*, meaning basic service without optional features (such as call waiting, conference calling, call forwarding, etc.).

potted circuit A circuit embedded in plastic or wax to protect it against the environment, and/or to minimize the effects of physical vibration (see POTTING).

potted component An electronic part embedded in a suitable plastic or wax to protect it against the environment, and/or to minimize the effects of physical vibration (see POTTING).

potting A process of embedding a component or circuit in a solid mass of plastic or wax held in a container. The process is similar to encapsulation, except that in potting, the container (envelope) remains as part of the assembly. Compare ENCAPSULATION.

potting material A substance, such as a resin or wax, used for potting electronic gear. Also called *potting compound.*

pound **1.** Abbreviations, lb, p. A unit of weight equal to 16 avoirdupois ounces. **2.** Abbreviation, lbf. A unit of force approximately equal to 4.448 newtons. **3.** Abbreviation, lbm. A unit of mass approximately equal to 0.4536 kilogram.

poundal A unit of force equal to approximately 13825.5 dynes or 0.138255 newton. One poundal is the force that, when acting for one second, will impart a speed of one foot per second to a one-pound mass.

pound-foot Abbreviation, lb-ft. A unit of torque equal to the product of a force of one pound and a moment arm of one foot. Compare OUNCE-INCH.

pounds per square inch absolute Abbreviation, psia. Absolute pressure (i.e., the sum of atmospheric pressure and the pressure indicated by a gauge). Compare POUNDS PER SQUARE INCH GAUGE.

pounds per square inch gauge Abbreviation, PSIG. The value of pressure indicated by a gauge, without correction for atmospheric pressure. Compare POUNDS PER SQUARE INCH ABSOLUTE.

powdered-iron core A magnetic core consisting of minute particles of iron, each coated with a film to insulate it from others, molded into a solid mass. Because of its low eddy-current loss, this type of core is usable in radio-frequency transformers and coils, where it increases the inductance of the winding.

power **1.** Symbol, *P.* Unit, watt. The rate of doing work, or producing or transmitting energy. In direct-current circuits, and in alternating-current circuits containing no reactance, power is the product of the root-mean-square current and voltage. See, for example, AC POWER, APPARENT POWER, DC POWER, KILOVOLT-AMPERE, POWER FACTOR, REACTIVE KILOVOLT-AMPERE, REACTIVE VOLT-AMPERE, TRUE POWER, VOLT-AMPERE, WATT, and WATTLESS POWER. **2.** The product obtained by multiplying a quantity x by itself n times, written xn. For example, $2^4 = 2 \times 2 \times 2 \times 2 = 8$; here, 8 is the *fourth power* of 2. Compare ROOT, **1**. **3.** The exponent in an expression, as defined in **2**.

power amplification **1.** The amplification of a signal having a certain power (wattage) to produce a signal having greater power. **2.** The signal power increase, expressed as a ratio or as a figure in decibels, resulting from the process defined in **1**. Also called POWER GAIN.

power amplification ratio See POWER AMPLIFICATION, **2** and POWER GAIN.

power amplifier An amplifier that delivers useful amounts of power to a load, such as one or more speakers. Compare CURRENT AMPLIFIER and VOLTAGE AMPLIFIER.

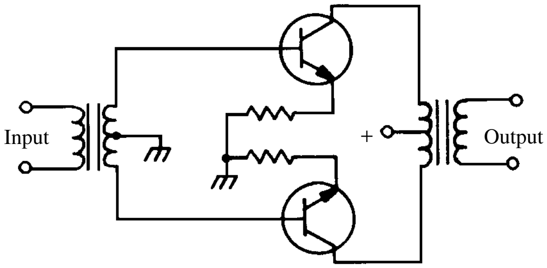

power amplifier

power-amplifier device A high-current tube or transistor designed especially for high power output. Such a device does not always provide significant voltage amplification, but always provides power amplification. Compare VOLTAGE-AMPLIFIER DEVICE.

power at peak torque Symbol, *Pp*. For a torque motor, the input power in watts needed for peak torque at stall at a winding temperature of 25 degrees Celsius.

power attenuation **1.** A reduction of power level. **2.** See POWER LOSS.

power bandwidth For a high-fidelity audio amplifier, the difference between the maximum and minimum frequencies at which the amplifier can produce at least 50 percent of its maximum power output, with less than a certain amount of total harmonic distortion (usually 10 percent).

power blackout A situation in which all electric power is lost to utility subscribers in a defined region.

power consumption **1.** For a direct-current device, the normal operating voltage multiplied by the normal drawn current. **2.** For an alternating-current circuit, the root-mean-square voltage multiplied by the root-mean-square current.

power control The adjustment of the output voltage of a power supply, usually by means of a variable autotransformer, silicon-controlled rectifier, thyratron, or similar device.

power cutoff frequency Symbol, f_{co}. The frequency at which the power gain of a transistor drops 3 dB below its low-frequency value.

power derating For a temperature higher than the specified ambient temperature, a deliberate reduction of the power dissipated by a component or device. This is done to prevent failure of the component or device. Also see DERATING, DERATING CURVE, and DERATING FACTOR.

power difference An expression of the power lost in a circuit when power is absorbed by a dielectric material.

power diode A heavy-duty diode that is usually used in power-supply service. Also called *rectifier diode.*

power dissipation Abbreviation, PD. The power consumed by a device during normal operation. This power is not available in the electrical output of the device. An example is the direct-current power dissipated in the collector circuit of a high-fidelity audio amplifier.

power divider A circuit that distributes power, in a predetermined manner, among various loads.

power drain The amount of power drawn by a device. It can be operating power or standby power.

power dump See DUMP, **2**.

power equations Variations of the basic power equation: $P = EI = E^2/R = I^2R$, where P is the power in watts, E is the voltage in volts, I is the current in amperes, and R is the resistance in ohms.

power factor Abbreviation, PF. In an alternating-current circuit, the ratio (expressed either as a decimal or a percentage) of true power (power actually consumed) to apparent power (simple product of voltage and current). The power factor is equal to the cosine of the phase angle. Also see AC POWER.

power-factor balance In a capacitance bridge, a separate null adjustment for the internal resistance component of a capacitor under measurement. The dial of the variable component for this adjustment reads directly in percent power factor in some bridges.

power-factor correction To raise the power factor of an inductive circuit by inserting a parallel capacitance. In power circuits, this affords improved economy of operation because the current drain is brought more in line with that of a resistive circuit.

power-factor meter An instrument that gives direct readings of power factor (lead or lag). One such meter uses a dynamometer-type movement (see ELECTRODYNAMOMETER) in which the rotating element consists of two coils fastened together at right angles.

power-factor regulator A device that regulates the power factor of an alternating-current line.

power-factor relay An alternating-current relay actuated by a rise or fall in power factor, with respect to a predetermined value.

power frequency **1.** See POWER-LINE FREQUENCY. **2.** The frequency of an alternating-current generator. **3.** The output frequency of a power inverter (see INVERTER, **1**).

power-frequency meter An instrument for measuring power-line frequency. It can use electromechanical devices, or can directly count the number of alternations per second.

power gain Abbreviations, PG or PG_{dB}. The extent to which power is increased by a power amplifier. It can be expressed as the ratio of power output to power input as $PG = P_o/P_i$, or in decibels as $PG_{dB} = 10 \log_{10}(P_o/P_i)$.

power grid An aggregation of power-generating stations, transmission lines, and associated equipment, usually extending over hundreds of miles and embracing several communities, so operated that individual members can deliver power to the system or draw power from it, according to local demand.

power ground The power-supply ground for a circuit or system.

power-handling capacity **1.** The amount of power that a device can dissipate, either continuously or intermittently, without suffering damage. **2.** The maximum input power that can be tolerated by an amplifier transistor or tube without overheating.

power hyperbola For a semiconductor device or vacuum tube, a curve plotted from the device's current and voltage values, which provide the power value when multiplied (e.g., a 2-watt curve for the direct-current collector input of a power transistor).

power input See INPUT POWER.

power-input control The adjustment of the output of a power supply by varying the alternating-current input to the power transformer. Usually, a variable autotransformer is operated ahead of the power transformer. See, for example, VARIABLE TRANSFORMER and VARIAC.

power inverter See CHOPPER POWER SUPPLY.

power-level indicator **1.** See DB METER. **2.** See OUTPUT POWER METER.

power line The line through which electrical energy is received by a subscriber.

power-line communication Abbreviation, PLC. Carrier-current telephony or telegraphy over power lines that are common to transmitting and receiving stations. Also see WIRED WIRELESS.

power-line filter **1.** A heavy-duty radio-frequency (RF) filter inserted in the power line close to a device that generates RF energy, such as a radio transmitter. It prevents transmission of RF energy via the power line. **2.** An RF filter inserted in the power line, where it enters the power supply of a sensitive electronic device, such as a computer or high-fidelity audio amplifier. It prevents RF energy on the power line from entering the device via the power supply.

power-line frequency The frequency of the alternating current and voltage available over commercial power lines. In the United States, it is 60 Hz; in some countries, it is 50 Hz.

power-line monitor An expanded-scale alternating-current voltmeter for the continuous monitoring of power-line voltage.

power-line pickup The interception of radio-frequency energy by utility power lines acting as receiving antennas. This energy can enter a sensitive electronic device, such as a computer or a high-fidelity audio amplifier, via the power supply.

power loss The power dissipated in a component. It generates heat while doing no useful work. Represents energy loss, except when the generation of heat is the end purpose.

power-loss factor Symbol, F_p. In interstage coupling, the ratio of available power (with the coupling network in place) to the available power when the network is disconnected.

power meter See WATTMETER.

power modulation factor In amplitude modulation, the ratio of the peak power to the average power.

power oscillator A heavy-duty oscillator delivering useful power output.

power output See OUTPUT POWER.

power-output meter See OUTPUT POWER METER.

power pack An external power-line-operated unit supplying alternating or direct current for the operation of electronic equipment.

power pentode A heavy-duty pentode vacuum tube designed to deliver relatively high output power.

power plug A plug for insertion into a power-line outlet.

power programmer A device that adjusts radar output power, in accordance with the target distance.

power rating **1.** The specified power required by an equipment for normal operation. **2.** The specified power output of a generator or amplifier.

power reactive See REACTIVE VOLT-AMPERE.

power rectifier A heavy-duty semiconductor diode used to rectify alternating current for power-supply purposes.

power relay A heavy-duty relay designed to switch significant amounts of power. The heavy contacts and armature require high actuating current; this necessitates a larger coil than is used in lighter-duty relays.

power resistor A heavy-duty resistor (i.e., one designed to carry large currents without overheating).

power stack A selenium rectifier consisting of a number of rectifier plates stacked in series for higher voltage handling.

power supply **1.** A device, such as a generator or a transformer-rectifier-filter arrangement, that produces the power needed to operate an electronic equipment. **2.** A reserve of available power (e.g., the power line, an installation of batteries, etc.).

power-supply filter A low-pass filter that is used to remove the ripple from the output of a power-supply rectifier. See, for example, BRUTE-FORCE FILTER.

power-supply rejection ratio The ratio of the output-voltage change for an amplifier, oscillator, or other circuit, to the change in power-supply voltage. It is determined on an instantaneous basis.

power-supply sensitivity In an operational amplifier, sensitivity of the offset to variations in the power-supply voltage.

power surge **1.** A momentary increase in the voltage on a utility line. **2.** An abnormally high voltage that sometimes exists for the first several milliseconds after utility power is restored following a blackout.

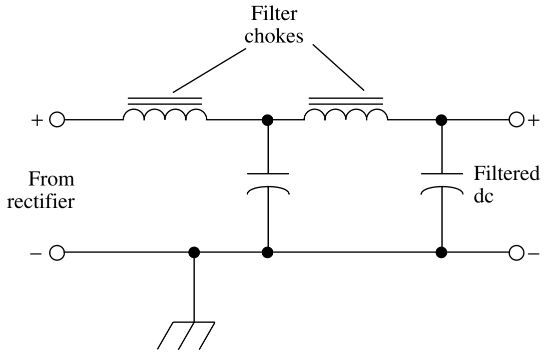

power-supply filter

power switch The switch for controlling power to a piece of equipment. Also see ON-OFF SWITCH.

power switching Switching operating power on and off. There are two principal methods: One involves making and breaking the connections between equipment and the power line; the other involves making and breaking the output of a line-operated or battery-type power supply.

power tetrode A heavy-duty tetrode vacuum tube designed to deliver relatively high output power.

power-to-decibel conversion Abbreviation, P_{dB}. Power expressed in decibels, with respect to a reference power level. Determined by the formula $P_{dB} = 10 \log_{10}(P_x/P_{ref})$, where P_x is the given power level and P_{ref} is the reference power level.

power transfer **1.** The passage of power from a generator to a load. **2.** The passage of power from one circuit to another.

power transfer theorem See MAXIMUM POWER TRANSFER THEOREM.

power transformer A transformer designed solely to supply operating power to electronic equipment—either directly or through a rectifier-filter circuit. Because a power transformer is used at low (power-line) frequencies, its core does not require the high-grade iron used in audio transformers, nor are special winding techniques needed to reduce the leakage inductance and interwinding capacitance.

power transistor A heavy-duty transistor designed for power-amplifier and power-control service.

power triode A heavy-duty triode vacuum tube designed to deliver relatively high output power.

power tube A heavy-duty electron tube designed to deliver useful amounts of power. See, for example, POWER PENTODE, POWER TETRODE, and POWER TRIODE.

power unit **1.** A power supply (see POWER SUPPLY, **1**). **2.** A unit of power measurement. See, for example, KILOWATT, MEGAWATT, MICROWATT, MILLIWATT, PICOWATT, and WATT.

power user In personal or business computing, a serious user who has extensive, up-to-date knowledge of hardware and software.

power winding In a magnetic amplifier or saturable reactor, the output winding (i.e., the winding through which the controlled current flows).

Poynting vector In an electromagnetic wave, the vector product of instantaneous electric intensity and magnetic intensity.

PP Abbreviation of *peripheral processor.*

PP **1.** Symbol for PLATE POWER. **2.** Symbol for PEAK POWER.

ppb Abbreviation of *parts per billion.*

PPI Abbreviation of PLAN POSITION INDICATOR.

pp junction In a semiconductor wafer, the boundary between two p-type regions that have somewhat different properties.

PPM Abbreviation of PULSE-POSITION MODULATION.

ppm **1.** Abbreviation of *parts per million.* **2.** Abbreviation of *pulses per minute.*

pps Abbreviation of *pulses per second.*

ppt Abbreviation of *parts per thousand.*

PPV Abbreviation of PAY PER VIEW.

Pr Symbol for PRASEODYMIUM.

practical component A circuit component considered in proper combination with the stray components inherent in it. Thus, a resistor has residual inductance and capacitance, an inductor has residual capacitance and resistance. Compare IDEAL COMPONENT.

practical units A set of physical/electrical units especially suited to a particular application. For example, in direct-current electrical applications, practical units are the AMPERE, OHM, VOLT, and WATT.

praetersonics See ACOUSTOELECTRONICS; ACOUSTIC DELAY LINE; SURFACE-WAVE AMPLIFIER; SURFACE-WAVE FILTER.

pragilbert The unit of MAGNETOMOTIVE FORCE in the absolute mks (Giorgi) system.

pragilbert per weber The unit of RELUCTANCE in the absolute mks (Giorgi) system.

praoersted The unit of MAGNETIZING FORCE in the absolute mks (Giorgi) system.

praseodymium Symbol, Pr. A metallic element of the rare-earth group. Atomic number, 59. Atomic weight, 140.908.

preaccelerating electrode In the electron gun of a cathode-ray tube, the high-voltage electrode that provides initial acceleration to the electron beam.

pre-alarm signal An audio and/or visual indicator that an alarm will sound if an area is not cleared within a short time. An example is a voice recording in a car alarm system (e.g., “Stand back!”).

preamplifier **1.** A high-sensitivity, low-noise amplifier that usually uses a high-gain field-effect transistor (FET) and is used to enhance the sensitivity of a radio communications receiver. They are especially used at frequencies above

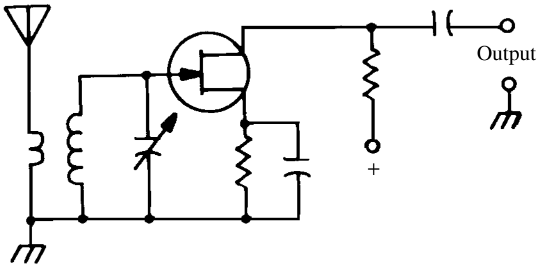

preamplifier

approximately 15 MHz, where most of the noise comes from the receiver, rather than from outside sources. **2.** A low-noise amplifier used for boosting weak signals for television reception. **3.** A low-noise, low-level amplifier used for boosting signal levels from transducers, such as microphones or photocells.

prebiased relay A relay through which is maintained a steady current that is just lower than that needed to close the relay. The actuating signal, then, need only be a small amount of additional current.

precedence effect See FUSION, **1**.

precipitation **1.** Water falling from the atmosphere in some form (rain, snow, hail, or sleet). See PRECIPITATION STATIC. **2.** The amount of precipitation occurring in a given period of time at a specific location. **3.** Separation of a solid material from a solution, as a result of a chemical or physical action.

precipitation static **1.** Radio noise that sometimes occurs when it rains or snows. It can be mistaken for artificially generated noise. **2.** Radio noise caused by atmospheric electricity arising from rain, snow, ice crystals, hail, or dust clouds, through which an aircraft carrying the radio flies.

precipitator See DUST PRECIPITATOR.

precipitron See DUST PRECIPITATOR.

precision **1.** Pertaining to electronic hardware, especially test instruments and measuring devices, designed and built to function with a high degree of accuracy. **2.** The relative accuracy of a meter or other indicating device. **3.** The accuracy of the results of an experiment, test, or measurement.

precision approach radar A radar aimed along the approach path to guide an aircraft during approach.

precision instrument An instrument possessing high accuracy and stability (i.e., one capable of reproducing readings or settings for various trials under set circumstances).

precision potentiometer **1.** A POTENTIOMETER possessing highly accurate resistance calibration, linearity, and repeatability of settings. **2.** A potentiometer-type voltage-measuring instrument.

preconduction current **1.** The cutoff current in a transistor. **2.** In a thyratron, the small (anode) current flowing before the tube is fired.

predetermined counter A counter programmed to count to a desired number and stop.

predistortion See PREEMPHASIS.

pre-Dolby **1.** To record a tape with DOLBY compression. **2.** A tape that has been recorded with Dolby compression.

preemphasis In frequency modulation, the introduction of a rising-response characteristic (response rises as modulation frequency increases). Compare DEEMPHASIS.

preequalization **1.** In sound recording and reproduction, equalization during recording. **2.** See PREEMPHASIS.

preferred values of components A number system used by the Electronics Industries Association (EIA) for establishing the values of composition resistors and small fixed capacitors.

prefix multiplier See MULTIPLIER PREFIX.

prefix notation As used with complex expressions involving many operators and operands, a type of notation in which the expressions, rather than containing brackets, are given a value, according to the relative positions of operators and operands.

preform **1.** A small wafer, usually dry-pressed from powdered plastic, from which the body of a component, such as a capacitor or resistor, is heat-molded. Also called a *pill* or *biscuit.* **2.** The preformed slab used in molding a phonograph disc. **3.** To shape a moldable circuit before fixing the final configuration or package.

preliminary information For manufactured electronic components, data that is released prior to the actual availability of the device. Subject to change when units are produced.

premix A molding compound of reinforced plastic.

prerecorded disc A phonograph disc on which a recording has been made (i.e., a recorded disc).

prerecorded tape Magnetic tape on which a program or data has been recorded. Also called RECORDED TAPE.

p region See P LAYER.

prescaler A device operated ahead of a counter to establish a new, usually higher-frequency, range over which frequency measurements can be made.

preselector A tuned or untuned radio-frequency amplifier operated ahead of a radio or television receiver to boost the sensitivity of the receiver.

presence **1.** In sound reproduction, the quality of being true to life. **2.** The effect of boosted upper-midrange frequencies in music.

preset counter A pulse counter that delivers one output pulse for a number of successive input pulses, determined by the settings of counter-circuit controls. Thus, a preset counter might give an output pulse for each train of 125 input pulses.

preset element In automation and control, an element that can be preset to a given level or value, and to which other elements can then be referred.

preset switch In the circuit of a PRESET COUNTER, a multiposition rotary switch that can be set to determine the number of input pulses that must be received for the circuit to deliver one output pulse.

preshoot A downward-moving transient pip that sometimes precedes the rise of a pulse.

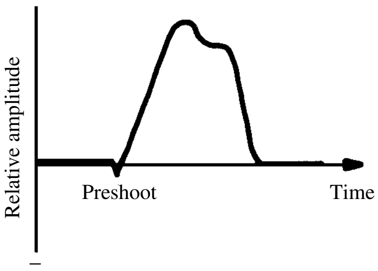

preshoot

preshoot amplitude The peak voltage of a PRESHOOT, measured from the zero line to the valley of the preshoot.

preshoot time The width of a PRESHOOT, measured along the horizontal base line (time axis).

pressing **1.** A process by which phonographic discs are fabricated from plastic. **2.** A disc pressed from plastic.

press-to-talk microphone A microphone that uses a PRESS-TO-TALK SWITCH for actuation.

press-to-talk switch A switch in a microphone or on the end of a control cord. It is used to actuate a transmitter, telephone, or recorder when the operator wishes to speak.

pressure Abbreviation, P or p. **1.** Force per unit area. It can be expressed in any appropriate units of force and area (e.g., newtons per square meter, pounds per square inch, grams per square centimeter, etc.). **2.** The application of force over part or all of a surface. **3.** Compression. **4.** See TENSION, **1**.

pressure amplitude The pressure caused by an acoustic disturbance. It is usually measured in dynes per square centimeter.

pressure capacitor An enclosed fixed or variable capacitor, whose breakdown voltage increases when the air pressure rises inside the container.

pressure contact **1.** Electrical contact made by pressing two conducting surfaces together (to complete a circuit). **2.** A contact for obtaining the condition described in **1**.

pressure-gradient microphone See PRESSURE MICROPHONE.

pressure microphone A microphone that receives sound waves at only one side of its diaphragm. This one-sided exposure results in the displacement of the diaphragm by an amount proportional to the instantaneous pressure of the sound waves.

pressure pad In a tape recorder, a small pad that holds the tape against one of the heads.

pressure pickup See PRESSURE TRANSDUCER.

pressure roller In a tape recorder, a rubber-tired roller that presses the tape against the capstan.

pressure sensor A device that detects the presence of, and/or measures, physical force within a specific area. One simple device uses two metal plates separated by electrically resistive foam. Pressure compresses the foam and reduces the resistance between the plates. This resistance change can be detected and measured.

pressure switch A switch that is opened or closed by a change in pressure within a system.

pressure transducer A sensor for converting pressure into proportionate current or voltage. Some use strain gauges; others use piezoelectric crystals, potentiometers, and other variable elements.

pressure zone A region of high air pressure that is immediately adjacent to a surface reflecting an acoustic (sound) wave.

pressure-zone microphone A microphone equipped with a deflector that helps to guide acoustic energy toward the diaphragm.

prestore To place data in memory before it is intended for use.

pretuned stage A stage, such as one in an intermediate-frequency amplifier or single-frequency receiver, that is preset to a frequency, rather than being continuously tuned.

prf Abbreviation of PULSE REPETITION FREQUENCY.

pri Abbreviation of PRIMARY.

primaries See PRIMARY COLORS.

primary **1.** See PRIMARY WINDING. **2.** See PRIMARY STANDARD.

primary battery A battery composed of primary cells.

primary block A fundamental group of channels in pulse-code modulation, combined by means of time-division multiplexing.

primary capacitance **1.** The distributed capacitance of the primary winding of a transformer whose secondary winding is unloaded. Compare SECONDARY CAPACITANCE, **1**. **2.** A series or shunt capacitance used to tune the primary coil of a radio-frequency transformer. Compare SECONDARY CAPACITANCE, **2**.

primary cell An electrochemical cell that does not require, and generally will not accept, an electrical charge in order to function. Once it has been discharged, the cell must usually be thrown away. Compare STORAGE CELL. Also see CELL, DRY CELL, and STANDARD CELL.

primary circuit **1.** The circuit associated with the primary winding of a transformer. **2.** The circuitry associated with the input to a device or system.

primary coil See PRIMARY WINDING.

primary colors See COLOR PRIMARY.

primary current The current flowing in the primary winding of a transformer. Also called TRANSFORMER INPUT CURRENT. Compare SECONDARY CURRENT.

primary electron The electron possessing the greater energy after a collision between two electrons. Compare SECONDARY ELECTRON.

primary emission Emission arising directly from a source, such as the cathode of an electron tube. Compare SECONDARY EMISSION.

primary frequency standard A device that generates unmodulated signals at precise frequencies. It generally uses a highly stable crystal oscillator that can be referred to a time standard and periodically corrected. A string of multivibrators, together with harmonic amplifiers and buffers, divide, and multiply the fundamental crystal frequency. The resulting signals provide markers for calibrating receivers and test equipment. Compare SECONDARY FREQUENCY STANDARD. Also see PRIMARY STANDARD.

primary impedance **1.** The impedance of the primary winding of a transformer whose secondary winding is unloaded. Compare SECONDARY IMPEDANCE, **1**. **2.** An external impedance presented to the primary winding of a transformer. Compare SECONDARY IMPEDANCE, **2**.

primary inductance The inductance of the primary winding of a transformer whose secondary winding is unloaded. Compare SECONDARY INDUCTANCE.

primary kVA The kilovolt-amperes in the primary circuit of a transformer. Compare SECONDARY KVA.

primary measuring element A detector, sensor, or transducer that performs the initial conversion in a measurement or control system. Such an element converts a phenomenon into a signal that can be transmitted to appropriate instruments for translation and evaluation.

primary power Power in the primary circuit of a transformer. Also see PRIMARY KVA and PRIMARY VA. Compare SECONDARY POWER.

primary radiator **1.** The driven element of a directive antenna system that incorporates parasitic elements. **2.** The driven element of a directive antenna that uses a reflector, such as a screen or dish.

primary resistance The direct-current resistance of the primary winding of a transformer. Compare SECONDARY RESISTANCE.

primary standard A usually stationary source of a quantity (e.g., capacitance, frequency, time, inductance, resistance, etc.). This source is so precise, and is maintained with such care, that it can be used as a universal reference. Compare SECONDARY STANDARD.

primary turns Symbol, N_p. The number of turns in the primary winding of a transformer. Compare SECONDARY TURNS.

primary utilization factor Abbreviation, UF_p. For a transformer in a rectifier circuit, the ratio of direct-current power output to primary volt-amperes. Numerically, the primary utilization factor is higher than the secondary utilization factor, but is less than 1. Also see SECONDARY UTILIZATION FACTOR and UTILITY FACTOR.

primary VA The volt-amperes in the input circuit of a transformer. Compare SECONDARY VA.

primary voltage The voltage across the primary winding of a transformer. Also called *transformer input voltage.* Compare SECONDARY VOLTAGE.

primary winding The normal or usual input winding of a transformer. Also called *primary coil.* Compare SECONDARY WINDING.

prime meridian See ZERO MERIDIAN.

prime mover A machine, such as a gas engine, steam engine, or water turbine, that converts a natural force or material into mechanical power.

primitive oscillation period In a complex oscillation waveform, the shortest period for which a definite repetition occurs; the highest fundamental frequency.

principal axis The line passing through the center of the spherical part of a lens, mirror, or dish reflector.

principal focus The focal point of rays arriving parallel to the principal axis of a lens, mirror, or dish reflector.

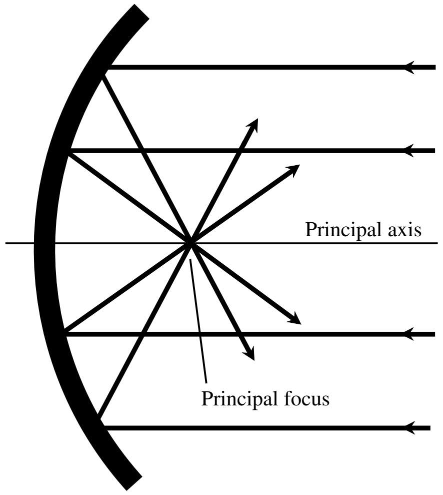

principal focus

principal mode See DOMINANT MODE.

principal ray The path described by an electron entering an electron lens parallel to the lens' axis, or by an electron leaving this lens parallel to the axis.

print **1.** The material transferred from a typewriter onto paper. **2.** The command, in a computer system, that causes data to be placed on paper or onto the output screen. **3.** The alphanumeric output of a computer or data terminal.

printed capacitor A two-plate capacitor formed on a printed circuit.

printed circuit A pattern of conductors (corresponding to the wiring of an electronic circuit) formed on a board of insulating material, such as a phenolic, by photo-etching, silk-screening of metallic paint, or by the use of pressure-sensitive preforms. The leads or pins of discrete components are soldered to the printed metal lines at the proper places in the circuit, or the components can be formed along with the conductors. Also see ETCHED CIRCUIT.

printed-circuit board A usually copper-clad plastic board used to make a printed circuit.

printed-circuit lamp A baseless lamp having flexible leads for easy soldering or welding to a printed circuit.

printed-circuit relay A usually small relay provided with pins or lugs for easy solder connection to a printed circuit.

printed-circuit switch A rotary switch whose contacts and contact leads are printed on a substrate.

printed-circuit template Also called *etching pattern.* A drawing for the purpose of making printed-circuit boards by photographic means.

printed coil A flat, spiral coil formed on a printed circuit.

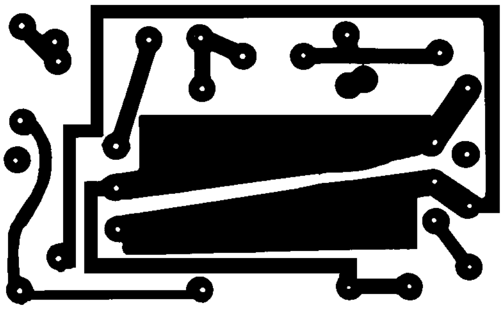

printed-circuit template

printed component A component formed on the substrate of a printed circuit. See, for example, PRINTED CAPACITOR, PRINTED COIL, and PRINTED RESISTOR.

printed display See DATA PRINTOUT, **2**.

printed element See PRINTED COMPONENT.

printed inductor See PRINTED COIL.

printed resistor A resistor printed or painted on a printed circuit.

printed wiring The printed or etched metal lines that serve as the conductors in a printed circuit.

printer In computer and calculator operations and in measurement procedures, a readout device that prints a permanent record of output data. There are several types, including the *dot-matrix printer*, the *daisy-wheel printer*, the *inkjet printer*, and the *laser printer.* Some printers, such as the daisy-wheel type, can render only text data; others, such as the laser type, can print high-resolution graphic images, sometimes in color.

printing calculator **1.** An electronic calculator that supplies a printed record of the results of a calculation. **2.** For a programmable calculator, the results, a record of program steps, and plots of curves.

printing digital voltmeter Abbreviation, PDVM. A digital voltmeter that delivers a printed record of a voltage reading, in addition to the usual digital readout of the voltage.

printing telegraph **1.** A telegraph that prints the received message on a tape or page. **2.** See TELETYPEWRITER.

printing wheel See PRINT WHEEL.

print format The form of data transmitted by a computer program to a printer (e.g., plain text, graphics, color graphics, etc.).

printout See DATA PRINTOUT, **1**, **2**.

print-through In prerecorded magnetic tape on a reel or cassette, the transfer of magnetism between layers of the rolled-up tape.

print wheel In a daisy-wheel printer, the rotatable wheel on whose rim the letters, numbers, and other symbols are inscribed in relief.

priority indicator In data transmission, a code that specifies the order of importance of a message in a group of messages to be sent.

priority processing In multiple programming operations, a system for ascertaining the order of processing for different programs.

privacy code **1.** A subaudible tone used in cordless telephone systems to reduce the chances of interference between phones operating on the same channel in close proximity. **2.** A subaudible tone used in radio transmissions, especially in conjunction with repeaters, to allow only those stations with the proper code to be received. **3.** A tone-burst sequence at the beginning of a transmission that actuates a receiver, allowing only those stations with the proper code to be received.

privacy equipment Devices, such as speech scramblers and digital encryption programs, that provide some measure of secrecy in communications.

privacy switch In a telephone amplifier, a switch (usually a pushbutton) for muting outgoing messages.

private automatic exchange Abbreviation, PAX. A dial telephone system for use within an organization and having no connection to the central office. Compare PRIVATE BRANCH EXCHANGE.

private branch exchange Abbreviation, PBX. A telephone system, complete with a private manually operated switchboard and individual telephone sets, installed and operated on private premises but having trunk-line connection to the central office. Compare PRIVATE AUTOMATIC EXCHANGE.

private line **1.** A communication circuit in which the use is limited, by electronic means, to certain subscribers. **2.** A subaudible-tone system used to restrict access to a communications system. The tone frequency is predetermined. For access to the system, a transmitted signal must contain the tone of the appropriate subaudible frequency, in addition to the voice or other information.

probability **1.** The branch of mathematics concerned with the likelihood of an event's occurrence. It has many applications in quality control and physics. **2.** The mathematical likelihood that an event will occur.

probable error Abbreviation, PE. The value of error above and below which all other error values are equally likely to occur.

probe **1.** A usually slender pencil-like implement with a pointed metal tip and a flexible, insulated lead. It is used to contact live points in a circuit under test (e.g., *voltmeter probe* and *oscilloscope probe*). **2.** A device used to sample a radio-frequency voltage or current at a desired point (e.g., WAVEGUIDE PROBE). **3.** A pickup device shaped like a probe for insertion into close quarters (e.g., PROBE THERMISTOR).

probe meter See PROBE-TYPE VOLTMETER.

probe thermistor A thermistor of slender construction for insertion into an area in which the temperature is to be monitored or controlled.

probe thermocouple A thermocouple in the form of a slender probe for insertion into close quarters for temperature sensing or temperature control.

probe tip See PROD.

probe-type voltmeter A voltmeter installed in a long probe or wand. Kilovoltmeters are sometimes constructed in this fashion, with a long multiplier resistor housed in the probe.

probing A process for locating, or determining the existence of, external artificial interference (e.g., power-line noise) in a radio communications circuit.

problem-oriented language Any high-level computer programming language that allows the user to write programs as statements in terms applicable to the field of interest (e.g., COBOL's statements in English for problems relating to business).

problem reduction In artificial intelligence, a process in which problems are made easier by breaking them down into smaller logical parts.

process control The control of a process, such as one of manufacturing, by means of computers.

processor **1.** A circuit or device used to modify a signal in response to certain requirements (e.g., *clipper* and *waveshaper*). **2.** See DATA-PROCESSING MACHINE. **3.** See CENTRAL PROCESSING UNIT. **4.** See MICROPROCESSOR.

prod The metal tip of a probe (see PROBE, **1**).

product **1.** The result of mixing or heterodyning of two or more signals. **2.** The result of modulating one signal with another. **3.** The result of combining or processing a signal or signals in a specified manner. **4.** The saleable end result of a manufacturing process. **5.** The result obtained when two or more quantities are multiplied by each other.

product detector A detector circuit whose output is the product of two signals applied simultaneously to the circuit. In a single-sideband receiver, for example, one of the signals is the incoming signal; the other, the signal from the local beat-frequency oscillator.

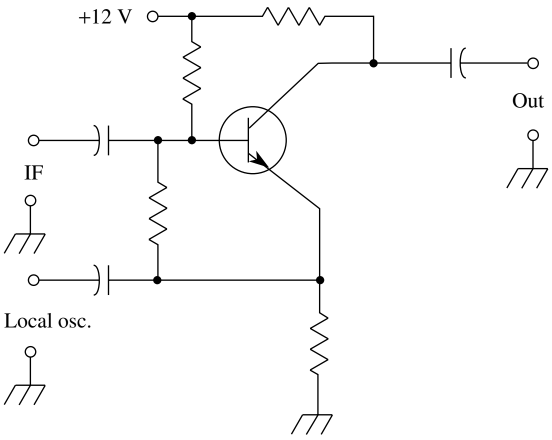

product detector

production lot A manufactured set of components, circuits, or systems, intended for sale. All of the units in the lot are identical. The finished product is suitable (presumably) for consumer use.

production unit One unit in a production lot; a finished unit, ready for use by a consumer.

product modulator A modulator whose output is equal or proportional to the product of carrier voltage and modulating voltage.

product of sine waves The result of multiplying one sine wave by another with attention being paid to the power factor. In the case of a resistive circuit, where the power factor is equal to 1, all voltage-current (*EI*) products are positive, and are equal to the true power. A product wave has negative half-cycles when the circuit contains reactance.

professional engineer A person licensed by a state board of examiners to work independently as an engineering. Also see PE and REGISTERED PROFESSIONAL ENGINEER.

program **1.** In computer operations, a detailed sequence of instructions representing an algorithm (the necessary steps in solving a problem) that can be implemented by a computer. **2.** The content of a radio or television broadcast during a specified period of time. **3.** In audio recording, the composite output from the mixer, used to make the master tape or disc.

program address counter See INSTRUCTION REGISTER.

program amplifier A broadcast preamplifier used at the studio or a remote location.

programatics The study of computer programming.

program circuit In wire telephony, a circuit capable of handling music and other audio data that covers a wide band of frequencies.

program compatibility The condition in which a program written for one computer can be used with another computer having a different architecture.

program controller In a central processor, a unit that controls the sequence and execution of program instructions.

program counter See CONTROL REGISTER.

program file A flexible reference system for software library maintenance.

program flowchart A representation of a computer program in the form of a flowchart. Each function and transition point is indicated by a box in the chart. A user can follow the flowchart and determine the outcome of the program for any given set of input parameters.

program library A collection of computer or programmable-calculator programs. Usually, it means the collection of programs used in a given computer system, often a software package supplied by the hardware vendor. It might also be a catalog of programs with instructions for their use.

programmable calculator A calculator that can be programmed to perform a chain of operations in a given order repetitively.

programmable read-only memory In a computer, a read-only memory (ROM) that can store a program.

program maintenance The ongoing correcting, updating, and modification of computer programs belonging to a system.

programmed dump A dump that occurs during a program run, according to a program instruction.

programmed halt During a computer program run, a temporary cessation resulting from an interrupt or halt instruction.

programmed instruction See MACRO INSTRUCTION.

programmed timer See CYCLE TIMER.

programmer A person who writes computer programs.

program modification **1.** In computer programming operations, a change in the effect of instructions and addresses during a program run by performing arithmetic and logical operations on them. **2.** Rewriting, or adding a patch to, a computer program. Also see PATCH, **3**.

program register See CONTROL REGISTER.

program segment A unit within a computer program that is stored with others in memory at the time of the program's execution, or sometimes, as overlays loaded individually when the entire program exceeds memory capacity.

program specification A description of the steps involved in the solution of a problem, from which a programmer devises a computer program.

program step An instruction in a computer program.

program tape In computer operations, a magnetic or paper tape that contains programs for a system or application.

program timer **1.** A programmed timer (see CYCLE TIMER). **2.** A timing unit that controls the duration of a program.

progressive scanning Non-interlaced television raster scanning, in which the lines are traced from top to bottom in succession. Conventional television broadcasting uses INTERLACED SCANNING.

progressive wave A wave disturbance that travels through a theoretically perfect homogeneous medium. This can be a compression (longitudinal) wave, a transverse wave, or an electromagnetic wave.

projected cutoff For an amplifier circuit, the operating point at which crossover distortion vanishes. The direct-current bias voltage (grid or gate) required for projected cutoff is somewhat lower than the value corresponding to conventional cutoff of plate or drain current.

project engineering A field of engineering dealing with the coordination of a complete project.

projection television Large-screen television for viewing by a relatively large group, usually accomplished via a projection tube and optical system.

projection tube A cathode-ray tube, especially a television picture tube, capable of producing a bright image that can be projected onto a large screen by means of a lens system.

projector **1.** A device that transmits a visible image onto a surface for reproduction. **2.** In general, any device that transmits a signal into space.

PROLOG Acronym for *programming in logic.* A high-level computer programming language, similar to LISP, used in artificial intelligence. The operator inputs facts and rules; the computer, in effect, derives theorems from the facts by following the logical rules.

PROM Abbreviation of PROGRAMMABLE READ-ONLY MEMORY.

promethium Symbol, Pm. A metallic element of the rare-earth group, produced artificially. Atomic number, 61. Atomic weight, approximately 145. Formerly called *illinium.*

promethium cell A radioactive battery cell using an isotope of promethium. Radioactive particles from this substance strike a phosphor, causing it to glow. Self-generating photocells then convert this light into electricity.

PROM programmer An electronic device that can store a computer program in a PROGRAMMABLE READ-ONLY MEMORY (PROM). It uses a built-in keyboard.

prompt In computer operations, a message received by an operator from an operating system or an individual program. For example, in disk operating system (DOS), it could be the statement "Bad command or file name."

prompting In computer or programmed-calculator operations, the entry of a special, required variable when the machine halts and awaits such entry.

prong See PIN.

prony brake An arrangement for measuring the mechanical power output of a rotating machine. It is a special form of friction brake consisting of a band passed around a pulley on the rotating shaft of the machine under test and held at each end by a spring balance.

propagation **1.** The extension of energy into and through space. Thus, radiant energy is *propagated* from and by its source. **2.** A phenomenon resulting from the extension of energy into and through space. Thus, radio waves can be spoken of as a *propagation.*

propagation constant For waves transmitted along a line, a number showing the effect the line has on the wave. This is a complex figure [i.e., one containing a real-number component (the attenuation constant) and an imaginary-number component (the phase constant)].

propagation delay **1.** Symbol, t_{pd}. In an integrated-circuit logic gate, the time taken for a logic signal to be propagated across the gate. **2.** In digital-circuit operation, the time required for a logic-level change to be transmitted through one or more elements.

propagation delay-power product See DELAY-POWER PRODUCT.

propagation factor The ratio E/E_o, where E is the complex electric-field strength at a point to which a wave has been propagated, and E_o is the complex electric-field strength at the point of origin. Also called *propagation ratio.*

propagation loss The path loss of an electromagnetic disturbance between the transmitting and receiving antennas.

propagation mode See WAVEGUIDE MODE.

propagation ratio See PROPAGATION FACTOR.

propagation time In digital-circuit operation, the time required for a binary bit to be transferred from one point to another in the system.

propagation velocity See VELOCITY OF PROPAGATION.

proportional action An action, such as amplification or conversion, that produces an output signal proportional to the input signal.

proportional amplifier An amplifier in which the instantaneous output amplitude is proportional to the instantaneous input amplitude.

proportional control A voltage-regulation system in which the feedback correction voltage is proportional to the output-voltage error.

proportional counter A Geiger tube having a pointed-wire (or ball-tipped-wire) anode. The voltage developed across the load resistor is proportional to the number of ions created by the radioactive particles entering the tube.

proprioceptor A set of transducers and associated circuitry that allows a computerized robot to constantly sense the positions of its end effectors, and use this data in carrying out programmed tasks.

prosodic features Variations in voice tone and emphasis that lend meaning and implication to spoken statements. It is important in advanced computer speech recognition and speech synthesis systems. Two sentences with identical wording can have greatly different meanings, depending on these factors (e.g., "*You* are!" versus "You *are?*").

prosthesis An electromechanical artificial human limb or body part. Examples: artificial legs, artificial hands, and artificial respirators. Some such devices are computer-controlled; others can be manipulated by nerve impulses.

protactinium Symbol, Pa. A relatively short-lived radioactive metallic element. Atomic number, 91. Atomic weight, 231.04 Formerly called *protoactinium.*

protected area A region to which access is restricted, and that is secured by an alarm system, surveillance cameras, or other intrusion-prevention systems.

protected location In computer storage, a location whose contents are protected from mutilation or erasure by making the location usable only by following a special procedure (e.g., using a password).

protection In a multiple processing computer system, preventing interference between data or programs.

protective bias In the final power amplifier of a radio transmitter, external direct-current bias applied to the base, gate, or grid. Prevents runaway in collector, drain, or plate current when the bias caused by the driving signal is lost.

protective capacitor A power-line bypass capacitor.

protective device **1.** A component that breaks a circuit in the event of excessive voltage or current from the power supply. **2.** A device that prevents excessive power from being delivered to a load by a driving circuit.

protective gap **1.** A spark gap connected in parallel with a component, or between a line and ground as protection against high-voltage transients and surges. **2.** A spark-gap-type lightning arrester.

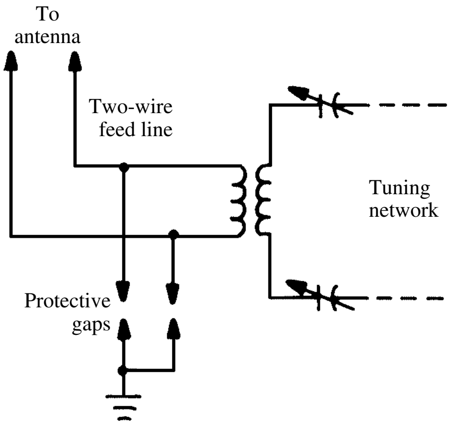

protective gap

protective resistor **1.** A bleeder resistor connected in parallel with a filter capacitor in a high-voltage direct-current power supply to discharge the capacitor automatically, thus preventing electric shock. **2.** A series resistor that limits the current going through a device.

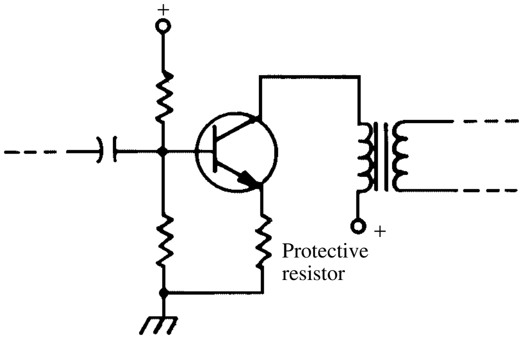

protective resistor, 2.

protector **1.** A fast-acting power-disconnect device, such as a circuit breaker or fuse, that acts to protect electronic equipment. **2.** A device or connection, such as a safety ground or ground-fault interrupter, that protects an operator from electric shock. **3.** See CONTACT PROTECTOR.

protium The light isotope of hydrogen, having an atomic mass of 1.

protoactinium See PROTACTINIUM.

protocol **1.** A set of parameters for a digital communications signal. **2.** The method by which a procedure is followed; a uniform set of governing regulations. It ensures proper operation of a system or network.

proton A positively charged particle in the nucleus of an atom. The mass of a proton is approximately 1840 times the mass of an electron.

proton rest mass See MASS OF PROTON AT REST.

proton-synchrotron A synchrotron that uses frequency modulation of the radio-frequency accelerating voltage. It can accelerate protons to energies of several billion electronvolts.

prototype The preliminary design or model of a device or system. It is often modified numerous times before the final design is attained. Compare PILOT MODEL.

proustite Crystalline silver arsenide trisulfide. Artificial crystals of this compound are used in tunable infrared-ray instruments.

proximity alarm A capacitance relay used to actuate an alerting-signal device when an area is intruded upon or a person is too close to a protected object. Also called INTRUSION ALARM.

proximity detector See PROXIMITY SENSOR.

proximity effect **1.** The influence of high-frequency current flowing in one conductor on the distribution of current flowing in an adjacent conductor. **2.** In an audio system, the result of placing a microphone too close to a person's mouth. Under these conditions, some spoken consonants (e.g., B, F, P, and T) produce clapping or booming sounds.

proximity fuse An electronic device situated in the nose of a missile. When the missile is near the target, the fuse transmits a signal that is reflected back from the target; this reflected signal detonates the missile.

proximity relay See CAPACITANCE RELAY.

proximity sensing The ability of a machine, especially a robot, to detect when an object is near. This is an aid in robot navigation because it prevents collisions. Some devices can measure the distance from a robot, or from a robotic end effector, to a nearby object.

proximity sensor A device that indicates the presence of a nearby body. Such a device uses some form of circuit, such as that of a CAPACITANCE RELAY, that changes its operating characteristics when an object enters its field.

proximity switch See CAPACITANCE RELAY.

PRR Abbreviation of PULSE REPETITION RATE.

PRV Abbreviation of PEAK REVERSE VOLTAGE.

PS Abbreviation of POWER SUPPLY.

ps Abbreviation of PICOSECOND. (Also, psec.)

PSD Abbreviation of PHASE-SENSITIVE DETECTOR.

psec Abbreviation of PICOSECOND. (Also, ps.)

pseudocode In a computer system, an instruction or code symbol that affects the operation of the programming in an indirect manner.

pseudo-instruction In computer programming operations, data representing an instruction and requiring translation by a compiler or assembler.

pseudo-offlining During input/output operations in a computer system, maximizing hardware by disconnecting slow devices from the process in question.

pseudo-operation In computer operations, an operation that, rather than being performed by hardware, is carried out by special software or by macroinstruction.

pseudo-random numbers Numbers that, although produced by a computer operating on an algorithm for their generation, are useful for an application requiring random numbers.

pseudo-stereophonic effect A somewhat heightened binaural effect obtained when two loudspeakers are situated, relative to the listener, so that a transit-time difference of 1 to 30 milliseconds results.

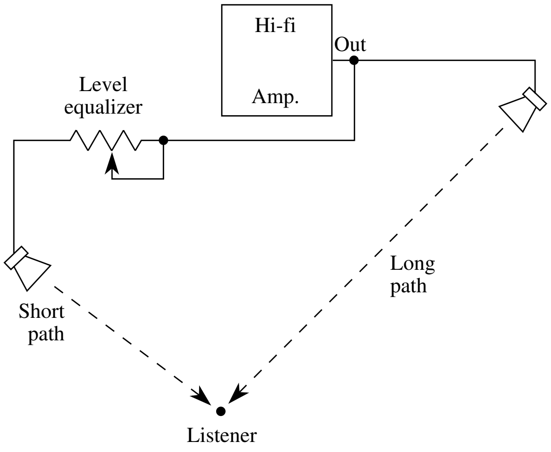

pseudo-stereophonic effect

psf Abbreviation of *pounds per square foot.* (Also, lb per sq ft, lb/ft^2.)

psi **1.** Abbreviation of *pounds per square inch.* (Also, lb per sq in, lb/in^2.)

psia Abbreviation of POUNDS PER SQUARE INCH ABSOLUTE.

psig Abbreviation of POUNDS PER SQUARE INCH GAUGE.

psi particle A massive elementary particle that represents a resonance in an electron-positron interaction.

PSK Abbreviation of PHASE-SHIFT KEYING.

PSM Abbreviation of *pulse-spacing modulation,* more commonly called PULSE-INTERVAL MODULATION.

psophometer A device used to measure noise in a wire communications system. It provides quantitative readings based on typical human observations.

psvm Abbreviation of *phase-sensitive voltmeter.*

PSWR Abbreviation of *power standing-wave ratio.*

psychoacoustics A field of acoustics, overlapping with psychology, concerned with the effects of sounds on human beings.

PT Abbreviation of *Pacific Time.*

Pt Symbol for PLATINUM.

PTC Abbreviation of POSITIVE TEMPERATURE COEFFICIENT.

PtIr Symbol for PLATINIRIDIUM.

PTM Abbreviation of PULSE-TIME MODULATION.

PTO Abbreviation of PERMEABILITY-TUNED OSCILLATOR.

PTT Abbreviation for *press-to-talk.* See PRESS-TO-TALK MICROPHONE; PRESS-TO-TALK SWITCH.

PTV Abbreviation of *public television.*

p-type conduction In a semiconductor, current flow consisting of the movement of holes. Compare N-TYPE CONDUCTION.

p-type material Semiconductor material that has been doped with an acceptor-type impurity and, consequently, conducts current mainly via hole migration. Germanium, for example, when doped with indium, becomes p-type. Compare N-TYPE MATERIAL.

p-type semiconductor An acceptor-type semiconductor (i.e., one containing an excess of holes in its crystal lattice).

PU Abbreviation of PICKUP.

Pu Symbol for PLUTONIUM.

public-address amplifier A high-gain, high-power audio amplifier designed especially for the reproduction of speech and music at large gatherings.

public-address system A system of sound reproduction especially designed for use at large gatherings indoors or outdoors. The system includes microphones, a public-address amplifier, loudspeakers, and sometimes recorders and playback devices. Also called *PA system.*

puck drive In a tape recorder, a speed-reduction system for driving the flywheel from the shaft of the (high-speed) motor. In some machines, a rubber tire mounted on the flywheel is driven, through friction, by the motor shaft. In others, an intermediate rubber-tired wheel is placed between the motor shaft and the rim of the flywheel.

puffer A meter or bridge for measuring small values of capacitance. The name comes from the spoken sound of *pF,* the abbreviation of PICOFARAD.

pulldown Descriptive of a circuit, device, or individual component used to lower the value (e.g., impedance) of a circuit to which it is connected.

pull-in current The current required to close a relay.

pulling **1.** The abnormal tendency of one circuit to cause another to slip into tune with it. This often results from coupling (intended or accidental) that is too tight. Thus, when two oscillators feed a common circuit, such as a mixer, one might pull the other into tune with itself. **2.** Lowering of a crystal frequency by an external reactance.

pull-in voltage The voltage required to close a relay.

pull switch A mechanical switch actuated by a pulling action.

pullup Descriptive of a circuit or component used to raise the value (e.g., impedance) of a circuit to which it is connected.

pulsar An extremely dense, rapidly rotating collapsed star that produces radio signals at regular intervals. The pulse frequency varies from less than one hertz to several tens or hundreds of hertz.

pulsating direct current A direct current that periodically rises and falls between zero and a maximum value (or between two positive or negative values) without changing polarity. Thus, it is possible to have either a pulsating positive current or a pulsating negative current. Also see DIRECT CURRENT.

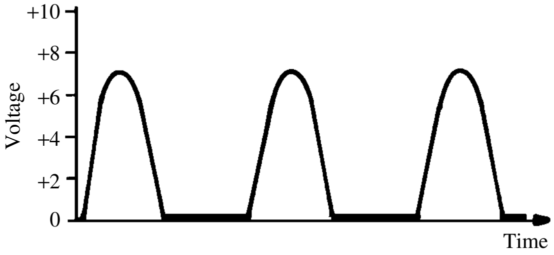

pulsating direct current

pulsating wave See PULSATING DIRECT CURRENT.

pulse A transient signal that is usually of short duration, constant amplitude, and one polarity. A typical example is a narrow positive or negative spike.

pulse amplifier An amplifier having wide frequency response and low distortion, used for amplifying steep-sided pulses of short duration.

pulse-amplitude modulation Abbreviation, PAM. A method of conveying information in wireless communications. A train of pulses is transmitted. The strength of each individual pulse varies according to the modulating waveform. Normally, the pulse amplitude increases as the instantaneous modulating-signal level increases (positive modulation). However, this can be reversed so that higher audio levels cause the pulse amplitude to go down (negative modulation).

pulse bandwidth For an amplitude pulse, the minimum bandwidth occupied. The faster the rise and/or decay times of a pulse, the greater the bandwidth. The greater the pulse frequency, the greater the bandwidth.

pulse code A code in which groups of pulses represent digits.

pulse-code modulation Abbreviation, PCM. A method of conveying information in wireless communications. A train of pulses is transmitted. The intelligence-carrying signal is sampled periodically and the amplitude is converted into binary code. The code might allow for eight levels (000 to 111), 16 levels (0000 to 1111), 32 levels (00000 to 11111), or 64 levels (000000 to 111111).

pulse-code-modulation binary code A pulse code used in communications not in the form of line transmission. Individual values are denoted by binary numbers.

pulse-code modulation multiplex equipment A multiplexer/demultiplexer for signal conversion between a single signal and multiple-channel signals. It uses both pulse-code modulation and time-division multiplexing.

pulse-count divider A circuit or device that receives an input of a certain number of pulses (or pulses per second) and delivers an output that is a function of that quantity. See, for example, DIVIDE-BY-SEVEN CIRCUIT and DIVIDE-BY-TWO CIRCUIT.

pulse counter A circuit or device that indicates the number of pulses presented to it in a given time interval.

pulse counting Counting pulses in a sequence. At low speed (pulse repetition rate), this can be done with an electromechanical dial-type counter. At high speed, a fully electronic circuit is required.

pulse delay circuit A monostable multivibrator adapted to deliver its single output pulse a predetermined time after the input pulse has been applied.

pulse dialing A form of telephone dialing in which each digit is formed by a series of pulses, usually at 10 to 20 Hz. The pulses are the equivalent of disconnecting the line for a few milliseconds. Each digit has the corresponding number of pulses, except digit 0, which is formed by 10 pulses.

pulse droop Distortion observable as a downward-sloping top on the oscilloscope trace of a pulse. It can be quantified in volts, millivolts, microvolts, amperes, milliamperes, or microamperes.

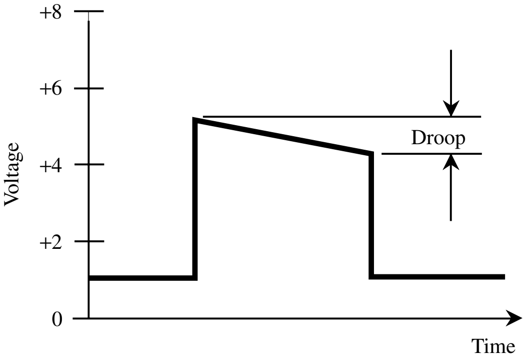

pulse droop

pulsed laser A laser in which flashes (pulses) of high-intensity light excite the lasing medium.

pulse duration The time period during which a pulse exists (i.e., its width on an oscilloscope display).

pulse-duration modulation Abbreviation, PDM. See PULSE WIDTH MODULATION.

pulse equalizer A MONOSTABLE MULTIVIBRATOR adapted to deliver pulses of equal amplitude, shape, and width—even when it receives trigger pulses of different kinds.

pulse fall time The time required for the trailing edge of a pulse to fall from 90 to 10 percent of its peak amplitude. Compare PULSE RISE TIME.

pulse-forming line A circuit used in radar for producing high-intensity pulses. Inductances and capacitances are combined in a long string, and the effect is to generate high-amplitude radio-frequency pulses.

pulse-frequency modulation Abbreviation, PFM. See PULSE-INTERVAL MODULATION.

pulse generator A signal generator that produces pulses. A general-purpose generator of this sort will produce pulses of adjustable amplitude, duration, shape, and repetition rate.

pulse-height discriminator A circuit or device that passes only pulses whose amplitudes exceed a predetermined level.

pulse interval The interval between successive pulses.

pulse-interval modulation Abbreviation, PIM. Also called *pulse-frequency modulation* (PFM) or *pulse-numbers modulation* (PNM). A method of conveying information in wireless communications. A train of pulses is transmitted. Every pulse has the same amplitude and the same duration, but their rate fluctuates with the modulating waveform. When there is no modulation, the pulses are evenly spaced with respect to time. An increase in the instantaneous data amplitude might cause pulses to be sent more often (positive modulation) or less often (negative modulation).

pulse inverter A single-stage, wideband, low-distortion, common-emitter, or common-source amplifier. The output-pulse waveforms are therefore inverted, with respect to the input-pulse waveforms.

pulse jitter In a pulse train, a disturbance characterized by random changes in the spacing between pulses.

pulse-length modulation See PULSE-DURATION MODULATION.

pulse load The load impedance for a pulse generator.

pulse mode See PULSE MODULATION.

pulse modulation See PULSE-AMPLITUDE MODULATION, PULSE-CODE MODULATION, PULSE-INTERVAL MODULATION, PULSE-POSITION MODULATION, PULSE-WIDTH MODULATION.

pulse modulator **1.** A modulator that delivers power or voltage pulses for modulating a carrier. **2.** A device that modulates pulses (see PULSE MODULATION, **2**).

pulse narrower A circuit or device that reduces the duration (width) of a pulse.

pulse-numbers modulation Abbreviation, PNM. See PULSE-INTERVAL MODULATION.

pulse operation Intermittent operation of a circuit, in the form of discrete pulses.

pulse oscillator Any oscillator with an output that consists of a series of pulses.

pulse-position modulation Abbreviation, PPM. A method of conveying information in wireless communications. A train of pulses is transmitted. The timing of each individual pulse varies according to the modulating waveform. The pulses occur earlier or later than the nominal (zero-modulation) time, depending on the instantaneous amplitude of the modulating signal.

pulse rate See PULSE REPETITION RATE.

pulse ratio The ratio of pulse height (amplitude) to pulse width (duration).

pulse regeneration Restoration of the original waveform and frequency to a pulse. It eliminates distortion caused by circuits or propagation conditions.

pulse repetition frequency Abbreviation, PRF. See PULSE-REPETITION RATE.

pulse-repetition rate Abbreviation, PRR. The number of pulses per unit time; usually pulses per second (pps).

pulse rise time The time required for the leading edge of a pulse to rise from 10 to 90 percent of its maximum amplitude. Compare PULSE FALL TIME.

pulse scaler A circuit actuated by the reception of a definite, predetermined number of input pulses.

pulse-shaping circuit **1.** A circuit for producing a pulse from a wave of some other shape (e.g., sine wave). **2.** A circuit for tailoring a pulse to desired shape, amplitude, and duration.

pulse spacing The interval between successive pulses.

pulse-spacing modulation Abbreviation, PSM. See PULSE-INTERVAL MODULATION.

pulse-steering diode In a flip-flop circuit, a diode through which the trigger pulse must pass to switch the circuit. Because of the unidirectional conductivity of a diode, pulses of only one polarity are passed.

pulse stretcher **1.** A shaping circuit that widens a pulse (i.e., increases its duration). **2.** A circuit, such as a special monostable multivibrator, that generates a pulse that is wider than the trigger pulse.

pulse stuffing See JUSTIFICATION, **2**.

pulse tilt The sloping of the normally flat top of a pulse either up or down. Also see PULSE DROOP.

pulse time See PULSE DURATION.

pulse-time modulation See PULSE-POSITION MODULATION.

pulse train A series of successive pulses of usually one kind.

pulse transformer A transformer designed to accommodate the fast rise and fall times of pulses and similar nonsinusoidal waveforms. Such transformers often use special core materials and are made using special winding techniques.

pulse transmitter **1.** A device that transmits a series of pulses. **2.** A pulse-modulated transmitter. **3.** See PULSE MODULATOR.

pulse waveform The general shape of a pulse as it appears on an oscilloscope display. The various forms range from positive or negative half-sinusoids, through rectangles, to thin-line spikes.

pulse width The horizontal dimension of a pulse (i.e., its duration).

pulse-width modulation Abbreviation, PWM. Also called *pulse-duration modulation* (PDM). A method of conveying information in wireless communications. A train of pulses is transmitted. The width (duration) of each individual pulse varies according to the modulating waveform. Normally, the pulse width increases as the instantaneous modulating-signal level increases (positive modulation). However, this can be reversed so that higher audio levels cause the pulse width to decrease (negative modulation).

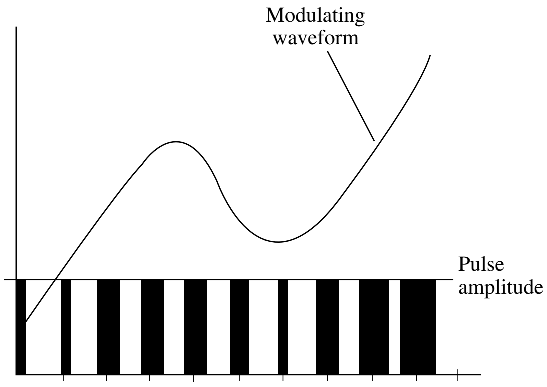

pulse-width modulation

pump **1.** In a parametric amplifier, the oscillator that supplies the signal that periodically varies the reactance of the varactor. **2.** The pumping signal in **1**. **3.** To perform the operation (pumping) described in **1**. **4.** To increase the energy level of an atom or molecule (by exposing it to electromagnetic radiation) to such an extent that oscillation or amplification occurs. A ruby laser, for example, produces its intense, coherent beam as a result of pumping. **5.** The radiation used to pump an atom or molecule. **6.** The device producing the radiation required to pump an atom or molecule.

pump frequency The frequency of a PUMP voltage.

pumping A method of laser actuation. A series of pulses, at the resonant frequency of the lasing material, is injected to cause laser output.

pump oscillator An oscillator for producing a pump voltage.

pump voltage The voltage of a pumping signal. Also see PARAMETRIC AMPLIFIER and PUMP, **1**, **2**.

punch **1.** A tool for cutting holes in metal chassis, panels, and boxes for electronic equipment. **2.** High signal strength.

punch-in editing In audio recording, a feature that allows convenient insertion of new material on a tape. The tape recorder can be switched instantly from Play to Record, and back again, whenever the operator wants to add material.

punchthrough In a bipolar transistor, the potentially damaging condition resulting when the reverse bias of the collector is increased to a voltage high enough to spread the depletion layer entirely through the base. This tends to effectively connect the emitter to the collector.

punchthrough region The conduction region associated with higher-than-punchthrough voltage, in which bipolar-transistor current is excessive. Also see PUNCHTHROUGH.

punchthrough voltage The voltage that causes PUNCHTHROUGH in a given bipolar transistor.

puncture voltage See BREAKDOWN VOLTAGE, **1**.

Pupin coil One of several loading coils that can be inserted at intervals in series with a telephone line to cancel line-capacitance effects and, thus, improve the clarity of speech.

pure tone An audio-frequency (AF) tone having essentially no harmonic content; a sine-wave AF tone.

pure wave A wave containing no distortion products.

purging The removal of an undesired gas or other substance from a system by introducing a material to displace it.

purifier A power-line operated alternating-current electromagnet that can be manually rotated in front of a color-television picture tube to demagnetize the tube. Also called a DEGAUSSER.

purity **1.** In color television, complete saturation of a hue. **2.** In a waveform, complete freedom from distortion. **3.** The extent to which spurious signals are attenuated in the output of a radio or television transmitter. Also called *spectral purity.*

purity adjustment In a color-television picture tube, adjustment of each purity control for pure color.

purity coil A variable-current coil around the neck of a color-television picture tube that is used to adjust color purity.

purity control For a purity coil, the variable resistor that controls the current for color correction.

purity magnet A ring-magnet collar around the neck of a color-television picture tube to adjust, by rotation, color purity.

purple plague Corrosion that occurs when aluminum and gold are placed in contact.

pushbutton tuner A radio or television tuner utilizing pushbutton tuning.

pushbutton tuning The tuning of a circuit to various frequencies in single steps by means of pushbutton switches.

pushdown list In data processing, a method of amending a list. A new item entered at the top moves each existing item one position down.

pushdown stack Also called *first-in/last-out.* A digital read-write memory in which data bits emerge in reverse sequence from the order they go in. If data bit *x* enters the pushdown stack before data bit *y*, then *x* will come out after *y*. Compare FIRST-IN/FIRST-OUT.

push-in terminal A circuit contact or tie point, usually of thin, springy material, that can be inserted into a hole in a perforated board.

push-pull Pertaining to a circuit in which two active devices are used, with the inputs and outputs both placed in phase opposition. In the output circuit, even harmonics are canceled, and odd harmonics are reinforced.

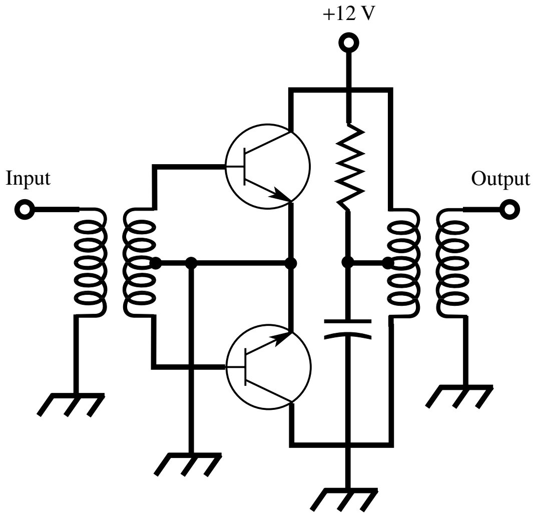

push-pull

push-pull amplifier An amplifier stage in which, for increased power output, two active devices are operated 180 degrees out of phase with each other in opposite halves of a symmetrical circuit. Also see PUSH-PULL CIRCUIT.

push-pull circuit A symmetrical circuit in which two active devices operate on separate halves of the input-signal cycle and deliver a combined output signal.

push-pull deflection In an oscilloscope, the application of deflection voltage to a pair of deflecting plates 180 degrees out of phase with each other. For this purpose, the output amplifier in the horizontal or vertical deflection channel is a push-pull stage.

push-pull doubler See PUSH-PULL MULTIPLIER.

push-pull microphone A set of two microphones, in which the audio-frequency outputs are in phase opposition.

push-pull multiplier A push-pull amplifier with its output circuit tuned to an odd-numbered multiple of the input frequency. This circuit is unsuitable for even-harmonic operation, but has some merit as an odd-harmonic multiplier (e.g., a *tripler* or *quintupler*). Also see PUSH-PUSH MULTIPLIER.

push-pull oscillator An oscillator stage in which, for increased power output, two active devices are operated 180 degrees out of phase with each other in opposite halves of a symmetrical circuit. Also see PUSH-PULL CIRCUIT.

push-pull/parallel amplifier An amplifier stage in which tubes or transistors are connected in push-pull/parallel for increased power output. Also see PARALLEL-COMPONENT AMPLIFIER, PUSH-PULL AMPLIFIER, and PUSH-PULL/PARALLEL CIRCUIT.

push-pull/parallel circuit A push-pull circuit in which two or more active devices are connected in parallel on each side of the circuit. This arrangement gives increased power output over that of the conventional push-pull circuit. See, for example, PUSH-PULL/PARALLEL AMPLIFIER and PUSH-PULL/PARALLEL OSCILLATOR.

push-pull/parallel oscillator An oscillator stage in which active devices are connected in push-pull/parallel for increased power output. Also see PARALLEL-COMPONENT OSCILLATOR, PUSH-PULL OSCILLATOR, and PUSH-PULL/PARALLEL CIRCUIT.

push-pull recording A type of film sound track consisting of two side-by-side images 180 degrees out of phase with each other.

push-pull transformer A transformer having a center-tapped winding for operation in a push-pull circuit.

push-push Pertaining to a circuit in which two active devices are used, with the inputs connected in phase opposition, and the outputs connected in parallel. The result is reinforcement of the even harmonics, and cancellation of the fundamental frequency and all odd harmonics.

push-push circuit See PUSH-PUSH MULTIPLIER.

push-push multiplier An amplifier circuit containing two active devices with their inputs connected in phase opposition and their outputs connected in parallel. This circuit is unsuitable for fundamental-frequency and odd-harmonic operation,

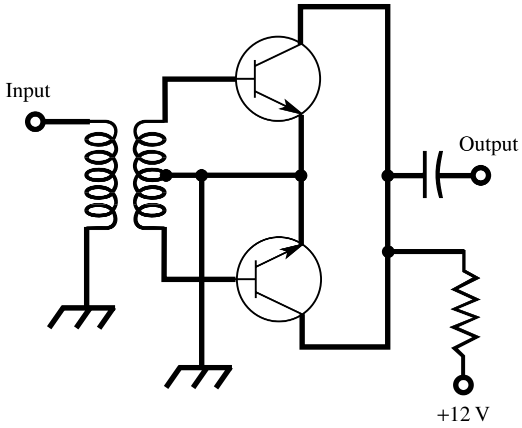

push-push

but has some merit as an even-harmonic multiplier (e.g., a *doubler* or *quadrupler*). Also see PUSH-PULL MULTIPLIER.

push-to-talk switch See PRESS-TO-TALK SWITCH.

pushup list In data processing, a method of amending a list, whereby new items are added at the end of the list; all other items retain their original positions. Compare PUSHDOWN LIST.

pV Abbreviation of PICOVOLT.

PVC Abbreviation of POLYVINYL CHLORIDE.

pW Abbreviation of PICOWATT.

PWM **1.** Abbreviation of PULSE-WIDTH MODULATION. **2.** Abbreviation of PLATED-WIRE MEMORY.

pwr Abbreviation of POWER.

Pyralin See CELLULOSE NITRATE.

pyramidal horn antenna A rectangular horn antenna that is flared in two dimensions. The horn width and height both increase linearly with increasing distance (in the direction of maximum radiation/response) from the feed point.

pyramidal wave See BACK-TO-BACK SAWTOOTH.

Pyrex A heat-resistant glass having numerous applications in electronics and chemistry.

pyrheliometer An instrument used to measure infrared radiation.

pyroelectricity In certain crystals, electricity generated by temperature change, and in particular, by the application of heat.

pyroelectric lamp See NERNST LAMP.

pyroelectric material A crystalline material that generates an output voltage when it is heated.

pyrolysis The process whereby heat changes a substance into one of several different substances by rearranging its atoms.

pyromagnetic effect In a material or circuit, the combined effect of heat and magnetism.

pyrometer An instrument, other than a thermometer, used for the measurement of temperature. See, for example, OPTICAL PYROMETER.

Pythagorean scale A sound scale defining a specific type of relationship among audio tones. If x and y are related by the Pythagorean scale and are adjacent in frequency, then a specific frequency (f) exists, so $x = f^2$ and $y = f^3$.

Pythagorean theorem A theorem of plane geometry. For a right triangle, with sides of lengths a, b, and c, where c is the side opposite the right angle, it is always true that $a^2 + b^2 = c^2$.

p-zone See P LAYER.

PZM Abbreviation of PRESSURE-ZONE MICROPHONE.

PZT Abbreviation of LEAD ZIRCONATE TITANATE.

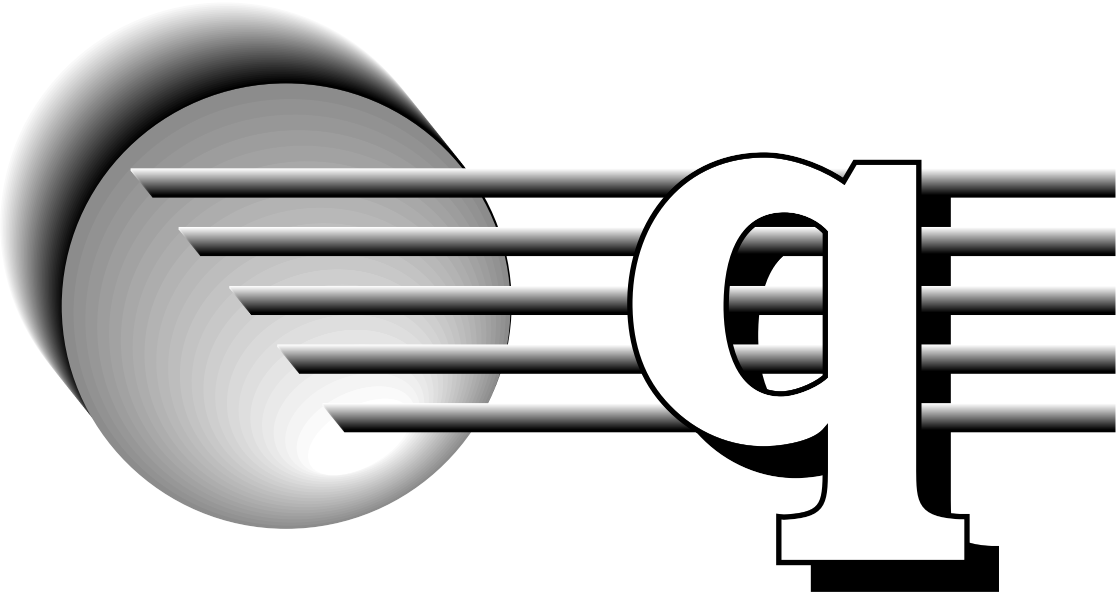

Q **1.** The figure of merit of a capacitor, inductor, or inductance-capacitance (LC) circuit, equal to the reactance divided by the resistance. **2.** Symbol for electrical charge. **3.** Occasional symbol for SELECTIVITY. **4.** See Q BAND. **5.** See *Q* OUTPUT.

q **1.** Symbol for quantity of electricity (in coulombs). **2.** Symbol for the charge carried by an electron (the charge carried by a hole is represented by –q). **3.** Symbol for the value of a quantum.

QA Abbreviation of QUALITY ASSURANCE.

***Q* adjustment** The separate null adjustment for the *Q* value of a component being tested in an impedance bridge having separate resistive and reactive balances.

***Q*-antenna** An antenna in which the transmission line (feeder) is matched in impedance to the center of the radiator by means of a *Q*-matching section.

QAVC Abbreviation of QUIET AUTOMATIC VOLUME CONTROL.

Q band The radio-frequency band 36 to 46 GHz. It is subdivided as follows: Q_a, 36 to 38 GHz; Q_b, 38 to 40 GHz; Q_c, 40 to 42 GHz, Q_d, 42 to 44 GHz; and Q_e, 44 to 46 GHz.

***Q* bar** One of the parallel metal tubes in a *Q*-matching section. Also see *Q* ANTENNA.

***Q* booster** See *Q* MULTIPLIER.

***Q* bridge** An alternating-current bridge used principally to determine the *Q* of capacitors and inductors. Bridges are usually used for audio-frequency *Q* determinations; resonant-type *Q* meters are generally used for measurement of radio-frequency *Q*.

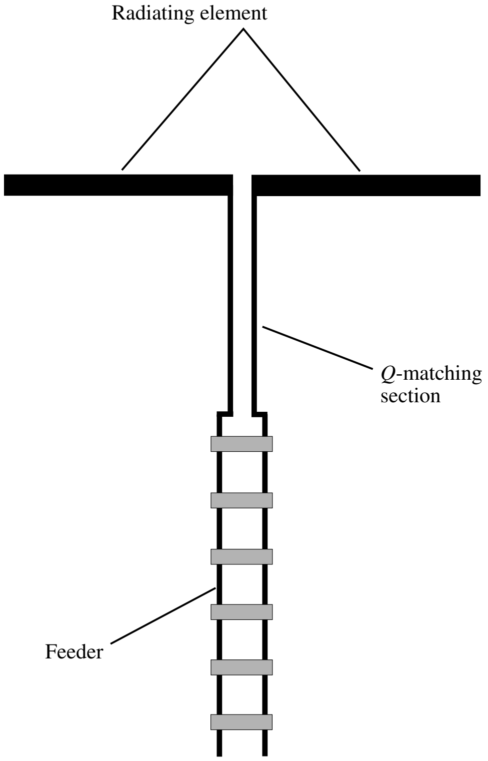

***Q*-antenna**

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

QC Abbreviation of QUALITY CONTROL.

QCE Abbreviation of QUALITY-CONTROL ENGINEERING or QUALITY-CONTROL ENGINEER.

Q channel In American (NTSC) color television, the 508-kHz-wide green-magenta color information transmission band.

QCT Abbreviation of QUALITY-CONTROL TECHNICIAN.

QCW In the local oscillator and associated circuitry of a color television receiver, a 3.85-MHz CW signal of Q PHASE.

QCW signal In a color television receiver, the component of the chrominance signal that is 90 degrees out of phase with the in-phase component.

Q demodulator In a color television receiver, the demodulator that combines the chrominance signal and the color-burst oscillator signal to recover the Q signal (see Q SIGNAL, **2**).

QED **1.** Abbreviation of QUANTUM ELECTRODYNAMICS. **2.** Abbreviation of *quod erat demonstrandum*, Latin for "which was to be demonstrated." Also, Q.E.D. Often written at the conclusions of valid logical proofs and derivations.

***Q* factor** See *Q*.

QFM Abbreviation of QUADRATURE MODULATION.

QM Abbreviation of QUADRATURE MODULATION.

***Q*-matching section** A linear radio-frequency impedance-matching transformer consisting of two straight, rigid, parallel metal conductors that are used to match a feeder to an antenna. The section is ¼ wavelength long at the operating frequency. The diameters and center-to-center spacing of the conductors are such that the characteristic impedance of the matching section is equal to the geometric mean of the feeder characteristic impedance and the radiation resistance of the radiator. Also see *Q* ANTENNA and QUARTER-WAVELENGTH MATCHING STUB.

***Q* meter** A usually direct-reading instrument for determining the *Q* of a capacitor, inductor, or inductance-capacitance (LC) circuit. Most *Q* meters are operated at radio frequencies, but audio-frequency instruments are available.

***Q* modulation** Amplitude modulation obtained by varying the effective *Q* of a radio-frequency tank circuit in step with a modulating component. See ABSORPTION MODULATION.

QMQB Abbreviation of *quick make/quick break*.

***Q* multiplier** A positive-feedback (regenerative) amplifier that increases the effective *Q* of a tuned circuit, and thereby sharpens its response, when its input is connected across the tuned circuit.

Q output The reference output of a flip-flop.

Q phase A color-television carrier signal that is 147 degrees out of phase with the color subcarrier.

Q point The point or points at which a load line intersects a device characteristic (such as the collector curve of a transistor or plate curve of a tube) and that identifies the quiescent operating point.

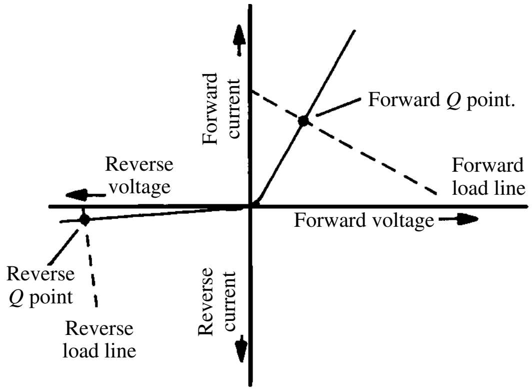

***Q* points on diode characteristic**

QRD Abbreviation of QUADRATIC RESIDUE DIFFUSOR.

***Q*-section transformer** See *Q*-MATCHING SECTION.

Q signal In color television, the quadrature component of the chrominance signal, equal to +0.48(R-Y) and to +0.41(B-Y), where B is the blue camera signal, R is the red camera signal, and Y is the luminance signal.

Q signals A set of three-letter groups, each beginning with the letter Q, used for simplified telegraph and radiotelegraph communication, and sometimes rapid voice communication (in radiotelephony). Each signal represents a commonly used phrase or message.

QSL card A card verifying communication with, or the reception of signals from, the station sending the card. Such verification is common in the amateur radio service and with some shortwave broadcast and CB stations.

QSO Amateur radiotelegraph abbreviation for TWO-WAY COMMUNICATION.

Q spoiler A device or circuit that produces Q SPOILING in a laser.

Q spoiling The technique of inhibiting laser action during an interval when an ion population excess is pumped up. When the laser is subsequently triggered by Q switching, a more powerful pulse of light results than would be otherwise obtained.

Q switching A laser-switching action obtainable with Kerr cells or rotating reflecting prisms, which consists of holding the Q of the laser cavity to a low value during an ion-population buildup, then abruptly switching the Q to a higher value.

***Q* transformer** See *Q*-MATCHING SECTION.

qty Abbreviation of QUANTITY.

quad **1.** A combination of four components, such as diodes, transistors, etc., in a single housing. The components are usually carefully matched. **2.** In a cable, a combination of four separately insulated conductors (sometimes, two twisted pairs) twisted together. **3.** Abbreviation of QUADRANT. **4.** See QUAD ANTENNA. **5.** See QUADROPHONIC.

quad antenna A directional antenna similar to the Yagi, except that full-wavelength loops are used instead of straight elements. A two-element array can consist of a driven element and a reflector, or it can have a driven element and a director. A three-element system has one driven element, one director, and one reflector. The director has a perimeter of about 0.97 electrical wavelength, the driven element measures exactly 1 electrical wavelength around, and the reflector has a perimeter of about 1.03 electrical wavelength. Additional director elements can be added to the basic three-element quad design to form arrays having any desired numbers of elements. The gain and directivity increase as the number of elements increases. Compare YAGI ANTENNA.

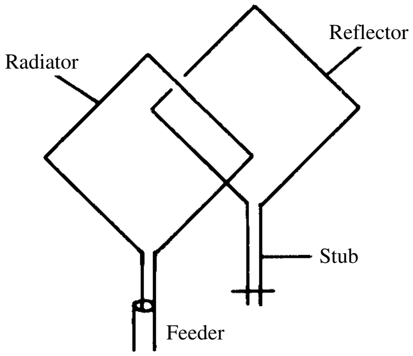

quad antenna

quadded cable See QUAD, **2**.

quadding Redundancy obtained by connecting components in series-parallel for enhanced reliability and/or increased power-handling capacity.

quad latch A set of four interconnected flip-flops that is used for digital data storage.

quadrant **1.** A specific 90-degree arc of a circle. **2.** One of the four parts formed on a plane surface by rectangular coordinates and designated I, II, III, and IV in a counterclockwise direction, starting with the upper-right quadrant. **3.** An altitude-measuring instrument.

quadrantal deviation The part of magnetic-compass deviation caused by the induction of transient magnetism into the horizontal soft iron of a vessel by the horizontal component of terrestrial magnetism.

quadrantal error See QUADRANTAL DEVIATION.

quadrant electrometer An electrometer whose principal parts are quadrants (a pillbox-shaped brass chamber split into four parts) and a needle (a flat, bowtie-shaped aluminum vane) suspended by a platinized quartz fiber between the quadrants.

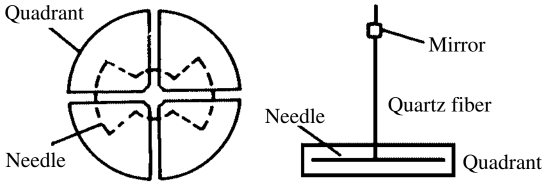

quadrant electrometer

quadraphonic sound Sound recording and reproduction involving four channels.

quadraphony Music recording or playback in which four distinct information channels are used. Also called *four-channel stereo.*

quadratic equation A second-degree equation [i.e., one in which the highest exponent is 2 (the square of an unknown) (e.g., *ax2* + *bx* + *c* = 0)].

quadratic residue diffusor Abbreviation, QRD. In acoustics, a sound-reflection grating that scatters (diffuses) sound waves almost uniformly in all directions. The depths of the indentations in the grating are determined according to a QUADRATIC EQUATION.

quadrature The state of (cyclic events or points) being 90 degrees out of phase.

quadrature amplifier An amplifier circuit that introduces a 90-degree phase shift. Such amplifiers are used in control devices, test instruments, transmitters, and color television receivers.

quadrature axes The vertical axes in the complex-number plane (i.e., the +*j* and –*j* axes).

quadrature carrier See Q PHASE.

quadrature component **1.** The reactive component of an alternating current or voltage. **2.** A vector perpendicular to a reference vector. **3.** The imaginary-number component in a complex-number expression.

quadrature current Reactive current in an alternating-current circuit.

quadrature modulation In-phase modulation of two carrier components having a 90-degree phase difference.

quadrature number See IMAGINARY NUMBER.

quadrature-phase subcarrier signal See QCW SIGNAL.

quadrature portion In color television, the portion of the chrominance signal having the same (or opposite) phase as that of the Q-signal-modulated subcarrier, and that is 90 degrees out of phase with the in-phase portion.

quadrature sensitivity The sensitivity of a transducer to motions in a direction that is perpendicular to the normal axis of response.

quadrature voltage A voltage 90 degrees out of phase with another (reference) voltage.

quadrilateral **1.** Pertaining to an object having four sides. **2.** A four-sided plane polygon.

quadrillion The number 1,000,000,000,000,000 (10^{15}).

quadripartite Having four parts.

quadripole network A four-terminal network, usually with input- and output-terminal pairs.

quadrivalent Having a valence of 4. Tin, for example, is quadrivalent. Also called TETRAVALENT.

quadruped robot A robot with four legs that can move independently. It offers better stability than three-legged designs. Functions well in mobile machines that must navigate irregular terrain.

quadrupler **1.** A rectifier circuit that delivers a direct-current output voltage approximately equal to four times the peak value of the alternating-current input voltage. **2.** An amplifier or other circuit that delivers an output signal of four times the frequency of the input signal.

quadruplex circuit A data circuit in which two messages are carried in each direction simultaneously.

quadrupole **1.** A combination of two dipoles, producing a force that varies inversely with the fourth power of distance. **2.** A four-pole magnet used in some synchrotrons and linear accelerators to focus and bend a particle beam. **3.** A system consisting of two dipoles of equal and opposite direct moment.

qualification The quality-control or quality-assurance scheme used in the production of components, circuits, or systems. Certain minimum requirements must be met for a device to obtain qualification.

qualitative test A test performed to determine the general mode of operation or the presence of certain factors, without regard to numerical values. Compare QUANTITATIVE TEST.

quality **1.** In audio-frequency applications, fidelity of transmission or reproduction. **2.** The degree of conformity of a product to specifications.

quality assurance The outcome of measures taken to bring performance into conformity with specifications. See QUALITY, **2**.

quality control The surveillance of selection, manufacturing, and testing operations to ensure conformity of a product to specifications.

quality-control engineering The branch of engineering concerned principally with the technical methods of quality control and statistical methods of assessing quality (see QUALITY, **2**).

quality-control technician A technician whose principal duty is the performance of operations in the areas of incoming inspection, manufacturing support, and product testing. Sometimes statistical evaluations are required.

quality engineering A field of engineering that deals with quality assurance and quality control in the production of components, circuits, and systems.

quality factor See *Q*.

quality-factor bridge See *Q* BRIDGE.

quality-factor meter See *Q* METER.

quanta Plural of QUANTUM.

quantimeter An instrument used to measure the quantity of X rays to which a body has been exposed.

quantitative test A test performed to determine the numerical values (and their relationships) connected with observable phenomena. Compare QUALITATIVE TEST.

quantity **1.** A parameter (e.g., collector current, grid voltage, etc.). **2.** In calculations, a positive or negative real number. **3.** Electrical charge, usually specified or measured in coulombs (see COULOMB). Also called *electrical quantity.*

quantization The conversion of a quantity having infinitely many possible values or levels (such as an analog signal) into one that can attain only a finite number of defined values or levels (such as a digital signal). The number of levels is usually some integral power of 2 (i.e., 2, 4, 8, 16, 32, etc.). This allows the levels to be represented as binary numbers.

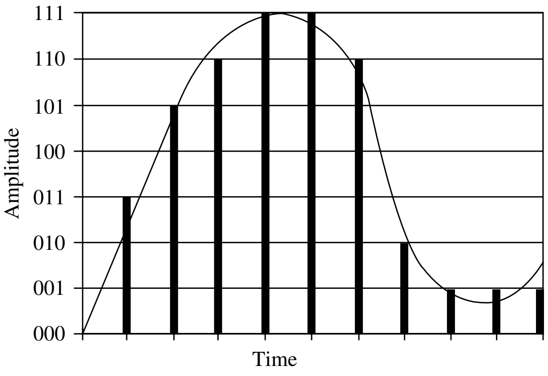

quantization

quantization distortion Distortion introduced by the process of QUANTIZATION in a communications or broadcast signal.

quantization distortion power The level of the distortion in a signal resulting from QUANTIZATION. It is expressed in microwatts, milliwatts, or watts. It can also be expressed as a percentage, or as a level in decibels, relative to the power level of the input signal.

quantization error The difference between the actual values of quantities and their quantized values.

quantization noise Noise introduced by the process of QUANTIZATION in a communications or broadcast signal.

quantize **1.** To perform the process of QUANTIZATION. **2.** To split a quantitative commodity, such as energy into its smallest measurable incremental units.

quantized pulse modulation Pulse modulation involving QUANTIZATION. Examples are PULSE-CODE MODULATION and PULSE-NUMBERS MODULATION.

quantizer A circuit or device that selects the digital subdivision into which an analog quantity is placed (e.g., an analog-to-digital converter).

quantizing See QUANTIZATION.

quantometer An instrument for measuring magnetic flux.

quantum **1.** Abbreviation, Q. Plural, *quanta.* In physics, the elemental unit or particle of electromagnetic energy. The energy contained in one such particle is directly proportional to the frequency, and inversely proportional to the wavelength. **2.** See PHOTON. **3.** Any discrete unit derived by QUANTIZATION.

quantum chromodynamics A term coined by Professor Murray Gell-Mann for the theory of *quarks* and *gluons.*

quantum counter A radiation-counter tube with a window for the admission of light to the cathode.

quantum efficiency See QUANTUM YIELD.

quantum electrodynamics A branch of quantum mechanics that involves the motions of electrons, photons, and muons caused by electromagnetic action. Quantum electrodynamics takes relativistic effects into account.

quantum electronics The branch of electronics concerned with energy states in matter.

quantum equivalence The principle that one electron is emitted for each photon absorbed by a material (when the photon has the necessary energy).

quantum jump The abrupt movement of a particle from one discrete energy state to another.

quantum level The orbit or ring occupied by an electron in an atom.

quantum mechanics A branch of physics concerned with the behavior of matter and energy, on the basis of observable data.

quantum noise A noise signal arising from random variations in the average rate at which quanta impinge upon a detector.

quantum number A number that describes the energy level, or change in energy level, for a particle.

quantum statistics A branch of QUANTUM MECHANICS concerned with the distribution of elementary particles through various quantized energy levels.

quantum theory The theory that the emission or absorption of energy by atoms or molecules occurs in discrete packages or units, rather than continuously. Each unit is the emission or absorption of an energy packet called a QUANTUM. Thus, radiant energy is thought to be divided into *quanta.*

quantum transition The movement of an electron from one energy level to another within an atom.

quantum yield The photoelectric efficiency of a light-sensitive surface in terms of the number of electrons emitted for each absorbed quantum of light.

quark A hypothetical particle having a fractional electrical charge; quarks are thought to be constituents of other subatomic particles.

quarter-deflection method A method of measuring high-frequency resistance, involving the use of a sine-wave signal source, a standard noninductive variable resistor, and a square-law radio-frequency ammeter.

quarter-phase See TWO-PHASE.

quarter wave **1.** The length of time corresponding to 90 electrical degrees in a wave disturbance. **2.** The distance in space, or along a wire or feed line that corresponds to 90 electrical degrees in a wave disturbance.

quarter-wave antenna An antenna in which the radiator is an electrical quarter-wavelength long at the operating frequency.

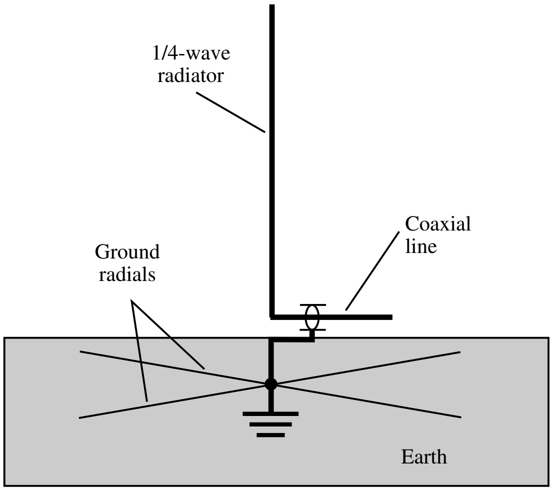

quarter-wave antenna

quarter-wave attenuator In a transmission line or waveguide, two energy-absorbing structures separated by an odd number of quarter wavelengths so that the reflection from one structure is canceled by that from the other.

quarter-wave balun A balun using quarter-wave elements. One form of this device consists of a grounded quarter-wavelength-long cylinder closed at one end and open at the other, for matching an unbalanced low-impedance line to a balanced high-impedance line.

quarter wavelength Symbol, $\lambda/4$. The distance that corresponds to 90 degrees of phase as an electromagnetic (EM) field is propagated. In free space, it is related to the frequency by a simple equation

$$L_{ft} = 246/f$$

where L_{ft} represents $\lambda/4$ in feet, and f represents the frequency in megahertz. If $\lambda/4$ is expressed in meters, then the formula is

$$L_m = 75/f$$

where L_m represents the displacement in meters. In general, if v is the velocity factor (expressed as a ratio) in a given medium, then:

$$L_{ft} = 246v/f$$

and

$$L_m = 75v/f$$

Compare FULL WAVELENGTH, HALF WAVELENGTH.

quarter-wavelength line A transmission line or feeder that is a quarter-wavelength long at the operating frequency. Also called *quarter-wave line.*

quarter-wavelength matching stub An arrangement consisting of a quarter-wavelength-long parallel-wire section of transmission line, used for matching the impedance of a nonresonant feeder to that of an antenna. It is similar to a *Q*-MATCHING SECTION, except that flexible transmission line (e.g., ladder line) is used, rather than rigid metal rods or tubing.

quarter-wave monopole A nondirectional UHF vertical antenna requiring no ground. The radiator is ¼ wavelength long, and is an enlarged-diameter outer sleeve connected to the outer conductor of the coaxial feeder. The two sections simulate a half-wave antenna.

quarter-wave plate A plate of double-refracting crystalline material whose thickness allows the introduction of a quarter-cycle phase difference between the ordinary and extraordinary components of light transmitted by it.

quarter-wave radiator An antenna consisting of a single, usually straight, active element that measures an electrical quarter wavelength from end to end. When operated against electrical ground, the element exhibits resonance. A simple quarter-wavelength ($\lambda/4$) conductor with a high length-to-diameter ratio measures approximately 95 percent of $\lambda/4$ in free space. The element can be much shorter than free-space $\lambda/4$ yet remain resonant when operated against electrical ground if inductance is inserted in series with the radiator. The element can be much longer than free-space $\lambda/4$ yet remain resonant when operated against electrical ground if distributed capacitances are inserted in series with the radiator.

quarter-wave resonance Resonance at the operating frequency in a quarter-wave antenna.

quarter-wave resonant line A section of transmission line (such as open-wire line or coaxial cable) that is a quarter-wavelength long at the operating frequency. Such a section is useful in impedance matching and in various radio-frequency tests and measurements.

quarter-wave stub See QUARTER-WAVE TRANSFORMER.

quarter-wave support In a coaxial line, a quarter-wave metal stub that can be used, instead of an insulator, to separate the inner and outer conductors.

quarter-wave termination In a waveguide, a set of two metal barriers separated by 90 electrical degrees. One barrier totally reflects the energy striking it. The other barrier allows some energy to pass through. Resonance occurs in the space between the barriers.

quarter-wave transformer A quarter-wave resonant line short-circuited at one end by an adjustable slider. This arrangement is useful for radio-frequency impedance matching.

quarter-wave transmission line See QUARTER-WAVE LINE.

quartic equation A fourth-degree equation of the form $ax^4 + bx^3 + cx^2 + dx + e = 0$, where a, b, c, d, and e are constants. Also called BIQUADRATIC EQUATION.

quartz A mineral that is a variety of natural silicon dioxide, or an artificially grown material of the same sort. In the natural state, quartz occurs in hexagonal crystals having pyramidal ends. It has various uses in electronics; one of the most common is the manufacture of piezoelectric crystals.

quartz bar A comparatively large, thick piezoelectric quartz plate used in standard-frequency oscillators and in sharply tuned low-frequency filters. Common resonant frequencies are 50 kHz, 100 kHz, and 1000 kHz.

quartz crystal A natural or artificial piece of quartz cut to specific dimensions, usually self-contained in a solder-in or plug-in enclosure. The device acts as a highly stable selective circuit. It exhibits a sharp resonance at the frequency for which it is cut, and at harmonics of this frequency. It is used as the frequency-determining element in precision oscillators.

quartz-crystal oscillator See CRYSTAL OSCILLATOR.

quartz-crystal resonator See CRYSTAL RESONATOR.

quartz delay line An acoustic delay line using quartz to transmit the sound waves.

quartz-fiber electroscope An electroscope using a gold-plated quartz fiber, instead of gold leaves.

quartz-halogen lamp An incandescent, usually low-voltage lamp used in automotive headlights, and in some home and office lighting appliances. It provides greater efficiency than conventional incandescent lamps.

quartz lamp A mercury-vapor lamp with a transparent quartz (instead of glass) envelope. Unlike glass, quartz readily passes the ultraviolet rays generated by the mercury discharge.

quartz lock A circuit that uses a CRYSTAL OSCILLATOR to regulate frequency, timing, or speed. It is used in electronic clocks and watches, television receivers, synthesized radio receivers, transmitters, transceivers, high-fidelity turntables, etc.

quartz oscillator See CRYSTAL OSCILLATOR.

quartz plate A piezoelectric plate cut from a quartz crystal. The plate is itself often called a *crystal.* Also see CRYSTAL AXES and CRYSTAL CUTS.

quartz resonator See CRYSTAL RESONATOR.

quartz timepiece A watch or clock having as its control element a time-determining quartz crystal.

quasi- A prefix meaning "to some extent" or "similar to," as in *quasi-optical radio wave* (a radio wave that behaves like a light ray).

quasi-bistable circuit A trigger-operated multivibrator. It operates as a flip-flop when the trigger frequency is sufficiently high.

quasi-instruction In a computer program, a data item appearing as an encoded instruction, but that is not acted upon.

quasi-linear feedback system A system in which the feedback elements are nearly linear, but not entirely linear.

quasi-negative Pertaining to a voltage that is negative, with respect to some other voltage, but whose absolute polarity is positive. For example, +0.5 volt is quasi-negative, with respect to +5.5 volts.

quasi-optical Behaving like light. The term is used to describe certain extremely short radio waves or other radiations that, like light rays, follow line-of-sight paths and can be directed, reflected, refracted, or diffused.

quasi-optical path A line-of-sight path followed by very short radio waves, such as microwaves.

quasi-positive Pertaining to a voltage that is positive, with respect to some other voltage, but whose absolute polarity is negative. For example, –0.5 volt is quasi-positive, with respect to –5.5 volts.

quasi-random A set of numbers considered to be random, but chosen according to an algorithm.

quasi-rectangular wave A wave whose shape approaches that of a rectangular wave, but that possesses a small amount of tilt and/or curvature.

quasi-scientific A term that is sometimes applied to the design of electronic systems or to the appraisal of circuit behavior, using an intuitive, rather than analytical approach.

quasi-sine wave A waveform that is not a perfect sine curve, but is close enough to be considered sinusoidal, for all practical purposes.

quasi-single sideband A modulated waveform that somewhat resembles single sideband, in which parts of both sidebands are present.

quasi-square wave A waveform that is not a perfect square, but is close enough to be considered square for all practical purposes. It is sometimes applied to a rectangular wave when a square wave is desired.

quasi-technical A term sometimes applied to qualitative tests, as opposed to quantitative tests.

quaternary **1.** Pertaining to a base-4 number system. **2.** Of an atom, joined to carbon atoms for four bonds. **3.** The fourth member of a 4-unit set.

Quebec Phonetic alphabet code word for the letter Q.

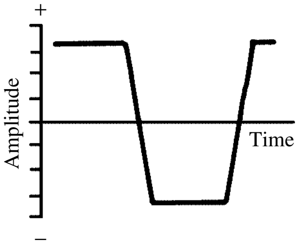

quasi square wave

quench **1.** To suddenly bring to an end (e.g., to quench an oscillation). **2.** To cool quickly, as in the quenching of a heated metal object. **3.** To extinguish the discharge in a gas tube.

quench capacitor A capacitor that prevents a spark from arcing across an inductor when current flow abruptly stops.

quench frequency See QUENCHING FREQUENCY.

quenching action Typical operation of a superregenerative circuit, in which regeneration is increased to nearly the point of oscillation and then reduced; this action is repeated at an ultrasonic frequency and results in very high sensitivity. Also see QUENCHING FREQUENCY, QUENCH OSCILLATOR, and SUPERREGENERATIVE CIRCUIT.

quenching frequency The frequency at which regeneration in a superregenerative circuit is increased and decreased.

quench oscillator In some superregenerative circuits, the separate ultrasonic oscillator that produces the required quenching action.

quench resistor A resistor in a quenching network that prevents a spark from occurring across an inductor when current flow stops.

queue A list of data, steps in a process, or commands awaiting execution in a specific order.

queuing theory A branch of mathematical electronics, dealing with the optimum order in which steps should be executed to obtain a particular end result.

quibinary code In computer operations, a binary-coded decimal system in which each decimal digit is represented by seven bits occupying places whose values are 8, 6, 4, 2, 0, 1, and 0.

quibinary decade circuit A decade counter consisting of a ring-of-5, followed by a single binary stage.

quick break An operating characteristic of a switch, relay, or circuit breaker whereby the contacts open rapidly—even when the actuating current or mechanical force is slow-acting.

quick-break fuse A fuse in which the wire melts and breaks almost instantly when the current rating is exceeded. Also called *quick-blow fuse.* Compare SLOW-BLOW FUSE.

quick-break switch A switch that opens rapidly—even if its handle or lever is moved slowly by the operator. This action minimizes arcing and prevents chatter. Compare QUICK-MAKE SWITCH.

quick charge The process of charging a battery, such as a nickel-cadmium (NICAD) or nickel-metal-hydride (NiMH) type, at a relatively rapid rate, at high charging current. It is sometimes used to charge a battery from a state of almost total discharge. Compare TRICKLE CHARGE.

quick-disconnect The characteristic of a connector that enables its mating halves to be separated quickly and simply, to break the circuit in which it is situated.

quickening liquid A solution of mercuric cyanide or mercuric nitrate, into which objects can be dipped prior to electroplating with silver. The process ensures good adhesion of the silver to the object.

quick make An operating characteristic of a switch, relay, or circuit breaker, whereby the contacts close rapidly—even when the actuating current or mechanical force is slow acting.

quick-make switch A switch that closes rapidly—even if its handle or lever is moved slowly by the operator. Compare QUICK-BREAK SWITCH.

quick printer A high-speed printer, used with a data terminal or computer. A relative term, depending on the user and the application.

quicksilver See MERCURY.

quick-stop control A control on tape recorders and some dictating machines that allows the operator to stop the tape, but keep the machine in the play or record mode. Also called *pause control.*

QuickTime Trade name (Apple Computer, Inc.) for system software commonly used in MULTIMEDIA applications with personal computers.

QUICKTRAN For multiaccess computer systems, a computer programming language based on FORTRAN and offering facilities, through the use of remote terminals, for running, testing, debugging, and compiling programs.

quiescent carrier operation A modulation system in which the carrier is present only during modulation (i.e., it is suppressed at all other times). Also called *controlled-carrier transmission.*

quiescent-carrier telephony A carrier-current (wired-wireless) telephone system in which the carrier is suppressed when there is no voice or alerting signal.

quiescent component In an electronic device, a component that is momentarily nonfunctional.

quiescent current Operating current (usually a direct current) flowing in a circuit or component during zero-signal or no-drain intervals. Also called IDLING CURRENT.

quiescent operation Zero-signal operation of a device, such as a transistor, diode, magnetic amplifier, or similar component.

quiescent period The no-signal interval during which equipment is not operating—even though it is energized.

quiescent point The point on the characteristic curve of a transistor, diode, or similar device, denoting the zero-signal operating conditions.

quiescent push-pull Denoting a push-pull stage, especially an audio power-output amplifier, in which the direct-current signal is essentially zero.

quiescent state The inactive, or resting, state of an active component, such as a transistor or vacuum tube.

quiescent value The zero-signal value of current or voltage for any component supplied with operating power.

quiet AGC See DELAYED AUTOMATIC GAIN CONTROL.

quiet automatic gain control See DELAYED AUTOMATIC GAIN CONTROL.

quiet automatic volume control See DELAYED AUTOMATIC GAIN CONTROL.

quiet AVC See DELAYED AUTOMATIC GAIN CONTROL.

quiet battery A direct-current source specially designed and filtered to minimize noise components in its output.

quieting Noise-voltage reduction in the output of a frequency-modulation (FM) receiver when an unmodulated carrier is received. Also called *noise quieting.*

quieting level In a frequency-modulation (FM) receiver, the limiter threshold point.

quieting sensitivity In a frequency-modulation (FM) receiver, the lowest input-signal amplitude at which the output signal-to-noise ratio is below the specified limit.

quiet tuning A system of tuning in which the output of a receiver is muted until a station is tuned in properly.

quinary code See BIQUINARY CODE.

quinary counter A decade counter consisting of a five-stage ring.

quinhydrone electrode A pH meter electrode consisting of a platinum wire in a solution of quinhydrone ($C_{12}H_{10}O_4$). Also see PH METER.

quintillion The number 1,000,000,000,000,000,000 (10^{18}).

quintupler **1.** A rectifier circuit that delivers a direct-current output voltage equal to about five times the peak value of the alternating-current input voltage. **2.** A circuit that delivers an output signal at the fifth harmonic of the input signal.

QWERTY The standard typewriter and computer keyboard layout. The name is derived from the first several letters in the top letter row: Q, W, E, R, T, and Y.

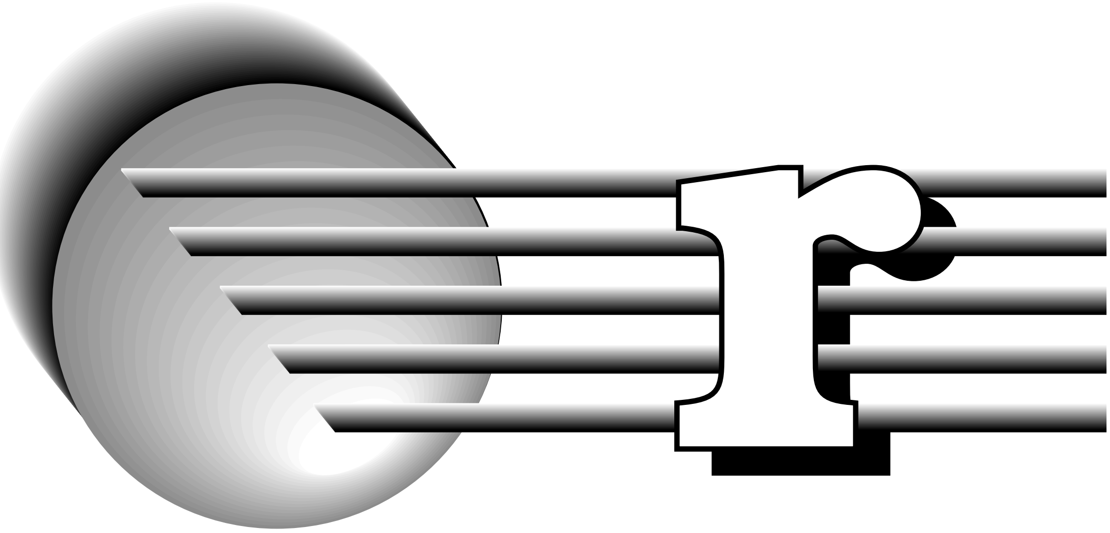

R **1.** Symbol for RESISTANCE. (Also, r.) **2.** Radiotelegraph abbreviation for ROGER. **3.** Symbol for RELUCTANCE. **4.** Abbreviation of RADIUS.

r **1.** Symbol for ROENTGEN. **2.** Symbol for *correlation coefficient.* **3.** Abbreviation of RADIUS. (Also, abbreviated R.)

r_e Symbol for CLASSICAL ELECTRON RADIUS.

RA **1.** Abbreviation of *right ascension.* **2.** Abbreviation of RANDOM ACCESS.

rabbit ears An indoor antenna, sometimes used with a television receiver, consisting of two vertical whips (usually telescoping), the angle between which is adjustable.

RAC Abbreviation of RECTIFIED ALTERNATING CURRENT.

R_{ac} Symbol for AC RESISTANCE. (Also, r_{ac}.)

race Incorrect interpretation of the clock pulses by a digital circuit. Also called *racing.* The circuit improperly attempts to do many operations during one clock pulse, rather than a single operation.

RACES Abbreviation of *Radio Amateur Civil Emergency System.*

raceway See WIRE DUCT and WIREWAYS.

rack An upright frame for holding equipment of RACK-AND-PANEL CONSTRUCTION.

rack-and-panel construction A method of building electronic equipment on a chassis attached horizontally or vertically to a vertical panel. After completion of a unit, the panel is fastened in place on a RACK. Several such panels fill the rack.

rack and pinion A device used for mechanical adjustment of a control, such as the tuning control in a radio receiver. A gear engages a serrated rod. As the gear is turned, the rod moves lengthwise.

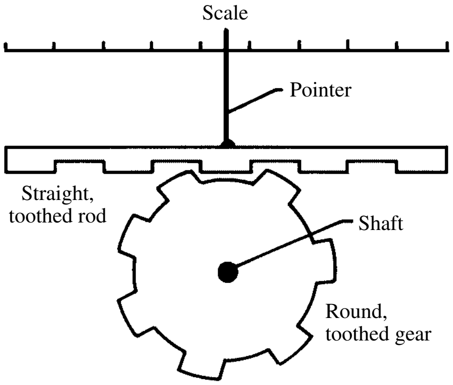

rack and pinion

rack up In computer operations, a way of displaying data, in which a new line added to the already completely occupied screen bumps up what has forgone, thus eliminating the top line.

racon Acronym for *radar beacon.*

rad **1.** A unit of ionizing radiation received by a body (dose) equal to 0.001 J/kg. **2.** Abbreviation of RADIAC. **3.** Abbreviation of RADIAN. **4.** Abbreviation of RADIO. **5.** Abbreviation of RADIX.

radar **1.** A microwave system for detecting objects and determining their distance, direction, heading, speed, and other characteristics. Signals from the transmitter are reflected back to the transmitter site by the object, and the reflection (sometimes along with the transmission) is displayed on a cathode-ray screen. The name is an

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

acronym for *radio detection and ranging.* **2.** The theory and application of radio detection and ranging systems as defined in **1**.

radar altitude The distance of an aircraft above the surface of the earth, as determined by radar. This value varies with the terrain over which the aircraft passes.

radar antenna Any antenna used for transmitting and/or receiving radar signals.

radar astronomy The use of radar equipment to observe and map planets, moons, and asteroids, and to measure their distance from the earth or from a spacecraft.

radar beacon A radar transceiver that, on receipt of radar signals, transmits encoded signals from which the operator can take a bearing.

radar beam The cone-shaped main lobe of energy emitted by a radar antenna. The narrower the beam, the greater the resolution of the radar system.

radar clutter Visual interference on a radar screen caused by reflections from ground or sea.

radar countermeasures Abbreviations, RCM and rad CM. In wartime, any method of interfering with enemy radar, such as jamming or use of decoys.

radar detector **1.** A device used in automobiles and trucks to detect the proximity of police or highway-patrol radar. **2.** A device used in military applications, especially aviation, to indicate the presence of radar.

radar display The scheme via which a radar set portrays the relative positions of the objects that produce echoes. The most common is the azimuth/range display, showing compass bearings (usually in degrees clockwise from geographic north) and radius from the transmitter (usually in miles or kilometers). Altitude above mean sea level can be displayed for individual echoes.

radar homing A method of missile homing in which radar is used to track a target.

radar speed trap A radar system used by traffic police to spot speeding vehicles.

radar telescope The transmission and reception unit used in radar astronomy. Compare RADIO TELESCOPE.

RadCM Abbreviation of RADAR COUNTERMEASURES. (Also, RCM.)

radial **1.** One of several conductors used to enhance the performance of an unbalanced, vertical antenna. These can be constructed from wire or metal tubing, and generally measure one-quarter wavelength or more. When a vertical antenna is mounted at the earth's surface, the ground conductivity is improved by these conductors, which run outward from the base of the radiator, and are connected to the shield of a coaxial feed line. The greater the number of radials of a given length, the more the ground loss is reduced. Also, the longer the radials for a given number, the better. If a vertical radiator is mounted with its base more than one-quarter wavelength above the earth's surface, there need only be three or four conductors measuring exactly one-quarter wavelength. See GROUND-MOUNTED VERTICAL ANTENNA, GROUND-PLANE ANTENNA. **2.** Pertaining to the distance from the center of a circle to its periphery. **3.** Pertaining to the distance from the center of a sphere to its surface. **4.** Extending or emanating outward in a straight line from a defined point.

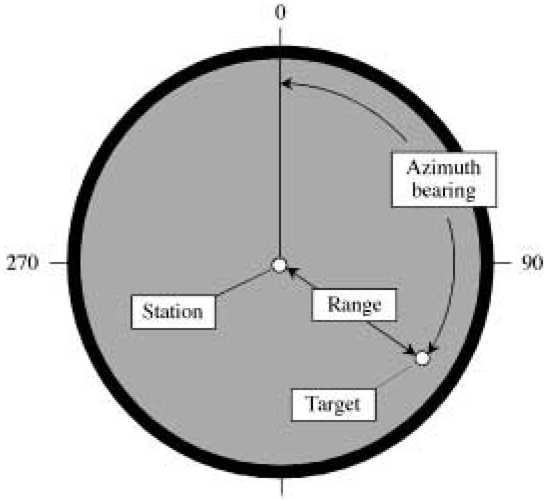

radar display

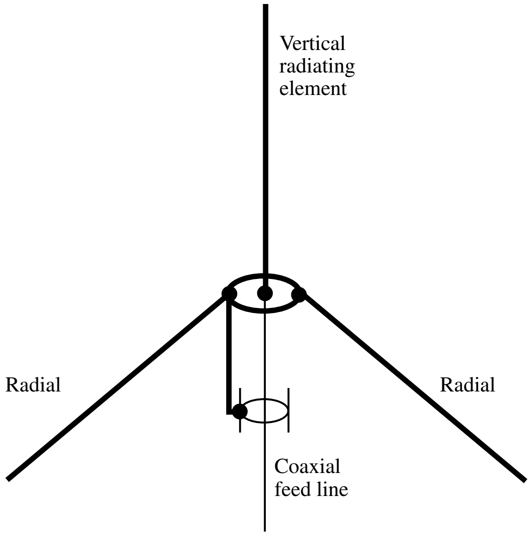

radials

radial ground An earth connection composed of radials buried in the ground.

radial lead A lead (pigtail) attached perpendicular to the axis of a component, such as a resistor or capacitor.

radian Abbreviation, rad. The angle at the center of a circle subtended by an arc whose length is equal to the radius. Equal to approximately 57.2958 degrees.

radiance The radiant flux emitted by an object. Radiance is measured in terms of the amount of energy contained in a unit solid angle (steradian) with the source at the apex.

radians-to-degrees conversion The conversion of radian angular measure into degrees. To change radians to degrees, multiply the number of degrees by 57.2958. Thus, 0.7854 radian = 45 degrees. Compare DEGREES-TO-RADIANS CONVERSION.

radiant efficiency The ratio of the radiant energy emitted by a source to the energy consumed by the source. The radiant energy is generally specified within a certain range of wavelengths. An example is the incandescent light bulb, which has relatively low radiant efficiency in the visible spectrum between about 750 and 390 nanometers.

radiant energy **1.** Any form of energy emitted by a source and propagated through space as an electromagnetic disturbance. Included are radio waves, infrared, visible light, ultraviolet, X rays, and gamma rays. **2.** Electromagnetic disturbances at infrared and shorter wavelengths.

radiant flux The rate at which radiant energy is emitted.

radiation **1.** The emission of energy or particles (e.g., waves from an antenna, X rays from an X-ray tube, energy from a radioactive material, heat from a body, etc.). **2.** Radio waves, infrared, visible light, ultraviolet, X rays, or gamma rays. **3.** Ionizing emissions from radioactive substances (e.g., alpha particles, beta particles, neutrons, gamma rays, etc.).

radiation angle The horizontal or vertical angle at which electromagnetic waves are radiated from an antenna. Measured between the central axis of the main lobe and the horizon, or between the central axis of the main lobe and geographic north.

radiation belts See VAN ALLEN RADIATION BELTS.

radiation counter An instrument used for determining the intensity of atomic-particle radiation, X rays, or gamma rays. It operates by means of ionization of a gas in a sealed tube.

radiation field The portion of the electromagnetic field that is propagated by a radiator, as opposed to the induction field.

radiation intensity For a directional radio transmitting antenna, the radiated power per steradian in a given direction.

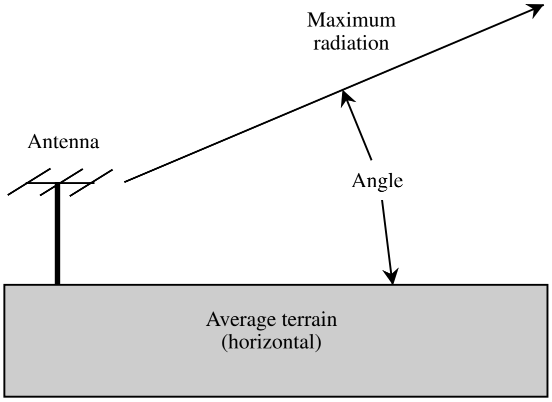

radiation angle

radiation loss Loss of energy through radiation from a conductor. Also see LOSS.

radiation pattern A graphical representation of the intensity of the electromagnetic field in various directions from a radiator, such as a transmitting antenna. It is usually shown in either the horizontal plane or a specific vertical plane containing the antenna. Also see LOBE.

radiation physics The study of radiation and its effects on matter. Radiation physics is especially concerned with ionizing radiation, but it can involve any kind of particle or electromagnetic energy.

radiation pressure Pressure exerted on a surface by impinging electromagnetic radiation.

radiation resistance The inherent resistance at the feed point of a resonant radio antenna.

radiation sickness General physiological symptoms resulting from a short-term overdose of X rays, gamma rays, or atomic-particle radiation.

radiator **1.** The element of an antenna from which radio energy is directly radiated, as opposed to the transmission line, lead-in, reflector, or director. **2.** See LOUDSPEAKER.

radio **1.** Wireless electrical communication, i.e., by means of electromagnetic waves. **2.** See RADIO RECEIVER. **3.** See RADIO TRANSCEIVER. **4.** See RADIO TRANSMITTER. **5.** To communicate by radio.

radio- **1.** A prefix meaning "pertaining to wireless electrical communication." Examples: *radiotelephone* and *radiotelegraph.* **2.** A prefix meaning "using radio waves." Examples: *radiosonde*, *radiolocator*, and *radiothermics.* **3.** A prefix meaning "pertaining to using or possessing radioactivity," or "pertaining to X rays." Examples: *radiograph*, *radioisotope*, and *radiologist.*

radioactive Having the property of emitting alpha, beta, and (sometimes) gamma rays as the result of nuclear disintegration. Also see HALF-LIFE.

radioactive element A chemical element that is RADIOACTIVE (e.g., uranium). Also called *radioelement.*

radioactive isotope See RADIOISOTOPE.

radioactive tracer A quantity of radioactive material put into a system so that its path can be monitored by means of a radiation detector. An example is the introduction of radioactive barium into the large intestine. The flow and concentration of the barium gives an indication of the functioning of the lower intestine.

radioactive transducer A pickup device for detecting and measuring radioactivity (e.g., *Geiger-Mueller tube*).

radioactivity counter See GEIGER COUNTER and SCINTILLATION COUNTER.

radio altitude See RADAR ALTITUDE.

radio amateur An electronics hobbyist licensed to operate two-way wireless communications stations in various assigned frequency bands, without receiving payment for services rendered.

Radio Amateur Civil Emergency System Abbreviation, RACES. A civil-defense organization of licensed amateur radio stations. Also see RADIO AMATEUR.

radio astronomy The observation, study, and analysis of radio-frequency electromagnetic emissions from bodies or points in space, and the study of these bodies through their radiations.

radioautograph See AUTORADIOGRAPH.

radio beacon **1.** A radio transmitter of direction-finding or guidance signals. **2.** Also called *radio beam.* The signals transmitted by a radio beacon, as defined in **1**.

radio beam **1.** Antenna radiation focused in one direction. **2.** See RADIO BEACON, **2**.

radiobiology A field of biology concerned with the influence of radiant energy or radioactivity on living organisms.

radio broadcast A radio transmission directed to numerous, nonspecific receivers—especially by a station in the *broadcast service.* Also called RADIOCAST. Also see BROADCAST SERVICE, **1**, **2**.

radio car An automobile equipped with a two-way radio.

radio carbon Radioactive carbon (i.e., carbon 14).

radiocast See RADIO BROADCAST.

radio channel A single, usually narrow radio-frequency band within a larger band, in which stations are authorized to transmit signals of a specified type. Also see CHANNEL, **1**; CHANNEL SEPARATION; and CHANNEL WIDTH.

radiochemistry The chemistry of radioactive substances.

radio communication Wireless communication carried on by means of radio-frequency electromagnetic waves.

radio compass See DIRECTION FINDER.

radioconductor A substance or body whose electrical conductivity is affected by radio waves, and that can be used as a sensor of such waves.

radio control See REMOTE CONTROL.

radio direction finder See DIRECTION FINDER.

radio Doppler **1.** A change in the frequency of a radio signal emitted by a source having radial motion, with respect to the receiver. **2.** An electronic device used to measure radial speed by means of the Doppler effect at radio frequencies.

radio-electronics The branch of electronics specifically involved with wireless communications.

radioelement See RADIOACTIVE ELEMENT.

radio engineer A trained professional skilled in the physics and mathematics of radio communications, and in the theory and application of basic electronics engineering and related subjects. Also see RADIO ENGINEERING.

radio engineering The branch of electronics engineering devoted to the theory and operations of radio communication.

radio field strength The intensity of radio waves at a given point. Also see FIELD INTENSITY, **2** and RADIO MAP.

radio frequency Abbreviation, RF. **1.** Consisting of, or pertaining to, alternating currents at frequencies above about 9 kHz (the lowest allocated radio communications frequency). **2.** Consisting of, or pertaining to, electromagnetic fields whose wavelengths are longer than those of infrared, but shorter than about 33 kilometers (corresponding to a frequency of 9 kHz). Also see RADIO SPECTRUM.

radio-frequency amplifier **1.** In a superheterodyne circuit, the channel in which the incoming signal is amplified. Compare INTERMEDIATE-FREQUENCY AMPLIFIER. **2.** Broadly, an amplifier of radio-frequency signals.

radio-frequency choke Abbreviation, RFC. A low-inductance coil used to block radio-frequency (RF) alternating currents. Many RF chokes have air cores; some have cores of ferrite or powdered iron.

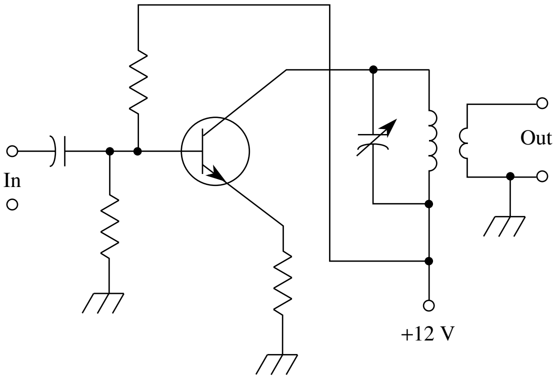

(General bipolar circuit)

radio-frequency amplifier, 2

radio-frequency current Symbol, I_{RF}. **2.** The intensity of a generated radio-frequency (RF) signal, usually expressed in microamperes. **2.** Loosely, any measurable RF signal.

radio-frequency heating The generation of heat in an object by an intense radio-frequency electromagnetic field. See, for example, DIATHERMY, **1**; DIELECTRIC HEATING; and INDUCTION HEATING.

radio-frequency interference Abbreviation, RFI. **1.** Annoying electrical noise in radio-frequency (RF) amplifiers, detectors, and instruments. **2.** Undesired RF signals that compete with desired ones in amplifiers, receivers, and instruments. **3.** The unwanted interception and demodulation of a strong RF signal by an audio-frequency (AF) device, such as a telephone set or high-fidelity stereo amplifier.

radio-frequency meter An instrument for measuring signals of RADIO FREQUENCY (9 kHz and above).

radio-frequency oscillator Abbreviation, RFO. An oscillator (self-excited or crystal-controlled) for operation at radio frequencies. In such an oscillator, stray components, efficiency, and general losses are of primary concern. Also see RADIO FREQUENCY.

radio-frequency power Symbol, P_{RF}. Alternating-current power at radio frequencies.

radio-frequency resistance The total in-phase resistance exhibited by a conductor at radio frequencies. This opposition to current includes direct-current resistance and the in-phase components caused by skin effect, shielding, and the presence of dielectrics.

radio-frequency selectivity The selectivity of a radio-frequency (RF) channel, such as the RF amplifier and first detector of a superheterodyne circuit.

radio-frequency transformer Abbreviation, RF transformer. A device used for the purpose of impedance matching, antenna tuning, or interstage coupling at frequencies above the range of human hearing (approximately 20 kHz and up). The device might consist of solenoidal windings with an air core, solenoidal windings with a powdered-iron or ferrite core, or toroidal windings with a powdered-iron or ferrite core. The windings are designed to minimize distributed capacitance and direct-current resistance. When no reactance is present, the impedance-transfer ratio is equal to the square of the turns ratio. Compare AUDIO-FREQUENCY TRANSFORMER.

radio-frequency transistor **1.** A transistor capable of providing significant amplification at radio frequencies. **2.** A transistor operable at frequencies above 100 kHz.

radiogenic Produced by radioactivity.

radiogoniometer A radio compass (see DIRECTION FINDER and GONIOMETER, **1**).

radiogram A (usually printed out) message transmitted and received via radiotelegraphy or radioteletype. The term is an acronym for *radio telegram.*

radiograph **1.** To contact by sending a RADIOGRAM. **2.** An X-ray photograph.

radio homing **1.** A method of homing that uses the tracking of a target on the basis of a radio signal emitted by that target. **2.** A method of keeping a missile on track via radio remote control.

radio interference **1.** Interference to radio communication, from whatever cause. **2.** See RADIO-FREQUENCY INTERFERENCE.

radioisotope A radioactive isotope (natural or artificial) of a normally nonradioactive chemical element (e.g., radioactive carbon). Also see ISOTOPE.

radio jamming See JAMMING.

radio knife A surgical instrument consisting essentially of a needle that forms a high-frequency arc. The arc simultaneously cuts and cauterizes tissue.

radiolocation A process whereby the position of a vehicle, aircraft, or ocean-going vessel is determined. The simplest method is the directional method. Two or three fixed receiving stations are used. Radio direction-finding (RDF) equipment is employed at each station, in conjunction with a transmitter aboard the vessel, to establish the bearings of the vessel with respect to each station. Radar can also be used to locate vessels. In wartime, enemy craft can sometimes be located by visual or infrared apparatus. Satellites can locate enemy ships and missiles, in some cases with an error smaller than the length of the vessel itself. Compare RADIONAVIGATION.

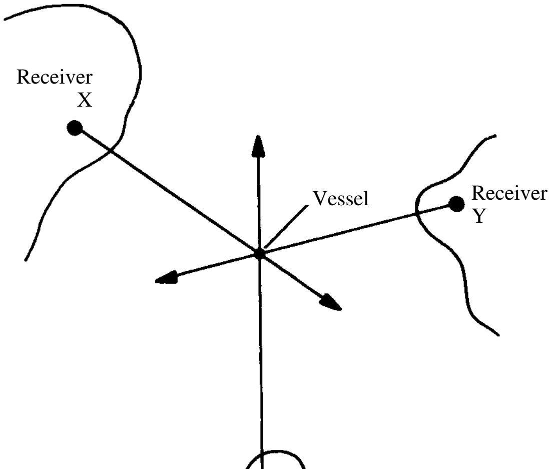

radiolocation

radiolocator See RADAR.

radiological system See X-RAY THERAPY SYSTEM.

radiologist A specialist skilled in RADIOLOGY.

radiology The science embracing the theory and use of X rays and radioactive substances in the diagnosis and treatment of diseases and ailments.

radiolucency **1.** The property of a material that allows ionizing radiation to pass through it with little or no absorption. **2.** The extent to which a material transmits ionizing radiation.

radioluminescence Visible light emitted from a radioactive material. A good example is radium; it was once used on wristwatch dials so that they could be seen in the dark.

radiolysis Chemical decomposition brought about by radiation.

radioman A radio technician or operator.

radio map A map of a geographic area, on which lines are drawn connecting measured points of equal field strength for signals from a radio station at the approximate center of the area.

radiometeorograph See RADIOSONDE.

radiometer A device for detecting and measuring the strength of radiant energy. One form consists of a set of vanes blackened on one side and mounted on pivots in a partially evacuated glass bulb. Visible light or infrared causes the vane assembly to rotate, the speed being proportional to the intensity of the light.

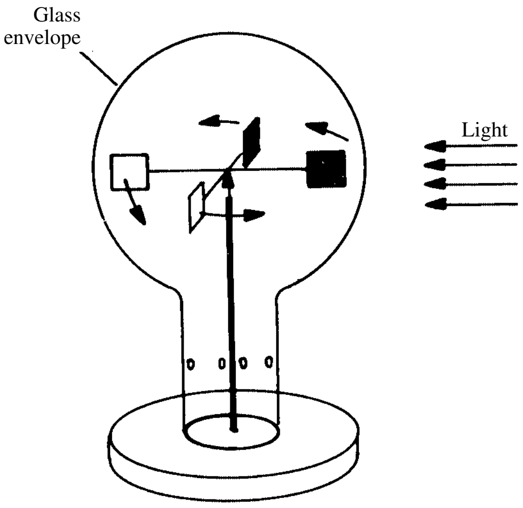

radiometer

radiometry The science and art of measuring radiation in the infrared, visible, and ultraviolet regions of the electromagnetic frequency spectrum. Compare PHOTOMETRY.

radio micrometer See MICRORADIOMETER.

radionavigation The use of radio apparatus, usually in conjunction with computers, by personnel aboard moving vessels, for the purpose of plotting, correcting, and maintaining a course. The intersecting-line method is simplest. Two or three land-based transmitters are needed. Their locations must be accurately known. A direction-finding device on the vessel is used to determine the bearings of each of the transmitters. Aircraft radionavigation can be performed with the aid of radar. The most sophisticated radionavigation techniques employ the Global Positioning System (GPS). Computers are used to project the course of a craft based on its current position, its speed, and the direction of its movement. Course corrections are made by choosing the desired course and having the computer calculate speed and/or direction changes. Compare RADIOLOCATION.

radio net A group of radio stations operating together in an organization, often on or near the same frequency.

radio network See RADIO NET.

radio operator A technician licensed to operate a transmitter in the radio, television, or radar services.

radiopaque Opaque to X rays or other ionizing radiation. Compare RADIOPARENT.

radioparent Transparent to X rays or other ionizing radiation. Compare RADIOPAQUE.

radiophone See RADIOTELEPHONE.

radiophoto A photograph transmitted and received by radio. Also see FACSIMILE.

radio pill See ENDORADIOSONDE.

radio prospecting The use of radio-frequency devices to locate underground or underwater metals and mineral deposits. Also see METAL LOCATOR.

radio range A radio station providing navigational aid to airplanes.

radio receiver The complete apparatus that selects, amplifies, demodulates, and reproduces a radio signal for purposes of communication, as distinct from *facsimile receiver*, *remote-control receiver*, *telemetry receiver*, *television receiver*, etc.

radiosensitivity **1.** The property of being sensitive to ionizing radiation. Most photographic films have this property. **2.** The extent to which a substance or device is sensitive to ionizing radiation.

radio service technician An electronics technician skilled in the repair and maintenance of radio equipment—especially receivers.

radiosonde A balloon-carried combination of radio transmitter and transducers, for sending to a ground monitoring station signals revealing such atmospheric conditions as temperature, humidity, and pressure. It is used mainly for gathering meteorological data at high altitudes.

radiosonobuoy See SONOBUOY.

radio spectroscope A device used by radio astronomers to obtain the radio-frequency profile of a distant star or galaxy. It generally consists of a graph, obtained by scanning the radio spectrum and plotting signal intensity as a function of frequency or wavelength.

radio spectrum The continuum of frequencies useful for radio communication and control. Classified in the following manner: *Very low frequency (VLF)*, 9 to 30 kHz; *low frequency (LF)*, 30 to 300 kHz; *medium frequency (MF)*, 300 to 3000 kHz; *high frequency (HF)*, 3 to 30 MHz; *very high frequency (VHF)*, 30 to 300 MHz; *ultrahigh frequency (UHF)*, 300 to 3000 MHz; *super high frequency (SHF)*, 3 to 30 GHz; *extremely high frequency (EHF)*, 30 to 300 GHz.

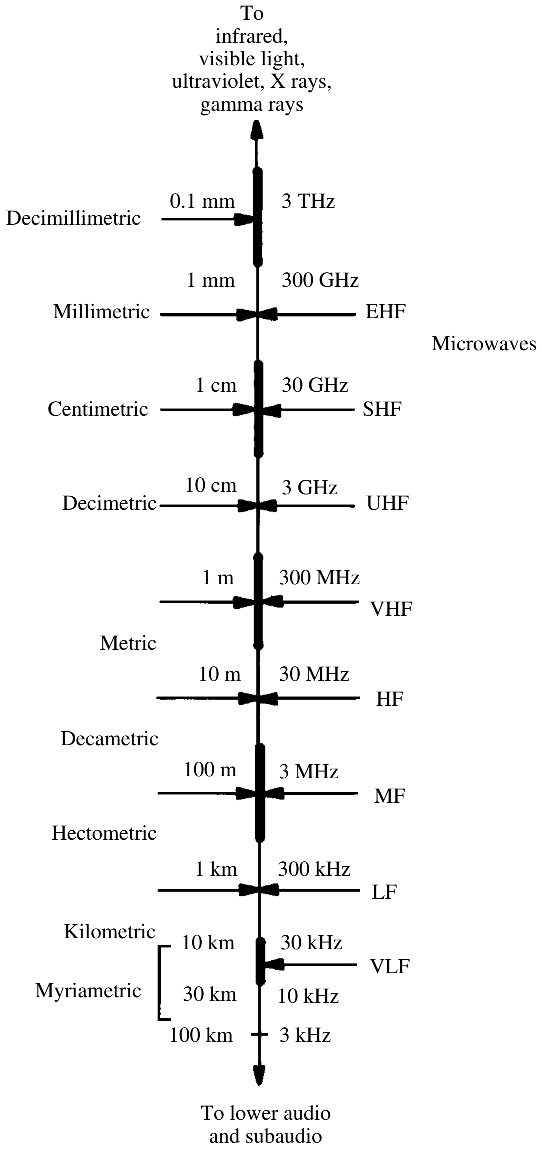

radio spectrum

radiostat See CRYSTAL FILTER.

radio station **1.** The location at which a radio transmitter and/or receiver is/are installed. **2.** The complete set of equipment for a radio receiving and/or transmitting installation, including the studio, linking apparatus, and antennas. **3.** A standard broadcast station.

radio technician A professional skilled in the construction, testing, repair, and maintenance of radio equipment, and sometimes in its design, and who usually works under the supervision of a radio engineer. Also see RADIO SERVICE TECHNICIAN.

radiotelegram See RADIOGRAM.

radiotelegraph **1.** Pertaining to the theory and application of, and the equipment for, Morse code transmission and/or reception via radio. **2.** An installation for Morse code transmission and/or reception via radio. **3.** The transmission and/or reception of Morse code signals via radio.

radiotelegraph code See CONTINENTAL CODE.

radiotelegraph distress signal See SOS.

radiotelegraph monitor See KEYING MONITOR.

radiotelegraphy The transmission and/or reception of telegraphic communications, usually Morse code, by means of radio.

radiotelephone **1.** Pertaining to the theory and application of, and the equipment for, voice transmission and/or reception via radio. **2.** An installation for voice transmission and/or reception via radio. **3.** The transmission and/or reception of voice signals via radio.

radiotelephone distress signal See MAYDAY.

radio/telephone patch See PHONE PATCH.

radiotelephony The transmission and/or reception of audio signals, usually human voices, by means of radio.

radio telescope A directional antenna and associated equipment for receiving and evaluating the radio-frequency electromagnetic radiation from space—especially from celestial objects (such as the sun, planets, stars, nebulae, galaxies, quasars, etc.). See RADIO ASTRONOMY.

radioteletype **1.** Pertaining to the theory and application of, and the equipment for, text data transmission and/or reception via radio. **2.** An installation for text data transmission and/or reception via radio. **3.** The transmission and/or reception of text data signals via radio.

radioteletypewriter A teletypewriter adapted to radio, rather than wire service; it is used in some RADIOTELETYPE installations. In recent years, personal computers and terminals have largely replaced adapted teletypewriters for this purpose.

radiotherapy The use of X rays and/or radioactive substances in the treatment of disease and disorders.

radiothermics The science of the generation of heat by means of radio-frequency current.

radiothermy See DIATHERMY.

radio thorium Radioactive THORIUM.

radiotracer See TRACER.

radio transceiver A RADIO RECEIVER and RADIO TRANSMITTER built into a single unit, and generally intended for use in two-way communication.

radio transmitter The complete apparatus that generates radio-frequency power, modifies it with the data needed for communication, and delivers the product to an antenna for radiation into space. Here, the radio transmitter is distinguished from similar equipment: *facsimile transmitter*, *remote-control transmitter*, *telemetry transmitter*, *television transmitter*, etc.

radio-transparent material 1. A substance through which radio waves pass with little or no attenuation. **2.** A substance through which X rays, gamma rays, or high-speed subatomic particles can pass with little or no attenuation.

radiotrician Acronym for *radio electrician.* See RADIO SERVICE TECHNICIAN.

radio tube 1. A VACUUM TUBE used at radio frequencies. **2.** A vacuum tube used as an amplifier, local oscillator, detector, or mixer in an early radio receiver.

radiovision See TELEVISION.

radio watch See WATCH.

radio waves Electromagnetic waves in the RADIO SPECTRUM.

radio window That portion of the radio-frequency electromagnetic spectrum that passes through the atmosphere, rather than being refracted or absorbed. The wavelength range is about 20 meters to 5 millimeters, or 15 MHz to 60 GHz. The lower limit of this range is affected by ionospheric conditions. The upper frequency limit depends on various factors, including relative humidity and dust content of the air.

radium Symbol, Ra. A rare radioactive metallic element. Atomic number, 88. Atomic weight, 226.025.

radius The straight-line distance from the center of a circle or sphere to its periphery.

radius vector In spherical or polar coordinates, a line segment drawn from the pole, or origin, and representing the vector magnitude.

radix The number indicating the number of symbols in a system of numerical notation, and the powers of which give the place values of the system. Thus, 10 is the radix of the decimal system, and 2 is the radix of the binary system. Also called BASE.

radix point In a number, the point (dot or period) separating the integral and fractional digits. Its specific name depends on the system of notation involved: *binary point*, *decimal point*, etc.

radome A plastic shell housing a radar antenna—especially aboard an aircraft.

radon Symbol Rn. A gaseous radioactive element that results from the disintegration of radium. Atomic number, 86. Atomic weight, 222.

rad/s Abbreviation of *radians per second*, the SI unit of angular velocity.

rad/s^2 Abbreviation of *radians per second squared*, the SI unit of angular acceleration.

radux A continuous-wave, low-frequency radionavigation system. Position is determined by comparing the phase of two signals sent from different locations.

RAID Acronym for REDUNDANT ARRAY OF INDEPENDENT DISKS.

rainbow generator A test-signal generator that produces a full color spectrum, a pattern resembling the successive coloration of a rainbow, on the screen of a color-television receiver. Also see RAINBOW PATTERN.

rainbow pattern A test pattern for servicing a color-television receiver. It consists of a full color spectrum, thus taking its name from its resemblance to a rainbow. Also see RAINBOW GENERATOR.

RAM Abbreviation of RANDOM-ACCESS MEMORY.

ramp A sawtooth wave with a linear rise and a practically instantaneous decay; its name was derived from its resemblance to an incline.

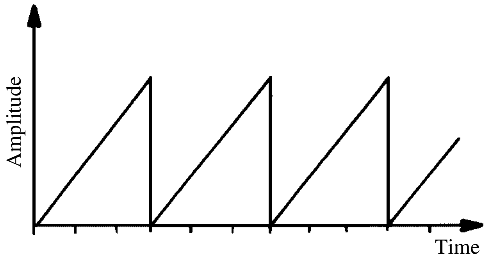

ramp

ramp generator A test-signal generator that produces sawtooth-wave signals. Also see RAMP.

R and D Abbreviation of research and development or research and design. (Also, R&D.)

random access Abbreviation, RA. In computer and data-processing operations, pertaining to storage or memory in which data can be recovered in any order.

random-access memory In computer and data-processing systems, a memory providing access time that is independent of the address.

random deviation Irregular RIPPLE.

random-fed antenna An antenna that uses RANDOM FEED.

random feed A method of connecting a transmission line to an antenna, wherein the feed point is not necessarily at the center and not necessarily at a current loop or voltage loop. This technique is rarely used; it generally results in some radiation from the feed line.

random noise Electrical noise in which the pulses or fluctuations have no discernible pattern of occurrence (i.e., they are haphazard in frequency and amplitude).

random number A number derived by chance. It is used in statistical analysis for various purposes.

random number generator Hardware or software that provides a sequence of numbers or digits that are random for the purpose of a given statistical application.

random occurrence See CHANCE OCCURRENCE.

random variable In statistics, a variable that can have a number of values, each of the same probability.

random winding A coil winding in which the turns are wound haphazardly to reduce distributed capacitance.

range **1.** The limits within which a circuit or device operates (i.e., the territory defined by such limits). Examples: *current range*, *frequency range*, and *voltage range*. **2.** The difference between the upper and lower limits of deflection of a meter. **3.** The distance over which a transmitter operates reliably. **4.** A clear area for testing antennas. **5.** The distance between a radar station and a target. **6.** The possible values for a quantity or function that lie between given limits.

range capacitor See TRIMMER CAPACITOR.

range-height indicator Abbreviation, RHI. A radar display in which the horizontal axis shows distance to the target, and the vertical axis shows elevation of the target.

range mark See DISTANCE MARK.

range plotting The creation of a graph of the distance (range) to objects, as a function of direction or orientation in two or three dimensions. Commonly used in robot guidance systems.

range resistor See TRIMMER RESISTOR.

range sensing The measurement of distances to objects via electronic methods such as radar, sonar, vision systems, etc. Commonly used in robot guidance systems.

ranging **1.** Any means of determining the distance from a station or vehicle to an object or objects. **2.** Any of several methods for a vehicle, vessel, aircraft, spacecraft, or robot to navigate in its environment by measuring, and keeping track of changes in, the distance between itself and one or more objects or beacons.

rank **1.** To arrange in a specific sequence according to significance. **2.** A place in such a sequence.

Rankine scale A temperature scale on which the freezing point of water is 491.69 degrees, and the boiling point 671.69 degrees. Absolute zero is represented by 0 degrees. For conversion to kelvins, multiply degrees Rankine by 5/9.

rapid drift A fast change of a quantity or setting (usually in one direction) with time.

rapid printer See QUICK PRINTER.

raser A device that produces coherent electromagnetic waves at radio frequencies; the radio-frequency equivalent of a LASER.

raster The rectangle of light (composed of unmodulated lines) seen on the screen of a television picture tube when no signal is present.

ratchet circuit See COMMUTATOR, **2** and ELECTRONIC RATCHET.

rate action See DERIVATIVE ACTION.

rate effect In a four-layer semiconductor device, the tendency for the switch to conduct undesirably as a result of a transient spike.

rate-grown transistor See GRADED-JUNCTION TRANSISTOR.

rate gyro A special gyroscope for measuring angular rates.

rate of change **1.** The extent to which the value of a dependent variable changes in accordance with a specified change in an independent variable (usually time). **2.** A quantitative expression of the speed with which a dependent variable changes, with respect to an independent variable (usually time).

rate signal A signal whose amplitude is proportional to the derivative of a variable, with respect to time.

rate time In automatic-control operations, the time over which the addition of DERIVATIVE ACTION advances PROPORTIONAL ACTION.

ratio-arm bridge A simple four-arm bridge in which the balancing potentiometer supplies the two arms, one on each side of the slider at all settings.

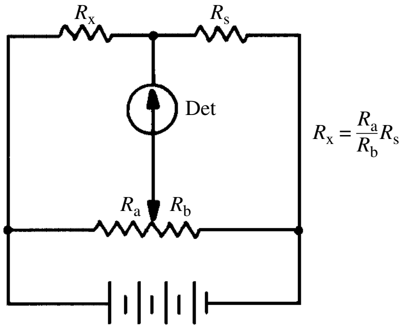

ratio-arm bridge

ratio arms Two impedance arms serving to establish the numerical ratio of a bridge circuit.

ratio control In automatic-control operations, a system in which the controlled variable is in a prescribed ratio to another variable.

ratio detector A frequency-modulation (FM) second detector resembling the FOSTER-SEELEY DISCRIMINATOR, except that one of the two diodes is reversed and the junction point of the load resistors is grounded. In an FM circuit using a ratio detector, no limiter is required. The ratio of the direct-current outputs is proportional to the ratio of the instantaneous signal voltages applied to the two diodes.

ratio meter An instrument that compares two different signals (and indirectly their sources) and delivers a reading of their ratio.

rational number A number that can be expressed in the form a/b, where a and b are integers and b is not equal to zero.

rational operation Any of the conventional arithmetic operations: multiplication, division, addition, or subtraction.

ratio of geometric progression In a geometric progression, the ratio of one value to the next.

ratio of similitude The ratio of the lengths of corresponding sides in similar geometric figures.

rat race See HYBRID RING.

raven red A variety of red oxide of iron, a commercial red paint used as the magnetic coating of early recording tapes.

raw ac Unrectified alternating current (ac) or voltage.

raw data Data that has not been processed in any way.

rawinsonde A RADIOSONDE tracked by a radio direction finder to determine wind velocity. The name is an acronym from *radar wind radiosonde.*

raw tape See BLANK TAPE.

ray **1.** A line of radiant energy. Such a line (e.g., the path of a single photon of visible light) is imagined to arise from a point source and have zero width. **2.** A thin beam of radiant energy (e.g., the beam of electrons in a cathode-ray tube). **3.** A quantity of radiant energy or ionizing radiation (e.g, *gamma ray*). **4.** One of numerous lines converging toward, or emanating from, a specific point. **5.** A vector representing the direction in which an electromagnetic field or acoustic disturbance travels. **6.** Also called *half line.* The set of points on a line consisting of a defined origin and all the points on one side of the origin. Example: the positive reactance axis in an ARGAND DIAGRAM.

Raydist A continuous-wave, medium-frequency radionavigation system. The position is determined according to the phase difference between two signals transmitted from different locations.

Rayleigh-Carson theorem An expression of the reciprocal relationship between the transmitting and receiving properties of an antenna. If voltage E applied to antenna A causes current I to flow at a given point in antenna B, then the same voltage (E) applied at that point in antenna B will produce identical current I (same magnitude and phase) at the point in antenna A, where voltage E originally was applied. Also see RECIPROCITY THEOREM.

Rayleigh distribution A probability-density function, used to describe the behavior of sky-wave electromagnetic signals.

Rayleigh's law The hysteresis loss in a magnetic material varies in proportion to the cube of the magnetic induction.

Rb Symbol for RUBIDIUM.

R_B Symbol for BASE RESISTANCE. (Also, r_B.)

RC **1.** Abbreviation of RESISTANCE-CAPACITANCE. **2.** Abbreviation of RADIO-CONTROLLED. **3.** Abbreviation of REMOTE CONTROL.

R_c **1.** Symbol for COLLECTOR RESISTANCE. (Also, r_C.) **2.** Symbol for COLD RESISTANCE.

RCA jack See PHONO JACK.

RCA plug See PHONO PLUG.

RC circuit See RESISTANCE-CAPACITANCE CIRCUIT.

RC-coupled amplifier See RESISTANCE-CAPACITANCE-COUPLED AMPLIFIER.

RC coupling See RESISTANCE-CAPACITANCE COUPLING.

RC filter See RESISTANCE-CAPACITANCE FILTER.

RCL **1.** Abbreviation of RECALL. **2.** Abbreviation of RESISTANCE-CAPACITANCE-INDUCTANCE.

RCM Abbreviation of RADAR COUNTERMEASURES. (Also, radCM.)

RC phase shifter See RESISTANCE-CAPACITANCE PHASE SHIFTER.

RC time constant See RESISTANCE-CAPACITANCE TIME CONSTANT.

RCTL Abbreviation of RESISTOR-CAPACITOR-TRANSISTOR LOGIC.

RC tuning See RESISTANCE-CAPACITANCE TUNING.

RCV Abbreviation for *receive.* (Also, rcv.)

RCVR Abbreviation for RECEIVER. (Also, rcvr, rx.)

rd Abbreviation for *rutherford.*

R & D See R AND D.

R_D Symbol for DRAIN RESISTANCE.

R_d **1.** Symbol for DIODE RESISTANCE. (Also, r_d.) **2.** Symbol for DISTRIBUTED RESISTANCE.

R-DAT Abbreviation of ROTARY DIGITAL AUDIO TAPE.

R_{dc} Symbol for DC RESISTANCE. (Also, r_{dc}.)

RDF Abbreviation of RADIO DIRECTION FINDER.

Re Symbol for RHENIUM.

R_e Symbol for EMITTER RESISTANCE. (Also, r_E.)

REA Abbreviation of *Rural Electrification Administration.*

reachthrough See PUNCHTHROUGH.

reachthrough region See PUNCHTHROUGH REGION.

reachthrough voltage See PUNCHTHROUGH VOLTAGE.

reactance Symbol, X. Unit, ohm. The opposition offered to the flow of alternating current by pure capacitance, pure inductance, or a combination of the two. Reactance introduces phase shift. Also see CAPACITIVE REACTANCE and INDUCTIVE REACTANCE. Compare RESISTANCE.

reactance chart A nomograph for capacitance, inductance, and frequency.

reactance factor The ratio of the alternating-current resistance of a conductor to the direct-current resistance. The reactance factor generally increases as the frequency increases because of skin effect and because the length of the conductor might be a sizable part of the wavelength of the transmitted energy.

reactance modulator A frequency modulator using a variable reactance, usually a varactor diode in the oscillator.

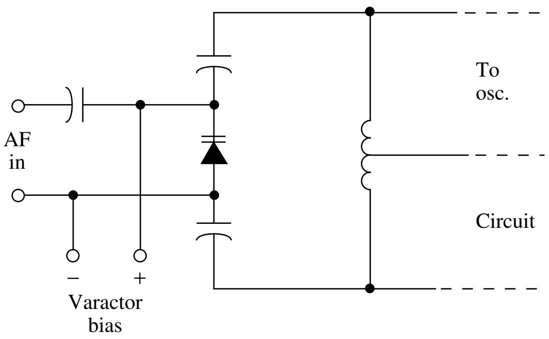

reactance modulator

reactance transistor A transistor used as a REACTANCE MODULATOR.

reaction-time meter See NEOMATACHOGRAPH and NEOMATACHOMETER.

reactive absorber In acoustics, a device that dissipates impinging sound waves by means of reflection, resonance, and other effects, besides dissipation in the form of heat.

reactive attenuator An attenuator that functions by means of reactance, rather than by means of resistance.

reactive current The component of alternating current that is not in phase with the voltage. Compare RESISTIVE CURRENT.

reactive kilovolt-ampere Abbreviation, kVAR. A unit of high apparent power; it is the product of kilovolts and amperes in a reactive component of a circuit. Also see APPARENT POWER, KILOVOLT-AMPERE, REACTIVE VOLT-AMPERE, and VOLT-AMPERE.

reactive load **1.** A load device that is capacitive or inductive, rather than resistive. **2.** A load device that contains reactance as well as resistance.

reactive power See REACTIVE KILOVOLTAMPERE and REACTIVE VOLT-AMPERE.

reactive volt-ampere Abbreviation, VAR. A unit of apparent power; it is the product of volts and amperes in a reactive component of a circuit. Also see APPARENT POWER, KILOVOLT-AMPERE, REACTIVE KILOVOLT-AMPERE, and VOLT-AMPERE.

reactor **1.** An inductor, especially one having very low internal resistance, used principally for its inductive reactance. **2.** A chamber in which the nuclei of atoms are split to provide atomic energy. Also see NUCLEAR REACTOR. **3.** In industrial chemistry, a vat in which reactions take place.

read **1.** In computer operations, to extract data from memory or a storage medium and (usually) transfer it to another area of memory or other medium. Compare WRITE. **2.** In digital communications, to transcribe data into printed form. **3.** In radiotelegraphy, to listen to Morse-code signals and comprehend the text without necessarily writing it down. **4.** To observe and note the indication of an instrument, such as a meter.

readability In electronic communications, the degree to which a desired signal can be recognized and interpreted in a given context.

readback In a multiplexer, a feature that facilitates inspection of the contents of the control latch.

reader A device that transcribes digital signals or markings into meaningful data. Examples: *Morse-code reader* and *bar-code reader.*

read head In a magnetic memory or in a tape recorder or wire recorder used for data recording, the head that picks up the magnetic pulses from the drum, tape, disk, or wire. Compare WRITE HEAD.

reading rate The number of input characters per second that a computer or other data-processing device handles.

read-only memory Abbreviation, ROM. In a computer or calculator, a memory unit in which instructions or data are permanently stored for use by the machine or for reference by the user. The stored information is read out nondestructively.

readout lamp An electron tube containing several cathodes, filled with a gas (such as neon), and used as a numeric or alphanumeric display device. Each cathode is connected to a separate pin on the base. A single anode is common to all cathodes. The cathode(s) to which a voltage is applied glow(s), showing the shape of a numeral, letter of the alphabet, or other symbol. In recent years, this type of display has been replaced by light-emitting diodes (LEDs) and liquid-crystal displays (LCDs).

readout pulse In a random-access memory (RAM), a pulse applied to the word line, facilitating readout of the information in a certain storage slot.

read pulse In computer operations, a pulse that activates the read function (see READ). Compare WRITE PULSE.

read rate The number of data units an input read device can transcribe per unit of time [e.g., *bits per second (bps)* and *words per minute (wpm)*].

readthrough **1.** The reception of signals between transmitted pulses at the same frequency. **2.** The continuous monitoring of a signal being jammed. Any change in the frequency, modulation, or other characteristics of the signal can then be detected, and the jamming signal adjusted accordingly.

read time The period during which data is being transferred from a computer storage unit.

read-write channel In computer operations, a channel over which activity between a central processing unit and a specific peripheral occurs.

read-write head An electromagnetic transducer used for both reading and writing data. See READ and WRITE.

read-write memory **1.** A small data storage bank for short-term use. The contents of the memory are easily changed. **2.** See RANDOM ACCESS MEMORY.

real address See ABSOLUTE ADDRESS.

real axis The axis of the real-number component of a COMPLEX NUMBER (i.e., the horizontal axis in an ARGAND DIAGRAM).

real component The real-number part of a COMPLEX NUMBER.

real image The image formed on a screen when rays from the object converge on passing through a lens. Compare VIRTUAL IMAGE.

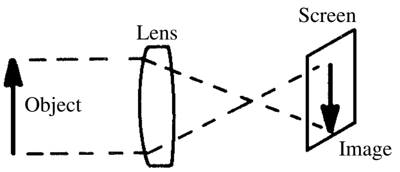

real image

real number A number in the category that includes zero, all rational numbers, and all irrational numbers. Also see COMPLEX NUMBER, IMAGINARY NUMBER, IRRATIONAL NUMBER, and RATIONAL NUMBER.

real power The apparent power multiplied by the power factor in an alternating-current circuit containing reactance. Real power is the difference between the apparent power and the reactive power. Actual radiated or dissipated power cannot exceed the real power.

real time Pertaining to the operation of a computer, communications, or data processing system in which events are represented or acted on as they occur. Data is processed as it becomes available, usually through the use of time-sharing, direct-access storage devices, and remote terminals.

real-time clock A device that produces periodic signals that reflect the interval between events. It is sometimes used to indicate time of day.

rear end The low-frequency portion of a superheterodyne receiver (i.e., the intermediate-frequency amplifier, second detector, and audio-frequency amplifier). Compare FRONT END.

rear projection A method of big-screen television picture reproduction. The image is focused onto a translucent, flat surface. The viewer is positioned on the side of the screen opposite to the projecting beam.

Reaumur scale A thermometer scale on which zero is the freezing point of water and 80 degrees is the boiling point of water. Compare ABSOLUTE SCALE, CELSIUS SCALE, FAHRENHEIT SCALE, and RANKINE SCALE.

rebecca The airborne interrogator in the British REBECCA-EUREKA SYSTEM.

rebecca-eureka system A British 90-mile-hovering radar navigation system that consists of an airborne interrogator (rebecca) and a ground transponder beacon (eureka).

rebroadcast The retransmission of a radio broadcast simultaneously by a station other than the originator. Also see AUTOMATIC RELAY STATION.

rebroadcast station See AUTOMATIC RELAY STATION.

recalescence During the cooling of a metal, the sudden release of heat. Also see RECALESCENT POINT. Compare DECALESCENCE.

recalescent point In a metal whose temperature is being lowered from a higher value, the temperature at which heat is suddenly released. Compare DECALESCENT POINT.

recall Abbreviation, RCL. In computers and calculators, an instruction that brings material from the memory for examination or use. The opposite instruction is STORE.

receiver **1.** A device or system operated at the destination end of a communication link; it accepts a signal and processes or converts it for local use. Also see specific entries for various types of receiver. **2.** The earpiece of a telephone. **3.** A radio broadcast-band tuner integrated with a general-purpose preamplifier and power amplifier, and containing standard jacks for input and output of audio signals to and from peripheral equipment.

receiver detector In a wireless communications or broadcast receiver, a circuit that extracts the information from the signal. The design depends on the type of emission to be received. See also DISCRIMINATOR, ENVELOPE DETECTOR, PHASE-LOCKED LOOP, PRODUCT DETECTOR, RATIO DETECTOR.

receiver dynamic range A quantitative measure of the ability of a wireless receiver to maintain a fairly constant output, and yet to maintain its rated sensitivity, in the presence of signals ranging from very weak to extremely strong. This figure is specified in decibels. It is typically 100 dB or more in a well-engineered communications receiver.

receiver front end The portion of a wireless communications or broadcast receiver consisting of the first radio-frequency (RF) amplifier, and often also including bandpass filters between this amplifier and the antenna. The dynamic range and sensitivity of a receiver are determined by the performance of this stage. These two characteristics are among the most important for any receiver.

Low-noise, high-gain amplifiers are the rule. Field-effect transistors are commonly used.

receiver IF chain In a wireless communications or broadcast receiver, the series of radio-frequency (RF) amplifier stages in which most of the gain takes place. These stages are also where the best possible RF selectivity is obtained. The intermediate frequency (IF) on which the amplifiers works is a constant frequency. This simplifies the design of the amplifiers to produce optimum gain and selectivity. Crystal-lattice filters or mechanical filters are commonly used in these stages to obtain the desired bandwidth and response.

receiver mixer In a superheterodyne wireless communications or broadcast receiver, a stage that converts the variable input signal frequency to a constant intermediate frequency (IF), or a stage that converts the first IF to a second IF usually having a lower frequency. This type of circuit is nonlinear, and combines the signal with a carrier from a local oscillator (LO). The output is either the sum or the difference of the signal frequency and the LO frequency.

receiver muting See MUTING, **1**.

receiver noise figure A quantitative measure of the ability of a wireless receiver to respond to desired signals while rejecting unwanted noise. This can be quantified in various ways. In general, the lower the noise figure, the better the sensitivity. Gallium–arsenide field-effect transistors (GaAsFETs) are well known for the low levels of noise they generate, even at quite high frequencies. Other types of FETs can be used at lower frequencies. Bipolar transistors tend to be noisy. See also NOISE FACTOR, NOISE FIGURE.

receiver post-detector stages In a wireless communications or broadcast receiver, one or more stages of amplification and/or filtering employed to refine the detector output to a form suitable for feeding a speaker, headset, printer, fax machine, slow-scan television picture tube, computer, electromechanical device, or other peripheral equipment.

receiver primaries See DISPLAY PRIMARIES.

receiver selectivity A quantitative measure of the ability of a wireless receiver to respond to a desired signal, but not to undesired ones. The frequency window is established by a preselector in the early RF amplification stages, and is honed to precision by bandpass filters in later amplifier stages. The preselector passes energy within a range of about plus-or-minus 10 percent of the signal frequency; other frequencies are attenuated. This reduces the chance for strong, out-of-band signals to impair the performance of the receiver. The narrowband filter in the final intermediate-frequency (IF) stage responds only to energy within the actual signal band. This minimizes adjacent-channel interference. In some receivers, yet another bandpass filter is used in the audio-amplifier stages.

receiver sensitivity A quantitative measure of the ability of a wireless receiver to recover weak signals and process them into readable data. The most common expression is the number of signal microvolts that must exist at the antenna terminals to produce a certain signal-to-noise ratio (S/N). Sometimes, the signal-plus-noise-to-noise ratio (abbreviated S+N/N) is given. The front end, or first RF amplifier stage, of a receiver is the most important stage with regard to sensitivity. Sensitivity is directly related to the gain of this stage, but the amount of noise the stage generates is even more significant. A good front end should produce the best possible S/N or S+N/N ratio at its output. All subsequent stages amplify the front-end noise output as well as the front-end signal output.

receiving set RADIO RECEIVER.

receiving station A station that ordinarily only receives signals (i.e., it makes no type of transmission). Compare TRANSMITTING STATION.

receptacle **1.** See SOCKET. **2.** The half of a connector that is mounted on a support, such as a panel, and that is therefore stationary.

recharge In certain cells and batteries, the restoration of chemical energy following use so that the device is ready to deliver its full rated electric current. Also see RECHARGEABLE.

rechargeable Pertaining to a secondary cell or battery that can accept a restoration of chemical energy following use, and thus can be completely charged and discharged numerous times. Examples: *nickel-metal-hydride (NiMH) battery* and *lead-acid battery.*

reciprocal impedances See INVERSE IMPEDANCES.

reciprocal ohm See SIEMENS and MHO.

reciprocation **1.** The determination of a mathematical reciprocal value from a given value. **2.** The transmission of a message in response to a received message.

reciprocity in antennas See RAYLEIGH-CARSON THEOREM.

reciprocity theorem When a voltage E across branch A of a network causes a current I to flow in branch B of the network, the voltage can be applied across branch B to cause the same value of current to flow in branch A. Compare COMPENSATION THEOREM, MAXIMUM POWER TRANSFER THEOREM, NORTON'S THEOREM, SUPERPOSITION THEOREM, and THEVENIN'S THEOREM.

recombination The refilling of holes by electrons in a semiconductor.

recombination current In a transistor circuit, base current resulting from recombination.

recombination rate In a semiconductor material, the speed at which the electrons and holes recombine. It can be expressed as the time required for a certain proportion of charge carriers to recombine.

recompile In computer operations, to COMPILE again, usually according to program amendments following debugging, or to create a different form of a program so that it will be compatible with other hardware.

record **1.** See PHONOGRAPH DISC. **2.** A chart delivered by a graphic recorder. **3.** To make one of the foregoing. **4.** In data processing, a constituent of a file. **5.** In data processing, a data unit portraying a specific transaction.

record blocking In data processing operations, making data blocks from groups of records so that the blocks can, in a single operation, be transferred to a nonvolatile storage medium, such as diskette or tape.

record count A usually running total of a file's records.

recorded disc A phonograph disc on which a recording has been made. Also called PRERECORDED DISC.

recorded tape Magnetic tape containing recorded material. Also called PRERECORDED TAPE. Compare BLANK TAPE.

recorder **1.** A machine for preserving sound, video, or data signals in the sequence in which they occur (e.g., DISC RECORDER, TAPE RECORDER, and WIRE RECORDER). **2.** A machine for making a permanent visual record (photographically or by stylus) of an electrical phenomenon. Examples: DRUM RECORDER and OSCILLOGRAPH.

record head See RECORDING HEAD.

recording density In a magnetic storage medium, the number of information units (bits, bytes, etc.) represented by magnetized areas, per unit area or length.

recording disc A phonograph record on which material has not been recorded, or from which recorded material has been removed. Compare PRERECORDED DISC.

recording head In a magnetic recorder/reproducer, the head that magnetizes the medium in accordance with sounds or other signals. Also called RECORD HEAD and WRITE HEAD. Compare PLAYBACK HEAD.

Recording Industry Association of America Abbreviation, RIAA. An organization that sets standards for audio recording and reproduction in the United States.

recording instrument A measuring instrument, such as a voltmeter or ammeter, that makes a permanent record of its deflections. Also see RECORDER, **2**.

recording loss **1.** Loss of data during a recording process. **2.** Loss resulting from recording efficiency of less than 100 percent; audio power loss.

recording tape Magnetic tape on which nothing has been recorded, or from which all data has been erased. Compare PRERECORDED TAPE.

recovery time **1.** Symbol, t_r. The time required for a semiconductor pn junction to attain its high-resistance state when the bias voltage is suddenly switched from forward to reverse. **2.** The time required for a circuit to recover from momentary overdrive. **3.** The time required for a computer system to stabilize following a degenerative operation. **4.** The time required for switching a memory from the write to the read mode. It is measured as the length of time from switching out of the write mode until meaningful signals occur at the output. **5.** In a transceiver, the time required from the completion of a transmitted signal until the receiver is activated.

rect Abbreviation of RECTIFIER.

rectangular coordinates See CARTESIAN COORDINATES.

rectangular scan **1.** A method of beam scanning in a cathode-ray tube, in which the beam moves sequentially in parallel lines to cover a rectangular region. Used in television. **2.** In radar, a two-dimensional scan, covering a specific rectangular region.

rectangular wave An alternating or pulsating current or voltage whose rise and decay times are essentially zero, and whose maxima and minima are essentially flat, but not necessarily of equal duration. The SQUARE WAVE is a special type of rectangular wave.

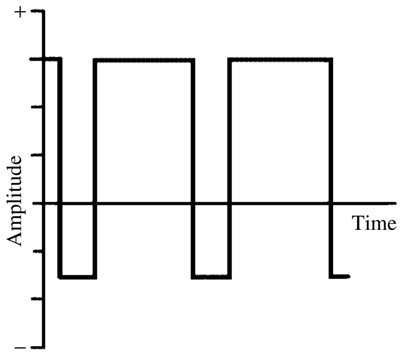

rectangular wave

rectangular waveguide A waveguide having a rectangular cross section.

rectification The conversion of alternating current into pulsating direct current by any means other than the use of a motor-generator. Also see RECTIFIER.

rectification efficiency The ratio (expressed as a percentage) of the direct-current output voltage to the peak alternating-current input voltage of a rectifier.

rectified alternating current The unfiltered, pulsating direct-current output of a rectifier. It consists of the unidirectional half-cycles passed by the rectifier (one per cycle for half-wave rectification, and two per cycle for full-wave rectification).

rectifier Abbreviation, rect. An electronic or electromechanical device that converts alternating current into pulsating direct current.

rectifier diode A heavy-duty tube or semiconductor diode designed primarily to change alternating current to pulsating direct current in power supplies.

rectifier filter A circuit containing parallel capacitance, sometimes in combination with series inductance, intended for smoothing out the ripple in the output of a power-supply rectifier.

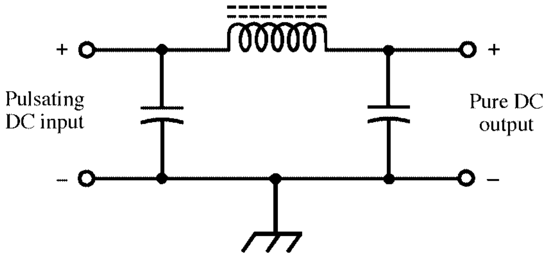

rectifier filter

rectifier-filter system The rectifier plus power-supply-filter combination for converting alternating current into direct current.

rectifier photocell A photovoltaic cell consisting of two layers of material with a semiconductor junction between them. The device produces direct current when exposed to visible light, infrared, or ultraviolet radiation.

rectifier probe A diode-type probe used with a direct-current (dc) voltmeter to measure radio-frequency (RF) voltage. The diode rectifies the RF signal and presents to the meter a dc voltage proportional to the peak RF voltage.

rectifier stack An assembly of separate rectifier disks or plates in series on a central bolt, as in most selenium rectifiers.

rectifier tube A two-element electron tube, once commonly used for converting alternating current into pulsating direct current in high-voltage, high-current power supplies.

rectifier-type meter See DIODE-TYPE METER.

rectilinear chart A graphic-recorder chart in which the crossing coordinates are arcs, rather than straight lines, to correspond to the swing of the pen. Also see STRIP CHART.

rectilinear scan See RECTANGULAR SCAN, **1**.

recurrent network A circuit in which several sections of identical configuration (e.g., L sections) are cascaded.

recurrent phenomenon A phenomenon that repeats itself periodically.

recurrent sweep In an oscilloscope, a repetitive horizontal sweep of the beam occurring at a frequency determined by the settings of the sweep controls. Also called REPETITIVE SWEEP. Compare NONREPETITIVE SWEEP.

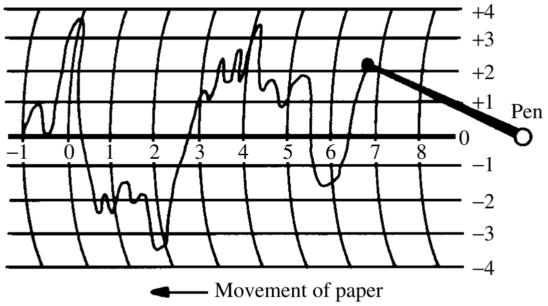

rectilinear chart

recursion **1.** Generating a complete sequence of functions or numbers by applying an algorithm to initial values in the sequence. **2.** In computer programming and artificial intelligence, a logical process containing loops in calculation or reasoning.

recursive Relating to a procedure or set of steps that repeat endlessly.

Red Book The first format developed for compact-disc data storage media, developed by Sony and Philips. It is commonly used in digital audio systems. See also CD-ROM, GREEN BOOK, ORANGE BOOK, and YELLOW BOOK.

red-green-blue Abbreviation, RGB. In video applications, the three primary colors from which all other colors are derived. Also see COLOR TELEVISION.

red gun In a three-gun color-television picture tube, the electron gun whose (correctly adjusted) beam strikes only the red phosphor dots on the screen.

red oxide of iron An iron oxide of the general formula Fe_2O_3, used as the magnetic coating of recording tape. Also see IRON OXIDE.

red oxide of zinc See ZINCITE.

red-tape operation An operation or function needed for organizational purposes, but that does not directly contribute to the completion of the task at hand.

reduced instruction set computer Abbreviation, RISC. A computer architecture in which program instructions are simplified to obtain enhanced processing speed. It is useful especially in complex graphics, animation, multimedia, and scientific work requiring many calculations.

reductio ad absurdum A method of obtaining a conclusion by proving that its negation results in a contradiction. It is sometimes used in computer programming involving mathematical proofs.

reduction In an electrochemical cell or battery, a transfer of electrons to the active chemical.

reductionism The theory that all human thought processes, including emotion and intuition, can be reduced to digital logic, and thus can be duplicated by a sufficiently powerful computer. It is of interest to researchers in artificial intelligence.

reductionist A person who subscribes to the theory of REDUCTIONISM.

redundancy **1.** The repetition of components in a circuit (e.g., series or parallel connection of them) so that one will be available for circuit operation if the other fails. **2.** Having available more than one method for performing a function. **3.** Having on hand several copies of data as a safeguard against data loss.

redundancy check A check for the integrity of digitized data to which extra bits have been added for the purpose (e.g., *parity check*).

redundant **1.** Pertaining to any two units of data that resemble each other in such a manner that if either unit is removed, no information is lost from the system. **2.** A unit of data that contains information already present in the system.

redundant array of independent disks Acronym, RAID. A set of data storage media used to store video programs.

red video voltage In a three-gun color-television circuit, the red-signal voltage that actuates the red gun.

reed A usually thin metal blade, leaf, or strip used in vibrators, reed-type relays, reed-type oscillators, and similar devices.

reed oscillator See REED-TYPE OSCILLATOR.

reed relay See DRY-REED SWITCH and MERCURY-WETTED REED RELAY.

reed-relay logic Logic circuits using reed relays. Also see RELAY LOGIC.

reed switch **1.** A frequency-sensitive switch in which the movable contact is mounted on the tip of a thin, metal strip (reed). The reed is actuated by an alternating-current (ac) coil. The reed closes the contacts when the ac excitation is at its natural frequency. **2.** See DRY-REED SWITCH.

reed-type oscillator An electromechanical audio-frequency oscillator whose frequency is controlled by a vibrating metal strip (reed) instead of a tuning fork. Also see HUMMER.

reed-type switch See REED SWITCH.

reel **1.** The spool around which a magnetic tape or video film is wound. **2.** A spool containing magnetic tape or video film.

reentrant cavity A resonant cavity in which one or more sections are directed inward to confine the electric field to a small volume.

reentrant winding A winding of wire that returns to its starting point—especially in a motor armature.

ref Abbreviation of REFERENCE.

reference address As a point of reference, an address for instructions having relative addresses.

reference amplifier A voltage-regulation device consisting of a transistor and Zener diode in the same envelope.

reference angle In radar, the angle of incidence of the beam against a target surface, measured with respect to the normal (perpendicular) line at the surface.

reference antenna A standard antenna, such as an isotropic radiator or a half-wave dipole, used to establish a reference for determining the relative gain of another antenna.

reference bias current In a reference amplifier, the input current that subtracts from the reference current. It is generally measured in microamperes.

reference current range In a digital-to-analog converter, the difference between the maximum and minimum reference current for which the device is within specifications for resolution.

reference diode A Zener diode whose constant voltage drop is used as a direct-current reference potential in calibrator circuits and voltage regulators.

reference dipole See REFERENCE ANTENNA.

reference electrode For use with a pH meter, an electrode that provides a reference potential.

reference input slew rate In a digital-to-analog converter, the average rate of change in output for a given change in the reference input. It is expressed in milliamperes or microamperes per microsecond.

reference level A specific value of a quantity (e.g., current, frequency, power, or voltage) to which other values of the same quantity are referred.

reference time The point at which a trigger pulse attains 10 percent of its maximum amplitude.

reference tone A standard audible tone of known frequency [e.g., 440 Hz (representing A below middle C)]. Sometimes the intensity as well, as the frequency, is specified.

reference white level The television picture signal value representing the uppermost limit for peak white signals.

R_{eff} Symbol for EFFECTIVE RESISTANCE.

reflectance **1.** See MISMATCH FACTOR. **2.** The reflected part of the radiant flux striking a surface. It is expressed as a fraction of the total incident radiation.

reflected binary code See CYCLIC CODE.

reflected electromagnetic field In a transmission line, the electromagnetic energy not absorbed by the load when an impedance mismatch exists between the load and the line. See INCIDENT POWER and REFLECTED POWER.

reflected impedance In a coupled circuit, the impedance in the secondary that appears in the primary circuit, or vice-versa, as if it were reflected through the coupling transformer.

reflected power In a transmission line not perfectly matched to a load at the feed point, an expression of the amount of electromagnetic field reflected from the feed point, rather than absorbed by the load. In general, this can be expressed in watts or as a percentage of the incident power. It is not a true indicator of the loss caused by the mismatch because the reflected field is usually all returned again when it arrives back at the transmitter.

reflected-power meter A radio-frequency instrument, connected between a source and a load, that can measure INCIDENT POWER and REFLECTED POWER.

reflected ray The ray that is reflected by the surface of a body or region it strikes. Compare INCIDENT RAY and REFRACTED RAY.

reflected resistance **1.** In a transformer, the effective resistance across the primary winding when a resistive load is connected to the secondary. **2.** In a transmission line, the resistance at the input end when a load is connected to the output end.

reflected wave **1.** An electromagnetic wave reflected by the ionosphere or by the surface of the earth. Compare INCIDENT WAVE and REFRACTED WAVE. Also see IONOSPHERE and IONOSPHERIC PROPAGATION. **2.** A wave that is bounced off an obstruction, such as a building or mountain.

reflecting galvanometer A galvanometer having a light-beam pointer.

reflecting shell See IONOSPHERE.

reflection **1.** The turning back of a ray by a surface it strikes. Examples of reflecting media are the surface of the earth, the polished surface of a material, and a layer of the ionosphere. Compare REFRACTION. **2.** The return of energy to the source by the mismatched end of a transmission line or by the end of a radiator.

reflection error In a radar, radionavigation, or radiolocation system, an error in the reading caused by reflections of the signal from objects other than the intended signal source or object.

reflection factor See MISMATCH FACTOR.

reflection law When a ray strikes a smooth reflecting surface, the angle of incidence is equal to the angle of reflection.

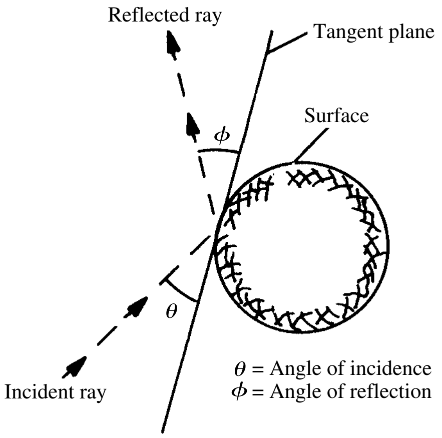

reflection law

reflection loss **1.** Loss caused by the reflection of an electromagnetic field at a discontinuity in a transmission line. **2.** Loss that occurs when an electromagnetic wave is reflected from a surface or object.

reflection phase grating A device that diffuses sound waves by diffraction effects. The acoustic equivalent of an electromagnetic-wave DIFFRACTION GRATING.

reflective code See GRAY CODE.

reflectivity **1.** See MISMATCH FACTOR. **2.** The degree to which a point, plane, or surface reflects the radiation (light, for example) that strikes it.

reflectometer **1.** See REFLECTED-POWER METER. **2.** A type of photometer used to measure reflection.

reflector **1.** A smooth, metal surface or wire screen for reflecting radio waves. See, for example, PARABOLIC REFLECTOR. **2.** A length of wire, rod, or tubing used in a parasitic antenna to reflect radio waves. Compare DIRECTOR and RADIATOR. **3.** A polished surface for reflecting visible light or infrared rays (i.e., a mirror). **4.** See REPELLER.

reflector element See REFLECTOR, **2**.

reflector satellite A satellite whose skin reflects radio waves.

reflector voltage In a reflex Klystron, the reflector-to-cathode voltage.

reflex baffle A loudspeaker BAFFLE constructed so that some of the sound radiated to the rear of the diaphragm is transmitted forward (after phase shift) to boost acoustic radiation at some frequencies.

reflex bunching In a Klystron, electron bunching following direct-current-field-induced reversal of the velocity-modulated electrons. Also see REFLEX KLYSTRON.

reflex circuit A radio receiver circuit in which a single transistor is used successively for different functions. For example, one active device can act as a mixer and as a radio-frequency amplifier.

reflex Klystron A Klystron having only one cavity. This cavity serves first as the buncher and then, as the electrons are turned around and caused to pass through again, as the catcher.

refracted ray The ray that is refracted by a body or region through which it passes. Compare INCIDENT RAY and REFLECTED RAY.

refracted wave An electromagnetic wave that is refracted by the ionosphere. Compare INCIDENT WAVE and REFLECTED WAVE. Also see IONOSPHERE and IONOSPHERIC PROPAGATION.

refraction The bending of an energy ray as it passes through media that cause a change in the speed of propagation. It can occur with radio waves, infrared, visible light, ultraviolet, X rays, gamma rays, and sound waves.

refractive index See INDEX OF REFRACTION.

refractivity The extent of the ability to refract, given as the quantity $(v_1/v_2) - 1$, where v_1 is the phase velocity in free space, and v_2 is the phase velocity in the medium through which a wave passes.

refractory A heat-resistant, nonmetallic ceramic material.

refrigerator A chamber used to maintain a circuit or component at a constant temperature that is lower than the ambient temperature. This device is analogous to the oven, which maintains a higher temperature than the surrounding medium. A refrigerator can be used to maintain precise frequency for a reference oscillator.

regeneration **1.** The processing of a distorted signal so that it has its original characteristics. **2.** Positive feedback generally used for the purpose of causing oscillation, or for detection in a regenerative receiver. See POSITIVE FEEDBACK.

regeneration period The period during which the electron beam scans a cathode-ray tube screen to restore changes to the screen surface.

regenerative amplifier An amplifier that uses regeneration to increase its gain and/or selectivity.

regenerative detector A detector provided with regenerative feedback. Although such a detector is sensitive, it can be unstable. Compare NONREGENERATIVE DETECTOR.

regenerative feedback Feedback producing regeneration (i.e., positive feedback). Compare DEGENERATIVE FEEDBACK.

regenerative IF amplifier An intermediate-frequency amplifier in which regeneration is introduced to boost sensitivity and, sometimes, selectivity.

regenerative reading A method of reading data (see READ) so that it is automatically restored, by writing, to locations from which it came.

register In computer systems, an arrangement of several storage devices, such as flip-flops, for storing a certain number of digits (a two-bit register, for example, requires two flip-flops).

register capacity The range of values for quantities that can be handled by a register.

registered professional engineer A title granted by a state board of examiners to a person licensed to work as an engineer.

register length The number of characters or bits that can be held in a register, according to its capacity.

registration The accurate alignment of terminals or other points on different components or on opposite sides of a board so that when the surfaces containing those points are overlaid, all points mate precisely.

regulated power supply A power supply whose output is held automatically to a constant level or within a narrow range, regardless of loading variations.

regulating transformer See VOLTAGE-REGULATING TRANSFORMER.

regulation **1.** In general, the adjustment or control of a component, device, or system. **2.** Automatic control. See, for example, SELF-REGULATION. **3.** See CURRENT REGULATION. **4.** See VOLTAGE REGULATION.

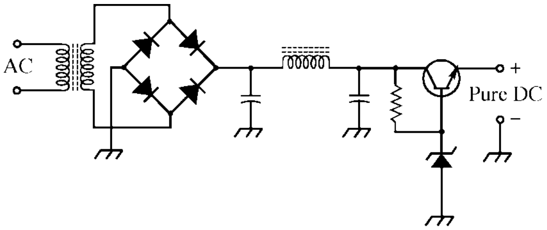

regulated power supply

regulator **1.** A device that automatically holds a quantity to a constant value (e.g., a voltage regulator). **2.** A device via which a quantity can be varied (e.g., *potentiometer*, *rheostat*, and *variable autotransformer*).

regulator diode A semiconductor diode—especially a Zener diode used as a two-terminal voltage regulator.

reinitialization The setting of all lines in a microcomputer or microprocessor to logic zero automatically when power is removed, then reapplied.

reject amplifier A tuned amplifier having the response of a band-suppression filter. Like the filter, the amplifier rejects or severely attenuates one frequency (or band of frequencies) while readily passing lower and higher frequencies. Compare PASS AMPLIFIER.

reject filter See REJECTION FILTER.

rejection circuit A circuit performing the function of a REJECTION FILTER.

rejection filter A filter that suppresses one frequency (or band of frequencies) while passing all other frequencies.

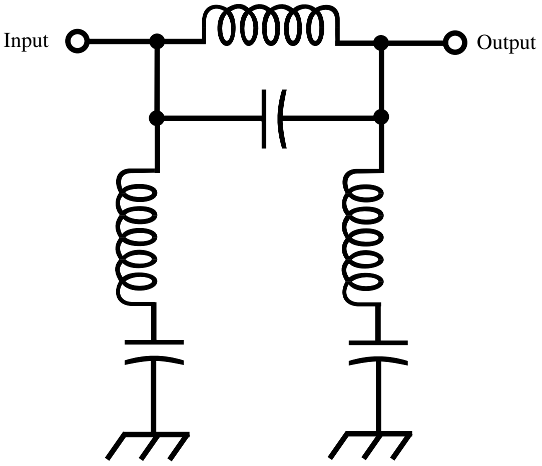

rejection filter

rejection notch A sharp dip in the transmission characteristic of a crystal filter. It provides rejection-filter action at the notch frequency. Also see CRYSTAL RESONATOR and REJECTION FILTER.

rejectivity The degree to which a selective circuit rejects an unwanted signal. Compare TRANSMITTIVITY.

rejuvenation See REACTIVATION.

rel Symbol, R. The cgs unit of reluctance, equivalent to gilberts per maxwell.

relative accuracy In a measuring instrument, the error determined as a percentage of the actual value; the difference between the actual and measured values, divided by the actual value, then multiplied by 100.

relative address In the address part of a computer program instruction, a number specifying a location relative to a BASE ADDRESS. When the base address is added to the relative address, it yields the ABSOLUTE ADDRESS.

relative error The ratio of the absolute error to the exact value of a quantity.

relative gain The current, voltage, or power gain, measured, with respect to a reference standard.

relative humidity Abbreviation, rh. The ratio, as a percentage, of the amount of moisture in the air to the amount the air could contain at a given temperature. Compare ABSOLUTE HUMIDITY.

relative luminosity Luminosity measured with respect to a reference level.

relative permeability The ratio of the permeability of a given material to the permeability of another material (or of the same material under different conditions).

relative power Power level specified with respect to another (often reference) power level.

relative uncertainty The uncertainty of a measurement divided by the measured value. The maximum value that this quotient can have is 1. Also see UNCERTAINTY IN MEASUREMENT.

relative visibility Response of the human eye to light. This is relative because the eye does not see equally well throughout the visible spectrum. The peak response of the human eye is around 5.4×10^{14} Hz; this represents yellow-green light. Photoelectric devices have peak responses that can differ considerably from this value.

relativity theory See EINSTEIN'S THEORY.

relaxation A delayed change in circuit conditions, as a result of change in the input.

relaxation inverter An inverter circuit in which the direct-to-alternating-current conversion device is a RELAXATION OSCILLATOR.

relaxation oscillator An oscillator whose operation results from the buildup of a charge in a capacitor, followed by sudden discharge of the capacitor, the sequence being repeated periodically. In one circuit, a capacitor is connected in series with a resistor and a direct-current power supply, and a neon bulb is connected in parallel with the capacitor. The output is a sawtooth wave.

relaxation time **1.** The time required for an exponentially decreasing variable to fall to $1/e$ (approximately 36.8 percent) of its initial value, where e is the natural-logarithm base (approximately 2.71828). **2.** For a gas, the time required for it to return to its original state after having been disturbed.

relay **1.** A signal-actuated switching device. In most instances, a relatively weak current or voltage is used to make the relay switch a higher current or voltage. A relay can be electromechanical or fully electronic (no moving parts). See, for example, ELECTROMECHANICAL RELAY and ELECTRONIC RELAY. **2.** A repeater station. **3.** In communications, to receive a message and retransmit it en route from a source to a destination.

relay amplifier See RELAY DRIVER.

relay booster See RELAY DRIVER.

relay driver A direct-current amplifier (usually one stage) used to actuate an electromechanical relay in response to a low-powered signal.

relay flip-flop See BISTABLE RELAY.

relay logic Abbreviation, RL. In computer and industrial-control operations, a logic system using electromechanical relays as flip-flops (see BISTABLE RELAY).

relay transmitter See AUTOMATIC REPEATER STATION.

release time **1.** The interval between the instant power is removed from a relay and the instant the armature is released sufficiently to operate the contacts. **2.** The time between one control input becoming inactive and another becoming active. **3.** The time required for reception to resume in a transceiver, once transmission has stopped. **4.** An expression for the rapidity with which an automatic gain control reverts to maximum sensitivity following the reception of a strong signal.

reliability **1.** The dependability of operation of a device or circuit under specified conditions. **2.** The proportion of units that still work after a set of units has been in use for a specified length of time.

reliability engineering The branch of engineering devoted to the theory and application of reliability; based on fundamental engineering and advanced statistical concepts.

reluctance Symbol, R. SI unit, A/Wb; cgs unit, rel. In a magnetic circuit, the opposition to the establishment of a magnetic field; it is analogous to resistance in electric circuits.

reluctance motor An electric motor having a squirrel-cage rotor with some of its teeth ground down, and a shaded-pole or split-phase type of stator that supplies a rotating magnetic field. When starting, this motor comes up to speed like an induction motor, but the protruding teeth of the rotor then follow the field in the manner of the poles of a hysteresis motor.

reluctivity Specific reluctance (i.e., the reluctance of a sample of magnetic material one centimeter long and one square centimeter in cross section). Reluctivity is the reciprocal of permeability.

rem Acronym for *roentgen equivalent man*, an amount of ionizing radiation having the same effect on the body as a one-roentgen dose of gamma or X radiation.

remagnetizer A magnetizer used principally to restore weakened permanent magnets.

remainder **1.** The result of subtracting one quantity (the subtrahend) from another (the minuend). Also called DIFFERENCE. **2.** In division, the numerical value left over after the integral part of the quotient has been determined; it becomes the fractional part when divided by the divisor. For example, in 25/3, the remainder is 1.

remanence See RESIDUAL MAGNETISM.

remanent flux density See REMANENCE.

remodulator Any device that changes the modulation of a signal from one form to another, such as from frequency modulation to amplitude modulation, without loss of intelligence.

remote alarm In security systems, an alarm that occurs at a location different from where an intrusion occurs (e.g., at the headquarters of a security company).

remote control Control of distant devices by mechanical means or by radio-frequency signals sent from a transmitter especially designed for the purpose; in the latter case, it is sometimes called *radio control.*

remote-control receiver The complete device that selects, amplifies, and demodulates or rectifies a radio signal for control of a circuit or mechanism at a distance from the transmitter of the control signal. Some receivers have self-contained antennas.

remote-control transmitter The complete device that generates radio-frequency power, adds to it the signals needed for remote control, and radiates the modified power.

remote-control system The complete set of hardware units and software programs facilitating the operation of a computer or robot from a distance.

remote data terminal In a computer system, a terminal connected to the central processor by a telephone line or radio link. It is used for the transfer of data without providing control of the system. Also called *remote data station.*

remote error sensing A method of regulation used in some power supplies. The voltage across the load, or the current through the load, is determined by remote control. The power-supply output is adjusted to compensate for losses in the system.

remote job entry In computer operations, the keying-in of input data at a site physically distant from the central processor.

remote tuning The electrical or radio tuning of a circuit or device from a distance.

rendering In three-dimensional computer animation, the software process in which all the aspects of the model are combined to obtain the presentation.

rep **1.** Acronym for *roentgen equivalent physical*, an amount of ionizing radiation that, upon absorption by body tissue, will develop the energy of a one-roentgen dose of gamma or X radiation. **2.** Colloquial abbreviation for *repetition*, as in *rep rate.* **3.** Colloquial abbreviation for *representative*, as in *service representative.*

repeatability The ability of an instrument, system, or method to give identical performance or results in successive instances.

repeater A receiver/transmitter device that retransmits a signal it receives from another source, often simultaneously. In this way, a signal can be transmitted on several frequencies, or the service area of the original station can be extended. Also see ONE-WAY REPEATER and TWO-WAY REPEATER.

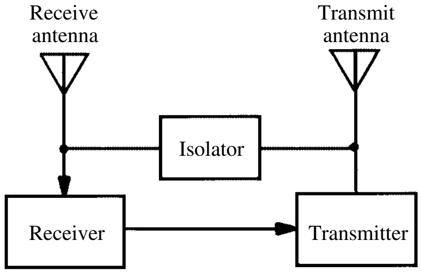

repeater

repeater station See AUTOMATIC REPEATER STATION and REPEATER.

repeating decimal A decimal fraction in which groups of digits recur endlessly (e.g., 25/99 = 0.252525. . .).

repeller An electrode, especially in a velocity-modulated tube, for reversing the direction of an electron beam.

repertoire The instruction set for a particular object or source computer programming language.

repetition instruction In a loop in a computer program, an instruction that causes the repetitive implementation of one or more instructions.

repetitive phenomenon See RECURRENT PHENOMENON.

repetitive sweep See RECURRENT SWEEP.

replication In an electronic system, especially in a computer, the redundancy of hardware units to provide standby facilities in case of failure.

replacement A component or circuit that can be substituted directly for another; it fits exactly into place and functions exactly like the component it replaces, without modification to the equipment.

report **1.** The results of testing and evaluation of a device, organized into a written document. **2.** The output of a computer, printed on paper for permanent reference.

report program generator Abbreviation, RPG. A computer programming language with which programs can be produced for the generation of business reports.

reproduce head See PLAYBACK HEAD.

reproducing stylus A stylus for the playback of material from a phonograph disc.

reproduction **1.** The recovery of data from storage, and its presentation in original form. **2.** Data obtained by the process defined in **1**. **3.** See PLAYBACK.

reproduction loss See PLAYBACK LOSS.

repulsion A force that pushes objects away from each other, as between similar electric charges or similar magnetic poles. Compare ATTRACTION. Also see LAW OF REPULSION.

repulsion-induction motor An alternating-current motor arranged to start as a REPULSION MOTOR and run as an INDUCTION MOTOR, but with better regulation than that of the latter.

repulsion motor An alternating-current motor having an armature and commutator similar to those of a direct-current motor, and a stator similar to that of a split-phase motor, without the auxiliary starting winding. Repulsion caused by the negative half-cycle of torque is utilized to drive the armature, by placing the brushes in such a way that they close the coils only when the latter are in position to receive this repulsive action.

repulsion-start motor An alternating-current motor that starts as a REPLUSION MOTOR but at approximately 75 percent of full speed. Its commutator is automatically short-circuited and the motor runs as an INDUCTION MOTOR. Also see REPULSION INDUCTION MOTOR.

request slip In computer operations, peripheral and memory needs for a program given in a written statement.

reradiation Radiation of energy by a body that has been exposed to radiation, as when a receiving antenna retransmits a signal.

rerecording A recording of played-back material.

reroute **1.** In computer operations, to establish new channels between peripherals and main memory. **2.** To establish new circuit paths, physically (as by changing conductor orientation) or electronically (as by selecting an alternate signal bus).

rerun See ROLLBACK.

res **1.** Abbreviation of RESISTANCE or RESISTOR. (Also, R and r.) **2.** Abbreviation of RESEARCH. **3.** Abbreviation of RESOLUTION.

reset **1.** The clearing of a flip-flop of data in storage (i.e., the setting of the flip-flop to its zero state). **2.** In a computer program, an instruction to initialize the value of a variable. **3.** In a security system, a function that terminates an alarm signal following an intrusion, and renders the system operational again so that it can detect subsequent intrusions should they occur.

reset action **1.** The return of a circuit or device to its normal operating condition. **2.** A method of adjusting a circuit to compensate for the severity of an abnormal condition. The extent of readjustment is determined by the extent of the departure from normal conditions.

reset generator A circuit or device that generates a pulse for resetting a flip-flop or counter. Also see RESET and RESET TERMINAL.

reset pulse A pulse that resets (see RESET, **1**) a storage cell in a computer memory.

reset terminal In a flip-flop, the zero-input terminal. Compare SET TERMINAL.

reset time The elapsed time between a malfunction and the completion of the reset action.

reset timer A device that returns a circuit or device to its initial state after a specified time delay.

reserve In multiple programming computer operations, to allocate memory areas and peripherals for a program.

reserve battery A battery in which the electrolyte is in a special standby chamber outside of the interelectrode section while the battery is on the shelf. When the battery is readied for service, the electrolyte is caused to flow into position between the electrodes, either by heating the battery, shocking it mechanically, or inverting it.

residual amplitude modulation See INCIDENTAL AM.

residual charge The electric charge remaining in a capacitor after it has been initially discharged. It results from dielectric absorption.

residual current A current that continues to flow in a circuit after removing power. The duration is measured in nanoseconds or microseconds.

residual frequency modulation **1.** See INCIDENTAL FM. **2.** Frequency modulation of the fundamental frequency of a Klystron by noise or alternating-current heater voltage.

residual gas Minute quantities of gas remaining in a vacuum tube after evacuation.

residual magnetism Magnetism remaining in a material, such as iron, after the magnetizing force has been removed.

residual modulation **1.** Modulation of a signal by hum or noise. **2.** See INCIDENTAL AM. **3.** See INCIDENTAL FM.

residual voltage In the output of a null device, such as a bridge, a usually small voltage still present at null and preventing zero balance.

residue check In computer operations, the verification of the result of an arithmetic operation using the remainders generated when each operand is divided by a special number; the remainder is transmitted along with the operand as a check digit.

resilience Also called *fault resilience.* The ability of an electronic device or system, especially a computer, to keep functioning after part of it has failed.

resin A natural or synthetic organic substance that is polymeric in structure and largely amorphous. Various plastics are made from synthetic resins.

resistance 1. Symbol, *R* or *r*. Unit, ohm. In a device, component, or circuit, the simple opposition to current flow. Resistance by itself causes no phase shift. In a purely resistive circuit, $R = E/I$, where R is the resistance in ohms, E is the voltage in volts, and I is the current in amperes. **2.** A property of circuits, devices, or substances that causes impinging energy to be dissipated by conversion to heat. Compare REACTANCE.

resistance

Per kilometer of solid copper wire for American Wire Gauge (AWG) 1 through 40.

AWG	*ohms/km*	*AWG*	*ohms/km*
1	0.42	21	43
2	0.52	22	54
3	0.66	23	68
4	0.83	24	86
5	1.0	25	110
6	1.3	26	140
7	1.7	27	170
8	2.1	28	220
9	2.7	29	270
10	3.3	30	350
11	4.2	31	440
12	5.3	32	550
13	6.7	33	690
14	8.4	34	870
15	11	35	1100
16	13	36	1400
17	17	37	1700
18	21	38	2200
19	27	39	2800
20	34	40	3500

resistance alloys Metallic alloys used in the manufacture of resistance wire and resistance elements. Such alloys include CONSTANTAN, GERMAN SILVER, MANGANIN, MONEL METAL, and NICHROME.

resistance balance A device used to balance a circuit, by means of the insertion of resistances.

resistance brazing A method of brazing in which metal is heated by passing a current through it. The I^2R loss, or dissipated power, occurs in the form of heat.

resistance bridge A bridge (see BRIDGE, **2**) for measuring resistance only.

resistance-capacitance Abbreviation, *RC*. Pertaining to a combination of resistance and capacitance (e.g., RESISTANCE-CAPACITANCE CIRCUIT).

resistance-capacitance bridge 1. A four-arm null circuit containing only resistors and capacitors. Also see BRIDGE, **1**. **2.** An alternating-current bridge (see BRIDGE, **2**) for measuring resistance and capacitance.

resistance-capacitance circuit A circuit containing only resistors and capacitors. There are no inductors.

resistance-capacitance-coupled amplifier A multistage amplifier circuit in which RESISTANCE-CAPACITANCE COUPLING is used between stages and at the input and output points of the circuit.

resistance-capacitance coupling Coupling, especially between stages in a circuit, using blocking capacitors and supply-path resistors.

resistance-capacitance filter A power-supply filter or wave filter containing only resistors and capacitors. The resistors are in the positions occupied by inductors in inductance-capacitance filters.

resistance-capacitance-inductance Abbreviation, *RCL*. Pertaining to a combination of resistance, capacitance, and inductance.

resistance-capacitance phase shifter A phase shifter containing only resistors and capacitors to obtain the desired shift.

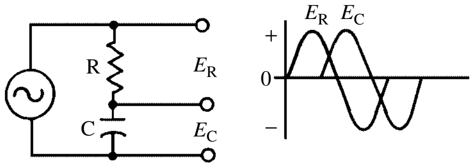

resistance-capacitance phase shifter

resistance-capacitance time constant Symbol, *t*. The time constant (see ELECTRICAL TIME CONSTANT) of a circuit containing (ideally) only resistance and capacitance; $t = RC$, where t is in seconds, R is in ohms, and C is in farads. Compare RESISTANCE-INDUCTANCE TIME CONSTANT.

resistance-capacitance tuning Tuning of a circuit, such as that of an amplifier or oscillator, by means of a variable resistor or ganged units of this type. See, for example, PARALLEL-TEE AMPLIFIER, PARALLEL-TEE OSCILLATOR, and WIEN-BRIDGE OSCILLATOR.

resistance-coupled amplifier See RESISTANCE-CAPACITANCE-COUPLED AMPLIFIER.

resistance drop The voltage drop across a resistor, or across the inherent resistance of a device.

resistance-inductance Abbreviation, *RL*. Pertaining to a combination of resistance and inductance (e.g., RESISTANCE-INDUCTANCE CIRCUIT).

resistance-inductance bridge 1. A four-arm null circuit containing only resistors and inductors. Also see BRIDGE, **1**. **2.** An alternating-current bridge (see BRIDGE, **2**) for measuring resistance and inductance only.

resistance-inductance circuit A circuit containing only resistors and inductors. There are no capacitors.

resistance-inductance phase shifter A phase shifter containing only resistors and inductors to obtain the desired phase shift.

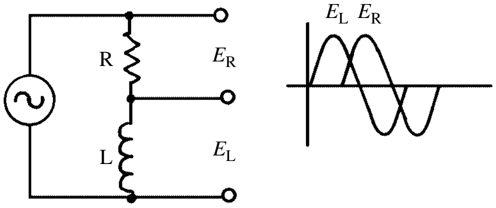

resistance-inductance phase shifter

resistance lamp An incandescent bulb inserted in series with a circuit to provide a dropping resistance. Such a lamp is capable of dissipating a large amount of power, shows very little reactance at low frequencies, and is inexpensive.

resistance magnetometer A magnetometer whose operation is based upon the change of electrical resistance of a material (such as bismuth wire) placed in the magnetic field under test.

resistance material A substance, such as carbon or German silver, whose resistivity is high enough to enable its use as a lumped resistor. See, for example, RESISTANCE ALLOYS and RESISTANCE METAL.

resistance metal A metal, such as iron, whose resistivity is high enough to enable its use as a lumped resistor. Also see RESISTANCE ALLOYS.

resistance pad An attenuator composed of noninductive resistors.

resistance standard A highly accurate and stable resistor used in precision measurements of resistance. Also see PRIMARY STANDARD and SECONDARY STANDARD.

resistance strain gauge An electrical strain gauge in which the stressed element is a thin resistance wire.

resistance strip A strip of metallic or nonmetallic resistance material. Also see RESISTANCE ALLOYS and RESISTANCE METAL.

resistance temperature detector A transducer consisting of a specially made resistor whose resistance varies linearly with temperature.

resistance thermometer An electronic thermometer whose operation is based on the change of resistance of a wire as it is heated or cooled.

resistance transducer See RESISTIVE TRANSDUCER.

resistance tuning See VARIABLE-RESISTANCE TUNING.

resistance welding An electrical or electronic welding process in which the workpieces are heated by current flowing through the inherent resistance of their junction.

resistance wire Wire made of a metal or alloy that exhibits significant resistivity. See, for example, RESISTANCE ALLOYS and RESISTANCE METAL.

resistance-wire sensor A specific length of resistance wire, properly mounted, whose resistance is proportional to a sensed phenomenon (such as strain, temperature, presence of gas, pressure, etc.). See, for example, ELECTRICAL STRAIN GAUGE, GAS DETECTOR, and PRESSURE TRANSDUCER.

resistive current The component of alternating current that is in phase with voltage. Also called WATT CURRENT. Compare REACTIVE CURRENT.

resistive cutoff frequency Symbol, f_{rco}. The frequency beyond which a tunnel diode ceases to exhibit negative resistance.

resistive load A load device that is essentially a pure resistance.

resistive losses Losses resulting from the resistance of a circuit or device, and usually appearing as heat.

resistive transducer A transducer in which the sensed phenomenon causes a change in resistance, which in turn produces a corresponding change in output current or voltage. Compare CAPACITIVE TRANSDUCER, CRYSTAL TRANSDUCER, INDUCTIVE TRANSDUCER, MAGNETIC TRANSDUCER, and PHOTOELECTRIC TRANSDUCER.

resistive trimmer See TRIMMER RESISTOR.

resistive voltage The voltage across the resistance component in a circuit. In an alternating-current circuit, the resistive voltage is in phase with the current.

resistivity Symbol, r. Resistance per unit volume or per unit area. It can be expressed in terms of ohms per cubic meter or ohms per square meter. Also see MICROHM-CENTIMETER and OHM-CENTIMETER.

resistor A device having resistance concentrated in lumped form. Also see RESISTANCE and RESISTIVITY.

resistor-capacitor-transistor logic Abbreviation, RCTL. A form of RESISTOR-TRANSISTOR LOGIC in which capacitors are used to enhance switching speed.

resistor color code See COLOR CODE.

resistor core A form around which a resistance wire can be wound for the purpose of constructing a high-power resistor.

resistor decade See DECADE RESISTOR.

resistor diode A usually forward-biased semiconductor diode that acts as a VOLTAGE-DEPENDENT RESISTOR.

resistor FET See ELECTRONIC RESISTOR.

resistor fuse See FUSIBLE RESISTOR.

resistors in parallel See PARALLEL RESISTORS.

resistors in parallel-series See PARALLEL-SERIES RESISTORS.

resistors in series See SERIES RESISTORS.

resistors in series-parallel See SERIES-PARALLEL RESISTORS.

resistor substitution box A self-contained assortment of common-value resistors arranged to be switched one at a time to a pair of terminals. In troubleshooting and circuit development, any of several useful fixed resistance values can thus be obtained.

resistor transistor See ELECTRONIC RESISTOR.

resistor-transistor logic Abbreviation, RTL. A circuit in which the logic function is performed by resistors, and an inverted output is provided by transistors.

resnatron A form of vacuum tube that is used as an oscillator and amplifier at ultra-high and microwave frequencies. It is essentially a cavity resonator.

resolution **1.** The degree to which closely adjacent parts of an image can be differentiated. **2.** The reduction of a problem by means of logical analysis. **3.** The ability of a vision or ranging system to distinguish between objects that are close together in terms of radial distance, direction, or absolute separation.

resolution ratio In a television image, the ratio of horizontal resolution to vertical resolution.

resonance **1.** The state in which the natural response frequency of a circuit coincides with the frequency of an applied signal, or vice versa, yielding intensified response. **2.** The state in which the natural vibration frequency of a body coincides with an applied vibration force, or vice versa, yielding reinforced vibration of the body.

resonance bridge An alternating-current bridge (see BRIDGE, **2**) in which one or two arms are series-resonant or parallel-resonant, the other arms being resistances. Also see SERIES-TYPE RESONANCE BRIDGE and SHUNT-TYPE RESONANCE BRIDGE.

resonance curve A graph of the insertion gain or loss of a component, device, circuit, or system to variations in the frequency of an applied sound or signal. Such curves are almost always plotted in rectangular coordinates with frequency as the independent variable on the horizontal axis. The dependent variable, plotted on the vertical axis, can be any characteristic that displays a peak or dip at the resonant frequency or frequencies. In radio-frequency circuits, such parameters include current, voltage, attenuation, gain, and impedance.

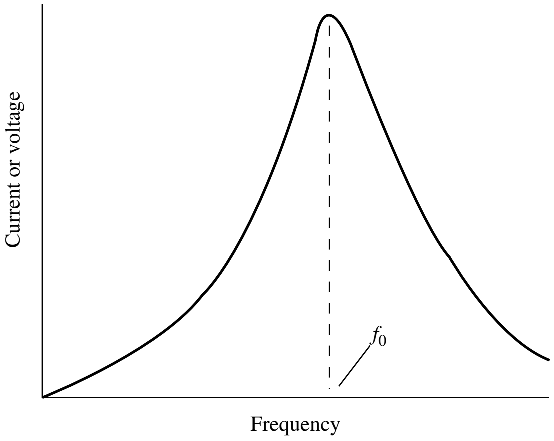

resonance curve

resonance theory of hearing The theory that sound waves pass down the *auditory canal* and cause the *eardrum* to vibrate. Behind the eardrum is a system of three bones leading to the *cochlea.* The cochlea consists of fibers that resonate. Therefore, they vary in length and tension. Various groups of fibers are activated by different sounds, and the vibrations are transmitted to nerves leading to the brain.

resonance radiation Electromagnetic radiation from an energized substance, resulting from movement of electrons from a higher to lower energy level. When an electron moves from a higher to a lower orbit, a photon, having a definite wavelength, is emitted.

resonant cavity A chamber whose size reinforces energy injected into it at a natural frequency, which is determined by the chamber's dimensions. Such cavities can be used with acoustic or electromagnetic waves.

resonant circuit A circuit whose constants are chosen for maximum circuit response at a given frequency. Examples: *parallel-resonant circuit* and *series-resonant circuit.* Also see RESONANCE and RESONANT FREQUENCY.

resonant current Current flowing in a tuned circuit at resonance.

resonant feeder An antenna feeder that is resonant at the operating frequency.

resonant filter A filter containing at least one series- or parallel-resonant arm for sharp response. Thus, a power-supply filter of this kind might have a parallel-resonant arm acting as a wavetrap at the ripple frequency.

resonant frequency Symbol, f_r or f_o. The natural frequency at which a circuit oscillates or a device vibrates. In an inductance-capacitance circuit (series-resonant or parallel-resonant), the reactances cancel at the resonant frequency, leaving only resistance.

resonant-gate transistor A transistor embodying a tiny tuning fork for resonance at low frequencies, thus eliminating bulky coils and capacitors.

resonant line A transmission line that is resonant at the operating frequency.

resonant-line amplifier See LINE-TYPE AMPLIFIER.

resonant-line circuit A circuit using resonant lines as a tank. See, for example, LINE-TYPE AMPLIFIER and LINE-TYPE OSCILLATOR.

resonant-line oscillator See LINE-TYPE OSCILLATOR.

resonant-line wavemeter See LECHER WIRES.

resonant rise See VOLTAGE RISE.

resonant-slope amplifier See DIELECTRIC AMPLIFIER.

resonant-slope detector See SLOPE DETECTOR.

resonant suckout The drawing of radio-frequency energy out of the energized part of a coil or transmission line by the part not intended to be energized, when the latter resonates at the same frequency.

resonant-voltage rise See VOLTAGE RISE.

resonant-voltage stepup See VOLTAGE RISE.

resonate **1.** To exhibit RESONANCE—either electrically or acoustically. **2.** To adjust a variable-frequency electrical or acoustical device so that it exhibits RESONANCE at a specific frequency.

resonator A device that produces or undergoes resonance. See, for example, HELMHOLTZ RESONATOR and RESONANT CAVITY.

resource A part of a computer system that can be used for a specific application as a unit (e.g., printer, PCMCIA standard adapter card, tape drive, etc.).

responder The transmitting section of a transponder.

response The behavior of a circuit or device (especially in terms of its dependent variables), in accordance with an applied signal (e.g., *frequency response* and *current-vs.-voltage response*).

response curve A graph depicting the performance of a circuit or device. Examples: *attenuation-vs.-frequency curve* and *current-vs.-voltage curve.*

response time The interval between the instant a signal is applied to or removed from a circuit or device and the instant that the circuit acts accordingly.

restart Following a malfunction or error occurring during a computer program run, to go back to an earlier point in the program.

resting state See QUIESCENT STATE.

restore See RESET.

resultant **1.** The vector that results from the addition of two or more vectors. **2.** A quantity that results from mathematical operations performed on other quantities.

retarding magnet See DRAG MAGNET.

retentivity **1.** The property whereby a material retains magnetism imparted to it. **2.** A quantitative measure of the extent to which a material retains magnetism imparted to it.

retention period In computer operations, the time during which the data on a magnetic medium (disk or tape) must be kept intact.

retrace **1.** In a cathode-ray tube, the return of the scanning beam to its starting point. **2.** In a cathode-ray tube, a line traced on the screen by the scanning beam as it returns to its starting point, if RETRACE BLANKING is not used.

retrace blanking Obliteration of the RETRACE of the electron beam in a cathode-ray tube. It renders the retrace line invisible on the screen so that it will not interfere with the display.

retrace line See RETRACE, **2**.

retrace ratio For the swept beam in a cathode-ray tube, the ratio of the scanning velocity in the trace direction to the scanning velocity in the RETRACE direction.

retrace time In a cathode-ray tube, the amount of time required for the scanning beam to move from the end of one trace or line to the beginning of the next.

retrofit To supply something with specially designed or adapted parts that were not available when it was made.

retrograde orbit For an artificial earth satellite, an orbit whose direction is east-to-west, relative to the earth's surface.

return **1.** See RETRACE. **2.** See RETURN CIRCUIT. **3.** See RETURN POINT. **4.** In an electronic circuit, the electrical ground and ground current path.

return circuit The circuit through which current returns to a generator.

return instruction In a computer program, an instruction in a subroutine directing operation back to a specific point in the main program.

return interval See RETRACE TIME.

return line See RETRACE, **2**.

return point **1.** The point to which circuits are returned (e.g., a common ground point). **2.** The terminal point of a return circuit.

return ratio See FEEDBACK FACTOR.

return time See RETRACE TIME.

return to zero **1.** Abbreviation RZ or RTZ. In the magnetic recording of data, a method in which the write current returns to zero following the write pulse. Compare NONRETURN-TO-ZERO. **2.** A logic system in which the zero and one states are represented by zero voltage and a discrete voltage.

return trace See RETRACE, **1**, **2**.

REV **1.** Abbreviation of REENTRY VEHICLE. **2.** Abbreviation of REVERSE.

rev **1.** Abbreviation of REVOLUTION. **2.** To quickly and substantially increase the angular velocity of a motor.

reverberation **1.** Multiple reflections of sound waves from the inside surfaces of an enclosed chamber. **2.** The dying-out of sound waves in an enclosed chamber as the waves reflect repeatedly from the inside surfaces. **3.** In sound recording and/or reproduction, an electronically produced echo. It is used for special effects—especially in electronic music systems. **4.** See RESONANCE, **2**.

reverberation chamber A room in which the walls, floor, and ceiling absorb very little sound, resulting in echoes. To avoid standing waves, the room is designed so that no two surfaces are exactly parallel.

reverberation system A system of devices operated with an electronic organ to simulate the effect of reverberation in a large room, such as a church auditorium.

reverberation time In an enclosed chamber, the time required for a sound to die down to a specified level (usually -60 dB) after the disturbance has stopped.

reverberation unit A device for producing artificial echoes—especially in the operation of electronic music systems.

reverse **1.** To alter the direction of a current or process or motion of an object so that the new direction is exactly opposite the previous direction. **2.** In a directional wattmeter, the reflected-power indication or switch position.

reverse AGC Automatic gain control in which a signal-dependent voltage is fed back to an earlier stage to adjust its gain automatically. Compare FORWARD AGC.

reverse bias Reverse voltage or current in a transistor or a semiconductor diode. Compare FORWARD BIAS.

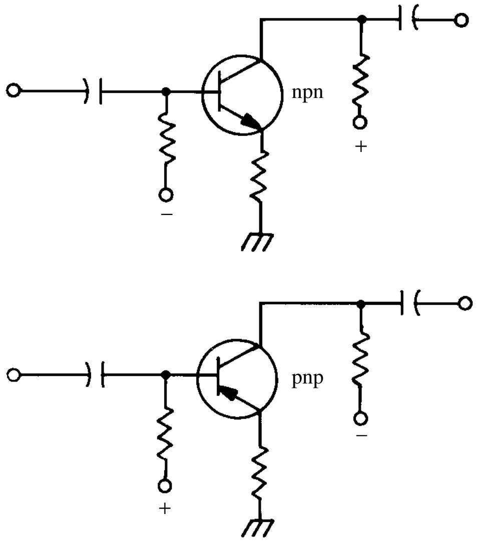

reverse bias

reverse breakdown See AVALANCHE.

reverse breakdown voltage See AVALANCHE VOLTAGE.

reverse characteristic The current-vs.-voltage response of a semiconductor junction that is biased in the reverse (low-conduction) direction. Compare FORWARD CHARACTERISTIC.

reverse conduction The very small current conduction through a pn junction when it is reverse-biased. Compare FORWARD CONDUCTION.

reverse current Symbol, I_r. The current that flows through a pn junction when it is reverse-biased. Also called *back current.* Compare FORWARD CURRENT.

reverse engineering A design process in which a specific device or system is copied functionally, but not literally.

reverse Polish notation Abbreviation, RPN. A system of notation for expressing mathematical operations in which the operators follow the operands being manipulated. It is a mode of entry for some calculators (e.g., the operation 7 × 2 might be entered by pressing keys in this order: 7, ENTER, 2, ×).

reverse recovery time See RECOVERY TIME, **1**.

reverse resistance Symbol, R_r. The resistance of a reverse-biased pn junction. Also called BACK RESISTANCE. Compare FORWARD RESISTANCE.

reverse voltage Symbol, E_r or V_r. Direct-current voltage applied to a pn junction so that the p-type material is electrically more negative than the n-type material. Also called BACK VOLTAGE. Compare VOLTAGE.

reverse-voltage capacitance The internal capacitance of a reverse-biased semiconductor pn junction.

reverse voltage drop The voltage drop across a semiconductor pn junction that is biased in the reverse (low-conduction) direction.

reversible counter A counter that, by a control signal, can have the value it is holding either increased or decreased.

reversible permeability The permeability of a ferromagnetic substance when the magnitude of the alternating-current field is arbitrarily small.

reversing switch **1.** A switch that reverses the polarity of a direct-current voltage. **2.** A switch that reverses the direction of motor rotation.

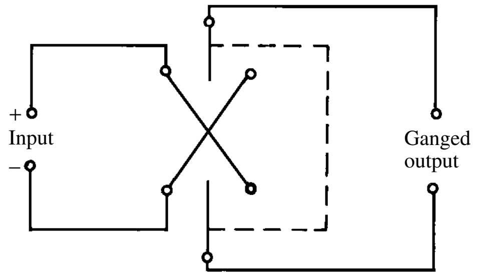

reversing switch

revolute geometry A method by which a robot arm can move freely in three dimensions. The entire assembly rotates from the base in a horizontal plane through a complete circle (360 degrees). An elevation joint at the base moves the arm from horizontal to vertical (90 degrees). A joint in the middle of the arm can bend through about 180 degrees.

revolution Abbreviation, r or rev. One complete rotation (i.e., 360 degrees of circular travel).

revolving field See ROTATING FIELD.

rewind To run a magnetic tape on a transport at a high speed, in the direction opposite to that associated with the play mode.

rewrite In computer operations, to return information read from a storage location to that location by recording.

R_F Symbol for FILAMENT RESISTANCE.

RF Abbreviation of RADIO FREQUENCY.

RF amplifier See RADIO-FREQUENCY AMPLIFIER.

RFC Abbreviation of RADIO-FREQUENCY CHOKE.

RF heating See RADIO-FREQUENCY HEATING.

RFI Abbreviation of RADIO-FREQUENCY INTERFERENCE.

RF inverse feedback A negative-feedback system for radiophone transmitters, in which a portion of the modulated radio-frequency (RF) signal is rectified, and the resulting direct-current voltage is filtered and applied as bias to one of the audio stages in the proper polarity for degeneration.

RF lamp A lighting lamp, used with radio-frequency (RF) alternating current, rather than the conventional 60-Hz utility current. This results in better efficiency, and allows much more light to be generated with a given filament lamp, as compared with 60-Hz current.

RF motion detector In security systems, an intrusion detection and alarm system that senses Doppler-effect-induced changes in the frequency or phase of a radio-frequency (RF) electromagnetic field. The Doppler effect results from motion of objects in the secured area.

RFO Abbreviation of *radio-frequency oscillator.*

RF power supply See OSCILLATOR-TYPE POWER SUPPLY.

RF preamplifier A sensitive, radio-frequency amplifier circuit intended for improving the signal-to-noise (S/N) ratio in a wireless receiver. Generally placed between the receiver and the antenna or feed line. Some such devices are tunable; others are broadbanded. See also PREAMPLIFIER.

RF probe See RECTIFIER PROBE.

RF resistance See RADIO-FREQUENCY RESISTANCE.

RF selectivity See RADIO-FREQUENCY SELECTIVITY.

RF transformer See RADIO-FREQUENCY TRANSFORMER.

RF transistor See RADIO-FREQUENCY TRANSISTOR.

R_g Symbol for GRID RESISTANCE.

R_G Symbol for GATE RESISTANCE.

RGB Abbreviation of RED-GREEN-BLUE.

RGT Abbreviation of RESONANT-GATE TRANSISTOR.

Rh Symbol for RHODIUM.

R/h Abbreviation of ROENTGENS PER HOUR.

R_H **1.** Symbol for HEATER RESISTANCE. **2.** Symbol for HOT RESISTANCE.

rh Abbreviation of RELATIVE HUMIDITY.

rhenium Symbol, Re. A metallic element. Atomic number, 75. Atomic weight, 186.207. It is used in some thermocouples.

rheostat A wirewound variable dropping resistor of the rotary type or slider type.

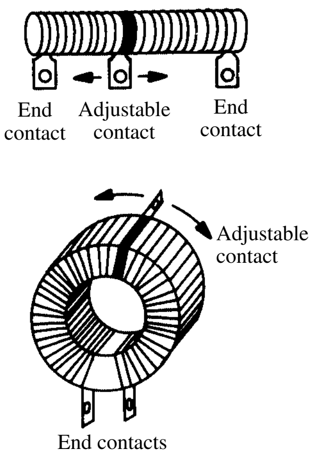

rheostat

R_{HF} Symbol for *high-frequency resistance* (see RADIO-FREQUENCY RESISTANCE).

RHI Abbreviation of RANGE-HEIGHT INDICATOR.

rhodium Symbol, Rh. A metallic element. Atomic number, 45. Atomic weight, 102.906.

rhombus A four-sided geometric plane figure, in which all four sides have equal length, and opposite angles have equal measure.

rhombic antenna See DIAMOND ANTENNA.

rho-theta A radio-navigation system in which a single transmitting station is used, and the position is determined, according to polar coordinates (distance and direction).

rhumbatron A RESONANT CAVITY—especially one in a KLYSTRON.

RI Abbreviation of RADIO INTERFERENCE.

R_i Symbol for INPUT RESISTANCE. (Also, R_{in}.)

RIAA Abbreviation of RECORDING INDUSTRY ASSOCIATION OF AMERICA.

RIAA curve The amplitude-versus-frequency function used in recording and reproduction of long-playing (33.3 rpm) phonograph discs, and specified by the Recording Industry Association of America (RIAA). The RIAA curve takes advantage of the sensitivity of the human ear at various frequencies to reduce the level of audible noise.

ribbon microphone See VELOCITY MICROPHONE.

ride gain In broadcasting, the operations of constantly adjusting the audio modulation of the transmitter for optimum operation.

Rieke chart A visual aid, similar to the SMITH CHART, used with traveling-wave tubes in the ultra-high-frequency (UHF) and microwave frequency bands for determining the optimum load impedance.

rig Colloquialism for a radio communications installation—especially a transmitter or transceiver. It is commonly used among amateur radio operators.

right-angle line section See ELL.

right-hand lay See DIRECTION OF LAY.

right-hand polarized wave See CLOCKWISE-POLARIZED WAVE.

right-hand rule for induced emf See FLEMING'S RIGHT-HAND RULE.

right-hand rule for wire A simple rule for indicating the direction of the magnetic field surrounding a straight wire that carries current. When the wire is grasped in the right hand with the thumb pointing in the direction of current flow, the fingers curl in the direction of the magnetic field.

right-hand taper Potentiometer or rheostat taper in which most of the resistance is in the clockwise half of rotation, as viewed from the front. Compare LEFT-HAND TAPER.

right justified In a computer memory location, a data item that takes up consecutive bit positions, from right to left.

Right-Leduc effect A phenomenon somewhat analogous to the Hall effect. When a metal strip conducting heat is placed in a magnetic field perpendicular to the plane of the strip, a temperature difference develops across the strip.

right shift In computer operation, a shift whereby word bits are displayed to the right; the effect is division in a right arithmetic shift.

rim drive **1.** In a tape recorder, a driving method in which the motor shaft is provided with a smooth pulley that transfers motion directly to the rubber-tired rim of the flywheel. **2.** A driving method for a phonograph turntable in which a rotating wheel contacts the outer edge of the platter.

R_{in} Symbol for INPUT RESISTANCE. (Also, R_i.)

ring **1.** The core of a toroidal coil. **2.** See HYBRID RING. **3.** See RING MODULATOR. **4.** See RING INDUCTOR. **5.** See RING MAGNET. **6.** See RINGING.

ring armature A motor or generator armature having a ring winding.

ringback In a telephone system, the signal sent from the receiving (destination) telephone set back to the sending (source) set, indicating that the signal is being received. This consists of a tone, interrupted by pulses at intervals of about 0.05 second. The signal stops when the destination set is taken off the hook. Also see RINGDOWN.

ring circuit **1.** See RING MODULATOR. **2.** A waveguide hybrid-tee resembling a ring having radial branches. **3.** In amateur radio operations, a circuit connected to a telephone line and radio transmitter. The radio transmitter is energized and modulated with an identifiable signal each time the telephone rings, so the telephone can be answered via remote control.

ring counter An electronic counter in which successive cascaded stages form a ring (i.e., the last stage in the chain is connected to the first stage so that the counter advances through the cycle, stage by stage, repetitively).

ringdown In a telephone system, the signal sent from the transmitting (source) set to the receiving (destination) set, causing the destination set to ring. Also see RINGBACK.

ring head In tape recorders, a recording and playback head that consists essentially of a metal ring with a gap at one point, and on which the coils are wound.

ring inductor **1.** An inductor consisting of a single turn of wire, or of a conductor bent into a loop. **2.** See SHADING COIL.

ringing Self oscillation in a pulsed inductance-capacitance circuit, sustained by the circuit's flywheel action (hysteresis), and usually producing a damped wave.

ringing coil In the horizontal oscillator in a television circuit, a small, adjustable coil (shunted by a capacitor) used to produce a sharp rise in input-signal voltage.

ringing current In wire telephony, an alternating current superimposed on the direct operating current. Produces RINGDOWN.

ringing time See RING TIME, **1**.

ring magnet A permanent magnet in the shape of a ring or donut.

ring main An electric power main that is closed to form a ring. This results in two independent electrical paths between any two points in the circuit. If one path is interrupted, power can still be transmitted to any other point in the circuit from a power station.

ring modulator A double-balanced diode-type modulator circuit; its name is derived from the ring-like arrangement of the four diodes.

ring oscillator A self-excited oscillator in which two sets of two transistors are operated in push-pull/parallel.

ring shift In computer operation, the cyclic shifting of digits from one end of a register to the other.

ring time **1.** The period of a damped oscillation—especially one set up in an inductance-capacitance circuit by a pulse. **2.** The time required for an ECHO BOX signal to decay below the display level.

ring winding A winding in which the turns of the coil are laid on the outside of a ring-shaped core and passed through its center, resulting in a donut coil with a core.

ripple **1.** A small alternating-current component in the output of a direct-current power supply with inadequate filtering. **2.** In computer and data-processing operations, the serial transmission of data.

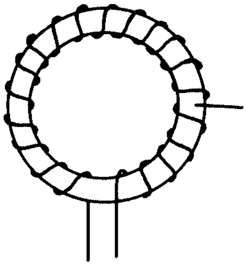

ring winding

ripple amplitude The peak or peak-to-peak value of ripple in a power supply (see RIPPLE, **1**).

ripple counter A binary counter consisting of flip-flops cascaded in series. A pulse must pass sequentially from the input, through each stage, to the output of the chain.

ripple current Current flowing in a circuit as the result of ripple voltage (see RIPPLE AMPLITUDE).

ripple factor The ratio of the RIPPLE AMPLITUDE to the direct-current voltage output of a power supply.

ripple frequency The frequency of a ripple component (see RIPPLE, **1**). In power supplies using half-wave rectification, this frequency is normally 60 Hz (the line frequency); in full-wave supplies, it is normally 120 Hz (twice the line frequency).

ripple percentage See PERCENT RIPPLE.

ripple torque Symbol, T_R. In a torque motor, the small fluctuation in torque resulting from commutator switching action.

ripple voltage See RIPPLE AMPLITUDE.

RISC Abbreviation of REDUCED INSTRUCTION SET COMPUTER.

rise **1.** See VOLTAGE RISE. **2.** See RISE TIME. **3.** An increase in the amplitude of a pulse or wave.

rise cable **1.** A vertical feeder cable. **2.** A vertical section of a feeder cable.

rise time The time required for a pulse to rise from 10 percent to 90 percent of its peak amplitude. Compare FALL TIME.

RJE Abbreviation of REMOTE JOB ENTRY.

R_k Symbol for cathode resistance.

RL **1.** Abbreviation of RESISTANCE-INDUCTANCE. **2.** Abbreviation of RELAY LOGIC.

R_L Abbreviation of LOAD RESISTANCE.

RL bridge See RESISTANCE-INDUCTANCE BRIDGE.

RL circuit See RESISTANCE-INDUCTANCE CIRCUIT.

R_{LF} Symbol for low-frequency resistance.

RL phase shifter See RESISTANCE-INDUCTANCE PHASE SHIFTER.

R_m Symbol for METER RESISTANCE.

r_m Symbol for *emitter-collector transresistance* of a bipolar transistor.

rms Abbreviation of ROOT MEAN SQUARE.

rms converter See ROOT-MEAN-SQUARE CONVERTER.

rms current See EFFECTIVE CURRENT.

rms meter A current meter or voltmeter whose deflection is proportional to the root-mean-square (rms) value of current or voltage. In most meters, the deflection is proportional to either the peak value or the average value, but the scale of an rms unit is graduated on the basis of sine-wave input.

rms value See EFFECTIVE VALUE.

rms voltage See EFFECTIVE VOLTAGE.

Rn Symbol for RADON.

R_n **1.** Symbol for NEGATIVE RESISTANCE. (Also, –R.) **2.** Symbol for *null resistance.*

R_o Symbol for OUTPUT RESISTANCE. (Also, R_{out}.)

roaming **1.** In cellular communications, an ideal condition in which the user of a mobile or portable wireless telephone set can travel throughout a specified geographic region with no "dead zones" or interruption in service. **2.** For a mobile or portable cellular telephone subscriber, the act of traveling throughout a large geographic region and, in particular, among different states or countries.

robot **1.** An electromechanical device or system capable of reliably performing complex and/or repetitive tasks. It can be controlled by a human operator or by a computer. **2.** A usually autonomous device, as defined in **1**, built to physically resemble a human being, with a head, two arms, and some form of locomotion.

robot generations Agreed-on milestones or criteria in the evolution of robots and smart machines. *First generation:* Before 1980. Mainly mechanical, stationary, physically rugged, no external sensors, no artificial intelligence. *Second generation:* 1980–1990. Tactile sensors, vision systems, position sensors, pressure sensors, microcomputer control, programmable. *Third generation:* After 1990. Mobile, autonomy or group control, artificial intelligence, speech recognition/synthesis, teleoperation, navigation systems. *Fourth generation:* In conceptual stages. Highly intelligent, capable of building other robots, capable of doing many human tasks.

robot gripper A robotic END EFFECTOR designed specifically to grasp objects. The two basic designs are: hand-like and specialized. Hand-like grippers are engineered according to the notion that the human hand has evolved to near perfection, and should be mimicked in robots. Specialized grippers are built by trial-and-error methods and often bear little resemblance to human hands.

robotics The science and art of robot design, construction, operation, and maintenance.

Rochelle salt Sodium potassium tartrate, a substance whose crystals are piezoelectric. Such crystals are used in some microphones, loudspeakers, and transducers. Also called SEIGNETE SALT.

rock Slang for QUARTZ CRYSTAL.

rockbound Pertaining to an oscillator or radio transmitter whose frequency is determined by a quartz crystal, and is, therefore, not variable over a significant range.

rocker switch A toggle switch whose lever is a specially shaped bar. The bar is rocked back and forth to operate the switch. Compare BAT-HANDLE SWITCH, PADDLE SWITCH, and SLIDE SWITCH.

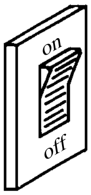

rocker switch

rod **1.** A unit of length or distance; 1 rod = 5.029 meters. **2.** A bar of material with special properties.

rod antenna See FERRITE-ROD ANTENNA.

rod magnet A permanent magnet in the shape of a rod with circular or elliptical cross section.

roentgen Abbreviation, r. A unit of ionizing radiation; 1 r is the quantity of radiation that produces 1 esu of electricity (positive or negative) per cubic centimeter of air at standard temperature and pressure. In average tissue, 1 r produces ionization equivalent to an energy concentration of 2.58×10^{-4} C/kg (93 ergs per gram). Also see MILLIROENTGEN.

roentgen equivalent man See REM.

roentgen equivalent physical See REP.

roentgen ray See X RAY.

roger A communications signal meaning "Acknowledged."

Roget spiral A spring-like wire device that contracts in proportion to the magnitude of the current flowing through it.

role indicator In computer operations, a code classifying a keyword as a part of the speech (e.g., noun).

roll **1.** In a display terminal having a line length of less than the standard 80 characters, an operating feature that allows the operator to follow the text along. The cursor remains fixed near the center of the displayed line, while the text moves from right to left. **2.** Vertical movement of a television picture, resulting from lack of vertical synchronization.

rollback In computer operations, the running again of a computer program or portion of the program. Also called RERUN.

roller inductor A variable inductor, usually of the air-core type, with a shaft attached that allows the whole assembly to be rotated; a wheel contact provides a movable tap. Such an indicator is continuously variable and is often used in such devices as antenna-tuning networks.

rolling In television, the apparent continuous upward or downward movement of the picture, resulting from lack of vertical synchronization between the transmitter and receiver.

rolloff **1.** The rate at which a dependent variable (e.g., output amplitude) diminishes above or below a certain critical value of the independent variable (e.g., frequency). It pertains especially to frequency response in audio devices and systems. **2.** Attenuation of the bass and/or treble response or output in an audio system.

ROM Abbreviation of READ-ONLY MEMORY.

Romeo Phonetic alphabet code word for the letter R.

Romex cable A form of wire cable with a covering that is highly resistant to the environment.

roof mount A metal bracket for fastening an antenna mast to a roof.

room noise Ambient acoustic noise in a room.

room resonance Acoustic resonance caused by the geometry and contents of a room.

room temperature Abbreviation, RT. **1.** The temperature of the chamber in which a test or fabrication is carried out. It is commonly used to distinguish between operations that can be performed at the ambient temperature and those that require an oven or a cold chamber. **2.** A temperature typical of an indoor environment, approximately 21 degrees Celsius (70 degrees Fahrenheit).

room tone A qualitative expression for the suitability or behavior of an enclosed area for a given acoustic application. It affected by factors (such as resonances or lack thereof, echoes or lack thereof, overall room size and proportions, and background noise).

root mean square Abbreviation, rms. The square root of the arithmetic mean (average) of the squares of a set of values.

root-mean-square converter A device that converts an input signal of any waveform into a direct-current signal of the same value as the EFFECTIVE VALUE of the input signal.

root-mean-square current See EFFECTIVE CURRENT.

root-mean-square value See EFFECTIVE VALUE.

root-mean-square voltage See EFFECTIVE VOLTAGE.

ROP Abbreviation of *record of production.*

rosin A substance derived chemically from an extract of pine wood, and used in some solders.

rotameter A fluid flow gauge consisting of a float within a glass tube having incremental markings.

rotary amplifier See AMPLIDYNE.

rotary antenna See ROTATABLE ANTENNA.

rotary beam A beam antenna, such as a Yagi, that can be rotated in a (usually horizontal) plane to allow transmission and/or reception in various directions. Also see ROTATABLE ANTENNA.

rotary-beam antenna See ROTARY BEAM.

rotary converter A dynamo (electric machine) having a direct-current armature connected to a commutator on one end of the shaft and to slip rings on the other end. When the machine is operated as a direct-current motor, it delivers an alternating-current output, and vice versa. Also called DOUBLE-CURRENT GENERATOR.

rotary dialing An older style of telephone dialing in which a rotary pulse generator is used to dial the digits.

rotary digital audio tape Digital audio tape used with a recording/playback system that uses a rotating head or heads.

rotary encoder An optical analog-to-digital (A/D) converter that operates by passing a light beam through a rotating disk. The amplitude of the analog input signal at any moment causes a certain angular rotation of the disk, cutting off the light beam, according to the nature of the pattern on the disk. The resulting modulated light beam has digital characteristics and can be detected using photocells.

rotary inverter A motor-generator used to change a direct-current input voltage into an alternating-current output voltage.

rotary-motion sensor A transducer that delivers an output voltage proportional to the arc over which its shaft has been turned.

rotary power amplifier See DC GENERATOR AMPLIFIER.

rotary relay An electromechanical relay in which a pivoted armature rotates to open or close the contacts. The *meter relay* is an example.

rotary selector switch See ROTARY SWITCH.

rotary stepping relay See STEPPING SWITCH.

rotary stepping switch See STEPPING SWITCH.

rotary switch A switch in which a blade moves in a circle or in arcs over the contacts.

rotary transformer A motor-generator used to change an input voltage into a lower or higher output voltage.

rotatable antenna An antenna that can be turned to change the orientation of its main lobe (direction of greatest forward gain) in a specific plane (usually horizontal).

rotating amplifier See DC GENERATOR AMPLIFIER.

rotating antenna An antenna that constantly turns to scan a given area (e.g., a RADAR ANTENNA).

rotating-capacitor modulator A frequency modulator consisting of a motor-driven variable capacitor. Also see WOBBULATOR.

rotating field An alternating-current electric or magnetic field, such as that generated by the stators of some motors, that revolves between poles.

rotating interrupter A commutator (see COMMUTATOR, **1**).

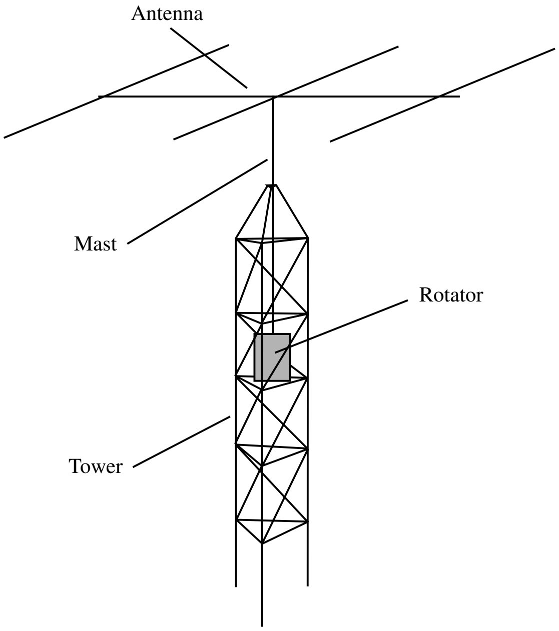

rotatable antenna

rotating machines Electromechanical devices (such as motors, generators, amplidynes, rotary converters, etc.) that utilize magnetic flux to convert angular motion or displacement into electrical energy, or vice-versa.

rotating memory See DISK MEMORY and DRUM MEMORY.

rotating voltmeter See GENERATING VOLTMETER.

rotator A motor-driven, remotely controlled mechanism for turning a directional antenna in a specific plane.

rotor **1.** A rotatable coil—especially in a motor or generator. Compare STATOR, **1**. **2.** The rotating member of a motor or generator. Compare STATOR, **2**. **3.** The rotating-plate assembly of a variable capacitor. Compare STATOR, **3**.

rotor blade The wiper arm of a rheostat or potentiometer.

rotor coil See ROTOR, **1**.

rotor plate The rotating plate(s) of a variable capacitor. Compare STATOR PLATE.

roulette pattern A circular pattern for frequency identification with an oscilloscope, consisting of loops around the screen's circumference. Compare GEAR-WHEEL PATTERN, LISSAJOUS FIGURE, and SPOT-WHEEL PATTERN.

rounding **1.** A method of approximating a quantity by reducing the number of significant digits. For example, rounding 3.44 to two significant digits yields 3.4; rounding 3.46 to two significant digits yields 3.5. Compare TRUNCATION. **2.** The approximation of a value to a specified number of decimal places or significant digits. **3.** Smoothing of the corners of a square wave or sawtooth wave, resulting in lengthening of the transition time from one state to another.

rounding error The error resulting from the rounding of a number (see ROUNDING, **1**, **2**).

round off To shorten an otherwise lengthy number by replacing numerical digits with zeros and increasing the final nonreplaced digit by 1 if the leftmost replaced digit is 5 or greater. Thus, 3.141592653 can be rounded off to 3.1416 or 3.14.

round-up A form of numerical approximation, in which a number with a value of n.5 or greater is assigned the value $n + 1$. This is a feature of many calculators using scientific notation or a fixed number of decimal places.

R_{out} Symbol for OUTPUT RESISTANCE. (Also, R_o.)

route **1.** To physically position wires or conducting circuit paths by planning and deliberation. **2.** The path over which conductors are positioned. **3.** A path over which signals or information can be carried.

routine **1.** In computer operations, the complete sequence of instructions for performing an operation (i.e., a program or program segment). **2.** A test or measurement procedure. **3.** An assembly or manufacturing procedure. **4.** A standard troubleshooting procedure.

row In a matrix, a horizontal arrangement or set of values.

R_P **1.** Symbol for PLATE RESISTANCE. (Also, r_P.) **2.** Symbol for POSITIVE RESISTANCE. **3.** Symbol for PARALLEL RESISTANCE. (Also, R_{par}.) **4.** Symbol for PRIMARY RESISTANCE. (Also, R_{pri}.)

R_{par} Symbol for PARALLEL RESISTANCE. (Also, R_P.)

r parameters **1.** Device or network parameters expressed as resistances. **2.** Transistor parameters in terms of resistance values in the equivalent tee network. Compare G PARAMETERS and H PARAMETERS.

RPG Abbreviation of REPORT PROGRAM GENERATOR.

rpm Abbreviation of *revolutions per minute.*

rpm meter See ELECTRONIC TACHOMETER and STROBOSCOPE.

RPN Abbreviation of REVERSE POLISH NOTATION.

R_{pri} Symbol for PRIMARY RESISTANCE. (Also, R_P.)

rps Abbreviation of *revolutions per second.*

RPT Radiotelegraphic abbreviation of *repeat.*

R_{req} **1.** Symbol for REQUIRED RESISTANCE. **2.** Symbol for SCREEN RESISTANCE.

R_{RF} Symbol for RADIO-FREQUENCY RESISTANCE.

R_S Symbol for SOURCE RESISTANCE in a field-effect transistor.

R_s **1.** Symbol for SERIES RESISTANCE. (Also, R_{ser}.) **2.** Symbol for SECONDARY RESISTANCE. (Also, R_{sec}.)

R_{sec} Symbol for SECONDARY RESISTANCE. (Also, R_s.)

R_{ser} Symbol for SERIES RESISTANCE. (Also, R_s.)

RST flip-flop A conventional flip-flop subject to the operations of reset, set, and trigger.

RST system In the amateur radio service, a method of reporting signal quality in terms of *readability*, *strength*, and *tone.*

R sweep In an oscilloscope or spectrum analyzer, an expanded portion of the trace produced by a long triggered sweep. It permits detailed analysis of a small portion of a displayed waveform or frequency band.

RT **1.** Abbreviation of RADIOTELEPHONE. **2.** Abbreviation of ROOM TEMPERATURE.

rt Abbreviation of RIGHT.

R_T **1.** Symbol for THERMAL RESISTANCE. **2.** Symbol for *total resistance.* (Also, R_t.)

R_t Symbol for *total resistance.*

RTD Abbreviation of RESISTANCE TEMPERATURE DETECTOR.

RTL Abbreviation of RESISTOR-TRANSISTOR LOGIC.

RTTY Abbreviation of RADIOTELETYPE.

RTZ Abbreviation of RETURN TO ZERO. (Also, RZ.)

Ru Symbol for RUTHENIUM.

rubber A natural insulating material; an elastomer exhibiting rapid elastic recovery. Dielectric constant, 2 to 3.5. Dielectric strength, 16 to 50 kV/mm. Also called *India rubber.* Compare HARD RUBBER.

rubber-covered wire Wire insulated with a jacket of rubber.

rubidium Symbol, Rb. A metallic element. Atomic number, 37. Atomic weight, 85.468.

ruby laser A laser using a ruby rod as the resonant element.

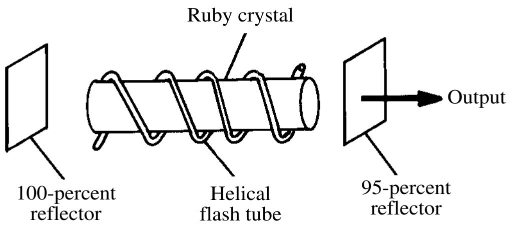

ruby laser

ruby maser A maser in which the resonant material is ruby.

Ruhmkorff coil See INDUCTION COIL.

rumble **1.** Low-frequency acoustical noise of a frequency below about 50 Hz. **2.** Low-frequency electrical noise in an audio circuit of a frequency below about 50 Hz. **3.** Vibrations that can occur in a poorly designed or malfunctioning phonograph turntable.

rumble filter An audio high-pass filter having sharp cutoff below 50 Hz, for eliminating rumble arising from irregularities in the rotation of a phonograph turntable. Also see RUMBLE, **3**.

run **1.** The execution of a computer routine or program. **2.** To execute a routine or program. **3.** To cause a routine or program to be executed. **4.** A command that causes a routine or program to be executed.

runaway In a current-carrying circuit or device, especially a semiconductor, a rapid increase in current that causes the temperature to rise, in turn resulting in a further increase in current. Unless preventive measures are taken, this will ultimately damage or destroy the component.

run chart In computer operations, a flowchart showing the organization and order of pertinent programs to be run.

running accumulator A computer storage unit having registers linked so that data is transferred unidirectionally from one to the other, and in which only one register is accessible from the outside.

running open **1.** The condition of a mechanical teleprinter running continuously in the absence of a signal. The teleprinter operates, but nothing is printed; this keeps the machine in synchronization. **2.** Operation of a transmitter at the maximum rated level of input or output power.

running-time meter See ELAPSED-TIME METER.

run time **1.** The period of time during which a computer program is executed. **2.** The length of time required for a computer program to be executed.

rupture **1.** The usually rapid and violent tearing apart, or breaking through, of an insulating material subjected to excessive voltage. **2.** The clean opening of relay, circuit-breaker, or switch contacts to interrupt a current-carrying circuit.

rush Broadband audio background noise, such as that arising from superheterodyne receivers and high-gain amplifiers. Its name is derived from resemblance to the gentle rushing of wind. Compare HISS, **1**, **2**.

ruthenium Symbol, Ru. A rare metallic element. Atomic number, 44. Atomic weight, 101.07.

rutherford Abbreviation, rd. A unit of radioactivity equal to 10^6 disintegrations per second (2.7×10^{-5} curie). Also see KILORUTHERFORD, MEGARUTHERFORD, MICRORUTHERFORD, and MILLIRUTHERFORD.

Rutherford atom An early concept of the nature of the atom, proposed by Rutherford in 1912. In this model, negatively charged electrons orbit a central, positively charged nucleus in a manner similar to the way planets orbit the sun. Compare BOHR ATOM.

rutherfordium Symbol, Rf. Also called *unnilquadium* (Unq) and *kurchatovium* (Ku). Atomic number, 104. The most common isotope has atomic weight 261. Classified as a transition metal. It has a half-life on the order of a few seconds to a few tenths of a second (depending on the isotope), is human-made, and is not known to occur in nature.

RW Abbreviation of *radiological warfare.*

R_x Symbol for unknown resistance.

RY Abbreviation of RELAY.

ryotron A form of inductive semiconductor switch, operated at cold temperatures to maximize conductivity.

R-Y signal In a color-television circuit, the signal representing primary red (R) minus luminance (Y). A primary red signal is obtained when the R-Y signal is combined with the luminance (Y) signal. Compare B-Y SIGNAL and G-Y SIGNAL.

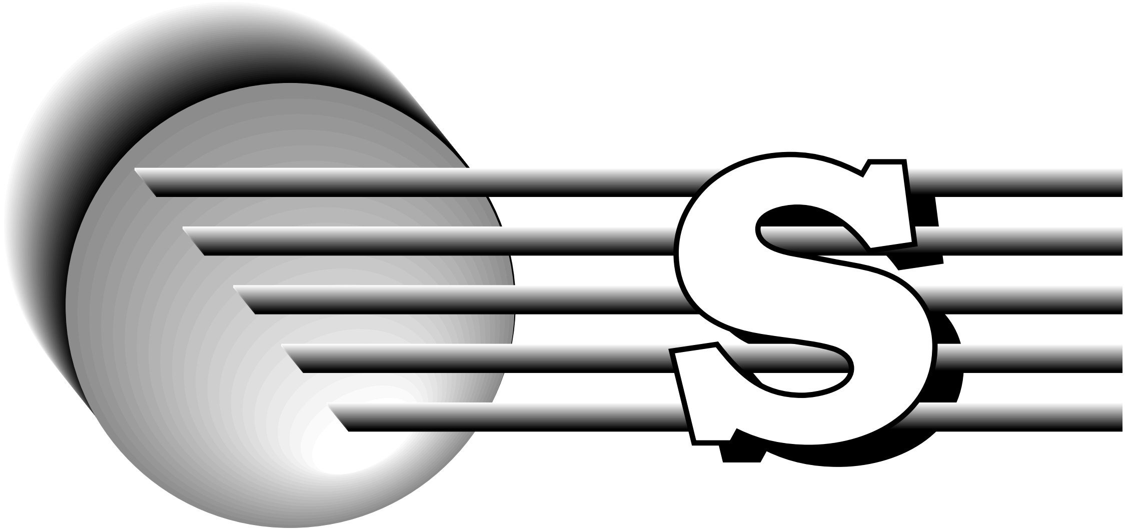

S **1.** Symbol for SCREEN GRID of a vacuum tube. **2.** Symbol for SHELL of a tube or semiconductor device. **3.** Symbol for SULFUR. **4.** Symbol for DEFLECTION SENSITIVITY. **5.** Symbol for SWITCH. **6.** Symbol for ELASTANCE. **7.** Abbreviation of SYNC. **8.** Symbol for SECONDARY. **9.** Abbreviation for SIEMENS. **10.** Abbreviation of SINE. **11.** Symbol for ENTROPY.

s **1.** Symbol for distance or DISPLACEMENT. **2.** Symbol for SCREEN GRID of a vacuum tube. **3.** Symbol for STANDARD DEVIATION. **4.** Abbreviation for SECOND.

SA Abbreviation of SUBJECT TO APPROVAL.

S_A band A section of the S band, extending from 3100 to 3400 MHz.

sabin A unit of sound absorption; 1 sabin represents a surface that can absorb sound at the same rate as 1 square foot of a perfectly absorbent surface.

SADT Abbreviation of *surface alloy diffused-base transistor.*

SAE **1.** Abbreviation of SHAFT-ANGLE ENCODER. **2.** Abbreviation of *Society of Automotive Engineers.*

safe noise level A level of acoustic intensity equal to 85 dB above the threshold of hearing. At sound levels higher than this, eardrum damage is possible.

safety factor A figure denoting the extent of overload that a device can withstand before breaking down.

safety ground A connection made between equipment (usually the metal chassis, panel, case, or B-minus circuit) and the earth as a protective measure against fire and electric shock.

safety switch See ELECTRICAL INTERLOCK.

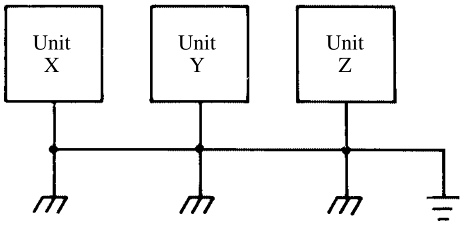

safety ground

sal ammoniac Formula, NH_4Cl. Ammonium chloride, the principal ingredient in the gelatinous electrolyte of some dry cells and batteries.

salient pole A pole, such as the polepiece of a motor or generator, that projects from the rest of the structure (rotor assembly or motor frame).

Salisbury chamber A radio-frequency test chamber in which the walls are non-reflective at various frequencies, thus simulating free space.

salt-spray test A test to assess the life and performance of electronic equipment in a marine environment. The equipment is sprayed, usually with a saltwater mist, and various electrical parameters are measured at prescribed time intervals.

samarium Symbol, Sm. A metallic element of the rare-earth group. Atomic number, 62. Atomic weight, 150.36.

sample **1.** A selection of quantities, events, or objects, taken at a specific time or time interval for analysis or testing. **2.** To take a sample, as defined in **1**. **3.** To test a quantity (current, voltage, temperature, pressure, etc.) or a material

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

(electrolyte, insulant, corrosion, rust, etc.) taken from a larger group or body.

sample and hold A method of storing a variable signal for detailed examination.

sampled data system A system that can be either analog or digital, and that works from samples of the input signals.

sampler In audio systems, a device that digitizes and stores sound for a brief interval of time.

sample size In statistics, the number of items in the sample space chosen for analysis.

sample space In statistics, the set of events, numbers, or other items chosen for analysis.

sampling **1.** Observation of a signal at various points in a circuit, without affecting the operation of the circuit. **2.** The conversion of analog signals to binary signals—especially for use in digital communications systems and in computers. **3.** In statistics and probability, a set of function values corresponding to specifically chosen points in the domain.

sampling rate The frequency with which samples are taken [e.g., 1/hr (one sample per hour) or 10/min (10 samples per minute)].

sampling window See WINDOW, **2**.

sand load A microwave power dissipator in which the absorptive material is a mixture of sand and carbon.

sapphire needle See SAPPHIRE STYLUS.

sapphire stylus A jewel-tipped stylus for disc recording and playback. It is noted for durability.

sat **1.** Abbreviation of *saturate.* **2.** Abbreviation of SATURATION. **3.** Abbreviation of SATELLITE.

satd Abbreviation of *saturated.*

satellite An artificial object sent into orbit around the earth or another planet. See, for example, ACTIVE COMMUNICATIONS SATELLITE and PASSIVE COMMUNICATIONS SATELLITE.

satellite communication Communication via one or more satellite transponders. Usually both stations are on the ground, although sometimes one or both stations are airborne or in space.

satellite processor **1.** In a computer, a microprocessor that is subsidiary to the central processing unit (CPU). **2.** In a data-processing system, a CPU used to handle the running of programs of secondary importance to the system's main application.

satellite television The broadcasting and reception of television (TV) signals via earth-orbiting satellite. Usually, the satellite is in a geostationary orbit. The receiving station employs a dish antenna, a tuner, a digital-to-analog (D/A) converter (if the signals are digital), and a TV receiver.

satisfy To make a statement of inequality or an equation true (e.g., $x = 2$ satisfies the equation $2x = x^2$).

saturable capacitor A voltage-variable ceramic or semiconductor capacitor in which variations in capacitance stabilize at a reasonably constant value after a particular voltage level is reached.

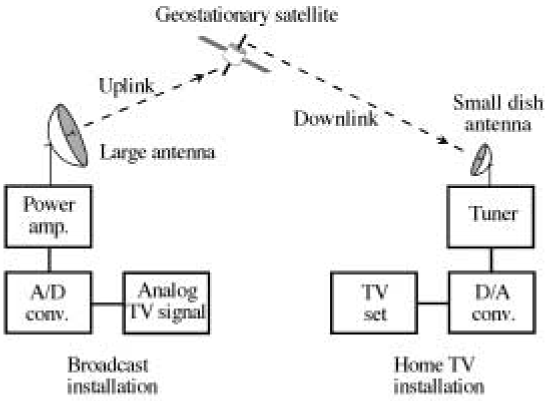

satellite television

saturable-core magnetometer A MAGNETOMETER in which the sensor is a saturable magnetic core with a winding. The readout is proportional to the change in permeability of the core produced by the magnetic field under test.

saturable reactor An inductor consisting essentially of a coil wound on a core of magnetic material whose magnetic flux can easily reach saturation level. The inductance and, accordingly, the reactance of the device can be varied by passing a direct current through the coil simultaneously with the alternating current to be controlled.

saturable transformer A transformer having a saturable core that permits automatic regulation of an alternating-current voltage.

saturated color A visible color whose energy is concentrated at a single wavelength or in a narrow band of wavelengths. Also called *pure color.*

saturated logic Any form of digital-logic circuit in which the transistors are either completely cut off or completely saturated. It is characterized by relative immunity to noise, high speed, and high input-level requirements.

saturated operation **1.** The operation of a magnetic core at or beyond its saturation point (i.e., in the region where an increase in coil current produces no change in core magnetization). **2.** The operation of a transistor or vacuum tube beyond its saturation point (i.e., in the region where an increase in voltage produces no change in current, or vice versa). Compare UNSATURATED OPERATION.

saturated solution A solution, such as an electrolyte, that contains all of the solute that it ordinarily will hold at a given temperature and pressure. Compare SUPERSATURATED SOLUTION. Also see SOLUTE; SOLUTION, **1** and SOLVENT, **2**.

saturating current See SATURATION CURRENT.

saturation **1.** See SATURATION POINT. **2.** The state of purity of a color. In general, the greater the saturation, the narrower the bandwidth of the electromagnetic energy. The highest possible saturation is represented by energy at a single wavelength [e.g., light of 700 nanometers (nm) appears as highly saturated red, and light of 400 nm appears as highly saturated violet]. Compare HUE. **3.** The condition of a ferromagnetic material in which it can accommodate no additional magnetic flux. **4.** The condition of a dielectric material in which it can accommodate no additional electric flux.

saturation current In a device, the current flowing at and beyond the SATURATION POINT.

saturation flux **1.** The magnetic flux density that will saturate a given sample of magnetic material. **2.** The electric flux density that will saturate a given sample of dielectric material.

saturation flux density See SATURATION INDUCTION.

saturation induction For a magnetic material, the maximum possible flux density.

saturation limiting Output peak clipping that occurs when a transistor or vacuum tube is driven into saturation during part of the input cycle. Compare CUTOFF LIMITING.

saturation point On a voltage-current conduction curve, the point beyond which a further increase in voltage produces no appreciable increase in current.

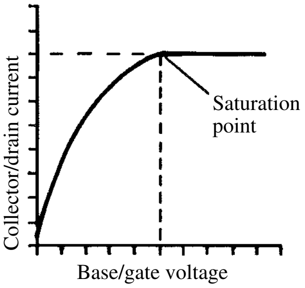

saturation point

saturation resistance The voltage-to-current ratio for a saturated semiconductor.

saturation switching The on/off switching operation in which a transistor is in its saturated state when conducting.

saturation value **1.** In a transistor, field-effect transistor, or vacuum tube, the lowest level of the input current, voltage, or power that results in saturation. **2.** The maximum obtainable output level for a given circuit. **3.** In a magnetic material, the smallest level of magnetizing force that results in the maximum possible flux density.

saturation voltage The (usually direct-current output) voltage appearing across a device operating in its saturation region (e.g., the collector voltage of a switching transistor in its switched-on state).

SAVOR Abbreviation of SIGNAL-ACTUATED VOICE RECORDER.

sawtooth An alternating or pulsating wave of current or voltage that is characterized by a gradual rise in amplitude followed by a rapid fall, or vice versa; its name is derived from its graphic resemblance to the teeth of a saw.

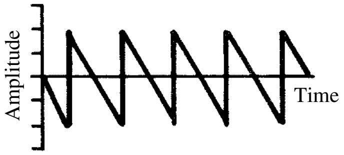

sawtooth

SB **1.** Abbreviation of SIDEBAND. **2.** Abbreviation of SIMULTANEOUS BROADCAST.

Sb Symbol for ANTIMONY.

S band A radio-frequency band extending from 1550 to 5200 MHz. For subdivisions of this band, see S_A BAND, S_C BAND, S_D BAND, S_F BAND, S_G BAND, S_H BAND, S_Q BAND, S_S BAND, S_T BAND, S_W BAND, S_Y BAND, and S_Z BAND.

SBC Abbreviation of SINGLE-BOARD COMPUTER.

SBDT Abbreviation of SURFACE-BARRIER DIFFUSED TRANSISTOR.

SBT Abbreviation of SURFACE-BARRIER TRANSISTOR.

SC **1.** Abbreviation of SUPPRESSED CARRIER. **2.** Abbreviation of SHORT CIRCUIT. **3.** On drawings, the abbreviation for *silk-covered.*

Sc Symbol for SCANDIUM.

sc **1.** Abbreviation of SINGLE CRYSTAL. **2.** Abbreviation of SCALE. **3.** Abbreviation of *sine-cosine.* **4.** Abbreviation of *science.*

SCA Abbreviation of SUBSIDIARY COMMUNICATION AUTHORIZATION.

SCA adapter An auxiliary tuner unit for separating the SCA subcarrier from a main frequency-modulated signal on which it is superimposed. Also see SCA SUBCARRIER and SUBSIDIARY COMMUNICATIONS AUTHORIZATION.

scalar quantity A quantity having magnitude, but for which direction is not specified. Compare VECTOR QUANTITY.

scale **1.** A graduated line or curve for indicating values of a quantity. **2.** An ordered set of values. **3.** An ordered series of quantities, such as tones, frequencies, voltages, etc. (e.g., *musical scale*).

scale division The space between consecutive graduations on a scale (see SCALE, **1**).

scale down In computer operations, to adjust a group of quantities according to a fixed factor so that it can be accommodated by hardware or software.

scale expansion Spreading out the divisions in part of a scale (see SCALE, **1**).

scale factor **1.** A figure by which the readings from a particular scale must be multiplied or divided to give the true values of measured quantities. **2.** A figure via which values in one system of notation are converted to those in another system. **3.** In scaling down (see SCALE DOWN), the factor by which a group of quantities is adjusted. **4.** The ratio of output frequency to input voltage for a voltage-to-frequency converter.

scale-factor adjustment In some meters, an adjustment that allows full-scale deflection to be set at any desired value (within certain limits) of applied-signal amplitude.

scale-factor error The absolute value of the difference between the actual scale factor and the ideal scale factor for a multiplier circuit.

scale-factor tolerance The extent to which a measured value for the scale factor differs from the computed value. It is generally given as a percentage.

scale length The end-to-end dimension of a scale (see SCALE, **1**), in inches, centimeters, geometric degrees, or number of divisions.

scale multiplier See SCALE FACTOR, **1**.

scale-of-two counter A circuit that delivers one output pulse for two successive input pulses.

scale-of-10 counter A circuit that delivers one output pulse for 10 successive input pulses.

scale-of-ten scaler See SCALE-OF-10 COUNTER.

scaler A circuit or device for extending the frequency range of another device (e.g., a circuit that extends the range of a 1-MHz counter to 100 MHz).

scale range The difference between the lowest and highest values on a scale.

scale span See SCALE RANGE.

scaling The fact, and the implications of the fact, that the mechanical strength of a structure increases in proportion to the square of linear dimension while the weight increases, according to the cube of linear dimension. Thus, weight increases more rapidly than strength as a structure, composed of a given material, is made larger. It is important in the design of large antennas and support structures.

scaling adder An inverting OPERATIONAL AMPLIFIER used to weight and sum multiple voltages.

scaling circuit A circuit, such as one or more flip-flops, that will deliver one output pulse after a predetermined number of input pulses have been received; therefore, it will provide pulse or frequency division. See, for example, SCALE-OF-TWO COUNTER.

scaling factor For a scaler, the number of input pulses required for one output pulse.

scaling ratio See SCALING FACTOR.

scan **1.** To traverse a range, field, or dimension. **2.** The amount of traversal in **1**. **3.** See SWEEP. **4.** To sample or reproduce an image in a single-line element, as in facsimile or television. **5.** A single line resulting from 4. **6.** In information retrieval operations, to inspect each record in a file or constituent of a list. **7.** To check communications or data channels for availability.

scan conversion In television reception, the scanning of each line twice to convert a conventional image into one that can be displayed on a high-definition picture tube.

scan-converter tube A face-to-face assembly of a cathode-ray tube and a vidicon in one envelope.

scandium Symbol, Sc. A metallic element. Atomic number, 21. Atomic weight, 44.956.

scan frequency See SCANNING FREQUENCY.

scanner A device, especially a radio receiver, equipped with a circuit that searches communications or data channels for signals.

scanner amplifier An amplifier for boosting a scanning signal. Also see SCAN, **1**, **3**, **4**.

scanning **1.** In a cathode-ray tube or camera tube, the synchronized movement of the electron beam (or other marker) from right to left and/or from top to bottom. **2.** The intermittent, but repetitive, monitoring of two or more communications channels in rotating sequence. **3.** The movement of a radar beam for the purpose of obtaining coverage over a specified area.

scanning antenna A transmitting or receiving antenna (such as a rotating one) that covers a generally circular region.

scanning beam The deflected electron beam in a cathode-ray tube. Also see SCAN, **1**, **2**, **3**, **4**, **5**.

scanning circuit A circuit for producing a scan (see SCAN, **2**, **4**).

scanning frequency The number of scans per unit time, usually expressed in lines, sweeps, or channels per second or per minute. Also called SCANNING RATE.

scanning line A single line sampled or produced by a scanning process, as in facsimile, television, and graphic recording.

scanning line frequency See SCANNING FREQUENCY.

scanning linearity Uniformity of scanning rate. In a linear scan, for example, scan speed is the same at all points along a line.

scanning loss The effective reduction in radar sensitivity that occurs as the beam scans a given area, rather than remaining in a fixed orientation.

scanning rate See SCANNING FREQUENCY.

scanning receiver A receiver whose tuning is automatically and continuously swept through a frequency band to detect all signals in the band.

scanning sonar A form of distance-measuring or depth-finding sonar, in which the receiving transducer scans to find the direction of the echo or echoes.

scanning speed The rate at which a line, region, or quantity is scanned or at which samples are taken.

scanning yoke See YOKE, **2**.

scan rate **1.** The rate at which a controlled quantity is checked periodically by a control computer. **2.** See SCANNING FREQUENCY.

scan tuning Repetitive, automatic sweeping of a frequency band by a tuned circuit containing a varactor, whose capacitance is periodically varied by a sawtooth voltage.

SCARA Abbreviation of SELECTIVE COMPLIANCE ASSEMBLY ROBOT ARM.

SCA subcarrier An auxiliary carrier (commonly 67 kHz) superimposed on a main frequency-modulated carrier to convey subsidiary communications, such as music without commercials. Also see SUBSIDIARY COMMUNICATION AUTHORIZATION.

scatter To disperse or diffuse transmitted electromagnetic radiation.

scattering **1.** The tendency of a concentrated beam of energy to be spread out when it passes through a given medium or substance. **2.** The spreading out of radio waves as they pass through the ionosphere or troposphere.

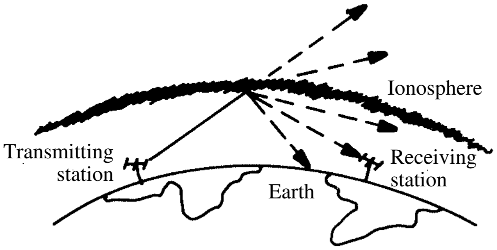

scattering

scatter read In data processing, to distribute data from an input record to several storage areas.

scatter transmission See BACK SCATTER and FORWARD SCATTER.

S_C band A section of the S BAND extending from 2000 to 2400 MHz.

scc Abbreviation of SINGLE-COTTON-COVERED (WIRE).

SCDSB Abbreviation of SUPPRESSED CARRIER DOUBLE-SIDEBAND (see DOUBLE-SIDEBAND SUPPRESSED CARRIER).

SCE Abbreviation of *saturated calomel electrode.*

sce Abbreviation of SINGLE-COTTON-ENAMELED (WIRE).

SCEPTRON Acronym for SPECTRAL COMPARATIVE PATTERN RECOGNIZER.

schedule In computer operations, to establish the order of importance of jobs to be run, and assign the necessary resources for those jobs.

schedule method A method of waveform analysis involving the evaluation of instantaneous amplitudes at numerous points in time. The values are obtained at specific intervals from the image of one complete wave cycle, as displayed on an oscilloscope or plotted on a graph.

schematic diagram An electronic circuit diagram. Also called WIRING DIAGRAM and *schematic.*

schematic symbol A graphic symbol used to represent electronic components in a circuit diagram.

Schering bridge A four-arm capacitance bridge in which the unknown capacitance is compared with a standard capacitance. This bridge is frequently used in testing electrolytic capacitors, to which a direct-current polarizing voltage is applied during the measurement.

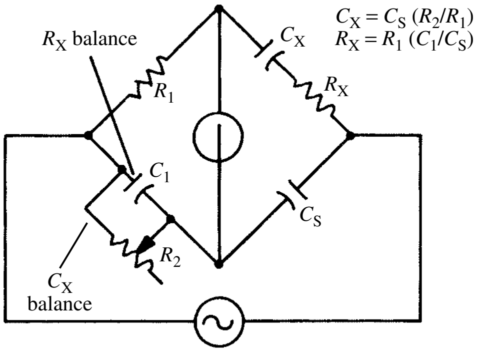

Schering bridge

Schmidt optical system In projection television, a lens system used between the bright-image picture tube and the screen.

Schmitt limiter See SCHMITT TRIGGER.

Schmitt trigger A multivibrator that produces uniform-amplitude output pulses from a variable-amplitude input signal. This circuit has many applications, one being the conversion of a sine wave into a square wave.

Schottky diode A solid-state diode in which a metal and a semiconductor form the pn junction. Electrons injected into the metal have a higher energy level than the charge carriers in a semiconductor, and energy storage at the junction is low because current flow is not accompanied by hole movement. Also known as HOT-CARRIER DIODE.

Schottky logic A form of integrated-injection logic with enhanced operating speed.

Schottky noise The random noise resulting from the emission of charged particles, usually electrons or holes, from an electrode in an amplifying device. This noise is usually of a wideband nature.

scientific language Any computer programming language used primarily for mathematical or scientific applications.

scientific notation The expression of very large and very small numbers as a fixed-point part (mantissa) and a power of the radix (usually 10).

Generally, the mantissa is greater than or equal to 1, but less than 10; the power of 10 is adjusted accordingly. Thus, for example, $203{,}700 = 2.037 \times 10^5$; $0.000533 = 5.33 \times 10^{-4}$. See also SIGNIFICANT FIGURES.

scintillating crystal A crystal, such as one of sodium iodide, that sparkles or flashes when exposed to radioactive particles or rays.

scintillation **1.** In radar operations, the apparent rapid displacement of a target from its mean position. **2.** A momentary flash of light produced in a phosphor or scintillating crystal when a high-velocity particle strikes it. **3.** A small fluctuation in radio field intensity at a receiving point.

scintillation counter A radiation counter consisting essentially of a scintillating crystal in combination with a photomultiplier tube. Flashes from the excited crystal cause the tube to deliver output pulses that are totaled and indicated.

scintillator material A substance, such as crystalline sodium iodide, that scintillates under certain stimuli.

scissoring A method of interrupting the electron beam in a cathode-ray tube when the beam would not land on the phosphor screen.

SCLC Abbreviation of SPACE-CHARGE-LIMITED CURRENT.

sco Abbreviation of SUBCARRIER OSCILLATOR.

scope Colloquialism for OSCILLOSCOPE.

Scott oscillator See PARALLEL-TEE OSCILLATOR.

scp Abbreviation of SPHERICAL CANDLEPOWER.

SCR Abbreviation of SILICON-CONTROLLED RECTIFIER.

scrambled signal Any signal in which (for secrecy or exclusivity) the elements are disarranged according to an encryption algorithm. Thus, intelligent reception is possible only if the signal is processed via a suitable decryption algorithm. Example: SCRAMBLED SPEECH.

scrambled speech Voice transmission in which the frequencies have been inverted to prevent eavesdropping. It is automatically unscrambled (by reinversion) at the receiver to restore intelligibility.

scrambler circuit A circuit containing filters and frequency inverters for scrambling speech.

scratch filter An audio-frequency low-pass filter that suppresses high-frequency noise caused by friction between a phonograph disc and the needle.

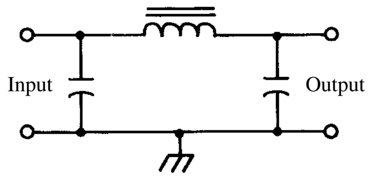

scratch filter

scratchpad memory In computers, a low-capacity memory that stores an intermediate result of a calculation.

scratch tape Magnetic data tape that can be overwritten for any purpose.

screen **1.** See SCREEN GRID. **2.** See SHIELD. **3.** The front surface of a cathode-ray tube. **4.** The surface of a computer or terminal display.

screen angle In radar, the angular difference between the actual horizon and the plane perpendicular to the line connecting the radar set with the center of the earth.

screen current Symbol, I_S or I_{SG}. The current flowing in the screen circuit of an electron tube.

screen grid In a vacuum tube, a grid element between the control grid and plate. It reduces the internal grid-plate capacitance, and consequently prevents self-oscillation when the tube is used in a straight-through amplifier.

screen-grid neutralization Neutralization of an amplifier that uses a tetrode vacuum tube. Such circuits require smaller neutralizing capacitances than those used in triode amplifiers because of the lower interelectrode capacitance of the screen-grid tube.

screen illumination Edge lighting of the transparent screen of an oscilloscope, to make the lines of the graticule more clearly visible.

screen material See PHOSPHOR.

screen resistance Symbol, R_S or R_{SG}. The internal resistance presented by the screen-grid/cathode path of an electron tube.

screen room See CAGE.

screen voltage Symbol, E_S or E_{SG}. The voltage at the screen grid of an electron tube.

scribing The etching of a semiconductor wafer to facilitate breaking the wafer into smaller pieces.

ScriptX A high-level programming language used in the writing of software for personal computers—especially in multimedia. Developed by Kaleida Laboratories.

SCS Abbreviation of SILICON-CONTROLLED SWITCH.

SCT Abbreviation of SURFACE-CHARGE TRANSISTOR.

S curve **1.** The voltage-versus-current curve for a negative-resistance device. Compare N CURVE. **2.** The response curve for a frequency-modulation discriminator or ratio detector.

SD Abbreviation of STANDARD DEVIATION.

S_D band A section of the S BAND extending from 4200 to 5200 MHz.

SDF Abbreviation of static direction finder.

SDR See SOFTWARE-DEFINED RADIO.

Se Symbol for SELENIUM.

seaborgium Symbol, Sg. Also called *unnilhexium* (Unh). Atomic number, 106. The most common isotope has atomic weight 263. Classified as a transition metal. It has a half-life on the order of 1 second or less, is human-made, and is not known to occur in nature.

sea clutter Collectively, the radar echoes that the sea reflects.

seal **1.** The point at which a lead or electrode enters or leaves and is secured to an envelope, case, or housing. Such a point is often tightly closed against the passage of air in or out of the envelope. **2.** To close off a circuit or component from tampering.

sealed dry battery A set of electrochemical dry cells that can be installed without concern for orientation or position. Example: 9-volt “transistor battery.”

sealed meter **1.** A meter that is tightly closed against the entry of moisture and foreign materials. **2.** A meter that is locked or otherwise protected against tampering.

sealing compound A substance (such as wax, pitch, or plastic) used to enclose and protect electronic devices.

search **1.** To scan or sweep through a range of quantities or through a region of interest. **2.** To examine (usually in some prescribed order) items of information in a computer memory to find those satisfying a given criterion.

search coil An inductive probe (exploring coil) used to sample magnetic fields.

search oscillator A variable-frequency oscillator used to locate and identify signals by the heterodyne method.

search probe **1.** See SEARCH COIL. **2.** A capacitive probe used to sample electric fields.

search radar A radar that displays a target almost immediately after that target enters a scanned area.

search time The time needed to test items during a search (see SEARCH, **2**).

sea return See SEA CLUTTER.

seasonal effects In ionospheric propagation, the changes produced as a result of the revolution of the earth around the sun. The path of the sun across the sky, and the length of the day, are primarily responsible for such effects.

seasonal static Atmospheric electrical interference, most prevalent during the summer.

S_E band A section of the S BAND extending from 1550 to 1650 MHz.

sec **1.** Abbreviation of SECOND. **2.** Abbreviation of SECTION. (Also, sect.) **3.** Abbreviation of SECONDARY. **4.** Abbreviation of SECANT.

secant Abbreviation, sec. The trigonometric function representing the ratio of the hypotenuse of a right triangle to the adjacent side (c/b). The secant is the reciprocal of the cosine; $\sec x = 1/\cos x$.

second **1.** Abbreviation, s and sec. A unit of time. The mean solar second is 1/86,400 of a mean solar day, and is 1/60 minute or 1/3600 hour. **2.** Symbol ("). A unit of arc measure. 1" = 1/3600 geometric degree.

secondaries See SECONDARY COLORS.

secondary **1.** See SECONDARY WINDING. **2.** See SECONDARY STANDARD.

secondary battery See STORAGE BATTERY.

secondary calibration The calibration of an instrument, based on a reference instrument calibrated against an absolute source.

secondary capacitance **1.** The distributed capacitance of the secondary winding of a transformer whose primary winding is unloaded. Compare PRIMARY CAPACITANCE, **1**. **2.** A series or shunt capacitance used to tune the secondary coil of a radio-frequency transformer. Compare PRIMARY CAPACITANCE, **2**.

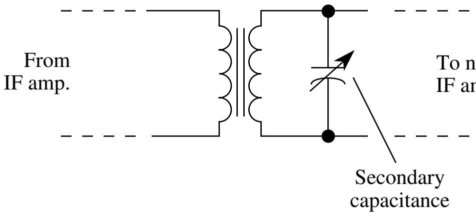

secondary capacitance, 2

secondary cell See STORAGE CELL.

secondary circuit **1.** The circuit associated with the secondary winding of a transformer. **2.** See OUTPUT CIRCUIT.

secondary coil See SECONDARY WINDING.

secondary color **1.** A color prepared by mixing two primary colors. **2.** In television operations, any displayed color composed of two or more color primaries.

secondary current The current flowing in the secondary winding of a transformer. Also called TRANSFORMER OUTPUT CURRENT. Compare PRIMARY CURRENT.

secondary electron **1.** The electron possessing the lesser energy after a collision between two electrons. Compare PRIMARY ELECTRON. **2.** An electron ejected by secondary emission.

secondary emission The action whereby electrons in the atoms at the surface of a target are ejected as a result of bombardment by a beam of (primary) electrons. Thus, in an electron tube, electrons from the cathode strike the plate with a force that drives secondary electrons out of the plate, into the surrounding space.

secondary emitter A source of secondary electrons (e.g., the plate of an electron tube or a dynode in a photomultiplier tube).

secondary failure The failure of a component or circuit, resulting from the failure of some other component. For example, the pass transistor in a power supply might burn out, causing the output voltage to increase; this increased voltage can damage equipment connected to the supply.

secondary frequency standard A device for generating signals of accurate frequency, but that does not possess the very high stability and extreme accuracy of a primary frequency standard. The secondary standard is periodically checked against a PRIMARY FREQUENCY STANDARD and appropriately corrected.

secondary impedance **1.** The impedance of the secondary winding of a transformer whose primary winding is unloaded. Compare PRIMARY IMPEDANCE, **1**. **2.** An external impedance presented to the secondary winding of a transformer. Compare PRIMARY IMPEDANCE, **2**.

secondary inductance The inductance of the secondary winding of a transformer whose primary winding is unloaded. Compare PRIMARY INDUCTANCE.

secondary kVA The kilovolt-amperes in the secondary circuit of an operating transformer. Compare PRIMARY KVA.

secondary power The power in the secondary circuit of a transformer. Also see SECONDARY KVA and SECONDARY VA. Compare PRIMARY POWER.

secondary radiation **1.** The (sometimes random) reradiation of electromagnetic waves, as from a receiving antenna. **2.** Rays emitted by atoms or molecules when the latter are struck by other radiation.

secondary rays Rays emitted by atoms or molecules that have been bombarded by other rays of the same general nature. Examples: *secondary X rays* and *secondary beta rays.*

secondary resistance The direct-current resistance of the secondary winding of a transformer. Compare PRIMARY RESISTANCE.

secondary standard An accurate source of a quantity (capacitance, frequency, inductance, resistance, etc.), that is referred periodically to a primary standard for correction.

secondary storage In computer and data-processing operations, storage that is auxiliary to the main storage. Also called *backing storage.*

secondary turns The number of turns in the secondary winding of a transformer. Compare PRIMARY TURNS.

secondary utilization factor For a transformer in a rectifier circuit, the utility factor of the secondary winding (ratio of direct-current power output to secondary volt-amperes).

secondary VA The volt-amperes in the secondary circuit of a transformer. Compare PRIMARY VA.

secondary voltage The voltage across the secondary winding of a transformer. Also called TRANSFORMER OUTPUT VOLTAGE. Compare PRIMARY VOLTAGE.

secondary winding The normal output winding of a transformer. Also called SECONDARY COIL. Compare PRIMARY WINDING.

second breakdown In a large-area power transistor, a destructive breakdown caused by thermal runaway.

second-breakdown voltage The collector voltage at which second breakdown occurs in a transistor.

second-channel attenuation See SELECTANCE, **2**.

second-channel interference In a given channel, interference arising from authorized signals two channels removed.

second detector In a superheterodyne receiver, the intermediate-frequency detector. Compare FIRST DETECTOR.

second-level address In a computer program instruction, an address giving the location of the address of a required operand. Also called *indirect address.*

sect Abbreviation of SECTION. (Also, sec.)

section **1.** A subcircuit or stage of a larger circuit (e.g., the *oscillator section* of a receiver). **2.** The smaller unit described in **1**, when self-contained and operated independently (e.g., *filter section*). **3.** See PROGRAM SEGMENT.

sectionalized antenna A set of collinear radiating elements, placed end-to-end with reactances between them, for the purpose of modifying the radiation pattern.

sectionalized winding **1.** A method of winding a coil in complete, multilayer sections that are stacked side by side or top to bottom, a technique that reduces distributed capacitance. **2.** A coil wound, as defined in **1**.

sector On a magnetic disk, a specific portion of a track.

sectoral horn antenna An (usually sheetmetal) antenna with the shape of a horn of rectangular cross section. It is flared in one dimension only.

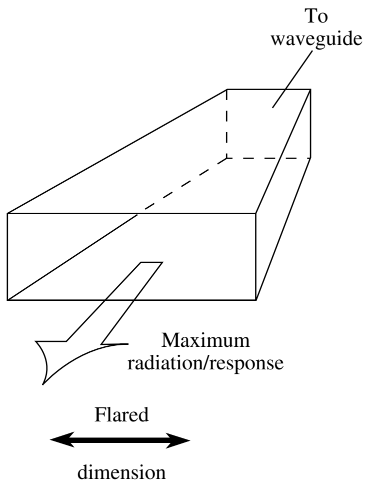

sectoral horn antenna

sector display In a radar, a display that allows the continuous observation of a portion of the scanned area.

secular equilibrium The state in which a radioactive substance changing into another substance is decaying as fast as the second substance is being formed.

secular variation A slow change in the intensity of the terrestrial magnetic field.

secure mode In a security system, the condition of being fully activated.

securite In radiotelephony, a spoken word (pronounced say-kyoor-ee-tay) identifying a transmission concerning safety. Equivalent to TT in radiotelegraphy.

security code **1.** A set of alphanumeric characters (letters and/or numbers) or switch settings that activates or deactivates a security system. **2.** See PASSWORD.

security robot In a smart home, a robot that assists in protection of a home or business, and its occupants, from intrusion, burglary, or attack. If an intruder enters the property, the security robot might drive the offender away or detain the offender until police arrive. The robot might employ iris-print detection or facial feature detection to identify an intruder from a distance. The robot would be linked to the home computer via a broad-bandwidth, high-speed wireless system. Compare SENTRY ROBOT. Also see SMART HOME OR BUSINESS.

security system A set of electronic (and sometimes also electromechanical) devices designed to do one or more of the following: restrict access to a premises or computer system, detect abnormal conditions, detect unauthorized entry, alert human operators of abnormal conditions, alert owners and/or authorities of unauthorized entry to a premises, and (in some cases) provide physical protection of property.

Seebeck effect The development of an electromotive force in a junction of two dissimilar metals (a thermocouple) when the temperature of the junction is different from that of the rest of the metal.

Seebeck emf The electromotive force resulting from the Seebeck effect.

seed crystal A small single crystal from which a larger single crystal (e.g., germanium or silicon) is grown. Also see CZOCHRALSKI METHOD.

seeing-eye robot A robot that serves in the capacity of a seeing-eye dog, to help visually impaired people in their daily activities.

seek See SEARCH.

seek area An area of direct-access storage to which are assigned specific records and from which the records can be accessed quickly.

segment **1.** The portion of a line or curve lying between two points. **2.** See PROGRAM SEGMENT.

segmental meter An expanded-scale meter (see SCALE EXPANSION).

segmented encoding law An approximation of a smooth encoding law, done by means of linear partitions or segments. The greater the number of segments, the more accurate the approximation.

segmenting See PARTITIONING.

segment mark A character indicating the division between tape file sections.

segue (Pronounced SEG-way.) A smooth transition from one sound or channel to another (e.g., between musical selections in a radio broadcast). The first selection decreases in volume as the second selection increases. There is some overlap.

Seignette salt See ROCHELLE SALT.

seismic detector A vibration sensor used in some types of intrusion-detection systems.

seismogram A record produced by a SEISMOGRAPH.

seismograph An instrument for detecting and recording earth tremors. Indicates the direction, magnitude, and time of a quake.

seismometer See SEISMOGRAPH.

seismoscope An instrument that shows the occurrence and time of an earthquake. Compare SEISMOGRAPH.

select To accept or separate a unit, quantity, or course of action from all those available (in a group, mixture, or series).

selectance **1.** For a resonant circuit, the ratio Er/Ex, where Er is the voltage at resonance, and Ex is the voltage at a specified nonresonant frequency. **2.** For a receiver, the figure $S2/S1$, where $S1$ is the sensitivity of the receiver in a given frequency channel, and $S2$ is the sensitivity in another specified channel.

selected mode In an encoder, a mode in which one output is read and the others are ignored.

selective Pertaining to a device or system, such as a radio receiver, with a high degree of SELECTIVITY.

selective absorption The attenuation or absorption of some frequencies or bands of frequencies, with little or no attenuation at other frequencies or bands of frequencies.

selective amplifier An amplifier that can be tuned, with the desired degree of sharpness, to a single frequency or band of frequencies. Radio-frequency amplifiers are generally tuned by means of inductance-capacitance (LC) circuits; audio-frequency amplifiers are usually tuned by means of resistance-capacitance (RC) circuits.

selective calling The calling, alerting, or alarming of a desired station without interfering with other stations.

selective compliance assembly robot arm Abbreviation, SCARA. An electromechanical device designed especially for use in assembly lines. It uses cylindrical coordinate geometry to allow precise, programmable movements in three dimensions.

selective dump In computer operations, a dump (see DUMP) affecting a small, specific memory area.

selective fading Fading caused by propagation conditions, whose effects differ at slightly different frequencies. In an amplitude-modulated signal, this effect causes the sidebands and carrier to arrive in various phase relationships, with resulting distortion in the received signal.

selective interference Interference confined to a narrow band of frequencies.

selective polarization See POLARIZATION SELECTIVITY.

selective reflection In the reflection of electrons directed into a crystal by means of an electron gun, the tendency of the electrons to be reflected more readily when they strike the crystal at certain speeds.

selective relay **1.** A relay or relay circuit tuned to open or close at one signal frequency. **2.** A relay or relay circuit adjusted to open or close at one value of current or voltage.

selective trace In computer operations, a diagnostic program used to analyze certain areas of memory or specific kinds of program instructions, for debugging purposes.

selectivity **1.** The ability of a circuit or device to pass signals of one frequency and reject signals at other frequencies. **2.** The degree to which a circuit or device passes signals of one frequency and rejects signals at other frequencies.

selectivity control In some equipment (such as receivers, crystal filters, wave analyzers, and vibration meters) an adjustment that permits variation of selectivity.

Selectoject A fully electronic, continuously tunable, notcher-peaker that is resistance-capacitance tuned. The name is an acronym for *select or reject.*

selector **1.** A channel switch in a radio or television receiver. **2.** See SELECTOR SWITCH.

selector channel In data-processing and computer systems, a data transmission channel controlling the information flow between peripherals and a central processing unit.

selector pulse In digital communications, an identifying pulse that represents a certain group of bits or data.

selector relay A device, such as a stepping switch, that actuates one of a number of available circuits on receipt of a predetermined number of pulses.

selector switch A (usually rotary) multiposition switch that allows an operator to select from among several options (such as frequency channels, frequency bands, or selective filters).

selenium Symbol, Se. A nonmetallic element. Atomic number, 34. Atomic weight, 78.96. It is used in the manufacture of some diodes, rectifiers, and photocells.

selenium cell A photoelectric cell that uses specially processed selenium as the light-sensitive material. It can be operated as a photoconductive cell or a photovoltaic cell.

selenium diode A junction diode in which the semiconductor material is specially processed selenium. Also see JUNCTION DIODE.

selenium photocell See SELENIUM CELL.

selenium rectifier A disk- or plate-type power rectifier utilizing the junction between selenium and aluminum or selenium and iron.

self-bias For a transistor or vacuum tube, input-electrode bias voltage resulting from the flow of output-electrode current through a resistor common to both circuits. Also called AUTOMATIC BIAS.

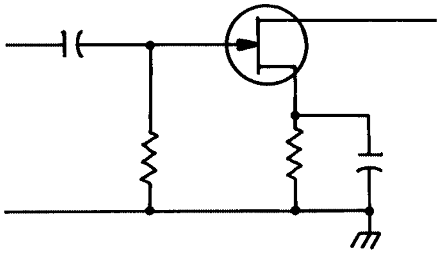

self-bias

self-capacitance The inherent internal capacitance of a device other than a capacitor.

self-checking number A number whose digits have a value that determines the check digit attached to it; thus, it can be verified following its transfer between storage locations or peripherals.

self-cleaning contacts Switch or relay contacts that clean themselves automatically by means of wiping action.

self-contained device A device containing all parts and sections (e.g., main circuit, power supply, meter, loudspeaker, etc.) needed for full operation (i.e., no auxiliary equipment is needed).

self-controlled oscillator See SELF-EXCITED OSCILLATOR.

self discharge The tendency of an electrochemical cell or battery to gradually lose stored energy when not in use.

self-discharge rate A quantitative expression of the speed with which SELF DISCHARGE occurs in an electrochemical cell or battery when it is stored without being used.

self-energy Symbol, *E.* The energy mc^2, in joules, of a particle traveling at the speed of light c (2.9979×10^8 meters per second) and whose mass is m kilograms.

self erasing In a magnetic tape, the unwanted erasing of data near a highly magnetized region.

self-excited generator An alternating-current generator in which the field coils are supplied with direct current produced by the machine itself. Compare SEPARATELY EXCITED GENERATOR.

self-excited oscillator An oscillator consisting of an amplifier that supplies its own input signal through positive feedback, and whose oscillation frequency depends entirely on circuit constants, such as the capacitance and inductance in a tank circuit. Compare CRYSTAL OSCILLATOR, FORK OSCILLATOR, HUMMER, and MAGNETOSTRICTION OSCILLATOR.

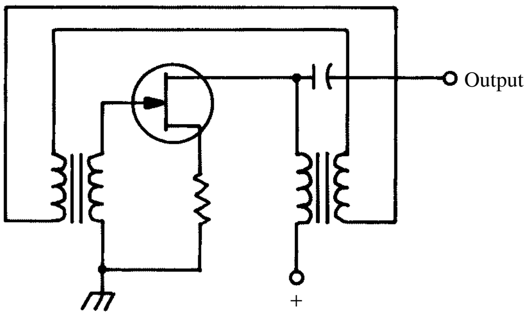

self-excited oscillator

self-focus picture tube A television picture tube in which the electron gun has an automatic, electrostatic focusing arrangement.

self-generating photocell See PHOTOVOLTAIC CELL.

self-generating transducer A voltage-producing transducer, such as a piezoelectric pickup or dynamic microphone.

self-healing capacitor A capacitor, such as a wet electrolytic unit, in which the dielectric is restored to its normal condition after a high-voltage breakdown.

self-heated thermocouple A thermocouple in which the passage of an applied current produces the heat necessary for the activation of the thermocouple.

self-heating thermistor A thermistor heated to above ambient temperature by the current passing through it. Also called DIRECTLY HEATED THERMISTOR. Compare INDIRECTLY HEATED THERMISTOR.

self impedance The effective or measured impedance at a circuit point.

self-inductance **1.** The inductance of an inductor. **2.** The inherent internal inductance of a device other than an inductor.

self-induction Induction that occurs in a single circuit. An instance is the generation of an opposing voltage across a coil by an alternating current flowing through it. Compare INDUCTION. Also see INDUCTANCE.

self-latching relay A relay that remains in the state that it has been switched (i.e., locked open or closed) until a subsequent signal is received.

self-modulated oscillator A circuit, such as a blocking oscillator, in which oscillation occurs simultaneously at two frequencies, one modulating the other.

self-organizing Pertaining to a computer or system that can change the arrangement of data files for particular purposes.

self-powered device A device that requires no external source of power (i.e., it is equipped with a self-contained battery or a generator).

self-pulsing blocking oscillator A blocking oscillator that produces a train of radio-frequency pulses. Compare SINGLE-SWING BLOCKING OSCILLATOR.

self-quenching detector A super-regenerative detector (see SUPER-REGENERATIVE CIRCUIT) in which the low-frequency quenching voltage is supplied by the regenerative detector itself. Also see QUENCHING ACTION and QUENCH OSCILLATOR. Compare SEPARATELY QUENCHED DETECTOR.

self-quenching oscillator A circuit, such as a blocking oscillator, in which oscillation is periodically switched off automatically, resulting in a self-interrupted wave train.

self-rectifying vibrator A vibrator-type power supply in which one vibrator reed chops the direct-current input to the primary winding of the transformer, and a second vibrator reed rectifies the alternating-current output delivered by the secondary winding. Also see VIBRATOR RECTIFIER.

self-rectifying X-ray tube An X-ray tube operated with alternating-current anode voltage.

self-regulation The ability of a circuit or device to control its output automatically, according to some predetermined plan, by using output error to correct operation or to vary the input.

self reset **1.** The action of a circuit breaker to reapply power after a certain elapsed time. **2.** The action of any device, returning a circuit or system to normal automatically.

self-resetting loop In a computer program, a loop in which instructions cause locations used in the loop to assume their condition prior to the loop's execution.

self-resistance The inherent internal resistance of a device other than a resistor.

self-resonant frequency The frequency at which a device will resonate naturally (without external tuning). Thus, an inductor will self-resonate with its distributed capacitance; similarly, a capacitor will resonate with its stray inductance.

self-saturation In a magnetic amplifier, saturation resulting from rectification of the saturable-reactor output current.

self-starting motor An alternating-current motor that starts running as soon as voltage is applied (i.e., no external mechanical force is needed). Also see SHADING COIL.

self-sustained oscillations Oscillations maintained by means of positive feedback (inductive or capacitive) from the output to the input of a circuit. See, for example, SELF-EXCITED OSCILLATOR.

self-test Any arrangement whereby a device or system determines, without the aid of an external operator, whether or not it is operating correctly.

self-ventilated motor See OPEN MOTOR.

self-wiping contacts See SELF-CLEANING CONTACTS.

selsyn See AUTOSYN and SYNCHRO.

SEM **1.** Abbreviation of SINGLE-ELECTRON MEMORY. **2.** Abbreviation of SCANNING ELECTRON MICROSCOPE.

semantic network A reasoning scheme sometimes used in artificial intelligence. Logical state-ments or sentences are broken down into nodes (generally nouns) and relationships (generally verbs and modifiers). This allows statements to be mapped in a way that is easy for computers to store and modify.

semiautomatic key Also called *bug.* A telegraph key that mechanically produces a string of Morse-code dots (short pulses) when its lever is pressed to one side (usually toward the right), and continuous circuit closure when the lever is pressed to the opposite side (usually toward the left). Dashes are manually sent by the operator.

semiconductor A material whose natural resistivity lies between that of conductors and insulators (e.g., GERMANIUM, SILICON, SELENIUM, and GALLIUM ARSENIDE).

semiconductor capacitor A miniaturized component that takes advantage of reverse biasing in a semiconductor P-N junction. When a voltage source is connected across a diode so that it does not conduct, the diode acts as a capacitor. The capacitance varies depending on how much reverse voltage is applied to the diode. The greater the reverse voltage, the smaller the capacitance. Some diodes are especially manufactured to serve this function. This phenomenon can be useful in the fabrication of integrated circuits (ICs). See also INTEGRATED CIRCUIT, VARACTOR.

semiconductor counter A device for measuring the intensity of ionizing radiation (such as alpha particles, beta particles, or gamma rays) using a photodiode and sensing circuit.

semiconductor device A component (such as a diode, photocell, rectifier, or transistor) that exploits the properties of a semiconductor.

semiconductor diode A solid-state diode, as opposed to a vacuum-tube diode or gas-tube diode. Examples: *germanium diode*, *selenium diode*, and *silicon diode.*

semiconductor junction Within a body of semiconductor material, the area of physical contact between two regions (usually n and p) having opposite electrical properties.

semiconductor laser See LASER DIODE.

semiconductor material See SEMICONDUCTOR.

semiconductor-metal junction The area of physical contact between a metal and a semiconductor.

semiconductor photosensor A semiconductor photodiode or phototransistor, as opposed to a phototube.

semiconductor rectifier A heavy-duty semiconductor diode (or assembly of such diodes) designed primarily to change alternating current to direct current in power-supply units. Rectifiers commonly are made from copper oxide, germanium, magnesium-copper sulfide, selenium, or silicon. Also see JUNCTION DIODE and METER RECTIFIER.

semiconductor resistor A tiny resistor manufactured from semiconductor material, especially one that is etched onto the chip of an integrated circuit (IC). The thickness, and the types and concentrations of impurities added, determine the resistance of the component. Such resistors can handle only a tiny amount of power because of their small size. But because IC circuits are designed to consume minimal power, this is not a problem. The small signals produced by ICs can be amplified, using circuits made from discrete components, if it is necessary to obtain higher signal power. See also INTEGRATED CIRCUIT.

semidirectional Pertaining to a transducer that exhibits different directional characteristics at different frequencies.

semiduplex operation A two-frequency communication system that operates in duplex at one end of the link, and in simplex at the other end. Also see DUPLEX OPERATION and SIMPLEX TELEGRAPHY.

semilogarithmic graph Also called *semilog graph.* A graph in which one axis is logarithmic and the other axis is linear.

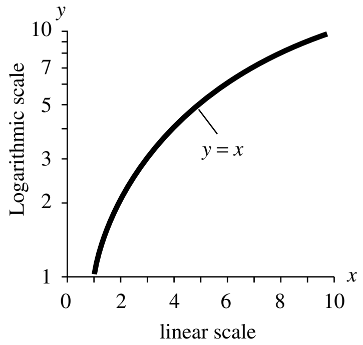

semilogarithmic graph

semimetal An elemental substance that exhibits some, but not all, of the properties of a metal (e.g., antimony and arsenic). Also called METALLOID.

semiresonant line An open-wire transmission line cut approximately to resonant length at the frequency of operation.

semitone See HALF STEP.

sender See TRANSMITTER, **1**, **3**.

sending-end impedance See DRIVING-POINT IMPEDANCE.

sending set **1.** See RADIO TRANSMITTER. **2.** An equipment for transmitting electromagnetic waves. Also see TRANSMITTER, **1**.

sensation level The level of sound that produces a tingling or noticeable sensation in the ear.

sense **1.** To check the condition of a switching device, such as a gate. **2.** See READ.

sense amplifier A device that produces a control signal when some characteristic of the input signal changes.

sense determination In a direction finder that provides an ambiguous indication (two readings 180 degrees apart), the process of deducing, or the ability of the apparatus to deduce, the true direction from which the signal is coming.

sense resistor A (usually low-value) resistor used to sense current in a circuit without introducing a significant loss. The voltage drop across this resistor is proportional to the current and can be applied to a voltmeter, oscilloscope, or other instrument for measurement or observation.

sensing circuit **1.** A circuit that samples a quantity. **2.** In a voltage regulator, the circuit that monitors the output voltage and delivers a control voltage proportional to the output-voltage error.

sensing window See WINDOW, **2**.

sensitive communications Two-way communications of an emergency or priority nature, or involving the security of government operations.

sensitive device A device that responds to a signal of low amplitude.

sensitivity **1.** The ability of a circuit or device to respond to a low-level applied stimulus. **2.** For a receiver, the input-signal (in microvolts or millivolts) required for a specified output level. **3.** For a galvanometer, microamperes or milliamperes per scale division. **4.** The ohms-per-volt rating of a voltmeter. Also see VOLTMETER SENSITIVITY.

sensitivity adjustment **1.** An input gain control in an amplifier circuit. **2.** The radio-frequency gain control of a receiver. **3.** A control or switch that is used to select the range or threshold of a piece of test equipment.

sensitivity control A manual or automatic device for adjusting the sensitivity of a circuit or device.

sensitometer An instrument used to measure the sensitivity of certain materials to light.

sensor **1.** A device that samples a phenomenon, and delivers a proportionate current or voltage in terms of which the intensity of the phenomenon can be measured, or with which control action can be initiated. **2.** An electronic device that detects abnormal conditions (e.g., smoke and heat) and delivers a warning signal to human operators and/or computers. **3.** An electronic device that detects intrusion to a premises and delivers a warning signal to human operators and/or computers.

sentry robot In a smart home or business, a robot that alerts the owner to abnormal conditions. It can detect fire, burglars, or water in places it should not be. It might also detect abnormal temperature, barometric pressure, wind speed, humidity, or air pollution. A wireless link alerts the owner via a device similar to a common beeper. Compare SECURITY ROBOT. Also see SMART HOME OR BUSINESS.

separately excited generator An alternating-current generator whose field coils are supplied with direct current from another generator or from a battery. Compare SELF-EXCITED GENERATOR.

separately quenched detector A superregenerative detector (see SUPERREGENERATIVE CIRCUIT) in which the quenching voltage is supplied by a separate low-frequency oscillator. Also see QUENCHING ACTION and QUENCH OSCILLATOR. Compare SELF-QUENCHED OSCILLATOR.

separation See CHANNEL SEPARATION.

separation energy The energy required to remove a proton or neutron from the nucleus of an atom. The separation energy depends on the atomic number.

separator **1.** See FILTER, **1**. **2.** A perforated or porous plate of insulating material (usually plastic or wood) for holding active plates apart in a storage cell. **3.** In computer operations, a character marking the division between logical data units. Also called *data delimiter.*

septate cavity A coaxial cavity containing a SEPTUM between the inner and outer conductors.

septate waveguide A waveguide containing one or more septa (see SEPTUM) to control power transmission.

septum A thin metal vane used as a reflector in a waveguide or cavity.

sequence **1.** A succession of objects, parameters, or numbers. **2.** An ordered set of numbers—each of which is related to its predecessor by a specific mathematical function.

sequence checking routine In computer operations, a routine that verifies the order of items of data.

sequence control register In a computer memory, a register whose contents determine the instruction to be implemented next.

sequence programmer A timing device that can be preset to start or stop various operations at predetermined times.

sequencer A device that initiates or terminates events in a desired sequence.

sequence relay A relay whose several contacts close in a predetermined order.

sequence timer A timer in which separate delay circuits are actuated in a predetermined sequence.

sequential In computer operations, a term denoting operations on data items in which the items (e.g., records in a file) are taken in an order determined by key values, rather than in the order in which the items are physically arranged.

sequential access memory Any semiconductor memory in which data can be recalled or addressed only in a certain specified order.

sequential analysis In statistics, using an unspecified number of observations as samples from which a result is derived. Each observation is accepted or rejected, or another observation is made.

sequential color television The successive transmission of the three primary colors in a television system, and their reproduction at the receiver in the same order. Also see DOT-SEQUENTIAL SYSTEM, FIELD-SEQUENTIAL SYSTEM, and LINE-SEQUENTIAL SYSTEM.

sequential control Computer operation in which the order of instruction implementation is the same as the order of instruction storage.

sequential relay See SEQUENCE RELAY.

sequential scanning Rectilinear television scanning in which the center-to-center distance between successive lines is the same as the nominal line width.

sequential switch **1.** A switch that provides selection of two or more ports in a rotating succession. **2.** In a television system, a switch that allows the monitoring technician to select any of the cameras for viewing.

sequential timer See SEQUENCE TIMER.

ser **1.** Abbreviation of SERIES. **2.** Abbreviation of SERIAL.

serial **1.** Pertaining to the performance of steps, or the occurrence of elements (such as data items on magnetic tape), in succession. **2.** An order, row, or sequence in which one item follows another (as opposed to parallel).

serial access Access to data file records in their order in a storage medium.

serial adder See SERIAL ARITHMETIC UNIT.

serial arithmetic unit In computer operations, an arithmetic unit in which digits are handled in order. Compare PARALLEL ADDER.

serial bit Data in which the bits of each byte or word are sent or received one at a time.

serial memory A register in which the input and output data is stored and retrieved one bit at a time.

serial-parallel **1.** Pertaining to data transfer that is serial in one sense and parallel in another sense. For example, entire words might be serially transmitted within a system, but the constituent bits of each word might be transferred in parallel. **2.** See PARALLEL-SERIES.

serial processing In computer operations, the sequential processing of several different programs through a single channel. Compare PARALLEL PROCESSING.

serial storage In computer operation, storage in which elements are entered in order and are available only in the same order. Compare PARALLEL STORAGE.

serial transfer The propagation of information along a single path, in which data bits are sent one after the other.

series **1.** The sum of a mathematical sequence (see SEQUENCE, **2**). **2.** Pertaining to the connection of elements or components end-to-end (see SERIES CIRCUIT).

series addition See SERIES-AIDING.

series-aiding The condition in which two series voltages or magnetic fields are added together. Compare SERIES-BUCKING.

series antenna tuning Antenna-feeder tuning in which a separate tuning capacitor is connected in series with each wire. Compare PARALLEL ANTENNA TUNING.

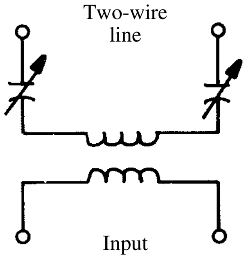

series antenna tuning

series bucking The condition in which two series voltages or magnetic fields oppose each other. Compare SERIES-AIDING.

series capacitance Capacitance acting, or connected, in series with another capacitance or other quantity.

series capacitors Capacitors connected in series. If the individual capacitors have values $C1$, $C2$, . . . , Cn, then the total capacitance Ct is equal to $1/(1/C1 + 1/C2 + \dots + 1/Cn)$. Also see SERIES CIRCUIT.

series circuit A circuit whose components are, in effect, connected in a string (i.e., end-to-end). Compare PARALLEL CIRCUIT.

series-diode half-wave rectifier See SERIES-DIODE RECTIFIER.

series-diode rectifier A rectifier circuit in which the diode is connected in series with the source and load. Compare SHUNT-DIODE RECTIFIER.

series dropping resistor See DROPPING RESISTOR.

series equivalent impedance A series impedance that will draw the same current (magnitude and phase) drawn by a given parallel circuit connected across the same single-phase source.

series equivalent of parallel circuit See SERIES EQUIVALENT IMPEDANCE.

series-fed amplifier An amplifier circuit in which the operating voltages are applied in series with the alternating-current signal voltages. Also see SERIES FEED.

series-fed oscillator An oscillator circuit in which the direct-current operating voltage is applied in series with the alternating-current output voltage. Also see SERIES FEED.

series feed The application of alternating-current (ac) and direct-current (dc) voltages in series to a device. Example: the presentation of the dc operating voltages for an amplifier in series with the ac signal voltages (see SERIES-FED AMPLIFIER).

series feedback A feedback system in which the feedback signal is presented to the input point in series with the input signal. Compare SHUNT FEEDBACK.

series field A magnetic field produced by a series winding in a motor or generator.

series generator An electric generator in which the armature and field windings are connected in series. Compare SHUNT GENERATOR.

series inductance **1.** Inductance acting, effectively, in series with some other quantity (e.g., the inherent inductance of a wirewound resistor). **2.** An inductance connected in series with other inductances, or with some other quantity (e.g., capacitance and resistance).

series inductors Inductors connected in series, and separated or oriented in a way that minimizes the effects of mutual inductance. Assuming zero mutual inductance, if the individual inductors have values $L1$, $L2$, . . . , Ln, then the total inductance Lt is equal to $L1 + L2 + \ldots + Ln$. Also see SERIES CIRCUIT.

series limiter A limiter (clipper) circuit in which the diode is essentially in series with the signal. Compare PARALLEL LIMITER.

series loading The series insertion of reactances in a circuit for the purpose of impedance matching.

series magnetic circuits A combination of several magnetic paths in line so that flux extends through each path in sequence; this is analogous to the passage of electric current successively through series-connected resistors.

series motor An electric motor whose armature and field windings are connected in series. Compare SHUNT MOTOR.

series operation The operation of units in succession, necessitating sequential current flow through each. Also see SERIES CIRCUIT.

series opposition See SERIES BUCKING.

series-parallel See PARALLEL-SERIES.

series-parallel capacitors See PARALLEL-SERIES CAPACITORS.

series-parallel inductors See PARALLEL-SERIES INDUCTORS.

series-parallel resistors See PARALLEL-SERIES RESISTORS.

series regulator A voltage regulator circuit in which the controlled device is in series with the load. Compare SHUNT REGULATOR.

series resistance **1.** Resistance acting in series with another resistance or with another quantity (e.g., capacitance and inductance). **2.** The inherent resistance that acts effectively in series with the plates of a capacitor. **3.** The resistance of the wire in a coil, acting effectively in series with the inductance.

series resistors Resistors connected in series with each other. If the individual resistors have values $R1$, $R2$, . . . , Rn, then the total resistance Rt is equal to $R1 + R2 + \ldots + Rn$. Also see SERIES CIRCUIT.

series resonance Resonance in a circuit consisting of a capacitor, inductor, and an alternating-current generator in series. At the resonant frequency, the inductive reactance and the capacitive reactance cancel, so the net reactance is zero. The capacitor current and inductor current are maximum and equal, and the circuit impedance is minimum. Compare PARALLEL RESONANCE.

series-resonant circuit A resonant circuit in which the capacitor, inductor, and generator are connected in series. Also see SERIES RESONANCE. Compare PARALLEL-RESONANT CIRCUIT.

series-resonant trap A wavetrap consisting of a series-resonant inductance-capacitance (LC) circuit. Compare PARALLEL-RESONANT TRAP.

series-resonant wavetrap See SERIES-RESONANT TRAP.

series-shunt circuit See PARALLEL-SERIES.

series tee junction See E-PLANE TEE JUNCTION.

series tracking capacitor See OSCILLATOR PADDER.

series-type frequency multiplier A varactor frequency-multiplier circuit in which the varactor is in series with the input and output. Compare SHUNT-TYPE FREQUENCY MULTIPLIER.

series-type resonance bridge A resonance bridge in which the impedance arm is a series-resonant circuit. Compare SHUNT-TYPE RESONANCE BRIDGE.

series winding **1.** In a motor or generator, a winding connected in series with the armature. **2.** A method of motor or generator construction in which the field winding is connected in series with the armature.

series-wound generator See SERIES GENERATOR.

series-wound motor See SERIES MOTOR.

serrated pulse A pulse having a notched or slotted top. An example is the vertical sync pulse in television.

serrated rotor plate In a variable capacitor, an external rotor plate that is slotted radially. This alters the capacitance-variation curve of the capacitor to allow alignment of sensitive apparatus (e.g., the tracking of radio-frequency tuned circuits in a radio receiver).

serrated vertical sync pulse In television, the vertical sync pulse notched at twice the horizontal sweep frequency.

service **1.** To maintain or repair electronic equipment. **2.** To provide maintenance or repair of electronic equipment.

serviceability ratio For a device or system, the ratio *ts*/(*ts* + *td*), where *ts* is serviceable (operational) time, and *td* is downtime (non-operational time).

serviceable time The cumulative time during which an operator-monitored (but not necessarily operated) device or system is capable of normal operation.

service area For a broadcast or communications station, the useful coverage area.

service band **1.** For a communications system, the band of frequencies in which operation is normally carried out. **2.** A band of frequencies specifically assigned, by government regulation, to a certain communications service or services.

service channel The band of frequencies that a particular broadcast or communications station occupies, when the carrier frequency is held constant.

service charge The amount charged by a technician for installation, maintenance, or repair of equipment. It is often performed on a per-hour basis.

service maintenance For a cell or battery, the relative amount of energy capacity (percentage of full-charge capacity) available at a given time, or after a certain length of time in normal use.

service meter **1.** An energy ("power") meter. Also see KILOWATT-HOUR METER. **2.** A rugged multimeter used by a service technician.

service oscillator A signal generator designed expressly for troubleshooting and repair service.

service switch **1.** The main switch controlling the electric service to a building or other place of installation. **2.** In television repair, a switch on the rear of a chassis. The switch facilitates adjustment of screen controls by removing vertical deflection temporarily.

service-type instrument An instrument having reasonable accuracy and a degree of ruggedness so that it is suitable for field or shop use. Examples: SERVICE METER and SERVICE OSCILLATOR. Compare LABORATORY-GRADE INSTRUMENT.

servo amplifier A highly stable amplifier designed expressly for use in a SERVOMECHANISM.

servo loop In a control system (particularly a servo amplifier), the output-to-input feedback loop, through which automatic control is performed.

servomechanism Also called *servo.* A self-correcting, closed-loop control system. It usually uses an electromechanical device, such as a motor, that controls some electronic device. Error signals, supplied by the controlled electronic device, cause the motor to run in such a way as to optimize or stabilize the system.

servomotor A motor operated by the output signal of a servo amplifier. Depending on the end application of the servo system, the motor signal might or might not be corrected.

servo oscillation In a servo system, a back-and-forth movement or fluctuation, relative to the optimum setting or position. It Results from improper system adjustment. Sometimes the system stabilizes at the optimum after a short period of oscillation; in some cases, the oscillation continues indefinitely.

servo robot A (usually industrial) robot whose motion sequence is programmed into a computer. The robot follows the instructions given by the computer, and makes precise, timed movements on that basis. Different computer programs allow different motion sequences, so a single robot can be used for various tasks.

servo system An automatic control system using one or more servomechanisms.

set **1.** A piece of equipment or a system (e.g., *radio set*). **2.** In a flip-flop circuit, an input that is not controlled by the clock. **3.** To adjust a circuit or device, such as a flip-flop, to a desired operating point or condition. **4.** A class of numbers, things, or events. **5.** In computer programming, to initialize a variable (i.e., to assign a label to a location).

set analyzer A combination test instrument designed originally for troubleshooting radio receivers. It consists of a multimeter and transistor tester or vacuum-tube tester.

set noise Electrical noise arising inside a radio or television receiver, as opposed to that picked up from the outside.

set pulse A pulse used to adjust a device to a certain state (see SET, **3**).

set terminal In a flip-flop, the one-input terminal. Compare RESET TERMINAL.

setting The position or value to which an adjustable device is set for a particular purpose.

settling time **1.** In a digital voltmeter, the time required between the application of a test voltage and the final display of an accurate readout. **2.** In a digital-to-analog converter, the time between half of the level change over all inputs and the arrival of the output to a level within a certain tolerance of its specified final level. It is defined for either full-scale to zero or zero to full-scale.

set up To arrange and prepare equipment for operation.

setup See SET, **1**.

set-up time **1.** The time required to install and test an electronic system, and to ready the system for operation. **2.** In a digital gate, the length of time that a pulse must be held to produce a change of state.

sexadecimal number system See HEXADECIMAL NUMBER SYSTEM.

sexagesimal number system A number system whose radix is 60.

SF **1.** Abbreviation of SAFETY FACTOR. **2.** Abbreviation of SINGLE FREQUENCY. **3.** Abbreviation of STANDARD FREQUENCY. **4.** Abbreviation of STABILITY FACTOR.

SFA Abbreviation of SINGLE-FREQUENCY AMPLIFIER.

S_F band A section of the S BAND extending from 1650 to 1850 MHz.

sferics In wireless broadcast or communications reception, random electromagnetic noise generated by the earth's atmosphere. Some of this noise is thermal in origin; some originates in thunderstorms. Sometimes this noise is called *static.*

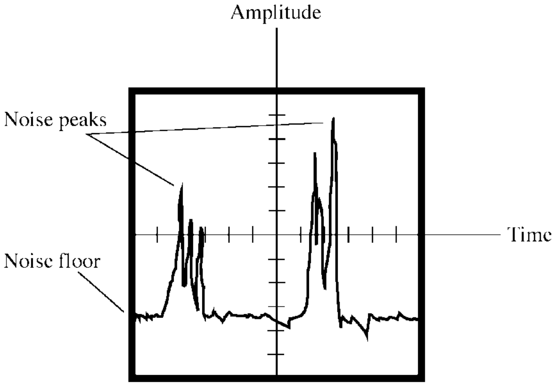

sferics

SFO **1.** Abbreviation of SINGLE-FREQUENCY OSCILLATOR. **2.** Abbreviation of STANDARD-FREQUENCY OSCILLATOR.

SFR Abbreviation of SINGLE-FREQUENCY RECEIVER.

SFR-Chireix-Mesny antenna See CHIREIX-MESNY ANTENNA.

SG Abbreviation of SCREEN GRID.

S_G band A section of the S BAND extending from 2700 to 2900 MHz.

SGCS Abbreviation of *silicon gate-controlled switch* (see SILICON-CONTROLLED SWITCH).

shaded-pole motor An induction-type alternating-current motor using shading coils on the field poles for self-starting with a single-phase supply.

shading Electronic enhancement of a television picture, resulting in a different brightness over various portions of the background, as compared with the actual situation. It can be used, for example, to make certain subjects stand out from the background.

shading coil **1.** A single, short-circuited turn (copper ring) encircling the tip of the core of a coil that carries alternating current (ac), such as the field pole of a motor. Current induced in the coil causes a momentary flux shift that approximates a rotating field that self-starts a simple single-phase induction motor. **2.** A coil used in a simple ac relay to prevent chatter.

shading ring See SHADING COIL.

shading signal In a television camera, a signal that increases the gain of the amplifier while the beam scans a dark part of the image.

shadow area A region in which signal attenuation or the absence of a signal results from the SHADOW EFFECT.

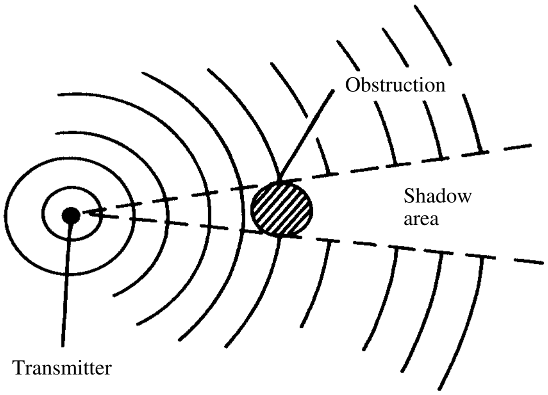

shadow area

shadow attenuation **1.** The attenuation of electromagnetic energy caused by an obstacle. It is generally measured in decibels. **2.** The attenuation of electromagnetic energy produced by the curvature of the earth.

shadow effect The obstruction of radio waves by objects in their path.

shadow mask See APERTURE MASK.

shadow meter See SHADOW TUNING INDICATOR.

shadow region See SHADOW AREA.

shadow tuning indicator A tuning meter in which the indicating medium is a shadow whose width is proportional to meter current.

shaft The rodlike part of an adjustable component (such as a potentiometer or variable capacitor) to which a rotating (or turning) member is attached.

shaft-angle encoder An electronic system for converting shaft rotation into direct binary or decimal readings.

shaft lock A device for fastening the shaft of an adjustable component (such as a potentiometer, rotary switch, or variable capacitor) in position at a particular setting.

shaft-position encoder See SHAFT-ANGLE ENCODER.

shaft-position indicator A device that delivers an analog or digital output signal proportional to the arc of rotation of a shaft.

shaker See VIBRATOR, **2**.

shake table A platform, actuated by a vibrator, on which components can be mounted for a vibration test.

shallow-diffused junction A pn junction made by diffusing the impurity material for a short distance into the semiconductor wafer. Compare DEEP-DIFFUSED JUNCTION.

shape factor **1.** For a tuned circuit, the ratio of the 60-dB bandwidth to the 6-dB bandwidth. **2.** For a filter, the ratio of bandwidth at high attenuation to that at low attenuation.

shaping network A combination of components for changing the natural response of a circuit to a desired response (i.e., a curve-changing circuit).

shared file A data file that is available for use by more than one system simultaneously.

shared files system A data-processing system having one direct-access storage device from which information can be accessed by more than one computer.

sharpener **1.** A circuit or device for increasing the selectivity of another circuit or device. **2.** A circuit or device for decreasing the rise or fall time of a pulse or square wave. **3.** A circuit or device for steepening the response of a filter.

sharpness See SELECTIVITY.

sharp pulse A pulse having extremely fast rise and fall times and narrow width (i.e., a spike).

shaving The physical modification of a phonograph disc, or other permanent recording surface, in preparation for rerecording.

S_H band A section of the S BAND extending from 3700 to 3900 MHz.

sheath See POSITIVE-ION SHEATH.

shelf corrosion In a dry cell in storage, deterioration of the negative electrode because of local action in the zinc.

shelf life **1.** The longest period of time that electronic equipment can be continuously kept in storage before deterioration of materials or degradation of performance occurs. **2.** The longest period of time that a battery can be stored without use before it must be discarded or recharged.

shell **1.** An electronic orbit (imaginary shell) in an atom. **2.** The envelope of a component (e.g., the outer casing of a power transistor or the housing of a plug). **3.** The rigid case in an audio or video tape cassette.

shell-type choke See SHELL-TYPE INDUCTOR.

shell-type core A core that completely surrounds the coil(s) of a choke or transformer.

shell-type inductor An inductor in which the core completely surrounds the coil.

shell-type transformer A transformer in which the core completely surrounds the coils.

shf Abbreviation of SUPERHIGH FREQUENCY.

shield A metallic partition or box for confining an electric or magnetic field.

shield baffle A sheet-type shield. Also see BAFFLE, **2** and SHIELD PARTITION.

shield box A shield having a general box shape, and which is usually enclosed on all sides.

shield braid Tubing woven from wire, through which an insulated wire is passed and thus shielded.

shield can A cylindrical shield, usually enclosed on all sides.

shield disk A flat shield having a disk shape. Also see BAFFLE, **2**; SHIELD BAFFLE; and SHIELD PARTITION.

shielded cable Cable completely enclosed within a metal sheath that is either flexible or rigid.

shielded wire A single strand of insulated wire completely enclosed in a flexible or rigid shield.

shield partition A wall-type shield usually consisting of a single, flat sheet of metal, sometimes bent into an angle. Also called BAFFLE SHIELD (see BAFFLE, **2**).

shield plate See BAFFLE, **2**; SHIELD BAFFLE; SHIELD DISK; and SHIELD PARTITION.

shield room See CAGE.

shield wire A (usually grounded) wire, near and parallel to another wire that it shields.

shift **1.** To move from one operating point to another in a characteristic curve, or in the operation of an equipment. **2.** To transfer data from one point to another in a system, or move it left or right in a register.

shift flip-flop circuit A flip-flop designed especially as a stage in a shift register.

shift pulse In a shift register, a drive pulse that initiates the shifting of characters.

shift register In computers, calculators, and storage systems, a circuit (usually composed of flip-flops in cascade) in which pulses can be shifted from stage to stage and finally out of the circuit.

shingle-type photocell A device in which several separate photocells are series connected by slightly overlapping the ends of adjacent cells.

ship station A radio or radar station installed aboard a ship that is not in port.

ship-to-shore communication Radio communication between a ship at sea and a land-based station.

shock **1.** See ELECTRIC SHOCK. **2.** A signal applied momentarily to a circuit, as in the *shock excitation* of a tank. **3.** A sudden, dramatic change in an environmental variable (such as temperature). **4.** Physical blows or vibration.

shock absorber Any object or device intended for reducing physical vibration of a component, set of components, circuit, or system.

shock device **1.** A device for administering shock therapy (see ELECTROSHOCK, **1**). **2.** An induction coil and associated primary supply for applying high voltage to a wire fence.

shock excitation Driving an inductance-capacitance (LC) tuned circuit into damped oscillation by momentarily applying a pulse.

shock-excited oscillator A type of self-excited oscillator in which the transistor is suddenly cut off

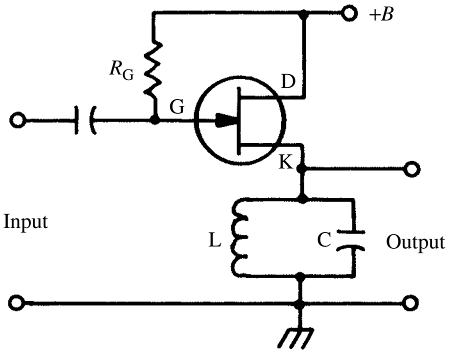

shock-excited oscillator

by applying a cutoff voltage to the gate or base electrode. This abrupt interruption of steady drain or collector current shocks the tank into damped oscillations.

shock hazard **1.** Any situation that presents the danger of electric shock to attendant personnel. **2.** The existence of a potential difference that will cause a current of at least 5 mA to flow through a resistance of 500 ohms or more, for a prolonged period of time.

Shockley diode See FOUR-LAYER DIODE.

shock mount A structure that secures a microphone while minimizing the pickup of vibrations through the table, floor, or other surface on which the microphone is placed.

shock therapy See ELECTROSHOCK, **1**.

shoran Contraction of SHORT-RANGE NAVIGATION.

shore effect The tendency of radio waves traveling along a shore to be bent either toward or away from the shore. It can occur because of differences in surface conductivity and/or atmospheric temperature over land, as compared with water.

shore station A fixed, land-based radio station that communicates with ships at sea.

shore-to-ship communication See SHIP-TO-SHORE COMMUNICATION.

short circuit An often unintended low-resistance path through which current bypasses a component or circuit.

short-circuit current In a power supply, the current that flows when the output is directly shorted. Many power supplies have shutdown devices that cause the current to stop flowing when the output terminals are short-circuited; other supplies effectively insert resistance in series with the load, if necessary, to limit the current.

short-circuiting switch A rotary selector switch in which unused contacts are automatically short-circuited.

short-circuit parameter A parameter for which zero resistance is assumed in the part of the circuit under consideration. The current amplification factor (alpha) of a common-base-connected transistor is such a parameter because its collector load resistance is assumed to be zero.

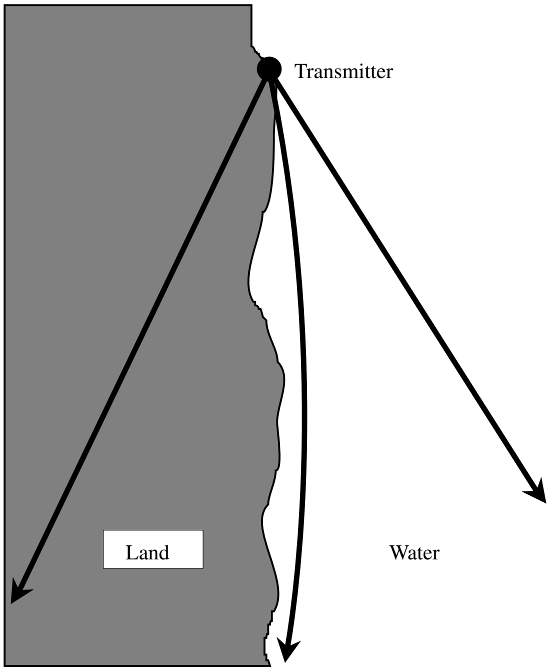

shore effect

shorted-stub tuning Tuning a stub to match a feeder to an antenna by sliding a short-circuiting bar along the two wires of the feeder.

shorting bar A thick, metal strap for short-circuiting two binding posts.

shorting link A sheet-metal strip for connecting together two binding posts.

shorting loop In a telephone system, a device that short-circuits two specified points for the purpose of testing or line fault location.

shorting stick A metal rod with an insulating handle, used to short-circuit a charged capacitor to remove the shock hazard.

shorting switch See SHORT-CIRCUITING SWITCH.

short-line tuning Use of a parallel capacitance to tune a transmission line that is less than a quarter-wave long.

short-range navigation Contraction, shoran. Navigation by means of SHORT-RANGE RADAR.

short-range radar A radar having a 50- to 150-mile maximum line-of-sight range for a 1-square-meter reflecting target that is perpendicular to the radar beam.

short skip Skip of only a few hundred miles range. Also see SHORT-SKIP COMMUNICATION.

short-skip communication Radio communication via the ionosphere over relatively short distances (400 to 1300 miles). See, for example, SPORADIC-E SKIP.

short-term drift The gradual change in the value of a quantity, such as frequency or voltage, observed over a comparatively brief interval, as opposed to change occurring over a long period. Compare LONG-TERM DRIFT.

short-term effect The variation of any electrical parameter over a relatively brief time interval. Example: frequency drift over a short time period. Also called *short-time effect.*

short-term stability Stability reckoned over a comparatively brief time interval, as opposed to stability for a long period. Compare LONG-TERM STABILITY.

short-time effect See SHORT-TERM EFFECT.

shortwave **1.** Pertaining to wavelengths shorter than 200 meters (i.e., frequencies higher than 1.50 MHz). **2.** Pertaining to the frequencies above the standard amplitude-modulation broadcast band (above 1.605 MHz), but below 30 MHz.

shortwave converter A superheterodyne converter for adapting a longwave receiver (such as a broadcast receiver) for shortwave reception.

shortwave listener Abbreviation, SWL. A radio hobbyist who receives, but does not transmit, shortwave signals.

shortwave receiver Any radio receiver capable of intercepting and demodulating signals in the range 1.705 MHz to 30 MHz.

shortwave transmitter Any radio transmitter capable of producing energy in the range 1.705 MHz to 30 MHz.

shot-effect noise Electrical noise caused by random fluctuations in a current, as in a diode or transistor. Also see EQUIVALENT NOISE RESISTANCE. Compare THERMAL NOISE.

shotgun microphone A highly directional microphone sensitive only to sounds coming from a specific direction; the response pattern has a narrow main lobe. It name results from its long, cylindrical configuration.

shot noise Electrical noise arising from intermittent impulses, such as those produced by spark discharges, make-and-break contacts, etc. Its name results from its resemblance to pistol shots.

shrink The amount by which a material being measured with an electronic instrument decreases in surface dimension. Compare STRETCH.

shrink tubing Plastic sleeving placed over a conductor or at a conductor/connector joint, and made to shrink tightly with the application of heat.

shunt Synonym, *parallel.* **1.** Pertaining to the connection of one component across (in parallel with) another (e.g., *shunt resistor*). **2.** Pertaining to connection of components in such a manner that they each (or all) are subjected to identical voltages. **3.** A deliberately produced short circuit between two specific points in a device or system. **4.** To deliberately bypass some part of a system by means of a short circuit.

shunt circuit See PARALLEL CIRCUIT.

shunt-diode rectifier A rectifier circuit in which the diode is connected in parallel with the source and load. Compare SERIES-DIODE RECTIFIER.

shunt-fed **1.** Pertaining to a circuit or device in which the direct-current operating voltage and alternating-current signal voltage are applied in parallel to an electrode. **2.** Pertaining to a base-grounded vertical antenna excited at some point above ground.

shunt feed See PARALLEL FEED.

shunt feedback A feedback system in which the fed-back signal is presented to the input of the network in parallel with the input signal. Compare SERIES FEEDBACK.

shunt generator An electric generator in which the armature and field windings are connected in parallel. Compare SERIES GENERATOR.

shunting effect The condition in which a quantity, such as stray capacitance or resistance, acts in parallel with another quantity. Example: the shunting (parallel) resistance of an electrolytic capacitor.

shunt leads Interconnecting wires used for the purpose of attaching a shunting component to a test instrument.

shunt limiter See PARALLEL LIMITER.

shunt loading The parallel insertion of reactance in a circuit, for the purpose of impedance matching.

shunt motor An electric motor whose armature and field windings are connected in parallel. Compare SERIES MOTOR.

shunt regulator A voltage-regulator circuit in which the controlled transistor or vacuum tube is in parallel with the output (load) terminals. Compare SERIES REGULATOR.

shunt resistor **1.** A resistor connected in parallel with a meter or recorder to increase its current range. **2.** A resistor connected in parallel with a voltmeter to convert it into a current meter. Compare MULTIPLIER RESISTOR.

shunt-series circuit See PARALLEL-SERIES CIRCUIT.

shunt tee junction A waveguide H-PLANE TEE JUNCTION.

shunt-type frequency multiplier A varactor frequency multiplier circuit in which the varactor is in parallel with the input and output. Compare SERIES-TYPE FREQUENCY MULTIPLIER.

shunt-type resonance bridge A resonance bridge in which the impedance arm is a parallel-resonant circuit. Compare SERIES-TYPE RESONANCE BRIDGE.

shunt-wound generator See SHUNT GENERATOR.

shunt-wound motor See SHUNT MOTOR.

SI Abbreviation of (Standard) INTERNATIONAL SYSTEM OF UNITS.

S/I Abbreviation of *signal-to-intermodulation ratio.*

Si Symbol for SILICON.

sibilants **1.** High-frequency (hissing) components of speech. **2.** High-frequency sounds or audio signals.

SIC Abbreviation of *specific inductive capacity* (see DIELECTRIC CONSTANT).

SiC Formula for SILICON CARBIDE.

sideband **1.** With respect to a carrier, one of the additional frequencies generated by the modulation process. In simple amplitude modulation, the two sidebands are *fc* + *fm* and *fc* – *fm*, where *fc* is the carrier frequency, and *fm* is the modulation frequency. **2.** Pertaining to sidebands as defined in **1**.

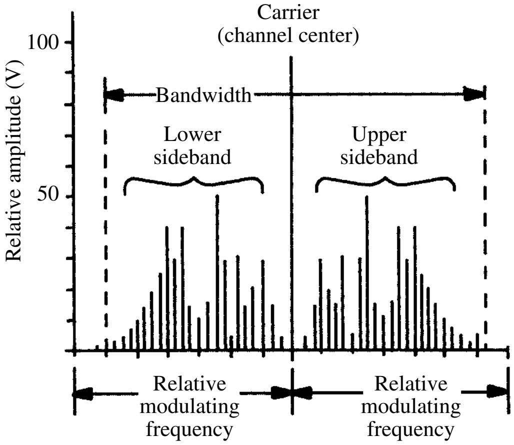

sideband

sideband attenuation See SIDEBAND CUTTING.

sideband cutting Elimination or attenuation of the sidebands of a modulated signal by a circuit having insufficient bandwidth.

sideband frequency The frequency of the modulation-generated signal accompanying a carrier. One sideband frequency is that of the carrier minus that of the modulating signal; another is the sum of the carrier and the modulation frequency. See also SIDEBAND, **1**.

sideband interference **1.** Interference arising from one or both of the normal sidebands of a modulated signal. **2.** Interference caused by spurious sidebands, resulting from overmodulation.

sideband power The power contained in the sideband(s) of a signal.

sideband slicing See SIDEBAND CUTTING.

sideband splatter In an amplitude-modulated or single-sideband signal, the emission of sideband energy at frequencies other than within the designated channel. Also simply called splatter.

sideband technique A method of using, for communications or other purposes, one or both of the sidebands of a modulated signal without the carrier.

side-chain amplifier An auxiliary amplifier that is external to a main amplifier. Such an amplifier might be used, for example, in a feedback channel or in a volume-compression or volume-expansion channel.

side frequency See SIDEBAND.

side lobe In certain directional antenna systems, a minor lobe in the horizontal-plane directivity pattern that appears at right angles, or nearly at right angles, to the main lobe. Such a lobe represents reduced sensitivity and/or power gain relative to the main lobe. Also see MAIN LOBE, MINOR LOBE.

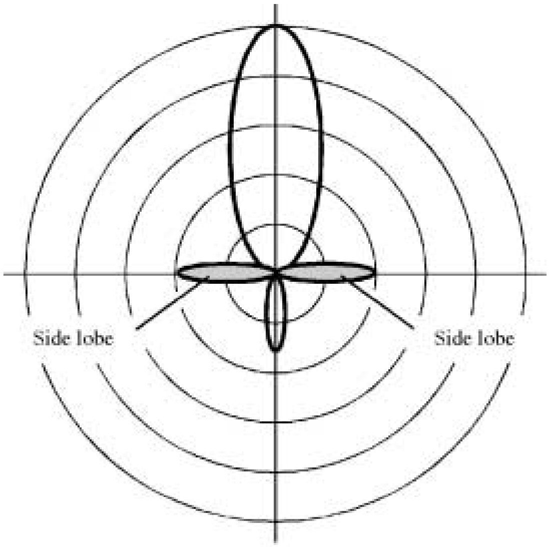

side lobe

sidelobe suppression Elimination of the sidelobe(s) from the radiation pattern of an antenna.

sidestacked antennas Antennas mounted in a horizontal line, parallel to each other, and connected by a common coupler to a transmitter or receiver.

sideswiper A manual telegraph key operated by moving the lever sideways, rather than up and down.

sidetone **1.** In wire telephony, the reproduction by the receiver of sounds picked up by the transmitter of the same telephone. **2.** In radiotelegraphy, an audible tone actuated when the carrier is transmitted. It allows the sending operator to hear Morse code elements as they are sent.

sidetone telephone A telephone set with no provision for canceling the sidetone.

siemens Symbol, S. The SI unit of conductance. The conductance of a component or medium in siemens is equal to the reciprocal of the resistance in ohms.

Siemen's electrodynamometer A spring-tension meter that operates by means of torque, with zero current through the device representing zero torque. It can be used for measurements of current, voltage, or power.

Sierra Phonetic alphabet code word for the letter S.

sig Abbreviation of SIGNAL.

sign **1.** Any indicator denoting whether a value is positive or negative. **2.** A graphic device indicating an operation. Examples: + (addition), × (multiplication). **3.** Any symbol. An ampersand, for example, is an "and" sign. **4.** A characteristic symptom of malfunction or improper operation (e.g., a high standing-wave ratio in an antenna system is a sign of an impedance mismatch).

signal An electrical quantity, such as a current or voltage, that can be used to convey information for communication, control, calculation, etc.

signal-actuated voice recorder Abbreviation, SAVOR. A recorder that goes into operation automatically when the speaker starts talking and stops when the speaker finishes.

signal amplitude The intensity of a signal quantity (see SIGNAL).

signal booster See PREAMPLIFIER.

signal channel In a system, a channel through which only signals flow, control and modifying impulses being accommodated by other channels.

signal circuit A circuit handling signal currents and voltages to the exclusion of control and operating currents and voltages.

signal conditioner Any accessory device (such as a peak probe, demodulator probe, current shunt, etc.) used to modify or change the function of a basic instrument (such as an electronic voltmeter).

signal converter See CONVERTER, **1**.

signal current The current component of a signal, as opposed to operating current in a system.

signal diode A diode designed primarily for light-duty signal applications (detection, demodulation, modulation, curve changing), as opposed to the heavy-duty applications of power diodes and rectifiers.

signal distance In two words (bit groups) of the same length, the number of corresponding bit positions whose states differ. For example, the signal distance between 01001 and 10011 is 3.

signal-flow analysis A graphic method of analyzing circuits, particularly those using feedback, through the use of diagrams in which straight arrows represent transmission paths, dots represent nodes, and curved arrows represent feedback paths.

signal-flow diagram The transmission-path diagram used in SIGNAL-FLOW ANALYSIS.

signal gain The gain of an amplifier circuit—especially if used in small-signal applications. See also AMPLIFICATION, and GAIN.

signal generator An instrument that produces signals of precise frequency and amplitude, usually over a wide range.

signal ground **1.** Any circuit point that remains at zero signal potential. **2.** A connection to a point that is deliberately maintained at zero signal potential.

signal/image ratio See SIGNAL-TO-IMAGE RATIO.

signaling In a communications system, the exchange of data in electrical form, either analog or digital.

signaling rate In data communications, the speed at which data is transmitted. It is commonly expressed in bits per second (bps). Also, it is sometimes expressed in baud or in words per minute (wpm).

signaling time slot In a communications signal, a specified interval of time, starting at a certain instant in each signal frame. This interval is used exclusively for the purpose of signaling.

signal injection **1.** The introduction of a signal into a circuit. **2.** A method of troubleshooting in communications receivers. A signal generator is used to introduce a test signal into each stage, starting with the output and proceeding stage-by-stage toward the input, until the defective stage or component is located.

signal injector A simple (usually single-frequency) signal generator used in troubleshooting to introduce a test signal at selected points in a circuit, to locate malfunctioning stages or components. Also see SIGNAL INJECTION, **2**.

signal intensity See SIGNAL STRENGTH.

signal inversion Phase reversal of a signal passing through a circuit, device or medium.

signal level At a given point in a circuit, the strength of a signal, with respect to a reference amplitude.

signal loss **1.** A reduction in the amplitude of a signal as it passes through a system. **2.** The complete disappearance of a signal. **3.** See FRACTIONAL GAIN.

signal mixer See MIXER.

signal/noise ratio See SIGNAL-TO-NOISE RATIO.

signal notcher See NOTCH FILTER.

signal peaker See PEAK FILTER.

signal power The amplitude of a signal expressed in watts, milliwatts, or microwatts, as opposed to amplitude expressed as a current or voltage.

signal processor Any device, (e.g., preamplifier, expander, amplitude limiter, delay network) inserted into or added onto a system to modify an input or output signal.

signal rectification The conversion of an alternating-current signal into a proportionate direct-current signal, usually by means of a diode circuit.

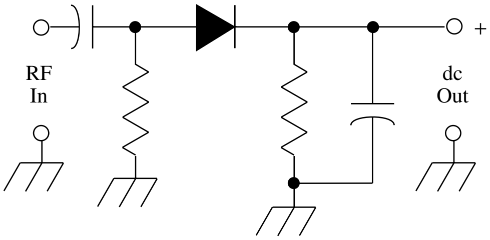

signal rectification

signal rectifier See SIGNAL DIODE.

signal regeneration See SIGNAL RESHAPING.

signal reshaping **1.** The processing of a signal so that it acquires its original waveform. Also called *signal regeneration.* **2.** Passing a digital signal of any type through a circuit that delivers a uniform output pulse on a real-time one-to-one basis.

signal shifter **1.** A device used for quickly changing the frequency of a transmitted signal. **2.** A device that automatically causes a transmitted signal to be sent on a frequency that differs from the receiver frequency by a known and predetermined amount. **3.** See MIXER, **1**. **4.** See CONVERTER, **1**.

signal squirter See SIGNAL INJECTOR.

signal strength The amplitude of a signal, usually in terms of voltage. Current or power is specified in some applications.

signal-strength meter **1.** See FIELD-STRENGTH METER. **2.** See S-METER.

signal synthesizer A special signal generator delivering signals whose frequency, amplitude, and waveshape can be adjusted at will.

signal time delay The time required for an element of a signal to be transmitted through a circuit or network. This delay results in phase shift in an amplifier.

signal-to-distortion ratio In a receiver, the ratio of the desired signal to the level of distortion other than the specified signal. Usually expressed in decibels (dB).

signal-to-image ratio Abbreviation, S/I. In a receiver, the ratio of signal amplitude to image amplitude, both being measured in the same units. It is usually expressed in decibels (dB).

signal-to-noise-plus-noise ratio Abbreviation, (S+N)/N. In a receiver, the ratio of the combined signal and noise amplitude to the amplitude of the noise alone. It is usually expressed in decibels (dB). Peak voltages are used to determine this ratio in the case of pulse noise; root-mean-square (rms) voltages are used in the case of random noise.

signal-to-noise ratio Abbreviations: S/N, SNR. In a receiver, the ratio of signal amplitude to noise amplitude. It is usually expressed in decibels (dB). Peak voltages are used to determine this ratio in the case of pulse noise; root-mean-square (rms) voltages are used in the case of random noise.

signal-to-noise-and-distortion ratio Abbreviation, SINAD. In a receiver, the ratio of the desired signal to the level of noise and distortion, other than the specified signal. It is usually expressed in decibels (dB).

signal tracer A tuned or untuned detector/amplifier having an input probe and an output indicator (meter, loudspeaker, or both), for following a test signal through a circuit.

signal voltage The voltage component of a signal, as opposed to the operating voltage of the circuit generating or passing the signal.

signal wave **1.** Any electromagnetic disturbance of a periodic nature that is modulated to convey information. **2.** The visual illustration or rendition of an electromagnetic disturbance that is modulated to convey information.

signal winding In a magnetic amplifier or saturable reactor, the coil that receives the control current.

signal window See WINDOW, **2**.

signal wobbulator A frequency modulator used with an unmodulated signal generator to provide sweep signals for visual alignment. Also see WOBBULATOR.

sign bit A one-bit SIGN DIGIT.

sign digit A character indicating the sign (positive or negative) of the value of the field or word to which it is attached (usually at the end).

signed field In a computer record, a field having a number whose sign is indicated by a SIGN DIGIT.

significant digits See SIGNIFICANT FIGURES.

significant figures In a numerical quantity, especially one expressed in scientific (power of 10) notation, those figures (digits) that depict a quantity to a required, relevant, or justifiable degree of precision. For example, 173,201 expressed to three significant figures is 1.73×10^5; 0.00477583 expressed to four significant figures is 4.776×10^{-3}. See also SCIENTIFIC NOTATION.

silencer See AUTOMATIC NOISE LIMITER.

silent alarm In security systems, the transmission of a warning signal to attendant human operators and/or computers, without producing an audible or visible warning to intruders.

silent alignment See VISUAL ALIGNMENT.

silent piano See ELECTRONIC PIANO.

silica pencil A rod of silicon dioxide heated to emit infrared rays.

silicon Symbol, Si. A metalloidal element. Atomic number, 14. Atomic weight, 28.086. Silicon is abundant in the earth's crust. It is used in many semiconductor devices, including integrated circuits, diodes, photocells, rectifiers, and transistors.

silicon capacitor See VOLTAGE-VARIABLE CAPACITOR.

silicon carbide Formula, SiC. A compound of silicon and carbon valued as a semiconductor, an abrasive material, and a refractory substance. The commercial product is made by heating carbon and sand to a high temperature in an electric resistance furnace. Also called CARBORUNDUM.

silicon cell A type of photovoltaic cell using specially processed silicon as the light-sensitive material. This cell has a comparatively high voltage output.

silicon-controlled rectifier Abbreviation, SCR. A four-layer semiconductor device commonly used in power control applications (e.g., light dimmers and motor-speed controls). The electrodes are called the anode, the cathode, and the gate. The control signal is applied to the gate.

silicon-controlled switch Abbreviation, SCS. A four-terminal semiconductor switching device similar to the SILICON-CONTROLLED RECTIFIER. It is used for light-duty switching.

silicon crystal detector **1.** See SILICON JUNCTION DIODE. **2.** See SILICON POINT-CONTACT DIODE. **3.** A point-contact diode in which a lump of silicon is contacted by either a fine wire (cat whisker) or a blunt-tipped steel screw under pressure.

silicon detector See SILICON CRYSTAL DETECTOR.

silicon-diffused transistor A silicon bipolar transistor fabricated by diffusion techniques. It is characterized by high power-dissipation tolerance.

silicon diode A semiconductor diode in which the semiconductor material is specially processed silicon. Also see SILICON JUNCTION DIODE and SILICON POINT-CONTACT DIODE.

silicon dioxide Formula, SiO_2. A compound of silicon and oxygen. In the passivation of transistors and integrated circuits, a thin layer of silicon dioxide is grown on the surface of the wafer to protect the otherwise exposed junctions.

silicone A polymeric material characterized by a recurring chemical group containing oxygen and silicon atoms in the main chain as links. Various silicone compounds have numerous uses in electronics.

silicon junction diode A semiconductor diode using a pn junction in a silicon wafer. Compare SILICON POINT-CONTACT DIODE.

silicon on sapphire Abbreviation, SOS. Pertaining to integrated-circuit fabrication in which a silicon epitaxial layer is grown on a sapphire substrate.

silicon oxide A compound containing both silicon monoxide and silicon dioxide, and having dielectric properties. It is used in the manufacture of metal-oxide-semiconductor (MOS) devices.

silicon photocell A photocell using a silicon pn junction as the light-sensitive medium.

silicon photodiode A silicon diode constructed so that radiant energy can strike the barrier between the P- and N-type materials. A reverse bias is applied. When light falls on the P-N junction, current flows. The current is proportional to the intensity of the impinging energy, within certain limits. The greatest sensitivity is in the near infrared (IR). When energy of variable brightness falls on the P-N junction under conditions of reverse bias, the output current follows the intensity variations. This makes the device useful in fiberoptic communications systems.

silicon point contact The contact between a pointed metal wire (cat whisker) and a silicon wafer.

silicon point-contact diode A diode in which a tungsten wire (cat whisker) contacts a wafer of single-crystal silicon. It is useful at ultra high frequencies (UHF). Compare SILICON JUNCTION DIODE.

silicon rectifier A semiconductor rectifier consisting essentially of a junction between n- and p-type silicon inside a specially processed wafer or plate of single-crystal silicon.

silicon resistor See CRYSTAL RESISTOR.

silicon solar cell A relatively heavy-duty photovoltaic cell using specially processed silicon as the light-sensitive material.

silicon steel A high-permeability, high-resistance steel containing 2 to 3 percent silicon. It is used as core material in transformers and other electromagnetic devices.

silicon transistor A transistor in which the semiconductor material is single-crystal silicon.

silk-enameled wire Wire whose insulation is a layer of silk on top of an enamel coating.

silver Symbol, Ag. A precious metallic element. Atomic number, 47. Atomic weight, 107.87. It is used in circuits where low resistance and high Q are mandatory.

silver arsenide trisulfide See PROUSTITE.

silver-dollar construction Printed-circuit assembly on a disk-shaped board, about the size of a U.S. silver dollar.

silver-mica capacitor A fixed capacitor made by painting or depositing a silver layer (capacitor plate) on both faces of a thin mica film (dielectric separator).

silver migration The undesirable tendency of silver to be removed from one location and deposited in another under adverse environmental conditions.

silver-oxide battery A set of two or more silver-oxide cells stacked one atop the other, electrically connected in series. The resulting battery has a cylindrical shape. A set of four cells provides approximately 6 volts under no-load conditions; a battery of six cells provides 9 volts; a battery of eight cells provides 12 volts. See also SILVER-OXIDE CELL.

silver-oxide cell An electrochemical cell having a button-like shape, small enough to fit inside a wristwatch. There are several available sizes and thicknesses, all with similar appearance. The generated potential difference under no-load conditions is 1.5 volts, with a high ratio of stored energy per unit mass. The cell has a flat discharge curve; the voltage remains essentially constant until the charge is almost depleted, and then the voltage drops rapidly.

silver solder A solder consisting of an alloy of silver, copper, and zinc. It has a comparatively high melting temperature. Also see HARD SOLDER.

silverstat A multiconductor device used to adjust the balance of a resistance or reactance bridge.

similar decimals Two or more decimal numbers that have the same number of digits to the right of the radix point (e.g., 3.14 and 6.39, or 1.234 and 1.000).

simple quad A combination of two parallel paths—each containing two elements in series.

simple tone A pure sine-wave tone (i.e., one having negligible harmonic content).

simplex channel An information channel for unidirectional transmission.

simplex system **1.** In data communications, a system that transmits data in only one direction. Compare FULL DUPLEX SYSTEM and HALF DUPLEX SYSTEM. **2.** In voice communications via radio, a direct path over a single channel, used alternately for transmitting and receiving at each station.

simplex telegraphy Wire telegraphy in which only one message at a time can be sent over a line.

simplification of circuits See CIRCUIT SIMPLIFICATION.

simulation **1.** Imitation of the performance of a process, device, or system. **2.** The use of a mathematical model to represent a physical process, device, or system. **3.** The use of a computer, sometimes with virtual reality hardware and software, to mimic a real-life situation.

simulator **1.** A software or hardware system capable of simulation (see SIMULATION, **2**). **2.** A computer program whose implementation allows programs written for one computer to be compatible with another computer. **3.** A system of equipment for simulation (see SIMULATION, **1**).

simulcast **1.** To broadcast a program over two or more different channels at the same time. **2.** To broadcast a program over two or more different types of mode, for example, television and radio, at the same time. **3.** A program broadcast over two or more channels or modes at the same time.

simultaneous access See PARALLEL ACCESS.

simultaneous broadcasting See SIMULTANEOUS TRANSMISSION.

simultaneous computer See PARALLEL COMPUTER.

simultaneous transmission The transmission of the same information in two or more channels, or by means of two or more processes, at the same time.

sin Abbreviation of SINE.

sin^{-1} Symbols for the inverse of the sine function, also called the *arc sine.*

sine Abbreviation, sin. The trigonometric function a/c, the ratio of the opposite side of a right triangle to the hypotenuse.

SINAD See SIGNAL-TO-NOISE-AND-DISTORTION RATIO.

sine galvanometer A galvanometer in which the sine of the angle of deflection is proportional to the current. Compare TANGENT GALVANOMETER.

sine law The variation in radiation intensity in any direction from a linear source is proportional to the sine of the angle between the axis of the source and the direction of interest.

sine potentiometer A POTENTIOMETER whose output is proportional to the sine of the angle through which the shaft has rotated.

sine wave A periodic wave that can be represented by a sine curve (i.e., its amplitude is directly proportional to the sine of a linear quantity, such as displacement or time). Compare COSINE WAVE.

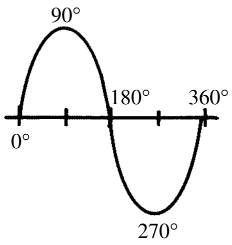

sine wave

singing Audible oscillation in a circuit or device, such as the low-level buzz emanating from the filament of a lamp dimmed with a phase-control circuit.

single-address coding In computer programming, the use of instruction words that contain the address for the location of the data to be operated on, and no other addresses.

single-board computer Abbreviation, SBC. A computer built entirely on one circuit board.

single-button microphone A carbon microphone having only one button attached to the diaphragm. Also see BUTTON MICROPHONE.

single-channel codec A form of CODEC intended for operation on a single signal source, rather than in a multiplexed system.

single-chip codec An integrated circuit contained entirely on one chip and in one package that can accomplish all CODEC functions. It can be a single-channel device or multiplexed.

single-cotton-covered wire Wire insulated with one layer of cotton.

single cotton enameled wire Wire insulated with one layer of cotton on top of an enamel coating.

single-crystal Pertaining to the internal structure of a crystalline material, in which the characteristic lattice is continuous throughout any size piece of the material. Also called MONOCRYSTALLINE.

single-crystal material A substance, such as a semiconductor, of which a sample, regardless of size, consists of only one crystal (i.e., there are no grain boundaries). Also see SINGLE-CRYSTAL. Compare POLYCRYSTALLINE MATERIAL.

single-crystal pulling See CZOCHRALSKI METHOD.

single-dial control Adjustment of a multistage system via one rotatable, calibrated control attached to a ganged arrangement that tunes all stages simultaneously.

single-diffused transistor A transistor in which only one diffusion of an impurity substance is made. Thus, in a diffused-base transistor, a single diffusion provides the base region and at the same time creates the emitter-base and collector-base junctions. Compare DOUBLE-DIFFUSED TRANSISTOR.

single-electron memory Abbreviation, SEM. A computer memory in which the movement of one electron can change a logic bit from 1 (high) to 0 (low) or vice versa.

single-element rotary antenna See ONE-ELEMENT ROTARY ANTENNA.

single-ended circuit A circuit that has one end grounded, as opposed to a double-ended circuit and push-pull circuit.

single-ended deflection In an oscilloscope or similar device, horizontal or vertical deflection provided by a single-ended deflection channel. Compare PUSH-PULL DEFLECTION.

single-ended input An input circuit with one terminal grounded (or the equivalent ungrounded input circuit). Also called *unbalanced input.* Compare BALANCED INPUT.

single-ended multiplexer A group of analog switches that selects from several analog signals.

single-ended output An output circuit with one terminal grounded (or the equivalent ungrounded output). Also called *unbalanced output.* Compare BALANCED OUTPUT.

single-ended push-pull circuit An arrangement, such as a complementary symmetry circuit, that provides push-pull output with single-ended input, but does not require transformers.

single-frequency Also called *fixed-frequency.* Pertaining to circuits or devices that normally operate at one frequency only (e.g., *single-frequency oscillator*).

single-frequency amplifier An amplifier that normally operates at only one frequency (or within a very narrow band of frequencies) (e.g., an intermediate-frequency amplifier, or a selective audio-frequency amplifier used for harmonic analysis or bridge balancing).

single-frequency duplex Two-way communications over one medium or frequency. Voice-actuated (VOX) or break-in devices are used at both ends of the circuit.

single-frequency oscillator An oscillator that normally delivers a signal at only one frequency until it is switched to another frequency (e.g., *crystal-controlled oscillator*).

single-frequency receiver A radio or television receiver that normally operates at one carrier frequency, rather than being tunable. Such receivers are used in monitoring specific programs, picking up standard-frequency signals, and in similar applications.

single-gun color picture tube A color-television picture tube in which the image is produced by a single beam that scans the red, green, and blue color-phosphor dots sequentially.

single-hop propagation Long-distance radio-wave propagation involving only one encounter with the ionosphere, and involving no intermediate reflections from the earth's surface.

single-hop return distance The return distance, as a function of the angle of departure from a radio transmission, from a layer of the ionosphere. The illustration shows the maximum possible distance under average conditions and assuming an angle of departure of zero degrees.

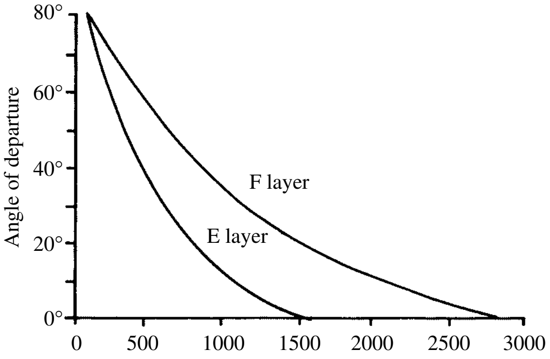

single-hop return distance (miles)

single-image response In an oscilloscope presentation, a single pattern, as opposed to a double-trace pattern.

single-inline package Abbreviation, SIP. A flat, molded component package having terminal pins along one edge. All the pins lie along a common line.

single-junction transistor See UNIJUNCTION TRANSISTOR.

single-layer coil A coil whose turns are wound side by side in one layer.

single-layer solenoid See SINGLE-LAYER COIL.

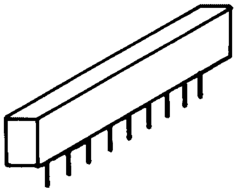

single-inline package

single-line tap In a telephone system, a connection that provides or designates a separate line (e.g., to serve a single household).

single-loop feedback Feedback through only one path.

single phase Pertaining to the presence or generation of one alternating-current phase only. Compare POLYPHASE.

single-phase full-wave Pertaining to a rectifier operated from a single-phase alternating-current (ac) power line, and rectifying both half-cycles of ac voltage. Compare SINGLE-PHASE/HALF WAVE.

single-phase/full-wave bridge A bridge rectifier operated from a single-phase alternating-current supply, usually from the untapped secondary winding of a transformer. Compare SINGLE-PHASE/FULL-WAVE CIRCUIT and SINGLE-PHASE/HALF-WAVE CIRCUIT.

single-phase/full-wave circuit A rectifier circuit in which each half-cycle of single-phase alternating current is rectified by a separate diode supplied from the ends of a center-tapped winding of a transformer. Compare SINGLE-PHASE/FULL-WAVE BRIDGE and SINGLE-PHASE/HALF-WAVE CIRCUIT.

single-phase/half-wave Pertaining to a rectifier operated from a single-phase alternating-current (ac) power line, and rectifying only one half-cycle of ac voltage.

single-phase/half-wave circuit A rectifier circuit in which a diode, output load, and single-phase alternating-current supply are connected in series, only one half-cycle of the cycle being passed by the diode. Compare SINGLE-PHASE/FULL-WAVE CIRCUIT and SINGLE-PHASE/FULL-WAVE BRIDGE.

single-phase rectifier See SINGLE-PHASE/FULL-WAVE BRIDGE, SINGLE-PHASE/FULL-WAVE CIRCUIT, and SINGLE-PHASE/HALF-WAVE CIRCUIT.

single-point ground One ground connection to which all channels of a circuit are returned. Such a common connection eliminates or greatly minimizes the common coupling often encountered when separate ground points are used.

single-pole double-throw Abbreviation, SPDT. Descriptive of an electrical, electronic, or mechanical switch with a pole that can be connected to either of two adjacent poles, but not to both.

single-pole single-throw Abbreviation, SPST. Descriptive of an electrical, electronic, or mechanical switch with a pole that can be connected to an adjacent pole (or disconnected from it) at will. It is used to provide the make and break function in a single circuit.

single rail **1.** A one-conductor communications medium with a ground return. **2.** A one-conductor data line, with a ground return.

single shot Also called *one shot.* Pertaining to circuit operation in which a single input pulse applied to a switching device (such as a multivibrator) causes it to deliver a single output pulse, rather than switch to a stable "on" state. A MONOSTABLE MULTIVIBRATOR operates in this mode.

single-shot multivibrator See MONOSTABLE MULTIVIBRATOR.

single sideband Abbreviation, SSB. Pertaining to a system of modulation in which one of the sidebands from an amplitude-modulated signal is attenuated or canceled out, leaving only one sideband. The carrier is generally suppressed also.

single-sideband suppressed-carrier Abbreviation, SSSC or SSBSC. Pertaining to a system of modulation in which the carrier and one sideband from an amplitude-modulated signal are suppressed; only the remaining sideband is transmitted. Sometimes this mode is simply called SINGLE SIDEBAND.

single signal Pertaining to a mode of reception in which signals appear on only one side of zero beat, enhancing selectivity and reducing interference among received signals. Most superheterodyne receivers have this feature; most direct-conversion receivers do not.

single-signal receiver A superheterodyne receiver that achieves high selectivity via a selective filter in the intermediate-frequency amplifier chain. Signals appear on only one side of zero beat, in a band whose width can be adjusted or selected for various values from about 200 Hz to 3 kHz.

single silk-covered wire Wire insulated with one layer of silk.

single-skip propagation See SINGLE-HOP PROPAGATION.

single-step operation See STEP-THROUGH OPERATION.

single sweep In an oscilloscope, a single time-axis deflection of the electron beam. Also see SWEEP, **1**, **2**. Compare RECURRENT SWEEP.

single-sweep blocking oscillator A blocking oscillator that cuts off after generating a single cycle or pulse.

single-throw switch A single-action switch with two or more poles.

single-tone keying Modulated continuous-wave keying. A single audio-frequency tone is used to amplitude-modulate or frequency-modulate the carrier.

single-track recorder A recorder, such as a magnetic-tape recorder or a graphic recorder, that permits recording along only one track.

single-trip multivibrator See MONOSTABLE MULTIVIBRATOR.

single-tuned circuit A circuit tuned by varying only one of its components [e.g., an intermediate-frequency transformer in which only the secondary coil (rather than both primary and secondary) is tuned].

single-turn coil **1.** A coil consisting of a single turn of wire, tubing, or strip. **2.** See RING INDUCTOR. **3.** See SHADING COIL.

single-turn potentiometer A potentiometer that can be adjusted through its entire resistance range by no more than one full rotation of the shaft. Usually, the turning range is somewhat less than a full circle (e.g., 300 degrees).

single-wire-fed antenna See WINDOM ANTENNA.

single-wire line **1.** See SINGLE-WIRE TRANSMISSION LINE. **2.** A single wire used for communication or control purposes. The earth furnishes the return path.

single-wire transmission line An antenna transmission line or feeder consisting of one wire only (see, for example, WINDOM ANTENNA).

sink **1.** A device or circuit into which current drains. **2.** See HEATSINK.

sink circuit The circuit associated with a load or other sink. Compare SOURCE CIRCUIT, **2**.

sinker A piece of semiconductor material used to reduce the base-collector junction resistance in a bipolar transistor.

sintering A process in which various solid bodies are formed from fusible powders at temperatures below their melting points. Example: *sintered magnetic core.*

sinusoidal Having the shape and properties of a SINE WAVE.

SIO Abbreviation of *serial input/output.*

SIP Abbreviation of SINGLE-INLINE PACKAGE.

six-phase rectifier A polyphase rectifier circuit operated from a three-phase supply. The output ripple frequency is six times the supply frequency.

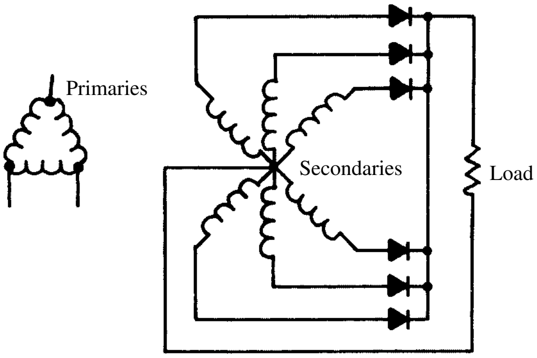

six-phase rectifier

SJD Abbreviation of SILICON JUNCTION DIODE.

skating In a phonograph turntable, the tendency of the tone arm to swing toward the spindle during record play, independent of the action produced by the stylus following the groove.

skeletal code A generalized computer routine needing only certain parameters to be usable for a specific application.

skeleton bridge A bridge consisting of an adjustable arm (potentiometer) and a pair of binding posts for each of the other three arms. Suitable resistors, capacitors, or inductors are connected to the binding posts to set up the bridge circuit desired.

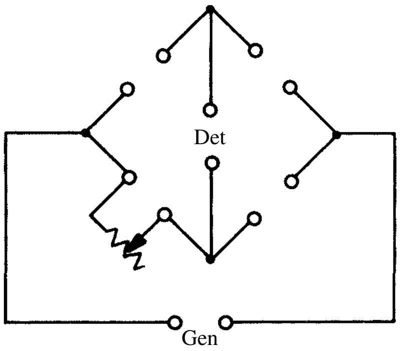

skeleton bridge

skeleton-type assembly **1.** A method of electronic-equipment construction in which a minimum of supporting members is used. An example is the use of an open framework, instead of a chassis, to support components. **2.** An assembly of electronic equipment, consisting essentially of a foundation unit (containing the basic circuitry) and plug-in units for setting up various complete equipments.

skew **1.** A condition resulting from failure of the horizontal synchronization in facsimile or television. The picture appears distorted, and appears as a non-rectangular parallelogram. **2.** In a print display, nonalignment of columns resulting from an incorrect number of line spaces in each line. **3.** In a probability function, an accumulation of values toward either side of center.

skewing **1.** The bending of a curve away from its normal shape. **2.** In a differential amplifier, the offset between two signals. Also see OFFSET.

Skiatron A special form of cathode-ray tube, with the fluorescent coating replaced by a screen of halide crystals that darken, instead of glow, when exposed to the electron stream.

skin depth The depth to which current penetrates below the surface of a conductor, as a result of the SKIN EFFECT.

skin effect The tendency of high-frequency alternating current to travel along the surface of a conductor; the high-frequency reactance is lower along the outside than at the center of a conductor. This tends to increase the resistivity of solid conductors at high alternating-current frequencies, as compared with low frequencies and direct current.

skip **1.** Ionosphere-reflected radio transmissions. **2.** In a computer program, an instruction whose sole function is that of causing a jump to the next instruction.

skip distance For a signal propagated via the ionosphere, the distance from the transmitter to the point at which the returned skywave strikes the earth.

skip fading For a signal propagated via the ionosphere, changes in signal strength caused by fluctuations in the altitude and/or contour of the ionized layer(s).

skip zone See ZONE OF SILENCE.

skirt selectivity **1.** The bandwidth between points of high attenuation (usually 30 dB or 60 dB) on the selectivity curve in a communications receiver. **2.** The relative steepness of the attenuation-vs.-frequency curve in a communications receiver.

SKM Abbreviation of *sine-cosine multiplier.*

skyhook **1.** Colloquialism for ANTENNA. **2.** A wire antenna supported by a captive balloon or kite.

sky noise Radio noise originating in outer space.

skywave A radio wave propagated by ionospheric reflections and/or refractions. Compare GROUND WAVE.

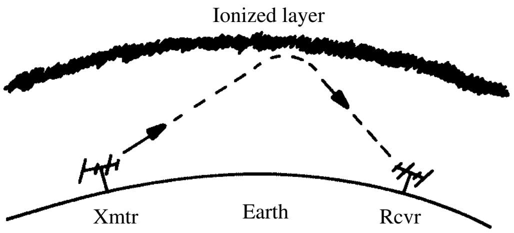

skywave

skywave correction A factor applied to long-range radionavigation signals to account for the time delay resulting from ionospheric propagation.

skywire See OUTSIDE ANTENNA.

slab **1.** A relatively thick body of quartz, ceramic, semiconductor, or dielectric. **2.** See SUBSTRATE.

slap back The return of sounds by an acoustically reflective object or surface a short distance away, resulting in almost immediate echoes.

slashed-field-gun CRT A straight-gun television picture tube (see STRAIGHT-GUN CRT). Because the gap between the anodes in this tube is slanted, the electrostatic field is diagonal, causing the electron and ion beams to be diverted at an angle.

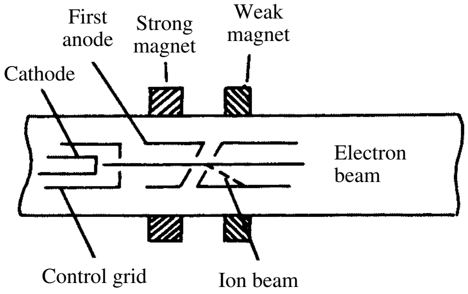

slashed-field-gun CRT

slave flash A photoflash operated by the light flash from another such unit.

slave relay A relay operated by, and whose action follows, a MASTER RELAY.

SLC Abbreviation of STRAIGHT-LINE CAPACITANCE.

sleeping sickness A gradual increase in transistor leakage current.

sleep machine An electronic device sometimes used as an aid for relaxation or sleeping. It consists of a wideband audio-frequency noise generator that produces low-level sounds similar to the noise of waves on a beach or a light wind through trees.

sleeve antenna A vertical antenna in which the upper half is a quarter-wave rod connected to the inner conductor of a coaxial feeder, and the lower half is a quarter-wave metal sleeve connected to the outer conductor of the feeder. Also called COAXIAL ANTENNA.

sleeving A material in tubular form that can be slipped over another material [e.g., insulating sleeving for wires (spaghetti)].

slew rate In an operational amplifier, the rate at which the output can be driven between its limits.

SLF Abbreviation of STRAIGHT-LINE FREQUENCY.

slice A semiconductor wafer cut from a single-crystal ingot.

slicer See CLIPPER-LIMITER.

slide-back meter An electronic voltmeter in which an unknown alternating-current signal voltage applied to the input of an amplifier stage is bucked by an internal, adjustable, accurately known signal voltage. The internal voltage is adjusted until a null occurs, indicating that its magnitude is equal to that of the unknown voltage.

slider A flat-spring contact that slides along the turns of a resistance or inductance coil to vary the coil's resistance or inductance. Also called a WIPER.

slide-rule dial A dial mechanism having a straight scale that resembles a slide rule.

slide switch A switch actuated by sliding a block-shaped button. Compare BAT-HANDLE SWITCH, PADDLE SWITCH, and ROCKER SWITCH.

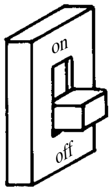

slide switch

slide wire A simple potentiometer consisting of a single, straight piece of resistance wire with a sliding contact. Also see SLIDE-WIRE RESISTOR.

slide-wire bridge A simple four-arm bridge in which the adjustable element is a single, straight resistance wire along which a clip or slider is moved, and that supplies two arms of the bridge (one on each side of the slider).

slide-wire resistor A variable resistor consisting of a single wire (straight or coiled) along whose length a slider is moved to vary the resistance.

sliding contact A contact that mates with another contact, or moves along a contacted surface, with a sliding motion. Also called SELF-CLEANING CONTACT and WIPING CONTACT.

slip **1.** In an eddy-current brake, coupling, or drive, the difference in speed between the field magnets and the iron eddy-current ring. **2.** In a synchronous motor, the difference between rotor speed and stator speed.

slip clutch In a gear or rack-and-pinion drive system, a device that releases the load if the torque becomes excessive. The gears then slip instead of being damaged.

slip ring See COLLECTOR RING, **1**.

slip speed See SLIP, **2**.

slope **1.** The slant of a line (graph) in rectangular coordinates, depicted as the ratio of the change in the dependent variable y to the change in the independent variable x. If (x_1,y_1) and (x_2,y_2) are two points on the line, then slope m is determined by $m = (y_2 - y_1)/(x_2 - x_1)$. **2.** The slant of a line in rectangular coordinates as defined in **1**, when the line is tangent to a curve (graph) at a specified point. **3.** The skirt(s) of a selectivity curve, particularly in a communications receiver, where a small change in frequency results in a significant change in gain or attenuation. **4.** The ratio of the extent of change in a quantity to the extent of change in some other quantity, when a causal relation exists between the magnitudes of the quantities. Example: See SLOPE RESISTANCE.

slope detector An amplitude-modulation (AM) receiving circuit detuned to one side of resonance (i.e., to a point along the skirt of the selectivity curve) to detect a frequency-modulated (FM) signal. The FM swing occurs along the slope of the resonance curve. Slope detection is useful in narrowband FM when conventional FM circuitry is not available.

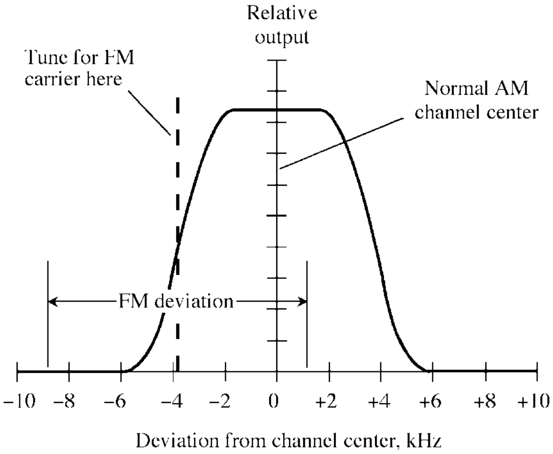

slope detector

slope resistance The ratio of a small change in voltage to a small change in current at an electrode or in a component.

slop-jar capacitor See WATER CAPACITOR.

slop-jar rectifier See ELECTROLYTIC RECTIFIER.

slot **1.** In the armature of a motor or generator, a groove in which the windings are laid. **2.** The notch in the response curve of a crystal filter.

slot antenna A microwave antenna that radiates energy through a slot cut in a surface, such as the metal skin of an aircraft.

slot cell A reinforcing, dielectric material (such as plastic) placed in the slot of a ferromagnetic core.

slot coupling Coupling microwave energy between a waveguide and a coaxial cable by means of two slots, one in the waveguide and the other in the outer conductor of the cable.

slot-discharge resistance See CORONA RESISTANCE.

slot insulation **1.** Insulation of wires in the slots of the armature of a motor or generator (see SLOT, **1**). **2.** A material in the form of tape or sheets, used for the purpose defined in **1**.

slot radiator See SLOT ANTENNA.

slotted line A device consisting of a section of air-dielectric coaxial line arranged for microwave measurements. The outer conductor is a metal cylinder, and the inner conductor is a concentric metal rod. The cylinder is provided with a lengthwise slot through which a small pickup probe extends for sampling the electromagnetic field inside the device. The probe is attached to a carriage that slides along a graduated scale on the outside of the cylinder. Radio-frequency energy is injected into one end of the line through a coaxial cable; as the probe moves along, response points are indicated by an external detector connected to the probe. The scale is read at these points to determine frequency, standing-wave ratio, impedance, and power. An alternate form of slotted line uses a section of slotted waveguide, instead of a section of coaxial line.

slotted rotor See SERRATED ROTOR PLATE.

slotted section See SLOTTED LINE.

slotted waveguide See SLOTTED LINE.

slot width **1.** The width of a slot in the armature of a motor or generator (see SLOT, **1**). **2.** The bandwidth of the notch in the response curve of a band-suppression filter of any kind. See, for example, NOTCH FILTER.

slow-acting relay Any relay designed to operate at some finite period following the application of the actuation voltage.

slow-blow fuse A fuse in which the melting wire breaks apart slowly. The time delay allows the fuse to withstand momentary current surges that do not damage the protected equipment, but that would cause a fast-blow fuse to break the circuit needlessly.

slow-break, fast-make relay A relay that opens slowly and closes rapidly.

slow-break, slow-make relay A relay that opens slowly and closes slowly.

slow charge Storage-battery charging in which a low current is passed through the battery over a long period of time. It ensures that the rated ampere-hour capacity will be restored to the battery.

slow death **1.** The gradual deterioration of transistor performance. **2.** The gradual deterioration in the performance of a component, circuit, device, or system.

slow drift The gradual change of a quantity or setting (usually in one direction). Compare FAST DRIFT.

slow-make, fast-break relay A relay that closes slowly and opens rapidly.

slow-make, slow-break relay A relay that closes slowly and opens slowly.

slow-operate, fast-release relay See SLOW-MAKE-FAST-BREAK RELAY.

slow-operate, slow-release relay See SLOW-MAKE-SLOW-BREAK RELAY.

slow-release, fast-operate relay See SLOW-BREAK-FAST-MAKE RELAY.

slow-release, slow-operate relay See SLOW-BREAK-SLOW-MAKE RELAY.

slow storage A form of memory with long storage and recovery time.

slow time scale An extended time scale, i.e., one larger than the time unit of the system under consideration.

SLS Abbreviation of SIDELOBE SUPPRESSION.

slug **1.** A movable core of ferromagnetic material, used to tune (varying the inductance of) coil by changing its position along the axis of the coil. Also see SLUG-TUNED COIL. **2.** A copper ring attached to the core of a relay for time-delay purposes (see SLUG-TYPE DELAY RELAY).

slug-tuned coil A coil whose inductance is varied by means of a ferromagnetic slug that slides in and out of the coil.

slug tuner A tuner for a radio or television receiver or test instrument, using slug-tuned coils.

slug-type delay relay A delayed-response relay that achieves time delay through the action of a heavy copper slug on the core. The slug forms a low-resistance, short-circuited single turn in which a current is induced by the magnetic flux, resulting from energizing the relay. The resulting flux of the slug opposes the buildup of relay-coil flux.

slumber switch An alarm-reset switch on an electronic clock radio. If the alarm activates, the slumber switch (usually a pushbutton device) can be pressed to turn off the alarm for a predetermined length of time. Also called *snooze button.*

SLW Abbreviation of STRAIGHT-LINE WAVELENGTH.

Sm Symbol for SAMARIUM.

small-current amplifier **1.** A direct-current (dc) amplifier for low-level input currents (i.e., currents of 1 milliampere or less). **2.** An amplifier (such as a silicon-transistor unit) requiring very low dc operating current.

small-scale frequency response For an analog circuit, the output frequency at which the level is –3 dB, relative to the maximum level, with a small signal at the input, normally 1 volt peak-to-peak.

small loop antenna A closed loop antenna with one to several turns and a circumference less than 0.1 wavelength at the highest operating frequency. This antenna is suitable for wireless reception, but generally not for transmission because the radiation resistance is extremely low. The antenna is least responsive along its axis, and is most responsive in the plane defined by its turn(s). The null along the axis is sharp and deep. A capacitor can be connected in series or parallel with the loop to provide a resonant response. This type of antenna can reduce interference caused by human-made noise or strong local signals. It is also useful for radio direction finding (RDF) at frequencies up to approximately 20 MHz. Compare FERRITE-ROD ANTENNA, LARGE LOOP ANTENNA.

small signal A low-amplitude signal. Such a signal covers so small a part of the operating characteristic of a device that operation is essentially linear. Compare LARGE SIGNAL.

small-signal analysis Analysis of circuit or component operation in which it is assumed that the signals deviate from (fluctuate to either side of) the steady bias levels by only a small amount. Also see SMALL SIGNAL.

small-signal bandwidth The frequency at which the output signal of an analog circuit decreases to –3 dB, relative to the value for direct current. The output voltage is generally set at 0.1 volt peak-to-peak for testing this value so that the circuit is not overdriven.

small-signal component **1.** A coefficient or parameter (such as amplification, transconductance, dynamic resistance, etc.) calculated or measured under conditions of small-signal operation. Also see SMALL SIGNAL and SMALL-SIGNAL EQUIVALENT CIRCUIT. **2.** A device designed for operation at low signal levels.

small-signal diode See SIGNAL DIODE.

small-signal equivalent circuit For a given transistor circuit, the equivalent circuit for low signal levels (i.e., at amplitudes lower than saturation and cutoff levels). Also see EQUIVALENT CIRCUIT.

small-signal operation Operation at low signal amplitudes (i.e., at signal levels that do not extend into the saturation or cutoff levels of a transistor, diode, or other component).

small-signal transistor A transistor designed for low-level applications, such as the amplification of small voltages and currents and low-voltage switching. Compare POWER TRANSISTOR.

SmallTalk A high-level computer programming language that uses a graphical user interface (GUI). It is used in complex design and research, and in robotics.

smart home or business An electronically controlled home or business, in which computers and robots take care of cooking, dish washing, floor scrubbing, waste removal, laundry, yard maintenance, snow removal, and other mundane chores. Intrusion detection and fire prevention are constantly maintained. In some cases, intruders can actually be identified or detained; fires can be extinguished by controlled sprinkler systems or remotely controlled robots. Some tasks can be monitored and controlled directly by the owner from remote locations, via telephone lines or wireless.

smartness The ability of an electronic system, especially a computer or control system, to perform a complete series of operations, substituting alternative steps, where necessary—all with a minimum of instructions from, and supervision by, human operators.

smearing In television or facsimile, a form of picture distortion caused by an excessively narrow receiving bandpass. The image appears fattened and horizontally blurred. Contrast might also be lost.

smectic crystal A liquid crystal in which the molecules are arranged in parallel layers and cannot slide past each other.

S meter In a radio communications receiver, a meter graduated in S units and/or decibels to indicate the strength of a received signal.

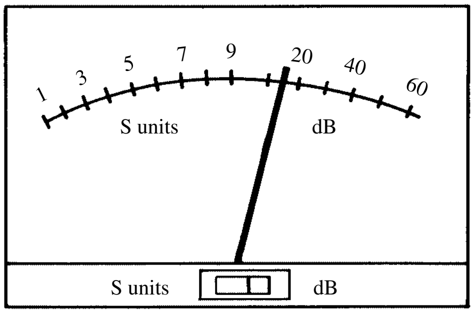

S meter

Smith chart A curvilinear graph on which complex-number impedance values can be plotted. It is useful in evaluating the behavior of radio-frequency circuits, transmission lines, and antenna systems—especially with regard to impedance mismatches and standing-wave ratio.

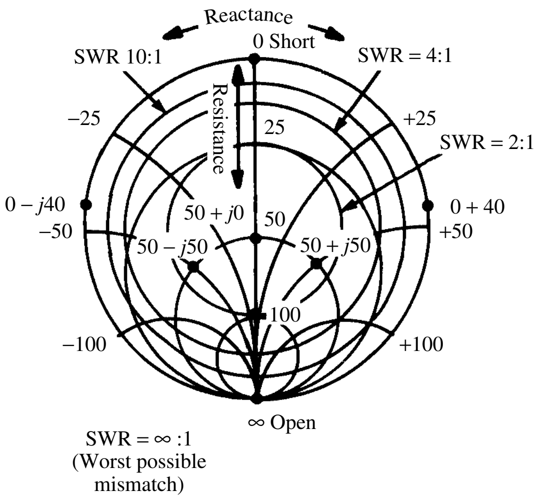

Smith chart

smoke alarm A device that produces audible and/or visible signals in the presence of smoke or unusual gases in the air. Also see PHOTOELECTRIC SMOKE ALARM.

smoke control See PHOTOELECTRIC SMOKE CONTROL.

smoke detector Any circuit or device used to sense the presence of smoke or noxious gases.

Some types detect changes in the ionization potential of the air; others sense changes in the dielectric constant of the air. Also see PHOTOELECTRIC SMOKE DETECTOR.

smoke sensor See SMOKE DETECTOR.

smooth **1.** Relatively free from surface irregularity. **2.** To reduce or eliminate irregularities in the voltage or current from a direct-current power source. **3.** To reduce or eliminate irregularities in data or signal amplitude.

smoothing choke A power-supply filter choke having a core with an air gap that prevents saturation at maximum rated direct current. Compare SWINGING CHOKE.

smoothing factor For a power-supply filter, a quantity approximately equal to $6.28fRC$, where f is the alternating-current frequency in Hertz, R is the filter resistance (in an RC filter) or the series reactance of the choke (in an LC filter) in ohms, and C is the filter capacitance in farads.

smoothing filter **1.** A filter for smoothing the alternating-current ripple component of a direct-current power supply following rectification. It can consist of one or more parallel capacitors of large value, and one or more series chokes of large inductance. **2.** A low-pass filter used at the output of a digital-to-analog (D/A) converter for eliminating high-frequency components generated by sampling.

SMPTE Abbreviation of SOCIETY OF MOTION PICTURE AND TELEVISION ENGINEERS.

smudge See SQUEEZEOUT.

SN Abbreviation of semiconductor network.

S/N Abbreviation of SIGNAL-TO-NOISE RATIO.

Sn Symbol for TIN.

snake **1.** A long, strong, flexible wire or strip used to pull other wires through electrical conduits. **2.** To route wires or cables through a group of circuits, components, or boards.

snap-action switch A switch that snaps quickly into the on or off position to prevent arcing and consequent premature contact deterioration.

snap diode A semiconductor diode in which switch-off time after carrier storage is extremely short (e.g., under 1 nanosecond).

snap magnet A magnet that reduces the tendency for arcing in relay-control instruments, thereby minimizing electromagnetic interference and prolonging contact life.

snap-on connector An electrical connector that locks in place, reducing the chance that it will detach unless it is deliberately removed.

snapshot dump During a computer program run, a dump, for debugging purposes, of certain storage areas.

snap switch See SNAP-ACTION SWITCH.

sneak current Unintended current flow through a path that is auxiliary to a main circuit.

sneak path A path through which current is accidentally detoured; it is usually a leakage path.

Snell's law A rule of physics that applies to visible light passing from air (or a vacuum) to some medium with an index of refraction c. If the light ray strikes the medium at an angle u, relative to a line normal (perpendicular) to the surface, and passes into the medium at angle v relative to the normal, then $(\sin u)/(\sin v) = c$.

sniffer See EXPLORING COIL.

sniperscope A telescope, snooperscope, or starlight scope for a carbine or rifle.

(S+N)/N Abbreviation of SIGNAL-PLUS-NOISE-TO-NOISE RATIO.

SNOBOL Acronym for *string-oriented symbolic language*, a computer-programming language for manipulating character strings.

snooperscope **1.** An infrared-sensitive device that permits viewing objects and surroundings in total darkness. It presents the image on a fluorescent screen. **2.** A rifle-mounted starlight scope.

snow A type of television picture interference that typically occurs when the signal-to-noise ratio is low (the reception is marginal or poor). Characterized by countless tiny out-of-focus light spots, whose rapid, random motion mimics the appearance of falling snow.

SNR Abbreviation of SIGNAL-TO-NOISE-RATIO.

soak value The smallest value of current that will cause saturation of a relay core.

Society of Motion Picture and Television Engineers Abbreviation, SMPTE. A group that decides on various procedures in video recording and reproduction, both on magnetic media (tape or disk) and on film.

socket A (usually female) fixture into which a plug, integrated circuit, or other component is inserted for easy installation in, or removal from, a circuit.

socket punch See PUNCH, **2**.

sodium Symbol, Na. A metallic element of the alkali-metal group. Atomic number, 11. Atomic weight, 22.9898.

sodium silicate See WATER GLASS.

sodium-vapor lamp A gas-discharge lamp containing neon and a small amount of sodium. After the filaments of the lamp are lighted for a short time, the heat vaporizes the sodium, and the filaments are disconnected by an automatic switch. Under the influence of the voltage across the lamp, the sodium vapor glows with a characteristic yellow light.

sofar A system for pinpointing the source of underwater sounds (coming from as far away as 2000 miles) through triangulation. The name is an acronym for *sound fixing and ranging.*

soft-drawn wire Wire that is highly malleable, and is therefore easily bent and unbent. Compare HARD-DRAWN WIRE.

soft iron A grade of iron, used in some cores, that is easily demagnetized.

soft solder A low-melting-point solder.

software **1.** Vendor-supplied or user-generated programs or groups of programs for a computer or computer system. **2.** The detailed instructions for performing a particular operation with a calculator or a computer.

software-defined radio Abbreviation, SDR. **1.** Wireless communications in which the modulation waveforms and protocols are generated and decoded by computer software. This allows a single, microcomputer-controlled radio receiver, transmitter, or transceiver to operate in a variety of services that use different protocols. Changing the modulation waveform or protocol requires only a change in the program run by a microcomputer that controls the radio. **2.** A receiver, transmitter, or transceiver that employs the technology defined in (1). **3.** The use of wireless equipment that employs the technology defined in (1).

soft X rays Low-frequency (long-wavelength) X rays. Such radiation has relatively poor penetrating power. Compare HARD X RAYS.

sol **1.** Abbreviation of SOLUTION. **2.** Abbreviation of SOLUBLE.

solar access For a specific property or location, the availability of direct exposure to the sun's rays as a source of energy.

solar absorption index A quantitative measure of the effect of the sun on the ionospheric absorption of radio waves.

solar activity See SOLAR RADIATION and SUNSPOT cycle.

solar battery A battery composed of solar cells connected in series and/or parallel for increased output.

solar cell A photovoltaic power transducer that converts visible light to electricity. It is called a cell because its output is a low direct-current voltage. Such cells can be connected in series and/or parallel to provide useful electric power output.

solar cycle See SUNSPOT CYCLE.

solar energy **1.** The total energy arriving from the sun, over a given period of time and within a specific surface area, at a given location on the surface of the earth. **2.** Any energy derived entirely from the sun.

solar-energy conversion Any process that changes solar radiant energy into another useful form.

solar flare A violent storm on the surface of the sun. These events tend to occur near the peak of the 11-year sunspot cycle. They cause an increase in the level of radio noise that comes from the sun, and they emit high-speed subatomic particles that reach earth a few hours after the first appearance of the flare. Because the particles are electrically charged, they are accelerated by the geomagnetic field. Sometimes a geomagnetic storm results, producing aurora near the poles and deterioration of ionospheric radio-propagation conditions.

solar flux An indicator of general solar activity. The solar noise level is measured at a particular frequency, such as 2800 MHz. The solar flux tends to be highest during periods of the greatest sunspot activity and immediately following a solar flare.

solar laser See SUNLIGHT-POWERED LASER.

solar noise Broadband electromagnetic noise generated by the sun. In particular, noise that occurs at radio frequencies and affects satellite, moonbounce, and space communications. This noise varies in intensity with the 11-year sunspot cycle, being generally higher at and near sunspot maxima. A solar flare can cause a sudden and dramatic increase in the noise at all wavelengths. See SOLAR FLARE.

solar panel An array consisting of a number of series-connected or series-parallel-connected solar cells mounted on a flat plate.

solar power Useful amounts of electricity obtained from suitable arrays of solar cells.

solar radiation Electromagnetic energy of various wavelengths originating in the sun. Such radiation, after passing through the earth's atmosphere, consists mostly of infrared rays and visible light. Some ultraviolet rays also reach the earth's surface.

solar relay See SUN SWITCH.

solar switch See SUN SWITCH.

solar wind Continuous emission of high-speed subatomic particles by the sun. It causes distortion of the lines of flux in the earth's magnetic field. It becomes more intense following a solar flare.

solder **1.** A metal alloy (usually of tin and lead) that is melted to electrically and mechanically join pieces of other metals. Also see HARD SOLDER and SOFT SOLDER. **2.** To join metals with solder.

solder

Solder type	*Melting point (°F/°C)*	*Principal uses*
Tin-lead 50 : 50, rosin-core	430/220	Electronic circuits
Tin-lead 60 : 40, rosin-core	370/190	Electronic circuits, low-heat
Tin-lead 63 : 37, rosin-core	360/180	Electronic circuits, low-heat
Tin-lead 50 : 50, acid-core	430/220	Non-electronic metal bonding
Silver	600/320	High-current, high-heat

soldering Joining (usually nonferrous) metal parts with solder, a lead-alloy substance. Compare BRAZING.

soldering gun An electric soldering iron having the general shape of a handgun. The element heats and cools more rapidly than the element in a typical soldering iron. The element is heated by pressing a device similar to the trigger on a pistol.

soldering iron An electric or nonelectric tool having a heated tip for melting solder.

solderless breadboard A foundation (see BREADBOARD, **1**) on which a circuit can be assembled by plugging components into tiny jacks without the use of solder.

solderless connection A connection between leads, or between leads and terminals, accomplished entirely through crimping, pinching, splicing, or wire wrapping. Solder is not used. Also see WIRE-WRAP CONNECTION.

solderless terminal A terminal to which a solderless connection can be made. Also see WRAP POST.

solenoid **1.** A coil of wire having a single layer, wound on a cylindrical form. **2.** A multilayer coil used as an electromagnet, and usually having a straight, iron core.

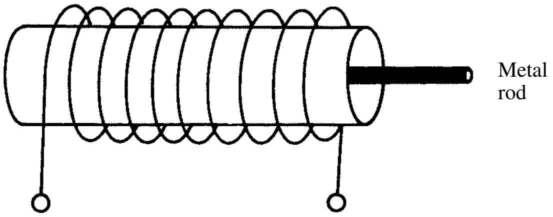

solenoid

solenoid switch A switch consisting of a solenoid coil (see SOLENOID, **2**) into which a core is pulled by the magnetic field to close a pair of contacts.

solid **1.** One of the states of matter. It is characterized by a definite shape and volume, and by atoms that maintain a fixed position, relative to each other. Compare GAS, LIQUID, and PLASMA. **2.** An enclosed, defined volume of three-dimensional space. **3.** In communications, descriptive of error-free reception of a series of coded signals. **4.** In printing and data transmission, a large print area whose entire surface is of equal and maximum intensity (of ink, light, or darkness).

solid angle Unit, steradian. The angle within the apex of a cone formed by all line segments between the center of a sphere and a defined circle on the surface of the sphere.

solid circuit Any circuit consisting of a single piece of hardware that is not normally separated into smaller parts.

solid conductor See SOLID WIRE.

solid electrolyte A solid substance affording ionic action similar to that in a liquid electrolyte.

solid electrolytic capacitor A capacitor using a solid electrolyte.

solid ground See DIRECT GROUND.

solid-state Pertaining to devices and circuits in which the flow of charge carriers (electrons and holes) is controlled in specially prepared blocks, wafers, rods, or disks of solid materials. Semiconductor devices, such as transistors and integrated circuits, are solid-state components.

solid-state battery An atomic battery consisting essentially of a photovoltaic cell in combination with a quantity of radioactive material, whose radiation causes the cell to generate electricity.

solid-state camera A video camera device that makes use of solid-state technology. The target is a matrix of charge-coupled devices (CCDs). When light strikes these devices, charge carriers are separated in a manner similar to that in a photovoltaic cell. The matrix is scanned according to a particular scheme, and the voltages developed in each CCD combine to produce the video signal output.

solid-state capacitor See SOLID ELECTROLYTIC CAPACITOR.

solid-state chronometer Any semiconductor device to indicate or measure time.

solid-state circuit See MONOLITHIC INTEGRATED CIRCUIT.

solid-state lamp **1.** See LIGHT-EMITTING DIODE. **2.** See LASER DIODE. **3.** See ELECTROLUMINESCENT CELL.

solid-state maser A device, such as the ruby maser, in which the stimulated medium is a solid material.

solid-state photosensor A semiconductor photodiode or phototransistor, as opposed to a phototube.

solid-state physics The branch of physics concerned with the nature and applications of such solids as electronic semiconductors.

solid-state relay **1.** A sensitive relay consisting of a conventional electromagnetic relay preceded by a transistorized amplifier. **2.** A completely electronic relay (i.e., one without moving parts) in which switching transistors provide the on and off states. **3.** See THYRISTOR.

solid-state thermometer An electronic thermometer utilizing one or more solid-state components, such as transistors, integrated circuits, or thermistors.

solid-state thyratron See SILICON-CONTROL-LED RECTIFIER and SILICON-CONTROLLED SWITCH.

solid-state tube A semiconductor device (diode, rectifier, transistor, SCR, etc.) whose housing and base allow it to replace directly an electron tube.

solid tantalum capacitor A capacitor using tantalum as a solid electrolyte.

solid wire Wire consisting of a single strand of metal. Compare STRANDED WIRE.

solute A substance that is dissolved in some other substance. Also see SOLUTION, **1**.

solution **1.** A well-diffused mixture of two or more substances. It can consist of a gas in a liquid, a gas in a solid, a gas in a gas, a liquid in a solid, or a solid in a solid. A solution, typically, is molecular (i.e., there is no chemical reaction between its constituents). Also see SATURATED SOLUTION; SOLUTE; SOLVENT, **1**; and SUPERSATURATED SOLUTION. **2.** The result of solving a problem or making a calculation. Also called *answer* or *result.*

solution conductivity The electrical conductivity of a solution, such as an electrolyte. The conductivity (and conversely, the resistance) depends on the number and mobility of ions in the solution.

solution-conductivity bridge A direct-current bridge specially designed and calibrated to measure the conductivity of chemical solutions.

solution pressure In an electrolyte into which a metal body is immersed, the force that causes the metal to tend to pass into solution as positive ions and to form a Helmholtz double layer.

solvent **1.** A fluid that dissolves other materials. **2.** The constituent of a solution that dissolves one or more other constituents. Thus, in a saltwater solution, water is the solvent and salt the solute. Also see SOLUTION, **1**.

SOM Abbreviation of *start of message.*

Sonalert Tradename for a small, but loud, sound reproducer used with solid-state circuits for alarm purposes.

sonar A system of detection and ranging by means of sonic and ultrasonic signals. In this system of echo ranging, the distance to an underwater object is determined from the time it takes a sound signal to reach the object and be reflected back to the transmitter. The name is an acronym for *sound navigation and ranging.*

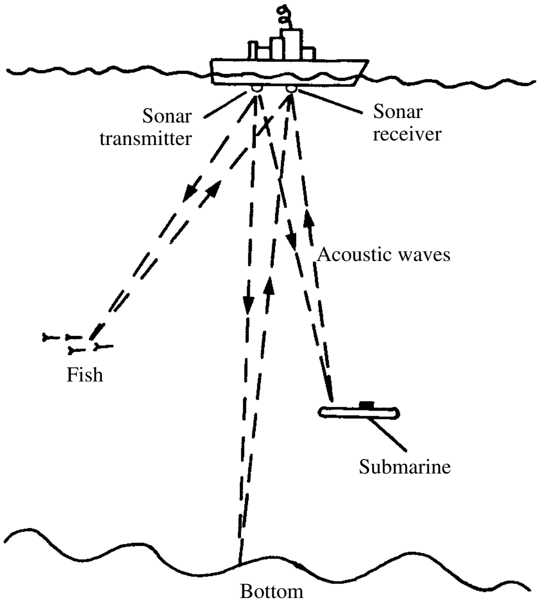

sonar

sonde A device for automatically gathering meterological data at high altitudes. An example is the *radiosonde.*

sone A unit of loudness for an individual listener. The level of 1 sone is the loudness of a 1000-Hz tone that is 40 dB above the particular listener's threshold of hearing.

sonic altimeter An altimeter (see ABSOLUTE ALTIMETER) using sound waves. The time required for a transmitted wave to reach a target and be reflected back to the transmitter is proportional to the distance between the transmitter and the target.

sonic boom An explosive sound occurring when the shock wave produced by an aircraft flying at supersonic speed strikes the earth.

sonic delay line A delay line using electroacoustic transducers and an intervening medium through which a sound wave is transmitted.

sonic depth finder See ACOUSTIC DEPTH FINDER.

sonic thermocouple A thermocouple whose heat-absorbing properties are enhanced by subjecting it to acoustic vibrations.

sonobuoy A buoy equipped with an acoustic receiver and radio transmitter. The device is parachuted into the water, where it receives submarine sounds and transmits them to a monitoring station. Several sonobuoys communicating with a computer will track the path of a submarine.

sonovox An electronic device used to produce special sound effects when it is held against the throat of the operator. The special sounds are formed into words by the operator's mouth.

SOP Abbreviation of *standard operating procedure.*

sophisticated electronics Advanced electronics theory and operations, usually dealing with complex devices or systems and requiring rigorous analysis to describe their operation and devise applications. Compare UNSOPHISTICATED ELECTRONICS.

sorption processes Processes whereby certain substances (e.g., activated charcoal) occlude and retain gases and vapors. A chamber containing such a substance is often useful in the production of a vacuum. Sorption includes both absorption and adsorption.

sort **1.** To group information items using their keys. Also see KEY. **2.** To group information items according to some system of classification, as to print an alphabetical list of words stored in a random sequence.

sorting routine A computer program for sequencing data items according to key words (values in specific fields) of the different records.

SOS **1.** The international radiotelegraph distress signal; equivalent to *mayday* in radiotelephony. It consists of three dits ("dots"), followed by three dahs ("dashes"), followed by three more dits. **2.** Abbreviation of SILICON ON SAPPHIRE.

sound The vibratory or wave phenomenon to which the sense of hearing is responsive. Conducted by waves in solids, liquids, and gases; not propagated through a vacuum.

sound absorption The nonreflection and nontransmission of acoustic energy by a body or medium, and the attendant conversion of the acoustic energy into another form of energy (usually heat).

sound absorption coefficient A quantitative expression of the extent to which a surface absorbs acoustic energy (as opposed to reflecting, transmitting, diffusing, or scattering it).

sound amplifier **1.** An audio amplifier—especially the sound channel of a television system. **2.** A device, such as a horn or reflector, that directly boosts the intensity of sound at a given listening point.

sound analyzer An instrument, often a wave analyzer equipped with a microphone, for measuring such characteristics of sound as amplitude, frequency (pitch), and harmonic content (timbre).

sound articulation See ARTICULATION.

sound bars In a television picture, horizontal bars resulting from interference between the audio and video channels of the receiver.

sound carrier In a television signal, the frequency-modulated carrier that transmits the audio part of the program. Compare VIDEO CARRIER.

sound chamber An air enclosure, usually a box or can, for modifying the acoustic qualities of sound or of an audio signal.

sound detector The discriminator or ratio detector that demodulates the sound signal in a television receiver circuit.

sound-energy density Sound energy per unit volume, expressed in joules per cubic meter or ergs per cubic centimeter.

sound-energy flux The average rate of flow of sound energy through a specified area, as expressed in ergs or joules per second.

sound field A volume of space or material containing sound waves.

sound film Motion-picture film on which a sound track is recorded. Also see OPTICAL SOUND RECORDING.

sound flux The rate of flow of sound energy, usually expressed in terms of sound pressure at a point or over a unit area normal to the direction of sound propagation.

sound gate An optical device used to convert the sound track of a movie film into electrical impulses.

sound generator Any combination of oscillator, amplifier, and transducer (loudspeaker or headphones) for producing sound waves.

sound-hazard integrator An instrument used to measure cumulative noise exposure received by persons in noisy environments. One such instrument provides direct readings in percent of permissible exposure.

sound IF amplifier In a television receiver circuit, the separate amplifier for the sound intermediate frequency. See, for illustration, INTERCARRIER RECEIVER and SPLIT-SOUND RECEIVER.

sound-level meter See SOUND SURVEY METER.

sound marker A marker indicating the sound-carrier point on a television alignment curve displayed on an oscilloscope screen.

sound-marker generator A special radio-frequency signal generator (or a special circuit in a television-alignment generator) for the production of a sound marker.

sound mirage See ACOUSTIC MIRAGE.

sound mix In sound recording or reproduction, the composite output from an audio mixer circuit.

sound-on-film recording See OPTICAL SOUND RECORDING.

sound-on-sound recording The simultaneous recording (on a single track on magnetic tape) of new material with previously recorded material. The old recording is not erased.

sound-operated relay **1.** A relay operated indirectly from sound, through the medium of a pickup microphone and amplifier. **2.** A relay having a delicately poised armature that operates directly from sound vibrations.

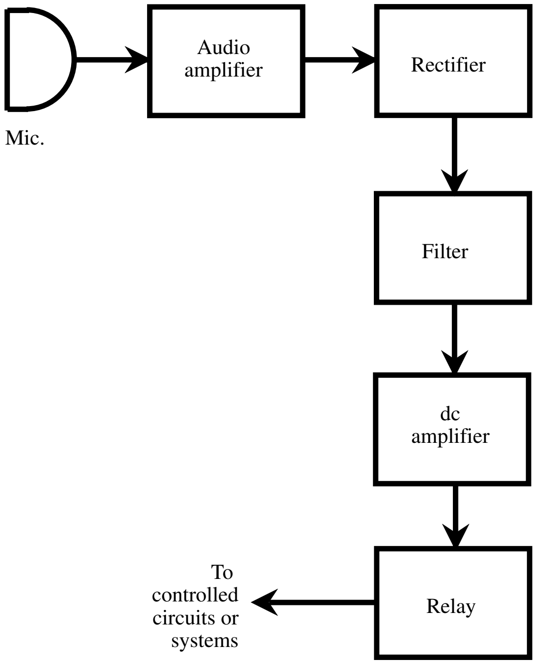

sound-operated relay

sound power The total sound energy per unit time produced by a sound source, as expressed in ergs per second or in watts.

sound power level The extent, in decibels, by which SOUND POWER exceeds one picowatt (10^{-12} watt).

sound pressure **1.** The force exerted by sound waves on a surface area, expressed in dynes per square centimeter (as an rms value over 1 cycle). The sound pressure is proportional to the square root of sound-energy density. **2.** The instantaneous difference between actual air pressure and average air pressure at a given point.

sound pressure level The extent, in decibels, via which SOUND PRESSURE exceeds 20 micropascals (2.0×10^{-5} pascal).

sound probe A transducer used to receive acoustic vibrations for detection or measurement purposes.

sound recording The electrical recording of sound, using cylinder, disc, tape, wire, or other comparable storage medium.

sound-recording system A complete, integrated array of equipment for recording sound, including such components as microphones, amplifiers, pickups, filters and other shaping networks, attenuators, level indicators, and recording mechanisms. Compare SOUND-REPRODUCTION SYSTEM.

sound reinforcement Intensification of sound by horns, resonant chambers, or other acoustical devices.

sound relay See SOUND-OPERATED RELAY.

sound reproduction The electrical reproduction of sound from recordings on vinyl discs, magnetic tapes, magnetic discs, compact optical disbs, etc.

sound-reproduction system A complete, integrated array of equipment for the playback of recorded sound, including such components as tape or record players, amplifiers, filters and other shaping networks, attenuators, level indicators, loudspeakers, and headphones. Compare SOUND-RECORDING SYSTEM.

sound spectrograph A device that produces a display of sound amplitude vs. frequency. Similar to a SPECTRUM ANALYZER, except that it operates at audio frequencies (about 20 Hz to 20 kHz) and is actuated by acoustic disturbances, rather than by electromagnetic signals.

sound spectrum The continuous band of frequencies (about 20 Hz to 20 kHz) constituting audible sounds, and sometimes the immediately adjacent (subaudible and superaudible) frequencies.

sound stage The apparent dimensions of a sound source.

sound survey meter A portable instrument for measuring the intensity and other characteristics of sound.

sound sweetening In audio recording or reproduction, the modification of the sound to achieve some desired effect.

sound system A sound-recording system, sound-reproduction system, or a combination of the two.

sound takeoff In a television receiver circuit, the point at which the frequency-modulated sound signal is extracted from the complex signal.

sound track The variable-density or variable-width recording on one side of the film in sound-on-film recording and reproduction. Also see OPTICAL SOUND RECORDING.

sound transducer See ACOUSTIC TRANSDUCER.

sound-transmission coefficient See ACOUSTICAL TRANSMITTIVITY.

sound trap In a television receiver circuit, a wavetrap that prevents the sound signal from entering the picture channels.

sound unit See PHONE, **2**.

sound wave The vibratory phenomenon produced in a medium by acoustic energy. A sound wave in air consists of alternate compressions and rarefactions of the air. Also see ACOUSTIC WAVE.

source **1.** The origin of a signal or electrical energy (e.g., a transmitting station). **2.** In a field-effect transistor, the electrode that is equivalent to the emitter of a common-emitter-connected bipolar transistor, or the cathode of a vacuum tube. **3.** That which is being transcribed to magnetic tape. **4.** Manufacturer, wholesaler, or retailer.

source circuit **1.** The circuit associated with the source electrode of a field-effect transistor. **2.** A generator circuit. Compare SINK CIRCUIT.

source code See SOURCE LANGUAGE.

source computer A computer for compiling a source program.

source-coupled multivibrator A multivibrator circuit using field-effect transistors, in which feedback coupling is achieved with a common source resistor for the two FETs. This circuit is equivalent to the emitter-coupled bipolar-transistor-type multivibrator.

source data automation A means of storing a master data file for easy duplication, whenever necessary.

source deck An audio or video tape player that reproduces original recordings in the editing process.

source follower A field-effect-transistor circuit in which the output is taken across a resistor between source and ground. This circuit is equivalent to the emitter follower, and is a unity-gain stage whose impedance-transformation characteristics make it ideal in signal conditioning, buffering, and impedance-matching applications.

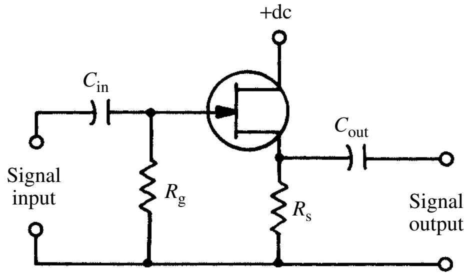

source follower

source impedance **1.** The impedance of a generator in a circuit. **2.** The impedance of the source electrode of a field-effect transistor.

source language A computer programming language from which is derived (by a compiler) the machine (object) language on which the computer operates.

source program A computer program written in a source language.

south magnetic pole The south pole of the equivalent bar magnet constituted by earth's magnetic field (see EARTH'S MAGNETIC FIELD). The south magnetic pole lies close to the geographic south pole. Compare NORTH MAGNETIC POLE.

south-seeking pole Symbol, S. The so-called south pole of a magnet. When a bar magnet is suspended horizontally, this pole points to earth's south magnetic pole. Compare NORTH-SEEKING POLE.

SP Abbreviation of STACK POINTER.

sp **1.** Abbreviation of *single pole.* **2.** Abbreviation of *special.*

space-charge field **1.** The electric field existing within a group of charged particles. **2.** The electric field existing in a plasma.

space diversity See DIVERSITY RECEPTION.

space-diversity reception See DIVERSITY RECEPTION.

space division A method of data transfer in which different paths are used for the transmission of different signals.

space-division switch A switch having two or more ports and different paths connecting the ports.

space lattice The three-dimensional, redundant pattern formed by atoms and molecules in a crystal and having a shape that is characteristic of a particular crystalline material.

spacer An insulating rod or bar that serves to hold apart the conductors of a two-wire, four-wire, or coaxial air-dielectric transmission line.

space suppression Following the printing of a line by a printer, the prevention of normal paper travel.

space-time-space switch Abbreviated STS switch or STSS. A large switching array with two space switch blocks and a time switch block between them.

space wave One of the components of an electromagnetic ground wave. The space wave, unlike the *surface wave*, is not earth-guided. It has two components: the *direct wave* and the *ground-reflected wave.*

spaghetti Slender, varnished-cambric tubing used as slipover insulation for wires and busbars.

span On an instrument scale, the difference between the highest value and the lowest value.

spark See ELECTRIC SPARK.

spark absorber **1.** See SPARK SUPPRESSOR. **2.** See KEY-CLICK FILTER.

spark coil A small induction coil. Its name is derived from its initial purpose of supplying the high voltage for spark plugs in gas engines.

spark energy The energy dissipated by an electric arc or spark.

spark gap A device consisting essentially of two metal points, tips, or balls that are separated by a small air gap. A high voltage applied to the electrodes causes a spark to jump across the gap.

sparking distance The maximum separation of the electrodes of a spark gap at which a given voltage will produce a spark.

sparking voltage The lowest voltage that will cause a spark to jump across a gap of a given width.

spark killer See SPARK SUPPRESSOR.

sparkover A discharge in air, a vacuum, or a dielectric. It is characterized by sparking between electrodes in the medium.

spark plate In some automobile radios, a noise-interference-eliminating bypass capacitor in which the chassis is one plate.

spark-plug suppressor A small resistive device connected in series with a spark plug to suppress electrical noise arising from the ignition in an internal combustion engine.

spark quencher See SPARK SUPPRESSOR.

spark suppressor A resistor, capacitor, and/or diode used to eliminate or minimize sparking between make-and-break contacts.

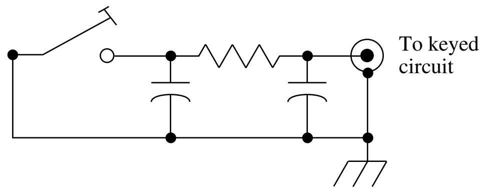

spark suppressor

spark-suppressor diode See SUPPRESSOR DIODE.

spatial distribution The three-dimensional directional pattern of a transducer (such as an antenna, microphone, or speaker).

spc **1.** Abbreviation of *silicon point contact.* **2.** Abbreviation of *silver-plated copper.*

SPDT Abbreviation of SINGLE-POLE/DOUBLE-THROW (switch or relay).

speaker See LOUDSPEAKER.

speaker damping See DAMPED LOUDSPEAKER.

speaker-level audio In a sound reproduction system, radio receiver, or other audio circuit, an audio-frequency signal of sufficient amplitude to drive a speaker or speaker system directly, without the need for additional amplification.

speaking arc A method of modulated-light transmission. An electric arc is modulated by audio-frequency signals.

special character A printed, displayed, or encoded character other than a numeral or letter, such as an ampersand (&) or a pound sign (#). Also called SYMBOL.

special effects Various techniques used in filmmaking, computer animation, and videotape recording for achieving certain visual scenes or images.

special-purpose computer A computer designed to handle problems or be suitable for applications of a specific category; a dedicated computer.

special-purpose calculator An electronic calculator intended for essentially "nonmathematical" purposes, such as biorhythm data, astrological information, metric conversions, musical composing, etc.

specific address See ABSOLUTE ADDRESS.

specification **1.** For an electronic device, a statement of performance over specific parameters. Example: for a high-fidelity stereo amplifier, 50 watts per channel over a frequency range of 10 Hz to 30 kHz, with less than 1 percent total harmonic distortion. **2.** A precise listing of requirements or expectations.

specific conductivity Conductance per unit volume. In SI units, this is expressed in siemens per cubic centimeter (S/cm^3).

specific dielectric strength For an insulant, the dielectric strength per millimeter of thickness.

specific gravity Abbreviation, sp gr. The ratio of the density of a material to the density of a substance accepted as a standard (usually water at 4 degrees Celsius or 39.2 degrees Fahrenheit).

specific inductive capacity See DIELECTRIC CONSTANT.

specific resistance See RESISTIVITY.

specific sound-energy flux See SOUND INTENSITY.

spectral comparative pattern recognizer Acronym, SCEPTRON. Equipment used to classify automatically complex signals obtained from information that has been converted into electrical signals.

spectral density For a complex signal, the amount of energy contained within a given band of frequencies.

spectral energy distribution The occurrence of different amounts of energy in different areas of a spectrum (as in a visible-light spectrum, sound spectrum, or radio-frequency spectrum).

spectral response The characteristic of a device, such as a photocell or the human eye, that describes the device's sensitivity to radiations of various frequencies in a given spectrum.

spectral sensitivity The color response of a light-sensitive device.

spectrograph A recording SPECTROMETER.

spectrometer **1.** An instrument used to measure spectral wavelengths. **2.** An instrument used to measure the index of refraction. **3.** See MASS SPECTROMETER.

spectrophotometer A photoelectric instrument for chemical analysis. In the device, light passing through the material under analysis is broken up into a spectrum that is examined with a photoelectric circuit, which, in turn, plots a spectrogram.

spectroscope An instrument that resolves a radiation into its various frequency components and permits measurement of each.

spectrum A band of frequencies or wavelengths (e.g., *radio spectrum*, *visible-light spectrum*, and *sound spectrum*).

spectrum analyzer **1.** An automatic wave analyzer with a visual display (oscilloscope). **2.** A scanning receiver with a screen that shows a plot of signals and their bandwidths over a specific frequency band.

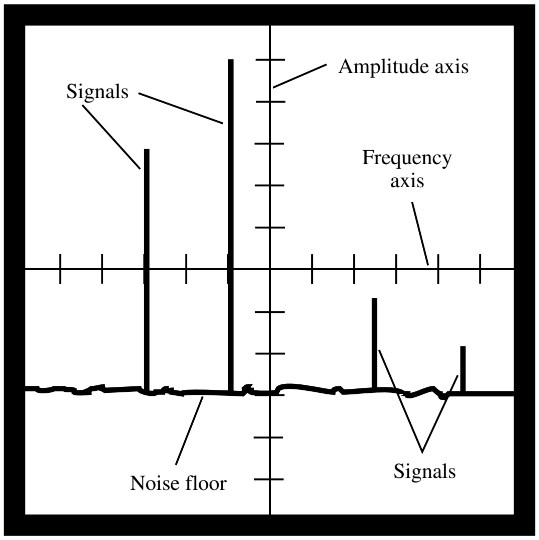

spectrum analyzer, 2.

specularity A qualitative or experimentally derived expression of the efficiency with which a DIFFRACTION GRATING works.

specular reflection Reflection in which the reflected ray is in the same plane as the incident ray, as in reflection from an extremely smooth surface.

speech amplifier A (usually low-level) audio amplifier designed especially for speech frequencies. It is generally used to amplify the signals from a microphone.

speech clipper A device, such as a biased diode, (see POSITIVE PEAK CLIPPER) for holding speech signals to a constant amplitude. Compare SPEECH COMPRESSOR. Also see SPEECH CLIPPING.

speech clipping The use of a limiting circuit to maintain the output-signal amplitude of a speech amplifier against fluctuations in the intensity of speech input. The resulting signal requires filtering to remove harmonics generated by the process. Compare SPEECH COMPRESSION.

speech compression Automatic regulation of the gain of a speech amplifier to maintain its output-signal amplitude against speech-input fluctuations. Compare SPEECH CLIPPING.

speech compressor A circuit or device, such as an automatic-gain-control (agc) system, for performing speech compression. Compare SPEECH CLIPPER.

speech digit signaling A method of digital signaling, where time slots generally used for encoded audio or video are used alternately for signaling.

speech frequencies See VOICE FREQUENCIES.

speech intelligibility The quality of reproduced speech that makes it easily understood by a reasonably proficient user of the language. For good speech intelligibility, a circuit should transmit frequencies between 300 Hz and 3000 Hz with minimal distortion. Increased bandwidth improves fidelity, but does not provide a significant increase in intelligibility for normal speech.

speech inverter See SCRAMBLER CIRCUIT.

speech power **1.** The alternating-current power in an electric wave corresponding to speech, as opposed to that in a sine wave. **2.** Sound power in a speech transmission.

speech recognition The ability of a device to translate audible spoken words, phrases, or sentences into binary digital signals that can be used by machines, such as computers and robots.

speech recognizer An electronic device that translates audible spoken words, phrases, or sentences into binary digital signals that can be used by machines, such as computers and robots.

speech scrambler See SCRAMBLER CIRCUIT.

speech synthesis The ability of a device to translate binary digital signals from a machine, such as a computer or robot, into audible, coherent spoken words, phrases, or sentences.

speech synthesizer An electronic device that translates binary digital signals from a machine, such as a computer or robot, into audible, coherent spoken words, phrases, or sentences.

speed key **1.** A semiautomatic key for manual generation of Morse code characters. **2.** A similar, fully automatic key used especially for high-speed telegraphy.

speed of light Symbol, *c*. The speed at which electromagnetic waves propagate in a vacuum; approximately 299,792 kilometers per second (186,282 miles per second).

speed of sound The speed at which acoustic waves propagate. It depends on the nature of the medium. In air, at ordinary temperatures, it is approximately 344 meters (1129 feet) per second. In fresh water, it is approximately 1463 meters (4800 feet) per second.

speed of transmission The amount of data sent in a given unit of time. It is generally measured in

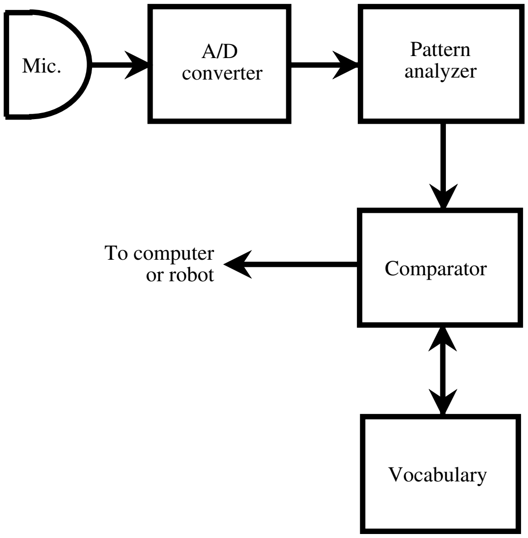

speech recognizer

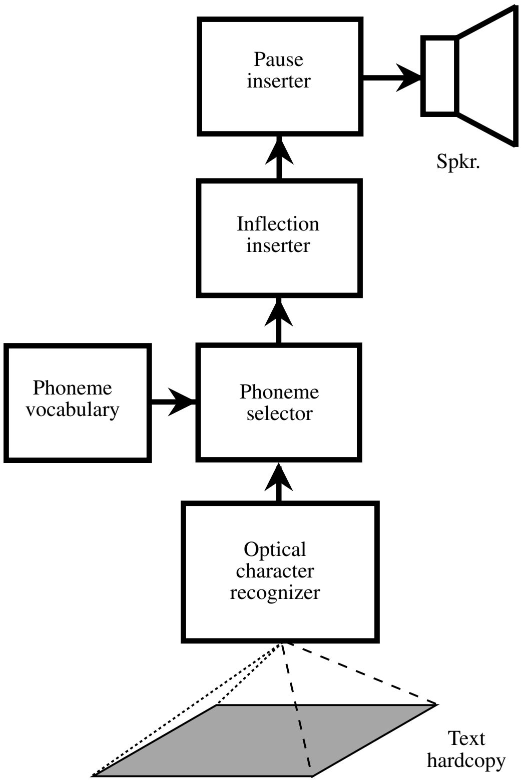

speech synthesizer

bits per second (bps), characters per second (cps), characters per minute (cpm), or words per minute (wpm). It is used primarily for digital codes.

speedup capacitor See COMMUTATING CAPACITOR.

SPFW Abbreviation of SINGLE-PHASE FULL-WAVE.

sp gr Abbreviation of SPECIFIC GRAVITY.

sphere **1.** A closed surface in three-dimensional space, represented by the set of all points equidistant from a specified center point. **2.** A solid in three-dimensional space, represented by the set of all points on or within a closed surface, as defined in **1**.

sphere gap A spark gap in which the spark passes between two polished metal spheres. When the air gap is adjustable, unknown high voltages can be measured in terms of the largest gap width at which sparking occurs. Compare NEEDLE GAP.

sphere-gap voltmeter See SPHERE VOLTMETER.

sphere voltmeter A gap voltmeter using a SPHERE GAP.

spherical aberration In a spherical lens, mirror, or reflecting dish, distortion as a result of the spherical (as opposed to paraboloidal) shape of the surface. This causes the focus to be elongated into a short line segment along the principal axis.

spherical angle An angle formed by the intersection of two arcs on the surface of a sphere.

spherical coordinate geometry A scheme for guiding a robot arm in three dimensions via SPHERICAL COORDINATES. The length (radius) of the arm can be varied, as can the elevation (latitude) and azimuth (longitude).

spherical coordinates A method of defining a point (P) in three-space using two angles (latitude and longitude) and a radial distance (r) from the origin.

spherical degree A unit equal to $\frac{1}{720}$ of the surface area of a sphere.

spherical distance The length of the shortest arc (lying on a great circle) connecting two specified points on a sphere.

spherical divergence The manner in which energy normally propagates from a fixed point source in three dimensions. Wavefronts expand from the source in the form of spheres, whose centers are at the point source.

spherical reflector A microwave reflector (dish) whose contour is that of a sphere, rather than that of a paraboloid.

spherical wave A wave characterized by wavefronts that are concentric spheres.

spheroidal antenna A sheetmetal or wire-mesh antenna having the cross section of a sphere that is flattened at the ends of one axis.

SPHW Abbreviation of SINGLE-PHASE HALF-WAVE.

spider **1.** The flat, round, springy part that holds the apex of the vibrating cone of a dynamic speaker. **2.** A quickly assembled, chassisless hookup in which the pigtails of components form the wiring and the mechanical support for a circuit.

spiderweb antenna A set of dipole antennas for different frequencies, arranged in a common (usually horizontal) plane. The result is a broadband antenna.

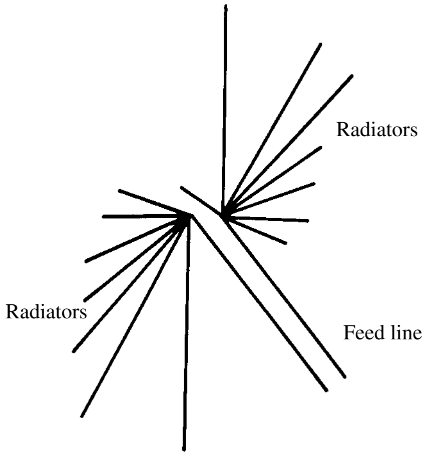

spiderweb antenna

spiderweb coil A flat, single-layer coil in which a strand of wire is woven over and under the spokes of a wheel-like form, and having the general appearance of a spiderweb. The criss-cross winding reduces distributed capacitance by breaking up the parallelism of adjacent turns.

spike **1.** A current or voltage pulse of extremely short duration. **2.** A sharp transient, such as an overshoot on a pulse or square wave.

spike suppressor A clipper or similar device for removing a spike from a signal voltage.

spin A quantity of angular momentum possessed by a subatomic particle. It can be positive, negative, or zero, but it always exists in integral multiples of ½.

spindle **1.** The pivoted shaft that carries the movable element and rotates between the pivots in a meter movement. **2.** The rotor of an alternator—especially when the rotor is a permanent magnet. **3.** The rotating shaft in a motor, generator, or similar electric machine.

spinning electron An electron having nonzero angular momentum.

spinthariscope An optical device for observing the effect of alpha particles emitted by a radioactive substance, from the scintillations they produce upon striking a small, fluorescent screen.

spiral coil See DISK WINDING.

spiral distortion A form of television camera-tube distortion caused by spiraling of the electrons as they are emitted from the photocathode. The result is a twisted image.

spiral-rod oscillator A parallel-line-type oscillator in which the lines are rods that are rolled up into spirals to conserve space. Also see LINE-TYPE OSCILLATOR.

spiral sweep **1.** A means of sweeping the electron beam in a cathode-ray tube to produce a spiral trace on the screen. **2.** The circuit for producing such a sweep.

spiral trace See SPIRAL SWEEP.

spiral winding See DISK WINDING.

spkr Abbreviation of *speaker* (see LOUDSPEAKER).

splatter See SIDEBAND SPLATTER.

splatter-suppression filter In an amplitude-modulated (AM) or single-sideband (SSB) radio transmitter, a low-pass filter inserted between the output of the audio amplifier and the audio input of the modulator. It suppresses high-frequency audio components that would otherwise cause sideband splatter.

splaying The construction of a room or auditorium so that the walls do not meet at right angles. It is useful in optimizing the acoustic characteristics of the enclosure because it tends to reduce acoustic resonant effects.

splice **1.** A physical or electrical connection between two wires, made by tightly winding the ends together. Solder is often applied for additional mechanical strength and electrical continuity. **2.** A physical connection between two lengths of magnetic tape, made in such a way as to cause minimal disturbance in reproduced audio, video, or data. **3.** To prepare a joint, as defined in **1** or **2**.

splicer A device for making a SPLICE.

splicing block A device specifically designed to facilitate easy splicing of audio, video, or digital magnetic tape.

splicing tape A durable, flexible adhesive designed to hold a magnetic-tape splice together.

spline-based modeling In video animation and advanced computer graphics, a scheme that uses curve-generating formulas (e.g., Bezier curves) to create lifelike images.

split-anode magnetron A magnetron in which the plate (anode) consists of two semicylinders with the cathode at their axis.

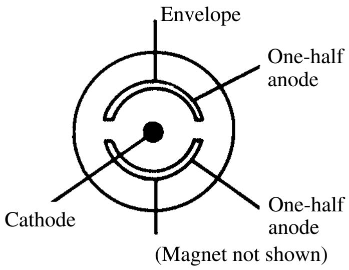

split-anode magnetron

split-load phase inverter See PARAPHASE INVERTER.

split-phase motor A fractional-horsepower, alternating-current motor that starts like a two-phase motor and runs like a single-phase motor. After the machine approaches approximately 75 percent of full speed, a centrifugal switch cuts out the starting winding.

split projector An acoustic transmission device with several independently operated transducers.

split-reed vibrator See SELF-RECTIFYING VIBRATOR.

split-rotor plate See SERRATED ROTOR PLATE.

split-stator capacitor A variable capacitor having two separate stator sections and a common rotor section.

split-sound receiver A television receiver circuit in which the sound signal is separated from the composite signal shortly after pickup by the antenna, and is amplified separately from the video signal. Compare INTERCARRIER RECEIVER.

splitter A device used to couple two or more television receivers to a single antenna.

spool See REEL.

sporadic-E layer ionization Occasional, scattered ionization in the E-layer of the ionosphere.

sporadic-E propagation At certain radio frequencies, the long-distance propagation of electromagnetic (EM) waves via the E layer of the ionosphere. This layer exists at an altitude of approximately 50 miles (80 km) above sea level. This mode of propagation tends to be intermittent, and conditions can change rapidly. It is most likely to occur between 20 MHz and 150 MHz. Occasionally, it is observed at frequencies as high as 200 MHz. The propagation range is normally several hundred miles, but occasionally can occur over distances of 1,000 to 1,500 miles. The standard frequency modulation (FM) broadcast band is sometimes affected by this propagation. The same is true of the lowest television (TV) broadcast channels, especially channels 2 and 3. See IONOSPHERE.

SPOT Abbreviation of *satellite position and tracking.*

spot brightness In a cathode-ray tube, the relative brilliance of the glowing dot or trace produced on the screen by the electron beam.

spot check **1.** To take a random sample or to make a random test by arbitrarily selecting a single item from a run of similar items and subjecting it to analysis, examination, performance, or parametric evaluation, etc. **2.** A random sample or test.

spot frequency **1.** A single frequency or signal to which other frequencies can be referred. **2.** A frequency or signal that acts as a limit marker (e.g., to define the edges of an allocated frequency band).

spot jamming Deliberate interference to a radio signal at some frequency and at some specific time.

spot modulation In a cathode-ray tube, modulation of the brightness of the spot (and, accordingly, of the image) produced on the screen by the electron beam. Also see INTENSITY MODULATION.

spot welding A method of electrical welding in which the parts to be joined are held together, overlapping, between the points of two electrodes, between which a current is passed to heat the parts at the spot of contact.

spot-wheel pattern A frequency-identifying wheel pattern produced on an oscilloscope screen by intensity-modulating a circular trace. The circular trace is produced by applying a standard-frequency signal to the horizontal and vertical input terminals 90 degrees out of phase. A square-wave signal of unknown frequency is applied to the intensity-modulation (*z*-axis) input terminals. The square wave chops the circle into a number of bright sectors or spots. The unknown frequency (*fx*) equals *Nfs*, where *N* is the number of spots in the circle, and *fs* is the standard frequency. Compare GEAR-WHEEL PATTERN.

sprat Acronym for *small portable radar torch.* A portable radar unit that uses a Gunn diode to generate microwave energy. The range is about 500 meters (⅓ mile).

spray coating **1.** Applying a protective coat of insulating material to a conductor or component by spraying it with a liquid substance and allowing it to dry. Compare DIP COATING. **2.** The coat applied, as defined in **1**.

spreader **1.** An insulator used to separate the wires of an air-spaced transmission line. **2.** Any of the rods composing the supporting structure in a cubical quad antenna.

spreading current In a semiconductor, current caused by the movement of charge carriers by circuitous routes, that is, in paths that deviate significantly from straight lines.

spreading loss Energy lost during the transmission of radiation.

spreading resistance In a semiconductor device, the resistance that is a consequence of electrical paths through material that is not along straight lines between electrodes.

spread spectrum **1.** A method of transmission in which the occupied bandwidth of the signal is deliberately increased, or spread out, over a much wider range than it would normally occupy with conventional modulation. **2.** A signal transmitted, as defined in **1**.

spring coil See SOLENOID, **1**.

spring contact See FLEXIBLE CONTACT.

sprite In video and animated computer graphics, a brief insert, such as a little insect that scurries across the screen, or a face that pops in and smiles. It is used primarily for effect.

SPST Abbreviation of SINGLE-POLE/SINGLE-THROW (switch or relay).

spurious emission From a radio or television transmitter, an unintended and unwanted output signal on a frequency other than the fundamental signal frequency. It can be generated by faulty modulation, amplification, and/or oscillation.

spurious oscillation **1.** Oscillation in a normally nonoscillatory circuit. **2.** In an oscillator, simultaneous oscillation at a frequency other than the normal one.

spurious response In a communications receiver, a signal that appears to be on a certain frequency, when, in fact, the received signal is not on that frequency. It often results from inadequate image rejection.

spurious-response ratio The ratio of the transmission (or gain) of a circuit of a desired signal to its transmission (or gain) for a spurious signal at the same setting of the circuit (e.g., *signal-to-image ratio*).

spurious sidebands In an amplitude-modulated (AM) or single-sideband (SSB) radio signal, sideband energy at frequencies outside the nominal signal band, usually resulting from improper modulation, inadequate filtering, improper envelope clipping, or nonlinear amplification.

Sputnik The first orbiting artificial earth satellite. It was launched by Russia (then known as the Union of Soviet Socialist Republics) in 1957.

sputter **1.** A layer of material obtained by sputtering (see SPUTTERING, **1**). **2.** To carry out the process of sputtering (see SPUTTERING, **1**).

sputtering **1.** A technique for electrically depositing a film of metal on a metallic or nonmetallic surface. In a vacuum chamber, the piece of metal to be deposited is made the cathode of a high-voltage circuit, with respect to a nearby anode plate. The high voltage causes atoms to be ejected from the surface of the cathode and strike the surface of an object placed in their path, becoming deposited on it as a film of cathode metal. Compare EVAPORATION, **1**. **2.** The disintegration of a vacuum-tube cathode through ejection of surface atoms from the cathode by impinging positive ions.

sq Abbreviation of SQUARE.

S_Q band A section of the S BAND, from 2400 to 2600 MHz.

SQC Abbreviation of STATISTICAL QUALITY CONTROL.

SQR **1.** Abbreviation of SQUARE ROOTER. **2.** In the BASIC computer-programming language, a function that computes the square root of a positive number.

S quad See SIMPLE QUAD.

square-law demodulator See SQUARE-LAW DETECTOR.

square-law detector A detector whose output is proportional to the square of the root-mean-square (rms) value of the input. Also called WEAK-SIGNAL DETECTOR.

square-law meter **1.** A meter whose deflection is proportional to the square of the quantity applied to it. Also see CURRENT-SQUARED METER. **2.** A high-impedance electronic voltmeter, whose deflection is proportional to the square of the applied voltage.

square-law modulator A circuit or device that accomplishes amplitude modulation of one signal current by another, by simultaneously passing the two currents through a component, such as a nonlinear resistor, having a square-law response.

square-law response Circuit or component operation that results in an output signal, proportional to the square of the input.

square rooter An analog or digital device used to extract the square root of a number.

square wave An alternating or pulsating current or voltage whose rise and decay times are essentially zero, and whose maxima and minima are essentially flat. The duration of the maxima is equal to the duration of the minima. A special form of RECTANGULAR WAVE.

square-wave amplifier An amplifier designed to operate with square waves.

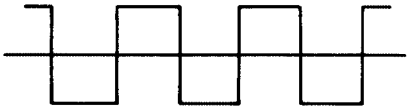

square wave

square-wave converter See SQUARING CIRCUIT.

square-wave generator A signal generator delivering an output signal that has a square waveform. Compare SQUARING CIRCUIT, **1**.

square-wave testing Testing the response of a circuit or device, such as an amplifier, by observing the extent to which it distorts a square-wave signal passing through it (a measure of high-frequency response).

squaring circuit **1.** A circuit (such as a twin-diode clipper, overdriven amplifier, or Schmitt trigger) that converts a sine wave or pulse into a square wave. **2.** A circuit whose instantaneous output-signal amplitude is equal to the square of the instantaneous input-signal amplitude.

squarish wave **1.** A signal whose oscilloscope trace is nearly, but not exactly, the same as that of a square wave. **2.** A rectangular wave that is not a square wave; that is, whose maxima and minima are not of the same duration. See RECTANGULAR WAVE and SQUARE WAVE.

squawker **1.** In a three-way speaker system, the midrange speaker. **2.** Any slave station in a multistation intercom network.

squeal A high-pitched interferential sound, such as that encountered in spuriously oscillating systems.

squeezeout In optical character recognition (OCR), a condition in which errors occur because the printed characters have excessive ink at the edges. Also called *smudge.*

squegging A choking-type cutoff action in a circuit caused by an excessively strong signal.

squegging oscillator An oscillator that starts and stops oscillating intermittently as a result of SQUEGGING.

squelch See SQUELCH CIRCUIT.

squelch amplifier An amplifier that can be controlled by a squelch signal. Also see SQUELCH CIRCUIT.

squelch circuit One of several circuits that automatically disable a receiver or amplifier, except when incoming signals exceed a predetermined threshold amplitude. This action mutes the equipment, eliminating annoying background noise and unwanted signals. Also called MUTING CIRCUIT.

squelch signal The activating or deactivating signal delivered by a SQUELCH CIRCUIT.

squiggle See BLIP, **2**.

squint **1.** The angular resolution of a radar antenna. **2.** The angular difference between the antenna axis and the major lobe of a radar transmitter.

squirrel-cage induction motor See SQUIRREL-CAGE MOTOR.

squirrel-cage motor An induction-type alternating-current motor using a squirrel-cage rotor.

squirrel-cage rotor In an alternating-current motor, a rotor composed of straight copper bars embedded in a laminated soft-iron core and short-circuited at the ends by rings. Its name is derived from its resemblance to a revolving squirrel cage.

squirter See SIGNAL INJECTOR.

SR **1.** Abbreviation of SILICON RECTIFIER. **2.** Abbreviation of *silicone rubber* (see SILICONE). **3**. Abbreviation of SHIFT REGISTER.

S-R Abbreviation of SEND-RECEIVE.

Sr Symbol for STRONTIUM.

sr Abbreviation of STERADIAN.

SRAM Abbreviation of *static random-access memory.*

S-rays See SECONDARY RAYS.

SRF Abbreviation of SELF-RESONANT FREQUENCY.

S-RF meter A dual-function meter in a radio transceiver. In the receiving mode, the meter indicates S units. In the transmit mode, the meter indicates relative output power.

SRR Abbreviation of SHORT-RANGE RADAR.

SS **1.** Abbreviation of SOLID STATE. **2.** Abbreviation of SINGLE SHOT. **3.** Abbreviation of *small signal.* **4.** Abbreviation of SINGLE SIGNAL. **5.** Abbreviation of *same size.* **6.** Abbreviation of *stainless steel.*

SSB Abbreviation of SINGLE SIDEBAND.

S_S band A section of the S BAND extending from 2900 to 3100 MHz.

SSBSC Abbreviation of SINGLE-SIDEBAND SUPPRESSED CARRIER.

ssc Abbreviation of *single-silk-covered* (wire).

S scale A scale of numbers used in radio communications, and especially in amateur radio, to report the approximate strength of signals: S1, faint signals; S2, very weak signals; S3, weak signals; S4, fair signals; S5, fairly good signals; S6, good signals; S7, moderately strong signals; S8, strong signals; S9, extremely strong signals. Also see S METER.

sse Abbreviation of *single-silk-enameled* (wire).

SSI Abbreviation of SMALL-SCALE INTEGRATION.

SSL Abbreviation of SOLID-STATE LAMP.

SSSC Abbreviation of SINGLE-SIDEBAND SUPPRESSED CARRIER. (Also, SSBSC.)

ST Abbreviation of SINGLE THROW.

sta **1.** Abbreviation of STATION. (Also *stn.*) **2.** Abbreviation of STATIONARY.

stab Abbreviated form of stabilizer (see STABILIZER, 4).

stability **1.** The condition in which an equipment or device is able to maintain a particular mode of operation without deviation. **2.** The condition in which the setting or adjustment of a device remains at a particular point without movement. **3.** The condition in which a quantity remains constant, with respect to time, temperature, or another variable. **4.** The ability of inks used in optical character recognition to retain their color after exposure to light or heat.

stability factor Abbreviation, SF. For a bipolar transistor, the derivative dI_c/dI_{co}, where I_c is the steady-state collector current and I_{co} is the collector cutoff current.

stabilized platform See STABLE PLATFORM.

stabilizer **1.** See DAMPING DIODE. **2.** See DAMPING RESISTOR. **3.** A device or circuit for the self-regulation of current or voltage. **4.** A chemical used to control or arrest a reaction.

stabilizing windings Auxiliary field windings used to prevent speed runaway in shunt motors.

Stabistor Trade name for a type of voltage-regulating semiconductor diode.

stable device A device whose characteristics and performance remain substantially unchanged with time or variations in temperature, applied power, or other quantities.

stable element **1.** A component that maintains its value or ratings, despite widely variable environmental conditions. **2.** A navigational instrument that maintains its orientation at all times.

stable platform A gyro-type device used to stabilize objects in space, and to provide accurate information regarding attitude (pitch, roll, and yaw).

stable state A stable condition, such as the high and low states of a flip-flop. The flip-flop has two stable states and will remain in one until it is switched to the other, whereupon it will then remain in that latter state until switched back to the former. Compare UNSTABLE STATE.

stack **1.** A piled assembly of capacitor plates and separating dielectric films. **2.** An assembly of selenium rectifier plates (see POWER STACK). **3.** To assemble a stacked array. **4.** A temporary storage area consisting of a small group of registers in a computer memory.

stacked array An antenna system in which two or more identical antennas, such as dipoles, Yagis, or halos, are placed one above the other or side-by-side. It provides additional forward gain, and, in some cases, enhances the front-to-back ratio and/or front-to-side ratio.

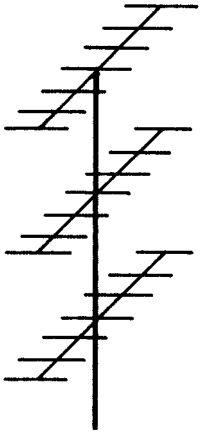

stacked array

stacked-dipole antenna A stacked array of half-wave dipole antennas.

stacking The combination of two or more identical antennas, such as dipoles, Yagis, or halos, in a STACKED ARRAY to provide enhanced forward gain. It can also enhance the front-to-back ratio and/or front-to-side ratio. Stacking can be done vertically or horizontally.

stack pointer Abbreviation, SP. A register indicating the last data item to be entered in a stack (see STACK, 4).

stage A complete functional unit of a system (e.g., *amplifier stage*, *oscillator stage*, *modulator stage*, etc.).

stage-by-stage elimination See SIGNAL INJECTION.

stage gain The amplification provided by a single stage in a system.

stage loss The loss introduced by a single stage in a system.

stagger **1.** An error in facsimile reception, occurring as a constant discrepancy in the position of the received dot. **2.** To deliberately tune a set of resonant circuits, especially in a bandpass filter, to one side or the other of the center frequency.

staggered tuning The tuning of the input and output circuits of a single stage, or the tuning of successive stages to slightly different frequencies to obtain flat-top response.

stagger tuning See STAGGERED TUNING.

stagger-wound coil See BASKET-WEAVE COIL and SPIDERWEB COIL.

staircase circuit See STAIR-STEP CIRCUIT.

staircase generator A circuit or device for generating a STAIR-STEP WAVE.

staircase wave See STAIR-STEP WAVE.

stair-step circuit A circuit that converts a series of equal-amplitude pulses into a stair-step wave.

stair-step wave A nonsinusoidal wave characterized by a multistep rise and a steep fall. It is so called from its resemblance to the cross section of a staircase.

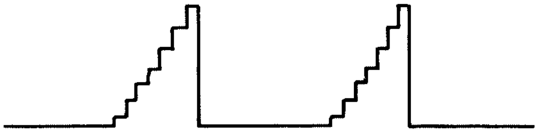

stair-step wave

stall torque The torque produced when a motor shaft is prevented from turning.

stalo Acronym for *stabilized oscillator.*

stand-alone photovoltaic system A solar-power plant that uses large banks of rechargeable electrochemical batteries, such as the lead–acid type, to store electric energy as it is supplied by photovoltaics during hours of bright sunshine. The energy is released by the batteries at night or in gloomy daytime weather. This system does not depend on the electric utility companies. Although this scheme offers independence from the utility companies, a blackout will occur if the system goes down. Compare INTERACTIVE PHOTOVOLTAIC SYSTEM.

standard **1.** A precise specification governing the dimensions and characteristics of a device or system (e.g., *military standard*). **2.** A highly accurate physical or electrical quantity to which similar quantities can be compared (e.g., *standard frequency*). **3.** The device or system that produces a standard quantity as defined in **2** (e.g., *frequency standard*). **4.** Having conventional and widely accepted characteristics.

standard atmosphere Abbreviation, atm. Air pressure at sea level (1.013 Pascals, or about 14.7 pounds per square inch). Also called ATMOSPHERE.

standard broadcast band Any of numerous frequency bands allocated to conventional broadcast stations. In the United States, the amplitude-modulation (AM) radio broadcast band extends from 535 to 1705 kHz, and the frequency-modulation (FM) radio broadcast band extends from 88 to 108 MHz. The television (TV) broadcast bands range from 54 MHz to 806 MHz in several sections, designated in channels from 2 through 69. Also see BROADCAST SERVICE, **1**.

standard candle See CANDELA.

standard cell A highly refined primary cell used to supply a precise direct-current voltage for electronic measurements. The *Weston standard cell* contains a mercury positive electrode, cadmium amalgam negative electrode, and cadmium sulfate electrolyte, and delivers 1.0183 volts at 20 degrees Celsius. Also see ZINC STANDARD CELL.

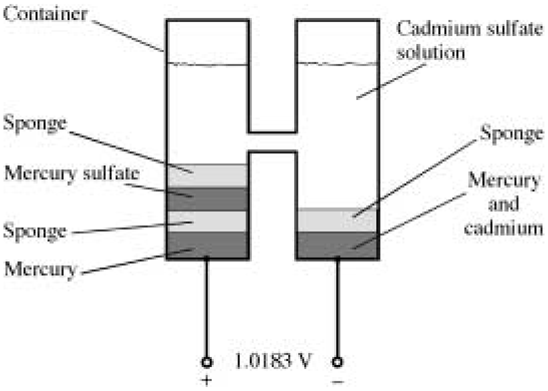

standard cell

standard deviation In statistical analysis, the square root of the mean of squares of deviation from the mean.

standard frequency A highly precise frequency to which other frequencies can be compared for identification or measurement.

standard-frequency oscillator A stable, precise oscillator that delivers a standard frequency. Also see PRIMARY FREQUENCY STANDARD and SECONDARY FREQUENCY STANDARD.

standard-frequency signal A calibration and reference signal that is broadcast on a standard frequency, such as those transmitted on 2.5, 5, 10, and 15 MHz by the National Bureau of Standards.

standard pitch The tone corresponding to the frequency 440 Hz (in music, the note A above middle C).

standard signal generator A (usually continuously variable) high-grade generator of modulated and unmodulated radio-frequency test signals. A general-purpose instrument of this type usually covers a wide range (e.g., 15 kHz to 100 MHz) in several tuning bands. For calibration, a standard signal generator is referred to a primary frequency standard or secondary frequency standard.

standard subroutine A usually vendor-supplied computer program segment applicable to more than one program and used as needed as a subroutine.

standard temperature and pressure The condition where the temperature is zero degrees Celsius and the pressure is one atmosphere. Abbreviated STP.

standard time Official civil time in a particular region. See TIME ZONE.

standby **1.** The state in which equipment is out of operation, but can be immediately activated. Also called IDLING. **2.** A state of readiness on the part of personnel, equipment, or systems.

standby battery An emergency power source for a battery-powered installation.

standby current The CURRENT DRAIN of a circuit, device, or system when in the standby condition.

standby equipment See EMERGENCY EQUIPMENT.

standby operation Keep-alive operation during a standby interval (see STANDBY).

standby power The power drawn by an equipment connected to the power supply, but out of operation.

standby power supply A circuit containing a battery, an automatic switch, and sometimes a power inverter. When utility power fails, the switch actuates the supply, and the battery supplies power to essential devices or systems. Similar to an UNINTERRUPTIBLE POWER SUPPLY.

standing wave A stationary distribution of current or voltage along a line because of the interactions between a wave transmitted down the line and a wave reflected back; it is characterized by maximum-amplitude points (loops) and minimum-amplitude or zero points (nodes).

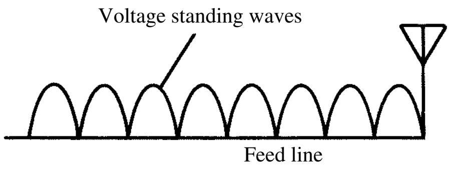

standing wave

standing-wave distortion Distortion of current or voltage caused by standing waves on a transmission line terminated in an impedance that contains reactance, and/or that differs from the characteristic impedance of the line.

standing-wave indicator **1.** A device, such as a lamp or meter, used to detect standing waves. **2.** Standing-wave meter (see SWR BRIDGE).

standing-wave loss The additional loss, over the matched-line loss, that occurs in a transmission line when the standing-wave ratio (SWR) is not 1.

standing-wave meter See SWR BRIDGE.

standing-wave ratio Abbreviation, SWR. **1.** The ratio between the maximum and minimum voltage along a transmission line. This quantity is sometimes specifically called *voltage standing-wave ratio (VSWR).* **2.** The ratio between the maximum and minimum current along a transmission line. **3.** The ratio of load impedance to feed-line characteristic impedance or vice versa, whichever is greater than or equal to 1. Ideally, the SWR is equal to 1 or 1:1, representing a load impedance that is purely resistive and has the same value as the characteristic impedance of the feed line. A high standing-wave ratio causes increased loss in the line and can also result in excessive conductor heating or dielectric breakdown.

standoff insulator An insulator (usually of the post type) that is used to hold a wire or component away from a chassis or base.

star **1.** In a gravity-battery cell, the copper electrode. The name is derived from its star shape. **2.** A star-shaped circuit of three-phase components. Also see WYE CONNECTION.

star connection See WYE CONNECTION.

Stark effect The influence of a strong transverse electric field on the spectrum lines of a gas.

starlight scope A device capable of viewing in apparent total darkness. Its operation depends on its ability to provide high amplification of extremely low light levels, such as that of objects reflecting the light from a moonless, but starlit, sky.

star rectifier See WYE RECTIFIER.

starter **1.** An ignitor electrode in an ignitron (see IGNITOR). **2.** See STARTING BOX.

starting box A special rheostat for starting a motor gradually in steps. The device is provided with an electromagnet for holding the arm in the maximum-speed position and releasing it when power is interrupted.

starting rod An ignitor electrode in an ignitron (see IGNITOR).

starting voltage **1.** For a gas tube, the minimum voltage that will initiate the glow discharge. **2.** In appropriate solid-state devices (e.g., a diac), the voltage at which conduction between electrodes occurs.

start lead The lead attached to the first turn of a coil. Also called *inside lead.* Compare FINISH LEAD.

start/stop multivibrator See MONOSTABLE MULTIVIBRATOR.

stat- A prefix denoting ELECTROSTATIC.

statampere The cgs electrostatic unit of current; 1 statampere = 3.335640×10^{-10} ampere.

statcoulomb The cgs electrostatic unit of charge; 1 statcoulomb = 3.335640×10^{-10} coulomb.

state **1.** The present condition (i.e., on or off, true or false, 1 or 0, high or low) of a bistable device, such as a flip-flop. **2.** The physical or electrial condition or status of a component, device, circuit, or system.

statement The contents of a line in a source-language computer program.

state of charge The amount of charge, measured in coulombs or ampere hours, in a storage cell or battery at a given time. A measure of the available remaining energy in the cell or battery.

statfarad The cgs electrostatic unit of capacitance; 1 statfarad = 1.112650×10^{-12} farad.

stathenry The cgs electrostatic unit of inductance; 1 stathenry = 8.987554×10^{11} henry.

static **1.** Pertaining to that which is constant in quantity (e.g., *static collector current* of a transistor). **2.** Pertaining to that which is at rest (e.g., *static electricity*). **3.** The radio noise (sferics) produced by electric discharges in the atmosphere, usually lightning. **4.** Pertaining to a test-and-measurement mode for a unit or device, without subjecting the unit or device to regular operation. Compare DYNAMIC.

static base current See DC BASE CURRENT.

static base resistance See DC BASE RESISTANCE.

static base voltage See DC BASE VOLTAGE.

static cathode current See DC CATHODE CURRENT.

static cathode resistance See DC CATHODE RESISTANCE.

static cathode voltage See DC CATHODE VOLTAGE.

static characteristic An operating characteristic determined from constant, rather than alternating or fluctuating, values of independent and dependent variables. Examples: the direct-current (dc) characteristics of diodes, transistors, and integrated circuits. Compare DYNAMIC CHARACTERISTIC.

static charge Energy stored in a stationary electric field; electricity at rest.

static collector A device that grounds the rotating wheels of a motor vehicle, thereby removing the static electricity generated by the friction of the tires on the roadway.

static collector current See DC COLLECTOR CURRENT.

static collector resistance See DC COLLECTOR RESISTANCE.

static collector voltage See DC COLLECTOR VOLTAGE.

static convergence In a color-television picture tube, the convergence of the three undeflected electron beams at the center of the aperture mask.

static device A device with no moving parts.

static discharge resistor A fixed resistor connected between the earth and the high side of the power line in a television receiver to drain off atmospheric electric charge.

static drain current See DC DRAIN CURRENT.

static drain resistance See DC DRAIN RESISTANCE.

static drain voltage See DC DRAIN VOLTAGE.

static dump In computer operations, a dump occurring at a predetermined point in a program run or at the end of the run.

static electricity Energy in the form of a stationary electric charge, such as that stored in capacitors or produced by friction or induction.

static emitter current See DC EMITTER CURRENT.

static emitter resistance See DC EMITTER RESISTANCE.

static emitter voltage See DC EMITTER VOLTAGE.

static flip-flop A flip-flop (see BISTABLE MULTIVIBRATOR) using direct-current operating voltages. A single pulse switches the unit from on to off, and vice versa. Compare DYNAMIC FLIP-FLOP.

static forward current transfer ratio Symbol, H_{FE}. An expression of gain in a bipolar transistor. It can range from a factor of just a few times up to hundreds of times. Mathematically,

$$H_{FE} = I_C / I_B$$

where I_C is the collector current and I_B is the base current. The H_{FE} rating is important because it gives engineers an indication of the greatest current amplification that can be obtained with a particular transistor.

static frequency multiplier A magnetic-core device, similar to a magnetic amplifier or peaking transformer, that provides harmonics by distorting a sine-wave signal.

static gate current See DC GATE CURRENT.

static gate resistance See DC GATE RESISTANCE.

static gate voltage See DC GATE VOLTAGE.

static grid current See DC GRID CURRENT.

static grid voltage See DC GRID VOLTAGE.

static hysteresis The condition in which the magnetization of a material (when it has the same intensity as the magnetizing force) is different when the force is increasing than when the force is decreasing, regardless of the time lag. Compare VISCOUS HYSTERESIS.

static induction See ELECTROSTATIC INDUCTION.

static machine See ELECTROSTATIC GENERATOR.

static memory Also called *nonvolatile memory*. In a computer, a data memory medium (such as programmable read-only memory, or PROM) in which information is stored until it is altered or erased. It does not require a source of power to maintain the integrity of the data. Compare VOLATILE MEMORY.

static mutual conductance See STATIC TRANSCONDUCTANCE.

static plate current See DC PLATE CURRENT.

static plate resistance See DC PLATE RESISTANCE.

static plate voltage See DC PLATE VOLTAGE.

statics The study of forces, bodies, poles, charges, or fields at rest or in equilibrium. Compare DYNAMICS.

static skew In magnetic tape recording or playback, the amount of lead or lag time of one track, with respect to another. Ideally, the static skew should be zero or practically zero.

static source current See DC SOURCE CURRENT.

static source resistance See DC SOURCE RESISTANCE.

static source voltage See DC SOURCE VOLTAGE.

static storage Also called *nonvolatile storage.* In a computer, a data storage medium (such as magnetic or optical disk) in which information is stored until it is altered or erased. It does not require a source of power to maintain the integrity of the data. Virtually all data storage media are of this type, as contrasted with memory, which is often volatile (see STATIC MEMORY and VOLATILE MEMORY).

static stability The ability of a robot to maintain its balance while standing still. A robot with two legs is generally poor in this respect. This is one of the reasons why humanoid robots (androids) are difficult to engineer. A minimum of three legs is necessary for good static stability.

static subroutine In computer programming, a subroutine that always serves the same purpose [i.e., it does not need to be tailored (according to parameters) for a specific application].

station **1.** An installation consisting of a transmitter, receiver, or both. **2.** A test-equipment installation or position. **3.** A computer installation including peripherals.

stationary battery A (usually wet storage) battery not normally moved when in use.

stationary state A particular energy state for an atom represented by its electrons being in shells at specific energy levels.

stationary wave See STANDING WAVE.

station authorization The legal privilege assigned to a broadcast or communications station, allowing that station to be used for the purpose of transmitting electromagnetic signals.

station license See STATION AUTHORIZATION.

statistical quality control Quality control based upon the techniques of probability and statistics in analyzing findings, making predictions, and formulating procedures for sampling.

statmho The cgs electrostatic unit of direct-current conductance; 1 statmho = 1.112650 × 10^{-12} siemens.

statoersted The cgs electrostatic unit of magnetizing force; 1 statoersted = 265.458 A/m (3.33585 × 10^{-11} oersted).

statohm The cgs electrostatic unit of direct-current resistance; 1 statohm = 8.987554 × 10^{11} ohm.

stator **1.** A stationary coil. Compare ROTOR, **1**. **2.** The stationary member of a motor or generator. Compare ROTOR, **2**. **3.** The stationary-plate section of a variable capacitor. Compare ROTOR, **3**.

stator coil A stationary coil (see STATOR, **1**, **2**).

stator plate(s) The stationary plate(s) of a variable capacitor. Compare ROTOR PLATE.

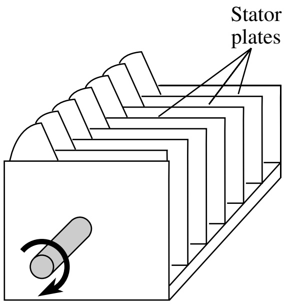

stator plate

stator section See STATOR, **3**.

statoscope An aircraft altimeter that shows small changes in altitude.

statvolt The cgs electrostatic unit of electromotive force; 1 statvolt = 299.7925 volts.

statweber The cgs electrostatic unit of magnetic flux; 1 statweber = 299.7925 webers (2.997925 × 10^{10} maxwells).

S_T band A section of the S BAND, extending from 1850 to 2000 MHz.

std Abbreviation of STANDARD.

steady-state component A quantity whose value remains constant during normal operation of a circuit or device, as opposed to an alternating, fluctuating, or transient component.

steerable antenna A directional antenna having a rotatable major lobe.

steering diode See DIRECTIONAL DIODE.

Stefan-Boltzmann constant Value, 5.67051 × 10^{-8} $Wm^{-2}K^{-4}$.

Stefan-Boltzmann law The thermal-radiation law that shows the total emissive power of a blackbody to be proportional to the fourth power of the absolute temperature of the body.

stenode See CRYSTAL FILTER.

step **1.** A computer program instruction. **2.** A single action in the operation, maintenance, or troubleshooting of equipment. **3.** A specific increment in a quantity (such as frequency, voltage, current, etc.). **4.** A sharp or rapid change in the value of a quantity.

step-by-step operation See STEPTHROUGH OPERATION.

step change A single-increment change in a value.

step circuit A circuit that produces a step (sharp change of slope) in the response curve of an amplifier.

step counter **1.** A stair-step circuit arranged to count input pulses. The output capacitor of the circuit discharges when a predetermined number of input pulses has raised the capacitor voltage to the level required to trigger a counter. **2.** In a computer or calculator, a circuit or device that counts the steps in an operation (such as division, multiplication, or shifting) called for by an instruction.

step-down ratio In a circuit or device, such as a step-down transformer or cathode follower, the ratio of the low output voltage to the high input voltage. Compare STEP-UP RATIO.

step-down transformer A transformer delivering an output voltage that is lower than the input voltage. In such a transformer, the secondary (output) winding contains fewer turns than the primary (input) winding. Compare STEP-UP TRANSFORMER.

step function See UNIT FUNCTION.

step generator **1.** A signal generator that delivers a step function (see UNIT FUNCTION). **2.** A circuit or device that generates a STAIR-STEP WAVE.

stepped leader The probing flow of electrons through the atmosphere preceding a lightning stroke. Once the path has been established, the discharge takes occurs along the ionized path determined by the stepped leader. It is so called because the electrons move in hesitations, jumping several meters with each advance or step.

stepper motor A motor in which the shaft advances in uniform angular steps, instead of rotating continuously. These motors are extensively used in robotic devices. When such a motor is stopped and its coils are carrying current, the shaft resists turning.

stepping relay See STEPPING SWITCH.

stepping switch A multiposition rotary switch in which an electromechanical ratchet mechanism advances to the next contact position each time that a pulse of current is received.

step-through operation A way of operating a computer, usually during a debugging operation, in which program instructions are executed one at a time by direction of the user. Also called *single-step operation* and *step-by-step operation.*

step-up ratio In a circuit or device, such as a step-up transformer or voltage amplifier, the ratio of the high output voltage to the low input voltage. Compare STEP-DOWN RATIO.

step-up transformer A transformer delivering an output voltage that is higher than the input voltage. In such a transformer, the secondary (output) winding contains more turns than the primary (input) winding. Compare STEP-DOWN TRANSFORMER.

steradian A unit of solid-angle measure. A cone-shaped solid angle that has a vertex at the center of a sphere (of radius r), that cuts off a portion of the sphere's surface whose outer perimeter is a circle, and that has an area (as measured on the sphere's surface) of r^2. Also see SOLID ANGLE.

Sterba array See BARRAGE ARRAY.

Sterba curtain See BARRAGE ARRAY.

stereo **1.** Contraction of STEREOPHONIC. **2.** General term for a two-channel high-fidelity audio reproduction system.

stereo-adaptable Pertaining to a television receiver or videocassette recorder (VCR) that requires a special circuit to obtain stereo sound from Multichannel Television Sound signals. Compare STEREO-READY.

stereo amplifier A two-channel amplifier for binaural reproduction (see BINAURAL).

stereophonic Pertaining to equipment or techniques for producing a (somewhat) three-dimensional perspective of sound reproduction.

stereophonic sound system See STEREO SYSTEM.

stereo-irrelevant Pertaining to sound components in a stereo system that are of equal magnitude in both (or all) channels. Thus, these components sound the same whether the system is reproducing stereophonic sound or monaural sound.

stereo phono cartridge A phono cartridge capable of reproducing sound from stereo discs.

stereo-ready Pertaining to a television receiver or videocassette recorder (VCR) that can deliver stereo sound from Multichannel Television Sound signals without the need for a decoding circuit. Compare STEREO-ADAPTABLE.

stereo recording A method of recording in which two independent sound channels are transferred to some medium simultaneously, with the intention that the two channels be reproduced at the same time.

stereoscopic television Television in which the reproduced image appears three-dimensional.

stereo system A multichannel, high-fidelity sound reproduction system including an amplifier and various other components, such as a radio receiver, compact-disc (CD) player, tape player, turntable, and speakers.

stereotape Magnetic tape bearing more than one channel (usually two channels) for the recording and reproduction of stereophonic sound.

sterilizer Any electronic device, such as an ultraviolet generator, used to kill germs.

stethoscope An electronic or nonelectronic instrument used by physicians to listen to the heartbeat and other body sounds, and by technicians to listen to mechanical sounds.

still **1.** A stationary picture on television. **2.** A picture transmitted or received by means of facsimile. **3.** A print on photographic paper of a negative.

still television See FACSIMILE.

stinger A brief, loud sound burst, such as a musical chord, sometimes used for effect in recorded audio or audio-visual presentations.

stn Abbreviation of STATION. (Also, sta.)

STO Abbreviation of STORAGE FUNCTION.

stochastic The condition in which, at any instant, a given variable can assume a state dependent on previous states, as well as chance elements (e.g., words uttered extemporaneously by a speaker are random in that they cannot be predicted), while also being dependent, by the application of grammar, on previously spoken words (i.e., previous states).

stop amplifier See REJECT AMPLIFIER.

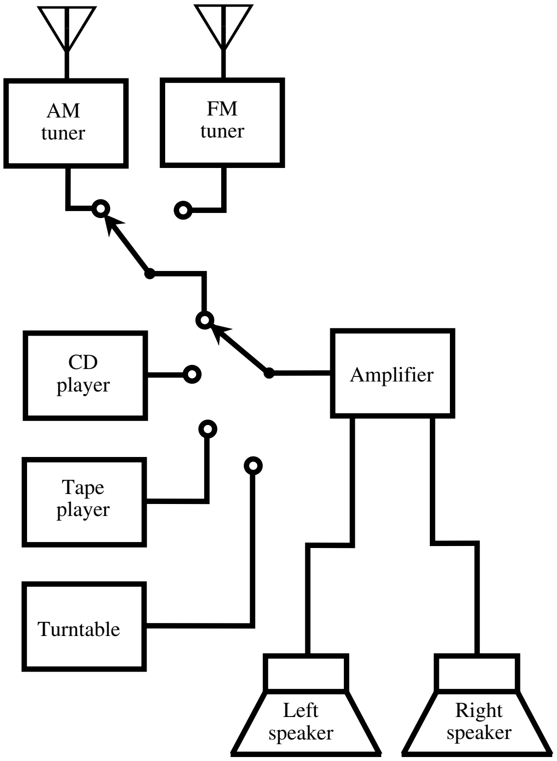

stereo system

stopband The continuous spectrum of frequencies rejected by a filter, selective amplifier, or other band-suppression device.

stopband ripple Single or multiple attenuation peaks within the stopband of an elliptic filter, where the frequency of the attenuation peak(s) is/are associated with the resonant circuit(s) in the filter.

stop code See HALT INSTRUCTION.

stop element In digital transmission, a bit or set of bits indicating the end of a character, and serving to inform the receiving device of the end of the character.

stop filter See BAND-SUPPRESSION FILTER.

stopper resistor See STOPPING RESISTOR.

stopping capacitor See BLOCKING CAPACITOR.

stopping coil See CHOKE COIL.

stopping resistor A parasitic-suppressing resistor, usually inserted in series with the input and/or output of a power amplifying device. Also called *stopper resistor.*

storage **1.** In computer operations, a medium on which data can be kept for an extended period of time. Examples: magnetic disk, magnetic tape, optical disk. **2.** The transferring of data from memory to a more permanent medium. Compare MEMORY. **3.** The retention of data of any kind, such as an oscilloscope image or video display, for use at a later time. **4.** The retention of electric energy or charge, as in a capacitor or electrochemical cell. **5.** The retention of energy in the form of a magnetic field, as in an inductor. **6.** The retention of potential energy in any form.

storage allocation The assignment of computer memory areas to certain kinds of information, as outlined in a source program and implemented by a compiler.

storage battery A rechargeable battery; the technical term is *secondary battery.* Also see STORAGE CELL.

storage capacity **1.** The amount of data that can be stored in a specific medium, such as a hard disk, diskette, or tape. Generally measured in bytes, kilobytes, megabytes, gigabytes, or terabytes. **2.** The energy-delivering capability of a storage battery in terms of ampere-hours, milliampere-hours, or other current-time units for a specific rated voltage.

storage cell **1.** An electrochemical cell whose potential can be restored by charging it with electricity. Compare PRIMARY CELL. Also see CELL. **2.** The smallest part of a computer storage medium. **3.** In computer data storage, the part that can hold a data unit (e.g., a bit).

storage CRT See STORAGE TUBE.

storage cycle In computer operation, the period during which a location of a cyclic storage device cannot be accessed.

storage density The number of data units (e.g., bytes, kilobytes, or megabytes) that can be stored in a given length or area of a storage medium.

storage device A medium into which data can be placed and kept for later use. Examples: magnetic disk, magneto-optical disk, optical disk, and magnetic tape.

storage dump See DUMP.

storage function Abbreviation, STO. The so-called "memory" function of a microcomputer chip. It causes data to be inserted into memory or storage for later use. It is commonly used in programmable calculators, automatic-dialing telephones, and radio receivers. Also called *memory function.*

storage laser A laser that stores intense energy before flashing.

storage life See SHELF LIFE.

storage mesh In the cathode-ray tube (CRT) of a storage oscilloscope, a fine metal mesh electrode that serves as the target on which the image is electrostatically stored. Also see STORAGE TUBE.

storage oscilloscope An oscilloscope that retains a displayed image until the display is erased. Also see STORAGE TUBE.

storage register In computers and calculators, a storage unit composed of flip-flops. In computers, it is independent of the central processing unit (CPU).

storage temperature **1.** The recommended temperature for storing specified electronic components. **2.** The particular temperature at which electronic components have been stored.

storage time **1.** The interval during which carriers remain in a semiconductor-junction device after the bias has been removed. Also see DIODE RECOVERY TIME. **2.** For a switching semiconductor device, the time required for the amplitude of the output pulse to fall from maximum to 90% of maximum after the input pulse has fallen to zero. **3.** In a computer, the time required for data to be transferred from random-access memory (RAM) to nonvolatile storage (e.g., hard disk).

storage tube A cathode-ray tube that retains information in the form of images on a special electrode until erased by a signal.

store **1.** To place data in a nonvolatile medium (such as a hard disk, diskette, optical disk, magnetic tape, etc.). **2.** To place in the memory of a calculator or computer. **3.** In computing, a command that causes data to be placed in a nonvolatile medium. In some applications, this is called *save.* **4.** A nonvolatile medium on which data has been placed for future use or for archival purposes.

stored base charge The carriers that remain in the base layer of a bipolar transistor immediately after the forward bias has been interrupted. This charge maintains collector current momentarily.

stored-energy welding A method of electric welding in which electrical energy is stored slowly, then released at the rate required for the welding.

STP See STANDARD TEMPERATURE AND PRESSURE.

straight adapter An inline coaxial fitting for joining two fixture-terminated coaxial lines in series.

straight angle An angle measuring 180 degrees.

straight dipole A (usually center-fed) dipole antenna having only one radiator. Also see DIPOLE ANTENNA.

straightforward Pertaining to data transmission in one direction only.

straight-gun CRT A cathode-ray tube (CRT) in which the electron gun projects the beam in a straight line through the deflecting elements to the screen. Compare BENT-GUN CRT.

straight-line capacitance Abbreviation, SLC. Pertaining to a variable capacitor for which the setting-vs.-capacitance curve is a straight line; the capacitance variation is linear. Compare STRAIGHT-LINE FREQUENCY and STRAIGHT-LINE WAVELENGTH.

straight-line coding See STRAIGHT-LINE PROGRAMMING.

straight-line frequency Abbreviation, SLF. Pertaining to a variable capacitor in a tuned circuit for which the setting-vs.-frequency curve is a straight line; the frequency variation is linear. Compare STRAIGHT-LINE CAPACITANCE and STRAIGHT-LINE WAVELENGTH.

straight-line programming During the writing of a computer program, avoiding the creation of loops by repeating a series of instructions to reduce execution time.

straight-line tracking In a phonograph turntable, linear lateral stylus movement (as opposed to motion along an arc) as the disc is played. This ensures that the stylus is always at the optimum angle in the disc groove. The result is improved sound reproduction, and longer disc and stylus life because of minimal friction between the stylus and groove.

straight-line wavelength Abbreviation, SLW. Pertaining to a variable capacitor in a tuned circuit for which the setting-vs.-wavelength curve is a straight line; the wavelength variation is linear. Compare STRAIGHTLINE CAPACITANCE and STRAIGHT-LINE FREQUENCY.

straight-through amplifier An amplifier in which the input and output circuits are tuned to the same frequency. Compare MULTIPLIER AMPLIFIER.

strain A force that compresses or squeezes a body. Compare TENSION, **1**.

strain gauge See ELECTRIC STRAIN GAUGE.

strain-gauge bridge A four-arm resistance bridge in which an ELECTRIC STRAIN GAUGE forms one arm. The resistance of the gauge changes because of strain. The amount of strain can be determined by balancing the bridge.

strain-gauge transducer A transducer, other than a strain sensor, that uses strain gauges to convert values of pressure into their electrical analogs (e.g., *pressure transducer* and *strain-gauge phonograph pickup*).

strain pickup A phonograph pickup using a strain gauge to convert sound vibrations into a varying electric current.

strand A single solid conductor in a STRANDED WIRE.

stranded wire A conductor composed of several non-insulated wires twisted together to provide mechanical flexibility. Compare SOLID WIRE.

stratosphere The portion of earth's upper atmosphere beginning at a height of approximately 10 miles and extending to the ionosphere.

stray capacitance Inherent capacitance in a place where it can be detrimental, such as that between the turns of a coil or between adjacent areas in a circuit. Also see STRAY COMPONENT.

stray component An electrical property that exists as an inherent, and usually undesirable, side effect in a circuit or device. Thus, for example, STRAY CAPACITANCE unavoidably exists between parallel conductors, and STRAY INDUCTANCE is present in all wiring.

stray field The portion of an electric or magnetic field that extends beyond the immediate vicinity of the circuit with which it is associated, and which is, therefore, capable of interfering with other circuits or devices.

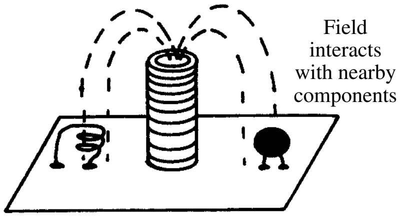

stray field

stray inductance Inherent inductance in a place where it can be detrimental (e.g., inductance in the coil of a wirewound resistor). Also see STRAY COMPONENT.

stray resistance Inherent resistance in a place where it can be detrimental, such as leakage resistance in a dielectric, and wire resistance in an inductor.

streaking In television or facsimile, a form of distortion in which the image appears enlarged in the horizontal, but not the vertical.

stress **1.** See STRAIN. **2.** See TENSION, **1**. **3.** The force per unit area that produces STRAIN or TENSION on a body.

stretch The amount by which a material being measured with an electronic device increases in surface dimensions. Compare SHRINK.

stretched string A long, thin wire or string that vibrates at a certain frequency, causing standing waves. It generally exhibits a specific fundamental frequency and integral harmonics of this frequency. As the wavelength is cut in half, the frequency doubles.

strike To initiate a discharge, as in striking a gas tube.

striking voltage See STARTING VOLTAGE.

string **1.** In computer operations, a set of items in a sequence determined by the order of keys. **2.** In a computer memory, a sequence of bits or characters. **3.** Any group of series-connected components or circuits.

string electrometer See BIFILAR ELECTROMETER.

string variable A string of characters, usually forming a word or phrase, represented by a variable name and character string symbol (BASIC's $, for example) in a computer program.

strip chart A longitudinal, as opposed to circular, chart for graphic recording. In a rectilinear chart, both coordinates are straight; in a curvilinear chart, the crosswise coordinates are arcs.

strip core A ferromagnetic core material made from a strip of the substance. The method of manufacture results in superior ferromagnetic qualities, but also imparts a polarization to the material.

strip fuse A fuse in which the fusible element is a flat strip of low-melting-point metal. Compare WIRE FUSE.

strobe **1.** See ELECTRONIC FLASH, **1**. **2.** See STROBOSCOPE.

strobe light **1.** See ELECTRONIC FLASH, **1**. **2.** See STROBOSCOPE.

Strobolume Trade name for a type of high-output stroboscope.

stroboscope **1.** An instrument that emits bright, adjustable-rate flashes of light. When this light illuminates an object that is rotating or vibrating at a fixed period, and the flash rate is made to match that period, the object seems to stand still and can be examined for flaws or faulty operation (and its speed can be measured). **2.** A rotatable, slotted disk for producing the effect defined in **1**.

stroboscopic disk A rotatable disk with alternating white and black radial regions, used in conjunction with a strobe light for precise measurement of the speed of a phonograph turntable.

strobotron A gas tetrode tube used as the flashing light source in a stroboscope.

stroke speed See SCANNING FREQUENCY.

strong coupling See CLOSE COUPLING.

strontium Symbol, Sr. A metallic element of the alkaline-earth group. Atomic number, 38. Atomic weight, 87.62. It is used in some ceramic dielectrics, such as *barium-strontium titanate.*

Strowger exchange A telephone system incorporating Strowger switches.

Strowger switch A switch with one input and 100 individually selectable outputs. It is used with telephone switching networks. The telephone dial code causes a contact to move vertically and horizontally in such a way that a particular output is connected to the input. Each output has a unique dial code and each dial code has a unique output.

structured programming Computer programming using a limited number of procedural sets, while minimizing branches to make the program as forward-going as possible. This allows it to be easily modified or debugged.

STS switch See SPACE-TIME-SPACE SWITCH.

stub A (usually short) section of transmission line that is patched onto a longer line for tuning or impedance matching.

Stubs gauge See BIRMINGHAM WIRE GAUGE.

stub tuner A tuning unit consisting of a stub with a short circuit that can be moved along the stub.

stub trap See INTERFERENCE STUB.

stub-type wavetrap See INTERFERENCE STUB.

stuffing bits In a digital communications system, extra bits inserted into some words so that all the words are the same length.

styli Plural of STYLUS.

stylus **1.** The "needle" that conveys vibrations to or from the disk in phonograph-disc recording or playback. **2.** One of the pins in the print head of a dot-matrix printer.

stylus drag See NEEDLE DRAG.

stylus friction Rubbing of the stylus against the record groove in phonograph-disc playback.

stylus pressure See VERTICAL STYLUS FORCE.

stylus printer See WIRE PRINTER.

stylus scratch See NEEDLE SCRATCH.

sub Abbreviation of *subtract.*

sub- Prefix denoting *under*, *below*, *less than*, or *lower than*, with respect to size, value, or rank. Compare SUPER-.

subassembly A completely fabricated unit that forms part of a larger unit into which it easily fits.

subatomic particle **1.** Any of various particles that comprise atoms of matter. **2.** A particle smaller than an atom. See, for example, ANTI-PARTICLE, ELECTRON, MESON, NEUTRETTO, NEUTRINO, NEUTRON, NUCLEON, POSITRON, and PROTON.

subaudible **1.** Pertaining to any frequency falling below the limit of human hearing, that is, less than about 20 Hz. **2.** Any sound that is too low in amplitude to be heard.

subaudible tone A signal, usually a steady, sine-wave tone, sent along with a radio signal. The tone frequency varies from about 20 Hz to 200 Hz, below the audio cutoff frequency of most voice communications systems. Subaudible tones are used mainly for privacy. The receiver is programmed to receive only signals having the correct subaudible tone frequency.

suballocation A portion of a radio-frequency broadcast or communications band that is legally set aside for specific purposes or users, e.g., the Extra-class segment of the 40-meter amateur-radio band.

subband **1.** A portion of a frequency band with specific characteristics. **2.** A portion of a radio-frequency broadcast or communications band that is set aside, legally or by convention, for specific purposes [e.g., the single-sideband (SSB) portion of the 20-meter amateur-radio band].

subcarrier A modulated carrier wave that composes the modulating signal for another carrier wave.

subcarrier band The band of frequencies in which a subcarrier signal is transmitted.

subcarrier frequency modulation In a system in which a carrier frequency is obtained by beating a low-frequency radio-frequency (RF) signal with a high-frequency RF signal, the application of frequency modulation to the low-frequency component. The technique is sometimes used in sweep-frequency signal generators.

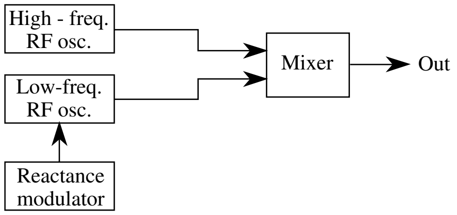

subcarrier frequency modulation

subcarrier oscillator In a color-television receiver, the oscillator operating at the burst (chrominance-subcarrier) frequency of 3.579545 MHz.

subchassis An auxiliary chassis on which one section of a larger piece of equipment is completely assembled and wired.

subfrequency See SUBHARMONIC.

subharmonic An integral submultiple of a fundamental frequency. Thus, for example, the 10th subharmonic of 15 MHz is 1.5 MHz.

submarine cable An underwater cable designed to withstand continuous immersion.

submarine robot A robot designed for underwater operation. It can be operated via telepresence or by simple remote control, usually using a conventional or fiberoptic cable. Some underwater robots have manipulators attached; others are equipped only with cameras, lights, and propulsion devices.

subminiature jack A female connector with an inside diameter of $\frac{3}{32}$ inch.

subminiature plug A male connector with an outside diameter of $\frac{3}{32}$ inch.

submultiple A fractional multiple, usually in reference to a frequency. For example, 7.2 MHz is a submultiple of 14.4 MHz. See SUBHARMONIC.

subpanel The front panel of a removable unit or module that forms a part of a larger unit.

subroutine In a computer program, a sequence of instructions for carrying out a section of the program's function. It is usually entered (led to) by a conditional branch (jump) instruction in the main program.

subscriber An individual user of a communications network or service.

subscript A small number or letter written to the lower right (and occasionally to the lower left) of another number or letter to identify the latter from others of the same designation (e.g., a_5, S_n). Compare SUPERSCRIPT.

subscription TV A television (TV) service paid for by subscribing viewers. The signals are scrambled so as to be useless to nonsubscribers, and legitimate subscribers are provided with a decoder to unscramble the telecasts.

subset **1.** In statistics and set theory, a set whose members are all contained in a larger set. **2.** A telephone handset or deskset (subscriber's set). **3.** A modulator/demodulator for making business machines compatible with telephone circuits.

subsidiary communication authorization Abbreviation, SCA. An authorization provided by the Federal Communications Commission (FCC) for a frequency-modulation (FM) broadcast station to transmit, in addition to its main program, a second program within the assigned bandwidth usually consisting of commercial-free background music on a subcarrier that can be detected only with a special receiver or with a special adapter attached to a standard FM receiver.

subsidiary feedback Feedback other than the main feedback in a system.

subsonic frequencies Frequencies below the range of hearing, that is, less than about 20 Hz. Also called ULTRALOW FREQUENCIES and SUBAUDIBLE FREQUENCIES.

substation An intermediate electricity-distributing location from which electrical energy is transformed and transmitted to users within a given geographical area.

substitution capacitor A capacitor used temporarily in place of another of usually the same value, as in troubleshooting. Also see CAPACITOR SUBSTITUTION BOX.

substitution inductor An inductor used temporarily in place of another of usually the same value, as in troubleshooting. Also see INDUCTOR SUBSTITUTION BOX.

substitution method **1.** A method of measuring a quantity (such as capacitance, inductance, or resistance) in which the value of the unknown quantity is determined in terms of the amount of a standard quantity that must be removed to restore the test circuit to its original state of balance. **2.** A method of troubleshooting in which good components are substituted for bad ones in a circuit (see, for example, SUBSTITUTION CAPACITOR, SUBSTITUTION INDUCTOR, SUBSTITUTION RESISTOR, SUBSTITUTION SPEAKER, and SUBSTITUTION TRANSFORMER).

substitution resistor A resistor used temporarily in place of another of usually the same value, as in troubleshooting. Also see RESISTOR SUBSTITUTION BOX.

substitution speaker A loudspeaker used temporarily in place of another, as in troubleshooting.

substitution transformer A transformer used temporarily in place of another having the same characteristics, as in troubleshooting.

substrate A plate, wafer, panel, or disk of suitable material on (or in) which the components of a unit, such as an integrated or printed circuit, are deposited or formed.

subterranean **1.** Pertaining to components, systems, or devices installed underground. It is applicable especially to cables. **2.** Pertaining to a phenomenon, such as the propagation of electric currents or acoustical waves, that occurs underground.

subterranean acoustical communication A method of communication that uses low-frequency sound waves, such as SONAR, to communicate via conduction through earth or water.

subtracter See ELECTRONIC SUBTRACTER.

subtractive color A color formed by mixing subtractive primary pigments.

subtractive primaries Broad-spectrum pigments used in printing to produce a wide variety of colors through filtering. These primaries are cyan (blue-green), magenta (pink-red), yellow, and sometimes black. They are used to print images that have been filtered through additive-primary lenses.

subtrahend In the process of subtraction, the quantity that is subtracted from another (the minuend) to produce the remainder or difference.

subwoofer A speaker designed to effectively reproduce extremely low audio frequencies, in some cases, subaudible (below 20 Hz).

successive derivatives Successive repetition of the operation of differentiating a function, which yields the first derivative, second derivative, and so on to the *n*th derivative.

successive integration The operation of double or triple integration.

Suhl effect A reduction in hole life that occurs in a semiconductor material in the presence of a magnetic field.

suite **1.** A group of computer programs run successively as a job. **2.** A bundled, high-end software package used especially in business computing.

sulfate Contraction of lead sulfate.

sulfation In a lead-acid storage cell, the formation of disabling lead sulfate during discharge of the cell.

sulfur Symbol, S. A nonmetallic element. Atomic number, 16. Atomic weight 32.06.

sulfur hexafluoride A gas used as a coolant and insulant in some power transformers.

sulfuric acid Formula, H_2SO_4. An acid used in dilute solution as the electrolyte of a lead-acid battery. This highly corrosive fluid also has many industrial uses.

sum The result obtained by adding two or more terms. Compare SUMMATION.

sumcheck See SUMMATION CHECK.

sum frequency **1.** In an amplitude-modulated carrier, the upper sideband frequency (i.e., the sideband equal to the carrier frequency plus the modulating frequency). Compare DIFFERENCE FREQUENCY. **2.** In superheterodyne operation, an intermediate frequency equal to the signal frequency plus the local-oscillator frequency.

summation **1.** The sum of a finite number of terms. Thus, the total resistance of n resistors connected in series is the summation of all R (resistance) terms. **2.** A frequency equal to the sum of two other frequencies.

summation check In computer operations, a check carried out on a group of digits. The result of adding the digits, and disregarding any overflow, is a check digit that can be compared with a standard value for the operation to verify the integrity of data.

summer **1.** See ADDER. **2.** See SUMMING AMPLIFIER.

summing amplifier An operational amplifier whose output is the sum, or is proportional to the sum, of several inputs.

sun battery A set of photovoltaic cells connected in series, parallel, or series-parallel to produce useful output voltages and currents.

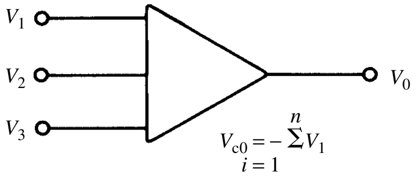

summing amplifier

S units In radio (especially amateur radio), gradations reflecting the strength of received signals. Typically, a value of S9 (9 s-units), representing "extremely strong signals," is equal to a strength of 50 microvolts. The next lower S unit (S8) is 6 dB lower in voltage (i.e., 25 microvolts); S7 is 12 dB below S9 (i.e., 12.5 microvolts), etc.

sun lamp An incandescent or fluorescent lamp with high ultraviolet output, used in medicine for the treatment of certain skin disorders. It can also be used for skin tanning; most physicians discourage habitual use of the lamps for this purpose.

sunlight lamp A lamp that produces visible light whose spectral distribution is similar to that of typical daylight. It generally produces more blue and violet light than a conventional lamp. It is sometimes used for indoor lighting in winter at high latitudes, and/or to enhance plant growth.

sunlight-powered laser A laser whose action is stimulated by sunlight collected by a system of mirrors and lenses. The life of the device is long, compared with that of conventional lasers.

sun-pumped laser See SUNLIGHT-POWERED LASER.

sun relay See SUN SWITCH.

sunspot An area on the sun's surface that is visible as a dark, irregular region of variable size, generally several thousand miles across. Sunspots are believed to be comparatively cool regions associated with solar magnetic disturbances. The number of sunspots is correlated with the frequency and intensity of solar flares (see SOLAR FLARE).

sunspot cycle Regular periodic variation of sunspot activity. The time between peaks in activity is approximately 11 years.

sun switch A photoelectric switch or relay actuated by sunlight and used for various domestic and industrial purposes, such as switching lights, operating window shades, etc.

sup Abbreviation of SUPPRESSOR.

super Contraction of *supersonic.*

super- Prefix denoting *over*, *above*, *greater than*, or *higher than*, with respect to size, value, or rank. Compare SUB-.

superaudible frequency See ULTRASONIC FREQUENCY.

superbeta transistor A transistor or transistor combination, such as a Darlington pair (see COMPOUND CONNECTION), that provides a very high current amplification factor (beta).

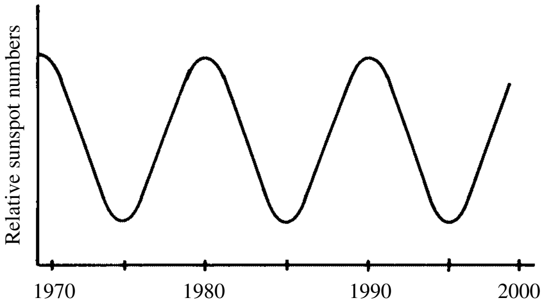

sunspot cycle

supercardioid microphone A microphone that is highly sensitive in one direction and insensitive in all other directions. The directional pattern is similar to a CARDIOID PATTERN, but is exaggerated along the axis of optimum response.

superconducting cable A cable in which superconductivity is achieved by surrounding the cable with liquid helium to lower its temperature to near absolute zero.

superconductivity The virtual disappearance of resistance in some metals cooled to temperatures in the vicinity of absolute zero. Also see CRYOGENICS, CRYOSTAT, and CRYOTRON.

superconductor A material or device that displays superconductivity.

super flatpack An integrated-circuit package of the flatpack type having considerably more components and leads than those in the conventional flatpack.

superhet Contraction of *superheterodyne.*

superheterodyne circuit A circuit in which the incoming signal in a first detector (or mixer) beats with the signal of a local oscillator, resulting in a lower (intermediate) frequency, which then is amplified by an intermediate-frequency (IF) amplifier. This IF signal is detected by a second detector whose output is amplified by an audio-frequency (AF) amplifier. Because the IF amplifier operates at a single (fixed) frequency, it can be adjusted for optimum selectivity and gain. Also called *superhet circuit.*

superheterodyne receiver A radio or television receiver using a SUPERHETERODYNE CIRCUIT.

superhigh frequency See RADIO SPECTRUM.

Supermalloy An alloy having a maximum permeability of 10^6.

supermodulation A type of amplitude modulation (AM) in which one radio-frequency (RF) power stage continuously generates the carrier, and a second (usually identical) RF power stage is gated into full operation at the proper instant by the audio modulation to add additional RF power (corresponding to 100% modulation) to the signal. At the same time, the carrier amplitude is decreased by the proper amount to fulfill the conditions of a signal amplitude swing between zero and twice maximum for 100% modulation.

superposition In a complex wave, the manner in which the constituent waves combine. The instantaneous value of the complex wave is equal to the vector sum of the instantaneous values of all the constituent waves.

superposition theorem In a network of linear elements, if a voltage E_1 in branch A causes a current I_1 to flow through branch C, and if a voltage E_2 in branch B (which might be identical with branch A) causes a current I_2 to flow through branch C, then E_1 in branch A and E_2 in branch B applied simultaneously will cause a current equal to $I_1 + I_2$ to flow through branch C. Compare COMPENSATION THEOREM, MAXIMUM POWER TRANSFER THEOREM, NORTON'S THEOREM, RECIPROCITY THEOREM, and THEVENIN'S THEOREM.

superpower An arbitrary term denoting very high power. In the rating of standard broadcast stations, it has come to signify 1,000,000 watts (one megawatt) radio-frequency (RF) power output.

superradiance In a laser, a rapid increase in intensity of fluorescent-line emission with increasing excitation power.

superregenerative circuit A regenerative detector circuit in which regeneration is periodically increased almost to the point of oscillation, then decreased. This quenching action takes place at a supersonic rate (typically at 50 or 100 kHz) so that the quenching is inaudible. The result is that much more regeneration is afforded, without the detector going into oscillation, than is possible by simply increasing the regeneration manually. An extremely sensitive detector is the result.

supersaturated solution A solution that contains more solute than it normally would hold. Supersaturated solutions are obtained through special techniques and are extremely unstable. Compare SATURATED SOLUTION. Also see SOLUTE; SOLUTION, **1**; and SOLVENT, **2**.

superscript A small number or letter written to the upper right of another number or letter, the BASE, to indicate the power to which the base must be raised. Example: 10^5, e^x, y^2. Also called EXPONENT. Compare SUBSCRIPT.

supersensitive relay A relay that operates with a current of less than one milliampere, or with a voltage of less than one millivolt.

supersonic flow In a gas or liquid, movement of the medium at a speed greater than the speed of sound in that medium. Supersonic flow results in a greatly increased resistance or drag because of shock waves that form in the medium.

supersonic frequency See ULTRASONIC FREQUENCY.

supersonics See ULTRASONICS.

supervised line In a security system, a wire or foil strip that carries electrical current. If the current changes in such a line, an alarm is actuated.

supervisor **1.** In a computer, a set of routines that oversees the operation of the system. The supervisor routines are coordinated by the central processing unit. **2.** The execution of such a set of routines. **3.** A microcomputer that oversees the operation of a security system.

supervisory circuit In a security system, a link between a sensor and the central computer or control device. This link can be via electric current through a wire or cable, but other methods can be used, such as fiberoptics, line-of-sight optics, infrared, ultrasonic, or radio.

supply **1.** See CURRENT SUPPLY. **2.** See POWER SUPPLY. **3.** See VOLTAGE SUPPLY.

supply current Alternating or direct current available for operating a circuit, device, or system.

supply frequency The frequency of an alternating-current power supply.

supply power The maximum power that can be reliably delivered by an alternating-current or direct-current power supply.

supply reel In a reel-to-reel tape recorder or player, the reel that is initially full, and that gradually empties as the tape moves through the machine.

supply voltage The voltage of an alternating-current or direct-current power supply.

suppressed carrier A carrier that has been canceled or filtered out of a carrier/sideband combination.

suppressed-carrier double sideband See DOUBLE-SIDEBAND and SUPPRESSED CARRIER.

suppressed-zero instrument A meter or graphic recorder in which the zero point is off-scale or upscale, but has been brought to scale-zero by means of mechanical adjustment or use of a bucking voltage.

suppressor **1.** A filter used to suppress radio interference. **2.** See AUTOMATIC NOISE LIMITER. **3.** See SPARK SUPPRESSOR. **4.** In a pentode vacuum tube, a gridlike element between the screen grid and the plate, used to suppress secondary emission. Also see GRID, **2** and PENTODE.

suppressor circuit The circuit associated with the suppressor electrode of a vacuum tube.

suppressor diode A semiconductor diode used to prevent inductive kickback in circuits, to eliminate or reduce transients, or to prevent arcing between make-and-break contacts.

suppressor grid See SUPPRESSOR, **4**.

suppressor modulation A method of modulation in which a modulating voltage is superimposed on the suppressor voltage of a pentode radio-frequency power amplifier tube.

suppressor pulse A pulse that prevents electron flow.

surface analyzer A device designed for the measurement of surface flatness or uniformity.

surface-barrier diffused transistor See MICROALLOY DIFFUSED TRANSISTOR.

surface-barrier transistor Abbreviation, SBT. A pnp transistor made by means of electrolysis and electroplating: Two fine streams of indium sulfate

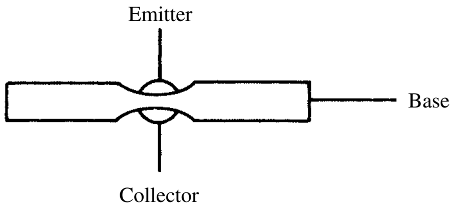

surface-barrier transistor

solution are placed on axially opposite points on the faces of an n-type wafer. At the same time, a direct current is passed through the wafer and solution in such a direction as to remove semiconductor material electrolytically from the faces; the tiny sprayed areas are etched away. When the desired wafer thickness is reached at the points of impact, the etching process is arrested by reversing the direction of current flow. This reversal causes an indium dot to be plated on each opposite face in the etched-out pit. Leads are attached to the collector and emitter dots and to the wafer (base).

surface-charge transistor A semiconductor device consisting essentially of two narrowly separated plates (source electrode and receiver electrode) deposited on the film-insulated surface of a silicon chip, and a third, overlapping electrode (the transfer gate) deposited on, but insulated from, the other electrodes. An input signal stores a charge in the capacitor formed by the source electrode and chip. A subsequent trigger signal applied to the transfer gate transfers the charge to the receiver electrode, where it becomes an output signal (often amplified, with respect to the input signal).

surface effect An effect (such as current, resistance, or resistivity) observed on the surface of a sample of material, rather than throughout the body of the material. Compare BULK EFFECT.

surface insulation A coating applied to the surfaces of core laminations to prevent the passage of currents between laminations.

surface leakage Leakage of current over the surface of a dielectric material, as opposed to leakage through the interior of the material.

surface noise See NEEDLE SCRATCH.

surface recombination rate For a semiconductor, the rate at which electrons and holes recombine at the surface. Compare VOLUME RECOMBINATION RATE.

surface resistivity The resistance of a unit area of a material, measured between opposite edges. Compare VOLUME RESISTIVITY. Also see RESISTIVITY.

surface tension The tendency of the surface of a liquid to “shrink.” This property varies with different liquids and is caused by a net molecular force directed inward from the surface.

surface wave **1.** The earth-guided component of a ground wave. (The other component is the SPACE WAVE.) **2.** An acoustic wave traveling along the surface of a plate in a surface-wave amplifier or surface-wave filter.

surface-wave amplifier An amplifying device consisting essentially of a surface-wave filter to which has been added a direct-current-biased n-type silicon electrode, which is separated from the crystal substrate of the filter by a very thin oxide layer. Amplification is produced by interaction between the electron current in the silicon and the piezoelectric field of the filter. Also see ACOUSTOELECTRONICS.

surface-wave filter An acoustoelectronic device consisting essentially of a crystal plate having electrodes at each end. An alternating-current (ac) input signal applied to one electrode sets up acoustic waves that travel along the surface of the plate to the other electrode, where they generate an ac output voltage by piezoelectric action. The resonant frequency of the device is governed by the dimensions of the crystal plate. Also see ACOUSTOELECTRONICS.

surge A sudden rise or flow of current or voltage.

surge absorber See SURGE SUPPRESSOR.

surge arrester See SURGE SUPPRESSOR.

surge current A heavy current that flows initially into a capacitor when a charging voltage is applied.

surge impedance Symbol, Z_o. The impedance seen by a pulse applied to a transmission line; $Z_o = L/C$ (approximately), where L and C are the inductance and capacitance, in microhenrys and microfarads, per unit length of the line. Also called CHARACTERISTIC IMPEDANCE.

surge protector Misnomer for SURGE SUPPRESSOR.

surge suppressor A semiconductor device used to absorb potentially destructive transients or overvoltages on a utility power line. It has a three-wire cord for plugging into a 117-volt outlet, a power switch, and several three-wire outlets for connection to sensitive electronic equipment (such as personal computers, videocassette recorders, television sets, hi-fi amplifiers, etc.).

surround See SUSPENSION, **1**.

Surround Sound The trade name for a multichannel sound system for use with television receivers and videocassette players. Some televised movies, especially on cable and satellite networks, deliver multichannel sound through receiving/recording systems equipped with special decoders.

surveillance **1.** A method of monitoring a specific area or volume for intrusion or other disturbance. **2.** A means of monitoring a specified portion of the electromagnetic spectrum for unauthorized signals.

surveillance radar An air-traffic-control radar that supplies continuous information regarding the azimuth and distance of aircraft inside a selected radius around an airport.

susceptance Symbol, *B.* Unit, siemens. The reactive component of admittance, as distinguished from conductance.

susceptibility The capacity of a substance to become magnetized, expressed as the ratio of magnetization to the strength of the magnetizing force.

suspension **1.** In a speaker, the flexible, circular or elliptical structure via which the cone is attached to the frame. **2.** The wire or metallized fiber supporting the movable coil of a galvanometer. **3.** Particles of a substance and the liquid in which it is mixed, but not dissolved. **4.** The substance, as defined in **3**.

suspension galvanometer A meter with a light-beam apparatus for lengthening the arc through which the pointer travels. When the beam of light is cast a long distance, a tiny movement of the coil will cause considerable movement of the image.

sustained oscillations Oscillations that continue as long as power is supplied to the oscillation generator. Also see CONTINUOUS WAVE. Compare DAMPED OSCILLATIONS.

sustaining voltage The voltage at which second-collector breakdown occurs in a transistor (see SECOND BREAKDOWN).

S video In animation, a scheme that separates brightness and color. It can enhance the video in some applications.

SW Abbreviation of SHORTWAVE.

sw Abbreviation of SWITCH. (Also, S or s.)

swamping resistor **1.** A noninductive resistor connected in parallel with the input circuit of a class-B linear amplifier for automatic regulation of the excitation. **2.** A resistor connected in series with the emitter of a bipolar transistor to minimize the effects of temperature-induced variations in junction resistance.

swarf The string of material that threads off a disc during sound recording.

S_W band A section of the S BAND, extending from 3400 to 3700 MHz.

sweep **1.** To deflect the electron beam in a cathode-ray tube, usually horizontally, to provide a time base. **2.** The circuit for achieving the particular deflection described in **1**.

sweep circuit A circuit, such as a deflection generator (e.g., a sawtooth oscillator), for producing a sweep signal. Also see SWEEP.

sweep delay In an oscilloscope, the process of initiating the sweep of the electron beam at some selected instant after the signal has started.

sweep-delay circuit In an oscilloscope or radar, the circuit for delaying the sweep until the start of the signal. Also see DELAYED SWEEP.

sweeper **1.** See SWEEP GENERATOR. **2**. See SWEEP-SIGNAL GENERATOR.

sweep frequency **1.** The frequency at which the electron beam in a cathode-ray tube is deflected along the reference axis. **2.** The frequency at which the carrier frequency is increased and decreased by a sweep-signal generator. **3.** In an oscilloscope, the number of times that the trace moves across the screen in one second. It is equal to the reciprocal of the SWEEP PERIOD.

sweep generator **1.** A device that causes the electron beam in a cathode-ray tube to scan at a known speed. **2.** An oscillator that generates a signal that rapidly varies in frequency. It is used for the testing and adjustment of bandpass filters and other selective circuits.

sweeping receiver See SCANNING RECEIVER.

sweep magnification In an oscilloscope, increasing or multiplying the sweep frequency, thus reducing the time per horizontal division. This increases the maximum frequency of waveforms that can be analyzed, and allows closer inspection of high-frequency signal components.

sweep magnifier In an oscilloscope, a circuit for achieving sweep magnification.

sweep oscillator See SWEEP GENERATOR.

sweep period The duration, in seconds, of one complete cycle of sweep signal in an oscilloscope. It is equal to the reciprocal of the SWEEP FREQUENCY.

sweep signal The (usually linear, sawtooth) signal used to sweep the beam of an oscilloscope tube. Also see SWEEP, **1**, **2**.

sweep-signal generator A signal generator that supplies a signal whose frequency varies automatically and periodically throughout a given band.

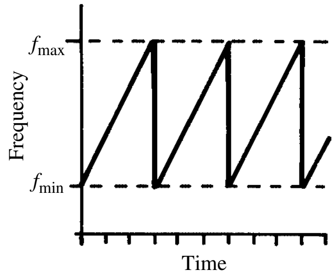

sweep-signal generator

sweep test A method of testing the attenuation-vs.-frequency characteristics of a selective circuit, using a radio-frequency sweep generator.

sweep time The actual time required for a single sweep by a deflecting signal; $t = 1/f$, where t is sweep time in seconds, and f is sweep frequency in hertz.

sweep voltage The peak voltage amplitude of the sweep signal.

SWG Abbreviation of *standard wire gauge.*

swing The maximum change exhibited by a varying quantity (e.g., *amplitude swing* and *frequency swing*).

swinging choke A filter choke that exhibits relatively high inductance when low current flows through it, and lower inductance when high current flows through it. This inductance, which swings under conditions of varying load current, permits the use of a high-resistance bleeder resistor. Compare SMOOTHING CHOKE.

Swiss-cheese packaging A method of packaging an electronic circuit, in which components are inserted into the assembly through holes drilled or punched in parallel, stacked printed-circuit boards.

switch **1.** A circuit or device (electronic, electromechanical, or mechanical) for opening and closing a circuit or for connecting a line to one of several different lines (e.g., *rotary selector switch*). **2.** To change the logic state of a circuit or device. **3.** In a computer program, a branch instruction directing the program to a line number dependent on the value of a variable or result (e.g., BASIC's GOTO). **4.** To cause an electrical circuit to change state, as from low to high or vice versa.

switch current **1.** The current flowing through a switch. **2.** The current flowing through a switching diode or transistor. **3.** The minimum current necessary to produce switching of a transistor, specified in milliamperes or microamperes.

switchgear Collectively, devices and systems for making and breaking circuits—either automatically or manually.

switchhook In a telephone set, the spring-and-switch device that engages the line when the receiver is lifted.

switching characteristics Technical data describing the performance and capabilities of switching devices and circuits.

switching circuit An on-off type of circuit containing electronic or mechanical switches.

switching diode See COMPUTER DIODE.

switching frequency The frequency at which a repetitive switch operates. Also see SWITCHING RATE.

switching mode Operation in which a device, such as a transistor or diode functions as a binary digital device, rather than as an analog device. The current is generally either zero (cutoff or pinchoff) or some value that depends on the bias and on the applied voltage.

switching rate The rate (e.g., closures per second) at which a repetitive switch operates. Also see SWITCHING FREQUENCY.

switching speed The time required for a switch to open or close or for a switching device to change states (as from cutoff to saturation). Also see SWITCHING TIME.

switching time The time required, after the application of a pulse, for an electronic switch to change state. Also see SWITCHING SPEED.

switching transistor A transistor designed especially for on-off operation. Such units exhibit short recovery time and low capacitance.

switching voltage The largest voltage that a switching device can handle without malfunctioning.

switch leakage current **1.** The current flowing through a switching device when it is supposed to be nonconducting. **2.** In a switching transistor, for a given voltage, the leakage current between the emitter and collector when the device is supposed to be nonconducting.

SWL Abbreviation of SHORTWAVE LISTENER.

SWR Abbreviation of STANDING-WAVE RATI0.

SWR bridge A four-arm resistance bridge for measuring voltage standing-wave ratio. This radio-frequency bridge has noninductive resistors in three of its arms and the device under test in the fourth arm. The bridge is balanced first with an equivalent noninductive resistor that replaces the device, and the output voltage is noted. Then the device is substituted for the test resistor, and the change in voltage is noted. The standing-wave ratio is determined from the voltage ratio.

SWR meter See SWR BRIDGE.

S_Y band A section of the S BAND, extending from 2600 to 2700 MHz.

syllable compandor A device that compresses or expands the amplitude of an audio signal. The time constant is fast enough to allow response to individual syllables. Compression is generally used at the transmitting station, and expansion at the receiving station.

sym **1.** Abbreviation of *symmetrical.* **2.** Abbreviation of SYMBOL.

symbol **1.** A letter or graphic device representing a quantity or term [e.g., I (current), f (frequency), etc.]. **2.** A conventional device denoting a mathematical operation (e.g., +, /). **3.** In a circuit diagram, a pictorial device representing a component.

symbolic address An address in a source-language computer program (i.e., the arbitrary label used by the programmer).

symbolic language See SOURCE LANGUAGE.

symbolic logic A system for representing logical relationships, such as those acted upon by computer and switching circuits, by means of symbols that are usually nonnumerical. Also see BOOLEAN ALGEBRA.

symmetrical circuit A circuit having identical configurations on each side of a dividing line, such as the ground bus. A push-pull circuit is an example.

symmetrical communications **1.** Two-way communications in which the volume of transmitted data is the same, or nearly the same, in both directions. **2.** Two-way communications in which the speed of transmitted data is the same, or nearly the same, in both directions. Compare ASYMMETRICAL COMMUNICATIONS.

symmetrical conductivity Identical conductivity for both positive and negative electricity. Compare ASYMMETRICAL CONDUCTIVITY.

symmetrical FET See SYMMETRICAL FIELD-EFFECT TRANSISTOR.

symmetrical field-effect transistor A field-effect transistor whose source and drain terminals can be interchanged without affecting circuit operation. Also called BIDIRECTIONAL TRANSISTOR. Compare UNILATERAL FIELD-EFFECT TRANSISTOR.

symmetrical input See BALANCED INPUT.

symmetrical output See BALANCED OUTPUT.

symmetrical transistor See BIDIRECTIONAL TRANSISTOR.

symmetrical wave A wave whose positive and negative half-cycles are identical in shape and peak amplitude.

symmetry **1.** The condition of having the same shape on each side of an axis. **2.** The condition of conducting positive and negative currents equally well. **3.** The condition in which a circuit is identical on both sides of a reference line, such as the ground line.

sympathetic vibration Resonant vibration of one body in response to the vibration of another body.

sync **1.** Contraction of SYNCHRONIZATION. **2.** Contraction of SYNCHRONISM.

sync amplifier In a television circuit, the amplifier used to increase the amplitude of the sync pulses after they are separated from the composite video signal.

sync generator A circuit that produces the synchronization pulses in a television transmitter.

synchro A dynamo-electric-control device that, when connected to a similar device and the alternating-current power line, permits remote control. Thus, when the rotor of one synchro is turned to a certain position, the rotor of the other assumes the same position. Also see AUTOSYN and SELSYN.

synchrocyclotron A type of cyclotron in which the variation in mass, because of increased velocity, of the charged particles is compensated, resulting in higher energy for the particles.

synchro differential A synchro that receives two input signals and delivers a single output signal. The inputs can be two electrical signals, or one electrical signal and one mechanical signal.

synchrodyne receiver A direct-conversion receiver in which the local oscillator frequency or phase is locked into synchronism with the incoming signal carrier frequency or phase.

synchroflash A flash that is synchronized with the shutter of a camera.

synchro generator The transmitting member of a synchro system.

synchro motor The receiving member of a synchro system.

synchronism **1.** The condition of being in step, as when two motors are running in synchronism with each other and the power frequency, or when two relays open and close in step. **2.** The condition of being in phase, as when two pulses occur simultaneously.

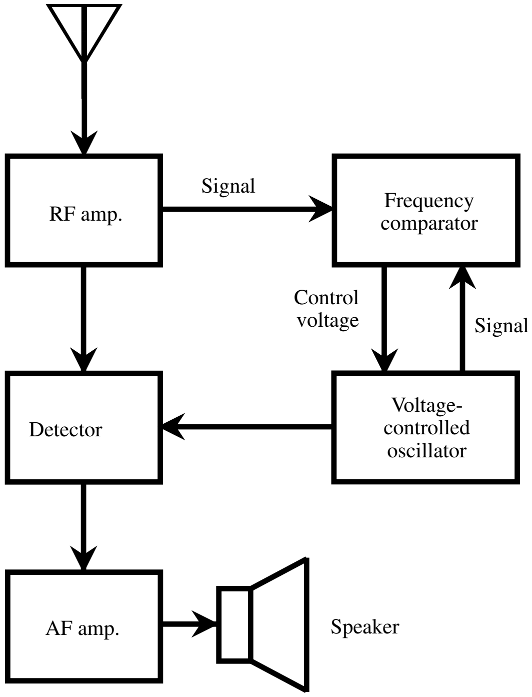

synchrodyne receiver

synchronization The coincidence of one process or operation with another, as in the synchronization of an oscillator frequency by means of an applied standard-frequency voltage, in which case the oscillator frequency becomes that of the standard signal.

synchronized clamping A type of clamping in which an output voltage is maintained at a predetermined fixed value until a synchronizing pulse is applied, whereupon the output follows the input.

synchronized multivibrator See DRIVEN MULTIVIBRATOR.

synchronizer A computer storage device used between two devices transmitting data at different speeds, to counteract this differential (as a buffer).

synchronizing signal A signal used to synchronize another signal, usually in frequency.

synchronous The condition of operating in step (phase) with some reference signal.

synchronous clock **1.** An alternating-current clock driven by a synchronous motor. Although 60-Hz models are common, such clocks are not restricted to low-frequency ac operation; 1-kHz types, for example, are used in some primary frequency standards. **2.** The timing source in a synchronous computer.

synchronous computer A computer whose operations are timed by single-frequency clock signals.

synchronous contacts The rectifying contacts of a synchronous vibrator (see VIBRATOR-TYPE RECTIFIER).

synchronous converter A synchronous machine that can run on alternating current and generate direct current, or vice versa. Also called ROTARY CONVERTER.

synchronous gate A gate whose output is synchronized, according to the input signal.

synchronous generator An alternator operating in synchronism with one or more other alternators. Also see SYNCHROSCOPE, **2**.

synchronous induction motor An alternating-current motor that is intermediate between the fractional-horsepower reluctance motor and the multiple-horsepower three-phase, synchronous motor. The synchronous induction machine starts like an induction motor and runs like a synchronous motor.

synchronous inputs In a computer flip-flop, inputs that accept pulses only at the command of the clock.

synchronous machine See SYNCHRONOUS INDUCTION MOTOR.

synchronous motor See SYNCHRONOUS INDUCTION MOTOR.

synchronous network A communications network in which all clocks are set so that they run at the same rate, their increments are identical in duration, and transitions occur simultaneously or with a specified phase difference. Such a system allows for greatly enhanced signal-to-noise ratio for a given amount of transmitter power, and also reduces the bandwidth necessary for a single signal so that many more signals can be placed in a given frequency band.

synchronous orbit See GEOSTATIONARY ORBIT.

synchronous satellite See GEOSTATIONARY SATELLITE.

synchronous speed For an alternating-current (ac) machine, the speed corresponding to the ac frequency.

synchronous transmission A method of signal transmission in which individual symbols are sent at a specified rate, according to a clock that also governs the receiver.

synchronous vibrator See VIBRATOR-TYPE RECTIFIER.

synchroscope **1.** An oscilloscope having a high-speed sweep triggered by a synchronizing signal. Such an instrument is valuable for viewing high-speed pulses. **2.** A pointer-type instrument used to indicate the synchronism between two power alternators.

synchro system A circuit or system using synchros for the transmission and reception of positioning signals. Also see SYNCHRO.

synchrotron A particle accelerator that uses a high-frequency electrostatic field and a low-frequency magnetic field to impart very high velocity to the particles.

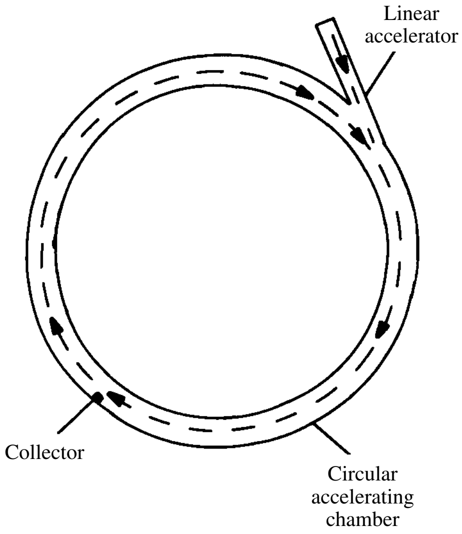

synchrotron

sync pulse **1.** A pulse used to control the frequency or repetition rate of an oscillator or other generator. **2.** In a television system, a pulse transmitted as part of the composite video signal to control scanning. Also see HORIZONTAL SYNC PULSE and VERTICAL SYNC PULSE.

sync separator In a television receiver circuit, a stage used to separate and deliver the sync pulses from the composite video signal. See, for example, DIODE SYNC SEPARATOR.

sync signal See SYNCHRONIZING SIGNAL.

sync takeoff The point in the video amplifier circuit of a television receiver at which the composite video signal is sampled to extract the sync pulses.

syntax **1.** The rules by which computer program statements are structured. **2.** The way that a written or spoken sentence is constructed. It is important in speech recognition and speech synthesis.

synthesis The rigorous (usually mathematical) design of an electronic circuit or device and the accurate prediction of its performance. Compare ANALYSIS.

synthesizer **1.** See SIGNAL SYNTHESIZER. **2.** A circuit synthesizer (i.e., a device that allows a wide variety of circuits to be set up temporarily or simulated, for testing and evaluating). Sometimes, a specially programmed computer serves this purpose. **3.** A keyboard on which music can be played, and whose output can be adjusted to simulate the sounds of various musical instruments. **4.** See MOOG SYNTHESIZER.

synthetic bass An apparent accentuation of bass notes resulting from intermodulation distortion in an amplifier.

synthetic crystal An artificially produced crystal, such as synthetic quartz.

synthetic resin An artificially produced resin. Also see THERMOPLASTIC MATERIAL and THERMOSETTING MATERIAL.

syntony See RESONANCE.

syst Abbreviation of SYSTEM.

system **1.** An integrated assemblage of hardware and/or software elements operating together to accomplish a prescribed end purpose (e.g., *servo system*, *operating system*, and *communications system*). **2.** A methodology incorporating fixed and ordered procedures for accomplishing an end purpose. **3.** A self-contained computer workstation.

systematic error See CUMULATIVE ERROR.

system engineering See SYSTEMS ENGINEERING.

system of units A set of fundamental units for defining the magnitudes of all physical variables. The most common system of units is the standard international (SI) system.

systems analysis In computer system operation, analyzing the way something is done and devising a better alternative by isolating the problem area, scrutinizing the system as it stands, studying what is thereby disclosed, devising the alternate application of software and/or hardware, disseminating the revised operational procedure, and overseeing the implementation of the new method.

systems engineering The branch of engineering devoted to the design, development, and application of complete systems. The approach takes into consideration all elements in a system or process and their integration.

systems flowchart A flowchart showing the interrelationship of activities in a system.

S_Z band A section of the S BAND, extending from 3900 to 4200 MHz.

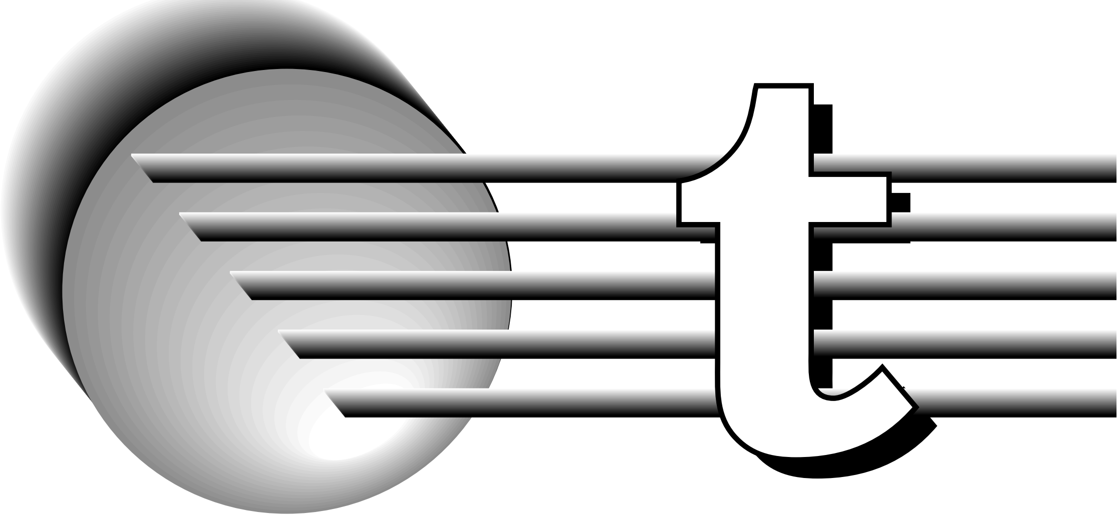

T **1.** Symbol for TRANSFORMER. **2.** Abbreviation of prefix TERA-. **3.** Symbol for *thermodynamic temperature.* **4.** Symbol for TRITIUM. **5.** Abbreviation of TON. (Also, t and tn.) **6.** Abbreviation of TESLA. **7.** Symbol for KINETIC ENERGY. **8.** Symbol for PERIOD. **9.** Symbol for *true.*

t **1.** Symbol for TIME. **2.** Abbreviation of TON. (Also, T and tn.) **3.** Symbol for CELSIUS TEMPERATURE. (Also, T.) **4.** Abbreviation of TARGET. **5.** Abbreviation of *technical.* **6.** Abbreviation of TENSION.

Ta Symbol for TANTALUM.

tab On the keyboard for a computer, typewriter, terminal, or word-processing system, a key that moves the cursor a specified number of spaces toward the right. It also performs various other functions in menu-driven or graphical computer interfaces.

table In an internal or external computer memory, an array (i.e., a list or matrix) of data that can be recalled using keys (e.g., single- or double-subscripted variables).

table look-at Abbreviation, TLA. In computer operations, finding the position of a data item in a table by implementing an algorithm.

table look-up Abbreviation, TLU. In computer operations, locating items in a table by inspecting what is in the table, key by key.

tabulate In data processing, to combine the totals for data item groups having the same key.

tabulation **1.** The printout of what has been tabulated (see TABULATE). **2.** The computer-program-directed movement of the cursor on a cathode-ray-tube display, or of a typewriter carriage, to certain positions in a line.

tabulator See TAB.

tacan A pulse-type UHF air navigation system in which a station is interrogated by signals from an aircraft to provide bearing and range information. The name is an acronym for *tactical air navigation.*

tach Abbreviated form of *tachometer.*

tachometer See ELECTRONIC TACHOMETER.

tachometer generator A small, dynamo-type electric generator that delivers a voltage proportional to the rotational speed of a shaft to which it is attached.

tachyon A high-speed subatomic particle thought to move faster than the speed of light.

tactical air navigation See TACAN.

tactical radar A radar system used in military operations.

tactile sensor A device that provides an intelligent machine with a sense of "touch": temperature, pressure, force, texture, and torque. It is important in robotics, and also in some computer applications, such as virtual reality (VR).

T-adapter See TEE-JUNCTION.

tag **1.** In data-processing and computer operations, the identification of digits or characters forming part of a record. **2.** An encoded price tag (i.e., a passive transponder or barcode strip). It is commonly used in retail stores.

tag converter A device that senses the information on tags (see TAG, **2**) and transfers it to a computer system.

tail **1.** The decay of a waveform from maximum amplitude to zero amplitude. **2.** Any pulse that follows a main pulse as a result of the main pulse.

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

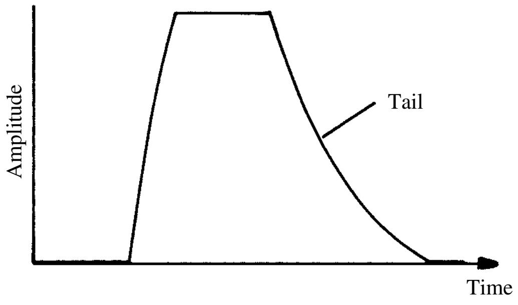

tail

taillight monitor An electronic device for warning a motorist of taillight failure.

tail pulse **1.** A pulse with a fast rise time, but a slow decay time. **2.** See TAIL, **2**.

takedown In computer operations, the process of clearing output peripherals for upcoming program runs.

take-down time The duration of the TAKEDOWN process.

take-up reel In a reel-to-reel tape recorder/reproducer, the reel on which the tape accumulates during recording or reproduction.

talk-back circuit See INTERPHONE.

talk-listen switch A transmit-receive switch in an intercommunication system (see INTERCOM).

talk power See SPEECH POWER.

tally **1.** To obtain a sum or total. **2.** The rows of operands, subtotals, and totals that an adding machine prints.

tally reader A device that, via optical character recognition (OCR), can read the digits and symbols on a tally (see TALLY, **2**).

tamper switch A device that closes a circuit or actuates an alarm when a certain set of conditions is altered.

tan Abbreviation of TANGENT.

tan^{-1} Arc tan (inverse tangent function).

tandem transistor An assembly of two series-connected transistors in the same envelope.

tangent **1.** Abbreviation, tan. The ratio of the side opposite to the side adjacent to an acute angle in a right triangle. **2.** A line that intersects a curve at a single point without crossing the curve. **3.** A plane that intersects a curved surface at a single point.

tangent galvanometer A galvanometer in which the current is proportional to the tangent of the angle of deflection. Compare SINE GALVANOMETER.

tangential mode In acoustics, the reflection of sound waves from four surfaces in a room having six interior surfaces.

tanh Abbreviation of HYPERBOLIC TANGENT.

tank **1.** A parallel-resonant inductance-capacitance circuit in the output of a radio-frequency amplifier. **2.** A circuit in which is stored electrical energy of frequencies in a range whose midpoint is resonance for the circuit. **3.** See MERCURY DELAY LINE.

tantalum Symbol, Ta. A metallic element of the vanadium family. Atomic number, 73. Atomic weight, 180.95. Tantalum is used in the elements of some electron tubes and in some electrolytic capacitors.

tantalum capacitor A type of electrolytic capacitor that uses tantalum rather than aluminum. The tantalum can be foil, as is the aluminum in a conventional electrolytic capacitor. It might also take the form of a porous pellet, the irregular surface of which provides a large area in a small volume. An extremely thin oxide layer forms on the tantalum. These capacitors have high reliability and excellent efficiency, and are used in military applications because they have a low failure rate. They can be used in audio and digital circuits in place of aluminum electrolytic capacitors. The main disadvantage of tantalum capacitors is that they are comparatively expensive. Compare CERAMIC CAPACITOR, ELECTROLYTIC CAPACITOR, MICA CAPACITOR, PAPER CAPACITOR, PLASTIC-FILM CAPACITOR.

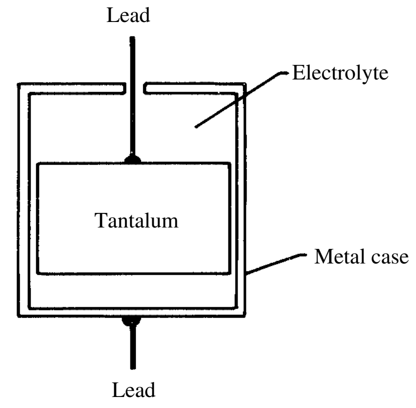

tantalum capacitor

tantalum detector A radio-frequency contact detector consisting essentially of a fine tantalum wire whose point lightly touches the surface of a small pool of mercury.

tantalum-nitride resistor A resistor consisting of a thin film of tantalum nitride deposited on a substrate. Also see THIN FILM.

T-antenna See TEE-ANTENNA.

tap A connection made to an intermediate point on a coil, resistor, or other device. See, for example, CENTER TAP and TAPPED COMPONENT.

tap changer A device that facilitates adjustment of the turns ratio of a transformer by changing the tap on the primary winding, the secondary winding, or both. It also facilitates the adjustment of the impedance-transfer ratio. It can be used for power supplies or audio- or radio-frequency transformers.

tape **1.** Magnetic tape. **2.** Insulating tape. **3.** To make a magnetic tape recording (audio, video, and/or digital data).

tape cable **1.** A form of cable in which all of the conductor centers lie in the same plane. **2.** A flat cable, commonly used in situations, where repeated flexing occurs.

tape cartridge A holder and the reel of blank or prerecorded magnetic (⅛-inch audio or wider video) tape it contains, which can be inserted into a recorder/reproducer without having to thread or otherwise handle the tape for either playing or rewinding.

tape comparator In a data-processing or computer system, a machine that compares tapes generated from the same input, for differences in the data thereon; it is a character-by-character process.

tape core A ring-type magnetic core made by tightly winding metal tape in several layers for the desired thickness.

tape counter See POSITION INDICATOR.

tape deck In a tape recorder/reproducer, the complete tape-transport mechanism (drive, heads, equalization circuitry, and preamplifiers).

tape drive **1.** See TAPE TRANSPORT. **2.** See TAPE DECK. **3.** A tape recorder/reproducer for computer data, used for backup and/or archiving.

tape file A data file recorded on magnetic tape.

tape group An assembly of two or more tape decks.

tape label On a reel or cassette of magnetic tape containing a data file, a record at the beginning or end that contains information about the file.

tape loop An endless loop of magnetic tape.

tape magazine See TAPE CARTRIDGE.

tape mark **1.** A character that subdivides the magnetic tape file on which it is recorded. Also called CONTROL MARK. **2.** A character marking the end of a length of magnetic tape on a reel. Also called END-OF-TAPE MARK.

tape plotting system In computer operations, a system for operating a digital incremental plotter using the information on magnetic tape.

taper In a potentiometer or rheostat, the rate of change in resistance during uniform rotation of the shaft. See, for example, LINEAR TAPER and LOG TAPER.

tape recorder A machine for recording audio, video, or data signals on magnetic tape; it can usually also play back the recorded material.

tapered potentiometer A potentiometer having a tapered resistance winding (see TAPERED WINDING).

tapered winding A resistance winding (in a rheostat or potentiometer) in which the resistance change per unit length of winding is nonuniform (see, for example, LOG TAPER). However, a winding having uniform resistance change is often called LINEAR TAPER.

tape skew A condition in which a magnetic tape passes between the recording or playback heads in an irregular way. The result is that the various channels, or tracks, are not always perfectly aligned.

tape sort The (computer) operation of sorting data in a magnetic tape file into a record-key-determined sequence.

tape splicer A mechanism that aligns and secures the overlapping ends of broken magnetic tape so that they can be cut (often with an integral cutter) to form butted ends, and taped into a splice.

tape station See TAPE DECK.

tape-to-head contact See HEAD-TO-TAPE CONTACT.

tape transmitter A transmitter that receives its signal input from a recorded tape.

tape transport In a tape deck or reproducer, the device that moves the tape past the heads.

tape unit **1.** See TAPE DECK. **2.** See TAPE GROUP.

tape verifier In computer operations, a device that checks the integrity of data on paper tape through comparison with an original document.

tape width In magnetic tape, the dimension perpendicular to tape travel; in general, the greater the tape width, the more tracks the tape can contain.

tape-wound core See TAPE CORE.

tapped coil An inductor to which one or more intermediate connections (taps) are made at appropriate turns to provide intermediate values of inductance.

tapped component A coil, transformer, choke, resistor, or other component in which an intermediate connection is made. See, for example TAPPED COIL.

tapped inductor See TAPPED COIL.

tapped resistor A resistor in which one or more intermediate connections (taps) are made to appropriate parts of the resistance element to provide steps of resistance.

tapped transformer A transformer having one or more tapped windings.

tapped winding A transformer or choke winding with one or more taps. Also see TAP and TAPPED COIL.

tap switch A multiposition switch used to connect an external circuit to various taps on a component. Also see SELECTOR SWITCH and TAP.

target **1.** The bombarded electrode in an X-ray tube. **2.** The scanned storage element in a television camera tube. **3.** In radar operations, the scanned object. **4.** An object intended for nuclear-particle bombardment. **5.** A goal—especially in a production process (deadline, desired number of units, etc.).

target acquisition **1.** The moment at which a target comes within the range of a radar system. **2.** The observation of a new target on a radar screen.

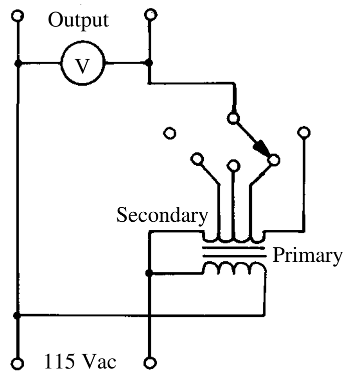

tapped transformer

target discrimination The extent to which a radar system can distinguish between two targets that are close together. Also called *target resolution.* It is specified in linear units (such as meters, kilometers, or miles).

target identification Any method by which the identity of a radar target is determined.

target voltage In a television camera tube, the backplate-to-cathode voltage.

task environment The characteristics of the space in which an autonomous robot works. It depends on such factors as the intended application(s), the required speed at which the robot(s) must work, and the human/robotic or computer/robotic interface.

task-level programming In robotics, the writing of programs to perform sequences of actions. A complex process, it is a primitive level of artificial intelligence (AI). An example is a program that directs a robot to prepare and serve a meal.

taut-band meter A movable-coil meter in which the conventional spiral springs of the coil are replaced by two tightly stretched, thin, straight metal ribbons, whose twist provides torque that returns the pointer to zero after a deflection.

Tb Symbol for TERBIUM.

BS Abbreviation of *talk between ships.*

Tc Symbol for TECHNETIUM.

TCCO Abbreviation of TEMPERATURE-CONTROLLED CRYSTAL OSCILLATOR.

T circuit See TEE NETWORK.

T circuit parameters See R PARAMETERS.

T circulator See TEE CIRCULATOR.

TCL Abbreviation of TRANSISTOR-COUPLED LOGIC.

TCM Abbreviation of *thermocouple meter* (see THERMOCOUPLE-TYPE METER).

TDM Abbreviation of TIME-DIVISION MULTIPLEX.

TDR **1.** Abbreviation of TIME-DELAY RELAY (see DELAY RELAY). **2.** Abbreviation of TIME-DOMAIN REFLECTOMETRY.

TDS Abbreviation of TIME-DELAY SPECTROMETRY.

TE Abbreviation of *transverse electric* (see, for example, TRANSVERSE ELECTRIC MODE).

Te Symbol for TELLURIUM.

teach box A robot-control device via which an operator can program the machine to carry out specific movements. The operator manipulates controls, and the robot's computer stores the data. When the data is accessed, the robot reproduces the motions.

tearing In a television picture, the abnormal condition in which poor synchronization causes the horizontal lines to be irregularly displaced. The effect resembles cloth being torn.

technetium Symbol, Tc. A metallic element produced artificially. Atomic number, 43. Atomic weight, 98. Formerly called *masurium.*

technician **1.** A person who repairs faulty electronic equipment. **2.** A person who assists with the design and debugging of a system prototype. **3.** A person who operates an electronic system.

technocentrism Overdependence on, and/or obsession with, the products of technology, particularly computers.

tee Pertaining to a network, connection, or configuration whose geometric shape or schematic representation resembles an uppercase letter T.

tee adapter See TEE JUNCTION.

tee antenna An antenna consisting of a horizontal radiator with a vertical lead-in or feeder connected to its center point.

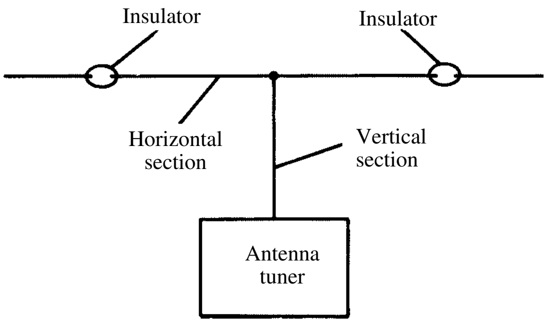

tee antenna

tee circuit See TEE NETWORK.

tee circulator In microwave systems, a tee-shaped junction of three waveguides with a ferrite post at the junction.

tee-equivalent circuit An equivalent circuit in which the components are arranged in the form of a tee. See, for example, TEE NETWORK and R PARAMETERS.

tee junction **1.** A tee-shaped splice between two wires. **2.** A tee-shaped fixture for splicing one coaxial line perpendicularly to another. **3.** A tee-shaped section for joining one waveguide perpendicularly to another. Also called WAVE GUIDE TEE.

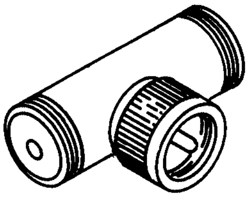

tee junction

tee network A three-terminal network resembling a tee.

tee pad A three-resistor pad in which two series resistors and a shunt resistor are arranged to form a tee.

tee switch A combination of three switches arranged to form a tee; two switches are in series with a shunt switch in between. If the series switches are open and the shunt switch is closed, isolation is greatly improved, compared to a single series switch.

Teflon FEP A plastic insulating material. Dielectric constant, 2.1. Dielectric strength, 2800 V/mil.

Teflon TFE A plastic insulating material. Dielectric constant, 2.2. Dielectric strength, 600 V/mil.

tel **1.** Abbreviation of TELEPHONE. **2.** Abbreviation of TELEGRAPH. **3.** Abbreviation of TELEGRAM.

telautograph A device that transmits and receives handwriting, drawings, and similar material. At the receiver, a pen follows the movements of a similar pen at the transmitter.

teleammeter A TELEMETER for measuring current generated at a remote point.

telecamera See TELEVISION CAMERA.

telecast A television program for general reception. The term a contraction of *television broadcast.*

telechir A remotely controlled, autonomous robot. See TELEOPERATION and TELEPRESENCE.

telecommunication Communication, usually between widely separated points, by electrical or electronic means.

telecontrol See REMOTE CONTROL and TELEOPERATION.

telefacsimile See FACSIMILE.

telegram Abbreviation, tel. A (usually printed-out) message transmitted and received via telegraph or teletypewriter. Compare CABLEGRAM and RADIOGRAM.

telegraph Abbreviation, tel. An instrument for transmitting and receiving messages by means of telegraphy. In its simplest form, it consists of a key and a sounder powered by a battery. Also see PRINTING TELEGRAPH, **1**, **2**.

telegraph channel **1.** The frequency band assigned to a particular telegraph station. **2.** The frequency band occupied by a telegraph signal.

telegraph code Broadly, Morse code. Wire telegraphy often uses a special version, such as the American Morse code.

telegraph key See KEY, **1**.

telegraph sounder See SOUNDER.

telegraph system A complete, integrated, coordinated arrangement of equipment for communication by means of telegraphy. Included are telegraph keys or keyers, sounders or printers, relays, switchboards, wire lines and cables, and power supplies.

telegraphy The branch of electrical communications that deals with the transmission and reception of messages by means of prearranged codes—especially over wires. Also see MORSE, **1**, **2**, **3**; MORSE CODE; and WIRE TELEGRAPHY.

telemeter **1.** An indicating instrument that measures the value of a quantity generated at a distant point or measures and transmits the value. **2.** The action afforded by the device in **1**.

telemetering See TELEMETRY.

telemeter receiver See TELEMETRIC RECEIVER.

telemeter transmitter See TELEMETRIC TRANSMITTER.

telemetric receiver A system that selects, amplifies, and demodulates or rectifies a radio signal, and actuates indicating instruments, recorders, or data processors.

telemetric transmitter A specialized transmitter that generates radio-frequency (RF) power, adds to it signals delivered by data transducers, and delivers the modulated power to an antenna for transmission to a distant telemetric receiver.

telemetry The transmission of data signals over a distance, either by radio or wire, and the reception and application of the signals to indicating instruments, recorders, etc.

teleoperation The remote control of autonomous robots. A human operator can control the speed, direction, and other movements of a robot from some distance away. See also TELEPRESENCE.

telephone Abbreviation, tel. An instrument for transmitting and receiving messages by means of telephony. In its simplest form, it consists of a microphone, earphone, switching and ringing devices, wire line or cable, and power supply. Also see HANDSET.

telephone accessories Devices (such as answering machines, speaker phones, facsimile machines, etc.) used in conjunction with a telephone set.

telephone amplifier A small audio amplifier, usually with a self-contained loudspeaker, for increasing the sound volume of a telephone output. Some amplifiers of this kind are connected to the telephone line, and others receive sound from the telephone receiver.

telephone-answering machine See ANSWERING MACHINE.

telephone bypass capacitors A set of fixed capacitors installed between each wire in a telephone line and an electrical ground to bypass radio-frequency energy. It can reduce or prevent radio-frequency interference (RFI) to telephone sets and data terminals.

telephone data set A device that converts signals from a data terminal for passage over a telephone circuit to a data-processing center.

telephone dialer In a security system, a circuit that automatically dials one or more telephone numbers, alerting the recipient(s) that an emergency exists at a given location.

telephone induction coil A small telephone-to-line impedance-matching transformer used in telephone systems.

telephone patch See PHONE PATCH.

telephone pickup A device for receiving conversations from a telephone to which it isn't directly connected.

telephone plug See PHONE PLUG.

telephone-radio patch See PHONE PATCH.

telephone receiver The handheld part of a telephone set, containing the microphone and earphone, and, in some cases, the dialing keypad.

telephone repeater An amplifier and associated equipment used to boost the amplitude of a telephone signal at an appropriate point along the line.

telephone service entrance The point at which the telephone wiring in a house or building connects to the outside telephone line.

telephone silencer A device for muting a telephone or its bell.

telephone system A complete, integrated, and coordinated arrangement of equipment for communication by means of telephony. Included are telephones, switchboards and associated equipment, wire lines and cables, and power supplies. Also see DIAL TELEPHONE SYSTEM, HANDSET, PRIVATE AUTOMATIC EXCHANGE (PAX), PRIVATE BRANCH EXCHANGE (PBX), and TELEPHONE.

telephone test set See PHONE TEST SET.

telephone transmitter The sound pickup unit (microphone) of a telephone. Also see TRANSMITTER, **2**.

telephony The branch of electrical communication dealing with the transmission and reception of sounds—especially over wires. Also see WIRE TELEPHONY.

Telephoto **1.** The transmission and/or reception of photographs by means of FACSIMILE. **2.** A photograph transmitted and/or received by means of FACSIMILE.

telephoto lens A camera lens providing a telescopic effect.

Telephoto receiver See FACSIMILE RECEIVER.

Telephoto transmitter See FACSIMILE TRANSMITTER.

telepresence An advanced form of robotic TELEOPERATION, in which a human operator has the impression of being at the robot's location. It includes vision systems, pressure sensors, sound sensors, tactile sensors, and electromechanical control devices. The operator uses, or wears, equipment similar to that used for VIRTUAL REALITY (VR).

teleprinter A terminal telegraph printing machine. Also see PRINTING TELEGRAPH, **1**, **2**.

TelePrompTer A device that presents a running display on a screen before a television announcer, performer, or speaker, of dialogue.

teleran A ground-to-air communications system. Ground-based radar pictures are transmitted, via television, to aircraft.

telescoping antenna A vertical antenna consisting of separate lengths of metal tubing of progressively smaller diameter so that one can slide into another. The antenna can be pulled out to full length or compressed to the length of the largest-diameter section.

telescoping antenna

teletext A method of communication in which a video image is sent of a page of printed material.

telethermoscope A device for measuring the temperature of distant objects.

Teletorque See AUTOSYN and SYNCHRO.

Teletype **1.** The transmission and/or reception of text messages in digital form, displayed or printed at a distant location. **2.** See TELETYPEWRITER.

teletype grade A term descriptive of a circuit having the quality necessary for communication via telegraphy.

Teletypesetter An electronic system for operating a distant Linotype.

teletypewriter A variety of printing telegraph using electric typewriters and associated equipment. The message is typed on the keyboard at the transmitting station and is typed out in corresponding letters at the receiving station. The same typewriter is able to send and receive messages. Also see RADIOTELETYPEWRITER.

teletypewriter exchange Abbreviation, TWX. A center for switching and routing teletypewriter communications. Also see TELETYPEWRITER.

teleview To observe a scene or program via television.

televise **1.** To convert a scene into a television signal. **2.** To broadcast a scene or program via television.

television **1.** Abbreviation, TV. The transmission and/or reception, via electromagnetic fields, wire cable, and/or fiberoptic cable, of images, usually with sound. **2.** A system for receiving signals, as defined in **1**. **3.** Video programs or data, with or without sound, transmitted via electromagnetic fields, wire cable, and/or fiberoptic cable.

television band See UHF TELEVISION CHANNELS and VHF TELEVISION CHANNELS.

television camera The pickup device that scans a scene and delivers a series of electrical signals that can be used to reconstruct the image on the screen of a picture tube.

television-camera tube See CAMERA TUBE, ICONOSCOPE, and ORTHICON.

television channel A radio-frequency band allocated exclusively for the transmission of a television signal. In conventional broadcasting, this band is usually 6 MHz wide. See, specifically, UHF TELEVISION CHANNELS and VHF TELEVISION CHANNELS.

television engineer A trained professional skilled in video electronics engineering, as well as in basic engineering and associated subjects.

television engineering The branch of electronics engineering devoted to the theory and application of television.

television interference Abbreviation, TVI. Interference to the reception of television signals, usually occasioned by signals from radio services or computers, or by electrical noise.

televisor **1.** A television transmitter or receiver. **2.** A person or entity responsible for the broadcast of television programs or signals.

televoltmeter A TELEMETER for measuring voltage generated at a remote point.

telewriter See TELAUTOGRAPH.

Telex **1.** A teleprinter system that operates via the telephone lines, and was once commonly used by businesses for sending and receiving short messages. It has largely been supplanted by computerized data communications systems. **2.** A hard-copy message sent or received by such a system.

telluric currents Also called *terrestrial currents.* A flow of electrical charge carriers, primarily electrons, in the earth.

tellurium Symbol, Te. A rare, metalloidal element related to selenium. Atomic number, 52. Atomic weight, 127.60.

TE mode See TRANSVERSE ELECTRIC MODE.

temp **1.** Abbreviation of *temperature.* (Also, T.) **2.** Abbreviation of *temporary.* **3.** Abbreviation of TEMPLATE.

temperament The tuning of a keyboard-type musical instrument to produce a nearly perfect diatonic scale.

temperature Symbol, T. A quantitative measure of the heat exhibited by an object or phenomenon. Also see THERMOMETER SCALE.

temperature coefficient A figure that states the extent to which a quantity drifts or varies under the influence of changing temperature. It is generally expressed in percent per degree (%/°C) or in parts per million per degree (ppm/°C).

temperature-compensated crystal oscillation Oscillation in a crystal oscillator, in which the crystal and/or circuit is automatically compensated against frequency drift caused by temperature change.

temperature-compensating component A circuit component, such as a capacitor or resistor, whose temperature coefficient is equal in magnitude and opposite in sign to that of a conventional component to which it is connected to cancel temperature-induced variation in the latter's value.

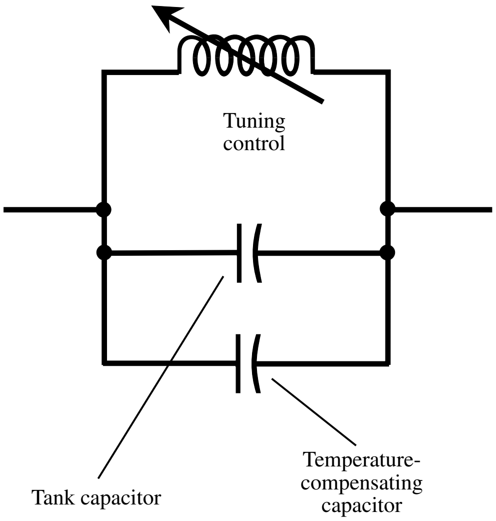

temperature-compensating component

temperature compensation The use of a device, such as a temperature-compensating component, to correct a temperature-induced deviation in the value of a conventional component.

temperature control **1.** The adjustment of temperature. **2.** The automatic maintenance of temperature at a desired level, as in a temperature-controlled oven. **3.** A device for controlling temperature, as defined in **1** or **2**.

temperature-controlled crystal oscillator Abbreviation, TCCO. A high-precision crystal oscillator in which the crystal plate (and sometimes the circuitry, as well) is held at constant temperature.

temperature degree See DEGREE, **2** and THERMOMETER SCALE.

temperature derating Deliberate reduction of operating current and/or voltage of a device to a specific temperature to ensure proper operation. Also see DERATING, DERATING CURVE, and DERATING FACTOR.

temperature gradient A range of temperature variation, such as the rate of change of temperature, with respect to change of power dissipation, or the rate of change of temperature with spatial displacement.

temperature inversion See INVERSION, **1**.

temperature meter An indicator (usually a direct-current voltmeter or millivoltmeter) whose scale reads directly in degrees.

temperature scale See THERMOMETER SCALE.

temperature-sensitive resistor See THERMISTOR.

temperature shock See THERMAL SHOCK.

temperature-to-voltage converter A circuit or device, such as a thermocouple, that delivers an output voltage proportional to an applied temperature.

template **1.** A diagram, usually drawn on paper, to show the locations at which components should be placed or tasks should be performed. It is taped or cemented temporarily to the work, and the points are transferred to the latter by prick-punching. **2.** A stencil-like plate with alphanumeric and circuit symbols, used as a drafting aid. Sometimes called "drafting stencil."

temporary magnet **1.** A body that exhibits magnetism only briefly after it has been exposed to another magnet. Compare PERMANENT MAGNET. **2.** See ELECTROMAGNET.

temporary storage **1.** In computer and data-processing operations, the storage of data or instructions only until they are needed. Also called INTERIM STORAGE. **2.** Locations in a computer memory set aside during a program run for holding intermediate results of operations.

TEM wave See TRANSVERSE ELECTROMAGNETIC WAVE.

ten code A set of abbreviations used by two-way radio operators. Each "ten signal" represents a specific statement or query.

tension **1.** A force that tends to stretch, pull tight, or pull apart. Compare STRAIN. **2.** A term referring to VOLTAGE—especially in a utility power transmission line.

ten-turn potentiometer A precision potentiometer whose shaft must be turned through 10 complete revolutions to cover the entire resistance range. Also see HELICAL POTENTIOMETER and MULTITURN POTENTIOMETER.

T-equivalent circuit See TEE-EQUIVALENT CIRCUIT.

tera- Abbreviation, T. **1.** A prefix meaning trillion(s), (i.e., 10^{12}). **2.** A prefix meaning 2^{40}.

tera-electronvolt Abbreviation, TeV. A large unit of energy; 1 TeV = 10^{12} electronvolts. Also see ELECTRONVOLT.

terahertz Abbreviation, THz. A unit of extremely high frequency equal to 10^{12} Hz. Also called *Fresnel.*

teraohm A unit of high resistance, reactance, or impedance equal to 10^{12} ohms.

terawatt Abbreviation, TW. A large unit of power; 1 TW = 10^{12} W.

terbium Symbol, Tb. A metallic element of the rare-earth group. Atomic number, 65. Atomic weight, 158.93.

terbium metals A group of rare-earth metals, including *europium*, *gadolinium*, *terbium*, and occasionally *dysprosium.*

term In an algebraic expression, constants, variables, or combinations of these, separated by operation signs (e.g., the expression $4xy + z$ has two terms).

terminal **1.** A connection point at the input, output, or an intermediate point of a device, or a point at which a voltage is to be applied. **2.** A metal tab or lug attached to the end of a lead for connection purposes. **3.** Pertaining to the end of a series of events, etc. (e.g., terminal tests). **4.** In a data-communications system, a point of data input or output. Also called *data terminal.* If it does not have computing capability of its own, it is often called a *dumb terminal.*

terminal block A group of several terminals, intended for interconnection of circuits, mounted on a solid insulating block.

terminal board An insulating board carrying several lugs, tabs, or screws as terminals (see TERMINAL, **2**). Also see TERMINAL STRIP.

terminal guidance The navigation of a missile or aircraft to help it reach its target or destination.

terminal impedance The internal impedance of a device measured at the input or output terminals.

terminal point of degradation The point at which degradation of a circuit or component is complete. Also see DEGRADATION FAILURE.

terminal repeater A telephone repeater operated at the end of a line.

terminal strip A strip of insulating material, such as plastic or ceramic, on which are mounted one or more screws, lugs, or other terminals. Also see TERMINAL, **2**.

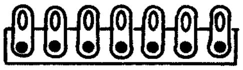

terminal strip (lug-type)

terminal voltage The voltage at the output terminals of an unloaded battery or generator.

ternary code See TRINARY NUMBER SYSTEM.

ternary fission The splitting of an atomic nucleus into three nuclear pieces. Also see FISSION.

ternary number system See TRINARY NUMBER SYSTEM.

terrain echoes Radar images caused by reflections from hills, mountains, and other natural terrestrial surface features. Also see GROUND CLUTTER.

terrestrial magnetism See EARTH'S MAGNETIC FIELD.

tertiary coil A third winding (see TERTIARY WINDING).

tertiary winding A third winding on a transformer or magnetic amplifier.

tesla Symbol, T. A unit of magnetic flux density; 1 T = 1 weber per square meter = 10^4 gauss.

Tesla coil A special type of air-core step-up transformer for developing high voltage at radio frequencies. It consists essentially of a low-turn primary coil, through which radio-frequency (RF) current flows, and a multiturn secondary coil, across which the high voltage, is developed.

test A procedure consisting of one or several steps, in which (1) the mode of operation of a circuit or device is established, (2) the value of a component is ascertained, or (3) the behavior of a circuit or device is observed.

test bench An equipment installation intended for the testing, repair, or debugging of electronic devices by a technician.

test data Data used to test a computer program, including samples within limits that might be encountered during the program's implementation.

tester **1.** See TEST INSTRUMENT. **2.** A technician who primarily makes tests and measurements.

testing window See WINDOW, **2**.

test instrument A device for checking the operation of a circuit or the value of a component. This class of instrument is usually less accurate than measurement instruments. Also see TEST SET.

test lead The flexible, insulated wire attached to a test prod.

test pattern A picture-and-line display on the screen of a television picture tube, used to check such features as aspect ratio, linearity, contrast, etc.

test point A terminal intended for connection of test equipment in the repair or debugging of a circuit. Often, test points are labeled by the letters TP followed by numerals (such as TP1, TP2, etc.).

test probe See PROBE, **1**.

test prod A stick-type probe (see PROBE, **1**) with a flexible, insulated lead terminating in a plug or lug for attachment to an instrument.

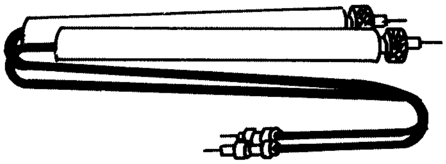

test prods

test program In computer operations, a program devised to check the functioning of hardware. Also called *utility* and *test routine.*

test run In computer operations, using test data to check the operation of a program, by comparing the results obtained thereby with what should ideally result.

test set A combination of instruments assembled on a single panel, and usually enclosed in a carrying case, for convenience in making tests.

test signal **1.** A signal used for conducting a test of a component, circuit, or system. **2.** In radiotelegraphy, a special signal signifying that the transmitting station is testing equipment. Also see VEE SIGNAL.

test-signal generator A device, such as an oscillator, for producing a signal for testing equipment (see TEST SIGNAL, **1**).

test tape A magnetic tape containing signals for testing equalization, frequency response, head adjustment, stereo balance, etc. in an audiotape recorder, high-fidelity sound system, computer tape drive, or videocassette recorder.

tetravalent See QUADRIVALENT.

tetrode An electron tube in which the principal electrodes are cathode, control grid, screen, and plate.

tetrode transistor **1.** A bipolar transistor with two emitters. **2.** A dual-gate field-effect transistor.

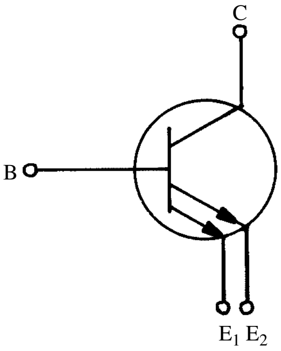

tetrode transistor

TeV Abbreviation of TERA-ELECTRONVOLT.

T_e value The temperature at which a centimeter cube of glass or ceramic exhibits 1 megohm of resistance.

TE wave See TRANSVERSE ELECTRIC MODE.

text editor A computer program for finding and changing data in a file.

texture map In computer graphics, the topographical qualities of the surface of a three-dimensional rendition. For example, the image of an orange would have a "bumpy" texture map; the image of a peach would have a "fuzzy" texture map.

texture sensing The ability of a robotic end effector to determine the relative smoothness or roughness of a surface. One common scheme uses reflected light or infrared. Scattered reflections indicate a rough or matte surface.

tgt Abbreviation of TARGET.

TH Abbreviation of *true heading.*

Th Symbol for THEORIUM.

TH Symbol for *heater temperature.*

thallium Symbol, Tl. A metallic element. Atomic number, 81. Atomic weight, 204.38.

thalofide cell An evacuated photoconductive cell using thallium oxysulfide as the light-sensitive material.

THD Abbreviation of TOTAL HARMONIC DISTORTION.

theory A reasonable proposition put forth to account for the behavior of, or the relationships between, bodies and forces, or to explain concepts and their relations. When a theory has stood up under exhaustive tests, it might reveal a scientific law.

therm A gas heating unit. 1 therm = 100,000 British thermal units (10^5 Btu).

thermal agitation Random movement of particles (such as electrons) in a substance, because of heat.

thermal-agitation noise See THERMAL NOISE.

thermal ammeter See HOT-WIRE AMMETER.

thermal anemometer See HOT-WIRE ANEMOMETER.

thermal conductivity The heat-conducting ability of a material. Compare electrical conductivity (see CONDUCTIVITY).

thermal conductivity

Element	*Thermal conductivity (mW/m/°C)*
Aluminum	22.0
Carbon	2.4
Chrominum	6.9
Cooper	39.0
Gold	30.0
Iron	7.9
Lead	3.5
Magnesium	16.0
Mercury	0.8
Nickel	8.9
Platinum	6.9
Silicon	8.4
Silver	41.0
Thorium	4.1
Tin	6.4
Tungsten	20.0
Zinc	11.0

thermal-conductivity device An instrument or control unit using a heated filament whose temperature and, accordingly, conductivity is varied by some sensed phenomenon. See, for example, GAS DETECTOR, HEATED-WIRE SENSOR, HOT-WIRE ANEMOMETER, HOT-WIRE FLOW-METER, and HOT-WIRE MICROPHONE.

thermal-conductivity gasmeter See GAS DETECTOR.

thermal detector **1.** See BOLOMETER. **2.** In a security or fire-protection system, a device that closes a circuit or actuates an alarm if the temperature rises to a specific level.

thermal emf See SEEBECK EFFECT.

thermal gasmeter See GAS DETECTOR.

thermal imaging See THERMOGRAPHY.

thermal instrument See HOT-WIRE METER and THERMOCOUPLE-TYPE METER.

thermally sensitive resistor See THERMISTOR.

thermal meter See HOT-WIRE METER and THERMOCOUPLE-TYPE METER.

thermal neutron A neutron that is essentially in thermal equilibrium with the surrounding medium or environment.

thermal noise Frequency-independent electrical noise caused by the agitation of particles (e.g., atoms and electrons) in a material by heat. Thermal noise is proportional to bandwidth, resistance, and absolute temperature.

thermal radiation See HEAT.

thermal recorder A graphic recorder in which a strip of paper coated with a thin layer of opaque wax is drawn between a knife-edge platen and a heated writing stylus that melts the wax beneath its point, exposing the underlying black paper as a fine line.

thermal resistance Symbol, *RT*. For a semiconductor device, the rate of change of junction temperature, with respect to power dissipation; $RT = \mathrm{d}T/\mathrm{d}P$, where *RT* is in degrees Celsius per milliwatt, *T* is the temperature in degrees Celsius, and *P* is the power in milliwatts.

thermal resistor A resistor that is sufficiently temperature-sensitive to be used as a heat sensor. Examples: thermistor and germanium diode.

thermal response time For a power-dissipating component, the elapsed time between the initial change in power dissipation and the moment at which the temperature has changed by a specified percentage (usually 90%) of the total value.

thermal runaway A destructive process resulting from cumulative temperature effects. In bipolar transistors, this can occur if the collector current increases as the temperature rises. As the unit gets hotter, the collector-base junction dissipation increases, generating still more heat. The ultimate result, if the process continues unchecked, is destruction of the component. The process can also occur in certain batteries when they are charged too rapidly or at excessively high temperatures.

thermal shock The effect of applying heat or cold to a device so rapidly that abnormal reactions occur, such as rapid (often catastrophic) expansions and contractions.

thermal switch A switch actuated by a temperature change. Types vary from the simple thermostat to complex servosystem switches with a temperature-transducer input.

thermal time-delay relay A delay relay utilizing the slow-heating and slow-cooling property of one of its components.

thermion An ion or electron emitted by a hot body, such as the heated cathode of a vacuum tube.

thermionic Pertaining to thermions and their applications.

thermionic cathode A heated cathode, in contrast to a cold cathode, used as an emitter of electrons or ions. Also see THERMION.

thermionic current Current caused by thermionic emission—especially in an electron tube.

thermionic detector A vacuum-tube detector. Also see THERMION and THERMIONIC EMISSION.

thermionic diode A hot-cathode diode tube.

thermionic emission The emission of electrons by a hot body, such as the filament or cathode of a vacuum tube. Also see THERMION and HOT CATHODE.

thermionics The study of electron emission from objects or materials at high temperature.

thermionic tube An electron tube (i.e., one in which electrons or ions are emitted by a heated cathode). Also see THERMION, THERMIONIC CURRENT, and THERMIONIC EMISSION.

thermionic work function The energy required to force an electron from inside a heated cathode into the surrounding space (in thermionic emission). Also see WORK FUNCTION.

thermistor A temperature-sensitive resistor, usually made from specially processed oxides of cobalt, magnesium, manganese, nickel, uranium, or mixtures of such substances. Thermistors are available with either a positive or negative temperature coefficient of resistance. The name is a contraction of *thermally sensitive resistor.*

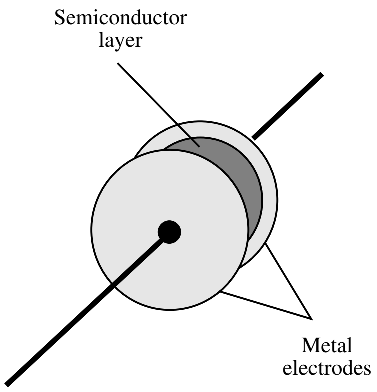

thermistor

thermistor bridge A four-arm bridge (see BRIDGE, **1**) in which one arm is a thermistor and, therefore, is temperature sensitive.

thermistor power meter A radio-frequency power-measuring instrument based on a thermistor bridge.

thermistor probe A temperature probe containing a thermistor as the sensing element.

thermistor thermometer An electronic thermometer in which the temperature-sensitive element is a thermistor.

thermoammeter See THERMOCOUPLE-TYPE METER.

thermocouple A device consisting essentially of a bond between two wires or strips of dissimilar metals (such as antimony and bismuth). When the bond is heated, a direct-current voltage appears across it.

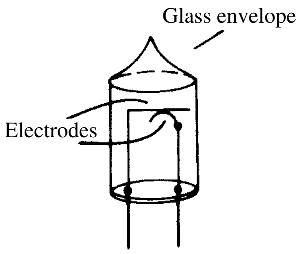

thermocouple

thermocouple bridge A four-arm bridge (see BRIDGE, **1**) in which one arm is a thermocouple; therefore, it is temperature sensitive, as well as being voltage-productive.

thermocouple meter See THERMOCOUPLE-TYPE METER.

thermocouple-type meter A radio-frequency (RF) meter consisting of a thermocouple and direct-current (dc) ammeter, milliammeter, or microammeter, connected in series. The thermocouple is heated, directly or indirectly, by an applied RF current, and the resulting dc output deflects the meter pointer.

thermodynamics The science dealing with the relationships between heat and mechanical energy and their interconversion. Also see CARNOT THEOREM, FIRST LAW OF THERMODYNAMICS, SECOND LAW OF THERMODYNAMICS, and THIRD LAW OF THERMODYNAMICS.

thermoelectric cooler A cooling device based on the Peltier effect.

thermoelectric effect The production of thermoelectricity by certain materials.

thermoelectricity Heat-produced electricity, as in thermocouple operation.

thermoelectric junction A junction between two conductors that exhibits variable characteristics under conditions of changing temperature.

thermoelectric series A series of conducting metals, sequenced so that if a thermocouple is made from two of the metals, current flows at the hot junction from the metal occurring earlier in the series.

thermoelectron An electron emitted by a hot cathode. Also see THERMION and THERMIONIC EMISSION.

thermoelement A thermocouple, especially a miniature one, used in a THERMOCOUPLE-TYPE METER.

thermogalvanometer See THERMOCOUPLE TYPE METER.

thermography A means of locating or mapping by detecting infrared images. It can be in the form of live video, similar to television, or in the form of photographs, using infrared film. Many common cameras can be used with infrared film to make infrared photographs (thermographs).

thermojunction The junction of the two metals in a thermocouple. A voltage appears when the dissimilar metals are heated.

thermoluminescence Luminescence resulting from the moderate heating of certain materials.

thermomagnetic effect The effects of temperature on the magnetism of a body, or vice versa.

thermometer A (usually direct reading) device for measuring temperature. Also see ELECTRONIC THERMOMETER and THERMOMETER SCALE.

thermometer scale The scale on a thermometer, graduated in degrees, from which temperature is read. For a description of different scales, see ABSOLUTE SCALE, CENTIGRADE SCALE, FAHRENHEIT SCALE, REAUMUR SCALE, and RANKINE SCALE. The Kelvin scale is the same as the absolute scale, and the Celsius scale is the same as the centigrade scale.

thermonuclear reaction A nuclear reaction in which energy is released when lighter atoms are converted into heavier atoms at temperatures in the millions of degrees Celsius. Also see FISSION, FUSION, and NUCLEAR REACTOR.

thermopile A device consisting of two or more thermocouples connected in series for increased voltage output.

thermoplastic material A plastic that can be resoftened by applying heat after having been molded into a desired shape. Example: polystyrene. Compare THERMOSETTING MATERIAL.

thermorelay See THERMOSTAT.

thermosensitivity Sensitivity of a circuit or device to heat.

thermosetting material A plastic that cures chemically (will not ordinarily soften again when heat is applied) after having been heat-molded into a desired shape. Example: Bakelite. Compare THERMOPLASTIC MATERIAL.

thermostat A temperature-sensitive switch. In one common form, a movable contact is carried by a strip of bimetal and the stationary contact is mounted nearby.

thermostatic switch See THERMOSTAT.

thermoswitch See THERMOSTAT.

theta wave A form of brain wave that occurs at extremely low frequencies, and is associated with sleep or mental incoherence.

Thevenin's theorem The proposition that, with reference to a particular set of terminals, a network containing a number of generators and constant impedances can be simplified to a single generator in series with a single impedance. The equivalent circuit will deliver to a given load, the same current, voltage, and power delivered by the original network. Compare COMPENSATION THEOREM, MAXIMUM POWER TRANSFER THEOREM, NORTON'S THEOREM, RECIPROCITY THEOREM, and SUPERPOSITION THEOREM.

thick film A film of selected material (conductive, resistive, dielectric, etc.) applied to a substrate by painting, photography, or similar process. See, for example, PRINTED CIRCUIT. Compare THIN FILM. Typically, thick films are 1 mil or more in thickness.

thick-film component A unit, such as a capacitor or resistor, fabricated by thick-film techniques. See, for example, PRINTED COMPONENT. Also see THICK FILM. Compare THIN-FILM COMPONENT.

thick-film resistor A resistor fabricated by thick-film techniques. See, for example, PRINTED RESISTOR. Also see THICK FILM. Compare THIN-FILM COMPONENT.

thick magnetic film See MAGNETIC THICK FILM.

thin film An extremely thin layer (less than 1 mil) of a selected material (conductive, resistive, semiconductive, dielectric, etc.) electrodeposited or grown on a substrate. Compare THICK FILM.

thin-film capacitor A capacitor made by electrodepositing a thin film of metal on each side of a grown layer of oxide, as in an integrated circuit. Also see THIN FILM.

thin-film component A unit (such as a capacitor, resistor, diode, or transistor), fabricated by thin-film techniques. Also see THIN FILM. Compare THICK-FILM COMPONENT.

thin-film integrated circuit An integrated circuit in which the components and "wiring" are produced by depositing (or growing) and processing materials on a semiconductor slab or wafer (the substrate). Compare HYBRID INTEGRATED CIRCUIT and MONOLITHIC INTEGRATED CIRCUIT.

thin-film memory In a computer, a storage medium that is a magnetic thin film (see THIN FILM) on a nonmagnetic substrate (often glass) and that can be magnetized to represent digital data.

thin-film microelectronic circuit An integrated circuit that occupies (essentially) two dimensions; that is, a very thin integrated circuit.

thin-film resistor A resistor fabricated by thin-film techniques (see THIN FILM) (e.g., TANTALUM-NITRIDE RESISTOR). Compare THICK-FILM RESISTOR.

thin-film semiconductor A very shallow layer of semiconductor material, such as single-crystal silicon, electrodeposited on a substrate. Also see THIN FILM.

thin-film transistor A transistor fabricated by thin-film techniques. Also see THIN FILM and THIN-FILM COMPONENT.

thin magnetic film See MAGNETIC THIN FILM.

third-generation computer A computer in which the active discrete components are integrated circuits.

Third Law of Thermodynamics As the temperature of absolute zero is approached in an isothermal process involving a solid or liquid, the change in entropy approaches zero; and the entropy of a substance is zero at absolute zero.

third-octave band See ONE-THIRD-OCTAVE BAND.

thirty-channel multiplex A form of pulse-code-modulated (PCM) multiplex using eight digits and the A-law. There are 30 speech channels and two utility channels.

Thomson bridge See KELVIN BRIDGE.

Thomson effect The liberation or absorption of heat (depending on the material of interest) when an electric current flows from a warmer to a cooler part of a conductor.

Thomson heat The amount of thermal energy transferred because of Thomson effect.

Thomson voltage The voltage drop between two points on a conductor that are at different temperatures.

thoriated-tungsten filament In a vacuum tube, a filament made of tungsten to which thorium oxide has been added to increase the emission of electrons. Also see THERMION and THERMIONIC EMISSION.

thorium Symbol, Th. A radioactive metallic element. Atomic number, 90. Atomic weight, 232.04. Thorium, when heated, is a copious emitter of electrons, so the filaments or cathode cylinders of some electron tubes are coated with one of its compounds.

thoron Symbol, Tn. A radioactive isotope of RADON.

three-address instruction A computer program instruction having three addresses, two for operands and one for the result (of the operation called for).

three-channel stereo A form of stereophonic sound recording and reproduction in which three distinct channels are used; these are usually designated the left, right, and center channels.

three-coil meter See ELECTRODYNAMOMETER.

three-conductor jack A female connector in which two separate conductors are provided, in addition to the ground conductor.

three-conductor plug A male connector in which two separate conductors are provided, in addition to the ground conductor.

three-dimensional television See STEREOSCOPIC TELEVISION.

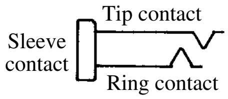

three-conductor jack

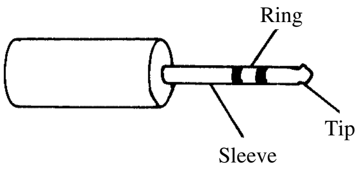

three-conductor plug

three-electrode tube See TRIODE.

three-element tube See TRIODE.

three-gun picture tube A color-television picture tube having a separate gun for each primary color (red, green, and blue).

three-halves-power law See CHILD'S LAW.

three-junction transistor A pnpn or npnp transistor. Also see NPNP DEVICE.

three-phase bridge rectifier A bridge-rectifier circuit for three-phase alternating current. Two diodes are provided for each phase. The ripple frequency is six times the line frequency. Also see BRIDGE RECTIFIER, POLYPHASE RECTIFIER, and THREE-PHASE CIRCUIT.

three-phase circuit The circuit of a three-phase system. See THREE-PHASE SYSTEM and, specifically, DELTA CONNECTION and WYE CONNECTION.

three-phase current Current in a three-phase circuit. The currents in the three legs differ in phase by 120°.

three-phase four-wire system See FOUR-WIRE WYE SYSTEM.

three-phase generator A (usually dynamo-type) generator of three-phase current or voltage. See THREE-PHASE SYSTEM.

three-phase half-wave rectifier A half-wave rectifier circuit for three-phase alternating current. One diode is provided for each phase. The ripple frequency is three times the line frequency. Also see HALF-WAVE RECTIFIER, POLYPHASE RECTIFIER, and THREE-PHASE CIRCUIT. Compare THREE-PHASE BRIDGE RECTIFIER.

three-phase motor An alternating-current motor operating on three-phase power. Above fractional-horsepower size, the three-phase motor is smoother running and more simply structured than the single-phase counterpart.

three-phase power The total power dissipated or delivered in a three-phase alternating-current circuit.

three-phase rectifier A rectifier for three-phase alternating current. At least one diode is included for each phase. Also see POLYPHASE RECTIFIER, THREE-PHASE BRIDGE RECTIFIER, THREE-PHASE CIRCUIT, and THREE-PHASE HALF-WAVE RECTIFIER.

three-phase system An alternating-current electrical system in which three currents or voltages exist simultaneously. They are of equal amplitude, but are 120° out of phase with each other.

three-phase, three-wire system An electrical system having three conductors, with an alternating-current phase difference of 120° between conductor pairs.

three-phase-to-single-phase transformer An alternating-current transformer with three-phase input and single-phase output (one terminal is grounded).

three-phase-to-two-phase transformer An alternating-current transformer with three-phase input and two-phase output, the output currents are 180° out of phase with each other.

three-phase voltage Voltage in a three-phase alternating-current circuit. The voltages across the three legs differ by 120°.

three-quarter bridge A bridge rectifier having diodes in three arms and a resistor in the fourth.

three-space A mathematical continuum in which each point is uniquely defined by three variables in an ordered triple, such as (x,y,z), and each ordered triple corresponds to exactly one point in the space. A common coordinate system is the Cartesian system consisting of three axes, x, y, and z—each mutually perpendicular and intersecting at the origin (0,0,0).

three-space coordinates Any set of coordinates used for locating points or plotting graphs in three dimensions.

three-state logic See TRI-STATE LOGIC.

three-way speaker A set of three individual speakers contained in a single cabinet: a woofer (low-frequency speaker), a midrange speaker, and a tweeter (high-frequency speaker). It is common in high-fidelity music reproduction systems. Compare TRIAXIAL SPEAKER.

three-wire system **1.** An electric-power feed system using three wires, the center one (neutral) being at a potential midway between the potential across the other (outer) two. **2.** See THREE-PHASE and THREE-WIRE SYSTEM. **3.** See TWO-PHASE and THREE-WIRE CIRCUIT.

threshold **1.** The initial (observable) point of an effect (e.g., *threshold of hearing*). **2.** A predetermined point, such as of minimum current or voltage, for the start of operation of a circuit or device.

threshold component A value of current, voltage, sound intensity, etc., selected as the minimum level at which a circuit or device is to operate in some prescribed manner, or beyond which a certain condition will prevail. Also see THRESHOLD, **1**, **2**.

threshold current **1.** The minimum value of current at which a certain effect takes place. **2.** The smallest amount of forward current that flows through a diode. **3.** In a gas, the smallest level of current for which a discharge will maintain itself under variable conditions.

threshold detector A device that prevents a signal from passing until its peak amplitude reaches a certain value.

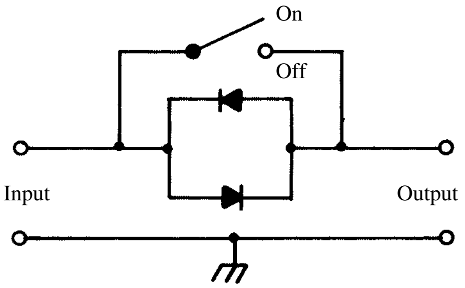

threshold detector

threshold frequency **1.** The cutoff frequency or frequencies in a band-pass, band-rejection, low-pass, or high-pass filter. **2.** The lowest frequency for a metal or semiconductor junction at which incident radiation gives rise to the photoelectric effect.

threshold of hearing The minimum intensity level at which sounds are audible in an environment containing essentially no background acoustic noise.

threshold of pain The intensity level at which hearing a sound causes physical discomfort. This is approximately 120 dB above the THRESHOLD OF HEARING for most people.

threshold signal The weakest signal that can be detected in a receiving system.

throat See HORN THROAT.

throat microphone A small microphone operated in contact with the user's throat.

throttle The feedback control device in a regenerative detector or amplifier.

throughput In computer operations, an overall quantitative indicator of processing power, expressed in terms of the amount of data processed in a given period of time.

throw-out spiral On a phonograph disc, a lead-out groove.

thulium Symbol, Tm. A metallic element of the rare-earth group. Atomic number, 69. Atomic weight, 168.93.

thumbwheel potentiometer A potentiometer operated by means of a knurled knob (usually protruding perpendicularly through a panel) that is turned with the thumb or with a finger.

thumbwheel switch A switch operated by means of a knurled knob (usually protruding perpendicularly through a panel) that is turned with the thumb or with a finger.

thy Abbreviation of THYRATRON.

thyratron A gas triode or gas tetrode used principally for switching and control purposes. Thyratron action differs from that of the vacuum tube in the unique behavior of the thyratron control electrode (grid): Anode current starts to flow abruptly when grid voltage reaches a particular value, at which point the grid provides no further control; anode current continues to flow until the anode voltage is either interrupted or reversed.

thyratron inverter An inverter circuit (see INVERTER, **1**) using thyratrons as the switching devices. Also see THYRATRON.

Thyrector A silicon diode exhibiting very high resistance (approaching that of an insulator) up to a certain voltage, beyond which the unit switches to a low-resistance conducting state.

thyristor **1.** A pnpn-type bistable semiconductor device having anode, cathode, and gate terminals that is used as an electronic switch. **2.** The generic term for all thyratronlike solid-state devices, such as the silicon-controlled rectifier.

Thyrite Ceramic silicon carbide, a nonlinear resistance material, or a resistor made of this material. The resistance of Thyrite decreases sharply as the applied voltage is increased. Thyrite resistors are used in voltage regulators, equipment protectors, lightning arresters, curve changers, and similar devices.

THz Abbreviation of TERAHERTZ.

Ti Symbol for TITANIUM.

tickler A (usually small) coil, through which energy is inductively fed back from the output to the input of a circuit to induce oscillation.

tickler coil See TICKLER.

tickler-coil regeneration Positive feedback obtained via inductive coupling between a small coil (tickler) in the output circuit of an amplifier, and a (usually larger) coil in the input circuit. Also see TICKLER.

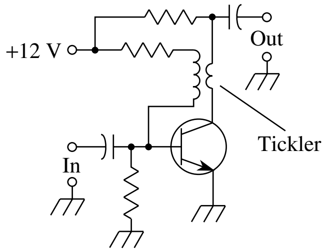

tickler-coil regeneration

tickler oscillator An oscillator circuit in which positive feedback is obtained through inductive coupling between an output (tickler) coil and an input coil.

tie A bracket, clamp, clip, ring, or strip for holding several wires tightly as a cable or bundle.

tie cable **1.** A cable that connects two distributing points in a telephone system. **2.** Any cable that interconnects two circuits.

tie point A lug, screw, or other terminal to which wires are connected at a junction.

tie-point strip A terminal strip with lugs to which conductors can be soldered.

tight coupling See CLOSE COUPLING.

tilt switch A device, such as a mercury switch, that is actuated by tilting it to a certain angle.

timbre The quality that distinguishes the sound of one voice or instrument from that of another, largely because of harmonic content.

time Symbol, *t.* **1.** The instant at which an event occurs. **2.** The instant at which a time-base variable reaches a given value. **3.** The interval between two instants, commonly called *duration* or *length of time.* Also see STANDARD TIME, TIME BASE, TIME ZONE, and ZERO TIME.

time base Time as the independent variable in a physical relation or function. It appears in expressions such as *pulses per second*, *feet per minute*, *watts per hour*, etc.

time compressor In audio recording and reproduction, a device that speeds up or slows down the tempo without changing the audio frequencies. It is used for special effects.

time constant See ELECTRICAL TIME CONSTANT and MECHANICAL TIME CONSTANT.

time delay See DELAY, **1**, **2**.

time-delay relay See DELAY RELAY.

time-delay spectrometry In acoustics, a method of simulating an echo-free environment within an enclosure that is actually not echo-free.

time-division multiplexing Abbreviation, TDM. In data and computer communications, a time-sharing technique in which several terminals use the same channel by transmitting data at regular, staggered intervals (i.e., one is active while the others are idle). This gives the appearance of simultaneous real-time operation.

time-division-multiplex switch A switch with multiple ports—each port corresponding to a certain time slot in a time-division-multiplex scheme. The input and output (send and receive) ports are connected by selecting the same TIME SLOT for each.

time-domain reflectometry Abbreviation, TDR. Measuring the reflective characteristics of a device or system by superimposing the direct and reflected components of a step-formed test signal on a time-calibrated oscilloscope screen.

time duration See TIME, **3**.

time-duration modulation See PULSE-DURATION MODULATION.

time factor The ratio *ta*/*tr*, where *ta* is analog time (the relativistic duration of an event as simulated by a computer), and *tr* is real time (the actual duration of the event). Also called TIME SCALE.

time-interval mode In computer operations, operation that allows a number of events to be counted between two points on a waveform.

time modulation Any form of modulation in which the instantaneous characteristics of a signal are varied.

timeout The expiration of an allotted time period for a given operation.

timer **1.** A device for automatically controlling the duration of an operation. See ELECTRONIC TIMER. **2.** A device for measuring the duration of an operation.

time sharing In computer operations, a method for interlaced (i.e., nearly simultaneous) use of a machine or facility by two or more persons or agencies. As the cost of computers diminishes, so does the need for time sharing. Also see TIME-DIVISION MULTIPLEXING.

time-sharing dynamic allocator In computer storage, a program that allocates memory areas and peripherals to programs being entered into a time-sharing system; it also controls program execution.

timeshift **1.** To receive a message at a significantly later time than when it was sent. **2.** To use a videocassette or videodisc recorder to view a television program at a later time than when it was transmitted.

time-shifting communications Any form of communications in which the recipient reads or views the message(s) at a significantly later time than the message is sent from the source. A common example is electronic mail (e-mail) using online computer networks.

time signals Special radio transmissions made under the auspices of the National Bureau of Standards, for indicating Coordinated Universal Time (UTC).

time slot A specifically defined time interval in a data signal. It is of importance primarily in digital communications, where a given time interval can be high or low or at some discrete value.

time-space-time switch A large switch consisting of a space block between two time blocks.

time zone One of the 24 zones into which the global map is divided for the purpose of standardizing time. Within these zones, mean solar time is determined in terms of distance east or west of the zero meridian at Greenwich (near London, England). Each zone is equal to 15 degrees of longitude, or 1 hour. Four zones fall within the continental United States: Eastern Standard Time (zone of the 75th meridian), Central Standard Time (zone of the 90th meridian), Mountain Standard Time (zone of the 105th meridian), and Pacific Standard Time (zone of the 120th meridian). Also see MERIDIAN, ZERO MERIDIAN, and ZONE TIME.

timing extraction The retrieval of a timing signal from incoming data.

timing pin current Measured in microamperes. The current that generates a timing waveform in an integrated-circuit voltage regulator.

timing signal A repeating signal sent along with data to control the synchronization of transmitter and receiver.

tin **1.** Symbol, Sn. A metallic element. Atomic number, 50. Atomic weight, 118.71. Tin is widely used in electronics as a structural material, a constituent of solder, and (in foil form) as the plates of some fixed capacitors. **2.** To prepare the tip of a soldering gun or iron, or the stripped end of a wire or cable, by applying a coat of solder.

tin-oxide resistor See METAL-OXIDE RESISTOR.

tinsel Metal film strips interwoven with fabric threads to provide a flexible cord, particularly for headphones.

tint control In a color-television receiver, the control for changing color hue.

tip jack The mating connector for a tip plug.

tip plug **1.** A prod terminating in a phone tip. **2.** A plug-type connector terminating in a phone tip.

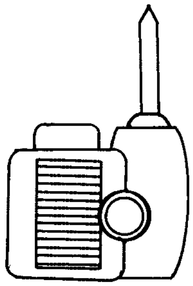

tip plug, 2

titanium Symbol, Ti. A metallic element. Atomic number, 22. Atomic weight, 47.88. Titanium enters into some dielectric compounds (e.g., titanium dioxide).

titanium dioxide Formula, TiO_2. A ceramic dielectric material. Dielectric constant, 90 to 170. Dielectric strength, 100 to 210 V/mil.

T junction See TEE JUNCTION.

Tl Symbol for THALLIUM.

T^2L Abbreviation of TRANSISTOR-TRANSISTOR LOGIC. (Also, TTL.)

TLA Abbreviation of TABLE LOOK-AT.

TLC Abbreviation of *thin-layer chromatography.*

TLU Abbreviation of TABLE LOOK-UP.

TM **1.** Abbreviation of TRANSVERSE MAGNETIC. **2.** Abbreviation of *technical manual.*

Tm Symbol for THULIUM.

TM mode See TRANSVERSE MAGNETIC MODE.

TM wave See TRANSVERSE MAGNETIC MODE.

T network See TEE NETWORK.

TNS Abbreviation of TRANSCUTANEOUS NERVE STIMULATOR.

toggle A bistable device.

toggle switch A switch having a mechanism that snaps into the on or off position at the opposite extremes to which a lever is moved.

tolerance The amount by which error is allowed in a value, rating, dimension, etc. It is usually

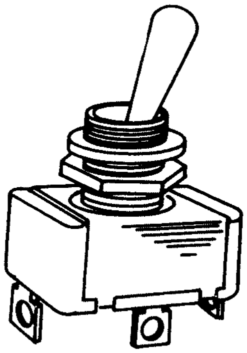

toggle switch

expressed as a percent of nominal value, plus and minus so many units of measurement (or parts per million).

toll call In a telephone system, a call that is charged on a per-minute or per-second basis.

ton Abbreviation, T, t, or tn. A unit of avoirdupois weight; in the United States, it is usually taken to mean *short ton,* a unit equal to 907.20 kilograms (2000 lb). Compare METRIC TON.

tone **1.** The pitch (frequency) and timbre (relative harmonic content) of a sound other than noise. **2.** A sound consisting of a periodic waveform having constant frequency; also called *note.*

tone arm See PICKUP ARM.

tone burst A test signal consisting of a single-frequency sine wave sustained for a brief period of time, usually having a rectangular envelope (rapid rise and decay).

tone-burst entry In repeater systems, a technique whereby a short tone signal is used at the start of a transmission to trigger a particular repeater so that all repeaters in the system will not go into operation simultaneously.

tone-burst generator An oscillator and associated circuitry for producing a tone burst.

tone control An adjustable device or circuit for modifying the frequency response of an amplifier (i.e., for emphasizing bass, treble, or midrange pitches).

tone dialing A method of telephone dialing that uses standard tone pairs actuated via a keypad with 12 keys representing digits 0 through 9 and symbols # and *. Some keypads have four additional keys: A, B, C, and D.

tone generator An oscillator for producing simple audio-frequency signals for communications, control, or testing.

tone keying In wire and radio telegraphy, the representation of code characters by audio-frequency tones. Also see MODULATED CONTINUOUS WAVE.

tone localizer A localizer that provides lateral guidance for an aircraft by comparing the amplitudes of two modulating frequencies.

tone modulation **1.** The transmission of Morse, Baudot, or ASCII signals by audio-frequency modulation of a radio-frequency carrier. **2.** Any rapid variation of the amplitude or frequency of an audio tone.

top cap A small metal cap on top of some electron tubes, serving as a direct, low-capacitance connection to one of the internal elements—usually the control grid, but sometimes the plate.

top loading A method of feeding a vertical antenna at or near the top.

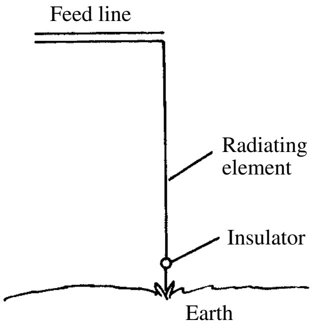

top loading

topology **1.** A branch of mathematics concerned with the properties of surfaces and spaces. **2.** The details of layout of an integrated circuit. **3.** The characteristics of a surface.

toroid See TOROIDAL COIL.

toroidal coil A coil wound on a form that is shaped like a donut. The form is made of powdered-iron or ferrite. Toroidal coils have certain advantages over solenoidal coils: greater inductance for a given physical size, better isolation properties, and higher *Q* factor. A disadvantage is that an air core is not practical.

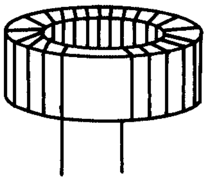

toroidal coil

torque **1.** The force that tends to produce a rotating motion. **2.** Rotation of the plane of polarization of light by some crystals.

torque amplifier A device having rotating input and output shafts and that delivers greater torque at the output shaft than that required to turn the input shaft.

torque sensitivity Symbol, *KT*. For a torque motor, torque output per ampere of input current.

torr A unit of pressure equal to the pressure required to support a column of mercury 1 millimeter high at 0°C and standard gravity; 1 torr = 133.322 Pa (1333.2 microbars).

torsion The effect on an object by torque applied to one end while the other is being held fast or torqued in the opposite direction.

torsion delay line A delay line in which the delay is manifested in a material that is torqued by mechanical vibrations.

torsion waves Waves that travel by means of torque, instead of displacement or compression. The velocity of propagation depends on the modulus of rigidity and the density of the propagating medium.

tot **1.** Abbreviation of *total.* **2.** To derive a sum or total.

total breakaway torque For a torque motor, the sum of magnetic retarding torque and brush-commutator friction.

total harmonic distortion Abbreviation, THD. The distortion caused by the combined action of all the harmonics present in a complex waveform. An important specification in high-fidelity audio amplifiers.

total distortion See TOTAL HARMONIC DISTORTION.

total internal reflection **1.** The reflection of visible light from a boundary between two substances having different indices of refraction. When the angle of incidence, relative to the tangent to the boundary, is smaller than a certain value, as light travels through the more dense medium total reflection occurs at the boundary. **2.** Reflection of electromagnetic waves from an ionized layer in the atmosphere. This occurs at angles smaller than a certain angle, relative to the tangent of the plane of the ionizing layer. In some cases, total internal reflection does not occur. Actually, most ionospheric reflection is really refraction; the electromagnetic waves are bent by the ionized layer rather than reflected.

total reflection Full return of a ray by a reflector, none of the energy being transmitted by or absorbed in the reflecting material.

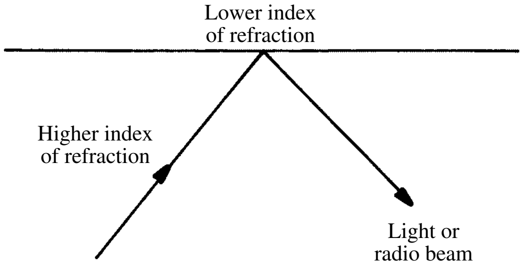

total internal reflection

touchplate relay A capacitance relay in which the pickup element is a small, metal plate that actuates the relay when it is touched.

tough dog A malfunctioning circuit that seemingly defies all attempts to diagnose and correct its trouble. Also see DOG.

tourmaline crystal A piezoelectric and pyroelectric crystal; a transparent plate of tourmaline has the ability to polarize light passing through it. Also see POLARIZATION, **3**; POLARIZED LIGHT; and POLARIZER.

tower **1.** A usually tall and self-supporting openwork structure used to support an antenna, and usually having three or four sides. **2.** A metal structure, as defined in **1**, used as a vertical antenna.

Townsend discharge In a glow-discharge tube, the discharge that begins after the applied voltage reaches a given level. It is a low-current, non-self-sustaining discharge. Compare ABNORMAL GLOW and NORMAL GLOW DISCHARGE.

touch screen A special cathode-ray tube (CRT) or video display unit that allows input of data via physical contact with the screen surface. Items are selected by simply touching the appropriate spot ("button" or icon) on the screen.

TP Abbreviation of TRANSACTION PROCESSING.

T pad See TEE PAD.

TR Abbreviation of transmit-receive.

tr **1.** Symbol for RECOVERY TIME. **2.** Symbol for RISE TIME.

trace **1.** A tiny or insignificant quantity. **2.** The movement of the electron beam across the face of a cathode-ray tube. **3.** A routine used for testing of, or for locating a fault in, a circuit or computer program. **4.** The process of implementing such a routine.

trace element See MICROELEMENT.

tracer A suitable substance or object introduced into the human body and whose progress through the body can be followed (or its state monitored) by means of electronic equipment. Tracers are sometimes used also in nonbiological systems, such as pipelines.

track **1.** A discrete information band on a magnetic disk or tape. **2.** To follow, as by a stylus, a phono disc groove. **3.** To follow, as by radar, a target.

trackability An expression of the accuracy with which a phonograph stylus follows the irregularities in a disc.

trackball A device for guiding the cursor or pointer in a computer. It is often used in laptop (notebook) computers. It resembles a ball bearing; the operator moves the cursor or pointer by pushing on the bearing with a finger.

track-drive locomotion A method of robotic locomotion using two or more wheels that drive a belt (track). It uses the same principle as a military tank.

tracking Following in step, as when ganged circuits resonate at the same frequency (or some frequency difference) at all settings, or when a missile closely follows its guiding signal.

tracking force See VERTICAL STYLUS FORCE.

track label On a magnetic storage medium, a record that identifies a track.

tracking mode In tracking supplies, the usual manner of operation in which the output of each of the separate supplies automatically follows that of the one being adjusted. Compare INDEPENDENT MODE.

tracking supplies Adjustable power supplies, packaged two or more to the unit, in which the output of each will automatically follow adjustment made to one of them.

track pitch The distance between tracks (see TRACK, **2**).

traffic **1.** Collectively, messages handled by a communications station. **2.** Collectively, data and instructions handled by a computer system in continual use.

trailer label At the end of a magnetic-tape or floppy-disk file, a record signaling the end of the file and often giving such information as the number of records in the file.

trailer record At the end of a group of computer records, a record containing information relevant to the group's processing.

trailing edge The falling edge of a pulse. Compare LEADING EDGE.

trans Abbreviation of TRANSVERSE.

transaction The exchange of activity that occurs between a computer, via a terminal, and the user, including any processing required (e.g., that involved in adding records to, or deleting them from, a file).

transaction file In data processing, a group of records used to update a master file.

transaction processing In computer operations, the use of a central processor for handling, modifying, or otherwise acting on information by transactions.

transaction tape Magnetic tape on which a transaction file has been recorded.

transadmittance For an active device, the ratio $dI2/dE1$, where $I2$ is the alternating-current (ac) component of the current in a second electrode (such as the drain), and $E1$ is the ac component of voltage on a first electrode (such as the gate), with constant direct-current operating voltages.

transceiver **1.** A combination transmitter and receiver, housed in a single enclosure, with frequency control and some ancillary stages common to both units. This design is economical because it eliminates redundancy. Such a system is more easily tuned than a separate transmitter and receiver if the operating frequency must be changed often, but it can be difficult to carry out split-frequency or split-band communication on two frequencies that differ greatly. Compare TRANSMITTER–RECEIVER. **2.** In computer practice, a read/write data terminal capable of transmitting and receiving information to and from a channel.

transcendental functions Nonalgebraic functions. These include logarithmic functions, exponential functions, trigonometric functions, and inverse trigonometric functions.

transconductance **1.** Symbol, gm. Unit, micromho. In an electron tube, the extent to which plate current (IP) changes in response to a change in grid voltage (EG); $gm = dIP/dEG$. **2.** Symbol, gfs. Unit, micromho. In a field-effect transistor, the extent to which drain current (ID) changes in response to a change in gate-to-source voltage (VGS); $gfs = dID/dVGS$.

transconductance amplifier An amplifier in which the output current is a linear function of the input voltage.

transcribe **1.** To record material, such as a radio program, for future transmission. **2.** In computer operations, the intermedia transfer of data, as from tape to disk.

transcriber A device used for intermedia data transfer (e.g., one that can move the data on magnetic tape to magnetic or optical disc).

transcription A recording of material, such as a record or tape of a radio program, for later use in a transmitter. See ELECTRICAL TRANSCRIPTION.

transcutaneous nerve stimulator Abbreviation, TNS. An electronic device for the temporary relief of pain. In its use, electrodes taped to the skin over the painful area are connected to a portable generator of suitable pulse energy.

transducer A device that converts one quantity into another quantity, specifically when one of the quantities is electrical. Thus, a loudspeaker converts electrical impulses into sound, a microphone converts sound into electrical impulses, a photocell converts light into electricity, a thermocouple converts heat into electricity, etc.

transducer amplifier An amplifier used expressly to boost the output of a transducer.

transducer efficiency For a transducer, the ratio of the output power to the input power.

transductor See MAGNETIC AMPLIFIER.

transfer **1.** To move a signal from one point to another—especially through a modifying circuit or device. **2.** To transmit information or data from one point or device to another, inside or outside a system.

transfer characteristic A figure or plot expressing the output-input signal relationship in a circuit or device. Also see TRANSFER, **1**.

transfer efficiency In a charge-coupled device, the proportion of charge that is transferred under given conditions.

transfer function **1.** See TRANSFER CHARACTERISTIC. **2.** An expression that mathematically shows how two entities or events occurring in different places or at different times are related.

transfer rate The speed at which data can be moved between a computer's internal memory and a peripheral.

transferred charge In a circuit containing a capacitor, the net electric charge that moves around the external circuit from one plate of the capacitor to the other.

transform **1.** See LAPLACE TRANSFORM. **2.** To change the voltage or nature of an electrical parameter (e.g., high to low voltage and alternating to direct current). **3.** To change the form, but not the content, of data.

transformation constant See DISINTEGRATION CONSTANT.

transformer **1.** A device using electromagnetic induction to transfer electrical energy from one circuit to another (i.e., without direct connection between them). In its simplest form, a transformer consists of separate primary and secondary coils wound on a common core of ferromagnetic material, such as iron. When an alternating current flows through the primary coil, the resulting magnetic flux in the core induces an alternating voltage across the secondary coil; the induced voltage can cause a current to flow in an external circuit. Also see AIR-CORE TRANSFORMER, INDUCTION, INDUCTIVE COUPLING, IRON-CORE TRANSFORMER, and TURNS RATIO. **2.** A section of radio-frequency (RF) transmission line used to match impedances. Also see LINEAR TRANSFORMER.

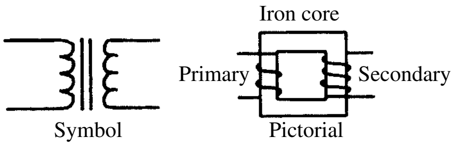

transformer, 1

transformer-coupled amplifier An amplifier using coupling transformers between its stages or at its input and output points. Compare RESISTANCE-CAPACITANCE-COUPLED AMPLIFIER.

transformer coupling The inductive coupling of circuits through a transformer. Also see COEFFICIENT OF COUPLING; INDUCTIVE COUPLING; MUTUAL INDUCTANCE; and TRANSFORMER, **1**.

transformer equivalent circuit An equivalent circuit depicting the various parameters of a transformer (such as primary and secondary resistances, primary and secondary reactances, core losses, etc.) and their relationship to each other.

transformer feedback Inductively coupled feedback (positive or negative) through a transformer.

transformer input current See PRIMARY CURRENT.

transformer input voltage See PRIMARY VOLTAGE.

transformer iron See SILICON STEEL.

transformer loss The expression (in decibels) 10 $\log_{10}(P1/P2)$, where $P1$ is the calculated power that a given transformer should deliver to a particular load impedance, and $P2$ is the actual power delivered.

transformer oil A petroleum product in which high-voltage, high-current transformers are sometimes immersed. The oil protects the windings from environmental damage.

transformer output current See SECONDARY CURRENT.

transformer output voltage See SECONDARY VOLTAGE.

transformer-type voltage regulator See VOLTAGE-REGULATING TRANSFORMER.

transformer utilization factor See UTILITY FACTOR.

transient **1.** A sudden, high-voltage spike in an alternating-current system, caused by arcing or lightning. **2.** A spurious signal in a hardcopy-receiving system. **3.** Any short pulse attributable to external causes. **4.** Existing for a short period of time, or intermittently for short periods of time.

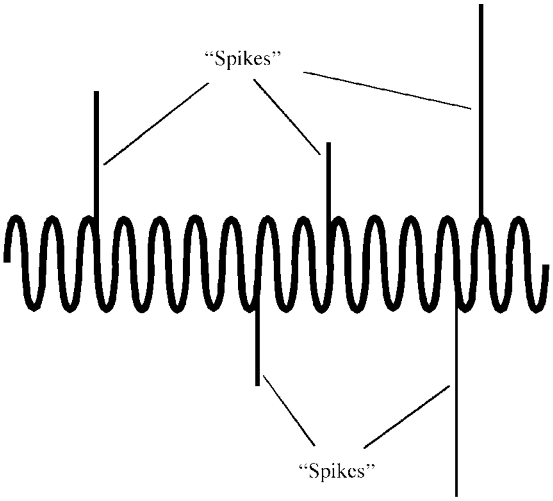

transient, 1.

transient absorber See SURGE ARRESTER.

transient arrester See SURGE ARRESTER.

transient-based amplifier See CRYSTAL AMPLIFIER, **2**.

transient current A momentary pulse of current. Also see TRANSIENT.

transient response The response of a circuit to a transient, as opposed to its steady-state response, usually evaluated in terms of its ability to reproduce a square wave.

transient suppressor See SURGE ARRESTOR.

transient voltage A momentary pulse of voltage. Also see TRANSIENT.

transient voltmeter An instrument that indicates the voltage of momentary signals. Usually, a peak-reading meter in the instrument displays the highest positive or negative value the transient attains (sometimes holding the deflection for reading later) and can respond to pulses of 1 microsecond in duration.

transistor An active (commonly three-terminal) semiconductor device capable of amplification, oscillation, and switching action. The name is a contraction of *transfer resistor.* Also see ALLOY-DIFFUSED TRANSISTOR, ALLOY TRANSISTOR, BIPOLAR TRANSISTOR, DIFFUSED-BASE TRANSISTOR, DIFFUSED EMITTER-AND-BASE TRANSISTOR, DIFFUSED-JUNCTION TRANSISTOR, DIFFUSED MESA TRANSISTOR, DIFFUSED PLANAR TRANSISTOR, DIFFUSED TRANSISTOR, DIFFUSION TRANSISTOR, DOUBLE-BASE JUNCTION TRANSISTOR, FIELD-EFFECT TRANSISTOR, FIELDISTOR, GERMANIUM TRANSISTOR, GROWN-DIFFUSED TRANSISTOR, GROWN-JUNCTION TRANSISTOR, JUNCTION TRANSISTOR, MESA TRANSISTOR, METAL-OXIDE-SEMICONDUCTOR FIELD-EFFECT TRANSISTOR, MICROALLOY DIFFUSED TRANSISTOR, MICROALLOY TRANSISTOR, PHOTOTRANSISTOR, PLANAR EPITAXIAL PASSIVATED TRANSISTOR, PLANAR TRANSISTOR, POINT-CONTACT TRANSISTOR, POWER TRANSISTOR, SILICON TRANSISTOR, SURFACE-BARRIER TRANSISTOR, SURFACE-CHARGE TRANSISTOR, TANDEM TRANSISTOR, THIN-FILM TRANSISTOR, THREE-JUNCTION TRANSISTOR, and UNIJUNCTION TRANSISTOR.

transistor amplifier An amplifier containing only transistors as the active components. Also called *transistorized amplifier.*

transistor analyzer An instrument for measuring the electrical characteristics of transistors. Compare TRANSISTOR TESTER.

transistor battery A 9-volt electrochemical battery consisting of six tiny zinc–carbon or alkaline cells in series. Each of the six cells supplies 1.5 volts. The ampere-hour capacity is small. The battery has a characteristic box shape with adjacent clip-on terminals at one end. These batteries are used in low-current electronic devices such as remote-control garage-door openers, television channel changers, remote videocassette recorder (VCR) controls, electronic calculators, and smoke detectors.

transistor-coupled logic Abbreviation, TCL. In computer and automatic-control operations, logic circuitry and systems using multi-emitter-coupled transistors. Also see RESISTOR-TRANSISTOR LOGIC and TRANSISTOR-TRANSISTOR LOGIC.

transistor current meter An ammeter, milliammeter, or microammeter circuit containing an amplifier using only transistors. Also see ELECTRONIC CURRENT METER.

transistorized voltmeter See TRANSISTOR VOLTMETER.

transistorized voltohmmeter An electronic voltmeter-ohmmeter combination using a transistorized circuit similar to that of a vacuum-tube voltohmmeter.

transistor keyer A power transistor acting as a keying device.

transistor oscillator An oscillator using only transistors as the active components. Also called *transistorized oscillator.*

transistor power supply A high-current, well-filtered, direct-current power supply for operating transistor circuits.

transistor radio Any small, portable, battery-powered radio receiver whose active components are transistors and/or other semiconductor devices.

transistor-resistor logic See RESISTOR-TRANSISTOR LOGIC.

transistor tester An instrument for checking the condition of transistors (i.e., whether good or bad). Compare TRANSISTOR ANALYZER.

transistor tetrode See DOUBLE-BASE JUNCTION TRANSISTOR.

transistor thyratron See SOLID-STATE THYRATRON.

transistor-transistor logic Abbreviation, TTL or T^2L. In computer operations, a circuit in which the multiple-diode cluster of the diode-transistor logic circuit has been replaced by a multiple-emitter transistor.

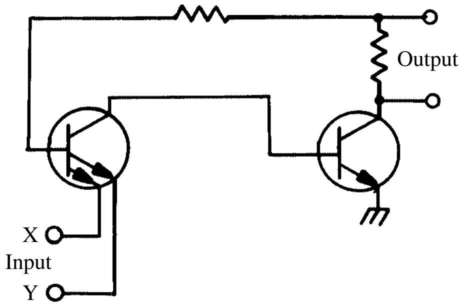

transistor-transistor logic

transistor voltmeter Abbreviation, TVM. A voltmeter containing an amplifier that uses only bipolar transistors. Also called *transistorized voltmeter.* Also see ELECTRONIC VOLTMETER. Compare FET VOLTMETER and VACUUM-TUBE VOLTMETER.

transistor VOM See TRANSISTORIZED VOLTOHMMETER.

transit angle For an electron, the angular frequency multiplied by the time required to travel from one point to another.

transit time The time taken by an electron to travel from one electrode to another—especially from the cathode to the plate in a vacuum tube.

transistor element A metallic element whose atoms have valence electrons in two shells. Examples: chromium, iron, and nickel.

transition factor See MISMATCH FACTOR.

transition region See BARRIER, **1**.

translation loss See PLAYBACK LOSS.

translator **1.** See COMPILER. **2.** See ASSEMBLER.

transliteration **1.** To change the characters in one alphabet or code system to the characters in a different system. **2.** The function that maps the characters in one alphabet or code system to those in another.

translucence **1.** The transmission of radiation, especially visible light, through a material. **2.** The extent to which a substance can transmit radiation—especially visible light. **3.** Pertaining to a material that partially or totally transmits radiation—especially visible light.

transmission gain Current amplification, power amplification, or voltage amplification.

transmission line **1.** A single conductor or group of conductors for carrying electrical energy from one point to another. **2.** A correctly dimensioned conductor or pair of conductors for carrying radio-frequency energy from a transmitter to an antenna or coupling device.

transmission mode **1.** In a transceiver, the condition in which the transmitter is enabled and the receiver is disabled. **2.** In a waveguide, propagation via transverse waves.

transmission speed The number of information elements (words, code groups, data symbols, bits, and bytes) that can be generated or received per unit time (second or minute) by a system or operator.

transmission wavemeter A (usually simple) inductance-capacitance-tuned wavemeter that provides peak response when tuned to the frequency of a signal passing through it; also, the comparable microwave device. Compare ABSORPTION WAVEMETER.

transmit-receive switch A manually or electrically operated switch for transferring a single antenna between a transmitter and receiver.

transmittancy The relative ability of a substance to transmit radiation. Transmittancy depends on the frequency of the radiation, as well as the substance.

transmitter **1.** An equipment for producing and sending signals or data. **2.** See MICROPHONE. **3.** One who originates signals or data.

transmitter–receiver A transmitter and receiver usually contained in a single enclosure, but with separate and independent oscillators, mixers, and amplifiers. The main asset of this scheme is the fact that it is possible to conduct split-frequency communication on bands whose wavelengths differ greatly. Another asset is the ability of the system, with proper design, to operate in full-duplex mode. Installations of this type are often used in satellite communications. Compare TRANSCEIVER, **1.**

transmitting antenna An antenna designed expressly for the efficient radiation of electromagnetic energy into space.

transmitting station A station that only transmits signals (i.e., it engages in no official form of reception). Compare RECEIVING STATION.

transmittivity The degree to which a selective circuit transmits a desired signal. Compare REJECTIVITY.

transmultiplexer A device that changes a signal from one multiplexed form to another while maintaining all of the information contained in the signal. For example, a transmultiplexer might convert time-division-multiplex data to frequency-division multiplex or vice-versa.

transonic Equal to, or approximating, the speed of sound in air (approximately 1100 feet per second).

transparence The practically unimpeded transmission of radiation, such as light, through a material. Compare OPACITY and TRANSLUCENCE.

transponder An acronym for transmitter and responder. **1.** A combination transmitter–receiver that automatically transmits an identification signal whenever it receives an interrogating signal. **2.** In a communications satellite, a broadband repeater that receives signals sent up from the earth, converts them to another frequency, and retransmits them back to the earth. Some satellites have units that employ multiplexing, data storage and retrieval systems, and other schemes to maximize the amount and variety of data they can handle.

transponder overload A condition that occurs when a satellite transponder is heavily used or when an excessively strong uplink signal is received. During heavy usage, all downlink signals are consistently weaker than they are when the transponder is not dealing with many signals. If an overly powerful uplink signal is received, all downlink signals are attenuated intermittently; the greater the instantaneous power of the offending uplink signal, the greater the instantaneous attenuation of all the downlink signals. See TRANSPONDER, **2.**

transport See TAPE TRANSPORT.

transportable equipment Portable electronic equipment. See, for example, PORTABLE TRANSMITTER.

transuranium An element whose atomic number is higher than that of uranium.

transverse Occurring in a direction or directions perpendicular to the direction of propagation.

transverse electric mode In a waveguide, the condition in which the electric lines of flux are perpendicular to the direction of propagation. Compare TRANSVERSE MAGNETIC MODE. Also see WAVEGUIDE MODE.

transverse electromagnetic wave An electromagnetic wave having electric-field vectors and magnetic-field vectors perpendicular to the direction of propagation.

transverse magnetic mode In a waveguide, the condition in which the magnetic lines of flux are perpendicular to the direction of propagation. Compare TRANSVERSE ELECTRIC MODE. Also see WAVEGUIDE MODE.

trap **1.** See WAVETRAP. **2.** In a semiconductor crystal, an imperfection capable of trapping current carriers.

trapezoid **1.** A polygon having four sides, of which only two are parallel. **2.** See TRAPEZOIDAL PATTERN. **3.** See TRAPEZOIDAL WAVE.

trapezoidal distortion In television or facsimile, a form of distortion in which the frame is wider at the top than at the bottom, or vice versa.

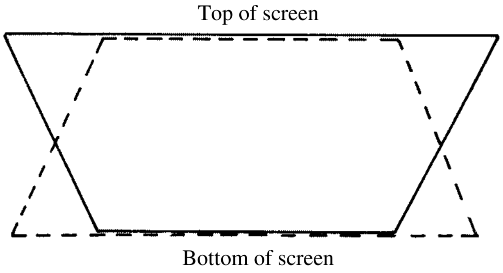

trapezoidal distortion

trapezoidal pattern An oscilloscope pattern used to check the percentage of modulation of an amplitude-modulated wave. Its name is derived from its trapezoidal shape.

trapezoidal wave A nonsinusoidal wave that is a combination of a rectangular component and a sawtooth component. It is the required waveform of the voltage applied to a magnetic deflecting coil (oscilloscope or television) to ensure a sawtooth wave of current in the coil.

traveling-wave amplifier Abbreviation, TWA. An amplifier based on the unique operation of the traveling-wave tube.

traveling-wave tube Abbreviation, TWT. A microwave tube containing an electron gun, helical transmission line, collector, and input and output couplers. A microwave signal is coupled into the helix, through which it travels while the gun projects an electron beam through the helix. When wave and electron velocities are equal, a power gain is obtained in the signal coupled out of the helix. Also see BACKWARD-WAVE OSCILLATOR.

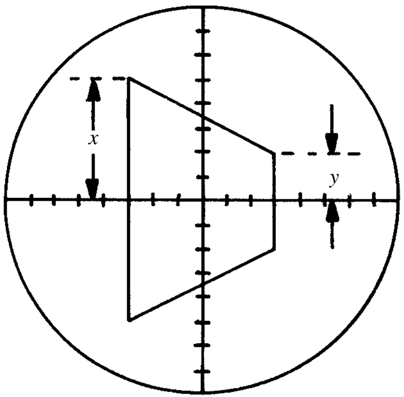

$$\text{Modulation (\%)} = 100\left(\frac{x-y}{x+y}\right)$$

trapezoidal pattern

Travis discriminator A discriminator circuit in which the diodes are operated from separately tuned halves of the secondary winding of the input transformer. Compare FOSTER-SEELEY DISCRIMINATOR.

treasure locator See METAL LOCATOR.

treble The higher portion of the sound spectrum—especially the upper end of the musical scale (middle C and above). Compare BASS.

treble boost **1.** Amplification of high audio frequencies (the treble notes) in an audio system—especially in high-fidelity music reproduction. **2.** The extent of amplification of high audio frequencies in audio applications—especially in music reproduction. **3.** A control that allows adjustment of the relative treble gain in an audio system.

treble control See TREBLE BOOST, **3**.

tree **1.** A cause-and-effect chain with two or more independent branches. **2.** A circuit with two or more branches but no meshes.

TRF Abbreviation of *tuned radio frequency.*

TRF amplifier See TUNED RADIO-FREQUENCY AMPLIFIER.

TRF receiver See TUNED RADIO-FREQUENCY RECEIVER.

triac A three-terminal, gate-controlled, semiconductor switching device. Compare DIAC.

triangular wave See BACK-TO-BACK SAWTOOTH.

triangulation A method of radiolocation in which directional bearings are obtained for the signals from two transmitters whose positions are accurately known. The position of the receiver is indicated by the intersection of two lines on a

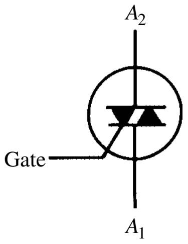

triac

map, drawn emanating from the transmitter locations 180 degrees opposite their respective bearings.

Triax A double-shielded coaxial cable. A center conductor is surrounded by two concentric, independent shield conductors.

triaxial connector A male or female connector having three concentric contacting surfaces. By contrast, the coaxial jack and coaxial plug have only two such surfaces.

triaxial speaker A dynamic loudspeaker containing three coils and three cones, for high-, low-, and middle-frequency ranges. It is, in effect, three speakers in one. It is commonly used in high-fidelity music reproduction systems. Also called *triaxial driver.* Compare THREE-WAY SPEAKER.

triboelectric Pertaining to frictional electricity.

triboelectric series See ELECTROSTATIC SERIES.

triboluminescence Luminescence produced by means of friction.

trickle charge A continuous slow charge of a storage battery, in which the charging rate is just sufficient to compensate for internal losses or normal discharge.

trickle charger A light-duty unit for charging a battery gradually at low current.

tricorner reflector A device that reflects a ray of incident energy at a 180-degree angle, regardless of the direction from which the ray approaches. Thus, the ray will return to its source, unless some obstruction intervenes. It consists of three mutually perpendicular reflecting plane surfaces that intersect in a common point. The geometry is identical to that in a typical room, where the ceiling meets two walls. It is useful at microwave, infrared, visible, and ultraviolet wavelengths.

triethanolamine An amino alcohol that precipitates metallic silver from a silver-nitrate solution in the deposition of a silver surface on a substrate, such as glass or a ceramic.

trig Abbreviation of TRIGONOMETRY.

trigatron A form of electrically operated switch. The circuit is closed by the breakdown of an electrical gap.

trigger **1.** A pulse used to start or stop the operation of a circuit or device, such as a flip-flop. **2.** To place a circuit or device into or out of operation with a start or stop pulse, as defined in **1**.

trigger diode See DIAC and FOUR-LAYER DIODE.

triggered multivibrator See DRIVEN MULTIVIBRATOR.

triggered sweep In an oscilloscope or similar device, a driven sweep.

triggering point The voltage level at which the action of an electronic switching device is initiated. Also see BREAKOVER POINT.

trigistor A three-junction semiconductor device that exhibits two-state operation, and is useful as a flip-flop or switch. Also see NPNP DEVICE.

trigonometric functions See CIRCULAR FUNCTIONS and HYPERBOLIC FUNCTIONS.

trigonometry The branch of mathematics devoted to the application of CIRCULAR FUNCTIONS in a plane. It is useful in electronics for determining impedance and phase.

trim To make a fine adjustment, as of a tuning control, balance control, output adjustment, etc.

trimmer A low-valued variable capacitor, inductor, or resistor operated in conjunction with a main unit (usually of the same sort) for vernier adjustment or range setting. See, for example, HIGH-FREQUENCY TRIMMER and OSCILLATOR TRIMMER.

trimmer capacitor A variable capacitor used as a trimmer.

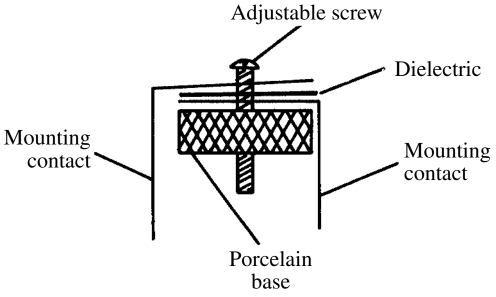

trimmer capacitor

trimmer coil See TRIMMER INDUCTOR.

trimmer inductor A variable inductor used as a trimmer.

trimmer resistor A variable resistor used as a trimmer.

trinary number system The base-3 system of notation. This system uses the digits 0, 1, and 2, the positional values being successive powers of 3 (e.g., decimal 14 equals trinary 112). Also called *ternary number system.*

Trinitron **1.** A single-gun color-television picture tube developed by Sony. **2.** A television set using this tube. The gun has three cathodes that modulate the three color beams (red, green, and blue). The beams are accelerated by common grids and are focused at different angles by convergence plates.

Trinoscope See TRINITRON.

triode A three-electrode tube or transistor, embodying an anode, cathode, and a control electrode as the principal elements.

triple-diffused transistor A diffused transistor in which the base and emitter are diffused into the top face of the chip, and the collector into the bottom face.

triple diode An assembly of three (often closely matched) semiconductor diodes in a single housing.

tripler **1.** A rectifier that delivers a direct-current output voltage at approximately triple the peak value of the alternating-current input voltage. **2.** An amplifier or other circuit that delivers an output signal at triple the frequency of the input signal.

triplexer In radar, a device that facilitates the use of two receivers at the same time.

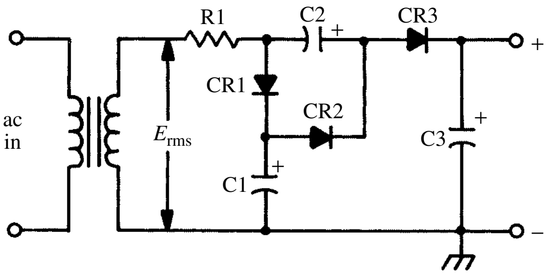

tripler, 1

Tri-state logic Digital logic in which there are three possible states, rather than the usual two. The conditions are defined as 0, 1, and undecided. Trade name of National Semiconductor.

tritium Symbol, T or H_3. An isotope of hydrogen whose nucleus contains two neutrons and one proton. Compare DEUTERIUM.

TRL Abbreviation of *transistor-resistor logic* (see RESISTOR-TRANSISTOR LOGIC).

Trojan horse A program written with the malicious intent of sabotaging the operating system and/or files in a computer. It is somewhat similar to a VIRUS; it is occasionally spread via software whose origin is questionable or unknown.

troposcatter See TROPOSPHERIC-SCATTER PROPAGATION.

troposphere The portion of the atmosphere in which virtually all weather phenomena occur. It extends from the surface to an altitude of 8 to 12 miles above sea level.

tropospheric bending A common type of radiowave propagation that takes place when radio waves are refracted in the lower atmosphere. This effect is most dramatic near weather fronts, where warm, relatively light air lies above cool, heavy air. The cooler air has a higher index of refraction than the warm air, causing radiofrequency electromagnetic fields to be bent downward at a considerable distance from the transmitter. It is often responsible for anomalies in reception of FM and TV broadcast signals. Compare TROPOSPHERIC DUCTING, TROPOSPHERIC-SCATTER PROPAGATION.

tropospheric ducting Also called *duct effect.* A form of tropospheric propagation that occurs close to the earth's surface at very-high frequencies (VHF) and ultra-high frequencies (UHF). A duct forms when a layer of cool air becomes sandwiched between two layers of warmer air. Total internal reflection of the electromagnetic (EM) field takes place inside the duct. For this phenomenon to provide communications, both the transmitting and receiving antennas must be located within the same duct, and this duct must be unbroken and unobstructed between the two locations. A duct might measure only a few feet from top to bottom, but cover thousands of square miles parallel to the surface. Compare TROPOSPHERIC BENDING, TROPOSPHERIC-SCATTER PROPAGATION.

tropospheric propagation Also called *tropo.* Over-the-horizon propagation of radio-frequency electromagnetic waves that occurs as a result of reflection, scattering, and/or refraction in the troposphere, the lowest 10 or 12 miles of the earth's atmosphere. At wavelengths shorter than about 15 m (frequencies above 20 MHz), refraction and reflection can take place within and between air masses of different density. The air also produces some scattering of EM energy at wavelengths shorter than about 3 m (frequencies above 100 MHz). "Tropo" can often allow

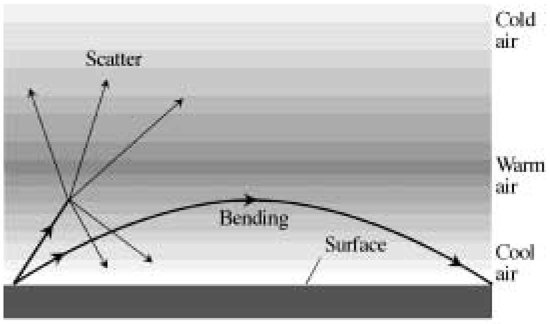

tropospheric propagation

communications over distances of several hundred miles. Also see TROPOSPHERIC BENDING, TROPOSPHERIC DUCTING, and TROPOSPHERIC-SCATTER PROPAGATION.

tropospheric-scatter propagation Also called *troposcatter*. A form of tropospheric propagation that occurs at frequencies above about 100 MHz, where the atmosphere has a scattering effect on radio waves. Dust in the air increases the scattering effect, but some troposcatter occurs regardless of the weather. Troposcatter takes place mostly at low altitudes where the air is the most dense. This mode can provide reliable communication over distances of several hundred miles when the appropriate equipment is used. Communication via troposcatter requires the use of high-gain antennas. Scatter is often observed along with other modes of tropospheric propagation. Compare TROPOSPHERIC BENDING, TROPOSPHERIC DUCTING.

troubleshoot **1.** To look for the cause of equipment failure. **2.** To look for flaws in computer software or to debug a program.

troubleshooting test A test that is part of the procedure for finding the cause of faulty electronic equipment operation. Also see DIAGNOSTIC TEST. Compare PERFORMANCE TEST.

trough value The minimum amplitude of a composite current or voltage.

true complement See RADIX COMPLEMENT.

true ground The earth, as opposed to an artificial ground, such as that provided by the radials of a ground-plane antenna or an equipment chassis.

true power In an alternating-current (ac) circuit, the power actually dissipated in the resistive component; the reactive component consumes no power. In ac circuits containing reactance, the true power is less than the product of the voltage and current. Also see AC POWER and POWER FACTOR.

truncation A method of approximating a quantity by cutting off its digits beyond a certain point. For example, 3.44 and 3.46, when truncated to two significant digits, both become 3.4. Compare ROUNDING.

trunk **1.** A communications link between two points, one usually being external. **2.** The interface between a central processing unit and a peripheral device.

trunk link In computer systems, an interface permitting access to main storage via a peripheral.

truth table In logic analysis and logic circuit design, a table in which are listed all combinations of input values and the corresponding output values for a given logic function.

ts Abbreviation of tensile strength.

Tschebyscheff filter See CHEBYSHEV FILTER.

TSS Abbreviation of TIME-SHARING SYSTEM.

T switch See TEE SWITCH.

TST switch See TIME-SPACE-TIME SWITCH.

truth table

$(x + y)' + xz$

x	y	z	$x + y$	$(x + y)'$	xz	$(x + y)' + xz$
0	0	0	0	1	0	1
0	0	1	0	1	0	1
0	1	0	1	0	0	0
0	1	1	1	0	0	0
1	0	0	1	0	0	0
1	0	1	1	0	1	1
1	1	0	1	0	0	0
1	1	1	1	0	1	1

TTL Abbreviation of TRANSISTOR-TRANSISTOR LOGIC. (Also, T^2L.)

TTT In radiotelegraphy, a signal indicating that a message concerning safety is to follow (equivalent to *securite* in radiotelephony).

TTY Abbreviation of TELETYPE.

T-type antenna See TEE ANTENNA.

T-type attenuator See TEE PAD.

tube **1.** Generic term for any electron tube (e.g., vacuum tube, gas tube, cathode-ray tube, X-ray tube, etc.). **2.** Glow lamp: argon bulb, neon bulb, mercury-vapor lamp, etc.

tube capacitances The internal capacitances between the elements of an electron tube.

tube diode A two-element (cathode and anode) electron tube for current rectification. Also see DIODE.

tube of flux A group of flux lines within a circular cross section. It can vary in diameter as the density of the flux lines changes.

tube parameters Operating coefficients of electron tubes (e.g., plate current, grid voltage, screen current, transconductance, amplification factor, etc.).

tube tester An instrument for checking one or more of the parameters of an electron tube.

tubular capacitor A fixed capacitor consisting of a wound section enclosed in a cylindrical can.

tune **1.** To adjust a selective circuit to accept or reject a signal. **2.** To correct the natural frequency of vibration of a body. **3.** To adjust a radio transmitter for optimum output. **4.** To adjust the frequency of an oscillator—especially in a radio transmitter or receiver. **5.** To adjust the resonant frequency of an antenna, antenna coupler, or antenna system.

tuned AF amplifier **1.** An audio amplifier that is continuously tunable over a band of frequencies. **2.** An audio oscillator that is set to a fixed, precise frequency. See, for example, PARALLEL-TEE AMPLIFIER.

tuned-base oscillator A self-excited, common-emitter connected, bipolar-transistor oscillator in which the tuned tank is in the base circuit.

tuned circuit A (usually series-resonant or parallel-resonant) circuit adjusted to accept or reject a signal. Also see RESONANCE.

tuned-collector oscillator A self-excited, common-emitter connected, bipolar-transistor oscillator in which the tuned tank is in the collector circuit.

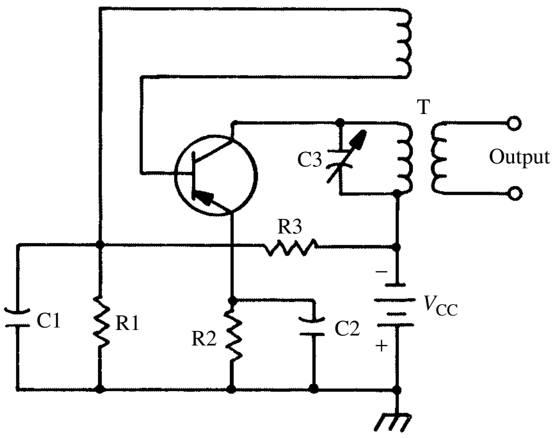

tuned-collector oscillator

tuned coupler An antenna coupler (transmatch) that can be adjusted independently of the transmitter or receiver with which it is used.

tuned dipole A half-wave, center-fed, resonant antenna.

tuned feeders An antenna feed line that is adjusted or trimmed so that the entire system (feeders and radiating element) is resonant at the transmitted-signal frequency.

tuned headphones Headphones used in radio-telegraphy that are fix-tuned to a single audio frequency (e.g., 1 kHz) by means of a small parallel capacitor.

tuned line An antenna wire or transmission line that provides a resistive load at a specific resonant frequency.

tuned pickup A pickup circuit or device (such as a radio-frequency sampling coil) that is tuned to the signal frequency.

tuned radio-frequency amplifier An amplifier circuit that is continuously tunable over a specific band of radio frequencies.

tuned radio-frequency receiver A radio receiver consisting only of a tuned-radio-frequency amplifier, detector, and audio amplifier. Compare SUPERHETERODYNE RECEIVER.

tuned reed A vibrating reed whose length, width, and/or thickness have been adjusted so that it vibrates naturally at a desired frequency.

tuned-reed frequency meter An audio frequency meter using tuned metal reeds as the indicators. Also see POWER FREQUENCY METER, **2**.

tuned relay An electronic or electromechanical relay that closes at one frequency. See, for example, REED SWITCH, **1**.

tuned signal tracer A signal tracer that can be tuned sharply to the frequency of the test signal being traced.

tuned transformer An audio-frequency or radio-frequency transformer tuned via a capacitor in parallel with its primary winding and/or a capacitor in parallel with it secondary winding.

tuner A circuit or device that can be set to select one signal from a number of signals in a frequency band.

tungsten Symbol, W. A metallic element. Atomic number, 74. Atomic weight, 183.85. Also called *wolfram.* It is used in switch and relay contacts, in the elements of some electron tubes, and in incandescent-lamp filaments.

tuning **1.** Adjustment of the frequency of a receiver to intercept a signal on a given frequency. **2.** Adjustment of a transmitter oscillator to a desired frequency. **3.** Adjustment of an inductance-capacitance circuit for resonance on a desired frequency. **4.** Adjustment of an antenna or antenna system to a desired frequency. **5.** Adjustment of a radio-frequency amplifier for optimum performance. **6.** Alignment of a musical instrument for correct tone frequency.

tuning capacitor A variable capacitor used to tune an inductance-capacitance circuit (series-resonant or parallel-resonant).

tuning coil A variable inductor used to tune an inductance-capacitance circuit (series-resonant or parallel-resonant).

tuning core See TUNING SLUG.

tuning diode See VOLTAGE-VARIABLE CAPACITOR, **1**.

tuning fork A metal device that vibrates at a precise audio frequency when struck physically. It usually has two prongs and looks something like a fork.

tuning-fork oscillator See FORK OSCILLATOR.

tuning indicator **1.** A meter or display that indicates when a receiver is properly tuned to the frequency of an incoming signal. **2.** Sometimes, a bridge null detector.

tuning meter A meter-type resonance indicator.

tuning potentiometer A single or ganged potentiometer (depending on the circuit in which it is used) used to vary the frequency response of a resistance-capacitance-tuned circuit, such as the bridge-tee, parallel-tee or Wien-bridge circuit.

tuning slug A powdered-iron core that slides or screws in and out of a coil to vary the inductance.

tuning voltage An adjustable direct-current voltage used to vary the capacitance of a varactor diode (serving as the capacitor in a tuned circuit).

tuning wand See NEUTRALIZING TOOL.

tunnel diode A specially processed junction diode whose forward conduction characteristic displays negative resistance.

tunnel-diode amplifier A circuit that uses the negative-resistance properties of a tunnel diode to produce gain. It is used primarily at ultra-high and microwave frequencies.

tunnel-diode oscillator A circuit that uses the negative-resistance properties of a tunnel diode to produce oscillation at a specific frequency. It is used primarily at ultra-high and microwave frequencies.

turbidimeter A device used to measure the turbidity (relative opaqueness) of liquids.

turbidity meter See TURBIDIMETER.

Turing machine A hypothetical model, conceived by Alan Turing in the 1930s, devised as an exercise in problem solving. In effect, it was a computer programming language with a limited instruction set, operating in a computer with infinite memory and storage capacity.

Turing test A scheme devised by Alan Turing, designed to test computers for artificial intelligence (AI). If the machine can fool a person into thinking it is another person most of the time, then the machine in effect "passes" the test, indicating that it has some AI.

turn One complete loop of a coil.

turnaround time **1.** The length of time required for a repair to be performed. **2.** The length of time required for a computer program to be completed. **3.** The time required for a trans-ceiver to switch from the full-transmit to the full-receive condition.

turn factor For a given inductor core, the number of turns (in a standard configuration) that results in an inductance of 1 H.

turn-off time The time required for the operation of a circuit or device to cease completely after a signal or operating power has been removed. Compare TURN-ON TIME.

turn-on time The time required for a circuit or device to come up to full (normal) operation after application of a trigger or operating power. Compare TURN-OFF TIME.

turnover pickup See DUAL PICKUP.

turns ratio For a transformer, the ratio of the number of turns in the primary winding to the number of turns in the secondary winding.

turnstile antenna A polyphase antenna that resembles a turnstile gate.

turntable **1.** The motor-driven, rotating platter on which a phonograph disc rests during recording or reproduction. **2.** In a high-fidelity phonograph system, the assembly containing the platter, as defined in **1**.

turret tuner A television front-end tuner having a separate tuned-circuit section for each channel. Turning a knob rotates a desired section into position against switch contacts. Also called DETENT TUNER.

TV **1.** Abbreviation of TELEVISION. **2.** Abbreviation of *terminal velocity.*

T/V Abbreviation of *temperature to voltage.*

TV decoder A device for unscrambling a television broadcast that has been intentionally encrypted to prevent its use by unauthorized viewers. See SUBSCRIPTION TV.

TV ghost See GHOST.

TVI Abbreviation of TELEVISION INTERFERENCE.

TVL Amateur radio abbreviation of *television listener* (looker).

TVM Abbreviation of TRANSISTORIZED VOLTMETER.

TVO Abbreviation of TRANSISTORIZED VOLT-OHMMETER. (Also, TVOM.)

TVOM Abbreviation of TRANSISTORIZED VOLT-OHMMETER. (Also, TVO.)

TV projector A combination electronic and optical device for projecting television images on a large screen.

TV sound marker See SOUND MARKER.

TVT Abbreviation of *television terminal* (computer peripheral).

TW Abbreviation of TERAWATT.

TWA Abbreviation of TRAVELING-WAVE AMPLIFIER.

tweeter A loudspeaker that favors the extreme treble (high) audio frequencies, ranging from 4000 or 5000 Hz to more than 20,000 Hz. Compare WOOFER.

twin diode See DUAL DIODE.

twinlead See FLAT-RIBBON LINE.

twin line See FLAT-RIBBON LINE.

twin meter See DUAL METER.

twin-T measuring circuit See PARALLEL-TEE MEASURING CIRCUIT.

twin-T network See PARALLEL-TEE NETWORK.

twisted pair A simple wire line used for communications. It consists of two wires twisted together to form a helix.

twister A piezoelectric crystal, such as one of Rochelle salt, that generates a voltage when torqued.

twistor A magnetic-memory element consisting of a winding of magnetic wire on a length of nonmagnetic wire.

two-channel amplifier See DUAL-CHANNEL AMPLIFIER.

two-channel recorder See DOUBLE-CHANNEL RECORDER.

two-electrode amplifier See DIODE AMPLIFIER.

two-electrode tube See TUBE DIODE.

two-element amplifier See DIODE AMPLIFIER.

two-element beam A Yagi antenna consisting of a driven element and one parasitic element. The parasitic element can be either a reflector, about 5% longer than the resonant half wavelength at the operating frequency, or it can be a director, about 5% shorter than a half wavelength at the operating frequency.

two-element tube An electron tube having two principal electrodes (i.e., a tube diode).

two-five code See BIQUINARY CODE.

two-phase Pertaining to circuits or devices in which two components (two voltages, two

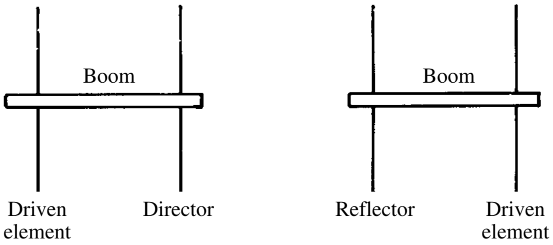

two-element beam

currents, or a current and a voltage) are 90° out of phase with each other.

two-phase system See QUARTER-PHASE SYSTEM.

two-phase, three-wire circuit A circuit using three conductors, one of which (the common return) is 90° out of phase with the other (outer) two.

two-pincher gripper A simple robotic end effector using two tongs, and resembling large tweezers or needle-nosed pliers. It grasps objects by "pinching" them.

two-point tuning See DOUBLE-SPOT TUNING.

two-space A two-dimensional mathematical continuum in which each point is uniquely defined by two variables in an ordered pair, such as (x,y), and each ordered pair corresponds to exactly one point in the continuum. A common scheme is the Cartesian coordinate system; another scheme is the polar coordinate system. See CARTESIAN COORDINATES and POLAR COORDINATES.

two-space coordinates A coordinate system for defining points in two-space. See CARTESIAN COORDINATES and POLAR COORDINATES.

two-state device A device having two stable states (e.g., a flip-flop).

two-state logic Digital logic in which there are two possible conditions (called high and low, true and false, or 1 and 0).

two-tone keying See FREQUENCY-SHIFT KEYING.

two-track recording Recording on two adjacent tracks on magnetic tape. The separate recordings can be in the same direction (as in stereo) or in opposite directions.

two-way amplifier An amplifier whose input and output terminals can be used interchangeably.

two-way communication The exchange of messages between two or more stations that transmit as well as receive. Compare ONE-WAY COMMUNICATION, **1**.

two-way radio **1.** Any form of TWO-WAY COMMUNICATION using electromagnetic waves. **2.** A radio transceiver—especially one used for voice communication at very-high or ultra-high frequencies.

two-way repeater In telephony, a device that amplifies and retransmits a signal in either direction. Also see TWO-WAY AMPLIFIER.

two-way speaker A woofer/midrange speaker and a tweeter occupying the same enclosure, and interconnected by a crossover network for wideband frequency response.

TWT Abbreviation of TRAVELING-WAVE TUBE.

TWX Abbreviation of TELETYPEWRITER EXCHANGE.

Twystron A form of microwave-oscillator tube, similar to the Klystron.

Type A telephone line A telephone line serving one subscriber (as opposed to a party line).

u

U **1.** Symbol for URANIUM. **2.** Abbreviation of UNIT.

u Symbol (ital.) for UNIFIED ATOMIC MASS UNIT.

Ua Abbreviation of *unit of activity* (electroencephalography).

UAM Abbreviation of *underwater-to-air missile.*

ubiquitous carrier A current carrier (especially an electron) whose velocity is so high, or mechanism of transfer so subtle, that it appears to be in two places simultaneously. Thus, in the tunnel diode, during the interval of tunneling a single electron, appears to be on both sides of the barrier at the same time.

Ubitron A tube in which a periodic magnetic field causes an electron beam to undulate. Through its transverse velocity component, the beam interacts with a radio-frequency wave, and its kinetic energy is converted into radio-frequency energy. The ubitron can be used as an amplifier or oscillator.

UCT Abbreviation of UNIVERSAL COORDINATED TIME.

UDC Abbreviation of *universal decimal classification.*

UDOP Abbreviation of *ultra-high-frequency Doppler.* A 440-MHz phase-coherent tracking system used to determine the velocity and position of missiles and space vehicles.

UEP Abbreviation of *underwater electric potential.*

UFET Abbreviation of unipolar field-effect transistor. (Also, UNIFET.) See UNIPOLAR TRANSISTOR.

UFO See UNIDENTIFIED FLYING OBJECT.

UG Abbreviation of underground.

UHF Abbreviation of ULTRAHIGH FREQUENCY.

UHF adaptor See UHF CONVERTER, **1**.

UHF capacitor A button-type capacitor. Because of its unique design, it is an efficient bypass capacitor at ultrahigh frequencies.

UHF converter **1.** A circuit, usually consisting of a radio-frequency amplifier and mixer, for converting ultra-high-frequency (UHF) signals to a lower band of frequencies. **2.** A circuit for converting UHF television signals to very-high-frequency (VHF) signals so that they can be accommodated by an older (VHF only) television receiver. See UHF TELEVISION CHANNELS and VHF TELEVISION CHANNELS.

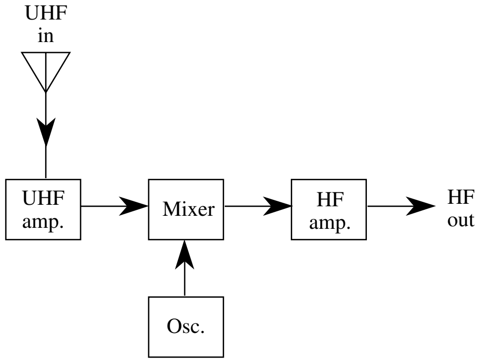

UHF converter, 1

UHF diode A semiconductor diode whose rectification efficiency is good at ultra-high frequencies.

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

UHF generator **1.** An oscillating device (such as a transistor, tunnel diode, Klystron, or magnetron) used to produce radio-frequency energy at ultra-high frequencies (UHF). **2.** The equipment in which such a device is used. **3.** An ultra-high-frequency test-signal generator.

UHF loop A (usually single-turn) loop antenna with a toroidal radiation pattern perpendicular to the loop; the antenna has a natural wavelength in the ultra-high-frequency (UHF) range.

UHF receiver A receiver tunable to frequencies in any band in the range 300 MHz to 3000 MHz.

UHF television channels Television channels 14 through 69, whose frequencies lie in the ultra-high range.

UHF transistor A transistor specially designed and fabricated for ultra-high-frequency (UHF) operation. It is characterized by extended beta cutoff frequency, low junction capacitance, and fast recovery time.

UHF translator A television broadcast translator station transmitting in an ultra-high-frequency (UHF) channel.

UHF transmitter A transmitter that is specially designed for operation at ultra-high frequencies. In such a transmitter, stray parameters are minimized and special tubes or transistors are required.

UHF tube A vacuum tube specially designed for ultra-high-frequency operation. It is characterized by low interelectrode capacitance, low input and output capacitances, short electron transit time, and low lead inductance.

UHF tuner **1.** A superheterodyne receiver front end tunable in the ultra-high-frequency (UHF) range. It is usually arranged to deliver the mixer output to a lower-frequency receiver. **2.** A television receiver front end tunable to the UHF channels.

UHR Abbreviation of ULTRA-HIGH RESISTANCE.

U index The difference between consecutive daily mean values of the horizontal component of the geomagnetic field.

UJT Abbreviation of UNIJUNCTION TRANSISTOR.

UL **1.** Abbreviation of UNDERWRITERS LABORATORIES. **2.** Abbreviation of ULTRALINEAR.

ULD Abbreviation of ULTRA-LOW DISTORTION.

ULF Abbreviation of ULTRA-LOW FREQUENCY.

U-line section A short section of coaxial cable or hardline having the shape of a squared U.

ULSI See ULTRA-LARGE-SCALE INTEGRATION.

ultimate attenuation The maximum attenuation in a circuit or device when it is operated outside its passband, or within its stopband.

ultimately controlled variable In an automated system, the quantity or operation (e.g., liquid level, temperature, pressure, precision drilling, etc.), whose control is the end purpose of the system.

ultimate-range ballistic missile A ballistic missile whose range is greater than half the earth's circumference (approximately 12,000 miles).

ultimate ratio The limiting value of a ratio (i.e., the value approached by a ratio).

ultimate sensitivity **1.** In an instrument or system, the maximum degree of perception of, or response to, a quantity or condition. **2.** In a graphic recorder, half the deadband.

ultimate threshold See ULTIMATE SENSITIVITY, **2**.

ultimate trip current For a specified set of environmental conditions, the lowest value of current that will trip a circuit breaker.

ultimate trip limits For a circuit breaker, the maximum and minimum current at which the breaker trips and drops out.

ultor anode The second anode of a television picture tube or oscilloscope tube.

ultor element In a television picture tube, the element that receives the highest direct-current voltage. Also called *ultor electrode.* Also see ULTOR ANODE.

ultor voltage The high direct-current voltage applied to the second anode of an oscilloscope tube or television picture tube. Also see ULTOR ANODE.

ultra- A prefix meaning *above*, *larger than*, *greater than*, or *higher than.* See the following several definitions for examples.

ultra-atomic Pertaining to particles smaller than atoms (i.e., SUBATOMIC).

ultra-audible frequency A frequency higher than the highest audible frequency. Also called ULTRASONIC FREQUENCY and *supersonic frequency.*

ultra-audion oscillator See ULTRAUDION OSCILLATOR.

ultracentrifuge A centrifuge that spins at a very high speed.

ultrafast switch An electronic switch capable of microsecond operation.

Ultrafax A system for high-speed transmission of printed matter by facsimile, radio, or television.

ultra-high frequency A radio frequency in the range 300 MHz to 3000 MHz, corresponding to free-space wavelengths between 1 meter and 10 centimeters.

ultra-high-frequency band See UHF BAND.

ultra-high resistance Resistance of 100 megohms or higher.

ultra-large-scale integration An integrated-circuit process that goes several orders of magnitude beyond VERY-LARGE-SCALE INTEGRATION. More than 4×10^6 transistors could be on a single chip.

ultra-linear Pertaining to the most linear operation obtainable with state-of-the-art electronic equipment.

ultra-linear amplifier A high-fidelity audio-frequency power amplifier having wide frequency response and very low distortion.

ultra-low distortion Total harmonic distortion (THD) of less than 0.1 percent.

ultra-low frequencies **1.** Radio frequencies between 300 Hz and 3 kHz, pertaining to wavelengths between 1000 kilometers and 100 kilometers. **2.** Audio frequencies lower than those in the human hearing range.

ultramicrometer An electronic instrument for extremely small linear measurements.

ultramicroscope An optical microscope using refracted light to illuminate minute particles.

ultramicrowave Pertaining to wavelengths in the range 10^{-3} to 10^{-6} meter (300 GHz to 300 THz).

ultraminiature See SUBMINIATURE.

ultraphotic rays Collectively, infrared and ultraviolet rays.

ultrared rays See INFRARED RAYS.

ultra-short waves Waves whose length correspond to ultra-high frequencies (i.e., wavelengths between 0.1 and 1 meter).

ultrasonic Of higher frequency than those that are audible (supersonic). See, for example, ULTRASONIC FREQUENCY.

ultrasonic bonding A method of bonding metal by means of physical vibration at frequencies above the human hearing range.

ultrasonic brazing Forming a nonporous bond between metal parts through the use of ultrasonic energy and a second, different metal (or alloy) having a lower melting point.

ultrasonic cleaning A method of cleaning delicate or intricately structured items (such as dentures, contact lenses, or jewelry) in which the soiled items are immersed in a fluid that is agitated by ultrasonic transducers; the foreign particles are shaken away.

ultrasonic cleaning tank A thick-walled stainless-steel tank with ultrasonic transducers mounted within its walls and used for ultrasonic cleaning.

ultrasonic coagulation The coagulation of a substance through ultrasonic agitation.

ultrasonic communication Underwater telegraphic communication between ships and/or submarines by keying the echo-ranging sonar equipment.

ultrasonic delay line A delay line, such as the mercury type, through which an ultrasonic signal is propagated. The delay results from the relatively slow propagation of the ultrasonic wave through the substance in the line.

ultrasonic densitometer A density-measuring apparatus whose operation is based on the time required for an ultrasonic signal to penetrate the material under test, or for the signal to penetrate the material and be reflected back to the transmitter.

ultrasonic depth finder See ACOUSTIC DEPTH FINDER.

ultrasonic detector A device that responds to ultrasonic waves by indicating their presence, intensity, and/or frequency.

ultrasonic diagnosis The medical determination of the condition of tissues or structures within the body, in terms of reflection of ultrasonic waves by those tissues or structures. The technique is useful in situations where other techniques, such as radiology or the use of catheter probes, are risky. Also see ULTRASOUND.

ultrasonic disintegrator In biology and related fields, a device that uses ultrasonic energy to rupture or shatter cells, tissues, or foreign bodies, such as kidney stones.

ultrasonic drill A drill running at speeds corresponding to ultrasonic frequencies. The tool is valuable in drilling hard or brittle materials and in dentistry.

ultrasonic filter **1.** A bandpass filter whose operation is based on the natural (ultrasonic) vibration frequency of small disks or rods of magnetostrictive metal. See MAGNETOSTRICTION. Also called *mechanical filter.* **2.** Generally, a filter operating at ultrasonic frequencies.

ultrasonic flaw detector A system analogous to radar, in which an ultrasonic wave is transmitted through a solid material and is reflected back to a detector and display device to reveal flaws, cracks, and strain in the material.

ultrasonic frequency For an acoustic disturbance, any frequency above the limit of human hearing (higher than 20 kHz).

ultrasonic generator **1.** An oscillator that operates at frequencies above the range of human hearing, and the output of which is intended for coupling to an electroacoustic transducer. **2.** Any device that produces ultrasonic waves.

ultrasonic grating constant For a sound wave producing a diffraction spectrum, a figure indicating the distance between diffracting centers of the wave.

ultrasonic heating The production of heat in a specimen by means of ultrasonic energy directed into or through it.

ultrasonic image converter A device that makes visible acoustic field patterns.

ultrasonic inspection The use of ultrasonic waves to detect internal flaws in solid materials, such as metals. It is valued particularly because it is nondestructive. Also see ULTRASONIC FLAW DETECTOR.

ultrasonic intrusion alarm A security system activated when an intruder disturbs a pattern of ultrasonic waves in the protected area; a sensitive relay closes, setting off an alarm.

ultrasonic level detector A level-monitoring system in which an ultrasonic transmitter and detector are mounted together on one wall of a tank or chamber. When the tank is empty, the transmitted signal is reflected by the opposite wall back to the detector. When the liquid rises, the reflection time is reduced; this change is used to operate a device that indicates that the tank has been filled to a desired level.

ultrasonic light diffraction Diffraction resulting from the periodic variation of light refraction when a beam of light passes through a longitudinal sound-wave field.

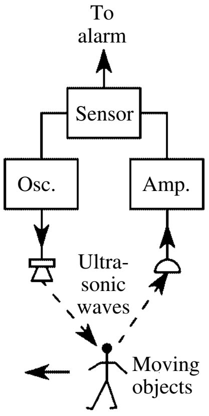

ultrasonic intrusion alarm

ultrasonic light modulator A device that modulates a light beam passing transversely through a fluid agitated by sound waves.

ultrasonic material dispersion The use of the agitative action of high-intensity ultrasonic waves to disperse or emulsify one substance in another.

ultrasonic plating The use of ultrasonic energy to deposit or bond one material onto the surface of another.

ultrasonic probe A rodlike director of ultrasonic energy.

ultrasonic relay An electronic relay actuated by ultrasound.

ultrasonics **1.** The use of high-frequency acoustic energy in industry, medicine, research, and the home. See ULTRASONIC FREQUENCY. **2.** The use of high-frequency acoustic energy for medical diagnosis and treatment. **3.** The branch of physics dealing with the effects and behavior of acoustic disturbances at frequencies above the range of human hearing.

ultrasonic soldering See ULTRASONIC BRAZING.

ultrasonic sounding Determining the depth of a body of water in terms of the time taken for a transmitted ultrasonic signal to be reflected back to a transmitting point on the surface of the water.

ultrasonic space grating A periodic variation in the index of refraction when acoustic waves are present in a light-transmitting medium. Also see ULTRASONIC LIGHT DIFFRACTION.

ultrasonic storage cell See ULTRASONIC DELAY LINE.

ultrasonic stroboscope A stroboscope that uses an ultrasonically modulated beam of light.

ultrasonic switch An electronic switch actuated by ultrasound.

ultrasonic therapy The use of ultrasonic energy in medicine for treatment of certain disorders.

ultrasonic thickness gauge An instrument that determines the thickness of a specimen in terms of the propagation time (or echo time) of ultrasonic waves through the specimen.

ultrasonic transducer A transducer that converts electrical energy into ultrasonic energy, or vice-versa. Common types are the quartz crystal, ceramic crystal, and magnetostrictive disk.

ultrasonic waves Acoustic waves whose length corresponds to ultrasonic frequencies (i.e., frequencies above 20 kHz).

ultrasonic welding A below-melting-point technique of joining two metallic bodies by clamping them tightly together and applying ultrasonic energy, rather than heat, in the plane of the desired weld. Compare ULTRASONIC BRAZING.

ultrasonic whistle A whistle whose pitch is beyond the range of human hearing. Although some are of the simile (blown) type, others are (usually miniature) electronic sound generators. These devices can be used for the remote control of television receivers, garage-door openers, and other equipment.

ultrasonography A method of examining human tissues or organs by transmitting ultrasonic waves into the body and receiving the echoes.

ultrasound **1.** Acoustic disturbances at frequencies above 20 kHz. **2.** See ULTRASONIC FREQUENCY. **3.** See ULTRASONOGRAPHY.

ultraviolet **1.** Electromagnetic radiation at wavelengths somewhat shorter than those of visible light. Longwave or *near ultraviolet* extends from approximately 390 nanometers (nm) down to 50 nm. Shortwave or *far ultraviolet* extends from 50 nm down to 4 nm. **2.** Pertaining to the behavior and effects of electromagnetic radiation in the range of approximately 390 nm to 4 nm.

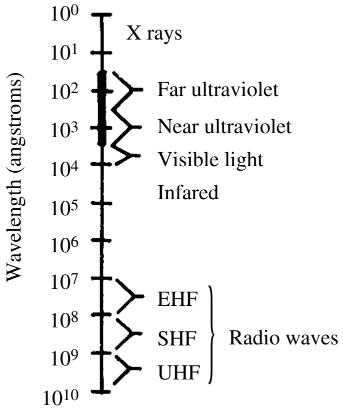

ultraviolet

ultraviolet altimeter An altimeter that uses ultraviolet rays instead of radio waves to determine the altitude of an aircraft or spacecraft.

ultraviolet lamp A lamp that delivers ultraviolet rays. Common types are arc lamps and mercury-vapor lamps.

ultraviolet light See ULTRAVIOLET RAYS.

ultraviolet power The power (in watts) of ultraviolet radiation.

ultraviolet rays Radiation at frequencies in the ultraviolet region (i.e., between the highest visible-light frequencies and the lowest X-ray frequencies).

ultraviolet therapy The use of ultraviolet rays by physicians and other health personnel in the treatment of certain disorders.

ultraviolet wavelength The wavelength range of ultraviolet radiation [e.g., 390 nanometers (nm) to 4 nm].

ultraviolet waves See ULTRAVIOLET RAYS.

umbilical cord **1.** A cord through which missiles and rockets are controlled and powered until they are launched. **2.** A cord connecting an astronaut to his space vehicle during extravehicular activity (e.g., during a space “walk”).

umbilical tower The vertical tower supporting the umbilical cords extending to a rocket in the launch position.

umbra **1.** The region of total (conical) shadow behind an object situated in the path of a radiation. **2.** The comparatively dark central region in a sunspot.

umbrella antenna An antenna consisting of a number of wires, extending from the top of a vertical mast to points on a circle below, at the center of which the mast is mounted. The wires are usually fed in parallel from the top.

Umklapp process A process responsible for thermal resistance in nonconducting materials. It results from collision between phonons, or phonons and electrons.

UMW Abbreviation of ULTRAMICROWAVE.

unabsorbed field strength The field strength with no absorption between transmitter and detector. The unabsorbed field strength is an ideal quantity; the actual field strength over a given distance is always less. It is expressed in volts per meter.

unamplified back bias A negative-feedback voltage developed across a fast-time-constant circuit within a single amplifier stage.

unamplified feedback A positive or negative current or voltage taken from the output of a system and presented to the input without being boosted through auxiliary amplification.

unamplified ALC Automatic level control (ALC) in which the control-signal voltage is taken from a point in the circuit and fed to the controlled point without being boosted by an auxiliary amplifier. Compare AMPLIFIED ALC.

unattended operation Operation, as of an electronically programmed machine, with minimal (or no) human supervision.

unattended time The period, excluding down time for repair or checkout, during which a computer is unpowered.

unattenuated Not reduced in intensity of amplitude. Thus, an unattenuated signal is one that has retained its original strength during its transmission through a system.

unavailable energy The difference between the quantity of heat energy supplied to a system and the available energy of the system.

unbalance **1.** Lack of balance or symmetry in a circuit, line, or system. **2.** The condition in which a bridge (or the equivalent) is not nulled.

unbalanced circuit See SINGLE-ENDED CIRCUIT.

unbalanced delta system A three-phase circuit in which the elements are connected in a triangular (delta) configuration, but are of unequal impedance. In an unbalanced delta system, no definite relationship exists between line and phase currents. Compare BALANCED DELTA SYSTEM.

unbalanced input See SINGLE-ENDED INPUT.

unbalanced line **1.** A transmission line in which one side (conductor) is grounded (e.g., a coaxial line). **2.** A normally balanced transmission line in which the currents in the two halves are out of balance (i.e., they are not equal in amplitude and opposite in phase).

unbalanced multivibrator A multivibrator in which the time constant of one base or gate circuit differs from that of the other base or gate circuit. The asymmetry results in short-duration output pulses separated by long time periods.

unbalanced output See SINGLE-ENDED OUTPUT.

unbalanced two-phase system A two-phase circuit in which the load elements are of unequal impedance.

unbalanced wire circuit A wire line whose (two) sides are dissimilar.

unbalanced-wye system A three-phase circuit in which the elements are connected in the familiar wye, or star, configuration, but are of unequal impedance.

unbalanced-Y system See UNBALANCED-WYE SYSTEM.

unbalance-to-balance transformer A transformer for matching an unbalanced transmission line to a balanced line (e.g., a coaxial line to a parallel-wire line).

unbiased limiting Peak limiting that results from overdriving an unbiased active device.

unbiased unit A device or circuit operated without bias [e.g., a transistor without base bias (common emitter and emitter follower) or emitter bias (common base)].

unblanking Removal of the blanking pulse in a cathode-ray-tube circuit (i.e., turning on the beam).

unblanking generator A (usually pulse-type) signal source for turning on the beam of a blanked cathode-ray tube.

unblanking interval The period during which the beam of a cathode-ray tube is turned on. Compare BLANKING INTERVAL.

unblanking pulse A pulse that turns on the beam of a blanked cathode-ray tube. Compare BLANKING PULSE.

unblanking time **1.** See UNBLANKING INTERVAL. **2.** The instant at which unblanking begins.

unbonded strain gauge A strain gauge that is not directly attached (by cement, for example) to the strained surface.

unbound electron A free electron (i.e., an electron not confined to a shell within an atom).

unbuffered output An output (signal, power, etc.) that is delivered directly from the generating device without the benefit of an isolating stage, such as a buffer amplifier. Compare BUFFERED OUTPUT.

unbypassed emitter resistor In a common-emitter transistor circuit, an emitter resistor without a bypass capacitor. The flow of output-signal current through the resistor produces negative feedback within a single stage.

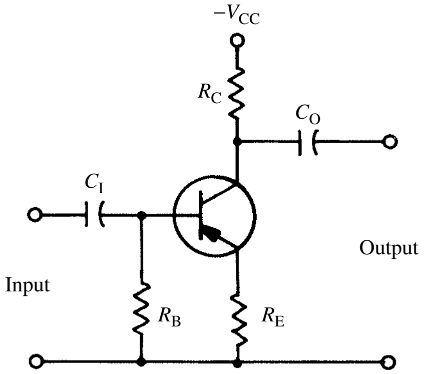

unbypassed emitter resistor (R_E)

unbypassed source resistor In a common-source field-effect transistor circuit, a source resistor without a bypass capacitor. The flow of output-signal current through the resistor produces negative feedback within a single stage.

uncage In a displacement gyroscope system, to disconnect the erection circuit.

uncalibrated unit A component, circuit, or instrument that has never been calibrated or has not recently been calibrated and, hence, is of questionable accuracy.

uncanny valley According to a theory put forth by certain Japanese roboticists, a phenomenon that can occur among people as technology advances. As machines rapidly get more sophisticated, a point is reached at which their human users become unnerved, partially offsetting the advantages the technology has to offer.

uncased choke See UNSHIELDED CHOKE.

uncased transformer See UNSHIELDED TRANSFORMER.

uncertainty in computation The degree of doubt concerning the exactness of computations. This uncertainty is always greater than that of the roughest measurement used in obtaining data for the computation. Uncertainty is closely related to the tolerance of the instruments and formulas used.

uncertainty in measurement The estimated maximum amount whereby the numerical value of a measured quantity can differ from the true value.

uncertainty principle The observation that high precision in the location of an electron is obtainable only at a sacrifice in the accuracy with which the momentum of the electron can be determined, and vice versa.

uncharged Without an electric charge, as opposed to discharged (depleted of a charge).

unclamp To switch off clamping action in a circuit. See CLAMPER and CLAMPING.

uncoated filament **1.** A plain filament (i.e., one without a coating of electron-increasing material). **2.** A stripped filament (i.e., one from which the electron-increasing coating has been burned off).

uncompensated Not modified to produce a desired type of performance, such as increased bandwidth or reduced temperature sensitivity. Example: an uncompensated amplifier put into service in the (wide) video band. Compare COMPENSATED AMPLIFIER, COMPENSATED CAPACITOR, and COMPENSATED DIODE DETECTOR.

uncompensated amplifier An amplifier without a provision for the automatic or manual modification or correction of its response. Compare COMPENSATED AMPLIFIER.

uncompensated capacitor A fixed or variable capacitor without a provision for the automatic or manual correction or modification of its capacitance or range or the improvement of its temperature coefficient. Compare COMPENSATED CAPACITOR.

uncompensated inductor A fixed or variable inductor without a provision for the automatic or manual correction or modification of its inductance or range or the improvement of its temperature coefficient.

uncompensated resistor A fixed or variable resistor without a provision for the automatic or manual correction or modification of its resistance or range or the improvement of its temperature coefficient.

unconditional branch See UNCONDITIONAL JUMP.

unconditional jump In computer operations, a program instruction that interrupts, without a relational test being made, the normal, ordered sequence of instructions and gives the address of (usually) the first instruction in a subroutine [e.g., BASIC's GOTO (line number)]. Also called *unconditional branch.* Compare CONDITIONAL JUMP.

unconditional stability Stability of a system at all frequencies and ages. Also called *absolute stability.*

unconditional transfer See UNCONDITIONAL JUMP.

uncontrolled multivibrator A multivibrator that is not synchronized with a signal source. Also called ASTABLE MULTIVIBRATOR and FREE-RUNNING MULTIVIBRATOR.

uncontrolled oscillator See SELF-EXCITED OSCILLATOR.

uncorrected power In an alternating-current circuit, a power value calculated without regard for power factor (i.e., the volt-ampere product).

uncorrected time Local standard time that has not been corrected in terms of the distance of the locality from the nearest standard-time meridian.

uncorrected tube A television picture tube operated without antipincushioning magnets.

uncoupled mode A mode of vibration existing in a system concurrent with (but independent of) other modes.

undamped galvanometer A galvanometer for which no provision has been made to limit overswing or prevent oscillation.

undamped meter **1.** A meter for which no provision has been made to limit overswing or prevent oscillation. **2.** A meter that is unprotected by a short circuit for limiting vibratory movement of the pointer during transport.

undamped natural frequency In the absence of damping, the oscillation frequency of a system having one degree of freedom when a transient force displaces the system momentarily from its position of rest.

undamped oscillation Continuous-wave oscillation. Also see CONTINUOUS WAVE.

undamped output circuit An audio power-amplifier output circuit that has not been designed so that overswing of the loudspeaker cone is prevented.

undamped speaker enclosure A loudspeaker cabinet for which provision has not been made to deaden resonances and other undesired vibration.

undamped wave See CONTINUOUS WAVE.

undeflected beam A cathode-ray beam or a light beam in its normal position of rest (quiescence).

underbiased unit A component, such as a transistor or vacuum tube, whose bias voltage or current is lower than the prescribed value. Compare OVERBIASED UNIT.

underbunching Less-than-optimum bunching of electrons in a velocity-modulated tube because of lowered buncher voltage.

undercompounded generator A generator in which the output voltage varies inversely with load resistance.

undercoupling Loose coupling, usually of an amount insufficient for optimum signal transfer.

undercurrent A current of lower than specified strength. Compare OVERCURRENT.

undercurrent relay A relay that is actuated when coil current falls below a predetermined level.

undercut The removal of the metal under the edge of the resist in a printed circuit by the etchant; thus the cross section of the conductor is reduced.

undercutting A phonograph-disc groove that is cut too shallow or with reduced internal movement of the recording stylus.

underdamping Insufficient damping of a system (i.e., not enough to prevent output oscillation following application of a transient force).

underdriven unit An amplifier, oscillator, or transducer whose driving signal (current, voltage, power, or other quantity) is lower than the prescribed value. Compare OVERDRIVEN UNIT.

underexcited Receiving lower-than-normal excitation, as in an underdriven final amplifier of a radio transmitter. Compare OVEREXCITED.

underflow In computer operations, the condition in which a quantity entered into storage is shorter than the space provided for it (e.g., a 12-digit quantity in a 16-position register).

underground antenna A transmitting or receiving antenna installed and operated below ground. Included are buried antennas and antennas on equipment operated in tunnels, cellars, and similar areas.

underground cable A cable that is buried in the earth.

underground communication Communication between a transmitter and receiver, both below the surface of the earth.

underground image The below-ground mirror image of an antenna; it combines with the actual antenna to form the complete radiation pattern.

underground line A power line laid below the surface of the earth.

underground receiver **1.** A receiver situated completely below the surface of the earth. **2.** A clandestine receiver.

underground reception Reception of an above-ground station's transmissions by an underground receiver. Compare UNDERGROUND COMMUNICATION.

underground transmitter **1.** A transmitter situated completely below the surface of the earth. **2.** A clandestine transmitter.

under insulation The insulation (usually a strip of tape) laid under a wire brought up from the center of a coil.

underinsulation Inadequate or insufficient insulation.

underlap In a facsimile or television picture, a crowding of the scanned lines.

underload circuit breaker See CIRCUIT BREAKER.

underloaded amplifier **1.** An amplifier whose load resistance (impedance) is less than the prescribed value. **2.** A power amplifier delivering less than its rated output power. Compare OVERLOADED AMPLIFIER.

underload relay A relay actuated when circuit current drops below a predetermined value. Compare OVERLOAD RELAY.

undermodulation Incomplete modulation of a carrier wave. Compare COMPLETE MODULATION and OVERMODULATION.

underpass The crossing of two conductors on a semiconductor wafer, without an electrical connection.

underpower relay A relay actuated when power drops below a predetermined level. Compare OVERPOWER RELAY.

underrate To assign a rating (e.g., current or power) lower than the quantity of the rating an equipment can handle, or tolerate. For safety or reliability, apparatus sometimes is deliberately underrated—especially in power output and maximum current or voltage.

undershoot On an oscilloscope screen or graph, a momentary swing of a current or voltage below the reference axis. Compare OVERSHOOT.

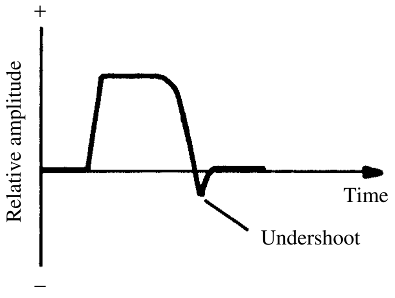

undershoot

undershoot distortion Distortion caused by reduction of the maximum amplitude of a signal wavefront below the steady-state amplitude that would be reached by a prolonged signal wave.

understudy See BACKUP.

underthrow A form of signal distortion that occurs when the modulating-waveform frequency is too high in proportion to the frequency of the wave itself.

undervoltage A voltage of lower than specified value. Compare OVERVOLTAGE.

undervoltage protection The automatic disconnection of a load device from its driving source when the driving voltage falls below a predetermined (threshold) level. This action is sometimes accomplished with a Zener diode (whose breakdown voltage is equal to the threshold voltage) in series with a disconnect relay.

undervoltage relay A relay actuated when voltage drops below a predetermined level. Compare OVERVOLTAGE RELAY.

underwater antenna A transmitting or receiving antenna normally operated in a body of water. It is usually associated with a submarine.

underwater sound projector A hydroacoustic transducer that converts audio-frequency power into sound waves, which it radiates through a body of water to a receiver.

Underwriters Code The advisory National Electrical Code adopted by the National Fire Protection Association; it is often enforced by courts of law and inspection agencies.

Underwriters Laboratories Abbreviation, UL. A private corporation that issues standards of safety for electrical components and equipment, and for their operation.

undistorted power output For active amplifier devices, a specified maximum audio power output level at which the total distortion does not exceed a specified low value (i.e., at which operation is practically distortionless).

undistorted wave **1.** A sine wave that contains essentially no harmonic energy. **2.** A nonsinusoidal wave whose shape corresponds exactly to the equation for the wave.

undisturbed-one output In digital-memory operations, the one-output of a magnetic cell or other memory unit that has received a full, rather than partial, read pulse. Compare UNDISTURBED OUTPUT SIGNAL.

undisturbed output signal In digital-memory operations, the output of a magnetic cell or other memory unit previously set to one or zero, and that has received a full, rather than partial, read pulse. Also called *undisturbed response voltage.*

undisturbed response voltage See UNDISTURBED OUTPUT SIGNAL.

undisturbed-zero output In digital-memory operations, the zero-output of a magnetic cell or other memory unit that has received a full, rather than partial, read pulse. Compare UNDISTURBED OUTPUT SIGNAL.

Universal Truth Machine (UTM) A hypothetical computer capable of proving any logically true statement. According to the Incompleteness Theorem proved in 1930, such a machine cannot exist for a first-order logical system.

undoped Pertaining to a pure semiconductor material (i.e., one containing no lattice-altering additives).

undulating current A current, such as an alternating or composite current, whose value oscillates in the manner of a wave.

undulation A wavelike alternation.

undulatory theory The theory that light is transmitted by means of an undulating (wavelike) movement between the luminous object and the eye.

unequal alternation In an alternating-current waveform, a form of asymmetry in which the positive and negative half-cycles are not exact mirror images; they might vary in peak amplitude, duration, shape, etc.

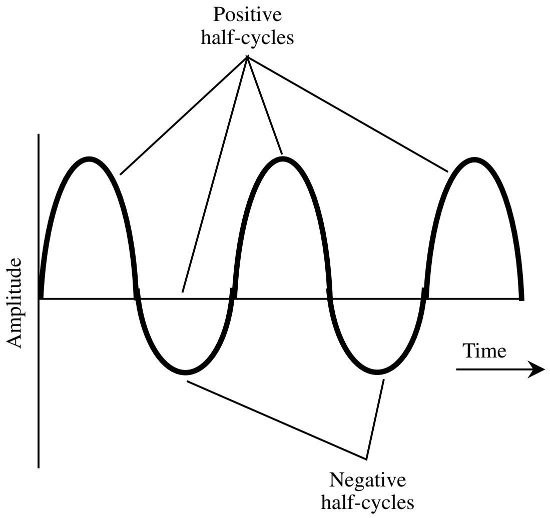

unequal alternation

unexpected halt During a computer program run, an undirected (unplanned) halt (e.g., one caused by a machine fault or program bug).

unfiltered Not having been subjected to filtering action. An example is the pulsating-direct-current output of an alternating-current-operated rectifier circuit.

unfocused light source In a photoelectric system, a light source that delivers diffused light.

unformed rectifier A newly fabricated semiconductor rectifier (especially selenium) or an electrolytic rectifier before it has been electroformed for improved characteristics.

unfurlable antenna An antenna that can be unrolled to increase its length, thereby lowering its fundamental resonant frequency; it can also be rolled up to decrease its length and increase its fundamental resonant frequency.

ungrounded item A component, circuit, or circuit point having no connection to ground, except an inadvertent one through common impedances or leakage paths.

ungrounded system A system operated entirely above ground, any path to ground being accidental.

unguided Without electronic guidance (pertaining to missiles, rockets, satellites, etc.).

uni- A prefix (combining form) meaning *one* or *single* and appearing in a number of electronic terms. See the following several definitions for examples.

uniaxial **1.** Having one axis. **2.** Referred to one axis.

uniaxial crystal A crystal having one optical axis.

unibivalent electrolyte An electrolyte, such as sodium carbonate (Na_2CO_3), that dissociates into two univalent ions and one bivalent ion.

Unicode Trade name for a 16- or 32-bit binary digital code that allows for the representation of characters and glyphs in all the written languages of the world. The 16-bit code contains 65,536 code points; the 32-bit version contains more than 1 million code points.

uniconductor cable A cable having a single conductor, usually of braided or twisted wires.

uniconductor waveguide A waveguide consisting of a metal sheath deposited on a solid dielectric of cylindrical or rectangular cross section.

unidentified flying object Abbreviation, UFO. An object claimed to have been seen in flight, but not identified by reliable authorities as any known type of vehicle.

unidirect To commutate (see COMMUTATOR).

unidirectional **1.** Flowing or acting in one direction only. **2.** Having a radiation or response sensitivity that is maximum in primarily one direction in space.

unidirectional antenna See UNIDIRECTIONAL ARRAY.

unidirectional array A beam antenna that radiates in one direction only, or principally in one direction, unless rotated.

unidirectional conductivity Electrical conductivity in only one direction; it characterizes an ideal diode.

unidirectional coupler A directional coupler sampling only one direction of transmission.

unidirectional current A current that flows always in the same direction, although it might fluctuate in intensity.

unidirectional elements Circuit elements, such as diodes or transistors, that transmit energy in only (or best in) one direction.

unidirectional field-effect transistor See UNILATERAL FIELD-EFFECT TRANSISTOR.

unidirectional hydrophone An underwater unidirectional microphone.

unidirectional loudspeaker A loudspeaker that radiates sound substantially in one direction.

unidirectional microphone A microphone that receives sound waves from one direction, usually from the front, minimum response usually being from the sides and back.

unidirectional network A network that transmits signals in only one direction (i.e., the input and output terminals are not interchangeable).

unidirectional pattern For a transducer (such as an antenna, speaker, or microphone), a radiation or response pattern that shows a pronounced maximum in one direction only.

unidirectional pulse A single-polarity pulse.

unidirectional pulse train A series of unidirectional pulses.

unidirectional response For a receiving transducer, such as an antenna or microphone, a response pattern that shows a pronounced maximum in one direction only.

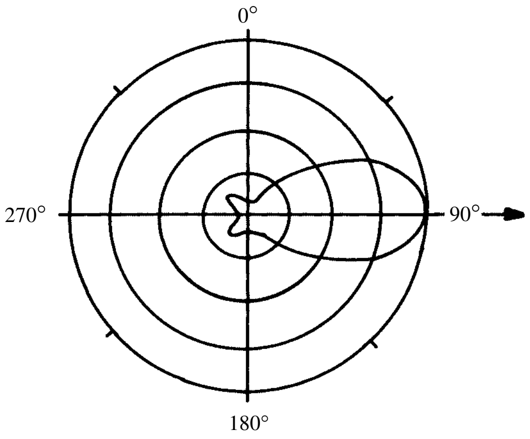

unidirectional pattern

unidirectional speaker See UNIDIRECTIONAL LOUDSPEAKER.

unidirectional transducer A transducer that operates in one direction only (i.e., its input and output cannot be interchanged).

unidirectional voltage A voltage that never changes polarity, although it might fluctuate in value.

UNIFET Abbreviation of UNIPOLAR FIELD-EFFECT TRANSISTOR. (Also, UFET.)

unified-field theory A theory, as yet unproved, that all forces in nature are interrelated and are controlled by the same causes and factors. Albert Einstein worked on this theory after he successfully developed the theories of relativity.

unifilar Having (or wound as) one fiber, wire, filament, or thread.

unifilar magnetometer A magnetometer in which a bar is suspended at its center of gravity by a single thread.

unifilar suspension A method of meter construction in which the moving part is suspended by a single filament. The torque might be provided by this filament or by a spring or other resisting device; the greater the displacement, the greater the torque.

uniform circular motion Motion at a uniform rate and describing a circle (e.g., a motor armature rotating at a constant speed).

uniform electric field An electric field in which all the lines of flux are straight and parallel, and in which the electrostatic force has the same value at all points (e.g., the field between two oppositely charged, flat, parallel plates).

uniform field See UNIFORM ELECTRIC FIELD and UNIFORM MAGNETIC FIELD.

uniform frequency response Frequency response that is flat throughout a specified range. Such response is characterized by the transmission of a signal with no introduced amplitude or phase distortion.

uniform line A transmission line having identical electrical properties over its entire length.

uniform magnetic field A magnetic field in which all the lines of flux are straight and parallel, and in which the magnetic force has the same value at all points.

uniform plane wave A free-space plane wave at an infinite distance from the generator, having constant-amplitude electric and magnetic field vectors over the equiphase surfaces.

uniform precession In regions of the uniform magnetic field of a sample of material, the state in which the magnetic moments of all atoms are parallel and precess in phase around the magnetic field.

uniform waveguide A waveguide having constant electrical and physical characteristics along its axis.

uniground **1.** The grounding of a circuit at one point to reduce susceptibility to hum and noise. **2.** The point at which such a connection is made.

unijunction transistor A semiconductor device consisting of a thin silicon bar on which a single pn junction acting as an emitter is formed near one end. Two bases are provided—each an ohmic connection made to one end of the bar. Also called *double-base diode.*

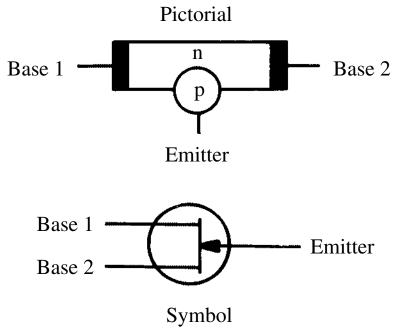

unijunction transistor

unilateral area track A film sound track having modulation on only one edge of the opaque area.

unilateral bearing In radio direction finding, a bearing obtained with a finder that produces unilateral response and thereby eliminates 180-degree error.

unilateral conductivity See UNIDIRECTIONAL CONDUCTIVITY.

unilateral element See UNIDIRECTIONAL ELEMENTS.

unilateral field-effect transistor A field-effect transistor whose source and drain terminals cannot be interchanged. Also called ASYMMETRICAL FIELD-EFFECT TRANSISTOR. Compare SYMMETRICAL FIELD-EFFECT TRANSISTOR.

unilateralization A system of neutralization used in transistor amplifiers at very-high and ultrahigh frequencies, in which the internal capacitive and resistive components are compensated by the neutralization feedback.

unilateralized amplifier A transistor amplifier in which both the internal resistive and capacitive components are compensated by the neutralizing circuit. Also see UNILATERALIZATION.

unilateral network See UNIDIRECTIONAL NETWORK.

unilateral transducer See UNIDIRECTIONAL TRANSDUCER.

uninterruptible power supply Abbreviation, UPS. A device that provides a continuous utility current to electronic devices in the event of a utility power dip or blackout. It contains a battery and a power inverter. These devices are used extensively in personal computing to prevent loss of data that could otherwise result from a utility power failure.

union **1.** The logical inclusive-OR operation. **2.** See BOND, **1**.

union catalog In computer operations, the compiled list of the contents of two or more tape or disk libraries.

unionic material A material having no ions (i.e., one in which all atoms are neutral).

uniphase antenna See COLLINEAR ANTENNA.

UNIPOL Acronym for *universal problem-oriented language*, a high-level computer-programming language.

unipolar **1.** Having or using a single pole or polarity. **2.** Operating with one class of current carrier.

unipolar armature An electric-motor armature that maintains its polarity throughout a complete revolution.

unipolar field-effect transistor See UNIPOLAR TRANSISTOR.

unipolar induction Induction by only one pole of a magnet.

unipolar input The input circuit of an instrument or device designed for input signals of one polarity only.

unipolar pulse A pulse in which the current flows in only one direction, or in which the voltage occurs with only one polarity.

unipolar transistor A field-effect transistor (FET). It utilizes only one kind of carrier (electrons in the n-channel FET, holes in the p-channel FET). Compare BIPOLAR TRANSISTOR.

unipolar winding The winding of a UNIPOLAR ARMATURE.

unipole **1.** An all-pass filter having one pole and for which there is one zero. Also see ALL-PASS FILTER, POLES OF IMPEDANCE, and ZEROS OF IMPEDANCE. **2.** A hypothetical, omnidirectional antenna.

unipotential cathode An indirectly heated tube cathode. Also called *equipotential cathode.* See INDIRECTLY HEATED CATHODE.

unit **1.** A named, single magnitude adopted as a standard by physical measurement. Thus, the unit of current is the ampere, the unit of frequency is the hertz, the unit of capacitance is the farad, etc. Also see ABSOLUTE SYSTEM OF UNITS, INTERNATIONAL SYSTEM OF UNITS, and CENTIMETER-GRAM-SECOND. **2.** A single piece or assemblage of equipment, such as amplifier, converter, power supply, etc. **3.** A quantity of 1 implied when unit is the adjective describing a quantity (e.g., *unit length* means a distance of 1 meter, 1 foot, etc.).

unitary code In computer operations, a code based on one digit. The number of times the digit is repeated indicates a given number.

unit cell In crystals, the simplest polyhedron exhibiting all the structural features and which is repeated to form the crystal lattice.

unit charge See UNIT ELECTROSTATIC CHARGE.

unit electric charge See UNIT ELECTROSTATIC CHARGE.

unit electrostatic charge An electrostatic point charge that will attract or repel a point charge of equal strength 1 centimeter away in a vacuum, with a force of 1 dyne (10^{-5} newton).

unit function Symbol, *H* or *1.* A time-dependent quantity that is zero before the start of a period (when time *t* is zero) and 1 for all values of *t* greater than zero. It is approximated by a square wave, and is useful in solving problems involving transients.

unit length **1.** A fundamental unit of distance or time, used for reference in a measuring system. For example, in the mks (meter-kilogram-second) system, the distance unit length is the meter, and the time unit length is the second. **2.** The duration of a fundamental element, or bit, in a binary-code transmission system.

unitized construction The fabrication of an electronic equipment in subassemblies (such as modules) that can be tested separately, and that can be easily replaced (plugged in) in the event of individual failure.

unit line A line of electric or magnetic flux.

unit magnetic pole See UNIT POLE.

unit matrix A matrix whose main diagonal terms (all) are all at unity, the other terms being zero.

unitooth Pertaining to the use in electrical machinery of one slot per pole per phase.

unitor See OR GATE.

unit pole The strength (magnetic flux) of a (hypothetical) magnetic pole that will attract or repel a pole of equal strength 1 centimeter away in a vacuum, with a force of 1 dyne; 1 unit pole = 1.257×10^{-7} weber.

unit-ramp function A function whose value is zero before time *t*, and becomes equal to the time measured from *t* at all other instants. The integral of the UNIT FUNCTION.

unit record In computer operations, a complete record consisting of many data elements and contained within one storage medium, such as a magnetic disk.

unitrivalent electrolyte An electrolyte, such as sodium phosphate (Na_3PO_4), that dissociates into three univalent ions and one trivalent ion.

units position In a numbering system, the rightmost position in a multidigit whole number; or, if the number contains a radix point, the digit immediately to the left of the radix point.

unit-step function See UNIT FUNCTION.

unitunnel diode A special form of semiconductor diode, used as an oscillator at ultra-high and microwave frequencies.

unit vector A vector that is 1 unit long in terms of the scale representing a factor of interest, and has the same direction as the vector of interest. For example, if *u* is a unit vector, then 5*u* represents a vector having the same direction and five times the magnitude of *u*.

unity **1.** The figure 1 implied. **2.** A ratio of 1:1. **3.** A gain of 0 dB.

unity coupling Tight coupling between two coils, the turns ratio often being 1 to 1 and the coils always being closely interwound.

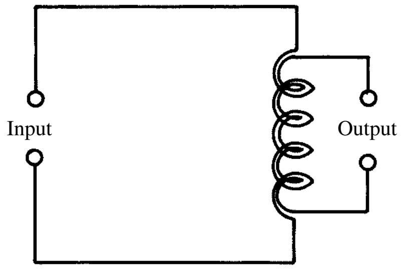

unity coupling

unity gain The condition in which the output amplitude is the same as the input amplitude; that is, a gain factor of 1, or 0 dB.

unity-gain bandwidth **1.** For an active filter, the bandwidth between the frequencies at which the gain is 0 dB. **2.** For an operational amplifier having a rolloff of 6 dB per octave, the frequency at which the open-loop gain is 0 dB.

uni-univalent electrolyte An electrolyte that dissociates into two univalent ions [e.g., sodium chloride (NaCl)].

univ Abbreviation of *universal.*

UNIVAC The name given to a family of digital computers and data-processing systems. The name is an acronym for *Universal Automatic Computer.*

univalent Having a valence of 1. For example, in the compound sodium chloride (NaCl), the sodium ion (Na+) and the chlorine ion (Cl–) are univalent.

universal bridge A bridge for the measurement of capacitance (*C*), inductance (*L*), and resistance (*R*). Such a bridge often is a skeletal foundation circuit with a provision for plugging L, C, or R standards in the various arms, as required. Also called *general-purpose bridge.*

universal coordinated time Abbreviation, UTC. Greenwich mean time coordinated by the International Time Bureau for broadcast as signals (on WWV, for example).

universal coupler **1.** An arrangement of one or more inductors and variable capacitors for matching a transmitter to virtually any antenna. One such device is the Collins coupler. **2.** A device for matching numerous generator output impedances to numerous load impedances.

universal filter An active filter that is continuously tunable over a wide frequency range and that offers low-pass, high-pass, bandpass, and band-suppression functions.

universal frequency counter A digital frequency meter usable at radio and audio frequencies.

universal motor A small series-wound motor that runs on direct current or single-phase alternating current. This type of motor is used in many household appliances and in motor-driven tools, such as portable electric drills, polishers, etc.

universal output transformer An output transformer having a number of taps on its primary and secondary windings for use with a wide variety of impedances.

universal product code Abbreviation, UPC. The variable-width-bar code appearing on price tags or product labels, and yielding such information as cost, size, and color when read by supermarket (or other retailer) optical character-recognition equipment. Also called *universal vendor marking (UVM).*

universal receiver A radio receiver that can be operated either from alternating current or direct current (i.e., from utility power or battery power).

universal shunt See AYRTON-MATHER GALVANOMETER SHUNT.

universal time See GREENWICH MEAN TIME.

universal transformer See UNIVERSAL OUTPUT TRANSFORMER.

universal transmitter A radio transmitter that can be operated either from alternating current or direct current (i.e., from utility power or battery power).

universal Turing machine In computer theory, a TURING MACHINE capable of simulating other Turing machines.

universal vendor marking See UNIVERSAL PRODUCT CODE.

universal winding A zigzag winding used to reduce the distributed capacitance of multilayer coils. Also called *lattice winding* and *honeycomb winding.*

universal-wound coil A coil using a universal winding. Such coils are common in intermediate-frequency (IF) transformers and in radio-frequency (RF) chokes.

univibrator See MONOSTABLE MULTIVIBRATOR.

unknown quantity Any variable quantity sought in calculations; it is usually represented by lowercase italic letters from the second half of the alphabet (e.g., x, y, z).

unlike charges Dissimilar electric charges (i.e., positive and negative).

unlike poles Dissimilar poles, such as the north and south poles of a magnet.

unload **1.** To remove data from a file. **2.** To disconnect the load from a circuit.

unloaded amplifier See UNTERMINATED AMPLIFIER.

unloaded antenna An antenna operated without added inductance or capacitance.

unloaded-applicator impedance In dielectric-heating operations, the unloaded impedance of applicator electrodes placed in their normal working position without the dielectric-material load in place.

unloaded battery **1.** A battery in the standby condition. **2.** A battery tested for open-circuit terminal voltage (i.e., with no load, except the voltmeter).

unloaded generator See UNTERMINATED GENERATOR.

unloaded line See UNTERMINATED LINE.

unloaded potentiometer A potentiometer or voltage divider with an open-circuited output.

unloaded Q The Q factor (degree of selectivity) of a coil or tuned circuit that is activated by a signal, but that delivers no output to a load.

unloaded transmitter A transmitter operated under open-circuit conditions (i.e., without an external load, such as an antenna or dummy resistor).

unmanned factory A manufacturing plant run largely by robots and computers. There must generally be at least one human being, whose job(s) is/are to oversee and maintain the operation of the machines.

unmatched elements **1.** Components (such as resistors, capacitors, semiconductors, etc.) having different values. **2.** Mating elements or devices not having the same impedance. Also called MISMATCH.

unmodulated carrier See CONTINUOUS WAVE.

unmodulated current A current whose characteristics (amplitude, phase, or frequency) are not varied by a signal or by noise.

unmodulated voltage A voltage whose characteristics (amplitude, phase, or frequency) are not varied by a signal or by noise.

unmodulated wave See CONTINUOUS WAVE.

unnilennium See MEITNERIUM.

unnilhexium See SEABORGIUM.

unniloctium See HASSIUM.

unnilpentium See DUBNIUM.

unnilquadium See RUTHERFORDIUM.

unnilseptium See BOHRIUM.

unode See MONODE.

unpack In computer operations, to remove data from a storage location into which, with other data, it has been packed (as a memory-conservation measure).

unpolarized light Light in which the wave orientation is random around the axis of the beam.

unpolarized plug A plug that can be inserted into a socket in two or more different ways, thus increasing the likelihood of a wrong connection. Compare POLARIZED PLUG.

unpolarized receptacle A receptacle into which a plug can be inserted in two or more different ways, thus increasing the likelihood of a wrong connection. Also called UNPOLARIZED SOCKET. Compare POLARIZED RECEPTACLE.

unpolarized relay A relay that always closes in the same direction, regardless of the direction of current in its coil. Also called *nonpolarized relay.*

unpolarized socket A socket into which a component or plug can be inserted in two or more different ways, thus increasing the likelihood of wrong connections. Also called UNPOLARIZED RECEPTACLE.

unprotected antenna An outside antenna operated without a lightning arrester or grounding switch. Increases the risk of damage to equipment, such as radio receivers and transmitters, connected to the antenna. It also increases the danger to personnel using the equipment.

unrationalized systems of units Collectively, the absolute centimeter-gram-second (cgs) electrostatic system of units, the absolute cgs electromagnetic system of units, and the absolute meter-kilogram-second (mks) system of units.

unreflected ray See UNREFLECTED WAVE.

unreflected wave **1.** See DIRECT WAVE. **2.** See GROUND WAVE. **3.** An electromagnetic wave that penetrates the ionosphere and passes into space.

unregulated Not held to a constant value. For example, an unregulated voltage is free to fluctuate in value.

unregulated power supply A power supply whose output current or voltage is not automatically held to a constant value. Compare CONSTANT-CURRENT SOURCE and CONSTANT-VOLTAGE SOURCE.

uns Abbreviation of UNSYMMETRICAL.

unsaturated core A magnetic core operated below the point of saturation (i.e., below the point at which an increase in coil current produces no further increase in magnetization of the core).

unsaturated logic A digital-logic scheme in which the switching devices operate between the saturated and cut-off conditions during either or both of the high and low states.

unsaturated operation Operation of a device at a point below saturation [i.e., below the point at which (1) an increase in voltage produces no further increase in current, or vice versa, or (2) an increase in coil current produces no further increase in magnetization of a core]. Compare SATURATED OPERATION.

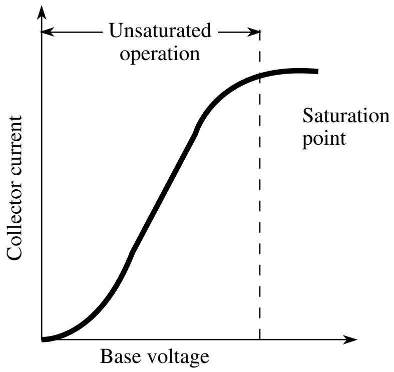

unsaturated operation

unsaturated standard cell See STANDARD CELL.

unshielded cable A cable, such as a twisted pair or a multiwire twist, lacking a shielding jacket. Unless special precautions are taken, such as transposing conductors, such a cable is susceptible to stray pickup and is capable of radiation.

unshielded choke An uncased choke (i.e., one without a protective and shielding metal housing).

unshielded coil An inductor without a field-confining enclosure.

unshielded probe An instrument probe that has no shielding enclosure. Such a probe is desirable for some tests, but it is subject to antenna-pickup effect, body capacitance, and stray-field pickup.

unshielded stage In electronic equipment, a stage operating entirely in the open (i.e., without electrostatic or magnetic shielding enclosures). It is, therefore, susceptible to stray pickup and is capable of radiation.

unshielded transformer **1.** An uncased transformer (i.e., one without a protective and shielding metal housing). **2.** A transformer having no electrostatic shield between its windings.

unshift The mechanical action in a teletypewriter when the carriage moves from the figures position to the letters position.

unshunted current meter A single-range D'Arsonval milliammeter or microammeter that has no shunt resistor. The resistance of the instrument is, therefore, equal to the resistance of the movable coil.

unsolder To separate wires, contacts, or fastenings by melting the solder that holds them together. Also called DESOLDER.

unsophisticated electronics Electronics theory and operations of the simplest kind, involving little or no mathematics. It is generally intended for a hobbyist audience. Compare SOPHISTICATED ELECTRONICS.

unstable oscillation **1.** Relaxation oscillation. **2.** The periodically interrupted oscillation characteristic of a blocking oscillation.

unstable region The negative-resistance portion of an N- or S-shaped response curve.

unstable servo A servo having unlimited output drift.

unstable state A condition, such as a negative-resistance region, that is difficult to maintain and often results in oscillation. The negative-resistance region of the tunnel diode forward-conduction curve is evidence of such a state. Compare STABLE STATE.

untapped winding A winding having two terminals, one at either end, and no intermediate terminals or connections. Also called *untapped coil.*

unterminated amplifier An amplifier operated without a load device (i.e., under open-circuit output conditions).

unterminated generator A generator operated without a load device (i.e., under open-circuit output conditions).

unterminated line A transmission line that is not terminated with an impedance (i.e., an open-ended line).

untriggered flip-flop A flip-flop that, at a particular instant, receives no actuating pulse, and that, therefore, does not change state.

untuned amplifier A radio-frequency amplifier that is not tuned to a single frequency, but has useful gain over a wide band of frequencies. Examples are the *distributed amplifier* and the *video amplifier.*

untuned filter See ALL-PASS FILTER.

untuned line An aperiodic transmission line (i.e., one that is only tuned to a particular frequency by its own distributed characteristics).

untuned transformer A transformer having simple primary and secondary windings and no tuning devices (such as capacitors in series or parallel with the windings), and designed so that its natural resonant frequency (because of distributed capacitance) lies outside the specified range of operation.

ununbium Symbol, Uub. Atomic number, 112. The most common isotope has atomic weight 277. Classified as a transition metal. It is human-made and not known to occur in nature.

ununhexium Symbol, Uuh. Atomic number, 116. The most common isotope has atomic weight 289. First reported in January 1999. It is a decomposition product of ununoctium, and it in turn decomposes into ununquadium. It is not known to occur in nature.

ununnilium Symbol, Uun. Atomic number, 110. The most common isotope has atomic weight 269. Classified as a transition metal. It is human-made and not known to occur in nature.

ununoctium Symbol, Uuo. Atomic number, 118. The most common isotope has atomic weight 293. It is the result of the fusion of krypton and lead, and decomposes into ununhexium. It is not known to occur in nature.

ununquadium Symbol, Uuq. Atomic number, 114. The most common isotope has atomic weight 285. First reported in January 1999. It is human-made, and not known to occur in nature.

unununium Symbol, Uuu. Atomic number, 111. The most common isotope has atomic weight 272. Classified as a transition metal. It is human-made, and not known to occur in nature.

unused time See UNATTENDED TIME.

unweighted average In the statistical analysis of data, an average calculated from terms not weighted by a factor. Compare WEIGHTED AVERAGE. Also see UNWEIGHTED TERM.

unweighted term In the statistical analysis of data, a term not operated upon by a weighting factor. Therefore, it has no extraordinary influence on a final result, such as an average. Compare WEIGHTED TERM.

unwind To eliminate all redundant or unnecessary operations in a computer system.

up-and-downer amplifier In broadcasting and sound amplification, an auxiliary amplifier that allows the level of sound effects and background music to be controlled automatically without disturbing the dialogue levels.

UPC Abbreviation of UNIVERSAL PRODUCT CODE.

upconv Abbreviation of UP CONVERTER.

up convert In superheterodyne conversion, to heterodyne a signal to an intermediate frequency higher than the signal frequency. Compare DOWN CONVERT.

update In data-processing operations, to make a file reflect the current status of pertinent information by using transactions for its modification.

up-down counter See BIDIRECTIONAL COUNTER.

uplink **1.** The signal sent up to a satellite transponder. **2.** The frequency of the signal sent up to a satellite transponder.

uplink frequency **1.** The specific frequency of a given signal sent up to a satellite transponder. **2.** The band of frequencies sent up to a satellite transponder. There can be many signals in an uplink; there can also be many different stations sending uplink signals within the band.

uplink power **1.** The power output of the transmitter sending the signal up to a satellite transponder. **2.** The effective radiated power of the station sending the uplink signal.

uplink station Any station that transmits a signal, or group of signals, up to a satellite transponder.

upper atmosphere See IONOSPHERE, STRATOSPHERE, and TROPOSPHERE.

upper sideband Abbreviation, USB. In an amplitude-modulated wave, the component whose frequency is the sum of the carrier frequency and the modulation frequency. Compare LOW-ER SIDEBAND.

upper sideband, suppressed carrier Abbreviation, USSC. A single-sideband transmission technique in which the upper sideband is transmitted, the lower sideband is suppressed, and the carrier is suppressed. Compare DOUBLE SIDEBAND, SUPPRESSED CARRIER, LOWER SIDEBAND, and SUPPRESSED CARRIER.

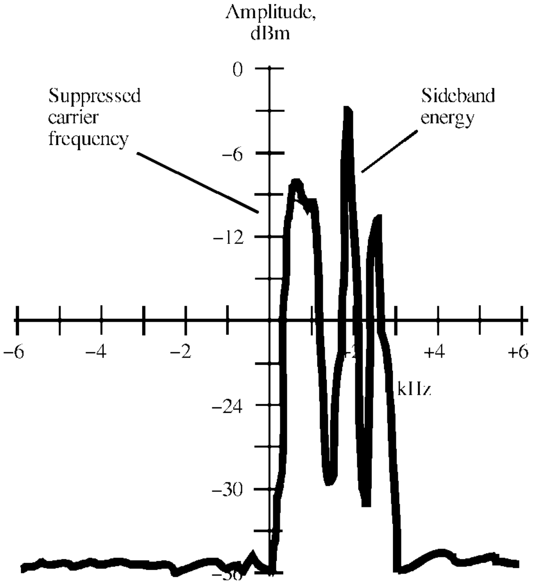

upper sideband

UPS Abbreviation for UNINTERRUPTIBLE POWER SUPPLY.

upside down **1.** In radioteletype, a signal in which the MARK and SPACE signals are reversed in frequency. **2.** In single sideband, reception of a signal on the wrong sideband, resulting in "monkey chatter."

up time See SERVICEABLE TIME.

upturn A usually sudden rise in a performance curve. Compare DOWNTURN.

upward modulation Modulation in which the output current of the modulated circuit increases during modulation. Compare DOWNWARD MODULATION.

UR Abbreviation of ULTRARED.

uranium Symbol, U. A radioactive metallic element. Atomic number, 92. Atomic weight, 238.03.

uranium 235 Symbol, U-235. An isotope of uranium of mass number 235, used in atomic bombs and atomic power plants.

uranium 238 Symbol, U-238. An isotope of uranium of mass number 238, which, under fast-neutron bombardment, decays to become neptunium and then plutonium (of mass number 239).

uranium metals The group of heavy, radioactive elements of atomic numbers 89 through 103. Also called the *actinide series*.

uranium rays The ionizing radiation emitted by uranium. In 1896, Antoine Henri Becquerel observed that uranium salts emit rays (later identified as alpha, beta, and gamma) that can pass through bodies opaque to light and are capable of exposing a photographic plate. Also called *Becquerel rays.*

urea formaldehyde resin A thermosetting, synthetic resin used in the manufacture of a number of plastic dielectric bodies and formed by the reaction of urea with formaldehyde.

urea plastic Any of several thermosetting plastics having a urea base and used as a dielectric and as a molding material in electronics (e.g., *urea formaldehyde resin*).

URSI Abbreviation of *Union Radio Scientifique Internationale (International Radio Scientific Union).*

ursigram A broadcast message giving information on sunspots, radio propagation, terrestrial magnetism, and related subjects. Also see URSI.

usable frequency **1.** Any frequency at which communications can be maintained between two points via ionospheric propagation. **2.** Any frequency at which a communications system is operational.

usable sample **1.** The portion of an oscilloscope or monitor display that is visible on the screen. **2.** In statistics, a sample that is considered valid for calculation purposes.

USB Abbreviation of UPPER SIDEBAND.

user friendliness **1.** For electronic equipment and systems, the quality of being easy for people to operate. **2.** The relative ease with which a machine, especially a computer, can be operated by people.

useful life The elapsed time between the installation of an expendable component, circuit, or system, and the time it must be replaced.

useful line In television, the portion of a scanning line that can be used for picture signals. Also called AVAILABLE LINE.

user group A group of hobbyists or company representatives having in common the ownership and/or use of a specific brand of computer, and who meet or otherwise interact to share their expertise with, or programs, for, the machine. Also applicable to similar groups of programmable-calculator owners.

user's library A compilation of programs supplied to (by the vendor) or generated by a user group.

USM Abbreviation of *underwater-to-surface missile.*

USSC Abbreviation of UPPER SIDEBAND, SUPPRESSED CARRIER.

UTC Abbreviation of UNIVERSAL COORDINATED TIME.

utility **1.** An organization providing a public service, such as electric power or electronic communications. **2.** In computer operations, a program or set of programs intended for diagnosing and/or correcting hardware and software problems.

utility box A general-purpose aluminum or steel box, supplied in various convenient sizes (painted or unpainted), used as a housing or shield for electronic equipment.

utility factor For a transformer used in a direct-current (dc) power supply, the ratio of dc output to required kilovolt-ampere (kVA) capacity.

utilization factor See UTILITY FACTOR.

UTL Abbreviation of *unit transmission loss.*

UUM Abbreviation of *underwater-to-underwater missile.*

UV **1.** Abbreviation of ULTRAVIOLET. **2.** Abbreviation of UNDERVOLTAGE.

Uvicon A television camera tube in which a conventional vidicon scanning section is preceded by an ultraviolet-sensitive photocathode, an electron-accelerating section, and a special target.

Uviol lamp An ultraviolet lamp using uviol glass—a special glass that is transparent to ultraviolet rays.

UVM Abbreviation of UNIVERSAL VENDOR MARKING.

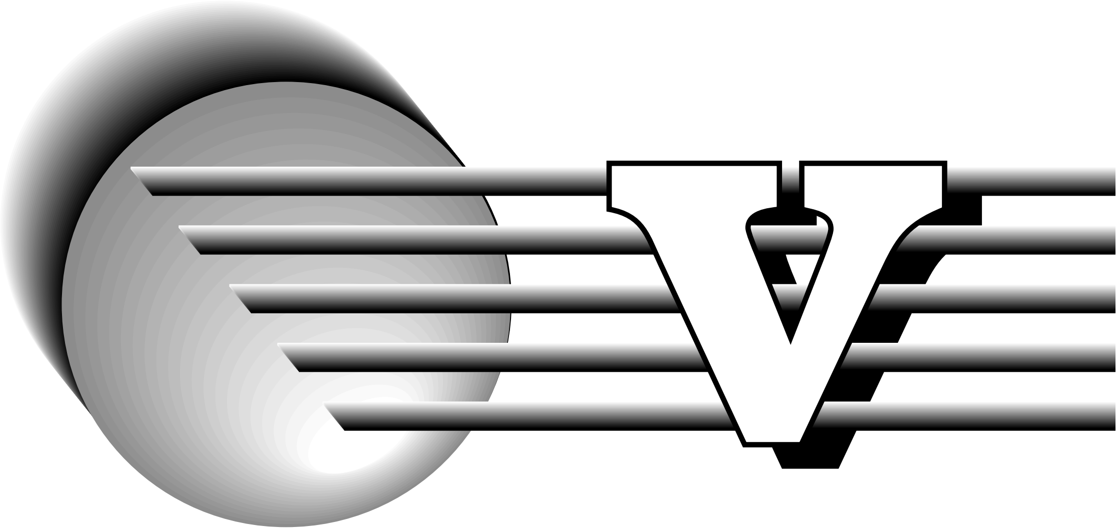

V **1.** Abbreviation of VOLT. **2.** Symbol (ital) for VOLTAGE or POTENTIAL. **3.** Symbol for VANADIUM. **4.** Symbol for VOLUME. **5.** Symbol (ital) for RELUCTIVITY. **6.** Abbreviation of VERTICAL. (Also, vert.)

v **1.** Abbreviation of VELOCITY. **2.** Abbreviation of VECTOR.

VA Abbreviation of VOLT-AMPERE.

V/A Abbreviation of *volts per ampere* (ohms).

VAC Abbreviation of *vector analog computer.*

vac Abbreviation of VACUUM.

V_{ac} Symbol for AC VOLTS. (Also, Vac.)

vacancy In a crystal lattice, a position not occupied by a nucleus.

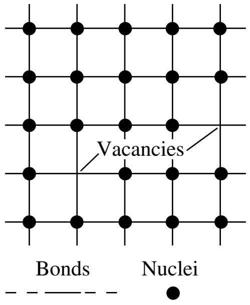

vacancy

vacuum A space from which all air and other gases have been removed to the greatest practicable extent. Some electronic parts, such as the elements of an electron tube, are housed in an evacuated space to protect them from the deterioration that would result from open-air operation.

vacuum capacitor A plate- or concentric-cylinder-type capacitor sealed in an evacuated glass tube or bulb. The vacuum acts as the dielectric and provides a dielectric constant of 1 and very high voltage breakdown.

vacuum deposition The electrical application of a layer of one material (such as a metal) to the surface of another (the substrate), carried out in a vacuum chamber (e.g., evaporation and sputtering). Also see DEPOSITION and EVAPORATION, **1**.

vacuum envelope The shell or tube of an electron tube, X-ray tube, or other electron device that requires a vacuum.

vacuum evaporation A method of manufacturing thin-film circuits by vaporizing a substance, and letting it accumulate or condense on a base.

vacuum gauge An instrument for measuring the vacuum in a device being evacuated. One of the several varieties of this gauge uses the elements of a triode, which are sealed in part of the vacuum system, some characteristic of the tube being continuously monitored as the vacuum pumping progresses. Another uses a thermistor sealed in part of the system; the resistance of the thermistor changes proportionately during the vacuum pumping.

vacuum impregnation The impregnation of a device (such as a capacitor, transformer, or choke coil) in a vacuum chamber. The process causes the pores in the device and its insulating materials to be completely filled with the impregnant.

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

vacuum level **1.** The pressure, in millimeters of mercury (mm/Hg). Normal atmospheric pressure is 760 mm/Hg. A perfect vacuum would be represented by 0 mm/Hg. **2.** The proportion, or percentage, of normal atmospheric pressure in a given environment.

vacuum phototube A phototube enclosed in an evacuated envelope, as opposed to one that is gas-filled.

vacuum pump A pump for removing air and gases from electron tubes, X-ray tubes, lamp bulbs, etc. Such pumps include mechanical types, diffusion types, and combinations of these. Also see DIFFUSION PUMP.

vacuum range The range of a communications system if propagation takes place through a perfect-vacuum (theoretical) medium.

vacuum seal An airtight seal between adjoining parts in an evacuation system.

vacuum switch A switch that is enclosed in a vacuum bulb or tube to reduce contact sparking.

vacuum thermocouple A thermocouple enclosed in a vacuum bulb with a small heater element. Radio-frequency current passed through the heater raises its temperature and causes the thermocouple to generate a proportional direct-current voltage.

vacuum tube An electron tube from which virtually all air and gases are removed. Also see ELECTRON TUBE.

vacuum-tube amplifier An amplifier using one or more vacuum tubes, rather than semiconductor devices.

vacuum-tube bridge A special bridge for determining vacuum-tube characteristics by null methods.

vacuum-tube characteristics The operating parameters of a vacuum tube, such as plate current, grid voltage, input resistance, interelectrode capacitances, amplification factor, transconductance, etc., that describe tube performance.

vacuum-tube coefficients See VACUUM-TUBE CHARACTERISTICS.

vacuum-tube current meter An ammeter, milliammeter, microammeter, or nanoammeter embodying an amplifier that uses one or more vacuum tubes. Also see ELECTRONIC CURRENT METER.

vacuum-tube modulator A circuit using one or more vacuum tubes to impress a modulating signal on a carrier.

vacuum-tube rectifier **1.** An alternating-to-direct-current converter circuit using one or more vacuum tubes, rather than gas tubes or semiconductor devices. **2.** A rectifier tube.

vacuum-tube sweep A sweep oscillator using a vacuum tube, rather than a gas tube. Also called HARD-TUBE SWEEP.

vacuum-tube voltmeter Abbreviation, vtvm. A voltmeter using a tube-type amplifier. Also see ELECTRONIC VOLTMETER.

val Abbreviation of VALUE.

valence A unit showing the degree to which elements or radicals (replaceable atoms or groups of atoms) will combine to form compounds.

valence band In the energy diagram for a semiconductor, the band of lowest energy. This band lies below the forbidden band (energy gap), which is below the conduction band. Also see ENERGY BAND DIAGRAM.

valence bond In a semiconductor material, an interatomic path over which shared electrons travel.

valence electrons Electrons in the outermost orbits of an atom. These electrons determine the chemical and physical properties of a material. Also see FREE ELECTRON.

valence shells See ELECTRON SHELLS.

validate To check data for correctness.

validity **1.** Correctness or accuracy of data. **2.** The logical truth of a derivation or statement, based on a given set of propositions.

valley A dip between adjacent peaks in a curve or wave.

valley current In a tunnel diode, the current at the valley point.

valley point The lowest point of finite current on the current-voltage response curve of a tunnel diode. Immediately before this point, current decreases with increasing applied voltage (an indication of negative resistance). Beyond this point, however, the current again increases with increasing voltage. Compare PEAK POINT.

valley voltage In a tunnel diode, the voltage at the valley point.

value **1.** The level or magnitude of a quantity (e.g., voltage value). **2.** The worth of a system, procedure, device, etc., in terms of goal fulfillment or other criterion.

valve Variation (Brit.) of ELECTRON TUBE. The term was applied to the first tube, a diode (the Fleming valve), and is descriptive of the action of a tube (a controller of electric current), rather than its appearance.

vanadium Symbol, V. A metallic element. Atomic number, 23. Atomic weight, 50.94.

Van Allen radiation belts Two high-altitude zones that surround the earth and consist of high-energy subatomic particles. These belts were once thought to preclude space travel by humans beyond the immediate vicinity of the earth.

Van de Graaff generator See BELT GENERATOR.

vane-anode magnetron A magnetron having plane parallel walls between adjacent cavities.

vane attenuator A waveguide attenuator consisting essentially of a slab (vane) of resistive material that slides laterally through the waveguide.

vane instrument See IRON-VANE METER.

vane-type magnetron See VANE-ANODE MAGNETRON.

vane-type meter See IRON-VANE METER.

V antenna See VEE ANTENNA.

vapor lamp A discharge lamp consisting essentially of a glass tube filled with a small amount of gas under low pressure and some element, such as mercury or sodium. A high voltage is applied between electrodes sealed into each end of the tube. The voltage causes the element to vaporize. This, in turn, ionizes the gas, causing it to glow.

vapor pressure In a confined medium, the pressure of a gas, measured in atmospheres, pounds per square inch, or millimeters of mercury.

VAR Abbreviation of VOLT-AMPERES REACTIVE.

var Abbreviation of VARIABLE.

varactor A semiconductor-type voltage-variable capacitor. Sometimes called a *varactor diode.*

varactor amplifier A dielectric amplifier using a varactor as the voltage-variable capacitor.

varactor flip-flop A bistable multivibrator based on the nonlinear performance of one or two varactors.

varactor frequency multiplier A frequency multiplier (doubler, tripler, etc.) in which multiples of a fundamental frequency result from the nonlinear action of a varactor.

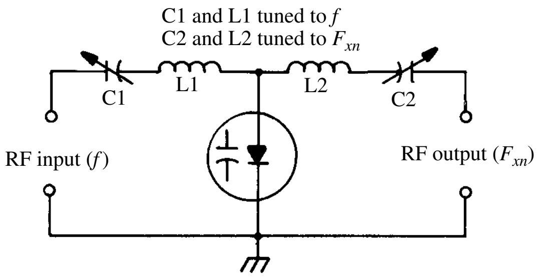

varactor frequency multiplier

varactor tuning A method of tuning a circuit or adjusting the frequency of an oscillator, using an inductor and a varactor to obtain the desired frequency. The varactor acts as a variable capacitor.

VAR-hour Abbreviation, VARh. Short for *volt-ampere-reactive-hour*, a unit of reactive energy; 1 VARh can be represented by 1 VAR for 1 hour, 2 VAR for 0.5 hour, 0.5 VAR for 2 hours, etc.

VAR-hour meter An instrument that measures reactive energy.

variable Abbreviation, var. A quantity whose value changes at some stated or calculable rate, and is given names in expressions or equations, such as *x*, *y*, or *z*. Compare CONSTANT. Also see DEPENDENT VARIABLE and INDEPENDENT VARIABLE.

variable-area sound track See VARIABLE-WIDTH SOUND RECORD.

variable block In computer operations, a unit of data, such as a group of records, whose size is dependent on data requirements (i.e., it is not fixed).

variable-capacitance diode See VARACTOR.

variable-capacitance transducer See CAPACITIVE TRANSDUCER.

variable capacitor A capacitor that can be adjusted from a low value (practically zero) to some maximum value. A step-type unit contains a number of fixed capacitors that can be switched in parallel with each other until, at the last step, all are in parallel. A continuously variable unit has a provision for moving one plate or set of plates, relative to another plate or set of plates; or one plate might be moved, with respect to another, so that the distance between them is changed. In a voltage-variable capacitor (such as a varactor), capacitance varies in accordance with an applied direct-current voltage.

variable-carrier modulation See QUIESCENT-CARRIER OPERATION.

variable connector In a flowchart, a path leading from a box in which a decision is made, to a number of other points on the diagram.

variable coupling Coupling that is adjustable. In an inductively coupled circuit, the distance or angle between the coils is usually the adjustable factor. In a capacitively coupled circuit, a variable capacitor is generally used.

variable-density sound record In photographic sound-on-film recording, a sound track made by varying, in sympathy with the sound frequency, the amount of light reaching the film. Compare VARIABLE-WIDTH SOUND RECORD.

variable-depth sonar A sonar device that can be placed far below the surface, for the purpose of detecting objects or terrain at greater depths than would be possible with a surface-located sonar device.

variable-efficiency modulation See EFFICIENCY MODULATION.

variable-erase recording In magnetic-tape operations, recording a signal by selectively erasing a previously recorded signal.

variable field **1.** Any field with an intensity that changes with time. An electromagnetic field is a common example. **2.** In a computer record, a field that is a variable block with a terminal symbol at the end.

variable-frequency oscillator Abbreviation, VFO. An oscillator (usually self-excited) whose frequency is continuously variable.

variable-inductance pickup A phonograph pickup in which vibration of the stylus causes the inductance of a small coil to vary in sympathy with the sound frequency.

variable-inductance transducer A transducer in which a monitored quantity causes the inductance of a coil to vary proportionately. The coil thereby offers a varying impedance to an alternating-current supply voltage.

variable inductor An inductor whose value can be adjusted from zero to a certain maximum. The variability might be in steps, provided by taps on the inductor, or a sliding contact can be used.

Several different inductors can be connected in series and parallel combinations, with switches to facilitate variability. One type of variable inductor uses a roller contact to allow continuous adjustment. See ROLLER INDUCTOR. Another method of varying inductance is to change the permeability of the core, such as by moving a rod of ferrite or powdered iron in and out of the coil. See VARICOUPLER and VARIOMETER.

variable-pitch indication An audible signal used in lieu of, or in conjunction with, a meter to indicate voltage, current, logic state, etc. The higher the value of the measured quantity, the higher the pitch.

variable-reluctance microphone A microphone in which the impinging sound waves cause corresponding variations in the reluctance of an internal magnetic circuit.

variable-reluctance pickup A phonograph pickup in which the stylus causes an armature to vibrate in a magnetic field, and consequently the reluctance of the magnetic circuit varies in sympathy with the audio frequency.

variable-reluctance transducer A transducer in which the monitored quantity causes the reluctance of an internal magnetic circuit to vary proportionately.

variable-resistance pickup A phonograph pickup in which the vibration of the stylus causes the resistance of an internal resistive element to vary in sympathy with the sound frequency.

variable-resistance transducer A transducer in which a monitored quantity causes the resistance of an internal resistive element to vary proportionately.

variable-resistance tuning Tuning of a selective resistance-capacitance (RC) circuit, such as a Wien bridge or parallel-tee network, by varying one or more of its resistance arms.

variable resistor A resistor whose value can be varied either continuously or in steps. Also see POTENTIOMETER and RHEOSTAT.

variable selectivity In a circuit or device, selectivity that is adjustable. Example: a variable-selectivity intermediate-frequency (IF) amplifier.

variable-speed motor Any motor whose speed is adjustable, smoothly or in steps, under load.

variable transformer A transformer (often an autotransformer) whose output voltage is adjustable from zero (or some minimum value) to maximum. For this purpose, one winding (usually the secondary in a two-winding transformer) can have a number of taps. Smooth variation can be provided by a wiper arm that slides over the turns of the winding.

variable-width wound record In photographic sound-on-film recording, a sound track made by varying the width of the track in sympathy with the sound variations. Compare VARIABLE-DENSITY SOUND RECORD.

Variac See VARIABLE TRANSFORMER.

Varicap A voltage-variable capacitor of the semiconductor-diode type. Also see VOLTAGE-VARIABLE CAPACITOR.

varicoupler An adjustable radio-frequency transformer consisting of a primary coil (usually the rotor) and a secondary coil (usually the stator), the former being rotatable to vary the coupling between the coils. The inductance of the stator is varied via a set of taps.

varindor A coil having a special core and whose inductance varies with the amount of direct current flowing through the winding. Also see SATURABLE REACTOR.

variocoupler A radio-frequency transformer in which one winding is rotatable, for the purpose of adjusting the mutual inductance between the primary and secondary. It is used in certain antenna-coupling applications.

variolosser A variable attenuator.

variometer A continuously variable inductor consisting of two coils connected in series or parallel and mounted concentrically—one rotating inside the other. Inductance is maximum when the coils are perpendicular to each other, and minimum when they are parallel.

varioplex A time-sharing method of transmitting and receiving wire telegraph signals. It allows the optimum usage of available lines.

varistor See VOLTAGE-DEPENDENT RESISTOR.

Varley-loop bridge A four-arm, direct-current bridge circuit in which one of the arms, a two-wire line, is accidentally grounded at a distant point. By adjustment of the bridge, the resulting resistance indicates the distance to the fault.

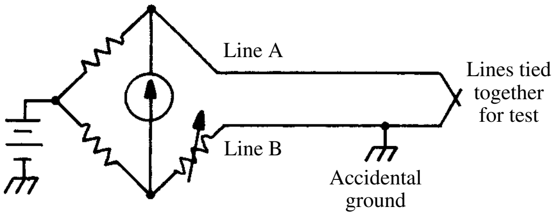

Varley loop bridge

VAR meter An instrument for measuring the apparent (VAR) power in a reactive circuit.

varnished cambric Cotton or linen fabric in sheets, tape or tubes, that has been varnish-impregnated and baked. A common form of this insulating material, the slender tubing known as *spaghetti*, has largely been replaced by plastic tubing.

varnished-cambric tubing Slender tubes of varnished cambric slipped over bare wires and busbars to insulate them. Also called *spaghetti*.

VATE Abbreviation of VERSATILE AUTOMATIC TEST EQUIPMENT.

V_B Symbol for BASE VOLTAGE of a bipolar transistor.

V_{BB} Symbol for *base-voltage supply* for a bipolar transistor.

V beam system See VEE BEAM SYSTEM.

V_C Symbol for COLLECTOR VOLTAGE of a bipolar transistor.

VCA Abbreviation of VOLTAGE-CONTROLLED AMPLIFIER.

V_{CC} Symbol for *collector-voltage supply* of a bipolar transistor.

VCCO Abbreviation of VOLTAGE-CONTROLLED CRYSTAL OSCILLATOR. (Also, lowercase.)

VCD Abbreviation of VARIABLE-CAPACITANCE DIODE.

VCG Abbreviation of VOLTAGE-CONTROLLED GENERATOR.

VCO Abbreviation of VOLTAGE-CONTROLLED OSCILLATOR.

V connection of transformers See VEE CONNECTION OF TRANSFORMERS.

VCR Abbreviation of VIDEO CASSETTE RECORDER.

VCSR Abbreviation of *voltage-controlled shift register.*

V-cut crystal See VEE-CUT CRYSTAL.

VCXO Abbreviation of VOLTAGE-CONTROLLED CRYSTAL OSCILLATOR. (Also, lowercase.)

VD **1.** Abbreviation of VOLTAGE DROP. **2.** Abbreviation of *vapor density.*

V_D Symbol for drain voltage of a field-effect transistor.

Vdc Symbol for DC VOLTS.

VDU Abbreviation of VISUAL DISPLAY UNIT.

vdcw Symbol for DC WORKING VOLTAGE. (Also, dcwv.)

VDIG Abbreviation of VIDEO DIGITIZER.

VDR **1.** Abbreviation of VOLTAGE-DEPENDENT RESISTOR. **2.** Abbreviation of VIDEODISC RECORDER.

V_{drive} Abbreviation for DRIVING VOLTAGE or OUTPUT VOLTAGE.

VE Abbreviation of value engineering.

V_E Symbol for EMITTER VOLTAGE of a bipolar transistor.

vectograph A graphic three-dimensional representation of a scene composed of superimposed images photographed through polarizing filters of different orientation, and reconstructed by a similar viewing technique.

vector **1.** A quantity having direction (in two or three dimensions) and magnitude. **2.** A graphical representation of a quantity, as defined in **1**. It consists of a straight arrow indicating the direction, and whose length is proportional to the magnitude.

vector addition The summation of two vectors, determined by adding their corresponding component values. In two dimensions, for example, suppose **A** = (x_1,y_1) and **B** = (x_2,y_2). Then **A** + **B** = [(x_1 + x_2),(y_1 + y_2)]. The sum can be found geometrically by constructing a parallelogram from the vectors, corresponding to the two known sides, as determined by the vectors. The sum is the vector originating at the point where the two vectors originate, and extending the diagonal of the parallelogram.

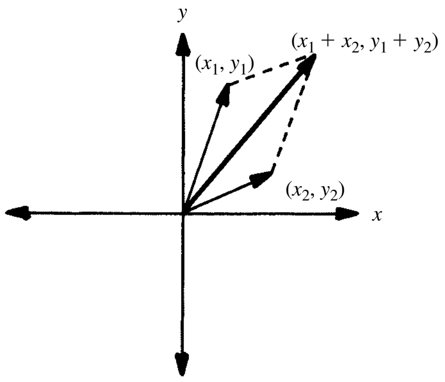

vector addition

vector admittance The reciprocal of VECTOR IMPEDANCE.

vector cardiograph An electrocardiograph that indicates the magnitude and the direction of a heart signal.

vector components **1.** The Cartesian coordinates of a vector, or its component angle and radius (polar coordinates). **2.** Quantities that can be represented by means of vectors (e.g., impedance and velocity).

vector diagram A graphical representation of vector quantities.

vector function A function having both magnitude and direction. Also see VECTOR; VECTOR COMPONENTS, **2**; and VECTOR QUANTITY.

vector generator A device that graphically illustrates vectors.

vectorial angle **1.** The angle between a vector and the horizontal axis. **2.** The angle between two vectors.

vector impedance Complex impedance (i.e., an impedance of the form $R + jX$, where R is resistance, X is reactance, and j is an operator indicating that the reactance is 90 degrees out of phase, relative to the resistance).

vector-mode display On a cathode-ray-tube display, the data representation in which points on the screen are connected by straight lines.

vector power The vector quantity $(P_a^2 + P_x^2)^{1/2}$, where P_a is the active power and Px is the reactive power.

vector power factor The ratio $P_a/(P_a^2 + P_x^2)^{1/2}$, where P_a is the active power and Px is the reactive power. In sine-wave situations, this ratio (of active power to vector power) produces a figure identical to the conventional power factor.

vector quantity A quantity having both magnitude and direction, and that can, therefore, be represented by a vector.

vectorscope A special oscilloscope for visual adjustment of a color-television receiver by means of a color-phase diagram.

vector sum The resultant of two nonparallel vectors; therefore, the resultant of the forces or quantities represented by them. For example, reactance and resistance can be represented by two perpendicular vectors.

vector voltmeter A voltmeter that indicates the phase as well as the amplitude of an alternating-current voltage.

V_{EE} Symbol for *emitter-voltage supply* of a bipolar transistor.

vee antenna A center-fed antenna in which the two halves of the radiator form an angle considerably less than 180 degrees.

vee-beam system An elevation-measuring radar system in which fan-shaped vertical and inclined beams, intersecting at ground and rotating continuously around the vertical axis, are radiated by one antenna. Target elevation is determined from the interval between successive echoes from the target.

vee connection of transformers In a three-phase system, a method of connecting two transformers so that the line current and voltage equal the phase current and voltage. Also called *open-delta connection.*

vee-cut crystal A piezoelectric plate cut from a quartz crystal so that its faces are not parallel to the *x*, *y*, or *z* axis of the crystal. Also see CRYSTAL AXES and CRYSTAL CUTS.

vee particle A short-lived elementary particle that results when high-energy neutrons or protons collide with nuclei. The particle can be positive, negative, or neutral, and gets its name from its cloud-chamber track.

vee signal In radiotelegraphy, the letter V (di-di-di-dah) transmitted as a test signal, usually three times in rapid succession. It is used during on-the-air transmitter tests, followed by the station's call letters.

vehicle An inert substance, usually a liquid, that acts as a solvent, carrier, or binder for some other, more active substance. Thus, shellac can be the vehicle for the metallic powder in a silver paint used in silk-screening electronic circuits.

vehicle smog-control device See COMPUTER-CONTROLLED CATALYTIC CONVERTER and EXHAUST ANALYZER.

vel Abbreviation of VELOCITY.

velocimeter **1.** An instrument for measuring the velocity of sound in various materials. **2.** An electronic velocity meter—especially a radial-velocity meter using Doppler radar. **3.** An electronic flow meter.

velocity Symbol, v. Abbreviation, vel. Unit, distance per unit time [e.g., meters per second (m/s)]. **1.** The change in the position of a body over a specified period of time. **2.** The derivative of displacement with respect to time. Compare ACCELERATION.

velocity constant See VELOCITY FACTOR.

velocity error For a servomechanism in which the input and output shafts rotate at the same speed, the angular displacement between them.

velocity factor Abbreviation, v. For a transmission line, the ratio of the speed of electromagnetic wave propagation in the line to the speed of electromagnetic waves in a vacuum (299,792 kilometers per second). It can be expressed as a number between 0 and 1, or as a percentage between 0 and 100. In practical feed lines, v ranges from about 0.66 (for coaxial line with a solid polyethylene dielectric) to about 0.95 (for open-wire line with widely separated spacers).

velocity hydrophone A hydrophone whose output, like that of the velocity microphone, is proportional to the instantaneous particle velocity in the sound wave impinging on the device.

velocity-lag error A lag (proportional to the input-variation rate) between the input and output of a device, such as a servomechanism.

velocity level For a sound, the ratio v_0, expressed in decibels:

$$v_0 = 20 \log_{10}(v_s/v_r)$$

where v_s is the particle velocity of the sound, and v_r is a reference particle velocity.

velocity microphone A microphone in which the vibratory element is a thin, aluminum or Duralumin ribbon suspended loosely between the poles of a strong permanent magnet. Vibration of the corrugated ribbon in the magnetic field causes an audio-frequency voltage to be induced across the ribbon. The microphone is so called because its output is proportional to the instantaneous particle velocity in the sound wave impinging on the ribbon. Also called *ribbon microphone.*

velocity-modulated amplifier A circuit in which radio-frequency amplification is obtained by velocity modulation.

velocity-modulated oscillator A vacuum-tube device in which an electron stream is velocity-modulated (see VELOCITY MODULATION) as it passes through a resonant cavity (the buncher); the subsequent energy, of a higher intensity, is extracted from the bunched stream as it passes through another resonant cavity (the catcher). Feedback from catcher to buncher sustains oscillations. See, for example, KLYSTRON OSCILLATOR.

velocity-modulated tube A vacuum tube utilizing velocity modulation. See, for example, KLYSTRON.

velocity modulation The process in which the input signal of a vacuum tube varies the velocity of the electrons in a constant-current electron beam in sympathy with the input signal.

velocity resonance A state of resonance in which a 90-degree phase difference is between the fundamental frequency of oscillation of a system and the fundamental frequency of the applied signal.

velocity sorting Selecting and separating electrons, according to their velocity.

velocity spectrograph A device that uses electric or magnetic deflection to separate charged particles into various streams, according to their velocity.

velocity variation See VELOCITY MODULATION.

velocity-variation amplifier See VELOCITY-MODULATED AMPLIFIER.

velocity-variation oscillator See VELOCITY-MODULATED OSCILLATOR.

velometer An instrument used to measure the velocity of air (e.g., to determine wind speed).

venetian-blind effect A television-display malfunction in which the picture seems to appear through horizontal slits.

Venn diagram A method of illustrating the relationship among various subsets within a specified universal set. The subsets are shown as geometric figures, usually circles.

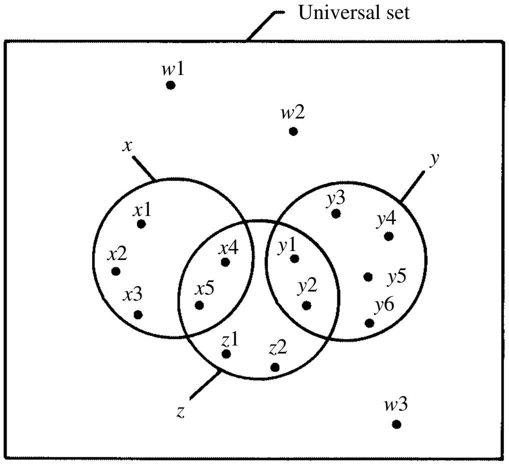

Venn diagram

vent **1.** An opening in an equipment enclosure for the entry of cool air or the escape of warm air. **2.** In an electrolytic capacitor, a lightly covered blowout hole for the relief of gas pressure in the event of severe overload. **3.** A small opening for relieving gas pressure in an automobile battery. **4.** In a loudspeaker enclosure, such as the bass-reflex type, an auxiliary opening that extends the low-frequency range of the speaker.

vented baffle A loudspeaker enclosure having the proper auxiliary opening(s) for coupling the speaker to the surrounding air.

vented-baffle loudspeaker See ACOUSTICAL PHASE INVERTER and BASS-REFLEX LOUDSPEAKER.

vent mount A metal bracket for fastening an antenna mast to a (plumbing) vent pipe on a roof.

ventricular fibrillation See FIBRILLATION.

verification **1.** The process of ensuring that two sets of data are identical. **2.** The process of validating (see VALIDATE).

vernier **1.** An auxiliary scale along which a regular, linear scale (such as that of a tuning dial) slides. The vernier scale is graduated so that when the main scale is set to an unmarked point between two of its graduations, and the "0" point on the vernier scale is used as the index, "1" on the vernier scale will exactly coincide with "1" on the main scale. The corresponding number on the vernier scale indicates the exact number of subdivisions between two points on the main scale. Vernier arrangements are provided with the dials of some precision analog instruments. **2.** See VERNIER CAPACITOR. **3.** See VERNIER DIAL. **4.** See VERNIER RESISTOR.

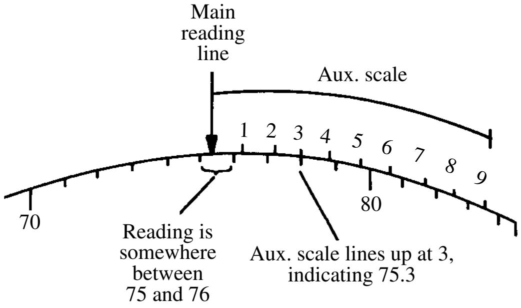
Main reading line
Aux. scale
1
2
3
4
5
6
7
8
9
70
80
Reading is somewhere between 75 and 76
Aux. scale lines up at 3, indicating 75.3

vernier, 1

vernier capacitor A low-capacitance variable capacitor connected in parallel with a higher-capacitance fixed or variable capacitor for precise adjustment of the total capacitance.

vernier dial **1.** A slow moving dial for fine tuning an adjustable device. The required reduction ratio is obtained with gears, friction wheels, or a belt-and-pulley combination. **2.** A dial provided with an accessory vernier scale.

vernier resistor A low-resistance variable resistor connected in series with a higher-resistance fixed or variable resistor to precisely adjust the total resistance.

vernier rheostat See VERNIER RESISTOR.

vers Abbreviation of VERSED SINE.

versatile automatic test equipment Abbreviation, VATE. For troubleshooting the electronic systems of missiles, a computer-controlled tester that isolates faults through logical operations.

versed sine Abbreviation, vers. A trigonometric function equal to the difference between the cosine of an angle and one; vers $x = 1 - \cos x$.

vert Abbreviation of VERTICAL.

vertex **1.** The terminal point at which two or more branches of a network meet. Also see NODE, **1**. **2.** The point of intersection of two lines that form an angle.

vertex plate In the reflector of a microwave antenna, a matching plate mounted at the vertex.

vertical **1.** The direction indicated by a line perpendicular to the plane of the horizon (e.g., a vertical axis). **2.** In the 45-45 recording process, the signal resulting from sound that arrives simultaneously and 180 degrees out of phase at the two microphones and causes up-and-down movement of the cutting stylus.

vertical amplification Gain provided by the vertical channel of a device, such as an oscilloscope, cathode-ray electrocardiograph, or television receiver. Compare HORIZONTAL AMPLIFICATION.

vertical amplifier The circuit or device that provides gain in the vertical channel of a device. Compare HORIZONTAL AMPLIFIER.

vertical-amplitude control **1.** See VERTICAL-GAIN CONTROL. **2.** In a color television receiver, one of the three controls by which the amplitude of the parabolic voltages applied to the coils of the magnetic-convergence assembly are adjusted.

vertical angle of radiation The angle, with respect to the plane of the horizon, at which radio-frequency energy is propagated from a transmitting antenna.

vertical antenna **1.** An antenna that consists essentially of a single, straight, vertical conductor. **2.** An antenna that is mounted vertically, instead of horizontally.

vertical blanking See VERTICAL RETRACE BLANKING.

vertical-blanking pulse In a television signal, the pulse at the end of each field that accomplishes vertical retrace blanking. Compare HORIZONTAL-BLANKING PULSE.

vertical-centering control See CENTERING CONTROL.

vertical channel The system of amplifiers, controls, and terminations that constitutes the path of the vertical signal applied to a device, such as an oscilloscope. Compare HORIZONTAL CHANNEL.

vertical compliance In disc sound reproduction, the ease with which the stylus can move vertically while it is in position on the disc.

vertical coordinates See CARTESIAN COORDINATES.

vertical deflection **1.** In an oscilloscope or television receiver, the movement of the spot up and down on the screen. Compare HORIZONTAL DEFLECTION. **2.** In a direct-writing recorder, up or down deflection of the pen.

vertical deflection coils In a cathode-ray tube, the pair of coils in a deflection yoke that provides the electromagnetic field for the vertical deflection of the electron beam. Compare HORIZONTAL DEFLECTION COILS.

vertical deflection electrodes See VERTICAL DEFLECTION COILS and VERTICAL DEFLECTION PLATES.

vertical deflection plates In an oscilloscope (and in some early television picture tubes), a pair of plates that provides the electrostatic field for vertical deflection. Compare HORIZONTAL DEFLECTION PLATES.

vertical dipole An antenna consisting of a straight, center-fed, half-wave conductor oriented vertically.

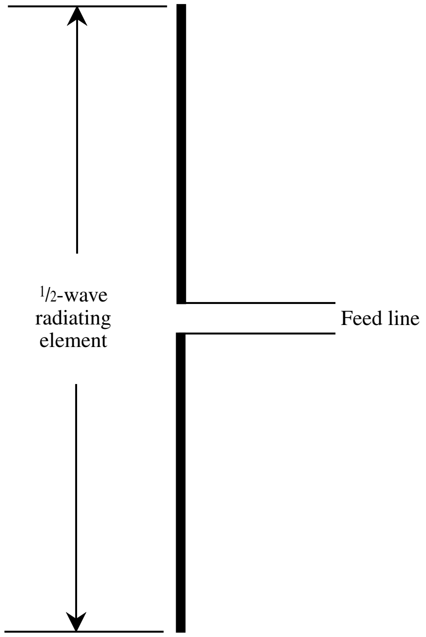

vertical dipole

vertical dynamic convergence During the scanning of points along a vertical line through the center of a color-television picture tube, the convergence of the electron beams at the aperture mask. Compare HORIZONTAL DYNAMIC CONVERGENCE.

vertical field strength The field strength of signals in a vertical plane passing through an antenna. Compare HORIZONTAL FIELD STRENGTH.

vertical-field-strength diagram A plot of vertical field strength.

vertical frequency response The gain-frequency characteristic of the vertical channel of an instrument, such as an oscilloscope. Compare HORIZONTAL FREQUENCY RESPONSE.

vertical gain The overall amplification (gain) provided by the vertical channel of an instrument, such as an oscilloscope, graphical recorder, or television receiver. Compare HORIZONTAL GAIN.

vertical-gain control A control, such as a potentiometer, for adjusting vertical amplification. Compare HORIZONTAL-GAIN CONTROL.

vertical-hold control In a television receiver, the control for adjusting the frequency of the vertical oscillator so that the picture can be locked to prevent vertical roll. Compare HORIZONTAL-HOLD CONTROL.

vertical-incidence transmission Transmission of a wave vertically to the ionosphere and the subsequent reflection of the wave to the earth.

vertical-lateral recording In stereophonic disc recording, the recording of one signal vertically (see VERTICAL RECORDING) and the other laterally (see LATERAL RECORDING).

vertical linearity Linearity of response (gain and deflection) of the vertical channel of an oscilloscope or television receiver. A linear picture is neither contracted nor expanded vertically in any part of the screen. Compare HORIZONTAL LINEARITY.

vertical-linearity control In an oscilloscope or television receiver, the control whereby vertical linearity is adjusted. Compare HORIZONTAL-LINEARITY CONTROL.

vertically polarized wave An electromagnetic wave whose electric lines of flux are perpendicular to the plane of the horizon. Compare HORIZONTALLY POLARIZED WAVE.

vertical-metal-oxide semiconductor field-effect transistor Abbreviation, VMOSFET. A metal-oxide semiconductor field-effect transistor fabricated so that the current flow within the device is vertical, instead of the usual horizontal, affording several advantages over the conventional MOSFET.

vertical oscillator In a television receiver, the oscillator that generates the vertical sweep signal. Compare HORIZONTAL OSCILLATOR.

vertical-output regulator In a television receiver, a voltage-dependent resistor used to stabilize the sweep voltage across the horizontal deflection coils—especially during warmup.

vertical output stage In a television receiver, an output amplifier following the vertical oscillator. Compare HORIZONTAL OUTPUT STAGE.

vertical polarization Pertaining to an electromagnetic wave whose electric lines of flux are perpendicular to the plane of the horizon. In general, when the radiating element of an antenna is vertical, the electric lines of flux in the transmitted waves are vertical, and the antenna is most sensitive to incoming signals whose electric lines of flux are vertical. Compare HORIZONTAL POLARIZATION.

vertical positioning control See CENTERING CONTROL.

vertical quantity The quantity measured along the *y*-axis of a graph, or represented by the vertical deflection of an electron beam. Compare HORIZONTAL QUANTITY.

vertical radiator See VERTICAL ANTENNA.

vertical recording Disc recording in which the depth of the groove is varied in sympathy with the sound. Also called HILL-AND-DALE RECORDING. Compare LATERAL RECORDING.

vertical redundance In computer operation, the error state when a character has an odd number of bits in an even-parity system, or vice versa.

vertical resolution The number of horizontal wedge-lines that can be easily seen in a television test pattern before they blend. Compare HORIZONTAL RESOLUTION.

vertical retrace In a cathode-ray device, such as an oscilloscope or television receiver, the rapid return of the beam to its starting point at the top of the screen after completely traversing the screen from top to bottom. Compare HORIZONTAL RETRACE.

vertical retrace blanking In a television receiver, the automatic shutoff of the electron beam during a vertical retrace period, to prevent an extraneous line being traced on the screen during this period. Also see BLACKOUT, **2**; BLANKING INTERVAL, **1**, **2**; BLANKING PEDESTAL; BLANKING TIME; VERTICAL RETRACE; and VERTICAL RETRACE PERIOD. Compare HORIZONTAL RETRACE BLANKING.

vertical retrace period In a television receiver, the interval during which the spot returns from the bottom to the top of the screen after having traced all horizontal lines from top to bottom. Compare HORIZONTAL RETRACE PERIOD.

vertical sensitivity The signal voltage required at the input of a vertical channel to produce full vertical deflection. Also see VERTICAL GAIN. Compare HORIZONTAL SENSITIVITY.

vertical signal A signal serving as a vertical quantity (e.g., one that deflects the beam in a cathode-ray tube upward and/or downward). Compare HORIZONTAL SIGNAL.

vertical-speed transducer A transducer whose electrical output is proportional to the vertical speed of an aircraft or missile carrying the transducer.

vertical stylus force In disc sound reproduction, the downward force (in grams or ounces) that the stylus exerts on the disc.

vertical sweep **1.** In a cathode-ray tube, especially a television picture tube, the movement of the spot up or down on the screen. Compare HORIZONTAL SWEEP. **2.** The circuit that produces this sweep.

vertical sweep frequency The frequency at which vertical sweep occurs. In a television receiver, it is 60 Hz. Also called *vertical sweep rate.* Compare HORIZONTAL SWEEP FREQUENCY.

vertical sweep rate See VERTICAL SWEEP FREQUENCY.

vertical synchronization In a television receiver, synchronization of the vertical component of scanning with that of the camera. Also see VERTICAL SYNC PULSE. Compare HORIZONTAL SYNCHRONIZATION.

vertical sync pulse In a video signal, the pulse that synchronizes the vertical component of scanning in a television receiver with that of the camera, and that triggers vertical retrace and blanking. Compare HORIZONTAL SYNC PULSE.

vertical wave See VERTICALLY POLARIZED WAVE.

vertical width control See WIDTH CONTROL, **1**, **2**.

very high frequency Abbreviation, VHF. Radio frequencies in the range 30 MHz to 300 MHz, corresponding to free-space wavelengths of 10 meters to 1 meter. It is sometimes divided into *VHF low band* (30 MHz to 50 MHz) and *VHF high band* (50 MHz to 300 MHz). Also see RADIO SPECTRUM, **1**.

very high resistance Abbreviation, VHR. Large values of resistance; usually expressed in megohms, gigohms, or teraohms.

very-high-resistance voltmeter Abbreviation, VHRVM. A voltmeter using a low-range microammeter or picoammeter and a very high value of multiplier resistance (see VOLTMETER MULTIPLIER).

very-high-speed integrated circuit An integrated circuit used for switching or other digital functions at thousands, millions, or billions of changes of state per second.

very-large-scale integration Abbreviation, VLSI. The inclusion of several complete systems, such as computers, on a single integrated-circuit chip. This can extend several orders of magnitude beyond large-scale integration (LSI).

very long range Abbreviation, VLR. Pertaining to ground radar sets having a maximum slant range of over 250 miles. Compare VERY SHORT RANGE.

very low frequency Abbreviation VLF. A radio frequency in the range 10 kHz to 30 kHz, corresponding to wavelengths between 30 kilometers and 10 kilometers. Also see RADIO SPECTRUM, **1**.

very low resistance Abbreviation, VLR. Values of resistance less than 1 ohm, usually expressed in milliohms or microhms.

very short range Abbreviation, VSR. Pertaining to ground radar sets having a maximum slant range of less than 25 miles. Compare VERY LONG RANGE.

vestigial **1.** An effect that remains as a by-product, but that serves no directly applicable purpose. See, for example, VESTIGIAL SIDEBAND. **2.** Unnecessary; extraneous.

vestigial sideband **1.** A portion of one sideband in an amplitude-modulated signal, remaining after passage through a selective filter. **2.** An amplitude-modulated signal in which one sideband has been partially or largely suppressed. **3.** The small amount of energy emitted in the unused sideband in a single-sideband transmitter.

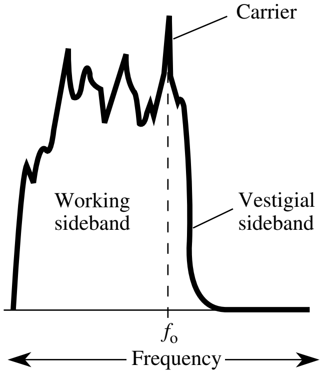

vestigial sideband, 2

vestigial-sideband filter A filter operated between an amplitude-modulated transmitter and an antenna to obtain a vestigial-sideband signal.

vestigial-sideband signal An amplitude-modulated signal in which one of the sidebands has been partially suppressed.

vestigial-sideband transmission Transmission of a signal containing a vestigial sideband. In television, for example, the upper sideband is transmitted fully, while the lower sideband is almost completely suppressed. The lower sideband is, therefore, a vestigial sideband.

vestigial-sideband transmitter An amplitude-modulated transmitter equipped with the filters or other subcircuits necessary for emitting a vestigial-sideband signal.

VF Abbreviation of VIDEO FREQUENCY.

V_{FB} Abbreviation for *feedback voltage* in an integrated circuit device. The term applies especially to operational amplifiers.

VFO Abbreviation of VARIABLE-FREQUENCY OSCILLATOR.

V_g Abbreviation of GENERATOR VOLTAGE. (Also, *Eg*.)

VGA Abbreviation of VARIABLE-GAIN AMPLIFIER.

V_{GD} Symbol for GATE-DRAIN VOLTAGE of a field-effect transistor.

V_{GS} Symbol for GATE-SOURCE VOLTAGE of a field-effect transistor.

VHF Abbreviation of VERY HIGH FREQUENCY.

VHF high band See VERY HIGH FREQUENCY.

VHF low band See VERY HIGH FREQUENCY.

VHF omnirange Abbreviation, VOR. A very-high-frequency radionavigation system that provides radial lines of position.

VHF oscillator A radio-frequency oscillator specially designed to operate in the range 30 MHz to 300 MHz.

VHF receiver **1.** A radio receiver specially designed to operate in the range of 30 MHz to 300 MHz. **2.** A television receiver for the VHF channels (see VHF TELEVISION CHANNELS).

VHF television channels Twelve 6-MHz-wide channels (television channels 2 through 13) between 54 and 216 MHz.

Channel	*Frequency (MHz)*
2	54.0 – 60.0
3	60.0 – 66.0
4	66.0 – 72.0
5	76.0 – 82.0
6	82.0 – 88.0
7	174 – 180
8	180 – 186
9	186 – 192
10	192 – 198
11	198 – 204
12	204 – 210
13	210 – 216

VHF television channels (United States)

VHF transmitter A radio or television transmitter specially designed to operate in the range 30 MHz to 300 MHz.

VHR Abbreviation of VERY HIGH RESISTANCE.

VHRVM Abbreviation of VERY-HIGH-RESISTANCE VOLTMETER.

VHS videocassette recorder A popular scheme for videocassette recording, used in millions of households in the United States alone. Typical VHS cassette play times range from about two hours to more than eight hours.

VHSIC Abbreviation for VERY HIGH SPEED INTEGRATED CIRCUIT.

VI **1.** Abbreviation of VOLUME INDICATOR. **2.** Abbreviation of VISCOSITY INDEX.

V_i Symbol for INPUT VOLTAGE. (Also, V_{in}.)

vibrating bell A bell with a striking mechanism that oscillates, causing a continuous ringing sound.

vibrating-reed frequency meter See POWER-FREQUENCY METER, **2**.

vibrating-reed oscillator An audio-frequency oscillator in which an iron, steel, or alloy reed vibrating in a magnetic field acts like a tuning fork to control the oscillation frequency.

vibrating-reed relay An electromechanical relay in which the movable contact is carried by the end of a thin, metal strip (reed) of iron or magnetic alloy. The strip is supported within the magnetic field of a coil carrying an alternating control current; when the frequency of this current corresponds to the natural (resonant) frequency of the reed, it vibrates vigorously enough to close the contacts. Such a relay, consequently, is frequency selective.

vibrating-wire oscillator An oscillator similar to the vibrating-reed oscillator, but using a wire, instead of a reed.

vibrating-wire transducer A transducer in which the fluctuating tension of a thin wire suspended in a magnetic field frequency-modulates the operating voltage.

vibration **1.** The changing of the position or dimensions of a body back and forth, usually at a rapid rate, an action seen in the repetitive movement of a musical string, headphone diaphragm, loudspeaker cone, loose machine, etc. **2.** See OSCILLATION.

vibration analyzer See VIBRATION METER.

vibration calibrator A device that generates a standard vibration, usually at a fixed frequency, for calibrating vibration meters, pickups, transducers, etc.

vibration galvanometer A type of alternating-current galvanometer. The natural frequency of the movable element of the instrument is made equal to that of the alternating current under test, to obtain a reading.

vibration isolator In an electronic-equipment assembly, a cushioning support that protects the equipment from vibration.

vibration machine See VIBRATOR, **2**.

vibration meter An instrument for measuring the amplitude and frequency of vibration (see VIBRATION, **1**). It consists essentially of a vibration pickup followed by a selective amplifier and an indicating voltmeter graduated in vibration units.

vibration pickup A transducer that senses mechanical vibration and converts it into a proportionate output voltage, or changes resistance in sympathy with the vibration.

vibrato In electronic musical instruments, a circuit or device that modulates a note by varying its frequency, amplitude, or both, at an extremely low frequency (a few hertz).

vibrator **1.** See INTERRUPTER. **2.** See VIBRATING-REED RELAY. **3.** A device for shaking something under test at selected frequencies and amplitudes.

vibrator-type power supply A battery-operated, high-voltage power supply in which one vibrator (see INTERRUPTER) makes and breaks direct current flowing from the battery into the primary winding of a step-up transformer, and another vibrator rectifies the high voltage delivered by the secondary winding.

vibrator-type rectifier A vibrator (see INTERRUPTER) that connects one terminal of the secondary winding of a transformer to an output terminal each time that terminal is positive. When the negative half-cycle appears at the transformer terminal, the vibrator is open. In this way, the alternating-current output of the transformer is rectified.

vibrocardiography A method of monitoring and recording the movement of the chest cavity as a result of the beating of the heart.

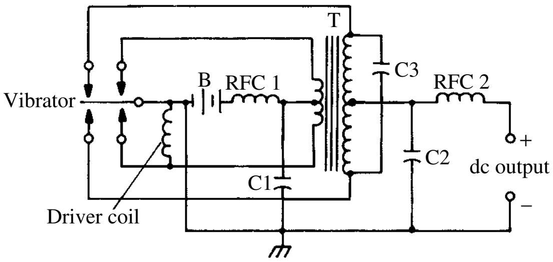

vibrator-type power supply

vibrograph A device for observing and recording vibration.

vibrometer See VIBRATION METER.

vibroscope An instrument for observing (and sometimes recording) vibration.

Victor Phonetic alphabet communications code word for the letter V.

Victron A brand of polystyrene, a low-loss plastic insulant that is especially useful at high radio frequencies.

video **1.** Pertaining to television—especially to its picture circuitry or to circuits and devices related to television, but used for other purposes. **2.** The picture portion of a television broadcast, as opposed to the sound (audio) portion. **3.** Sometimes, a cathode-ray-tube terminal or display. **4.** Generally, television in adjective sense (e.g., *video play*). **5.** The images on a computer display or monitor.

video amplifier **1.** In television, the wideband stage (or stages) that amplifies the picture signal and presents it to the picture tube. **2.** A similar wideband amplifier, such as an instrument amplifier or preamplifier having at least a 4-MHz bandwidth.

video animation Movement of graphic images on a cathode-ray-tube display by the use of computer-manipulation through special software, under the direction of an operator using a keyboard, joystick, or digitizer.

video buffer A unity-gain circuit that minimizes loading effects caused by having several video multiplexer inputs connected to the same signal source.

video carrier The television carrier amplitude-modulated by the picture information, synchronization pulses, and blanking pulses.

videocassette recorder Abbreviation, VCR. A machine for recording television programs on magnetic-tape cassettes (see CASSETTE, **1**).

videocast **1.** Television broadcast. Also called TELECAST. **2.** To make a television broadcast.

videocaster See TELECASTER.

video CD-ROM A digital optical VIDEODISC, resembling an audio or computer compact disc. A typical disc can store more than an hour's worth of conventional television programming.

video correlator A device used with radar to locate and identify a target with high precision.

video detector In a television receiver, the demodulator for the video signal. This detector follows the video intermediate-frequency amplifier and precedes the video amplifier. It usually embodies a relatively simple, wideband diode circuit.

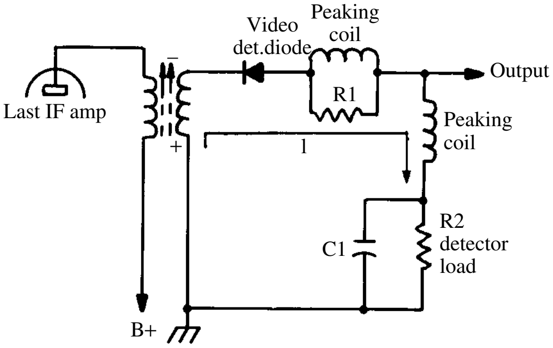

video detector

Video Digitizer Abbreviation, VDIG. A program developed for multimedia video in personal computer systems.

videodisc A flat magnetic or optical disc on which audio-visual programs are recorded, and from which they can be played back into a television receiver or multimedia personal computer.

videodisc recorder Abbreviation, VDR. A machine for recording television programs on discs comparable to those used for sound or data recording.

video discrimination In radar, the use of a circuit or device to decrease the width of the video-amplifier passband.

video frequency **1.** Pertaining to signals in the wide passband of a video amplifier: 30 Hz to 4 MHz. **2.** Pertaining to a device capable of operating over the wide passband of a video amplifier (e.g., a video-frequency alternating-current voltmeter).

video-frequency amplifier An amplifier capable of handling signals of a wide frequency range (e.g., direct current to 4 MHz). Also see VIDEO FREQUENCY, **1**.

video gain control In a television receiver, a gain control for adjusting the video signal amplitude. In a color-television receiver, a pair of these controls permits adjustment of the three color signals to the proper amplitude ratio.

video game A game, such as football, basketball, slot machine, tic-tac-toe, etc., played on the screen of a television receiver or on a computer.

videogenic Suitable for television.

video IF amplifier In a television receiver, the broadband intermediate-frequency (IF) amplifier of the video signal. In modern receivers, this IF amplifier also handles the sound signal.

video mapping A system of surveillance or mapping in which, for reference, the outlines of the area being surveyed are superimposed electronically on the radar display of the area.

video masking A form of radar signal processing in which ground clutter and other unwanted echoes are removed, making the desired targets more readily visible.

video mixer A circuit or device for mixing the signals from two or more television cameras.

video modulation In television transmission, amplitude modulation of the carrier wave with pulses and waves corresponding to the picture elements.

video on demand Television service in which subscribers can choose the programs they want to watch, and also the specific times for viewing.

video random-access memory Abbreviation, VRAM. In computer systems, RANDOM-ACCESS MEMORY used primarily for enhancing the performance of the display.

video recording **1.** Recording wideband material (such as a video signal) on a tape or disc. **2.** The recording made of a telecast. Also see VIDEOTAPE RECORDER.

video signal **1.** In television, the amplitude-modulated signal containing picture information and pulses. Also see VIDEO, **2** and VIDEO MODULATION. **2.** Broadly, a telecast signal, including sound.

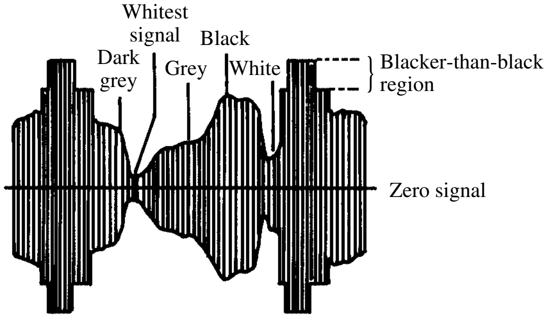

video signal

video stretching In some electronic navigation systems, increasing the duration of a video pulse.

video synthesizer A computerized device that produces graphical renditions of objects or circuits in three dimensions.

videotape **1.** A special magnetic tape for video recording (see VIDEO RECORDING, **2**). **2.** To make a video recording.

videotape recorder Abbreviation, VTR. A wideband, magnetic-tape recorder for producing video recordings with a camera, or for making a record of television programs, for subsequent reproduction (playback).

videotape recording **1.** Abbreviation, VTR. The technique of recording video signals on magnetic tape. Also see VIDEO RECORDING, **1**, **2** and VIDEOTAPE. **2.** A tape on which a video signal has been recorded.

videotext A system that allows television viewers to dial up special material, such as stock market quotations, weather data, sports scores, etc.

vidicon A television camera tube in which the electron beam scans a charged-density pattern that has been formed and stored on the surface of the photoconductor. It is commonly used in camcorders, closed-circuit television systems, and robot vision systems. Its assets include compactness and sensitivity.

viewfinder An accessory or integral device in which an image (formed optically or electronically) corresponds to the image viewed by the camera with which it is used.

viewing mesh In a cathode-ray storage tube, the mesh on which the image is presented for viewing by the operator. Also see CHARGE-STORAGE TUBE.

viewing mirror In an oscilloscope-camera assembly, a flat, slant-mounted mirror that reflects the image on the oscilloscope screen to the viewer's eye.

viewing screen In a cathode-ray device, such as an oscilloscope tube or television picture tube, the face on which the image appears.

viewing time In a storage type cathode-ray tube, the length of time for which the image persists.

viewing window See WINDOW, **2**.

Villard circuit **1.** A voltage-doubler circuit using only one diode and two capacitors. The open-circuit direct-current output voltage is approximately twice the peak value of the alternating-current input voltage. Also see VOLTAGE DOUBLER. **2.** See SELECTOJECT.

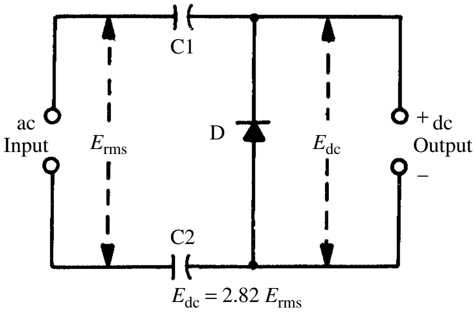

Villard circuit, 1

Villari effect In certain magnetostrictive materials, the change in magnetic induction that accompanies the application of a mechanical stress in a given direction. Also see MAGNETOSTRICTION.

vinyl A general name for *vinyl copolymer resins.*

vinylidene chloride A plastic insulant. Dielectric constant, 3 to 5. Dielectric strength, 20 kV/mm.

Vinylite-A A brand of the plastic *polyvinyl acetate.*

Vinylite-Q A brand of the plastic *polyvinyl chloride.*

vinyl resin A synthetic resin resulting from the polymerization of compounds that contain the group $CH_2 = CH–$.

violation In a security system, a condition that results in the actuation of an alarm. Example: an intruder breaks the light beam in an electric eye.

virgin magnetic material Core or shield material that has never been magnetized. When virgin material is first subjected to magnetization, the hysteresis loop starts at zero, but never returns to zero (see BOX-SHAPED LOOP).

virgin neutron Any neutron that has not been involved in a collision with another particle subsequent to its initial generation.

virgin record See BLANK RECORD.

virgin tape See BLANK TAPE.

virtual address In computer operations, an address that must be modified to refer to a location in main memory.

virtual decision value In quantizing or encoding, a method of expressing the maximum input amplitude. Two values are extrapolated from actual decision values—one at either end of the working range used.

virtual height The altitude that a vertically propagated electromagnetic wave would reach before reflection if its path in the ionosphere were a straight line. The actual distance at which the wave penetrates the ionosphere before reflection is less than the virtual height.

virtual image The image formed when rays from a scene diverge after passing beyond the focal point of a convex lens. The scene appears inverted to an observer. Compare REAL IMAGE.

virtual memory **1.** In a computer system, a means of using two or more memory stores simultaneously. **2.** Auxiliary memory used in conjunction with the main, or core, memory.

virtual ppi reflectoscope A device used to superimpose a virtual image of a chart or map onto a plan-position indicator (ppi) radar display. Also see VIDEO MAPPING.

virtual reality Abbreviation, VR. A general term for any of various high-level computer simulation or remote-control programs. The user often wears a head-mounted display that provides vivid, three-dimensional imagery and binaural sound. It can be used in robotic telepresence systems for precision remote control.

virus See COMPUTER VIRUS.

vis **1.** Abbreviation of *visibility.* **2.** Abbreviation of *visual.*

viscometer An instrument for measuring viscosity. There are several electronic versions of this device. In one, a steel ball falls through a material (such as an oil) under test and distorts the magnetic field of a pickup coil, causing the deflection of a meter by an amount proportional to the speed of the ball and, consequently, to the viscosity of the fluid.

viscosimeter See VISCOMETER.

viscosity The resistance offered by a fluid (liquid or gas) to objects passing through it. The viscosity of pure water is low; that of heavy oil is high. Expressed in newton-seconds per meter squared.

viscosity index Abbreviation, VI. A number indicating how well an oil retains its viscosity with temperature changes; larger indexes are assigned to oils that are little influenced by variations in temperature.

viscous-damped arm A phonograph pickup arm with an oil dashpot to prevent arm resonance and to slow the descent of the arm to the disc.

viscous damping The use of a viscous fluid in the dashpot of a device (such as a relay, timer, or pickup arm) to provide damping. See, for example, DASHPOT RELAY.

viscous hysteresis A slow, slight increase in the magnetization of a material when the magnetizing field is constant. Compare STATIC HYSTERESIS.

visibility factor See DISPLAY LOSS.

visible radiation Electromagnetic radiation that is perceptible to the eye. Also see VISIBLE SPECTRUM.

visible spectrum The band of electromagnetic wavelengths that the human eye perceives as visible light. For most people, this band extends from approximately 750 nanometers (nm), representing red light, down to 390 nm, representing violet light. The visibility curve peaks in the yellow-green region at about 560 nm; 1 nm = 0.000

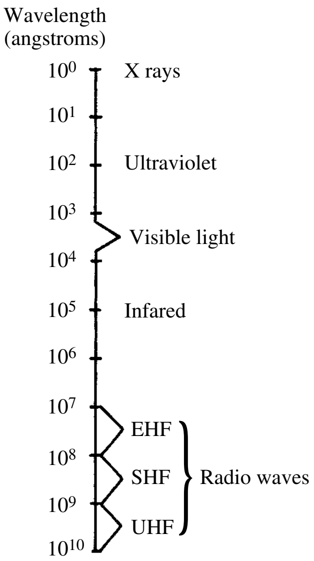

visible spectrum

000 001 meter = 10^{-9} m. Also see ELECTROMAGNETIC THEORY OF LIGHT.

vision system A set of devices that allows a machine (computer or robot) to observe and respond to visual stimuli. Typically, it includes a camera [such as a vidicon or charge-coupled device (CCD)], a cable, electromagnetic or fiberoptic link, a receiver, and software for interpreting the images.

visual alignment Optimization of a circuit (such as a radio receiver) with the aid of meter deflections, rather than with audible indications. Also called *silent alignment.*

visual-aural radio range A radio aid to air navigation (characterized by an audible signal, meter deflection, or both) used by the pilot of an aircraft to determine if the course is being followed.

visual-aural range See VISUAL-AURAL RADIO RANGE.

visual-aural signal tracer A troubleshooting instrument for following a signal through a circuit by sensing it at successive points in the circuit. The signal strength is indicated by a meter and a loudspeaker.

visual carrier frequency See VIDEO CARRIER and VIDEO-FREQUENCY, **1**.

visual communication **1.** Transmission and reception of signals by visible-light means, such as modulated beams of flashing lights. **2.** Transmission and reception of messages by direct visual observation, such as signal lights.

visual display unit Abbreviation, VDU. **1.** A computer peripheral consisting of a cathode-ray tube, a keyboard, and often a pointing device. **2.** A dumb terminal for operation of a computer from a remote site.

visual horizon **1.** The distance from a given point to the farthest visible point on the surface of the earth in a particular direction. **2.** The enclosed geometric plane figure on the surface of the earth, representing the set of farthest-visible points for a particular location.

visual telegraphy Telegraphy in which the received signals are read from a visual device, such as a blinking light, meter, or swinging pen.

visual transmitter The equipment used to transmit the picture portion of a television broadcast.

visual transmitter power The peak power output of a visual transmitter during normal operation.

vitreous Pertaining to a material or surface resembling, or related to, glass (e.g., the *vitreous enamel* on certain types of resistors).

VLF Abbreviation of VERY LOW FREQUENCY.

VLR **1.** Abbreviation of VERY LOW RESISTANCE. **2.** Abbreviation of VERY LONG RANGE.

VLSI Abbreviation of VERY-LARGE-SCALE INTEGRATION.

V/m Abbreviation of VOLTS PER METER (a unit of electric field strength).

VMOS Abbreviation of *vertical-metal-oxide semiconductor.*

VMOSFET Abbreviation of VERTICAL-METAL-OXIDE SEMICONDUCTOR FIELD-EFFECT TRANSISTOR.

V_o Symbol for OUTPUT VOLTAGE. Also, V_{out}.

VOA Abbreviation of VOLT-OHM-AMMETER.

vocabulary In computer-programming operations, a list of available operating codes and instructions for the computer. Also called INSTRUCTION SET.

Vocoder A device for reducing speech to frequencies low enough for efficient transmission through a limited-bandwidth channel. The term is a contraction of *voice coder.*

Voda In a telephone system that utilizes a radio link (using the same frequency for transmission in both directions) and land-line links, an automatic, voice-operated switching system for enabling the transmitter and disabling the receiver, and vice versa. The name is an acronym for *voice-operated device, anti-sing.*

Vodacom Contraction of *voice data communications.*

Voder An electronic device that synthesizes speech. The term is an acronym for *voice operation demonstrator.*

vogad A type of automatic gain control for audio amplifiers and modulators. In a radio transmitter, it maintains 100% modulation of the carrier—even when the speaker's voice level varies widely. The name is an acronym for *voice-operated gain adjusting device.*

voice actuation See VOICE-OPERATED CONTROL.

voice analyzer A circuit that evaluates the characteristics of human voices, such as the amplitude-vs.-frequency function or the amplitude-vs.-time function.

voice band See VOICE FREQUENCIES.

voice coil The moving coil of a dynamic microphone or dynamic speaker.

voice-controlled break-in A type of break-in operation for radiotelephony, in which the transmitter automatically switches on and the receiver switches off immediately when the operator starts talking; the transmitter switches off and the receiver switches on a moment after the operator stops talking.

voice-controlled relay An electronic relay that is actuated by the human voice.

voice filter **1.** A filter for passing, suppressing, or modifying voice frequencies. **2.** A parallel-resonant circuit inserted into a line that feeds several loudspeakers, to trap certain frequencies and thereby improve the sound of the reproduced voice.

voice frequencies **1.** The audio-frequency range of human speech, from about 60 Hz to 8000 Hz. **2.** The frequencies within the audio passband of a typical single-sideband voice transmitter or receiver. The lower limit is generally 200 Hz to 300 Hz; the upper limit is about 2500 Hz to 3000 Hz.

voice-frequency carrier telegraphy A type of carrier-current telegraphy (see WIRED WIRELESS) in which the modulated carrier can be transmitted over a telephone line having a voice-frequency bandwidth.

voice-frequency dialing A system of telephone dialing involving the conversion of direct-current pulses into voice-frequency alternating-current pulses.

voice-frequency telephony Wire telephone communication in which the frequencies of the electric waves are identical to the frequencies of the sound waves (or very nearly so).

voice grade Pertaining to a communications system having a bandpass capable of transferring a human voice with reasonable intelligibility. See VOICE FREQUENCIES, **2**.

voice-grade channel **1.** A telephone line and attendant equipment suitable for the transmission of speech and certain other information, such as control signals, digital data, etc. **2.** In a radiotelephone transmitter, a speech amplifier-modulator channel suitable only for voice frequencies.

voice-operated control Abbreviation, vox. Pertaining to a device (such as a relay, automatic modulation control, transmit-receive switch, etc.) that is actuated when the operator speaks into a microphone.

voice-operated device, anti-sing See VODA.

voice-operated gain-adjusting device See VOGAD.

voice-operated loss control and suppressor In wire telephony, a device that switches the loss from the transmitting line to the receiving line when the subscriber speaks, and vice versa.

voice-over The simultaneous recording of a human voice (such as that of a narrator) along with other sounds (such as music, children playing, wind in trees, ocean waves, chirping birds, etc.). The voice is generally louder than the other sounds.

voice print A graphic recording of the speech frequencies produced by an individual and used as a means of identifying that individual.

voice-recognition device See SPEECH RECOGNIZER.

voice security In voice communications system, the use of encryption and decryption, usually in the form of digital algorithms, to “scramble” speech at the source and “unscramble” it at the destination.

voice-stress analyzer Abbreviation, VSA. An instrument that samples the spoken voice and produces a display from which the relative amount of stress experienced by the speaker can be determined, and from which, in turn, probable truth or falsity of statements or answers can be inferred.

voice synthesis device See SPEECH SYNTHESIZER.

void space See VACUUM.

vol Abbreviation of VOLUME.

volatile **1.** Capable of evaporating (e.g., *volatile solvents* used in the encapsulation of electronic equipment). **2.** Explosive (noun and verb). **3.** Pertaining to a state which is difficult to maintain (e.g., a VOLATILE MEMORY).

volatile memory Memory, usually of the random-access type, whose data vanishes when power is removed unless some provision is made for memory backup. The most common means of memory backup is the use of an electrochemical cell or battery. Modern memory chips need so little current to store their data that a backup battery lasts almost as long in the circuit as it would on the shelf. The main advantage of this type of memory is the fact that it can easily be erased if desired. Compare NONVOLATILE MEMORY.

volatile storage See VOLATILE MEMORY.

volatile store See VOLATILE MEMORY.

Voldicon A form of semiconductor logic device used for analysis of analog signals. Trade name of Adage, Inc.

volt Abbreviation, V. The basic practical unit of difference of potential (i.e., of electrical pressure); 1 volt is the difference of potential produced across a resistance of 1 ohm by a current of 1 ampere. Also see KILOVOLT, MEGAVOLT, MICROVOLT, MILLIVOLT, NANOVOLT, and PICOVOLT.

Volta effect See VOLTA'S PRINCIPLE.

voltage Symbols, E, e, V, v. Electromotive force, or difference of potential; $E = IR$, where I is current and R is resistance. Also see VOLT.

voltage-actuated device An electronic device, such as a field-effect transistor, that amplifies a voltage signal or is controlled by a voltage, and draws virtually no signal current or control current. The opposite is a current-actuated device, such as a bipolar transistor.

voltage amplification **1.** Abbreviation, A_v. Amplification of an input-signal voltage to provide a higher output-signal voltage. **2.** Abbreviation, A_v. The signal increase (V_{out}/V_{in}) resulting from this process. Also called *voltage gain.*

voltage-amplification device A low-current device designed especially for voltage amplification. It provides little or no power amplification.

voltage amplifier An amplifier operated primarily to increase a signal voltage. Compare CURRENT AMPLIFIER and POWER AMPLIFIER.

voltage at peak torque Symbol, V_p. For a torque motor operated at 25°C, the voltage required to produce peak torque at standstill.

voltage attenuation **1.** The reduction in voltage at a given point in a circuit. **2.** For a device, the ratio of input voltage to output voltage when the output voltage is the lower quantity.

voltage-balance relay A relay actuated by a voltage differential.

voltage breakdown See BREAKDOWN VOLTAGE.

voltage-breakdown test **1.** A test in which the measured voltage applied to an insulating material is continuously increased until the breakdown point is reached. **2.** A test in which the measured reverse voltage applied to a semiconductor junction is continuously increased until the reverse breakdown point is reached (see AVALANCHE BREAKDOWN).

voltage burden The voltage drop across a CURRENT SHUNT.

voltage calibrator A device used to calibrate, in terms of voltage, a meter scale, oscilloscope screen, graphic-recorder chart, etc.

voltage-capacitance curve A plot depicting the variation of capacitance with applied voltage for a voltage-variable capacitor. For a varactor (variable-capacitance diode), capacitance varies inversely with reverse direct-current voltage.

voltage coefficient A figure that shows the extent to which a quantity drifts under the influence of voltage. It is generally expressed in percent per volt or in parts per million per volt (ppm/V).

voltage coefficient of capacitance For a voltage-dependent capacitor, the capacitance change per unit change in applied voltage.

voltage coefficient of resistance For a voltage-dependent resistor, the resistance change per unit change in applied voltage.

voltage coil See POTENTIAL COIL.

voltage comparator A device for comparing (usually only two) voltages. The various types range from simple, manually balanced potentiometers to analog or digital devices that automatically compare the applied voltages and provide a direct readout of either their difference or the percent of unbalance.

voltage control **1.** A component or circuit that allows the adjustment of the output voltage of a power supply within a given range. **2.** The adjustment of the output voltage of a power supply to optimize the performance of a circuit connected to the supply. **3.** Any form of circuit control that is accomplished by the adjustment of the voltage at a given circuit point.

voltage-controlled amplifier Abbreviation, VCA. An amplifier in which gain is controlled by means of a voltage applied to a control terminal.

voltage-controlled attenuator An attenuator circuit in which a transistor serves as a voltage-variable resistor. The output resistance of the transistor varies inversely with the direct-current bias voltage applied to the input electrode (base, emitter, or gate).

voltage-controlled capacitor See VOLTAGE-DEPENDENT CAPACITOR.

voltage-controlled crystal oscillator Abbreviation, VCCO or VCXO. A voltage-controlled oscillator of the crystal-stabilized type.

voltage-controlled generator Abbreviation, VCG. Any signal-generating device whose output frequency is varied by changing one of the direct-current operating voltages of the device.

voltage-controlled oscillator Abbreviation, VCO. An oscillator of the inductance-capacitance (LC) type that employs a varactor diode in the capacitive portion of the tuned circuit. The varactor is placed in series or parallel with the tuning capacitor and is isolated for direct current by blocking capacitors. The frequency can be adjusted, within certain limits, by nonmechanical means. The signal can also be frequency modulated by applying the modulating voltage across the varactor.

voltage-controlled resistor See VOLTAGE-DEPENDENT RESISTOR.

voltage corrector See VOLTAGE REGULATOR.

voltage crest See VOLTAGE PEAK.

voltage-current characteristic See VOLTAGE-CURRENT CURVE.

voltage-current crossover The point at which a voltage-regulated supply becomes a current-regulated supply.

voltage-current curve For a circuit or device, the plot of the interrelationship of current and voltage, with voltage as the independent variable.

voltage-current feedback See CURRENT-VOLTAGE FEEDBACK.

voltage decay The exponential decrease of voltage across a discharging capacitor. Also see EXPONENTIAL DECREASE.

voltage-dependent capacitor A capacitor (such as a varactor) whose capacitance varies with applied voltage.

voltage-dependent resistor A nonlinear resistor whose value varies inversely with the voltage drop across it. Also called *varistor.*

voltage detector A circuit or device that delivers an output voltage only when the input voltage is of a prescribed value. Compare VOLTAGE DISCRIMINATOR.

voltage-determined property A property (capacitance, current, frequency, or resistance) whose magnitude depends on the value of an applied voltage. See, for example, VOLTAGE-DEPENDENT CAPACITOR, VOLTAGE-DEPENDENT RESISTOR, and VOLTAGE-CONTROLLED OSCILLATOR.

voltage differential Symbol, dE or dV. An infinitesimal change in voltage. Compare VOLTAGE INCREMENT.

voltage-directional relay **1.** A relay that is actuated when voltage exceeds a certain value in a given direction. **2.** A relay that closes only when the applied voltage is in a specific direction.

voltage discriminator A circuit or device whose output voltage is zero when the input voltage is of a prescribed value. When the input voltage is greater than this value, the output is positive; when it is less, the output is negative.

voltage distribution **1.** The delivery of operating voltage throughout a circuit (e.g., high and low direct-current voltages in the various stages of a control circuit). **2.** Sometimes, the distribution of electrical energy ("power" distribution).

voltage divider A resistive or capacitive potentiometer or network used to divide an applied voltage by a desired amount.

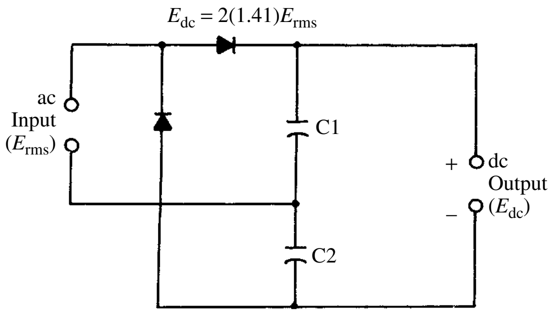

voltage doubler

voltage doubler A power-supply circuit that supplies a direct-current output voltage of about twice the peak value of the alternating-current input voltage.

voltage drift See DRIFT VOLTAGE.

voltage drop Abbreviation, VD. The voltage (E) set up across a resistance (R) carrying a current (I); $E = IR$. For alternating current, reactance X and impedance Z can be substituted for resistance, where applicable.

voltage-equalizing resistors In a power-supply filter, resistors connected across each capacitor in a string of electrolytics connected in series to withstand a high voltage. The resistors protect the capacitors by equalizing the voltage across them.

voltage-fed antenna An antenna in which the feeder is attached to the radiator at a voltage loop (current node). Compare CURRENT-FED ANTENNA.

voltage feed The delivery of voltage to a device or circuit at a point where voltage, rather than current, is dominant. Compare CURRENT FEED.

voltage feedback A feedback signal consisting of voltage fed from the output to the input circuit of an amplifier or other device. Compare CURRENT FEEDBACK and CURRENT-VOLTAGE FEEDBACK.

voltage-frequency transducer **1.** A device whose output voltage is a function of the frequency of a signal at the input. **2.** A device whose output frequency is a function of the voltage at the input.

voltage gain See VOLTAGE AMPLIFICATION.

voltage generator See VOLTAGE SUPPLY.

voltage gradient The voltage per unit length along a defined path.

voltage increment A change in voltage represented by $E_2 - E_1$, where E_2 and E_1 are the voltages at two defined points on a curve, and the difference is finite (nonzero). Compare VOLTAGE DIFFERENTIAL.

voltage input encoder An analog-to-digital encoder for which the input is an analog voltage. See, for example, VOLTAGE-TO-SHAFT-POSITION ENCODER.

voltage inverter A circuit or device whose voltage output has the sign opposite that of the input voltage. The device might or might not provide amplification.

voltage jump **1.** An upward transient in voltage. **2.** In a glow-discharge tube, a sudden break or increase in voltage drop across the tube.

voltage lag The condition in which voltage changes occur after corresponding current changes. Compare CURRENT LAG and VOLTAGE LEAD.

voltage lead The condition in which voltage changes precede corresponding current changes. Compare CURRENT LEAD and VOLTAGE LAG.

voltage level **1.** A prescribed reference value of voltage (e.g., the black level in a television picture signal). **2.** The discrete value of a steady voltage, or the average value of a fluctuating voltage, as observed or measured in a circuit.

voltage limit The maximum or minimum level in a voltage range.

voltage loop In a standing-wave system, such as an antenna or transmission line, a maximum-voltage point. Compare VOLTAGE NODE.

voltage loss **1.** Reduction of a voltage across a load, occurring because of a series resistance. **2.** The ratio, in decibels, between the input voltage to a transmission line and the output voltage at the load end of the line, assuming that the impedance is the same at both points.

voltage maximum See VOLTAGE PEAK.

voltage minimum See VOLTAGE TROUGH.

voltage-mode switching circuit A resistor-transistor-logic (RTL) NAND or NOR circuit in which (in the off state) the transistor is cut off by the V_{BB} bias voltage. The output is then approximately equal to the collector supply voltage, V_{CC}. The proper combination of input pulses overrides the cutoff bias, and the transistor switches on. The output then drops to a level equal to V_{CC} minus the voltage drop across the external collector resistor.

voltage multiplier A special type of rectifier circuit that delivers a direct-current output voltage that is a multiple of the peak value of the alternating-current input voltage, thus affording voltage step-up without a transformer. See, for example, VOLTAGE DOUBLER; VOLTAGE QUADRUPLER; QUINTUPLER, **1**; and VOLTAGE TRIPLER.

voltage node In a standing-wave system, such as an antenna or transmission line, a minimum-voltage point. Compare VOLTAGE LOOP.

voltage of self-induction The voltage drop across an inductor, resulting from the flow of alternating current through the inductor; it is caused by self-induction.

voltage peak The highest value attained by a voltage during an excursion. Also called *voltage crest* or *voltage maximum.* Compare VOLTAGE TROUGH.

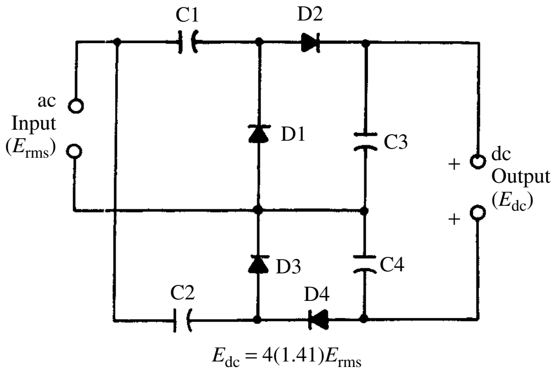

voltage quadrupler

voltage-power directional relay A relay system in which two circuits are connected when their voltage difference exceeds a predetermined value in one direction, and are disconnected when the voltage in the opposite direction exceeds a predetermined level.

voltage quadrupler A special rectifier circuit that supplies a direct-current output voltage of approximately four times the peak alternating-current input voltage.

voltage quintupler See QUINTUPLER, **1**.

voltage-range multiplier **1.** A multiplier resistor (see VOLTMETER MULTIPLIER) connected in series with a voltmeter that has an internal multiplier, to increase the range of the instrument. **2.** For an alternating-current voltmeter, an input amplifier used to increase the sensitivity of the instrument.

voltage rating **1.** For a circuit or device, the recommended maximum voltage that can be applied, or the recommended working voltage, as specified. **2.** For a generator, the specified output voltage.

voltage ratio The quotient of the voltages, E_1/E_2, at two specific points in a circuit, device, or system. Examples: ratio of input voltage to output voltage and ratio of primary voltage to secondary voltage.

voltage-ratio box See VOLT BOX.

voltage reference See STANDARD CELL and ZENER-DIODE VOLTAGE REFERENCE.

voltage-reference cell See STANDARD CELL.

voltage-reference diode A Zener diode biased into its Zener region. The voltage drop across the diode is comparatively constant. Under proper conditions, it can be used for reference purposes. Also see ZENER-DIODE VOLTAGE REFERENCE.

voltage reflection coefficient In a reflected-wave situation, the ratio E_r/E_i, where E_i is the field-strength voltage of the incident wave, and Er is the field-strength voltage of the reflected wave.

voltage-regulated supply See CONSTANT-VOLTAGE SOURCE.

voltage-regulating transformer A special transformer in which a resonant circuit and core saturation (see SATURATED OPERATION, **1**) are used to provide a constant output voltage.

voltage regulation The stabilization of a voltage against fluctuations in source or load.

voltage-regulation constant Symbol, K_v. For a voltage-regulated power supply, the ratio dE_L/dE_i, where dE_i is a change in input voltage, and dE_L is the corresponding change in load voltage.

voltage regulator A circuit or device that holds an output voltage constant during variations in the output load or input voltage.

voltage-regulator diode See ZENER-DIODE VOLTAGE REGULATOR.

voltage-regulator transformer See VOLTAGE-REGULATING TRANSFORMER.

voltage relay A relay or relay circuit that is actuated by a discrete voltage, rather than by current or power.

voltage-responsive device See VOLTAGE-ACTUATED DEVICE.

voltage rise The normal condition in a series-resonant circuit, in which the voltage across the coil or capacitor is higher than the voltage applied to the circuit.

voltage saturation In a current-actuated device, such as a bipolar transistor, the situation in which an increase in current provides no increase in voltage drop beyond a certain point (the saturation point).

voltage-sensitive bridge A bridge having a nonlinear element (such as a voltage-dependent resistor) as one of the arms. Because of this element, the bridge can be balanced (with a given set of other arms) at only one value of applied voltage. At lower voltages, the bridge becomes unbalanced in one direction; at higher voltage, in the opposite direction.

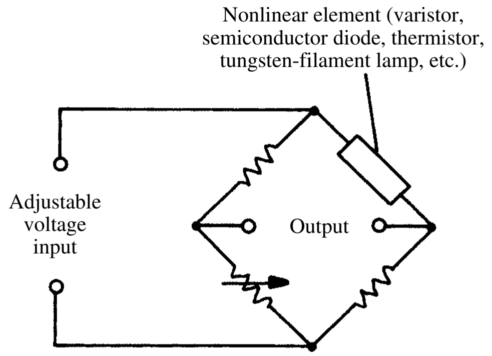
Nonlinear element (varistor, semiconductor diode, thermistor, tungsten-filament lamp, etc.)
Adjustable voltage input
Output

voltage-sensitive bridge

voltage-sensitive capacitor See VOLTAGE-DEPENDENT CAPACITOR.

voltage-sensitive resistor See VOLTAGE-DEPENDENT RESISTOR.

voltage sensitivity **1.** Particular responsiveness of a circuit or device to voltage, rather than current or power. See, for example, VOLTAGE-DEPENDENT CAPACITOR, VOLTAGE-DEPENDENT RESISTOR, and VOLTAGE RELAY. **2.** See VOLTMETER SENSITIVITY.

voltage-spectrum function The voltage-vs.-frequency curve at the output of a circuit or transducer.

voltage-stabilized supply See CONSTANT-VOLTAGE SOURCE.

voltage stabilizer See VOLTAGE REGULATOR.

voltage-stabilizing diode See ZENER-DIODE VOLTAGE REGULATOR.

voltage standard A device that delivers a voltage of accuracy and stability so that it can be used to calibrate other voltage generators and test instruments. See, for example, STANDARD CELL and ZENER-DIODE VOLTAGE REFERENCE.

voltage standing-wave ratio Abbreviation, VSWR. In a standing-wave system, the ratio of the maximum voltage to the minimum voltage.

voltage supply A power supply whose usable output is voltage, rather than current or power. When such a supply is not voltage-regulated, it can only be used reliably with a very light load.

voltage to chassis In electronic equipment mounted on a metal chassis, the voltage between the chassis and a given point in the circuit.

voltage-to-frequency converter A device or circuit that delivers an output frequency proportional to an input voltage (usually direct current). Compare FREQUENCY-TO-VOLTAGE CONVERTER.

voltage to ground **1.** In a circuit, the voltage measured between a given point and the ground point. **2.** The voltage measured between the earth and a line or piece of equipment.

voltage to panel In electronic equipment mounted on a metal panel, the voltage between the panel and a given point in the circuit.

voltage-to-shaft-position encoder An encoder for which the output is the rotation of a motor shaft over an arc proportional to an input voltage.

voltage transformer **1.** A transformer used primarily to supply voltage with little or no current. **2.** A small step-up transformer for increasing the sensitivity of an alternating-current voltmeter. Also called POTENTIAL TRANSFORMER.

voltage tripler A rectifier circuit that, without a transformer, supplies a direct-current output voltage of approximately three times the peak value of the alternating-current input voltage.

voltage trough The lowest value reached by a voltage during an excursion. Compare VOLTAGE PEAK.

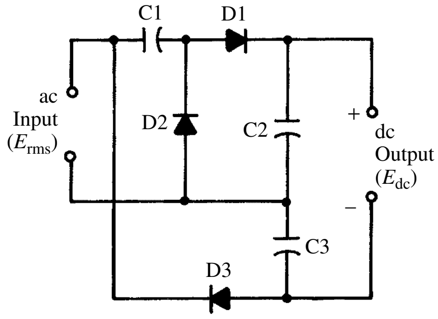

$E_{dc} = 3(1.41)E_{rms}$

voltage tripler

voltage tuning A method of adjusting the frequency of an oscillator or resonant circuit by means of a variable, direct-current voltage.

voltage-tunable magnetron A magnetron oscillator in which the output frequency is varied by adjusting the direct-current voltage on the tube.

voltage-tunable oscillator See VOLTAGE-CONTROLLED OSCILLATOR.

voltage-turns ratio The turns ratio of a transformer, indicated by the ratio of primary voltage to secondary voltage, or vice-versa.

voltage-type telemetry Telemetry based upon the variation of a single voltage in sympathy with the changes in a sensed phenomenon.

voltage-variable capacitor **1.** A specially processed semiconductor diode of which the voltage-variable capacitance of the junction is utilized. Also called VARACTOR. **2.** A capacitor having a specially processed nonlinear dielectric, such as barium strontium titanate, whose capacitance varies inversely with the applied direct-current bias voltage.

voltage-variable resistor See VOLTAGE-DEPENDENT RESISTOR.

voltage vector In a vector diagram, a vector showing the magnitude and phase of a voltage. Compare CURRENT VECTOR.

voltaic Pertaining to chemically produced direct current. Sometimes interchangeable with *galvanic.*

voltaic cell For the generation of a direct-current voltage, a primary cell consisting of two electrodes made of different metals and immersed in an electrolyte. Also called GALVANIC CELL. Also see CELL, ELECTROMOTIVE SERIES, and PRIMARY CELL.

voltaic couple A pair of dissimilar metals (or other substances) that generate a direct-current voltage when they contact an electrolyte.

voltaic pile A rudimentary primary battery consisting of a series of disks made of two different metals, stacked alternately with electrolyte-soaked cloth or paper disks.

voltaic series See ELECTROMOTIVE SERIES.

voltameter An electrolytic cell for determining the value of an unknown current, or of an unknown quantity of electricity, from the weight of metal deposited out of an electrolyte onto the cathode by the passage of current over an accurately timed interval.

volt-ammeter A combination meter for measuring electrical voltages and currents.

volt-ampere Symbol, VA. Unit, watt. The simple product of voltage and current in volts and amperes, yielding the true power in a direct-current circuit and the apparent power in an alternating-current circuit. Also see APPARENT POWER and TRUE POWER.

volt-ampere-hour meter See VAR-HOUR METER.

volt-ampere meter See VAR METER.

volt-amperes reactive Abbreviation, VAR. The product of volts and amperes for a purely reactive circuit. This product does not produce the true power because the power factor is neglected; it produces only the apparent power. A true reactance absorbs power during one half-cycle of alternating current, and returns it to the generator during the next half-cycle.

Volta's law See VOLTA'S PRINCIPLE.

Volta's pile See VOLTAIC PILE.

Volta's principle Two dissimilar metals brought into contact (even in air) will generate a difference of potential whose value is characteristic of the metals. Also see ELECTROMOTIVE SERIES.

volt box A precision, potentiometer-type voltage divider used in the calibration of meters and other instruments. The device is usually provided with a set of terminal posts for selecting various ratios of output voltage to input voltage.

volt-electron See ELECTRONVOLT.

voltmeter A usually direct-reading instrument used to measure voltage. Also see ELECTRONIC VOLTMETER, FET VOLTMETER, TRANSISTOR VOLTMETER, and VACUUM-TUBE VOLTMETER.

voltmeter-ammeter See VOLT-AMMETER.

voltmeter amplifier A wideband, flat-frequency-response, low-distortion preamplifier used to boost the sensitivity of an alternating-current voltmeter.

voltmeter-millivoltmeter A voltmeter that provides several low ranges, as well as several high ones. A familiar example is an alternating-current voltmeter, which has various full-scale ranges from about 1 millivolt to about 1000 volts.

voltmeter multiplier A resistor connected in series with a current meter (usually a milliammeter or microammeter) to convert it into a voltmeter.

voltmeter sensitivity Unit, ohm per volt. For a voltmeter, the total resistance of the instrument (multiplier resistance plus the resistance of the meter movement) divided by the full-scale deflection of the meter. Thus, a 0-to-10 direct-current voltmeter with an input resistance of 100 kilohms has a sensitivity of 10,000 ohms per volt.

volt-milliammeter A combination meter for measuring volts and milliamperes.

volt-ohm-ammeter A multimeter for measuring voltage, resistance, and current (in amperes).

volt-ohmmeter A combination meter for measuring voltage and resistance.

volt-ohm-milliammeter Abbreviation VOM. A multimeter for measuring voltage, resistance, and current (in milliamperes and microamperes).

voltsensor See VOLTAGE DETECTOR.

volts per meter Abbreviation, V/m or Vpm. The unit of electric field strength.

volume **1.** Intensity of sound. Also called LOUDNESS. **2.** A circumscribed portion of space, either imaginary or actually occupied, and described by three dimensions (e.g., sphere, ellipsoid, cube, pyramid, etc.). **3.** In a computer system, a unit of magnetic storage medium (e.g., a diskette).

volume compression The automatic reduction of the gain of an audio amplifier. The process differs from clipping (which "slices off" the tops of waves) in that compression (ideally) reduces the amplitude while preserving the waveform. Compare VOLUME EXPANSION.

volume compressor A circuit or device (such as an automatic-gain-control amplifier) for achieving volume compression. Compare VOLUME EXPANDER.

volume conductivity The reciprocal of VOLUME RESISTIVITY.

volume control A variable resistor (usually a potentiometer) or a network of resistors (such as a pad) for adjusting the gain, and, consequently, the output-signal loudness, of an amplifier.

volume equivalent In wire telephony, speech loudness throughout the system, expressed in terms of the trunk loss in a reference system and adjusted for equal loudness.

volume expander A circuit or device for automatically boosting the volume of an audio-frequency signal. Also see EXPANDER. Compare VOLUME COMPRESSOR.

volume expansion The technique of automatically increasing the gain, and consequently the output-signal volume, of an audio amplifier. Also see VOLUME EXPANDER. Compare VOLUME COMPRESSION.

volume indicator A fast-acting alternating-current meter used to monitor the volume level in an audio channel in which the signal level is fluctuating. The scale is graduated in VOLUME UNITS. Also called *VU meter.*

volume lifetime In a homogeneous semiconductor, the interval between minority-carrier generation and recombination.

volume limiter A circuit or device that automatically holds the volume level of an audio channel to a predetermined maximum. Also see VOLUME COMPRESSION and VOLUME COMPRESSOR.

volume-limiting amplifier An amplifier that functions as a volume limiter through the action of volume-limiting subcircuits.

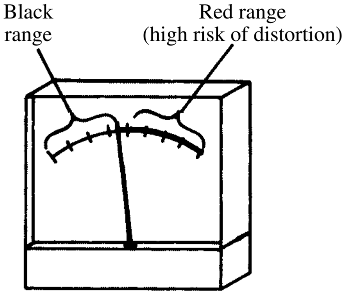

volume indicator

volume magnetostriction **1.** The decrease in the total volume of certain ferromagnetic substances under the influence of a magnetic field. **2.** The increase in the total volume of certain ferromagnetic substances under the influence of a magnetic field.

volume range Unit, decibel (dB). The difference between the maximum and minimum volume levels that a device or system can accommodate.

volume recombination rate Within the volume of a semiconductor, the rate at which electrons and holes recombine.

volume resistance The effective resistance, through a given medium, between two electrodes placed within that medium.

volume resistivity The resistance of a specific volume of a material (e.g., the resistance between opposite faces of a 1-centimeter cube of the material). Also see MICROHM-CENTIMETER; OHM-CENTIMETER; RESISTIVITY.

volumetric efficiency In an electronic assembly, the ratio of the volume of parts to the total volume of the assembly.

volumetric radar A radar providing a three-dimensional display.

volumetric sensor In security systems, a device that detects effects in a specifically defined three-dimensional region, such as an entire room (from floor to ceiling).

volume unit Abbreviation, VU. The unit of measurement of fluctuating alternating-current power, such as that of speech or music. Zero VU corresponds to a reference power of 2.51 milliwatts, or +4 dBm (four decibels above 1 milliwatt). Volume units are measured with a VOLUME INDICATOR.

volume-unit indicator See VOLUME INDICATOR.

volunteer examiner In amateur radio, a person who works with the Federal Communications Commission (FCC) to administer license examinations. Such a person is not actually an employee of the FCC, but serves on an independent, volunteer basis.

VOM Abbreviation of VOLT-OHM-MILLIAMMETER.

von Hippel breakdown theory The theory that, assuming no electron-energy distribution, breakdown occurs at field intensities for which the recombination rate of electrons and holes is lower than the rate of ionization by collision.

von Neumann architecture In computer systems, a design scheme in which the bits are transmitted serially (one by one) along a single line or bus.

VOR Abbreviation of *very-high-frequency omnirange* (see VHF OMNIRANGE).

voter See MAJORITY LOGIC.

vox Abbreviated form of VOICE-OPERATED CONTROL.

V_P **1.** Symbol for GATE-SOURCE PINCHOFF VOLTAGE of a field-effect transistor. **2.** Symbol for PLATE VOLTAGE.

V_p Symbol for VOLTAGE AT PEAK TORQUE.

V particle See VEE PARTICLE.

Vpm Abbreviation of VOLTS PER METER. (Also, V/m.)

VR **1.** Abbreviation of VIRTUAL REALITY. **2.** Abbreviation of VOLTAGE REGULATOR.

VRAM Abbreviation of VIDEO RANDOM-ACCESS MEMORY.

V_{ref} Abbreviation of *reference voltage.*

vrr Abbreviation of *visual radio range.*

V•s Abbreviation of *volt-seconds* (webers).

V•s/A Abbreviation of *volt-seconds per ampere* (henrys).

VSA Abbreviation of VOICE-STRESS ANALYZER.

VSB Abbreviation of VESTIGIAL SIDEBAND.

VSF Abbreviation of VESTIGIAL-SIDEBAND FILTER.

V signal See VEE SIGNAL.

vsr Abbreviation of VERY SHORT RANGE.

VSWR Abbreviation of VOLTAGE STANDING-WAVE RATIO.

vt **1.** Abbreviation of VACUUM TUBE. **2.** Abbreviation of *variable time.*

VT fuse See PROXIMITY FUSE.

VTL Abbreviation of VARIABLE-THRESHOLD LOGIC.

vtm Abbreviation of VOLTAGE-TUNABLE MAGNETRON.

VTO Abbreviation of VOLTAGE-TUNABLE OSCILLATOR.

VTR Abbreviation of VIDEOTAPE RECORDER or VIDEOTAPE RECORDING.

VTVM Abbreviation of VACUUM-TUBE VOLTMETER.

vt voltmeter See VACUUM-TUBE VOLTMETER.

VU Abbreviation of VOLUME UNIT.

vulcanized fiber A tough insulating material derived from cellulose. Dielectric constant, 5 to 8.

VU meter See VOLUME INDICATOR.

VVCD Abbreviation of *voltage-variable capacitor diode* (see VOLTAGE-VARIABLE CAPACITOR, **1**).

VVV signal See VEE SIGNAL.

VW Abbreviation of *volts working* (see WORKING VOLTAGE).

vy Abbreviation for *very.*

W

W **1.** Symbol for WORK. **2.** Abbreviation of WATT. **3.** Symbol for TUNGSTEN. **4.** Symbol for ENERGY. **5.** Abbreviation of *west.* **6.** Abbreviation of WIDTH.

w **1.** Abbreviation of WEIGHT. (Also, wt.) **2.** Abbreviation of *week.*

WAC Amateur radio abbreviation of *Worked All Continents*, an award given to operators who have engaged in verified two-way communication with stations on all continents.

wafer **1.** Semiconductor die. **2.** A thin, flat disk, ring, or plate around which the contacts of a rotary switch are spaced. **3.** A thin square or rectangle of dielectric material used as the dielectric member in a fixed capacitor. **4.** A plate cut from a crystal (e.g., a quartz wafer).

wafer fabrication The various processes used in the manufacture of semiconductor integrated circuits.

wafer slicing Cutting plates from a mother crystal, as when piezoelectric plates are cut from a quartz crystal.

wafer socket A component socket consisting of a plastic or ceramic wafer with spring-type contacts for gripping the pins.

wafer switch A rotary switch whose contacts are arranged around the periphery of a plastic or ceramic wafer.

Wagner ground A circuit (often a single potentiometer) that facilitates cancellation of stray reactance in an alternating-current bridge. The bridge is balanced alternately with the bridge-balance control and the Wagner control, until there is no further shift of null point when changing from one to the other.

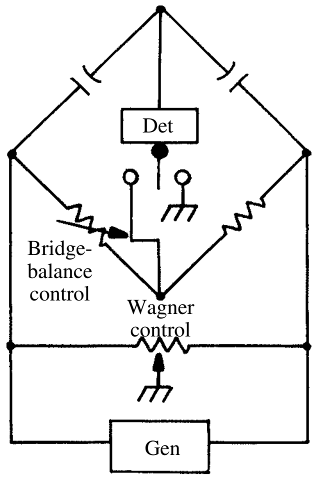

Wagner ground (with capacitance bridge)

waiting time **1.** See WARMUP TIME. **2.** The delay between the time a data transfer to or from a computer memory is called for and the actual transfer of the data. Also called LATENCY.

walkie-lookie A portable combination camera and transmitter for remote television pickup. At sports events and other gatherings, the unit is strapped to the shoulder of the camera operator.

walkie-talkie A portable, compact transceiver (or transmitter-receiver).

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

wall absorption Diminished radioactive emission (especially of beta and gamma rays) because of absorption by the radiating substance.

wall box A wall-mounted (usually metal) box enclosing circuit breakers, fuses, switches, etc.

wall effect In an ionization chamber, the rise in ionization because of electrons being released by the walls of the chamber.

wall energy In a ferromagnetic substance, the energy per unit area stored in a domain wall between two regions of opposite magnetization.

Wallman amplifier A cascode amplifier (see CASCODE).

wall mount A metal bracket for fastening an antenna to a wall.

wall outlet A plug or socket, usually mounted in a protective box or can and recessed in a wall that can be accessed from the front. It is commonly used for supplying alternating-current power to appliances at 117 volts or 234 volts. It is also used in telephone and cable-television systems.

wall plaque A loudspeaker so thin that, mounted in a frame (sometimes behind grill cloth), it can be hung on a wall.

wall plate A (usually rectangular) plate of metal or plastic for holding a wall outlet or wall switch.

wall plug A male or female plug usually mounted in a protective box or can and recessed in a wall. Such a device can provide easy access to an antenna, telephone line, or load; or it can deliver alternating-current, direct-current, or radio-frequency power.

walls The sides of the groove cut into a record disc.

wall socket A male or female socket usually mounted in a protective box or can and recessed in a wall. Such a device can provide easy access to an antenna, telephone line, or load; or it can deliver alternating-current, direct-current, or radio-frequency power.

wall speaker See WALL PLAQUE.

wall switch A switch usually mounted in a protective box or can and recessed in a wall.

wall telephone A wall-mounted telephone set. It generally fits over the jack, so the only cord is between the main unit and the receiver.

wall-through tube See LEAD-IN TUBE.

Walmsley antenna A phased array consisting of several full-wavelength loops.

wamoscope A radar display tube that performs several microwave-receiver functions (detection, oscillation, amplification, etc.) as well as displaying an image. The name is an acronym for *wave-modulation oscilloscope.*

wander See SCINTILLATION, **1**.

WAP Acronym for WIRELESS ACCESS PROTOCOL.

warble **1.** A periodic rise and fall in the pitch of a musical tone or combination of tones. **2.** To rise and fall in pitch with a definite period.

warble-tone generator An audio-frequency oscillator whose frequency is varied at a subaudible rate over a fixed frequency range.

warm junction The heated junction in a two-junction thermocouple circuit. Also called HOT JUNCTION. Compare COLD JUNCTION.

warmup The process of stabilizing an electronic equipment by allowing its temperature to rise to the optimum level.

warmup time The time required for an electronic circuit to become fully operational, or to stabilize, after the power has been switched on.

warning bell (buzzer) A bell (or buzzer) used as an audible alarm in a WARNING DEVICE.

warning device A device, such as an electronic signaler, for alerting a person to an emergency (intrusion on premises, danger to life and safety, etc.) or to the existence of an intrusion into a secured area.

warning lamp See WARNING LIGHT.

warning light A lamp used as a visual alarm in a warning device.

warpage Distortion of the normally straight sides of a triangular wave.

warping In multimedia computer graphics, a change in the shape of an object that occurs smoothly over a period of time.

WAS Amateur radio abbreviation of *Worked All States*, an award given to operators who have engaged in verified two-way communication with stations in all states of the United States.

washer capacitor A very thin donut capacitor.

washer resistor A resistor made in the general shape of a washer or ring and having a center hole for a mounting screw or stacking rod.

washer thermistor A thermistor made in the general shape of a washer and having a center hole for a mounting screw or stacking rod.

washer varistor A varistor made in the general shape of a washer and having a center hole for a mounting screw or stacking rod.

washout process A method of fabricating bipolar transistors. The contact metal is deposited in the diffusion hole.

waste instruction In a computer program, an instruction not meant to be acted upon (e.g., one used to take up space in the listing for some reason). Also called *null instruction* and *dummy instruction.*

watch A work shift, as of electronic personnel (e.g., radio station operators).

watchcase receiver An earphone enclosed in a small, round case with a screw-on cap. It is derived from its resemblance to a large pocket watch.

water absorption For a solid material, such as a dielectric, the ratio of the weight of water absorbed by the material to the weight of the material.

water-activated battery A battery that contains all the ingredients of its electrolyte, except water, which must be added when the battery is put into service.

water adsorption The formation of a thin layer of water molecules on the surface of normally dry material, but not by absorption. Also see ADSORPTION.

water analogy The useful, but not wholly accurate, teaching device of comparing an electric current with the flow of water. In such a comparison, voltage is shown equivalent to water pressure, and current to the quantity (e.g., gallons) of water flowing in unit time.

water battery A primary battery or cell using water as the electrolyte.

water calorimeter A calorimeter used to measure power in terms of the increase in temperature of water heated by the electrical energy.

water capacitor An emergency capacitor made by setting one glass of water in another larger jar of water so that the two bodies of water are separated by the walls of the smaller jar. The bodies of water, in which an electrolyte has been dissolved, form the "plates" of the capacitor, and the wall of the smaller jar serves as the dielectric between them.

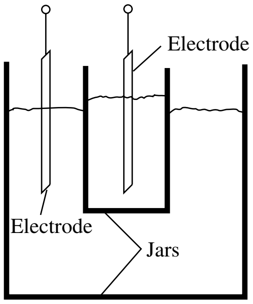

water capacitor

water-cooled tube A power tube, such as a large radio transmitting tube, cooled by the circulation of water in the space between the outside of the tube envelope and a surrounding jacket.

water cooling A method of cooling components by pumping water through pipes surrounding them.

water-flow alarm An electronic circuit that actuates an alarm when the flow of water through pipes or other channels changes from a predetermined rate.

water-flow control A servo system for automatically maintaining or adjusting the flow of water through pipes or other channels.

water-flow gauge See WATER-FLOW METER.

water-flow indicator See WATER-FLOW METER.

water-flow meter An instrument used to monitor the flow of water through pipes or other channels, sometimes showing its direction as well as its rate.

water-flow switch In water-cooled systems (e.g., water-cooled tubes), a switch that actuates an alarm when the water slows or stops.

water glass Sodium silicate, a substance used as a fireproofing agent and protective coating.

water ground An earth connection made by dropping a weighted wire into a body of water.

water jacket In a water-cooled device, the outer jacket that, along with the outer wall of the cooled component, forms a space through which the cooling water flows.

water-level control A servo system that automatically maintains or adjusts the level of water inside a tank or other container.

water-level gauge An electronic system that gives direct water level readings inside a tank or other container.

water-level indicator See WATER-LEVEL GAUGE.

water load **1.** A makeshift, power-dissipating resistive load (see DUMMY LOAD) consisting of a container of tap water or saltwater into which two wires are immersed. **2.** A waveguide termination containing water heated by the microwave energy. It is usable as a water calorimeter.

water monitor A sensitive electronic instrument for checking radioactivity in a water supply.

water-pipe ground An earth connection made by running a wire to the nearest cold-water pipe. This scheme works only with metal pipes devoid of insulating joints or splices.

water power Hydroelectric power (i.e., electrical energy produced by generators driven by water).

water-pressure alarm An electronic circuit that actuates an alarm when water in pipes or other channels changes from a predetermined level.

water-pressure control A servo system for automatically maintaining or adjusting water pressure in pipes or other channels.

water-pressure gauge See WATER-PRESSURE METER.

water-pressure indicator See WATER-PRESSURE METER.

water-pressure meter An instrument that directly indicates water pressure in a pipe or tank.

water-pressure switch A switch that actuates an alarm when water pressure rises or falls.

water pump In a water-cooled electronic system, the (usually rotary) pump that circulates the water.

water resistor An electrolytic resistor in which the electrolyte is tap water or diluted saltwater.

water rheostat A variable water resistor. Usually, the resistance is varied by moving the immersed electrodes closer together or farther apart.

water tester An instrument for checking pH, electrical resistance, and other properties of water.

water witching Locating underground water by electronic methods.

WATS Acronym for *Wide Area Telephone Service.* A form of long-distance telephone service. Rates are charged on a different basis than normal long-distance service. The system is especially favored by businesses because it saves money for subscribers making a large number of calls in each billing period.

watt Abbreviation, W. The practical unit of electric and other power. One watt is dissipated by a resistance of 1 ohm through which a current of 1 ampere flows. See also KILOWATT, MEGAWATT, MICROWATT, and MILLIWATT.

wattage Electrical power, especially when expressed in watts.

wattage rating **1.** The recommended output power of a device. **2.** The recommended power dissipation of a device.

watt current The component of alternating current that is in phase with voltage. Also called RESISTIVE CURRENT. Compare REACTIVE CURRENT.

watt-decibel conversion The conversion of a power level, such as the power output of an amplifier, in watts to the corresponding power level in decibels, with respect to a reference level. Thus, n dB = $10 \log_{10}(P/P_{ref})$, where P is the power of interest (watts), and P_{ref} is the reference level (e.g., one milliwatt).

watt-hour Abbreviation, WH. The unit of electrical energy or work; 1 WH = 3600 joules = 10^{-3} kWH. Also see ENERGY, KILOWATT-HOURS, POWER, and WATT-SECOND.

watt-hour capacity The number of watt-hours that a storage battery can deliver reliably and safely under specified operating conditions.

watt-hour-demand meter A combination watt-hour meter and demand meter.

watt-hour efficiency For a storage battery, the ratio of watt-hours output to watt-hours of recharge.

watt-hour constant In an electric-energy meter, the number of watt-hours in one revolution of the indicating disk.

watt-hour meter An instrument for measuring electrical energy in watt-hours. One well-known type consists essentially of a small motor geared to a row of four dial indicators. An eddy-current disk keeps the motor speed proportional to the watt-hours consumed by a load, a value which is the sum of the readings of the dials. Also called SERVICE METER and KILOWATT-HOUR METER.

wattless current The component of alternating current that is out of phase with voltage. Also called REACTIVE CURRENT. Compare RESISTIVE CURRENT.

wattless power The apparent power in a reactive circuit, indicated by the product of volts and amperes. There is no actual power consumption because the power taken by a reactance during a half-cycle is returned to the generator during the next half-cycle. Also see AC POWER, APPARENT POWER, REACTIVE KILOVOLT-AMPERES, and REACTIVE VOLT-AMPERES.

wattless volt-amperes See WATTLESS POWER.

wattless watts See WATTLESS POWER.

wattmeter Abbreviation, WM. An instrument used to measure electrical power. The scale usually reads directly in watts, kilowatts, milliwatts, or microwatts. Also see ELECTRONIC WATTMETER.

watt-second Abbreviation, Ws. A small unit of electrical energy or work; 1 watt-second = 1 joule = 1/3600 watt-hour. Also see ENERGY and POWER.

watt-second constant In an electric-energy meter, the number of watt seconds in one revolution of the indicating disk.

wave **1.** A single oscillation of some property of matter or space, such as density, displacement, or field strength. It moves outward from a point of disturbance and grows weaker as it travels farther. Wave motion is associated with mechanical vibration, sound, radio, heat, light, X rays, gamma rays, and cosmic rays. See SOUND and WAVELENGTH. **2.** A single cycle of alternating or pulsating current or voltage. Also see AC VOLTAGE, ALTERNATING CURRENT, PULSATING DC VOLTAGE, and PULSATING DIRECT CURRENT.

wave absorption The removal of energy from electromagnetic waves as they pass through certain media, such as solid bodies, water, and the atmosphere. Compare POLARIZATION, WAVE REFLECTION, and WAVE REFRACTION.

wave amplitude The peak value of a wave. Also see WAVE CREST and WAVE TROUGH.

wave analyzer An instrument consisting essentially of a continuously tunable bandpass filter and an electronic alternating-current voltmeter. As the filter is tuned successively to the fundamental frequency of a complex wave and to its various harmonics, the voltmeter shows the amplitude of each of the components. Also see HETERODYNE WAVE ANALYZER.

wave angle The angle, measured with respect to the horizon, at which a radio wave is transmitted or received.

wave antenna See BEVERAGE ANTENNA.

wave attenuation The reduction of wave amplitude, with respect to distance from the source.

waveband A band of radio frequencies. Also called FREQUENCY BAND.

waveband switch See BANDSWITCH.

wave beam Unidirectional radiation from a directive antenna.

wave bounce See WAVE REFLECTION.

wave clutter Radar interference caused by waves on a body of water, particularly large swells on the ocean.

wave converter A waveguide part, such as a baffle-plate or grating, that changes a wave pattern from one type to another.

wave crest The maximum value of a wave envelope. Compare WAVE TROUGH.

wave cycle A complete single alternation of a wave.

wave direction The direction in which an electromagnetic wave travels. It is perpendicular to the wave front and depends (whether it is forward or backward) upon the direction of the electric and magnetic components. If either is reversed, wave direction reverses; if both are reversed, the direction remains unchanged.

wave duct **1.** See DUCT, **1**. **2.** A tubular waveguide in which wave propagation is concentrated.

wave envelope The outline described by the various amplitude peaks of the cycles in an amplitude-modulated wave. The envelope frequency is equal to the modulating frequency.

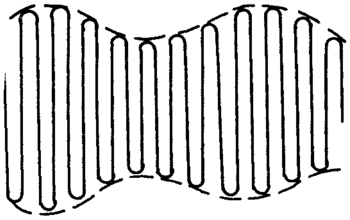

wave envelope

wave equation A second-degree partial differential equation whose solution describes wave phenomena.

wave filter A circuit or device that offers different amounts of attenuation to signals of different frequencies. See BANDPASS FILTER, BAND-SUPPRESSION FILTER, HIGH-PASS FILTER, and LOW-PASS FILTER.

waveform The shape of a wave described in terms of its resemblance to some well-known figure or to its conformity to the curve of the applicable wave equation (e.g., sinusoidal, square, sawtooth, cosine, rectangular, and triangular).

waveform-amplitude distortion See AMPLITUDE DISTORTION.

waveform analyzer See WAVE ANALYZER.

waveform converter A circuit or device for changing a signal of one waveform (such as a sine wave) into one of another waveform (such as a pulse or square wave).

waveform distortion The malfunction evidenced by a change of the waveshape of a signal passing through a circuit.

waveform error In a quantity displayed by an alternating-current test instrument, an error caused by the waveform of the measured signal. Thus, a voltmeter calibrated with a sine-wave voltage is subject to error when a measured signal is nonsinusoidal. Also called *waveform effect.*

waveform generator See FUNCTION GENERATOR.

waveform influence See WAVEFORM ERROR.

waveform monitor In television operations, an oscilloscope that continuously displays the video waveform.

waveform synthesizer A variable-frequency signal generator that allows the tailoring of waveshape to suit individual applications. A *function generator* (see FUNCTION GENERATOR, **1**) is such an instrument, but it usually provides only a choice of common waveshapes.

wave front For a radio wave, the plane that is parallel to the perpendicular electric and magnetic lines of flux. The wave propagates at right angles to this plane.

wave function A point function in a wave equation, specifying wave amplitude.

wave group The resultant of several different-frequency waves traveling over a common path.

waveguide A transmission line used at ultra-high and microwave radio frequencies. It is a hollow metal pipe with a rectangular or circular cross section. A rectangular waveguide must have sides measuring at least 0.5 wavelength and preferably more than 0.7 wavelength. A circular waveguide should be at least 0.6 wavelength in diameter, and preferably 0.7 wavelength or more. The characteristic impedance (Z_0) varies with frequency. In this sense, it differs from coaxial or parallel-wire lines, whose Z_0 values are independent of frequency.

waveguide apparatus See WAVEGUIDE COMPONENTS.

waveguide attenuator A device, such as an interposed energy-absorbing plate, for signal attenuation in a waveguide.

waveguide choke flange A waveguide flange that presents no impedance to the signal, and which need not be metallic for continuity.

waveguide component A device adapted for connection to, or insertion into, a waveguide system. Such components include waveguide parts and accessories (e.g., splicing hardware, attenuators, loads, wavemeters, etc.).

waveguide connector A fitting for joining waveguides for the efficient propagation of a signal.

waveguide coupling See WAVEGUIDE CONNECTOR.

waveguide critical dimension The cross-sectional dimension that determines the cutoff frequency for a waveguide.

waveguide cutoff In a waveguide, the highest or lowest frequency that can be propagated with less than a specified amount of attenuation per unit length.

waveguide directional coupler A directional coupler made of two parallel waveguides with a common wall. Two slots cut in the wall allow part of the microwave energy propagated in one direction in the main waveguide to be extracted, and energy traveling in the opposite direction to be rejected.

waveguide dummy load A waveguide section that dissipates the microwave energy entering it.

waveguide elbow **1.** A curved bend in a waveguide. **2.** A waveguide connector with a bend.

waveguide flange A flat, liplike fitting at the end of the pipe of a waveguide. It fastens waveguide sections together or attaches a waveguide

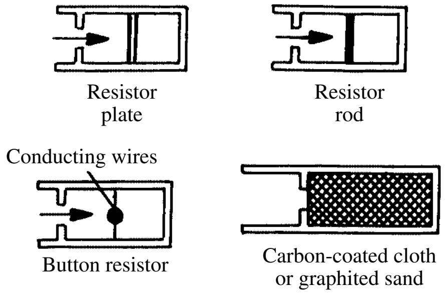

waveguide dummy load

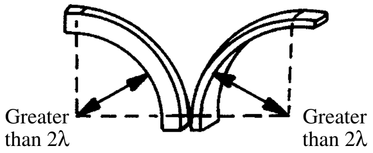

waveguide elbow

component, equipped with an identical flange, to the end of a waveguide.

waveguide frequency meter See WAVEGUIDE WAVEMETER.

waveguide gasket A gasket that provides electrical continuity between mating waveguide sections.

waveguide grating An array of wires mounted inside a waveguide, and that passes signals at some frequencies while obstructing others.

waveguide impedance Where the power P is known, and voltage E and current I are defined, with respect to a type of wave and waveguide, the value is either E^2/P or P/I^2.

waveguide junction A fitting that allows one waveguide section to be joined at an angle to another section. See, for example, WAVEGUIDE TEE and WAVEGUIDE WYE.

waveguide lens A microwave lens consisting of waveguide sections that provide the required phase shifts.

waveguide load See WAVEGUIDE DUMMY LOAD.

waveguide mode The form of propagation indicated by the field pattern in a plane transverse to the direction in which energy is propagated through a waveguide. Common modes are TRANSVERSE ELECTRIC MODE, also called *TE mode*, and TRANSVERSE MAGNETIC MODE, also called *TM mode.*

waveguide mode suppressor A filter that suppresses undesired propagation modes in a waveguide.

waveguide phase shifter A shifter for adjusting the phase of waveguide output energy, with respect to input energy.

waveguide plunger A plunger-like device that reflects incident microwave energy in a waveguide.

waveguide post A transverse rod inside a waveguide that acts as a parallel susceptance.

waveguide probe A pickup probe (tip or loop, as required) for sampling the field inside a waveguide, or a similar injection probe for introducing energy into a waveguide. Also see WAVEGUIDE SLOTTED LINE.

waveguide propagation **1.** The transmission of microwave energy through a waveguide by successive reflections between the inner walls. **2.** Propagation of very-high-frequency (VHF), ultra-high-frequency (UHF), or microwave electromagnetic fields through an atmospheric duct (see DUCT, **1**), as if through a waveguide. **3.** Propagation of very-low-frequency (VLF) electromagnetic fields through a waveguide-like duct between the ionosphere and the earth's surface.

waveguide radiator An antenna consisting of an open-ended waveguide with or without a horn. It radiates microwave energy into space or to a reflector.

waveguide resonator A waveguide section used as a cavity resonator (see RESONANT CAVITY).

waveguide seal A protective cover for the end of a waveguide. The seal introduces very little microwave attenuation, while preventing entry of moisture and debris.

waveguide shim A thin, pliable metal sheet inserted between mating waveguide components for electrical continuity. Also see WAVEGUIDE GASKET.

waveguide shutter An adjustable mechanical barrier, such as a rotatable vane, inserted into a waveguide to block or divert microwave energy.

waveguide slotted line A section of waveguide having a slot that accommodates a movable probe or coupling element.

waveguide slug tuner A quarter-wave dielectric slug inserted into a waveguide so that its amount of penetration and position can be adjusted for tuning purposes.

waveguide stub A stub consisting of a waveguide section joined to a main waveguide at an angle and provided with a nondissipative termination.

waveguide stub tuner An adjustable piston in a waveguide stub for tuning purposes.

waveguide switch A switch consisting of a movable section of waveguide that can be positioned for coupling to one of several other waveguide sections; it thus passes the energy it receives to any of the other sections.

waveguide system Microwave "plumbing" consisting of waveguides, their fittings and accessories, and associated components (such as attenuators, loads, wavemeters, etc.).

waveguide taper A connector that is flared to allow coupling between two waveguide sections having different cross-sectional sizes.

waveguide tee In a waveguide assembly, a tee-shaped junction used to connect a section of waveguide in series or parallel with another section.

waveguide transformer A waveguide component that functions as an impedance transformer.

waveguide tuner In a waveguide system, an adjustable tuner providing impedance transformation.

waveguide twist A length of waveguide whose cross section is rotated around the longitudinal axis (e.g., from vertical to horizontal).

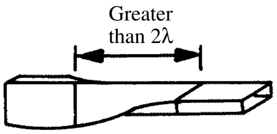

waveguide twist

waveguide wavelength In a uniform waveguide operating at a given frequency and in a particular mode, the distance between similar points for 360° phase shift.

waveguide wavemeter A waveguide component that acts as an absorption wavemeter or transmission wavemeter for identifying microwave frequencies.

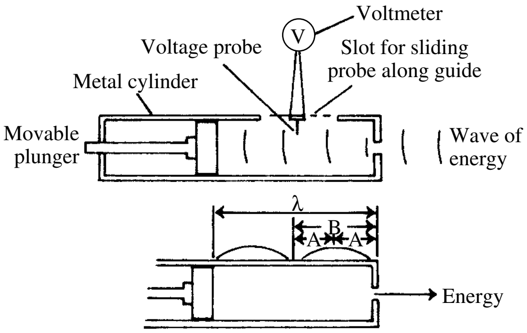

waveguide wavemeter

waveguide wedge See WEDGE, **1**.

waveguide window A thin metal opening mounted transversely inside a waveguide for impedance-matching purposes. The edges of the slit in a capacitive window are perpendicular to the electric field; in an inductive window, they are parallel to the electric field.

waveguide wye In a waveguide assembly, a wye-shaped junction for joining three waveguide sections.

wave heating Heating a material by energy absorbed from traveling electromagnetic waves.

wave interference Interaction between two or more waves, resulting in reinforcements and cancellations of energy.

wavelength Unit, meter. The displacement in one complete wave of an alternating or vibrating phenomenon, generally measured from crest to crest or from trough to trough of successive waves. For electromagnetic waves in free space, the wavelength in meters is equal to 3×10^8 divided by the frequency in Hertz. Also see WAVELENGTH-PERIOD-FREQUENCY RELATIONSHIPS.

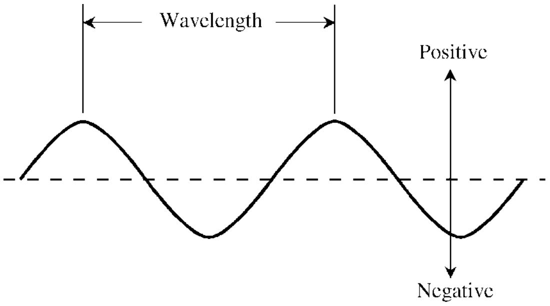

wavelength

wavelength constant The imaginary-number component of the propagation constant.

wavelength shifter **1.** A frequency shifter whose performance is indicated in units of wavelength, rather than in units of frequency. **2.** In certain photosensitive cells and tubes, a photofluorescent substance that raises the efficiency of the device by absorbing photons and then releasing ones of longer wavelength.

wave mechanics A theory of matter that views subatomic particles as complex wave patterns, and attempts to account for all physical processes in terms of wave phenomena.

wavemeter An instrument for measuring the wavelength or frequency of radio waves. One form consists of a series-resonant circuit containing an inductor, variable capacitor, and diode-type meter. The dial of the capacitor is calibrated to read in MHz. The inductor picks up energy from the radio-frequency source of unknown frequency, the capacitor is tuned for peak deflection of the meter, and the unknown frequency is read from the dial. This instrument is often called an *absorption wavemeter* because it absorbs a certain amount of power from the signal source under test. See also CAVITY WAVEMETER, COAXIAL WAVEMETER, LECHER WIRES, and SLOTTED LINE.

wave motion Undulating motion (e.g., up and down, and side to side). An electromagnetic wave has undulating electric and magnetic components that are both in phase and perpendicular to each other and to the direction of propagation of the wave.

wave normal **1.** The direction of propagation of an electromagnetic wave. **2.** A unit vector directed at a right angle to both the electric and magnetic lines of flux in an electromagnetic wave.

wave number The reciprocal of wavelength. This number denotes the number of waves per unit distance.

wave packet A short pulse composed of waves.

wave packets Radiant energy resulting from a number of wave trains of different wavelength.

wave path The line along which a WAVE TRAIN is propagated.

wave polarization The direction (horizontal or vertical) of wave undulations (i.e., the plane of the undulations, with respect to the direction of propagation). In general, a vertical antenna radiates a vertically polarized wave, and a horizontal antenna radiates a horizontally polarized wave.

wave propagation The movement of waves through space or through some medium. Electromagnetic waves travel through space at the speed of light (approximately 3×10^8 meters, or 186,000 miles, per second) and, like light, can be reflected and refracted.

wave reflection The reflection of electromagnetic waves by an obstruction, such as a solid body or a layer of the ionosphere. Compare WAVE ABSORPTION, WAVE POLARIZATION, and WAVE REFRACTION.

wave refraction Bending of the line of propagation of electromagnetic waves as they pass through various media, such as the troposphere or the ionosphere. Compare WAVE ABSORPTION, WAVE POLARIZATION, and WAVE REFLECTION.

waveshape The overall contour of a wave—especially as revealed by a curve plotted for the particular wave equation. Also see WAVEFORM.

waveshaping circuit A circuit that receives an input signal having a certain waveshape, and delivers an output signal having a different waveshape. For example, a squaring circuit converts a sine wave into a square wave at the same frequency.

wave surface See WAVE FRONT.

wave tail In a decaying pulse or signal envelope, the interval between the beginning of the decay and the point at which the amplitude reaches zero.

wave telegraphy See RADIOTELEGRAPHY.

wave telephony See RADIOTELEPHONY.

wave theory of matter A physical theory that the charge of an electron is distributed in space, rather than being focused at a point. Also see WAVE MECHANICS.

wave tilt A slight forward tilt of the electric flux lines in a radio wave radiated at the surface of the earth by a vertical antenna.

wave train A series of identical electromagnetic wave cycles propagated at equal intervals; an electromagnetic energy burst lasting at least several cycles.

wavetrap A resonant circuit consisting of an inductor and capacitor, either or both of which can be adjustable for tuning, used to remove (trap) a signal at the resonant frequency from a signal mixture.

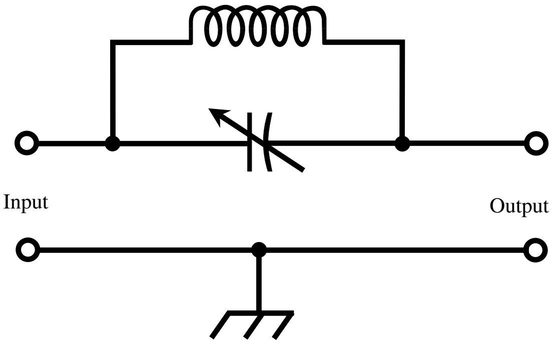

wavetrap

wave trough The minimum value of a wave envelope. Compare WAVE CREST.

wave velocity The distance per unit time traversed by a wave passing through a given medium.

wave winding See DRUM WINDING.

wa-wa pedal A foot-operated device used with an electronic musical instrument to produce a “wah-wah” sound fluctuation.

wax **1.** Any of a series of organic materials having important uses as dielectrics, impregnants, sealers, and lubricants in electronics. They are usually solid or semisolid, waterproof, and easily melted. **2.** In certain phonograph record discs, a blend of wax (see **1**, above) and metallic soaps. Also see WAX MASTER, **2**.

wax cake See WAX MASTER, **1**, **2**.

wax capacitor A fixed capacitor that has been dipped in or impregnated with a wax, such as halowax.

wax-dipped capacitor A fixed capacitor that has been dipped in a wax for sealing against moisture.

waxed paper See WAX PAPER.

wax-filled capacitor A fixed capacitor impregnated with a wax for enhancing the properties of its dielectric (usually paper) and sealing the capacitor unit.

wax master **1.** In disc-recording operations, the original recording made on a wax-surface disc. **2.** To make an original recording on a wax-surface disc.

wax original See WAX MASTER.

wax paper Wax-saturated paper used as a dielectric film in fixed capacitors and as an insulator.

way-operated circuit A single or duplex circuit shared by three or more party stations.

way point An important point selected on a radionavigational course line.

way station A station consisting of a teletypewriter connected at an intermediate point in a line (i.e., between, and in series with, other teletypewriter stations).

WAZ Amateur radio abbreviation of *Worked All Zones*, an award given to operators who have carried on verified two-way communication with stations in all communications zones of the world.

WB Abbreviation of *weather bureau.*

W_B Abbreviation of *base-region width* (in a transistor).

Wb Symbol for WEBER.

Wb/m^2 Abbreviation of *Webers per square meter* (see TESLA).

W_C Abbreviation of *collector-region width* (in a transistor).

WCEMA Abbreviation of *West Coast Electronic Manufacturers' Association.*

W/cm^2 Abbreviation of *watts per square centimeter.*

WE Abbreviation of *write enable.*

W_E Abbreviation of *emitter-region width* (in a transistor).

weak battery **1.** A battery that has been depleted to the point that its output (no-load or full-load) is too low to be useful. **2.** A battery specially designed for low-voltage output.

weak color Lack of color vividness or poor contrast between colors in a color-television picture. The condition is often caused by some malfunction in the chroma demodulator(s).

weak contrast In a television picture, poor differentiation of adjacent tonal areas.

weak coupling See LOOSE COUPLING.

weak current An extremely small electric current. The term is relative; generally, it refers to currents of a few microamperes or less.

weak magnet **1.** A magnet whose power has deteriorated considerably below a prescribed level. **2.** A body that normally is only slightly magnetic.

weak signal A signal whose amplitude is very low compared with that of signals considered satisfactory in a given application. Although the term is relative, it usually implies a signal that is noncompetitive with other signals in a given environment.

weak-signal detector A detector in which, at low input-signal amplitudes (weak-signal levels), the direct-current output is proportional to the square of the root-mean-square (rms) value of the input-signal voltage.

wearout The complete deterioration of a component or system (i.e., beyond restoration to useful service).

wearout failure Failure because of wearout, which can be predicted on the basis of known lifetime and the deterioration characteristics of components and equipment.

wearout point The instant of wearout, in terms of power output, watt-hour capacity, or some other specification.

weather antenna An antenna dimensioned for reception exclusively in the 162.4- to 162.55-MHz weather band. See WEATHER TRANSMISSION.

weathering Deterioration of electronic equipment as a result of exposure to outdoor heat, cold, moisture, wind, and similar conditions.

weather-protected machine A machine (usually a generator or motor) whose vent holes are designed to prevent entry of dust, water, and debris.

weather protection The coating, sealing, or treating of electronic equipment for protection against corrosion, humidity, and temperature changes in outdoor use.

weather satellite A satellite designed to photograph weather systems in infrared and/or visible light, and relay the pictures to earth via facsimile or television.

weather sonde See RADIOSONDE.

weather transmission The radio transmission of meteorological reports. Sometimes the transmissions are combined with guidance transmissions, from which they can be separated by means of a filter in the receiver.

weber Abbreviation, Wb. The SI unit of magnetic flux and of the magnetic flux quantum; $1 \text{ Wb} = 10^8 \text{ maxwells} = 1.257 \times 10^{-7}$ unit pole.

Weber-Fechner law The law expressing the relationship between a stimulus and the physiological reaction it produces: The sensation is proportional to the logarithm of the stimulus.

weber per square meter Symbol, Wb/m^2. See TESLA.

weber turn A unit of magnetic flux linkage equal to 10^8 maxwell turns.

wedge **1.** In a waveguide, a termination consisting of a tapered block or plate of carbon (or other dissipative material). **2.** In a television test pattern, convergent, equally spaced lines for checking resolution.

wedge bonding In integrated-circuit fabrication, a method of bonding in which a thermocompression bond (see COLD-COMPRESSION WELDING) is obtained through pressure from a wedge-shaped tool.

Wehnelt cathode An oxide-coated cathode in an electron tube.

Wehnelt cylinder In a cathode-ray tube, the cathode-enclosing cylinder that concentrates the electrons emitted by the cathode.

weight **1.** The amount of gravitational pull on a body or particle. **2.** Extra significance given to a term or value. See, for example, WEIGHTED TERM. **3.** The dot-to-space ratio in a Morse-code signal.

weight-density Symbol, *d.* Unit, kg/m^3. The weight per unit volume of a liquid, such as an electrolyte or insulating oil; also called *density.*

weighted distortion factor In the measurement of harmonic distortion, a factor whose use allows the harmonics in the complex waveform to be weighted in proportion to their relationship.

weighted noise level Unit, dBm. The noise level weighted with respect to the 70-dB equal-loudness contour of hearing.

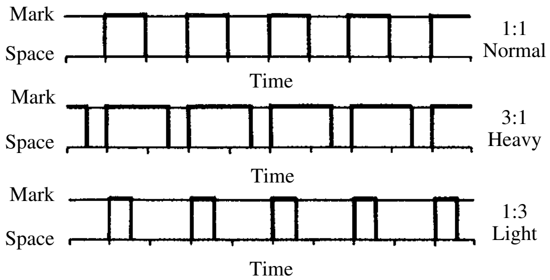

weight, 3

weighting **1.** Adjustment of a parameter to compensate for some imbalance in a system. **2.** Adjustment of the dot-to-space ratio in a Morse code signal. **3.** Adjustment of the mark-to-space ratio in a digital communications signal.

weighting filter A filter used in a communications network to represent the characteristics of the transmission passband.

weighting network A network that weighs differently (in a prescribed ratio) the frequency components appearing in an output signal by offering unequal attenuation to those frequencies.

weightlessness switch See ZERO-GRAVITY SWITCH.

Weir circuit In frequency-modulated signal transmission, a circuit used to stabilize the carrier wave. It compares the average carrier frequency with the frequency of a standard crystal oscillator, obtaining a direct-current compensating voltage (proportional to frequency deviation) that is applied to the frequency modulator. Also called *Weir stabilization circuit.*

Weiss constant In paramagnetism, a constant also known as *paramagnetic Curie temperature.* It can be positive or negative, depending on the particular paramagnetic material. It is important in defining the behavior of certain paramagnetic substances.

weld A strong bond of materials (usually metals) obtained by applying heat to areas to be joined while they are held or pressed together. No foreign metal is used, as is the case in brazing and soldering. The required heat is sometimes obtained by passing a high electric current through the materials.

welder **1.** An electrical device, often electronically controlled, for welding materials. **2.** A person who operates a device, as defined in **1**.

weldgate pulse In a welding device, the pulse that affects the arc current; therefore, it also affects the intensity of the heat produced by the device.

welding control An electronic system for controlling the interval during which current is passed through a workpiece in spot welding or seam welding. In this system, an electronic timer circuit determines the conduction time of thyratrons, ignitrons, or silicon-controlled rectifiers in the welding circuit. The control system can also regulate the welding current.

welding current The high electric current passed through a workpiece to produce the heat required for welding.

welding cycle The required sequence of steps (and the time required) in making a weld electronically.

welding time See WELD TIME.

welding transformer For electronic welding, a special very-high-current step-down transformer.

weld junction See WELD.

weld polarity The polarity of welding current. Some materials require a certain direction of current flow for a good weld.

weld time The interval during which welding current flows through the bodies to be bonded together.

well counter A radiation-counter setup in which a radioactive sample and detector are enclosed together in a thick-walled (usually lead) cylinder to minimize background count.

well-structured language An advanced form of high-level computer programming language. It is used in graphical and control applications.

Wenner element An adjustable, dual-slidewire balancing resistor used in constant-current, laboratory potentiometers to eliminate the necessity for sliding contacts in the measuring circuit.

Wenner winding A low-capacitance, low-inductance winding for high-frequency wirewound resistors in which the direction of the wire is reversed by looping alternate turns along the form.

Wertheim effect The tendency for a potential difference to develop between opposite ends of a length of wire, when the wire is placed parallel to magnetic lines of flux and rotated.

Western Union joint A strong splice of two wires made by tightly twisting a short portion of the tip of each wire along the body of the other. For increased ruggedness, the joint is often soldered. Also called *Western Union splice.*

Weston cell See STANDARD CELL.

Westrex system A system of sound recording in which signals from two separate microphone channels are recorded on opposite walls of a groove on a disc.

wet battery A battery of cells having a liquid electrolyte.

wet Liquid, especially pertaining to the electrolyte material in an electrochemical cell.

wet cell A battery cell having a liquid electrolyte. Compare DRY CELL.

wet-charged stand The length of time that a fully charged, wet storage cell can stand idle before its capacity drops by a specified amount.

wet electrolytic capacitor An electrolytic capacitor in which the electrolyte is a liquid. The leakage current in this type is higher than in the dry electrolytic, but it is self-healing after momentary voltage breakdown. Compare DRY ELECTROLYTIC CAPACITOR.

wet grab The adherence of a pressure-sensitive tape or sheet to a surface when very little pressure is used.

wet rectifier See ELECTROLYTIC RECTIFIER.

wet shelf life The specified shelf life of a discharged, wet storage cell. Compare DRY SHELF LIFE.

wetted-contact relay See MERCURY-WETTED REED RELAY.

wetting Applying a mercury coating to a contact surface.

wetting agent A substance (such as an alcohol or ester) that promotes the spreading and adhesion of a liquid or its absorption by a porous material.

WG Abbreviation of WIRE GAUGE.

WH Abbreviation of WATT-HOUR.

wh Abbreviation of WHITE.

Wheatstone bridge A four-arm balancing circuit (see BRIDGE), having resistors in each arm and used to measure an unknown resistance in terms of a standard resistance. The bridge supply is usually direct current, but alternating current can be used if all four resistances are nonreactive.

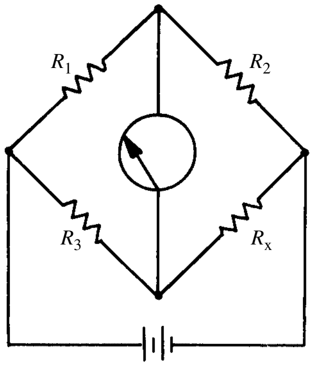

$R_x = R_3(R_2/R_1)$

Wheatstone bridge

wheel-drive locomotion The use of wheels for moving mobile apparatus such as robots. Offers simplicity and low cost. The main disadvantage is the inability to negotiate irregular terrain.

Wheeler's formula A formula for calculating the inductance of a multilayer air-core coil:

$$L = (0.8a^2n^2)/(6a + 9l + 10b)$$

where L is in microhenrys, a is the mean radius (inches) of the winding (axis to center of cross section), b is the depth of winding (inches), l is the length of winding (inches), and n is the number of turns.

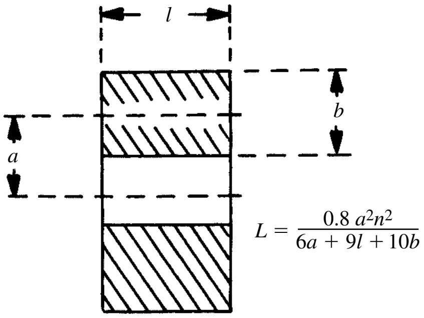

Wheeler's formula (dimensions illustrated)

wheel pattern A frequency test pattern produced on an oscilloscope screen by *z*-axis modulation of a circular trace. A sinusoidal axis signal produces a *gear-wheel pattern*, and a square-wave or pulse *z*-axis signal produces a *spot-wheel pattern.*

wheel printer A printout device for computers and calculators. It consists essentially of a rotating metal wheel around whose rim letters and numbers stand in relief. When the desired character comes into position, a hammer strikes it through the recording paper and carbon paper, printing the character on the recording paper. Also called *daisy-wheel printer.*

wheel static Static electricity (and the resulting radio interference) generated by friction between automobile tires and the road.

whiffletree switch In computer operations, a multiposition electronic switching circuit, so called from its circuit configuration, which resembles the contrivance used between a wagon and the horse team pulling it.

whip antenna A small-diameter, vertical rod (often telescoping) used as an antenna—especially in mobile communications, portable radio and television receivers, field-strength meters, etc.

whirl One of the circular, magnetic lines of flux around a straight wire carrying current.

whisker **1.** The pointer-wire electrode of a point-contact diode, point-contact transistor, or crystal detector. **2.** A slender filament of metal or ceramic, having high purity and high tensile strength.

whisker resistance In a semiconductor material, the resistance of a whisker component (see WHISKER, **1**).

Whiskey Phonetic alphabet code word for the letter W.

whistle A high-pitched tone (e.g., a beat note or acoustic feedback).

whistle filter A notch filter used to eliminate a whistle in an amplifier or other audio-frequency circuit. This filter can be of several versions, ranging from a simple resistance-capacitance

(RC) null circuit tuned to remove the offending frequency to a feedback-type band-suppression amplifier.

whistle interference The appearance of extraneous whistles in audio-frequency circuits. In some equipment, whistles result from oscillating amplifiers; in radio receivers, they are usually audible beat notes (heterodynes) produced by interfering carriers.

whistler A type of very-low-frequency (VLF) radio noise. They are thought to be caused by electromagnetic fields from distant lightning strokes, circulating in the earth's magnetic field. The name is derived from the peculiar sound the noise makes in a VLF radio receiver.

whistler mode propagation Radio transmission from the northern hemisphere to the southern hemisphere along the flux lines of earth's magnetic field.

whistlestop See WHISTLE FILTER.

whistling atmospheric See WHISTLER.

white The color that results from mixing all the additive primary colors: red, green, and blue.

white acid Hydrofluoric acid. Formula, HF. Used as an etchant, especially of glass.

white brass Brass that is more than 49% zinc.

white compression In television transmission, reduction of gain at highlight levels in the picture.

white-dot pattern In color-television tests with a dot generator, the one-white-dot pattern obtained when beam convergence has been secured.

white lamp See DAYLIGHT LAMP.

white level The lower-voltage point in a video signal, corresponding to full brilliance of the line on the screen (i.e., to the condition of whiteness in the picture).

white light See WHITE RADIATION, **2**.

white noise Random noise (acoustic or electric) equally distributed over a given frequency band, an example being the noise resulting from the random motion of free electrons in conductors and semiconductors.

white-noise generator A test device that generates electrical noise over a wide frequency spectrum. A simple type uses a reverse-biased silicon diode, the output of which is useful for testing audio amplifiers and radio receivers.

white-noise record A phonograph record containing recorded bands of white noise, each accompanied by voice announcements (e.g., instructions), and used to test the frequency response of a sound system.

white object A body that reflects and diffuses light of all wavelengths equally well.

white peak In a television picture signal, the maximum excursion in the white direction.

white radiation **1.** See WHITE NOISE. **2.** Visible light radiated with more or less equal intensity throughout the visible spectrum; seen as gray or white.

white raster See CHROMA-CLEAR RASTER.

white recording **1.** In amplitude modulation, recording characterized by a correspondence of maximum received power to minimum recording-medium density. **2.** In frequency modulation, recording characterized by a correspondence of lowest received frequency to minimum recording-medium density.

white room See CLEAN ROOM.

white saturation See WHITE COMPRESSION.

white signal The facsimile signal (see FACSIMILE) corresponding to scanning the copy area having maximum density.

white-to-black amplitude range **1.** At a point in a positive amplitude-modulated (AM) facsimile system (see FACSIMILE), the ratio (in dB) of signal current or voltage corresponding to white in the picture to that for black in the picture. **2.** At a point in a negative AM facsimile system, the ratio (in dB) of signal current or voltage corresponding to black in the picture to that for white in the picture.

white-to-black frequency swing At a point in a frequency-modulated facsimile system, the frequency difference $fw - fb$, where fb is the frequency corresponding to black in the picture, and fw is the frequency corresponding to white.

white transmission A system of picture or facsimile transmission in which the maximum copy darkness corresponds to the smallest amplitude (in an amplitude-modulated transmitter) or the highest instantaneous frequency (in a frequency-modulated transmitter). The opposite of BLACK TRANSMISSION.

white X radiation X rays of the continuous, or general, type.

whizzer An attachment that can improve the high-frequency reproduction in some audio loudspeakers.

whole-number division Arithmetic division (as in the division of binary numbers) in which the quotient is a whole number (i.e., division in which the divisor is contained in the dividend an integral number of times).

whorl See WHIRL.

WHP Abbreviation of *water horsepower.*

wick action **1.** The absorption of a liquid by, and its flow through, a cloth or thread, such as a lamp wick or lubricating wick. **2.** The flow of molten solder along and under the insulation of a wire.

wicking See WICK ACTION.

wide-angle diffusion A form of diffusion characterized by the wide-angle scattering of light; causes the source to have the same brightness at all viewing angles.

wide-area network A group of computers linked together, but separated by large geographic distances. Links can be made via telephone lines or radio.

wide-area service A teletype network that operates over long-distance wire lines.

Wide Area Telephone Service See WATS.

wideband **1.** Having a bandwidth greater than the minimum necessary to transmit a signal with acceptable intelligibility. **2.** For a voice signal, having a bandwidth greater than 6 kHz. **3.** Having the capability to operate, without adjustment, over a broad and continuous range of frequencies or wavelengths. **4.** In a digital network, a channel having a bandwidth of more than 64 kbps but less than 2 Mbps.

wideband amplifier An amplifier that exhibits reasonably flat response to a broad band of frequencies. The term is relative, depending on the application.

wideband antenna An antenna that transmits or receives signals over a broad frequency range, usually without the need for tuning.

wideband axis In a color-television signal, the direction of the fine chrominance primary phasor.

wideband communications **1.** Communications carried out over a band of frequencies wider than the minimum necessary for effective transfer of the information. **2.** A method of transmitting and receiving signals by deliberately varying the channel frequency over a wide range. Also called *spread-spectrum communications.*

wideband generator A signal generator covering a wide frequency range. Typical coverage in a laboratory-type instrument is 10 kHz to 1000 MHz.

wideband oscilloscope An oscilloscope whose horizontal, vertical, and sweep channels operate over a wide band of frequencies. Although the term wideband is relative, a wideband oscilloscope is usually assumed to be capable of displaying both radio and audio frequencies.

wideband ratio The ratio *B1*/*B2*, where *B1* is frequency bandwidth and *B2* is intelligence bandwidth.

wideband receiver A radio receiver that can tune in signals over a broad range of frequencies. An example is a communications receiver that can cover 10 kHz to 30 MHz continuously.

wideband repeater A repeater capable of operating over a wide range of input and output frequencies. Such repeaters are used in active communications satellites handling many different channels at the same time.

wideband signal generator See WIDEBAND GENERATOR.

wideband sweep **1.** In the operation of an oscilloscope, a repetitive sweep of the electron beam, the frequency of which is adjustable to any desired point within a wide range. The basic sweep rate in simple oscilloscopes is restricted to the audio-frequency spectrum (up to about 20 kHz), but a wideband sweep extends to much higher frequencies, typically several tens of MHz. **2.** A sweep circuit that produces the wideband sweep action described in **1**.

wideband test meter An alternating-current (ac) meter that can measure quantities over a wide frequency range in its basic form (i.e., without special external probes or converters). An example is an electronic ac voltmeter with a range extending from 10 Hz to 2.5 MHz.

wide-base diode A junction diode in which the p region is considerably wider than the n region.

wide-open **1.** Pertaining to wideband, untuned response. **2.** Pertaining to maximum-gain operation (e.g., a wide-open amplifier or receiver).

wide-range ammeter An ammeter that employs one or more shunt resistances to increase the full-scale deflection, usually by a power of 10 (10, 100, 1000, etc.). The resistor must be capable of carrying the current without burning out. Shunts are used when it is necessary to measure very large currents, such as hundreds of amperes. Shunts also allow a microammeter or milliammeter to be used as a multimeter with many current ranges.

wide-range reproduction High-fidelity audio-frequency reproduction.

width **1.** The horizontal dimension of a pulse, usually corresponding to its effective duration; also called PULSE DURATION. **2.** The horizontal dimension of an image, such as a television picture.

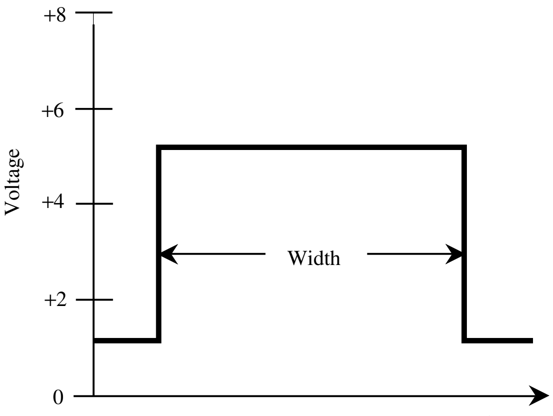

width, 1.

width coding The modification of pulse duration according to a code.

width coil See WIDTH CONTROL, **1**.

width control **1.** In a television receiver, the variable component for adjusting the swing of the horizontal deflection voltage and, therefore, the width of the picture. It is often a slug-tuned coil connected in parallel with a portion of the secondary winding of the horizontal output transformer. **2.** Sometimes, the horizontal gain control in an oscilloscope.

width mode In electronic-counter operations, a time-interval mode in which the signal is measured.

width modulation Modulation of the interval during which a gate circuit is open.

Wiedemann effect See DIRECT WIEDEMANN EFFECT.

Wiedemann-Franz law For metals that are good electrical conductors, the ratio of thermal conductivity to electrical conductivity is nearly constant, and is proportional to the absolute temperature.

Wien bridge A frequency-sensitive bridge in which two adjacent arms are resistances and the other two adjacent arms are resistance-capacitance (RC) combinations. One of the latter contains resistance and capacitance in series; the other contains resistance and capacitance in parallel. Because the bridge can be balanced at only one frequency at a time, it is useful as a simple audio frequency meter (see BRIDGE-TYPE AF METER). It is used also for capacitance and resistance measurements. When inductors are substituted for the capacitors, the Wien bridge can be used for inductance measurements (see WIEN INDUCTANCE BRIDGE).

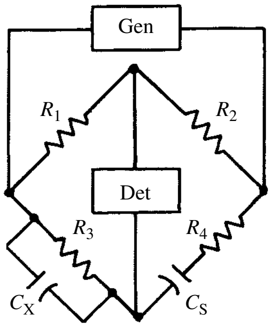

Wien bridge

Wien-bridge audio frequency meter See WIEN-BRIDGE FREQUENCY METER.

Wien-bridge distortion meter A distortion meter in which a Wien bridge circuit is used to remove the fundamental frequency from the complex waveform. The bridge is inserted between amplifier stages, and its notch response is sharpened by means of overall negative feedback.

Wien-bridge equivalent A resistance-capacitance null circuit, such as the parallel-tee (twin-tee) network, which has the same balance equation as the Wien bridge.

Wien-bridge filter A Wien bridge used as a band-suppression filter (notch circuit).

Wien-bridge frequency meter A bridge-type audio frequency meter in which the continuously variable frequency-selective circuit is a Wien bridge. Also see BRIDGE-TYPE AF METER.

Wien-bridge heterodyne eliminator A notch filter composed of a continuously variable Wien-bridge circuit (see WIEN BRIDGE). The device is connected in the audio-frequency channel of a radio receiver, and is tuned to remove a troublesome heterodyne whistle.

Wien-bridge oscillator An oscillator in which the frequency-selective feedback network is a Wien-bridge circuit (see WIEN BRIDGE). Also see BRIDGE-TYPE OSCILLATOR.

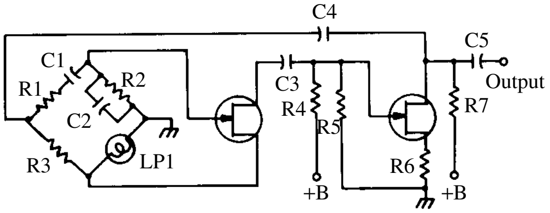

Wien-bridge oscillator

Wien capacitance bridge A Wien bridge arranged for the measurement of unknown capacitance.

Wien effect For an electrolytic substance, an increase in conductivity (decrease in resistivity) that occurs when a very large voltage is placed across the material. This voltage must be greater than approximately 1 megavolt per 50 cm.

Wien inductance bridge A Wien bridge containing inductors in place of capacitors and used for the measurement of unknown inductance.

Wien's displacement law The wavelength of maximum radiation of a black body is inversely proportional to the absolute temperature.

Wien's first law See WIEN'S DISPLACEMENT LAW.

Wien's laws See WIEN'S FIRST LAW, WIEN'S SECOND LAW, and WIEN'S THIRD LAW.

Wien's radiation law See WIEN'S SECOND LAW.

Wien's second law The emissive power of a black body is proportional to the fifth power of the absolute temperature.

Wien's third law An empirical law for the spectral distribution of energy radiated from a black body at a specified temperature. The distribution is a curve with a peak (maximum) at a wavelength that depends on the temperature. The higher the temperature, the shorter the wavelength at which maximum radiation occurs.

willemite See ZINC ORTHOSILICATE PHOSPHOR and ZINC SILICATE PHOSPHOR.

Williams tube A type of electrostatic cathode-ray storage tube.

Wilson chamber See WILSON CLOUD CHAMBER.

Wilson cloud chamber An airtight chamber containing water vapor or alcohol vapor at low pressure, and provided with a viewing window.

Radioactive particles penetrating the chamber are made visible as droplets form trails when the vapor condenses on them.

Wilson effect Electric polarization of a dielectric material being moved in a magnetic field.

Wilson electroscope A leaf-type electroscope (see ELECTROSCOPE) using a single leaf hanging vertically.

Wilson experiment An experiment demonstrating the WILSON EFFECT. It consists of rotating about its axis a hollow dielectric cylinder (whose inner and outer surfaces are metallized) in a magnetic field parallel to the axis. The result: An alternating voltage appears between the metallized surfaces.

Wimshurst machine A rotating machine used to produce high-voltage static electricity. The machine contains two glass disks, each having separate sectors of metal foil spaced around its face. The disks rotate in opposite directions, and the foil sectors passing each other form variable capacitors. Metal brushes pick up charges from the sectors passing under them and deliver this energy to one or more Leyden jars for storage.

wind charger A wind-driven generator used specifically to supply direct current for charging storage batteries.

wind-driven generator A dynamo-type generator, either stationary or mobile, powered by a windmill-like device.

wind gauge See ANEMOMETER.

winding **1.** A coil in an inductor or transformer (e.g., *primary winding*, *secondary winding*, *output winding*, etc.). **2.** A coil in a motor or generator.

winding arc The winding length of a coil, expressed in degrees.

winding cross section The cross section of a multilayer coil.

winding depth The depth of a multilayer coil, measured from the outermost layer to the innermost layer.

winding factor For a transformer or choke coil (or for a toroidal coil), the ratio of the total area of wire in the window to the window area.

winding length The length of a coil from the first turn to the last in a single-layer coil, or from the first turn to the last in one layer of a multilayer coil when all layers are identical.

winding space The window area of the core of a transformer or choke (see WINDOW, **3**).

wind loading **1.** The total wind pressure on an antenna system, generally measured in pounds or kilograms. The greater the wind speed, the greater the wind loading for a given antenna. The greater the exposed surface of the antenna, the greater the wind loading for a given wind speed. **2.** The highest wind speed, in miles per hour or meters per second, that an antenna system can safely withstand, assuming no accumulation of ice on the antenna structure.

Windom antenna **1.** A multiband antenna in which a single-wire feeder is attached to a horizontal half-wave radiator wire somewhat off center, at a point about ⅙ wavelength from one end of the radiator. The antenna operates with reasonable efficiency at the first several harmonics of the frequency at which it measures ½ wavelength. **2.** A similar antenna using a parallel-wire feeder.

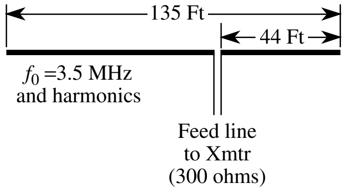

Windom antenna, 2

window **1.** Radar-interference material (see CHAFF). **2.** An interval during which a circuit is gated open to permit signal sampling. During this interval, a window is figuratively open to the signal. **3.** The open spaces between the legs of an iron core for a transformer or choke coil. **4.** An electromagnetic frequency band easily transmitted by earth's atmosphere. **5.** The period during which conditions are ideal for a complex operation, such as a rocket or spacecraft launch. **6.** An application space in a computer program that uses a graphical interface.

window area See WINDOW, **3**.

window comparator A comparator that detects voltage levels within a certain range of values, rather than simply indicating whether a voltage is more or less than a certain specified value.

window corridor An area where window (see WINDOW, **1**) has been dispersed.

window jamming The disturbance of electronic communications, especially radar, by dumping reflective material, such as metal foil, from an aircraft. Also see CHAFF; JAMMING; and WINDOW, **1**.

Windows Trade name (Microsoft) for a personal-computer interface scheme. The several versions all use selectable icons and menus.

window strip An insulated, flat, metal strip for bringing an antenna lead-in through a window. The window can be closed on the strip.

wind screen A foam covering that can minimize the roar caused by wind blowing against or across a microphone.

wind shield A radio-transparent cover placed over a radar antenna to protect it against damage from high winds. It is used especially in radar systems aboard aircraft.

wing In an antenna or other radiator, a (usually flat) member attached to, and sticking out from, another member. Its name is derived from its characteristic shape.

winterization The protective treatment and preparation of electronic equipment for winter storage or for operation in frigid regions.

wipeout Severe interference that obliterates all desired signals.

wipeout area An area in which WIPEOUT occurs.

wiper **1.** A thin metal blade or strip in sliding contact with a coil or other element over which it is turned to vary some quantity (such as resistance, inductance, voltage, current, etc.). **2.** A brush in a motor or generator.

wiper arm See WIPER, **1**.

wiper blade See WIPER, **1**.

wiping action The movement of contacts against each other when they slide as they mate or withdraw.

wiping contact A contact that makes and breaks with a sliding motion.

wire **1.** A metal strand or thread serving as a conductor of electricity. **2.** To connect wires between points in a circuit. **3.** See TELEGRAM. **4.** To send a TELEGRAM.

wire bonding **1.** The interconnection of components within a discrete package, by means of fine wire conductors welded to the individual components. **2.** A method of temporarily splicing the outer conductors of two coaxial cables. **3.** In general, any solderless method of splicing between two conductors.

wire broadcasting Broadcasting by means of carrier-current communication (see WIRED RADIO).

wire cloth A net of fine wire, used as an electrical shield when air circulation is necessary.

wire communication Communication carried on by means of signals transmitted over wire lines, as opposed to wireless (radio) communication.

wire control The control of remote devices by means of signals transmitted over wire lines, as opposed to radio control.

wire core A magnetic core consisting of a bundle of iron (or magnetic alloy) wires.

wired AND See DOT AND.

wired-in component A component to which wires are permanently connected in a circuit, as opposed to a plug-in component.

wired OR See DOT OR.

wired-program computer A computer in which instructions are wired-in by making appropriate connections to a panel with patch cords.

wired radio The transmission and reception of voice and other sounds, telegraph signals, pictures, control signals, telemetry signals, and the like by means of radio-frequency energy conducted by wire lines. Usually, the lines are used primarily for some other purpose (e.g., power lines or telephone lines).

wired-radio receiver The complete device or system that selects, amplifies, and demodulates or rectifies a radio signal picked up from wires that conduct it from a transmitting station.

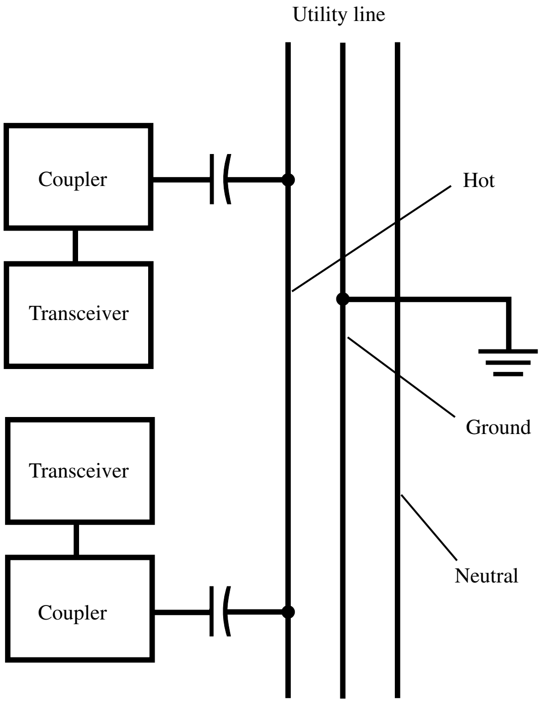

wired radio

wired-radio transmitter The complete device or system that generates radio-frequency power, adds signals (for communication, remote control, telemetry, etc.), and delivers the power to wires or to a cable for reception at a distant point.

wire drawing In the manufacture of wire, pulling metal through special dies to form wire of selected diameter.

wire dress The careful arrangement of wires in a chassis to ensure optimum performance. Also called *lead dress* (see DRESS).

wire duct A conduit through which wires are run. Such ducts have several shapes, but generally are rectangular.

wired wireless See WIRED RADIO.

wire-equivalent security See LEVEL-1 SECURITY.

wire fuse A fuse in which the fusible element is a wire of low-melting-point metal. Compare STRIP FUSE.

wire gauge **1.** A system for specifying the characteristics of wire. See AMERICAN WIRE GAUGE and BIRMINGHAM WIRE GAUGE. **2.** A wire number governed by diameter (e.g., 18-gauge wire for #18 wire). **3.** A tool or instrument for measuring wire diameter or for determining wire number.

wire gauze A fine screen of thin wires.

wire grating See WAVEGUIDE GRATING.

wire-guided Pertaining to guiding a machine, such as a robot, by means of signals sent by wire from the control point.

wire leads Leads that are (usually thin, flexible) wires, rather than pins, bars, strips, etc. An example is the wire leads of some transistors (as opposed to terminal pins).

wireless **1.** Pertaining to data communications and control systems that operate without wires (e.g., an infrared link between a notebook computer and a desktop computer). **2.** An early name for radio; it is still used in some countries. Sometimes for specificity, radio is referred to as *wireless telegraphy* or *wireless telephony.* See RADIOTELEPHONY.

wireless access protocol Acronym, WAP. A communications standard used on a worldwide basis with portable cellular telephone sets incorporating electronic mail, online service access, and/or Internet access.

wireless broadcaster See WIRELESS MICROPHONE.

wireless compass A radio compass (see DIRECTION FINDER).

wireless device A device that operates over a distance, without the use of interconnecting wires.

wireless e-mail **1.** The use of electronic mail (e-mail) in conjunction with a wireless Internet service. **2.** The use of a specialized portable wireless communications device, resembling a stripped-down computer or an enhanced cell phone, to send and receive e-mail messages.

wireless intercom An intercom using wired-radio for transmission and reception over the power line from which it is operated. Also see WIRED RADIO.

wireless Internet **1.** The use of a personal computer, usually a notebook or portable, in conjunction with a wireless modem connected to the Internet. Services of this type are provided with high-end cellular telephone subscriptions. Assets and limitations are similar to those associated with cell phone services. **2.** The use of a specialized portable or mobile wireless communications device, resembling a stripped-down computer or an enhanced cell phone, for basic Internet communications and information retrieval.

wireless microphone A device consisting of a small radio-frequency oscillator modulated by the microphone to which it is attached, and provided with a tiny antenna. The modulated signal is picked up and reproduced by a radio receiver.

wireless modem **1.** A modem that facilitates connection of a computer to the Internet using a cell phone set. **2.** A wireless transceiver that facilitates connection of a computer to the Internet. A cellular telephone connection can be included as an option. **3.** Also called *cordless modem.* A pair of short-range (on the order of a few feet) radio or infrared transceivers. One unit connects to a telephone jack or cable port, and the other unit connects to a computer. Can protect a portable computer against damage by lightning, assuming the computer uses batteries and is not connected to utility mains.

wireless tap A method of eavesdropping on communications in which a portion of the circuit makes use of a radio-frequency wireless link. An eavesdropping receiver can be positioned within range of the wireless transmitting antenna, and the signals intercepted. The existence of such a tap causes no change in the electronic characteristics of any equipment in the system, so its deployment is difficult to detect even if eavesdropping is anticipated.

wireless telegraphy See RADIOTELEGRAPHY.

wireless telephone **1.** A radio transceiver interconnected with the telephone lines. **2.** A telephone in which a short-range radio link replaces the cord between the receiver and the base unit. It normally has a range of several hundred feet.

wireless telephony See RADIOTELEPHONY.

wire line One or more wires or cables conducting currents for communication, control, or measuring purposes.

wire link A line for wire communication or wire control.

wire-link telemetry Telemetry carried on over a wire link.

wireman **1.** An electrician who specializes in the installation and servicing of electrical wiring. **2.** See LINEMAN.

wire mile A unit of measure equal to the product *ns*, where *n* is the number of separate, equal-length conductors in a line, and *s* is the length of a conductor in miles.

Wirephoto **1.** A system for transmitting and receiving photographs over wire lines. Also called *Telephoto.* See FACSIMILE. **2.** A photograph transmitted and/or received, as defined in **1**. **3.** A trademark for a photograph transmitted and received over telephone lines.

wire printer A printout device consisting of a wire array in which each wire is activated by an electromagnet. When a wire is selected (electrically), it is pushed against, and makes a dot on the recording paper. The letters and figures, then, are dot patterns. Compare WHEEL PRINTER.

wire radio See WIRED RADIO.

wire recorder In audio applications and data recording, a machine that records sounds or data pulses in the form of magnetized spots on a thin wire.

wiresonde A system for transmitting weather information data through a cable that holds a captive balloon. Compare RADIOSONDE.

wire splice A strong, low-resistance joint between (usually two) wires. See, for example, WESTERN UNION JOINT.

wire stripper A hand tool or power machine for cutting the insulating jacket on a wire and removing it without cutting or nicking the wire.

wiretap **1.** An instance of wiretapping. **2.** A wiretapping device.

wiretapper **1.** A device for wiretapping. **2.** A person who practices wiretapping.

wiretapping The act of making direct or indirect connections to a communications line to overhear or record a conversation.

wire telegraphy Telegraphic communication carried on by means of signals transmitted over wire lines, as opposed to wireless (radio) telegraphy.

wire telephony Voice communication carried on by means of signals transmitted over wire lines, as opposed to wireless (radio) telephony.

wire tie See TIE.

wire wave communication See WIRED RADIO.

wireways Hinged-cover metal troughs for containing and protecting wires and cables. Also see WIRE DUCT.

wirewound potentiometer A potentiometer in which the resistance element is a coil of resistance wire wound on a cord bent into a cylinder, or on a rigid, circular core.

wirewound resistor A resistor made from a coil of wire that is a poor conductor. The wire is usually wound around a cylindrical form; some components use toroidal forms. The resistance is determined by how well the wire metal conducts, by its diameter (gauge), and by its length. This type of resistor can be manufactured to precision tolerance. If the wire gauge is heavy, it can dissipate large amounts of power. The component has inductive reactance, making it unsuitable for use at radio frequencies. Compare CARBON-COMPOSITION RESISTOR, FILM RESISTOR.

wirewound rheostat A rheostat in which the resistance element is a coil of resistance wire wound on a card bent into a cylinder, or on a rigid, circular core.

wire wrap A method of circuit-board wiring in which components are interconnected by individual wire conductors, the ends of which are wrapped around terminal posts using a special tool.

wire-wrap connection An electrical connection made by tightly wrapping a bare wire around a special terminal.

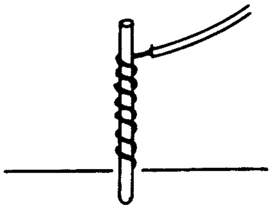

wire wrap

wire-wrapping tool A device used to make a wire-wrap connection.

wiring **1.** The system of wires or similar conductors connected between circuit components. **2.** Connecting and dressing such wires, as in wiring a circuit. **3.** Collectively, the connections (either actual wire or printed metal lines or processed semiconductor paths) between terminals and components in an electronic circuit. **4.** The process of installing or making such connections.

wiring board The control panel of a computer (i.e., a plugboard).

wiring capacitance Unavoidable capacitance between wires in a circuit, or between the wires and nearby metal bodies.

wiring connector A small (usually metal) fitting used to tie wires together.

wiring diagram See CIRCUIT DIAGRAM.

wiring impedance Unavoidable impedance in the wires in a circuit, and in associated terminals and hardware conducting alternating current. This impedance is the vector sum of WIRING REACTANCE and WIRING RESISTANCE.

wiring inductance Unavoidable inductance in the wires in a circuit, and in associated terminals and other hardware, through which alternating currents flow.

wiring reactance Unavoidable reactance in the wires in a circuit, and in associated terminals and other hardware, conducting alternating current. This reactance is caused by WIRING CAPACITANCE and/or WIRING INDUCTANCE.

wiring resistance Unavoidable resistance in the wires in a circuit and in associated terminals and other hardware.

Witka circuit A voltage-tripler circuit using only two diodes and two capacitors. Its no-load direct-current output voltage is approximately three times the peak value of the alternating-current input voltage. Also see VOLTAGE TRIPLER.

wk **1.** Abbreviation of WORK. **2.** Abbreviation of *week.*

WL **1.** Abbreviation of WAVELENGTH. **2.** Abbreviation of *waterline.*

WM Abbreviation of WATTMETER.

$\mathbf{Wm^2}$ Symbol for *watt square meter* (unit of the first radiation constant).

W/(MK) Symbol for *watts per meter kelvin* (unit of thermal conductivity).

WMO Abbreviation of *World Meteorological Organization.*

wobbulator A device that sweeps the frequency of an oscillator. There are several types, from a motor-driven, rotating, tank capacitor to sophisticated, fully electronic variable capacitors or variable inductors.

wolfram See TUNGSTEN.

Wolf's equation A sunspot-number equation used to forecast maximum usable frequency: $R = k(10G + N)$, where R is relative sunspot number, k is a constant for the telescope used, G is the number of sunspot groups observed, and *N* is the total number of sunspots.

Wollaston wire An extremely fine wire made by coating platinum with silver, drawing it thinner, and dissolving the silver coat.

womp A sudden surge in amplitude that causes a flare up of brightness in a television picture.

woodpecker A slang term used to describe the sound of the signal from over-the-horizon radar. Operating in the high-frequency part of the spectrum and changing in wavelength, roughly following the maximum usable frequency, the signal consists of repeated "spikes" that can interfere with routine communications.

woodpecker filter A special type of blanking filter used to quiet a shortwave receiver during strong, brief "spikes." It is similar to a noise blanker, but is designed to filter out the longer peaks characteristic of the over-the-horizon radar, known as the WOODPECKER.

Wood's alloy See WOOD'S METAL.

Wood's metal A low-melting-point (159.8°F) alloy containing bismuth (50%), cadmium (12.5%), lead (25%), and tin (12.5%). The alloy looks like lead and is used to mount rectifying crystals (such as galena) whose electrical sensitivity would be destroyed by the high temperatures required to melt most soft metals.

woofer A large loudspeaker, often a foot or more in diameter, designed specifically to reproduce very low (bass) audio frequencies. It is commonly used in high-fidelity stereo sound systems. Compare TWEETER.

word **1.** In computer operations, a group of bits or characters treated as a unit. **2.** In telegraphy, a data unit consisting of five characters plus one space.

word-address format A method of addressing a word by means of a single character, such as the first letter of the word.

word code **1.** A cipher in which word meanings are interchanged, rather than letter symbols. **2.** A word having an altered meaning in a cipher system, as defined in **1**. For example, "word" might mean "code."

word format The specific sequence of characters that forms a word of data.

word generator A special signal generator that delivers pulses in selected combinations corresponding to digital words (see WORD, **1**). It is used in testing computers and digital systems.

word length **1.** In computer operations, the number of bits in a word. **2.** In telegraphic communications (radio or wire), the average number of letters in a word.

word pattern The shortest meaningful word (see WORD, **1**, **2**) that can be recognized by a machine.

word processor **1.** An electronic device, similar to a typewriter, used for writing. Words, phrases, sentences, or paragraphs can be changed, replaced, or deleted prior to the final printout. **2.** A computer used for writing, as defined in **1**. **3.** Software that allows a computer to be used for writing, as defined in **1**.

word rate In a communications or computer system, the number of words per unit time (e.g., WORDS PER MINUTE).

word size See WORD LENGTH.

words per minute In telegraphy, a measure of data speed. It is approximately equal to the number of characters (including spaces) per minute divided by 6.

word time In computer operations, the time required to process one word that is in storage.

work **1.** Symbol, *W*. Units: joule, erg, foot-pound, kilogram-meter. That which is accomplished by the transfer of energy from one body to another, as when an exerted force causes a displacement. The amount of work performed is equal to force times distance: $W = Fd$. **2.** An amateur radio term meaning to engage in two-way communication with another station.

work area In computer operations, a temporary area of memory for data items being processed. Also called INTERMEDIATE STORAGE, WORKING MEMORY, and WORKING STORAGE.

work coil The alternating-current-carrying coil that induces energy in the workpiece in induction heating.

work envelope The range of motion over which a robot arm can move. It can be two-dimensional or three-dimensional.

work function Unit, eV. The energy required to bring an internal particle to the surface of a material and out into space, as when an electron is emitted by the hot cathode of a vacuum tube. The work function is the voltage required to extract 1 electrostatic unit of electricity from the material.

working data file A temporary accumulation of data sets that is erased or otherwise discarded after its transferal to another medium.

working life The expected or guaranteed lifetime of a material, device, or system in actual operation or use. Compare SHELF LIFE.

working memory See WORK AREA.

working point The operating point of an active device (i.e., the point along one of the characteristic curves around which operation is fixed).

working *Q* The *Q* of a loaded circuit or device (e.g., tank-circuit *Q* of a radio transmitter loaded with an antenna or dummy load).

working range **1.** The usable maximum distance between a transmitter and receiver in a wireless communications circuit. **2.** The allowed range over which specified parameters can vary in a particular system while facilitating normal operation, or operation within rated specifications.

working storage See WORK AREA.

working voltage The (usually maximum) voltage at which a circuit or device can be operated continuously with safety and reliability.

workout **1.** A dry run; a test. Example: checking a computer's processing power using a set of benchmark programs. **2.** An exacting test of equipment; a burn-in procedure.

workpiece The object heated by an induction or dielectric heater.

worm A cylindrical gear whose spiral "teeth" resemble a screw thread.

worm drive A mechanism for transferring motion from a tuning-knob shaft through a right angle to the shaft of an adjustable component, such as a potentiometer or variable capacitor. The knob turns a threaded shaft that mates with a gear wheel.

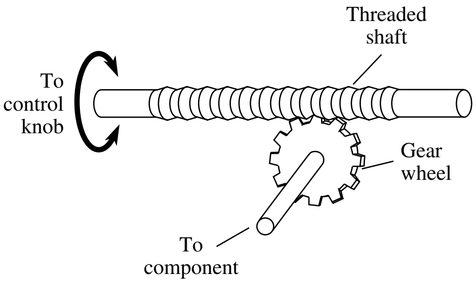

worm drive

worst-case circuit analysis Analysis that seeks the worst possible effects of variations in circuit parameters on circuit performance—especially variations in component characteristics.

worst-case noise pattern See DOUBLE-CHECKERBOARD PATTERN.

worst-case design The design of electronic equipment in such a way that normal operation is obtained—even though the characteristics of circuit components might vary widely.

woven resistor A resistance element made of strands of resistance material woven in the form of gauze.

wow Slow, periodic variations in the pitch of reproduced sound because of variations in the speed of the drive mechanism. Its name is derived from its characteristic sound. Compare FLUTTER.

wow meter An instrument that indicates the amount of wow produced by a turntable or other moving part.

Wpc Abbreviation of *watts per candle.*

wpm Abbreviation of *words per minute.*

wrap See WIRE-WRAP CONNECTION.

wrap-around **1.** The extent of curvature in magnetic tape passing over the heads during recording or playback. **2.** An enclosure that resembles a sheet (wrapped around a piece of electronic equipment). **3.** In computer and data-processing operations, a technique in which the display is cleared of data once it is filled (all available lines are occupied) and additional data is traced from top to bottom.

wrap joint See WIRE-WRAP CONNECTION.

wrapper A paper or tape wound around a component, such as a coil or capacitor, for insulation and protection.

wrapping **1.** See WRAPPER. **2.** Insulating a wire or other conductor by wrapping insulating tape around it.

wrap post A terminal post, tip, pin, or lug used in a wire-wrap connection.

wrap-up The (usually successful) completion of a design, fabrication, test, or investigation.

Wratten filter A light filter for separating colors. It is available in transparent sheets of various colors and is useful in photography and in several phases of electronics, including the operation of color meters and color matchers.

wrinkle finish A pattern of fine wrinkles created by special paint when it dries on a surface, such as that of a metal cabinet for a piece of electronic equipment.

wrist-force sensor In robotics, a set of strain gauges that detects the various forces in the joint connecting an end effector to an arm, and sends signals back to the robot controller. The controller can use the signals to direct the movements of the arm and end effector.

write **1.** In computer operations, to transfer data from one form of memory or storage to another form. Example: To transfer data in a computer from random-access memory (RAM) to the hard disk. Compare READ. **2.** To produce an image on the storage mesh in the cathode-ray tube of a storage oscilloscope.

write enable ring See WRITE PERMIT RING.

write gun See WRITING GUN.

write head In a magnetic memory or in a tape recorder or wire recorder used for recording data, the head that magnetizes the drum, tape, disk, or wire. Compare READ HEAD.

write pulse In computer operation, the pulse that causes information to be recorded in a magnetic cell, or sets it to the one-state. Compare READ PULSE.

write time The time taken to write data to a storage device.

writing gun In a storage oscilloscope, the electron gun that produces the image electronically on the storage mesh. It is mounted in the rear of the cathode-ray tube. Compare FLOOD GUN.

writing head See WRITE HEAD.

writing rate In photographing the image on a cathode-ray tube screen, the highest spot speed at which an acceptable picture can be made.

writing speed See WRITING RATE.

writing telegraph See TELAUTOGRAPH.

wrong color The instance of an undesired color (e.g., purple instead of red) in a color-television picture. The condition is often caused by a malfunction of the chroma demodulator(s).

W/(sr•m^2) Abbreviation of *watts per steradian square meter.*

WT **1.** Abbreviation of wireless telegraphy. **2.** Abbreviation of WATERTIGHT.

wt Abbreviation of WEIGHT.

WUI Abbreviation of *Western Union International.*

Wulf electrometer See BIFILAR ELECTROMETER.

Wullenweber antenna An electronically steerable antenna composed of two concentric circular arrays of masts connected to the steering circuitry.

WUX Abbreviation of *Western Union telegram.*

WVdc Abbreviation of *direct-current working voltage:* the maximum continuous direct-current voltage that can safely be placed across a component.

ww Abbreviation of *wirewound.*

W/sr Abbreviation of *watts per steradian* (unit of radiant intensity).

WWV The call letters of a standard-frequency/standard-time broadcasting station operated by the National Bureau of Standards and located in the continental United States. Also see WWVH.

WWVH The call letters of a standard-frequency/standard-time broadcasting station operated by the National Bureau of Standards and located in Hawaii. Also see WWV.

wx Radiotelegraph abbreviation of weather.

WXD International Telecommunications Union symbol for *meteorological radar station.*

WXR International Telecommunications Union symbol for *radiosonde station.*

wye adapter A connector that provides two outputs for a single input, or vice versa. It is commonly used in audio applications.

wye box In a three-phase power-measuring setup, a special arrangement of two impedances, each of which is equal to the impedance of the potential element of the wattmeter used in the setup. The box permits a single wattmeter to be used, which indicates one-third the total power. Without the box, three wattmeters would be needed.

wye connection A method of connecting three windings in a three-phase system so that one terminal of each winding is connected to the neutral point. It is shaped like the letter Y. Also called *star connection.*

wye current Current through one of the branches of a three-phase star (wye) connection.

wye delta starter A starter circuit for a three-phase squirrel-cage induction motor. When the starter switch is thrown in one direction, the stator of the motor is wye-connected for starting; thrown in the opposite direction, the stator is delta-connected for continued operation.

wye equivalent circuit A wye-connected three-phase circuit equivalent to a delta-connected circuit when the impedance of any pair of corresponding lines for both wye and delta are the same; the third line is unconnected.

wye junction See WAVEGUIDE WYE.

wye-matched-impedance antenna An antenna in which the impedance of a nonresonant open-wire feed line is matched to that of the center of the radiator by spreading the end of the feeder wires out into a Y-shaped or delta-shaped matching section. Also called DELTA-MATCHED-IMPEDANCE ANTENNA.

wye point The star point in a three-phase system. Also see WYE CONNECTION.

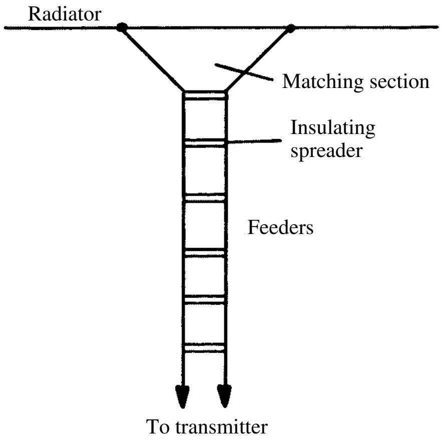

wye-matched-impedance antenna

wye potential In a three-phase armature, the voltage between a terminal and the neutral point.

wye rectifier A three-phase rectifier circuit in which the transformer or generator windings are arranged in a wye connection.

wye-wye circuit A circuit consisting a wye-connected generator and a wye-connected load.

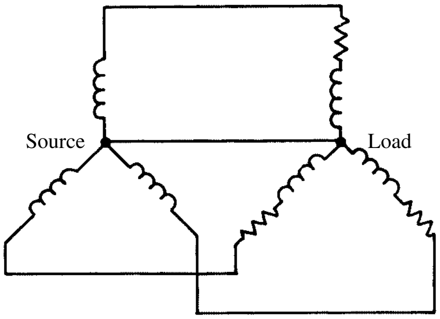

wye-wye circuit

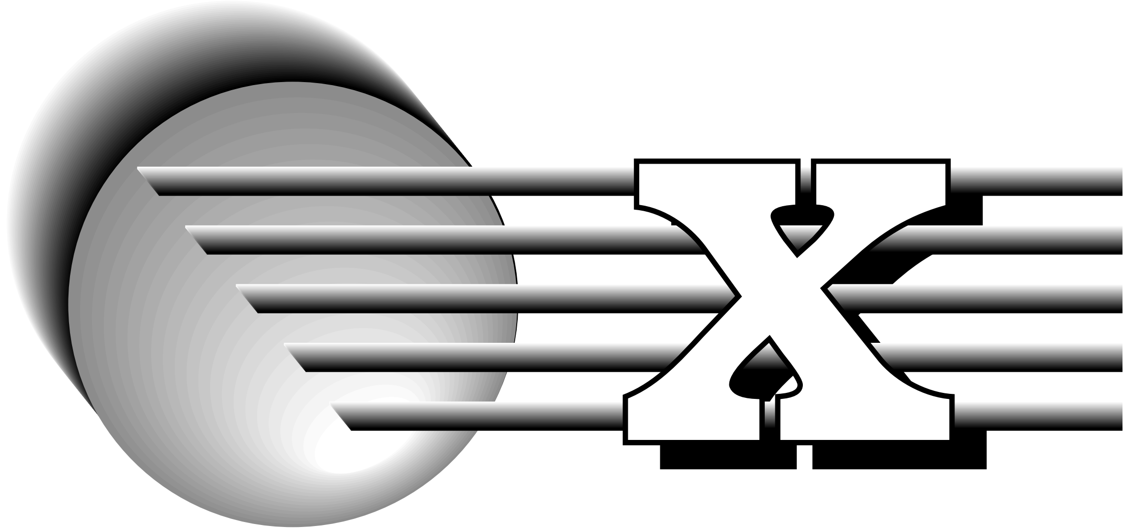

X **1.** Symbol for REACTANCE. **2.** Symbol for *no connection.* (Also, NC.) **3.** Symbol for the HORIZONTAL AXIS of a graph or screen in the rectangular (Cartesian) coordinate system. **4.** Symbol for an unknown quantity.

x **1.** Symbol for number of carriers drawn from collector to base of a transistor, for each carrier collected. **2.** Symbol for an unknown quantity. **3.** Symbol for the HORIZONTAL AXIS of a graph or screen in the rectangular (Cartesian) coordinate system.

x amplifier The horizontal amplifier of an oscilloscope or recorder. Compare Y AMPLIFIER and Z AMPLIFIER.

X and Z demodulation Color television demodulation based on the 60-degree difference between the two reinserted 3.58-MHz subcarrier signals. The R-Y, B-Y, and G-Y voltages derived from the demodulated signals control the three guns of the picture tube. Compare Q MODULATION.

x-axis **1.** The horizontal axis of a chart, graph, or screen in the rectangular (Cartesian) coordinate system. **2.** In a quartz crystal, the axis drawn through the corners of the hexagon.

x-axis amplifier See X AMPLIFIER.

XB Abbreviation of *crossbar.*

X balance The reactance balance (either the variable component or the adjustment of it) in an impedance bridge in which separate resistance (*R*) and reactance (*X*) balancing is provided.

X band The frequency band extending from 5.2 to 11 GHz.

X bar A rectangular, piezoelectric quartz bar cut from a Z-section, whose faces are parallel to the X-axis and whose edges are parallel to the X-, Y-, and Z-axes. See CRYSTAL CUTS.

X bridge An alternating-current bridge for measuring reactance.

X_C Symbol for CAPACITIVE REACTANCE.

X channel The horizontal channel of an oscilloscope or recorder. Compare Y CHANNEL and Z CHANNEL.

X-channel gain See X GAIN.

X component In a complex impedance, the reactive component (either inductive or capacitive).

x coordinate See ABSCISSA.

X-cut crystal A piezoelectric plate cut with its faces perpendicular to the X-axis of a quartz crystal. See CRYSTAL CUTS.

xcvr Abbreviation of TRANSCEIVER. Often capitalized.

X deflection Horizontal deflection of the spot on the screen of a cathode-ray tube. Compare Y DEFLECTION.

X diode The decoding diode in each of the X lines of a memory matrix. Compare Y DIODE.

X direction The horizontal direction in deflections and in graphical presentations of data.

X drive The driving source or energy for the X lines of a computer memory matrix.

xducer Abbreviation of TRANSDUCER.

Xe Symbol for XENON.

xenon Symbol, Xe. An inert gaseous element. Atomic number, 54. Atomic weight, 131.29. Xenon is present in trace amounts in the earth's atmosphere and is used in some thyratrons, electric lamps, and lasers.

xenon tube A flash tube filled with xenon.

xerography A system for the reproduction of printed matter, drawings, and other graphic matter. It is an electrostatic process in which a charged photoconductive surface is exposed to an

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

image of the material to be copied. A latent image is formed on the surface and is developed (by dusting with black powder attracted to the charged image) to make the image visible.

xeroradiography Xerography using X RAYS.

xerothermic A condition characterized by both heat and dryness.

Xerox **1.** Name of a company that manufactures a wide variety of office and computer equipment. **2.** Trade name for a XEROGRAPHY machine. **3.** The reproduction obtained by means of XEROGRAPHY (generic term). **4.** To make a xerographic reproduction (generic term).

X factor **1.** An unknown or unidentifiable quantity or parameter. **2.** See X COMPONENT.

xformer Abbreviation of TRANSFORMER.

X gain The gain (or gain control) of the horizontal channel of an oscilloscope or X-Y recorder. Compare Y GAIN and Z GAIN.

X-H array See LAZY-H ANTENNA.

XHV Abbreviation of EXTREMELY HIGH VACUUM.

x intercept The *x* coordinate of the point at which a line or plane intersects the X-AXIS.

x irradiate To expose to X RAYS.

xistor Abbreviation of TRANSISTOR.

X_L Symbol for INDUCTIVE REACTANCE.

X line A horizontal line in a memory matrix. Compare Y LINE.

XLR connector A microphone connector with a locking device to prevent unintentional disconnection. It has three pins and is commonly used with balanced audio systems.

X meter **1.** An instrument for measuring reactance and phase angle. **2.** A form of simple capacitance meter whose dial is direct-reading in microfarads, although the deflection is actually proportional to the reactance of the capacitor under measurement.

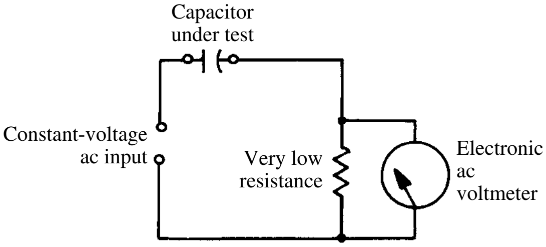

X meter, 2

xmission Abbreviation of *transmission.*

xmit Abbreviation of *transmit.* Also, xmt.

xmitter Abbreviation of TRANSMITTER. Also, xmtr.

XMODEM In data transmission, an error-correcting mode in which data is sent in blocks of 128 kilobytes (128K), or 131,072 bytes. The source and destination tally up the bytes in each block. If they agree, the next data block is sent. If they do not agree, the current data block is retransmitted. Compare YMODEM and ZMODEM.

xmt Abbreviation of *transmit.* Also, xmit.

xmtr Abbreviation of TRANSMITTER. Also, xmitter.

XOR Abbreviation of EXCLUSIVE-OR.

xover Abbreviation of CROSSOVER.

X particle See MESON.

xponder Abbreviation of TRANSPONDER.

X position **1.** The alignment of a cathode-ray beam along the horizontal axis of an oscilloscope screen. **2.** The position of a point, with respect to the horizontal axis of a graph in the rectangular (Cartesian) coordinate system.

XR Abbreviation of INDEX REGISTER.

X radiation Electromagnetic energy in the form of X rays.

X-ray **1.** Phonetic alphabet communications code word for the letter X. **2.** Pertaining to, or consisting of, X rays.

X-ray astronomy The science of observing celestial objects and the sky in the X-ray band of wavelengths. Generally, this must be done from above the earth's atmosphere because the atmosphere absorbs X rays.

X-ray crystallography The science of observing atomic patterns in a crystal by means of X rays.

X-ray detecting device **1.** An X-ray instrument for spotting flaws in solid bodies. **2.** A device, such as a fluoroscope, for showing the presence of X rays.

X-ray diagnosis The use of X rays in the diagnosis of disease and in the observation of internal parts of the body.

X-ray diffraction The diffraction of X rays by a material into or onto which they are directed. Also see X-RAY DIFFRACTION CAMERA and X-RAY DIFFRACTION PATTERN.

X-ray diffraction camera A special camera that furnishes a photograph of the pattern created by X rays diffracted by a material. See X-RAY DIFFRACTION and X-RAY DIFFRACTION PATTERN.

X-ray diffraction pattern The pattern produced on film exposed to X rays diffracted by a material. Also see X-RAY DIFFRACTION and X-RAY DIFFRACTION CAMERA.

X-ray goniometer An instrument used to find the position of the axes of a quartz crystal. It uses X rays reflected from the atomic planes of the crystal.

X-ray inspection The use of X rays in the examination and study of the internal features of materials and devices.

X-ray laser A laser designed to emit coherent X rays in a narrow beam.

X-ray load **1.** An X-ray tube that is the terminal member of an electronic system. **2.** A body, mass, or material exposed to X rays.

X-ray machine A device that generates X rays for a specific purpose, such as medical diagnosis.

X-ray photograph A photograph made by exposure to X rays, especially an *X-ray shadowgram,* a picture made without a camera by interposing an object (such as a part of the human body) between an X-ray tube and photographic film.

X-ray radiation See X RADIATION.

X rays Invisible, electromagnetic radiation having wavelengths ranging from approximately 0.01 nanometer (10^{-11} meter) to 0.15 nanometer (1.5×10^{-10} meter). These waves are shorter than ultraviolet, but longer than gamma rays. They can be produced by bombarding a target of heavy metal (such as tungsten) with a stream of high-speed electrons in a vacuum tube. X rays have high penetrating power, can expose photographic film, and cause some substances to fluoresce. Also see X-RAY DIAGNOSIS, X-RAY INSPECTION, and X-RAY THERAPY.

X-ray spectra The continuous band of ionizing electromagnetic radiation having wavelengths ranging from approximately 0.01 nanometers to 0.15 nanometers.

X-ray spectrograph A spectrograph used to disperse and measure the wavelength of X rays.

X-ray spectrometer A spectrometer with which the diffraction angle of X rays reflected from the surface of a crystal can be measured. The device allows the characteristics and composition of almost any material to be studied.

X-ray star A collapsed star that emits high-intensity energy concentrated mainly in the X-ray region of the electromagnetic spectrum.

X-ray system Collectively, an X-ray tube and the associated equipment required for a specific application, such as crystallography, irradiation, or medical therapy.

X-ray technician **1.** A professional skilled in the operation and maintenance of X-ray systems, who usually works under the supervision of an engineer. **2.** A professional skilled in the medical use of X rays, who works under the supervision of a medical doctor.

X-ray therapy The use of X rays in the treatment of certain physiological diseases.

X-ray therapy system An X-ray system designed for X-ray therapy.

X-ray thickness gauge An instrument used to measure the thickness of metal, such as a continuously moving sheet of steel. The measurement is in terms of the amount of absorption of an applied X-ray beam by the metal.

X-ray tube A specialized, very-high-voltage diode in which a high-speed electron beam bombards an anode (target) of heavy metal, such as tungsten, causing it to emit X rays. The target is tilted to reflect the X rays through the glass wall of the tube.

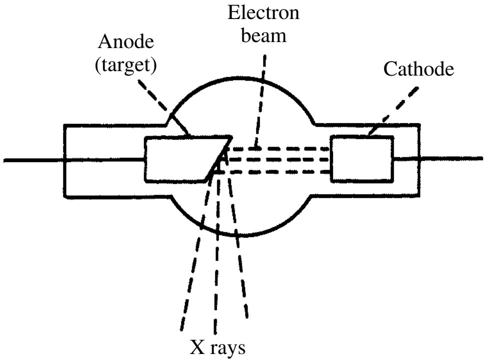

X-ray tube

X section Abbreviation of *cross section.*

X sink The circuit or device into which the X lines of a memory matrix feed. Compare Y SINK.

X_T Symbol for total reactance.

xtal Abbreviation of CRYSTAL.

xtalk Abbreviation of CROSSTALK.

x unit A small unit of length, equal to approximately 10^{-13} meter or 10^{-4} nanometer. Now seldom used, the x unit was once often used in the expression of ultraviolet and X-ray wavelengths.

X wave One (the extraordinary) of the pair of components into which an ionospheric radio wave is divided by the earth's magnetic field. Compare O-WAVE.

XXX In radiotelegraphy, a signal meaning *urgent.*

X-Y counting A technique used with electronic counters to determine the ratio of one frequency (X) to another (Y). When X is the higher frequency, the counter readout shows the number of pulses of X producing one cycle of Y.

XY-cut crystal A piezoelectric plate cut from a quartz crystal at such an angle that its electrical characteristics fall between those of the X-cut and Y-cut crystal.

x-y plotter An output device similar to the x-y recorder. It allows a digital computer to plot graphs. Also called *data plotter.*

x-y recorder An instrument that produces a permanent record (photographic print or directly inked paper) of a variable quantity on a chart having Cartesian (rectangular) coordinates.

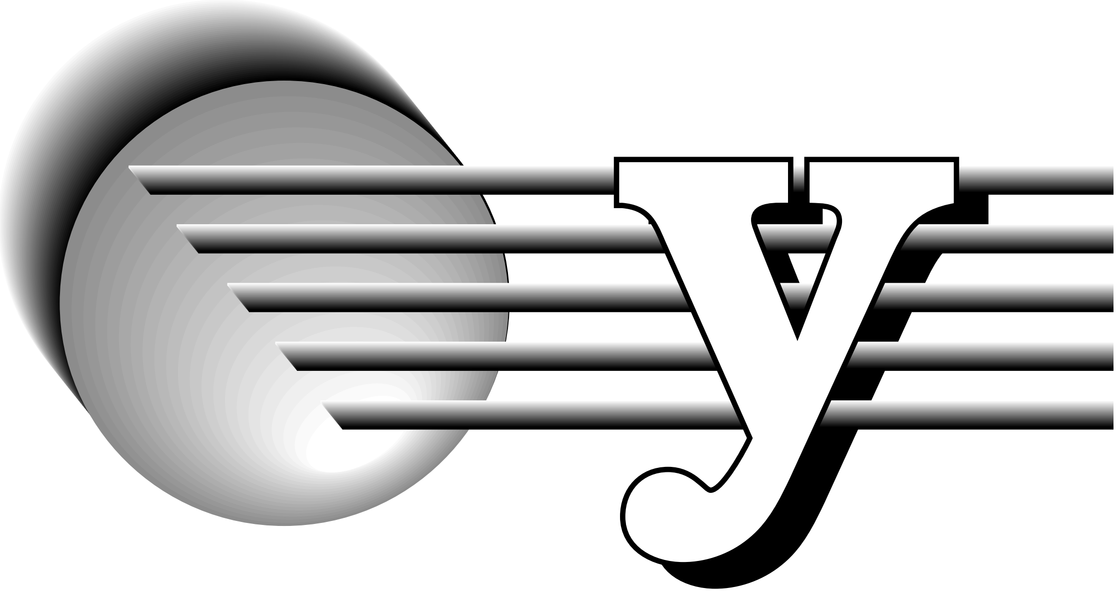

Y **1.** Symbol for ADMITTANCE. **2.** Symbol for YTTRIUM. **3.** Symbol for YOUNG'S MODULUS.

y **1.** Abbreviation of YEAR. (Also, yr.) **2.** Abbreviation of YARD. (Also, yd.) **3.** Symbol for the vertical axis of a graph or screen in the rectangular (Cartesian) coordinate system.

Y adapter See WYE ADAPTER.

YAG Abbreviation of *yttrium-aluminum-garnet*, the stimulated substance in some lasers.

Yagi antenna Also called *Yagi-Uda array* or *beam.* An antenna consisting of two or more parallel, straight elements, including at least one parasitic element and at least one driven element. The elements all lie in the same plane. Each driven element is connected to the feed line, half-wave resonant, and center-fed. A two-element array can be formed by adding a director or a reflector alongside a single driven element. An array with one director and one reflector, along with the driven element, increases the gain and directivity compared with a two-element system. The gain and directivity increase further as elements are added. This is usually done by placing additional directors in front of a three-element array. Compare QUAD ANTENNA.

y amplifier The vertical amplifier of an oscilloscope or recorder. Compare X AMPLIFIER and **Z** AMPLIFIER.

Yankee Phonetic alphabet communications code word for the letter Y.

yard Abbreviation, yd. A unit of linear measure in the English system. 1 yd = 3 ft = 36 inches = 0.9144 meter.

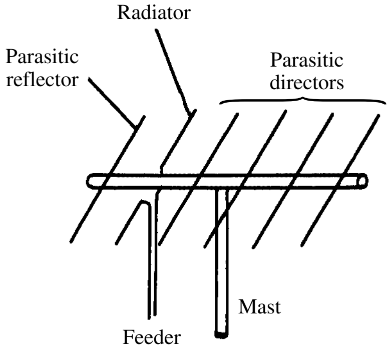

yagi antenna

yaw Side-to-side movement in a vehicle or robotic end effector. Essentially horizontal displacement about the vertical axis.

yaw amplifier In an aircraft servo system, the unit that amplifies the yaw signal (the signal proportional to the deviation of the aircraft from the line of flight).

yaw meter An instrument for measuring the angle of yaw of an aircraft.

y-axis **1.** The vertical axis of a chart, graph, or screen in the rectangular coordinate system.

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

2. In a quartz crystal, the axis drawn perpendicular to the faces of the hexagon.

y-axis amplifier See Y AMPLIFIER.

Yb Symbol for YTTERBIUM.

Y box See WYE BOX.

Y channel See Y AMPLIFIER.

Y-channel gain See Y GAIN.

Y circulator An interconnection among three waveguides. When power is applied to the junction through one waveguide, the wave is transferred to the adjacent waveguide immediately to the right or left.

Y connection See WYE CONNECTION.

y coordinate See ORDINATE.

Y current See WYE CURRENT.

Y-cut crystal A piezoelectric plate cut from a quartz crystal in such a way that the plane of the plate is perpendicular to the Y-axis of the crystal. See CRYSTAL AXES and Y-AXIS, **2**.

yd Abbreviation of YARD.

Y deflection Vertical deflection of the spot on the screen of a cathode-ray tube. Compare X DEFLECTION.

Y diode The decoding diode in each of the Y lines of a memory matrix. Compare X DIODE.

y direction The vertical direction in deflections and in graphical presentations of data.

Y drive The driving source or energy for the y lines of a computer memory matrix. Compare X DRIVE.

year Abbreviation, y or yr. The period of the earth's revolution around the sun, with respect to distant stars: approximately 365.25 days or 3.15576×10^7 seconds.

Yellow Book A specialized format for *compact-disk read-only memory (CD-ROM)* computer data storage media, developed by Sony and Philips. It is similar to RED BOOK, but it includes superior error correction and storage capacity. See also CD-ROM, GREEN BOOK, ORANGE BOOK, and RED BOOK.

yellow copper ore See CHALCOPYRITE.

yellow metal A copper-zinc alloy that is 60% copper.

yellow pyrites See CHALCOPYRITE.

Y-equivalent circuit See WYE-EQUIVALENT CIRCUIT.

Y gain The gain (or gain control) of the vertical channel of an oscilloscope or x-y recorder. Compare X GAIN and Z GAIN.

yield In a phototube or photocell, the photoelectric current per unit intensity of light.

yield map A diagram of an integrated circuit showing the locations of faulty components.

yield strength The lowest stress for plastic deformation of a material, below which the material is elastic and above which it is viscous.

yield value The amount of physical stress that causes a substance to become stretched permanently out of shape.

YIG Abbreviation of *yttrium-iron-garnet*, a crystalline material used in certain types of acoustic delay lines, parametric amplifiers, and filters.

YIG crystal A crystal of yttrium-iron-garnet. See YIG.

YIG filter A filter using a YIG crystal and tuned electromagnetically.

YIG-tuned parametric amplifier A parametric amplifier in which tuning is accomplished by continuously varying direct current through the solenoid of a YIG filter.

y intercept The y coordinate of the point at which a line or plane intersects the y-axis.

ylem The material from which the primordial fireball is thought to have been made, and from which the entire known universe is believed to have originated approximately 10 billion (10^{10}) years ago.

Y line A vertical line of a memory matrix. Compare X LINE.

Y-matched-impedance antenna See WYE-MATCHED-IMPEDANCE ANTENNA.

YMODEM In data transmission, an error-correcting mode in which data is sent in blocks of one megabyte (1MB), or 1,048,576 bytes. The source and destination tally up the bytes in each block. If they agree, the next data block is sent. If they do not agree, the current data block is retransmitted. Compare XMODEM and ZMODEM.

yoke **1.** The ferromagnetic ring or cylinder that holds the pole pieces of a dynamo-type generator and acts as part of the magnetic circuit. **2.** The system of coils used for magnetic deflection of the electron beam in cathode-ray tubes.

yoke method A method of measuring the permeability of a sample of magnetic material. It involves completing the magnetic circuit with a heavy yoke of soft iron.

Young's modulus Symbol, Y. The ratio of tensile stress to tensile strain.

Y parameters The admittance parameters of a four-terminal network or device.

Y point See WYE POINT.

Y position **1.** On a cathode-ray screen, the vertical position of the beam spot. **2.** On a graph, the position of a point along the vertical axis.

Y potential See WYE POTENTIAL.

yr Abbreviation of YEAR. Also, y.

Y rectifier See WYE RECTIFIER.

Y signal In color television, the monochromatic signal conveying brightness information.

Y sink The circuit or device into which the Y lines of a memory matrix feed. Compare X SINK.

ytterbium Symbol, Yb. A metallic element of the rare-earth group. Atomic number, 70. Atomic weight, 173.04.

ytterbium metals The rare-earth metals dysprosium, erbium, lutetium, holmium, thulium, and ytterbium.

Yttralox A transparent polycrystalline ceramic composed primarily of yttrium oxide, which has many applications in electro-optics.

yttria Formula, Y_2O_3. Yttrium oxide, a white powder used in Nernst lamps.

yttrium Symbol, Y. A metallic element of the rare-earth-metals group. Atomic number, 39. Atomic weight, 88.906.

yttrium-iron-garnet See YIG.

yttrium metal Any of the group of metals, including yttrium, erbium, holmium, lutetium, thulium, ytterbium, and sometimes dysprosium, gadolinium, and terbium.

Yukon Standard Time Standard time of the ninth time zone west of Greenwich, embracing the Yukon Territory and a portion of southern Alaska.

Y winding See WYE CONNECTION.

Y-Y circuit See WYE-WYE CIRCUIT.

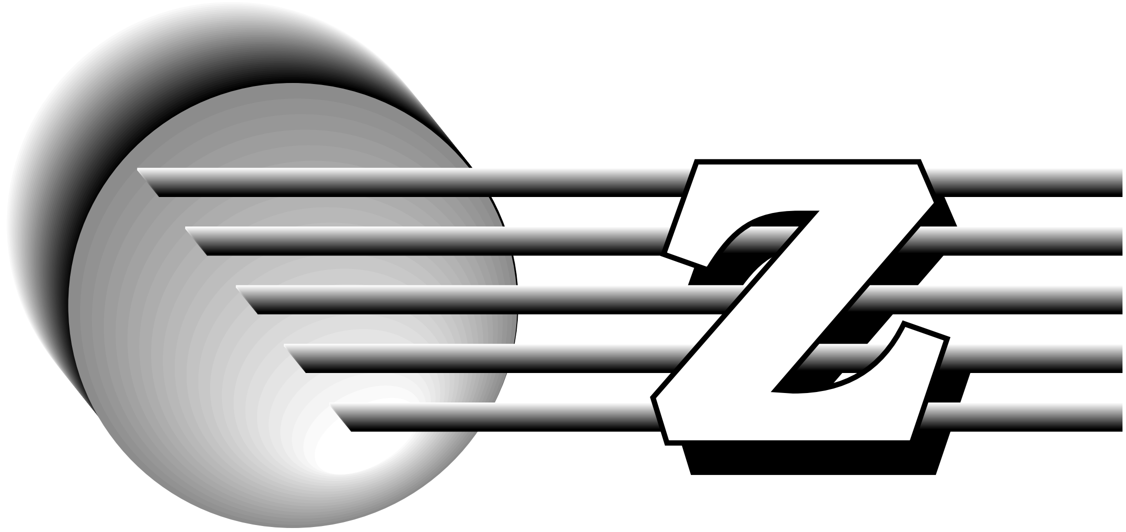

Z **1.** Symbol for IMPEDANCE. **2.** Symbol for ATOMIC NUMBER. **3.** Symbol for *zenith distance* (astronomy).

z **1.** Abbreviation of ZERO. **2.** Symbol for ELECTROCHEMICAL EQUIVALENT. **3.** Abbreviation of ZONE. **4.** Symbol for the vertical axis of a three-dimensional Cartesian graph.

zag The short, straight deflection of a point or particle, or of waves along a jagged path in a direction opposite that of a ZIG; one of the components of ZIGZAG DEFLECTION.

Zamboni pile A high-voltage electrochemical cell consisting of an aluminum anode, a manganese-dioxide cathode, and an aluminum-chloride electrolyte.

Z amplifier The intensity-modulation amplifier in an oscilloscope. Compare X AMPLIFIER and Y AMPLIFIER.

Z-angle meter An instrument, commonly of the null-balance type, that indicates the impedance and phase angle of capacitors, inductors, and sometimes of inductance-capacitance-resistance (ICR) combinations.

z-axis **1.** The intensity-modulation input of an oscilloscope, including the associated circuit. **2.** The video-signal input of a television picture tube, including the associated circuit. **3.** The third axis in a three-dimensional coordinate system. **4.** The lengthwise axis of a quartz crystal.

z-axis amplifier See Z AMPLIFIER.

z-axis modulation See INTENSITY MODULATION.

Z bar See Z-CUT CRYSTAL.

z channel See Z AMPLIFIER.

z-channel gain See Z GAIN.

Z-cut crystal A plate cut from a quartz crystal so that the plane of the plate is perpendicular to the Z-axis of the crystal and parallel to the X-axis. See CRYSTAL AXES.

ZD Abbreviation of *zero defects.*

Z deflection Deflection of a cathode-ray beam beyond the (ordinarily defined) deflection area on the tube screen.

Zeeman effect The splitting of a spectral line of a gas into closely spaced, polarized frequency components by an applied magnetic field.

Zenely electroscope A sensitive alpha-ray electroscope (see ELECTROSCOPE) whose leaf is attracted by a metal plate biased at 50 to 200 volts. On touching the plate, the leaf becomes charged and is repelled. But the gas ions around the leaf neutralize the charge, and the leaf returns to the plate to repeat the action. This sequence causes the leaf to oscillate, the number of oscillations per second being proportional to the strength of the ionizing source.

Zener See ZENER DIODE.

Zener breakdown See AVALANCHE BREAKDOWN.

Zener current See AVALANCHE CURRENT.

Zener diode Also sometimes called *avalanche diode* or *Zener.* A semiconductor diode (usually silicon) specially fabricated to take advantage of the effects of avalanche breakdown. It is available in a wide variety of configurations and ratings; it is commonly used as a voltage regulating device in low-voltage power supplies.

Zener-diode clipper A signal-amplitude clipper that does not require direct-current bias because it uses a Zener diode. Clipping takes place at the avalanche voltage. Zener diodes can be connected back to back for alternating-current clipping.

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

Zener-diode coupling element A Zener diode used as a direct coupling element between amplifier stages in an electronic system. When no signal is present, the resistance of the diode is extremely high because the direct-current voltages of the diode-coupled amplifier stages reverse-bias the diode to below the avalanche point, and no current flows from one stage to the next. The signal raises the voltage enough to cause avalanche breakdown; the signal is thus readily transmitted from one stage to the following one.

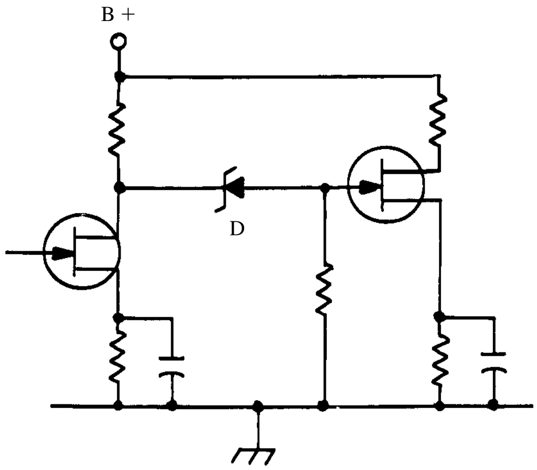

zener-diode coupling element

Zener-diode voltage reference A device that utilizes the constant voltage drop across a Zener diode operated in its breakdown region, to provide a standard voltage for reference purposes. It consists essentially of a direct-current (dc) voltage source, limiting resistor, and Zener diode. For an alternating-current (ac) reference voltage, an ac voltage source, limiting resistor, and two Zener diodes (connected in parallel, back to back) are required. Zener diodes can be connected in series to supply a higher reference voltage than can be delivered by a single diode.

Zener-diode voltage regulator A simple voltage regulator whose output is the constant voltage drop developed across a Zener diode conducting in the reverse breakdown region. The simple regulator circuit consists of a Zener diode and an appropriate limiting resistor connected in series across the output terminals of an unregulated direct-current power supply. For alternating-current voltage regulation, two Zener diodes can be connected in parallel, back to back. Zener diodes can be connected in series to regulate a higher voltage than can be accommodated by a single diode.

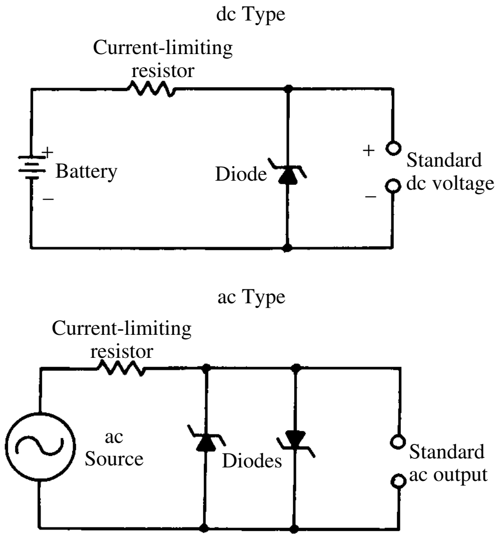
dc Type

ac Type
Current-limiting resistor
+
Battery
−
Diode
+
−
Standard dc voltage
Current-limiting resistor
ac Source
Diodes
Standard ac output

zener-diode voltage reference

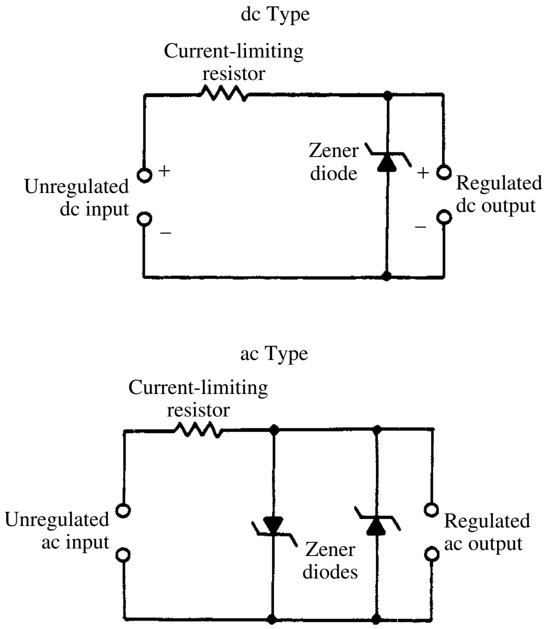
dc Type

ac Type
Current-limiting resistor
Unregulated dc input
+
−
Zener diode
+
−
Regulated dc output
Current-limiting resistor
Unregulated ac input
Zener diodes
Regulated ac output

zener-diode voltage regulator

Zener effect See AVALANCHE BREAKDOWN.

Zener knee In the response of a reverse-biased Zener diode, the point of abrupt transition from low current (high resistance) to high current (low resistance). For voltage regulation and voltage standardization, the knee should be as sharp as possible.

Zener knee current In a Zener diode, the current that flows when the reverse bias reaches the avalanche voltage.

Zener-protected MOSFET See GATE-PROTECTED MOSFET.

Zener voltage The particular value of reverse voltage at which a Zener diode or other reverse-biased pn junction abruptly exhibits the avalanche effect. Depending on the Zener diode, this potential can be less than 10 volts, or as much as several hundred volts.

zenith In the sky, the point directly overhead, exactly 180 degrees opposite the direction of the earth's center.

Zeolite process A method using certain artificial Zeolites (hydrous silicates) to soften water used in some electronic manufacturing and testing operations.

zeppelin antenna Also called *zepp.* **1.** A half-wavelength radiator, fed at one end with a quarter-wavelength section of open-wire line. The feed line can come away from the radiator at any angle, usually 90 degrees or more. The antenna was originally used aboard zeppelins; the entire system was dangled in flight and the feed line was collinear with the radiating element. The impedance at the feed point is high; the impedance at the transmitter end of the feed line is low. The antenna will operate satisfactorily at all harmonics of the design frequency. **2.** A radiator that is an integral multiple of a half wavelength, fed with open-wire line of any length. A transmatch is employed at the transmitter and of the feed line. This arrangement is popular among radio amateurs. The primary advantage of this antenna is ease of installation; the main drawback is the fact that the feed line radiates to some extent because the system is not perfectly balanced.

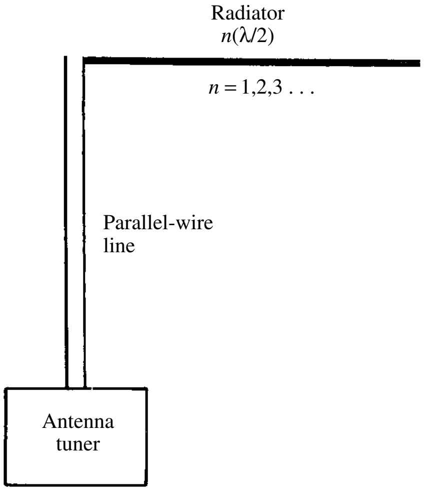

zeppelin antenna

zero **1.** The number represented by the cipher symbol (0). **2.** To set a meter or other instrument to its zero reading or condition. **3.** To align one element of a system precisely with another. **4.** In computer operations, to set a register to zero. **5.** In computer operations, using zero pulses to replace what is in a storage area.

zero-access memory See ZERO-ACCESS STORAGE.

zero-access storage Computer storage requiring negligible waiting time (latency).

zero-address instruction A computer instruction requiring no address; the operation specified by the instruction defines the locations of the operands. Also called *addressless instruction.*

zero adjust In an analog metering device, a mechanical or electrical control that allows precise setting of the reading to zero, when the parameter to be measured is actually zero.

zero adjustment **1.** The act of setting an instrument or circuit to its zero reading or zero-operating condition. **2.** A device or subcircuit used to set a meter to its zero reading. Also see ZERO SET, **1**, **2**.

zero-angle cut An alternate term for *X cut*, as applied to quartz crystal plates. Also see CRYSTAL CUTS.

zero axis In a plot of a variable quantity, the (usually horizontal) reference line that indicates the zero value of the quantity.

zero-axis symmetry The condition in which a waveshape is symmetrical about a zero axis.

zero beat Complete absence of a beat note (i.e., silence), a condition occurring when the frequencies of the beating signals (or their harmonics) are equal. See BEAT NOTE.

zero-beat detector A device or circuit used to sense and indicate the condition of zero beat.

zero-beat indicator See ZERO-BEAT DETECTOR.

zero-beat method A means of adjusting the frequency of one signal exactly to that of another (usually standard) signal by setting the first signal frequency to zero beat with the second signal frequency.

zero-beat reception A system of reception entailing zero beating an incoming signal with the signal from a local oscillator. Also called *homodyne reception.*

zero-bias operation Operation of a transistor or vacuum tube without direct-current base, gate, or grid bias.

zero-bias tube A vacuum tube designed for operation (particularly in a class-B power amplifier) without direct-current grid bias. In such a tube, the zero-signal plate current is very low because of the large amplification factor.

zero capacitance **1.** Absence of capacitance (a theoretically ideal condition). **2.** In some circuits,

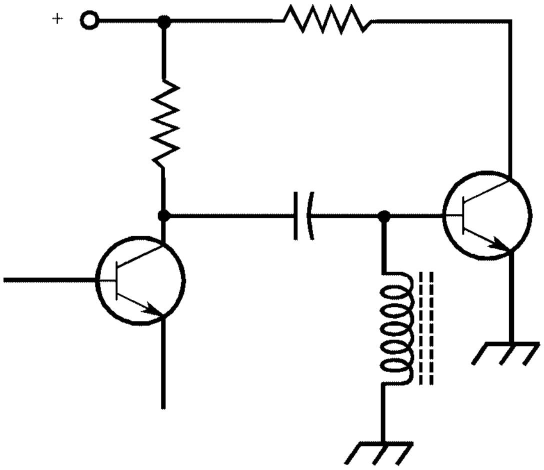

zero-bias operation

the lowest capacitance point, to which all other capacitances are referred.

zero compensation The elimination or minimizing of the zero-signal output of a transducer or similar device.

zero compression See ZERO SUPPRESSION, **1**.

zero condition See ZERO STATE.

zero current **1.** Absence of movement of electrical charge carriers. **2.** In some circuits, the lowest current level, to which all other currents are referred.

zero-cut crystal A piezoelectric plate cut from a quartz crystal in such a way that the frequency-temperature coefficient is zero.

zero-dB reference level An agreed-upon zero level for decibel ratings (which are by nature relative). Zero dBm, for example, corresponds to 1 milliwatt. Compare VOLUME UNIT.

zero drift **1.** The (usually gradual) drift of a zero point, such as the zero setting of an electronic voltmeter. **2.** The condition of no change in the value of a quantity.

zero elimination See ZERO SUPPRESSION, **1**.

zero energy The condition in which energy is neither generated, consumed, nor dissipated.

zero error **1.** In instruments and measurements, an error so small that it can be considered inconsequential. **2.** In a radar system, the inherent delay in the transmitter and receiver circuits.

zero field emission Thermionic emission from a cathode or hot conductor within a uniform electrostatic field.

zero fill See ZERO, **5**.

zero gravity The condition of weightlessness (i.e., the state in which gravitational pull by a celestial body is completely absent or has been nullified).

zero-gravity switch A switch actuated automatically when the condition of zero gravity is reached.

zero impedance **1.** Absence of impedance (a theoretically ideal condition). **2.** In some circuits, the lowest impedance level, to which all other impedances are referred.

zero inductance **1.** Absence of inductance (a theoretically ideal condition). **2.** In some circuits, the lowest inductance level, to which all other inductances are referred.

zero input **1.** Complete absence of input (signal or noise). **2.** Absence of operating voltage or power to a system. **3.** In a flip-flop, the input terminal that is not receiving a trigger signal.

zero-input terminal In a flip-flop, the input-signal terminal at which a trigger signal will switch the flip-flop output to zero. Also called *zero terminal* and RESET TERMINAL.

zero instant See ZERO TIME.

zero level The reference level for the comparison of quantities. For example, it might be a voltage or current level; in audio measurements, it is the zero-dB reference level.

zero-line stability In an analog metering device, the ability of the instrument to maintain proper zero adjustment over a period of time.

zero magnet A permanent magnet used to set the pointer of a meter to zero.

zero magnitude **1.** For a quantity, the state of being valueless (i.e., a complete absence of the quantity). **2.** In some tests, measurements, or calculations, the lowest value of a quantity, to which all other values are referred.

zero meridian The meridian at Greenwich (near London), England, from which longitude and standard time are reckoned. Also see TIME ZONE and ZONE TIME.

zero method A method of measurement entailing adjustment of a circuit or device (such as a bridge, tee network, or potentiometer) for zero response of the detector. Also called *null method.*

zero modulation The momentary lack of modulation in a communications or broadcast system, as during a pause in speech.

zero-modulation noise Noise produced by previously erased tape that is run under specified operating conditions.

zero output **1.** Complete absence of output signal or output power, sometimes disregarding noise output. **2.** In a flip-flop, the normal condition of no signal pulse at a particular output terminal.

zero-output terminal In a flip-flop, the output terminal that is not delivering an output pulse.

zero phase In an alternating-current circuit, the condition in which the current and voltage are in step. That is, the current peaks occur at exactly the same times as the voltage peaks; the phase difference between the current and voltage is zero degrees.

zero pole In a multiconductor line, the common or reference conductor.

zero potential **1.** Complete absence of voltage. **2.** In some circuits, the lowest voltage, to which all other voltages are referred. **3.** The potential of the earth as a reference point.

zero power **1.** Complete absence of dissipated power. **2.** In some circuits and systems, the lowest power level, to which all other power values are referred.

zero-power resistance In a thermistor, the resistance at which power dissipation is zero.

zero-power resistance-temperature characteristic For a thermistor, a figure that reveals the extent to which zero-power resistance varies with the temperature of the thermistor body.

zero-power temperature coefficient of resistance A temperature coefficient that reveals the extent to which the temperature of the thermistor body causes the zero-power resistance to change (expressed in ohms per ohm per degree Celsius).

zero reactance **1.** Absence of reactance (a theoretically ideal condition in alternating-current circuits). **2.** In some circuits, the lowest reactance, to which all other reactances are referred.

zero resistance **1.** Absence of resistance (a theoretically ideal condition). **2.** In some circuits, the lowest resistance, to which all other resistances are referred.

zero scale current In a digital-to-analog (D/A) converter, the current into the output when all logic inputs are low (off) and the output is at a certain predetermined value, in microamperes or milliamperes.

zero screw The mechanical zero adjuster of a meter.

zero set **1.** A (usually screwdriver-adjusted) mechanism for setting the pointer of a meter to zero. **2.** An electrical circuit consisting of a resistance bridge or adjustable bucking voltage for setting a meter to read zero under specific conditions.

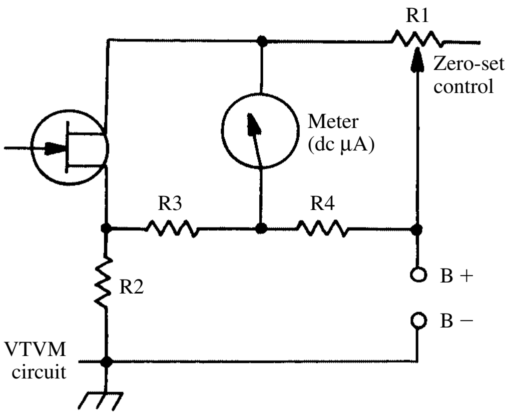

zero set, 2

zero shift See ZERO DRIFT.

zero signal **1.** The condition of complete absence of signal. **2.** A finite signal level used as a reference point against which all other signal levels are measured.

zeros of impedance For an alternating-current network, the frequencies at which the impedance is zero. Also called *zeros.*

zeros of network function The real or complex values at which the network function is zero. Compare POLES OF NETWORK FUNCTION.

zeros of transfer function The frequencies at which a transfer function becomes zero. Compare POLES OF TRANSFER FUNCTION.

zero stability Constancy of the zero condition in an instrument or system (i.e., absence of zero drift).

zero state The low, zero, off, or false logic state of a bistable device, such as a flip-flop or magnetic cell. It might be actual zero output or a low-voltage output. Compare ONE STATE. In binary notation and calculation, the zero state is represented by a cipher.

zero suppression **1.** In computer operation, automatic nonsignificant leading-zero cancellation. **2.** Absence or removal of a restraining medium, such as a noise-suppression voltage. **3.** In an audio recording system, the introduction of a voltage to cancel the steady-state component of the input signal.

zero temperature **1.** The point from which temperatures are reckoned on a thermometer scale. On the Celsius (centigrade) scale, zero degrees corresponds to the freezing point of pure water at standard atmospheric pressure. On the Fahrenheit scale, zero degrees is 32 degrees colder than the freezing point of pure water at standard atmospheric pressure. On the Kelvin or Rankine scales, zero degrees corresponds to the complete absence of thermal energy; it is the coldest possible temperature. **2.** A temperature point relative to which all other temperatures are reckoned.

zero temperature coefficient A temperature coefficient having the value zero (i.e., one that indicates there is no temperature-dependent drift of a given quantity).

zero terminal See ZERO-INPUT TERMINAL and ZERO-OUTPUT TERMINAL.

zero time **1.** In some measurements, the first instant of time, to which all other instants are referred. **2.** See ZERO PHASE.

zero time reference During one cycle of radar operation, the time reference of the schedule of events.

zero vector A vector whose magnitude is zero.

zero voltage See ZERO POTENTIAL.

zero voltage coefficient A voltage coefficient having the value of zero (i.e., one that indicates there is no voltage-dependent drift of a given quantity).

zero-zero The atmospheric condition in which the ceiling and visibility both are zero (i.e., extremely dense fog).

Z gain The gain (or gain control) of the intensity channel of an oscilloscope. Compare X GAIN and Y GAIN.

zig The short, straight deflection of a point or particle, or of a wave along a jagged path in a direction opposite that of ZAG; a component of ZIGZAG DEFLECTION.

zigzag deflection Deflection of a particle, point, or object in a path that contains side-to-side motion, as well as forward motion. Also see ZAG and ZIG.

zigzag rectifier A special version of the three-phase star (three-phase, half-wave) rectifier circuit. Direct-current (dc) saturation of the transformer secondary is avoided by winding half the turns of each secondary on a separate core (i.e., each core carries two half-windings). The opposing flux resulting from different phases in the half-windings causes cancellation of the dc component of flux in each core.

zigzag reflections Multihop reflections of waves along a zigzag path, resulting from repeated reflections within the ionosphere.

zinc Symbol, Zn. A metallic element. Atomic number, 30. Atomic weight, 65.39. It is used as the negative-electrode material in dry cells and as a protective coating for some metals used in electronics.

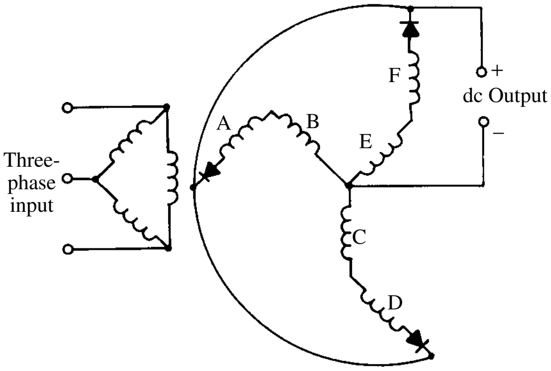

zigzag rectifier

zinc aluminate phosphor Either of two similar substances used as a phosphor coating for cathode-ray tube screens. One form glows blue; the other form glows red.

zinc beryllium silicate phosphor A substance used as a phosphor coating for cathode-ray-tube screens. It glows yellow.

zinc beryllium zirconium silicate phosphor A substance used as a phosphor coating for cathode-ray-tube screens. It glows white.

zinc borate phosphor A substance used as a phosphor coating for cathode-ray-tube screens. It glows yellow-orange.

zinc cadmium sulfide phosphor Either of two similar substances used as a phosphor coating for cathode-ray-tube screens. One form glows blue; the other form glows red.

zinc–carbon cell A common non-rechargeable electrochemical cell. Produces approximately 1.5 volts under no-load conditions. Zinc forms the outer case or shell, and is the negative electrode. A carbon rod serves as the positive electrode. The electrolyte is a paste of manganese dioxide and carbon. The cell is inexpensive, and is commercially available in various sizes. The shelf life is fairly long. Cells of this type work well at moderate temperatures, and in applications where the current drain is moderate to high. They function poorly at low temperatures. Compare ALKALINE CELL.

zinc germanate phosphor A substance used as a phosphor coating for cathode-ray-tube screens. It glows yellow-green.

zinc magnesium fluoride phosphor A substance used as a phosphor coating for cathode-ray-tube screens. It glows orange.

zincolysis Electrolysis in a cell having a zinc anode.

zinc orthoscilicate phosphor Also called *Willemite.* A substance used as a phosphor coating for cathode-ray-tube screens. It glows yellow-green.

zinc oxide A substance used as a phosphor coating for cathode-ray-tube screens. It glows blue-green. It is also used in the manufacture of certain electronic components, such as resistors.

zinc-oxide resistor A voltage-dependent resistor whose active ingredient is zinc oxide.

zinc silicate phosphor A substance used as a phosphor coating for cathode-ray-tube screens. It glows blue.

zinc standard cell A standard cell using zinc and mercury electrodes, and a mercurous sulfate excitant and depolarizer. Produces 1.4322 volts at 15°C. Also called *Clark cell.* Compare WESTON CELL.

zinc sulfide phosphor A substance used as a phosphor coating for cathode-ray-tube screens. Glows blue-green or yellow-green.

ZIP Abbreviation of *zinc-impurity photodetector.*

zip cord A simple two-conductor, flexible power cord.

zirconia Preparations of zirconium (especially $Zr0_2$), valued for their high-temperature dielectric properties.

zirconium Symbol, Zr. A metallic element. Atomic number, 40. Atomic weight, 91.22.

Z marker A vertically radiating marker beacon defining the zone above a radio-range station.

Z meter A device for measuring impedances. Instruments of this kind take four principal forms: **(1)** a direct-reading meter resembling an ohmmeter; **(2)** an adjustable circuit manipulated somewhat like a bridge and that compares an unknown impedance with a standard resistance;

(3) an impedance bridge for evaluating the reactive and resistive components of an unknown impedance; **(4)** a section of transmission line used with a signal source and voltmeter for measuring impedance in terms of a standard resistor and/or standing waves.

ZMODEM In data communications, an error-correction mode similar to XMODEM, except that when an error is found during transmission of a block of data, the source retransmits only that portion of the block following the error. This improves data transmission speed because, when an error occurs, the number of bytes retransmitted is generally fewer than the 128K block size standard in XMODEM. Compare XMODEM and YMODEM.

Zn **1.** Symbol for ZINC. **2.** Symbol for AZIMUTH.

Z_0 Symbol for CHARACTERISTIC IMPEDANCE.

zone **1.** On a magnetic disk, a group of tracks whose associated transfer rate is not the same as that for the rest of the disk. **2.** In computer operations, the area of a storage medium containing digits. **3.** In a computer system, a main memory area set aside for a particular purpose. **4.** In a security system, a specified area or region under surveillance.

zone blanking In a radar system, a method of extinguishing the cathode-ray-tube beam during a portion of the antenna sweep.

zone candle power In a given zone, the luminous flux per steradian, emitted by a light source under test.

zoned circuit In a security system, a circuit in which some areas are protected at all times, while the protection in other areas can be temporarily disabled for entry or exit.

zone leveling See ZONE REFINING.

zone marker See Z MARKER.

zone melting See ZONE REFINING.

zone of silence The region between alternate reflections of a radio wave, in which no signal is detectable. Also called *skip zone.*

zone plate antenna A rapid-scan radar antenna having a reflector composed of confocal parabolas arranged in a circle.

zone-position indicator A radar that reveals the position of an object to a second radar having a restricted field.

zone purification See ZONE REFINING.

zone refining A method of purifying semiconductor materials (such as germanium and silicon) in which a molten zone in an ingot of the material moves along the length of the ingot, dissolving impurities as it travels, eventually depositing them at the end of the ingot, which is sawed off. This concentrated and segregated melting is accomplished by means of radio-frequency heating.

zone time In a given time zone, standard time in terms of the number of hours that must be added to local time to equal the time at the zero (Greenwich) meridian.

zoning **1.** A method of fabricating a microwave reflector in concentric, flat regions, producing the same practical results as a continuous paraboloid. **2.** In a communications system, the division of the coverage area into different geographic regions for a specific purpose.

zoom **1.** To rapidly change the focal length of a television or motion-picture camera lens so that the object seems to advance toward or recede from the viewer, remaining in focus as it does so. **2.** See ZOOM LENS. **3.** To magnify the image in a computer graphical user interface. A user can enlarge a specific portion of the display, at the expense of other portions. It is generally measured in percent (e.g., 200% zoom represents a magnification factor of 2).

zoom lens A continuously adjustable lens system that allows zoom effects to be obtained with a television or motion-picture camera, or a similar arrangement for still cameras that obviates the need for lens interchange when different focal lengths are needed. The lens system, which can be operated electronically, allows either narrow- or wide-angle views to be obtained without losing focus at any time.

Z parameters Device or network parameters expressed as impedances.

ZPI Abbreviation of ZONE-POSITION INDICATOR.

Z-plunger In a waveguide, a combination choke and bucket plunger for radiation leakage reduction.

Zr Symbol for ZIRCONIUM.

Z signals A collection of letter groups, each starting with the letter Z, used for simplified telegraphy and radiotelegraphy by the military services.

Zulu Phonetic alphabet communications code word for the letter Z.

Zulu time Greenwich mean time. Also see ZEBRA TIME.

zwitterion An ion that carries both a positive and a negative charge.

This page intentionally left blank.

Appendix A

Schematic symbols

Ammeter	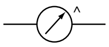
Amplifier general	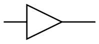
Amplifier, inverting	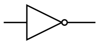
Amplifier, operational	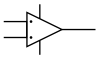
AND gate	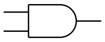
Antenna, balanced	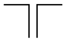
Antenna, general	
Antenna, loop	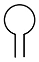
Antenna, loop, multiturn	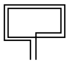
Battery	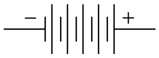
Capacitor, feedthrough	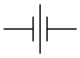
Capacitor, fixed	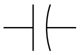
Capacitor, variable	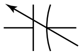
Capacitor, variable, split-rotor	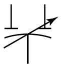
Capacitor, variable, split-stator	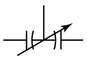

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

Cathode, electron-tube, cold

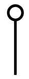

Cathode, electron-tube, directly heated

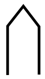

Cathode, electron-tube indirectly heated

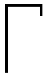

Cavity resonator

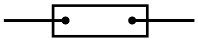

Cell, electrochemical

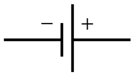

Circuit breaker

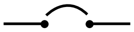

Coaxial cable

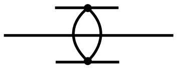

Crystal, piezoelectric

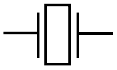

Delay line

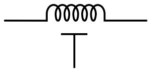

Diac

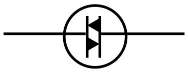

Diode, field-effect

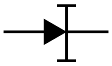

Diode, general

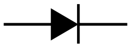

Diode, Gunn

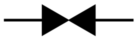

Diode, light-emitting

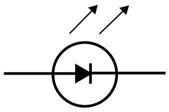

Diode, photosensitive

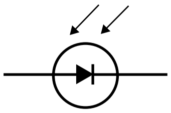

Diode, PIN	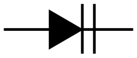
Diode, Schottky	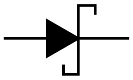
Diode, tunnel	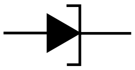
Diode, varactor	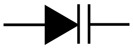
Diode, zener	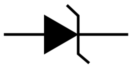
Directional coupler	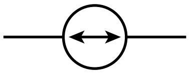
Directional wattmeter	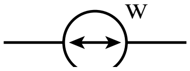
Exclusive-OR gate	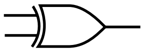
Female contact, general	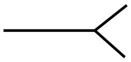
Ferrite bead	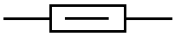
Filament, electron-tube	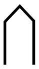
Fuse	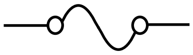
Galvanometer	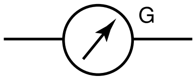
Grid, electron-tube	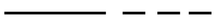
Ground, chassis	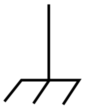

Ground, earth	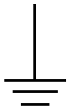
Headset	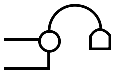
Handset, double	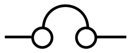
Headset, single	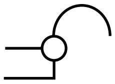
Headset, stereo	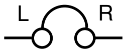
Inductor, air core	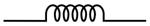
Inductor, air core, bifilar	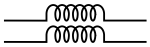
Inductor, air core, tapped	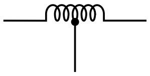
Inductor, air core, variable	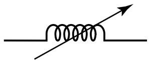
Inductor, iron core	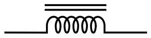
Inductor, iron core, bifilar	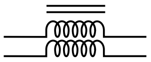
Inductor, iron core, tapped	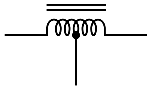
Inductor iron core, variable	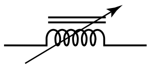
Inductor, powdered-iron core	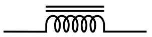
Inductor, powdered-iron core, bifilar	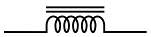

Inductor, powdered-iron core, tapped	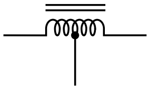
Inductor, powdered-iron core, variable	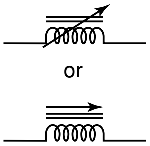
Integrated circuit, general	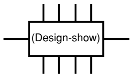
Jack, coaxial or phono	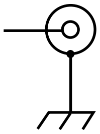
Jack, phone, two-conductor	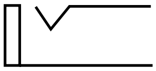
Jack, phone, three-conductor	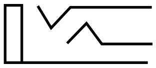
Key, telegraph	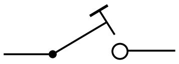
Lamp, incandescent	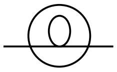
Lamp, neon	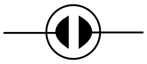
Male contact, general	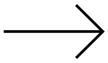
Meter, general	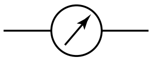
Microammeter	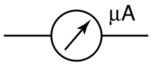
Microphone	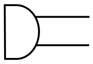

Microphone, directional	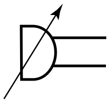
Milliammeter	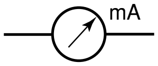
NAND gate	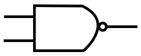
Negative voltage connection	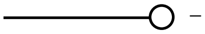
NOR gate	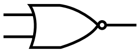
NOT gate	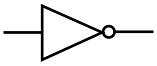
Optoisolator	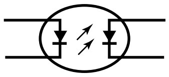
OR gate	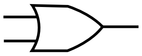
Outlet, two-wire, nonpolarized	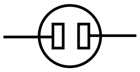
Outlet, two-wire, polarized	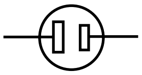
Outlet, three-wire	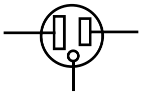
Outlet, 234-V	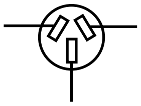
Plate, electron-tube	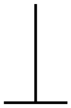
Plug, two-wire, nonpolarized	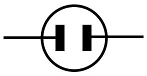

Plug, two-wire, polarized	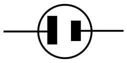
Plug, three-wire	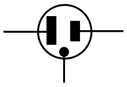
Plug, 234-V	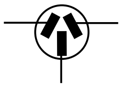
Plug, coaxial or phono	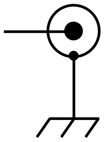
Plug, phone, two-conductor	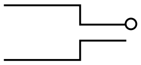
Plug, phone, three-conductor	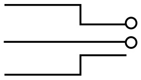
Positive voltage connection	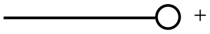
Potentiometer	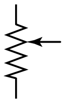
Probe, radio-frequency	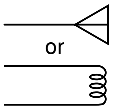
Rectifier, gas-filled	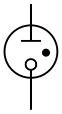
Rectifier, high-vacuum	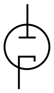
Rectifier, semiconductor	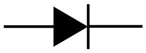
Rectifier, silicon-controlled	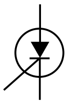

Relay, double-pole, double-throw	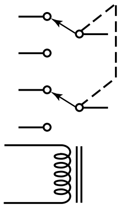
Relay, double-pole, single-throw	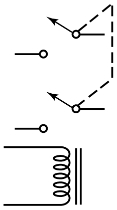
Relay, single-pole, double-throw	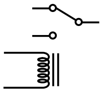
Relay, single-pole, single-throw	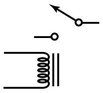
Resistor, fixed	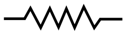
Resistor, preset	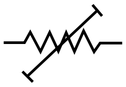
Resistor, tapped	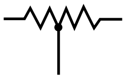
Resonator	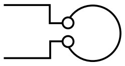
Rheostat	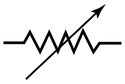
Saturable reactor	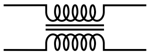
Signal generator	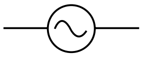
Solar battery	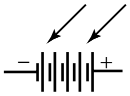

Solar cell

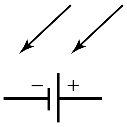

Source, constant-current

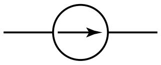

Source, constant-voltage

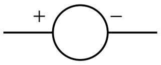

Speaker

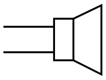

Switch, double-pole, double-throw

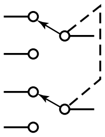

Switch, double-pole, rotary

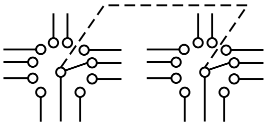

Switch, double-pole, single-throw

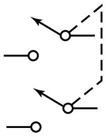

Switch, momentary-contact

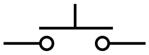

Switch, silicon-controlled

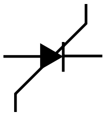

Switch, single-pole, rotary

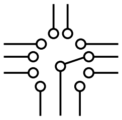

Switch, single-pole, double-throw

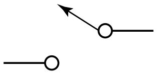

Switch, single-pole, single-throw

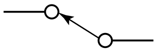

Terminals, general, balanced

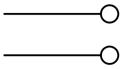

Terminals, general, unbalanced

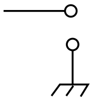

Test point

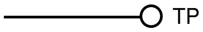

Thermocouple

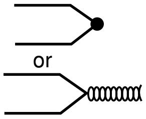

Transformer, air core

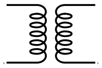

Transformer, air core, step-down

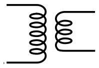

Transformer, air core, step-up

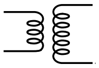

Transformer, air core, tapped primary

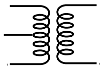

Transformer, air core, tapped secondary

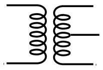

Transformer, iron core

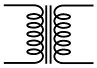

Transformer, iron core, step-down

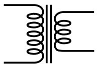

Transformer, iron core, step-up

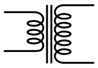

Transformer, iron core, tapped primary

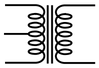

Transformer, iron core, tapped secondary

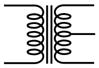

Transformer, powdered-iron core	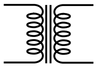
Transformer, powdered-iron core, step-down	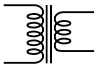
Transformer, powdered-iron core, step-up	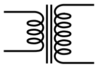
Transformer, powdered-iron core, tapped primary	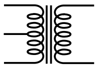
Transformer, powdered-iron core, tapped secondary	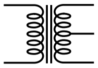
Transistor, bipolar, *NPN*	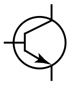
Transistor, bipolar, *PNP*	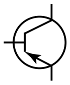
Transistor, field-effect, *N*-channel	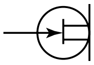
Transistor, field-effect, *P*-channel	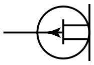
Transistor, MOS field-effect, *N*-channel	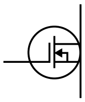
Transistor, MOS field-effect, *P*-channel	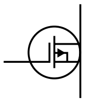
Transistor, photosensitive, *NPN*	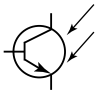
Transistor, photosensitive, *PNP*	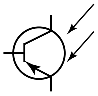

Transistor, photosensitive, field-effect, *N*-channel

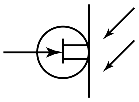

Transistor, photosensitive, field-effect, *P*-channel

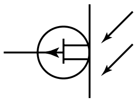

Transistor, unijunction

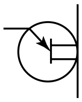

Triac

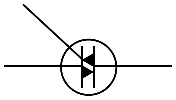

Tube, diode

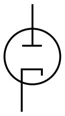

Tube, heptode

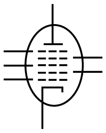

Tube, hexode

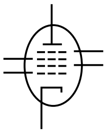

Tube, pentode

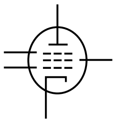

Tube, photosensitive

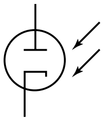

Tube, tetrode

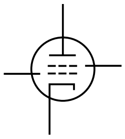

Tube, triode

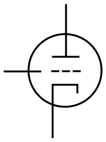

Voltmeter

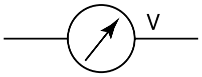

Wattmeter

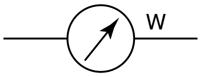

Waveguide, circular

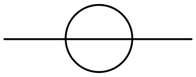

Waveguide, flexible

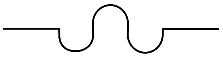

Waveguide, rectangular

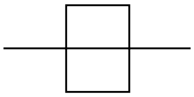

Waveguide, twisted

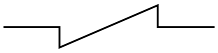

Wires, crossing, connected

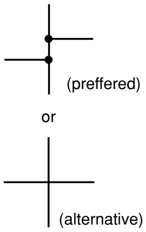
(preffered)
or
(alternative)

Wires, crossing, not connected

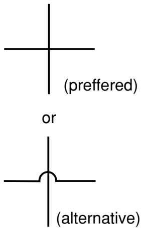
(preffered)
or
(alternative)

Appendix B

Tables and data

Conversion between Electrical Systems

		System	
Property	**mks**	**cgs electromagnetic**	**cgs electrostatic**
Capacitance	1 farad	10^{-9} abfarad	9×10^{11} statfarad
	10^{9} F	1 abF	9×10^{20} statF
	$10^{-11}/9$ F	$10^{-20}/9$ abF	1 statF
Charge	1 coulomb	0.1 abcoulomb	3×10^{9} statC
	10 C	1 abC	3×10^{10} statC
	$10^{-9}/3$ C	$10^{-10}/3$aC	1 statC
Charge density	1 coulomb/m^{3}	10^{-7} abcoulomb/cm^{3}	3×10^{3} statcoulomb/cm^{3}
	10^{7} C/m^{3}	1 abC/cm^{3}	3×10^{10} statCcm^{3}
	$10^{-3}/3$ C/m^{3}	$10^{-10}/3$ aC/cm^{3}	1 statC/cm^{3}
Conductivity	1 siemens/m	10^{-11} absiemens/cm	9×10^{9} statSiemens/cm
	10^{11} S/m	1 abS/cm	9×10^{20} statS/cm
	$10^{-9}/9$ S/m	$10^{-20}/9$ abS/cm	1 statS/cm
Current	1 ampere	10^{-1} abampere	3×10^{9} statampere
	10 a	1 abA	3×10^{10} statA
	$10^{-9}/3$ a	$10^{-10}/3$ abA	1 statA
Current density	1 ampere/m^{2}	10^{-5} abampere/cm^{2}	3×10^{5} statampere/cm^{2}
	10^{5} A/m^{2}	1 abA/cm^{2}	3×10^{10} statA/cm^{2}
	$10^{-5}/3$ A/m^{2}	$10^{-10}/3$ aA/cm^{2}	1 statA/cm^{2}
Electric field intensity	1 volt/m	10^{6} abvolt/cm	$10^{-4}/3$ statvolt/cm
	10^{-6} V/m	1 abV/cm	$10^{-10}/3$ statV/cm
	3×10^{4} V/m	3×10^{10} aV/cm	1 statV/cm
Electric potential	1 volt	10^{8} abvolt	$10^{-2}/3$ statvolt
	10^{-8} V	1 abV	$10^{-10}/3$ statV
	3×10^{2} V	3×10^{10} aV	1 statV
Electric dipole moment	1 coulomb-m	10 abC-cm	3×10^{11} statC-cm
	0.1 C-m	1 abC-cm	3×10^{10} statC-cm
	$10^{-11}/3$ C-m	$10^{-10}/3$ abC-cm	1 statC-cm

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

	System		
Property	**mks**	**cgs electromagnetic**	**cgs electrostatic**
Energy	1 joule	10^7 erg	10^7 erg
	10^{-7} J	1 e	1 e
	10^{-7} J	1 e	1 e
Force	1 newton	10^5 dyne	10^5 dyne
	10^{-5} N	1 d	1 d
	10^{-5} N	1 d	1 d
Flux density	1 Weber/m^2	10^4 gauss (or abtesla)	$10^{-6}/3$ electrostatic unit
	10^{-4} Wb/m^2	1 G	$10^{-10}/3$ esu
	3×10^6 Wb/m^2	3×10^{10} G	1 esu
Inductance	1 henry	10^9 abhenry	$10^{-11}/9$ stathenry
	10^{-9} H	1 abH	$10^{-20}/9$ statH
	9×10^{11} H	9×10^{20} abH	1 statH
Inductive capacity	1 farad/m	10^{-11} abfarad/cm	9×10^9 statfarad/cm
	10^{11} F/m	1 abF/cm	9×10^{20} statF/cm
	$10^{-9}/9$ F/m	$10^{-20}/9$ abF/cm	1 statF/cm
Magnetic flux	1 weber	10^8 Maxwell	$10^{-2}/3$ electrostatic unit
	10^{-8} W	1 Mx	$10^{-10}/3$ esu
	3×10^2 W	3×10^{10} Mx	1 esu
Magnetic dipole moment	1 ampere-m^2	10^3 abampere-cm^2	3×10^{13} statampere-cm^2
	10^{-3} A-m^2	1 abA-cm^2	3×10^{10} statA-cm^2
	$10^{-13}/3$ A-m^2	$10^{-10}/3$ abA-cm^2	1 statA-cm^2
Permeability	1 henry/m	10^7 abhenry/cm	$10^{-13}/9$ stathenry/cm
	10^{-7} H/m	1 abH/cm	$10^{-20}/9$ statH/cm
	9×10^{13} H/m	9×10^{20} abH/cm	1 statH/cm
Power	1 watt	10^7 erg/s	10^7 erg/s
	10^{-7} W	1 e/s	1e/s
	10^{-7} W	1 e/s	1e/s
Resistance	1 ohm	10^9 abohm	$10^{-11}/9$ statohm
	10^{-9} ohm	1 abohm	$10^{-20}/9$ statohm
	9×10^{11} ohm	9×10^{20} abohm	1 statohm

GREEK ALPHABET

Capital	Lowercase	Name
Α	α	alpha
Β	β	beta
Γ	γ	gamma
Δ	δ or ∂	delta
Ε	ε	epsilon
Ζ	ζ	zeta
Η	η	eta
Θ	θ or ϑ	theta
Ι	ι	iota
Κ	κ	kappa
Λ	λ	lambda
Μ	μ	mu
Ν	ν	nu
Ξ	ξ	xi
Ο	ο	omicron
Π	π	pi
Ρ	ρ	rho
Σ	σ or ζ	sigma
Τ	τ	tau
Υ	υ	upsilon
Φ	ϕ or φ	phi
Χ	χ	chi
Ψ	ψ	psi
Ω	ω	omega

Mathematical Functions

Signs and symbols

·	radix (base) point	≠	not equal to
•	multiplication symbol; logic AND function	~	similar to
∞	infinity	<	less than
+	plus; positive; logic OR function	≮	not less than
−	minus; negative	<<	much less than
±	plus or minus; positive or negative	>	greater than
∓	minus or plus; negative or positive	≯	not greater than
×	times	>>	much greater than
÷	divided by	≤	equal to or less than
/	divided by (expressive of a ratio)	≥	equal to or greater than
=	equal to	∝	proportional to; varies directly as
≡	identical to; is defined by	≐ or →	approaches
≃ or ≅	approximately equal to, congruent to	:	is to; proportional to
≈	approximately equal to	∴	therefore

@	at the rate of; at cost of
e	natural number = 2.71828
π	pi ≈ 3.14159 . . .
( )	parentheses (use to enclose a group of terms)
[ ]	brackets (use to enclose a group of terms that includes one or more groups in parentheses)
{ }	braces (use to enclose a group of terms that includes one or more groups in brackets.
$\angle$	angle
$^\circ$	degrees (arc or temperature)
$'$	minutes; prime
$''$	seconds; double prime
$\|\|$	parallel to
$\perp$	perpendicular to
. . .	and beyond
∇	(del or nabla) vector differential operator

Operations

$x + y$	x added to y; x OR y
$x - y$	y subtracted from x
$x \cdot y$, $x \times y$, or xy	multiplied by y; x AND y
$x \div y$	x divided by y
x/y or $\frac{x}{y}$	x divided by y
$1/x$	reciprocal of x
x^n	x raised to the indicated power of n
$\sqrt[n]{x}$	indicated root n of x
$x : y$	x is to y
$\|x\|$	absolute value of x, magnitude of x
$\bar{X}$, $\dot{X}$, or X	vector X
$\bar{x}$	average value of x; x not
$f(x)$ or $F(x)$	function of x
i	$\sqrt{-1}$
j	operator, equal to $\sqrt{-1}$
Δx	increment of x
$\mathrm{d}x$	differential of x
∂x	partial differential of x
$\frac{\Delta x}{\Delta y}$	change in x with respect to y
$\frac{\mathrm{d}x}{\mathrm{d}y}$	derivative of x with respect to y
$\frac{\mathrm{d}}{\mathrm{d}y}(x)$	derivative of x with respect to y
$\mathrm{d}_y x$	derivative of x with respect to y
$\frac{\partial x}{\partial y}$	partial derivative of x with respect to y
$\sum$	summation
$\sum_{b}^{a}$	summation between limits (from a to b)
$\prod$	product
$\prod_{b}^{a}$	product between limits (from a to b)
$\int$	integral
$\int_a^b$	integral between limits (from a to b)
$\int x\,\mathrm{d}y$	integral of x with respect to y
$\|_a$	evaluated at a
$\|_a^b$	evaluated between limits (from a to b)
$n!$	factorial of n

Prefixes

Prefix	Quantity	Symbol
atto-	10^{-18}	a
femto-	10^{-15}	f
pico-	10^{-12}	p
nano-	10^{-9}	n
micro-	10^{-6}	μ
milli-	10^{-3}	m
centi-	10^{-2}	c
deci-	10^{-1}	d
deka-	10	da
hecto-	10^{2}	h
kilo-	10^{3}	k
mega-	10^{6}	M
giga-	10^{9}	G
tera-	10^{12}	T
peta-	10^{15}	P
exa-	10^{18}	E

RESISTOR COLOR CODE

The first three color bands supply the total resistance. The fourth color (if any) gives the tolerance. Example: A 5,600-ohm resistor would have green (first color, 5), blue (second color, 6) and red

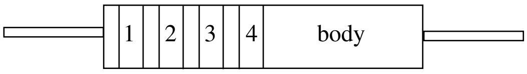

1

black = 0
brown = 1
red = 2
orange = 3
yellow = 4
green = 5
blue = 6
violet = 7
gray = 8
white = 9

2

black = 0
brown = 1
red = 2
orange = 3
yellow = 4
green = 5
blue = 6
violet = 7
gray = 8
white = 9

3

black =
brown = 0
red = 00
orange = 000
yellow = 0000
green = 00000
blue = 000000
gold = multiply by 0.1
silver = multiply by 0.01

4

gold ± 5%
silver ± 10%
no band ± 20%

This page intentionally left blank.

About the author

Stan Gibilisco has authored or coauthored dozens of nonfiction books about electronics and science. He first attracted attention with *Understanding Einstein's Theories of Relativity* (TAB Books, 1983). His *Encyclopedia of Electronics* (TAB Professional and Reference Books, 1985) and *Encyclopedia of Personal Computing* (McGraw-Hill, 1996) were annotated by the American Library Association as among the best reference volumes published in those years. Stan's work has gained reading audiences in the Far East, Europe, and South America.

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

This page intentionally left blank.

SOFTWARE AND INFORMATION LICENSE

The software and information on this diskette (collectively referred to as the "Product") are the property of The McGraw-Hill Companies, Inc. ("McGraw-Hill") and are protected by both United States copyright law and international copyright treaty provision. You must treat this Product just like a book, except that you may copy it into a computer to be used and you may make archival copies of the Products for the sole purpose of backing up our software and protecting your investment from loss.

By saying "just like a book," McGraw-Hill means, for example, that the Product may be used by any number of people and may be freely moved from one computer location to another, so long as there is no possibility of the Product (or any part of the Product) being used at one location or on one computer while it is being used at another. Just as a book cannot be read by two different people in two different places at the same time, neither can the Product be used by two different people in two different places at the same time (unless, of course, McGraw-Hill's rights are being violated).

McGraw-Hill reserves the right to alter or modify the contents of the Product at any time.

This agreement is effective until terminated. The Agreement will terminate automatically without notice if you fail to comply with any provisions of this Agreement. In the event of termination by reason of your breach, you will destroy or erase all copies of the Product installed on any computer system or made for backup purposes and shall expunge the Product from your data storage facilities.

LIMITED WARRANTY

McGraw-Hill warrants the physical diskette(s) enclosed herein to be free of defects in materials and workmanship for a period of sixty days from the purchase date. If McGraw-Hill receives written notification within the warranty period of defects in materials or workmanship, and such notification is determined by McGraw-Hill to be correct, McGraw-Hill will replace the defective diskette(s). Send request to:

Customer Service
McGraw-Hill
Gahanna Industrial Park
860 Taylor Station Road
Blacklick, OH 43004-9615

The entire and exclusive liability and remedy for breach of this Limited Warranty shall be limited to replacement of defective diskette(s) and shall not include or extend to any claim for or right to cover any other damages, including but not limited to, loss of profit, data, or use of the software, or special, incidental, or consequential damages or other similar claims, even if McGraw-Hill has been specifically advised as to the possibility of such damages. In no event will McGraw-Hill's liability for any damages to you or any other person ever exceed the lower of suggested list price or actual price paid for the license to use the Product, regardless of any form of the claim.

THE McGRAW-HILL COMPANIES, INC. SPECIFICALLY DISCLAIMS ALL OTHER WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO, ANY IMPLIED WARRANTY OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Specifically, McGraw-Hill makes no representation or warranty that the Product is fit for any particular purpose and any implied warranty of merchantability is limited to the sixty day duration of the Limited Warranty covering the physical diskette(s) only (and not the software or in-formation) and is otherwise expressly and specifically disclaimed.

This Limited Warranty gives you specific legal rights; you may have others which may vary from state to state. Some states do not allow the exclusion of incidental or consequential damages, or the limitation on how long an implied warranty lasts, so some of the above may not apply to you.

This Agreement constitutes the entire agreement between the parties relating to use of the Product. The terms of any purchase order shall have no effect on the terms of this Agreement. Failure of McGraw-Hill to insist at any time on strict compliance with this Agreement shall not constitute a waiver of any rights under this Agreement. This Agreement shall be construed and governed in accordance with the laws of New York. If any provision of this Agreement is held to be contrary to law, that provision will be enforced to the maximum extent permissible and the remaining provisions will remain in force and effect.

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

Suggested additional references

Crowhurst, N. and Gibilisco, S., *Mastering Technical Mathematics—2nd Edition.* New York: McGraw-Hill, 1999.
Dorf, R., *Electrical Engineering Handbook—2nd Edition.* Boca Raton, FL: CRC Press, 1997.
Gibilisco, S., *Electronics Portable Handbook.* New York: McGraw-Hill, 1999.
Gibilisco, S., *Electronics Formulas Pocket Reference.* New York: McGraw-Hill, 1999
Gibilisco, S., *Mathematical and Physical Data, Equations, and Rules of Thumb.* New York: McGraw-Hill, 2001.
Van Valkenburg, M., *Reference Data for Engineers: Radio, Electronics, Computer and Communications.* Indianapolis: Howard W. Sams & Co., 1998.
Veley, V., *The Benchtop Electronics Reference Manual.* New York: McGraw-Hill, 1994.

Copyright 2001 The McGraw-Hill Companies, Inc. Click Here for Terms of Use

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.
6238262 2013/12/10 12.27.66.5

CHAPTER 1

SYSTEMS, QUANTITIES, AND UNITS

OUTLINE

1–1 The Electronics Industry
1–2 Introduction to Electronic Systems
1–3 Types of Circuits
1–4 Scientific and Engineering Notation
1–5 Units and Metric Prefixes
1–6 Metric Unit Conversions
1–7 Measured Numbers
1–8 Electrical Safety

OBJECTIVES

- Describe the electronics industry in general terms
- Describe attributes of a system
- Describe attributes of circuits in general
- Use scientific notation to represent quantities
- Work with electrical units and metric prefixes
- Convert from one unit with a metric prefix to another
- Express measured data with the proper number of significant digits
- Recognize electrical hazards and practice proper safety procedures

KEY TERMS

Circuit
Vertical organization
Horizontal organization
Integrated circuit
Mechatronics
System
Boundary
Input
Output
Block diagram
Transfer curve
Gain
Passive component
Active component
Oscillator
Alternator
Digital
Analog
Transducer
Scientific notation
Power of ten
Exponent
Engineering notation
SI
Metric prefix
Error
Accuracy
Precision
Significant digit
Round off
Electrical shock

INTRODUCTION

Technical electronics has changed from less emphasis on troubleshooting discrete circuits at the component level into more work whereby technicians install and test systems. Technicians need to become more oriented to system integration than in the past. In order to understand how systems operate, anyone entering a technical field should be grounded in fundamental laws. This chapter begins with an overview of the electronics industry and attributes of systems and circuits. It then covers the basic units, electrical quantities, and engineering and scientific notation used in the field of electronics.

VISIT THE WEBSITE
Study aids for this chapter are available at
http://pearsonhighered.com/floyd

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

1–1 THE ELECTRONICS INDUSTRY

The electronics industry is made up of companies that are competing for business in manufacturing and marketing electronics products. Because of this fierce competition, the level of innovation and new products has been spectacular and continues to this day. One change in the industry has been the growth of small specialized companies with a narrow market area.

After completing this section, you should be able to

- **Describe the electronics industry in general terms**
  - **Describe briefly the processes engineers use to design and test a new circuit**
  - **Explain the difference between vertical and horizontal company organization**
  - **Cite an example of skills that service technicians need to apply to systems work**
  - **Discuss advantages to having a certification in a particular field**

The electronics industry has witnessed revolutionary technological advances, especially in semiconductors. These advances have changed the industry and the way circuits are developed, constructed, and repaired. A **circuit**[*] is an interconnection of electrical components designed to produce a desired result. Circuits have become more compact and with more reliability than in the past. Today, engineers use circuit design software at a workstation to begin the design process for a new circuit. The software simulates the performance of the circuit and looks for potential problems (such as timing issues, and thermal or noise problems). Computer simulations can enable a circuit to be optimized with minimal cost. When the engineer is satisfied that the simulation meets the requirements for the circuit, the computer automatically lays out the circuit on a printed circuit (PC) board and forms a complete parts list. From this, a prototype is constructed in 48 hours or less either in-house or by a company that specializes in prototyping PC boards. An engineer tests the prototype circuit, and the design is either modified or approved for production by a review team.

In addition to computer-aided design and simulation, in recent years companies have tended to shift their focus away from **vertical organization** with a definite hierarchy to a focus that is more specialized. A vertical organization may go from raw material to finished product in one facility. A company with a **horizontal organization** is one that has a structure where decisions are decentralized. The focus can be more on a specialty. Companies tend to outsource any work not directly related to their specialty. One aspect of horizontal organization is that many smaller companies are involved in producing intermediate components that go into the final product. A specialized production company then assembles the final product from the components. These smaller specialized companies are often centered near each other in regional areas to save on shipping and to lower inventory costs. For certain goods, many specialized suppliers and assembly companies are located in countries where manufacturing costs are low. The result is that many products do not have a single country of origin.

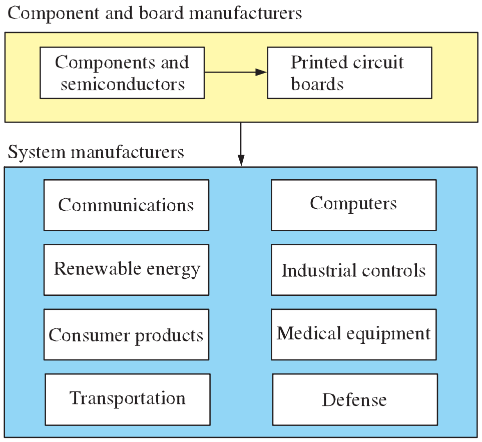

FIGURE 1–1 **The electronics manufacturing industry.**

Major Divisions

Two major divisions of the electronics industry are the service sector and the manufacturing sector. The service sector supports all of the manufacturing sectors as well as customers. The manufacturing sector is illustrated in Figure 1–1. Fundamental to all manufacturing is component and semiconductor manufacturers and printed circuit board manufacturers. System manufacturers make systems from fundamental components and/or printed circuit boards. System manufacturers (shown in the blue area) assemble these into completed systems. These manufacturers include companies making communications equipment (including broadcast), computers, renewable energy systems, and more.

---

[*] All bold terms are in the end-of-book glossary. The bold terms in color are key terms and are also defined at the end of the chapter.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

All of the electronic products that manufacturers produce today involve highly automated processes. These processes can assemble a range of items from integrated circuits to printed circuit boards. An **integrated circuit** is a complex circuit consisting of resistors, transistors, and other components fabricated as a single unit that performs the function of multiple discrete components. The result of automation for consumers has been cheaper products and constant improvements in technology (speed and reliability as well as new features).

As mentioned, the increased reliability of electronic products has caused a shift in the service sector from troubleshooting and repair to a system approach. In the service sector, most technicians need to be able to identify the faulty board, make a quick verification that it is bad, and replace it. Skills that technicians need tend to be broader than the traditional skills required in the past. In the manufacturing sector, technicians may have to work on PLCs (programmable logic controllers), computers, robotics, and mechanical assemblies. Skilled technicians are required to be able to work on a system as a whole for installation, maintenance, and troubleshooting regardless of where a fault may lie. These jobs often involve the new applied engineering classification of **mechatronics**, which is a synergistic combination of mechanics and electronics.

Perhaps surprisingly, the service sector provides more jobs than the manufacturing sector, although there is overlap (as in servicing a machine in a manufacturing plant). Many service sector jobs require a broad range of skills. For example, consider the skills required for the installation and maintenance of solar electrical systems. The technician must have knowledge of electrical and electronic systems and perform electrical and mechanical installations of modules, junction boxes, overcurrent devices, grounding equipment, and pumps. He or she must be able to test and troubleshoot the systems and work safely with both the electrical and nonelectrical hazards associated with these systems.

Manufacturers require people to work on the machines and systems that are used in their industries and reprogram them if necessary. For example, in the food-processing industry, companies commonly utilize many specialized machines for sorting, weighing, and packaging that are controlled by electronic systems. Figure 1–2 shows a check-weight scale that is common in the food-processing industry. The scale is only one element in the overall packaging system. A subassembly may be connected to the scale to acquire data and send it to a computer to determine if the product is within a certain specified range. If the package size were changed, a technician would set the new limits of an acceptable range in the controller.

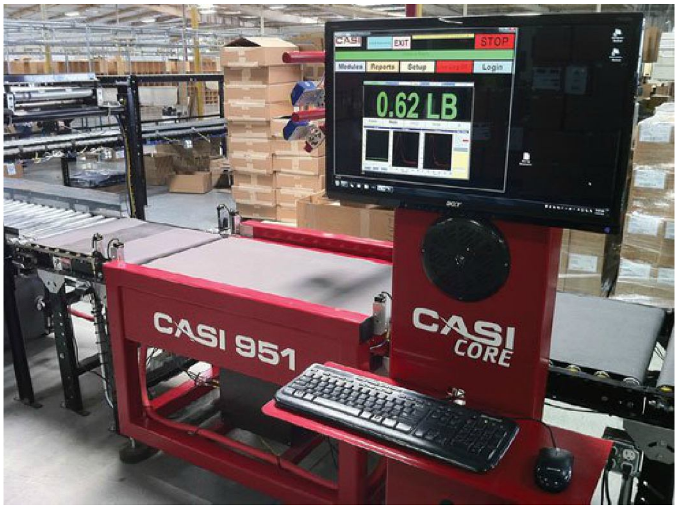

FIGURE 1–2 An automated check-weight scale in the food-processing industry. (Courtesy of Cornerstone Automation Systems, LLC.)

The service sector also requires technicians with electronics skills for a diverse range of jobs that includes the installation and repair of highly technical communication equipment (such as found in the broadcast industry), system integrators, and solar and wind energy systems to name a few. Other related positions in the service field are for technical trainers and writers, customer service representatives, and sales.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Certification

In virtually every field, there are professional organizations that issue certifications to show competency in that field. Certification allows individuals to showcase their skills in a given technical area and at the same time offers a means for employers and/or the public for judging that competency. In the broadcast field, a certification agency is iNARTE (interNational Association for Radio, Telecommunications, and Electromagnetics). This organization also administers FCC Commercial Operator License Examinations. In the energy field, the North American Board of Certified Energy Practitioners (NABCEP) is a certifying agency for several technical jobs. It can issue installer certification for PV systems, solar thermal systems, or small wind systems. Certification in any field does not replace any license requirements required by a state or governing agency but is useful for potential job applicants to show skill in that field.

6238262 2015/12/10 12.27.66.5

SECTION 1–1 CHECKUP*

1. What are the advantages to having a computer simulation of a new circuit before it is constructed?
2. What are the two major divisions of the electronics industry?
3. Name six product categories in the electronics systems manufacturing industry.
4. What is the purpose of certification?

***Answers are at the end of the chapter.**

1–2 INTRODUCTION TO ELECTRONIC SYSTEMS

An electronic system is an assembly of components and circuits designed to accomplish a specific function. Examples of electronic systems range from a simple garage door opener to a complex system such as a radar system.

After completing this section, you should be able to

- **Describe attributes of a system**
  - **Define the word *system* as it applies to electrical and electronic systems**
  - **Explain the purpose of a block diagram and read a basic block diagram**
  - **Give an example of a transfer curve as applied to a block within an electronic system.**

System Concepts

A **system** is a group of interrelated parts that perform a specific function. A **boundary** is the dividing line between what is part of the system and its environment, which is everything else, as illustrated in Figure 1–3.

A system communicates to the outside world through its inputs and outputs. An **input** is the voltage, current, or power that is applied to an electrical circuit to achieve a desired result. An **output** is the result obtained from the system after processing one or more inputs. Systems can have multiple inputs and multiple outputs. No system is completely isolated from its environment; however, for simplifying the analysis of a system, it is often treated as if it were isolated. Even a system that seems isolated, such as the International Space Station (ISS), receives solar energy, radiates heat into space, and is affected by the gravitational field of the earth.

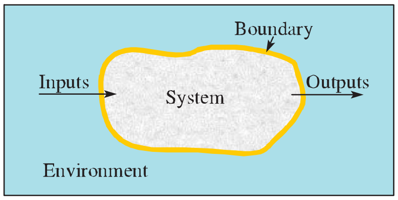

FIGURE 1–3 **A system and its environment.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

ELECTRICAL SYSTEMS Electrical systems deal with electric power. As an example of an electrical system, consider the wiring for a house. Normally, the outside walls and roof are defined as the boundary; and the interior space, including the basement and attic, is defined as the system. The connection to the utility grid is done at a panel and is considered to be an input to the system. The outputs represent specific points within the house (or outside) where loads are connected. The boundary can be changed to suit the analysis. For the house-wiring system, a portion of this (such as just the kitchen circuits) could be considered a system or a subsystem.

ELECTRONIC SYSTEMS Generally, electronic systems deal with signals rather than power. In the context of an electronic system, a *signal* is a changing electrical or electromagnetic quantity that carries information. If a particular input to a system never changes, it is completely predictable. In this case, the input is not considered a signal because no information is present. Most electronic systems process the information contained in a signal. The distinction between electronic and electrical circuits is blurred by the fact that electrical systems frequently use electronic components for regulation. Electronic systems will normally have an electrical subsystem for supplying necessary power.

Block Diagrams

Electronic systems will generally involve a logical sequence of processes. To simplify complicated systems and indicate the order of processing, systems are frequently drawn in block diagram form. A **block diagram** is a model of a system that represents its structure in a graphical format using labeled blocks to represent functions and lines to represent the signal flow. For example, the block diagram of a digital thermometer system is shown in Figure 1–4. Assume that the temperature sensor is one that changes its output voltage based on the temperature of its environment. The signal from the sensor is increased by the amplifier and sent to the ADC. (ADC stands for analog-to-digital converter, which changes the continuous voltage from the amplifier into a digital number.) In addition to basic processing, digital data can be stored and recovered accurately and is much less subject to noise than analog data. The processor converts the digital data to a temperature reading that is displayed. The display for this system is a lighted number. The block diagram shows the key parts of the system and the signal flow without showing the details of the circuits.

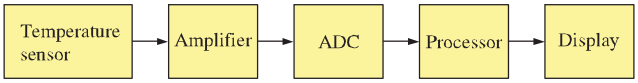

FIGURE 1–4 **A block diagram of a digital thermometer system.**

SYSTEM EXAMPLE 1–1

BLOCK DIAGRAMS AND FLOWCHARTS

Data acquisition systems are used for a variety of applications. Usually, these systems need to convert an analog quantity to digital data for processing. The analog-to-digital converter (ADC) is a key part of this type of system. System analysts can explain the relationship of the ADC, called a successive approximation ADC, in a system by using both a block diagram and a flowchart. A block diagram would be used to illustrate signal flow for the overall system; whereas a flowchart would be used to show the logical process that the control logic block must perform.

The difference between a block diagram and a flowchart is illustrated in Figure 1–5. Part (a) shows a block diagram of the hardware parts, showing the signal flow; it does not show details of the process that the control logic must perform. The control logic block is

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

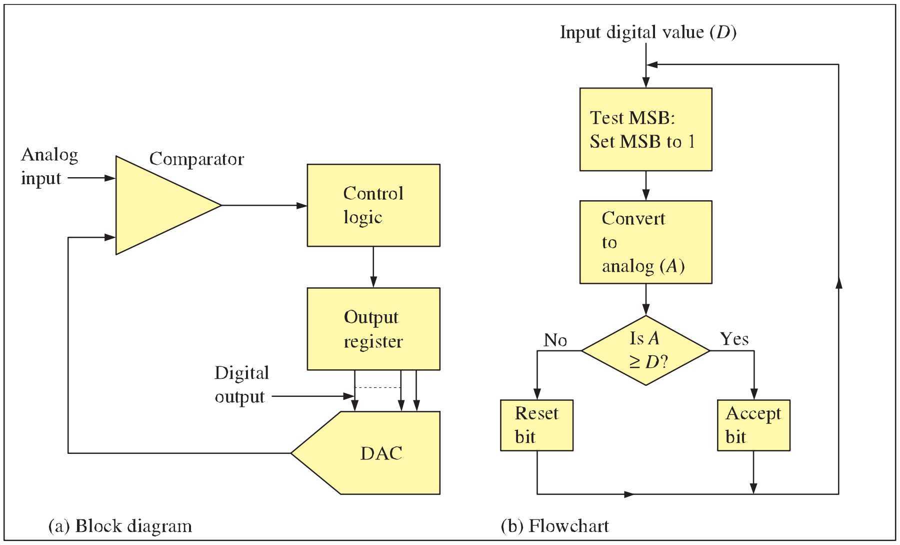

FIGURE 1–5 **Comparison of a block diagram and a flowchart.**

one part of the overall system. Figure 1–5(b) shows a flowchart diagram that breaks out the logic that the control logic block performs.

Block diagrams show the flow of a signal. The circuit uses a control logic circuit to determine an action to take based on the result of comparing the input (analog) and output (digital) signals. Flowcharts show the logic required to process the signal. The total system is best explained by showing both types of diagrams to clarify the signal flow and the logic.

Transfer Curve

A **transfer curve** is a plot that shows the ratio of the output to the input. In electronic systems, the transfer curve (or response curve) is useful because it describes how the system behaves for a given input. It can be used to illustrate the behavior of a given block or group of blocks. For example, the purpose of the amplifier block in the digital thermometer in Figure 1–4 is to make the small signal from the temperature sensor larger for the ADC. Ideally, the transfer curve for a linear amplifier is a straight line as shown in Figure 1–6. The input is a small voltage that is plotted along the x-axis. The output is a proportionally larger voltage that is plotted on the y-axis. For any small input voltage, the output voltage is larger by a factor known as the **gain**. In the example shown, the gain is 10 because for any given input voltage, the output is 10× larger.

The amplifier transfer curve is a ratio of two voltages; hence it is dimensionless and represents the voltage gain of the amplifier (V_{out}/V_{in}). Ideally, the transfer curve for a linear amplifier is a straight line, indicating constant gain for any input. In practice, amplifiers often approach this ideal for a certain range of inputs, but no amplifier is "perfect."

In some cases, the input and output may have different units; in this case, the transfer curve will not be dimensionless. For example, the transfer curve for a sensor might show a resistance change (a change in opposition to current) as a function of temperature. This type of transfer is shown in Figure 1–7 for a typical thermistor (a type of temperature sensor). Notice that the sensor has a nonlinear response, a point that is easy to spot on the transfer curve. If it is used as a sensor in the digital thermometer system, the data would need some form of processing to convert its output to a temperature reading that can then be displayed.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

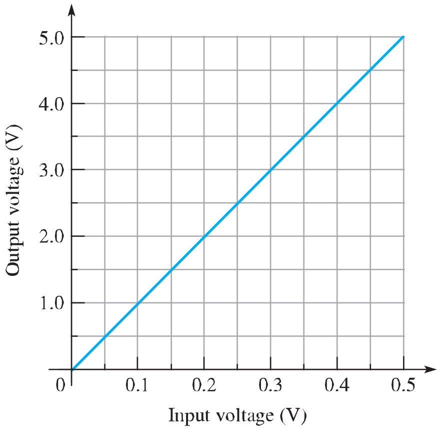

FIGURE 1–6 **Ideal transfer curve for an amplifier that has a gain of 10.**

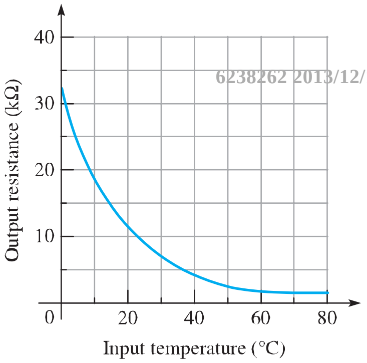

6238262 2013/12/10 12.27.66.5

FIGURE 1–7 **Transfer curve for a typical thermistor.**

SECTION 1–2 CHECKUP

1. What is the input and output of a digital thermometer system?
2. Consider the International Space Station as a system. What constitutes the environment of this system?
3. What is the purpose of a block diagram?
4. What name is given to a plot of the ratio of the output to the input for a given block?

1–3 TYPES OF CIRCUITS

Electrical systems deal with power, whereas electronic systems deal with signals. Electrical systems can be either dc (direct current) or ac (alternating current). Electronic systems generally require some form of power, so it is important to have an understanding of both electrical and electronic systems. Fortunately, most of the fundamental laws for both types of systems are the same. This section focuses on the types of circuits used in electrical and electronic systems.

After completing this section, you should be able to

- **Describe circuits in general terms**
  - **Explain the difference between active and passive components**
  - **Name applications for ac and dc circuits in electrical power distribution**
  - **Compare digital and analog electronic circuits**
  - **Define transducer**

Components

The word *circuit* is based on the Latin word *circuitus*, meaning to go around. Electrical circuits must have a complete path (like a circle) that starts at a source, includes a load, and returns to the source. When a circuit has a complete path, it is said to be closed; if the path is broken, it is said to be open. In addition to one or more sources, circuits use components, which are devices that alter the electrical properties of the circuit. Components that do not require electrical power to function are called **passive components**; those that require a source of power are called **active components**. Passive components cannot

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

increase the power in a signal; active components can. Most systems contain both types of components.

Three fundamental passive components are resistors, capacitors, and inductors. Other passive components are diodes, transformers, batteries, and motors. A battery is considered to be a passive component because it cannot increase the power in a signal (although it does supply electrical power).

Active components are generally classified as those that can increase signal power. They include transistors, operational amplifiers, and microprocessors. Sometimes the signal is not an input to the system but is produced internally. An electronic **oscillator** circuit is an example where the output is internally generated, but the device generating the output is considered an active device. Oscillators are circuits that generate a continuous output that can be one of several different shapes. They are commonly used in telecommunications, test instruments, and computers.

Electrical Circuits

Electrical circuits generate, deliver, and control power, which can be in the form of ac (alternating current) or dc (direct current).

AC CIRCUITS Alternating current (ac) changes polarity a certain number of times each second (the frequency). The power grids in the world are generally ac (with some exceptions). The most common way to produce ac for power delivery is to use a device called an **alternator** (an ac generator). Alternators are discussed in detail in Section 8–6.

AC can be transformed from low-to-high voltage and back easily and with lower expense than dc (however, some very high voltage systems use dc.) High voltages are much more efficient to transmit over long distances, so nearly all power distribution systems use ac. Historically, power companies developed and supplied power to their customers with no thought to interconnections. As a result, a hodgepodge of systems developed with different frequencies and voltages. Eventually, different parts of the world selected one of two frequencies as a standard. For most of the world, the standard frequency is 50 Hz (hertz is the unit of frequency); in the U.S. and North America, it is 60 Hz. In aircraft and certain military applications where weight is critical, 400 Hz is used because it allows smaller and lighter weight components to be used.

DC CIRCUITS Direct current (dc) is the form of electrical current that does not change polarity (unlike ac). Although most utility power is ac, there is high voltage dc (HVDC) technology used for some long distance transmission and underwater transmission. Although the conversion of dc to the very high voltages required for efficient transmission is more costly than ac, the tower and transmission lines can be designed for a lower cost per mile. For very long runs, dc can be cost effective. It is also useful for interfacing independent ac networks that are not synchronized. This occurs in Japan, where two completely different frequencies are in use (both 50 Hz and 60 Hz). DC is also generated for some low-voltage military and aircraft systems for supplying electrical power; typically 28 V is used.

Electronic Circuits

As you know, electronic circuits work with signals. Pure dc does not have signal information but is very important in electronic applications because it is required by active electronic circuits. For this reason, the study of dc circuits forms a foundation for anyone who wants to have an understanding of electrical or electronic systems. DC sources include batteries, fuel cells, solar cells, and generators. These sources are discussed in Section 2–3. A common method of obtaining dc is to convert ac to dc with a power supply. Power supplies are discussed in Section 3–7.

A signal carries information because of an intentional change in a parameter such as voltage or frequency. There are two major divisions for signal types. Signals that have

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

discrete levels are **digital** signals; continuous signals are **analog** signals. In some cases, the signal may be internally generated in the system. For example, Figure 1–8 shows a bicycle warning light. This is an example of a simple digital system in which a flashing signal is internally generated.

Analog signals vary continuously over an allowed range. As an example of an analog signal, consider an AM radio signal. The radio station transmits a high-frequency signal (called the *carrier*) whose amplitude is varied (or modulated) by a lower frequency carrying the information of interest. Figure 1–9 illustrates a simulated AM signal that is an example of an analog signal. In this case, the information is contained in the envelope.

Frequently, you will see electronic circuits that have analog and digital portions to the same circuit (such as the digital thermometer described in the previous section). The conversion from an analog signal to a digital signal is done with a converter (ADC), which often receives its input from a transducer. A **transducer** is a device that transforms energy from one form to another; for electronic systems one of the forms is an electrical signal. The temperature sensor in the digital thermometer system is an example of an input transducer. The transducer converts the temperature to a voltage, which is then converted to a digital signal for processing.

In some cases, such as audio, a digital signal may need to be converted to an analog signal. This is accomplished with a digital-to-analog converter (DAC). For example, a CD player takes a digital signal from the CD, processes it, and converts it to an analog output. The analog signal is amplified and sent to a speaker. The speaker is an output transducer, converting electrical energy to sound.

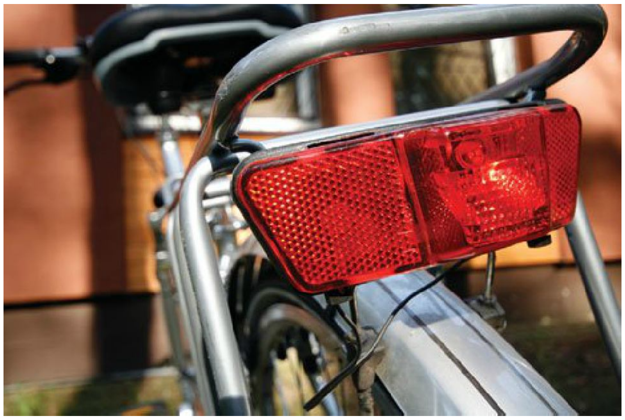

FIGURE 1–8 **Bicycle warning light.** (natenn/iStockphoto.com)

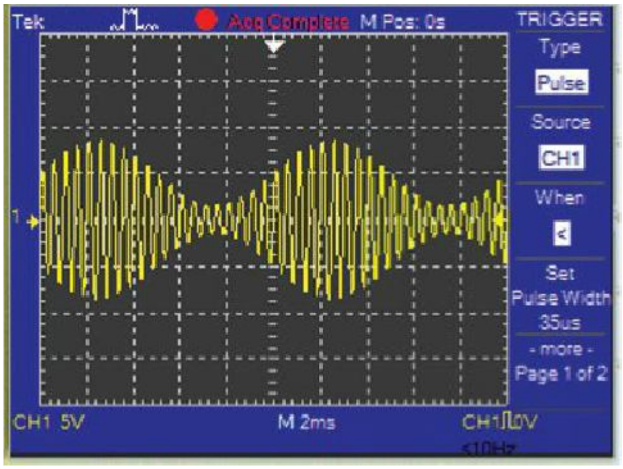

FIGURE 1–9 **A simulated AM radio wave.**

SECTION 1–3 CHECKUP

1. What is the difference between a passive and an active component?
2. Why is a battery not considered to be an electronic device?
3. What are the two frequencies that are used in most of the world's power grids?
4. Why is dc important to active circuits?
5. What is the difference between a digital and an analog circuit?
6. What is a transducer?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

1–4 SCIENTIFIC AND ENGINEERING NOTATION

Working with electrical and electronics systems, you will encounter both very small and very large quantities. For example, electrical current can range from hundreds of amperes in power applications to a few thousandths or millionths of an ampere in many electronic circuits. This range of values is typical of many other electrical quantities also. Engineering notation is a specialized form of scientific notation. It is used widely in technical fields to express large and small quantities. In electronics, engineering notation is used to express values of voltage, current, power, resistance, and other quantities.

After completing this section, you should be able to

- **Use scientific notation to represent quantities**
  - **Express any number using a power of ten**
  - **Perform calculations with powers of ten**

Scientific notation provides a convenient method for expressing large and small numbers and for performing calculations involving such numbers. In scientific notation, a quantity is expressed as a product of a number between 1 and 10 (one digit to the left of the decimal point) and a power of ten. For example, the quantity 150,000 is expressed in scientific notation as 1.5×10^5, and the quantity 0.00022 is expressed as 2.2×10^{-4}.

Powers of Ten

Table 1–1 lists some powers of ten, both positive and negative, and the corresponding decimal numbers. The **power of ten** is expressed as an *exponent* of the *base* 10 in each case.

$$\underset{\text{Base}\searrow}{10}{}^{\overset{\swarrow\text{Exponent}}{x}}$$

An **exponent** is a number to which a base number is raised. The exponent indicates the number of places that the decimal point is moved to the right or left to produce the decimal number. For a positive power of ten, move the decimal point to the right to get the equivalent decimal number. As an example, for an exponent of 4,

$$10^4 = 1 \times 10^4 = 1\underset{\curvearrowright}{.0000.} = 10{,}000.$$

TABLE 1–1 • Some positive and negative powers of ten.

$10^6 = 1{,}000{,}000$	$10^{-6} = 0.000001$
$10^5 = 100{,}000$	$10^{-5} = 0.00001$
$10^4 = 10{,}000$	$10^{-4} = 0.0001$
$10^3 = 1{,}000$	$10^{-3} = 0.001$
$10^2 = 100$	$10^{-2} = 0.01$
$10^1 = 10$	$10^{-1} = 0.1$
$10^0 = 1$	

For a negative power of ten, move the decimal point to the left to get the equivalent decimal number. As an example, for an exponent of −4,

6238262 2013/12/10 12.27.66.5

$$10^{-4} = 1 \times 10^{-4} = \underset{\curvearrowleft}{.0001.} = 0.0001$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The negative exponent does not indicate that a number is negative; it simply moves the decimal point to the left.

EXAMPLE 1–1

Express each number in scientific notation:

(a) 240 **(b)** 5100 **(c)** 85,000 **(d)** 3,350,000

SOLUTION

In each case, move the decimal point an appropriate number of places to the left to determine the positive power of ten.

(a) $240 = \mathbf{2.4 \times 10^2}$ **(b)** $5100 = \mathbf{5.1 \times 10^3}$

(c) $85{,}000 = \mathbf{8.5 \times 10^4}$ **(d)** $3{,}350{,}000 = \mathbf{3.35 \times 10^6}$

RELATED PROBLEM*

Express 750,000,000 in scientific notation.

***Answers are at the end of the chapter.**

EXAMPLE 1–2

Express each number in scientific notation:

(a) 0.24 **(b)** 0.005 **(c)** 0.00063 **(d)** 0.000015

SOLUTION

In each case, move the decimal point an appropriate number of places to the right to determine the negative power of ten.

(a) $0.24 = \mathbf{2.4 \times 10^{-1}}$ **(b)** $0.005 = \mathbf{5 \times 10^{-3}}$

(c) $0.00063 = \mathbf{6.3 \times 10^{-4}}$ **(d)** $0.000015 = \mathbf{1.5 \times 10^{-5}}$

RELATED PROBLEM

Express 0.00000093 in scientific notation.

EXAMPLE 1–3

Express each of the following numbers as a normal decimal number:

(a) 1×10^5 **(b)** 2.9×10^3 **(c)** 3.2×10^{-2} **(d)** 2.5×10^{-6}

SOLUTION

Move the decimal point to the right or left a number of places indicated by the positive or the negative power of ten respectively.

(a) $1 \times 10^5 = \mathbf{100{,}000}$ **(b)** $2.9 \times 10^3 = \mathbf{2900}$

(c) $3.2 \times 10^{-2} = \mathbf{0.032}$ **(d)** $2.5 \times 10^{-6} = \mathbf{0.0000025}$

RELATED PROBLEM

Express 8.2×10^8 as a normal decimal number.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Calculations With Powers of Ten

The advantage of scientific notation is in addition, subtraction, multiplication, and division of very small or very large numbers.

ADDITION The steps for adding numbers in powers of ten are as follows:

1. Express the numbers to be added in the same power of ten.
2. Add the numbers without their powers of ten to get the sum.
3. Bring down the common power of ten, which becomes the power of ten of the sum.

EXAMPLE 1–4

Add 2×10^6 and 5×10^7 and express the result in scientific notation.

SOLUTION

1. Express both numbers in the same power of ten: $(2 \times 10^6) + (50 \times 10^6)$.
2. Add $2 + 50 = 52$.
3. Bring down the common power of ten (10^6); the sum is $52 \times 10^6 = \mathbf{5.2 \times 10^7}$.

RELATED PROBLEM

Add 4.1×10^3 and 7.9×10^2.

SUBTRACTION The steps for subtracting numbers in powers of ten are as follows:

1. Express the numbers to be subtracted in the same power of ten.
2. Subtract the numbers without their powers of ten to get the difference.
3. Bring down the common power of ten, which becomes the power of ten of the difference.

EXAMPLE 1–5

Subtract 2.5×10^{-12} from 7.5×10^{-11} and express the result in scientific notation.

SOLUTION

1. Express each number in the same power of ten: $(7.5 \times 10^{-11}) - (0.25 \times 10^{-11})$.
2. Subtract $7.5 - 0.25 = 7.25$.
3. Bring down the common power of ten (10^{-11}); the difference is $\mathbf{7.25 \times 10^{-11}}$.

RELATED PROBLEM

Subtract 3.5×10^{-6} from 2.2×10^{-5}.

MULTIPLICATION The steps for multiplying numbers in powers of ten are as follows:

1. Multiply the numbers directly without their powers of ten.
2. Add the powers of ten algebraically (the exponents do not have to be the same).

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 1–6

Multiply 5×10^{12} by 3×10^{-6} and express the result in scientific notation.

SOLUTION

Multiply the numbers, and algebraically add the powers.

$$(5 \times 10^{12})(3 \times 10^{-6}) = 15 \times 10^{12+(-6)} = 15 \times 10^{6} = \mathbf{1.5 \times 10^{7}}$$

RELATED PROBLEM

Multiply 1.2×10^{3} by 4×10^{2}.

DIVISION The steps for dividing numbers in powers of ten are as follows:

1. Divide the numbers directly without their powers of ten.
2. Subtract the power of ten in the denominator from the power of ten in the numerator (the exponents do not have to be the same).

EXAMPLE 1–7

6238262 2013/12/10 12 27 66 5

Divide 5.0×10^{8} by 2.5×10^{3} and express the result in scientific notation.

SOLUTION

Write the division problem with a numerator and denominator.

$$\frac{5.0 \times 10^{8}}{2.5 \times 10^{3}}$$

Divide the numbers and subtract the powers of ten (3 from 8).

$$\frac{5.0 \times 10^{8}}{2.5 \times 10^{3}} = 2 \times 10^{8-3} = \mathbf{2 \times 10^{5}}$$

RELATED PROBLEM

Divide 8×10^{-6} by 2×10^{-10}.

SCIENTIFIC NOTATION ON A CALCULATOR Entering a number in scientific notation is accomplished on most calculators using the EE key as follows: Enter the number with one digit to the left of the decimal point, press EE, and enter the power of ten. This method requires that the power of ten be determined before entering the number. Some calculators can be placed in a mode that will automatically convert any decimal number entered into scientific notation.

EXAMPLE 1–8

Enter 23,560 in scientific notation using the EE key.

SOLUTION

Move the decimal point four places to the left so that it comes after the digit 2. This results in the number expressed in scientific notation as

$$2.3560 \times 10^{4}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Enter this number on your calculator as follows:

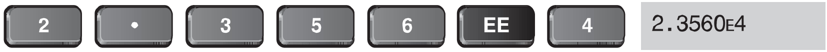

Note that it is not necessary to enter the zero.

RELATED PROBLEM

Enter the number 573,946 using the EE key.

Engineering Notation

Engineering notation is similar to scientific notation. However, in **engineering notation** a number can have from one to three digits to the left of the decimal point and the power-of-ten exponent must be a multiple of three. For example, the number 33,000 expressed in engineering notation is 33×10^3. In scientific notation, it is expressed as 3.3×10^4. As another example, the number 0.045 is expressed in engineering notation as 45×10^{-3}. In scientific notation, it is expressed as 4.5×10^{-2}. Engineering notation is useful in electrical and electronic calculations that use metric prefixes (discussed in Section 1–5).

EXAMPLE 1–9

Express the following numbers in engineering notation:

(a) 82,000 **(b)** 243,000 **(c)** 1,956,000

SOLUTION

In engineering notation,

(a) 82,000 is expressed as $\mathbf{82 \times 10^3}$.
(b) 243,000 is expressed as $\mathbf{243 \times 10^3}$.
(c) 1,956,000 is expressed as $\mathbf{1.956 \times 10^6}$.

RELATED PROBLEM

Express 36,000,000,000 in engineering notation.

EXAMPLE 1–10

Convert each of the following numbers to engineering notation:

(a) 0.0022 **(b)** 0.000000047 **(c)** 0.00033

SOLUTION

In engineering notation,

(a) 0.0022 is expressed as $\mathbf{2.2 \times 10^{-3}}$.
(b) 0.000000047 is expressed as $\mathbf{47 \times 10^{-9}}$.
(c) 0.00033 is expressed as $\mathbf{330 \times 10^{-6}}$.

RELATED PROBLEM

Express 0.0000000000056 in engineering notation.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

ENGINEERING NOTATION ON A CALCULATOR Use the EE key to enter the number with one, two, or three digits to the left of the decimal point, press EE, and enter the power of ten that is a multiple of three. This method requires that the appropriate power of ten be determined before entering the number.

EXAMPLE 1–11

Enter 51,200,000 in engineering notation using the EE key.

SOLUTION

Move the decimal point six places to the left so that it comes after the digit 1. This results in the number expressed in engineering notation as

$$51.2 \times 10^6$$

Enter this number on your calculator as follows:

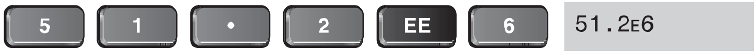

RELATED PROBLEM

Enter the number 273,900 in engineering notation using the EE key.

SECTION 1–4 CHECKUP

1. Scientific notation uses powers of ten. (True or False)
2. Express 100 as a power of ten.
3. Express the following numbers in scientific notation:
   - **(a)** 4350
   - **(b)** 12,010
   - **(c)** 29,000,000
4. Express the following numbers in scientific notation:
   - **(a)** 0.760
   - **(b)** 0.00025
   - **(c)** 0.000000597
5. Do the following operations:
   - **(a)** $(1 \times 10^5) + (2 \times 10^5)$
   - **(b)** $(3 \times 10^6)(2 \times 10^4)$
   - **(c)** $(8 \times 10^3) \div (4 \times 10^2)$
   - **(d)** $(2.5 \times 10^{-6}) - (1.3 \times 10^{-7})$
6. Enter the numbers expressed in scientific notation in Problem 3 into your calculator.
7. Express the following numbers in engineering notation:
   - **(a)** 0.0056
   - **(b)** 0.0000000283
   - **(c)** 950,000
   - **(d)** 375,000,000,000
8. Enter the numbers in Problem 7 into your calculator using engineering notation.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

1–5 UNITS AND METRIC PREFIXES

In electronics, you must deal with measurable quantities. For example, you must be able to express how many volts are measured at a certain test point in a circuit, how much current there is through a conductor, or how much power a certain amplifier delivers. In this section, you are introduced to the units and symbols for most of the electrical quantities that are used throughout the book. Metric prefixes are used in conjunction with engineering notation as a "shorthand" for the certain powers of ten that commonly are used.

After completing this section, you should be able to

- **Work with electrical units and metric prefixes**
  - **Name the units for twelve electrical quantities**
  - **Specify the symbols for the electrical units**
  - **List the metric prefixes**
  - **Change a power of ten in engineering notation to a metric prefix**
  - **Use metric prefixes to express electrical quantities**

Electrical Units

Letter symbols are used in electronics to represent both quantities and their units. One symbol is used to represent the name of the quantity, and another is used to represent the unit of measurement of that quantity. Table 1–2 lists the most important electrical quantities, along with their SI units and symbols. For example, italic P stands for *power* and nonitalic (roman) W stands for *watt,* which is the unit of power. In general, italic letters represent quantities and nonitalic letters represent units. Notice that energy is abbreviated with an italic W that represents *work;* and both *energy* and *work* have the same unit (the joule). The term **SI** is the French abbreviation for *International System* (*Système International* in French).

TABLE 1–2 • Electrical quantities and their corresponding units with SI symbols.

QUANTITY	SYMBOL	SI UNIT	SYMBOL
capacitance	C	farad	F
charge	Q	coulomb	C
conductance	G	siemens	S
current	I	ampere	A
energy or work	W	joule	J
frequency	f	hertz	Hz
impedance	Z	ohm	Ω
inductance	L	henry	H
power	P	watt	W
reactance	X	ohm	Ω
resistance	R	ohm	Ω
voltage	V	volt	V

In addition to the common electrical units shown in Table 1–2, the SI system has many other units that are defined in terms of certain fundamental units. In 1954, by international agreement, *meter, kilogram, second, ampere, degree kelvin,* and *candela* were

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

adopted as the basic SI units (*degree kelvin* was later changed to just *kelvin*). These units form the basis of the mks (for meter-kilogram-second) units that are used for derived quantities and have become the preferred units for nearly all scientific and engineering work. An older metric system, called the cgs system, was based on the centimeter, gram, and second as fundamental units. There are still a number of units in common use based on the cgs system; for example, the gauss is a magnetic flux unit in the cgs system and is still in common usage. In keeping with preferred practice, this text uses mks units, except when otherwise noted.

Metric Prefixes

In engineering notation **metric prefixes** represent each of the most commonly used powers of ten. These metric prefixes are listed in Table 1–3 with their symbols and corresponding powers of ten.

TABLE 1–3 • Metric prefixes with their symbols and corresponding powers of ten and values.

METRIC PREFIX	SYMBOL	POWER OF TEN	VALUE
femto	f	10^{-15}	one-quadrillionth
pico	p	10^{-12}	one-trillionth
nano	n	10^{-9}	one-billionth
micro	μ	10^{-6}	one-millionth
milli	m	10^{-3}	one-thousandth
kilo	k	10^{3}	one thousand
mega	M	10^{6}	one million
giga	G	10^{9}	one billion
tera	T	10^{12}	one trillion

Metric prefixes are used only with numbers that have a unit of measure, such as volts, amperes, and ohms, and precede the unit symbol. For example, 0.025 amperes can be expressed in engineering notation as 25×10^{-3} A. This quantity expressed using a metric prefix is 25 mA, which is read 25 milliamps. The metric prefix *milli* has replaced 10^{-3}. As another example, the Alpha Ventus offshore wind power system has 12 wind generators, each capable of delivering 5,000,000 W, which is expressed as 5.0 MW. The system generates 60,000,000 W, which is 60 MW. The metric prefix *mega* has replaced 10^{6} in expressing the power in both cases.

EXAMPLE 1–12

Express each quantity using a metric prefix:

(a) 50,000 V **(b)** 25,000,000 Ω **(c)** 0.000036 A

SOLUTION

(a) $50{,}000\ \text{V} = 50 \times 10^{3}\ \text{V} = \mathbf{50\ kV}$
(b) $25{,}000{,}000\ \Omega = 25 \times 10^{6}\ \Omega = \mathbf{25\ M\Omega}$
(c) $0.000036\ \text{A} = 36 \times 10^{-6}\ \text{A} = \mathbf{36\ \mu A}$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

RELATED PROBLEM

Express each quantity using metric prefixes:

(a) 56,000,000 Ω **(b)** 0.000470 A

SECTION 1–5 CHECKUP

1. List the metric prefix for each of the following powers of ten: 10^6, 10^3, 10^{-3}, 10^{-6}, 10^{-9}, and 10^{-12}.
2. Use a metric prefix to express 0.000001 A.
3. Use a metric prefix to express 250,000 W.

1–6 METRIC UNIT CONVERSIONS

It is sometimes necessary or convenient to convert a quantity from one unit with a metric prefix to another, such as from milliamperes (mA) to microamperes (μA). Moving the decimal point in the number an appropriate number of places to the left or to the right, depending on the particular conversion, results in a metric unit conversion.

After completing this section, you should be able to

- **Convert from one unit with a metric prefix to another**
  - **Convert between milli, micro, nano, and pico**
  - **Convert between kilo and mega**

The following basic rules apply to metric unit conversions:

1. When converting from a larger unit to a smaller unit, move the decimal point to the right.
2. When converting from a smaller unit to a larger unit, move the decimal point to the left.
3. Determine the number of places to move the decimal point by finding the difference in the powers of ten of the units being converted.

For example, when converting from milliamperes (mA) to microamperes (μA), move the decimal point three places to the right because there is a three-place difference between the two units (mA is 10^{-3} A and μA is 10^{-6} A). The following examples illustrate a few conversions.

EXAMPLE 1–13

Convert 0.15 milliampere (0.15 mA) to microamperes (μA).

SOLUTION

Move the decimal point three places to the right.

$$0.15 \text{ mA} = 0.15 \times 10^{-3} \text{ A} = 150 \times 10^{-6} \text{ A} = \mathbf{150\ \mu A}$$

RELATED PROBLEM

Convert 1 mA to microamperes.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 1–14

Convert 4500 microvolts (4500 μV) to millivolts (mV).

SOLUTION

Move the decimal point three places to the left.

$$4500\ \mu\text{V} = 4500 \times 10^{-6}\ \text{V} = 4.5 \times 10^{-3}\ \text{V} = \mathbf{4.5\ mV}$$

RELATED PROBLEM

Convert 1000 μV to millivolts.

EXAMPLE 1–15

Convert 5000 nanoamperes (5000 nA) to microamperes (μA).

SOLUTION

Move the decimal point three places to the left.

$$5000\ \text{nA} = 5000 \times 10^{-9}\ \text{A} = 5 \times 10^{-6}\ \text{A} = \mathbf{5\ \boldsymbol{\mu}A}$$

RELATED PROBLEM

Convert 893 nA to microamperes.

EXAMPLE 1–16

Convert 47,000 picofarads (47,000 pF) to microfarads (μF).

6238262 2013/12/10 12.27.66.5

SOLUTION

Move the decimal point six places to the left.

$$47{,}000\ \text{pF} = 47{,}000 \times 10^{-12}\ \text{F} = 0.047 \times 10^{-6}\ \text{F} = \mathbf{0.047\ \boldsymbol{\mu}F}$$

RELATED PROBLEM

Convert 10,000 pF to microfarads.

EXAMPLE 1–17

Convert 0.00022 microfarad (0.00022 μF) to picofarads (pF).

SOLUTION

Move the decimal point six places to the right.

$$0.00022\ \mu\text{F} = 0.00022 \times 10^{-6}\ \text{F} = 220 \times 10^{-12}\ \text{F} = \mathbf{220\ pF}$$

RELATED PROBLEM

Convert 0.0022 μF to picofarads.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 1–18

Convert 1800 kilohms (1800 kΩ) to megohms (MΩ).

SOLUTION

Move the decimal point three places to the left.

$$1800\ \text{k}\Omega = 1800 \times 10^3\ \Omega = 1.8 \times 10^6\ \Omega = \mathbf{1.8\ M\Omega}$$

RELATED PROBLEM

Convert 2.2 kΩ to megohms.

When adding (or subtracting) quantities with different metric prefixes, first convert one of the quantities to the same prefix as the other quantity.

EXAMPLE 1–19

Add 15 mA and 8000 μA and express the result in milliamperes.

SOLUTION

Convert 8000 μA to 8 mA and add.

$$15\ \text{mA} + 8000\ \mu\text{A} = 15\ \text{mA} + 8\ \text{mA} = \mathbf{23\ mA}$$

RELATED PROBLEM

Add 2873 mA and 10,000 μA.

SECTION 1–6 CHECKUP

1. Convert 0.01 MV to kilovolts (kV).
2. Convert 250,000 pA to milliamperes (mA).
3. Add 0.05 MW and 75 kW and express the result in kW.
4. Add 50 mV and 25,000 μV and express the result in mV.

1–7 MEASURED NUMBERS

Whenever a quantity is measured, there is uncertainty in the result due to limitations of the instruments used. When a measured quantity contains approximate numbers, the digits known to be correct are called significant digits. When reporting measured quantities, the number of digits that should be retained are the significant digits and no more than one uncertain digit.

After completing this section, you should be able to

- **Express measured data with the proper number of significant digits**
  - **Define *accuracy, error,* and *precision***
  - **Round numbers properly**

Error, Accuracy, and Precision

Data taken in experiments are not perfect because the accuracy of the data depends on the accuracy of the test equipment and the conditions under which the measurement was made. In order to properly report measured data, the error associated with the measurement

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

should be taken into account. Experimental error should not be thought of as a mistake. All measurements that do not involve counting are approximations of the true value. The difference between the true or best-accepted value of some quantity and the measured value is the **error**. A measurement is said to be accurate if the error is small. **Accuracy** is an indication of the range of error in a measurement. For example, if you measure thickness of a 10.00 mm gauge block with a micrometer and find that it is 10.8 mm, the reading is not accurate because a gauge block is considered to be a working standard. If you measure 10.02 mm, the reading is accurate because it is in reasonable agreement with the standard.

Calibrating Equipment

Precision and accuracy are important to understand if you are required to calibrate any piece of electronic equipment. *Calibration* is the process of comparing a given instrument to a standard. The standard is generally required to be more *precise* by a factor of 4 over the instrument being calibrated, but this alone does not assure accuracy. To assure accuracy, the calibration should be certifiable, and it needs to be traceable to national standards (in the U.S., this is National Institute of Standards and Technology). The standard requires periodic recertifications to assure its accuracy.

SYSTEM NOTE

Another term associated with the quality of a measurement is *precision*. **Precision** is a measure of the repeatability (or consistency) of a measurement of some quantity. It is possible to have a precise measurement in which a series of readings are not scattered, but each measurement is inaccurate because of an instrument error. For example, a meter may be out of calibration and produce inaccurate but consistent (precise) results. However, it is not possible to have an accurate instrument unless it is also precise.

Significant Digits

The digits in a measured number that are known to be correct are called **significant digits**. Most measuring instruments show the proper number of significant digits, but some instruments can show digits that are not significant, leaving it to the user to determine what should be reported. This may occur because of an effect called *loading* (discussed in Section 6–4). A meter can change the actual reading in a circuit by its very presence. It is important to recognize when a reading may be inaccurate; you should not report digits that are known to be inaccurate.

Another problem with significant digits occurs when you perform mathematical operations with numbers. The number of significant digits should never exceed the number in the original measurement. For example, if 1.0 V is divided by 3.0 Ω, a calculator will show 0.33333333. Since the original numbers each contain 2 significant digits, the answer should be reported as 0.33 A, the same number of significant digits.

The rules for determining if a reported digit is significant are

1. Nonzero digits are always considered to be significant.
2. Zeros to the left of the first nonzero digit are never significant.
3. Zeros between nonzero digits are always significant.
4. Zeros to the right of the decimal point for a decimal number are significant.
5. Zeros to the left of the decimal point with a whole number may or may not be significant depending on the measurement. For example, the number 12,100 Ω can have 3, 4, or 5 significant digits. To clarify the significant digits, scientific notation (or a metric prefix) should be used. For example, 12.10 kΩ has 4 significant digits.

When a measured value is reported, one uncertain digit may be retained but other uncertain digits should be discarded. To find the number of significant digits in a number, ignore the decimal point, and count the number of digits from left to right starting with the first nonzero digit and ending with the last digit to the right. All of the digits counted are significant except zeros to the right end of the number, which may or may not be significant.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

In the absence of other information, the significance of the right-hand zeros is uncertain. Generally, zeros that are placeholders, and not part of a measurement, are considered to be not significant. To avoid confusion, numbers should be shown using scientific or engineering notation if it is necessary to show the significant zeros.

EXAMPLE 1–20

Express the measured number 4300 with 2, 3, and 4 significant digits.

SOLUTION

Zeros to the right of the decimal point in a decimal number are significant. Therefore, to show two significant digits, write

$$\mathbf{4.3 \times 10^3}$$

To show three significant digits, write

$$\mathbf{4.30 \times 10^3}$$

To show four significant digits, write

$$\mathbf{4.300 \times 10^3}$$

RELATED PROBLEM

6238262 2013/12/10 12.27.66.5

How would you show the number 10,000 showing three significant digits?

EXAMPLE 1–21

Underline the significant digits in each of the following measurements:

(a) 40.0 **(b)** 0.3040 **(c)** 1.20×10^5 **(d)** 120,000 **(e)** 0.00502

SOLUTION

(a) $\underline{40.0}$ has three significant digits; see rule 4.

(b) $0.\underline{3040}$ has four significant digits; see rules 2 and 3.

(c) $\underline{1.20} \times 10^5$ has three significant digits; see rule 4.

(d) $\underline{120}{,}000$ has at least two significant digits. Although the number has the same value as in (c), zeros in this example are uncertain; see rule 5. This is *not* a recommended method for reporting a measured quantity; use scientific notation or a metric prefix in this case. See Example 1–20.

(e) $0.00\underline{502}$ has three significant digits; see rules 2 and 3.

RELATED PROBLEM

What is the difference between a measured quantity of 10 and 10.0?

Rounding Off Numbers

Since they always contain approximate numbers, measurements should be shown only with those digits that are significant plus no more than one uncertain digit. The number of digits shown is indicative of the precision of the measurement. For this reason, you should **round off** a number by dropping one or more digits to the right of the last significant digit. Use only the most significant dropped digit to decide how to round off. The rules for rounding off are

1. If the most significant digit dropped is greater than 5, increase the last retained digit by 1.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

2. If the digit dropped is less than 5, do not change the last retained digit.
3. If the digit dropped is 5, increase the last retained digit *if* it makes it an even number, otherwise do not. This is called the "round-to-even" rule.

EXAMPLE 1–22

Round each of the following numbers to three significant digits:

(a) 10.071 **(b)** 29.961 **(c)** 6.3948 **(d)** 123.52 **(e)** 122.52

SOLUTION

(a) 10.071 rounds to **10.1**.
(b) 29.961 rounds to **30.0**.
(c) 6.3948 rounds to **6.39**.
(d) 123.52 rounds to **124**.
(e) 122.52 rounds to **122**.

RELATED PROBLEM

Round 3.2850 to three significant digits using the round-to-even rule.

In most electrical and electronics systems and circuits, components have tolerances greater than 1% (5% and 10% are common). Most measuring instruments have accuracy specifications better than this, but it is unusual for measurements to be made with higher accuracy than 1 part in 1000. For this reason, three significant digits are appropriate for numbers that represent measured quantities in all but the most exacting work. If you are working with a problem with several intermediate results, keep all digits in your calculator, but round the answers to three when reporting a result.

SECTION 1–7 CHECKUP

1. What is the rule for showing zeros to the right of the decimal point?
2. What is the round-to-even rule?
3. On schematics, you will frequently see a 1000 Ω resistor listed as 1.0 kΩ. What does this imply about the value of the resistor?
4. If a power supply is required to be set to 10.00 V, what does this imply about the accuracy needed for the measuring instrument?
5. How can scientific or engineering notation be used to show the correct number of significant digits in a measurement?

1–8 ELECTRICAL SAFETY

Safety is a major concern when working with electrical systems. The possibility of an electric shock or a burn is always present, so caution should always be used. You provide a current path when voltage is applied across two points on your body, and current produces electrical shock. Electrical components often operate at high temperatures, so you can sustain skin burns when you come in contact with them. Also, the presence of electricity creates a potential fire hazard.

After completing this section, you should be able to

- **Recognize electrical hazards and practice proper safety procedures**
  - **Describe the cause of electrical shock**
  - **List the groups of current paths through the body**
  - **Discuss the effects of current on the human body**
  - **List the safety precautions that you should observe when you work with electricity**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Electrical Shock

Current through your body, not the voltage, is the cause of **electrical shock**. Of course, it takes voltage across a resistance to produce current. When a point on your body comes in contact with a voltage and another point comes in contact with a different voltage or with ground, such as a metal chassis, there will be current through your body from one point to the other. The path of the current depends on the points across which the voltage occurs. The severity of the resulting electrical shock depends on the amount of voltage and the path that the current takes through your body.

The current path through the body determines which tissues and organs will be affected. The current paths can be placed into three groups which are referred to as *touch potential, step potential,* and *touch/step potential.* These are illustrated in Figure 1–10.

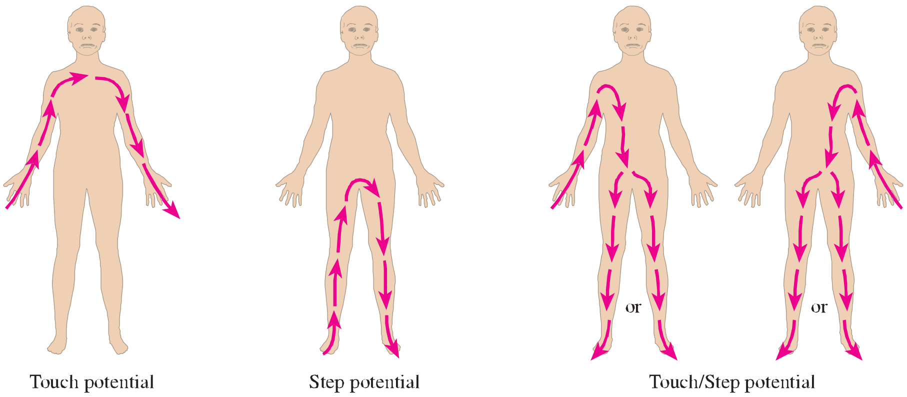

FIGURE 1–10 **Shock hazard in terms of three basic current path groups.**

EFFECTS OF CURRENT ON THE HUMAN BODY The amount of current is dependent on voltage and resistance. The human body has resistance that depends on many factors, which include body mass, skin moisture, and points of contact of the body with a voltage potential. Table 1–4 shows the effects for various values of current in milliamperes.

TABLE 1–4 • Physical effects of electrical current. Values vary depending on body mass.

CURRENT (mA)	PHYSICAL EFFECT
0.4	Slight sensation
1.1	Perception threshold
1.8	Shock, no pain, no loss of muscular control
9	Painful shock, no loss of muscular control
16	Painful shock, let-go threshold
23	Severe painful shock, muscular contractions, breathing difficulty
75	Ventricular fibrillation, threshold
235	Ventricular fibrillation, usually fatal for duration of 5 seconds or more
4,000	Heart paralysis (no ventricular fibrillation)
5,000	Tissue burn

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

BODY RESISTANCE Resistance of the human body is typically between 10 kΩ and 50 kΩ and depends on the two points between which it is measured. The moisture of the skin also affects the resistance between two points. The resistance determines the amount of voltage required to produce each of the effects listed in Table 1–4. For example, if you have a resistance of 10 kΩ between two given points on your body, 90 V across those two points will produce enough current (9 mA) to cause painful shock.

HANDS ON TIP
Receptacle testers are designed for use with specific receptacle types including specialized outlets. They can pinpoint problems such as open lines, faulty wiring, or reversed polarity; they show results with a lighted LED or neon bulb. Some testers are designed to test ground fault circuit interrupters (GFCI) for proper operation.

Utility Voltages

We tend to take utility voltages for granted, but they can be and have been lethal. It is best to be careful around any source of voltage (even low voltages can present a serious burn hazard). As a general rule, you should avoid working on any energized circuit, and check that the power is off with a known good meter. Most work in educational labs uses low voltages, but you should still avoid touching any energized circuit. If you are working on a circuit that is connected to utility voltages, the service should be disconnected, a notice should be placed on the equipment or place where the service is disconnected, and a padlock should be used to prevent someone from accidentally turning on the power. This procedure is called *lockout/tagout* and is widely used in industry. There are specific OSHA and industry standards for lockout/tagout.

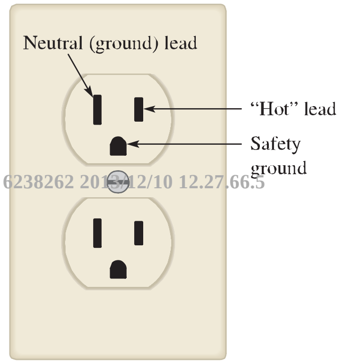

6238262 201[illegible]12/10 12.27.66.5

FIGURE 1–11 Standard receptacle and connections.

Most laboratory equipment is connected to the utility line ("ac") and in North America, this is 120 V rms (rms is discussed in Section 8–2). A faulty piece of equipment can cause the "hot" lead to inadvertently become exposed. You should inspect cords for exposed wires and check equipment for missing covers or other potential safety problems. The single-phase utility lines in homes and electrical laboratories use three insulated wires that are referred to as the "hot" (black or red wire), neutral (white wire), and safety ground (green wire). The hot and neutral wires will have current, but the green safety line should never have current in normal operation. The safety wire is connected to the metal exterior of encased equipment and is also connected to conduit and the metal boxes for housing receptacles. Figure 1–11 shows the location of these conductors on a standard receptacle. Notice on the receptacle that the neutral lead is larger than the hot lead.

The safety ground should be connected to the neutral at the service panel. The metal chassis of an instrument or appliance is also connected to ground. In the event that the hot wire is accidentally in contact with ground, the resulting high current should trip the circuit breaker or open a fuse to remove the hazard. However, a broken or missing ground lead may not have high current until it is contacted by a person. This danger is one obvious reason for ensuring that line cords have not been altered by removing the ground pin.

Many circuits are further protected with a special device called a ground-fault circuit interrupter (GFCI, which is sometimes called just GFI). If a fault occurs in a GFCI circuit, a sensor detects that the current in the hot line and the neutral line are not equal as they should be and trips the circuit breaker. The GFCI breaker is very fast acting and can trip faster than the breaker on the main panel. GFCI breakers are required in areas where a shock hazard exists such as wherever there is water or moisture. Pools, bathrooms, kitchens, basements, and garages should all have GFCI outlets. Figure 1–12 shows a ground-fault receptacle with reset and test buttons. When the test button is pressed, the circuit should immediately open. The reset button restores power.

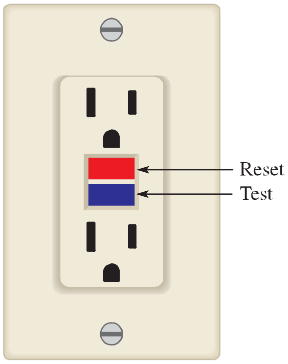

FIGURE 1–12 GFCI receptacle.

Safety Precautions

There are many practical things that you should do when you work with electrical and electronic equipment. Some important precautions are listed here.

- Avoid contact with any voltage source. Turn power off before you work on circuits when you need to touch circuit parts.
- Do not work alone. A telephone should be available for emergencies.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

- Do not work when tired or taking medications that make you drowsy.
- Remove rings, watches, and other metallic jewelry when you work on circuits.
- Do not work on equipment until you know proper procedures and are aware of potential hazards.
- Make sure power cords are in good condition and grounding pins are not missing or bent.
- Keep your tools properly maintained. Make sure the insulation on metal tool handles is in good condition.
- Handle tools properly and maintain a neat work area.
- Wear safety glasses when appropriate, particularly when soldering, clipping wires, or working with power tools.
- Always shut off power and discharge capacitors before you touch any part of a circuit with your hands.
- Know the location of the emergency power-off switch and emergency exits.
- Never try to override or tamper with safety devices such as an interlock switch or ground pin on a three-prong plug.
- Always wear shoes and keep them dry. Do not stand on metal or wet floors when working on electrical circuits.
- Never handle instruments when your hands are wet.
- Never assume that a circuit is off. Double-check it with a reliable meter before handling.
- Set the limiter on electronic power supplies to prevent currents larger than necessary to supply the circuit under test.
- Some devices such as capacitors can store a lethal charge for long periods after power is removed. They must be properly discharged before you work with them.
- When making circuit connections, always make the connection to the point with the highest voltage as your last step.
- Avoid contact with the terminals of power supplies.
- Always use wires with insulation and connectors or clips with insulating shrouds.
- Keep cables and wires as short as possible. Connect polarized components properly.
- Report any unsafe condition.
- Be aware of and follow all workplace and laboratory rules. Do not have drinks or food near equipment.
- If another person cannot let go of an energized conductor, switch the power off immediately. If that is not possible, use any available nonconductive material to try to separate the body from the contact.
- Use a lockout/tagout procedure to avoid someone turning power on while you are working on a circuit.

SAFETY NOTE

A GFCI outlet does not prevent shock or injury in all cases. If you are touching the hot and neutral wires without being grounded, no ground fault is detected and the GFCI breaker will not trip. In another case, the GFCI may prevent electrocution but not the initial electric shock before it interrupts the circuit. The initial shock could cause a secondary injury, such as from a fall.

SECTION 1–8 CHECKUP

1. What causes physical pain and/or damage to the body when electrical contact is made?
2. It's OK to wear a ring when working on an electrical circuit. (T or F)
3. Standing on a wet floor presents no safety hazard when working with electricity. (T or F)
4. A circuit can be rewired without removing the power if you are careful. (T or F)
5. Electrical shock can be extremely painful or even fatal. (T or F)
6. What does *GFCI* stand for?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SUMMARY

- Steps for design of a new product include using circuit design software, computer lay out, prototype construction, and testing.
- Many companies have tended to move away from vertical organization (raw material to finished product in one facility) to horizontal organization (focus on a specialty).
- The electronics industry can be divided into the manufacturing sector and the service sector.
- Many technical service sector jobs also require a broad range of skills including skills related to both electrical/electronic and mechanical installations.
- A system is a group of interrelated parts that perform a specific function. It is characterized by a boundary that separates the system from its environment. The system communicates to the outside world through inputs and outputs.
- A block diagram is a model of a system that represents the system's structure in a graphical format using labeled blocks to represent functions and lines to represent the signal flow.
- A transfer curve is the ratio of the output to the input for a system or circuit.
- Electrical and electronic circuits must have a complete path, one or more sources, and a load. Electrical and electronic circuits contain either passive components or active components.
- Passive components cannot increase the power in a signal; active components can.
- Signals that have discrete levels are called digital signals; continuous signals are analog signals.
- A transducer is a device that transforms energy from one form to another.
- Scientific notation is a method for expressing very large and very small numbers as a number between one and ten (one digit to left of decimal point) times a power of ten.
- Engineering notation is a form of scientific notation in which quantities are expressed with one, two, or three digits to the left of the decimal point times a power of ten that is a multiple of three.
- Metric prefixes are symbols used to represent powers of ten that are multiples of three.
- The uncertainty of a measured quantity depends on the accuracy and precision of the measurement.
- The number of significant digits in the result of a mathematical operation should never exceed the significant digits in the original numbers.
- Standard connections to electrical plugs include a hot wire, a neutral, and a safety ground.
- GFCI breakers sense the current in the hot wire and in the neutral wire and trip the breaker if they are different, indicating a ground fault.

KEY TERMS

Key terms and other bold terms in the chapter are defined in the end-of-book glossary.

Accuracy	An indication of the range of error in a measurement.
Active component	A component that requires electrical power to function correctly; active components can deliver more signal power than they receive from the input signal.
Alternator	An ac generator; alternators convert mechanical energy to electrical energy.
Analog	A continuous signal.
Block diagram	A model of a system that represents its structure in a graphical format using labeled blocks to represent functions and lines to represent the signal flow.
Boundary	The dividing line between what is part of the system and its environment.
Circuit	An interconnection of electrical components designed to produce a desired result.
Digital	A signal that has discrete levels.
Electrical shock	The physical sensation resulting from current through the body.
Engineering notation	A system for representing any number as a one-, two-, or three-digit number times a power of ten with an exponent that is a multiple of 3.
Error	The difference between the true or best-accepted value of some quantity and the measured value.
Exponent	The number to which a base number is raised.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Gain The ratio of the output to the input for an amplifier.

Horizontal organization A business structure that is decentralized, to allow managers to focus on a specialty and streamline decision making.

Input The voltage, current, or power applied to an electrical circuit to achieve a desired result.

Integrated circuit An active complex circuit consisting of resistors, transistors, and other components fabricated as a single unit that performs the function of multiple discrete components.

Mechatronics A synergistic combination of mechanics and electronics that includes instrumentation and control systems.

Metric prefix A symbol that is used to replace the power of ten in numbers expressed in engineering notation.

Oscillator A circuit where the output signal is internally generated; the output is a continuous signal that can be one of several different shapes.

Output The result obtained from the system after processing the inputs.

Passive component A component that does not require power; passive components cannot increase signal power.

Power of ten A numerical representation consisting of a base of 10 and an exponent; the number 10 raised to a power.

Precision A measure of the repeatability (or consistency) of a series of measurements.

Round off The process of dropping one or more digits to the right of the last significant digit in a number.

Scientific notation A system for representing any number as a number between 1 and 10 times an appropriate power of ten.

SI Standardized international system of units used for all engineering and scientific work; abbreviation for French *Le Systeme International d'Unites.*

Significant digit A digit known to be correct in a number.

System A group of interrelated parts that perform a specific function.

Transducer A device that transforms energy from one form to another.

Transfer curve A plot showing the ratio of the output to the input.

Vertical organization A centralized business structure that has an organized top-down approach. The tendency is less toward a specialization.

TRUE/FALSE QUIZ

Answers are at the end of the chapter.

6238262 2013/12/10 12.27.66.5

1. A vertical organization is characterized by decentralized decision making.
2. Technicians working in service and repair usually need to identify a faulty component and replace it.
3. Mechatronics is an applied engineering classification that is a synergistic combination of mechanics and electronics.
4. It is not possible to completely isolate a system from its environment.
5. A block diagram is the same thing as a flowchart.
6. The units for a transfer curve must be the same for input and output.
7. A passive device cannot have power gain.
8. An example of a transducer is a loudspeaker.
9. The number 3300 is written as 3.3×10^3 in both scientific and engineering notation.
10. A negative number that is expressed in scientific notation will always have a negative exponent.
11. When you multiply two numbers written in scientific notation, the exponents need to be the same.
12. When you divide two numbers written in scientific notation, the exponent of the denominator is subtracted from the exponent of the numerator.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

13. The metric prefix *micro* has an equivalent power of ten equal to 10^6.

14. To express 56×10^6 with a metric prefix, the result is 56 M.

15. 0.047 μF is equal to 47 nF.

16. The number of significant digits in the number 0.0102 is three.

17. When you apply the *round-to-even* rule to round off 26.25 to three digits, the result is 26.3.

18. The white neutral lead for ac power should have the same current as the hot lead.

SELF-TEST

Answers are at the end of the chapter.

1. A PLC is a
(a) printed logic circuit **(b)** power limiting circuit
(c) programmable logic controller **(d)** peak limit controller

2. The boundary of a system separates
(a) inputs from outputs
(b) the system from its environment
(c) only the inputs from the environment
(d) only the outputs from the environment

3. For a system, a block diagram is a graphical tool to illustrate
(a) functions and signal flow (b) details of the system
(c) power supplies (d) logical processes

4. The transfer curve always
(a) is a straight line **(b)** is the ratio of output to input
(c) is dimensionless **(d)** all of the above

5. An active component is a device that
(a) supplies its own power **(b)** supplies power to other components
(c) does not have an output **(d)** can increase signal power

6. An example of an ac source is a
(a) battery **(b)** alternator **(c)** solar cell **(d)** fuel cell

7. The quantity 4.7×10^3 is the same as
(a) 470 **(b)** 4700 **(c)** 47,000 **(d)** 0.0047

8. The quantity 56×10^{-3} is the same as
(a) 0.056 **(b)** 0.560 **(c)** 560 **(d)** 56,000

9. The number 3,300,000 can be expressed in engineering notation as
(a) 3300×10^3 **(b)** 3.3×10^{-6} **(c)** 3.3×10^6 **(d)** either (a) or (c)

10. Ten milliamperes can be expressed as
(a) 10 MA **(b)** 10 μA **(c)** 10 kA **(d)** 10 mA

11. Five thousand volts can be expressed as
(a) 5000 V **(b)** 5 MV **(c)** 5 kV **(d)** either (a) or (c)

12. Twenty million ohms can be expressed as
(a) 20 mΩ **(b)** 20 MW **(c)** 20 MΩ **(d)** 20 $\mu\Omega$

13. 15,000 W is the same as
(a) 15 mW **(b)** 15 kW **(c)** 15 MW **(d)** 15 μW

14. Which of the following is not an electrical quantity?
(a) current **(b)** voltage **(c)** time **(d)** power

15. The unit of current is
(a) volt **(b)** watt **(c)** ampere **(d)** joule

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

16. The unit of voltage is
(a) ohm **(b)** watt **(c)** volt **(d)** farad

17. The unit of resistance is
(a) ampere **(b)** henry **(c)** hertz **(d)** ohm

18. Hertz is the unit of
(a) power **(b)** inductance **(c)** frequency **(d)** time

19. The number of significant digits in the number 0.1050 is
(a) two **(b)** three **(c)** four **(d)** five

PROBLEMS

Answers to odd-numbered problems are at the end of the book.

BASIC PROBLEMS AND QUESTIONS

SECTION 1–1 The Electronics Industry

1. If an engineer has an idea for improving a circuit board, what steps should be taken before the change is implemented?

2. Cite examples of products made by system manufacturers.

3. Cite the fundamental reasons that technician work has shifted from troubleshooting and repair at the component level to a system approach.

SECTION 1–2 Introduction to Electronic Systems

4. What are key differences between an electrical system and an electronic system?

5. What are two advantages to converting an analog signal into a digital signal?

6. What is the primary difference between a block diagram and a flowchart?

SECTION 1–3 Types of Circuits

7. (a) What is an oscillator? (b) How does it differ from most circuits?

8. (a) What do the initials HVDC stand for? (b) Where is it used?

9. What is a carrier?

10. The transfer function for a device is a plot of the output to the input. What does the transfer function look like for an ADC?

SECTION 1–4 Scientific and Engineering Notation

11. Express each of the following numbers in scientific notation:
(a) 3000 **(b)** 75,000 **(c)** 2,000,000

12. Express each fractional number in scientific notation:
(a) 1/500 **(b)** 1/2000 **(c)** 1/5,000,000

13. Express each of the following numbers in scientific notation:
(a) 8400 **(b)** 99,000 **(c)** 0.2×10^6

14. Express each of the following numbers in scientific notation:
(a) 0.0002 **(b)** 0.6 **(c)** 7.8×10^{-2}

15. Express each of the following as a regular decimal number:
(a) 2.5×10^{-6} **(b)** 5.0×10^2 **(c)** 3.9×10^{-1}

16. Express each number in regular decimal form:
(a) 4.5×10^{-6} **(b)** 8×10^{-9} **(c)** 4.0×10^{-12}

17. Add the following numbers:
(a) $(9.2 \times 10^6) + (3.4 \times 10^7)$ **(b)** $(5 \times 10^3) + (8.5 \times 10^{-1})$
(c) $(5.6 \times 10^{-8}) + (4.6 \times 10^{-9})$

18. Perform the following subtractions:
(a) $(3.2 \times 10^{12}) - (1.1 \times 10^{12})$ **(b)** $(2.6 \times 10^8) - (1.3 \times 10^7)$
(c) $(1.5 \times 10^{-12}) - (8 \times 10^{-13})$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

19. Perform the following multiplications:
(a) $(5 \times 10^3)(4 \times 10^5)$ **(b)** $(1.2 \times 10^{12})(3 \times 10^2)$
(c) $(2.2 \times 10^{-9})(7 \times 10^{-6})$

20. Divide the following:
(a) $(1.0 \times 10^3) \div (2.5 \times 10^2)$ **(b)** $(2.5 \times 10^{-6}) \div (5.0 \times 10^{-8})$
(c) $(4.2 \times 10^8) \div (2 \times 10^{-5})$

21. Express each number in engineering notation:
(a) 89,000 **(b)** 450,000 **(c)** 12,040,000,000,000

22. Express each number in engineering notation:
(a) 2.35×10^5 **(b)** 7.32×10^7 **(c)** 1.333×10^9

23. Express each number in engineering notation:
(a) 0.000345 **(b)** 0.025 **(c)** 0.00000000129

6238262 2013/12/10 12.27.66.5

24. Express each number in engineering notation:
(a) 9.81×10^{-3} **(b)** 4.82×10^{-4} **(c)** 4.38×10^{-7}

25. Add the following numbers and express each result in engineering notation:
(a) $2.5 \times 10^{-3} + 4.6 \times 10^{-3}$ **(b)** $68 \times 10^6 + 33 \times 10^6$
(c) $1.25 \times 10^6 + 250 \times 10^3$

26. Multiply the following numbers and express each result in engineering notation:
(a) $(32 \times 10^{-3})(56 \times 10^3)$ **(b)** $(1.2 \times 10^{-6})(1.2 \times 10^{-6})$
(c) $100(55 \times 10^{-3})$

27. Divide the following numbers and express each result in engineering notation:
(a) $50 \div (2.2 \times 10^3)$ **(b)** $(5 \times 10^3) \div (25 \times 10^{-6})$
(c) $(560 \times 10^3) \div (660 \times 10^3)$

SECTION 1–5 Units and Metric Prefixes

28. Express each number in Problem 11 in ohms using a metric prefix.

29. Express each number in Problem 13 in amperes using a metric prefix.

30. Express each of the following as a quantity having a metric prefix:
(a) 31×10^{-3} A **(b)** 5.5×10^3 V **(c)** 20×10^{-12} F

31. Express the following using metric prefixes:
(a) 3×10^{-6} F **(b)** $3.3 \times 10^6\ \Omega$ **(c)** 350×10^{-9} A

32. Express each quantity with a power of ten:
(a) 5 μA **(b)** 43 mV **(c)** 275 kΩ **(d)** 10 MW

SECTION 1–6 Metric Unit Conversions

33. Perform the indicated conversions:
(a) 5 mA to microamperes **(b)** 3200 μW to milliwatts
(c) 5000 kV to megavolts **(d)** 10 MW to kilowatts

34. Determine the following:
(a) The number of microamperes in 1 milliampere
(b) The number of millivolts in 0.05 kilovolt
(c) The number of megohms in 0.02 kilohm
(d) The number of kilowatts in 155 milliwatts

35. Add the following quantities:
(a) 50 mA + 680 μA **(b)** 120 kΩ + 2.2 MΩ **(c)** 0.02 μF + 3300 pF

36. Do the following operations:
(a) 10 kΩ ÷ (2.2 kΩ + 10 kΩ) **(b)** 250 mV ÷ 50 μV **(c)** 1 MW ÷ 2 kW

SECTION 1–7 Measured Numbers

37. How many significant digits are in each of the following numbers:
(a) 1.00×10^3 **(b)** 0.0057 **(c)** 1502.0
(d) 0.000036 **(e)** 0.105 **(f)** 2.6×10^2

38. Round each of the following numbers to three significant digits. Use the "round-to-even" rule.
(a) 50,505 **(b)** 220.45 **(c)** 4646 **(d)** 10.99 **(e)** 1.005

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

ANSWERS TO SECTION CHECKUPS

SECTION 1–1 The Electronics Industry

1. Simulation software can enable a circuit to be optimized with minimal cost. It can look for potential problems such as timing, noise, or thermal problems.
2. Manufacturing and service
3. Communications, computers, renewable energy, industrial controls, consumer products, medical equipment, transportation, defense.
4. Certification allows individuals to showcase their skills and gives employers a means of judging a person's competency.

SECTION 1–2 Introduction to Electronic Systems

1. The digital thermometer extracts heat from its surroundings, so the input is heat; the output is the displayed temperature.
2. The environment of the ISS includes the gravitational field of the earth and the other orbiting material around earth, and the radiated heat and light from the sun.
3. A block diagram shows the relationship of functional blocks and the signal flow paths.
4. A transfer curve is a plot that shows the ratio of the output to the input.

SECTION 1–3 Types of Circuits

1. A passive component does not require electrical power to function and cannot increase the power in a signal. An active component does require electrical power to function correctly and can increase the power in a signal.
2. A battery can only supply electrical power; it cannot increase the power in a signal.
3. 50 Hz and 60 Hz
4. Normally it is dc that is converted to signal power by an active circuit.
5. A digital circuit is one that processes discrete signals; an analog circuit is one that processes continuous signals.
6. A transducer is a device that transforms energy from one form to another.

SECTION 1–4 Scientific and Engineering Notation

1. True
2. 10^2
3. **(a)** 4.35×10^3 **(b)** 1.201×10^4 **(c)** 2.9×10^7
4. **(a)** 7.6×10^{-1} **(b)** 2.5×10^{-4} **(c)** 5.97×10^{-7}
5. **(a)** 3×10^5 **(b)** 6×10^{10} **(c)** 2×10^1 **(d)** 2.37×10^{-6}
6. Enter the digits, press EE, and enter the power of ten.
7. **(a)** 5.6×10^{-3} **(b)** 28.3×10^{-9} **(c)** 950×10^3 **(d)** 375×10^9
8. Enter the digits, press EE, and enter the power of ten.

SECTION 1–5 Units and Metric Prefixes

1. Mega (M), kilo (k), milli (m), micro (μ), nano (n), and pico (p)
2. 1 μA (one microampere)
3. 250 kW (250 kilowatts)

SECTION 1–6 Metric Unit Conversions

1. 0.01 MV = 10 kV
2. 250,000 pA = 0.00025 mA
3. 125 kW
4. 75 mV

SECTION 1–7 Measured Numbers

1. Zeros should be retained only if they are significant because if they are shown, they are considered significant.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

2. If the digit dropped is 5, increase the last retained digit *if* it makes it even, otherwise do not.
3. A zero to the right of the decimal point implies that the resistor is accurate to the nearest 100 Ω (0.1 kΩ).
4. The instrument must be accurate to four significant digits.
5. Scientific and engineering notation can show any number of digits to the right of a decimal. Numbers to the right of the decimal are always considered significant.

SECTION 1–8 Electrical Safety

1. Current
2. F
3. F
4. F
5. T
6. Ground-fault circuit interrupter

ANSWERS TO RELATED PROBLEMS FOR EXAMPLES

1–1 7.5×10^8
1–2 9.3×10^{-7}
1–3 820,000,000
1–4 4.89×10^3
1–5 1.85×10^{-5}
1–6 4.8×10^5
1–7 4×10^4
1–8 Enter 5.73946; press EE, enter 5.
1–9 36×10^9
1–10 5.6×10^{-12}
1–11 Enter 273.9, press EE, enter 3.
1–12 **(a)** 56 MΩ **(b)** 470 μA
1–13 1000 μA
1–14 1 mV
1–15 0.893 μA
1–16 0.01 μF
1–17 2200 pF
1–18 0.0022 MΩ
1–19 2883 mA
1–20 10.0×10^3
1–21 The number 10 has two significant digits; the number 10.0 has three.
1–22 3.28

ANSWERS TO TRUE/FALSE QUIZ

1. F **2.** F **3.** T **4.** T **5.** F **6.** F **7.** T **8.** T **9.** T
10. F **11.** F **12.** T **13.** F **14.** T **15.** T **16.** T **17.** F **18.** T

ANSWERS TO SELF-TEST

1. (c) **2.** (b) **3.** (a) **4.** (b) **5.** (d) **6.** (b) **7.** (b)
8. (a) **9.** (c) **10.** (d) **11.** (d) **12.** (c) **13.** (b) **14.** (c)
15. (c) **16.** (c) **17.** (d) **18.** (c) **19.** (c)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

CHAPTER 2

VOLTAGE, CURRENT, AND RESISTANCE

6238262 2013/12/10 12.27.66.5

OUTLINE

2–1 Atoms
2–2 Electrical Charge
2–3 Voltage
2–4 Current
2–5 Resistance
2–6 The Electric Circuit
2–7 Basic Circuit Measurements

OBJECTIVES

- Describe the basic structure of an atom
- Explain the concept of electrical charge
- Define *voltage* and discuss its characteristics
- Define *current* and discuss its characteristics
- Define *resistance* and discuss its characteristics
- Describe a basic electric circuit
- Make basic circuit measurements

KEY TERMS

Atom
Electron
Free electron
Conductor
Semiconductor
Insulator
Charge
Coulomb's law
Coulomb (C)
Voltage
Volt (V)
Voltage source
Fuel cell
Current
Ampere (A)
Current source
Resistance
Ohm (Ω)
Conductance
Siemens (S)
Resistor
Potentiometer
Rheostat
Load
Schematic
Closed circuit
Open circuit
Switch
Fuse
Circuit breaker
AWG
Ground
Voltmeter
Ammeter
Ohmmeter
DMM

INTRODUCTION

Three basic electrical quantities presented in this chapter are voltage, current, and resistance. No matter what type of electrical or electronic system you may work with, these quantities will always be of primary importance. This is true for dc and ac circuits, but our focus will be dc circuits in Part 1 of this book. Because of its importance in electrical systems, an ac circuit may be used occasionally to illustrate a particular concept; however, in these special cases the analysis and calculations are the same as for an equivalent dc circuit.

To help you understand voltage, current, and resistance, the basic structure of the atom is discussed and the concept of charge is introduced. The basic electric circuit is studied, along with techniques for measuring voltage, current, and resistance.

VISIT THE WEBSITE
Study aids for this chapter are available at
http://pearsonhighered.com/floyd

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

2–1 ATOMS

All matter is made of atoms; and all atoms consist of electrons, protons, and neutrons. The configuration of certain electrons in an atom is the key factor in determining how well a conductive or semiconductive material conducts electric current.

After completing this section, you should be able to

- **Describe the basic structure of an atom**
  - **Define *nucleus, proton, neutron,* and *electron***
  - **Define *atomic number***
  - **Define *shell***
  - **Explain what a valence electron is**
  - **Describe ionization**
  - **Explain what a free electron is**
  - **Define *conductor, semiconductor,* and *insulator***

An **atom** is the smallest particle of an **element** that retains the characteristics of that element. Each of the known 118 elements has atoms that are different from the atoms of all other elements. This gives each element a unique atomic structure. According to the classic Bohr model, an atom is visualized as having a planetary type of structure that consists of a central nucleus surrounded by orbiting electrons, as illustrated in Figure 2–1. The **nucleus** consists of positively charged particles called **protons** and uncharged particles called **neutrons**. The basic particles of negative charge are called **electrons**. Electrons orbit the nucleus.

Each type of atom has a certain number of protons that distinguishes it from the atoms of all other elements. For example, the simplest atom is that of hydrogen, which has one proton and one electron, as pictured in Figure 2–2(a). As another example, the helium atom, shown in Figure 2–2(b), has two protons and two neutrons in the nucleus and two electrons orbiting the nucleus.

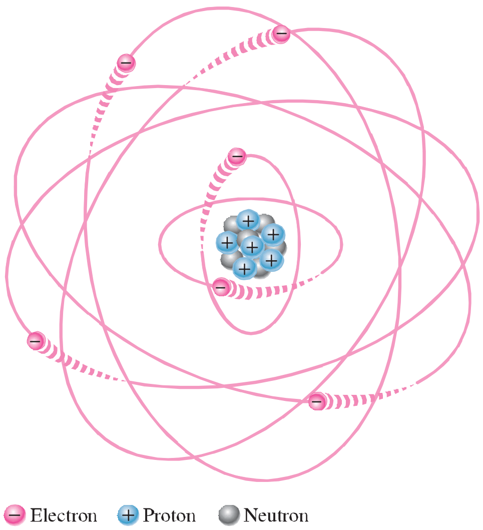

FIGURE 2–1 **The Bohr model of an atom showing electrons in circular orbits around the nucleus. The "tails" on the electrons indicate they are moving.**

Atomic Number

All elements are arranged in the periodic table of the elements in order according to their **atomic number**. The atomic number equals the number of protons in the nucleus. For example, hydrogen has an atomic number of 1 and helium has an atomic number of 2. In their normal (or neutral) state, all atoms of a given element have the same number of electrons as protons; the positive charges cancel the negative charges, and the atom has a net charge of zero, making it electrically balanced.

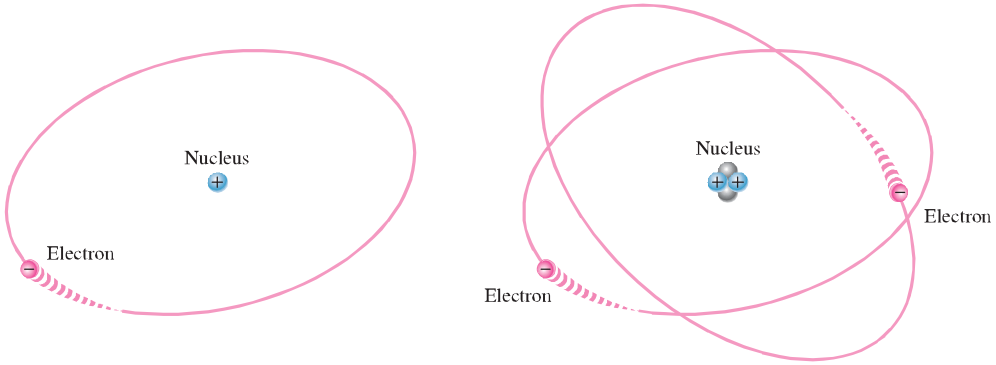

(a) Hydrogen atom

(b) Helium atom

FIGURE 2–2 **The two simplest atoms, hydrogen and helium.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

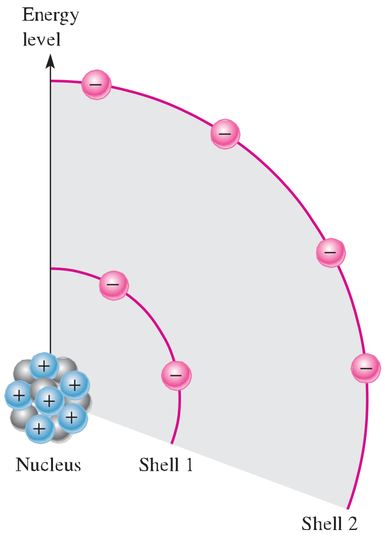

FIGURE 2–3 **Energy levels increase as the distance from the nucleus increases.**

Electron Shells and Orbits

Electrons orbit the nucleus of an atom at certain distances from the nucleus. Electrons near the nucleus have less energy than those in more distant **orbits**. It is known that only discrete (separate and distinct) values of electron energies exist within atomic structures. Therefore, electrons must orbit only at discrete distances from the nucleus.

ENERGY LEVELS Each discrete distance (orbit) from the nucleus corresponds to a certain energy level. In an atom, the orbits are grouped into energy bands known as **shells**. A given atom has a fixed number of shells. Each shell has a fixed maximum number of electrons at permissible energy levels (orbits). The shells are designated 1, 2, 3, and so on, with 1 being closest to the nucleus. This energy band concept is illustrated in Figure 2–3, which shows two energy levels. Additional shells may exist in other types of atoms, depending on the element.

The number of electrons in each shell follows a predictable pattern according to the formula, $2N^2$, where N is the number of the shell. The first shell of any atom ($N = 1$) can have up to two electrons, the second shell ($N = 2$) up to eight electrons, the third shell up to 18 electrons, and the fourth shell up to 32 electrons. In many elements, electrons start filling the fourth shell after eight electrons are in the third shell.

Valence Electrons

Electrons that are in orbits farther from the nucleus have higher energy and are less tightly bound to the atom than those closer to the nucleus. This is because the force of attraction between the positively charged nucleus and the negatively charged electron decreases with increasing distance from the nucleus. Electrons with the highest energy levels exist in the outermost shell of an atom and are relatively loosely bound to the atom. This outermost shell is known as the **valence** shell, and electrons in this shell are called **valence electrons**. These valence electrons contribute to chemical reactions and bonding within the structure of a material, and they determine a material's electrical properties.

Free Electrons and Ions

If an electron absorbs a **photon** of sufficient energy, the electron escapes from the atom and becomes a **free electron**. Any time an atom or group of atoms is left with a net charge, the atom or group of atoms is called an **ion**. When an electron escapes from the neutral hydrogen atom (designated H), the atom is left with a net positive charge and becomes a *positive ion* (designated H^+). In other cases, an atom or group of atoms can acquire one or more electrons, in which case it is called a *negative ion*.

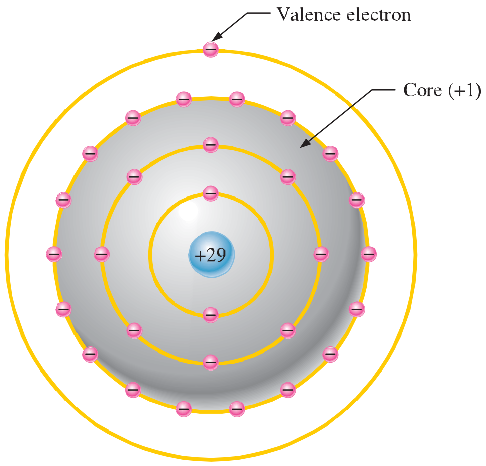

FIGURE 2–4 **The copper atom.**

The Copper Atom

Copper is the most commonly used metal in electrical systems. The copper atom has 29 electrons that orbit the nucleus in four shells, as shown in Figure 2–4. Notice that the fourth or outermost shell, the valence shell, has only one valence electron. The inner shells are called the core. When the valence electron in the outer shell of the copper atom gains sufficient thermal energy, it can break away from the parent atom and become a free electron. In a piece of copper at room temperature, a "sea" of these free electrons is present. These electrons are not bound to a given atom but are free to move in the copper material. Free electrons make copper an excellent conductor and make electrical current possible.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Categories of Materials

Three categories of materials are used in electronics: conductors, semiconductors, and insulators.

CONDUCTORS **Conductors** are materials that readily allow current. They have a large number of free electrons and are characterized by one to three valence electrons in their structure. Most metals are good conductors. Silver is the best conductor, and copper is next. Copper is the most widely used conductive material because it is less expensive than silver. Copper wire is commonly used as a conductor in electric circuits.

SEMICONDUCTORS **Semiconductors** are classed below the conductors in their ability to carry current because they have fewer free electrons than do conductors. Semiconductors have four valence electrons in their atomic structures. However, because of their unique characteristics, certain semiconductor materials are the basis for electronic devices such as the diode, transistor, and integrated circuit. Silicon and germanium are common semiconductive materials.

INSULATORS **Insulators** are nonmetallic materials that are poor conductors of electric current; they are used to prevent current where it is not wanted. Insulators have no free electrons in their structure. The valence electrons are bound to the nucleus and not considered "free." Although nonmetal elements are generally considered to be insulators, most practical insulators used in electrical and electronic systems are compounds such as glass, porcelain, Teflon, and polyethylene, to name a few.

SECTION 2–1 CHECKUP*

1. What is the basic particle of negative charge?

2. Define *atom*.

3. What does an atom consist of?

4. Define *atomic number*.

5. Do all elements have the same types of atoms?

6. What is a free electron?

7. What is a shell in the atomic structure?

8. Name two conductive materials.

6238262 2013/12/10 12.27.66.5

***Answers are at the end of the chapter.**

2–2 ELECTRICAL CHARGE

As you know, an electron is the smallest particle that exhibits negative electrical charge. When an excess of electrons exists in a material, there is a net negative electrical charge. When a deficiency of electrons exists, there is a net positive electrical charge.

After completing this section, you should be able to

- **Explain the concept of electrical charge**
  - **Name the unit of charge**
  - **Name the types of charge**
  - **Describe the forces between charges**
  - **Determine the amount of charge on a given number of electrons**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The charge of an electron and that of a proton are equal in magnitude and opposite in sign. Electrical **charge** is an electrical property of matter that exists because of an excess or deficiency of electrons. Charge is symbolized by Q. Static electricity is the presence of a net positive or negative charge in a material. Everyone has experienced the effects of static electricity from time to time, for example, when attempting to touch a metal surface or another person or when the clothes in a dryer cling together.

Materials with charges of opposite polarity are attracted to each other; materials with charges of the same polarity are repelled, as indicated symbolically in Figure 2–5. A force acts between charges, as evidenced by the attraction or repulsion. This force, called an *electric field,* consists of invisible lines of force as represented in Figure 2–6.

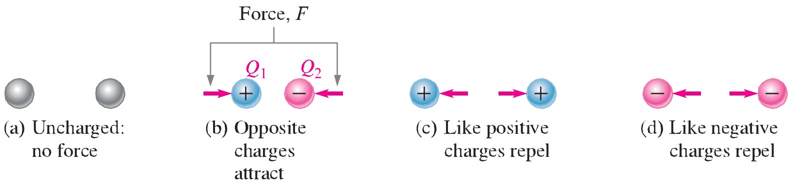
Force, F
Q1
Q2
+
−
(a) Uncharged: no force
(b) Opposite charges attract
(c) Like positive charges repel
(d) Like negative charges repel

FIGURE 2–5 **Attraction and repulsion of electrical charges.**

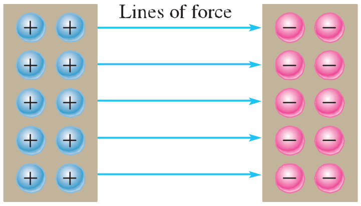

FIGURE 2–6 **Electric field between two oppositely charged surfaces.**

Coulomb's law states

A force (F) exists between two point-source charges (Q_1, Q_2) that is directly proportional to the product of the two charges and inversely proportional to the square of the distance (d) between the charges.

Coulomb: The Unit of Charge

Electrical charge is measured in coulombs, symbolized by C.

One coulomb is the total charge possessed by 6.25×10^{18} electrons.

A single electron has a charge of 1.6×10^{-19} C. The total charge Q, expressed in coulombs, for a given number of electrons is found by the following formula:

$$Q = \frac{\text{number of electrons}}{6.25 \times 10^{18}\ \text{electrons/C}} \qquad (2\text{–}1)$$

Positive and Negative Charge

Consider a neutral atom—that is, one that has the same number of electrons and protons and thus has no net charge. As you know, when a valence electron is pulled away from the atom by the application of energy, the atom is left with a net positive charge (more protons than electrons) and becomes a positive ion. A **positive ion** is defined as an atom or group of atoms with a net positive charge. If an atom acquires an extra electron in its outer shell, it has a net negative charge and becomes a negative ion. A **negative ion** is defined as an atom or group of atoms with a net negative charge.

The amount of energy required to free a valence electron is related to the number of electrons in the outer shell. An atom can have up to eight valence electrons. The more complete the outer shell, the more stable the atom and thus the more energy is required to remove an electron. Figure 2–7 illustrates the creation of a positive ion and a negative ion when a hydrogen atom gives up its single valence electron to a chlorine atom, forming gaseous hydrogen chloride (HCl). When the gaseous HCl is dissolved in water, hydrochloric acid is formed.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

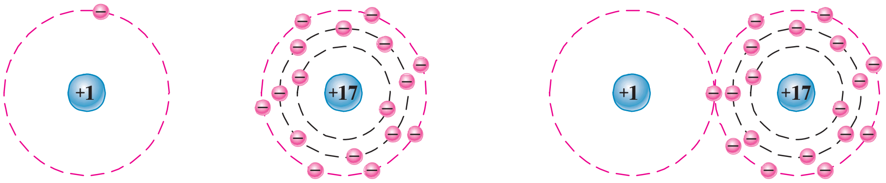

Hydrogen atom
(1 proton, 1 electron)

Chlorine atom
(17 protons, 17 electrons)

(a) The neutral hydrogen atom has a single valence electron.

(b) The atoms combine by sharing the valence electron to form gaseous hydrogen chloride (HCl).

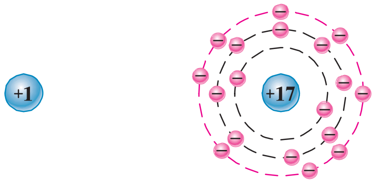

Positive hydrogen ion
(1 proton, no electrons)

Negative chloride ion
(17 protons, 18 electrons)

(c) When dissolved in water, hydrogen chloride gas separates into positive hydrogen ions and negative chloride ions. The chlorine atom retains the electron given up by the hydrogen atom forming both positive and negative ions in the same solution.

FIGURE 2–7 **Example of the formation of positive and negative ions.**

EXAMPLE 2–1

How many coulombs of charge do 93.8×10^{16} electrons represent?

SOLUTION

$$Q = \frac{\text{number of electrons}}{6.25 \times 10^{18}\ \text{electrons/C}} = \frac{93.8 \times 10^{16}\ \text{electrons}}{6.25 \times 10^{18}\ \text{electrons/C}} = 15 \times 10^{-2}\ \text{C} = \mathbf{0.15\ C}$$

RELATED PROBLEM*

How many electrons does it take to have 3 C of charge?

***Answers are at the end of the chapter.**

SECTION 2–2 CHECKUP

1. What is the symbol used for charge?
2. What is the unit of charge, and what is the unit symbol?
3. What are the two types of charge?
4. How much charge, in coulombs, is there in 10×10^{12} electrons?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

2–3 VOLTAGE

As you have learned, a force of attraction exists between a positive and a negative charge. A certain amount of energy must be exerted in the form of work to overcome the force and move the charges a given distance apart. All opposite charges possess a certain potential energy because of the separation between them. The difference in potential energy of the charges is the potential difference or *voltage*. Voltage is the driving force in electric circuits and is what establishes current.

6238262 2013/12/10 12.27.66.5

After completing this section, you should be able to

- **Define *voltage* and discuss its characteristics**
  - **State the formula for voltage**
  - **Name and define the unit of voltage**
  - **Describe the basic sources of voltage**

Voltage is defined as energy per unit of charge and is expressed as

$$V = \frac{W}{Q} \qquad (2\text{–}2)$$

where V is voltage in volts (V), W is energy in joules (J), and Q is charge in coulombs (C). As a simple analogy, you can think of voltage as corresponding to the pressure difference created by a pump that causes water to flow through a pipe in a closed water system.

Volt: The Unit of Voltage

The unit of voltage is the **volt**, symbolized by V.

> **One volt is the potential difference (voltage) between two points when one joule of energy is used to move one coulomb of charge from one point to the other.**

EXAMPLE 2–2

If 50 J of energy are required to move 10 C of charge, what is the voltage?

SOLUTION

$$V = \frac{W}{Q} = \frac{50\ \text{J}}{10\ \text{C}} = \mathbf{5\ V}$$

RELATED PROBLEM

How much energy is required to move 50 C from one point in a circuit to another when the voltage between the two points is 12 V?

The DC Voltage Source

A **voltage source** provides electrical energy or electromotive force (emf), more commonly known as voltage. Voltage is produced by means of chemical energy, light energy, and magnetic energy combined with mechanical motion.

THE IDEAL DC VOLTAGE SOURCE An ideal dc voltage source can provide a constant voltage for any current required by a circuit. The ideal voltage source does not exist but can be closely approximated in practice. We will assume ideal unless otherwise specified.

Voltage sources can be either dc or ac. A common symbol for a dc voltage source is shown in Figure 2–8.

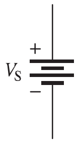

FIGURE 2–8 **Symbol for dc voltage sources.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

A graph showing current versus voltage for an ideal dc voltage source is shown in Figure 2–9. As you can see, the voltage is constant for any current from the source. For a practical (nonideal) voltage source connected in a circuit, the voltage decreases slightly as the current increases as shown by the dashed line. Current is always drawn from a voltage source when a load such as a resistance is connected to it.

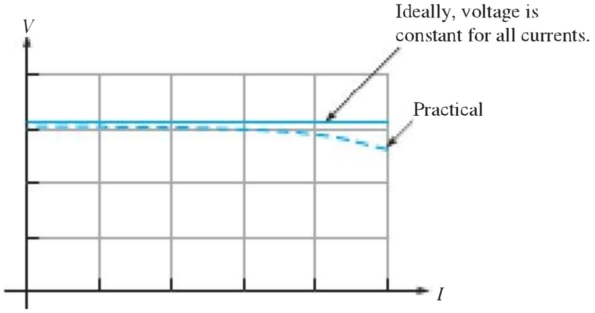

FIGURE 2–9 **Voltage source graph.**

Types of DC Voltage Sources

BATTERIES A **battery** is a type of voltage source that is composed of one or more cells that convert chemical energy directly into electrical energy. As you know, work (or energy) per charge is the basic unit for voltage, and a battery adds energy to each unit of charge. It is something of a misnomer to talk about "charging a battery" because a battery does not store charge but rather stores chemical potential energy. All batteries use a specific type of chemical reaction called an *oxidation-reduction* reaction. In this type of reaction, electrons are transferred from one reactant to the other. If the chemicals used in the reaction are separated, it is possible to cause the electrons to travel in the external circuit, creating current. As long as there is an external path for the electrons, the reaction can proceed, and stored chemical energy is converted to electrical current. If the path is broken, the reaction stops and the battery is said to be in equilibrium. In a battery, the terminal that supplies electrons has a surplus of electrons and is the negative electrode or **anode**. The electrode that acquires electrons has a positive potential and is the **cathode**.

Figure 2–10 shows a nonrechargeable single-cell copper-zinc battery that we will use for illustration of battery operation. The copper-zinc cell is simple to construct and illustrates concepts common to all nonrechargeable batteries. A zinc electrode and a copper electrode are immersed in solutions of zinc sulfate ($ZnSO_4$) and copper sulfate ($CuSO_4$), which are separated by a salt bridge that prevents the Cu^{2+} ions from reacting directly with the Zn metal. The zinc metal electrode supplies Zn^{2+} ions to the solution and electrons to the external circuit, so this electrode is constantly eaten away as the reaction proceeds. The salt bridge allows ions to pass through it to maintain charge balance in the cell. There are no free electrons in the solutions, so an external path for electrons is provided through an ammeter (in our case) or other load. On the cathode side, the electrons that were given up by the zinc combine with copper ions from the solution to form copper metal, which deposits on the copper electrode. The chemical reactions (shown in the diagram) occur at

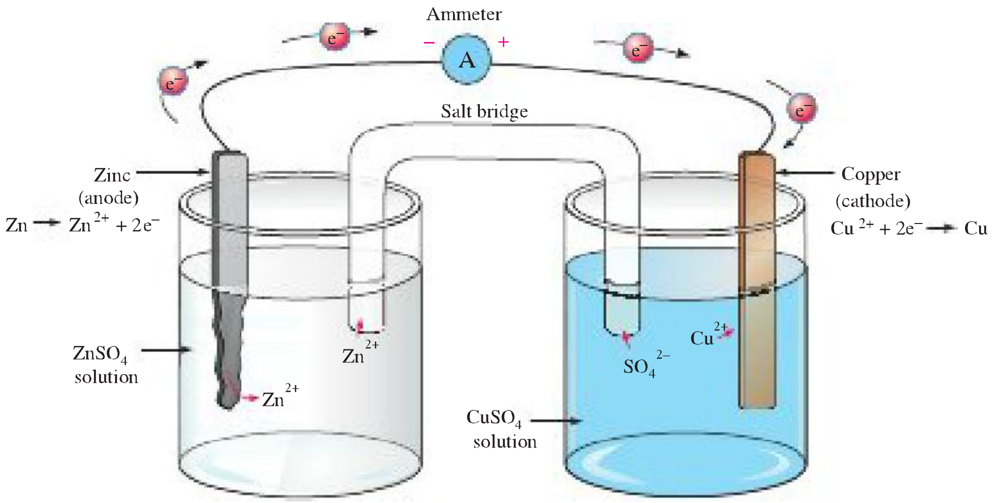

FIGURE 2–10 **A copper-zinc battery. The reaction can only occur if an external path is provided for the electrons. As the reaction proceeds, the Zn anode is eaten away and Cu^{2+} ions combine with electrons to form copper metal on the cathode.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

the electrode. Different types of batteries have different reactions, but all involve transfer of electrons in the external circuit and migration of ions internally as the battery discharges.

A single cell will have a certain fixed voltage. In the copper-zinc cell, the voltage is 1.1 V. In a lead-acid cell, the kind used in car batteries, a potential difference of about 2.1 V is between the anode and cathode. A typical car battery has six such cells connected in series. The voltage of any cell depends on the cell chemistry. Nickel-cadmium cells are about 1.2 V and lithium cells can be as high as almost 4 V, depending on the second reactant. Cell chemistry also determines the shelf life and discharge characteristic for a battery. For example, a lithium-MnO_2 battery typically has five times the shelf life as a comparable carbon-zinc battery.

Although the voltage of a battery cell is fixed by its chemistry, the capacity is variable and depends on the quantity of materials in the cell. Essentially, the *capacity* of a cell is the number of electrons that can be obtained from it and is measured by the amount of current (defined in Section 2–4) that can be supplied over time.

SAFETY NOTE

Lead-acid batteries can be dangerous because sulfuric acid is highly corrosive and battery gases (primarily hydrogen) are explosive. The acid in the battery can cause serious eye damage if it contacts the eye and can cause skin burns or destroy clothing. You should always wear eye protection when working on or around batteries and wash well after handling batteries.

When taking a battery out of service, disconnect the ground end first (this is the negative side on all modern cars). This prevents current and avoids a potential spark when the positive lead is disconnected.

Battery Backup

Batteries are used in many systems to provide redundancy, such as in the case of alarm systems where the main power may be off at a critical time, or to back up utility power in case of failure. In solar-energy systems, batteries are used to supply power when the sun is not available. Deep-cycle batteries are constructed with heavier plates than normal batteries and are designed to be discharged further than regular automotive batteries. However, deep-cycle batteries tend to be expensive, inefficient, and require regular maintenance.

Photo courtesy of NREL.

SYSTEM NOTE

Batteries normally consist of multiple cells that are electrically connected together internally. The way that the cells are connected and the type of cells determine the voltage and current capacity of the battery. If the positive electrode of one cell is connected to the negative electrode of the next and so on, as illustrated in Figure 2–11(a), the battery voltage is the sum of the individual cell voltages. This is called a *series connection.* To increase

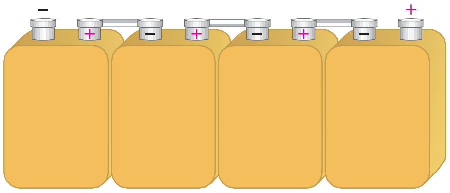

(a) Series-connected cells increase voltage.

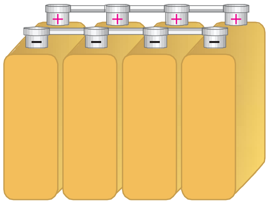

(b) Parallel-connected cells increase current capacity.

FIGURE 2–11 **Cells connected to form batteries.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

battery current capacity, the positive electrodes of several cells are connected together and all the negative electrodes are connected together, as illustrated in Figure 2–11(b). This is called a *parallel connection.*

There are many sizes of batteries; large batteries, with more material, can supply more current. In addition to the many sizes and shapes, batteries are classified according to their chemical makeup and if they are rechargeable or not. Primary batteries are not rechargeable and are discarded when they run down because their chemical reactions are irreversible; secondary batteries are reusable because their chemical reactions are reversible. The following are some important types of batteries:

- **Alkaline-MnO_2.** This is a primary battery that is commonly used in palm-type computers, photographic equipment, toys, radios, and recorders. It has a longer shelf life and a higher power density than a carbon-zinc battery.
- **Carbon-zinc.** This is a primary multiuse battery for flashlights and small appliances. It is available in a variety of sizes such as AAA, AA, C, and D.
- **Lead-acid.** This is a secondary (rechargable) battery that is commonly used in automotive, marine, and other similar applications. Deep-cycle lead-acid batteries are commonly used as system backup types.
- **Lithium-ion.** This is a secondary battery that is commonly used in all types of portable electronics. This type of battery is increasingly being used in defense, aerospace, and automotive applications.
- **Lithium-MnO_2.** This is a primary battery that is commonly used in photographic and electronic equipment, smoke alarms, personal organizers, memory backup, and communications equipment.
- **Nickel-metal hydride.** This is a secondary (rechargable) battery that is commonly used in portable computers, cell phones, camcorders, and other portable consumer electronics.
- **Silver oxide.** This is a primary battery that is commonly used in watches, photographic equipment, hearing aids, and electronics requiring high-capacity batteries.
- **Zinc air.** This is a primary battery that is commonly used in hearing aids, medical monitoring instruments, pagers, and similar applications.

HANDS ON TIP

To store a lead-acid battery for an extended period, it should be fully charged and placed in a cool, dry location that is protected from freezing or excessive heat. A battery will self-discharge over time, so it needs to be periodically checked and recharged when it is less than 70% fully charged. Battery manufacturers will have specific recommendations for storage on their websites.

FUEL CELLS A **fuel cell** is a device that converts electrochemical energy into dc voltage directly. Fuel cells combine a fuel (usually hydrogen) with an oxidizing agent (usually oxygen). In the hydrogen fuel cell, hydrogen and oxygen react to form water, which is the only by-product. The process is clean, quiet, and more efficient than burning. Fuel cells and batteries are similar in that they both use an oxidation-reduction chemical reaction in which electrons are forced to travel in the external circuit. However, a battery is a closed system with all its chemicals stored inside, whereas in a fuel cell, the hydrogen and oxygen constantly flow into the cell where they combine and produce electricity.

Hydrogen fuel cells are usually classified by their operating temperature and the type of electrolyte they use. Some types work well for use in stationary power generation plants. Others may be useful for small portable applications or for powering cars. For example, the type that holds the most promise for automotive applications is the polymer exchange membrane fuel cell (PEMFC), which is a type of hydrogen fuel cell. A simplified diagram is shown in Figure 2–12 to illustrate the basic operation.

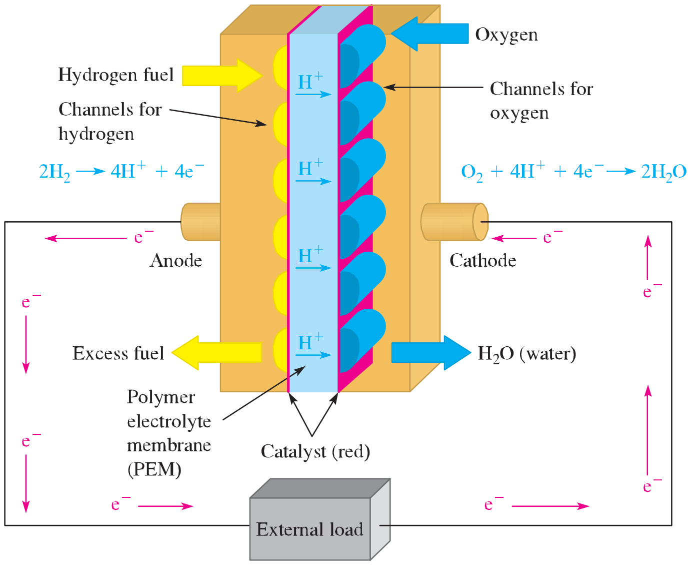

FIGURE 2–12 **Simplified diagram of a fuel cell.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The channels disperse pressurized hydrogen gas and oxygen gas equally over the surface of the catalyst, which facilitates the reaction of the hydrogen and oxygen. When an H_2 molecule comes in contact with the platinum catalyst on the anode side of the fuel cell, it splits into two H^+ ions and two electrons (e^-). The hydrogen ions are passed through the polymer electrolyte membrane (PEM) onto the cathode. The electrons pass through the anode and into the external circuit to create current.

When an O_2 molecule comes in contact with the catalyst on the cathode side, it breaks apart, forming two oxygen ions. The negative charge of these ions attracts two H^+ ions through the electrolyte membrane and together they combine with electrons from the external circuit to form a water molecule (H_2O), which is passed from the cell as a by-product. In a single fuel cell, this reaction produces only about 0.7 V. To get higher voltages, multiple fuel cells are connected in series.

Current research on fuel cells is ongoing and is focused on developing reliable, smaller, and cost-effective components for vehicles and other applications. The conversion to fuel cells also requires research on how best to obtain and provide hydrogen fuel where it is needed. Potential sources for hydrogen include using solar, geothermal, or wind energy to break apart water. Hydrogen can also be obtained by breaking down coal or natural gas molecules, which are rich in hydrogen. Various demonstration projects for using hydrogen fuel cells have been completed. For example, the mail-processing center at Anchorage, AK, uses five 200 kW fuel cells to power the building.

SOLAR CELLS The operation of solar cells is based on the **photovoltaic effect**, which is the process whereby light energy is converted directly into electrical energy. A basic solar cell consists of two layers of different types of semiconductive materials joined together to form a junction. When one layer is exposed to light, many electrons acquire enough energy to break away from their parent atoms and cross the junction. This process forms negative ions on one side of the junction and positive ions on the other, and thus a potential difference (voltage) is developed. Figure 2–13 shows the construction of a basic solar cell.

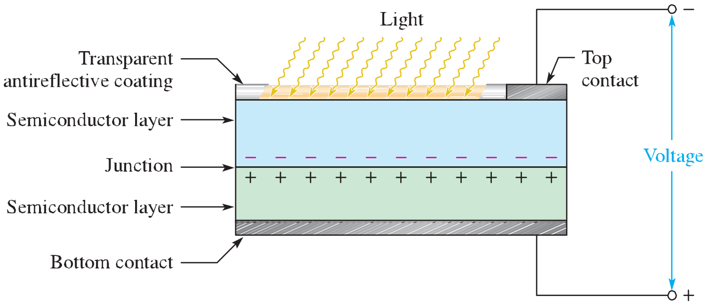

FIGURE 2–13 **Construction of a basic solar cell.**

Although solar cells can be used in room light for powering a calculator, research is focusing more on converting sunlight to electricity. There is considerable research in increasing the efficiency of solar cells and photovoltaic (PV) modules today because they are a very clean source of energy using sunlight. A complete system for continuous power generally requires a battery backup to provide energy when the sun is not shining. Solar cells are well suited for remote locations where energy sources are unavailable and are used in providing power to satellites.

Scientists are working to develop flexible solar cells that can be printed by a banknote printing process. The technology uses organic cells that can be mass-produced cheaply by printing them in the same manner as money is printed. “Printable electronics” is at the forefront of polymer technology research.

GENERATOR Electrical **generators** convert mechanical energy into electrical energy using a principle called *electromagnetic induction* (see Chapter 7). A coil is rotated

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

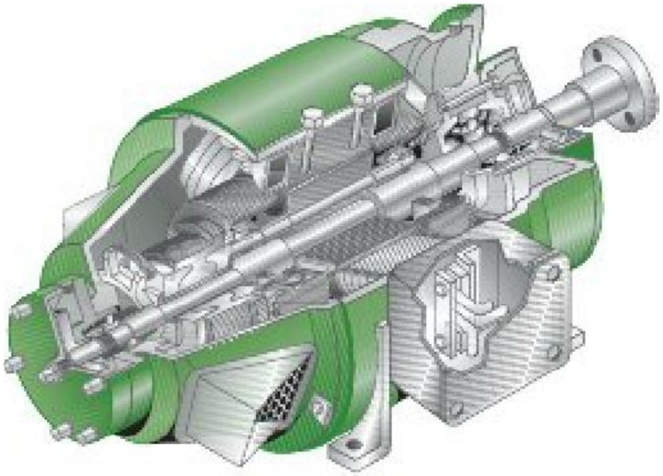

FIGURE 2–14 **Cutaway view of a dc voltage generator.**

through a magnetic field, and a voltage is produced across the coil. A typical generator is pictured in Figure 2–14.

THE POWER SUPPLY **Power supplies** convert the ac voltage from the wall outlet to a constant (dc) voltage that is available across two terminals. A basic commercial power supply is shown in Figure 2–15. Power supplies are covered in more detail in Section 3–7.

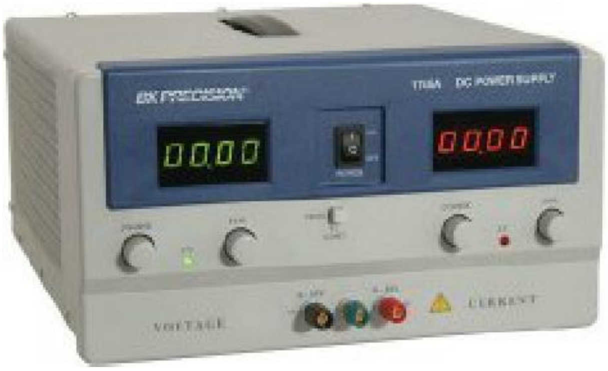

FIGURE 2–15 **A basic power supply.** **(Courtesy of B+K Precision)**

THERMOCOUPLES The **thermocouple** is a thermoelectric type of voltage source that is commonly used to sense temperature. A thermocouple is formed by the junction of two dissimilar metals, and its operation is based on the **Seebeck effect** that describes the voltage generated at the junction of the metals as a function of temperature.

Standard types of thermocouples are characterized by the specific metals used. These standard thermocouples produce predictable output voltages for a range of temperatures. The most common is type K, made of chromel and alumel. Other types are also designated by letters as E, J, N, B, R, and S. Most thermocouples are available in wire or probe form.

PIEZOELECTRIC SENSORS These sensors act as very low-power voltage sources and are based on the **piezoelectric effect** where a voltage is generated when a piezoelectric material is mechanically deformed by an external force. Quartz and ceramic are two types of piezoelectric material. Piezoelectric sensors are used in various systems, such as pressure sensors, force sensors, accelerometers, microphones, and ultrasonic devices.

SECTION 2–3 CHECKUP

1. Define *voltage*.
2. What is the unit of voltage?
3. What is the voltage when 24 J of energy are required to move 10 C of charge?
4. List seven sources of voltage.
5. What type of chemical reaction occurs in all batteries and fuel cells?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

2–4 CURRENT

Voltage provides energy to electrons that allows them to move through a circuit. This movement of electrons is the current, which results in work being done in an electric circuit.

After completing this section, you should be able to

- **Define *current* and discuss its characteristics**
  - **Explain the movement of electrons**
  - **State the formula for current**
  - **Name and define the unit of current**

As you have learned, free electrons are available in all conductive and semiconductive materials. These outer-shell electrons drift randomly in all directions, from atom to atom, within the structure of the material, as indicated in Figure 2–16. These electrons are loosely bound to the positive metal ions in the material; but because of thermal energy, they are free to move about the crystalline structure of the metal.

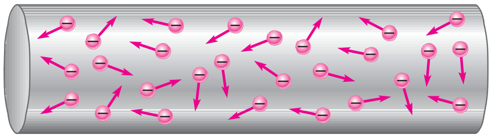

FIGURE 2–16 **Random motion of free electrons in a material.**

If a voltage is placed across a conductive or semiconductive material, one end becomes positive and the other negative, as indicated in Figure 2–17. The repulsive force produced by the negative voltage at the left end causes the free electrons (negative charges) to move toward the right. The attractive force produced by the positive voltage at the right end pulls the free electrons to the right. The result is a net movement of the free electrons from the negative end of the material to the positive end, as shown in Figure 2–17.

6238262 2013/12/10 12.27.66.5

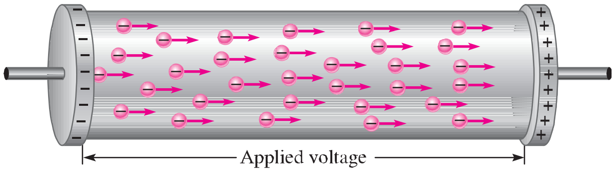

FIGURE 2–17 **Electrons flow from negative to positive when a voltage is applied across a conductive or semiconductive material.**

The movement of the free electrons from the negative end of a material to the positive end is the electrical current, symbolized by *I*.

Electrical current is the rate of flow of charge.

Current in a conductive material is measured by the number of electrons (amount of charge) that flow past a point in a unit of time.

$$I = \frac{Q}{t} \qquad (2\text{–}3)$$

where *I* is current in amperes (A), *Q* is the charge of the electrons in coulombs (C), and *t* is time in seconds (s). As a simple analogy, you can think of current as corresponding to

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

water flowing through a pipe in a water system when pressure (corresponding to voltage) is applied by a pump (corresponding to a voltage source). *Voltage causes current.*

Ampere: The Unit of Current

Current is measured in a unit called the *ampere* or *amp* for short, symbolized by A.

> **One ampere (1 A) is the amount of current that exists when a number of electrons having a total charge of one coulomb (1 C) move through a given cross-sectional area in one second (1 s).**

See Figure 2–18. Remember, one coulomb is the charge carried by 6.25×10^{18} electrons.

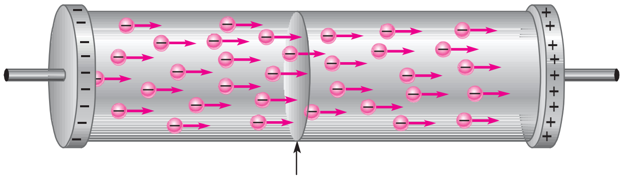

When a number of electrons having a total charge of 1 C pass through a cross-sectional area in 1 s, there is 1 A of current.

FIGURE 2–18 **Illustration of 1 A of current (1 C/s) in a material.**

EXAMPLE 2–3

Ten coulombs of charge flow past a given point in a wire in 2 s. What is the current in amperes?

SOLUTION

$$I = \frac{Q}{t} = \frac{10\text{ C}}{2\text{ s}} = \mathbf{5\ A}$$

RELATED PROBLEM

If there are 8 A of direct current through the filament of a light bulb, how many coulombs have moved through the filament in 1.5 s?

The Current Source

THE IDEAL CURRENT SOURCE As you know, an ideal voltage source can provide a constant voltage for any load. An ideal **current source** can provide a constant current in any load. Just as in the case of a voltage source, the ideal current source does not exist but can be approximated in practice. We will assume ideal unless otherwise specified.

The symbol for a current source is shown in Figure 2–19(a). The graph for an ideal current source is a horizontal line as illustrated in Figure 2–19(b). This graph is called the *IV* characteristic curve. Notice that the current is constant for any voltage across the current source. In a nonideal current source, the current drops off as shown by the dashed line in Figure 2–19(b).

ACTUAL CURRENT SOURCES Power supplies are normally thought of as voltage sources because they are the most common type of source in the laboratory. However, current sources are another type of power supply. Current sources may be "stand-alone"

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

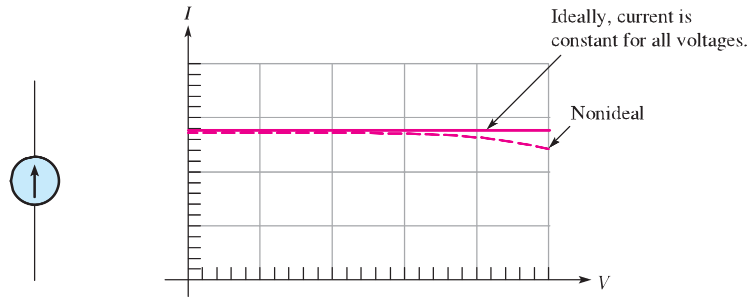

(a) Symbol (b) *IV* characteristic

FIGURE 2–19 **The current source.**

instruments or may be combined with other instruments, such as a voltage source, DMM, or function generator. Examples of combination instruments are the source-measurement units shown in Figure 2–20. These units can be set up as voltage or current sources and include a built-in DMM, as well as other instruments. They are used primarily for testing transistors and other semiconductors.

SAFETY NOTE

Current sources change the output voltage in order to supply a constant current to the load. For example, a meter calibrator can have a different output voltage that depends on the meter under test. You should never touch the leads from a current source; the voltage can be high, and a shock will result, particularly if the load is a high resistance load or the load is disconnected when the current source is turned on.

KEITHLEY
2636 SYSTEM SourceMeter
-000.01pA -000.01pA
SrcA:+000.000mV SrcB:+000.000mV
2635 SYSTEM SourceMeter
-000.123pA

FIGURE 2–20 **Source-measurement units that have current and voltage sources.** (Courtesy of Keithley Instruments)

In most transistor circuits, the transistor acts as a current source because part of the *IV* characteristic curve is a horizontal line as shown by the transistor characteristic in Figure 2–21. The flat part of the graph indicates where the transistor current is constant over a range of voltages. The constant-current region is used to form a constant-current source.

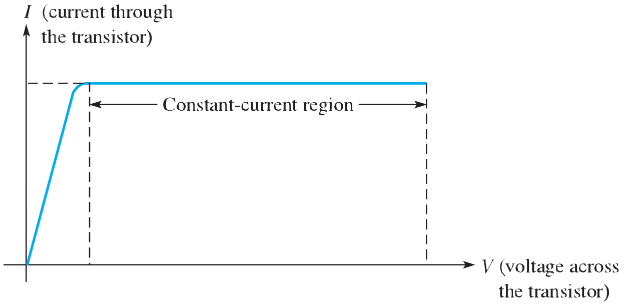

FIGURE 2–21 **Characteristic curve of a transistor showing the constant-current region.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SYSTEM EXAMPLE 2–1

A CURRENT SOURCE

Current sources are useful in a variety of automotive lighting requirements (taillights, turn signals, brakes, daytime running lights, interior lights) and in requirements like studio lighting where constant illumination is required. High brightness light-emitting diodes (LEDs) are popular in these applications because they are energy efficient, reliable, small, and can turn on rapidly. Many integrated circuits are specifically designed to provide the high currents needed for constant illumination. An example is the BP5843A constant current LED driver for illumination. The BP5843A can accept either a dc source (from 113 V to 170 V) or an ac source (from 80 V to 120 V) and can supply a range of currents from 250 mA to 350 mA at a constant level as set by the user.

A basic constant-current illumination lighting system is shown in Figure 2–22. The input ac is filtered, transformed, and rectified (converted to pulsing dc). The components labeled *C* are capacitors, which are for filtering and noise removal. (Capacitors are discussed in Chapter 9.) The resistor, labeled *R*, is used to set the amount of current through the high brightness LEDs. (Resistors are covered in Section 2–5.)

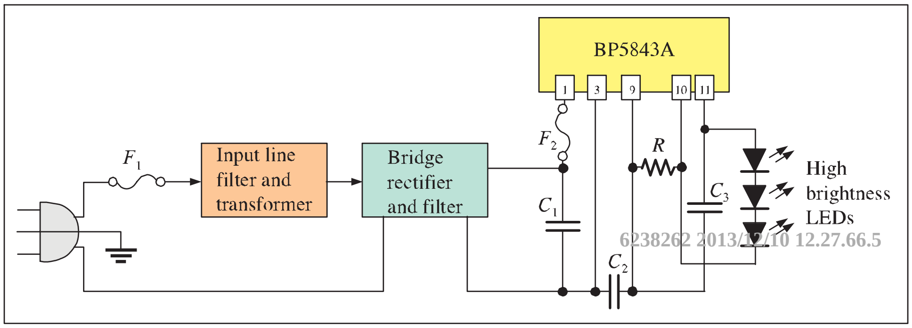

6238262 2013/12/10 12.27.66.5

FIGURE 2–22 A basic constant-current illumination lighting system.

SECTION 2–4 CHECKUP

1. Define *current* and state its unit.
2. How many electrons make up one coulomb of charge?
3. What is the current in amperes when 20 C flow past a point in a wire in 4 s?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

2–5 RESISTANCE

When there is current through a material, the free electrons move through the material and occasionally collide with atoms. These collisions cause the electrons to lose some of their energy, and thus their movement is restricted. The more collisions, the more the flow of electrons is restricted. This restriction varies and is determined by the type of material. The property of a material that restricts the flow of electrons is called *resistance*, designated with an *R*.

After completing this section, you should be able to

- **Define *resistance* and discuss its characteristics**
  - **Name and define the unit of resistance**
  - **Describe the basic types of resistors**
  - **Determine resistance value by color code or labeling**

Resistance is the opposition to current.

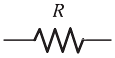

FIGURE 2–23 **Resistance/resistor symbol.**

The schematic symbol for resistance is shown in Figure 2–23.

When there is current through any material that has resistance, heat is produced by the collisions of free electrons and atoms. Therefore, wire, which typically has a very small resistance, can become warm or even hot when there is sufficient current through it.

As a simple analogy, you can think of a resistor as corresponding to a partially open valve in a closed water system that restricts the amount of water flowing through a pipe. If the valve is opened more (corresponding to less resistance), the water flow (corresponding to current) increases. If the valve is closed a little (corresponding to more resistance), the water flow (corresponding to current) decreases.

Ohm: The Unit of Resistance

Resistance, R, is expressed in the unit of ohms, which is symbolized by the Greek letter omega (Ω).

One ohm (1 Ω) of resistance exists when there is one ampere (1 A) of current in a material with one volt (1 V) applied across the material.

CONDUCTANCE The reciprocal of resistance is **conductance**, symbolized by G. It is a measure of the ease with which current is established. The formula is

$$G = \frac{1}{R} \tag{2–4}$$

The unit of conductance is the **siemens**, symbolized by S. For example, the conductance of a 22 kΩ resistor is

$$G = \frac{1}{22\text{ k}\Omega} = 45.5\ \mu\text{S}$$

Occasionally, the obsolete unit of *mho* is still used for conductance.

Resistors

Components that are specifically designed to have a certain amount of resistance are called **resistors**. The principal applications of resistors are to limit current, divide voltage, and, in certain cases, generate heat. Although different types of resistors come in many shapes and sizes, they can all be placed in one of two main categories: fixed or variable.

FIXED RESISTORS Fixed resistors are available with a large selection of resistance values that are set during manufacturing and cannot be changed easily. Fixed resistors are constructed using various methods and materials. Figure 2–24 shows several common types.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

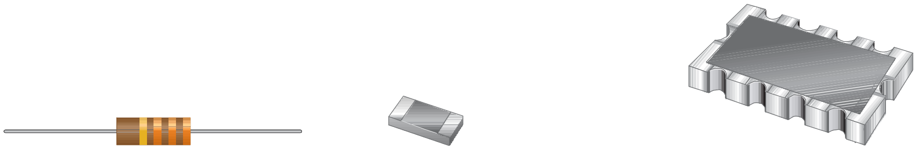

(a) Carbon-composition (b) Metal film chip resistor (c) Chip resistor array

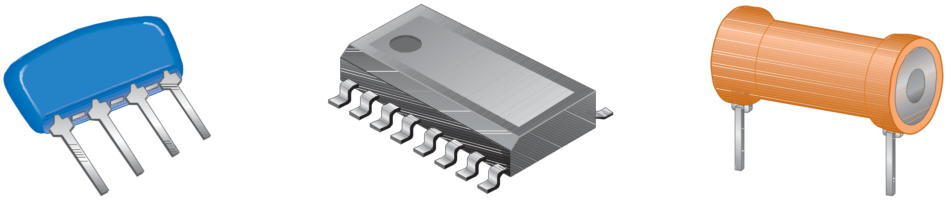

(d) Resistor network (simm) (e) Resistor network (surface mount) (f) Radial-lead for PC board insertion

FIGURE 2–24 **Typical fixed resistors.**

One common fixed resistor is the carbon-composition type, which is made with a mixture of finely ground carbon, insulating filler, and a resin binder. The ratio of carbon to insulating filler sets the resistance value. The mixture is formed into rods which are cut into short lengths, and lead connections are made. The entire resistor is then encapsulated in an insulated coating for protection. Figure 2–25(a) shows the construction of a typical carbon-composition resistor.

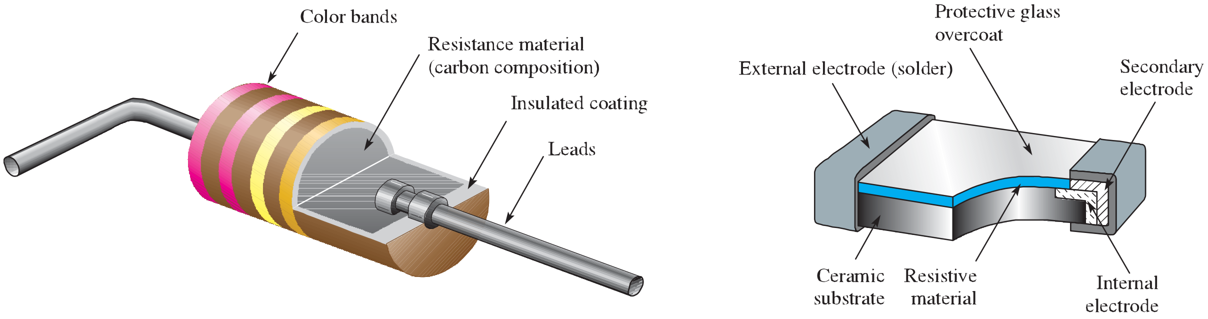

(a) Cutaway view of a carbon-composition resistor (b) Cutaway view of a tiny chip resistor

FIGURE 2–25 **Two types of fixed resistors (not to scale).**

The chip resistor is another type of fixed resistor and is in the category of SMT (surface-mount technology) components. It has the advantage of a very small size for compact assemblies. Small-value chip resistors ($<1\ \Omega$) can be made with very close tolerances ($\pm 0.5\%$) and have application as current-sensing resistors. Figure 2–25(b) shows the construction of a chip resistor.

Other types of fixed resistors include carbon film, metal film, metal-oxide film, and wirewound. In film resistors, a resistive material is deposited evenly onto a high-grade ceramic rod. The resistive film may be carbon (carbon film) or nickel chromium (metal film). In these types of resistors, the desired resistance value is obtained by removing part of the resistive material in a helical pattern along the rod using a spiraling technique, as shown in Figure 2–26(a). Very close **tolerance** can be achieved with this method. Film resistors are also available in the form of resistor networks, as shown in Figure 2–26(b).

Wirewound resistors are constructed with resistive wire wound around an insulating rod and then sealed. Normally, wirewound resistors are used because of their relatively high power ratings. Since they are constructed with a coil of wire, wirewound resistors have significant inductance and are not used at higher frequencies. Some typical wirewound resistors are shown in Figure 2–27.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

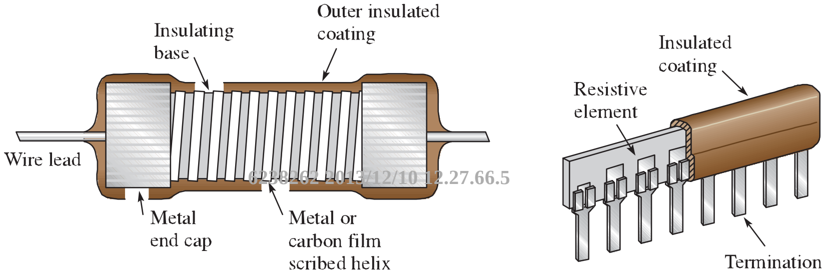
Outer insulated coating
Insulating base
Wire lead
Metal end cap
Metal or carbon film scribed helix
Insulated coating
Resistive element
Termination

(a) Film resistor showing spiraling technique

(b) Resistor network

FIGURE 2–26 **Construction views of typical film resistors.**

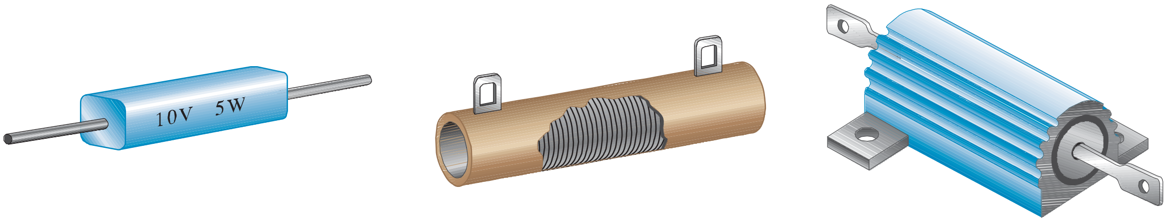

FIGURE 2–27 **Typical wirewound power resistors.**

RESISTOR COLOR CODES Some types of fixed resistors with value tolerances of 5% or 10% are color coded with four bands to indicate the resistance value and the tolerance. This color-code band system is shown in Figure 2–28, and the **color code** is listed in Table 2–1. The bands are always located closer to one end.

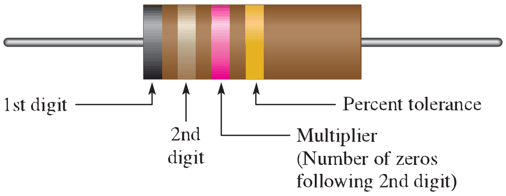
1st digit
2nd digit
Multiplier (Number of zeros following 2nd digit)
Percent tolerance

FIGURE 2–28 **Color-code bands on a 4-band resistor.**

The 4-band color code is read as follows:

1. Start with the band closest to one end of the resistor. The first band is the first digit of the resistance value. If it is not clear which is the banded end, start from the end that does not begin with a gold or silver band.
2. The second band is the second digit of the resistance value.
3. The third band is the number of zeros following the second digit, or the *multiplier*.
4. The fourth band indicates the percent tolerance and is usually gold or silver. If there is no fourth band, the tolerance is ±20%.

For example, a 5% tolerance means that the *actual* resistance value is within ±5% of the color-coded value. Thus, a 100 Ω resistor with a tolerance of ±5% can have an acceptable range of values from a minimum of 95 Ω to a maximum of 105 Ω.

As indicated in the table, for resistance values less than 10 Ω, the third band is either gold or silver. Gold in the third band represents a multiplier of 0.1, and silver represents 0.01. For example, a color code of red, violet, gold, and silver represents 2.7 Ω with a tolerance of ±10%. A table of standard resistance values is in Appendix A.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

TABLE 2–1 • 4-band resistor color code.

	Digit	Color	
Resistance value, first three bands: First band—1st digit Second band—2nd digit *Third band—multiplier (number of zeros following the 2nd digit)	0	⬛	Black
	1	🟫	Brown
	2	🟥	Red
	3	🟧	Orange
	4	🟨	Yellow
	5	🟩	Green
	6	🟦	Blue
	7	🟪	Violet
	8	◼	Gray
	9	⬜	White
Fourth band—tolerance	±5%	🟨	Gold
	±10%	▨	Silver

* For resistance values less than 10 Ω, the third band is either gold or silver. Gold is for a multiplier of 0.1 and silver is for a multiplier of 0.01.

EXAMPLE 2–4

Find the resistance values in ohms and the percent tolerance for each of the color-coded resistors shown in Figure 2–29.

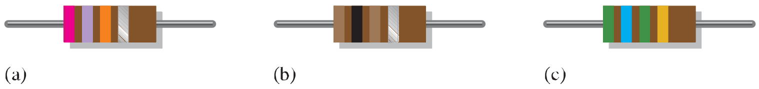

FIGURE 2–29

SOLUTION

(a) First band is red = 2, second band is violet = 7, third band is orange = 3 zeros, fourth band is silver = ±10% tolerance.

$$R = \mathbf{27{,}000\ \Omega \pm 10\%}$$

(b) First band is brown = 1, second band is black = 0, third band is brown = 1 zero, fourth band is silver = ±10% tolerance.

$$R = \mathbf{100\ \Omega \pm 10\%}$$

(c) First band is green = 5, second band is blue = 6, third band is green = 5 zeros, fourth band is gold = ±5% tolerance.

$$R = \mathbf{5{,}600{,}000\ \Omega \pm 5\%}$$

RELATED PROBLEM

A certain resistor has a yellow first band, a violet second band, a red third band, and a gold fourth band. Determine the value in ohms and its percent tolerance.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

FIVE-BAND COLOR CODE Certain precision resistors with tolerances of 2%, 1%, or less are generally color coded with five bands, as shown in Figure 2–30. Begin at the band closest to one end. The first band is the first digit of the resistance value, the second band is the second digit, the third band is the third digit, the fourth band is the multiplier (number of zeros after the third digit), and the fifth band indicates the tolerance. Table 2–2 shows the 5-band color code.

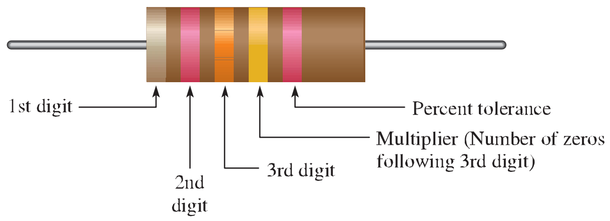

FIGURE 2–30 Color-code bands on a 5-band resistor.

TABLE 2–2 • 5-band resistor color code.

	DIGIT	COLOR
Resistance value, first four bands:	0	Black
	1	Brown
	2	Red
First band—1st digit	3	Orange
Second band—2nd digit	4	Yellow
Third band—3rd digit	5	Green
Fourth band—multiplier	6	Blue
(number of zeros following 3rd digit)	7	Violet
	8	Gray
	9	White
Fourth band—multiplier	0.1	Gold
	0.01	Silver
	±2%	Red
	±1%	Brown
Fifth band—tolerance	±0.5%	Green
	±0.25%	Blue
	±0.1%	Violet

EXAMPLE 2–5

Find the resistance value in ohms and the percent tolerance for each of the color-coded resistors shown in Figure 2–31.

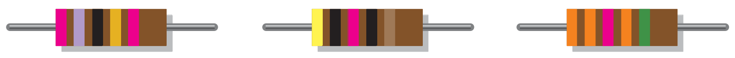

(a) (b) (c)

FIGURE 2–31

SOLUTION

(a) First band is red = 2, second band is violet = 7, third band is black = 0, fourth band is gold = ×0.1, fifth band is red = ±2% tolerance.

$$R = 270 \times 0.1 = \mathbf{27\ \Omega \pm 2\%}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

(b) First band is yellow $= 4$, second band is black $= 0$, third band is red $= 2$, fourth band is black $= 0$, fifth band is brown $= \pm 1\%$ tolerance.

$$R = \mathbf{402\ \Omega \pm 1\%}$$

(c) First band is orange $= 3$, second band is orange $= 3$, third band is red $= 2$, fourth band is orange $= 3$, fifth band is green $= \pm 0.5\%$ tolerance.

$$R = \mathbf{332{,}000\ \Omega \pm 0.5\%}$$

RELATED PROBLEM

A certain resistor has a yellow first band, a violet second band, a green third band, a gold fourth band, and a red fifth band. Determine its value in ohms and its percent tolerance.

RESISTOR LABEL CODES Not all types of resistors are color coded. Many, including surface-mount resistors, use typographical marking to indicate the resistance value and tolerance. These label codes consist of either all numbers (numeric) or a combination of numbers and letters (alphanumeric). In some cases when the body of the resistor is large enough, the entire resistance value and tolerance are stamped on it in standard form. For example, a 33,000 Ω resistor may be labeled as 33 kΩ.

Numeric labeling uses three digits to indicate the resistance value, as shown in Figure 2–32 using a specific example. The first two digits give the first two digits of the resistance value, and the third digit gives the multiplier or number of zeros that follow the first two digits. This code is limited to values of 10 Ω or greater.

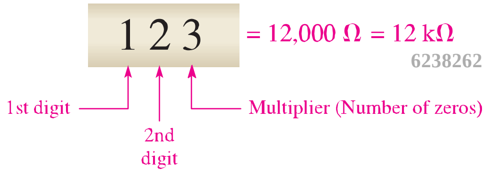

6238262 2013/12/10 12.27.66.5

FIGURE 2–32 **Example of three-digit labeling for a resistor.**

Another common type of marking is a three- or four-character label that uses both digits and letters. An alphanumeric label typically consists of only three digits or two or three digits and one of the letters R, K, or M. The letter is used to indicate the multiplier, and the position of the letter indicates the decimal point placement. The letter R indicates a multiplier of 1 (no zeros after the digits), the K indicates a multiplier of 1000 (three zeros after the digits), and the M indicates a multiplier of 1,000,000 (six zeros after the digits). In this format, values from 100 to 999 consist of three digits and no letter to represent the three digits in the resistance value. Figure 2–33 shows three examples of this type of resistor label.

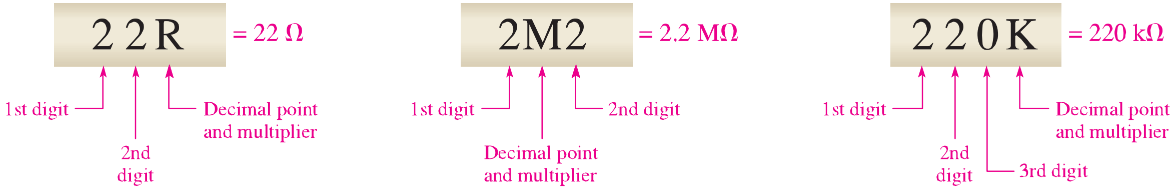

FIGURE 2–33 **Examples of the alphanumeric resistor label.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 2–6

Interpret the following alphanumeric resistor labels:

(a) 470 **(b)** 5R6 **(c)** 68K **(d)** 10M **(e)** 3M3

SOLUTION

(a) 470 = **470 Ω** **(b)** 5R6 = **5.6 Ω** **(c)** 68K = **68 kΩ**
(d) 10M = **10 MΩ** **(e)** 3M3 = **3.3 MΩ**

RELATED PROBLEM

What is the resistance indicated by 1K2?

One system of labels for resistance tolerance values uses the letters F, G, and J:

F = ±1% G = ±2% J = ±5%

For example, 620F indicates a 620 Ω resistor with a tolerance of ±1%, 4R6G is a 4.6 Ω ±2% resistor, and 56KJ is a 56 kΩ ±5% resistor.

VARIABLE RESISTORS Variable resistors are designed so that their resistance values can be changed easily. Two basic uses for variable resistors are to divide voltage and to control current. The variable resistor used to divide voltage is called a **potentiometer**. The variable resistor used to control current is called a **rheostat**. Schematic symbols for these types are shown in Figure 2–34. The potentiometer is a three-terminal device, as indicated in part (a). Terminals 1 and 2 have a fixed resistance between them, which is the total resistance. Terminal 3 is connected to a moving contact (**wiper**). You can vary the resistance between 3 and 1 or between 3 and 2 by moving the contact.

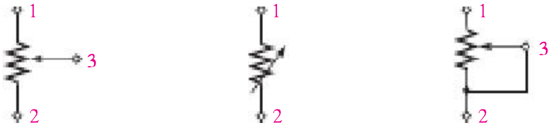

(a) Potentiometer (b) Rheostat (c) Potentiometer connected as a rheostat

FIGURE 2–34 **Potentiometer and rheostat symbols.**

Figure 2–34(b) shows the rheostat as a two-terminal variable resistor. Part (c) shows how you can use a potentiometer as a rheostat by connecting terminal 3 to either terminal 1 or terminal 2. Parts (b) and (c) are equivalent symbols. Some typical potentiometers are pictured in Figure 2–35. The construction view is shown under the normal view.

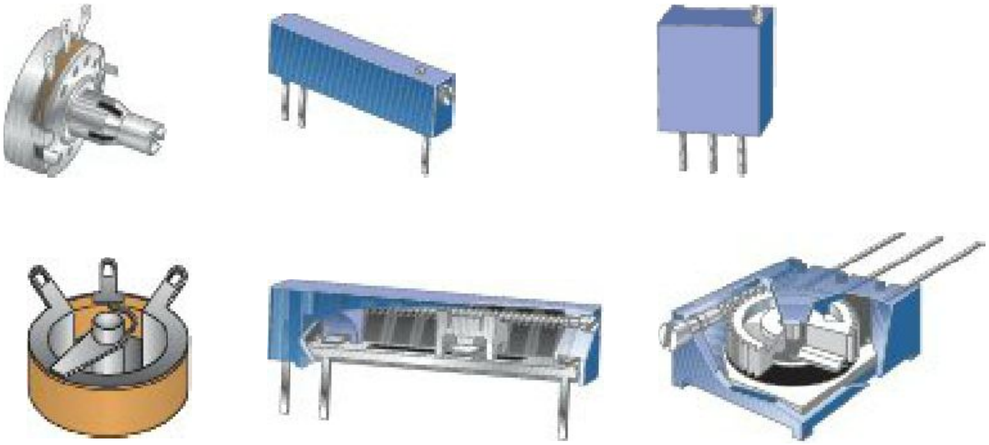

FIGURE 2–35 **Typical potentiometers and construction views.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Potentiometers and rheostats can be classified as linear or tapered, as shown in Figure 2–36, where a potentiometer with a total resistance of 100 Ω is used as an example. As shown in part (a), in a linear potentiometer, the resistance between either terminal and the moving contact varies linearly with the position of the moving contact. For example, one-half of a turn results in one-half the total resistance. Three-quarters of a turn results in three-quarters of the total resistance between the moving contact and one terminal, or one-quarter of the total resistance between the other terminal and the moving contact.

In a **tapered** potentiometer, the resistance varies nonlinearly with the position of the moving contact, so that one-half of a turn does not necessarily result in one-half the total resistance. This concept is illustrated in Figure 2–36(b), where the nonlinear values are arbitrary.

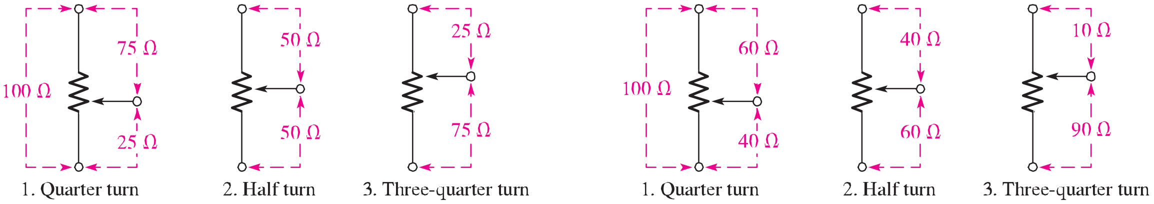

1. Quarter turn 2. Half turn 3. Three-quarter turn

(a) Linear

1. Quarter turn 2. Half turn 3. Three-quarter turn

(b) Tapered (nonlinear)

FIGURE 2–36 **Examples of (a) linear and (b) tapered potentiometers.**

The potentiometer is used as a voltage-control device. When a fixed voltage is applied across the end terminals, a variable voltage is obtained at the wiper contact with respect to either end terminal. The rheostat is used as a current-control device; the current can be changed by changing the wiper position.

VARIABLE RESISTANCE SENSORS Many sensors operate on the concept of a variable resistance, in which a physical quantity alters the electrical resistance. Depending on the sensor and the measurement requirements, the change in resistance may be determined directly or indirectly using the resistance change to alter a voltage or current.

Examples of resistance sensors include **thermistors** that change resistance as a function of temperature, **photoconductive cells** that change resistance as a function of light, and **strain gages** that change resistance when a force is applied to them. Thermistors are commonly used in thermostats. Photocells have many applications; for example, they are used to turn on street lights at dusk and turn them off at dawn. Strain gages are widely used in scales and applications where mechanical motion needs to be sensed. The measuring instruments for strain gages need to be very sensitive because the change in resistance is very small. Figure 2–37 shows symbols for these types of resistance sensors.

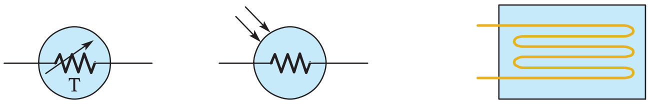

(a) Thermistor (b) Photoconductive cell (c) Strain gage

FIGURE 2–37 **Symbols for resistance devices with sensitivities to temperature, light, and force.**

Truck Scales

Many large scales, such as those that are used for weighing trucks, use strain gages as the basic sensor to detect the force. The strain gage is part of a complete electromechanical assembly consisting of four gages. Typically, the strain gages are composed of a back-and forth pattern of thin wires that are bonded to a steel beam or plate. The steel is bent by the weight, stretching or compressing the wire in the strain gage and changing its resistance. The change is detected and converted to a number that represents the weight.

Courtesy of Cardinal Scale Manufacturing Co. www.CardinalScale.com

SYSTEM NOTE

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 2–5 CHECKUP

1. Define *resistance* and name its unit.

2. What are the two main categories of resistors? Briefly explain the difference between them.

3. In the 4-band resistor color code, what does each band represent?

4. Determine the resistance and percent tolerance for each of the following color codes:

- **(a)** yellow, violet, red, gold
- **(b)** blue, red, orange, silver
- **(c)** brown, gray, black, gold
- **(d)** red, red, blue, red, green

5. What resistance value is indicated by each alphanumeric label:

- **(a)** 33R
- **(b)** 5K6
- **(c)** 900
- **(d)** 6M8

6238262 2013/12/10 12.27.66.5

6. What is the basic difference between a rheostat and a potentiometer?

7. Name three resistance sensors and the physical quantity that affects their resistance.

2–6 THE ELECTRIC CIRCUIT

A basic electric circuit is an arrangement of physical components that use voltage, current, and resistance to perform some useful function.

After completing this section, you should be able to

- **Describe a basic electric circuit**
  - **Relate a schematic to a physical circuit**
  - **Define *open circuit* and *closed circuit***
  - **Describe various types of protective devices**
  - **Describe various types of switches**
  - **Explain how wire sizes are related to gauge numbers**
  - **Define *ground* or *common***

Basically, an electric circuit consists of a voltage source, a load, and a path for current between the source and the load. The **load** is a device on which work is done by the current through it. Figure 2–38 shows an example of a simple electric circuit: a battery connected to a lamp with two conductors (wires). The battery is the voltage source, the lamp is the load on the battery because it draws current from the battery, and the two wires provide the current path from the negative terminal of the battery to the lamp and back to the positive terminal of the battery, as indicated by the red arrows. There is current through the filament of the lamp (which has a resistance), causing it to become hot enough to emit visible light. Current through the battery is produced by chemical action.

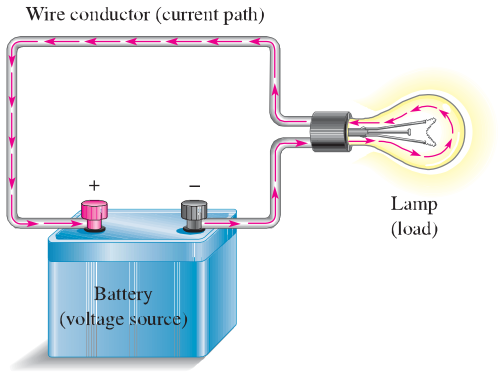

FIGURE 2–38 **A simple electric circuit.**

In many practical cases, one terminal of the battery is connected to a ground (common) point. For example, in automobiles, the negative battery terminal is generally connected to the metal chassis of the car. The chassis is the ground for the automobile electrical system and provides a current path for the circuit. (The concept of *ground* is covered later in this chapter.)

As you know, block diagrams are a basic way to show the signal path and functional blocks in a system. When details of a circuit are required for troubleshooting or analysis, in large systems, for example,

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

it is useful to have a schematic of the circuit. A **schematic** is a diagram that uses standard symbols and designators to represent components and lines to represent conductors; it shows the circuit in a logical organized manner so that circuit operation can be determined. The schematic may be quite different from the physical layout and is usually drawn to show signal flow from left to right. In general, the symbols used in schematics have been standardized by various standards organizations. For the U.S., the standard is IEEE STD 315-1975, available from the American National Standards Institute (ANSI). In this textbook, symbols are introduced as they are needed. Figure 2–39 is a basic schematic for the circuit in Figure 2–38.

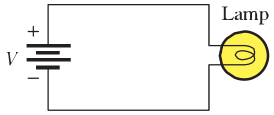

FIGURE 2–39 **A schematic for the circuit in Figure 2–38.**

SAFETY NOTE

To avoid electrical shock, never touch a circuit while it is connected to a voltage source. If you need to handle a circuit, remove a component, or change a component, first make sure the voltage source is disconnected.

Circuit Current Control and Protection

The circuit in Figure 2–40(a) illustrates a **closed circuit**—that is, a circuit in which the current has a complete path. An **open circuit** is a circuit in which the current path is broken so that there is no current, as shown in part (b). An open circuit is considered to have infinite resistance (*infinite* means immeasurably large).

MECHANICAL SWITCHES **Switches** are commonly used for controlling the opening or closing of circuits. For example, a switch is used to turn a lamp on or off, as illustrated in Figure 2–40. Each circuit pictorial is shown with its associated schematic. The type of switch indicated is a *single-pole–single-throw* (SPST) toggle switch. The term *pole* refers to the movable arm in a switch, and the term *throw* indicates the number of contacts that are affected (either opened or closed) by a single switch action (a single movement of a pole).

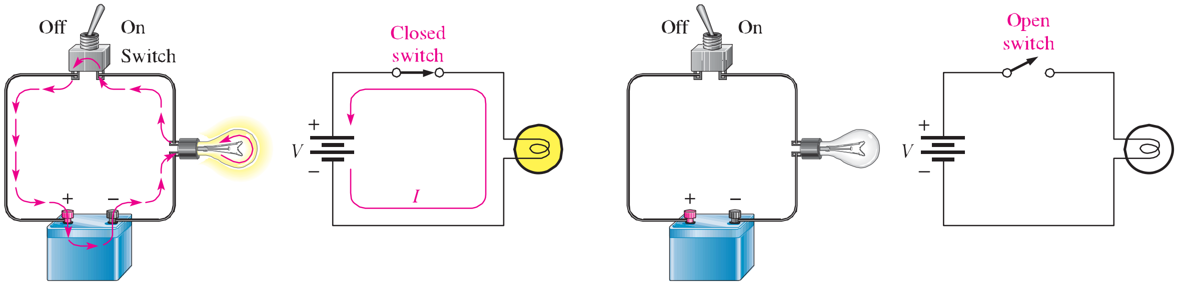

(a) There is current in a *closed* circuit because there is a complete current path (switch is ON or in the *closed* position). Current is always indicated by a red arrow in this text.

(b) There is no current in an *open* circuit because the path is broken (switch is OFF or in the *open* position).

FIGURE 2–40 **Illustration of closed and open circuits using an SPST switch for control.**

Figure 2–41 shows a somewhat more complicated circuit using a *single-pole–double-throw* (SPDT) type of switch to control the current to two different lamps. When one lamp is on, the other is off, and vice versa, as illustrated by the two schematics that represent each of the switch positions.

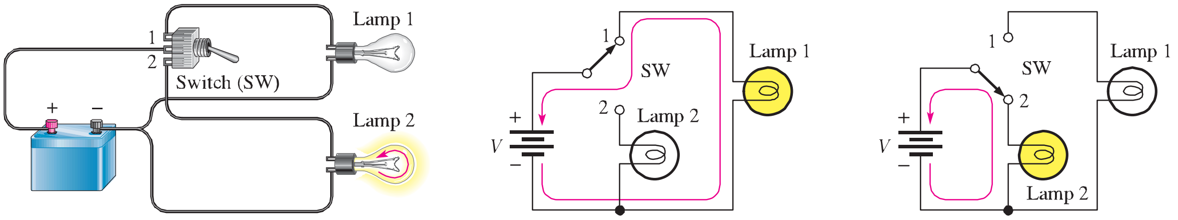

(a) Pictorial

(b) A schematic showing Lamp 1 on and Lamp 2 off

(c) A schematic showing Lamp 2 on and Lamp 1 off

FIGURE 2–41 **An example of an SPDT switch controlling two lamps.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

HANDS ON TIP

Small electronic parts can be damaged easily if you apply too much heat to them when soldering. The small switches shown in Figure 2–43 frequently are constructed with plastic that can melt and render the switch useless. Manufacturers usually supply a maximum time and temperature that can be applied to a part without damage. A small heat sink can be temporarily connected between the point where solder is applied and the sensitive region of the component.

In addition to the SPST and the SPDT switches (symbols are shown in Figure 2–42(a) and (b)), the following other types of switches are also important:

- **Double-pole–single-throw (DPST).** The DPST switch permits simultaneous opening or closing of two sets of contacts. The symbol is shown in Figure 2–42(c). The dashed line indicates that the contact arms are mechanically linked so that both move with a single switch action.
- **Double-pole–double-throw (DPDT).** The DPDT switch provides connection from one set of contacts to either of two other sets. The schematic symbol is shown in Figure 2–42(d).
- **Push button (PB).** In the normally open push-button switch (NOPB), shown in Figure 2–42(e), connection is made between two contacts when the button is depressed, and connection is broken when the button is released. In the normally closed push-button switch (NCPB), shown in Figure 2–42(f), connection between the two contacts is broken when the button is depressed.
- **Rotary.** In a rotary switch, a knob is turned to make a connection between one contact and any one of several others. A symbol for a simple six-position rotary switch is shown in Figure 2–42(g).

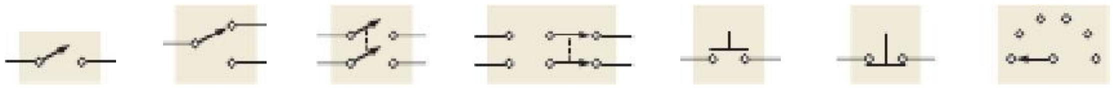

(a) SPST (b) SPDT (c) DPST (d) DPDT (e) NOPB (f) NCPB (g) Single-pole rotary (6-position)

FIGURE 2–42 **Switch symbols.**

Figure 2–43 shows several varieties of switches, and Figure 2–44 shows a construction view of a typical toggle switch. In addition to mechanical switches, a transistor can be used as the equivalent of a single-pole–single-throw switch in certain applications.

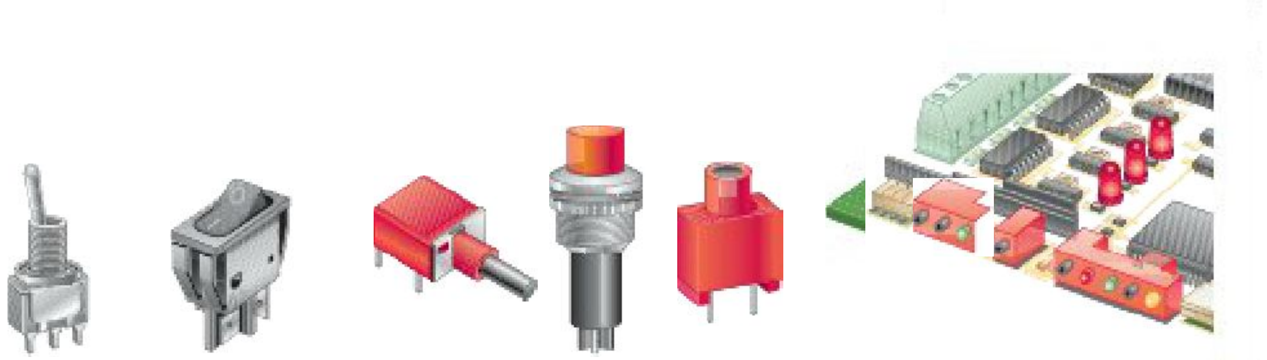

Toggle switch Push-button switches PC board mounted push-button switches

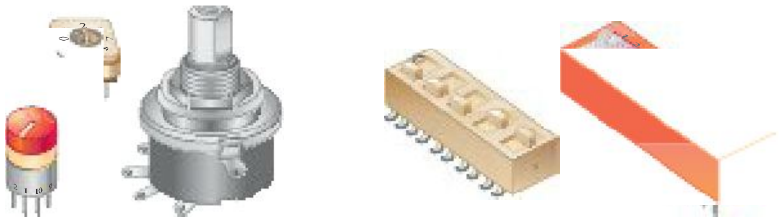

Rotary switches DIP switches for mounting on PC boards

FIGURE 2–43 **Typical mechanical switches.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

PROTECTIVE DEVICES **Fuses** and **circuit breakers** are placed in the current path and are used to deliberately create an open circuit when the current exceeds a specified number of amperes due to a malfunction or other abnormal condition in a circuit. For example, a fuse or circuit breaker with a 20 A rating will open a circuit when the current exceeds 20 A.

The basic difference between a fuse and a circuit breaker is that when a fuse is "blown," it must be replaced; but when a circuit breaker opens, it can be reset and reused repeatedly. Both of these devices protect against damage to a circuit due to excess current or prevent a hazardous condition created by the overheating of wires and other components when the current is too great. Because fuses cut off excess current more quickly than circuit breakers, fuses are used whenever delicate electronic equipment needs to be protected. Several typical fuses and circuit breakers, along with their schematic symbols, are shown in Figure 2–45.

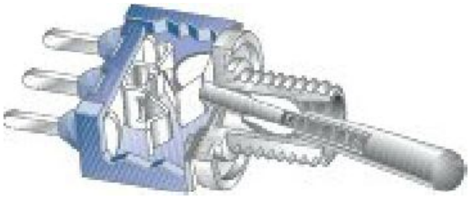

FIGURE 2–44 **Construction view of a typical toggle switch.**

6238262 2013/12/10 12.27.66.5

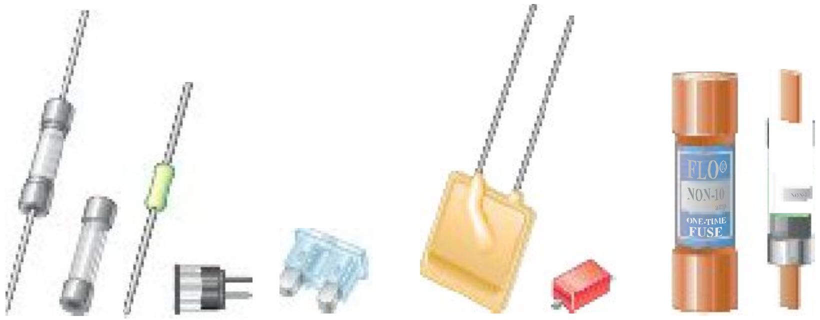

(a) Cartridge fuses

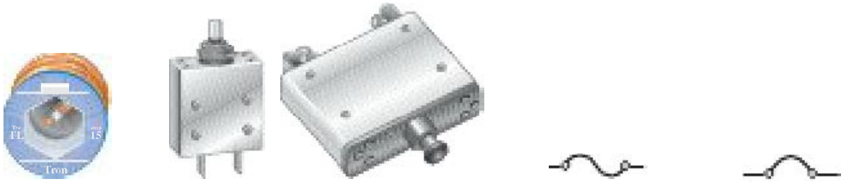

(b) Plug fuse

(c) Circuit breakers

(d) Fuse symbol

(e) Circuit breaker symbol

FIGURE 2–45 **Typical fuses and circuit breakers and their symbols.**

Two basic categories of fuses in terms of their physical configuration are cartridge type and plug type (screw in). Cartridge type fuses have various shaped housings with leads or other types of contacts as shown in Figure 2–45(a). A typical plug type fuse is shown in part (b). Fuse operation is based on the melting temperature of a wire or other metal element. As current increases, the fuse element heats up and when the rated current is exceeded, the element reaches its melting temperature and opens, thus removing power from the circuit.

Two common types of fuses are the fast-acting and the time-delay (slow blow). Fast-acting fuses are type F and time-delay fuses are type T. In normal operation, fuses are often subjected to intermittent current surges that may exceed the rated current, such as when power to a circuit is turned on. Over time, this reduces the fuse's ability to withstand short surges or even current at the rated value. A slow blow fuse can tolerate greater and longer duration surges of current than the typical fast-acting fuse. A fuse symbol is shown in Figure 2–45(d).

Typical circuit breakers are shown in Figure 2–45(c) and the symbol is shown in part (e). Generally, a circuit breaker detects excess current either by the heating effect of the current or by the magnetic field it creates. In a circuit breaker based on the heating effect, a bimetallic spring opens the contacts when the rated current is exceeded. Once opened, the contact is held open by mechanical means until manually reset. In a circuit breaker based on a magnetic field, the contacts are opened by a sufficient magnetic force created by excess current and must be mechanically reset.

SAFETY NOTE

Always use fully insulated fuse pullers to remove and replace fuses in an electrical box. Even if the disconnect switch is in the off position, line voltage is still present in the box. Never use metal tools to remove and replace fuses.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SYSTEM EXAMPLE 2–2

A QUIZ BOARD

Figure 2–46 shows a quiz board for a science fair project; it is an example of a small system that uses components introduced in this chapter. The operation of the circuit is for the user to rotate the switch into one of four positions to select a battery type (shown with a light). After selecting the battery type, the user presses a pushbutton corresponding to the correct cell voltage for that battery. The "correct" light will illuminate only if the selected pushbutton is correct. The quiz board is constructed with the sequence of correct answers as B, D, A, C for batteries 1, 2, 3, and 4. A rheostat is added to control the brightness of all lights and the circuit should be fused.

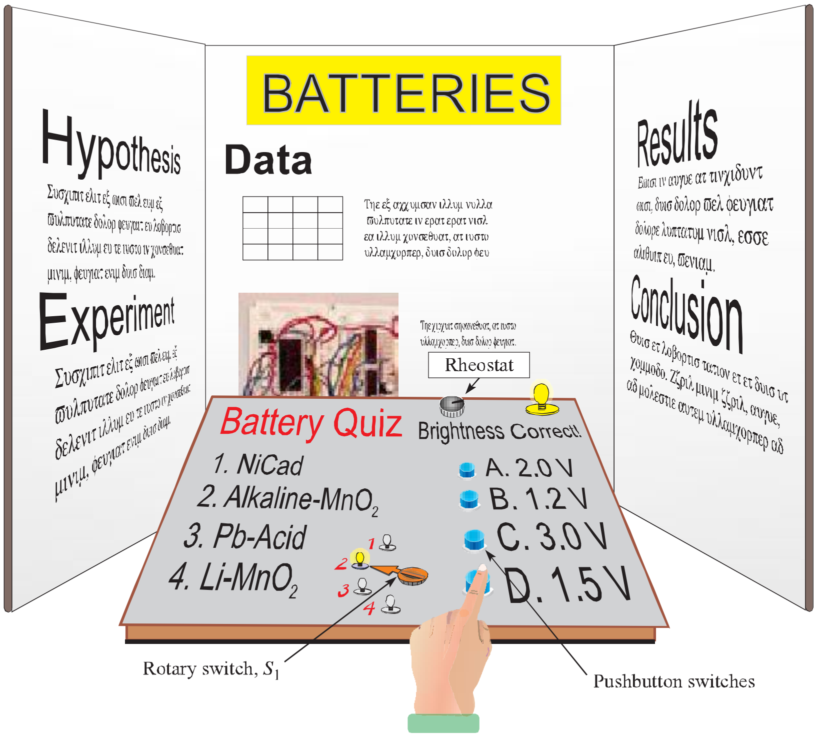
BATTERIES
Hypothesis
Data
Experiment
Results
Conclusion
Rheostat
Battery Quiz
Brightness Correct!
1. NiCad
2. Alkaline-MnO$_2$
3. Pb-Acid
4. Li-MnO$_2$
A. 2.0 V
B. 1.2 V
C. 3.0 V
D. 1.5 V
1
2
3
4
Rotary switch, S_1
Pushbutton switches

FIGURE 2–46 A quiz board.

The schematic for the circuit is shown in Figure 2–47. Notice that each position of the rotary switch goes to a light and that only one pushbutton switch corresponds to the correct answer.

FIGURE 2–47 Quiz board schematic.

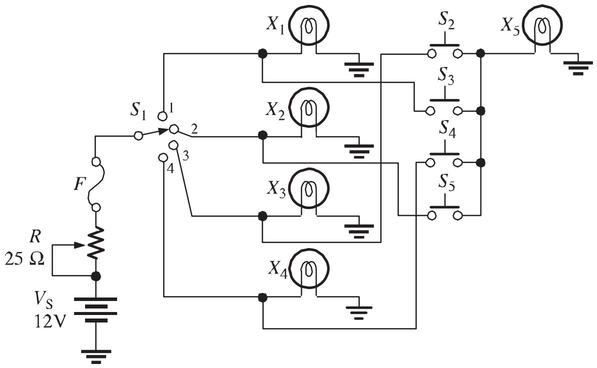
X_1
X_2
X_3
X_4
X_5
S_1
1
2
3
4
S_2
S_3
S_4
S_5
F
R
25 Ω
V_S
12 V

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Wires

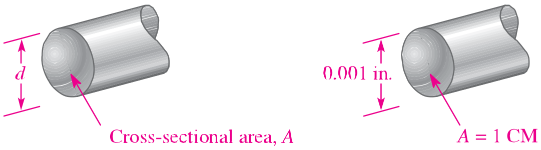

FIGURE 2–48 **Cross-sectional area of a wire.**

Wires are the most common form of conductive material used in electrical applications. They vary in diameter size and are arranged according to standard gauge numbers, called **AWG** (American Wire Gauge) sizes. As the gauge number increases, the wire diameter decreases. The size of a wire is also specified in terms of its cross-sectional area, as illustrated in Figure 2–48. The unit of cross-sectional area is the **circular mil,** abbreviated CM. One circular mil is the area of a wire with a diameter of 0.001 inch (0.001 in., or 1 mil). You can find the cross-sectional area in circular mils by expressing the diameter in thousandths of an inch (mils) and squaring it, as follows:

$$A = d^2 \qquad (2\text{–}5)$$

where A is the cross-sectional area in circular mils and d is the diameter in mils. Table 2–3 lists the AWG sizes with their corresponding cross-sectional area and resistance in ohms per 1000 ft at 20°C.

TABLE 2–3 • AWG (American Wire Gauge) sizes and resistances for solid round copper.

AWG #	AREA (CM)	RESISTANCE (Ω/1000 FT AT 20°C)	AWG #	AREA (CM)	RESISTANCE (Ω/1000 FT AT 20°C)
0000	211,600	0.0490	19	1,288.1	8.051
000	167,810	0.0618	20	1,021.5	10.15
00	133,080	0.0780	21	810.10	12.80
0	105,530	0.0983	22	642.40	16.14
1	83,694	0.1240	23	509.45	20.36
2	66,373	0.1563	24	404.01	25.67
3	52,634	0.1970	25	320.40	32.37
4	41,742	0.2485	26	254.10	40.81
5	33,102	0.3133	27	201.50	51.47
6	26,250	0.3951	28	159.79	64.90
7	20,816	0.4982	29	126.72	81.83
8	16,509	0.6282	30	100.50	103.2
9	13,094	0.7921	31	79.70	130.1
10	10,381	0.9989	32	63.21	164.1
11	8,234.0	1.260	33	50.13	206.9
12	6,529.0	1.588	34	39.75	260.9
13	5,178.4	2.003	35	31.52	329.0
14	4,106.8	2.525	36	25.00	414.8
15	3,256.7	3.184	37	19.83	523.1
16	2,582.9	4.016	38	15.72	659.6
17	2,048.2	5.064	39	12.47	831.8
18	1,624.3	6.385	40	9.89	1049.0

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 2–7

What is the cross-sectional area of a wire with a diameter of 0.005 inch?

SOLUTION

$$d = 0.005\text{ in.} = 5\text{ mils}$$
$$A = d^2 = (5\text{ mils})^2 = \mathbf{25\ CM}$$

RELATED PROBLEM

What is the cross-sectional area of a 0.0201 in. diameter wire? What is the AWG # for this wire from Table 2–3?

WIRE RESISTANCE Although copper wire conducts current extremely well, it still has some resistance, as do all conductors. The resistance of a wire depends on three physical characteristics: (a) type of material, (b) length of wire, and (c) cross-sectional area. In addition, temperature can also affect the resistance.

Each type of conductive material has a characteristic called its *resistivity,* which is represented by the Greek letter rho (ρ). For each material, ρ is a constant value at a given temperature. The formula for the resistance of a wire of length l and cross-sectional area A is

$$\boldsymbol{R = \frac{\rho l}{A}} \qquad (2\text{–}6)$$

Connectors

Connectors route wiring to and from system assemblies. The size and type of connector needed depends on the number and type of signals, power requirements, environment, and physical requirements. Connectors should be mechanically strong, reliable, and provide a low-resistance contact. A loose or corroded connection will have higher than normal resistance and can lead to internal arcing and breakdown.

6238262 2013/12/10 12.27.66.5

In most systems, connectors are critical for reliability. NASA has experienced delays in missions due to connector problems and has developed certain reliability tests for connectors for space missions. For this reason, life support systems, space, and military systems all have unique requirements and specifications for system connectors.

SYSTEM NOTE

This formula shows that resistance increases with an increase in resistivity and length and decreases with an increase in cross-sectional area. For resistance to be calculated in ohms, the length must be in feet (ft), the cross-sectional area in circular mils (CM), and the resistivity in CM-Ω/ft.

EXAMPLE 2–8

Find the resistance of a 100 ft length of copper wire with a cross-sectional area of 810.1 CM. The resistivity of copper is 10.37 CM-Ω/ft at 20°C.

SOLUTION

$$R = \frac{\rho l}{A} = \frac{(10.37\text{ CM-}\Omega/\text{ft})\,(100\text{ ft})}{810.1\text{ CM}} = \mathbf{1.280\ \Omega}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

RELATED PROBLEM

Use Table 2–3 to determine the resistance of 100 ft of copper wire with a cross-sectional area of 810.1 CM. Compare with the calculated value.

As mentioned, Table 2–3 lists the resistance of the various standard wire sizes in ohms per 1000 ft at 20°C. For example, a 1000 ft length of 14-gauge copper wire has a resistance of 2.525 Ω. A 1000 ft length of 22-gauge wire has a resistance of 16.14 Ω. For a given length, the smaller-gauge wire has more resistance. Thus, for a given voltage, larger-gauge wires can carry more current than smaller ones.

SPLICING WIRES In some systems it is necessary to splice wires, especially when the run is long. Specific rules for splicing of electrical cables are given in the National Electric Code (NEC). Spicing is generally permissible provided that the wire is clean and a good electrical and mechanical connection is made. Most of the time, the splice is within an electrical box to protect it. For small-diameter wires, the crimp-type connectors are useful and quick. A crimping tool is used to make the connections.

Soldering the wires together in a splice is also permitted except on conductors used for grounding, provided that there is a good mechanical connection prior to soldering. After splicing, the splice should be encased in an insulation that is as good as the wire insulation.

One problem to avoid is splicing wires made from different conductors such as aluminum and copper. In this case, a chemical reaction called electrolysis can occur at the junction, leading to a high resistance connection that fails. Another problem with splicing different conductors is that they can expand differently if they are warmed, causing the splice to loosen and fail.

Ground

Ground is the reference point in an electric circuit. The term *ground* originated from the fact that one conductor of a circuit was typically connected with an 8-foot long metal rod driven into the earth itself. Today, this type of connection is referred to as an *earth ground*. In household wiring, earth ground is indicated with a green or bare copper wire. Earth ground is normally connected to the metal chassis of an appliance or a metal electrical box for safety. Unfortunately, there have been exceptions to this rule, which can present a safety hazard if a metal chassis is not at earth ground. It is a good idea to confirm that a metal chassis is actually at earth ground potential before doing any work on an instrument or appliance.

Another type of ground is called a *reference ground*. Voltages are always specified with respect to another point. If that point is not stated explicitly, the reference ground is understood. Reference ground defines 0 V for the circuit. The reference ground can be at a completely different potential than the earth ground. Reference ground is also called **common** and labeled COM or COMM because it represents a common conductor. When you are wiring a protoboard in the laboratory, you will normally reserve one of the bus strips (a long line along the length of the board) for this common conductor.

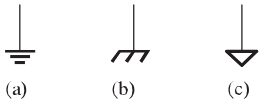

FIGURE 2–49 **Symbols for ground.**

Three ground symbols are shown in Figure 2–49. Unfortunately, there is not a separate symbol to distinguish between earth ground and reference ground. The symbol in (a) represents either an earth ground or a reference ground, (b) shows a chassis ground, and (c) is an alternate reference symbol typically used when there is more than one common connection (such as analog and digital ground in the same circuit). In this book, the symbol in part (a) will be used throughout.

Figure 2–50 illustrates a simple circuit with ground connections. The current is from the negative terminal of the 12 V source, through the common ground connection, through the lamp, and through the wire back to the positive terminal of the source. Ground provides a path for the current from the source because all of the ground points are electrically the same point and provide a zero resistance (ideally) current path. The voltage at the top of the circuit is +12 V with respect to ground. You can think of all the ground points in a circuit as being connected together by a conductor.

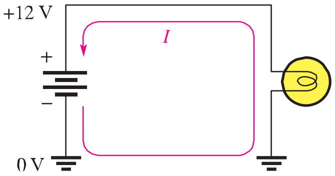

FIGURE 2–50 **A simple circuit with ground connections.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 2–6 CHECKUP

1. What are the basic elements of an electric circuit?
2. Define *open circuit*.
3. Define *closed circuit*.
4. What is the resistance of an open switch? Ideally, what is the resistance of a closed switch?
5. What is the purpose of a fuse?
6. What is the difference between a fuse and a circuit breaker?
7. Which wire is larger in diameter, AWG #3 or AWG #22?
8. What is ground in an electric circuit?

2–7 BASIC CIRCUIT MEASUREMENTS

In working on electrical or electronic circuits, you will frequently need to measure voltage, current, and resistance and to use meters safely and correctly.

After completing this section, you should be able to

- **Make basic circuit measurements**
  - **Properly measure voltage in a circuit**
  - **Properly measure current in a circuit**
  - **Properly measure resistance**
  - **Set up and read basic meters**

Voltage, current, and resistance measurements are commonly required in electrical and electronics work. The instrument used to measure voltage is a **voltmeter**, the instrument used to measure current is an **ammeter**, and the instrument used to measure resistance is an **ohmmeter**. Commonly, all three instruments are combined into a single instrument known as a **multimeter,** in which you can choose the specific quantity to measure by selecting the appropriate function with a switch.

Typical portable multimeters are shown in Figure 2–51. Part (a) shows a digital multimeter (DMM), which provides a digital readout of the measured quantity; and part (b) shows an analog meter with a needle pointer. Many digital multimeters also include a bar graph display.

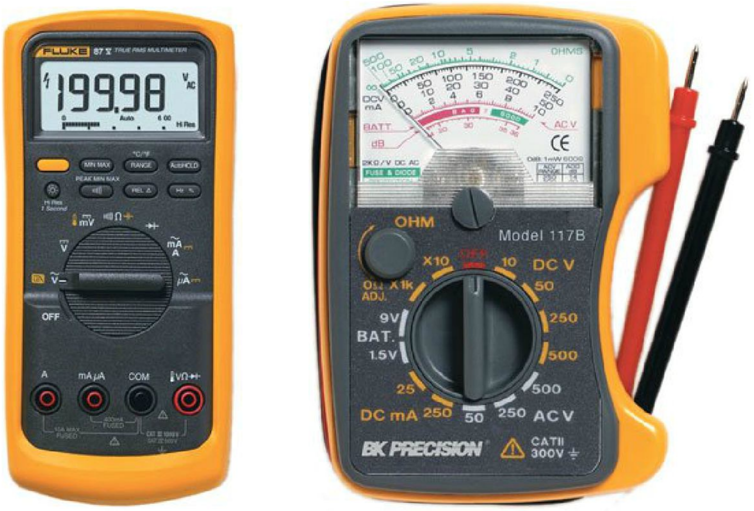

FIGURE 2–51 Typical portable multimeters. (a) Courtesy of Fluke Corporation. Reproduced with permission. (b) Courtesy of B+K Precision.

(a) Digital multimeter (b) Analog multimeter

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Meter Symbols

Throughout this book certain symbols will be used in circuits to represent meters, as shown in Figure 2–52. You may see any of four types of symbols for voltmeters, ammeters, and ohmmeters, depending on which symbol most effectively conveys the information required. The digital meter symbol is used when specific values are to be indicated in a circuit. The bar graph meter symbol and sometimes the analog meter symbol are used to illustrate the operation of a circuit when *relative* measurements or changes in quantities, rather than specific values, need to be depicted. A changing quantity may be indicated by an arrow in the display showing an increase or decrease. The generic symbol is used to indicate placement of meters in a circuit when no values or value changes need to be shown.

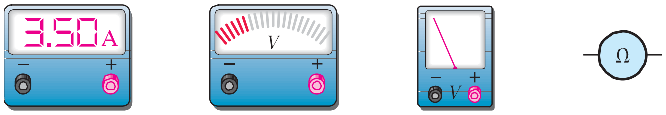

(a) Digital (b) Bar graph (c) Analog (d) Generic

FIGURE 2–52 **Examples of meter symbols used in this book. Each of the symbols can be used to represent either an ammeter (A), a voltmeter (V), or an ohmmeter (Ω).**

Measuring Current

Figure 2–53 illustrates how to measure current with an ammeter. Part (a) shows a simple circuit in which the current through a resistor is to be measured. Connect an ammeter in the current path by first opening the circuit, as shown in part (b). Then insert the meter as shown in part (c). Such a connection is a *series* connection. The polarity of the meter must be such that the current is in at the negative terminal and out at the positive terminal. If you are using a DMM to measure current, the leads must be moved to a special current jack.

6258202 2015/12/10 12:27:08.5

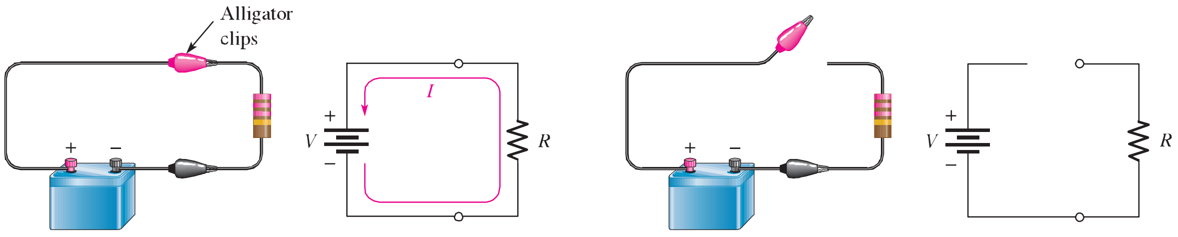

(a) Circuit in which the current is to be measured

(b) Open the circuit either between the resistor and the positive terminal or between the resistor and the negative terminal of source.

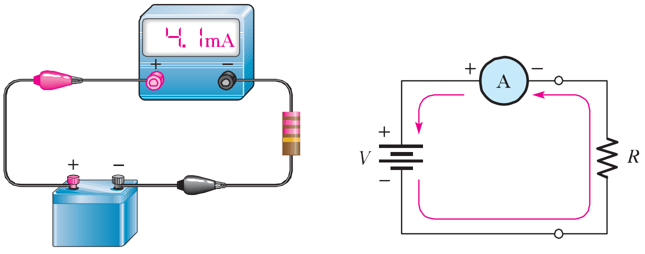

(c) Install the ammeter in the current path with polarity as shown (negative to negative, positive to positive).

FIGURE 2–53 **Example of an ammeter connection to measure current in a simple circuit.**

SAFETY NOTE

Never wear rings or any type of metallic jewelry while you work on a circuit. These items may accidentally come in contact with the circuit, causing shock and/or damage to the circuit. With high-energy sources, such as a car battery, a short across jewelry (a watch or ring) can become hot instantly, causing burns to the wearer.

Measuring Voltage

To measure voltage, connect a voltmeter across the component for which the voltage is to be found. Such a connection is a *parallel* connection. The negative terminal of the meter must be connected to the negative side of the circuit, and the positive terminal of the meter must be connected to the positive side of the circuit. Figure 2–54 shows a voltmeter connected to measure the voltage across the resistor.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

FIGURE 2–54 **Example of a voltmeter connection to measure voltage in a simple circuit.**

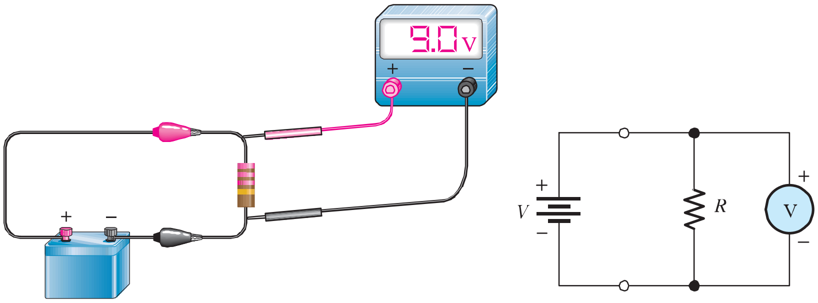

Measuring Resistance

To measure resistance, connect an ohmmeter across the resistor. *The resistor must first be removed or disconnected from the circuit.* This procedure is shown in Figure 2–55.

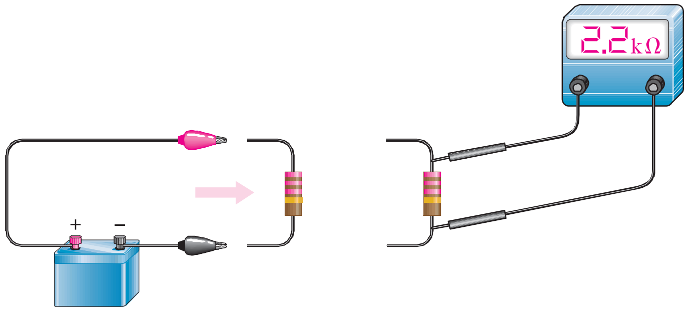

(a) Disconnect the resistor from the circuit to avoid damage to the meter and/or incorrect measurement.

(b) Measure the resistance. (Polarity is not important.)

FIGURE 2–55 **Example of using an ohmmeter to measure resistance.**

Digital Multimeters (DMMs)

A **DMM** is a multifunction electronic instrument that can measure voltage, current, or resistance. DMMs are the most widely used type of electronic measuring instrument. Generally, DMMs provide more functions, better accuracy, greater ease of reading, and greater reliability than do analog meters, which are covered next. Analog meters have at least one advantage over DMMs, however. They can track short-term variations and trends in a measured quantity that many DMMs are too slow to respond to. Typical DMMs are shown in Figure 2–56. Many DMMs are autoranging types in which the proper range is automatically selected by internal circuitry.

FIGURE 2–56 **Typical digital multimeters (DMMs).** (Courtesy of B+K Precision)

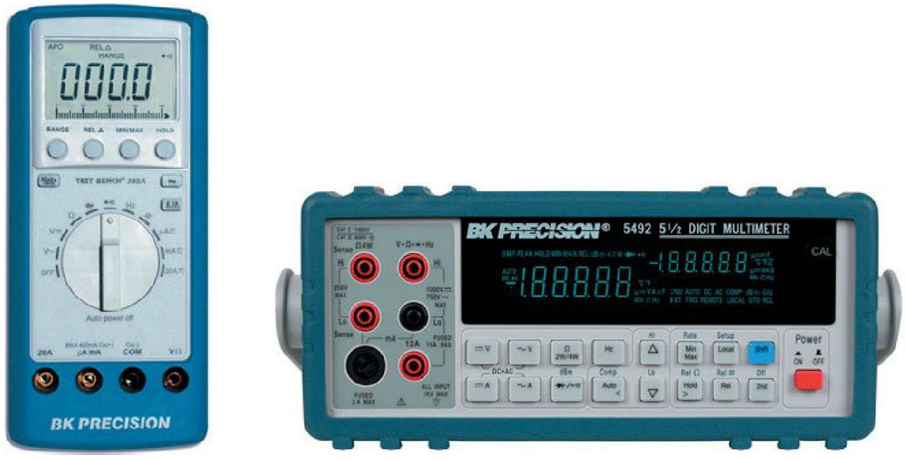

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

DMM FUNCTIONS The basic functions found on most DMMs are

- Ohms
- DC voltage and current
- AC voltage and current

Some DMMs provide additional functions such as analog bar graph displays, transistor or diode tests, power measurement, and decibel measurement for audio amplifier tests.

DMM DISPLAYS DMMs are available with either LCD (liquid-crystal display) or LED (light-emitting diode) readouts. The LCD is the most commonly used readout in battery-powered instruments because it requires only very small amounts of current. A typical battery-powered DMM with an LCD readout operates on a 9 V battery that will last from a few hundred hours to 2000 hours and more. The disadvantages of LCD readouts are that (a) they are difficult or impossible to see in low-light conditions and (b) they are relatively slow to respond to measurement changes. LEDs, on the other hand, can be seen in the dark and respond quickly to changes in measured values. LED displays require much more current than LCD displays; and, therefore, battery life is shortened when LEDs are used in portable equipment.

SAFETY NOTE

Manufacturers use the CE mark to indicate that a product meets all the essential requirements of the European Directives for health and safety. For example, DMMs that meet the directive IEC 61010-1 are in compliance and can use the CE mark on the DMM. This mark is used on many products in various countries much like the UL (underwriters lab) mark that originated in the U.S. It also shows compliance with safety standards.

Both LCD and LED DMM displays are in a seven-segment format. Each digit in the display consists of seven separate segments, as shown in Figure 2–57(a). Each of the ten decimal digits is formed by the activation of appropriate segments, as illustrated in Figure 2–57(b). In addition to the seven segments, there is also a decimal point.

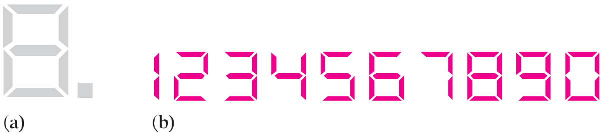

(a) (b)

FIGURE 2–57 **Seven-segment display.**

RESOLUTION The **resolution** of a DMM is the smallest increment of a quantity that the DMM can measure. The smaller the increment, the better the resolution. One factor that determines the resolution of a meter is the number of digits in the display.

Because many DMMs have 3½ digits in their display, we will use this case for illustration. *A* 3½-digit multimeter has three digit positions that can indicate from 0 through 9, and one digit position that can indicate only a value of 1. This latter digit, called the *half-digit,* is always the most significant digit in the display. For example, suppose that a DMM is reading .999 V, as shown in Figure 2–58(a). If the voltage increases by 0.001 V to 1 V, the display correctly shows 1.000 V, as shown in part (b). The "1" is the half-digit. Thus, with 3½ digits, a variation of 0.001 V, which is the resolution, can be observed.

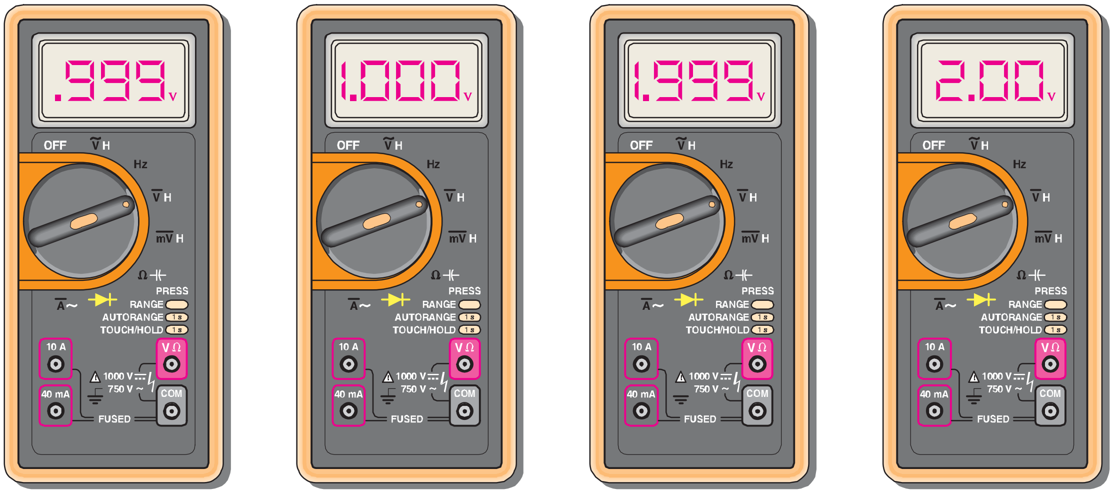

(a) Resolution: 0.001 V (b) Resolution: 0.001 V (c) Resolution: 0.001 V (d) Resolution: 0.01 V

FIGURE 2–58 **A 3½-digit DMM illustrates how the resolution changes with the number of digits in use.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

HANDS ON TIP

When reading the scales on an analog meter, always view the scale and "needle" directly from the front and not from an angle. This practice will avoid *parallax,* which is an apparent change in position of the needle relative to the meter scale and results in inaccurate readings.

Now, suppose that the voltage increases to 1.999 V. This value is indicated on the meter as shown in Figure 2–58(c). If the voltage increases by 0.001 V to 2 V, the half-digit cannot display the "2," so the display shows 2.00. The half-digit is blanked and only three digits are active, as indicated in part (d). With only three digits active, the resolution is 0.01 V rather than 0.001 V as it is with 3½ active digits. The resolution remains 0.01 V up to 19.99 V. The resolution goes to 0.1 V for readings of 20.0 V to 199.9 V. At 200 V, the resolution goes to 1 V, and so on.

The resolution capability of a DMM is also determined by the internal circuitry and the rate at which the measured quantity is sampled. DMMs with displays of 4½ through 8½ digits are also available.

ACCURACY As defined in Chapter 1, accuracy is an indication, usually expressed in percentage, of the range of error which is the difference in the measured and true or accepted value of a quantity. The accuracy of a DMM is established strictly by its internal circuitry and calibration. For typical meters, accuracies range from 0.01% to 0.5%, with some precision laboratory-grade meters going to 0.002%.

Reading Analog Multimeters

Although the DMM is predominately used, you may have to use on analog meter on occasion.

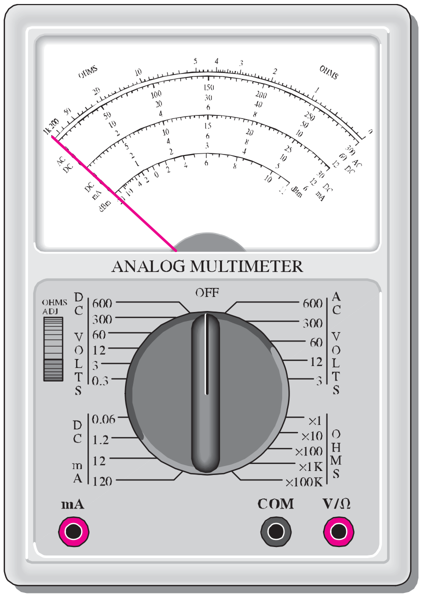

FIGURE 2–59 **A typical analog multimeter.**

FUNCTIONS The face of a typical analog (needle-type) multimeter is represented in Figure 2–59. This particular instrument can be used to measure both direct current (dc) and alternating current (ac) quantities as well as resistance values. It has four selectable functions: dc volts (DC VOLTS), dc milliamperes (DC mA), ac volts (AC VOLTS), and OHMS. Many analog multimeters are similar to this one, although range selections and scales may vary.

RANGES Within each function there are several ranges, as indicated by the brackets around the selector switch. For example, the DC VOLTS function has 0.3 V, 3 V, 12 V, 60 V, 300 V, and 600 V ranges. Thus, dc voltages from 0.3 V full-scale to 600 V full-scale can be measured. On the DC mA function, direct currents from 0.06 mA full-scale to 120 mA full-scale can be measured. On the ohm scale, the range settings are ×1, ×10, ×100, ×1000, and ×100,000.

THE OHM SCALE Ohms are read on the top scale of the meter. This scale is nonlinear; that is, the values represented by each division (large or small) vary as you go across the scale. In Figure 2–59, notice how the scale becomes more compressed as you go from right to left.

To read the actual value in ohms, multiply the number on the scale as indicated by the pointer by the factor selected by the switch. For example, when the switch is set at ×100 and the pointer is at 20, the reading is 20 × 100 = 2000 Ω.

As another example, assume that the switch is at ×10 and the pointer is at the seventh small division between the 1 and 2 marks, indicating 17 Ω (1.7 × 10). Now, if the meter remains connected to the same resistance and the switch setting is changed to ×1, the pointer will move to the second small division between the 15 and 20 marks. This, of course, is also a 17 Ω reading, illustrating that a given resistance value can often be read at more than one switch setting. However, the meter should be *zeroed* each time the range is changed by touching the leads together and adjusting the needle.

THE AC-DC AND DC mA SCALES The second, third, and fourth scales from the top (labeled "AC" and "DC") are used in conjunction with the DC VOLTS and AC VOLTS functions. The upper ac-dc scale ends at the 300 mark and is used with range settings such as 0.3, 3, and 300. For example, when the switch is at 3 on the DC VOLTS function, the

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

300 scale has a full-scale value of 3 V; at the range setting of 300, the full-scale value is 300 V. The middle ac-dc scale ends at 60. This scale is used in conjunction with range settings such as 0.06, 60, and 600. For example, when the switch is at 60 on the DC VOLTS function, the full-scale value is 60 V. The lower ac-dc scale ends at 12 and is used in conjunction with switch settings such as 1.2, 12, and 120. The three DC mA scales are used in a similar way to measure current.

HANDS ON TIP

When you are using a multimeter, such as the analog multimeter illustrated in Figure 2–59, where you manually select the voltage and current ranges, it is good practice to always set the multimeter on the maximum range before you measure an unknown voltage or current. You can then reduce the range until you get an acceptable reading.

EXAMPLE 2–9

In Figure 2–60, determine the quantity (voltage, current, or resistance) that is being measured and its value for each of the following specified switch settings on the meter in Figure 2–59.

(a) DC volts: 60 **(b)** DC mA: 12 **(c)** OHMS: ×1K

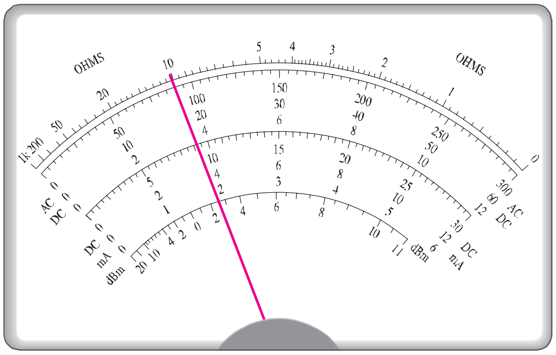

FIGURE 2–60

SOLUTION

(a) The reading taken from the middle AC-DC scale is **18 V**.

(b) The reading taken from the lower AC-DC scale is **3.8 mA**.

(c) The reading taken from the ohms scale (top) is **10 kΩ**.

RELATED PROBLEM

For part (c) in this example, the switch is moved to the ×100 ohms setting. Assuming that the same resistance is being measured, what will the needle do?

SECTION 2–7 CHECKUP

1. Name the multimeter functions for measuring
   - **(a)** current
   - **(b)** voltage
   - **(c)** resistance
2. Show how to place two ammeters in the circuit of Figure 2–41 to measure the current through either lamp (be sure to observe the polarities). How can the same measurements be accomplished with only one ammeter?
3. Show how to place a voltmeter to measure the voltage across lamp 2 in Figure 2–41.
4. List two common types of DMM displays, and discuss the advantages and disadvantages of each.
5. Define *resolution* in a DMM.
6. The analog multimeter in Figure 2–59 is set on the 3 V range to measure dc voltage. Assume the pointer is at 150 on the upper ac-dc scale. What voltage is being measured?
7. How do you set up the multimeter in Figure 2–59 to measure 275 V dc, and on what scale do you read the voltage?
8. If you expect to measure a resistance in excess of 20 kΩ with the multimeter in Figure 2–59, where do you set the switch?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SUMMARY

- An atom is the smallest particle of an element that retains the characteristics of that element.
- The electron is the basic particle of negative electrical charge.
- The proton is the basic particle of positive charge.
- An ion is an atom that has gained or lost an electron and is no longer neutral.
- When electrons in the outer orbit of an atom (valence electrons) break away, they become free electrons.
- Free electrons make current possible.
- Like charges repel each other, and opposite charges attract each other.
- Voltage must be applied to a circuit before there can be current.
- Fuel cells and batteries convert chemical energy to electrical energy using oxidation-reduction reaction.
- Resistance limits the current.
- Basically, an electric circuit consists of a source, a load, and a current path.
- An open circuit is one in which the current path is broken.
- A closed circuit is one which has a complete current path.
- An ammeter is connected in line (series) with the current path to measure current.
- A voltmeter is connected across (parallel) the current path to measure voltage.
- An ohmmeter is connected across a resistor to measure resistance. The resistor must be disconnected from the circuit.
- Figure 2–61 shows the electrical symbols introduced in this chapter.

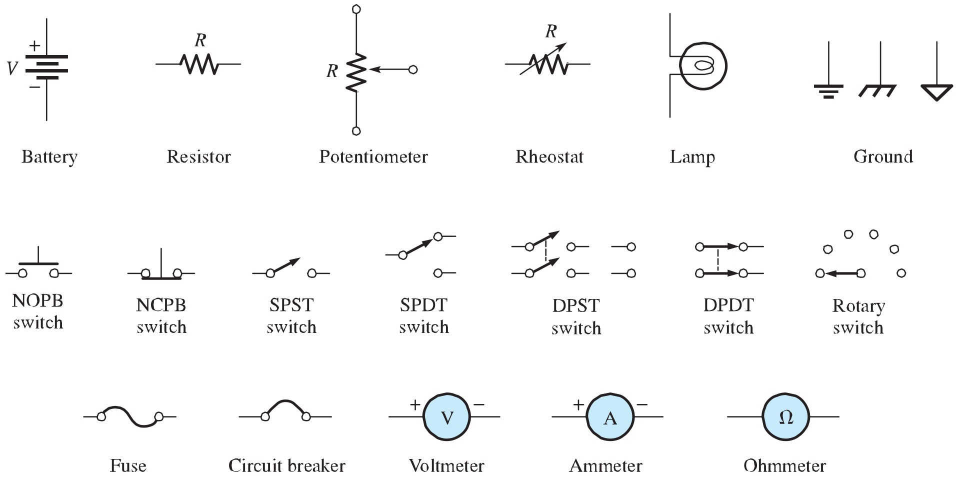

FIGURE 2–61

- One coulomb is the charge on 6.25×10^{18} electrons.
- One volt is the potential difference (voltage) between two points when one joule of energy is used to move one coulomb of charge from one point to the other.
- One ampere is the amount of current that exists when one coulomb of charge moves through a given cross-sectional area of a material in one second.
- One ohm is the resistance when there is one ampere of current in a material with one volt applied across the material.

KEY TERMS

Key terms and other bold terms in the chapter are defined in the end-of-book glossary.

Ammeter An electrical instrument used to measure current.

Ampere (A) The unit of electrical current.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Atom The smallest particle of an element possessing the unique characteristics of that element.

AWG American Wire Gauge; a standardization based on wire diameter.

Charge An electrical property of matter that exists because of an excess or a deficiency of electrons. Charge can be either positive or negative.

Circuit breaker A resettable protective device used for interrupting excessive current in an electric circuit.

Closed circuit A circuit with a complete current path.

Conductance The ability of a circuit to allow current. The unit is the siemens (S).

Conductor A material in which electric current is easily established. An example is copper.

Coulomb (C) The unit of electrical charge; the total charge possessed by 6.25×10^{18} electrons.

Coulomb's law A law that states a force exists between two charged bodies that is directly proportional to the product of the two charges and inversely proportional to the square of the distance between them.

Current The rate of flow of charge (free electrons).

Current source A device that produces a constant current for a varying load.

DMM Digital multimeter; an electronic instrument that combines meters for measurement of voltage, current, and resistance.

Electron A basic particle of electrical charge in matter. The electron possesses negative charge.

Free electron A valence electron that has broken away from its parent atom and is free to move from atom to atom within the atomic structure of a material.

Fuel cell A device that converts electrochemical energy from an external source into dc voltage. The hydrogen fuel cell is the most common type.

Fuse A protective device that burns open when there is excessive current in a circuit.

Ground The common or reference point in a circuit.

Insulator A material that does not allow current under normal conditions.

Load An element (resistor or other component) connected across the output terminals of a circuit that draws current from the source and upon which work is done.

Ohm (Ω) The unit of resistance.

Ohmmeter An instrument for measuring resistance.

Open circuit A circuit in which there is not a complete current path.

Potentiometer A three-terminal variable resistor.

Resistance Opposition to current. The unit is the ohm (Ω).

Resistor An electrical component designed specifically to have a certain amount of resistance.

Rheostat A two-terminal variable resistor.

Schematic A symbolized diagram of an electrical or electronic circuit.

Semiconductor A material that has a conductance value between that of a conductor and an insulator. Silicon and germanium are examples.

Siemens (S) The unit of conductance.

Switch An electrical or electronic device for opening and closing a current path.

Volt (V) The unit of voltage or electromotive force.

Voltage The amount of energy per charge available to move electrons from one point to another in an electric circuit.

Voltage source A device that produces a constant voltage for a varying load.

Voltmeter An instrument used to measure voltage.

KEY FORMULAS

6238262 2013/12/10 12.27.66.5

(2–1)	$Q = \dfrac{\text{number of electrons}}{6.25 \times 10^{18}\ \text{electrons/C}}$	Charge
(2–2)	$V = \dfrac{W}{Q}$	Voltage in volts equals energy in joules divided by charge in coulombs.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

(2–3)	$I = \frac{Q}{t}$	Current in amperes equals charge in coulombs divided by time in seconds.
(2–4)	$G = \frac{1}{R}$	Conductance in siemens is the reciprocal of resistance in ohms.
(2–5)	$A = d^2$	Cross-sectional area in circular mils equals the diameter in mils squared.
(2–6)	$R = \frac{\rho l}{A}$	Resistance is resistivity in CM Ω/ft times length in feet divided by cross-sectional area in circular mils.

TRUE/FALSE QUIZ

Answers are at the end of the chapter.

1. The number of neutrons in the nucleus is the atomic number of that element.

2. The unit of charge is the ampere.

3. Energy in a battery is stored in the form of chemical energy.

4. A volt can be defined in terms of energy per charge.

5. In a five-band precision resistor, the fourth band is the tolerance band.

6. A rheostat performs the same function as a potentiometer.

7. A strain gage changes resistance in response to an applied force.

8. Soldering is never allowed on a splice.

9. All circuits must have a complete path for current.

10. A circular mil is a unit of area.

11. The three basic measurements that can be done by a DMM are voltage, current, and power.

12. To measure current with a meter, the meter should be placed in series.

SELF-TEST

Answers are at the end of the chapter.

1. A neutral atom with an atomic number of three has how many electrons?
(a) 1 **(b)** 3 **(c)** none **(d)** depends on the type of atom

2. Electron orbits are called
(a) shells **(b)** nuclei **(c)** waves **(d)** valences

3. Materials in which current cannot be established are called
(a) filters **(b)** conductors **(c)** insulators **(d)** semiconductors

4. When placed close together, a positively charged material and a negatively charged material will
(a) repel **(b)** become neutral **(c)** attract **(d)** exchange charges

5. The charge on a single electron is
(a) 6.25×10^{-18} C **(b)** 1.6×10^{-19} C **(c)** 1.6×10^{-19} J **(d)** 3.14×10^{-6} C

6. *Potential difference* is another term for
(a) energy **(b)** voltage **(c)** distance of an electron from the nucleus **(d)** charge

7. The unit of energy is the
(a) watt **(b)** coulomb **(c)** joule **(d)** volt

8. Which one of the following is not a type of energy source?
(a) battery **(b)** solar cell **(c)** generator **(d)** potentiometer

9. Which one of the following is the byproduct of a hydrogen fuel cell?
(a) oxygen **(b)** carbon dioxide **(c)** hydrochloric acid **(d)** water

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

10. Which one of the following is generally not a possible condition in an electric circuit?
(a) voltage and no current **(b)** current and no voltage
(c) voltage and current **(d)** no voltage and no current

11. Electrical current is defined as
(a) free electrons **(b)** the rate of flow of free electrons
(c) the energy required to move electrons **(d)** the charge on free electrons

12. There is no current in a circuit when
(a) a series switch is closed **(b)** a series switch is open **(c)** there is no source voltage
(d) both (a) and (c) **(e)** both (b) and (c)

13. The primary purpose of a resistor is to
(a) increase current **(b)** limit current **(c)** produce heat **(d)** resist current change

14. Potentiometers and rheostats are types of
(a) voltage sources **(b)** variable resistors **(c)** fixed resistors **(d)** circuit breakers

15. The current in a given circuit is not to exceed 22 A. Which value of fuse is best?
(a) 10 A **(b)** 25 A **(c)** 20 A **(d)** a fuse is not necessary

PROBLEMS

Answers to odd-numbered problems are at the end of the book.

BASIC PROBLEMS

SECTION 2–2 Electrical Charge

1. How many coulombs of charge do 50×10^{31} electrons possess?

2. How many electrons does it take to make 80 μC of charge?

3. What is the charge in coulombs of the nucleus of a copper atom?

4. What is the charge in coulombs of the nucleus of a chlorine atom?

SECTION 2–3 Voltage

5. Determine the voltage in each of the following cases:
(a) 10 J/C **(b)** 5 J/2 C **(c)** 100 J/25 C

6. Five hundred joules of energy are used to move 100 C of charge through a resistor. What is the voltage across the resistor?

7. What is the voltage of a battery that uses 800 J of energy to move 40 C of charge through a resistor?

8. How much energy does a 12 V battery in your car use to move 2.5 C through the electrical circuit?

9. Assume that a solar battery charger delivers 2.5 J of energy when 0.2 C of charge has been moved. What is the voltage?

SECTION 2–4 Current

10. If the solar cell in Problem 9 has moved the charge in 10 s, what is the current?

11. Determine the current in each of the following cases:
(a) 75 C in 1 s **(b)** 10 C in 0.5 s **(c)** 5 C in 2 s

12. Six-tenths coulomb passes a point in 3 s. What is the current in amperes?

13. How long does it take 10 C to flow past a point if the current is 5 A?

14. How many coulombs pass a point in 0.1 s when the current is 1.5 A?

SECTION 2–5 Resistance

15. Figure 2–62(a) shows color-coded resistors. Determine the resistance value and the tolerance of each.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

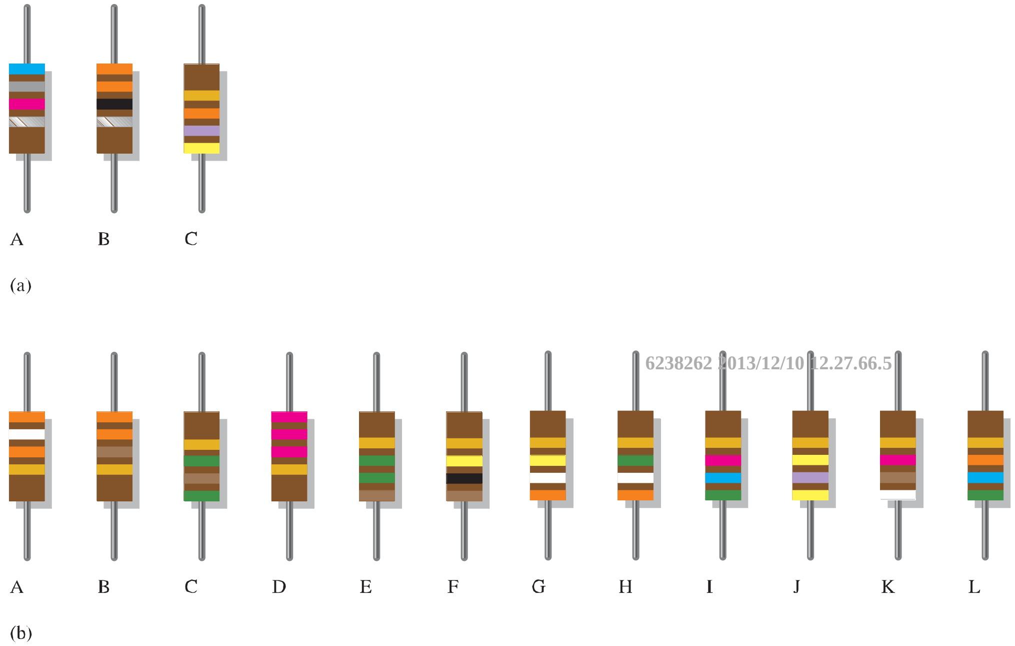

6238262 2013/12/10 12.27.66.5

FIGURE 2–62

16. Find the minimum and the maximum resistance within the tolerance limits for each resistor in Figure 2–62(a).

17. **(a)** If you need a 270 Ω resistor with 5% tolerance, what color bands should you look for?
(b) From the selection of resistors in Figure 2–62(b), choose the following values: 330 Ω, 2.2 kΩ, 39 kΩ, 56 kΩ, and 100 kΩ.

18. Determine the resistance value and tolerance for each resistor in Figure 2–63.

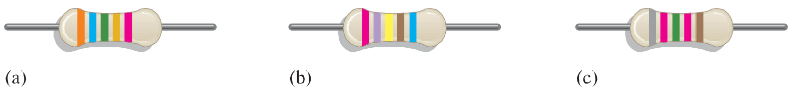

FIGURE 2–63

19. Determine the resistance and tolerance of each of the following 4-band resistors:
(a) brown, black, black, gold **(b)** green, brown, green, silver **(c)** blue, gray, black, gold

20. Determine the color bands for each of the following 4-band resistors. Assume each has a 5% tolerance.
(a) 0.47 Ω **(b)** 270 kΩ **(c)** 5.1 MΩ

21. Determine the resistance and tolerance of each of the following 5-band resistors:
(a) red, gray, violet, red, brown **(b)** blue, black, yellow, gold, brown
(c) white, orange, brown, brown, brown

22. Determine the color bands for each of the following 5-band resistors. Assume each has a 1% tolerance.
(a) 14.7 kΩ **(b)** 39.2 Ω **(c)** 9.76 kΩ

23. Determine the resistance values represented by the following labels:
(a) 220 **(b)** 472 **(c)** 823 **(d)** 3K3 **(e)** 560 **(f)** 10M

24. The adjustable contact of a linear potentiometer is set at the mechanical center of its adjustment. If the total resistance is 1000 Ω, what is the resistance between each end terminal and the adjustable contact?

SECTION 2–6 The Electric Circuit

25. Trace the current path in the lamp circuit of Figure 2–41(a) with the switch making contact between the middle and lower pins.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

26. With the switch in either position, redraw the lamp circuit in Figure 2–41(b) with a fuse connected to protect the circuit against excessive current.

SECTION 2–7 Basic Circuit Measurements

27. Show the placement of an ammeter and a voltmeter to measure the current and the source voltage in Figure 2–64.

28. Show how you would measure the resistance of R_2 in Figure 2–64.

29. In Figure 2–65 what does each voltmeter indicate when the switch (SW) is in position 1? In position 2?

30. In Figure 2–65, show how to connect an ammeter to measure the current from the voltage source regardless of the switch (SW) position.

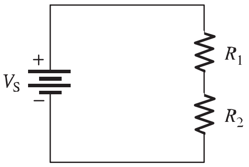

FIGURE 2–64

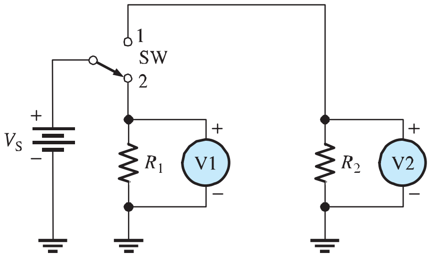
1
SW
2
+
V_S
−
R_1
+
V1
−
R_2
+
V2
−

FIGURE 2–65

31. What is the voltage reading of the meter in Figure 2–66?

32. How much resistance is the meter in Figure 2–67 measuring?

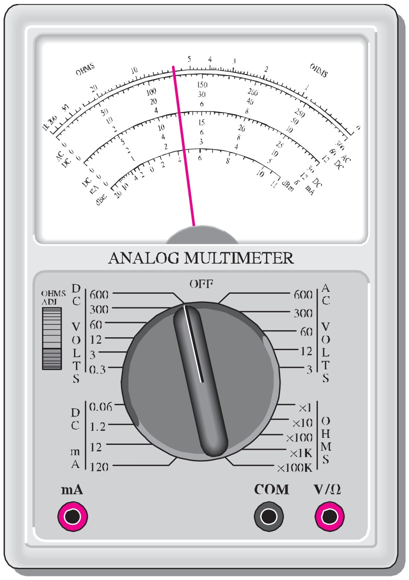
OHMS
ANALOG MULTIMETER
OFF
OHMS ADJ
DC VOLTS
600
300
60
12
3
0.3
AC VOLTS
600
300
60
12
3
DC mA
0.06
1.2
12
120
OHMS
×1
×10
×100
×1K
×100K
mA
COM
V/Ω

FIGURE 2–66

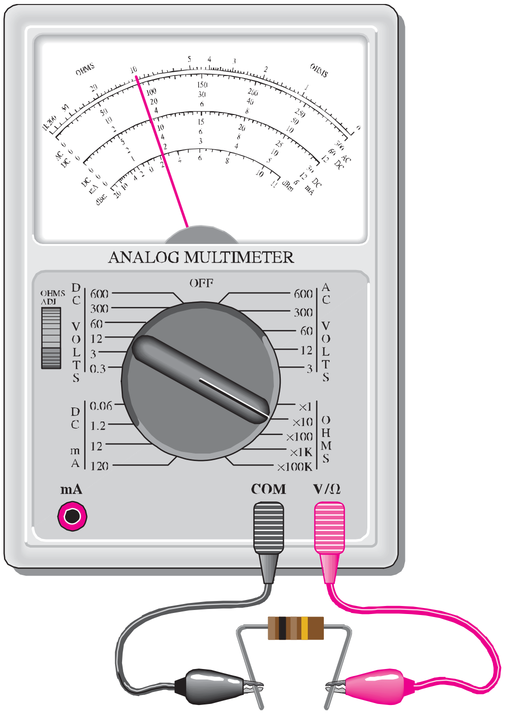
OHMS
ANALOG MULTIMETER
OFF
OHMS ADJ
DC VOLTS
600
300
60
12
3
0.3
AC VOLTS
600
300
60
12
3
DC mA
0.06
1.2
12
120
OHMS
×1
×10
×100
×1K
×100K
mA
COM
V/Ω

FIGURE 2–67

33. Determine the resistance indicated by each of the following ohmmeter readings and range settings:
(**a**) pointer at 2, range setting at R × 100 (**b**) pointer at 15, range setting at R × 10M
(**c**) pointer at 45, range setting at R × 100

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

34. A multimeter has the following ranges: 1 mA, 10 mA, 100 mA; 100 mV, 1 V, 10 V; R × 1, R × 10, R × 100. Indicate schematically how you would connect the multimeter in Figure 2–68 to measure the following quantities:

(a) I_{R1} **(b)** V_{R1} **(c)** R_1

In each case indicate the *function* to which you would set the meter and the *range* that you would use.

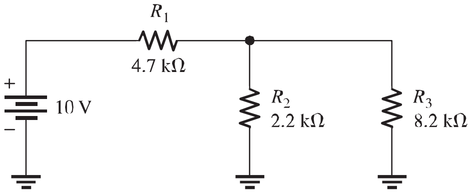

FIGURE 2–68

ADVANCED PROBLEMS

35. A resistor with a current of 2 A through it in an amplifier circuit converts 1000 J of electrical energy to heat energy in 15 s. What is the voltage across the resistor?

36. If 574 × 10^{15} electrons flow through a wire in 250 ms, what is the current in amperes?

37. A 120 V source is to be connected to a 1500 Ω resistive load by two lengths of wire as shown in Figure 2–69. The voltage source is to be located 50 ft from the load. Using Table 2–3, determine the gauge number of the *smallest* wire that can be used if the total resistance of the two lengths of wire is not to exceed 6 Ω.

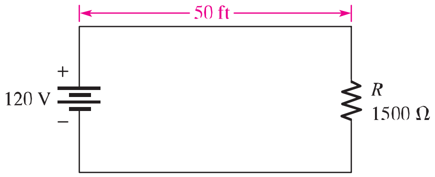

FIGURE 2–69

38. Determine the resistance and tolerance of each resistor labeled as follows:

(a) 4R7J **(b)** 560KF **(c)** 1M5G

39. There is only one circuit in Figure 2–70 in which it is possible to have all lamps on at the same time. Determine which circuit it is.

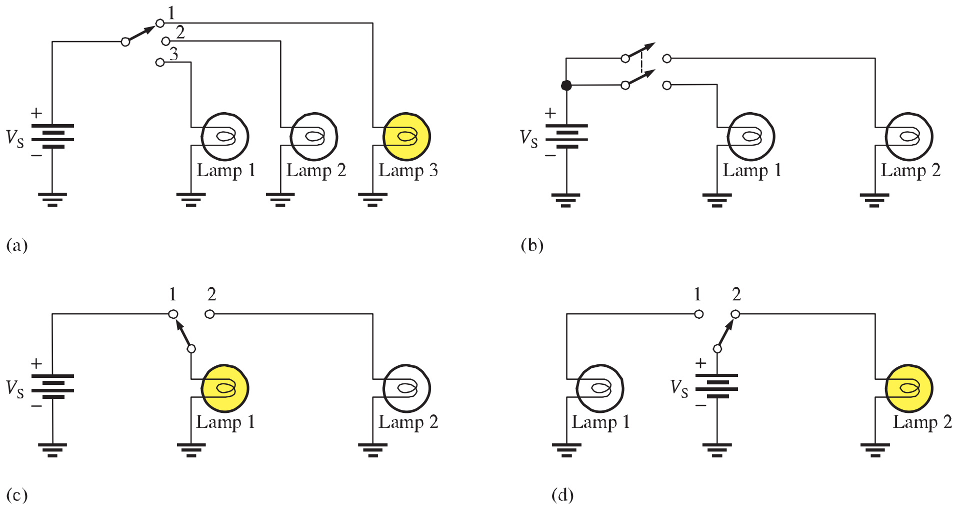

FIGURE 2–70

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

40. Through which resistor in Figure 2–71 is there always current, regardless of the position of the switches?

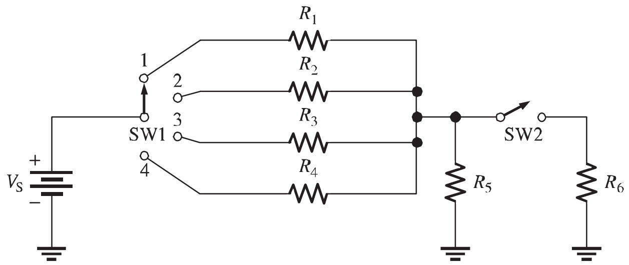

FIGURE 2–71

41. In Figure 2–71, show the proper placement of ammeters to measure the current through each resistor and the current out of the battery.

42. Show the proper placement of voltmeters to measure the voltage across each resistor in Figure 2–71.

43. Devise a switch arrangement whereby two voltage sources V_{S1} and V_{S2} can be connected simultaneously to either of two resistors (R_1 and R_2) as follows:

V_{S1} connected to R_1 and V_{S2} connected to R_2
V_{S1} connected to R_2 and V_{S2} connected to R_1

44. Show how to wire the quiz board in Figure 2–46 if the correct sequence of answers is D-B-C-A.

ANSWERS TO SECTION CHECKUPS

SECTION 2–1 Atoms

1. The electron is the basic particle of negative charge.
2. An atom is the smallest particle of an element that retains the unique characteristics of the element.
3. An atom is a positively charged nucleus surrounded by orbiting electrons.
4. Atomic number is the number of protons in a nucleus.
5. No, each element has a different type of atom.
6. A free electron is an outer-shell electron that has drifted away from the parent atom.
7. Shells are energy bands in which electrons orbit the nucleus of an atom.
8. Copper and silver

SECTION 2–2 Electrical Charge

1. Q is the symbol for charge.
2. Coulomb is the unit of charge, and C is the symbol for coulomb.
3. The two types of charge are positive and negative.
4. $Q = \dfrac{10 \times 10^{12}\text{ electrons}}{6.25 \times 10^{18}\text{ electrons/C}} = 1.6 \times 10^{-6}\text{ C} = 1.6\ \mu\text{C}$

SECTION 2–3 Voltage

6238262 2013/12/10 12.27.66.5

1. Voltage is energy per unit charge.
2. The unit of voltage is the volt.
3. $V = W/Q = 24\text{ J}/10\text{ C} = 2.4\text{ V}$
4. Battery, fuel cell, power supply, solar cell, generator, thermocouple, and piezoelectric sensors are voltage sources.
5. Oxidation-reduction reactions

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 2–4 Current

1. Current is the rate of flow of charge; **the** unit of current is the ampere (A).
2. There are 6.25×10^{18} electrons in one coulomb.
3. $I = Q/t = 20\text{ C}/4\text{ s} = 5\text{ A}$

SECTION 2–5 Resistance

1. Resistance is opposition to current and its unit is the ohm (Ω).
2. Two resistor categories are fixed and variable. The value of a fixed resistor cannot be changed, but that of a variable resistor can.
3. *First band:* first digit of resistance value.

   Second band: second digit of resistance value.

   Third band: number of zeros following the 2nd digit.

   Fourth band: percent tolerance.
4. **(a)** Yellow, violet, red, gold = 4700 Ω ± 5%
   (b) Blue, red, orange, silver = 62,000 Ω ± 10%
   (c) Brown, gray, black, gold = 18 Ω ± 5%
   (d) Red, red, blue, red, green = 22.6 kΩ ± 0.5%
5. **(a)** 33R = 33 Ω **(b)** 5K6 = 5.6 kΩ **(c)** 900 = 900 Ω **(d)** 6M8 = 6.8 MΩ
6. A rheostat has two terminals; a potentiometer has three terminals.
7. Thermistor—temperature; photoconductor cell—light; strain gage—force.

SECTION 2–6 The Electric Circuit

1. A basic electric circuit consists of source, load, and current path between source and load.
2. An open circuit is one that has no path for current.
3. A closed circuit is one that has a complete path for current.
4. $R = \propto$ (infinite); $R = 0\ \Omega$
5. A fuse protects a circuit against excessive current.
6. A fuse must be replaced once blown. A circuit breaker can be reset once tripped.
7. AWG #3 is larger than AWG #22.
8. Ground is the reference point with zero voltage with respect to other points.

SECTION 2–7 Basic Circuit Measurements

1. **(a)** Ammeter measures current.
   (b) Voltmeter measures voltage.
   (c) Ohmmeter measures resistance.
2. See Figure 2–72.
3. See Figure 2–73.

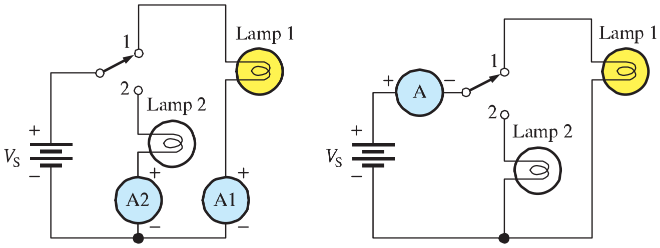

(a) Two ammeters (b) One ammeter

FIGURE 2–72

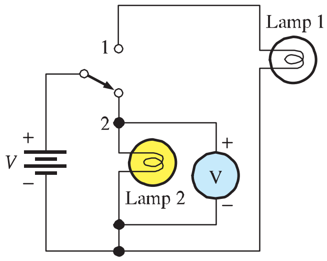

FIGURE 2–73

4. Two types of DMM displays are LED and LCD. The LCD requires little current, but it is difficult to see in low light and is slow to respond. The LED can be seen in the dark, and it responds quickly; however, it requires much more current than does the LCD.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

5. Resolution is the smallest increment of a quantity that a meter can measure.
6. The voltage being measured is 1.5 V.
7. DC VOLTS, 600 setting; 275 V is read on the 60 scale near midpoint.
8. OHMS × 1000

ANSWERS TO RELATED PROBLEMS FOR EXAMPLES

2–1 1.88×10^{19} electrons
2–2 600 J
2–3 12 C
2–4 4700 Ω ± 5%
2–5 47.5 Ω ± 2%
2–6 1.2 kΩ
2–7 404.01 CM; #24
2–8 1.280 Ω; same as calculated result
2–9 The needle will indicate 100 on the top scale.

ANSWERS TO TRUE/FALSE QUIZ

1. F **2.** F **3.** T **4.** T **5.** F **6.** F **7.** T **8.** F **9.** T
10. T **11.** F **12.** T

ANSWERS TO SELF-TEST

1. (b) **2.** (a) **3.** (c) **4.** (c) **5.** (b) **6.** (b) **7.** (c) **8.** (d)
9. (d) **10.** (b) **11.** (b) **12.** (e) **13.** (b) **14.** (b) **15.** (c)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

CHAPTER 3

OHM’S LAW, ENERGY, AND POWER

OUTLINE

3–1 Ohm’s Law
3–2 Application of Ohm’s Law
3–3 Energy and Power
3–4 Power in an Electric Circuit
3–5 The Power Rating of Resistors
3–6 Energy Conversion and Voltage Drop in a Resistance
3–7 Power Supplies and Batteries
3–8 Introduction to Troubleshooting

OBJECTIVES

- Explain Ohm’s law
- Use Ohm’s law to determine voltage, current, or resistance
- Define *energy* and *power*
- Calculate power in a circuit
- Properly select resistors based on power consideration
- Explain energy conversion and voltage drop
- Discuss characteristics of power supplies and batteries
- Describe a basic approach to troubleshooting

KEY TERMS

Ohm’s law
Linear
Energy
Power
Joule (J)
Watt (W)
Kilowatt-hour (kWh)
Watt’s law
Power rating
Voltage drop
Power supply
Efficiency
Ah rating
Troubleshooting
Half-splitting

INTRODUCTION

Georg Simon Ohm (1787–1854) experimentally found that voltage, current, and resistance are all related in a specific way. This basic relationship, known as *Ohm’s law,* is one of the most fundamental and important laws in the fields of electricity and electronics. In this chapter, Ohm’s law is examined, and its use in practical circuit applications is discussed and demonstrated by numerous examples.

In addition to Ohm’s law, the concepts and definitions of energy and power in electric circuits and systems are introduced, and the Watt’s law power formulas are given. A general approach to troubleshooting using the analysis, planning, and measurement (APM) method is also introduced.

6238262 2013/12/10 12.27.66.5

VISIT THE WEBSITE
Study aids for this chapter are available at
http://pearsonhighered.com/floyd

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

3–1 OHM'S LAW

Ohm's law describes mathematically how voltage, current, and resistance in a circuit are related. Ohm's law can be written in three equivalent forms; the formula you use depends on the quantity you need to determine.

After completing this section, you should be able to

- **Explain Ohm's law**
  - **Describe how voltage (*V*), current (*I*), and resistance (*R*) are related**
  - **Express *I* as a function of *V* and *R***
  - **Express *V* as a function of *I* and *R***
  - **Express *R* as a function of *V* and *I***

Ohm determined experimentally that if the voltage across a resistor is increased, the current through the resistor will increase; and, likewise, if the voltage is decreased, the current will decrease. For example, if the voltage is doubled, the current will double. If the voltage is halved, the current will also be halved. This relationship is illustrated in Figure 3–1, with relative meter indications of voltage and current.

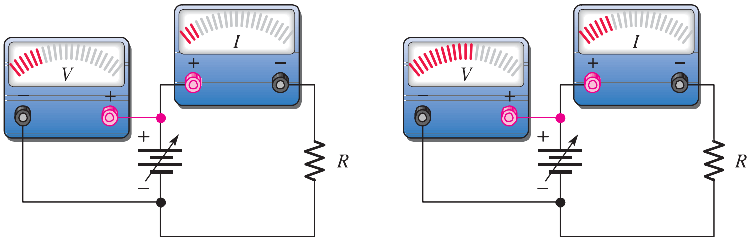

(a) Less *V*, less *I*   (b) More *V*, more *I*

FIGURE 3–1 **Effect on the current of changing the voltage with the resistance at a constant value.**

Ohm also determined that if the voltage is held constant, less resistance results in more current, and more resistance results in less current. For example, if the resistance is halved, the current doubles. If the resistance is doubled, the current is halved. This concept is illustrated by the meter indications in Figure 3–2, where the resistance is increased and the voltage is held constant.

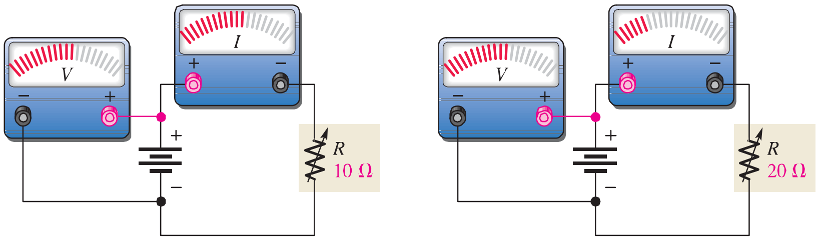

(a) Less *R*, more *I*   (b) More *R*, less *I*

FIGURE 3–2 **Effect on the current of changing the resistance with the voltage at a constant value.**

Ohm's law states that current is directly proportional to voltage and inversely proportional to resistance.

$$I = \frac{V}{R} \qquad (3\text{–}1)$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

where I is current in amperes (A), V is voltage in volts (V), and R is resistance in ohms (Ω). This formula describes the relationship illustrated by the action in the circuits of Figures 3–1 and 3–2.

For a constant resistance, if the voltage applied to a circuit is increased, the current will increase; and if the voltage is decreased, the current will decrease.

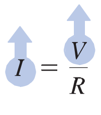

Increase V, I increases

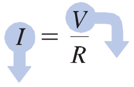

Decrease V, I decreases

R constant

For a constant voltage, if the resistance in a circuit is increased, the current will decrease; and if the resistance is decreased, the current will increase.

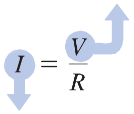

Increase R, I decreases

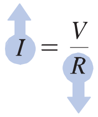

Decrease R, I increases

V constant

EXAMPLE 3–1

Using the Ohm's law formula in Equation 3–1, verify that the current through a 10 Ω resistor increases when the voltage is increased from 5 V to 20 V.

SOLUTION

For $V = 5\text{ V}$,

$$I = \frac{V}{R} = \frac{5\text{ V}}{10\ \Omega} = \mathbf{0.5\ A}$$

For $V = 20\text{ V}$,

$$I = \frac{V}{R} = \frac{20\text{ V}}{10\ \Omega} = \mathbf{2\ A}$$

RELATED PROBLEM*

Show that the current decreases when the resistance is increased from 5 Ω to 20 Ω and the voltage is a constant 10 V.

***Answers are at the end of the chapter.**

Ohm's law can also be stated another equivalent way. By multiplying both sides of Equation 3–1 by R and transposing terms, you obtain an equivalent form of Ohm's law, as follows:

$$\boxed{V = IR} \qquad (3\text{–}2)$$

With this formula, you can calculate voltage in volts if you know the current in amperes and resistance in ohms.

EXAMPLE 3–2

Use the Ohm's law formula in Equation 3–2 to calculate the voltage across a 1.0 kΩ resistor when the current is 5.0 mA.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

$$V = IR = (5.0\text{ mA})(1.0\text{ k}\Omega) = \mathbf{5.0\ V}$$

RELATED PROBLEM

6238262 2013/12/10 12.27.66.5

Find the voltage across a 1.0 kΩ resistor when the current is 1 mA.

There is a third equivalent way to state Ohm's law. By dividing both sides of Equation 3–2 by I and transposing terms, you obtain the following formula:

$$R = \frac{V}{I} \qquad (3\text{–}3)$$

This expression of the Ohm's law formula is used to determine resistance in ohms if you know the values of voltage in volts and current in amperes.

Remember, the three expressions—Equations 3–1, 3–2, and 3–3—are all equivalent. They are simply three ways of stating Ohm's law.

EXAMPLE 3–3

Use the Ohm's law formula in Equation 3–3 to calculate the resistance of a rear window defroster grid in a certain vehicle. When it is connected to 12.6 V, it draws 15.0 A from the battery. What is the resistance of the defroster grid?

SOLUTION

$$R = \frac{V}{I} = \frac{12.6\text{ V}}{15.0\text{ A}} = \mathbf{840\ m\Omega}$$

RELATED PROBLEM

If one of the grid wires opens, the current drops to 13.0 A. What is the new resistance?

The Linear Relationship of Current and Voltage

In resistive circuits, current and voltage are linearly proportional. **Linear** means that if one is increased or decreased by a certain percentage, the other will increase or decrease by the same percentage, assuming that the resistance is constant in value. For example, if the voltage across a resistor is tripled, the current will triple. If the voltage is reduced by half, the current will decrease by half.

EXAMPLE 3–4

Show that if the voltage in the circuit of Figure 3–3 is increased to three times its present value, the current will triple in value.

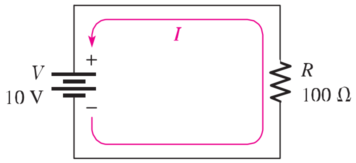

FIGURE 3–3

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

With 10 V, the current is

$$I = \frac{V}{R} = \frac{10\text{ V}}{100\ \Omega} = 0.1\text{ A}$$

If the voltage is increased to 30 V, the current will be

$$I = \frac{V}{R} = \frac{30\text{ V}}{100\ \Omega} = 0.3\text{ A}$$

The current went from 0.1 A to 0.3 A when the voltage was tripled to 30 V.

RELATED PROBLEM

If the voltage in Figure 3–3 is quadrupled, will the current also quadruple?

Let's take a constant value of resistance, for example, 10 Ω, and calculate the current for several values of voltage ranging from 10 V to 100 V in the circuit in Figure 3–4(a). The current values obtained are shown in Figure 3–4(b). The graph of the I values versus the V values is shown in Figure 3–4(c). Note that it is a straight line graph. This graph shows that a change in voltage results in a linearly proportional change in current. No matter what value R is, assuming that R is constant, the graph of I versus V will always be a straight line.

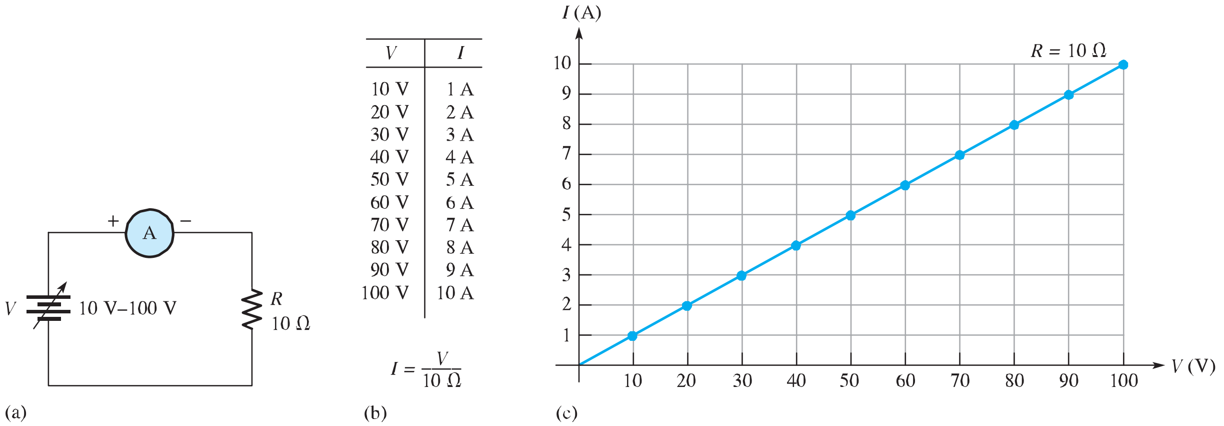

V	I
10 V	1 A
20 V	2 A
30 V	3 A
40 V	4 A
50 V	5 A
60 V	6 A
70 V	7 A
80 V	8 A
90 V	9 A
100 V	10 A

$I = \frac{V}{10\ \Omega}$

FIGURE 3–4 Graph of current versus voltage for the circuit in part (a).

A Graphic Aid for Ohm's Law

You may find the graphic aid in Figure 3–5 helpful for applying Ohm's law. It is a way to remember the formulas.

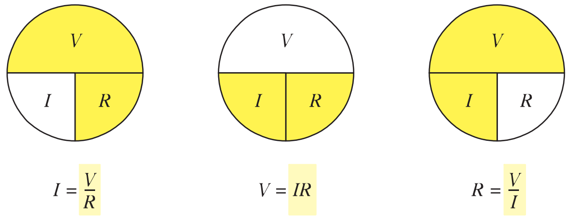

FIGURE 3–5 A graphic aid for the Ohm's law formulas.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 3–1 CHECKUP*

1. Briefly state Ohm's law in words.
2. Write the Ohm's law formula for calculating current.
3. Write the Ohm's law formula for calculating voltage.
4. Write the Ohm's law formula for calculating resistance.
5. If the voltage across a resistor is tripled, does the current increase or decrease? By how much?
6. There is a fixed voltage across a variable resistor, and you measure a current of 10 mA. If you double the resistance, how much current will you measure?
7. What happens to the current in a linear circuit where both the voltage and the resistance are doubled?

***Answers are at the end of the chapter.**

3–2 APPLICATION OF OHM'S LAW

This section provides examples of the application of Ohm's law for calculating voltage, current, and resistance in electric circuits. You will also see how to use quantities expressed with metric prefixes in circuit calculations.

After completing this section, you should be able to

- **Use Ohm's law to determine voltage, current, or resistance**
  - **Use Ohm's law to find current when you know voltage and resistance**
  - **Use Ohm's law to find voltage when you know current and resistance**
  - **Use Ohm's law to find resistance when you know voltage and current**
  - **Use quantities with metric prefixes**

Current Calculations

In these examples you will learn to determine current values when you know the values of voltage and resistance. In these problems, the formula $I = V/R$ is used. In order to get current in amperes, you must express the value of V in volts and the value of R in ohms.

EXAMPLE 3–5

An indicator light requires a 330 Ω resistor to limit current. The voltage across the current limiting resistor is 3 V. What is the current in the resistor?

SOLUTION

$$I = \frac{V}{R} = \frac{3.0\text{ V}}{330\ \Omega} = \mathbf{9.09\ mA}$$

RELATED PROBLEM

How will the current change if a 270 Ω resistor is used instead and 3.0 V still is across the resistor?

In electronics, resistance values of thousands or millions of ohms are common. The metric prefixes *kilo* (k) and *mega* (M) are used to indicate large values. Thus, thousands of ohms are expressed in kilohms (kΩ), and millions of ohms are expressed in megohms (MΩ). The following examples illustrate how to use kilohms and megohms when you use Ohm's law to calculate current.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 3–6

Calculate the current in milliamperes for the circuit of Figure 3–6.

FIGURE 3–6

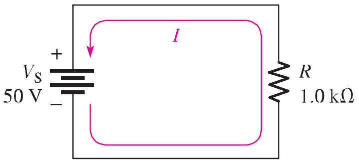

SOLUTION

Remember that 1.0 kΩ is the same as $1.0 \times 10^3\ \Omega$. Use the formula $I = V/R$ and substitute 50 V for V and $1.0 \times 10^3\ \Omega$ for R.

$$I = \frac{V_S}{R} = \frac{50\text{ V}}{1.0\text{ k}\Omega} = \frac{50\text{ V}}{1.0 \times 10^3\ \Omega} = 50 \times 10^{-3}\text{ A} = \mathbf{50\ mA}$$

RELATED PROBLEM

If the resistance in Figure 3–6 is increased to 10 kΩ, what is the current?

6238262 2013/12/10 12.27.66.5

MULTISIM

Open Multisim file E03-06; files are found at www.pearsonhighered.com/floyd. Connect the multimeter to the circuit and verify the value of current calculated in this example.

In Example 3–6, the current is expressed as 50 mA. Thus, *when volts (V) are divided by kilohms (kΩ), the current is in milliamperes (mA).*

When volts (V) are divided by megohms (MΩ), the current is in microamperes (μA), as Example 3–7 illustrates.

EXAMPLE 3–7

Determine the amount of current in microamperes for the circuit of Figure 3–7.

FIGURE 3–7

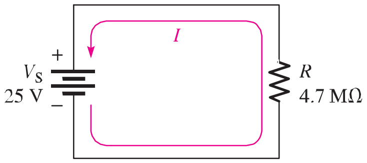

SOLUTION

Recall that 4.7 MΩ equals $4.7 \times 10^6\ \Omega$. Use the formula $I = V/R$ and substitute 25 V for V and $4.7 \times 10^6\ \Omega$ for R.

$$I = \frac{V_S}{R} = \frac{25\text{ V}}{4.7\text{ M}\Omega} = \frac{25\text{ V}}{4.7 \times 10^6\ \Omega} = 5.32 \times 10^{-6}\text{ A} = \mathbf{5.32\ \mu A}$$

RELATED PROBLEM

If the resistance in Figure 3–7 is decreased to 1.0 MΩ, what is the current?

MULTISIM

Open Multisim file E03-07. Connect the multimeter to the circuit and verify the value of current calculated in this example.

Small voltages, usually less than 50 V, are common in electronic circuits. Occasionally, however, large voltages are encountered. For example, the high-voltage supply in radio transmitters, plasma guns, ion motors, and x-ray machines are commonly well above 1000 V. Transmission voltages generated by the power companies may be as high as 345,000 V (345 kV).

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 3–8

How much current in microamperes is there through a 100 MΩ resistor when 50 kV are applied across it?

SOLUTION

Divide 50 kV by 100 MΩ to get the current. Substitute 50×10^3 V for 50 kV and $100 \times 10^6\ \Omega$ for 100 MΩ in the formula for current. V_R is the voltage across the resistor.

$$I = \frac{V_R}{R} = \frac{50\text{ kV}}{100\text{ M}\Omega} = \frac{50 \times 10^3\text{ V}}{100 \times 10^6\ \Omega} = 0.5 \times 10^{-3}\text{ A}$$

$$= 500 \times 10^{-6} = \mathbf{500\ \mu A}$$

RELATED PROBLEM

How much current is there through 10 MΩ when 2 kV are applied?

Voltage Calculations

In these examples you will learn how to determine voltage values when you know the current and resistance using the formula $V = IR$. To obtain voltage in volts, you must express the value of I in amperes and the value of R in ohms.

EXAMPLE 3–9

In the circuit of Figure 3–8, how much voltage is needed to produce 5 A of current?

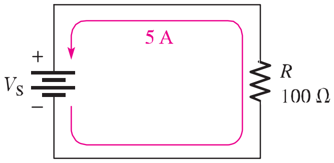

FIGURE 3–8

SOLUTION

Substitute 5 A for I and 100 Ω for R into the formula $V = IR$.

$$V_S = IR = (5\text{ A})(100\ \Omega) = \mathbf{500\ V}$$

Thus, 500 V are required to produce 5 A of current through a 100 Ω resistor.

RELATED PROBLEM

How much voltage is required to produce 8 A in the circuit of Figure 3–8?

EXAMPLE 3–10

How much voltage will be measured across the resistor in Figure 3–9?

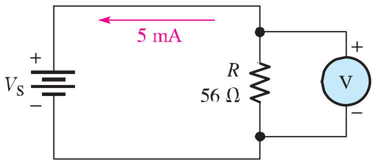

FIGURE 3–9

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

Note that 5 mA equals 5×10^{-3} A. Substitute the values for I and R into the formula $V = IR$.

$$V_R = IR = (5\text{ mA})(56\ \Omega) = (5 \times 10^{-3}\text{ A})(56\ \Omega) = \mathbf{280\ mV}$$

When milliamperes are multiplied by ohms, the result is millivolts.

RELATED PROBLEM

Change the resistor in Figure 3–9 to 22 Ω and determine the voltage required to produce 10 mA.

EXAMPLE 3–11

The circuit in Figure 3–10 has a current of 10 mA. What is the source voltage?

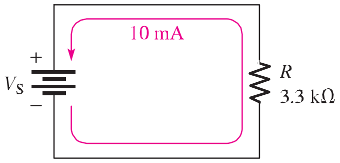

FIGURE 3–10

SOLUTION

Note that 10 mA equals 10×10^{-3} A and that 3.3 kΩ equals $3.3 \times 10^{3}\ \Omega$. Substitute these values into the formula $V = IR$.

$$V_S = IR = (10\text{ mA})(3.3\text{ k}\Omega) = (10 \times 10^{-3}\text{ A})(3.3 \times 10^{3}\ \Omega) = \mathbf{33\ V}$$

When milliamperes and kilohms are multiplied, the result is volts.

RELATED PROBLEM

What is the voltage in Figure 3–10 if the current is 5 mA?

EXAMPLE 3–12

A small solar cell is connected to a 27 kΩ resistor. In bright sunlight, the solar cell looks like a current source that can supply 180 μA to the resistor, as shown in Figure 3–11. What is the voltage across the resistor?

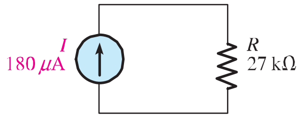

FIGURE 3–11

SOLUTION

$$V_R = IR = (180\ \mu\text{A})(27\text{ k}\Omega) = \mathbf{4.86\ V}$$

RELATED PROBLEM

How does the voltage change when in cloudy conditions the current drops to 40 μA?

MULTISIM

Open Multisim file E03-12. Connect a voltmeter across the resistor and confirm the voltage calculated in this example.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Resistance Calculations

In these examples you will learn how to determine resistance values when you know the voltage and current using the formula $R = V/I$. To find resistance in ohms, you must express the value of V in volts and the value of I in amperes.

EXAMPLE 3–13

An automotive lamp draws 2 A from the 13.2 V battery. What is the resistance of the bulb?

SOLUTION

$$R = \frac{V}{I} = \frac{13.2\text{ V}}{2.0\text{ A}} = \mathbf{6.6\ \Omega}$$

RELATED PROBLEM

When operated at 6.6 V, the same bulb has a current of 1.1 A. What is the resistance of the bulb in this case?

EXAMPLE 3–14

A cadmium sulfide cell (CdS cell) is a photosensitive resistor that changes resistance when light strikes it. It is used in applications such as turning on lights at dusk. You could monitor the resistance indirectly from an ammeter when the cell is in an active circuit with the setup shown in Figure 3–12. What resistance is implied by the readings shown?

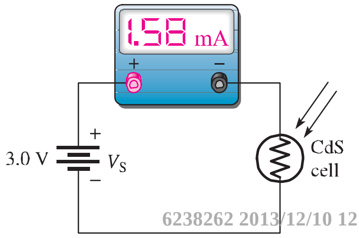

6238262 2013/12/10 12.27 66.5

FIGURE 3–12

SOLUTION

$$R = \frac{V}{I} = \frac{3.0\text{ V}}{1.58\text{ mA}} = \mathbf{1.90\ k\Omega}$$

RELATED PROBLEM

In the dark, the current drops to 76 μA. What resistance does this imply?

SYSTEM EXAMPLE 3–1

CURRENT-SENSING RESISTORS

Current-sensing resistors are precision low-value resistors that are widely used in various systems, including power supplies, battery-charging circuits, motor controllers, automotive systems, telecommunications, and computers. As stated by Ohm's law, there is a voltage drop across the resistor when current is in it. The current-sensing resistor converts this current into a small voltage that can be easily measured and monitored. The majority of applications use precision surface-mount resistors such as the one shown in Figure 3–13.

Current-sensing resistors have low-resistance values in order to minimize power loss and self-heating effects. Low values also ensure there is a minimum effect on the circuit due to insertion loss. Resistance values of 20–25 mΩ are common.

One application for current-sensing resistors is for a battery charger, which is a type of power supply. Figure 3–14 shows a simplified drawing of a type of battery charger that

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

is designed to regulate current. The current-sensing resistor is shown in red and is the focus of this example. The current from the supply is converted to a small voltage that is monitored by the controller, which is a small integrated circuit. The controller can adjust the amount of current based on an optimum charging profile. The controller enables the charger to be highly efficient and reliable.

FIGURE 3–13 **Surface mount current-sensing resistor.**

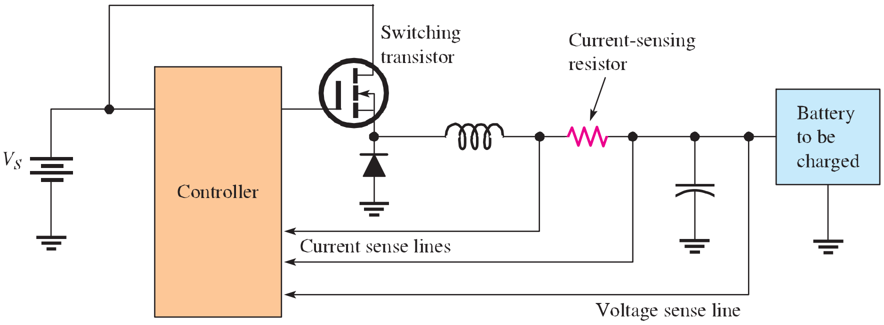

FIGURE 3–14 **A power supply system for charging batteries that uses a current-sensing resistor to monitor the charging current.**

SECTION 3–2 CHECKUP

1. $V = 10$ V and $R = 4.7\ \Omega$. Find I.
2. If a 4.7 MΩ resistor has 20 kV across it, how much current is there?
3. How much current will 10 kV across a 2 kΩ resistance produce?
4. $I = 1$ A and $R = 10\ \Omega$. Find V.
5. What voltage do you need to produce 3 mA of current in a 3 kΩ resistance?
6. A battery produces 2 A of current through a 6 Ω resistive load. What is the battery voltage?
7. $V = 10$ V and $I = 2$ A. Find R.
8. In a stereo amplifier circuit there is a resistor across which you measure 25 V, and your ammeter indicates 50 mA of current through the resistor. What is the resistor's value in kilohms? In ohms?

3–3 ENERGY AND POWER

When there is current through a resistance, electrical energy is converted to heat or other form of energy, such as light. A common example of this is an incandescent light bulb that becomes too hot to touch. The current through the filament that produces light also produces unwanted heat because the filament has resistance. Electrical components must be able to dissipate a certain amount of energy in a given period of time.

After completing this section, you should be able to

- **Define *energy* and *power***
  - **Express power in terms of energy and time**
  - **State the unit of power**
  - **State the common units of energy**
  - **Perform energy and power calculations**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Energy is the ability to do work, and power is the rate at which energy is used.

In other words, power, P, is a certain amount of energy, W, used in a certain length of time (t), expressed as follows:

$$\boldsymbol{P = \frac{W}{t}} \qquad (3\text{–}4)$$

where P is power in watts (W), W is energy in joules (J), and t is time in seconds (s). Note that an italic W is used to represent energy in the form of work and a nonitalic W is used for watts, the unit of power. The **joule** is the SI unit for energy.

Energy in joules divided by time in seconds gives power in watts. For example, if 50 J of energy are used in 2 s, the power is $50\text{ J}/2\text{ s} = 25\text{ W}$. By definition,

One watt is the amount of power when one joule of energy is used in one second.

Thus, the number of joules used in 1 s is always equal to the number of watts. For example, if 75 J are used in 1 s, the power is

$$P = \frac{W}{t} = \frac{75\text{ J}}{1\text{ s}} = 75\text{ W}$$

Amounts of power much less than one watt are common in certain areas of electronics. As with small current and voltage values, metric prefixes are used to designate small amounts of power. Thus, milliwatts (mW) and microwatts (μW) are commonly found in some applications.

In the electrical utilities field, kilowatts (kW) and megawatts (MW) are common units. Radio and television stations also use large amounts of power to transmit signals. Electric motors are commonly rated in horsepower (hp) where 1 hp = 746 W.

Since power is the rate at which energy is used, power utilized over a period of time represents energy consumption. If you multiply power in watts and time in seconds, you have energy in joules, symbolized by W.

$$W = Pt$$

EXAMPLE 3–15

An amount of energy equal to 100 J is used in 5 s. What is the power in watts?

SOLUTION

$$P = \frac{\text{energy}}{\text{time}} = \frac{W}{t} = \frac{100\text{ J}}{5\text{ s}} = \mathbf{20\text{ W}}$$

RELATED PROBLEM

If 100 W of power occurs for 30 s, how much energy in joules is used?

The Kilowatt-hour (kWh) Unit of Energy

The joule has been defined as the unit of energy. However, there is another way to express energy. Since power is expressed in watts and time can be expressed in hours, a unit of energy called the kilowatt-hour (kWh) can be used.

When you pay your electric bill, you are charged on the basis of the amount of energy you use. Because power companies deal in huge amounts of energy, the most practical unit is the kilowatt-hour. You use a **kilowatt-hour** of energy when you use the equivalent of 1000 W of power for 1 h. For example, a 100 W for television that is on 10 h uses 1 kWh of energy.

$$W = Pt = (100\text{ W})(10\text{ h}) = 1000\text{ Wh} = 1\text{ kWh}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 3–16

Determine the number of kilowatt-hours (kWh) for each of the following energy consumptions:

(a) 1400 W for 1 hr **(b)** 2500 W for 2 h **(c)** 100,000 W for 5 h

SOLUTION

(a) 1400 W $=$ 1.4 kW

$W = Pt = (1.4\text{ kW})(1\text{ h}) = \mathbf{1.4\text{ kWh}}$

(b) 2500 W $=$ 2.5 kW

$\text{Energy} = (2.5\text{ kW})(2\text{ h}) = \mathbf{5\text{ kWh}}$

(c) 100,000 W $=$ 100 kW

$\text{Energy} = (100\text{ kW})(5\text{ h}) = \mathbf{500\text{ kWh}}$

RELATED PROBLEM

How many kilowatt-hours of energy are used by a 250 W light bulb burning for 8 h?

Table 3–1 lists the typical power rating in watts for several household appliances. You can determine the maximum kWh for various appliances by using the power rating in Table 3–1 converted to kilowatts times the number of hours it is used.

6238262 2013/12/10 12.27.66.5

TABLE 3–1

APPLIANCE	POWER RATING (WATTS)
Air conditioner	860
Blow dryer	1000
Clock	2
Clothes dryer	4000
Dishwasher	1200
Heater	1322
Microwave oven	800
Range	12,200
Refrigerator	500
Television	250
Washing machine	400
Water heater	2500

EXAMPLE 3–17

During a typical 24-hour period, you use the following appliances for the specified lengths of time:

air conditioner: 15 hours
blow dryer: 10 minutes
clock: 24 hours

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

clothes dryer: 1 hour
dishwasher: 45 minutes
microwave oven: 15 minutes
refrigerator: 12 hours
television: 2 hours
water heater: 8 hours

Determine the total kilowatt-hours and the electric bill for the time period. The rate is 11 cents per kilowatt-hour.

SOLUTION

Determine the kWh for each appliance used by converting the watts in Table 3–1 to kilowatts and multiplying by the time in hours:

air conditioner: $0.860\text{ kW} \times 15\text{ h} = 12.9\text{ kWh}$
blow dryer: $1.0\text{ kW} \times 0.167\text{ h} = 0.167\text{ kWh}$
clock: $0.002\text{ kW} \times 24\text{ h} = 0.048\text{ kWh}$
clothes dryer: $4.0\text{ kW} \times 1\text{ h} = 4.0\text{ kWh}$
dishwasher: $1.2\text{ kW} \times 0.75\text{ h} = 0.9\text{ kWh}$
microwave: $0.8\text{ kW} \times 0.25\text{ h} = 0.2\text{ kWh}$
refrigerator: $0.5\text{ kW} \times 12\text{ h} = 6\text{ kWh}$
television: $0.25\text{ kW} \times 2\text{ h} = 0.5\text{ kWh}$
water heater: $2.5\text{ kW} \times 8\text{ h} = 20\text{ kWh}$

Now, add up all the kilowatt-hours to get the total energy for the 24-hour period.

$$\text{Total energy} = (12.9 + 0.167 + 0.048 + 4.0 + 0.9 + 0.2 + 6.0 + 0.5 + 20)\text{ kWh} = \mathbf{44.7\text{ kWh}}$$

At 11 cents/kilowatt-hour, the cost of energy to run the appliances for the 24-hour period is

$$\text{Energy cost} = 44.7\text{ kWh} \times 0.11\ \$/\text{kWh} = \mathbf{\$4.92}$$

RELATED PROBLEM

In addition to the appliances, suppose you used one 200 W humidifier for 2 hours and one 75 W heating pad for 3 hours. Calculate your cost for the 24-hour period for everything.

SECTION 3–3 CHECKUP

1. Define *power*.
2. Write the formula for power in terms of energy and time.
3. Define *watt*.
4. Express each of the following values of power in the most appropriate units:
   (a) 68,000 W **(b)** 0.005 W **(c)** 0.000025 W
5. If you use 100 W of power for 10 h, how much energy (in kilowatt-hours) have you used?
6. Convert 2000 W to kilowatts.
7. How much does it cost to run a heater (1322 W) for 24 hours if energy cost is 11¢ per kilowatt-hour?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

3–4 POWER IN AN ELECTRIC CIRCUIT

The generation of heat, which occurs when electrical energy is converted to heat energy, in an electric circuit is often an unwanted by-product of current through the resistance in the circuit. In some cases, however, the generation of heat is the primary purpose of a circuit as, for example, in an electric resistive heater. In any case, you must frequently deal with power in electrical and electronic circuits.

After completing this section, you should be able to

- **Calculate power in a circuit**
  - **Determine power when you know *I* and *R***
  - **Determine power when you know *V* and *I***
  - **Determine power when you know *V* and *R***

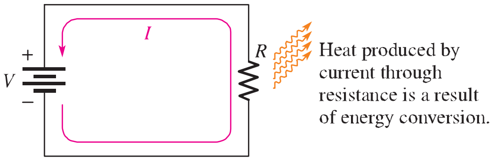
I
R
+
V
−
Heat produced by current through resistance is a result of energy conversion.

FIGURE 3–15 **Power dissipation in an electric circuit is seen as heat given off by the resistance. The power dissipation is equal to the power supplied by the voltage source.**

When there is current through a resistance, the collisions of the electrons as they move through the resistance give off heat, resulting in a conversion of electrical energy to thermal energy as indicated in Figure 3–15. The amount of power dissipated in an electrical circuit is dependent on the amount of resistance and on the amount of current, expressed as follows:

$$P = I^2R \tag{3–5}$$

where P is power in watts (W), V is voltage in volts (V), and I is current in amperes (A). You can get an equivalent expression for power in terms of voltage and current by substituting $I \times I$ for I^2 and V for IR.

$$P = I^2R = (I \times I)R = I(IR) = (IR)I$$

$$P = VI \tag{3–6}$$

You can obtain another equivalent expression by substituting V/R for I (Ohm's law).

$$P = VI = V\left(\frac{V}{R}\right)$$

$$P = \frac{V^2}{R} \tag{3–7}$$

The three power expressions in Equations 3–5, 3–6, and 3–7 are known as **Watt's law**. To calculate the power in a resistance, you can use any one of the three equivalent Watt's law power formulas, depending on what information you have. For example, assume that you know the values of current and voltage; in this case, calculate the power with the formula $P = VI$. If you know I and R, use the formula $P = I^2R$. If you know V and R, use the formula $P = V^2/R$.

An aid for using both Ohm's law and Watt's law is found in the chapter summary, Figure 3–27.

EXAMPLE 3–18

Calculate the power in each of the three circuits of Figure 3–16.

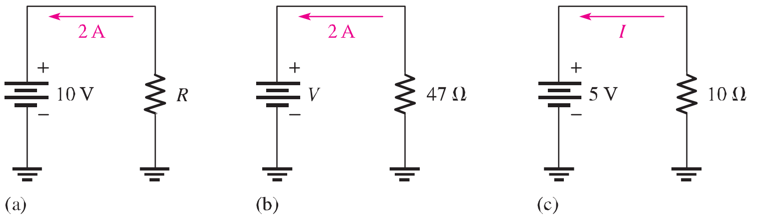

FIGURE 3–16

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

In circuit (a), you know V and I. The power is determined as follows:

$$P = VI = (10\ \text{V})(2\ \text{A}) = \mathbf{20\ W}$$

In circuit (b), you know I and R. Therefore,

$$P = I^2R = (2\ \text{A})^2(47\ \Omega) = \mathbf{188\ W}$$

In circuit (c), you know V and R. Therefore,

$$P = \frac{V^2}{R} = \frac{(5\ \text{V})^2}{10\ \Omega} = \mathbf{2.5\ W}$$

RELATED PROBLEM

Determine the power in each circuit of Figure 3–16 for the following changes: In circuit (a), I is doubled and V remains the same; in circuit (b), R is doubled and I remains the same; in circuit (c), V is halved and R remains the same.

EXAMPLE 3–19

A solar yard light such as the one shown in Figure 3–17 has a solar collector that can provide 1.0 W of power for charging the 3.0 V batteries. What is the maximum charging current the solar collector could supply to fully discharged 3.0 V batteries?

623826 2013/12/10 12.27.66.5

FIGURE 3–17

SOLUTION

$$I = \frac{P}{V} = \frac{1.0\ \text{W}}{3.0\ \text{V}} = \mathbf{0.33\ A}$$

RELATED PROBLEM

What is the power dissipated by the light at night if the current is 30 mA?

SECTION 3–4 CHECKUP

1. Assume a car window defroster is connected to 13.0 V and has a current of 12 A. What power is dissipated in the defroster?
2. If there is a current of 5 A through a 47 Ω resistor, what is the power dissipated?
3. Many oscilloscopes have a 50 Ω input position that places a 2 W, 50 Ω resistor between the input and ground. What is the maximum voltage that could be applied to the input before exceeding the power rating of this resistor?
4. Assume a car seat heater has an internal resistance of 3.0 Ω. If the battery voltage is 13.4 V, what power is dissipated by the heater when it is on?
5. How much power does a 2.2 kΩ resistor with 8 V across it produce?
6. What is the resistance of a 55 W bulb that draws 0.5 A?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

3–5 THE POWER RATING OF RESISTORS

As you know, a resistor gives off heat when there is current through it. There is a limit to the amount of heat that a resistor can give off, which is specified by its power rating.

After completing this section, you should be able to

- **Properly select resistors based on power consideration**
  - **Define *power rating***
  - **Explain how physical characteristics of resistors determine their power rating**
  - **Check for resistor failure with an ohmmeter**

SAFETY NOTE

Some resistors can become very hot in normal operation. To avoid a burn, do not touch a circuit component while the power is connected to the circuit. After power has been turned off, allow time for the components to cool down.

The **power rating** is the maximum amount of power that a resistor can dissipate without being damaged by excessive heat buildup. The power rating is not related to the ohmic value (resistance) but rather is determined mainly by the physical composition, size, and shape of the resistor. All else being equal, the larger the surface area of a resistor, the more power it can dissipate. *The surface area of a cylindrically shaped resistor is equal to the length (l) times the circumference (c),* as indicated in Figure 3–18. The area of the ends is not included.

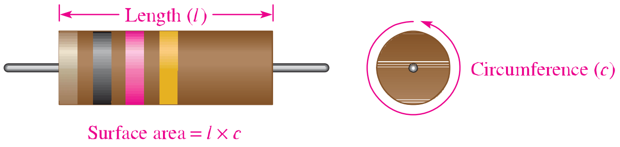

FIGURE 3–18 **The power rating of a resistor is directly related to its surface area.**

Metal-film resistors are available in standard power ratings from ⅛ W to 1 W, as shown in Figure 3–19. Available power ratings for other types of resistors vary. For example, wirewound resistors have ratings up to 225 W or greater. Figure 3–20 shows some of these resistors.

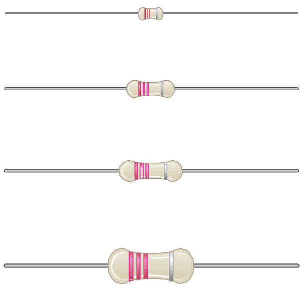

FIGURE 3–19 **Relative sizes of metal-film resistors with standard power ratings of ⅛ W, ¼ W, ½ W, and 1 W.**

Braking Resistors

Power resistors that are rated for several thousand watts are used as braking resistors for motors that need to be stopped completely. When a motor is turning, it has kinetic energy. For example, there is a large amount of energy in a moving elevator that must be dissipated when the elevator is stopped at a floor. When the electrical input is removed from an elevator motor, it would try to continue spinning due to inertia. To stop the motor quickly and bring it to a smooth stop, a special high-power braking resistoris switched across the motor. The braking resistor dissipates the energy from the moving system as heat. The braking resistor is normally mounted to a heat sink to help remove the heat.

Photo courtesy of Glover Resistors, Inc.

SYSTEM NOTE

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

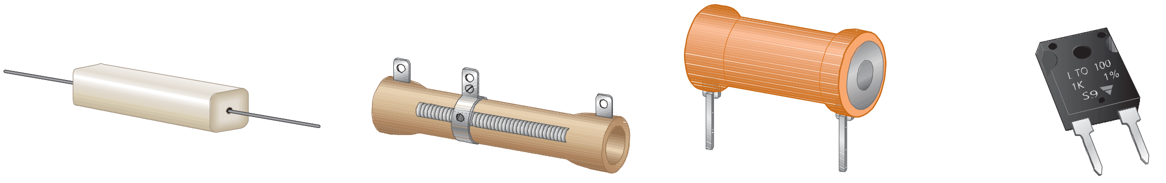

(a) Axial-lead wirewound (b) Adjustable wirewound (c) Radial-lead for PC board insertion (d) Thick-film power

FIGURE 3–20 **Typical resistors with high power ratings.**

When a resistor is used in a circuit, its power rating should be greater than the maximum power that it will have to handle to create a safety margin. Generally, the next higher standard value is used. For example, if a metal-film resistor is to dissipate 0.75 W in a circuit application, its rating should be the next higher standard value which is 1 W.

SYSTEM EXAMPLE 3–2

A LOAD TEST BOX

Most manufacturers run some quality control on a sample of their products to assure each tested product meets specifications before shipping it to a customer. In many cases, a specialized test fixture is designed to perform the quality test. Assume a manufacturer needs to test four different supplies at their maximum rated output, which are as follows: 5.0 V @ 1.0 A, 8.0 V at 500 mA, 10 V @ 100 mA, and 15 V at 100 mA. A set of selectable resistors is needed that will allow a technician to check that the power supply can operate properly under load. For this example, a resistor test box is designed that will meet the requirements, using resistors selected from stock.

Step 1 of the design process is to determine the resistance values that will provide the required loads and the power dissipated by the resistor when the power supply is connected to it. From Ohm's law, the required resistances for each supply are

5.0 V supply: $R = V/I = 5.0\text{ V}/1.0\text{ A} = 5.0\ \Omega$
8.0 V supply: $R = V/I = 8.0\text{ V}/0.50\text{ A} = 16\ \Omega$
10 V supply: $R = V/I = 10\text{ V}/0.10\text{ A} = 100\ \Omega$
15 V supply: $R = V/I = 15\text{ V}/0.10\text{ A} = 150\ \Omega$

Step 2 is to determine the amount of power expected in each resistor. From Watt's law, the expected power dissipation for each power supply is

5.0 V supply: $P = VI = (5.0\text{ V})(1.0\text{ A}) = 5.0\text{ W}$
8.0 V supply: $P = VI = (8.0\text{ V})(0.50\text{ A}) = 4.0\text{ W}$
10 V supply: $P = VI = (10\text{ V})(0.10\text{ A}) = 1.0\text{ W}$
15 V supply: $P = VI = (15\text{ V})(0.10\text{ A}) = 1.5\text{ W}$

To be safe, specify resistors that have at least a 20% margin above their rated power in case a power supply has more voltage than expected. The power rating of the resistors depends on the type of resistor and the power tolerance required. Assume the choices for power ratings are 1.0 W, 2.0 W, 5.0 W, or 10 W. The selected resistors are then

5.0 V supply: $R = 5.0\ \Omega$; $P = 5.0\text{ W}$, choose $P_{(rating)} = 10\text{ W}$
8.0 V supply: $R = 16\ \Omega$; $P = 4.0\text{ W}$, choose $P_{(rating)} = 5.0\text{ W}$
10 V supply: $R = 100\ \Omega$; $P = 1.0\text{ W}$, choose $P_{(rating)} = 2.0\text{ W}$
15 V supply: $R = 150\ \Omega$; $P = 1.5\text{ W}$, choose $P_{(rating)} = 2.0\text{ W}$

The final step is to specify the fuse and list the parts. The fuse needs to be large enough to accept a more than the maximum expected current (1A). A reasonable fuse is 1.5 A. The final design is shown in Figure 3–21. To complete the box, a parts list is prepared including the pc board, components, and hardware items (box, screws, knob, etc.).

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

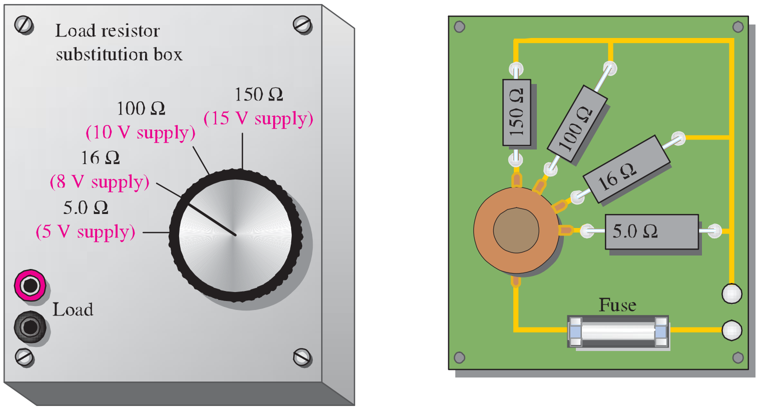

FIGURE 3–21 A load test box showing top view and the pc board.

When the power dissipated in a resistor is greater than its rating, the resistor will become excessively hot. As a result, the resistor may burn open or its resistance value may be greatly altered.

A resistor that has been damaged because of overheating can often be detected by the charred or altered appearance of its surface. If there is no visual evidence, a resistor that is suspected of being damaged can be checked with an ohmmeter for an open or increased resistance value. Recall that a resistor should be disconnected from the circuit to measure resistance. Sometimes an overheated resistor is due to another failure in the circuit. After replacing an overheated resistor, the underlying cause should be investigated before restoring power to the circuit.

EXAMPLE 3–20

Determine whether the resistor in each circuit of Figure 3–22 has possibly been damaged by overheating.

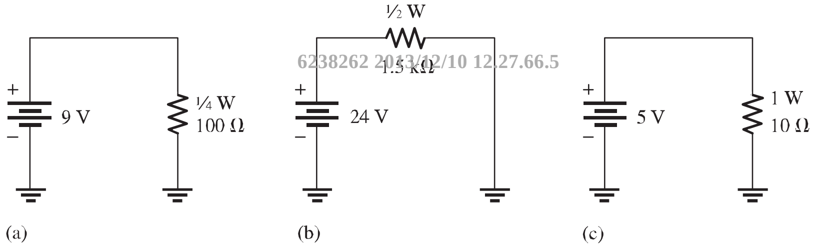

6238262 2013/12/10 12.27.66.5

FIGURE 3–22

SOLUTION

For the circuit in Figure 3–22(a),

$$P = \frac{V^2}{R} = \frac{(9\text{ V})^2}{100\ \Omega} = 0.81\text{ W}$$

The rating of the resistor is ¼ W (0.25 W), which is insufficient to handle the power. The resistor has been overheated and may be burned out, making it an open.

For the circuit in Figure 3–22(b),

$$P = \frac{V^2}{R} = \frac{(24\text{ V})^2}{1.5\text{ k}\Omega} = 0.384\text{ W}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The rating of the resistor is ½ W (0.5 W), which is sufficient to handle the power.

For the circuit in Figure 3–22(c),

$$P = \frac{V^2}{R} = \frac{(5\text{ V})^2}{10\ \Omega} = 2.5\text{ W}$$

The rating of the resistor is 1 W, which is insufficient to handle the power. The resistor has been overheated and may be burned out, making it an open.

RELATED PROBLEM

A ¼ W, 1.0 kΩ resistor is connected across a 12 V battery. Will it overheat?

SECTION 3–5 CHECKUP

1. Name two important parameters associated with a resistor.
2. How does the physical size of a resistor determine the amount of power that it can handle?
3. List the standard power ratings of metal-film resistors.
4. A resistor must handle 0.3 W. What standard size metal-film resistor should be used to dissipate the energy properly?
5. What is the maximum voltage that can be applied to a ¼ W, 100 Ω resistor if the power rating is not to be exceeded?

3–6 ENERGY CONVERSION AND VOLTAGE DROP IN A RESISTANCE

As you have learned, when there is current through a resistance, electrical energy is converted to heat energy. This heat is caused by collisions of the free electrons within the atomic structure of the resistive material. When a collision occurs, heat is given off; and the electron gives up some of its acquired energy as it moves through the material.

After completing this section, you should be able to

- **Explain energy conversion and voltage drop**
  - **Discuss the cause of energy conversion in a circuit**
  - **Define *voltage drop***
  - **Explain the relationship between energy conversion and voltage drop**

Figure 3–23 illustrates charge in the form of electrons flowing from the negative terminal of a battery, through a circuit, and back to the positive terminal. As they emerge from the negative terminal, the electrons are at their highest energy level. The electrons flow through each of the resistors that are connected together to form a current path (this type of connection is called series, as you will learn in Chapter 4). As the electrons flow through each resistor, some of their energy is given up in the form of heat. Therefore, the electrons have more energy when they enter a resistor than when they exit the resistor, as illustrated in the figure by the decrease in the intensity of the red color. When they have traveled through the circuit back to the positive terminal of the battery, the electrons are at their lowest energy level.

Recall that voltage equals energy per charge ($V = W/Q$) and that charge is a property of electrons. Based on the voltage of the battery, a certain amount of energy is imparted to all of the electrons that flow out of the negative terminal. The same number of electrons flow at each point throughout the circuit, but their energy decreases as they move through the resistance of the circuit.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

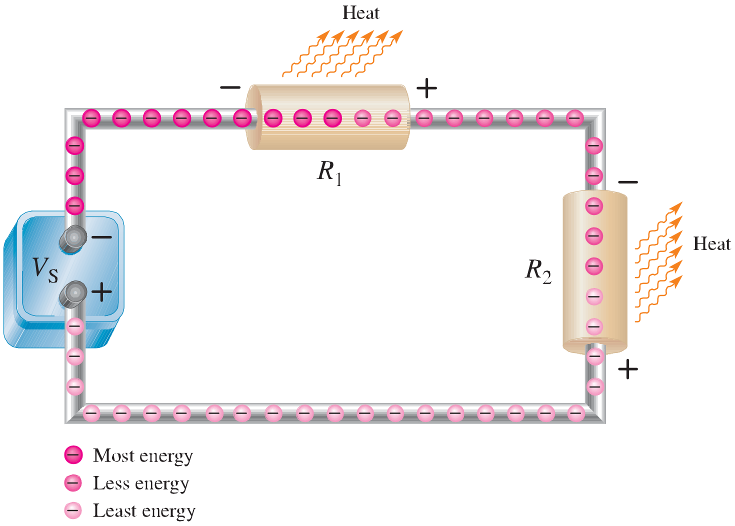

FIGURE 3–23 **A loss of energy by electrons (charge) as they flow through a resistance creates a voltage drop because voltage equals energy divided by charge.**

In Figure 3–23, the voltage at the left end of R_1 is equal to W_{enter}/Q, and the voltage at the right end of R_1 is equal to W_{exit}/Q. The same number of electrons that enter R_1 also exit R_1, so Q is constant. However, the energy W_{exit} is less than W_{enter}, so the voltage at the right end of R_1 is less than the voltage at the left end. This decrease in voltage across the resistor due to a loss of energy is called a **voltage drop**. The voltage at the right end of R_1 is less negative (or more positive) than the voltage at the left end. The voltage drop is indicated by − and + signs (the + implies a more positive voltage).

The electrons have lost some energy in R_1 and now they enter R_2 with a reduced energy level. As they flow through R_2, they lose more energy, resulting in another voltage drop across R_2.

Integrated Circuit Thermal Sensors

Heat is the bane of many electronic systems. Noncontact integrated circuit (IC) temperature sensors that measure infrared energy are available for applications where heat can be a problem and space is limited. Temperature sensors are available that include the sensor, signal conditioning, analog-to-digital converter (ADC) and a reference all in a single IC package. This simplifies data collection and recording of temperature data over a continuous time period.

Photo/image provided courtesy of Fluke Corporation.

SYSTEM NOTE

SECTION 3–6 CHECKUP

1. What is the basic reason for energy conversion in a resistor?
2. What is a voltage drop?
3. What is the polarity of a voltage drop in relation to current direction?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

3–7 POWER SUPPLIES AND BATTERIES

Recall that power supplies and batteries were briefly introduced as types of voltage sources in Section 2–3. A power supply is generally defined as a device that converts ac (alternating current) from the utility lines to a dc (direct current) voltage that virtually all electronic circuits and some transducers require. Batteries are also capable of supplying dc; in fact, many systems, such as laptop computers, can run from a power supply or internal battery. In this section, both types of voltage sources are described.

After completing this section, you should be able to

- **Discuss characteristics of power supplies and batteries**
  - **Describe controls on typical laboratory power supplies**
  - **Determine the efficiency of a power supply given the input and output power**
  - **Define ampere-hour rating of batteries**

Utilities universally have adopted ac for transmitting electricity from the generating station to the user because it can be readily transformed to high voltages for transmission and low voltages for the end user. High voltages are much more efficient and cost-effective to transmit over long distances. In the United States, the standard voltage supplied to outlets is approximately 120 V or 240 V at 60 Hz, but in Europe and other countries, the outlet voltage is 240 V at 50 Hz.

Virtually all electronic systems require stable dc for the integrated circuits and other devices to work properly. Power supplies fulfill this function by converting ac to stable dc and are usually built into the product. Many electronic systems have a recessed and protected switch that allows the internal power supply to be set for either the 120 V standard or for the 240 V standard. That switch must be set correctly, or serious damage can occur to the equipment.

HANDS ON TIP

A power supply provides both output voltage and current. You should make sure the voltage range is sufficient for your applications. Also, you must have sufficient current capacity to assure proper circuit operation. The current capacity is the maximum current that a power supply can provide to a load at a given voltage.

6238262 2015/12/10 12:27:66.5

In the laboratory, circuits are developed and tested. The purpose of a laboratory power supply is to provide the required stable dc to the circuit under test. The test circuit can be anything from a simple resistive network to a complex amplifier or logic circuit. To meet the requirement for a constant voltage, with almost no noise or ripple, laboratory power supplies are **regulated power supplies**, meaning the output is constantly sensed and automatically adjusted if it tries to change because of a change in the line voltage or the load.

Many circuits require multiple voltages, as well as the ability to set the voltage to a precise value or change it a small amount for testing. For this reason, laboratory power supplies usually have two or three outputs that are independent of each other and can be controlled separately. Output metering is normally part of a good laboratory power supply, in order to set and monitor the output voltage or current. Control may include fine and coarse controls or digital inputs to set precise voltages.

Figure 3–24 shows a triple output bench supply such as the type used in many electronic laboratories. The model shown has two 0–30 V independent supplies and a 4 V–6.5 V high current supply (commonly referred to a logic supply). Voltages can be precisely set using coarse and fine controls. The 0–30 V supplies have floating outputs, meaning they are not referenced to ground. This allows the user to set them up as a positive or negative supply or even connect them to another external supply. Another feature of this supply is that it can be set up as a current source, with a maximum voltage set for constant-current applications.

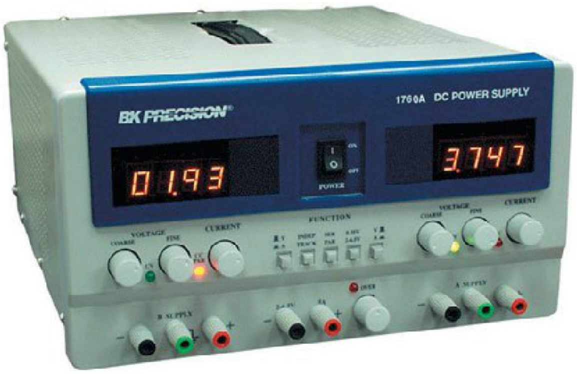

FIGURE 3–24 **A triple output power supply.** (Photo courtesy of B&K Precision Corp.)

As in many power supplies, there are three output banana jacks for each of the 0–30 V supplies. The output is taken between the red (more positive) and black terminals. The green jack is referenced to the chassis, which is earth ground. Ground can be connected to either the red or black

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

jacks, which are normally "floating" (not referenced to ground). In addition, current and voltage can be monitored using the built-in digital meters.

The power delivered by a power supply is the product of the absolute voltage and current. For example, if a power supply is providing −15.0 V at 3.0 A, the supplied power is 45 W. For a triple output supply, the total power supplied by all three supplies is the sum of the power from each one individually.

EXAMPLE 3–21

What is the total power delivered by a triple output power supply if the output voltage and current are as follows?

Source 1: 18 V at 2.0 A

Source 2: −18 V at 1.5 A

Source 3: 5.0 V at 1.0 A

SOLUTION

Power delivered from each supply is the product of voltage and current (ignoring the sign).

Source 1: $P_1 = V_1 I_1 = (18\text{ V})(2.0\text{ A}) = 36\text{ W}$

Source 2: $P_2 = V_2 I_2 = (18\text{ V})(1.5\text{ A}) = 27\text{ W}$

Source 3: $P_3 = V_3 I_3 = (5.0\text{ V})(1.0\text{ A}) = 5.0\text{ W}$

The total power is

$$P_T = P_1 + P_2 + P_3 = 36\text{ W} + 27\text{ W} + 5.0\text{ W} = \mathbf{68\text{ W}}$$

RELATED PROBLEM

How will the total power delivered change if the current from Source 1 increases to 2.5 A?

Power Supply Efficiency

An important characteristic of power supplies is efficiency. **Efficiency** is the ratio of the output power, P_{OUT}, to the input power, P_{IN}.

$$\textbf{Efficiency} = \frac{P_{\text{OUT}}}{P_{\text{IN}}} \qquad (3\text{–}8)$$

Efficiency is often expressed as a percentage. For example, if the input power is 100 W and the output power is 50 W, the efficiency is (50 W/100 W) × 100% = 50%.

All electronic power supplies are energy converters and require that power be put into them in order to get power out. For example, an electronic dc power supply might use the ac power from a wall outlet as its input. Its output is usually regulated dc voltage. The output power is *always* less than the input power because some of the total power is used internally to operate the power supply circuitry. This internal power dissipation is normally called the *power loss*. The output power is the input power minus the power loss.

$$P_{\text{OUT}} = P_{\text{IN}} - P_{\text{LOSS}} \qquad (3\text{–}9)$$

High efficiency means that very little power is dissipated in the power supply and there is a higher proportion of output power for a given input power.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 3–22

A certain electronic power supply requires 25 W of input power. It can produce an output power of 20 W. What is its efficiency, and what is the power loss?

SOLUTION

$$\text{Efficiency} = \frac{P_{\text{OUT}}}{P_{\text{IN}}} = \left(\frac{20\text{ W}}{25\text{ W}}\right) = \mathbf{0.8}$$

Expressed as a percentage,

$$\text{Efficiency} = \frac{20\text{ W}}{25\text{ W}} = 80\%$$

$$P_{\text{LOSS}} = P_{\text{IN}} - P_{\text{OUT}} = 25\text{ W} - 20\text{ W} = \mathbf{5\text{ W}}$$

RELATED PROBLEM

A power supply has an efficiency of 92%. If P_{IN} is 50 W, what is P_{OUT}?

Ampere-Hour Ratings of Batteries

Batteries convert stored chemical energy to electrical energy. They are widely used to power small systems, such as laptop computers and cell phones, to supply the stable dc required. The batteries used in these small systems are normally rechargeable, meaning that the chemical reaction can be reversed from an external source. The capacity for any battery is measured in ampere-hours (Ah). For a rechargeable battery, the Ah rating is the capacity before it needs to be recharged. The **Ah rating** determines the length of time a battery can deliver a certain amount of current at the rated voltage.

A rating of one ampere-hour means that a battery can deliver an average of one ampere of current to a load for one hour at the rated voltage output. This same battery can deliver an average of two amperes for one-half hour. The more current the battery is required to deliver, the shorter the life of the battery. In practice, a battery usually is rated for a specified current level and output voltage. For example, a 12 V automobile battery may be rated for 70 Ah at 3.5 A. This means that it can supply an average of 3.5 A for 20 h at the rated voltage.

EXAMPLE 3–23

For how many hours can a battery deliver 2 A if it is rated at 70 Ah?

SOLUTION

The ampere-hour rating is the current times the number of hours, x.

$$70\text{ Ah} \times (2\text{ A})(x\text{ h})$$

Solving for the number of hours, x, yields

$$x = \frac{70\text{ Ah}}{2\text{ A}} = \mathbf{35\text{ h}}$$

RELATED PROBLEM

A certain battery delivers 10 A for 6 h. What is its minimum Ah rating?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 3–7 CHECKUP

1. When a loading device draws an increased amount of current from a power supply, does this change represent a greater or a smaller load on the supply?
2. A power supply produces an output voltage of 10 V. If the supply provides 0.5 A to a load, what is the power output?
3. If a battery has an ampere-hour rating of 100 Ah, how long can it provide 5 A to a load?
4. If the battery in Question 3 is a 12 V device, what is its power output for the specified value of current?
5. A power supply used in the lab operates with an input power of 1 W. It can provide an output power of 750 mW. What is its efficiency?

3–8 INTRODUCTION TO TROUBLESHOOTING

Technicians must be able to diagnose and repair malfunctioning circuits and systems. In this section, you learn a general approach to troubleshooting using a simple example. Troubleshooting coverage is an important part of this textbook, so you will find a troubleshooting section in many of the chapters as well as troubleshooting problems for skill building.

6238262 2013/12/10 12:27:66.5

After completing this section, you should be able to

- **Describe a basic approach to troubleshooting**
    - **List three steps in troubleshooting**
    - **Explain what is meant by half-splitting**
    - **Discuss and compare the three basic measurements of voltage, current, and resistance**

Troubleshooting is the application of logical thinking combined with a thorough knowledge of circuit or system operation to correct a malfunction. The basic approach to troubleshooting consists of three steps: *analysis, planning,* and *measuring*. We will refer to this 3-step approach as APM.

HANDS ON TIP

When you troubleshoot a newly manufactured circuit on a pc board, certain failures are possible that are not likely on a circuit that has been working. For example, a wrong part may have accidentally been used in the new circuit or a pin may have been bent when it was installed. Cases occur where a series of newly manufactured boards have identical faults because of a wrong part.

Analysis

The first step in troubleshooting a circuit is to analyze clues or symptoms of the failure. The analysis can begin by determining the answer to certain questions:

1. Has the circuit ever worked?
2. If the circuit once worked, under what conditions did it fail?
3. What are the symptoms of the failure?
4. What are the possible causes of failure?

Planning

The second step in the troubleshooting process, after analyzing the clues, is formulating a logical plan of attack. Much time can be saved by proper planning. A working knowledge of the circuit is a prerequisite to a plan for troubleshooting. If you are not certain how the circuit is supposed to operate, take time to review circuit diagrams (schematics), operating instructions, and other pertinent information. A schematic with proper voltages marked at various test points is particularly useful. Although logical thinking is perhaps the most important tool in troubleshooting, it rarely can solve the problem by itself.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Measuring

The third step is to narrow the possible failures by making carefully thought out measurements. These measurements usually confirm the direction you are taking in solving the problem, or they may point to a new direction that you should take. Occasionally, you may find a totally unexpected result.

An APM Example

The thought process that is part of the APM approach can be illustrated with a simple example. Suppose you have a string of eight decorative 15 V bulbs connected in series to a 120 V source V_S, as shown in Figure 3–25. Assume that this circuit worked properly at one time but stopped working after it was moved to a new location. When plugged in at the new location, the bulbs fail to turn on. How do you go about finding the trouble?

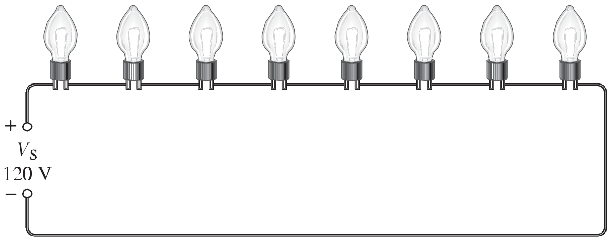

FIGURE 3–25 **A string of bulbs connected to a voltage source.**

THE ANALYSIS THOUGHT PROCESS You may think like this as you proceed to analyze the situation:

- Since the circuit worked before it was moved, the problem could be that there is no voltage at the new location.
- Perhaps the wiring was loose and pulled apart when moved.
- It is possible that a bulb is burned out or loose in its socket.

With this reasoning, you have considered possible causes and failures that may have occurred. The thought process continues:

- The fact that the circuit once worked eliminates the possibility that the original circuit was improperly wired.
- If the fault is due to an open path, it is unlikely that there is more than one break which could be either a bad connection or a burned out bulb.

You have now analyzed the problem and are ready to plan the process of finding the fault in the circuit.

THE PLANNING THOUGHT PROCESS The first part of your plan is to measure for voltage at the new location. If the voltage is present, then the problem is in the string of lights. If voltage is not present, check the circuit breaker at the distribution box in the house. Before resetting breakers, you should think about why a breaker may be tripped. Let's assume that you find the voltage is present. This means that the problem is in the string of lights.

The second part of your plan is to measure either the resistance in the string of lights or to measure voltages across the bulbs. The decision whether to measure resistance or voltage is a toss-up and can be made based on the ease of making the test. Seldom is a troubleshooting plan developed so completely that all possible contingencies are included. You will frequently need to modify the plan as you go along.

THE MEASUREMENT PROCESS You proceed with the first part of your plan by using a multimeter to check the voltage at the new location. Assume the measurement shows a voltage of 120 V. Now you have eliminated the possibility of no voltage. You know that, since you have voltage across the string and there is no current because no bulb is on, there must be an open in the current path. Either a bulb is burned out, a connection at the lamp socket is broken, or the wire is broken.

Next, you decide to locate the break by measuring resistance with your multimeter. Applying logical thinking, you decide to measure the resistance of each half of the string instead of measuring the resistance of each bulb. By measuring the resistance of half the

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

bulbs at once, you can usually reduce the effort required to find the open. This technique is a type of troubleshooting procedure called **half-splitting**.

Once you have identified the half in which the open occurs, as indicated by an infinite resistance, you use half-splitting again on the faulty half and continue until you narrow the fault down to a faulty bulb or connection. This process is shown in Figure 3–26, assuming for purposes of illustration that the seventh bulb is burned out.

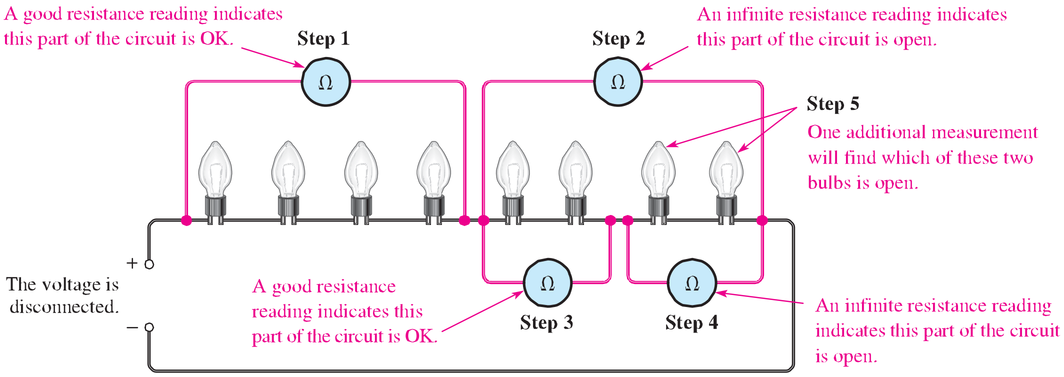
A good resistance reading indicates this part of the circuit is OK.
Step 1
Ω
Step 2
An infinite resistance reading indicates this part of the circuit is open.
Step 5
One additional measurement will find which of these two bulbs is open.
+
The voltage is disconnected.
−
A good resistance reading indicates this part of the circuit is OK.
Step 3
Step 4
An infinite resistance reading indicates this part of the circuit is open.

FIGURE 3–26 Illustration of the half-splitting method of troubleshooting. The numbered steps indicate the sequence in which the multimeter is moved from one position to another.

As you can see in the figure, the half-splitting approach in this particular case takes a maximum of five measurements to identify the open bulb. If you had decided to measure each bulb individually and had started at the left, you would have needed seven measurements. So, sometimes half-splitting saves steps; sometimes it doesn't. The number of steps required depends on where you make your measurements and in what sequence.

Unfortunately, most troubleshooting is more difficult than this example. However, analysis and planning are essential for effective troubleshooting in any situation. As measurements are made, the plan is often modified; the experienced troubleshooter narrows the search by fitting the symptoms and measurements into a probable cause. In some cases, low-cost equipment is simply discarded or recycled when troubleshooting and repair costs are comparable to replacement costs.

Comparison of *V*, *R*, and *I* Measurements

As you know from Section 2–7, you can measure voltage, current, or resistance in a circuit. To measure voltage, place the voltmeter in parallel across the component; that is, place one lead on each side of the component. This makes voltage measurements the easiest of the three types of measurements.

To measure resistance, connect the ohmmeter across a component; however, the voltage must be first disconnected, and sometimes the component itself must be removed from the circuit. Therefore, resistance measurements are generally more difficult than voltage measurements.

To measure current, place the ammeter in series with the component; that is, the ammeter must be in line with the current path. To do this you must disconnect a component lead or a wire before you connect the ammeter. This usually makes a current measurement the most difficult to perform.

SECTION 3–8 CHECKUP

1. Name the three steps in the APM approach to troubleshooting.
2. Explain the basic idea of the half-splitting technique.
3. Why are voltages easier to measure than currents in a circuit?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SUMMARY

- Voltage and current are linearly proportional.
- Ohm's law gives the relationship of voltage, current, and resistance.
- Current is inversely proportional to resistance.
- A kilohm (kΩ) is one thousand ohms.
- A megohm (MΩ) is one million ohms.
- A microampere (μA) is one-millionth of an ampere.
- A milliampere (mA) is one-thousandth of an ampere.
- Use $I = V/R$ to calculate current.
- Use $V = IR$ to calculate voltage.
- Use $R = V/I$ to calculate resistance.
- One watt equals one joule per second.
- Watt is the unit of power, and joule is the unit of energy.
- The power rating of a resistor determines the maximum power that it can handle safely.
- Resistors with a larger physical size can dissipate more power in the form of heat than smaller ones.
- A resistor should have a power rating as high or higher than the maximum power that it is expected to handle in the circuit.
- Power rating is not related to resistance value.
- A resistor usually opens when it overheats and fails.
- Energy is equal to power multiplied by time.
- The kilowatt-hour is a unit of energy.
- An example of one kilowatt-hour is one thousand watts used for one hour.
- A power supply is an energy source used to operate electrical and electronic devices.
- A battery is one type of power supply that converts chemical energy into electrical energy.
- A power supply converts commercial energy (ac from the power company) to regulated dc at various voltage levels.
- The output power of a power supply is the output voltage times the load current.
- A load is a device that draws current from the power supply.
- The capacity of a battery is measured in ampere-hours (Ah).
- One ampere-hour equals one ampere used for one hour, or any other combination of amperes and hours that has a product of one.
- A power supply with a high efficiency has a smaller percentage power loss than one with a lower efficiency.
- The formula wheel in Figure 3–27 gives the Ohm's law and Watt's law relationships.
- APM (analysis, planning, and measurement) provides a logical approach to troubleshooting.
- The half-splitting method of troubleshooting generally results in fewer measurements.

6238262 2013/12/10 12.27.66.5

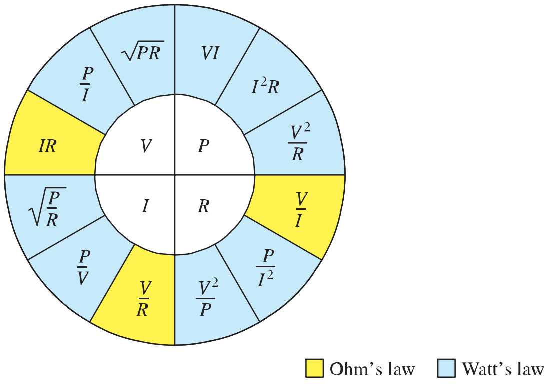

FIGURE 3–27

KEY TERMS

Key terms and other bold terms in the chapter are defined in the end-of-book glossary.

Ah rating A capacity rating for batteries determined by multiplying the current (A) times the length of time (h) a battery can deliver that current to a load.

Efficiency The ratio of the output power to the input power of a circuit, usually expressed as a percent.

Energy The ability to do work. The unit is the joule (J).

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Half-splitting A troubleshooting procedure where one starts in the middle of a circuit or system and, depending on the first measurement, works toward the output or toward the input to find the fault.

Joule (J) The SI unit of energy.

Kilowatt-hour (kWh) A large unit of energy used mainly by utility companies.

Linear Characterized by a straight-line relationship.

Ohm's law A law stating that current is directly proportional to voltage and inversely proportional to resistance.

Power The rate of energy usage. The unit is the watt (W).

Power rating The maximum amount of power that a resistor can dissipate without being damaged by excessive heat buildup.

Power supply A device that converts ac from the utility lines to a dc voltage.

Troubleshooting A systematic process of isolating, identifying, and correcting a fault in a circuit or system.

Voltage drop The decrease in voltage across a resistor due to a loss of energy.

Watt (W) The unit of power. One watt is the power when 1 J of energy is used in 1 s.

Watt's law A law that states the relationships of power to current, voltage, and resistance.

KEY FORMULAS

(3–1)	$I = \frac{V}{R}$	Form of Ohm's law for calculating current
(3–2)	$V = IR$	Form of Ohm's law for calculating voltage
(3–3)	$R = \frac{V}{I}$	Form of Ohm's law for calculating resistance
(3–4)	$P = \frac{W}{t}$	Power equals energy divided by time.
(3–5)	$P = I^2R$	Power equals current squared times resistance.
(3–6)	$P = VI$	Power equals voltage times current.
(3–7)	$P = \frac{V^2}{R}$	Power equals voltage squared divided by resistance.
(3–8)	$\text{Efficiency} = \frac{P_{OUT}}{P_{IN}}$	Power supply efficiency
(3–9)	$P_{OUT} = P_{IN} - P_{LOSS}$	Output power

TRUE/FALSE QUIZ

Answers are at the end of the chapter.

1. If the total resistance of a circuit increases, current decreases.
2. Ohm's law for finding resistance is $R = I/V$.
3. When milliamps and kilohms are multiplied together, the result is volts.
4. If a 10 kΩ resistor is connected to a 10 V source, the current in the resistor will be 1 A.
5. The kilowatt-hour is a unit of power.
6. One watt is equal to one joule per second.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

7. The power rating of a resistor should always be less than the required power dissipation in the circuit.

8. Within limits, a regulated power supply can automatically keep the output voltage constant even if the load changes.

9. A power supply that has a negative output voltage absorbs power from the load.

10. When analyzing a circuit problem, you should consider the conditions under which it failed.

SELF-TEST

Answers are at the end of the chapter.

1. Ohm's law states that

(a) current equals voltage times resistance
(b) voltage equals current times resistance
(c) resistance equals current divided by voltage
(d) voltage equals current squared times resistance

2. When the voltage across a resistor is doubled, the current will

(a) triple **(b)** halve **(c)** double **(d)** not change

3. When 10 V are applied across a 20 Ω resistor, the current is

(a) 10 A **(b)** 0.5 A **(c)** 200 A **(d)** 2 A

4. When there are 10 mA of current through a 1.0 kΩ resistor, the voltage across the resistor is

(a) 100 V **(b)** 0.1 V **(c)** 10 kV **(d)** 10 V

5. If 20 V are applied across a resistor and there are 6.06 mA of current, the resistance is

(a) 3.3 kΩ **(b)** 33 kΩ **(c)** 330 Ω **(d)** 3.03 kΩ

6. A current of 250 μA through a 4.7 kΩ resistor produces a voltage drop of

(a) 53.2 V **(b)** 1.175 mV **(c)** 18.8 V **(d)** 1.175 V

7. A resistance of 2.2 MΩ is connected across a 1 kV source. The resulting current is approximately

(a) 2.2 mA **(b)** 455 μA **(c)** 45.5 μA **(d)** 0.455 A

8. Power can be defined as

(a) energy **(b)** heat
(c) the rate at which energy is used **(d)** the time required to use energy

9. For 10 V and 50 mA, the power is

(a) 500 mW **(b)** 0.5 W
(c) 500,000 μW **(d)** answers (a), (b), and (c)

10. When the current through a 10 kΩ resistor is 10 mA, the power is

(a) 1 W **(b)** 10 W **(c)** 100 mW **(d)** 1 mW

11. A 2.2 kΩ resistor dissipates 0.5 W. The current is

(a) 15.1 mA **(b)** 227 μA **(c)** 1.1 mA **(d)** 4.4 mA

12. A 330 Ω resistor dissipates 2 W. The voltage is

(a) 2.57 V **(b)** 660 V **(c)** 6.6 V **(d)** 25.7 V

13. The power rating of a resistor that is to handle up to 1.1 W should be

(a) 0.25 W **(b)** 1 W **(c)** 2 W **(d)** 5 W

14. A 22 Ω half-watt resistor and a 220 Ω half-watt resistor are connected across a 10 V source. Which one(s) will overheat?

(a) 22 Ω **(b)** 220 Ω **(c)** both **(d)** neither

15. When the needle of an analog ohmmeter indicates infinity, the resistor being measured is

(a) overheated **(b)** shorted **(c)** open **(d)** reversed

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

TROUBLESHOOTING: SYMPTOM AND CAUSE

The purpose of these exercises is to help develop thought processes essential to troubleshooting. Answers are at the end of the chapter.

Determine the cause for each set of symptoms. Refer to Figure 3–28.

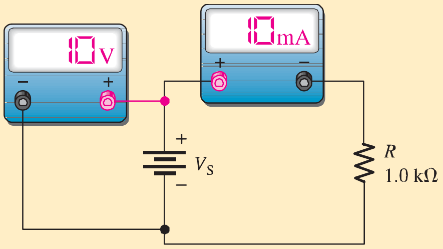

FIGURE 3–28 The meters indicate the correct readings for this circuit.

1. *Symptom:* The ammeter reading is zero, and the voltmeter reading is 10 V.
   Cause:
   (a) *R* is shorted.
   (b) *R* is open.
   (c) The voltage source is faulty.

6238262 2013/12/10 12.27.66.5

2. *Symptom:* The ammeter reading is zero, and the voltmeter reading is 0 V.
   Cause:
   (a) *R* is open.
   (b) *R* is shorted.
   (c) The voltage source is turned off or faulty.
3. *Symptom:* The ammeter reading is 10 mA, and the voltmeter reading is 0 V.
   Cause:
   (a) The voltmeter is defective.
   (b) The ammeter is faulty.
   (c) The voltage source is turned off or faulty.
4. *Symptom:* The ammeter reading is 1 mA, and the voltmeter reading is 10 V.
   Cause:
   (a) The voltmeter is defective.
   (b) The resistor value is higher than it should be.
   (c) The resistor value is lower than it should be.
5. *Symptom:* The ammeter reading is 100 mA, and the voltmeter reading is 10 V.
   Cause:
   (a) The voltmeter is defective.
   (b) The resistor value is higher than it should be.
   (c) The resistor value is lower than it should be.

PROBLEMS

Answers to odd-numbered problems are at the end of the book.

BASIC PROBLEMS

SECTION 3–1 Ohm's Law

1. The current in a circuit is 1 A. Determine what the current will be when
   (a) the voltage is tripled
   (b) the voltage is reduced by 80%
   (c) the voltage is increased by 50%

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

2. The current in a circuit is 100 mA. Determine what the current will be when
   - **(a)** the resistance is increased by 100%
   - **(b)** the resistance is reduced by 30%
   - **(c)** the resistance is quadrupled
3. The current in a circuit is 10 mA. What will the current be if the voltage is tripled and the resistance is doubled?

SECTION 3–2 Application of Ohm's Law

4. Determine the current in each case.
   - **(a)** $V = 5\text{ V}, R = 1.0\ \Omega$
   - **(b)** $V = 15\text{ V}, R = 10\ \Omega$
   - **(c)** $V = 50\text{ V}, R = 100\ \Omega$
   - **(d)** $V = 30\text{ V}, R = 15\text{ k}\Omega$
   - **(e)** $V = 250\text{ V}, R = 4.7\text{ M}\Omega$
5. Determine the current in each case.
   - **(a)** $V = 9\text{ V}, R = 2.7\text{ k}\Omega$
   - **(b)** $V = 5.5\text{ V}, R = 10\text{ k}\Omega$
   - **(c)** $V = 40\text{ V}, R = 68\text{ k}\Omega$
   - **(d)** $V = 1\text{ kV}, R = 2\text{ k}\Omega$
   - **(e)** $V = 66\text{ kV}, R = 10\text{ M}\Omega$
6. A 10 Ω resistor is connected across a 12 V battery. How much current is there through the resistor?
7. A resistor is connected across the terminals of a dc voltage source in each part of Figure 3–29. Determine the current in each resistor.

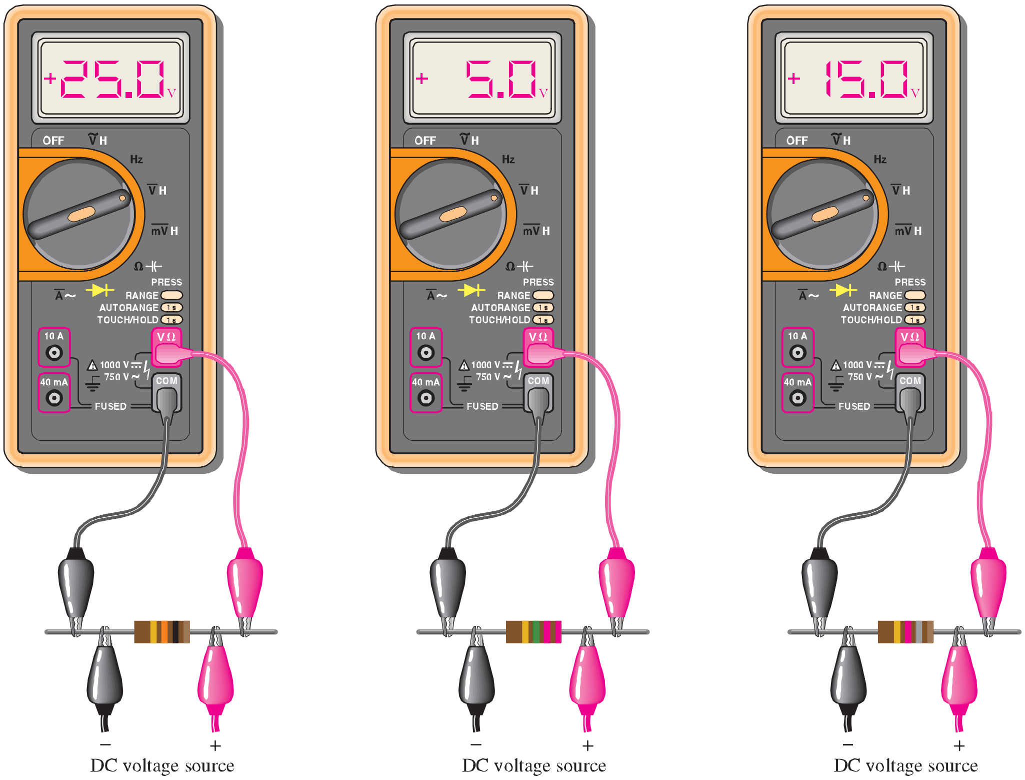

(a) (b) (c)

FIGURE 3–29

8. A 5-band resistor is connected across a 12 V source. Determine the current if the resistor color code is orange, violet, yellow, gold, brown.
9. If the voltage in Problem 8 is doubled, will a 0.5 A fuse blow? Explain your answer.
10. Calculate the voltage for each value of I and R.
    - **(a)** $I = 2\text{ A}, R = 18\ \Omega$
    - **(b)** $I = 5\text{ A}, R = 47\ \Omega$
    - **(c)** $I = 2.5\text{ A}, R = 620\ \Omega$
    - **(d)** $I = 0.6\text{ A}, R = 47\ \Omega$
    - **(e)** $I = 0.1\text{ A}, R = 470\ \Omega$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

11. Calculate the voltage for each value of I and R.
(a) $I = 1\text{ mA}, R = 10\ \Omega$
(b) $I = 50\text{ mA}, R = 33\ \Omega$
(c) $I = 3\text{ A}, R = 4.7\text{ k}\Omega$
(d) $I = 1.6\text{ mA}, R = 2.2\text{ k}\Omega$
(e) $I = 250\ \mu\text{A}, R = 1.0\text{ k}\Omega$
(f) $I = 500\text{ mA}, R = 1.5\text{ M}\Omega$
(g) $I = 850\ \mu\text{A}, R = 10\text{ M}\Omega$
(h) $I = 75\ \mu\text{A}, R = 47\ \Omega$

12. A 20 mΩ current-sensing resistor has 3 A. What is the voltage across the current-sensing resistor?

13. Assign a voltage value to each source in the circuits of Figure 3–30 to obtain the indicated amounts of current.

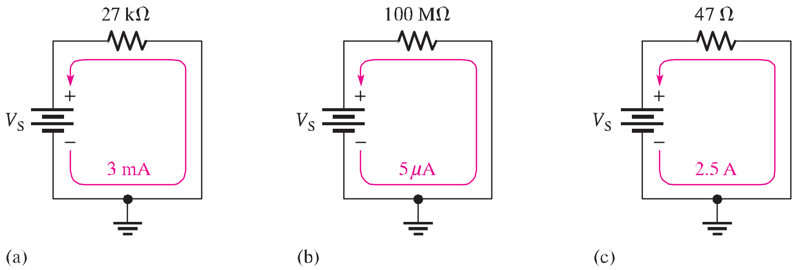

FIGURE 3–30

14. Calculate the resistance for each value of V and I.
(a) $V = 10\text{ V}, I = 2\text{ A}$
(b) $V = 90\text{ V}, I = 45\text{ A}$
(c) $V = 50\text{ V}, I = 5\text{ A}$
(d) $V = 5.5\text{ V}, I = 10\text{ A}$
(e) $V = 150\text{ V}, I = 0.5\text{ A}$

15. Calculate R for each set of V and I values.
(a) $V = 10\text{ kV}, I = 5\text{ A}$
(b) $V = 7\text{ V}, I = 2\text{ mA}$
(c) $V = 500\text{ V}, I = 250\text{ mA}$
(d) $V = 50\text{ V}, I = 500\ \mu\text{A}$
(e) $V = 1\text{ kV}, I = 1\text{ mA}$

16. Six volts are applied across a resistor. A current of 2 mA is measured. What is the value of the resistor?

17. Choose the correct value of resistance to get the current values indicated in each circuit of Figure 3–31.

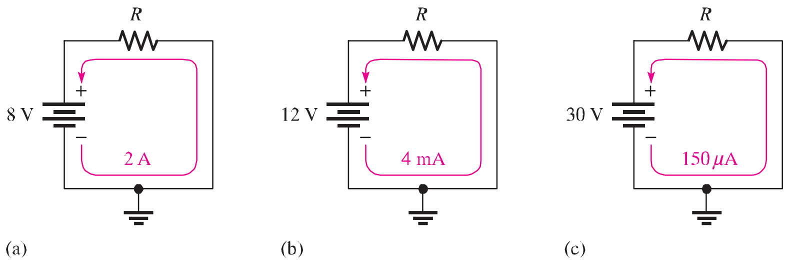

FIGURE 3–31

18. A flashlight is operated from 3.2 V and has a resistance of 3.9 Ω when the filament is hot. What current is supplied by the batteries?

SECTION 3–3 Energy and Power

19. The flashlight in Problem 18 uses 26 J in 10 s. What is the power in watts?

20. What is the power when energy is used at the rate of 350 J/s?

21. What is the power in watts when 7500 J of energy are used in 5 h?

22. Convert the following to kilowatts:
(a) 1000 W (b) 3750 W (c) 160 W (d) 50,000 W

23. Convert the following to megawatts:
(a) 1,000,000 W (b) 3×10^6 W (c) 15×10^7 W (d) 8700 kW

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

24. Convert the following to milliwatts:
(a) 1 W
(b) 0.4 W
(c) 0.002 W
(d) 0.0125 W

25. Convert the following to microwatts:
(a) 2 W
(b) 0.0005 W
(c) 0.25 mW
(d) 0.00667 mW

26. Convert the following to watts:
(a) 1.5 kW
(b) 0.5 MW
(c) 350 mW
(d) 9000 μW

27. Prove that the unit for power (the watt) is equivalent to 1 V × 1 A.

28. Show that there are 3.6×10^6 joules in a kilowatt-hour.

SECTION 3–4 Power in an Electric Circuit

29. If a resistor has 5.5 V across it and 3 mA through it, what is the power dissipation?

30. An electric heater works on 115 V and draws 3 A of current. How much power does it use?

31. What is the power when there are 500 mA of current through a 4.7 kΩ resistor?

32. A 25 mΩ current-sensing resistor has 5 A of current. What power is dissipated?

33. If there are 60 V across a 620 Ω resistor, what is the power dissipation?

34. A 56 Ω resistor is connected across the terminals of a 1.5 V battery. What is the power dissipation in the resistor?

35. If a resistor is to carry 2 A of current and handle 100 W of power, how many ohms must it be? Assume that the voltage can be adjusted to any required value.

36. Convert 5×10^6 watts used for 1 minute to kWh.

37. Convert 6700 watts used for 1 second to kWh.

38. How many kilowatt-hours do 50 W used for 12 h equal?

39. Assume that an alkaline D-cell battery can maintain an average voltage of 1.25 V for 90 h in a 10 Ω load before becoming unusable. What average power is delivered to the load during the life of the battery?

40. What is the total energy in joules that is delivered during the 90 hours for the battery in Problem 39?

SECTION 3–5 The Power Rating of Resistors

41. A 6.8 kΩ resistor has burned out in a circuit. You must replace it with another resistor with the same resistance value. If the resistor carries 10 mA, what should its power rating be? Assume that you have available resistors in all the standard power ratings.

42. A certain type of power resistor comes in the following ratings: 3 W, 5 W, 8 W, 12 W, 20 W. Your particular application requires a resistor that can handle approximately 8 W. Which rating would you use for a minimum safety margin of 20% above the rated value? Why?

SECTION 3–6 Energy Conversion and Voltage Drop in a Resistance

43. For each circuit in Figure 3–32, assign the proper polarity for the voltage across the resistor.

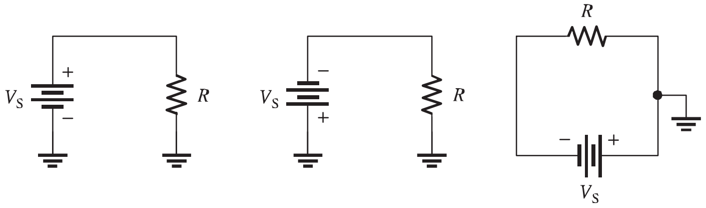

(a) (b) (c)

FIGURE 3–32

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 3–7 Power Supplies and Batteries

44. A 50 Ω load dissipates 1 W of power. What is the output voltage of the power supply?

45. A battery can provide an average of 1.5 A of current for 24 h. What is its ampere-hour rating?

46. How much average current can be drawn from an 80 Ah battery for 10 h?

47. If a battery is rated at 650 mAh, how much average current will it provide for 48 h?

48. If the input power is 500 mW and the output power is 400 mW, what is the power loss? What is the efficiency of this power supply?

49. To operate at 85% efficiency, how much output power must a source produce if the input power is 5 W?

SECTION 3–8 Introduction to Troubleshooting

50. In the light circuit of Figure 3–33, identify the faulty bulb based on the series of ohmmeter readings shown.

51. Assume you have a 32-light string and one of the bulbs is burned out. Using the half-splitting approach and starting in the left half of the circuit, how many resistance measurements will it take to find the faulty bulb if it is seventeenth from the left.

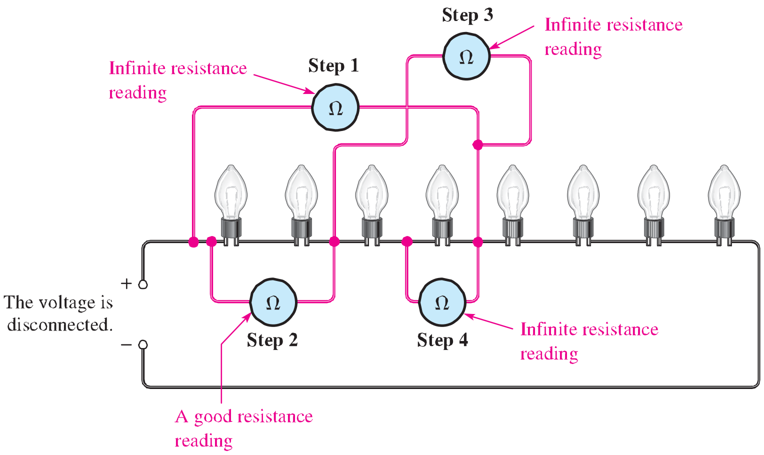
Step 3
Infinite resistance reading
Infinite resistance reading
Step 1
Ω
+
The voltage is disconnected.
–
Step 2
Step 4
Infinite resistance reading
A good resistance reading

FIGURE 3–33

ADVANCED PROBLEMS

52. A certain power supply provides a continuous 2 W to a load. It is operating at 60% efficiency. In a 24 h period, how many kilowatt-hours does the power supply use?

53. The filament of a light bulb in the circuit of Figure 3–34(a) has a certain amount of resistance, represented by an equivalent resistance in Figure 3–34(b). If the bulb operates with 120 V and 0.8 A of current, what is the resistance of its filament?

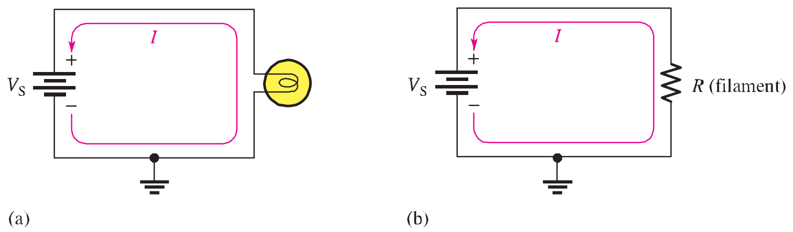

FIGURE 3–34

54. A certain electrical device has an unknown resistance. You have available a 12 V battery and an ammeter. How would you determine the value of the unknown resistance? Draw the necessary circuit connections.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

55. A variable voltage source is connected to the circuit of Figure 3–35. Start at 0 V and increase the voltage in 10 V steps up to 100 V. Determine the current at each voltage value, and plot a graph of V versus I. Is the graph a straight line? What does the graph indicate?

56. In a certain circuit, $V_S = 1$ V and $I = 5$ mA. Determine the current for each of the following voltages in a circuit with the same resistance:
(a) $V_S = 1.5$ V **(b)** $V_S = 2$ V **(c)** $V_S = 3$ V
(d) $V_S = 4$ V **(e)** $V_S = 10$ V

57. Figure 3–36 is a graph of current versus voltage for three resistance values. Determine R_1, R_2, and R_3.

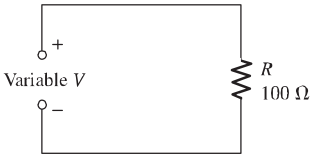

FIGURE 3–35

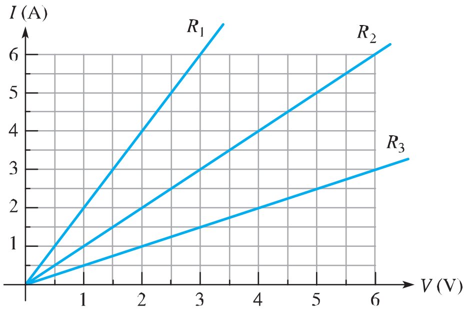

FIGURE 3–36

58. You are measuring the current in a circuit that is operated on a 10 V battery. The ammeter reads 50 mA. Later, you notice that the current has dropped to 30 mA. Eliminating the possibility of a resistance change, you must conclude that the voltage has changed. How much has the voltage of the battery changed, and what is its new value?

59. If you wish to increase the amount of current in a resistor from 100 mA to 150 mA by changing the 20 V source, by how many volts should you change the source? To what new value should you set it?

60. A 6 V source is connected to a 100 Ω resistor by two 12 ft lengths of 18-gauge copper wire. Refer to Table 2–3 to determine the following:
(a) current **(b)** resistor voltage **(c)** voltage across each length of wire

61. If a 300 W bulb is allowed to burn continuously for 30 days, how many kilowatt-hours of energy does it use?

62. At the end of a 31-day period, your utility bill shows that you have used 1500 kWh. What is the average daily power?

63. A certain type of power resistor comes in the following ratings: 3 W, 5 W, 8 W, 12 W, 20 W. Your particular application requires a resistor that can handle approximately 10 W. Which rating would you use? Why?

64. A 12 V source is connected across a 10 Ω resistor for 2 min.
(a) What is the power dissipation?
(b) How much energy is used?
(c) If the resistor remains connected for an additional minute, does the power dissipation increase or decrease?

65. The rheostat in Figure 3–37 is used to control the current to a heating element. When the rheostat is adjusted to a value of 8 Ω or less, the heating element can burn out. What is the rated value of the fuse needed to protect the circuit if the voltage across the heating element at the point of maximum current is 100 V?

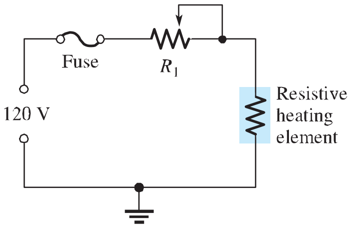

FIGURE 3–37

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

66. A 30 mΩ current-sensing resistor is rated for ½ W. What is the maximum current it can have before the rating is exceeded?

6238262 2013/12/10 12.27.66.5

67. What happens to the power dissipated in a resistor if the current is doubled?

68. Develop a complete materials list for the load test box in Figure 3–21.

69. Draw the schematic for the load test box in Figure 3–21.

MULTISIM

MULTISIM TROUBLESHOOTING PROBLEMS

70. Open file P03-70; files are found at www.pearsonhighered.com/floyd. Determine whether or not the circuit is operating properly. If not, find the fault.

71. Open file P03-71. Determine whether or not the circuit is operating properly. If not, find the fault.

72. Open file P03-72. Determine whether or not the circuit is operating properly. If not, find the fault.

73. Open file P03-73. Determine whether or not the circuit is operating properly. If not, find the fault.

74. Open file P03-74. Determine whether or not the circuit is operating properly. If not, find the fault.

ANSWERS TO SECTION CHECKUPS

SECTION 3–1 Ohm's Law

1. Ohm's law states that current varies directly with voltage and inversely with resistance.
2. $I = V/R$
3. $V = IR$
4. $R = V/I$
5. The current increases; three times when V is tripled
6. Doubling R cuts I in half to 5 mA.
7. No change in I if V and R are doubled.

SECTION 3–2 Application of Ohm's Law

1. $I = 10\text{ V}/4.7\ \Omega = 2.13\text{ A}$
2. $I = 20\text{ kV}/4.7\text{ M}\Omega = 4.26\text{ mA}$
3. $I = 10\text{ kV}/2\text{ k}\Omega = 5\text{ A}$
4. $V = (1\text{ A})(10\ \Omega) = 10\text{ V}$
5. $V = (3\text{ mA})(3\text{ k}\Omega) = 9\text{ V}$
6. $V = (2\text{ A})(6\ \Omega) = 12\text{ V}$
7. $R = 10\text{ V}/2\text{ A} = 5\ \Omega$
8. $R = 25\text{ V}/50\text{ mA} = 0.5\text{ k}\Omega = 500\ \Omega$

SECTION 3–3 Energy and Power

1. Power is the rate at which energy is used.
2. $P = W/t$
3. Watt is the unit of power. One watt is the power when 1 J of energy is used in 1 s.
4. **(a)** 68,000 W $=$ 68 kW **(b)** 0.005 W $=$ 5 mW **(c)** 0.000025 $=$ 25 μW
5. (100 W)(10 h) $=$ 1 kWh
6. 2000 W $=$ 2 kW
7. (1.322 kW)(24 h) $=$ 31.73 kWh; (0.11 \$/kWh)(31.73 kWh) $=$ \$3.49

SECTION 3–4 Power in an Electric Circuit

1. $P = IV = (12\text{ A})(13\text{ V}) = 156\text{ W}$
2. $P = (5\text{ A})^2(47\ \Omega) = 1175\text{ W}$
3. $V = \sqrt{PR} = \sqrt{(2\text{ W})(50\ \Omega)} = 10\text{ V}$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

4. $P = \frac{V^2}{R} = \frac{(13.4\ \text{V})^2}{3.0\ \Omega} = 60\ \text{W}$
5. $P = (8\ \text{V})^2/2.2\ \text{k}\Omega = 29.1\ \text{mW}$
6. $R = 55\ \text{W}/(0.5\ \text{A})^2 = 220\ \Omega$

SECTION 3–5 The Power Rating of Resistors

1. Two resistor parameters are resistance and power rating.
2. A larger resistor physical size dissipates more energy.
3. Standard ratings of metal-film resistors are 0.125 W, 0.25 W, 0.5 W, and 1 W.
4. At least a 0.5 W rating to handle 0.3 W
5. 5.0 V

SECTION 3–6 Energy Conversion and Voltage Drop in a Resistance

1. Energy conversion in a resistor is caused by collisions of free electrons with the atoms in the material.
2. Voltage drop is a decrease in voltage across a resistor due to a loss of energy.
3. Voltage drop is negative to positive in the direction of current.

SECTION 3–7 Power Supplies and Batteries

1. An increased amount of current represents a greater load.
2. $P_{\text{OUT}} = (10\ \text{V})(0.5\ \text{A}) = 5\ \text{W}$
3. $100\ \text{Ah}/5\ \text{A} = 20\ \text{h}$
4. $P_{\text{OUT}} = (12\ \text{V})(5\ \text{A}) = 60\ \text{W}$
5. $\text{Efficiency} = (750\ \text{mW}/1000\ \text{mW})100\% = 75\%$

SECTION 3–8 Introduction to Troubleshooting

1. Analysis, Planning, and Measuring
2. Half-splitting identifies the fault by successively isolating half of the remaining circuit.
3. Voltage is measured across a component. Current is measured in series with the component.

ANSWERS TO RELATED PROBLEMS FOR EXAMPLES

3–1 $I_1 = 10\ \text{V}/5\ \Omega = 2\ \text{A}; I_2 = 10\ \text{V}/20\ \Omega = 0.5\ \text{A}$

3–2 1 V

3–3 970 mΩ

3–4 Yes

3–5 11.1 mA

3–6 5 mA

3–7 25 μA

3–8 200 μA

3–9 800 V

3–10 220 mV

3–11 16.5 V

3–12 Voltage drops to 1.08 V.

3–13 6.0 Ω

3–14 39.5 kΩ

3–15 3000 J

3–16 2 kWh

3–17 \$4.92 + \$0.07 = \$4.99

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

3–18 **(a)** 40 W **(b)** 376 W **(c)** 625 mW
3–19 26 W
3–20 No
3–21 77 W
3–22 46 W
3–23 60 Ah

ANSWERS TO TRUE/FALSE QUIZ

1. T **2.** F **3.** T **4.** F **5.** F **6.** T **7.** F **8.** T
9. F **10.** T

ANSWERS TO SELF-TEST

1. (b) **2.** (c) **3.** (b) **4.** (d) **5.** (a) **6.** (d) **7.** (b) **8.** (c)
9. (d) **10.** (a) **11.** (a) **12.** (d) **13.** (c) **14.** (a) **15.** (c)

ANSWERS TO TROUBLESHOOTING: SYMPTOM AND CAUSE

1. (b)
2. (c)
3. (a)
4. (b)
5. (c)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

CHAPTER 4

SERIES CIRCUITS

OUTLINE

4–1 Resistors in Series
4–2 Total Series Resistance
4–3 Current in a Series Circuit
4–4 Application of Ohm's Law
4–5 Voltage Sources in Series
4–6 Kirchhoff's Voltage Law
4–7 Voltage Dividers
4–8 Power in Series Circuits
4–9 Voltage Measurements
4–10 Troubleshooting

OBJECTIVES

- Identify a series resistive circuit
- Determine total series resistance
- Determine the current throughout a series circuit
- Apply Ohm's law in series circuits
- Determine the total effect of voltage sources connected in series
- Apply Kirchhoff's voltage law
- Use a series circuit as a voltage divider
- Determine power in a series circuit
- State how to measure voltage with respect to ground
- Troubleshoot series circuits

KEY TERMS

Series
Series-aiding
Series-opposing
Kirchhoff's voltage law
Voltage divider
Reference ground
Open
Short

INTRODUCTION

Resistive circuits can be of two basic forms: series or parallel. In this chapter series circuits are discussed. Parallel circuits are covered in Chapter 5, and combinations of series and parallel circuits are examined in Chapter 6. In this chapter you will see how Ohm's law is used in series circuits; and you will study another important law, Kirchhoff's voltage law. Also, important system applications of series circuits are presented.

6238262 2013/12/10 12.27.66.5

VISIT THE WEBSITE
Study aids for this chapter are available at
http://pearsonhighered.com/floyd

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

4–1 RESISTORS IN SERIES

When connected in series, resistors form a "string" in which there is only one path for current.

After completing this section, you should be able to

- **Identify a series resistive circuit**
  - **Translate a physical arrangement of resistors into a schematic diagram**

Figure 4–1(a) shows two resistors connected in series between point A and point B. Part (b) shows three resistors in series, and part (c) shows four in series. Of course, there can be any number of resistors in series.

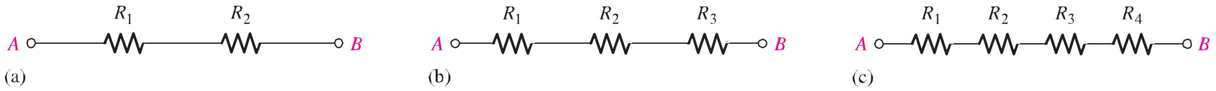

FIGURE 4–1 **Resistors in series.**

When a voltage source is connected from point A to point B, the only way for current to get from one point to the other in any of the connections of Figure 4–1 is to go through each of the resistors. A series circuit is identified as follows:

> **A series circuit provides only one path for current between two points so that the current is the same through each series resistor.**

In an actual circuit diagram, a series circuit may not always be as easy to identify as those in Figure 4–1. For example, Figure 4–2 shows series resistors drawn in other ways with a voltage applied. Remember, if there is only one current path between two points, the resistors between those two points are in series, no matter how they appear in a diagram.

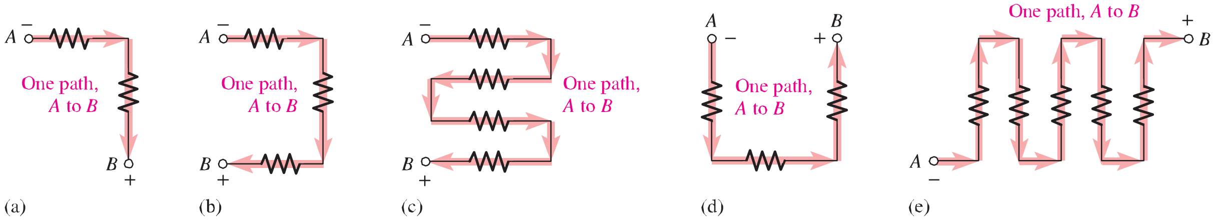

FIGURE 4–2 **Some examples of series connections of resistors. Notice that the current must be the same at all points because the current has only one path.**

EXAMPLE 4–1

Five resistors are positioned on a protoboard as shown in Figure 4–3. Wire them together in series so that, starting from the negative (−) terminal, R_1 is first, R_2 is second, R_3 is third, and so on. Draw a schematic showing this connection.

FIGURE 4–3

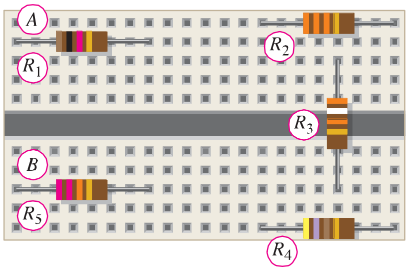

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

The wires are connected as shown in Figure 4–4(a), which is the assembly diagram. The schematic is shown in Figure 4–4(b). Note that the schematic does not necessarily show the actual physical arrangement of the resistors as does the assembly diagram. The schematic shows how components are connected electrically; the assembly diagram shows how components are arranged and interconnected physically.

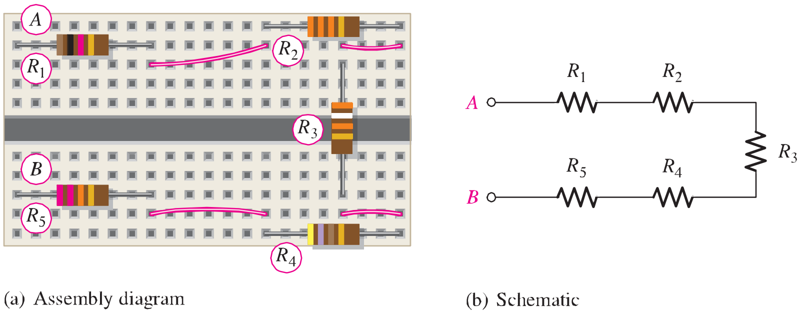

(a) Assembly diagram

(b) Schematic

FIGURE 4–4

RELATED PROBLEM*

(a) Show how you would rewire the circuit board in Figure 4–4(a) so that all the odd-numbered resistors come first starting at point *A* followed by the even-numbered ones.

(b) Determine the value of each resistor.

*Answers are at the end of the chapter.

EXAMPLE 4–2

Describe how the resistors on the printed circuit (PC) board in Figure 4–5 are related electrically. Determine the resistance value of each resistor.

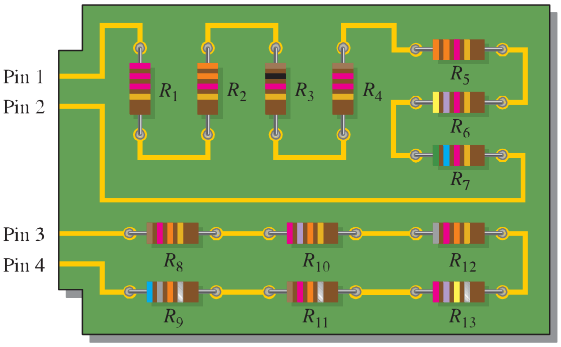

FIGURE 4–5

SOLUTION

Resistors R_1 through R_7 are in series with each other. This series combination is connected between pins 1 and 2 on the PC board.

Resistors R_8 through R_{13} are in series with each other. This series combination is connected between pins 3 and 4 on the PC board.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The values of the resistors are $R_1 = 2.2\ \text{k}\Omega$, $R_2 = 3.3\ \text{k}\Omega$, $R_3 = 1.0\ \text{k}\Omega$, $R_4 = 1.2\ \text{k}\Omega$, $R_5 = 3.3\ \text{k}\Omega$, $R_6 = 4.7\ \text{k}\Omega$, $R_7 = 5.6\ \text{k}\Omega$, $R_8 = 12\ \text{k}\Omega$, $R_9 = 68\ \text{k}\Omega$, $R_{10} = 27\ \text{k}\Omega$, $R_{11} = 12\ \text{k}\Omega$, $R_{12} = 82\ \text{k}\Omega$, and $R_{13} = 270\ \text{k}\Omega$.

6238262 2013/12/10 12.27.66.5

RELATED PROBLEM

How is the circuit in Figure 4–5 changed when pin 2 and pin 3 are connected?

SECTION 4–1 CHECKUP*

1. How are the resistors connected in a series circuit?
2. How can you identify a series circuit?
3. Complete the schematics for the circuits in each part of Figure 4–6 by connecting the resistors in series in numerical order from *A* to *B*.
4. Now connect each group of series resistors in Figure 4–6 in series.

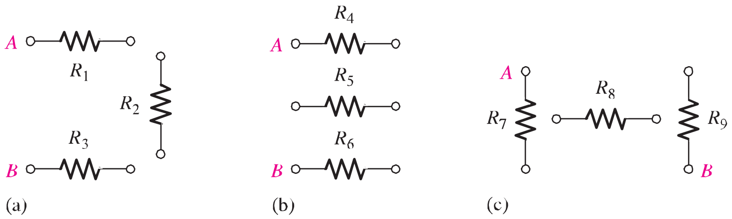

FIGURE 4–6

*Answers are at the end of the chapter.

4–2 TOTAL SERIES RESISTANCE

The total resistance of a series circuit is equal to the sum of the resistances of each individual resistor.

After completing this section, you should be able to

- **Determine total series resistance**
  - **Explain why resistance values add when resistors are connected in series**
  - **Apply the series resistance formula**

Series Resistor Values Add

When resistors are connected in series, the resistor values add because each resistor offers opposition to the current in direct proportion to its resistance. A greater number of resistors connected in series creates more opposition to current. More opposition to current implies a higher value of resistance. Thus, for every resistor that is added in series, the total resistance increases.

Figure 4–7 illustrates how series resistances add together to increase the total resistance. Part (a) has a single 10 Ω resistor. Part (b) shows another 10 Ω resistor connected in

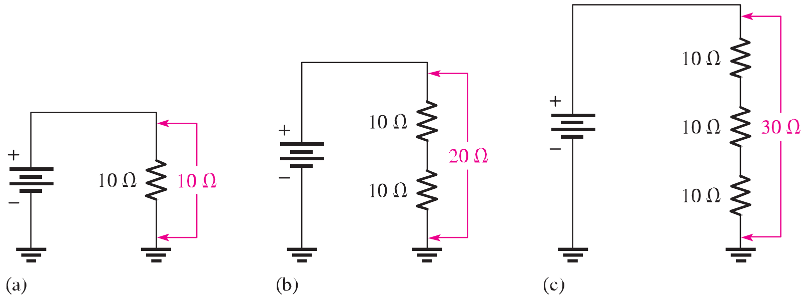

FIGURE 4–7 **Total resistance increases with each additional series resistor. The ground symbol used here was introduced in Section 2–6.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

series with the first one, making a total resistance of 20 Ω. If a third 10 Ω resistor is connected in series with the first two, as shown in part (c), the total resistance becomes 30 Ω.

Series Resistance Formula

For any number of individual resistors connected in series, the total resistance is the sum of each of the individual values.

$$\boldsymbol{R_T = R_1 + R_2 + R_3 + \cdots + R_n} \qquad (4\text{–}1)$$

where R_T is the total resistance and R_n is the last resistor in the series string (n can be any positive integer equal to the number of resistors in series). For example, if there are four resistors in series ($n = 4$), the total resistance formula is

$$R_T = R_1 + R_2 + R_3 + R_4$$

If there are six resistors in series ($n = 6$), the total resistance formula is

$$R_T = R_1 + R_2 + R_3 + R_4 + R_5 + R_6$$

To illustrate the calculation of total series resistance, let's determine R_T of the circuit of Figure 4–8, where V_S is the source voltage. This circuit has five resistors in series. To find the total resistance, simply add the values.

$$R_T = 56\ \Omega + 100\ \Omega + 27\ \Omega + 10\ \Omega + 47\ \Omega = 240\ \Omega$$

Note in Figure 4–8 that the order in which the resistances are added does not matter. Also, you can physically change the positions of the resistors in the circuit without affecting the total resistance or the current.

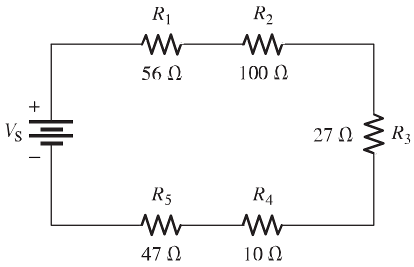

FIGURE 4–8 **Example of five resistors in series. V_S stands for source voltage.**

EXAMPLE 4–3

Connect the resistors on the protoboard in Figure 4–9 in series, and determine the total resistance, R_T, from the color codes.

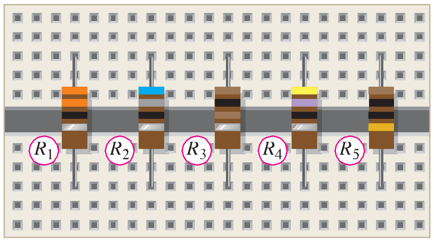

FIGURE 4–9

SOLUTION

The resistors are connected as shown in Figure 4–10. Find the total resistance by adding all the values.

$$R_T = R_1 + R_2 + R_3 + R_4 + R_5 = 33\ \Omega + 68\ \Omega + 100\ \Omega + 47\ \Omega + 10\ \Omega = \mathbf{258\ \Omega}$$

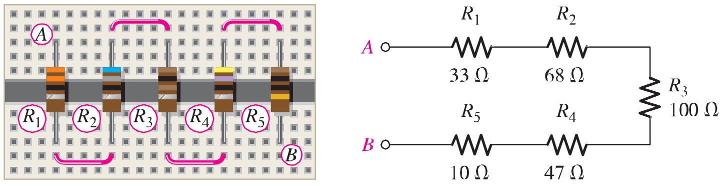

(a) Circuit assembly

(b) Schematic

FIGURE 4–10

RELATED PROBLEM

Determine the total resistance in Figure 4–10(a) if the positions of R_2 and R_4 are interchanged.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 4–4

Calculate the total resistance, R_T, for each circuit in Figure 4–11.

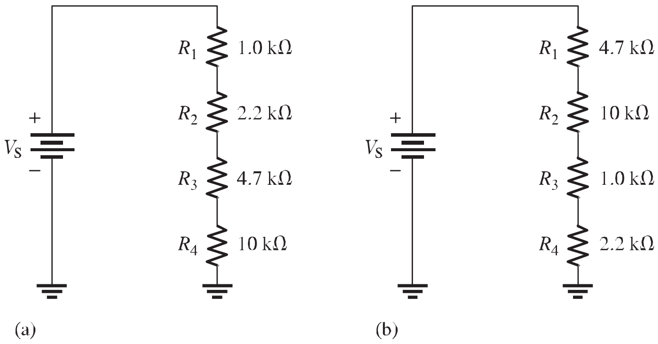

FIGURE 4–11

SOLUTION

For circuit (a),

$$R_T = 1.0\,\text{k}\Omega + 2.2\,\text{k}\Omega + 4.7\,\text{k}\Omega + 10\,\text{k}\Omega = \mathbf{17.9\,k\Omega}$$

For circuit (b),

$$R_T = 4.7\,\text{k}\Omega + 10\,\text{k}\Omega + 1.0\,\text{k}\Omega + 2.2\,\text{k}\Omega = \mathbf{17.9\,k\Omega}$$

Notice that the total resistance does not depend on the position of the resistors. Both circuits are identical in terms of total resistance.

RELATED PROBLEM

What is the total resistance for the following series resistors: 1.0 kΩ, 2.2 kΩ, 3.3 kΩ, and 5.6 kΩ?

EXAMPLE 4–5

Determine the value of R_4 in the circuit of Figure 4–12.

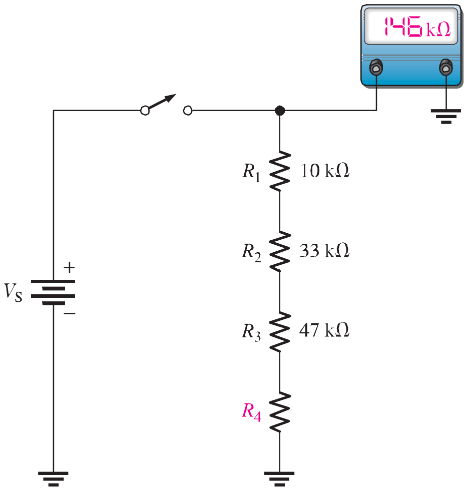

FIGURE 4–12

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

From the ohmmeter reading, $R_T = 146\ \text{k}\Omega$.

$$R_T = R_1 + R_2 + R_3 + R_4$$

Solving for R_4 yields

$$R_4 = R_T - (R_1 + R_2 + R_3) = 146\ \text{k}\Omega - (10\ \text{k}\Omega + 33\ \text{k}\Omega + 47\ \text{k}\Omega) = \mathbf{56\ k\Omega}$$

RELATED PROBLEM

Determine the value of R_4 in Figure 4–12 if the ohmmeter reading is 112 kΩ.

Equal-Value Series Resistors

When a circuit has more than one resistor of the same value in series, there is a shortcut method to obtain the total resistance: Simply multiply the resistance value of the resistors having the same value by the number of equal-value resistors that are in series. This method is essentially the same as adding the values. For example, five 100 Ω resistors in series have an R_T of $5(100\ \Omega) = 500\ \Omega$. In general, the formula is expressed as

$$\boldsymbol{R_T = nR} \tag{4–2}$$

where n is the number of equal-value resistors and R is the resistance value.

EXAMPLE 4–6

Find the R_T of eight 22 Ω resistors in series.

SOLUTION

Find R_T by adding the values.

$$\begin{aligned} R_T &= 22\ \Omega + 22\ \Omega + 22\ \Omega + 22\ \Omega + 22\ \Omega + 22\ \Omega + 22\ \Omega + 22\ \Omega \\ &= \mathbf{176\ \Omega} \end{aligned}$$

However, it is much easier to multiply.

$$R_T = 8(22\ \Omega) = \mathbf{176\ \Omega}$$

RELATED PROBLEM

Find R_T for three 1.0 kΩ and two 680 Ω resistors in series.

SYSTEM EXAMPLE 4–1

A MOTOR STARTER

Shunt dc motors (discussed in Section 7–7) are a type of motor that has a torque-speed characteristic that can be varied over a wide range while still retaining high efficiency. For variable loads, such as an electric hoist motor, a dc shunt motor has an advantage of allowing control of both speed and torque. Speed is controlled by voltage; torque is controlled by current.

The shunt motor has two interacting electromagnets: one is called the field coil and the other is called the rotor, which is a rotating electromagnet. If full voltage is applied without a current-limiting device to the motor, the initial current can be very high, which runs the risk of burning out the motor or overloading the line and blowing a fuse or circuit breaker. To avoid this, a shunt dc motor should be started by limiting current with resistors

6238262 2013/12/10 12.27.66.5

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

(or other means) until the motor begins to speed up. After that, the resistance can be dropped out to adjust speed and torque to the load requirement.

A motor starter circuit, such as the one shown in Figure 4–13, provides the means to limit current. The resistors are effectively removed as required so that there is less resistance as the rotary switch is moved clockwise. For this example, a manual rotary switch is shown, but automated methods are also used. To start the motor, the rotary switch is moved from the off position to position 1. This applies dc to the field coil and the series solenoid (an electromagnet, which keeps the switch in place). If power drops out for any reason, the solenoid releases its pull and the switch is pulled by the spring back to the off position. This prevents the motor from restarting with no resistance in place.

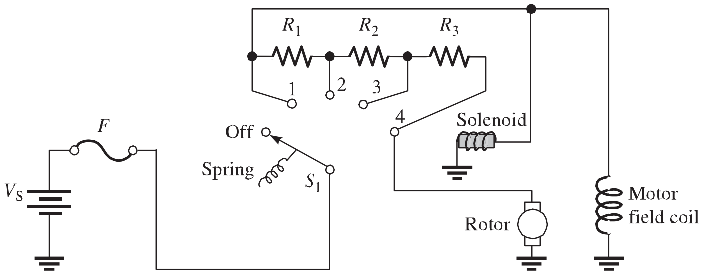

FIGURE 4–13 **A manual motor starter for a dc shunt motor.**

SECTION 4–2 CHECKUP

1. Calculate R_T between points A and B for each circuit in Figure 4–14.
2. The following resistors are in series: one 100 Ω, two 47 Ω, four 12 Ω, and one 330 Ω. What is the total resistance?
3. Suppose that you have one resistor for each of the following values: 1.0 kΩ, 2.7 kΩ, 3.3 kΩ, and 1.8 kΩ. To get a total resistance of 10 kΩ, you need one more resistor. What should its value be?
4. What is the R_T for twelve 47 Ω resistors in series?

FIGURE 4–14

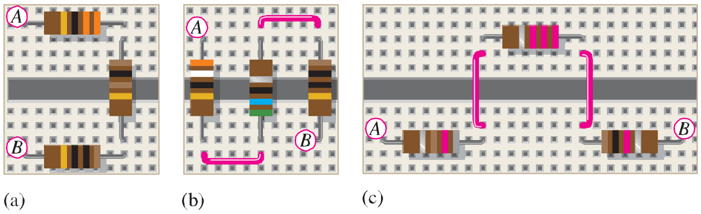

(a) (b) (c)

4–3 CURRENT IN A SERIES CIRCUIT

The current is the same through all points in a series circuit. The current through each resistor in a series circuit is the same as the current through all the other resistors that are in series with it.

After completing this section, you should be able to

- **Determine the current throughout a series circuit**
  - **Show that the current is the same at all points in a series circuit**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Figure 4–15 shows three resistors connected in series to a voltage source. *At any point in this circuit, the current into that point must equal the current out of that point,* as illustrated by the current directional arrows. Notice also that the current out of each of the resistors must equal the current in because there is no place where part of the current can branch off and go somewhere else. Therefore, the current in each section of the circuit is the same as the current in all other sections. It has only one path going from the negative (−) side of the source to the positive (+) side.

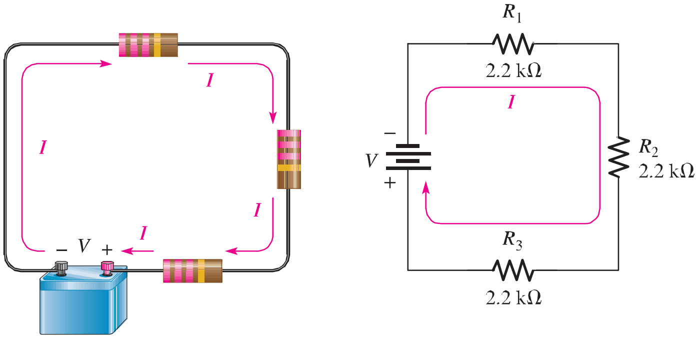

(a) Pictorial (b) Schematic

FIGURE 4–15 **Current entering any point in a series circuit is the same as the current leaving that point.**

In Figure 4–16, the battery supplies 1.82 mA of current to the series resistors. There are 1.82 mA from the battery's negative terminal. As shown, the same current is measured at several points in the series circuit.

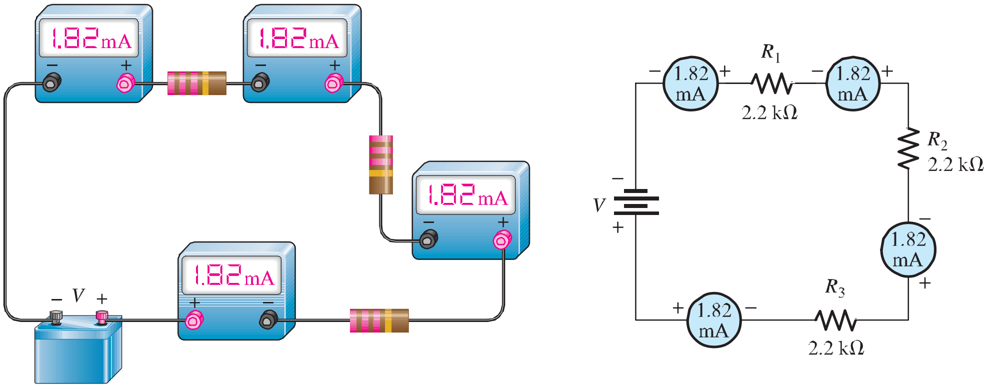

(a) Pictorial (b) Schematic

FIGURE 4–16 **Current is the same at all points in a series circuit.**

SECTION 4–3 CHECKUP

1. What statement can you make about the amount of current at any point in a series circuit?
2. In a series circuit with a 100 Ω and a 47 Ω resistor, there are 20 mA through the 100 Ω resistor. How much current is there through the 47 Ω resistor?
3. A milliammeter is connected between points *A* and *B* in Figure 4–17. It measures 50 mA. If you move the meter and connect it between points *C* and *D*, how much current will it indicate? Between *E* and *F*?
4. In Figure 4–18, how much current does ammeter 1 indicate? Ammeter 2?

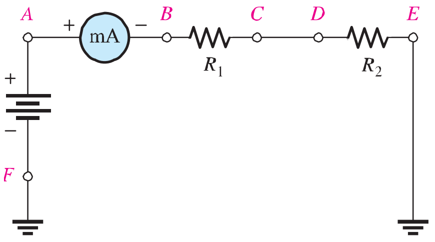

FIGURE 4–17

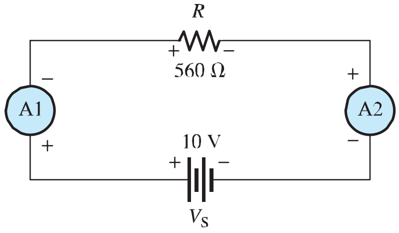

FIGURE 4–18

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

4–4 APPLICATION OF OHM'S LAW

The basic concepts of series circuits and Ohm's law can be applied to series circuit analysis.

After completing this section, you should be able to

- **Apply Ohm's law in series circuits**
  - **Find the current in a series circuit**
  - **Find the voltage across each resistor in series**

The following are key points to remember when you analyze series circuits:

1. Current through any of the series resistors is the same as the total current.
2. If you know the total applied voltage and the total resistance, you can determine the total current by Ohm's law.

$$I_T = \frac{V_T}{R_T}$$

3. If you know the voltage drop across one of the series resistors, R_x, you can determine the total current by Ohm's law.

$$I_T = \frac{V_x}{R_x}$$

4. If you know the total current, you can find the voltage drop across any of the series resistors by Ohm's law.

$$V_x = I_T R_x$$

5. The polarity of a voltage drop across a resistor is positive at the end of the resistor that is closest to the positive terminal of the voltage source.
6. The current through a resistor is defined to be in a direction from the negative end of the resistor to the positive end.
7. An open in a series circuit prevents current; and, therefore, there is zero voltage drop across each series resistor. The total voltage appears across the points between which there is an open.

Now let's look at several examples that use Ohm's law for series circuit analysis.

EXAMPLE 4–7

Find the current in the circuit of Figure 4–19.

FIGURE 4–19

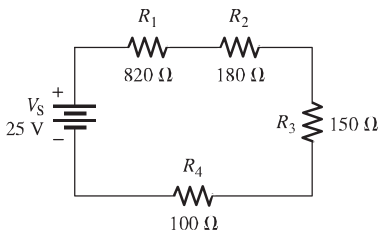

SOLUTION

The current is determined by the source voltage and the total resistance. First, calculate the total resistance.

6238262 2013/12/10 12.27.66.5

$$R_T = R_1 + R_2 + R_3 + R_4 = 820\ \Omega + 180\ \Omega + 150\ \Omega + 100\ \Omega = 1.25\ \text{k}\Omega$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Next, use Ohm's law to calculate the current.

$$I = \frac{V_S}{R_T} = \frac{25\text{ V}}{1.25\text{ k}\Omega} = \mathbf{20\text{ mA}}$$

Remember, the same current exists at all points in the circuit. Thus, each resistor has 20 mA through it.

RELATED PROBLEM

What is the current in the circuit of Figure 4–19 if R_4 is changed to 200 Ω?

MULTISIM

Open Multisim file E04-07; files are found at www.pearsonhighered.com/floyd. Connect the multimeter and verify the value of current calculated in this example. Change the value of R_4 to 200 Ω and verify the value of current calculated in the related problem.

EXAMPLE 4–8

The current in the circuit of Figure 4–20 is 1 mA. For this amount of current, what must the source voltage V_S be?

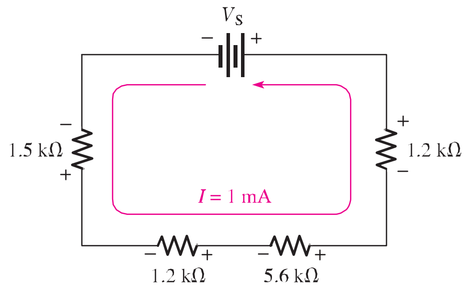

FIGURE 4–20

SOLUTION

In order to calculate V_S, first determine R_T.

$$R_T = 1.2\text{ k}\Omega + 5.6\text{ k}\Omega + 1.2\text{ k}\Omega + 1.5\text{ k}\Omega = 9.5\text{ k}\Omega$$

Now use Ohm's law to determine V_S.

$$V_S = IR_T = (1\text{ mA})(9.5\text{ k}\Omega) = \mathbf{9.5\text{ V}}$$

RELATED PROBLEM

If the 5.6 kΩ resistor is changed to 3.9 kΩ, what value of V_S is necessary to keep the current at 1 mA?

MULTISIM

Open Multisim file E04-08. Verify that the calculated source voltage establishes the current in Figure 4–20. Verify the value of V_S determined in the related problem.

EXAMPLE 4–9

Calculate the voltage drop across each resistor in Figure 4–21, and find the value of V_S. To what value can V_S be raised if the current is to be limited to 5 mA?

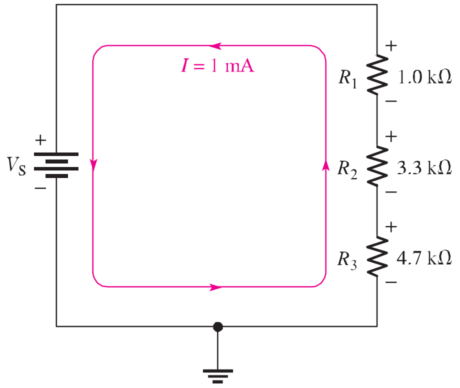

FIGURE 4–21

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

By Ohm's law, the voltage drop across each resistor is equal to its resistance multiplied by the current through it. Use the Ohm's law formula $V = IR$ to determine the voltage drop across each of the resistors. Keep in mind that there is the same current through each series resistor. The voltage drop across R_1 (designated V_1) is

$$V_1 = IR_1 = (1\text{ mA})(1.0\text{ k}\Omega) = \mathbf{1\ V}$$

The voltage drop across R_2 is

$$V_2 = IR_2 = (1\text{ mA})(3.3\text{ k}\Omega) = \mathbf{3.3\ V}$$

The voltage drop across R_3 is

$$V_3 = IR_3 = (1\text{ mA})(4.7\text{ k}\Omega) = \mathbf{4.7\ V}$$

To calculate the value of V_S, first determine the total resistance.

$$R_T = 1.0\text{ k}\Omega + 3.3\text{ k}\Omega + 4.7\text{ k}\Omega = 9\text{ k}\Omega$$

The source voltage V_S is equal to the current times the total resistance.

$$V_S = IR_T = (1\text{ mA})(9\text{ k}\Omega) = \mathbf{9\ V}$$

Notice that if you add the voltage drops of the resistors, they total 9 V, which is the same as the source voltage.

V_S can be increased to a maximum value where $I = 5$ mA. Therefore the maximum value of V_S is

$$V_{S(max)} = IR_T = (5\text{ mA})(9\text{ k}\Omega) = \mathbf{45\ V}$$

MULTISIM

Open Multisim file E04-09. Using the multimeter, verify that the voltages across the resistors agree with the values calculated in this example.

RELATED PROBLEM

Repeat the calculations for V_1, V_2, V_3, V_S, and $V_{S(max)}$ if $R_3 = 2.2\text{ k}\Omega$ and I is maintained at 1 mA.

Indicator Lights

Indicator lights are a common way to ascertain a specific condition of a machine or relay, such as a motor starter. Incandescent indicator lights are available for common voltages used in industrial plants (120 V, 240 V, 480 V, and 600 V). In small systems, low voltages are common, and LEDs are a good choice for indicating a condition (such as a power supply is *on*). A typical indicating LED will drop about 2 V and has an *on* current of 20 mA. To prevent current from exceeding the specification, a current-limiting resistor is used in series (in-line) with the LED. Example 4–10 shows how to calculate the size of current limiting resistors when a variable brightness is needed to account for different ambient conditions.

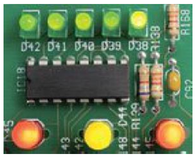

(yukosourov/Fotolia.com)

SYSTEM NOTE

EXAMPLE 4–10

LEDs are commonly used in systems to indicate a certain condition. The LED indicators have specific current requirements. Too little current, and the LED is too dim; too much and it burns out. In some systems, a control is needed to adjust the light intensity based on ambient conditions. Figure 4–22 shows this basic

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

application, where the red LED is used as a system indicator light. In this case, a rheostat and series fixed resistor limit current to the LED. The rheostat is included to dim the LED, depending on ambient conditions. We will focus on just these two current-limiting resistors.

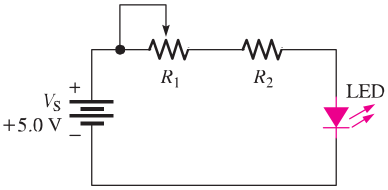

FIGURE 4–22

A red LED will always have a voltage across it of about +1.7 V when it is on and working within its normal operating range. The remaining voltage from the power supply will be across the two series resistors. Together, the rheostat and the fixed resistor will have a total of 3.3 V across them.

Assume you want current in the LED to range from a minimum of 2.5 mA (dim) to a maximum of 10 mA (bright). What values of R_1 and R_2 would you choose to accomplish this?

SOLUTION

Start with the brightest condition when the resistance of the rheostat is adjusted for 0 Ω. In this case, there will be no voltage across R_1 and 3.3 V will be across R_2. Because it is a series circuit, the same current is in R_2 as the LED. Therefore,

$$R_2 = \frac{V}{I} = \frac{3.3\text{ V}}{10\text{ mA}} = \mathbf{330\ \Omega}$$

Now determine the total resistance required to limit the current to 2.5 mA. The total resistance is $R_T = R_1 + R_2$, and the voltage drop across R_T is 3.3 V. From Ohm's law,

$$R_T = \frac{V}{I} = \frac{3.3\text{ V}}{2.5\text{ mA}} = 1.32\text{ k}\Omega$$

To find R_1, subtract the value of R_2 from the total resistance.

$$R_1 = R_T - R_2 = 1.32\text{ k}\Omega - 330\ \Omega = 990\ \Omega$$

Choose a **1.0 kΩ** rheostat as the nearest standard value.

RELATED PROBLEM

What is the value of R_2 if the highest current is 12 mA?

SECTION 4–4 CHECKUP

1. A 6 V battery is connected across three 100 Ω resistors in series. What is the current through each resistor?
2. How much voltage is required to produce 5 mA through the circuit of Figure 4–23?

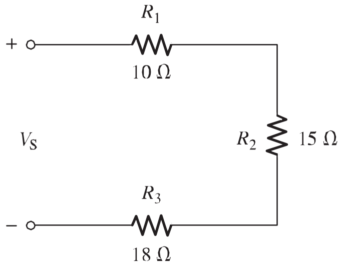

FIGURE 4–23

3. How much voltage is dropped across each resistor in Figure 4–23 when the current is 5 mA?
4. Four equal-value resistors are connected in series with a 5 V source. The measured current is 4.63 mA. What is the value of each resistor?
5. What size should a series current-limiting resistor be to limit the current to a red LED to 10 mA from a 3.0 V source? Assume the LED drops 1.7 V.

6238262 2013/12/10 12 27 66.5

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

4–5 VOLTAGE SOURCES IN SERIES

Recall that a voltage source is an energy source that provides a constant voltage to a load. Batteries and power supplies are practical examples of dc voltage sources. When two or more voltage sources are in series, the total voltage is equal to the algebraic sum of the individual source voltages.

After completing this section, you should be able to

- **Determine the total effect of voltage sources connected in series**
  - **Determine the total voltage of series sources with the same polarities**
  - **Determine the total voltage of series sources with opposite polarities**

When batteries are placed in a flashlight, they are connected in a **series-aiding** arrangement to produce a larger voltage, as illustrated in Figure 4–24. In this example, three 1.5 V batteries are placed in series to produce a total voltage ($V_{S(tot)}$).

$$V_{S(tot)} = V_{S1} + V_{S2} + V_{S3} = 1.5\text{ V} + 1.5\text{ V} + 1.5\text{ V} = 4.5\text{ V}$$

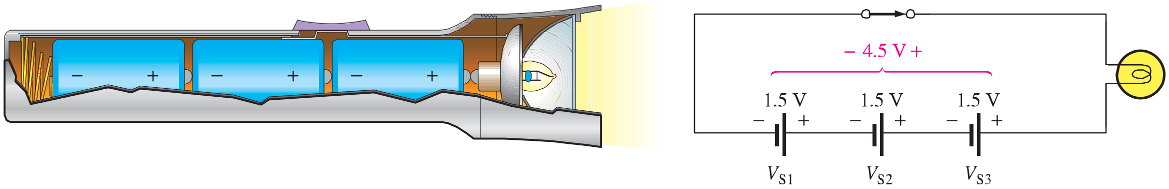

(a) Flashlight with series batteries

(b) Schematic of flashlight circuit

FIGURE 4–24 **Example of series-aiding voltage sources.**

HANDS ON TIP
When you replace batteries in a flashlight or other device, it is best to use all of the same type of battery and not mix old batteries with new batteries. In particular, do not mix alkaline with nonalkaline batteries. Incorrect use of batteries can cause hydrogen gas to build inside the battery and cause the casing to rupture. Worse, the mixing of hydrogen gas with oxygen can be an explosion hazard. Cases are reported of flashlights that have exploded in hazardous situations.

Series voltage sources (batteries in this case) are added when their polarities are in the same direction, or series-aiding, and are subtracted when their polarities are in opposite directions, or **series-opposing**. For example, if one of the batteries in the flashlight is turned around, as indicated in the schematic of Figure 4–25, its voltage subtracts because it has a negative value and reduces the total voltage.

$$V_{S(tot)} = V_{S1} - V_{S2} + V_{S3} = 1.5\text{ V} - 1.5\text{ V} + 1.5\text{ V} = 1.5\text{ V}$$

There is no valid reason for reversing a battery, and it can lead to high current and reduced life if it happens accidentally. However, an opposing source does occur naturally in the case of motors. An opposing voltage to the source voltage is generated within the motor, which reduces the current as you will see in Section 7–7.

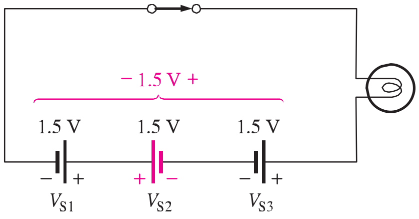

FIGURE 4–25 **When batteries are connected in opposite directions, the total voltage is the algebraic sum of the voltages. As noted, this is not a valid configuration for batteries.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 4–11

What is the total source voltage ($V_{S(tot)}$) in Figure 4–26?

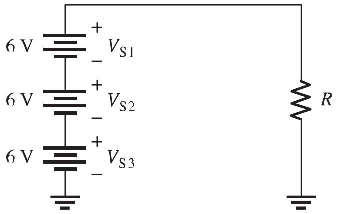

FIGURE 4–26

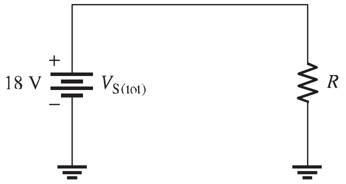

FIGURE 4–27

SOLUTION

The polarity of each source is the same (the sources are connected in the same direction in the circuit). Therefore, add the three voltages to get the total.

$$V_{S(tot)} = V_{S1} + V_{S2} + V_{S3} = 6\text{ V} + 6\text{ V} + 6\text{ V} = \mathbf{18\text{ V}}$$

The three individual sources can be replaced by a single equivalent source of 18 V with its polarity as shown in Figure 4–27.

RELATED PROBLEM

If the battery V_{S3} in Figure 4–26 is accidentally installed backwards, what is the total voltage?

MULTISIM

Open Multisim file E04-11. Verify the total source voltage. Repeat for the related problem.

SYSTEM EXAMPLE 4–2

SOLAR PANELS IN SERIES

Solar panels (modules) are composed of many solar cells in an array, wired to produce a certain voltage in full sunlight. Within the module, each solar cell produces only about 0.5 V in full sun. Therefore, an 18 V panel requires at least 36 solar cells wired in series, and a 24 V panel requires at least 48 solar cells in series. For battery charging, this is sufficient voltage for most applications. In general, module voltages range from 16.5 V to 72.3 V as specified under standard test conditions (STC), which allow comparisons between modules. Actual operating conditions are different.

Figure 4–28 is a block diagram of a basic solar electric system that is independent of the utility grid (in this case). It has a battery bank for storing energy when there is no solar

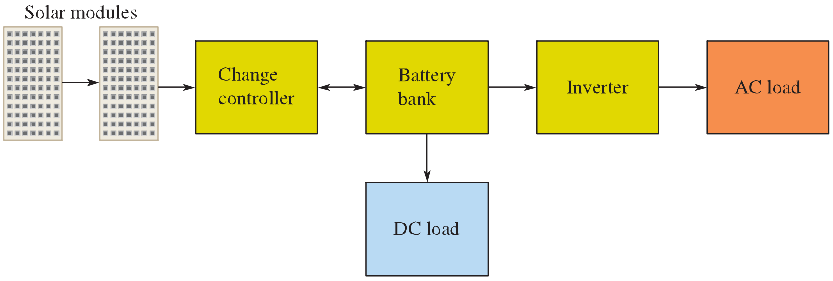

FIGURE 4–28 **Typical small solar electric system.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

input. The solar modules convert sunlight to dc electrical power, which is sent to a charge controller. The charge controller regulates charging current to prevent overcharging the batteries or draining the batteries back through the solar modules when they are needed for the load. The inverter converts the dc output to ac for powering small appliances. The inverter is a basic one in this system rather than a more expensive grid-tied inverter. Grid-tied inverters are special inverters that are required for any renewable energy system that is connected to a utility grid.

For most systems, a number of modules are connected together to provide a high voltage dc for converting to ac. The voltage rating of the module and the inverter voltage ratings determine how many modules can be put in series. All components in the system (including modules, wiring, and inverters) must be rated for the maximum system voltage. In the United States, the maximum permitted voltage is 600 V for most installations.

For this example, assume a single panel has a maximum voltage of 34.6 V and is rated for 250 W. The number of these panels that can be installed in series depends on the maximum permitted voltage. For a permitted voltage of 600 V, 17 panels in series is the maximum that can be connected without exceeding 600 V. In this case, the maximum voltage will be 588 V. The advantage of the series connection for these panels is that the higher voltage is more efficient and can use smaller wire because the current is the same going into and from each panel. The maximum rated power from the array is 17 panels × 250 W/panel = 4250 W.

EXAMPLE 4–12

Many circuits use positive and negative supply voltages. A dual-power supply will normally have two independent outputs, such as those shown in Figure 4–29. Show how to connect the two 12 V outputs from the power supply so that there is both a positive and a negative output and the two sources are series-aiding.

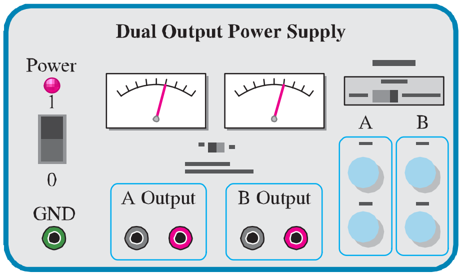

FIGURE 4–29

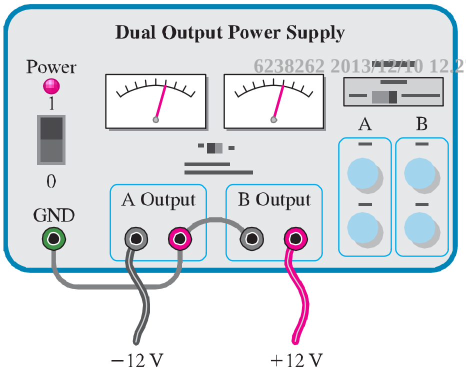

6238262 2013/12/10 12.27.66.5

FIGURE 4–30

SOLUTION

See Figure 4–30. The positive output of one supply is connected to the negative output of the second supply. This makes the connection series-aiding. The ground terminal is connected to the common point between the supplies, forcing the *A* output to be below ground (meaning it is a negative voltage with respect to ground) and the *B* output to be above ground (positive with respect to ground).

RELATED PROBLEM

Draw the schematic of the setup in Figure 4–30.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 4–5 CHECKUP

1. How many 12 V batteries must be connected in series to produce 60 V? Draw a schematic that shows the battery connections.
2. Four 1.5 V flashlight batteries are connected in series-aiding. What is the total voltage across the bulb?
3. The resistive circuit in Figure 4–31 is used to bias a transistor amplifier. Show how to connect the resistors to two 15 V power supplies in order to get 30 V across the series resistors.

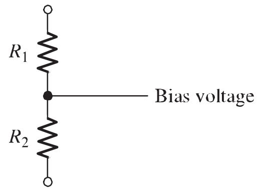

FIGURE 4–31

4. Determine the total source voltage for the circuit in Figure 4–32.

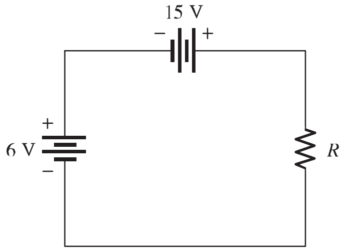

FIGURE 4–32

5. Assume that in a four-cell flashlight one of the four 1.5 V batteries was accidentally installed in the wrong direction. In this case, what voltage would be across the bulb when the light was turned on?

4–6 KIRCHHOFF'S VOLTAGE LAW

Kirchhoff's voltage law is a fundamental circuit law that states that the algebraic sum of all the voltages around a single closed path is zero or, in other words, the sum of the voltage drops equals the total source voltage.

After completing this section, you should be able to

- **Apply Kirchhoff's voltage law**
  - **State Kirchhoff's voltage law**
  - **Determine the source voltage by adding the voltage drops**
  - **Determine an unknown voltage drop**

In an electric circuit, the voltages across the resistors (voltage drops) *always* have polarities opposite to the source voltage polarity. For example, in Figure 4–33, follow a counterclockwise loop around the circuit. Note that the source polarity is plus-to-minus and each voltage drop is minus-to-plus.

In Figure 4–33, the current is out of the negative side of the source and through the resistors as the arrows indicate. The current is into the negative side of each resistor and out the positive side. As you learned in Chapter 3, when electrons flow through a resistor, they lose energy and are therefore at a lower energy level when they emerge. The lower energy side is less negative (more positive) than the higher energy side. The drop in energy level across a resistor creates a potential difference, or voltage drop, with a minus-to-plus polarity in the direction of the current.

The voltage from point A to point B in the circuit of Figure 4–33 is the source voltage, V_S. Also, the voltage from A to B is the sum of the series resistor voltage drops. Therefore, the source voltage is equal to the sum of the three voltage drops, as stated by **Kirchhoff's voltage law**.

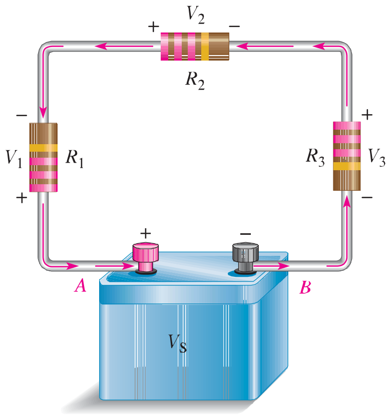

FIGURE 4–33 **Illustration of voltage polarities in a closed-loop circuit.**

The sum of all the voltage drops around a single closed path in a circuit is equal to the total source voltage in that closed path.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

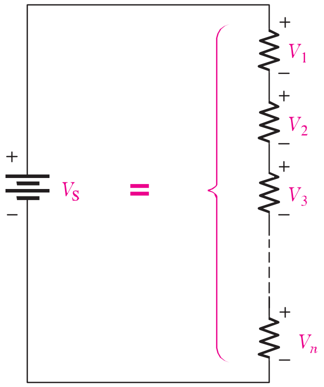

FIGURE 4–34 **Sum of n voltage drops equals the source voltage.**

Kirchhoff's voltage law applied to a series circuit is illustrated in Figure 4–34. For this case, Kirchhoff's voltage law can be expressed by the equation

$$V_S = V_1 + V_2 + V_3 + \cdots + V_n \qquad (4\text{–}3)$$

where the subscript n represents the number of voltage drops.

If all the voltage drops around a closed path are added and then this total is subtracted from the source voltage, the result is zero. This result occurs because the sum of the voltage drops in a series circuit always equals the source voltage.

Equation 4–3 can be applied to any circuit that has a voltage source and resistors or other loads. The key idea is that a single closed path is followed from an arbitrary starting point and back to this point. In a series circuit, the closed path will always include a voltage source and one or more resistors or other loads. In this case, the source represents a voltage *rise,* and each load represents a voltage *drop.* Another way of stating Kirchhoff's voltage law for a series circuit is that the sum of all voltage rises is equal to the sum of all voltage drops.

You can verify Kirchhoff's voltage law in a series circuit by connecting the circuit and measuring each resistor voltage and the source voltage, as illustrated in Figure 4–35. When the resistor voltages are added together, their sum will equal the source voltage. Any number of resistors can be used.

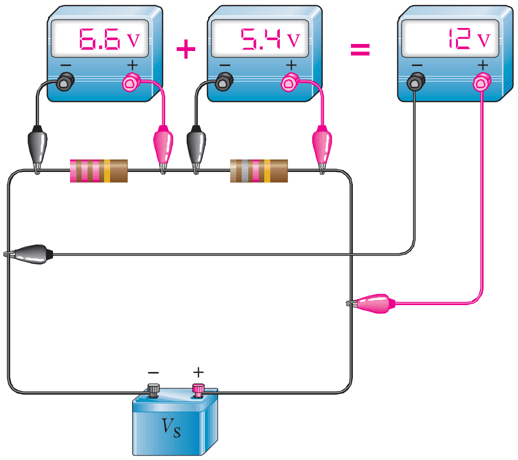

FIGURE 4–35 **Illustration of a verification of Kirchhoff's voltage law.**

Kirchhoff's voltage law was developed here for a series circuit, but the concept can be applied to any circuit. In complex circuits, you can still write Kirchhoff's voltage law, but there are cases where there is no source voltage in a given closed loop. Even so, Kirchhoff's voltage law still applies. This leads to a more general statement of Kirchhoff's voltage law:

> **The algebraic sum of the voltages around any closed path in a circuit is equal to zero.**

If a voltage source is present, it is just treated as one of the terms in the summation. It is important to assign the correct algebraic sign to the terms, noting if the voltage is rising or dropping as one traverses the loop. In equation form,

$$V_1 + V_2 + V_3 + \cdots + V_n = 0 \qquad (4\text{–}4)$$

Any of the variables in Equation 4–4 can represent a voltage rise or a voltage drop. This more-general form of Kirchhoff's voltage law can be expressed more concisely as

$$\sum_{i=1}^{n} V_i = 0$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

This equation is a shorthand way of writing Equation 4–4. The capital sigma ($\sum$) means to add the voltages from the first ($i = 1$) to the last ($i = n$).

This form of Kirchhoff's voltage law can be applied to circuits other than the series circuits, provided you follow a single closed path. In applying Equation 4–4, you need to assign an algebraic sign to each voltage in the path. In circuits other than series circuits, a voltage across a resistor can appear as either a rise or a drop, depending on the path chosen. As you traverse the path, you will need to write voltage rises and voltage drops in a consistent manner. For series circuits such as given in Figure 4–34, the main idea is that the source voltage (rise) is equal to the sum of the voltages (drops) across the loads (resistors).

6238262 2013/12/10 12.27.66.5

EXAMPLE 4–13

Determine the source voltage, V_S, in Figure 4–36 where the two voltage drops are given.

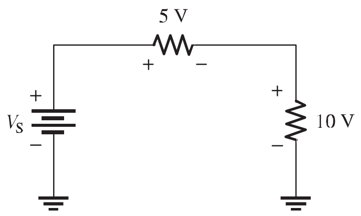

FIGURE 4–36

SOLUTION

By Kirchhoff's voltage law (Eq. 4–3), the source voltage (applied voltage) must equal the sum of the voltage drops. Adding the voltage drops gives the value of the source voltage.

$$V_S = 5\text{ V} + 10\text{ V} = \mathbf{15\text{ V}}$$

RELATED PROBLEM

If V_S is increased to 30 V in Figure 4–36, what are the two voltage drops?

MULTISIM

Open Multisim file E04-13. Verify that the voltage drops equal to V_S. Repeat for the related problem.

EXAMPLE 4–14

Determine the unknown voltage drop, V_3, in Figure 4–37.

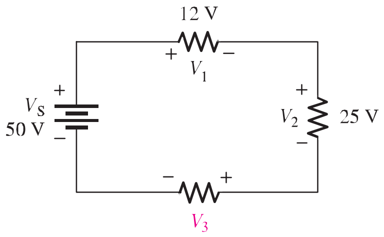

FIGURE 4–37

SOLUTION

By Kirchhoff's voltage law (Equation 4–4), the algebraic sum of the voltages around the circuit is zero.

$$V_1 + V_2 + V_3 + V_4 = 0$$

Substitute $-V_S$ for V_4. Then,

$$V_1 + V_2 + V_3 - V_S = 0$$

Solve for V_3.

$$V_3 = V_S - V_1 - V_2 = 50\text{ V} - 12\text{ V} - 25\text{ V} = \mathbf{13\text{ V}}$$

The voltage drop, V_3, across R_3 is 13 V, and its polarity is as shown in Figure 4–37.

RELATED PROBLEM

Determine V_3 in Figure 4–37 if the source is changed to 25 V.

MULTISIM

Open Multisim file E04-14. Verify that V_3 agrees with the calculated value. Repeat for the related problem.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 4–15

Find the value of R_4 in Figure 4–38.

FIGURE 4–38

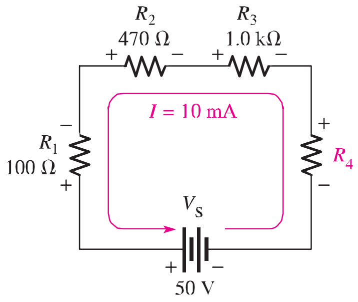

SOLUTION

In this problem you will use both Ohm's law and Kirchhoff's voltage law. First, use Ohm's law to find the voltage drop across each of the known resistors.

$$V_1 = IR_1 = (10\text{ mA})(100\ \Omega) = 1.0\text{ V}$$
$$V_2 = IR_2 = (10\text{ mA})(470\ \Omega) = 4.7\text{ V}$$
$$V_3 = IR_3 = (10\text{ mA})(1.0\text{ k}\Omega) = 10\text{ V}$$

Next, use Kirchhoff's voltage law to find V_4, the voltage drop across the unknown resistor.

$$V_S - V_1 - V_2 - V_3 - V_4 = 0\text{ V}$$
$$50\text{ V} - 1.0\text{ V} - 4.7\text{ V} - 10\text{ V} - V_4 = 0\text{ V}$$
$$34.3\text{ V} - V_4 = 0\text{ V}$$
$$V_4 = 34.3\text{ V}$$

Now that you know V_4, you can use Ohm's law to calculate R_4.

$$R_4 = \frac{V_4}{I} = \frac{34.3\text{ V}}{10\text{ mA}} = \mathbf{3.43\text{ k}\Omega}$$

R_4 is most likely a 3.3 kΩ standard-value resistor because 3.43 kΩ is within a standard tolerance range (±5%) of 3.3 kΩ.

MULTISIM

Open Multisim file E04-15. Verify that the calculated value of R_4 produces the current in Figure 4–38. Repeat for the related problem.

RELATED PROBLEM

Determine the value of R_4 in Figure 4–38 if $V_S = 20$ V and $I = 10$ mA.

SECTION 4–6 CHECKUP

1. State Kirchhoff's voltage law in two ways.
2. A 50 V source is connected to a series resistive circuit. What is the sum of the voltage drops in this circuit?
3. Two equal-value resistors are connected in series across a 10 V battery. What is the voltage drop across each resistor?
4. In a series circuit with a 25 V source, there are three resistors. One voltage drop is 5 V, and the other is 10 V. What is the value of the third voltage drop?
5. The individual voltage drops in a series circuit are 1 V, 3 V, 5 V, 7 V, and 8 V. What is the total voltage applied across the series circuit?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

4–7 VOLTAGE DIVIDERS

A series circuit acts as a voltage divider. The voltage divider is an important application of series circuits.

After completing this section, you should be able to

- **Use a series circuit as a voltage divider**
  - **Apply the voltage-divider formula**
  - **Use the potentiometer as an adjustable voltage divider**
  - **Describe some voltage-divider applications**

A circuit consisting of a series string of resistors connected to a voltage source acts as a **voltage divider**. Figure 4–39(a) shows a circuit with two resistors in series, although there can be any number. As you already know, there are two voltage drops: one across R_1 and one across R_2. These voltage drops are labeled V_1 and V_2, respectively, as indicated in the schematic. Since each resistor has the same current through it, the voltage drops are proportional to the resistance values. For example, if the value of R_2 is twice that of R_1, then the value of V_2 is twice that of V_1.

The total voltage drop around a single closed path divides among the series resistors in amounts directly proportional to the resistance values. The smallest resistance has the least voltage, and the largest resistance has the most voltage ($V = IR$). For example, in Figure 4–39(b), if V_S is 10 V, R_1 is 100 Ω, and R_2 is 200 Ω, then V_1 is one-third the total voltage, or 3.33 V, because R_1 is one-third the total resistance. Likewise, V_2 is two-thirds V_S, or 6.67 V, because R_2 is two-thirds the total resistance.

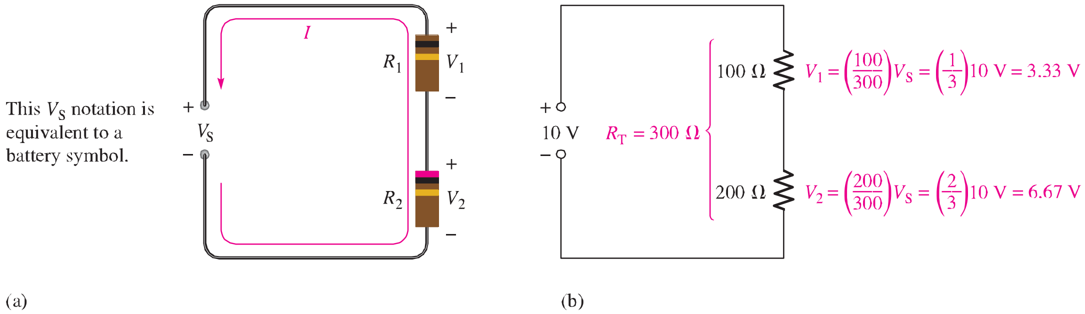
I
R1
V1
+
−
+
VS
−
R2
V2
+
−
+
10 V
−
RT = 300 Ω
100 Ω
200 Ω
V1 = (100/300)VS = (1/3)10 V = 3.33 V
V2 = (200/300)VS = (2/3)10 V = 6.67 V

This V_S notation is equivalent to a battery symbol.

(a)

(b)

FIGURE 4–39 **Example of a two-resistor voltage divider.**

Voltage-Divider Formula

With a few calculations, you can develop a formula for determining how the voltages divide among series resistors. Assume a circuit with n resistors in series as shown in Figure 4–40, where n can be any number.

Let V_x represent the voltage drop across any one of the resistors and R_x represent the number of a particular resistor or combination of resistors. By Ohm's law, you can express the voltage drop across R_x as follows:

$$V_x = IR_x$$

The current through the circuit is equal to the source voltage divided by the total resistance ($I = V_S/R_T$). In the circuit of Figure 4–40, the total resistance is

$$R_T = R_1 + R_2 + R_3 + \cdots + R_n$$

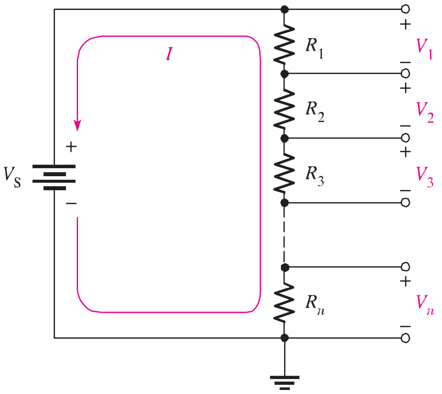

FIGURE 4–40 **Generalized voltage divider with *n* resistors.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

By substitution of V_S/R_T for I in the expression for V_x,

$$V_x = \left(\frac{V_S}{R_T}\right)R_x$$

Rearranging the terms you get

$$\boldsymbol{V_x = \left(\frac{R_x}{R_T}\right)V_S} \qquad (4\text{–}5)$$

Equation 4–5 is the general voltage-divider formula, which can be stated as follows:

> **The voltage drop across any resistor or combination of resistors in a series circuit is equal to the ratio of that resistance value to the total resistance, multiplied by the source voltage.**

Flash Converters

In many systems, an analog signal is converted to digital for processing. This process uses an analog-to-digital converter (ADC). For applications where speed is critical, a flash converter is the fastest type of ADC.

The standard flash converter uses a large voltage-divider string composed of hundreds of resistors to set up a series of reference voltages. For example, an 8-bit flash ADC uses 255 resistors in a divider string. The resistors must be very accurate, so they are carefully controlled within the integrated circuit. The input analog signal is compared to all of the different voltages simultaneously, and the level closest to the signal level is digitized.

SYSTEM NOTE

6238262 2013/12/10 12.27.66.5

EXAMPLE 4–16

Determine V_1 (the voltage across R_1) and V_2 (the voltage across R_2) in the circuit in Figure 4–41.

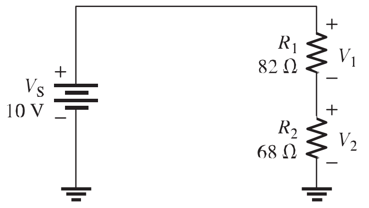

FIGURE 4–41

SOLUTION

To determine V_1, use the voltage-divider formula, $V_x = (R_x/R_T)V_S$, where $x = 1$. The total resistance is

$$R_T = R_1 + R_2 = 82\ \Omega + 68\ \Omega = 150\ \Omega$$

R_1 is 82 Ω and V_S is 10 V. Substitute these values into the voltage-divider formula.

$$V_1 = \left(\frac{R_1}{R_T}\right)V_S = \left(\frac{82\ \Omega}{150\ \Omega}\right)10\ \text{V} = \mathbf{5.47\ V}$$

There are two ways to find the value of V_2: Kirchhoff's voltage law or the voltage-divider formula. If you use Kirchhoff's voltage law ($V_S = V_1 + V_2$), substitute the values for V_S and V_1 and solve for V_2.

$$V_2 = V_S - V_1 = 10\ \text{V} - 5.47\ \text{V} = \mathbf{4.53\ V}$$

To determine V_2, use the voltage-divider formula where $x = 2$.

$$V_2 = \left(\frac{R_2}{R_T}\right)V_S = \left(\frac{68\ \Omega}{150\ \Omega}\right)10\ \text{V} = \mathbf{4.53\ V}$$

MULTISIM

Open Multisim file E04-16. Use the multimeter to verify the calculated values of V_1 and V_2. Repeat for the related problem.

RELATED PROBLEM

Find the voltage drops across R_1 and R_2 if R_2 is changed to 180 Ω in Figure 4–41.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 4–17

Calculate the voltage drop across each resistor in the circuit of Figure 4–42.

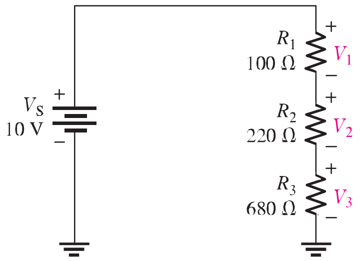

FIGURE 4–42

SOLUTION

Look at the circuit and consider the following: The total resistance is 1000 Ω. Ten percent of the total voltage is across R_1 because it is 10% of the total resistance (100 Ω is 10% of 1000 Ω). Likewise, 22% of the total voltage is dropped across R_2 because it is 22% of the total resistance (220 Ω is 22% of 1000 Ω). Finally, R_3 drops 68% of the total voltage (680 Ω is 68% of 1000 Ω).

Because of the convenient values in this problem, it is easy to mentally calculate the voltage drops ($V_1 = 0.10 \times 10\text{ V} = 1\text{ V}$, $V_2 = 0.22 \times 10\text{ V} = 2.2\text{ V}$, and $V_3 = 0.68 \times 10\text{ V} = 6.8\text{ V}$). Such is usually not the case, but sometimes a little thinking will produce a result more efficiently.

Although you have already reasoned through this problem, the calculations are

$$V_1 = \left(\frac{R_1}{R_T}\right)V_S = \left(\frac{100\ \Omega}{1000\ \Omega}\right)10\text{ V} = \mathbf{1\ V}$$

$$V_2 = \left(\frac{R_2}{R_T}\right)V_S = \left(\frac{220\ \Omega}{1000\ \Omega}\right)10\text{ V} = \mathbf{2.2\ V}$$

$$V_3 = \left(\frac{R_3}{R_T}\right)V_S = \left(\frac{680\ \Omega}{1000\ \Omega}\right)10\text{ V} = \mathbf{6.8\ V}$$

Notice that the sum of the voltage drops is equal to the source voltage, in accordance with Kirchhoff's voltage law. This check is a good way to verify your results.

RELATED PROBLEM

If R_1 and R_2 are both changed to 680 Ω in Figure 4–42, what are all the voltage drops?

MULTISIM

Open Multisim file E04-17. Verify the values of V_1, V_2, and V_3. Repeat for the related problem.

EXAMPLE 4–18

Determine the voltages between the following points in the circuit of Figure 4–43:

(a) *A* to *B* **(b)** *A* to *C* **(c)** *B* to *C* **(d)** *B* to *D* **(e)** *C* to *D*

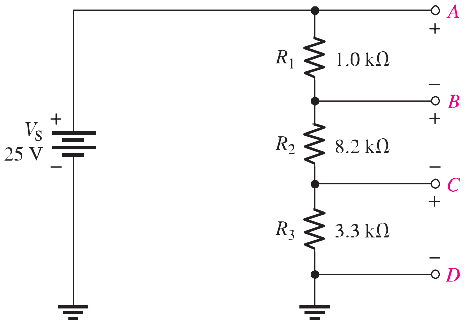

FIGURE 4–43

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

First determine R_T.

$$R_T = 1.0\text{ k}\Omega + 8.2\text{ k}\Omega + 3.3\text{ k}\Omega = 12.5\text{ k}\Omega$$

Then apply the voltage-divider formula to obtain each required voltage.

(a) The voltage A to B is also the voltage drop across R_1.

$$V_{AB} = \left(\frac{R_1}{R_T}\right)V_S = \left(\frac{1.0\text{ k}\Omega}{12.5\text{ k}\Omega}\right)25\text{ V} = \mathbf{2\text{ V}}$$

(b) The voltage from A to C is the combined voltage drop across both R_1 and R_2. In this case, R_x in the general formula given in Equation 4–5 is $R_1 + R_2$.

$$V_{AC} = \left(\frac{R_1 + R_2}{R_T}\right)V_S = \left(\frac{9.2\text{ k}\Omega}{12.5\text{ k}\Omega}\right)25\text{ V} = \mathbf{18.4\text{ V}}$$

(c) The voltage from B to C is the voltage drop across R_2.

$$V_{BC} = \left(\frac{R_2}{R_T}\right)V_S = \left(\frac{8.2\text{ k}\Omega}{12.5\text{ k}\Omega}\right)25\text{ V} = \mathbf{16.4\text{ V}}$$

(d) The voltage from B to D is the combined voltage drop across both R_2 and R_3. In this case, R_x in the general formula is $R_2 + R_3$.

$$V_{BD} = \left(\frac{R_2 + R_3}{R_T}\right)V_S = \left(\frac{11.5\text{ k}\Omega}{12.5\text{ k}\Omega}\right)25\text{ V} = \mathbf{23\text{ V}}$$

(e) Finally, the voltage from C to D is the voltage drop across R_3.

$$V_{CD} = \left(\frac{R_3}{R_T}\right)V_S = \left(\frac{3.3\text{ k}\Omega}{12.5\text{ k}\Omega}\right)25\text{ V} = \mathbf{6.6\text{ V}}$$

If you connect this voltage divider, you can verify each of the calculated voltages by connecting a voltmeter between the appropriate points in each case.

RELATED PROBLEM

Determine each of the previously calculated voltages if V_S is doubled.

MULTISIM

Open Multisim file E04-18. Verify the values of V_{AB}, V_{AC}, V_{BC}, V_{BD}, and V_{CD}. Repeat for the related problem.

The Potentiometer as an Adjustable Voltage Divider

Recall from Chapter 2 that a potentiometer is a variable resistor with three terminals. A linear potentiometer connected to a voltage source is shown in Figure 4–44. Notice that the two end terminals are labeled 1 and 2. The adjustable terminal or wiper is labeled 3. The

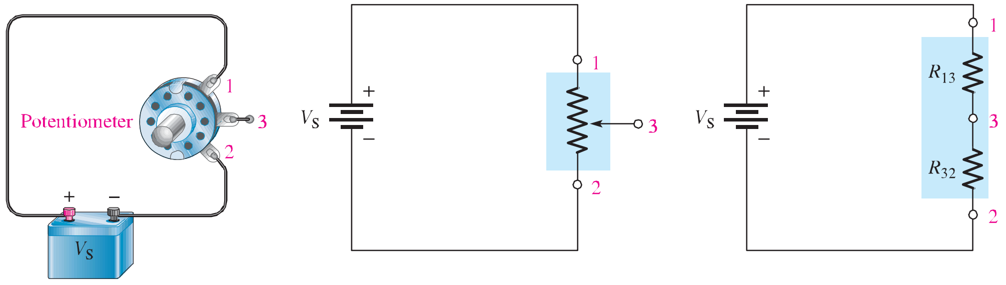

(a) Pictorial (b) Schematic (c) Equivalent schematic

FIGURE 4–44 **The potentiometer used as a voltage divider.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

potentiometer acts as a voltage divider, which can be illustrated by separating the total resistance into two parts as shown in Figure 4–44(c). The resistance between terminal 1 and terminal 3 (R_{13}) is one part, and the resistance between terminal 3 and terminal 2 (R_{32}) is the other part. The potentiometer acts like a two-resistor voltage divider that can be manually adjusted.

Figure 4–45 shows what happens when the wiper contact (3) is moved. In part (a), the wiper is exactly centered, making the two resistances equal. If you measure the voltage across terminals 3 to 2 as indicated, you have one-half of the total source voltage. When the wiper is moved up from the center position, as in part (b), the resistance between terminals 3 and 2 increases, and the voltage across it increases proportionally. When the wiper is moved down from the center position, as in part (c), the resistance between terminals 3 and 2 decreases, and the voltage decreases proportionally.

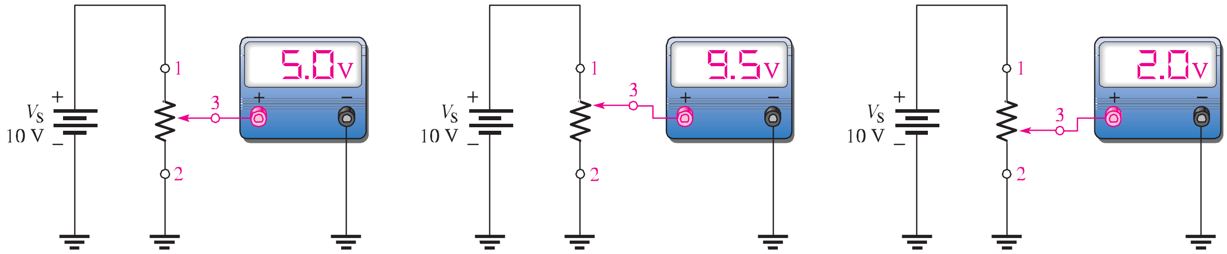

(a) (b) (c)

FIGURE 4–45 **Adjusting the voltage divider.**

6238262 2013/12/10 12.27.66.5

Applications

There are many applications of voltage dividers in systems. A radio receiver can be thought of as a small system. The volume control is a common application of a potentiometer used as a voltage divider. Since the loudness of the sound is dependent on the amount of voltage associated with the audio signal, you can increase or decrease the volume by adjusting the potentiometer, that is, by turning the knob of the volume control. The block diagram in Figure 4–46 shows how a potentiometer can be used for volume control.

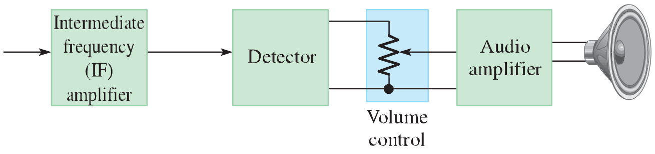

FIGURE 4–46 **A variable voltage divider used for volume control in a radio receiver.**

Figure 4–47 illustrates how a potentiometer voltage divider can be used as a level sensor in a storage tank. As shown in part (a), the float moves up as the tank is filled and moves down as the tank empties. The float is mechanically linked to the wiper arm of a potentiometer, as shown in part (b). The output voltage varies proportionally with the position of the wiper arm. As the liquid in the tank decreases, the sensor output voltage also decreases. The output voltage goes to the indicator circuitry, which controls a digital readout to show the amount of liquid in the tank. The schematic of this system is shown in part (c).

Another application for voltage dividers is in setting the gain of operational amplifiers and setting up reference voltages in power supplies. In certain precision applications, voltage dividers are available in IC form. In addition, voltage dividers are used for setting the dc operating voltage (bias) in transistor amplifiers. Figure 4–48 shows a voltage divider used for this purpose.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

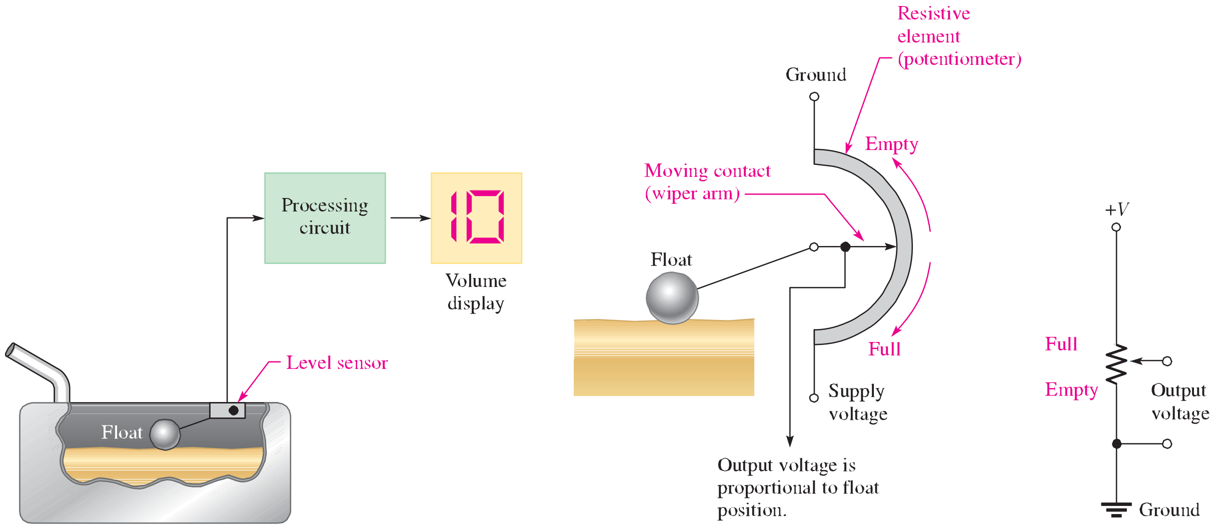
Processing circuit
10
Volume display
Level sensor
Float
Resistive element (potentiometer)
Ground
Empty
Moving contact (wiper arm)
Float
Full
Supply voltage
Output voltage is proportional to float position.
+V
Full
Empty
Output voltage
Ground

(a) Storage tank

(b) Detail of level sensor

(c) Schematic of level sensor

FIGURE 4–47 **A potentiometer voltage divider used as a bias level sensor.**

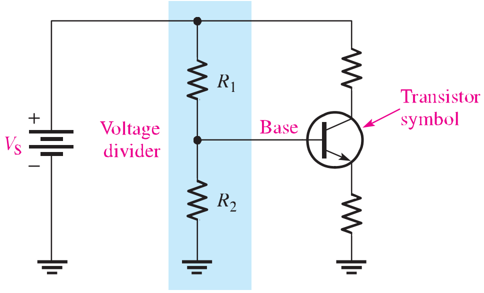
R_1
Transistor symbol
+
V_S
−
Voltage divider
Base
R_2

FIGURE 4–48 **The voltage divider used as a bias circuit for a transistor amplifier.**

A voltage divider is useful for converting a resistance sensor output to a voltage. Resistance sensors were described in Section 2–5 and include thermistors, photoconductive cells, and strain gages. To convert a change in resistance to an output voltage, the resistance sensor can be used in place of one of the resistors of a voltage divider.

EXAMPLE 4–19

Assume you have a CdS cell (a type of photoconductive cell) configured as shown in Figure 4–49 that uses three AA batteries for a voltage source (4.5 V). At dusk, the resistance of the cell rises from a low resistance to above 90 kΩ. The cell is used to trigger a logic circuit that will turn on lights if V_{OUT} is greater than approximately 1.5 V. What value of R will produce an output voltage of 1.5 V when the cell resistance is 90 kΩ?

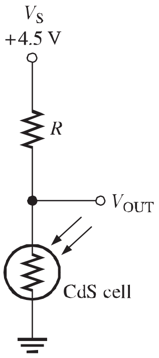
V_S
+4.5 V
R
V_{OUT}
CdS cell

FIGURE 4–49

SOLUTION

Notice that the threshold voltage (1.5 V) is one-third of the supply voltage. You know that 90 kΩ represents one-third of the total resistance at this point. Therefore, the total resistance is

$$R_T = 3(90\text{ k}\Omega) = 270\text{ k}\Omega$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The resistance needed to produce an output voltage of 1.5 V is

$$R = R_T - 90\text{ k}\Omega = 270\text{ k}\Omega - 90\text{ k}\Omega = \mathbf{180\text{ k}\Omega}$$

RELATED PROBLEM

Starting with Equation 4–5, prove that the required resistance is 180 kΩ to produce an output of 1.5 V when the cell resistance is 90 kΩ.

SECTION 4–7 CHECKUP

1. What is a voltage divider?
2. How many resistors can there be in a series voltage-divider circuit?
3. Write the general voltage-divider formula.
4. If two series resistors of equal value are connected across a 20 V source, how much voltage is there across each resistor?
5. A 56 kΩ resistor and an 82 kΩ resistor are connected as a voltage divider. The source voltage is 10 V. Draw the circuit, and determine the voltage across each of the resistors.
6. The circuit of Figure 4–50 is an adjustable voltage divider. If the potentiometer is linear, where would you set the wiper in order to get 5 V from *B* to *A* and 5 V from *C* to *B*?

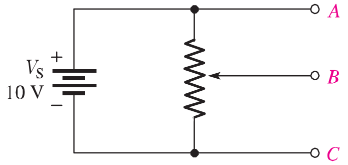

FIGURE 4–50

4–8 POWER IN SERIES CIRCUITS

The power dissipated by each individual resistor in a series circuit contributes to the total power in the circuit. The individual powers are additive.

After completing this section, you should be able to

- **Determine power in a series circuit**
  - **Apply any of the power formulas**

The total amount of power in a series resistive circuit is equal to the sum of the powers in each resistor in series.

$$\boldsymbol{P_T = P_1 + P_2 + P_3 + \cdots + P_n} \quad (4\text{–}6)$$

where P_T is the total power and P_n is the power in the last resistor in series (n can be any positive integer equal to the number of resistors in series).

The power formulas that you learned in Chapter 3 are, applicable to series circuits. Since each resistor in series has the same current through it, the following formulas are used to calculate the total power:

$$P_T = V_S I$$
$$P_T = I^2 R_T$$
$$P_T = \frac{V_S^2}{R_T}$$

where I is the current through the circuit, V_S is the source voltage across the series circuit, and R_T is the total resistance.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 4–20

Determine the total amount of power in the series circuit in Figure 4–51.

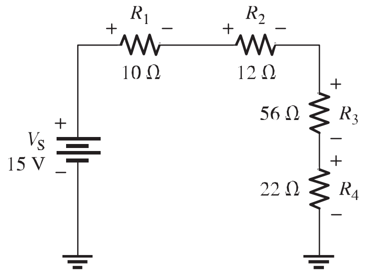

FIGURE 4–51

SOLUTION

The source voltage is 15 V. The total resistance is

$$R_T = 10\ \Omega + 12\ \Omega + 56\ \Omega + 22\ \Omega = 100\ \Omega$$

The easiest formula to use is $P_T = V_S^2/R_T$ since you know both V_S and R_T.

$$P_T = \frac{V_S^2}{R_T} = \frac{(15\ \text{V})^2}{100\ \Omega} = \frac{225\ \text{V}^2}{100\ \Omega} = 2.25\ \text{W}$$

If you determine the power of each resistor separately and add all these powers, you obtain the same result. Another calculation will illustrate. First, find the current.

$$I = \frac{V_S}{R_T} = \frac{15\ \text{V}}{100\ \Omega} = 150\ \text{mA}$$

Next, calculate the power for each resistor using $P = I^2R$.

$$P_1 = (150\ \text{mA})^2(10\ \Omega) = 225\ \text{mW}$$
$$P_2 = (150\ \text{mA})^2(12\ \Omega) = 270\ \text{mW}$$
$$P_3 = (150\ \text{mA})^2(56\ \Omega) = 1.26\ \text{W}$$
$$P_4 = (150\ \text{mA})^2(22\ \Omega) = 495\ \text{mW}$$

6238262 2013/12/10 12:27:56.5

Now, add these powers to get the total power.

$$P_T = 225\ \text{mW} + 270\ \text{mW} + 1.260\ \text{W} + 495\ \text{mW} = \mathbf{2.25\ W}$$

This result shows that the sum of the individual powers is equal to the total power as determined by the formula $P_T = V_S^2/R_T$.

RELATED PROBLEM

What is the total power in the circuit of Figure 4–51 if V_S is increased to 30 V?

SECTION 4–8 CHECKUP

1. If you know the power in each resistor in a series circuit, how can you find the total power?
2. The resistors in a series circuit have the following powers: 1 W, 2 W, 5 W, and 8 W. What is the total power in the circuit?
3. A circuit has a 100 Ω, a 330 Ω, and a 680 Ω resistor in series. There is a current of 1 mA through the circuit. What is the total power?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

4–9 VOLTAGE MEASUREMENTS

The concept of *reference ground* was introduced in Chapter 2 and designated as the 0 V reference point for a circuit. Voltage is always measured with respect to another point in the circuit. *Ground* is discussed in more detail in this section.

After completing this section, you should be able to

- **State how to measure voltage with respect to ground**
  - **Define the term *reference ground***
  - **Explain the use of single and double subscripts for indicating voltages**

The term *ground* has its origin in telephone systems in which one of the conductors was the earth itself. The term was also used in early radio receiving antennas (called *aerials*) where one part of the system was connected to a metal pipe driven into the earth. Today, *ground* can mean different things and is not necessarily at the same potential as the earth. In electronic systems, **reference ground** (or **common**) refers to a conductor that is the comparison point for voltage measurements in a circuit. Frequently, it is the conductor that carries the power supply return current. Most electronic circuit boards have a larger conducting surface area for ground. For many multilayer boards the surface area is a separate internal layer, which is referred to as a **ground plane**.

In electrical wiring, the reference ground is usually the same as the earth potential because neutral and earth ground are connected at the entrance point to a building. In this case, reference ground and earth ground are at the same potential. (Section 517 of the National Electric Code (NEC) specifies certain exceptions to this grounding method for hospital operating rooms and health care facilities.)

The concept of *reference ground* is also used in automotive electrical systems. In most automotive systems, the chassis of the automobile is the ground reference (even though the tires insulate it from the earth, which can be a different potential). In nearly all modern automobiles, the negative post of the battery is connected to the chassis with a solid low-resistance connection. This makes the chassis of the vehicle serve as the return path for all of the electrical circuits in the vehicle, as illustrated in the simplified drawing in Figure 4–52. In some vintage cars, the positive terminal was connected to the chassis in an arrangement called *positive ground*. In both cases, the chassis represents the reference ground point.

SAFETY NOTE

When removing a battery from an automobile, remove the ground wire first. If a tool should then accidentally come in contact with the chassis and the positive terminal, a spark will be avoided; there is no current because there is no return path. When installing a battery, always install the ground connection last to avoid a spark if a tool accidentally touches the chassis.

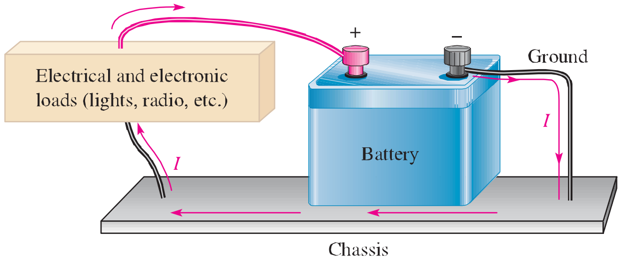

FIGURE 4–52 **The chassis serves as a return path for the electrical circuits in a vehicle.**

Measuring Voltages with Respect to Ground

When voltages are measured with respect to ground, they are indicated with a single-letter subscript. For example, V_A is voltage at point A with respect to ground. Each circuit in Figure 4–53 consists of three 1.0 kΩ series resistors and four lettered reference points. Reference ground represents a potential of 0 V with respect to all other points in each circuit as illustrated. In part (a), the reference point is D, and all voltages are positive with respect to D. In part (b), the reference point is A in the circuit, and all other points have a negative voltage.

Many circuits use both positive and negative voltages and as mentioned, the return path for these supplies is designated as reference ground. Figure 4–53(c) shows the same

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

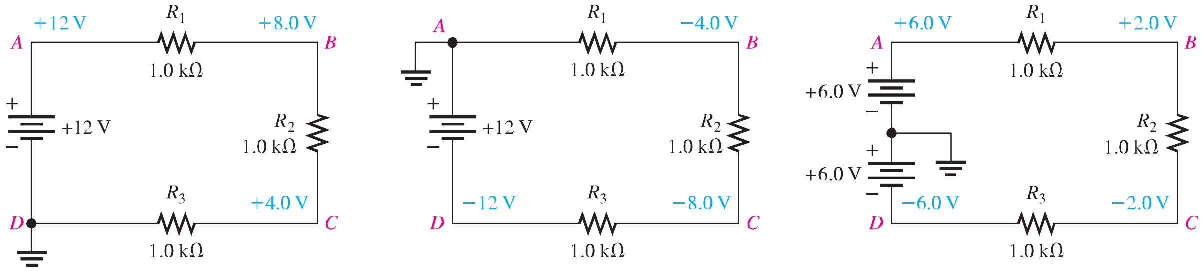

(a) Positive voltages with respect to ground (b) Negative voltages with respect to ground (c) Positive and negative voltages with respect to ground

FIGURE 4–53 **The ground point does not affect the current in the circuit or the voltage drops across the resistors.**

circuit but with two 6 V series-aiding sources replacing the 12 V source. In this case, the reference point is arbitrarily designated between the two voltage sources. There is exactly the same current in all three circuits, but now voltages are referenced to the new ground point. As you can see from these examples, the reference ground point is arbitrary and does not change the current.

Not all voltages are measured with respect to ground. If you wish to specify the voltage drop across an ungrounded resistor, you can name the resistor in the subscript or use two-lettered subscripts. When two different subscripts are used, the voltage represents the difference between the points. For example, V_{BC} means $V_B - V_C$. In Figure 4–53, V_{BC} is the same in all three circuits (+4.0 V) as you can confirm by performing the subtraction in each case. Another way of designating V_{BC} is to write simply V_{R2}.

There is one more convention commonly used to express voltages using subscripts. Power supply voltages are usually given with a double-letter subscript. The reference point is ground or common. For example, a voltage written as V_{CC} is a positive power supply voltage with respect to ground. Other common power supply voltages are V_{DD} (positive), V_{EE} (negative), and V_{SS} (negative).

To measure voltages with a digital meter, the meter leads can be connected across any two points and the meter will indicate the voltage, either positive or negative. The meter reference jack is labeled "COM" (normally black). This is common only to the meter, not to the circuit. Figure 4–54 shows a DMM used to measure the voltage across the ungrounded

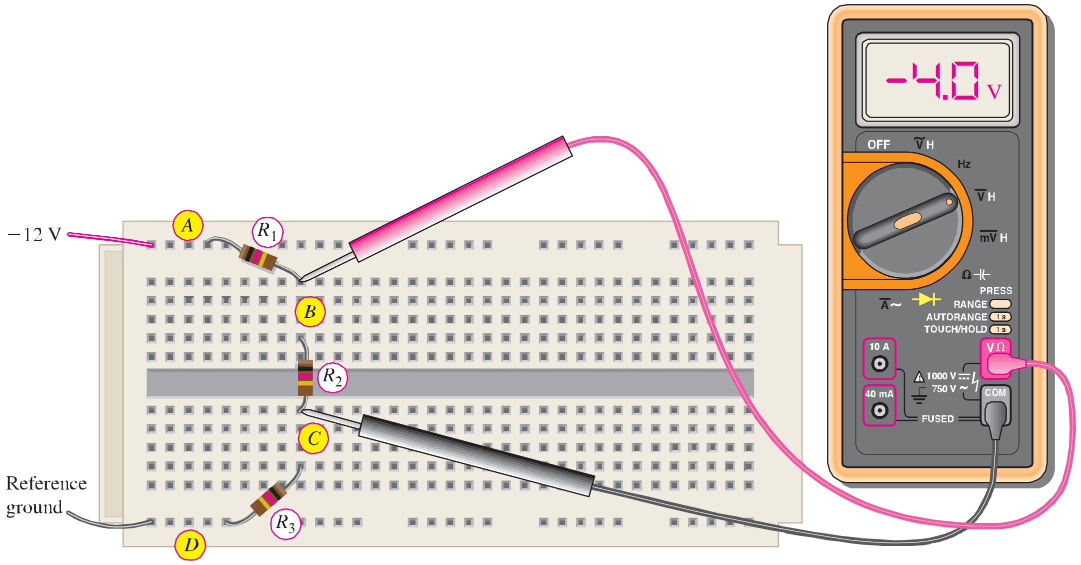

FIGURE 4–54 **A DMM has a "floating" common, so the leads can be connected to any points in the circuit and it will indicate the correct voltage between the two leads.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

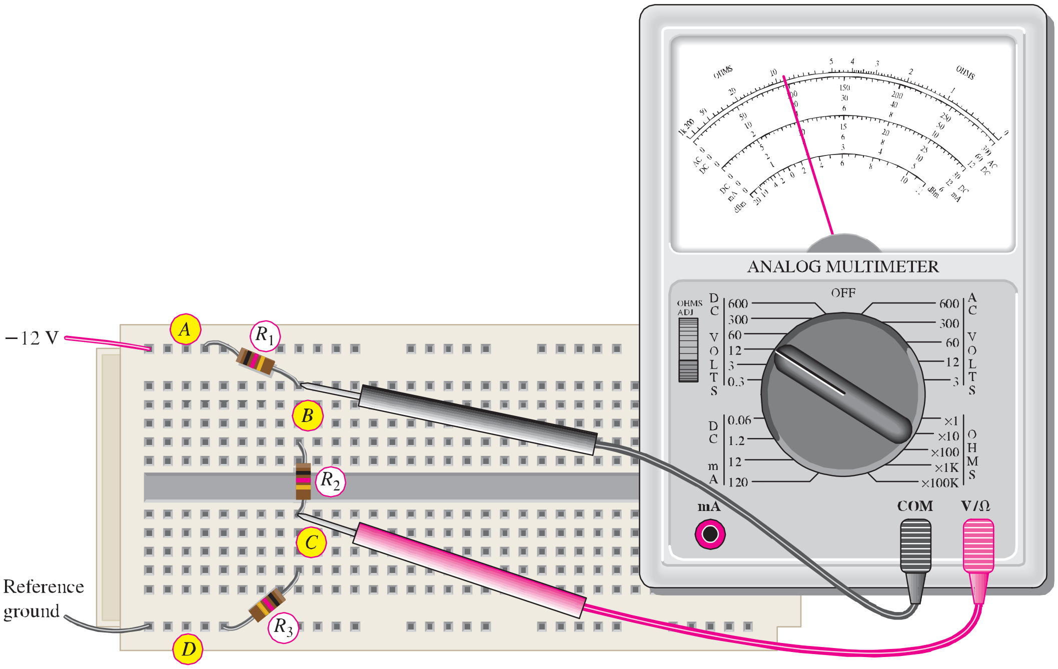

FIGURE 4–55 **An analog meter needs to be connected so that the positive lead goes to the more positive point in the circuit.**

(floating) resistor R_2. The circuit is the same one given in Figure 4–53(b), which has a negative power supply. It is shown as it might be constructed in a lab. Notice that the meter indicates a negative voltage, which means that the meter's COM lead is the more positive lead. If you wanted to measure voltages with respect to the circuit's reference ground, you would connect the COM on the meter to the circuit's reference ground. The voltage with respect to the reference ground will be indicated.

0258262 2015/12/10 12.27.66.5

If you are using an analog meter to make circuit measurements, you must connect the meter so that its common lead is connected to the most negative point in the circuit, or else the meter movement will try to move backward. Figure 4–55 shows an analog meter connected to the same circuit as before; notice that the meter leads are reversed. To measure voltage across R_2, the leads must be connected so that the meter will deflect in the positive direction. The positive lead on the meter is connected to the more positive point in the circuit, which in this case is the circuit's ground. The user appends a minus sign to the reading when recording the reading.

MULTISIM

Open Multisim file E04-21. For each circuit, verify the values of the voltages at each point with respect to ground. Repeat for the related problem.

EXAMPLE 4–21

Determine the voltages at each of the indicated points in each circuit of Figure 4–56 with respect to ground. Since each of the four resistors has the same value, 25 V are dropped across each one.

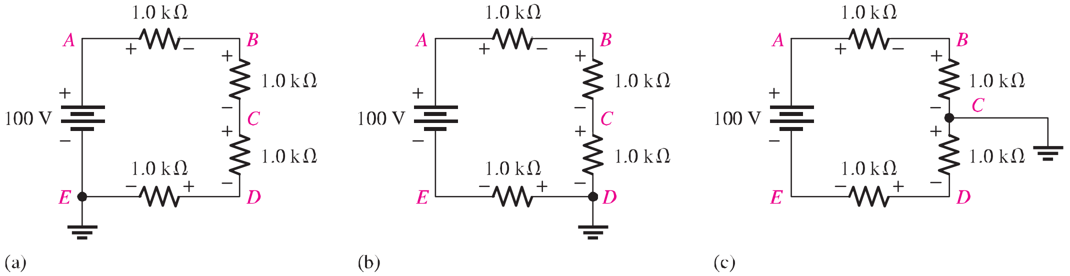

(a) (b) (c)

FIGURE 4–56

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

For the circuit in Figure 4–56(a), the voltage polarities are as shown. Point *E* is ground. Single-letter subscripts denote voltage at a point with respect to ground. The voltages with respect to ground are as follows:

$$V_E = \mathbf{0\ V}, \quad V_D = \mathbf{+25\ V}, \quad V_C = \mathbf{+50\ V}, \quad V_B = \mathbf{+75\ V}, \quad V_A = \mathbf{+100\ V}$$

For the circuit in Figure 4–56(b), the voltage polarities are as shown. Point *D* is ground. The voltages with respect to ground are as follows:

$$V_E = \mathbf{-25\ V}, \quad V_D = \mathbf{0\ V}, \quad V_C = \mathbf{+25\ V}, \quad V_B = \mathbf{+50\ V}, \quad V_A = \mathbf{+75\ V}$$

For the circuit in Figure 4–56(c), the voltage polarities are as shown. Point *C* is ground. The voltages with respect to ground are as follows:

$$V_E = \mathbf{-50\ V}, \quad V_D = \mathbf{-25\ V}, \quad V_C = \mathbf{0\ V}, \quad V_B = \mathbf{+25\ V}, \quad V_A = \mathbf{+50\ V}$$

RELATED PROBLEM

If the ground is moved to point *A* in Figure 4–56(a), what are the voltages at each of the other points with respect to ground?

SECTION 4–9 CHECKUP

1. What is the reference point in a circuit called?
2. If V_{AB} in a circuit is +5.0 V, what is V_{BA}?
3. Voltages in a circuit are generally referenced to ground. (True or False)
4. The housing or chassis can be used as reference ground. (True or False)

4–10 TROUBLESHOOTING

Open components or contacts and shorts between conductors are common problems in all circuits. An open produces an infinite resistance. A short produces nearly zero resistance.

After completing this section, you should be able to

- **Troubleshoot series circuits**
  - **Check for an open circuit**
  - **Check for a short circuit**
  - **Identify primary causes of opens and shorts**

Open Circuit

The most common failure in a series circuit is an **open**. For example, when a resistor or a lamp burns out, it creates an open, causing a break in the current path, as illustrated in Figure 4–57.

An open in a series circuit prevents current.

TROUBLESHOOTING AN OPEN In Chapter 3, you were introduced to the analysis, planning, and measurement (APM) approach to troubleshooting. You also learned

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

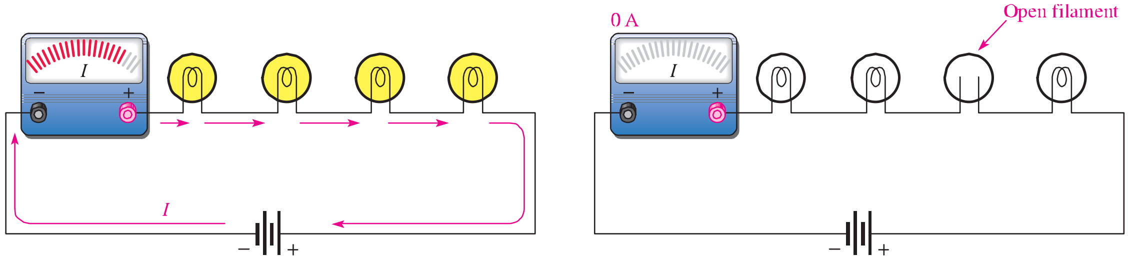

(a) A complete series circuit has current.

(b) An open series circuit has no current.

FIGURE 4–57 **Current ceases when an open occurs.**

about the half-splitting method and saw an example using an ohmmeter. Now, the same principles will be applied using voltage measurements instead of resistance measurements. As you know, voltage measurements are generally the easiest to make because you do not have to disconnect anything.

As a beginning step, prior to analysis, it is a good idea to make a visual check of the faulty circuit. Occasionally, you can find a charred resistor, a broken lamp filament, a loose wire, or a loose connection this way. However, it is possible (and probably more common) for a resistor or other component to open without showing visible signs of damage. When a visual check reveals nothing, then proceed with the APM approach.

When an open occurs in a series circuit, all of the source voltage appears across the open. The reason for this is that the open condition prevents current through the series circuit. With no current, there can be no voltage drop across any of the other resistors (or other component). Since $IR = (0\text{ A})R = 0\text{ V}$, the voltage on each end of a good resistor is the same. Therefore, the voltage applied across a series string also appears across the open component because there are no other voltage drops in the circuit, as illustrated in Figure 4–58. The source voltage will appear across the open resistor in accordance with Kirchhoff's voltage law as follows:

$$
\begin{aligned}
V_S &= V_1 + V_2 + V_3 + V_4 + V_5 + V_6 \\
V_4 &= V_S - V_1 - V_2 - V_3 - V_5 - V_6 \\
&= 10\text{ V} - 0\text{ V} - 0\text{ V} - 0\text{ V} - 0\text{ V} - 0\text{ V} \\
V_4 &= V_S = 10\text{ V}
\end{aligned}
$$

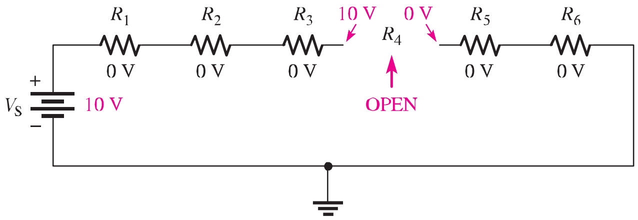

FIGURE 4–58 **The source voltage appears across the open series resistor.**

EXAMPLE OF HALF-SPLITTING USING VOLTAGE MEASUREMENTS

Suppose a circuit has four resistors in series. You have determined by *analyzing* the symptoms (there is voltage but no current) that one of the resistors is open, and you are *planning* to find the open resistor using a voltmeter for *measuring* by the half-splitting method. A sequence of measurements for this particular case is illustrated in Figure 4–59.

Step 1: Measure across R_1 and R_2 (the left half of the circuit). A 0 V reading indicates that neither of these resistors is open.

Step 2: Move the meter to measure across R_3 and R_4; the reading is 10 V. This indicates there is an open in the right half of the circuit, so either R_3 or R_4 is the faulty resistor (assume no bad connections).

HANDS ON TIP

When measuring a resistance, make sure that you do not touch the meter leads or the resistor leads. If you hold both ends of a high-value resistor in your fingers along with the meter probes, the measurement will be inaccurate because your body resistance can affect the measured value. When body resistance is placed in parallel with a high-value resistor, the measured value will be less than the actual value of the resistor.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

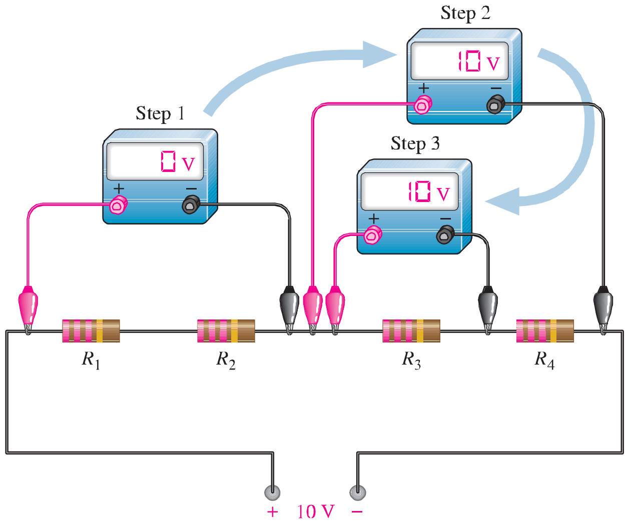

FIGURE 4–59 **Troubleshooting a series circuit for an open using half-splitting.**

Step 3: Move the meter to measure across R_3. A measurement of 10 V across R_3 identifies it as the open resistor. If you had measured across R_4, it would have indicated 0 V. This would have also identified R_3 as the faulty component because it would have been the only one left that could have 10 V across it.

Short Circuit

Sometimes an unwanted short circuit occurs when two conductors touch or a foreign object such as solder or a wire clipping accidentally connects two sections of a circuit together. This situation is more common in circuits with a high component density. Three potential causes of short circuits are illustrated on the PC board in Figure 4–60.

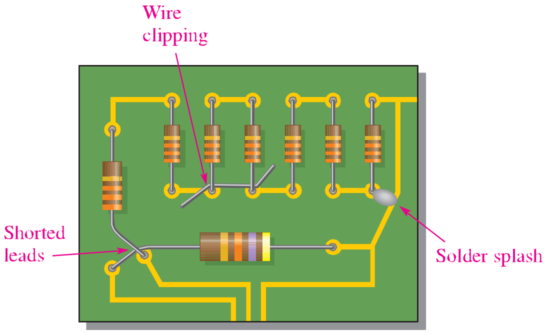

FIGURE 4–60 **Examples of shorts on a PC board.**

When there is a **short**, a portion of the series resistance is bypassed (all of the current goes through the short), thus reducing the total resistance as illustrated in Figure 4–61. Notice that the current increases as a result of the short.

A short in a series circuit causes more current.

TROUBLESHOOTING A SHORT A short is generally very difficult to troubleshoot. As in any troubleshooting situation, it is a good idea to make a visual check of the faulty circuit. In the case of a short in the circuit, a wire clipping, solder splash, or touching leads is often found to be the culprit. In terms of component failure, shorts are less common than opens in many types of components. Furthermore, a short in one part of a circuit can

6238262 2013/12/10 12.27.66.5

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

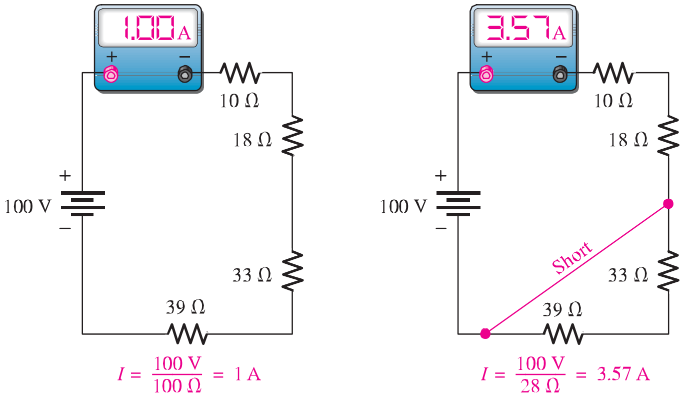
1.00 A
+
−
10 Ω
18 Ω
100 V
33 Ω
39 Ω
$I = \frac{100\ V}{100\ \Omega} = 1\ A$
3.57 A
Short
$I = \frac{100\ V}{28\ \Omega} = 3.57\ A$

(a) Before short (b) After short

FIGURE 4–61 **The effect of a short in a series circuit.**

cause overheating in another part due to the higher current caused by the short. As a result two failures, an open and a short, may occur together.

When a short occurs in a series circuit, there is essentially no voltage across the shorted part. A short has zero or near zero resistance, although shorts with significant resistance values can occur from time to time. These are called *resistive shorts.* For purposes of illustration, zero resistance is assumed for all shorts.

In order to troubleshoot a short, measure the voltage across each resistor until you get a reading of 0 V. This is the straightforward approach and does not use half-splitting. In order to apply the half-splitting method, you must know the correct voltage at each point in the circuit and compare it to measured voltages. Example 4–22 illustrates using half-splitting to find a short.

EXAMPLE 4–22

Assume you have determined that there is a short in a circuit with four series resistors because the current is higher than it should be. You know that the voltage at each point in the circuit should be as shown in Figure 4–62 if the circuit is working properly. The voltages are shown relative to the negative terminal of the source. Find the location of the short.

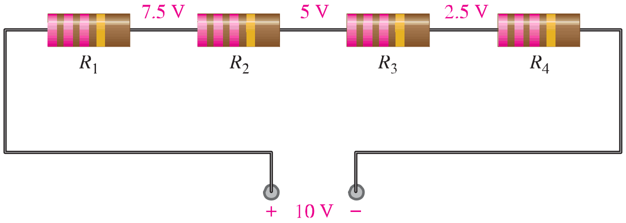

FIGURE 4–62 **Series circuit (without a short) with correct voltages marked.**

SOLUTION

Use the half-splitting method to troubleshoot the short.

Step 1: Measure across R_1 and R_2. The meter shows a reading of 6.67 V, which is higher than the normal voltage (it should be 5 V). Look elsewhere for a voltage that is lower than normal because a short will make the voltage less across that part of the circuit.

Step 2: Move the meter and measure across R_3 and R_4; the reading of 3.33 V is incorrect and lower than normal (it should be 5 V). This shows that the short is in the right half of the circuit and that either R_3 or R_4 is shorted.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Step 3: Again move the meter and measure across R_3. A reading of 3.3 V across R_3 tells you that R_4 is shorted because it must have 0 V across it. Figure 4–63 illustrates this troubleshooting technique.

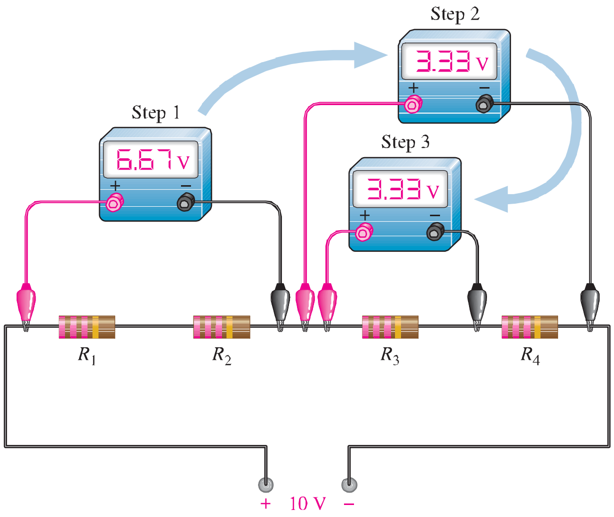

FIGURE 4–63 **Troubleshooting a series circuit for a short using half-splitting.**

RELATED PROBLEM

Assume that R_1 is shorted in Figure 4–63. What would the Step 1 measurement be?

SECTION 4–10 CHECKUP

1. Define *open*.
2. Define *short*.
3. What happens when a series circuit opens?
4. Name two general ways in which an open circuit can occur in practice. How can a short circuit occur?
5. When a resistor fails, it will normally open. (True or False)
6. The total voltage across a string of series resistors is 24 V. If one of the resistors is open, how much voltage is there across it? How much is there across each of the good resistors?
7. Explain why the voltage measured in Step 1 of Figure 4–63 is higher than normal.

SUMMARY

- The total series resistance is the sum of all resistors in the series circuit.
- The total resistance between any two points in a series circuit is equal to the sum of all resistors connected in series between those two points.
- If all of the resistors in a series circuit are of equal value, the total resistance is the number of resistors multiplied by the resistance value of one resistor.
- The current is the same at all points in a series circuit.
- Voltage sources in series add algebraically.
- Kirchhoff's voltage law: The sum of the voltage drops in a series circuit equals the total source voltage.
- Kirchhoff's voltage law: the algebraic sum of all the voltages around a closed single path is zero.
- The voltage drops in a circuit are always opposite in polarity to the total source voltage.
- Current is defined to be out of the negative side of a source and into the positive side.
- Current is defined to be into the negative side of each resistor and out of the positive side.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

- A voltage divider is a series arrangement of resistors connected to a voltage source.
- A voltage divider is so named because the voltage drop across any resistor in the series circuit is divided down from the total voltage by an amount proportional to that resistance value in relation to the total resistance.
- A potentiometer can be used as an adjustable voltage divider.
- The total power in a resistive circuit is the sum of all the individual powers of the resistors making up the series circuit.
- Voltages given with a single-lettered subscript are referenced to ground. When two different letters are used in the subscript, the voltage is the difference in the two points.
- Ground (common) is zero volts with respect to all points referenced to it in the circuit.
- *Negative ground* is the term used when the negative side of the source is grounded.
- *Positive ground* is the term used when the positive side of the source is grounded.
- The voltage across an open component always equals the source voltage.
- The voltage across a shorted component is always 0 V.

KEY TERMS

Key terms and other bold terms in the chapter are defined in the end-of-book glossary.

Kirchhoff's voltage law A law stating that (1) the sum of the voltage drops around a single closed path equals the source voltage in that loop or (2) the algebraic sum of all the voltages around a single closed path in a circuit is zero.

Open A circuit condition in which the current path is interrupted.

Reference ground The metal chassis that houses the assembly or a large conductive area on a printed circuit board used as the common or reference point; also called common (com).

6238262 2013/12/10 12.27.66.5

Series In an electric circuit, a relationship of components in which the components are connected such that they provide a single current path between two points.

Series-aiding An arrangement of two or more series voltage sources with polarities in the same direction.

Series-opposing An arrangement of two series voltage sources with polarities in the opposite direction; not a normal configuration.

Short A circuit condition in which there is a zero or abnormally low resistance path between two points: usually an inadvertent condition.

Voltage divider A circuit consisting of series resistors across which one or more output voltages are taken.

KEY FORMULAS

(4–1)	$R_T = R_1 + R_2 + R_3 + \cdots + R_n$	Total resistance of *n* resistors in series
(4–2)	$R_T = nR$	Total resistance of *n* equal-value resistors in series
(4–3)	$V_S = V_1 + V_2 + V_3 + \cdots + V_n$	Kirchhoff's voltage law in a series circuit
(4–4)	$V_1 + V_2 + V_3 + \cdots + V_n = 0$	Kirchhoff's voltage law
(4–5)	$V_x = \left(\frac{R_x}{R_T}\right)V_S$	Voltage-divider formula
(4–6)	$P_T = P_1 + P_2 + P_3 + \cdots + P_n$	Total power

TRUE/FALSE QUIZ

Answers are at the end of the chapter.

1. A series circuit can have more than one path for current.
2. The total resistance of a series circuit can be less than the largest resistor in that circuit.
3. If two series resistors are different sizes, the larger resistor will have the larger current.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

4. If two series resistors are different sizes, the larger resistor will have the larger voltage.
5. If three equal resistors are used in a voltage divider, the voltage across each one will be one-third of the source voltage.
6. There is no valid electrical reason for installing flashlight batteries so that they are series-opposing.
7. Kirchhoff's voltage law is valid only if a loop contains a voltage source.
8. The voltage-divider equation can be written as $V_x = (R_x/R_T)V_S$.
9. The power dissipated by the resistors in a series circuit is the same as the power supplied by the source.
10. If point A in a circuit has a voltage of $+10$ V and point B has a voltage of -2 V, then V_{AB} is $+8$ V.

SELF-TEST

Answers are at the end of the chapter.

1. Five equal-value resistors are connected in series and there is a current of 2 mA into the first resistor. The amount of current out of the second resistor is
   (a) 2 mA **(b)** 1 mA **(c)** 4 mA **(d)** 0.4 mA
2. To measure the current out of the third resistor in a circuit consisting of four series resistors, an ammeter can be placed
   (a) between the third and fourth resistors **(b)** between the second and third resistors
   (c) at the positive terminal of the source **(d)** at any point in the circuit
3. When a third resistor is connected in series with two series resistors, the total resistance
   (a) remains the same **(b)** increases **(c)** decreases **(d)** increases by one-third
4. When one of four series resistors is removed from a circuit and the circuit reconnected, the current
   (a) decreases by the amount of current through the removed resistor
   (b) decreases by one-fourth **(c)** quadruples **(d)** increases
5. A series circuit consists of three resistors with values of 100 Ω, 220 Ω, and 330 Ω. The total resistance is
   (a) less than 100 Ω **(b)** the average of the values **(c)** 550 Ω **(d)** 650 Ω
6. A 9 V battery is connected across a series combination of 68 Ω, 33 Ω, 100 Ω, and 47 Ω resistors. The amount of current is
   (a) 36.3 mA **(b)** 27.6 A **(c)** 22.3 mA **(d)** 363 mA
7. While putting four 1.5 V batteries in a flashlight, you accidentally put one of them in backward. The light will be
   (a) brighter than normal **(b)** dimmer than normal **(c)** off **(d)** the same
8. If you measure all the voltage drops and the source voltage in a series circuit and add them together, taking into consideration the polarities, you will get a result equal to
   (a) the source voltage **(b)** the total of the voltage drops
   (c) zero **(d)** the total of the source voltage and the voltage drops
9. There are six resistors in a given series circuit and each resistor has 5 V dropped across it. The source voltage is
   (a) 5 V **(b)** 30 V
   (c) dependent on the resistor values **(d)** dependent on the current
10. A series circuit consists of a 4.7 kΩ, a 5.6 kΩ, and a 10 kΩ resistor. The resistor that has the most voltage across it is
    (a) the 4.7 kΩ **(b)** the 5.6 kΩ
    (c) the 10 kΩ **(d)** impossible to determine from the given information
11. Which of the following series combinations dissipates the most power when connected across a 100 V source?
    (a) one 100 Ω resistor **(b)** two 100 Ω resistors
    (c) three 100 Ω resistors **(d)** four 100 Ω resistors

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

12. The total power in a certain circuit is 1 W. Each of the five equal-value series resistors making up the circuit dissipates

(a) 1 W **(b)** 5 W **(c)** 0.5 W **(d)** 0.2 W

13. When you connect an ammeter in a series resistive circuit and turn on the source voltage, the meter reads zero. You should check for

(a) a broken wire **(b)** a shorted resistor **(c)** an open resistor **(d)** both (a) and (c)

14. While checking out a series resistive circuit, you find that the current is higher than it should be. You should look for

(a) an open circuit **(b)** a short **(c)** a low resistor value **(d)** both (b) and (c)

TROUBLESHOOTING: SYMPTOM AND CAUSE

The purpose of these exercises is to help develop thought processes essential to troubleshooting. Answers are at the end of the chapter.

Determine the cause for each set of symptoms. Refer to Figure 4–64.

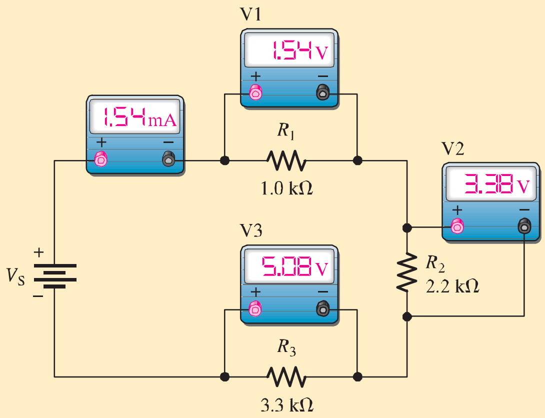

FIGURE 4–64 **The meters indicate the correct readings for this circuit.**

1. *Symptom:* The ammeter reading is zero, the voltmeter 1 and voltmeter 3 readings are zero, and the voltmeter 2 reading is 10 V.

Cause:

(a) R_1 is open. **(b)** R_2 is open. **(c)** R_3 is open.

2. *Symptom:* The ammeter reading is zero, and all the voltmeter readings are zero.

Cause:

(a) A resistor is open. **(b)** The voltage source is turned off or faulty.
(c) One of the resistor values is too high.

3. *Symptom:* The ammeter reading is 2.33 mA, and the voltmeter 2 reading is zero.

Cause:

(a) R_1 is shorted. **(b)** The voltage source is set too high. **(c)** R_2 is shorted.

4. *Symptom:* The ammeter reading is zero, voltmeter 1 reads 0 V, voltmeter 2 reads 5 V, and voltmeter 3 reads 5 V.

Cause:

(a) R_1 is shorted. **(b)** R_1 and R_2 are open. **(c)** R_2 and R_3 are open.

5. *Symptom:* The ammeter reading is 0.645 mA, the voltmeter 1 reading is too high, and the other two voltmeter readings are too low.

Cause:

(a) R_1 has an incorrect value of 10 kΩ. **(b)** R_2 has an incorrect value of 10 kΩ.
(c) R_3 has an incorrect value of 10 kΩ.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

PROBLEMS

6238262 2013/12/10 12.27.66.5

Answers to odd-numbered problems are at the end of the book.

BASIC PROBLEMS

SECTION 4–1 Resistors in Series

1. Connect each set of resistors in Figure 4–65 in series between points A and B.

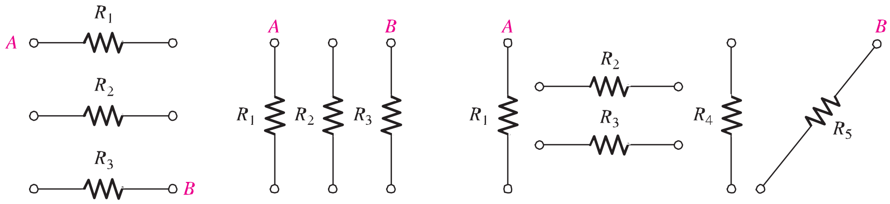

(a) (b) (c)

FIGURE 4–65

2. Determine which resistors in Figure 4–66 are in series. Show how to interconnect the pins to put all the resistors in series.
3. Determine the resistance between pins 1 and 8 in the circuit board in Figure 4–66.
4. Determine the resistance between pins 2 and 3 in the circuit board in Figure 4–66.

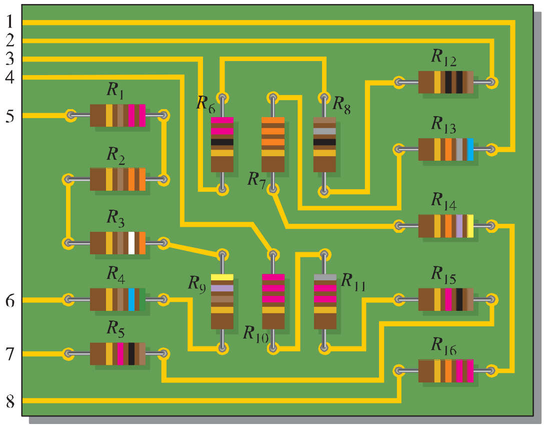

FIGURE 4–66

SECTION 4–2 Total Series Resistance

5. An 82 Ω resistor and a 56 Ω resistor are connected in series. What is the total resistance?
6. Find the total resistance of each group of series resistors shown in Figure 4–67.

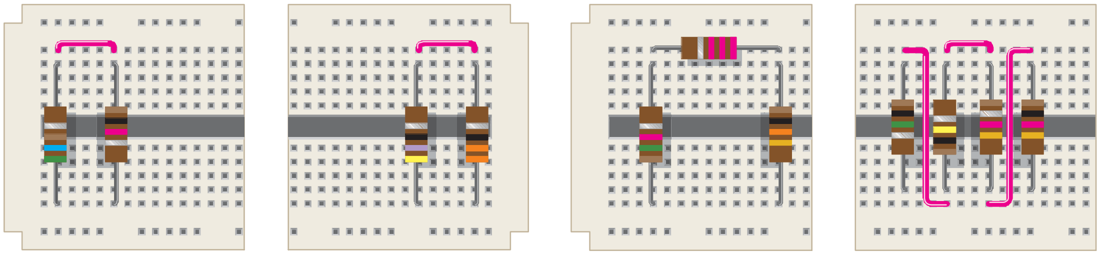

(a) (b) (c) (d)

FIGURE 4–67

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

7. Determine R_T for each circuit in Figure 4–68. Show how to measure R_T with an ohmmeter.

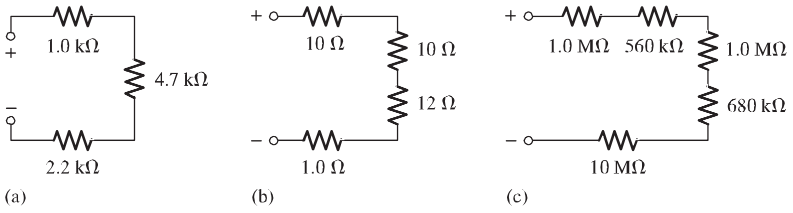

FIGURE 4–68

8. What is the total resistance of twelve 5.6 kΩ resistors in series?

9. Six 47 Ω resistors, eight 100 Ω resistors, and two 22 Ω resistors are in series. What is the total resistance?

10. The total resistance in Figure 4–69 is 20 kΩ. What is the value of R_5?

11. Determine the resistance between each of the following sets of pins on the PC board in Figure 4–66.

(a) pin 1 and pin 8 **(b)** pin 2 and pin 3
(c) pin 4 and pin 7 **(d)** pin 5 and pin 6

12. If all the resistors in Figure 4–66 are connected in series, what is the total resistance?

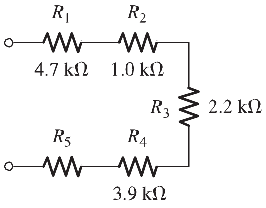

FIGURE 4–69

SECTION 4–3 Current in a Series Circuit

13. What is the current through each of four resistors in a series circuit if the source voltage is 12 V and the total resistance is 120 Ω?

14. The current from the source in Figure 4–70 is 5 mA. How much current does each milliammeter in the circuit indicate?

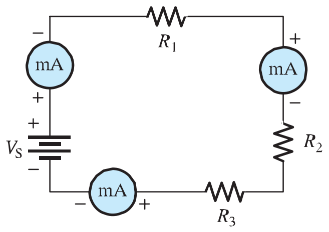

FIGURE 4–70

SECTION 4–4 Application of Ohm's Law

15. What is the current in each circuit of Figure 4–71? Show how to connect an ammeter in each case.

16. Determine the voltage across each resistor in Figure 4–71.

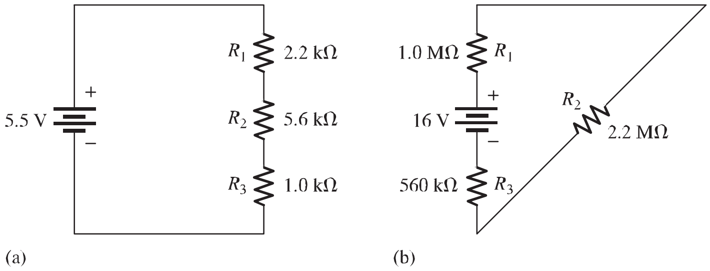

FIGURE 4–71

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

17. Three 470 Ω resistors are in series with a 48 V source.
   (a) How much current is there?
   (b) What is the voltage across each resistor?
   (c) What is the minimum power rating of the resistors?

18. Four equal-value resistors are in series with a 5 V source, and a current of 1 mA is measured. What is the value of each resistor?

SECTION 4–5 Voltage Sources in Series

19. Show how to connect four 6 V batteries to achieve a voltage of 24 V.

20. What happens if one of the batteries in Problem 19 is accidentally connected in reverse?

SECTION 4–6 Kirchhoff's Voltage Law

21. The following voltage drops are measured across each of three resistors in series: 5.5 V, 8.2 V, and 12.3 V. What is the value of the source voltage to which these resistors are connected?

22. Five resistors are in series with a 20 V source. The voltage drops across four of the resistors are 1.5 V, 5.5 V, 3 V, and 6 V. How much voltage is across the fifth resistor?

23. Determine the unspecified voltage drop(s) in each circuit of Figure 4–72. Show how to connect a voltmeter to measure each unknown voltage drop.

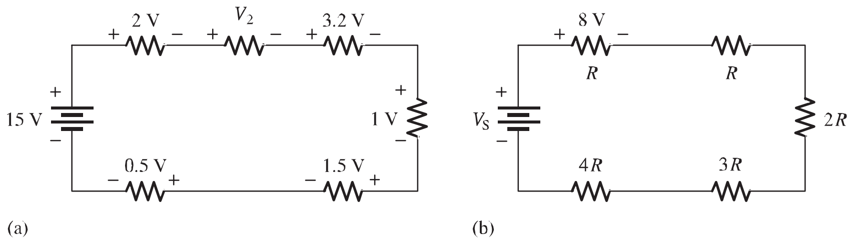

(a) (b)

FIGURE 4–72

SECTION 4–7 Voltage Dividers

24. The total resistance of a series circuit is 500 Ω. What percentage of the total voltage appears across a 22 Ω resistor in the series circuit?

25. Find the voltage between A and B in each voltage divider of Figure 4–73.

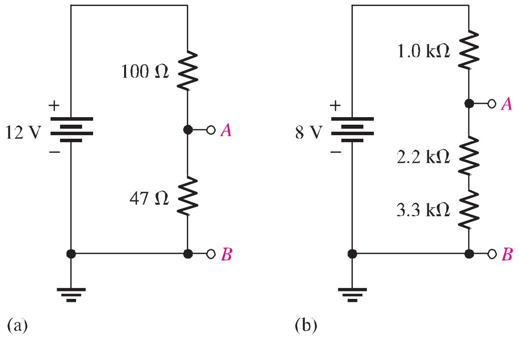

(a) (b)

FIGURE 4–73

26. Determine the voltage with respect to ground for output A, B, and C in Figure 4–74(a).

27. Determine the minimum and maximum output voltage from the voltage divider in Figure 4–74(b).

28. What is the voltage across each resistor in Figure 4–75? R is the lowest value and all others are multiples of that value as indicated.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

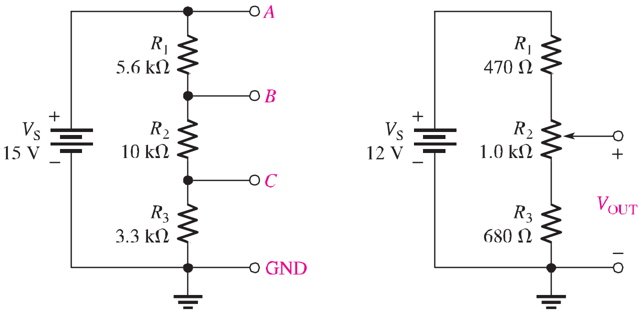
A
R_1
5.6 kΩ
B
V_S
15 V
+
–
R_2
10 kΩ
C
R_3
3.3 kΩ
GND
R_1
470 Ω
V_S
12 V
R_2
1.0 kΩ
+
V_{OUT}
R_3
680 Ω
–

(a)

(b)

FIGURE 4–74

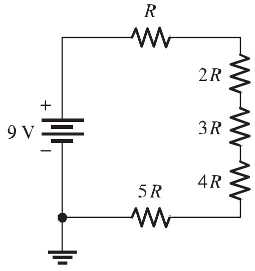
R
2R
3R
4R
5R
9 V
+
–

FIGURE 4–75

29. What is the voltage across each resistor on the protoboard in Figure 4–76(b)?

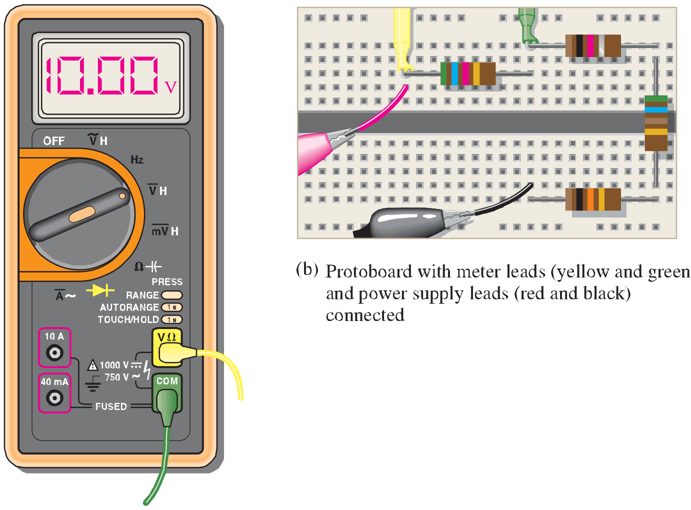
10.00 V
OFF
Hz
mV
PRESS
RANGE
AUTORANGE
TOUCH/HOLD
10 A
40 mA
1000 V
750 V
FUSED
VΩ
COM

(a) Meter with leads going to protoboard

(b) Protoboard with meter leads (yellow and green) and power supply leads (red and black) connected

FIGURE 4–76

SECTION 4–8 Power in Series Circuits

30. Five series resistors each dissipate 50 mW of power. What is the total power?

31. Find the total power in Figure 4–76.

SECTION 4–9 Voltage Measurements

32. Determine the voltage at each point with respect to ground in Figure 4–77.

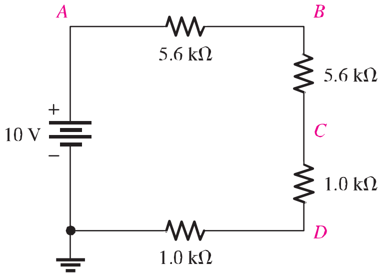
A
B
5.6 kΩ
5.6 kΩ
10 V
+
–
C
1.0 kΩ
D
1.0 kΩ

FIGURE 4–77

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

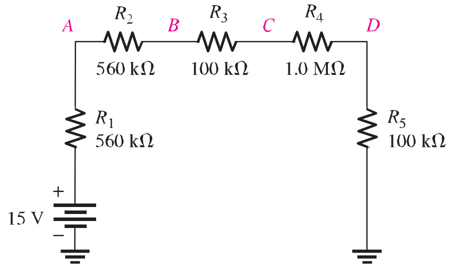
A
R_2
B
R_3
C
R_4
D
560 kΩ
100 kΩ
1.0 MΩ
R_1
560 kΩ
R_5
100 kΩ
+
15 V
−

FIGURE 4–78

33. In Figure 4–78, how would you determine the voltage across R_2 by measurement, without connecting a meter directly across the resistor?

34. Determine the voltage at each point with respect to ground in Figure 4–78.

35. In Figure 4–78, what is V_{AC}?

36. In Figure 4–78, what is V_{CA}?

SECTION 4–10 Troubleshooting

37. By observing the meters in Figure 4–79, determine the types of failures in the circuits and which components have failed.

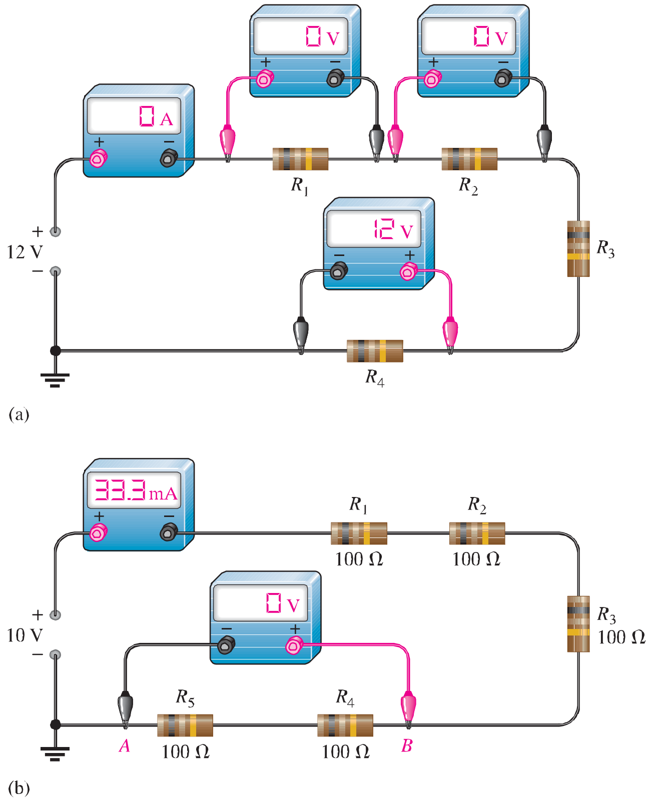
0 V
0 V
0 A
+
−
R_1
R_2
R_3
+
12 V
−
12 V
R_4
33.3 mA
R_1
R_2
100 Ω
100 Ω
R_3
100 Ω
+
10 V
−
0 V
R_5
R_4
A
100 Ω
100 Ω
B

(a)

(b)

FIGURE 4–79

38. Is the multimeter reading in Figure 4–80 correct? If not, what is wrong?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

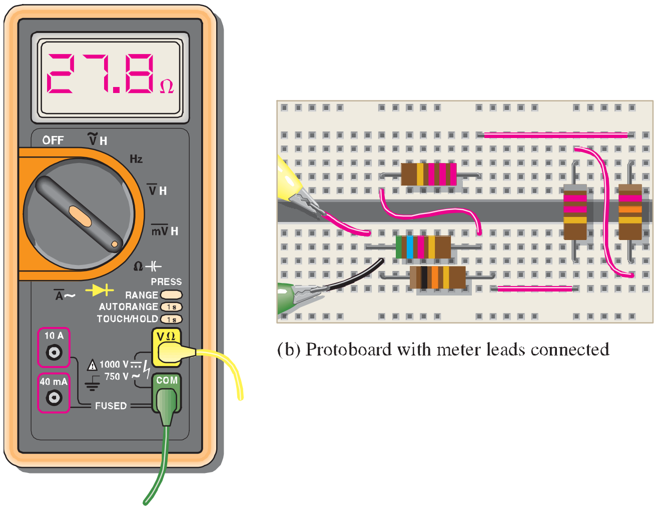
27.8 Ω
OFF
Ṽ H
Hz
V̄ H
mV̄ H
Ω ⊣⊢
PRESS
Ā~
RANGE
AUTORANGE
TOUCH/HOLD
10 A
40 mA
VΩ
1000 V
750 V ~
COM
FUSED

(b) Protoboard with meter leads connected

(a) Meter with leads going to protoboard

FIGURE 4–80

ADVANCED PROBLEMS

39. Determine the unknown resistance (R_3) in the circuit of Figure 4–81.

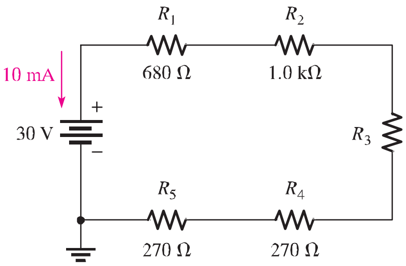
R_1
680 Ω
R_2
1.0 kΩ
10 mA
+
30 V
–
R_3
R_5
270 Ω
R_4
270 Ω

FIGURE 4–81

40. You have the following resistor values available to you in the lab in unlimited quantities: 10 Ω, 100 Ω, 470 Ω, 560 Ω, 680 Ω, 1.0 kΩ, 2.2 kΩ, and 5.6 kΩ. All of the other standard values are out of stock. A project that you are working on requires an 18 kΩ resistance. What combination of available values can you use to obtain the needed value?

41. Determine the voltage at each point in Figure 4–82 with respect to ground.

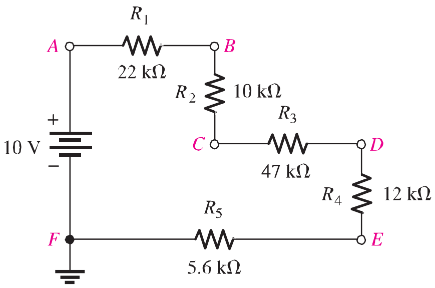
R_1
A
22 kΩ
B
R_2
10 kΩ
R_3
+
10 V
–
C
47 kΩ
D
R_4
12 kΩ
R_5
F
5.6 kΩ
E

FIGURE 4–82

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

42. Find all the unknown quantities (shown in red) in Figure 4–83.

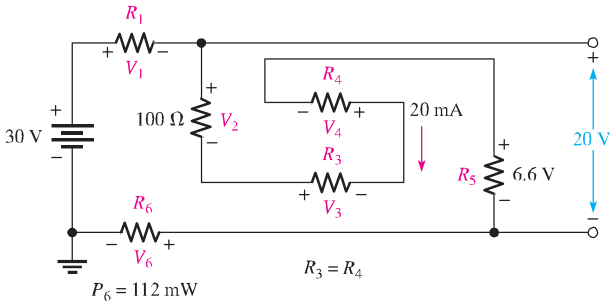

FIGURE 4–83

43. There are 250 mA in a series circuit with a total resistance of 1.5 kΩ. The current must be reduced by 25%. Determine how much resistance must be added in order to accomplish this reduction in current.

6238262 2013/12/10 12.27.66.5

44. Four $\frac{1}{2}$ W resistors are in series: 47 Ω, 68 Ω, 100 Ω, and 120 Ω. To what maximum value can the current be raised before the power rating of one of the resistors is exceeded? Which resistor will burn out first if the current is increased above the maximum?

45. A certain series circuit is made up of a $\frac{1}{8}$ W resistor, a $\frac{1}{4}$ W resistor, and a $\frac{1}{2}$ W resistor. The total resistance is 2400 Ω. If each of the resistors is operating at its maximum power level, determine the following:

(a) I **(b)** V_S **(c)** the value of each resistor

46. Using 1.5 V batteries, a switch, and three lamps, devise a circuit to apply 4.5 V across one lamp, two lamps in series, or three lamps in series with a single control switch. Draw the schematic.

47. Develop a variable voltage divider to provide output voltages ranging from a minimum of 10 V to a maximum of 100 V using a 120 V source. The maximum voltage must be at the maximum resistance setting of the potentiometer. The minimum voltage must be at the minimum resistance (zero ohms) setting. The current is to be 10 mA.

48. Using the standard resistor values given in Appendix A, design a voltage divider to provide the following approximate voltages with respect to the negative terminal of a 30 V source: 8.18 V, 14.7 V, and 24.6 V. The current drain on the source must be limited to no more than 1 mA. The number of resistors, their resistance values, and their power ratings must be specified. Draw a schematic showing the circuit with all resistor values indicated.

49. On the double-sided PC board in Figure 4–84, identify each group of series resistors and determine its total resistance. Note that many of the interconnections feed through the board from the top side to the bottom side.

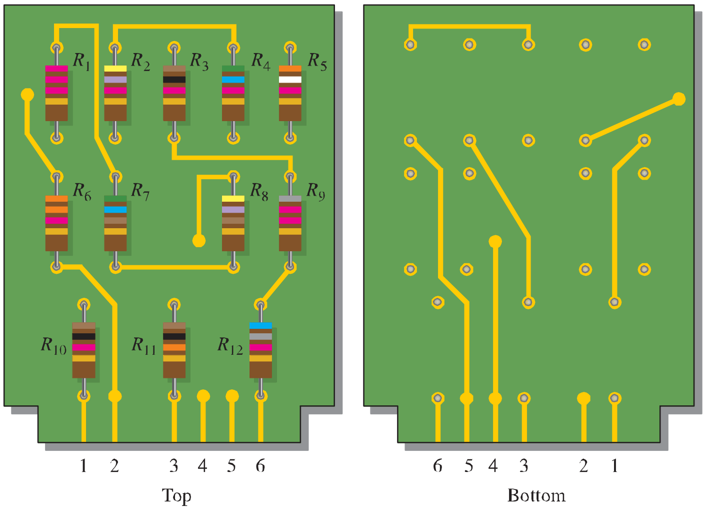

FIGURE 4–84

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

50. What is the total resistance from *A* to *B* for each switch position in Figure 4–85?

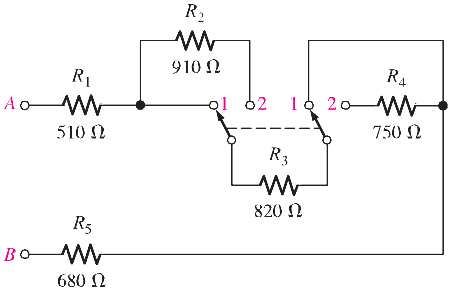
R_2
910 Ω
R_1
A
510 Ω
1
2
1
2
R_4
750 Ω
R_3
820 Ω
R_5
B
680 Ω

FIGURE 4–85

51. Determine the current measured by the meter in Figure 4–86 for each switch position.

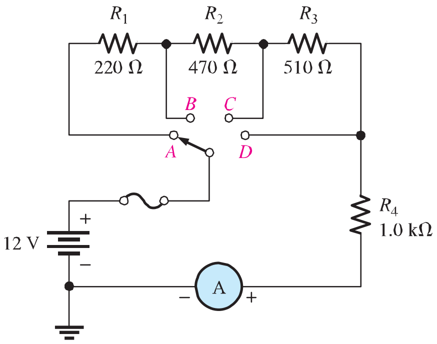
R_1
R_2
R_3
220 Ω
470 Ω
510 Ω
B
C
A
D
+
12 V
−
R_4
1.0 kΩ
A
−
+

FIGURE 4–86

52. Determine the current measured by the meter in Figure 4–87 for each position of the ganged switch.

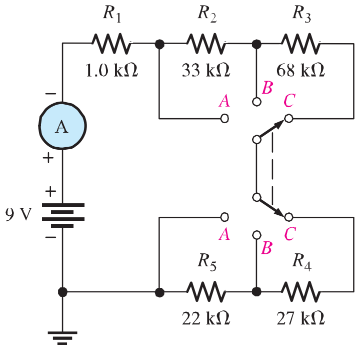
R_1
R_2
R_3
1.0 kΩ
33 kΩ
68 kΩ
−
A
+
B
A
C
+
9 V
−
A
B
C
R_5
R_4
22 kΩ
27 kΩ

FIGURE 4–87

53. Determine the voltage across each resistor in Figure 4–88 for each switch position if the current through R_5 is 6 mA when the switch is in the *D* position.

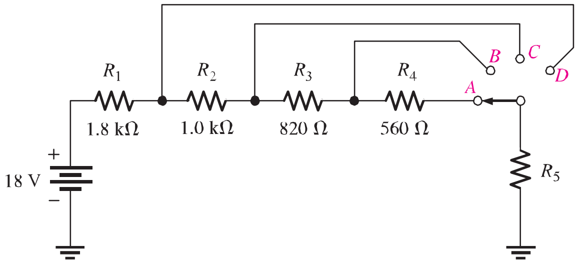
R_1
R_2
R_3
R_4
B
C
D
A
1.8 kΩ
1.0 kΩ
820 Ω
560 Ω
+
18 V
−
R_5

FIGURE 4–88

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

54. Table 4–1 shows the results of resistance measurements on the PC circuit board in Figure 4–84. Are these results correct? If not, identify the possible problems.

TABLE 4–1

BETWEEN PINS	RESISTANCE
1 and 2	∝
1 and 3	∝
1 and 4	4.23 kΩ
1 and 5	∝
1 and 6	∝
2 and 3	23.6 kΩ
2 and 4	∝
2 and 5	∝
2 and 6	∝
3 and 4	∝
3 and 5	∝
3 and 6	∝
4 and 5	∝
4 and 6	∝
5 and 6	19.9 kΩ

55. You measure 15 kΩ between pins 5 and 6 on the PC board in Figure 4–84. Does this indicate a problem? If so, identify it.

56. In checking out the PC board in Figure 4–84, you measure 17.83 kΩ between pins 1 and 2. Also, you measure 13.6 kΩ between pins 2 and 4. Does this indicate a problem on the PC board? If so, identify the fault.

57. The three groups of series resistors on the PC board in Figure 4–84 are connected in series with each other to form a single series circuit by connecting pin 2 to pin 4 and pin 3 to pin 5. A voltage source is connected across pins 1 and 6 and an ammeter is placed in series. As you increase the source voltage, you observe the corresponding increase in current. Suddenly, the current drops to zero and you smell smoke. All resistors are $\frac{1}{2}$ W.

(a) What has happened?
(b) Specifically, what must you do to fix the problem?
(c) At what voltage did the failure occur?

58. Refer to System Example 4–1. If the field coil should open while the motor is running, what happens?

59. Refer to System Example 4–1. Assume the rotor has a resistance of 2 Ω and resistors R_1, R_2, and R_3 are each 10 Ω. Assuming V_S is 12 V, what is the current when the motor first starts (in position 1)?

60. Refer to System Example 4–2. Assume the system voltage is maximum (588 V for 17 panels) and power is 4250 W.

(a) What is the current in the first module?
(b) How does it compare to the current in the last module?

61. Refer to System Example 4–2. If one of the panels is shaded and can only source 5.0 A, what will happen to the total current from the array?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

MULTISIM TROUBLESHOOTING PROBLEMS

MULTISIM

62. Open file P04-62: Files are found at www.pearsonhighered.com/floyd. Determine if there is a fault and, if so, identify the fault.

63. Open file P04-63. Determine if there is a fault and, if so, identify the fault.

64. Open file P04-64. Determine if there is a fault and, if so, identify the fault.

65. Open file P04-65. Determine if there is a fault and, if so, identify the fault.

66. Open file P04-66. Determine if there is a fault and, if so, identify the fault.

67. Open file P04-67. Determine if there is a fault and, if so, identify the fault.

6238262 2013/12/10 12.27.66.5

ANSWERS TO SECTION CHECKUPS

SECTION 4–1 Resistors in Series

1. Series resistors are connected end-to-end in a "string."
2. There is a single current path in a series circuit.
3. See Figure 4–89.

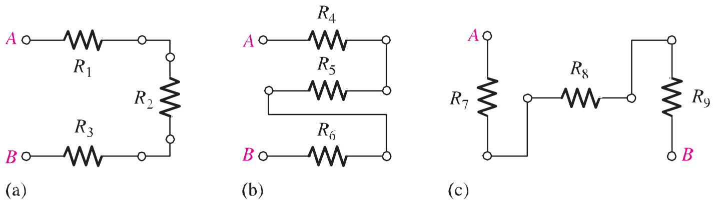

(a) (b) (c)

FIGURE 4–89

4. See Figure 4–90.

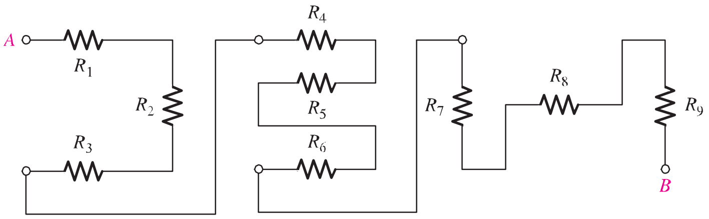

FIGURE 4–90

SECTION 4–2 Total Series Resistance

1. **(a)** $R_T = 33\ \Omega + 100\ \Omega + 10\ \Omega = 143\ \Omega$
   (b) $R_T = 39\ \Omega + 56\ \Omega + 10\ \Omega = 105\ \Omega$
   (c) $R_T = 820\ \Omega + 2200\ \Omega + 1000\ \Omega = 4020\ \Omega$
2. $R_T = 100\ \Omega + 2(47\ \Omega) + 4(12\ \Omega) + 330\ \Omega = 572\ \Omega$
3. $10\ \text{k}\Omega - 8.8\ \text{k}\Omega = 1.2\ \text{k}\Omega$
4. $R_T = 12(47\ \Omega) = 564\ \Omega$

SECTION 4–3 Current in a Series Circuit

1. The current is the same at all points in a series circuit.
2. There are 20 mA through the 47 Ω resistor.
3. 50 mA between *C* and *D*, 50 mA between *E* and *F*
4. Ammeter 1 indicates 17.9 mA. Ammeter 2 indicates 17.9 mA.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 4–4 Ohm's Law in Series Circuits

1. $I = 6\text{ V}/300\ \Omega = 0.020\text{ A} = 20\text{ mA}$
2. $V = (5\text{ mA})(43\ \Omega) = 125\text{ mV}$
3. $V_1 = (5\text{ mA})(10\ \Omega) = 50\text{ mV}$,
   $V_2 = (5\text{ mA})(15\ \Omega) = 75\text{ mV}$,
   $V_3 = (5\text{ mA})(18\ \Omega) = 90\text{ mV}$
4. $R = 1.25\text{ V}/4.63\text{ mA} = 270\ \Omega$
5. $R = 130\ \Omega$

SECTION 4–5 Voltage Sources in Series

1. $60\text{ V}/12\text{ V} = 5$; see Figure 4–91.
2. $V_T = (4)(1.5\text{ V}) = 6.0\text{ V}$
3. See Figure 4–92.

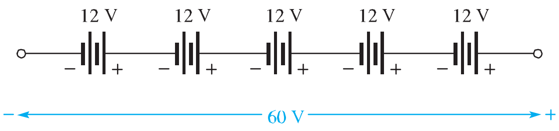

FIGURE 4–91

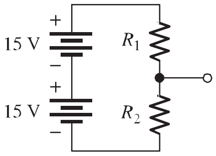

FIGURE 4–92

4. $V_{S(tot)} = 6\text{ V} + 15\text{ V} = 21\text{ V}$
5. 3.0 V

SECTION 4–6 Kirchhoff's Voltage Law

1. Kirchhoff's voltage law states:
   (a) The algebraic sum of the voltages around a closed path is zero.
   (b) The sum of the voltage drops equals the total source voltage.
2. $V_{R(tot)} = V_S = 50\text{ V}$
3. $V_{R1} = V_{R2} = 10\text{ V}/2 = 5\text{ V}$
4. $V_{R3} = 25\text{ V} - 5\text{ V} - 10\text{ V} = 10\text{ V}$
5. $V_S = 1\text{ V} + 3\text{ V} + 5\text{ V} + 7\text{ V} + 8\text{ V} = 24\text{ V}$

SECTION 4–7 Voltage Dividers

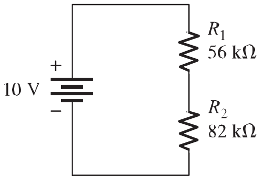

FIGURE 4–93

1. A voltage divider is a series circuit with two or more resistors in which the voltage taken across any resistor or combination of resistors is proportional to the value of that resistance.
2. Two or more resistors form a voltage divider.
3. $V_x = (R_x/R_T)V_S$ is the general voltage-divider formula.
4. $V_R = 20\text{ V}/2 = 10\text{ V}$
5. See Figure 4–93. $V_{R1} = (56\text{ k}\Omega/138\text{ k}\Omega)10\text{ V} = 4.06\text{ V}$,
   $V_{R2} = (82\text{ k}\Omega/138\text{ k}\Omega)10\text{ V} = 5.94\text{ V}$.
6. Set the potentiometer at the midpoint.

SECTION 4–8 Power in Series Circuits

1. Add the powers in each resistor to get total power.
2. $P_T = 1\text{ W} + 2\text{ W} = 5\text{ W} + 8\text{ W} = 16\text{ W}$
3. $P_T = (1\text{ mA})^2(100\ \Omega + 330\ \Omega + 680\ \Omega) = 1.11\text{ mW}$

SECTION 4–9 Voltage Measurements

1. Ground
2. −5.0 V

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

3. True
4. True

SECTION 4–10 Troubleshooting

1. An open is a break in the current path.
2. A short is a zero resistance path that bypasses a portion of a circuit.
3. Current ceases when a series circuit opens.
4. An open can be created by a component failure or a faulty contact. A short can be created by wire clippings, solder splashes, etc.
5. True
6. 24 V across the open R; 0 V across the other resistors.
7. Because R_4 is shorted, more voltage is dropped across the other resistors than normal. The total voltage is divided across three equal-value resistors.

ANSWERS TO RELATED PROBLEMS FOR EXAMPLES

4–1 **(a)** Left end of R_1 to terminal A, right end of R_1 to top end of R_3, bottom end of R_3 to right end of R_5, left end of R_5 to left end of R_2, right end of R_2 to right end of R_4, left end of R_4 to terminal B

(b) $R_1 = 1.0\text{ k}\Omega, R_2 = 33\text{ k}\Omega, R_3 = 39\text{ k}\Omega, R_4 = 470\ \Omega, R_5 = 22\text{ k}\Omega$

4–2 The two series circuits are connected in series, so all the resistors on the board are in series.

4–3 258 Ω (No change)

4–4 12.1 kΩ

4–5 22 kΩ

4–6 4.36 kΩ

4–7 18.5 mA

4–8 7.8 V

4–9 $V_1 = 1\text{ V}, V_2 = 3.3\text{ V}, V_3 = 2.2\text{ V}, V_S = 6.5\text{ V}, V_{S(max)} = 32.5\text{ V}$

4–10 275 Ω

4–11 6 V

4–12 See Figure 4–94.

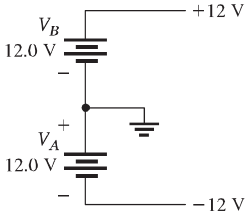

FIGURE 4–94

4–13 10 V, 20 V

4–14 6.5 V

4–15 430 Ω

4–16 $V_1 = 3.13\text{ V}; V_2 = 6.87\text{ V}$

4–17 $V_1 = V_2 = V_3 = 3.33\text{ V}$

4–18 $V_{AB} = 4\text{ V}; V_{AC} = 36.8\text{ V}; V_{BC} = 32.8\text{ V}; V_{BD} = 46\text{ V}; V_{CD} = 13.2\text{ V}$

4–19 $V_x = \left(\frac{R_x}{R_T}\right)V_S$

$$\frac{V_x}{V_S} = \frac{R_x}{R_T} = \frac{R_x}{R + R_x}$$

$$\frac{1.5\text{ V}}{4.5\text{ V}} = \frac{90\text{ k}\Omega}{R + 90\text{ k}\Omega}$$

$$1.5\text{ V}(R + 90\text{ k}\Omega) = (4.5\text{ V})(90\text{ k}\Omega)$$

$$1.5R = 270\ \Omega$$

$$R = 180\ \Omega$$

4–20 9 W

4–21 $V_A = 0\text{ V}; V_B = -25\text{ V}; V_C = -50\text{ V}; V_D = -75\text{ V}; V_E = -100\text{ V}$

4–22 3.33 V

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

ANSWERS TO TRUE/FALSE QUIZ

1. F **2.** F **3.** F **4.** T **5.** T
6. T **7.** F **8.** T **9.** T **10.** F

ANSWERS TO SELF-TEST

6238262 2013/12/10 12.27.66.5

1. (a) **2.** (d) **3.** (b) **4.** (d) **5.** (d) **6.** (a) **7.** (b)
8. (c) **9.** (b) **10.** (c) **11.** (a) **12.** (d) **13.** (d) **14.** (d)

ANSWERS TO TROUBLESHOOTING: SYMPTOM AND CAUSE

1. (b) **2.** (b) **3.** (c) **4.** (c) **5.** (a)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

CHAPTER 5

PARALLEL CIRCUITS

OUTLINE

5–1 Resistors in Parallel
5–2 Total Parallel Resistance
5–3 Voltage in a Parallel Circuit
5–4 Application of Ohm's Law
5–5 Kirchhoff's Current Law
5–6 Current Dividers
5–7 Power in Parallel Circuits
5–8 Troubleshooting

OBJECTIVES

- Identify a parallel resistive circuit
- Determine total parallel resistance
- Determine the voltage across each parallel branch
- Apply Ohm's law in a parallel circuit
- Apply Kirchhoff's current law
- Use a parallel circuit as a current divider
- Determine power in a parallel circuit
- Troubleshoot parallel circuits

KEY TERMS

Parallel
Branch
Kirchhoff's current law
Node
Current divider
Radiometric image

INTRODUCTION

In this chapter, you will see how Ohm's law is used in parallel circuits; and you will learn Kirchhoff's current law. Also, several system applications of parallel circuits, including automotive lighting, residential wiring, and control circuits are presented. You will learn how to determine total parallel resistance and how to troubleshoot for open resistors.

When resistors are connected in parallel and a voltage is applied across the parallel circuit, each resistor provides a separate path for current. The total resistance of a parallel circuit is reduced as more resistors are connected in parallel. The voltage across each of the parallel resistors is equal to the voltage applied across the entire parallel circuit.

VISIT THE WEBSITE
Study aids for this chapter are available at
http://pearsonhighered.com/floyd

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

5–1 RESISTORS IN PARALLEL

When two or more resistors are individually connected between the same two points, they are in parallel with each other. A parallel circuit provides more than one path for current.

After completing this section, you should be able to

- **Identify a parallel resistive circuit**
  - **Translate a physical arrangement of parallel resistors into a schematic**

Each parallel path in a circuit is called a **branch**. Two resistors connected in parallel are shown in Figure 5–1(a). As shown in part (b), the current out of the source (I_T) divides when it gets to point B. I_1 goes through R_1 and I_2 goes through R_2. The two currents come back together at point A. If additional resistors are connected in parallel with the first two, more current paths are provided, as shown in Figure 5–1(c). All points along the top shown in blue are electrically the same as point A, and all the points along the bottom shown in green are electrically the same as point B.

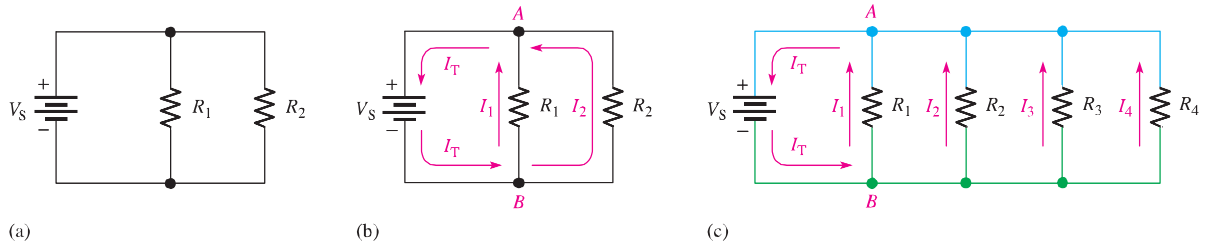

FIGURE 5–1 **Resistors in parallel.**

In Figure 5–1, it is obvious that the resistors are connected in parallel. However, in actual circuit diagrams, the parallel relationship often is not as clear. It is important that you learn to recognize parallel circuits regardless of how they may be drawn.

A rule for identifying parallel circuits is as follows:

> **If there is more than one current path (branch) between two points, and if the voltage between those two points also appears across each of the branches, then there is a parallel circuit between those two points.**

Figure 5–2 shows parallel resistors drawn in different ways between two points labeled A and B. Notice that in each case, the current has two paths going from A to B, and the voltage across each branch is the same. Although these figures show only two parallel paths, there can be any number of resistors in parallel.

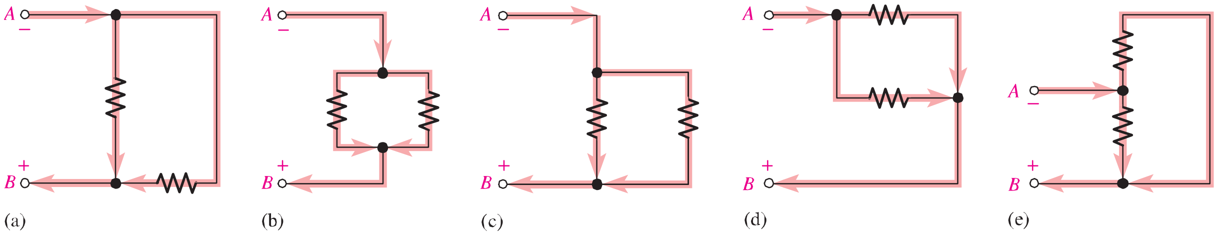

FIGURE 5–2 **Examples of circuits with two parallel paths.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 5–1

Five resistors are positioned on a protoboard as shown in Figure 5–3. Show the wiring required to connect all the resistors in parallel between *A* and *B*. Draw a schematic and label each of the resistors with its value.

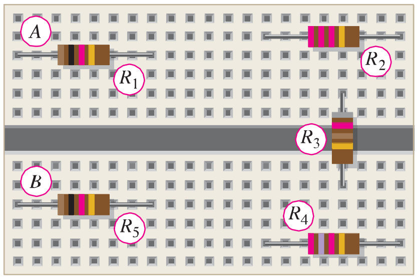

FIGURE 5–3

SOLUTION

Wires are connected as shown in the assembly wiring diagram of Figure 5–4(a). The schematic is shown in Figure 5–4(b) with resistance values from the color bands. Again, note that the schematic does not necessarily have to show the actual physical arrangement of the resistors. The schematic shows how components are connected electrically.

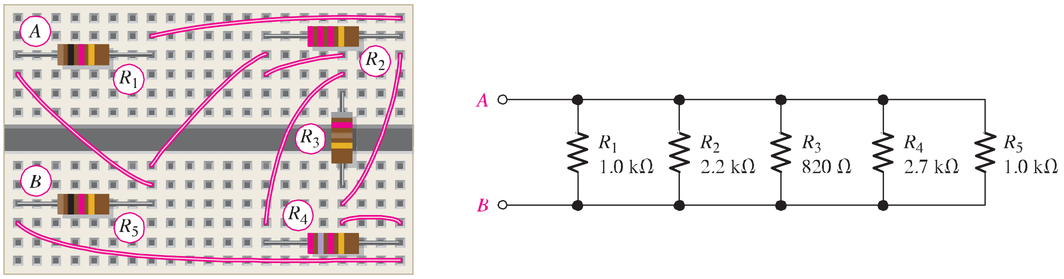

(a) Assembly wiring diagram

(b) Schematic

FIGURE 5–4

RELATED PROBLEM*

Would the circuit have to be rewired if R_2 is removed?

6238262 2013/12/10 12.27.66.5

*Answers are at the end of the chapter.

EXAMPLE 5–2

Determine the parallel groupings in Figure 5–5 and the value of each resistor.

SOLUTION

Resistors R_1 through R_4 and R_{11} and R_{12} are all in parallel. This parallel combination is connected to pins 1 and 4. Each resistor in this group is 56 kΩ.

Resistors R_5 through R_{10} are all in parallel. This combination is connected to pins 2 and 3. Each resistor in this group is 100 kΩ.

RELATED PROBLEM

How would you connect all the resistors on the PC board in parallel?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

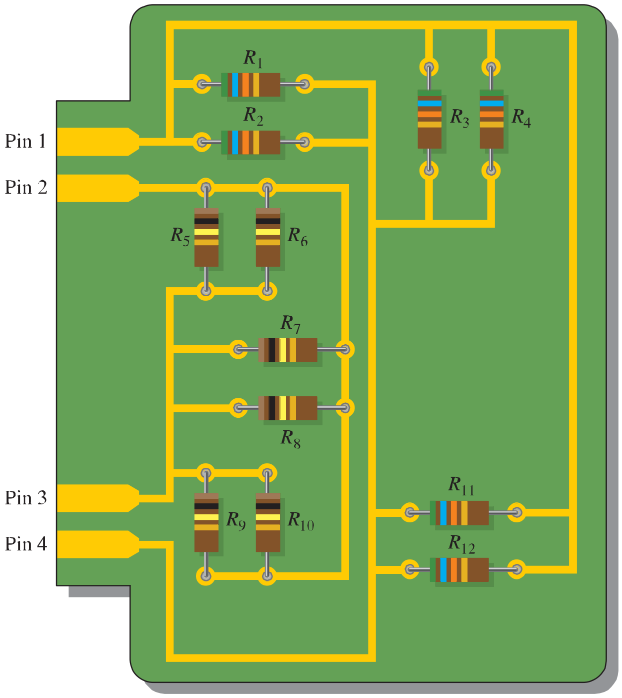

FIGURE 5–5

SECTION 5–1 CHECKUP*

1. How are the resistors connected in a parallel circuit?
2. How do you identify a parallel circuit?
3. Complete the schematics for the circuits in each part of Figure 5–6 by connecting the resistors in parallel between points *A* and *B*.
4. Connect each group of parallel resistors in Figure 5–6 in parallel with each other.

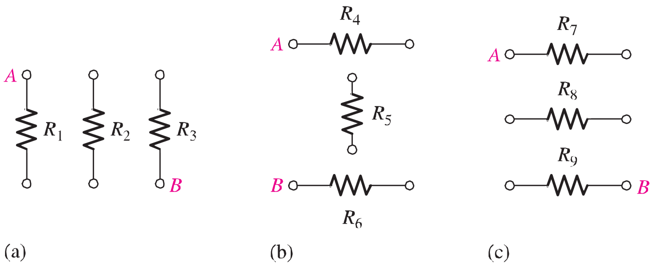

(a) (b) (c)

FIGURE 5–6

---

*Answers are at the end of the chapter.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

5–2 TOTAL PARALLEL RESISTANCE

When resistors are connected in parallel, the total resistance of the circuit decreases. The total resistance of a parallel circuit is always less than the value of the smallest resistor. For example, if a 10 Ω resistor and a 100 Ω resistor are connected in parallel, the total resistance is less than 10 Ω.

After completing this section, you should be able to

- **Determine total parallel resistance**
  - **Explain why resistance decreases as resistors are connected in parallel**
  - **Apply the parallel-resistance formula**
  - **Describe two applications of a parallel circuit**

When resistors are connected in parallel, the current has more than one path. The number of current paths is equal to the number of parallel branches.

In Figure 5–7(a), there is only one current path because it is a series circuit. There is a certain amount of current, I_1, through R_1. If resistor R_2 is connected in parallel with R_1, as shown in Figure 5–7(b), there is an additional amount of current, I_2, through R_2. The total current coming from the source has increased with the addition of the parallel branch. Assuming that the source voltage is constant, an increase in the total current from the source means that the total resistance has decreased, in accordance with Ohm's law. Additional resistors connected in parallel will further reduce the resistance and increase the total current.

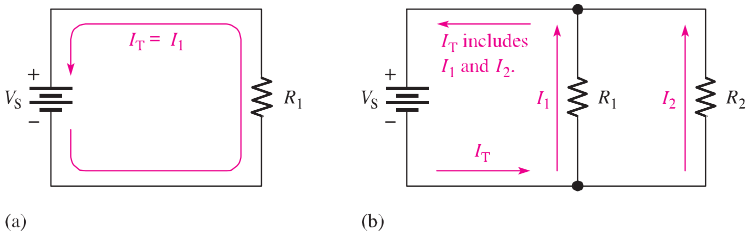

(a) (b)

FIGURE 5–7 **Connecting resistors in parallel reduces total resistance and increases total current.**

Formula for Total Parallel Resistance, R_T

The circuit in Figure 5–8 shows a general case of n resistors in parallel (n can be any integer greater than 1). As resistors are added, there are more paths for current; hence, there is increased conductance. Recall that conductance (G) was defined in Section 2–5 as the reciprocal of resistance, ($1/R$). It has the unit of siemens (S).

With parallel resistors, it is simple to think in terms of the conductive paths; each resistor adds to the total conductance as given by the following equation:

$$G_T = G_1 + G_2 + G_3 + \cdots + G_n$$

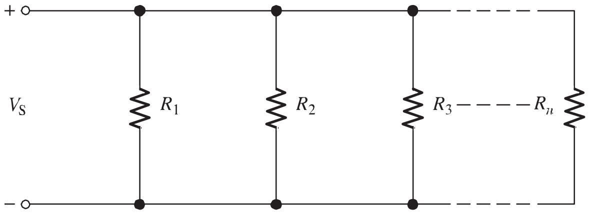

FIGURE 5–8 **Circuit with n resistors in parallel.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Substituting $1/R$ for G,

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots + \frac{1}{R_n}$$

Solve for R_T by taking the reciprocal of (that is, by inverting) both sides of the equation.

$$\boldsymbol{R_T = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots + \frac{1}{R_n}}} \qquad (5\text{–}1)$$

Equation 5–1 shows that to find the total parallel resistance, add all the $1/R$ (or conductance, G) terms and then take the reciprocal of the sum.

$$R_T = \frac{1}{G_T}$$

EXAMPLE 5–3

Calculate the total parallel resistance between points A and B of the circuit in Figure 5–9.

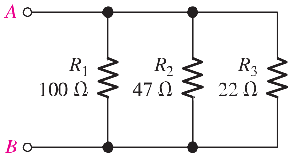

FIGURE 5–9

SOLUTION

Use Equation 5–1 to calculate the total resistance when you know the individual resistances. First, find the conductance, which is the reciprocal of the resistance, of each of the three resistors.

$$G_1 = \frac{1}{R_1} = \frac{1}{100\ \Omega} = 10\text{ mS}$$

$$G_2 = \frac{1}{R_2} = \frac{1}{47\ \Omega} = 21.3\text{ mS}$$

$$G_3 = \frac{1}{R_3} = \frac{1}{22\ \Omega} = 45.5\text{ mS}$$

Next, calculate R_T by adding G_1, G_2, and G_3 and taking the reciprocal of the sum.

$$R_T = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}} = \frac{1}{G_1 + G_2 + G_3}$$

$$= \frac{1}{10\text{ mS} + 21.3\text{ mS} + 45.5\text{ mS}} = \frac{1}{76.8\text{ mS}} = \mathbf{13.0\ \Omega}$$

For a rough accuracy check, notice that the value of R_T (13.0 Ω) is smaller than the smallest value in parallel, which is R_3 (22 Ω), as it should be.

RELATED PROBLEM

If a 33 Ω resistor is connected in parallel in Figure 5–9, what is R_T?

6238262 2013/12/10 12.27.66.5

CALCULATOR TIP The parallel-resistance formula is easily solved on a calculator using Equation 5–1. The general procedure is to enter the value of R_1 and then take its reciprocal by pressing the x^{-1} key (x^{-1} is a secondary function on some calculators). The notation x^{-1} means $1/x$. Some calculators use $1/x$ instead of x^{-1}. Next press the + key; then enter the value of R_2 and take its reciprocal. Repeat this procedure until all of the resistor values have been entered and the reciprocal of each has been added. The final step is to press the x^{-1} key to convert $1/R_T$ to R_T. The total parallel resistance is now on the display. The display format may vary, depending on the particular calculator.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 5–4

Show the steps required for a typical calculator solution of Example 5–3.

SOLUTION

1. Enter 100. Display shows 100.
2. Press the x^{-1} key (or the 2nd then x^{-1} keys).
3. Press the + key.
4. Enter 47.
5. Press the x^{-1} key.
6. Press the + key.
7. Enter 22.
8. Press the x^{-1} key.
9. Press the ENTER key. Display shows 76.7311411992E^{-3}.
10. Press the x^{-1} and ENTER keys. Display shows 13.0325182758E^{0}.

The number displayed in Step 10 is the total resistance, which is rounded to 13.0 Ω.

RELATED PROBLEM

Show the additional calculator steps for R_T when a 33 Ω resistor is placed in parallel in the circuit of Example 5–3.

THE CASE OF TWO RESISTORS IN PARALLEL Equation 5–1 is a general formula for finding the total resistance of any number of resistors in parallel. It is often useful to consider only two resistors in parallel because this situation occurs commonly in practice.

Derived from Equation 5–1, the formula for two resistors in parallel is

$$R_T = \frac{R_1 R_2}{R_1 + R_2} \qquad (5\text{–}2)$$

Equation 5–2 states

> **The total resistance for two resistors in parallel is equal to the product of the two resistors divided by the sum of the two resistors.**

This equation is sometimes referred to as the "product over the sum" formula.

EXAMPLE 5–5

Calculate the total resistance connected to the voltage source of the circuit in Figure 5–10.

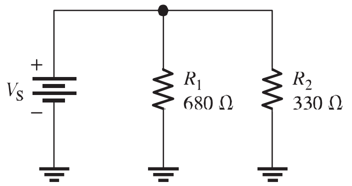

FIGURE 5–10

SOLUTION

Use Equation 5–2.

$$R_T = \frac{R_1 R_2}{R_1 + R_2} = \frac{(680\ \Omega)(330\ \Omega)}{680\ \Omega + 330\ \Omega} = \frac{224{,}000\ \Omega^2}{1{,}010\ \Omega} = \mathbf{222\ \Omega}$$

RELATED PROBLEM

Find R_T if a 220 Ω resistor replaces R_1 in Figure 5–10.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

THE CASE OF EQUAL-VALUE RESISTORS IN PARALLEL Another special case of parallel circuits is the parallel connection of several resistors having the same value. The following is a shortcut method of calculating R_T when this case occurs:

$$\boldsymbol{R_T = \frac{R}{n}} \qquad (5\text{–}3)$$

Equation 5–3 says that when any number of resistors (n), all having the same resistance (R), are connected in parallel, R_T is equal to the resistance divided by the number of resistors in parallel.

EXAMPLE 5–6

Find the total resistance between points A and B in Figure 5–11.

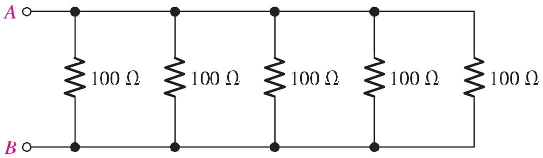

FIGURE 5–11

SOLUTION

There are five 100 Ω resistors in parallel. Use Equation 5–3.

$$R_T = \frac{R}{n} = \frac{100\ \Omega}{5} = \mathbf{20\ \Omega}$$

RELATED PROBLEM

Find R_T for three 100 kΩ resistors in parallel.

NOTATION FOR PARALLEL RESISTORS Sometimes, for convenience, parallel resistors are designated by two parallel vertical marks. For example, R_1 in parallel with R_2 can be written as $R_1 \| R_2$. Also, when several resistors are in parallel with each other, this notation can be used. For example,

$$R_1 \| R_2 \| R_3 \| R_4 \| R_5$$

indicates that R_1 through R_5 are all in parallel.

This notation is also used with resistance values. For example,

$$10\ \text{k}\Omega \| 5\ \text{k}\Omega$$

means that a 10 kΩ resistor is in parallel with a 5 kΩ resistor.

Applications of Parallel Circuits

AUTOMOTIVE One advantage of a parallel circuit over a series circuit is that when one branch opens, the other branches are not affected. For example, Figure 5–12 shows a simplified diagram of an automobile lighting system. When one headlight on your car goes out, it does not cause the other lights to go out because they are all in parallel.

Notice that the brake lights are switched on independently of the headlights and taillights. They come on only when the driver closes the brake light switch by depressing the brake pedal. When the lights switch is closed, both headlights and both taillights are on. When the headlights are on, the parking lights are off and vice versa, as indicated by the dashed line connecting the switches. If any one of the lights burns out (opens), there is still current in each of the other lights. The back-up lights are switched on when the reverse gear is engaged.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

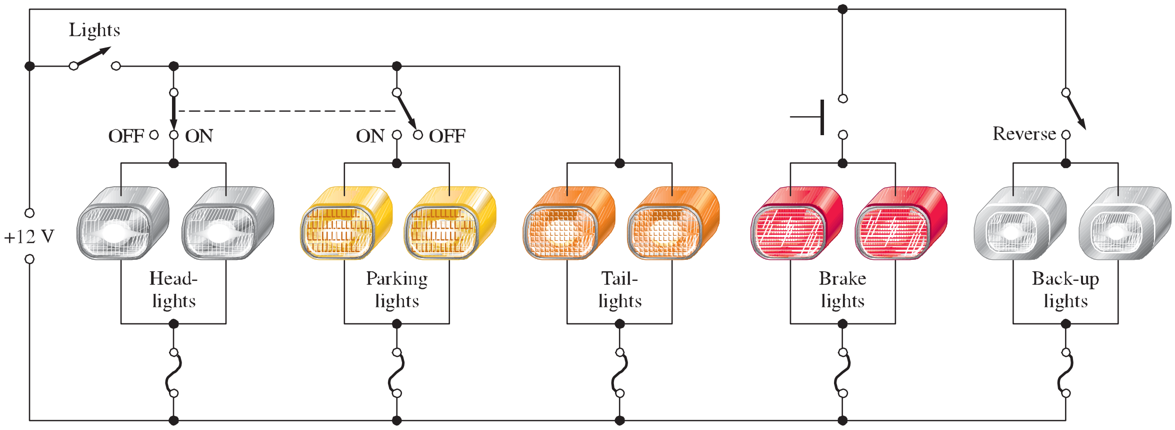

FIGURE 5–12 **Simplified diagram of the exterior light system of an automobile.**

Another automotive application of parallel resistances is the rear window defroster. As you know, power is dissipated in resistance in the form of heat. The defroster consists of a group of parallel resistance wires that warm the glass when power is applied. A typical defroster can dissipate over 100 W across the window. Although resistance heating is not as efficient as other forms of heat, in applications like this it is simple and cost effective.

RESIDENTIAL Another common use of parallel circuits is in residential electrical systems. All the lights and appliances in a home are wired in parallel. Figure 5–13 shows a typical room wiring arrangement with two switch-controlled lights and three wall outlets in parallel.

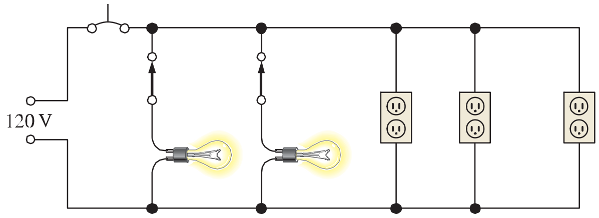

FIGURE 5–13 **Example of a parallel circuit in residential wiring.**

CONTROL CIRCUITS Many control systems use parallel circuits or equivalent circuits for control and monitoring an industrial process such as a production line. Most complex control applications are implemented on a dedicated computer called a **programmable logic controller (PLC)**. A PLC shows an equivalent parallel circuit, which is displayed on a computer screen. Internally, the circuit may exist only as a computer code, written in a computer programming language. However, the displayed circuit is one that could in principle be constructed with hardware. These circuits can be drawn to resemble a ladder—with rungs representing the load (and a source) and the rails representing the two conductors that are connected to the voltage source. (For example, notice that Figure 5–13 resembles a ladder on its side, with the lights and wall outlets as loads.) Parallel control circuits use **ladder diagrams** but add additional control elements (switches, relays, timers, and more). The result of adding the control elements creates a logical diagram, which is referred to as **ladder logic**. Ladder logic is easy to understand, so it is popular for showing control logic in industrial environments, such as factories or food processors. Entire books are written about ladder logic, but a ladder diagram (parallel circuit) is at the core of all ladder logic. The most important aspect is that ladder logic shows the basic functionality of the circuit in an easy-to-read format.

6238262 2013/12/10 12.27.66.5

In addition to ladder diagrams used in industrial control problems, many auto repair manuals use these diagrams for showing circuits for troubleshooting purposes. Ladder diagrams are logical, so a technician can organize a test procedure by reading the diagram.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 5–2 CHECKUP

1. Does the total resistance increase or decrease as more resistors are connected in parallel?
2. The total parallel resistance is always less than what value?
3. Determine R_T (between pin 1 and pin 4) for the circuit in Figure 5–14. Note that pins 1 and 2 are connected and pins 3 and 4 are connected.

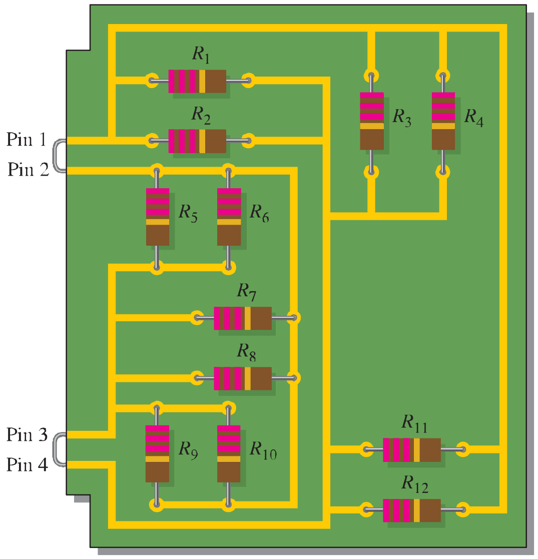

FIGURE 5–14

5–3 VOLTAGE IN A PARALLEL CIRCUIT

The voltage across any given branch of a parallel circuit is equal to the voltage across each of the other branches in parallel. As you know, each current path in a parallel circuit is called a *branch.*

After completing this section, you should be able to

- **Determine the voltage across each branch in a parallel circuit**
- **Explain why the voltage is the same across all parallel resistors**

To illustrate voltage in a parallel circuit, let's examine Figure 5–15(a). Points *A*, *B*, *C*, and *D* along the left side of the parallel circuit are electrically the same point because the voltage is the same along this line. You can think of all of these points as being connected

FIGURE 5–15 **Voltage across parallel branches is the same.**

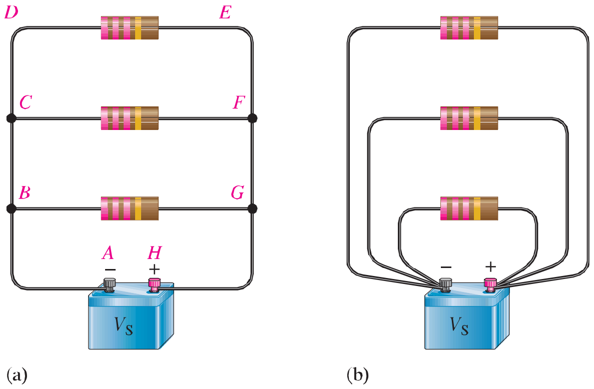

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

by a single wire to the negative terminal of the battery. The points *E*, *F*, *G*, and *H* along the right side of the circuit are all at a voltage equal to that of the positive terminal of the source. Thus, voltage across each parallel resistor is the same, and each is equal to the source voltage. Notice that the parallel circuit in Figure 5–15 resembles a ladder as described earlier.

Figure 5–15(b) is the same circuit as in part (a), drawn in a slightly different way. Here the left side of each resistor is connected to a single point, which is the negative battery terminal. The right side of each resistor is connected to a single point, which is the positive battery terminal. The resistors are still all in parallel across the source.

In Figure 5–16, a 12 V battery is connected across three parallel resistors. When the voltage is measured across the battery and then across each of the resistors, the readings are the same. As you can see, the same voltage appears across each branch in a parallel circuit.

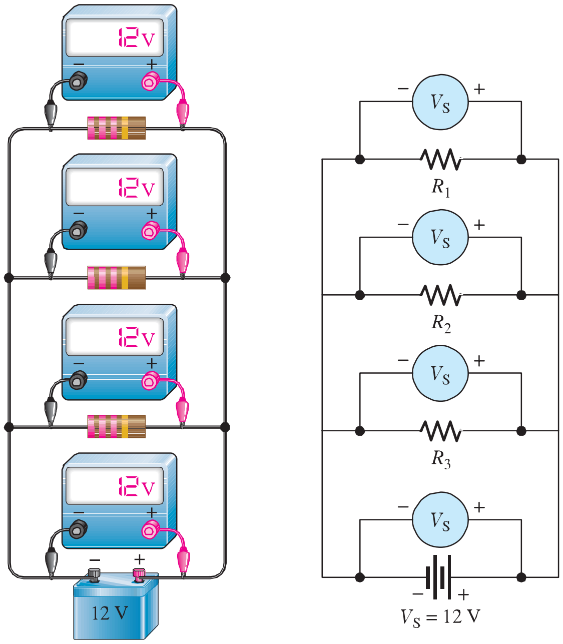

(a) Pictorial (b) Schematic

FIGURE 5–16 **The same voltage appears across each resistor in parallel.**

EXAMPLE 5–7

Determine the voltage across each resistor in Figure 5–17.

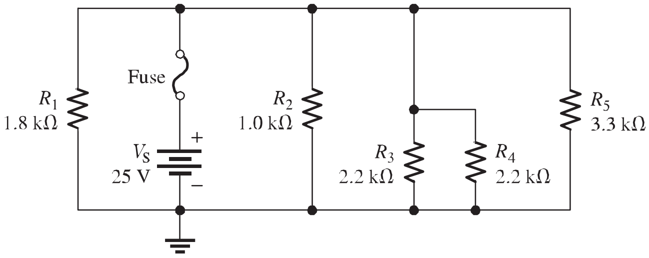

FIGURE 5–17

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

MULTISIM

Open Multisim file E05-07; files are found at www.pearsonhighered.com/floyd. Verify that each resistor voltage equals the source voltage. Repeat for the related problem.

SOLUTION

The five resistors are in parallel; so the voltage across each one is equal to the source voltage, V_S. There is no voltage across the fuse.

$$V_1 = V_2 = V_3 = V_4 = V_5 = V_S = \mathbf{25\ V}$$

RELATED PROBLEM

If R_4 is removed from the circuit, what is the voltage across R_3?

6238262 2013/12/10 12 27 66 5

SYSTEM EXAMPLE 5–1

DUMMY LOADS AND LOAD BANKS

A **dummy load** is a precision high-power resistor and may include multiple resistors in one enclosure. A dummy load is used to test a system such as a radio transmitter without the actual load (the antenna) in place. A *load bank* performs the same function electronically, but it absorbs a certain amount of current, which is controlled to represent a specific resistance. Both a dummy load and a load bank absorb energy from the system. For radio transmitters, the energy from the transmitter that is normally routed to the antenna system as electromagnetic energy is instead routed to the dummy load, where it is dissipated in the form of heat. Without an antenna or a dummy load, the transmitter could be damaged. The dummy load enables the transmitter to be tested under simulated conditions without causing interference to other systems as adjustments are made. In addition to testing radio transmitters, dummy loads can allow testing of other systems. For example, load cells can be tested with a dummy load for calibration purposes.

Figure 5–18(a) shows the schematic for a dummy load for testing a small radio transmitter and is composed of 1 W resistors (R_{11} and R_{12} can be 0.5 W each). Resistors R_{11} and R_{12} form a voltage divider, which is in parallel with the other resistors in the array. The output from the divider is rectified and measured with a voltmeter that can be calibrated for power. To match a 50 Ω antenna system, the total resistance is 50 Ω as you can determine from the values given. The entire assembly is contained in a shielded enclosure to avoid radiating any electromagnetic energy. The enclosure contains a meter to read the power in the load, as shown in Figure 5–18(b).

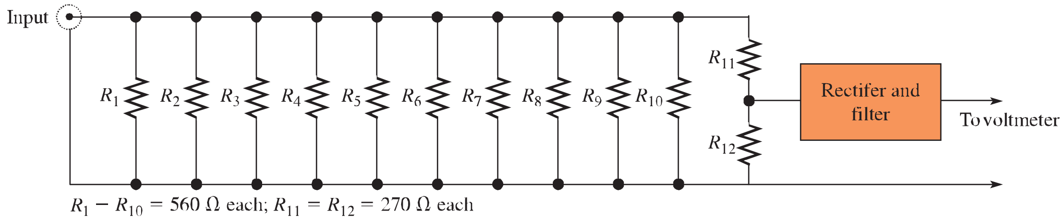

(a) Schematic

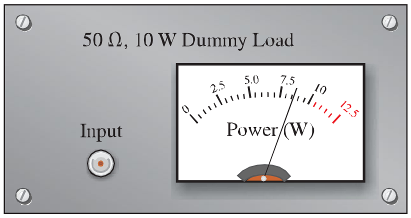

(b) Front panel of dummy load

FIGURE 5–18 **A dummy load for testing a small transmitter.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 5–3 CHECKUP

1. A 100 Ω and a 220 Ω resistor are connected in parallel with a 5 V source. What is the voltage across each of the resistors?
2. A voltmeter connected across R_1 in Figure 5–19 measures 118 V. If you move the meter and connect it across R_2, how much voltage will it indicate? What is the source voltage?

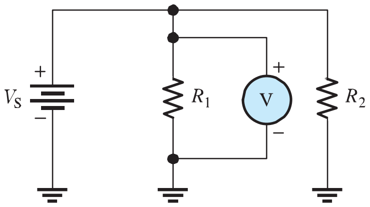

FIGURE 5–19

3. In Figure 5–20, how much voltage does voltmeter 1 indicate? Voltmeter 2?

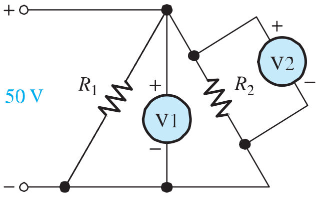

FIGURE 5–20

4. How are voltages across each branch of a parallel circuit related?

5–4 APPLICATION OF OHM'S LAW

Ohm's law can be applied to parallel circuit analysis.

After completing this section, you should be able to

- **Apply Ohm's law in a parallel circuit**
  - **Find the total current in a parallel circuit**
  - **Find the branch currents, voltage, and resistance in a parallel circuit**

The following examples illustrate how to apply Ohm's law in parallel circuits.

EXAMPLE 5–8

Find the total current from the battery in Figure 5–21.

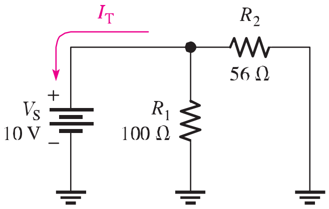

FIGURE 5–21

SOLUTION

The battery "sees" a total parallel resistance that determines the amount of current. First, calculate R_T.

$$R_T = \frac{R_1 R_2}{R_1 + R_2} = \frac{(100\ \Omega)(56\ \Omega)}{100\ \Omega + 56\ \Omega} = \frac{5600\ \Omega^2}{156\ \Omega} = 35.9\ \Omega$$

The battery voltage is 10 V. Use Ohm's law to find I_T.

$$I_T = \frac{V_S}{R_T} = \frac{10\ \text{V}}{35.9\ \Omega} = \mathbf{279\ mA}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

MULTISIM

Open Multisim file E05-08. Using the multimeter, verify the calculated values of total current and the branch currents.

RELATED PROBLEM

Find the currents through R_1 and R_2 in Figure 5–21. Show that the sum of the currents through R_1 and R_2 equals the total current.

EXAMPLE 5–9

Determine the current through each resistor in the parallel circuit of Figure 5–22.

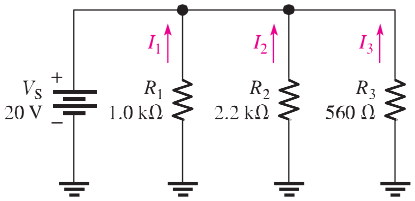

FIGURE 5–22

SOLUTION

The voltage across each resistor (branch) is equal to the source voltage. That is, the voltage across R_1 is 20 V, the voltage across R_2 is 20 V, and the voltage across R_3 is 20 V. Determine the current through each resistor as follows:

$$I_1 = \frac{V_S}{R_1} = \frac{20\text{ V}}{1.0\text{ k}\Omega} = \mathbf{20.0\text{ mA}}$$

$$I_2 = \frac{V_S}{R_2} = \frac{20\text{ V}}{2.2\text{ k}\Omega} = \mathbf{9.09\text{ mA}}$$

$$I_3 = \frac{V_S}{R_3} = \frac{20\text{ V}}{560\ \Omega} = \mathbf{35.7\text{ mA}}$$

MULTISIM

Open Multisim file E05-09. Measure the current through each resistor. Connect a 910 Ω resistor in parallel with the other resistors and measure the branch currents. How much does the total current from the source change when the new resistor is added?

RELATED PROBLEM

If an additional resistor of 910 Ω is connected in parallel to the circuit in Figure 5–22, determine all the branch circuits.

EXAMPLE 5–10

Find the voltage across the parallel circuit shown in Figure 5–23.

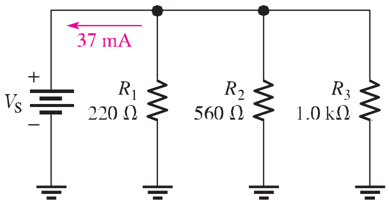

FIGURE 5–23

SOLUTION

The total current into the parallel circuit is 37 mA. If you know the total resistance, then you can apply Ohm's law to get the voltage. The total resistance is

$$R_T = \frac{1}{\dfrac{1}{R_1} + \dfrac{1}{R_2} + \dfrac{1}{R_3}} = \frac{1}{\dfrac{1}{220\ \Omega} + \dfrac{1}{560\ \Omega} + \dfrac{1}{1.0\text{ k}\Omega}}$$

$$= \frac{1}{4.55\text{ mS} + 1.79\text{ mS} + 1\text{ mS}} = \frac{1}{7.34\text{ mS}} = 136\ \Omega$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Therefore, the source voltage and the voltage across each branch is

$$V_S = I_T R_T = (37\text{ mA})(136\ \Omega) = \mathbf{5.05\ V}$$

RELATED PROBLEM

Find the total current if R_3 opens in Figure 5–23. Assume V_S remains the same.

EXAMPLE 5–11

Sometimes a direct measurement of resistance is not practical. For example, tungsten filament bulbs get hot after turning them on, causing the resistance to increase. An ohmmeter can only measure the cold resistance. Assume you want to know the equivalent hot resistance of the two headlight bulbs and the two taillight bulbs in a car. The two headlights normally operate on 12.6 V and draw 2.8 A each when they are on.

(a) What is the total equivalent hot resistance when the two headlights are on?

(b) Assume the total current for headlights and taillights is 8.0 A when all four bulbs are on. What is the equivalent resistance of each taillight?

SOLUTION

(a) Use Ohm's law to calculate the equivalent resistance of one headlight.

$$R_{\text{HEAD}} = \frac{V}{I} = \frac{12.6\text{ V}}{2.8\text{ A}} = 4.5\ \Omega$$

Because the two bulbs are in parallel and have equal resistances,

$$R_{\text{T(HEAD)}} = \frac{R_{\text{HEAD}}}{n} = \frac{4.5\ \Omega}{2} = \mathbf{2.25\ \Omega}$$

(b) Use Ohm's law to find the total resistance with two taillights and two headlights on.

$$R_{\text{T(HEAD+TAIL)}} = \frac{12.6\text{ V}}{8.0\text{ A}} = 1.58\ \Omega$$

Apply the formula for parallel resistors to find the resistance of the two taillights together.

$$\frac{1}{R_{\text{T(HEAD+TAIL)}}} = \frac{1}{R_{\text{T(HEAD)}}} + \frac{1}{R_{\text{T(TAIL)}}}$$

$$\frac{1}{R_{\text{T(TAIL)}}} = \frac{1}{R_{\text{T(HEAD+TAIL)}}} - \frac{1}{R_{\text{T(HEAD)}}} = \frac{1}{1.58\ \Omega} - \frac{1}{2.25\ \Omega}$$

$$R_{\text{T(TAIL)}} = 5.25\ \Omega$$

The two taillights are in parallel, so each light has a resistance of

6238262 2013/12/10 12.27.66.5

$$R_{\text{TAIL}} = nR_{\text{T(TAIL)}} = 2(5.25\ \Omega) = \mathbf{10.5\ \Omega}$$

RELATED PROBLEM

What is the total equivalent resistance of two headlights that each draw 3.15 A?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 5–4 CHECKUP

1. A 12 V battery is connected across three 680 Ω resistors that are in parallel. What is the total current from the battery?
2. How much voltage is required to produce 20 mA through the circuit of Figure 5–24?
3. How much current is there through each resistor of Figure 5–24?
4. There are four equal-value resistors in parallel with a 12 V source and 6 mA of current from the source. What is the value of each resistor?
5. A 1.0 kΩ and a 2.2 kΩ resistor are connected in parallel. There is a total of 100 mA through the parallel combination. How much voltage is dropped across the resistors?

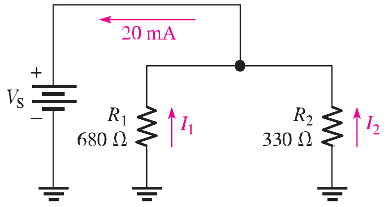

FIGURE 5–24

5–5 KIRCHHOFF'S CURRENT LAW

Kirchhoff's voltage law deals with voltages in a single closed path. Kirchhoff's current law applies to currents in multiple paths.

After completing this section, you should be able to

- **Apply Kirchhoff's current law**
  - **State Kirchhoff's current law**
  - **Define *node***
  - **Determine the total current by adding the branch currents**
  - **Determine an unknown branch current**

Kirchhoff's current law, often abbreviated KCL, can be stated as follows:

> **The sum of the currents into a node (total current in) is equal to the sum of the currents out of that node (total current out).**

A **node** is any point or junction in a circuit where two or more components are connected. In a parallel circuit, a node is a point where the parallel branches come together. For example, in the circuit of Figure 5–25, there are two nodes as indicated, node *A* and node *B*. Let's

FIGURE 5–25 **Kirchhoff's current law: The current into a node equals the current out of that node.**

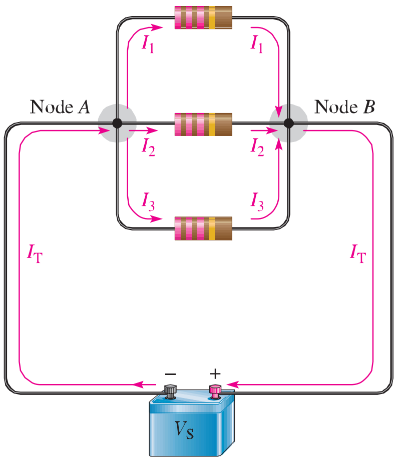

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

start at the negative terminal of the source and follow the current. The total current I_T from the source is *into* node *A*. At this point, the current splits up among the three branches as indicated. Each of the three branch currents (I_1, I_2 and I_3) is *out of* node *A*. Kirchhoff's current law says that the total current into node *A* is equal to the total current out of node *A*; that is,

$$I_T = I_1 + I_2 + I_3$$

Now, following the currents in Figure 5–25 through the three branches, you see that they come back together at node *B*. Currents I_1, I_2, and I_3 are into node *B*, and I_T is out of node *B*. Kirchhoff's current law formula at node *B* is therefore the same as at node *A*.

$$I_1 + I_2 + I_3 = I_T$$

Kirchhoff's current law can be applied to anything that acts as a load for an electrical system. For example, the power connections to a circuit board will have the same current supplied to the board as leaving the board. Likewise, a building has the same current on the incoming "hot" line as leaving on the neutral. The supply and return currents are always the same, but a fault in a system can cause return current to follow a different path other than the neutral line (a condition known as a "ground fault").

SYSTEM NOTE

By Kirchhoff's current law, the sum of the currents into a node must equal the sum of the currents out of that node. Figure 5–26 illustrates the general case of Kirchhoff's current law and can be written as a mathematical relationship.

$$\begin{aligned} &\boldsymbol{I}_{\mathbf{IN(1)}} + \boldsymbol{I}_{\mathbf{IN(2)}} + \boldsymbol{I}_{\mathbf{IN(3)}} + \cdots + \boldsymbol{I}_{\mathbf{IN}(\boldsymbol{n})} \\ &= \boldsymbol{I}_{\mathbf{OUT(1)}} + \boldsymbol{I}_{\mathbf{OUT(2)}} + \boldsymbol{I}_{\mathbf{OUT(3)}} + \cdots + \boldsymbol{I}_{\mathbf{OUT}(\boldsymbol{m})} \end{aligned} \qquad (5\text{–}4)$$

Moving the terms on the right side to the left side and changing the sign results in the following equivalent equation:

$$I_{\text{IN}(1)} + I_{\text{IN}(2)} + I_{\text{IN}(3)} + \cdots + I_{\text{IN}(n)} - I_{\text{OUT}(1)} - I_{\text{OUT}(2)} - I_{\text{OUT}(3)} - \cdots - I_{\text{OUT}(m)} = 0$$

This equation shows that all current into and out of the junction sums to zero. It is equivalent to "balancing the books" by accounting for all current into and out of a node.

An equivalent way of writing Kirchhoff's current law can be expressed using the mathematical summation shorthand as was done for Kirchhoff's voltage law in Section 4–6. To express Kirchhoff's current law with this method, we will assign all currents with a sequential subscript (1, 2, 3, and so on) no matter if the current is into or out of the node.

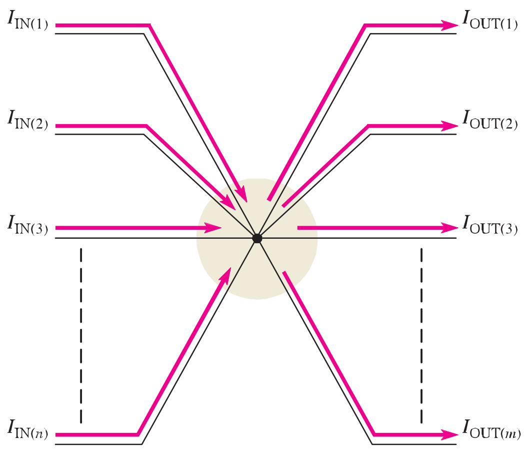

$$I_{\text{IN}(1)} + I_{\text{IN}(2)} + I_{\text{IN}(3)} + \cdots + I_{\text{IN}(n)} = I_{\text{OUT}(1)} + I_{\text{OUT}(2)} + I_{\text{OUT}(3)} + \cdots I_{\text{OUT}(m)}$$

FIGURE 5–26 **Generalized circuit node illustrates Kirchhoff's current law.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Currents into the node are assigned a positive sign and currents leaving the node are assigned a negative sign. Using this notation, Kirchhoff's current law can be stated this way:

$$\sum_{i=1}^{n} I_i = 0$$

This equation uses the mathematical symbol for summation, $\sum$, which means to add the terms from $i = 1$ to $i = n$, taking into account their signs, and set them equal to zero. It can be expressed as a statement as

> **The algebraic sum of all of the currents entering and leaving a node is equal to zero.**

You can verify Kirchhoff's current law by connecting a circuit and measuring each branch current and the total current from the source, as illustrated in Figure 5–27. *When the branch currents are added together, their sum will equal the total current.* This rule applies for any number of branches.

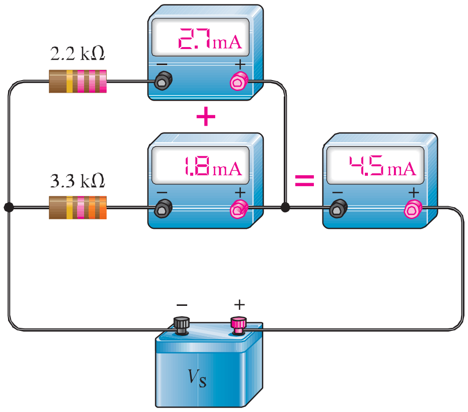

FIGURE 5–27 **An illustration of Kirchhoff's current law.**

6238262 2013/12/10 12.27.66.5

EXAMPLE 5–12

In Example 5–11, you found the equivalent resistance of the headlights and taillights of a car. Using Kirchhoff's current law, find the current in each of the two taillights, given that the total current is 8.0 A and together the headlights draw 5.6 A. Assume the lights are the only load on the battery.

SOLUTION

The current from the battery equals the current to the lights.

$$I_{\text{BAT}} = I_{\text{T(HEAD)}} + I_{\text{T(TAIL)}}$$

The total current from the battery is +8.0 A and the currents to the bulbs are negative. Rearranging,

$$I_{\text{T(TAIL)}} = I_{\text{BAT}} - I_{\text{T(HEAD)}} = 8.0\text{ A} - 5.6\text{ A} = 2.4\text{ A}$$

Because the taillights are identical, each draws $2.4\text{ A}/2 = \mathbf{1.2\text{ A}}$.

RELATED PROBLEM

Using the resistance of the taillights from Example 5–11, show that Ohm's law gives the same result.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 5–13

You know the branch currents in the circuit of Figure 5–28. Determine the total current entering node *A* and the total current leaving node *B*.

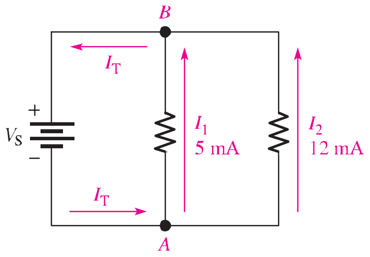

FIGURE 5–28

SOLUTION

The total current out of node *A* is the sum of the two branch currents. Therefore, the total current into node *A* is

$$I_T = I_1 + I_2 = 5\text{ mA} + 12\text{ mA} = \mathbf{17\text{ mA}}$$

The total current entering node *B* is the sum of the two branch currents. Thus, the total current out of node *B* is

$$I_T = I_1 + I_2 = 5\text{ mA} + 12\text{ mA} = \mathbf{17\text{ mA}}$$

RELATED PROBLEM

If a third resistor is connected in parallel to the circuit of Figure 5–28 and its current is 3 mA, what is the total current into node *A* and out of node *B*?

MULTISIM

Open Multisim file E05-13. Verify that the total current is the sum of the branch currents. Repeat for the related problem.

EXAMPLE 5–14

Determine the current through R_2 in Figure 5–29.

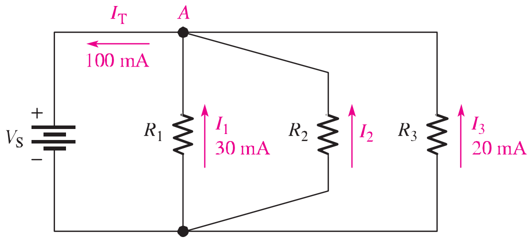

FIGURE 5–29

SOLUTION

The total current out of node *A* is $I_T = I_1 + I_2 + I_3$. From Figure 5–29, you know the total current and the branch currents through R_1 and R_3. Solve for I_2.

$$I_2 = I_T - I_1 - I_3 = 100\text{ mA} - 30\text{ mA} - 20\text{ mA} = \mathbf{50\text{ mA}}$$

RELATED PROBLEM

Determine I_T and I_2 if a fourth branch is added to the circuit in Figure 5–29 and it has 12 mA through it.

MULTISIM

Open Multisim file E05-14. Verify the calculated value for I_2. Repeat for the related problem.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 5–15

Use Kirchhoff's current law to find the current measured by ammeters A3 and A5 in Figure 5–30.

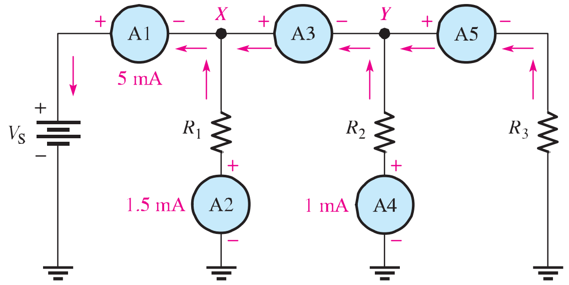

FIGURE 5–30

SOLUTION

The total current out of node X is 5 mA. Two currents are into node X: 1.5 mA through resistor R_1 and the current through A3. Kirchhoff's current law at node X is

$$5\text{ mA} = 1.5\text{ mA} + I_{A3}$$

Solving for I_{A3} yields

$$I_{A3} = 5\text{ mA} - 1.5\text{ mA} = \mathbf{3.5\text{ mA}}$$

The total current out of node Y is $I_{A3} = 3.5$ mA. Two currents are into node Y: 1 mA through resistor R_2 and the current through A5 and R_3. Kirchhoff's current law applied to node Y gives

$$3.5\text{ mA} = 1\text{ mA} + I_{A5}$$

Solving for I_{A5} yields

$$I_{A5} = 3.5\text{ mA} - 1\text{ mA} = \mathbf{2.5\text{ mA}}$$

RELATED PROBLEM

How much current will an ammeter measure when it is placed in the circuit in Figure 5–30 right below R_3? Below the negative battery terminal?

SECTION 5–5 CHECKUP

1. State Kirchhoff's current law in two ways.

2. A total current of 2.5 A is into a node and the current out of the node divides into three parallel branches. What is the sum of all three branch currents?

3. In Figure 5–31, 100 mA and 300 mA are into the node. What is the amount of current out of the node?

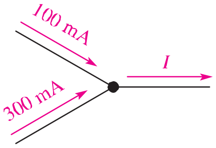

FIGURE 5–31

4. Each of the two taillights on a trailer draws 1 A and the two brake lights each draw another 1 A. What is the current in the ground lead to the trailer when all lights are on?

5. A underground pump has 10 A of current on the hot line. What should be the current on the neutral line?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

5–6 CURRENT DIVIDERS

A parallel circuit acts as a current divider because the current entering the junction of parallel branches "divides" up into several individual branch currents.

After completing this section, you should be able to

- **Use a parallel circuit as a current divider**
  - **Apply the current-divider formula**
  - **Determine an unknown branch current**

In a parallel circuit, the total current into the junction of the parallel branches (node) divides among the branches. Thus, a parallel circuit acts as a current divider. This current-divider principle is illustrated in Figure 5–32 for a two-branch parallel circuit in which part of the total current I_T goes through R_1 and part through R_2.

Since the same voltage is across each of the resistors in parallel, the branch currents are inversely proportional to the values of the resistors. For example, if the value of R_2 is twice that of R_1, then the value of I_2 is one-half that of I_1. In other words,

> **The total current divides among parallel resistors into currents with values inversely proportional to the resistance values.**

The branches with higher resistance have less current, and the branches with lower resistance have more current, in accordance with Ohm's law. If all the branches have the same resistance, the branch currents are all equal.

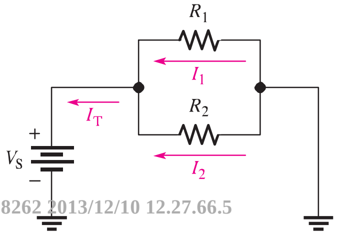

6238262 2013/12/10 12.27.66.5

FIGURE 5–32 **Total current divides between the two branches.**

Power Resistors in the TO Case

In high-current systems (up to about 40 A), a shunt resistor normally has a very small resistance. Small-value resistors with a large power rating are available in various case configurations that are designed to be mounted to a heat sink. The resistor shown is in a TO style case. Resistors as small as 5 mΩ with 1% tolerance are available in the TO case.

LTO 100
1K 1%
59

The TO case configuration allows heat to be dissipated by the heat sink and away from the sensitive electronics assembly. Power specifications depends on the thermal design and range from 1 W to 100 W. An advantage to resistors used in TO cases is that they can operate in fast-switching circuits and in high-frequency applications because there is little effect on the signal.

SYSTEM NOTE

Figure 5–33 shows specific values to demonstrate how the currents divide according to the branch resistances. Notice that in this case the resistance of the upper branch is one-tenth the resistance of the lower branch, but the upper branch current is ten times the lower branch current.

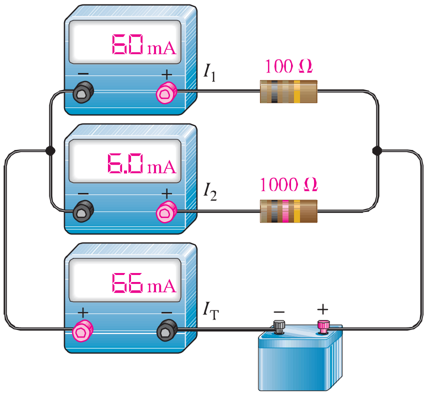

FIGURE 5–33 **The branch with the lowest resistance has the most current, and the branch with the highest resistance has the least current.**

Current-Divider Formula

You can develop a formula for determining how currents divide among any number of parallel resistors, as shown in Figure 5–34, where n is the total number of resistors.

The current through any one of the parallel resistors is I_x, where x represents the number of a particular resistor (1, 2, 3, and so on). By Ohm's law, you can express the current through any one of the resistors in Figure 5–34 as follows:

$$I_x = \frac{V_S}{R_x}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

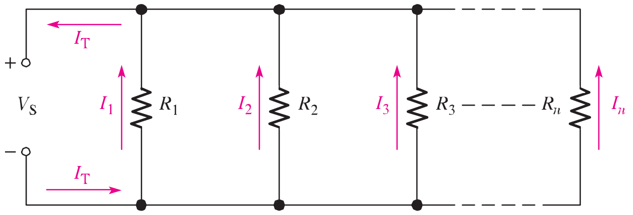

FIGURE 5–34 **A parallel circuit with *n* branches.**

The source voltage, V_S, appears across each of the parallel resistors, and R_x represents any one of the parallel resistors. The total source voltage, V_S, is equal to the total current times the total parallel resistance.

$$V_S = I_T R_T$$

Substituting $I_T R_T$ for V_S in the expression for I_x results in

$$I_x = \frac{I_T R_T}{R_x}$$

Rearranging terms yields

$$\boldsymbol{I_x = \left(\frac{R_T}{R_x}\right) I_T} \qquad (5\text{–}5)$$

where $x = 1, 2, 3$, etc. Equation 5–5 is the general current-divider formula and applies to a parallel circuit with any number of branches.

The current (I_x) through any branch equals the total parallel resistance (R_T) divided by the resistance (R_x) of that branch, and then multiplied by the total current (I_T) into the junction of parallel branches.

EXAMPLE 5–16

Determine the current through each resistor in the circuit of Figure 5–35.

FIGURE 5–35

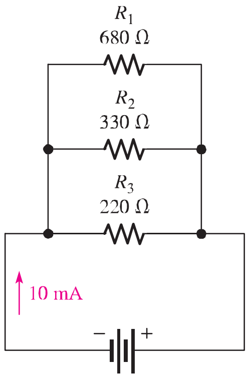

SOLUTION

First, calculate the total parallel resistance.

$$R_T = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}} = \frac{1}{\frac{1}{680\ \Omega} + \frac{1}{330\ \Omega} + \frac{1}{220\ \Omega}} = 111\ \Omega$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The total current is 10 mA. Use Equation 5–5 to calculate each branch current.

$$I_1 = \left(\frac{R_T}{R_1}\right)I_T = \left(\frac{111\ \Omega}{680\ \Omega}\right)10\text{ mA} = \mathbf{1.63\ mA}$$

$$I_2 = \left(\frac{R_T}{R_2}\right)I_T = \left(\frac{111\ \Omega}{330\ \Omega}\right)10\text{ mA} = \mathbf{3.36\ mA}$$

$$I_3 = \left(\frac{R_T}{R_3}\right)I_T = \left(\frac{111\ \Omega}{220\ \Omega}\right)10\text{ mA} = \mathbf{5.05\ mA}$$

RELATED PROBLEM

Determine the current through R_1 and R_2 in Figure 5–35 if R_3 is removed. Assume the source voltage remains the same.

CURRENT-DIVIDER FORMULAS FOR TWO BRANCHES You already know how to use Ohm's law ($I = V/R$) to determine the current in any branch in a parallel circuit when you know the voltage and resistance. When you do not know the voltage but you know total current, you can find both of the branch currents (I_1 and I_2) by using the following formulas:

$$I_1 = \left(\frac{R_2}{R_1 + R_2}\right)I_T \qquad (5–6)$$

$$I_2 = \left(\frac{R_1}{R_1 + R_2}\right)I_T \qquad (5–7)$$

These formulas show that the current in either branch is equal to the opposite branch resistance divided by the sum of the two resistances and then multiplied by the total current.

EXAMPLE 5–17

Find I_1 and I_2 in Figure 5–36.

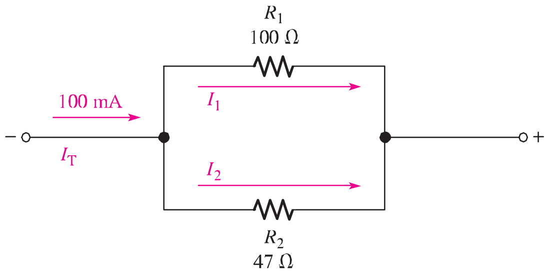

FIGURE 5–36

SOLUTION

Use Equation 5–6 to determine I_1.

$$I_1 = \left(\frac{R_2}{R_1 + R_2}\right)I_T = \left(\frac{47\ \Omega}{147\ \Omega}\right)100\text{ mA} = \mathbf{32.0\ mA}$$

Use Equation 5–7 to determine I_2.

$$I_2 = \left(\frac{R_1}{R_1 + R_2}\right)I_T = \left(\frac{100\ \Omega}{147\ \Omega}\right)100\text{ mA} = \mathbf{68.0\ mA}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

RELATED PROBLEM

6238262 2013/12/10 12.27.66.5

If $R_1 = 56\ \Omega$ and $R_2 = 82\ \Omega$ in Figure 5–36 and I_T stays the same, what will each branch current be?

SYSTEM EXAMPLE 5–2

AMMETER SHUNTS

Many systems incorporate basic analog ammeters; they are built into a control panel to give a visual indication of current. Most ammeters are set up to read more than one scale. Often, the meter is sensitive and will be destroyed with large currents. In order to avoid this and to read higher currents, parallel resistors called *shunts* are switched in, allowing most of the current to go around the meter.

The mechanism in an analog ammeter that causes the pointer to move in proportion to the current is called the *meter movement*. A meter movement has a certain resistance and a maximum current. This maximum current, called the *full-scale deflection current*, causes the pointer to go all the way to the right side of the scale. The meter in this example has 50 Ω internal resistance and a full-scale deflection current of 1 mA. A basic meter with this particular movement can measure currents of 1 mA or less, but if a current greater than 1 mA is in the meter movement, the pointer will "peg" (or stop) at full scale, which can damage the sensitive movement.

Parallel resistors are used to extend the range of an ammeter. The resistor values can be determined by applying the current-divider rule or Ohm's law. Figure 5–37 illustrates a 50 Ω, 1 mA meter with three ranges: 1 mA, 10 mA, and 100 mA. In the 1 mA setting, there is no shunt. In the 10 mA setting, the shunt resistance is 1/9 of the meter resistance or 5.56 Ω.

The maximum current in the meter and maximum shunt current are shown in Table 5–1 for each setting of the range switch. Notice that the maximum meter current never exceeds its rated 1.0 mA. The meter in Figure 5–37 is shown measuring the maximum current. In this case, 99 mA goes through the shunt and 1.0 mA goes through the meter movement. If the meter reads less than full scale, the currents in the meter and the shunt are proportionately less.

TABLE 5–1 • Meter current.

RANGE SWITCH SETTING	MAXIMUM METER CURRENT	MAXIMUM SHUNT CURRENT
1 mA	1.0 mA	0
10 mA	1.0 mA	9.0 mA
100 mA	1.0 mA	99.0 mA

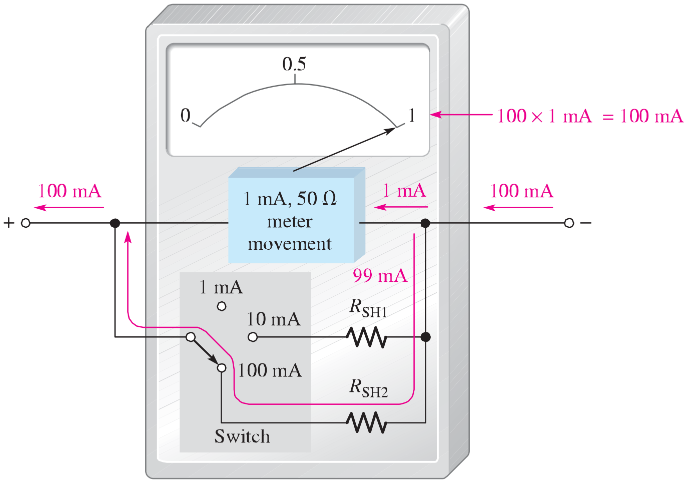

FIGURE 5–37 **A milliammeter with three ranges.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 5–6 CHECKUP

1. A circuit has the following resistors in parallel with a voltage source: 220 Ω, 100 Ω, 68 Ω, 56 Ω, and 22 Ω. Which resistor has the most current through it? The least current?
2. Determine the current through R_3 in Figure 5–38.

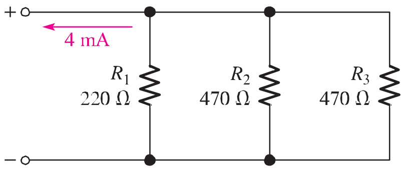

FIGURE 5–38

3. Find the currents through each resistor in the circuit of Figure 5–39.

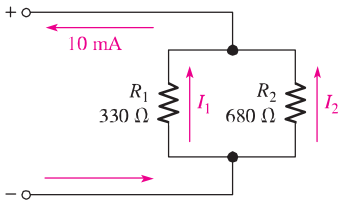

FIGURE 5–39

5–7 POWER IN PARALLEL CIRCUITS

Total power in a parallel circuit is found by adding up the powers of all the individual resistors, the same as for series circuits.

After completing this section, you should be able to

- **Determine power in a parallel circuit**

Equation 5–8 states the formula for total power for any number of resistors in parallel.

$$P_T = P_1 + P_2 + P_3 + \cdots + P_n \tag{5–8}$$

where P_T is the total power and P_n is the power in the last resistor in parallel. As you can see, the powers are additive, just as in a series circuit.

The power formulas from Chapter 3 are directly applicable to parallel circuits. The following formulas are used to calculate the total power P_T:

$$P_T = V_S I_T$$
$$P_T = I_T^2 R_T$$
$$P_T = \frac{V_S^2}{R_T}$$

where V_S is the voltage across the parallel circuit, I_T is the total current into the parallel circuit, and R_T is the total resistance of the parallel circuit. Examples 5–18 and 5–19 show how total power can be calculated in a parallel circuit.

EXAMPLE 5–18

Determine the total amount of power in the parallel circuit in Figure 5–40.

SOLUTION

The total current is 200 mA. The total resistance is

$$R_T = \frac{1}{\dfrac{1}{68\ \Omega} + \dfrac{1}{33\ \Omega} + \dfrac{1}{22\ \Omega}} = 11.1\ \Omega$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

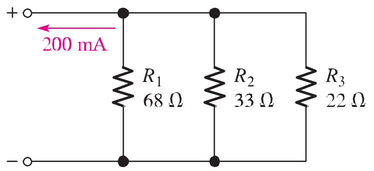

FIGURE 5–40

The easiest formula to use is $P_T = I_T^2 R_T$ since you know both I_T and R_T. Thus,

$$P_T = I_T^2 R_T = (200 \text{ mA})^2 (11.1 \ \Omega) = \mathbf{444 \ mW}$$

Let's demonstrate that if you determine the power in each resistor and if you add all of these values together, you get the same result. First, find the voltage across each branch of the circuit.

$$V_S = I_T R_T = (200 \text{ mA})(11.1 \ \Omega) = 2.22 \text{ V}$$

Remember that the voltage across all branches is the same.

Next, use $P = V_S^2/R$ to calculate the power for each resistor.

$$P_1 = \frac{(2.22 \text{ V})^2}{68 \ \Omega} = 72.5 \text{ mW}$$

$$P_2 = \frac{(2.22 \text{ V})^2}{33 \ \Omega} = 149 \text{ mW}$$

$$P_3 = \frac{(2.22 \text{ V})^2}{22 \ \Omega} = 224 \text{ mW}$$

Now, add these powers to get the total power.

$$P_T = 72.5 \text{ mW} + 149 \text{ mW} + 224 \text{ mW} = \mathbf{446 \ mW}$$

This calculation shows that the sum of the individual powers is equal (approximately) to the total power as determined by one of the power formulas. Rounding to three significant figures accounts for the difference.

RELATED PROBLEM

Find the total power in Figure 5–40 if the total voltage is doubled.

EXAMPLE 5–19

The amplifier in one channel of a stereo system as shown in Figure 5–41 drives two parallel speakers. If the maximum voltage* to the speakers is 15 V, how much power must the amplifier be able to deliver to the speakers?

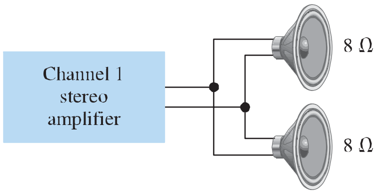

FIGURE 5–41

---

*Voltage is ac in this case; but power is determined the same for ac voltage as for dc voltage, as you will learn later.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

The speakers are connected in parallel to the amplifier output, so the voltage across each is the same. The maximum power to each speaker is

$$P_{\max} = \frac{V_{\max}^2}{R} = \frac{(15\text{ V})^2}{8\ \Omega} = 28.1\text{ W}$$

The total power that the amplifier must be capable of delivering to the speaker system is twice the power in an individual speaker because the total power is the sum of the individual powers.

6238262 2013/12/10 12.27.66.5

$$P_{\text{T(max)}} = P_{\max} + P_{\max} = 2P_{\max} = 2(28.1\text{ W}) = \mathbf{56.2\text{ W}}$$

RELATED PROBLEM

If the amplifier can produce a maximum of 18 V, what is the maximum total power to the speakers?

SECTION 5–7 CHECKUP

1. If you know the power in each resistor in a parallel circuit, how can you find the total power?
2. The resistors in a parallel circuit dissipate the following powers: 1 W, 2 W, 5 W, and 8 W. What is the total power in the circuit?
3. A circuit has a 1.0 kΩ, a 2.7 kΩ, and a 3.9 kΩ resistor in parallel. There is a total current of 1 mA into the parallel circuit. What is the total power?
4. Normally, circuits are protected with a fuse that is at least 120% of the maximum current expected. What size of fuse should be installed for an automotive rear window defroster that is rated for 100 W? Assume the battery voltage is 12.6 V.

5–8 TROUBLESHOOTING

Recall that an open circuit is one in which the current path is interrupted and there is no current. In this section, you will see how an open in a parallel branch affects the parallel circuit.

After completing this section, you should be able to

- **Troubleshoot parallel circuits**
  - **Check for an open in a circuit**

Open Branches

If a switch is connected in a branch of a parallel circuit, as shown in Figure 5–42, an open or a closed path can be made by the switch. When the switch is closed, as in Figure 5–42(a), R_1 and R_2 are in parallel. The total resistance is 50 Ω (two 100 Ω resistors in parallel). Current is through both resistors. If the switch is opened, as in Figure 5–42(b), R_1 is effectively removed

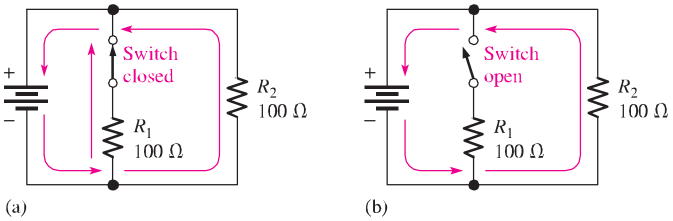

FIGURE 5–42 **When the switch opens, total current decreases and current through R_2 remains unchanged.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

from the circuit, and the total resistance is 100 Ω. There is still the same voltage across R_2 and the same current through it although the total current from the source is reduced by the amount of the R_1 current.

In general,

> **When an open occurs in a parallel branch, the total resistance increases, the total current decreases, and the same current continues through each of the remaining parallel paths.**

Consider the lamp circuit in Figure 5–43. There are four bulbs in parallel with a 12 V source. In part (a), there is current through each bulb. Now suppose that one of the bulbs burns out, creating an open path as shown in Figure 5–43(b). This light will go out because there is no current through the open path. Notice, however, that current continues through all the other parallel bulbs, and they continue to glow. The open branch does not change the voltage across the remaining parallel branches; it remains at 12 V, and the current through each branch remains the same.

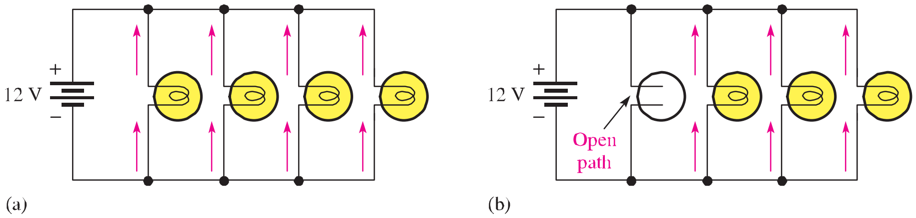

FIGURE 5–43 **When a lamp filament opens, total current decreases by the amount of current in the lamp that opened. The other branch currents remain unchanged.**

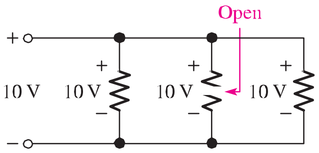

FIGURE 5–44 **All parallel branches (open or not) have the same voltage.**

You can see that a parallel circuit has an advantage over a series connection in lighting systems because if one or more of the parallel bulbs burn out, the others will stay on. In a series circuit, when one bulb goes out, all of the others go out also because the current path is completely interrupted.

When a resistor in a parallel circuit opens, the open resistor cannot be located by measurement of the voltage across the branches because the same voltage exists across all the branches. Thus, there is no way to tell which resistor is open by simply measuring voltage (the half-splitting method is not applicable). The good resistors will always have the same voltage as the open one, as illustrated in Figure 5–44 (note that the middle resistor is open). However, a temperature measurement with a thermal-imaging camera (see System Note on page 202) can reveal which resistors have current (good) and which do not (bad).

If a visual inspection does not reveal the open resistor, it can be located by resistance or current measurements. In practice, measuring resistance or current is more difficult than measuring voltage because you must disconnect a component to measure the resistance and you must insert an ammeter in series to measure the current. Thus, a wire or a printed circuit connection must usually be cut or disconnected, or one end of a component must be lifted off the circuit board, in order to connect a DMM to measure resistance and current. This procedure, of course, is not required when voltage measurements are made because the meter leads are simply connected across a component.

HANDS ON TIP

An alternate way to measure current without having to break a circuit to connect an ammeter is to place, as part of the original circuit, a 1 Ω series resistor in each line through which current is to be measured. This small "sense" resistor will generally not affect the total resistance. By measuring the voltage across the sense resistor, you automatically have the current reading since

$$I = \frac{V}{R} = \frac{V}{1\ \Omega} = V$$

Finding an Open Branch by Current Measurement

In a parallel circuit with a suspected open branch, the total current can be measured to find the open. *When a parallel resistor opens, I_T is always less than its normal value.* Once you know I_T and the voltage across the branches, a few calculations will determine the open resistor when all the resistors are of different values.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Consider the two-branch circuit in Figure 5–45(a). If one of the resistors opens, the total current will equal the current in the good resistor. Ohm's law quickly tells you what the current in each resistor should be.

$$I_1 = \frac{50\text{ V}}{560\ \Omega} = 89.3\text{ mA}$$

$$I_2 = \frac{50\text{ V}}{100\ \Omega} = 500\text{ mA}$$

$$I_T = I_1 + I_2 = 589.3\text{ mA}$$

If R_2 is open, the total current is 89.3 mA, as indicated in Figure 5–45(b). If R_1 is open, the total current is 500 mA, as indicated in Figure 5–45(c).

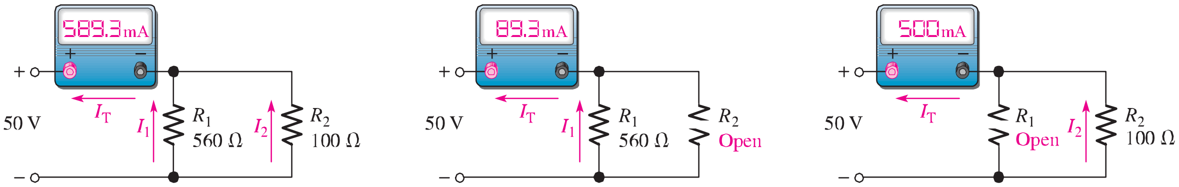

(a) Current with no open branch (b) Current with R_2 open (c) Current with R_1 open

FIGURE 5–45 **Finding an open path by current measurement.**

This procedure can be extended to any number of branches having unequal resistances. If the parallel resistances are all equal, the current in each branch must be checked until a branch is found with no current.

EXAMPLE 5–20

In Figure 5–46, there is a total current of 31.09 mA, and the voltage across the parallel branches is 20 V. Is there an open resistor, and if so, which one is it?

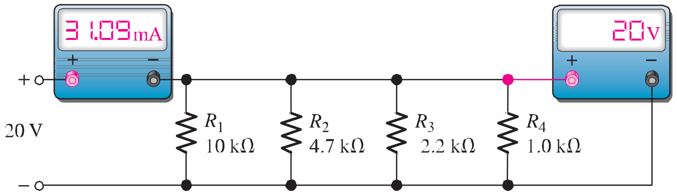

FIGURE 5–46

SOLUTION

Calculate the current in each branch.

$$I_1 = \frac{V}{R_1} = \frac{20\text{ V}}{10\text{ k}\Omega} = 2\text{ mA}$$

$$I_2 = \frac{V}{R_2} = \frac{20\text{ V}}{4.7\text{ k}\Omega} = 4.26\text{ mA}$$

$$I_3 = \frac{V}{R_3} = \frac{20\text{ V}}{2.2\text{ k}\Omega} = 9.09\text{ mA}$$

$$I_4 = \frac{V}{R_4} = \frac{20\text{ V}}{1.0\text{ k}\Omega} = 20\text{ mA}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The total current should be

$$I_T = I_1 + I_2 + I_3 + I_4 = 2\text{ mA} + 4.26\text{ mA} + 9.09\text{ mA} + 20\text{ mA}$$
$$= 35.35\text{ mA}$$

The actual measured current is 31.09 mA, as stated, which is 4.26 mA less than normal, indicating that the branch with 4.26 mA is open. Thus, R_2 **must be open**.

RELATED PROBLEM

What is the total current measured in Figure 5–46 if R_4 and not R_2 is open?

MULTISIM

Open Multisim file E05-20. Measure the total current and the current in each resistor. There are no faults in the circuit.

Shorted Branches

When a branch in a parallel circuit shorts, the current increases to an excessive value, usually causing the resistor to burn open or a fuse or circuit breaker to blow. This results in a difficult troubleshooting problem because it is hard to isolate the shorted branch.

A pulser and a current tracer are tools often used to find shorts in a circuit. They are not restricted to use in digital circuits but can be effective in any type of circuit. The pulser is a pen-shaped tool that applies pulses to a selected point in a circuit, causing pulses of current through the shorted path. The current tracer is also a pen-shaped tool that senses pulses of current. By following the current with the tracer, the current path can be identified.

Thermal Imaging

6238262 2013/12/10 12.27.66.5

Components are always designed to handle a certain amount of power as specified by the manufacturer. Sometimes a fault occurs (such as a loose or oxidized connection) and a component or junction overheats. A tool for detecting overheating in any device is thermal imaging, which is sensitive to the infrared radiation emitted by all objects. Thermal imaging is useful in a number of fields (e.g., mechanical, electrical, medical).

Many manufacturers routinely do thermal imaging of equipment as part of a preventive maintenance program. Thermal imaging has advantages in that it can often spot potential problems before they cause a complete malfunction. A thermal inspection is noncontacting and very sensitive to temperature differences.

Courtesy of Fluke Corporation

SYSTEM NOTE

Thermography

Figure 5–47 shows the main part of the electromagnetic spectrum. Electromagnetic energy is a continuous band of frequencies that includes visible light. We respond to the visible region with our eyes, but this is only a small fraction of the electromagnetic spectrum. The infrared region has longer wavelengths than the visible region. The infrared region is subdivided into the *near-infrared, mid-infrared,* and the *far-infrared*. It is this far-infrared region that the temperature-sensitive nerve endings in our skin sense as thermal energy when we are in sunlight or stand next to a hot stove. At wavelengths that are longer than the far-infrared are microwave frequencies and radio waves. At wavelengths shorter than the visible light region are the ultraviolet and x-ray regions.

The surfaces of all bodies radiate electromagnetic energy. If all bodies are constantly emitting energy, why don't they eventually radiate all of their energy and cool to absolute zero? The answer is because they continuously absorb energy from their surroundings. If

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

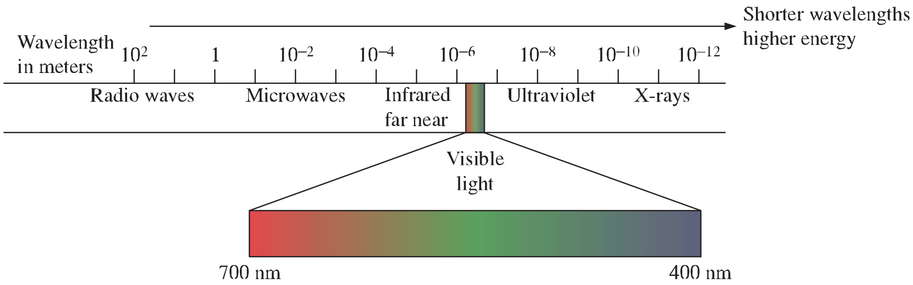

FIGURE 5–47 **The electromagnetic spectrum.**

the amount of energy they absorb is greater than what is radiated, the temperature rises; if it is less, the temperature falls. The wavelength that is emitted depends on the temperature of the body. If a body is hot enough, it radiates energy in the visible spectrum. If it is cooler, the radiated energy is primarily in the infrared region.

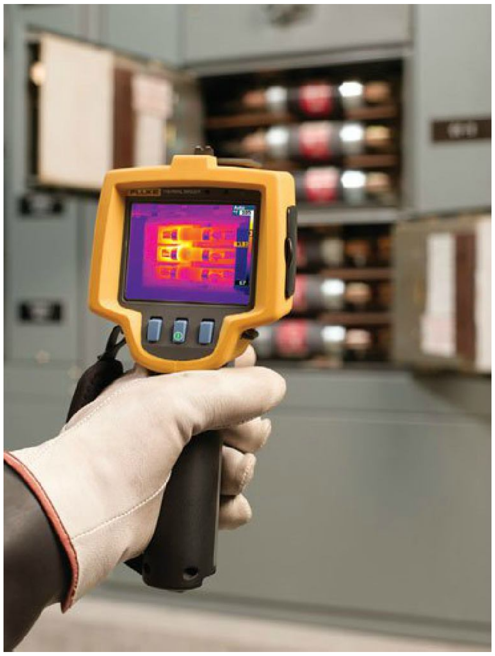

FIGURE 5–48 **Thermal imager showing the temperature profile of the target (a set of three fuses).** (Courtesy of Fluke Corporation)

Although our skin is very sensitive to heat, humans have physical limitations on detecting specific direction and quantizing the amount of heat. Thermal imaging instruments extend our senses and enable us to measure heat and provide a radiometric image of a hot source. A **radiometric image** is a thermal image that contains temperature-measurement calculations for various points within the image. The image is displayed on a screen where colors correspond to the temperature of the infrared radiation emitted by the target. Figure 5–48 shows a thermal imager that is viewing the infrared energy emitted by three fuses. The color corresponds to the temperature within the image. From this, it is obvious which fuses are carrying the largest current because they are warmer.

Thermal imaging is used in a variety of applications and is widely used in electronics and industrial operations for both troubleshooting and preventive maintenance. It can immediately show a component that is too hot or one that is not carrying current because it is cooler than normal. In the case of the dummy load in System Example 5–1, the set of parallel resistors needs to all be functioning correctly to present the required resistance that simulates an antenna. If one resistor were to open, the total resistance would increase and there would be a reflected signal back toward the transmitter. A thermal imager will show the open resistor as a cooler component than the others with no need to disconnect it in the circuit to test it. The test is fast and noncontacting.

SECTION 5–8 CHECKUP

1. If a parallel branch opens, what changes can be detected in the circuit's voltage and the currents, assuming that the parallel circuit is across a constant voltage source?
2. What happens to the total resistance if one branch opens?
3. If several light bulbs are connected in parallel and one of the bulbs opens (burns out), will the others continue to glow?
4. There is one ampere of current in each branch of a parallel circuit. If one branch opens, what is the current in each of the remaining branches?
5. Assume a load draws a hot current of 1.00 A, and the neutral return current is 0.90 A. If the source voltage is 120 V, what is the apparent faulty resistance in the ground path?
6. What do the colors represent in a radiometric image?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SUMMARY

- Resistors in parallel are connected between two nodes in a circuit.
- A parallel circuit provides more than one path for current.
- The total parallel resistance is less than the lowest-value parallel resistor.
- The voltages across all branches of a parallel circuit are the same.
- Kirchhoff's current law: The sum of the currents into a node (total current in) equals the sum of the currents out of the node (total current out).

  The algebraic sum of all the currents entering and leaving a node is zero.
- A parallel circuit is a current divider, so called because the total current entering a node divides up into each of the branches connected to the node.
- If all of the branches of a parallel circuit have equal resistance, the currents through all of the branches are equal.
- The total power in a parallel-resistive circuit is the sum of all of the individual powers of the resistors making up the parallel circuit.
- The total power for a parallel circuit can be calculated with the power formulas using values of total current, total resistance, or total voltage.
- If one of the branches of a parallel circuit opens, the total resistance increases, and therefore the total current decreases.
- If a branch of a parallel circuit opens, there is still the same current through the remaining branches.
- A radiometric image of a circuit is useful for preventive maintenance and for troubleshooting.

KEY TERMS

Key terms and other bold terms in the chapter are defined in the end-of-book glossary.

Branch One current path in a parallel circuit.

Current divider A parallel circuit in which the currents divide inversely proportional to the parallel branch resistances.

Kirchhoff's current law A law stating that the total current into a node equals the total current out of the node. Equivalently, it can be stated that the algebraic sum of all the currents entering and leaving a node is zero.

Node A point or junction in a circuit at which two or more components are connected.

Parallel The relationship between two circuit components that exists when they are connected between the same pair of nodes.

Radiometric image A thermal picture in which colors represent temperature.

KEY FORMULAS

(5–1)	$R_T = \dfrac{1}{\dfrac{1}{R_1} + \dfrac{1}{R_2} + \dfrac{1}{R_3} + \cdots + \dfrac{1}{R_n}}$	Total parallel resistance
(5–2)	$R_T = \dfrac{R_1 R_2}{R_1 + R_2}$	Special case for two resistors in parallel
(5–3)	$R_T = \dfrac{R}{n}$	Special case for n equal-value resistors in parallel
(5–4)	$I_{IN(1)} + I_{IN(2)} + I_{IN(3)} + \cdots + I_{IN(n)} = I_{OUT(1)} + I_{OUT(2)} + I_{OUT(3)} + \cdots + I_{OUT(m)}$	Kirchhoff's current law
(5–5)	$I_x = \left(\dfrac{R_T}{R_x}\right) I_T$	General current-divider formula
(5–6)	$I_1 = \left(\dfrac{R_2}{R_1 + R_2}\right) I_T$	Two-branch current-divider formula

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

(5–7) $$I_2 = \left(\frac{R_1}{R_1 + R_2}\right)I_T$$ Two-branch current-divider formula

(5–8) $$P_T = P_1 + P_2 + P_3 + \cdots + P_n$$ Total power

TRUE/FALSE QUIZ

Answers are at the end of the chapter.

1. To find the total conductance of parallel resistors, you can add the conductance of each of the resistors.
2. The total resistance of parallel resistors is always smaller than the smallest resistor.
3. The product-over-sum rule works for any number of parallel resistors.
4. In a parallel circuit, the voltage is larger on a larger resistor and smaller on a smaller resistor.
5. When a new path is added to a parallel circuit, the total resistance goes up.
6. When a new path is added to a parallel circuit, the total current goes up.
7. The total current entering a node is always equal to the total current leaving the node.
8. In the current-divider formula, $I_x = (R_T/R_x)I_T$, the quantity in parentheses is a fraction that is never greater than 1.
9. When two resistors are in parallel, the smaller resistor will have the smaller power dissipation.
10. The total power dissipated by parallel resistors can be larger than the power supplied by the source.

SELF-TEST

Answers are at the end of the chapter.

1. In a parallel circuit, each resistor has
   (a) the same current **(b)** the same voltage
   (c) the same power **(d)** all of the above
2. When a 1.2 kΩ resistor and a 100 Ω resistor are connected in parallel, the total resistance is
   (a) greater than 1.2 kΩ **(b)** greater than 100 Ω but less than 1.2 kΩ
   (c) less than 100 Ω but greater than 90 Ω **(d)** less than 90 Ω
3. A 330 Ω resistor, a 270 Ω resistor, and a 68 Ω resistor are all in parallel. The total resistance is approximately
   (a) 668 Ω **(b)** 47 Ω **(c)** 68 Ω **(d)** 22 Ω
4. Eight resistors are in parallel. The two lowest-value resistors are both 1.0 kΩ. The total resistance
   (a) cannot be determined **(b)** is greater than 1.0 kΩ
   (c) is less than 1.0 kΩ **(d)** is less than 500 Ω
5. When an additional resistor is connected across an existing parallel circuit, the total resistance
   (a) decreases **(b)** increases
   (c) remains the same **(d)** increases by the value of the added resistor
6. If one of the resistors in a parallel circuit is removed, the total resistance
   (a) decreases by the value of the removed resistor **(b)** remains the same
   (c) increases **(d)** doubles
7. The currents into a node are along two paths. One current is 5 A and the other is 3 A. The total current out of the node is
   (a) 2 A **(b)** unknown
   (c) 8 A **(d)** the larger of the two

6238262 2013/12/10 12.27.66.5

8. The following resistors are in parallel across a voltage source: 390 Ω, 560 Ω, and 820 Ω. The resistor with the least current is
   (a) 390 Ω **(b)** 560 Ω
   (c) 820 Ω **(d)** impossible to determine without knowing the voltage

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

9. A sudden decrease in the total current into a parallel circuit may indicate

(a) a short **(b)** an open resistor
(c) a drop in source voltage **(d)** either (b) or (c)

10. In a four-branch parallel circuit, there are 10 mA of current in each branch. If one of the branches opens, the current in each of the other three branches is

(a) 13.33 mA **(b)** 10 mA **(c)** 0 A **(d)** 30 mA

11. In a certain three-branch parallel circuit, R_1 has 10 mA through it, R_2 has 15 mA through it, and R_3 has 20 mA through it. After measuring a total current of 35 mA, you can say that

(a) R_1 is open **(b)** R_2 is open
(c) R_3 is open **(d)** the circuit is operating properly

12. If there are a total of 100 mA into a parallel circuit consisting of three branches and two of the branch currents are 40 mA and 20 mA, the third branch current is

(a) 60 mA **(b)** 20 mA **(c)** 160 mA **(d)** 40 mA

13. A complete short develops across one of five parallel resistors on a PC board. The most likely result is

(a) the smallest-value resistor will burn out
(b) one or more of the other resistors will burn out
(c) the fuse in the power supply will blow
(d) the resistance values will be altered

14. The power dissipation in each of four parallel branches is 1 mW. The total power dissipation is

(a) 1 mW **(b)** 4 mW **(c)** 0.25 mW **(d)** 16 mW

TROUBLESHOOTING: SYMPTOM AND CAUSE

The purpose of these exercises is to help develop thought processes essential to troubleshooting. Answers are at the end of the chapter.

Determine the cause for each set of symptoms. Refer to Figure 5–49.

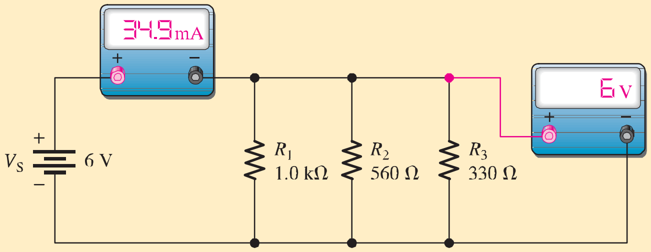

FIGURE 5–49 **The meters indicate the correct readings for this circuit.**

1. *Symptom:* The ammeter and voltmeter readings are zero.
Cause:
(a) R_1 is open. **(b)** The voltage source is turned off or faulty. **(c)** R_3 is open.

2. *Symptom:* The ammeter reading is 16.7 mA, and the voltmeter reading is 6 V.
Cause:
(a) R_1 is open. **(b)** R_2 is open. **(c)** R_3 is open.

3. *Symptom:* The ammeter reading is 28.9 mA, and the voltmeter reading is 6 V.
Cause:
(a) R_1 is open. **(b)** R_2 is open. **(c)** R_3 is open.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

4. *Symptom:* The ammeter reading is 24.2 mA, and the voltmeter reading is 6 V.
   Cause:
   (a) R_1 is open.
   (b) R_2 is open.
   (c) R_3 is open.

5. *Symptom:* The ammeter reading is 34.9 mA, and the voltmeter reading is 0 V.
   Cause:
   (a) A resistor is shorted.
   (b) The voltmeter is faulty.
   (c) The voltage source is turned off or faulty.

PROBLEMS

Answers to odd-numbered problems are at the end of the book.

BASIC PROBLEMS

SECTION 5–1 Resistors in Parallel

1. Connect the resistors in Figure 5–50 in parallel across the battery.
2. Determine whether or not all the resistors in Figure 5–51 are connected in parallel on the printed circuit board. Draw the schematic including resistor values.

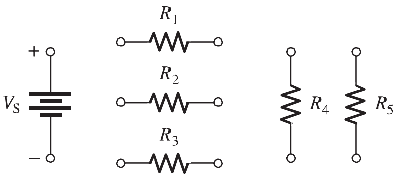

FIGURE 5–50

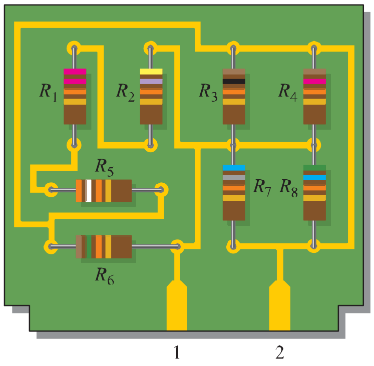

FIGURE 5–51

SECTION 5–2 Total Parallel Resistance

3. Determine the total resistance between pins 1 and 2 in Figure 5–51.
4. The following resistors are connected in parallel: 1.0 MΩ, 2.2 MΩ, 4.7 MΩ, 12 MΩ, and 22 MΩ. Determine the total resistance.
5. Find the total resistance between nodes A and B for each group of parallel resistors in Figure 5–52.

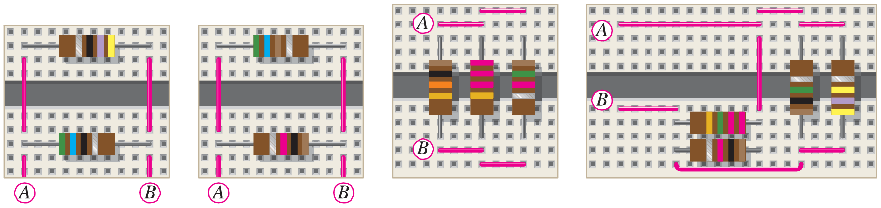

(a) (b) (c) (d)

FIGURE 5–52

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

6. Calculate R_T for each circuit in Figure 5–53.

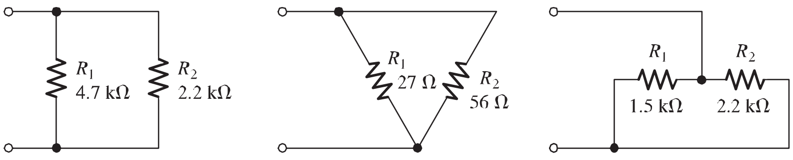

(a) (b) (c)

FIGURE 5–53

7. What is the total resistance of eleven 22 kΩ resistors in parallel?

8. Five 15 Ω, ten 100 Ω, and two 10 Ω resistors are all connected in parallel. What is the total resistance?

SECTION 5–3 Voltage in a Parallel Circuit

9. Determine the voltage across and the current through each parallel resistor if the total voltage is 12 V and the total resistance is 600 Ω. There are four resistors, all of equal value.

10. The source voltage in Figure 5–54 is 100 V. How much voltage does each of the three meters read?

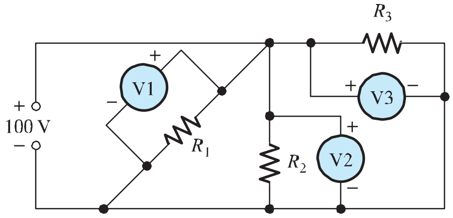

FIGURE 5–54

SECTION 5–4 Application of Ohm's Law

11. What is the total current, I_T, in each circuit of Figure 5–55?

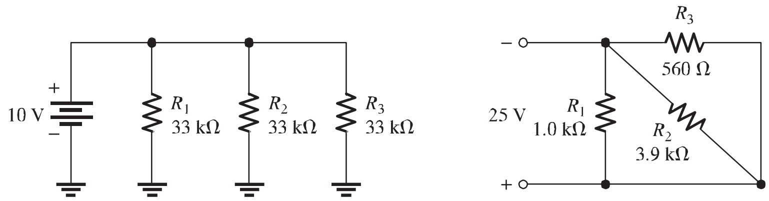

(a) (b)

FIGURE 5–55

12. The resistance of a 60 W bulb is approximately 240 Ω. What is the current from the source when three bulbs are on in a 120 V parallel circuit?

13. What is the current in each resistor for the circuits shown in Figure 5–56?

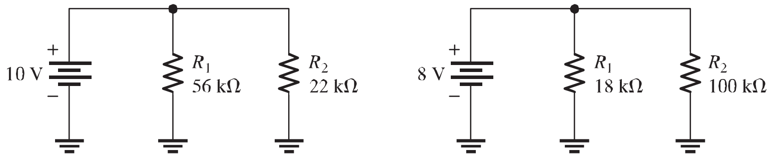

(a) (b)

FIGURE 5–56

14. Four equal-value resistors are connected in parallel. Five volts are applied across the parallel circuit, and 2.5 mA are measured from the source. What is the value of each resistor?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 5–5 Kirchhoff's Current Law

15. The following currents are measured in the same direction in a three-branch parallel circuit: 250 mA, 300 mA, and 800 mA. What is the value of the current into the node to which these three branches are connected?

16. There is a total of 500 mA of current into five parallel resistors. The currents through four of the resistors are 50 mA, 150 mA, 25 mA, and 100 mA. What is the current through the fifth resistor?

17. How much current is through R_2 and R_3 in Figure 5–57 if R_2 and R_3 have the same resistance? Show how to connect ammeters to measure these currents.

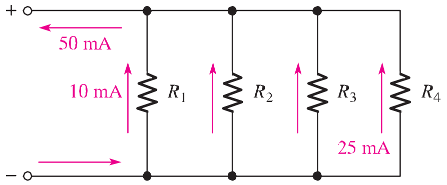

FIGURE 5–57

18. A trailer has four running lights that draw 0.5 A each and two taillights that draw 1.2 A each. What is the current supplied to the trailer when the taillights and running lights are on?

19. Assume the trailer in Problem 18 has two brake lights that draw 1 A each.
(**a**) When all lights are on, what current is supplied to the trailer?
(**b**) For this condition, what is the ground return current from the trailer?

SECTION 5–6 Current Dividers

20. A 10 kΩ resistor and a 15 kΩ resistor are in parallel with a voltage source. Which resistor has the highest current?

21. How much branch current should each meter in Figure 5–58 indicate?

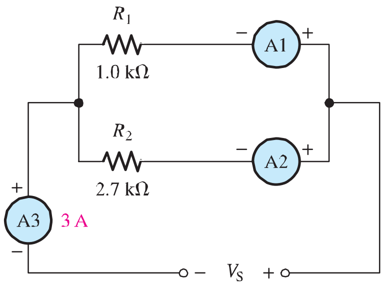

FIGURE 5–58

22. Using the current-divider formula, determine the current in each branch of the circuits of Figure 5–59.

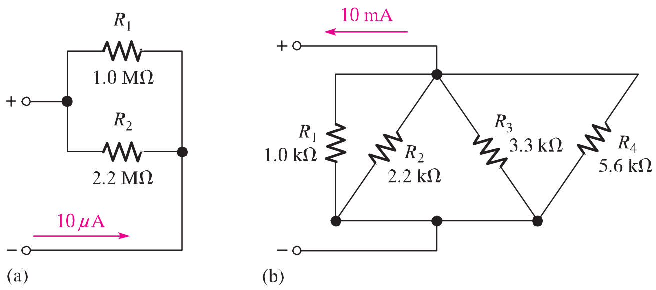

(a) (b)

FIGURE 5–59

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 5–7 Power in Parallel Circuits

23. Five parallel resistors each handle 40 mW. What is the total power?

24. Determine the total power in each circuit of Figure 5–59.

25. Six light bulbs are connected in parallel across 120 V. Each bulb is rated at 75 W. How much current is through each bulb, and what is the total current?

SECTION 5–8 Troubleshooting

26. If one of the bulbs burns out in Problem 25, how much current will be through each of the remaining bulbs? What will be the total current?

27. In Figure 5–60, the current and voltage measurements are indicated. Has a resistor opened, and, if so, which one?

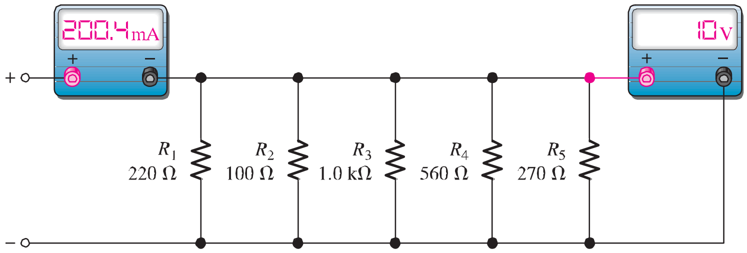

FIGURE 5–60

28. What is wrong with the circuit in Figure 5–61?

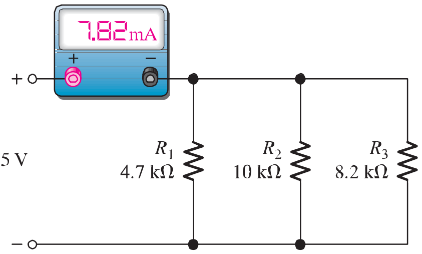

FIGURE 5–61

29. Find the open resistor in Figure 5–62.

30. From the ohmmeter reading in Figure 5–63, can you tell if a resistor is open? If so, can you identify it?

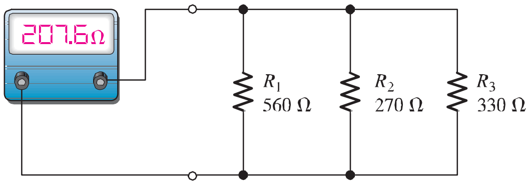

FIGURE 5–62

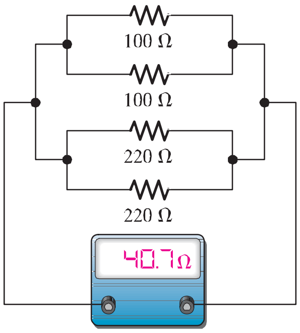

FIGURE 5–63

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

ADVANCED PROBLEMS

31. In the circuit of Figure 5–64, determine resistances R_2, R_3, and R_4.

32. The total resistance of a parallel circuit is 25 Ω. If the total current is 100 mA, how much current is through a 220 Ω resistor that makes up part of the parallel circuit?

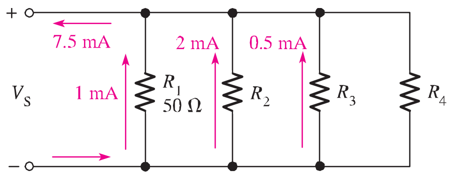
+
7.5 mA
2 mA
0.5 mA
V_S
1 mA
R_1
50 Ω
R_2
R_3
R_4
–

FIGURE 5–64

33. What is the current through each resistor in Figure 5–65? *R* is the lowest-value resistor, and all others are multiples of that value as indicated.

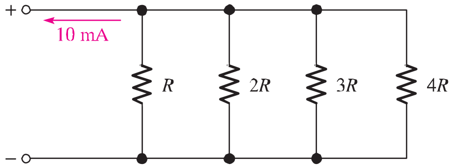
+
10 mA
R
2R
3R
4R
–

FIGURE 5–65

34. A certain parallel circuit consists of only ½ W resistors each having the same resistance. The total resistance is 1 kΩ, and the total current is 50 mA. If each resistor is operating at one-half its maximum power level, determine the following:

(a) the number of resistors **(b)** the value of each resistor
(c) the current in each branch **(d)** the applied voltage

35. Find the values of the unspecified quantities (shown in color) in each circuit of Figure 5–66.

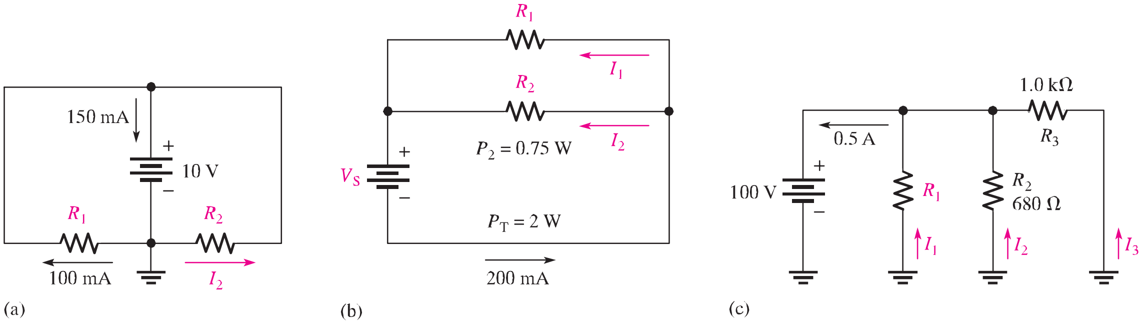
150 mA
+
10 V
–
R_1
R_2
100 mA
I_2
(a)
R_1
I_1
R_2
I_2
+
V_S
–
P_2 = 0.75 W
P_T = 2 W
200 mA
(b)
1.0 kΩ
R_3
0.5 A
+
100 V
–
R_1
R_2
680 Ω
I_1
I_2
I_3
(c)

FIGURE 5–66

36. What is the total resistance between terminal *A* and ground in Figure 5–67 for the following conditions?

(a) SW1 and SW2 open **(b)** SW1 closed, SW2 open
(c) SW1 open, SW2 closed **(d)** SW1 and SW2 closed

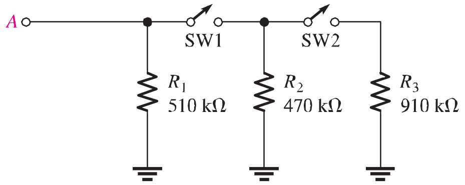
A
SW1
SW2
R_1
510 kΩ
R_2
470 kΩ
R_3
910 kΩ

6238262 2013/12/10 12.27.66.5

FIGURE 5–67

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

37. What value of R_2 in Figure 5–68 will cause excessive current?

38. Determine the total current from the source and the current through each resistor for each switch position in Figure 5–69.

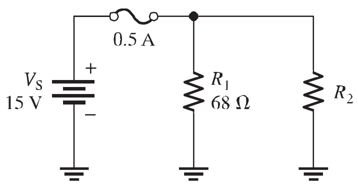

FIGURE 5–68

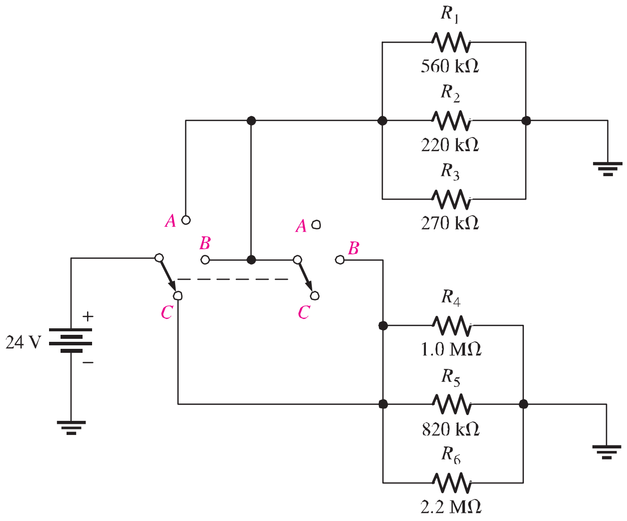

FIGURE 5–69

39. The electrical circuit in a room is protected with a 15 A circuit breaker. A space heater that draws 8.0 A is plugged into one wall outlet, and two table lamps that draw 0.833 A each are plugged into another outlet. Can a vacuum cleaner that uses 5.0 A be plugged into this same circuit without exceeding the breaker current? Explain your answer.

40. The total resistance of a parallel circuit is 25 Ω. What is the current through a 220 Ω resistor that makes up part of the parallel circuit if the total current is 100 mA?

41. Identify which groups of resistors are in parallel on the double-sided PC board in Figure 5–70. Determine the total resistance of each group.

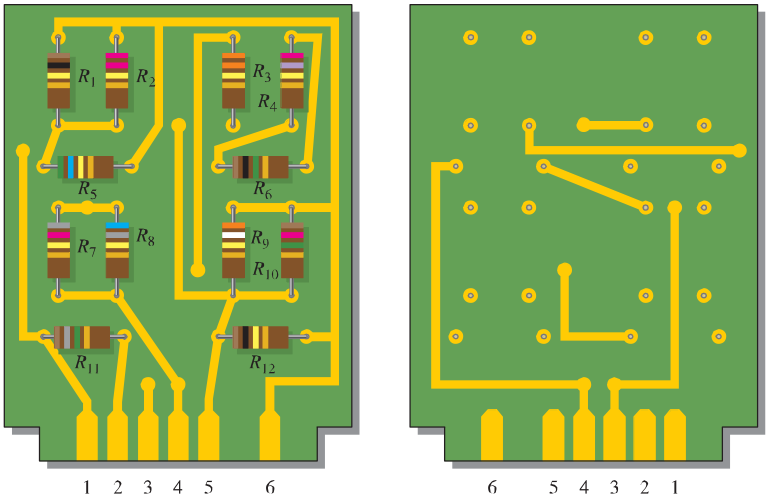

Side 1

Side 2

FIGURE 5–70

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

42. If the total resistance in Figure 5–71 is 200 Ω, what is the value of R_2?

43. Determine the unknown resistances in Figure 5–72.

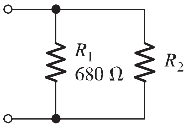

FIGURE 5–71

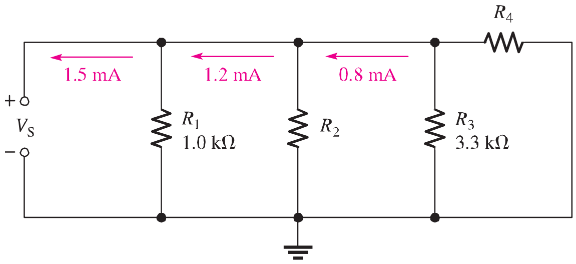
R_4
1.5 mA
1.2 mA
0.8 mA
+
V_S
−
R_1
1.0 kΩ
R_2
R_3
3.3 kΩ

FIGURE 5–72

44. There is a total of 250 mA into a parallel circuit with a total resistance of 1.5 kΩ. The current must be increased by 25%. Determine how much resistance to add in parallel to accomplish this increase in current.

45. Draw the schematic for the setup in Figure 5–73 and determine what is wrong with the circuit if 25 V are applied across the red and black leads.

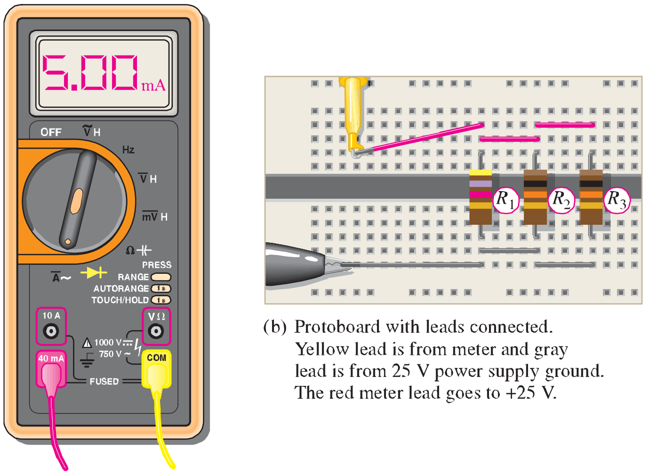
5.00 mA
OFF
V H
Hz
V H
mV H
Ω
PRESS
RANGE
AUTORANGE
TOUCH/HOLD
A~
10 A
VΩ
40 mA
1000 V
750 V ~
COM
FUSED
R_1
R_2
R_3

(b) Protoboard with leads connected. Yellow lead is from meter and gray lead is from 25 V power supply ground. The red meter lead goes to +25 V.

(a) Meter with yellow lead going to protoboard and red lead going to the positive terminal of the 25 V power supply.

FIGURE 5–73

46. Refer to System Example 5–1. If you needed to design the load for a 10 W, 75 Ω resistance using standard values of 1 W, 1% resistors, what values would you choose? Draw your completed design.

47. A dummy load has the values shown in Figure 5–74. All resistors are rated for 2 W except R_{14} and R_{15} are rated for ½ W each. (a) What is the resistance of the load? (b) What is its power rating?

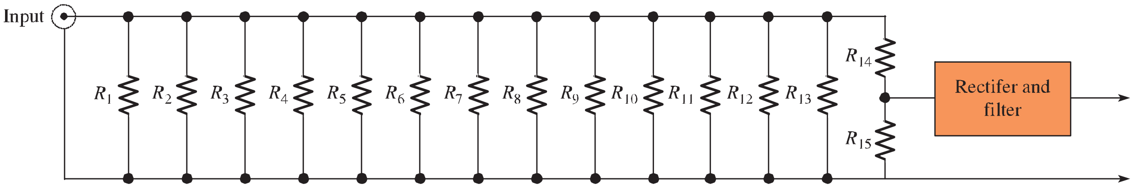
Input
R_1
R_2
R_3
R_4
R_5
R_6
R_7
R_8
R_9
R_{10}
R_{11}
R_{12}
R_{13}
R_{14}
R_{15}
Rectifier and filter

$R_1 - R_{13} = 1.00\ \text{k}\Omega$ each; $R_{14} = R_{15} = 1.5\ \text{k}\Omega$

FIGURE 5–74

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

48. Refer to System Example 5–2. For this example, what is the correct value of R_{SH2}?

49. Assume you have a meter that has a 100 μA full-scale deflection current and an internal resistance of 80 Ω. What shunt resistance would you specify to cause the meter to read 1 mA full scale?

MULTISIM

MULTISIM TROUBLESHOOTING PROBLEMS

50. Open file P05-50; files are found at www.pearsonhighered.com/floyd. Using current measurements, determine if there is a fault in the circuit. If so, identify the fault.

51. Open file P05-51. Using current measurements, determine if there is a fault in the circuit. If so, identify the fault.

52. Open file P05-52. Using resistance measurements, determine if there is a fault in the circuit. If so, identify the fault.

53. Open file P05-53. Measure the total resistance of each circuit and compare to the calculated values.

ANSWERS TO SECTION CHECKUPS

SECTION 5–1 Resistors in Parallel

1. Parallel resistors are connected between the same two points.
2. A parallel circuit has more than one current path between two given points.
3. See Figure 5–75.

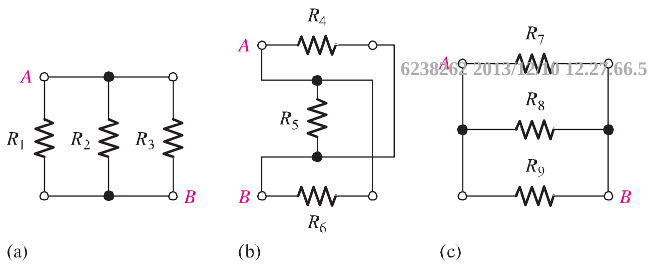

6238262 2013/12/10 12.27.66.5

(a) (b) (c)

FIGURE 5–75

4. See Figure 5–76.

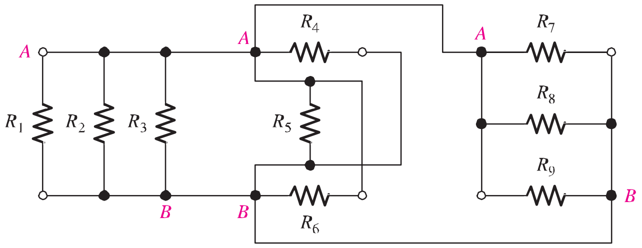

FIGURE 5–76

SECTION 5–2 Total Parallel Resistance

1. As more resistors are connected in parallel, the total resistance decreases.
2. R_T is always less than the smallest resistance value.
3. $R_T = 2.2\ k\Omega/12 = 183\ \Omega$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 5–3 Voltage in a Parallel Circuit

1. 5 V
2. $V_{R2} = 118\text{ V}$; $V_S = 118\text{ V}$
3. $V_{R1} = 50\text{ V}$; $V_{R2} = 50\text{ V}$
4. Voltage is the same across all parallel branches.

SECTION 5–4 Application Ohm's Law

1. $I_T = 12\text{ V}/(680\ \Omega/3) = 53\text{ mA}$
2. $V_S = 20\text{ mA}(680\ \Omega \| 330\ \Omega) = 4.44\text{ V}$
3. $I_1 = 4.44\text{ V}/680\ \Omega = 6.53\text{ mA}$; $I_2 = 4.44\text{ V}/330\ \Omega = 13.5\text{ mA}$
4. $R_T = 4(12\text{ V}/6\text{ mA}) = 8\text{ k}\Omega$
5. $V = (1.0\text{ k}\Omega \| 2.2\text{ k}\Omega)100\text{ mA} = 68.8\text{ V}$

SECTION 5–5 Kirchhoff's Current Law

1. Kirchhoff's current law: The algebraic sum of all currents at a node is zero; The sum of the currents entering a node equals the sum of all the currents leaving that node.
2. $I_1 + I_2 + I_3 = 2.5\text{ A}$
3. $100\text{ mA} + 300\text{ mA} = 400\text{ mA}$
4. 4 A
5. 10 A

SECTION 5–6 Current Dividers

1. The 22 Ω has most current; the 220 Ω has least current.
2. $I_3 = (R_T/R_3)4\text{ mA} = (113.6\ \Omega/470\ \Omega)4\text{ mA} = 967\ \mu\text{A}$
3. $I_2 = (R_T/680\ \Omega)10\text{ mA} = 3.27\text{ mA}$; $I_1 = (R_T/330\ \Omega)10\text{ mA} = 6.73\text{ mA}$

SECTION 5–7 Power in Parallel Circuits

1. Add the powers in each resistor to get P_T.
2. $P_T = 1\text{ W} + 2\text{ W} + 5\text{ W} + 8\text{ W} = 16\text{ W}$
3. $P_T = (1\text{ mA})^2 R_T = 615\ \mu\text{W}$
4. The normal current is $I = P/V = 100\text{ W}/12.6\text{ V} = 7.9\text{ A}$. Choose a 10 A fuse.

SECTION 5–8 Troubleshooting

1. When a parallel branch opens, there is no change in voltage and the total current decreases.
2. When the total resistance is greater than it should be, a branch is open.
3. Yes, all other bulbs continue to glow.
4. The current in each nonopen branch remains 1 A.
5. By Kirchhoff's current law, the ground return current is 0.10 A and represents a parallel return path. From Ohm's law, $R = 120\text{ V}/0.1\text{ A} = 1200\ \Omega$.
6. Temperature

ANSWERS TO RELATED PROBLEMS FOR EXAMPLES

5–1 No rewiring is necessary.

5–2 Connect pin 1 to pin 2 and pin 3 to pin 4.

5–3 9.34 Ω

5–4 Replace Step 9 with "Press the + key."
Display shows $100^{-1} + 47^{-1} + 22^{-1} +$.
Replace Step 10 with "Enter 33."
Display shows $100^{-1} + 47^{-1} + 22^{-1} + 33$.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Step 11: Press the x^{-1} key.

Display shows $100^{-1} + 47^{-1} + 22^{-1} + 33^{-1}$.

Step 12: Press the ENTER key. Display shows $107.034171502\text{e}^{-3}$.

Step 13: Press the x^{-1} and ENTER keys. Display shows 9.34281067406E0.

5–5 132 Ω
5–6 33.3 kΩ
5–7 25 V
5–8 $I_1 = 100$ mA; $I_2 = 179$ mA; 100 mA + 179 mA = 279 mA
5–9 $I_1 = 20.0$ mA; $I_2 = 9.09$ mA; $I_3 = 35.7$ mA; $I_4 = 22.0$ mA
5–10 31.9 mA
5–11 2.0 Ω
5–12 $I = V/R_{\text{TAIL}} = 12.6\text{ V}/10.5\ \Omega = 1.2\text{ A}$
5–13 20 mA
5–14 $I_T = 112$ mA; $I_2 = 50$ mA
5–15 2.5 mA; 5 mA
5–16 $I_1 = 1.63$ mA; $I_2 = 3.35$ mA
5–17 $I_1 = 59.4$ mA; $I_2 = 40.6$ mA
5–18 1.78 W
5–19 81 W
5–20 15.4 mA

ANSWERS TO TRUE/FALSE QUIZ

1. T **2.** T **3.** F **4.** F **5.** F **6.** T **7.** T **8.** T **9.** F **10.** F

ANSWERS TO SELF-TEST

1. (b) **2.** (c) **3.** (b) **4.** (d) **5.** (a) **6.** (c) **7.** (c)
8. (c) **9.** (d) **10.** (b) **11.** (a) **12.** (d) **13.** (c) **14.** (b)

ANSWERS TO TROUBLESHOOTING: SYMPTOM AND CAUSE

1. (b)
2. (c)
3. (a)
4. (b)
5. (b)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

CHAPTER 6

SERIES-PARALLEL CIRCUITS

OUTLINE

6–1 Identifying Series-Parallel Relationships
6–2 Analysis of Series-Parallel Resistive Circuits
6–3 Voltage Dividers with Resistive Loads
6–4 Loading Effect of a Voltmeter
6–5 The Wheatstone Bridge
6–6 Thevenin's Theorem
6–7 The Maximum Power Transfer Theorem
6–8 The Superposition Theorem
6–9 Troubleshooting

OBJECTIVES

- Identify series-parallel relationships
- Analyze series-parallel circuits
- Analyze loaded voltage dividers
- Determine the loading effect of a voltmeter on a circuit
- Analyze and apply a Wheatstone bridge
- Apply Thevenin's theorem to simplify a circuit for analysis
- Apply the maximum power transfer theorem
- Apply the superposition theorem to circuit analysis
- Troubleshoot series-parallel circuits

KEY TERMS

Loading
Load current
Bleeder current
Wheatstone bridge
Balanced bridge
Unbalanced bridge
Thevenin's theorem
Terminal equivalency
Maximum power transfer
Superposition

INTRODUCTION

Various combinations of both series and parallel resistors are often found in electronic systems. In this chapter, examples of such series-parallel arrangements are examined and analyzed. An important circuit called the *Wheatstone bridge* is introduced, which is widely used in measurement systems. You will learn how complex circuits can be simplified using Thevenin's theorem. The maximum power transfer theorem, used in applications where it is important for a given circuit to provide maximum power to a load, is discussed. Also, circuits with more than one voltage source are analyzed in simple steps using the superposition theorem. Troubleshooting series-parallel circuits for shorts and opens is also covered.

6238262 2013/12/10 12.27.66.5

VISIT THE WEBSITE
Study aids for this chapter are available at
http://pearsonhighered.com/floyd

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6–1 IDENTIFYING SERIES-PARALLEL RELATIONSHIPS

A series-parallel circuit consists of combinations of both series and parallel current paths. It is important to be able to identify how the components in a circuit are arranged in terms of their series and parallel relationships.

After completing this section, you should be able to

- **Identify series-parallel relationships**
  - **Recognize how each resistor in a given circuit is related to the other resistors**
  - **Determine series and parallel relationships on a PC board**

Figure 6–1(a) shows an example of a simple series-parallel combination of resistors. Notice that the resistance from A to B is R_1. The resistance from B to C is R_2 and R_3 in parallel ($R_2 \| R_3$) because these resistors are connected to the same pair of nodes (node B and node C). The total resistance from A to C is R_1 in series with the parallel combination of R_2 and R_3, as indicated in Figure 6–1(b).

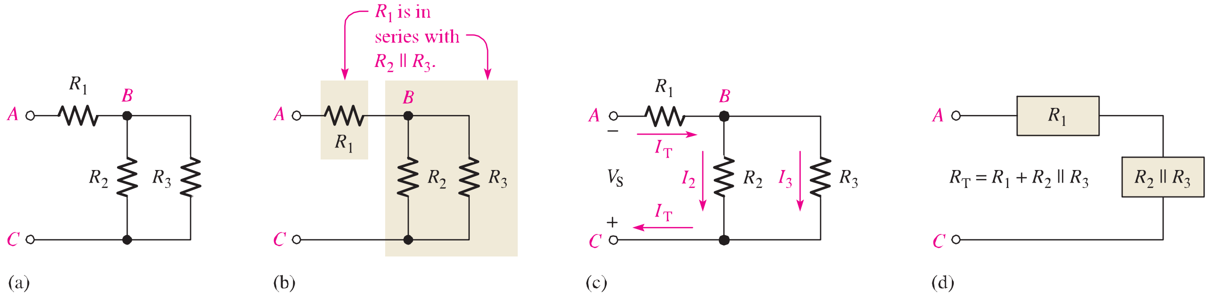

FIGURE 6–1 **A simple series-parallel resistive circuit.**

When the resistors in Figure 6–1 are connected to a voltage source as shown in part (c) the total current through R_1 divides at point B into the two parallel paths. These two branch currents then recombine. The total current is into the positive source terminal as shown. The resistor relationships are shown in block form in part (d).

Now, to further illustrate series-parallel relationships, let's increase the complexity of the arrangement in Figure 6–1(a) step-by-step.

1. In Figure 6–2(a), R_4 is connected in series with R_1. The resistance between points A and B is now $R_1 + R_4$. This series combination is in series with the parallel combination of R_2 and R_3, as illustrated in Figure 6–2(b). Part (c) shows the resistor relationships in block form.

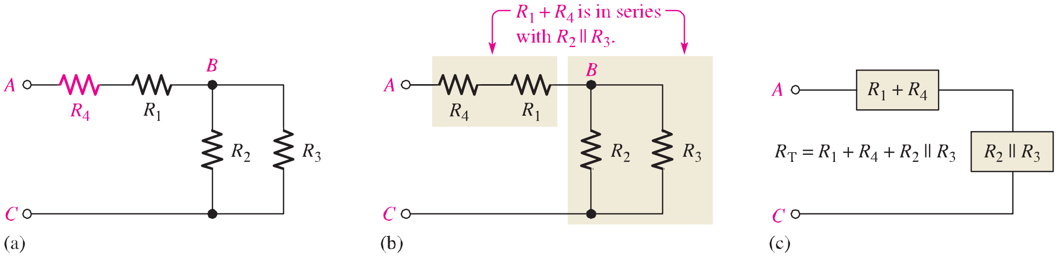
A
B
C
R_4
R_1
R_2
R_3
$R_1 + R_4$ is in series with $R_2 \| R_3$.
$R_1 + R_4$
$R_T = R_1 + R_4 + R_2 \| R_3$
$R_2 \| R_3$
(a)
(b)
(c)

FIGURE 6–2 **R_4 is added to the circuit in series with R_1.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

2. In Figure 6–3(a), R_5 is connected in series with R_2. The series combination of R_2 and R_5 is in parallel with R_3. This entire series-parallel combination is in series with the $R_1 + R_4$ combination, as illustrated in Figure 6–3(b). The block diagram is shown in part (c).

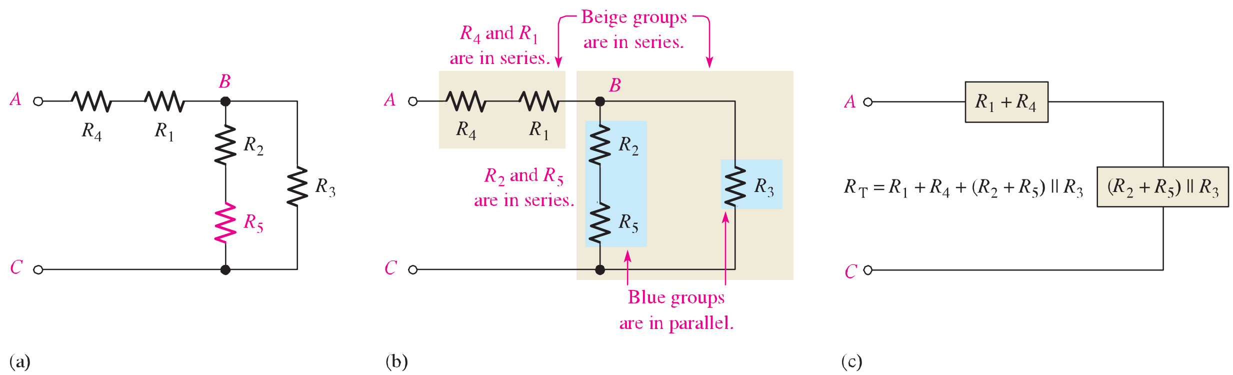
A
B
C
R_4
R_1
R_2
R_3
R_5
R_4 and R_1 are in series.
Beige groups are in series.
R_2 and R_5 are in series.
Blue groups are in parallel.
$R_1 + R_4$
$R_T = R_1 + R_4 + (R_2 + R_5) \parallel R_3$
$(R_2 + R_5) \parallel R_3$

(a) (b) (c)

FIGURE 6–3 **R_5 is added to the circuit in series with R_2.**

3. In Figure 6–4(a), R_6 is connected in parallel with the series combination of R_1 and R_4. The series-parallel combination of R_1, R_4, and R_6 is in series with the series-parallel combination of R_2, R_3, and R_5, as indicated in Figure 6–4(b). The block diagram is shown in part (c).

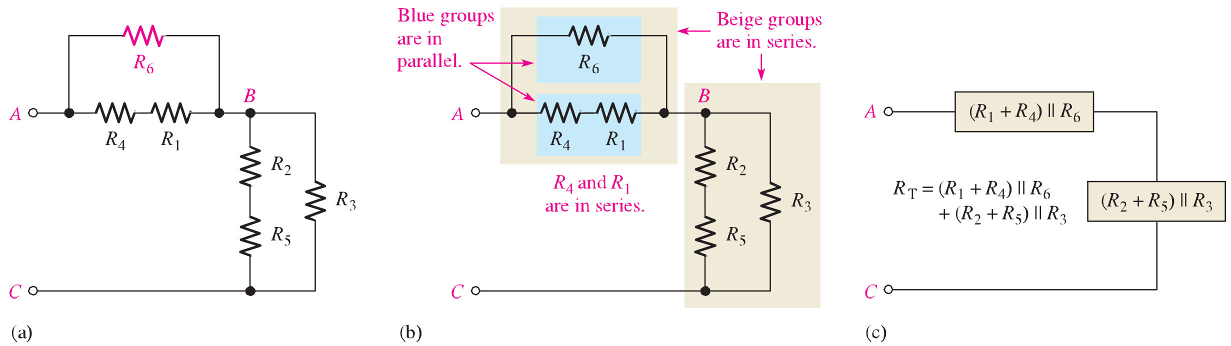
A
B
C
R_6
R_4
R_1
R_2
R_3
R_5
Blue groups are in parallel.
Beige groups are in series.
R_4 and R_1 are in series.
$(R_1 + R_4) \parallel R_6$
$R_T = (R_1 + R_4) \parallel R_6 + (R_2 + R_5) \parallel R_3$
$(R_2 + R_5) \parallel R_3$

(a) (b) (c)

FIGURE 6–4 **R_6 is added to the circuit in parallel with the series combination of R_1 and R_4.**

EXAMPLE 6–1

Identify the series-parallel relationships in Figure 6–5.

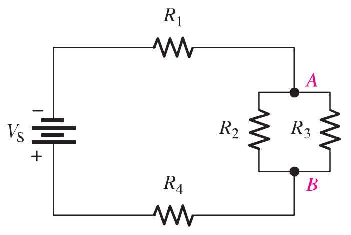
R_1
A
V_S
−
+
R_2
R_3
B
R_4

FIGURE 6–5

SOLUTION

Starting at the negative terminal of the source, follow the current paths.

1. All the current produced by the source must go through R_1, which is in series with the rest of the circuit.
2. The total current takes two paths when it gets to node A. Part of it is through R_2, and part of it is through R_3.
3. Resistors R_2 and R_3 are in parallel with each other because they are connected to the same pair of nodes. This parallel combination is in series with R_1.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

4. At node B, the currents through R_2 and R_3 come together again into a single path. Thus, the total current is through R_4.
5. Resistor R_4 is in series with both R_1 and the parallel combination of R_2 and R_3.

The currents are shown in Figure 6–6, where I_T is the total current. In summary, R_1 and R_4 are in series with the parallel combination of R_2 and R_3.

$$R_1 + R_4 + R_2 \| R_3$$

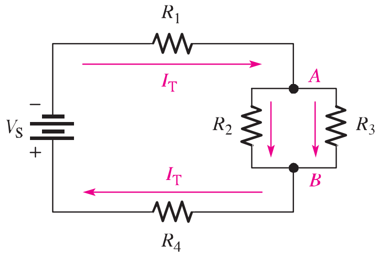

FIGURE 6–6

RELATED PROBLEM*

If another resistor, R_5, is connected from node A to the positive side of the source in Figure 6–6, what is its relationship to the other resistors?

0258262 2015/12/10 12.27.66.5

*Answers are at the end of the chapter.

EXAMPLE 6–2

Describe the series-parallel combination between terminals A and D in Figure 6–7.

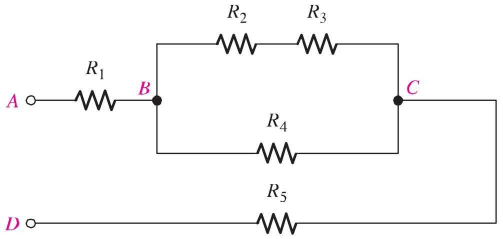

FIGURE 6–7

SOLUTION

Between nodes B and C, there are two parallel paths.

1. The lower path consists of R_4.
2. The upper path consists of a series combination of R_2 and R_3.

This parallel combination is in series with both R_1 and R_5. In summary, R_1 and R_5 are in series with the parallel combination of R_4 and $(R_2 + R_3)$.

$$R_1 + R_5 + R_4 \| (R_2 + R_3)$$

RELATED PROBLEM

If a resistor is connected from node C to node D in Figure 6–7, describe its relationship in the circuit.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 6–3

Describe the total resistance between each pair of terminals in Figure 6–8.

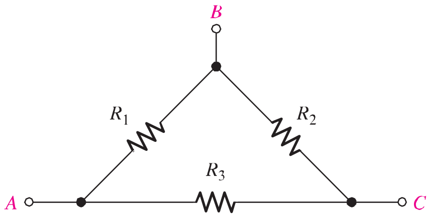

FIGURE 6–8

SOLUTION

1. From *A* to *B:* R_1 is in parallel with the series combination of R_2 and R_3.
2. From *A* to *C:* R_3 is in parallel with the series combination of R_1 and R_2.
3. From *B* to *C:* R_2 is in parallel with the series combination of R_1 and R_3.

RELATED PROBLEM

In Figure 6–8, describe the total resistance between each terminal and an added ground if a new resistor, R_4, is connected from terminal *C* to ground. None of the existing resistors connect directly to the ground.

When connecting a circuit on a protoboard from a schematic, it is easier to check the circuit if the resistors and connections on the protoboard are oriented so that they approximately match the way the schematic is drawn. In some cases, it is difficult to see the series-parallel relationships on a schematic because of the way in which it is drawn. In such a situation, it helps to redraw the diagram so that the relationships become clear. The physical arrangement of components on a PC or protoboard usually bears no resemblance to the actual electrical relationships. By tracing out the circuit and rearranging the components on paper into a recognizable form, you can determine the series-parallel relationships.

EXAMPLE 6–4

Identify the series-parallel relationships in Figure 6–9.

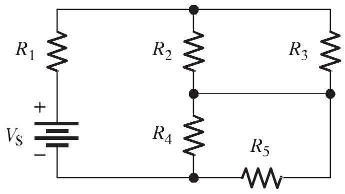

FIGURE 6–9

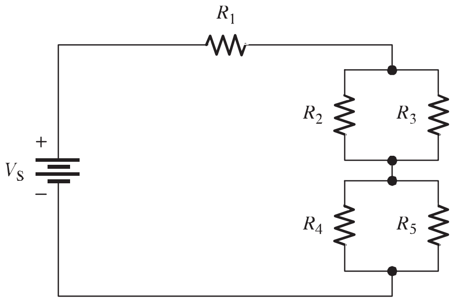

FIGURE 6–10

SOLUTION

The circuit schematic is redrawn in Figure 6–10 to better illustrate the series-parallel relationships. Now you can see that R_2 and R_3 are in parallel with each

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

other and also that R_4 and R_5 are in parallel with each other. Both parallel combinations are in series with each other and with R_1.

$$R_1 + R_2 \| R_3 + R_4 \| R_5$$

RELATED PROBLEM

If a resistor is connected from the bottom end of R_3 to the top end of R_5 in Figure 6–10, what effect does it have on the circuit? Explain.

SECTION 6–1 CHECKUP*

1. A certain series-parallel circuit is described as follows: R_1 and R_2 are in parallel. This parallel combination is in series with another parallel combination of R_3 and R_4. Draw the circuit.
2. In the circuit of Figure 6–11, describe the series-parallel relationships of the resistors.
3. Which resistors are in parallel in Figure 6–12?
4. Identify the parallel resistors in Figure 6–13.
5. Are the parallel combinations in Figure 6–13 in series?

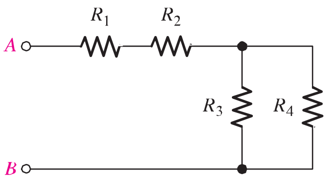

FIGURE 6–11

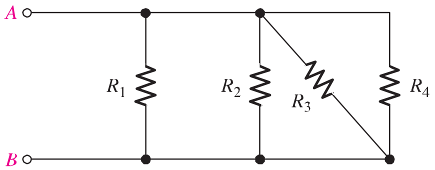

FIGURE 6–12

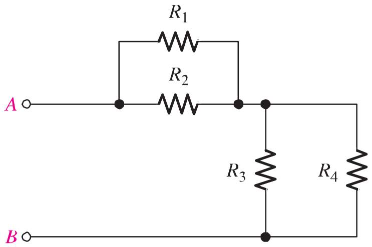

FIGURE 6–13

*Answers are at the end of the chapter.

6–2 ANALYSIS OF SERIES-PARALLEL RESISTIVE CIRCUITS

The analysis of series-parallel circuits can be approached in many ways, depending on what information you need and what circuit values you know. The examples in this section do not represent an exhaustive coverage, but they give you an idea of how to approach series-parallel circuit analysis.

After completing this section, you should be able to

- **Analyze series-parallel circuits**
  - **Determine total resistance**
  - **Determine all the currents**
  - **Determine all the voltage drops**

If you know Ohm's law, Kirchhoff's laws, the voltage-divider formula, and the current-divider formula, and if you know how to apply these laws, you can solve most resistive circuit analysis problems. The ability to recognize series and parallel combinations is, of course, essential. There is no standard "cookbook" approach that can be applied to all situations. Logical thought is the most powerful tool you can apply to problem solving.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Total Resistance

In Chapter 4, you learned how to determine total series resistance. In Chapter 5, you learned how to determine total parallel resistance. To find the total resistance (R_T) of a series-parallel combination, first identify the series and parallel relationships, and then simplify the circuit by finding the total resistance. The following two examples illustrate the general approach.

EXAMPLE 6–5

Determine R_T between terminals *A* and *B* of the circuit in Figure 6–14.

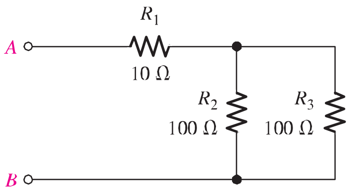

FIGURE 6–14

SOLUTION

Resistors R_2 and R_3 are in parallel, and this parallel combination is in series with R_1. First find the parallel resistance of R_2 and R_3. Since R_2 and R_3 are equal in value, divide the value by 2.

$$R_{2\|3} = \frac{R}{n} = \frac{100\ \Omega}{2} = 50\ \Omega$$

Now, since R_1 is in series with $R_{2\|3}$, add their values.

$$R_T = R_1 + R_{2\|3} = 10\ \Omega + 50\ \Omega = \mathbf{60\ \Omega}$$

RELATED PROBLEM

Determine R_T in Figure 6–14 if R_3 is changed to 82 Ω.

MULTISIM

Open Multisim file E06-05; files are found at www.pearsonhighered.com/floyd. Using the multimeter, verify the calculated value of total resistance. Change R_1 to 18 Ω, R_2 to 82 Ω, and R_3 to 82 Ω and measure the total resistance.

EXAMPLE 6–6

Find R_T of the circuit in Figure 6–15.

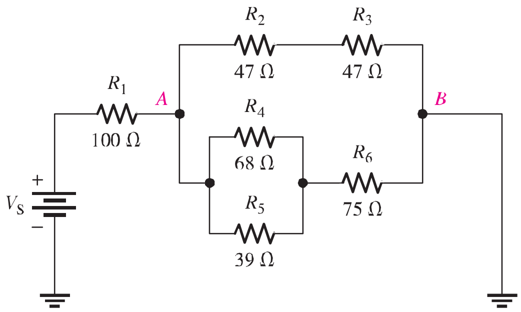

FIGURE 6–15

SOLUTION

6238262 2013/12/10 12.27.66.5

1. In the upper branch between nodes *A* and *B*, R_2 is in series with R_3. The series combination is designated R_{2+3} and is equal to $R_2 + R_3$.

$$R_{2+3} = R_2 + R_3 = 47\ \Omega + 47\ \Omega = 94\ \Omega$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

2. In the lower branch, R_4 and R_5 are in parallel with each other. This parallel combination is designated $R_{4\|5}$.

$$R_{4\|5} = \frac{R_4R_5}{R_4 + R_5} = \frac{(68\ \Omega)(39\ \Omega)}{68\ \Omega + 39\ \Omega} = 24.8\ \Omega$$

3. Also in the lower branch, the parallel combination of R_4 and R_5 is in series with R_6. This series-parallel combination is designated $R_{4\|5+6}$.

$$R_{4\|5+6} = R_6 + R_{4\|5} = 75\ \Omega + 24.8\ \Omega = 99.8\ \Omega$$

Figure 6–16 shows the original circuit in a simplified equivalent form.

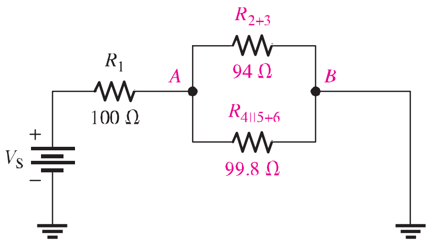

FIGURE 6–16

4. Now you can find the resistance between nodes A and B. It is R_{2+3} in parallel with $R_{4\|5+6}$. Calculate the equivalent resistance as follows:

$$R_{AB} = \frac{1}{\dfrac{1}{R_{2+3}} + \dfrac{1}{R_{4\|5+6}}} = \frac{1}{\dfrac{1}{94\ \Omega} + \dfrac{1}{99.8\ \Omega}} = 48.4\ \Omega$$

5. Finally, the total circuit resistance is R_1 in series with R_{AB}.

$$R_T = R_1 + R_{AB} = 100\ \Omega + 48.4\ \Omega = \mathbf{148\ \Omega}$$

MULTISIM

Open Multisim file E06-06. Verify the total calculated resistance. Remove R_5 from the circuit and measure the total resistance. Then check your measured value by calculating the total resistance.

RELATED PROBLEM

Determine R_T if a 68 Ω resistor is connected to the circuit in Figure 6–16 from node A to node B.

Total Current

Once you know the total resistance and the source voltage, you can apply Ohm's law to find the total current in a circuit. Total current is the source voltage divided by the total resistance.

$$I_T = \frac{V_S}{R_T}$$

For example, let's find the total current in the circuit of Example 6–6 (Figure 6–15). Assume that the source voltage is 10 V. The calculation is

$$I_T = \frac{V_S}{R_T} = \frac{10\ \text{V}}{148\ \Omega} = 67.6\ \text{mA}$$

Branch Currents

Using the current-divider formula, Kirchhoff's current law, Ohm's law, or combinations of these, you can find the current in any branch of a series-parallel circuit. In some cases, it may take repeated application of the formula to find a given current.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 6–7

Determine the current through R_4 in Figure 6–17 if $V_S = 5.0$ V.

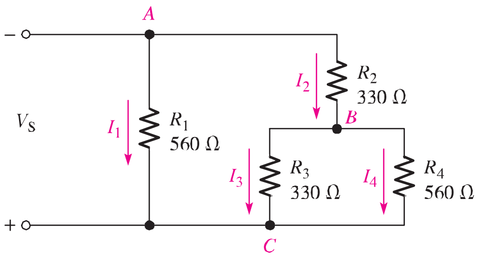

FIGURE 6–17

SOLUTION

First, find the current (I_2) into node *B*. Once you know this current, you can use the current-divider formula to find I_4, the current through R_4.

Notice that there are two main branches in the circuit. The left-most branch consists of only R_1. The right-most branch has R_2 in series with the parallel combination of R_3 and R_4. The voltage across both of these main branches is the same and equal to 5.0 V. Calculate the equivalent resistance ($R_{2+3\|4}$) of the right-most main branch and then apply Ohm's law; I_2 is the total current through this main branch. Thus,

$$R_{2+3\|4} = R_2 + \frac{R_3R_4}{R_3 + R_4} = 330\ \Omega + \frac{(330\ \Omega)(560\ \Omega)}{890\ \Omega} = 538\ \Omega$$

$$I_2 = \frac{V_S}{R_{2+3\|4}} = \frac{5.0\text{ V}}{538\ \Omega} = 9.29\text{ mA}$$

Use the two-resistor current-divider formula to calculate I_4.

$$I_4 = \left(\frac{R_3}{R_3 + R_4}\right)I_2 = \left(\frac{330\ \Omega}{890\ \Omega}\right)9.29\text{ mA} = \mathbf{3.45\ mA}$$

RELATED PROBLEM

Find I_1, I_3, and I_T in Figure 6–17.

MULTISIM

Open Multisim file E06-07. Measure the current in each resistor. Compare the measurements with the calculated values.

Voltage Relationships

The circuit in Figure 6–18 illustrates voltage relationships in a series-parallel circuit. Voltmeters are connected to measure each of the resistor voltages, and the readings are indicated.

Some general observations about Figure 6–18 are as follows:

1. V_{R1} and V_{R2} are equal because R_1 and R_2 are in parallel. (Recall that voltages across parallel branches are the same.) V_{R1} and V_{R2} are the same as the voltage from *A* to *B*.
2. V_{R3} is equal to $V_{R4} + V_{R5}$ because R_3 is in parallel with the series combination of R_4 and R_5. (V_{R3} is the same as the voltage from *B* to *C*.)
3. V_{R4} is about one-third of the voltage from *B* to *C* because R_4 is about one-third of the resistance $R_4 + R_5$ (by the voltage-divider principle).
4. V_{R5} is about two-thirds of the voltage from *B* to *C* because R_5 is about two-thirds of $R_4 + R_5$.
5. $V_{R1} + V_{R3} - V_S = 0$ because, by Kirchhoff's voltage law, the algebraic sum of the voltage drops around a single closed path must equal zero.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

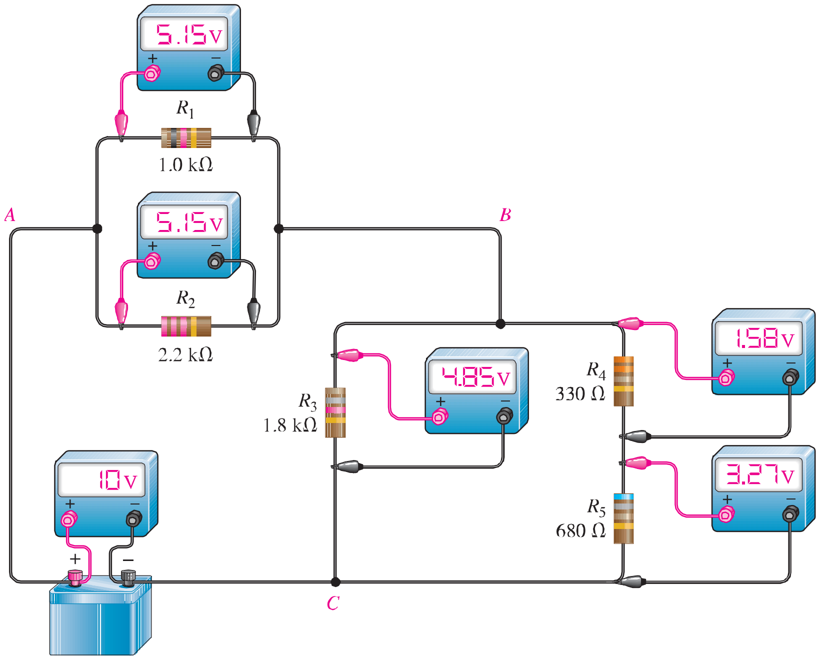

FIGURE 6–18 **Illustration of voltage relationships.**

6238262 2013/12/10 12.27.66.5

Example 6–8 will verify the meter readings in Figure 6–18.

EXAMPLE 6–8

Verify that the voltmeter readings in Figure 6–18 are correct. The circuit is redrawn as a schematic in Figure 6–19.

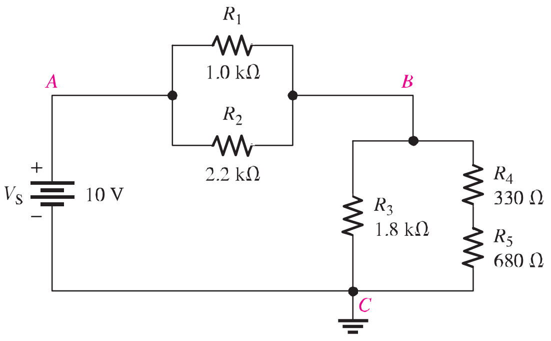

FIGURE 6–19

SOLUTION

The resistance from A to B is the parallel combination of R_1 and R_2.

$$R_{AB} = \frac{R_1 R_2}{R_1 + R_2} = \frac{(1.0\ \text{k}\Omega)(2.2\ \text{k}\Omega)}{3.2\ \text{k}\Omega} = 688\ \Omega$$

The resistance from B to C is R_3 in parallel with the series combination of R_4 and R_5.

$$R_4 + R_5 = 330\ \Omega + 680\ \Omega = 1010\ \Omega = 1.01\ \text{k}\Omega$$

$$R_{BC} = \frac{R_3(R_4 + R_5)}{R_3 + R_4 + R_5} = \frac{(1.8\ \text{k}\Omega)(1.01\ \text{k}\Omega)}{2.81\ \text{k}\Omega} = 647\ \Omega$$

The resistance from A to B is in series with the resistance from B to C, so the total circuit resistance is

$$R_T = R_{AB} + R_{BC} = 688\ \Omega + 647\ \Omega = 1335\ \Omega$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Use the voltage-divider principle to calculate the voltages.

$$V_{AB} = \left(\frac{R_{AB}}{R_T}\right)V_S = \left(\frac{688\ \Omega}{1335\ \Omega}\right)10\text{ V} = 5.15\text{ V}$$

$$V_{BC} = \left(\frac{R_{BC}}{R_T}\right)V_S = \left(\frac{647\ \Omega}{1335\ \Omega}\right)10\text{ V} = 4.85\text{ V}$$

$$V_{R1} = V_{R2} = V_{AB} = \mathbf{5.15\ V}$$

$$V_{R3} = V_{BC} = \mathbf{4.85\ V}$$

$$V_{R4} = \left(\frac{R_4}{R_4 + R_5}\right)V_{BC} = \left(\frac{330\ \Omega}{1010\ \Omega}\right)4.85\text{ V} = \mathbf{1.58\ V}$$

$$V_{R5} = \left(\frac{R_5}{R_4 + R_5}\right)V_{BC} = \left(\frac{680\ \Omega}{1010\ \Omega}\right)4.85\text{ V} = \mathbf{3.27\ V}$$

RELATED PROBLEM

Determine each voltage drop in Figure 6–19 if the source voltage is doubled.

MULTISIM

Open Multisim file E06-08. Measure the voltage across each resistor and compare to the calculated values. Verify by measurement that each voltage drop doubles if the source voltage is doubled and is halved if the source voltage is halved.

EXAMPLE 6–9

Determine the voltage drop across each resistor in Figure 6–20.

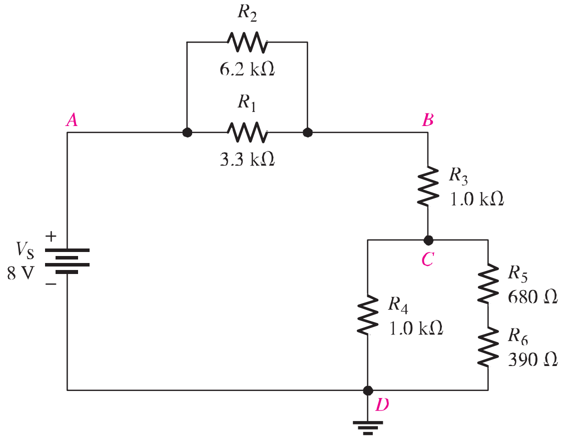

FIGURE 6–20

SOLUTION

Because you know the total voltage, you can solve this problem by forming an equivalent series circuit and applying the voltage-divider formula.

Step 1: Reduce each parallel combination to an equivalent resistance. Since R_1 and R_2 are in parallel between nodes A and B, combine their values.

$$R_{AB} = \frac{R_1R_2}{R_1 + R_2} + \frac{(3.3\text{ k}\Omega)(6.2\text{ k}\Omega)}{9.5\text{ k}\Omega} = 2.15\text{ k}\Omega$$

Since R_4 is in parallel with the R_5 and R_6 series combination (R_{5+6}) between nodes C and D, combine these values.

$$R_{CD} = \frac{R_4(R_{5+6})}{R_4 + R_{5+6}} + \frac{(1.0\text{ k}\Omega)(1.07\text{ k}\Omega)}{2.07\text{ k}\Omega} = 517\ \Omega$$

Step 2: Draw the equivalent circuit as shown in Figure 6–21. The total circuit resistance is

$$R_T = R_{AB} + R_3 + R_{CD} = 2.15\text{ k}\Omega + 1.0\text{ k}\Omega + 517\ \Omega = 3.67\text{ k}\Omega$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

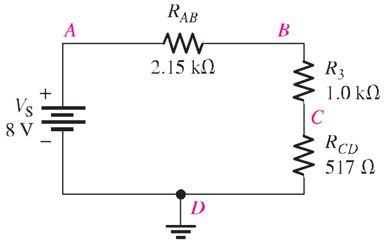
FIGURE 6–21

Step 3: Apply the voltage-divider formula to solve for the voltages in the equivalent series circuit.

$$V_{AB} = \left(\frac{R_{AB}}{R_T}\right)V_S = \left(\frac{2.15\ \text{k}\Omega}{3.67\ \text{k}\Omega}\right)8\ \text{V} = 4.69\ \text{V}$$

$$V_{BC} = \left(\frac{R_3}{R_T}\right)V_S = \left(\frac{1.0\ \text{k}\Omega}{3.67\ \text{k}\Omega}\right)8\ \text{V} = 2.18\ \text{V}$$

$$V_{CD} = \left(\frac{R_{CD}}{R_T}\right)V_S = \left(\frac{517\ \Omega}{3.67\ \text{k}\Omega}\right)8\ \text{V} = 1.13\ \text{V}$$

Refer to Figure 6–20. V_{AB} equals the voltage across both R_1 and R_2.

$$V_{R1} = V_{R2} = V_{AB} = \mathbf{4.69\ V}$$

V_{BC} is the voltage across R_3.

$$V_{R3} = V_{BC} = \mathbf{2.18\ V}$$

V_{CD} is the voltage across R_4 and also across the series combination of R_5 and R_6.

$$V_{R4} = V_{CD} = \mathbf{1.13\ V}$$

MULTISIM

Open Multisim file E06-09. Measure the voltage across each resistor and compare to the calculated values. If R_4 is increased to 2.2 kΩ, specify which voltage drops increase and which ones decrease. Verify this by measurement.

Step 4: Apply the voltage-divider formula to the series combination of R_5 and R_6 to get V_{R5} and V_{R6}.

$$V_{R5} = \left(\frac{R_5}{R_5 + R_6}\right)V_{CD} = \left(\frac{680\ \Omega}{1070\ \Omega}\right)1.13\ \text{V} = \mathbf{718\ mV}$$

$$V_{R6} = \left(\frac{R_6}{R_5 + R_6}\right)V_{CD} = \left(\frac{390\ \Omega}{1070\ \Omega}\right)1.13\ \text{V} = \mathbf{412\ mV}$$

RELATED PROBLEM

Determine the current and power in each resistor in Figure 6–20.

SECTION 6–2 CHECKUP

1. Find the total resistance between A and B in the circuit of Figure 6–22.
2. Find the current through R_3 in Figure 6–22.
3. Find V_{R2} in Figure 6–22.
4. Determine R_T and I_T in Figure 6–23.

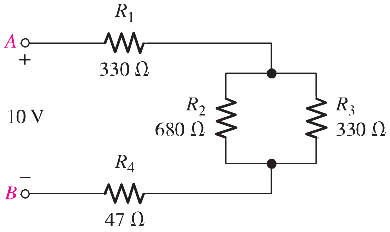
FIGURE 6–22

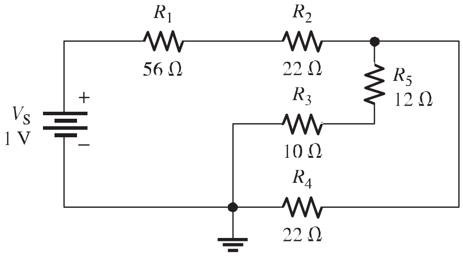
FIGURE 6–23

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6–3 VOLTAGE DIVIDERS WITH RESISTIVE LOADS

Voltage dividers were introduced in Chapter 4. In this section, you will learn how resistive loads affect the operation of voltage-divider circuits.

After completing this section, you should be able to

- **Analyze loaded voltage dividers**
  - **Determine the effect of a resistive load on a voltage-divider circuit**
  - **Define *bleeder current***

The voltage divider in Figure 6–24(a) produces an output voltage (V_{OUT}) of 5 V because the input is 10 V and the two resistors are of equal value. This voltage is the unloaded output voltage. When a load resistor, R_L, is connected from the output to ground as shown in Figure 6–24(b), the output voltage is reduced by an amount that depends on the value of R_L. This effect is called **loading**. The load resistor is in parallel with R_2, reducing the resistance from node A to ground and, as a result, also reducing the voltage across the parallel combination. This is one effect of loading a voltage divider. Another effect of a load is that more current is drawn from the source because the total resistance of the circuit is reduced.

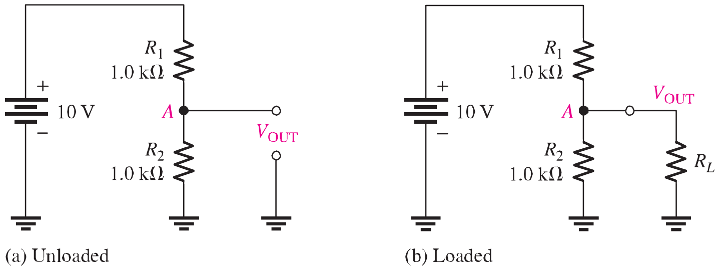

(a) Unloaded (b) Loaded

FIGURE 6–24 **A voltage divider with both unloaded and loaded outputs.**

The effect of loading on a voltage divider is important to take into account when choosing resistors for a divider. When R_L is large compared to the divider resistors, the loading effect is small and the output voltage will change only a small amount from its unloaded value. If the loading effect is small, the divider is said to be a *stiff voltage divider*. This term is relative but generally means that the difference between the unloaded and loaded output voltage is small. As a rule of thumb, a **stiff voltage divider** is one in which the load resistor is at least ten times larger than the divider resistors. Stiff voltage dividers are more stable but use more power, so the choice of divider resistors is a trade-off between these two parameters. Figure 6–25 illustrates the effect of a load resistor on the output voltage.

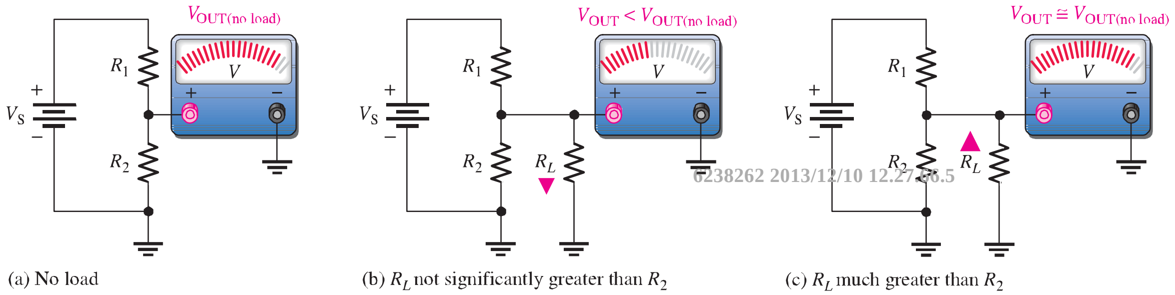

6238262 2013/12/10 12.27.86.5

(a) No load (b) R_L not significantly greater than R_2 (c) R_L much greater than R_2

FIGURE 6–25 **The effect of a load resistor. The circuit in part (c) illustrates a stiff voltage divider.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 6–10

(a) Determine the unloaded output voltage of the voltage divider in Figure 6–26.

(b) Find the loaded output voltages of the voltage divider in Figure 6–26 for the following two values of load resistance: $R_L = 10\ \text{k}\Omega$ and $R_L = 100\ \text{k}\Omega$.

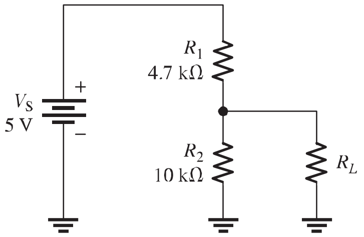

FIGURE 6–26

SOLUTION

(a) The unloaded output voltage is

$$V_{\text{OUT(unloaded)}} = \left(\frac{R_2}{R_1 + R_2}\right)V_S = \left(\frac{10\ \text{k}\Omega}{14.7\ \text{k}\Omega}\right)5\ \text{V} = \mathbf{3.40\ V}$$

(b) With the $10\ \text{k}\Omega$ load resistor connected, R_L is in parallel with R_2, which gives

$$R_2 \| R_L = \frac{R_2 R_L}{R_2 + R_L} = \frac{(10\ \text{k}\Omega)(10\ \text{k}\Omega)}{20\ \text{k}\Omega} = 5.0\ \text{k}\Omega$$

The equivalent circuit is shown in Figure 6–27(a). The loaded output voltage is

$$V_{\text{OUT(loaded)}} = \left(\frac{R_2 \| R_L}{R_1 + R_2 \| R_L}\right)V_S = \left(\frac{5.0\ \text{k}\Omega}{9.7\ \text{k}\Omega}\right)5\ \text{V} = \mathbf{2.58\ V}$$

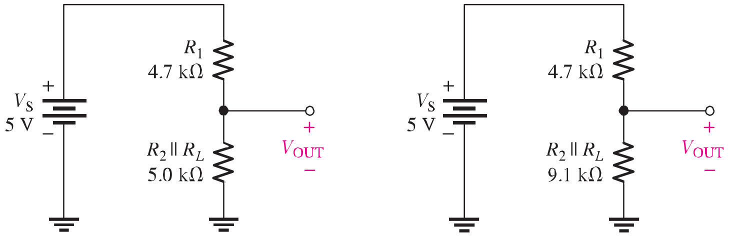

(a) $R_L = 10\ \text{k}\Omega$ (b) $R_L = 100\ \text{k}\Omega$

FIGURE 6–27

With the $100\ \text{k}\Omega$ load, the resistance from output to ground is

$$R_2 \| R_L = \frac{R_2 R_L}{R_2 + R_L} = \frac{(10\ \text{k}\Omega)(100\ \text{k}\Omega)}{110\ \text{k}\Omega} = 9.1\ \text{k}\Omega$$

The equivalent circuit is shown in Figure 6–27(b). The loaded output voltage is

$$V_{\text{OUT(loaded)}} = \left(\frac{R_2 \| R_L}{R_1 + R_2 \| R_L}\right)V_S = \left(\frac{9.1\ \text{k}\Omega}{13.8\ \text{k}\Omega}\right)5\ \text{V} = \mathbf{3.30\ V}$$

For the smaller value of R_L, the reduction in V_{OUT} is

$$3.40\ \text{V} - 2.58\ \text{V} = 0.82\ \text{V} \qquad \text{(a 24\% drop in output voltage)}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

For the larger value of R_L, the reduction in V_{OUT} is

$$3.40\text{ V} - 3.30\text{ V} = 0.10\text{ V} \qquad \text{(a 3\% drop in output voltage)}$$

This illustrates the loading effect of R_L on the voltage divider.

RELATED PROBLEM

Determine V_{OUT} in Figure 6–26 for a 1.0 MΩ load resistance.

MULTISIM

Open Multisim file E06-10. Measure the voltage at the output terminal with respect to ground. Connect a 10 kΩ load resistor from the output to ground and measure the output voltage. Change the load resistor to 100 kΩ and measure the output voltage. Do these measurements agree closely with your calculated values?

Load Current and Bleeder Current

In a multiple-tap loaded voltage-divider circuit, the total current drawn from the source consists of currents through the load resistors, called **load currents**, and the divider resistors. Figure 6–28 shows a voltage divider with two voltage outputs or taps. Notice that the total current, I_T, is through R_1. The total current is composed of the two branch currents, I_{RL1} and I_2. The current I_2 is composed of two additional branch currents, I_{RL2} and I_3. Current I_3 is called the **bleeder current**, which is the current left after the total load current is subtracted from the total current in the circuit.

$$I_{\text{BLEEDER}} = I_T - I_{RL1} - I_{RL2} \qquad (6\text{–}1)$$

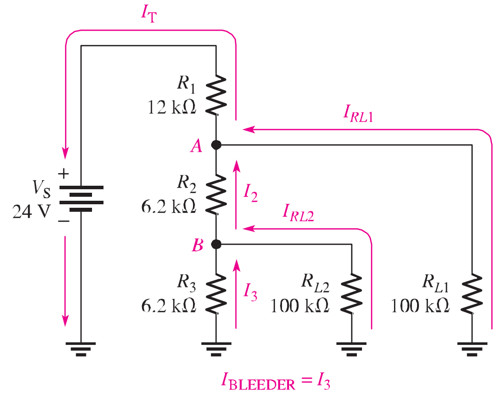

FIGURE 6–28 **Currents in a two-tap loaded voltage divider.**

EXAMPLE 6–11

Determine the load currents I_{RL1} and I_{RL2} and the bleeder current I_3 in the two-tap loaded voltage divider in Figure 6–28.

SOLUTION

The equivalent resistance from node A to ground is the 100 kΩ load resistor R_{L1} in parallel with the combination of R_2 in series with the parallel combination of R_3 and R_{L2}. Determine the resistance values first. R_3 in parallel with R_{L2} is designated R_B. The resulting equivalent circuit is shown in Figure 6–29(a).

$$R_B = \frac{R_3 R_{L2}}{R_3 + R_{L2}} = \frac{(6.2\text{ k}\Omega)(100\text{ k}\Omega)}{106.2\text{ k}\Omega} = 5.84\text{ k}\Omega$$

R_2 in series with R_B is designated R_{2+B}. The resulting equivalent circuit is shown in Figure 6–29(b).

$$R_{2+B} = R_2 + R_B = 6.2\text{ k}\Omega + 5.84\text{ k}\Omega = 12.0\text{ k}\Omega$$

R_{L1} in parallel with R_{2+B} is designated R_A. The resulting equivalent circuit is shown in Figure 6–29(c).

$$R_A = \frac{R_{L1} R_{2+B}}{R_{L1} + R_{2+B}} = \frac{(100\text{ k}\Omega)(12.0\text{ k}\Omega)}{112\text{ k}\Omega} = 10.7\text{ k}\Omega$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

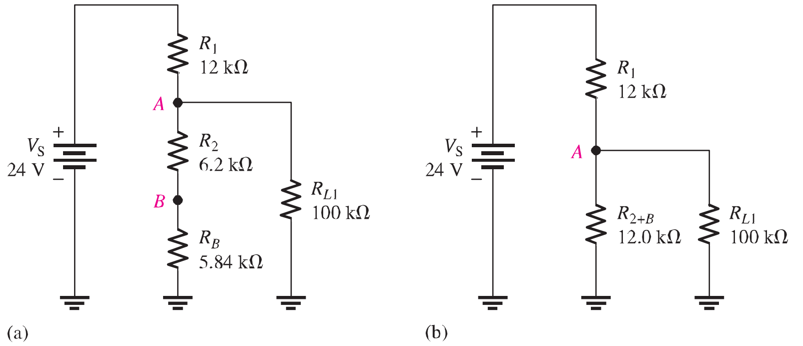

FIGURE 6–29

R_A is the total resistance from node A to ground. The total resistance for the circuit is

$$R_T = R_A + R_1 = 10.7\text{ k}\Omega + 12\text{ k}\Omega = 22.7\text{ k}\Omega$$

Determine the voltage across R_{L1} as follows, using the equivalent circuit in Figure 6–29(c):

$$V_{RL1} = V_A = \left(\frac{R_A}{R_T}\right)V_S = \left(\frac{10.7\text{ k}\Omega}{22.7\text{ k}\Omega}\right)24\text{ V} = 11.3\text{ V}$$

The load current through R_{L1} is

$$I_{RL1} = \frac{V_{RL1}}{R_{L1}} = \left(\frac{11.3\text{ V}}{100\text{ k}\Omega}\right) = \mathbf{113\ \mu A}$$

Determine the voltage at node B by using the equivalent circuit in Figure 6–29(a) and the voltage at node A.

$$V_B = \left(\frac{R_B}{R_{2+B}}\right)V_A = \left(\frac{5.84\text{ k}\Omega}{12.0\text{ k}\Omega}\right)11.3\text{ V} = 5.50\text{ V}$$

The load current through R_{L2} is

$$I_{RL2} = \frac{V_{RL2}}{R_{L2}} = \frac{V_B}{R_{L2}} = \frac{5.50\text{ V}}{100\text{ k}\Omega} = \mathbf{55\ \mu A}$$

The bleeder current is

$$I_3 = \frac{V_B}{R_3} = \frac{5.50\text{ V}}{6.2\text{ k}\Omega} = \mathbf{887\ \mu A}$$

RELATED PROBLEM

What will happen to the load current in R_{L2} if R_{L1} is disconnected?

MULTISIM

Open Multisim file E06-11. Measure the voltage across and the current through each load resistor, R_{L1} and R_{L2}.

SYSTEM EXAMPLE 6–1

DISCRETE AMPLIFIERS WITH VOLTAGE-DIVIDER BIAS

Amplifiers are fundamental to electronic systems because they are the basis of many other circuits. In linear circuits, bipolar transistors are often selected because of their linear characteristics. For a small-signal bipolar transistor amplifier, a voltage-divider is commonly used to develop the correct dc conditions (called *bias*) in order for the transistor to amplify. Figure 6–30 shows a simplified bipolar transistor voltage amplifier. The transistor requires a certain amount of base current to operate, which is set up by the voltage-divider bias network composed of R_1 and R_2. In an amplifier in which bias is set by a voltage divider, the base current should normally be much smaller ($<10\%$) than the bleeder current. When

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

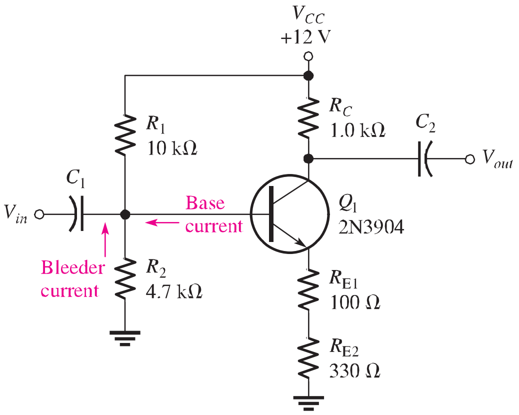

V_{CC}
+12 V
R_C
1.0 kΩ
C_2
V_{out}
R_1
10 kΩ
C_1
V_{in}
Base
current
Q_1
2N3904
Bleeder
current
R_2
4.7 kΩ
R_{E1}
100 Ω
R_{E2}
330 Ω

FIGURE 6–30 **A voltage amplifier.**

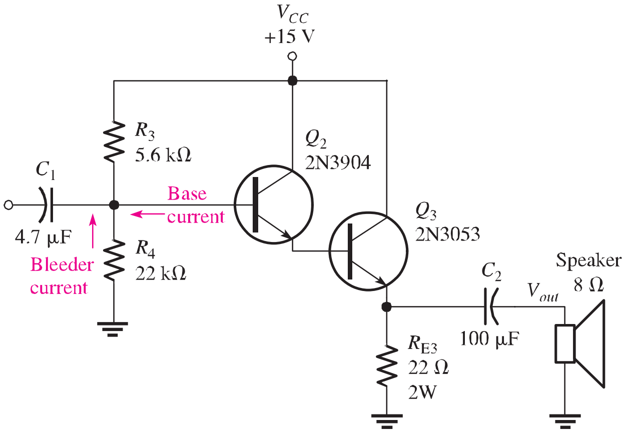

V_{CC}
+15 V
R_3
5.6 kΩ
Q_2
2N3904
C_1
4.7 µF
Base
current
Bleeder
current
R_4
22 kΩ
Q_3
2N3053
C_2
Speaker
8 Ω
V_{out}
100 µF
R_{E3}
22 Ω
2W

FIGURE 6–31 **A power amplifier.**

this condition is met, the basic calculation of base voltage is simple; simply apply the unloaded voltage-divider rule you already learned, substituting V_{CC} for the source voltage, V_S, in the original equation:

$$V_{BASE} = \left(\frac{R_2}{R_1 + R_2}\right)V_{CC} = \left(\frac{4.7\text{ k}\Omega}{10\text{ k}\Omega + 4.7\text{ k}\Omega}\right)12\text{ V} = 3.84\text{ V}$$

The actual base voltage in this circuit is 3.7 V, showing a small loading effect.

Figure 6–31 shows another example with voltage-divider bias, but the load is a speaker and the amplifier is a power amplifier. The divider for a power amplifier will usually have different values for the divider string depending on the type of amplifier and the bias current required for proper operation. If the bias current requirement is large, the bias resistors will be smaller. The simplifying assumption of an unloaded voltage divider can be used whenever the base current is small compared to the bleeder current, and the result is a reasonable estimate of the base voltage.

SECTION 6–3 CHECKUP

1. A load resistor is connected to an output on a voltage divider. What effect does the load resistor have on the output voltage?
2. A larger-value load resistor will cause the output voltage of a voltage divider to change less than a small-value one will. (True or False)
3. For the voltage divider in Figure 6–32, determine the unloaded output voltage. Also determine the output voltage with a 10 MΩ load resistor connected from the output to ground.

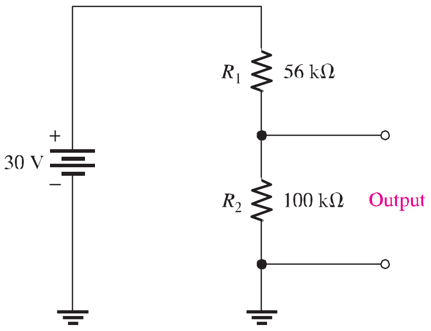

R_1
56 kΩ
+
30 V
−
R_2
100 kΩ
Output

FIGURE 6–32

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6–4 LOADING EFFECT OF A VOLTMETER

As you have learned, voltmeters must be connected in parallel with a resistor in order to measure the voltage across the resistor. Because of its internal resistance, a voltmeter, or any other measuring instrument for that matter, puts a load on the circuit and will affect, to a certain extent, the voltage that is being measured. Until now, we have ignored the loading effect because the internal resistance of a voltmeter is very high, and normally it has negligible effect on the circuit that is being measured. However, if the internal resistance of the voltmeter is not sufficiently greater than the circuit resistance across which it is connected, the loading effect will cause the measured voltage to be less than its actual value.

After completing this section, you should be able to

- **Determine the loading effect of a voltmeter on a circuit**
  - **Explain why a voltmeter can load a circuit**
  - **Discuss the internal resistance of a voltmeter**

When a voltmeter is connected to a circuit as shown, for example, in Figure 6–33(a), its internal resistance appears in parallel with R_3, as shown in part (b). The resistance from A to B is altered by the loading effect of the voltmeter's internal resistance, R_M, and is equal to $R_3 \| R_M$, as indicated in part (c).

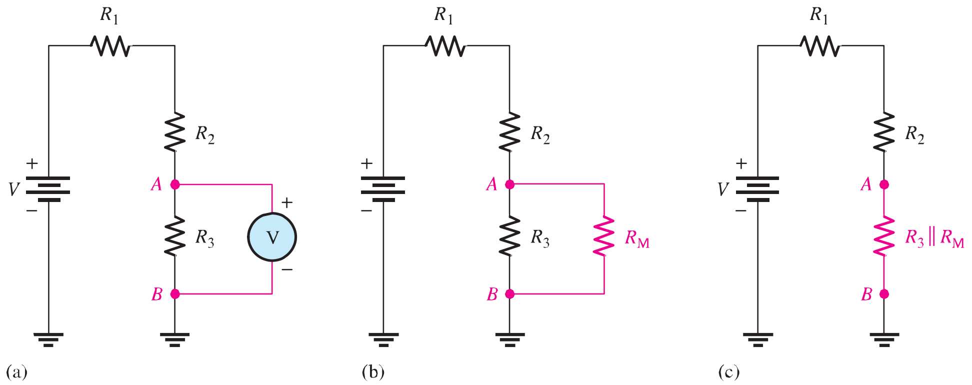

(a) (b) (c)

FIGURE 6–33 **The loading effect of a voltmeter.**

If R_M is much greater than R_3, the resistance from A to B changes very little, and the meter reading is very close to the actual voltage when no meter is present. If R_M is not sufficiently greater than R_3, the resistance from A to B is reduced significantly, and the voltage across R_3 is altered by the loading effect of the meter. A good rule of thumb for troubleshooting work is that *if the meter resistance is at least ten times greater than the resistance across which it is connected, the loading effect can be neglected (measurement error is less than 10%).*

Most voltmeters are part of a multifunction instrument such as the DMM or the analog multimeter discussed in Section 2–7. The voltmeter in a DMM will typically have an internal resistance of 10 MΩ or more, so the loading effect is important only in very high-resistance circuits. DMMs have a constant resistance on all ranges because the input is connected to an internal fixed voltage divider. For analog multimeters, the internal resistance depends on the range selected for making a measurement. To determine the loading effect, you need to know the **sensitivity** of the meter, a value given by the manufacturer on the meter or in the manual. Sensitivity is expressed in ohms/volt and is typically about 20,000 Ω/V. To determine the internal series resistance, multiply the sensitivity by the maximum voltage on the range selected. For example, a 20,000 Ω/V meter will have an internal resistance of 20,000 Ω on the 1 V range and 200,000 Ω on the 10 V range. As you can see, there is a smaller loading effect for higher voltage ranges than for lower ones on the analog multimeter.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 6–12

How much does the digital voltmeter affect the voltage being measured for each circuit indicated in Figure 6–34? Assume the meter has an input resistance (R_M) of 10 MΩ.

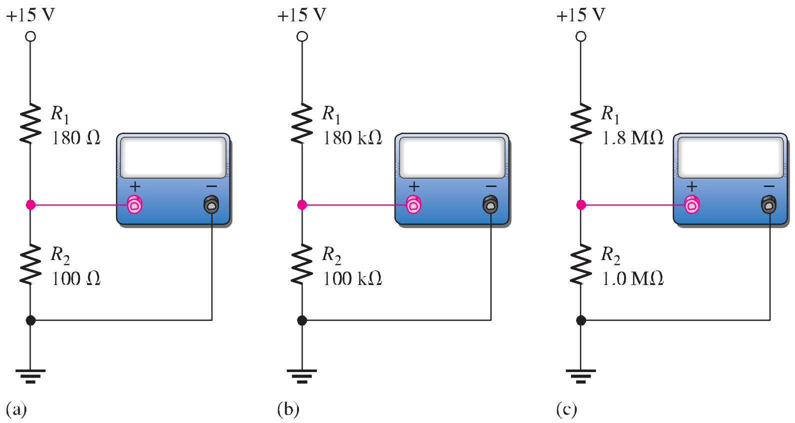

FIGURE 6–34

SOLUTION

To show the small differences more clearly, the results are expressed in more than three significant digits in this example.

(a) Refer to Figure 6–34(a). The unloaded voltage across R_2 in the voltage-divider circuit is

$$V_{R2} = \left(\frac{R_2}{R_1 + R_2}\right)V_S = \left(\frac{100\ \Omega}{280\ \Omega}\right)15\ \text{V} = 5.357\ \text{V}$$

The meter's resistance in parallel with R_2 is

$$R_2 \| R_M = \left(\frac{R_2 R_M}{R_2 + R_M}\right) = \frac{(100\ \Omega)(10\ \text{M}\Omega)}{10.0001\ \text{M}\Omega} = 99.999\ \Omega$$

The voltage actually measured by the meter is

$$V_{R2} = \left(\frac{R_2 \| R_M}{R_1 + R_2 \| R_M}\right)V_S = \left(\frac{99.999\ \Omega}{279.999\ \Omega}\right)15\ \text{V} = 5.357\ \text{V}$$

The voltmeter has no measurable loading effect.

(b) Refer to Figure 6–34(b).

$$V_{R2} = \left(\frac{R_2}{R_1 + R_2}\right)V_S = \left(\frac{100\ \text{k}\Omega}{280\ \text{k}\Omega}\right)15\ \text{V} = 5.357\ \text{V}$$

$$R_2 \| R_M = \frac{R_2 R_M}{R_2 + R_M} = \frac{(100\ \text{k}\Omega)(10\ \text{M}\Omega)}{10.1\ \text{M}\Omega} = 99.01\ \text{k}\Omega$$

The voltage actually measured by the meter is

$$V_{R2} = \left(\frac{R_2 \| R_M}{R_1 + R_2 \| R_M}\right)V_S = \left(\frac{99.01\ \text{k}\Omega}{279.01\ \text{k}\Omega}\right)15\ \text{V} = 5.323\ \text{V}$$

The loading effect of the voltmeter reduces the voltage by a very small amount.

0298262 2013/12/10 12.27.66.5

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

(c) Refer to Figure 6–34(c).

$$V_{R2} = \left(\frac{R_2}{R_1 + R_2}\right)V_S = \left(\frac{1.0\text{ M}\Omega}{2.8\text{ M}\Omega}\right)15\text{ V} = 5.357\text{ V}$$

$$R_2 \| R_M = \frac{R_2 R_M}{R_2 + R_M} = \frac{(1.0\text{ M}\Omega)(10\text{ M}\Omega)}{11\text{ M}\Omega} = 909.09\text{ k}\Omega$$

The voltage actually measured is

$$V_{R2} = \left(\frac{R_2 \| R_M}{R_1 + R_2 \| R_M}\right)V_S = \left(\frac{909.09\text{ k}\Omega}{2.709\text{ M}\Omega}\right)15\text{ V} = 5.034\text{ V}$$

The loading effect of the voltmeter reduces the voltage by a noticeable amount. As you can see, the higher the resistance across which a voltage is measured, the more the loading effect.

RELATED PROBLEM

Calculate the voltage across R_2 in Figure 6–34(c) if the meter resistance is 20 MΩ.

SECTION 6–4 CHECKUP

1. Explain why a voltmeter can potentially load a circuit.
2. If a voltmeter with a 10 MΩ internal resistance is measuring the voltage across a 1.0 kΩ resistor, should you normally be concerned about the loading effect?
3. If a voltmeter with a 10 MΩ resistance is measuring the voltage across a 3.3 MΩ resistor, should you be concerned about the loading effect?
4. What is the internal series resistance of a 20,000 Ω/V VOM if it is on the 200 V range?

6–5 THE WHEATSTONE BRIDGE

The Wheatstone bridge circuit can be used to precisely measure resistance. However, the bridge is most commonly used in an automated circuit in conjunction with transducers to measure physical quantities such as strain, temperature, and pressure. Transducers are devices that sense a change in a physical parameter and convert that change into an electrical quantity such as a change in resistance. For example, a strain gage exhibits a change in resistance when it is exposed to mechanical factors such as force, pressure, or displacement. A thermistor exhibits a change in its resistance when it is exposed to a change in temperature. The Wheatstone bridge can be operated in a balanced or an unbalanced condition. The condition of operation depends on the type of application.

After completing this section, you should be able to

- **Analyze and apply a Wheatstone bridge**
  - **Determine when a bridge is balanced**
  - **Determine an unknown resistance with a balanced bridge**
  - **Determine when a bridge is unbalanced**
  - **Discuss measurements using an unbalanced bridge**

A **Wheatstone bridge** circuit is shown in its most common diamond-shaped configuration in Figure 6–35(a). It consists of four resistors and a dc voltage source connected across the top and bottom points of the diamond. The output voltage is taken across the left and right points of the diamond between *A* and *B*. In part (b), the circuit is drawn in a slightly different way to more clearly show its series-parallel configuration.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

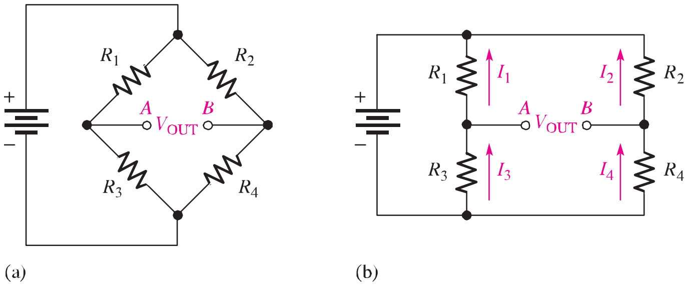

FIGURE 6–35 Wheatstone bridge. Notice that the bridge forms two back-to-back voltage dividers.

The Balanced Wheatstone Bridge

The Wheatstone bridge in Figure 6–35 is in the **balanced bridge** condition when the output voltage (V_{OUT}) between terminals A and B is equal to zero.

$$V_{OUT} = 0\text{ V}$$

When the bridge is balanced, the voltages across R_1 and R_2 are equal ($V_1 = V_2$) and the voltages across R_3 and R_4 are equal ($V_3 = V_4$). Therefore, the voltage ratios can be written as

$$\frac{V_1}{V_3} = \frac{V_2}{V_4}$$

Substituting IR for V by Ohm's law gives

$$\frac{I_1R_1}{I_3R_3} = \frac{I_2R_2}{I_4R_4}$$

Since $I_1 = I_3$ and $I_2 = I_4$, all the current terms cancel, leaving the resistor ratios.

$$\frac{R_1}{R_3} = \frac{R_2}{R_4}$$

Solving for R_1 results in the following formula:

$$R_1 = R_3\left(\frac{R_2}{R_4}\right)$$

This formula allows you to find the value of resistor R_1 in terms of the other resistor values when the bridge is balanced. You can also find the value of any other resistor in a similar way.

USING THE BALANCED WHEATSTONE BRIDGE TO FIND AN UNKNOWN RESISTANCE

Assume that R_1 in Figure 6–35 has an unknown value, which we call R_X. Resistors R_2 and R_4 have fixed values so that their ratio, R_2/R_4, also has a fixed value. Since R_X can be any value, R_3 must be adjusted to make $R_1/R_3 = R_2/R_4$ in order to create a balanced condition. Therefore, R_3 is a variable resistor, which we will call R_V. When R_X is placed in the bridge, R_V is adjusted until the bridge is balanced as indicated by a zero output voltage. Then, the unknown resistance is found as

$$\boldsymbol{R_X = R_V\left(\frac{R_2}{R_4}\right)} \qquad (6\text{–}2)$$

The ratio R_2/R_4 is the scale factor.

An older type of measuring instrument called a *galvanometer* can be connected between the output terminals A and B to detect a balanced condition. The galvanometer is essentially a very sensitive ammeter that senses current in either direction. It differs from a regular ammeter in that the midscale point is zero. Most Wheatstone bridges are now automated and use an amplifier connected across the bridge output to indicate a balanced condition when its output is 0 V. Also, high precision micro-adjustable resistors can be used in

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

demanding applications. The micro-adjustable resistors enable fine adjustment of the bridge resistors during manufacture for applications such as medical sensors, scales, and precision measurements.

From Equation 6–2, the value of R_V at balance multiplied by the scale factor R_2/R_4 is the actual resistance value of R_X. If $R_2/R_4 = 1$, then $R_X = R_V$, if $R_2/R_4 = 0.5$, then $R_X = 0.5R_V$, and so on. In a practical bridge circuit, the position of the R_V adjustment can be calibrated to indicate the actual value of R_X on a scale or with some other method of display.

EXAMPLE 6–13

Determine the value of R_X in the balanced bridge shown in Figure 6–36. The bridge is balanced ($V_{OUT} = 0$ V) when R_V is set at 1200 Ω.

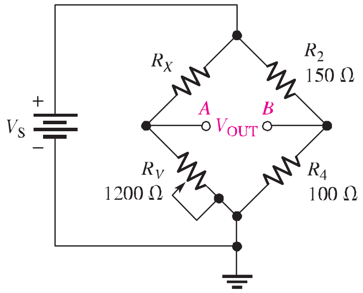

FIGURE 6–36

SOLUTION

The scale factor is

$$\frac{R_2}{R_4} = \frac{150\ \Omega}{100\ \Omega} = 1.5$$

The unknown resistance is

$$R_X = R_V\left(\frac{R_2}{R_4}\right) = (1200\ \Omega)(1.5) = \mathbf{1800\ \Omega}$$

RELATED PROBLEM

If R_V must be adjusted to 2.2 kΩ to balance the bridge in Figure 6–36, what is R_X?

The Unbalanced Wheatstone Bridge

An **unbalanced bridge** condition occurs when V_{OUT} is not equal to zero. The unbalanced bridge is used to measure several types of physical quantities such as mechanical strain, temperature, or pressure. This can be done by connecting a transducer in one leg of the bridge, as shown in Figure 6–37. The resistance of the transducer changes proportionally to the changes in the parameter that it is measuring. If the bridge is balanced at a known point, then the amount of deviation from the balanced condition, as indicated by the output voltage, indicates the amount of change in the parameter being measured. Therefore, the value of the parameter being measured can be determined by the amount that the bridge is unbalanced.

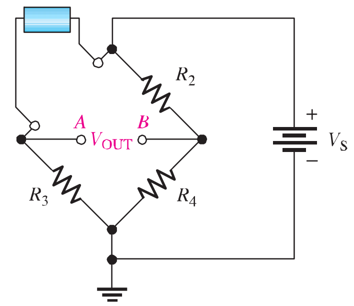

FIGURE 6–37 **A bridge circuit for measuring a physical parameter using a transducer.**

A BRIDGE CIRCUIT FOR MEASURING TEMPERATURE If temperature is to be measured, the transducer can be a thermistor, which is a temperature-sensitive resistor. The thermistor resistance changes in a predictable way as the

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

temperature changes. A change in temperature causes a change in thermistor resistance, which causes a corresponding change in the output voltage of the bridge as it becomes unbalanced. The output voltage is proportional to the temperature; therefore, either a voltmeter connected across the output can be calibrated to show the temperature or the output voltage can be amplified and converted to digital form to drive a readout display of the temperature.

A bridge circuit used to measure temperature is designed so that it is balanced at a reference temperature and becomes unbalanced at a measured temperature. For example, let's say the bridge is to be balanced at 25°C. A thermistor will have a known value of resistance at 25°C.

EXAMPLE 6–14

Determine the output voltage of the temperature-measuring bridge circuit in Figure 6–38 if the thermistor is exposed to a temperature of 50°C and its resistance at 25°C is 1.0 kΩ. Assume the resistance of the thermistor decreases to 900 Ω at 50°C.

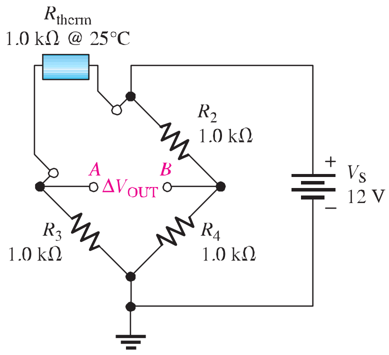

FIGURE 6–38

SOLUTION

Apply the voltage-divider formula to the left side of the bridge at 50°C.

$$V_A = \left(\frac{R_3}{R_3 + R_{\text{therm}}}\right)V_S = \left(\frac{1\text{ k}\Omega}{1\text{ k}\Omega + 900\ \Omega}\right)12\text{ V} = 6.32\text{ V}$$

Apply the voltage-divider formula to the right side of the bridge.

$$V_B = \left(\frac{R_4}{R_2 + R_4}\right)V_S = \left(\frac{1\text{ k}\Omega}{2\text{ k}\Omega}\right)12\text{ V} = 6.00\text{ V}$$

The output voltage at 50°C is the difference between V_A and V_B.

$$V_{\text{OUT}} = V_A - V_B = 6.32\text{ V} - 6.00\text{ V} = \mathbf{0.32\text{ V}}$$

Node *A* is positive with respect to node *B*.

RELATED PROBLEM

If the temperature is increased to 60°C, causing the thermistor resistance in Figure 6–38 to decrease to 850 Ω, what is V_{OUT}?

STRAIN GAGE APPLICATION OF THE WHEATSTONE BRIDGE A Wheatstone bridge with a strain gage can be used to measure forces. A strain gage is a device that exhibits a change in resistance when it is compressed or stretched by the application of an external force. As the resistance of the strain gage changes, the previously

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

balanced bridge becomes unbalanced. This unbalance causes the output voltage to change from zero, and this change can be measured to determine the amount of strain. In strain gages, the resistance change is extremely small. This tiny change unbalances a Wheatstone bridge and can be detected because of its high sensitivity. For example, Wheatstone bridges with strain gages are commonly used in weight scales.

Some resistive transducers have extremely small resistance changes, and these changes are difficult to measure accurately with a direct measurement. In particular, strain gages are one of the most useful resistive transducers that convert the stretching or compression of a fine wire into a change in resistance. When strain causes the wire in the gage to stretch, the resistance increases a small amount and when it compresses, the resistance of the wire decreases.

Strain gages are used in many types of scales, from those that are used for weighing small parts to those for weighing huge trucks. Typically, the gages are mounted on a special block of aluminum that deforms when a weight is on the scale. The strain gages are extremely delicate and must be mounted properly, so the entire assembly is generally prepared as a single unit called a *load cell.* A **load cell** is a transducer that uses strain gages to convert mechanical force into an electrical signal. A wide variety of load cells with different shapes and sizes are available from manufacturers depending on the application. A typical S-shaped load cell for a weighing application that has four strain gages is illustrated in Figure 6–39(a). The gages are mounted so that two of the gages stretch (tension) when a load is placed on the scale and two of the gages compress.

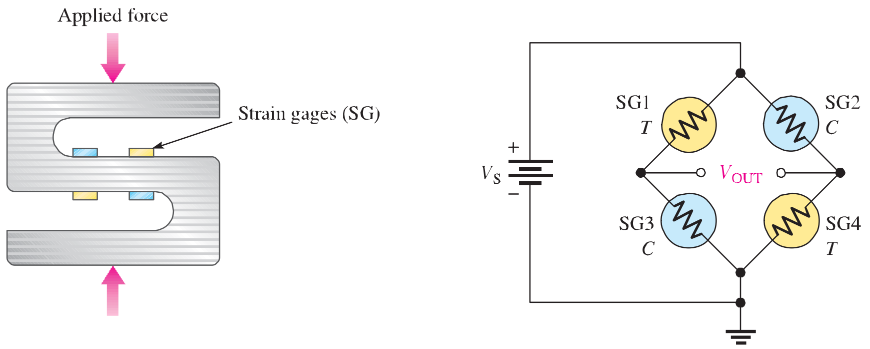

(a) A typical load cell with four active strain gages (two compression and two tension)

(b) Wheatstone bridge

FIGURE 6–39 **Example of a load cell.**

The 4-20 mA Current Loop

Many systems use transducers including strain gages and load cells that have very small signal levels, which are subject to corruption due to interference. Signal conditioning is the process of converting the signal to a usable level that is not likely to pick up unwanted noise. Typically, the signal is converted from an analog signal to a digital signal close to the source.

An older, but widely used, standard for sending information in an industrial environment is the 4-20 mA current loop shown in the block diagram. It is particularly useful when the information needs to be transmitted a long distance (1000 feet or more). The sensor's output is converted to a signal that uses 4 mA to signal the lowest level and 20 mA to signal the highest level. Because the signaling method is current rather than voltage, the line resistance is not a factor, as long as the transmitter can overcome the voltage drop in the line. The major advantage of the method is that it is virtually immune to noise pickup.

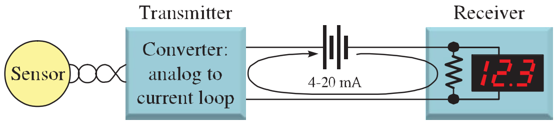

SYSTEM NOTE

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Load cells are usually connected to a Wheatstone bridge as shown in Figure 6–39(b) with strain gages (SG) in tension (*T*) and compression (*C*) in opposite diagonal legs as shown. The output of the bridge is normally digitized and converted to a reading for a display or sent to a computer for processing. The major advantage of the Wheatstone bridge circuit is that it is capable of accurately measuring very small differences in resistance. The use of four active transducers increases the sensitivity of the measurement and makes the bridge the ideal circuit for instrumentation. The Wheatstone bridge circuit has the added benefit of compensating for temperature variations and wire resistance of connecting wires that would otherwise contribute to inaccuracies.

In addition to scales, strain gages are used with Wheatstone bridges in other types of measurements including pressure measurements, displacement and acceleration measurements to name a few. In pressure measurements, the strain gages are bonded to a flexible diaphragm that stretches when pressure is applied to the transducer. The amount of flexing is related to the pressure, which again converts to a very small resistance change.

SYSTEM EXAMPLE 6–2

A LIQUID LEVEL-SENSING SYSTEM

Load cells were described in the text as force-sensing transducers. Weight is the force gravity exerts on an object, so load cells are widely used in scales. An application that converts the force to another quantity is a liquid-level detection system for a large tank. Cylindrical tanks will usually use three load cells; other shapes may have four cells. A tank containing hazardous materials or subject to wind loading might use more than four load cells.

Typical methods for measuring the liquid-level in a tank include floats, capacitive sensors, and ultrasonic sensors. Certain liquids, like paint, hazardous liquids or those at a high temperature can be a problem for contacting sensors, so weighing systems are an attractive alternative in these cases. Essentially the system described in this example is a weighing system that indicates the liquid level by a basic calculation of net weight. The weight is translated by the computer into a liquid-level indication.

6238262 2013/12/10 12.27.66.5

Figure 6–40 shows a liquid-level sensor composed of three load cells arranged so that each load cell has the same load. The tank needs to be initially set level and be able to freely move up and down for this method to work properly. Because the tank needs to move freely in the vertical direction, flexible fittings are used. The load cells must not experience any horizontal motion, and the tank needs to be restrained in the horizontal direction. The tank is fully supported by the load cells, so that the sum of the weights represents the gross weight (tank plus fluid). Each load cell is instrumented with a Wheatstone bridge. A computer calculates the percentage of liquid in the tank and the result is displayed. Figure 6–41 shows a block diagram of the system.

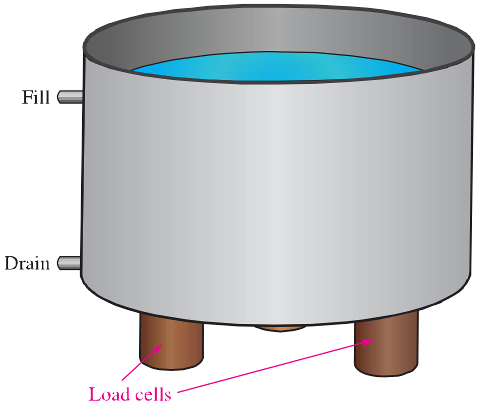

FIGURE 6–40 Tank liquid-level sensing system.

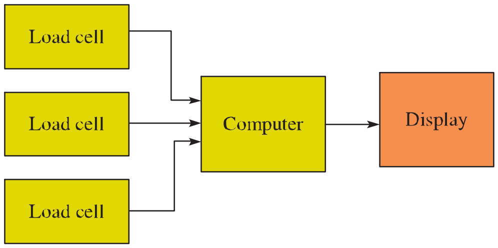

FIGURE 6–41 Block diagram of system.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 6–5 CHECKUP

1. Draw a basic Wheatstone bridge circuit.
2. Under what condition is a bridge balanced?
3. What is the unknown resistance in Figure 6–36 when $R_V = 3.3\ \text{k}\Omega$, $R_2 = 10\ \text{k}\Omega$, and $R_4 = 2.2\ \text{k}\Omega$?
4. How is a Wheatstone bridge used in the unbalanced condition?
5. What is a load cell?

6–6 THEVENIN'S THEOREM

Thevenin's theorem provides a method for simplifying a circuit to a standard equivalent form with respect to two output terminals. In many cases, this theorem can be used to simplify the analysis of series-parallel circuits. Another method for simplifying a circuit to an equivalent form is Norton's theorem, which is covered in Appendix C.

After completing this section, you should be able to

- **Apply Thevenin's theorem to simplify a circuit for analysis**
  - **Describe the form of a Thevenin equivalent circuit**
  - **Obtain the Thevenin equivalent voltage source**
  - **Obtain the Thevenin equivalent resistance**
  - **Explain terminal equivalency in the context of Thevenin's theorem**
  - **Thevenize a portion of a circuit**
  - **Thevenize a Wheatstone bridge**

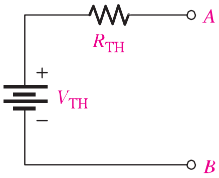

FIGURE 6–42 **The general form of a Thevenin equivalent circuit is a voltage source in series with a resistance.**

The Thevenin equivalent form of any two-terminal resistive circuit consists of an equivalent voltage source (V_{TH}) and an equivalent resistance (R_{TH}), arranged as shown in Figure 6–42. The values of the equivalent voltage and resistance depend on the values in the original circuit. Any two-terminal resistive circuit can be simplified to a Thevenin equivalent regardless of its complexity.

The equivalent voltage, V_{TH}, is one part of the complete Thevenin equivalent circuit. The other part is R_{TH}.

> **The Thevenin equivalent voltage (V_{TH}) is the open circuit (no-load) voltage between two specified output terminals in a circuit.**

Any component connected between these two terminals effectively "sees" V_{TH} in series with R_{TH}. As defined by **Thevenin's theorem**,

> **The Thevenin equivalent resistance (R_{TH}) is the total resistance appearing between two specified output terminals in a circuit with all sources replaced by their internal resistances (which for an ideal voltage source is zero).**

Although a Thevenin equivalent circuit is not of the same form as the original circuit, it acts the same in terms of the output voltage and current. Consider the following demonstration as illustrated in Figure 6–43. A resistive circuit of any complexity is placed in a box with only the output terminals exposed. The Thevenin equivalent of that circuit is placed in an identical box with, again, only the output terminals exposed. Identical load resistors are connected across the output terminals of each box. Next, a voltmeter and an ammeter are connected to measure the voltage and current for each load as shown in the figure. The measured values will be identical (neglecting tolerance variations), and you will not be able to determine which box contains the original circuit and which contains the Thevenin equivalent of the original circuit. That is, in terms of what you can observe by electrical measurement, both circuits appear to be the same. This condition is sometimes

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

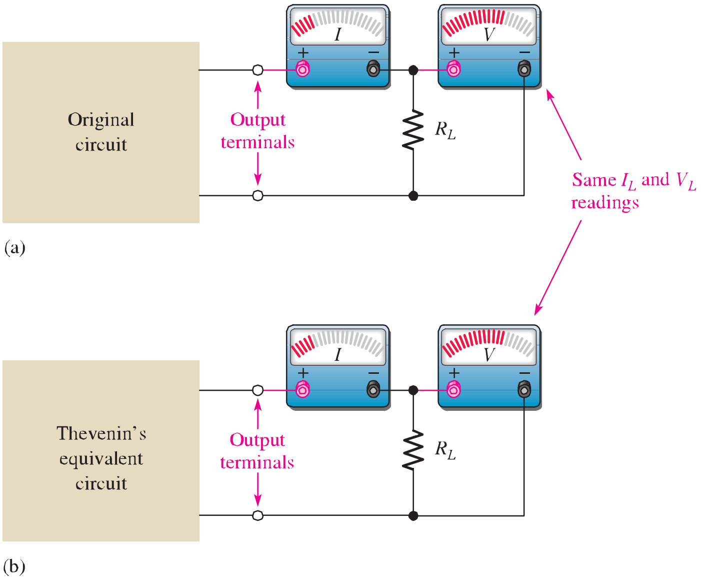
Original circuit
Output terminals
I
V
+
−
R_L
Same I_L and V_L readings
(a)
Thevenin's equivalent circuit
Output terminals
I
V
+
−
R_L
(b)

FIGURE 6–43 **Which box contains the original circuit and which contains the Thevenin equivalent circuit? You cannot tell by observing the meters because the circuits have terminal equivalency.**

known as **terminal equivalency** because both circuits look the same from the "viewpoint" of the two output terminals.

To find the Thevenin equivalent of any circuit, you must determine the equivalent voltage, V_{TH}, and the equivalent resistance, R_{TH}. For example, the Thevenin equivalent for the circuit between output terminals A and B is developed in Figure 6–44.

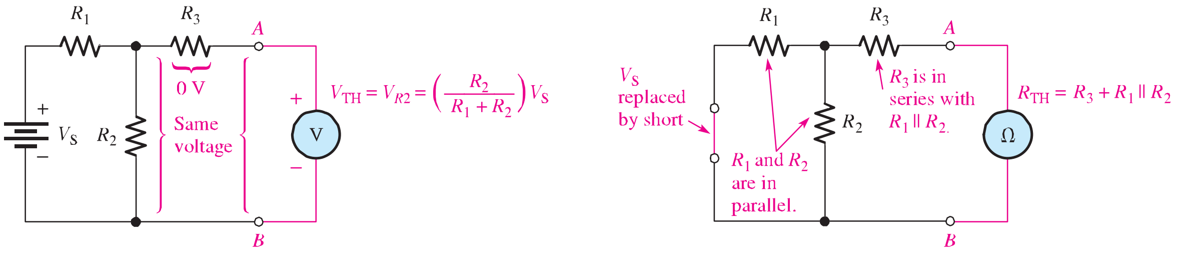
R_1
R_3
A
+
V_S
−
R_2
0 V
Same voltage
+
V
−
B
$V_{TH} = V_{R2} = \left(\frac{R_2}{R_1 + R_2}\right)V_S$
R_1
R_3
A
V_S replaced by short
R_3 is in series with $R_1 \parallel R_2$.
R_2
R_1 and R_2 are in parallel.
$R_{TH} = R_3 + R_1 \parallel R_2$
Ω
B

(a) Finding V_{TH} (b) Finding R_{TH}

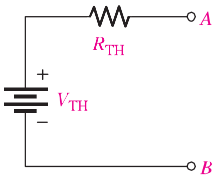

(c) Thevenin equivalent circuit

FIGURE 6–44 **Example of the simplification of a circuit by Thevenin's theorem.**

In Figure 6–44(a), the voltage across the designated terminals A and B is the Thevenin equivalent voltage. In this particular circuit, the voltage from A to B is the same as the voltage across R_2 because there is no current through R_3 and, therefore, no voltage drop across it. V_{TH} is expressed as follows for this particular example:

$$V_{TH} = \left(\frac{R_2}{R_1 + R_2}\right)V_S$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

In Figure 6–44(b), the resistance between terminals A and B with the source replaced by a short to represent its zero internal resistance is the Thevenin equivalent resistance. In this particular circuit, the resistance from A to B is R_3 in series with the parallel combination of R_1 and R_2. Therefore, R_{TH} is expressed as follows:

$$R_{TH} = R_3 + \frac{R_1R_2}{R_1 + R_2}$$

The Thevenin equivalent circuit is shown in Figure 6–44(c).

EXAMPLE 6–15

Find the Thevenin equivalent circuit between the output terminals A and B of the circuit in Figure 6–45. If there were a load resistance connected across terminals A and B, it would first have to be removed.

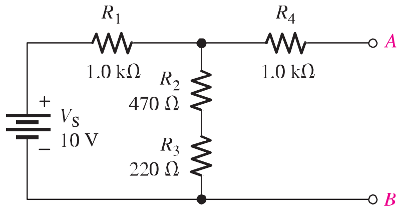

FIGURE 6–45

SOLUTION

Since there is no voltage drop across R_4, V_{AB}, equals the voltage across $R_2 + R_3$ and $V_{TH} = V_{AB}$, as shown in Figure 6–46(a). Use the voltage-divider principle to find V_{TH}.

$$V_{TH} = \left(\frac{R_2 + R_3}{R_1 + R_2 + R_3}\right)V_S = \left(\frac{690\ \Omega}{1.69\ \text{k}\Omega}\right)10\ \text{V} = \mathbf{4.08\ V}$$

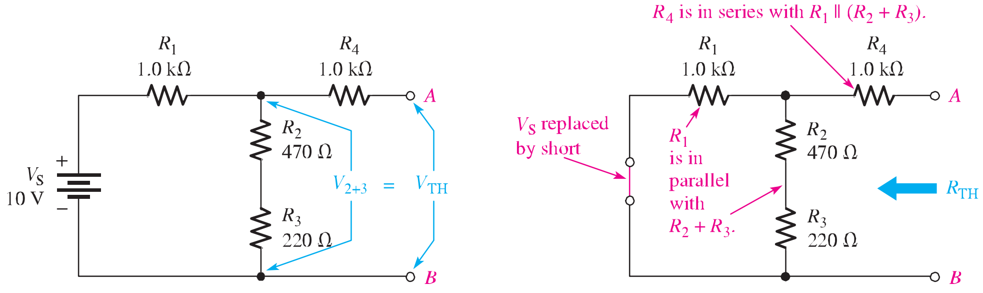

(a) The voltage from A to B is V_{TH} and equals V_{2+3}.

(b) Looking from terminals A and B, R_4 appears in series with the combination of R_1 in parallel with $(R_2 + R_3)$.

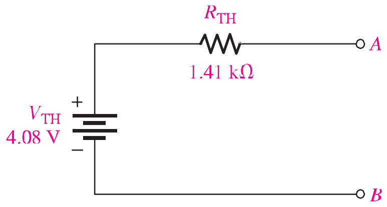

(c) Thevenin equivalent circuit

FIGURE 6–46

To find R_{TH}, first replace the source with a short to simulate a zero internal resistance. Then R_1 appears in parallel with $R_2 + R_3$, and R_4 is in series with the series-parallel combination of R_1, R_2, and R_3 as indicated in Figure 6–46(b).

0238262 2013/12/10 12.23.66.5

$$R_{TH} = R_4 + \frac{R_1(R_2 + R_3)}{R_1 + R_2 + R_3} = 1.0\ \text{k}\Omega + \frac{(1.0\ \Omega)(690\ \Omega)}{1.69\ \text{k}\Omega} = \mathbf{1.41\ k\Omega}$$

The resulting Thevenin equivalent circuit is shown in Figure 6–46(c).

RELATED PROBLEM

Determine V_{TH} and R_{TH} if a 560 Ω resistor is connected in parallel across R_2 and R_3 in Figure 6–45.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Thevenin Equivalency Depends on the Viewpoint

The Thevenin equivalent for any circuit depends on the location of the two output terminals from which the circuit is "viewed." In Figure 6–45, you viewed the circuit from between the two terminals labeled *A* and *B*. Any given circuit can have more than one Thevenin equivalent, depending on how the output terminals are designated. For example, if you view the circuit in Figure 6–47 from between terminals *A* and *C*, you obtain a completely different result than if you view it from between terminals *A* and *B* or from between terminals *B* and *C*.

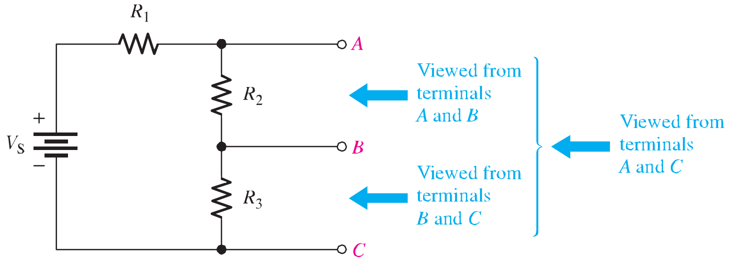
R_1
R_2
R_3
V_S
+
−
A
B
C
Viewed from terminals A and B
Viewed from terminals B and C
Viewed from terminals A and C

FIGURE 6–47 **Thevenin's equivalent depends on the output terminals from which the circuit is viewed.**

In Figure 6–48(a), when viewed from between terminals *A* and *C*, V_{TH} is the voltage across $R_2 + R_3$ and can be expressed using the voltage-divider formula as

$$V_{TH(AC)} = \left(\frac{R_2 + R_3}{R_1 + R_2 + R_3}\right)V_S$$

Also, as shown in Figure 6–48(b), the resistance between terminals *A* and *C* is $R_2 + R_3$ in parallel with R_1 (the source is replaced by a short) and can be expressed as

$$R_{TH(AC)} = R_1 \| (R_2 + R_3) = \frac{R_1(R_2 + R_3)}{R_1 + R_2 + R_3}$$

The resulting Thevenin equivalent circuit is shown in Figure 6–48(c).

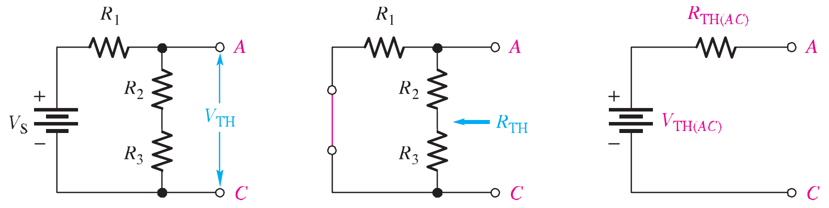

(a) $V_{TH(AC)} = \left(\frac{R_2 + R_3}{R_1 + R_2 + R_3}\right)V_S$ (b) $R_{TH(AC)} = R_1 \| (R_2 + R_3)$ (c) Thevenin equivalent

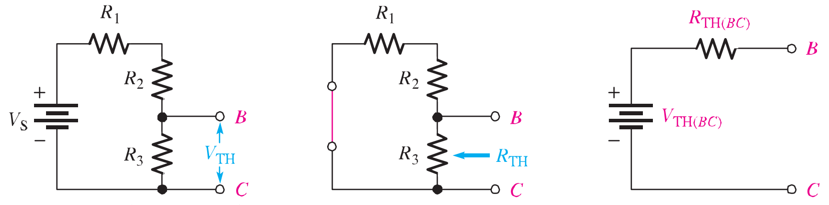

(d) $V_{TH(BC)} = \left(\frac{R_3}{R_1 + R_2 + R_3}\right)V_S$ (e) $R_{TH(BC)} = R_3 \| (R_1 + R_2)$ (f) Thevenin equivalent

FIGURE 6–48 **Example of a circuit thevenized from two different sets of terminals. Parts (a), (b), and (c) illustrate one set of terminals and parts (d), (e), and (f) illustrate another set of terminals. (The V_{TH} and R_{TH} values are different for each case.)**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

When viewed from between terminals *B* and *C* as indicated in Figure 6–48(d), $V_{TH(BC)}$ is the voltage across R_3 and can be expressed as

$$V_{TH(BC)} = \left(\frac{R_3}{R_1 + R_2 + R_3}\right)V_S$$

As shown in Figure 6–48(e), the resistance between terminals *B* and *C* is R_3 in parallel with the series combination of R_1 and R_2.

$$R_{TH(BC)} = R_3 \| (R_1 + R_2) = \frac{R_3(R_1 + R_2)}{R_1 + R_2 + R_3}$$

The resulting Thevenin equivalent is shown in Figure 6–48(f).

EXAMPLE 6–16

(a) Determine the Thevenin equivalent circuit for the circuit in Figure 6–49 viewed from terminals *A* and *C*.

(b) Determine the Thevenin equivalent circuit for the circuit in Figure 6–49 viewed from terminals *B* and *C*.

FIGURE 6–49

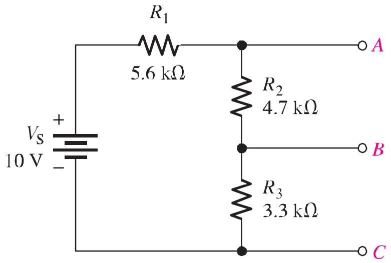

SOLUTION

(a) $$V_{TH(AC)} = \left(\frac{R_2 + R_3}{R_1 + R_2 + R_3}\right)V_S = \left(\frac{4.7\text{ k}\Omega + 3.3\text{ k}\Omega}{5.6\text{ k}\Omega + 4.7\text{ k}\Omega + 3.3\text{ k}\Omega}\right)10\text{ V} = \mathbf{5.88\text{ V}}$$

$$R_{TH(AC)} = R_1 \| (R_2 + R_3) = 5.6\text{ k}\Omega \| (4.7\text{ k}\Omega + 3.3\text{ k}\Omega) = \mathbf{3.29\text{ k}\Omega}$$

The Thevenin equivalent circuit is shown in Figure 6–50(a).

(b) $$V_{TH(BC)} = \left(\frac{R_3}{R_1 + R_2 + R_3}\right)V_S = \left(\frac{3.3\text{ k}\Omega}{5.6\text{ k}\Omega + 4.7\text{ k}\Omega + 3.3\text{ k}\Omega}\right)10\text{ V} = \mathbf{2.43\text{ V}}$$

$$R_{TH(BC)} = R_3 \| (R_1 + R_2) = 3.3\text{ k}\Omega \| (5.6\text{ k}\Omega + 4.7\text{ k}\Omega) = \mathbf{2.5\text{ k}\Omega}$$

The Thevenin equivalent circuit is shown in Figure 6–50(b).

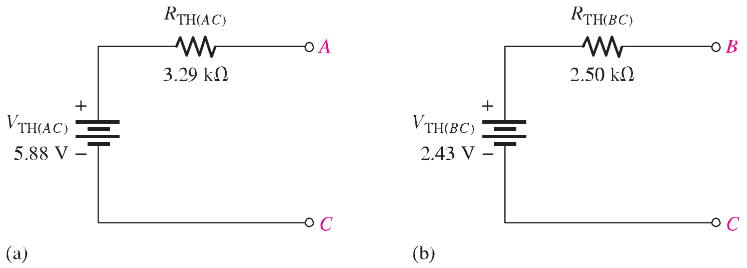

(a) (b)

FIGURE 6–50

RELATED PROBLEM

Determine the Thevenin equivalent circuit viewed from terminals *A* and *B* in Figure 6–49.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Thevenizing a Bridge Circuit

The usefulness of Thevenin's theorem can be illustrated when it is applied to a Wheatstone bridge circuit. For example, consider the case when a load resistor is connected to the output terminals of a Wheatstone bridge, as shown in Figure 6–51. The bridge circuit is difficult to analyze because it is not a straightforward series-parallel arrangement when a load resistance is connected between the output terminals *A* and *B*. There are no resistors that are in series or in parallel with another resistor.

Using Thevenin's theorem, you can simplify the bridge circuit to an equivalent circuit viewed from the load resistor as shown step-by-step in Figure 6–52. Study carefully the steps in this figure. Once the equivalent circuit for the bridge is found, the voltage and current for any value of load resistor can easily be determined by Ohm's law.

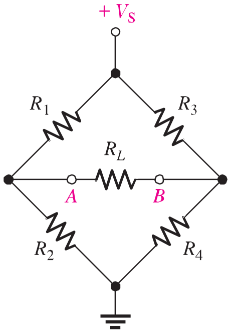

FIGURE 6–51 **A Wheatstone bridge with a load resistor connected between the output terminals is not a straightforward series-parallel circuit.**

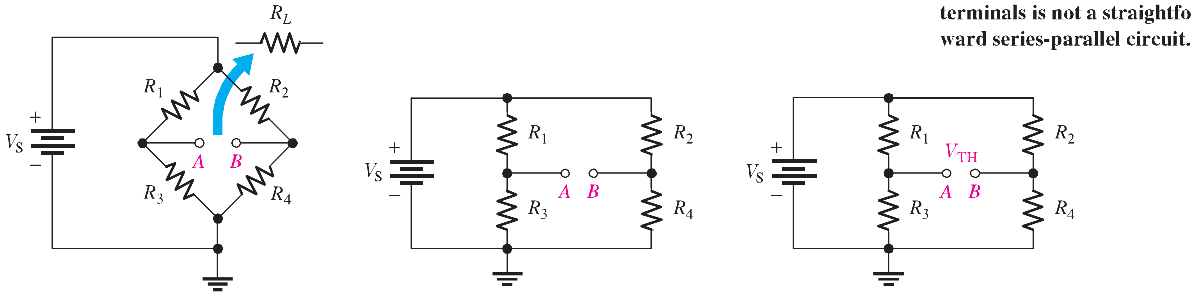

(a) Remove R_L to create an open circuit between the output terminals *A* and *B*.

(b) Redraw (if you wish).

(c) Find V_{TH}:

$$V_{TH} = V_A - V_B = \left(\frac{R_3}{R_1 + R_3}\right)V_S - \left(\frac{R_4}{R_2 + R_4}\right)V_S$$

6238262 2013/12/10 12.27.66.5

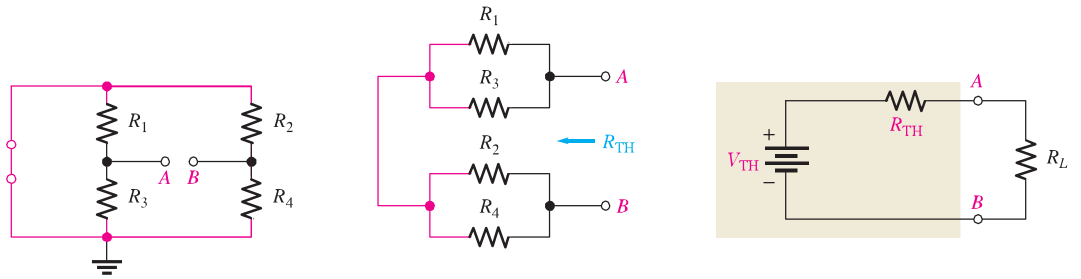

(d) Replace V_S with a short to represent its zero internal resistance. *Note*: The red lines represent the same electrical point as the red lines in Part (e).

(e) Redraw (if you wish) and find R_{TH}:

$$R_{TH} = R_1 \| R_3 + R_2 \| R_4$$

(f) Thevenin's equivalent circuit (beige block) with R_L reconnected

FIGURE 6–52 **Simplifying a Wheatstone bridge with Thevenin's theorem.**

EXAMPLE 6–17

Determine the voltage and current for the load resistor, R_L, in the bridge circuit of Figure 6–53.

FIGURE 6–53

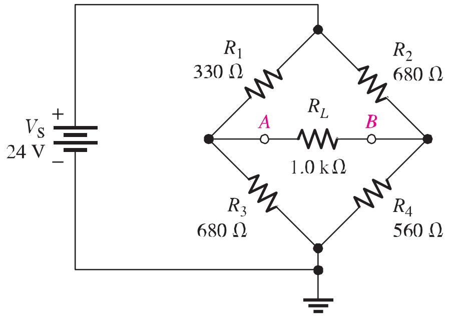

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

Step 1: Remove R_L to create an open circuit between A and B.

Step 2: To thevenize the bridge as viewed from between terminals A and B, as was shown in Figure 6–52, first determine V_{TH}.

$$V_{\mathrm{TH}} = V_A - V_B = \left(\frac{R_3}{R_1 + R_3}\right)V_{\mathrm{S}} - \left(\frac{R_4}{R_2 + R_4}\right)V_{\mathrm{S}}$$

$$= \left(\frac{680\ \Omega}{1010\ \Omega}\right)24\ \mathrm{V} - \left(\frac{560\ \Omega}{1240\ \Omega}\right)24\ \mathrm{V} = 16.16\ \mathrm{V} - 10.84\ \mathrm{V} = 5.32\ \mathrm{V}$$

Step 3: Determine R_{TH}.

$$R_{\mathrm{TH}} = \frac{R_1R_3}{R_1 + R_3} + \frac{R_2R_4}{R_2 + R_4}$$

$$= \frac{(330\ \Omega)(680\ \Omega)}{1010\ \Omega} + \frac{(680\ \Omega)(560\ \Omega)}{1240\ \Omega} = 222\ \Omega + 307\ \Omega = 529\ \Omega$$

Step 4: Place V_{TH} and R_{TH} in series to form the Thevenin equivalent circuit.

Step 5: Connect the load resistor from terminals A to B of the equivalent circuit, and determine the load voltage and current as illustrated in Figure 6–54.

$$V_L = \left(\frac{R_L}{R_L + R_{\mathrm{TH}}}\right)V_{\mathrm{TH}} = \left(\frac{1.0\ \mathrm{k}\Omega}{1.529\ \mathrm{k}\Omega}\right)5.32\ \mathrm{V} = \mathbf{3.48\ V}$$

$$I_L = \frac{V_L}{R_L} = \frac{3.48\ \mathrm{V}}{1.0\ \mathrm{k}\Omega} = \mathbf{3.48\ mA}$$

FIGURE 6–54

Thevenin's equivalent for the Wheatstone bridge

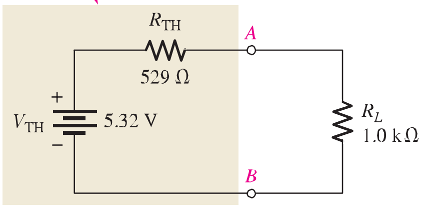

RELATED PROBLEM

Calculate I_L for $R_1 = 2.2\ \mathrm{k}\Omega$, $R_2 = 3.9\ \mathrm{k}\Omega$, $R_3 = 3.3\ \mathrm{k}\Omega$, and $R_4 = 2.7\ \mathrm{k}\Omega$ in Figure 6–53.

MULTISIM

Open Multisim file E06-17. Determine the voltage and current for R_L using a multimeter. Change the resistor values to those specified in the related problem and measure the voltage and current for R_L.

Summary of Thevenin's Theorem

Remember, the Thevenin equivalent circuit for any resistive circuit is *always* an equivalent voltage source in series with an equivalent resistance regardless of the original circuit that it replaces. The significance of Thevenin's theorem is that the equivalent circuit can replace the original circuit as far as any external load is concerned. Any load resistor connected between the terminals of a Thevenin equivalent circuit will have the same current through it and the same voltage across it as if it were connected to the terminals of the original circuit.

A summary of steps for applying Thevenin's theorem is as follows:

Step 1: Open the two terminals (remove any load) between which you want to find the Thevenin equivalent circuit.

Step 2: Determine the voltage (V_{TH}) across the two open terminals.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Step 3: Determine the resistance (R_{TH}) between the two terminals with all sources replaced by their internal resistance. (An ideal voltage source is replaced by a short.)

Step 4: Connect V_{TH} and R_{TH} in series to produce the complete Thevenin equivalent for the original circuit.

Step 5: Replace the load removed in Step 1 across the terminals of the Thevenin equivalent circuit. You can now calculate the load current and load voltage using only Ohm's law, and they have the same value as the load current and load voltage in the original circuit.

Two additional theorems are sometimes used in circuit analysis. One of these is Norton's theorem, which is similar to Thevenin's theorem except that it deals with current sources instead of voltage sources. The other is Millman's theorem, which deals with parallel voltage sources. See Appendix C for a coverage of Norton's theorem and Millman's theorem.

HANDS ON TIP

The Thevenin resistance of a circuit can be measured by connecting a variable resistance on the output of the circuit and adjusting the resistance until the output voltage of the circuit is one-half of its open circuit value. Now, if you remove the variable resistance and measure it, the value equals the Thevenin equivalent resistance of the circuit.

SECTION 6–6 CHECKUP

1. What are the two components of a Thevenin equivalent circuit?
2. Draw the general from of a Thevenin equivalent circuit.
3. Define V_{TH}.
4. Define R_{TH}.
5. For the original circuit in Figure 6–55, determine the Thevenin equivalent circuit as viewed from the output terminals *A* and *B*.

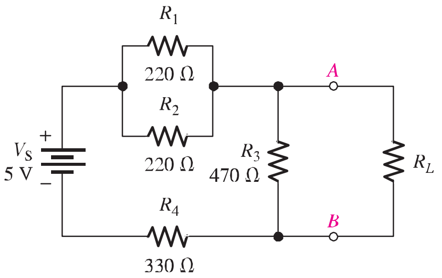

FIGURE 6–55

6–7 THE MAXIMUM POWER TRANSFER THEOREM

The maximum power transfer theorem is important when you need to know the value of the load at which the most power is delivered from the source.

After completing this section, you should be able to

- **Apply the maximum power transfer theorem**
  - **State the theorem**
  - **Determine the value of load resistance for which maximum power is transferred from a given circuit**

The **maximum power transfer** theorem is stated as follows:

> **For a given source voltage, maximum power is transferred from a source to a load when the load resistance is equal to the internal source resistance.**

The source resistance, R_S, of a circuit is the equivalent resistance as viewed from the output terminals using Thevenin's theorem. A Thevenin equivalent circuit with its output resistance and load is shown in Figure 6–56. When $R_L = R_S$, the maximum power possible is transferred from the voltage source to R_L for a given value of V_S.

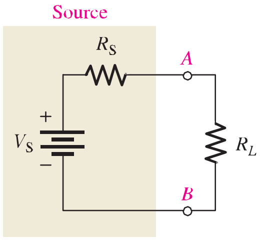

FIGURE 6–56 **Maximum power is transferred to the load when $R_L = R_S$.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

Practical applications of the maximum power transfer theorem include audio systems such as stereo, radio, and public address. In these systems the resistance of the speaker is the load. The circuit that drives the speaker is a power amplifier. The systems are typically optimized for maximum power to the speakers. Thus, the resistance of the speaker must equal the internal source resistance of the amplifier.

Example 6–18 shows that maximum power occurs when $R_L = R_S$.

EXAMPLE 6–18

The source in Figure 6–57 has an internal resistance of 75 Ω. Determine the load power for each of the following values of the variable load resistance:

(a) 0 Ω **(b)** 25 Ω **(c)** 50 Ω

(d) 75 Ω **(e)** 100 Ω **(f)** 125 Ω

Draw a graph showing the load power versus the load resistance.

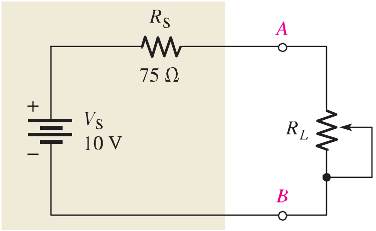

FIGURE 6–57

SOLUTION

Use Ohm's law ($I = V/R$) and the power formula ($P = I^2R$) to find the load power, P_L, for each value of load resistance.

(a) For $R_L = 0\ \Omega$,

$$I = \frac{V_S}{R_S + R_L} = \frac{10\text{V}}{75\ \Omega + 0\ \Omega} = 133\text{ mA}$$

$$P_L = I^2R_L = (133\text{ mA})^2(0\ \Omega) = \mathbf{0\ mW}$$

(b) For $R_L = 25\ \Omega$,

$$I = \frac{V_S}{R_S + R_L} = \frac{10\text{ V}}{75\ \Omega + 25\ \Omega} = 100\text{ mA}$$

$$P_L = I^2R_L = (100\text{ mA})^2(25\ \Omega) = \mathbf{250\ mW}$$

(c) For $R_L = 50\ \Omega$,

$$I = \frac{V_S}{R_S + R_L} = \frac{10\text{ V}}{125\ \Omega} = 80\text{ mA}$$

$$P_L = I^2R_L = (80\text{ mA})^2(50\ \Omega) = \mathbf{320\ mW}$$

(d) For $R_L = 75\ \Omega$,

$$I = \frac{V_S}{R_S + R_L} = \frac{10\text{ V}}{150\ \Omega} = 66.7\text{ mA}$$

$$P_L = I^2R_L = (66.7\text{ mA})^2(75\ \Omega) = \mathbf{334\ mW}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

(e) For $R_L = 100\ \Omega$,

$$I = \frac{V_S}{R_S + R_L} = \frac{10\ \text{V}}{175\ \Omega} = 57.1\ \text{mA}$$

$$P_L = I^2R_L = (57.1\ \text{mA})^2(100\ \Omega) = \mathbf{326\ mW}$$

(f) For $R_L = 125\ \Omega$,

$$I = \frac{V_S}{R_S + R_L} = \frac{10\ \text{V}}{200\ \Omega} = 50\ \text{mA}$$

$$P_L = I^2R_L = (50\ \text{mA})^2(125\ \Omega) = \mathbf{313\ mW}$$

Notice that the load power is greatest when $R_L = R_S = 75\ \Omega$, which is the same as the internal source resistance. When the load resistance is less than or greater than this value, the power drops off, as the curve in Figure 6–58 graphically illustrates.

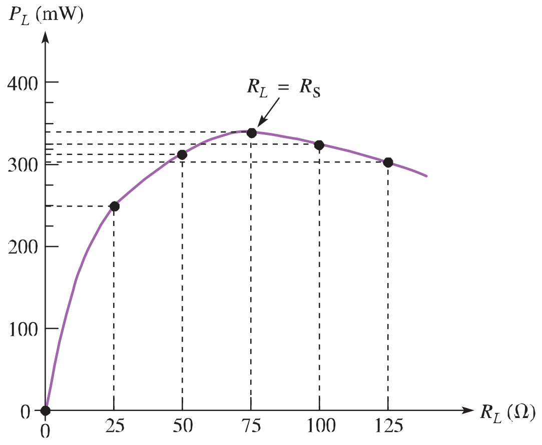

FIGURE 6–58 **Curve showing that the load power is maximum when $R_L = R_S$.**

RELATED PROBLEM

If the source resistance in Figure 6–57 is 600 Ω, what is the maximum power that can be delivered to a load?

SECTION 6–7 CHECKUP

1. State the maximum power transfer theorem.
2. When is maximum power delivered from a source to a load?
3. A given circuit has an internal source resistance of 50 Ω. What will be the value of the load to which the maximum power is delivered?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6–8 THE SUPERPOSITION THEOREM

Some circuits require more than one voltage or current source. For example, most amplifiers operate with two voltage sources: an ac and a dc source. Additionally, some amplifiers require both a positive and a negative dc voltage source for proper operation. When multiple sources are used in a circuit, the superposition theorem provides a method for analysis.

After completing this section, you should be able to

- **Apply the superposition theorem to circuit analysis**
  - **State the superposition theorem**
  - **List the steps in applying the theorem**

The **superposition** theorem is a way to determine currents in a circuit with multiple sources by leaving one source at a time and replacing the other sources by their internal resistances. Recall that the ideal voltage source has a zero internal resistance and the ideal current source has infinite internal resistance. All sources will be treated as ideal in order to simplify the coverage.

The Superposition Theorem in Physics

Anja Kaiser/Fotolia.com

The superposition theorem is not restricted to electrical and electronic systems. It can be applied to any linear system with two or more sources (stimuli). In physics, it is applied to various types of waves (including sound and water waves), mechanical structures (beams), field theory, quantum mechanics, and even overlapping magnetic or electric fields.

For example, when water, sound, or light waves are combined from independent sources, the result can be predicted within certain limits by the superposition theorem. Where two crests overlap, the wave amplitude increases; when a crest and a trough overlap, they tend to cancel. Another familiar example of superposition is the superimposing of an airplane's forward motion with any wind motion. The vector sum of the two is the net motion of the plane.

SYSTEM NOTE

A general statement of the superposition theorem is as follows:

> **The current in any given branch of a multiple-source linear circuit can be found by determining the currents in that particular branch produced by each source acting alone, with all other sources replaced by their internal resistances. The total current in the branch is the algebraic sum of the individual source currents in that branch.**

The steps in applying the superposition theorem are as follows:

Step 1: Leave one voltage (or current) source at a time in the circuit and replace each of the other voltage (or current) sources with its internal resistance. For ideal sources, a short represents zero internal resistance and an open represents infinite internal resistance.

Step 2: Determine the particular current (or voltage) that you want to find just as if there were only one source in the circuit. This is a component of the total current or voltage that you are looking for.

Step 3: Take the next source in the circuit and repeat Steps 1 and 2 for each source.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Step 4: To find the actual current in a given branch (with all sources active), algebraically add the results for all sources. Once you find this current, you can determine the voltage using Ohm's law.

The approach to superposition is illustrated in Figure 6–59 for a series-parallel circuit with two ideal voltage sources. Study the steps in this figure.

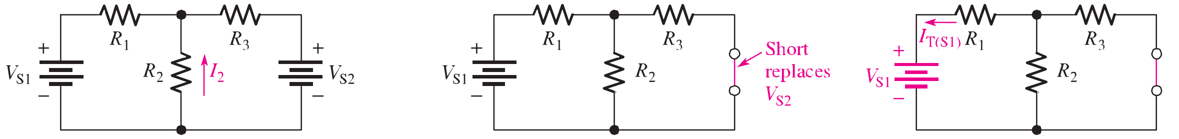

(a) Problem: Find I_2.

(b) Replace V_{S2} with zero resistance (short).

(c) Find R_T and I_T looking from V_{S1}:
$R_{T(S1)} = R_1 + R_2 \parallel R_3$
$I_{T(S1)} = V_{S1}/R_{T(S1)}$

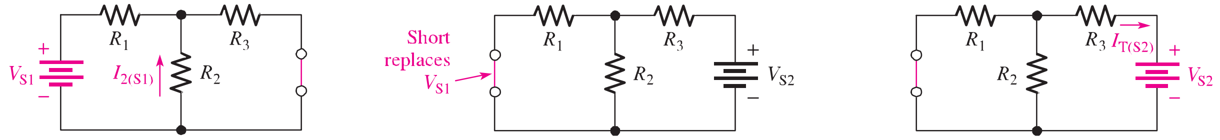

(d) Find I_2 due to V_{S1}:

$$I_{2(S1)} = \left(\frac{R_3}{R_2 + R_3}\right) I_{T(S1)}$$

(e) Replace V_{S1} with zero resistance (short).

(f) Find R_T and I_T looking from V_{S2}:
$R_{T(S2)} = R_3 + R_1 \parallel R_2$
$I_{T(S2)} = V_{S2}/R_{T(S2)}$

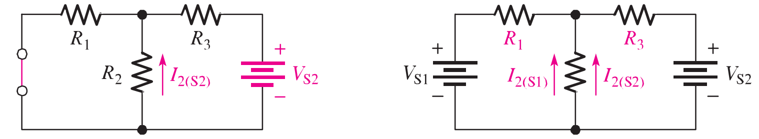

(g) Find I_2 due to V_{S2}:

$$I_{2(S2)} = \left(\frac{R_1}{R_1 + R_2}\right) I_{T(S2)}$$

(h) Restore the original sources. Algebraically add $I_{2(S1)}$ and $I_{2(S2)}$ to get the actual I_2 (they are in same direction):
$I_2 = I_{2(S1)} + I_{2(S2)}$

FIGURE 6–59 **Illustration of the superposition theorem.**

6238262 2013/12/10 12.27.66.5

EXAMPLE 6–19

Use the superposition theorem to find the current through R_2 and the voltage across it in Figure 6–60.

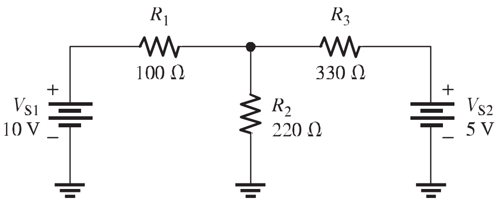

FIGURE 6–60

SOLUTION

Step 1: Replace V_{S2} with a short to represent its zero internal resistance and find the current through R_2 due to voltage source V_{S1}, as shown in

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Figure 6–61. To find I_2, use the current-divider formula. Looking from V_{S1},

$$R_{T(S1)} = R_1 + R_2 \| R_3 = 100\ \Omega + 220\ \Omega \| 330\ \Omega = 232\ \Omega$$

$$I_{T(S1)} = \frac{V_{S1}}{R_{T(S1)}} = \frac{10\text{ V}}{232\ \Omega} = 43.1\text{ mA}$$

The component of the total current through R_2 due to V_{S1} is

$$I_{2(S1)} = \left(\frac{R_3}{R_2 + R_3}\right)I_{T(S1)} = \left(\frac{330\ \Omega}{220\ \Omega + 330\ \Omega}\right)43.1\text{ mA} = 25.9\text{ mA}$$

Note that this current is upward through R_2.

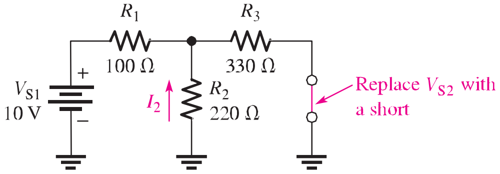

FIGURE 6–61

Step 2: Find the current through R_2 due to voltage source V_{S2} by replacing V_{S1} with a short, as shown in Figure 6–62. Looking from V_{S2},

$$R_{T(S2)} = R_3 + R_1 \| R_2 = 330\ \Omega + 100\ \Omega \| 220\ \Omega = 399\ \Omega$$

$$I_{T(S2)} = \frac{V_{S2}}{R_{T(S2)}} = \frac{5\text{ V}}{399\ \Omega} = 12.5\text{ mA}$$

The component of the total current through R_2 due to V_{S2} is

$$I_{2(S2)} = \left(\frac{R_1}{R_1 + R_2}\right)I_{T(S2)} = \left(\frac{100\ \Omega}{100\ \Omega + 220\ \Omega}\right)12.5\text{ mA} = 3.90\text{ mA}$$

Note that this current is upward through R_2.

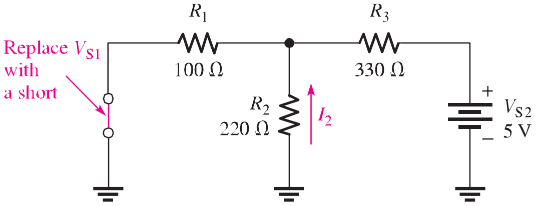

FIGURE 6–62

Step 3: Both component currents are upward through R_2, so they have the same algebraic sign. Therefore, add the values to get the total current through R_2.

$$I_{2(tot)} = I_{2(S1)} + I_{2(S2)} = 25.9\text{ mA} + 3.90\text{ mA} = \mathbf{29.8\text{ mA}}$$

The voltage across R_2 is

$$V_{R2} = I_{2(tot)}R_2 = (29.8\text{ mA})(220\ \Omega) = \mathbf{6.56\text{ V}}$$

RELATED PROBLEM

Determine the total current through R_2 if the polarity of V_{S2} in Figure 6–60 is reversed.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 6–20

Find the total current through and voltage across R_3 in Figure 6–63.

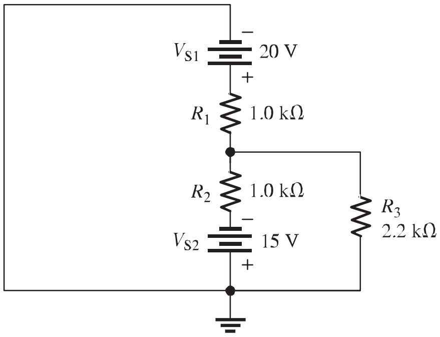

FIGURE 6–63

SOLUTION

Step 1: Find the current through R_3 due to source V_{S1} by replacing source V_{S2} with a short to represent its zero internal resistance, as shown in Figure 6–64. Looking from V_{S1},

$$R_{T(S1)} = R_1 + \frac{R_2R_3}{R_2 + R_3} = 1.0\,\text{k}\Omega + \frac{(1.0\,\text{k}\Omega)(2.2\,\text{k}\Omega)}{3.2\,\text{k}\Omega} = 1.69\,\text{k}\Omega$$

$$I_{T(S1)} = \frac{V_{S1}}{R_{T(S1)}} = \frac{20\text{ V}}{1.69\,\text{k}\Omega} = 11.8\text{ mA}$$

Now apply the current-divider formula to get the current through R_3 due to source V_{S1}.

$$I_{3(S1)} = \left(\frac{R_2}{R_2 + R_3}\right)I_{T(S1)} = \left(\frac{1.0\,\text{k}\Omega}{3.2\,\text{k}\Omega}\right)11.8\text{ mA} = 3.69\text{ mA}$$

Notice that this current is upward through R_3.

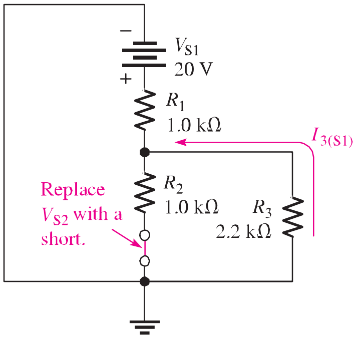

FIGURE 6–64

Step 2: Find the current through R_3 due to source V_{S2} by replacing source V_{S1} with a short, as shown in Figure 6–65. Looking from V_{S2},

$$R_{T(S2)} = R_2 + \frac{R_1R_3}{R_1 + R_3} = 1.0\,\text{k}\Omega + \frac{(1.0\,\text{k}\Omega)(2.2\,\text{k}\Omega)}{3.2\,\text{k}\Omega} = 1.69\,\text{k}\Omega$$

$$I_{T(S2)} = \frac{V_{S2}}{R_{T(S2)}} = \frac{15\text{ V}}{1.69\,\text{k}\Omega} = 8.88\text{ mA}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

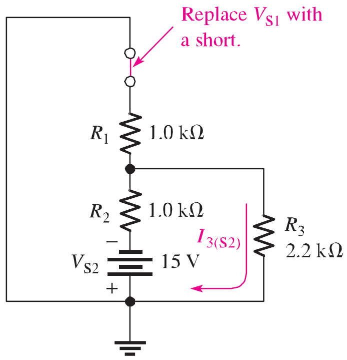
Replace V_{S1} with a short.

FIGURE 6–65

Now apply the current-divider formula to find the current through R_3 due to source V_{S2}.

$$I_{3(S2)} = \left(\frac{R_1}{R_1 + R_3}\right)I_{T(S2)} = \left(\frac{1.0\text{ k}\Omega}{3.2\text{ k}\Omega}\right)8.88\text{ mA} = 2.78\text{ mA}$$

Notice that this current is downward through R_3.

Step 3: Calculate the total current through R_3 and the voltage across it.

$$I_{3(tot)} = I_{3(S1)} - I_{3(S2)} = 3.69\text{ mA} - 2.78\text{ mA} = 0.91\text{ mA} = \mathbf{910\ \mu A}$$
$$V_{R3} = I_{3(tot)}R_3 = (910\ \mu\text{A})(2.2\text{ k}\Omega) \cong \mathbf{2\ V}$$

The current is upward through R_3.

6238262 2013/12/10 12.27.66.5

RELATED PROBLEM

Find the total current through R_3 in Figure 6–63 if V_{S1} is changed to 12 V and its polarity reversed.

Although regulated dc power supplies are close to ideal voltage sources, many ac sources are not. For example, function generators generally have 50 Ω or 600 Ω of internal resistance, which appears as a resistance in series with an ideal source. Also, batteries can look ideal when they are fresh, but as they age, the internal resistance increases. When applying the superposition theorem, it is important to recognize when a source is not ideal and replace it with its equivalent internal resistance.

Current sources are not as common as voltage sources and are also not always ideal. If a current source is not ideal, as in the case of transistors, it should be replaced by its equivalent internal resistance when the superposition theorem is applied.

SECTION 6–8 CHECKUP

1. State the superposition theorem.
2. Why is the superposition theorem useful for analysis of multiple-source linear circuits?
3. Why is an ideal voltage source shorted when the superposition theorem is applied?
4. If, as a result of applying the superposition theorem, two component currents are in opposing directions through a branch of a circuit, in which direction is the net current?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6–9 TROUBLESHOOTING

Troubleshooting is the process of identifying and locating a failure or problem in a circuit. Some troubleshooting techniques and the application of logical thought have already been discussed in relation to both series circuits and parallel circuits. A basic premise of troubleshooting is that you must know what to look for before you can successfully troubleshoot a circuit.

After completing this section, you should be able to

- **Troubleshoot series-parallel circuits**
  - **Determine the effects of an open in a circuit**
  - **Determine the effects of a short in a circuit**
  - **Locate opens and shorts**

Opens and shorts are typical problems that occur in electric circuits. As mentioned in Chapter 4, if a resistor burns out, it will normally produce an open. Bad solder connections, broken wires, and poor contacts can also be causes of open paths. Pieces of foreign material, such as solder splashes, broken insulation on wires, and so on, can lead to shorts in a circuit. A short is considered to be a zero resistance path between two points.

In addition to complete opens or shorts, partial opens or partial shorts can develop in a circuit. A partial open would be a much higher than normal resistance, but not infinitely large. A partial short would be a much lower than normal resistance, but not zero.

The following three examples illustrate troubleshooting series-parallel circuits.

EXAMPLE 6–21

From the indicated voltmeter reading in Figure 6–66, determine if there is a fault by applying the APM approach. If there is a fault, identify it as either a short or an open.

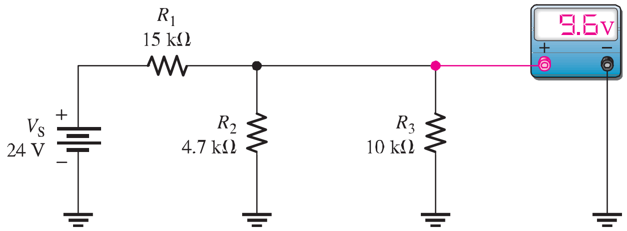

FIGURE 6–66

SOLUTION

Step 1: Analysis

Determine what the voltmeter should be indicating as follows. Since R_2 and R_3 are in parallel, their combined resistance is

$$R_{2\|3} = \frac{R_2R_3}{R_2 + R_3} = \frac{(4.7\text{ k}\Omega)(10\text{ k}\Omega)}{14.7\text{ k}\Omega} = 3.20\text{ k}\Omega$$

Determine the voltage across the parallel combination by the voltage-divider formula.

$$V_{2\|3} = \left(\frac{R_{2\|3}}{R_1 + R_{2\|3}}\right)V_S = \left(\frac{3.2\text{ k}\Omega}{18.2\text{ k}\Omega}\right)24\text{ V} = 4.22\text{ V}$$

This calculation shows that 4.22 V is the voltage reading that you should get on the meter. However, the meter reads 9.6 V across $R_{2\|3}$. This value is incorrect, and, because it is higher than it should be, either

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

R_2 or R_3 is probably open. Why? Because if either of these two resistors is open, the resistance across which the meter is connected is larger than expected. A higher resistance will drop a higher voltage in this circuit.

Step 2: Planning

Start trying to find the open resistor by assuming that R_2 is open. If it is, the voltage across R_3 is

$$V_3 = \left(\frac{R_3}{R_1 + R_3}\right)V_S = \left(\frac{10\text{ k}\Omega}{25\text{ k}\Omega}\right)24\text{ V} = 9.6\text{ V}$$

Since the measured voltage is also 9.6 V, this calculation shows that R_2 is probably open.

Step 3: Measurement

Disconnect power and remove R_2. Measure its resistance to verify it is open. If it is not, inspect the wiring, solder, or connections around R_2, looking for the open.

RELATED PROBLEM

What would be the voltmeter reading if R_3 were open in Figure 6–66? If R_1 were open?

MULTISIM

Open Multisim file E06-21. Determine if a fault exists in the circuit and, if so, isolate the fault to a single component.

EXAMPLE 6–22

Suppose that you measure 24 V with the voltmeter in Figure 6–67. Determine if there is a fault, and, if there is, identify it.

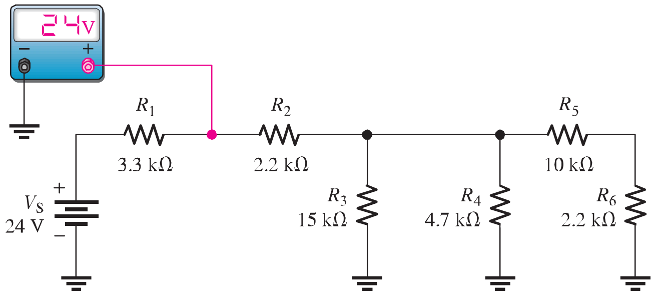

FIGURE 6–67

SOLUTION

Step 1: Analysis

There is no voltage drop across R_1 because both sides of the resistor are at +24 V. Either there is no current through R_1 from the source, which tells you that R_2 is open in the circuit, or R_1 is shorted.

Step 2: Planning

The most probable failure is an open R_2. If it is open, then there will be no current from the source. To verify this, plan to measure across R_2 with the voltmeter. If R_2 is open, the meter will indicate 24 V. The right side of R_2 will be at zero volts because there is no current through any of the other resistors to cause a voltage drop across them.

Step 3: Measurement

The measurement to verify that R_2 is open is shown in Figure 6–68.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

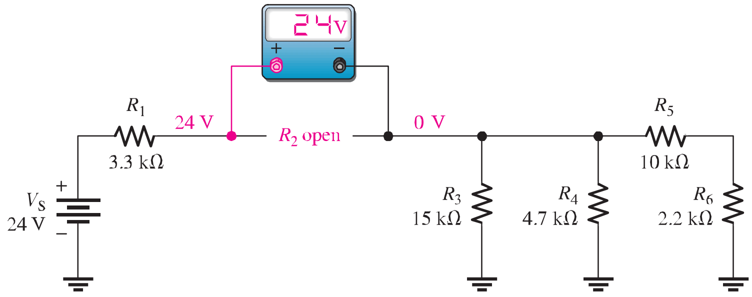

FIGURE 6–68

RELATED PROBLEM

What would be the voltage across an open R_5 in Figure 6–67 assuming no other faults?

MULTISIM

Open Multisim file E06-22. Determine if a fault exists in the circuit and, if so, isolate the fault to a single component.

EXAMPLE 6–23

The two voltmeters in Figure 6–69 indicate the voltages shown. Apply logical thought and your knowledge of circuit operation to determine if there are any opens or shorts in the circuit and, if so, where they are located.

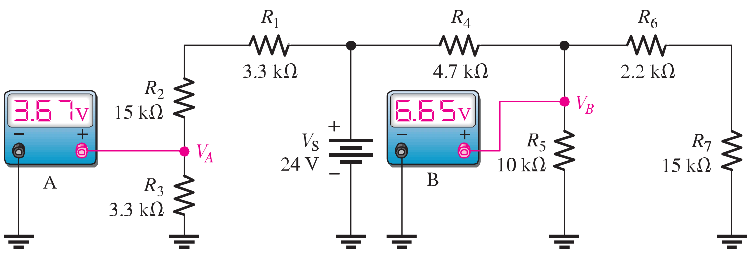

FIGURE 6–69

SOLUTION

6238262 2013/12/10 12.27.66.5

Step 1: Determine if the voltmeter readings are correct. R_1, R_2, and R_3 act as a voltage divider. Calculate the voltage (V_A) across R_3 as follows:

$$V_A = \left(\frac{R_3}{R_1 + R_2 + R_3}\right)V_S = \left(\frac{3.3\text{ k}\Omega}{21.6\text{ k}\Omega}\right)24\text{ V} = 3.67\text{ V}$$

The voltmeter A reading is correct. This indicates that R_1, R_2, and R_3 are connected and are not faulty.

Step 2: See if the voltmeter B reading is correct. $R_6 + R_7$ is in parallel with R_5. The series-parallel combination of R_5, R_6, and R_7 is in series with R_4. Calculate the resistance of the R_5, R_6, and R_7 combination as follows:

$$R_{5\|(6+7)} = \frac{R_5(R_6 + R_7)}{R_5 + R_6 + R_7} = \frac{(10\text{ k}\Omega)(17.2\text{ k}\Omega)}{27.2\text{ k}\Omega} = 6.32\text{ k}\Omega$$

$R_{5\|(6+7)}$ and R_4 form a voltage divider, and voltmeter B is measuring the voltage across $R_{5\|(6+7)}$. Is it correct? Check as follows:

$$V_B = \left(\frac{R_{5\|(6+7)}}{R_4 + R_{5\|(6+7)}}\right)V_S = \left(\frac{6.32\text{ k}\Omega}{11\text{ k}\Omega}\right)24\text{ V} = 13.8\text{ V}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Thus, the actual measured voltage (6.65 V) at this point is incorrect. Some logical thinking will help to isolate the problem.

Step 3: R_4 is not open, because if it were, the meter would read 0 V. If there were a short across it, the meter would read 24 V. Since the actual voltage is much less than it should be, $R_{5\|(6+7)}$ must be less than the calculated value of 6.32 kΩ. The most likely problem is a short across R_7. If there is a short from the top of R_7 to ground, R_6 is effectively in parallel with R_5. In this case,

$$R_5 \| R_6 = \frac{R_5 R_6}{R_5 + R_6} = \frac{(10\text{ k}\Omega)(2.2\text{ k}\Omega)}{12.2\text{ k}\Omega} = 1.80\text{ k}\Omega$$

Then V_B is

$$V_B = \left(\frac{1.80\text{ k}\Omega}{6.5\text{ k}\Omega}\right)24\text{ V} = 6.65\text{ V}$$

This value for V_B agrees with the voltmeter B reading. So there is a short across R_7. If this were an actual circuit, you would try to find the physical cause of the short.

MULTISIM

Open Multisim file E06-23. Determine if a fault exists in the circuit and, if so, isolate the fault to a single component.

RELATED PROBLEM

If the only fault in Figure 6–69 is that R_2 is shorted, what will voltmeter A read? What will voltmeter B read?

SECTION 6–9 CHECKUP

1. Name two types of common circuit faults.
2. For the following faults in Figure 6–70, determine what voltage would be measured at node A:
   (a) no faults
   (b) R_1 open
   (c) short across R_5
   (d) R_3 and R_4 open
   (e) R_2 open
3. In Figure 6–71, one of the resistors in the circuit is open. Based on the meter reading, determine which is the open resistor.

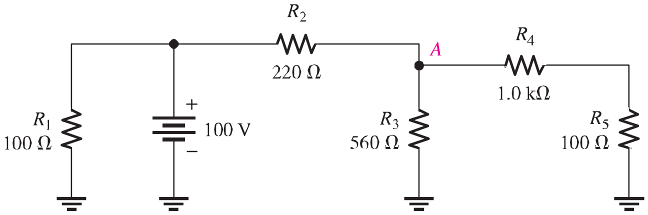

FIGURE 6–70

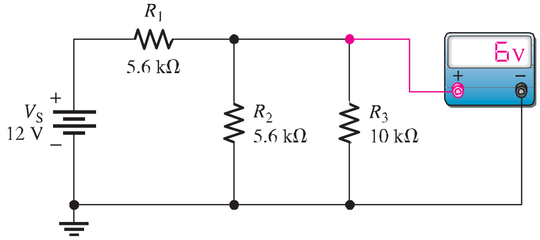

FIGURE 6–71

SUMMARY

- A series-parallel circuit is a combination of both series current paths and parallel current paths.
- To determine total resistance in a series-parallel circuit, identify the series and parallel relationships, and then apply the formulas for series resistance and parallel resistance from Chapters 4 and 5.
- To find the total current, divide the total voltage by the total resistance.
- To determine branch currents, apply the current-divider formula, Kirchhoff's current law, or Ohm's law. Consider each circuit problem individually to determine the most appropriate method.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

- To determine voltage drops across any portion of a series-parallel circuit, use the voltage-divider formula, Kirchhoff's voltage law, or Ohm's law. Consider each circuit problem individually to determine the most appropriate method.
- When a load resistor is connected across a voltage-divider output, the output voltage decreases.
- The load resistor should be large compared to the resistance across which it is connected, in order that the loading effect may be minimized. A *10-times* value is sometimes used as a rule of thumb, but the value depends on the accuracy required for the output voltage.
- To find any current or voltage in a circuit with two or more voltage sources, take the sources one at a time using the superposition theorem.
- A balanced Wheatstone bridge can be used to measure an unknown resistance.
- A bridge is balanced when the output voltage is zero. The balanced condition produces zero current through a load connected across the output terminals of the bridge.
- An unbalanced Wheatstone bridge can be used to measure physical quantities using transducers.
- Any two-terminal resistive circuit, no matter how complex, can be replaced by its Thevenin equivalent.
- The Thevenin equivalent circuit is made up of an equivalent resistance (R_{TH}) in series with an equivalent voltage source (V_{TH}).
- The maximum power transfer theorem states that maximum power is transferred from a source to a load when $R_S = R_L$.
- Opens and shorts in circuits are typical faults.
- Resistors normally open when they fail.

KEY TERMS

Key terms and other bold terms in the chapter are defined in the end-of-book glossary.

Balanced bridge A bridge circuit that is in the balanced state as indicated by zero volts across the bridge.

Bleeder current The current left after the total load current is subtracted from the total current into the circuit.

Load current The output current of a circuit supplied to a load.

Loading The effect on a circuit when an element that draws current from the circuit is connected across the output terminals.

Maximum power transfer The condition, when the load resistance equals the source resistance, under which maximum power is transferred from source to load.

Superposition A method for analyzing circuits with two or more sources by examining the effects of each source by itself and then combining the effects.

Terminal equivalency A condition that occurs when two circuits produce the same load voltage and load current where the same value of load resistance is connected to either circuit.

Thevenin's theorem A circuit theorem that provides for reducing any two-terminal resistive circuit to a single equivalent voltage source in series with an equivalent resistance.

Unbalanced bridge A bridge circuit that is in the unbalanced state as indicated by a voltage across the bridge proportional to the amount of deviation from the balanced state.

Wheatstone bridge A 4-legged type of bridge circuit with which an unknown resistance can be accurately measured using the balanced state of the bridge. Deviations in resistance can be measured using the unbalanced state.

KEY FORMULAS

(6–1)	$I_{BLEEDER} = I_T - I_{RL1} - I_{RL2}$	Bleeder current
(6–2)	$R_X = R_V\left(\frac{R_2}{R_4}\right)$	Unknown resistance in a Wheatstone bridge

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

TRUE/FALSE QUIZ

Answers are at the end of the chapter.

1. Parallel resistors are always connected between the same pair of nodes.
2. If one resistor is connected in series with a parallel combination, the series resistor will always have a larger voltage drop than the parallel resistors.
3. In a series-parallel combinational circuit, the same current will always be in parallel resistors.
4. A larger load resistor has a smaller loading effect on a circuit.
5. When measuring dc voltage, a DMM will normally have a small loading effect on a circuit.
6. When measuring dc voltage, the input resistance of a DMM is the same no matter on what scale it is used.
7. When measuring dc voltage, the input resistance of an analog multimeter is the same no matter on what scale it is used.
8. A Thevenin circuit consists of a voltage source with a parallel resistor.
9. The internal resistance of an ideal voltage source is zero.
10. To transfer maximum power to a load, the load resistor should be twice the Thevenin resistance of the source.

SELF-TEST

Answers are at the end of the chapter.

1. Which of the following statements are true concerning Figure 6–72?
   (a) R_1 and R_2 are in series with R_3, R_4, and R_5.
   (b) R_1 and R_2 are in series.
   (c) R_3, R_4, and R_5 are in parallel.
   (d) The series combination of R_1 and R_2 is in parallel with the series combination of R_3, R_4, and R_5.
   (e) answers (b) and (d)

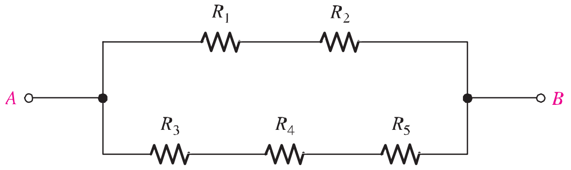

FIGURE 6–72

2. The total resistance of Figure 6–72 can be found with which of the following formulas?
   (a) $R_1 + R_2 + R_3 \| R_4 \| R_5$ **(b)** $R_1 \| R_2 + R_3 \| R_4 \| R_5$
   (c) $(R_1 + R_2) \| (R_3 + R_4 + R_5)$ **(d)** none of these answers
3. If all of the resistors in Figure 6–72 have the same value, when voltage is applied across terminals A and B, the current is
   (a) greatest in R_5 **(b)** greatest in R_3, R_4, and R_5
   (c) greatest in R_1 and R_2 **(d)** the same in all the resistors
4. Two 1.0 kΩ resistors are in series and this series combination is in parallel with a 2.2 kΩ resistor. The voltage across one of the 1.0 kΩ resistors is 6 V. The voltage across the 2.2 kΩ resistor is
   (a) 6 V **(b)** 3 V **(c)** 12 V **(d)** 13.2 V

6238262 2013/12/10 12:27:66.5

5. The parallel combination of a 330 Ω resistor and a 470 Ω resistor is in series with the parallel combination of four 1.0 kΩ resistors. A 10 V source is connected across the circuit. The resistor with the most current has a value of
   (a) 1.0 kΩ **(b)** 330 Ω **(c)** 470 Ω

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6. In the circuit described in Question 5, the resistor(s) with the most voltage has a value of

(a) 1.0 kΩ **(b)** 470 Ω **(c)** 330 Ω

7. In the circuit described in Question 5, the percentage of the total current through any single 1.0 kΩ resistor is

(a) 100% **(b)** 25%

(c) 50% **(d)** 31.25%

8. The output of a certain voltage divider is 9 V with no load. When a load is connected, the output voltage

(a) increases
(b) decreases
(c) remains the same
(d) becomes zero

9. A certain voltage divider consists of two 10 kΩ resistors in series. Which of the following load resistors will have the most effect on the output voltage?

(a) 1.0 MΩ **(b)** 20 kΩ

(c) 100 kΩ **(d)** 10 kΩ

10. When a load resistance is connected to the output of a voltage-divider circuit, the current drawn from the source

(a) decreases
(b) increases
(c) remains the same
(d) is cut off

11. The output voltage of a balanced Wheatstone bridge is

(a) equal to the source voltage
(b) equal to zero
(c) dependent on all of the resistor values in the bridge
(d) dependent on the value of the unknown resistor

12. The primary method of analyzing a circuit with two or more voltage sources is usually

(a) Thevenin's theorem
(b) Ohm's law
(c) superposition
(d) Kirchhoff's law

13. In a certain two-source circuit, one source acting alone produces 10 mA through a given branch. The other source acting alone produces 8 mA in the opposite direction through the same branch. With both sources, the total current through the branch is

(a) 10 mA
(b) 8 mA
(c) 18 mA
(d) 2 mA

14. A Thevenin equivalent circuit consists of

(a) a voltage source in series with a resistance
(b) a voltage source in parallel with a resistance
(c) a current source in parallel with a resistance
(d) two voltage sources and a resistance

15. A voltage source with an internal source resistance of 300 Ω transfers maximum power to a

(a) 150 Ω load
(b) 50 Ω load
(c) 300 Ω load
(d) 600 Ω load

16. You are measuring the voltage at a given point in a circuit that has very high resistance values and the measured voltage is a little lower than it should be. This is possibly because of

(a) one or more of the resistance values being off
(b) the loading effect of the voltmeter
(c) the source voltage is too low
(d) all of these answers

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

TROUBLESHOOTING: SYMPTOM AND CAUSE

The purpose of these exercises is to help develop thought processes essential to troubleshooting. Answers are at the end of the chapter.

Determine the cause for each set of symptoms. Refer to Figure 6–73.

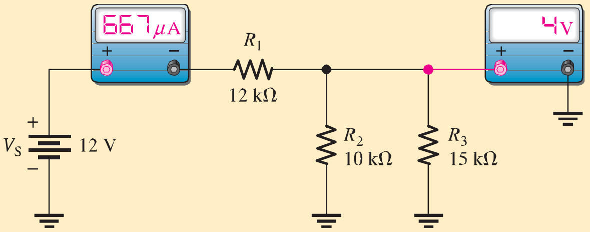

FIGURE 6–73 **The meters indicate the correct readings for this circuit.**

1. *Symptom:* The ammeter reading is too low, and the voltmeter reading is 5.45 V.

   Cause:

   (a) R_1 is open.
   (b) R_2 is open.
   (c) R_3 is open.

2. *Symptom:* The ammeter reading is 1 mA, and the voltmeter reading is 0 V.

   Cause:

   (a) There is a short across R_1.
   (b) There is a short across R_2.
   (c) R_3 is open.

3. *Symptom:* The ammeter reading is near zero, and the voltmeter reading is 12 V.

   Cause:

   (a) R_1 is open.
   (b) R_2 is open.
   (c) Both R_2 and R_3 are open.

4. *Symptom:* The ammeter reading is 444 μA, and the voltmeter reading is 6.67 V.

   Cause:

   (a) R_1 is shorted.
   (b) R_2 is open.
   (c) R_3 is open.

5. *Symptom:* The ammeter reading is 2 mA, and the voltmeter reading is 12 V.

   Cause:

   (a) R_1 is shorted.
   (b) R_2 is shorted.
   (c) Both R_2 and R_3 are open.

PROBLEMS

Answers to odd-numbered problems are at the end of the book.

BASIC PROBLEMS

SECTION 6–1 Identifying Series-Parallel Relationships

1. Identify the series and parallel relationships in Figure 6–74 as seen from the source terminals.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12:27:06.5

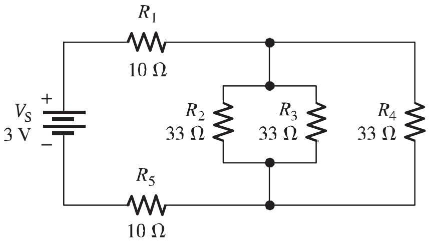

FIGURE 6–74

2. Visualize and draw the following series-parallel combinations:
   - **(a)** R_1 in series with the parallel combination of R_2 and R_3.
   - **(b)** R_1 in parallel with the series combination of R_2 and R_3.
   - **(c)** R_1 in parallel with a branch containing R_2 in series with a parallel combination of four other resistors
3. Visualize and draw the following series-parallel circuits:
   - **(a)** a parallel combination of three branches, each containing two series resistors
   - **(b)** a series combination of three parallel circuits, each containing two parallel resistors
4. In each circuit of Figure 6–75 identify the series and parallel relationships of the resistors viewed from the source.

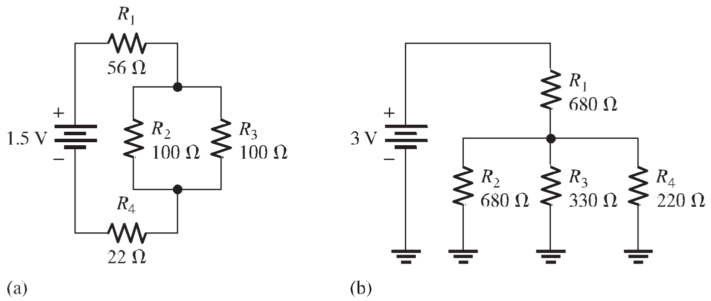

(a) (b)

FIGURE 6–75

SECTION 6–2 Analysis of Series-Parallel Resistive Circuits

5. A certain circuit is composed of two parallel resistors. The total resistance is 667 Ω. One of the resistors is 1.0 kΩ. What is the other resistor?
6. For the circuit in Figure 6–76, determine the total resistance between *A* and *B*.
7. Determine the total resistance for each circuit in Figure 6–75.
8. Determine the current through each resistor in Figure 6–74; then calculate each voltage drop.
9. Determine the current through each resistor in both circuits of Figure 6–75; then calculate each voltage drop.
10. In Figure 6–77, find the following:
    - **(a)** total resistance between terminals *A* and *B*
    - **(b)** total current drawn from a 6 V source connected from *A* to *B*

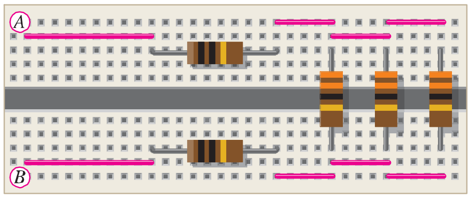

FIGURE 6–76

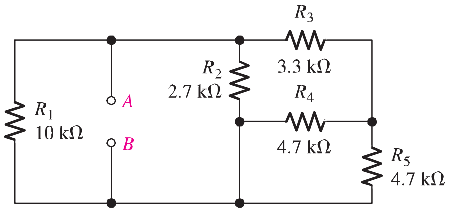

FIGURE 6–77

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

(c) current through R_5
(d) voltage across R_2

11. In Figure 6–77, determine the current through R_2 for $V_{AB} = 6\text{ V}$.

12. In Figure 6–77, determine the current through R_4 for $V_{AB} = 6\text{ V}$.

SECTION 6–3 Voltage Dividers with Resistive Loads

13. A voltage divider consists of two 56 kΩ resistors and a 15 V source. Calculate the unloaded output voltage taken across one of the 56 kΩ resistors. What will the output voltage be if a load resistor of 1.0 MΩ is connected across the output?

14. A 12 V battery output is divided down to obtain two output voltages. Three 3.3 kΩ resistors are used to provide the two outputs with only one output at a time loaded with 10 kΩ. Determine the output voltages in both cases.

15. Which will cause a smaller decrease in output voltage for a given voltage divider, a 10 kΩ load or a 56 kΩ load?

16. In Figure 6–78, determine the current drain on the battery with no load on the output terminals. With a 10 kΩ load, what is the current from the battery?

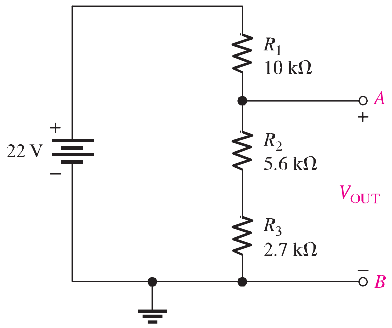

FIGURE 6–78

SECTION 6–4 Loading Effect of a Voltmeter

17. Across which one of the following resistances will a voltmeter with a 10 MΩ internal resistance present the minimum load on a circuit?
(a) 100 kΩ
(b) 1.2 MΩ
(c) 22 kΩ
(d) 8.2 MΩ

18. A certain voltage divider consists of three 1.0 MΩ resistors connected in series to a 100 V source. Determine the voltage across one of the resistors measured by a 10 MΩ voltmeter.

19. What is the difference between the measured and the actual unloaded voltage in Problem 18?

20. By what percentage does the voltmeter in Problem 18 alter the voltage which it measures?

21. A 10,000 Ω/V VOM is used on the 10 V scale to measure the output of a voltage divider. If the divider consists of two series 100 kΩ resistors, what fraction of the source voltage will be measured across one of the resistors?

22. If a DMM with 10 MΩ input resistance is used instead of the VOM in Problem 21, what percentage of the source voltage will be measured by the DMM?

SECTION 6–5 The Wheatstone Bridge

23. A resistor of unknown value is connected to a Wheatstone bridge circuit. The bridge parameters for a balanced condition are set as follows: $R_V = 18\text{ k}\Omega$ and $R_2/R_4 = 0.02$. What is R_X?

24. A bridge network is shown in Figure 6–79. To what value must R_V be set in order to balance the bridge?

25. Determine the value of R_X in the balanced bridge in Figure 6–80.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

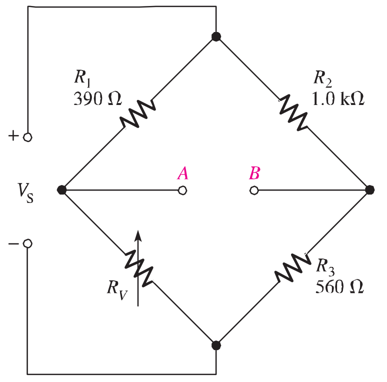
R_1
390 Ω
R_2
1.0 kΩ
+
V_S
−
A
B
R_V
R_3
560 Ω

FIGURE 6–79

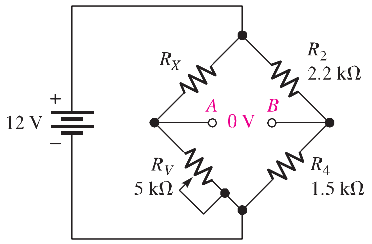
+
12 V
−
R_X
R_2
2.2 kΩ
A
0 V
B
R_V
5 kΩ
R_4
1.5 kΩ

FIGURE 6–80

26. Determine the output voltage of the unbalanced bridge in Figure 6–81 for a temperature of 65°C. The thermistor has a nominal resistance of 1 kΩ at 25°C and a positive temperature coefficient. Assume that its resistance changes 5 Ω for each C° change in temperature.

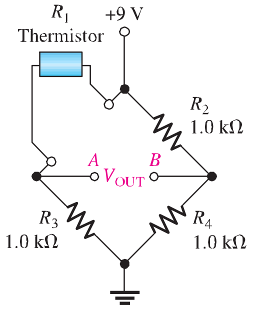
R_1
Thermistor
+9 V
R_2
1.0 kΩ
A
V_{OUT}
B
R_3
1.0 kΩ
R_4
1.0 kΩ

FIGURE 6–81

SECTION 6–6 Thevenin's Theorem

27. Reduce the circuit in Figure 6–82 to its Thevenin equivalent as viewed from terminals *A* and *B*.

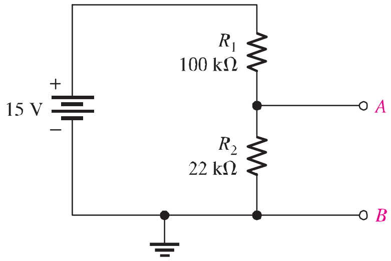
R_1
100 kΩ
+
15 V
−
A
R_2
22 kΩ
B

FIGURE 6–82

28. For each circuit in Figure 6–83, determine the Thevenin equivalent as seen from terminals *A* and *B*.

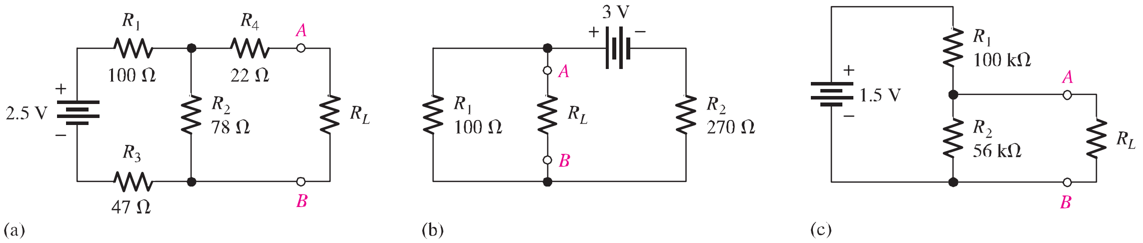
R_1
100 Ω
R_4
22 Ω
A
+
2.5 V
−
R_2
78 Ω
R_L
R_3
47 Ω
B
(a)
3 V
+
−
A
R_1
100 Ω
R_L
B
R_2
270 Ω
(b)
R_1
100 kΩ
+
1.5 V
−
A
R_2
56 kΩ
R_L
B
(c)

FIGURE 6–83

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

29. Determine the voltage and current for R_L in Figure 6–84.

FIGURE 6–84

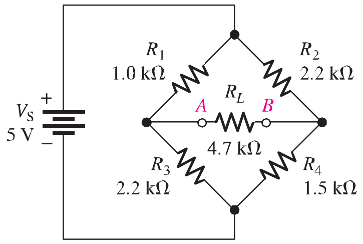
R_1
1.0 kΩ
R_2
2.2 kΩ
R_L
A
B
4.7 kΩ
V_S
5 V
+
−
R_3
2.2 kΩ
R_4
1.5 kΩ

SECTION 6–7 The Maximum Power Transfer Theorem

30. Determine the value of a load resistor connected between terminals A and B in Figure 6–82 for maximum power transfer to the load resistor.

31. A certain Thevenin equivalent circuit has a $V_{TH} = 5.5$ V and an $R_{TH} = 75\ \Omega$. To what value of load resistor will maximum power be transferred?

32. Determine the value of R_L in Figure 6–83(a) for which R_L dissipates maximum power.

SECTION 6–8 The Superposition Theorem

33. In Figure 6–85, use the superposition theorem to find the current in R_3.

34. In Figure 6–85, what is the current through R_2?

FIGURE 6–85

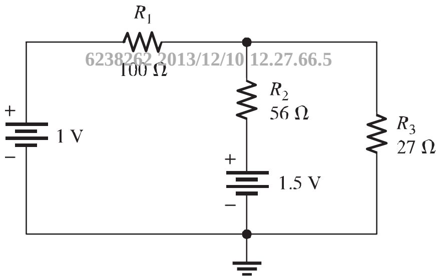
R_1
100 Ω
R_2
56 Ω
R_3
27 Ω
1 V
1.5 V
+
−

6238262 2013/12/10 12.27.66.5

SECTION 6–9 Troubleshooting

35. Is the voltmeter reading in Figure 6–86 correct? If not, what is the problem?

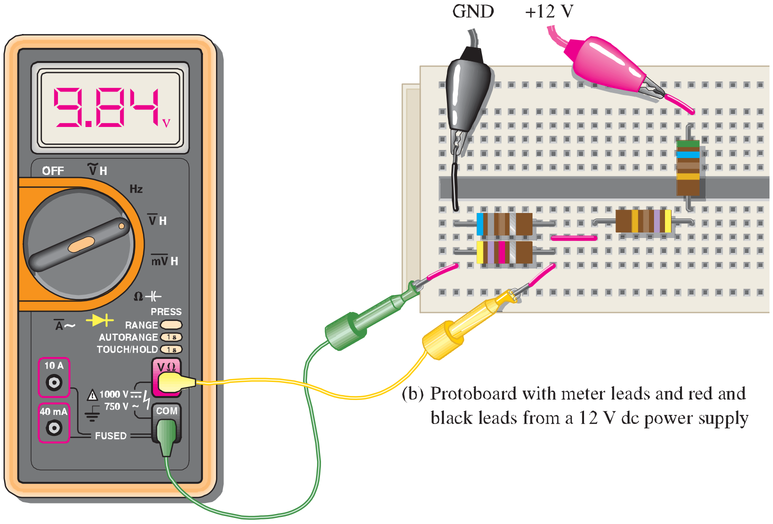
GND
+12 V
9.84 V
OFF
V H
Hz
V H
mV H
Ω
PRESS
RANGE
AUTORANGE
TOUCH/HOLD
10 A
40 mA
VΩ
1000 V
750 V ~
COM
FUSED

(a) Meter with leads going to protoboard

(b) Protoboard with meter leads and red and black leads from a 12 V dc power supply

FIGURE 6–86

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

36. If R_2 in Figure 6–87 opens, what voltages will be read at points A, B, and C?

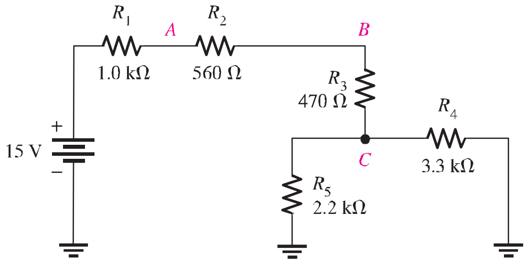

R1
A
R2
B
1.0 kΩ
560 Ω
R3
470 Ω
R4
+
15 V
−
C
3.3 kΩ
R5
2.2 kΩ

FIGURE 6–87

37. Check the meter readings in Figure 6–88 and locate any fault that may exist.

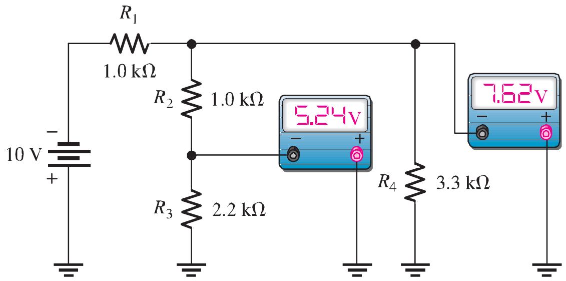

R1
1.0 kΩ
R2
1.0 kΩ
5.24V
7.62V
−
+
10 V
R4
3.3 kΩ
R3
2.2 kΩ

FIGURE 6–88

38. Determine the voltage you would expect to measure across each resistor in Figure 6–87 for each of the following faults. Assume the faults are independent of each other.
(a) R_1 open **(b)** R_3 open **(c)** R_4 open **(d)** R_5 open
(e) point C shorted to ground

39. Determine the voltage you would expect to measure across each resistor in Figure 6–88 for each of the following faults:
(a) R_1 open **(b)** R_2 open **(c)** R_3 open **(d)** a short across R_4

ADVANCED PROBLEMS

40. In each circuit of Figure 6–89, identify the series and parallel relationships of the resistors viewed from the source.

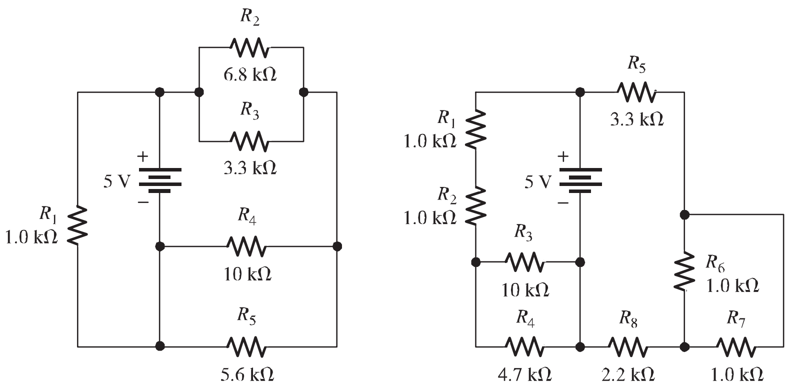

R2
6.8 kΩ
R3
3.3 kΩ
+
5 V
−
R1
1.0 kΩ
R4
10 kΩ
R5
5.6 kΩ
R5
3.3 kΩ
R1
1.0 kΩ
+
5 V
−
R2
1.0 kΩ
R3
10 kΩ
R6
1.0 kΩ
R4
R8
R7
4.7 kΩ
2.2 kΩ
1.0 kΩ

(a) (b)

FIGURE 6–89

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

41. Draw the schematic of the PC board layout in Figure 6–90 showing resistor values and identify the series-parallel relationships. Which resistors, if any, can be removed with no effect on R_T?

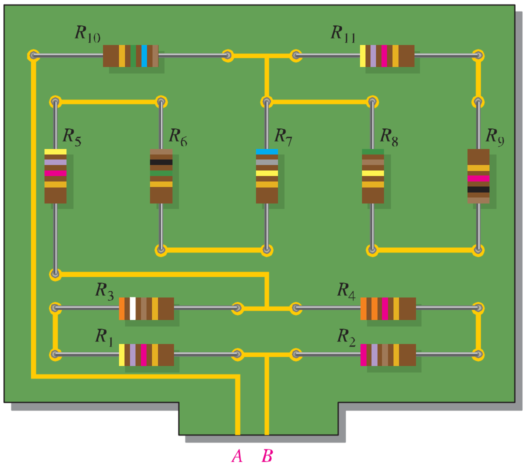
R_{10}
R_{11}
R_5
R_6
R_7
R_8
R_9
R_3
R_4
R_1
R_2
A
B

FIGURE 6–90

42. For the circuit shown in Figure 6–91, calculate the following:

(a) total resistance across the source

(b) total current from the source

(c) current through the 910 Ω resistor

(d) voltage from point A to point B

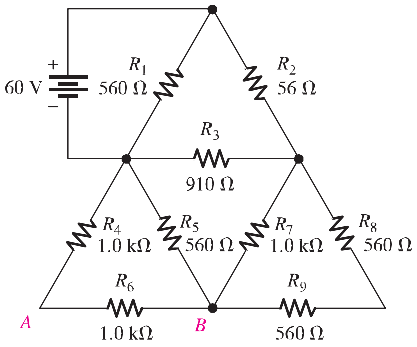
+
60 V
−
R_1
560 Ω
R_2
56 Ω
R_3
910 Ω
R_4
1.0 kΩ
R_5
560 Ω
R_7
1.0 kΩ
R_8
560 Ω
R_6
1.0 kΩ
R_9
560 Ω
A
B

FIGURE 6–91

43. Determine the total resistance and the voltage at points A, B, and C in the circuit of Figure 6–92.

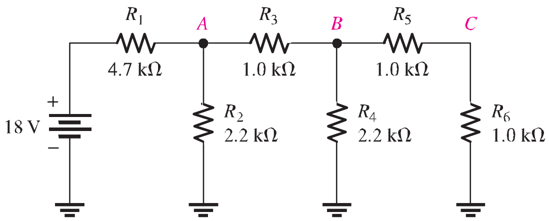
R_1
4.7 kΩ
A
R_3
1.0 kΩ
B
R_5
1.0 kΩ
C
+
18 V
−
R_2
2.2 kΩ
R_4
2.2 kΩ
R_6
1.0 kΩ

FIGURE 6–92

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

44. Determine the total resistance between terminals *A* and *B* of the circuit in Figure 6–93. Also calculate the current in each branch with 10 V between *A* and *B*.

45. What is the voltage across each resistor in Figure 6–93? There are 10 V between *A* and *B*.

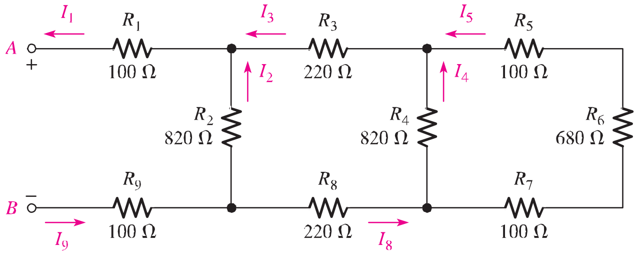
I_1
R_1
I_3
R_3
I_5
R_5
A
+
100 Ω
I_2
220 Ω
I_4
100 Ω
R_2
820 Ω
R_4
820 Ω
R_6
680 Ω
R_9
R_8
R_7
B
–
100 Ω
220 Ω
100 Ω
I_9
I_8

FIGURE 6–93

46. Determine the voltage, V_{AB}, in Figure 6–94.

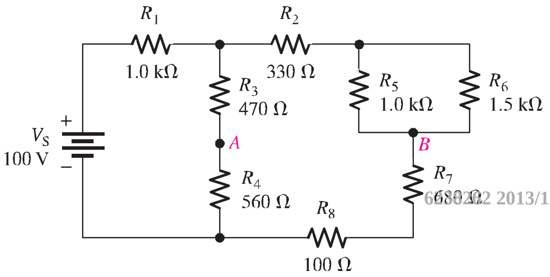
R_1
R_2
1.0 kΩ
330 Ω
R_3
470 Ω
R_5
1.0 kΩ
R_6
1.5 kΩ
+
V_S
100 V
–
A
B
R_7
680 Ω
R_4
560 Ω
R_8
100 Ω

6680282 2013/12/10 12.27.66.5

FIGURE 6–94

47. Find the value of R_2 in Figure 6–95.

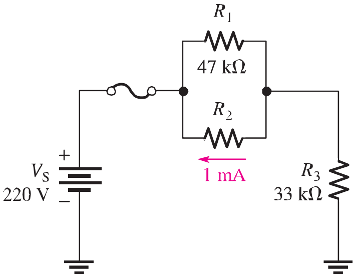
R_1
47 kΩ
R_2
1 mA
+
V_S
220 V
–
R_3
33 kΩ

FIGURE 6–95

48. Determine the total resistance and the voltage at points *A*, *B*, and *C* in the circuit of Figure 6–96.

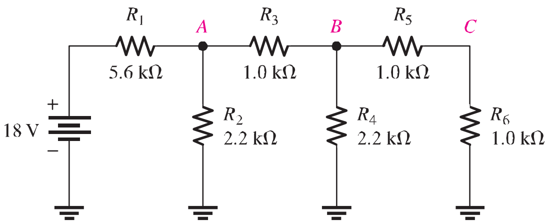
R_1
A
R_3
B
R_5
C
5.6 kΩ
1.0 kΩ
1.0 kΩ
+
18 V
–
R_2
2.2 kΩ
R_4
2.2 kΩ
R_6
1.0 kΩ

FIGURE 6–96

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

49. Develop a voltage divider to provide a 6 V output with no load and a minimum of 5.5 V across a 1.0 kΩ load. The source voltage is 24 V, and the unloaded current is not to exceed 100 mA.

50. Determine the resistance values for a voltage divider that must meet the following specifications: The current under an unloaded condition is not to exceed 5 mA. The source voltage is to be 10 V. A 5 V output and a 2.5 V output are required. Draw the circuit. Determine the effect on the output voltages if a 1.0 kΩ load is connected to each output.

51. Using the superposition theorem, calculate the current in the right-most branch of Figure 6–97.

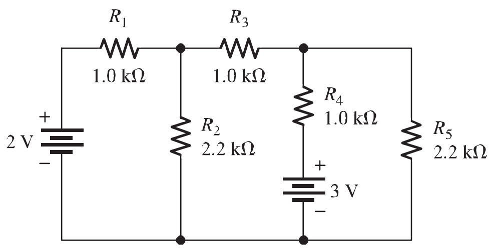
R_1
1.0 kΩ
R_3
1.0 kΩ
R_4
1.0 kΩ
+
2 V
−
R_2
2.2 kΩ
R_5
2.2 kΩ
+
3 V
−

FIGURE 6–97

52. Determine V_{OUT} for the circuit in Figure 6–98 for the following conditions:
(a) Switch SW2 connected to +12 V and the rest to ground
(b) Switch SW1 connected to +12 V and the rest to ground

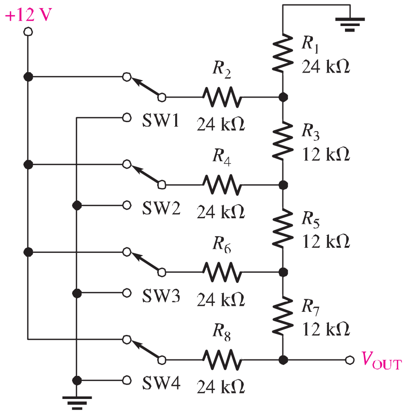
+12 V
R_1
24 kΩ
R_2
SW1
24 kΩ
R_3
12 kΩ
R_4
SW2
24 kΩ
R_5
12 kΩ
R_6
SW3
24 kΩ
R_7
12 kΩ
R_8
V_{OUT}
SW4
24 kΩ

FIGURE 6–98

53. The voltage divider in Figure 6–99 has a switched load. Determine the voltage at each tap (V_1, V_2, and V_3) for each position of the switch.

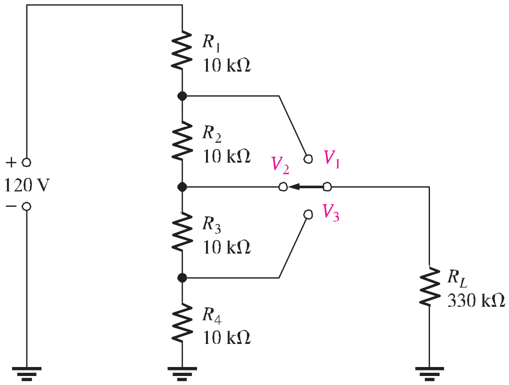
R_1
10 kΩ
R_2
10 kΩ
V_2
V_1
+
120 V
−
V_3
R_3
10 kΩ
R_L
330 kΩ
R_4
10 kΩ

FIGURE 6–99

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

54. Figure 6–100 shows a dc biasing arrangement for a field-effect transistor amplifier. Biasing is a common method for setting up certain dc voltage levels required for proper amplifier operation. Although you may not be familiar with transistor amplifiers at this point, the dc voltages and currents in the circuit can be determined using methods that you already know.

(a) Find V_G and V_S with respect to ground

(b) Determine I_1, I_2, I_D, and I_S

(c) Find V_{DS} and V_{DG}

FIGURE 6–100

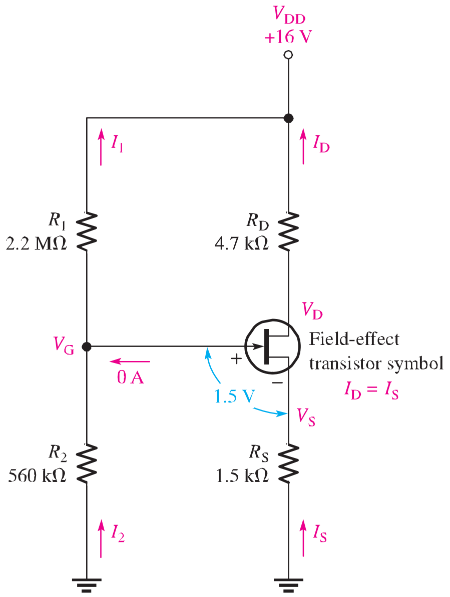

55. For the circuit in Figure 6–100, what is V_G if R_1 is open?

56. For the circuit in Figure 6–100, what is V_G if R_1 and R_2 are reversed?

57. There is one fault in Figure 6–101. Based on the voltmeter indications, determine what the fault is.

FIGURE 6–101

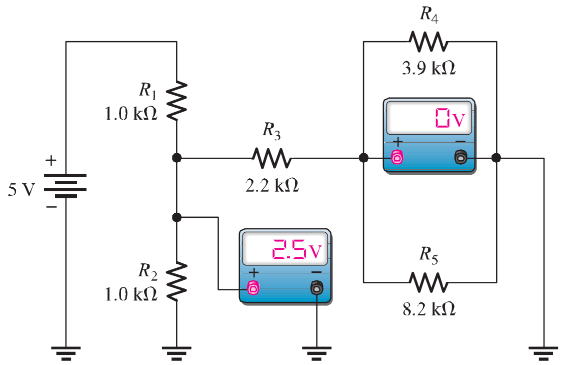

58. Assume a student decided to increase resistors R_1 and R_2 in Figure 6–30 to ten times the values. The result was a drop in the voltage from 3.7 V to 2.8 V. Explain why this happened.

59. Calculate the base voltage for the circuit in Figure 6–31. Assume there is no loading effect.

60. Repeat problem 59 if the transistors have the same loading effect as a 470 kΩ resistor.

61. Assume a 4-20 mA current loop is terminated with a 150 Ω resistor. What voltage is represented by a 4 mA current? What voltage is represented by a 20 mA current?

62. For the liquid-level sensing system in Figure 6–40, assume the gross weight of the tank is 1700 pounds when full and 300 pounds when empty. What percentage should the display indicate if the load cells indicate a total weight of 700 pounds?

63. For the liquid-level sensing system in Figure 6–40, what is the advantage of using three load cells over four?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

MULTISIM

MULTISIM TROUBLESHOOTING PROBLEMS

64. Open file P06-64; files are found at www.pearsonhighered.com/floyd. Determine if there is a fault in the circuit. If so, identify the fault.

65. Open file P06-65 and determine if there is a fault in the circuit. If so, identify the fault.

66. Open file P06-66 and determine if there is a fault in the circuit. If so, identify the fault.

67. Open file P06-67 and determine if there is a fault in the circuit. If so, identify the fault.

68. Open file P06-68 and determine if there is a fault in the circuit. If so, identify the fault.

69. Open file P06-69 and determine if there is a fault in the circuit. If so, identify the fault.

70. Open file P06-70 and determine if there is a fault in the circuit. If so, identify the fault.

71. Open file P06-71 and determine if there is a fault in the circuit. If so, identify the fault.

ANSWERS TO SECTION CHECKUPS

SECTION 6–1 Identifying Series-Parallel Relationships

1. See Figure 6–102.

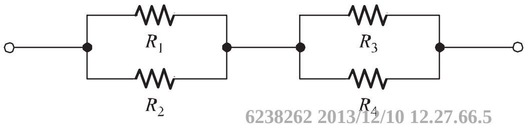

6238262 2013/12/10 12.27.66.5

FIGURE 6–102

2. R_1 and R_2 are in series with the parallel combination of R_3 and R_4.
3. All resistors are in parallel.
4. R_1 and R_2 are in parallel; R_3 and R_4 are in parallel.
5. Yes, the two parallel combinations are in series with each other.

SECTION 6–2 Analysis of Series-Parallel Resistive Circuits

1. $R_T = R_1 + R_4 + R_2 \| R_3 = 599\ \Omega$
2. $I_3 = 11.2\ \text{mA}$
3. $V_{R2} = I_2 R_2 = 3.7\ \text{V}$
4. $R_T = 89\ \Omega$; $I_T = 11.2\ \text{mA}$

SECTION 6–3 Voltage Dividers with Resistive Loads

1. The load resistor decreases the output voltage.
2. True
3. $V_{\text{OUT(unloaded)}} = 19.23\ \text{V}$, $V_{\text{OUT(loaded)}} = 19.16\ \text{V}$

SECTION 6–4 Loading Effect of a Voltmeter

1. A voltmeter loads a circuit because the internal resistance of the meter appears in parallel with the circuit resistance across which it is connected, reducing the resistance between those two points of the circuit and drawing current from the circuit.
2. No, because the meter resistance is much larger than 1.0 kΩ.
3. Yes.
4. 4.0 MΩ

SECTION 6–5 The Wheatstone Bridge

1. See Figure 6–103.
2. A bridge is balanced when the output voltage is zero.
3. $R_X = 15\ \text{k}\Omega$

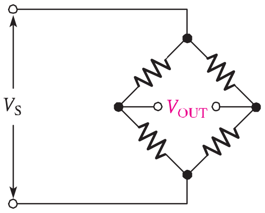

FIGURE 6–103

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

4. An unbalanced bridge is used to measure transducer-sensed quantities.
5. A load cell is a transducer that uses strain gauges to convert mechanical force into an electrical signal.

SECTION 6–6 Thevenin's Theorem

1. A Thevenin equivalent circuit consists of V_{TH} and R_{TH}.
2. See Figure 6–104.
3. V_{TH} is the open circuit voltage between two points in a circuit.
4. R_{TH} is the resistance as viewed from two terminals in a circuit, with all sources replaced by their internal resistances.
5. See Figure 6–105.

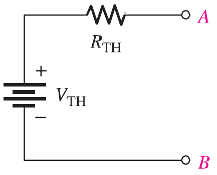

FIGURE 6–104

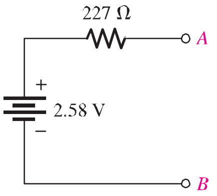

FIGURE 6–105

SECTION 6–7 The Maximum Power Transfer Theorem

1. The maximum power transfer theorem states that maximum power is transferred from a source to a load when the load resistance is equal to the internal source resistance.
2. Maximum power is delivered to a load when $R_L = R_S$.
3. $R_L = R_S = 50\ \Omega$

SECTION 6–8 The Superposition Theorem

1. The total current in any branch of a multiple-source linear circuit is equal to the algebraic sum of the currents due to the individual sources acting alone, with the other sources replaced by their internal resistances.
2. The superposition method allows each source to be treated independently.
3. A short simulates the zero internal resistance of an ideal voltage source.
4. The net current is in the direction of the larger current.

SECTION 6–9 Troubleshooting

1. Opens and short are two common faults.
2. **(a)** 62.8 V **(b)** 62.8 V **(c)** 62 V **(d)** 100 V **(e)** 0 V
3. The 10 kΩ resistor is open.

ANSWERS TO RELATED PROBLEMS FOR EXAMPLES

6–1 The added resistor would be in parallel with the combination of R_4 in series with R_2 and R_3 in parallel.

6–2 The added resistor is in parallel with R_5.

6–3 A to gnd: $R_T = R_3 \| (R_1 + R_2) + R_4$
B to gnd: $R_T = R_2 \| (R_1 + R_3) + R_4$
C to gnd: $R_T = R_4$

6–4 None. The new resistor will be shorted by the existing connection between those points.

6–5 55.1 Ω

6–6 128.3 Ω

6–7 $I_1 = 8.93$ mA; $I_3 = 5.85$ mA; $I_T = 18.2$ mA

6–8 $V_1 = V_2 = 10.3$ V; $V_3 = 9.70$ V; $V_4 = 3.16$ V; $V_5 = 6.54$ V

6–9 $I_1 = 1.42$ mA, $P_1 = 6.67$ mW; $I_2 = 756\ \mu A$, $P_2 = 3.55$ mW;
$I_3 = 2.18$ mA, $P_3 = 4.75$ mW; $I_4 = 1.13$ mA, $P_4 = 1.28$ mW;
$I_5 = 1.06$ mA, $P_5 = 758\ \mu$W; $I_6 = 1.06$ mA, $P_6 = 435\ \mu$W

6–10 3.39 V

6–11 The current will increase because there is a smaller loading effect on the circuit. The new current in R_{L2} is 59 μA.

6–12 5.19 V

6–13 3.3 kΩ

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6–14 0.49 V
6–15 2.36 V; 124 Ω
6–16 $V_{\text{TH}(AB)} = 3.46$ V; $R_{\text{TH}(AB)} = 3.08$ kΩ
6–17 1.17 mA
6–18 41.7 mW
6–19 22.0 mA
6–20 5 mA
6–21 5.73 V, 0 V
6–22 9.46 V
6–23 $V_A = 12$ V; $V_B = 13.8$ V

ANSWERS TO TRUE/FALSE QUIZ

1. T **2.** F **3.** F **4.** T **5.** T **6.** T **7.** F
8. F **9.** T **10.** F

ANSWERS TO SELF-TEST

1. (e) **2.** (c) **3.** (c) **4.** (c) **5.** (b) **6.** (a) **7.** (b)
8. (b) **9.** (d) **10.** (b) **11.** (b) **12.** (c) **13.** (d) **14.** (a)
15. (c) **16.** (d)

ANSWERS TO TROUBLESHOOTING: SYMPTOM AND CAUSE

1. (c) **2.** (b) **3.** (c) **4.** (b) **5.** (a)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

CHAPTER 7

MAGNETISM AND ELECTROMAGNETISM

OUTLINE

7–1 The Magnetic Field
7–2 Electromagnetism
7–3 Electromagnetic Devices
7–4 Magnetic Hysteresis
7–5 Electromagnetic Induction
7–6 DC Generators
7–7 DC Motors

OBJECTIVES

- Explain the principles of the magnetic field
- Explain the principles of electromagnetism
- Describe the principle of operation for several types of electromagnetic devices
- Explain magnetic hysteresis
- Discuss the principle of electromagnetic induction
- Explain how a dc generator works
- Explain how a dc motor works

KEY TERMS

Magnetic field
Lines of force
Magnetic flux
Weber (Wb)
Tesla (T)
Gauss
Hall effect
Electromagnetism
Electromagnetic field
Permeability
Reluctance ($\mathscr{R}$)
Magnetomotive force (mmf)
Ampere-turn (At)
Solenoid
Relay
Speaker
Magnetic field intensity
Hysteresis
Retentivity
Induced voltage (v_{ind})
Electromagnetic induction
Induced current (i_{ind})
Faraday's law
Lenz's law

INTRODUCTION

This chapter departs from the previous six chapters and introduces two new concepts: magnetism and electromagnetism. The operation of many types of electrical devices is based partially on magnetic or electromagnetic principles. Electromagnetic induction is important in an electrical component called an inductor or coil (covered in Chapter 11).

Two types of magnets are the permanent magnet and the electromagnet. The permanent magnet maintains a constant magnetic field between its two poles with no external excitation. The electromagnet produces a magnetic field only when there is current through it. The electromagnet is basically a coil of wire wound around a magnetic core material. The chapter includes an introduction to dc generators and to dc motors. Motors are widely used in industrial systems, so a basic understanding of motors is important.

VISIT THE WEBSITE
Study aids for this chapter are available at
http://pearsonhighered.com/floyd

6238262 2013/12/10 12.27.66.5

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

7–1 THE MAGNETIC FIELD

A permanent magnet has a magnetic field surrounding it. A magnetic field is visualized by lines of force that radiate from the north pole (N) to the south pole (S) and back to the north pole through the magnetic material.

After completing this section, you should be able to

- **Explain the principles of the magnetic field**
  - **Define *magnetic flux***
  - **Define *magnetic flux density***
  - **Discuss how materials are magnetized**
  - **Explain how a magnetic switch works**

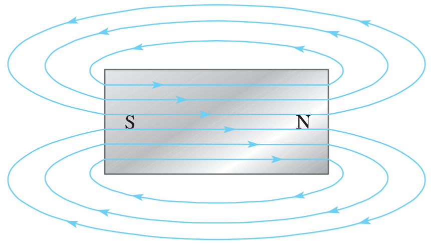

Blue lines represent only a few of the many magnetic lines of force in the magnetic field.

FIGURE 7–1 **Magnetic lines of force around a bar magnet.**

A permanent magnet, such as the bar magnet shown in Figure 7–1, has a magnetic field surrounding it. All magnetic fields have their origin in moving charge, which in solid materials is caused by moving electrons. In certain materials such as iron, atoms can be aligned so that the electron motion is reinforced, creating an observable field that extends in three dimensions. Even some electrical insulators can exhibit this behavior; ceramics make excellent magnets but are electrical insulators.

To explain and illustrate magnetic fields, Michael Faraday drew "lines of force" or flux lines to represent the unseen field. Flux lines are widely used as a description of that field, showing the strength and direction of the field. The flux lines never cross. When lines are close together, the field is more intense; when they are farther apart, the field is weaker. The flux lines are always drawn from the north pole (N) to the south pole (S) of a magnet. Even in small magnets, the number of lines based on the mathematical definition is extremely large, so for clarity, only a few lines are generally shown in drawings of magnetic fields.

When unlike poles of two permanent magnets are placed close together, their magnetic fields produce an attractive force, as indicated in Figure 7–2(a). When two like poles are brought close together, they repel each other, as shown in part (b).

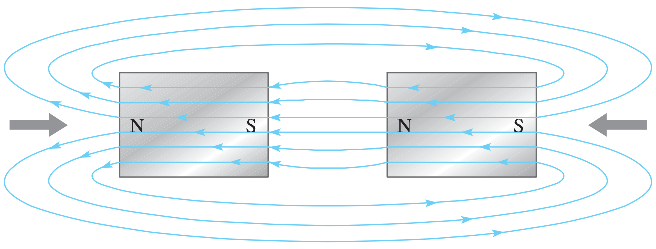

(a) Unlike poles attract.

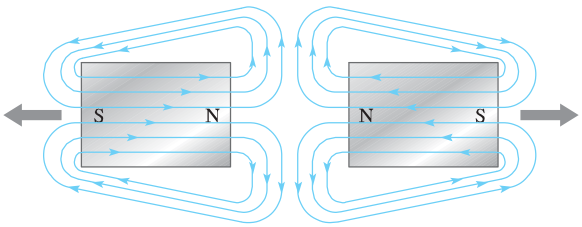

(b) Like poles repel.

FIGURE 7–2 **Magnetic attraction and repulsion.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

When a nonmagnetic material such as paper, glass, wood, or plastic is placed in a magnetic field, the lines of force are unaltered, as shown in Figure 7–3(a). However, when a magnetic material such as iron is placed in the magnetic field, the lines of force tend to change course and pass through the iron rather than through the surrounding air. They do so because the iron provides a magnetic path that is more easily established than that of air. Figure 7–3(b) illustrates this principle. The fact that magnetic lines of force follow a path through iron or other materials is a consideration in the design of shields that prevent stray magnetic fields from affecting sensitive circuits.

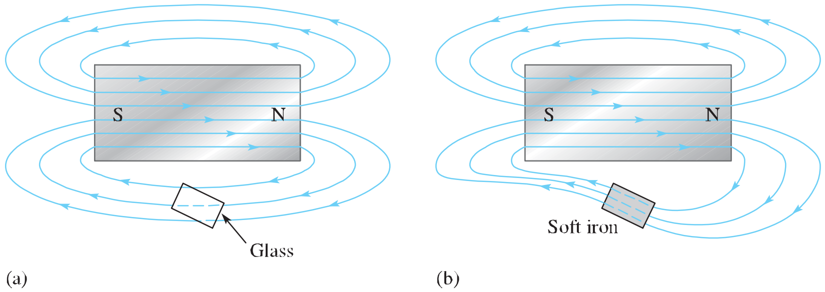

FIGURE 7–3 **Effect of (a) nonmagnetic and (b) magnetic materials on a magnetic field.**

Magnetic Flux (ϕ)

The group of force lines going from the north pole to the south pole of a magnet is called the **magnetic flux**, symbolized by ϕ (the lowercase Greek letter phi). A stronger field is represented by more lines of force. Several factors determine the strength of a magnet, including the material and physical geometry as well as the distance from the magnet. Magnetic field lines tend to be more concentrated at the poles.

The unit of magnetic flux is the **weber (Wb)**. One weber equals 10^8 lines. The weber is a very large unit; thus, in most practical situations, the microweber (μWb) is used. One microweber equals 100 lines of magnetic flux.

Magnetic Flux Density (B)

The **magnetic flux density** is the amount of flux per unit area perpendicular to the magnetic field. Its symbol is B, and its unit is the **tesla (T)**. One tesla equals one weber per square meter (Wb/m^2). The following formula expresses the flux density:

$$B = \frac{\phi}{A} \tag{7–1}$$

where ϕ is the flux in webers (Wb) and A is the cross-sectional area in square meters (m^2) of the magnetic field.

EXAMPLE 7–1

Compare the flux and the flux density in the two magnetic cores shown in Figure 7–4. The diagram represents the cross section of a magnetized material. Assume that each dot represents 100 lines or 1 μWb.

SOLUTION

The flux is simply the number of lines. In Figure 7–4(a) there are 49 dots. Each represents 1 μWb, so the flux is 49 μWb. In Figure 7–4(b) there are 72 dots, so the flux is 72 μWb.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

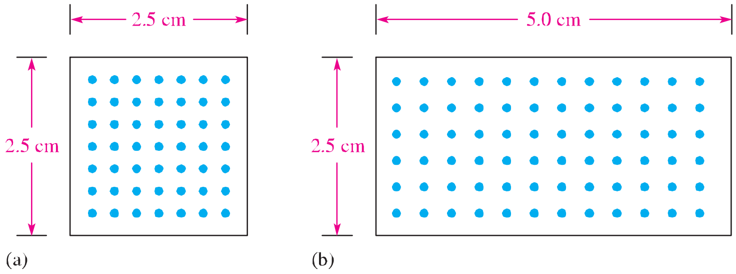

FIGURE 7–4

To calculate the flux density, first calculate the area in m^2. For Figure 7–4(a) the area is

$$A = l \times w = 0.025 \text{ m} \times 0.025 \text{ m} = 6.25 \times 10^{-4} \text{ m}^2$$

For Figure 7–4(b) the area is

$$A = l \times w = 0.025 \text{ m} \times 0.050 \text{ m} = 1.25 \times 10^{-3} \text{ m}^2$$

Use Equation 7–1 to calculate the flux density. For Figure 7–4(a) the flux density is

6238262 2013/12/10 12.27.66.5

$$B = \frac{\phi}{A} = \frac{49 \ \mu\text{Wb}}{6.25 \times 10^{-4} \text{ m}^2} = 78.4 \times 10^{-3} \text{ Wb/m}^2 = 78.4 \times 10^{-3} \text{ T}$$

For Figure 7–4(b) the flux density is

$$B = \frac{\phi}{A} = \frac{72 \ \mu\text{Wb}}{1.25 \times 10^{-3} \text{ m}^2} = 57.6 \times 10^{-3} \text{ Wb/m}^2 = 57.6 \times 10^{-3} \text{ T}$$

The data in Table 7–1 compares the two cores. Note that the core with the largest flux does not necessarily have the highest flux density.

TABLE 7–1

	FLUX (Wb)	AREA (m^2)	FLUX DENSITY (T)
Figure 7–4(a):	49 μWb	$6.25 \times 10^{-4} \text{ m}^2$	78.4×10^{-3} T
Figure 7–4(b):	72 μWb	$1.25 \times 10^{-3} \text{ m}^2$	57.6×10^{-3} T

RELATED PROBLEM*

What is the flux density if the same flux shown in Figure 7–4(a) is in a core that is 5.0 cm × 5.0 cm?

*Answers are at the end of the chapter.

EXAMPLE 7–2

If the flux density in a certain magnetic material is 0.23 T and the area of the material is 0.38 in.^2, what is the flux through the material?

SOLUTION

First, 0.38 in.^2 must be converted to square meters. 39.37 in. = 1 m; therefore,

$$A = 0.38 \text{ in.}^2 [1 \text{ m}^2/(39.37 \text{ in.})^2] = 245 \times 10^{-6} \text{ m}^2$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The flux through the material is

$$\phi = BA = (0.23\ \text{T})(245 \times 10^{-6}\ \text{m}^2) = \mathbf{56.4\ \mu Wb}$$

RELATED PROBLEM

Calculate B if $A = 0.05\ \text{in.}^2$ and $\phi = 1000\ \mu\text{Wb}$.

THE GAUSS Although the tesla (T) is the SI unit for flux density, another unit called the **gauss (G)** from the CGS (centimeter-gram-second) system, is used ($10^4\ \text{G} = 1\ \text{T}$). In fact, the instrument used to measure flux density is the gaussmeter. A typical gaussmeter is shown in Figure 7–5. This particular gaussmeter is a portable unit with four ranges that can measure magnetic fields as small as the earth's field (about 0.5 G, but changes depending on location) to strong fields such as in an MRI unit (about 10,000 G). The unit *gauss* is still in widespread use, so you should be familiar with it as well as the *tesla*.

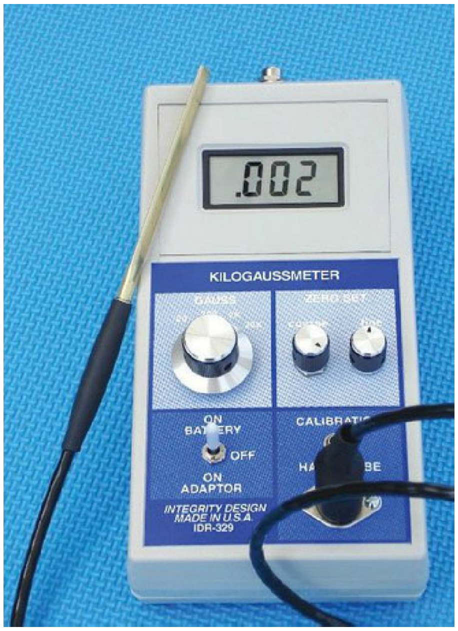

FIGURE 7–5 **A dc gaussmeter.**
(Integrity Model IDR-329 distributed by Less EMF Inc.)

How Materials Become Magnetized

Ferromagnetic materials such as iron, nickel, and cobalt become magnetized when placed in the magnetic field of a magnet. We have all seen a permanent magnet pick up paper clips, nails, or iron filings. In these cases, the object becomes magnetized (that is, it actually becomes a magnet itself) under the influence of the permanent magnetic field and becomes attracted to the magnet. When removed from the magnetic field, the object tends to lose its magnetism.

The magnetic material affects not only the magnetic flux density at the poles but also how the magnetic flux density falls off as distance from the poles increases. The physical size also affects the flux density. For example, two disk magnets (both made from sintered Alnico) have very similar densities near the pole, but the larger magnet has much higher flux density, as you move away from the pole, as shown in Figure 7–6. Notice that the flux density falls off rapidly as you move away from the pole. This type of plot can illustrate if a given magnet is effective for a specific application that depends on the distance in which the magnet must work.

Ferromagnetic materials have minute magnetic domains created within their atomic structure by the orbital motion and spin of electrons. These domains can be thought of as

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

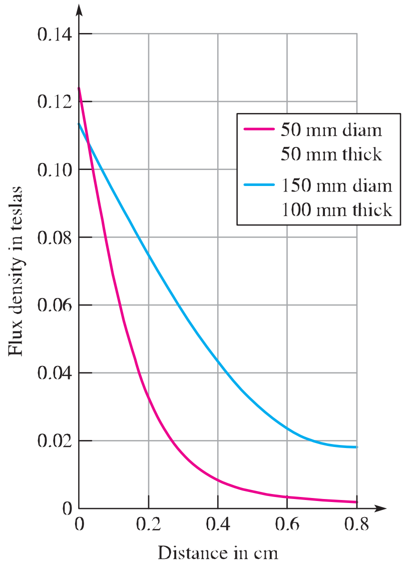

FIGURE 7–6 **Example of magnetic flux densities for two disk magnets as a function of distance. The blue curve represents the larger magnet.**

very small bar magnets with north and south poles. When the material is not exposed to an external magnetic field, the magnetic domains are randomly oriented, as shown in Figure 7–7(a). When the material is placed in a magnetic field, the domains align themselves, as shown in part (b). Thus, the object itself effectively becomes a magnet.

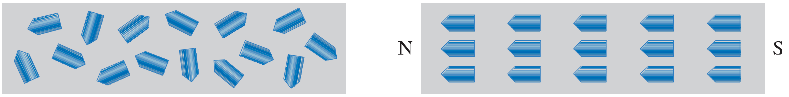

(a) The magnetic domains (N ◀ S) are randomly oriented in the unmagnetized material.

(b) The magnetic domains become aligned when the material is magnetized.

FIGURE 7–7 **Ferromagnetic domains in (a) an unmagnetized and (b) a magnetized material.**

The type of material is an important parameter for the actual flux density of magnets. Table 7–2 lists the flux densities of typical magnetic fields in teslas. For permanent magnets, the numbers given are based on the flux density of the field that is typical if measured close to the pole. As previously discussed, these values can drop significantly as distance from the poles is increased. The strongest field most people will ever experience is about 1 T (10,000 G) if they have an MRI exam. The strongest commercially available permanent

TABLE 7–2 • Flux density of various magnetic fields.

SOURCE	TYPICAL FLUX DENSITY IN TESLAS (T)
Earth's magnetic field	4×10^{-5} (varies with location)
Small "refrigerator" magnets	0.08 to 0.1
Ceramic magnets	0.2 to 0.3
Alnico 5 reed switch magnet	0.1 to 0.2
Neodymium magnets	0.3 to 0.52
Magnetic resonance imaging (MRI)	1
The strongest steady magnetic field ever achieved in a laboratory	45

SAFETY NOTE

Many strong magnets are very brittle and can shatter on impact. Eye protection should always be worn when you work with strong magnets. Strong magnets are not toys and should not be given to children. People with pacemakers should stay away from strong magnetic fields.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

magnets are neodymium-iron-boron composites (NdFeB). To find the flux density in gauss, multiply the values in teslas by 10^4 (10,000).

Sensitive Magnetic Sensor

A new sensor developed by the National Institute of Standards and Technology (NIST) has the ability to measure the very tiny magnetic field associated with a human heartbeat. The heart's magnetic signature is measured in picoteslas (trillionths of a tesla), which is a million times smaller than Earth's magnetic field. The measurement had to be made in a building with magnetic shielding to block out effects of Earth's magnetic field.

The miniature sensors may eventually find application as manetocardiograms, as a supplement to electrocardiograms.

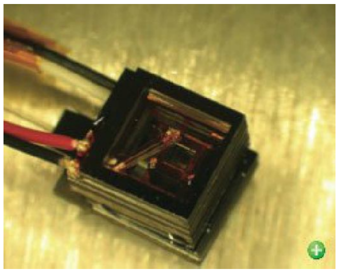

Courtesy of S. Knappe/NIST.

SYSTEM NOTE

Applications

Permanent magnets are widely used in brushless motors (discussed in Section 7–7), magnetic separators, speakers, microphones, automobiles, and magnetic resonance-imaging devices. They are also commonly used in switches, such as the normally closed switch illustrated in Figure 7–8. When the magnet is near the switch mechanism, as in Figure 7–8 (a), the switch is closed. When the magnet is moved away, as in part (b), the spring pulls the arm open. Magnetic switches are widely used in security systems.

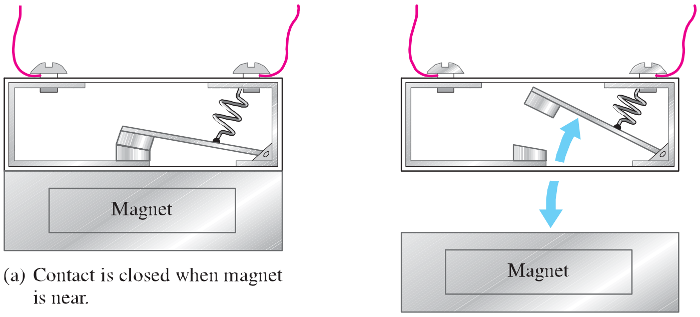

(a) Contact is closed when magnet is near.

(b) Contact opens when magnet is moved away.

FIGURE 7–8 Operation of a magnetic switch.

System Applications for Magnetic Switches

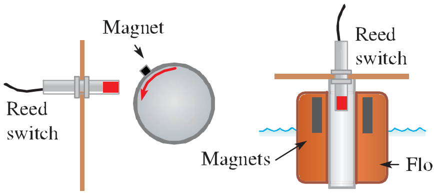

Magnetic switches are widely used in systems. One important application is for proximity detectors, which are useful for counting operations or measuring the speed of movement in rotating machinery. Reed type magnetic switches sense a moving magnet as it moves by the switch. Reed switches can open and close up to 1000 times per second and can have an expected lifetime of three billion cycles.

Magnetic switches are also used in float detectors for monitoring liquid level. The float has a magnet that moves up and down with the fluid. As the magnets pass by the switch, the reeds are pulled together closing the switch.

6238262 2013/12/10 12.27.66.5

SYSTEM NOTE

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

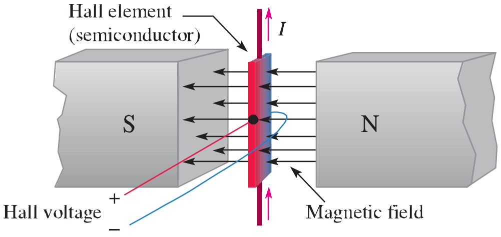

FIGURE 7–9 **The Hall effect. The Hall voltage is induced across the Hall element—the red side is positive; the blue side is negative.**

Another important application of permanent magnets is in sensors that take advantage of an effect known as the Hall effect. The **Hall effect** is the occurrence of a small voltage (a few μV) that is generated on opposite sides of a thin current-carrying conductor or semiconductor (the Hall element) that is in a magnetic field. A voltage called the Hall voltage appears across the Hall element, as illustrated in Figure 7–9. The Hall voltage is due to the forces exerted on the electrons as they traverse the magnetic field, causing an excess of charge on one side of the Hall element. Although the effect was first noticed in a conductor, it is more pronounced in semiconductors, which are normally used in Hall-effect sensors. Notice that the magnetic field, the electric current, and the Hall voltage are all at right angles to each other. This voltage is amplified and can be used to detect the presence of the magnetic field. The detection of a magnetic field is useful in sensor applications.

Hall-effect sensors are widely used because they are small, inexpensive, and have no moving parts. In addition, they are noncontacting sensors, so they can last for billions of repeated operations, a clear advantage over contacting sensors that can wear out. Hall-effect sensors can detect the nearby presence of a magnet by sensing its magnetic field. For this reason, they can be used for position measurements or for sensing motion. They are used in conjunction with other sensor elements to measure current, temperature, or pressure.

There are many system applications for Hall-effect sensors. In automobiles, Hall-effect sensors are used to measure various parameters such as throttle angle, crankshaft and camshaft positions, distributor position, tachometer, power seat and rear-view mirror positions. A few other applications for Hall-effect sensors include measuring parameters for rotating devices such as drills, fans, flow meter vanes, and disk speed detection. They are also used in dc motors as will be seen in Section 7–7.

SECTION 7–1 CHECKUP*

1. When the north poles of two magnets are placed close together, do they repel or attract each other?
2. What is the difference between magnetic flux and magnetic flux density?
3. What are two units for measuring magnetic flux density?
4. What is the flux density when $\phi = 4.5\ \mu\text{Wb}$ and $A = 5 \times 10^{-3}\ \text{m}^2$?

***Answers are at the end of the chapter.**

7–2 ELECTROMAGNETISM

Electromagnetism is the production of a magnetic field by current in a conductor.

After completing this section, you should be able to

- **Explain the principles of electromagnetism**
  - **Determine the direction of the magnetic lines of force**
  - **Define *permeability***
  - **Define *reluctance***
  - **Define *magnetomotive force***
  - **Describe a basic electromagnet**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Current produces a magnetic field, called an **electromagnetic field**, around a conductor, as illustrated in Figure 7–10. The invisible lines of force of the magnetic field form a concentric circular pattern around the conductor and are continuous along its length. The direction of the lines of force surrounding a conductor for the given current direction is as indicated. The lines of force are in a clockwise direction. When current direction is reversed, the magnetic field lines are in a counterclockwise direction.

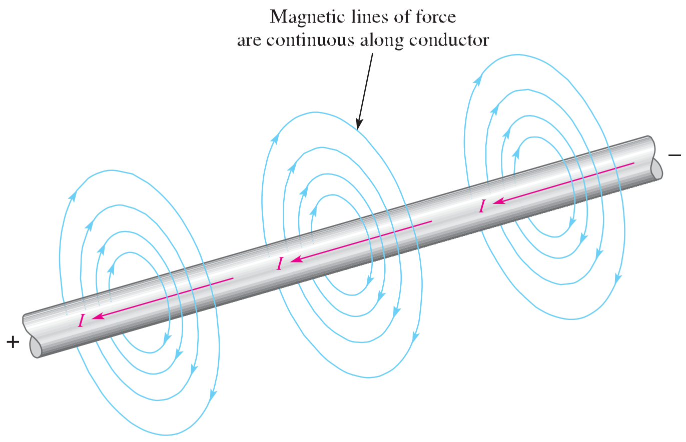

FIGURE 7–10 **Magnetic field around a current-carrying conductor. The red arrows indicate the direction of electron flow current (− to +).**

Although the magnetic field cannot be seen, it is capable of producing visible effects. For example, if a current-carrying wire is inserted through a sheet of paper in a perpendicular direction, iron filings placed on the surface of the paper arrange themselves along the magnetic lines of force in concentric rings, as illustrated in Figure 7–11(a). Part (b) of the figure illustrates that the needle of a compass placed in the electromagnetic field will point in the direction of the lines of force. The field is stronger closer to the conductor and becomes weaker with increasing distance from the conductor.

LEFT-HAND RULE An aid to remembering the direction of the lines of force is illustrated in Figure 7–12. Imagine that you are grasping the conductor with your left hand, with your thumb pointing in the direction of current. Your fingers indicate the direction of the magnetic lines of force.

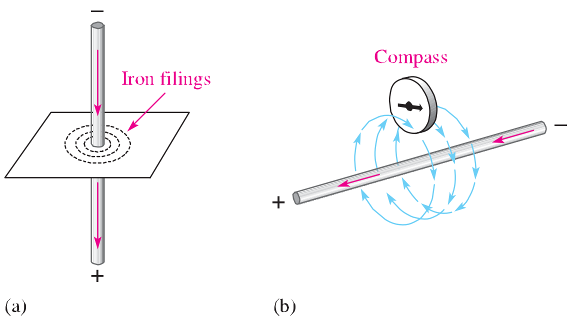

FIGURE 7–11 **Visible effects of an electromagnetic field.**

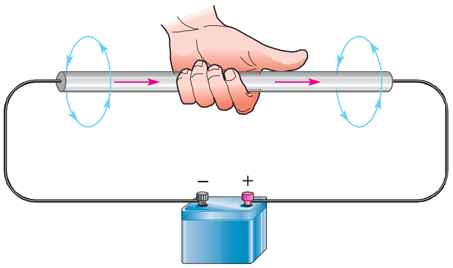

FIGURE 7–12 **Illustration of left-hand rule. The left hand rule is used for electron flow current (− to +).**

Electromagnetic Properties

Several important properties related to electromagnetic fields are now discussed.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

PERMEABILITY (μ) The ease with which a magnetic field can be established in a given material is measured by the **permeability** of that material. The higher the permeability, the more easily a magnetic field can be established. The symbol of permeability is μ (the Greek letter mu).

A material's permeability depends on its type. The permeability of a vacuum (μ_0) is $4\pi \times 10^{-7}$ Wb/At · m (webers/ampere-turn · meter) and is used as a reference. Ferromagnetic materials typically have permeabilities hundreds of times larger than that of a vacuum, indicating that a magnetic field can be set up with relative ease in these materials. Ferromagnetic materials include iron, steel, nickel, cobalt, and their alloys.

The *relative permeability* (μ_r) of a material is the ratio of its absolute permeability (μ) to the permeability of a vacuum (μ_0).

$$\mu_r = \frac{\mu}{\mu_0} \tag{7–2}$$

Because it is a ratio of permeabilities, μ_r has no units. Typical magnetic materials, such as iron, have a relative permeability of a few hundred. Highly permeable materials can have a relative permeability as high as 100,000.

RELUCTANCE ($\mathcal{R}$) The opposition to the establishment of a magnetic field in a material is called **reluctance ($\mathcal{R}$)**. The value of reluctance is directly proportional to the length (l) of the magnetic path, and inversely proportional to the permeability (μ) and to the cross-sectional area (A) of the material, as expressed by the following equation:

$$\mathcal{R} = \frac{l}{\mu A} \tag{7–3}$$

Reluctance in magnetic circuits is analogous to resistance in electric circuits. The unit of reluctance can be derived using l in meters, A (area) in square meters, and μ in Wb/At · m as follows:

$$\mathcal{R} = \frac{l}{\mu A} = \frac{\cancel{\text{m}}}{(\text{Wb/At}\cdot\cancel{\text{m}})(\cancel{\text{m}}^2)} = \frac{\text{At}}{\text{Wb}}$$

At/Wb is ampere-turns/weber.

Equation 7–3 is similar to Equation 2–6 for determining wire resistance. Recall Equation 2–6 is

$$R = \frac{\rho l}{A}$$

The reciprocal of resistivity (ρ) is conductivity (σ). By substituting $1/\sigma$ for ρ, Equation 2–6 can be written as

$$R = \frac{l}{\sigma A}$$

Compare this last equation for wire resistance with Equation 7–3. The length (l) and the area (A) have the same meaning in both equations. The conductivity (σ) in electrical circuits is analogous to permeability (μ) in magnetic circuits. Also, resistance (R) in electric circuits is analogous to reluctance ($\mathcal{R}$) in magnetic circuits; both are oppositions. Typically, the reluctance of a magnetic circuit is 50,000 At/Wb or more, depending on the size and type of material.

EXAMPLE 7–3

Calculate the reluctance of a torus (a doughnut-shaped core) made of low-carbon steel. The inner radius of the torus is 1.75 cm and the outer radius of the torus is 2.25 cm. Assume the permeability of low-carbon steel is 2×10^{-4} Wb/At · m.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

You must convert centimeters to meters before you calculate the area and length. From the dimensions given, the thickness (diameter) is 0.5 cm = 0.005 m. Thus, the cross-sectional area is

$$A = \pi r^2 = \pi(0.0025)^2 = 1.96 \times 10^{-5}\ \text{m}^2$$

The length is equal to the circumference of the torus measured at the average radius of 2.0 cm or 0.020 m.

$$l = C = 2\pi r = 2\pi(0.020\ \text{m}) = 0.125\ \text{m}$$

Substituting values into Equation 7–3,

$$\mathcal{R} = \frac{l}{\mu A} = \frac{0.125\ \text{m}}{(2 \times 10^{-4}\ \text{Wb/At}\cdot\text{m})(1.96 \times 10^{-5}\ \text{m}^2)} = \mathbf{31.9 \times 10^6\ At/Wb}$$

RELATED PROBLEM

What happens to the reluctance if cast steel with a permeability of 5×10^{-4} Wb/At · m is substituted for the cast iron core?

EXAMPLE 7–4

Mild steel has a relative permeability of 800. Calculate the reluctance of a mild steel core that has a length of 10 cm, and has a cross section of 1.0 cm × 1.2 cm.

SOLUTION

First, determine the permeability of mild steel.

$$\mu = \mu_0\mu_r = (4\pi \times 10^{-7}\ \text{Wb/At}\cdot\text{m})(800) = 1.00 \times 10^{-3}\ \text{Wb/At}\cdot\text{m}$$

Next, convert the length to meters and the area to square meters.

$$l = 10\ \text{cm} = 0.10\ \text{m}$$
$$A = 0.010\ \text{m} \times 0.012\ \text{m} = 1.2 \times 10^{-4}\ \text{m}^2$$

Substituting values into Equation 7–3,

$$\mathcal{R} = \frac{l}{\mu A} = \frac{0.10\ \text{m}}{(1.00 \times 10^{-3}\ \text{Wb/At}\cdot\text{m})(1.2 \times 10^{-4}\ \text{m}^2)} = \mathbf{8.33 \times 10^5\ At/Wb}$$

RELATED PROBLEM

What happens to the reluctance if the core is made from 78 Permalloy with a relative permeability of 4000?

MAGNETOMOTIVE FORCE (mmf) As you have learned, current in a conductor produces a magnetic field. The cause of a magnetic field is called the **magnetomotive force (mmf)**. Magnetomotive force is something of a misnomer because in a physics sense mmf is not really a force, but rather it is a direct result of the movement of charge (current). The unit of mmf, the **ampere-turn (At)**, is established on the basis of the current in a coil of wire. The formula for mmf is

$$F_m = NI \tag{7–4}$$

where F_m is the magnetomotive force, N is the number of turns of wire, and I is the current in amperes.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Figure 7–13 illustrates that a number of turns of wire carrying a current around a magnetic material creates a force that sets up flux lines through the magnetic path. The amount of flux depends on the magnitude of the mmf and on the reluctance of the material, as expressed by the following equation:

$$\phi = \frac{F_m}{\mathscr{R}} \qquad (7\text{–}5)$$

Equation 7–5 is known as the *Ohm's law for electromagnetic circuits* because the flux (ϕ) is analogous to current, the mmf (F_m) is analogous to voltage, and the reluctance ($\mathscr{R}$) is analogous to resistance. Like other phenomena in science, the flux is an effect, the mmf is a cause, and the reluctance is an opposition.

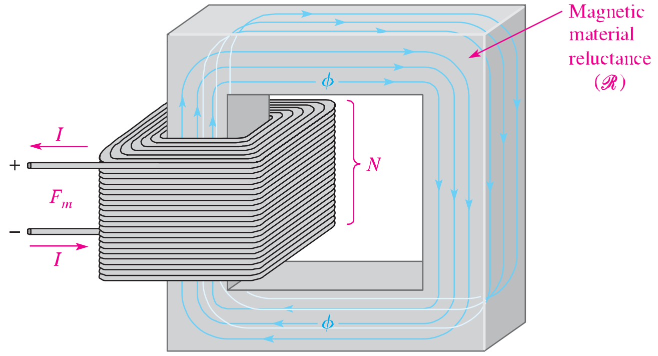

FIGURE 7–13 **A basic electromagnetic circuit.**

One important difference between an electric circuit and a magnetic circuit is that in magnetic circuits, Equation 7–5 is only valid up to a certain point before the magnetic material saturates (flux becomes a maximum). You will see this when you look at magnetization curves in Section 7–4. Another difference that was already noted is that flux does occur in permanent magnets with no source of mmf. In a permanent magnet, the flux is due to internal electron motion rather than to an external current. No equivalent effect occurs in electric circuits.

EXAMPLE 7–5

How much flux is established in the magnetic path of Figure 7–14 if the reluctance of the material is 2.8×10^5 At/Wb?

FIGURE 7–14

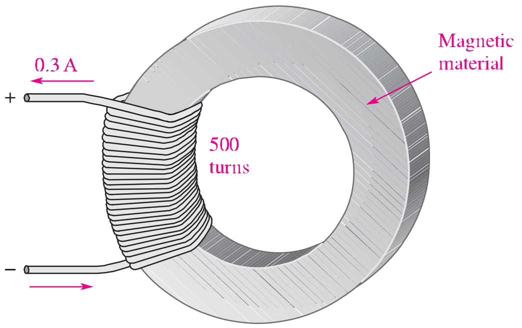

SOLUTION

$$\phi = \frac{F_m}{\mathscr{R}} = \frac{NI}{\mathscr{R}} = \frac{(500\text{ t})(0.3\text{ A})}{2.8 \times 10^5\text{ At/Wb}} = 5.36 \times 10^{-4}\text{ Wb} = \mathbf{536\ \mu Wb}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

RELATED PROBLEM

How much flux is established in the magnetic path of Figure 7–14 if the reluctance is 7.5×10^3 At/Wb, the number of turns is 300, and the current is 0.18 A?

EXAMPLE 7–6

There is 0.1 ampere of current through a coil with 400 turns.

(a) What is the mmf?

(b) What is the reluctance of the circuit if the flux is 250 μWb?

SOLUTION

6238262 2013/12/10 12.27.66.5

(a) $N = 400$ and $I = 0.1$ A

$$F_m = NI = (400\text{ t})(0.1\text{ A}) = \mathbf{40\ At}$$

(b) $$\mathcal{R} = \frac{F_m}{\phi} = \frac{40\text{ At}}{250\ \mu\text{Wb}} = \mathbf{1.60 \times 10^5\ At/Wb}$$

RELATED PROBLEM

Rework the example for $I = 85$ mA, $N = 500$, and $\phi = 500\ \mu$Wb.

In many magnetic circuits, the core is not continuous. For example, if an air gap is cut into the core, it will increase the reluctance of the magnetic circuit. This means that more current is required to establish the same flux as before because an air gap represents a significant opposition to establishing flux. The situation is analogous to a series electrical circuit; the total reluctance of the magnetic circuit is the sum of the reluctance of the core and the reluctance of the air gap.

The Electromagnet

An electromagnet is based on the properties that you have just learned. A basic electromagnet is simply a coil of wire wound around a core material that can be easily magnetized.

The shape of the electromagnet can be designed for various applications. For example, Figure 7–15 shows a U-shaped magnetic core. When the coil of wire is connected to a battery and there is current, as shown in part (a), a magnetic field is established as indicated. If the current direction is reversed, as shown in part (b), the direction of the magnetic field is also reversed. The closer the north and south poles are brought together, the smaller the air gap between them becomes, and the easier it becomes to establish a magnetic field because the reluctance is reduced.

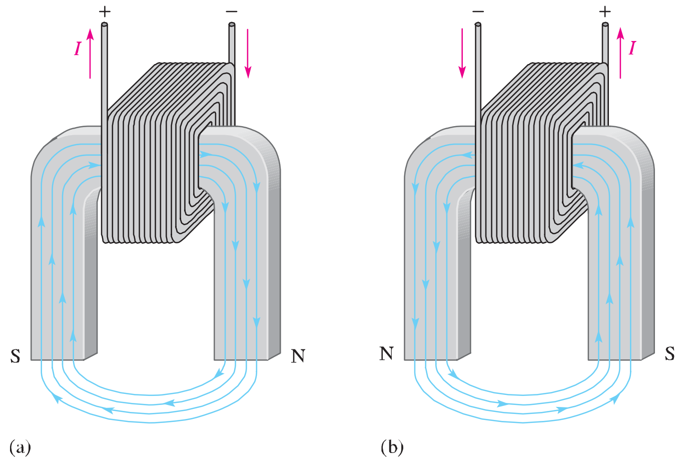

FIGURE 7–15 Reversing the current in the coil causes the electromagnetic field to reverse.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 7–2 CHECKUP

1. Explain the difference between magnetism and electromagnetism.
2. What happens to the magnetic field in an electromagnet when the current through the coil is reversed?
3. State Ohm's law for electromagnetic circuits.
4. Compare each quantity in Question 3 to its electrical counterpart.

7–3 ELECTROMAGNETIC DEVICES

Many types of useful devices such as tape recorders, electric motors, speakers, solenoids, and relays are based on electromagnetism. The transformer is another important example and will be covered in Chapter 14.

After completing this section, you should be able to

- **Describe the principle of operation for several types of electromagnetic devices**
  - **Discuss how a solenoid and a solenoid valve work**
  - **Discuss how a relay works**
  - **Discuss how a speaker works**
  - **Discuss the basic analog meter movement**
  - **Explain a magnetic disk and tape read/write operation**
  - **Explain the concept of the magneto-optical disk**

The Solenoid

The **solenoid** is a type of electromagnetic device that has a movable iron core called a *plunger*. The movement of this iron core depends on both an electromagnetic field and a mechanical spring force. The basic structure of a solenoid is shown in Figure 7–16. It consists of a cylindrical coil of wire wound around a nonmagnetic hollow form. A stationary iron core is fixed in position at the end of the shaft, and a sliding iron core is attached to the stationary core with a spring.

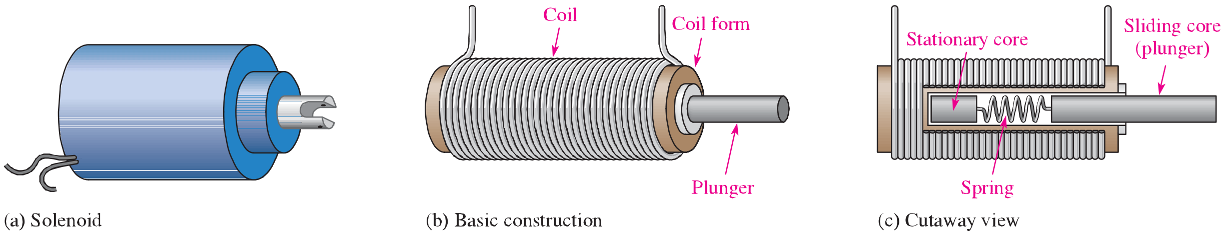

(a) Solenoid (b) Basic construction (c) Cutaway view

FIGURE 7–16 **Basic solenoid structure.**

In the at-rest (or unenergized) state, the plunger is extended as shown in Figure 7–17(a). The solenoid is energized by current through the coil, as shown in part (b). The current sets up an electromagnetic field that magnetizes both iron cores as indicated. The south pole of the stationary core attracts the north pole of the movable core, which causes it to slide inward, thus retracting the plunger and compressing the spring. As long as there is coil current, the plunger remains retracted by the attractive force of the magnetic fields. When the current is cut off, the magnetic fields collapse; and the force of the compressed spring pushes the

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

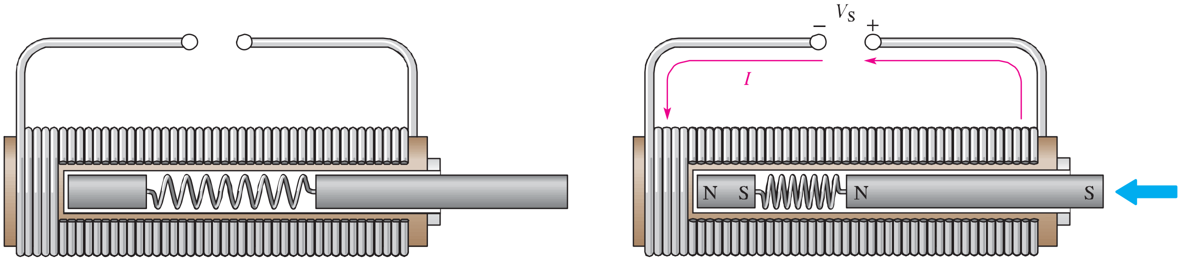

(a) Unenergized (no voltage or current)–plunger extended

(b) Energized–plunger retracted

FIGURE 7–17 **Basic solenoid operation.**

plunger back out. The solenoid is used for applications such as opening and closing valves and automobile door locks.

SOLENOID VALVES In industrial controls, **solenoid valves** are widely used to control the flow of air, water, steam, oils, refrigerants, and other fluids. Solenoid valves are used in both pneumatic (air) and hydraulic (oil) systems, common in machine controls. Solenoid valves are also common in the aerospace and medical fields. Solenoid valves can either move a plunger to open or close a port or can rotate a blocking flap a fixed amount.

A solenoid valve consists of two functional units: a solenoid coil that provides the magnetic field to provide the required movement to open or close the valve and a valve body, which is isolated from the coil assembly via a leakproof seal and includes a pipe and butterfly valve. Figure 7–18 shows a cutaway of one type of solenoid valve. When the solenoid is energized, the butterfly valve is turned to open a normally closed (NC) valve or to close a normally open (NO) valve.

Solenoid valves are available with a wide variety of configurations including normally open or normally closed valves. They are rated for different types of fluids (for example, gas or water), pressures, number of pathways, sizes, and more. The same valve may control more than one line and may have more than one solenoid to move.

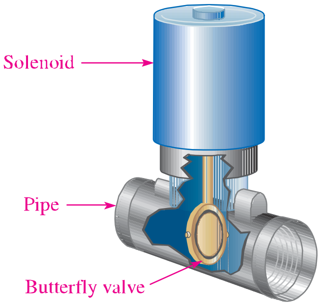

FIGURE 7–18 **A basic solenoid valve structure.**

The Relay

Relays differ from solenoids in that the electromagnetic action is used to open or close electrical contacts rather than to provide mechanical movement. Figure 7–19 shows the basic operation of an armature-type relay with one normally open (NO) contact and one normally closed (NC) contact (single pole–double throw). When there is no coil current,

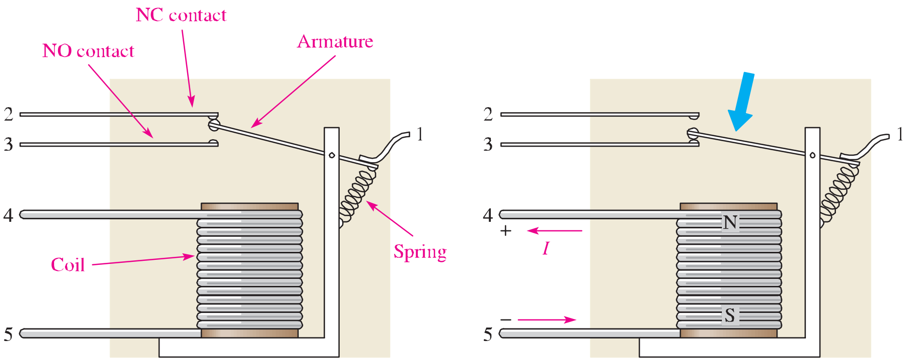

(a) Unenergized: continuity from terminal 1 to terminal 2

(b) Energized: continuity from terminal 1 to terminal 3

FIGURE 7–19 **Basic structure of a single-pole–double-throw armature relay.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

the armature is held against the upper contact by the spring, thus providing continuity from terminal 1 to terminal 2, as shown in Figure 7–19(a). When energized with coil current, the armature is pulled down by the attractive force of the electromagnetic field and makes connection with the lower contact to provide continuity from terminal 1 to terminal 3, as shown in Figure 7–19(b).

A typical armature relay and its schematic symbol are shown in Figure 7–20.

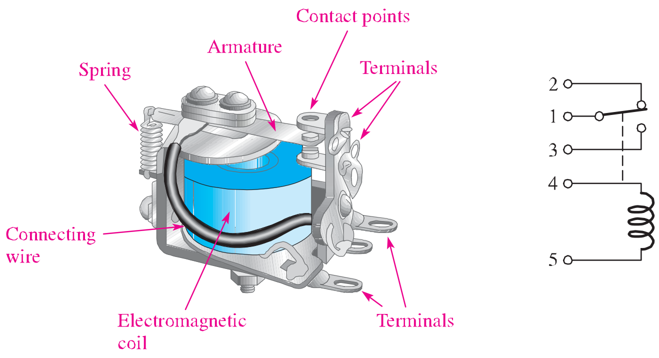

(a) Typical relay structure

(b) Symbol

FIGURE 7–20 **A typical armature relay.**

Another widely used type of relay is the *reed relay*, which is shown in Figure 7–21. The reed relay, like the armature relay, uses an electromagnetic coil. The contacts are thin reeds of magnetic material and are usually located inside the coil. When there is no coil current, the reeds are in the open position as shown in Figure 7–21(b). When there is current through the coil, the reeds make contact because they are magnetized and attract each other as shown in part (c).

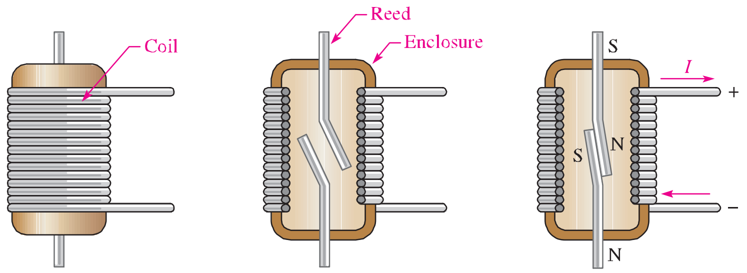

(a)

(b) Unenergized

(c) Energized

FIGURE 7–21 **Basic structure of a reed relay.**

Reed relays are faster, are more reliable, and produce less contact arcing than armature relays. However, they have less current-handling capability than armature relays and are more susceptible to mechanical shock.

SYSTEM EXAMPLE 7–1

AN ALARM SYSTEM

Figure 7–22 shows a simplified intrusion detection system that uses a relay to turn on an audible alarm and lights. The relay uses a battery that allows it to function and sound the audible alarm even if ac power is off. The relay uses 50 mA to close and the audible alarm

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

draws 2 A from the battery. Notice that the contacts for the relay are physically close but drawn separated for clarity of reading the schematic. If any one of the magnetic detection switches closes when the system is *on*, the relay is energized, providing current to the audible alarm and turning on ac power. The system is designed such that the high current for the audible alarm is not in the detection switches; most magnetic switches are low-current devices, so the relay provides a path for the high-current output.

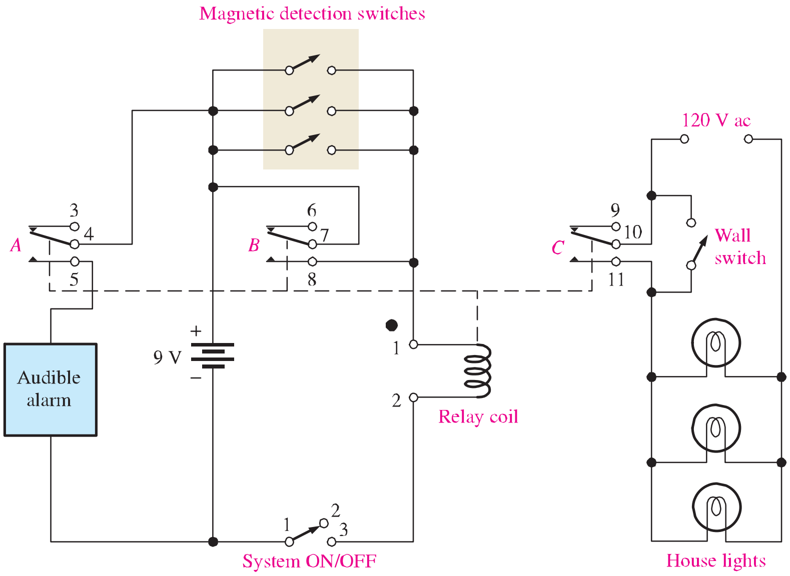

FIGURE 7–22 **Simplified security alarm system.**

The Speaker

A **speaker** is an electromagnetic device that converts electrical signals into sound. Essentially, it is a linear motor that alternately attracts and repels an electromagnet to and from a permanent magnet called a donut magnet. Figure 7–23 shows the key parts of a speaker. The audio signal is connected using very flexible wires to a cylindrical coil called the voice coil. The voice coil and its movable core form an electromagnet, which is suspended in an accordion-like structure called the spider. The spider acts like an accordion spring, keeping the voice coil in the center and restoring it to the rest position when there is no input signal.

Current from the audio input alternates back and forth and powers the electromagnet; when there is more current, the attraction or repulsion is greater. When the input current reverses direction, the polarity of the electromagnet reverses direction also, faithfully following the input signal. The voice coil and its moving magnet are firmly attached to the cone. The cone is a flexible diaphragm that vibrates to produce sound.

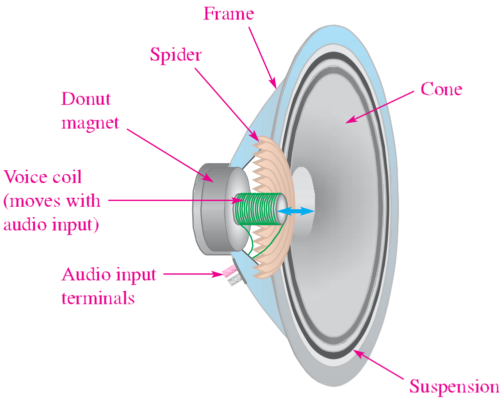

FIGURE 7–23 **Key parts of a speaker (cutaway view).**

Meter Movement

The d'Arsonval meter movement is the most common type used in analog multimeters. In this type of meter movement, the pointer is deflected in proportion to the amount of current through a coil. Figure 7–24 shows a basic d'Arsonval meter movement. It consists of a coil

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

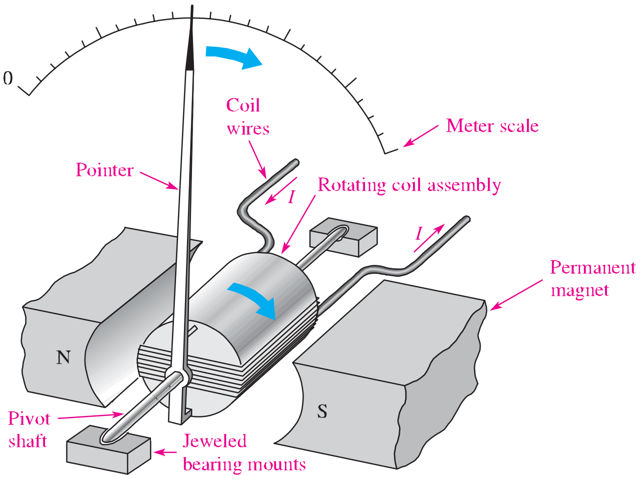

FIGURE 7–24 The basic d'Arsonval meter movement.

of wire wound on a bearing-mounted assembly that is placed between the poles of a permanent magnet. A pointer is attached to the moving assembly. With no current through the coil, a spring mechanism keeps the pointer at its left-most (zero) position. When there is current through the coil, electromagnetic forces act on the coil, causing a rotation to the right. The amount of rotation depends on the amount of current.

Figure 7–25 illustrates how the interaction of magnetic fields produces rotation of the coil assembly. Current is inward at the "cross" and outward at the "dot" in the single winding shown. The inward current produces a counterclockwise electromagnetic field that reinforces the permanent magnetic field below it. The result is an upward force on the left side of the coil as shown. A downward force is developed on the right side of the coil, where the current is outward. These forces produce a clockwise rotation of the coil assembly and are opposed by a spring mechanism. The indicated forces and the spring force are balanced at the value of the current. When current is removed, the spring force returns the pointer to its zero position.

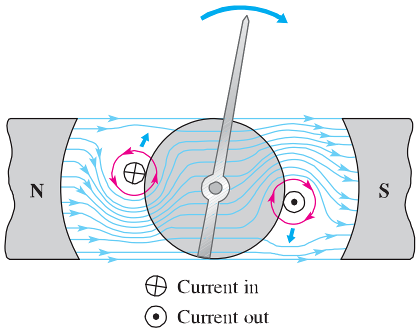

FIGURE 7–25 When the electromagnetic field interacts with the permanent magnetic field, forces are exerted on the rotating coil assembly, causing it to move clockwise and thus deflecting the pointer.

Magnetic Disk and Tape Read/Write Head

A simplified diagram of a magnetic disk or tape surface read/write operation is shown in Figure 7–26. A data bit (1 or 0) is written on the magnetic surface by the magnetization of a small segment of the surface as it moves by the write head. The direction of the magnetic

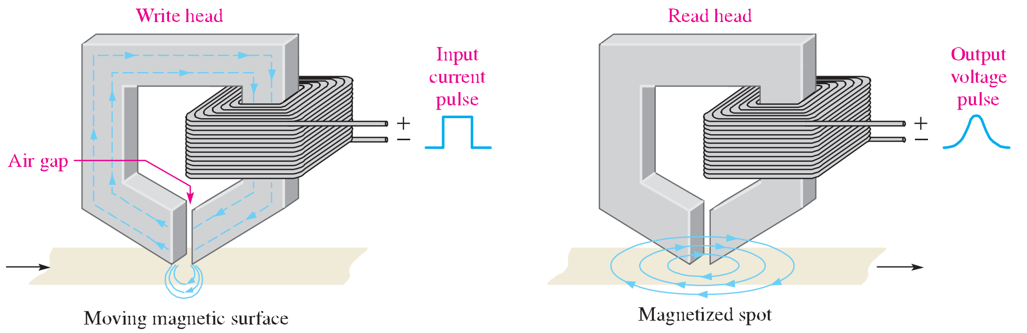

(a) The magnetic flux from the write head follows the low reluctance path through the moving magnetic surface.

(b) When read head passes over magnetized spot, an induced voltage appears at the output.

FIGURE 7–26 Read/write function on a magnetic surface.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

flux lines is controlled by the direction of the current pulse in the winding, as shown in Figure 7–26(a) for the case of a positive pulse. At the air gap in the write head, the magnetic flux takes a path through the surface of the storage device. This magnetizes a small spot on the surface in the direction of the field. A magnetized spot of one polarity represents a binary 1, and one of the opposite polarity represents a binary 0. Once a spot on the surface is magnetized, it remains until written over with an opposite magnetic field.

When the magnetic surface passes a read head, the magnetized spots produce magnetic fields in the read head, which induce voltage pulses in the winding. The polarity of these pulses depends on the direction of the magnetized spot and indicates whether the stored bit is a 1 or a 0. This process is illustrated in Figure 7–26(b). Often the read and write heads are combined into a single unit.

The Magneto-Optical Disk

The magneto-optical disk uses an electromagnet and laser beams to read and write (record) data on a magnetic surface. Magneto-optical disks are formatted in tracks and sectors similar to hard disks. However, because of the ability of a laser beam to be precisely directed to an extremely small spot, magneto-optical disks are capable of storing much more data than standard magnetic hard disks.

Figure 7–27(a) illustrates a small cross-sectional area of a disk before recording, with an electromagnet positioned below it. Tiny magnetic particles, represented by the arrows, are all magnetized in the same direction.

6238262 2013/12/10 12.27.66.5

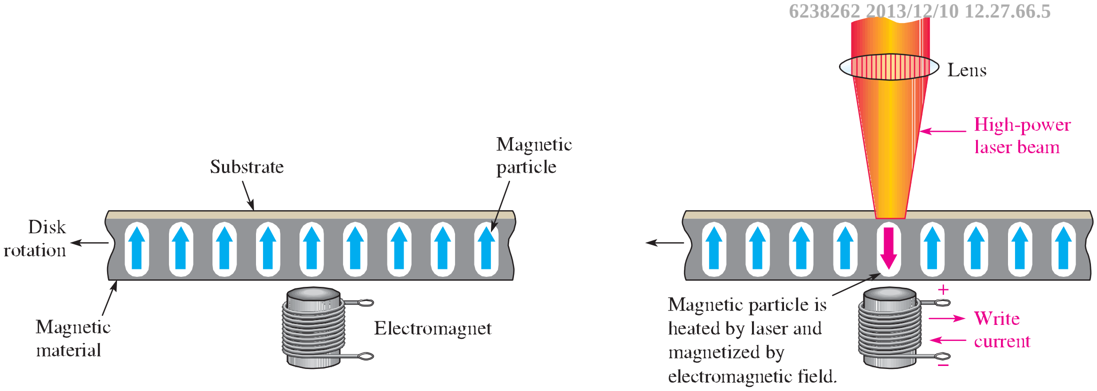
Substrate
Magnetic particle
Disk rotation
Magnetic material
Electromagnet
Lens
High-power laser beam
Magnetic particle is heated by laser and magnetized by electromagnetic field.
+
Write current
−

(a) Small cross-section of unrecorded disk

(b) Writing: A high-power laser beam heats the spot, causing the magnetic particle to align with the electromagnetic field.

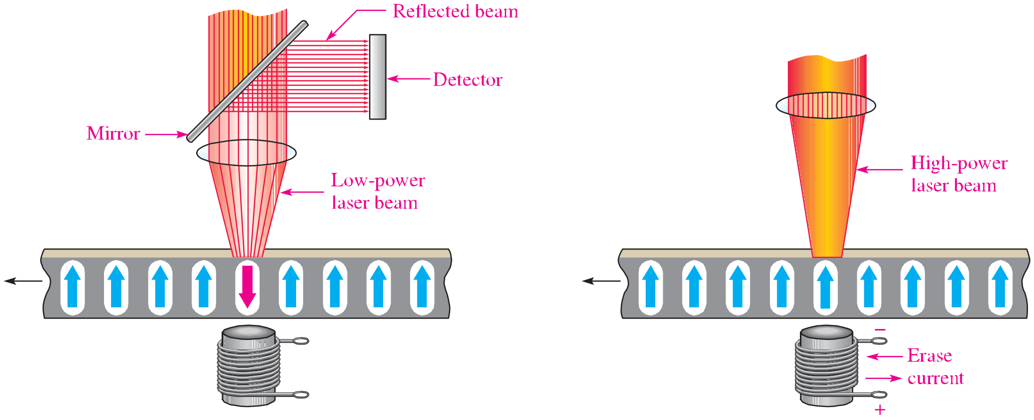
Reflected beam
Detector
Mirror
Low-power laser beam
High-power laser beam
−
Erase current
+

(c) Reading: A low-power laser beam reflects off of the reversed-polarity magnetic particle and its polarization shifts. If the particle is not reversed, the polarization of the reflected beam is unchanged.

(d) Erasing: The electromagnetic field is reversed as the high-power laser beam heats the spot, causing the magnetic particle to be restored to the original polarity.

FIGURE 7–27 **Basic concept of the magneto-optical disk.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Writing (recording) on the disk is accomplished by applying an external magnetic field opposite to the direction of the magnetic particles as indicated in Figure 7–27(b) and then directing a high-power laser beam to heat the disk at a precise spot where a binary 1 is to be stored. The disk material, a magneto-optic alloy, is highly resistant to magnetization at room temperature; but at the spot where the laser beam heats the material, the inherent direction of magnetism is reversed by the external magnetic field produced by the electromagnet. At points where binary 0s are to be stored, the laser beam is not applied and the inherent upward direction of the magnetic particle remains.

As illustrated in Figure 7–27(c), reading data from the disk is accomplished by turning off the external magnetic field and directing a low-power laser beam at a spot where a bit is to be read. Basically, if a binary 1 is stored at the spot (reversed magnetization), the low-power laser beam is reflected and its polarization is shifted; but if a binary 0 is stored, the polarization of the reflected laser beam is unchanged. A detector senses the difference in the polarity of the reflected laser beam to determine if the bit being read is a 1 or a 0.

Figure 7–27(d) shows that the disk is erased by restoring the original magnetic direction of each particle by reversing the external magnetic field and applying the high-power laser beam.

SECTION 7–3 CHECKUP

1. Explain the difference between a solenoid and a relay.
2. What is the movable part of a solenoid called?
3. What is the movable part of a relay called?
4. Upon what basic principle is the d'Arsonval meter movement based?

7–4 MAGNETIC HYSTERESIS

When a magnetizing force is applied to a material, the magnetic flux density in the material changes in a certain way.

After completing this section, you should be able to

- **Explain magnetic hysteresis**
  - **State the formula for magnetic field intensity**
  - **Discuss a hysteresis curve**
  - **Define *retentivity***

Magnetic Field Intensity (H)

The **magnetic field intensity** (also called *magnetizing force*) in a material is defined to be the magnetomotive force (F_m) per unit length (l) of the material, as expressed by the following equation. The unit of magnetic field intensity (H) is ampere-turns per meter (At/m).

$$H = \frac{F_m}{l} \tag{7–6}$$

where $F_m = NI$. Note that the magnetic field intensity (H) depends on the number of turns (N) of the coil of wire, the current (I) through the coil, and the length (l) of the material. It does not depend on the type of material.

Since $\phi = F_m/\mathfrak{R}$, as F_m increases, the flux increases. Also, the magnetic field intensity (H) increases. Recall that the flux density (B) is the flux per unit cross-sectional area

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

($B = \phi/A$), so B is also proportional to H. The curve showing how these two quantities (B and H) are related is called the B-H curve, or the hysteresis curve. The parameters that influence both B and H are illustrated in Figure 7–28.

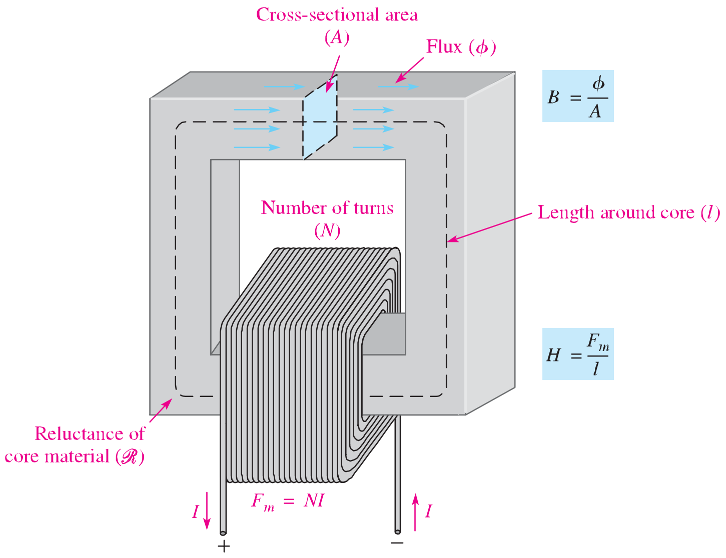

FIGURE 7–28 **Parameters that determine the magnetic field intensity (H) and the flux density (B).**

The Hysteresis Curve and Retentivity

Hysteresis is a characteristic of a magnetic material whereby a change in magnetization lags the application of the magnetic field intensity. The magnetic field intensity (H) can be readily increased or decreased by varying the current through the coil of wire, and it can be reversed by reversing the voltage polarity across the coil.

Figure 7–29 illustrates the development of the hysteresis curve. Let's start by assuming a magnetic core is unmagnetized so that $B = 0$. As the magnetic field intensity (H) is increased from zero, the flux density (B) increases proportionally as indicated by the curve in Figure 7–29(a). When H reaches a certain value, the value of B begins to level off. As H continues to increase, B reaches a saturation value (B_{sat}) when H reaches a value (H_{sat}), as illustrated in Figure 7–29(b). Once saturation is reached, a further increase in H will not increase B.

Now, if H is decreased to zero, B will fall back along a different path to a residual value (B_R), as shown in Figure 7–29(c). This indicates that the material continues to be magnetized even when the magnetic field intensity is removed ($H = 0$). The ability of a material, once magnetized, to maintain a magnetized state without magnetic field intensity is called **retentivity**. The retentivity of a material is indicated by the ratio of B_R to B_{sat}.

Reversal of the magnetic field intensity is represented by negative values of H on the curve and is achieved by reversing the current in the coil of wire. An increase in H in the negative direction causes saturation to occur at a value ($-H_{\text{sat}}$) where the flux density is at its maximum negative value, as indicated in Figure 7–29(d).

When the magnetic field intensity is removed ($H = 0$), the flux density goes to its negative residual value ($-B_R$), as shown in Figure 7–29(e). From the $-B_R$ value, the flux density follows the curve indicated in part (f) back to its maximum positive value when the magnetic field intensity equals H_{sat} in the positive direction.

The complete B-H curve is shown in Figure 7–29(g) and is called the *hysteresis curve*. The magnetic field intensity required to make the flux density zero is called the *coercive force*, H_C.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

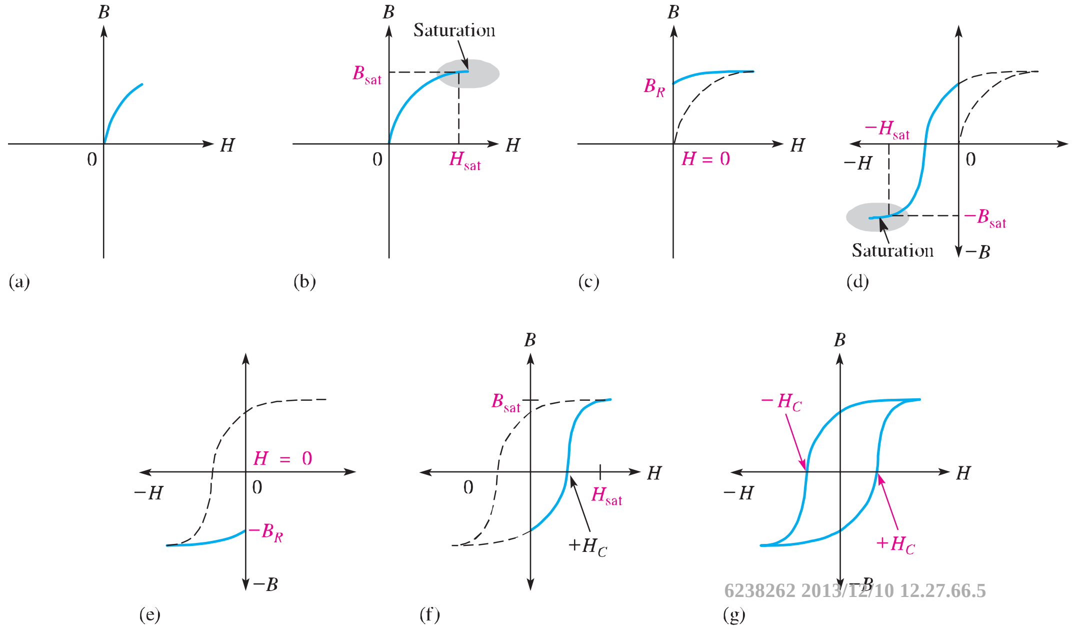

6238262 2013/12/10 12.27.66.5

FIGURE 7–29 **Development of a magnetic hysteresis (*B-H*) curve.**

Materials with a low retentivity do not retain a magnetic field very well while those with high retentivities exhibit values of B_R very close to the saturation value of B. Depending on the application, retentivity in a magnetic material can be an advantage or a disadvantage. In permanent magnets and memory cores, for example, high retentivity is required. In ac motors, retentivity is undesirable because the residual magnetic field must be overcome each time the current reverses, thus wasting energy.

SECTION 7–4 CHECKUP

1. For a given wirewound core, how does an increase in current through the coil affect the flux density?
2. Define *retentivity.*

7–5 ELECTROMAGNETIC INDUCTION

In this section, you are introduced to electromagnetic induction. Electromagnetic induction is what makes transformers, electrical generators, electrical motors, and many other devices possible.

After completing this section, you should be able to

- **Discuss the principle of electromagnetic induction**
  - **Explain how voltage is induced in a conductor in a magnetic field**
  - **Determine polarity of an induced voltage**
  - **Discuss forces on a conductor in a magnetic field**
  - **State Faraday's law**
  - **State Lenz's law**
  - **Explain how a crankshaft position sensor works**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Relative Motion

When a straight conductor is moved perpendicular to a magnetic field, there is a relative motion between the conductor and the magnetic field. Likewise, when a magnetic field is moved past a stationary conductor, there is also relative motion. In either case, this relative motion results in an **induced voltage (v_{ind})** across the conductor, as Figure 7–30 indicates. This principle is known as **electromagnetic induction**. The lowercase v stands for instantaneous voltage. The amount of the induced voltage depends on the rate at which the conductor and the magnetic field move with respect to each other: The faster the relative motion, the greater the induced voltage.

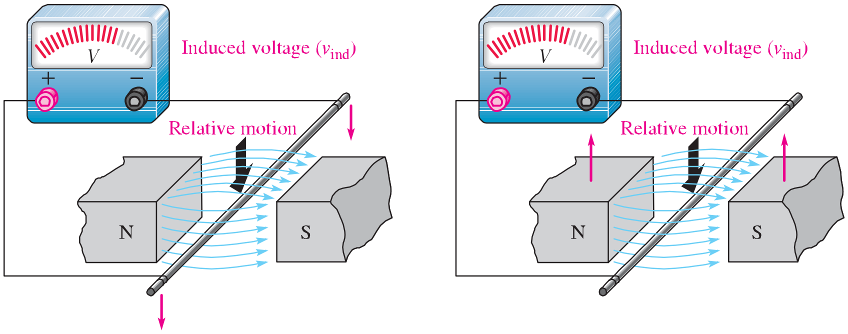

(a) Conductor moving downward

(b) Magnetic field moving upward

FIGURE 7–30 **Relative motion between a straight conductor and a magnetic field.**

Vibrating Wire Strain Gage

A strain gage that works on the principle of magnetic induction is the vibrating wire strain gage. It consists of a strong magnetic wire stretched between two blocks and encased in a tube. A sensor in the tube has a permanent magnet and a plucking coil.

When the wire is plucked by the sensor, it vibrates at a natural frequency. The frequency is picked up by the coil and is sent to instruments that record it. When the assembly is subject to strain, the wire is stretched, changing its frequency. Thus, the frequency recorded is related to strain. The gage is well suited to systems that have wet and damp environments because it is a sealed unit.

SYSTEM NOTE

Polarity of the Induced Voltage

If the conductor in Figure 7–30 is moved first one way and then another in the magnetic field, a reversal of the polarity of the induced voltage will be observed. When the relative motion of the conductor is downward, a voltage is induced with the polarity indicated in Figure 7–31(a). When the relative motion of the conductor is upward, the polarity is as indicated in part (b) of the figure.

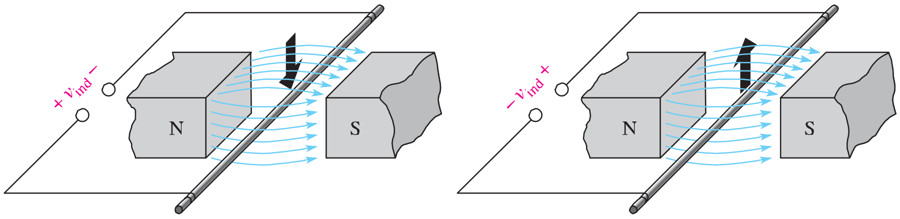

(a) Downward relative motion

(b) Upward relative motion

FIGURE 7–31 **Polarity of induced voltage depends on direction of motion of the conductor relative to the magnetic field.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

When a straight conductor moves perpendicular to a constant magnetic field, the induced voltage is given by

$$v_{\text{ind}} = B_{\perp}lv \tag{7–7}$$

where v_{ind} is the induced voltage in volts, $B_{\perp}$ is the component of the magnetic flux density that is perpendicular to the moving conductor (in teslas), l is the length of the conductor that is exposed to the magnetic field, and v is the velocity of the conductor in m/s.

EXAMPLE 7–7

Assume the conductor in Figure 7–31 is 10 cm long and the pole face of the magnet is 5.0 cm wide. The flux density is 0.5 T, and the conductor is moved upward at a velocity of 0.8 m/s. What voltage is induced in the conductor?

SOLUTION

Although the conductor is 10 cm, only 5.0 cm (0.05 m) is in the magnetic field because of the size of the pole faces. Therefore,

$$v_{\text{ind}} = B_{\perp}lv = (0.5\text{ T})(0.05\text{ m})(0.8\text{ m/s}) = \mathbf{20\text{ mV}}$$

RELATED PROBLEM

What is the induced voltage if the velocity is doubled?

Induced Current

When a load resistor is connected to the conductor in Figure 7–31, the voltage induced by the relative motion between the conductor and the magnetic field will cause a current in the load, as shown in Figure 7–32. This current is called the **induced current** **(i_{ind})**. The lowercase i stands for instantaneous current.

The action of producing a voltage and a current in a load by moving a conductor across a magnetic field is the basis for electrical generators. A single conductor will have a small induced current, so practical generators use coils with many turns. Also, the existence of a conductor in a moving magnetic field is fundamental to the concept of inductance in an electric circuit.

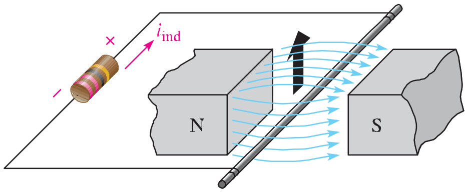

FIGURE 7–32 Induced current (i_{ind}) in a load as the conductor moves through the magnetic field.

Faraday's Law

Michael Faraday discovered the principle of electromagnetic induction in 1831. The key idea behind Faraday's law is that a changing magnetic field can induce a voltage in a conductor. Sometimes Faraday's law is stated as Faraday's law of induction. Faraday experimented with coils, and his law is an extension of the principle of electromagnetic induction for straight conductors discussed previously.

When a conductor is coiled into multiple turns, more conductors can be exposed to the magnetic field, increasing the induced voltage. When the flux is caused to change by any means, an induced voltage will result. The change in the magnetic field can be caused by relative motion between the magnetic field and the coil. Faraday's observations can be stated as follows:

1. The amount of voltage induced in a coil is directly proportional to the rate of change of the magnetic field with respect to the coil.
2. The amount of voltage induced in a coil is directly proportional to the number of turns of wire in the coil.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Faraday's first observation is demonstrated in Figure 7–33, where a bar magnet is moved through a coil, thus creating a changing magnetic field. In part (a) of the figure, the magnet is moved at a certain rate, and a certain induced voltage is produced as indicated. In part (b), the magnet is moved at a faster rate through the coil, creating a greater induced voltage.

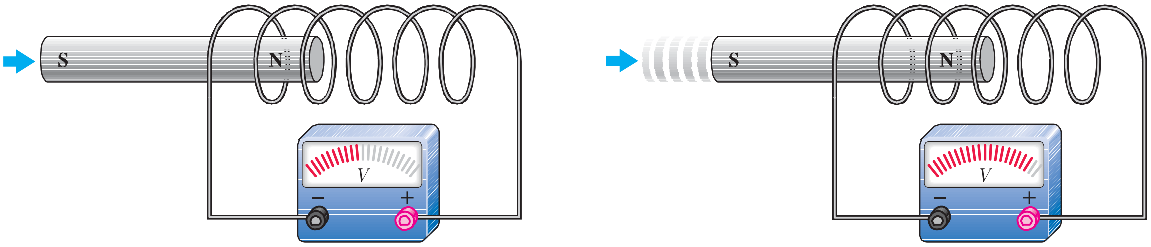

(a) As the magnet moves slowly to the right, its magnetic field is changing with respect to the coil, and a voltage is induced.

(b) As the magnet moves more rapidly to the right, its magnetic field is changing more rapidly with respect to the coil, and a greater voltage is induced.

FIGURE 7–33 **A demonstration of Faraday's first observation: The amount of induced voltage is directly proportional to the rate of change of the magnetic field with respect to the coil.**

6238262 2013/12/10 12.27.66.5

Faraday's second observation is demonstrated in Figure 7–34. In part (a), the magnet is moved through the coil and a voltage is induced as shown. In part (b), the magnet is moved at the same rate through a coil that has a greater number of turns. The greater number of turns creates a greater induced voltage.

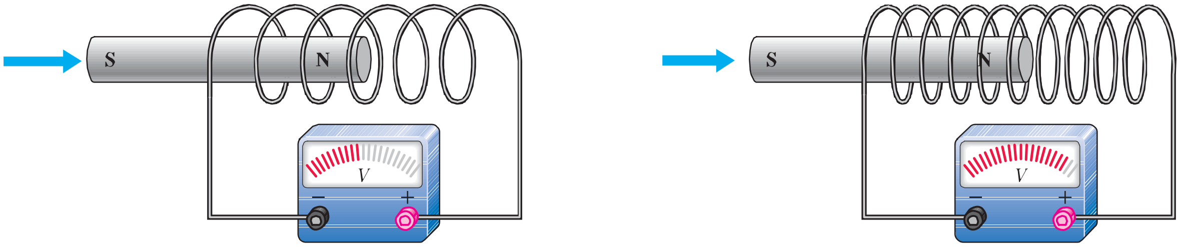

(a) Magnet moves through a coil and induces a voltage.

(b) Magnet moves at same rate through a coil with more turns (loops) and induces a greater voltage.

FIGURE 7–34 **A demonstration of Faraday's second observation: The amount of induced voltage is directly proportional to the number of turns in the coil.**

Faraday's law is stated as follows:

> **The voltage induced across a coil equals the number of turns in the coil times the rate of change of the magnetic flux.**

Any *relative* motion between the magnetic field and the magnet will produce the changing magnetic field that will induce a voltage in the coil. The changing magnetic field can even be induced by ac applied to an electromagnet, just as if there were motion. This type of changing magnetic field is the basis for transformer action in ac circuits, as you will see in Chapter 14.

Lenz's Law

You have learned that a changing magnetic field induces a voltage in a coil that is directly proportional to the rate of change of the magnetic field and the number of turns in the coil. **Lenz's law** defines the polarity or direction of the induced voltage.

> **When the current through a coil changes, the polarity of the induced voltage created by the changing magnetic field is such that it always opposes the change in current that caused it.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

An Application of Electromagnetic Induction

Automotive control systems include a large number of sensors. For optimum engine performance, it is necessary to know the position and/or speed of the crankshaft to control engine ignition timing, fuel mixture, tachometer, and antilock brakes. The most common method for sensing the crankshaft (or camshaft) position is with a Hall-effect sensor, which was discussed in Section 7–1.

Another widely used method is to detect a change in the magnetic field as a metallic tab moves though the air gap in a magnetic assembly. A basic concept is shown in Figure 7–35. A steel disk with protruding tabs is connected to the end of the crankshaft. As the crankshaft turns, the tabs move through the magnetic field. Steel has a much lower reluctance than does air, so the magnetic flux increases when a tab is in the air gap. This change in the magnetic flux causes an induced voltage to appear across the coil, indicating the position of the crankshaft.

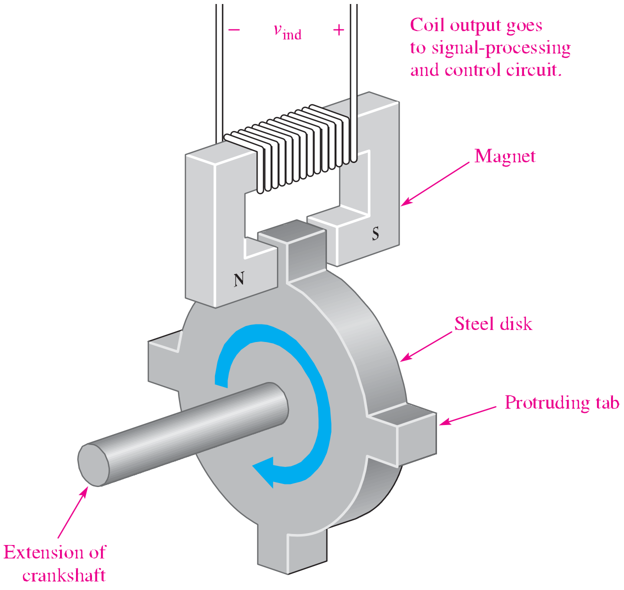

FIGURE 7–35 **A crankshaft position sensor that produces a voltage when a tab passes through the air gap of the magnet.**

Forces on a Current-Carrying Conductor in a Magnetic Field (Motor Action)

Figure 7–36(a) shows current inward through a wire in a magnetic field. The electromagnetic field set up by the current interacts with the permanent magnetic field; as a result, the permanent lines of force above the wire tend to be deflected down under the wire because

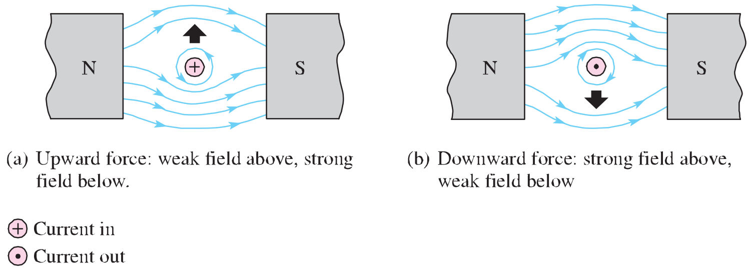

(a) Upward force: weak field above, strong field below.

(b) Downward force: strong field above, weak field below

FIGURE 7–36 **Forces on a current-carrying conductor in a magnetic field.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

they are opposite in direction to the electromagnetic lines of force. Therefore, the flux density above is reduced, and the magnetic field is weakened. The flux density below the conductor is increased, and the magnetic field is strengthened. An upward force on the conductor results, and the conductor tends to move toward the weaker magnetic field. Figure 7–36(b) shows the current outward, resulting in a force on the conductor in the downward direction. These upward and downward forces on a conductor are the basis for electric motors.

The force on a current-carrying conductor is given by the equation

$$F = BIl \qquad (7\text{–}8)$$

where F is the force in newtons, B is the magnetic flux density in teslas, I is the current in amperes, and l is the length of the conductor exposed to the magnetic field in meters.

EXAMPLE 7–8

Assume a magnetic pole face is a square that is 3.0 cm on a side. Find the force on a conductor that has a current of 2 A if the conductor is perpendicular to the field and the flux density is 0.35 T.

SOLUTION

The length of conductor exposed to the magnetic flux is 3.0 cm (0.030 m). Therefore,

$$F = BIl = (0.35\text{ T})(2.0\text{ A})(0.03\text{ m}) = \mathbf{0.21\text{ N}}$$

RELATED PROBLEM

What is the direction of the force if the field is directed up (along the y axis), and the current (electron flow) is directed inward (along the z axis)?

SECTION 7–5 CHECKUP

1. What is the induced voltage across a stationary conductor in a stationary magnetic field?
2. When the rate at which a conductor is moved through a magnetic field is increased, does the induced voltage increase, decrease, or remain the same?
3. When there is current through a conductor in a magnetic field, what happens?
4. If the steel disk in the crankshaft position sensor has stopped with a tab in the magnet's air gap, what is the induced voltage?

7–6 DC GENERATORS

DC generators produce a voltage that is proportional to the magnetic flux and the rotational speed of the armature.

After completing this section, you should be able to

- **Explain how a dc generator works**
  - **Draw an equivalent circuit for a self-excited shunt-type dc generator**
  - **Discuss the parts of a dc generator**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Figure 7–37 shows a greatly simplified dc generator consisting of a single loop of wire that rotates in a magnetic field. Notice that each end of the loop is connected to a split-ring arrangement. This conductive metal ring is called a *commutator.* As the wire loop rotates in the magnetic field, the split commutator ring also rotates. Each half of the split ring rubs against the fixed contacts, called *brushes,* and connects the wire to an external circuit.

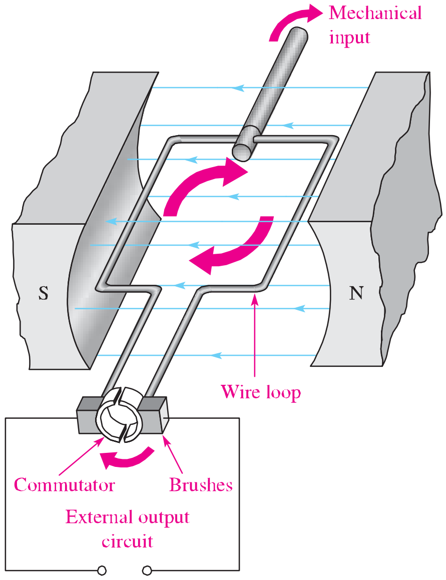

FIGURE 7–37 **A simplified dc generator.**

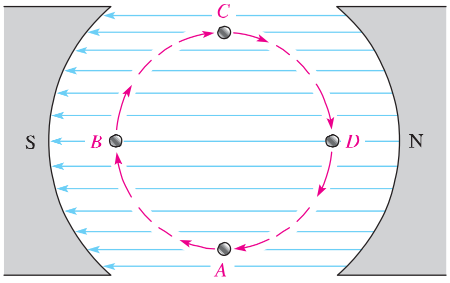

FIGURE 7–38 **End view of loop of wire cutting through the magnetic field.**

Driven by an external mechanical force, the loop of wire rotates through the magnetic field and cuts through the flux lines at varying angles, as illustrated in Figure 7–38. At position *A* in its rotation, the loop is effectively moving parallel with the magnetic field. Therefore, at this instant, the rate at which it is cutting through the magnetic flux lines is zero. As the loop moves from position *A* to position *B*, it cuts through the flux lines at an increasing rate. At position *B*, it is moving effectively perpendicular to the magnetic field and thus is cutting through a maximum number of lines. As the loop rotates from position *B* to position *C*, the rate at which it cuts the flux lines decreases to minimum (zero) at *C*. From position *C* to position *D*, the rate at which the loop cuts the flux lines increases to a maximum at *D* and then back to a minimum again at *A*.

6238262 2013/12/10 12.27.66.5

As you have learned, when a wire moves through a magnetic field, a voltage is induced, and by Faraday's law, the amount of induced voltage is proportional to the number of loops (turns) in the wire and the rate at which it is moving with respect to the magnetic field. The angle at which the wire moves with respect to the magnetic flux lines determines the amount of induced voltage because the rate at which the wire cuts through the flux lines depends on the angle of motion.

Figure 7–39 illustrates how a voltage is induced in the external circuit as a single loop of wire rotates in the magnetic field. Assume that the loop is in its instantaneous horizontal position, so the induced voltage is zero. As the loop continues in its rotation, the induced voltage builds up to a maximum at position *B*, as shown in part (a) of the figure. Then, as the loop continues from *B* to *C*, the voltage decreases to zero at *C*, as shown in part (b).

During the second half of the revolution, shown in Figure 7–39(c) and (d), the brushes switch to opposite commutator sections, so the polarity of the voltage remains the same across the output. Thus, as the loop rotates from position *C* to *D* and then back to *A*, the voltage increases from zero at *C* to a maximum at *D* and back to zero at *A*.

Figure 7–40 shows how the induced voltage varies as a wire loop in a dc generator goes through several rotations (three in this case). This voltage is a dc voltage because its polarity does not change. However, the voltage is pulsating between zero and its maximum value.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

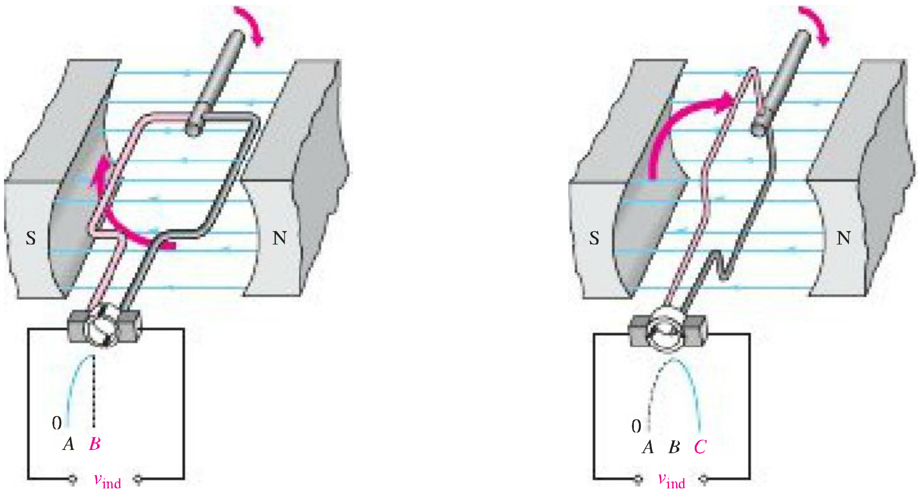

Position *B*: Loop is moving perpendicular to flux lines, and voltage is maximum.

Position *C*: Loop is moving parallel with flux lines, and voltage is zero.

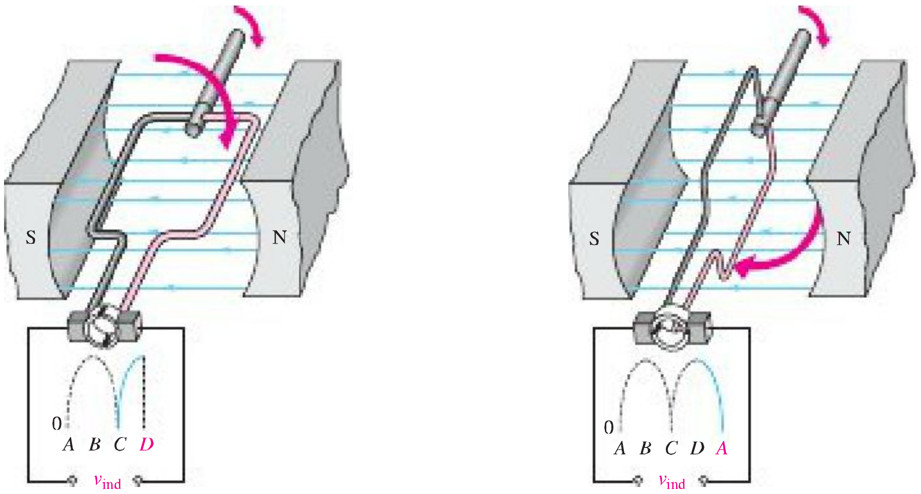

Position *D*: Loop is moving perpendicular to flux lines, and voltage is maximum.

Position *A*: Loop is moving parallel with flux lines, and voltage is zero.

FIGURE 7–39 **Basic operation of a dc generator.**

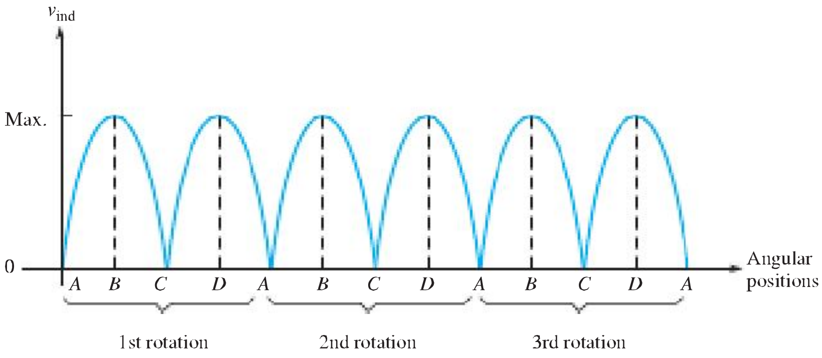

FIGURE 7–40 **Induced voltage over three rotations of a wire loop in a dc generator.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

In practical generators, multiple coils are pressed into slots in a ferromagnetic-core assembly. The entire assembly, called the **rotor**, is connected to bearings and rotates in the magnetic field. Figure 7–41 is a diagram of the rotor core with no wire loops (coils) shown. The commutator is divided into segments, with each pair of segments connected to the end of a coil. With more coils, the voltages from several coils are combined because the brushes can contact more than one of the commutator segments at once. The loops do not reach maximum voltage at the same time, so the pulsating output voltage is much smoother than is the case with only one coil or loop shown previously. The variations can be further smoothed by filters to produce a nearly constant dc output. (Filters are discussed in Chapter 13.)

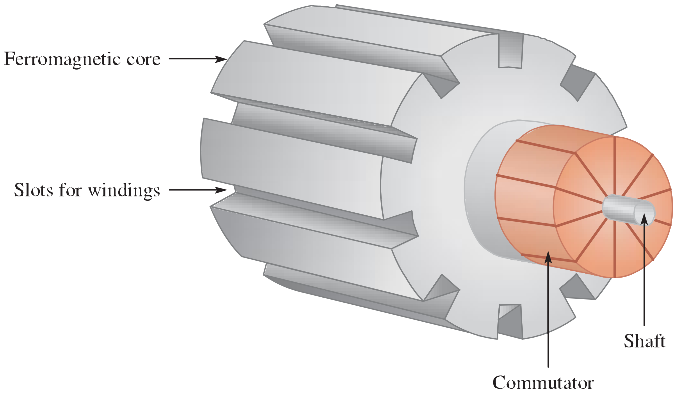

FIGURE 7–41 **A simplified rotor core. The coils are pressed into the slots and connected to the commutator.**

Instead of permanent magnets, most generators use electromagnets to provide the required magnetic field. One advantage to this is that flux density can be controlled, thus controlling the output voltage of the generator. The windings for the electromagnets are appropriately called the field windings. The field windings require current to produce the magnetic field.

The current for the field windings can be provided from a separate voltage source, but this is a disadvantage. A better method is to use the generator itself to provide the current for the electromagnets; this is called a **self-excited generator**. The generator starts because there is normally enough residual magnetism in the field magnets due to hysteresis that causes a small initial field and allows the generator to start producing a voltage. In cases where a generator has not been used for a long time, it may be necessary to provide an external source to the field windings to start it.

The stationary part of a generator (or a motor) includes all of the nonmoving parts and is called the **stator**. Figure 7–42 illustrates a simplified two-pole dc generator showing

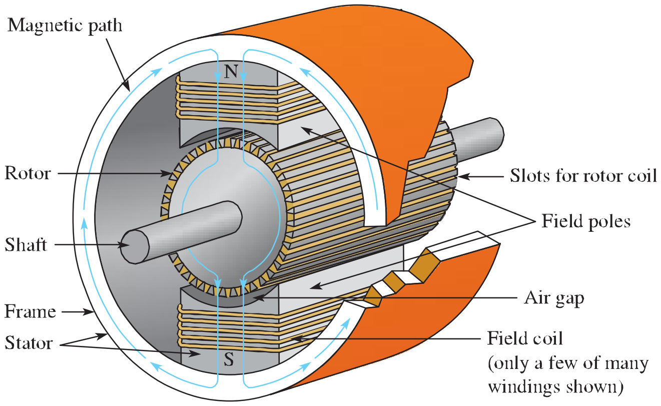

FIGURE 7–42 **The magnetic structure of a generator (or motor). In this case, the rotor is also the armature because it produces the power.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

the magnetic paths (end caps, bearings, and commutator are not shown). Notice that the frame is part of the magnetic path for the field magnets. To make a generator have high efficiency, the air gap is kept as small as possible. The **armature** is the power-producing component and can be either on the rotor or the stator. In the dc generator described previously, the armature is the rotor because power is produced in the moving conductors and taken from the rotor via the commutator.

Equivalent Circuit for a DC Generator

A self-excited generator can be represented by a basic dc circuit with a coil to produce the magnetic field and a mechanically driven generator, as shown in Figure 7–43. There are other configurations of dc generators, but this represents a common one. In the case shown, the field windings are in parallel with the source; this configuration is called a *shunt-wound generator*. The resistance of the field windings is shown as R_F. For the equivalent circuit, this resistance is shown in series with the field windings. The armature is driven by a mechanical input, causing it to spin; the armature is the generator voltage source, V_G. The armature resistance is shown as the series resistance, R_A. The rheostat, R_{REG}, is in series with the field winding resistance and regulates the output voltage by controlling the current to the field windings and thus the flux density.

When a load is connected to the output, current in the armature is shared between the load and the field windings. The efficiency of the generator can be calculated as the ratio of the power delivered to the load (P_L) to the total power (P_T), which includes the losses in the armature and the resistance in the field circuit.

6238262 2013/12/10 12.27.66.5

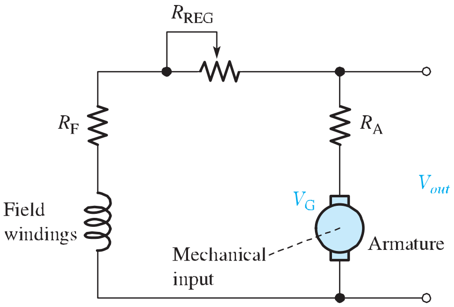

FIGURE 7–43 **Equivalent circuit for a self-excited shunt dc generator.**

SECTION 7–6 CHECKUP

1. What is the moving part of a generator called?
2. What is the purpose of a commutator?
3. How does greater resistance in the field windings of a generator affect the output voltage?
4. What is meant by a self-excited generator?

7–7 DC MOTORS

Motors convert electrical energy to mechanical motion by taking advantage of the force produced when a current-carrying conductor is in a magnetic field. A dc motor operates from a dc source and can use either an electromagnet or a permanent magnet to supply the field.

After completing this section, you should be able to

- **Explain how a dc motor works**
  - **Draw an equivalent circuit for a series and a shunt-type dc motor**
  - **Discuss back emf and how it reduces armature current**
  - **Discuss power rating of motors**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Basic Operation

As in the case of generators, motor action is the result of the interaction of magnetic fields. In a dc motor, the rotor field interacts with the magnetic field set up by current in the stator windings. The rotor in all dc motors contains the armature winding, which sets up a magnetic field. The rotor moves because of the attractive force between opposite poles and the repulsive force between like poles, as illustrated in the simplified diagram of Figure 7–44. The rotor moves because of the attraction of its north pole with the south pole of the stator (and vice-versa). As the two poles near each other, the polarity of the rotor current is suddenly switched by the commutator, reversing the magnetic poles of the rotor. The commutator serves as a mechanical switch to reverse the current in the armature just as the unlike poles are near each other, continuing the rotation of the rotor.

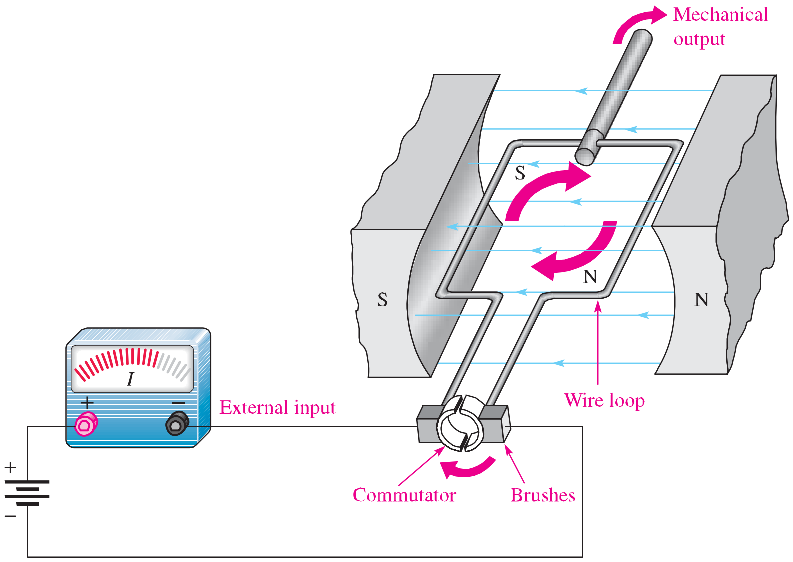

FIGURE 7–44 **Simplified dc motor.**

Motor Maintenance

Industrial systems typically have many motors to perform functions such as pumps, machine operations, conveyers, and robotic control. A motor that is running too hot may be indicating that there is lubrication or bearing issues that need to be addressed. A thermal image can reveal such problems, so in many companies thermal imaging is part of a routine inspection and maintenance program.

Courtesy of Fluke Corporation.

SYSTEM NOTE

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Brushless DC Motors

Many dc motors do not use a commutator to reverse the polarity of the current. Instead of supplying current to a moving armature, the magnetic field is rotated in the stator windings using an electronic controller. The direction of current in the field coils is periodically reversed by the controller by producing an ac waveform (or modified ac waveform) from the dc input. This causes the stator field to rotate, and the permanent magnet rotor moves in the same direction to keep up with the rotating field. A common way to sense the position of the rotating magnet is to use a Hall-effect sensor, which provides the controller with position information by sending pulses every time a magnet passes it. Brushless motors have higher reliability than traditional brushed motors because they do not need to have periodic brush replacement, but have the added complexity of the electronic controller. Figure 7–45 shows a cutaway view of a brushless dc motor that includes a pulse width modulation control within the motor housing as well as an optical encoder to indicate the shaft position.

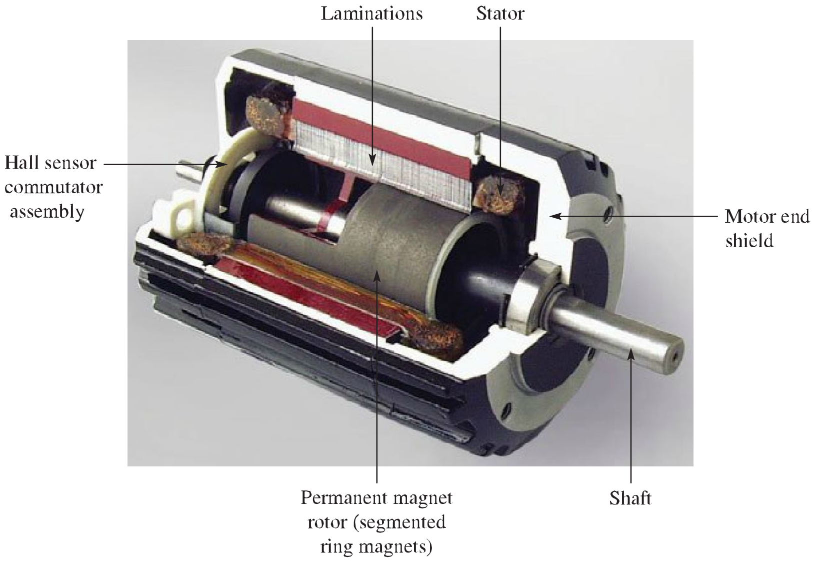
Laminations
Stator
Hall sensor commutator assembly
Motor end shield
Permanent magnet rotor (segmented ring magnets)
Shaft

FIGURE 7–45 **Cutaway view of a brushless dc motor.** **(Courtesy of Bodine Electric Company)**

Back EMF

When a dc motor is first started, a magnetic field is present from the field windings. Armature current develops another magnetic field that interacts with the one from the field windings and starts the motor turning. The armature windings are now spinning in the presence of the magnetic field, so generator action occurs. In effect, the spinning armature has a voltage developed across it that opposes the original applied voltage in accordance with Lenz's law. This self-generated voltage is called **back emf** (electromotive force). The term *emf* was once common for voltage but is not favored because voltage is not a "force" in the physics sense, but back emf is still applied to the self-generated voltage in motors. Back emf, also called *counter emf,* serves to significantly reduce the armature current when the motor is turning at constant speed.

HANDS ON TIP
One characteristic of dc motors is that if they are allowed to run without a load, the torque can cause the motor to "run away" to a speed beyond the manufacturer's rating. Therefore, dc motors should always be operated with a load to prevent self-destruction.

Motor Ratings

Some motors are rated by the torque they can provide; others are rated by the power they produce. Torque and power are important parameters for any motor. Although torque and power are different physical parameters, if one is known, the other can be obtained.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Torque tends to rotate an object. In a dc motor, the torque is proportional to the amount of flux and to the armature current. Torque, *T*, in a dc motor can be calculated from the equation

6238262 2013/12/10 12.27.66.5

$$T = k\phi I_A \qquad (7\text{–}9)$$

where *T* is torque in newton-meters (N-m) *k* is a constant that depends on physical parameters of the motor, ϕ is magnetic flux in webers (Wb), and I_A is armature current in amperes (A).

Recall that power is defined as the rate of doing work. To calculate power from torque, you must know the speed of the motor in rpm for the torque that you measured. The equation to determine the power, given the torque at a certain speed, is

$$P = 0.105Ts \qquad (7\text{–}10)$$

where *P* is power in W, *T* is torque in N-m, and *s* is speed of motor in rpm.

EXAMPLE 7–9

What is the power developed by a motor that turns at 350 rpm when the torque is 3.6 N-m?

SOLUTION

Substitute into Equation 7–10.

$$P = 0.105Ts = 0.105(3.6\text{ N-m})(350\text{ rpm}) = \mathbf{132\ W}$$

RELATED PROBLEM

There are 746 W in one hp. What is the hp of this motor under these conditions?

SYSTEM EXAMPLE 7–2

DC MOTOR REVERSING

DC motors have many applications in systems ranging from small portable drills to large cranes. In many systems that use a dc motor, the motor needs to have the ability to be reversed. For example, a motor that operates a garage door needs to be reversed to open or close the door. To reverse a dc motor, you need to be able to connect separate power leads to the field (stator) and the armature (rotor) and switch the polarity on *either* the rotor or the stator. Only one of the windings is reversed; if both were reversed, the motor would turn in the same direction.

In some cases, the forward or reverse direction may be controlled by a logic operation. In the system described here, a relay is used to interface logic that is used to specify the direction with the motor-reversing switches, which are the relay contacts. A simplified reversing circuit is shown in Figure 7–46. In addition to its role in reversing the motor, the relay provides a second benefit. It completely isolates the low voltage logic circuits from the motor circuit.

The transistor (Q_1) acts as a switch to interface the logic to the relay coil. When the logic is LOW, the transistor is *off*, and the relay is in its unenergized position. When the logic is HIGH, the transistor conducts and activates the relay. The logic circuits are isolated from the motor power by the relay. The purpose of the diode is to avoid noise spikes when the relay is deenergized. The On/Off switch is controlled by the logic to remove power from the motor before changing its direction.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

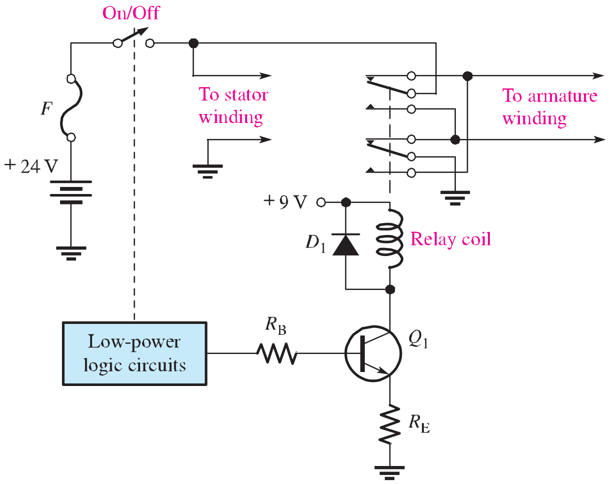

FIGURE 7–46 **DC motor reversing circuit using a relay.**

Series DC Motors

The series dc motor has the field coil windings and the armature coil windings in series. A schematic of this arrangement is shown in Figure 7–47(a). The internal resistance is generally small and consists of field coil resistance, armature winding resistance, and brush resistance. As in the case of generators, dc motors may also contain an interpole winding, as shown, and current limiting for speed control. In a series dc motor, the armature current, field current, and line current are all the same.

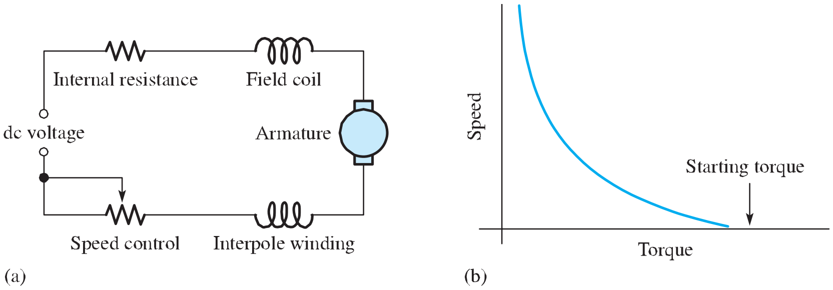

FIGURE 7–47 **Simplified schematic for a series dc motor and torque-speed characteristic.**

As you know, magnetic flux is proportional to the current in a coil. The magnetic flux created by the field windings is proportional to armature current because of the series connection. Thus, when the motor is loaded, armature current rises, and the magnetic flux rises too. Recall that Equation 7–9 showed that the torque in a dc motor is proportional to both the armature current and the magnetic flux. Thus, the series-wound motor will have a very high starting torque when the current is high because flux and armature current are high. For this reason, series dc motors are used when high starting torques are required (such as a starter motor in a car).

The plot of the torque and motor speed for a series dc motor is shown in Figure 7–47(b). The starting torque is at its maximum value. At low speeds, the torque is still very high, but drops off dramatically as the speed increases. As you can see, the speed can be very high if the torque is low; for this reason, the series-wound dc motor always is operated with a load.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Shunt DC Motors

A shunt dc motor has the field coil in parallel with the armature, as shown in the equivalent circuit in Figure 7–48(a). In the shunt motor, the field coil is supplied by a constant voltage source, so the magnetic field set up by the field coils is constant. The armature resistance and the back emf produced by generator action in the armature determine the armature current.

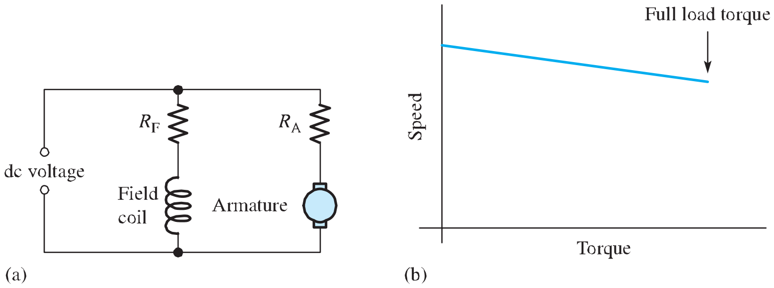

FIGURE 7–48 **Simplified schematic for a shunt dc motor and torque-speed characteristic.**

The torque-speed characteristic for a shunt dc motor is quite different than for a series dc motor. When a load is applied, the shunt motor will slow down, causing the back emf to be reduced and the armature current to increase. The increase in armature current tends to compensate for the added load by increasing the torque of the motor. Although the motor has slowed because of the additional load, the torque-speed characteristic is nearly a straight line for the shunt dc motor as shown in Figure 7–48(b). At full load, the shunt dc motor still has high torque.

SECTION 7–7 CHECKUP

1. What creates back emf?
2. How does back emf affect current in a rotating armature as it comes up to speed?
3. What type of dc motor has the highest starting torque?
4. What is the major advantage of a brushless motor over a brushed motor?

SUMMARY

- Unlike magnetic poles attract each other, and like poles repel each other.
- Materials that can be magnetized are called *ferromagnetic*.
- When there is current through a conductor, it produces an electromagnetic field around the conductor.
- You can use the left-hand rule to establish the direction of the electromagnetic lines of force around a conductor.
- An electromagnet is basically a coil of wire around a magnetic core.
- When a conductor moves within a magnetic field, or when a magnetic field moves relative to a conductor, a voltage is induced across the conductor.
- The faster the relative motion between a conductor and a magnetic field, the greater is the induced voltage.
- Table 7–3 summarizes the quantities and SI units used in this chapter.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

TABLE 7–3

SYMBOL	QUANTITY	SI UNIT
B	Magnetic flux density	Tesla (T)
ϕ	Flux	Weber (Wb)
μ	Permeability	Webers/ampere-turn · meter (Wb/At · m)
$\mathcal{R}$	Reluctance	Ampere-turns/weber (At/Wb)
F_m	Magnetomotive force (mmf)	Ampere-turn (At)
H	Magnetic field intensity	Ampere-turns/meter (At/m)
F	Force	Newton (N)
T	Torque	Newton-meter (N-m)

6238262-2013/12/10 12.27.66.5

- Hall-effect sensors use current to sense the presence of a magnetic field.
- DC generators convert mechanical power to dc electrical power.
- The moving part of a generator or motor is called the rotor; the stationary part is called the stator.
- DC motors convert electrical power to mechanical power.
- Brushless dc motors use a permanent magnet as the rotor, and the stator is the armature.

KEY TERMS

Key terms and other bold terms in the chapter are defined in the end-of-book glossary.

Ampere-turn (At) The SI unit of magnetomotive force (mmf).

Electromagnetic field A formation of a group of magnetic lines of force surrounding a conductor created by electrical current in the conductor.

Electromagnetic induction The phenomenon or process by which a voltage is produced in a conductor when there is relative motion between the conductor and a magnetic or electromagnetic field.

Electromagnetism The production of a magnetic field by current in a conductor.

Faraday's law A law stating that the voltage induced across a coil equals the number of turns in the coil times the rate of change of the magnetic flux.

Gauss A CGS unit of flux density.

Hall effect A change in current density across a conductor or semiconductor when current in the material is perpendicular to a magnetic field. The change in current density produces a small transverse voltage in the material, called the *Hall voltage*.

Hysteresis A characteristic of a magnetic material whereby a change in magnetization lags the application of the magnetic field intensity.

Induced current (i_{ind}) A current induced in a conductor as a result of a changing magnetic field.

Induced voltage (v_{ind}) Voltage produced as a result of a changing magnetic field.

Lenz's law A physical law that states when the current through a coil changes, the polarity of the induced voltage created by the changing magnetic field is such that it always opposes the change in current that caused it. The current cannot change instantaneously.

Lines of force Magnetic flux lines in a magnetic field radiating from the north pole to the south pole.

Magnetic field A force field radiating from the north pole to the south pole of a magnet.

Magnetic flux The lines of force between the north and south poles of a permanent magnet or an electromagnet.

Magnetic field intensity The amount of mmf per unit length of magnetic material.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Magnetomotive force (mmf) The cause of a magnetic field, measured in ampere-turns.

Permeability The measure of ease with which a magnetic field can be established in a material.

Relay An electromagnetically controlled mechanical device in which electrical contacts are opened or closed by a magnetizing current.

Reluctance ($\mathfrak{R}$) The opposition to the establishment of a magnetic field in a material.

Retentivity The ability of a material, once magnetized, to maintain a magnetized state without the presence of a magnetizing force.

Solenoid An electromagnetically controlled device in which the mechanical movement of a shaft or plunger is activated by a magnetizing current.

Speaker An electromagnetic device that converts electrical signals to sound waves.

Tesla (T) The SI unit of flux density.

Weber (Wb) The SI unit of magnetic flux, which represents 10^8 lines.

KEY FORMULAS

(7–1)	$B = \dfrac{\phi}{A}$	Magnetic flux density
(7–2)	$\mu_r = \dfrac{\mu}{\mu_0}$	Relative permeability
(7–3)	$\mathfrak{R} = \dfrac{l}{\mu A}$	Reluctance
(7–4)	$F_m = NI$	Magnetomotive force
(7–5)	$\phi = \dfrac{F_m}{\mathfrak{R}}$	Magnetic flux
(7–6)	$H = \dfrac{F_m}{l}$	Magnetic field intensity
(7–7)	$v_{\text{ind}} = B_{\perp}lv$	Induced voltage
(7–8)	$F = BIl$	Force on current-carrying conductor
(7–9)	$T = k\phi I_{\text{A}}$	Torque in a dc motor
(7–10)	$P = 0.105Ts$	Power from torque

TRUE/FALSE QUIZ

Answers are at the end of the chapter.

1. The tesla (T) and the gauss (G) are both units for magnetic flux density.
2. The unit for measuring magnetomotive force (mmf) is the volt.
3. Ohm's law for a magnetic circuit gives the relationship between flux density, magnetomotive force, and reluctance.
4. A solenoid is a form of electromagnetic switch that opens and closes mechanical contacts.
5. A hysteresis curve is a plot of flux density (B) as a function of field intensity (H).
6. To produce an induced voltage in a coil, the magnetic field surrounding it can be changed.
7. The speed of a generator can be controlled with a rheostat in the field windings.
8. A self-excited dc generator will normally have enough residual magnetism in the field magnets to start the generator, producing voltage at the output when it is first turned on.
9. The power developed by a motor is proportional to its torque.
10. In a brushless motor, the magnetic field is supplied by permanent magnets.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SELF-TEST

Answers are at the end of the chapter.

1. When the south poles of two bar magnets are brought close together, there will be

(a) a force of attraction **(b)** a force of repulsion
(c) an upward force **(d)** no force

2. A magnetic field is made up of

(a) positive and negative charges **(b)** magnetic domains
(c) flux lines **(d)** magnetic poles

3. The direction of a magnetic field is from

(a) north pole to south pole **(b)** south pole to north pole
(c) inside to outside the magnet **(d)** front to back

4. Reluctance in a magnetic circuit is analogous to

(a) voltage in an electric circuit **(b)** current in an electric circuit
(c) power in an electric circuit **(d)** resistance in an electric circuit

5. The unit of magnetic flux is the

(a) tesla **(b)** weber **(c)** gauss **(d)** ampere-turn

6. The unit of magnetomotive force is the

(a) tesla **(b)** weber **(c)** ampere-turn **(d)** electron-volt

7. A unit of magnetic flux density is the

(a) tesla **(b)** weber **(c)** ampere-turn **(d)** ampere-turns/meter

8. The electromagnetic activation of a movable shaft is the basis for

(a) relays **(b)** circuit breakers
(c) magnetic switches **(d)** solenoids

9. When there is current through a wire placed in a magnetic field,

(a) the wire will overheat **(b)** the wire will become magnetized
(c) a force is exerted on the wire **(d)** the magnetic field will be cancelled

10. A coil of wire is placed in a changing magnetic field. If the number of turns in the coil is increased, the voltage induced across the coil will

(a) remain unchanged **(b)** decrease
(c) increase **(d)** be excessive

11. If a conductor is moved back and forth at a constant rate in a constant magnetic field, the voltage induced in the conductor will

(a) remain constant **(b)** reverse polarity
(c) be reduced **(d)** be increased

12. In the crankshaft position sensor in Figure 7–35, the induced voltage across the coils is caused by

(a) current in the coil
(b) rotation of the disk
(c) a tab passing through the magnetic field
(d) acceleration of the disk's rotational speed

13. The purpose of the commutator in a generator or motor is to

(a) change the direction of the current to the rotor as it spins
(b) change the direction of the current to the stator windings
(c) support the motor or generator shaft
(d) provide the magnetic field for the motor or generator

14. In a motor, back emf serves to

(a) increase the power from the motor **(b)** decrease the flux
(c) increase the current in the field windings **(d)** decrease the current in the armature

15. The torque of a motor is proportional to the

(a) amount of flux **(b)** armature current
(c) both of the above **(d)** none of the above

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

PROBLEMS

Answers to odd-numbered problems are at the end of the book.

BASIC PROBLEMS

SECTION 7–1 The Magnetic Field

1. The cross-sectional area of a magnetic field is increased, but the flux remains the same. Does the flux density increase or decrease?

6238262 2013/12/10 12.27.66.5

2. In a certain magnetic field the cross-sectional area is 0.5 m^2 and the flux is 1500 μWb. What is the flux density?
3. What is the flux in a magnetic material when the flux density is 2500×10^{-6} T and the cross-sectional area is 150 cm^2?
4. At a given location, the earth's magnetic field is 0.6 G (gauss). Express this flux density in tesla.
5. A very strong permanent magnet has a field of 100,000 μT. Express this flux density in gauss.

SECTION 7–2 Electromagnetism

6. What happens to the compass needle in Figure 7–11 when the current through the conductor is reversed?
7. What is the relative permeability of a ferromagnetic material whose absolute permeability is 750×10^{-6} Wb/At · m?
8. Determine the reluctance of a material with a length of 0.28 m and a cross-sectional area of 0.08 m^2 if the absolute permeability is 150×10^{-7} Wb/At · m.
9. What is the magnetomotive force in a 500 turn coil of wire with 3 A through it?

SECTION 7–3 Electromagnetic Devices

10. Typically, when a solenoid is activated, is the plunger extended or retracted?
11. **(a)** What force moves the plunger when a solenoid is activated?
    (b) What force causes the plunger to return to its at-rest position?
12. Explain the sequence of events in the circuit of Figure 7–49 starting when switch 1 (SW1) is closed.

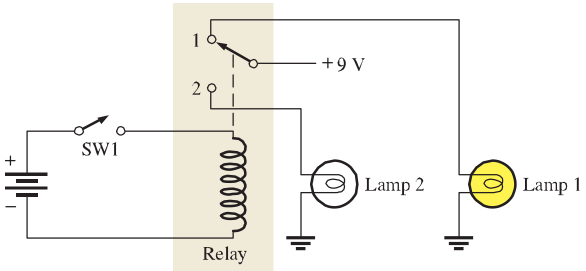

FIGURE 7–49

13. What causes the pointer in a d'Arsonval meter movement to deflect when there is current through the coil?

SECTION 7–4 Magnetic Hysteresis

14. What is the magnetizing force in Problem 9 if the length of the core is 0.2 m?
15. How can the flux density in Figure 7–50 be changed without altering the physical characteristics of the core?
16. In Figure 7–50, determine the following if the winding has 100 turns:
    (a) H **(b)** ϕ **(c)** B
17. Determine from the hysteresis curves in Figure 7–51 which material has the most retentivity.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

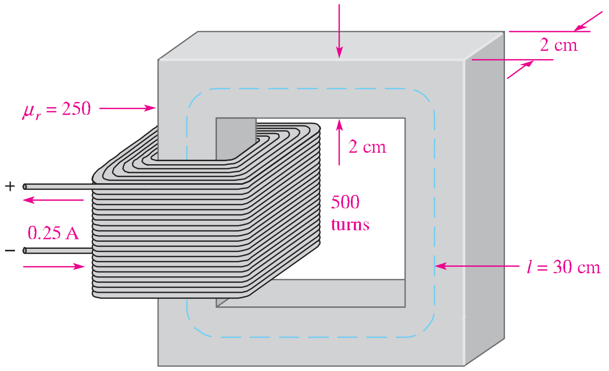

FIGURE 7–50

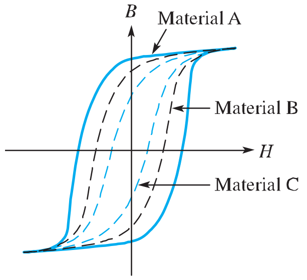

FIGURE 7–51

SECTION 7–5 Electromagnetic Induction

18. According to Faraday's law, what happens to the induced voltage across a given coil if the rate of change of magnetic flux doubles?

19. The voltage induced across a certain coil is 100 mV. A 100 Ω resistor is connected to the coil terminals. What is the induced current?

20. In Figure 7–35, why is there no induced voltage when the steel disk is not rotating?

21. A 20 cm long conductor is moving upward between the poles of a magnet as shown in Figure 7–52. The pole faces are 8.5 cm on each side and the flux is 1.24 mWb. The motion produces an induced voltage across the conductor of 44 mV. What is the speed of the conductor?

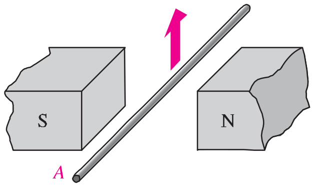

FIGURE 7–52

22. **(a)** For the conductor shown in Figure 7–52, what is the polarity of the end marked with the letter *A*?
(b) Assuming a complete path is provided and current is in the direction you indicated, what direction is the induced force on the conductor?

SECTION 7–6 DC Generators

23. If a generator is 80% efficient and delivers 45 W to a load, what is the input power?

24. Assume the self-excited shunt dc generator in Figure 7–43 has a load connected to it that draws 12 A. If the field windings draw 1.0 A, what is the armature current?

25. **(a)** If the output voltage in Problem 24 is 14 V, what power is supplied to the load?
(b) What power is dissipated in the field resistance?

SECTION 7–7 DC Motors

26. **(a)** What power is developed by a motor that turns at 1200 rpm and has a torque of 3.0 N-m?
(b) What is the horsepower rating of the motor? (746 W = 1 hp)

27. Assume a motor dissipates 12 W internally when it delivers 50 W to a load. What is the efficiency?

ADVANCED PROBLEMS

28. A basic one-loop dc generator is rotated at 60 rps. How many times each second does the dc output voltage peak (reach a maximum)?

29. Assume that another loop of wire, 90° from the first loop, is added to the dc generator in Problem 28. Describe the output voltage. Let the maximum voltage be 10 V.

30. Explain the purpose of each of the three sets of relay contacts (*A*, *B*, and *C*) in Figure 7–22.

31. For the DC motor reversing circuit in Figure 7–46, assume the polarity to the armature is opposite to the polarity of the stator. Is the relay coil energized or not? Explain your answer.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

MULTISIM

MULTISIM TROUBLESHOOTING PROBLEMS

32. Open file P07-32; files are found at www.pearsonhighered.com/floyd. Determine if there is a fault in the circuit. If so, identify the fault.

33. Open file P07-33 and determine if there is a fault in the circuit. If so, identify the fault.

ANSWERS TO SECTION CHECKUPS

SECTION 7–1 The Magnetic Field

1. The north poles repel.
2. Magnetic flux is the group of lines of force that make up a magnetic field; magnetic flux density is the flux per area.
3. The gauss and the tesla
4. $B = \phi/A = 900\ \mu\text{T}$

SECTION 7–2 Electromagnetism

1. Electromagnetism is produced by current through a conductor. An electromagnetic field exists only when there is current. A magnetic field exists independently of current.
2. The direction of the magnetic field also reverses when the current is reversed.
3. Flux equals magnetomotive force divided by reluctance.
4. Flux is analogous to current, mmf is analogous to voltage, and reluctance is analogous to resistance.

SECTION 7–3 Electromagnetic Devices

1. A solenoid provides mechanical movement of a shaft. A relay provides an electrical contact closure.
2. The movable part of a solenoid is the plunger.
3. The movable part of a relay is the armature.
4. The d'Arsonval meter movement is based on the interaction of magnetic fields.

SECTION 7–4 Magnetic Hysteresis

1. In a wirewound core, an increase in current increases the flux density.
2. Retentivity is the ability of a material to remain magnetized after removal of the magnetizing force.

SECTION 7–5 Electromagnetic Induction

1. The induced voltage is zero.
2. The induced voltage increases.
3. A force is exerted on the current-carrying conductor in a magnetic field.
4. The induced voltage is zero.

SECTION 7–6 DC Generators

1. The rotor
2. The commutator reverses the current in the rotating coil.
3. Greater resistance will reduce the magnetic flux, causing the output voltage to drop.
4. A generator in which the field windings derive their current from the output

SECTION 7–7 DC Motors

1. Back emf is a voltage developed in a motor because of generator action as the rotor turns. It opposes the original supply voltage.
2. Back emf reduces the armature current.
3. A series-wound motor
4. Higher reliability because there are no brushes to wear out

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

ANSWERS TO RELATED PROBLEMS FOR EXAMPLES

7–1 19.6×10^{-3} T

7–2 31.0 T

7–3 The reluctance is reduced to 12.8×10^{6} At/Wb.

7–4 Reluctance is 165.7×10^{3} At/Wb

7–5 7.2 mWb

7–6 $F_m = 42.5$ At; $\mathcal{R} = 8.5 \times 10^{4}$ At/Wb

7–7 40 mV

7–8 The direction is along the negative x axis.

7–9 0.18 hp

ANSWERS TO SELF-TEST

1. (b) **2.** (c) **3.** (a) **4.** (d) **5.** (b) **6.** (c) **7.** (a) **8.** (d)
9. (c) **10.** (c) **11.** (b) **12.** (c) **13.** (a) **14.** (d) **15.** (c)

ANSWERS TO TRUE/FALSE QUIZ

1. T **2.** F **3.** F **4.** F **5.** T **6.** T **7.** F **8.** T
9. T **10.** F

6238262 2013/12/10 12.27.66.5

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

CHAPTER 8

INTRODUCTION TO ALTERNATING CURRENT AND VOLTAGE

OUTLINE

8–1 The Sinusoidal Waveform
8–2 Voltage and Current Values of Sine Waves
8–3 Angular Measurement of a Sine Wave
8–4 The Sine Wave Formula
8–5 Analysis of AC Circuits
8–6 Alternators (AC Generators)
8–7 AC Motors
8–8 Nonsinusoidal Waveforms
8–9 The Oscilloscope
8–10 Signal Sources

OBJECTIVES

- Identify a sinusoidal waveform and measure its characteristics
- Determine the voltage and current values of sine waves
- Describe angular relationships of sine waves
- Mathematically analyze a sinusoidal waveform
- Apply the basic circuit laws to resistive ac circuits
- Describe how an alternator generates electricity
- Explain how ac motors convert electrical energy into rotational motion
- Identify the characteristics of basic nonsinusoidal waveforms
- Use the oscilloscope to measure waveforms
- Describe the types of signal sources and explain the purpose of typical controls

KEY TERMS

Sine wave
Waveform
Cycle
Period (T)
Frequency (f)
Hertz (Hz)
Oscillator
Function generator
Instantaneous value
Peak value
Amplitude
Peak-to-peak value
rms value
Degree
Radian
Phase
Alternator
Induction motor
Synchronous motor
Squirrel cage
Slip
Pulse
Pulse width (t_W)
Rise time (t_r)
Fall time (t_f)
Periodic
Duty cycle
Ramp
Fundamental frequency
Harmonics
Oscilloscope
Modulation

VISIT THE WEBSITE
Study aids for this chapter are available at
http://pearsonhighered.com/floyd

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

INTRODUCTION

This chapter provides an introduction to alternating current (ac) circuits. Alternating voltages and currents fluctuate with time and periodically change polarity and direction according to certain patterns called *waveforms*. Particular emphasis is given to the sinusoidal waveform (sine wave) because of its basic importance in ac circuits. Alternators, which generate sine waves, and ac motors are introduced. Other types of waveforms are also introduced, including pulse, triangular, and sawtooth. The use of an oscilloscope for displaying and measuring waveforms is discussed. Frequently, specialized waveforms are needed to test systems. The chapter concludes with a discussion on signal generating.

8–1 THE SINUSOIDAL WAVEFORM

The sinusoidal waveform or sine wave is the fundamental type of alternating current (ac) and alternating voltage. It is also referred to as a sinusoidal wave, or, simply, sinusoid. The electrical service provided by the power companies is in the form of sinusoidal voltage and current. In addition, other types of repetitive waveforms are composites of many individual sine waves called harmonics.

After completing this section, you should be able to

- **Identify a sinusoidal waveform and measure its characteristics**
  - **Define and determine the period**
  - **Define and determine the frequency**
  - **Relate the period and the frequency**
  - **Describe two types of electronic signal generators**

Sine waves, or sinusoids, are produced by two types of sources: rotating electrical machines (ac generators) or electronic oscillator circuits, which are used in instruments commonly known as electronic signal generators. Electronic signal generators are introduced in this section and covered in more detail in Section 8–10. Alternators, which generate ac by electromechanical means, are covered in Section 8–6. Figure 8–1 shows the symbol used to represent a source of sinusoidal voltage.

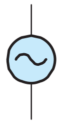

FIGURE 8–1 Symbol for a sinusoidal voltage source.

Figure 8–2 is a graph that shows the general shape of a sine wave, which can be either an alternating current or an alternating voltage. Voltage (or current) is displayed on the vertical axis, and time (t) is displayed on the horizontal axis. Notice how the voltage (or current) varies with time. Starting at zero, the voltage (or current) increases to a positive maximum (peak), returns to zero, and then increases to a negative maximum (peak) before returning again to zero, thus completing one full cycle.

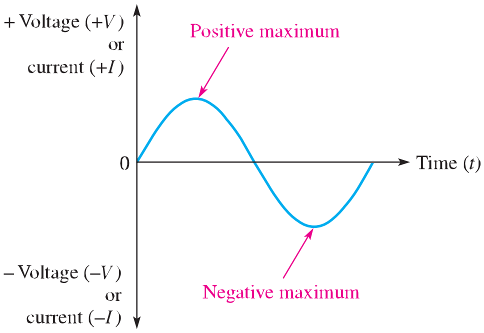

FIGURE 8–2 Graph of one cycle of a sine wave.

Polarity of a Sine Wave

As mentioned, a sine wave changes polarity at its zero value; that is, it alternates between positive and negative values. When a sinusoidal voltage source (V_s) is applied to a resistive circuit, as in Figure 8–3, an alternating sinusoidal current results. When the voltage changes polarity, the current correspondingly changes direction as indicated.

During the positive alternation of the source voltage V_s, the current is in the direction shown in Figure 8–3(a). During a negative alternation of the source voltage, the current is in the opposite direction, as shown in Figure 8–3(b). The combined positive and negative alternations make up one cycle of a sine wave.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

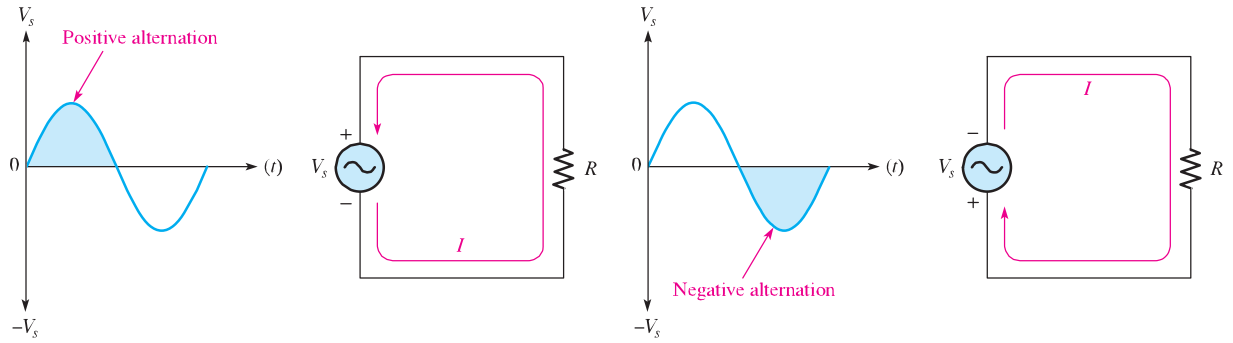

(a) During a positive alternation of the voltage, current is in the direction shown.

(b) During a negative alternation of the voltage, current reverses direction, as shown.

FIGURE 8–3 **Alternating current and voltage.**

Period of a Sine Wave

6238262 2013/12/10 12.27.66.5

A sine wave varies with time (t) in a definable manner.

> **The time required for a given sine wave to complete one full cycle is called the period (T).**

Figure 8–4(a) illustrates the period of a sine wave. Typically, a sine wave continues to repeat itself in identical cycles, as shown in Figure 8–4(b). Since all cycles of a repetitive sine wave are the same, the period is always a fixed value for a given sine wave. The period of a sine wave can be measured from a zero crossing to the next corresponding zero crossing, as indicated in Figure 8–4(a). The period can also be measured from any peak in a given cycle to the corresponding peak in the next cycle.

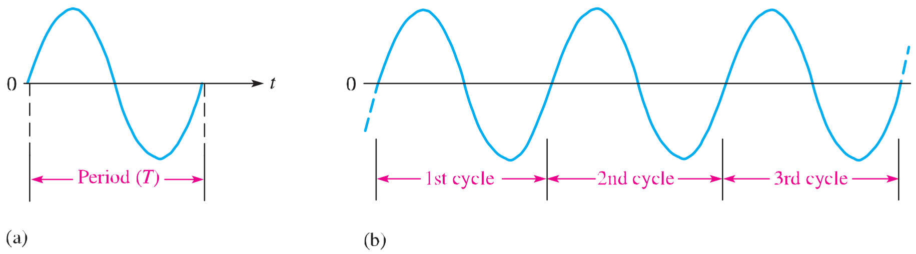

FIGURE 8–4 **The period of a given sine wave is the same for each cycle.**

EXAMPLE 8–1

What is the period of the sine wave in Figure 8–5?

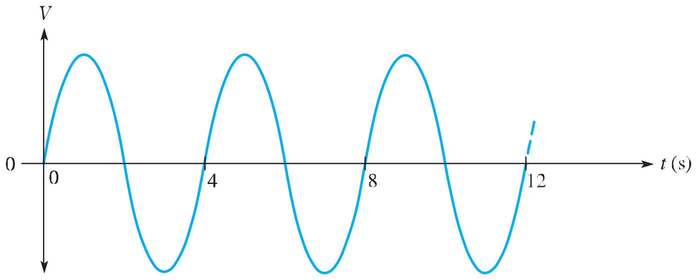

FIGURE 8–5

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

As shown in Figure 8–5, it takes twelve seconds (12 s) to complete three cycles. Therefore, to complete one cycle it takes four seconds (4 s), which is the period.

$$T = \mathbf{4\ s}$$

RELATED PROBLEM*

What is the period if a given sine wave goes through five cycles in 12 s?

***Answers are at the end of the chapter.**

EXAMPLE 8–2

Show three possible ways to measure the period of the sine wave in Figure 8–6. How many cycles are shown?

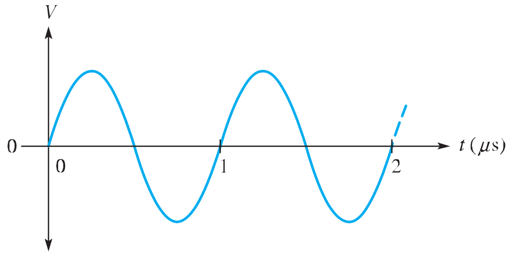

FIGURE 8–6

SOLUTION

Method 1: The period can be measured from one zero crossing to the corresponding zero crossing in the next cycle (the slope must be the same at the corresponding zero crossings).

Method 2: The period can be measured from the positive peak in one cycle to the positive peak in the next cycle.

Method 3: The period can be measured from the negative peak in one cycle to the negative peak in the next cycle.

These measurements are indicated in Figure 8–7, where **two cycles** of the sine wave are shown. Keep in mind that you obtain the same value for the period no matter which corresponding peaks or corresponding zero crossings on the waveform you use.

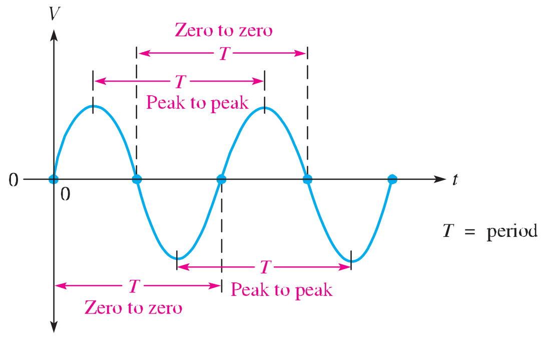

FIGURE 8–7 **Measurement of the period of a sine wave.**

RELATED PROBLEM

If a positive peak occurs at 1 ms and the next positive peak occurs at 2.5 ms, what is the period?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Frequency of a Sine Wave

Frequency is the number of cycles that a sine wave completes in one second.

The more cycles completed in one second, the higher the frequency. Frequency (f) is measured in units of **hertz (Hz)**. One hertz is equivalent to one cycle per second; for example, 60 Hz is 60 cycles per second. Figure 8–8 shows two sine waves. The sine wave in part (a) completes two full cycles in one second. The one in part (b) completes four cycles in one second. Therefore, the sine wave in part (b) has twice the frequency of the one in part (a).

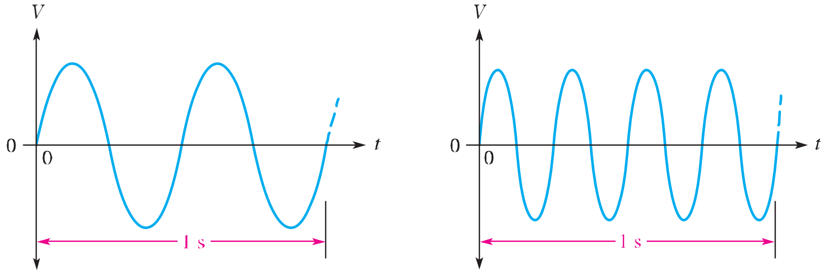

(a) Lower frequency: fewer cycles per second

(b) Higher frequency: more cycles per second

FIGURE 8–8 **Illustration of frequency.**

Relationship of Frequency and Period

The formulas for the relationship between frequency and period are

$$f = \frac{1}{T} \qquad (8\text{–}1)$$

$$T = \frac{1}{f} \qquad (8\text{–}2)$$

There is a reciprocal relationship between f and T. Knowing one, you can calculate the other with the x^{-1} or $1/x$ key on your calculator. (On some calculators, the reciprocal key is a secondary function.) This inverse relationship indicates that a sine wave with a longer period goes through fewer cycles in one second than one with a shorter period.

EXAMPLE 8–3

Which sine wave in Figure 8–9 has the higher frequency? Determine the frequency and the period of both waveforms.

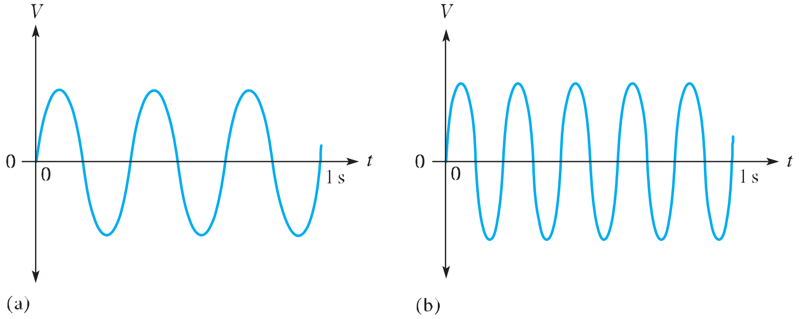

(a)

(b)

FIGURE 8–9

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

SOLUTION

The sine wave in Figure 8–9(b) has the higher frequency because it completes more cycles in 1 s than does the one in part (a).

In Figure 8–9(a), three cycles are completed in 1 s. Therefore,

$$f = \mathbf{3\ Hz}$$

One cycle takes 0.333 s (one-third second), so the period is

$$T = 0.333\text{ s} = \mathbf{333\ ms}$$

In Figure 8–9(b), five cycles are completed in 1 s. Therefore,

$$f = \mathbf{5\ Hz}$$

One cycle takes 0.2 s (one-fifth second), so the period is

$$T = 0.2\text{ s} = \mathbf{200\ ms}$$

RELATED PROBLEM

If the time interval between consecutive negative peaks of a given sine wave is 50 μs, what is the frequency?

EXAMPLE 8–4

The period of a certain sine wave is 10 ms. What is the frequency?

SOLUTION

Use Equation 8–1.

$$f = \frac{1}{T} = \frac{1}{10\text{ ms}} = \frac{1}{10 \times 10^{-3}\text{ s}} = \mathbf{100\ Hz}$$

RELATED PROBLEM

A certain sine wave goes through four cycles in 20 ms. What is the frequency?

EXAMPLE 8–5

The frequency of a sine wave is 60 Hz. What is the period?

SOLUTION

Use Equation 8–2.

$$T = \frac{1}{f} = \frac{1}{60\text{ Hz}} = \mathbf{16.7\ ms}$$

RELATED PROBLEM

If $f = 1$ kHz, what is T?

Electronic Signal Generators

The signal generator is an instrument that electronically produces sine waves for use in testing or controlling electronic circuits and systems. There are a variety of signal generators, ranging from special-purpose instruments that produce only one type of waveform in a limited frequency range to programmable instruments that produce a wide range of frequencies

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

and a variety of waveforms. All signal generators consist basically of an **oscillator**, which is an electronic circuit that produces a repetitive waveform on its output with only the dc supply voltage as an input.

Electronic signal generators are important in testing various types of systems. One type of signal generator is the **function generator**, so called because it can generate a variety of waveforms such as sine, triangle, and square waves. An **arbitrary waveform generator** is another generator that can generate complex signals that simulate different types of inputs. For example, a laboratory that does testing of structures for earthquake safety may need to simulate the complex signal from an earthquake; an arbitrary waveform generator such as shown in Figure 8–10 can easily simulate this. Section 8–10 covers these generators and other types of electronic signal generators in more detail.

FIGURE 8–10 **A typical arbitrary waveform generator. Copyright © Tektronix, Inc. Reproduced by permission.**

SECTION 8–1 CHECKUP*

1. Describe one cycle of a sine wave.
2. At what point does a sine wave change polarity?
3. How many maximum points does a sine wave have during one cycle?
4. How is the period of a sine wave measured?
5. Define *frequency*, and state its unit.
6. Determine f when $T = 5\ \mu s$.
7. Determine T when $f = 120$ Hz.

***Answers are at the end of the chapter.**

8–2 VOLTAGE AND CURRENT VALUES OF SINE WAVES

Five ways to express and measure the value of a sine wave in terms of its voltage or its current magnitude are instantaneous, peak, peak-to-peak, rms, and average values.

After completing this section, you should be able to

- **Determine the voltage and current values of sine waves**
  - **Find the instantaneous value at any point**
  - **Find the peak value**
  - **Find the peak-to-peak value**
  - **Define *rms* and find the rms value**
  - **Explain why the average value of an alternating sine wave is always zero over a complete cycle**
  - **Find the half-cycle average value**

Instantaneous Value

Figure 8–11 illustrates that at any point in time on a sine wave, the voltage (or current) has an **instantaneous value**. This instantaneous value is different at different points along the curve. Instantaneous values are positive during the positive alternation and negative during the negative alternation. Instantaneous values of voltage and current are symbolized by lowercase v and i, respectively. The curve in part (a) shows voltage only, but it applies equally for current when v's are replaced with i's. An example of instantaneous values is shown in part (b), where the instantaneous voltage is 3.1 V at 1 μs, 7.07 V at 2.5 μs, 10 V at 5 μs, 0 V at 10 μs, −3.1 V at 11 μs, and so on.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

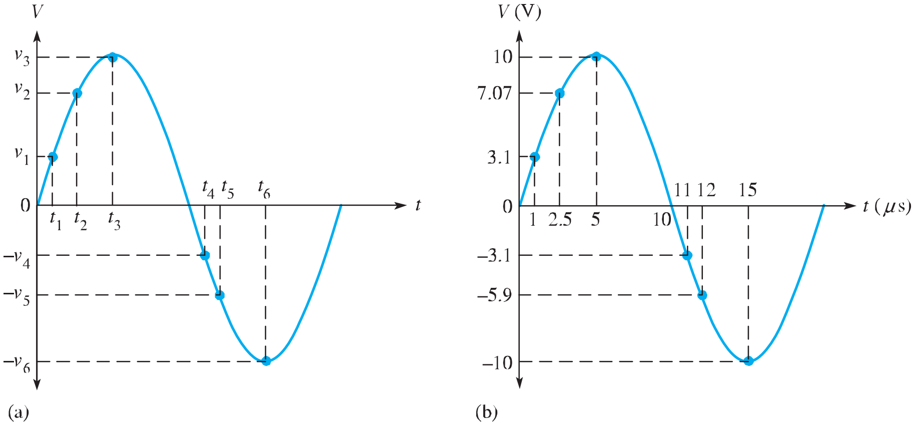

FIGURE 8–11 **Example of instantaneous values of a sinusoidal voltage.**

Peak Value

The **peak value** of a sine wave is the value of voltage (or current) at the positive or the negative maximum (peaks) with respect to zero. Since positive and negative peak values are equal in **magnitude**, a sine wave is characterized by a single peak value, as is illustrated in Figure 8–12. For a given sine wave, the peak value is constant and is represented by V_p or I_p. The maximum or peak value of a sine wave is also called its **amplitude**. The amplitude is measured from the 0 V line to the peak. In the figure, the peak voltage is 8 V, which is also its amplitude.

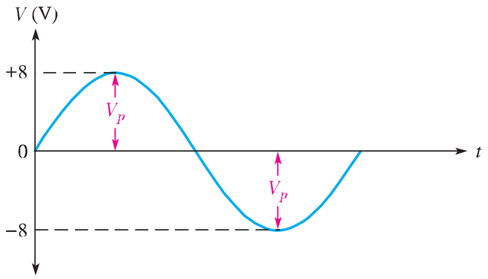

FIGURE 8–12 **Peak voltage.**

Peak-to-Peak Value

The **peak-to-peak value** of a sine wave, as illustrated in Figure 8–13, is the voltage (or current) from the positive peak to the negative peak. It is always twice the peak value as expressed by the following equations. Peak-to-peak values are represented by V_{pp} or I_{pp}.

$$V_{pp} = 2V_p \tag{8–3}$$

$$I_{pp} = 2I_p \tag{8–4}$$

In Figure 8–13, the peak-to-peak voltage is 16 V.

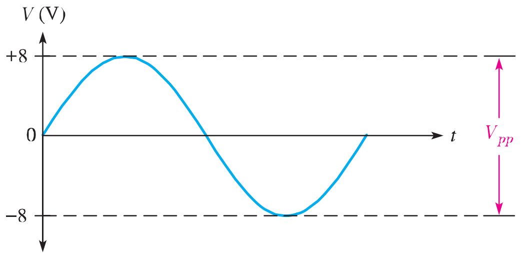

FIGURE 8–13 **Peak-to-peak voltage.**

RMS Value

The term *rms* stands for *root mean square*. Most ac voltmeters display the rms voltage. The 120 V at your wall outlet is an rms value. The **rms value**, also referred to as the **effective value**, of a sinusoidal voltage is actually a measure of the heating effect of the sine wave. For example, when a resistor is connected across an ac (sinusoidal) voltage source, as shown in Figure 8–14(a), it dissipates a certain amount of heat. Figure 8–14(b) shows the same resistor

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

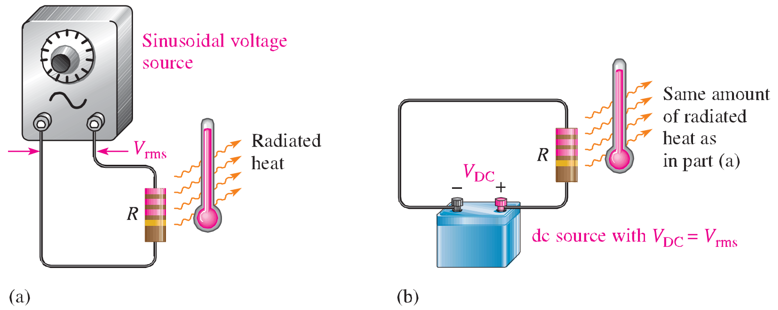
Sinusoidal voltage source
V_{rms}
Radiated heat
R
(a)
Same amount of radiated heat as in part (a)
R
V_{DC}
−
+
dc source with $V_{DC} = V_{rms}$
(b)

FIGURE 8–14 When the same amount of heat is being produced by the resistor in both setups, the sinusoidal voltage has an rms value equal to the dc voltage.

connected across a dc voltage source. The value of the ac voltage can be adjusted so that the resistor gives off the same amount of heat as it does when connected to the dc source.

The rms value of a sinusoidal voltage is equal to the dc voltage that produces the same amount of heat in a resistance as does the sinusoidal voltage.

You can convert the peak value of a sine wave to the corresponding rms value by using the following relationships for either voltage or current:

$$V_{rms} = 0.707V_p \tag{8–5}$$

$$I_{rms} = 0.707I_p \tag{8–6}$$

Using these formulas, you can also determine the peak value if you know the rms value.

$$V_p = \frac{V_{rms}}{0.707}$$

$$V_p = 1.414V_{rms} \tag{8–7}$$

Similarly,

$$I_p = 1.414I_{rms} \tag{8–8}$$

To find the peak-to-peak values, simply double the peak value, which is equivalent to multiplying the rms value by 2.828.

6238262 2013/12/10 12.27.66.5

$$V_{pp} = 2.828V_{rms} \tag{8–9}$$

and

$$I_{pp} = 2.828I_{rms} \tag{8–10}$$

SYSTEM EXAMPLE 8–1

CLAMP METERS

Most industrial electrical and electronic systems have ac motors and other high current loads used for fans, pumps, and conveyors. It is often impractical, time consuming, and a safety hazard to break open these circuits to check current, particularly when a process is ongoing. The traditional clamp meter is a noncontacting meter designed to measure large ac current by sensing the changing magnetic field surrounding the conductor through transformer action (transformers are covered in Chapter 14). The user selects the scale and clamps the jaws or flexible current probe around the conductor. The meter then displays the current based in the selected scale. It is a quick and easy way to check motor current to verify that loading is not excessive or to check for phase balance in three-phase motors.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Originally, clamp meters were designed to only measure ac current; today they also frequently have all of the traditional DMM functions available (resistance, voltage, dc current and sometimes frequency). Meters that can measure direct current (dc) use a Hall-effect sensor. The meter in Figure 8–15 is a multifunction meter that is measuring dc current from a car battery.

Some clamp meters have min/max capability. The min/max function will show the current range. This is useful for reading the currents in a 4-20 mA current loop. (See System Note on page 240.) The min/max capability is also useful for troubleshooting a problem such as determining the maximum current for a circuit breaker that trips intermittently.

FIGURE 8–15 **A multifunction clamp meter with detachable jaw.** Courtesy of Fluke Corporation.

Average Value

The average value of a sine wave when taken over one complete cycle is always zero because the positive values (above the zero crossing) offset the negative values (below the zero crossing).

To be useful for comparison purposes and in determining the average value of a rectified voltage such as found in power supplies, the **average value** of a sine wave is defined over a half-cycle rather than over a full cycle. The average value is expressed in terms of the peak value as follows for both voltage and current sine waves:

$$V_{avg} = 0.637V_p \qquad (8\text{–}11)$$

$$I_{avg} = 0.637I_p \qquad (8\text{–}12)$$

The half-cycle average value of a voltage sine wave is illustrated in Figure 8–16.

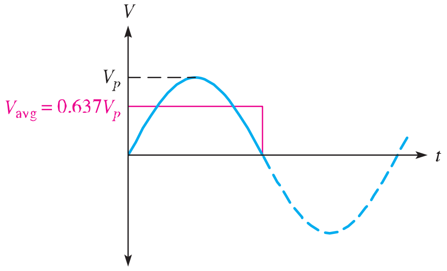

FIGURE 8–16 **Half-cycle average value.**

EXAMPLE 8–6

Determine V_p, V_{pp}, V_{rms}, and the half-cycle V_{avg} for the sine wave in Figure 8–17.

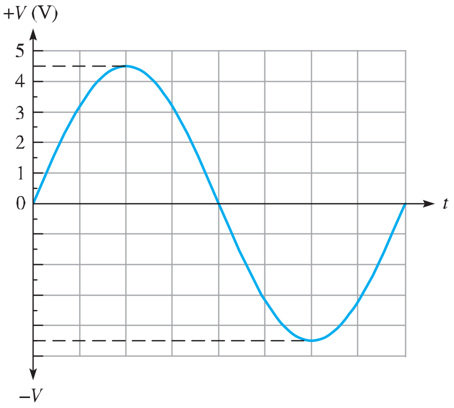

FIGURE 8–17

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

As taken directly from the graph, $V_p = \mathbf{4.5\ V}$. From this value, calculate the other values.

$$V_{pp} = 2V_p = 2(4.5\text{ V}) = \mathbf{9\ V}$$

$$V_{\text{rms}} = 0.707V_p = 0.707(4.5\text{ V}) = \mathbf{3.18\ V}$$

$$V_{\text{avg}} = 0.637V_p = 0.637(4.5\text{ V}) = \mathbf{2.87\ V}$$

RELATED PROBLEM

If $V_p = 25$ V, determine V_{pp}, V_{rms}, and V_{avg} for a sine wave.

SECTION 8–2 CHECKUP

1. Determine V_{pp} in each case when
   (a) $V_p = 1$ **(b)** $V_{\text{rms}} = 1.414$ V **(c)** $V_{\text{avg}} = 3$ V
2. Determine V_{rms} in each case when
   (a) $V_p = 2.5$ V **(b)** $V_{pp} = 10$ V **(c)** $V_{\text{avg}} = 1.5$ V
3. Determine the half-cycle V_{avg} in each case when
   (a) $V_p = 10$ V **(b)** $V_{\text{rms}} = 2.3$ V **(c)** $V_{pp} = 60$ V

8–3 ANGULAR MEASUREMENT OF A SINE WAVE

As you have seen, sine waves can be measured along the horizontal axis on a time basis; however, since the time for completion of one full cycle or any portion of a cycle is frequency-dependent, it is often useful to specify points on the sine wave in terms of an angular measurement expressed in degrees or radians. Angular measurement is independent of frequency.

After completing this section, you should be able to

- **Describe angular relationships of sine waves**
  - **Show how to measure a sine wave in terms of angles**
  - **Define *radian***
  - **Convert radians to degrees**
  - **Determine the phase of a sine wave**

A sinusoidal voltage can be produced by an alternator, which is an ac generator. There is a direct relationship between the rotation of the rotor in an alternator and the sine wave output. Thus, the angular measurement of the rotor's position is directly related to the angle assigned to the sine wave.

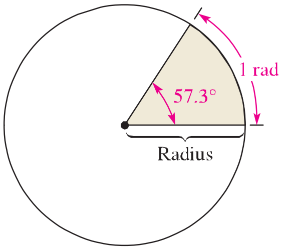

FIGURE 8–18 Angular measurement showing relationship of the radian to degrees.

Angular Measurement

A **degree** is an angular measurement corresponding to 1/360 of a circle or a complete revolution. A **radian** (rad) is the angle formed when the distance along the circumference of a circle is equal to the radius of the circle. One radian is equivalent to 57.3°, as illustrated in Figure 8–18. In a 360° revolution, there are 2π radians.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The Greek letter π (pi) represents the ratio of the circumference of any circle to its diameter and has a constant value of approximately 3.1416.

Scientific calculators have a π function so that the actual numerical value does not have to be entered.

Table 8–1 lists several values of degrees and the corresponding radian values. These angular measurements are illustrated in Figure 8–19.

TABLE 8–1

DEGREES (°)	RADIANS (RAD)
0	0
45	$\pi/4$
90	$\pi/2$
135	$3\pi/4$
180	π
225	$5\pi/4$
270	$3\pi/2$
315	$7\pi/4$
360	2π

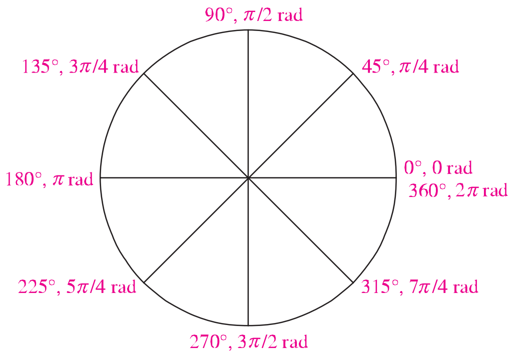

FIGURE 8–19 **Angular measurements starting at 0° and going counterclockwise.**

Radian/Degree Conversion

Degrees can be converted to radians using Equation 8–13.

$$\textbf{rad} = \left(\frac{\pi\ \textbf{rad}}{\textbf{180°}}\right) \times \textbf{degrees} \tag{8–13}$$

Similarly, radians can be converted to degrees with Equation 8–14.

$$\textbf{degrees} = \left(\frac{\textbf{180°}}{\pi\ \textbf{rad}}\right) \times \textbf{rad} \tag{8–14}$$

EXAMPLE 8–7

(a) Convert 60° to radians. **(b)** Convert $\pi/6$ radian to degrees.

SOLUTION

(a) $\text{Rad} = \left(\frac{\pi\ \text{rad}}{180°}\right)60° = \boldsymbol{\frac{\pi}{3}}\ \textbf{rad}$ **(b)** $\text{Degrees} = \left(\frac{180°}{\pi\ \text{rad}}\right)\left(\frac{\pi}{6}\ \text{rad}\right) = \mathbf{30°}$

RELATED PROBLEM

(a) Convert 15° to radians. **(b)** Convert $5\pi/8$ radians to degrees.

6238262 2013/12/10 12.27.66.5

Sine Wave Angles

The angular measurement of a sine wave is based on 360° or 2π rad for a complete cycle. A half-cycle is 180° or π rad; a quarter-cycle is 90° or $\pi/2$ rad; and so on. Figure 8–20(a) shows angles in degrees over a full cycle of a sine wave; part (b) shows the same points in radians.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

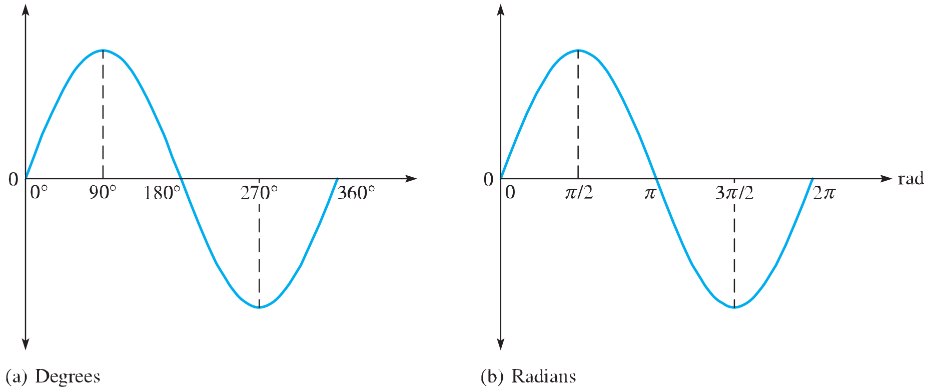
0
0°
90°
180°
270°
360°
(a) Degrees
0
0
π/2
π
3π/2
2π
rad
(b) Radians

FIGURE 8–20 **Sine wave angles.**

Phase of a Sine Wave

The **phase** of a sine wave is an angular measurement that specifies the position of that sine wave relative to a reference. Figure 8–21 shows one cycle of a sine wave to be used as the reference. Note that the first positive-going crossing of the horizontal axis (zero crossing) is at 0° (0 rad), and the positive peak is at 90° ($\pi/2$ rad). The negative-going zero crossing is at 180° (π rad), and the negative peak is at 270° ($3\pi/2$ rad). The cycle is completed at 360° (2π rad). When the sine wave is shifted left or right with respect to this reference, there is a phase shift.

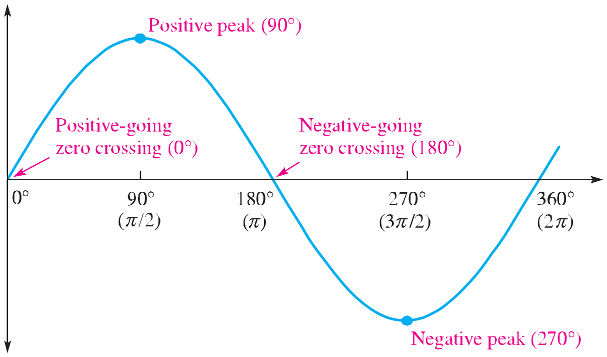
Positive peak (90°)
Positive-going zero crossing (0°)
Negative-going zero crossing (180°)
0°
90° (π/2)
180° (π)
270° (3π/2)
360° (2π)
Negative peak (270°)

FIGURE 8–21 **Phase reference.**

Figure 8–22 illustrates phase shifts of a sine wave. In part (a), sine wave B is shifted to the right by 90° ($\pi/2$ rad) with respect to sine wave A. Thus, there is a phase angle of 90° between sine wave A and sine wave B. In terms of time, the positive peak of sine wave B occurs later than the positive peak of sine wave A because time increases to the right

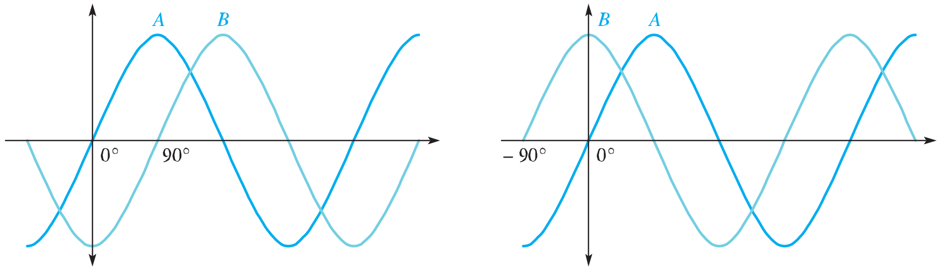
A
B
0°
90°
B
A
– 90°
0°

(a) *A* leads *B* by 90°, or *B* lags *A* by 90°.

(b) *B* leads *A* by 90°, or *A* lags *B* by 90°.

FIGURE 8–22 **Illustration of a phase shift.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

along the horizontal axis. In this case, sine wave *B* is said to **lag** sine wave *A* by 90° or $\pi/2$ radians. In other words, sine wave *A* leads sine wave *B* by 90°.

In Figure 8–22(b), sine wave *B* is shown shifted left by 90° with respect to sine wave *A*. Thus, again there is a phase angle of 90° between sine wave *A* and sine wave *B*. In this case, the positive peak of sine wave *B* occurs earlier in time than that of sine wave *A*; therefore, sine wave *B* is said to **lead** sine wave *A* by 90°. In both cases there is a 90° phase angle between the two waveforms.

EXAMPLE 8–8

What are the phase angles between sine waves *A* and *B* in Figure 8–23(a) and 8–23(b)?

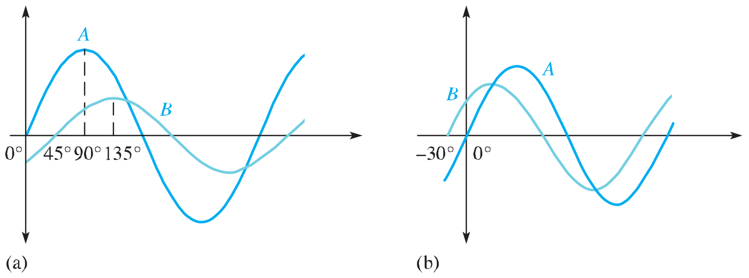

(a) (b)

FIGURE 8–23

SOLUTION

In Figure 8–23(a), the zero crossing of sine wave *A* is at 0°, and the corresponding zero crossing of sine wave *B* is at 45°. There is a **45°** phase angle between the two waveforms with sine wave *A* leading.

In Figure 8–23(b) the zero crossing of sine wave *B* is at −30°, and the corresponding zero crossing of sine wave *A* is at 0°. There is a **30°** phase angle between the two waveforms with sine wave *B* leading.

RELATED PROBLEM

If the positive-going zero crossing of one sine wave is at 15° relative to the 0° reference and that of the second sine wave is at 23° relative to the 0° reference, what is the phase angle between the sine waves?

When you measure the phase shift between two waveforms on an oscilloscope, you should align the waveforms vertically and make them appear to have the same amplitude. This is done by taking one of the channels out of vertical calibration and adjusting the corresponding waveform until its apparent amplitude equals that of the other waveform. This procedure eliminates the error caused if both waveforms are not measured at their exact center.

Polyphase Power

One important application of phase-shifted sine waves is in electrical power systems. Electrical utilities generate ac with three phases that are separated by 120° as shown in Figure 8–24. The reference is called neutral. Normally, three-phase power is delivered to the user with four lines—three hot lines and neutral. There are important advantages to three-phase power for ac motors. Three-phase motors are more efficient and simpler than an equivalent single-phase motor. Motors are discussed further in Section 8–7.

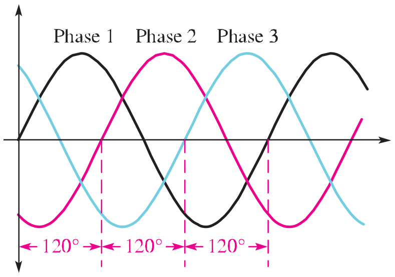

FIGURE 8–24 **Three-phase power waveforms.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Phase Converters

Many industrial motors and other devices use three-phase motors, which are simpler in construction, more efficient, and more reliable than single-phase motors. Three-phase motors are particularly useful for operations that require frequent starting and stopping or reversal such as used with machine tools.

Because of costs, some smaller industrial plants do not have three-phase power installed. For these plants, phase converters can supply three-phase power from a single-phase source. Several different technologies are available for converting single-phase to three-phase power. The solid-state phase converter shown here uses a digital signal processor (DSP) to monitor the load and generate the required three-phase power. If the load changes, the converter responds to keep constant voltage in all three phases.

Courtesy of Phase Technologies LLC.

SYSTEM NOTE

The three phases can be split up by the utility company to supply three separate single-phase systems. If only one of the three phases plus neutral is supplied, the result is standard 120 V, which is single-phase power. Single-phase power is distributed to residential and small commercial buildings; it consists of two 120 V hot lines that are 180° out of phase with each other and a neutral, which is grounded at the service entrance. The two hot lines allow for connecting to 240 V for high-power appliances (dryers, air conditioners).

6238262 2013/12/10 12.27.66.5

SECTION 8–3 CHECKUP

1. When the positive-going zero crossing of a sine wave occurs at 0°, at what angle does each of the following points occur?

   (a) positive peak **(b)** negative-going zero crossing

   (c) negative peak **(d)** end of first complete cycle

2. A half cycle is completed in _____ degrees, or _____ radians.
3. A full cycle is completed in _____ degrees, or _____ radians.
4. Determine the phase angle between the sine waves *B* and *C* in Figure 8–25.

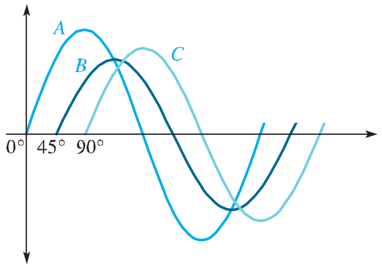

FIGURE 8–25

8–4 THE SINE WAVE FORMULA

A sine wave can be graphically represented by voltage or current values on the vertical axis and by angular measurement (degrees or radians) along the horizontal axis. This graph can be expressed mathematically, as you will see.

After completing this section, you should be able to

- **Mathematically analyze a sinusoidal waveform**
  - **State the sine wave formula**
  - **Find instantaneous values using the sine wave formula**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

A generalized graph of one cycle of a sine wave is shown in Figure 8–26. The sine wave amplitude, A, is the maximum value of the voltage or current on the vertical axis; angular values run along the horizontal axis. The variable y is an instantaneous value that can represent either voltage or current at a given angle, θ. The symbol θ is the Greek letter *theta*.

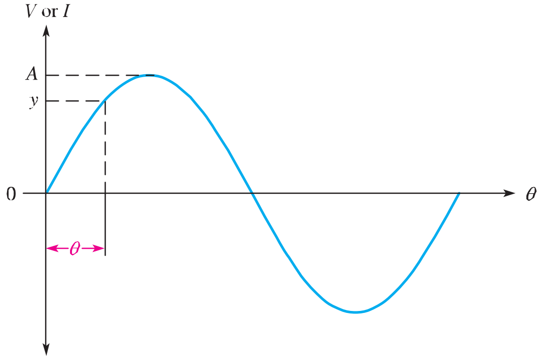

FIGURE 8–26 **One cycle of a generic sine wave showing amplitude and phase.**

All electrical sine waves follow a specific mathematical formula. The general expression for the sine wave curve in Figure 8–26 is

$$y = A \sin \theta \qquad (8\text{–}15)$$

This formula states that any point on the sine wave, represented by an instantaneous value y, is equal to the maximum value A times the sine (sin) of the angle θ at that point. For example, a certain voltage sine wave has a peak value of 10 V. The instantaneous voltage at a point 60° along the horizontal axis can be calculated as follows, where $y = v$ and $A = V_p$:

$$v = V_p \sin \theta = (10\text{ V})\sin 60^\circ = (10\text{ V})0.866 = 8.66\text{ V}$$

Figure 8–27 shows this particular instantaneous value on the curve. You can find the sine of any angle on most calculators by first entering the value of the angle and then pressing the SIN key. The calculator must be in the degree mode.

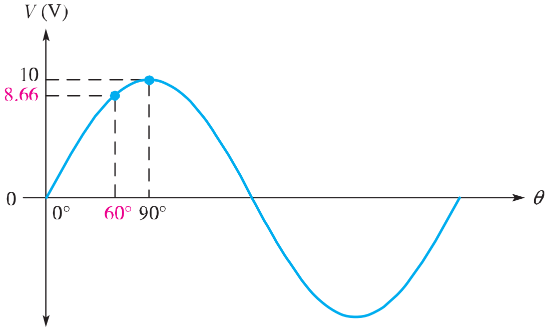

FIGURE 8–27 **Illustration of the instantaneous value of a voltage sine wave at $\theta = 60°$.**

Derivation of the Sine Wave Formula

As you move along the horizontal axis of a sine wave, the angle increases and the magnitude (height along the y axis) varies. At any given instant, the magnitude of a sine wave can be described by the values of the phase angle and the amplitude (maximum height) and can, therefore, be represented as a **phasor** quantity. A phasor quantity is one that has both

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

magnitude and direction (phase angle). A phasor is represented graphically as an arrow that rotates around a fixed point. The length of a sine wave phasor is the peak value (amplitude), and its angular position as it rotates is the phase angle. One full cycle of a sine wave can be viewed as the rotation of a phasor through 360°.

Figure 8–28 illustrates a phasor rotating counterclockwise through a complete revolution of 360°. If the tip of the phasor is projected over to a graph with the phase angles running along the horizontal axis, a sine wave is "traced out," as shown in the figure. At each angular position of the phasor, there is a corresponding magnitude value. As you can see, at 90° and at 270°, the amplitude of the sine wave is maximum and equal to the length of the phasor. At 0° and at 180°, the sine wave is equal to zero because the phasor lies horizontally at these points.

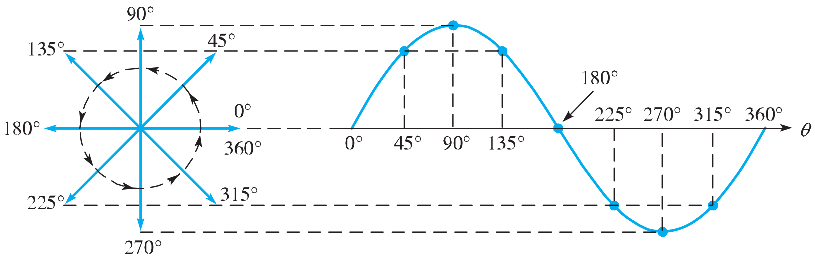

FIGURE 8–28 Sine wave represented by a rotating phasor.

Let's examine a phasor representation at one specific angle. Figure 8–29 shows the voltage phasor at an angular position of 45° and the corresponding point on the sine wave. The instantaneous value, v, of the sine wave at this point is related to both the position (angle) and the length (amplitude) of the phasor. The vertical distance from the phasor tip down to the horizontal axis represents the instantaneous value of the sine wave at that point.

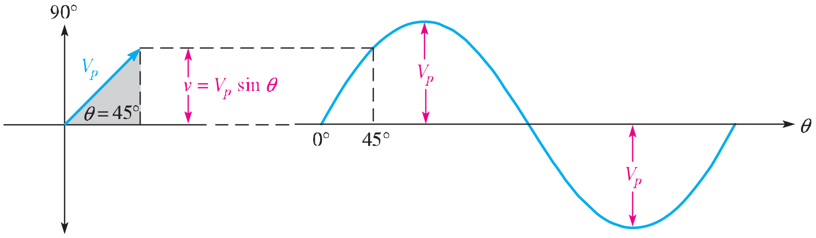

FIGURE 8–29 Right triangle derivation of sine wave formula, $v = V_p \sin \theta$.

Notice that when a vertical line is drawn from the phasor tip down to the horizontal axis, a **right triangle** is formed, as shown in Figure 8–29. The length of the phasor is the **hypotenuse** of the triangle, and the vertical projection is the opposite side. From trigonometry, *the opposite side of a right triangle is equal to the hypotenuse times the sine of the angle* θ. In this case, the length of the phasor is the peak value of the voltage sine wave, V_p. Thus, the opposite side of the triangle, which is the instantaneous value, can be expressed as

$$v = V_p \sin \theta \qquad (8\text{–}16)$$

This formula also applies to a current sine wave.

$$i = I_p \sin \theta \qquad (8\text{–}17)$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Expressions for Phase-Shifted Sine Waves

When a sine wave is shifted to the right of the reference (lags) by a certain angle, ϕ, (Greek letter phi) as illustrated in Figure 8–30(a), the general expression is

$$y = A\sin(\theta - \phi) \tag{8–18}$$

where y represents instantaneous voltage or current, and A represents the peak value (amplitude). When a sine wave is shifted to the left of the reference (leads) by a certain angle, ϕ, as shown in Figure 8–30(b), the general expression is

$$y = A\sin(\theta + \phi) \tag{8–19}$$

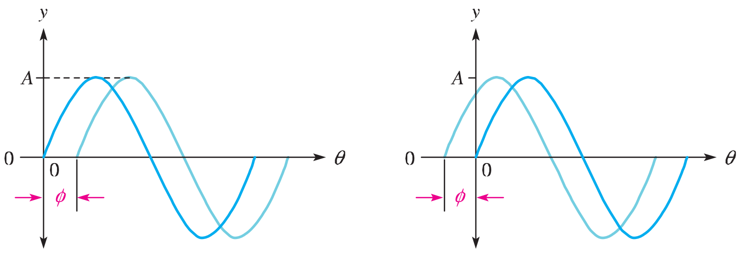

(a) $y = A\sin(\theta - \phi)$ (b) $y = A\sin(\theta + \phi)$

FIGURE 8–30 **Shifted sine waves.**

EXAMPLE 8–9

Determine the instantaneous value at 90° on the horizontal axis for each voltage sine wave in Figure 8–31.

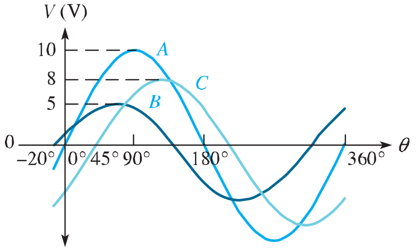

6238262 2013/12/10 12.27.66.5

FIGURE 8–31

SOLUTION

Sine wave A is the reference. Sine wave B is shifted left 20° with respect to A, so B leads. Sine wave C is shifted right 45° with respect to A, so C lags.

$$v_A = V_p\sin\theta = (10\text{ V})\sin 90° = \mathbf{10\text{ V}}$$
$$v_B = V_p\sin(\theta + \phi_B) = (5\text{ V})\sin(90° + 20°) = (5\text{ V})\sin 110° = \mathbf{4.70\text{ V}}$$
$$v_C = V_p\sin(\theta - \phi_C) = (8\text{ V})\sin(90° - 45°) = (8\text{ V})\sin 45° = \mathbf{5.66\text{ V}}$$

RELATED PROBLEM

A sine wave has a peak value of 20 V. What is the instantaneous value at +65° from the 0° reference?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 8–4 CHECKUP

1. Determine the sine of the following angles:
- **(a)** 30°
- **(b)** 60°
- **(c)** 90°

2. Calculate the instantaneous value at 120° for the sine wave in Figure 8–27.

3. Determine the instantaneous value at 45° on a voltage sine wave that leads the zero reference by 10° ($V_p = 10$ V).

8–5 ANALYSIS OF AC CIRCUITS

When a time-varying ac voltage such as a sinusoidal voltage is applied to a circuit, the circuit laws and power formulas that you studied earlier still apply. Ohm's law, Kirchhoff's laws and the power formulas apply to ac circuits in the same way that they apply to dc circuits.

After completing this section, you should be able to

- **Apply the basic circuit laws to resistive ac circuits**
  - **Apply Ohm's law to resistive circuits with ac sources**
  - **Apply Kirchhoff's voltage law and current law to resistive circuits with ac sources**
  - **Determine power in resistive ac circuits**
  - **Determine total voltages that have both ac and dc components**

If a sinusoidal voltage is applied across a resistor, as shown in Figure 8–32, there is a sinusoidal current. The current is zero when the voltage is zero and is maximum when the voltage is maximum. When the voltage changes polarity, the current reverses direction. As a result, the voltage and current are said to be in phase with each other.

When you use Ohm's law in ac circuits, remember that both the voltage and the current must be expressed consistently, that is, both as peak values, both as rms values, both as average values, and so on.

Kirchhoff's voltage and current laws apply to ac circuits as well as to dc circuits. Figure 8–33 illustrates Kirchhoff's voltage law in a resistive circuit that has a sinusoidal voltage source. As you can see, the source voltage is the sum of all the voltage drops across the resistors, just as in a dc circuit.

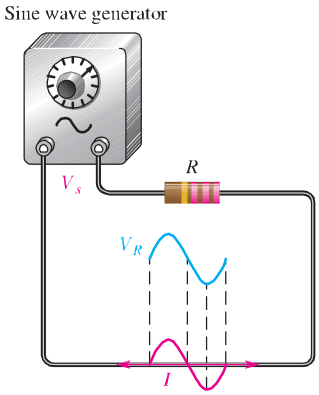

FIGURE 8–32 A sinusoidal voltage produces a sinusoidal current.

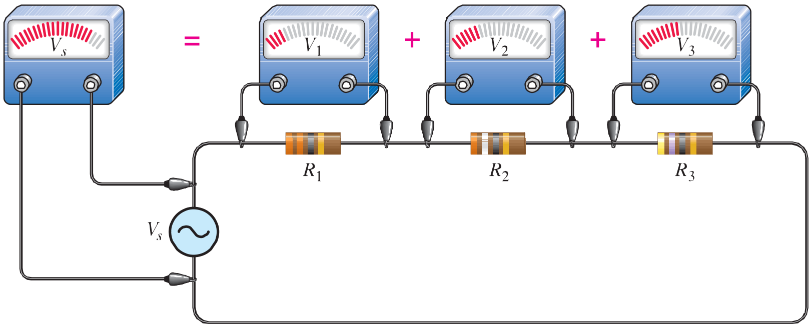

FIGURE 8–33 Illustration of Kirchhoff's voltage law in an ac circuit.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Power in resistive ac circuits is determined the same as for dc circuits except that you must use rms values of current and voltage. Recall that the rms value of a sine wave voltage is equivalent to a dc voltage of the same value in terms of its heating effect. The general power formulas are restated for a resistive ac circuit as

$$P = V_{\text{rms}} I_{\text{rms}}$$

$$P = \frac{V_{\text{rms}}^2}{R}$$

$$P = I_{\text{rms}}^2 R$$

EXAMPLE 8–10

Determine the rms voltage across each resistor and the rms current in Figure 8–34. The source voltage is given as an rms value. Also, determine the total power.

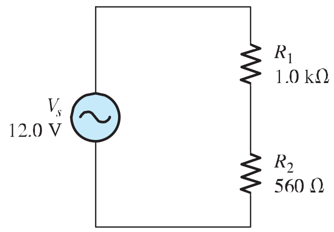

FIGURE 8–34

SOLUTION

The total resistance of the circuit is

$$R_{tot} = R_1 + R_2 = 1.0\ \text{k}\Omega + 560\ \Omega = 1.56\ \text{k}\Omega$$

Use Ohm's law to find the rms current.

$$I_{\text{rms}} = \frac{V_{s(\text{rms})}}{R_{tot}} = \frac{12.0\ \text{V}}{1.56\ \text{k}\Omega} = \mathbf{7.69\ mA}$$

The rms voltage drop across each resistor is

$$V_{1(\text{rms})} = I_{\text{rms}} R_1 = (7.69\ \text{mA})(1.0\ \text{k}\Omega) = \mathbf{7.69\ V}$$

$$V_{2(\text{rms})} = I_{\text{rms}} R_2 = (7.69\ \text{mA})(560\ \Omega) = \mathbf{4.31\ V}$$

The total power is

$$P_{tot} = I_{\text{rms}}^2 R_{tot} = (7.69\ \text{mA})^2 (1.56\ \text{k}\Omega) = \mathbf{92.3\ mW}$$

RELATED PROBLEM

Repeat this example for a source voltage of 10 V peak.

MULTISIM

Open Multisim file E08-10; files are found at www.pearsonhighered.com/floyd. Measure the rms voltage across each resistor and compare to the calculated values. Change the source voltage to a peak value of 10 V, measure each resistor voltage, and compare to your calculated values.

EXAMPLE 8–11

All values in Figure 8–35 are given in rms.

(a) Find the unknown peak voltage drop in Figure 8–35(a).

(b) Find the total rms current in Figure 8–35(b).

(c) Find the total power in Figure 8–35(b).

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

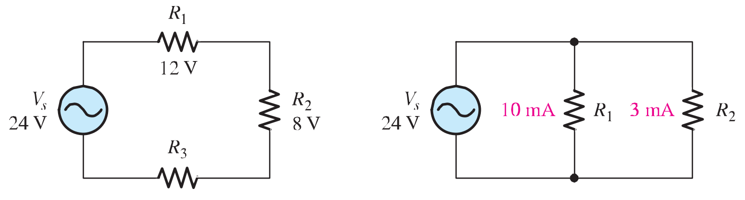

(a)

(b)

FIGURE 8–35

SOLUTION

(a) Use Kirchhoff's voltage law to find V_3.

$$V_s = V_1 + V_2 + V_3$$

$$V_{3(\text{rms})} = V_{s(\text{rms})} - V_{1(\text{rms})} - V_{2(\text{rms})} = 24\text{ V} - 12\text{ V} - 8\text{ V} = 4\text{ V}$$

Convert rms to peak.

$$V_{3(p)} = 1.414V_{3(\text{rms})} = 1.414(4\text{ V}) = \mathbf{5.66\text{ V}}$$

(b) Use Kirchhoff's current law to find I_{tot}.

$$I_{tot(\text{rms})} = I_{1(\text{rms})} + I_{2(\text{rms})} = 10\text{ mA} + 3\text{ mA} = \mathbf{13\text{ mA}}$$

(c) $P_{tot} = V_{\text{rms}}I_{\text{rms}} = (24\text{ V})(13\text{ mA}) = \mathbf{312\text{ mW}}$

RELATED PROBLEM

A series circuit has the following voltage drops: $V_{1(\text{rms})} = 3.50$ V, $V_{2(p)} = 4.25$ V, $V_{3(\text{avg})} = 1.70$ V. Determine the peak-to-peak source voltage.

Superimposed DC and AC Voltages

In many practical circuits, you will find both dc and ac voltages combined. An example of this is in amplifier circuits where ac signal voltages are superimposed on dc operating voltages. Figure 8–36 shows a dc source and an ac source in series. These two voltages will add algebraically to produce an ac voltage "riding" on a dc level, as measured across the resistor.

If V_{DC} is greater than the peak value of the sinusoidal voltage, the combined ac and dc voltage is a sine wave that never reverses polarity and is therefore nonalternating. That is, the sine wave is riding on a dc level, as shown in Figure 8–37(a). If V_{DC} is less than the peak value of the sine wave, the sine wave will be negative during a portion of its lower half-cycle, as illustrated in Figure 8–37(b), and is therefore alternating. In either case, the sine wave will reach a maximum voltage equal to $V_{\text{DC}} + V_p$, and it will reach a minimum voltage equal to $V_{\text{DC}} - V_p$.

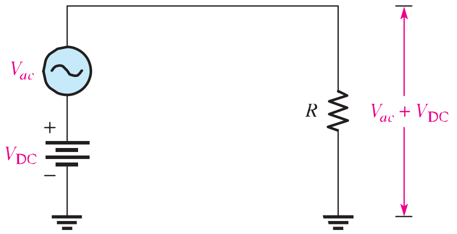

FIGURE 8–36 **Superimposed dc and ac voltages.**

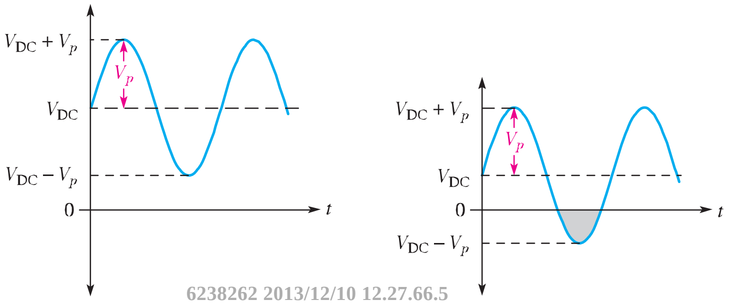

6238262 2013/12/10 12.27.66.5

(a) $V_{\text{DC}} > V_p$. The sine wave never goes negative.

(b) $V_{\text{DC}} < V_p$. The sine wave reverses polarity during a portion of its cycle.

FIGURE 8–37 **Sine waves with dc levels.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 8–12

Determine the maximum and minimum voltages across the resistor in each circuit of Figure 8–38 and show the resulting waveforms.

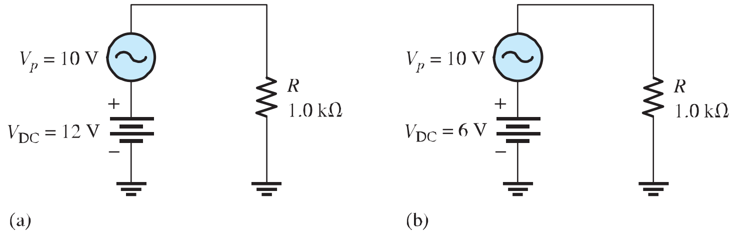

FIGURE 8–38

SOLUTION

In Figure 8–38(a), the maximum voltage across *R* is

$$V_{max} = V_{DC} + V_p = 12\text{ V} + 10\text{ V} = \mathbf{22\ V}$$

The minimum voltage across *R* is

$$V_{min} = V_{DC} - V_p = 12\text{ V} - 10\text{ V} = \mathbf{2\ V}$$

Therefore, $V_{R(tot)}$ is a nonalternating sine wave that varies from +22 V to +2 V, as shown in Figure 8–39(a).

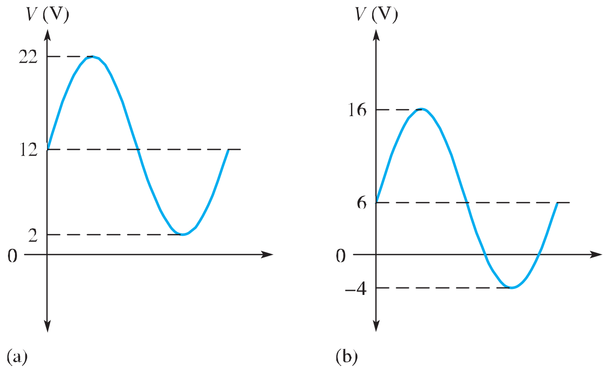

FIGURE 8–39

In Figure 8–38(b), the maximum voltage across *R* is

$$V_{max} = V_{DC} + V_p = 6\text{ V} + 10\text{ V} = \mathbf{16\ V}$$

The minimum voltage across *R* is

$$V_{min} = V_{DC} - V_p = \mathbf{-4\ V}$$

Therefore, $V_{R(tot)}$ is an alternating sine wave that varies from +16 V to −4 V, as shown in Figure 8–39(b).

RELATED PROBLEM

Explain why the waveform in Figure 8–39(a) is nonalternating but the waveform in part (b) is considered to be alternating.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 8–5 CHECKUP

1. A sinusoidal voltage with a half-cycle average value of 12.5 V is applied to a circuit with a resistance of 330 Ω. What is the peak current in the circuit?
2. The peak voltage drops in a series resistive circuit are 6.2 V, 11.3 V, and 7.8 V. What is the rms value of the source voltage?
3. What is the maximum positive value of the resulting total voltage when a sine wave with $V_p = 5$ V is added to a dc voltage of +2.5 V?
4. Will the resulting voltage in Question 3 alternate polarity?
5. If the dc voltage in Question 3 is −2.5 V, what is the maximum positive value of the resulting total voltage?

8–6 ALTERNATORS (AC GENERATORS)

An alternator is an ac generator that converts energy of motion into electrical energy. Although it is similar to a dc generator, the alternator is more efficient than the dc generator. Alternators are widely used in vehicles, boats, and other applications even when dc is the final output.

After completing this section, you should be able to

- **Describe how an alternator generates electricity**
  - **Identify the main parts of an alternator, including the rotor, stator, and slip rings**
  - **Explain why the output of a rotating-field alternator is taken from the stator**
  - **Describe the purpose of the slip rings**
  - **Explain how an alternator can be used to produce dc**

Simplified Alternator

Both the dc generator and the alternator, which generates ac voltage, are based on the principle of electromagnetic induction that produces a voltage when there is relative motion between a magnetic field and a conductor. For a simplified alternator, a single rotating loop passes permanent magnetic poles. The natural voltage that is generated by a rotating loop is ac. In an alternator, instead of the split rings used in a dc generator, solid rings called slip rings are used to connect to the rotor, and the output is ac. The simplest form of an alternator has the same appearance as a dc generator (see Figure 7–37) except for the slip rings, as shown in Figure 8–40.

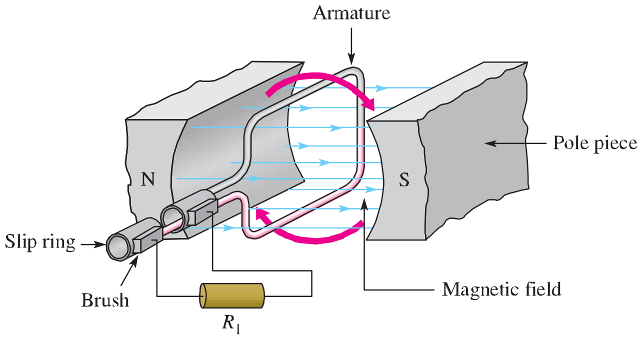

FIGURE 8–40 **A simplified alternator.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Frequency

In the simplified alternator in Figure 8–40, each revolution of the loop produces one cycle of a sine wave. The positive and negative peaks occur when the loop cuts the maximum number of flux lines. The rate the loop spins determines the time for one complete cycle and the frequency. If it takes 1/60 of a second to make a revolution, the period of the sine wave is 1/60 of a second and the frequency is 60 Hz. Thus, the faster the loop rotates, the higher the frequency.

Another way of achieving a higher frequency is to use more magnetic poles. When four poles are used instead of two, as shown in Figure 8–41, the conductor passes under a north and a south pole during one-half a revolution, which doubles the frequency. Alternators can have many more poles, depending on the requirements; some have as many as 100. The number of poles and the speed of the rotor determine the frequency in accordance with the following equation:

$$f = \frac{Ns}{120} \tag{8–20}$$

where f is the frequency in hertz, N is the number of poles, and s is the rotational speed in revolutions per minute.

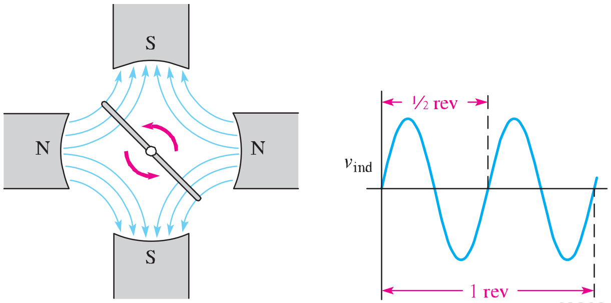

6238262 2013/12/10 12.27.66.5

FIGURE 8–41 **Four poles produce twice the frequency of two poles for the same rotational speed.**

EXAMPLE 8–13

Assume a large alternator is turned by a turbine at 300 rpm and has 24 poles. What is the output frequency?

SOLUTION

$$f = \frac{Ns}{120} = \frac{(24)(300\text{ rpm})}{120} = \mathbf{60\ Hz}$$

RELATED PROBLEM

At what speed must the rotor move to produce a 50 Hz output?

Practical Alternators

The single loop in our simplified alternator produces only a tiny voltage. In a practical alternator, hundreds of loops are wound on a magnetic core, which forms the rotor. Practical alternators usually have fixed windings surrounding the rotor instead of permanent magnets. Depending on the type of alternator, these fixed windings can either provide the magnetic field (in which case they are called field windings) or act as the fixed conductors that produce the output (in which case they are the armature windings).

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

ROTATING-ARMATURE ALTERNATORS In a rotating-armature alternator, the magnetic field is stationary and is supplied by permanent magnets or electromagnets operated from dc. With electromagnets, field windings are used instead of permanent magnets and provide a fixed magnetic field that interacts with the rotor coils. Power is generated in the rotating assembly and supplied to the load through the slip rings.

In a rotating-armature alternator, the rotor is the component from which power is taken. In addition to hundreds of windings, the practical rotating-armature alternator usually has many pole pairs in the stator that alternate as north and south poles, which serve to increase the output frequency.

ROTATING-FIELD ALTERNATOR The rotating-armature alternator is generally limited to low-power applications because all output current must pass through the slip rings and brushes. To avoid this problem, rotating-field alternators take the output from the stator coils and use a rotating magnet, hence the name. Small alternators may have a permanent magnet for a rotor, but most use an electromagnet formed by a wound rotor. A relatively small amount of dc is supplied to the rotor (through the slip rings) to power the electromagnet. As the rotating magnetic field sweeps by the stator windings, power is generated in the stator. The stator is therefore the armature in this case.

Figure 8–42 shows how a rotating-field alternator can generate three-phase sine waves. (For simplicity, a permanent magnet is shown for the rotor.) AC is generated in each windings as the north pole and the south pole of the rotor alternately sweep by a stator winding. If the north pole generates the positive portion of the sine wave, the south pole will generate the negative portion; thus, one rotation produces a complete sine wave. Each winding has a sine wave output; but because the windings are separated by 120°, the three sine waves are also shifted by 120°. This produces the three-phase output as shown. Most alternators generate three-phase voltage because it is more efficient to produce and is widely used in industry. If the final output is dc, three-phase is easier to convert to dc.

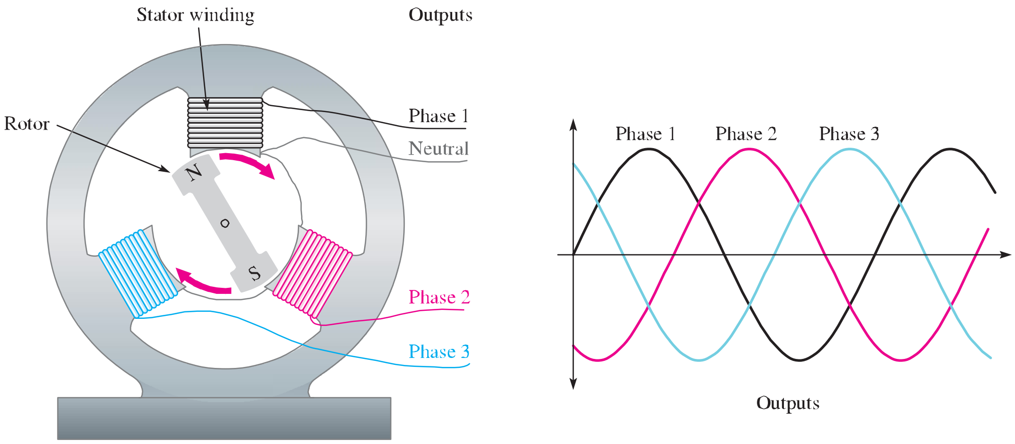

FIGURE 8–42 **The rotor shown is a permanent magnet that produces a strong magnetic field. As it sweeps by each stator winding, a sine wave is produced across that winding. The neutral is the reference.**

Rotor Current

A wound rotor offers important control advantages to alternators. A wound rotor enables control over the strength of the magnetic field by controlling the rotor current and hence the output voltage. For wound rotors, dc must be supplied to the rotor. This current is usually supplied through brushes and slip rings, which are made of a continuous ring of material, (unlike a commutator, which is segmented). Because the brushes need to pass only the

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

magnetizing rotor current, they last longer and are smaller than the brushes in an equivalent dc generator, which pass all of the output current.

In wound-rotor alternators, the only current through the brushes and slip rings is the dc that is used to maintain its magnetic field. The dc is usually derived from a small portion of the output current, which is taken from the stator and converted to dc. Large alternators, such as in power stations, may have a separate dc generator, called an **exciter**, to supply current to the field coils. An exciter can respond very fast to changes in output voltage to keep the alternator's output constant, an important consideration in high-power alternators. Some exciters are set up using a stationary field with the armature on the rotating main shaft. The result is a brushless system because the exciter output is on the rotating shaft. A brushless system eliminates the primary maintenance issue with large alternators in cleaning, repairing, and replacing brushes.

An Application

Alternators are used in nearly all modern automobiles, trucks, tractors, and other vehicles. In vehicles, the output is usually three-phase ac taken from stator windings and then converted to dc with diodes that are housed inside the alternator case. (Diodes are solid-state devices that allow current in only one direction.) Current to the rotor is controlled by a voltage regulator, which is also internal to the alternator. The voltage regulator keeps the output voltage relatively constant for engine speed changes or changing loads. Alternators have replaced dc generators in automobiles and most other applications because they are more efficient and more reliable.

Important parts of a small alternator, such as you might find in an automobile are shown in Figure 8–43. Like the self-excited generator discussed in Section 7–6, the rotor has a small residual magnetism to begin with, so an ac voltage is generated in the stator as soon as the rotor starts spinning. This ac is converted to dc by a set of rectifier diodes. A portion of the dc is used to provide current to the rotor; the rest is available for the loads. The amount of current required by the rotor is much less than the total current from the alternator, so it can easily provide the required current to the load.

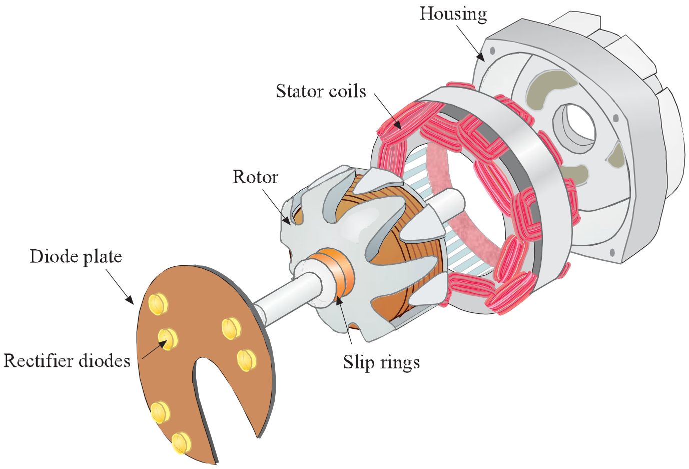

FIGURE 8–43 **Simplified view of a rotor, stator, and diode plate for a small alternator that produces dc.**

In addition to the fact that three-phase voltage is more efficient to produce, it can produce a stable dc output easily by using two diodes in each winding. Since vehicles require dc for the charging system and loads, the output of the alternator is converted to dc internally using a rectifier diode array mounted on a diode plate. Thus, a standard three-phase automotive alternator will normally have six diodes inside to convert the output to dc. (Some alternators have six independent stator coils and twelve diodes.)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SYSTEM EXAMPLE 8–2

AUTOMOTIVE CHARGING SYSTEM

The schematic for a basic automotive charging system is shown in Figure 8–44. The alternator is at the heart of the system. The alternator provides ac which is rectified (changed to dc) by the diode array. The dc is sent to the battery through a heavy wire.

There are normally four terminals on an alternator that are connected to the rest of the charging system. The terminals are labeled B, L, IG, and S. Terminal B is connected to the battery and carries the charging current. Terminal L is connected to the charge warning lamp. The charge lamp is located on the dash panel and is illuminated when there is a difference in voltage across it; if the lamp has equal positive voltages on both sides, it is off. Terminal IG is connected to the run side of the ignition switch, which activates the regulator. If voltage is removed from either side of the lamp, it illuminates to indicate a problem with the charging system. Terminal S is a "sense" connection, which is connected to the battery to provide a battery voltage monitor for the voltage regulator.

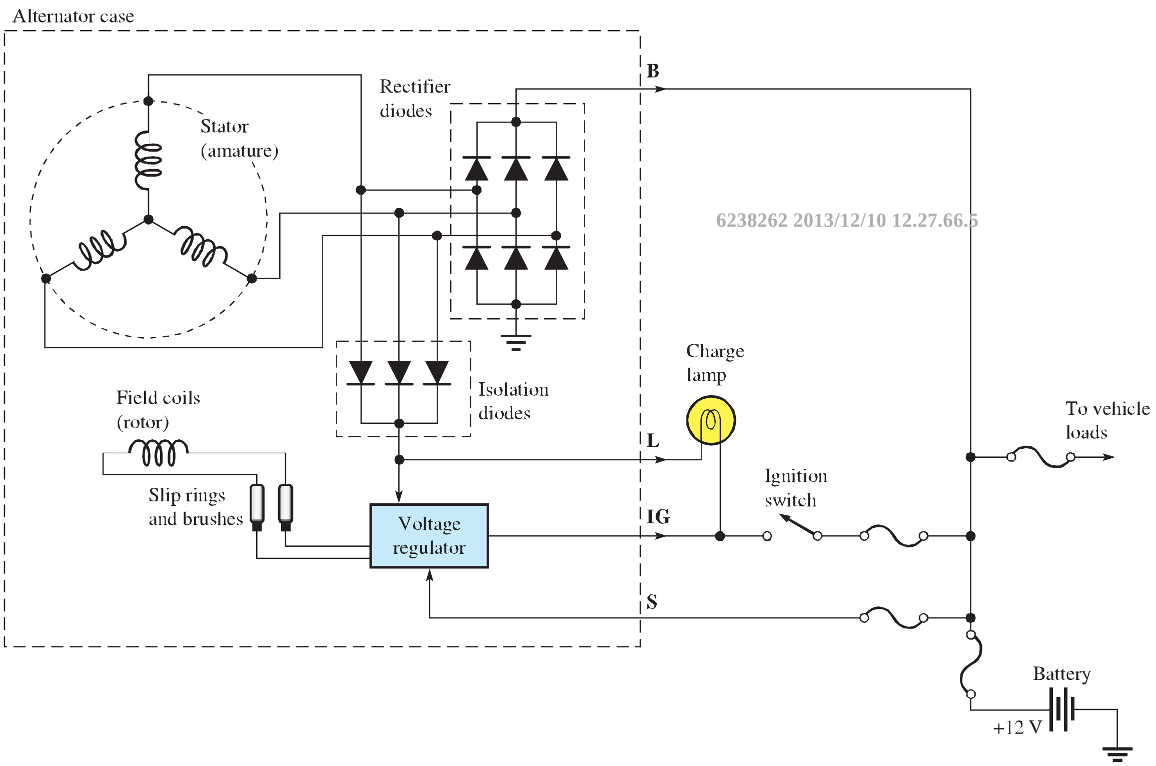

6238262 2013/12/10 12.27.66.5

FIGURE 8–44 **A basic automotive charging system.**

SECTION 8–6 CHECKUP

1. What two factors affect the frequency of an alternator?
2. What is the advantage of taking the output from the stator in a rotating-field alternator?
3. What is an exciter?
4. What is the purpose of the diodes in an automobile alternator?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

8–7 AC MOTORS

Motors are electromagnetic devices that represent the most common loads for ac in power applications. AC motors are used to operate household appliances such as heat pumps, refrigerators, washers, dryers, and vacuums. In industry, ac motors are used in many applications to move and process materials as well as refrigeration and heating units, machining operations, pumps, and much more. In this section, the two major types of ac motors, induction motors and synchronous motors, are introduced.

After completing this section, you should be able to

- **Explain how ac motors convert electrical energy into rotational motion**
  - **Cite the main differences between induction and synchronous motors**
  - **Explain how the magnetic field rotates in an ac motor**
  - **Explain how an induction motor develops torque**

AC Motor Classification

The two major classifications of ac motors are induction motors and synchronous motors. Several considerations determine which of these types are best for any given application. These considerations include the speed and power requirements, voltage rating, load characteristics (such as starting torque required), efficiency requirements, maintenance requirements, and operating environment (such as underwater operation or temperature).

An **induction motor** is so named because a magnetic field *induces* current in the rotor, creating a magnetic field that interacts with the stator field. Normally, there is no electrical connection to the rotor[1], so there is no need for slip rings or brushes, which tend to wear out. The rotor current is caused by electromagnetic induction, which also occurs in transformers (covered in Chapter 14), so induction motors are said to work by transformer action.

In a **synchronous motor**, the rotor moves in sync (at the same rate) as the rotating field of the stator. Synchronous motors are used in applications where maintaining constant speed is important. Synchronous motors are not self-starting and must receive starting torque from an external source or from built-in starting windings. Like alternators, synchronous motors use slip rings and brushes to provide current to the rotor.

Rotating Stator Field

Both synchronous and induction ac motors have a similar arrangement for the stator windings, which allow the magnetic field of the stator to rotate. The rotating stator field is equivalent to moving a magnet in a circle except that the rotating field is produced electrically, with no moving parts.

How can the magnetic field in the stator rotate if the stator itself does not move? The rotating field is created by the changing ac itself. Let's look at a rotating field with a three-phase stator, as shown in Figure 8–45. Notice that one of the three phases "dominates" at different times. When phase 1 is at 90°, the current in the phase 1 winding is at a maximum and current in the other windings is smaller. Therefore, the stator magnetic field will be oriented toward the phase-1 stator winding. As the phase-1 current declines, the phase-2 current increases, and the field rotates toward the phase-2 winding. The magnetic field will be oriented toward the phase-2 winding when current in it is a maximum. As the phase-2 current declines, the phase-3 current increases, and the field rotates toward the phase-3 winding. The process repeats as the field returns to the phase-1 winding. Thus, the field rotates at a rate determined by the frequency of the applied voltage. With a more detailed analysis, it can be shown that the magnitude of the field is unchanged; only the direction of the field changes.

HANDS ON TIP

New motor technology for high power motors will use high temperature superconductors (HTS). The new technology significantly reduces the size and weight of large motors by hundreds of tons. One application is for ships' propulsion motors, which will reduce fuel consumption and free up space on board. A 49,000 HP HTS motor has passed initial tests for the Navy and will be installed in the near future on Navy ships.

---

[1] An exception is the wound-rotor motor, which is a type of large induction motor that generally has limited application.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

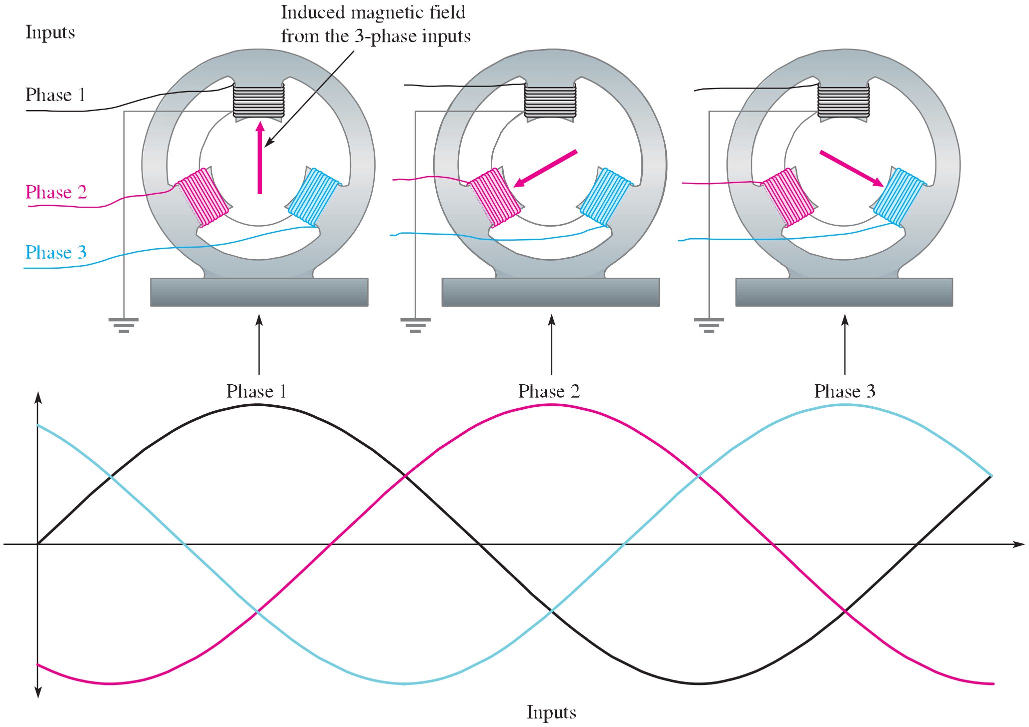

FIGURE 8–45 **The application of three phases to the stator produces a net magnetic field as shown by the red arrow. The rotor (not shown) moves in response to this field.**

As the stator field moves, the rotor moves in sync with it in a synchronous motor but lags behind in an induction motor. The rate the stator field moves is called the *synchronous speed* of the motor.

Induction Motors

The theory of operation is essentially the same for both single-phase and three-phase induction motors. Both types use the rotating field described previously, but the single-phase motor requires starting windings or another method to produce torque for starting the motor, whereas the three-phase motor is self-starting. When starting windings are employed in a single-phase motor, they are removed from the circuit by a mechanical centrifugal switch as the motor speeds up.

The core of the induction motor's rotor consists of an aluminum frame that forms the conductors for the circulating current in the rotor. (Some larger induction motors use copper bars.) The aluminum frame is similar in appearance to the exercise wheel for pet squirrels (common in the early 20th century), so it is aptly called a **squirrel cage**, illustrated in Figure 8–46. The aluminum squirrel cage itself is the *electrical* path; it is embedded within

FIGURE 8–46 **Diagram of a squirrel-cage rotor.**

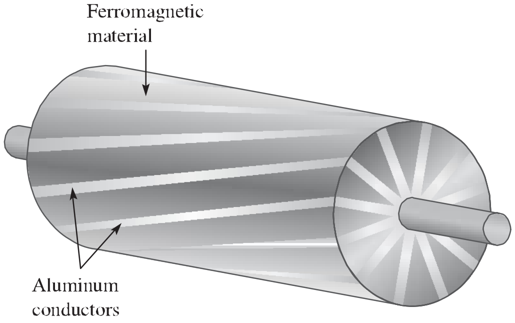

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

a ferromagnetic material to provide a low reluctance *magnetic* path through the rotor. In addition, the rotor has cooling fins that may be molded into the same piece of aluminum as the squirrel cage. The entire assembly must be balanced so that it spins easily and without vibrating.

OPERATION OF AN INDUCTION MOTOR When the magnetic field from the stator moves across the squirrel cage of the inductor, a current is generated in the squirrel cage. This current creates a magnetic field that reacts with the moving field of the stator, causing the rotor to start turning. The rotor will try to "catch-up" with the moving field, but cannot, in a condition known as slip. **Slip** is defined as the difference between the synchronous speed of the stator and the rotor speed. The rotor can never reach the synchronous speed of the stator field because, if it did, it would not cut any field lines and the torque would drop to zero. Without torque, the rotor could not turn itself.

Initially, before the rotor starts moving, there is no back emf, so the stator current is high. As the rotor speeds up, it generates a back emf that opposes the stator current. As the motor speeds up, the torque produced balances the load and the current is just enough to keep the rotor turning. The running current is significantly lower than the initial start-up current because of the back emf. If the load on the motor is then increased, the motor will slow down and generate less back emf. This increases the current to the motor and increases the torque it can apply to the load. Thus, an induction motor can operate over a range of speeds and torque. Maximum torque occurs when the rotor is spinning at about 75% of the synchronous speed.

Synchronous Motors

Recall that an induction motor develops no torque if it runs at the synchronous speed, so it must run slower than the synchronous speed, depending on the load. The synchronous motor will run at the synchronous speed and still develop the required torque for different loads. The only way to change the speed of a synchronous motor is to change the frequency.

6238262 2013/12/10 12:27:66.5

The fact that synchronous motors maintain a constant speed for all load conditions is a major advantage in certain industrial operations and in applications where clock or timing requirements are involved (such as a telescope drive motor or a chart recorder). In fact, the first application of synchronous motors was in electric clocks (in 1917).

Another important advantage to large synchronous motors is their efficiency. Although their original cost is higher than a comparable induction motor, the savings in power will often pay for the cost difference in a few years.

OPERATION OF A SYNCHRONOUS MOTOR Essentially, the rotating stator field of the synchronous motor is identical to that of an induction motor. The primary difference in the two motors is in the rotor. The induction motor has a rotor that is electrically isolated from a supply, and the synchronous motor uses a magnet to follow the rotating stator field. Small synchronous motors use a permanent magnet for the rotor; larger motors use an electromagnet. When an electromagnet is used, dc is supplied from an external source via slip rings as in case of the alternator.

SECTION 8–7 CHECKUP

1. What is the main difference between an induction motor and a synchronous motor?
2. What happens to the magnitude of the rotating stator field as it moves?
3. What is the purpose of a squirrel cage?
4. With reference to motors, what does the term *slip* mean?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

8–8 NONSINUSOIDAL WAVEFORMS

Sine waves are important in electronics, but they are not the only type of ac or time-varying waveform. Two other major types of waveforms are the pulse waveform and the triangular waveform.

After completing this section, you should be able to

- **Identify the characteristics of basic nonsinusoidal waveforms**
  - **Discuss the properties of a pulse waveform**
  - **Define *duty cycle***
  - **Discuss the properties of triangular and sawtooth waveforms**
  - **Discuss the harmonic content of a waveform**

Pulse Waveforms

Basically, a **pulse** can be described as a very rapid transition (**leading edge**) from one voltage or current level (**baseline**) to another level; and then, after an interval of time, a very rapid transition (**trailing edge**) back to the original baseline level. The transitions in level are called *steps*. An ideal pulse consists of two opposite-going steps of equal amplitude. When the leading or trailing edge is positive-going, it is called a **rising edge**. When the leading or trailing edge is negative-going, it is called a **falling edge**.

Figure 8–47(a) shows an ideal positive-going pulse consisting of two equal but opposite instantaneous steps separated by an interval of time called the **pulse width**. Figure 8–47(b) shows an ideal negative-going pulse. The height of the pulse measured from the baseline is its voltage (or current) amplitude. Commonly, analysis is simplified by treating all pulses as ideal (composed of instantaneous steps and perfectly rectangular in shape).

FIGURE 8–47 **Ideal pulses.**

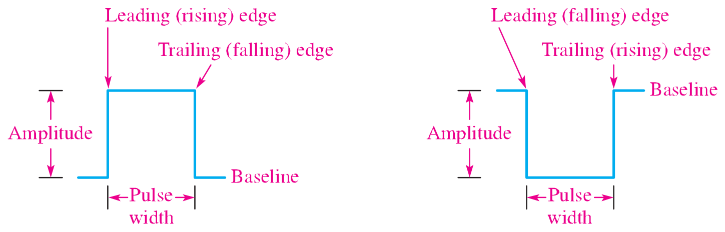

(a) Positive-going pulse

(b) Negative-going pulse

Actual pulses, however, are never ideal. Pulses cannot change from one level to another instantaneously. Time is always required for a transition (step), as illustrated in Figure 8–48(a). As you can see, there is an interval of time during which the pulse is rising from its lower value to its higher value. This interval is called the *rise time*, t_r.

> **Rise time is the time required for the pulse to go from 10% of its amplitude to 90% of its amplitude.**

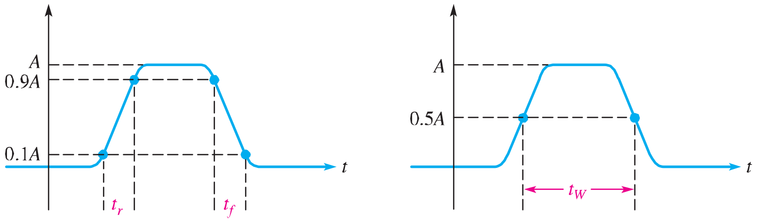

(a) Rise and fall times

(b) Pulse width

FIGURE 8–48 **Nonideal pulse.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The interval of time during which the pulse is falling from its higher value to its lower value is called the *fall time*, t_f.

> **Fall time is the time required for the pulse to go from 90% of its amplitude to 10% of its amplitude.**

Pulse width (t_W) also requires a precise definition for the nonideal pulse because the leading and trailing edges are not vertical.

> **Pulse width is the time between the point on the leading edge where the value is 50% of the amplitude and the point on the trailing edge where the value is 50% of the amplitude.**

Pulse width is shown in Figure 8–48(b).

REPETITIVE PULSES Any waveform that repeats itself at fixed intervals is periodic. Figure 8–49 shows some examples of periodic pulse waveforms. Notice that in each case, the pulses repeat at regular intervals. The rate at which the pulses repeat is the **pulse repetition frequency**, which is the fundamental frequency of the waveform. The frequency can be expressed in hertz or in pulses per second. The time from one pulse to the corresponding point on the next pulse is the period (T). The relationship between frequency and period is the same as with the sine wave, $f = 1/T$.

An important characteristic of periodic pulse waveforms is the duty cycle.

> **The duty cycle is the ratio of the pulse width (t_W) to the period (T) and is usually expressed as a percentage.**

$$\textbf{Percent duty cycle} = \left(\frac{t_W}{T}\right)100\% \qquad (8\text{–}21)$$

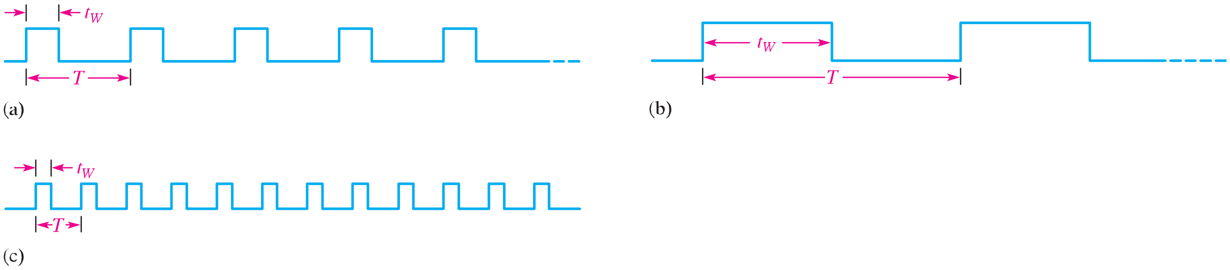

FIGURE 8–49 Repetitive pulse waveforms.

EXAMPLE 8–14

Determine the period, frequency, and duty cycle for the pulse waveform in Figure 8–50.

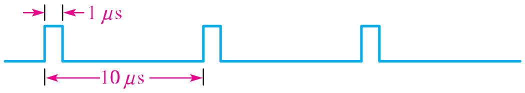

FIGURE 8–50

SOLUTION

As indicated in Figure 8–50, the period is

$$T = \mathbf{10\ \mu s}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Use Equations 8–1 and 8–21 to determine the frequency and duty cycle.

$$f = \frac{1}{T} = \frac{1}{10\ \mu\text{s}} = \mathbf{100\ kHz}$$

$$\text{Percent duty cycle} = \left(\frac{t_W}{T}\right)100\% = \left(\frac{1\ \mu\text{s}}{10\ \mu\text{s}}\right)100\% = \mathbf{10\%}$$

RELATED PROBLEM

A certain pulse waveform has a frequency of 200 kHz and a pulse width of 0.25 μs. Determine the duty cycle expressed as a percentage.

SQUARE WAVES A square wave is a pulse waveform with a duty cycle of 50%. Thus, the pulse width is equal to one-half of the period. A square wave is shown in Figure 8–51.

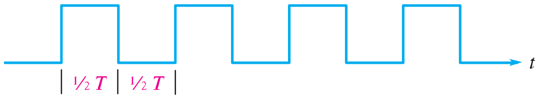

FIGURE 8–51 **Square wave.**

THE AVERAGE VALUE OF A PULSE WAVEFORM The **average value** of a pulse waveform is equal to its baseline value plus the product of its duty cycle and its amplitude. The lower level of a positive-going waveform or the upper level of a negative-going waveform is taken as the baseline. The formula for average voltage is as follows:

$$V_{\text{avg}} = \textbf{baseline} + \textbf{(duty cycle)(amplitude)} \tag{8–22}$$

The following example illustrates the calculation of the average voltage of pulse waveforms.

EXAMPLE 8–15

Determine the average voltage of each of the positive-going waveforms in Figure 8–52.

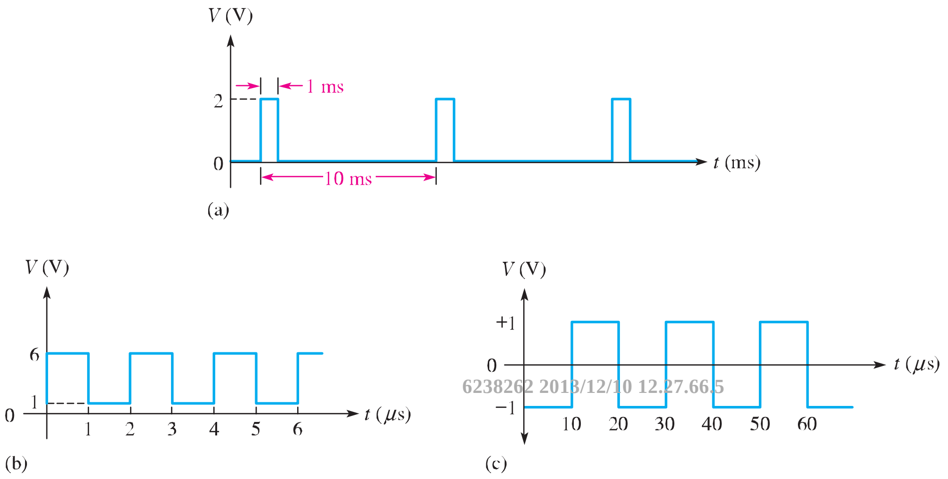

6238262 2018/12/10 12.27.66.5

FIGURE 8–52

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

In Figure 8–52(a), the baseline is at 0 V, the amplitude is 2 V, and the duty cycle is 10%. The average voltage is

$$\begin{aligned} V_{\text{avg}} &= \text{baseline} + (\text{duty cycle})(\text{amplitude}) \\ &= 0\text{ V} + (0.1)(2\text{ V}) = \mathbf{0.2\ V} \end{aligned}$$

The waveform in Figure 8–52(b) has a baseline of +1 V, an amplitude of 5 V, and a duty cycle of 50%. The average voltage is

$$\begin{aligned} V_{\text{avg}} &= \text{baseline} + (\text{duty cycle})(\text{amplitude}) \\ &= 1\text{ V} + (0.5)(5\text{ V}) = 1\text{ V} + 2.5\text{ V} = \mathbf{3.5\ V} \end{aligned}$$

The waveform in Figure 8–52(c) is a square wave with a baseline of −1 V and an amplitude of 2 V. The duty cycle is 50%. The average voltage is

$$\begin{aligned} V_{\text{avg}} &= \text{baseline} + (\text{duty cycle})(\text{amplitude}) \\ &= -1\text{ V} + (0.5)(2\text{ V}) = -1\text{ V} + 1\text{ V} = \mathbf{0\ V} \end{aligned}$$

This is an alternating square wave, and, like an alternating sine wave, it has an average voltage of zero over a full cycle.

RELATED PROBLEM

If the baseline of the waveform in Figure 8–52(a) is shifted to +1 V, what is the average voltage?

Triangular and Sawtooth Waveforms

Triangular and sawtooth waveforms are formed by voltage or current ramps. A **ramp** is a linear increase or decrease in the voltage or current. Figure 8–53 shows both positive- and negative-going ramps. In part (a), the ramp has a positive slope; in part (b), the ramp has a negative slope. The slope of a voltage ramp is $\pm V/t$ and the slope of a current ramp is $\pm I/t$.

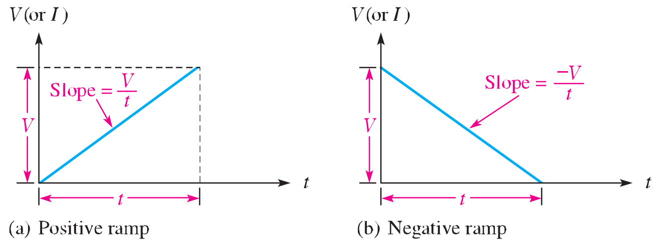

(a) Positive ramp (b) Negative ramp

FIGURE 8–53 **Voltage ramps.**

EXAMPLE 8–16

What are the slopes of the voltage ramps in Figure 8–54?

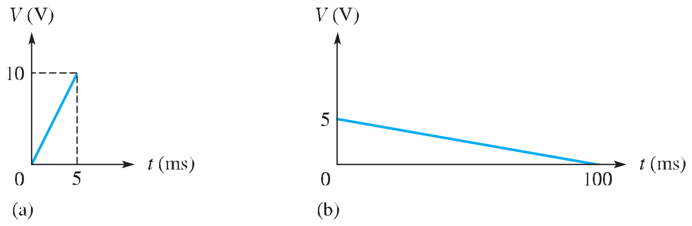

FIGURE 8–54

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

In Figure 8–54(a), the voltage increases from 0 V to +10 V in 5 ms. Thus, $V = 10\text{ V}$ and $t = 5\text{ ms}$. The slope is

$$\frac{V}{t} = \frac{10\text{ V}}{5\text{ ms}} = \mathbf{2\ V/ms}$$

In Figure 8–54(b), the voltage decreases from +5 V to 0 V in 100 ms. Thus, $V = -5\text{ V}$ and $t = 100\text{ ms}$. The slope is

$$\frac{V}{t} = \frac{-5\text{ V}}{100\text{ ms}} = \mathbf{-0.05\ V/ms}$$

RELATED PROBLEM

A certain voltage ramp has a slope of +12 V/μs. If the ramp starts at zero, what is the voltage at 0.01 ms?

TRIANGULAR WAVEFORMS Figure 8–55 shows that a **triangular waveform** is composed of positive-going and negative-going ramps having equal slopes. The period of this waveform can be measured from one peak to the next corresponding peak, as illustrated. This particular triangular waveform is alternating and has an average value of zero.

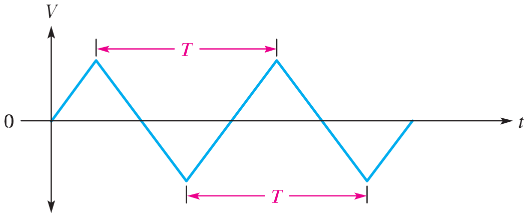

FIGURE 8–55 **Alternating triangular waveform.**

Figure 8–56 depicts a triangular waveform with a nonzero average value. The frequency for triangular waves is determined in the same way as for sine waves, that is, $f = 1/T$.

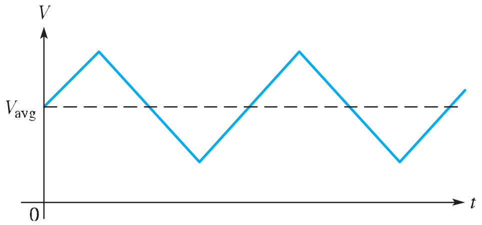

FIGURE 8–56 **Nonalternating triangular waveform.**

SAWTOOTH WAVEFORMS The **sawtooth waveform** is actually a special case of the triangular waveform consisting of two ramps, one of much longer duration than the other. Sawtooth waveforms are used in many electronic systems. For example, a sawtooth waveform is used in automatic test equipment, control systems, and certain types of displays, including analog oscilloscopes.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Figure 8–57 is an example of a sawtooth waveform. Notice that it consists of a positive-going ramp of relatively long duration, followed by a negative-going ramp of relatively short duration.

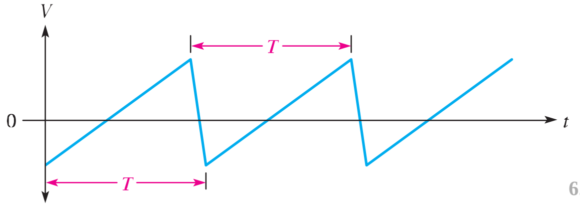

6238262 2013/12/10 12.27.66.5

FIGURE 8–57 **Alternating sawtooth waveform.**

Harmonics

A repetitive nonsinusoidal waveform contains sinusoidal waveforms with a fundamental frequency and harmonic frequencies. The **fundamental frequency** is the repetition rate of the waveform, and the **harmonics** are higher-frequency sine waves that are multiples of the fundamental.

ODD HARMONICS *Odd harmonics* are frequencies that are odd multiples of the fundamental frequency of a waveform. For example, a 1 kHz square wave consists of a fundamental of 1 kHz and odd harmonics of 3 kHz, 5 kHz, 7 kHz, and so on. The 3 kHz frequency in this case is the third harmonic; the 5 kHz frequency is the fifth harmonic; and so on.

EVEN HARMONICS *Even harmonics* are frequencies that are even multiples of the fundamental frequency. For example, if a certain wave has a fundamental of 200 Hz, the second harmonic is 400 Hz, the fourth harmonic is 800 Hz, the sixth harmonic is 1200 Hz, and so on. These are even harmonics.

COMPOSITE WAVEFORM Any variation from a pure sine wave produces harmonics. A nonsinusoidal wave is a composite of the fundamental and the harmonics. Some types of waveforms have only odd harmonics, some have only even harmonics, and some contain both. The shape of the wave is determined by its harmonic content. Generally, only the fundamental frequency and the first few harmonics are of significant importance in determining the wave shape.

A square wave is an example of a waveform that consists of a fundamental frequency and only odd harmonics. When the instantaneous values of the fundamental and each odd harmonic are added algebraically at each point, the resulting curve will have the shape of a square wave, as illustrated in Figure 8–58. In part (a), the fundamental and the third harmonic

HANDS ON TIP
The frequency response of an oscilloscope limits the accuracy with which waveforms can be accurately displayed. To view pulse waveforms, the frequency response must be high enough for all significant harmonics of the waveform. For example, a 100 MHz oscilloscope distorts a 100 MHz pulse waveform because the third, fifth, and higher harmonics are greatly attenuated.

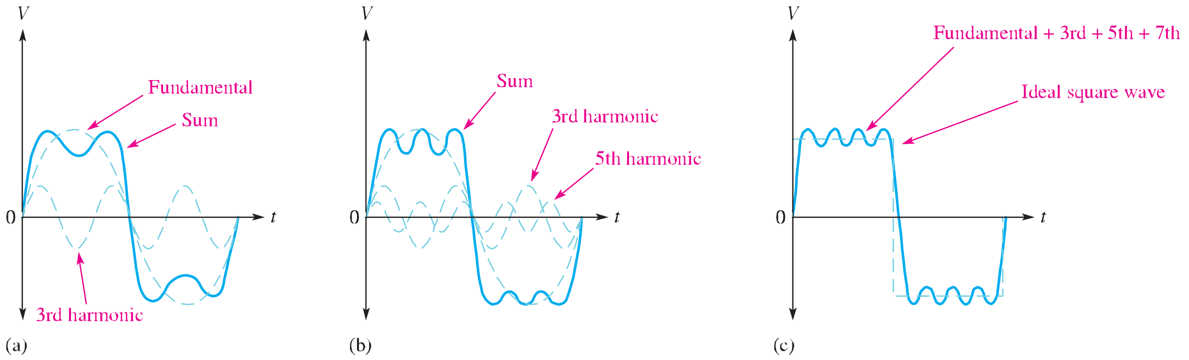

FIGURE 8–58 **Odd harmonics combine to produce a square wave. The plot is for a continuous pulse that theoretically extends from $-\infty$ to $+\infty$.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

produce a wave shape that begins to resemble a square wave. In part (b), the fundamental, third, and fifth harmonics produce a closer resemblance. When the seventh harmonic is included, as in part (c), the resulting wave shape becomes even more like a square wave. As more harmonics are included, a square wave is approached.

SYSTEM EXAMPLE 8–3

ANALYZING SIGNALS IN THE FREQUENCY DOMAIN

As you have learned, signals can be thought of as being composed of various sine waves. All waves can be broken into sine waves that are related to each other. A plot of the amplitude of the sine waves as a function of their frequency is called a frequency domain plot. (The domain refers to the independent variable.) The same wave can be plotted with time as the independent variable, which is referred to as a time domain plot. Figure 8–59 shows the time and frequency domain views of a periodic signal. Notice that the time domain view shows a summation of all of the frequencies in the signal and the frequency domain view breaks the signal into its constituent frequencies.

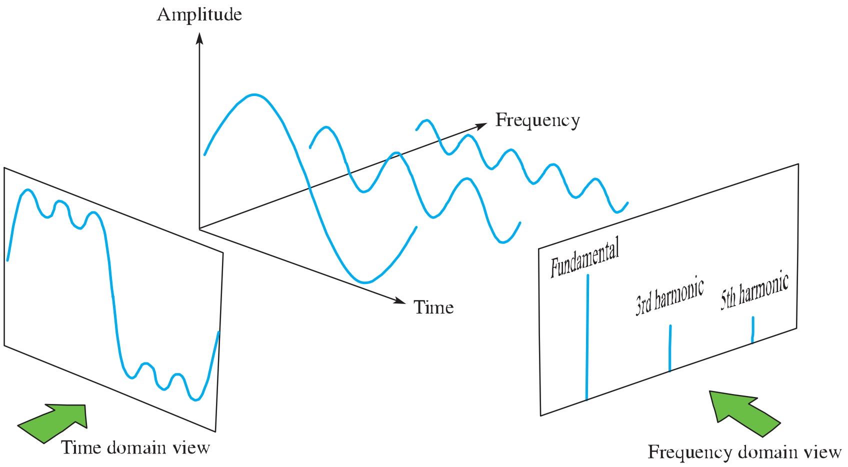

FIGURE 8–59 Time and frequency domain views of a signal.

The time and frequency domains are the two principal domains for describing electronic signals. The relationship between the time and frequency domains was first described mathematically by Joseph Fourier in 1807 in describing heat flow, so the conversion between domains is called Fourier analysis.

The spectrum analyzer is an instrument that is widely used in radio-frequency (rf) and microwave systems. A spectrum analyzer breaks the signal of interest into its sinusoidal components and displays these components to the user. The oscilloscope, which is introduced in Section 8–9, is a time domain instrument (although some specialized oscilloscopes can display the frequency domain).

Many systems use radio-frequency devices including cell phones, wireless LANs, certain remote controls, medical devices, and various tracking devices (for tracking packages or even livestock!). As a result of these applications, the demand for bandwidth has increased significantly and so have problems with interference between various systems. Testing for interfering frequencies is the type of frequency domain measurement ideally suited to the spectrum analyzer. It is also useful for measurements of frequency, power, modulation, distortion, and noise levels.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 8–8 CHECKUP

1. Define the following parameters:

(a) rise time **(b)** fall time **(c)** pulse width

2. In a certain repetitive positive-going pulse waveform, the pulses are 200 μs wide and occur once every millisecond. What is the frequency of this waveform?

3. Determine the duty cycle, amplitude, and average value of the waveform in Figure 8–60(a).

4. What is the period of the triangular wave in Figure 8–60(b)?

5. What is the frequency of the sawtooth wave in Figure 8–60(c)?

6. Define *fundamental frequency*.

7. What is the second harmonic of a fundamental frequency of 1 kHz?

8. What is the fundamental frequency of a square wave having a period of 10 μs?

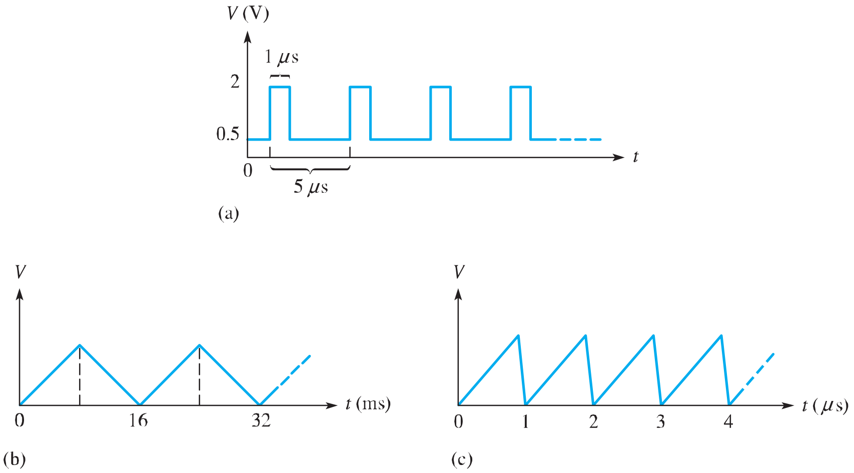

FIGURE 8–60

8–9 THE OSCILLOSCOPE

The oscilloscope, or (scope for short) is a widely used and versatile test instrument for observing and measuring waveforms.

After completing this section, you should be able to

- **Use an oscilloscope to measure waveforms**
  - **Recognize common oscilloscope controls**
  - **Measure the amplitude of a waveform**
  - **Measure the period and frequency of a waveform**

The **oscilloscope** is basically a graph-displaying device that traces a graph of a measured electrical signal on its screen. In most applications, the graph shows how signals change over time. The vertical axis of the display screen represents voltage and the horizontal axis represents time. You can measure amplitude, period, and frequency of a signal using an oscilloscope. Also, you can determine the pulse width, duty cycle, rise time, and fall time of a pulse waveform. Most scopes can display at least two signals on the screen at one time, enabling you to observe their time relationship. Two different digital oscilloscopes are shown in Figure 8–61.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

FIGURE 8–61 **Digital oscilloscopes.** Copyright © Tektronix. Reproduced by permission.

Two basic types of oscilloscopes, analog and digital, can be used to view digital waveforms. The older analog scope works by applying the measured waveform directly to control the up and down motion of the electron beam in the cathode-ray tube (CRT) as it sweeps across the screen. As a result, the beam traces out the waveform pattern on the screen. The digital scope converts the measured waveform to digital information by a sampling process in an analog-to-digital converter (ADC). The digital information is then used to reconstruct the waveform on the screen.

The digital scope is more widely used than the analog scope. However, either type can be used in many applications; each has characteristics that make it more suitable for certain situations. An analog scope displays waveforms as they occur in “real time.” Digital scopes are useful for measuring transient pulses that may occur randomly or only once. Also, because information about the measured waveform can be stored in a digital scope, it may be viewed at some later time, printed out, or thoroughly analyzed by a computer or other means.

Basic Operation of Analog Oscilloscopes

To measure a voltage, a probe must be connected from the scope to the point in a circuit at which the voltage is present. Generally, a ×10 probe is used that reduces (attenuates) the signal amplitude by ten. The signal goes through the probe into the vertical circuits where it is either further attenuated (reduced) or amplified depending on the actual amplitude and on where you set the vertical control of the scope. The vertical circuits then drive the vertical deflection plates of the CRT. Also, the signal goes to the trigger circuits that trigger the horizontal circuits to initiate repetitive horizontal sweeps of the electron beam across the screen using a sawtooth waveform. There are many sweeps per second so that the beam appears to form a solid line across the screen in the shape of the waveform. This basic operation is illustrated in Figure 8–62.

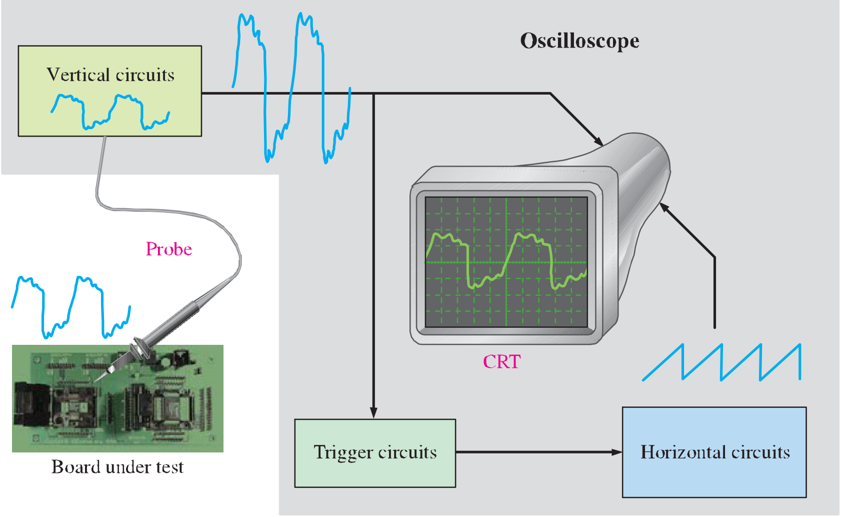

FIGURE 8–62 **Block diagram of an analog oscilloscope.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Basic Operation of Digital Oscilloscopes

Some parts of a digital scope are similar to the analog scope. However, the digital scope is more complex than an analog scope and typically has an LCD screen rather than a CRT. Rather than displaying a waveform as it occurs, the digital scope first acquires the measured analog waveform and converts it to a digital format using an analog-to-digital converter (ADC). The digital data is stored and processed. The data then goes to the reconstruction and display circuits for display in its original analog form. Figure 8–63 shows a block diagram for a digital oscilloscope.

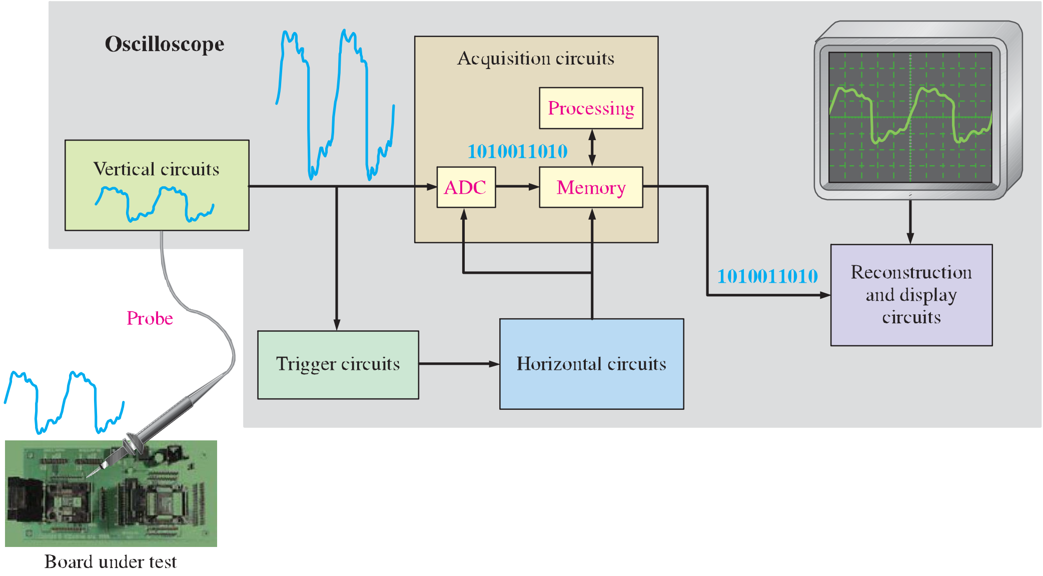
Oscilloscope
Acquisition circuits
Processing
1010011010
ADC
Memory
Vertical circuits
1010011010
Reconstruction and display circuits
Probe
Trigger circuits
Horizontal circuits
Board under test

FIGURE 8–63 Block diagram of a digital oscilloscope.

Oscilloscope Controls

A front panel view of a typical dual-channel oscilloscope is shown in Figure 8–64. Instruments vary depending on model and manufacturer, but most have certain common features.

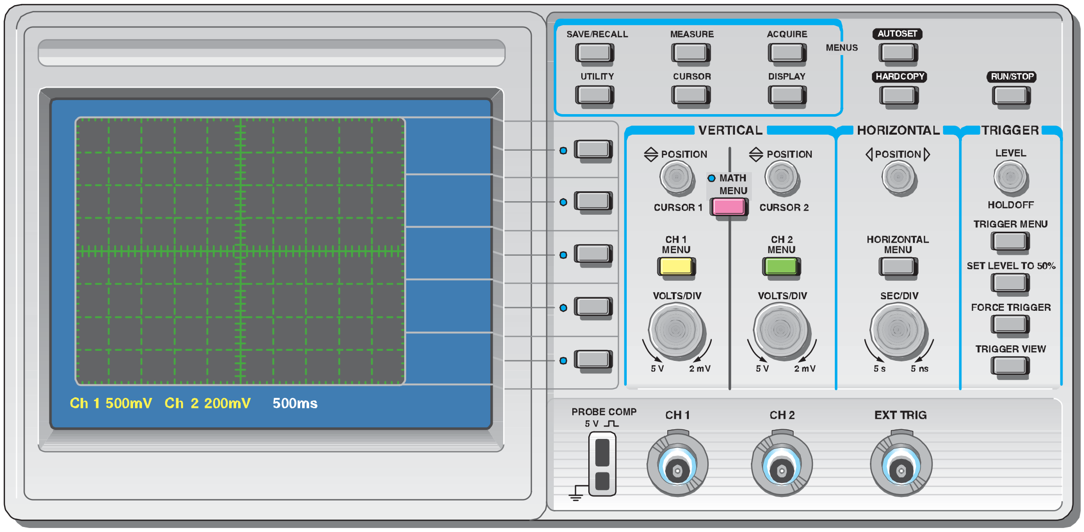
SAVE/RECALL
MEASURE
ACQUIRE
MENUS
AUTOSET
UTILITY
CURSOR
DISPLAY
HARDCOPY
RUN/STOP
VERTICAL
HORIZONTAL
TRIGGER
POSITION
POSITION
POSITION
LEVEL
MATH MENU
CURSOR 1
CURSOR 2
HOLDOFF
TRIGGER MENU
CH 1 MENU
CH 2 MENU
HORIZONTAL MENU
SET LEVEL TO 50%
VOLTS/DIV
VOLTS/DIV
SEC/DIV
FORCE TRIGGER
5 V
2 mV
5 V
2 mV
5 s
5 ns
TRIGGER VIEW
Ch 1 500mV Ch 2 200mV 500ms
PROBE COMP 5 V
CH 1
CH 2
EXT TRIG

FIGURE 8–64 A typical dual-channel oscilloscope. Numbers below screen indicate the values for each division on the vertical (voltage) and horizontal (time) scales and can be varied using the vertical and horizontal controls on the scope.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

For example, the two vertical sections contain a Position control, a channel menu button, and a Volts/Div control. The horizontal section contains a Sec/Div control.

Some of the main controls are now discussed. Refer to the user manual for complete details of your particular scope.

VERTICAL CONTROLS In the Vertical section of the scope in Figure 8–64, there are identical controls for each of the two channels (CH1 and CH2). The position control lets you move a displayed waveform up or down vertically on the screen. The Menu button provides for the selection of several items that appear on the screen, such as the coupling modes (ac, dc, or ground) and coarse or fine adjustment for the Volts/Div. The Volts/Div control adjusts the number of volts represented by each vertical division on the screen. The Volts/Div setting for each channel is displayed on the bottom of the screen. The Math Menu button provides a selection of operations that can be performed on the input waveforms, such as subtraction and addition of signals.

HORIZONTAL CONTROLS In the Horizontal section, the controls apply to both channels. The Position control lets you move a displayed waveform left to right horizontally on the screen. The Horizontal Menu button provides for the selection of several items that appear on the screen such as the main time base, expanded view of a portion of a waveform, and other parameters. The Sec/Div control adjusts the time represented by each horizontal division or main time base. The Sec/Div setting is displayed at the bottom of the screen.

TRIGGER CONTROLS In the Trigger section, the Level control determines the point on the triggering waveform where triggering occurs to initiate the sweep to display input waveforms. The Trigger Menu button provides for the selection of several items that appear on the screen including edge or slope triggering, trigger source, trigger mode, and other parameters. There is also an input for an external trigger signal.

Triggering stabilizes a waveform on the screen and properly triggers on a pulse that occurs only one time or randomly. Also it allows you to observe time delays between two waveforms. Figure 8–65 compares a triggered to an untriggered signal. The untriggered signal tends to drift across the screen producing what appears to be multiple waveforms.

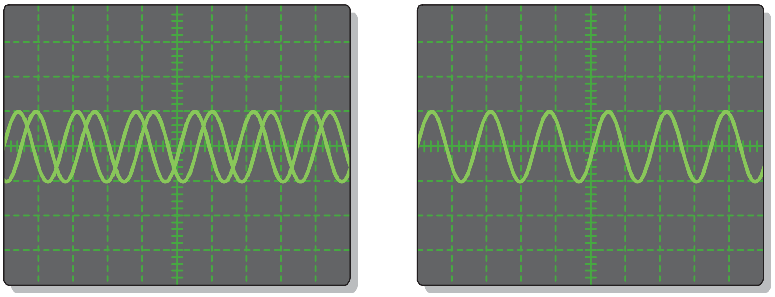

(a) Untriggered waveform display (b) Triggered waveform display

FIGURE 8–65 **Comparison of an untriggered and a triggered waveform on an oscilloscope.**

COUPLING A SIGNAL INTO THE SCOPE Coupling is the method used to connect a signal voltage to be measured into the oscilloscope. The DC and AC coupling modes are selected from the Vertical menu. DC coupling allows a waveform including its dc component to be displayed. AC coupling blocks the dc component of a signal so that you see the waveform centered at 0 V. The Ground mode allows you to connect the channel input to ground to see where the 0 V reference is on the screen. Figure 8–66 illustrates the result of DC and AC coupling using a sinusoidal waveform that has a dc component.

The voltage probe, shown in Figure 8–67, is used for connecting a signal to the scope. Since all instruments tend to affect the circuit being measured due to loading, most scope probes provide an attenuation network to minimize loading effects. Probes that attenuate

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

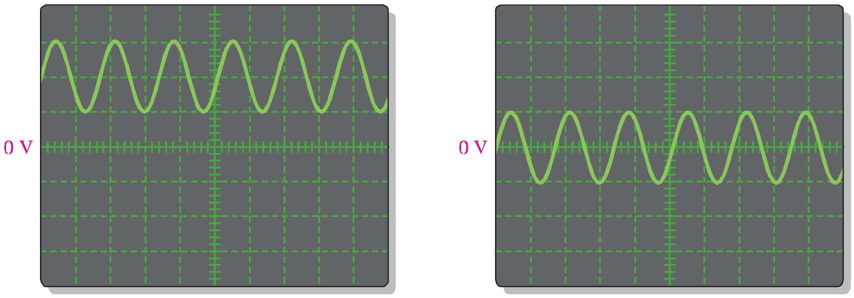

(a) DC coupled waveform (b) AC coupled waveform

FIGURE 8–66 **Displays of the same waveform having a dc component.**

the measured signal by a factor of 10 are called ×10 (times ten) probes. Probes with no attenuation are called ×1 (times one) probes. Most oscilloscopes automatically adjust the calibration for the attenuation of the type of probe being used. For most measurements, the ×10 probe should be used. However, if you are measuring very small signals, a ×1 may be the best choice.

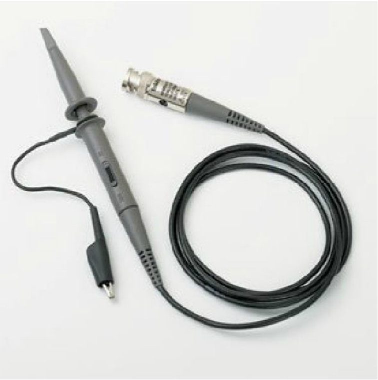

FIGURE 8–67 **An oscilloscope voltage probe.** Copyright © Tektronix, Inc. Reproduced by permission.

The probe has an adjustment that allows you to compensate for the input capacitance of the scope. Most scopes have a probe compensation output that provides a calibrated square wave for probe compensation. Before making a measurement, you should make sure that the probe is properly compensated to eliminate any distortion introduced. Typically, there is a screw or other means of adjusting compensation on a probe. Figure 8–68 shows scope waveforms for three probe conditions: properly compensated, undercompensated, and overcompensated. If the waveform appears either over- or undercompensated, adjust the probe until the properly compensated square wave is achieved.

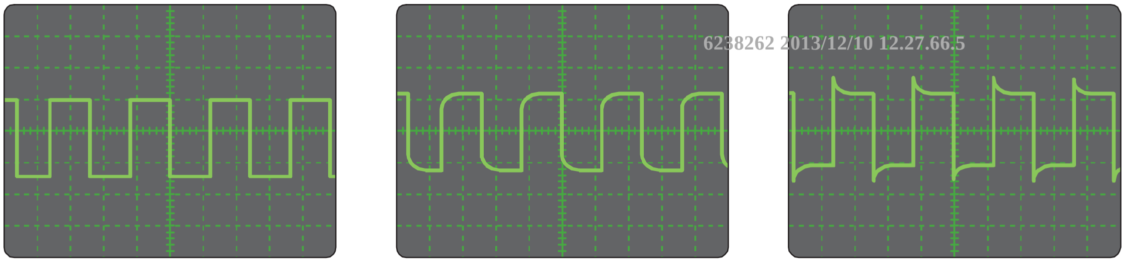
6238262 2013/12/10 12.27.66.5

Properly compensated Undercompensated Overcompensated

FIGURE 8–68 **Probe compensation conditions.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 8–17

Determine the peak-to-peak value and period of each sine wave in Figure 8–69 from the digital scope screen displays and the settings for Volts/Div and Sec/Div, which are indicated under the screens. Sine waves are centered vertically on the screens.

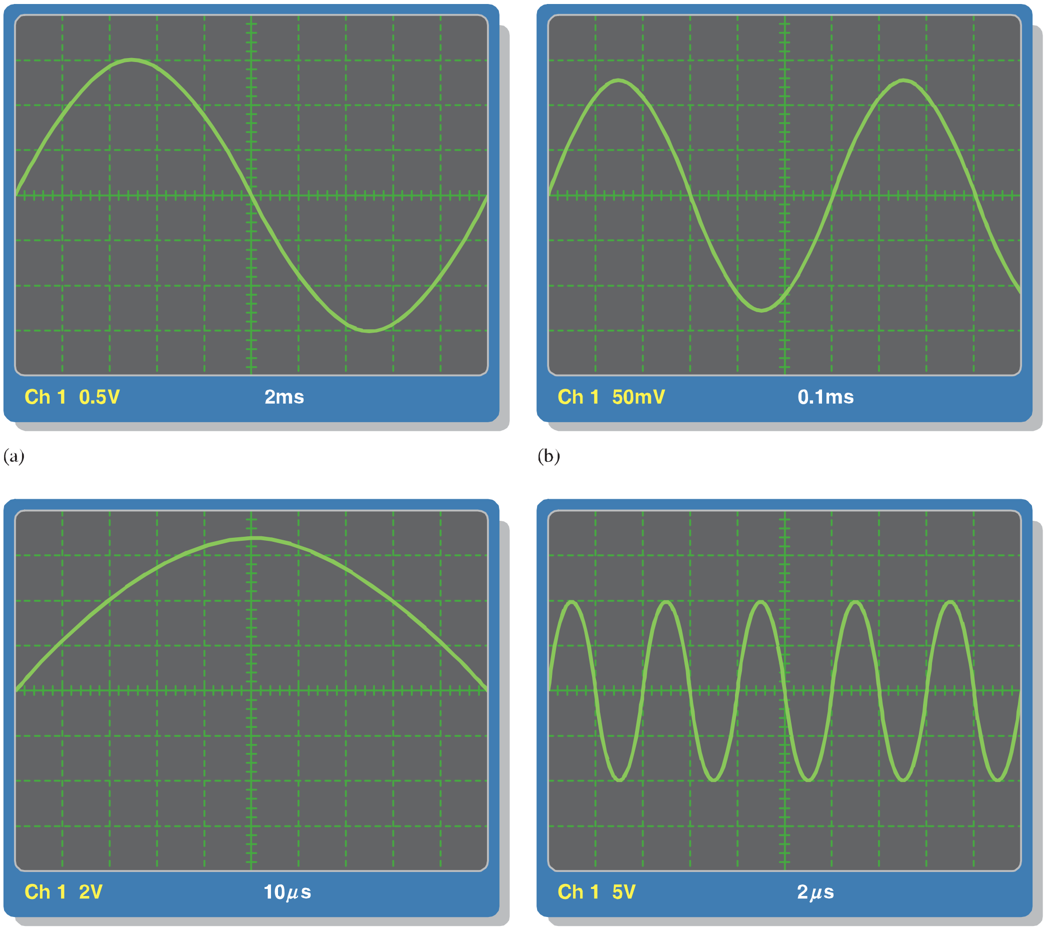

(a) (b) (c) (d)

FIGURE 8–69

SOLUTION

Looking at the vertical scale in Figure 8–69(a),

$$V_{pp} = 6 \text{ divisions} \times 0.5 \text{ V/division} = \mathbf{3.0 \text{ V}}$$

From the horizontal scale (one cycle covers ten divisions),

$$T = 10 \text{ divisions} \times 2 \text{ ms/division} = \mathbf{20 \text{ ms}}$$

Looking at the vertical scale in Figure 8–69(b),

$$V_{pp} = 5 \text{ divisions} \times 50 \text{ mV/division} = \mathbf{250 \text{ mV}}$$

From the horizontal scale (one cycle covers six divisions),

$$T = 6 \text{ divisions} \times 0.1 \text{ ms/division} = 0.6 \text{ ms} = \mathbf{600\ \mu s}$$

Looking at the vertical scale in Figure 8–69(c),

$$V_{pp} = 6.8 \text{ divisions} \times 2 \text{ V/division} = \mathbf{13.6 \text{ V}}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

From the horizontal scale (one-half cycle covers ten divisions),

$$T = 20 \text{ divisions} \times 10\ \mu\text{s/division} = \mathbf{200\ \mu s}$$

Looking at the vertical scale in Figure 8–69(d),

$$V_{pp} = 4 \text{ divisions} \times 5 \text{ V/division} = \mathbf{20\ V}$$

From the horizontal scale (one cycle covers two divisions),

$$T = 2 \text{ divisions} \times 2\ \mu\text{s/division} = \mathbf{4\ \mu s}$$

RELATED PROBLEM

Determine the rms value and the frequency for each waveform displayed in Figure 8–69.

SECTION 8–9 CHECKUP

1. What is the main difference between a digital and an analog oscilloscope?
2. Is voltage read horizontally or vertically on a scope screen?
3. What does the Volts/Div control on an oscilloscope do?
4. What does the Sec/Div control on an oscilloscope do?
5. When should you use a ×10 probe for making a voltage measurement?

8–10 SIGNAL SOURCES

The function generator was introduced in Section 8–1. In this section, we'll look in more depth at function generators and other signal sources, which are commonly used for testing circuits. The function generator is found at nearly every electronic workbench. Other signal sources include specialized generators and digital pattern generators.

After completing this section, you should be able to

- **Describe the types of signal sources and explain the purpose of typical controls**
  - **Discuss types of signal sources and their application**
  - **Describe important specifications for signal sources**
  - **Describe the function selection, frequency/amplitude adjustment controls, and the dc offset/duty cycle control for a typical function generator**

Types of Signal Sources

Signals can be generated from a variety of low-power electronic instruments, ranging from basic sine wave oscillators to generators that have a selection of waveforms including sine waves, ramps, pulses and even arbitrary waveforms from specialized generators. Except for low-end instruments, the output frequency and amplitude are calibrated and can be varied over a specified range. Instruments for producing waveforms for test are generally categorized as low frequency, radio frequency, and microwave. Traditionally, low-frequency generators extend from dc to about 1 MHz. Radio-frequency sources generally fall in the region from 100 kHz to 1 GHz, and microwave sources are from 1 GHz up. The wide variety of generators on the market has blurred the distinction between these traditional ranges; the range may vary depending on the wave shape.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Some important types of signal sources are function generators, arbitrary function generators (AFGs), signal generators (sometimes referred to as rf generators), swept-frequency oscillators (or sweepers), pulse generators, and digital-pattern generators (DPGs).

Function generators Function generatorsuse a free-running oscillator to provide a selection of sine, square, and triangle waveforms in a single instrument. Function generators may also produce pulses, ramps, and other waveforms.

Arbitrary function generators *(AFGs)* Arbitrary function generators are the most versatile signal generators. They are digital synthesizers that allow the user to create custom waveforms from a mathematical formula, a captured waveform, or simulation software. AFGs create an analog waveform using a controller and a digital-to-analog converter (DAC).

Signal generators Signal generators produce high-frequency sine waves and modulated sine waves. A large variety of signal generators are available with frequencies that go from dc to microwave, and many use frequency synthesizers to generate precise waveforms. A frequency synthesizer generates waveforms digitally from a fixed-reference oscillator.

Swept-frequency oscillators Swept-frequency oscillators generate a sine wave output that varies at a cyclic rate between two selected frequencies. They allow the user to view signals directly in the frequency domain. Some models produce a logarithmic or other nonlinear sweep output.

Pulse generators Pulse generators are specialized instruments for producing pulses with fast rise times over a wide frequency range and with varying duty cycles. They frequently have outputs for specific logic levels. Most pulse generators allow the user to control a variety of pulse parameters, such as amplitude, rise time, offsets, triggering, and polarity. In addition, some units produce pulse pairs and trains.

Digital-pattern generators *(DPGs)* Digital-pattern generators produce a digital sequence for testing digital equipment. Some DPGs can generate a series of standard test patterns for performance checks.

Swept-Frequency Measurements

Many times a system needs to be tested for its frequency response. Swept-frequency oscillators are a type of signal generator that has an output that can be set for a given range depending on the system under test. For example, a radar transmitter may have multiple stages that need to work together to produce a desired output. The frequencies are in the microwave region, so the generator is a specialized microwave oscillator.

A swept-frequency oscillator can generate a small high-frequency signal that is used for tuning purposes. The response is converted to a signal that can be observed on an oscilloscope while the system is tuned in real time.

SYSTEM NOTE

Specifications for Signal Generators

The most important specifications for signal generators are those for mode, frequency range, spectral purity, amplitude range, modulation, and output impedance. These parameters are discussed here:

Mode The mode is a specification of signal types that can be output. Specialized generators may only have one mode, such as a sine wave. Function generators will have several modes (waveform types) to choose from. Typically these are sine, square, triangle, positive and negative pulse, and positive and negative ramps. The duty cycle, which is the on/off time for a pulse, may be adjustable.

Frequency range This is the span of frequencies over which instrument performance is specified. Accuracy and resolution limits are also generally given with the frequency range

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

specification. The accuracy specification depends on the internal oscillator's accuracy as well as the type of dial controls (mechanical or digital). A precise digital control is of little value if the internal oscillator is not accurate, nor is it meaningful to have an accurate oscillator if the control is an imprecise mechanical dial.

Spectral purity Recall that a pure sine wave consists of a single fundamental line in the frequency domain. One type of distortion called *phase noise* can broaden the fundamental line, meaning the signal is not a pure sine wave. Another distortion is called *harmonic distortion*, which means multiples of the fundamental frequency are present in the output. These and other noise sources can create problems with certain system tests, particularly for verifying communication equipment. Even if the equipment is working properly, a signal generator with distortion can make the circuit under test to look worse than it is.

Amplitude range This specification gives the output voltage amplitude range or the maximum and minimum power delivered to a specified load. In addition to the ac voltage amplitude range, output specifications include the dc offset range as well as the accuracy, resolution, and flatness of the output across the frequency range. A related adjustment is the amount of dc offset. The dc offset range is either a positive or negative dc voltage that can be added to the output; it is typically a variable amount up to about 10 V.

Modulation **Modulation** is the process in which a signal containing information is used to modify the characteristic of another signal, such as amplitude frequency or pulse width, so that the information on the first is also contained in the second. A signal generator with modulation allows the user to change a high-frequency waveform (called a *carrier* in communications systems) by a lower-frequency waveform called the *modulating signal.* The modulation signal can be supplied from an external source or may be generated within the generator; typically, a sine, square, triangle, or ramp can be selected as the modulation signal. The parameter being modulated can be the high-frequency signal's amplitude, frequency, or phase. The various modulation methods are useful for testing the performance of different communication systems.

Output impedance No matter how complicated the internal circuitry, signal generators can be modeled as a Thevenin circuit, which consists of a voltage source in series with a resistance. The output impedance is equivalent to the resistance in series with an ideal source. Generators are normally marked with the output impedance on the output terminals. Typical values are 50 Ω, 75 Ω, or 600 Ω. Keep in mind that connecting a load to a source that has a finite output resistance will change the output voltage amplitude of the generator.

The change in output voltage amplitude can affect the desired measurement to the point of making it meaningless if it is not taken into account. For example, to measure the frequency response of a circuit without having the generator affect the results, the amplitude must be maintained at a constant value from the generator. The output of the circuit under test is monitored while the frequency is changed. Each change in frequency changes the input impedance of the amplifier and causes a different loading effect to occur. To prevent input loading from affecting the output, the amplitude of the generator should be readjusted to the same value each time a new frequency is tested.

Waveform Modes

Depending on the complexity, waveform generators can have from one to several waveform modes from which to choose. Common waveform modes are as follows.

- **Continuous** The output is steady at a specific frequency, amplitude, and offset. This is the basic output of signal generators.
- **Triggered** The output is initiated by an internal or external (user-supplied) trigger. The trigger can be generated by external equipment, a manual pushbutton, or sent over a control bus.
- **Burst** This is the same as the triggered mode, except the output is programmed for a specific number of cycles at a time, which can be repeated as set by a separate oscillator. The burst can be triggered manually or with a command in programmable instruments.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

- **Gated** The output is enabled for the duration of an external gate signal.
- **Swept** The frequency of the output changes in some predetermined way. Sweeps can be triggered or programmed (start-and-stop frequency), linear, or logarithmic.
- **Modulation** The output is modulated by an internal or external waveform. Modulated outputs are useful for testing communication systems.

The Basic Function Generator

Function generators are characterized by the various waveforms they produce; they all produce sine, square, and triangle waves and may also include pulses and ramp (sawtooth) waveforms. Sine and square waves are commonly used for general-purpose testing of circuits such as amplifiers. Pulses that have a duty cycle that is not near 50% are useful for digital testing because it is easy to tell if the pulses have been inverted; they don't have the same appearance when inverted as when not. (If a square wave is inverted, it still looks like a square wave.) A versatile function generator is the Agilent 33220A, shown in Figure 8–70. This generator includes more functions than a basic generator, such as the ability to store waveforms in a memory. To set up a standard waveform on the Agilent 33220A, you can set the controls as indicated in the figure.

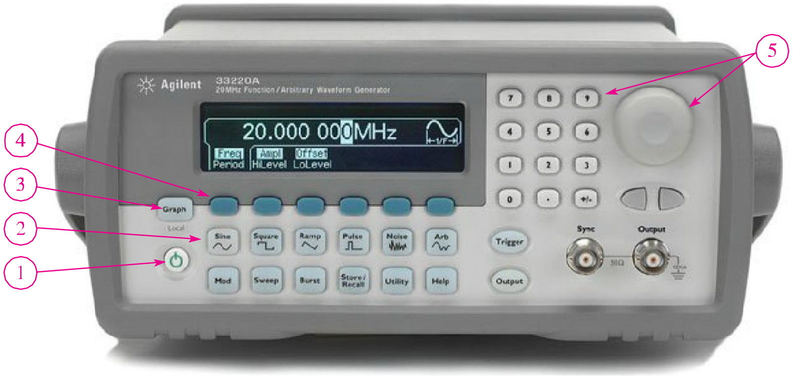

1. Press the power switch to turn on power.
2. Choose a function switch to select the type of waveform (sinusoidal, pulse, etc.). The function key will be lighted.
3. Use graph mode to view all setup parameters at once.
4. Select a pushbutton to adjust a function (frequency, amplitude, etc.)
5. Adjust the selected function with the keypad or the circular knob.

FIGURE 8–70 **The Agilent 33220A function generator.** (© Agilent Technologies, Inc. 2012, Reproduced with Permission, Courtesy of Agilent Technologies, Inc.)

An option for function generators is a display. Some models do not include a display, and the user verifies the generator signal using an oscilloscope. Others, such as the Agilent 33220A have a digital readout that indicates the precise frequency (or period), offset, etc. and shows a visual indication of the waveform mode. It also has a graph mode that shows a graphical view of the selected waveform with a soft key menu of available parameters that can be displayed (frequency, period, rise time, etc.).

Function generators often have outputs that are designed for testing digital systems. Digital systems are characterized by two voltage levels, called HIGH and LOW, which are determined by the type of logic to be tested. The pulse output may be set by the manufacturer or may be set by the user depending on the generator. Digital systems use fast signals, so the rise time specification is important. This specification is the time required for the signal to make a 10% to 90% change in the output level.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Another option available for function generators is the ability to control the output with a computer or controller. This is particular useful in automatic test systems. The system can change the output based on the particular requirements for a test.

EXAMPLE 8–18

Assume a function generator is showing the signal shown on the oscilloscope display in Figure 8–71. The oscilloscope controls are set as shown. Determine the type of waveform that is selected on the function generator, the frequency setting, and the peak-to-peak amplitude. Often the peak-to-peak is the simplest measurement with an oscilloscope and, as the name implies, is measured between the maximum excursions of the waveform.

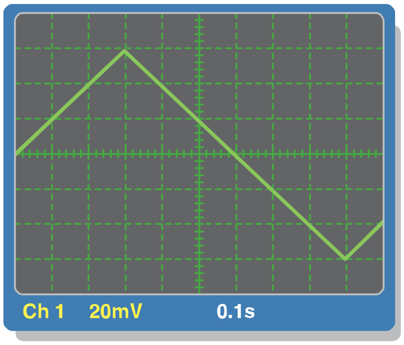

FIGURE 8–71 **The horizontal axis represents 0 V.**

SOLUTION

The waveform that is selected is a **triangle wave**.

The period is

$$T = (10\text{ div})(0.1\ \mu\text{s/div}) = 1.0\ \mu\text{s}$$

The frequency setting is

$$f = \frac{1}{T} = \frac{1}{1.0\ \mu\text{s}} = \mathbf{1.0\ MHz}$$

The peak-to-peak amplitude is

$$V_{\text{pp}} = (6.0\text{ div})(20\text{ mV/div}) = \mathbf{1.20\ V_{pp}}$$

RELATED PROBLEM

What setting of dc offset will produce a triangle waveform that goes from 0 V to 1.20 V? (With no dc offset, the waveform is centered on 0 V.)

SECTION 8–10 CHECKUP

1. What type of signals can be generated with an arbitrary function generator?
2. What is a swept-frequency oscillator?
3. Why is spectral purity important for a function generator?
4. What is modulation?
5. What is burst mode for a function generator?

6238262 2013/12/10 12:27:66.5

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SUMMARY

- The sine wave is a time-varying, periodic waveform.
- The sine wave is the fundamental type of alternating current (ac) and alternating voltage.
- Alternating current changes direction in response to changes in the polarity of the source voltage.
- One cycle of an alternating sine wave consists of a positive alternation and a negative alternation.
- The average value of a sine wave determined over one half-cycle is 0.637 times the peak value. The average value of a sine wave determined over a full cycle is zero.
- A full cycle of a sine wave is 360°, or 2π radians. A half-cycle is 180°, or π radian. A quarter-cycle is 90° or $\pi/2$ radians.
- Phase angle is the difference in degrees (or radians) between two sine waves or between a sine wave and a reference wave.
- The angular position of a phasor represents the angle of a sine wave, and the length of the phasor represents the amplitude.
- Voltages and currents must all be expressed with consistent units when you apply Ohm's or Kirchhoff's laws in ac circuits.
- Power in a resistive ac circuit is determined using rms voltage and/or current values.
- Alternators (ac generators) produce power when there is relative motion between a magnetic field and a conductor.
- Most alternators take the output from the stator. The rotor provides a moving magnetic field.
- Two major types of ac motors are induction motors and synchronous motors.
- Induction motors have a rotor that turns in response to a rotating field from the stator.
- Synchronous motors move at a constant speed in sync with the field of the stator.
- A pulse consists of a transition from a baseline level to an amplitude level, followed by a transition back to the baseline level.
- A triangular or sawtooth wave consists of positive-going and negative-going ramps.
- Harmonic frequencies are odd or even multiples of the repetition rate (or fundamental frequency) of a nonsinusoidal waveform.
- Conversions of sine wave values are summarized in Table 8–2.

TABLE 8–2

TO CHANGE FROM	TO	MULTIPLY BY
Peak	rms	0.707
Peak	Peak-to-peak	2
Peak	Average	0.637
rms	Peak	1.414
Peak-to-peak	Peak	0.5
Average	Peak	1.57

- Signal sources include function generators and a number of specialized instruments.
- Key specifications for signal generators include mode, frequency range, spectral purity, amplitude range, modulation, and output impedance.

KEY TERMS

Key terms and other bold terms in the chapter are defined in the end-of-book glossary.

Alternator An ac generator, alternators convert mechanical energy to electrical energy.

Amplitude The maximum value of a voltage or current.

Cycle One repetition of a periodic waveform.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Degree The unit of angular measure corresponding to 1/360 of a complete revolution.

Duty cycle A characteristic of a pulse waveform that indicates the percentage of time that a pulse is present during a cycle; the ratio of pulse width to period expressed as either a fraction or as a percentage.

Fall time (t_f) The time interval required for a pulse to change from 90% to 10% of its amplitude.

Frequency (f) A measure of the rate of change of a periodic function; the number of cycles completed in 1 s. The unit of frequency is the hertz.

Function generator An instrument that produces more than one type of waveform.

Fundamental frequency The repetition rate of a waveform.

Harmonics The frequencies contained in a composite waveform, which are integer multiples of the pulse repetition frequency (fundamental).

Hertz (Hz) The unit of frequency. One hertz equals one cycle per second.

Induction motor An ac motor that achieves excitation to the rotor by transformer action.

Instantaneous value The voltage or current value of a waveform at a given instant in time.

Modulation The process in which a signal containing information is used to modify the characteristics of another signal such as amplitude frequency or pulse width so that the information on the first is also contained in the second.

Oscillator An electronic circuit that produces a repetitive waveform on its output with only the dc supply voltage as an input.

Oscilloscope A measurement instrument that displays signal waveforms on a screen.

Peak-to-peak value The voltage or current value of a waveform measured from its minimum to its maximum points.

Peak value The voltage or current value of a waveform at its maximum positive or negative points.

Period (T) The time interval of one complete cycle of a periodic waveform.

Periodic Characterized by a repetition at fixed-time intervals.

Phase The relative angular displacement of a time-varying waveform in terms of its occurrence with respect to a reference.

Pulse A type of waveform that consists of two equal and opposite steps in voltage or current separated by a time interval.

Pulse width (t_W) The time interval between the opposite steps of an ideal pulse. For a nonideal pulse, the time between the 50% points on the leading and trailing edges.

Radian A unit of angular measurement. There are 2π radians in one complete 360° revolution. One radian equals 57.3°.

Ramp A type of waveform characterized by a linear increase or decrease in voltage or current.

Rise time (t_r) The time interval required for a pulse to change from 10% to 90% of its amplitude.

rms value The value of a sinusoidal voltage that indicates its heating effect, also known as the effective value. It is equal to 0.707 times the peak value. *rms* stands for root mean square.

Sine wave A type of waveform that follows a cyclic sinusoidal pattern defined by the formula $v = A \sin \theta$.

Slip The difference between the synchronous speed of the stator field and the rotor speed in an induction motor.

Squirrel cage An aluminum frame within the rotor of an induction motor that forms the electrical conductors for a rotating current.

Synchronous motor An ac motor in which the rotor moves at the same rate as the rotating magnetic field of the stator.

Waveform The pattern of variations of a voltage or current showing how the quantity changes with time.

KEY FORMULAS

(8–1)	$f = \dfrac{1}{T}$	Frequency
(8–2)	$T = \dfrac{1}{f}$	Period

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

(8–3)	$V_{pp} = 2V_p$	Peak-to-peak voltage (sine wave)
(8–4)	$I_{pp} = 2I_p$	Peak to peak current (sine wave)
(8–5)	$V_{\text{rms}} = 0.707V_p$	Root-mean-square voltage (sine wave)
(8–6)	$I_{\text{rms}} = 0.707I_p$	Root-mean-square current (sine wave)
(8–7)	$V_p = 1.414V_{\text{rms}}$	Peak voltage (sine wave)
(8–8)	$I_p = 1.414I_{\text{rms}}$	Peak current (sine wave)
(8–9)	$V_{pp} = 2.828V_{\text{rms}}$	Peak-to-peak voltage (sine wave)
(8–10)	$I_{pp} = 2.828I_{\text{rms}}$	Peak-to-peak current (sine wave)
(8–11)	$V_{\text{avg}} = 0.637V_p$	Half-cycle average voltage (sine wave)
(8–12)	$I_{\text{avg}} = 0.637I_p$	Half-cycle average current (sine wave)
(8–13)	$\text{rad} = \left(\dfrac{\pi\ \text{rad}}{180°}\right) \times \text{degrees}$	Degrees to radian conversion
(8–14)	$\text{degrees} = \left(\dfrac{180°}{\pi\ \text{rad}}\right) \times \text{rad}$	Radian to degrees conversion
(8–15)	$y = A \sin\theta$	General formula for a sine wave
(8–16)	$v = V_p \sin\theta$	Sinusoidal voltage
(8–17)	$i = I_p \sin\theta$	Sinusoidal current
(8–18)	$y = A\sin(\theta - \phi)$	Lagging sine wave
(8–19)	$y = A\sin(\theta + \phi)$	Leading sine wave
(8–20)	$f = \dfrac{Ns}{120}$	Output frequency of an alternator
(8–21)	$\text{Percent duty cycle} = \left(\dfrac{t_W}{T}\right)100\%$	Duty cycle
(8–22)	$V_{\text{avg}} = \text{baseline} + (\text{duty cycle})(\text{amplitude})$	Average voltage of pulse waveform

6239262 2013/12/10 13.27.66.5

TRUE/FALSE QUIZ

Answers are at the end of the chapter.

1. The period of a 60 Hz sine wave is 16.7 ms.
2. The rms and average value of a sine wave are the same.
3. A sine wave with a peak value of 10 V has the same heating effect as a 10 V dc source.
4. The peak value of a sine wave is the same as its amplitude.
5. The number of radians in 360° is 2π.
6. In a three-phase electrical system, the phases are separated by 60°.
7. The purpose of an exciter is to supply dc rotor current to an alternator.
8. In an automotive alternator, the output current is taken from the rotor through slip rings.
9. A maintenance issue with induction motors is brush replacement.
10. A synchronous motor can be used when constant speed is required.
11. Clamp meters are available that can measure dc.
12. The purpose of the diode array in an automobile alternator is to convert dc to ac.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SELF-TEST

Answers are at the end of the chapter.

1. The difference between alternating current (ac) and direct current (dc) is

(a) ac changes value and dc does not **(b)** ac changes direction and dc does not
(c) both (a) and (b) **(d)** neither (a) nor (b)

2. During each cycle, a sine wave reaches a peak value

(a) one time **(b)** two times **(c)** four times
(d) a number of times depending on the frequency

3. A sine wave with a frequency of 12 kHz is changing at a faster rate than a sine wave with a frequency of

(a) 20 kHz **(b)** 15,000 Hz **(c)** 10,000 Hz **(d)** 1.25 MHz

4. A sine wave with a period of 2 ms is changing at a faster rate than a sine wave with a period of

(a) 1 ms **(b)** 0.0025 s **(c)** 1.5 ms **(d)** 1000 μs

5. When a sine wave has a frequency of 60 Hz, in 10 s it goes through

(a) 6 cycles **(b)** 10 cycles **(c)** 1/16 cycle **(d)** 600 cycles

6. If the peak value of a sine wave is 10 V, the peak-to-peak value is

(a) 20 V **(b)** 5 V **(c)** 100 V **(d)** none of these

7. If the peak value of a sine wave is 20 V, the rms value is

(a) 14.14 V **(b)** 6.37 V **(c)** 7.07 V **(d)** 0.707 V

8. The average value of a 10 V peak sine wave over one complete cycle is

(a) 0 V **(b)** 6.37 V **(c)** 7.07 V **(d)** 5 V

9. The average half-cycle value of a sine wave with a 20 V peak is

(a) 0 V **(b)** 6.37 V **(c)** 12.74 V **(d)** 14.14 V

10. One sine wave has a positive-going zero crossing at 10° and another sine wave has a positive-going zero crossing at 45°. The phase angle between the two waveforms is

(a) 55° **(b)** 35° **(c)** 0° **(d)** none of these

11. The instantaneous value of a 15 A peak sine wave at a point 32° from its positive-going zero crossing is

(a) 7.95 A **(b)** 7.5 A **(c)** 2.13 A **(d)** 7.95 V

12. If the rms current through a 10 kΩ resistor is 5 mA, the rms voltage drop across the resistor is

(a) 70.7 V **(b)** 7.07 V **(c)** 5 V **(d)** 50 V

13. Two series resistors are connected to an ac source. If there is 6.5 V rms across one resistor and 3.2 V rms across the other, the peak source voltage is

(a) 9.7 V **(b)** 9.19 V **(c)** 13.72 V **(d)** 4.53 V

14. An advantage of a three-phase induction motor is that it

(a) maintains constant speed for any load **(b)** does not require starting windings
(c) has a wound rotor **(d)** all of the above

15. The difference in the synchronous speed of the stator field and the rotor speed of a motor is called

(a) differential speed **(b)** loading **(c)** lag **(d)** slip

16. A 10 kHz pulse waveform consists of pulses that are 10 μs wide. Its duty cycle is

(a) 100% **(b)** 10% **(c)** 1% **(d)** not determinable

17. The duty cycle of a square wave

(a) varies with the frequency **(b)** varies with the pulse width
(c) both (a) and (b) **(d)** is 50%

18. A waveform mode that changes the output between a start-and-stop frequency is called

(a) triggered **(b)** burst
(c) swept **(d)** modulation

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

TROUBLESHOOTING: SYMPTOM AND CAUSE

The purpose of these exercises is to help develop thought processes essential to troubleshooting. Answers are at the end of the chapter.

Determine the cause for each set of symptoms. Refer to Figure 8–72.

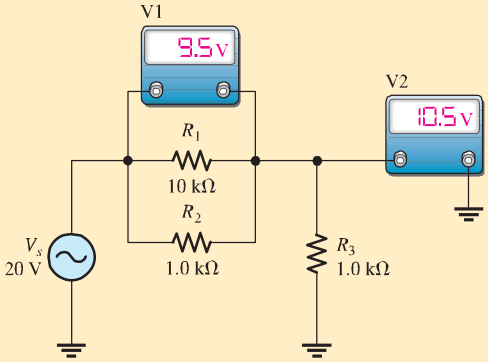

FIGURE 8–72 **The ac meters indicate the correct readings for this circuit.**

1. *Symptom:* The voltmeter 1 reading is 0 V, and the voltmeter 2 reading is 20 V.
   Cause:
   (a) R_1 is open. **(b)** R_2 is open. **(c)** R_3 is open.
2. *Symptom:* The voltmeter 1 reading is 20 V, and the voltmeter 2 reading is 0 V.
   Cause:
   (a) R_1 is open. **(b)** R_2 is shorted. **(c)** R_3 is shorted.
3. *Symptom:* The voltmeter 1 reading is 18.2 V, and the voltmeter 2 reading is 1.8 V.
   Cause:
   (a) R_1 is open. **(b)** R_2 is open. **(c)** R_1 is shorted.
4. *Symptom:* Both voltmeter readings are 10 V.
   Cause:
   (a) R_1 is open. **(b)** R_1 is shorted. **(c)** R_2 is open.
5. *Symptom:* The voltmeter 1 reading is 16.7 V, and the voltmeter 2 reading is 3.3 V.
   Cause:
   (a) R_1 is shorted. **(b)** R_2 is 10 kΩ instead of 1 kΩ. **(c)** R_3 is 10 kΩ instead of 1 kΩ.

PROBLEMS

Answers to odd-numbered problems are at the end of the book.

BASIC PROBLEMS

SECTION 8–1 The Sinusoidal Waveform

1. Calculate the frequency for each of the following periods:
   (a) 1 s **(b)** 0.2 s **(c)** 50 ms **(d)** 1 ms **(e)** 500 μs **(f)** 10 μs
2. Calculate the period for each of the following frequencies:
   (a) 1 Hz **(b)** 60 Hz **(c)** 500 Hz **(d)** 1 kHz **(e)** 200 kHz **(f)** 5 MHz
3. A sine wave goes through 5 cycles in 10 μs. What is its period?
4. A sine wave has a frequency of 50 kHz. How many cycles does it complete in 10 ms?
5. How long does it take a 10 kHz sine wave to complete 100 cycles?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 8–2 Voltage and Current Values of Sine Waves

6. A sine wave has a peak value of 12 V. Determine the following voltage values:
(a) rms **(b)** peak-to-peak **(c)** half-cycle average

7. A sinusoidal current has an rms value of 5 mA. Determine the following current values:
(a) peak **(b)** half-cycle average **(c)** peak-to-peak

8. For the sine wave in Figure 8–73, determine the peak, peak-to-peak, rms, and half-cycle average values.

9. If each horizontal division in Figure 8–73 is 1 ms, determine the instantaneous voltage value at
(a) 1 ms **(b)** 2 ms **(c)** 4 ms **(d)** 7 ms

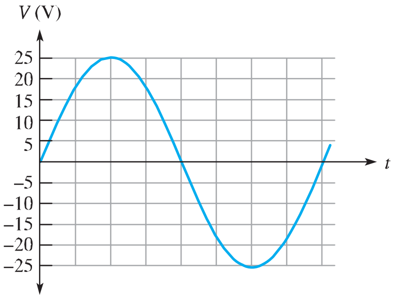

FIGURE 8–73

SECTION 8–3 Angular Measurement of a Sine Wave

10. In Figure 8–73, what is the instantaneous voltage at
(a) 45° **(b)** 90° **(c)** 180°

11. Sine wave A has a positive-going zero crossing at 30° from a reference. Sine wave B has a positive-going zero crossing at 45° from the same reference. Determine the phase angle between the two signals. Which signal leads?

12. One sine wave has a positive peak at 75° and another has a positive peak at 100°. How much is each sine wave shifted in phase from the 0° reference? What is the phase angle between them?

13. Draw two sine waves as follows: Sine wave A is the reference, and sine wave B lags A by 90°. Both have equal amplitudes.

14. Convert the following angular values from degrees to radians:
(a) 30° **(b)** 45° **(c)** 78° **(d)** 135° **(e)** 200° **(f)** 300°

15. Convert the following angular values from radians to degrees:
(a) $\pi/8$ **(b)** $\pi/3$ **(c)** $\pi/2$ **(d)** $3\pi/5$ **(e)** $6\pi/5$ **(f)** 1.8π

6238262 2013/12/10 12.27.66.5

SECTION 8–4 The Sine Wave Formula

16. A certain sine wave has a positive-going zero crossing at 0° and an rms value of 20 V. Calculate its instantaneous value at each of the following angles:
(a) 15° **(b)** 33° **(c)** 50° **(d)** 110°
(e) 70° **(f)** 145° **(g)** 250° **(h)** 325°

17. For a particular 0° reference sinusoidal current, the peak value is 100 mA. Determine the instantaneous value at each of the following points:
(a) 35° **(b)** 95° **(c)** 190° **(d)** 215° **(e)** 275° **(f)** 360°

18. For a 0° reference sine wave with an rms value of 6.37 V, determine its instantaneous value at each of the following points:
(a) $\pi/8$ rad **(b)** $\pi/4$ rad **(c)** $\pi/2$ rad **(d)** $3\pi/4$ rad
(e) π rad **(f)** $3\pi/2$ rad **(g)** 2π rad

19. Sine wave A lags sine wave B by 30°. Both have peak values of 15 V. Sine wave A is the reference with a positive-going crossing at 0°. Determine the instantaneous value of sine wave B at 30°, 45°, 90°, 180°, 200°, and 300°.

20. Repeat Problem 19 for the case when sine wave A leads sine wave B by 30°.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 8–5 Analysis of AC Circuits

21. A sinusoidal voltage is applied to the resistive circuit in Figure 8–74. Determine the following:
(a) I_{rms} **(b)** I_{avg} **(c)** I_p **(d)** I_{pp} **(e)** i at the positive peak

22. Find the half-cycle average values of the voltages across R_1 and R_2 in Figure 8–75. All values shown are rms.

23. Determine the rms voltage across R_3 in Figure 8–76.

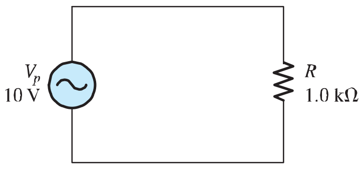

FIGURE 8–74

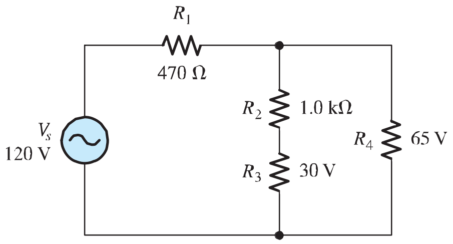

FIGURE 8–75

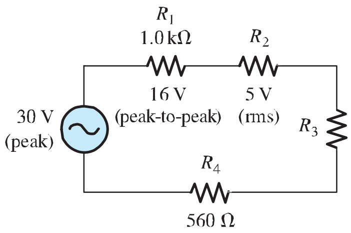

FIGURE 8–76

24. A sine wave with an rms value of 10.6 V is riding on a dc level of 24 V. What are the maximum and minimum values of the resulting waveform?

25. How much dc voltage must be added to a 3 V rms sine wave in order to make the resulting voltage nonalternating (no negative values)?

26. A 6 V peak sine wave is riding on a dc voltage of 8 V. If the dc voltage is lowered to 5 V, how far negative will the sine wave go?

SECTION 8–6 Alternators (AC Generators)

27. The conductive wire loop on the rotor of a simple two-pole, single-phase generator rotates at a rate of 250 rps. What is the frequency of the induced output voltage?

28. A certain four-pole generator has a speed of rotation of 3600 rpm. What is the frequency of the voltage produced by this generator?

29. At what speed of rotation must a four-pole generator be operated to produce a 400 Hz sinusoidal voltage?

30. A common frequency for alternators on aircraft is 400 Hz. How many poles does a 400 Hz alternator have if the speed of rotation is 3000 rpm?

SECTION 8–7 AC Motors

31. What is the main difference between a one-phase induction motor and a three-phase induction motor?

32. Explain how the field in a three-phase motor rotates if there are no moving parts to the field coils.

SECTION 8–8 Nonsinusoidal Waveforms

33. From the graph in Figure 8–77, determine the approximate values of t_r, t_f, t_W, and amplitude.

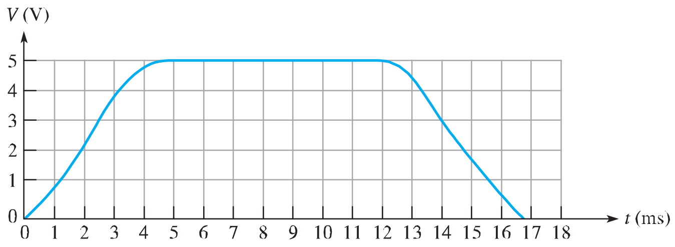

FIGURE 8–77

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

34. Determine the duty cycle for each pulse waveform in Figure 8–78.

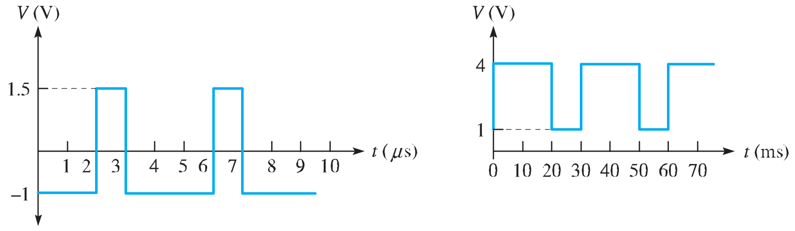

(a) (b)

FIGURE 8–78

35. Find the average value of each positive-going pulse waveform in Figure 8–78.

36. What is the frequency of each waveform in Figure 8–78.

37. What is the frequency of each sawtooth waveform in Figure 8–79?

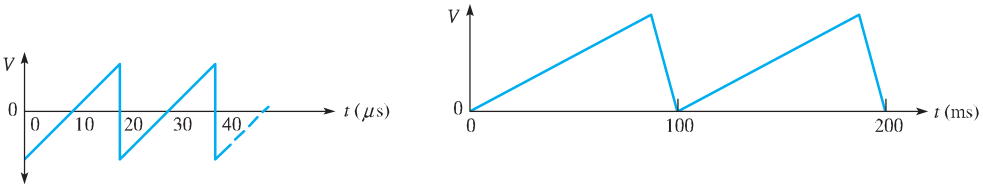

(a) (b)

FIGURE 8–79

38. A square wave has a period of 40 μs. List the first six odd harmonics.

39. What is the fundamental frequency of the square wave mentioned in Problem 38?

SECTION 8–9 The Oscilloscope

40. Determine the peak value and the period of the sine wave displayed on the scope screen in Figure 8–80. The horizontal axis is 0 V.

41. Determine the rms value and the frequency of the sine wave displayed on the scope screen in Figure 8–80.

42. Determine the rms value and the frequency of the sine wave displayed on the scope screen in Figure 8–81. The horizontal axis is 0 V.

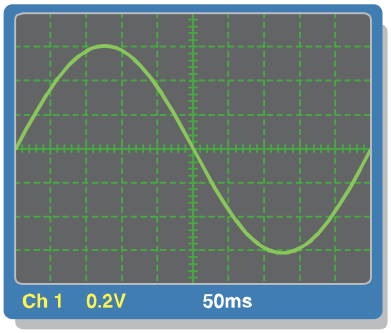

FIGURE 8–80

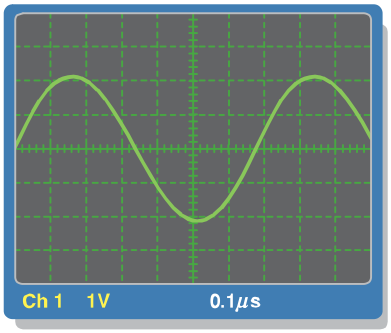

FIGURE 8–81

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

43. Find the amplitude, pulse width, and duty cycle for the pulse waveform displayed on the scope screen in Figure 8–82. The horizontal axis is 0 V.

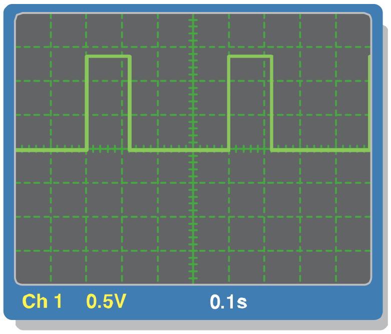

FIGURE 8–82

SECTION 8–10 Signal Sources

44. Why is spectral purity important for a sinusoidal signal generator?

45. What is a modulating signal and how is it applied to an analog signal generator?

46. Assume you are testing a circuit with different frequencies to measure its response. Why is it important to check the signal level from the generator each time the frequency is changed?

47. What is the difference between burst mode and gated mode for a generator?

ADVANCED PROBLEMS

48. A certain sine wave has a frequency of 2.2 kHz and an rms value of 25 V. Assuming a given cycle begins (zero crossing) at $t = 0$ s, what is the change in voltage from 0.12 ms to 0.2 ms?

49. Figure 8–83 shows a sinusoidal voltage source in series with a dc source. Effectively, the two voltages are superimposed. Draw the total voltage across R_L. Determine the maximum current through R_L and the average voltage across R_L.

50. A nonsinusoidal waveform called a *stairstep* is shown in Figure 8–84. Determine its average value.

6238262 2013/12/10 12.27.66.5

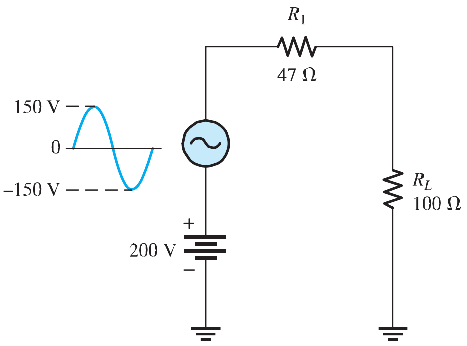

FIGURE 8–83

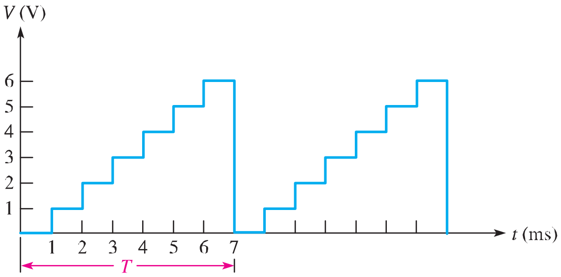

FIGURE 8–84

51. Refer to the oscilloscope screen in Figure 8–85.
(a) How many cycles are displayed?
(b) What is the rms value of the sine wave?
(c) What is the frequency of the sine wave?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

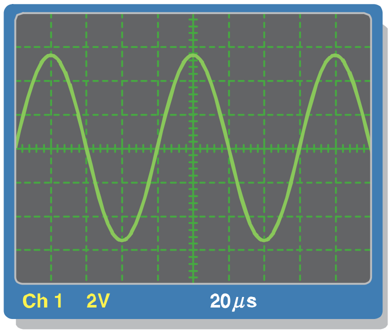

FIGURE 8–85

52. Accurately draw on a grid representing the scope screen in Figure 8–85 how the sine wave will appear if the Volts/Div control is moved to a setting of 5 V.

53. Accurately draw on a grid representing the scope screen in Figure 8–85 how the sine wave will appear if the Sec/Div control is moved to a setting of 10 μs.

54. Based on the instrument settings and an examination of the scope display and the circuit board in Figure 8–86, determine the frequency and peak value of the input signal and output signal. The waveform shown is channel 1. Draw the channel 2 waveform as it would appear on the scope with the indicated settings.

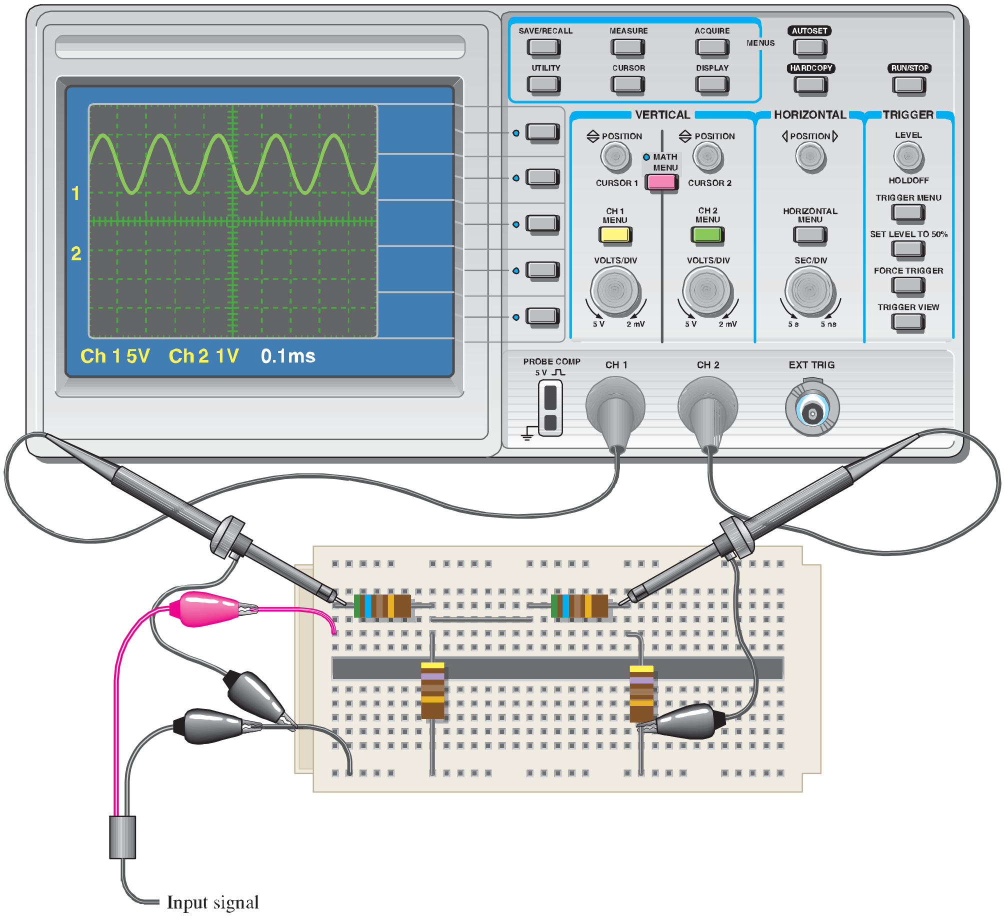
1
2
Ch 1 5V
Ch 2 1V
0.1ms
SAVE/RECALL
MEASURE
ACQUIRE
MENUS
UTILITY
CURSOR
DISPLAY
AUTOSET
HARDCOPY
RUN/STOP
VERTICAL
POSITION
MATH MENU
POSITION
CURSOR 1
CURSOR 2
CH 1 MENU
CH 2 MENU
VOLTS/DIV
VOLTS/DIV
5 V
2 mV
5 V
2 mV
HORIZONTAL
POSITION
HORIZONTAL MENU
SEC/DIV
5 s
5 ns
TRIGGER
LEVEL
HOLDOFF
TRIGGER MENU
SET LEVEL TO 50%
FORCE TRIGGER
TRIGGER VIEW
PROBE COMP
5 V
CH 1
CH 2
EXT TRIG
Input signal

FIGURE 8–86

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

55. Examine the circuit board and the oscilloscope display in Figure 8–87 and determine the peak value and the frequency of the unknown input signal.

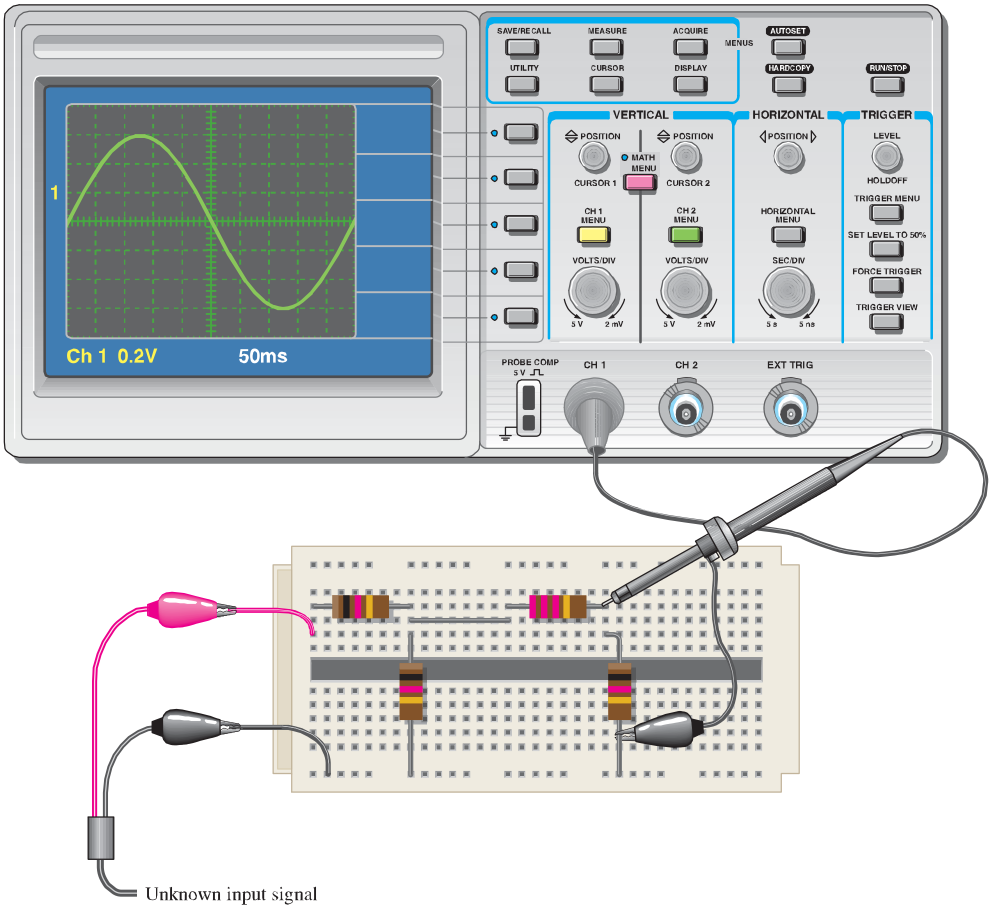
SAVE/RECALL
MEASURE
ACQUIRE
MENUS
AUTOSET
UTILITY
CURSOR
DISPLAY
HARDCOPY
RUN/STOP
VERTICAL
HORIZONTAL
TRIGGER
POSITION
MATH MENU
POSITION
POSITION
LEVEL
CURSOR 1
CURSOR 2
HOLDOFF
TRIGGER MENU
CH 1 MENU
CH 2 MENU
HORIZONTAL MENU
SET LEVEL TO 50%
VOLTS/DIV
VOLTS/DIV
SEC/DIV
FORCE TRIGGER
5 V
2 mV
5 V
2 mV
5 s
5 ns
TRIGGER VIEW
1
Ch 1 0.2V
50ms
PROBE COMP 5 V
CH 1
CH 2
EXT TRIG
Unknown input signal

FIGURE 8–87

MULTISIM

MULTISIM TROUBLESHOOTING PROBLEMS

56. Open file P08–56; files are found at www.pearsonhighered.com/floyd. Measure the peak value and period of the voltage sine wave using the oscilloscope.
57. Open file P08–57 and determine if there is a fault. If so, identify it.
58. Open file P08–58 and determine if there is a fault. If so, identify it.
59. Open file P08–59 and measure the amplitude and period of the pulse waveform using the oscilloscope.
60. Open file P08–60 and determine if there is a fault. If so, identify it.

ANSWERS TO SECTION CHECKUPS

SECTION 8–1 The Sinusoidal Waveform

1. One cycle goes from the zero crossing through a positive peak, then through zero to a negative peak and back to the zero crossing.
2. A sine wave changes polarity at the zero crossings.
3. A sine wave has two maximum points (peaks) in one cycle.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

4. The period is measured from one zero crossing to the next corresponding zero crossing, or from one peak to the next corresponding peak.

5. Frequency is the number of cycles completed in one second and its unit is the hertz.

6. $f = 1/5\ \mu\text{s} = 200\text{ kHz}$

7. $T = 1/120\text{ Hz} = 8.33\text{ ms}$

6238262 2013/12/10 12.27.66.5

SECTION 8–2 Voltage and Current Values of Sine Waves

1. **(a)** $V_{pp} = 2(1\text{ V}) = 2\text{ V}$
(b) $V_{pp} = 2(1.414)(1.414\text{ V}) = 4\text{ V}$
(c) $V_{pp} = 2(1.57)(3\text{ V}) = 9.42\text{ V}$

2. **(a)** $V_{\text{rms}} = (0.707)(2.5\text{ V}) = 1.77\text{ V}$
(b) $V_{\text{rms}} = (0.5)(0.707)(10\text{ V}) = 3.54\text{ V}$
(c) $V_{\text{rms}} = (0.707)(1.57)(1.5\text{ V}) = 1.66\text{ V}$

3. **(a)** $V_{\text{avg}} = (0.637)(10\text{ V}) = 6.37\text{ V}$
(b) $V_{\text{avg}} = (0.637)(1.414)(2.3\text{ V}) = 2.07\text{ V}$
(c) $V_{\text{avg}} = (0.637)(0.5)(60\text{ V}) = 19.1\text{ V}$

SECTION 8–3 Angular Measurement of a Sine Wave

1. **(a)** The positive peak is at 90°.
(b) The negative-going zero crossing is at 180°.
(c) The negative peak is at 270°.
(d) The cycle ends at 360°.

2. A half-cycle is completed in 180° or π radians.

3. A full cycle is completed in 360° or 2π radians.

4. $\theta = 90^\circ - 45^\circ = 45^\circ$

SECTION 8–4 The Sine Wave Formula

1. **(a)** $\sin 30^\circ = 0.5$
(b) $\sin 60^\circ = 0.866$
(c) $\sin 90^\circ = 1$

2. $v = 10 \sin 120^\circ = 8.66\text{ V}$

3. $v = 10 \sin(45^\circ + 10^\circ) = 8.19\text{ V}$

SECTION 8–5 Analysis of AC Circuits

1. $I_p = V_p/R = (1.57)(12.5\text{ V})/330\ \Omega = 59.5\text{ mA}$

2. $V_{s(\text{rms})} = (0.707)(25.3\text{ V}) = 17.9\text{ V}$

3. $+V_{max} = 5\text{ V} + 2.5\text{ V} = 7.5\text{ V}$

4. Yes, it will alternate.

5. $+V_{max} = 5\text{ V} - 2.5\text{ V} = 2.5\text{ V}$

SECTION 8–6 Alternators (AC Generators)

1. The number of poles and the rotor speed.

2. The brushes do not have to handle the output current.

3. A dc generator that supplies rotor current to larger alternators.

4. The diodes convert the ac from the stator to dc for the final output.

SECTION 8–7 AC Motors

1. The difference is the rotors. In an induction motor, the rotor obtains current by transformer action; in a synchronous motor, the rotor is a permanent magnet or an electromagnet that is supplied current from an external source through slip rings and brushes.

2. The magnitude is constant.

3. The squirrel cage is composed of the electrical conductors that generate current in the rotor.

4. Slip is the difference between the synchronous speed of the stator field and the rotor speed.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 8–8 Nonsinusoidal Waveforms

1. **(a)** Rise time is the time interval from 10% to 90% of the amplitude;
   (b) Fall time is the time interval from 90% to 10% of the amplitude;
   (c) Pulse width is the time interval from 50% of the leading pulse edge to 50% of the trailing pulse edge.
2. $f = 1/1\text{ ms} = 1\text{ kHz}$
3. $\text{d} = (1/5)100\% = 20\%$; Ampl. 1.5 V; $V_{\text{avg}} = 0.5\text{ V} + 0.2(1.5\text{ V}) = 0.8\text{ V}$
4. $T = 16\text{ ms}$
5. $f = 1/T = 1/1\,\mu\text{s} = 1\text{ MHz}$
6. Fundamental frequency is the repetition rate of the waveform.
7. 2nd harmonic: 2 kHz
8. $f = 1/10\,\mu\text{s} = 100\text{ kHz}$

SECTION 8–9 The Oscilloscope

1. Digital: Signal is converted to digital for processing and then reconstructed for display.
   Analog: Signal drives display directly.
2. Voltage is measured vertically.
3. The Volts/Div control sets the amount of volts represented by each division on the vertical scale.
4. The Sec/Div control sets the amount of time represented by each division on the horizontal scale.
5. Use a ×10 probe most of the time except when measuring very small voltages.

SECTION 8–10 Signal Sources

1. Any waveform within the limits (amplitude and frequency) of the generator. Waves can be specified by various methods.
2. A generator that has a sine wave output that is varied at a cyclic rate between two selected frequencies.
3. Certain measurements of a unit under test can be distorted if the generator does not have good spectral purity.
4. The process in which a signal containing information is used to modify the characteristic of another signal.
5. A mode where the output is triggered for a specific number of cycles at a time.

ANSWERS TO RELATED PROBLEMS FOR EXAMPLES

8–1 2.4 s

8–2 1.5 ms

8–3 20 kHz

8–4 200 Hz

8–5 1 ms

8–6 $V_{pp} = 50\text{ V}$; $V_{\text{rms}} = 17.7\text{ V}$; $V_{\text{avg}} = 15.9\text{ V}$

8–7 **(a)** $\pi/12$ rad
(b) 112.5°

8–8 8°

8–9 18.1 V

8–10 $I_{\text{rms}} = 4.53\text{ mA}$; $V_{1(\text{rms})} = 4.53\text{ V}$; $V_{2(\text{rms})} = 2.54\text{ V}$; $P_{tot} = 32\text{ mV}$

8–11 23.7 V

8–12 The waveform in part (a) never goes negative. The waveform in part (b) goes negative for a portion of its cycle.

8–13 250 rpm

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

8–14 5%

8–15 1.2 V

8–16 120 V

8–17 **(a)** $V_{\text{rms}} = 1.06\text{ V}; f = 50\text{ Hz}$
(b) $V_{\text{rms}} = 88.4\text{ mV}; f = 1.67\text{ kHz}$
(c) $V_{\text{rms}} = 4.81\text{ V}; f = 5\text{ kHz}$
(d) $V_{\text{rms}} = 7.07\text{ V}; f = 250\text{ kHz}$

8–18 0.6 V

ANSWERS TO TRUE/FALSE QUIZ

1. T **2.** F **3.** F **4.** T **5.** T **6.** F
7. T **8.** F **9.** F **10.** T **11.** T **12.** F

ANSWERS TO SELF-TEST

1. (b) **2.** (b) **3.** (c) **4.** (b) **5.** (d) **6.** (a)
7. (a) **8.** (a) **9.** (c) **10.** (b) **11.** (a) **12.** (d)
13. (c) **14.** (b) **15.** (d) **16.** (b) **17.** (d) **18.** (c)

ANSWERS TO TROUBLESHOOTING: SYMPTOM AND CAUSE

1. (c)
2. (c)
3. (b)
4. (a)
5. (b)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.
6238262 2013/12/10 12.27.66.5

CHAPTER 9

CAPACITORS

OUTLINE

9–1 The Basic Capacitor
9–2 Types of Capacitors
9–3 Series Capacitors
9–4 Parallel Capacitors
9–5 Capacitors in DC Circuits
9–6 Capacitors in AC Circuits
9–7 Capacitor Applications

OBJECTIVES

- Describe the basic structure and characteristics of a capacitor
- Discuss various types of capacitors
- Analyze series capacitors
- Analyze parallel capacitors
- Describe how a capacitor operates in a dc switching circuit
- Describe how a capacitor operates in an ac circuit
- Discuss some capacitor applications

KEY TERMS

Capacitor
Dielectric
Capacitance
Farad (F)
Dielectric strength
Temperature coefficient
Dielectric constant
Charging
***RC* time constant**
Exponential
Transient time
Capacitive reactance
Instantaneous power
True power
Reactive power
VAR (volt-ampere reactive)
Filter
Ripple voltage
Coupling
Decoupling
Bypass

INTRODUCTION

The *capacitor* is a device that can store electrical charge, thereby creating an electric field which in turn stores energy. The measure of the charge-storing ability of a capacitor is its *capacitance*.

In this chapter, the basic capacitor is introduced and its characteristics are studied. The physical construction and electrical properties of various types of capacitors are discussed. Series and parallel combinations are analyzed, and the basic behavior of capacitors in both dc and ac circuits is studied. Representative circuits that are found in many systems also are discussed. One example, found in nearly all systems that are not battery powered, is the power supply that is used to convert ac utility voltage to dc. Capacitors are important elements of power supplies, and they are discussed in this chapter.

VISIT THE WEBSITE
Study aids for this chapter are available at
http://pearsonhighered.com/floyd

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

9–1 THE BASIC CAPACITOR

A capacitor is a passive electrical component that stores electrical charge and has the property of capacitance.

After completing this section, you should be able to

- **Describe the basic structure and characteristics of a capacitor**
  - **Explain how a capacitor stores charge**
  - **Define *capacitance* and state its unit**
  - **Explain how a capacitor stores energy**
  - **Discuss voltage rating and temperature coefficient**
  - **Explain capacitor leakage**
  - **Specify how the physical characteristics of a capacitor affect the capacitance**

Basic Construction

In its simplest form, a basic capacitor is an electrical device constructed of two parallel conductive plates separated by an insulating material called the **dielectric**. Connecting leads are attached to the parallel plates. A basic capacitor is shown in Figure 9–1(a), and the schematic symbol for this type of capacitor is shown in part (b).

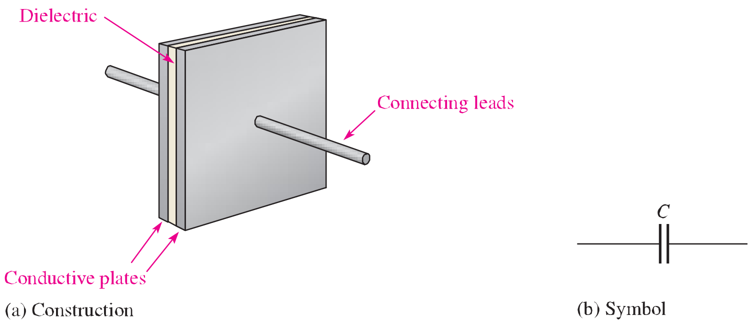

(a) Construction

(b) Symbol

FIGURE 9–1 **A basic capacitor.**

How a Capacitor Stores Charge

In the neutral state, both plates of a capacitor have an equal number of free electrons, as indicated in Figure 9–2(a). When the capacitor is connected to a dc voltage source through a resistor, as shown in part (b), electrons (negative charge) are removed from plate A, and an equal number are deposited on plate B. As plate A loses electrons and plate B gains electrons, plate A becomes positive with respect to plate B. During this charging process, electrons flow only through the connecting leads and the source. No electrons flow through the dielectric of the capacitor because it is an insulator. The movement of electrons ceases when the voltage across the capacitor equals the source voltage, as indicated in Figure 9–2(c). If the capacitor is disconnected from the source, it retains the stored charge for a long period of time (the length of time depends on the type of capacitor) and still has the voltage across it, as shown in Figure 9–2(d). A charged capacitor can act as a temporary battery.

Capacitance

The amount of charge that a capacitor can store per unit of voltage across its plates is its capacitance, designated C. That is, **capacitance** is a measure of a capacitor's ability to store

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

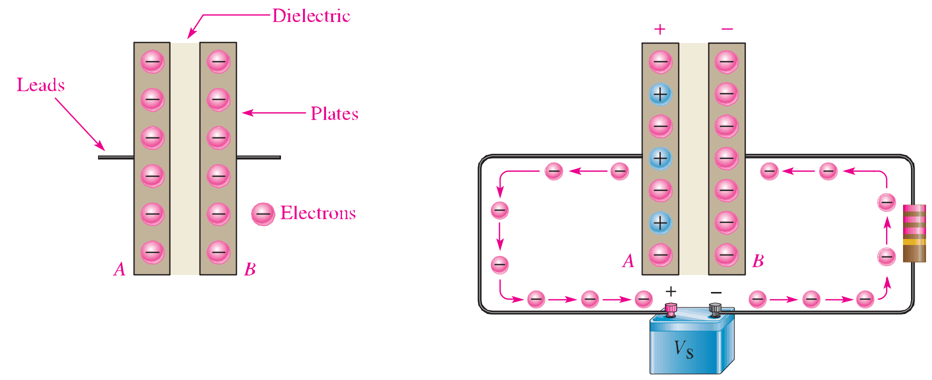

(a) Neutral (uncharged) capacitor (same charge on both plates)

(b) When connected to a voltage source, electrons flow from plate *A* to plate *B* as the capacitor charges.

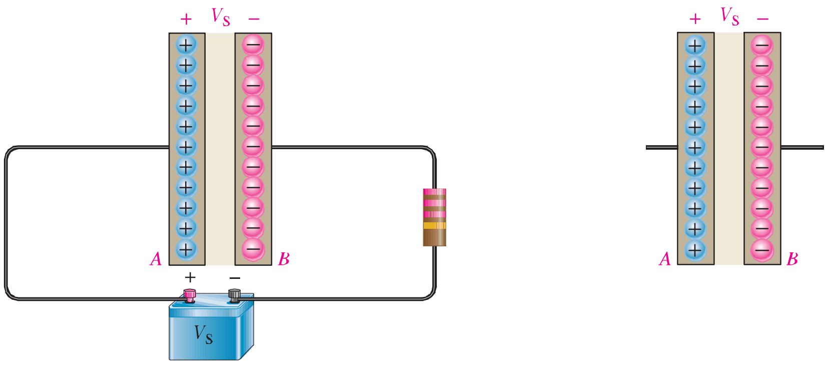

(c) After the capacitor charges to V_S, no electrons flow.

(d) Ideally, the capacitor retains charge when disconnected from the voltage source.

FIGURE 9–2 **Illustration of a capacitor storing charge.**

charge. The more charge per unit of voltage that a capacitor can store, the greater its capacitance, as expressed by the following formula:

$$C = \frac{Q}{V} \qquad (9\text{–}1)$$

where C is capacitance, Q is charge, and V is voltage.

By rearranging the terms in Equation 9–1, you can obtain two other formulas.

$$Q = CV \qquad (9\text{–}2)$$

$$V = \frac{Q}{C} \qquad (9\text{–}3)$$

THE UNIT OF CAPACITANCE The **farad (F)** is the basic unit of capacitance. Recall that the coulomb (C) is the unit of electrical charge.

> **One farad is the amount of capacitance when one coulomb of charge is stored with one volt across the plates.**

Most capacitors that are used in electronics work have capacitance values in microfarads (μF) and picofarads (pF). Occasionally, a schematic will show a capacitance in nanofarads (nF). A microfarad is one-millionth of a farad (1 μF $= 1 \times 10^{-6}$ F); and a picofarad is one-trillionth of a farad (1 pF $= 1 \times 10^{-12}$ F). Conversions for farads, microfarads, and picofarads are given in Table 9–1.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

TABLE 9–1 • Conversions for farads, microfarads, and picofarads.

TO CONVERT FROM	TO	MOVE THE DECIMAL POINT
Farads	Microfarads	right 6 places ($\times 10^6$)
Farads	Picofarads	right 12 places ($\times 10^{12}$)
Microfarads	Farads	left 6 places ($\times 10^{-6}$)
Microfarads	Picofarads	right 6 places ($\times 10^6$)
Nanofarads	Farads	left 9 places ($\times 10^{-9}$)
Nanofarads	Microfarads	left 3 places ($\times 10^{-3}$)
Picofarads	Farads	left 12 places ($\times 10^{-12}$)
Picofarads	Microfarads	left 6 places ($\times 10^{-6}$)

EXAMPLE 9–1

(a) A certain capacitor stores 50 microcoulombs (50 μC) when 10 V are applied across its plates. What is its capacitance?

(b) A 2.2 μF capacitor has 100 V across its plates. How much charge does it store?

(c) Determine the voltage across a 100 pF capacitor that is storing 2 μC of charge.

SOLUTION

(a) $C = \dfrac{Q}{V} = \dfrac{50\ \mu\text{C}}{10\ \text{V}} = \mathbf{5\ \mu F}$

(b) $Q = CV = (2.2\ \mu\text{F})(100\ \text{V}) = \mathbf{220\ \mu C}$

(c) $V = \dfrac{Q}{C} = \dfrac{2\ \mu\text{C}}{100\ \text{pF}} = \mathbf{20\ kV}$

RELATED PROBLEM*

Determine V if $C = 1000$ pF and $Q = 10\ \mu$C.

***Answers are at the end of the chapter.**

EXAMPLE 9–2

Convert the following values to microfarads:

(a) 0.00001 F **(b)** 0.0047 F **(c)** 1000 pF **(d)** 220 pF

SOLUTION

(a) $0.00001\ \text{F} \times 10^6\ \mu\text{F/F} = \mathbf{10\ \mu F}$

(b) $0.0047\ \text{F} \times 10^6\ \mu\text{F/F} = \mathbf{4700\ \mu F}$

(c) $1000\ \text{pF} \times 10^{-6}\ \mu\text{F/pF} = \mathbf{0.001\ \mu F}$

(d) $220\ \text{pF} \times 10^{-6}\ \mu\text{F/pF} = \mathbf{0.00022\ \mu F}$

RELATED PROBLEM

Convert 47,000 pF to microfarads.

6238262 2013/12/10 12.27.66.5

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 9–3

Convert the following values to picofarads:

(a) 0.1×10^{-8} F **(b)** 0.000027 F **(c)** 0.01 μF **(d)** 0.0047 μF

SOLUTION

(a) $0.1 \times 10^{-8}\text{ F} \times 10^{12}\text{ pF/F} = \mathbf{1000\ pF}$

(b) $0.000027\text{ F} \times 10^{12}\text{ pF/F} = \mathbf{27 \times 10^{6}\ pF}$

(c) $0.01\ \mu\text{F} \times 10^{6}\text{ pF}/\mu\text{F} = \mathbf{10{,}000\ pF}$

(d) $0.0047\ \mu\text{F} \times 10^{6}\text{ pF}/\mu\text{F} = \mathbf{4700\ pF}$

RELATED PROBLEM

Convert 100 μF to picofarads.

How a Capacitor Stores Energy

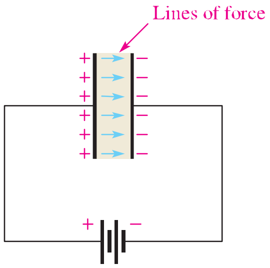

FIGURE 9–3 **The electric field stores energy in a capacitor. The beige area indicates the dielectric.**

A capacitor stores energy in an electric field that is established by the opposite charges stored on the two plates. The electric field is represented by lines of force between the positive and negative charges and is concentrated within the dielectric, as shown in Figure 9–3.

The plates in Figure 9–3 have acquired a charge because they are connected to a battery. This creates an electric field between the plates, which stores energy. The energy stored in the electric field is directly related to the size of the capacitor and to the square of the voltage as given by the following equation for the energy stored:

$$W = \frac{1}{2}CV^2 \tag{9–4}$$

When capacitance (C) is in farads and voltage (V) is in volts, the energy (W) is in joules.

Voltage Rating

Every capacitor has a limit on the amount of voltage that it can withstand across its plates. The voltage rating specifies the maximum dc voltage that can be applied without risk of damage to the device. If this maximum voltage, commonly called the *breakdown voltage* or *working voltage,* is exceeded, permanent damage to the capacitor can result.

You must consider both the capacitance and the voltage rating before you use a capacitor in a circuit application. The choice of capacitance value is based on particular circuit requirements. The voltage rating should always be above the maximum voltage expected in a particular application.

DIELECTRIC STRENGTH The breakdown voltage of a capacitor is determined by the **dielectric strength** of the dielectric material used. The dielectric strength is expressed in V/mil (1 mil = 0.001 in. = 2.54×10^{-5} meter). Table 9–2 lists typical values for several materials. Exact values vary depending on the specific composition of the material.

The dielectric strength of a capacitor can best be explained by an example. Assume that a certain capacitor has a plate separation of 1 mil and that the dielectric material is ceramic. This particular capacitor can withstand a maximum voltage of 1000 V because its dielectric strength is 1000 V/mil. If the maximum voltage is exceeded, the dielectric may break down and conduct current, causing permanent damage to the capacitor. If the ceramic capacitor has a plate separation of 2 mils, its breakdown voltage is 2000 V.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

TABLE 9–2 • Some common dielectric materials and their typical dielectric strengths.

MATERIAL	DIELECTRIC STRENGTH (V/MIL)
Air	80
Oil	375
Ceramic	1000
Paper (paraffined)	1200
Teflon®	1500
Mica	1500
Glass	2000

Temperature Coefficient

The **temperature coefficient** indicates the amount and direction of a change in capacitance value with temperature. A positive temperature coefficient means that the capacitance increases with an increase in temperature or decreases with a decrease in temperature. A negative coefficient means that the capacitance decreases with an increase in temperature or increases with a decrease in temperature.

Temperature coefficients typically are specified in parts per million per Celsius degree (ppm/°C). For example, a negative temperature coefficient of 150 ppm/°C for a 1 μF capacitor means that for every degree Celsius rise in temperature, the capacitance decreases by 150 pF (there are one million picofarads in one microfarad).

Leakage

No insulating material is perfect. The dielectric of any capacitor will conduct some very small amount of current. Thus, the charge on a capacitor will eventually leak off. Some types of capacitors have higher leakages than others. An equivalent circuit for a nonideal capacitor is shown in Figure 9–4. The parallel resistor R_{leak} represents the extremely high resistance (several hundred kΩ or more) of the dielectric material through which there is leakage current.

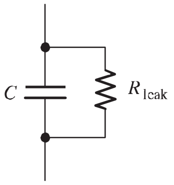

FIGURE 9–4 **Equivalent circuit for a nonideal capacitor.**

Capacitive Sensors

Capacitive sensors are common in systems that measure quantities like pressure, proximity, humidity, liquid level, and more. A capacitive sensor is a variable capacitor that responds to a physical stimulus. In the case of a pressure sensor, a conductive, flexible diaphragm forms one plate of the capacitor. The other plate is fixed. Pressure causes the diaphragm to move, thus changing the capacitance. The change in capacitance can be sensed by an external circuit.

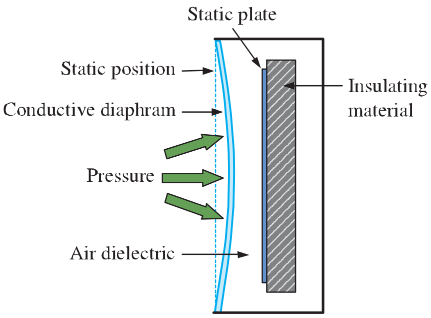

In the case of a proximity sensor, one of the plates can actually be a moving object (such as the cam on a rotating shaft). Other sensors can change the dielectric as in the case of a liquid-level sensor.

SYSTEM NOTE

Physical Characteristics of a Capacitor

The following parameters are important in establishing the capacitance and the voltage rating of a capacitor: plate area, plate separation, and dielectric constant.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

PLATE AREA *Capacitance is directly proportional to the physical size of the plates as determined by the plate area, A.* A large plate area produces more capacitance, and a smaller plate area produces less capacitance. Figure 9–5(a) shows that the plate area of a parallel plate capacitor is the area of one of the plates. If the plates are moved in relation to each other, as shown in Figure 9–5(b), the overlapping area determines the effective plate area. This variation in effective plate area is the basis for a certain type of variable capacitor.

PLATE SEPARATION *Capacitance is inversely proportional to the distance between the plates.* The plate separation is designated d, as shown in Figure 9–6. A greater separation of the plates produces a lesser capacitance, as illustrated in the figure. The breakdown voltage is directly proportional to the plate separation. The further the plates are separated, the greater the breakdown voltage.

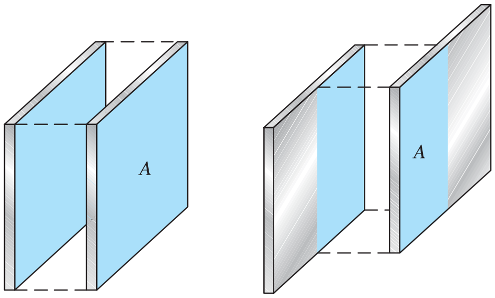

(a) Full plate area: more capacitance

(b) Reduced plate area: less capacitance

FIGURE 9–5 **Capacitance is directly proportional to plate area (A).**

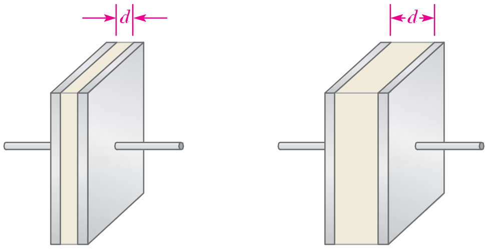

(a) Plates closer together: more capacitance

(b) Plates further apart: less capacitance

FIGURE 9–6 **Capacitance is inversely proportional to the distance d between the plates.**

DIELECTRIC CONSTANT As you know, the insulating material between the plates of a capacitor is called the *dielectric*. Every dielectric material has the ability to concentrate the lines of force of the electric field existing between the oppositely charged plates of a capacitor and thus increase the capacity for energy storage. The measure of a material's ability to establish an electric field is called the **dielectric constant** or *relative permittivity,* symbolized by ε_r (the Greek letter epsilon).

Capacitance is directly proportional to the dielectric constant. The dielectric constant of a vacuum is defined as 1, and that of air is very close to 1. These values are used as a reference, and all other materials have values of ε_r specified with respect to that of a vacuum or air. For example, a material with $\varepsilon_r = 5$ can have a capacitance five times greater than that of air, with all other factors being equal.

Table 9–3 lists several common dielectric materials and a typical dielectric constant for each. The values can vary because it depends on the specific composition of the material.

TABLE 9–3 • Some common dielectric materials and their typical dielectric constants.

MATERIAL	TYPICAL ε_r VALUE
Air (vacuum)	1.0
Teflon®	2.0
Paper (paraffined)	2.5
Oil	4.0
Mica	5.0
Glass	7.5
Ceramic	1200

6238262 2013/12/10 12.27.66.5

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The dielectric constant (relative permittivity) is dimensionless because it is a relative measure. It is a ratio of the absolute permittivity of a material, ε, to the absolute permittivity of a vacuum, ε_0, as expressed by the following formula:

$$\varepsilon_r = \frac{\varepsilon}{\varepsilon_0} \qquad (9\text{–}5)$$

The value of ε_0 is 8.85×10^{-12} F/m (farads per meter).

FORMULA FOR CAPACITANCE You have seen how capacitance is directly related to plate area, A, and the dielectric constant, ε_r, and inversely related to plate separation, d. An exact formula for calculating the capacitance in terms of these three quantities is

$$C = \frac{A\varepsilon_r(8.85 \times 10^{-12}\ \text{F/m})}{d} \qquad (9\text{–}6)$$

where A is in square meters (m^2), d is in meters (m), and C is in farads (F).

EXAMPLE 9–4

Determine the capacitance in μF of a parallel plate capacitor having a plate area of 0.01 m^2 and a plate separation of 0.5 mil (1.27×10^{-5} m). The dielectric is mica, which has a dielectric constant of 5.0.

SOLUTION

Use Equation 9–6.

$$C = \frac{A\varepsilon_r(8.85 \times 10^{-12}\ \text{F/m})}{d} = \frac{(0.01\ \text{m}^2)(5.0)(8.85 \times 10^{-12}\ \text{F/m})}{1.27 \times 10^{-5}\ \text{m}} = \mathbf{0.035\ \mu F}$$

RELATED PROBLEM

Determine C in μF where $A = 3.6 \times 10^{-5}$ m^2, $d = 1$ mil (2.54×10^{-5} m), and ceramic is the dielectric.

SECTION 9–1 CHECKUP*

1. Define *capacitance.*
2. **(a)** How many microfarads in one farad?
   (b) How many picofarads in one farad?
   (c) How many picofarads in one microfarad?
3. Convert 0.0015 μF to picofarads. To farads.
4. How much energy in joules is stored by a 0.01 μF capacitor with 15 V across its plates?
5. **(a)** When the plate area of a capacitor is increased, does the capacitance increase or decrease?
   (b) When the distance between the plates is increased, does the capacitance increase or decrease?
6. The plates of a ceramic capacitor are separated by 2 mils. What is the typical breakdown voltage?

*Answers are at the end of the chapter.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

9–2 TYPES OF CAPACITORS

Capacitors normally are classified according to the type of dielectric material. The most common types of dielectric materials are mica, ceramic, plastic-film, and electrolytic (aluminum oxide and tantalum oxide).

After completing this section, you should be able to

- **Discuss various types of capacitors**
  - **Describe the characteristics of mica, ceramic, plastic-film, and electrolytic capacitors**
  - **Describe types of variable capacitors**
  - **Identify capacitor labeling**
  - **Discuss capacitance measurement**

Fixed Capacitors

MICA CAPACITORS Two types of mica capacitors are stacked-foil and silver-mica. The basic construction of the stacked-foil type is shown in Figure 9–7. It consists of alternate layers of metal foil and thin sheets of mica. The metal foil forms the plate, with alternate foil sheets connected together to increase the plate area. More layers are used to increase the plate area, thus increasing the capacitance. The mica/foil stack is encapsulated in an insulating material such as Bakelite®, as shown in Figure 9–7(b). The silver-mica capacitor is formed in a similar way by stacking mica sheets with silver electrode material screened on them.

Mica capacitors are generally available with capacitance values ranging from 1 pF to 0.1 μF and voltage ratings from 100 V dc to 2500 V dc and higher. Mica has a typical dielectric constant of 5.

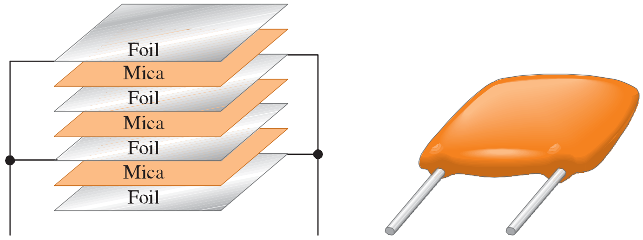

(a) Stacked layer arrangement

(b) Layers are pressed together and encapsulated.

FIGURE 9–7 **Construction of a typical radial-lead mica capacitor.**

CERAMIC CAPACITORS Ceramic dielectrics provide very high dielectric constants (1200 is typical). As a result, comparatively high capacitance values can be achieved in a small physical size. Ceramic capacitors are commonly available in a ceramic disk form, as shown in Figure 9–8; in a multilayer radial-lead configuration, as shown in Figure 9–9; or in a leadless ceramic chip, as shown in Figure 9–10.

Ceramic capacitors typically are available in capacitance values ranging from 1 pF to 100 μF with voltage ratings up to 6 kV.

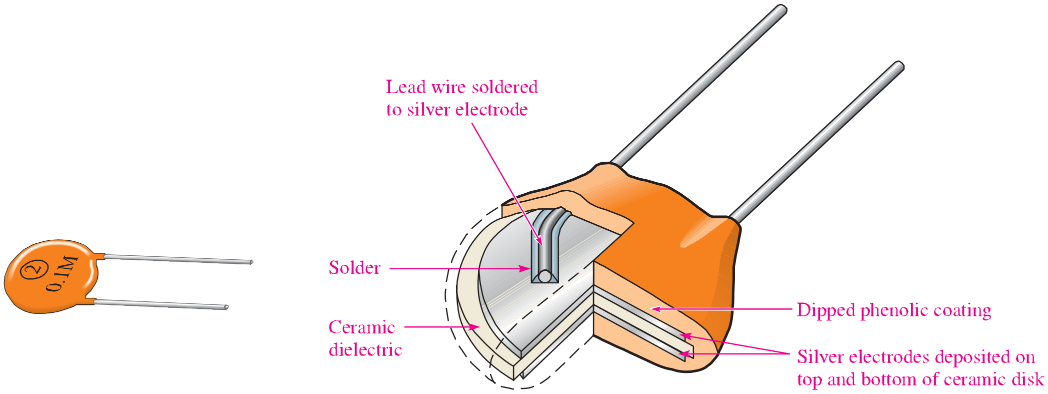

(a)

(b)

FIGURE 9–8 **A ceramic disk capacitor and its basic construction.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

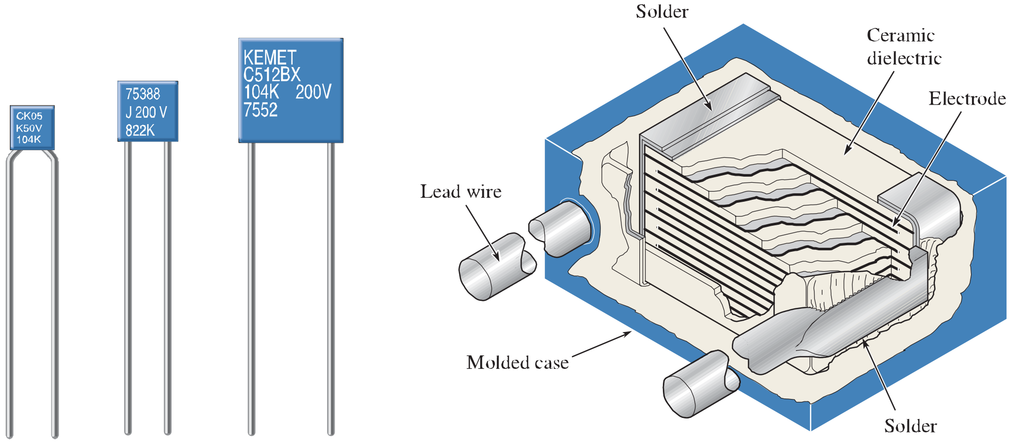
CK05
K50V
104K
75388
J 200 V
822K
KEMET
C512BX
104K 200V
7552
Solder
Ceramic dielectric
Electrode
Lead wire
Molded case
Solder

(a) Typical capacitors (b) Construction view

FIGURE 9–9 **Examples of ceramic capacitors.**

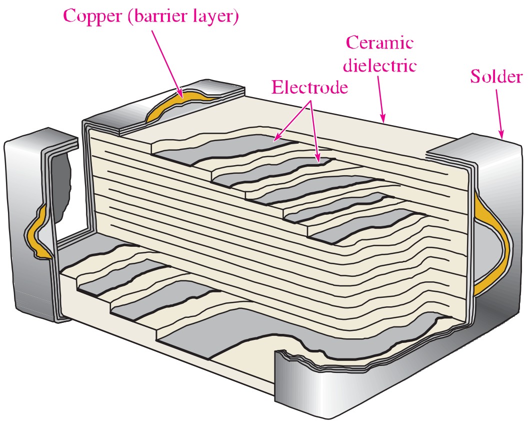
Copper (barrier layer)
Ceramic dielectric
Electrode
Solder

FIGURE 9–10 **Construction view of a typical ceramic chip capacitor used for surface mounting on printed circuit boards.**

HANDS ON TIP

Chip capacitors are for surface mounting on a printed circuit board and have conductive terminals plated on the ends. These capacitors will withstand the molten solder temperatures encountered in reflow and wave soldering used in automated circuit board assembly. Chip capacitors are in great demand because of miniaturization.

PLASTIC-FILM CAPACITORS Common dielectric materials used in plastic-film capacitors include polycarbonate, propylene, polyester, polystyrene, polypropylene, and mylar. Some of these types have capacitance values up to 100 μF but most are less than 1 μF.

Figure 9–11 shows a basic construction used in many plastic-film capacitors. A thin strip of plastic-film dielectric is sandwiched between two thin metal strips that act as plates. One lead is connected to the inner plate and one is connected to the outer plate as indicated. The strips are then rolled in a spiral configuration and encapsulated in a molded case. Thus, a large plate area can be packaged in a relatively small physical size, thereby achieving large capacitance values. Another method uses metal deposited directly on the film dielectric to form the plates.

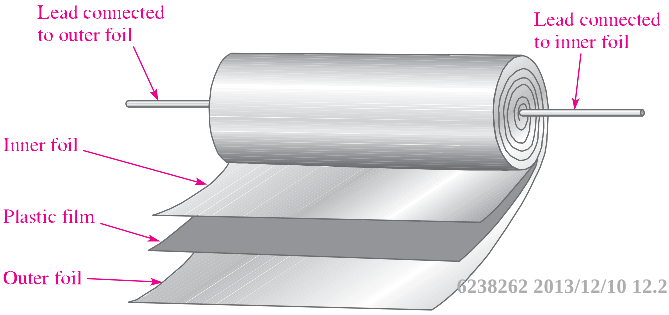
Lead connected to outer foil
Lead connected to inner foil
Inner foil
Plastic film
Outer foil

6238262 2013/12/10 12.27.66.5

FIGURE 9–11 **Basic construction of axial-lead tubular plastic-film dielectric capacitors.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Figure 9–12(a) shows typical plastic-film capacitors. Figure 9–12(b) shows a construction view for one type of an axial-lead plastic-film capacitor.

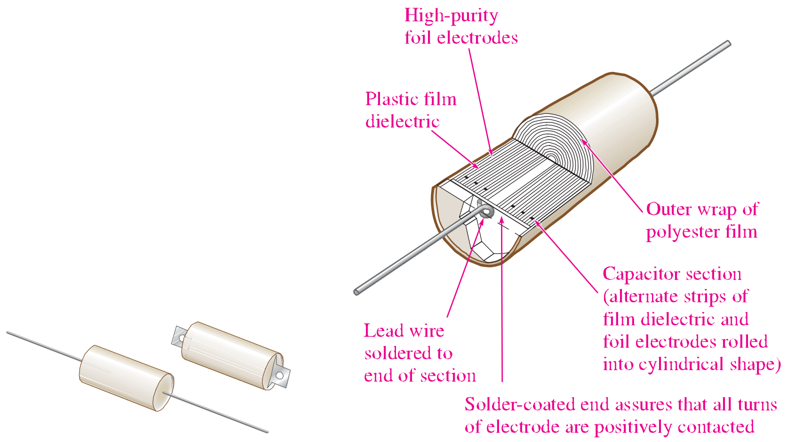
High-purity foil electrodes
Plastic film dielectric
Outer wrap of polyester film
Capacitor section (alternate strips of film dielectric and foil electrodes rolled into cylindrical shape)
Lead wire soldered to end of section
Solder-coated end assures that all turns of electrode are positively contacted

(a) Typical capacitors

(b) Construction view

FIGURE 9–12 **Examples of plastic-film capacitors.**

ELECTROLYTIC CAPACITORS Electrolytic capacitors are polarized so that one plate is positive and the other negative. These capacitors are generally used for high capacitance values from 1 μF up to over 200,000 μF, but they have relatively low breakdown voltages (350 V is a typical maximum but higher voltages are occasionally found) and high amounts of leakage.

Manufacturers have developed new electrolytic capacitors with much larger capacitance values; however, these new capacitors have lower voltage ratings than smaller-value capacitors and tend to be expensive. Super capacitors with capacitances of hundreds of farads are available. These capacitors are useful for battery backup and for applications like small motor starters that require a very large capacitance.

Electrolytic capacitors offer much higher capacitance values than mica or ceramic capacitors, but their voltage ratings are typically lower. While other capacitors use two similar plates, the electrolytic capacitor consists of one plate made of aluminum foil and another plate made of a conducting electrolyte applied to a material such as plastic film. These two "plates" are separated by a layer of aluminum oxide that forms on the surface of the aluminum plate. Figure 9–13(a) illustrates the construction of a typical aluminum electrolytic capacitor with axial leads. Other electrolytic capacitors with radial leads are shown in Figure 9–13(b); the symbol for an electrolytic capacitor is shown in part (c).

Tantalum electrolytic capacitors can be in either a tubular configuration similar to Figure 9–13 or "tear drop" shape as shown in Figure 9–14. In the tear drop configuration, the positive plate is actually a pellet of tantalum powder rather than a sheet of foil. Tantalum pentoxide forms the dielectric, and manganese dioxide forms the negative plate.

Because of the process used for the insulating pentoxide dielectric, the metallic (aluminum or tantalum) plate is always positive with respect to the electrolyte plate, and thus all electrolytic capacitors are polarized. The metal plate (positive lead) is usually indicated by a plus sign or some other obvious marking on the capacitor and must always be connected in a dc circuit where the voltage across the capacitor does not change polarity. Reversal of the polarity of the voltage can result in complete destruction of the capacitor.

SAFETY NOTE

Be extremely careful with electrolytic capacitors because it does make a difference which way an electrolytic capacitor is connected. Always observe the proper polarity. If a polarized capacitor is connected backwards, it may explode and cause injury.

HANDS ON TIP

Scientists are working with graphene, a carbon-based material that may be used for improved charge storage in rechargeable batteries and ultracapacitors. The ability to store large amounts of charge is important in many applications from office copiers to improving the efficiency of electric and hybrid vehicles. The new technology may speed development of renewable energies such as wind and solar power, which need to be able to store large amounts of energy.

Variable Capacitors

Variable capacitors are used in a circuit when there is a need to adjust the capacitance value either manually or automatically. These capacitors are generally less than 300 pF but

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

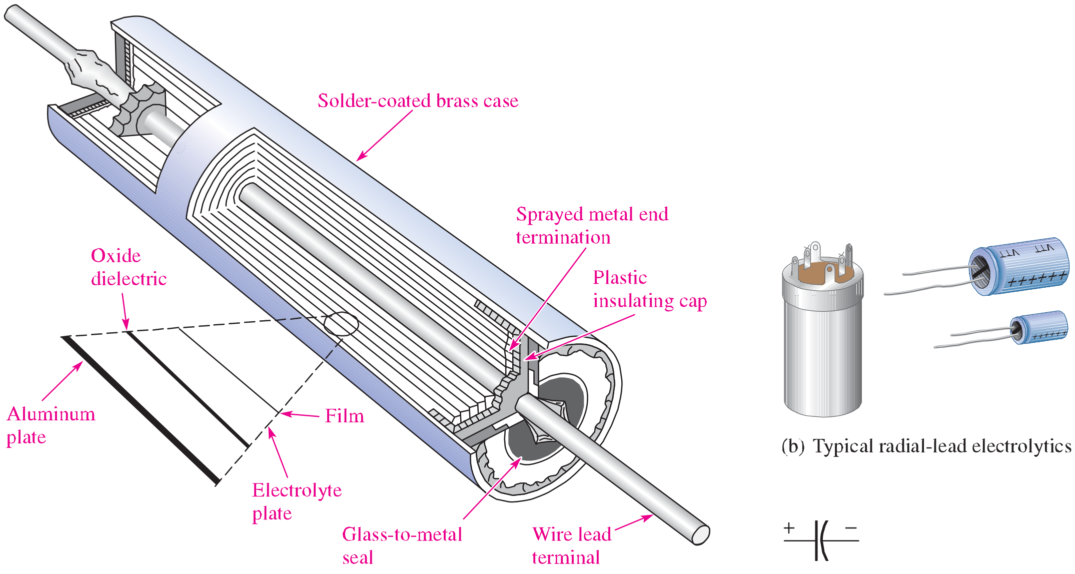
Solder-coated brass case
Sprayed metal end termination
Plastic insulating cap
Oxide dielectric
Aluminum plate
Film
Electrolyte plate
Glass-to-metal seal
Wire lead terminal
+
−

(a) Construction view of an axial-lead electrolytic capacitor

(b) Typical radial-lead electrolytics

(c) Symbol for an electrolytic capacitor.

FIGURE 9–13 **Examples of electrolytic capacitors.**

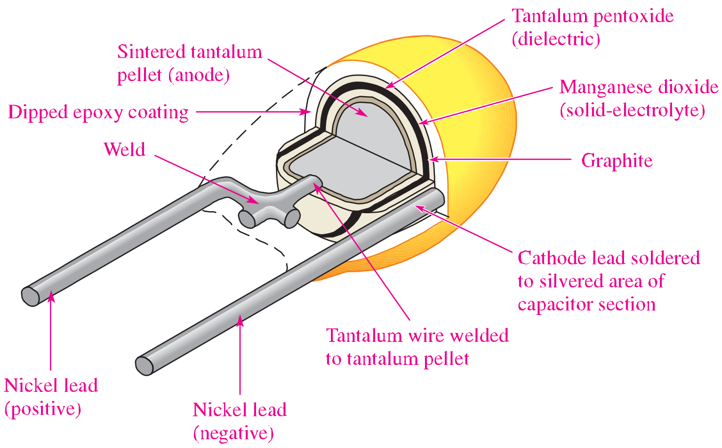
Tantalum pentoxide (dielectric)
Sintered tantalum pellet (anode)
Manganese dioxide (solid-electrolyte)
Dipped epoxy coating
Weld
Graphite
Cathode lead soldered to silvered area of capacitor section
Tantalum wire welded to tantalum pellet
Nickel lead (positive)
Nickel lead (negative)

FIGURE 9–14 **Construction view of a typical "tear drop" shaped tantalum electrolytic capacitor.**

are available in larger values for specialized application. The schematic symbol for a variable capacitor is shown in Figure 9–15.

Adjustable capacitors that normally have slotted screw-type adjustments and are used for very fine adjustments in a circuit are called **trimmers**. Ceramic or mica is a common dielectric in these types of capacitors, and the capacitance usually is changed by adjusting the plate separation. Generally, trimmer capacitors have values less than 100 pF. Figure 9–16 shows some typical devices.

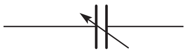

FIGURE 9–15 **Schematic symbol for a variable capacitor.**

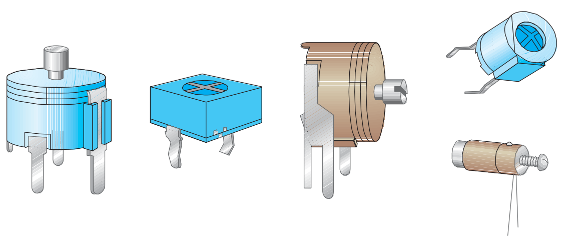

FIGURE 9–16 **Trimmer capacitors.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The **varactor** is a semiconductive device that exhibits a capacitance characteristic that is varied by changing the voltage across its terminals. This device usually is covered in detail in a course on electronic devices.

Capacitor Labeling

Capacitor values are indicated on the body of the capacitor either by numerical or alphanumerical labels and sometimes by color codes. Capacitor labels indicate various parameters such as capacitance, voltage rating, and tolerance.

6238262 2013/12/10 12.27.66.5

Some capacitors carry no unit designation for capacitance. In these cases, the units are implied by the value indicated and are recognized by experience. For example, a ceramic capacitor marked .001 or .01 has units of microfarads because picofarad values that small are not available. As another example, a ceramic capacitor labeled 50 or 330 has units of picofarads because microfarad units that large are not available in this type. In some cases a 3-digit designation is used. The first two digits are the first two digits of the capacitance value. The third digit is the multiplier or number of zeros after the second digit. For example, 103 means 10,000 pF. In some instances, the units are labeled as pF or μF; sometimes the microfarad unit is labeled as MF or MFD.

A voltage rating appears on some types of capacitors with WV or WVDC and is omitted on others. When it is omitted, the voltage rating can be determined from information supplied by the manufacturer. The tolerance of the capacitor is usually labeled as a percentage, such as $\pm 10\%$. The temperature coefficient is indicated by a *parts per million* marking. This type of label consists of a P or N followed by a number. For example, N750 means a negative temperature coefficient of 750 ppm/°C, and P30 means a positive temperature coefficient of 30 ppm/°C. An NP0 designation means that the positive and negative coefficients are zero; thus the capacitance does not change with temperature. Certain types of capacitors are color coded. Refer to Appendix B for additional capacitor labeling and color code information.

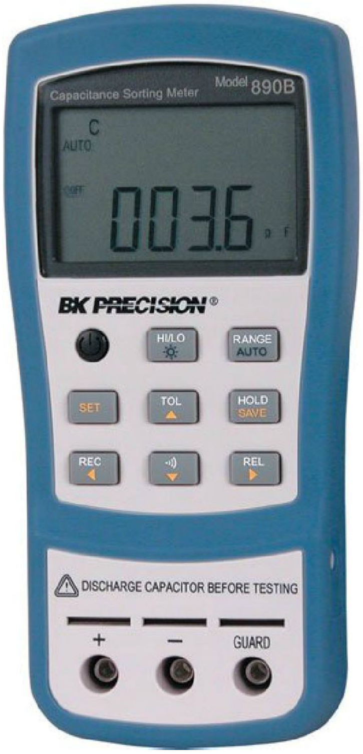

FIGURE 9–17 **A typical capacitance meter.** (Photo courtesy of B&K Precision Corp.)

Capacitance Measurement

A capacitance meter, such as the one shown in Figure 9–17, can be used to check the value of a capacitor. Also, many DMMs provide a capacitance measurement feature. All capacitors change value over a period of time, some more than others. Ceramic capacitors, for example, often exhibit a 10% to 15% change in the value during the first year. Electrolytic capacitors are particularly subject to value change due to drying of the electrolytic solution. In other cases, capacitors may be labeled incorrectly or the wrong value may have been installed in the circuit. Although a value change represents less than 25% of defective capacitors, a value check should be made to quickly eliminate this as a source of trouble when troubleshooting a circuit. Values from 200 pF to 50,000 μF can be measured by simply connecting the capacitor, setting the switch, and reading the value on the display.

Some capacitance meters can also be used to check for leakage current in capacitors. In order to check for leakage, a sufficient voltage must be applied across the capacitor to simulate operating conditions. This is automatically done by the test instrument. Over 40% of all defective capacitors have excessive leakage current and electrolytic capacitors are particularly susceptible to this problem.

SECTION 9–2 CHECKUP

1. How are capacitors commonly classified?
2. What is the difference between a fixed and a variable capacitor?
3. What type of capacitor normally is polarized?
4. What precautions must be taken when a polarized capacitor is installed in a circuit?
5. An electrolytic capacitor is connected between a negative supply voltage and ground. Which capacitor lead should be connected to ground?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

9–3 SERIES CAPACITORS

The total capacitance of a series connection of capacitors is less than the individual capacitance of any of the capacitors. Capacitors in series divide voltage across them inversely proportional to their capacitance. Thus, a large capacitor in series with a smaller one will have less voltage than a smaller one.

After completing this section, you should be able to

- **Analyze series capacitors**
  - **Determine total capacitance**
  - **Determine capacitor voltages**

When capacitors are connected in series, the total capacitance is less than the smallest capacitance value because the effective plate separation increases. The calculation of total series capacitance is analogous to the calculation of total resistance of parallel resistors (Chapter 5).

Two capacitors can be used in series to show how the total capacitance is determined. Figure 9–18 shows two capacitors, which initially are uncharged, connected in series with a dc voltage source. When the switch is closed, as shown in part (a), current begins.

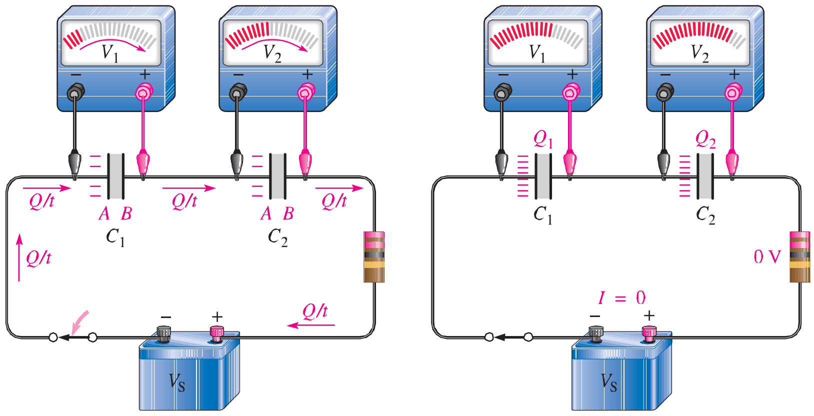

(a) While charging, $I = Q/t$ is the same at all points. Capacitor voltages are increasing.

(b) Both capacitors store the same amount of charge ($Q_T = Q_1 = Q_2$).

FIGURE 9–18 Capacitors in series produce a total capacitance that is less than the smallest value.

Recall that current is the same at all points in a series circuit and that current is defined as the rate of flow of charge ($I = Q/t$). In a certain period of time, a certain amount of charge moves through the circuit. Since current is the same everywhere in the circuit of Figure 9–18(a), the same amount of charge is moved from the negative side of the source to plate *A* of C_1, and from plate *B* of C_1 to plate *A* of C_2, and from plate *B* of C_2 to the positive side of the source. As a result, of course, the same amount of charge is deposited on the plates of both capacitors in a given period of time, and the total charge (Q_T) moved through the circuit in that period of time equals the charge stored by C_1 and also equals the charge stored by C_2.

$$Q_T = Q_1 = Q_2$$

As the capacitors charge, the voltage across each one increases as indicated.

Figure 9–18(b) shows the capacitors after they have been completely charged and the current has ceased. Both capacitors store an equal amount of charge (Q), and the voltage across each one depends on its capacitance value ($V = Q/C$). By Kirchhoff's voltage law,

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

which applies to capacitive circuits as well as to resistive circuits, the sum of the capacitor voltages equals the source voltage.

$$V_S = V_1 + V_2$$

Using the formula $V = Q/C$, you can substitute into the formula for Kirchhoff's law and get the following relationship (where $Q = Q_T = Q_1 = Q_2$):

$$\frac{Q}{C_T} = \frac{Q}{C_1} + \frac{Q}{C_2}$$

The Q can be factored out of the right side of the equation and canceled with the Q on the left side as follows:

$$\frac{\cancel{Q}}{C_T} = \cancel{Q}\left(\frac{1}{C_1} + \frac{1}{C_2}\right)$$

Thus, you have the following relationship for two capacitors in series:

$$\frac{1}{C_T} = \frac{1}{C_1} + \frac{1}{C_2}$$

Taking the reciprocal of both sides of this equation gives the formula for the total capacitance for two capacitors in series.

$$C_T = \frac{1}{\dfrac{1}{C_1} + \dfrac{1}{C_2}}$$

This equation can also be expressed equivalently as

$$\mathbf{C_T = \frac{C_1C_2}{C_1 + C_2}} \qquad (9\text{–}7)$$

EXAMPLE 9–5

Find the total capacitance C_T in Figure 9–19.

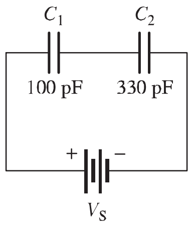

FIGURE 9–19

SOLUTION

$$C_T = \frac{C_1C_2}{C_1 + C_2} = \frac{(100 \text{ pF})(330 \text{ pF})}{100 \text{ pF} + 330 \text{ pF}} = \mathbf{76.7\ pF}$$

RELATED PROBLEM

Determine C_T if $C_1 = 470$ pF and $C_2 = 680$ pF in Figure 9–19.

GENERAL FORMULA Although more than two capacitors in series is rare in practical circuits, you should be aware of the general formula, should the situation arise. The equation for two series capacitors can be extended to any number of capacitors in series, as shown in Figure 9–20.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

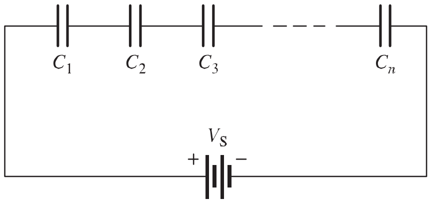

FIGURE 9–20 **General circuit with *n* capacitors in series.**

The formula for total capacitance for any number of capacitors in series is developed as follows; the subscript *n* can be any number.

$$\frac{1}{C_T} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots + \frac{1}{C_n}$$

$$\boldsymbol{C_T = \cfrac{1}{\cfrac{1}{C_1} + \cfrac{1}{C_2} + \cfrac{1}{C_3} + \cdots + \cfrac{1}{C_n}}} \tag{9–8}$$

Remember,

The total series capacitance is always less than the smallest capacitance.

EXAMPLE 9–6

Determine the total capacitance in Figure 9–21.

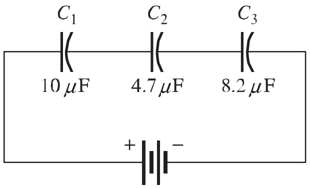

FIGURE 9–21

6238262 2013/12/10 12.27.66.5

SOLUTION

$$C_T = \cfrac{1}{\cfrac{1}{C_1} + \cfrac{1}{C_2} + \cfrac{1}{C_3}} = \cfrac{1}{\cfrac{1}{10\ \mu\text{F}} + \cfrac{1}{4.7\ \mu\text{F}} + \cfrac{1}{8.2\ \mu\text{F}}} = \mathbf{2.30\ \mu F}$$

RELATED PROBLEM

If another 4.7 μF capacitor is connected in series with the three existing capacitors in Figure 9–21, what is the value of C_T?

Capacitor Voltages

The voltage across each capacitor in a series connection depends on its capacitance value according to the formula $V = Q/C$. You can determine the voltage across any individual capacitor in series with the following formula:

$$\boldsymbol{V_x = \left(\frac{C_T}{C_x}\right)V_S} \tag{9–9}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

where C_x is any capacitor in series, such as C_1, C_2, and C_3, and so on, and V_x is the voltage across C_x.

The largest-value capacitor in a series connection will have the smallest voltage across it. The smallest-value capacitor will have the largest voltage across it.

EXAMPLE 9–7

Find the voltage across each capacitor in Figure 9–22.

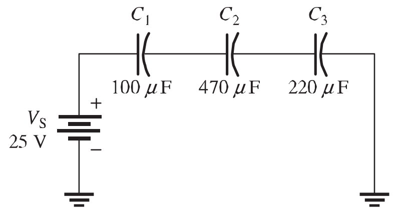

FIGURE 9–22

SOLUTION

Calculate the total capacitance.

$$C_T = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}} = \frac{1}{\frac{1}{100\,\mu\text{F}} + \frac{1}{470\,\mu\text{F}} + \frac{1}{220\,\mu\text{F}}} = 60\,\mu\text{F}$$

The voltages are as follows:

$$V_1 = \left(\frac{C_T}{C_1}\right)V_S = \left(\frac{60\,\mu\text{F}}{100\,\mu\text{F}}\right)25\text{ V} = \mathbf{15.0\text{ V}}$$

$$V_2 = \left(\frac{C_T}{C_2}\right)V_S = \left(\frac{60\,\mu\text{F}}{470\,\mu\text{F}}\right)25\text{ V} = \mathbf{3.19\text{ V}}$$

$$V_3 = \left(\frac{C_T}{C_3}\right)V_S = \left(\frac{60\,\mu\text{F}}{220\,\mu\text{F}}\right)25\text{ V} = \mathbf{6.82\text{ V}}$$

RELATED PROBLEM

Another 470 μF capacitor is connected in series with the existing capacitors in Figure 9–22. Determine the voltage across the new capacitor, assuming all the capacitors are initially uncharged.

MULTISIM

Open Multisim file E09-07; files are found at www.pearsonhighered.com/floyd. Measure the voltage across each of the capacitors and compare with the calculated values. Connect another 470 μF capacitor in series with the other three and measure the voltage across the new capacitor. Also measure the voltage across C_1, C_2, and C_3 and compare with the previous voltages. Did the values increase or decrease after the fourth capacitor was added? Why?

SYSTEM EXAMPLE 9–1

FEEDBACK IN A COLPITTS OSCILLATOR

Oscillators are circuits that generate a repetitive wave. Sinusoidal oscillators are used in many systems and for test signals. The circuit shown in Figure 9–23 is a sinusoidal oscillator designed for producing 550 kHz sine waves, which are used in AM radio receivers to mix with the incoming frequency from the antenna. This circuit is an example of one that uses a feedback network; a portion of the output is returned by the feedback network to the input to be amplified and reinforce the output. The feedback network is shown in the shaded box. Our focus is on the capacitors in the feedback network. Although the circuit is an ac circuit, the equations for capacitance and capacitor voltages developed with dc sources can be applied.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Capacitors C_1 and C_2 are in series and form a resonant circuit with L (resonant circuits are covered in Chapter 13). The total capacitance of C_1 in series with C_2 is found by Equation 9–7:

$$C_T = \frac{C_1C_2}{C_1 + C_2} = \frac{(2.2\text{ nF})(220\text{ pF})}{2.2\text{ nF} + 220\text{ pF}} = 200\text{ pF}$$

The amplifier's signals are referenced to ground, so the feedback signal, V_f, is the fraction of the output developed across C_1. Notice that V_f is returned to the input of the amplifier. You can find the fraction of output voltage that is fed back (V_f/V_{out}) by rearranging Equation 9–9 and substituting, so V_{out} for V_S and V_f for V_x.

$$\frac{V_x}{V_S} = \frac{V_f}{V_{out}} = \frac{C_T}{C_1} = \left(\frac{200\text{ pF}}{2.2\text{ nF}}\right) = 0.0909 = 9.1\%$$

As you can see from this result, the larger capacitor has the smaller fraction of the output voltage. The gain of the amplifier has to be high enough to keep oscillations going and restore the output to 100%. The feedback network determines the required gain for the amplifier portion of the Colpitts oscillator. The feedback has returned only 9.1% of the output, so the gain needs to be 11 (9.1% × 11 = 100%).

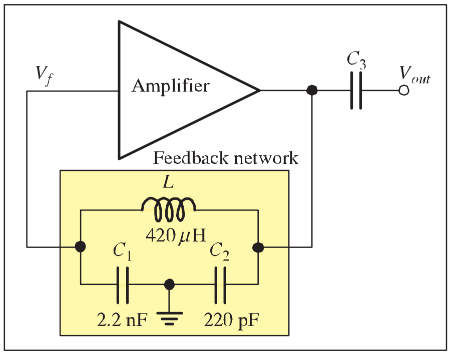

FIGURE 9–23 **A Colpitts oscillator showing the feedback network.**

SECTION 9–3 CHECKUP

1. Is the total capacitance of series capacitors less than or greater than the value of the smallest capacitor?
2. The following capacitors are in series: 100 pF, 220 pF, and 560 pF. What is the total capacitance?
3. A 0.01 μF and a 0.015 μF capacitor are in series. Determine the total capacitance.
4. Determine the voltage across the 0.01 μF capacitor in Question 3 if 10 V are connected across the two series capacitors.

9–4 PARALLEL CAPACITORS

Capacitances add when they are connected in parallel. As in all parallel circuits, the voltage across each capacitor is the same as the source voltage.

After completing this section, you should be able to

- **Analyze parallel capacitors**
  - **Determine total capacitance**

When capacitors are connected in parallel, the total capacitance is the sum of the individual capacitances because the effective plate area increases. The calculation of total parallel capacitance is analogous to the calculation of total series resistance (Chapter 4).

Figure 9–24 shows two parallel capacitors connected to a dc voltage source. When the switch is closed, as shown in part (a), current begins. A total amount of charge (Q_T) moves through the circuit in a certain period of time. Part of the total charge is stored by C_1 and part by C_2. The portion of the total charge that is stored by each capacitor depends on its capacitance value according to the relationship $Q = CV$.

Figure 9–24(b) shows the capacitors after they have been completely charged and the current has stopped. Since the voltage across both capacitors is the same, the larger capacitor

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

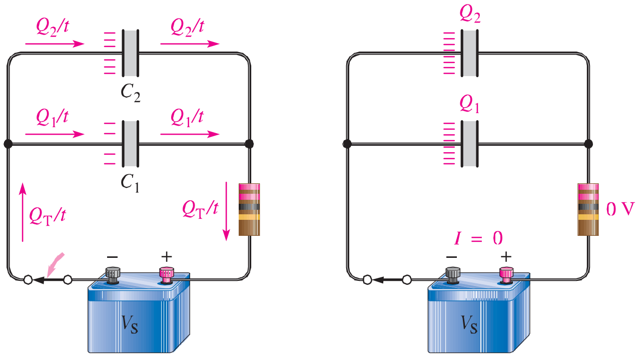

(a) The amount of charge on each capacitor is directly proportional to its capacitance value.

(b) $Q_T = Q_1 + Q_2$

FIGURE 9–24 **Capacitors in parallel produce a total capacitance that is the sum of the individual capacitances.**

stores more charge. If the capacitors are equal in value, they store an equal amount of charge. The charge stored by both of the capacitors together equals the total charge that was delivered from the source.

$$Q_T = Q_1 + Q_2$$

Since $Q = CV$, Equation 9–2, you can use substitution to get the following relationship:

$$C_TV_S = C_1V_S + C_2V_S$$

Because all the V_S terms are equal, they can be canceled. Therefore, the total capacitance for two capacitors in parallel is

$$\mathbf{C_T = C_1 + C_2} \qquad \text{(9–10)}$$

EXAMPLE 9–8

6208263 2013/12/10 12 27 66 5

What is the total capacitance in Figure 9–25? What is the voltage across each capacitor?

FIGURE 9–25

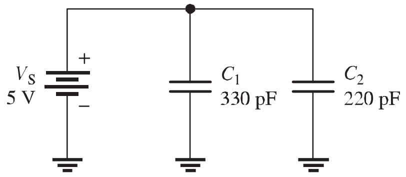

SOLUTION

The total capacitance is

$$C_T = C_1 + C_2 = 330\text{ pF} + 220\text{ pF} = \mathbf{550\ pF}$$

The voltage across each capacitor in parallel is equal to the source voltage.

$$V_S = V_1 = V_2 = \mathbf{5\ V}$$

RELATED PROBLEM

What is C_T if a 100 pF capacitor is connected in parallel with C_1 and C_2 in Figure 9–25?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

GENERAL FORMULA There are a few cases in practical circuits where you may find more than two parallel capacitors. In some industrial plants, parallel capacitors can be switched into the circuit depending on the load. For this reason, you should be aware of the general formula. Equation 9–10 can be extended to any number of capacitors in parallel, as shown in Figure 9–26. The expanded formula is as follows; the subscript n can be any number.

$$C_T = C_1 + C_2 + C_3 + \cdots + C_n \quad (9\text{–}11)$$

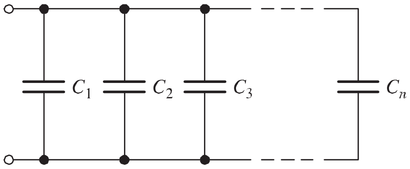

FIGURE 9–26 **General circuit with n capacitors in parallel.**

EXAMPLE 9–9

Determine C_T in Figure 9–27.

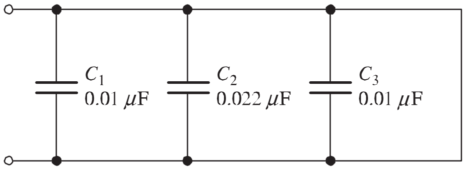

FIGURE 9–27

SOLUTION

$$C_T = C_1 + C_2 + C_3 = 0.01\ \mu\text{F} + 0.022\ \mu\text{F} + 0.01\ \mu\text{F} = \mathbf{0.042\ \mu F}$$

RELATED PROBLEM

If one more 0.01 μF capacitor is connected in parallel in Figure 9–27, what is C_T?

SECTION 9–4 CHECKUP

1. How is total parallel capacitance determined?
2. In a certain application, you need 0.05 μF. The only capacitor value available is 0.01 μF, and this capacitor is available in large quantities. How can you get the total capacitance that you need?
3. The following capacitors are in parallel: 10 pF, 56 pF, 33 pF, and 68 pF. What is C_T?

9–5 CAPACITORS IN DC CIRCUITS

A capacitor will charge up when it is connected to a dc voltage source. The buildup of charge across the plates occurs in a predictable manner that is dependent on the capacitance and the resistance in a circuit.

After completing this section, you should be able to

- **Describe how a capacitor operates in a dc switching circuit**
  - **Describe the charging and discharging of a capacitor**
  - **Define *RC time constant***
  - **Relate the time constant to charging and discharging**
  - **Write equations for the charging and discharging curves**
  - **Explain why a capacitor blocks constant dc**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Charging a Capacitor

A capacitor will charge when it is connected to a dc voltage source, as shown in Figure 9–28. The capacitor in part (a) is uncharged; that is, plate *A* and plate *B* have equal numbers of free electrons. When the switch is closed, as shown in part (b), the source moves electrons away from plate *A* through the circuit to plate *B* as the arrows indicate. As plate *A* loses electrons and plate *B* gains electrons, plate *A* becomes positive with respect to plate *B*. As this **charging** process continues, the voltage across the plates builds up rapidly until it is equal to the source voltage, V_S, but opposite in polarity, as shown in part (c). When the capacitor is fully charged, there is no current.

A capacitor blocks constant dc.

When the charged capacitor is disconnected from the source, as shown in Figure 9–28(d), it remains charged for long periods of time, depending on its leakage resistance. The charge on an electrolytic capacitor generally leaks off more rapidly than in other types of capacitors.

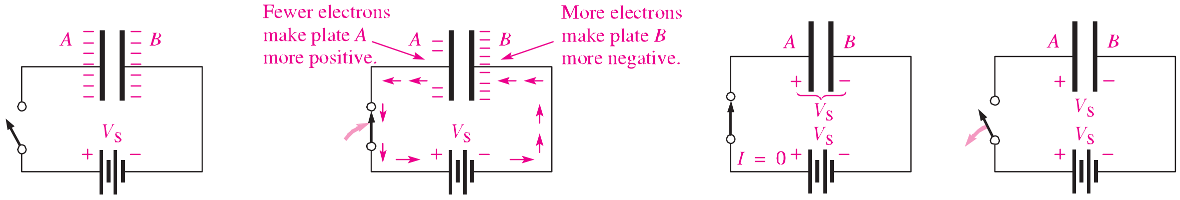
A
B
V_S
+
−
Fewer electrons make plate A more positive.
More electrons make plate B more negative.
$I = 0$

(a) Uncharged (b) Charging (c) Fully charged (d) Retains charge

FIGURE 9–28 **Charging a capacitor.**

Discharging a Capacitor

When a conductor is connected across a charged capacitor, as shown in Figure 9–29, the capacitor will discharge. In this particular case, a very low resistance path (the conductor) is connected across the capacitor with a switch. Before the switch is closed, the capacitor is charged to 50 V, as indicated in part (a). When the switch is closed, as shown in part (b), the excess electrons on plate *B* move through the circuit to plate *A* (indicated by the arrows); as a result of the current through the low resistance of the conductor, the energy stored by the capacitor is dissipated in the resistance of the conductor. The charge is neutralized when the numbers of free electrons on both plates are again equal. At this time, the voltage across the capacitor is zero, and the capacitor is completely discharged, as shown in part (c).

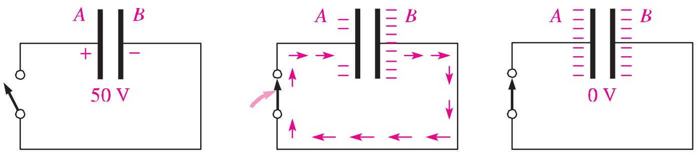

(a) Retains charge (b) Discharging (c) Uncharged

FIGURE 9–29 **Discharging a capacitor.**

Current and Voltage During Charging and Discharging

Notice in Figures 9–28 and 9–29 that the direction of electron flow during discharge is opposite to that during charging. It is important to understand that, ideally, *there is no current through the dielectric of the capacitor during charging or discharging because the*

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

dielectric is an insulating material. There is current from one plate to the other only through the external circuit.

Figure 9–30(a) shows a capacitor connected in series with a resistor and a switch to a dc voltage source. Initially, the switch is open and the capacitor is uncharged with zero volts across its plates. At the instant the switch is closed, the current jumps to its maximum value and the capacitor begins to charge. The current is maximum initially because the capacitor has zero volts across it and, therefore, effectively acts as a short; thus, the current is limited only by the resistance. As time passes and the capacitor charges, the current decreases and the voltage across the capacitor (V_C) increases. The resistor voltage is proportional to the current during this charging period.

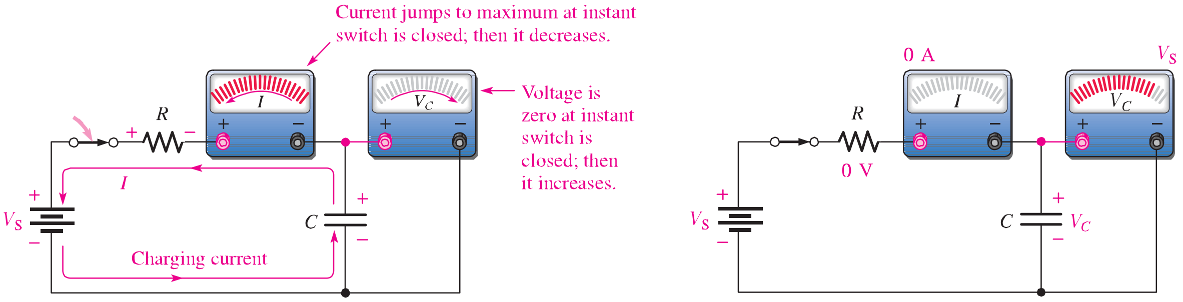

Current jumps to maximum at instant switch is closed; then it decreases.

Voltage is zero at instant switch is closed; then it increases.

(a) Charging: Capacitor voltage increases as the current and resistor voltage decrease.

(b) Fully charged: Capacitor voltage equals source voltage. The current is zero.

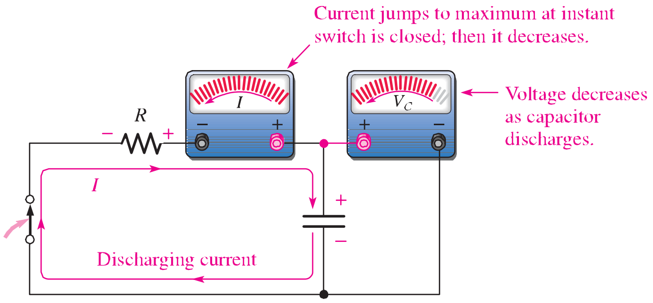

Current jumps to maximum at instant switch is closed; then it decreases.

Voltage decreases as capacitor discharges.

(c) Discharging: Capacitor voltage, resistor voltage, and the current decrease from their initial maximum values. Note that the discharge current is opposite to the charge current.

FIGURE 9–30 **Current and voltage in a charging and discharging capacitor.**

After a certain period of time, the capacitor reaches full charge. At this point, the current is zero and the capacitor voltage is equal to the dc source voltage, as shown in Figure 9–30(b). If the switch were opened now, the capacitor would retain its full charge (neglecting any leakage).

In Figure 9–30(c), the voltage source has been removed. When the switch is closed, the capacitor begins to discharge. Initially, the current jumps to a maximum but in a direction opposite to its direction during charging. As time passes, the current and capacitor voltage decrease. The resistor voltage is always proportional to the current. When the capacitor has fully discharged, the current and the capacitor voltage are zero.

Remember these two rules about capacitors in dc circuits:

1. A capacitor appears as an *open* to constant voltage.
2. A capacitor appears as a *short* to an instantaneous change in voltage.

Now let's examine in more detail how the voltage and current change with time in a capacitive circuit.

SAFETY NOTE

Capacitors can store charge for a long time after power has been turned off. Be careful when touching or handling capacitors in or out of a circuit. If you touch the leads, you may be in for a shock as the capacitor discharges through you! It is good practice to discharge a capacitor using a shorting tool with an insulated grip before handling the capacitor.

6238262 2017/12/10 12.27.66.5

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Surge Resistor

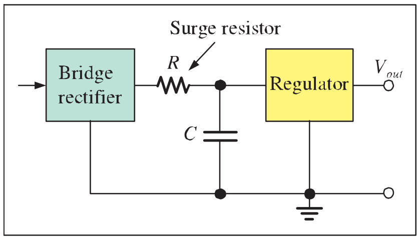

A typical small power supply for electronic systems has a bridge rectifier to change ac to pulsating dc, a filter capacitor to smooth the pulses, and a regulator to provide a constant dc output. Many times, you will see a small-value resistor between the rectifier and the regulator. This resistor limits inrush current to the capacitor when the supply is first turned on. The capacitor is typically a large value and has the highest current when it is initially uncharged. By limiting charging current, the surge resistor prevents blowing a fuse or tripping a circuit breaker.

SYSTEM NOTE

The *RC* Time Constant

In a practical situation, there cannot be capacitance without some resistance in a circuit. It may simply be the small resistance of a wire, or it may be a designed-in resistance. Because of this, the charging and discharging characteristics of a capacitor must always be considered with the associated series resistance included. The resistance introduces the element of *time* in the charging and discharging of a capacitor.

When a capacitor charges or discharges through a resistance, a certain time is required for the capacitor to charge fully or discharge fully. The voltage across a capacitor cannot change instantaneously because a finite time is required to move charge from one point to another. The time constant of a series *RC* circuit determines the rate at which the capacitor charges or discharges.

The *RC* time constant is a fixed time interval that equals the product of the resistance and the capacitance in a series *RC* circuit.

The time constant is expressed in units of seconds when resistance is in ohms and capacitance is in farads. It is symbolized by τ (Greek letter tau), and the formula is

$$\tau = RC \tag{9–12}$$

Recall that $I = Q/t$. The current depends on the amount of charge moved in a given time. When the resistance is increased, the charging current is reduced, thus increasing the charging time of the capacitor. When the capacitance is increased, the amount of charge increases; thus, for the same current, more time is required to charge the capacitor.

EXAMPLE 9–10

A series *RC* circuit has a resistance of 1.0 MΩ and a capacitance of 4.7 μF. What is the time constant?

SOLUTION

$$\tau = RC = (1.0 \times 10^6\ \Omega)(4.7 \times 10^{-6}\ \text{F}) = \mathbf{4.7\ s}$$

RELATED PROBLEM

A series *RC* circuit has a 270 kΩ resistor and a 3300 pF capacitor. What is the time constant in μs?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

When a capacitor is charging or discharging between two voltage levels, the charge on the capacitor changes by approximately 63% of the difference in the levels in one time constant. An uncharged capacitor charges to approximately 63% of its fully charged voltage in one time constant. When a fully charged capacitor is discharging, its voltage drops to approximately 100% − 63% = 37% of its initial voltage in one time constant. This change also corresponds to a 63% change.

The Charging and Discharging Curves

A capacitor charges and discharges following a nonlinear curve, as shown in Figure 9–31. In these graphs, the approximate percentage of full charge is shown at each time-constant interval. This type of curve follows a precise mathematical formula and is called an **exponential** curve. The charging curve is an increasing exponential, and the discharging curve is a decreasing exponential. It takes five time constants to reach 99% (considered 100%) of the final voltage. This five time-constant interval is generally accepted as the time to fully charge or discharge a capacitor and is called the **transient time**.

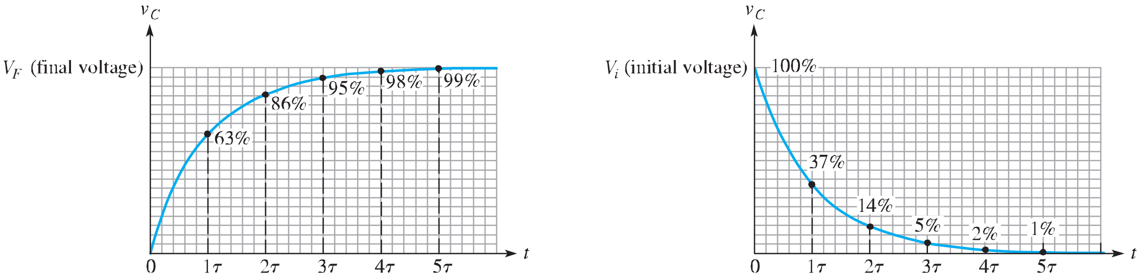

(a) Charging curve with percentages of the final voltage

(b) Discharging curve with percentages of the initial voltage

FIGURE 9–31 **Exponential voltage curves for the charging and discharging of a capacitor in an *RC* circuit.**

GENERAL FORMULA The general expressions for either increasing or decreasing exponential curves are given in the following equations for both instantaneous voltage and current:

$$v = V_F + (V_i - V_F)e^{-t/\tau} \qquad (9\text{–}13)$$

$$i = I_F + (I_i - I_F)e^{-t/\tau} \qquad (9\text{–}14)$$

where V_F and I_F are the final values of voltage and current and V_i and I_i are the initial values of voltage and current. The lowercase italic letters v and i are the instantaneous values of the capacitor voltage and current at time t, and e is the base of natural logarithms. The e^x key on a calculator makes it easy to work with this exponential term.

CHARGING FROM ZERO The formula for the special case in which an increasing exponential voltage curve begins at zero ($V_i = 0$), as shown in Figure 9–31(a), is given in Equation 9–15. It is developed as follows, starting with the general formula, Equation 9–13.

$$v = V_F + (V_i - V_F)e^{-t/\tau} = V_F + (0 - V_F)e^{-t/RC} = V_F - V_F e^{-t/RC}$$

Factoring out V_F, you have

$$v = V_F(1 - e^{-t/RC}) \qquad (9\text{–}15)$$

Using Equation 9–15, you can calculate the value of the charging voltage of a capacitor at any instant of time if it is initially uncharged. You can calculate an increasing current by substituting i for v and I_F for V_F in Equation 9–15.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

EXAMPLE 9–11

In Figure 9–32, determine the capacitor voltage 50 μs after the switch is closed if the capacitor initially is uncharged. Draw the charging curve.

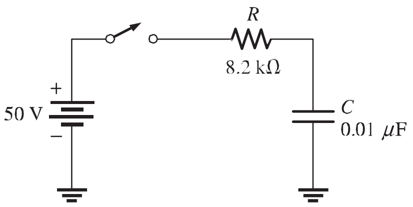

FIGURE 9–32

SOLUTION

The time constant is

$$\tau = RC = (8.2\ \text{k}\Omega)(0.01\ \mu\text{F}) = 82\ \mu\text{s}$$

The voltage to which the capacitor will fully charge is 50 V (this is V_F). The initial voltage is zero. Notice that 50 μs is less than one time constant; so the capacitor will charge less than 63% of the full voltage in that time.

$$v_C = V_F(1 - e^{-t/RC}) = (50\ \text{V})(1 - e^{-50\mu\text{s}/82\mu\text{s}}) = \mathbf{22.8\ V}$$

The charging curve for the capacitor is shown in Figure 9–33.

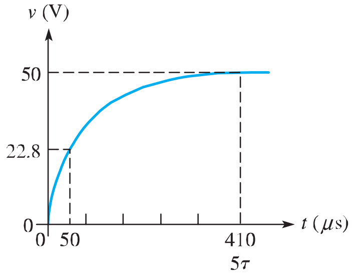

FIGURE 9–33

RELATED PROBLEM

Determine the capacitor voltage 15 μs after the switch is closed in Figure 9–32.

DISCHARGING TO ZERO The formula for the special case in which a decreasing exponential voltage curve ends at zero ($V_F = 0$), as shown in Figure 9–31(b), is derived from the general formula as follows:

$$v = V_F + (V_i - V_F)e^{-t/\tau} = 0 + (V_i - 0)e^{-t/RC}$$

This reduces to

$$\boxed{v = V_i e^{-t/RC}} \qquad \text{(9–16)}$$

where V_i is the voltage at the beginning of the discharge. You can use this formula to calculate the discharging voltage at any instant. The exponent $-t/RC$ can also be written as $-t/\tau$.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 9–12

Determine the capacitor voltage in Figure 9–34 at a point in time 6 ms after the switch is closed. Draw the discharging curve.

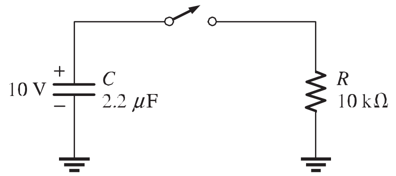

FIGURE 9–34

SOLUTION

The discharge time constant is

$$\tau = RC = (10\text{ k}\Omega)(2.2\ \mu\text{F}) = 22\text{ ms}$$

The initial capacitor voltage is 10 V. Notice that 6 ms is less than one time constant, so the capacitor will discharge less than 63%. Therefore, it will have a voltage greater than 37% of the initial voltage at 6 ms.

$$v_C = V_ie^{-t/RC} = (10\text{ V})e^{-6\text{ms}/22\text{ms}} = \mathbf{7.61\ V}$$

The discharging curve for the capacitor is shown in Figure 9–35.

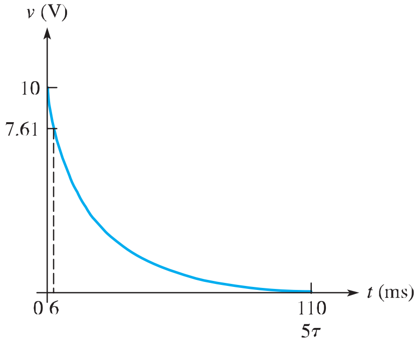

FIGURE 9–35

RELATED PROBLEM

In Figure 9–34, change R to 2.2 kΩ and determine the capacitor voltage 1 ms after the switch is closed.

GRAPHICAL METHOD USING UNIVERSAL EXPONENTIAL CURVES

The universal curves in Figure 9–36 provide a graphic solution of the charge and discharge of capacitors. Example 9–13 illustrates this graphical method.

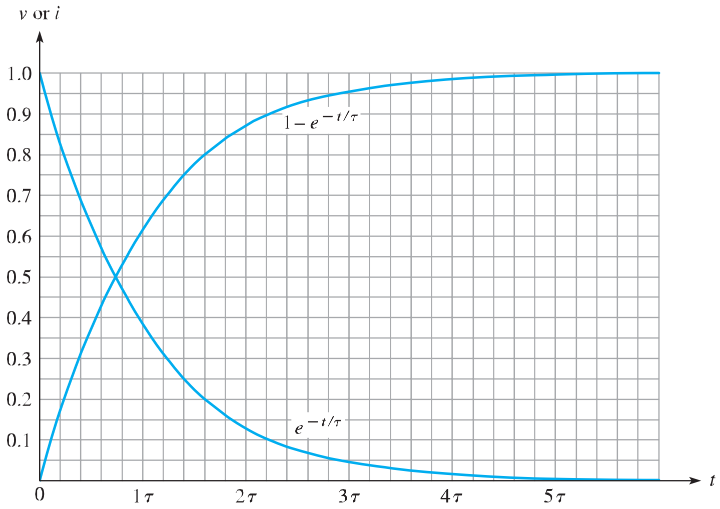

FIGURE 9–36 **Normalized universal exponential curves.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 9–13

How long will it take the capacitor in Figure 9–37 to charge to 75 V? What is the capacitor voltage 2 ms after the switch is closed? Use the normalized universal curves in Figure 9–36 to determine the answers.

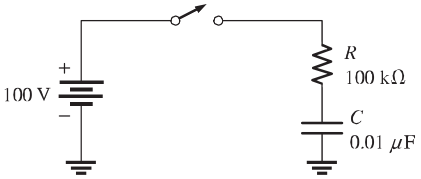

FIGURE 9–37

SOLUTION

The final voltage is 100 V, which is the 100% level (1.0) on the graph. The value 75 V is 75% of the maximum, or 0.75 on the graph. You can see that this value occurs at 1.4 time constants. In this circuit one time constant is $RC = (100\,\text{k}\Omega)(0.01\,\mu\text{F}) = 1\,\text{ms}$. Therefore, the capacitor voltage reaches 75 V at 1.4 ms after the switch is closed.

The capacitor is at approximately 86 V (0.86 on the vertical axis) in 2 ms. These graphic solutions are shown in Figure 9–38.

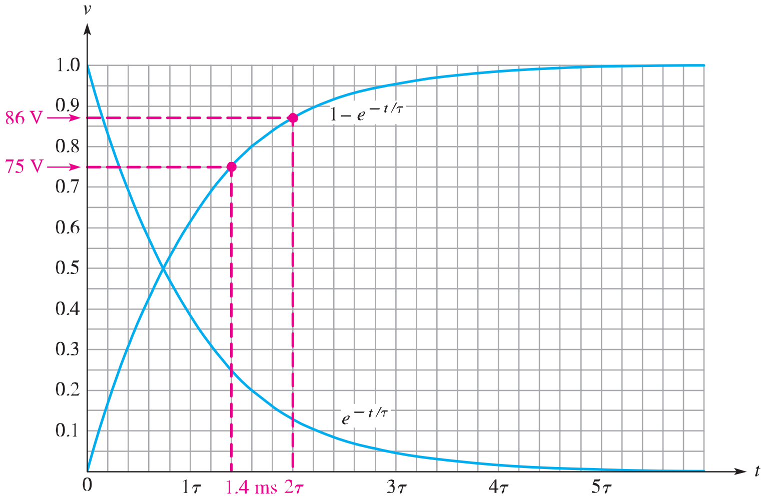

FIGURE 9–38

RELATED PROBLEM

Using the normalized universal exponential curves, determine how long it will take the capacitor in Figure 9–37 to charge to 50 V from zero. What is the capacitor voltage 3 ms after switch closure?

Response to a Square Wave

A common case that illustrates the rising and falling exponential occurs when an *RC* circuit is driven with a square wave that has a long period compared to the time constant. The square wave provides on and off action but, unlike a single switch, it provides a discharge path back through the generator when the wave drops back to zero.

When the square wave rises, the voltage across the capacitor rises exponentially toward the maximum value of the square wave in a time that depends on the time constant. When the square wave returns to the zero level, the capacitor voltage decreases exponentially, again depending on the time constant. The internal resistance of the square wave generator is part of the *RC* time constant; however, it can be ignored if it is small compared to *R*. The following example shows the waveforms for the case where the period is long compared to the time constant; other cases will be covered in detail in Chapter 15.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 9–14

For the circuit in Figure 9–39(a), calculate the voltage across the capacitor every 0.1 ms for one complete period of the input waveform shown in Figure 9–39(b). Then sketch the capacitor waveform. Assume the internal resistance of the generator is negligible.

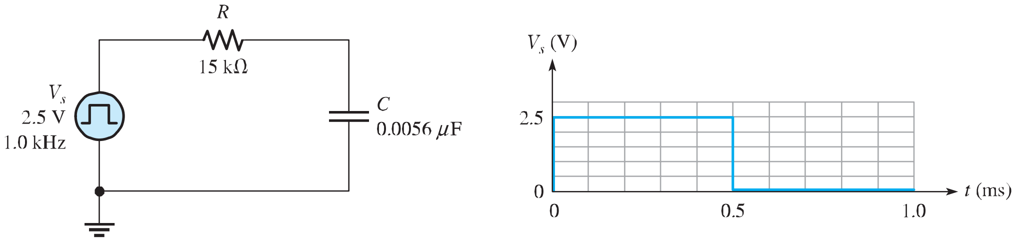

FIGURE 9–39

SOLUTION

$$\tau = RC = (15\text{ k}\Omega)(0.0056\ \mu\text{F}) = 0.084\text{ ms}$$

The period of the square wave is 1 ms, which is approximately 12τ. This means that 6τ will elapse after each change of the pulse, allowing the capacitor to fully charge and fully discharge.

For the increasing exponential,

$$v = V_F(1 - e^{-t/RC}) = V_F(1 - e^{-t/\tau})$$

At 0.1 ms: $v = 2.5\text{ V}(1 - e^{-0.1\text{ms}/0.084\text{ms}}) = 1.74\text{ V}$

At 0.2 ms: $v = 2.5\text{ V}(1 - e^{-0.2\text{ms}/0.084\text{ms}}) = 2.27\text{ V}$

At 0.3 ms: $v = 2.5\text{ V}(1 - e^{-0.3\text{ms}/0.084\text{ms}}) = 2.43\text{ V}$

At 0.4 ms: $v = 2.5\text{ V}(1 - e^{-0.4\text{ms}/0.084\text{ms}}) = 2.48\text{ V}$

At 0.5 ms: $v = 2.5\text{ V}(1 - e^{-0.5\text{ms}/0.084\text{ms}}) = 2.49\text{ V}$

For the decreasing exponential,

$$v = V_i(e^{-t/RC}) = V_i(e^{-t/\tau})$$

In the equation, time is shown from the point when the change occurs (subtracting 0.5 ms from the actual time). For example, at 0.6 ms, $t = 0.6\text{ ms} - 0.5\text{ ms} = 0.1\text{ ms}$.

At 0.6 ms: $v = 2.5\text{ V}(e^{-0.1\text{ms}/0.084\text{ms}}) = 0.76\text{ V}$

At 0.7 ms: $v = 2.5\text{ V}(e^{-0.2\text{ms}/0.084\text{ms}}) = 0.23\text{ V}$

At 0.8 ms: $v = 2.5\text{ V}(e^{-0.3\text{ms}/0.084\text{ms}}) = 0.07\text{ V}$

At 0.9 ms: $v = 2.5\text{ V}(e^{-0.4\text{ms}/0.084\text{ms}}) = 0.02\text{ V}$

At 1.0 ms: $v = 2.5\text{ V}(e^{-0.5\text{ms}/0.084\text{ms}}) = 0.01\text{ V}$

Figure 9–40 is a graph of these results.

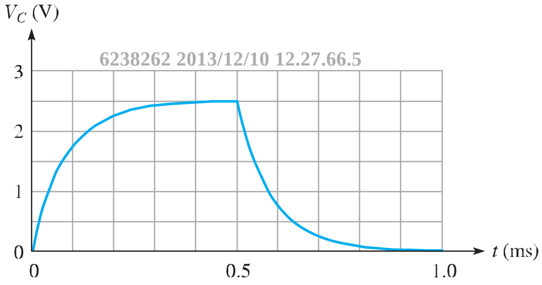

6238262 2013/12/10 12.27.66.5

FIGURE 9–40

RELATED PROBLEM

What is the capacitor voltage at 0.65 ms?

SECTION 9–5 CHECKUP

1. Determine the time constant when $R = 1.2\text{ k}\Omega$ and $C = 1000\text{ pF}$.
2. If the circuit in Question 1 is charged with a 5 V source, how long will it take the initially uncharged capacitor to reach full charge? At full charge, what is the capacitor voltage?
3. A certain circuit has a time constant of 1 ms. If it is charged with a 10 V battery, what will the capacitor voltage be at each of the following times: 2 ms, 3 ms, 4 ms, and 5 ms? The capacitor is initially uncharged.
4. A capacitor is charged to 100 V. If it is discharged through a resistor, what is the capacitor voltage at one time constant?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

9–6 CAPACITORS IN AC CIRCUITS

As you know, a capacitor blocks constant dc. A capacitor passes ac with an amount of opposition, called capacitive reactance, that depends on the frequency of the ac.

After completing this section, you should be able to

- **Describe how a capacitor operates in an ac circuit**
  - **Define *capacitive reactance***
  - **Determine the value of capacitive reactance in a given circuit**
  - **Calculate the capacitance for series and parallel capacitors**
  - **Explain why a capacitor causes a phase shift between voltage and current**
  - **Discuss instantaneous, true, and reactive power in a capacitor**

Capacitive Reactance, X_C

In Figure 9–41, a capacitor is shown connected to a sinusoidal voltage source. When the source voltage is held at a constant amplitude value and its frequency is increased, the amplitude of the current increases. Also, when the frequency of the source is decreased, the current amplitude decreases.

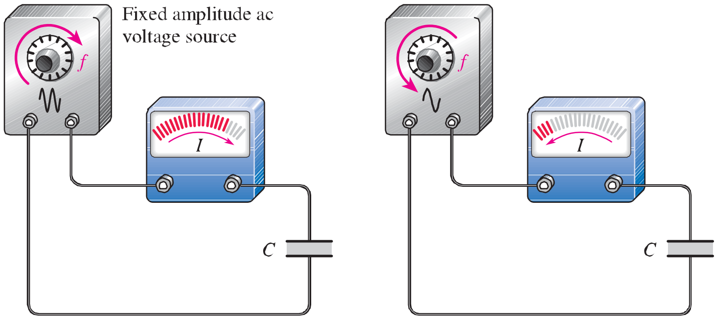
Fixed amplitude ac voltage source
f
I
C
f
I
C

(a) Current increases when the frequency increases.

(b) Current decreases when the frequency decreases.

FIGURE 9–41 **The current in a capacitive circuit varies directly with the frequency of the source voltage.**

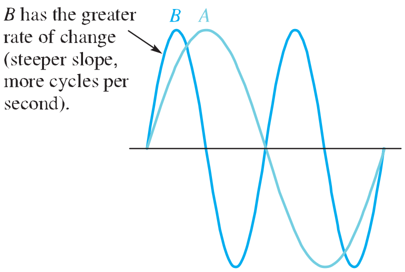

FIGURE 9–42 **Rate of change of a sine wave increases when frequency increases.**

When the frequency of the voltage increases, its rate of change also increases. This relationship is illustrated in Figure 9–42, where the frequency is doubled. Now, if the rate at which the voltage is changing increases, the amount of charge moving through the circuit in a given period of time must also increase. More charge in a given period of time means more current. For example, a tenfold increase in frequency means that the capacitor is charging and discharging 10 times more often in a given time interval. The rate of charge movement has increased 10 times. This means the current has increased by 10 because $I = Q/t$.

An increase in the amount of current with a fixed amount of voltage indicates that opposition to the current has decreased. Therefore, the capacitor offers opposition to current, and that opposition varies *inversely* with frequency.

The opposition to sinusoidal current in a capacitor is called capacitive reactance.

The symbol for capacitive reactance is X_C, and its unit is the ohm (Ω).

You have just seen how frequency affects the opposition to current (capacitive reactance) in a capacitor. Now let's see how the capacitance (C) itself affects the reactance.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Figure 9–43(a) shows that when a sinusoidal voltage with a fixed amplitude and fixed frequency is applied to a 1 μF capacitor, there is a certain amount of alternating current. When the capacitance value is increased to 2 μF, the current increases, as shown in Figure 9–43(b). Thus, when the capacitance is increased, the opposition to current (capacitive reactance) decreases. Therefore, not only is the capacitive reactance *inversely* proportional to frequency, but it is also *inversely* proportional to capacitance. This relationship can be stated as follows:

X_C is proportional to $\frac{1}{fC}$.

It can be proven that the constant of proportionality that relates X_C to $1/fC$ is $1/2\pi$. Therefore, the formula for capacitive reactance (X_C) is

$$X_C = \frac{1}{2\pi fC} \qquad (9\text{–}17)$$

X_C is in ohms when f is in hertz and C is in farads. The 2π term comes from the fact that, as you learned in Chapter 8, a sine wave can be described in terms of rotational motion, and one revolution contains 2π radians.

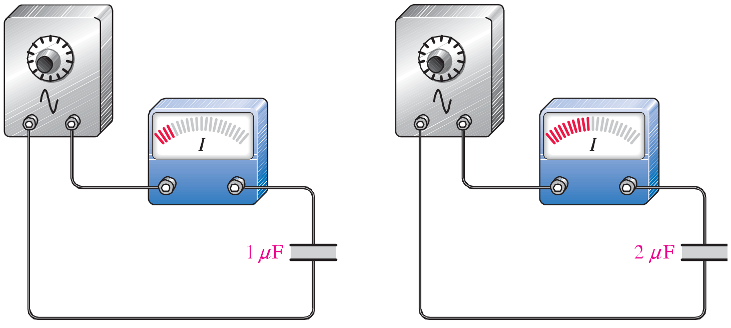

(a) Less capacitance, less current (b) More capacitance, more current

FIGURE 9–43 **For a fixed voltage and fixed frequency, the current varies directly with the capacitance value.**

EXAMPLE 9–15

A sinusoidal voltage is applied to a capacitor, as shown in Figure 9–44. The frequency of the sine wave is 1.0 kHz. Determine the capacitive reactance.

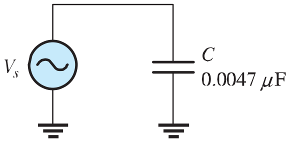

FIGURE 9–44

SOLUTION

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi(1.0 \times 10^3\ \text{Hz})(0.0047 \times 10^{-6}\ \text{F})} = \mathbf{33.9\ k\Omega}$$

RELATED PROBLEM

Determine the frequency required to make the capacitive reactance in Figure 9–44 equal to 10 kΩ.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Reactance for Series Capacitors

When capacitors are in series in an ac circuit, the total capacitance is smaller than the smallest individual capacitance. Because the total capacitance is smaller, the total capacitive reactance (opposition to current) must be larger than any individual capacitive reactance. With series capacitors, the total capacitive reactance ($X_{C(tot)}$) is the sum of the individual reactances.

$$X_{C(tot)} = X_{C1} + X_{C2} + X_{C3} + \cdots + X_{Cn} \tag{9–18}$$

Compare this formula with Equation 4–1 for finding the total resistance of series resistors. In both cases, you simply add the individual oppositions.

Reactance for Parallel Capacitors

In ac circuits with parallel capacitors, the total capacitance is the sum of the capacitances. Recall that the capacitive reactance is inversely proportional to the capacitance. Because the total parallel capacitance is larger than any individual capacitances, the total capacitive reactance must be smaller than the reactance of any individual capacitor. With parallel capacitors, the total reactance is found by

$$X_{C(tot)} = \frac{1}{\dfrac{1}{X_{C1}} + \dfrac{1}{X_{C2}} + \dfrac{1}{X_{C3}} + \cdots + \dfrac{1}{X_{Cn}}} \tag{9–19}$$

Compare this formula with Equation 5–1 for parallel resistors. As in the case of parallel resistors, the total opposition (resistance or reactance) is the reciprocal of the sum of the reciprocals of the individual oppositions.

For two capacitors in parallel, Equation 9–19 can be reduced to the product-over-sum form. This is useful because, for most practical circuits, more than two capacitors in parallel is not common.

$$X_{C(tot)} = \frac{X_{C1}X_{C2}}{X_{C1} + X_{C2}}$$

6238262 2013/12/10 12.27.66.5

EXAMPLE 9–16

What is the total capacitive reactance in each circuit in Figure 9–45?

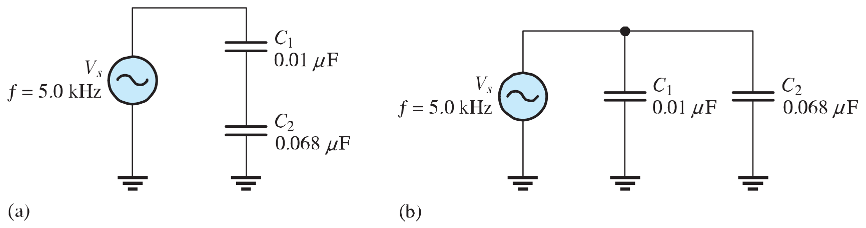

FIGURE 9–45

SOLUTION

The reactances of the individual capacitors are the same in both circuits.

$$X_{C1} = \frac{1}{2\pi f C_1} = \frac{1}{2\pi(5.0\text{ kHz})(0.01\ \mu\text{F})} = 3.18\text{ k}\Omega$$

$$X_{C2} = \frac{1}{2\pi f C_2} = \frac{1}{2\pi(5.0\text{ kHz})(0.068\ \mu\text{F})} = 468\ \Omega$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Series Circuit: For the capacitors in series in Figure 9–45(a), the total reactance is the sum of X_{C1} and X_{C2}, as given in Equation 9–18.

$$X_{C(tot)} = X_{C1} + X_{C2} = 3.18\text{ k}\Omega + 468\ \Omega = \mathbf{3.65\text{ k}\Omega}$$

Alternatively, you can obtain the total series reactance by first finding the total capacitance using Equation 9–7. Then substitute that value into Equation 9–17 to calculate the total reactance.

$$C_{tot} = \frac{C_1C_2}{C_1 + C_2} = \frac{(0.01\ \mu\text{F})(0.068\ \mu\text{F})}{0.01\ \mu\text{F} + 0.068\ \mu\text{F}} = 0.0087\ \mu\text{F}$$

$$X_{C(tot)} = \frac{1}{2\pi fC_{tot}} = \frac{1}{2\pi(5.0\text{ kHz})(0.0087\ \mu\text{F})} = \mathbf{3.65\text{ k}\Omega}$$

Parallel Circuit: For the capacitors in parallel in Figure 9–45(b), determine the total reactance from the product-over-sum rule using X_{C1} and X_{C2}.

$$X_{C(tot)} = \frac{X_{C1}X_{C2}}{X_{C1} + X_{C2}} = \frac{(3.18\text{ k}\Omega)(468\ \Omega)}{3.18\text{ k}\Omega + 468\ \Omega} = 408\ \Omega$$

RELATED PROBLEM

Determine the total parallel capacitive reactance by first finding the total capacitance.

OHM'S LAW The reactance of a capacitor is analogous to the resistance of a resistor. In fact, both are expressed in ohms. Since both R and X_C are forms of opposition to current, Ohm's law applies to capacitive circuits as well as to resistive circuits.

$$I = \frac{V}{X_C} \qquad (9\text{–}20)$$

When applying Ohm's law in ac circuits, you must express both the current and the voltage in the same way, that is, both in rms, both in peak, and so on.

EXAMPLE 9–17

Determine the rms current in Figure 9–46.

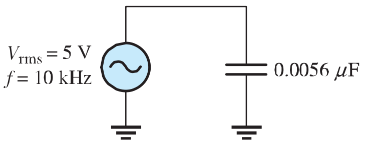

FIGURE 9–46

SOLUTION

First, find X_C.

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi(10 \times 10^3\text{ Hz})(0.0056 \times 10^{-6}\text{ F})} = 2.84\text{ k}\Omega$$

Then, apply Ohm's law.

$$I_{rms} = \frac{V_{rms}}{X_C} = \frac{5\text{ V}}{2.84\text{ k}\Omega} = \mathbf{1.76\text{ mA}}$$

RELATED PROBLEM

Change the frequency in Figure 9–46 to 25 kHz and determine the rms value of the current.

MULTISIM

Open Multisim file E09-17. Measure the rms current and compare with the calculated value. Change the frequency of the source voltage to 25 kHz and measure the current.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Capacitive Voltage Divider

In ac circuits, capacitors can be used in applications that require a voltage divider. The voltage across a series capacitor was given as Equation 9–9, which is repeated here.

$$V_x = \left(\frac{C_{tot}}{C_x}\right)V_s$$

A resistive voltage divider is expressed in terms of a resistance ratio, which is a ratio of oppositions. You can think of the capacitive voltage divider by applying this idea from a resistive divider, but using reactance in place of resistance. The equation for the voltage across a capacitor in a capacitive voltage divider can be written as

$$\boldsymbol{V_x = \left(\frac{X_{Cx}}{X_{C(tot)}}\right)V_s} \qquad \textbf{(9–21)}$$

where X_{Cx} is the reactance of capacitor C_x, $X_{C(tot)}$ is the total capacitive reactance, and V_x is the voltage across capacitor C_x. Compare this equation with Equation 4–5. Either Equation 9–9 or Equation 9–21 can be used to find the voltage from a divider as illustrated in the following example.

EXAMPLE 9–18

What is the voltage across C_2 in the circuit of Figure 9–47?

FIGURE 9–47

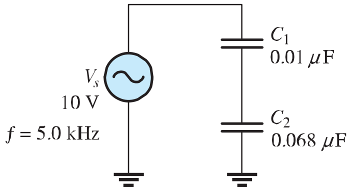

SOLUTION

The reactance of the individual capacitors and the total reactance were determined in Example 9–16. Substituting into Equation 9–21,

$$V_2 = \left(\frac{X_{C2}}{X_{C(tot)}}\right)V_s = \left(\frac{468\ \Omega}{3.65\ \text{k}\Omega}\right)10\ \text{V} = \mathbf{1.28\ V}$$

Notice that the voltage across the larger capacitor is the smaller fraction of the total. You can obtain the same result from Equation 9–9:

$$V_2 = \left(\frac{C_{tot}}{C_2}\right)V_s = \left(\frac{0.0087\ \mu\text{F}}{0.068\ \mu\text{F}}\right)10\ \text{V} = \mathbf{1.28\ V}$$

RELATED PROBLEM

Use Equation 9–21 to determine the voltage across C_1.

Current Leads Capacitor Voltage by 90°

A sinusoidal voltage is shown in Figure 9–48. Notice that the rate at which the voltage is changing varies along the sine wave curve, as indicated by the "steepness" of the curve. At the zero crossings, the curve is changing at a faster rate than anywhere else along the curve. At the peaks, the curve has a zero rate of change because it has just reached its maximum and is at the point of changing direction.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

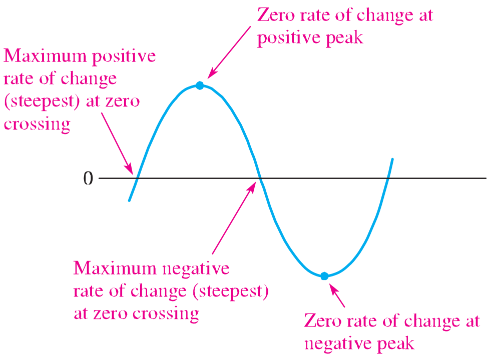
Zero rate of change at positive peak
Maximum positive rate of change (steepest) at zero crossing
0
Maximum negative rate of change (steepest) at zero crossing
Zero rate of change at negative peak

FIGURE 9–48 The rates of change of a sine wave.

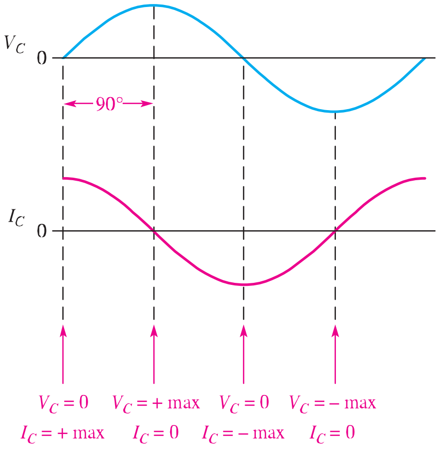
V_C
0
90°
I_C
0
$V_C = 0$
$I_C = +$ max
$V_C = +$ max
$I_C = 0$
$V_C = 0$
$I_C = -$ max
$V_C = -$ max
$I_C = 0$

FIGURE 9–49 Current is always leading the capacitor voltage by 90°.

The amount of charge stored by a capacitor determines the voltage across it. Therefore, the rate at which the charge is moved ($Q/t = I$) from one plate to the other determines the rate at which the voltage changes. When the current is changing at its maximum rate (at the zero crossings), the voltage is at its maximum value (peak). When the current is changing at its minimum rate (zero at the peaks), the voltage is at its minimum value (zero). This phase relationship is illustrated in Figure 9–49. As you can see, the current peaks occur a quarter of a cycle before the voltage peaks. Thus, the current leads the voltage by 90°.

Power in a Capacitor

As discussed earlier in this chapter, a charged capacitor stores energy in the electric field within the dielectric. An ideal capacitor does not dissipate energy; it only stores energy temporarily. When an ac voltage is applied to a capacitor, energy is stored by the capacitor during one-quarter of the voltage cycle. Then the stored energy is returned to the source during another quarter of the cycle. There is no net energy loss. Figure 9–50 shows the power curve that results from one cycle of capacitor voltage and current.

6238262 2013/12/10 12.27.66.5

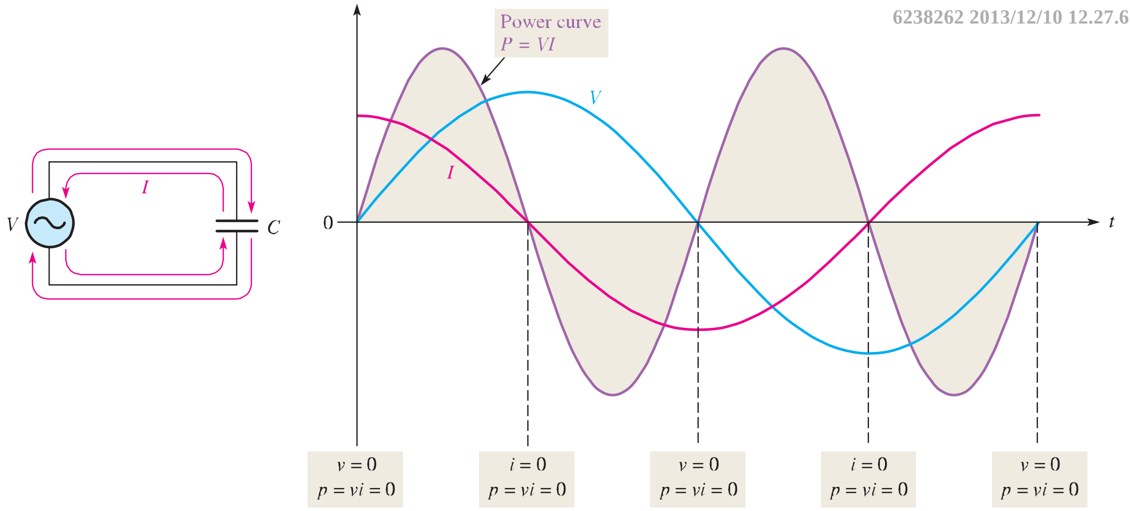

FIGURE 9–50 Power curve for a capacitor.

INSTANTANEOUS POWER (p) The product of instantaneous voltage, v, and instantaneous current, i, gives **instantaneous power**, p. At points where v or i is zero, p is also zero. When both v and i are positive, p is also positive. When either v or i is positive and the

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

other negative, p is negative. When both v and i are negative, p is positive. As you can see, the power follows a sinusoidal-shaped curve. Positive values of power indicate that energy is stored by the capacitor. Negative values of power indicate that energy is returned from the capacitor to the source. Note that the power fluctuates at a frequency twice that of the voltage or current, as energy is alternately stored and returned to the source.

TRUE POWER (P_{true}) Ideally, all of the energy stored by a capacitor during the positive portion of the power cycle is returned to the source during the negative portion. No net energy is lost due to conversion to heat in the capacitor, so the **true power** is zero. Actually, because of leakage and foil resistance in a practical capacitor, a small percentage of the total power is dissipated in the form of true power.

REACTIVE POWER (P_r) The rate at which a capacitor stores or returns energy is called its **reactive power**, P_r. The reactive power is a nonzero quantity, because at any instant in time, the capacitor is actually taking energy from the source or returning energy to it. Reactive power does not represent an energy loss. The following formulas apply:

$$P_r = V_{\text{rms}} I_{\text{rms}} \qquad (9\text{–}22)$$

$$P_r = \frac{V_{\text{rms}}^2}{X_C} \qquad (9\text{–}23)$$

$$P_r = I_{\text{rms}}^2 X_C \qquad (9\text{–}24)$$

Notice that these equations are of the same form as those introduced in Chapter 3 for true power in a resistor. The voltage and current are expressed as rms values. The unit of reactive power is **VAR (volt-ampere reactive)**.

EXAMPLE 9–19

Determine the true power and the reactive power in Figure 9–51.

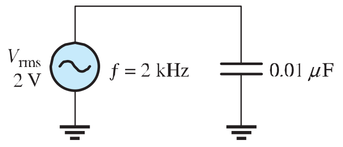

FIGURE 9–51

SOLUTION

The true power P_{true} is always **zero** for an ideal capacitor. Determine the reactive power by first finding the value for the capacitive reactance and then using Equation 9–23.

$$X_C = \frac{1}{2\pi f C} = \frac{1}{2\pi(2 \times 10^3\ \text{Hz})(0.01 \times 10^{-6}\ \text{F})} = 7.96\ \text{k}\Omega$$

$$P_r = \frac{V_{\text{rms}}^2}{X_C} = \frac{(2\ \text{V})^2}{7.96\ \text{k}\Omega} = 503 \times 10^{-6}\ \text{VAR} = \mathbf{503\ \mu VAR}$$

RELATED PROBLEM

If the frequency is doubled in Figure 9–51, what are the true power and the reactive power?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 9–6 CHECKUP

1. Calculate X_C for $f = 5$ kHz and $C = 47$ pF.
2. At what frequency is the reactance of a 0.1 μF capacitor equal to 2 kΩ?
3. Calculate the rms current in Figure 9–52.
4. State the phase relationship between current and voltage in a capacitor.
5. A 1 μF capacitor is connected to an ac voltage source of 12 V rms. What is the true power?
6. In Question 5, determine the reactive power at a frequency of 500 Hz.

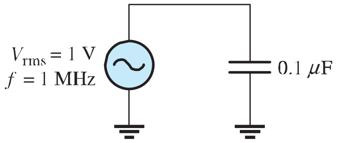

FIGURE 9–52

9–7 CAPACITOR APPLICATIONS

Capacitors are widely used in electrical and electronic systems.

After completing this section, you should be able to

- **Discuss some capacitor applications**
  - **Describe a power supply filter**
  - **Explain the purpose of coupling and bypass capacitors**
  - **Discuss the basics of capacitors applied to tuned circuits, timing circuits, and computer memories**

If you pick up any circuit board, open any power supply, or look inside any piece of electronic equipment, chances are you will find capacitors of one type or another. These components are used for a variety of purposes in both dc and ac systems.

Electrical Storage

One of the most basic applications of a capacitor is as a backup voltage source for low-power circuits such as certain types of semiconductor memories in computers. This particular application requires a very high capacitance value and negligible leakage.

A storage capacitor is connected between the dc power supply input to the circuit and ground. When the circuit is operating from its normal power supply, the capacitor remains fully charged to the dc power supply voltage. If the normal power source is disrupted, effectively removing the power supply from the circuit, the storage capacitor temporarily becomes the power source for the circuit.

A capacitor provides voltage and current to a circuit as long as its charge remains sufficient. As current is drawn by the circuit, charge is removed from the capacitor and the voltage decreases. For this reason, a storage capacitor can only be used as a temporary power source. The length of time that a capacitor can provide sufficient power to a circuit depends on the capacitance and the amount of current drawn by the circuit. The smaller the current and the higher the capacitance, the longer the time a capacitor can provide power to a circuit.

Power Supply Filtering

A basic dc power supply consists of a circuit known as a **rectifier** followed by a filter. The rectifier converts the 120 V, 60 Hz sinusoidal voltage available at a standard outlet to a

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

pulsating dc voltage that can be either a half-wave rectified voltage or a full-wave rectified voltage, depending on the type of rectifier circuit. As shown in Figure 9–53(a), a half-wave rectifier removes each negative half-cycle of the sinusoidal voltage. As shown in Figure 9–53(b), a full-wave rectifier actually reverses the polarity of the negative portion of each cycle. Both half-wave and full-wave rectified voltages are dc because, even though they are changing, they do not alternate polarity.

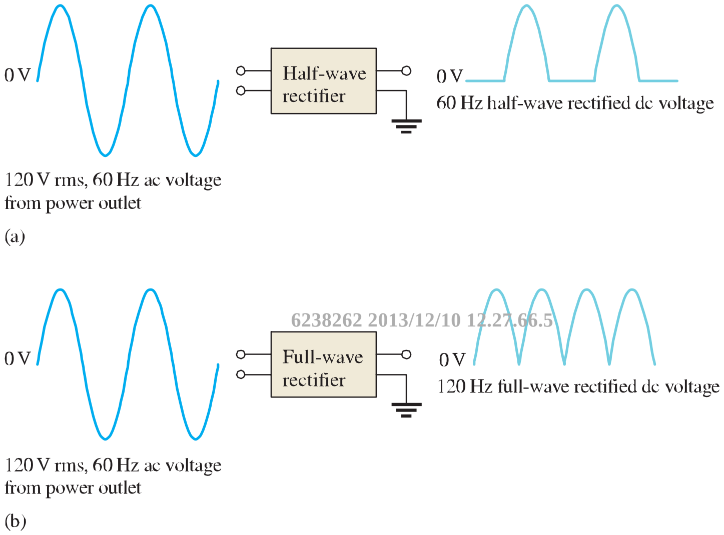
0 V
Half-wave
rectifier
0 V
60 Hz half-wave rectified dc voltage
120 V rms, 60 Hz ac voltage
from power outlet
(a)
0 V
Full-wave
rectifier
0 V
120 Hz full-wave rectified dc voltage
120 V rms, 60 Hz ac voltage
from power outlet
(b)

6238262 2013/12/10 12.27.66.5

FIGURE 9–53 **Half-wave and full-wave rectifier operation.**

To be useful for powering electronic circuits, the rectified voltage must be changed to constant dc voltage because all circuits require constant power. When connected to the rectifier output, the **filter** nearly eliminates the fluctuations in the rectified voltage and provides a smooth constant-value dc voltage to the load, which is the electronic circuit, as indicated in Figure 9–54.

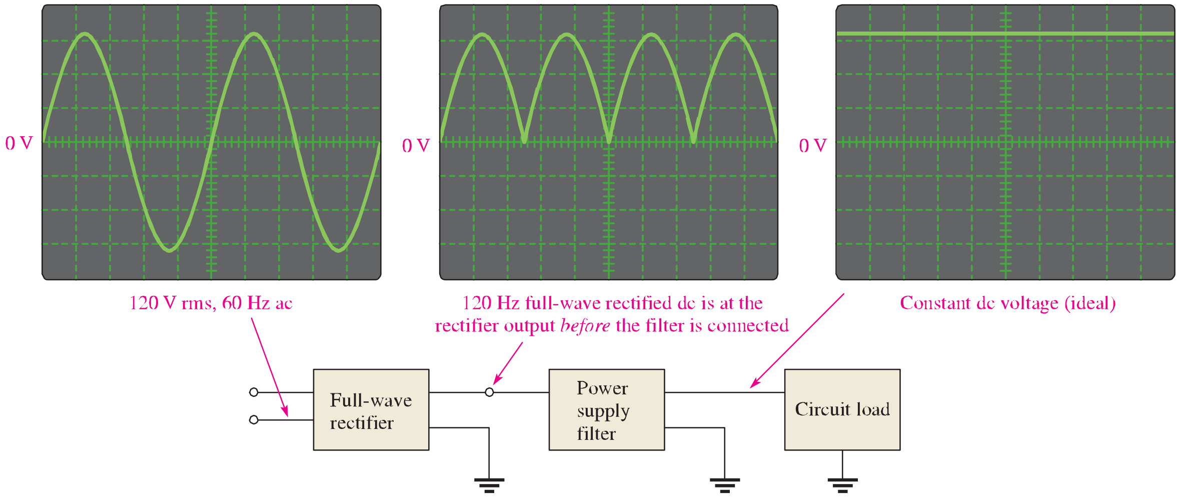
0 V
0 V
0 V
120 V rms, 60 Hz ac
120 Hz full-wave rectified dc is at the
rectifier output *before* the filter is connected
Constant dc voltage (ideal)
Full-wave
rectifier
Power
supply
filter
Circuit load

FIGURE 9–54 **Basic block diagram and operation of a dc power supply.**

THE CAPACITOR AS A POWER SUPPLY FILTER Capacitors are used as filters in dc power supplies because of their ability to store electrical charge. Figure 9–55(a) shows a dc power supply with a full-wave rectifier and a capacitor filter. The operation can be described from a charging and discharging point of view as follows. Assume the capacitor is initially

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

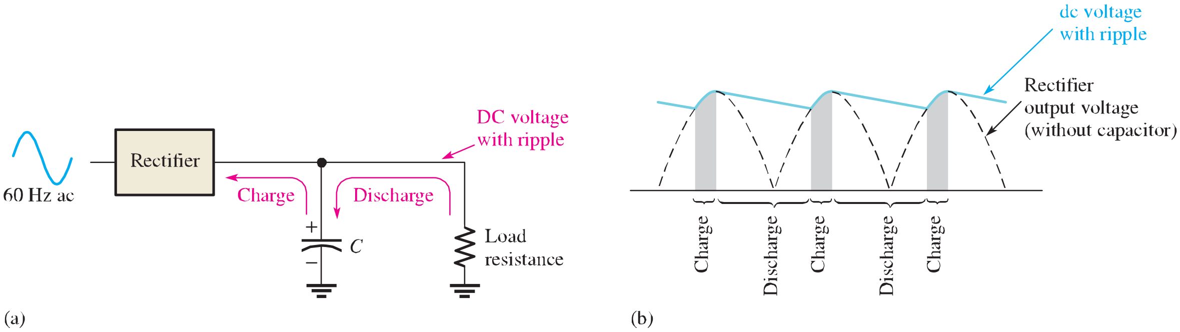

FIGURE 9–55 **Basic operation of a power supply filter capacitor.**

uncharged. When the power supply is first turned on and the first cycle of the rectified voltage occurs, the capacitor will quickly charge through the low resistance of the rectifier. The capacitor voltage will follow the rectified voltage curve up to the peak of the rectified voltage. As the rectified voltage passes the peak and begins to decrease, the capacitor will begin to discharge very slowly through the high resistance of the load circuit, as indicated in Figure 9–55(b). The amount of discharge is typically very small and is exaggerated in the figure for purposes of illustration. The next cycle of the rectified voltage will recharge the capacitor back to the peak value by replenishing the small amount of charge lost since the previous peak. This pattern of a small amount of charging and discharging continues as long as the power is on.

A rectifier is designed so that it allows current only in the direction to charge the capacitor. The capacitor will not discharge back through the rectifier but will only discharge a small amount through the relatively high resistance of the load. The small fluctuation in voltage due to the charging and discharging of the capacitor is called the **ripple voltage**. A good dc power supply has a very small amount of ripple on its dc output. The discharge time constant of a power supply filter capacitor depends on its capacitance and the resistance of the load; consequently, the higher the capacitance value, the longer the discharge time and, therefore, the smaller the ripple voltage.

DC Blocking and AC Coupling

Capacitors are commonly used to block the constant dc voltage in one part of a circuit from getting to another part. As an example of this, a capacitor is connected between two stages of an amplifier to prevent the dc voltage at the output of stage 1 from affecting the dc voltage at the input of stage 2, as illustrated in Figure 9–56. Assume that, for proper

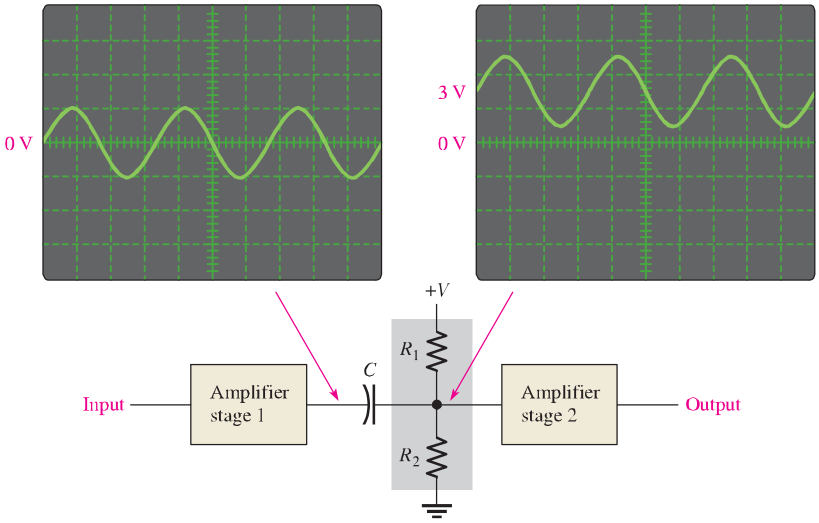

FIGURE 9–56 **An application of a capacitor used to block dc and couple ac in an amplifier.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

operation, the output of stage 1 has a zero dc voltage and the input to stage 2 has a 3 V dc voltage. The capacitor prevents the 3 V dc at stage 2 from getting to the stage 1 output and affecting its zero value, and vice versa.

If a sinusoidal signal voltage is applied to the input to stage 1, the signal voltage is increased (amplified) and appears on the output of stage 1, as shown in Figure 9–56. The amplified signal voltage is then coupled through the capacitor to the input of stage 2 where it is superimposed on the 3 V dc level and then again amplified by stage 2. In order for the signal voltage to be passed through the capacitor without being reduced, the capacitor must be large enough so that its reactance at the frequency of the signal voltage is negligible. In this type of application, the capacitor is known as a **coupling** capacitor, which ideally appears as an open to dc and as a short to ac. As the signal frequency is reduced, the capacitive reactance increases and, at some point, the capacitive reactance becomes large enough to cause a significant reduction in ac voltage between stage 1 and stage 2.

SYSTEM EXAMPLE 9–2

TRANSISTOR AMPLIFIER

Amplifiers form the backbone of many electronic systems. All transistor amplifiers require dc voltages to establish proper operating conditions for amplifying ac signals. These dc voltages are referred to as bias voltages and are superimposed on the ac signals. Figure 9–57 shows a discrete transistor amplifier with two coupling capacitors to connect the ac signal to and from the amplifier without affecting the bias voltages. The input bias voltage is set up by the voltage divider formed by R_1 and R_2, which divide the 24 V dc supply voltage.

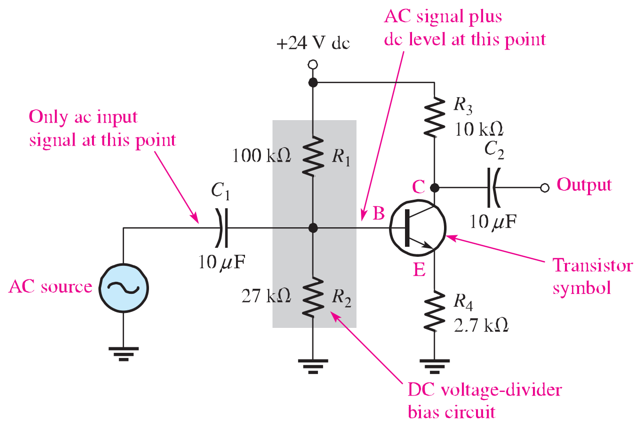

FIGURE 9–57 **A discrete transistor amplifier.**

When an ac signal voltage is applied to the amplifier, the input coupling capacitor, C_1, prevents ac source from affecting the dc bias voltage. Without the capacitor, the internal source resistance would appear in parallel with R_2 and drastically change the value of the dc voltage. The coupling capacitance is chosen so that its reactance, X_C, at the frequency of the ac signal is very small compared to the bias resistor values. The coupling capacitance, therefore, efficiently couples the ac signal from the source to the input of the amplifier. On the source side of the input coupling capacitor there is only ac, but on the amplifier side there is both the ac signal plus the dc bias voltage. Capacitor C_2 is the output coupling capacitor, which couples the amplified ac signal to another amplifier stage that would be connected to the output. It isolates the dc voltage at the transistor's output from the ac signal.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Power Line Decoupling

Capacitors connected from the dc supply voltage line to ground are used on circuit boards to decouple unwanted voltage transients or spikes that occur on the dc supply voltage because of fast switching digital circuits. A voltage **transient** contains higher frequencies that may affect the operation of the circuits. These transients are shorted to ground through the very low reactance of the decoupling capacitors. **Decoupling** capacitors are often used at various points along the supply voltage line on a circuit board, particularly near integrated circuits (ICs).

Bypassing

Bypass capacitors are used to bypass an ac voltage around a resistor in a circuit without affecting the dc voltage across the resistor. In amplifier circuits, for example, dc voltages called *bias voltages* are required at various points. For the amplifier to operate properly, certain bias voltages must remain constant and, therefore, any ac voltages must be removed. A sufficiently large capacitor connected from a bias point to ground provides a low reactance path to ground for ac voltages, leaving the constant dc bias voltage at the given point. This bypass application is illustrated in Figure 9–58. As frequencies decrease, the bypass capacitor becomes less effective because of its increased reactance.

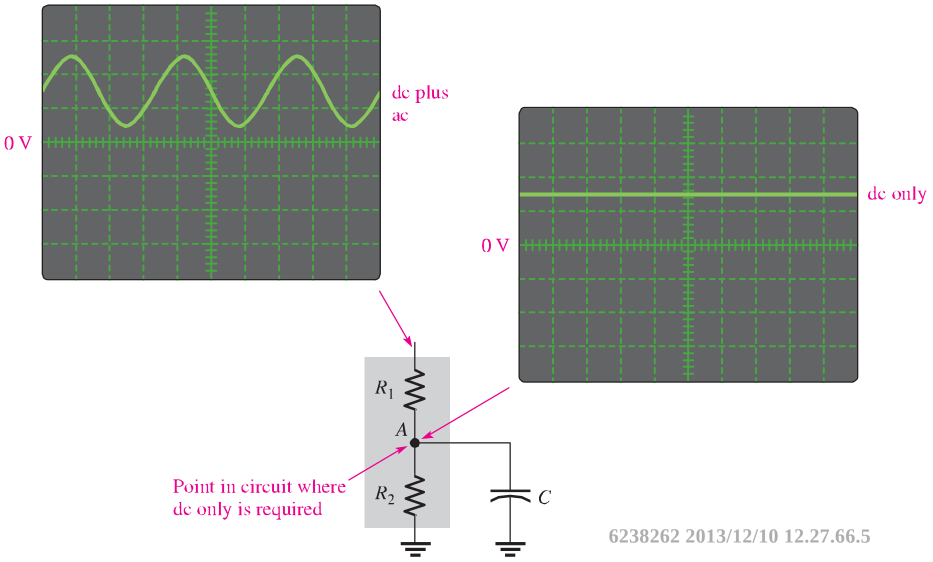

6238262 2013/12/10 12.27.66.5

FIGURE 9–58 **Example of the operation of a bypass capacitor.**

Signal Filters

Filters are used for selecting one ac signal with a certain specified frequency from a wide range of signals with many different frequencies or for selecting a certain band of frequencies and eliminating all others. A common example of this application is in radio and television receivers where it is necessary to select the signal transmitted from a given station and eliminate or filter out the signals transmitted from all the other stations in the area.

When you tune your radio or TV, you are actually changing the capacitance in the tuner circuit (which is a type of filter) so that only the signal from the station or channel you want passes through to the receiver circuitry. Capacitors are used in conjunction with resistors, inductors (covered in Chapter 11), and other components in these types of filters.

The main characteristic of a filter is its frequency selectivity, which is based on the fact that the reactance of a capacitor depends on frequency ($X_C = 1/2\pi fC$).

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Timing Circuits

Another important area in which capacitors are used is in timing circuits that generate specified time delays or produce waveforms with specific characteristics. Recall that the time constant of a circuit with resistance and capacitance can be controlled by selecting appropriate values for R and C. The charging time of a capacitor can be used as a basic time delay in various types of circuits. An example is the circuit that controls the turn indicators on your car where the light flashes on and off at regular intervals.

Computer Memories

Dynamic computer memories use capacitors as the basic storage element for binary information, which consists of two digits, 1 and 0. A charged capacitor can represent a stored 1 and a discharged capacitor can represent a stored 0. Patterns of 1s and 0s that make up binary data are stored in a memory that consists of an array of capacitors with associated circuitry. You will study this topic in a computer or digital fundamentals course.

SECTION 9–7 CHECKUP

1. Explain how half-wave or full-wave rectified dc voltages are smoothed out by a filter capacitor.
2. Explain the purpose of a coupling capacitor.
3. How large must a coupling capacitor be?
4. Explain the purpose of a decoupling capacitor.
5. Discuss how the relationship of frequency and capacitive reactance is important in frequency-selective circuits such as signal filters.

SUMMARY

- A capacitor is an electrical device that has the ability to store charge. It consists of one or more conductors separated by an insulating material called the *dielectric.*
- A capacitor stores electrical charge on its plates.
- Energy is stored in a capacitor by the electric field created between the charged plates in the dielectric.
- Capacitance is measured in units of farads (F).
- Capacitance is directly proportional to the plate area and the dielectric constant and inversely proportional to the distance between the plates (the dielectric thickness).
- The dielectric constant is an indication of the ability of a material to establish an electric field.
- The dielectric strength is one factor that determines the breakdown voltage of a capacitor.
- Capacitors are commonly classified according to the dielectric material. Typical materials are mica, ceramic, plastic-film, and electrolytic (aluminum oxide and tantalum oxide).
- The total capacitance of series capacitors is less than the smallest capacitance.
- The total capacitance of parallel capacitors is the sum of all the capacitances.
- A capacitor blocks constant dc.
- The time constant determines the charging and discharging time of a capacitor with resistance in series.
- In an RC circuit, the voltage and current during charging and discharging make an approximate 63% change during each time-constant interval.
- Five time constants are required for a capacitor to fully charge or discharge. This is called the *transient time.*
- The approximate percentage of final charge after each charging time-constant interval is given in Table 9–4.
- The approximate percentage of initial charge after each discharging time-constant interval is given in Table 9–5.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

TABLE 9–4

NUMBER OF TIME CONSTANTS	APPROXIMATE % OF FINAL CHARGE
1	63
2	86
3	95
4	98
5	99 (considered 100%)

TABLE 9–5

NUMBER OF TIME CONSTANTS	APPROXIMATE % OF INITIAL CHARGE
1	37
2	14
3	5
4	2
5	1 (considered 0)

- Alternating current in a capacitor leads the voltage by 90°.
- A capacitor passes ac to an extent that depends on its reactance and the resistance in the rest of the circuit.
- Capacitive reactance is the opposition to ac, expressed in ohms.
- Capacitive reactance (X_C) is inversely proportional to the frequency and to the capacitance value.
- The total capacitive reactance of series capacitors is the sum of the individual reactances.
- The total capacitive reactance of parallel capacitors is the reciprocal of the sum of the reciprocals of the individual reactances.
- Ideally, there is no energy loss in a capacitor and, thus, the true power (watts) is zero. However, in most capacitors there is some small energy loss due to leakage resistance.

KEY TERMS

Key terms and other bold terms in the chapter are defined in the end-of-book glossary.

Bypass A capacitor connected from a point to ground to remove the ac signal without affecting the dc voltage. A special case of decoupling.

Capacitance The ability of a capacitor to store electrical charge.

Capacitive reactance The opposition of a capacitor to sinusoidal current. The unit is the ohm.

Capacitor An electrical device that has the ability to store charge. It consists of one or more conductors separated by an insulating material called the *dielectric*.

Charging The process in which a current removes charge from one plate of a capacitor and deposits it on the other plate, making one plate more positive than the other.

Coupling The method of connecting a capacitor from between two points in a circuit to allow ac to pass from one point to the other while blocking dc.

Decoupling The method of connecting a capacitor from one point, usually the dc power supply line, to ground to short ac to ground without affecting the dc voltage.

Dielectric The insulating material between the plates of a capacitor.

Dielectric constant A measure of the ability of a dielectric material to establish an electric field.

Dielectric strength A measure of the ability of a dielectric material to withstand voltage without breaking down.

Exponential A mathematical function described by a natural logarithm (base). The charging and discharging of a capacitor are described by an exponential function.

Farad (F) The unit of capacitance.

Filter A type of circuit that passes certain frequencies and rejects all others.

Instantaneous power (p) The value of power in a circuit at any given instant of time.

RC time constant A fixed time interval set by R and C values that determines the time response of a series RC circuit. It equals the product of the resistance and the capacitance.

Reactive power (P_r) The rate at which energy is alternately stored and returned to the source by a capacitor. The unit is the VAR.

Ripple voltage The small fluctuation in voltage due to the charging and discharging of a capacitor.

Temperature coefficient A constant specifying the amount of change in the value of a quantity for a given change in temperature.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Transient time An interval equal to approximately five time constants.

True power (P_{true}) The power that is dissipated in a circuit, usually in the form of heat.

VAR (volt-ampere reactive) The unit of reactive power.

KEY FORMULAS

(9–1)	$C = \dfrac{Q}{V}$	Capacitance in terms of charge and voltage
(9–2)	$Q = CV$	Charge in terms of capacitance and voltage
(9–3)	$V = \dfrac{Q}{C}$	Voltage in terms of charge and capacitance
(9–4)	$W = \dfrac{1}{2}CV^2$	Energy stored by a capacitor
(9–5)	$\varepsilon_r = \dfrac{\varepsilon}{\varepsilon_0}$	Dielectric constant (relative permittivity)
(9–6)	$C = \dfrac{A\varepsilon_r(8.85 \times 10^{-12}\ \text{F/m})}{d}$	Capacitance in terms of physical parameters
(9–7)	$C_\text{T} = \dfrac{C_1C_2}{C_1 + C_2}$	Total series capacitance (two capacitors)
(9–8)	$C_\text{T} = \dfrac{1}{\dfrac{1}{C_1} + \dfrac{1}{C_2} + \dfrac{1}{C_3} + \cdots + \dfrac{1}{C_n}}$	Total series capacitance (general)
(9–9)	$V_x = \left(\dfrac{C_\text{T}}{C_x}\right)V_\text{S}$	Voltage across series capacitor
(9–10)	$C_\text{T} = C_1 + C_2$	Two capacitors in parallel
(9–11)	$C_\text{T} = C_1 + C_2 + C_3 + \cdots + C_n$	*n* capacitors in parallel
(9–12)	$\tau = RC$	*RC* time constant
(9–13)	$v = V_F + (V_i - V_F)e^{-t/\tau}$	Exponential voltage (general)
(9–14)	$i = I_F + (I_i - I_F)e^{-t/\tau}$	Exponential current (general)
(9–15)	$v = V_F(1 - e^{-t/RC})$	Increasing exponential voltage beginning at zero
(9–16)	$v = V_ie^{-t/RC}$	Decreasing exponential voltage ending at zero
(9–17)	$X_C = \dfrac{1}{2\pi fC}$	Capacitive reactance
(9–18)	$X_{C(tot)} = X_{C1} + X_{C2} + X_{C3} + \cdots + X_{Cn}$	Capacitive reactance for series capacitors
(9–19)	$X_{C(tot)} = \dfrac{1}{\dfrac{1}{X_{C1}} + \dfrac{1}{X_{C2}} + \dfrac{1}{X_{C3}} + \cdots + \dfrac{1}{X_{Cn}}}$	Capacitive reactance for parallel capacitors
(9–20)	$I = \dfrac{V}{X_C}$	Ohm's law for a capacitor
(9–21)	$V_x = \left(\dfrac{X_{Cx}}{X_{C(tot)}}\right)V_s$	Capacitive voltage divider
(9–22)	$P_r = V_\text{rms}I_\text{rms}$	Reactive power in a capacitor

6238262 2013/12/10 12.27.66.5

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

(9–23)	$P_r = \dfrac{V_{\text{rms}}^2}{X_C}$	Reactive power in a capacitor
(9–24)	$P_r = I_{\text{rms}}^2 X_C$	Reactive power in a capacitor

TRUE/FALSE QUIZ

Answers are at the end of the chapter.

1. The area of the plates of a capacitor is proportional to the capacitance.

2. A capacitance of 1200 pF is the same as 1.2 μF.

3. When two capacitors are in series with a voltage source, the smaller capacitor will have the larger voltage.

4. When two capacitors are in parallel with a voltage source, the smaller capacitor will have the larger voltage.

5. A capacitor appears as an open to a constant dc.

6. When a capacitor is charging or discharging between two levels, the charge on the capacitor changes by 63% of the difference in one time constant.

7. Capacitive reactance is proportional to the applied frequency.

8. The total reactance of series capacitors is the product-over-sum of the individual reactances.

9. Voltage leads current in a capacitor.

10. The unit of reactive power is the VAR.

SELF-TEST

Answers are at the end of the chapter.

1. Which of the following accurately describes a capacitor?
(a) The plates are conductive.
(b) The dielectric is an insulator between the plates.
(c) There is constant direct current (dc) through a fully charged capacitor.
(d) A capacitor stores charge indefinitely when disconnected from the source.
(e) none of the above answers
(f) all the above answers
(g) only answers (a) and (b)

2. Which one of the following statements is true?
(a) There is current through the dielectric of a charging capacitor.
(b) When a capacitor is connected to a dc voltage source, it will charge to the value of the source.
(c) A capacitor can be discharged by disconnecting it from the voltage source.

3. A capacitance of 0.01 μF is larger than
(a) 0.00001 F **(b)** 100,000 pF **(c)** 1000 pF **(d)** all of these answers

4. A capacitance of 1000 pF is smaller than
(a) 0.01 μF **(b)** 0.001 μF **(c)** 0.00000001 F **(d)** answers (a) and (c)

5. When the voltage across a capacitor is increased, the stored charge
(a) increases **(b)** decreases **(c)** remains constant **(d)** fluctuates

6. When the voltage across a capacitor is doubled, the stored charge
(a) stays the same **(b)** is halved **(c)** increases by four **(d)** doubles

7. The voltage rating of a capacitor is increased by
(a) increasing the plate separation **(b)** decreasing the plate separation
(c) increasing the plate area **(d)** answers (b) and (c)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

8. The capacitance value is increased by

(a) decreasing plate area **(b)** increasing plate separation
(c) decreasing plate separation **(d)** increasing plate area
(e) answers (a) and (b) **(f)** answers (c) and (d)

9. A 1 μF, a 2.2 μF, and a 0.047 μF capacitor are connected in series. The total capacitance is less than

(a) 1 μF **(b)** 2.2 μF **(c)** 0.047 μF **(d)** 0.001 μF

10. Four 0.022 μF capacitors are in parallel. The total capacitance is

(a) 0.022 μF **(b)** 0.088 μF **(c)** 0.011 μF **(d)** 0.044 μF

11. An uncharged capacitor and a resistor are connected in series with a switch and a 12 V battery. At the instant the switch is closed, the voltage across the capacitor is

(a) 12 V **(b)** 6 V **(c)** 24 V **(d)** 0 V

12. In Question 11, the voltage across the capacitor when it is fully charged is

(a) 12 V **(b)** 6 V **(c)** 24 V **(d)** −6 V

13. In Question 11, the capacitor will reach full charge in a time equal to approximately

(a) RC **(b)** $5RC$ **(c)** $12RC$ **(d)** cannot be predicted

14. A sinusoidal voltage is applied across a capacitor. When the frequency of the voltage is increased, the current

(a) increases **(b)** decreases **(c)** remains constant **(d)** ceases

15. A capacitor and a resistor are connected in series to a sine wave generator. The frequency is set so that the capacitive reactance is equal to the resistance and, thus, an equal amount of voltage appears across each component. If the frequency is decreased,

(a) $V_R > V_C$ **(b)** $V_C > V_R$ **(c)** $V_R = V_C$ **(d)** $V_C < V_R$

TROUBLESHOOTING: SYMPTOM AND CAUSE

The purpose of these exercises is to help develop thought processes essential to troubleshooting. Answers are at the end of the chapter.

Determine the cause for each set of symptoms. Refer to Figure 9–59.

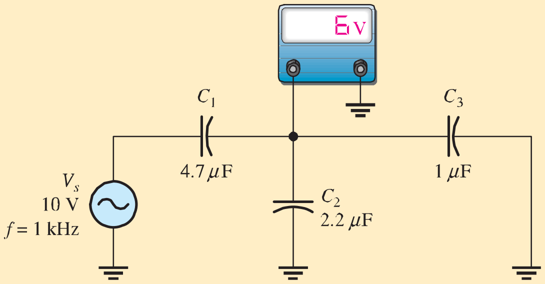

FIGURE 9–59 **The ac meter indicates the correct reading for this circuit.**

1. *Symptom:* The voltmeter reading is 0 V.

Cause:

(a) C_1 is shorted. **(b)** C_2 is shorted. **(c)** C_3 is open.

2. *Symptom:* The voltmeter reading is 10 V.

Cause:

(a) C_1 is shorted. **(b)** C_2 is open. **(c)** C_3 is open.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

3. *Symptom:* The voltmeter reading is 6.86 V.
Cause:
(a) C_1 is open. **(b)** C_2 is open. **(c)** C_3 is open.

4. *Symptom:* The voltmeter reading is 0 V.
Cause:
(a) C_1 is open. **(b)** C_2 is open. **(c)** C_3 is open.

5. *Symptom:* The voltmeter reading is 8.28 V.
Cause:
(a) C_1 is shorted. **(b)** C_2 is open. **(c)** C_3 is open.

PROBLEMS

Answers to odd-numbered problems are at the end of the book.

BASIC PROBLEMS

SECTION 9–1 The Basic Capacitor

1. **(a)** Find the capacitance when $Q = 50\ \mu C$ and $V = 10$ V.
(b) Find the charge when $C = 0.001\ \mu F$ and $V = 1$ kV.
(c) Find the voltage when $Q = 2$ mC and $C = 200\ \mu F$.

2. Convert the following values from microfarads to picofarads:
(a) 0.1 μF **(b)** 0.0025 μF **(c)** 5 μF

3. Convert the following values from picofarads to microfarads:
(a) 1000 pF **(b)** 3500 pF **(c)** 250 pF

4. Convert the following values from farads to microfarads:
(a) 0.0000001 F **(b)** 0.0022 F **(c)** 0.0000000015 F

5. What size capacitor is capable of storing 10 mJ of energy with 100 V across its plates?

6. A mica capacitor has a plate area of 20 cm^2 and a dielectric thickness of 2.5 mils. What is its capacitance?

7. An air capacitor has plates with an area of 0.1 m^2. The plates are separated by 0.01 m. Calculate the capacitance.

8. A student wants to construct a 1 F capacitor out of two square plates for a science fair project. He plans to use a paper dielectric ($\varepsilon_r = 2.5$) that is 8×10^{-5} m thick. The science fair is to be held in the Astrodome. Will his capacitor fit in the Astrodome? What would be the size of the plates if it could be constructed?

9. A student decides to construct a capacitor using two conducting plates 30 cm on a side. He separates the plates with a paper dielectric ($\varepsilon_r = 2.5$) that is 8×10^{-5} m thick. What is the capacitance of his capacitor?

10. At ambient temperature (25°C), a certain capacitor is specified to be 1000 pF. It has a negative temperature coefficient of 200 ppm/°C. What is its capacitance at 75°C?

11. A 0.001 μF capacitor has a positive temperature coefficient of 500 ppm/°C. How much change in capacitance will a 25°C increase in temperature cause?

6238262 2013/12/10 12.27.66.5

SECTION 9–2 Types of Capacitors

12. In the construction of a stacked-foil mica capacitor, how is the plate area increased?

13. What type of capacitor has the higher dielectric constant, mica or ceramic?

14. Show how to connect an electrolytic capacitor across points *A* and *B* in Figure 9–60.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

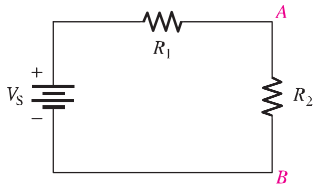

FIGURE 9–60

15. Determine the value of the typographically labeled ceramic disk capacitors in Figure 9–61.

16. Name two types of electrolytic capacitors. How do electrolytics differ from other capacitors?

17. Identify the parts of the ceramic disk capacitor shown in the cutaway view of Figure 9–62 by referring to Figure 9–8(b).

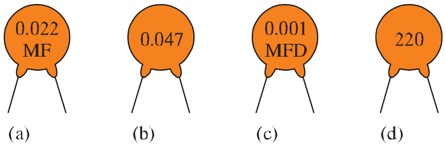
0.022
MF
0.047
0.001
MFD
220
(a)
(b)
(c)
(d)

FIGURE 9–61

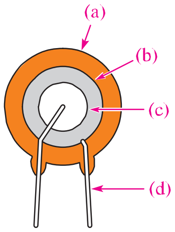

FIGURE 9–62

SECTION 9–3 Series Capacitors

18. Five 1000 pF capacitors are in series. What is the total capacitance?

19. Find the total capacitance for each circuit in Figure 9–63.

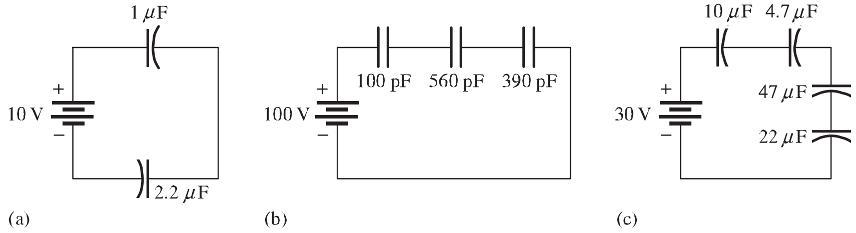
1 μF
+
10 V
−
2.2 μF
(a)
+
100 V
−
100 pF
560 pF
390 pF
(b)
10 μF
4.7 μF
+
30 V
−
47 μF
22 μF
(c)

FIGURE 9–63

20. For each circuit in Figure 9–63, determine the voltage across each capacitor.

21. The total charge stored by the series capacitors in Figure 9–64 is 10 μC. Determine the voltage across each of the capacitors.

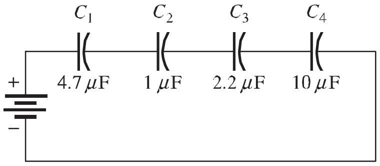
C_1
C_2
C_3
C_4
4.7 μF
1 μF
2.2 μF
10 μF
+
−

FIGURE 9–64

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 9–4 Parallel Capacitors

22. Determine C_T for each circuit in Figure 9–65.

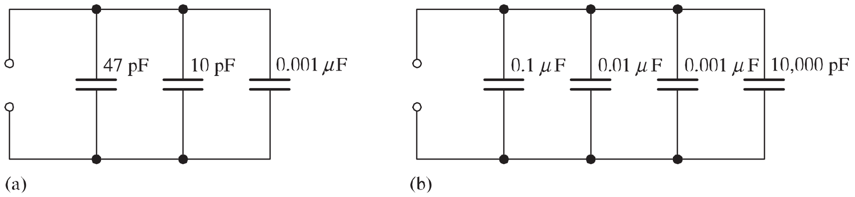

(a) (b)

FIGURE 9–65

23. Determine the total capacitance and total charge on the capacitors in Figure 9–66.

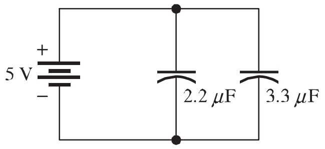

FIGURE 9–66

24. Assume you need a total capacitance of 2.1 μF in a certain timing application, but all that is available are 0.22 μF and 0.47 μF capacitors (in large quantities). How would you get the total capacitance that you need?

SECTION 9–5 Capacitors in DC Circuits

25. Determine the time constant for each of the following series *RC* combinations:
(a) $R = 100\ \Omega, C = 1\ \mu\text{F}$ **(b)** $R = 10\ \text{M}\Omega, C = 56\ \text{pF}$
(c) $R = 4.7\ \text{k}\Omega, C = 0.0047\ \mu\text{F}$ **(d)** $R = 1.5\ \text{M}\Omega, C = 0.01\ \mu\text{F}$

26. Determine how long it takes the capacitor to reach full charge for each of the following combinations:
(a) $R = 47\ \Omega, C = 47\ \mu\text{F}$ **(b)** $R = 3300\ \Omega, C = 0.015\ \mu\text{F}$
(c) $R = 22\ \text{k}\Omega, C = 100\ \text{pF}$ **(d** $R = 4.7\ \text{M}\Omega, C = 10\ \text{pF}$

27. In the circuit of Figure 9–67, the capacitor initially is uncharged. Determine the capacitor voltage at the following times after the switch is closed:
(a) 10 μs **(b)** 20 μs **(c)** 30 μs **(d)** 40 μs **(e)** 50 μs

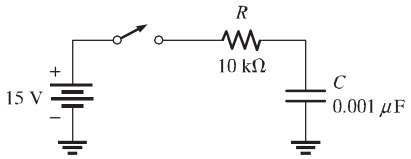

FIGURE 9–67

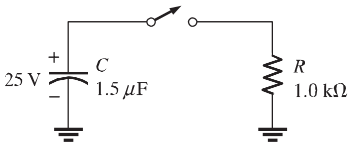

FIGURE 9–68

28. In Figure 9–68, the capacitor is charged to 25 V. Find the capacitor voltage at the following times when the switch is closed:
(a) 1.5 ms **(b)** 4.5 ms **(c)** 6 ms **(d)** 7.5 ms

29. Repeat Problem 27 for the following time intervals:
(a) 2 μs **(b)** 5 μs **(c)** 15 μs

30. Repeat Problem 28 for the following time intervals:
(a) 0.5 ms **(b)** 1 ms **(c)** 2 ms

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 9–6 Capacitors in AC Circuits

31. Determine X_C for a 0.047 μF capacitor at each of the following frequencies:
(a) 10 Hz **(b)** 250 Hz **(c)** 5 kHz **(d)** 100 kHz

32. What is the value of the total capacitive reactance in each circuit in Figure 9–69?

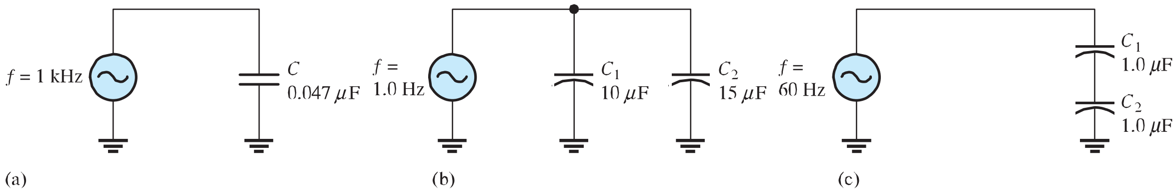

FIGURE 9–69

33. For the circuit in Figure 9–70, find the reactance of each capacitor, the total reactance, and the voltage across each capacitor.

FIGURE 9–70

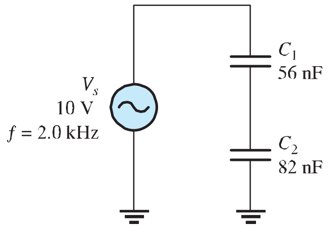

34. In each circuit of Figure 9–69, what frequency is required to produce an $X_{C(tot)}$ of 100 Ω? An $X_{C(tot)}$ of 1 kΩ?

35. A sinusoidal voltage of 20 V rms produces an rms current of 100 mA when connected to a certain capacitor. What is the reactance?

36. A 10 kHz voltage is applied to a 0.0047 μF capacitor, and 1 mA of rms current is measured. What is the rms value of the voltage?

37. Determine the true power and the reactive power in Problem 36.

SECTION 9–7 Capacitor Applications

38. If another capacitor is connected in parallel with the existing capacitor in the power supply filter of Figure 9–52, how is the ripple voltage affected?

39. Ideally, what should the reactance of a bypass capacitor be in order to eliminate a 10 kHz ac voltage at a given point in an amplifier circuit?

6238262 2013/12/10 12.27.66.5

ADVANCED PROBLEMS

40. Two series capacitors (one 1 μF, the other of unknown value) are charged from a 12 V source. The 1 μF capacitor is charged to 8 V, and the other to 4 V. What is the value of the unknown capacitor?

41. How long does it take C to discharge to 3 V in Figure 9–68?

42. How long does it take C to charge to 8 V in Figure 9–67?

43. Determine the time constant for the circuit in Figure 9–71.

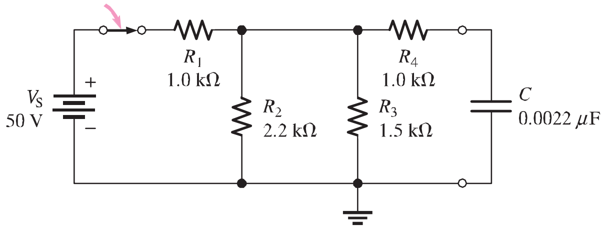

FIGURE 9–71

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

44. In Figure 9–72, the capacitor initially is uncharged. At $t = 10\ \mu\text{s}$ after the switch is closed, the instantaneous capacitor voltage is 7.2 V. Determine the value of R.

45. **(a)** The capacitor in Figure 9–73 is uncharged when the switch is thrown into position 1. The switch remains in position 1 for 10 ms and then is thrown into position 2, where it remains indefinitely. Draw the complete waveform for the capacitor voltage.
(b) If the switch is thrown back to position 1 after 5 ms in position 2, and then is left in position 1, how will the waveform appear?

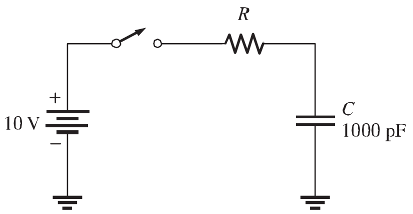

FIGURE 9–72

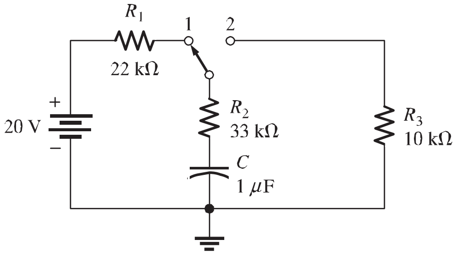

FIGURE 9–73

46. For the Colpitts oscillator in Figure 9–23, assume C_1 is changed to 3.3 nF.
(a) What is the new total capacitance?
(b) What is the new feedback fraction?

47. For the Colpitts oscillator in Figure 9–23, assume you need to change C_1 to obtain a feedback fraction of 15%. What value would you select?

48. Assume the discrete amplifier in Figure 9–57 has a 5 kHz, 1 V rms input. Sketch the complete signal as it would appear at the point in the circuit labeled B (the base of the transistor). Assume there is no loading effect of the transistor on the input voltage divider.

49. For the discrete amplifier in Figure 9–57, at what frequency will the reactance of $C_1 = 1.0\ \text{k}\Omega$?

MULTISIM TROUBLESHOOTING PROBLEMS

MULTISIM

50. Open file P09-50; files are found at www.pearsonhighered.com/floyd. Determine if there is a fault. If so, identify it.

51. Open file P09-51 and determine if there is a fault. If so, identify it.

52. Open file P09-52 and determine if there is a fault. If so, identify it.

53. Open file P09-53 and determine if there is a fault. If so, identify it.

54. Open file P09-54 and determine if there is a fault. If so, identify it.

ANSWERS TO SECTION CHECKUPS

SECTION 9–1 The Basic Capacitor

1. Capacitance is the ability (capacity) to store electrical charge.

2. **(a)** There are 1,000,000 microfarads in one farad.
(b) There are 1×10^{12} picofarads in one farad.
(c) There are 1,000,000 picofarads in one microfarad.

3. $0.0015\ \mu\text{F} \times 10^{6}\ \text{pF}/\mu\text{F} = 1500\ \text{pF}$; $0.0015\ \mu\text{F} \times 10^{-6}\ \text{F}/\mu\text{F} = 0.0000000015\ \text{F}$

4. $W = \frac{1}{2}CV^2 = \frac{1}{2}(0.01\ \mu\text{F})(15\ \text{V})^2 = 1.125\ \mu\text{J}$

5. **(a)** Capacitance increases when plate area is increased.
(b) Capacitance decreases when plate separation is increased.

6. $(1000\ \text{V/mil})(2\ \text{mils}) = 2\ \text{kV}$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 9–2 Types of Capacitors

1. Capacitors are commonly classified by the dielectric material.
2. A fixed capacitance cannot be changed; a variable capacitor can.
3. Electrolytic capacitors are polarized.
4. Be sure that the voltage rating is sufficient and connect the positive end to the positive side of the circuit when installing a polarized capacitor.
5. The positive lead should be connected to ground.

SECTION 9–3 Series Capacitors

1. C_T of series capacitors is less than the smallest value.
2. $C_T = 61.2\text{ pF}$
3. $C_T = 0.006\ \mu\text{F}$
4. $V = (0.006\ \mu\text{F}/0.01\ \mu\text{F})10\text{ V} = 6\text{ V}$

SECTION 9–4 Parallel Capacitors

1. The individual parallel capacitors are added to get C_T.
2. Use five 0.01 μF capacitors in parallel to get 0.05 μF.
3. $C_T = 167\text{ pF}$

SECTION 9–5 Capacitors in DC Circuits

1. $\tau = RC = 1.2\ \mu\text{s}$
2. $5\tau = 6\ \mu\text{s}$; V_C is approximately 5 V.
3. $v_{2\text{ms}} = (0.86)10\text{ V} = 8.6\text{ V}$; $v_{3\text{ms}} = (0.95)10\text{ V} = 9.5\text{ V}$;
   $v_{4\text{ms}} = (0.98)10\text{ V} = 9.8\text{ V}$; $v_{5\text{ms}} = (0.99)10\text{ V} = 9.9\text{ V}$
4. $v_C = (0.37)(100\text{ V}) = 37\text{ V}$

SECTION 9–6 Capacitors in AC Circuits

1. $X_C = 1/2\pi fC = 677\text{ k}\Omega$
2. $f = 1/2\pi CX_C = 796\text{ Hz}$
3. $I_{\text{rms}} = 1\text{ V}/1.59\ \Omega = 629\text{ mA}$
4. Current leads voltage by 90°.
5. $P_{\text{true}} = 0\text{ W}$
6. $P_r = (12\text{ V})^2/318\ \Omega = 0.453\text{ VAR}$

SECTION 9–7 Capacitor Applications

1. Once the capacitor charges to the peak voltage, it discharges very little before the next peak thus smoothing the rectified voltage.
2. A coupling capacitor allows ac to pass from one point to another, but blocks constant dc.
3. A coupling capacitor must be large enough to have a negligible reactance at the frequency that is to be passed without opposition.
4. A decoupling capacitor shorts power line voltage transients to ground.
5. X_C is inversely proportional to frequency and so is the filter's ability to pass ac signals.

ANSWERS TO RELATED PROBLEMS FOR EXAMPLES

9–1 10 kV
9–2 0.047 μF
9–3 100,000,000 pF
9–4 62.7 pF
9–5 278 pF

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

9–6 1.54 μF

9–7 2.83 V

9–8 650 pF

9–9 0.052 μF

9–10 891 μs

9–11 8.36 V

9–12 8.13 V

9–13 0.7 ms; 95 V

9–14 0.42 V

9–15 3.39 kHz

9–16 **(a)** 1.83 kΩ
(b) 408 Ω

9–17 4.40 mA

9–18 8.72 V

9–19 0 W; 1.01 mVAR

ANSWERS TO TRUE/FALSE QUIZ

1. T **2.** F **3.** T **4.** F **5.** T **6.** T **7.** F **8.** F **9.** F **10.** T

ANSWERS TO SELF-TEST

1. (g) **2.** (b) **3.** (c) **4.** (d) **5.** (a) **6.** (d) **7.** (a) **8.** (f)
9. (c) **10.** (b) **11.** (d) **12.** (a) **13.** (b) **14.** (a) **15.** (b)

ANSWERS TO TROUBLESHOOTING: SYMPTOM AND CAUSE

1. (b)
2. (a)
3. (c)
4. (a)
5. (b)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

CHAPTER 10

RC CIRCUITS

OUTLINE

10–1 Sinusoidal Response of Series *RC* Circuits
10–2 Impedance and Phase Angle of Series *RC* Circuits
10–3 Analysis of Series *RC* Circuits
10–4 Impedance and Phase Angle of Parallel *RC* Circuits
10–5 Analysis of Parallel *RC* Circuits
10–6 Analysis of Series-Parallel *RC* Circuits
10–7 Power in *RC* Circuits
10–8 Basic Applications
10–9 Troubleshooting

OBJECTIVES

- Describe the relationship between current and voltage in a series *RC* circuit
- Determine impedance and phase angle in a series *RC* circuit
- Analyze a series *RC* circuit
- Determine impedance and phase angle in a parallel *RC* circuit
- Analyze a parallel *RC* circuit
- Analyze series-parallel *RC* circuits
- Determine power in *RC* circuits
- Discuss some basic *RC* applications
- Troubleshoot *RC* circuits

KEY TERMS

Impedance (Z)
Phase angle
***RC* lag circuit**
***RC* lead circuit**
Capacitive susceptance (B_C)
Admittance (Y)
Apparent power (P_a)
Power factor
Frequency response
Cutoff frequency
Bandwidth

INTRODUCTION

An *RC* circuit, which contains both resistance and capacitance, is one of the basic types of reactive circuits. In this chapter, series and parallel *RC* circuits and their responses to sinusoidal voltages are covered. Series-parallel combinations are also examined. Power considerations in *RC* circuits are introduced, and practical aspects of power ratings are discussed. Three *RC* circuit applications are presented to give you an idea of how simple combinations of resistors and capacitors can be applied. Troubleshooting common faults in *RC* circuits is also covered.

The methods for analyzing reactive circuits are similar to those you studied in dc circuits. Reactive circuit problems can be solved at only one frequency at a time, and phasor math must be used.

VISIT THE WEBSITE
Study aids for this chapter are available at
http://pearsonhighered.com/floyd

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

10–1 SINUSOIDAL RESPONSE OF SERIES *RC* CIRCUITS

When a sinusoidal voltage is applied to a series *RC* circuit, each resulting voltage drop and the current in the circuit(s) are also sinusoidal and have the same frequency as the source voltage. The capacitance causes a phase shift between the voltage and current that depends on the relative values of the resistance and the capacitive reactance.

After completing this section, you should be able to

- **Describe the relationship between current and voltage in a series *RC* circuit**
  - **Discuss voltage and current waveforms**
  - **Discuss phase shift**

As shown in Figure 10–1 for the case of a series *RC* circuit, the resistor voltage (V_R), the capacitor voltage (V_C), and the current (I) are all sine waves with the frequency of the source. Phase shifts are introduced because of the capacitance. As you will learn, the resistor voltage and current are in phase with each other and lead the source voltage in phase. The capacitor voltage lags the source voltage. The phase angle between the current and the capacitor voltage is always 90°. These generalized phase relationships are indicated in Figure 10–1.

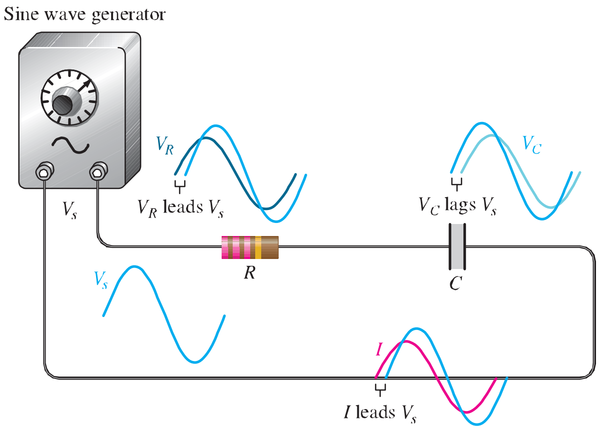

FIGURE 10–1 Illustration of sinusoidal response with general phase relationships of V_R, V_C, and I relative to the source voltage. V_R and I are in phase, and V_R and V_C are 90° out of phase.

The amplitudes and the phase relationships of the voltages and current depend on the values of the resistance and the capacitive reactance. When a circuit is purely resistive, the phase angle between the source voltage and the total current is zero. When a circuit is purely capacitive, the phase angle between the source voltage and the total current is 90°, with the current leading the voltage. When there is a combination of both resistance and capacitive reactance in a circuit, the phase angle between the source voltage and the total current is somewhere between zero and 90°, depending on the relative values of the resistance and the capacitive reactance.

SECTION 10–1 CHECKUP*

1. A 60 Hz sinusoidal voltage is applied to an *RC* circuit. What is the frequency of the capacitor voltage? What is the frequency of the current?
2. What determines the phase shift between V_s and I in a series *RC* circuit?
3. When the resistance in a series *RC* circuit is greater than the capacitive reactance, is the phase angle between the source voltage and the total current closer to 0° or to 90°?

*Answers are at the end of the chapter.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

10–2 IMPEDANCE AND PHASE ANGLE OF SERIES *RC* CIRCUITS

6238262 2013/12/10 12.27.66.5

In circuits where there is no reactance, the opposition to current is strictly resistance. In circuits having both resistance and reactance, the opposition to current is more complex because of the reactance and the resulting phase shift. Impedance, which is the total opposition to ac and includes the effect of phase shift, is introduced in this section.

After completing this section, you should be able to

- **Determine impedance and phase angle in a series *RC* circuit**
  - **Define *impedance***
  - **Define *phase angle***
  - **Draw an impedance triangle**
  - **Calculate the total impedance magnitude**
  - **Calculate the phase angle**

The **impedance** of a series *RC* circuit consists of resistance and capacitive reactance and is the total opposition to sinusoidal current. Its unit is the ohm. The **phase angle** is the phase difference between the total current and the source voltage.

In a purely resistive circuit, the impedance is simply equal to the total resistance. In a purely capacitive circuit, the impedance is the total capacitive reactance. The impedance of a series *RC* circuit is determined by both the resistance (R) and the capacitive reactance (X_C). These cases are illustrated in Figure 10–2. The magnitude of the impedance is symbolized by Z.

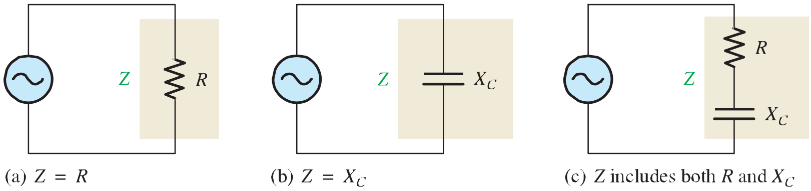

(a) $Z = R$ (b) $Z = X_C$ (c) Z includes both R and X_C

FIGURE 10–2 **Three cases of impedance.**

In ac analysis, both R and X_C are treated as phasor quantities, as shown in the phasor diagram of Figure 10–3(a), with X_C appearing at a $-90°$ angle with respect to R. This relationship comes from the fact that the capacitor voltage in a series *RC* circuit lags the current, and thus the resistor voltage, by 90°. Since Z is the phasor sum of R and X_C, its phasor representation is as shown in Figure 10–3(b). A repositioning of the phasors, as shown in Figure 10–3(c), forms a right triangle, which is called the *impedance triangle.* The length of each phasor represents the magnitude in ohms, and the angle θ (the Greek letter theta) is the phase angle of the *RC* circuit and represents the phase difference between the source voltage and the current.

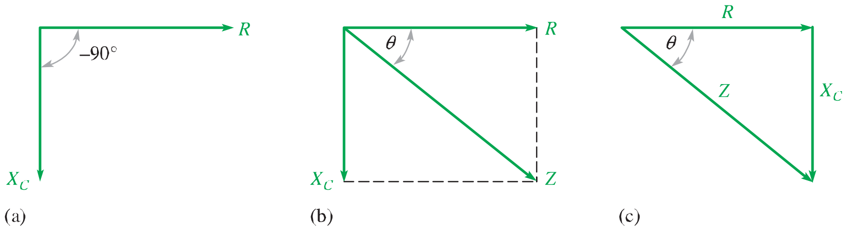

FIGURE 10–3 **Development of the impedance triangle for a series *RC* circuit.**

From right-triangle trigonometry (Pythagorean theorem), the magnitude (length) of the impedance can be expressed in terms of the resistance and capacitive reactance.

$$Z = \sqrt{R^2 + X_C^2} \qquad (10\text{–}1)$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The magnitude of the impedance (Z), as shown in the *RC* circuit in Figure 10–4, is expressed in ohms.

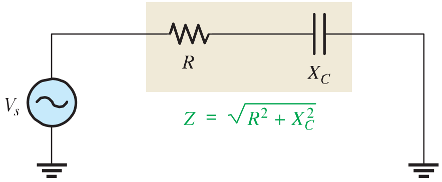

FIGURE 10–4 **Impedance in a series *RC* circuit.**

The value of the phase angle, θ, is expressed as

$$\boldsymbol{\theta = \tan^{-1}\left(\frac{X_C}{R}\right)} \qquad (10\text{–}2)$$

The symbol $\tan^{-1}$ stands for *inverse tangent* and can be found on most calculators by pressing the 2nd and TAN^{-1} keys. Another term for inverse tangent is *arctangent* (arctan).

EXAMPLE 10–1

Determine the impedance and the phase angle of the *RC* circuit in Figure 10–5. Draw the impedance triangle.

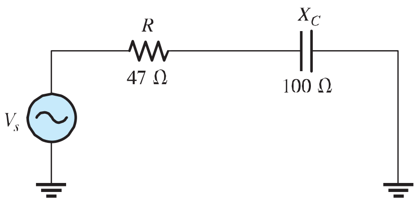

FIGURE 10–5

SOLUTION

The impedance is

$$Z = \sqrt{R^2 + X_C^2} = \sqrt{(47\ \Omega)^2 + (100\ \Omega)^2} = \mathbf{110\ \Omega}$$

The phase angle is

$$\theta = \tan^{-1}\left(\frac{X_C}{R}\right) = \tan^{-1}\left(\frac{100\ \Omega}{47\ \Omega}\right) = \tan^{-1}(2.13) = \mathbf{64.8^\circ}$$

The source voltage lags the current by 64.8°.

The impedance triangle is shown in Figure 10–6.

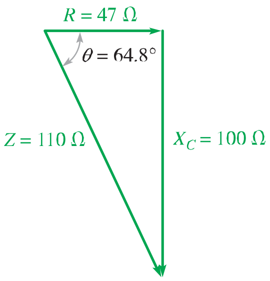

FIGURE 10–6

RELATED PROBLEM*

Find Z and θ for $R = 1.0\ \text{k}\Omega$ and $X_C = 2.2\ \text{k}\Omega$ in Figure 10–5.

*Answers are at the end of the chapter.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 10–2 CHECKUP

1. Define *impedance.*
2. Does the source voltage lead or lag the current in a series *RC* circuit?
3. What causes the phase angle in an *RC* circuit?
4. A series *RC* circuit has a resistance of 33 kΩ and a capacitive reactance of 50 kΩ. What is the value of the impedance? What is the phase angle?

10–3 ANALYSIS OF SERIES *RC* CIRCUITS

Ohm's law and Kirchhoff's voltage law are used in the analysis of series *RC* circuits to determine voltage, current, and impedance. Also, in this section *RC* lead and lag circuits are examined.

After completing this section, you should be able to

- **Analyze a series *RC* circuit**
  - **Apply Ohm's law and Kirchhoff's voltage law to series *RC* circuits**
  - **Determine the phase relationships of the voltages and current**
  - **Show how impedance and phase angle vary with frequency**
  - **Analyze the *RC* lag circuit**
  - **Analyze the *RC* lead circuit**

Ohm's Law

The application of Ohm's law to series *RC* circuits involves the use of the quantities of *Z*, *V*, and *I*. The three equivalent forms of Ohm's law are as follows:

$$V = IZ \tag{10–3}$$

$$I = \frac{V}{Z} \tag{10–4}$$

$$Z = \frac{V}{I} \tag{10–5}$$

The following two examples illustrate the use of Ohm's law.

EXAMPLE 10–2

The current in Figure 10–7 is 0.2 mA. Determine the source voltage and the phase angle. Draw the impedance triangle.

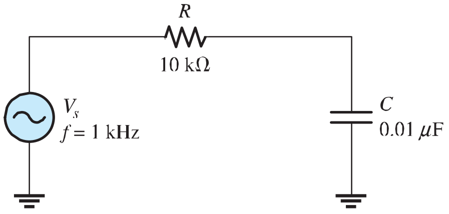

FIGURE 10–7

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

The capacitive reactance is

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi(1000\text{ Hz})(0.01\ \mu\text{F})} = 15.9\text{ k}\Omega$$

The impedance is

$$Z = \sqrt{R^2 + X_C^2} = \sqrt{(10\text{ k}\Omega)^2 + (15.9\text{ k}\Omega)^2} = 18.8\text{ k}\Omega$$

Applying Ohm's law yields

$$V_s = IZ = (0.2\text{ mA})(18.8\text{ k}\Omega) = \mathbf{3.76\ V}$$

The phase angle is

$$\theta = \tan^{-1}\left(\frac{X_C}{R}\right) = \tan^{-1}\left(\frac{15.9\text{ k}\Omega}{10\text{ k}\Omega}\right) = \mathbf{57.8^\circ}$$

The source voltage has a magnitude of 3.76 V and lags the current by 57.8°.

The impedance triangle is shown in Figure 10–8.

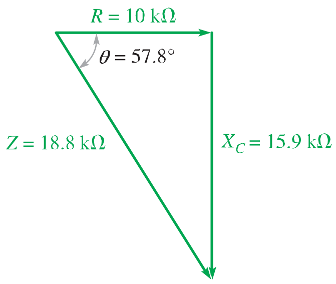

6238262 2013/12/10 12.27.66.5

FIGURE 10–8

RELATED PROBLEM

Determine V_s in Figure 10–7 if $f = 2$ kHz and $I = 200\ \mu$A.

EXAMPLE 10–3

Determine the current in the *RC* circuit of Figure 10–9.

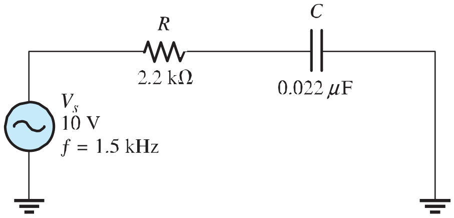

FIGURE 10–9

SOLUTION

The capacitive reactance is

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi(1.5\text{ kHz})(0.022\ \mu\text{F})} = 4.82\text{ k}\Omega$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The impedance is

$$Z = \sqrt{R^2 + X_C^2} = \sqrt{(2.2\text{ k}\Omega)^2 + (4.82\text{ k}\Omega)^2} = 5.30\text{ k}\Omega$$

Applying Ohm's law yields

$$I = \frac{V}{Z} = \frac{10\text{ V}}{5.30\text{ k}\Omega} = \mathbf{1.89\ mA}$$

RELATED PROBLEM

Determine the phase angle between V_s and I in Figure 10–9.

MULTISIM

Open Multisim file E10-03; files are found at www.pearsonhighered.com/floyd. Measure the current and the voltages across the resistor and the capacitor.

Phase Relationships of the Current and Voltages

In a series *RC* circuit, the current is the same through both the resistor and the capacitor. Thus, the resistor voltage is in phase with the current, and the capacitor voltage lags the current by 90°. Therefore, there is a phase difference of 90° between the resistor voltage, V_R, and the capacitor voltage, V_C, as shown in the waveform diagram of Figure 10–10.

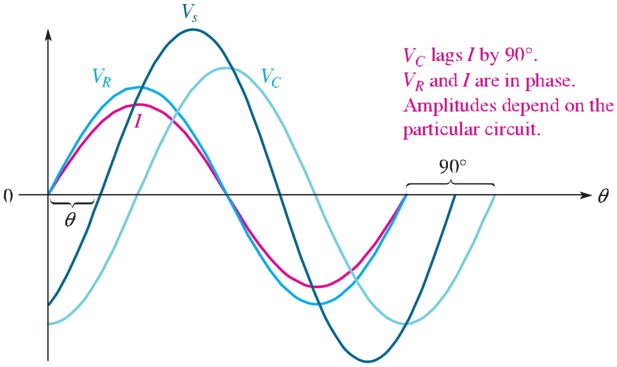

V_C lags I by 90°.
V_R and I are in phase.
Amplitudes depend on the particular circuit.

FIGURE 10–10 **Phase relation of the voltages and current in a series *RC* circuit.**

You know from Kirchhoff's voltage law that the sum of the voltage drops must equal the source voltage. However, since V_R and V_C are 90° out of phase with each other, they must be added as phasor quantities, with V_C lagging V_R, as shown in Figure 10–11(a). As shown in Figure 10–11(b), V_s is the phasor sum of V_R and V_C, as expressed in the following equation:

$$V_s = \sqrt{V_R^2 + V_C^2} \tag{10–6}$$

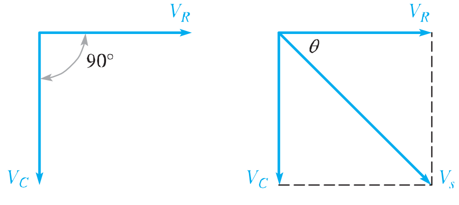

(a) (b)

FIGURE 10–11 **Voltage phasor diagram for the waveforms in Figure 10–10.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The value of the phase angle between the resistor voltage and the source voltage can be expressed as

$$\theta = \tan^{-1}\left(\frac{V_C}{V_R}\right) \qquad (10\text{–}7)$$

Since the resistor voltage and the current are in phase, θ in Equation 10–7 also represents the phase angle between the source voltage and the current and is equivalent to $\tan^{-1}(X_C/R)$.

Figure 10–12 shows the voltage and current phasor diagram that represents the waveform diagram of Figure 10–10.

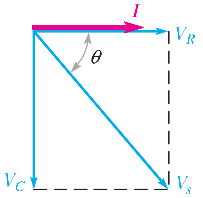

FIGURE 10–12 **Voltage and current phasor diagram for the waveforms in Figure 10–10.**

EXAMPLE 10–4

Determine the source voltage and the phase angle in Figure 10–13. Draw the voltage phasor diagram.

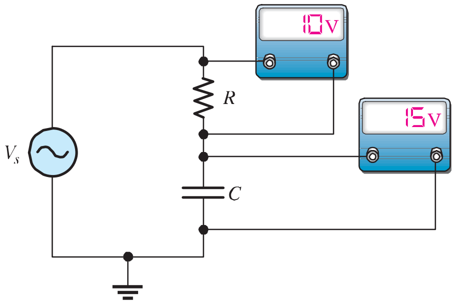

FIGURE 10–13

SOLUTION

Since V_R and V_C are 90° out of phase, you cannot add them directly. The source voltage is the phasor sum of V_R and V_C.

$$V_s = \sqrt{V_R^2 + V_C^2} = \sqrt{(10\text{ V})^2 + (15\text{ V})^2} = \mathbf{18\text{ V}}$$

The phase angle between the resistor voltage and the source voltage is

$$\theta = \tan^{-1}\left(\frac{V_C}{V_R}\right) = \tan^{-1}\left(\frac{15\text{ V}}{10\text{ V}}\right) = \mathbf{56.3^\circ}$$

The voltage phasor diagram is shown in Figure 10–14.

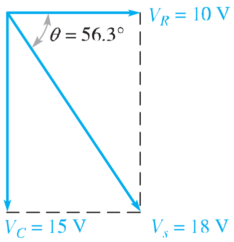

FIGURE 10–14

RELATED PROBLEM

In a certain series *RC* circuit, $V_s = 10$ V and $V_R = 7$ V. Find V_C.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Variation of Impedance and Phase Angle with Frequency

As you know, capacitive reactance varies inversely with frequency. Since $Z = \sqrt{R^2 + X_C^2}$, you can see that when X_C increases, the entire term under the square root sign increases and thus the total impedance also increases; and when X_C decreases, the total impedance also decreases. Therefore, *in a series RC circuit, Z is inversely dependent on frequency.*

Figure 10–15 illustrates how the voltages and current in a series *RC* circuit vary as the frequency increases or decreases, with the source voltage held at a constant value. Part (a) shows that as the frequency is increased, X_C decreases; so the voltage across the capacitor decreases. Also, Z decreases as X_C decreases, causing the current to increase. An increase in the current causes more voltage across R.

6238262 2013/12/10 12.27.66.5

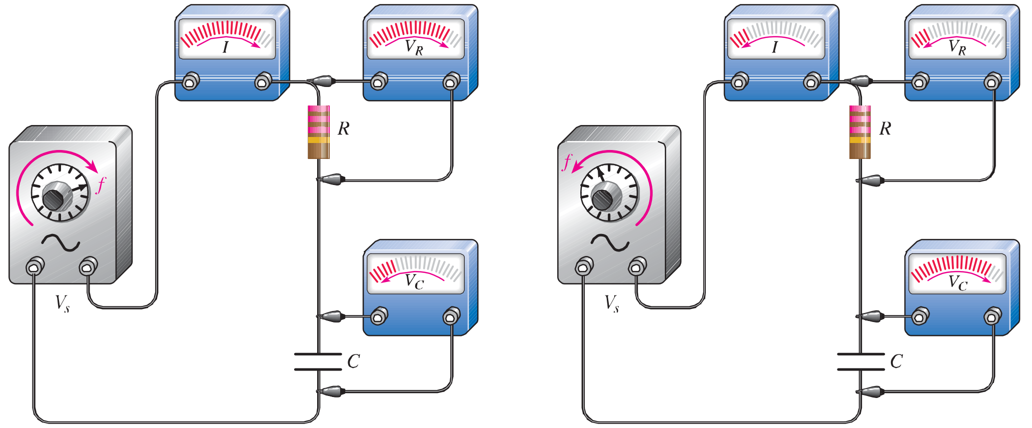

(a) As frequency is increased, Z decreases as X_C decreases, causing I and V_R to increase and V_C to decrease.

(b) As frequency is decreased, Z increases as X_C increases, causing I and V_R to decrease and V_C to increase.

FIGURE 10–15 **An illustration of how the variation of impedance affects the voltages and current as the source frequency is varied. The source voltage is held at a constant amplitude.**

Figure 10–15(b) shows that as the frequency is decreased, X_C increases; so more voltage is dropped across the capacitor. Also, Z increases as X_C increases, causing the current to decrease. A decrease in the current causes less voltage across R.

Changes in Z and X_C can be observed as shown in Figure 10–16. As the frequency increases, the voltage across Z remains constant because V_s is constant ($V_s = V_Z$). Also, the voltage across C decreases. The increasing current indicates that Z is decreasing. It

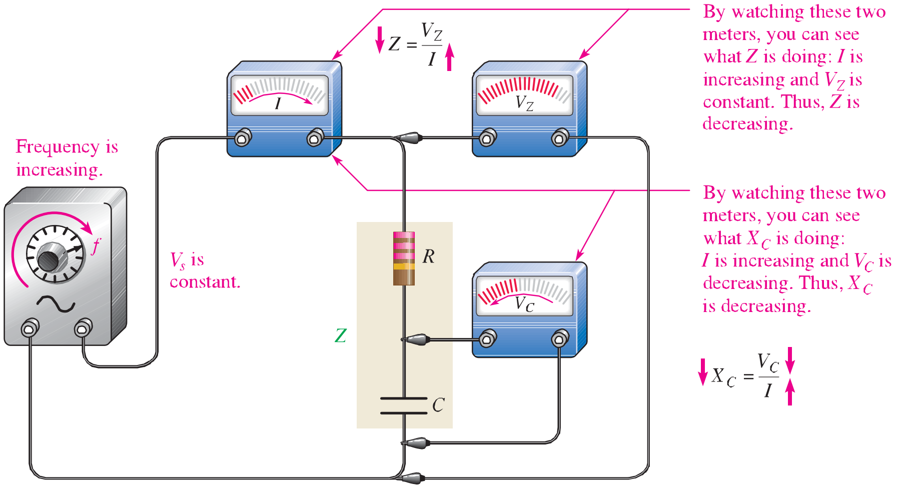

FIGURE 10–16 **An illustration of how Z and X_C change with frequency.**

$\downarrow Z = \frac{V_Z}{I\uparrow}$

By watching these two meters, you can see what Z is doing: I is increasing and V_Z is constant. Thus, Z is decreasing.

Frequency is increasing.

V_s is constant.

By watching these two meters, you can see what X_C is doing: I is increasing and V_C is decreasing. Thus, X_C is decreasing.

$\downarrow X_C = \frac{V_C\downarrow}{I\uparrow}$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

does so because of the inverse relationship stated in Ohm's law ($Z = V_Z/I$). The increasing current also indicates that X_C is decreasing ($X_C = V_C/I$). The decrease in V_C corresponds to the decrease in X_C.

Since X_C is the factor that introduces the phase angle in a series *RC* circuit, a change in X_C produces a change in the phase angle. As the frequency is increased, X_C becomes smaller, and thus the phase angle decreases. As the frequency is decreased, X_C becomes larger, and thus the phase angle increases. The angle between V_s and V_R is the phase angle of the circuit because I is in phase with V_R.

Figure 10–17 uses the impedance triangle to illustrate the variations in X_C, Z, and θ as the frequency changes. Of course, R remains constant. The key point is that because X_C varies inversely with the frequency, the magnitude of the total impedance and the phase angle also vary inversely with the frequency. Example 10–5 illustrates this.

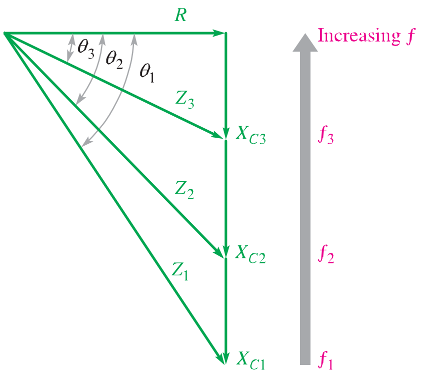

FIGURE 10–17 **As the frequency increases, X_C decreases, Z decreases, and θ decreases. Each value of frequency can be visualized as forming a different impedance triangle.**

EXAMPLE 10–5

For the series *RC* circuit in Figure 10–18, determine the impedance and phase angle for each of the following values of frequency:

(a) 10 kHz **(b)** 20 kHz **(c)** 30 kHz

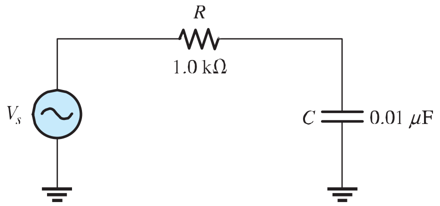

FIGURE 10–18

SOLUTION

(a) For $f = 10$ kHz, calculate the impedance as follows:

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi(10\text{ kHz})(0.01\ \mu\text{F})} = 1.59\text{ k}\Omega$$

$$Z = \sqrt{R^2 + X_C^2} = \sqrt{(1.0\text{ k}\Omega)^2 + (1.59\text{ k}\Omega)^2} = \mathbf{1.88\ k\Omega}$$

The phase angle is

$$\theta = \tan^{-1}\left(\frac{X_C}{R}\right) = \tan^{-1}\left(\frac{1.59\text{ k}\Omega}{1.0\text{ k}\Omega}\right) = \mathbf{57.8°}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

(b) For $f = 20\text{ kHz}$,

$$X_C = \frac{1}{2\pi(20\text{ kHz})(0.01\ \mu\text{F})} = 796\ \Omega$$
$$Z = \sqrt{(1.0\text{ k}\Omega)^2 + (796\ \Omega)^2} = \mathbf{1.28\ k\Omega}$$
$$\theta = \tan^{-1}\left(\frac{796\ \Omega}{1.0\text{ k}\Omega}\right) = \mathbf{38.5°}$$

(c) For $f = 30\text{ kHz}$,

$$X_C = \frac{1}{2\pi(30\text{ kHz})(0.01\ \mu\text{F})} = 531\ \Omega$$
$$Z = \sqrt{(1.0\text{ k}\Omega)^2 + (531\ \Omega)^2} = \mathbf{1.13\ k\Omega}$$
$$\theta = \tan^{-1}\left(\frac{531\ \Omega}{1.0\text{ k}\Omega}\right) = \mathbf{28.0°}$$

Notice that as the frequency increases, X_C, Z, and θ decrease.

RELATED PROBLEM

Find the total impedance and phase angle in Figure 10–18 for $f = 1\text{ kHz}$.

RC Lag Circuit

An ***RC* lag circuit** is a phase shift circuit in which the output voltage lags the input voltage by a specified angle, ϕ. Phase shift circuits are commonly used in electronic communication systems and for other applications.

A basic series *RC* lag circuit is shown in Figure 10–19(a). Keep in mind that phase angle θ is measured between the source (input) voltage and the current. In terms of voltages, this is equivalent to saying the phase angle is measured between V_{in} and V_R because current and voltage are in phase in a resistor. The output is taken across the capacitor; therefore, the phase lag between V_{in} and V_C is $90° - \theta$. This angle is designated as ϕ and represents the phase difference between the input (V_{in}) and the output (V_{out}), as shown in Figure 10–19(b).

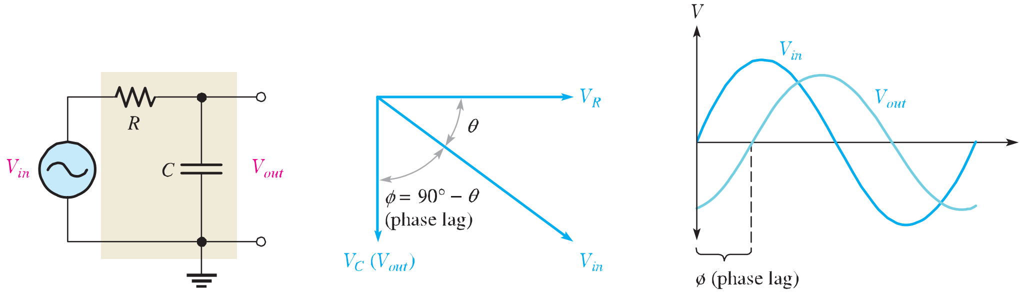

(a) A basic *RC* lag circuit

(b) Phasor voltage diagram showing the phase lag between V_{in} and V_{out}

(c) Input and output voltage waveforms

FIGURE 10–19 **The *RC* lag circuit ($V_{out} = V_C$).**

Since $\theta = \tan^{-1}(X_C/R)$, the value of the phase lag, ϕ, can be expressed as

$$\boldsymbol{\phi = 90° - \tan^{-1}\left(\frac{X_C}{R}\right)} \qquad \textbf{(10–8)}$$

The input and output voltage waveforms of the lag circuit are shown in Figure 10–19(c). The exact amount of phase lag between the input and the output depends on the values of the resistance and the capacitive reactance. The magnitude of the output voltage depends on these values also.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 10–6

Determine the amount of phase lag from input to output in the lag circuit in Figure 10–20.

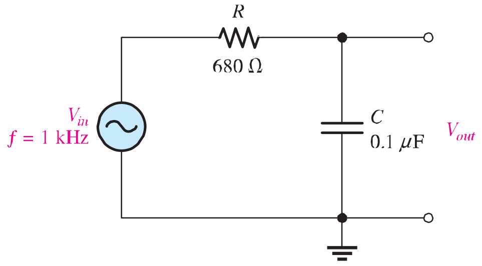

FIGURE 10–20

SOLUTION

First determine the capacitive reactance.

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi(1\text{ kHz})(0.1\ \mu\text{F})} = 1.59\text{ k}\Omega$$

The phase lag between the output voltage and the input voltage is

$$\phi = 90^\circ - \tan^{-1}\left(\frac{X_C}{R}\right) = 90^\circ - \tan^{-1}\left(\frac{1.59\text{ k}\Omega}{680\ \Omega}\right) = \mathbf{23.2^\circ}$$

The output voltage lags the input voltage by 23.2°.

RELATED PROBLEM

In a lag circuit, what happens to the phase lag if the frequency increases?

The phase-lag circuit can be considered as a voltage divider with a portion of the input voltage dropped across *R* and a portion across *C*. The output voltage can be determined with the following formula:

6238262 2013/12/10 12.27.66.5

$$V_{out} = \left(\frac{X_C}{\sqrt{R^2 + X_C^2}}\right)V_{in} \qquad (10\text{–}9)$$

EXAMPLE 10–7

For the lag circuit in Figure 10–20 of Example 10–6, determine the output voltage when the input voltage has an rms value of 10 V. Draw the input and output waveforms showing the proper relationships. The values for X_C (1.59 kΩ) and ϕ (23.2°) were found in Example 10–6.

SOLUTION

Use Equation 10–9 to determine the output voltage for the lag circuit in Figure 10–20.

$$V_{out} = \left(\frac{X_C}{\sqrt{R^2 + X_C^2}}\right)V_{in} = \left(\frac{1.59\text{ k}\Omega}{\sqrt{(680\ \Omega)^2 + (1.59\text{ k}\Omega)^2}}\right)10\text{ V} = \mathbf{9.2\ V\ rms}$$

The waveforms are shown in Figure 10–21. (Note that the rms value is converted to peak value on the plot.)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

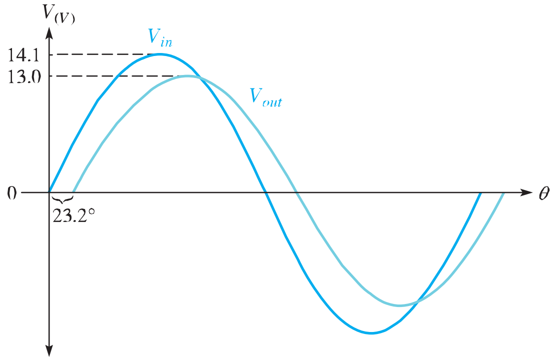

FIGURE 10–21

RELATED PROBLEM

In a lag circuit, what happens to the output voltage if the frequency increases?

EFFECTS OF FREQUENCY ON THE LAG CIRCUIT Since the circuit phase angle, θ, decreases as frequency increases, the phase lag ϕ between the input and the output voltages increases. You can see this relationship by examining Equation 10–8. Also, the magnitude of V_{out} decreases as the frequency increases because X_C becomes smaller and less of the total input voltage is dropped across the capacitor.

RC Lead Circuit

An ***RC* lead circuit** is a phase shift circuit in which the output voltage leads the input voltage by a specified angle, ϕ. A basic *RC* lead circuit is shown in Figure 10–22(a). Notice how it differs from the lag circuit. Here the output voltage is taken across the resistor. The relationship of the voltages is given in the phasor diagram in Figure 10–22(b). The output voltage, V_{out}, leads V_{in} by an angle that is the same as the circuit phase angle because V_R and *I* are in phase with each other.

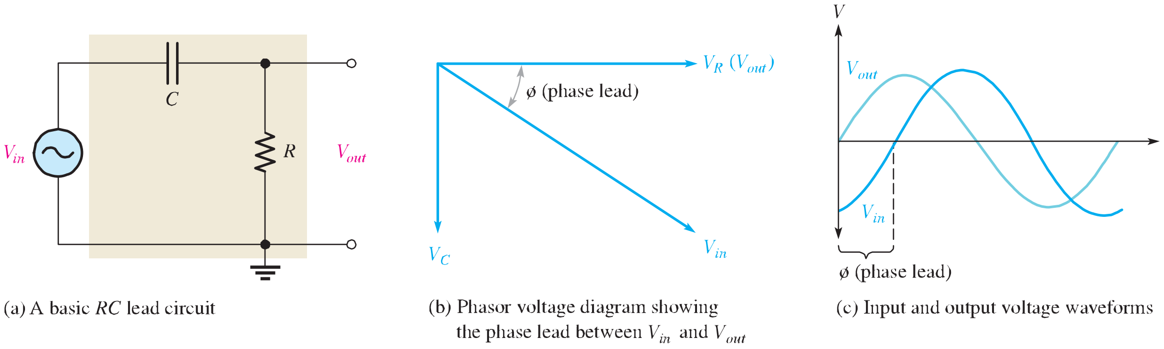

(a) A basic *RC* lead circuit

(b) Phasor voltage diagram showing the phase lead between V_{in} and V_{out}

(c) Input and output voltage waveforms

FIGURE 10–22 **The *RC* lead circuit ($V_{out} = V_R$).**

When the input and output waveforms are displayed on an oscilloscope, a relationship similar to that in Figure 10–22(c) is observed. Of course, the exact amount of phase lead and the output voltage magnitude depend on the values of *R* and X_C. The value of the phase lead, ϕ, is expressed as

$$\phi = \tan^{-1}\left(\frac{X_C}{R}\right) \quad (10\text{–}10)$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The output voltage is expressed as

$$V_{out} = \left(\frac{R}{\sqrt{R^2 + X_C^2}}\right)V_{in} \qquad (10\text{–}11)$$

EXAMPLE 10–8

Calculate the phase lead and the output voltage for the circuit in Figure 10–23.

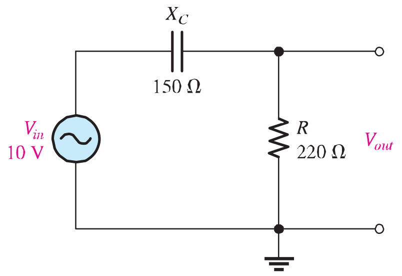

FIGURE 10–23

SOLUTION

The phase lead is

$$\phi = \tan^{-1}\left(\frac{X_C}{R}\right) = \tan^{-1}\left(\frac{150\ \Omega}{220\ \Omega}\right) = \mathbf{34.3^\circ}$$

The output leads the input by 34.3°.

Use Equation 10–11 to determine the output voltage.

$$V_{out} = \left(\frac{R}{\sqrt{R^2 + X_C^2}}\right)V_{in} = \left(\frac{220\ \Omega}{\sqrt{(220\ \Omega)^2 + (150\ \Omega)^2}}\right)10\text{ V} = \mathbf{8.26\ V}$$

RELATED PROBLEM

How does an increase in R affect the phase lead and the output voltage in Figure 10–23?

EFFECTS OF FREQUENCY ON THE LEAD CIRCUIT Since the phase lead is the same as the circuit phase angle θ, it decreases as frequency increases. The output voltage increases with frequency because as X_C becomes smaller, more of the input voltage is dropped across the resistor.

SYSTEM EXAMPLE 10–1

A BUFFER AMPLIFIER FOR A PREAMP

Often an analog system has a low-level input signal that needs to be amplified to avoid transmission problems before sending it to a controller or processor. For low signal levels, a field-effect transistor (FET) is often used as a buffer amplifier followed by a gain stage, forming a simple preamp system. A buffer amplifier for voltage is one that has a gain of 1 (output voltage equals input voltage) but presents a high impedance to the input signal thus preventing loading of a low-impedance source. Figure 10–24(a) shows a buffer amplifier, which then drives other gain stages. The equivalent input circuit for this amplifier is just the coupling capacitor, C_1, and the bias resistor, R_G, as shown in part (b). For calculating any effect on the driving circuit, this is the only part of the amplifier that needs to be considered because the transistor looks like an open to the input signal.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

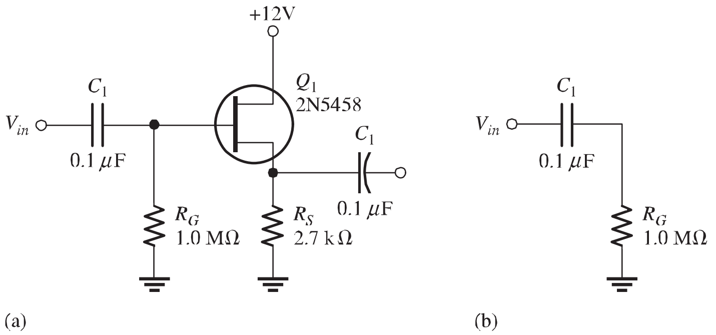

FIGURE 10–24 **A buffer amplifier.**

The input of the amplifier contributes to the lower cutoff frequency of the entire preamp. It can be simplified to the basic *RC* circuit that you studied in this section. In fact, the input to *any* capacitively coupled amplifier can be reduced to this same basic circuit. This simplification allows you to apply basic dc/ac theory to a more-complicated circuit to calculate a response. For example, you can easily determine the frequency at which the voltage across the capacitor is equal to the voltage across the resistor for the buffer amplifier. (This frequency is called the cutoff frequency, f_c, which is discussed in Section 10–8.) If the voltage across the capacitor is equal to the voltage across the resistor, then $X_C = R$. Then,

$$X_C = \frac{1}{2\pi f_c C} = R$$

$$f_c = \frac{1}{2\pi RC} = \frac{1}{2\pi(1.0\text{ M}\Omega)(0.1\ \mu\text{F})} = 1.59\text{ Hz}$$

SECTION 10–3 CHECKUP

1. In a certain series *RC* circuit, $V_R = 4$ V and $V_C = 6$ V. What is the magnitude of the source voltage?
2. In Question 1, what is the phase angle?
3. What is the phase difference between the capacitor voltage and the resistor voltage in a series *RC* circuit?
4. When the frequency of the source voltage in a series *RC* circuit is increased, what happens to each of the following?
   - **(a)** the capacitive reactance **(b)** the impedance
   - **(c)** the phase angle
5. A certain *RC* lag circuit consists of a 4.7 kΩ resistor and a 0.022 μF capacitor. Determine the phase lag between the input and output voltages at a frequency of 3 kHz.
6. An *RC* lead circuit has the same component values as the lag circuit in Question 5. What is the magnitude of the output voltage at 3 kHz when the input is 10 V rms?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

10–4 IMPEDANCE AND PHASE ANGLE OF PARALLEL *RC* CIRCUITS

In this section, you will learn how to determine the impedance and phase angle of a parallel *RC* circuit. Also conductance (G), capacitive susceptance (B_C), and total admittance (Y_{tot}) are discussed because of their usefulness in parallel circuit analysis.

After completing this section, you should be able to

- **Determine impedance and phase angle in a parallel *RC* circuit**
  - **Express total impedance in a product-over-sum form**
  - **Express the phase angle in terms of R and X_C**
  - **Determine the conductance, capacitive susceptance, and admittance**
  - **Convert admittance to impedance**

Figure 10–25 shows a basic parallel *RC* circuit.

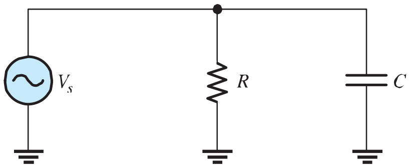

FIGURE 10–25 **Parallel *RC* circuit.**

The expression for the impedance in Equation 10–12 is given in a product-over-sum form similar to the way two resistors in parallel can be expressed. In this case, the denominator is the phasor sum of R and X_C.

$$Z = \frac{RX_C}{\sqrt{R^2 + X_C^2}} \qquad (10\text{–}12)$$

The value of the phase angle between the source voltage and the total current can be expressed in terms of R and X_C as shown in Equation 10–13.

$$\theta = \tan^{-1}\left(\frac{R}{X_C}\right) \qquad (10\text{–}13)$$

This formula is derived from an equivalent formula using branch currents (Eq. 10–22), which is introduced in Section 10–5.

EXAMPLE 10–9

For each circuit in Figure 10–26, determine the impedance and the phase angle.

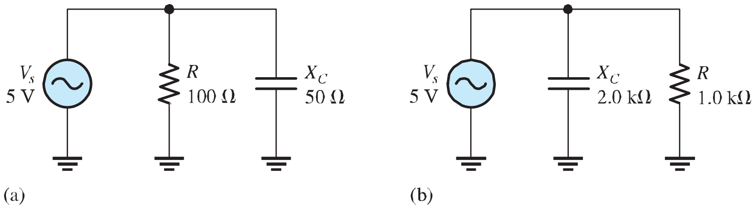

FIGURE 10–26

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

For the circuit in Figure 10–26(a), the impedance and phase angle are

$$Z = \frac{RX_C}{\sqrt{R^2 + X_C^2}} = \frac{(100\ \Omega)(50\ \Omega)}{\sqrt{(100\ \Omega)^2 + (50\ \Omega)^2}} = \mathbf{44.7\ \Omega}$$

$$\theta = \tan^{-1}\left(\frac{R}{X_C}\right) = \tan^{-1}\left(\frac{100\ \Omega}{50\ \Omega}\right) = \mathbf{63.4°}$$

For the circuit in Figure 10–26(b),

$$Z = \frac{(1.0\ \text{k}\Omega)(2.0\ \text{k}\Omega)}{\sqrt{(1.0\ \text{k}\Omega)^2 + (2.0\ \text{k}\Omega)^2}} = \mathbf{894\ \Omega}$$

$$\theta = \tan^{-1}\left(\frac{1.0\ \text{k}\Omega}{2.0\ \text{k}\Omega}\right) = \mathbf{26.6°}$$

RELATED PROBLEM

Determine Z in Figure 10–26(a) if the frequency is doubled.

Conductance, Capacitive Susceptance, and Admittance

Recall that **conductance (*G*)** is the reciprocal of resistance and is expressed as

$$\boxed{G = \frac{1}{R}} \qquad (10\text{–}14)$$

Two new terms are now introduced for use in parallel *RC* circuits. Susceptance is the reciprocal of reactance; therefore, **capacitive susceptance (B_C)** is the reciprocal of capacitive reactance and is a measure of the ability of a capacitor to permit current. It is expressed as

$$\boxed{B_C = \frac{1}{X_C}} \qquad (10\text{–}15)$$

Admittance (*Y*) is the reciprocal of impedance and is expressed as

$$\boxed{Y = \frac{1}{Z}} \qquad (10\text{–}16)$$

The unit of each of these three quantities is the siemens (S), which is the reciprocal of the ohm.

In working with parallel circuits, it is often easier to use conductance (G), capacitive susceptance (B_C), and admittance (Y) rather than resistance (R), capacitive reactance (X_C) and impedance (Z). In a parallel *RC* circuit, as shown in Figure 10–27(a), the total admittance is the phasor sum of the conductance and the capacitive susceptance, as shown in Figure 10–27(b).

$$\boxed{Y_{tot} = \sqrt{G^2 + B_C^2}} \qquad (10\text{–}17)$$

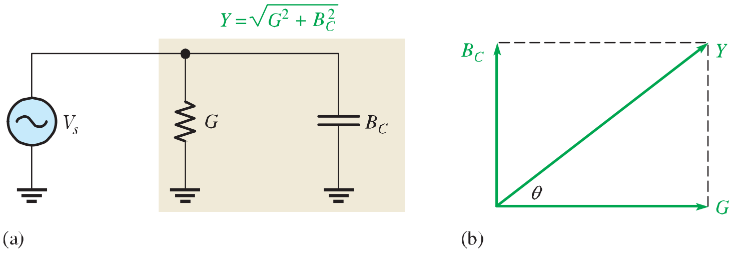

FIGURE 10–27 **Admittance in a parallel *RC* circuit.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 10–10

Determine the total admittance in Figure 10–28, and then convert it to impedance.

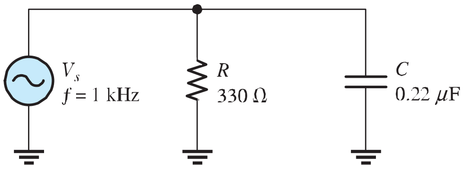

FIGURE 10–28

SOLUTION

To determine Y, first calculate the values for G and B_C. Since $R = 330\ \Omega$,

$$G = \frac{1}{R} = \frac{1}{330\ \Omega} = 3.03\text{ mS}$$

The capacitive reactance is

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi(1000\text{ Hz})(0.22\ \mu\text{F})} = 723\ \Omega$$

The capacitive susceptance is

$$B_C = \frac{1}{X_C} = \frac{1}{723\ \Omega} = 1.38\text{ mS}$$

6238262 2013/12/10 12.27.66.5

Therefore, the total admittance is

$$Y_{tot} = \sqrt{G^2 + B_C^2} = \sqrt{(3.03\text{ mS})^2 + (1.38\text{ mS})^2} = \mathbf{3.33\ mS}$$

Convert to impedance.

$$Z = \frac{1}{Y_{tot}} = \frac{1}{3.33\text{ mS}} = \mathbf{300\ \Omega}$$

RELATED PROBLEM

Calculate the admittance in Figure 10–28 if f is increased to 2.5 kHz.

SECTION 10–4 CHECKUP

1. Determine Z if a 1.0 kΩ resistance is in parallel with a 650 Ω capacitive reactance.

2. Define *conductance, capacitive susceptance,* and *admittance.*

3. If $Z = 100\ \Omega$, what is the value of Y?

4. In a certain parallel *RC* circuit, $R = 50\ \Omega$ and $X_C = 75\ \Omega$. Determine Y.

10–5 ANALYSIS OF PARALLEL *RC* CIRCUITS

Ohm's law and Kirchhoff's current law are used in the analysis of *RC* circuits. Current and voltage relationships in a parallel *RC* circuit are examined.

After completing this section, you should be able to

- **Analyze a parallel *RC* circuit**
    - **Apply Ohm's law and Kirchhoff's current law to parallel *RC* circuits**
    - **Show how impedance and phase angle vary with frequency**
    - **Convert from a parallel circuit to an equivalent series circuit**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

For convenience in the analysis of parallel circuits, the Ohm's law formulas using impedance—Equations 10–3, 10–4, and 10–5—can be rewritten for admittance using the relation $Y = 1/Z$.

$$\boldsymbol{V = \frac{I}{Y}} \qquad \text{(10–18)}$$

$$\boldsymbol{I = VY} \qquad \text{(10–19)}$$

$$\boldsymbol{Y = \frac{I}{V}} \qquad \text{(10–20)}$$

EXAMPLE 10–11

Determine the total current and the phase angle in Figure 10–29.

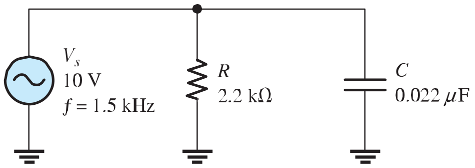

FIGURE 10–29

SOLUTION

First, determine the total admittance. The capacitive reactance is

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi(1.5\text{ kHz})(0.022\ \mu\text{F})} = 4.82\text{ k}\Omega$$

The conductance is

$$G = \frac{1}{R} = \frac{1}{2.2\text{ k}\Omega} = 455\ \mu\text{S}$$

The capacitive susceptance is

$$B_C = \frac{1}{X_C} = \frac{1}{4.82\text{ k}\Omega} = 207\ \mu\text{S}$$

Therefore, the total admittance is

$$Y_{tot} = \sqrt{G^2 + B_C^2} = \sqrt{(455\ \mu\text{S})^2 + (207\ \mu\text{S})^2} = 500\ \mu\text{S}$$

Next, use Ohm's law to calculate the total current.

$$I_{tot} = VY_{tot} = (10\text{ V})(500\ \mu\text{S}) = \mathbf{5.00\ mA}$$

The phase angle is

$$\theta = \tan^{-1}\left(\frac{R}{X_C}\right) = \tan^{-1}\left(\frac{2.2\text{ k}\Omega}{4.82\text{ k}\Omega}\right) = \mathbf{24.5^\circ}$$

The total current is 5.00 mA, and it leads the source voltage by 24.5°.

RELATED PROBLEM

What is the total current if the frequency is doubled?

MULTISIM

Open Multisim file E10-11. Verify the value of the total current that was calculated. Then measure each branch current. Double the frequency to 3 kHz and measure the total current.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Phase Relationships of the Currents and Voltages

Figure 10–30(a) shows all the currents and voltages in a basic parallel *RC* circuit. As you can see, the source voltage, V_s, appears across both the resistive and the capacitive branches, so V_s, V_R, and V_C are all in phase and of the same magnitude. The total current, I_{tot}, divides at the junction into the two branch currents, I_R and I_C.

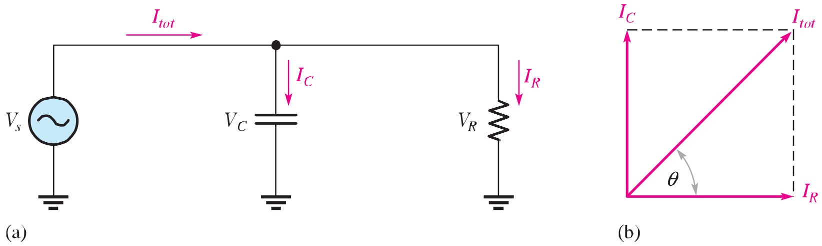

FIGURE 10–30 **Currents and voltages in a parallel *RC* circuit. The current directions shown in (a) are instantaneous and, of course, reverse when the source voltage reverses. The current phasors in (b) rotate once per cycle.**

Stray Capacitance

Stray capacitance is the unseen capacitance that is in all electronic circuits and systems. It occurs anytime conductors are separated by an insulator as in the case of a basic transistor. The effect of stray capacitance is to reduce the high-frequency response of analog circuits. In high-speed digital circuits, it limits switching speed. The stray capacitance for a transistor can be drawn as an equivalent parallel *RC* circuit, enabling high-frequency analysis of a transistor circuit by applying a basic circuit concept.

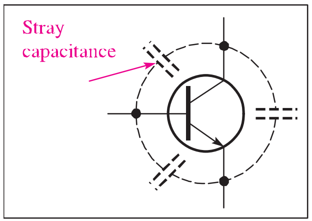

SYSTEM NOTE

The current through the resistor is in phase with the voltage. The current through the capacitor leads the voltage, and thus the resistive current, by 90°. By Kirchhoff's current law, the total current is the phasor sum of the two branch currents, as shown by the phasor diagram in Figure 10–30(b). The total current is expressed as

$$\boldsymbol{I_{tot} = \sqrt{I_R^2 + I_C^2}} \qquad \text{(10–21)}$$

The value of the phase angle between the resistor current and the total current is

$$\boldsymbol{\theta = \tan^{-1}\left(\frac{I_C}{I_R}\right)} \qquad \text{(10–22)}$$

Equation 10–22 is equivalent to Equation 10–13, $\theta = \tan^{-1}(R/X_C)$.

Figure 10–31 shows a complete current and voltage phasor diagram. Notice that I_C leads I_R by 90° and that I_R is in phase with the voltage ($V_s = V_R = V_C$).

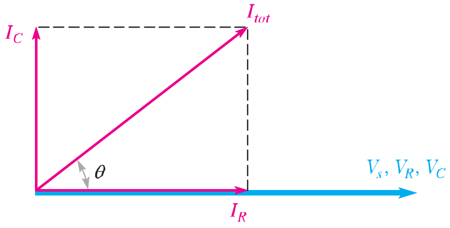

FIGURE 10–31 **Current and voltage phasor diagram for a parallel *RC* circuit (amplitudes depend on the particular circuit).**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 10–12

Determine the value of each current in Figure 10–32, and describe the phase relationship of I_R, I_C, and I_{tot}. Draw the current phasor diagram.

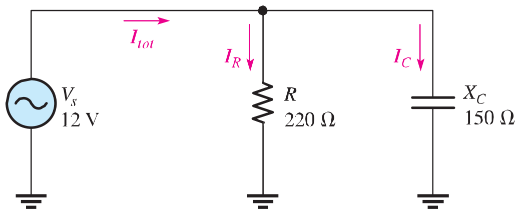

FIGURE 10–32

SOLUTION

The resistor current, the capacitor current, and the total current are expressed as follows:

$$I_R = \frac{V_s}{R} = \frac{12\text{ V}}{220\ \Omega} = \mathbf{54.5\ mA}$$

$$I_C = \frac{V_s}{X_C} = \frac{12\text{ V}}{150\ \Omega} = \mathbf{80\ mA}$$

$$I_{tot} = \sqrt{I_R^2 + I_C^2} = \sqrt{(54.5\text{ mA})^2 + (80\text{ mA})^2} = \mathbf{96.8\ mA}$$

The phase angle is

$$\theta = \tan^{-1}\left(\frac{I_C}{I_R}\right) = \tan^{-1}\left(\frac{80\text{ mA}}{54.5\text{ mA}}\right) = 55.7°$$

I_R is in phase with the source voltage, I_C leads the source voltage by 90°, and I_{tot} leads the source voltage by 55.7°. The current phasor diagram is shown in Figure 10–33.

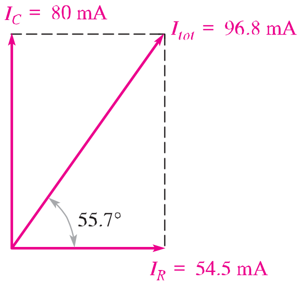

FIGURE 10–33

RELATED PROBLEM

In a certain parallel circuit, $I_R = 100$ mA and $I_C = 60$ mA. Determine the total current and the phase angle.

6238262 2013/12/10 12.27.66.5

Note that a parallel *RC* circuit becomes less reactive when X_C is increased. That is, the circuit phase angle becomes smaller. The reason for this effect is that when X_C is increased relative to *R*, less current is through the capacitive branch, and although the in-phase or resistive current does not increase, it becomes a greater percentage of the total current.

SECTION 10–5 CHECKUP

1. The admittance of a parallel *RC* circuit is 3.5 mS, and the source voltage is 6 V. What is the total current?
2. In a certain parallel *RC* circuit, the resistor current is 10 mA, and the capacitor current is 15 mA. Determine the phase angle and the total current.
3. What is the phase angle between the capacitor current and the source voltage in a parallel *RC* circuit?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

10–6 ANALYSIS OF SERIES-PARALLEL *RC* CIRCUITS

The concepts studied in the previous sections are used to analyze circuits with combinations of both series and parallel *R* and *C* components.

After completing this section, you should be able to

- **Analyze series-parallel *RC* circuits**
  - **Determine total impedance**
  - **Calculate currents and voltages**
  - **Measure impedance and phase angle**

As in the case of dc circuits, combinational ac circuits can be solved by combining series or parallel elements into equivalent circuits. The following example demonstrates the analysis of a series-parallel reactive circuit.

EXAMPLE 10–13

In the series-parallel *RC* circuit of Figure 10–34, determine the following:

(a) total impedance **(b)** total current **(c)** phase angle by which I_{tot} leads V_s

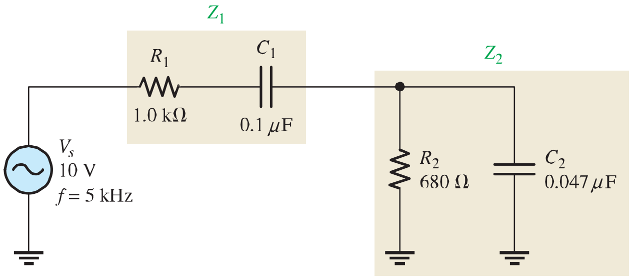

FIGURE 10–34

SOLUTION

(a) First, calculate the magnitudes of the capacitive reactances.

$$X_{C1} = \frac{1}{2\pi(5\text{ kHz})(0.1\ \mu\text{F})} = 318\ \Omega$$

$$X_{C2} = \frac{1}{2\pi(5\text{ kHz})(0.047\ \mu\text{F})} = 677\ \Omega$$

One approach is to find the series equivalent resistance and capacitive reactance for the parallel portion of the circuit; then add the resistances ($R_1 + R_{eq}$) to get total resistance and add the reactances ($X_{C1} + X_{C(eq)}$) to get total reactance. From these totals, you can determine the total impedance.

Find the impedance of the parallel portion (Z_2) by first finding the admittance.

$$G_2 = \frac{1}{R_2} = \frac{1}{680\ \Omega} = 1.47\text{ mS}$$

$$B_{C2} = \frac{1}{X_{C2}} = \frac{1}{677\ \Omega} = 1.48\text{ mS}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

$$Y_2 = \sqrt{G_2^2 + B_{C2}^2} = \sqrt{(1.47\text{ mS})^2 + (1.48\text{ mS})^2} = 2.09\text{ mS}$$

$$Z_2 = \frac{1}{Y_2} = \frac{1}{2.09\text{ mS}} = 478\ \Omega$$

The phase angle associated with the parallel portion of the circuit is

$$\theta_p = \tan^{-1}\left(\frac{R_2}{X_{C2}}\right) = \tan^{-1}\left(\frac{680\ \Omega}{677\ \Omega}\right) = 45.1^\circ$$

The series equivalent values for the parallel portion are

$$R_{\text{eq}} = Z_2\cos\theta_p = (478\ \Omega)\cos(45.1^\circ) = 337\ \Omega$$

$$X_{C(\text{eq})} = Z_2\sin\theta_p = (478\ \Omega)\sin(45.1^\circ) = 339\ \Omega$$

The total circuit resistance is

$$R_{tot} = R_1 + R_{\text{eq}} = 1000\ \Omega + 337\ \Omega = 1.34\text{ k}\Omega$$

The total circuit reactance is

$$X_{C(tot)} = X_{C1} + X_{C(\text{eq})} = 318\ \Omega + 339\ \Omega = 657\ \Omega$$

The total circuit impedance is

$$Z_{tot} = \sqrt{R_{tot}^2 + X_{C(tot)}^2} = \sqrt{(1.34\text{ k}\Omega)^2 + (657\ \Omega)^2} = \mathbf{1.49\text{ k}\Omega}$$

(b) Use Ohm's law to find the total current.

$$I_{tot} = \frac{V_s}{Z_{tot}} = \frac{10\text{ V}}{1.49\text{ k}\Omega} = \mathbf{6.71\text{ mA}}$$

(c) To find the phase angle, view the circuit as a series combination of R_{tot} and $X_{C(tot)}$. The phase angle by which I_{tot} leads V_s is

$$\theta = \tan^{-1}\left(\frac{X_{C(tot)}}{R_{tot}}\right) = \tan^{-1}\left(\frac{657\ \Omega}{1.34\text{ k}\Omega}\right) = \mathbf{26.1^\circ}$$

RELATED PROBLEM

Determine the voltages across Z_1 and Z_2 in Figure 10–34.

MULTISIM

Open Multisim file E10-13. Verify the calculated value of the total current. Measure the current through R_2 and the current through C_2. Measure voltages across Z_1 and Z_2.

SYSTEM EXAMPLE 10–2

PASSIVE OSCILLOSCOPE PROBES

Although there are many types of oscilloscope probes, the most common type is the ×10 passive probe. The ×10 refers to the fact that the probe attenuates the signal by a factor of 10. The purpose of the probe is to transmit as accurately as possible the signal while avoiding noise pickup or affecting the circuit under test (loading).

A typical oscilloscope input has about 20 pF of internal capacitance due to the input amplifier, cables, and stray capacitance. These are lumped into C_{in} in the circuit shown in Figure 10–35. The dotted lines are a reminder that this is not a physical capacitor but has all of the attributes of one. There is also about 1.0 MΩ of internal resistance for the input amplifier in the oscilloscope, shown as R_{in}. To complete the circuit, there is typically a series resistor in the probe itself, R_{probe}, and a small variable capacitor, C_{probe}, that is adjusted to optimize the overall probe/scope response.

At low frequencies, the small capacitors have very high reactance compared to the resistance. Ignoring the capacitors, the resistors form a 10:1 resistive divider. At high frequencies, the capacitive reactance of C_{probe} and C_{in} are both small compared to the resistors. Ignoring the resistances, C_{probe} forms a 10:1 capacitive divider with C_{in}, which keeps

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

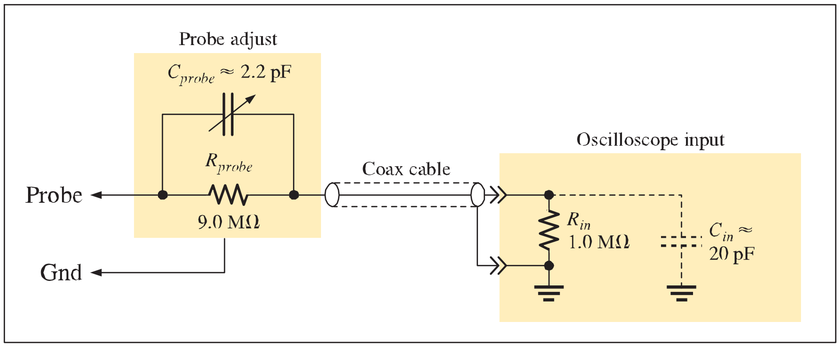
Probe adjust
$C_{probe} \approx 2.2$ pF
R_{probe}
9.0 MΩ
Probe
Gnd
Coax cable
Oscilloscope input
R_{in}
1.0 MΩ
$C_{in} \approx$
20 pF

FIGURE 10–35 A passive oscilloscope probe.

the frequency response flat. The net effect of the probe circuit for an oscilloscope is to reduce the input capacitance by the probe's ×10 ratio and increase the input impedance by the same amount, thus reducing circuit loading.

Circuit Measurements

DETERMINING Z_{tot} Now, let's see how the value of Z_{tot} for the circuit in Example 10–13 can be determined by measurement. First, the total impedance is measured as outlined in the following steps and as illustrated in Figure 10–36 (other ways are also possible).

Step 1: Using a sine wave generator, set the source voltage to a known value (10 V) and the frequency to 5 kHz. Check the voltage with an ac voltmeter and the frequency with a frequency counter or an oscilloscope rather than relying on the marked values on the generator controls.

Step 2: Connect an ac ammeter as shown in Figure 10–36, and measure the total current. Alternately, you can measure the voltage across R_1 and calculate the current.

Step 3: Calculate the total impedance by using Ohm's law.

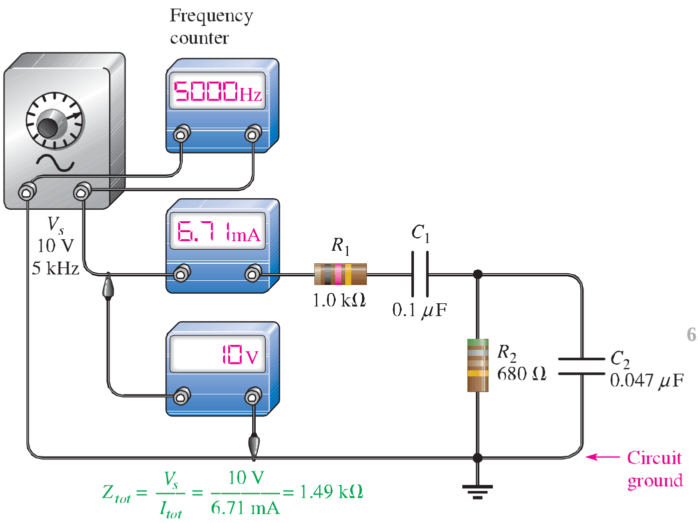
Frequency counter
5000 Hz
6.71 mA
10 V
V_s
10 V
5 kHz
R_1
1.0 kΩ
C_1
0.1 μF
R_2
680 Ω
C_2
0.047 μF
Circuit ground
$Z_{tot} = \frac{V_s}{I_{tot}} = \frac{10\text{ V}}{6.71\text{ mA}} = 1.49\text{ k}\Omega$

6238262 2013/12/10 12.27.66.5

FIGURE 10–36 Determining Z_{tot} by measurement of V_s and I_{tot}.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

DETERMINING *θ* To measure the phase angle, the source voltage and the total current must be displayed on an oscilloscope screen in the proper time relationship. Two basic types of scope probes are available to measure the quantities with an oscilloscope: the voltage probe and the current probe. The current probe is a convenient device, but it is often not as readily available as a voltage probe. We will confine our phase measurement technique to the use of voltage probes in conjunction with the oscilloscope. Although there are special isolation methods, a typical passive oscilloscope voltage probe has two points that are connected to the circuit: the probe tip and the ground lead. Thus, all voltage measurements must be referenced to ground.

Since only voltage probes are to be used, the total current cannot be measured directly. However, for phase measurement, the voltage across R_1 is in phase with the total current and can be used to establish the phase angle of the current.

Before proceeding with the actual phase measurement, there is a problem with displaying V_{R1}. If the scope probe is connected across the resistor, as indicated in Figure 10–37(a), the ground lead of the scope will short point *B* to ground, thus bypassing the rest of the components and effectively removing them from the circuit electrically, as illustrated in Figure 10–37(b) (assuming that the scope is not isolated from power line ground).

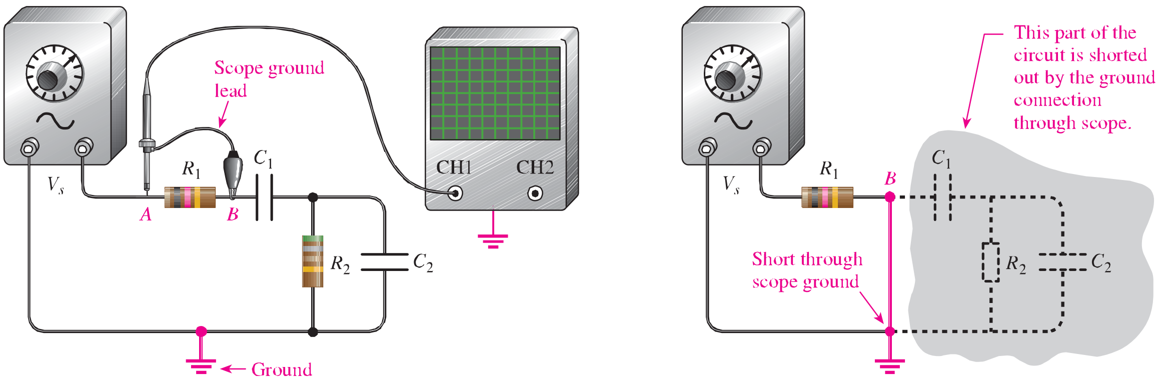
Scope ground lead
V_s
R_1
C_1
A
B
CH1
CH2
R_2
C_2
Ground
This part of the circuit is shorted out by the ground connection through scope.
V_s
R_1
B
C_1
R_2
C_2
Short through scope ground

(a) Ground lead on scope probe grounds point *B*.

(b) The effect of grounding point *B* is to short out the rest of the circuit.

FIGURE 10–37 Effects of measuring directly across a component when the instrument and the circuit are grounded. The scope ground has shorted out part of the circuit.

High-Frequency Differential Probing

Most oscilloscopes will allow you to combine two single-ended probes in order to make a measurement of an ungrounded component. This is a good method for low frequencies. For high-frequency work, the separate paths for each channel can lead to delay differences between the two signals, resulting in amplitude and timing errors. A better way to probe high-frequency and fast signals is to use a differential probe. A differential probe uses a differential amplifier at the probe tip to subtract the two signals, allowing one channel of the oscilloscope to make a measurement across an ungrounded component. The result is a more accurate measurement.

SYSTEM NOTE

To avoid this problem, you can switch the generator output terminals so that one end of R_1 is connected to the ground terminal, as shown in Figure 10–38(a). Now the scope can be connected across it to display V_{R1}, as indicated in Figure 10–38(b). The other probe is connected across the voltage source to display V_s as indicated. Now channel 1 of the scope has V_{R1} as an input, and channel 2 has V_s. The scope should be triggered from the source voltage (channel 2 in this case).

Before connecting the probes to the circuit, you should align the two horizontal lines (traces) so that they appear as a single line across the center of the scope screen. To do so, ground the probe tips and adjust the vertical position knobs to move the traces toward the

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

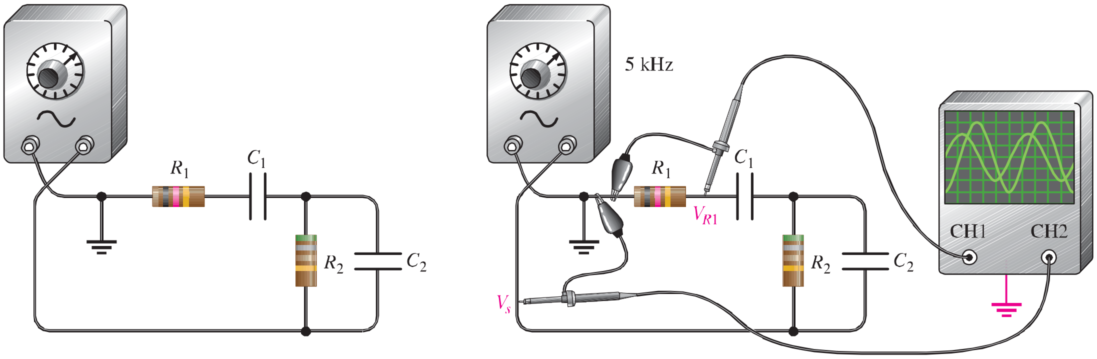
R_1
C_1
R_2
C_2
5 kHz
V_{R1}
V_s
CH1
CH2

(a) Ground repositioned so that one end of R_1 is grounded.

(b) The scope displays V_{R1} and V_s. V_{R1} represents the phase of the total current.

FIGURE 10–38 **Repositioning ground so that a direct voltage measurement can be made with respect to ground without shorting out part of the circuit.**

center line of the screen until they are superimposed. This procedure ensures that both waveforms have the same zero crossing so that an accurate phase measurement can be made.

Once you have stabilized the waveforms on the scope screen, you can measure the period of the source voltage. Next, use the Volts/Div controls to adjust the amplitudes of the waveforms until they both appear to have the same amplitude. Now, spread the waveforms horizontally by using the Sec/Div control to expand the distance between them. This horizontal distance represents the time between the two waveforms. The number of divisions between the waveforms along any horizontal lines times the Sec/Div setting is equal to the time between them, Δt. Also, you can use the cursors to determine Δt if your oscilloscope has this feature.

Once you have determined the period, T, and the time between the waveforms, Δt, you can calculate the phase angle with the following equation:

$$\theta = \left(\frac{\Delta t}{T}\right)360^\circ \qquad (10\text{–}23)$$

Figure 10–39 shows simulated screen displays for an oscilloscope in Multisim. In Figure 10–39(a), the waveforms are aligned and set to the same apparent amplitude by adjusting the fine Volts/Div control. The period of these waveforms is 200 μs. The Sec/Div control is adjusted to spread the waveforms out to read Δt more accurately. As shown in part (b), there are 3.0 divisions between the crossings on the center line. The Sec/Div control is set to 5.0 μs and there are 3.0 divisions between the waveforms.

$$\Delta t = 3.0 \text{ divisions} \times 5.0\ \mu\text{s/division} = 15\ \mu\text{s}$$

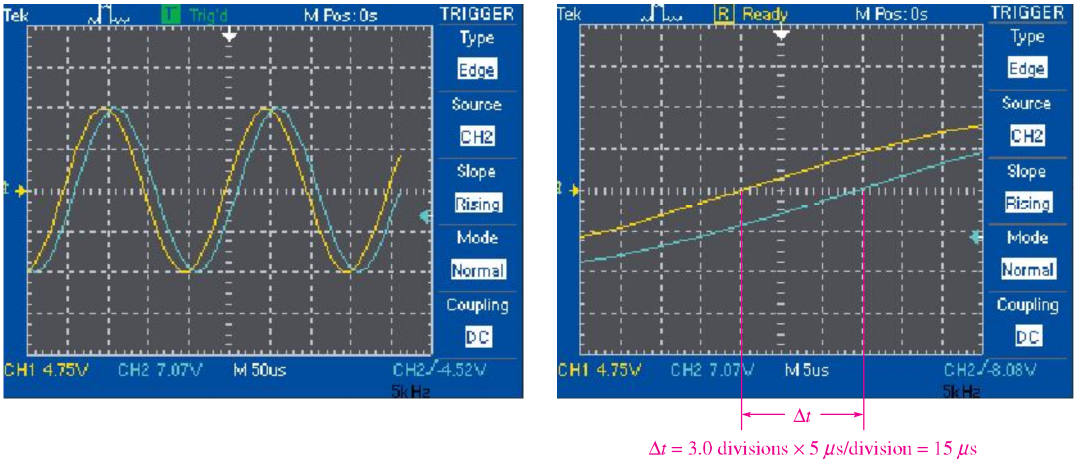
Tek
Trig'd
M Pos: 0s
TRIGGER
Type
Edge
Source
CH2
Slope
Rising
Mode
Normal
Coupling
DC
CH1 4.75V
CH2 7.07V
M 50us
CH2 ∫ 4.52V
5k Hz
Tek
Ready
M Pos: 0s
TRIGGER
Type
Edge
Source
CH2
Slope
Rising
Mode
Normal
Coupling
DC
CH1 4.75V
CH2 7.07V
M 5us
CH2 ∫ 8.08V
5k Hz
Δt

$\Delta t = 3.0 \text{ divisions} \times 5\ \mu\text{s/division} = 15\ \mu\text{s}$

(a)

(b)

FIGURE 10–39 **Determining the phase angle using an oscilloscope.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The phase angle is

$$\theta = \left(\frac{\Delta t}{T}\right)360^\circ = \left(\frac{15\ \mu\text{s}}{200\ \mu\text{s}}\right)360^\circ = 27^\circ$$

SECTION 10–6 CHECKUP

1. Explain why a scope probe ground should be connected to the circuit ground.
2. What is the voltage across R_1 in Figure 10–34?
3. To what value should S_{probe} in Figure 10–35 be adjusted if $C_{in} = 15$ pF?

6238262 2013/12/10 12.27.66.5

10–7 POWER IN *RC* CIRCUITS

In a purely resistive ac circuit, all of the energy delivered by the source is dissipated in the form of heat by the resistance. In a purely capacitive ac circuit, all of the energy delivered by the source is stored by the capacitor during a portion of the voltage cycle and then returned to the source during another portion of the cycle so that there is no net energy conversion to heat. When there is both resistance and capacitance, some of the energy is alternately stored and returned by the capacitance and some is dissipated by the resistance. The amount of energy converted to heat is determined by the relative values of the resistance and the capacitive reactance.

After completing this section, you should be able to

- **Determine power in *RC* circuits**
  - **Explain true and reactive power**
  - **Draw the power triangle**
  - **Define *power factor***
  - **Explain apparent power**
  - **Calculate power in an *RC* circuit**

When the resistance is greater than the capacitive reactance in a series *RC* circuit, more of the total energy delivered by the source is dissipated by the resistor than is stored by the capacitor. Likewise, when the reactance is greater than the resistance, more of the total energy is stored and returned than is converted to heat.

The formulas for the power dissipated in a resistor, sometimes called *true power* (P_{true}), and the power in a capacitor, called *reactive power* (P_r), are restated here. The unit of true power is the watt (W), and the unit of reactive power is the VAR (volt-ampere reactive).

$$P_{\text{true}} = I_{tot}^2 R \qquad (10\text{–}24)$$

$$P_r = I_{tot}^2 X_C \qquad (10\text{–}25)$$

Power Triangle for *RC* Circuits

The generalized impedance phasor diagram for a series *RC* circuit is shown in Figure 10–40(a). A relationship for power can also be represented by a similar diagram because the respective magnitudes of the powers, P_{true} and P_r, differ from R and X_C by a factor of I_{tot}^2, as shown in Figure 10–40(b).

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

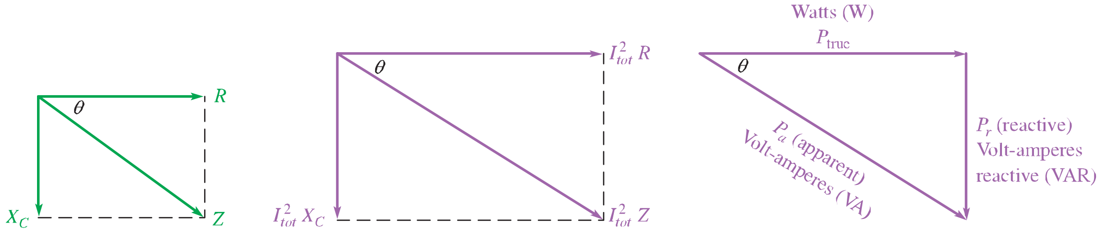

(a) Impedance phasors

(b) Impedance phasors are multiplied by I^2_{tot} to get power

(c) Power triangle

FIGURE 10–40 **Development of the power triangle for a series *RC* circuit.**

The resultant power, $I^2_{tot}Z$, represents the **apparent power, P_a**. At any instant in time, P_a is the total power that appears to be transferred between the source and the *RC* circuit. Part of the apparent power is true power, and part of it is reactive power. The unit of apparent power is the volt-ampere (VA). The expression for apparent power is

$$P_a = I^2_{tot}Z \tag{10–26}$$

The diagram in Figure 10–40(b) can be rearranged in the form of a right triangle as shown in Figure 10–40(c), which is called the *power triangle*. Using the rules of trigonometry, P_{true} can be expressed as

$$P_{true} = P_a \cos\theta$$

Since P_a equals $I^2_{tot}Z$ or V_sI_{tot}, the equation for true power can be written as

$$P_{true} = V_sI_{tot}\cos\theta \tag{10–27}$$

where V_s is the source voltage and I_{tot} is the total current.

For the case of a purely resistive circuit, $\theta = 0°$ and $\cos 0° = 1$, so P_{true} equals V_sI_{tot}. For the case of a purely capacitive circuit, $\theta = 90°$ and $\cos 90° = 0$, so P_{true} is zero. As you already know, there is no power dissipated in an ideal capacitor.

Power Factor

The term *cos θ* is called the **power factor** and is stated as

$$PF = \cos\theta \tag{10–28}$$

As the phase angle between the source voltage and the total current increases, the power factor decreases, indicating an increasingly reactive circuit. The smaller the power factor, the smaller the power dissipation.

The power factor can vary from 0 for a purely reactive circuit to 1 for a purely resistive circuit. In an *RC* circuit, the power factor is referred to as a leading power factor because the current leads the voltage.

EXAMPLE 10–14

Determine the power factor and the true power in the *RC* circuit of Figure 10–41.

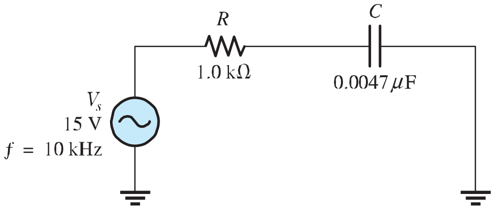

FIGURE 10–41

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

Calculate the capacitive reactance and phase angle.

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi(10\text{ kHz})(0.0047\ \mu\text{F})} = 3.39\text{ k}\Omega$$

$$\theta = \tan^{-1}\left(\frac{X_C}{R}\right) = \tan^{-1}\left(\frac{3.39\text{ k}\Omega}{1.0\text{ k}\Omega}\right) = 73.6^\circ$$

The power factor is

$$PF = \cos\theta = \cos(73.6^\circ) = \mathbf{0.282}$$

The impedance is

$$Z = \sqrt{R^2 + X_C^2} = \sqrt{(1.0\text{ k}\Omega)^2 + (3.39\text{ k}\Omega)^2} = 3.53\text{ k}\Omega$$

Therefore, the current is

$$I = \frac{V_s}{Z} = \frac{15\text{ V}}{3.53\text{ k}\Omega} = 4.25\text{ mA}$$

The true power is

$$P_{\text{true}} = V_s I\cos\theta = (15\text{ V})(4.25\text{ mA})(0.282) = \mathbf{18.0\ mW}$$

RELATED PROBLEM

What is the power factor if *f* is reduced by half in Figure 10–41?

MULTISIM

Open Multisim file E10–14. Measure the current and compare to the calculated value. Measure the voltages across *R* and *C* at 10 kHz, 5 kHz, and 20 kHz. Explain your observations.

Significance of Apparent Power

Apparent power is the power that *appears* to be transferred between the source and the load, and it consists of two components: a true power component and a reactive power component.

In all electrical and electronic systems, it is the true power that does the work. The reactive power is simply shuttled back and forth between the source and the load. Ideally, in terms of performing useful work, all of the power transferred to the load should be true power and none of it reactive power. However, in most practical situations the load has some reactance associated with it, and therefore you must deal with both power components.

For any reactive load, there are two components of the total current: the resistive component and the reactive component. If you consider only the true power (watts) in a load, you are dealing with only a portion of the total current that the load demands from a source. In order to have a realistic picture of the actual current that a load will draw, you must consider apparent power (in VA).

A source such as an ac generator can provide current to a load up to some maximum value. *If the load draws more than this maximum value, the source can be damaged.* Figure 10–42(a) shows a 120 V generator that can deliver a maximum current of 5 A to a load. Assume that the generator is rated at 600 W and is connected to a purely resistive load of 24 Ω (power factor of 1). The ammeter shows that the current is 5 A, and the wattmeter indicates that the power is 600 W. The generator has no problem under these conditions, although it is operating at maximum current and power.

Now, consider what happens if the load is changed to a reactive one with an impedance of 18 Ω and a power factor of 0.6, as indicated in Figure 10–42(b). The current is 120 V/18 Ω = 6.67 A, which exceeds the maximum. Even though the wattmeter reads 480 W, which is less than the power rating of the generator, the excessive current probably will cause damage. This example shows that a true power rating can be deceiving and is inappropriate for ac sources. This particular ac generator should be rated at 600 VA, a rating the manufacturer would use, rather than 600 W.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

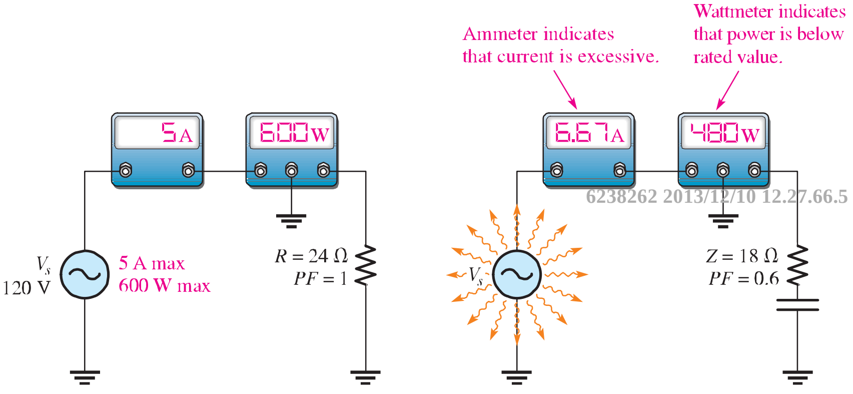
Ammeter indicates that current is excessive.
Wattmeter indicates that power is below rated value.
5 A
600 W
6.67 A
480 W
V_s 120 V
5 A max
600 W max
$R = 24\ \Omega$
$PF = 1$
V_s
$Z = 18\ \Omega$
$PF = 0.6$

6238262 2013/12/10 12.27.66.5

(a) Generator operating at its limits with a resistive load.

(b) Generator is in danger of internal damage due to excess current, even though the wattmeter indicates that the power is below the maximum wattage rating.

FIGURE 10–42 **The wattage rating of a source is inappropriate when the load is reactive. The rating should be in VA rather than in watts.**

EXAMPLE 10–15

For the circuit in Figure 10–43, find the true power, the reactive power, and the apparent power. X_C has been determined to be 2.0 kΩ.

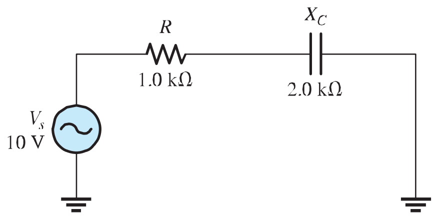
R
1.0 kΩ
X_C
2.0 kΩ
V_s
10 V

FIGURE 10–43

SOLUTION

First find the total impedance so that you can calculate the current.

$$Z_{tot} = \sqrt{R^2 + X_C^2} = \sqrt{(1.0\ \text{k}\Omega)^2 + (2.0\ \text{k}\Omega)^2} = 2.24\ \text{k}\Omega$$

$$I = \frac{V_s}{Z} = \frac{10\ \text{V}}{2.24\ \text{k}\Omega} = 4.46\ \text{mA}$$

The phase angle, θ, is

$$\theta = \tan^{-1}\left(\frac{X_C}{R}\right) = \tan^{-1}\left(\frac{2.0\ \text{k}\Omega}{1.0\ \text{k}\Omega}\right) = 63.4^\circ$$

The true power is

$$P_{\text{true}} = V_s I \cos\theta = (10\ \text{V})(4.46\ \text{mA})\cos(63.4^\circ) = \mathbf{20\ mW}$$

Note that the same result is realized if you use the formula $P_{\text{true}} = I^2R$.

The reactive power is

$$P_r = I^2X_C = (4.46\ \text{mA})^2(2.0\ \text{k}\Omega) = \mathbf{39.8\ mVAR}$$

The apparent power is

$$P_a = I^2Z = (4.46\ \text{mA})^2(2.24\ \text{k}\Omega) = \mathbf{44.6\ mVA}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The apparent power is also the phasor sum of P_{true} and P_r.

$$P_a = \sqrt{P_{\text{true}}^2 + P_r^2} = 44.6\text{ mVA}$$

RELATED PROBLEM

What is the true power in Figure 10–43 if $X_C = 10\text{ k}\Omega$?

SECTION 10–7 CHECKUP

1. To which component in an *RC* circuit is the power dissipation due?
2. If the phase angle, θ, is 45°, what is the power factor?
3. A certain series *RC* circuit has the following parameter values: $R = 330\ \Omega$, $X_C = 460\ \Omega$, and $I = 2\ A$. Determine the true power, the reactive power, and the apparent power.

10–8 BASIC APPLICATIONS

RC **circuits are found in a variety of applications, often as part of a more complex circuit. Three applications are phase shift circuits applied to an oscillator, frequency-selective circuits (filters), and ac coupling.**

After completing this section, you should be able to

- **Describe some basic *RC* applications**
  - **Discuss how the *RC* circuit is used in an oscillator**
  - **Discuss how the *RC* circuit operates as a filter**
  - **Discuss ac coupling**

The Phase Shift Oscillator

As you know, a series *RC* circuit will shift the phase of the output voltage by an amount that depends on the values of *R* and *C* and the frequency of the signal. This ability to shift phase depending on frequency is vital in certain feedback oscillator circuits. An oscillator is a circuit that generates a periodic waveform and is an important circuit for many electronic systems. You will study oscillators in a devices course, so the focus here is on the application of *RC* circuits for shifting phase. The requirement is that a fraction of the output of the oscillator is returned to the input (called *feedback*) in the proper phase to reinforce the input and sustain oscillations. Generally, the requirement is to feed back the signal with a total of 180° of phase shift.

A single *RC* circuit is limited to phase shifts that are smaller than 90°. The basic *RC* lag circuit discussed in Section 10–3 can be "stacked" to form a complex *RC* network, as shown in Figure 10–44, which shows a specific circuit called a *phase shift oscillator*. The phase shift oscillator typically uses three equal-component *RC* circuits that produce the required 180° phase shift at a certain frequency, which will be the frequency at which the oscillator works. The output of the amplifier is phase shifted by the *RC* network and returned to the input of the amplifier, which provides sufficient gain to maintain oscillations.

The process of putting several *RC* circuits together results in a loading effect, so the overall phase shift is not the same as simply adding the phase shifts of the individual *RC* circuits. The detailed calculation for this circuit involves a lot of tedious phasor math, but

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

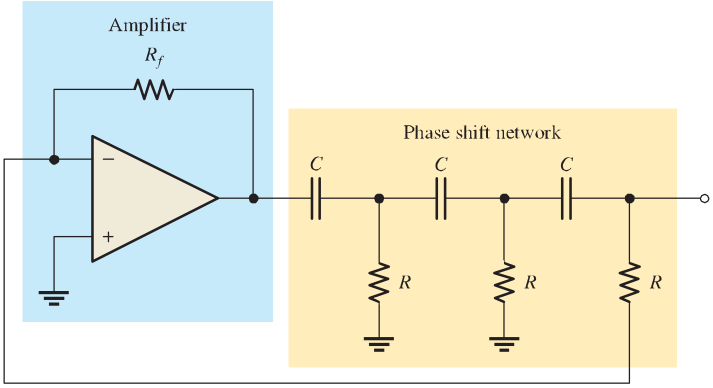

FIGURE 10–44 **Phase shift oscillator.**

the result is simple. With equal components, the frequency at which a 180° phase shift occurs is given by the following equation:

$$f_r = \frac{1}{2\pi\sqrt{6}RC}$$

It also turns out that the *RC* network attenuates (reduces) the signal from the amplifier by a factor of 29; the amplifier must make up for this attenuation by having a gain of −29 (the minus sign takes into account the phase shift).

EXAMPLE 10–16

In Figure 10–45, calculate the output frequency.

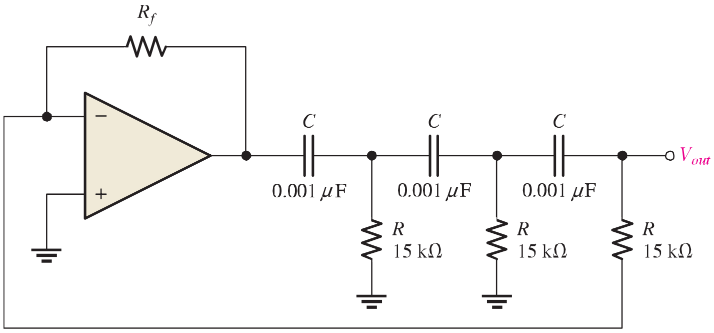

FIGURE 10–45

SOLUTION

$$f_r = \frac{1}{2\pi\sqrt{6}RC} = \frac{1}{2\pi\sqrt{6}(15\text{ k}\Omega)(0.001\ \mu\text{F})} = 4.33\text{ kHz}$$

RELATED PROBLEM

If all of the capacitors are changed to 0.0027 μF, what is the oscillator frequency?

The *RC* Circuit as a Filter

Frequency-selective circuits (filters) permit signals of certain frequencies to pass from the input to the output while blocking all others. That is, ideally all frequencies but the selected ones are filtered out. Frequency-selective circuits are important in many systems, particularly communication systems.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Series *RC* circuits exhibit a frequency-selective characteristic. There are two types. The first one that we examine is a **low-pass circuit** that is realized by taking the output across the capacitor, just as in a lag circuit. The second type is a **high-pass circuit**, which is implemented by taking the output across the resistor, as in a lead circuit. In practice, *RC* circuits are used in conjunction with operational amplifiers to create active filters, which are much more effective than passive *RC* circuits.

6238262 2013/12/10 12:27:665

LOW-PASS FILTER You have seen what happens to the phase angle and the output voltage in a series *RC* lag circuit. In terms of the filtering action of the series *RC* circuit, the variation in the magnitude of the output voltage as a function of frequency is important.

Figure 10–46 shows the filtering action of a series *RC* circuit using a specific series of measurements in which the frequency starts at 100 Hz and is increased in increments up to 20 kHz. At each value of frequency, the output voltage is measured. As you can see, the capacitive reactance decreases as frequency increases, thus dropping less voltage across the capacitor while the input voltage is held at a constant 10 V throughout each step. Table 10–1 summarizes the variation of the circuit parameters with frequency.

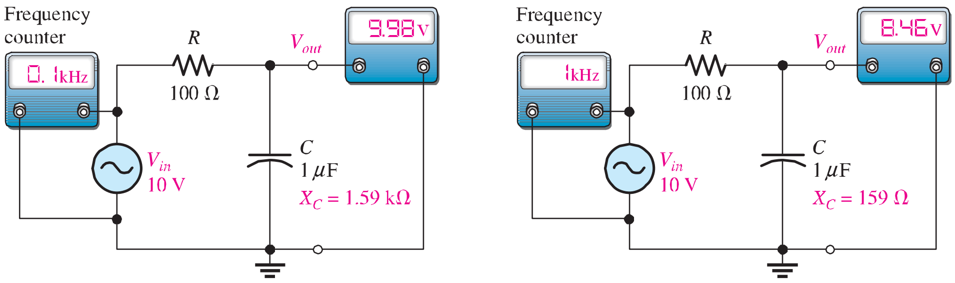
Frequency counter
0.1kHz
R
100 Ω
V_{out}
9.98V
V_{in}
10 V
C
1 μF
X_C = 1.59 kΩ
Frequency counter
1kHz
R
100 Ω
V_{out}
8.46V
V_{in}
10 V
C
1 μF
X_C = 159 Ω

(a) $f = 0.1$ kHz, $X_C = 1.59$ kΩ, $V_{out} = 9.98$ V (b) $f = 1$ kHz, $X_C = 159$ Ω, $V_{out} = 8.46$ V

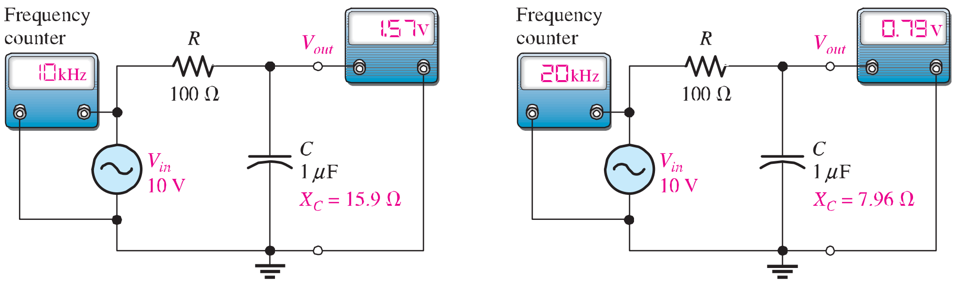
Frequency counter
10kHz
R
100 Ω
V_{out}
1.57V
V_{in}
10 V
C
1 μF
X_C = 15.9 Ω
Frequency counter
20kHz
R
100 Ω
V_{out}
0.79V
V_{in}
10 V
C
1 μF
X_C = 7.96 Ω

(c) $f = 10$ kHz, $X_C = 15.9$ Ω, $V_{out} = 1.57$ V (d) $f = 20$ kHz, $X_C = 7.96$ Ω, $V_{out} = 0.79$ V

FIGURE 10–46 **Example of low-pass filtering action. As frequency increases, V_{out} decreases.**

TABLE 10–1

f (kHz)	X_C (Ω)	Z_{tot} (Ω)	I (mA)	V_{out} (V)
0.1	1,590	1,590	6.29	9.98
1	159	188	53.2	8.46
10	15.9	101	99.0	1.57
20	7.96	100	100	0.79

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The **frequency response** of the low-pass *RC* circuit in Figure 10–46 is shown in Figure 10–47, where the measured values are plotted on a graph of V_{out} versus f, and a smooth curve is drawn connecting the points. This graph, called a *response curve,* shows that the output voltage is greater at the lower frequencies and decreases as the frequency increases. The frequency scale is logarithmic.

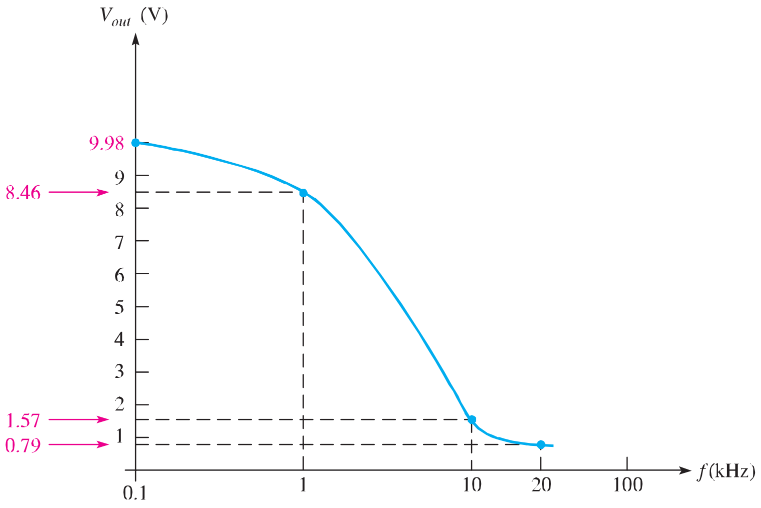

FIGURE 10–47 Frequency response curve for the low-pass *RC* circuit in Figure 10–46.

HIGH-PASS FILTER To illustrate *RC* high-pass filtering action, Figure 10–48 shows a series of specific measurements. The frequency starts at 10 Hz and is increased in increments up to 10 kHz. As you can see, the capacitive reactance decreases as the frequency increases, thus causing more of the total input voltage to be dropped across the resistor. Table 10–2 summarizes the variation of circuit parameters with frequency.

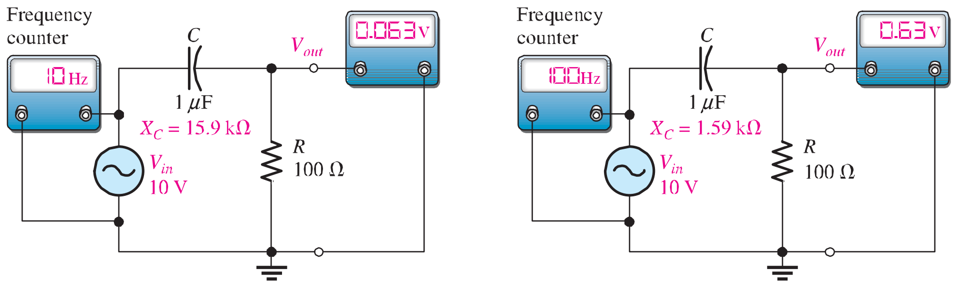
Frequency counter
10 Hz
C
1 μF
X_C = 15.9 kΩ
V_{in}
10 V
R
100 Ω
V_{out}
0.063 V
Frequency counter
100 Hz
C
1 μF
X_C = 1.59 kΩ
V_{in}
10 V
R
100 Ω
V_{out}
0.63 V

(a) $f = 10$ Hz, $X_C = 15.9$ kΩ, $V_{out} = 0.063$ V

(b) $f = 100$ Hz, $X_C = 1.59$ kΩ, $V_{out} = 0.63$ V

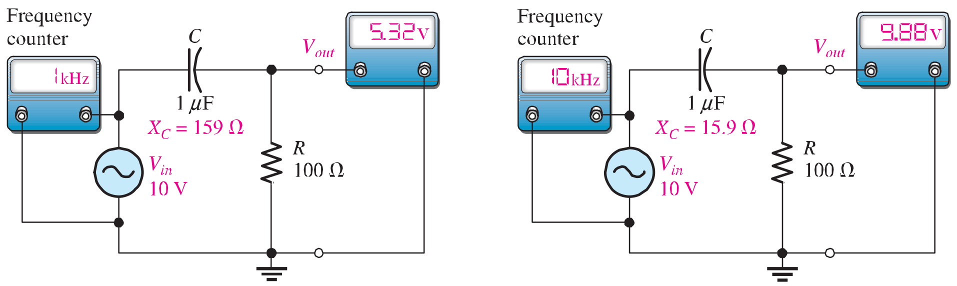
Frequency counter
1 kHz
C
1 μF
X_C = 159 Ω
V_{in}
10 V
R
100 Ω
V_{out}
5.32 V
Frequency counter
10 kHz
C
1 μF
X_C = 15.9 Ω
V_{in}
10 V
R
100 Ω
V_{out}
9.88 V

(c) $f = 1$ kHz, $X_C = 159$ Ω, $V_{out} = 5.32$ V

(d) $f = 10$ kHz, $X_C = 15.9$ Ω, $V_{out} = 9.88$ V

FIGURE 10–48 Example of high-pass filtering action. As frequency increases, V_{out} increases.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

TABLE 10–2

f (kHz)	X_C (Ω)	Z_{tot} (Ω)	I (mA)	V_{out} (V)
.01	15,900	15,900	0.629	0.063
.1	1590	1593	6.28	0.63
1	159	188	53.2	5.32
10	15.9	101	98.8	9.88

In Figure 10–49, the measured values for the high-pass *RC* circuit shown in Figure 10–48 have been plotted to produce a frequency response curve for this circuit. As you can see, the output voltage is greater at the higher frequencies and decreases as the frequency is reduced. The frequency scale is logarithmic.

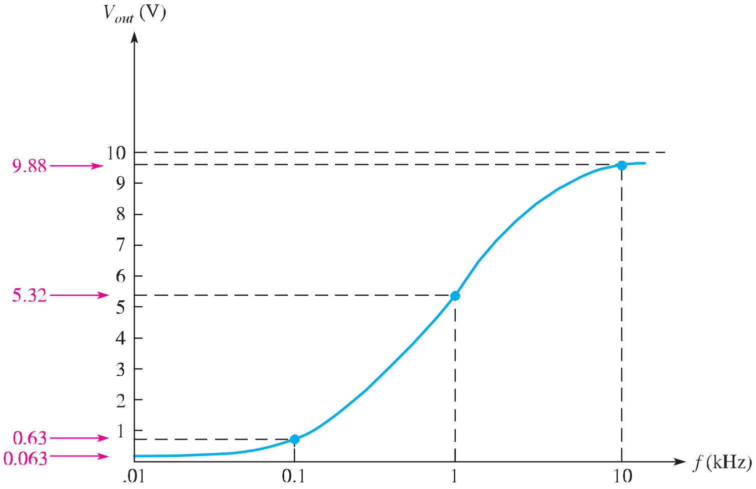

FIGURE 10–49 **Frequency response curve for the high-pass *RC* circuit in Figure 10–48.**

THE CUTOFF FREQUENCY AND THE BANDWIDTH OF AN *RC* CIRCUIT The frequency at which the capacitive reactance equals the resistance in a low-pass or high-pass *RC* circuit is called the **cutoff frequency** and is designated f_c. This condition is expressed as $1/(2\pi f_c C) = R$. Solving for f_c results in the following formula:

$$f_c = \frac{1}{2\pi RC} \tag{10–29}$$

At f_c, the output voltage of the *RC* circuit is 70.7% of its maximum value. It is standard practice to consider the cutoff frequency as the limit of a circuit's performance in terms of passing or rejecting frequencies. For example, in a high-pass circuit, all frequencies above f_c are considered to be passed from the input to the output, and all those below f_c are considered to be rejected. The reverse is true for a low-pass circuit.

The range of frequencies that is considered to be passed from the input to the output of a circuit is called the **bandwidth**. Figure 10–50 illustrates the bandwidth and the cutoff frequency for a low-pass circuit.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

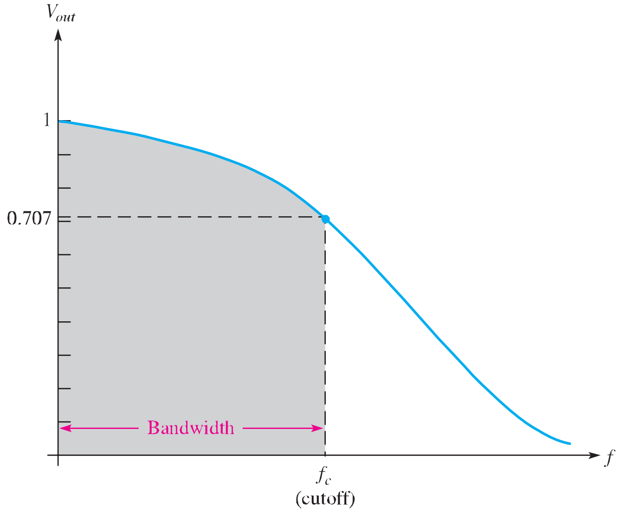

FIGURE 10–50 **Normalized general response curve of a low-pass *RC* circuit showing the cutoff frequency and the bandwidth.**

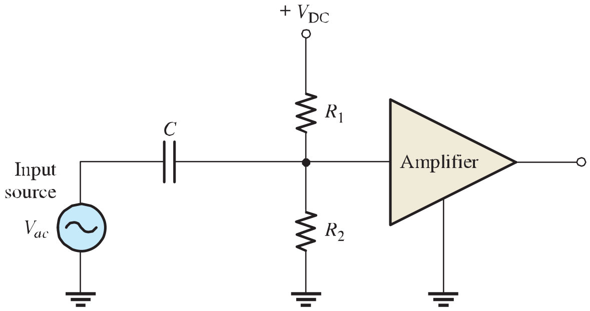

FIGURE 10–51 **Amplifier bias and signal-coupling circuit.**

Coupling an AC Signal into a DC Bias Circuit

Figure 10–51 shows an *RC* circuit that is used to create a dc voltage level with an ac voltage superimposed on it. This type of circuit is commonly found in amplifiers in which the dc voltage is required to **bias** the amplifier to the proper operating point and the signal voltage to be amplified is coupled through a capacitor and superimposed on the dc level. The capacitor prevents the low internal resistance of the signal source from affecting the dc bias voltage. This circuit is similar to the one in System Example 10–1 but uses a voltage-divider to set up the required bias voltage.

6238262 2013/12/10 12.27.66.5

In this type of application, a relatively large value of capacitance is selected so that for the frequencies to be amplified, the reactance is very small compared to the resistance of the bias network. When the reactance is very small (ideally zero), there is practically no phase shift or signal voltage dropped across the capacitor. Therefore, all of the signal voltage passes from the source to the input to the amplifier.

Figure 10–52 illustrates the application of the superposition principle to the circuit in Figure 10–51. In part (a), the ac source has been effectively removed from the circuit and

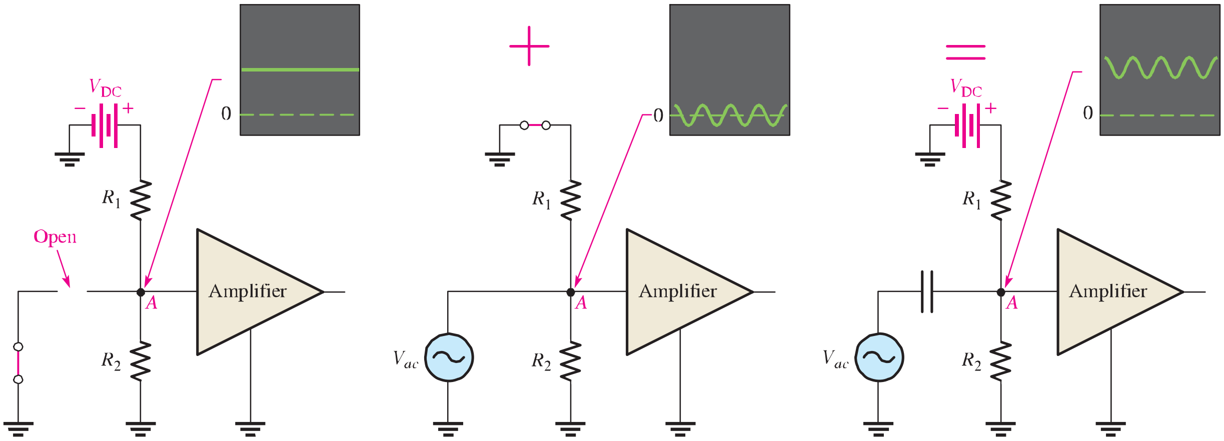

(a) dc equivalent: ac source replaced by short. C is open to dc. R_1 and R_2 act as dc voltage divider.

(b) ac equivalent: dc source is replaced by short. C is short to ac. All of V_{ac} is coupled to point A.

(c) dc + ac: Voltages are superimposed at point A.

FIGURE 10–52 **The superposition of dc and ac voltages in an *RC* bias and coupling circuit.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

replaced with a short to represent its ideal internal resistance. Since C is open to dc, the voltage at point A is determined by the voltage-divider action of R_1 and R_2 and the dc voltage source.

In Figure 10–52(b), the dc source has been effectively removed from the circuit and replaced with a short to represent its ideal internal resistance. Since C appears as a short at the frequency of the ac, the signal voltage is coupled directly to point A and appears across the parallel combination of R_1 and R_2. Figure 10–52(c) illustrates that the combined effect of the superposition of the dc and the ac voltages results in the signal voltage "riding" on the dc level.

SECTION 10–8 CHECKUP

1. In a phase shift oscillator, the *RC* circuit must provide a total phase shift of how many degrees?
2. To achieve a low-pass characteristic with a series *RC* circuit, across which component is the output taken?

10–9 TROUBLESHOOTING

Typical component failures or degradation have an effect on the response of basic *RC* circuits. The APM (analysis, planning, and measurement) method of troubleshooting can be used to locate problems in a circuit.

After completing this section, you should be able to

- **Troubleshoot *RC* circuits**
  - **Find an open resistor**
  - **Find an open capacitor**
  - **Find a shorted capacitor**
  - **Find a leaky capacitor**

EFFECT OF AN OPEN RESISTOR It is easy to see how an open resistor affects the operation of a basic series *RC* circuit, as shown in Figure 10–53. Obviously, there is no path for current, so the capacitor voltage remains at zero and the total voltage, V_s, appears across the open resistor.

EFFECT OF AN OPEN CAPACITOR When the capacitor is open, there is no current; thus, the resistor voltage drop is zero. The total source voltage appears across the open capacitor, as shown in Figure 10–54.

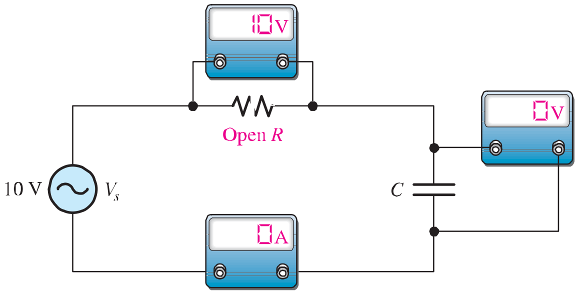

FIGURE 10–53 **Effect of an open resistor.**

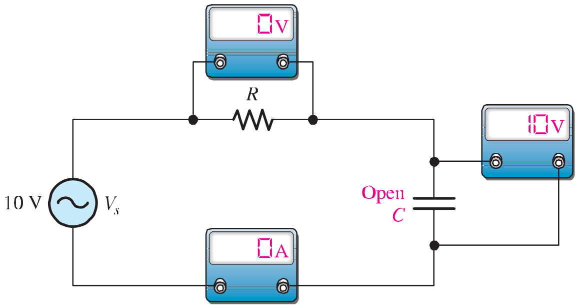

FIGURE 10–54 **Effect of an open capacitor.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EFFECT OF A SHORTED CAPACITOR When a capacitor shorts out, the voltage across it is zero, the current equals V_s/R, and the total voltage appears across the resistor, as shown in Figure 10–55.

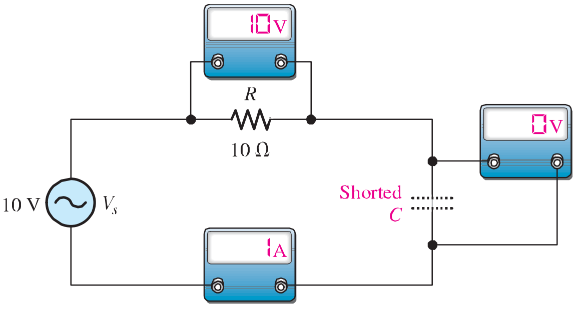

FIGURE 10–55 **Effect of a shorted capacitor.**

EFFECT OF A LEAKY CAPACITOR When a capacitor exhibits a high leakage current, the leakage resistance effectively appears in parallel with the capacitor, as shown in Figure 10–56(a). When the leakage resistance is comparable in value to the circuit resistance, R, the circuit response is drastically affected. The circuit, looking from the capacitor toward the source, can be thevenized, as shown in Figure 10–56(b). The Thevenin equivalent resistance is R in parallel with R_{leak} (the source appears as a short), and the Thevenin equivalent voltage is determined by the voltage-divider action of R and R_{leak}.

$$R_{th} = R \| R_{leak} = \frac{RR_{leak}}{R + R_{leak}}$$

$$V_{th} = \left(\frac{R_{leak}}{R + R_{leak}}\right)V_{in}$$

As you can see, the voltage to which the capacitor will charge is reduced since $V_{th} < V_{in}$. Also, the current is increased. The Thevenin equivalent circuit is shown in Figure 10–56(c).

HANDS ON TIP
Some multimeters have a relatively low frequency response of 1 kHz or less, while others can measure voltages or currents with frequencies up to about 2 MHz. Always check to make sure that your meter is capable of making accurate measurements at the particular frequency at which you are working.

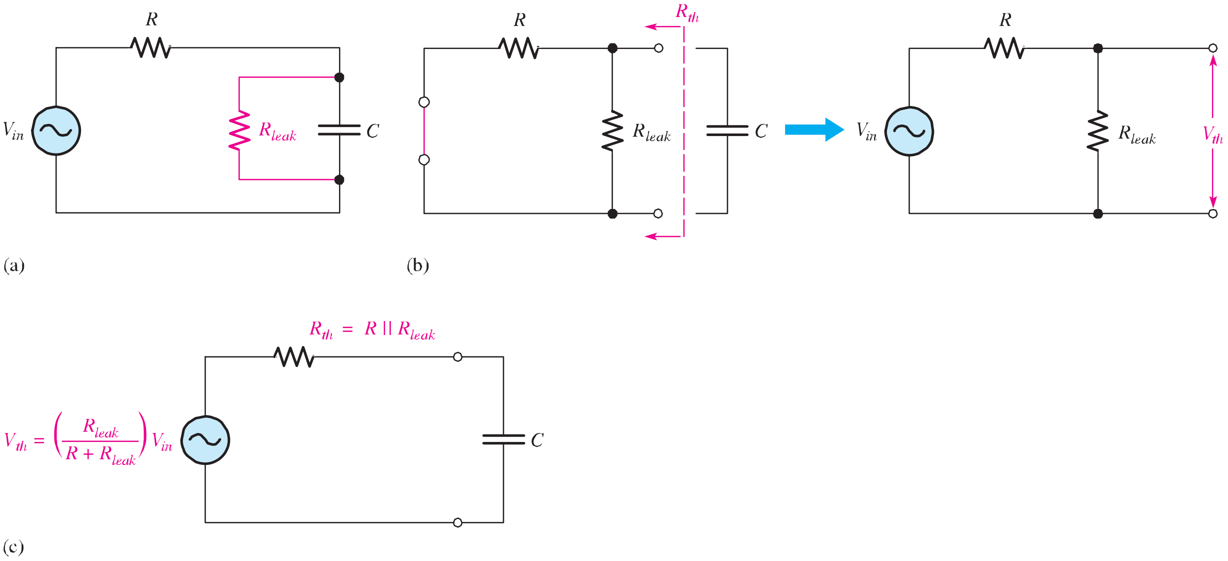

(a) (b) (c)

FIGURE 10–56 **Effect of a leaky capacitor.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 10–17

6238262 2013/12/10 12.27.66.5

Assume that the capacitor in Figure 10–57 is degraded to a point where its leakage resistance is 10 kΩ. Determine the phase shift from input to output and the output voltage under the degraded condition.

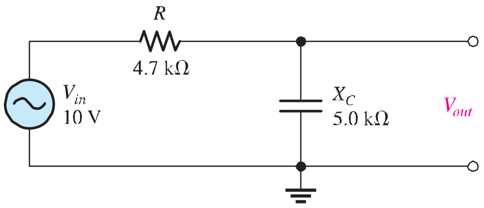

FIGURE 10–57

SOLUTION

The effective circuit resistance is

$$R_{th} = \frac{RR_{leak}}{R + R_{leak}} = \frac{(4.7\text{ k}\Omega)(10\text{ k}\Omega)}{14.7\text{ k}\Omega} = 3.2\text{ k}\Omega$$

The phase lag is

$$\phi = 90^\circ - \tan^{-1}\left(\frac{X_C}{R_{th}}\right) = 90^\circ - \tan^{-1}\left(\frac{5.0\text{ k}\Omega}{3.2\text{ k}\Omega}\right) = \mathbf{32.6^\circ}$$

To determine the output voltage, first calculate the Thevenin equivalent voltage.

$$V_{th} = \left(\frac{R_{leak}}{R + R_{leak}}\right)V_{in} = \left(\frac{10\text{ k}\Omega}{14.7\text{ k}\Omega}\right)10\text{ V} = 6.80\text{ V}$$

$$V_{out} = \left(\frac{X_C}{\sqrt{R_{th}^2 + X_C^2}}\right)V_{th} = \left(\frac{5.0\text{ k}\Omega}{\sqrt{(3.2\text{ k}\Omega)^2 + (5.0\text{ k}\Omega)^2}}\right)6.8\text{ V} = \mathbf{5.73\text{ V}}$$

RELATED PROBLEM

What would the output voltage be if the capacitor were not leaky?

Other Troubleshooting Considerations

Many times, the failure of a circuit to work properly is not the result of a faulty component. A loose wire, a bad contact, or a poor solder joint can cause an open circuit. A short can be caused by a wire clipping or solder splash. Things as simple as not plugging in a power supply or a function generator happen more often than you might think. Wrong values in a circuit (such as an incorrect resistor value), the function generator set at the wrong frequency, or the wrong output connected to the circuit can cause improper operation.

When you have problems with a circuit, always check to make sure that the instruments are properly connected to the circuits and to a power outlet. Also, look for obvious things such as a broken or loose contact, a connector that is not completely plugged in, or a piece of wire or a solder bridge that could be shorting something out.

The point is that you should consider all possibilities, not just faulty components, when a circuit is not working properly. The following example illustrates this approach with a simple circuit using the APM (analysis, planning, and measurement) method.

HANDS ON TIP

When connecting protoboard circuits, always consistently use standard colors for universal connections such as signals, power supply voltage, and ground. For example, you can use green wires for signals, red wires for power supply voltage, and black wires for ground connections. This helps you identify wiring during connecting and troubleshooting.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 10–18

The circuit represented by the schematic in Figure 10–58 has no output voltage, which is the voltage across the capacitor. You expect to see about 7.4 V at the output. The circuit is physically constructed on a protoboard. Use your troubleshooting skills to find the problem.

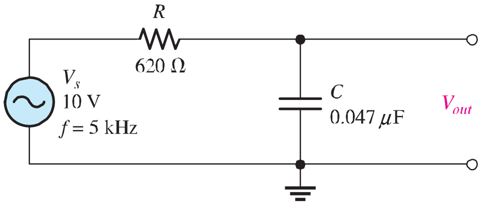

FIGURE 10–58

SOLUTION

Apply the APM method to this troubleshooting problem.

Analysis: First think of the possible causes for the circuit to have no output voltage.

1. There is no source voltage or the frequency is so high that the capacitive reactance is almost zero.
2. There is a short between the output terminals. Either the capacitor could be internally shorted, or there could be some physical short.
3. There is an open between the source and the output. This would prevent current and thus cause the output voltage to be zero. The resistor could be open, or the conductive path could be open due to a broken or loose connecting wire or a bad protoboard contact.
4. There is an incorrect component value. The resistor could be so large that the current and, therefore, the output voltage are negligible. The capacitor could be so large that its reactance at the input frequency is near zero.

Planning: You decide to make some visual checks for problems such as the function generator power cord not plugged in or the frequency set at an incorrect value. Also, broken leads, shorted leads, as well as an incorrect resistor color code or capacitor label often can be found visually. If nothing is discovered after a visual check, then you will make voltage measurements to track down the cause of the problem. You decide to use a digital oscilloscope and a DMM to make the measurements.

Measurement: Assume that you find that the function generator is plugged in and the frequency setting appears to be correct. Also, you find no visible opens or shorts during your visual check, and the component values are correct.

The first step in the measurement process is to check the voltage from the source with the scope. Assume a 10 V rms sine wave with a frequency of 5 kHz is observed at the circuit input as shown in Figure 10–59(a). The correct voltage is present, so *the first possible cause has been eliminated.*

Next, check for a shorted capacitor by disconnecting the source and placing a DMM (set on the ohmmeter function) across the capacitor. If the capacitor is good, an open will be indicated by an OL (overload) in the meter display after a short charging time. Assume the capacitor checks okay, as shown in Figure 10–59(b). *The second possible cause has been eliminated.*

Since the voltage has been "lost" somewhere between the input and the output, you must now look for the voltage. You reconnect the source and measure the voltage across the resistor with the DMM (set on the voltmeter function) from one resistor lead to the other. The voltage across the resistor is zero. This means there is no current, which indicates an open somewhere in the circuit.

Now, you begin tracing the circuit back toward the source looking for the voltage (you could also start from the source and work forward). You can use either the scope or the DMM, but decide to use the multimeter with one lead connected to ground and the other used to probe the circuit. As shown in Figure 10–59(c), the voltage on the right lead of the resistor, point ①, reads zero. Since you already have measured zero voltage across the resistor, the voltage on the left resistor lead at point ② must be zero as the meter indicates. Next, moving the meter probe to point ③, you read 10 V. You have found the voltage! Since there is zero volts on the left resistor lead, and there is 10 V at point ③, one of the two contacts in the protoboard hole into which the wire leads are inserted is bad. It could be that the small contacts were pushed in too far and were bent or broken so that the circuit lead does not make contact.

Move either or both the resistor lead and the wire to another hole in the same row. Assume that when the resistor lead is moved to the hole just above, you have voltage at the output of the circuit (across the capacitor).

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

FIGURE 10–59

RELATED PROBLEM

Suppose you had measured 10 V across the resistor before the capacitor was checked. What would this have indicated?

SECTION 10–9 CHECKUP

1. Describe the effect of a leaky capacitor on the dc response of an *RC* circuit.
2. In a series *RC* circuit, if all the applied voltage appears across the capacitor, what is the problem?
3. What faults can cause 0 V across the capacitor in a series *RC* circuit?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SUMMARY

- When a sinusoidal voltage is applied to an *RC* circuit, the current and all the voltage drops are also sine waves.
- Total current in a series or parallel *RC* circuit always leads the source voltage.
- The resistor voltage is always in phase with the current.
- The capacitor voltage always lags the current by 90°.
- In an *RC* circuit, the impedance is determined by both the resistance and the capacitive reactance combined.
- Impedance is expressed in units of ohms.
- The circuit phase angle is the angle between the total current and the source voltage.

6238262 2013/12/10 12.27.66.5

- The impedance of a series *RC* circuit varies inversely with frequency.
- The phase angle (θ) of a series *RC* circuit varies inversely with frequency.
- In an *RC* lag circuit, the output voltage lags the input voltage.
- In an *RC* lead circuit, the output voltage leads the input voltage.
- The impedance of a circuit can be determined by measuring the source voltage and the total current and then applying Ohm's law.
- In an *RC* circuit, part of the power is resistive and part reactive.
- The phasor combination of resistive power (true power) and reactive power is called *apparent power*.
- Apparent power is expressed in volt-amperes (VA).
- The power factor (*PF*) indicates how much of the apparent power is true power.
- A power factor of 1 indicates a purely resistive circuit, and a power factor of 0 indicates a purely reactive circuit.
- In a circuit that exhibits frequency selectivity, certain frequencies are passed to the output while others are rejected.

KEY TERMS

Key terms and other bold terms in the chapter are defined in the end-of-book glossary.

Admittance (Y) A measure of the ability of a reactive circuit to permit current; the reciprocal of impedance. The unit is the siemens (S).

Apparent power (P_a) The phasor combination of resistive power (true power) and reactive power. The unit is the volt-ampere (VA).

Bandwidth The range of frequencies that is considered to be passed from the input to the output of a circuit.

Capacitive susceptance (B_C) The ability of a capacitor to permit current: the reciprocal of capacitive reactance. The unit is the siemens (S).

Cutoff frequency The frequency at which the output voltage of a filter is 70.7% of the maximum output voltage.

Frequency response In electric circuits, the variation in the output voltage (or current) over a specified range of frequencies.

Impedance (Z) The total opposition to sinusoidal current expressed in ohms.

Phase angle The angle between the source voltage and the total current in a reactive circuit.

Power factor The relationship between volt-amperes and true power or watts. Volt-amperes multiplied by the power factor equals true power.

***RC* lag circuit** A phase shift circuit in which the output voltage, taken across the capacitor, lags the input voltage by a specified angle.

***RC* lead circuit** A phase shift circuit in which the output voltage, taken across the resistor, leads the input voltage by a specified angle.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

KEY FORMULAS

(10–1)	$Z = \sqrt{R^2 + X_C^2}$	Series *RC* impedance
(10–2)	$\theta = \tan^{-1}\left(\frac{X_C}{R}\right)$	Series *RC* phase angle
(10–3)	$V = IZ$	Ohm's law
(10–4)	$I = \frac{V}{Z}$	Ohm's law
(10–5)	$Z = \frac{V}{I}$	Ohm's law
(10–6)	$V_s = \sqrt{V_R^2 + V_C^2}$	Total voltage in series *RC* circuits
(10–7)	$\theta = \tan^{-1}\left(\frac{V_C}{V_R}\right)$	Series *RC* phase angle
(10–8)	$\phi = 90° - \tan^{-1}\left(\frac{X_C}{R}\right)$	Phase angle of lag circuit
(10–9)	$V_{out} = \left(\frac{X_C}{\sqrt{R^2 + X_C^2}}\right)V_{in}$	Output voltage of lag circuit
(10–10)	$\phi = \tan^{-1}\left(\frac{X_C}{R}\right)$	Phase angle of lead circuit
(10–11)	$V_{out} = \left(\frac{R}{\sqrt{R^2 + X_C^2}}\right)V_{in}$	Output voltage of lead circuit
(10–12)	$Z = \frac{RX_C}{\sqrt{R^2 + X_C^2}}$	Parallel *RC* impedance
(10–13	$\theta = \tan^{-1}\left(\frac{R}{X_C}\right)$	Parallel *RC* phase angle
(10–14)	$G = \frac{1}{R}$	Conductance
(10–15)	$B_C = \frac{1}{X_C}$	Capacitive susceptance
(10–16)	$Y = \frac{1}{Z}$	Admittance
(10–17)	$Y_{tot} = \sqrt{G^2 + B_C^2}$	Total admittance
(10–18)	$V = \frac{I}{Y}$	Ohm's law
(10–19)	$I = VY$	Ohm's law
(10–20)	$Y = \frac{I}{V}$	Ohm's law
(10–21)	$I_{tot} = \sqrt{I_R^2 + I_C^2}$	Total current in parallel *RC* circuits
(10–22)	$\theta = \tan^{-1}\left(\frac{I_C}{I_R}\right)$	Parallel *RC* phase angle

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

(10–23)	$\theta = \left(\frac{\Delta t}{T}\right)360^\circ$	Phase angle using time measurements
(10–24)	$P_{\text{true}} = I_{tot}^2R$	True power (W)
(10–25)	$P_r = I_{tot}^2X_C$	Reactive power (VAR)
(10–26)	$P_a = I_{tot}^2Z$	Apparent power (VA)
(10–27)	$P_{\text{true}} = V_sI_{tot}\cos\theta$	True power
(10–28)	$PF = \cos\theta$	Power factor
(10–29)	$f_c = \frac{1}{2\pi RC}$	Cutoff frequency of an *RC* circuit

TRUE/FALSE QUIZ

Answers are at the end of the chapter.

1. In a series *RC* circuit, the impedance increases when frequency increases.
2. In a series *RC* lag circuit, the output voltage is taken across the resistor.
3. Admittance is the reciprocal of susceptance.
4. In a parallel *RC* circuit, as frequency is increased, the conductance is unchanged.
5. The phase angle of an *RC* circuit is measured between the source voltage and current.
6. The impedance of a parallel *RC* circuit can be found by applying phasor arithmetic to the product-over-sum rule.
7. If $X_C = R$, the phase shift in a series *RC* circuit is 45° with current leading voltage.
8. The power factor is equal to the tangent of the phase angle.
9. A purely resistive circuit has a power factor of 0.
10. Apparent power is measured in watts.

SELF-TEST

Answers are at the end of the chapter.

1. In a series *RC* circuit, the voltage across the resistance is
   (a) in phase with the source voltage **(b)** lagging the source voltage by 90°
   (c) in phase with the current **(d)** lagging the current by 90°
2. In a series *RC* circuit, the voltage across the capacitor is
   (a) in phase with the source voltage **(b)** lagging the resistor voltage by 90°
   (c) in phase with the current **(d)** lagging the source voltage by 90°
3. When the frequency of the voltage applied to a series *RC* circuit is increased, the impedance
   (a) increases **(b)** decreases **(c)** remains the same **(d)** doubles
4. When the frequency of the voltage applied to a series *RC* circuit is decreased, the phase angle
   (a) increases **(b)** decreases **(c)** remains the same **(d)** becomes erratic
5. In a series *RC* circuit when the frequency and the resistance are doubled, the impedance
   (a) doubles **(b)** is halved
   (c) is quadrupled **(d)** cannot be determined without values
6. In a series *RC* circuit, 10 V rms is measured across the resistor and 10 V rms is also measured across the capacitor. The rms source voltage is
   (a) 20 V **(b)** 14.14 V **(c)** 28.28 V **(d)** 10 V

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

7. The voltages in Problem 6 are measured at a certain frequency. To make the resistor voltage greater than the capacitor voltage, the frequency
   (a) must be increased
   (b) must be decreased
   (c) is held constant
   (d) has no effect

8. When $R = X_C$, the phase angle is
   (a) 0° **(b)** +90° **(c)** −90° **(d)** 45°

9. To decrease the phase angle below 45°, the following condition must exist:
   (a) $R = X_C$ **(b)** $R < X_C$ **(c)** $R > X_C$ **(d)** $R = 10X_C$

10. When the frequency of the source voltage is increased, the impedance of a parallel *RC* circuit
   (a) increases
   (b) decreases
   (c) does not change

11. In a parallel *RC* circuit, there is 1 A rms through the resistive branch and 1 A rms through the capacitive branch. The total rms current is
   (a) 1 A **(b)** 2 A **(c)** 2.28 A **(d)** 1.414 A

12. A power factor of 1 indicates that the circuit phase angle is
   (a) 90° **(b)** 45° **(c)** 180° **(d)** 0°

6238262 2013/12/10 12.27.66.5

13. For a certain load, the true power is 100 W and the reactive power is 100 VAR. The apparent power is
   (a) 200 VA **(b)** 100 VA **(c)** 141.4 VA **(d)** 141.4 W

14. Energy sources are normally rated in
   (a) watts
   (b) volt-amperes
   (c) volt-amperes reactive
   (d) none of these

15. If the bandwidth of a certain low-pass filter is 1 kHz, the cutoff frequency is
   (a) 0 Hz **(b)** 500 Hz **(c)** 2 kHz **(d)** 1000 Hz

TROUBLESHOOTING: SYMPTOM AND CAUSE

The purpose of these exercises is to help develop thought processes essential to troubleshooting. Answers are at the end of the chapter.

Determine the cause for each set of symptoms. Refer to Figure 10–60.

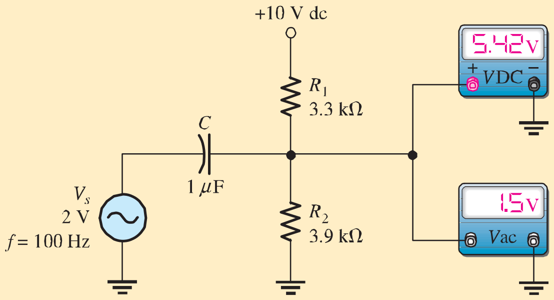

FIGURE 10–60 **The meters indicate the correct readings for this circuit.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

1. *Symptom:* The dc voltmeter reading is 0 V, and the ac voltmeter reading is 1.85 V.

   Cause:

   (a) C is shorted.
   (b) R_1 is open.
   (c) R_2 is open.

2. *Symptom:* The dc voltmeter reading is 5.42 V, and the ac voltmeter reading is 0 V.

   Cause:

   (a) C is shorted.
   (b) C is open.
   (c) A resistor is open.

3. *Symptom:* The dc voltmeter reading is approximately 0 V, and the ac voltmeter reading is 2 V.

   Cause:

   (a) C is shorted.
   (b) C is open.
   (c) R_1 is shorted.

4. *Symptom:* The dc voltmeter reading is 10 V, and the ac voltmeter reading is 0 V.

   Cause:

   (a) C is open.
   (b) C is shorted.
   (c) R_1 is shorted.

5. *Symptom:* The dc voltmeter reading is 10 V, and the ac voltmeter reading is 1.8 V.

   Cause:

   (a) R_1 is shorted.
   (b) R_2 is open.
   (c) C is shorted.

PROBLEMS

Answers to odd-numbered problems are at the end of the book.

BASIC PROBLEMS

SECTION 10–1 Sinusoidal Response of Series *RC* Circuits

1. An 8 kHz sinusoidal voltage is applied to a series RC circuit. What is the frequency of the voltage across the resistor? Across the capacitor?
2. What is the waveshape of the current in the circuit of Problem 1?

SECTION 10–2 Impedance and Phase Angle of Series *RC* Circuits

3. Find the impedance of each circuit in Figure 10–61.

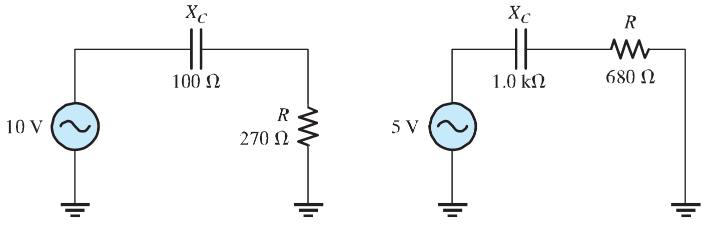

(a)

(b)

FIGURE 10–61

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

4. Determine the impedance and the phase angle in each circuit in Figure 10–62.

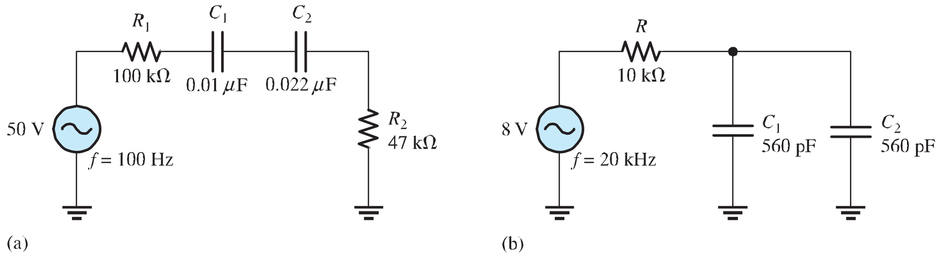
R_1
100 kΩ
C_1
0.01 μF
C_2
0.022 μF
50 V
f = 100 Hz
R_2
47 kΩ
(a)
R
10 kΩ
8 V
f = 20 kHz
C_1
560 pF
C_2
560 pF
(b)

FIGURE 10–62

5. For the circuit of Figure 10–63, determine the impedance for each of the following frequencies:
(a) 100 Hz **(b)** 500 Hz **(c)** 1.0 kHz **(d)** 2.5 kHz

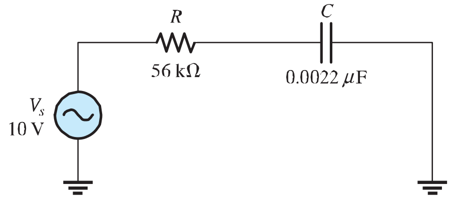
R
56 kΩ
C
0.0022 μF
V_s
10 V

FIGURE 10–63

6. Repeat Problem 5 for $C = 0.0047\ \mu\text{F}$.

SECTION 10–3 Analysis of Series *RC* Circuits

7. Calculate the total current in each circuit of Figure 10–61.

8. Repeat Problem 7 for the circuits in Figure 10–62.

9. For the circuit in Figure 10–64, draw the phasor diagram showing all voltages and the total current. Indicate the phase angles.

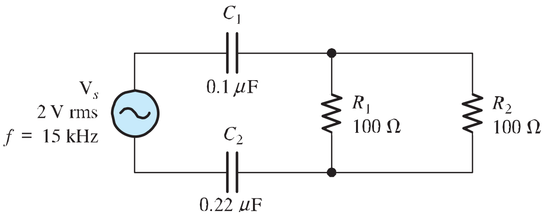
C_1
0.1 μF
V_s
2 V rms
f = 15 kHz
C_2
0.22 μF
R_1
100 Ω
R_2
100 Ω

FIGURE 10–64

10. For the circuit in Figure 10–65, determine the following:
(a) Z **(b)** I **(c)** V_R **(d)** V_C

11. To what value must the rheostat be set in Figure 10–66 to make the total current 10 mA? What is the resulting phase angle?

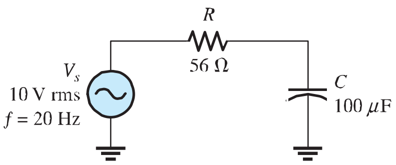
R
56 Ω
V_s
10 V rms
f = 20 Hz
C
100 μF

FIGURE 10–65

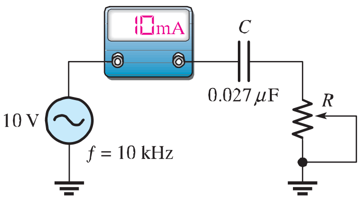
10mA
C
0.027 μF
R
10 V
f = 10 kHz

FIGURE 10–66

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

12. For the lag circuit in Figure 10–67, determine the phase lag between the input voltage and the output voltage for each of the following frequencies:
(a) 1 Hz **(b)** 100 Hz **(c)** 1.0 kHz **(d)** 10 kHz

13. Repeat Problem 12 for the lead circuit in Figure 10–68.

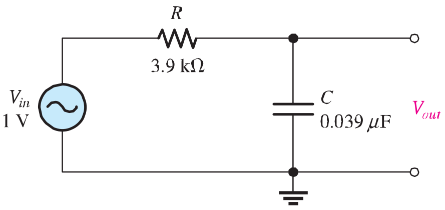
R
3.9 kΩ
V_{in}
1 V
C
0.039 μF
V_{out}

FIGURE 10–67

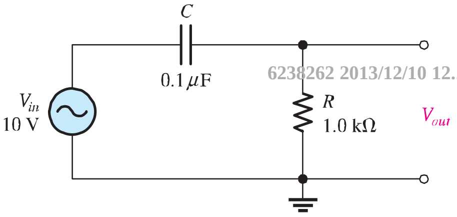
C
0.1 μF
V_{in}
10 V
R
1.0 kΩ
V_{out}

6238262 2013/12/10 12.27.66.5

FIGURE 10–68

SECTION 10–4 Impedance and Phase Angle of Parallel *RC* Circuits

14. Determine the impedance for the circuit in Figure 10–69.

15. Determine the impedance and the phase angle in Figure 10–70.

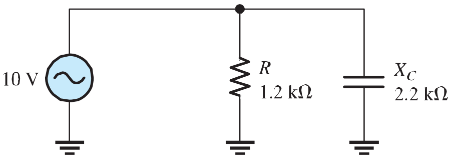
10 V
R
1.2 kΩ
X_C
2.2 kΩ

FIGURE 10–69

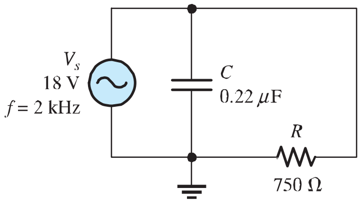
V_s
18 V
f = 2 kHz
C
0.22 μF
R
750 Ω

FIGURE 10–70

16. Repeat Problem 15 for the following frequencies:
(a) 1.5 kHz **(b)** 3.0 kHz **(c)** 5.0 kHz **(d)** 10 kHz

17. Determine the impedance and phase angle in Figure 10–71.

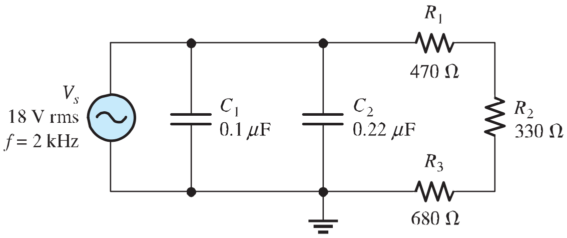
R_1
470 Ω
V_s
18 V rms
f = 2 kHz
C_1
0.1 μF
C_2
0.22 μF
R_2
330 Ω
R_3
680 Ω

FIGURE 10–71

SECTION 10–5 Analysis of Parallel *RC* Circuits

18. For the circuit in Figure 10–72, find all the currents and voltages.

19. For the parallel circuit in Figure 10–73, find each branch current and the total current. What is the phase angle between the source voltage and the total current?

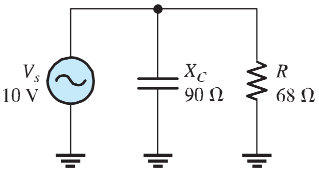
V_s
10 V
X_C
90 Ω
R
68 Ω

FIGURE 10–72

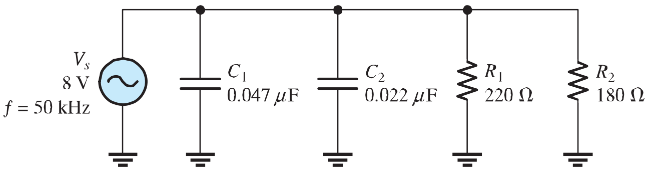
V_s
8 V
f = 50 kHz
C_1
0.047 μF
C_2
0.022 μF
R_1
220 Ω
R_2
180 Ω

FIGURE 10–73

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

20. For the circuit in Figure 10–74, determine the following:
(a) Z **(b)** I_R **(c)** I_C **(d)** I_{tot} **(e)** θ

21. Repeat Problem 20 for $R = 4.7\ \text{k}\Omega$, $C = 0.047\ \mu F$, and $f = 500$ Hz.

22. Find the total current for the circuit in Figure 10–75.

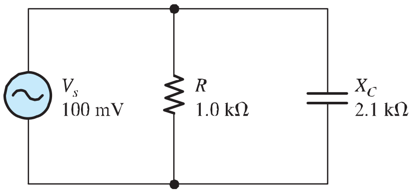

FIGURE 10–74

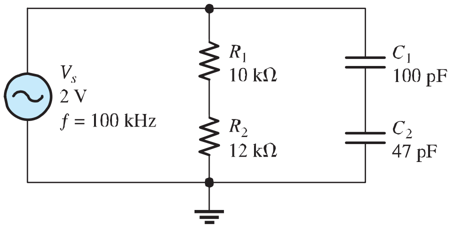

FIGURE 10–75

SECTION 10–6 Analysis of Series-Parallel *RC* Circuits

23. The circuit in Figure 10–76 is a complex circuit consisting of five components. With trigonometry it can be shown that R_1, R_2, C_2, and C_3 are equivalent to a resistance of 3.08 Ω in series with a capacitive reactance of 39.5 Ω. From this given information, determine the voltages across each component in the original circuit.

24. Is the circuit in Figure 10–76 predominantly resistive or predominantly capacitive?

25. Find the current through each branch and the total current in Figure 10–76.

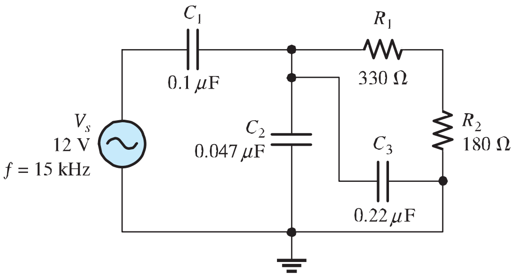

FIGURE 10–76

26. For the circuit in Figure 10–77, determine the following:
(a) I_{tot} **(b)** θ **(c)** V_{R1} **(d)** V_{R2} **(e)** V_{R3} **(f)** V_C

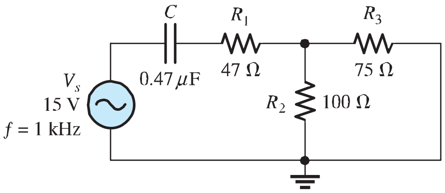

FIGURE 10–77

SECTION 10–7 Power in *RC* Circuits

27. In a certain series *RC* circuit, the true power is 2 W, and the reactive power is 3.5 VAR. Determine the apparent power.

28. In Figure 10–65, what is the true power and the reactive power?

29. What is the power factor for the circuit of Figure 10–75?

30. Determine P_{true}, P_r, P_a, and PF for the circuit in Figure 10–77. Draw the power triangle.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 10–8 Basic Applications

31. The lag circuit in Figure 10–67 also acts as a low-pass filter. Draw a response curve for this circuit by plotting the output voltage versus frequency for 0 Hz to 10 kHz in 1 kHz increments.

32. Plot the frequency response curve for the circuit in Figure 10–68 for a frequency range of 0 Hz to 10 kHz in 1 kHz increments.

33. Draw the voltage phasor diagram for each circuit in Figures 10–67 and 10–68 for a frequency of 5 kHz with $V_{in} = 1$ V rms.

34. The rms value of the signal voltage out of amplifier *A* in Figure 10–78 is 50 mV. If the input resistance to amplifier *B* is 10 kΩ, how much of the signal is lost due to the coupling capacitor (C_c) when the frequency is 3 kHz?

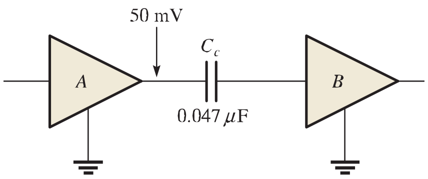

FIGURE 10–78

35. Determine the cutoff frequency for each circuit in Figures 10–67 and 10–68.

36. Determine the bandwidth of the circuit in Figure 10–68.

SECTION 10–9 Troubleshooting

37. Assume that the capacitor in Figure 10–79 is excessively leaky. Show how this degradation affects the output voltage and phase angle, assuming that the leakage resistance is 5 kΩ and the frequency is 10 Hz.

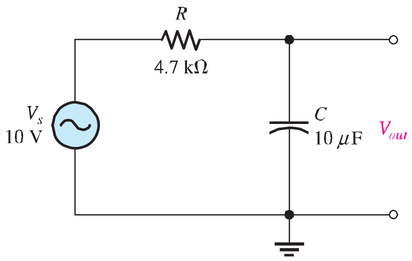

FIGURE 10–79

38. Each of the capacitors in Figure 10–80 has developed a leakage resistance of 2 kΩ. Determine the output voltages under this condition for each circuit.

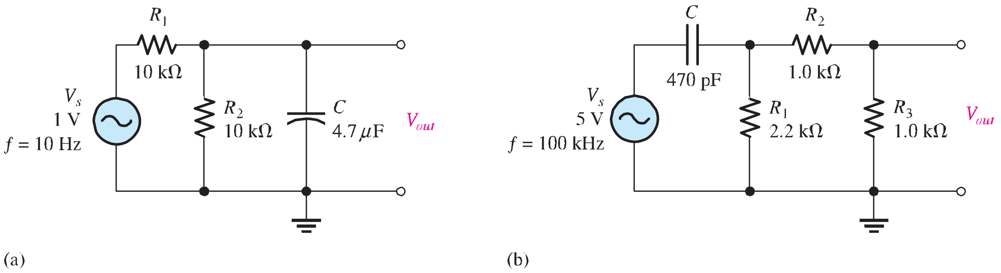

FIGURE 10–80

39. Determine the output voltage for the circuit in Figure 10–80(a) for each of the following failure modes, and compare it to the correct output:
(a) R_1 open **(b)** R_2 open **(c)** C open **(d)** C shorted

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

40. Determine the output voltage for the circuit in Figure 10–80(b) for each of the following failure modes, and compare it to the correct output:
(a) C open **(b)** C shorted **(c)** R_1 open **(d)** R_2 open **(e)** R_3 open

ADVANCED PROBLEMS

41. A single 240 V, 60 Hz source drives two loads. Load *A* has an impedance of 50 Ω and a power factor of 0.85. Load *B* has an impedance of 72 Ω and a power factor of 0.95.
- **(a)** How much current does each load draw?
- **(b)** What is the reactive power in each load?
- **(c)** What is the true power in each load?
- **(d)** What is the apparent power in each load?

42. What value of coupling capacitor is required in Figure 10–81 so that the signal voltage at the input of amplifier 2 is at least 70.7% of the signal voltage at the output of amplifier 1 when the frequency is 20 Hz? Disregard the input resistance of the amplifier.

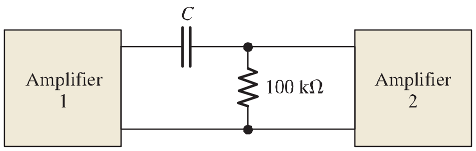

FIGURE 10–81

43. Determine the value of R_1 required to get a phase angle of 30° between the source voltage and the total current in Figure 10–82.

6238262 2013/12/10 12.27.66.5

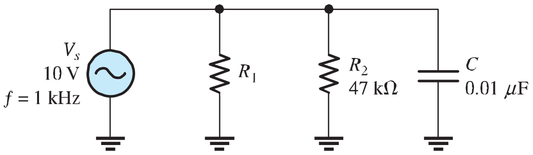

FIGURE 10–82

44. Draw the impedance phasor diagram for the input circuit of the amplifier in System Example 10–1 (Figure 10–24) at the cutoff frequency of 1.59 Hz.

45. A certain load dissipates 1.5 kW of power with an impedance of 12 Ω and a power factor of 0.75. What is its reactive power? What is its apparent power?

46. Determine the series element or elements that are in the block of Figure 10–83 to meet the following overall circuit requirements:
- **(a)** $P_{\text{true}} = 400$ W
- **(b)** leading power factor (I_{tot} leads V_s)

47. Determine the value of C_2 in Figure 10–84 when $V_A = V_B$.

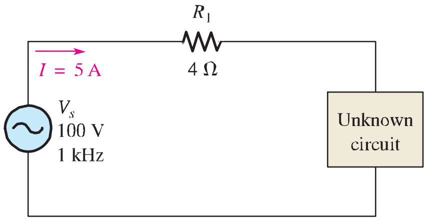

FIGURE 10–83

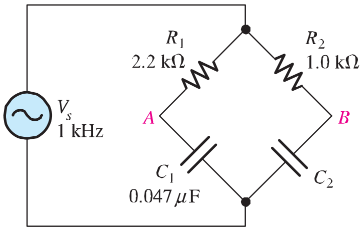

FIGURE 10–84

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

48. Draw the schematic for the circuit in Figure 10–85 and determine if the waveform on the scope is correct. If there is a fault in the circuit, identify it.

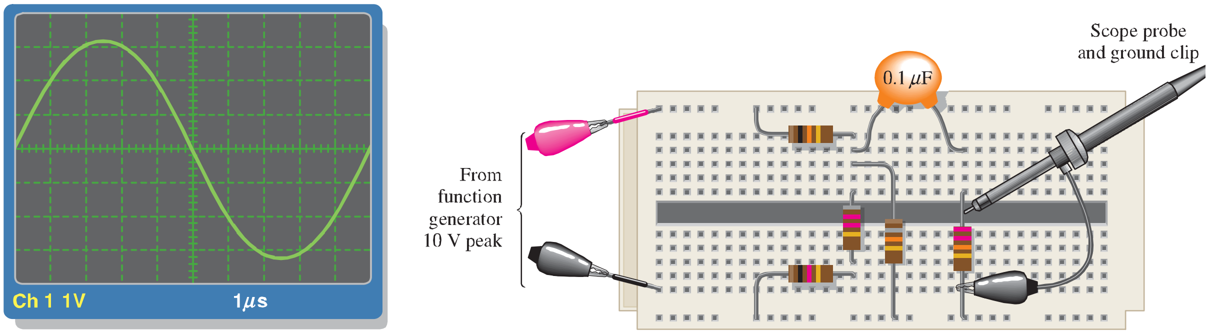
Ch 1 1V
1μs
0.1 μF
Scope probe and ground clip
From function generator 10 V peak

(a) Oscilloscope display (b) Circuit with leads connected

FIGURE 10–85

MULTISIM TROUBLESHOOTING PROBLEMS

MULTISIM

49. Open file P10-49; files are found at www.pearsonhighered.com/floyd. Determine if there is a fault and, if so, identify it.

50. Open file P10-50. Determine if there is a fault and, if so, identify it.

51. Open file P10-51. Determine if there is a fault and, if so, identify it.

52. Open file P10-52. Determine if there is a fault and, if so, identify it.

53. Open file P10-53. Determine if there is a fault and, if so, identify it.

54. Open file P10-54. Determine if there is a fault and, if so, identify it.

55. Open the Multisim file P10-55. Modify the circuit so that you can measure the phase shift as shown in Figure 10–38 by moving ground and using a Tektronix simulated scope as shown in Figure 10–39. You will need to trigger the scope with Normal mode triggering. You should be able to set up the display as shown in Figure 10–39 to measure the phase shift. Then simulate a shorted capacitor C_1 and determine the effect on phase shift. (You will need to reset the amplitudes to make an accurate check.)

ANSWERS TO SECTION CHECKUPS

SECTION 10–1 Sinusoidal Response of Series *RC* Circuits

1. V_C frequency is 60 Hz, I frequency is 60 Hz.
2. The capacitive reactance and resistance
3. θ is closer to 0° when $R > X_C$.

SECTION 10–2 Impedance and Phase Angle of Series *RC* Circuits

1. Impedance is the total opposition to sinusoidal current expressed in ohms.
2. V_s lags I.
3. The capacitive reactance produces the phase angle.
4. $Z = \sqrt{R^2 + X_C^2} = 59.9\text{ k}\Omega;\ \theta = \tan^{-1}(X_C/R) = 56.6^\circ$

SECTION 10–3 Analysis of Series *RC* Circuits

1. $V_s = \sqrt{V_R^2 + V_C^2} = 7.2\text{ V}$
2. $\theta = \tan^{-1}(V_C/V_R) = 56.3^\circ$ s
3. $\theta = 90^\circ$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

4. **(a)** X_C decreases when f increases.
   (b) Z decreases when f increases.
   (c) θ decreases when f increases.

5. $\phi = 90° - \tan^{-1}(X_C/R) = 62.8°$

6. $V_{out} = (R/\sqrt{R^2 + X_C^2})V_{in} = 8.9\text{ V rms}$

SECTION 10–4 Impedance and Phase Angle of Parallel *RC* Circuits

1. $Z = RX_C/\sqrt{R^2 + X_C^2} = 545\ \Omega$

2. *Conductance* is the reciprocal of resistance; *capacitive susceptance* is the reciprocal of capacitive reactance; and *admittance* is the reciprocal of impedance.

3. $Y = 1/Z = 10\text{ mS}$

4. $Y = \sqrt{G^2 + B_C^2} = 24\text{ mS}$

SECTION 10–5 Analysis of Parallel *RC* Circuits

1. $I_{tot} = V_sY = 21\text{ mA}$

2. $\theta = \tan^{-1}(I_C/I_R) = 56.3°$; $I_{tot} = \sqrt{I_R^2 + I_C^2} = 18\text{ mA}$

3. $\theta = 90°$

SECTION 10–6 Analysis of Series-Parallel *RC* Circuits

1. All grounds should be connected to an electrical common point to avoid shorting components.

2. $V_1 = I_{tot}R_1 = 6.71\text{ V}$

3. 1.67 pF

SECTION 10–7 Power in *RC* Circuits

1. Power dissipation is due to the resistance.

2. $PF = \cos 45° = 0.707$

3. $P_{\text{true}} = I_{tot}^2R = 1.32\text{ kW}$; $P_r = I_{tot}^2X_C = 1.84\text{ kVAR}$; $P_a = I_{tot}^2Z = 2.26\text{ kVA}$

SECTION 10–8 Basic Applications

1. 180°

2. Output is across the capacitor.

SECTION 10–9 Troubleshooting

1. The leakage resistance acts in parallel with *C*, which alters the circuit time constant.

2. The capacitor is open.

3. A shorted capacitor, open resistor, no source voltage, or open contact can cause 0 V across the capacitor.

ANSWERS TO RELATED PROBLEMS FOR EXAMPLES

10–1 2.42 kΩ; 65.6°

10–2 2.56 V

10–3 65.5°

10–4 7.14 V

10–5 15.9 kΩ; 86.4°

10–6 Phase lag ϕ increases.

10–7 Output voltage decreases.

10–8 Phase lead ϕ decreases; output voltage increases.

10–9 24.3 Ω

10–10 4.60 mS

10–11 6.16 mA

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

10–12 117 mA; 31°
10–13 $V_1 = 7.04$ V; $V_2 = 3.21$ V
10–14 0.146
10–15 990 μW
10–16 1.60 kHz
10–17 7.29 V
10–18 Resistor open

ANSWERS TO TRUE/FALSE QUIZ

1. F **2.** F **3.** F **4.** T **5.** T **6.** T **7.** T **8.** F **9.** F **10.** F

ANSWERS TO SELF-TEST

1. (c) **2.** (b) **3.** (b) **4.** (a) **5.** (d) **6.** (b) **7.** (a) **8.** (d)
9. (c) **10.** (b) **11.** (d) **12.** (d) **13.** (c) **14.** (b) **15.** (d)

ANSWERS TO TROUBLESHOOTING: SYMPTOM AND CAUSE

1. (b)
2. (b)
3. (a)
4. (c)
5. (b)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

CHAPTER 11

INDUCTORS

OUTLINE

11–1 The Basic Inductor
11–2 Types of Inductors
11–3 Series and Parallel Inductors
11–4 Inductors in DC Circuits
11–5 Inductors in AC Circuits
11–6 Inductor Applications

OBJECTIVES

- Describe the basic structure and characteristics of an inductor
- Discuss various types of inductors
- Analyze series and parallel inductors
- Analyze inductive dc switching circuits
- Analyze inductive ac circuits
- Discuss some inductor applications

KEY TERMS

Inductor
Winding
Coil
Induced voltage
Inductance (*L*)
Henry (H)
Winding resistance
***RL* time constant**
Inductive reactance
Quality factor (*Q*)

INTRODUCTION

Inductance is the property of a coil of wire that opposes a change in current. The basis for inductance is the electromagnetic field that surrounds any conductor when there is current through it. The electrical component designed to have the property of inductance is called an *inductor, coil* or in some applications *choke*. These terms refer to the same type of device. The term *choke* is usually associated with inductors that are used to block high frequencies which are widely used in communication systems and in radio-frequency circuits.

In this chapter, you will study the basic inductor and its characteristics. Various types of inductors are covered in terms of their physical construction and their electrical properties. The basic behavior of inductors in both dc and ac circuits is discussed, and series and parallel combinations are analyzed.

VISIT THE WEBSITE
Study aids for this chapter are available at
http://pearsonhighered.com/floyd

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

11–1 THE BASIC INDUCTOR

An inductor is a passive electrical component, formed by a coil of wire, that exhibits the property of inductance.

After completing this section, you should be able to

- **Describe the basic structure and characteristics of an inductor**
  - **Define *inductance* and state its unit**
  - **Discuss induced voltage**
  - **Explain how an inductor stores energy**
  - **Specify how the physical characteristics affect inductance**
  - **Discuss winding resistance and winding capacitance**
  - **Discuss how Faraday's law and Lenz's law explain the voltage induced across an inductor**

When a length of wire is formed into a coil, as shown in Figure 11–1, it becomes an inductor. Current through the coil produces an electromagnetic field. The magnetic lines of force around each loop (turn) in the **winding** of the coil effectively add to the lines of force around the adjoining loops, forming a stronger magnetic field within and around the coil, as shown. The net direction of the total magnetic field creates a north and a south pole. A schematic symbol for the inductor is shown in Figure 11–2.

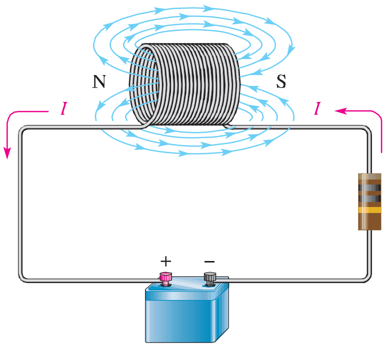

FIGURE 11–1 Current through a coil creates a three-dimensional electromagnetic field. The resistor limits the current.

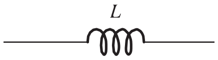

FIGURE 11–2 Symbol for an inductor.

Inductance

When there is current through an inductor, an electromagnetic field is established. When the current changes, the electromagnetic field also changes. An increase in current expands the field, and a decrease in current reduces it. Therefore, a changing current produces a changing electromagnetic field around the inductor (also known as **coil** and in some applications, **choke**). In turn, the changing electromagnetic field causes an **induced voltage** across the coil in a direction to oppose the change in current. This property is called *self-inductance* but is usually referred to as simply *inductance*, symbolized by *L*.

> **Inductance is a measure of a coil's ability to establish an induced voltage as a result of a change in its current, and that induced voltage is in a direction to oppose that change in current.**

THE UNIT OF INDUCTANCE The **henry**, symbolized by H, is the basic unit of inductance. By definition, the inductance of a coil is one henry when current through the coil, changing at the rate of one ampere per second, induces one volt across the coil. The

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

henry is a large unit, so in practical applications, millihenries (mH), and microhenries (μH) are the most common units. Nanohenries (nH) are also used for very small inductors.

ENERGY STORAGE An inductor stores energy in the magnetic field created by the current. The energy stored is expressed as

$$W = \frac{1}{2}LI^2 \tag{11–1}$$

As you can see, the energy stored is proportional to the inductance and the square of the current. When current (I) is in amperes and inductance (L) is in henries, the energy (W) is in joules.

Physical Characteristics of an Inductor

6238262 2013/12/10 12.27.66.5

The following characteristics are important in establishing the inductance of a coil: the permeability of the core material, the number of turns of wire, the core length, and the cross-sectional area of the core.

CORE MATERIAL As discussed earlier, an inductor is basically a coil of wire. The material around which the coil is formed is called the **core**. Coils are wound on either nonmagnetic or magnetic materials. Examples of nonmagnetic materials are air, wood, copper, plastic, and glass. The permeabilities of these materials are the same as for a vacuum. Examples of magnetic materials are iron, nickel, steel, cobalt, or alloys. These materials have permeabilities that are hundreds or thousands of times greater than that of a vacuum and are classified as *ferromagnetic*. A ferromagnetic core provides a lower reluctance path for the magnetic lines of force and thus permits a stronger magnetic field.

As you learned in Chapter 7, the permeability (μ) of the core material determines how easily a magnetic field can be established. The inductance is directly proportional to the permeability of the core material.

PHYSICAL PARAMETERS As indicated in Figure 11–3, the number of turns of wire, the length, and the cross-sectional area of the core are factors in setting the value of inductance. The inductance is inversely proportional to the length of the core and directly proportional to the cross-sectional area. Also, the inductance is directly related to the number of turns squared. This relationship is as follows:

$$L = \frac{N^2\mu A}{l} \tag{11–2}$$

where L is the inductance in henries (H), N is the number of turns, μ is the permeability in henries per meter (H/m), A is the cross-sectional area in meters squared (m^2), and l is the core length in meters (m).

FIGURE 11–3 Factors that determine the inductance of a coil.

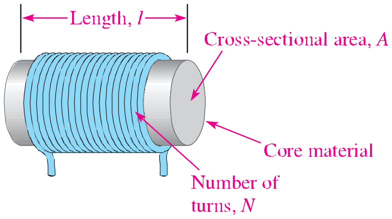

EXAMPLE 11–1

Determine the inductance of the coil in Figure 11–4. The permeability of the core is 0.25×10^{-3} H/m.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

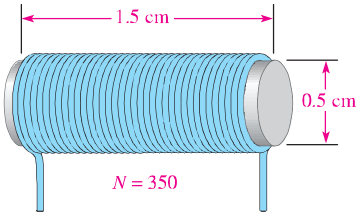

FIGURE 11–4

SOLUTION

1 cm = 0.01 m. Therefore, 1.5 cm = 0.015 m and 0.5 cm = 0.005 m.

$$A = \pi r^2 = \pi(0.25 \times 10^{-2}\ \text{m})^2 = 1.96 \times 10^{-5}\ \text{m}^2$$

$$L = \frac{N^2 \mu A}{l} = \frac{(350)^2(0.25 \times 10^{-3}\ \text{H/m})(1.96 \times 10^{-5}\ \text{m}^2)}{0.015\ \text{m}} = \mathbf{40\ mH}$$

RELATED PROBLEM*

Determine the inductance of a coil with 400 turns on a core that is 2 cm long and has a diameter of 1 cm. The permeability is 0.25×10^{-3} H/m.

***Answers are at the end of the chapter.**

Winding Resistance

When a coil is made of a certain material (for example, insulated copper wire), that wire has a certain resistance per unit of length. When many turns of wire are used to construct a coil, the total resistance may be significant. This inherent resistance is called the *dc resistance* or the **winding resistance (R_W)**.

Although this resistance is distributed along the length of the wire, as shown in Figure 11–5(a), it is sometimes indicated in a schematic as resistance appearing in series with the inductance of the coil, as shown in Figure 11–5(b). In many applications, the winding resistance can be ignored and the coil can be considered an ideal inductor. In other cases, the resistance must be considered.

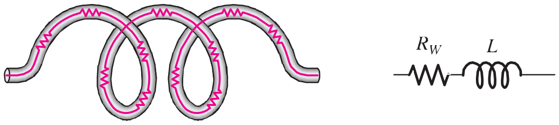

(a) The wire has resistance distributed along its length.

(b) Equivalent circuit

FIGURE 11–5 **Winding resistance of a coil.**

Winding Capacitance

When two conductors are placed side by side, there is always some capacitance between them. Thus, when many turns of wire are placed close together in a coil, a certain amount of stray capacitance, called *winding capacitance* (C_W), is a natural side effect. In many applications, this winding capacitance is very small and has no significant effect. In other cases, particularly at high frequencies, it may become quite important.

The equivalent circuit for an inductor with both its winding resistance (R_W) and its winding capacitance (C_W) is shown in Figure 11–6. The capacitance effectively acts in parallel. The total of the stray capacitances between each loop of the winding is indicated in a schematic as a capacitance appearing in parallel with the coil and its winding resistance, as shown in Figure 11–6(b).

SAFETY NOTE

Be careful when working with inductors because high induced voltages can be developed due to a rapidly changing magnetic field. This occurs when the current is interrupted or its value is abruptly changed.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

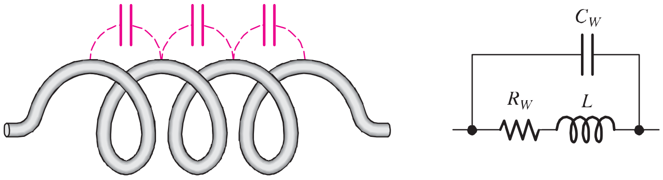

(a) Stray capacitance between each loop appears as a total parallel capacitance (C_W).

(b) Equivalent circuit

FIGURE 11–6 **Winding capacitance of a coil.**

Review of Faraday's Law

Faraday's law was introduced in Chapter 7 and is reviewed here because of its importance in the study of inductors. Faraday found that by moving a magnet through a coil of wire, a voltage was induced across the coil, and that when a complete path was provided, the induced voltage caused an induced current. He observed that

> **The amount of voltage induced in a coil is directly proportional to the rate of change of the magnetic field with respect to the coil.**

This principle is illustrated in Figure 11–7, where a bar magnet is moved through a coil of wire. An induced voltage is indicated by the voltmeter connected across the coil. The faster the magnet is moved, the greater the induced voltage.

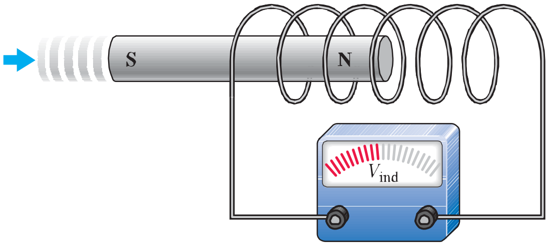

FIGURE 11–7 **Induced voltage is created by a changing magnetic field.**

When a wire is formed into a certain number of loops or turns and is exposed to a changing magnetic field, a voltage is induced across the coil. The induced voltage is proportional to the number of turns, N, of the wire in the coil and to the rate at which the magnetic field changes.

Lenz's Law

Lenz's law, also introduced in Chapter 7, adds to Faraday's law by defining the direction of induced voltage.

> **When the current through a coil changes and an induced voltage is created as a result of the changing magnetic field, the direction of the induced voltage is such that it always opposes the change in current.**

Figure 11–8 illustrates Lenz's law. In part (a), the current is constant and is limited by R_1. There is no induced voltage because the magnetic field is unchanging. In part (b), the switch suddenly is closed, placing R_2 in parallel with R_1 and thus reducing the resistance. Naturally, the current tries to increase and the magnetic field begins to expand, but the induced voltage opposes this attempted increase in current for an instant.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

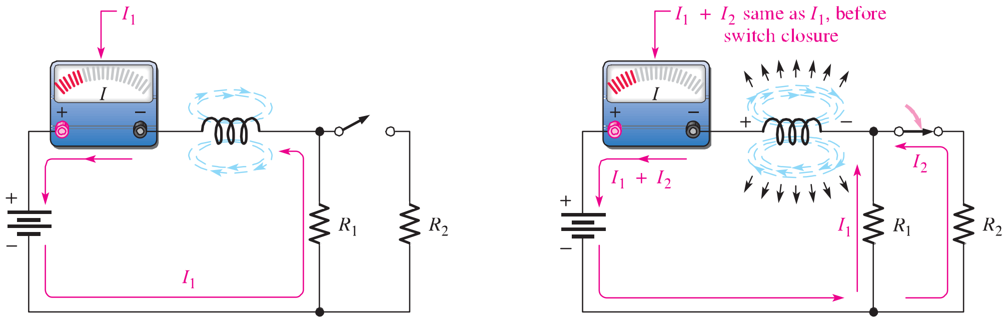

(a) Switch open: Constant current and constant magnetic field; no induced voltage.

(b) At instant of switch closure: Expanding magnetic field induces voltage, which opposes an increase in total current. The total current remains the same at this instant.

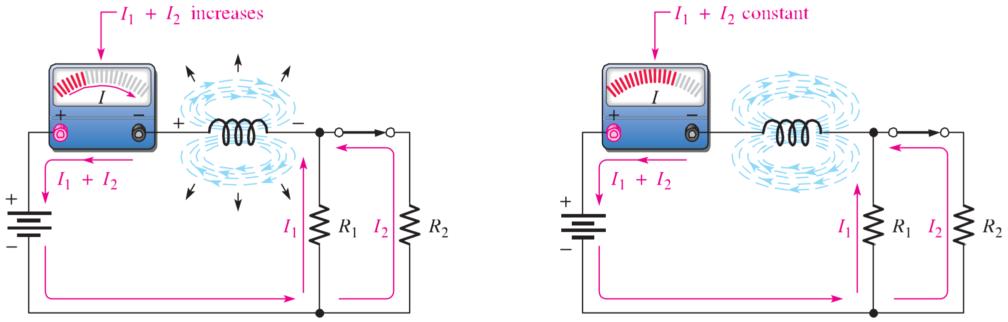

(c) Right after switch closure: The rate of expansion of the magnetic field decreases, allowing the current to increase exponentially as induced voltage decreases.

(d) Switch remains closed: Current and magnetic field reach constant value, no induced voltage.

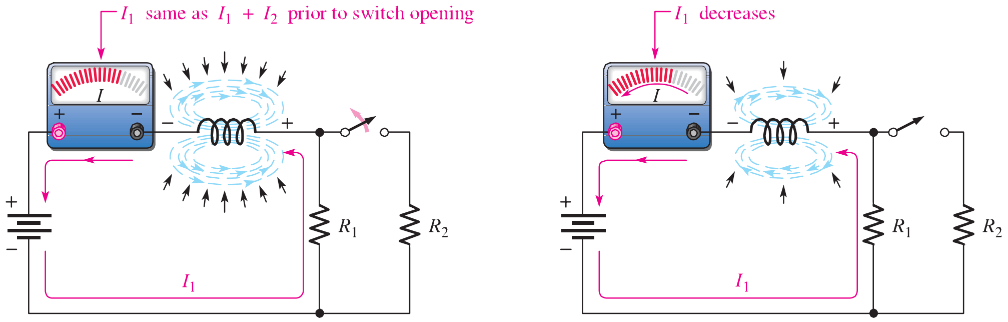

(e) At instant of switch opening: Magnetic field begins to collapse, creating an induced voltage, which opposes a decrease in current.

(f) After switch opening: Rate of collapse of magnetic field decreases, allowing current to decrease exponentially back to original value.

FIGURE 11–8 **Demonstration of Lenz's law in an inductive circuit: When the current tries to change suddenly, the electromagnetic field changes and induces a voltage in a direction that opposes that change in current.**

In Figure 11–8(c), the induced voltage gradually decreases, allowing the current to increase. In part (d), the current has reached a constant value as determined by the parallel resistors, and the induced voltage is zero. In part (e), the switch has been suddenly opened, and, for an instant, the induced voltage prevents any decrease in current. In part (f), the induced voltage gradually decreases, allowing the current to decrease back to a value determined by R_1. Notice that the induced voltage has a polarity that opposes any current change. The polarity of the induced voltage is opposite that of the battery voltage for an increase in current and aids the battery voltage for a decrease in current.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SYSTEM EXAMPLE 11–1

SWITCH-MODE POWER SUPPLY

Power supplies generally convert ac to dc; they are found in virtually all electronics equipment. Switch-mode power supplies (SMPS) became prevalent in the 1970s after the development of very fast switching transistors and IC pulse width modulators (PWM). The Apple II computer, released in 1977, became one of the first computer systems to use a switch-mode power supply, resulting in a much more efficient and cooler running supply. Today, nearly all computers use switch-mode power supplies.

A simplified schematic of a step-down SMPS is shown in Figure 11–9. The switch, SW, represents a switching MOSFET transistor that rapidly alternates between the open and closed position (this happens about 20,000 times per second). *V* represents an unregulated dc input voltage. The pulse-width modulator is the controller; it compares the output to a reference voltage and sets the pulse width to keep the output constant.

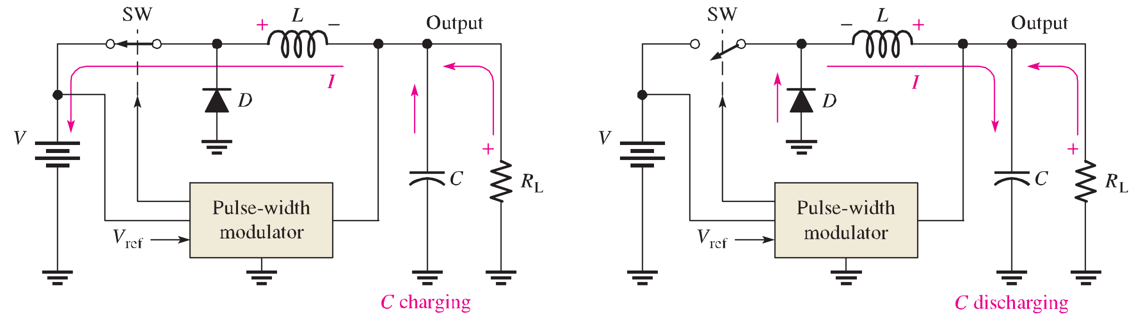

(a) Switch closed

(b) Switch open

FIGURE 11–9 **A simplified step-down switch-mode power supply.**

Our focus is on the inductor, *L*, a key component in a SMPS. Recall from Lenz's law that an inductor opposes a change in current. In part (a), the transistor is *on* as represented by the closed switch; the major current path to the load is shown with the red arrows. The induced voltage across the inductor is opposite to that of the dc input as the magnetic field builds and tends to keep the current constant in the load. In part (b), the switch is open (transistor *off*), and current is supplied through the diode and from the capacitor. Notice the inductor has the same polarity as the source when its magnetic field collapses. The induced voltage is opposite to what it was. The result is a nearly constant current in the load.

SECTION 11–1 CHECKUP*

1. List the parameters that contribute to the inductance of a coil.
2. Describe what happens to *L* when
   - **(a)** *N* is increased
   - **(b)** the core length is increased
   - **(c)** the cross-sectional area of the core is decreased
   - **(d)** a ferromagnetic core is replaced by an air core
3. Explain why inductors have some winding resistance.
4. Explain why inductors have some winding capacitance.

***Answers are at the end of the chapter.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

11–2 TYPES OF INDUCTORS

Inductors normally are classified according to the type of core material.

After completing this section, you should be able to

- **Discuss various types of inductors**
  - **Describe the basic types of fixed inductors**
  - **Distinguish between fixed and variable inductors**

Inductors are made in a variety of shapes and sizes. Basically, they fall into two general categories: fixed and variable. The standard schematic symbols are shown in Figure 11–10.

Both fixed and variable inductors can be classified according to the type of core material. Three common types are the air core, the iron core, and the ferrite core. Each has a unique symbol, as shown in Figure 11–11. Large-value iron-core inductors are rarely used.

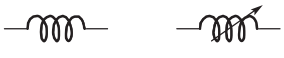
(a) Fixed (b) Variable

FIGURE 11–10 **Symbols for fixed and variable inductors.**

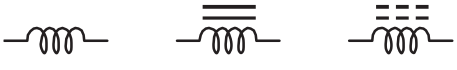
(a) Air core (b) Iron core (c) Ferrite core

FIGURE 11–11 **Inductor symbols.**

Adjustable (variable) inductors usually have a screw-type adjustment that moves a sliding core in and out, thus changing the inductance. A wide variety of inductors exists, and some are shown in Figure 11–12. Small fixed inductors are usually encapsulated in an insulating material that protects the fine wire in the coil. Encapsulated inductors have an appearance similar to a small resistor. Small-value inductors are available for high-frequency applications as surface-mount types and include some with ferrite-core construction.

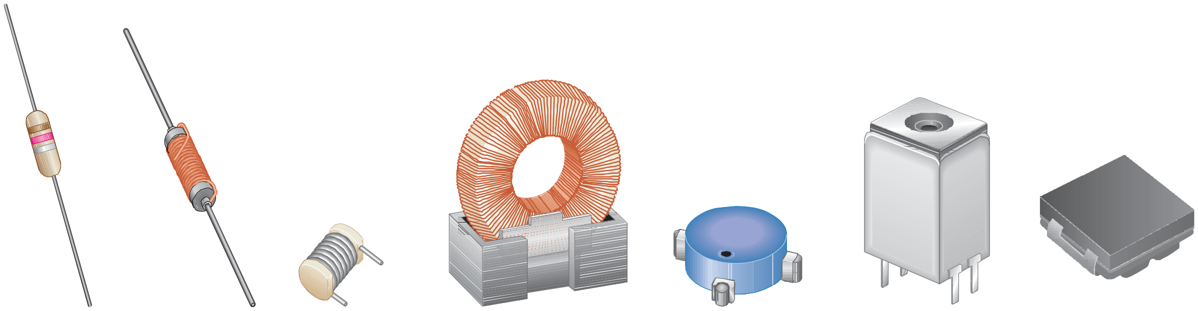

FIGURE 11–12 **Typical inductors.**

Shielded Inductors

Inductors can radiate electromagnetic energy in certain cases, causing interference in nearby circuits. To avoid this, engineers will sometimes specify shielded inductors. Shielded inductors are available in small surface-mount packages that are soldered in place. Shielded surface-mount inductors are available from 1 μH to 1 mH.

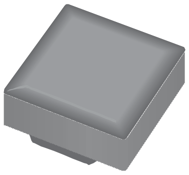

SYSTEM NOTE

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 11–2 CHECKUP

1. Name two general categories of inductors.
2. Identify the inductor symbols in Figure 11–13.

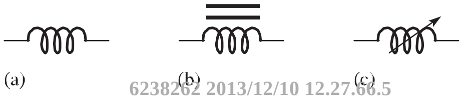

6238262 2013/12/10 12.27.66.5

FIGURE 11–13

11–3 SERIES AND PARALLEL INDUCTORS

When inductors are connected in series, the total inductance increases. When inductors are connected in parallel, the total inductance decreases.

After completing this section, you should be able to

- **Analyze series and parallel inductors**
  - **Determine total series inductance**
  - **Determine total parallel inductance**

Total Series Inductance

When inductors are connected in series, as in Figure 11–14, the total inductance, L_T, is the sum of the individual inductances. The formula for L_T is expressed in the following equation for the general case of n inductors in series:

$$L_T = L_1 + L_2 + L_3 + \cdots + L_n \quad (11\text{–}3)$$

Notice that the calculation of total inductance in series is analogous to the calculations of total resistance in series (Chapter 4) and total capacitance in parallel (Chapter 9).

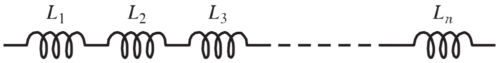

FIGURE 11–14 **Induction in series.**

EXAMPLE 11–2

Determine the total inductance for each of the series connections in Figure 11–15.

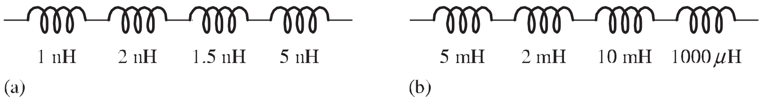

FIGURE 11–15

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

In Figure 11–15(a),

$$L_T = 1\text{ nH} + 2\text{ nH} + 1.5\text{ nH} + 5\text{ nH} = 9.5\text{ nH}$$

In Figure 11–15(b),

$$L_T = 5\text{ mH} + 2\text{ mH} + 10\text{ mH} + 1\text{ mH} = \mathbf{18\text{ mH}}$$

Note: $1000\ \mu\text{H} = 1\text{ mH}$

RELATED PROBLEM

What is the total inductance of three 50 μH inductors in series?

SYSTEM EXAMPLE 11–2

HARTLEY OSCILLATORS

Recall from System Example 9–1 that an oscillator is a circuit that generates a repetitive wave. The circuit shown in Figure 11–16 is a Hartley oscillator that uses inductors in series. As in the case of the Colpitts oscillator, this circuit uses positive feedback—a portion of the output is returned by the feedback network to the input to be amplified and reinforce the output. The feedback is across L_1 and is referenced to ground. Although the circuit is an ac circuit, the equations for inductors and inductive voltages developed with dc sources can be applied.

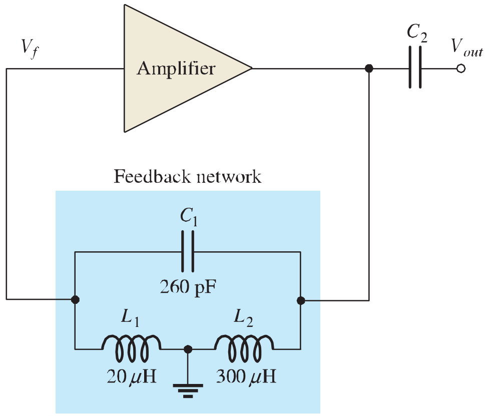

FIGURE 11–16 **A Hartley oscillator.**

Inductors L_1 and L_2 are in series and form part of a resonant circuit (resonant circuits are covered in Chapter 13). Find the total inductance by simply adding the inductors.

$$L_T = L_1 + L_2 = 20\ \mu\text{H} + 300\ \mu\text{H} = 320\ \mu\text{H}$$

The amplifier's signals are referenced to ground, so the feedback signal, V_f, is the fraction of the output developed across L_1. Notice that V_f is returned to the input of the amplifier. You can find the fraction of output voltage that is fed back (V_f/V_{out}) by applying the voltage divider idea to the series inductors.

$$\frac{V_f}{V_{out}} = \frac{L_1}{L_T} = \left(\frac{20\ \mu\text{H}}{320\ \mu\text{H}}\right) = 0.0625 = 6.2\%$$

Unlike capacitors, the smaller inductor has the smaller fraction of the output voltage.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Total Parallel Inductance

When inductors are connected in parallel, as in Figure 11–17, the total inductance is less than the smallest inductance. The general formula states that the reciprocal of the total inductance is equal to the sum of the reciprocals of the individual inductances.

$$\frac{1}{L_T} = \frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} + \cdots + \frac{1}{L_n}$$

You can calculate total inductance, L_T, by taking the reciprocals of both sides of the previous equation.

$$\mathbf{L_T = \frac{1}{\frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} + \cdots + \frac{1}{L_n}}} \qquad \text{(11–4)}$$

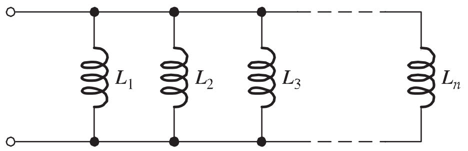

FIGURE 11–17 **Inductors in parallel.**

This calculation for total inductance in parallel is analogous to the calculations of total parallel resistance (Chapter 5) and total *series* capacitance (Chapter 9). For series-parallel combinations of inductors, determine the total inductance in the same way as total resistance in series-parallel resistive circuits (Chapter 6).

EXAMPLE 11–3

Determine L_T in Figure 11–18.

FIGURE 11–18

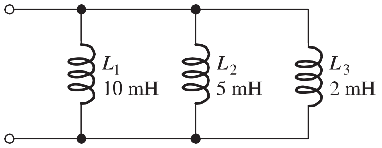

SOLUTION

$$L_T = \frac{1}{\frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3}} = \frac{1}{\frac{1}{10\text{ mH}} + \frac{1}{5\text{ mH}} + \frac{1}{2\text{ mH}}} = \frac{1}{0.8\text{ mH}} = \mathbf{1.25\text{ mH}}$$

RELATED PROBLEM

Determine L_T for the following inductors in parallel: 50 μH, 80 μH, 100 μH, and 150 μH.

SECTION 11–3 CHECKUP

1. State the rule for combining inductors in series.
2. What is L_T for a series connection of 100 μH, 500 μH, and 2 mH?
3. Five 100 mH coils are connected in series. What is the total inductance?
4. Compare the total inductance in parallel with the smallest-value individual inductor.
5. The calculation of total parallel inductance is analogous to that for total parallel resistance. (True or False)
6. Determine L_T for each parallel combination:
   (a) 100 mH, 50 mH, and 10 mH
   (b) 40 μH and 60 μH

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

11–4 INDUCTORS IN DC CIRCUITS

Energy is stored in the electromagnetic field of an inductor when it is connected to a dc voltage source. The buildup of current through the inductor occurs in a predictable manner, which is dependent on the time constant of the circuit. The time constant is determined by the inductance and resistance in a circuit.

After completing this section, you should be able to

- **Analyze inductive dc switching circuits**
  - **Define *RL time constant***
  - **Describe the increase and decrease of current in an inductor**
  - **Relate the time constant to the energizing and deenergizing of an inductor**
  - **Describe induced voltage**
  - **Write the exponential equations for current in an inductor**

When there is constant direct current in an inductor, there is no induced voltage. There is, however, a voltage drop due to the winding resistance of the coil. The inductance itself appears as a short to dc. Energy is stored in the magnetic field according to the formula $W = \frac{1}{2}LI^2$. The only energy conversion to heat occurs in the winding resistance ($P = I^2R_W$). This condition is illustrated in Figure 11–19.

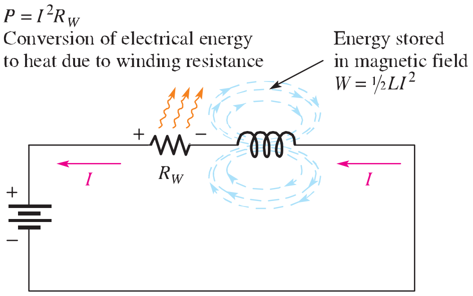
$P = I^2R_W$
Conversion of electrical energy to heat due to winding resistance
Energy stored in magnetic field $W = \frac{1}{2}LI^2$
+
–
R_W
I
I
+
–

FIGURE 11–19 **Energy storage and conversion to heat in an inductor.**

The *RL* Time Constant

Because the inductor's basic action is to develop a voltage that opposes a change in its current, it follows that current cannot change instantaneously in an inductor. A certain time is required for the current to make a change from one value to another. The rate at which the current changes is determined by the *RL* time constant.

> **The *RL* time constant is a fixed time interval that equals the ratio of the inductance to the resistance.**

The formula is

$$\tau = \frac{L}{R} \tag{11–5}$$

where τ is in seconds when inductance (L) is in henries and resistance (R) is in ohms.

EXAMPLE 11–4

A series *RL* circuit has a resistance of 1.0 kΩ and an inductance of 1.0 mH. What is the time constant?

6238262 2013/12/10 12.27.66.5

SOLUTION

$$\tau = \frac{L}{R} = \frac{1.0\text{ mH}}{1.0\text{ k}\Omega} = \frac{1.0 \times 10^{-3}\text{ H}}{1.0 \times 10^{3}\ \Omega} = 1.0 \times 10^{-6}\text{ s} = \mathbf{1.0\ \mu s}$$

RELATED PROBLEM

Find the time constant for $R = 2.2\text{ k}\Omega$ and $L = 500\ \mu\text{H}$.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Current in an Inductor

INCREASING CURRENT In a series RL circuit, the current will increase to approximately 63% of its full value in one time-constant interval after voltage is applied. This buildup of current (analogous to the buildup of capacitor voltage during the charging in an RC circuit) follows an exponential curve and reaches the approximate percentage of final current as indicated in Table 11–1 and as illustrated in Figure 11–20.

TABLE 11–1 • Percentage of final current after each time-constant interval during current buildup.

NUMBER OF TIME CONSTANTS	APPROXIMATE % OF FINAL CURRENT
1	63
2	86
3	95
4	98
5	99 (considered 100%)

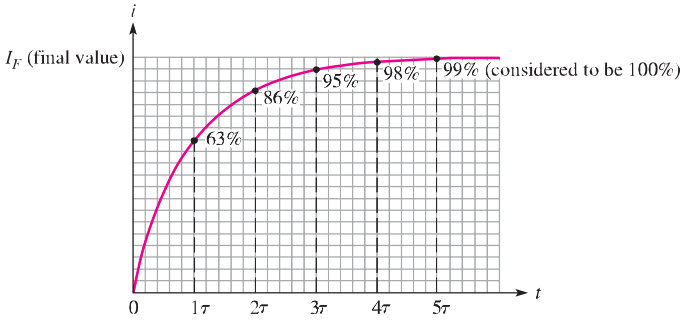

FIGURE 11–20 Increasing current in an inductor.

The change in current over five time-constant intervals is illustrated in Figure 11–21. When the current reaches its final value at approximately 5τ, it ceases to change. At this time, the inductor acts as a short (except for winding resistance) to the constant current. The final value of the current is

$$I_F = \frac{V_S}{R} = \frac{10\text{ V}}{1.0\text{ k}\Omega} = 10\text{ mA}$$

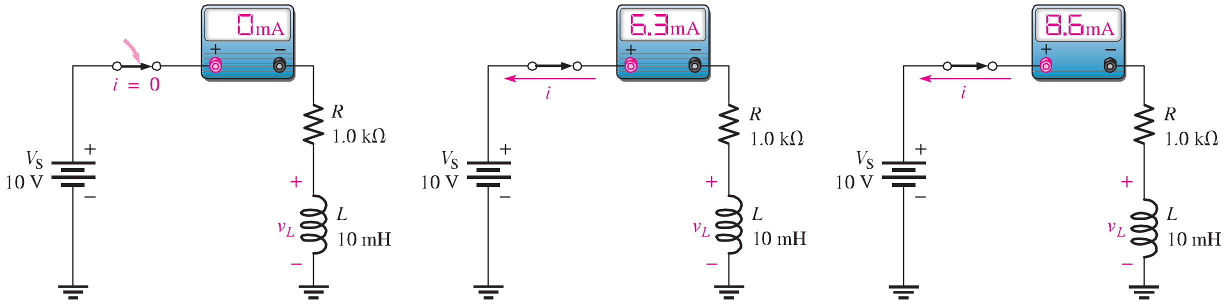

(a) Initially ($t = 0$) (b) At $t = 1\tau$ (c) At $t = 2\tau$

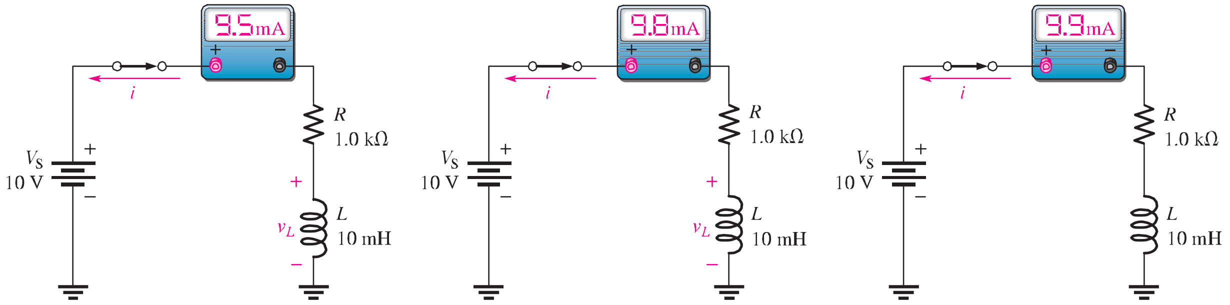

(d) At $t = 3\tau$ (e) At $t = 4\tau$ (f) At $t = 5\tau$

FIGURE 11–21 Illustration of the exponential buildup of current in an inductor. The current increases approximately 63% during each time-constant interval after the switch is closed. A voltage (v_L) is induced in the coil that tends to oppose the increase in current.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 11–5

Calculate the *RL* time constant for Figure 11–22. Then determine the current and the time at each time-constant interval, measured from the instant the switch is closed.

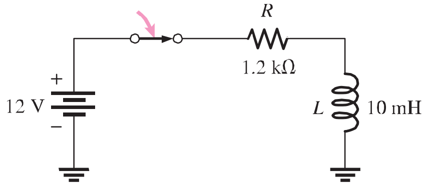

FIGURE 11–22

SOLUTION

The *RL* time constant is

$$\tau = \frac{L}{R} = \frac{10\text{ mH}}{1.2\text{ k}\Omega} = \mathbf{8.33\ \mu s}$$

The final current is

$$I_F = \frac{V_S}{R} = \frac{12\text{ V}}{1.2\text{ k}\Omega} = 10\text{ mA}$$

Use the time-constant percentage values from Table 11–1.

At 1τ: $i = 0.63(10\text{ mA}) = \mathbf{6.3\ mA};\ t = \mathbf{8.33\ \mu s}$
At 2τ: $i = 0.86(10\text{ mA}) = \mathbf{8.6\ mA};\ t = \mathbf{16.7\ \mu s}$
At 3τ: $i = 0.95(10\text{ mA}) = \mathbf{9.5\ mA};\ t = \mathbf{25.0\ \mu s}$
At 4τ: $i = 0.98(10\text{ mA}) = \mathbf{9.8\ mA};\ t = \mathbf{33.3\ \mu s}$
At 5τ: $i = 0.99(10\text{ mA}) = 9.9\text{ mA} \cong \mathbf{10\ mA};\ t = \mathbf{41.7\ \mu s}$

RELATED PROBLEM

Repeat the calculations if $R = 680\ \Omega$ and $L = 100\ \mu\text{H}$.

DECREASING CURRENT Current in an inductor decreases exponentially according to the approximate percentage values shown in Table 11–2 and in Figure 11–23.

TABLE 11–2 • Percentage of initial current after each time-constant interval while current is decreasing.

NUMBER OF TIME CONSTANTS	APPROXIMATE % OF INITIAL CURRENT
1	37
2	14
3	5
4	2
5	1 (considered 0)

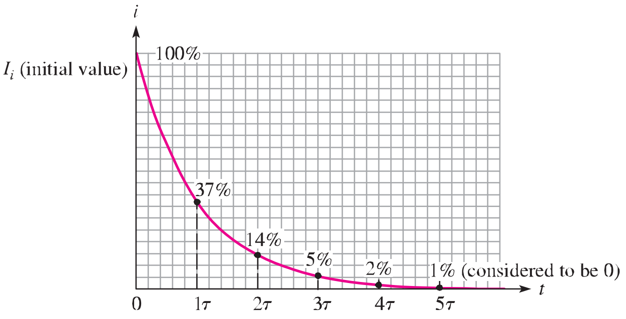

FIGURE 11–23 **Decreasing current in an inductor.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The change in current over five time-constant intervals is illustrated in Figure 11–24. When the current reaches its final value of approximately 0 A, it ceases to change. Before the switch is opened, the current through L is at a constant value of 10 mA, which is determined by R_1 because L acts ideally as a short. When the switch is opened, the induced inductor voltage initially provides 10 mA through R_2. The current then decreases by 63% during each time constant interval.

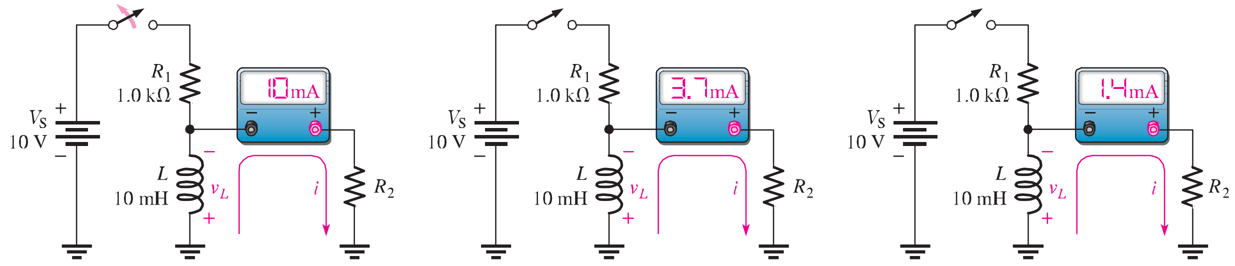

(a) Initially (t = 0) (b) At $t = 1\tau$ (c) At $t = 2\tau$

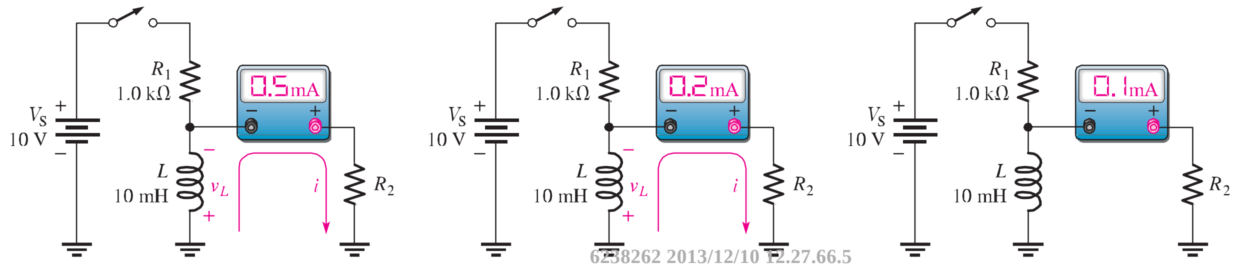

6238262 2013/12/10 12.27.66.5

(d) At $t = 3\tau$ (e) At $t = 4\tau$ (f) At $t = 5\tau$ (i is assumed to be 0)

FIGURE 11–24

Response to a Square Wave

A good way to demonstrate both increasing and decreasing current in an RL circuit is to use a square wave voltage as the input. The square wave is a useful signal for observing the dc response of a circuit because it provides on and off action similar to a switch. Pulse response will be covered further in Chapter 15. When the square wave goes from its low level to its high level, the current in the circuit responds by exponentially increasing to its final value. When the square wave returns to the zero level, the current in the circuit responds by exponentially decreasing to its zero value. Figure 11–25 shows input voltage and current waveforms. The current will have the same shape as the voltage across the resistor, so a two-channel difference measurement can show the waveform.

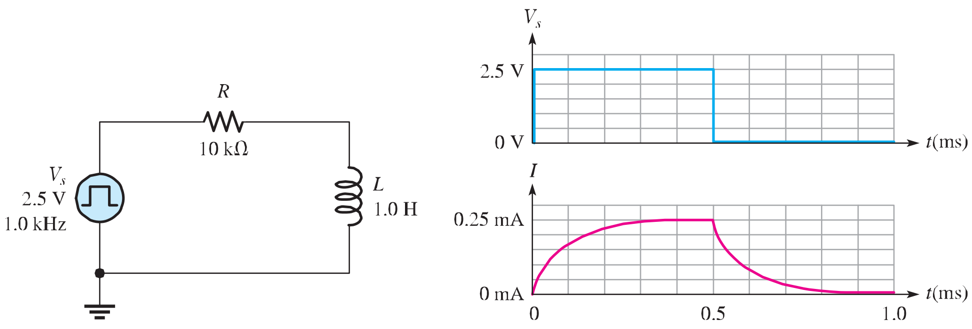

FIGURE 11–25

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 11–6

For the circuit in Figure 11–25, what is the current at 0.1 ms and at 0.6 ms?

SOLUTION

The *RL* time constant for the circuit is

$$\tau = \frac{L}{R} = \frac{1.0\text{ H}}{10\text{ k}\Omega} = 0.1\text{ ms}$$

If the square wave generator period is long enough for the current to reach its maximum value in 5τ, the current will increase exponentially and during each time constant interval will have a value equal to the percentage of the final current given in Table 11–1. The final current is

$$I_F = \frac{V_s}{R} = \frac{2.5\text{ V}}{10\text{ k}\Omega} = 0.25\text{ mA}$$

The current at 0.1 ms is

$$i = 0.63(0.25\text{ mA}) = \mathbf{0.158\text{ mA}}$$

At 0.6 ms, the square wave input has been at the 0 V level for 0.1 ms, or 1τ; and, the current decreases from the maximum value toward its final value of 0 mA by 63%. Therefore,

$$i = 0.25\text{ mA} - 0.63(0.25\text{ mA}) = \mathbf{0.092\text{ mA}}$$

RELATED PROBLEM

What is the current at 0.2 ms and 0.8 ms?

MULTISIM

Open the Multisim file E11-06; files are found at www.pearsonhighered.com/floyd. Notice that the inductor and resistor are reversed in order to ground one side of the resistor and make the measurement of the resistor voltage simpler. The shape of the current in the circuit is the same as the shape of the resistor voltage. By applying Ohm's law to the resistor voltage, you can find the current in the circuit at any instant in time. Confirm that the current at a time 0.1 ms is close to the calculated value.

Voltages in the Series *RL* Circuit

As you know, when current changes in an inductor, a voltage is induced. Let's examine what happens to the induced voltage across the inductor in the series circuit in Figure 11–26 during one complete cycle of a square wave input. Keep in mind that the generator produces a level that is like switching a dc source on and then puts an "automatic" low resistance (ideally zero) path across the source when it returns to its zero level.

An ammeter placed in the circuit shows the current in the circuit at any instant in time. The V_L waveform is the voltage across the inductor. In Figure 11–26(a), the square wave has just transitioned from zero to its maximum value of 2.5 V. In accordance with Lenz's law, a voltage is induced across the inductor that opposes this *change* as the magnetic field surrounding the inductor builds up. There is no current in the circuit due to the equal but opposing voltages.

As the magnetic field builds up, the induced voltage across the inductor decreases, and current is in the circuit. After 1τ, the induced voltage across the inductor has decreased by 63%, which causes the current to increase by 63% to 0.158 mA. This is shown in Figure 11–26(b) at the end of one time constant (0.1 ms).

The voltage on the inductor continues to exponentially decrease to zero, at which point the current is limited only by the circuit resistance. Then the square wave goes back to zero (at $t = 0.5$ ms) as shown in Figure 11–26(c). Again a voltage is induced across the inductor opposing this change. This time, the polarity of the inductor voltage is reversed due to the collapsing magnetic field. Although the source voltage is 0, the collapsing magnetic field maintains current in the same direction until the current decreases to zero, as shown in Figure 11–26(d).

The voltage across the resistor, V_R, in Figure 11–26(d) is found by subtracting the voltage across the inductor, V_L, from the source voltage, V_s, according to Kirchhoff's voltage law. The shape of the V_R waveform is the same as the current waveform in Figure 11–25.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

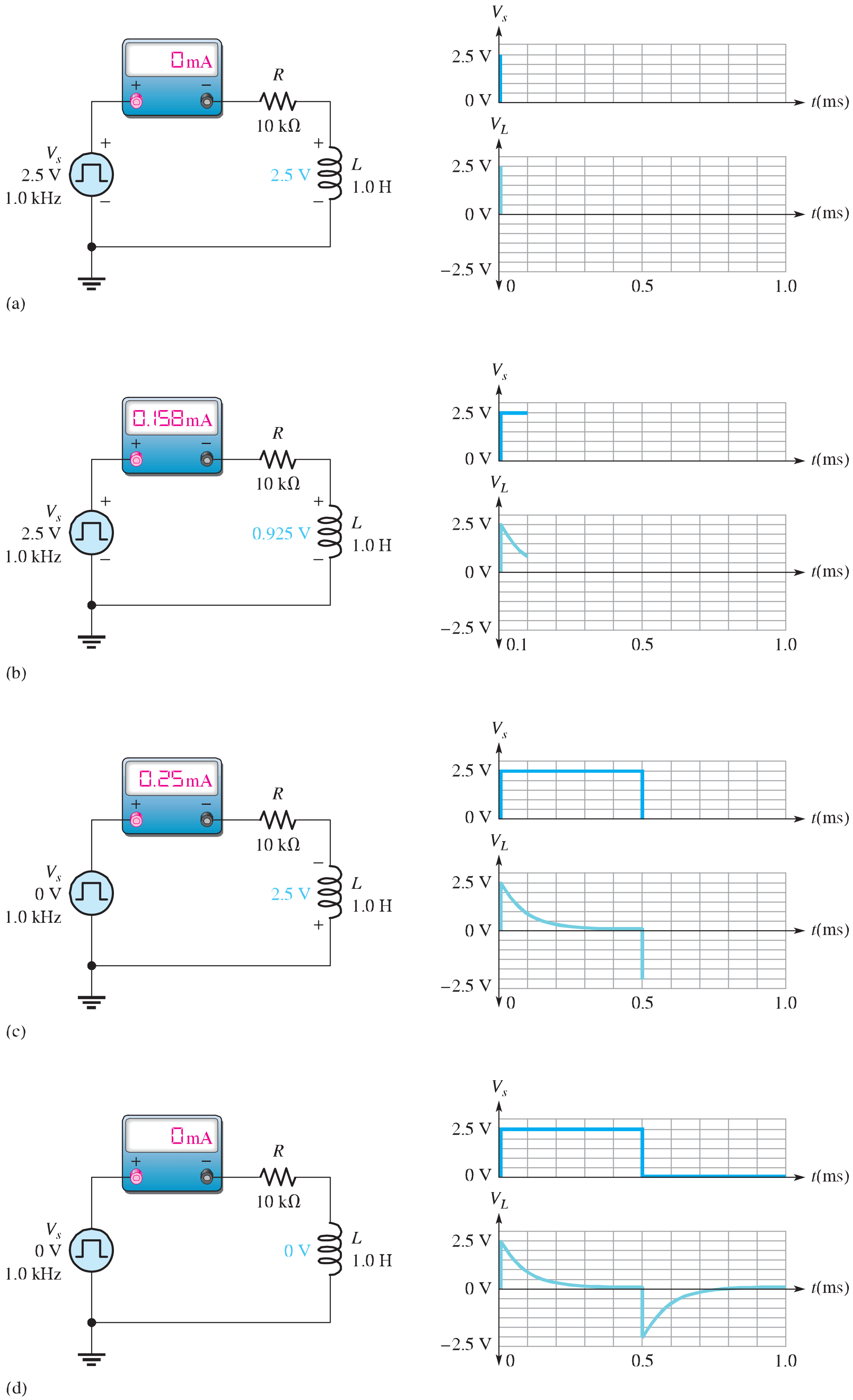
0 mA
R
10 kΩ
V_s
2.5 V
1.0 kHz
2.5 V
L
1.0 H
V_s
V_L
2.5 V
0 V
−2.5 V
t(ms)
0
0.5
1.0
(a)
0.158 mA
R
10 kΩ
V_s
2.5 V
1.0 kHz
0.925 V
L
1.0 H
0.1
(b)
0.25 mA
R
10 kΩ
V_s
0 V
1.0 kHz
2.5 V
L
1.0 H
(c)
0 mA
R
10 kΩ
V_s
0 V
1.0 kHz
0 V
L
1.0 H
(d)

FIGURE 11–26

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 11–7

(a) The circuit in Figure 11–27 has a square wave input. What is the highest frequency that can be used and still observe the complete waveform across the inductor?

6238262 2013/12/10 12.27.66.5

(b) Assume the generator is set to the frequency determined in (a). Describe the voltage waveform across the resistor?

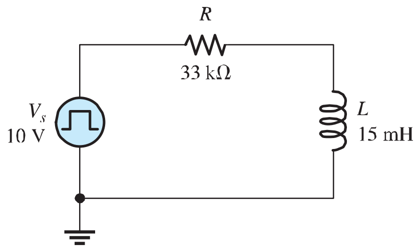

FIGURE 11–27

SOLUTION

(a) $\tau = \dfrac{L}{R} = \dfrac{15\text{ mH}}{33\text{ k}\Omega} = 0.454\ \mu\text{s}$

The period needs to be ten times longer than τ to observe the entire wave.

$$T = 10\tau = 4.54\ \mu\text{s}$$

$$f = \frac{1}{T} = \frac{1}{4.54\ \mu\text{s}} = \mathbf{220\ kHz}$$

(b) The voltage across the resistor has the same shape as the current waveform. The general shape was shown in Figure 11–26 and has a maximum value of 10 V (the same as V_s, assuming no winding resistance).

RELATED PROBLEM

What is the maximum voltage across the resistor for $f = 220$ kHz?

MULTISIM

Open Multisim file E11-07. Set the frequency to the calculated value and observe the voltage waveform across the inductor.

The Exponential Formulas

The formulas for the exponential current and voltage in an *RL* circuit are similar to those used in Chapter 9 for the *RC* circuit, and the universal exponential curves in Figure 9–36 apply to inductors as well as capacitors. The general formulas for *RL* circuits are stated as follows:

$$v = V_F + (V_i - V_F)e^{-Rt/L} \qquad (11\text{–}6)$$

$$i = I_F + (I_i - I_F)e^{-Rt/L} \qquad (11\text{–}7)$$

where V_F and I_F are the final values, V_i and I_i are the initial values, and v and i are the instantaneous values of the inductor voltage or current at time t.

INCREASING CURRENT The formula for the special case in which an increasing exponential current curve begins at zero ($I_i = 0$) is

$$i = I_F(1 - e^{-Rt/L}) \qquad (11\text{–}8)$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Using Equation 11–8, you can calculate the value of the increasing inductor current at any instant of time. You can calculate voltage by substituting v for i and V_F for I_F in Equation 11–8. The exponent $-Rt/L$ can also be written as $-t/(L/R) = -t/\tau$.

EXAMPLE 11–8

In Figure 11–28, determine the inductor current 30 μs after the switch is closed.

FIGURE 11–28

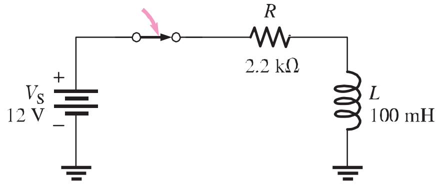

SOLUTION

The RL time constant is

$$\tau = \frac{L}{R} = \frac{100\text{ mH}}{2.2\text{ k}\Omega} = 45.5\ \mu\text{s}$$

The final current is

$$I_F = \frac{V_S}{R} = \frac{12\text{ V}}{2.2\text{ k}\Omega} = 5.45\text{ mA}$$

The initial current is zero. Notice that 30 μs is less than one time constant, so the current will reach less than 63% of its final value in that time.

$$i_L = I_F(1 - e^{-Rt/L}) = 5.45\text{ mA}(1 - e^{-0.66}) = \mathbf{2.63\ mA}$$

RELATED PROBLEM

In Figure 11–28, determine the inductor current 55 μs after the switch is closed.

DECREASING CURRENT The formula for the special case in which a decreasing exponential current has a final value of zero is

$$\boldsymbol{i = I_i e^{-Rt/L}} \qquad (11\text{–}9)$$

This formula can be used to calculate the decreasing inductor current at any instant.

EXAMPLE 11–9

In Figure 11–29, what is the current at each microsecond interval for one complete cycle of the input square wave as shown? After calculating the current at each time, sketch the current waveform.

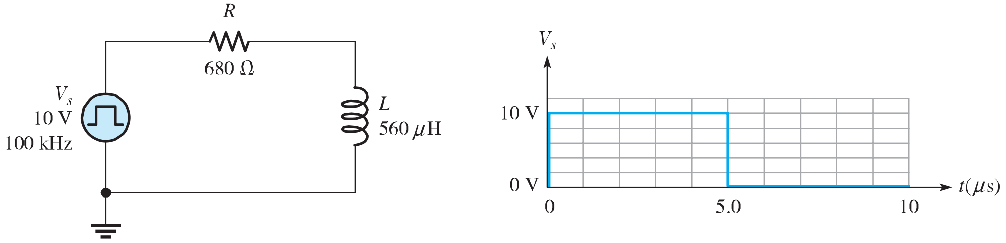

FIGURE 11–29

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

The *RL* time constant is

$$\tau = \frac{L}{R} = \frac{560\ \mu\text{H}}{680\ \Omega} = 0.824\ \mu\text{s}$$

When the pulse goes from 0 V to 10 V at $t = 0$, the current increases exponentially. The final current is

$$I_F = \frac{V_s}{R} = \frac{10\text{ V}}{680\ \Omega} = 14.7\text{ mA}$$

For the increasing current, $i = I_F(1 - e^{-Rt/L}) = I_F(1 - e^{-t/\tau})$.

At 1 μs: $i = 14.7\text{ mA}(1 - e^{-1\mu\text{s}/0.824\mu\text{s}}) = \mathbf{10.3\ mA}$

At 2 μs: $i = 14.7\text{ mA}(1 - e^{-2\mu\text{s}/0.824\mu\text{s}}) = \mathbf{13.4\ mA}$

At 3 μs: $i = 14.7\text{ mA}(1 - e^{-3\mu\text{s}/0.824\mu\text{s}}) = \mathbf{14.3\ mA}$

At 4 μs: $i = 14.7\text{ mA}(1 - e^{-4\mu\text{s}/0.824\mu\text{s}}) = \mathbf{14.6\ mA}$

At 5 μs: $i = 14.7\text{ mA}(1 - e^{-5\mu\text{s}/0.824\mu\text{s}}) = \mathbf{14.7\ mA}$

When the pulse goes from 10 V to 0 V at $t = 5\ \mu\text{s}$, the current decreases exponentially.

For the decreasing current,

$$i = I_i(e^{-Rt/L}) = I_i(e^{-t/\tau})$$

The initial current is the value at 5 μs, which is 14.7 mA.

At 6 μs: $i = 14.7\text{ mA}(e^{-1\mu\text{s}/0.824\mu\text{s}}) = \mathbf{4.37\ mA}$

At 7 μs: $i = 14.7\text{ mA}(e^{-2\mu\text{s}/0.824\mu\text{s}}) = \mathbf{1.30\ mA}$

At 8 μs: $i = 14.7\text{ mA}(e^{-3\mu\text{s}/0.824\mu\text{s}}) = \mathbf{0.38\ mA}$

At 9 μs: $i = 14.7\text{ mA}(e^{-4\mu\text{s}/0.824\mu\text{s}}) = \mathbf{0.11\ mA}$

At 10 μs: $i = 14.7\text{ mA}(e^{-5\mu\text{s}/0.824\mu\text{s}}) = \mathbf{0.03\ mA}$

Figure 11–30 is a graph of these results.

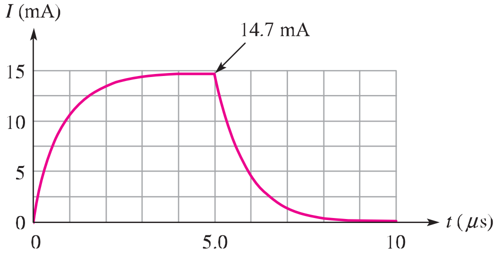

FIGURE 11–30

MULTISIM

Open Multisim file E11-09. Measure the current by placing a very small-value series resistor between the inductor and ground and observing the voltage across it.

RELATED PROBLEM

What is the current at 0.5 μs?

SECTION 11–4 CHECKUP

1. A 15 mH inductor with a winding resistance of 10 Ω has a constant direct current of 10 mA through it. What is the voltage drop across the inductor?
2. A 20 V dc source is connected to a series *RL* circuit with a switch. At the instant of switch closure, what are the values of *i* and v_L?
3. In the same circuit as in Question 2, after a time interval equal to 5τ from switch closure, what is v_L?
4. In a series *RL* circuit where $R = 1.0\text{ k}\Omega$ and $L = 500\ \mu\text{H}$, what is the time constant? Determine the current 0.25 μs after a switch connects 10 V across the circuit.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

11–5 INDUCTORS IN AC CIRCUITS

An inductor passes ac but with an amount of opposition that depends on the frequency of the ac.

After completing this section, you should be able to

- **Analyze inductive ac circuits**
  - **Define *inductive reactance***
  - **Determine the value of inductive reactance in a given circuit**
  - **Discuss instantaneous, true, and reactive power in an inductor**

Inductive Reactance, X_L

In Figure 11–31, an inductor is shown connected to a sinusoidal voltage source. When the source voltage is held at a constant amplitude value and its frequency is increased, the amplitude of the current decreases. Also, when the frequency of the source is decreased, the current amplitude increases.

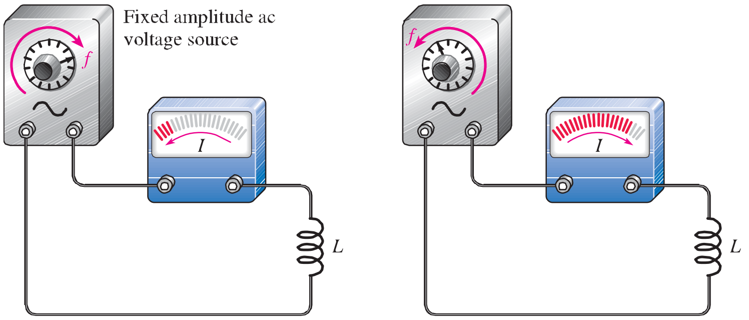

(a) Current decreases when the frequency increases.

(b) Current increases when the frequency decreases.

FIGURE 11–31 **The current in an inductive circuit varies inversely with the frequency of the source voltage.**

When the frequency of the source voltage increases, its rate of change also increases, as you already know. Now, if the frequency of the source voltage is increased, the frequency of the current also increases. According to Faraday's and Lenz's laws, this increase in frequency induces more voltage across the inductor in a direction to oppose the current and causes it to decrease in amplitude. Similarly, a decrease in frequency will cause an increase in current.

A decrease in the amount of current with an increase in frequency for a fixed amount of voltage indicates that opposition to the current has increased. Thus, the inductor offers opposition to current, and that opposition varies *directly* with frequency.

Inductive reactance is the opposition to sinusoidal current in an inductor.

The symbol for inductive reactance is X_L, and its unit is the ohm (Ω).

You have just seen how frequency affects the opposition to current (inductive reactance) in an inductor. Now let's see how the inductance, L, affects the reactance. Figure 11–32(a) shows that when a sinusoidal voltage with a fixed amplitude and fixed frequency is applied to a 1 mH inductor, there is a certain amount of alternating current. When the inductance value is increased to 2 mH, the current decreases, as shown in part (b). Thus, when the inductance is increased, the opposition to current (inductive reactance) increases. So not only is the inductive reactance *directly* proportional to frequency, but it is also *directly* proportional to inductance. This relationship can be stated as follows:

6238262 2013/12/10 12:27:66.5

X_L is proportional to fL.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

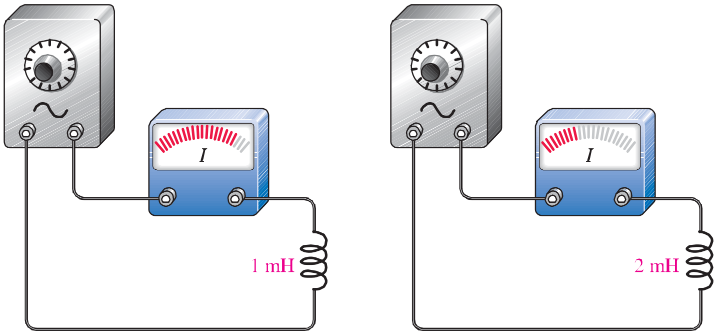

(a) Less inductance, more current. (b) More inductance, less current.

FIGURE 11–32 **For a fixed voltage and fixed frequency, the current varies inversely with the inductance value.**

Gyrators

Inductors tend to be nonideal components and larger inductors do not lend themselves to integrated circuits because of physical size. A circuit that mimics an inductor is useful for active filter applications and network synthesis. The circuit uses resistors and capacitors and two operational amplifiers to form an equivalent circuit for an inductor. A gyrator can be included within an IC.

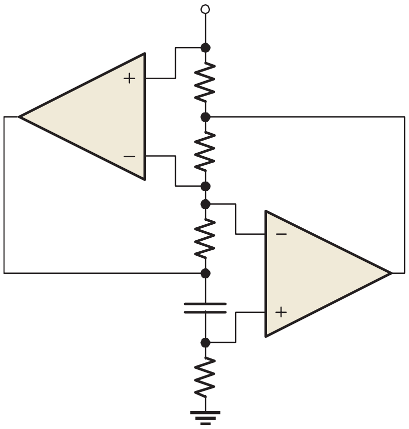

SYSTEM NOTE

It can be proven that the constant of proportionality is 2π, so the formula for inductive reactance (X_L) is

$$X_L = 2\pi f L \quad (11\text{–}10)$$

X_L is in ohms when f is in hertz and L is in henries. As with capacitive reactance, the 2π term comes from the relationship of the sine wave to rotational motion.

EXAMPLE 11–10

A sinusoidal voltage is applied to the circuit in Figure 11–33. The frequency is 10 kHz. Determine the inductive reactance.

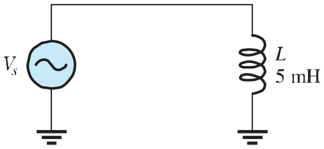

FIGURE 11–33

SOLUTION

Convert 10 kHz to 10×10^3 Hz and 5 mH to 5×10^{-3} H. Then, the inductive reactance is

$$X_L = 2\pi f L = 2\pi(10 \times 10^3\ \text{Hz})(5 \times 10^{-3}\ \text{H}) = \mathbf{314\ \Omega}$$

RELATED PROBLEM

What is X_L in Figure 11–33 if the frequency is increased to 35 kHz?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Reactance for Series Inductors

As given in Equation 11–3, the total inductance of series inductors is the sum of the individual inductances. Because reactance is directly proportional to the inductance, the total reactance of series inductors is the sum of the individual reactances.

$$X_{L(tot)} = X_{L1} + X_{L2} + X_{L3} + \cdots + X_{Ln} \qquad (11\text{–}11)$$

Notice that Equation 11–11 has the same form as Equation 11–3. It also has the same form as the formulas for finding the total opposition to current such as the total resistance of series resistors or the total reactance of series capacitors. When combining the resistance or reactance of the same type of component in series (resistors, inductors, or capacitors), you simply add the individual oppositions to obtain the total.

Reactance for Parallel Inductors

In an ac circuit with parallel inductors, Equation 11–4 was given to find the total inductance. It stated that the total inductance is the reciprocal of the sum of the reciprocals of the inductors. Likewise, the total inductive reactance is the reciprocal of the sum of the reciprocals of the individual reactances.

$$X_{L(tot)} = \frac{1}{\dfrac{1}{X_{L1}} + \dfrac{1}{X_{L2}} + \dfrac{1}{X_{L3}} + \cdots + \dfrac{1}{X_{Ln}}} \qquad (11\text{–}12)$$

Notice that Equation 11–12 has the same form as Equation 11–4. It also has the same form as the formulas for finding the total resistance of parallel resistors or the total reactance of parallel capacitors. When combining the resistance or reactance of the same type of component in parallel (resistors, inductors, or capacitors), you take the reciprocal of the sum of the reciprocals to obtain the total opposition.

For two inductors in parallel, Equation 11–12 can be reduced to the product-over-sum form.

$$X_{L(tot)} = \frac{X_{L1}X_{L2}}{X_{L1} + X_{L2}}$$

EXAMPLE 11–11

What is the total inductive reactance of each circuit of Figure 11–34?

FIGURE 11–34

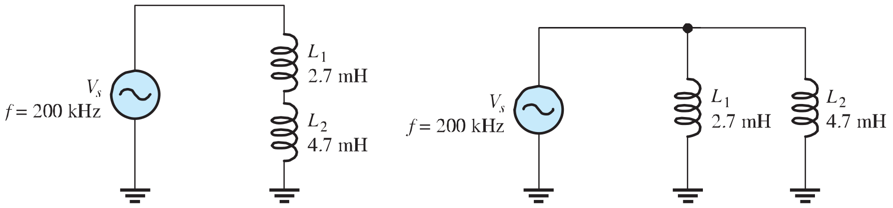

(a) (b)

SOLUTION

The reactances of the individual inductors are the same in both circuits.

$$X_{L1} = 2\pi f L_1 = 2\pi(200\text{ kHz})(2.7\text{ mH}) = 3.39\text{ k}\Omega$$

$$X_{L2} = 2\pi f L_2 = 2\pi(200\text{ kHz})(4.7\text{ mH}) = 5.91\text{ k}\Omega$$

For the series inductors in Figure 11–34(a), the total reactance is the sum of X_{L1} and X_{L2}, as given in Equation 11–11.

$$X_{L(tot)} = X_{L1} + X_{L2} = 3.39\text{ k}\Omega + 5.91\text{ k}\Omega = \mathbf{9.30\text{ k}\Omega}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

For the inductors in parallel in Figure 11–34(b), determine the total reactance from the product-over-sum rule using X_{L1} and X_{L2}.

$$X_{L(tot)} = \frac{X_{L1}X_{L2}}{X_{L1} + X_{L2}} = \frac{(3.39\text{ k}\Omega)(5.91\text{ k}\Omega)}{3.39\text{ k}\Omega + 5.91\text{ k}\Omega} = \mathbf{2.15\text{ k}\Omega}$$

You can also obtain the total reactance for either series or parallel inductors by first finding the total inductance and then substituting in Equation 11–10 to find the total reactance.

For the series inductors,

$$L_{\text{T}} = L_1 + L_2 = 2.7\text{ mH} + 4.7\text{ mH} = 7.4\text{ mH}$$

$$X_{L(tot)} = 2\pi fL_{\text{T}} = 2\pi(200\text{ kHz})(7.4\text{ mH}) = \mathbf{9.30\text{ k}\Omega}$$

For the parallel inductors,

$$L_{\text{T}} = \frac{L_1L_2}{L_1 + L_2} = \frac{(2.7\text{ mH})(4.7\text{ mH})}{2.7\text{ mH} + 4.7\text{ mH}} = 1.71\text{ mH}$$

$$X_{L(tot)} = 2\pi fL_{\text{T}} = 2\pi(200\text{ kHz})(1.71\text{ mH}) = \mathbf{2.15\text{ k}\Omega}$$

RELATED PROBLEM

What is the total inductive reactance for each circuit in Figure 11–34 if $L_1 = 1$ mH and L_2 is unchanged?

OHM'S LAW The reactance of an inductor is analogous to the resistance of a resistor. In fact, X_L, just like X_C and R, is expressed in ohms. Since inductive reactance is a form of opposition to current, Ohm's law applies to inductive circuits as well as to resistive circuits and capacitive circuits; and it is stated as follows:

$$\boldsymbol{I = \frac{V}{X_L}} \qquad (11\text{–}13)$$

When applying Ohm's law in ac circuits, you must express both the current and the voltage in the same way, that is, both in rms, both in peak, and so on.

EXAMPLE 11–12

Determine the rms current in Figure 11–35.

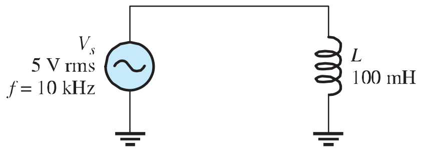

FIGURE 11–35

6238262 2013/12/10 12.27.66.5

SOLUTION

Convert 10 kHz to 10×10^3 Hz and 100 mH to 100×10^{-3} H. Then calculate X_L.

$$X_L = 2\pi fL = 2\pi(10 \times 10^3\text{ Hz})(100 \times 10^{-3}\text{ H}) = 6283\ \Omega$$

Apply Ohm's law.

$$I_{\text{rms}} = \frac{V_{\text{rms}}}{X_L} = \frac{5\text{ V}}{6283\ \Omega} = \mathbf{796\ \mu A}$$

RELATED PROBLEM

Determine the rms current in Figure 11–35 for the following values: $V_{\text{rms}} = 12$ V, $f = 4.9$ kHz, and $L = 680\ \mu$H.

MULTISIM

Open Multisim file E11-12. Measure the rms current and compare to the calculated value. Change the circuit values to those given in the related problem and measure the rms current.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Current Lags Inductor Voltage by 90°

As you know, a sinusoidal voltage has a maximum rate of change at its zero crossings and a zero rate of change at the peaks. From Faraday's law you know that the amount of voltage induced across a coil is directly proportional to the rate at which the current is changing. Therefore, the coil voltage is maximum at the zero crossings of the current where the rate of change of the current is the greatest. Also, the amount of voltage is zero at the peaks of the current where the rate of change is zero. This phase relationship is illustrated in Figure 11–36. As you can see, the current peaks occur a quarter cycle after the voltage peaks. Thus, the current *lags* the voltage by 90°. Recall that in a capacitor, the current *leads* the voltage by 90°.

FIGURE 11–36 **Current is always lagging the inductor voltage by 90°.**

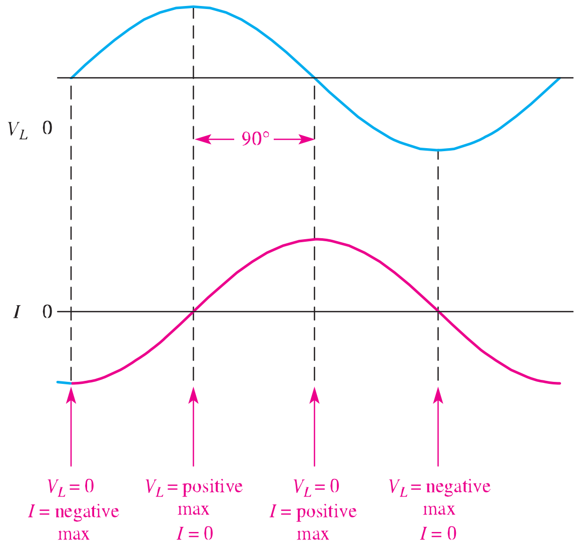

Power in an Inductor

As discussed earlier, an inductor stores energy in its magnetic field when there is current through it. An ideal inductor (assuming no winding resistance) does not dissipate energy; it only stores it. When an ac voltage is applied to an inductor, energy is stored by the inductor during a portion of the cycle; then the stored energy is returned to the source during another portion of the cycle. No net energy is lost in an ideal inductor due to conversion to heat. Figure 11–37 shows the power curve that results from one cycle of inductor current and voltage. Compare the power curve for an inductor with that of a capacitor in Figure 9–50. Notice that main difference is that voltage and current are interchanged between the plots.

FIGURE 11–37 **Power curve for an inductor.**

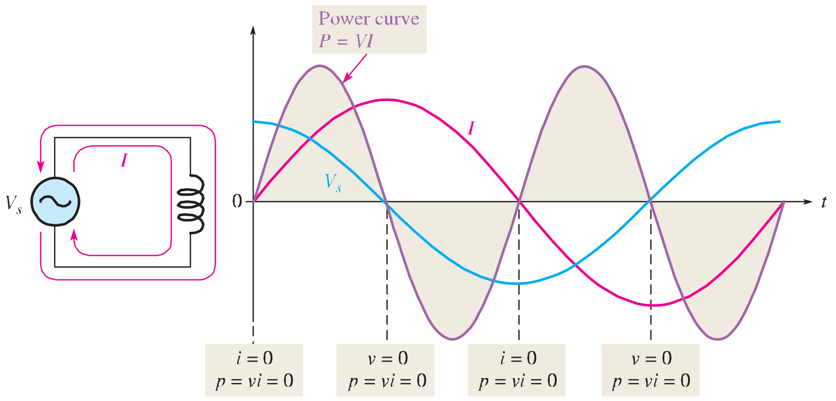

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

INSTANTANEOUS POWER (p) The product of instantaneous voltage, v, and instantaneous current, i, gives instantaneous power, p. At points where v or i is zero, p is also zero. When both v and i are positive, p is also positive. When either v or i is positive and the other negative, p is negative. When both v and i are negative, p is positive. As you can see in Figure 11–37, the power follows a sinusoidal-shaped curve. Positive values of power indicate that energy is stored by the inductor. Negative values of power indicate that energy is returned from the inductor to the source. Note that the power fluctuates at a frequency twice that of the voltage or current, as energy is alternately stored and returned to the source.

TRUE POWER (P_{true}) Ideally, all of the energy stored by an inductor during the positive portion of the power cycle is returned to the source during the negative portion. No net energy is lost due to conversion to heat in the inductor, so the power is zero. Actually, because of winding resistance in a practical inductor, some power is always dissipated, and there is a very small amount of true power, which can normally be neglected.

$$P_{\text{true}} = I_{\text{rms}}^2 R_W \tag{11–14}$$

REACTIVE POWER (P_r) The rate at which an inductor stores or returns energy is called its **reactive power**, P_r, with a unit of VAR (volt-ampere reactive). The reactive power is a nonzero quantity because at any instant in time the inductor is actually taking energy from the source or returning energy to it. Reactive power does not represent an energy loss due to conversion to heat. The following formulas apply:

$$P_r = V_{\text{rms}} I_{\text{rms}} \tag{11–15}$$

$$P_r = \frac{V_{\text{rms}}^2}{X_L} \tag{11–16}$$

$$P_r = I_{\text{rms}}^2 X_L \tag{11–17}$$

EXAMPLE 11–13

A 10 V rms signal with a frequency of 10 kHz is applied to a 10 mH coil with a winding resistance of 40 Ω. Determine the reactive power (P_r) and the true power (P_{true}).

SOLUTION

First, calculate the inductive reactance and current values.

$$X_L = 2\pi f L = 2\pi(10\text{ kHz})(10\text{ mH}) = 628\ \Omega$$

$$I = \frac{V_s}{X_L} = \frac{10\text{ V}}{628\ \Omega} = 15.9\text{ mA}$$

Then using Equation 11–17,

$$P_r = I^2 X_L = (15.9\text{ mA})^2(628\ \Omega) = \mathbf{159\ mVAR}$$

The true power is

$$P_{\text{true}} = I^2 R_W = (15.9\text{ mA})^2(40\ \Omega) = \mathbf{10.1\ mW}$$

RELATED PROBLEM

What happens to the reactive power if the frequency increases?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The Quality Factor (Q)

The **quality factor (Q)** is the ratio of the reactive power in the inductor to the true power in the winding resistance of the coil. It is a ratio of the power in L to the power in R_W. The quality factor is important in resonant circuits, which are studied in Chapter 13. A formula for Q is developed as follows:

6238262 2013/12/10 12.27.66.5

$$Q = \frac{\text{reactive power}}{\text{true power}} = \frac{P_r}{P_{\text{true}}} = \frac{I^2 X_L}{I^2 R_W}$$

The I^2 terms cancel, leaving

$$Q = \frac{X_L}{R_W} \qquad (11\text{–}18)$$

Note that Q is a ratio of like units and, therefore, has no unit itself. The quality factor is also known as unloaded Q because it is defined with no load across the coil. Notice also that Q depends on the frequency because X_L is dependent on frequency.

SECTION 11–5 CHECKUP

1. State the phase relationship between current and voltage in an inductor.
2. Calculate X_L for $f = 500$ kHz and $L = 1.0$ mH
3. At what frequency is the reactance of a 50 μH inductor equal to 800 Ω?
4. Calculate the rms current in Figure 11–38.
5. An ideal 50 mH inductor is connected to a 12 V rms source. What is the true power? What is the reactive power at a frequency of 1 kHz?

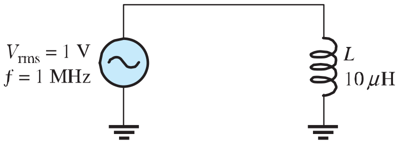

FIGURE 11–38

11–6 INDUCTOR APPLICATIONS

Inductors are not as versatile as capacitors and tend to be more limited in their application due, in part, to size, cost factors, and nonideal behavior (internal resistance, etc.). One of the most common applications for inductors is noise reduction applications.

After completing this section, you should be able to

- **Discuss some inductor applications**
  - **Discuss two ways in which noise enters a circuit**
  - **Describe the suppression of electromagnetic interference (EMI)**
  - **Explain how a ferrite bead is used**
  - **Discuss the basics of tuned circuits**

Noise Suppression

One of the most important applications of inductors has to do with suppressing unwanted electrical noise. The inductors used in these applications are generally wound on a closed core to avoid having the inductor become a source of radiated noise itself. Two types of noise are conductive noise and radiated noise.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

CONDUCTIVE NOISE Many systems have common conductive paths connecting different parts of the system, which can conduct high frequency noise from one part of the system to another. Consider the case of two circuits connected with common lines as shown in Figure 11–39(a). A path for high frequency noise exists though the common grounds, creating a condition known as a *ground loop*. Ground loops are particularly a problem in instrumentation systems, where a transducer may be located a distance from the recording system and noise current in the ground can affect the signal.

If the signal of interest changes slowly, a special inductor, called a *longitudinal choke,* can be installed in the signal line as shown in Figure 11–39(b). The longitudinal choke is a form of transformer (covered in Chapter 14) that acts as an inductor in each signal line. The ground loop sees a high impedance path, thus reducing the noise, while the low-frequency signal is coupled through the low impedance of the choke.

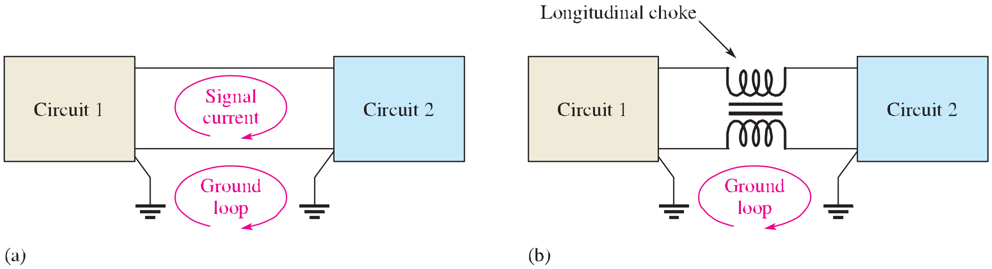

FIGURE 11–39

Switching circuits also tend to generate high-frequency noise (above 10 MHz) by virtue of the high-frequency components present. (Recall from Section 8–5 that a fast pulse contains many high-frequency harmonics.) Certain types of power supplies use high-speed switching circuits, which are a source of both conductive and radiated noise. Because an inductor's impedance increases with frequency, inductors are good for blocking electrical noise from these supplies, which should carry only dc. Inductors are frequently installed in the power supply lines to suppress this conductive noise so that one circuit does not adversely affect another circuit. One or more capacitors may also be used in conjunction with the inductor to improve filtering action.

HANDS ON TIP
When breadboarding circuits that include small inductors, it is best to use encapsulated inductors for structural strength. Inductors generally have extremely fine coil wire that is connected to a much larger size of lead wires.

In unencapsulated inductors, these contact points are very vulnerable to breaking if the inductor is frequently inserted and removed from a protoboard.

RADIATED NOISE Noise can also enter a circuit by way of an electromagnetic field. The noise source can be an adjacent circuit or a nearby power supply. There are several approaches to reducing the effects of radiated noise. Usually, the first step is to determine the cause of the noise and isolate it using shielding or filtering. Inductors are widely used in filters that are used to suppress radio-frequency noise. The inductor used for noise suppression must be carefully selected so as not to become a source of radiated noise itself. For high frequencies (>20 MHz), inductors wound on highly permeable toroidal cores are widely used, as they tend to keep the magnetic flux restricted to the core.

RF Chokes

Inductors used for the purpose of blocking very high frequencies are called *radio frequency (RF) chokes*. RF chokes are used for conductive or radiated noise. They are special inductors designed to block high frequencies from getting into or leaving parts of a system by providing a high impedance path for high frequencies. In general, the choke is placed in series with the line for which RF suppression is required. Depending on the frequency of the interference, different types of chokes are required. A common type of electromagnetic interference (EMI) filter wraps the signal line on a toroidal core several times. The toroidal configuration is desired because it contains the magnetic field so that the choke does not become a source of noise itself.

Another common type of RF choke is a ferrite bead. All wires have inductance, and the ferrite bead is a small ferromagnetic material that is strung onto the wire to increase its

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

inductance. The impedance presented by the bead is a function of both the material and the frequency, as well as the size of the bead. It is an effective and inexpensive "choke" for high frequencies. Ferrite beads are common in high-frequency communication systems. Sometimes several are strung together in series to increase the effective inductance.

Tuned Circuits

Inductors are used in conjunction with capacitors to provide frequency selection in communications systems. These tuned circuits allow a narrow band of frequencies to be selected while all other frequencies are rejected. The tuners in your TV and radio receivers are based on this principle and permit you to select one channel or station out of the many that are available.

Frequency selectivity is based on the fact that the reactances of both capacitors and inductors depend on the frequency and on the interaction of these two components when connected in series or parallel. Since the capacitor and the inductor produce opposite phase shifts, their combined opposition to current can be used to obtain a desired response at a selected frequency. Tuned (resonant) *RLC* circuits are covered in detail in Chapter 13.

SECTION 11–6 CHECKUP

1. Name two types of unwanted noise.
2. What do the letters EMI stand for?
3. How is a ferrite bead used?

SUMMARY

- Inductance is a measure of a coil's ability to establish an induced voltage as a result of a change in its current.
- An inductor opposes a change in its own current.
- Faraday's law states that relative motion between a magnetic field and a coil induces a voltage across the coil.
- Lenz's law states that the polarity of induced voltage is such that the resulting induced current is in a direction that opposes the change in the magnetic field that produced it.
- Energy is stored by an inductor in its magnetic field.
- One henry is the amount of inductance when current, changing at the rate of one ampere per second, induces one volt across the inductor.
- Inductance is directly proportional to the square of the turns, the permeability, and the cross-sectional area of the core. It is inversely proportional to the length of the core.
- The permeability of a core material is an indication of the ability of the material to establish a magnetic field.
- Inductors add in series.
- Total parallel inductance is less than that of the smallest inductor in parallel.
- The time constant for a series *RL* circuit is the inductance divided by the resistance.
- In an *RL* circuit, the increasing or decreasing voltage and current in an inductor make an approximately 63% change during each time-constant interval.
- Increasing and decreasing current and voltage follow exponential curves.
- Voltage leads current by 90° in an inductor.
- Inductive reactance (X_L) is directly proportional to frequency and inductance.
- The true power in an inductor is zero; that is, there is no energy conversion to heat in an ideal inductor, only in its winding resistance.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

KEY TERMS

Key terms and other bold terms in the chapter are defined in the end-of-book glossary.

Coil A common term for an inductor.

Henry (H) The unit of inductance.

Induced voltage Voltage produced as a result of a changing magnetic field.

Inductance The property of an inductor whereby a change in current causes the inductor to produce a voltage that opposes the change in current.

Inductive reactance The opposition of an inductor to sinusoidal current. The unit is the ohm.

Inductor A passive electrical component, formed by a coil of wire, that exhibits the property of inductance.

Quality factor (*Q*) The ratio of reactive power to true power.

***RL* time constant** A fixed time interval, set by the *L* and *R* values, that determines the time response of a circuit.

Winding The loops or turns of wire in an inductor.

Winding resistance The resistance of the length of wire that makes up a coil.

KEY FORMULAS

(11–1)	$W = \frac{1}{2}LI^2$	Energy stored by an inductor
(11–2)	$L = \frac{N^2\mu A}{l}$	Inductance in terms of physical parameters
(11–3)	$L_T = L_1 + L_2 + L_3 + \cdots + L_n$	Series inductance
(11–3)	$L_T = \dfrac{1}{\frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} + \cdots + \frac{1}{L_n}}$	Total parallel inductance
(11–5)	$\tau = \frac{L}{R}$	*RL* time constant
(11–6)	$v = V_F + (V_i - V_F)e^{-Rt/L}$	Exponential voltage (general)
(11–7)	$i = I_F + (I_i - I_F)e^{-Rt/L}$	Exponential current (general)
(11–8)	$i = I_F(1 - e^{-Rt/L})$	Increasing exponential current beginning at zero
(11–9)	$i = I_i e^{-Rt/L}$	Decreasing exponential current ending at zero
(11–10)	$X_L = 2\pi fL$	Inductive reactance
(11–11)	$X_{L(tot)} = X_{L1} + X_{L2} + X_{L3} + \cdots + X_{Ln}$	Series inductive reactance
(11–12)	$X_{L(tot)} = \dfrac{1}{\frac{1}{X_{L1}} + \frac{1}{X_{L2}} + \frac{1}{X_{L3}} + \cdots + \frac{1}{X_{Ln}}}$	Parallel inductive reactance
(11–13)	$I = \frac{V}{X_L}$	Ohm's law
(11–14)	$P_{\text{true}} = I^2_{\text{rms}}R_W$	True power
(11–15)	$P_r = V_{\text{rms}}I_{\text{rms}}$	Reactive power

6238262 2013/12/10 12.27.66.5

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

(11–16)	$P_r = \dfrac{V^2_{\mathrm{rms}}}{X_L}$	Reactive power
(11–17)	$P_r = I^2_{\mathrm{rms}} X_L$	Reactive power
(11–18)	$Q = \dfrac{X_L}{R_W}$	Quality factor

TRUE/FALSE QUIZ

Answers are at the end of the chapter.

1. Lenz's law states that the amount of voltage induced in a coil is proportional to the rate of change of the magnetic field with respect to the coil.

2. An ideal inductor has no winding resistance.

3. The total inductance of two parallel inductors is equal to the product-over-sum of the individual inductors.

4. The total inductance of parallel inductors is always less than that of the smallest inductor.

5. The time constant of an *RL* circuit is given by the formula $\tau = R/L$.

6. If a series *RL* circuit is connected to a dc source, the maximum current is limited by the total inductance.

7. Kirchhoff's voltage law does not apply to inductive circuits.

8. Inductive reactance is directly proportional to the frequency.

9. In an ac inductive circuit, the current lags the voltage in the inductor.

10. The frequency of the power curve for an inductive circuit is equal to the frequency of the applied voltage.

SELF-TEST

Answers are at the end of the chapter.

1. An inductance of 0.05 μH is larger than
(a) 0.0000005 H **(b)** 0.000005 H **(c)** 0.000000008 H **(d)** 0.00005 mH

2. An inductance of 0.33 mH is smaller than
(a) 33 μH **(b)** 330 μH **(c)** 0.05 mH **(d)** 0.0005 H

3. When the current through an inductor increases, the amount of energy stored in the electromagnetic field
(a) decreases **(b)** remains constant **(c)** increases **(d)** doubles

4. When the current through an inductor doubles, the stored energy
(a) doubles **(b)** quadruples **(c)** is halved **(d)** does not change

5. The winding resistance of a coil can be decreased by
(a) reducing the number of turns **(b)** using a larger wire
(c) changing the core material **(d)** either answer (a) or (b)

6. The inductance of an iron-core coil increases if
(a) the number of turns is increased **(b)** the iron core is removed
(c) the length of the core is increased **(d)** larger wire is used

7. Four 10 mH inductors are in series. The total inductance is
(a) 40 mH **(b)** 2.5 mH **(c)** 40,000 μH **(d)** answers (a) and (c)

8. A 1 mH, a 3.3 mH, and a 0.1 mH inductor are connected in parallel. The total inductance is
(a) 4.4 mH **(b)** greater than 3.3 mH
(c) less than 0.1 mH **(d)** answers (a) and (b)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

9. An inductor, a resistor, and a switch are connected in series to a 12 V battery. At the instant the switch is closed, the inductor voltage is

(a) 0 V **(b)** 12 V **(c)** 6 V **(d)** 4 V

10. A sinusoidal voltage is applied across an inductor. When the frequency of the voltage is increased, the current

(a) decreases **(b)** increases **(c)** does not change **(d)** momentarily goes to zero

11. An inductor and a resistor are in series with a sinusoidal voltage source. The frequency is set so that the inductive reactance is equal to the resistance. If the frequency is increased, then

(a) $V_R > V_L$ **(b)** $V_L < V_R$ **(c)** $V_L = V_R$ **(d)** $V_L > V_R$

TROUBLESHOOTING: SYMPTOM AND CAUSE

The purpose of these exercises is to help develop thought processes essential to troubleshooting. Answers are at the end of the chapter.

Determine the cause for each set of symptoms. Refer to Figure 11–40.

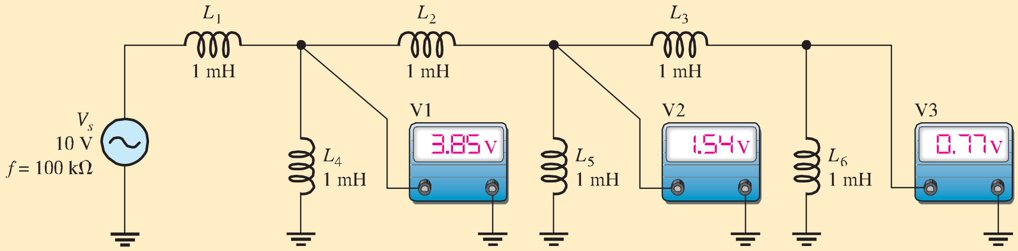

FIGURE 11–40 The ac meters indicate the correct readings for this circuit.

1. *Symptom:* All of the voltmeter readings are 0 V.

Cause:

(a) The source is off or faulty.
(b) L_1 is open.
(c) either (a) or (b)

2. *Symptom:* All of the voltmeter readings are 0 V.

Cause:

(a) L_4 is completely shorted.
(b) L_5 is completely shorted.
(c) L_6 is completely shorted.

3. *Symptom:* The voltmeter 1 reading is 5 V, and the voltmeter 2 and voltmeter 3 readings are 0 V.

Cause:

(a) L_4 is open.
(b) L_2 is open.
(c) L_5 is shorted.

4. *Symptom:* The voltmeter 1 reading is 4 V, the voltmeter 2 reading is 2 V, and the voltmeter 3 reading is 0 V.

Cause:

(a) L_3 is open.
(b) L_6 is shorted.
(c) either (a) or (b)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

5. *Symptom:* The voltmeter 1 reading is 4 V, the voltmeter 2 reading is 2 V, and the voltmeter 3 reading is 2 V.

   Cause:

   (a) L_3 is shorted.
   (b) L_6 is open.
   (c) either (a) or (b)

PROBLEMS

Answers to odd-numbered problems are at the end of the book.

BASIC PROBLEMS

SECTION 11–1 The Basic Inductor

1. Convert the following to millihenries:

   (a) 1 H
   (b) 250 μH
   (c) 10 μH
   (d) 0.0005 H

2. Convert the following to microhenries:

   (a) 300 mH
   (b) 0.08 H
   (c) 5 mH
   (d) 0.00045 mH

3. How many turns are required to produce 30 mH with a coil wound on a cylindrical core having a cross-sectional area of 10×10^{-5} m^2 and a length of 0.05 m? The core has a permeability of 1.26×10^{-6}.

4. A 12 V battery is connected across a coil with a winding resistance of 120 Ω. How much current is there in the coil?

5. How much energy is stored by a 100 mH inductor with a current of 1 A?

6. The current through a 100 mH coil is changing at a rate of 200 mA/s. How much voltage is induced across the coil?

SECTION 11–3 Series and Parallel Inductors

7. Five inductors are connected in series. The lowest value is 5 μH. If the value of each inductor is twice that of the preceding one, and if the inductors are connected in order of ascending values, what is the total inductance?

8. Suppose that you require a total inductance of 50 mH. You have available a 10 mH coil and a 22 mH coil. How much additional inductance do you need?

9. Determine the total parallel inductance for the following coils in parallel: 75 μH, 50 μH, 25 μH, and 15 μH.

10. You have a 12 mH inductor, and it is your smallest value. You need an inductance of 8 mH. What value can you use in parallel with the 12 mH to obtain 8 mH?

11. Determine the total inductance of each circuit in Figure 11–41.

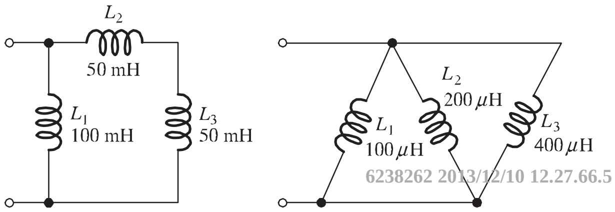

6238262 2013/12/10 12.27.66.5

(a) (b)

FIGURE 11–41

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

12. Determine the total inductance of each circuit in Figure 11–42.

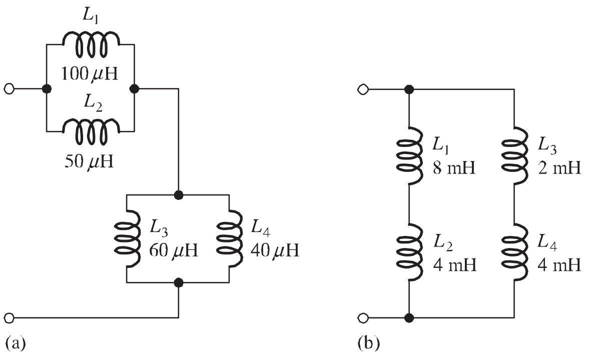

FIGURE 11–42

SECTION 11–4 Inductors in DC Circuits

13. Determine the time constant for each of the following series *RL* combinations:
(a) $R = 100\ \Omega, L = 100\ \mu\text{H}$ **(b)** $R = 4.7\ \text{k}\Omega, L = 10\ \text{mH}$
(c) $R = 1.5\ \text{k}\Omega, L = 3\ \text{mH}$

14. In a series *RL* circuit, determine how long it takes the current to build up to its full value for each of the following:
(a) $R = 56\ \Omega, L = 50\ \mu\text{H}$ **(b)** $R = 3300\ \Omega, L = 15\ \text{mH}$
(c) $R = 22\ \text{k}\Omega, L = 100\ \text{mH}$

15. In the circuit of Figure 11–43, there is initially no current. Determine the inductor voltage at the following times after the switch is closed:
(a) 10 μs **(b)** 20 μs **(c)** 30 μs **(d)** 40 μs **(e)** 50 μs

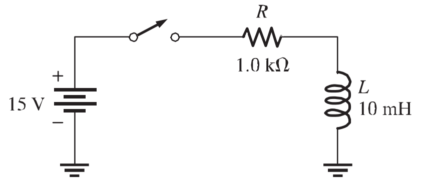

FIGURE 11–43

16. In Figure 11–44, calculate the current at each of the following times. Assume an ideal inductor and voltage source.
(a) 10 μs **(b)** 20 μs **(c)** 30 μs

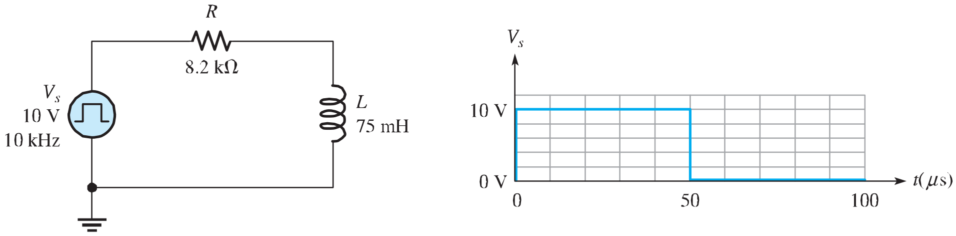

FIGURE 11–44

SECTION 11–6 Inductors in AC Circuits

17. Find the total reactance for each circuit in Figure 11–41 when a voltage with a frequency of 500 kHz is applied across the terminals.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

18. Find the total reactance for each circuit in Figure 11–42 when a 400 kHz signal is applied.

19. Determine the total rms current in Figure 11–45. What are the currents through L_2 and L_3?

FIGURE 11–45

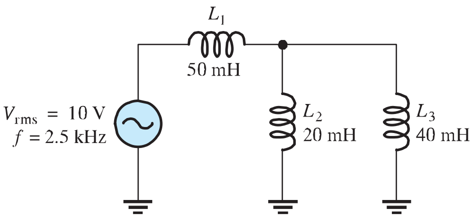

20. What frequency will produce a total rms current of 500 mA in each circuit of Figure 11–42 with an rms input voltage of 10 V?

21. Determine the reactive power in Figure 11–45, neglecting the winding resistance.

ADVANCED PROBLEMS

22. Determine the time constant for the circuit in Figure 11–46.

FIGURE 11–46

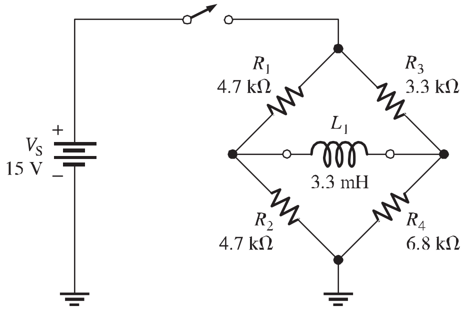

23. What is the voltage across the inductor in Figure 11–44 at each of the following times?
(a) 60 μs **(b)** 70 μs **(c)** 80 μs

24. What is the voltage across the resistor in Figure 11–44 at a time of 60 μs?

25. **(a)** What is the current in the inductor 1.0 μs after the switch closes in Figure 11–46.
(b) What is the current after 5τ have elapsed?

26. For the circuit in Figure 11–46, assume the switch has been closed for more than 5τ and is opened. What is the current in the inductor 1.0 μs after the switch is opened?

27. In System Example 11–2, assume f = 550 kHz. What is the reactance of each inductor? What is the total reactance of the two series inductors?

28. For System Example 11–2, assume you need to change L_1 to provide 8% feedback. What is the new value?

MULTISIM

MULTISIM TROUBLESHOOTING PROBLEMS

29. Open file P11-29; files are found at www.pearsonhighered.com/floyd. Test the circuit. If there is a fault, identify it.

30. Open file P11-30 and test the circuit. If there is a fault, identify it.

31. Determine if there is a fault in the circuit in file P11-31. If so, identify it.

32. Find and specify any faulty component in the circuit in file P11-32.

33. Is there a short or an open in the circuit in file P11-33? If so, identify the component that is faulty.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

ANSWERS TO SECTION CHECKUPS

SECTION 11–1 The Basic Inductor

1. Parameters that determine inductance are turns of wire, permeability, cross-sectional area, and core length.
2. (a) When N increases, L increases.
   (b) When the core length increases, L decreases.
   (c) When the cross-sectional area decreases, L decreases.
   (d) For an air core, L decreases.
3. All wire has some resistance, and because inductors are made from turns of wire, there is always resistance in the winding.
4. Adjacent turns of wire in a coil act as plates of a capacitor and produce a small capacitance.

SECTION 11–2 Types of Inductors

1. Two categories of inductors are fixed and variable.
2. (a) air core (b) iron core (c) variable

6238262 2013/12/10 12.27.66.5

SECTION 11–3 Series and Parallel Inductors

1. Inductances are added in series.
2. $L_T = 2600\ \mu H$
3. $L_T = 5 \times 100\ mH = 500\ mH$
4. The total parallel inductance is smaller than that of the smallest-value individual inductor in parallel.
5. True
6. (a) $L_T = 7.69\ mH$ (b) $L_T = 24\ \mu H$

SECTION 11–4 Inductors in DC Circuits

1. $V_L = (10\ mA)(10\ \Omega) = 100\ mV$
2. Initially, $i = 0\ V$, $v_L = 20\ V$
3. After 5τ, $v_L = 0\ V$
4. $\tau = 500\ \mu H/1.0\ k\Omega = 500\ ns$, $i_L = 3.93\ mA$

SECTION 11–5 Inductors in AC Circuits

1. Voltage leads current by 90° in an inductor.
2. $X_L = 2\pi fL = 3.14\ k\Omega$
3. $f = X_L/2\pi L = 2.55\ MHz$
4. $I_{rms} = 15.9\ mA$
5. $P_{true} = 0\ W$, $P_r = 458\ mVAR$

SECTION 11–6 Inductor Applications

1. Conductive and radiated
2. Electromagnetic interference
3. It is placed on a wire to increase its inductance, creating an RF choke.

ANSWERS TO RELATED PROBLEMS FOR EXAMPLES

11–1. 157 mH
11–2. 150 μH
11–3. 20.3 μH
11–4. 227 ns

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

11–5. $I_F = 17.6$ mA, $\tau = 147$ ns
At 1τ: $i = 11.1$ mA; $t = 147$ ns
At 2τ: $i = 15.1$ mA; $t = 294$ ns
At 3τ: $i = 16.7$ mA; $t = 441$ ns
At 4τ: $i = 17.2$ mA; $t = 588$ ns
At 5τ: $i = 17.4$ mA; $t = 735$ ns

11–6. At 0.2 ms: $i = 0.215$ mA
At 0.8 ms: $i = 0.0125$ mA

11–7. 10 V if R_W is neglected.

11–8. 3.83 mA

11–9. 6.7 mA

11–10. 1100 Ω

11–11. (a) 7.17 kΩ; (b) 1.04 kΩ

11–12. 573 mA

11–13. P_r decreases.

ANSWERS TO TRUE/FALSE QUIZ

1. F **2.** T **3.** T **4.** T **5.** F **6.** F **7.** F
8. F **9.** T **10.** F

ANSWERS TO SELF-TEST

1. (c) **2.** (d) **3.** (c) **4.** (b) **5.** (d) **6.** (a) **7.** (d)
8. (c) **9.** (b) **10.** (a) **11.** (d)

ANSWERS TO TROUBLESHOOTING: SYMPTOM AND CAUSE

1. (c) **2.** (a) **3.** (b) **4.** (a) **5.** (b)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

CHAPTER 12

RL CIRCUITS

OUTLINE

12–1 Sinusoidal Response of *RL* Circuits
12–2 Impedance and Phase Angle of Series *RL* Circuits
12–3 Analysis of Series *RL* Circuits
12–4 Impedance and Phase Angle of Parallel *RL* Circuits
12–5 Analysis of Parallel *RL* Circuits
12–6 Analysis of Series-Parallel *RL* Circuits
12–7 Power in *RL* Circuits
12–8 *RL* Filters
12–9 Troubleshooting

OBJECTIVES

- Describe the relationship between current and voltage in an *RL* circuit
- Determine impedance and phase angle in a series *RL* circuit
- Analyze a series *RL* circuit
- Determine impedance and phase angle in a parallel *RL* circuit
- Analyze a parallel *RL* circuit
- Analyze series-parallel *RL* circuits
- Determine power in *RL* circuits
- Discuss how the *RL* circuit operates as a filter
- Troubleshoot *RL* circuits

KEY TERMS

***RL* lag circuit**
***RL* lead circuit**
Inductive susceptance (B_L)

INTRODUCTION

An *RL* circuit contains both resistance and inductance. In this chapter, series and parallel *RL* circuits and their responses to sinusoidal voltages are covered. In addition, series-parallel combinations are examined. Power considerations in *RL* circuits are introduced, and practical aspects of the power factor are discussed. A method of improving the power factor is presented. Troubleshooting common faults in *RL* circuits is also covered.

The methods for analyzing reactive circuits are similar to those you studied in dc circuits. Reactive circuit problems can be solved at only one frequency at a time, and phasor math must be used. Frequently in systems, the frequency of interest is the utility power frequency (50 Hz or 60 Hz). The inductance is not an ordinary component but may be the winding for a motor, solenoid, or transformer. You can still apply basic *RL* theory to these cases.

As you study this chapter, note both the differences and the similarities in the response of *RL* circuits compared to *RC* circuits.

VISIT THE WEBSITE
Study aids for this chapter are available at
http://pearsonhighered.com/floyd

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

12–1 SINUSOIDAL RESPONSE OF *RL* CIRCUITS

As with an *RC* circuit, all currents and voltages in any type of *RL* circuit are sinusoidal when the input voltage is sinusoidal. The inductance causes a phase shift between the voltage and the current that depends on the relative values of the resistance and the inductive reactance. Because of its winding resistance, the inductor is generally not as "ideal" as a resistor or capacitor. However, it will usually be treated as ideal for purposes of illustration.

After completing this section, you should be able to

- **Describe the relationship between current and voltage in an *RL* circuit**
  - **Discuss voltage and current waveforms**
  - **Discuss phase shift**

6238262 2013/12/10 12.27.66.5

In an *RL* circuit, the resistor voltage and the current lag the source voltage. The inductor voltage leads the source voltage. Ideally, the phase angle between the current and the inductor voltage is always 90°. These generalized phase relationships are indicated in Figure 12–1. Notice how they differ from those of the *RC* circuit that was discussed in Chapter 10.

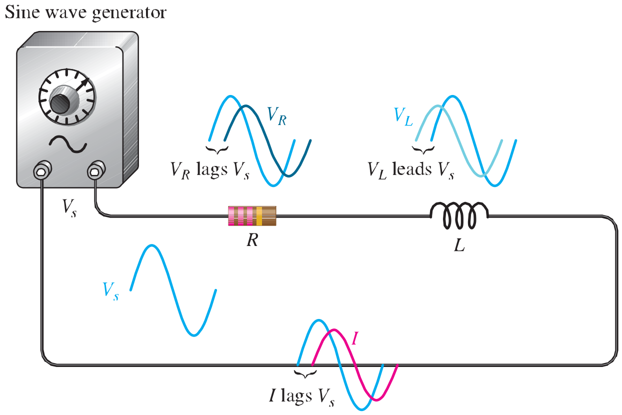

FIGURE 12–1 **Illustration of sinusoidal response with general phase relationships of V_R, V_L, and I relative to the source voltage. V_R and I are in phase; V_R and V_L are 90° out of phase with each other.**

The amplitudes and the phase relationships of the voltages and current depend on the values of the resistance and the inductive reactance. When a circuit is purely inductive, the phase angle between the source voltage and the total current is 90°, with the current lagging the voltage. When there is a combination of both resistance and inductive reactance in a circuit, the phase angle is somewhere between zero and 90°, depending on the relative values of the resistance and the inductive reactance. Because all inductors have winding resistance, ideal conditions may be approached but never reached in practice.

SECTION 12–1 CHECKUP*

1. A 1 kHz sinusoidal voltage is applied to an *RL* circuit. What is the frequency of the resulting current?
2. When the resistance in an *RL* circuit is greater than the inductive reactance, is the phase angle between the source voltage and the total current closer to 0° or to 90°?

*Answers are at the end of the chapter.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

12–2 IMPEDANCE AND PHASE ANGLE OF SERIES *RL* CIRCUITS

Impedance was introduced in Section 10–2 for *RC* circuits and represents the total opposition to sinusoidal current. As in the case with *RC* circuits, impedance is the combination of the resistive and reactive quantities, which can be represented by phasors. Because of phase differences, the total impedance must be treated as a phasor quantity.

After completing this section, you should be able to

- **Determine impedance and phase angle in a series *RL* circuit**
  - **Draw the impedance triangle**
  - **Calculate impedance magnitude**
  - **Calculate the phase angle**

The **impedance** of a series *RL* circuit is the total opposition to sinusoidal current and its unit is the ohm. The **phase angle** is the phase difference between the total current and the source voltage. Impedance (Z) is determined by the resistance (R) and the inductive reactance (X_L), as indicated in Figure 12–2.

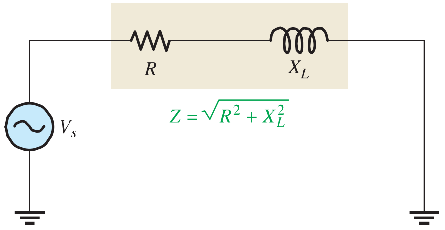

FIGURE 12–2 **Impedance of a series *RL* circuit.**

In ac analysis, both R and X_L are treated as phasor quantities, as shown in the phasor diagram of Figure 12–3(a), with X_L appearing at a +90° angle with respect to R. This relationship comes from the fact that the inductor voltage leads the current, and thus the resistor voltage, by 90°. Since Z is the phasor sum of R and X_L, its phasor representation is as shown in Figure 12–3(b). A repositioning of the phasors, as shown in part (c), forms a right triangle. This formation, as you have learned, is called the *impedance triangle*. The length of each phasor represents the magnitude of the quantity, and θ is the phase angle between the source voltage and the current in the *RL* circuit.

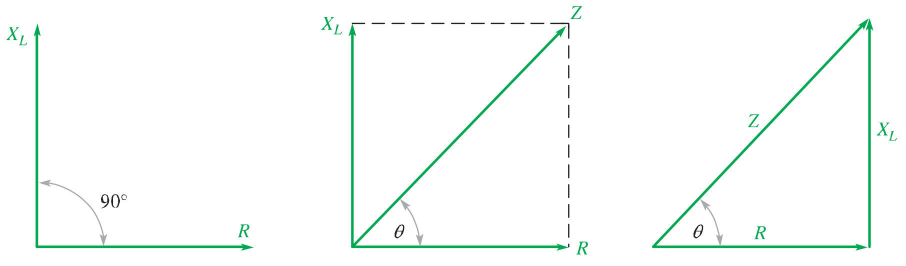

(a) (b) (c)

FIGURE 12–3 **Development of the impedance triangle for a series *RL* circuit.**

The magnitude of the impedance, Z, of the series *RL* circuit can be expressed in terms of the resistance and reactance as

$$Z = \sqrt{R^2 + X_L^2} \tag{12–1}$$

where Z is expressed in ohms.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The phase angle, θ, is expressed as

$$\boldsymbol{\theta = \tan^{-1}\left(\frac{X_L}{R}\right)} \qquad (12\text{–}2)$$

EXAMPLE 12–1

Determine the impedance and phase angle of the circuit in Figure 12–4. Draw the impedance triangle.

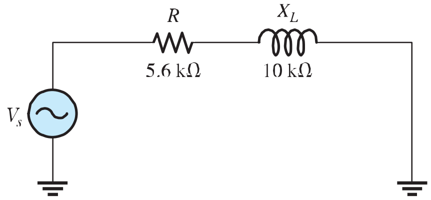

FIGURE 12–4

SOLUTION

The impedance is

$$Z = \sqrt{R^2 + X_L^2} = \sqrt{(5.6\ \text{k}\Omega)^2 + (10\ \text{k}\Omega)^2} = \mathbf{11.5\ k\Omega}$$

The value of the phase angle is

$$\theta = \tan^{-1}\left(\frac{X_L}{R}\right) = \tan^{-1}\left(\frac{10\ \text{k}\Omega}{5.6\ \text{k}\Omega}\right) = \mathbf{60.8^\circ}$$

The source voltage leads the current by 60.8°. The impedance triangle is shown in Figure 12–5.

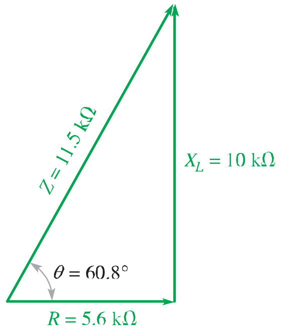

FIGURE 12–5

RELATED PROBLEM*

In a series *RL* circuit, $R = 1.8\ \text{k}\Omega$ and $X_L = 950\ \Omega$. Determine the impedance and phase angle.

***Answers are at the end of the chapter.**

SECTION 12–2 CHECKUP

1. Does the source voltage lead or lag the current in a series *RL* circuit?
2. What is the relationship between X_L and R when the phase angle is 45°?
3. How does the phase angle in an *RL* circuit differ from the phase angle in an *RC* circuit?
4. A series *RL* circuit has a resistance of 33 kΩ and an inductive reactance of 50 kΩ. Determine Z and θ.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

12–3 ANALYSIS OF SERIES *RL* CIRCUITS

Ohm's law and Kirchhoff's voltage law are used in the analysis of series *RL* circuits to determine voltage, current, and impedance. Also, in this section *RL* lead and lag circuits are examined.

After completing this section, you should be able to

- **Analyze a series *RL* circuit**
  - **Apply Ohm's law and Kirchhoff's voltage law to series *RL* circuits**
  - **Determine the phase relationships of the voltages and current**
  - **Show how impedance and phase angle vary with frequency**
  - **Discuss and analyze the *RL* lag circuit**
  - **Discuss and analyze the *RL* lead circuit**

Ohm's Law

The application of Ohm's law to series *RL* circuits involves the use of the quantities of Z, V, and I. The three equivalent forms of Ohm's law were stated in Chapter 10 for *RC* circuits. They apply also to *RL* circuits and are restated here:

$$V = IZ \qquad I = \frac{V}{Z} \qquad Z = \frac{V}{I}$$

The following example illustrates the use of Ohm's law.

EXAMPLE 12–2

The current in Figure 12–6 is 200 μA. Determine the source voltage.

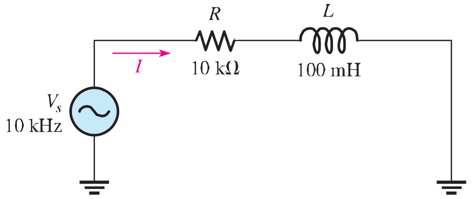

FIGURE 12–6

SOLUTION

From Equation 11–10, the inductive reactance is

$$X_L = 2\pi fL = 2\pi(10\text{ kHz})(100\text{ mH}) = 6.28\text{ k}\Omega$$

The impedance is

$$Z = \sqrt{R^2 + X_L^2} = \sqrt{(10\text{ k}\Omega)^2 + (6.28\text{ k}\Omega)^2} = 11.8\text{ k}\Omega$$

Applying Ohm's law yields

$$V_s = IZ = (200\ \mu\text{A})(11.8\text{ k}\Omega) = \mathbf{2.36\ V}$$

6238262 2013/12/10 12.27.66.5

RELATED PROBLEM

If the source voltage in Figure 12–6 is 5 V, what would be the current?

MULTISIM

Open Multisim file E12-02; files are found at www.pearsonhighered.com/floyd. Measure the current at 10 kHz, 5 kHz, and 20 kHz. Explain the results of your measurement.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Phase Relationships of the Current and Voltages

In a series *RL* circuit, the current is the same through both the resistor and the inductor. Thus, the resistor voltage is in phase with the current, and current lags the inductor voltage by 90°. Therefore, there is a phase difference of 90° between the resistor voltage, V_R, and the inductor voltage, V_L, as shown in the waveform diagram of Figure 12–7.

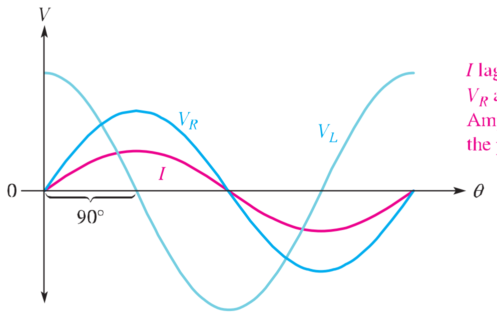

I lags V_L by 90°.
V_R and *I* are in phase.
Amplitudes depend on the particular circuit.

FIGURE 12–7 **Phase relation of current and voltages in a series *RL* circuit.**

From Kirchhoff's voltage law, the sum of the voltage drops must equal the source voltage. However, since V_R and V_L are not in phase with each other, they must be added as phasor quantities with V_L leading V_R by 90°, as shown in Figure 12–8(a). As shown in part (b), V_s is the phasor sum of V_R and V_L. This equation can be expressed as

$$V_s = \sqrt{V_R^2 + V_L^2} \tag{12–3}$$

The phase angle between the resistor voltage and the source voltage can be expressed as

$$\theta = \tan^{-1}\left(\frac{V_L}{V_R}\right) \tag{12–4}$$

Since the resistor voltage and the current are in phase, θ in Equation 12–4 also represents the phase angle between the source voltage and the current and is equivalent to $\tan^{-1}(X_L/R)$.

Figure 12–9 shows a voltage and current phasor diagram that represents the waveform diagram of Figure 12–7.

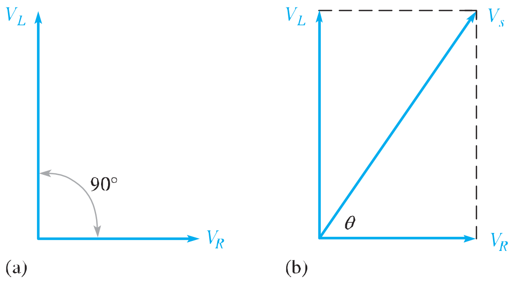

FIGURE 12–8 **Voltage phasor diagram for the waveforms in Figure 12–7.**

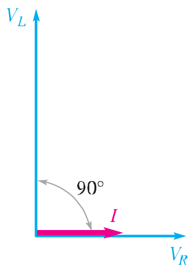

FIGURE 12–9 **Voltage and current phasor diagram for the waveforms in Figure 12–7.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 12–3

Determine the source voltage and the phase angle in Figure 12–10. Draw the voltage phasor diagram.

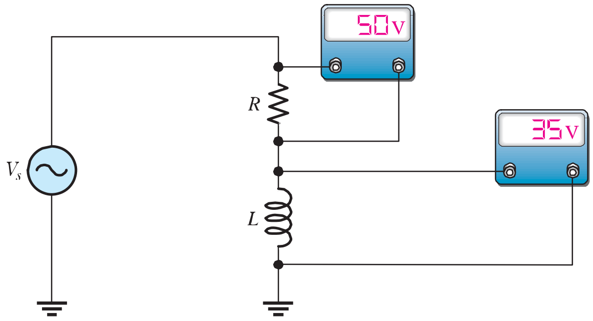

FIGURE 12–10

SOLUTION

Since V_R and V_L are 90° out of phase, you cannot add them directly. The source voltage is the phasor sum of V_R and V_L.

$$V_s = \sqrt{V_R^2 + V_L^2} = \sqrt{(50\text{ V})^2 + (35\text{ V})^2} = \mathbf{61\ V}$$

The phase angle between the resistor voltage and the source voltage is

$$\theta = \tan^{-1}\left(\frac{V_L}{V_R}\right) = \tan^{-1}\left(\frac{35\text{ V}}{50\text{ V}}\right) = \mathbf{35°}$$

The voltage phasor diagram is shown in Figure 12–11.

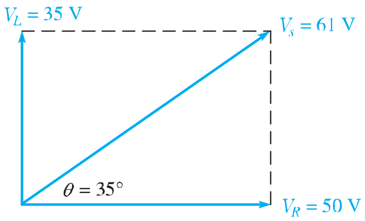

FIGURE 12–11

RELATED PROBLEM

With the information given, can you determine the current in Figure 12–10?

Variation of Impedance and Phase Angle with Frequency

As you know, inductive reactance varies directly with frequency. When X_L increases, the total impedance also increases; and when X_L decreases, the total impedance decreases. Thus, *in an RL circuit, Z is directly dependent on frequency.*

Figure 12–12 illustrates how the voltages and current in a series *RL* circuit vary as the frequency increases or decreases, with the source voltage held at a constant value. Part (a) shows that as frequency is increased, the increase in X_L will cause more voltage to appear across the inductor and less current will be in the circuit. Thus, V_L increases and V_R decreases. Also, Z is larger because there is less total current with the same total voltage across the circuit.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

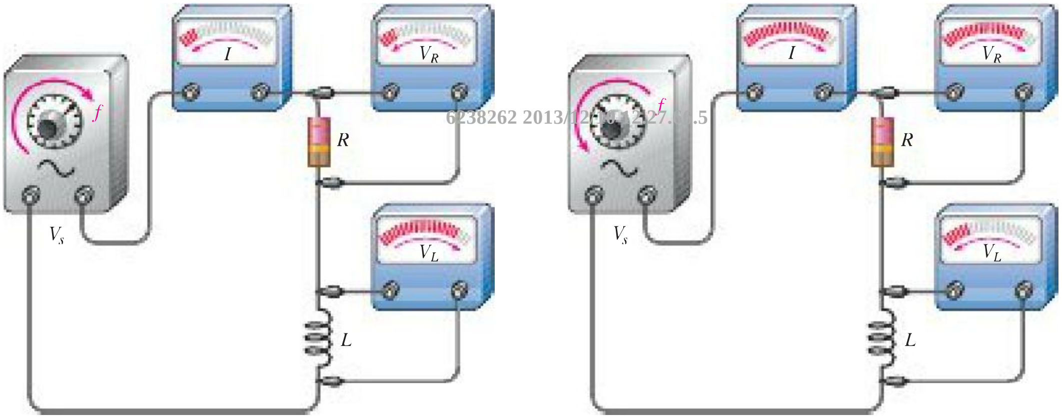

6238262 2013/12/10 12.27.66.5

(a) As frequency is increased, I and V_R decrease and V_L increases.

(b) As frequency is decreased, I and V_R increase and V_L decreases.

FIGURE 12–12 **An illustration of how the variation of impedance affects the voltages and current as the source frequency is varied. The source voltage is held at a constant amplitude.**

Figure 12–12(b) shows that as frequency decreases, the voltage across the inductor decreases because X_L decreases. Also, Z decreases, causing the current to increase. An increase in current causes more voltage across the resistor.

Changes in Z and X_L can be observed as shown in Figure 12–13. As the frequency increases, the voltage across Z remains constant because V_s is constant ($V_s = V_Z$), but the voltage across L increases. The decreasing current indicates that Z is increasing. It does so because of the inverse relationship stated in Ohm's law ($Z = V_Z/I$). The decreasing current also indicates that X_L is increasing. The increase in V_L corresponds to the increase in X_L.

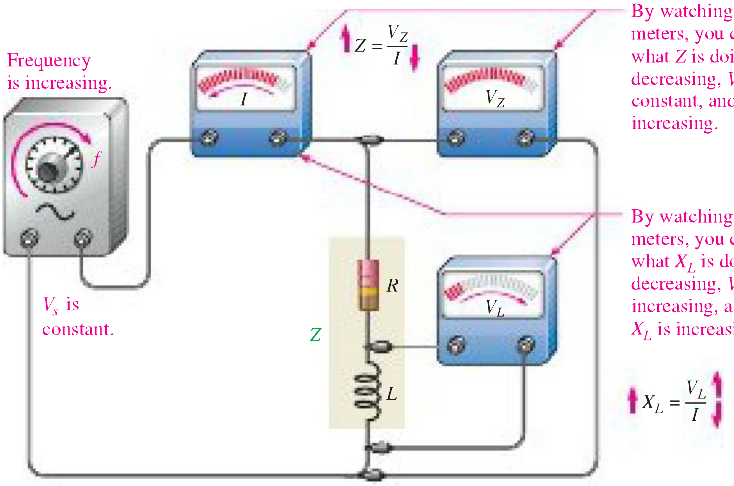

$\uparrow Z = \frac{V_Z}{I}\downarrow$
Frequency is increasing.
I
V_Z
f
V_s is constant.
R
Z
V_L
L
$\uparrow X_L = \frac{V_L}{I}\downarrow$

By watching these two meters, you can see what Z is doing: I is decreasing, V_Z is constant, and thus Z is increasing.

By watching these two meters, you can see what X_L is doing: I is decreasing, V_L is increasing, and thus X_L is increasing.

FIGURE 12–13 **Observing changes in Z and X_L with frequency by watching the meters and recalling Ohm's law.**

HANDS ON TIP

As you know, some multimeters have a relatively low frequency response. There are other things of which you should be aware. One is that most ac meters are accurate only if the waveform being measured is sinusoidal. Another is that accuracy when measuring small ac voltage is usually less than for dc measurements. Finally, loading can affect the accuracy of meter readings.

Since X_L is the factor that introduces the phase angle in a series *RL* circuit, a change in X_L produces a change in the phase angle. As the frequency is increased, X_L becomes greater, and thus the phase angle increases. As the frequency is decreased, X_L becomes smaller, and thus the phase angle decreases. The angle between V_s and V_R is the phase angle of the circuit because I is in phase with V_R. The variations of phase angle with frequency are illustrated with the impedance triangle as shown in Figure 12–14.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

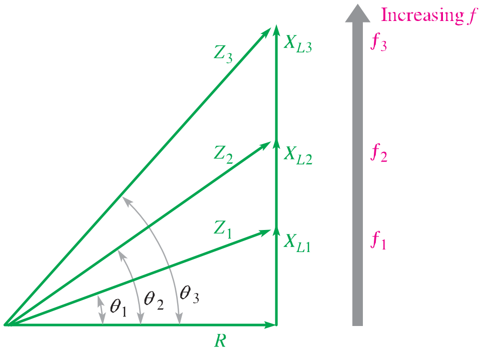

FIGURE 12–14 **As the frequency increases, the phase angle θ increases.**

EXAMPLE 12–4

For the series *RL* circuit in Figure 12–15, determine the impedance and the phase angle for each of the following frequencies:

(a) 10 kHz **(b)** 20 kHz **(c)** 30 kHz

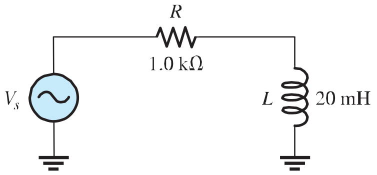

FIGURE 12–15

SOLUTION

(a) For $f = 10$ kHz, calculate the impedance as follows:

$$X_L = 2\pi fL = 2\pi(10\text{ kHz})(20\text{ mH}) = 1.26\text{ k}\Omega$$

$$Z = \sqrt{R^2 + X_L^2} = \sqrt{(1.0\text{ k}\Omega)^2 + (1.26\text{ k}\Omega)^2} = \mathbf{1.61\text{ k}\Omega}$$

The phase angle is

$$\theta = \tan^{-1}\left(\frac{X_L}{R}\right) = \tan^{-1}\left(\frac{1.26\text{ k}\Omega}{1.0\text{ k}\Omega}\right) = \mathbf{51.6^\circ}$$

(b) For $f = 20$ kHz,

$$X_L = 2\pi(20\text{ kHz})(20\text{ mH}) = 2.51\text{ k}\Omega$$

$$Z = \sqrt{(1.0\text{ k}\Omega)^2 + (2.51\text{ k}\Omega)^2} = \mathbf{2.70\text{ k}\Omega}$$

$$\theta = \tan^{-1}\left(\frac{2.51\text{ k}\Omega}{1.0\text{ k}\Omega}\right) = \mathbf{68.3^\circ}$$

(c) For $f = 30$ kHz,

$$X_L = 2\pi(30\text{ kHz})(20\text{ mH}) = 3.77\text{ k}\Omega$$

$$Z = \sqrt{(1.0\text{ k}\Omega)^2 + (3.77\text{ k}\Omega)^2} = \mathbf{3.90\text{ k}\Omega}$$

$$\theta = \tan^{-1}\left(\frac{3.77\text{ k}\Omega}{1.0\text{ k}\Omega}\right) = \mathbf{75.1^\circ}$$

Notice that as the frequency increases, X_L, Z, and θ also increase.

RELATED PROBLEM

Determine Z and θ in Figure 12–15 if $f = 100$ kHz.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

RL Lag Circuit

The ***RL* lag circuit** is a phase shift circuit in which the output voltage lags the input voltage by a specified angle, ϕ. A basic series *RL* lag circuit is shown in Figure 12–16(a). Notice that the output voltage is taken across the resistor and the input voltage is the total voltage applied across the circuit. The relationship of the voltages is shown in the phasor diagram in Figure 12–16(b), and a waveform diagram is shown in Figure 12–16(c). Notice that the output voltage, V_{out}, lags V_{in} by an angle, designated ϕ, that is the same as the circuit phase angle. The angles are equal, of course, because V_R and I are in phase with each other.

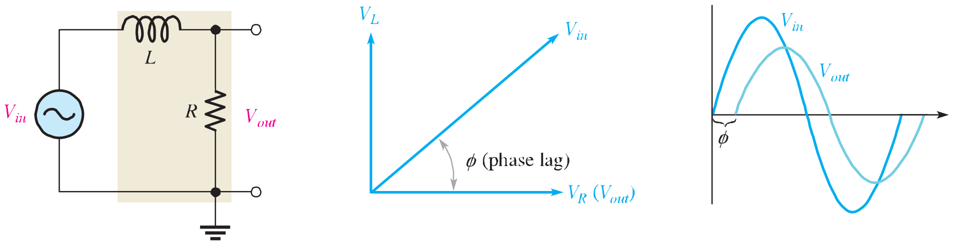

(a) A basic *RL* lag circuit

(b) Phasor voltage diagram showing phase lag between V_{in} and V_{out}

(c) Input and output waveforms

FIGURE 12–16 **The *RL* lag circuit ($V_{out} = V_R$).**

The phase lag, ϕ, can be expressed as

$$\phi = \tan^{-1}\left(\frac{X_L}{R}\right) \tag{12–5}$$

EXAMPLE 12–5

Calculate the phase lag for each circuit in Figure 12–17.

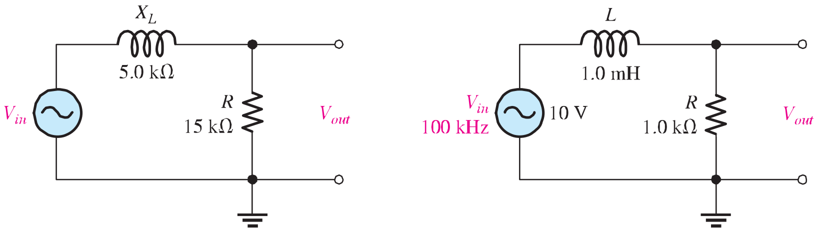

(a)

(b)

FIGURE 12–17

SOLUTION

For the circuit in Figure 12–17(a),

$$\phi = \tan^{-1}\left(\frac{X_L}{R}\right) = \tan^{-1}\left(\frac{5.0\text{ k}\Omega}{15\text{ k}\Omega}\right) = \mathbf{18.4^\circ}$$

The output lags the input by 18.4°.

For the circuit in Figure 12–17(b), first determine the inductive reactance.

$$X_L = 2\pi fL = 2\pi(100\text{ kHz})(1.0\text{ mH}) = 628\ \Omega$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The phase lag is

$$\phi = \tan^{-1}\left(\frac{X_L}{R}\right) = \tan^{-1}\left(\frac{628\ \Omega}{1.0\ \text{k}\Omega}\right) = \mathbf{32.1^\circ}$$

The output lags the input by 32.1°.

RELATED PROBLEM

In a certain lag circuit, $R = 5.6\ \text{k}\Omega$ and $X_L = 3.5\ \text{k}\Omega$. Determine the phase lag between input and output.

The phase-lag circuit can be considered as a voltage divider with a portion of the input voltage dropped across *L* and a portion across *R*. The output voltage can be determined with the following formula:

$$\boldsymbol{V_{out} = \left(\frac{R}{\sqrt{R^2 + X_L^2}}\right)V_{in}} \qquad \text{(12–6)}$$

EXAMPLE 12–6

The input voltage in Figure 12–17(b) of Example 12–5 has an rms value of 10 V. Determine the output voltage for the lag circuit shown in Figure 12–17(b). Draw the waveform relationships for the input and output voltages. The phase lag (32.1°) and X_L (628 Ω) were found in Example 12–5.

SOLUTION

Use Equation 12–6 to determine the output voltage for the lag circuit in Figure 12–17(b).

$$V_{out} = \left(\frac{R}{\sqrt{R^2 + X_L^2}}\right)V_{in} = \left(\frac{1.0\ \text{k}\Omega}{1.18\ \text{k}\Omega}\right)10\ \text{V} = \mathbf{8.47\ V\ rms}$$

6238262 2013/12/10 12.27.66.5

The waveforms are shown in Figure 12–18.

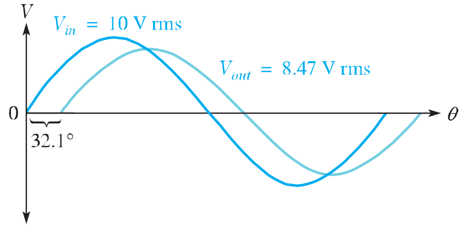

FIGURE 12–18

RELATED PROBLEM

In a lag circuit, $R = 4.7\ \text{k}\Omega$ and $X_L = 6\ \text{k}\Omega$. If the rms input voltage is 20 V, what is the output voltage?

MULTISIM

Open Multisim file E12-06. Measure the output voltage and compare to the calculated value.

EFFECTS OF FREQUENCY ON THE LAG CIRCUIT Since the circuit phase angle and the phase lag are the same, an increase in frequency causes an increase in phase lag. Also, an increase in frequency causes a decrease in the magnitude of the output voltage because X_L becomes greater and more of the total voltage is dropped across the inductor and less across the resistor.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

RL Lead Circuit

The ***RL* lead circuit** is a phase shift circuit in which the output voltage leads the input voltage by a specified angle, ϕ. A basic series *RL* lead circuit is shown in Figure 12–19(a). Notice how this circuit differs from the lag circuit. Here the output voltage is taken across the inductor rather than across the resistor. The relationship of the voltages is shown in the phasor diagram of Figure 12–19(b) and in the waveform plot of Figure 12–19(c). Notice that the output voltage, V_{out}, leads V_{in} by an angle (phase lead) that is the difference between 90° and the circuit phase angle θ.

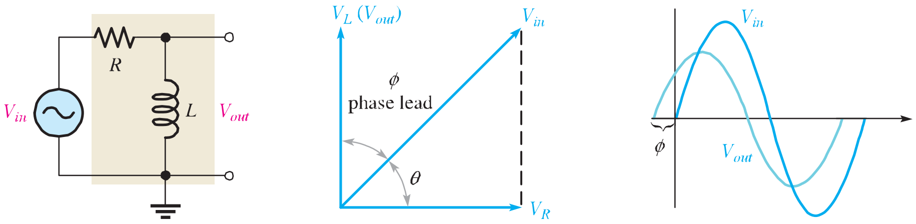

(a) A basic *RL* lead circuit

(b) Phasor voltage diagram showing V_{out} leading V_{in}

(c) Input and output voltage waveforms

FIGURE 12–19 **The *RL* lead circuit ($V_{out} = V_L$).**

Since $\theta = \tan^{-1}(X_L/R)$, the phase lead, ϕ, can be expressed as

$$\phi = 90° - \tan^{-1}\left(\frac{X_L}{R}\right) \quad (12\text{–}7)$$

Equivalently,

$$\phi = \tan^{-1}\left(\frac{R}{X_L}\right)$$

Again, the phase-lead circuit can be considered as a voltage divider with the voltage across L being the output. The expression for the output voltage is

$$V_{out} = \left(\frac{X_L}{\sqrt{R^2 + X_L^2}}\right)V_{in} \quad (12\text{–}8)$$

EXAMPLE 12–7

Determine the amount of phase lead and output voltage in the lead circuit in Figure 12–20.

FIGURE 12–20

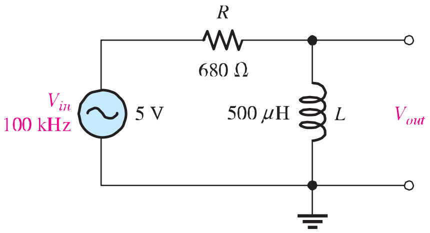

SOLUTION

First, determine the inductive reactance.

$$X_L = 2\pi fL = 2\pi(100\text{ kHz})(500\ \mu\text{H}) = 314\ \Omega$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The phase lead is

$$\phi = 90° - \tan^{-1}\left(\frac{X_L}{R}\right) = 90° - \tan^{-1}\left(\frac{314\ \Omega}{680\ \Omega}\right) = \mathbf{65.2°}$$

The output leads the input by 65.2°.

The output voltage is

$$V_{out} = \left(\frac{X_L}{\sqrt{R^2 + X_L^2}}\right)V_{in} = \left(\frac{314\ \Omega}{\sqrt{(680\ \Omega)^2 + (314\ \Omega)^2}}\right)5\text{ V} = \mathbf{2.1\ V}$$

RELATED PROBLEM

In a certain lead circuit, $R = 2.2\text{ k}\Omega$ and $X_L = 1\text{ k}\Omega$. What is the phase lead?

MULTISIM

Open Multisim file E12-07. Measure the output voltage and compare to the calculated value.

EFFECTS OF FREQUENCY ON THE LEAD CIRCUIT Since the circuit phase angle, θ, increases as frequency increases, the phase lead between the input and the output voltages decreases. Also, the amplitude of the output voltage increases as the frequency increases because X_L becomes greater and more of the total input voltage is dropped across the inductor.

SECTION 12–3 CHECKUP

1. In a certain series *RL* circuit, $V_R = 2$ V and $V_L = 3$ V. What is the magnitude of the total voltage?
2. In Question 1, what is the phase angle?
3. When the frequency of the source voltage in a series *RL* circuit is increased, what happens to the inductive reactance? What happens to the impedance? What happens to the phase angle?
4. A certain *RL* lead circuit consists of a 3.3 kΩ resistor and a 15 mH inductor. Determine the phase lead between input and output at a frequency of 5 kHz.
5. An *RL* lag circuit has the same component values as the lead circuit in Question 4. What is the magnitude of the output voltage at 5 kHz when the input is 10 V rms?

12–4 IMPEDANCE AND PHASE ANGLE OF PARALLEL *RL* CIRCUITS

In this section, you will learn how to determine the impedance and phase angle of a parallel *RL* circuit. Also, inductive susceptance and admittance of a parallel *RL* circuit are introduced.

After completing this section, you should be able to

- **Determine impedance and phase angle in a parallel *RL* circuit**
  - **Express total impedance in a product-over-sum form**
  - **Express the phase angle in terms of *R* and X_L**
  - **Determine inductive susceptance and admittance**
  - **Convert admittance to impedance**

A basic parallel *RL* circuit is shown in Figure 12–21. The expression for the impedance, using the product-over-sum rule, is

$$Z = \frac{RX_L}{\sqrt{R^2 + X_L^2}} \qquad (12\text{–}9)$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The phase angle between the source voltage and the total current can be expressed in terms of R and X_L as

$$\boldsymbol{\theta = \tan^{-1}\left(\frac{R}{X_L}\right)} \tag{12–10}$$

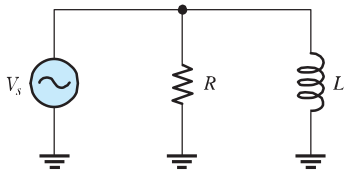

FIGURE 12–21 **Parallel *RL* circuit.**

EXAMPLE 12–8

For each circuit in Figure 12–22, determine the impedance and the phase angle.

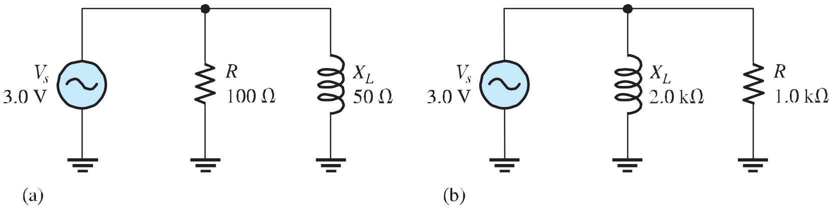

FIGURE 12–22

SOLUTION

For the circuit in Figure 12–22(a), the impedance and phase angle are

$$Z = \frac{RX_L}{\sqrt{R^2 + X_L^2}} = \frac{(100\ \Omega)(50\ \Omega)}{\sqrt{(100\ \Omega)^2 + (50\ \Omega)^2}} = \mathbf{44.7\ \Omega}$$

$$\theta = \tan^{-1}\left(\frac{R}{X_L}\right) = \tan^{-1}\left(\frac{100\ \Omega}{50\ \Omega}\right) = \mathbf{63.4^\circ}$$

For the circuit in Figure 12–22(b),

$$Z = \frac{(1.0\ \text{k}\Omega)(2.0\ \text{k}\Omega)}{\sqrt{(1.0\ \text{k}\Omega)^2 + (2.0\ \text{k}\Omega)^2}} = \mathbf{894\ \Omega}$$

$$\theta = \tan^{-1}\left(\frac{1.0\ \text{k}\Omega}{2.0\ \text{k}\Omega}\right) = \mathbf{26.6^\circ}$$

The voltage leads the current, compared to the parallel *RC* case where the voltage lags the current.

RELATED PROBLEM

6238262 2013/12/10 12.27.66.5

In a parallel *RL* circuit $R = 10\ \text{k}\Omega$ and $X_L = 14\ \text{k}\Omega$. Find Z and θ.

Conductance, Susceptance, and Admittance

As you know from Section 10–4, conductance (G) is the reciprocal of resistance, susceptance (B) is the reciprocal of reactance, and admittance (Y) is the reciprocal of impedance.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

For parallel *RL* circuits, **conductance** (*G*) is expressed as

$$G = \frac{1}{R} \quad (12\text{–}11)$$

Inductive susceptance (B_L) is expressed as

$$B_L = \frac{1}{X_L} \quad (12\text{–}12)$$

Admittance (*Y*) is expressed as

$$Y = \frac{1}{Z} \quad (12\text{–}13)$$

As with the *RC* circuit, the unit for G, B_L, and Y is the siemens (S).

In the basic parallel *RL* circuit shown in Figure 12–23(a), the total admittance is the phasor sum of the conductance and the inductive susceptance, as shown in part (b).

$$Y_{tot} = \sqrt{G^2 + B_L^2} \quad (12\text{–}14)$$

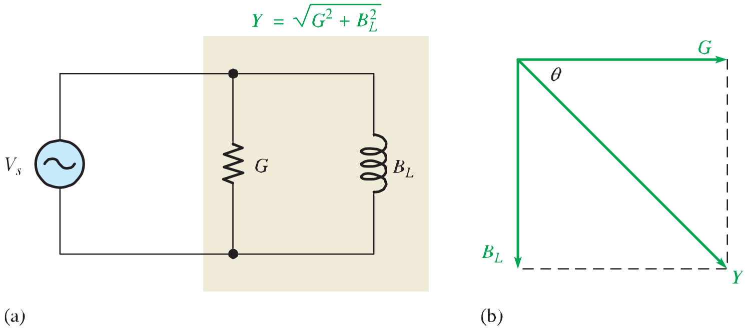

FIGURE 12–23 Admittance in a parallel *RL* circuit.

EXAMPLE 12–9

Determine the total admittance for the circuit in Figure 12–24; then convert it to impedance.

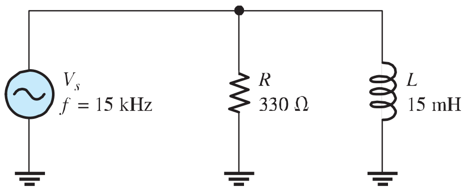

FIGURE 12–24

SOLUTION

To determine Y, first calculate the values for G and B_L. Since $R = 330\ \Omega$,

$$G = \frac{1}{R} = \frac{1}{330\ \Omega} = 3.03\ \text{mS}$$

The inductive reactance is

$$X_L = 2\pi fL = 2\pi(15\ \text{kHz})(15\ \text{mH}) = 1.41\ \text{k}\Omega$$

The inductive susceptance is

$$B_L = \frac{1}{X_L} = \frac{1}{1.41\ \text{k}\Omega} = 0.707\ \text{mS}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Therefore, the total admittance is

$$Y_{tot} = \sqrt{G^2 + B_L^2} = \sqrt{(3.03\text{ mS})^2 + (0.707\text{ mS})^2} = \mathbf{3.11\text{ mS}}$$

Convert to impedance.

$$Z = \frac{1}{Y_{tot}} = \frac{1}{3.11\text{ mS}} = \mathbf{321\ \Omega}$$

RELATED PROBLEM

What is the total admittance of the circuit in Figure 12–24 if f is decreased to 5 kHz?

SECTION 12–4 CHECKUP

1. If $Y = 50$ mS, what is the value of Z?
2. In a certain parallel *RL* circuit, $R = 470\ \Omega$ and $X_L = 750\ \Omega$. Determine the admittance.
3. In the circuit of Question 2, does the total current lead or lag the source voltage and by what phase angle?

12–5 ANALYSIS OF PARALLEL *RL* CIRCUITS

Ohm's law and Kirchhoff's current law are used in the analysis of parallel *RL* circuits. Current and voltage relationships in a parallel *RL* circuit are examined.

After completing this section, you should be able to

- **Analyze a parallel *RL* circuit**
  - **Apply Ohm's law and Kirchhoff's current law to parallel *RL* circuits**
  - **Determine total current and phase angle**

The following example applies Ohm's law to the analysis of a parallel *RL* circuit.

EXAMPLE 12–10

Determine the total current and the phase angle in the circuit of Figure 12–25.

FIGURE 12–25

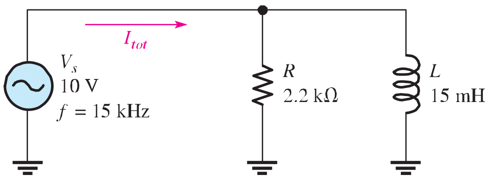

SOLUTION

First, determine the total admittance. The inductive reactance is

$$X_L = 2\pi fL = 2\pi(15\text{ kHz})(15\text{ mH}) = 1.41\text{ k}\Omega$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The conductance is

$$G = \frac{1}{R} = \frac{1}{2.2\ \text{k}\Omega} = 455\ \mu\text{S}$$

The inductive susceptance is

$$B_L = \frac{1}{X_L} = \frac{1}{1.41\ \text{k}\Omega} = 709\ \mu\text{S}$$

Therefore, the total admittance is

$$Y_{tot} = \sqrt{G^2 + B_L^2} = \sqrt{(455\ \mu\text{S})^2 + (709\ \mu\text{S})^2} = 842\ \mu\text{S}$$

Next, use Ohm's law to calculate the total current.

$$I_{tot} = VY_{tot} = (10\ \text{V})(842\ \mu\text{S}) = \mathbf{8.42\ mA}$$

The phase angle is

$$\theta = \tan^{-1}\left(\frac{R}{X_L}\right) = \tan^{-1}\left(\frac{2.2\ \text{k}\Omega}{1.41\ \text{k}\Omega}\right) = \mathbf{57.3°}$$

The total current is 8.42 mA, and it lags the source voltage by 57.3°.

RELATED PROBLEM

Determine the total current and the phase angle if the frequency is reduced to 8.0 kHz in Figure 12–25.

6238262 2013/12/10 12.27.66.5

MULTISIM

Open Multisim file E12-10. Measure the total current and each branch current. Change the frequency to 8.0 kHz and measure I_{tot}.

Phase Relationships of the Currents and Voltages

Figure 12–26(a) shows all the currents and voltages in a basic parallel *RL* circuit. As you can see, the source voltage, V_s, appears across both the resistive and the inductive branches, so V_s, V_R, and V_L are all in phase and of the same magnitude. The total current, I_{tot}, divides at the junction into the two branch currents, I_R and I_L. The current and voltage phasor diagram is shown in Figure 12–26(b).

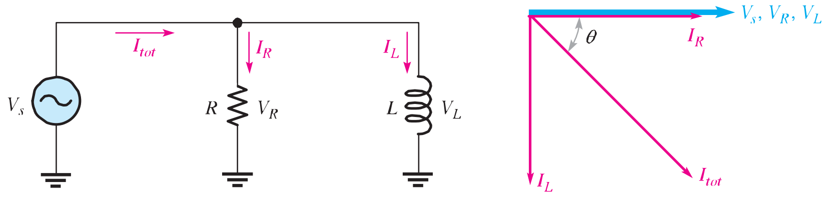

(a) (b)

FIGURE 12–26 **Currents and voltages in a parallel *RL* circuit. The current directions shown in part (a) are instantaneous and, of course, reverse when the source voltage reverses during each cycle.**

The current through the resistor is in phase with the voltage. The current through the inductor lags the voltage and the resistor current by 90°. By Kirchhoff's current law, the total current is the phasor sum of the two branch currents. The total current is expressed as

$$I_{tot} = \sqrt{I_R^2 + I_L^2} \qquad (12\text{–}15)$$

The phase angle between the resistor current and the total current is

$$\theta = \tan^{-1}\left(\frac{I_L}{I_R}\right) \qquad (12\text{–}16)$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 12–11

For the circuit in Figure 12–25, draw the current phasor diagram to show the phase relationships between the currents and the source voltage.

SOLUTION

The conductance, susceptance, admittance, total current, and phase angle were calculated in Example 12–10.

$$G = 455\ \mu\text{S}$$

$$B_L = 709\ \mu\text{S}$$

$$Y_{tot} = 842\ \mu\text{S}$$

$$I_{tot} = 8.42\ \text{mA}$$

$$\theta = 57.3^\circ$$

The branch currents are

$$I_R = GV = (455\ \mu\text{S})(10\ \text{V}) = 4.55\ \text{mA (in phase with } V_s)$$

$$I_L = B_L V = (709\ \mu\text{S})(10\ \text{V}) = 7.09\ \text{mA (lags } V_s \text{ by } 90^\circ)$$

The current phasor diagram is shown in Figure 12–27. The blue phasor represents the voltage in the circuit.

FIGURE 12–27

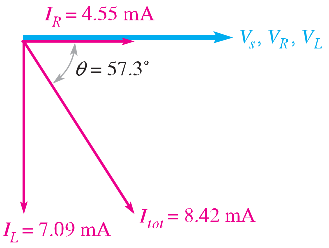

RELATED PROBLEM

Find I_{tot} for the circuit in Figure 12–25 if a second 15 mH inductor is in parallel with the first one.

SECTION 12–5 CHECKUP

1. The admittance of a parallel *RL* circuit is 4 mS, and the source voltage is 8 V. What is the total current?
2. In a certain parallel *RL* circuit, the resistor current is 12 mA, and the inductor current is 20 mA. Determine the phase angle and the total current.
3. What is the phase angle between the inductor current and the source voltage in a parallel *RL* circuit?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

12–6 ANALYSIS OF SERIES-PARALLEL *RL* CIRCUITS

The concepts studied in the previous sections are used to analyze circuits with combinations of both series and parallel *R* and *L* components.

After completing this section, you should be able to

- **Analyze series-parallel *RL* circuits**
  - **Determine total impedance and phase angle**
  - **Calculate currents and voltages**

The following two examples illustrate the procedures used to analyze a series-parallel reactive circuit.

EXAMPLE 12–12

In the series-parallel *RL* circuit of Figure 12–28, determine the values of the following:

(a) Z_{tot} **(b)** I_{tot} **(c)** θ

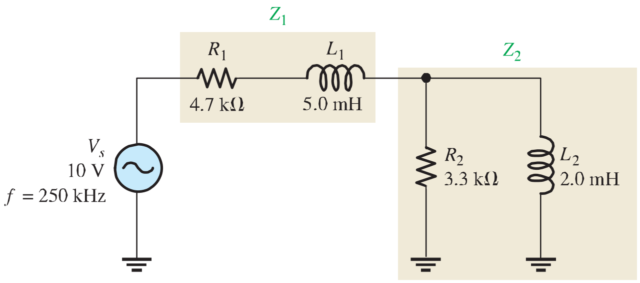

FIGURE 12–28

SOLUTION

(a) First, calculate the magnitudes of the inductive reactances.

$$X_{L1} = 2\pi f L_1 = 2\pi(250\text{ kHz})(5.0\text{ mH}) = 7.85\text{ k}\Omega$$

$$X_{L2} = 2\pi f L_2 = 2\pi(250\text{ kHz})(2.0\text{ mH}) = 3.14\text{ k}\Omega$$

One approach is to find the series equivalent resistance and inductive reactance for the parallel portion of the circuit; then add the resistances ($R_1 + R_{eq}$) to get the total resistance and add the reactances ($X_{L1} + X_{L(eq)}$) to get the total reactance. From these totals, you can determine the total impedance.

Determine the impedance of the parallel portion (Z_2) as follows:

$$G_2 = \frac{1}{R_2} = \frac{1}{3.3\text{ k}\Omega} = 303\ \mu\text{S}$$

$$B_{L2} = \frac{1}{X_{L2}} = \frac{1}{3.14\text{ k}\Omega} = 318\ \mu\text{S}$$

$$Y_2 = \sqrt{G_2^2 + B_L^2} = \sqrt{(303\ \mu\text{S})^2 + (318\ \mu\text{S})^2} = 439\ \mu\text{S}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Then

$$Z_2 = \frac{1}{Y_2} = \frac{1}{439\ \mu\text{S}} = 2.28\ \text{k}\Omega$$

The phase angle associated with the parallel portion of the circuit is

$$\theta_p = \tan^{-1}\left(\frac{R_2}{X_{L2}}\right) = \tan^{-1}\left(\frac{3.3\ \text{k}\Omega}{3.14\ \text{k}\Omega}\right) = 46.4^\circ$$

The series equivalent values for the parallel portion are found by breaking the impedance into an equivalent resistance in series with an equivalent inductive reactance as follows:

$$R_{\text{eq}} = Z_2\cos\theta_p = (2.28\ \text{k}\Omega)\cos(46.4^\circ) = 1.57\ \text{k}\Omega$$

$$X_{L(\text{eq})} = Z_2\sin\theta_p = (2.28\ \text{k}\Omega)\sin(46.4^\circ) = 1.65\ \text{k}\Omega$$

The total circuit resistance is

$$R_{tot} = R_1 + R_{\text{eq}} = 4.7\ \text{k}\Omega + 1.57\ \text{k}\Omega = 6.27\ \text{k}\Omega$$

The total circuit reactance is

$$X_{L(tot)} = X_{L1} + X_{L(\text{eq})} = 7.85\ \text{k}\Omega + 1.65\ \text{k}\Omega = 9.50\ \text{k}\Omega$$

The total circuit impedance is

$$Z_{tot} = \sqrt{R_{tot}^2 + X_{L(tot)}^2} = \sqrt{(6.27\ \text{k}\Omega)^2 + (9.50\ \text{k}\Omega)^2} = \mathbf{11.4\ k\Omega}$$

(b) Use Ohm's law to find the total current.

$$I_{tot} = \frac{V_s}{Z_{tot}} = \frac{10\ \text{V}}{11.4\ \text{k}\Omega} = \mathbf{877\ \mu A}$$

(c) To find the phase angle, view the circuit as a series combination of R_{tot} and $X_{L(tot)}$. The phase angle by which I_{tot} lags V_s is

$$\theta = \tan^{-1}\left(\frac{X_{L(tot)}}{R_{tot}}\right) = \tan^{-1}\left(\frac{9.50\ \text{k}\Omega}{6.27\ \text{k}\Omega}\right) = \mathbf{56.6^\circ}$$

MULTISIM

Open Multisim file E12-12. Measure the current through each component. Measure the voltage across Z_1 and Z_2.

RELATED PROBLEM

(a) Determine the voltage across the series part of the circuit in Figure 12–28.

(b) Determine the voltage across the parallel part of the circuit.

EXAMPLE 12–13

Determine the voltage across each component in Figure 12–29. Draw a current and voltage phasor diagram.

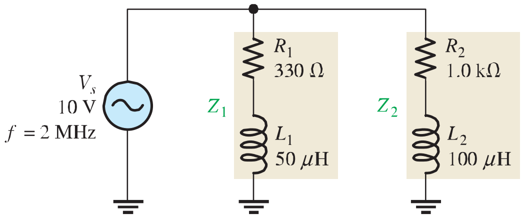

6238262 2013/12/10 12.27.66.5

FIGURE 12–29

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

First, calculate X_{L1} and X_{L2}.

$$X_{L1} = 2\pi f L_1 = 2\pi(2\text{ MHz})(50\ \mu\text{H}) = 628\ \Omega$$
$$X_{L2} = 2\pi f L_2 = 2\pi(2\text{ MHz})(100\ \mu\text{H}) = 1.26\text{ k}\Omega$$

Now, determine the impedance of each branch.

$$Z_1 = \sqrt{R_1^2 + X_{L1}^2} = \sqrt{(330\ \Omega)^2 + (628\ \Omega)^2} = 709\ \Omega$$
$$Z_2 = \sqrt{R_2^2 + X_{L2}^2} = \sqrt{(1.0\text{ k}\Omega)^2 + (1.26\text{ k}\Omega)^2} = 1.61\text{ k}\Omega$$

Calculate each branch current.

$$I_1 = \frac{V_s}{Z_1} = \frac{10\text{ V}}{709\ \Omega} = 14.1\text{ mA}$$
$$I_2 = \frac{V_s}{Z_2} = \frac{10\text{ V}}{1.61\text{ k}\Omega} = 6.21\text{ mA}$$

Now, use Ohm's law to find the voltage across each component.

$$V_{R1} = I_1R_1 = (14.1\text{ mA})(330\ \Omega) = \mathbf{4.65\ V}$$
$$V_{L1} = I_1X_{L1} = (14.1\text{ mA})(628\ \Omega) = \mathbf{8.85\ V}$$
$$V_{R2} = I_2R_2 = (6.21\text{ mA})(1.0\text{ k}\Omega) = \mathbf{6.21\ V}$$
$$V_{L2} = I_2X_{L2} = (6.21\text{ mA})(1.26\text{ k}\Omega) = \mathbf{7.82\ V}$$

Now determine the values of the angles associated with each branch current.

$$\theta_1 = \tan^{-1}\left(\frac{X_{L1}}{R_1}\right) = \tan^{-1}\left(\frac{628\ \Omega}{330\ \Omega}\right) = 62.3^\circ$$
$$\theta_2 = \tan^{-1}\left(\frac{X_{L2}}{R_2}\right) = \tan^{-1}\left(\frac{1.26\text{ k}\Omega}{1.0\text{ k}\Omega}\right) = 51.6^\circ$$

Thus, I_1 lags V_s by 62.3°, and I_2 lags V_s by 51.6°, as indicated in Figure 12–30(a).

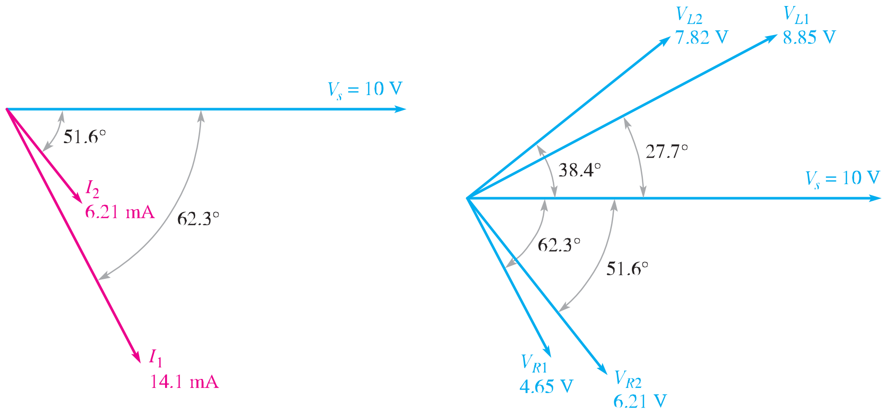

(a) (b)

FIGURE 12–30

The phase relationships of the voltages are determined as follows:

- V_{R1} is in phase with I_1 and therefore lags V_s by 62.3°.
- V_{L1} leads I_1 by 90°, so its angle is 90° − 62.3° = 27.7°.

MULTISIM

Open Multisim file E12-13. Measure the voltage across each component and compare to the calculated values.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

- V_{R2} is in phase with I_2 and therefore lags V_s by 51.6°.
- V_{L2} leads I_2 by 90°, so its angle is $90° - 51.6° = 38.4°$.

These phase relationships are shown in Figure 12–30(b).

RELATED PROBLEM

What effect does an increase in frequency have on the total current in Figure 12–30?

SECTION 12–6 CHECKUP

1. Determine the total current for the circuit in Figure 12–29. *Hint:* Find the sum of the horizontal components of I_1 and I_2 and the sum of the vertical components of I_1 and I_2. Then apply the Pythagorean theorem to get I_{tot}.
2. What is the total impedance of the circuit in Figure 12–29?

12–7 POWER IN *RL* CIRCUITS

In a purely resistive ac circuit, all of the energy delivered by the source is dissipated in the form of heat by the resistance. In a purely inductive ac circuit, all of the energy delivered by the source is stored by the inductor in its magnetic field during a portion of the voltage cycle and then returned to the source during another portion of the cycle so that there is no net energy conversion to heat. When there is both resistance and inductance, some of the energy is alternately stored and returned by the inductance and some is dissipated by the resistance. The amount of energy converted to heat is determined by the relative values of the resistance and the inductive reactance.

After completing this section, you should be able to

- **Determine power in *RL* circuits**
  - **Explain true and reactive power**
  - **Draw the power triangle**
  - **Define *power factor***
  - **Explain power factor correction**

When the resistance in a series *RL* circuit is greater than the inductive reactance, more of the total energy delivered by the source is dissipated by the resistance than is stored and returned by the inductor. When the reactance is greater than the resistance, more of the total energy is stored and returned than is converted to heat.

As you know, the power dissipated in a resistor is called the true power. The power in an inductor is **reactive power** and is expressed as

$$P_r = I^2X_L \qquad (12\text{–}17)$$

The generalized power triangle for a series *RL* circuit is shown in Figure 12–31. The **apparent power,** P_a, is the resultant of the true power, P_{true}, and the reactive power, P_r.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

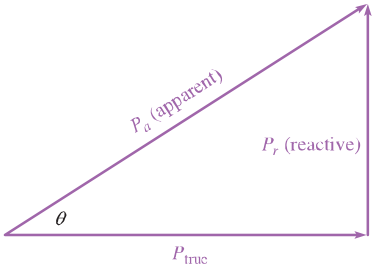

FIGURE 12–31 **Generalized power triangle for an *RL* circuit.**

6238262 2013/12/10 12.27.66.5

EXAMPLE 12–14

Determine the power factor, the true power, the reactive power, and the apparent power in Figure 12–32.

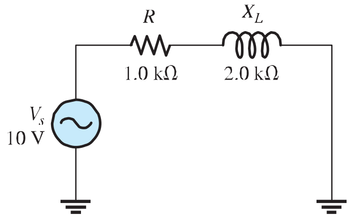

FIGURE 12–32

SOLUTION

The impedance of the circuit is

$$Z = \sqrt{R^2 + X_L^2} = \sqrt{(1.0\text{ k}\Omega)^2 + (2.0\text{ k}\Omega)^2} = 2.24\text{ k}\Omega$$

The current is

$$I = \frac{V_s}{Z} = \frac{10\text{ V}}{2.24\text{ k}\Omega} = 4.46\text{ mA}$$

The phase angle is

$$\theta = \tan^{-1}\left(\frac{X_L}{R}\right) = \tan^{-1}\left(\frac{2.0\text{ k}\Omega}{1.0\text{ k}\Omega}\right) = 63.4°$$

Power factor was defined in Equation 10–28. The power factor is

$$PF = \cos\theta = \cos(63.4°) = \mathbf{0.448}$$

The true power is

$$P_{true} = V_s I\cos\theta = (10\text{ V})(4.46\text{ mA})(0.448) = \mathbf{20\text{ mW}}$$

The reactive power is

$$P_r = I^2 X_L = (4.46\text{ mA})^2(2.0\text{ k}\Omega) = \mathbf{39.8\text{ mVAR}}$$

The apparent power is

$$P_a = I^2 Z = (4.46\text{ mA})^2(2.24\text{ k}\Omega) = \mathbf{44.6\text{ mVA}}$$

RELATED PROBLEM

If the frequency in Figure 12–32 is increased, what happens to P_{true}, P_r, and P_a?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Significance of the Power Factor

Recall that the **power factor** equals the cosine of θ ($PF = \cos\theta$). As the phase angle between the source voltage and the total current increases, the power factor decreases, indicating an increasingly reactive circuit. A smaller power factor indicates less true power and more reactive power. The power factor of inductive loads is called a *lagging power factor* because the current lags the source voltage.

As you learned in Chapter 10, the power factor (*PF*) is important in determining how much useful power (true power) is transferred to a load. The highest power factor is 1, which indicates that all of the current to a load is in phase with the voltage (resistive). When the power factor is 0, all of the current to a load is 90° out of phase with the voltage (reactive).

Motor Phase Shift

Large ac induction motors in industry can have a current that is five or six times the maximum rated current for a brief time. The starting impedance is purely inductive, meaning that while starting, the current lags the terminal voltage by 90° and the power factor is high. When the motor is up to speed, the current drops off to its rated current (or less), and the load looks much more resistive; thus, the phase shift and the power factor are reduced. Most large induction motors use a motor starter that inserts current-limiting resistors in series during startup to reduce the very high starting currents.

yukosourov/Fotolia.com

SYSTEM NOTE

Generally, a power factor as close to 1 as possible is desirable because most of the power transferred from the source to the load is useful or true power. True power goes only one way—from source to load—and performs work on the load in terms of energy dissipation. Reactive power simply goes back and forth between the source and the load with no net work being done. Energy must be used in order for work to be done.

Many practical loads have inductance as a result of their particular function, and it is essential for their proper operation. Examples are transformers, electric motors, and speakers, to name a few. Therefore, inductive (and capacitive) loads are important considerations.

To see the effect of the power factor on system requirements, refer to Figure 12–33. This figure shows a representation of a typical inductive load consisting effectively of

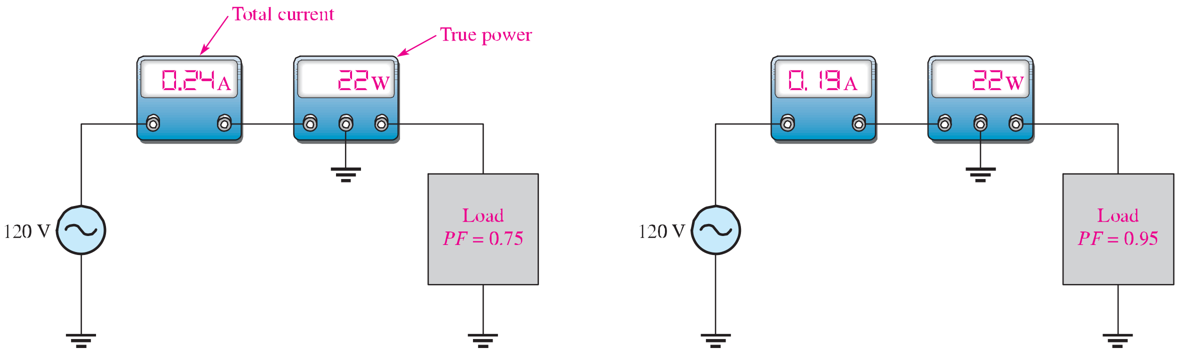

(a) A lower power factor means more total current for a given power dissipation (watts). A larger source is required to deliver the true power (watts).

(b) A higher power factor means less total current for a given power dissipation. A smaller source can deliver the same true power (watts).

FIGURE 12–33 **Illustration of the effect of the power factor on system requirements such as source rating (VA) and conductor size. Although the loads are different, they dissipate the same true power.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

inductance and resistance in parallel. Part (a) shows a load with a relatively low power factor (0.75), and part (b) shows a load with a relatively high power factor (0.95). Although both loads, as indicated by the wattmeters, dissipate the same amount of true power, the low power factor load in Figure 12–33(a) draws more current from the source than does the high power factor load in Figure 12–33(b), as indicated by the ammeters. Therefore, the source in part (a) must have a higher VA rating than the one in part (b). Also, the lines connecting the source to the load must be a larger wire gauge than those in part (b), a condition that becomes significant when very long transmission lines are required, such as in power distribution.

Figure 12–33 has demonstrated that a higher power factor is an advantage in delivering power more efficiently to a load. Also, because the electric company charges for apparent power, it is less expensive to have a high power factor.

POWER FACTOR CORRECTION The power factor of an inductive load can be increased by the addition of a capacitor in parallel, as shown in Figure 12–34. The capacitor compensates for the phase lag of the total current by creating a capacitive component of current that is 180° out of phase with the inductive component. This has a canceling effect and reduces the phase angle and the total current, as illustrated in the figure. This causes the power factor to increase.

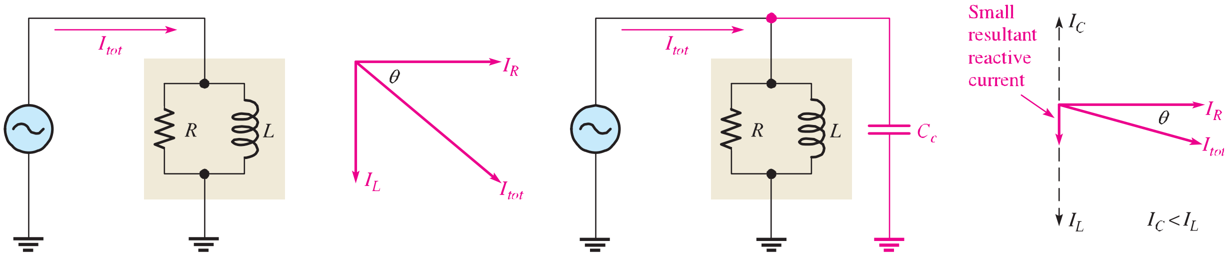

(a) Total current is the resultant of I_R and I_L.

(b) I_C subtracts from I_L, leaving only a small reactive current, thus decreasing I_{tot} and the phase angle.

FIGURE 12–34 Example of how the power factor can be increased by the addition of a compensating capacitor (C_c). As θ decreases, *PF* increases.

In industrial environments, three-phase motors, arc welders, and similar loads are common and tend to look like inductive loads, which cause the current to lag behind the voltage in each phase. The generalized power triangle, shown in Figure 12–31, can be applied to three-phase circuits by considering each phase as having true power and apparent power components. Ideally, the current waveform must track the voltage waveform for each phase (implying $PF = 1$), and the power drawn from each phase must be the same (this is referred to a *load balancing*.) With a balanced load, the capacitor-correction method can be used by connecting a compensating capacitor across each line, as was shown for the single-phase case.

In some cases, the load on each phase is unbalanced; different loads may cause distortion of the current waveform. Distortion is caused by nonlinear loads such as switched-mode power supplies. Loads that are intermittently on (such as arc welders) can also change the required power factor correction. These cases require more complicated solutions than just a compensating capacitor in each line; active circuits are available that can vary the phase angle.

SECTION 12–7 CHECKUP

1. To which component in an *RL* circuit is the power dissipation due?
2. Calculate the power factor when $\theta = 50°$.
3. A certain series *RL* circuit consists of a 470 Ω resistor and an inductive reactance of 620 Ω at the operating frequency. Determine P_{true}, P_r, and P_a when $I = 100$ mA.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

12–8 *RL* FILTERS

Like *RC* circuits, series *RL* circuits also exhibit a frequency-selective characteristic. An application of *RL* circuits is a basic frequency selective (filter) circuit.

After completing this section, you should be able to

- **Discuss how the *RL* circuit operates as a filter**

Low-Pass Characteristic

You have seen what happens to the phase angle and the output voltage in a series *RC* lag circuit. In terms of the filtering action of the series *RL* circuit, the variation in the magnitude of the output voltage as a function of frequency is important.

Figure 12–35 shows the filtering action of a series *RL* circuit using a specific series of measurements in which the frequency starts at 100 Hz and is increased in increments up to 20 kHz. At each value of frequency, the output voltage is measured. As you can see, the inductive reactance increases as frequency increases, thus causing less voltage to be dropped across the resistor while the input voltage is held at a constant 10 V throughout each step. The **frequency response** curve for these particular values would appear similar to the response curve in Figure 10–47 for the low-pass *RC* circuit.

6238262 2013/12/10 12.27.66.5

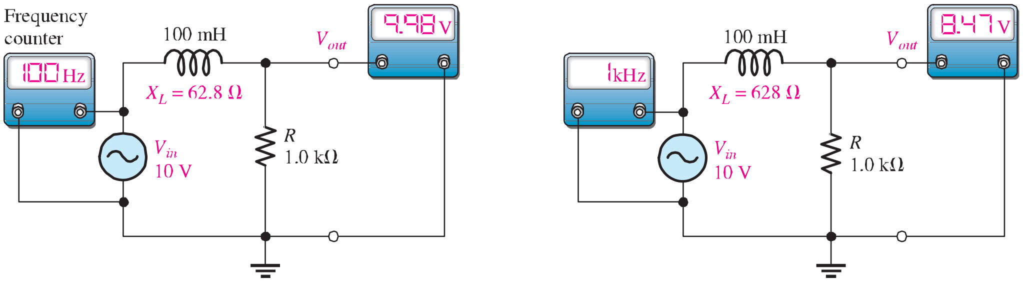

(a) $f = 100$ Hz, $X_L = 62.8\ \Omega$, $V_{out} = 9.98$ V (b) $f = 1$ kHz, $X_L = 628\ \Omega$, $V_{out} = 8.47$ V

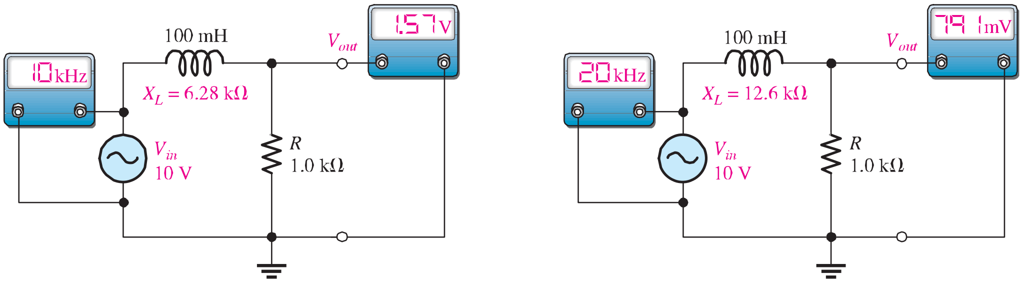

(c) $f = 10$ kHz, $X_L = 6.28$ kΩ, $V_{out} = 1.57$ V (d) $f = 20$ kHz, $X_L = 12.6$ kΩ, $V_{out} = 791$ mV

FIGURE 12–35 **Example of low-pass filtering action. Winding resistance has been neglected. As the input frequency increases, the output voltage decreases.**

High-Pass Characteristic

To illustrate *RL* high-pass filtering action, Figure 12–36 shows a series of specific measurements. The frequency starts at 10 Hz and is increased in increments up to 10 kHz. As you can see, the inductive reactance increases as the frequency increases, thus causing more voltage to be dropped across the inductor. Again, when the values are plotted, the response curve is similar to the one for the high-pass *RC* circuit that was shown in Figure 10–49.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

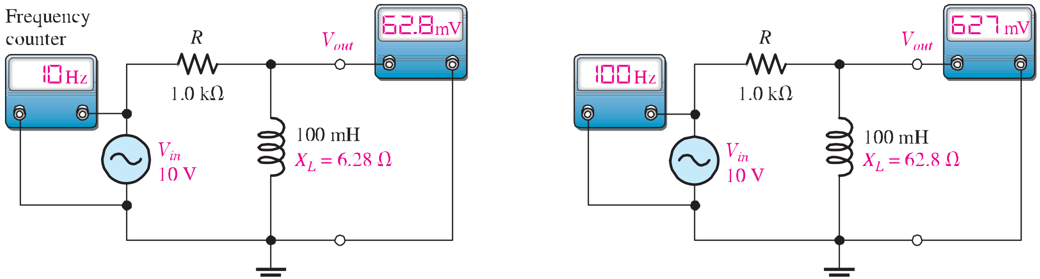
Frequency counter
10 Hz
R
1.0 kΩ
V_out
62.8 mV
V_in
10 V
100 mH
X_L = 6.28 Ω
100 Hz
R
1.0 kΩ
V_out
627 mV
V_in
10 V
100 mH
X_L = 62.8 Ω

(a) $f = 10$ Hz, $X_L = 6.28\ \Omega$, $V_{out} = 62.8$ mV (b) $f = 100$ Hz, $X_L = 62.8\ \Omega$, $V_{out} = 627$ mV

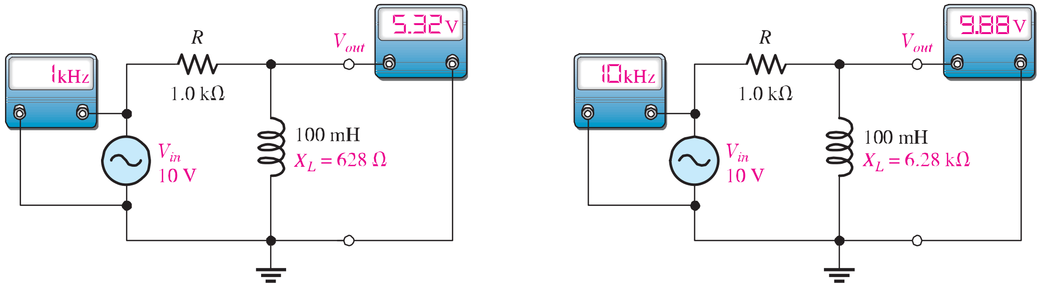
1 kHz
R
1.0 kΩ
V_out
5.32 V
V_in
10 V
100 mH
X_L = 628 Ω
10 kHz
R
1.0 kΩ
V_out
9.88 V
V_in
10 V
100 mH
X_L = 6.28 kΩ

(c) $f = 1$ kHz, $X_L = 628\ \Omega$, $V_{out} = 5.32$ V (d) $f = 10$ kHz, $X_L = 6.28$ kΩ, $V_{out} = 9.88$ V

FIGURE 12–36 **Example of high-pass filtering action. Winding resistance has been neglected. As the input frequency increases, the output voltage increases.**

The Cutoff Frequency of an *RL* Filter

The frequency at which the inductive reactance equals the resistance in a low-pass or high-pass *RL* filter is called the **cutoff frequency** and is designated f_c. This condition is expressed as $2\pi f_c L = R$. Solving for f_c results in the following formula:

$$f_c = \frac{R}{2\pi L} \qquad (12\text{–}18)$$

As with the *RC* filter, the output voltage is 70.7% of its maximum value at f_c. In a high-pass circuit, all frequencies above f_c are considered to be passed and all those below f_c are considered to be rejected. The reverse, of course, is true for a low-pass circuit. **Bandwidth**, defined in Chapter 10, applies to both *RC* and *RL* series circuits.

EXAMPLE 12–15

Calculate the cutoff frequency for the filters in Figure 12–37. Which is a high-pass filter and which is a low-pass filter?

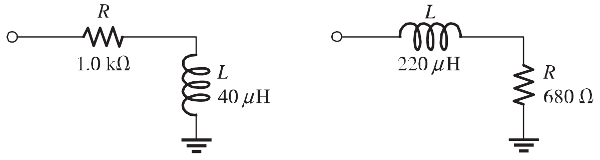
R
1.0 kΩ
L
40 μH
L
220 μH
R
680 Ω

(a) (b)

FIGURE 12–37

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

For Figure 12–32(a), the cutoff frequency is

$$f_c = \frac{R}{2\pi L} = \frac{1.0\,\text{k}\Omega}{2\pi(40\,\mu\text{H})} = \mathbf{3.98\ MHz}$$

This is a **high-pass filter**.

For Figure 12–32(b), the cutoff frequency is

$$f_c = \frac{R}{2\pi L} = \frac{680\ \Omega}{2\pi(220\,\mu\text{H})} = \mathbf{492\ kHz}$$

This is a **low-pass filter**.

RELATED PROBLEM

Assume you want to change the cutoff frequency of the high-pass filter to 10 MHz by changing the resistor. What value of resistance would you specify?

SYSTEM EXAMPLE 12–1

EMI FILTERS

An important application for inductors is for filters that select certain frequencies and reject others. A pervasive problem facing electronic engineers is the conducted or radiated interference called electromagnetic interference (EMI) that can create problems in other circuits and systems. In many systems, power supply inputs will contain an electromagnetic interference (EMI) input filter to limit potential interference from other circuits that are common to the supply through conduction.

The design of EMI filters starts by characterizing the problem, including the harmonic content of the interference, the type of noise, and the path (conductive or radiated). The interference can be any frequency but is typically in the radio frequency domain. The design of EMI filters can be quite complicated, depending on the sensitivity of the circuit being protected and the particular interference. Computers are often used as a design aid.

Figure 12–38 shows an actual EMI circuit that was designed for a dc-to-dc converter. Other EMI filters may look entirely different than this one but still have the goal of reducing interference.

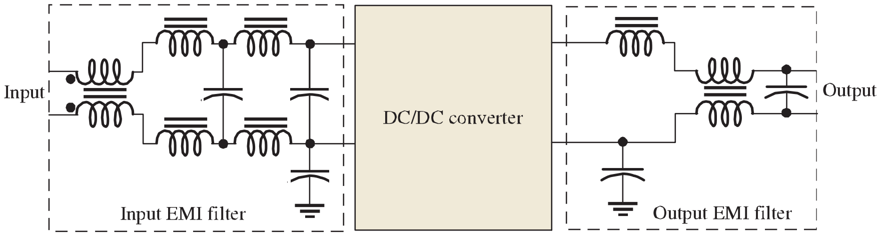

FIGURE 12–38 An EMI filter.

While the design of an EMI filter is beyond the scope of this course, it illustrates the application of inductors in complex filters. The filters in this section of text are basic *RL* filters, which are simplified frequency-selective circuits compared to the complicated EMI filter in this system example. An example of a basic filter (but not strictly *RL* or *RC*) is given in Figure 12–39, which represents a cross-over network to route high frequencies to one speaker (the "tweeter") and low frequencies to the other (the "woofer"). The speakers themselves are complex impedances, but an inductor will pass the low frequencies on to the woofer while attenuating the high frequencies, and a capacitor will pass the high frequencies on to the tweeter while attenuating the low frequencies.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

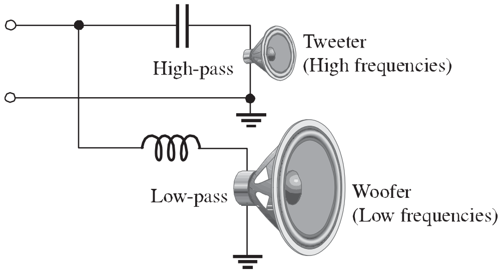

FIGURE 12–39 A cross-over network.

SECTION 12–8 CHECKUP

1. To achieve a low-pass characteristic with a series *RL* circuit, across which component is the output taken?
2. What is EMI?
3. What is the purpose of an EMI filter?

6238262 2013/12/10 12:27:00.5

12–9 TROUBLESHOOTING

Typical component failures have an effect on the frequency response of basic *RL* circuits.

After completing this section, you should be able to

- **Troubleshoot *RL* circuits**
  - **Find an open inductor**
  - **Find an open resistor**
  - **Find an open in a parallel circuit**

EFFECT OF AN OPEN INDUCTOR The most common failure mode for inductors occurs when the winding opens as a result of excessive current or a mechanical contact failure. For this reason, inductors are often encapsulated; this helps avoid breaking connections. It is easy to see how an open coil affects the operation of a basic series *RL* circuit, as shown in Figure 12–40. Obviously, there is no path for current, so the resistor voltage is zero and the total source voltage appears across the open inductor.

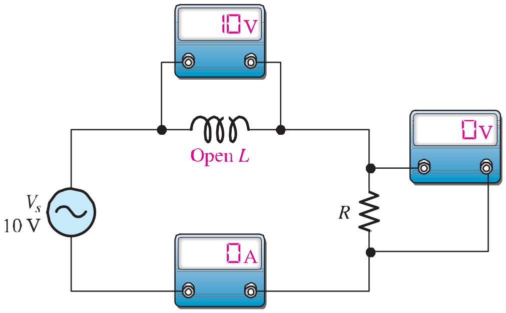

FIGURE 12–40 Effect of an open coil.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EFFECT OF AN OPEN RESISTOR When the resistor is open, there is no current, and the inductor voltage is zero. The total input voltage appears across the open resistor, as shown in Figure 12–41.

FIGURE 12–41 **Effect of an open resistor.**

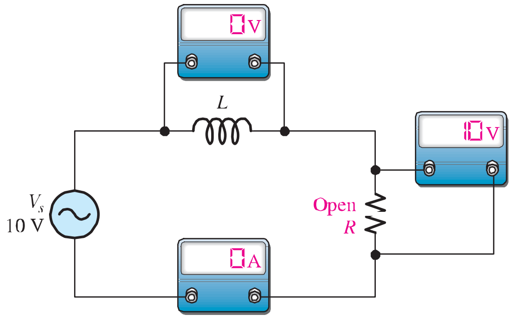

OPEN COMPONENTS IN PARALLEL CIRCUITS In a parallel *RL* circuit, an open resistor or inductor will cause the total current to decrease because the total impedance will increase. Obviously, the branch with the open component will have zero current. Figure 12–42 illustrates these conditions.

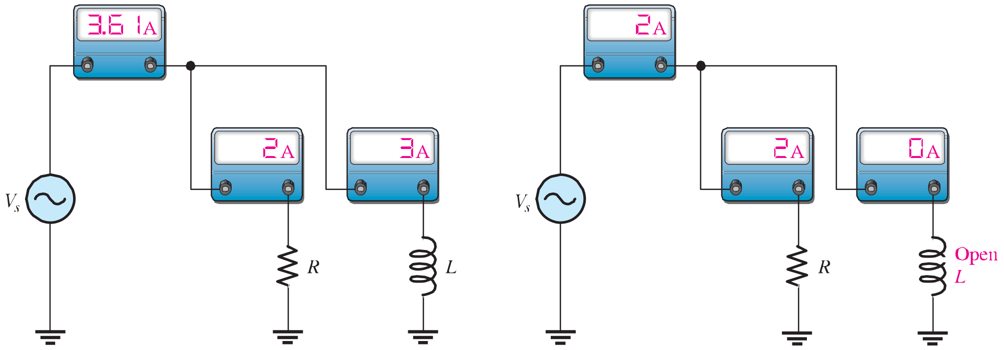

(a) Before failure

(b) After *L* opens

FIGURE 12–42 **Effect of an open component in a parallel circuit with V_s constant.**

EFFECT OF AN INDUCTOR WITH SHORTED WINDINGS Although a rare occurrence, it is possible for some of the windings of coils to short together as a result of damaged insulation. This failure mode is much less likely than the open coil. A sufficient number of shorted windings may result in a reduction in inductance because the inductance of a coil is proportional to the square of the number of turns.

Other Troubleshooting Considerations

The failure of a circuit to work properly is not always the result of a faulty component. A loose wire, a bad contact, or a poor solder joint can cause an open circuit. A short can be caused by a wire clipping or solder splash. Wrong values in a circuit (such as an incorrect resistor value), the function generator set at the wrong frequency, or the wrong output connected to the circuit can cause improper operation.

Always check to make sure that the instruments are properly connected to the circuits and to a power outlet. Also, look for obvious things such as a broken or loose contact, a connector that is not completely plugged in, or a piece of wire or a solder bridge that could be shorting something out.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 12–9 CHECKUP

1. Why are inductors frequently encapsulated?
2. Describe the effect of an inductor with shorted windings on the response of a series *RL* circuit.
3. In the circuit of Figure 12–43, indicate whether I_{tot}, V_{R1}, and V_{R2} increase or decrease as a result of *L* opening.

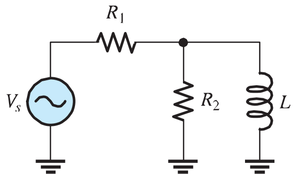

FIGURE 12–43

SUMMARY

- When a sinusoidal voltage is applied to an *RL* circuit, the current and all the voltage drops are also sine waves.
- Total current in an *RL* circuit always lags the source voltage.
- The resistor voltage is always in phase with the current.
- In an ideal inductor, the voltage always leads the current by 90°.
- In an *RL* lag circuit, the output voltage lags the input voltage in phase.
- In an *RL* lead circuit, the output voltage leads the input voltage in phase.
- In an *RL* circuit, the impedance is determined by both the resistance and the inductive reactance combined.
- Impedance is expressed in units of ohms.
- The impedance of an *RL* circuit varies directly with frequency.
- The phase angle (θ) of a series *RL* circuit varies directly with frequency.
- You can determine the impedance of a circuit by measuring the source voltage and the total current and then applying Ohm's law.
- In an *RL* circuit, part of the power is resistive and part reactive.
- The phasor combination of resistive power (true power) and reactive power is called *apparent power*.
- The power factor indicates how much of the apparent power is true power.
- A power factor of 1 indicates a purely resistive circuit, and a power factor of 0 indicates a purely reactive circuit.
- In a circuit that exhibits frequency selectivity, certain frequencies are passed to the output while others are rejected.

KEY TERMS

Key terms and other bold terms in the chapter are defined in the end-of-book glossary.

Inductive susceptance (B_L) The reciprocal of inductive reactance. The unit is the siemens.

***RL* lag circuit** A phase shift circuit in which the output voltage, taken across the resistor, lags the input voltage by a specified angle.

***RL* lead circuit** A phase shift circuit in which the output voltage, taken across the inductor, leads the input voltage by a specified angle.

KEY FORMULAS

12–1	$Z = \sqrt{R^2 + X_L^2}$	Series *RL* impedance
12–2	$\theta = \tan^{-1}\left(\frac{X_L}{R}\right)$	Series *RL* phase angle

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

12–3	$V_s = \sqrt{V_R^2 + V_L^2}$	Total voltage in a series *RL* circuit
12–4	$\theta = \tan^{-1}\left(\frac{V_L}{V_R}\right)$	Series *RL* phase angle
12–5	$\phi = \tan^{-1}\left(\frac{X_L}{R}\right)$	Phase angle of an *RL* lag circuit
12–6	$V_{out} = \left(\frac{R}{\sqrt{R^2 + X_L^2}}\right)V_{in}$	Output voltage of lag circuit
12–7	$\phi = 90° - \tan^{-1}\left(\frac{X_L}{R}\right)$	Phase angle of an *RL* lead circuit
12–8	$V_{out} = \left(\frac{X_L}{\sqrt{R^2 + X_L^2}}\right)V_{in}$	Output voltage of an *RL* lead circuit
12–9	$Z = \frac{RX_L}{\sqrt{R^2 + X_L^2}}$	Parallel *RL* impedance
12–10	$\theta = \tan^{-1}\left(\frac{R}{X_L}\right)$	Parallel *RL* phase angle
12–11	$G = \frac{1}{R}$	Conductance
12–12	$B_L = \frac{1}{X_L}$	Inductive susceptance
12–13	$Y = \frac{1}{Z}$	Admittance
12–14	$Y_{tot} = \sqrt{G^2 + B_L^2}$	Total admittance
12–15	$I_{tot} = \sqrt{I_R^2 + I_L^2}$	Total current in parallel *RL* circuit
12–16	$\theta = \tan^{-1}\left(\frac{I_L}{I_R}\right)$	Parallel *RL* phase angle
12–17	$P_r = I^2X_L$	Reactive power
12–18	$f_c = \frac{R}{2\pi L}$	Cutoff frequency of *RL* filter

TRUE/FALSE QUIZ

Answers are at the end of the chapter.

1. In an ac circuit where $R = X_L$, the phase angle is 45°.
2. In an ac series *RL* circuit where 3.0 V is dropped across the resistor and 4.0 V is dropped across the inductor, it follows that the source voltage is 5.0 V.
3. In an ac series *RL* circuit, the current and voltage are in phase for the inductor.
4. In an ac parallel *RL* circuit, the inductive susceptance is always less than the admittance.
5. In an ac parallel *RL* circuit, the voltage across the inductor is out of phase with the voltage across the resistor.
6. The unit *siemens* is used to measure both susceptance and admittance.
7. The reciprocal of impedance is susceptance.
8. If the power factor of a circuit is 0.5, the reactive power and true power are equal.
9. A purely resistive circuit has a power factor of 0.
10. A high-pass series *RL* filter has the output taken from across the resistor.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SELF-TEST

Answers are at the end of the chapter.

1. In a series *RL* circuit, the resistor voltage

(a) leads the source voltage **(b)** lags the source voltage
(c) is in phase with the source voltage **(d)** is in phase with the current
(e) answers (a) and (d) **(f)** answers (b) and (d)

2. When the frequency of the voltage applied to a series *RL* circuit is increased, the impedance

(a) decreases **(b)** increases **(c)** does not change

3. When the frequency of the voltage applied to a series *RL* circuit is decreased, the phase angle

(a) decreases **(b)** increases **(c)** does not change

4. If the frequency is doubled and the resistance is halved, the impedance of a series *RL* circuit

(a) doubles **(b)** halves
(c) remains constant **(d)** cannot be determined without values

5. To reduce the current in a series *RL* circuit, the frequency should be

(a) increased **(b)** decreased **(c)** constant

6. In a series *RL* circuit, 10 V rms is measured across the resistor, and 10 V rms is measured across the inductor. The peak value of the source voltage is

(a) 14.14 V **(b)** 28.28 V
(c) 10 V **(d)** 20 V

7. The voltages in Question 6 are measured at a certain frequency. To make the resistor voltage greater than the inductor voltage, the frequency is

(a) increased **(b)** decreased
(c) doubled **(d)** not a factor

8. When the resistor voltage in a series *RL* circuit becomes greater than the inductor voltage, the phase angle

(a) increases **(b)** decreases **(c)** is not affected

9. When the frequency is increased, the impedance of a parallel *RL* circuit

(a) increases **(b)** decreases **(c)** remains constant

10. In a parallel *RL* circuit, there are 2 A rms in the resistive branch and 2 A rms in the inductive branch. The total rms current is

(a) 4 A **(b)** 5.656 A **(c)** 2 A **(d)** 2.828 A

11. You are observing two voltage waveforms on an oscilloscope. The time base (sec/div) of the scope is set to 10 μs. One half-cycle of the waveforms covers the ten horizontal divisions. The positive-going zero crossing of one waveform is at the leftmost division, and the positive-going zero crossing of the other is three divisions to the right. The phase angle between these two waveforms is

(a) 18° **(b)** 36° **(c)** 54° **(d)** 180°

12. Which of the following power factors results in the least energy conversion to heat in an *RL* circuit?

(a) 1 **(b)** 0.9 **(c)** 0.5 **(d)** 0.1

13. If a load is purely inductive and the reactive power is 10 VAR, the apparent power is

(a) 0 VA **(b)** 10 VA
(c) 14.14 VA **(d)** 3.16 VA

14. For a certain load, the true power is 10 W and the reactive power is 10 VAR. The apparent power is

(a) 5 VA **(b)** 20 VA
(c) 14.14 VA **(d)** 100 VA

15. The cutoff frequency of a certain low-pass *RL* filter is 20 kHz. The bandwidth is

(a) 20 kHz **(b)** 40 kHz
(c) 0 kHz **(d)** unknown

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

TROUBLESHOOTING: SYMPTOM AND CAUSE

The purpose of these exercises is to help develop thought processes essential to troubleshooting. Answers are at the end of the chapter.

Determine the cause for each set of symptoms. Refer to Figure 12–44.

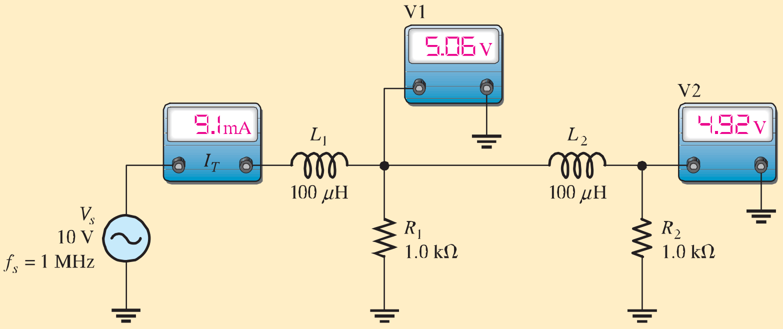

FIGURE 12–44 **The ac meters indicate the correct readings for this circuit.**

1. *Symptom:* The ammeter reading is 15.9 mA, and the voltmeter 1 and voltmeter 2 readings are zero.

   Cause:

   (a) L_1 is open.
   (b) L_2 is open.
   (c) R_1 is shorted.

2. *Symptom:* The ammeter reading is 8.47 mA, the voltmeter 1 reading is 8.47 V, and the voltmeter 2 reading is 0 V.

   Cause:

   (a) L_2 is open.
   (b) R_2 is open.
   (c) R_2 is shorted.

3. *Symptom:* The ammeter reading is slightly less than 20 mA, and both voltmeter readings are slightly less than 10 V.

   Cause:

   (a) L_1 is shorted.
   (b) R_1 is open.
   (c) The source frequency is incorrectly set very low.

4. *Symptom:* The ammeter reading is 4.55 mA, voltmeter 1 reads 2.53 V, and voltmeter 2 reads 2.15 V.

   Cause:

   (a) The source frequency is incorrectly set at 500 kHz.
   (b) The source voltage is incorrectly set at 5 V.
   (c) The source frequency is incorrectly set at 2 MHz.

5. *Symptom:* All meter readings are zero.

   Cause:

   (a) The voltage source is faulty or off.
   (b) L_1 is open.
   (c) Either (a) or (b)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

PROBLEMS

6238262 2013/12/10 12.27.66.5

Answers to odd-numbered problems are at the end of the book.

BASIC PROBLEMS

SECTION 12–1 Sinusoidal Response of *RL* Circuits

1. A 15 kHz sinusoidal voltage is applied to a series *RL* circuit. Determine the frequency of I, V_R, and V_L.

2. What are the waveshapes of I, V_R, and V_L in Problem 1?

SECTION 12–2 Impedance and Phase Angle of Series *RL* Circuits

3. Find the impedance of each circuit in Figure 12–45.

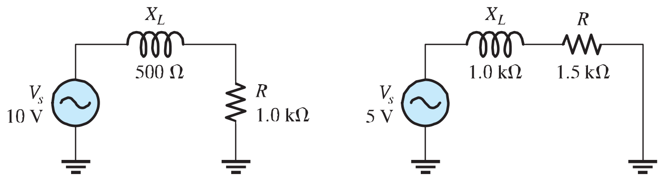

(a) (b)

FIGURE 12–45

4. Determine the impedance and phase angle for the circuit in Figure 12–46.

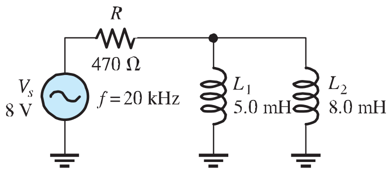

FIGURE 12–46

5. In Figure 12–47, determine the impedance at each of the following frequencies:
(a) 100 Hz **(b)** 500 Hz **(c)** 1 kHz **(d)** 2 kHz

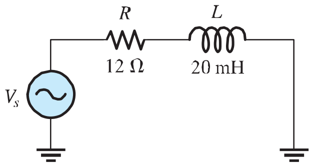

FIGURE 12–47

6. Determine the values of R and X_L in a series *RL* circuit for the following values of impedance and phase angle:
(a) $Z = 20\ \Omega, \theta = 45°$ **(b)** $Z = 500\ \Omega, \theta = 35°$
(c) $Z = 2.5\ \text{k}\Omega, \theta = 72.5°$ **(d)** $Z = 998\ \Omega, \theta = 45°$

SECTION 12–3 Analysis of Series *RL* Circuits

7. Assume that the inductance in Figure 12–45(a) is 796 μH. What is the source frequency?

8. Determine the voltage across the total resistance and across the total inductance in Figure 12–46.

9. Find the current for each circuit of Figure 12–45.

10. Calculate the total current for the circuit of Figure 12–46.

11. Determine θ for the circuit in Figure 12–48.

12. If the inductance in Figure 12–48 is doubled, does θ increase or decrease, and by how many degrees?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

13. Draw the waveforms for V_s, V_R, and V_L in Figure 12–48. Show the proper phase relationships.

14. For the circuit in Figure 12–49, find V_R and V_L for each of the following frequencies:
(a) 60 Hz **(b)** 200 Hz **(c)** 500 Hz **(d)** 1 kHz

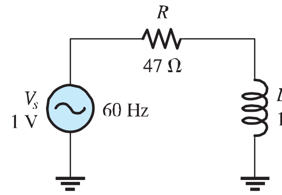

FIGURE 12–48

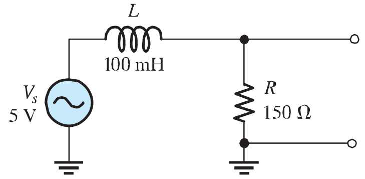

FIGURE 12–49

SECTION 12–4 Impedance and Phase Angle of Parallel *RL* Circuits

15. What is the impedance for the circuit in Figure 12–50?

16. Repeat Problem 15 for the following frequencies:
(a) 1.5 kHz **(b)** 3 kHz **(c)** 5 kHz **(d)** 10 kHz

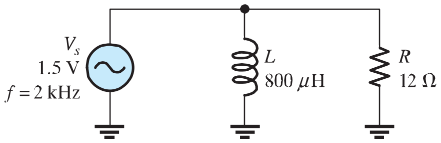

FIGURE 12–50

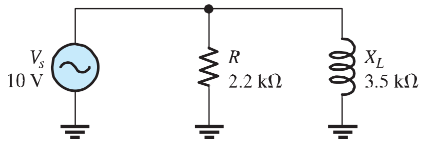

FIGURE 12–51

17. What is the admittance of the circuit in Figure 12–51?

18. What is the admittance of the circuit in Figure 15–51 if the frequency is doubled?

19. At what frequency does X_L equal R in Figure 12–50?

SECTION 12–5 Analysis of Parallel *RL* Circuits

20. Find the total current and each branch current in Figure 12–51.

21. Determine the following quantities in Figure 12–52:
(a) Z **(b)** I_R **(c)** I_L **(d)** I_{tot} **(e)** θ

22. Convert the circuit in Figure 12–53 to an equivalent series form.

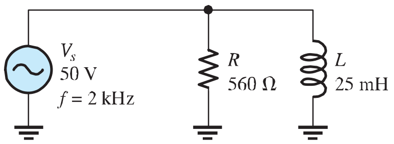

FIGURE 12–52

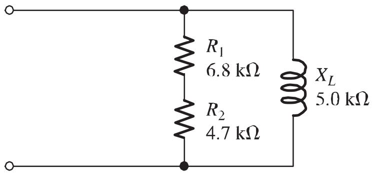

FIGURE 12–53

SECTION 12–6 Analysis of Series-Parallel *RL* Circuits

23. Determine the voltage across each element in Figure 12–54.

24. Is the circuit in Figure 12–54 predominantly resistive or predominantly inductive?

25. Find the current in each branch and the total current in Figure 12–54.

FIGURE 12–54

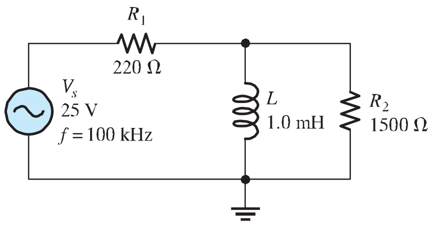

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 12–7 Power in *RL* Circuits

26. In a certain *RL* circuit, the true power is 100 mW, and the reactive power is 340 mVAR. What is the apparent power?

27. Determine the true power and the reactive power in Figure 12–48.

28. What is the power factor in Figure 12–51?

29. Determine P_{true}, P_r, P_a, and *PF* for the circuit in Figure 12–54. Draw the power triangle.

SECTION 12–8 *RL* Filters

30. Plot the response curve for the circuit in Figure 12–49. Show the output voltage versus frequency in 200 Hz increments from 0 to 1200 Hz.

31. Using the same procedure as in Problem 30, plot the response curve for Figure 12–55.

32. Draw the voltage phasor diagram for the circuit in Figure 12–55 for a frequency of 500 Hz.

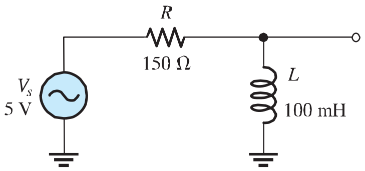

R
150 Ω
V_s
5 V
L
100 mH

FIGURE 12–55

SECTION 12–9 Troubleshooting

33. Determine the voltage across each component in Figure 12–56 if L_1 were open.

34. Determine if the voltage reading on the DMM in Figure 12–57 is correct or not. Explain your answer.

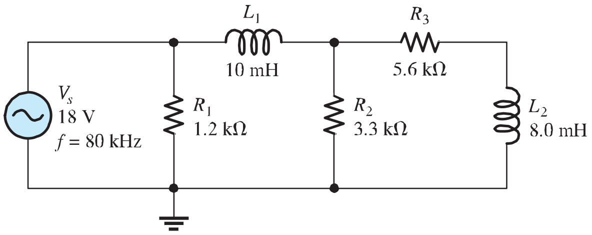

L_1
10 mH
R_3
5.6 kΩ
V_s
18 V
f = 80 kHz
R_1
1.2 kΩ
R_2
3.3 kΩ
L_2
8.0 mH

FIGURE 12–56

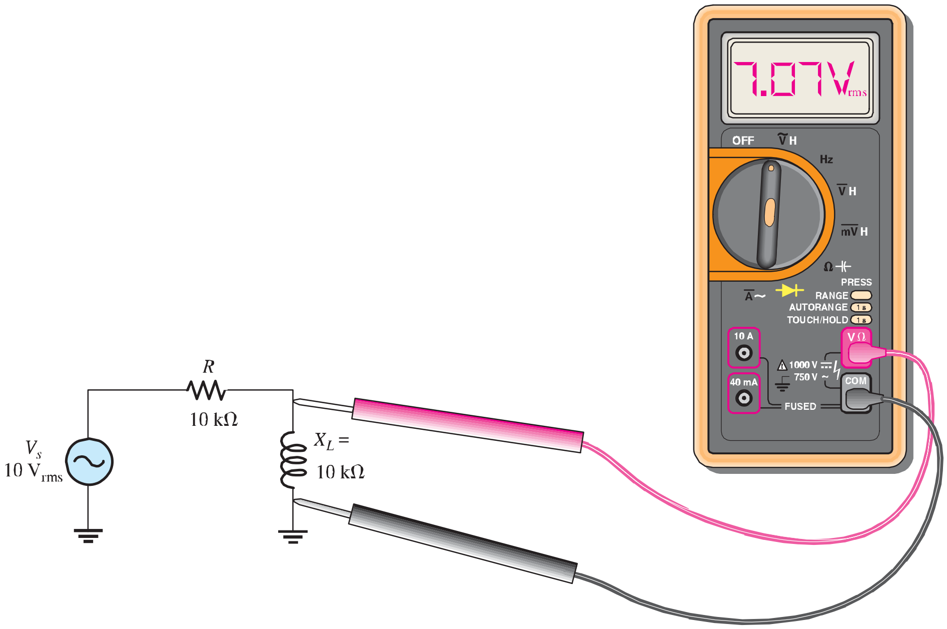

7.07V rms
OFF
V H
Hz
V H
mV H
Ω
PRESS
RANGE
AUTORANGE
TOUCH/HOLD
A~
10 A
40 mA
V Ω
1000 V
750 V ~
COM
FUSED
R
10 kΩ
V_s
10 V_{rms}
X_L =
10 kΩ

FIGURE 12–57

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

ADVANCED PROBLEMS

35. Determine the voltage across the inductors in Figure 12–58.

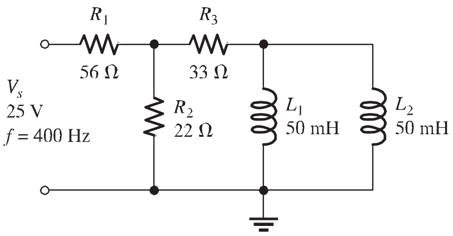

FIGURE 12–58

36. Is the circuit in Figure 12–58 predominantly resistive or predominantly inductive?

37. Find the total current in Figure 12–58.

38. For the circuit in Figure 12–59, determine the following:
(a) Z_{tot} **(b)** I_{tot} **(c)** θ **(d)** V_L **(e)** V_{R3}

39. For the circuit in Figure 12–60, determine the following:
(a) I_{R1} **(b)** I_{L1} **(c)** I_{L2} **(d)** I_{R2}

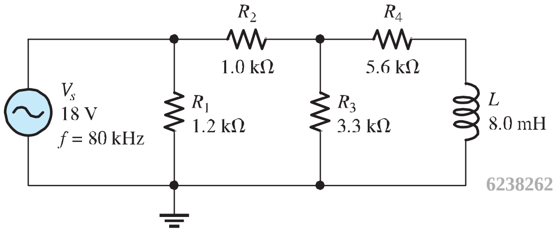

FIGURE 12–59

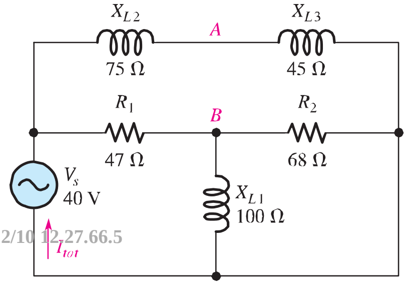

6238262 2013/12/10 12.27.66.5

FIGURE 12–60

MULTISIM

MULTISIM TROUBLESHOOTING PROBLEMS

40. Open file P12-40; files are found at www.pearsonhighered.com/floyd. Determine if there is a fault and, if so, identify it.

41. Open file P12-41. Determine if there is a fault and, if so, identify it.

42. Open file P12-42. Determine if there is a fault and, if so, identify it.

43. Open file P12-43. Determine if there is a fault and, if so, identify it.

44. Open file P12-44. Determine if there is a fault and, if so, identify it.

45. Open file P12-45. Determine if there is a fault and, if so, identify it.

ANSWERS TO SECTION CHECKUPS

SECTION 12–1 Sinusoidal Response of *RL* Circuits

1. The current frequency is at 1 kHz.
2. θ is closer to 0° when $R > X_L$.

SECTION 12–2 Impedance and Phase Angle of Series *RL* Circuits

1. V_s leads I.
2. $X_L = R$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

3. In an *RL* circuit, current lags voltage; in an *RC* circuit, current leads voltage.
4. $Z = \sqrt{R^2 + X_L^2} = 59.9\text{ k}\Omega;\ \theta = \tan^{-1}(X_L/R) = 56.6°$

SECTION 12–3 Analysis of Series *RL* Circuits

1. $V_s = \sqrt{V_R^2 + V_L^2} = 3.61\text{ V}$
2. $\theta = \tan^{-1}(X_L/V_R) = 56.3°$
3. When f increases, X_L increases, Z increases, and θ increases.
4. $\phi = 90° - \tan^{-1}(X_L/R) = 81.9°$
5. $V_{out} = \left(\dfrac{R}{\sqrt{R^2 + X_L^2}}\right)V_{in} = 9.90\text{ V}$

SECTION 12–4 Impedance and Phase Angle of Parallel *RL* Circuits

1. $Z = 1/Y = 20\ \Omega$
2. $Y = \sqrt{G^2 + B_L^2} = 2.5\text{ mS}$
3. I_{tot} lags V_s by 32.1°.

SECTION 12–5 Analysis of Parallel *RL* Circuits

1. $I_{tot} = V_sY = 32\text{ mA}$
2. $\theta = \tan^{-1}(I_L/I_R) = 59.0°;\ I_{tot} = \sqrt{I_R^2 + I_L^2} = 23.3\text{ mA}$
3. $\theta = 90°$

SECTION 12–6 Analysis of Series-Parallel *RL* Circuits

1. $I_{tot} = \sqrt{(I_1\cos\theta_1 + I_2\cos\theta_2)^2 + (I_1\sin\theta_1 + I_2\sin\theta_2)^2} = 20.2\text{ mA}$
2. $Z = V_s/I_{tot} = 494\ \Omega$

SECTION 12–7 Power in *RL* Circuits

1. Power dissipation is in the resistor.
2. $PF = \cos 50° = 0.643$
3. $P_{\text{true}} = I^2R = 4.7\text{ W};\ P_r = I^2X_L = 6.2\text{ VAR};\ P_a = \sqrt{P_{\text{true}}^2 + P_r^2} = 7.78\text{ VA}$

SECTION 12–8 *RL* Filters

1. Output is across the resistor.
2. Electromagnetic interference, caused when a circuit conducts or radiates electrical noise into another circuit
3. To reduce interference

SECTION 12–9 Troubleshooting

1. Encapsulation helps avoid open circuits due to stressing the leads on an inductor.
2. Shorted windings reduce L and thereby reduce X_L at any given frequency.
3. I_{tot} decreases, V_{R1} decreases, and V_{R2} increases when L opens.

ANSWERS TO RELATED PROBLEMS FOR EXAMPLES

12–1 2.04 kΩ; 27.8°
12–2 423 μA
12–3 No
12–4 12.6 kΩ; 85.5°
12–5 32°
12–6 12.3 V rms
12–7 65.6°

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

12–8 8.14 kΩ; 35.5°

12–9 3.70 mS

12–10 14.0 mA; 71.1°

12–11 14.9 mA

12–12 **(a)** 8.04 V
(b) 2.00 V

12–13 Current decreases.

12–14 P_{true}, P_r, and P_a decrease.

12–15 2.5 kΩ

ANSWERS TO TRUE/FALSE QUIZ

1. T **2.** T **3.** F **4.** T **5.** F **6.** T **7.** F
8. T **9.** F **10.** F

ANSWERS TO SELF-TEST

1. (f) **2.** (b) **3.** (a) **4.** (d) **5.** (a) **6.** (d) **7.** (b) **8.** (b)
9. (a) **10.** (d) **11.** (c) **12.** (d) **13.** (b) **14.** (c) **15.** (a)

ANSWERS TO TROUBLESHOOTING: SYMPTOM AND CAUSE

1. (c) **2.** (a) **3.** (c) **4.** (b) **5.** (c)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

CHAPTER 13

RLC CIRCUITS AND RESONANCE

OUTLINE

13–1 Impedance and Phase Angle of Series *RLC* Circuits
13–2 Analysis of Series *RLC* Circuits
13–3 Series Resonance
13–4 Series Resonant Filters
13–5 Parallel *RLC* Circuits
13–6 Parallel Resonance
13–7 Parallel Resonant Filters
13–8 Resonant Circuit Applications

OBJECTIVES

- Determine the impedance and phase angle of a series *RLC* circuit
- Analyze series *RLC* circuits
- Analyze a circuit for series resonance
- Analyze series resonant filters
- Analyze parallel *RLC* circuits
- Analyze a circuit for parallel resonance
- Analyze the operation of parallel resonant filters
- Discuss some applications of resonant circuits

KEY TERMS

Series resonance
Resonant frequency
Band-pass filter
Cutoff frequency
Half-power frequency
Decibel (dB)
Selectivity
Band-stop filter
Parallel resonance

INTRODUCTION

In this chapter, you will learn about frequency response of circuits with combinations of resistance, capacitance, and inductance (*RLC*). Both series and parallel *RLC* circuits, including the concepts of series and parallel resonance will be discussed.

Because it is the basis for frequency selectivity, resonance in electrical circuits is important to the operation of many types of electronic systems, particularly in the area of communications. For example, the ability of a radio or television receiver to select a certain frequency transmitted by a particular station and to eliminate frequencies from other stations is based on the principle of resonance.

6238262 2013/12/10 12.27.66.5

The operation of both band-pass and band-stop filters is based on resonance of circuits containing inductance and capacitance, and these filters are discussed in this chapter. Both types of filters are widely used in communication systems.

VISIT THE WEBSITE
Study aids for this chapter are available at
http://pearsonhighered.com/floyd

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

13–1 IMPEDANCE AND PHASE ANGLE OF SERIES *RLC* CIRCUITS

A series *RLC* circuit contains resistance, inductance, and capacitance. Since inductive reactance and capacitive reactance have opposite effects on the circuit phase angle, the total reactance is less than either individual reactance.

After completing this section, you should be able to

- **Determine the impedance and phase angle of a series *RLC* circuit**
  - **Calculate total reactance**
  - **Determine whether a circuit is predominately inductive or capacitive**

A series *RLC* circuit is shown in Figure 13–1. It contains resistance, inductance, and capacitance.

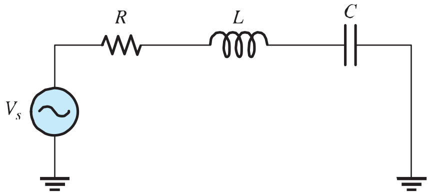

FIGURE 13–1 **Series *RLC* circuit.**

As you know, inductive reactance (X_L) causes the total current to lag the source voltage. Capacitive reactance (X_C) has the opposite effect: It causes the current to lead the voltage. Thus, X_L and X_C tend to offset each other. When they are equal, they cancel, and the total reactance is zero. In any case, the total reactance in a series circuit is

$$\boldsymbol{X_{tot} = |X_L - X_C|} \tag{13–1}$$

The term $|X_L - X_C|$ means the absolute value of the difference of the two reactances. That is, the sign of the result is considered positive no matter which reactance is greater. For example, $3 - 7 = -4$, but the absolute value is

$$|3 - 7| = 4$$

When $X_L > X_C$, the circuit is predominantly inductive; and when $X_C > X_L$, the circuit is predominantly capacitive.

The total impedance for a series *RLC* circuit is given by

$$\boldsymbol{Z_{tot} = \sqrt{R^2 + X_{tot}^2}} \tag{13–2}$$

and the value of the phase angle between V_s and I is

$$\boldsymbol{\theta = \tan^{-1}\left(\frac{X_{tot}}{R}\right)} \tag{13–3}$$

The phase angle is positive (current leading voltage) when the circuit is inductive and negative (current lagging voltage) when the circuit is capacitive.

EXAMPLE 13–1

Determine the total impedance and the phase angle for the series *RLC* circuit in Figure 13–2.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

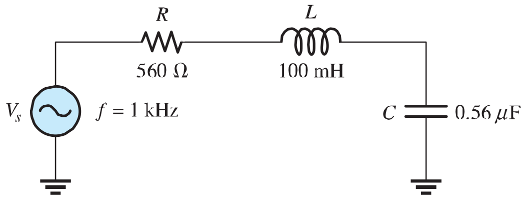

FIGURE 13–2

SOLUTION

First, find X_C and X_L.

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi(1\text{ kHz})(0.56\ \mu\text{F})} = 284\ \Omega$$

$$X_L = 2\pi fL = 2\pi(1\text{ kHz})(100\text{ mH}) = 628\ \Omega$$

In this case, X_L is greater than X_C, so the circuit is more inductive than capacitive. The magnitude of the total reactance is

$$X_{tot} = |X_L - X_C| = |628\ \Omega - 284\ \Omega| = 344\ \Omega \qquad \text{(inductive)}$$

The total circuit impedance is

$$Z_{tot} = \sqrt{R^2 + X_{tot}^2} = \sqrt{(560\ \Omega)^2 + (344\ \Omega)^2} = \mathbf{657\ \Omega}$$

The phase angle (between I and V_s) is

$$\theta = \tan^{-1}\left(\frac{X_{tot}}{R}\right) = \tan^{-1}\left(\frac{344\ \Omega}{560\ \Omega}\right) = \mathbf{31.6^\circ} \qquad \text{(current lagging } V_s\text{)}$$

RELATED PROBLEM*

Increase the frequency in Figure 13–2 to 2000 Hz and determine Z and θ.

***Answers are at the end of the chapter.**

As you have seen, when the inductive reactance is greater than the capacitive reactance, the circuit appears to be inductive, and the current lags the source voltage. When the capacitive reactance is greater, the circuit appears to be capacitive, and the current leads the source voltage.

SECTION 13–1 CHECKUP*

1. Explain how you can determine if a series *RLC* circuit is inductive or capacitive.
2. In a given series *RLC* circuit, X_C is 150 Ω and X_L is 80 Ω. What is the total reactance in ohms? Is the circuit inductive or capacitive?
3. Determine the impedance for the circuit in Question 2 when $R = 45\ \Omega$. What is the phase angle? Is the current leading or lagging the source voltage?

***Answers are at the end of the chapter.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

13–2 ANALYSIS OF SERIES *RLC* CIRCUITS

Recall that capacitive reactance varies inversely with frequency and that inductive reactance varies directly with frequency. In this section, the combined effects of the reactances as a function of frequency are examined.

After completing this section, you should be able to

- **Analyze series *RLC* circuits**
  - **Determine current in a series *RLC* circuit**
  - **Determine the voltages in a series *RLC* circuit**
  - **Determine the phase angle**

6238262 2013/12/10 12.27.66.5

Figure 13–3 shows that for a typical series *RLC* circuit the total reactance behaves as follows: Starting at a very low frequency, X_C is high, X_L is low, and the circuit is predominantly capacitive. As the frequency is increased, X_C decreases and X_L increases until a value is reached where $X_C = X_L$ and the two reactances cancel, making the circuit purely resistive. This condition is **series resonance** and will be studied in Section 13–3. As the frequency is increased further, X_L becomes greater than X_C, and the circuit is predominantly inductive. Example 13–2 illustrates how the impedance and phase angle change as the source frequency is varied.

FIGURE 13–3 **How X_C and X_L vary with frequency.**

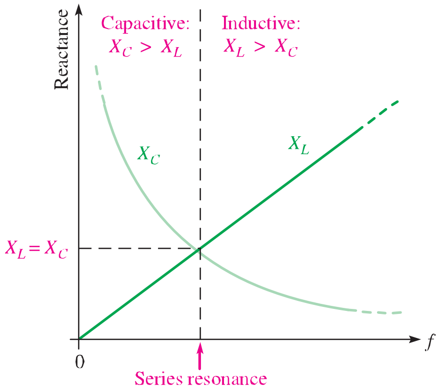

The graph of X_L is a straight line, and the graph of X_C is curved, as shown in Figure 13–3. The general equation for a straight line is $y = mx + b$, where m is the slope of the line and b is the y-axis intercept point. The formula $X_L = 2\pi fL$ fits this general straight-line formula, where $y = X_L$ (a variable), $m = 2\pi L$ (a constant), $x = f$ (a variable), and $b = 0$ as follows: $X_L = 2\pi Lf + 0$.

The X_C curve is called a *hyperbola*, and the general equation of a hyperbola is $xy = k$. The equation for capacitive reactance, $X_C = 1/2\pi fC$, can be rearranged as $X_C f = 1/2\pi C$ where $x = X_C$ (a variable), $y = f$ (a variable), and $k = 1/2\pi C$ (a constant).

EXAMPLE 13–2

For each of the following frequencies of the source voltage, find the impedance and the phase angle for the circuit in Figure 13–4. Note the change in the impedance and the phase angle with frequency.

(a) $f = 1\text{ kHz}$

(b) $f = 3.5\text{ kHz}$

(c) $f = 5\text{ kHz}$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

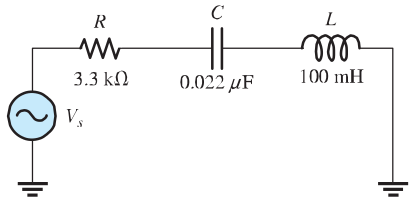

FIGURE 13–4

SOLUTION

(a) At $f = 1$ kHz,

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi(1\text{ kHz})(0.022\ \mu\text{F})} = 7.23\text{ k}\Omega$$

$$X_L = 2\pi fL = 2\pi(1\text{ kHz})(100\text{ mH}) = 628\ \Omega$$

The circuit is highly capacitive because X_C is much larger than X_L. The magnitude of the total reactance is

$$X_{tot} = |X_L - X_C| = |628\ \Omega - 7.23\text{ k}\Omega| = 6.6\text{ k}\Omega$$

The impedance is

$$Z = \sqrt{R^2 + X_{tot}^2} = \sqrt{(3.3\text{ k}\Omega)^2 + (6.6\text{ k}\Omega)^2} = \mathbf{7.38\text{ k}\Omega}$$

The phase angle is

$$\theta = \tan^{-1}\left(\frac{X_{tot}}{R}\right) = \tan^{-1}\left(\frac{6.60\text{ k}\Omega}{3.3\text{ k}\Omega}\right) = \mathbf{63.4°}$$

I leads V_s by 63.4°.

(b) At $f = 3.5$ kHz,

$$X_C = \frac{1}{2\pi(3.5\text{ kHz})(0.022\ \mu\text{F})} = 2.07\text{ k}\Omega$$

$$X_L = 2\pi(3.5\text{ kHz})(100\text{ mH}) = 2.20\text{ k}\Omega$$

The circuit is very close to being purely resistive but is slightly inductive because X_L is slightly larger than X_C. The total reactance, impedance, and phase angle are

$$X_{tot} = |2.20\text{ k}\Omega - 2.07\text{ k}\Omega| = 130\ \Omega$$

$$Z = \sqrt{(3.3\text{ k}\Omega)^2 + (130\ \Omega)^2} = \mathbf{3.30\text{ k}\Omega}$$

$$\theta = \tan^{-1}\left(\frac{130\ \Omega}{3.3\text{ k}\Omega}\right) = \mathbf{2.26°}$$

I lags V_s by 2.26°.

(c) At $f = 5$ kHz,

$$X_C = \frac{1}{2\pi(5\text{ kHz})(0.022\ \mu\text{F})} = 1.45\text{ k}\Omega$$

$$X_L = 2\pi(5\text{ kHz})(100\text{ mH}) = 3.14\text{ k}\Omega$$

The circuit is now predominantly inductive because $X_L > X_C$. The total reactance, impedance, and phase angle are

$$X_{tot} = |3.14\text{ k}\Omega - 1.45\text{ k}\Omega| = 1.69\text{ k}\Omega$$

$$Z = \sqrt{(3.3\text{ k}\Omega)^2 + (1.69\text{ k}\Omega)^2} = \mathbf{3.71\text{ k}\Omega}$$

$$\theta = \tan^{-1}\left(\frac{1.69\text{ k}\Omega}{3.3\text{ k}\Omega}\right) = \mathbf{27.1°}$$

I lags V_s by 27.1°.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Notice how the circuit changed from capacitive to inductive as the frequency increased. The phase condition changed from the current leading to the current lagging. It is important to note that both the impedance and the phase angle decreased to a minimum and then began increasing again as the frequency went up.

RELATED PROBLEM

Determine Z for $f = 7$ kHz in Figure 13–4.

In a series *RLC* circuit, the capacitor voltage and the inductor voltage are always 180° out of phase with each other. For this reason, V_C and V_L subtract from each other, and thus the voltage across L and C combined is always less than the larger individual voltage across either component, as illustrated in Figure 13–5 and in the waveform diagram of Figure 13–6.

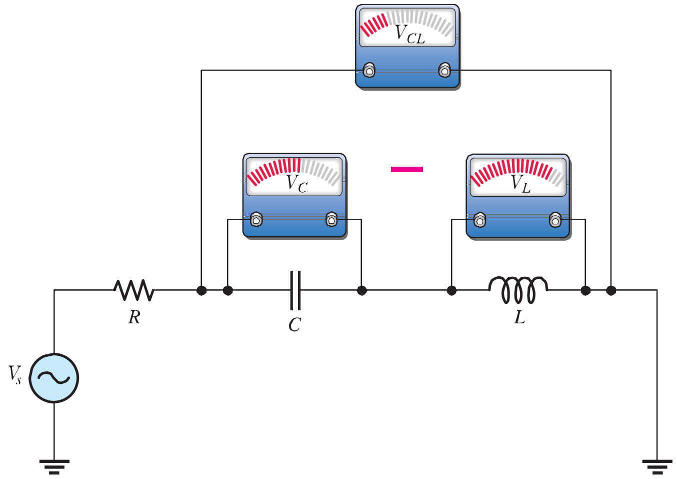

FIGURE 13–5 **The voltage across the series combination of *C* and *L* is always less than the larger individual voltage across either *C* or *L*.**

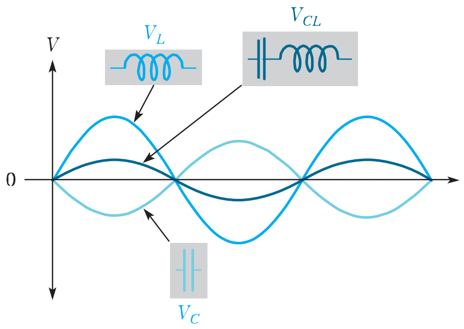

FIGURE 13–6 **Inductor voltage and capacitor voltage effectively subtract because they are out of phase.**

In the next example, Ohm's law is used to find the current and voltages in a series *RLC* circuit.

EXAMPLE 13–3

Find the voltage across each component in Figure 13–7, and draw a complete voltage phasor diagram. Also find the voltage across L and C combined.

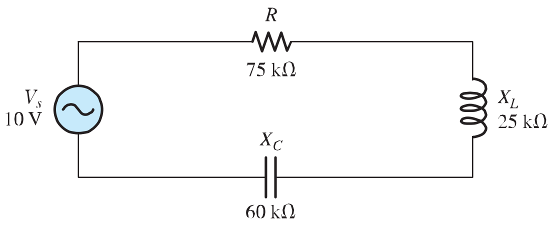

FIGURE 13–7

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12:27:06.5

SOLUTION

First, find the total reactance.

$$X_{tot} = |X_L - X_C| = |25\text{ k}\Omega - 60\text{ k}\Omega| = 35\text{ k}\Omega$$

The total impedance is

$$Z_{tot} = \sqrt{R^2 + X_{tot}^2} = \sqrt{(75\text{ k}\Omega)^2 + (35\text{ k}\Omega)^2} = 82.8\text{ k}\Omega$$

Apply Ohm's law to find the current.

$$I = \frac{V_s}{Z_{tot}} = \frac{10\text{ V}}{82.8\text{ k}\Omega} = 121\ \mu\text{A}$$

Now, apply Ohm's law to find the voltages across R, L, and C.

$$V_R = IR = (121\ \mu\text{A})(75\text{ k}\Omega) = \mathbf{9.08\text{ V}}$$

$$V_L = IX_L = (121\ \mu\text{A})(25\text{ k}\Omega) = \mathbf{3.03\text{ V}}$$

$$V_C = IX_C = (121\ \mu\text{A})(60\text{ k}\Omega) = \mathbf{7.26\text{ V}}$$

The voltage across L and C combined is

$$V_{CL} = V_C - V_L = 7.26\text{ V} - 3.03\text{ V} = \mathbf{4.23\text{ V}}$$

The circuit phase angle is

$$\theta = \tan^{-1}\left(\frac{X_{tot}}{R}\right) = \tan^{-1}\left(\frac{35\text{ k}\Omega}{75\text{ k}\Omega}\right) = 25^\circ$$

Since the circuit is capacitive ($X_C > X_L$), the current leads the source voltage by 25°.

The phasor diagram is shown in Figure 13–8. Notice that V_L is leading V_R by 90°, and V_C is lagging V_R by 90°. Also, there is a 180° phase difference between V_L and V_C. If the current phasor were shown, it would be at the same angle as V_R.

FIGURE 13–8

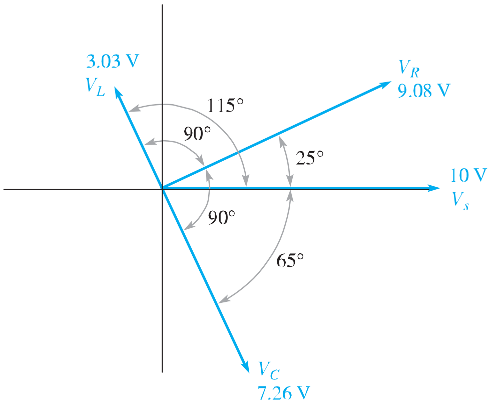

RELATED PROBLEM

What will happen to the current as the frequency of the source voltage in Figure 13–7 is increased?

SECTION 13–2 CHECKUP

1. The following voltages occur in a certain series *RLC* circuit: $V_R = 24$ V, $V_L = 15$ V, $V_C = 45$ V. Determine the source voltage.
2. When $R = 10\text{ k}\Omega$, $X_C = 18\text{ k}\Omega$, and $X_L = 12\text{ k}\Omega$, does the current lead or lag the source voltage? Why?
3. Determine the total reactance in Question 2.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

13–3 SERIES RESONANCE

In a series *RLC* circuit, series resonance occurs when $X_L = X_C$. The frequency at which resonance occurs is called the resonant frequency and is designated f_r.

After completing this section, you should be able to

- **Analyze a circuit for series resonance**
  - **Define *resonant frequency***
  - **Explain why the reactances cancel at resonance**
  - **Determine the series resonant frequency**
  - **Calculate the current and voltages at resonance**
  - **Determine the impedance at resonance**

Figure 13–9 illustrates the series resonant condition. Since the reactances are equal and effectively cancel, the impedance is purely resistive. These resonant conditions are stated as

$$X_L = X_C$$

$$Z_r = R$$

FIGURE 13–9 **At the resonant frequency (f_r), the reactances are equal in magnitude and effectively cancel, leaving $Z_r = R$.**

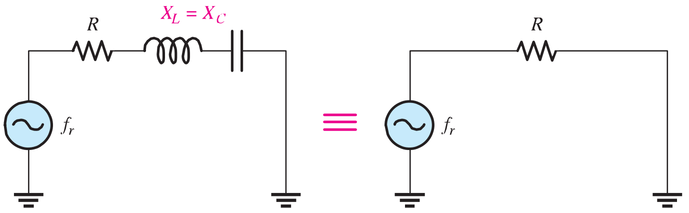

EXAMPLE 13–4

Assume the series *RLC* circuit in Figure 13–10 is at resonance. Determine X_C and Z.

FIGURE 13–10

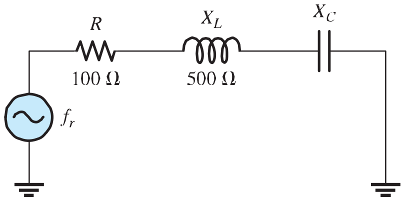

SOLUTION

X_L equals X_C at resonance; therefore,

$$X_C = X_L = \mathbf{500\ \Omega}$$

Since the reactances cancel each other, the impedance equals the resistance.

$$Z_r = R = \mathbf{100\ \Omega}$$

RELATED PROBLEM

Just below the resonant frequency, is the circuit more inductive or more capacitive?

At the series resonant frequency, the voltages across C and L are equal in magnitude because the reactances are equal. The same current is through both since they are in series ($IX_C = IX_L$). Also, V_L and V_C are always 180° out of phase with each other.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

During any given cycle of the input voltage, the polarities of the voltages across C and L are opposite, as shown in parts (a) and (b) of Figure 13–11. The equal and opposite voltages across C and L cancel, leaving zero volts from point A to point B as shown in the figure. Since there is no voltage drop from A to B but there is still current, the total reactance must be zero, as indicated in Figure 13–11(c). Also, the voltage phasor diagram in part (d) shows that V_C and V_L are equal in magnitude and 180° out of phase with each other.

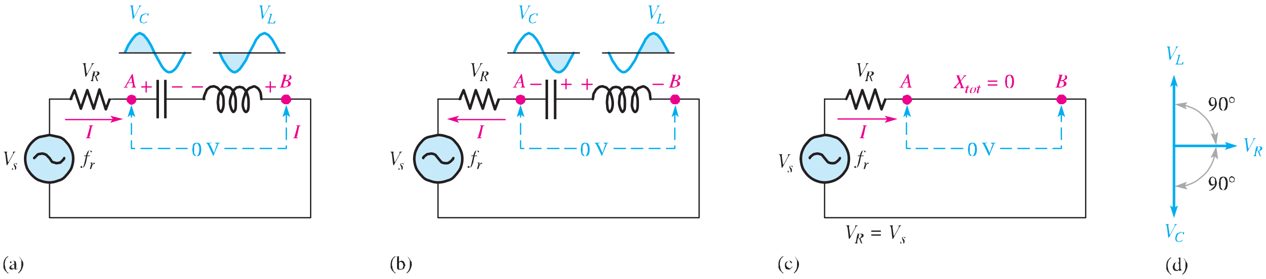

FIGURE 13–11 At the resonant frequency, f_r, the voltages across C and L are equal in magnitude. Since they are 180° out of phase with each other, they cancel, leaving 0 V across the CL combination (point A to point B). The section of the circuit from A to B effectively looks like a short at resonance (neglecting winding resistance).

Series Resonant Frequency

For a given series RLC circuit, resonance occurs at only one specific frequency. A formula for this resonant frequency for the ideal case is developed as follows:

$$X_L = X_C$$

Substitute the reactance formulas, and solve for the resonant frequency (f_r).

$$2\pi f_r L = \frac{1}{2\pi f_r C}$$

$$(2\pi f_r L)(2\pi f_r C) = 4\pi^2 f_r^2 LC = 1$$

$$f_r^2 = \frac{1}{4\pi^2 LC}$$

Take the square root of both sides. The formula for resonant frequency is

$$\boxed{f_r = \frac{1}{2\pi\sqrt{LC}}} \quad (13\text{–}4)$$

EXAMPLE 13–5

Find the series resonant frequency for the circuit in Figure 13–12.

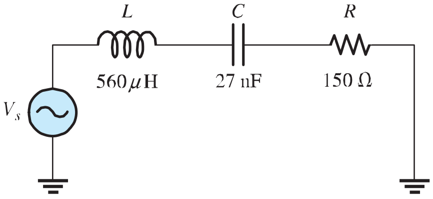

FIGURE 13–12

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

MULTISIM

Open Multisim file E13-05; files are found at www.pearsonhighered.com/floyd. Determine the series resonant frequency by measurement.

SOLUTION

The resonant frequency is

$$f_r = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{(560\ \mu\text{H})(27\ \text{nF})}} = \mathbf{40.9\ kHz}$$

RELATED PROBLEM

If $C = 0.01\ \mu\text{F}$ in Figure 13–12, what is the resonant frequency?

Voltage and Current Amplitudes in a Series *RLC* Circuit

The current and voltage amplitudes in a series *RLC* circuit vary as the frequency is increased from below the resonant frequency, through resonance, and then above resonance. Figure 13–13 illustrates the general response of a circuit in terms of changes in the

FIGURE 13–13 An illustration of how the voltage and current amplitudes respond in a series *RLC* circuit as the frequency is increased from below to above its resonant value. The source voltage is held at a constant amplitude.

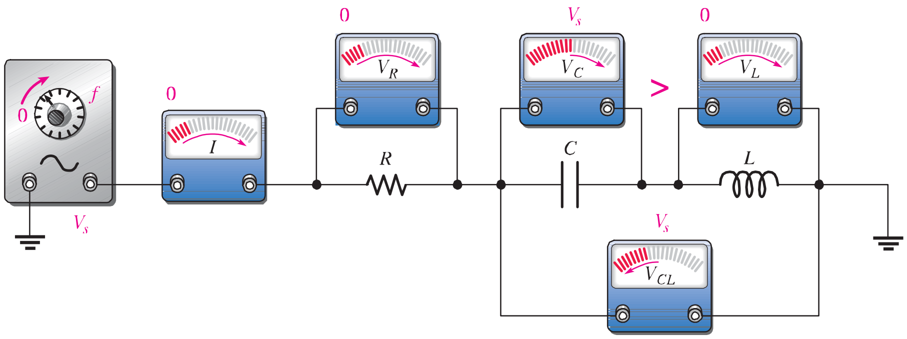

(a) As frequency is increased below resonance from 0: $X_C > X_L$, I increases from 0, V_R increases from 0, V_C increases from V_s, V_L increases from 0, and V_{CL} decreases from V_s.

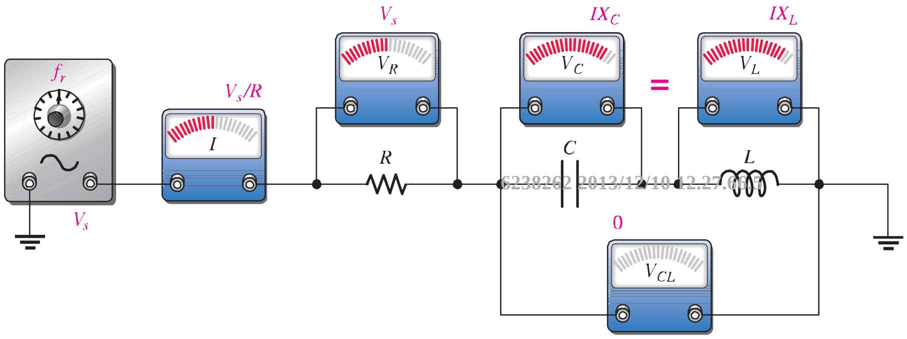

6238262 2013/12/10 12.27.66.5

(b) At the resonant frequency, $X_C = X_L$.

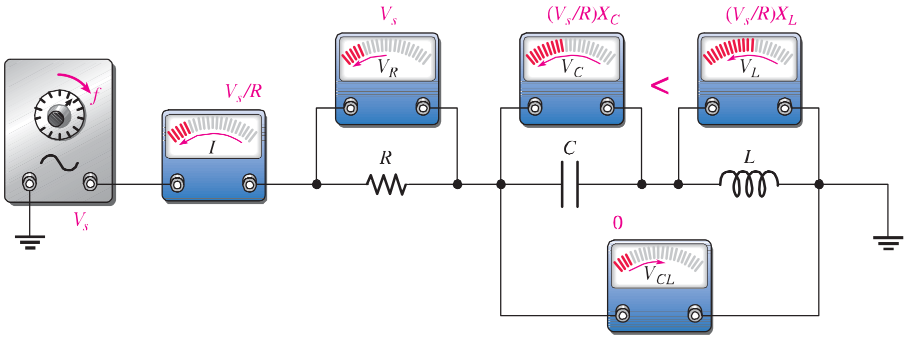

(c) As frequency is increased above resonance: $X_C < X_L$, I decreases from V_s/R, V_R decreases from V_s, V_C decreases from $(V_s/R)X_C$, V_L decreases from $(V_s/R)X_L$, and V_{CL} increases from 0.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

current and in the voltage drops. The Q (quality factor) of the circuit is assumed to be sufficiently high so it has no effect on the response. The Q is the ratio of reactive power to true power and is discussed in Section 13–4.

BELOW THE RESONANT FREQUENCY Figure 13–13(a) indicates the response as the source frequency is increased from zero toward f_r. At $f = 0$ Hz (dc), the capacitor appears open and blocks the current. Thus, there is no voltage across R or L, and the entire source voltage appears across C. The impedance of the circuit is infinitely large at 0 Hz because X_C is infinite (C is open). As the frequency begins to increase, X_C decreases and X_L increases, causing the total reactance, $X_C - X_L$, to decrease. As a result, the impedance decreases and the current increases. As the current increases, V_R also increases, and both V_C and V_L increase. The voltage across C and L combined decreases from its maximum value of V_s because as V_C and V_L approach the same value, their difference becomes less.

AT THE RESONANT FREQUENCY When the frequency reaches its resonant value, f_r, as shown in Figure 13–13(b), V_C and V_L can each be much larger than the source voltage. V_C and V_L cancel, leaving 0 V across the combination of C and L because the voltages are equal in magnitude but opposite in phase. At this point the total impedance is equal to R and is at its minimum value because the total reactance is zero. Thus, the current is at its maximum value, V_s/R, and V_R is at its maximum value, which is equal to the source voltage.

ABOVE THE RESONANT FREQUENCY As the frequency is increased above resonance, as indicated in Figure 13–13(c), X_L continues to increase and X_C continues to decrease, causing the total reactance, $X_L - X_C$, to increase. As a result, there is an increase in impedance and a decrease in current. As the current decreases, V_R also decreases and V_C and V_L decrease. As V_C and V_L decrease, their difference becomes greater, so V_{CL} increases. As the frequency becomes very high, the current approaches zero, both V_R and V_C approach zero, and V_L approaches V_s.

Figure 13–14(a) and (b) illustrate the responses of current and voltage to increasing frequency, respectively. As frequency is increased, the current increases below resonance, reaches a peak at the resonant frequency, and then decreases above resonance. The resistor voltage responds in the same way as the current.

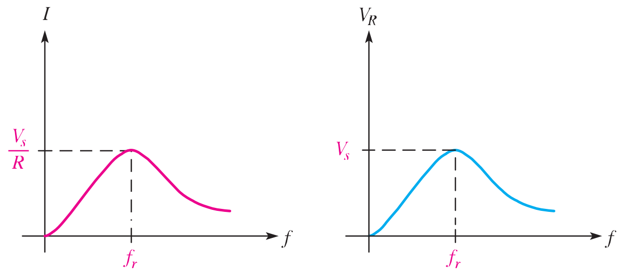

(a) Current (b) Resistor voltage

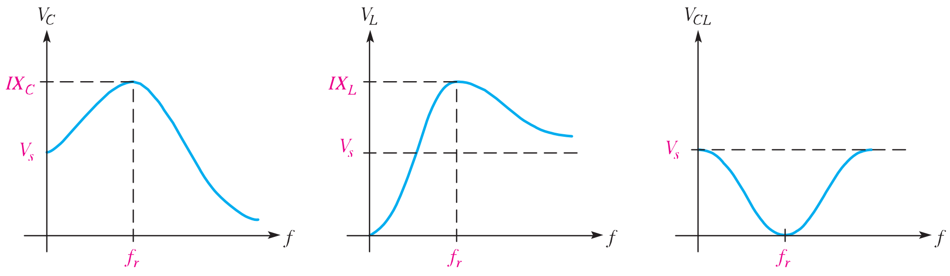

(c) Capacitor voltage (d) Inductor voltage (e) Voltage across C and L combined

FIGURE 13–14 **Generalized current and voltage magnitudes as a function of frequency in a series RLC circuit. V_C and V_L can be much larger than the source voltage. The shapes of the graphs depend on particular circuit values.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Capacitive and Inductive Sensors

Sensors are widely used by many systems to measure certain parameters such as temperature, pressure, or other variable. A common method of indirectly measuring the capacitance or inductance of a sensor is to measure the resonant frequency of an oscillator as the sensor variable changes. One unusual example is to infer the progress of a shock wave as it crushes a cable and changes its capacitance.

eplosione/Fotolia.com

SYSTEM NOTE

The general slopes of the V_C and V_L curves are indicated in Figure 13–14(c) and (d). The voltages are maximum at resonance but drop off above and below f_r. The voltages across L and C at resonance are exactly equal in magnitude but 180° out of phase, so they cancel. Thus, the total voltage across both L and C is zero, and $V_R = V_s$ at resonance. Individually, V_L and V_C can be much greater than the source voltage. Keep in mind that V_L and V_C are always opposite in polarity regardless of the frequency, but only at resonance are their magnitudes equal. The voltage across the C and L combination decreases as the frequency increases below resonance, reaching a minimum of zero at the resonant frequency; then it increases above resonance, as shown in Figure 13–14(e).

The curves in Figure 13–14 are generalized curves, which depend on the particular components in the circuit. The shape will depend greatly on Q. You can plot these curves for any circuit by constructing the circuit in Multisim and using the Bode plotter. The **Bode plotter** in Multisim is a fictitious instrument that plots the frequency response based on where the probes are placed. The response across the resistor in Example 13–5 is shown in Figure 13–15.

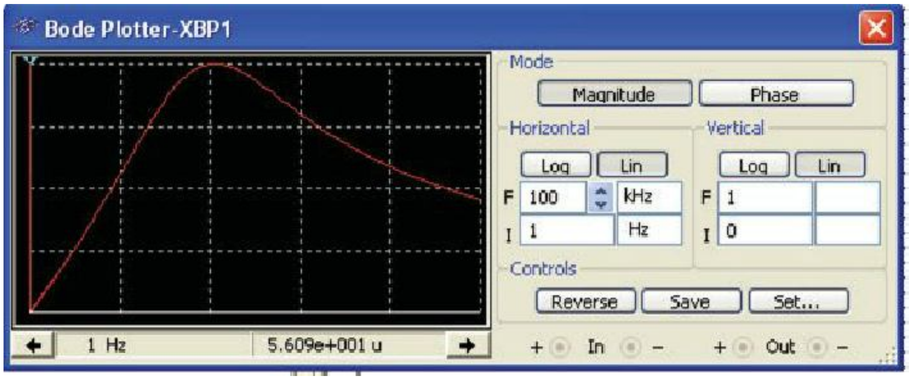

Bode Plotter-XBP1

Horizontal: F 100 kHz, I 1 Hz

Vertical: F 1, I 0

1 Hz 5.609e+001 u

FIGURE 13–15 **Multisim response across the resistor in Example 13–5.**

EXAMPLE 13–6

Find I, V_R, V_L, and V_C at resonance in Figure 13–16.

FIGURE 13–16

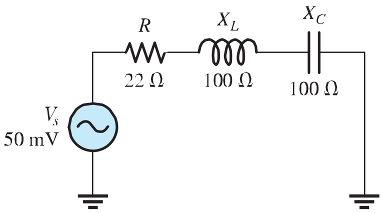

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

At resonance, I is maximum and equal to V_s/R.

$$I = \frac{V_s}{R} = \frac{50\text{ mV}}{22\ \Omega} = \mathbf{2.27\ mA}$$

Apply Ohm's law to obtain the voltages.

$$V_R = IR = (2.27\text{ mA})(22\ \Omega) = \mathbf{50\ mV}$$
$$V_L = IX_L = (2.27\text{ mA})(100\ \Omega) = \mathbf{227\ mV}$$
$$V_C = IX_C = (2.27\text{ mA})(100\ \Omega) = \mathbf{227\ mV}$$

Notice that all of the source voltage is dropped across the resistor. Also, of course, V_L and V_C are equal in magnitude but opposite in phase. This causes the voltages to cancel, making the total reactive voltage zero.

RELATED PROBLEM

What is the current at resonance in Figure 13–16 if $X_L = X_C = 1\text{ k}\Omega$?

The Impedance of a Series *RLC* Circuit

6238262 2013/12/10 12.27.66.5

Figure 13–17 shows a general graph of impedance versus frequency superimposed on the curves for X_C and X_L. At zero frequency, both X_C and Z are infinitely large and X_L is zero because the capacitor looks like an open at 0 Hz and the inductor looks like a short. As the frequency increases, X_C decreases and X_L increases. Since X_C is larger than X_L at frequencies below f_r, Z decreases along with X_C. At f_r, $X_C = X_L$ and $Z = R$. At frequencies above f_r, X_L becomes increasingly larger than X_C, causing Z to increase.

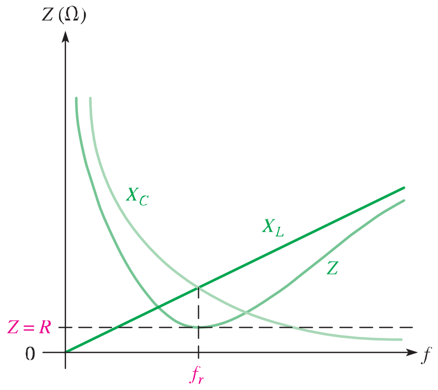

FIGURE 13–17 Series *RLC* impedance as a function of frequency.

EXAMPLE 13–7

For the circuit in Figure 13–18, determine the impedance at the following frequencies:

(a) f_r

(b) 1 kHz below f_r

(c) 1 kHz above f_r

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

FIGURE 13–18

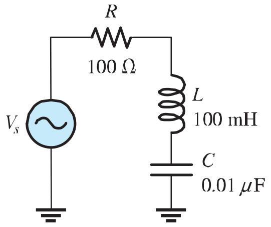

SOLUTION

(a) At f_r, the impedance is equal to R.

$$Z = R = \mathbf{100\ \Omega}$$

To determine the impedance above and below f_r, first calculate the resonant frequency.

$$f_r = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{(100\text{ mH})(0.01\ \mu\text{F})}} = 5.03\text{ kHz}$$

(b) At 1 kHz below f_r, the frequency and reactances are as follows:

$$f = f_r - 1\text{ kHz} = 5.03\text{ kHz} - 1\text{ kHz} = 4.03\text{ kHz}$$

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi(4.03\text{ kHz})(0.01\ \mu\text{F})} = 3.95\text{ k}\Omega$$

$$X_L = 2\pi fL = 2\pi(4.03\text{ kHz})(100\text{ mH}) = 2.53\text{ k}\Omega$$

Therefore, the total reactance and the impedance at $f_r - 1$ kHz are

$$X_{tot} = |X_L - X_C| = |2.53\text{ k}\Omega - 3.95\text{ k}\Omega| = 1.42\text{ k}\Omega$$

$$Z = \sqrt{R^2 + X_{tot}^2} = \sqrt{(100\ \Omega)^2 + (1.42\text{ k}\Omega)^2} = \mathbf{1.42\ k\Omega}$$

(c) At 1 kHz above f_r,

$$f = 5.03\text{ kHz} + 1\text{ kHz} = 6.03\text{ kHz}$$

$$X_C = \frac{1}{2\pi(6.03\text{ kHz})(0.01\ \mu\text{F})} = 2.64\text{ k}\Omega$$

$$X_L = 2\pi(6.03\text{ kHz})(100\text{ mH}) = 3.79\text{ k}\Omega$$

Therefore, the total reactance and the impedance at $f_r + 1$ kHz are

$$X_{tot} = |3.79\text{ k}\Omega - 2.64\text{ k}\Omega| = 1.15\text{ k}\Omega$$

$$Z = \sqrt{(100\ \Omega)^2 + (1.15\text{ k}\Omega)^2} = \mathbf{1.15\ k\Omega}$$

In part (b), Z is capacitive; in part (c), Z is inductive.

RELATED PROBLEM

What happens to the impedance if f is decreased below 4.03 kHz? Above 6.03 kHz?

MULTISIM

Open Multisim file E13-07. Measure the current and the voltage across each element at the resonant frequency, 1 kHz below resonance, and 1 kHz above resonance.

The Phase Angle of a Series *RLC* Circuit

At frequencies below resonance, $X_C > X_L$, and the current leads the source voltage, as indicated in Figure 13–19(a). The phase angle decreases as the frequency approaches the resonant value and is 0° at resonance, as indicated in part (b). At frequencies above resonance, $X_L > X_C$, and the current lags the source voltage, as indicated in part (c). As the frequency goes higher, the phase angle approaches 90°. A plot of phase angle versus frequency is shown in Figure 13–19(d).

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

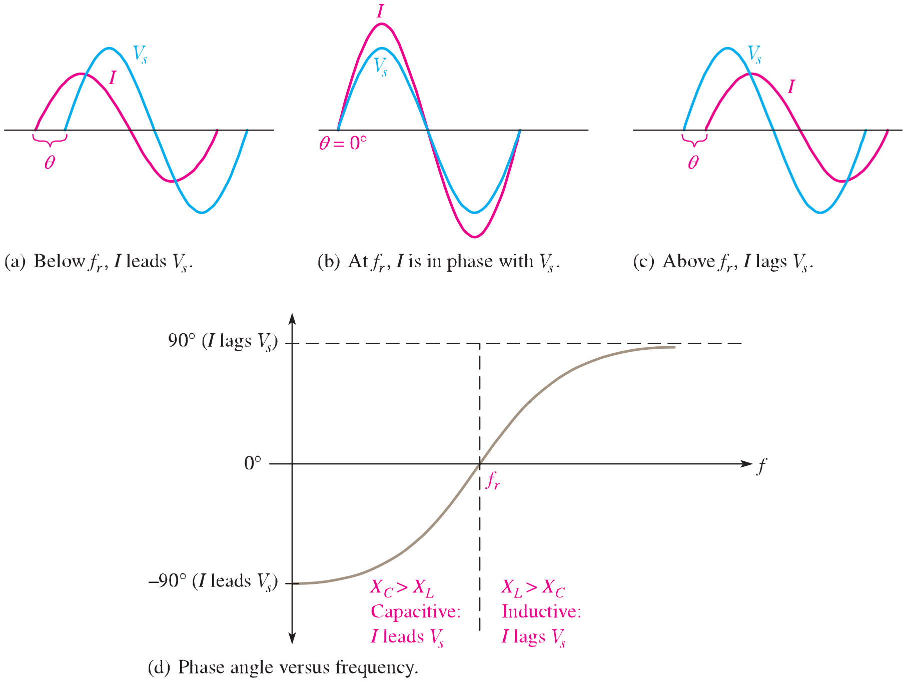

(a) Below f_r, I leads V_s.

(b) At f_r, I is in phase with V_s.

(c) Above f_r, I lags V_s.

(d) Phase angle versus frequency.

FIGURE 13–19 **The phase angle as a function of frequency in a series *RLC* circuit.**

SECTION 13–3 CHECKUP

1. What is the condition for series resonance?
2. Why is the current maximum at the resonant frequency?
3. Calculate the resonant frequency for $C = 1000$ pF and $L = 1000\ \mu$H.
4. In Question 3, is the circuit inductive, capacitive, or resistive at 50 kHz?

13–4 SERIES RESONANT FILTERS

A common use of series *RLC* circuits is in filter applications. In this section, you will learn the basic configurations for passive band-pass and band-stop filters and several important filter characteristics.

After completing this section, you should be able to

- **Analyze series resonant filters**
  - **Identify a basic series resonant band-pass filter**
  - **Define and determine *bandwidth***
  - **Define the *half-power frequency***
  - **Discuss dB measurement**
  - **Define *selectivity***
  - **Discuss filter quality factor (*Q*)**
  - **Identify a series resonant band-stop filter**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The Band-Pass Filter

A basic series resonant band-pass filter is shown in Figure 13–20. Notice that the series *LC* portion is placed between the input and the output and that the output is taken across the resistor.

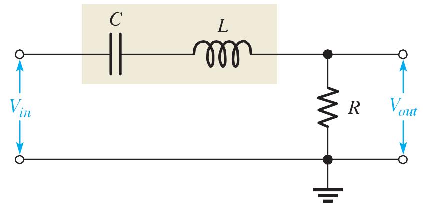

FIGURE 13–20 **A basic series resonant band-pass filter.**

A **band-pass filter** allows signals at the resonant frequency and at frequencies within a certain band (or range) extending below and above the resonant value to pass from input to output without a significant reduction in amplitude. Signals at frequencies lying outside this specified band (called the **passband**) are reduced in amplitude to below a certain level and are considered to be rejected by the filter.

The filtering action is the result of the impedance characteristic of the filter. As you learned in Section 13–3, the impedance is minimum at resonance and has increasingly higher values below and above the resonant frequency. At very low frequencies, the impedance is very high and tends to block the current. As the frequency increases, the impedance drops, allowing more current and thus more voltage across the output resistor. At the resonant frequency, the impedance is very low and equal to the total resistance of the circuit (winding resistance plus *R*). At this point there is maximum current and the resulting output voltage is maximum. As the frequency goes above resonance, the impedance again increases, causing the current and the resulting output voltage to drop. Figure 13–21 illustrates the general frequency response of a series resonant band-pass filter.

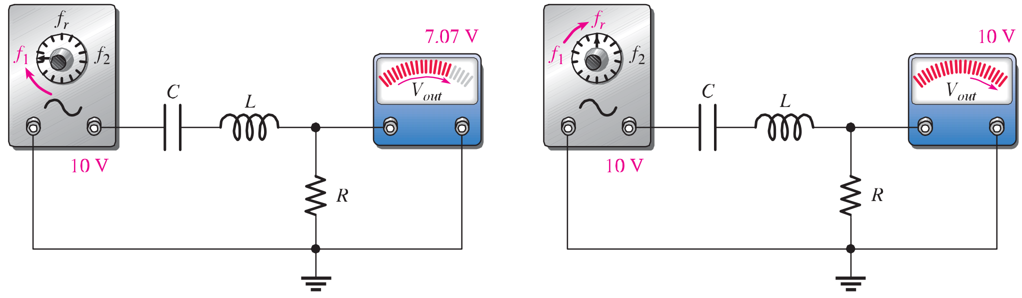

(a) As the frequency increases to f_1, V_{out} increases to 7.07 V.

(b) As the frequency increases from f_1 to f_r, V_{out} increases from 7.07 V to 10 V.

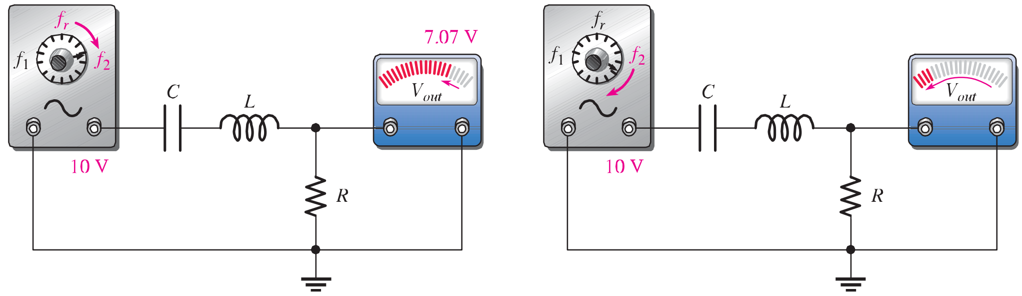

(c) As the frequency increases from f_r to f_2, V_{out} decreases from 10 V to 7.07 V.

(d) As the frequency increases above f_2, V_{out} decreases below 7.07 V.

FIGURE 13–21 **Example of the frequency response of a series resonant band-pass filter with the input voltage at a constant 10 V rms. The winding resistance of the coil is neglected.**

6238262 2013/12/10 12.27.66.5

Bandwidth of the Passband

The bandwidth (*BW*) of a band-pass filter is the range of frequencies for which the current (or output voltage) is equal to or greater than 70.7% of its value at the resonant frequency. Figure 13–22 shows the bandwidth on the response curve for a band-pass filter.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The frequencies at which the output of a filter is 70.7% of its maximum are the **cutoff frequencies**. Notice in Figure 13–22 that frequency f_1, which is below f_r, is the frequency at which I (or V_{out}) is 70.7% of the resonant value (I_{max}); f_1 is commonly called the *lower cutoff frequency*. At frequency f_2, above f_r, the current (or V_{out}) is again 70.7% of its maximum; f_2 is called the *upper cutoff frequency*. Other names for f_1 and f_2 are *−3 dB frequencies, critical frequencies, band frequencies,* and *half-power frequencies*. (The term *decibel,* abbreviated dB, is defined later in this section.)

The formula for calculating the bandwidth is

$$BW = f_2 - f_1 \tag{13–5}$$

The unit of bandwidth is the hertz (Hz), the same as for frequency.

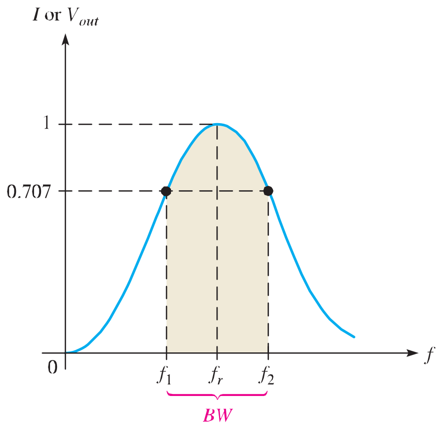

FIGURE 13–22 Generalized response curve of a series resonant band-pass filter.

EXAMPLE 13–8

A certain series resonant band-pass filter has a maximum current of 100 mA at the resonant frequency. What is the value of the current at the cutoff frequencies?

SOLUTION

Current at the cutoff frequencies is 70.7% of maximum.

$$I_{f1} = I_{f2} = 0.707I_{max} = 0.707(100\text{ mA}) = \mathbf{70.7\text{ mA}}$$

RELATED PROBLEM

Will a change in the cutoff frequencies affect the current at the new cutoff frequencies if the maximum current remains 100 mA?

EXAMPLE 13–9

A resonant circuit has a lower cutoff frequency of 8 kHz and an upper cutoff frequency of 12 kHz. Determine the bandwidth.

SOLUTION

$$BW = f_2 - f_1 = 12\text{ kHz} - 8\text{ kHz} = \mathbf{4\text{ kHz}}$$

RELATED PROBLEM

What is the bandwidth when $f_1 = 1.0\text{ MHz}$ and $f_2 = 1.2\text{ MHz}$?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Half-Power Points of the Filter Response

As previously mentioned, the upper and lower cutoff frequencies are sometimes called the **half-power frequencies**. This term is derived from the fact that the true power from the source at these frequencies is one-half the power delivered at the resonant frequency. The following steps show that this relationship is true for a series resonant circuit.

At resonance,

$$P_{max} = I_{max}^2 R$$

The power at f_1 (or f_2) is

$$P_{f1} = I_{f1}^2 R = (0.707 I_{max})^2 R = (0.707)^2 I_{max}^2 R = 0.5 I_{max}^2 R = 0.5 P_{max}$$

Decibel (dB) Measurement

Another common term for the upper and lower cutoff frequencies is *−3 dB frequencies*. The **decibel (dB)** is ten times the logarithmic ratio of two powers, which can be used to express the input-to-output relationship of a filter. The following equation expresses a power ratio in decibels:

$$\mathbf{dB = 10\log\left(\frac{P_{out}}{P_{in}}\right)} \quad (13\text{–}6)$$

The decibel formula using a voltage ratio is based on the previous formula for the power ratio with voltages measured across equal resistances.

$$\mathbf{dB = 20\log\left(\frac{V_{out}}{V_{in}}\right)} \quad (13\text{–}7)$$

EXAMPLE 13–10

At a certain frequency, the output power of a filter is 5 W and the input power is 10 W. Express the power ratio in decibels.

SOLUTION

$$10\log\left(\frac{P_{out}}{P_{in}}\right) = 10\log\left(\frac{5\text{ W}}{10\text{ W}}\right) = 10\log(0.5) = \mathbf{-3.01\ dB}$$

RELATED PROBLEM

Express in decibels the ratio $V_{out}/V_{in} = 0.2$.

THE −3 dB FREQUENCIES The output of a filter is said to be down 3 dB at the cutoff frequencies. As you know, this frequency is the point at which the output voltage is 70.7% of the maximum voltage at resonance. We can show that the 70.7% point is the same as 3 dB below maximum (or −3 dB) as follows. The maximum voltage is the 0 dB reference.

$$20\log\left(\frac{0.707V_{max}}{V_{max}}\right) = 20\log(0.707) = -3\text{ dB}$$

Selectivity of a Band-Pass Filter

The response curve in Figure 13–22 is also called a *selectivity curve*. **Selectivity** defines how well a resonant circuit responds to a certain frequency and discriminates against all others. *The narrower the bandwidth, the greater the selectivity.*

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

In the ideal response, we assume that a resonant circuit accepts frequencies within its bandwidth and completely eliminates frequencies outside the bandwidth. Such is not actually the case, however, because signals with frequencies outside the bandwidth are not completely eliminated. Their magnitudes, however, are greatly reduced. The further the frequencies are from the cutoff frequencies, the greater the reduction, as illustrated in Figure 13–23(a). An ideal selectivity curve is shown in Figure 13–23(b).

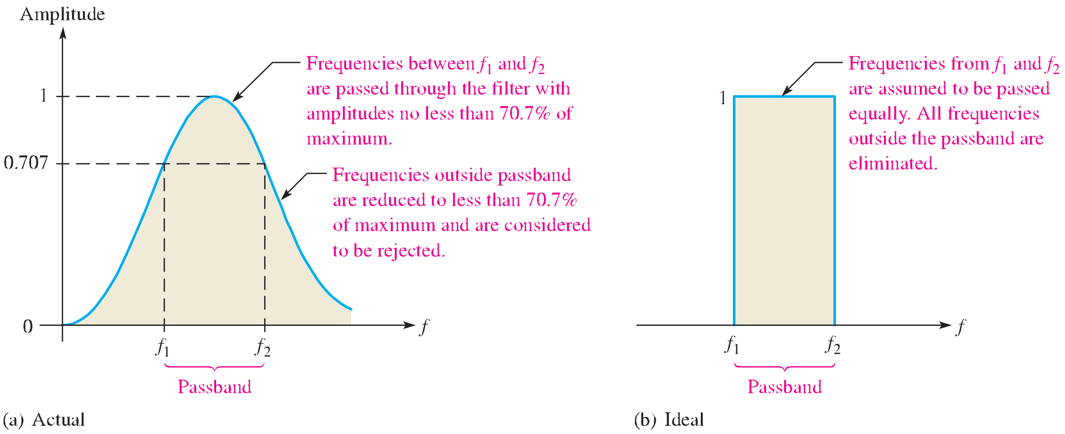

(a) Actual (b) Ideal

FIGURE 13–23 Generalized selectivity curve of a band-pass filter.

The factor that determines the selectivity of a practical band-pass filter is the steepness of the response curve, as illustrated in Figure 13–24. A more selective filter attenuates (reduces) frequencies outside the passband more quickly than a less selective filter. In communication systems, a highly selective filter will differentiate between a desired signal (such as from a radio station) and a nearby one.

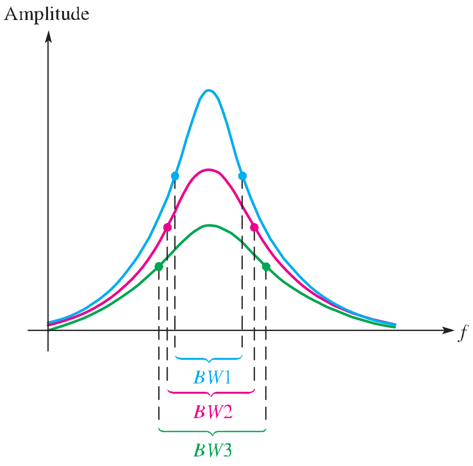

FIGURE 13–24 Comparative selectivity curves. The blue curve represents the greatest selectivity.

The Quality Factor (Q) of a Resonant Circuit

Recall that the quality factor (Q) for a coil was defined in Section 11–5 as the ratio of the reactive power (at a specified frequency) to the true power, which is dissipated in the winding resistance of the coil (L). In a series resonant circuit, the Q of the circuit includes any other resistance in series with the coil, so the circuit Q may be lower than the Q of the coil alone because R includes more than just the coil resistance. The Q of a series resonant

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

circuit is the ratio of the reactive power in L to the true power in R. The quality factor is important in resonant circuits. A formula for Q is developed as follows:

$$Q = \frac{\text{reactive power}}{\text{true power}} = \frac{I^2X_L}{I^2R}$$

The I^2 terms cancel, leaving

$$\boldsymbol{Q = \frac{X_L}{R}} \qquad \textbf{(13–8)}$$

Since Q varies with frequency because X_L varies, we are interested mainly in Q at resonance. Note that Q is a ratio of like units (ohms) which cancel and, therefore, Q has no unit itself. The quality factor is also known as *unloaded* Q because it is defined with no load across the coil.

EXAMPLE 13–11

Determine Q at resonance for the circuit in Figure 13–25.

FIGURE 13–25

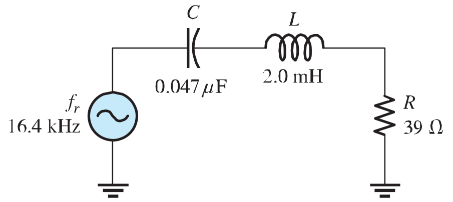

SOLUTION

Determine the inductive reactance.

$$X_L = 2\pi f_r L = 2\pi(16.4\text{ kHz})(2.0\text{ mH}) = 206\ \Omega$$

The quality factor is

$$Q = \frac{X_L}{R} = \frac{206\ \Omega}{39\ \Omega} = \mathbf{5.3}$$

RELATED PROBLEM

Calculate Q at resonance if C in Figure 13–25 is halved. The resonant frequency will increase.

HOW *Q* AFFECTS BANDWIDTH A higher value of circuit Q results in a smaller bandwidth. A lower value of Q causes a larger bandwidth. A formula for the bandwidth of a resonant circuit in terms of Q is

$$\boldsymbol{BW = \frac{f_r}{Q}} \qquad \textbf{(13–9)}$$

EXAMPLE 13–12

What is the bandwidth of the filter in Figure 13–26?

FIGURE 13–26

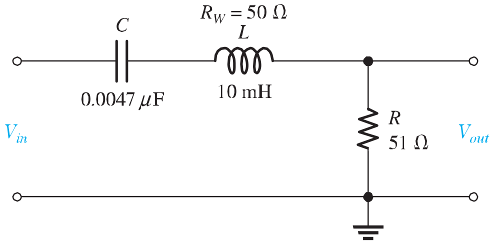

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

The total resistance is

$$R_{tot} = R + R_W = 51\ \Omega + 50\ \Omega = 101\ \Omega$$

Determine the bandwidth as follows:

$$f_r = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{(10\text{ mH})(0.0047\ \mu\text{F})}} = 23.2\text{ kHz}$$

$$X_L = 2\pi f_r L = 2\pi(23.2\text{ kHz})(10\text{ mH}) = 1.46\text{ k}\Omega$$

$$Q = \frac{X_L}{R_{tot}} = \frac{1.46\text{ k}\Omega}{101\ \Omega} = 14.5$$

$$BW = \frac{f_r}{Q} = \frac{23.2\text{ kHz}}{14.5} = \mathbf{1.60\text{ kHz}}$$

RELATED PROBLEM

Determine the bandwidth if L is changed to 50 mH with the some winding resistance in Figure 13–26.

MULTISIM

Open Multisim file E13-12. Determine the bandwidth by measurement. How closely does this result compare to the calculated value?

The Band-Stop Filter

The basic series resonant band-stop filter is shown in Figure 13–27. Notice that the output voltage is taken across the *LC* portion of the circuit. This filter is still a series *RLC* circuit, just as the band-pass filter is. The difference is that in this case, the output voltage is taken across the combination of *L* and *C* rather than across *R*.

The **band-stop filter** rejects signals with frequencies between the lower and upper cutoff frequencies and passes those signals with frequencies below and above the cutoff values, as shown in the response curve of Figure 13–28. The range of frequencies between the lower and upper cutoff points is called the **stopband**. This type of filter is also referred to as a *band-elimination filter, band-reject filter,* or a *notch filter.*

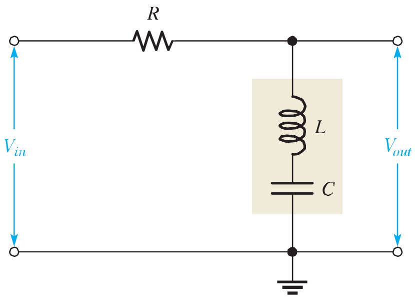

FIGURE 13–27 **A basic series resonant band-stop filter.**

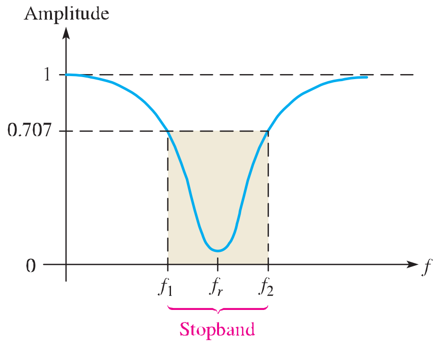

FIGURE 13–28 **Generalized response curve for a band-stop filter.**

All the characteristics that have been discussed in relation to the band-pass filter apply equally to the band-stop filter, with the exception that the response curve of the output voltage is opposite. For the band-pass filter, V_{out} is maximum at resonance. For the band-stop filter, V_{out} is minimum at resonance.

At very low frequencies, the *LC* combination appears as a near open due to the high X_C, thus allowing most of the input voltage to pass through to the output. As the frequency increases, the impedance of the *LC* combination decreases until, at resonance, it is zero (ideally). Thus, the input signal is shorted to ground, and there is very little output voltage. As the frequency goes above its resonant value, the *LC* impedance increases, allowing an increasing amount of voltage to be dropped across it. The general frequency response of a series resonant band-stop filter is illustrated in Figure 13–29.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

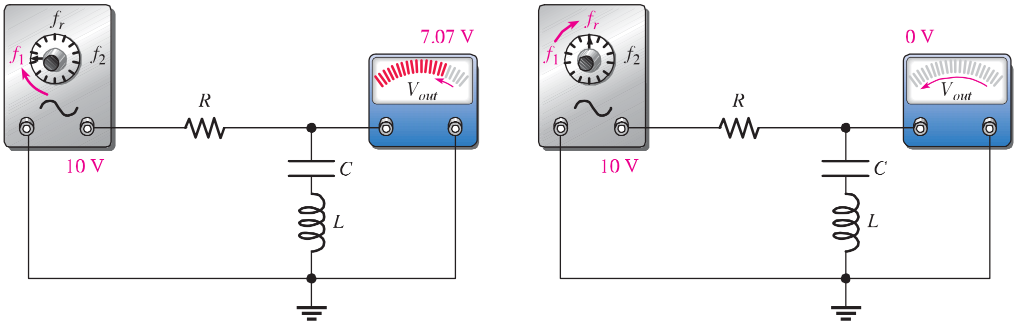
f_r
f_1
f_2
7.07 V
V_{out}
R
10 V
C
L
f_r
f_1
f_2
0 V
V_{out}
R
10 V
C
L

(a) As frequency increases to f_1, V_{out} decreases from 10 V to 7.07 V.

(b) As frequency increases from f_1 to f_r, V_{out} decreases from 7.07 V to 0 V.

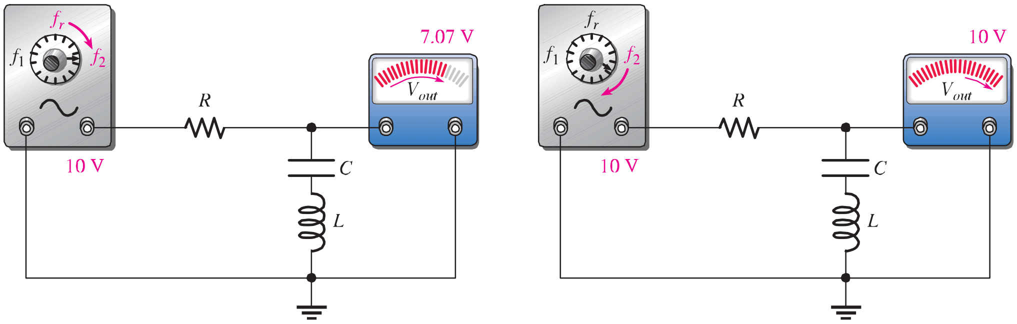
f_r
f_1
f_2
7.07 V
V_{out}
R
10 V
C
L
f_r
f_1
f_2
10 V
V_{out}
R
10 V
C
L

(c) As frequency increases from f_r to f_2, V_{out} increases from 0 V to 7.07 V.

(d) As frequency increases above f_2, V_{out} increases toward 10 V.

FIGURE 13–29 **Example of the frequency response of a series resonant band-stop filter with V_{in} at a constant 10 V rms. The winding resistance is neglected.**

EXAMPLE 13–13

Find the output voltage at f_r and the bandwidth in Figure 13–30.

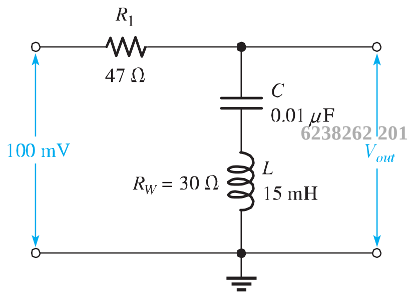
R_1
47 Ω
C
0.01 μF
100 mV
$R_W = 30\ \Omega$
L
15 mH
V_{out}

6238262 2013/12/10 12.27.66.5

FIGURE 13–30

SOLUTION

Since $X_L = X_C$ at resonance, we can apply the voltage-divider formula to find V_{out}.

$$V_{out} = \left(\frac{R_W}{R_1 + R_W}\right)V_{in} = \left(\frac{30\ \Omega}{77\ \Omega}\right)100\text{ mV} = \mathbf{39.0\ mV}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Determine the bandwidth as follows:

$$f_r = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{(15\text{ mH})(0.01\ \mu\text{F})}} = 13.0\text{ kHz}$$

$$X_L = 2\pi f_r L = 2\pi(13.0\text{ kHz})(15\text{ mH}) = 1.22\text{ k}\Omega$$

$$Q = \frac{X_L}{R} = \frac{X_L}{R_1 + R_W} = \frac{1.22\text{ k}\Omega}{77\ \Omega} = 15.9$$

$$BW = \frac{f_r}{Q} = \frac{1.22\text{ kHz}}{15.9} = \mathbf{810\text{ Hz}}$$

RELATED PROBLEM

What happens to V_{out} if the frequency is increased above resonance? Below resonance?

MULTISIM

Open Multisim file E13-13. Verify the calculated value of the resonant frequency and then measure the output voltage at resonance. Determine the bandwidth by measurement.

SECTION 13–4 CHECKUP

1. The output voltage of a certain band-pass filter is 15 V at the resonant frequency. What is its value at the cutoff frequencies?
2. For a certain band-pass filter, $f_r = 120$ kHz and $Q = 12$. What is the bandwidth of the filter?
3. In a band-stop filter, is the current minimum or maximum at resonance? Is the output voltage minimum or maximum at resonance?

13–5 PARALLEL *RLC* CIRCUITS

In this section, you will learn how to determine the impedance and phase angle of a parallel *RLC* circuit. Current relationships and series-parallel to parallel conversions are discussed.

After completing this section, you should be able to

- **Analyze parallel *RLC* circuits**
  - **Calculate the impedance**
  - **Calculate the phase angle**
  - **Determine all currents**
  - **Convert from a series-parallel *RLC* circuit to an equivalent parallel form**

Impedance and Phase Angle

The circuit in Figure 13–31 consists of the parallel combination of *R*, *L*, and *C*. To find the admittance, add the conductance (G) and the total susceptance (B_{tot}) as phasor quantities. B_{tot} is the absolute value of the difference of the inductive susceptance and the capacitive susceptance.

$$B_{tot} = |B_L - B_C|$$

Therefore, the formula for admittance is

$$\boldsymbol{Y = \sqrt{G^2 + B_{tot}^2}} \qquad (13\text{–}10)$$

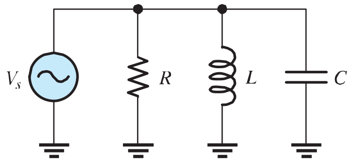

FIGURE 13–31 Parallel *RLC* circuit.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The total impedance is the reciprocal of the admittance.

$$Z_{tot} = \frac{1}{Y}$$

The phase angle of the circuit is given by the following formula:

$$\boldsymbol{\theta = \tan^{-1}\left(\frac{B_{tot}}{G}\right)} \tag{13–11}$$

When the frequency is above its resonant value ($X_C < X_L$), the impedance of the circuit in Figure 13–31 is predominantly capacitive because the capacitive current is greater, and the total current leads the source voltage. When the frequency is below its resonant value ($X_L < X_C$), the impedance of the circuit is predominantly inductive, and the total current lags the source voltage.

EXAMPLE 13–14

Determine the total impedance and the phase angle for the parallel *RLC* circuit in Figure 13–32.

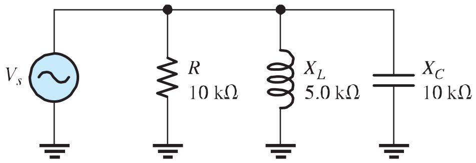

FIGURE 13–32

SOLUTION

First, determine the admittance as follows:

$$G = \frac{1}{R} = \frac{1}{10\ \text{k}\Omega} = 100\ \mu\text{S}$$

$$B_C = \frac{1}{X_C} = \frac{1}{10\ \text{k}\Omega} = 100\ \mu\text{S}$$

$$B_L = \frac{1}{X_L} = \frac{1}{5.0\ \text{k}\Omega} = 200\ \mu\text{S}$$

$$B_{tot} = |B_L - B_C| = 100\ \mu\text{S}$$

$$Y = \sqrt{G^2 + B_{tot}^2} = \sqrt{(100\ \mu\text{S})^2 + (100\ \mu\text{S})^2} = 141.4\ \mu\text{S}$$

From Y, you can get Z_{tot}.

$$Z_{tot} = \frac{1}{Y} = \frac{1}{141.4\ \mu\text{S}} = \mathbf{7.07\ k\Omega}$$

The phase angle is

$$\theta = \tan^{-1}\left(\frac{B_{tot}}{G}\right) = \tan^{-1}\left(\frac{100\ \mu\text{S}}{100\ \mu\text{S}}\right) = \mathbf{45°}$$

RELATED PROBLEM

Is the impedance of the circuit in Figure 13–32 predominantly inductive or predominantly capacitive?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Current Relationships

In a parallel *RLC* circuit, the current in the capacitive branch and the current in the inductive branch are always 180° out of phase with each other (neglecting any coil resistance). For this reason, I_C and I_L subtract from each other, and thus the total current into the parallel branches of L and C is always less than the largest individual branch current, as illustrated in Figure 13–33 and in the waveform diagram of Figure 13–34. Of course, the current in the resistive branch is always 90° out of phase with both reactive currents, as shown in the current phasor diagram of Figure 13–35. Notice that I_C is plotted on the positive y-axis and that I_L is plotted on the negative y-axis. The total current can be expressed as

$$I_{tot} = \sqrt{I_R^2 + I_{CL}^2} \tag{13–12}$$

Coaxial Cable

Coaxial cable (or "coax" for short) is shielded cable used for transmitting high frequencies and fast-pulsed waveforms. The shielding helps keep external signals from affecting the signal in the cable and forms part of the electrical circuit. In addition to the shielding, coax is characterized by its impedance, attenuation, and capacitance.

Albert Lozano-Nieto/Fotolia

SYSTEM NOTE

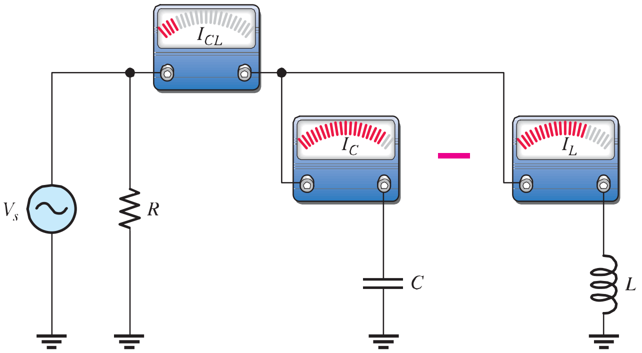

FIGURE 13–33 The total current into the parallel combination of C and L is the absolute value of the difference of the two branch currents ($I_{CL} = |I_C - I_L|$).

6238262 2013/12/10 12.27.66.5

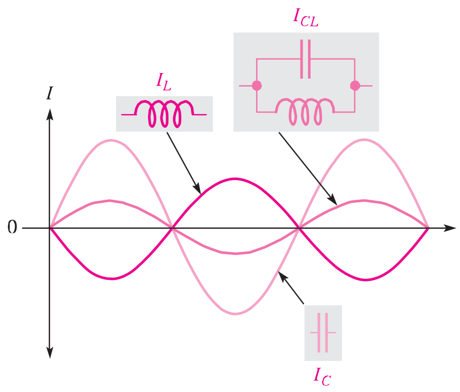

FIGURE 13–34 I_C and I_L effectively subtract from each other.

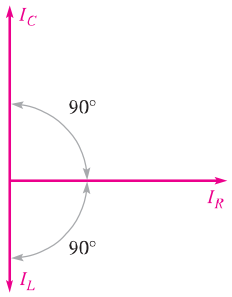

FIGURE 13–35 Current phasor diagram for a parallel *RLC* circuit.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

where I_{CL} is the absolute value of the difference between the two currents, $|I_C - I_L|$, and is the total current into the L and C branches.

The phase angle can also be expressed in terms of the branch currents as

$$\boldsymbol{\theta = \tan^{-1}\left(\frac{I_{CL}}{I_R}\right)} \tag{13–13}$$

EXAMPLE 13–15

Find each branch current and the total current in Figure 13–36. Draw a diagram of their relationship.

FIGURE 13–36

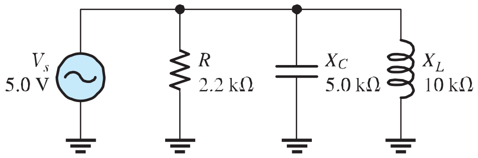

SOLUTION

Apply Ohm's law to find each branch current.

$$I_R = \frac{V_s}{R} = \frac{5.0\text{ V}}{2.2\text{ k}\Omega} = \mathbf{2.27\text{ mA}}$$

$$I_C = \frac{V_s}{X_C} = \frac{5.0\text{ V}}{5.0\text{ k}\Omega} = \mathbf{1.0\text{ mA}}$$

$$I_L = \frac{V_s}{X_L} = \frac{5.0\text{ V}}{10\text{ k}\Omega} = \mathbf{0.5\text{ mA}}$$

The total current is the phasor sum of the branch currents.

$$I_{CL} = |I_C - I_L| = 0.5\text{ mA}$$

$$I_{tot} = \sqrt{I_R^2 + I_{CL}^2} = \sqrt{(2.27\text{ mA})^2 + (0.5\text{ mA})^2} = \mathbf{2.32\text{ mA}}$$

The phase angle is

$$\theta = \tan^{-1}\left(\frac{I_{CL}}{I_R}\right) = \tan^{-1}\left(\frac{0.5\text{ mA}}{2.27\text{ mA}}\right) = 12.4^\circ$$

The total current is 2.32 mA, leading V_s by 12.4°. Figure 13–37 is the current phasor diagram for the circuit.

FIGURE 13–37

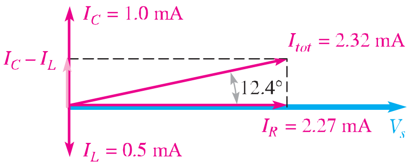

RELATED PROBLEM

Will the total current increase or decrease if the frequency in Figure 13–36 is increased? Why?

Conversion of Series-Parallel to Parallel

The particular series-parallel configuration shown in Figure 13–38 is important because it represents a circuit having parallel L and C branches, with the winding resistance of the coil taken into account as a series resistance in the L branch.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The series-parallel circuit in Figure 13–38 can be viewed in an equivalent purely parallel form, as indicated in Figure 13–39. This equivalent form will simplify the analysis of parallel resonant characteristics that will be discussed in Section 13–6.

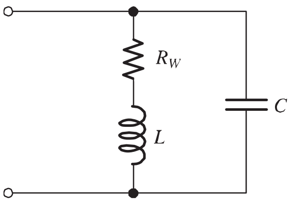

FIGURE 13–38 **A series-parallel *RLC* circuit ($Q = X_L/R_W$).**

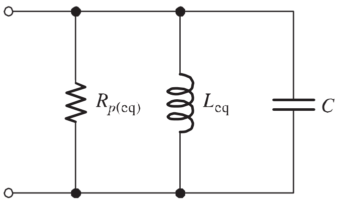

FIGURE 13–39 **Parallel equivalent form of the circuit in Figure 13–38.**

The equivalent inductance, L_{eq}, and the equivalent parallel resistance, $R_{p(\text{eq})}$, are given by the following formulas:

$$L_{\text{eq}} = L\left(\frac{Q^2 + 1}{Q^2}\right) \qquad \textbf{(13–14)}$$

$$R_{p(\text{eq})} = R_W(Q^2 + 1) \qquad \textbf{(13–15)}$$

where Q is the quality factor of the coil, X_L/R_W. Derivations of these formulas are quite involved and thus are not given here. Notice in the equations that for a $Q \geq 10$, the value of L_{eq} is approximately the same as the original value of L. For example, if $L = 10$ mH and $Q = 10$, then

$$L_{\text{eq}} = 10\text{ mH}\left(\frac{10^2 + 1}{10^2}\right) = 10\text{ mH}(1.01) = 10.1\text{ mH}$$

The equivalency of the two circuits means that at a given frequency, when the same value of voltage is applied to both circuits, the same total current is in both circuits and the phase angles are the same. Basically, an equivalent circuit simply makes circuit analysis more convenient.

EXAMPLE 13–16

Convert the series-parallel circuit in Figure 13–40 to an equivalent parallel form at the given frequency.

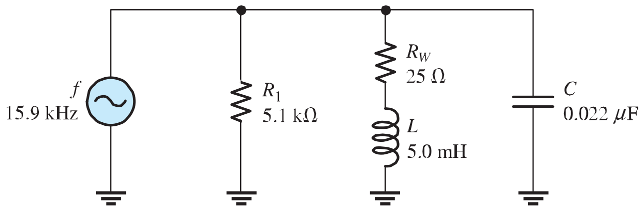

FIGURE 13–40

SOLUTION

Determine the inductive reactance.

$$X_L = 2\pi fL = 2\pi(15.9\text{ kHz})(5.0\text{ mH}) = 500\ \Omega$$

The Q of the coil is

$$Q = \frac{X_L}{R_W} = \frac{500\ \Omega}{25\ \Omega} = 20$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

Since $Q > 10$, then $L_{eq} \cong L = 5.0\text{ mH}$.

The equivalent parallel resistence is

$$R_{p(\text{eq})} = R_W(Q^2 + 1) = (25\ \Omega)(20^2 + 1) = 10.0\text{ k}\Omega$$

This equivalent resistance ($R_{p(\text{eq})}$) appears in parallel with R_1 as shown in Figure 13–41(a). When combined, they give a total parallel resistance ($R_{p(tot)}$) of 3.38 kΩ, as indicated in Figure 13–41(b).

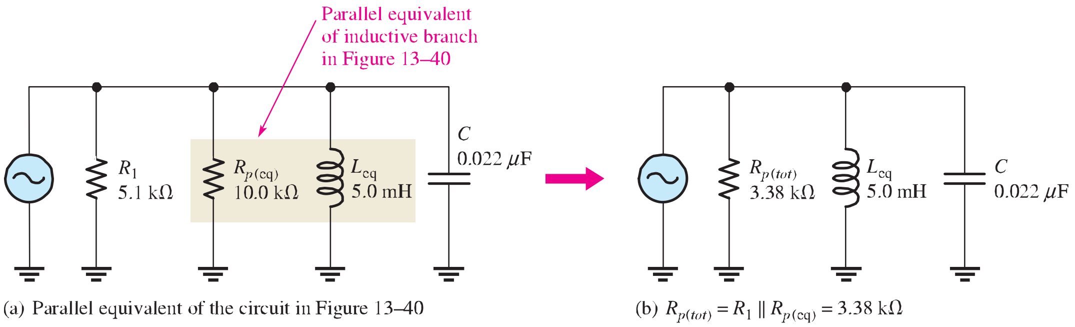

(a) Parallel equivalent of the circuit in Figure 13–40

(b) $R_{p(tot)} = R_1 \parallel R_{p(\text{eq})} = 3.38\text{ k}\Omega$

FIGURE 13–41

RELATED PROBLEM

Find the equivalent parallel circuit if $R_W = 10\ \Omega$ in Figure 13–40.

SECTION 13–5 CHECKUP

1. In a three-branch parallel circuit, $R = 1500\ \Omega$, $X_C = 1000\ \Omega$, and $X_L = 500\ \Omega$ at a certain frequency. Determine the current in each branch when $V_s = 12$ V.
2. Is the circuit in Question 1 capacitive or inductive? Why?
3. Find the equivalent parallel inductance and resistance for a 20 mH coil with a winding resistance of 10 Ω at a frequency of 1 kHz.

13–6 PARALLEL RESONANCE

In this section, we will first look at the resonant condition in an ideal parallel *LC* circuit. Then, we will examine the more realistic case where the resistance of the coil is taken into account.

After completing this section, you should be able to

- **Analyze a circuit for parallel resonance**
  - **Describe parallel resonance in an ideal circuit**
  - **Describe parallel resonance in a nonideal circuit**
  - **Explain how impedance varies with frequency**
  - **Determine current and phase angle at resonance**
  - **Determine parallel resonant frequency**
  - **Discuss the effects of loading a parallel resonant circuit**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Condition for Ideal Parallel Resonance

Ideally, **parallel resonance** occurs when $X_L = X_C$. The frequency at which resonance occurs is called the *resonant frequency,* just as in the series resonant circuit. When $X_L = X_C$, the two branch currents, I_C and I_L, are equal in magnitude, and, of course, they are always 180° out of phase with each other. Thus, the two currents cancel and the total current is zero, as presented in Figure 13–42. In this ideal case, the winding resistance of the coil is assumed to be zero.

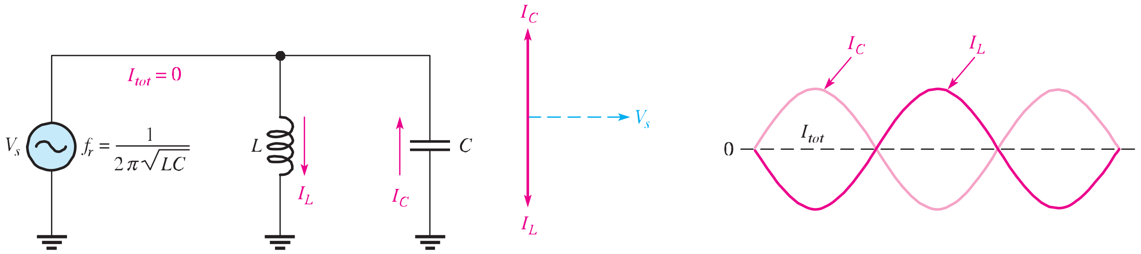

(a) Parallel circuit at resonance ($X_C = X_L$, $Z = \infty$) (b) Current phasors (c) Current waveforms

FIGURE 13–42 **An ideal parallel *LC* circuit at resonance.**

Since the total current is zero, the impedance of the parallel LC circuit is infinitely large (∞). These ideal resonant conditions are stated as follows:

$$X_L = X_C$$
$$Z_r = \infty$$

Parallel Resonant Frequency

For an ideal parallel resonant circuit, the frequency at which resonance occurs is determined by the same formula as in series resonant circuits.

$$f_r = \frac{1}{2\pi\sqrt{LC}} \qquad (13\text{–}16)$$

Currents in a Parallel Resonant Circuit

The currents in a parallel LC circuit vary as the frequency is increased from below the resonant value, through resonance, and then above the resonant value. Figure 13–43 illustrates the general response of an ideal circuit in terms of changes in the currents.

BELOW THE RESONANT FREQUENCY Figure 13–43(a) indicates the response as the source frequency is increased from zero toward f_r. At very low frequencies, X_C is very high and X_L is very low, so most of the current is through L. As the frequency increases, the current through L decreases and the current through C increases, causing the total current to decrease. At all times, I_L and I_C are 180° out of phase with each other, and thus the total current is the difference of the two branch currents. During this time, the impedance is increasing, as indicated by the decrease in total current.

AT THE RESONANT FREQUENCY When the frequency reaches its resonant value, f_r, as shown in Figure 13–43(b), X_C and X_L are equal, so I_C and I_L cancel because they are equal in magnitude but opposite in phase. At this point the total current is at its minimum value of zero. Since I_{tot} is zero, Z is infinite. Thus, the ideal parallel LC circuit appears as an open at f_r.

HANDS ON TIP

When you check the resonant frequency of an LC circuit using a function generator, the Thevenin resistance of the generator increases the Q of the circuit. A generator with an internal resistance of 600 Ω can significantly broaden the bandwidth of the resonant circuit. To reduce the Q and obtain a narrower bandwidth, a small parallel resistor can be placed across the generator output. The drawback of this method is that the signal amplitude from the generator is reduced.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

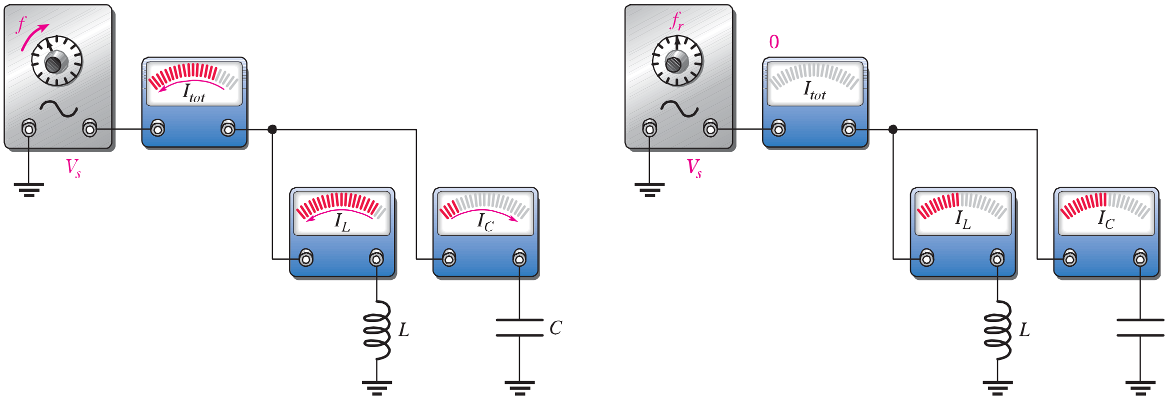

(a) As frequency is increased below resonance: $X_C > X_L$, $I_L > I_C$, I_L decreases, I_C increases, and I_{tot} decreases.

(b) At the resonant frequency: $X_C = X_L$, $I_L = I_C$, and $I_{tot} = 0$.

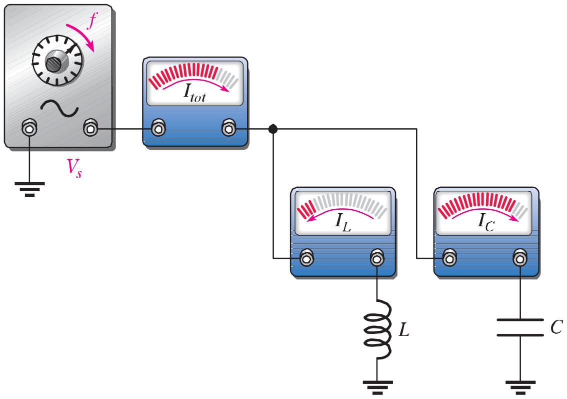

(c) As frequency is increased above resonance: $X_L > X_C$, $I_L < I_C$, I_L decreases, I_C increases, and I_{tot} increases.

FIGURE 13–43 **An illustration of how the currents respond in a parallel *LC* circuit as the frequency is varied from below resonance to above resonance. The source voltage amplitude is constant.**

ABOVE THE RESONANT FREQUENCY As the frequency is increased above resonance, as indicated in Figure 13–43(c), X_C continues to decrease and X_L continues to increase, causing the branch currents again to be unequal, with I_C being larger. As a result, total current increases and impedance decreases. As the frequency becomes very high, the impedance becomes very small due to the dominance of a very small X_C in parallel with a very large X_L.

In summary, the current dips to a minimum as the impedance peaks to a maximum at parallel resonance. The expression for the total current into the L and C branches is

$$I_{tot} = |I_L - I_C| \tag{13–17}$$

EXAMPLE 13–17

A resonant circuit in a communication receiver has a 680 μH inductor in parallel with a 180 pF capacitor as shown in Figure 13–44.

(a) What is the resonant frequency?

(b) If 2.0 V is across the parallel combination, what is the current in each component and the total current?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

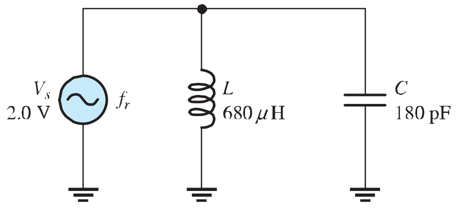

FIGURE 13–44

SOLUTION

(a) The resonant frequency is

$$f_r = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{(680\ \mu\text{H})(180\ \text{pF})}} = \mathbf{455\ kHz}$$

(b) $X_L = 2\pi fL = 2\pi(455\ \text{kHz})(680\ \mu\text{H}) = 1.94\ \text{k}\Omega$

$$X_C = X_L = 1.94\ \text{k}\Omega$$

$$I_L = \frac{V_s}{X_L} = \frac{2.0\ \text{V}}{1.94\ \text{k}\Omega} = \mathbf{1.03\ mA}$$

$$I_C = I_L = \mathbf{1.03\ mA}$$

The total current at resonance is

$$I_{tot} = |I_L - I_C| = \mathbf{0\ A}$$

RELATED PROBLEM

Find the current in the inductor and capacitor if the circuit is tuned to 470 kHz.

6238262 2013/12/10 12:27:06.5

MULTISIM

Open Multisim file E13-17. Determine the resonant frequency by measurement. Determine the currents through L and C at the resonant frequency.

Tank Circuit

The parallel resonant LC circuit is often called a **tank circuit**. The term *tank circuit* refers to the fact that the parallel resonant circuit stores energy in the magnetic field of the coil and in the electric field of the capacitor. The stored energy is transferred back and forth between the capacitor and the coil on alternate half-cycles as the current goes first one way and then the other when the inductor deenergizes and the capacitor charges, and vice versa. This concept is illustrated in Figure 13–45.

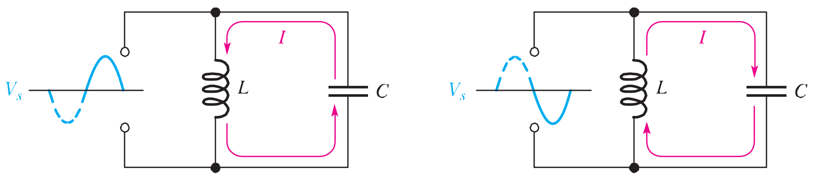

(a) The coil deenergizes as the capacitor charges.

(b) The capacitor discharges as the coil energizes.

FIGURE 13–45 **Energy storage in an ideal parallel resonant tank circuit.**

Parallel Resonant Conditions in a Nonideal Circuit

So far, the resonance of an ideal parallel LC circuit has been examined. Now let's consider resonance in a tank circuit with the winding resistance taken into account. Figure 13–46 shows a nonideal tank circuit and its parallel RLC equivalent.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

FIGURE 13–46 **A practical treatment of parallel resonant circuits includes the winding resistance.**

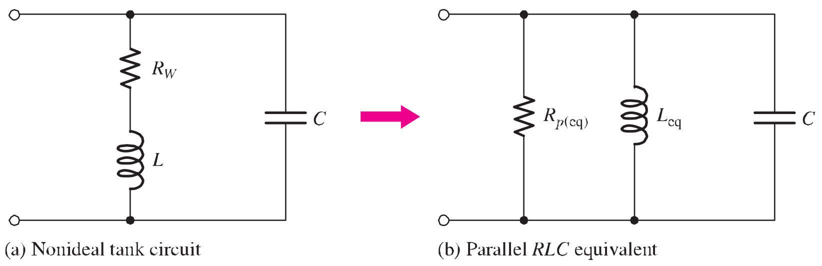

(a) Nonideal tank circuit (b) Parallel *RLC* equivalent

If the winding resistance is the only resistance in the circuit, the quality factor, Q, of the circuit at resonance is simply the Q of the coil.

$$Q = \frac{X_L}{R_W}$$

In terms of circuit component values, the Q can also be expressed as

$$Q = \frac{1}{R_W}\sqrt{\frac{L}{C}}$$

These equations for Q include only the winding resistance, R_W, for the coil and neglect other resistance that may be present. The internal resistance of the voltage source and any load resistance can lower the Q. One common way to measure the actual Q is to insert a relatively small resistor, called a *sense resistor* into the circuit. The sense resistor introduces minimal loading but develops a small voltage that can be used to determine the current. Experiment 27 in the lab manual[1] illustrates measuring the Q of a resonant circuit using a sense resistor.

The expressions for the equivalent inductance and the equivalent parallel resistance were given in Equations 13–14 and 13–15 as

$$L_{\text{eq}} = L\left(\frac{Q^2 + 1}{Q^2}\right)$$

$$R_{p(\text{eq})} = R_W(Q^2 + 1)$$

Recall that for $Q \geq 10$, $L_{\text{eq}} \cong L$.

At parallel resonance,

$$X_{L(\text{eq})} = X_C$$

In the parallel equivalent circuit, $R_{p(\text{eq})}$ is in parallel with an ideal coil and a capacitor, so the L and C branches act as an ideal tank circuit that has an infinite impedance at resonance, as shown in Figure 13–47. Therefore, the total impedance of the nonideal tank circuit at resonance can be expressed as simply the equivalent parallel resistance.

$$Z_r = R_W(Q^2 + 1)$$

FIGURE 13–47 **At resonance, the parallel *LC* portion appears open and the source sees only $R_{p(\text{eq})}$, which equals $R_W(Q^2 + 1)$.**

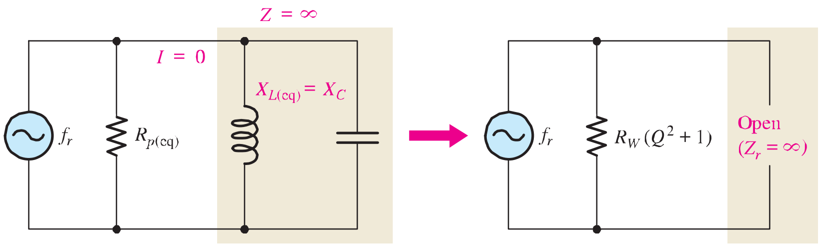

---

[1] *Experiments for DC/AC Fundamentals: A System Approach* by David M. Buchla.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 13–18

Determine the impedance of the circuit in Figure 13–48 at the resonant frequency ($f_r \cong 17{,}794$ Hz).

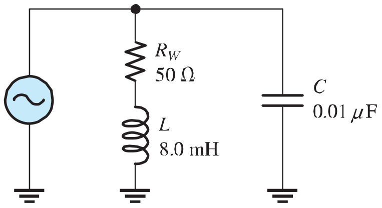

FIGURE 13–48

SOLUTION

Before you can calculate the impedance, you must find the quality factor. To get Q, first find the inductive reactance.

$$X_L = 2\pi f_r L = 2\pi(17{,}794\ \text{Hz})(8.0\ \text{mH}) = 894\ \Omega$$

$$Q = \frac{X_L}{R_W} = \frac{894\ \Omega}{50\ \Omega} = 17.9$$

$$Z_r = R_W(Q^2 + 1) = 50\ \Omega(17.9^2 + 1) = \mathbf{16.0\ k\Omega}$$

RELATED PROBLEM

Determine Z_r in Figure 13–48 if the winding resistance is 10 Ω.

Variation of the Impedance with Frequency

The impedance of a parallel resonant circuit is maximum at the resonant frequency and decreases at lower and higher frequencies, as indicated by the curve in Figure 13–49.

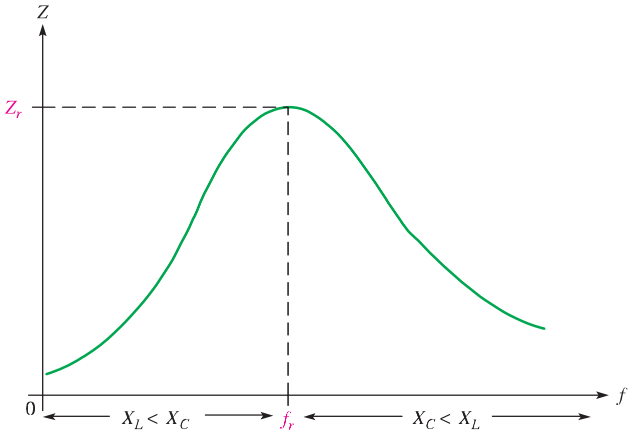

FIGURE 13–49 **Generalized impedance curve for a parallel resonant circuit. The circuit is inductive below f_r, resistive at f_r, and capacitive above f_r.**

At very low frequencies, X_L is very small and X_C is very high, so the total impedance is essentially equal to that of the inductive branch. As the frequency goes up, the impedance also increases, and the inductive reactance dominates (because it is less than X_C) until the resonant frequency is reached. At this point, of course, $X_L \cong X_C$ (for $Q > 10$) and the impedance is at its maximum. As the frequency goes above resonance, the capacitive reactance dominates (because it is less than X_L) and the impedance decreases.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Current and Phase Angle at Resonance

In the ideal tank circuit, the total current from the source at resonance is zero because the impedance is infinite. In the nonideal case, there is some total current at the resonant frequency, and it is determined by the impedance at resonance.

$$\boldsymbol{I_{tot} = \frac{V_s}{Z_r}} \tag{13–18}$$

The phase angle of the parallel resonant circuit is 0° because the impedance is purely resistive at the resonant frequency.

Parallel Resonant Frequency in a Nonideal Circuit

As you know, when the winding resistance is considered, the resonant condition is

$$X_{L(\text{eq})} = X_C$$

which can be expressed as

$$2\pi f_r L\left(\frac{Q^2 + 1}{Q^2}\right) = \frac{1}{2\pi f_r C}$$

Solving for f_r in terms of Q yields

$$f_r = \frac{1}{2\pi\sqrt{LC}}\sqrt{\frac{Q^2}{Q^2 + 1}}$$

When $Q \geq 10$, the term with the Q factors is approximately 1.

$$\sqrt{\frac{Q^2}{Q^2 + 1}} = \sqrt{\frac{100}{101}} = 0.995 \cong 1$$

Therefore, the parallel resonant frequency is approximately the same as the series resonant frequency as long as Q is equal to or greater than 10.

$$f_r \cong \frac{1}{2\pi\sqrt{LC}} \qquad \text{for } Q \geq 10$$

For the case where the R_W of the coil is the only resistance in the circuit, a precise expression for f_r in terms of the circuit component values is

$$\boldsymbol{f_r = \frac{\sqrt{1 - (R_W^2 C/L)}}{2\pi\sqrt{LC}}} \tag{13–19}$$

This precise formula is seldom necessary and for most practical situations the simpler equation $f_r = 1/(2\pi\sqrt{LC})$ is sufficient. However, the next example illustrates the use of Equation 13–19.

EXAMPLE 13–19

Find the frequency, impedance, and total current at resonance for the circuit in Figure 13–50 using Equation 13–19.

FIGURE 13–50

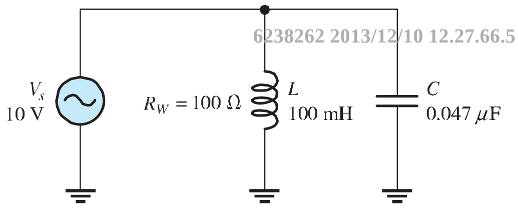

6238262 2013/12/10 12.27.66.5

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

The precise resonant frequency is

$$f_r = \frac{\sqrt{1 - (R_W^2 C/L)}}{2\pi\sqrt{LC}} = \frac{\sqrt{1 - [(100\ \Omega)^2(0.047\ \mu\text{F})/100\ \text{mH}]}}{2\pi\sqrt{(0.047\ \mu\text{F})(100\ \text{mH})}} = \mathbf{2.32\ kHz}$$

Calculate impedance as follows:

$$X_L = 2\pi f_r L = 2\pi(2.32\ \text{kHz})(100\ \text{mH}) = 1.46\ \text{k}\Omega$$

$$Q = \frac{X_L}{R_W} = \frac{1.46\ \text{k}\Omega}{100\ \Omega} = 14.6$$

$$Z_r = R_W(Q^2 + 1) = 100\ \Omega(14.6^2 + 1) = \mathbf{21.4\ k\Omega}$$

The total current is

$$I_{tot} = \frac{V_s}{Z_r} = \frac{10\ \text{V}}{21.4\ \text{k}\Omega} = \mathbf{467\ \mu A}$$

RELATED PROBLEM

Repeat this example using the formula $f_r = 1/(2\pi\sqrt{LC})$ and compare the results.

MULTISIM

Open Multisim file E13-19. Determine the resonant frequency by measurement. Determine the total current and the currents through L and C at the resonant frequency. How do these results compare with the calculated values?

An External Load Resistance Affects a Tank Circuit

In a basic parallel resonant circuit with no load resistor, the Q of the circuit is determined only by the Q of the coil and some loading due to the source resistance. However, the source loading effect is complicated by frequency dependency. At resonance, the source resistance has a small current, which is due to the nonideal characteristic of the tank circuit. As the circuit moves away from resonance, this current increases due to phase shifts. To simplify the discussion, we will ignore the effect of source loading.

In most practical circuits, as illustrated in Figure 13–51(a), an external load (R_L) dissipates most of the power delivered by the source. In this case, the load resistor will lower the Q of the circuit—the smaller the load resistor, the lower the Q. The load resistor, R_L, effectively is in parallel with the equivalent parallel resistance of the coil, $R_{p(\text{eq})}$; both are combined to determine the total parallel resistance, $R_{p(tot)}$, as indicated in Figure 13–51(b) and given as

$$\boldsymbol{R_{p(tot)} = R_L \| R_{p(\text{eq})}} \qquad (13\text{–}20)$$

In the case of an unloaded tank circuit, $R_{p(tot)}$ reduces to $R_{p(\text{eq})}$.

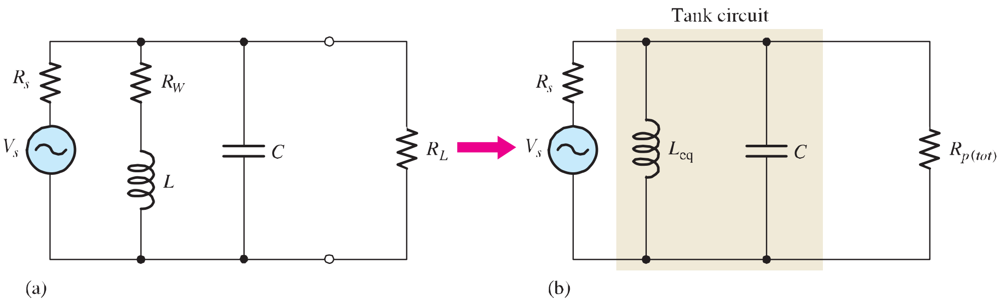

FIGURE 13–51 Tank circuit with a load resistor and its equivalent circuit.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The overall Q, designated Q_O, is given by the approximate equation

$$Q_O \cong \frac{R_{p(tot)}}{X_L} \qquad (13\text{–}21)$$

Notice that this equation is reversed from the Q of a series circuit or the Q of an inductor alone. However, the unloaded case for a parallel circuit will give the same result as the Q of the inductor alone, so either equation can be used with no load.

SYSTEM EXAMPLE 13–1

A BASIC METAL DETECTOR

Resonant circuits are used in many oscillators. Figure 13–52 shows a circuit that uses two oscillators to make a basic metal detector. The resonant circuit of oscillator-1 is the focus of this example and is shown in the yellow highlighted box. The oscillator is a Colpitts type first introduced in System Example 9–1. This particular oscillator will change frequencies when it is near metal because of the change in inductance of the sensing coil (L_1).

To calculate the oscillation frequency of oscillator-1, we apply

$$f_r = \frac{1}{2\pi\sqrt{LC}}$$

Capacitor C_1 has almost no effect on the resonant frequency and can be ignored; however, it is in the circuit to block dc. C_2 and C_3 are in series; together with L_1 they form an equivalent tank circuit (or parallel resonant circuit).

Oscillator-2 is designed to oscillate at a *slightly* different frequency from oscillator-1. It develops a fixed reference frequency that interacts with the frequency generated by oscillator-1. The output of each oscillator is mixed in a nonlinear mixer. (Recall that the superposition theorem allows you to add sources *linearly*, so it does not apply in this case.) By mixing the signals in a nonlinear circuit, two new frequencies appear at the output: the sum of oscillator-1 plus oscillator-2 and the difference in their frequencies; this lower difference frequency is known as a beat frequency. A low-pass filter, consisting of L_2 and R_4, allows only the beat frequency through, which is an audio tone. It is amplified and sent to headphones.

The frequency of oscillator-1 will change when it is near metal due to the magnetic coupling, causing reduced inductance. As a result, the beat frequency also changes and is heard in the headphones as a changing squeal.

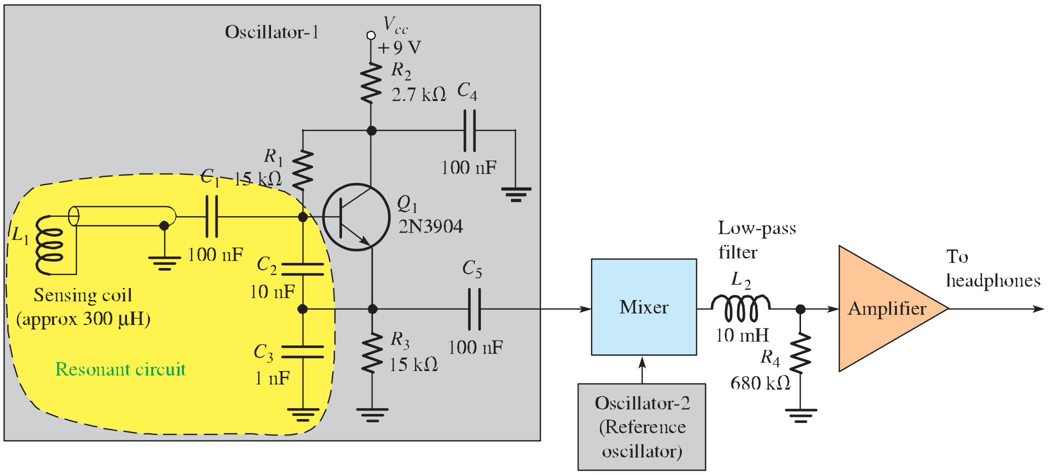

FIGURE 13–52 **A metal detector circuit.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 13–6 CHECKUP

1. Is the impedance minimum or maximum at parallel resonance?
2. Is the current minimum or maximum at parallel resonance?
3. At ideal parallel resonance, $X_L = 1.5\ \text{k}\Omega$. What is X_C?
4. A tank circuit has the following values: $R_W = 5.0\ \Omega$, $L = 220\ \mu\text{H}$, and $C = 0.10\ \mu\text{F}$. Calculate f_r and Z_r.
5. If $Q = 25$, $L = 50$ mH, and $C = 1000$ pF, what is f_r?
6. In Question 5, if $Q = 2.5$, what is f_r?
7. In a certain tank circuit, the coil resistance is $20\ \Omega$. What is the total impedance at resonance if $Q = 20$?

13–7 PARALLEL RESONANT FILTERS

Parallel resonant circuits are commonly applied to band-pass and band-stop filters.

After completing this section, you should be able to

- **Analyze the operation of parallel resonant filters**
  - **Show how a band-pass filter is implemented**
  - **Define *bandwidth***
  - **Explain how loading affects selectivity**
  - **Show how a band-stop filter is implemented**
  - **Determine the resonant frequency, bandwidth, and output voltage of band-pass and band-stop parallel resonant filters**

The Band-Pass Filter

A basic parallel resonant band-pass filter is shown in Figure 13–53. Notice that the output is taken across the tank circuit in this application.

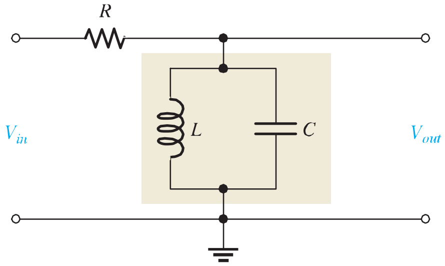

FIGURE 13–53 **A basic parallel resonant band-pass filter.**

6238262 2013/12/10 12.27.66.5

The bandwidth and cutoff frequencies for a parallel resonant band-pass filter are defined in the same way as for the series resonant circuit, and the formulas given in Section 13–4 still apply. General band-pass frequency response curves showing both V_{out} and I_{tot} versus frequency are given in Figures 13–54(a) and 13–54(b), respectively.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

FIGURE 13–54 **Generalized frequency response curves for a parallel resonant band-pass filter.**

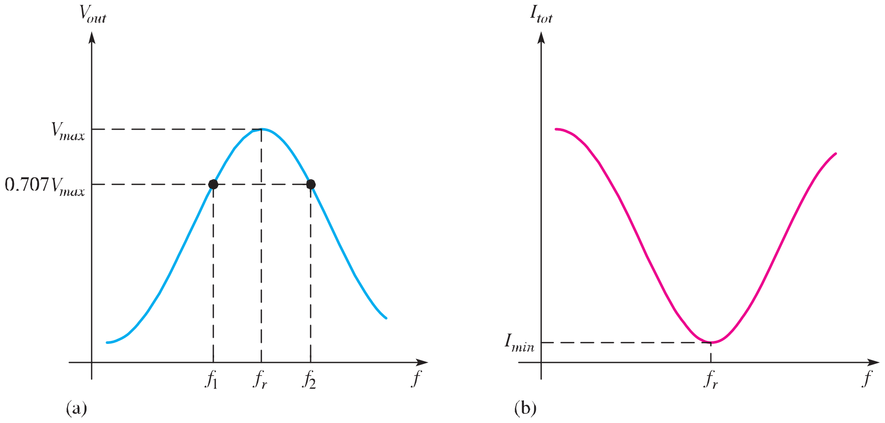

The filtering action is as follows: At very low frequencies, the impedance of the tank circuit is very low, and therefore only a small amount of the input voltage is dropped across it; the rest is dropped across R. As the frequency increases, the impedance of the tank circuit increases, and, as a result, the output voltage increases. When the frequency reaches its resonant value, the impedance is at its maximum and most of the input voltage is developed across the tank circuit. As the frequency goes above resonance, the impedance begins to decrease, causing the output voltage to decrease. The general response of a parallel resonant band-pass filter is illustrated in Figure 13–55. This illustration shows how the current and the output voltage change with frequency.

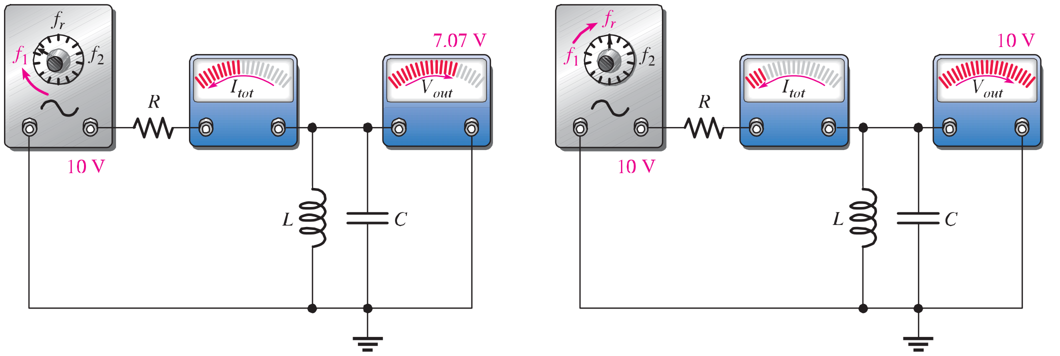

(a) As the frequency increases to f_1, V_{out} increases to 7.07 V, and I_{tot} decreases.

(b) As the frequency increases from f_1 to f_r, V_{out} increases from 7.07 V to 10 V, and I_{tot} decreases to its minimum value.

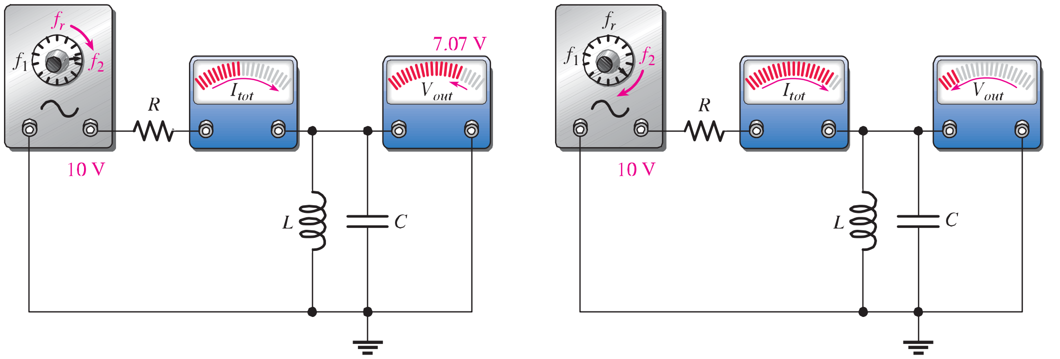

(c) As the frequency increases from f_r to f_2, V_{out} decreases from 10 V to 7.07 V, and I_{tot} increases from its minimum.

(d) As the frequency increases above f_2, V_{out} decreases below 7.07 V, and I_{tot} continues to increase.

FIGURE 13–55 **Example of the response of a parallel resonant band-pass filter with the input voltage at a constant 10 V rms.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 13–20

A certain parallel resonant band-pass filter has a maximum output voltage of 4 V at f_r. What is the value of V_{out} at the cutoff frequencies?

SOLUTION

V_{out} at the cutoff frequencies is 70.7% of maximum.

$$V_{out(1)} = V_{out(2)} = 0.707V_{out(max)} = 0.707(4\text{ V}) = \mathbf{2.83\text{ V}}$$

RELATED PROBLEM

What is V_{out} at the cutoff frequencies if V_{out} at the resonant frequency is 10 V?

EXAMPLE 13–21

A parallel resonant circuit has a lower cutoff frequency of 3.5 kHz and an upper cutoff frequency of 6 kHz. What is the bandwidth?

SOLUTION

$$BW = f_2 - f_1 = 6\text{ kHz} - 3.5\text{ kHz} = \mathbf{2.5\text{ kHz}}$$

RELATED PROBLEM

A filter has a lower cutoff frequency of 520 kHz and a bandwidth of 10 kHz. What is the upper cutoff frequency?

EXAMPLE 13–22

A certain parallel resonant band-pass filter has a resonant frequency of 12 kHz and a Q of 10. What is its bandwidth?

SOLUTION

$$BW = \frac{f_r}{Q} = \frac{12\text{ kHz}}{10} = \mathbf{1.2\text{ kHz}}$$

RELATED PROBLEM

A certain parallel resonant band-pass filter has a resonant frequency of 100 MHz and a bandwidth of 4 MHz. What is its Q?

LOADING EFFECTS When a resistive load is connected across the output of a filter as shown in Figure 13–56(a), the Q of the filter is reduced. Since $BW = f_r/Q$, the bandwidth is increased, thus reducing the selectivity. Also, the impedance of the filter at resonance is decreased because R_L effectively appears in parallel with $R_{p(\text{eq})}$. Thus, the maximum output voltage is reduced by the voltage-divider effect of $R_{p(tot)}$ and the internal source resistance R_s, as illustrated in Figure 13–56(b). Figure 13–56(c) shows the effect of a load on the specific resonant circuit described in Example 13–23. The effect of a load drops the peak and increases the bandwidth as shown.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

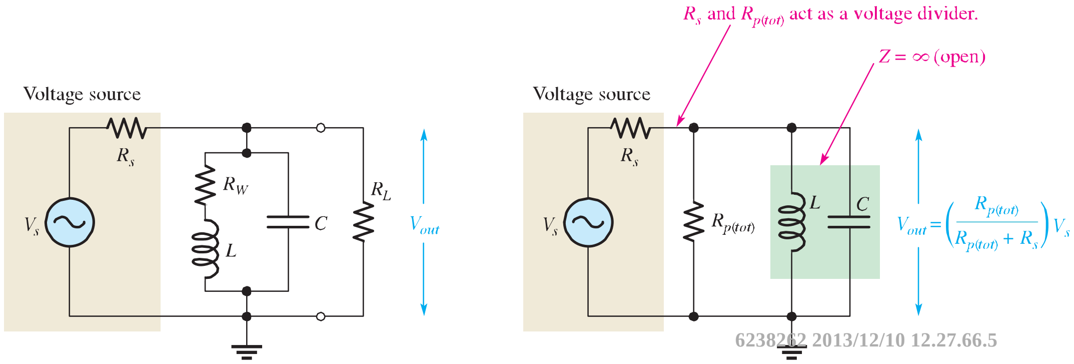
Voltage source
R_s
R_W
L
C
R_L
V_s
V_{out}
R_s and $R_{p(tot)}$ act as a voltage divider.
$Z = \infty$ (open)
Voltage source
R_s
V_s
$R_{p(tot)}$
L
C
$V_{out} = \left(\frac{R_{p(tot)}}{R_{p(tot)} + R_s}\right)V_s$

6238262 2013/12/10 12.27.66.5

(a) Tank circuit with load

(b) Equivalent circuit at resonance

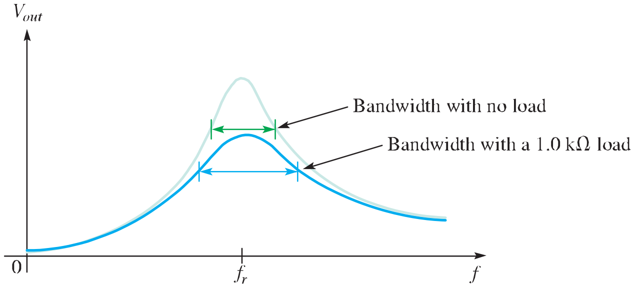

(c) Loading widens the bandwidth and reduces the output.

FIGURE 13–56 **Effects of loading on a parallel resonant band-pass filter. The resistor, *R*, in Figure 13–53 is now represented by the internal source resistance R_s.**

EXAMPLE 13–23

(a) Determine f_r, V_{out}, and *BW* at resonance for the filter in Figure 13–57(a). The inductor has a winding resistance of 10.5 Ω. The internal source resistance is 50 Ω.

(b) Repeat part (a) when the filter is loaded with a 1.0 kΩ resistance and compare the results.

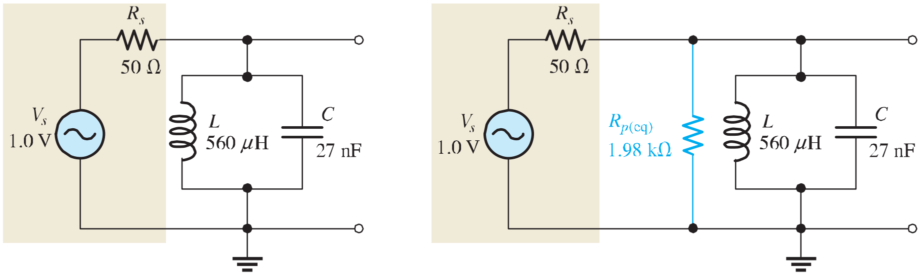
R_s
50 Ω
V_s
1.0 V
L
560 μH
C
27 nF
R_s
50 Ω
V_s
1.0 V
$R_{p(eq)}$
1.98 kΩ
L
560 μH
C
27 nF

(a) Original circuit

(b) Equivalent circuit with $R_{p(eq)}$ shown in blue

FIGURE 13–57

SOLUTION

(a) $$f_r = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{(560\ \mu\text{H})(27\ \text{nF})}} = \mathbf{40.9\ kHz}$$

The circuit is unloaded. To calculate the output voltage, begin by finding the *Q* of the coil and $R_{p(eq)}$.

$$X_L = 2\pi fL = 2\pi(40.9\ \text{kHz})(560\ \mu\text{H}) = 144\ \Omega$$

$$Q = \frac{X_L}{R_W} = \frac{144\ \Omega}{10.5\ \Omega} = 13.7$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Apply Equation 13–15 to calculate $R_{p(\text{eq})}$ (parallel equivalent resistance of the coil).

$$R_{p(\text{eq})} = R_W(Q^2 + 1) = 10.5\ \Omega(13.7 + 1) = 1.98\ \text{k}\Omega$$

Because the circuit is unloaded, $R_{p(tot)} = R_{p(\text{eq})} = 1.98\ \text{k}\Omega$, as shown in Figure 13–57(b). At resonance, the tank circuit looks like an open, so apply the voltage-divider formula to $R_{p(tot)}$ and R_s to find V_{out}.

$$V_{out} = \left(\frac{R_{p(tot)}}{R_{p(tot)} + R_s}\right)V_s = \left(\frac{1.98\ \text{k}\Omega}{1.98\ \text{k}\Omega + 50\ \Omega}\right)1.0\ \text{V} = \mathbf{975\ mV}$$

To calculate the *BW*, apply Equation 13–21.

$$Q_O \cong \frac{R_{p(tot)}}{X_L} \cong \frac{1.98\ \text{k}\Omega}{144\ \Omega} = 13.8$$

Notice that this is the same Q (within round off error) as the inductor alone, because the circuit is unloaded.

$$BW = \frac{f_r}{Q_O} = \frac{40.9\ \text{kHz}}{13.8} = \mathbf{2.96\ kHz}$$

(b) The resonant frequency is not significantly affected by the load and is equal to

$$f_r = \mathbf{40.9\ kHz}$$

The load drops the output voltage because it appears to be in parallel with $R_{p(\text{eq})}$. Apply Equation 13–20.

$$R_{p(tot)} = R_L \| R_{p(\text{eq})} = 1.0\ \text{k}\Omega \| 2.07\ \text{k}\Omega = 674\ \Omega$$

Applying the voltage-divider formula,

$$V_{out} = \left(\frac{R_{p(tot)}}{R_{p(tot)} + R_s}\right)V_s = \left(\frac{674\ \Omega}{674\ \Omega + 50\ \Omega}\right)1.0\ \text{V} = \mathbf{930\ mV}$$

To calculate the *BW*, apply Equation 13–21.

$$Q_O = \frac{R_{p(tot)}}{X_L} = \frac{674\ \Omega}{144\ \Omega} = 4.68$$

$$BW = \frac{f_r}{Q_O} = \frac{40.9\ \text{kHz}}{4.68} = \mathbf{8.74\ kHz}$$

The load decreased the output voltage and caused the bandwidth to increase.

RELATED PROBLEM

How is Q_O affected by a larger-value load resistor?

The Band-Stop Filter

A basic parallel resonant band-stop filter is shown in Figure 13–58. The output is taken across a load resistor in series with the tank circuit. The load resistor has only a small effect on Q in this configuration, and generally the effect can be ignored.

MULTISIM

Open the Multisim file E13-23. To view the response, the Bode plotter is connected across a "sense resistor" that acts as a current-to-voltage converter. Compare results in Multisim with the calculated values in this example. View the response on the Bode plotter. Try changing the winding resistance of the coil and see what happens.

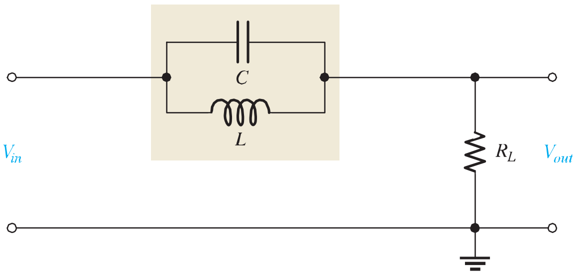

FIGURE 13–58 **A basic parallel resonant band-stop filter.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The variation of the tank circuit impedance with frequency produces the familiar current response that has been previously discussed; that is, the current is minimum at resonance and increases on both sides of resonance. Since the output voltage is across the series load resistor, the output voltage follows the current, thus creating the band-stop response characteristic, as indicated in Figure 13–59.

FIGURE 13–59 **Band-stop filter response.**

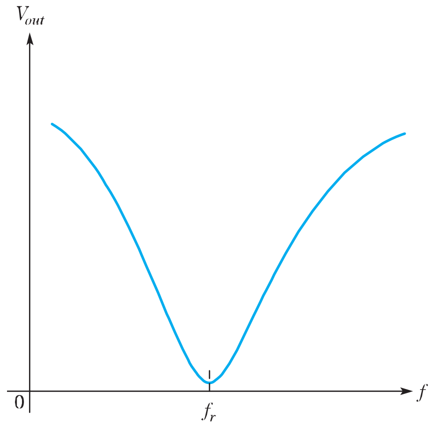

Actually, the band-stop filter in Figure 13–58 can be viewed as a voltage divider created by Z_r of the tank and the load resistance. Thus, the output voltage at f_r is

$$V_{out} = \left(\frac{R_L}{R_L + Z_r}\right)V_{in}$$

Wave Traps

A *wave trap* is a resonant circuit designed to significantly reduce interference from a specific source. For example, a wave trap to prevent interference from a citizen band radio at 27 MHz is a parallel resonant circuit tuned to that frequency. Other frequencies can pass through the filter with little effect. The circuit shown above is a wave trap designed to be inserted in 300 Ω twin lead, which is a commonly used radio frequency transmission line. The wave trap is installed in a shielded enclosure with the shield connected to ground.

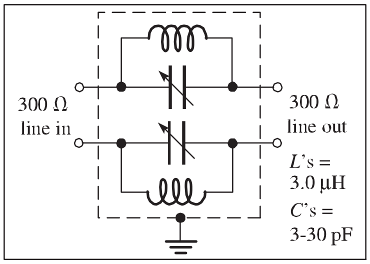

SYSTEM NOTE

EXAMPLE 13–24

Figure 13–60 shows the tank circuit from Example 13–23, which is used in a band-stop filter. Determine f_r and V_{out} at resonance. The winding resistance of the inductor is 10.5 Ω.

FIGURE 13–60

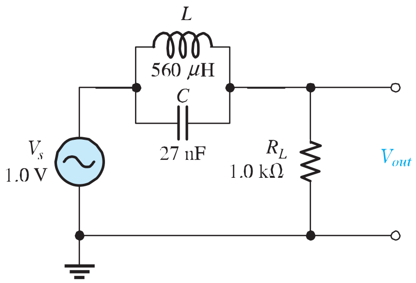

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

The resonant frequency is

$$f_r = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{(560\ \mu\text{H})(27\ \text{nF})}} = \mathbf{40.9\ kHz}$$

Determine V_{out} as follows:

$$X_L = 2\pi f L = 2\pi(40.9\ \text{kHz})(560\ \mu\text{H}) = 144\ \Omega$$

$$Q = \frac{X_L}{R_W} = \frac{144\ \Omega}{10.5\ \Omega} = 13.7$$

Applying Equation 13–15,

$$R_{p(\text{eq})} = R_W(Q^2 + 1) = 10.5\ \Omega(13.7 + 1) = 1.98\ \text{k}\Omega$$

At resonance,

$$V_{out} = \left(\frac{R_L}{R_L + R_{p(\text{eq})}}\right)V_s = \left(\frac{1.0\ \text{k}\Omega}{1.0\ \text{k}\Omega + 1.98\ \text{k}\Omega}\right)1.0\ \text{V} = \mathbf{335\ mV}$$

RELATED PROBLEM

What is the output at the resonant frequency if the load resistor is 470 Ω?

MULTISIM

Open the Multisim file E13-24. Compare results in Multisim with the calculations. You can read the resonant frequency and output voltage on the Bode plotter.

SECTION 13–7 CHECKUP

1. How can the bandwidth of a parallel resonant filter be increased?
2. The resonant frequency of a certain high-Q ($Q > 10$) filter is 5 kHz. If the Q is lowered to 2, does f_r change and, if so, to what value?
3. What is the impedance of a tank circuit at its resonant frequency if $R_W = 75\ \Omega$ and $Q = 25$?

13–8 RESONANT CIRCUIT APPLICATIONS

Resonant circuits are used in a wide variety of applications, particularly in communication systems. In this section, we will look at a few system applications to illustrate the importance of resonant circuits in electronic communication.

After completing this section, you should be able to

- **Discuss some applications of resonant circuits**
  - **Describe a tuned amplifier and double-tuned transformer coupling**
  - **Describe antenna coupling**
  - **Describe an AM radio receiver**

6238262 2013/12/10 12.27.66.5

Tuned Amplifiers

A *tuned amplifier* is a circuit that amplifies signals within a specified band. Typically, a parallel resonant circuit is used in conjunction with an amplifier to achieve the selectivity. In terms of the general operation, input signals with frequencies that range over a wide band are accepted on the amplifier's input and are amplified. The resonant circuit allows only a relatively narrow band of those frequencies to be passed on. The variable capacitor

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

allows tuning over the range of input frequencies so that a desired frequency can be selected, as indicated in Figure 13–61.

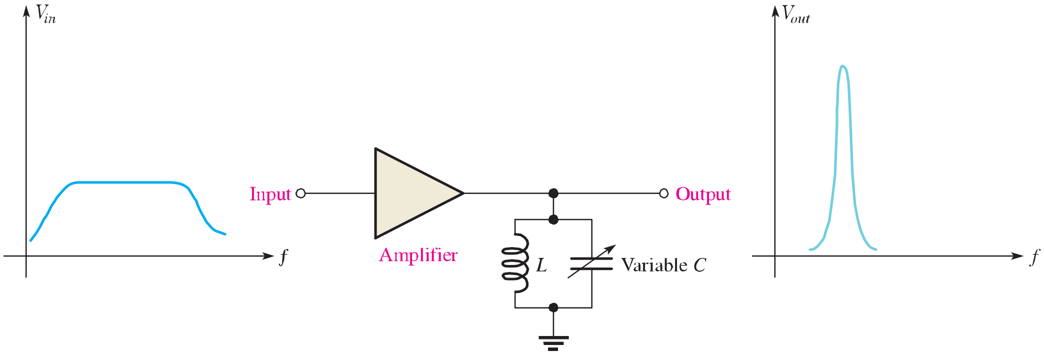

FIGURE 13–61 **A basic tuned band-pass amplifier.**

Double-Tuned Transformer Coupling in a Receiver

In some types of communication receivers, tuned amplifiers are transformer-coupled together to increase the amplification. Capacitors can be placed in parallel with the primary and secondary windings of the transformer, effectively creating two parallel resonant band-pass filters that are coupled together. This technique, illustrated in Figure 13–62, can result in a wider bandwidth and steeper slopes on the response curve, thus increasing the selectivity for a desired band of frequencies.

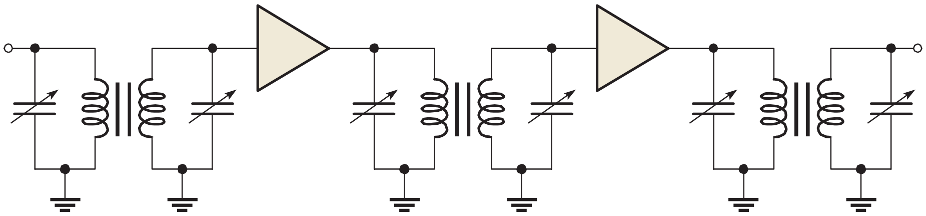

FIGURE 13–62 **Double-tuned amplifiers.**

Antenna Input to a Receiver

Radio signals are sent out from a transmitter via electromagnetic waves that propagate through the atmosphere. When the electromagnetic waves cut across the receiving antenna, small voltages are induced. Out of the wide range of electromagnetic frequencies, only one frequency or a limited band of frequencies must be extracted. Figure 13–63 shows a typical arrangement of an antenna coupled to the receiver input by a transformer (covered in

FIGURE 13–63 **Resonant coupling from an antenna.**

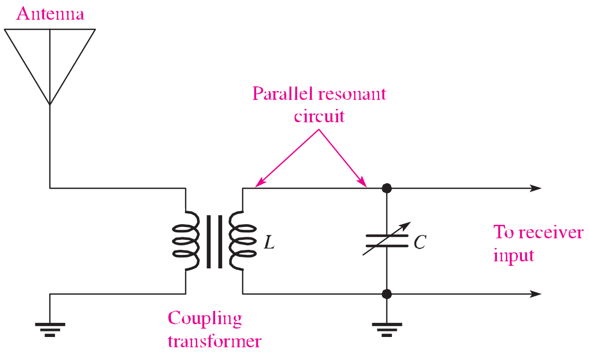

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Chapter 14). A variable capacitor is connected across the transformer secondary to form a parallel resonant circuit.

Superheterodyne Receiver

Another example of resonant circuit application is in the common AM (amplitude modulation) receiver. The AM broadcast band ranges from 535 kHz to 1605 kHz. Each AM station is assigned a 10 kHz bandwidth within that range. The tuned circuits are designed to pass only the signals from the desired radio station, rejecting all others. To reject stations outside the one that is tuned, the tuned circuits must be selective, passing on only the signals in the 10 kHz band and rejecting all others. Too much selectivity is not desirable either however. If the bandwidth is too narrow, some of the higher frequency modulated signals will be rejected, resulting in a loss of fidelity. Ideally, the resonant circuit must reject signals that are not in the desired passband. A simplified block diagram of a superheterodyne AM receiver is shown in Figure 13–64.

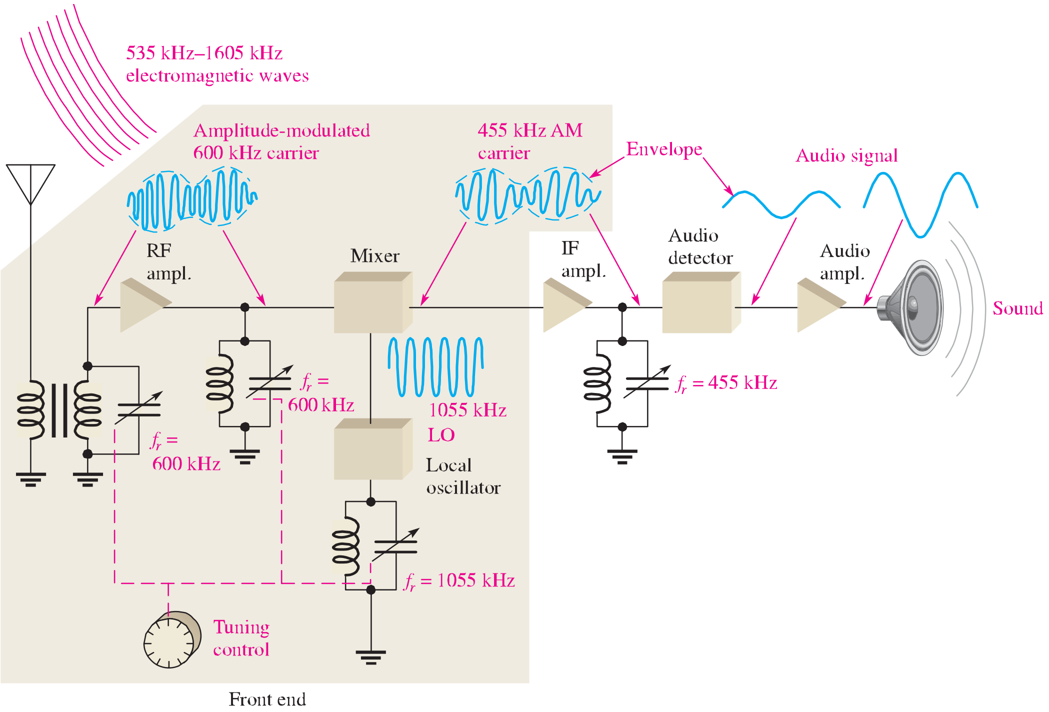

FIGURE 13–64 **A simplified diagram of a superheterodyne AM radio broadcast receiver showing the application of tuned resonant circuits.**

There are basically three parallel resonant circuits in the front end of the receiver. Each of these resonant circuits is gang-tuned by capacitors; that is, the capacitors are mechanically or electronically linked together so that they change together as a station is selected. The front end is tuned to receive a desired station, for example, one that transmits at 600 kHz. The input resonant circuit from the antenna and the RF (radio frequency) amplifier resonant circuit select only a frequency of 600 kHz out of all the frequencies crossing the antenna.

The actual audio (sound) signal is carried by the 600 kHz carrier frequency by modulating the amplitude of the carrier so that it follows the audio signal as indicated. The variation in the amplitude of the carrier corresponding to the audio signal is called the *envelope*. The 600 kHz is then applied to a circuit called the *mixer*.

The local oscillator (LO) is tuned to a frequency that is 455 kHz above the selected frequency (1055 kHz, in this case). By a process called *heterodyning* or *beating*, the AM signal and the local oscillator signal are mixed together, and the 600 kHz AM signal is converted by the mixer to a 455 kHz AM signal (1055 kHz − 600 kHz = 455 kHz).

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The 455 kHz is the intermediate frequency (IF) for standard AM receivers. No matter which station within the broadcast band is selected, its frequency is always converted to the 455 kHz IF. The amplitude-modulated IF is amplified by the IF amplifier which is tuned to 455 kHz. The output of the IF amplifier is applied to an audio detector which removes the IF, leaving only the envelope which is the audio signal. The audio signal is then amplified and applied to the speaker. 6238262 2013/12/10 12.27.66.5

SECTION 13–8 CHECKUP

1. Generally, why is a tuned resonant circuit necessary when a signal is coupled from an antenna to the input of a receiver?

2. What is the advantage of using an intermediate frequency in an AM radio?

3. What is meant by *ganged tuning?*

SUMMARY

- X_L and X_C have opposing effects in an *RLC* circuit.
- In a series *RLC* circuit, the larger reactance determines the net reactance of the circuit.
- In a parallel *RLC* circuit, the smaller reactance determines the net reactance of the circuit.

SERIES RESONANCE

- The reactances are equal.
- The impedance is minimum and equal to the resistance.
- The current is maximum.
- The phase angle is zero.
- The voltages across L and C are equal in magnitude and, as always, 180° out of phase with each other and thus they cancel.

PARALLEL RESONANCE

- The reactances are approximately equal for $Q \geq 10$.
- The impedance is maximum.
- The current is minimum and, ideally, equal to zero.
- The phase angle is zero.
- The currents in the L and C branches are equal in magnitude and, as always, 180° out of phase with each other and thus they cancel.
- A band-pass filter passes frequencies between the lower and upper critical frequencies and rejects all others.
- A band-stop filter rejects frequencies between its lower and upper critical frequencies and passes all others.
- The bandwidth of a resonant filter is determined by the quality factor (Q) of the circuit and the resonant frequency.
- Cutoff frequencies are also called -3 dB frequencies or critical frequencies.
- The output voltage is 70.7% of its maximum at the cutoff frequencies.

KEY TERMS

Key terms and other bold terms in the chapter are defined in the end-of-book glossary.

Band-pass filter A resonant circuit that passes a range of frequencies lying between two cutoff frequencies and rejects frequencies above and below the range.

Band-stop filter A resonant circuit that rejects a range of frequencies lying between two cutoff frequencies and passes frequencies above and below the range.

Cutoff frequency (f_c) The frequency at which the output voltage of a filter is 70.7% of the maximum output voltage.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Decibel (dB) A unit of logarithmic expression ten times the logarithm of the ratio of two powers or twenty times the logarithm of the ratio of two voltages measured across equal resistances.

Half-power frequency The frequency at which the output power of a filter is 50% of the maximum (the output voltage is 70.7% of maximum); another name for *critical frequency* or *cutoff frequency*.

Parallel resonance In a parallel *RLC* circuit, the condition where the impedance is maximum and the reactances are equal.

Resonant frequency The frequency at which a resonant condition occurs in a series or parallel *RLC* circuit.

Selectivity A measure of how effectively a resonant circuit passes certain frequencies and rejects others. The narrower the bandwidth, the greater the selectivity.

Series resonance In a series *RLC* circuit, the condition where the impedance is minimum and the reactances are equal.

KEY FORMULAS

(13–1)	$X_{tot} = \lvert X_L - X_C \rvert$	Total series reactance (absolute value)
(13–2)	$Z_{tot} = \sqrt{R^2 + X_{tot}^2}$	Total series *RLC* impedance
(13–3)	$\theta = \tan^{-1}\left(\frac{X_{tot}}{R}\right)$	Series *RLC* phase angle
(13–4)	$f_r = \frac{1}{2\pi\sqrt{LC}}$	Ideal series resonant frequency
(13–5)	$BW = f_2 - f_1$	Bandwidth
(13–6)	$\text{dB} = 10 \log\left(\frac{P_{out}}{P_{in}}\right)$	Decibel formula for voltage ratio
(13–7)	$\text{dB} = 20 \log\left(\frac{V_{out}}{V_{in}}\right)$	Decibel formula for power ratio
(13–8)	$Q = \frac{X_L}{R}$	Series resonant quality factor
(13–9)	$BW = \frac{f_r}{Q}$	Bandwidth
(13–10)	$Y = \sqrt{G^2 + B_{tot}^2}$	Parallel *RLC* admittance
(13–11)	$\theta = \tan^{-1}\left(\frac{B_{tot}}{G}\right)$	Parallel *RLC* phase angle
(13–12)	$I_{tot} = \sqrt{I_R^2 + I_{CL}^2}$	Total parallel *RLC* current
(13–13)	$\theta = \tan^{-1}\left(\frac{I_{CL}}{I_R}\right)$	Parallel *RLC* phase angle
(13–14)	$L_{eq} = L\left(\frac{Q^2 + 1}{Q^2}\right)$	Equivalent parallel inductance
(13–15)	$R_{p(eq)} = R_W(Q^2 + 1)$	Equivalent parallel resistance
(13–16)	$f_r = \frac{1}{2\pi\sqrt{LC}}$	Ideal parallel resonant frequency
(13–17)	$I_{tot} = \lvert I_L - I_C \rvert$	Total parallel *LC* current (absolute value)
(13–18)	$I_{tot} = \frac{V_s}{Z_r}$	Total current at parallel resonance

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

(13–19)	$f_r = \dfrac{\sqrt{1 - (R_W^2 C/L)}}{2\pi\sqrt{LC}}$	Parallel resonant frequency (exact)
(13–20)	$R_{p(tot)} = R_L \parallel R_{p(\text{eq})}$	Total parallel resistance
(13–21)	$Q_O \cong \dfrac{R_{p(tot)}}{X_{L(\text{eq})}}$	Overall Q of a parallel *RLC* circuit

TRUE/FALSE QUIZ

Answers are at the end of the chapter.

1. A series *RLC* circuit can have a higher voltage than the source voltage across the resistor.
2. The impedance of a series *RLC* circuit is dependent on the source voltage.
3. Above the resonant frequency, a series resonant circuit will look inductive and current will lag the voltage.
4. A band-pass filter can be constructed from a *RLC* circuit.
5. A parallel resonant band-stop filter has minimum impedance at the resonant frequency.
6. The upper and lower cutoff frequencies of a band-stop filter determine the bandwidth.
7. The Q of an inductor is dependent on the frequency at which it is measured.
8. The Q of a band-pass filter does not affect the bandwidth.
9. In a parallel *RLC* circuit, the total impedance is always greater than the resistance.
10. At resonance, the current in a parallel resonant circuit is the same in all components.

SELF-TEST

Answers are at the end of the chapter.

1. The total reactance of a series *RLC* circuit at resonance is
   (a) zero **(b)** equal to the resistance **(c)** infinity **(d)** capacitive
2. The phase angle of a series *RLC* circuit at resonance is
   (a) −90° **(b)** +90° **(c)** 0° **(d)** dependent on the reactance
3. The impedance at the resonant frequency of a series *RLC* circuit with $L = 15$ mH, $C = 0.015\ \mu$F, and $R_W = 80\ \Omega$ is
   (a) 15 kΩ **(b)** 80 Ω **(c)** 30 Ω **(d)** 0 Ω
4. In a series *RLC* circuit that is operating below the resonant frequency, the current
   (a) is in phase with the source voltage
   (b) lags the source voltage
   (c) leads the source voltage
5. If the value of C in a series *RLC* circuit is increased, the resonant frequency
   (a) is not affected **(b)** increases
   (c) remains the same **(d)** decreases
6. In a certain series resonant circuit, $V_C = 150$ V, $V_L = 150$ V, and $V_R = 50$ V. The value of the source voltage is
   (a) 150 V **(b)** 300 V **(c)** 50 V **(d)** 350 V
7. A certain series resonant band-pass filter has a bandwidth of 1 kHz. If the existing coil is replaced with one having a lower value of Q, the bandwidth will
   (a) increase **(b)** decrease **(c)** remain the same **(d)** be more selective
8. At frequencies below resonance in a parallel *RLC* circuit, the current
   (a) leads the source voltage
   (b) lags the source voltage
   (c) is in phase with the source voltage

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

9. The total current into the L and C branches of a parallel circuit at resonance is ideally
(a) maximum **(b)** low **(c)** high **(d)** zero

10. To tune a parallel resonant circuit to a lower frequency, the capacitance should be
(a) increased **(b)** decreased **(c)** left alone **(d)** replaced with inductance

11. The resonant frequency of a parallel circuit is the same as a series circuit using the same components when
(a) the Q is very low **(b)** the Q is very high **(c)** there is no resistance

12. If the resistance in parallel with a parallel resonant filter is reduced, the bandwidth
(a) disappears **(b)** decreases **(c)** becomes sharper **(d)** increases

PROBLEMS

Answers to odd-numbered problems are at the end of the book.

BASIC PROBLEMS

SECTION 13–1 Impedance and Phase Angle of Series *RLC* Circuits

1. A certain series RLC circuit operates at a frequency of 5 kHz and has the following values: $R = 10\ \Omega$, $C = 0.047\ \mu\text{F}$, and $L = 5$ mH. Determine the impedance and phase angle. What is the total reactance?

2. Find the impedance in Figure 13–65.

3. If the frequency of the source voltage in Figure 13–65 is doubled from the value that produces the indicated reactances, how does the impedance change?

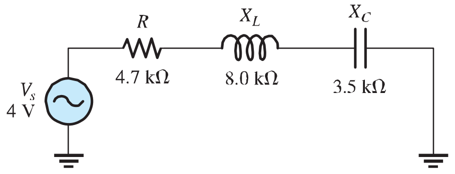

FIGURE 13–65

6238262 2013/12/10 12.27.66.5

SECTION 13–2 Analysis of Series *RLC* Circuits

4. For the circuit in Figure 13–65, find I_{tot}, V_R, V_L, and V_C.

5. Draw the voltage phasor diagram for the circuit in Figure 13–65.

6. Analyze the circuit in Figure 13–66 for the following ($f = 25$ kHz):
(a) I_{tot} **(b)** P_{true} **(c)** P_r **(d)** P_a

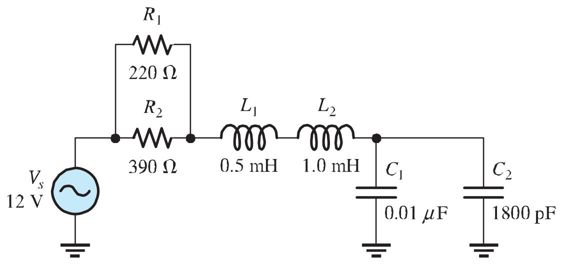

FIGURE 13–66

SECTION 13–3 Series Resonance

7. For the circuit in Figure 13–65, is the resonant frequency higher or lower than the value indicated by the reactances?

8. For the circuit in Figure 13–67, determine the voltage across R at resonance.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

FIGURE 13–67

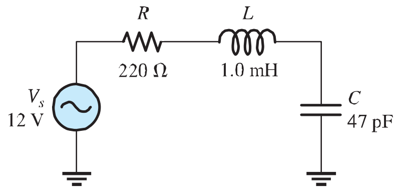
R
220 Ω
L
1.0 mH
V_s
12 V
C
47 pF

9. Find X_L, X_C, Z, and I at the resonant frequency in Figure 13–67.

10. A certain series resonant circuit has a maximum current of 50 mA and a V_L of 100 V. The source voltage is 10 V. What is Z? What are X_L and X_C?

11. For the *RLC* circuit in Figure 13–68, determine the resonant frequency and the cutoff frequencies.

12. What is the value of the current at the half-power points in Figure 13–68?

FIGURE 13–68

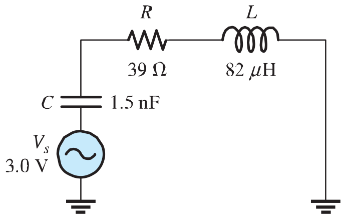
R
39 Ω
L
82 μH
C
1.5 nF
V_s
3.0 V

SECTION 13–4 Series Resonant Filters

13. Determine the resonant frequency for each filter in Figure 13–69. Are these filters band-pass or band-stop types?

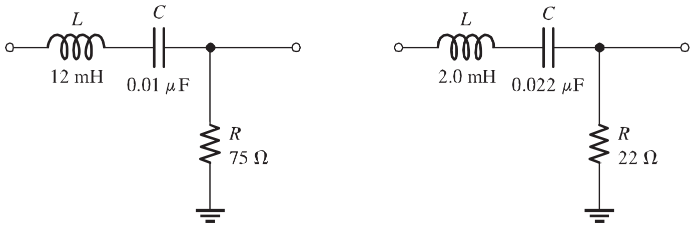
L
12 mH
C
0.01 μF
R
75 Ω
L
2.0 mH
C
0.022 μF
R
22 Ω

(a) (b)

FIGURE 13–69

14. Assuming that the coils in Figure 13–69 have a winding resistance of 10 Ω, find the bandwidth for each filter.

15. Determine f_r and BW for each filter in Figure 13–70.

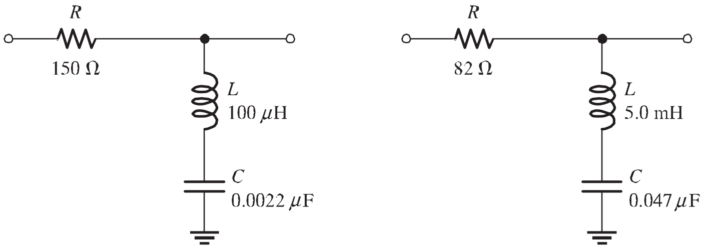
R
150 Ω
L
100 μH
C
0.0022 μF
R
82 Ω
L
5.0 mH
C
0.047 μF

(a) (b)

FIGURE 13–70

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 13–5 Parallel *RLC* Circuits

16. Find the total impedance of the circuit in Figure 13–71.

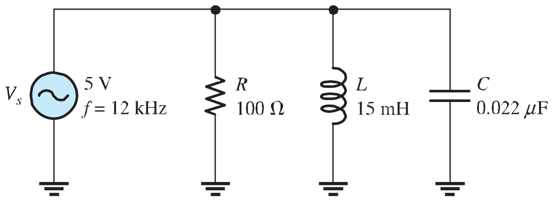

FIGURE 13–71

17. Is the circuit in Figure 13–71 capacitive or inductive? Explain.

18. For the circuit in Figure 13–71, find all the currents and voltages.

19. Find the total impedance for the circuit in Figure 13–72.

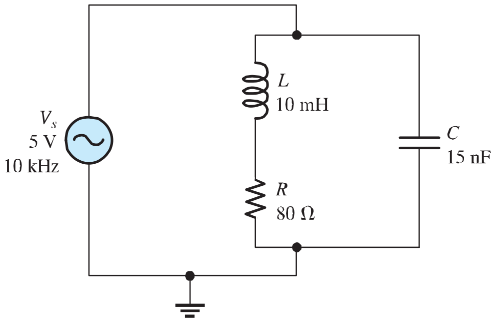

FIGURE 13–72

SECTION 13–6 Parallel Resonance

20. What is the impedance of an ideal parallel resonant circuit (no resistance in either branch)?

21. Find Z at resonance and f_r for the tank circuit in Figure 13–73.

22. How much current is drawn from the source in Figure 13–73 at resonance? What are the inductive current and the capacitive current at the resonant frequency?

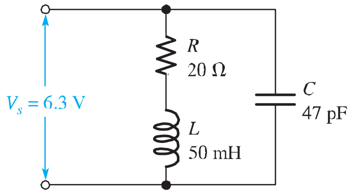

FIGURE 13–73

SECTION 13–7 Parallel Resonant Filters

23. At resonance, $X_L = 2\text{ k}\Omega$ and $R_W = 25\ \Omega$ in a parallel resonant band-pass filter. The resonant frequency is 5 kHz. Determine the bandwidth.

24. If the lower cutoff frequency is 2400 Hz and the upper cutoff frequency is 2800 Hz, what is the bandwidth?

25. In a certain resonant circuit, the power to the load at resonance is 2.75 W. What is the power at the lower and upper cutoff frequencies?

26. What values of L and C should be used in a tank circuit to obtain a resonant frequency of 8 kHz? The bandwidth must be 800 Hz. The winding resistance of the coil is 10 Ω.

27. A parallel resonant circuit has a Q of 50 and a BW of 400 Hz. If Q is doubled, what is the bandwidth for the same f_r?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

28. A parallel resonant band-stop filter is needed to reject 60 Hz power line noise. What size should the inductor be if the capacitor is 200 μF?

29. Draw the circuit described in Problem 28 if the output is taken across a 220 Ω resistor.

ADVANCED PROBLEMS

30. For each following case, express the voltage ratio in decibels:
(a) $V_{in} = 1$ V, $V_{out} = 1$ V **(b)** $V_{in} = 5$ V, $V_{out} = 3$ V
(c) $V_{in} = 10$ V, $V_{out} = 7.07$ V **(d)** $V_{in} = 25$ V, $V_{out} = 5$ V

31. Find the current through each component in Figure 13–74. Find the voltage across each component.

FIGURE 13–74

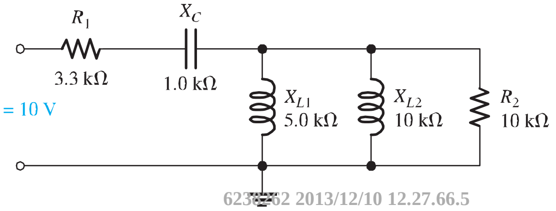

6238262 2013/12/10 12.27.66.5

32. Determine whether there is a value of C that will make $V_{ab} = 0$ V in Figure 13–75. If not, explain.

FIGURE 13–75

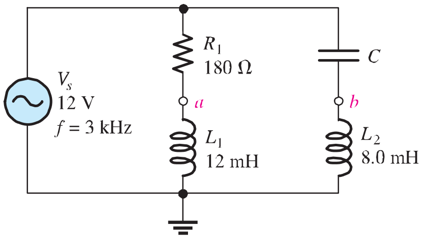

33. If the value of C is 0.22 μF, how much current is through each branch in Figure 13–75? What is the total current?

34. **(a)** Determine the frequency of oscillator-1 for the circuit in Figure 13–52. Assume L_1 is 300 μH.
(b) What is the cutoff frequency for the low-pass filter in Figure 13–52?

35. Design a band-pass filter using a parallel resonant circuit to meet the following specifications: $BW = 500$ Hz, $Q = 40$, $I_{C(max)} = 20$ mA, $V_{C(max)} = 2.5$ V.

36. Design a circuit in which the following series resonant frequencies are switch-selectable: 500 kHz, 1000 kHz, 1500 kHz, 2000 kHz.

37. Design a parallel-resonant network using a single coil and switch-selectable capacitors to produce the following resonant frequencies: 8 MHz, 9 MHz, 10 MHz, and 11 MHz. Assume a 10 μH coil with a winding resistance of 5 Ω.

MULTISIM

MULTISIM TROUBLESHOOTING PROBLEMS

38. Open file P13-38; files are found at www.pearsonhighered.com/floyd. Determine if there is a fault and, if so, identify it.

39. Open file P13-39. Determine if there is a fault and, if so, identify it.

40. Open file P13-40. Determine if there is a fault and, if so, identify it.

41. Open file P13-41. Determine if there is a fault and, if so, identify it.

42. Open file P13-42. Determine if there is a fault and, if so, identify it.

43. Open file P13-43. Determine if there is a fault and, if so, identify it.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

ANSWERS TO SECTION CHECKUPS

SECTION 13–1 Impedance and Phase Angle of Series *RLC* Circuits

1. The circuit is inductive if $X_L > X_C$ and capacitive if $X_C > X_L$.
2. $X_{tot} = |X_L - X_C| = 70\ \Omega$; The circuit is capacitive.
3. $Z = \sqrt{R^2 + X_{tot}^2} = 83.2\ \Omega$; $\theta = \tan^{-1}(X_{tot}/R) = 57.3°$; Current is leading V_s.

SECTION 13–2 Analysis of Series *RLC* Circuits

1. $V_s = \sqrt{V_R^2 + (V_C - V_L)^2} = 38.4\ V$
2. Current leads V_s because the circuit is capacitive.
3. $X_{tot} = |X_L - X_C| = 6\ \text{k}\Omega$

SECTION 13–3 Series Resonance

1. Series resonance occurs when $X_L = X_C$.
2. The current is maximum because the impedance is minimum.
3. $f_r = 1/(2\pi\sqrt{LC}) = 159\ \text{kHz}$
4. It is capacitive because $X_C > X_L$.

SECTION 13–4 Series Resonant Filters

1. $V_{out} = 0.707(15\ \text{V}) = 10.6\ \text{V}$
2. $BW = f_r/Q = 10\ \text{kHz}$
3. Current is maximum; output voltage is minimum.

SECTION 13–5 Parallel *RLC* Filters

1. $I_R = V_s/R = 8.0\ \text{mA}$; $I_C = V_s/X_C = 12.0\ \text{mA}$; $I_L = V_s/X_L = 24.0\ \text{mA}$
2. The circuit is inductive ($X_L < X_C$).
3. $L_{\text{eq}} = L[(Q^2 + 1)/Q^2] = 20.1\ \text{mH}$; $R_{p(\text{eq})} = R_W(Q^2 + 1) = 1589\ \Omega$

SECTION 13–6 Parallel Resonance

1. The impedance is maximum.
2. The current is minimum.
3. At ideal parallel resonance, $X_C = X_L = 1.5\ \text{k}\Omega$.
4. $f_r = \sqrt{1 - (R_W^2 C/L)}/2\pi\sqrt{LC} = 33.7\ \text{kHz}$; $Z_r = R_W(Q^2 + 1) = 440\ \Omega$
5. $f_r = 1/(2\pi\sqrt{LC}) = 22.5\ \text{kHz}$
6. $f_r = \sqrt{Q^2/(Q^2 + 1)}/2\pi\sqrt{LC} = 20.9\ \text{kHz}$
7. $Z_r = R_W(Q^2 + 1) = 8.02\ \text{k}\Omega$

SECTION 13–7 Parallel Resonant Filters

1. The bandwidth can be increased by reducing the parallel resistance.
2. f_r changes to 4.47 kHz.
3. $Z_r = R_W(Q^2 + 1) = 47.0\ \text{k}\Omega$

SECTION 13–8 Resonant Circuit Applications

1. A tuned resonant circuit is used to select a narrow band of frequencies.
2. The same tuned circuits can be used no matter what station is selected.
3. Several variable capacitors (or inductors) whose values can be varied simultaneously with a common control is an example of ganged tuning.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

ANSWERS TO RELATED PROBLEMS FOR EXAMPLES

13–1. 1.25 kΩ; 63.4°

13–2. 4.71 kΩ

13–3. Current increases, reaches a maximum at resonance, and then decreases.

13–4. More capacitive

13–5. 67.3 kHz

13–6. 2.27 mA

13–7. Z increases; Z increases

13–8. No

13–9. 200 kHz

13–10. −14 dB

13–11. 7.48

13–12. 322 Hz

13–13. V_{out} increases; V_{out} increases

13–14. Inductive

13–15. Increase. X_C will decrease and approach 0.

13–16. $R_{p(\text{eq})} = 25$ kΩ; $L_{\text{eq}} = 5$ mH; $C = 0.022$ μF, $R_{p(tot)} = 4.24$ kΩ

13–17. $I_L = 0.996$ mA, $I_C = 1.06$ mA

13–18. 80.0 kΩ

13–19. The differences are negligible.

13–20. 7.07 V

13–21. 530 kHz

13–22. 25

13–23. A larger-value load resistor has less effect on Q_O.

13–24. 185 mV

ANSWERS TO TRUE/FALSE QUIZ

1. T **2.** F **3.** T **4.** T **5.** F **6.** T **7.** T **8.** F **9.** F **10.** F

ANSWERS TO SELF-TEST

1. (a) **2.** (c) **3.** (b) **4.** (c) **5.** (d) **6.** (c)
7. (a) **8.** (b) **9.** (d) **10.** (a) **11.** (b) **12.** (d)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

CHAPTER 14

TRANSFORMERS

6238262 2013/12/10 12.27.66.5

OUTLINE

14–1 Mutual Inductance
14–2 The Basic Transformer
14–3 Step-Up and Step-Down Transformers
14–4 Loading the Secondary
14–5 Reflected Load
14–6 Impedance Matching
14–7 Transformer Ratings and Characteristics
14–8 Tapped and Multiple-Winding Transformers
14–9 Troubleshooting

OBJECTIVES

- Explain mutual inductance
- Describe how a transformer is constructed and how it works
- Describe how transformers increase and decrease voltage
- Discuss the effect of a resistive load across the secondary
- Discuss the concept of a reflected load in a transformer
- Discuss impedance matching with transformers
- Describe practical transformer ratings
- Describe several types of transformers
- Troubleshoot transformers

KEY TERMS

Electrical isolation
Mutual inductance (L_M)
Transformer
Primary winding
Secondary winding
Magnetic coupling
Turns ratio (n)
Reflected resistance
Impedance matching
Apparent power rating
Center tap

INTRODUCTION

In Chapter 11, you learned about self-inductance. In this chapter, you will study mutual inductance, which is the basis for the operation of transformers. Transformers are an important component in many systems, such as power supplies, electrical power distribution, and signal coupling in communications systems.

The operation of the transformer is based on the principle of mutual inductance, which occurs when two or more coils are in close proximity. A simple transformer is actually two coils that are electromagnetically coupled by their mutual inductance. Because there is no electrical contact between two magnetically coupled coils, the transfer of energy from one coil to the other can be achieved in a situation of complete electrical isolation. In relation to transformers, the term *winding* or *coil* is commonly used to describe the primary or secondary.

VISIT THE WEBSITE
Study aids for this chapter are available at
http://pearsonhighered.com/floyd

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

14–1 MUTUAL INDUCTANCE

When two coils are placed close to each other, a changing electromagnetic field produced by the current in one coil will cause an induced voltage in the second coil because of the mutual inductance between the two coils.

After completing this section, you should be able to

- **Explain mutual inductance**
  - **Discuss magnetic coupling**
  - **Define *electrical isolation***
  - **Define *coefficient of coupling***
  - **Identify the factors that affect mutual inductance and state the formula**

Recall that the electromagnetic field surrounding a coil of wire expands, collapses, and reverses as the current increases, decreases, and reverses. When a second coil is placed very close to the first coil so that the changing magnetic lines of force cut through the second coil, the coils are magnetically coupled and a voltage is induced, as indicated in Figure 14–1.

FIGURE 14–1 **A voltage is induced in the second coil as a result of the changing current in the first coil, producing a changing magnetic field that links the second coil.**

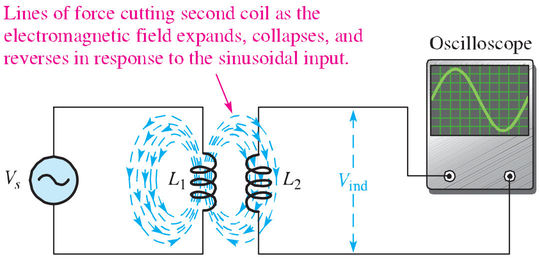
Lines of force cutting second coil as the electromagnetic field expands, collapses, and reverses in response to the sinusoidal input.
Oscilloscope
V_s
L_1
L_2
V_{ind}

When two coils are magnetically coupled, they provide electrical isolation because there is no electrical connection between them, only a magnetic link. **Electrical isolation** is the condition in which two circuits have no common conductive path between them. If the current in the first coil is a sine wave, the voltage induced in the second coil is also a sine wave. The amount of voltage induced in the second coil as a result of the current in the first coil is dependent on the **mutual inductance, L_M**.

The mutual inductance is established by the inductance of each coil (L_1 and L_2) and by the amount of coupling (k) between the two coils. To maximize coupling, the two coils are wound on a common core. The three factors that influence mutual inductance (k, L_1, and L_2) are shown in Figure 14–2. The formula for mutual inductance is

$$L_M = k\sqrt{L_1 L_2} \qquad (14–1)$$

FIGURE 14–2 **The mutual inductance of two coils.**

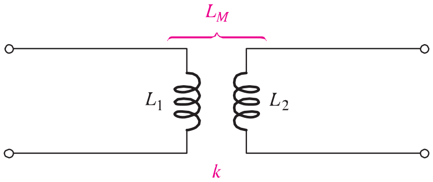

Coefficient of Coupling

The **coefficient of coupling, k**, between the two coils of a transformer is the ratio of the flux (lines of force) produced by coil 1 linking coil 2 ($\phi_{1\text{-}2}$) to the total flux produced by coil 1 (ϕ_1).

$$k = \frac{\phi_{1\text{-}2}}{\phi_1} \qquad (14–2)$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

For example, if half of the total flux produced by coil 1 links coil 2, then $k = 0.5$. A greater value of k means that more voltage is induced in coil 2 for a certain rate of change of current in coil 1. Note that k has no units. Recall that the unit of magnetic lines of force (flux) is the weber, abbreviated Wb.

The coefficient of coupling, k, depends on the physical proximity of the coils and the type of core material on which they are wound. Also, the construction and shape of the cores are factors.

EXAMPLE 14–1

Two coils are wound on a single core, and the coefficient of coupling is 0.3. The inductance of coil 1 is 10 μH, and the inductance of coil 2 is 15 μH. What is L_M?

SOLUTION

$$L_M = k\sqrt{L_1 L_2} = 0.3\sqrt{(10\ \mu\text{H})(15\ \mu\text{H})} = \mathbf{3.67\ \mu H}$$

RELATED PROBLEM*

Determine the mutual inductance when $k = 0.5$, $L_1 = 1$ mH, and $L_2 = 600$ μH.

***Answers are at the end of the chapter.**

EXAMPLE 14–2

One coil produces a total magnetic flux of 50 μWb, and 20 μWb link coil 2. What is k?

SOLUTION

$$k = \frac{\phi_{1\text{-}2}}{\phi_1} = \frac{20\ \mu\text{Wb}}{50\ \mu\text{Wb}} = \mathbf{0.4}$$

RELATED PROBLEM

Determine k when $\phi_1 = 500$ μWb and $\phi_{1\text{-}2} = 375$ μWb.

SECTION 14–1 CHECKUP*

1. Define *mutual inductance.*
2. Two 50 mH coils have $k = 0.9$. What is L_M?
3. If k is increased, what happens to the voltage induced in one coil as a result of a current change in the other coil?

***Answers are at the end of the chapter.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

14–2 THE BASIC TRANSFORMER

A transformer is an electrical device constructed of two or more coils of wire (windings) electromagnetically coupled to each other so that there is a mutual inductance for the transfer of power from one winding to the other. Although many transformers have more than two windings, we will restrict our discussion in this section to a basic two-winding transformer. Later, more complicated transformers are introduced.

After completing this section, you should be able to

6238262 2013/12/10 12.27.66.5

- **Describe how a transformer is constructed and how it works**
  - **Identify the parts of a basic transformer**
  - **Discuss the importance of the core material**
  - **Define *primary winding* and *secondary winding***
  - **Define *turns ratio***
  - **Discuss how the direction of windings affects voltage polarities**

A schematic of a transformer is shown in Figure 14–3(a). One coil is called the **primary winding** and the other coil is called the **secondary winding** as indicated. For standard operation, the source voltage is applied to the primary winding, and a load is connected to the secondary winding, as shown in Figure 14–3(b). The primary winding is the input winding, and the secondary winding is the output winding. It is common to refer to the side of the transformer that has the source voltage as the *primary*, and the side that has the induced voltage as the *secondary*.

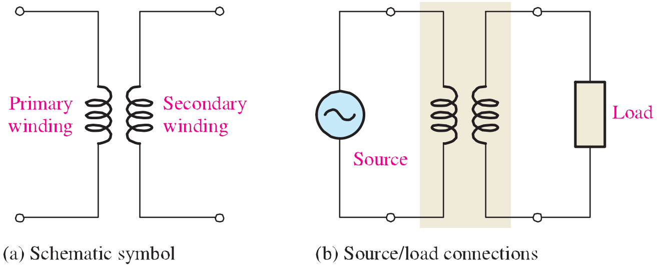

(a) Schematic symbol (b) Source/load connections

FIGURE 14–3 **The basic transformer.**

The windings of a transformer are formed around the core. The core provides both a physical structure for placement of the windings and a magnetic path so that the magnetic flux is concentrated close to the coils. Three general categories of core material are air, ferrite, and iron. The schematic symbol for each type is shown in Figure 14–4.

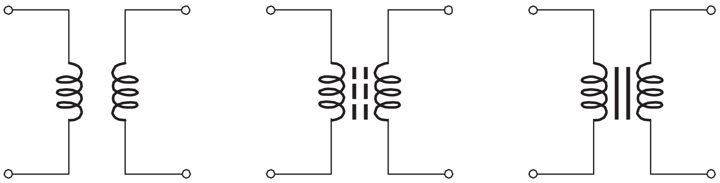

(a) Air core (b) Ferrite core (c) Iron core

FIGURE 14–4 **Schematic symbols specify the type of core.**

Iron-core transformers generally are used for audio frequency (AF) and power applications. These transformers consist of windings on a core constructed from laminated sheets of ferromagnetic material insulated from each other, as shown in Figure 14–5. This

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

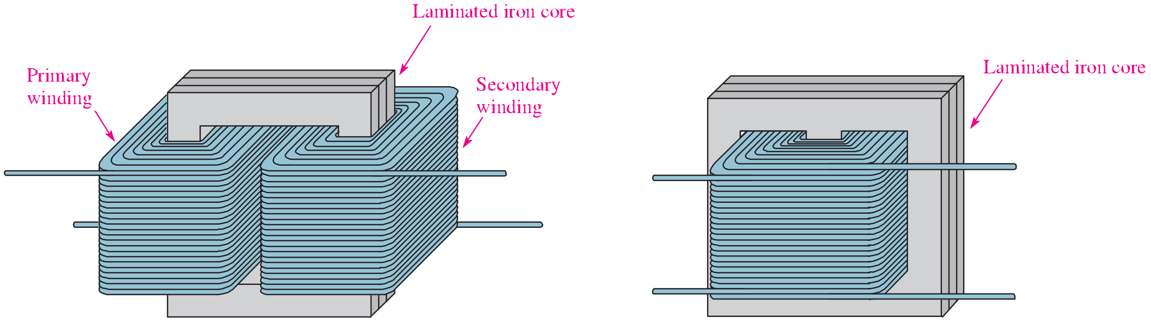

(a) Core type has each winding on a separate leg.

(b) Shell type has both windings on the same leg.

FIGURE 14–5 **Iron-core transformer construction with multilayer windings.**

construction provides an easy path for the magnetic flux and increases the amount of coupling between the windings. The figure also shows the basic construction of two major configurations of iron-core transformers. In the core-type construction, shown in Figure 14–5(a), the windings are on separate legs of the laminated core. In the shell-type construction, shown in part (b), both windings are on the same leg. Each type has certain advantages. In general, the core type has more room for insulation and can handle higher voltages. The shell type can produce higher core flux, so fewer turns are required.

Air-core and ferrite-core transformers generally are used for high-frequency applications and consist of windings on an insulating shell that is hollow (air) or constructed of ferrite, such as depicted in Figure 14–6. The wire is typically covered by a varnish-type coating to prevent the windings from shorting together. The amount of **magnetic coupling** between the primary winding and the secondary winding is set by the type of core material and by the relative positions of the windings. In Figure 14–6(a), the windings are loosely coupled because they are separated, and in part (b) they are tightly coupled because they are overlapping. The tighter the coupling, the greater the induced voltage in the secondary for a given current in the primary. Another common high-frequency transformer is the intermediate frequency (IF) transformer shown in Figure 14–6(c). The ferrite core can be adjusted with a screw that is used to tune the transformer to a specific frequency.

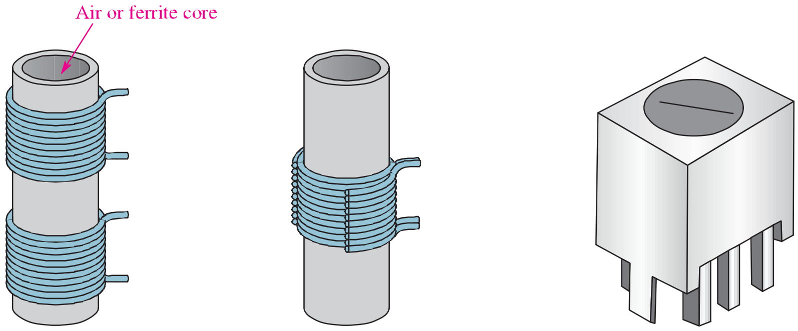

(a) Loosely coupled windings

(b) Tightly coupled windings. Cutaway view shows both windings.

(c) IF transformer in metal can

FIGURE 14–6 **Transformers with cylindrical-shaped cores.**

High-frequency transformers tend to have fewer windings and smaller inductances than power transformers. A type of high-frequency transformer that has become popular in recent years is the planar transformer. A representative planar transformer is shown in Figure 14–7(a). Planar transformers are constructed with printed circuit (PC) board assembly methods (rather than wire winding) that enable them to be produced

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

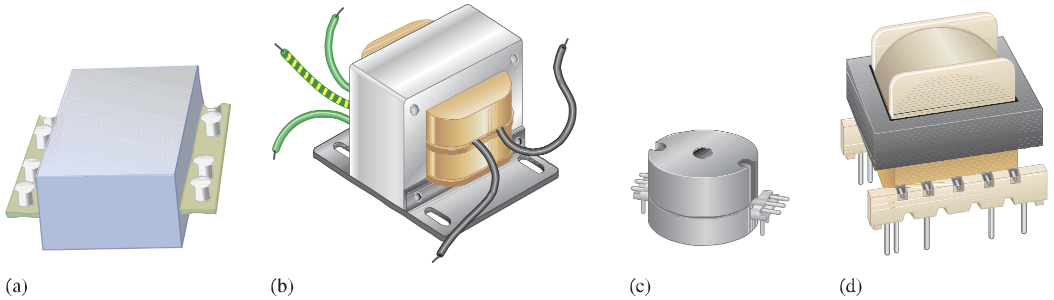

FIGURE 14–7 Some common types of transformers.

with high precision and low cost. The windings are actually traces laid out on stacked PC boards. Planar transformers are available in a variety of sizes and wattage ratings. The low profile of a planar transformer (typically less than 0.5 inches) makes it particularly suited to cases where space is critical. Figure 14–7(b) shows a low-voltage transformer commonly used in power supplies. Parts (c) and (d) show other common types of small transformers.

Intermediate Frequency Transformers

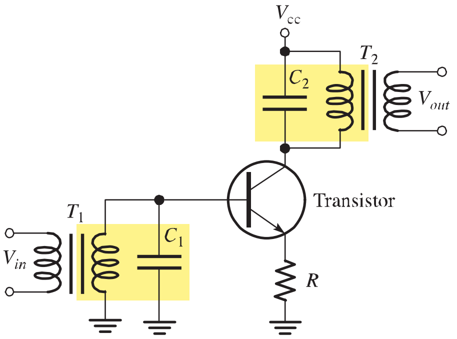

Communication systems and certain measurement systems use tuned transformers to couple a high-frequency signal from one stage to another. The signal is generally a predetermined frequency called the intermediate frequency (IF), which is a fixed frequency used in the system. The transistor circuit shown here is an intermediate frequency amplifier with parallel resonant tuned circuits shown in the yellow highlighted boxes on the input (C_1 and the secondary of T_1) and the output (C_2 and the primary of T_2). This causes the amplifier to selectively amplify the intermediate frequency.

SYSTEM NOTE

Turns Ratio

A transformer parameter that is useful in understanding how a transformer operates is the turns ratio. In this text, the **turns ratio, (n)** is defined as the ratio of the number of turns in the secondary winding (N_{sec}) to the number of turns in the primary winding (N_{pri}).

$$n = \frac{N_{sec}}{N_{pri}} \qquad (14\text{–}3)$$

This definition of turns ratio is based on the IEEE standard for electronics power transformers as specified in the IEEE dictionary. Other categories of transformer may have a different definition, so some sources define the turns ratio as N_{pri}/N_{sec}. Either definition is correct as long as it is clearly stated and used consistently. The turns ratio of a transformer is rarely if ever given as a transformer specification. Generally, the input and output voltages and the power rating are the key specifications. However, the turns ratio is useful in studying the operating principle of a transformer.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 14–3

A certain transformer used in a radar system has a primary winding with 100 turns and a secondary winding with 400 turns. What is the turns ratio?

SOLUTION

$N_{sec} = 400$ and $N_{pri} = 100$; therefore, the turns ratio is

$$n = \frac{N_{sec}}{N_{pri}} = \frac{400}{100} = \mathbf{4}$$

A turns ratio of 4 can be expressed as 1:4 on a schematic.

RELATED PROBLEM

A certain transformer has a turns ratio of 10. If $N_{pri} = 500$, what is N_{sec}?

Direction of Windings

Another important transformer parameter is the direction in which the windings are placed around the core. As illustrated in Figure 14–8, the direction of the windings determines the polarity of the voltage across the secondary winding (secondary voltage) with respect to the voltage across the primary winding (primary voltage). Phase dots can be used on the schematic symbols to indicate polarities, as shown in Figure 14–9.

6238262 2013/12/10 12.27.66.5

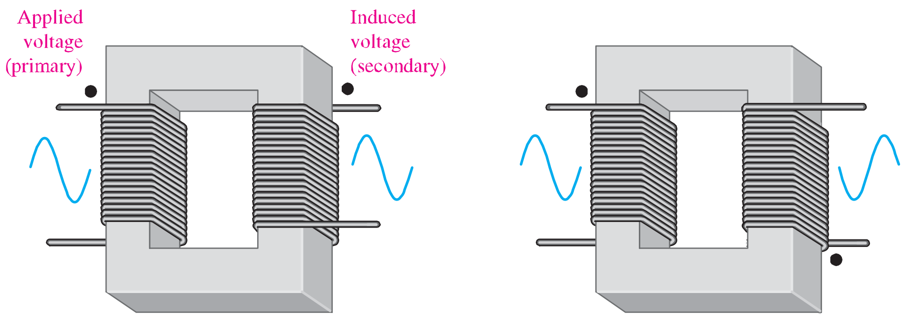

(a) The primary and secondary voltages are in phase when the windings are in the same effective direction around the magnetic path.

(b) The primary and secondary voltages are 180° out of phase when the windings are in the opposite direction.

FIGURE 14–8 **The direction of the windings determines the relative polarities of the voltages.**

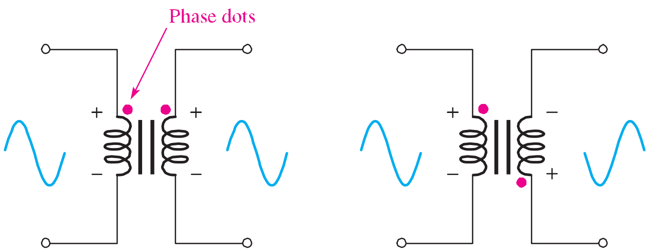

(a) Voltages are in phase.

(b) Voltages are out of phase.

FIGURE 14–9 **Phase dots indicate corresponding polarities of primary and secondary voltages.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 14–2 CHECKUP

1. Upon what principle is the operation of a transformer based?
2. Define *turns ratio*.
3. Why are the directions of the windings of a transformer important?
4. A certain transformer has a primary winding with 500 turns and a secondary winding with 250 turns. What is the turns ratio?
5. How do the windings in a planar transformer differ from other transformers?

14–3 STEP-UP AND STEP-DOWN TRANSFORMERS

A step-up transformer has more turns in its secondary winding than in its primary winding and is used to increase ac voltage. A step-down transformer has more turns in its primary winding than in its secondary winding and is used to decrease ac voltage.

After completing this section, you should be able to

- **Describe how transformers increase and decrease voltage**
  - **Explain how a step-up transformer works**
  - **Identify a step-up transformer by its turns ratio**
  - **State the relationship between primary and secondary voltages and the turns ratio**
  - **Explain how a step-down transformer works**
  - **Identify a step-down transformer by its turns ratio**
  - **Explain the purpose of ac line conditioning**

The Step-Up Transformer

A transformer in which the secondary voltage is greater than the primary voltage is called a **step-up transformer**. The amount that the voltage is stepped up depends on the turns ratio. For any transformer,

> **The ratio of secondary voltage (V_{sec}) to primary voltage (V_{pri}) is equal to the ratio of the number of turns in the secondary winding (N_{sec}) to the number of turns in the primary winding (N_{pri}).**

$$\frac{V_{sec}}{V_{pri}} = \frac{N_{sec}}{N_{pri}} \qquad (14\text{–}4)$$

Recall that N_{sec}/N_{pri} defines the turns ratio, n. Therefore, from this relationship, V_{sec} can be expressed as

$$V_{sec} = nV_{pri} \qquad (14\text{–}5)$$

Equation 14–5 shows that the secondary voltage is equal to the turns ratio times the primary voltage. This condition assumes that the coefficient of coupling is 1, and a good iron-core transformer approaches this value.

The turns ratio for a step-up transformer is always greater than 1 because the number of turns in the secondary winding (N_{sec}) is always greater than the number of turns in the primary winding (N_{pri}).

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 14–4

The transformer in Figure 14–10 has a turns ratio of 3. What is the voltage across the secondary? Voltages are in rms unless otherwise stated.

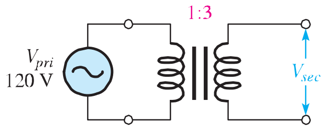

FIGURE 14–10

SOLUTION

The secondary voltage is

$$V_{sec} = nV_{pri} = 3(120\text{ V}) = \mathbf{360\ V}$$

Note that the turns ratio of 3 is indicated on the schematic as 1:3, meaning that there are three secondary turns for each primary turn.

RELATED PROBLEM

The transformer in Figure 14–10 is changed to one with a turns ratio of 4. Determine V_{sec}.

MULTISIM

Open Multisim file E14-04; files are found at www.pearsonhighered.com/floyd. Measure the secondary voltage.

The Step-Down Transformer

A transformer in which the secondary voltage is less than the primary voltage is called a **step-down transformer**. The amount by which the voltage is stepped down depends on the turns ratio. Equation 14–5 also applies to a step-down transformer.

The turns ratio of a step-down transformer is always less than 1 because the number of turns in the secondary winding (N_{sec}) is always less than the number of turns in the primary winding (N_{pri}).

EXAMPLE 14–5

The transformer in Figure 14–11 is part of a laboratory power supply and has a turns ratio of 0.2. What is the secondary voltage?

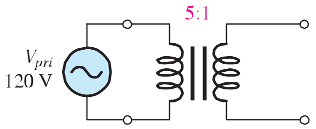

FIGURE 14–11

SOLUTION

The secondary voltage is

$$V_{sec} = nV_{pri} = 0.2(120\text{ V}) = \mathbf{24\ V}$$

RELATED PROBLEM

The transformer in Figure 14–11 is changed to one with a turns ratio of 0.48. Determine the secondary voltage.

MULTISIM

Open Multisim file E14-05 and measure the secondary voltage.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SYSTEM EXAMPLE 14–1

A BASIC POWER SUPPLY

One of the most important applications for transformers is in power supplies. Power supplies convert ac to dc, which is needed by nearly every electronic system. The transformer converts the ac to a higher or lower value before the next step of changing it to pulsating dc, which is then smoothed and regulated in the final step. Many systems use low-voltage dc for control circuits.

The basic power supply shown in Figure 14–12 provides a constant 5 V dc output over a range of loads and variations in the line voltage. This is accomplished by the 7805 voltage regulator, an integrated circuit regulator that senses the output, compares it to a reference, and adjusts it to keep it constant. The 7805 is designed for positive 5.0 V dc output.

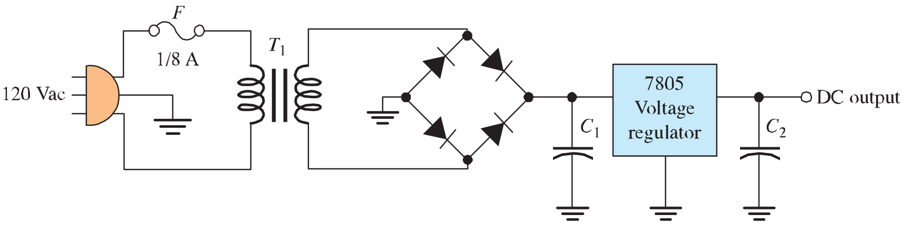

FIGURE 14–12 **A basic 5.0 V dc power supply.**

The operation of the power supply is as follows. The plug is connected to a wall socket that connects 120 V ac to the input of the transformer through a series fuse. The fuse is rated for a much smaller current than the maximum output current because the transformer will step *down* the voltage but step *up* the current. For the basic power supply here, the transformer output is about 6 V rms (8.5 V peak). The four diodes, which are arranged in a bridge configuration, change the voltage to a pulsating dc, which is smoothed by C_1. The 7805 regulator provides additional smoothing and keeps the output nearly constant. Final smoothing and filtering are provided by C_2.

If you work on any power supply, be aware that there are dangerous voltages present when a circuit is connected to ac; even low-voltage supplies can be lethal when they are connected to ac. In the case of high-voltage supplies, extra caution should be taken; even circuits that have been turned off remain dangerous because capacitors can retain a significant charge. Precautions for working safely around voltage sources and high energy sources (like batteries) are discussed in Chapter 1, Section 1–8, and should always be kept in mind.

DC Isolation

If there is dc through the primary circuit of a transformer, nothing happens in the secondary circuit, as indicated in Figure 14–13(a). A changing current in the primary winding is necessary in order to create a changing magnetic field. This will cause voltage to be induced in the secondary circuit, as indicated in Figure 14–13(b). Therefore, the transformer isolates

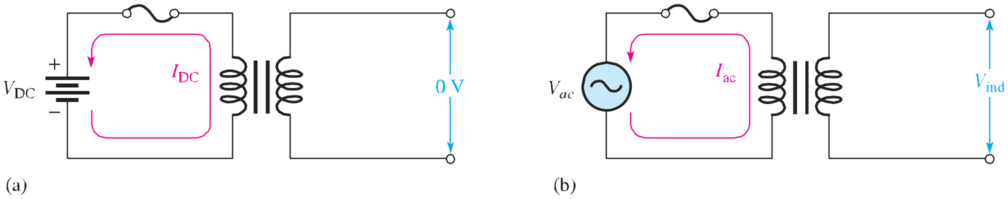

FIGURE 14–13 **DC isolation and ac coupling.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

the secondary circuit from any dc in the primary circuit. A transformer that is used strictly for isolation has a turns ratio of 1.

Isolation transformers are often packaged as part of a total ac line-conditioning device. In addition to the isolation transformer, the line conditioning includes surge protection, filters to eliminate interference, and sometimes automatic voltage regulation. Line conditioning is useful to isolate sensitive equipment such as microprocessor-based controllers. Specialized line conditioners for hospitals for patient-monitoring equipment provide a high degree of electrical isolation and protection from shock.

Small transformers that are used to isolate the dc bias from one stage of an amplifier to next stage are called coupling transformers because the ac signal is passed ("coupled") but the dc is blocked. Coupling transformers are widely used at high frequencies where they are designed to pass only a selected band of frequencies by making the primary and secondary coils part of a parallel resonant circuit. (Resonant circuits were discussed in Chapter 13.) A typical coupling transformer arrangement is shown in Figure 14–14 where the transformer is part of a resonant circuit on the input and output. Frequently, the core of a coupling transformer can be adjusted to fine-tune the frequency response. At audio frequencies, coupling transformers are primarily used to couple a signal from an amplifier to a speaker and to provide maximum power transfer. This type of transformer is called an impedance-matching transformer and is discussed in Section 14–6.

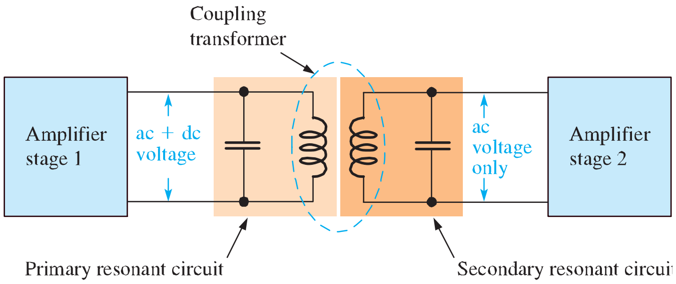

FIGURE 14–14 **A coupling transformer used to pass high frequencies in the band determined by the resonant circuit. The dc is not passed to the secondary.**

SECTION 14–3 CHECKUP

1. What does a step-up transformer do?
2. If the turns ratio is 5, how much greater is the secondary voltage than the primary voltage?
3. When 240 V ac are applied to a transformer with a turns ratio of 10, what is the secondary voltage?
4. What does a step-down transformer do?
5. A voltage of 120 V ac is applied to the primary winding of a transformer with a turns ratio of 0.5. What is the secondary voltage?
6. A primary voltage of 120 V ac is reduced to 12 V ac. What is the turns ratio?
7. What are typical features provided in an ac line conditioner?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

14–4 LOADING THE SECONDARY

When a resistive load is connected to the secondary winding of a transformer, the relationship of the load (secondary) current and the current in the primary circuit is determined by the turns ratio.

After completing this section, you should be able to

- **Discuss the effect of a resistive load across the secondary**
  - **Discuss power in a transformer**
  - **Determine the current delivered by the secondary when a step-up transformer is loaded**
  - **Determine the current delivered by the secondary when a step-down transformer is loaded**

If a transformer is operated without a load, the primary acts like an inductor. Ideally, the current in an inductor will lag the voltage by 90° and the power factor is 0. When a resistive load is connected to the secondary of the transformer, the primary no longer acts as an ideal inductor. The phase angle decreases and the power factor increases. The load has caused the primary to look resistive because the primary current and voltage will be nearly in phase. For purposes of discussion, we will assume this ideal condition when the transformer is loaded with a resistive load.

The power delivered to a load can never be greater than the power delivered to the primary winding. For an ideal transformer, the power delivered by the secondary winding (P_{sec}) equals the power delivered to the primary winding (P_{pri}). When losses are considered, the power delivered by the secondary is always less.

Power is dependent on voltage and current, and there can be no increase in power in a transformer. Therefore, if the voltage is stepped up, the current is stepped down and vice versa. In an ideal transformer, the power delivered by the secondary to the load is the same as the power delivered to the primary regardless of the turns ratio.

The power delivered to the primary is

$$P_{pri} = V_{pri}I_{pri}$$

The power delivered by the secondary is

$$P_{sec} = V_{sec}I_{sec}$$

Ideally, $P_{pri} = P_{sec}$; therefore,

$$V_{pri}I_{pri} = V_{sec}I_{sec}$$

Transposing terms,

$$\frac{I_{pri}}{I_{sec}} = \frac{V_{sec}}{V_{pri}}$$

From Equation 14–4,

$$\frac{V_{sec}}{V_{pri}} = \frac{N_{sec}}{N_{pri}}$$

Therefore, since N_{sec}/N_{pri} equals the turns ratio, n, the relationship of primary current to secondary current in a transformer is

$$\boldsymbol{\frac{I_{pri}}{I_{sec}} = n} \tag{14–6}$$

Inverting both sides of Equation 14–6 and solving for I_{sec},

$$\boldsymbol{I_{sec} = \left(\frac{1}{n}\right)I_{pri}} \tag{14–7}$$

Figure 14–15 illustrates the effects of voltages and currents in a transformer. In part (a), for a step-up transformer in which n is greater than 1, the secondary current is less than the primary current because $1/n$ is less than 1. For a step-down transformer, shown in Figure 14–15(b), n is less than 1, and I_{sec} is greater than I_{pri} because $1/n$ is greater than 1.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

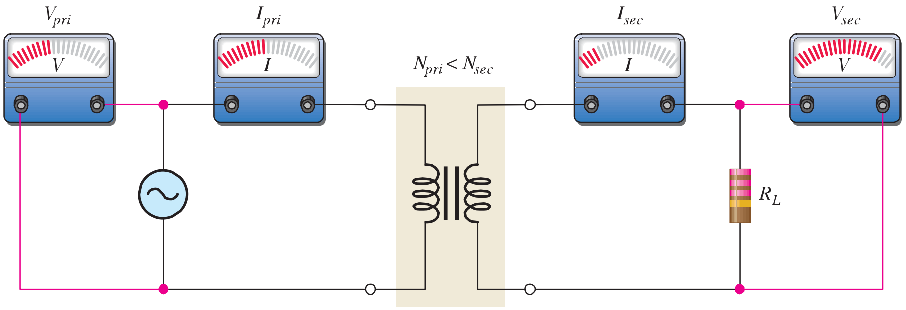

(a) Step-up transformer: $V_{sec} > V_{pri}$ and $I_{sec} < I_{pri}$

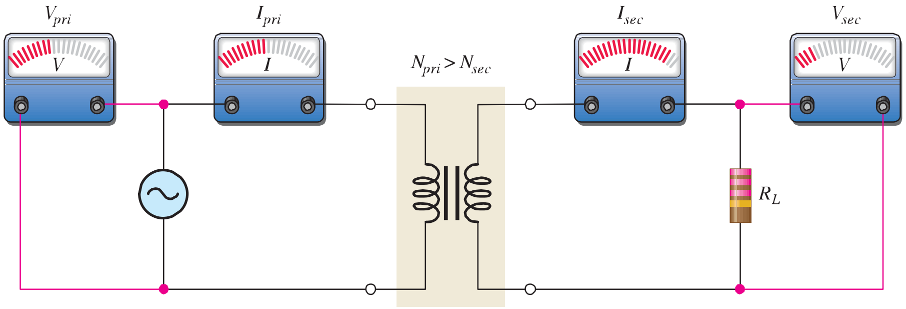

(b) Step-down transformer: $V_{sec} < V_{pri}$ and $I_{sec} > I_{pri}$

FIGURE 14–15 Illustration of voltages and currents in a transformer with a loaded secondary.

Current Transformers

A current transformer operates like a voltage transformer in that energy is transferred from the primary to the secondary through induction. Current transformers are often used to sense and monitor a large alternating current without having to open the line to make the measurement. One common application is to monitor current and supply the result to the watt-hour meter for customers with three-phase service or large-use single-phase customers.

One important consideration with current transformers is that the load must not be disconnected when current is in the primary. If this happens, the sudden opening will produce a very high voltage across the open secondary, leading to arcing and can permanently damage the transformer and equipment connected to it. In addition, it is a shock hazard to personnel working on it.

CT3
X
CT2
X
CT1
X

Photo courtesy of Thomas Kissell.

SYSTEM NOTE

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 14–6

The two ideal transformers shown in Figure 14–16 have loaded secondaries. If, as a result of the loaded secondary, the primary current is 100 mA in each case, what is the current through the load?

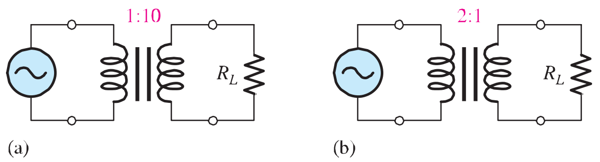

FIGURE 14–16

SOLUTION

In part (a), the turns ratio is 10. Therefore, the secondary load current is

$$I_L = I_{sec} = \left(\frac{1}{n}\right)I_{pri} = \left(\frac{1}{10}\right)I_{pri} = 0.1(100\text{ mA}) = \mathbf{10\text{ mA}}$$

In part (b), the turns ratio is 0.5. Therefore, the secondary load current is

$$I_L = I_{sec} = \left(\frac{1}{n}\right)I_{pri} = \left(\frac{1}{0.5}\right)I_{pri} = 2(100\text{ mA}) = \mathbf{200\text{ mA}}$$

RELATED PROBLEM

What is the secondary current in Figure 14–16(a) if the turns ratio is doubled? What is the secondary current in Figure 14–16(b) if the turns ratio is halved? Assume the load resistances are changed so that I_{pri} remains at 100 mA in both cases.

SECTION 14–4 CHECKUP

1. If the turns ratio of a transformer is 2, is the secondary current greater than or less than the primary current? By how much?
2. A transformer has 1000 primary turns and 250 secondary turns, and I_{pri} is 0.5 A. What is the turns ratio? What is the value of I_{sec}?
3. In Question 2, what is the primary current when there is a secondary load current of 10 A?
4. Why is it unsafe to disconnect the load on a current transformer when there is current in the primary?

14–5 REFLECTED LOAD

From the viewpoint of the primary, a load connected across the secondary winding of a transformer appears to have a resistance that is not necessarily equal to the actual resistance of the load. The actual load is "reflected" into the primary as determined by the turns ratio. This reflected load is what the source effectively sees, and it determines the amount of primary current.

After completing this section, you should be able to

- **Discuss the concept of a reflected load in a transformer**
  - **Define *reflected resistance***
  - **Explain how the turns ratio affects the reflected resistance**
  - **Calculate reflected resistance**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The concept of a reflected load is illustrated in Figure 14–17. The load (R_L) in the secondary of a transformer is reflected into the primary by transformer action. Ideally, the load appears to the source in the primary to be a resistance (R_{pri}) with a value determined by the turns ratio and the actual value of the load resistance. The resistance R_{pri} is called the **reflected resistance**.

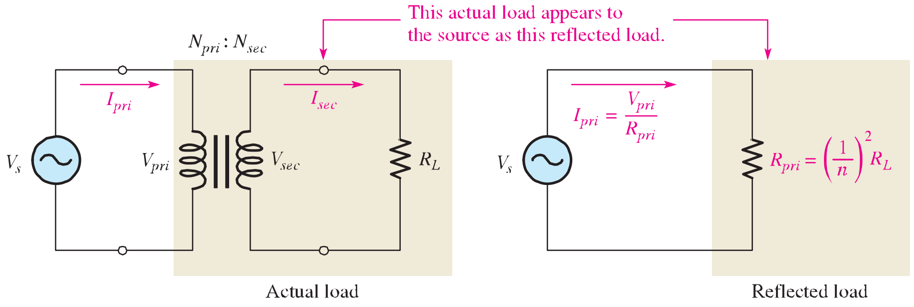

This actual load appears to the source as this reflected load.

FIGURE 14–17 **Reflected load in a transformer circuit.**

The resistance in the primary of Figure 14–17 is $R_{pri} = V_{pri}/I_{pri}$. The resistance in the secondary is $R_L = V_{sec}/I_{sec}$. From Equations 14–4 and 14–6, you know that $V_{sec}/V_{pri} = n$ and $I_{pri}/I_{sec} = n$. Using these relationships, a formula for R_{pri} in terms of R_L is determined as follows:

$$\frac{R_{pri}}{R_L} = \frac{V_{pri}/I_{pri}}{V_{sec}/I_{sec}} = \left(\frac{V_{pri}}{V_{sec}}\right)\left(\frac{I_{sec}}{I_{pri}}\right) = \left(\frac{1}{n}\right)\left(\frac{1}{n}\right) = \left(\frac{1}{n}\right)^2$$

Solving for R_{pri} yields

$$\boldsymbol{R_{pri} = \left(\frac{1}{n}\right)^2 R_L} \qquad (14\text{–}8)$$

Equation 14–8 shows that the resistance reflected into the primary is the square of the reciprocal of the turns ratio times the load resistance.

In a step-up transformer ($n > 1$), the reflected resistance is less than the actual load resistance; in a step-down transformer ($n < 1$), the reflected resistance is greater than the load resistance. This is illustrated in Examples 14–7 and 14–8, respectively.

EXAMPLE 14–7

Figure 14–18 shows a source that is transformer-coupled to a load resistor of 100 Ω. The transformer has a turns ratio of 4. What is the reflected resistance seen by the source?

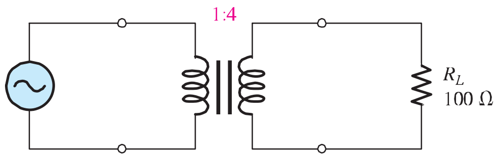

FIGURE 14–18

SOLUTION

The reflected resistance is

$$R_{pri} = \left(\frac{1}{n}\right)^2 R_L = \left(\frac{1}{4}\right)^2 100\ \Omega = \left(\frac{1}{16}\right)100\ \Omega = \mathbf{6.25\ \Omega}$$

The source sees a resistance of 6.25 Ω just as if it were connected directly, as shown in the equivalent circuit of Figure 14–19.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

FIGURE 14–19

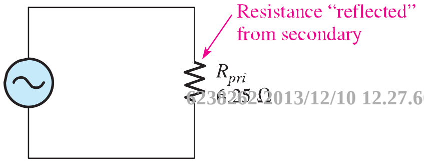

6238262 2013/12/10 12.27.66.5

RELATED PROBLEM

If the turns ratio in Figure 14–18 is 10 and R_L is 600 Ω, what is the reflected resistance?

EXAMPLE 14–8

In Figure 14–18, if a transformer that has a turns ratio of 0.25 is used, what is the reflected resistance?

SOLUTION

The reflected resistance is

$$R_{pri} = \left(\frac{1}{n}\right)^2 R_L = \left(\frac{1}{0.25}\right)^2 100\ \Omega = (4)^2 100\ \Omega = \mathbf{1600\ \Omega}$$

RELATED PROBLEM

To achieve a reflected resistance of 800 Ω, what turns ratio is required in Figure 14–18?

SECTION 14–5 CHECKUP

1. Define *reflected resistance.*
2. What transformer characteristic determines the reflected resistance?
3. A given transformer has a turns ratio of 10, and the load is 50 Ω. How much resistance is reflected into the primary?
4. What is the turns ratio required to reflect a 4 Ω load resistance into the primary as 400 Ω?

14–6 IMPEDANCE MATCHING

One application of transformers is in the matching of a load impedance to a source impedance to achieve maximum transfer of power or other results. This technique is called *impedance matching*. In audio systems, special wide-band transformers are often used to get the maximum amount of available power from the amplifier to the speaker by proper selection of the turns ratio. Transformers designed specifically for impedance matching usually show the input and output impedance they are designed to match.

After completing this section, you should be able to

- **Discuss impedance matching with transformers**
  - **Discuss the maximum power transfer theorem**
  - **Define *impedance matching***
  - **Explain the purpose of impedance matching**
  - **Describe a balun transformer**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Recall that the maximum power transfer theorem states that maximum power is transferred from a resistive source to a resistive load when the load resistance is equal to the source resistance (see Section 6–7). In ac circuits, the total opposition to current is called *impedance* and the process of making maximum power transfer between a source and load is called **impedance matching**. The simplest and most common case for impedance matching is when the source and load are both resistive, so we will restrict our discussion to this case.

Figure 14–20(a) shows an ac source with a fixed internal resistance. Some fixed internal resistance is inherent in all sources due to their internal circuitry. Part (b) shows a load connected to the source. In this case, the objective is often to transfer as much power to the load as possible.

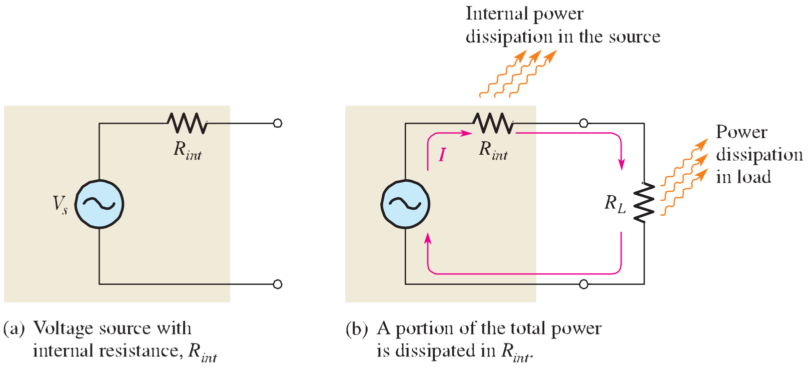
Internal power dissipation in the source
R_{int}
V_s
I
R_{int}
R_L
Power dissipation in load

(a) Voltage source with internal resistance, R_{int}

(b) A portion of the total power is dissipated in R_{int}.

FIGURE 14–20 **Power transfer from a nonideal voltage source to a load.**

In most practical situations, the internal resistance of various types of sources is fixed. Also, in many cases, the resistance of a device that acts like a load is fixed and cannot be altered. If you need to connect a given source to a given load, remember that only by chance will their resistances match. In this situation, a special type of wide-band transformer comes in handy. You can use the reflected-resistance characteristic provided by a transformer to make the load resistance appear to have the same resistance as the source resistance. This technique is called *impedance matching*, and the transformer is called an *impedance-matching transformer*.

Figure 14–21 illustrates a specific example of an impedance-matching transformer. In this example, the source resistance is driving a 300 Ω load. The impedance-matching transformer needs to make the load resistance look like a 75 Ω resistance to the source, thus delivering maximum power to the load. To select the right transformer, you need to know how the turns ratio affects the impedance. You can use Equation 14–8 to determine the turns ratio, n, when you know the value for R_L and R_{pri}.

$$R_{pri} = \left(\frac{1}{n}\right)^2 R_L$$

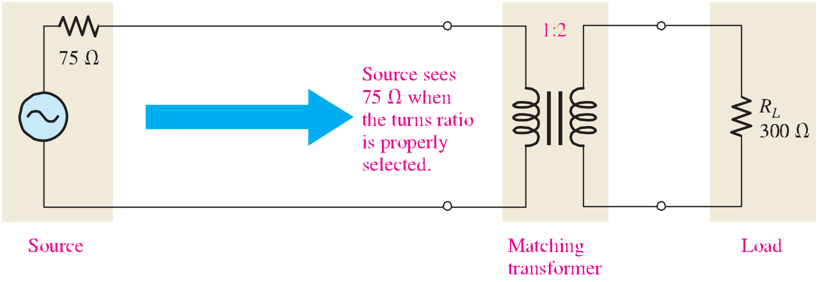

FIGURE 14–21 **Example of a load matched to a source by transformer coupling for maximum power transfer.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Transpose terms and divide both sides by R_L.

$$\left(\frac{1}{n}\right)^2 = \frac{R_{pri}}{R_L}$$

Then take the square root of both sides.

$$\frac{1}{n} = \sqrt{\frac{R_{pri}}{R_L}}$$

Invert both sides to get the following formula for the turns ratio:

$$\boldsymbol{n = \sqrt{\frac{R_L}{R_{pri}}}} \qquad (14\text{–}9)$$

Finally, solve for the particular turns ratio to match a 300 Ω load to a 75 Ω source.

$$n = \sqrt{\frac{300\ \Omega}{75\ \Omega}} = \sqrt{4} = 2$$

Therefore, a matching transformer with a turns ratio of 2 must be used in this application.

EXAMPLE 14–9

An amplifier has an 800 Ω internal resistance. In order to provide maximum power to an 8 Ω speaker, what turns ratio must be used in the coupling transformer?

SOLUTION

The reflected resistance must equal 800 Ω. Thus, from Equation 14–9, the turns ratio can be determined.

$$n = \sqrt{\frac{R_L}{R_{pri}}} = \sqrt{\frac{8\ \Omega}{800\ \Omega}} = \sqrt{0.01} = \mathbf{0.1}$$

The diagram and its equivalent reflected circuit are shown in Figure 14–22.

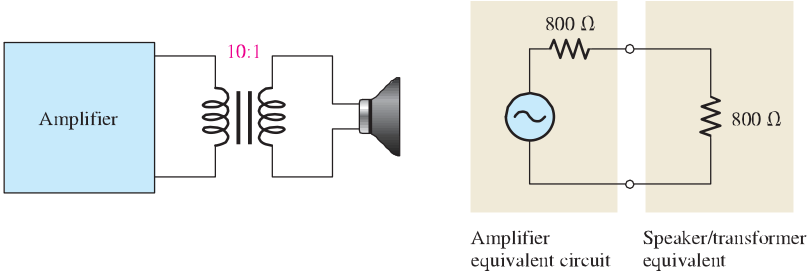

FIGURE 14–22

RELATED PROBLEM

What must be the turns ratio in Figure 14–22 to provide maximum power to two 8 Ω speakers in parallel?

THE BALUN TRANSFORMER An application for impedance matching is used in high-frequency antennas. In addition to impedance matching, many transmitting antennas also require that an unbalanced signal from the transmitter be converted to a balanced signal. A *balanced signal* is composed of two equal-amplitude signals that are 180° out-of-phase with each other. An *unbalanced signal* is one that is referenced to ground. A special type of transformer called a **balun**, a contraction of the words "*bal*anced-*un*balanced" is used to convert the unbalanced signal from the transmitter to a balanced signal at the antenna as shown in Figure 14–23.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

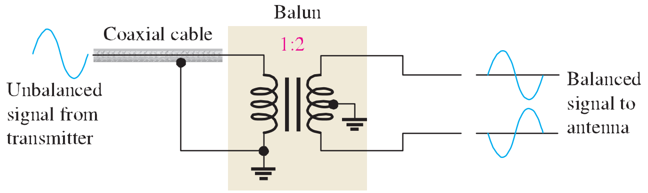
Balun
1:2
Coaxial cable
Unbalanced signal from transmitter
Balanced signal to antenna

FIGURE 14–23 **Illustration of a balun transformer converting an unbalanced signal to a balanced signal.**

The transmitter is usually connected to the balun with a coaxial cable (coax). A coax is basically a conductor surrounded by an insulating region and an outer conductor called the shield. In Figure 14–23, the signal from the transmitter is referenced to ground, which is connected to the outer shield of the coax, so it is an unbalanced signal. The shield on the coax minimizes pickup of radiated noise.

The coax has a certain characteristic impedance associated with it. The turns ratio of the balun is set to match the coax impedance to the antenna impedance. For example, if the transmitting antenna represents a 300 Ω impedance and the coax has a characteristic impedance of 75 Ω, the balun can provide the impedance match with a turns ratio of two, as was shown earlier. Baluns can also be used to convert from balanced signals to unbalanced signals.

SECTION 14–6 CHECKUP

1. What does impedance matching mean?
2. What is the advantage of matching the load resistance to the resistance of a source?
3. A transformer has a turns ratio of 0.5. What is the reflected resistance with 100 Ω across the secondary?
4. What is the purpose of a balun transformer?

14–7 TRANSFORMER RATINGS AND CHARACTERISTICS

Transformer operation has been discussed from an ideal point of view. That is, the winding resistance, the winding capacitance, and nonideal core characteristics were all neglected and the transformer was treated as if it had an efficiency of 100%. For studying the basic concepts and in many applications, the ideal model is valid. However, the practical transformer has several nonideal characteristics of which you should be aware.

After completing this section, you should be able to

- **Describe practical transformer ratings**
  - **List and describe the nonideal characteristics**
  - **Explain power rating of a transformer**
  - **Define *efficiency* of a transformer**

Ratings

POWER RATING A power transformer is typically rated in volt-amperes (VA), primary/secondary voltage, and operating frequency. For example, a given transformer rating may be specified as 2 kVA, 500/50, 60 Hz. The 2 kVA value is the **apparent power rating**. The 500 and the 50 can be either secondary or primary voltages. The 60 Hz is the operating frequency.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The transformer rating can be helpful in selecting the proper transformer for a given application. Let's assume, for example, that 50 V is the secondary voltage. In this case the load current is

$$I_L = \frac{P_{sec}}{V_{sec}} = \frac{2\text{ kVA}}{50\text{ V}} = 40\text{ A}$$

On the other hand, if 500 V is the secondary voltage, then

$$I_L = \frac{P_{sec}}{V_{sec}} = \frac{2\text{ kVA}}{500\text{ V}} = 4\text{ A}$$

These are the maximum currents that the secondary can handle in either case.

The reason that the power rating is in volt-amperes (apparent power) rather than in watts (true power) is as follows: If the transformer load is purely capacitive or purely inductive, the true power (watts) delivered to the load is ideally zero. However, the current for $V_{sec} = 500$ V and $X_C = 100\ \Omega$ at 60 Hz, for example, is 5 A. This current exceeds the maximum of 4 A that the 2 kVA secondary can handle, and even though the true power is zero, the transformer may be damaged. So it is meaningless to specify power in watts for transformers.

Resonant Charging

Resonant circuits can be used for charging by transformer action. Researchers are investigating wireless transfer of power by high-frequency magnetic coupling between two resonant coils. The idea dates back to Nicholas Tesla. It may be that electric cars would have a coil on board and when parked in the garage, would automatically be charged from a coil in the floor. Researchers still need to overcome efficiency problems, due to transformer losses, before it can be a viable charging method for electric cars.

SYSTEM NOTE

VOLTAGE AND FREQUENCY RATINGS In addition to the apparent power rating, most power transformers will have voltage and frequency ratings placed on the transformer. The voltage ratings include the primary voltage for which it is designed and the secondary voltage that is present when the rated load is connected to the secondary and the primary is connected to the rated input voltage. Frequently, there will be a small schematic drawing showing the windings and voltage rating for each winding. The frequency for which the transformer is designed will also be specified. A transformer operated on the wrong frequency can be damaged, so it important to note the frequency specification. These are minimum specifications you need to know in order to select a power transformer for an application.

Characteristics

WINDING RESISTANCE Both the primary and the secondary windings of a practical transformer have winding resistance. (You learned about the winding resistance of inductors in Chapter 11.) The winding resistances of a practical transformer are represented as resistors in series with the windings as shown in Figure 14–24.

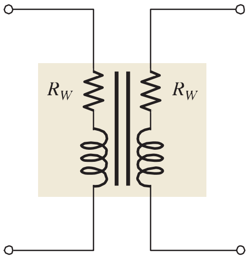

FIGURE 14–24 **Winding resistance in a practical transformer.**

Winding resistance in a practical transformer results in less voltage across a secondary load. Voltage drops due to the winding resistance effectively subtract from the primary and secondary voltages and result in a load voltage that is less than that predicted by the relationship $V_{sec} = nV_{pri}$. In most cases the effect is relatively small and can be neglected.

LOSSES IN THE CORE There is always some energy conversion in the core material of a practical transformer. This conversion is seen as a heating of ferrite and iron cores, but conversion does not occur in air cores. Part of this energy conversion is because of the continuous reversal of the magnetic field due to the changing direction of the primary

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

current; this component of the energy conversion is called *hysteresis loss.* The rest of the energy conversion to heat is caused by eddy currents produced when voltage is induced in the core material by the changing magnetic flux, according to Faraday's law. The eddy currents occur in circular patterns in the core resistance, thus producing heat. This conversion to heat is greatly reduced by the use of laminated construction of iron cores. The thin layers of ferromagnetic material are insulated from each other to minimize the buildup of eddy currents by confining them to a small area and to keep core losses to a minimum.

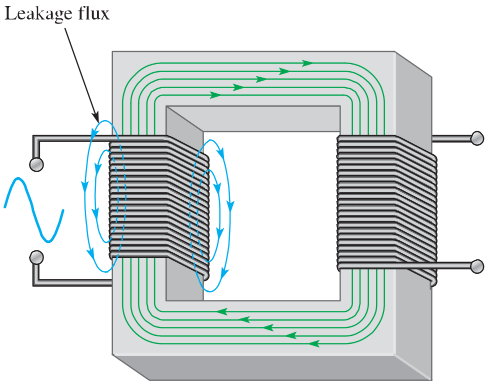

FIGURE 14–25 **Flux leakage in a practical transformer.**

MAGNETIC FLUX LEAKAGE In an ideal transformer, all the magnetic flux produced by the primary current is assumed to pass through the core to the secondary winding and vice versa. In a practical transformer, some of the magnetic flux lines break out of the core and pass through the surrounding air back to the other end of the winding, as illustrated in Figure 14–25 for the magnetic field produced by the primary current. Magnetic flux leakage results in a reduced secondary voltage.

The percentage of magnetic flux that actually reaches the secondary winding determines the coefficient of coupling of the transformer. For example, if nine out of ten flux lines remain inside the core, the coefficient of coupling is 0.90 or 90%. Most iron-core transformers have very high coefficients of coupling (greater than 0.99), whereas ferrite-core and air-core devices have lower values.

WINDING CAPACITANCE As you learned in Chapter 11, there is always some stray capacitance between adjacent turns of a winding. These stray capacitances result in an effective capacitance in parallel with each winding of a transformer, as indicated in Figure 14–26.

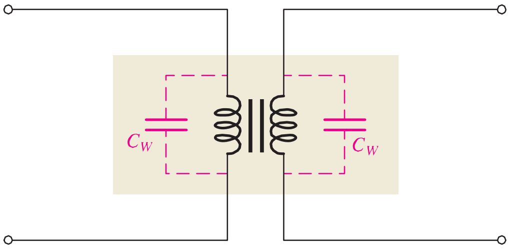

FIGURE 14–26 **Winding capacitance in a practical transformer.**

These stray capacitances have very little effect on the transformer's operation at low frequencies because the reactances (X_C) are very high. However, at higher frequencies, the reactances decrease and begin to produce a bypassing effect across the primary winding and across the secondary load. As a result, less of the total primary current is through the primary winding, and less of the total secondary current is through the load. This effect reduces the load voltage as the frequency goes up.

TRANSFORMER EFFICIENCY Recall that the secondary power is equal to the primary power in an ideal transformer. Because the nonideal characteristics just discussed result in a power loss in the transformer, the secondary (output) power is always less than the primary (input) power. The **efficiency**, symbolized by the Greek letter eta (η), of a transformer is a measure of the percentage of the input power that is delivered to the output.

$$\eta = \left(\frac{P_{out}}{P_{in}}\right)100\% \tag{14–10}$$

Most power transformers have efficiencies in excess of 95%.

HANDS ON TIP

If you have an unmarked small transformer, you can use a signal generator with a relatively low output voltage to check the voltage ratio and thus the turns ratio between the input (primary) and the output (secondary). This is safer than applying 120 V ac and holding your breath. Typically, the primary wires are black, a low-voltage secondary is green, and a high-voltage secondary is red. A striped wire usually indicates a center tap. Unfortunately, not all transformers have colored wires or the wires are not always standard colors.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 14–10

A certain type of transformer has a primary current of 5 A and a primary voltage of 4800 V. The secondary current is 95 A and the secondary voltage is 240 V. Determine the efficiency of this transformer.

SOLUTION

The input power is

$$P_{in} = V_{pri}I_{pri} = (4800\text{ V})(5\text{ A}) = 24\text{ kVA}$$

The output power is

$$P_{out} = V_{sec}I_{sec} = (240\text{ V})(95\text{ A}) = 22.8\text{ kVA}$$

The efficiency is

$$\eta = \left(\frac{P_{out}}{P_{in}}\right)100\% = \left(\frac{22.8\text{ kVA}}{24\text{ kVA}}\right)100\% = \mathbf{95\%}$$

RELATED PROBLEM

A transformer has a primary current of 9 A with a primary voltage of 440 V. The secondary current is 30 A and the secondary voltage is 120 V. What is the efficiency?

SYSTEM EXAMPLE 14–2

AN INDUCTION HEATING SYSTEM

A widely used method for heating electrical conductors is to use a coil that surrounds a conductive material to be heated. Heating is important for operations like brazing, soldering, heat treating, annealing, drying of adhesives and coatings, preheating metals prior to rubber molding, and many other processes. The hot coil is like the primary of a transformer; the conductive material to be heated acts as a secondary. Alternating current at frequencies from 60 Hz to over 1 MHz are used to induce a current in the material to be heated by transformer action. The heating element looks like a shorted secondary, which induces large eddy currents that are responsible for the heating.

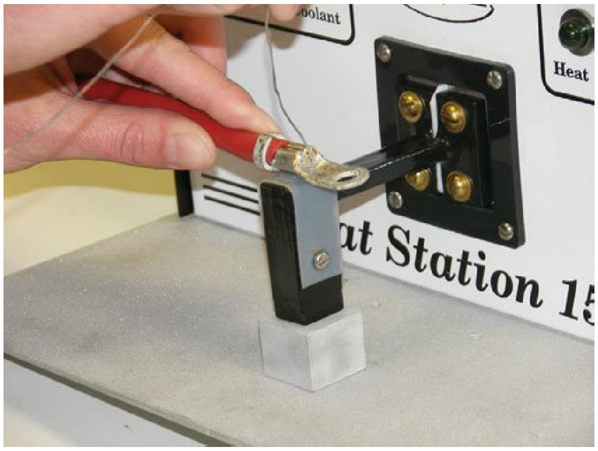

FIGURE 14–27 **A basic induction heating unit used for soldering applications** (Photo courtesy of MagneForce, Inc.).

Figure 14–27 shows a basic induction heating unit; heating units vary widely in size and heating characteristics that depend on the requirements. An important advantage of induction heating is that the material can be heated almost instantly (2000° F in less than 1 s). The heating pattern is repeatable and can provide a custom temperature profile depending on the application. For example, a soldering operation, such as done with cans, can be accomplished quickly in a continuous process, a major advantage in production lines.

6238262 2013/12/10 12.27.66.5

Figure 14–28 shows a block diagram of a basic heating system, which is part of a larger process (not shown). The ac input is converted to dc by a power supply, which is then converted to high-frequency ac (the frequency selected depends on the material and depth of heating required.) A controller senses when a part is ready for heat; it then closes a switch that applies power to the coil through an impedance-matching transformer.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

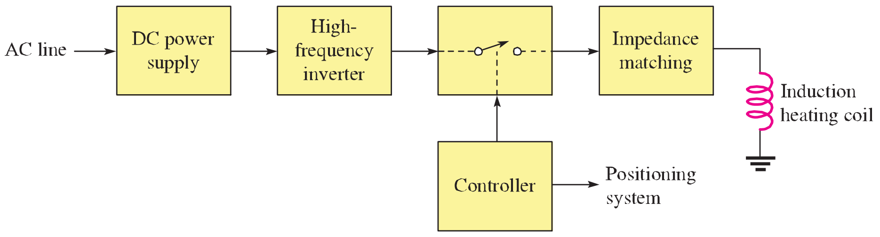
AC line
DC power supply
High-frequency inverter
Impedance matching
Induction heating coil
Controller
Positioning system

FIGURE 14–28 **Block diagram of heating unit.**

SECTION 14–7 CHECKUP

1. Explain how a practical transformer differs from the ideal model.
2. The coefficient of coupling of a certain transformer is 0.85. What does this mean?
3. A certain transformer has a rating of 10 kVA. If the secondary voltage is 250 V, how much load current can the transformer handle?

14–8 TAPPED AND MULTIPLE-WINDING TRANSFORMERS

The basic transformer has several important variations. They include tapped transformers, multiple-winding transformers, and autotransformers. Three-phase transformers, which are basically multiple-winding transformers, are covered in this section.

After completing this section, you should be able to

- **Describe several types of transformers**
  - **Describe center-tapped transformers**
  - **Describe multiple-winding transformers**
  - **Describe autotransformers**
  - **Explain how three-phase transformers are connected**

Tapped Transformers

A schematic of a transformer with a center-tapped secondary winding is shown in Figure 14–29(a). The **center tap (CT)** is equivalent to two secondary windings with half the total voltage across each.

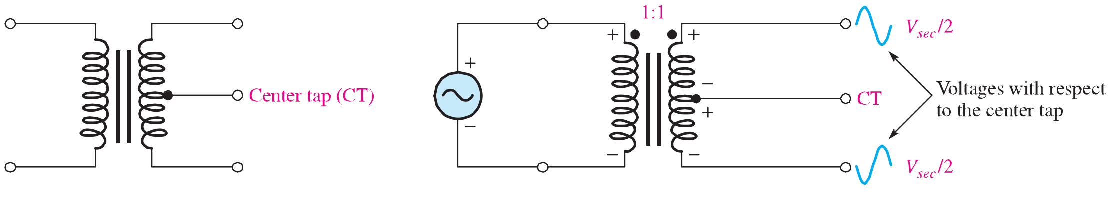
Center tap (CT)
1:1
+
−
+
+
−
+
−
−
CT
$V_{sec}/2$
Voltages with respect to the center tap
$V_{sec}/2$

(a) Center-tapped transformer

(b) Output voltages with respect to the center tap are 180° out of phase with each other and are one-half the magnitude of the secondary voltage.

FIGURE 14–29 **Operation of a center-tapped transformer.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The voltages between either end of the secondary winding and the center tap are, at any instant, equal in magnitude but opposite in polarity, as illustrated in Figure 14–29(b). Here, for example, at some instant on the sinusoidal voltage, the polarity across the entire secondary winding is as shown (top end +, bottom −). At the center tap, the voltage is less positive than the top end but more positive than the bottom end of the secondary. Therefore, measured with respect to the center tap, the top end of the secondary is positive, and the bottom end is negative. This center-tapped feature is used in many power supply rectifiers in which the ac voltage is converted to dc, as illustrated in Figure 14–30, and also in impedance-matching transformers.

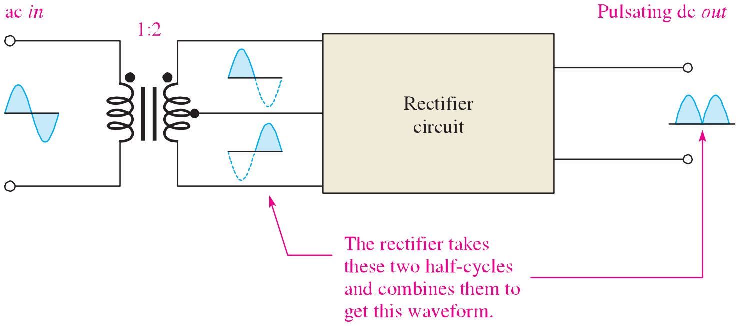
ac *in*
1:2
Rectifier circuit
Pulsating dc *out*
The rectifier takes these two half-cycles and combines them to get this waveform.

FIGURE 14–30 **Application of a center-tapped transformer in ac-to-dc conversion.**

Some tapped transformers have taps on the secondary winding at points other than the electrical center. Also, multiple primary and secondary taps are sometimes used in certain applications. Examples of these types of transformers are shown in Figure 14–31.

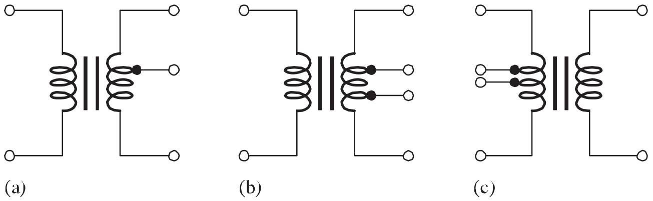

FIGURE 14–31 **Tapped transformers.**

Utility companies use many tapped transformers in distribution systems. Normally, power is generated and transmitted as three-phase power. At some point, the three-phase power is converted to single-phase for residential use. An example of a utility-pole transformer is shown in Figure 14–32 for the case where the high-voltage three-phase power has previously been converted to single-phase power (by tapping one of the three phases). It still needs to be converted to 120 V/240 V for residential customers, so a single-phase tapped transformer is used. By selecting the appropriate tap on the primary side, the utility company can make minor adjustments in the voltage that is delivered to the customer. The center tap on the secondary side is the neutral (usually uninsulated) conductor.

Multiple-Winding Transformers

Some transformers are designed to operate from either 120 V ac or 240 V ac lines. These transformers usually have two primary windings, each of which is designed for 120 V ac. When the two are connected in series, the transformer can be used for 240 V ac operation, as illustrated in Figure 14–33.

More than one secondary winding can be wound on a common core. Transformers with several secondary windings are often used to achieve several voltages by either stepping up or stepping down the primary voltage. These types are commonly used in power

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

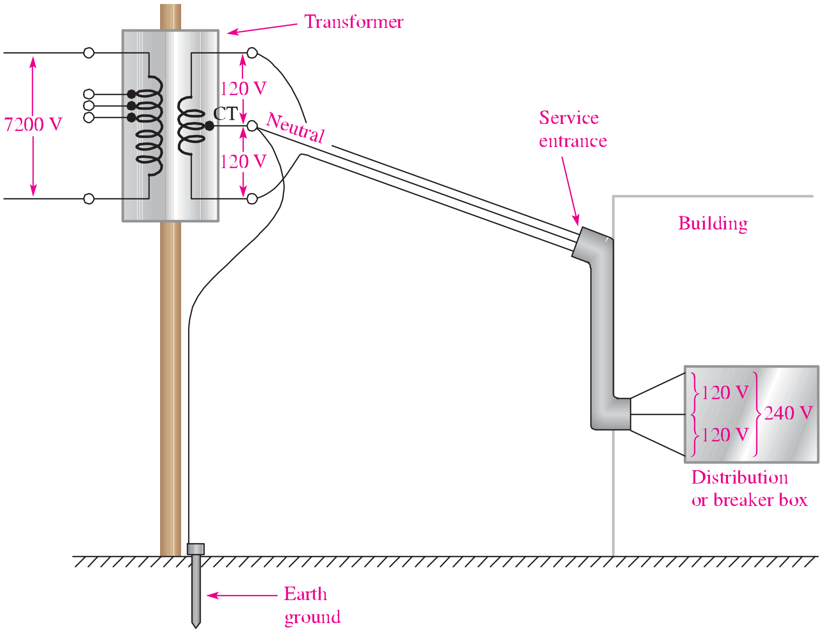
Transformer
7200 V
120 V
CT
Neutral
120 V
Service entrance
Building
120 V
240 V
120 V
Distribution or breaker box
Earth ground

FIGURE 14–32 **Utility-pole transformer in a typical power distribution system.**

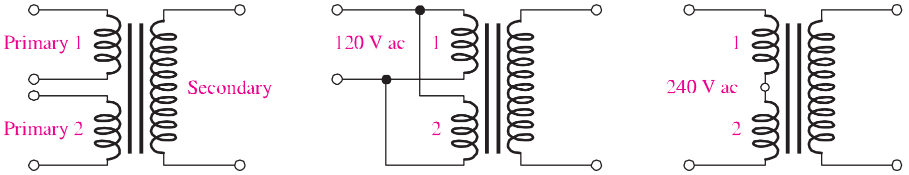
Primary 1
Secondary
Primary 2
120 V ac
1
2
1
240 V ac
2

(a) Two primary windings

(b) Primary windings in parallel for 120 V ac operation

(c) Primary windings in series for 240 V ac operation

6238262 2013/12/10 12.27.66.5

FIGURE 14–33 **Multiple-primary transformers.**

supply applications in which several voltage levels are required for the operation of an electronic system.

A typical schematic of a transformer with multiple secondary windings is shown in Figure 14–34; this transformer has three secondary windings. Sometimes you will find combinations of multiple primary windings, multiple secondary windings, and tapped transformers all in one unit.

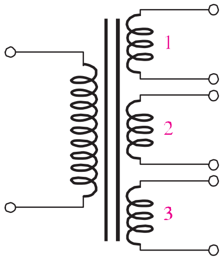

FIGURE 14–34 **A multiple-secondary transformer.**

EXAMPLE 14–11

The transformer shown in Figure 14–35 has the turns ratios for each secondary relative to the primary as indicated. One of the secondary windings is also center tapped. If 120 V ac are connected to the primary winding, determine each secondary voltage and the voltages with respect to the center tap (CT) on the middle secondary winding.

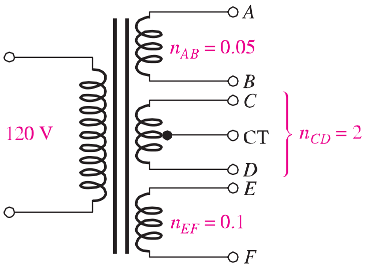
A
$n_{AB} = 0.05$
B
C
120 V
CT
$n_{CD} = 2$
D
E
$n_{EF} = 0.1$
F

FIGURE 14–35

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

$$V_{AB} = n_{AB}V_{pri} = (0.05)120\text{ V} = \mathbf{6.0\text{ V}}$$

$$V_{CD} = n_{CD}V_{pri} = (2)120\text{ V} = \mathbf{240\text{ V}}$$

$$V_{(\text{CT})C} = V_{(\text{CT})D} = \frac{240\text{ V}}{2} = \mathbf{120\text{ V}}$$

$$V_{EF} = n_{EF}V_{pri} = (0.1)120\text{ V} = \mathbf{12\text{ V}}$$

RELATED PROBLEM

Repeat the calculations if the primary winding is halved.

Autotransformers

Applications of the autotransformer include starting industrial induction motors and regulating transmission line voltages. In an **autotransformer**, one winding serves as both the primary and the secondary windings. The winding is tapped at the proper points to achieve the desired turns ratio for stepping up or stepping down the voltage.

Autotransformers differ from conventional transformers in that there is no electrical isolation between the primary and the secondary circuits because both are on one winding. Autotransformers normally are smaller and lighter than equivalent conventional transformers because they require a much lower kVA rating for a given load. Many autotransformers provide an adjustable tap using a sliding contact mechanism so that the output voltage can be varied (these are often called *variacs.*) Figure 14–36 shows schematic symbols for various types of autotransformers.

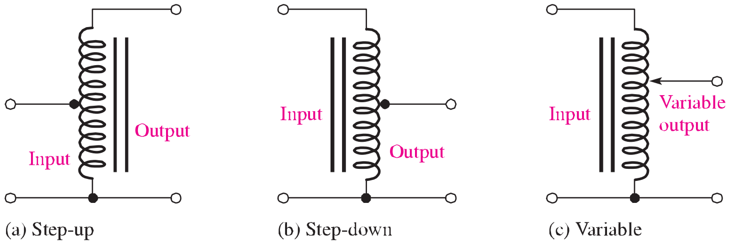

(a) Step-up (b) Step-down (c) Variable

FIGURE 14–36 **Variable autotransformers.**

Three-Phase Transformers

Three-phase power was introduced in Chapter 8 in relation to generators and motors. Three-phase transformers are widely used in power distribution systems. Three-phase is the most common way in which power is produced, transmitted, and used. Three-phase power is generally not used in residential applications, but it is particularly useful in industry because three-phase motors are more efficient.

A three-phase transformer consists of three sets of primary and secondary windings. Each set is wound on one leg of an iron-core assembly. Basically it is three single-phase transformers sharing a common core, as shown in Figure 14–37. It is possible (but more expensive) to connect three single-phase transformers together to achieve the same result. In a three-phase transformer, the three identical primary windings and the three identical secondary windings are connected in either of two ways, a delta (Δ) configuration or a wye (Y) configuration, to form a complete transformer unit. Delta and wye connections are shown in Figure 14–38.

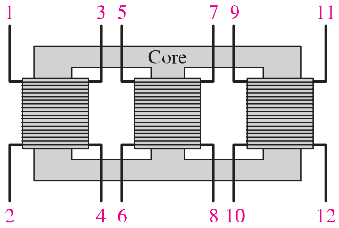

FIGURE 14–37 **Three-phase transformer.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

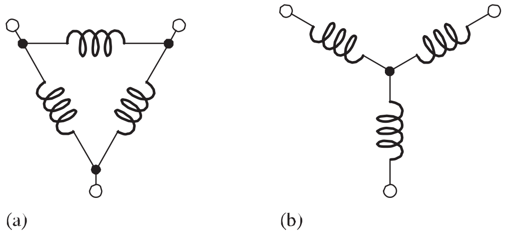

FIGURE 14–38 **Delta and wye configurations.**

In a three-phase transformer, the possible combinations of delta and wye configurations are

1. **Delta-to-wye (Δ-Y).** The primary winding is a delta and the secondary winding is a wye. This is the most common configuration used in commercial and industrial applications.
2. **Delta-to-delta (Δ-Δ).** The primary and secondary windings are both connected in the delta configuration. This is also commonly used in industrial applications.
3. **Wye-to-delta (Y-Δ).** The primary winding is a wye and the secondary winding is a delta. This is used in high-voltage transmission applications.
4. **Wye-to-wye (Y-Y).** The primary and secondary windings are both connected in the wye configuration. It is used in high-voltage, low kVA applications.

The delta-to-wye connections are shown in Figure 14–39. The proper winding phasing must be observed when connecting the transformers. The windings in the delta must be + to −; the polarity for each winding in the wye must be the same at the center point.

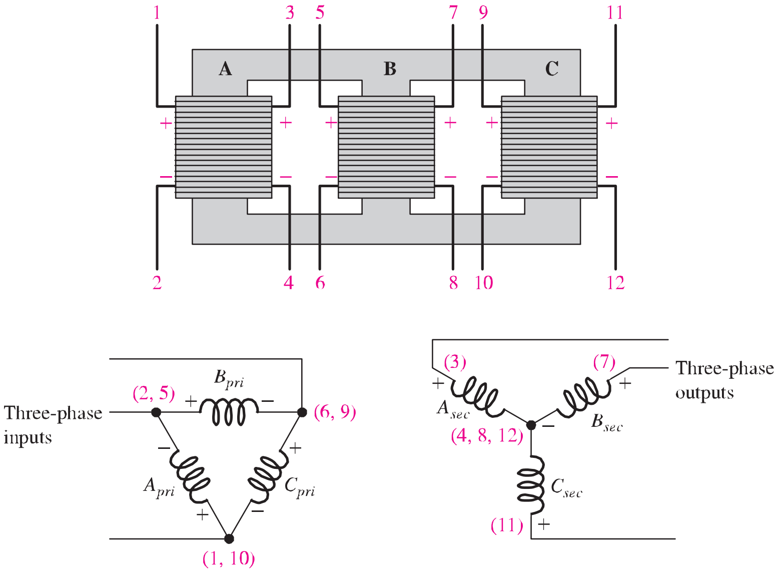

FIGURE 14–39 **Connections for a delta-to-wye transformer. The primary windings are designated A_{pri}, B_{pri} and C_{pri}; the secondary windings are designated A_{sec}, B_{sec}, and C_{sec}. The numbers in parentheses correspond to the transformer leads.**

The wye configuration has an advantage in that a neutral connection can be made at the center junction point. The delta configuration normally does not have a neutral. One exception is the special case for changing three-phase transmission-line voltage to single-phase power for residential use. In this case, a Y-Δ transformer with a center-tapped delta configuration is used, as shown in Figure 14–40. This configuration is referred to as a four-wire delta and can be used in cases where single phase is not available.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

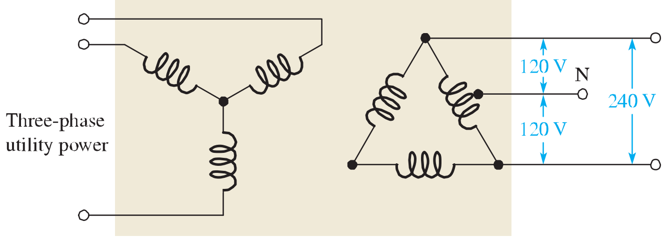

FIGURE 14–40 **A tapped wye-delta transformer that can be used to convert three-phase utility voltages to single-phase residential voltages.**

SECTION 14–8 CHECKUP

1. A certain transformer has two secondary windings. The turns ratio from the primary winding to the first secondary winding is 10. The turns ratio from the primary winding to the other secondary winding is 0.2. If 240 V ac are applied to the primary winding, what are the secondary voltages?
2. Name one advantage and one disadvantage of an autotransformer over a conventional transformer.
3. What is the most common configuration for a three-phase transformer?

14–9 TROUBLESHOOTING

6238262 2013/12/10 12.27.66.5

Transformers are reliable devices when operated within their specified range. Common failures in a transformer are opens in either the primary or the secondary windings. One cause of an open is the operation of the device under conditions that exceed its ratings. Normally, when a transformer fails, it is very difficult to repair, and therefore the simplest procedure is to replace it.

After completing this section, you should be able to

- **Troubleshoot transformers**
  - **Find an open primary or secondary winding**

When there is an open primary winding, there is no primary current and, therefore, no induced voltage or current in the secondary. This condition is illustrated in Figure 14–41(a), and the method of checking with an ohmmeter is shown in part (b).

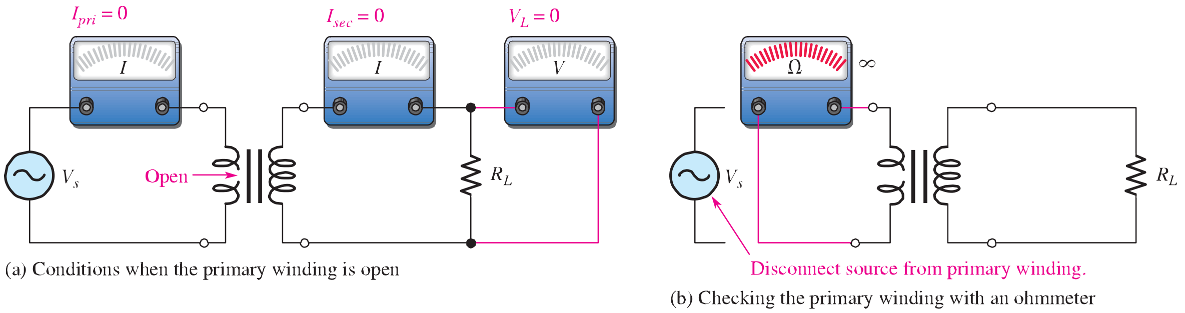

(a) Conditions when the primary winding is open

(b) Checking the primary winding with an ohmmeter

FIGURE 14–41 **Open primary winding.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

When there is an open secondary winding, there is no current in the secondary circuit and, as a result, no voltage across the load. Also, an open secondary winding causes the primary current to be very small (there is only a small magnetizing current). In fact, the primary current may be practically zero. This condition is illustrated in Figure 14–42(a), and the ohmmeter check is shown in part (b).

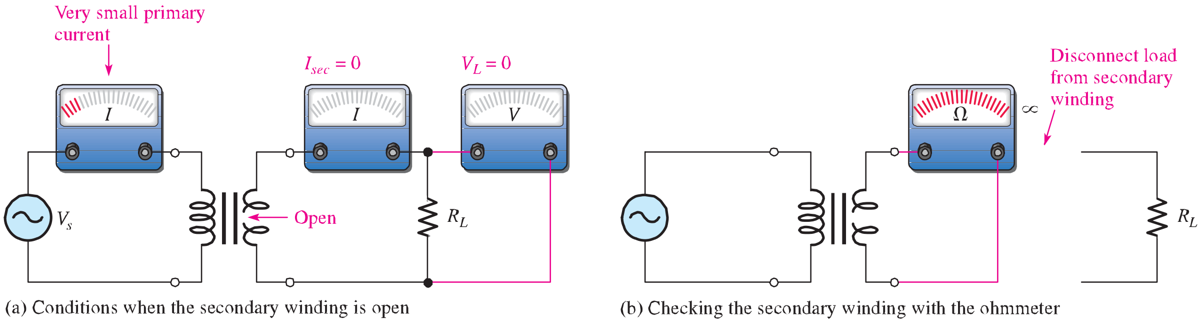
Very small primary current
I
$I_{sec} = 0$
I
$V_L = 0$
V
V_s
Open
R_L
Disconnect load from secondary winding
Ω
∞
R_L

(a) Conditions when the secondary winding is open

(b) Checking the secondary winding with the ohmmeter

FIGURE 14–42 **Open secondary winding.**

Shorted windings are very rare and if they do occur are difficult to find unless there is a visual indication or a large number of windings are shorted. A completely shorted primary winding will draw excessive current from the source; and unless there is a breaker or a fuse in the circuit, either the source or the transformer or both will burn out. A partial short in the primary winding can cause higher than normal or even excessive primary current.

SECTION 14–9 CHECKUP

1. What is the most probable failure in a transformer?
2. What is often the cause of transformer failure?

SUMMARY

- A transformer generally consists of two or more coils that are magnetically coupled on a common core.
- There is mutual inductance between two magnetically coupled coils.
- When current in one coil changes, voltage is induced in the other coil.
- The primary is the winding connected to the source, and the secondary is the winding connected to the load.
- The number of turns in the primary winding and the number of turns in the secondary winding determine the turns ratio.
- The relative polarities of the primary and secondary voltages are determined by the direction of the windings around the core.
- A step-up transformer has a turns ratio greater than 1.
- A step-down transformer has a turns ratio less than 1.
- A transformer cannot increase power.
- In an ideal transformer, the power from the source (input power) is equal to the power delivered to the load (output power).
- If the voltage is stepped up, the current is stepped down, and vice versa.
- A load connected across the secondary winding of a transformer appears to the source as a reflected load having a value dependent on the reciprocal of the turns ratio squared.
- An impedance-matching transformer can match a load resistance to an internal source resistance to achieve maximum power transfer to the load by selection of the proper turns ratio.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

- A balun is a type of transformer used to convert a balanced line (such as twisted pair wiring) to an unbalanced line (such as coaxial cable) or vice-versa.
- A transformer does not respond to constant dc.
- Conversion of electrical energy to heat in an actual transformer results from winding resistances, hysteresis loss in the core, eddy currents in the core, and flux leakage.
- Three-phase transformers are commonly used in power distribution applications.

KEY TERMS

Key terms and other bold terms in the chapter are defined in the end-of-book glossary.

Apparent power rating The method of rating transformers in which the power capability is expressed in volt-amperes (VA).

Center tap (CT) A connection at the midpoint of a winding in a transformer.

Electrical isolation The condition in which two circuits have no common conductive path between them.

Impedance matching A technique used to match a load resistance to a source resistance in order to achieve maximum transfer of power.

Magnetic coupling The magnetic connection between two coils as a result of the changing magnetic flux lines of one coil cutting through the second coil.

Mutual inductance (L_M) The inductance between two separate coils, such as in a transformer.

Primary winding The input winding of a transformer; also called *primary*.

Reflected resistance The resistance in the secondary circuit reflected into the primary circuit.

Secondary winding The output winding of a transformer; also called *secondary*.

Transformer An electrical device constructed of two or more coils of wire (windings) that are electromagnetically coupled to each other so that there is mutual inductance from one winding to another.

Turns ratio (n) The ratio of turns in the secondary winding to turns in the primary winding.

KEY FORMULAS

(14–1)	$L_M = k\sqrt{L_1L_2}$	Mutual inductance
(14–2)	$k = \dfrac{\phi_{1\text{-}2}}{\phi_1}$	Coefficient of coupling
(14–3)	$n = \dfrac{N_{sec}}{N_{pri}}$	Turns ratio
(14–4)	$\dfrac{V_{sec}}{V_{pri}} = \dfrac{N_{sec}}{N_{pri}}$	Voltage ratio
(14–5)	$V_{sec} = nV_{pri}$	Secondary voltage
(14–6)	$\dfrac{I_{pri}}{I_{sec}} = n$	Current ratio
(14–7)	$I_{sec} = \left(\dfrac{1}{n}\right)I_{pri}$	Secondary current
(14–8)	$R_{pri} = \left(\dfrac{1}{n}\right)^2 R_L$	Reflected resistance
(14–9)	$n = \sqrt{\dfrac{R_L}{R_{pri}}}$	Turns ratio for impedance matching
(14–10)	$\eta = \left(\dfrac{P_{out}}{P_{in}}\right)100\%$	Transformer efficiency

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

TRUE/FALSE QUIZ

Answers are at the end of the chapter.

1. An ideal transformer delivers the same power to the load as is delivered to the primary winding.
2. Dots on the schematic symbol of a transformer show the phase relationship between the input and the output.
3. A step-down transformer has more turns in the primary winding than in the secondary winding.
4. The primary current in a transformer is always larger than the secondary current.
5. When there is no load on a transformer, the power factor is 1.
6. Impedance-matching transformers allow maximum voltage to be passed from the source to the load.
7. Reflected resistance is the same as winding resistance.
8. A balun is a form of impedance-matching transformer.
9. Power transformers are typically rated in VA rather than in watts.
10. Transformer efficiency is the ratio of output voltage divided by input voltage.

6238262 2013/12/10 12.27.66.5

SELF-TEST

Answers are at the end of the chapter.

1. A transformer is used for
(a) dc voltages **(b)** ac voltages **(c)** both dc and ac

2. Which one of the following is affected by the turns ratio of a transformer?
(a) primary voltage **(b)** dc voltage **(c)** secondary voltage **(d)** none of these

3. If the windings of a certain transformer with a turns ratio of 1 are in opposite directions around the core, the secondary voltage is
(a) in phase with the primary voltage **(b)** less than the primary voltage
(c) greater than the primary voltage **(d)** out of phase with the primary voltage

4. When the turns ratio of a transformer is 10 and the primary ac voltage is 6 V, the secondary voltage is
(a) 60 V **(b)** 0.6 V **(c)** 6 V **(d)** 36 V

5. When the turns ratio of a transformer is 0.5 and the primary ac voltage is 100 V, the secondary voltage is
(a) 200 V **(b)** 50 V **(c)** 10 V **(d)** 100 V

6. A certain transformer has 500 turns in the primary winding and 2500 turns in the secondary winding. The turns ratio is
(a) 0.2 **(b)** 2.5 **(c)** 5 **(d)** 0.5

7. If 10 W of power are applied to the primary winding of an ideal transformer with a turns ratio of 5, the power delivered to the secondary load is
(a) 50 W **(b)** 0.5 W **(c)** 0 W **(d)** 10 W

8. In a certain loaded transformer, the secondary voltage is one-third the primary voltage. Ideally, the secondary current is
(a) one-third the primary current **(b)** three times the primary current
(c) equal to the primary current **(d)** less than the primary current

9. When a 1.0 kΩ load resistor is connected across the secondary winding of a transformer with a turns ratio of 2, the source "sees" a reflected load of
(a) 250 Ω **(b)** 2 kΩ **(c)** 4 kΩ **(d)** 1.0 kΩ

10. In Question 9, if the turns ratio is 0.5, the source "sees" a reflected load of
(a) 1.0 kΩ **(b)** 2 kΩ **(c)** 4 kΩ **(d)** 500 Ω

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

11. The turns ratio required to match a 50 Ω source to a 200 Ω load is
(a) 0.25 **(b)** 0.5 **(c)** 4 **(d)** 2

12. Maximum power is transferred from a source to a load when
(a) $R_L > R_{int}$ **(b)** $R_L < R_{int}$ **(c)** $R_L = R_{int}$ **(d)** $R_L = nR_{int}$

13. When a 12 V battery is connected across the primary winding of a transformer with a turns ratio of 4, the secondary voltage is
(a) 0 V **(b)** 12 V **(c)** 48 V **(d)** 3 V

14. A certain transformer has a turns ratio of 1 and a 0.95 coefficient of coupling. When 1 V ac is applied to the primary, the secondary voltage is
(a) 1 V **(b)** 1.95 V **(c)** 0.95 V

PROBLEMS

Answers to odd-numbered problems are at the end of the book.

BASIC PROBLEMS

SECTION 14–1 Mutual Inductance

1. What is the mutual inductance when $k = 0.75$, $L_1 = 1\ \mu H$, and $L_2 = 4\ \mu H$?
2. Determine the coefficient of coupling when $L_M = 1\ \mu H$, $L_1 = 8\ \mu H$, and $L_2 = 2\ \mu H$.

SECTION 14–2 The Basic Transformer

3. What is the turns ratio of a transformer having 120 turns in its primary winding and 360 turns in its secondary winding?
4. **(a)** What is the turns ratio of a transformer having 250 turns in its primary winding and 1000 turns in its secondary winding?
   (b) What is the turns ratio when the primary winding has 400 turns and the secondary winding has 100 turns?
5. Determine the phase of the secondary voltage with respect to the primary voltage for each transformer in Figure 14–43.

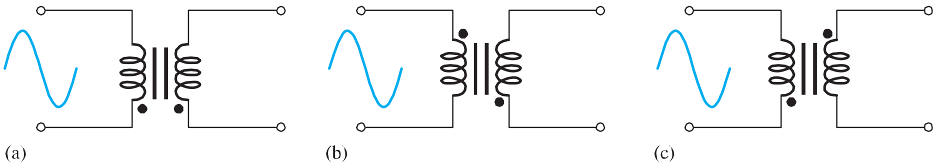
(a) (b) (c)

FIGURE 14–43

SECTION 14–3 Step-Up and Step-Down Transformers

6. If 120 V ac are connected to the primary of a transformer with a turns ratio of 1.5, what is the secondary voltage?
7. A certain transformer has 250 turns in its primary winding. In order to double the secondary voltage, how many turns must be in the secondary winding?
8. How many primary volts must be applied to a transformer with a turns ratio of 10 to obtain a secondary voltage of 60 V ac?
9. For each transformer in Figure 14–44, draw the secondary voltage showing its relationship to the primary voltage. Also indicate the amplitude.

FIGURE 14–44

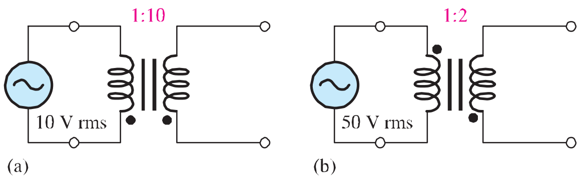

(a) (b)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

10. To step 120 V down to 30 V, what must be the turns ratio?

11. The primary winding of a transformer has 1200 V across it. What is the secondary voltage if the turns ratio is 0.2?

12. How many primary volts must be applied to a transformer with a turns ratio of 0.1 to obtain a secondary voltage of 6 V ac?

13. What is the voltage across the load in each circuit in Figure 14–45?

14. If the bottom of each secondary winding in Figure 14–45 were grounded, would the values of the load voltages be changed?

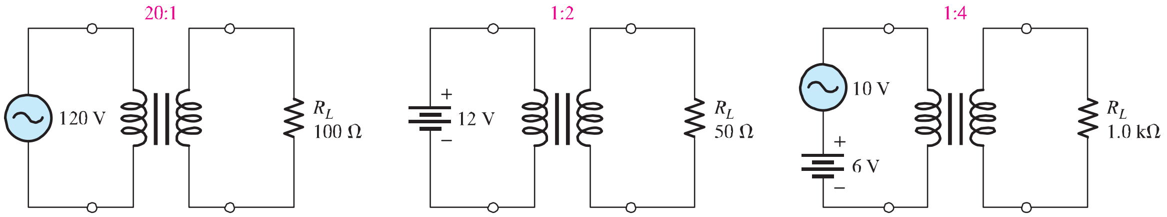

(a) (b) (c)

FIGURE 14–45

15. Determine the unspecified meter readings in Figure 14–46.

16. If R_L is doubled in the circuit of Figure 14–46(a), what would the secondary meter read?

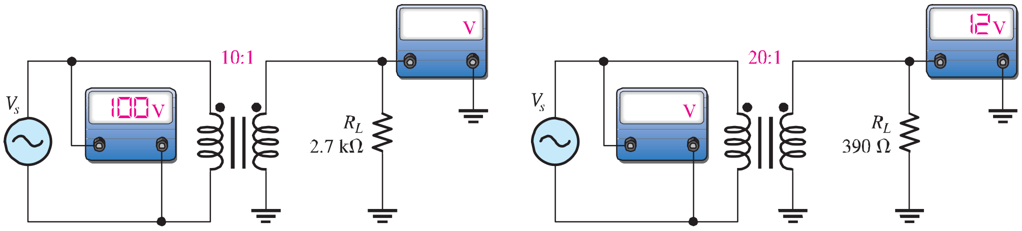

(a) (b)

FIGURE 14–46

SECTION 14–4 Loading the Secondary

17. Determine I_{sec} in Figure 14–47.

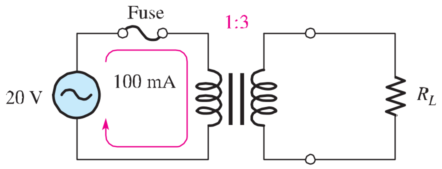

FIGURE 14–47

18. Determine the following quantities in Figure 14–48:

(a) secondary voltage **(b)** secondary current
(c) primary current **(d)** power in the load

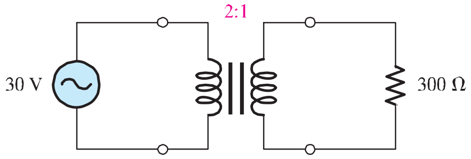

FIGURE 14–48

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 14–5 Reflected Load

19. What is the load resistance as seen by the source in Figure 14–49?

20. What is the resistance reflected into the primary circuit in Figure 14–50?

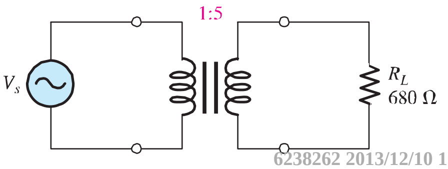

FIGURE 14–49

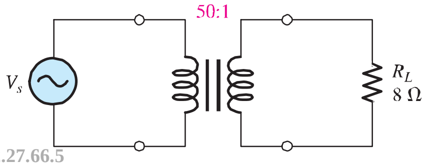

FIGURE 14–50

6238262 2013/12/10 12.27.66.5

21. What is the primary current (rms) in Figure 14–50 if the rms source voltage is 120 V?

22. What must be the turns ratio in Figure 14–51 in order to reflect 300 Ω into the primary circuit?

FIGURE 14–51

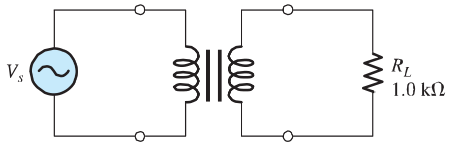

SECTION 14–6 Impedance Matching

23. For the circuit in Figure 14–52, find the turns ratio required to deliver maximum power to the 4 Ω speaker.

24. In Figure 14–52, what is the maximum power in watts delivered to the speaker?

25. Determine the value to which R_L must be adjusted in Figure 14–53 for maximum power transfer. The internal source resistance is 50 Ω.

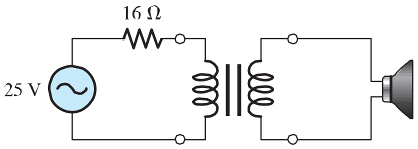

FIGURE 14–52

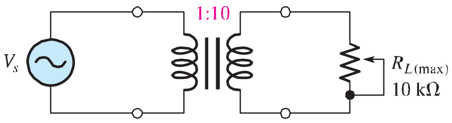

FIGURE 14–53

26. Plot the power curve for values of R_L in Figure 14–53 ranging from 1 kΩ to 10 kΩ in 1 kΩ increments. $R_L = 50\ \Omega$ and $V_s = 10\ \text{V}$.

SECTION 14–7 Transformer Ratings and Characteristics

27. In a certain transformer, the input power to the primary is 100 W. If 5.5 W are dissipated in the winding resistances, what is the output power to the load, neglecting any other losses?

28. What is the efficiency of the transformer in Problem 27?

29. Determine the coefficient of coupling for a transformer in which 2% of the total flux generated in the primary does not pass through the secondary.

30. A certain transformer is rated at 1 kVA. It operates on 60 Hz, 120 V ac. The secondary voltage is 600 V.
- **(a)** What is the maximum load current?
- **(b)** What is the smallest R_L that you can drive?
- **(c)** What is the largest capacitor that can be connected as a load?

31. What kVA rating is required for a transformer that must handle a maximum load current of 10 A with a secondary voltage of 2.5 kV?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 14–8 Tapped and Multiple-Winding Transformers

32. Determine each unknown voltage indicated in Figure 14–54.

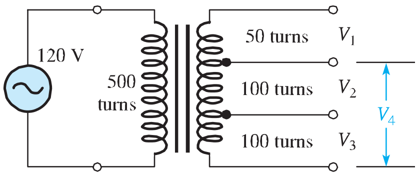

FIGURE 14–54

33. Using the indicated secondary voltages in Figure 14–55, determine the turns ratio of the primary winding to each tapped section.

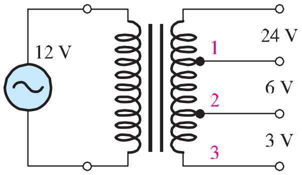

FIGURE 14–55

34. In Figure 14–56, each primary winding can accommodate 120 V ac. Show how the primaries should be connected for 240 V ac operation. Determine each secondary voltage.

35. Determine the turns ratios from each primary to each secondary in Figure 14–56.

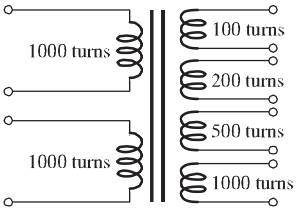

FIGURE 14–56

SECTION 14–9 Troubleshooting

36. When you apply 120 V ac across the primary winding of a transformer and check the voltage across the secondary winding, you get 0 V. Further investigation shows no primary or secondary current. List the possible faults. What is your next step in isolating the problem?

37. What is likely to happen if the primary winding of a transformer shorts?

ADVANCED PROBLEMS

38. The power supply in Figure 14–12 is designed to provide +5.0 V at 1 A, yet is fused for only $^1/_8$ A. Explain why the fuse does not blow at the rated output.

39. For the loaded, tapped-secondary transformer in Figure 14–57, determine the following:
- **(a)** all load voltages and currents
- **(b)** the resistance looking into the primary

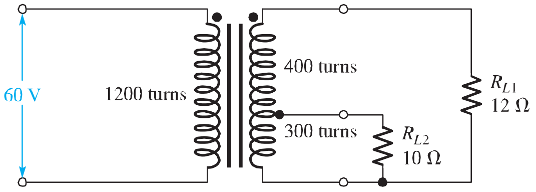

FIGURE 14–57

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

40. A certain transformer is rated at 5 kVA, 2400/120 V, at 60 Hz.
- **(a)** What is the turns ratio if the 120 V is the secondary voltage?
- **(b)** What is the current rating of the secondary if 2400 V is the primary voltage?
- **(c)** What is the current rating of the primary if 2400 V is the primary voltage?

41. Determine the voltage measured by each voltmeter in Figure 14–58. The bench-type meters have one terminal that connects to ground as indicated.

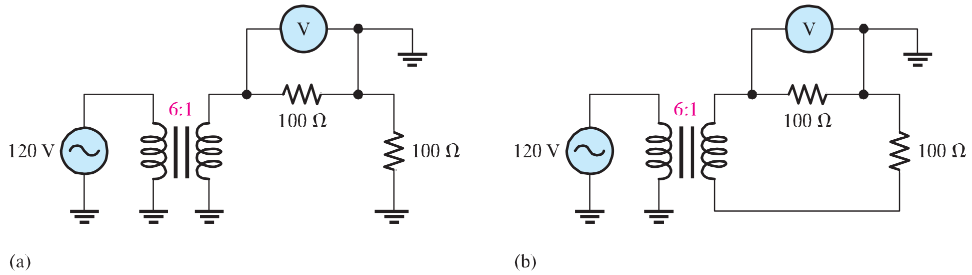

(a) (b)

FIGURE 14–58

42. Find the appropriate turns ratio for each switch position in Figure 14–59 in order to transfer the maximum power to each load when the internal source resistance is 10 Ω. Specify the number of turns for the secondary winding if the primary winding has 100 turns.

FIGURE 14–59

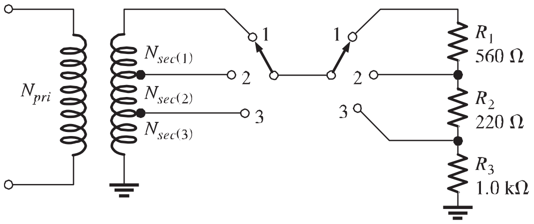

43. What must the turns ratio be in Figure 14–50 to limit the primary current to 3 mA with a source voltage of 120 V? Assume an ideal transformer and source.

44. Assume 120 V is applied to the primary of a transformer with a VA rating of 10 VA. The output voltage is 12.6 V. What is the smallest resistor that can be connected across the secondary?

45. A step-down transformer uses 120 V on the primary and 10 V on the secondary. If the secondary is rated for a maximum of 1 A, what fuse rating should be selected on the primary side?

MULTISIM

MULTISIM TROUBLESHOOTING PROBLEMS

46. Open file P14-46; files are found at www.pearsonhighered.com/floyd. Test the circuit. If there is a fault, identify it.

47. Open file P14-47 and test the circuit. If there is a fault, identify it.

48. Determine if there is a fault in the circuit in file P14-48. If so, identify it.

49. Find and specify any faulty component in the circuit in file P14-49.

ANSWERS TO SECTION CHECKUPS

SECTION 14–1 Mutual Inductance

1. Mutual inductance is the inductance between two coils, established by the amount of coupling between the coils.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

2. $L_M = k\sqrt{L_1L_2} = 45\text{ mH}$
3. Induced voltage increases when *k* is increased.

SECTION 14–2 The Basic Transformer

1. Transformer operation is based on the principle of mutual inductance.
2. Turns ratio is the ratio of turns in the secondary to turns in the primary.
3. Directions of the windings determine the relative voltage polarities.
4. $n = N_{sec}/N_{pri} = 0.5$
5. The windings are formed on a printed circuit board.

6238262 2013/12/10 12.27.66.5

SECTION 14–3 Step-Up and Step-Down Transformers

1. A step-up transformer increases voltage.
2. The secondary voltage is five times greater.
3. $V_{sec} = nV_{pri} = 2400\text{ V}$
4. A step-down transformer decreases voltage.
5. $V_{sec} = nV_{pri} = 60\text{ V}$
6. $n = 12\text{ V}/120\text{ V} = 0.1$
7. Electrical isolation, surge protection, and filters to eliminate interference

SECTION 14–4 Loading the Secondary

1. The secondary current is half the primary current.
2. $n = 0.25$; $I_{sec} = (1/n)I_{pri} = 2\text{ A}$
3. $I_{pri} = nI_{sec} = 2.5\text{ A}$
4. Disconnecting the load can lead to arcing, which can damage the transformer and equipment and present a shock hazard.

SECTION 14–5 Reflected Load

1. Reflected resistance is the resistance in the secondary circuit as seen from the primary circuit as a function of the turns ratio.
2. The reciprocal of the turns ratio determines reflected resistance.
3. $R_{pri} = (1/n)^2 R_L = 0.5\ \Omega$
4. $n = \sqrt{R_L/R_{pri}} = 0.1$

SECTION 14–6 Impedance Matching

1. Impedance matching is making maximum power transfer between a source and a load.
2. Maximum power is delivered to the load when $R_L = R_{int}$.
3. $R_{pri} = (1/n)^2 R_L = 400\ \Omega$
4. To connect unbalanced signals to balanced signals (and vice versa) and provide impedance matching.

SECTION 14–7 Transformer Ratings and Characteristics

1. In a practical transformer, conversion of electrical energy to heat reduces the efficiency. An ideal transformer has an efficiency of 100%.
2. When the coefficient of coupling is 0.85, 85% of the magnetic flux generated in the primary winding passes through the secondary winding.
3. $I_L = 10\text{ kVA}/250\text{ V} = 40\text{ A}$

SECTION 14–8 Tapped and Multiple-Winding Transformers

1. $V_{sec} = 10(240\text{ V}) = 2400\text{ V}$; $V_{sec} = 0.2(240\text{ V}) = 48\text{ V}$
2. Autotransformers are smaller and lighter for the same rating. Autotransformers provide no electrical isolation.
3. The delta-to-wye configuration.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 14–9 Troubleshooting

1. The most probable failure is an open winding.
2. Operating above rated values causes transformer failure.

ANSWERS TO RELATED PROBLEMS FOR EXAMPLES

14–1 387 μH

14–2 0.75

14–3 5000 turns

14–4 480 V

14–5 57.6 V

14–6 5 mA; 400 mA

14–7 6 Ω

14–8 0.354

14–9 0.0707 or 14.14:1

14–10 91%

14–11 $V_{AB} = 12$ V; $V_{CD} = 480$ V; $V_{(\mathrm{CT})C} = V_{(\mathrm{CT})D} = 240$ V; $V_{EF} = 24$ V

ANSWERS TO TRUE/FALSE QUIZ

1. T **2.** T **3.** T **4.** F **5.** F **6.** F **7.** F **8.** T **9.** T **10.** F

ANSWERS TO SELF-TEST

1. (b) **2.** (c) **3.** (d) **4.** (a) **5.** (b) **6.** (c) **7.** (d)
8. (b) **9.** (a) **10.** (c) **11.** (d) **12.** (c) **13.** (a) **14.** (c)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

CHAPTER 15

TIME RESPONSE OF REACTIVE CIRCUITS

OUTLINE

15–1 The *RC* Integrator
15–2 Response of *RC* Integrators to a Single Pulse
15–3 Response of *RC* Integrators to Repetitive Pulses
15–4 Response of *RC* Differentiators to a Single Pulse
15–5 Response of *RC* Differentiators to Repetitive Pulses
15–6 Response of *RL* Integrators to Pulse Inputs
15–7 Response of *RL* Differentiators to Pulse Inputs
15–8 Integrator and Differentiator Applications
15–9 Troubleshooting

OBJECTIVES

- Explain the operation of an *RC* integrator
- Analyze an *RC* integrator with a single-pulse input
- Analyze an *RC* integrator with a repetitive-pulse input
- Analyze an *RC* differentiator with a single-pulse input
- Analyze an *RC* differentiator with a repetitive-pulse input
- Analyze the operation of an *RL* integrator
- Analyze the operation of an *RL* differentiator
- Discuss some applications of integrators and differentiators
- Troubleshoot *RC* integrators and *RC* differentiators

KEY TERMS

Integrator
Transient time
Steady state
Differentiator

INTRODUCTION

In Chapters 10 and 12, the frequency response of *RC* and *RL* circuits was covered. In this chapter, the time response of *RC* and *RL* circuits with pulse inputs is examined.

Before starting this chapter, you should review the material in Sections 9–5 and 11–4. Understanding exponential changes in voltages and currents in capacitors and inductors is crucial to the study of the time response of reactive circuits. Exponential formulas that were given in Chapters 9 and 11 are used throughout this chapter.

With pulse inputs, the time response of the circuits is of primary importance. In the areas of pulse and digital circuits, technicians are often concerned with how a circuit responds over an interval of time to rapid changes in voltage or current. The relationship of the circuit time constant to the input pulse characteristics, such as pulse width and period, determines the wave shapes of the voltages in the circuit.

The terms *integrator* and *differentiator* refer to circuits that can, under certain conditions, approximate the mathematical process of integration and differentiation. Mathematical integration is a summation process. Under certain conditions, integrators are capable of averaging a waveform, as you will see in this chapter. Mathematical differentiation refers to the process of finding the rate of change of a quantity. Differentiator circuits produce an output that represents the rate of change of the input. Differentiator circuits are common in systems that require timing triggers, such as radar systems.

VISIT THE WEBSITE
Study aids for this chapter are available at
http://pearsonhighered.com/floyd

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

15–1 THE *RC* INTEGRATOR

In terms of time response, a series *RC* circuit in which the output voltage is taken across the capacitor is known as an integrator. Recall that in terms of frequency response, this particular series *RC* circuit is a low-pass filter. The term, *integrator,* is derived from a mathematical function which this type of circuit approximates under certain conditions.

After completing this section, you should be able to

- **Explain the operation of an *RC* integrator**
  - **Describe how the capacitor charges and discharges**
  - **Explain how a capacitor reacts to an instantaneous change in voltage or current**
  - **Describe the basic output voltage waveform**

Charging and Discharging of a Capacitor

When a pulse generator is connected to the input of an *RC* integrator, as shown in Figure 15–1, the capacitor will charge and discharge in response to the pulses. When the input goes from its low level to its high level, the capacitor charges toward the high level of the pulse through the resistor. This charging action is analogous to connecting a battery through a switch to the *RC* circuit, as illustrated in Figure 15–2(a). When the pulse goes from its high level back to its low level, the capacitor discharges back through the source. Compared to the resistance of the resistor, R, the resistance of the source is assumed to be negligible. This discharging action is analogous to replacing the source with a closed switch, as illustrated in Figure 15–2(b).

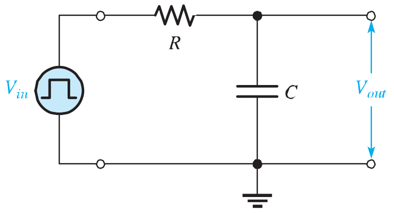

FIGURE 15–1 **An *RC* integrator with a pulse generator connected.**

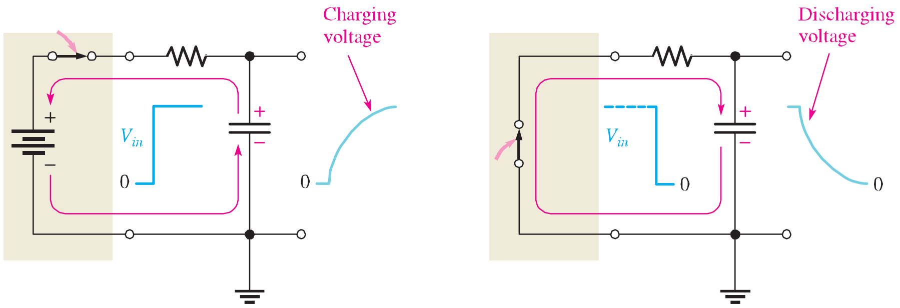

(a) When the input pulse goes HIGH, the source effectively acts as a battery in series with a closed switch, thereby charging the capacitor.

(b) When the input pulse goes back LOW, the source effectively acts as a closed switch, providing a discharge path for the capacitor.

6238262 2013/12/10 12.27.66.5

FIGURE 15–2 **The equivalent action when a pulse source charges and discharges the capacitor.**

A capacitor will charge and discharge following an exponential curve. Its rate of charging and discharging depends on the *RC* **time constant**, a fixed time interval determined by R and C ($\tau = RC$).

For an ideal pulse, both edges are considered to be instantaneous. Two basic rules of capacitor behavior help in understanding the response of *RC* circuits to pulse inputs.

1. The capacitor appears as a short to an instantaneous change in current and as an open to dc.
2. The voltage across the capacitor cannot change instantaneously—it can change only exponentially.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Capacitor Voltage

In an *RC* integrator, the output is the capacitor voltage. The capacitor charges during the time that the input pulse is high. If the pulse is at its high level long enough, the capacitor will fully charge to the voltage amplitude of the pulse, as illustrated in Figure 15–3. The capacitor discharges during the time that the pulse is low. If the low time between pulses is long enough, the capacitor will fully discharge to zero, as shown in the figure. Then when the next pulse occurs, it will charge again.

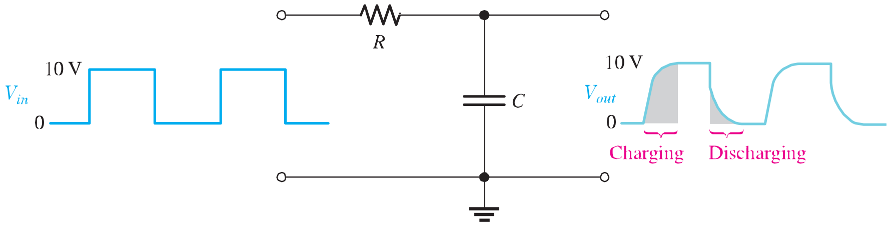

FIGURE 15–3 **Illustration of a capacitor fully charging and discharging in response to a pulse input. A pulse generator is connected to the input, but the symbol is not shown, only the waveform.**

SECTION 15–1 CHECKUP*

1. Define the term *integrator* in relation to an *RC* circuit.
2. What causes a capacitor in an *RC* circuit to charge and discharge?

***Answers are at the end of the chapter.**

15–2 RESPONSE OF *RC* INTEGRATORS TO A SINGLE PULSE

From the previous section, you have a general idea of how an *RC* integrator responds over time to a pulse input. In this section, the time response to a single pulse is examined in detail.

After completing this section, you should be able to

- **Analyze an *RC* integrator with a single-pulse input**
  - **Discuss the importance of the circuit time constant**
  - **Define *transient time***
  - **Determine the response when the pulse width is equal to or greater than five time constants**
  - **Determine the response when the pulse width is less than five time constants**

Two conditions of response to a single-pulse input must be considered:

1. When the input pulse width (t_W) is equal to or greater than five time constants ($t_W \geq 5\tau$)
2. When the input pulse width is less than five time constants ($t_W < 5\tau$)

Recall that five time constants is accepted as the time a capacitor needs to fully charge or fully discharge; this time is often called the **transient time**. A capacitor will fully charge if the pulse width is equal to or greater than five time constants (5τ). This condition is expressed as $t_W \geq 5\tau$. At the end of the pulse, the capacitor fully discharges back through the source.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Figure 15–4 illustrates the output waveforms for various *RC* transient times and a fixed input pulse width. Notice that as the transient time becomes shorter, compared to the pulse width, the shape of the output pulse approaches that of the input. In each case, the output reaches the full amplitude of the input.

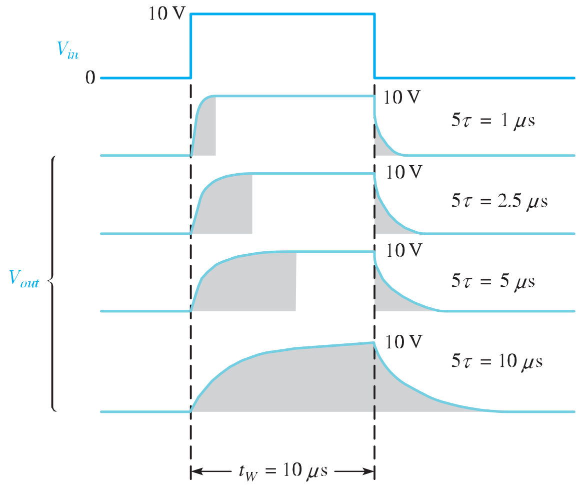

FIGURE 15–4 **Variation of an integrator's output pulse shape with transient time. The shaded areas indicate when the capacitor is charging and discharging.**

Figure 15–5 shows how a fixed time constant and a variable input pulse width affect the integrator output. Notice that as the pulse width is increased, the shape of the output pulse approaches that of the input. Again, this means that the transient time is short compared to the pulse width. The rise and fall times of the output remain constant.

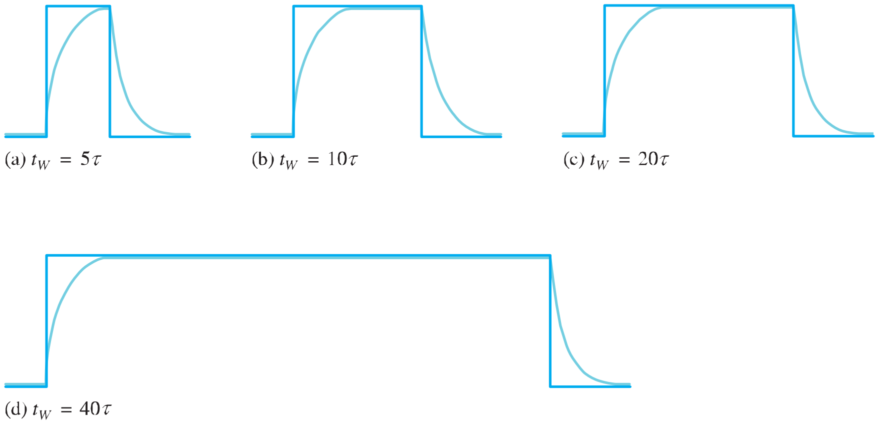
(a) $t_W = 5\tau$ (b) $t_W = 10\tau$ (c) $t_W = 20\tau$

(d) $t_W = 40\tau$

FIGURE 15–5 **Variation of an integrator's output pulse shape with input pulse width (the time constant is fixed). Dark blue is input and light blue is output.**

Now let's examine the case in which the width of the input pulse is less than five time constants of the *RC* integrator. This condition is expressed as $t_W < 5\tau$. As you know, a capacitor charges for the duration of the pulse and the pulse width is the time it has for charging. However, because the pulse width is less than the time the capacitor needs to fully charge (5τ), the output voltage will *not* reach the full input voltage before the end of the pulse. The capacitor only partially charges, as illustrated in Figure 15–6 for several values of *RC* time constants. Notice that for longer time constants, the output reaches a

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

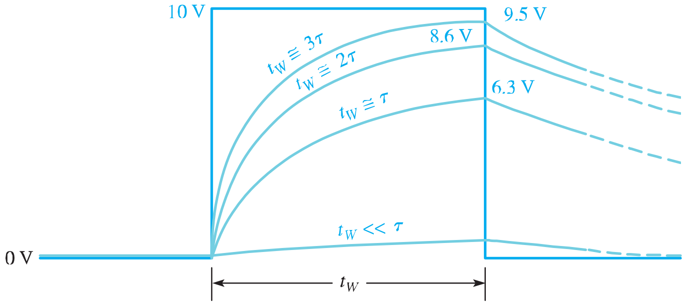

FIGURE 15–6 **Capacitor voltage for various time constants that are longer than the input pulse width. Dark blue is input and light blue is output.**

lower voltage because the capacitor cannot charge as much during the pulse width. Of course, in the examples with a single-pulse input, the capacitor fully discharges after the pulse ends.

When the time constant is much greater than the input pulse width, the capacitor charges very little, and, as a result, the output voltage becomes a very small almost constant value, as indicated in Figure 15–6.

Figure 15–7 illustrates the effect of reducing the input pulse width for a fixed time constant value. As the pulse width is reduced, the output voltage becomes smaller because the capacitor has less time to charge. However, it takes the capacitor the same length of time (5τ) to discharge back to zero for each condition after the pulse is removed.

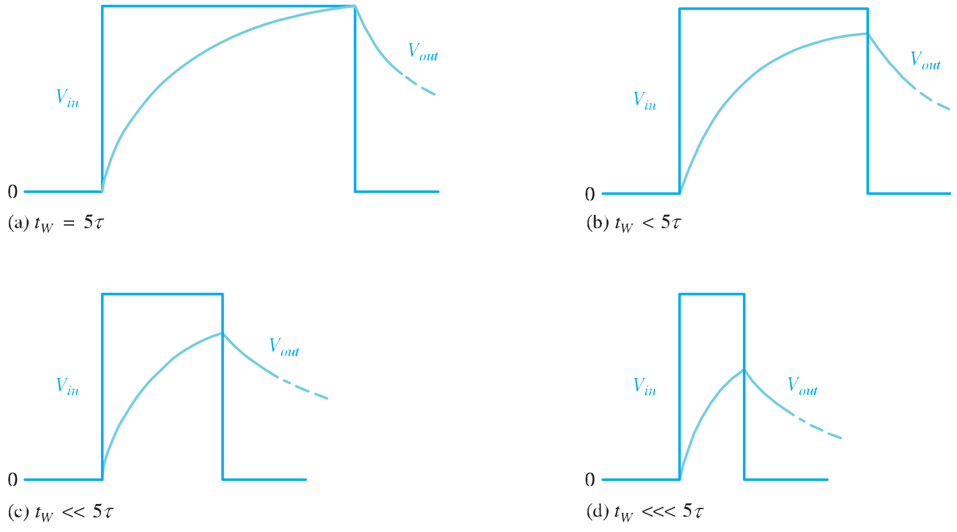

(a) $t_W = 5\tau$

(b) $t_W < 5\tau$

(c) $t_W << 5\tau$

(d) $t_W <<< 5\tau$

FIGURE 15–7 **The capacitor charges less and less as the input pulse width is reduced. The time constant is fixed.**

EXAMPLE 15–1

A single 10 V pulse with a width of 100 μs is applied to the *RC* integrator in Figure 15–8. The source resistance is assumed to be zero.

(a) To what voltage will the capacitor charge?

(b) How long will it take the capacitor to discharge?

(c) Show the output voltage waveform.

6238262 2013/12/10 12.27.66.5

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

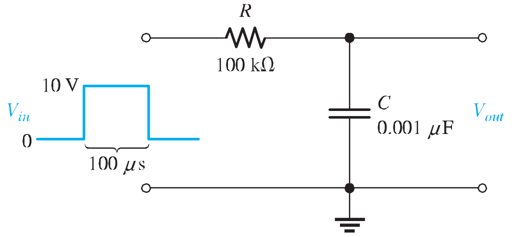

FIGURE 15–8

SOLUTION

(a) The circuit time constant is

$$\tau = RC = (100\text{ k}\Omega)(0.001\ \mu\text{F}) = 100\ \mu\text{s}$$

Notice that the pulse width is exactly equal to one time constant. Thus, the capacitor will charge approximately 63% of the full input amplitude in one time constant, so the output will reach a maximum voltage of

$$V_{out} = (0.63)10\text{ V} = \mathbf{6.3\ V}$$

(b) The capacitor discharges back through the source when the pulse ends. The total discharge time is

$$5\tau = 5(100\ \mu\text{s}) = \mathbf{500\ \mu s}$$

(c) The output charging and discharging curve is shown in Figure 15–9.

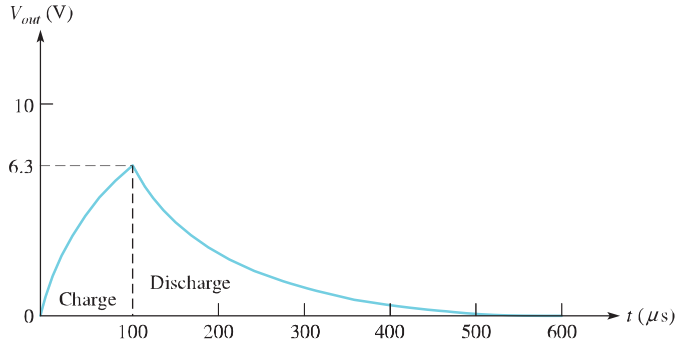

FIGURE 15–9

RELATED PROBLEM*

If the input pulse width in Figure 15–8 is increased to 200 μs, to what voltage will the capacitor charge?

*Answers are at the end of the chapter.

MULTISIM

Open Multisim file E15-01; files are found at www.pearsonhighered.com/floyd. Verify the calculated results in this example and the related problem.

EXAMPLE 15–2

Determine to what voltage the capacitor in Figure 15–10 will charge when a single pulse is applied to the input. Assume C is initially uncharged and the source resistance is zero.

SOLUTION

Calculate the time constant.

$$\tau = RC = (2.2\text{ k}\Omega)(1\ \mu\text{F}) = 2.2\text{ ms}$$

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

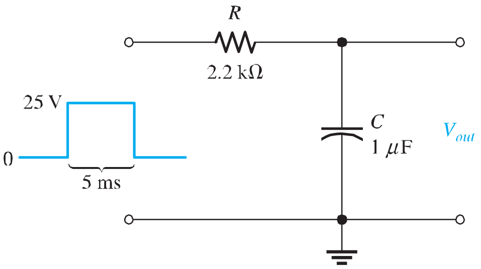

FIGURE 15–10

Because the pulse width is 5 ms, the capacitor charges for approximately 2.27 time constants (5 ms/2.2 ms = 2.27). Use the exponential formula, Equation 9–15, to find the voltage to which the capacitor will charge. With $V_F = 25$ V and $t = 5$ ms, the calculation is as follows:

$$v = V_F(1 - e^{-t/RC}) = (25\text{ V})(1 - e^{-5\text{ms}/2.2\text{ms}}) = \mathbf{22.4\text{ V}}$$

These calculations show that the capacitor charges to 22.4 V during the 5 ms duration of the input pulse. It will discharge back to zero when the pulse goes back to zero.

RELATED PROBLEM

Determine how much *C* will charge if the pulse width is increased to 10 ms.

SECTION 15–2 CHECKUP

1. When an input pulse is applied to an *RC* integrator, what condition must exist in order for the output voltage to reach the full amplitude of the input?
2. For the circuit in Figure 15–11, which has a single-pulse input, find the maximum output voltage and determine how long the capacitor will discharge.
3. For Figure 15–11, draw the approximate shape of the output voltage with respect to the input pulse.
4. If an integrator time constant equals the input pulse width, will the capacitor fully charge?
5. Describe the condition in an integrator under which the output voltage has the approximate shape of a rectangular input pulse.

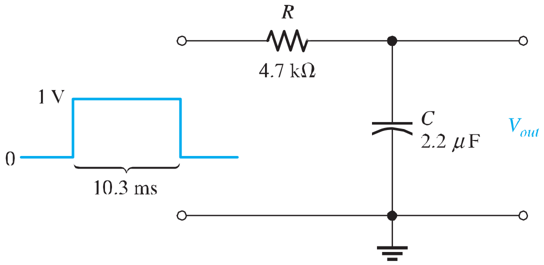

FIGURE 15–11

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

15–3 RESPONSE OF *RC* INTEGRATORS TO REPETITIVE PULSES

In electronic systems, you will encounter waveforms with repetitive pulses much more often than with single pulses. However, an understanding of the integrator's response to single pulses is necessary in order to understand how these circuits respond to repeated pulses.

After completing this section, you should be able to

- **Analyze an *RC* integrator with a repetitive-pulse input**
  - **Determine the response when the capacitor does not fully charge or discharge**
  - **Define *steady state***
  - **Describe the effect of a change in time constant on circuit response**

6238262 2013/12/10 12.27.66.5

If a periodic pulse waveform is applied to an *RC* integrator, as shown in Figure 15–12, the output waveshape depends on the relationship of the circuit time constant and the frequency of the input pulses. The capacitor, of course, charges and discharges in response to a pulse input. The amount of charge and discharge of the capacitor depends both on the circuit time constant and on the input frequency, as mentioned.

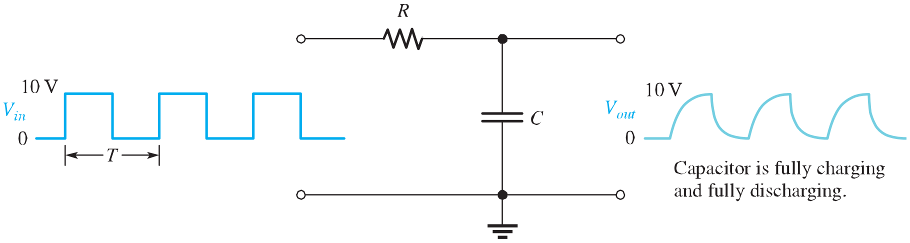

Capacitor is fully charging and fully discharging.

FIGURE 15–12 ***RC* integrator with a repetitive-pulse waveform input.**

If the pulse width and the time between pulses are each equal to or greater than five time constants, the capacitor will fully charge and fully discharge during each period of the input waveform. This case is shown in Figure 15–12.

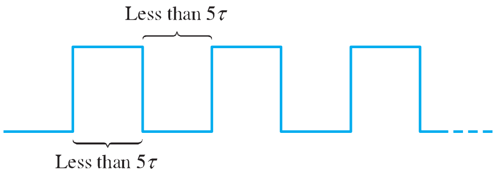

FIGURE 15–13 **Input waveform that does not allow full charge or discharge of the integrator capacitor.**

When the pulse width and the time between pulses are shorter than 5 time constants, as illustrated in Figure 15–13 for a square wave, the capacitor will *not* completely charge or discharge. We will now examine the effects of this situation on the output voltage of the *RC* integrator.

For illustration, let's use an *RC* integrator with a charging and discharging time constant equal to the pulse width of a 10 V square wave input, as in Figure 15–14. This choice will simplify the analysis and will demonstrate the basic action of the integrator under these conditions. At this point, we really do not care what the exact time constant value is because we know that an *RC* circuit charges 63.2% during one time constant interval.

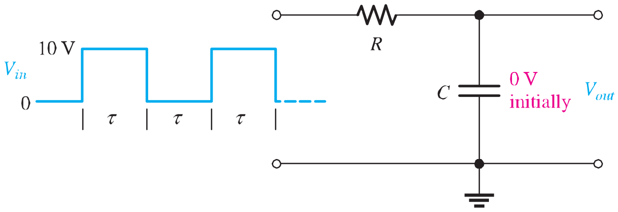

FIGURE 15–14 **Integrator with a square wave input having a period equal to two time constants ($T = 2\tau$).**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Let's assume that the capacitor in Figure 15–14 begins initially uncharged and examine the output voltage on a pulse-by-pulse basis. The results of this analysis are shown in Figure 15–15.

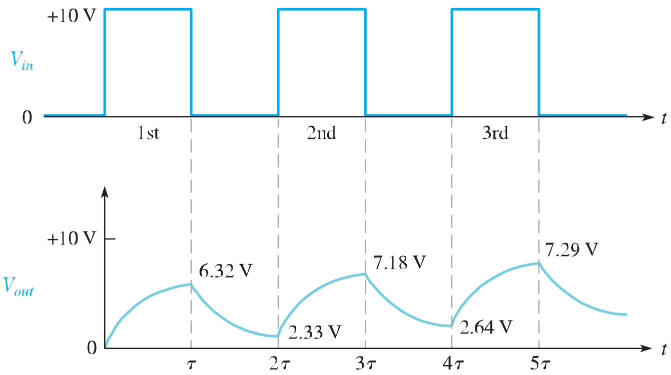

FIGURE 15–15 Input and output for the initially uncharged integrator in Figure 15–14.

First pulse During the first pulse, the capacitor charges. The output voltage reaches 6.32 V, which is 63.2% of 10 V, as shown in Figure 15–15.

Between first and second pulses The capacitor discharges, and the voltage decreases to 36.8% of the voltage at the beginning of this interval: 0.368(6.32 V) = 2.33 V.

Second pulse The capacitor voltage begins at 2.33 V and increases 63.2% of the way to 10 V. The total charging range is 10 V − 2.33 V = 7.67 V. The capacitor voltage will increase an additional 63.2% of 7.67 V, which is 4.85 V. Thus, at the end of the second pulse, the output voltage is 2.33 V + 4.85 V = 7.18 V, as indicated in Figure 15–15. Notice that the average is building up.

Between second and third pulses The capacitor discharges during this time, and therefore the voltage decreases to 36.8% of the initial voltage by the end of the second pulse: 0.368(7.18 V) = 2.64 V.

Third pulse At the start of the third pulse, the capacitor voltage is 2.64 V. The capacitor charges 63.2% of the way from 2.64 V to 10 V: 0.632(10 V − 2.64 V) = 4.65 V. Therefore, the voltage at the end of the third pulse is 2.64 V + 4.65 V = 7.29 V.

HANDS ON TIP
When you view pulse waveforms on an oscilloscope, set the scope to be DC coupled, not AC coupled. The reason is that AC coupling places a capacitor in series and can cause distortion in low-frequency pulse signals. Also, DC coupling allows you to observe the dc level of the pulse.

Steady-State Time Response

In the preceding discussion, the output voltage gradually built up and then began leveling off. It takes approximately 5τ for the output voltage to build up to 99% of its final value, regardless of the number of pulses that may occur during that time interval. This interval is the transient time of the circuit. Once the output voltage reaches the average value of the input voltage, a steady-state condition is reached, which continues as long as the periodic input continues. This condition is illustrated in Figure 15–16 based on the approximate values obtained after the first three pulses.

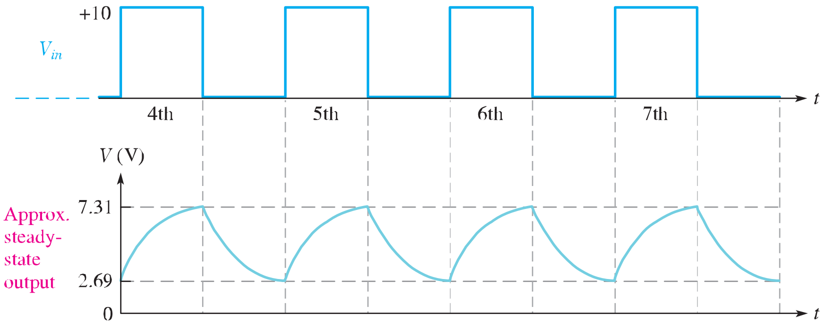

FIGURE 15–16 Output reaches steady state after 5τ and stabilizes at the values indicated.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

The transient time for our example circuit is the time from the beginning of the first pulse to the end of the third pulse. The capacitor voltage at the end of the third pulse is 7.29 V, which is about 99% of its final value.

The Effect of an Increase in Time Constant

What happens to the output voltage if the *RC* time constant of the integrator is increased with a variable resistor, as indicated in Figure 15–17? As the time constant is increased, the capacitor charges less during a pulse and discharges less between pulses. The result is a smaller fluctuation in the output voltage for increasing values of time constant, as shown in Figure 15–18.

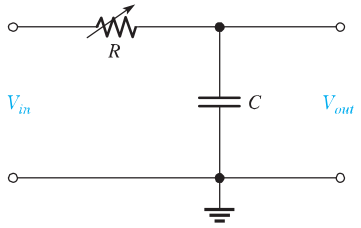

FIGURE 15–17 **Integrator with a variable resistor to control the time constant.**

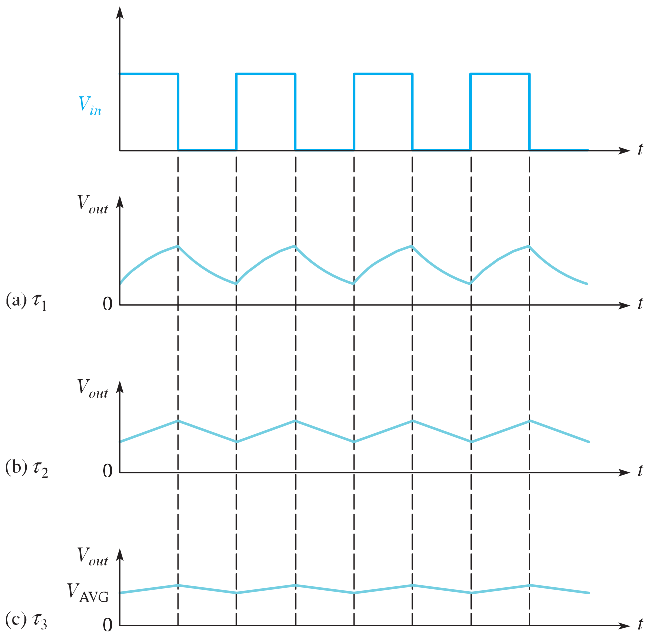

FIGURE 15–18 **Effect of increasing time constants on the output of an integrator ($\tau_3 > \tau_2 > \tau_1$).**

As the time constant becomes extremely long compared to the pulse width, the output voltage approaches a constant dc voltage, as shown in Figure 15–18(c). This value is the average value of the input. For a square wave, it is one-half the amplitude.

EXAMPLE 15–3

Determine the output voltage waveform for the first two pulses applied to the integrator circuit in Figure 15–19. Assume that the capacitor is initially uncharged.

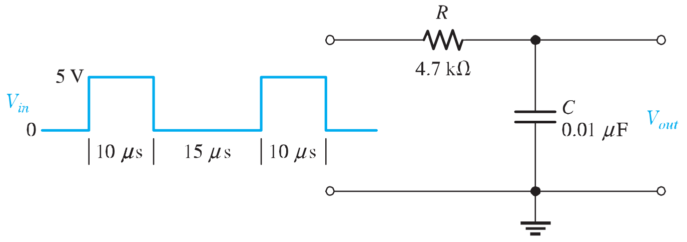

FIGURE 15–19

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

First, calculate the circuit time constant.

$$\tau = RC = (4.7\ \text{k}\Omega)(0.01\ \mu\text{F}) = 47\ \mu\text{s}$$

Obviously, the time constant is much longer than the input pulse width or the interval between pulses (notice that the input is not a square wave). Thus, in this case, the exponential formulas must be applied, and the analysis is relatively difficult. Follow the solution carefully.

1. *Calculation for first pulse:* Use the equation for an increasing exponential (Eq. 9–15) because C is charging. Note that V_F is 5 V, and t equals the pulse width of 10 μs. Therefore,

$$v_C = V_F(1 - e^{-t/RC}) = (5\ \text{V})(1 - e^{-10\mu\text{s}/47\mu\text{s}}) = 958\ \text{mV}$$

This result is plotted in Figure 15–20(a).

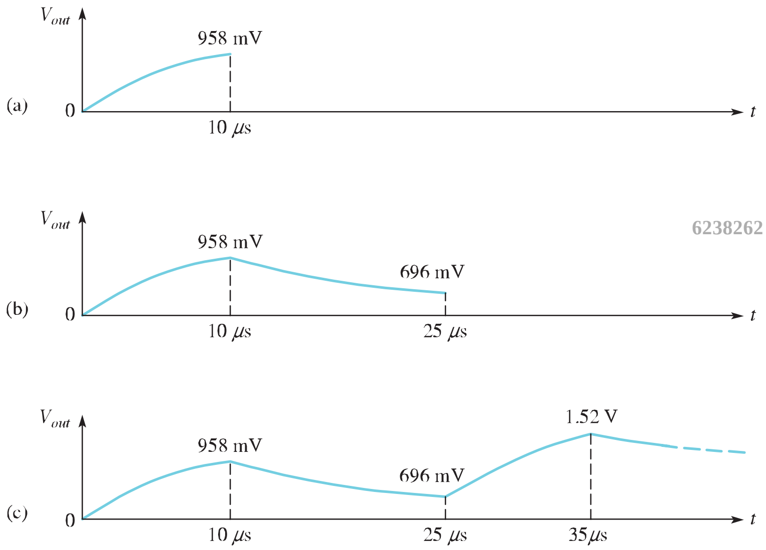

FIGURE 15–20

6238262 2013/12/10 12.27.66.5

2. *Calculation for interval between first and second pulse:* Use the equation for a decreasing exponential (Eq. 9–16) because C is discharging. Note that V_i is 958 mV because C begins to discharge from this value at the end of the first pulse. The discharge time is 15 μs. Therefore,

$$v_C = V_i e^{-t/RC} = (958\ \text{mV})e^{-15\mu\text{s}/47\mu\text{s}} = 696\ \text{mV}$$

This result is shown in Figure 15–20(b).

3. *Calculation for second pulse:* At the beginning of the second pulse, the output voltage is 696 mV. During the second pulse, the capacitor will again charge. In this case, it does not begin at zero volts. It already has 696 mV from the previous charge and discharge. To handle this situation, you must use the general exponential formula (Eq. 9–13).

$$v = V_F + (V_i - V_F)e^{-t/\tau}$$

Using this equation, you can calculate the voltage across the capacitor at the end of the second pulse.

$$v_C = V_F + (V_i - V_F)e^{-t/RC} = 5\ \text{V} + (696\ \text{mV} - 5\ \text{V})e^{-10\mu\text{s}/47\mu\text{s}} = 1.52\ \text{V}$$

This result is shown in Figure 15–20(c).

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Notice that the output waveform builds up on successive input pulses. After approximately 5τ, it will reach its steady state and will fluctuate between a constant maximum and a constant minimum, with an average equal to the average value of the input. You can demonstrate this pattern by carrying the analysis in this example further.

RELATED PROBLEM

Determine V_{out} at the beginning of the third pulse.

MULTISIM

Open Multisim file E15-03. Measure the steady-state output waveform in terms of its minimum, maximum, and average values.

SYSTEM EXAMPLE 15–1

A SWEPT-FREQUENCY TEST

Communication systems have specific requirements for the frequency response of the transmitter and receiver. To test the response of the transmitter or receiver, an oscilloscope can be set up to observe the frequency response in real time. The oscilloscope's horizontal axis (the *time base*) is converted to a frequency base by putting the scope in x-y mode. The oscilloscope can then plot the response of the unit under test to a test signal. A block diagram of the swept-frequency test system is shown in Figure 15–21.

The swept-frequency generator supplies the test signal. It consists of three blocks that generate a frequency-modulated (FM) signal. Pulses from a square wave generator are converted to a triangle wave by an integrator (an RC circuit in which $\gg T$.) The triangle wave is amplified and used to change the frequency of a voltage-controlled oscillator (VCO). Simultaneously, the triangle wave moves the x-axis of the scope in sync with the frequency change. This causes the oscilloscope time base to become a frequency base. This signal varies continuously between two selected radio frequencies (RF) that are used for the test. The unit under test is the portion of a communication system that is being aligned.

The unit under test is part of a communication system. It responds to the frequency modulated signal as determined by its frequency response. The output is detected (the rf signal is removed), and the envelope is observed on the oscilloscope, which shows the frequency response. The unit under test can be adjusted while observing the response in real time.

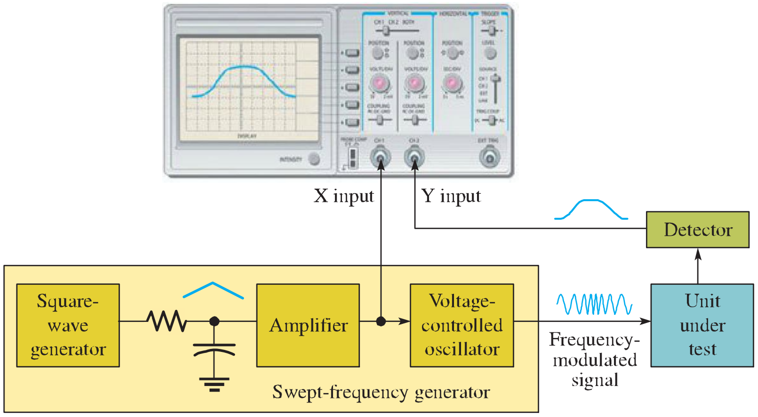
X input
Y input
Detector
Square-wave generator
Amplifier
Voltage-controlled oscillator
Unit under test
Frequency-modulated signal
Swept-frequency generator

FIGURE 15–21 **A swept-frequency test.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 15–3 CHECKUP

1. What conditions allow an *RC* integrator capacitor to fully charge and discharge when a periodic pulse waveform is applied to the input?
2. What will the output waveform look like if the circuit time constant is extremely small compared to the pulse width of a square wave input?
3. What is the time called that is required for the output voltage to build up to a constant average value when 5τ is greater than the pulse width of an input square wave?
4. Define *steady-state response.*
5. Describe the output of an *RC* integrator when the input is a square wave that has a period that is much less than 1τ.
6. What is the purpose of the integrator in the swept frequency measurement in System Example 15-1?

15–4 RESPONSE OF *RC* DIFFERENTIATORS TO A SINGLE PULSE

In terms of time response, a series *RC* circuit in which the output voltage is taken across the resistor is known as a differentiator. Recall that in terms of frequency response, this particular series *RC* circuit is a high-pass filter. The term, *differentiator*, is derived from a mathematical function which this type of circuit approximates under certain conditions.

After completing this section, you should be able to

- **Analyze an *RC* differentiator with a single-pulse input**
  - **Describe the response at the rising edge of the input pulse**
  - **Determine the response during and at the end of a pulse for various relationships of the pulse width and the time constant**

Figure 15–22 shows an *RC* differentiator with a pulse input. The same action occurs in the differentiator as in the integrator, except the output voltage is taken across the resistor rather than the capacitor. The capacitor charges exponentially at a rate depending on the *RC* time constant. The shape of the differentiator's resistor voltage is determined by the charging and discharging action of the capacitor.

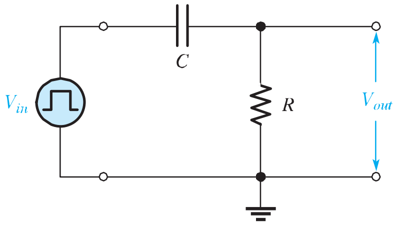

FIGURE 15–22 **An *RC* differentiator with a pulse generator connected.**

Pulse Response

To understand how the output voltage is shaped by a differentiator, you must consider the following:

1. The response to the rising pulse edge
2. The response between the rising and falling edges
3. The response to the falling pulse edge

Let's assume that the capacitor is initially uncharged prior to the rising pulse edge. Prior to the pulse, the input is zero volts. Thus, there are zero volts across the capacitor and also zero volts across the resistor, as indicated in Figure 15–23(a).

RESPONSE TO THE RISING EDGE OF THE INPUT PULSE

Let's also assume that a 10 V pulse is applied to the input. When the rising edge occurs, the voltage at point *A* goes to +10 V. Recall that the voltage across a capacitor cannot change instantaneously, and thus the capacitor appears instantaneously as a short. Therefore, if the

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

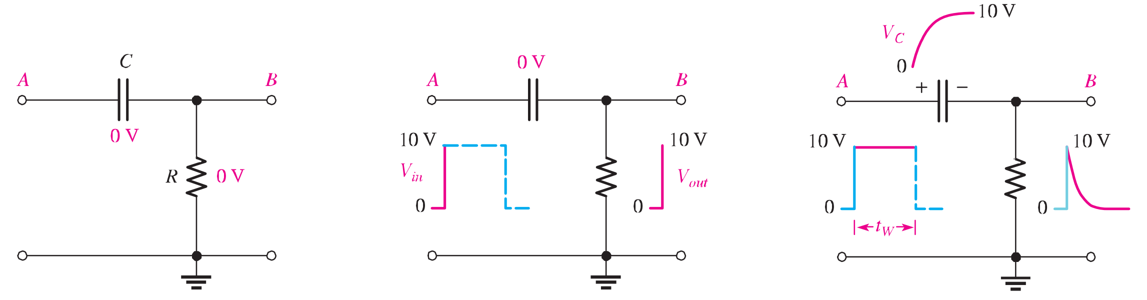

(a) Before pulse is applied

(b) At rising edge of input pulse

(c) During level part of pulse when $t_W \geq 5\tau$

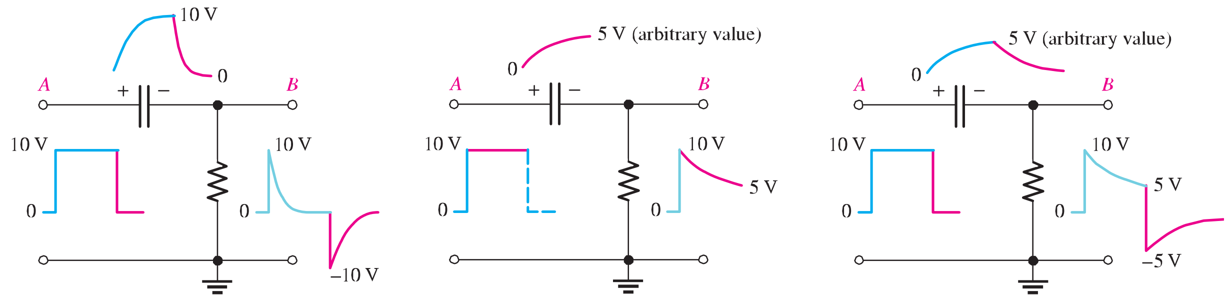

(d) At falling edge of pulse when $t_W \geq 5\tau$

(e) During level part of pulse when $t_W < 5\tau$

(f) At falling edge of pulse when $t_W < 5\tau$

FIGURE 15–23 **Examples of the response of a differentiator to a single-pulse input under two conditions: $t_W \geq 5\tau$ and $t_W < 5\tau$. A pulse generator is connected to the input, but the symbol is not shown, only the pulses.**

voltage at point *A* goes instantly to +10 V, then the voltage at point *B must* also go instantly to +10 V, keeping the capacitor voltage zero for the instant of the rising edge. The capacitor voltage is the voltage from point *A* to point *B*.

The voltage at point *B* with respect to ground is the voltage across the resistor (and the output voltage). Thus, the output voltage suddenly goes to +10 V in response to the rising pulse edge, as indicated in Figure 15–23(b).

RESPONSE DURING PULSE WHEN $t_W \geq 5\tau$ While the pulse is at its high level between the rising edge and the falling edge, the capacitor is charging. When the pulse width is equal to or greater than five time constants ($t_W \geq 5\tau$), the capacitor has time to fully charge.

As the voltage across the capacitor builds up exponentially, the voltage across the resistor decreases exponentially until it reaches zero volts at the time the capacitor reaches full charge (+10 V in this case). This decrease in the resistor voltage occurs because the sum of the capacitor voltage and the resistor voltage at any instant must be equal to the source voltage, in compliance with Kirchhoff's voltage law ($v_C + v_R = v_{in}$). This part of the response is illustrated in Figure 15–23(c).

RESPONSE TO FALLING EDGE WHEN $t_W \geq 5\tau$ Let's examine the case in which the capacitor is fully charged at the end of the pulse ($t_W \geq 5\tau$). Refer to Figure 15–23(d). On the falling edge, the input pulse suddenly goes from +10 V back to zero. An instant before the falling edge, the capacitor is charged to 10 V, so the voltage at point *A* is +10 V and the voltage at point *B* is 0 V. The voltage across a capacitor cannot change instantaneously, so when the voltage at point *A* makes a transition from +10 V to zero on the falling edge, the voltage at point *B must* also make a 10 V transition from zero to −10 V. This keeps the voltage across the capacitor at 10 V for the instant of the falling edge.

The capacitor now begins to discharge exponentially. As a result, the resistor voltage goes from −10 V to zero in an exponential curve, as indicated in Figure 15–23(d).

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

RESPONSE DURING PULSE WHEN $t_W < 5\tau$ When the pulse width is less than five time constants ($t_W < 5\tau$), the capacitor does not have time to fully charge. Its partial charge depends on the relationship of the time constant and the pulse width.

Because the capacitor does not reach the full +10 V, the resistor voltage will not reach zero volts by the end of the pulse. For example, if the capacitor charges to +5 V during the pulse interval, the resistor voltage will decrease to +5 V, as illustrated in Figure 15–23(e).

RESPONSE TO FALLING EDGE WHEN $t_W < 5\tau$ Now, let's examine the case in which the capacitor is only partially charged at the end of the pulse ($t_W < 5\tau$). For example, if the capacitor charges to +5 V, the resistor voltage at the instant before the falling edge is also +5 V because the capacitor voltage plus the resistor voltage must add up to +10 V, as illustrated in Figure 15–23(e).

When the falling edge occurs, the voltage at point *A* goes from +10 V to zero. As a result, the voltage at point *B* goes from +5 V to −5 V, as illustrated in Figure 15–23(f). This decrease occurs, of course, because the capacitor voltage cannot change at the instant of the falling edge. Immediately after the falling edge, the capacitor begins to discharge to zero. As a result, the resistor voltage goes from −5 V to zero, as shown.

Summary of *RC* Differentiator Response to a Single Pulse

A good way to summarize this section is to look at the general output waveforms of a differentiator as the time constant is varied from one extreme, when 5τ is much less than the pulse width, to the other extreme, when 5τ is much greater than the pulse width. These situations are illustrated in Figure 15–24. In part (a), the output consists of narrow positive and negative "spikes." These short pulses are useful in many systems to provide a timing signal called a "trigger." In part (e), the output approaches the shape of the input. Various conditions between these extremes are illustrated in parts (b), (c), and (d).

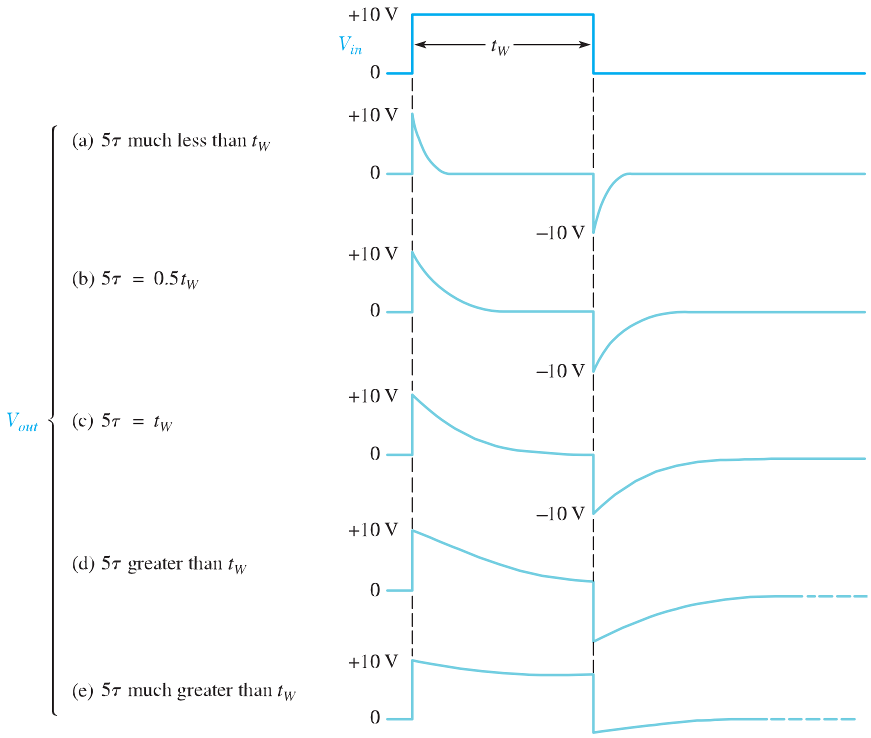

FIGURE 15–24 Effects of a change in time constant on the shape of the output voltage of an *RC* differentiator.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 15–4

Show the output voltage for the circuit in Figure 15–25.

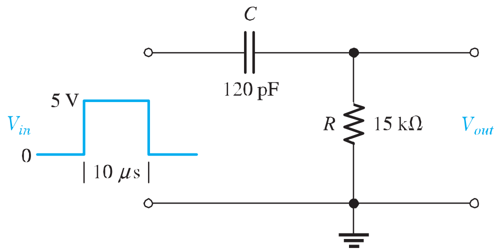

FIGURE 15–25

SOLUTION

First , calculate the time constant.

$$\tau = RC = (15\ \text{k}\Omega)(120\ \text{pF}) = 1.8\ \mu\text{s}$$

In this case, $t_W > 5\tau$, so the capacitor reaches full charge in 9 μs (before the end of the pulse).

On the rising edge, the resistor voltage jumps to +5 V and then decreases exponentially to zero before the end of the pulse. On the falling edge, the resistor voltage jumps to −5 V and then goes back to zero exponentially. The resistor voltage is, of course, the output, and its shape is shown in Figure 15–26.

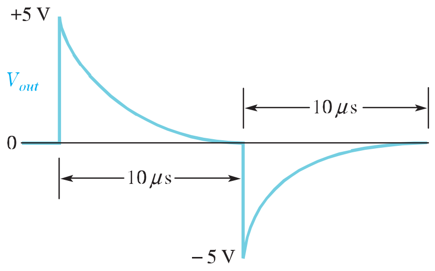

FIGURE 15–26

RELATED PROBLEM

Show the output voltage if $R = 18\ \text{k}\Omega$ and $C = 47\ \text{pF}$ in Figure 15–25.

EXAMPLE 15–5

Determine the output voltage waveform for the differentiator in Figure 15–27.

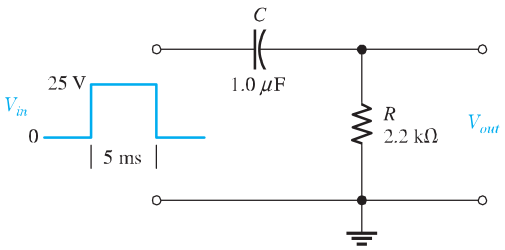

FIGURE 15–27

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

First, calculate the time constant.

$$\tau = (2.2\ \text{k}\Omega)(1.0\ \mu\text{F}) = 2.2\ \text{ms}$$

On the rising edge, the resistor voltage immediately jumps to +25 V. Because the pulse width is 5 ms, the capacitor charges for approximately 2.27 time constants during the pulse, and therefore does not reach full charge. Thus, you must use the formula for a decreasing exponential (Eq. 9–16) in order to calculate to what voltage the output decreases by the end of the pulse.

6238262/2013/12/10 12.27.66.5

$$v_{out} = V_i e^{-t/RC} = (25\ \text{V})e^{-5\text{ms}/2.2\text{ms}} = 2.58\ \text{V}$$

where $V_i = 25$ V and $t = 5$ ms. This calculation gives the resistor voltage at the end of the 5 ms pulse width interval.

On the falling edge, the resistor voltage immediately jumps from +2.58 V down to −22.4 V (a 25 V transition). The resulting waveform of the output voltage is shown in Figure 15–28.

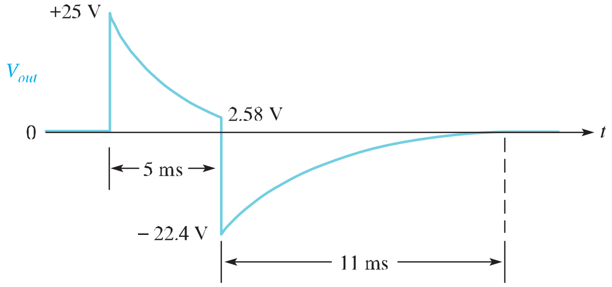

FIGURE 15–28

RELATED PROBLEM

Determine the voltage at the end of the pulse in Figure 15–27 if *R* is 1.5 kΩ.

SECTION 15–4 CHECKUP

1. Show the output of a differentiator for a 10 V input pulse when $5\tau = 0.5t_W$.
2. Under what condition does the output pulse shape most closely resemble the input pulse for a differentiator?
3. What does the differentiator output look like when 5τ is much less than the pulse width of the input?
4. If the resistor voltage in a differentiating circuit is down to +5 V at the end of a 15 V input pulse, to what negative value will the resistor voltage go in response to the falling edge of the input?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

15–5 RESPONSE OF *RC* DIFFERENTIATORS TO REPETITIVE PULSES

The *RC* differentiator response to a single pulse, covered in the last section, is extended in this section to repetitive pulses.

After completing this section, you should be able to

- **Analyze an *RC* differentiator with a repetitive-pulse input**
  - **Determine the response when the pulse width is less than five time constants**

If a periodic pulse waveform is applied to an *RC* differentiating circuit, two conditions again are possible: $t_W \geq 5\tau$ or $t_W < 5\tau$. Figure 15–29 shows the output when $t_W = 5\tau$. As the time constant is reduced, both the positive and the negative portions of the output become narrower. Notice that the average value of the output is zero; the waveform has equal positive and negative portions. The average value of a waveform is its **dc component**. Because a capacitor blocks dc, the dc component of the input is prevented from passing through to the output.

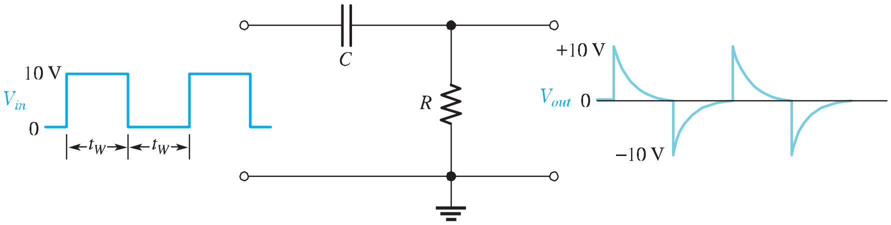

FIGURE 15–29 **Example of differentiator response when $t_W = 5\tau$.**

Figure 15–30 shows the steady-state output when $t_W < 5\tau$. As the time constant is increased, the positively and negatively sloping portions become flatter. For a very long time constant, the output approaches the shape of the input, but with an average value of zero.

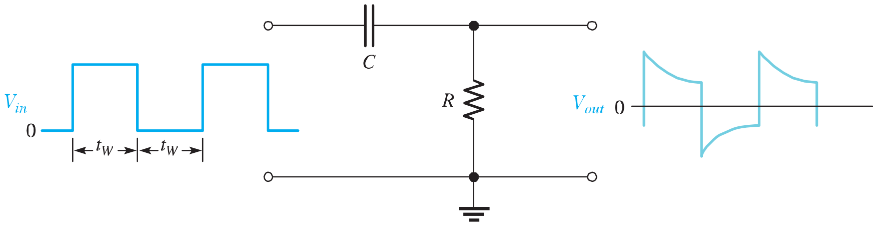

FIGURE 15–30 **Example of differentiator response when $t_W < 5\tau$.**

Analysis of a Repetitive Waveform

Like the integrator, the differentiator output takes time (5τ) to reach steady state. To illustrate the response, let's take an example in which the time constant equals the input pulse width. At this point, we do not care what the value of the circuit time constant is because we know that the resistor voltage will decrease to 36.8% of its maximum value during one pulse (1τ). Let's assume that the capacitor in Figure 15–31 begins initially uncharged and examine the output voltage on a pulse-by-pulse basis. The results of this analysis are shown in Figure 15–32.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

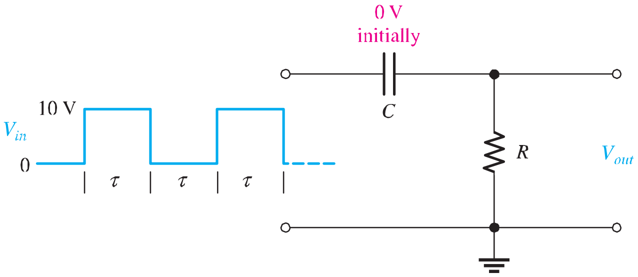

FIGURE 15–31 ***RC* differentiator with $T = 2\tau$.**

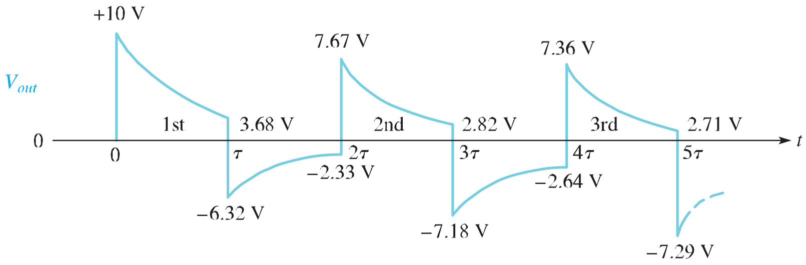

FIGURE 15–32 **Differentiator output waveform during transient time for the circuit in Figure 15–31.**

First pulse On the rising edge, the output instantaneously jumps to +10 V. Then the capacitor partially charges to 63.2% of 10 V, which is 6.32 V. Thus, the output voltage must decrease to 3.68 V, as shown in Figure 15–32.

On the falling edge, the output instantaneously makes a negative-going 10 V transition to −6.32 V because 3.68 V − 10 V = −6.32 V.

Between first and second pulses The capacitor discharges to 36.8% of 6.32 V, which is 2.33 V. Thus, the resistor voltage, which starts at −6.32 V, must increase to −2.33 V. Why? Because at the instant prior to the next pulse, the input voltage is zero. Therefore, the sum of v_C and v_R must be zero (+2.33 V − 2.33 V = 0). Remember that $v_C + v_R = v_{in}$ at all times, in accordance with Kirchhoff's voltage law.

Second pulse On the rising edge, the output makes an instantaneous, positive-going, 10 V transition from −2.33 V to 7.67 V. Then by the end of the pulse the capacitor charges 0.632 (10 V − 2.33 V) = 4.85 V. Thus, the capacitor voltage increases from 2.33 V to 2.33 V + 4.85 V = 7.18 V. The output voltage drops to 0.368(7.67 V) = 2.82 V.

On the falling edge, the output instantaneously makes a negative-going transition from 2.82 V to −7.18 V, as shown in Figure 15–32.

Between second and third pulses The capacitor discharges to 36.8% of 7.18 V, which is 2.64 V. Thus, the output voltage starts at −7.18 V and increases to −2.64 V because the capacitor voltage and the resistor voltage must add up to zero at the instant prior to the third pulse (the input is zero).

Third pulse On the rising edge, the output makes an instantaneous 10 V transition from −2.64 V to +7.36 V. Then the capacitor charges 0.632(10 V − 2.64 V) = 4.65 V to 2.64 V + 4.65 V = +7.29 V. As a result, the output voltage drops to 0.368(7.36 V) = 2.71 V. On the falling edge, the output instantly goes from +2.71 V down to −7.29 V.

After the third pulse, five time constants have elapsed, and the output voltage is close to its steady state. Thus, the waveform in Figure 15–32 will continue to vary from a positive maximum of about +7.3 V to a negative maximum of about −7.3 V, with an average value of zero.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6338622 2015/12/10 12.27.66.5

SYSTEM EXAMPLE 15–2

A WATCHDOG TIMER

Many systems require constant monitoring in order to take immediate action in case a power failure should occur. The circuit that monitors for a particular condition is called a *watchdog timer*, a colorful description of a monitoring circuit, usually one that can take a corrective action. For the case of a power failure monitor, there is usually enough stored power in the capacitors to take the action before the system shuts down. The action can be anything from switching to backup power, saving files, or sounding a battery-powered alarm. There is a wide variety of watchdog timer circuits, including some that are built into integrated circuits.

In this example, a watchdog timer is monitoring the input power and outputs a negative-going trigger signal when a failure occurs. The trigger can be used to start a corrective action. A key part of the circuit is two differentiating circuits. Figure 15–33 shows the basic circuit for this watchdog timer.

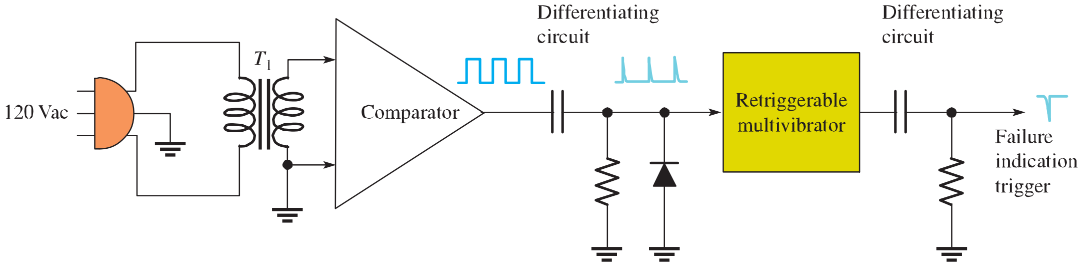

FIGURE 15–33 **A watchdog timer circuit for monitoring power.**

The circuit operation is to convert ac to a low voltage by T_1 and compare the low-voltage ac signal with ground. The comparator outputs a high signal when the ac is above ground and a low signal when it is below ground, thus producing a square wave. The first differentiating circuit and diode produce a series of triggers, which occur at the power line frequency. The triggers are sent to a *retriggerable multivibrator*, a circuit that has a high dc level output as long as triggers appear regularly at the input. If a power failure occurs, the dc drops to ground level and a second differentiating circuit converts this drop to a negative-going trigger, which can be used to take a corrective action.

SECTION 15–5 CHECKUP

1. What conditions allow an *RC* differentiator to fully charge and discharge when a periodic pulse waveform is applied to the input?
2. What will the output waveform look like if the circuit time constant is extremely small compared to the pulse width of a square wave input?
3. What does the average value of the differentiator output voltage equal during steady state?
4. Explain the purpose of each differentiator in the watchdog timer.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

15–6 RESPONSE OF *RL* INTEGRATORS TO PULSE INPUTS

A series *RL* circuit in which the output voltage is taken across the resistor is known as an integrator in terms of time response. The response to a single pulse is discussed but can be extended to repetitive pulses as was done for the case of *RC* integrators. Although the *RL* integrator has waveforms similar to the *RC* integrator, the *RL* integrator is not as commonly used; inductors tend to be more expensive than capacitors and have winding resistance that can adversely affect the circuit.

After completing this section, you should be able to

- **Analyze the operation of an *RL* integrator**
  - **Determine the response to a single-pulse input**

Figure 15–34 shows an *RL* integrator. The output waveform is taken across the resistor and, under equivalent conditions, is the same shape as that for the *RC* integrator. Recall that in the *RC* case, the output was across the capacitor.

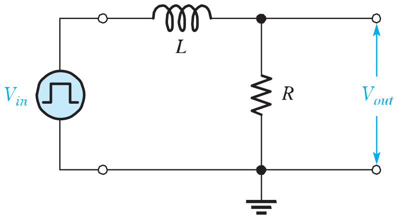

FIGURE 15–34 **An *RL* integrator with a pulse generator connected.**

As you know, each edge of an ideal pulse is considered to occur instantaneously. Two basic rules for inductor behavior will aid in analyzing *RL* circuit responses to pulse inputs:

1. The inductor appears as an open to an instantaneous change in current and as a short (ideally) to dc.
2. The current in an inductor cannot change instantaneously—it can change only exponentially when the input changes levels.

Response of the *RL* Integrator to a Single Pulse

When a pulse generator is connected to the input of an *RL* integrator and the voltage pulse goes from its low level to its high level, the inductor prevents a sudden change in current. As a result, the inductor acts as an open, and all the input voltage is across it at the instant of the rising pulse edge. This situation is indicated in Figure 15–35(a).

After the rising edge, the current builds up, and the output voltage follows the current as it increases exponentially, as shown in Figure 15–35(b). The current can reach a maximum of V_p/R if the transient time is shorter than the pulse width ($V_p = 10$ V in this example where V_p is pulse amplitude).

When the pulse goes from its high level to its low level, an induced voltage with reversed polarity is created across the coil in an effort to keep the current equal to V_p/R. The output voltage begins to decrease exponentially, as shown in Figure 15–35(c).

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

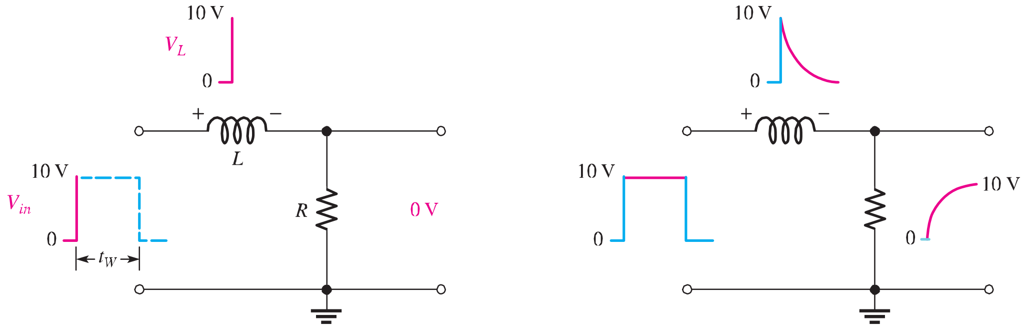

(a) At rising edge of pulse ($i = 0$)

(b) During flat portion of pulse

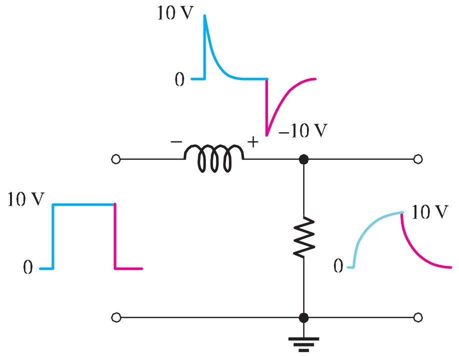

(c) At falling edge of pulse and after

FIGURE 15–35 **Illustration of the pulse response of an *RL* integrator ($t_W > 5\tau$). A pulse generator is connected to the input but its symbol is not shown, only the pulses.**

The exact shape of the output depends on the L/R time constant as summarized in Figure 15–36 for various relationships of the time constant and the pulse width. You should note that the response of this RL circuit in terms of the shape of the output is identical to that of the RC integrator. The relationship of the L/R time constant to the input pulse width has the same effect as the RC time constant that we discussed earlier in this chapter. For example, when $t_W < 5\tau$, the output voltage will not reach its maximum possible value.

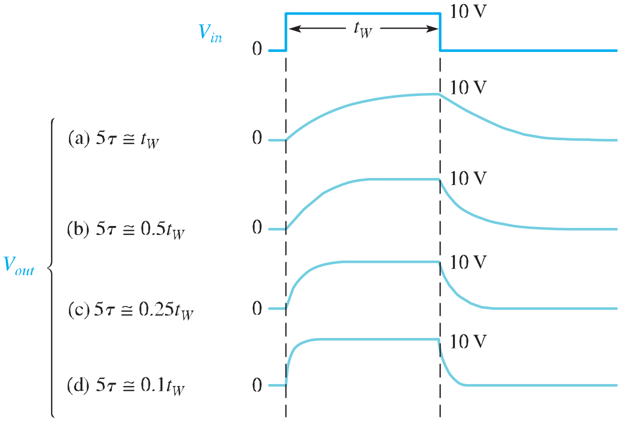

FIGURE 15–36 **Illustration of the variation in integrator output pulse shape with time constant.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 15–6

Determine the maximum output voltage for the integrator in Figure 15–37 when a single pulse is applied as shown.

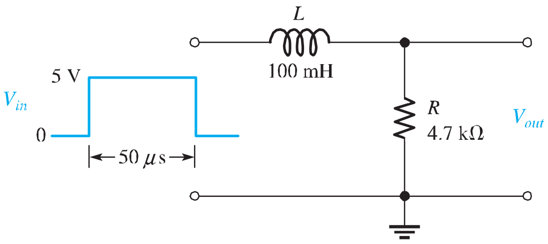

FIGURE 15–37

SOLUTION

Calculate the time constant.

$$\tau = \frac{L}{R} = \frac{100\text{ mH}}{4.7\text{ k}\Omega} = 21.3\ \mu\text{s}$$

Because the pulse width is 50 μs, the inductor's magnetic field increases (energizes) for approximately 2.35τ; $(50\ \mu\text{s}/21.3\ \mu\text{s} = 2.35)$. Use Equation 11–6 to calculate the voltage.

$$v = V_F + (V_i - V_F)e^{-Rt/L}$$

V_i is zero. The maximum output voltage occurs at the end of the pulse. Therefore,

$$v_L = V_F(1 - e^{-t/\tau}) = 5\text{ V}(1 - e^{-50\ \mu\text{s}/21.3\ \mu\text{s}}) = \mathbf{4.52\ V}$$

RELATED PROBLEM

What value must *R* be for the output voltage to reach 5 V by the end of the pulse?

MULTISIM

Open the file E15-06. Use the space bar to trigger the pulse generator. Confirm the calculations shown in the solution.

EXAMPLE 15–7

A pulse is applied to the *RL* integrator in Figure 15–38. Determine the waveforms and the values for I, V_R, and V_L.

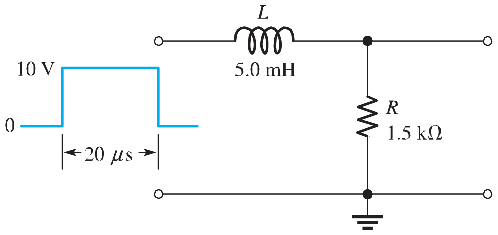

FIGURE 15–38

SOLUTION

The circuit time constant is

$$\tau = \frac{L}{R} = \frac{5.0\text{ mH}}{1.5\text{ k}\Omega} = 3.33\ \mu\text{s}$$

6238262 2013/12/10 12.27.66.5

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Since $5\tau = 16.7\ \mu s$ is less than t_W, the current will reach its maximum value and remain there until the end of the pulse.

At the rising edge of the pulse,

$$i = 0\text{ A}$$
$$v_R = 0\text{ V}$$
$$v_L = 10\text{ V}$$

The inductor appears as an open, so all of the input voltage appears across L.

During the pulse,

$$i \text{ increases exponentially to } \frac{V_p}{R} = \frac{10\text{ V}}{1.5\text{ k}\Omega} = 6.67\text{ mA in } 16.7\ \mu\text{s}$$

v_R increases exponentially to 10 V in 16.7 μs

v_L increases exponentially to zero in 16.7 μs

At the falling edge of the pulse,

$$i = 6.67\text{ mA}$$
$$v_R = 10\text{ V}$$
$$v_L = -10\text{ V}$$

After the pulse,

i decreases exponentially to zero in 16.7 μs

v_R decreases exponentially to zero in 16.7 μs

v_L decreases exponentially to zero in 16.7 μs

The waveforms are shown in Figure 15–39.

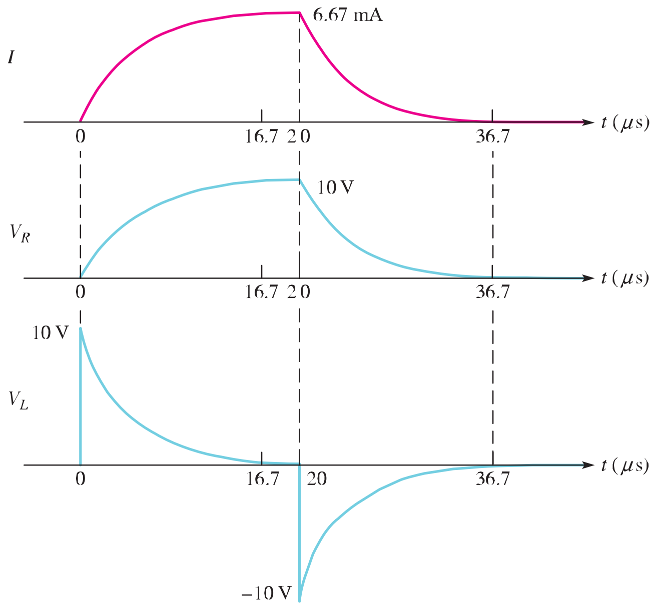

FIGURE 15–39

RELATED PROBLEM

What will be the maximum output voltage if the amplitude of the input pulse is increased to 20 V in Figure 15–38?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE 15–8

A 10 V pulse with a width of 10 μs is applied to the integrator in Figure 15–40. Determine the voltage level that the output will reach during the pulse. If the source has an internal resistance of 300 Ω, how long will it take the output to decrease to zero? Show the output voltage waveform.

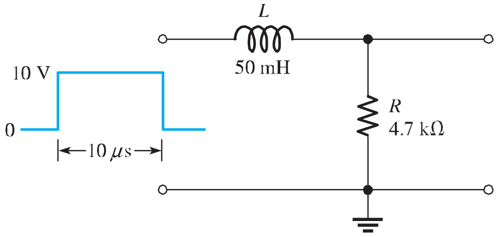

FIGURE 15–40

SOLUTION

The coil energizes through the 300 Ω source resistance plus the 4.7 kΩ external resistor. The time constant is

$$\tau = \frac{L}{R_{tot}} = \frac{50\text{ mH}}{4700\ \Omega + 300\ \Omega} = \frac{50\text{ mH}}{5.0\text{ k}\Omega} = 10\ \mu\text{s}$$

Notice that in this case the pulse width is exactly equal to τ. Thus, the output V_R will reach approximately 63% of the full input amplitude in 1τ. Therefore, the output voltage gets to **6.3 V** at the end of the pulse.

After the pulse is gone, the inductor de-energizes back through the 300 Ω source resistance and the 4.7 kΩ resistor. The output voltage takes 5τ to completely decrease to zero.

$$5\tau = 5(10\ \mu\text{s}) = \mathbf{50\ \mu s}$$

The output voltage is shown in Figure 15–41.

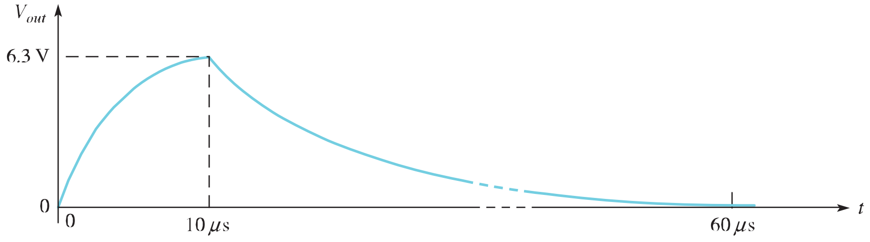

FIGURE 15–41

RELATED PROBLEM

To what approximate maximum value must *R* be changed in Figure 15–40 to allow the output voltage to reach the input level during the pulse?

SECTION 15–6 CHECKUP

1. In an *RL* integrator, across which component is the output voltage taken?
2. When a pulse is applied to an *RL* integrator, what condition must exist in order for the output voltage to reach the amplitude of the input?
3. Under what condition will the output voltage have the approximate shape of the input pulse?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

15–7 RESPONSE OF *RL* DIFFERENTIATORS TO PULSE INPUTS

A series *RL* circuit in which the output voltage is taken across the inductor is known as a differentiator in terms of time response. The response to a single pulse is discussed but can be extended to repetitive pulses as was done for the case of *RC* differentiators.

After completing this section, you should be able to

- **Analyze the operation of an *RL* differentiator**
  - **Determine the response to a single-pulse input**

Response of the *RL* Differentiator to a Single Pulse

Figure 15–42 shows an *RL* differentiator with a pulse generator connected to the input.

FIGURE 15–42 **An *RL* differentiator with a pulse generator connected.**

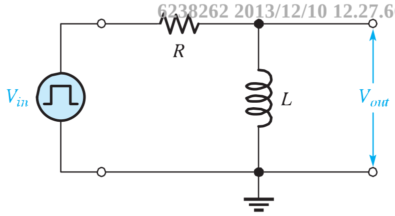

6238262 2013/12/10 12.27.66.5

Initially, before the pulse, there is no current in the circuit. When the input pulse goes from its low level to its high level, the inductor prevents a sudden change in current. It does this, as you know, with an induced voltage equal and opposite to the input. As a result, *L* looks like an open, and all of the input voltage appears across it at the instant of the rising edge, as shown in Figure 15–43(a) with a 10 V pulse.

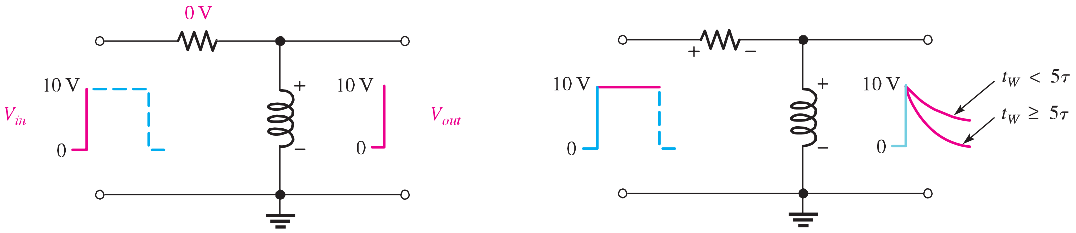

(a) At rising edge of pulse

(b) During flat portion of pulse

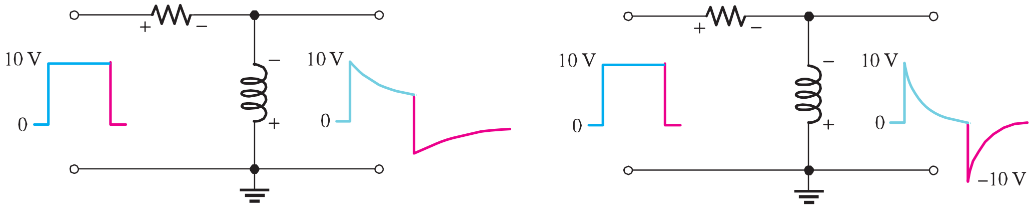

(c) At falling edge when $t_W < 5\tau$

(d) At falling edge when $t_W \geq 5\tau$

FIGURE 15–43 **Illustration of the response of an *RL* differentiator for both time constant conditions. The pulse generator is connected to the input, but its symbol is not shown, only the pulses.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

During the pulse, the current exponentially builds up. As a result, the inductor voltage decreases, as shown in Figure 15–43(b). The rate of decrease, as you know, depends on the L/R time constant. When the falling edge of the input occurs, the inductor reacts to keep the current as is, by creating an induced voltage in a direction as indicated in Figure 15–43(c). This reaction is seen as a sudden negative-going transition of the inductor voltage, as indicated in parts (c) and (d).

Two conditions are possible, as indicated in Figure 15–43(c) and (d). In part (c), 5τ is greater than the input pulse width, and the output voltage does not have time to decrease to zero. In part (d), 5τ is less than or equal to the pulse width, and so the output decreases to zero before the end of the pulse. In this case a full −10 V transition occurs at the falling edge.

Keep in mind that as far as the input and output waveforms are concerned, the *RL* integrator and differentiator perform the same as their *RC* counterparts.

A summary of the *RL* differentiator response for various relationships of time constants and pulse widths is shown in Figure 15–44. Compare this figure to Figure 15–24 for an *RC* differentiator and you will see that the responses are identical.

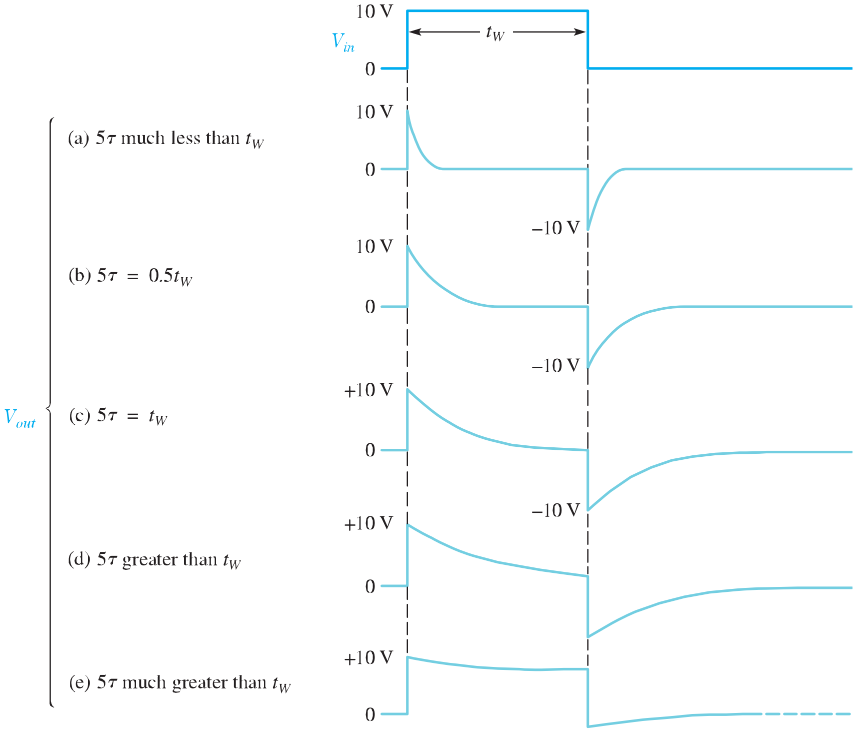

FIGURE 15–44 **Illustration of the variation in output pulse shape with the time constant.**

EXAMPLE 15–9

Show the output voltage waveform for the *RL* differentiator in Figure 15–45.

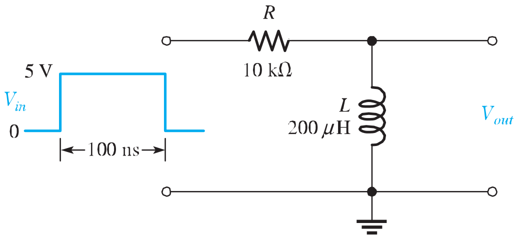

FIGURE 15–45

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

First, calculate the time constant.

$$\tau = \frac{L}{R} = \frac{200\ \mu\text{H}}{10\ \text{k}\Omega} = 20\ \text{ns}$$

In this case, $t_W = 5\tau$, so the output will decay to zero at the end of the pulse.

On the rising edge, the inductor voltage jumps to +5 V and then decays exponentially to zero. It reaches approximately zero at the instant of the falling edge. On the falling edge of the input, the inductor voltage jumps to −5 V and then goes back to zero. The output waveform is shown in Figure 15–46.

FIGURE 15–46

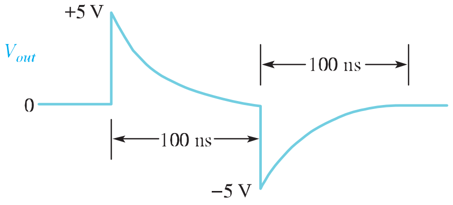

RELATED PROBLEM

Show the output voltage if the input pulse width is reduced to 50 ns in Figure 15–45.

EXAMPLE 15–10

Determine the output voltage waveform for the *RL* differentiator in Figure 15–47.

FIGURE 15–47

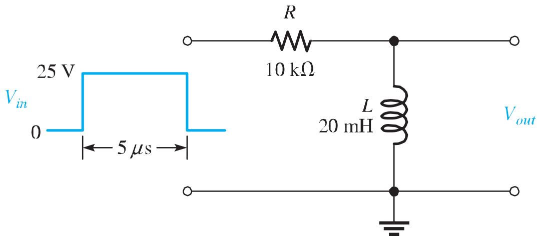

SOLUTION

First, calculate the time constant.

$$\tau = \frac{L}{R} = \frac{20\ \text{mH}}{10\ \text{k}\Omega} = 2\ \mu\text{s}$$

On the rising edge, the inductor voltage immediately jumps to +25 V, which is the initial value for the exponential decay. The final value is zero (assuming it could continue to decay). The exponential formula for the voltage at any instant on an inductor is Equation 11–6, which is

$$v = V_F + (V_i - V_F)e^{-Rt/L}$$

Substituting 0 for the final value gives the exponential decay formula:

$$v_L = V_i e^{-t/\tau} = (25\ \text{V})e^{-5\mu\text{s}/2/\mu\text{s}} = (25\ \text{V})e^{-2.5} = 2.05\ \text{V}$$

This result is the inductor voltage at the end of the 5 μs input pulse.

On the falling edge, the output immediately drops from +2.05 V to −22.95 V (a 25 V negative-going transition). The complete output waveform is shown in Figure 15–48.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

FIGURE 15–48

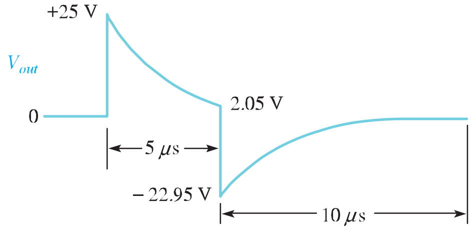

RELATED PROBLEM

What must be the value of R for V_{out} to reach zero by the end of the input pulse in Figure 15–47?

SECTION 15–7 CHECKUP

1. In an *RL* differentiator, across which component is the output taken?
2. Under what condition does the output pulse shape most closely resemble the input pulse?
3. If the inductor voltage in an *RL* differentiator is down to +2 V at the end of a +10 V input pulse, to what negative voltage will the output voltage go in response to the falling edge of the input?

15–8 INTEGRATOR AND DIFFERENTIATOR APPLICATIONS

Although integrators and differentiators are found in many different kinds of applications, only three basic ones are discussed in this section.

After completing this section, you should be able to

- **Discuss some applications of integrators and differentiators**
  - **Describe how *RC* integrators are used in timing circuits**
  - **Describe how an integrator can be used for converting pulse waveforms to dc**
  - **Describe how a differentiator can be used as a trigger pulse generator**

Timing Circuits

RC integrators can be used in timing circuits to set a specified time interval for various purposes. By changing the time constant, the interval can be adjusted. For example, the *RC* integrator circuit in Figure 15–49(a) is used to provide a delay between the time a power-on switch is closed and the time a certain event is activated by a threshold circuit. The threshold circuit is designed to respond to the input voltage only when the input reaches a certain specified level. Don't worry how the threshold circuit works at this point; you will learn about this type of circuit in a later course.

When the switch is thrown, the capacitor begins to charge at a rate set by the *RC* time constant. After a certain time, the capacitor voltage reaches the threshold value and triggers (turns on) the circuit to activate a circuit or a device such as a motor, relay, or lamp, depending on the particular application. A variable delay can be created by making *R* a rheostat. A circuit like this is useful in systems like radars to adjust certain features like "clutter filters." This action is illustrated by the waveforms in Figure 15–49(b).

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

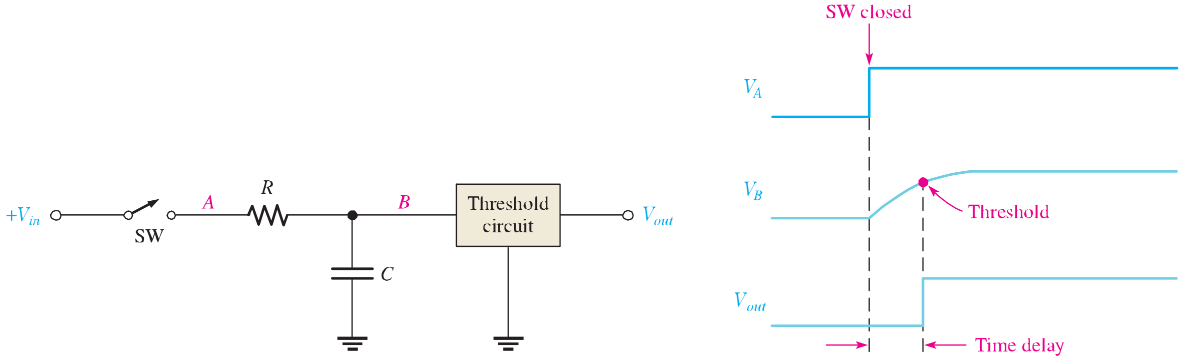

(a) Basic delay circuit

(b) Timing diagram

FIGURE 15–49 **A basic time delay application of an *RC* integrator.**

EXAMPLE 15–11

What is the time delay in the circuit of Figure 15–49(a) with $V_{in} = 9\text{ V}$, $R = 10\text{ M}\Omega$, and $C = 0.47\ \mu\text{F}$? Assume the threshold voltage is 5 V.

SOLUTION

Use the exponential formula, Equation 9–15, and solve for time as follows:

$$v = V_F(1 - e^{-t/RC}) = V_F - V_F e^{-t/RC}$$

$$V_F - v = V_F e^{-t/RC}$$

$$e^{-t/RC} = \frac{V_F - v}{V_F}$$

Taking the natural log (ln) of both sides of the equation yields

$$-\frac{t}{RC} = \ln\left(\frac{V_F - v}{V_F}\right)$$

$$t = -RC\ln\left(\frac{V_F - v}{V_F}\right)$$

V_F is the final voltage to which C will charge and equals V_{in}. Substituting values and solving for t yields

$$t = -(10\text{ M}\Omega)(0.47\ \mu\text{F})\ln\left(\frac{9\text{ V} - 5\text{ V}}{9\text{ V}}\right) = -(10\text{ M}\Omega)(0.47\ \mu\text{F})\ln\left(\frac{4\text{ V}}{9\text{ V}}\right) = \mathbf{3.8\text{ s}}$$

RELATED PROBLEM

Find the time delay for the circuit in Figure 15–49 for $V_{in} = 24\text{ V}$, $R = 10\text{ k}\Omega$, and $C = 1500\text{ pF}$.

A Pulse Waveform-to-DC Converter

An integrating circuit can be used to convert a pulse waveform to a constant dc value equal to the average of the waveform. This is accomplished by using a time constant that is extremely long compared to the period of the pulse waveform, as illustrated in Figure 15–50(a). Actually, there will be a slight ripple on the output as the capacitor charges and discharges a little. This ripple becomes less as the time constant becomes longer and can approach a constant level, as shown in Figure 15–50(b). In terms of sinusoidal response, this circuit is a low-pass filter.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

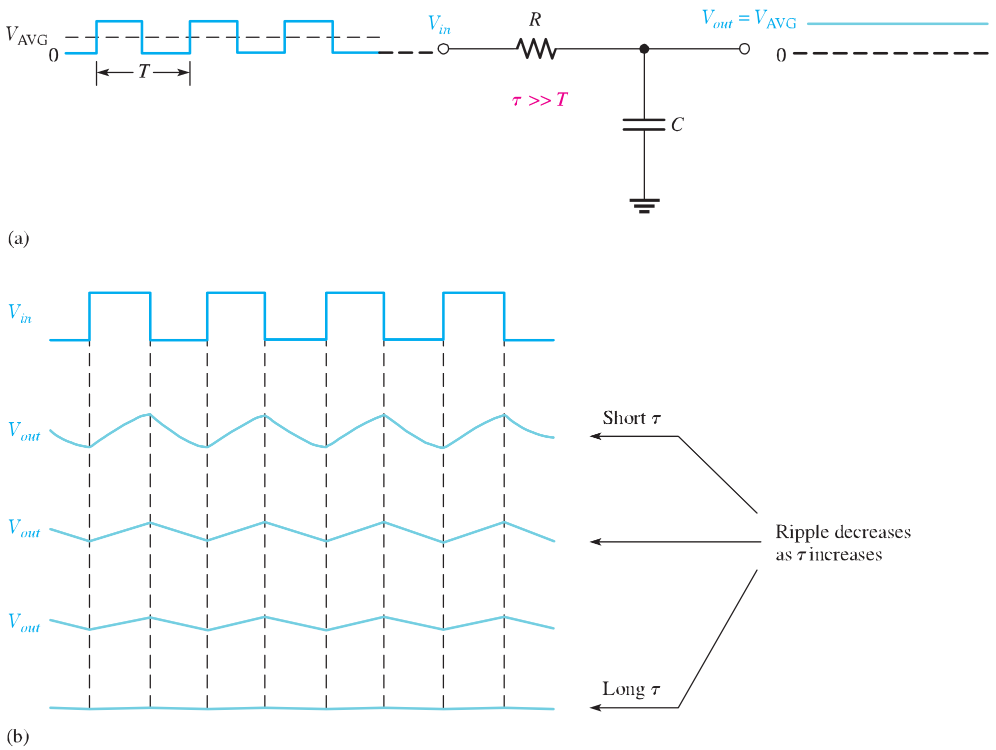

FIGURE 15–50 **A long time constant integrator acts as a pulse-to-dc converter.**

Trigger Pulse Generator and Waveshaping

A differentiator circuit can be used to produce very short duration pulses (spikes) with both positive and negative polarities, as you saw in System Example 15–2. When the time constant is short compared to the pulse width of the input, the differentiator produces a positive spike on each positive-going edge of the input and a negative spike on each negative-going edge, as shown in Figure 15–51. A diode can be used to remove either the positive trigger or the negative trigger.

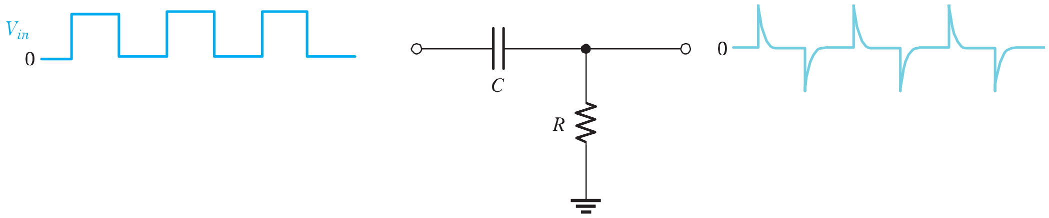

FIGURE 15–51 **A very short time constant differentiator generates positive and negative spikes.**

By changing the input to a triangle wave with a long period compared to the time constant, a wave-shaping circuit is produced. This time the output is a square wave that represents the rate of rise of the triangle. This idea is shown in Figure 15–52. There is a

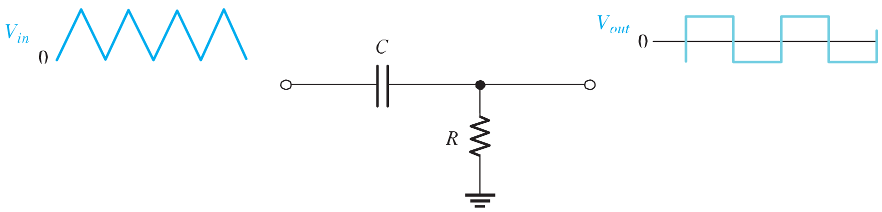

FIGURE 15–52 **Triangle to square wave circuit. The time constant of the circuit is long compared to the period.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

trade-off in this circuit between the rise time of the output square wave and the amplitude of the square wave. If the time constant is short, the square wave rises rapidly, but the amplitude is lower.

SECTION 15–8 CHECKUP

1. How can the delay time in the circuit of Figure 15–49 be increased?
2. Reducing the value of *R* in Figure 15–50 will (increase, decrease) the amount of ripple on the output.
3. What characteristic must the differentiator in Figure 15–51 have in order to produce very short duration spikes on its output?

15–9 TROUBLESHOOTING

In this section, *RC* circuits with pulse inputs are used to demonstrate the effects of common component failures in selected cases. The concepts can then be easily related to *RL* circuits.

After completing this section, you should be able to

- **Troubleshoot *RC* integrators and *RC* differentiators**
  - **Recognize the effect of an open capacitor**
  - **Recognize the effect of a shorted capacitor**
  - **Recognize the effect of an open resistor**

Open Capacitor

If the capacitor in an *RC* integrator opens, the output has the same waveshape as the input, as shown in Figure 15–53(a). If the capacitor in an *RC* differentiator opens, the output is zero because it is held at ground through the resistor, as illustrated in part (b).

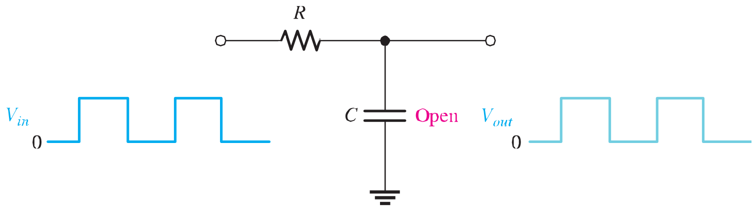

(a) Integrator

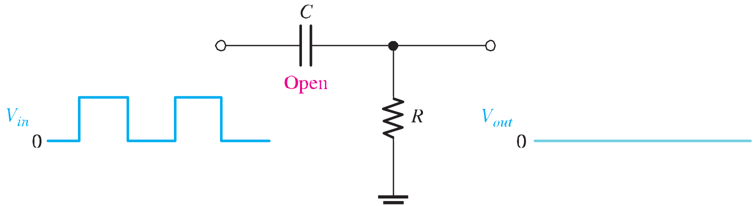

6238262 2013/12/10 12.27.66.5

(b) Differentiator

FIGURE 15–53 **Examples of the effect of an open capacitor.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Shorted Capacitor

If the capacitor in an *RC* integrator shorts, the output is at ground as shown in Figure 15–54(a). If the capacitor in an *RC* differentiator shorts, the output is the same as the input, as shown in part (b).

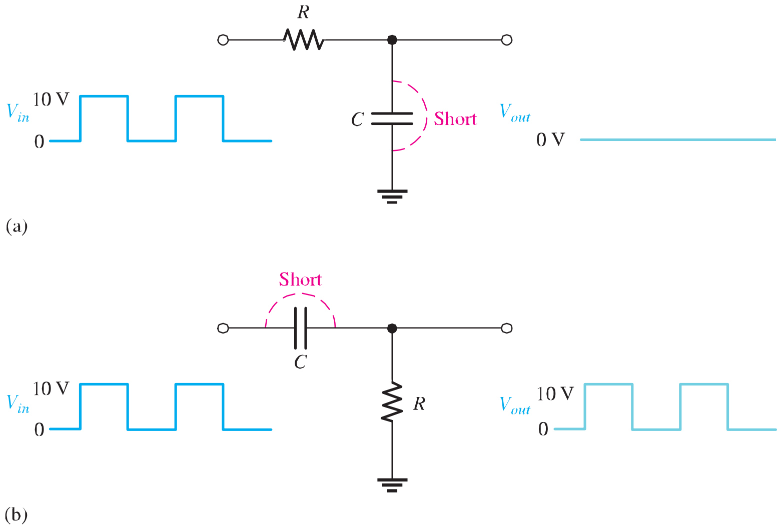
R
V_{in}
10 V
0
C
Short
V_{out}
0 V
(a)
Short
C
V_{in}
10 V
0
R
V_{out}
10 V
0
(b)

FIGURE 15–54 **Examples of the effect of a shorted capacitor.**

Open Resistor

If the resistor in an *RC* integrator opens, the capacitor has no discharge path, and, ideally, it will hold any charge it may have. In an actual situation, any charge will gradually leak off or the capacitor will discharge slowly through a measuring instrument connected to the output. This is illustrated in Figure 15–55(a).

If the resistor in an *RC* differentiator opens, the output looks like the input except for the dc level because the capacitor now must charge and discharge through the extremely high resistance of the oscilloscope, as shown in Figure 15–55(b).

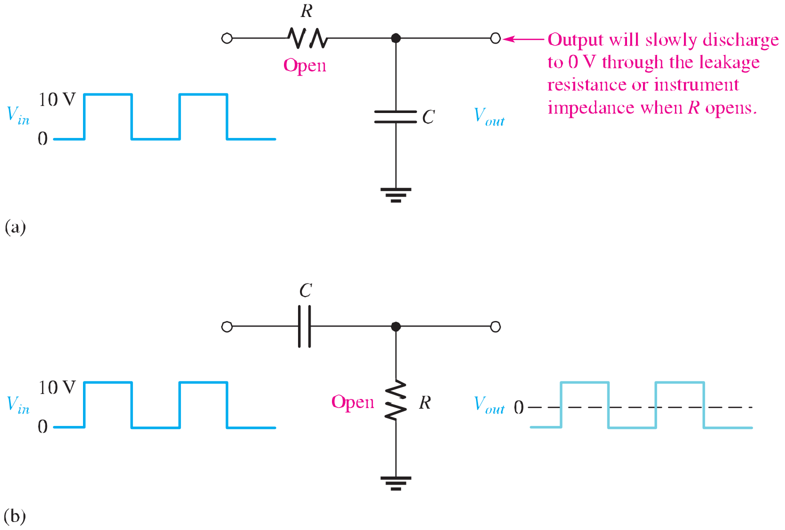
R
Open
Output will slowly discharge to 0 V through the leakage resistance or instrument impedance when R opens.
V_{in}
10 V
0
C
V_{out}
(a)
C
V_{in}
10 V
0
Open
R
V_{out}
0
(b)

FIGURE 15–55 **Examples of the effect of an open resistor.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 15–9 CHECKUP

1. An *RC* integrator has a zero output with a square wave input. What are the possible causes of this problem?
2. If the integrator capacitor is open, what is the output for a square wave input?
3. If the capacitor in a differentiator is shorted, what is the output for a square wave input?

SUMMARY

- In an *RC* integrator, the output voltage is taken across the capacitor.
- In an *RC* differentiator, the output voltage is taken across the resistor.
- In an *RL* integrator, the output voltage is taken across the resistor.
- In an *RL* differentiator, the output voltage is taken across the inductor.
- In an integrator, when the pulse width (t_W) of the input is much less than the transient time, the output voltage approaches a constant level equal to the average value of the input.
- In an integrator, when the pulse width of the input is much greater than the transient time, the output voltage approaches the shape of the input.
- In a differentiator, when the pulse width of the input is much less than the transient time, the output voltage approaches the shape of the input but with an average value of zero.
- In a differentiator, when the pulse width of the input is much greater than the transient time, the output voltage consists of narrow, positive-going and negative-going spikes occurring on the leading and trailing edges of the input pulses.
- *RC* integrators can be used to set up a specified time delay.

KEY TERMS

Key terms and other bold terms in the chapter are defined in the end-of-book glossary.

Differentiator	A circuit producing an output that approaches the mathematical derivative of the input.
Integrator	A circuit producing an output that approaches the mathematical integral of the input.
Steady state	The equilibrium condition of a circuit that occurs after an initial transient time.
Transient time	An interval equal to approximately five time constants.

TRUE/FALSE QUIZ

Answers are at the end of the chapter.

1. Transient time for an *RC* circuit is the same as the time constant.
2. The capacitor in a pulsed *RC* circuit will fully charge if the pulse width is equal or greater than 5τ.
3. If the source voltage is increased in a pulsed *RC* circuit, the time constant will increase.
4. Under certain conditions, the output of an *RC* integrator can have the same maximum and minimum voltage as the input.
5. Under certain conditions, the output of an *RC* integrator can be equal to the average voltage of the input.
6. The instant after a pulse is applied to an *RC* differentiator circuit, the output will be approximately equal to the peak value of the pulse.
7. In an *RL* integrator, the output is taken across the inductor.
8. If the inductor in an *RL* differentiator is open, the output will be zero.
9. The time constant for an *RL* differentiator is inversely proportional to the resistance.
10. A circuit that can develop positive and negative triggers from a square wave is a differentiator.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SELF-TEST

Answers are at the end of the chapter.

6238262 2013/12/10 12.27.66.5

1. The output of an *RC* integrator is taken across the
(a) resistor **(b)** capacitor **(c)** source **(d)** coil

2. When a 10 V input pulse with a width equal to one time constant is applied to an *RC* integrator, the capacitor charges to
(a) 10 V **(b)** 5 V **(c)** 6.3 V **(d)** 3.7 V

3. When a 10 V input pulse with a width equal to one time constant is applied to an *RC* differentiator, the capacitor charges to
(a) 6.3 V **(b)** 10 V **(c)** 5 V **(d)** 3.7 V

4. In an *RC* integrator, the output pulse closely resembles the input pulse when
(a) τ is much larger than the pulse width
(b) τ is equal to the pulse width
(c) τ is less than the pulse width
(d) τ is much less than the pulse width

5. In an *RC* differentiator, the output pulse closely resembles the input when
(a) τ is much larger than the pulse width
(b) τ is equal to the pulse width
(c) τ is less than the pulse width
(d) τ is much less than the pulse width

6. The positive and negative portions of a differentiator's output voltage are maximum when
(a) $5\tau < t_W$ **(b)** $5\tau > t_W$ **(c)** $5\tau = t_W$ **(d)** $5\tau > 0$

7. The output of an *RL* integrator is taken across the
(a) resistor **(b)** coil **(c)** source **(d)** capacitor

8. The maximum current in an *RL* differentiator is
(a) $I = \frac{V_p}{X_L}$ **(b)** $I = \frac{V_p}{Z}$ **(c)** $I = \frac{V_p}{R}$

9. The current in an *RL* differentiator reaches its maximum possible value when
(a) $5\tau = t_W$ **(b)** $5\tau < t_W$ **(c)** $5\tau > t_W$ **(d)** $\tau = 0.5t_W$

10. If you have an *RC* and an *RL* differentiator with equal time constants sitting side-by-side and you apply the same input pulse to both,
(a) the *RC* has the widest output pulse
(b) the *RL* has the most narrow spikes on the output
(c) the output of one is an increasing exponential and the output of the other is a decreasing exponential
(d) you can't tell the difference by observing the output waveforms

PROBLEMS

Answers to odd-numbered problems are at the end of the book.

BASIC PROBLEMS

SECTION 15–1 The *RC* Integrator

1. An integrator has $R = 2.2\ \text{k}\Omega$ in series with $C = 0.047\ \mu\text{F}$. What is the time constant?

2. Determine how long it takes the capacitor in an integrating circuit to reach full charge for each of the following series *RC* combinations:
(a) $R = 47\ \Omega, C = 47\ \mu\text{F}$
(b) $R = 3300\ \Omega, C = 0.015\ \mu\text{F}$
(c) $R = 22\ \text{k}\Omega, C = 100\ \text{pF}$
(d) $R = 4.7\ \text{M}\Omega, C = 10\ \text{pF}$

3. It is required that an integrator have a time constant of approximately 6 ms. If $C = 0.22\ \mu\text{F}$, what standard value of R must be used?

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

4. For the capacitor in the integrator of Problem 3 to fully charge during a pulse, what must be the minimum width of the pulse?

SECTION 15–2 Response of *RC* Integrators to a Single Pulse

5. A 20 V pulse is applied to an *RC* integrator. The pulse width equals one time constant. To what voltage does the capacitor charge during the pulse? Assume that it is initially uncharged.

6. Repeat Problem 5 for the following values of t_W:
(a) 2τ **(b)** 3τ **(c)** 4τ **(d)** 5τ

7. Show the approximate shape of an integrator output voltage where 5τ is much less than the pulse width of a 10 V square wave input. Repeat for the case in which 5τ is much larger than the pulse width.

8. Determine the output voltage for an integrator with a single-pulse input, as shown in Figure 15–56. For repetitive pulses, how long will it take this circuit to reach steady state?

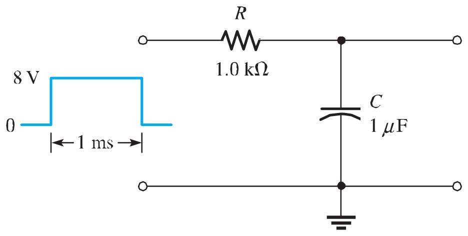

FIGURE 15–56

SECTION 15–3 Response of *RC* Integrators to Repetitive Pulses

9. Draw the integrator output in Figure 15–57 showing maximum voltages.

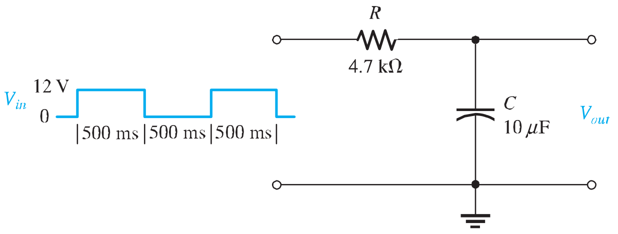

FIGURE 15–57

10. A 1 V, 10 kHz pulse waveform with a duty cycle of 25% is applied to an integrator with $\tau = 25\ \mu\text{s}$. Graph the output voltage for three initial pulses. *C* is initially uncharged.

11. What is the steady-state output voltage of the *RC* integrator with a square wave input shown in Figure 15–58?

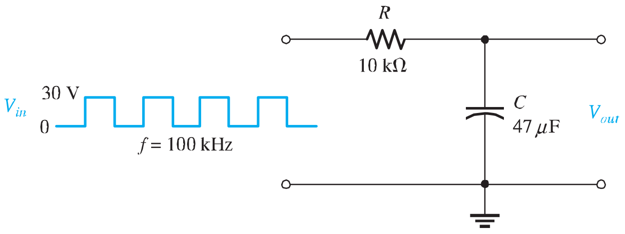

FIGURE 15–58

SECTION 15–4 Response of *RC* Differentiators to a Single Pulse

12. Repeat Problem 7 for an *RC* differentiator.

13. Redraw the circuit in Figure 15–56 to make it a differentiator, and repeat Problem 8.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 15–5 Response of *RC* Differentiators to Repetitive Pulses

14. Draw the differentiator output in Figure 15–59, showing maximum voltages.

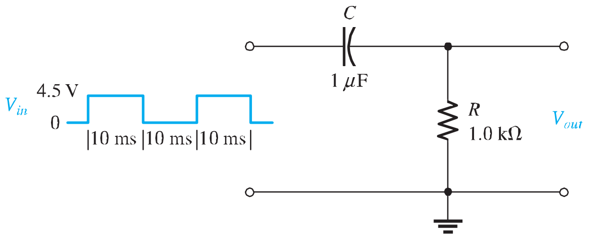
C
1 μF
4.5 V
V_{in}
0
|10 ms |10 ms |10 ms|
R
1.0 kΩ
V_{out}

FIGURE 15–59

15. What is the steady-state output voltage of the differentiator with the square wave input shown in Figure 15–60?

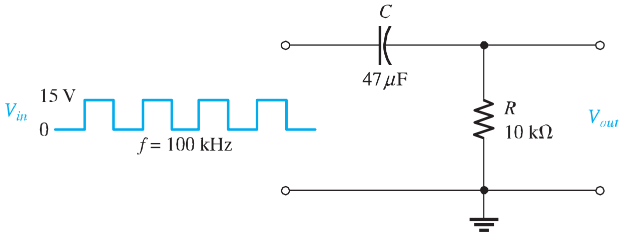
C
47 μF
15 V
V_{in}
0
f = 100 kHz
R
10 kΩ
V_{out}

FIGURE 15–60

SECTION 15–6 Response of *RL* Integrators to Pulse Inputs

16. Determine the output voltage for the circuit in Figure 15–61. A single-pulse input is applied as shown.

17. Draw the integrator output in Figure 15–62, showing maximum voltages.

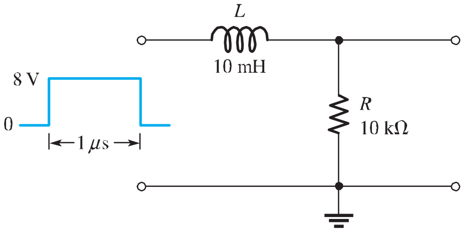
L
10 mH
8 V
0
1 μs
R
10 kΩ

FIGURE 15–61

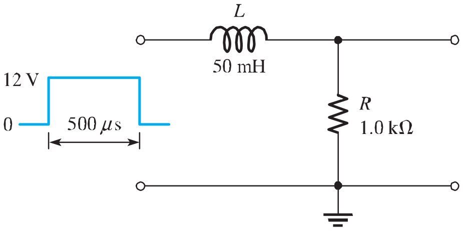
L
50 mH
12 V
0
500 μs
R
1.0 kΩ

FIGURE 15–62

SECTION 15–7 Response of *RL* Differentiators to Pulse Inputs

18. **(a)** What is τ in Figure 15–63?
(b) Draw the output voltage.

19. Draw the output waveform if a periodic pulse waveform with $t_W = 250$ ns and $T = 600$ ns is applied to the circuit in Figure 15–63.

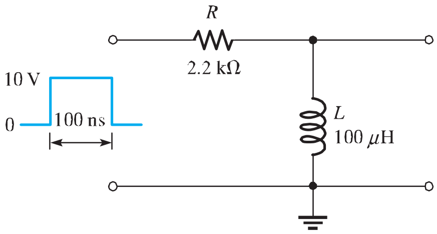
R
2.2 kΩ
10 V
0
100 ns
L
100 μH

FIGURE 15–63

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 15–8 Integrator and Differentiator Applications

20. What is the instantaneous voltage at point B in Figure 15–49 440 μs after the switch is closed for $R = 22\ \text{k}\Omega$ and $C = 0.001\ \mu\text{F}$? $V_{in} = 10$ V.

21. Ideally, what is the output of an RC integrator when the input is a square wave with an amplitude of 12 V? Assume that the time constant is much greater than the period of the input signal?

SECTION 15–9 Troubleshooting

22. Determine the most likely fault(s) in the circuit of Figure 15–64(a) for each set of waveforms in parts (b) and (c). V_{in} is a square wave with a period of 8 ms.

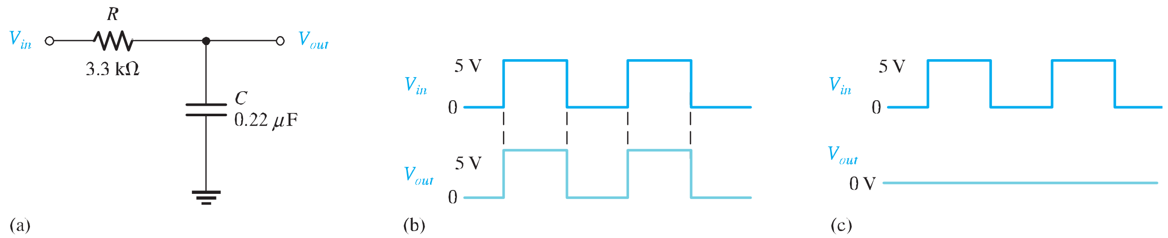

R
V_in
V_out
3.3 kΩ
C
0.22 μF
(a)
V_in
5 V
0
V_out
5 V
0
(b)
V_in
5 V
0
V_out
0 V
(c)

FIGURE 15–64

6238262 2013/12/10 12.27.66.5

23. Determine the most likely fault(s), if any, in the circuit of Figure 15–65(a) for each set of waveforms in parts (b) and (c). V_{in} is a square wave with a period of 8 ms.

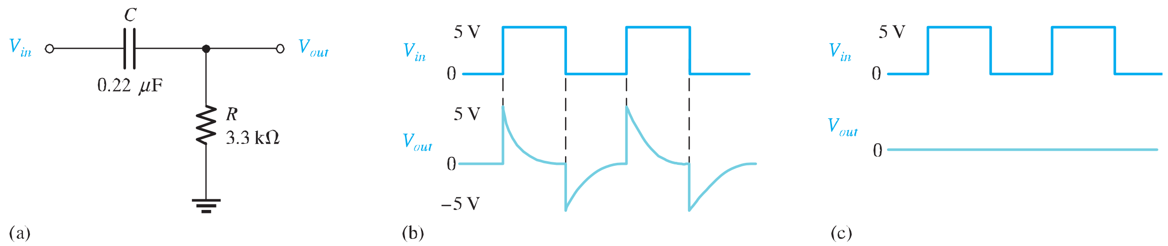

C
V_in
V_out
0.22 μF
R
3.3 kΩ
(a)
V_in
5 V
0
V_out
5 V
0
−5 V
(b)
V_in
5 V
0
V_out
0
(c)

FIGURE 15–65

ADVANCED PROBLEMS

24. **(a)** What is τ in Figure 15–66?
(b) Draw the output voltage.

FIGURE 15–66

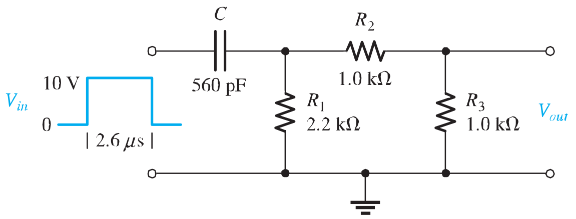

C
560 pF
R2
1.0 kΩ
V_in
10 V
0
2.6 μs
R1
2.2 kΩ
R3
1.0 kΩ
V_out

25. **(a)** What is τ in Figure 15–67?
(b) Draw the output voltage.

FIGURE 15–67

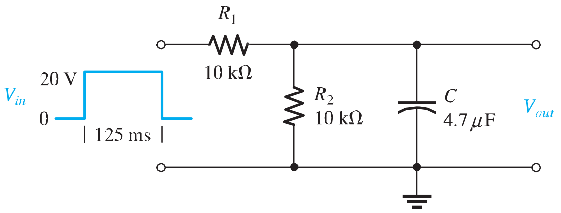

R1
10 kΩ
V_in
20 V
0
125 ms
R2
10 kΩ
C
4.7 μF
V_out

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

26. Determine the time constant in Figure 15–68. Is this circuit an integrator or a differentiator?

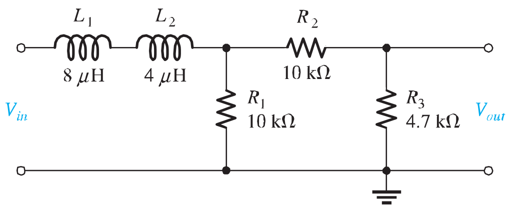

FIGURE 15–68

27. In a time-delay circuit like that of Figure 15–49, what time constant will produce a 1 s delay if the threshold of the circuit is 2.5 V and the amplitude of the input is 5 V?

28. Draw the schematic for the circuit in Figure 15–69 and determine if the oscilloscope presentation is correct.

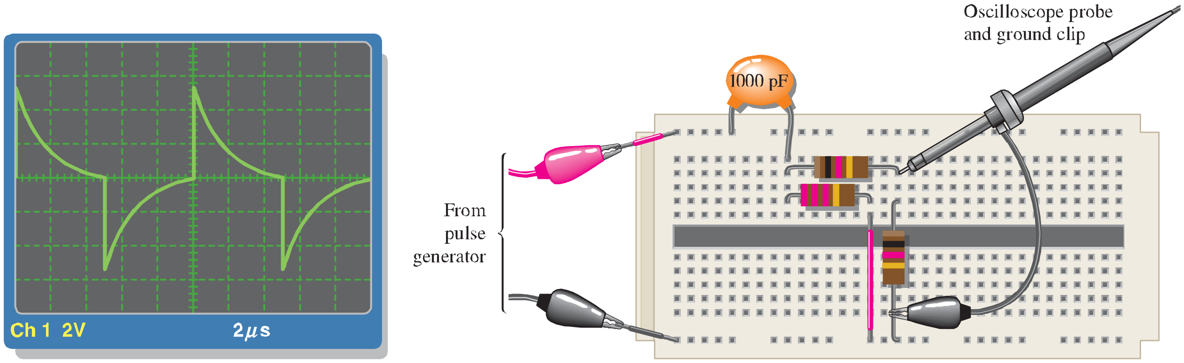

(a) Oscilloscope display

(b) Circuit board with instrument leads connected

FIGURE 15–69

29. Suggest a method that you could use to calibrate the time base for frequency in System Example 15–1.

30. Explain why the final output of System Example 15–2 is a negative-going trigger and not a positive-going trigger.

MULTISIM TROUBLESHOOTING PROBLEMS

MULTISIM

31. Open file P15-31; files are found at www.pearsonhighered.com/floyd. Test the circuit. If there is a fault, identify it.

32. Open file P15-32 and test the circuit. If there is a fault, identify it.

33. Determine if there is a fault in the circuit in file P15-33. If so, identify it.

34. Identify any faulty component in the circuit in file P15-34.

ANSWERS TO SECTION CHECKUPS

SECTION 15–1 The *RC* Integrator

1. An integrator is a series *RC* circuit with a pulse input in which the output is across the capacitor.

2. A voltage applied to the input causes the capacitor to charge. Zero volts across the input causes the capacitor to discharge.

SECTION 15–2 Response of *RC* Integrators to a Single Pulse

1. The output of an integrator reaches full amplitude when $5\tau \leq t_W$.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

2. $V_{out} = (0.632)1\text{ V} = 0.632\text{ V}$; $t_{\text{disch}} = 5\tau = 51.7\text{ ms}$.
3. See Figure 15–70.

FIGURE 15–70

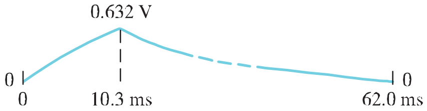

4. No, C will not fully charge.
5. The output is approximately the shape of the input when $5\tau \ll t_W$ (5τ much less than t_W).

SECTION 15–3 Response of *RC* Integrators to Repetitive Pulses

1. An integrator capacitor will fully charge and discharge when $5\tau \leq t_W$ and $5\tau \leq$ time between pulses.
2. The output looks like the input.
3. transient time
4. Steady state is the response after the transient time has passed.
5. The output is a dc voltage that is the average of the input.
6. The integrator converts the square wave to a triangle wave, which moves the beam in the x-axis.

SECTION 15–4 Response of *RC* Differentiators to a Single Pulse

1. See Figure 15–71.

FIGURE 15–71

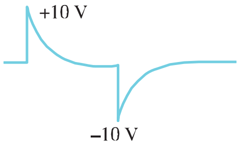

2. The output approximates the input when $5\tau \gg t_W$.
3. The output consists of a positive and a negative spike.
4. $V_R = +5\text{ V} - 15\text{ V} = -10\text{ V}$

SECTION 15–5 Response of *RC* Differentiators to Repetitive Pulses

1. C fully charges and discharges when $5\tau \leq t_W$ and $5\tau \leq$ time between pulses.
2. The output consists of positive and negative spikes.
3. $V_{out} = 0\text{ V}$
4. The first differentiator converts the comparator square wave output to a series of short pulses (spikes) which continuously retrigger the one-shot as long as power is on. The second differentiator detects when the output of the one-shot goes low as a result of power loss.

SECTION 15–6 Response of *RL* Integrators to Pulse Inputs

1. The output is across the resistor.
2. The output reaches the input amplitude when $5\tau \leq t_W$.
3. The output approximates the input when $5\tau \ll t_W$.

SECTION 15–7 Response of *RL* Differentiators to Pulse Inputs

1. The output is across the inductor.
2. The output approximates the input when $5\tau \gg t_W$.
3. $V_{out} = 2\text{ V} - 10\text{ V} = -8\text{ V}$

SECTION 15–8 Integrator and Differentiator Applications

1. Increase the time constant to increase delay time.
2. Reducing R will increase ripple.
3. A very short time constant produces very short duration spikes.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SECTION 15–9 Troubleshooting

1. Zero output of *RC* integrator indicates shorted capacitor, open resistor, no source voltage, or open contact.
2. An open integrator capacitor results in an output exactly like the input.
3. A shorted differentiator capacitor results in an output exactly like the input.

ANSWERS TO RELATED PROBLEMS FOR EXAMPLES

6238262 2013/12/10 12.27.66.5

15–1	8.65 V
15–2	24.7 V
15–3	1.10 V
15–4	See Figure 15–72.

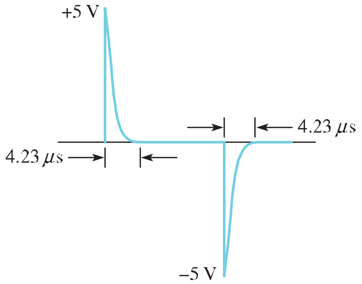

FIGURE 15–72

15–5	892 mV
15–6	10 kΩ
15–7	20 V
15–8	24.7 kΩ (assuming $R_s = 300\ \Omega$)
15–9	See Figure 15–73.

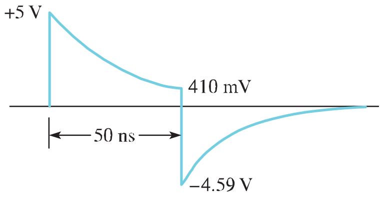

FIGURE 15–73

15–10	20 kΩ
15–11	3.5 μs

ANSWERS TO TRUE/FALSE QUIZ

1. F **2.** T **3.** F **4.** T **5.** T **6.** T **7.** F
8. F **9.** T **10.** T

ANSWERS TO SELF-TEST

1. (b) **2.** (c) **3.** (a) **4.** (d) **5.** (a) **6.** (a) **7.** (a)
8. (c) **9.** (b) **10.** (d)

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

This page intentionally left blank

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

APPENDIX A

Table of Standard Resistor Values

Resistance Tolerance (±%)

0.1% 0.25% 0.5%	1%	2% 5%	10%	0.1% 0.25% 0.5%	1%	2% 5%	10%	0.1% 0.25% 0.5%	1%	2% 5%	10%	0.1% 0.25% 0.5%	1%	2% 5%	10%	0.1% 0.25% 0.5%	1%	2% 5%	10%	0.1% 0.25% 0.5%	1%	2% 5%	10%
10.0	10.0	10	10	14.7	14.7	—	—	21.5	21.5	—	—	31.6	31.6	—	—	46.4	46.4	—	—	68.1	68.1	68	68
10.1	—	—	—	14.9	—	—	—	21.8	—	—	—	32.0	—	—	—	47.0	—	47	47	69.0	—	—	—
10.2	10.2	—	—	15.0	15.0	15	15	22.1	22.1	22	22	32.4	32.4	—	—	47.5	47.5	—	—	69.8	69.8	—	—
10.4	—	—	—	15.2	—	—	—	22.3	—	—	—	32.8	—	—	—	48.1	—	—	—	70.6	—	—	—
10.5	10.5	—	—	15.4	15.4	—	—	22.6	22.6	—	—	33.2	33.2	33	33	48.7	48.7	—	—	71.5	71.5	—	—
10.6	—	—	—	15.6	—	—	—	22.9	—	—	—	33.6	—	—	—	49.3	—	—	—	72.3	—	—	—
10.7	10.7	—	—	15.8	15.8	—	—	23.2	23.2	—	—	34.0	34.0	—	—	49.9	49.9	—	—	73.2	73.2	—	—
10.9	—	—	—	16.0	—	16	—	23.4	—	—	—	34.4	—	—	—	50.5	—	—	—	74.1	—	—	—
11.0	11.0	11	—	16.2	16.2	—	—	23.7	23.7	—	—	34.8	34.8	—	—	51.1	51.1	51	—	75.0	75.0	75	—
11.1	—	—	—	16.4	—	—	—	24.0	—	24	—	35.2	—	—	—	51.7	—	—	—	75.9	—	—	—
11.3	11.3	—	—	16.5	16.5	—	—	24.3	24.3	—	—	35.7	35.7	—	—	52.3	52.3	—	—	76.8	76.8	—	—
11.4	—	—	—	16.7	—	—	—	24.6	—	—	—	36.1	—	36	—	53.0	—	—	—	77.7	—	—	—
11.5	11.5	—	—	16.9	16.9	—	—	24.9	24.9	—	—	36.5	36.5	—	—	53.6	53.6	—	—	78.7	78.7	—	—
11.7	—	—	—	17.2	—	—	—	25.2	—	—	—	37.0	—	—	—	54.2	—	—	—	79.6	—	—	—
11.8	11.8	—	—	17.4	17.4	—	—	25.5	25.5	—	—	37.4	37.4	—	—	54.9	54.9	—	—	80.6	80.6	—	—
12.0	—	12	12	17.6	—	—	—	25.8	—	—	—	37.9	—	—	—	56.2	—	—	—	81.6	—	—	—
12.1	12.1	—	—	17.8	17.8	—	—	26.1	26.1	—	—	38.3	38.3	—	—	56.6	56.6	56	56	82.5	82.5	82	82
12.3	—	—	—	18.0	—	18	18	26.4	—	—	—	38.8	—	—	—	56.9	—	—	—	83.5	—	—	—
12.4	12.4	—	—	18.2	18.2	—	—	26.7	26.7	—	—	39.2	39.2	39	39	57.6	57.6	—	—	84.5	84.5	—	—
12.6	—	—	—	18.4	—	—	—	27.1	—	27	27	39.7	—	—	—	58.3	—	—	—	85.6	—	—	—
12.7	12.7	—	—	18.7	18.7	—	—	27.4	27.4	—	—	40.2	40.2	—	—	59.0	59.0	—	—	86.6	86.6	—	—
12.9	—	—	—	18.9	—	—	—	27.7	—	—	—	40.7	—	—	—	59.7	—	—	—	87.6	—	—	—
13.0	13.0	13	—	19.1	19.1	—	—	28.0	28.0	—	—	41.2	41.2	—	—	60.4	60.4	—	—	88.7	88.7	—	—
13.2	—	—	—	19.3	—	—	—	28.4	—	—	—	41.7	—	—	—	61.2	—	—	—	89.8	—	—	—
13.3	13.3	—	—	19.6	19.6	—	—	28.7	28.7	—	—	42.2	42.2	—	—	61.9	61.9	62	—	90.9	90.9	91	—
13.5	—	—	—	19.8	—	—	—	29.1	—	—	—	42.7	—	—	—	62.6	—	—	—	92.0	—	—	—
13.7	13.7	—	—	20.0	20.0	20	—	29.4	29.4	—	—	43.2	43.2	43	—	63.4	63.4	—	—	93.1	93.1	—	—
13.8	—	—	—	20.3	—	—	—	29.8	—	—	—	43.7	—	—	—	64.2	—	—	—	94.2	—	—	—
14.0	14.0	—	—	20.5	20.5	—	—	30.1	30.1	30	—	44.2	44.2	—	—	64.9	64.9	—	—	95.3	95.3	—	—
14.2	—	—	—	20.8	—	—	—	30.5	—	—	—	44.8	—	—	—	65.7	—	—	—	96.5	—	—	—
14.3	14.3	—	—	21.0	21.0	—	—	30.9	30.9	—	—	45.3	45.3	—	—	66.5	66.5	—	—	97.6	97.6	—	—
14.5	—	—	—	21.3	—	—	—	31.2	—	—	—	45.9	—	—	—	67.3	—	—	—	98.8	—	—	—

NOTE: These values are generally available in multiples of 0.1, 1, 10, 100, 1 k, and 1 M.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

This page intentionally left blank

6238262 2013/

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

APPENDIX B

Capacitor Color Coding and Marking

Capacitor Color

Some capacitors have color-coded designations. The color code used for capacitors is basically the same as that used for resistors. Some variations occur in tolerance designation. The basic color codes are shown in Table B–1, and some typical color-coded capacitors are illustrated in Figure B–1.

TABLE B–1 • Typical composite color codes for capacitors (picofarads).

COLOR	DIGIT	MULTIPLIER	TOLERANCE
Black	0	1	20%
Brown	1	10	1%
Red	2	100	2%
Orange	3	1000	3%
Yellow	4	10000	
Green	5	100000	5% (EIA)
Blue	6	1000000	
Violet	7		
Gray	8		
White	9		
Gold		0.1	5% (JAN)
Silver		0.01	10%
No color			20%

NOTE: EIA stands for Electronic Industries Association, and JAN stands for Joint Army-Navy, a military standard.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

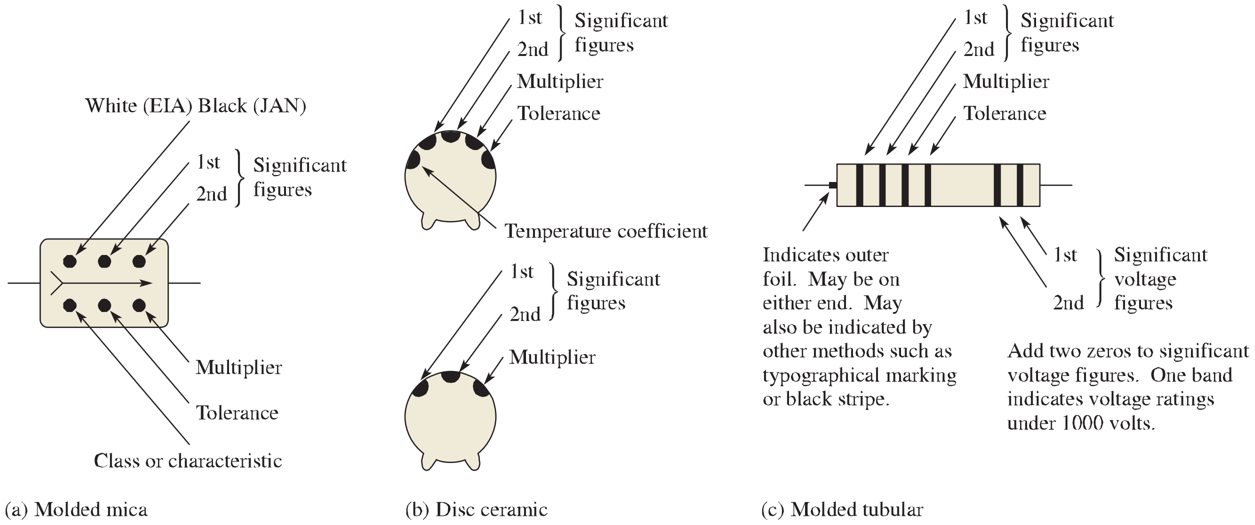
White (EIA) Black (JAN)
1st
2nd
Significant figures
Multiplier
Tolerance
Class or characteristic
(a) Molded mica
1st
2nd
Significant figures
Multiplier
Tolerance
Temperature coefficient
1st
2nd
Significant figures
Multiplier
(b) Disc ceramic
1st
2nd
Significant figures
Multiplier
Tolerance
Indicates outer foil. May be on either end. May also be indicated by other methods such as typographical marking or black stripe.
1st
2nd
Significant voltage figures
Add two zeros to significant voltage figures. One band indicates voltage ratings under 1000 volts.
(c) Molded tubular

FIGURE B–1 **Typical color-coded capacitors.**

Marking Systems

A capacitor, as shown in Figure B–2, has certain identifying features.

- Body of one solid color (off-white, beige, gray, tan or brown).
- End electrodes completely enclose ends of part.
- Many different sizes:
  1. Type 1206: 0.125 inch long by 0.063 inch wide (3.2 mm × 1.6 mm) with variable thickness and color.
  2. Type 0805: 0.080 inch long by 0.050 inch wide (2.0 mm × 1.25 mm) with variable thickness and color.
  3. Variably sized with a single color (usually translucent tan or brown). Sizes range from 0.059 inch (1.5 mm) to 0.220 inch (5.6 mm) in length and in width from 0.032 inch (0.8 mm) to 0.197 inch (5.0 mm).
- Three different marking systems:
  1. Two place (letter and number only).
  2. Two place (letter and number or two numbers).
  3. One place (letter of varying color).

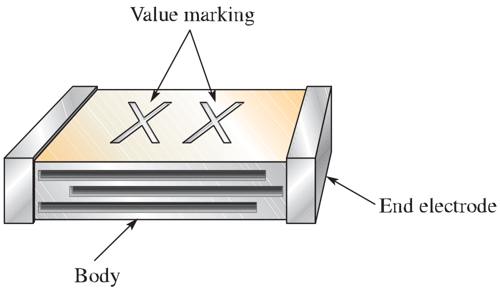
Value marking
X X
End electrode
Body

FIGURE B–2 **Capacitor marking.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Standard Two-Place Code

Refer to Table B–2.

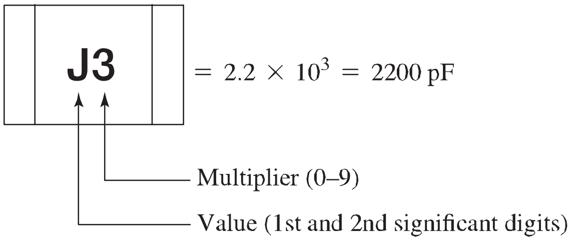
J3
Multiplier (0–9)
Value (1st and 2nd significant digits)

$= 2.2 \times 10^3 = 2200$ pF

Examples: S2 $= 4.7 \times 100 = 470$ pF
b0 $= 3.5 \times 1.0 = 3.5$ pF

TABLE B–2

VALUE*						MULTIPLIER
A	1.0	L	2.7	T	5.1	0 = ×1.0
B	1.1	M	3.0	U	5.6	1 = ×10
C	1.2	N	3.3	m	6.0	2 = ×100
D	1.3	b	3.5	V	6.2	3 = ×1000
E	1.5	P	3.6	W	6.8	4 = ×10000
F	1.6	Q	3.9	n	7.0	5 = ×100000
G	1.8	d	4.0	X	7.5	etc.
H	2.0	R	4.3	t	8.0	
J	2.2	e	4.5	Y	8.2	
K	2.4	S	4.7	y	9.0	
a	2.5	f	5.0	Z	9.1	

*Note uppercase and lowercase letters.

Alternate Two-Place Code

Refer to Table B–3.

- Values below 100 pF—Value read directly

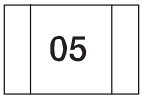

= 5 pF

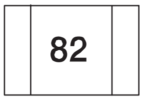

= 82 pF

- Values 100 pF and above—Letter/Number code

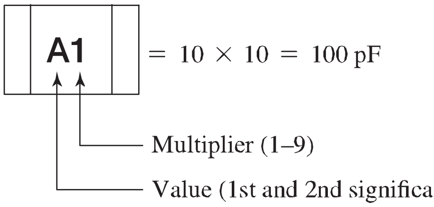
A1
Multiplier (1–9)
Value (1st and 2nd significant digits)

$= 10 \times 10 = 100$ pF

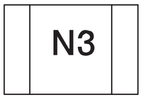

$= 33 \times 1000$
$= 33000$ pF $= .033$ μF

6238262 2013/12/10 12.27.66.5

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

TABLE B–3

VALUE*						MULTIPLIER
A	10	J	22	S	47	1 = ×10
B	11	K	24	T	51	2 = ×100
C	12	L	27	U	56	3 = ×1000
D	13	M	30	V	62	4 = ×10000
E	15	N	33	W	68	5 = ×100000
F	16	P	36	X	75	etc.
G	18	Q	39	Y	82	
H	20	R	43	Z	91	

*Note uppercase letters only.

Standard Single-Place Code

Refer to Table B–4.

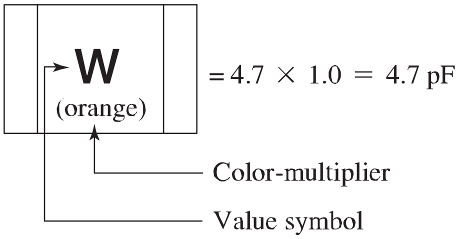

Examples: R (Green) = 3.3 × 100 = 330 pF
7 (Blue) = 8.2 × 1000 = 8200 pF

TABLE B–4

VALUE						MULTIPLIER (COLOR)
A	1.0	K	2.2	W	4.7	Orange = ×1.0
B	1.1	L	2.4	X	5.1	Black = ×10
C	1.2	N	2.7	Y	5.6	Green = ×100
D	1.3	O	3.0	Z	6.2	Blue = ×1000
E	1.5	R	3.3	3	6.8	Violet = ×10000
H	1.6	S	3.6	4	7.5	Red = ×100000
I	1.8	T	3.9	7	8.2	
J	2.0	V	4.3	9	9.1	

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

APPENDIX C

Norton's Theorem and Millman's Theorem

Norton's Theorem

Like Thevenin's theorem, Norton's theorem provides a method of reducing a more complex circuit to a simpler form. The basic difference is that Norton's theorem gives an equivalent current source in parallel with an equivalent resistance. The form of Norton's equivalent circuit is shown in Figure C–1. Regardless of how complex the original circuit is, it can always be reduced to this equivalent form. The equivalent current source is designated I_N, and the equivalent resistance is designated R_N.

To apply Norton's theorem, you must know how to find the two quantities I_N and R_N. Once you know them for a given circuit, simply connect them in parallel to get the complete Norton circuit.

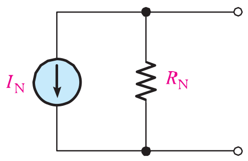

FIGURE C–1 **The form of Norton's equivalent circuit.**

Norton's Equivalent Current (I_N) As stated, I_N is one part of the complete Norton equivalent circuit; R_N is the other part. I_N is defined to be the short-circuit current between two points in a circuit. Any component connected between these two points effectively "sees" a current source of value I_N in parallel with R_N.

To illustrate, suppose that a resistive circuit of some kind has a resistor connected between two points in the circuit, as shown in Figure C–2(a). We wish to find the Norton circuit that is equivalent to the one shown as "seen" by R_L. To find I_N, calculate the current between points A and B with these two points shorted, as shown in Figure C–2(b). Example C–1 demonstrates how to find I_N.

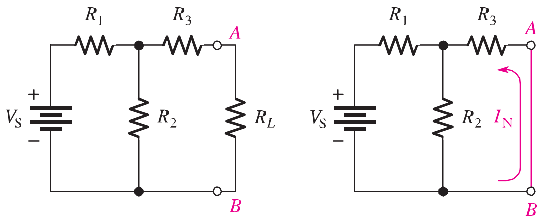

(a) Original circuit (b) Short the terminals to get I_N.

FIGURE C–2 **Determining the Norton equivalent current, I_N.**

EXAMPLE C–1

Determine I_N for the circuit within the shaded area in Figure C–3(a).

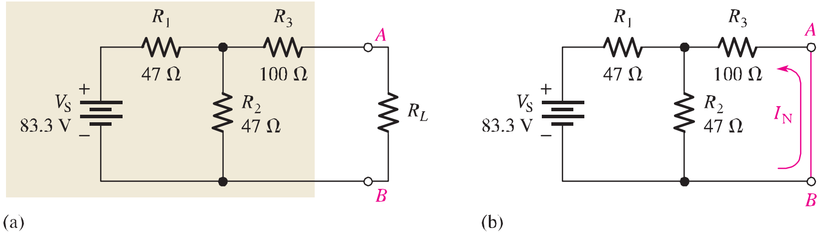

FIGURE C–3

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

SOLUTION

6238262 2013/12/10 12.27.66.5

Short terminals *A* and *B* as shown in Figure C–3(b). I_N is the current through the short and is calculated as follows: First, the total resistance seen by the voltage source is

$$R_T = R_1 + \frac{R_2R_3}{R_2 + R_3} = 47\ \Omega + \frac{(47\ \Omega)(100\ \Omega)}{147\ \Omega} = 79\ \Omega$$

The total current from the source is

$$I_T = \frac{V_S}{R_T} = \frac{83.3\ \text{V}}{79\ \Omega} = 1.05\ \text{A}$$

Now apply the current-divider formula to find I_N (the current through the short).

$$I_N = \left(\frac{R_2}{R_2 + R_3}\right)I_T = \left(\frac{47\ \Omega}{147\ \Omega}\right)1.05\ \text{A} = \mathbf{336\ mA}$$

This is the value for the equivalent Norton current source.

Norton's Equivalent Resistance (R_N) We define R_N in the same way as R_{TH}: it is the total resistance appearing between two terminals in a given circuit with all sources replaced by their internal resistances. Example C–2 demonstrates how to find R_N.

EXAMPLE C–2

Find R_N for the circuit within the shaded area of Figure C–3(a) (see Example C–1).

SOLUTION

First reduce V_S to zero by shorting it, as shown in Figure C–4.

Looking in at terminals *A* and *B*, you can see that the parallel combination of R_1 and R_2 is in series with R_3. Thus,

$$R_N = R_3 + \frac{R_1}{2} = 100\ \Omega + \frac{47\ \Omega}{2} = \mathbf{124\ \Omega}$$

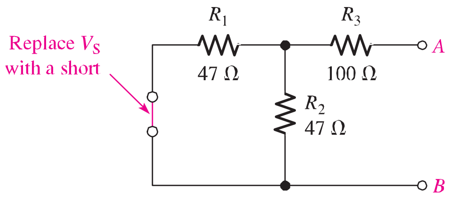

FIGURE C–4

The two examples have shown how to find the two equivalent components of a Norton equivalent circuit, I_N and R_N. Keep in mind that these values can be found for any linear circuit. Once these are known, they must be connected in parallel to form the Norton equivalent circuit, as illustrated in Example C–3.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

EXAMPLE C–3

Draw the complete Norton circuit for the original circuit in Figure C–3 (Example C–1).

SOLUTION

In Examples C–1 and C–2 you found that $I_N = 336$ mA and $R_N = 124\ \Omega$. The Norton equivalent circuit is shown in Figure C–5.

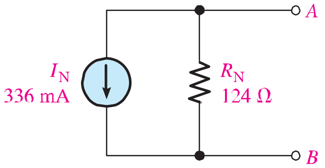

FIGURE C–5

Summary of Norton's Theorem Any load resistor connected between the terminals of a Norton equivalent circuit will have the same current through it and the same voltage across it as if it were connected to the terminals of the original circuit. A summary of steps for theoretically applying Norton's theorem is as follows:

1. Short the two terminals between which you want to find the Norton equivalent circuit.
2. Determine the current (I_N) through the shorted terminals.
3. Determine the resistance (R_N) between the two terminals (opened) with all voltage sources shorted and all current sources opened ($R_N = R_{TH}$).
4. Connect I_N and R_N in parallel to produce the complete Norton equivalent for the original circuit.

Norton's equivalent circuit can also be derived from Thevenin's equivalent circuit by use of the source conversion method.

Millman's Theorem

Millman's theorem permits any number of parallel voltage sources to be reduced to a single equivalent voltage source. It simplifies finding the voltage across or current through a load. Millman's theorem gives the same results as Thevenin's theorem for the special case of parallel voltage sources. A conversion by Millman's theorem is illustrated in Figure C–6.

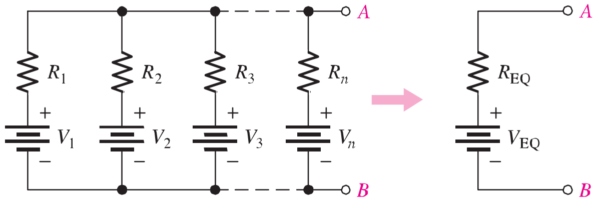

FIGURE C–6 **Reduction of parallel voltage sources to a single equivalent voltage source.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Millman's Equivalent Voltage (V_{EQ}) and Equivalent Resistance (R_{EQ})

Millman's theorem gives us a formula for calculating the equivalent voltage, V_{EQ}. To find V_{EQ}, convert each of the parallel voltage sources into the current sources, as shown in Figure C–7.

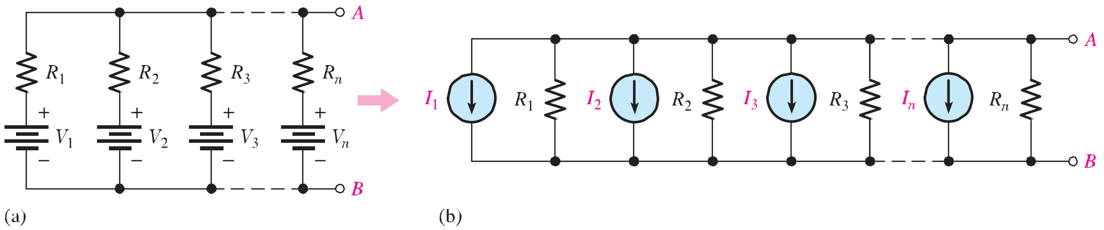

FIGURE C–7 **Parallel voltage sources converted to current sources.**

In Figure C–7 (b), the total current from the parallel current sources is

$$I_T = I_1 + I_2 + I_3 + \cdots + I_n$$

The total conductance between terminals *A* and *B* is

$$G_T = G_1 + G_2 + G_3 + \cdots + G_n$$

where $G_T = 1/R_T$, $G_1 = 1/R_1$, and so on. Remember, the current sources are effectively open. Therefore, by Millman's theorem, the equivalent resistance, R_{EQ}, is the total resistance R_T.

$$\boldsymbol{R_{EQ} = \frac{1}{G_T} = \frac{1}{(1/R_1) + (1/R_2) + (1/R_3) + \cdots + (1/R_n)}} \qquad \text{(C–1)}$$

By Millman's theorem, the equivalent voltage is $I_T R_{EQ}$, where I_T is expressed as follows:

$$I_T = I_1 + I_2 + I_3 + \cdots + I_n = \frac{V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_3}{R_3} + \cdots + \frac{V_n}{R_n}$$

The following is the formula for the equivalent voltage:

$$\boldsymbol{V_{EQ} = \frac{(V_1/R_1) + (V_2/R_2) + (V_3/R_3) + \cdots + (V_n/R_n)}{(1/R_1) + (1/R_2) + (1/R_3) + \cdots + (1/R_n)}} \qquad \text{(C–2)}$$

Equations C–1 and C–2 are the two Millman formulas. The equivalent voltage source has a polarity such that the total current through a load will be in the same direction as in the original circuit.

EXAMPLE C–4

Use Millman's theorem to find the voltage across R_L and the current through R_L in Figure C–8.

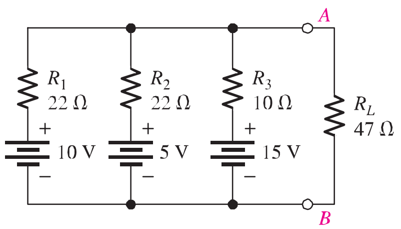

FIGURE C–8

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

SOLUTION

Apply Millman's theorem as follows:

$$R_{EQ} = \frac{1}{(1/R_1) + (1/R_2) + (1/R_3)}$$

$$= \frac{1}{(1/22\ \Omega) + (1/22\ \Omega) + (1/10\ \Omega)} = \frac{1}{0.19} = 5.24\ \Omega$$

$$V_{EQ} = \frac{(V_1/R_1) + (V_2/R_2) + (V_3/R_3)}{(1/R_1) + (1/R_2) + (1/R_3)}$$

$$= \frac{(10\ \text{V}/22\ \Omega) + (5\ \text{V}/22\ \Omega) + (15\ \text{V}/10\ \Omega)}{(1/22\ \Omega) + (1/22\ \Omega) + (1/10\ \Omega)} = \frac{2.18\ \text{A}}{0.19\ \text{S}} = 11.5\ \text{V}$$

The single equivalent voltage source is shown in Figure C–9.

Now calculate I_L and V_L for the load resistor.

$$I_L = \frac{V_{EQ}}{R_{EQ} + R_L} = \frac{11.5\ \text{V}}{52.2\ \Omega} = \mathbf{220\ mA}$$

$$V_L = I_L R_L = (220\ \text{mA})(47\ \Omega) = \mathbf{10.3\ V}$$

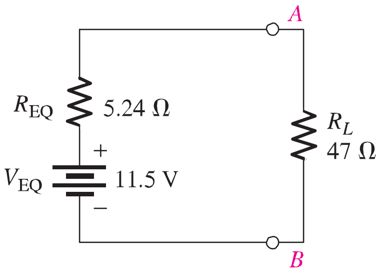

FIGURE C–9

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

This page intentionally left blank

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

APPENDIX D

NI Multisim for Circuit Simulation

Simulate, Prototype, and Test Circuits

Theory, Design, and Prototype

As electronic circuits and systems become more advanced, circuit designers rely on computers in the design process. It is essential for engineers and technicians to systematically design, simulate, prototype, and lay out the circuit. Students can also use the engineering and design process to reinforce concepts and theory in the classroom. The three stages of the student design process are illustrated in Figure D–1.

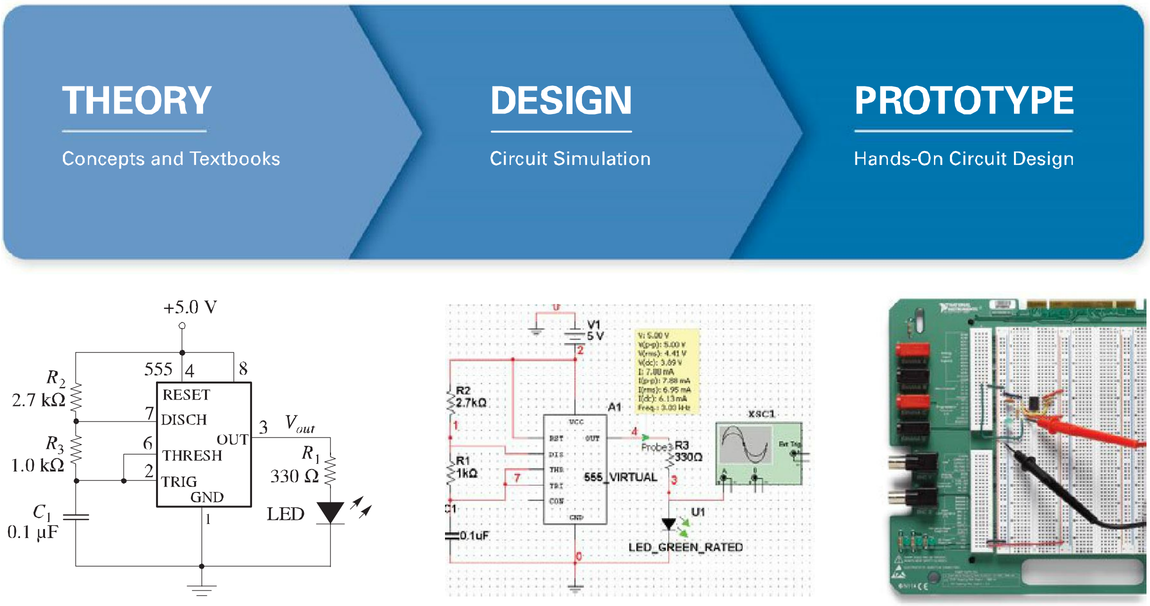
THEORY
Concepts and Textbooks
DESIGN
Circuit Simulation
PROTOTYPE
Hands-On Circuit Design
+5.0 V
555
4
8
R_2
2.7 kΩ
RESET
7
DISCH
OUT
3
V_{out}
R_3
1.0 kΩ
6
THRESH
R_1
330 Ω
2
TRIG
GND
1
C_1
0.1 µF
LED
V1
5 V
R2
2.7kΩ
R1
1kΩ
C1
0.1uF
A1
555_VIRTUAL
Probe3
R3
330Ω
XSC1
U1
LED_GREEN_RATED

FIGURE D–1 Three steps of the student design process for a practical circuit.

The National Instruments Electronics Education Platform is an end-to-end tool-chain designed to meet student and educator needs. The platform consists of NI Multisim simulation software, the NI Educational Laboratory Virtual Instrumentation Suite (NI ELVIS) prototyping workstation, and the NI LabVIEW graphical programming environment. NI Multisim provides intuitive schematic capture, SPICE simulation, and integration with the NI ELVIS to help students explore circuit theory and design circuits to investigate behavior. NI ELVIS is a prototyping platform that allows students to quickly and easily create

---

"Multisim™" is a registered trademark of National Instruments. This publication is independent of National Instruments, which is not affiliated with the publisher or the author, and does not authorize, sponsor, endorse, or otherwise approve this publication.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

their circuits. With NI LabVIEW, students can measure real-world signals and compare simulation results.

1. Investigate Theory Learn the fundamental theory of circuit design through the *Electric Circuits Fundamentals* textbook and course lectures. Reinforce important concepts in the easy-to-use Multisim environment by downloading the Multisim circuit files for the textbook. The Multisim circuit files help build the foundation for an in-depth understanding of circuit behavior. Using the prebuilt circuit files, simulate and analyze the circuit behavior of examples and problems in each chapter.

2. Design and Simulate Circuit simulation provides an interactive view into a circuit. Build a circuit from scratch and learn about the design performance using built-in circuit instruments and probes in an ideal pre-laboratory environment. With the 3D breadboard in Multisim, take the leap from circuit diagram to real-world physical implementation.

6238262 2013/12/10 12:27:66.5

3. Prototype, Measure, and Compare Hands-on experience building physical circuitry is essential. Move from the 3D breadboard in Multisim to a real-world breadboard on NI ELVIS to seamlessly complete the entire design process by prototyping the circuit. Within the LabVIEW environment, compare real-world measurements to the simulation values to reinforce theory, fully understand circuit behavior, and build the foundations of professional engineering analysis.

NI Multisim

Multisim software integrates powerful SPICE simulation and schematic capture into a highly intuitive electronics lab on the computer. Use the Multisim Student Edition to enhance learning by:

Building circuits in a simple, easy-to-learn environment

Simulating and analyzing circuits for homework and prelab assignments

Breadboarding in 3D at home before lab sessions

Creating designs of up to 50 components using a library of nearly 4,000 devices

Integrating with the NI ELVIS prototyping workstation

While using industry-standard SPICE in the background, you can take advantage of the Multisim drag-and-drop interface to make circuit drawing, wiring, and analysis simple and easy to use. You can build circuits from scratch and learn about design performances using built-in virtual instruments and probes in an ideal environment for experimentation. With the 3D breadboard, you can easily take the leap from circuit diagram to real-world implementation.

Using Multisim with Electric Circuits Fundamentals Create an example from the textbook to get familiar with the Multisim environment. First, launch Multisim and open a new schematic window (**File ≫ New ≫ Schematic Capture**). Design circuits in the circuit window by placing components from the component toolbar. Clicking on the component toolbar opens the component browser. Choose the family of components and select an individual component to place on the circuit window by double-clicking on it.

Once you have selected a component, it attaches itself and "ghosts" the mouse cursor. Place the component by clicking again on the desired location in the schematic. If you are new to Multisim, you should use the **BASIC_VIRTUAL** family of components, which you can assign any arbitrary value. As an example of a simple Multisim circuit, an *RLC* circuit is shown in Figure D–2. The circuit connects a 1 kΩ resistor, a 47 mH inductor, and a 10 nF capacitor. All these components are found in the Basic Group of the Component Database.

The next step is to wire the components together. Simply left-click on the source terminal and left-click on the destination terminal. Multisim automatically chooses the best path for the virtual wire between the two terminals. Always be sure you have fully captured your circuit; then you can simulate it. You can use your simulation results as a comparison with the physical circuit.

To analyze the circuit, use a measurement probe to measure the voltages and other characteristics from the circuit while the simulation is running. Use the virtual oscilloscope to analyze the output signal from our sample *RLC* circuit.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

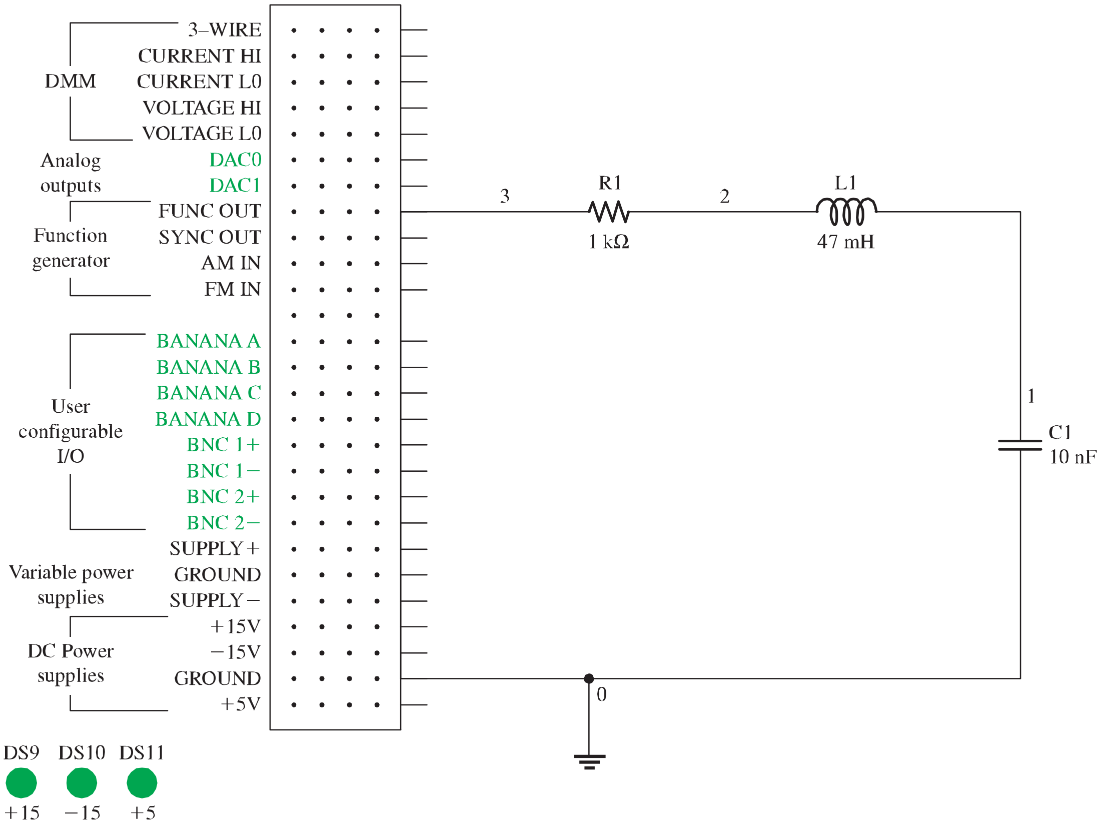
DMM
3–WIRE
CURRENT HI
CURRENT L0
VOLTAGE HI
VOLTAGE L0
Analog outputs
DAC0
DAC1
Function generator
FUNC OUT
SYNC OUT
AM IN
FM IN
User configurable I/O
BANANA A
BANANA B
BANANA C
BANANA D
BNC 1+
BNC 1−
BNC 2+
BNC 2−
Variable power supplies
SUPPLY+
GROUND
SUPPLY−
DC Power supplies
+15V
−15V
GROUND
+5V
DS9
DS10
DS11
+15
−15
+5
3
R1
1 kΩ
2
L1
47 mH
1
C1
10 nF
0

FIGURE D–2 **Multisim schematic of *RLC* circuit.**

NI ELVIS

NI ELVIS is a LabVIEW-based design and prototyping environment based on LabVIEW for university science and engineering laboratories. NI ELVIS consists of virtual instruments, or functions, based on LabVIEW; a multifunction data acquisition device; and a custom-designed benchtop workstation and prototyping board. This combination provides a ready-to-use suite of instruments for educational laboratories. The NI ELVIS workstation is shown in Figure D–3.

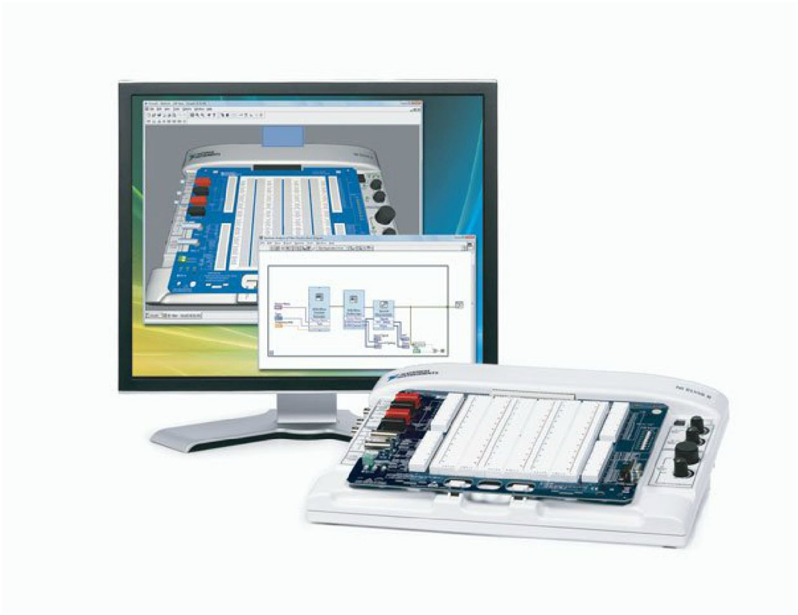

FIGURE D–3 **The NI ELVIS workstation.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

NI ELVIS is an ideal companion to any electronics lab that uses Multisim. It features USB connectivity for easy setup, maintenance, and portability, as well as a portable breadboard for real-world circuit prototyping.

NI ELVIS features a breadboard prototyping environment with built-in instruments including a function generator, digital multimeter (DMM), oscilloscope, and variable power supplies. The breadboard is detachable, so you can work on your projects and labs independently of the NI ELVIS unit. NI ELVIS provides software based on LabVIEW for interacting with virtual instruments.

Using NI ELVIS with Electric Circuits Fundamentals Returning to the *RLC* circuit introduced with Multisim as an example, you can continue the prototyping step. With prototyping hardware, you can quickly construct a circuit on a standard protoboard and test it with the laboratory instruments to complete the design.

Alternatively, you can repeat the simulation step but this time as a prototype using the Virtual 3D NI ELVIS from within Multisim. To construct a 3D NI ELVIS prototype, open the 3D breadboard by clicking **Tools ≫ Show Breadboard**. Place components and wires to build up your circuit. The corresponding connection points and symbols on the NI ELVIS schematic turn green, indicating the 3D connections are correct. If you create a traditional schematic, you see a standard breadboard. Figure D–4 shows the circuit on 3D Virtual NI ELVIS prototyped in Multisim and ready for construction. Once you have verified the layout of your circuit using the 3D virtual environment, you can physically build it on NI ELVIS.

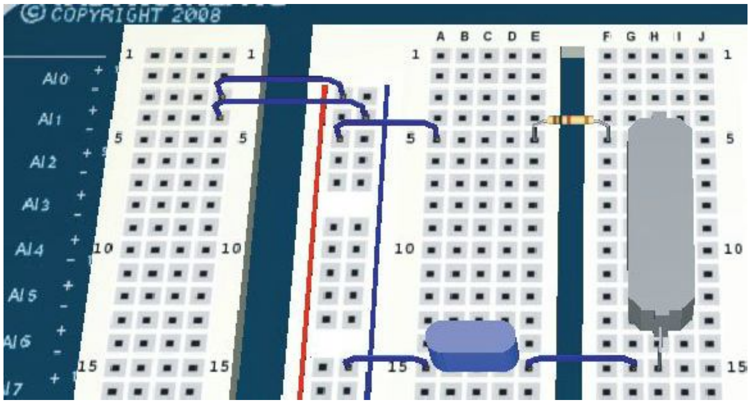

FIGURE D–4 **Prototyped circuit in Multisim.**

After constructing the circuit, you can use the NI ELVIS instrument launcher to measure the output signal of this circuit. Figure D–5 shows the simulated signal of the *RLC*

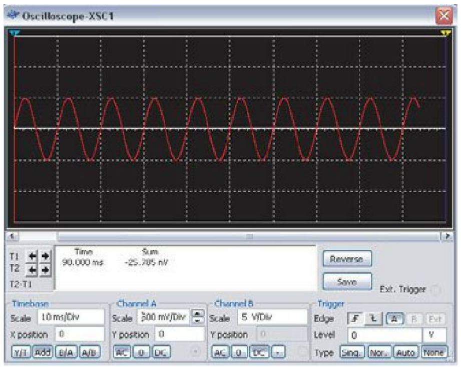

Time 90.000 ms
Sum -25.705 nV

Timebase
Scale 10 ms/Div
X position 0

Channel A
Scale 300 mV/Div
Y position 0

Channel B
Scale 5 V/Div
Y position 0

Trigger
Edge A
Level 0 V

FIGURE D–5 **Multisim simulated signal.**

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

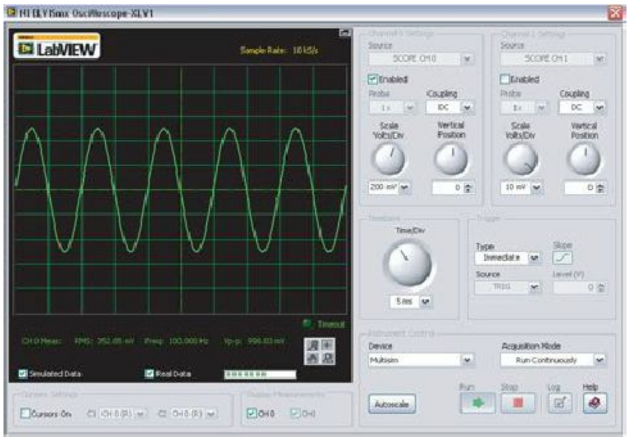

Sample Rate: 10 kS/s

Source: SCOPE CH0 — ☑ Enabled — Probe: 1x — Coupling: DC — Scale Volts/Div: 200 mV — Vertical Position: 0

Source: SCOPE CH1 — ☐ Enabled — Probe: 1x — Coupling: DC — Scale Volts/Div: 10 mV — Vertical Position: 0

Time/Div: 5 ms

Type: Immediate — Source: TRIG — Level (V): 0

Device: Multisim — Acquisition Mode: Run Continuously

☑ Simulated Data ☑ Real Data

☐ Cursors On — ☑ CH0 ☑ CH1

FIGURE D–6 **NI ELVIS measured signal.**

circuit on the Multisim oscilloscope. Figure D–6 shows the measured signal using the ELVIS II Oscilloscope. The most important step in the process is to compare the measurement of the prototyped circuit to the simulation. This helps you determine where potential errors exist in the design. After comparing the measurement with the theoretical values, you can revisit your design to improve it or prepare it for layout on a PCB board, using software such as NI Ultiboard.

NI Multisim Circuit Files Download the Multisim circuit files to develop in-depth understanding of circuit behavior. To download the pre-built circuit files and Multisim resources, visit: ni.com/academic/floyd

NI Multisim Resources The link provides references to the following resources to help you get started with NI Multisim:

Download the free Multisim 30-day evaluation version
View the Getting Started Guide to Multisim
Learn Multisim in 3-Hours tutorial
Discuss Multisim in an online discussion forum

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

This page intentionally left blank

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

ANSWERS TO ODD-NUMBERED PROBLEMS

Chapter 1

1. The circuit is first tested with a computer design and simulation program, which can simulate the performance and look for potential problems. When the simulation is satisfactory, a prototype circuit is constructed, tested, and modified as needed before putting it into production.

3. Electronic assemblies have become more complex but also more reliable, so there is less need for repair. It is generally cheaper for manufacturers to replace a board than troubleshoot it to the component level. Skills needed by technicians tend to be broader skills than in the past.

5. Advantages are that the digital signal can be processed and stored easily; it is also less subject to noise.

7. **(a)** An electronic oscillator generates a repetitive electronic signal.

(b) An oscillator does not have a signal input.

9. A carrier is a high frequency radio wave that can be modulated (changed) by a lower frequency signal.

11. **(a)** 3×10^3 **(b)** 7.5×10^4 **(c)** 2×10^6

13. **(a)** 8.4×10^3 **(b)** 9.9×10^4 **(c)** 2×10^5

15. **(a)** 0.0000025 **(b)** 500 **(c)** 0.39

17. **(a)** 4.32×10^7

(b) 5.00085×10^3

(c) 6.06×10^{-8}

19. **(a)** 2.0×10^9 **(b)** 3.6×10^{14} **(c)** 1.54×10^{-14}

21. **(a)** 89×10^3 **(b)** 450×10^3 **(c)** 12.04×10^{12}

23. **(a)** 345×10^{-6} **(b)** 25×10^{-3} **(c)** 1.29×10^{-9}

25. **(a)** 7.1×10^{-3} **(b)** 101×10^6 **(c)** 1.50×10^6

27. **(a)** 22.7×10^{-3} **(b)** 200×10^6 **(c)** 848×10^{-3}

29. **(a)** 345 μA **(b)** 25 mA **(c)** 1.29 nA

31. **(a)** 3 μF **(b)** 3.3 MΩ **(c)** 350 nA

33. **(a)** 5000 μA **(b)** 3.2 mW

(c) 5 MV **(d)** 10,000 kW

35. **(a)** 50.68 mA **(b)** 2.32 MΩ **(c)** 0.0233 μF

37. **(a)** 3 **(b)** 2 **(c)** 5 **(d)** 2 **(e)** 3 **(f)** 2

Chapter 2

1. 80×10^{12} C

3. 4.64×10^{-18} C

5. **(a)** 10 V **(b)** 2.5 V **(c)** 4 V

7. 20 V

9. 12.5 V

11. **(a)** 75 A **(b)** 20 A **(c)** 2.5 A

13. 2 s

15. A: 6800 Ω ± 10%

B: 33 Ω ± 10%

C: 47,000 Ω ± 5%

17. **(a)** Red, violet, brown, gold

(b) B: 330 Ω, D: 2.2 kΩ, A: 39 kΩ, L: 56 kΩ, F: 100 kΩ

19. **(a)** 10 Ω ± 5%

(b) 5.1 MΩ ± 10%

(c) 68 Ω ± 5%

21. **(a)** 28.7 kΩ ± 1%

(b) 60.4 Ω ± 1%

(c) 9.31 kΩ ± 1%

23. **(a)** 22 Ω **(b)** 4.7 kΩ **(c)** 82 kΩ

(d) 3.3 kΩ **(e)** 56 Ω **(f)** 10 MΩ

25. There is current through lamp 2.

27. Ammeter in series with resistors, with its negative terminal to the negative terminal of source and its positive terminal to one side of R_1. Voltmeter placed across (in parallel with) the source (negative to negative, positive to positive).

29. Position 1: V1 = 0, V2 = V_S

Position 2: V1 = V_S, V2 = 0

31. 250 V

33. **(a)** 200 Ω **(b)** 150 MΩ **(c)** 4500 Ω

35. 33.3 V

37. AWG #27

39. Circuit (b)

41. One ammeter in series with battery. One ammeter in series with each resistor (seven total).

43. See Figure P–1.

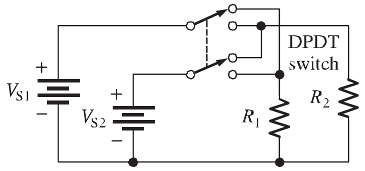

FIGURE P–1

Chapter 3

1. **(a)** 3 A **(b)** 0.2 A **(c)** 1.5 A

3. 15 mA

5. **(a)** 3.33 mA **(b)** 550 μA **(c)** 588 μA

(d) 500 mA **(e)** 6.60 mA

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

7. **(a)** 2.50 mA **(b)** 2.27 μA **(c)** 8.33 mA

9. $I = 0.642$ A, so the 0.5 A fuse will blow.

11. **(a)** 10 mV **(b)** 1.65 V **(c)** 14.1 kV
(d) 3.52 V **(e)** 250 mV **(f)** 750 kV
(g) 8.5 kV **(h)** 3.53 mV

13. **(a)** 81 V **(b)** 500 V **(c)** 117.5 V

15. **(a)** 2 kΩ **(b)** 3.5 kΩ **(c)** 2 kΩ
(d) 100 kΩ **(e)** 1 MΩ

17. **(a)** 4 Ω **(b)** 3 kΩ **(c)** 200 kΩ

19. 2.6 W

21. 417 mW

23. **(a)** 1 MW **(b)** 3 MW **(c)** 150 MW **(d)** 8.7 MW

25. **(a)** 2,000,000 μW **(b)** 500 μW
(c) 250 μW **(d)** 6.67 μW

27. $P = W/t$ in watts; $V = W/Q$; $I = Q/t$. $P = VI = W/t$, so (1 V)(1 A) = 1 watt

29. 16.5 mW

31. 1.18 kW

33. 5.81 W

35. 25 Ω

37. 0.00186 kWh

39. 156 mW

41. 1 W

43. **(a)** positive at top **(b)** positive at bottom
(c) positive at right

45. 36 Ah

47. 13.5 mA

49. 4.25 W

51. Five

53. 150 Ω

55. $V = 0$ V, $I = 0$ A; $V = 10$ V, $I = 100$ mA;
$V = 20$ V, $I = 200$ mA; $V = 30$ V, $I = 300$ mA;
$V = 40$ V, $I = 400$ mA; $V = 50$ V, $I = 500$ mA;
$V = 60$ V, $I = 600$ mA; $V = 70$ V, $I = 700$ mA;
$V = 80$ V, $I = 800$ mA; $V = 90$ V, $I = 900$ mA;
$V = 100$ V, $I = 1$ A

6238262 2013/12/10 12.27.66.5

57. $R_1 = 0.5\ \Omega$; $R_2 = 1\ \Omega$; $R_3 = 2\ \Omega$

59. 10 V; 30 V

61. 216 kWh

63. 12 W

65. 2.5 A

67. Power will increase by four times.

69. See Figure P–2.

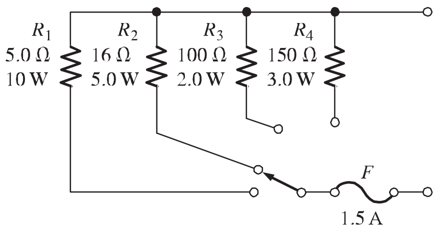

FIGURE P–2

71. No fault

73. Lamp 4 is shorted.

Chapter 4

1. See Figure P–3.

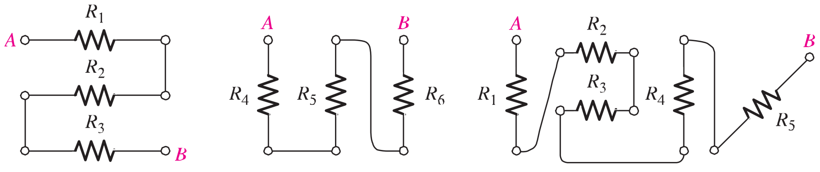

FIGURE P–3

3. 170 kΩ

5. 138 Ω

7. **(a)** 7.9 kΩ **(b)** 33 Ω **(c)** 13.24 MΩ
The series circuit is disconnected from the source, and the ohmmeter is connected across the circuit terminals.

9. 1126 Ω

11. **(a)** 170 kΩ **(b)** 50 Ω
(c) 12.4 kΩ **(d)** 1.97 kΩ

13. 0.1 A

15. **(a)** 625 μA
(b) 4.26 μA. The ammeter is connected in series.

17. **(a)** 34.0 mA **(b)** 16 V **(c)** 0.543 W

19. See Figure P–4.

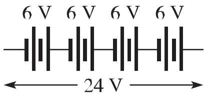

FIGURE P–4

21. 26 V

23. **(a)** $V_2 = 6.8$ V
(b) $V_R = 8$ V, $V_{2R} = 16$ V, $V_{3R} = 24$ V, $V_{4R} = 32$ V
The voltmeter is connected across (in parallel with) each resistor for which the voltage is unknown.

25. **(a)** 3.84 V **(b)** 6.77 V

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

27. 3.80 V; 9.38 V

29. $V_{5.6\text{ k}\Omega} = 10$ V; $V_{1\text{ k}\Omega} = 1.79$ V; $V_{560\ \Omega} = 1$ V; $V_{10\text{ k}\Omega} = 17.9$ V

31. 55 mW

33. Measure V_A and V_B individually with respect to ground; then $V_{R2} = V_A - V_B$.

35. 4.27 V

37. **(a)** R_4 open **(b)** R_4 and R_5 shorted

39. 780 Ω

41. $V_A = 10$ V; $V_B = 7.72$ V; $V_C = 6.68$ V; $V_D = 1.81$ V; $V_E = 0.57$ V; $V_F = 0$ V

43. 500 Ω

45. **(a)** 19.1 mA **(b)** 45.8 V

(c) $R(⅛\text{ W}) = 343\ \Omega$, $R(¼\text{ W}) = 686\ \Omega$, $R(½\text{ W}) = 1371\ \Omega$

47. See Figure P–5.

FIGURE P–5

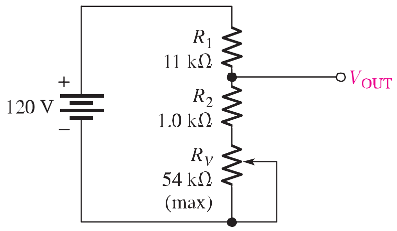

49. $R_1 + R_7 + R_8 + R_{10} = 4.23$ kΩ;

$R_2 + R_4 + R_6 + R_{11} = 23.6$ kΩ;

$R_3 + R_5 + R_9 + R_{12} = 19.9$ kΩ

51. A: 5.45 mA; B: 6.06 mA; C: 7.95 mA; D: 12 mA

53. A: $V_1 = 6.03$ V, $V_2 = 3.35$ V, $V_3 = 2.75$ V, $V_4 = 1.88$ V, $V_5 = 4.0$ V;

B: $V_1 = 6.71$ V, $V_2 = 3.73$ V, $V_3 = 3.06$ V, $V_5 = 4.5$ V;

C: $V_1 = 8.1$ V, $V_2 = 4.5$ V, $V_5 = 5.4$ V;

D: $V_1 = 10.8$ V, $V_5 = 7.2$ V

55. Yes, R_3 and R_5 are shorted.

57. **(a)** R_{11} has burned open due to excessive power.

(b) Replace R_{11} (10 kΩ). **(c)** 338 V

59. $I = 375$ mA

61. $I_T = 5.0$ A

63. R_6 is shorted.

65. Lamp 4 is open.

67. The 82 Ω resistor is shorted.

Chapter 5

1. See Figure P–6.

FIGURE P–6

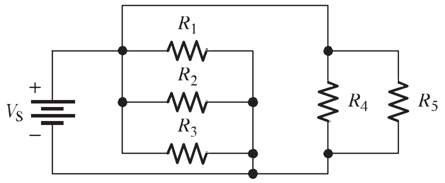

3. 3.43 kΩ

5. **(a)** 25.6 Ω **(b)** 359 Ω **(c)** 819 Ω **(d)** 996 Ω

7. 2 kΩ

9. 12 V; 5 mA

11. **(a)** 909 μA **(b)** 76 mA

13. **(a)** $I_1 = 179$ μA; $I_2 = 455$ μA

(b) $I_1 = 444$ μA; $I_2 = 80$ μA

15. 1350 mA

17. $I_2 = I_3 = 7.5$ mA. Connect an ammeter in series with each resistor in each branch.

19. 6.4 A; 6.4 A

21. $I_1 = 2.19$ A; $I_2 = 811$ mA

23. 200 mW

25. 0.625 A; 3.75 A

27. The 1.0 kΩ resistor is open.

29. R_2 is open.

31. $R_2 = 25\ \Omega$; $R_3 = 100\ \Omega$; $R_4 = 12.5\ \Omega$

33. $I_R = 4.8$ A; $I_{2R} = 2.4$ A; $I_{3R} = 1.6$ A; $I_{4R} = 1.2$ A

35. **(a)** $R_1 = 100\ \Omega$, $R_2 = 200\ \Omega$, $I_2 = 50$ mA

(b) $I_1 = 125$ mA, $I_2 = 74.9$ mA, $R_1 = 80\ \Omega$, $R_2 = 134\ \Omega$, $V_S = 10$ V

(c) $I_1 = 253$ mA, $I_2 = 147$ mA, $I_3 = 100$ mA, $R_1 = 395\ \Omega$

37. 53.7 Ω

39. Yes; total current = 14.7 A

41. $R_1 \| R_2 \| R_5 \| R_9 \| R_{10} \| R_{12} = 100\text{ k}\Omega \| 220\text{ k}\Omega \| 560\text{ k}\Omega \| 390\text{ k}\Omega \| 1.2\text{ M}\Omega \| 100\text{ k}\Omega = 33.6\text{ k}\Omega$

$R_4 \| R_6 \| R_7 \| R_8 = 270\text{ k}\Omega \| 1.0\text{ M}\Omega \| 820\text{ k}\Omega \| 680\text{ k}\Omega = 135.2\text{ k}\Omega$

$R_3 \| R_{11} = 330\text{ k}\Omega \| 1.8\text{ M}\Omega = 278.9\text{ k}\Omega$

43. $R_2 = 750\ \Omega$; $R_4 = 423\ \Omega$

45. The 4.7 kΩ resistor is open.

47. **(a)** 75.0 Ω **(b)** 26.6 W

49. 8.9 Ω

51. R_3 is open.

53. **(a)** R from pin 1 to pin 4 agrees with calculated value.

(b) R from pin 2 to pin 3 agrees with calculated value.

Chapter 6

1. R_2, R_3, and R_4 are in parallel, and this parallel combination is in series with both R_1 and R_5.

3. See Figure P–7.

5. 2003 Ω

7. **(a)** 128 Ω **(b)** 791 Ω

9. **(a)** $I_1 = I_4 = 11.7$ mA, $I_2 = I_3 = 5.85$ mA; $V_1 = 655$ mV, $V_2 = V_3 = 585$ mV, $V_4 = 257$ mV

(b) $I_1 = 3.8$ mA, $I_2 = 618$ μA, $I_3 = 1.27$ mA, $I_4 = 1.91$ mA; $V_1 = 2.58$ V, $V_2 = V_3 = V_4 = 420$ mV

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

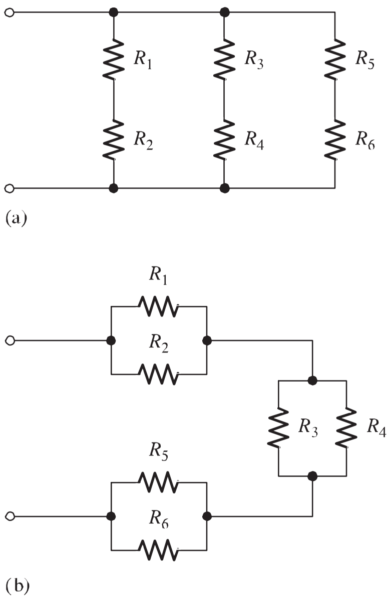
R_1
R_3
R_5
R_2
R_4
R_6
(a)
R_1
R_2
R_3
R_4
R_5
R_6
(b)

FIGURE P–7

11. 2.22 mA

13. 7.5 V unloaded; 7.29 V loaded

15. 56 kΩ load

17. 22 kΩ

19. 2 V

21. 33%

23. 360 Ω

25. 7.33 kΩ

27. $R_{TH} = 18$ kΩ; $V_{TH} = 2.7$ V

29. 1.06 V; 226 μA

31. 75 Ω

33. 21 mA

35. No, the meter should read 4.39 V. The 680 Ω resistor is open.

37. The 7.62 V and 5.24 V readings are incorrect, indicating that the 3.3 kΩ resistor is open.

39. **(a)** $V_1 = -10$ V, all others 0 V

(b) $V_1 = -2.33$ V, $V_4 = -7.67$ V, $V_2 = -7.67$ V, $V_3 = 0$ V

(c) $V_1 = -2.33$ V, $V_4 = -7.67$ V, $V_2 = 0$ V, $V_3 = -7.67$ V

(d) $V_1 = -10$ V, all others 0 V

41. See Figure P–8.

43. $R_T = 5.76$ kΩ; $V_A = 3.3$ V; $V_B = 1.7$ V; $V_C = 850$ mV

45. $V_1 = 1.61$ V; $V_2 = 6.77$ V; $V_3 = 1.72$ V; $V_4 = 3.33$ V; $V_5 = 378$ mV; $V_6 = 2.57$ V; $V_7 = 378$ mV; $V_8 = 1.72$ V; $V_9 = 1.61$ V

47. 110 kΩ

49. $R_1 = 180$ Ω; $R_2 = 60$ Ω. Output across R_2.

51. 845 μA

53. Pos 1: $V_1 = 88.0$ V, $V_2 = 58.7$ V, $V_3 = 29.3$ V

Pos 2: $V_1 = 89.1$ V, $V_2 = 58.2$ V, $V_3 = 29.1$ V

Pos 3: $V_1 = 89.8$ V, $V_2 = 59.6$ V, $V_3 = 29.3$ V

55. $V_G = 0$ V

57. The 2.2 kΩ resistor is open.

59. $V_B = 12.0$ V

61. $V_{4mA} = 0.60$ V; $V_{20mA} = 3.0$ V

63. Three load cells can be evenly loaded because three points define a plane.

65. R_2 is shorted.

67. No fault

69. R_4 is shorted.

71. R_5 is shorted.

Chapter 7

1. Decrease

3. 37.5 μWb

5. 1000 G

7. 597

9. 1500 At

11. **(a)** Electromagnetic force **(b)** Spring force

13. Electromagnetic force

15. Change the current.

17. Material A

19. 1 mA

21. 3.02 m/s

23. 56.3 W

25. **(a)** 168 W **(b)** 14 W

27. 80.6%

29. The output voltage has a 10 V dc peak with a 120 Hz ripple.

31. The coil is energized because the movable contacts will be in the lower position for the polarity to be reversed.

FIGURE P–8

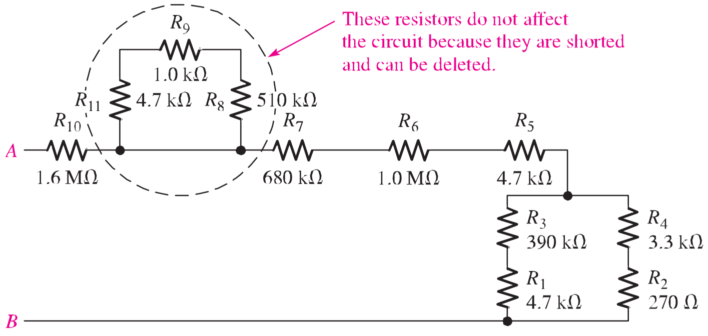
R_9
1.0 kΩ
R_{11}
4.7 kΩ
R_8
510 kΩ
These resistors do not affect the circuit because they are shorted and can be deleted.
R_{10}
A
1.6 MΩ
R_7
680 kΩ
R_6
1.0 MΩ
R_5
4.7 kΩ
R_3
390 kΩ
R_4
3.3 kΩ
R_1
4.7 kΩ
R_2
270 Ω
B

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

33. Design is flawed. Twelve volts is not enough voltage to operate two 12 V relays in series; 24 V is too much voltage to operate a 12 V lamp. Install a separate 12 V power supply for the lamp and change the 12 V to 24 V for the relays.

Chapter 8

1. **(a)** 1 Hz **(b)** 5 Hz **(c)** 20 Hz
(d) 1 kHz **(e)** 2 kHz **(f)** 100 kHz

3. 2 μs

5. 10 ms

7. **(a)** 7.07 mA **(b)** 4.5 mA **(c)** 14.14 mA

9. **(a)** 17.7 V **(b)** 25 V **(c)** 0 V **(d)** −17.7 V

11. 15°, *A* leads.

13. See Figure P–9.

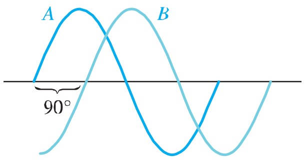

FIGURE P–9

15. **(a)** 22.5° **(b)** 60° **(c)** 90°
(d) 108° **(e)** 216° **(f)** 324°

17. **(a)** 57.4 mA **(b)** 99.6 mA **(c)** −17.4 mA
(d) −57.4 mA **(e)** −99.6 mA **(f)** 0 mA

19. 30°: 13.0 V; 45°: 14.5 V; 90°: 13.0 V; 180°: −7.5 V; 200°: −11.5 V; 300°: −7.5 V

21. **(a)** 7.07 mA **(b)** 0 A **(c)** 10 mA
(d) 20 mA **(e)** 10 mA

23. 7.38 V

25. 4.24 V

27. 250 Hz

29. 200 rps

31. A one phase motor requires a starting winding or other means to produce torque for starting the motor, whereas a three phase motor is self starting.

33. $t_r \cong 3.0$ ms; $t_f \cong 3.0$ ms; $t_W \cong 12.0$ ms; Ampl. $= 5$ V

35. **(a)** −0.375 V **(b)** 3.01 V

37. **(a)** 50 kHz **(b)** 10 Hz

39. 25 kHz

41. 0.424 V; 2 Hz

43. 1.4 V; 120 ms; 30%

45. A modulating signal is a lower frequency waveform that modifies a higher frequency waveform. It can be applied from an external source or it can be generated internally.

47. Burst mode outputs a specific number of cycles of a waveform; gated mode outputs the waveform for a set period of time as set by a gate signal.

49. $I_{\text{max}} = 2.38$ A; $V_{\text{avg}} = 136$ V; See Figure P–10.

FIGURE P–10

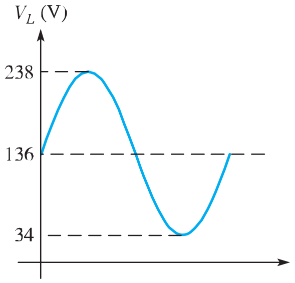

51. **(a)** 2.5 **(b)** 3.96 V **(c)** 12.5 kHz

53. See Figure P–11.

FIGURE P–11

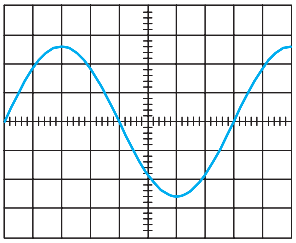

55. $V_{p(in)} = 4.44$ V; $f_{in} = 2$ Hz

57. R_3 is open.

59. 5 V; 1 ms

Chapter 9

1. **(a)** 5 μF **(b)** 1 μC **(c)** 10 V

3. **(a)** 0.001 μF **(b)** 0.0035 μF **(c)** 0.00025 μF

5. 2 μF

7. 88.5 pF

9. 0.0249 μF

11. A 12.5 pF increase

13. Ceramic

15. **(a)** 0.022 μF **(b)** 0.047 μF
(c) 0.001 μF **(d)** 22 pF

17. **(a)** Encapsulation **(b)** Dielectric (ceramic disk)
(c) Plate (metal disk) **(d)** Leads

19. **(a)** 0.69 μF **(b)** 69.7 pF **(c)** 2.6 μF

21. $V_1 = 2.13$ V; $V_2 = 10$ V; $V_3 = 4.55$ V; $V_4 = 1$ V

23. 5.5 μF, 27.5 μC

25. **(a)** 100 μs **(b)** 560 μs **(c)** 22.1 μs **(d)** 15 ms

27. **(a)** 9.48 V **(b)** 13.0 V **(c)** 14.3 V
(d) 14.7 V **(e)** 14.9 V

29. **(a)** 2.72 V **(b)** 5.90 V **(c)** 11.7 V

31. **(a)** 339 kΩ **(b)** 13.5 kΩ
(c) 677 Ω **(d)** 33.9 Ω

6238262 2013/12/10 12.27.66.5

33. $X_{C1} = 1.42$ kΩ; $X_{C2} = 970$ Ω; $X_{CT} = 2.39$ kΩ
$V_1 = 5.94$ V; $V_2 = 4.06$ V

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

35. 200 Ω

37. $P_{\text{true}} = 0$ W; $P_r = 3.39$ mVAR

39. 0 Ω

41. 3.18 ms

43. 3.24 μs

45. **(a)** Charges to 3.32 V in 10 ms, then discharges to 0 V in 215 ms.

(b) Charges to 3.32 V in 10 ms, then discharges to 2.96 V in 5 ms, then charges toward 20 V.

47. $C_1 = 1.25$ nF

49. 16 Hz

51. C_2 is open.

53. No fault

Chapter 10

1. 8 kHz; 8 kHz

3. **(a)** 288 Ω **(b)** 1209 Ω

5. **(a)** 726 kΩ **(b)** 155 kΩ

(c) 91.5 kΩ **(d)** 63.0 kΩ

7. **(a)** 34.7 mA **(b)** 4.14 mA

9. $I_{tot} = 12.3$ mA; $V_{C1} = 1.31$ V; $V_{C2} = 0.595$ V; $V_R = 0.616$ V; $\theta = 72.0°$ (V_s lagging I_{tot})

11. 808 Ω; −36.1°

13. **(a)** 90° **(b)** 86.4° **(c)** 57.8° **(d)** 9.04°

15. 326 Ω; 64.3°

17. 245 Ω; 80.5°

19. $I_{C1} = 118$ mA; $I_{C2} = 55.3$ mA; $I_{R1} = 36.4$ mA; $I_{R2} = 44.4$ mA; $I_{tot} = 191$ mA; $\theta = 65.0°$ (V_s lagging I_{tot})

21. **(a)** 3.86 kΩ **(b)** 21.3 μA **(c)** 14.8 μA

(d) 25.9 μA **(e)** 34.8° (V_s lagging I_{tot})

23. $V_{C1} = 8.74$ V; $V_{C2} = 3.26$ V; $V_{C3} = 3.26$ V; $V_{R1} = 2.11$ V; $V_{R2} = 1.15$ V;

25. $I_{tot} = 82.4$ mA; $I_{C2} = 14.4$ mA; $I_{C3} = 67.6$ mA; $I_{R1} = I_{R2} = 6.39$ mA

27. 4.03 VA

29. 0.915

31. Use the formula, $V_{out} = \left(\frac{X_C}{Z_{tot}}\right)1$ V. See Figure P–12.

FREQUENCY (KHZ)	X_C (kΩ)	Z_{tot} (kΩ)	V_{out} (V)
0			1.000
1	4.08	5.64	0.723
2	2.04	4.40	0.464
3	1.36	4.13	0.329
4	1.02	4.03	0.253
5	0.816	3.98	0.205
6	0.680	3.96	0.172
7	0.583	3.94	0.148
8	0.510	3.93	0.130
9	0.453	3.93	0.115
10	0.408	3.92	0.104

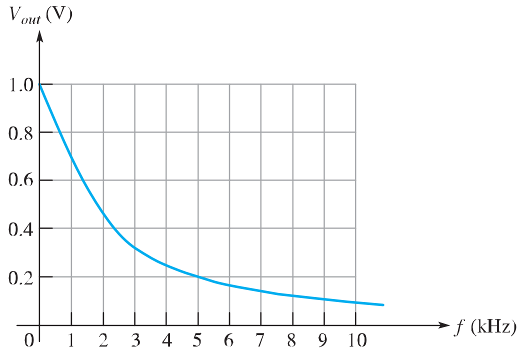

FIGURE P–12

33. See Figure P–13.

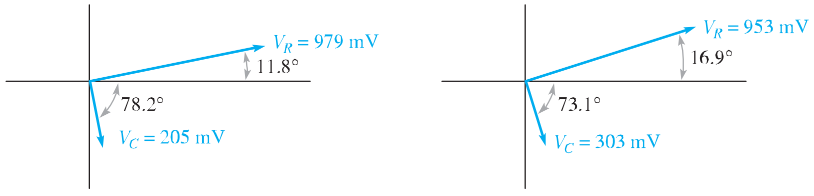

(a) For Figure 10–67

(b) For Figure 10–68

FIGURE P–13

35. Figure 10–67: 1.05 kHz; Figure 10–68: 1.59 kHz

37. No leakage: $V_{out} = 3.21$ V; $\theta = 18.7°$;

Leakage: $V_{out} = 2.83$ V, $\theta = 33.3°$

39. **(a)** 0 V **(b)** 0.321 V

(c) 0.5 V **(d)** 0 V

41. **(a)** $I_{L(A)} = 4.8$ A; $I_{L(B)} = 3.33$ A

(b) $P_{r(A)} = 606$ VAR; $P_{r(B)} = 250$ VAR

(c) $P_{\text{true}(A)} = 979$ W; $P_{\text{true}(B)} = 758$ W

(d) $P_{a(A)} = 1151$ VA; $P_{a(B)} = 798$ VA

43. 11.4 kΩ

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

45. $P_r = 1.32$ kVAR; $P_a = 2$ kVA

47. 0.103 μF

49. C is leaky.

51. No fault

53. R_2 is open.

55. Phase shift with C_1 shorted = 13.7°

Chapter 11

1. **(a)** 1000 mH **(b)** 0.25 mH
(c) 0.01 mH **(d)** 0.5 mH

3. 3450 turns

5. 50 mJ

7. 155 μH

9. 7.14 μH

11. **(a)** 50 mH **(b)** 57 μH

13. **(a)** 1 μs **(b)** 2.13 μs **(c)** 200 μs

15. **(a)** 5.52 V **(b)** 2.03 V **(c)** 0.747 V
(d) 0.275 V **(e)** 0.101 V

17. **(a)** 157 kΩ **(b)** 179 Ω

19. $I_{tot} = 10.1$ mA; $I_{L2} = 6.7$ mA; $I_{L3} = 3.37$ mA

21. 101 mVAR

23. **(a)** −3.35 V **(b)** −1.12 V **(c)** −0.37 V

25. **(a)** 0.427 mA **(b)** 0.569 mA

27. $X_{L1} = 69\ \Omega$; $X_{L2} = 1037\ \Omega$; $X_{LT} = 1106\ \Omega$

29. L_3 is open.

31. No fault

33. L_3 is shorted.

Chapter 12

1. 15 kHz

3. **(a)** 1.12 kΩ **(b)** 1.8 kΩ

5. **(a)** 17.4 Ω **(b)** 64 Ω **(c)** 127 Ω **(d)** 251 Ω

7. 100 kHz

9. **(a)** 8.94 mA **(b)** 2.77 mA

11. 38.7°

13. See Figure P–14.

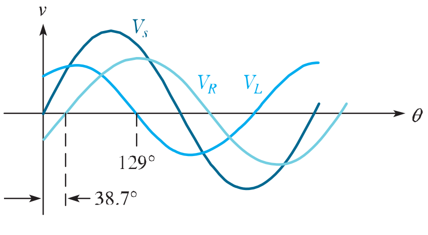

FIGURE P–14

15. 7.69 Ω

17. 537 μS

19. 2.39 kHz

21. **(a)** 274 Ω **(b)** 89.3 mA **(c)** 159 mA
(d) 183 mA **(e)** 60.7° (I_{tot} lagging V_s)

23. $V_{R1} = 7.92$ V; $V_{R2} = V_L = 20.8$ V

25. $I_{tot} = 36$ mA; $I_L = 33.2$ mA; $I_{R2} = 13.9$ mA

27. 13.0 mW; 10.4 mVAR

29. $PF = 0.386$; $P_{\text{true}} = 347$ mW; $P_r = 692$ mVAR;
$P_a = 900$ mVA

31. See Figure P–15.

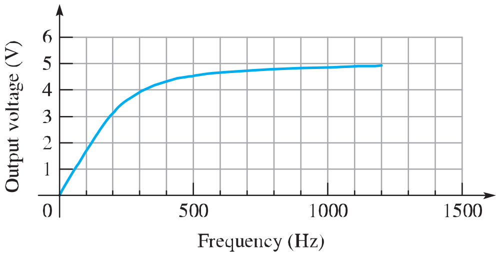

FIGURE P–15

33. $V_{R1} = V_{L1} = 18$ V; $V_{R2} = V_{R3} = V_{L2} = 0$ V

35. 5.57 V

37. 343 mA

39. **(a)** 405 mA **(b)** 228 mA
(c) 333 mA **(d)** 335 mA

41. L_2 is open.

43. R_2 is open.

45. L_1 is shorted.

Chapter 13

1. 520 Ω, 88.9° (V_s lagging I); 520 Ω capacitive

3. Impedance increases.

5. See Figure P–16.

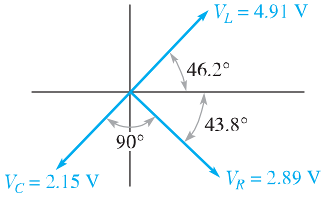

FIGURE P–16

7. f_r is lower.

9. $X_L = 4.61$ kΩ; $X_C = 4.61$ kΩ; $Z = 220\ \Omega$; $I = 54.5$ mA

11. $f_r = 454$ kHz; $f_1 = 416$ kHz; $f_2 = 492$ kHz

13. **(a)** 14.5 kHz; band-pass **(b)** 24.0 kHz; band-pass

15. **(a)** $f_r = 339$ kHz, $BW = 239$ kHz
(b) $f_r = 10.4$ kHz, $BW = 2.61$ kHz

17. Capacitive, $X_C < X_L$.

19. 1.47 kΩ

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

21. 53.1 MΩ; 104 kHz

23. 62.5 Hz

25. 1.38 W

27. 200 Hz

29. See Figure P–17.

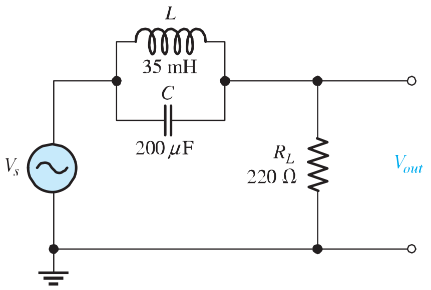

FIGURE P–17

31. $I_{R1} = I_C = 2.11$ mA; $I_{L1} = 1.33$ mA; $I_{L2} = 667$ μA; $I_{R2} = 667$ μA; $V_{R1} = 6.96$ V; $V_C = 2.11$ V; $V_{L1} = V_{L2} = V_{R2} = 6.67$ V

33. $I_{R1} = I_{L1} = 41.5$ mA; $I_C = I_{L2} = 133$ mA; $I_{tot} = 104$ mA

35. $L = 989$ μH; $C = 0.064$ μF

37. Neglecting R_W, 8 MHz: $C = 40$ pF; 9 MHz: $C = 31$ pF; 10 MHz: $C = 25$ pF; 11 MHz: $C = 21$ pF

39. No fault

41. L is shorted.

43. L is shorted.

Chapter 14

1. 1.5 μH

3. 3

5. **(a)** In phase **(b)** Out of phase **(c)** Out of phase

7. 500 turns

9. **(a)** Same polarity, 100 V rms

(b) Opposite polarity, 100 V rms

11. 240 V

13. **(a)** 6 V **(b)** 0 V **(c)** 40 V

15. **(a)** 10 V **(b)** 240 V

17. 33.3 mA

19. 27.2 Ω

21. 6.0 mA

23. 0.5

25. 5 kΩ

27. 94.5 W

29. 0.98

31. 25 kVA

33. Secondary 1: 2; Secondary 2: 0.5; Secondary 3: 0.25

35. Top secondary: $n = 100/1000 = 0.1$

Next secondary: $n = 200/1000 = 0.2$

Next secondary: $n = 500/1000 = 0.5$

Bottom secondary: $n = 1000/1000 = 1$

37. Excessive primary current is drawn, potentially burning out the source and/or the transformer, unless the primary is protected by a fuse.

39. **(a)** $V_{L1} = 35$ V, $I_{L1} = 2.92$ A, $V_{L2} = 15$ V, $I_{L2} = 1.5$ A

(b) 28.9 Ω

41. **(a)** 20 V **(b)** 10 V

43. 0.0141 (70.7:1)

45. 0.1 A

47. Secondary is open.

49. Primary is open.

Chapter 15

1. 103 μs

3. 27 kΩ

5. 12.6 V

7. See Figure P–18.

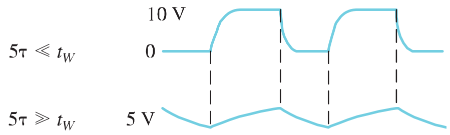

FIGURE P–18

9. See Figure P–19.

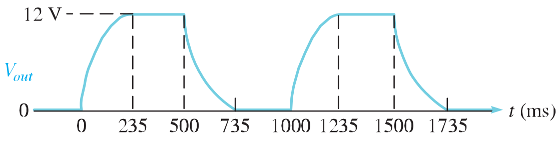

FIGURE P–19

11. 15 V dc level with a very small charge/discharge fluctuation

13. Exchange positions of R and C. See Figure P–20 for the output voltage. $5\tau = 5$ ms (repeating Problem 8)

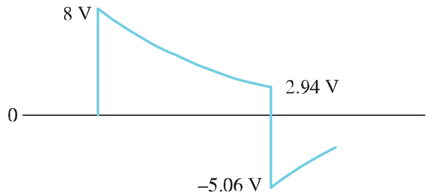

FIGURE P–20

15. Approximately the same shape as input, but with an average value of 0 V

17. See Figure P–21.

6238262 2013/12/10 13:27:66.5

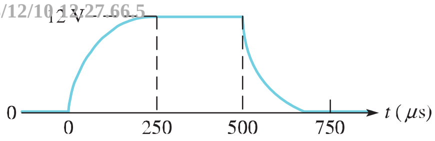

FIGURE P–21

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

19. See Figure P–22.

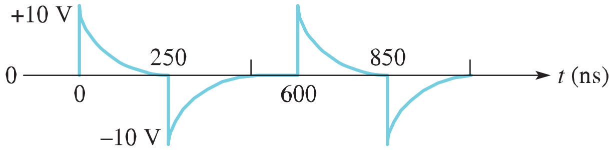

FIGURE P–22

21. A 6 V dc level

23. **(b)** No fault **(c)** Open *C* or shorted *R*

25. **(a)** 23.5 ms **(b)** See Figure P–23.

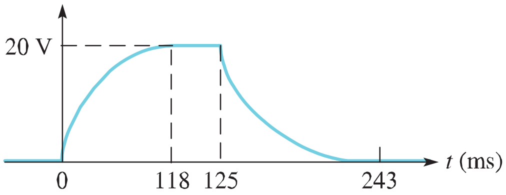

FIGURE P–23

27. 1.44 s

29. One method is as follows: Replace the input to the amplifier with a dc power supply set to the lowest voltage of the triangle wave. The VCO will have a constant output frequency, which is measured with a separate instrument and the oscilloscope horizontal axis will not move, so note the position. Increase the power supply voltage noting the frequency and the horizontal location on the oscilloscope (it will be in a new location on the screen). You can choose several points along the axis to calibrate the frequency as a function of horizontal position.

31. Capacitor is open.

33. No fault

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

This page intentionally left blank

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

GLOSSARY

accuracy An indication of the range of error in a measurement.

active component A component that requires electrical power to function correctly; active components can deliver more signal power than they receive from the input signal.

admittance (Y) A measure of the ability of a reactive circuit to permit current; the reciprocal of impedance. The unit is the siemens (S).

ah rating A capacity rating for batteries determined by multiplying the current in amps (A) times the length of time in hours (h) a battery can deliver that current to a load.

alternator An ac generator; alternators convert mechanical energy to electrical energy.

ammeter An electrical instrument used to measure current.

ampere (A or amp) The unit of electrical current.

ampere-turn The SI unit of magnetomotive force (mmf).

amplitude The maximum value of a voltage or current.

analog signal A continuous signal.

anode The terminal of a polarized device that electrons flow from. For a battery that is supplying current, this is the negative terminal.

apparent power (P_a) The product of the voltage times the current, expressed in volt-amperes (VA). In a purely resistive circuit it is the same as the true power.

apparent power rating The method of rating transformers in which the power capability is expressed in volt-amperes (VA).

armature The power-producing coil in an alternator, generator or motor. In dc generators, the armature is the rotor, but in alternators it can be either the rotor or the stator. In motors, the armature is the rotor.

atom The smallest particle of an element possessing the unique characteristics of that element.

atomic number The number of protons in the nucleus of an atom.

autotransformer A transformer in which the primary and secondary windings are in a single winding.

average value (1) The average value of a sine wave determined over one half-cycle as 0.637 times the peak value. (2) The average value of a pulse waveform is equal to its baseline value plus the product of its duty cycle and its amplitude.

AWG (American wire gauge) A standardization based on wire diameter.

back emf (electromotive force) A voltage developed across a spinning armature that opposes the original applied voltage.

balanced bridge A bridge circuit that is in the balanced state as indicated by zero volts across the bridge.

balun A type of transformer used to convert a balanced line (such as twisted pair wiring) to an unbalanced line (such as coaxial cable) or vice-versa.

band-pass filter A resonant circuit that passes a range of frequencies lying between two cutoff frequencies and rejects frequencies above and below the range.

band-stop filter A resonant circuit that rejects a range of frequencies lying between two cutoff frequencies and passes frequencies above and below the range.

bandwidth (BW) The characteristic of certain electronic circuits that specifies the usable range of frequencies for which signals pass from input to output without significant reduction in amplitude.

baseline The normal level of a pulse waveform; the voltage level in the absence of a pulse.

battery An energy source that uses a chemical reaction to convert chemical energy into electrical energy.

bias The application of a dc voltage to a diode or other electronic device to produce a desired mode of operation.

bleeder current The current left after the total load current is subtracted from the total current into the circuit.

block diagram A model of a system that represents its structure in a graphical format using labeled blocks to represent functions and lines to represent the signal flow.

boundary The dividing line between what is part of the system and its environment.

branch One current path in a parallel circuit.

bypass A capacitor connected from a point to ground to remove the ac signal without affecting the dc voltage. A special case of decoupling.

capacitance The ability of a capacitor to store electrical charge.

capacitive reactance The opposition of a capacitor to sinusoidal current. The unit is the ohm (Ω).

capacitive susceptance (B_C) The ability of a capacitor to permit current; the reciprocal of capacitive reactance. The unit is the siemens (S).

capacitor An electrical device that has the ability to store charge. It consists of one or more conductors separated by an insulating material called the *dielectric*.

6238262 2013/12/10 12.27.66.5

carrier A high-frequency radio wave that can be modulated (changed) by a lower frequency signal.

cathode The terminal of a polarized device that electrons flow to. For a battery that is supplying current, this is the positive terminal.

center tap (CT) A connection at the midpoint of the secondary winding of a transformer.

charge An electrical property of matter that exists because of an excess or a deficiency of electrons. Charge can be either positive or negative.

charging For a capacitor, the process in which a current removes charge from one plate of a capacitor and deposits it on the other plate, making one plate more positive than the other.

choke An inductor. The term is used more commonly in connection with inductors used to block or choke off high frequencies.

circuit An interconnection of electrical components designed to produce a desired result. A basic circuit consists of a source, a load, and an interconnecting current path.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

circuit breaker A resettable protective device used for interrupting excessive current in an electric circuit.

circular mil (CM) The unit of the cross-sectional area of a wire.

closed circuit A circuit with a complete current path.

coefficient of coupling (k) A constant associated with transformers that is the ratio of secondary magnetic flux to primary magnetic flux. The ideal value of 1 indicates that all the flux in the primary winding is coupled into the secondary winding.

coil A common term for an inductor.

color code A system of color bands or dots that identify the value of a resistor or other component.

common A reference point in a circuit; also called *reference ground*.

conductance (G) The ability of a circuit to allow current; the reciprocal of resistance. The unit is the siemens (S).

conductor A material in which electrical current is established with relative ease. An example is copper.

core The structure around which the winding of an inductor is formed. The core material influences the electromagnetic characteristics of the inductor.

coulomb (C) The unit of electrical charge; the total charge possessed by 6.25×10^{18} electrons.

Coulomb's law A law that states a force exists between two charged bodies that is directly proportional to the product of the two charges and inversely proportional to the square of the distance between them.

coupling The method of connecting a capacitor from between two points in a circuit to allow ac to pass from one point to the other while blocking dc.

covalent Related to the bonding of two or more atoms by the interaction of their valence electrons.

crystal The pattern or arrangement of atoms forming a solid material.

current The rate of flow of charge (free electrons).

current divider A parallel circuit in which the total current divides among the branches.

current source A device that produces a constant current for a varying load.

cycle One repetition of a periodic waveform.

dc component The average value of a pulse waveform.

dc power supply An electronic instrument that produces voltage, current, and power from the ac power line or batteries in a form suitable for use in powering electronic equipment.

decade A tenfold change in the value of a quantity. When a quantity becomes ten times less or ten times greater, it has changed a decade.

decibel (dB) A unit of logarithmic expression ten times the logarithm of the ratio of two powers or twenty times the logarithm of the ratio of two voltages measured across equal resistances.

decoupling The method of connecting a capacitor from one point, usually the dc power supply line, to ground to short ac to ground without affecting the dc voltage.

degree The unit of angular measure corresponding to $1/360$ of a complete revolution.

dielectric The insulating material between the plates of a capacitor.

dielectric constant A measure of the ability of a dielectric material to establish an electric field.

dielectric strength A measure of the ability of a dielectric material to withstand voltage without breaking down.

digital signal A signal that has discrete levels.

DMM Digital multimeter; an electronic instrument that combines meters for the measurement of voltage, current, and resistance.

dummy load A precision high-power resistor.

duty cycle A characteristic of a pulse waveform that indicates the percentage of time that a pulse is present during a cycle; the ratio of pulse width to period.

effective value A measure of the heating effect of a sine wave; also known as the *rms* (root mean square) *value*.

efficiency The ratio of the output power to the input power of a circuit, usually expressed as a percentage.

electrical Related to the use of electrical voltage and current to achieve desired results.

electrical isolation The condition in which two circuits have no common conductive path between them.

electrical shock The physical sensation resulting from current through the body.

electromagnetic field A formation of a group of magnetic lines of force surrounding a conductor created by electrical current in the conductor.

electromagnetic induction The phenomenon or process by which a voltage is produced in a conductor when there is relative motion between the conductor and a magnetic or electromagnetic field.

electromagnetism The production of a magnetic field by current in a conductor.

electron A basic particle of electrical charge in matter. The electron possesses negative charge.

electronic Related to the movement and control of free electrons in semiconductors or vacuum devices.

element One of the unique substances that make up the known universe. Each element is characterized by a unique atomic structure.

energy The ability to do work. The unit is the joule (J).

engineering notation A system for representing any number as a one-, two-, or three-digit number times a power of ten with an exponent that is a multiple of 3.

error The difference between the true or best-accepted value of some quantity and the measured value.

exciter A separate dc generator used supply current to the field coils of a large generator or alternator. Normally, excitation current is controlled automatically.

exponent The number to which a base number is raised.

exponential A mathematical function described by a natural logarithm (base). Increasing and decreasing current and voltage for a capacitor or inductor are described by exponential functions.

falling edge The negative-going transition of a pulse.

fall time (t_f) The time interval required for a pulse to change from 90% to 10% of its amplitude.

farad (F) The unit of capacitance.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Faraday's law A law stating that the voltage induced across a coil of wire equals the number of turns in the coil times the rate of change of the magnetic flux.

filter A type of circuit that passes certain frequencies and rejects all others.

free electron A valence electron that has broken away from its parent atom and is free to move from atom to atom within the atomic structure of a material.

frequency A measure of the rate of change of a periodic function; the number of cycles completed in 1 s. The unit of frequency is the hertz.

frequency response In electrical circuits, the variation in the output voltage (or current) over a specified range of frequencies.

fuel cell A device that converts electrochemical energy from an external source into dc voltage. The hydrogen fuel cell is the most common type.

function generator An instrument that produces more than one type of waveform.

fundamental frequency The repetition rate of a waveform.

fuse A protective device that burns open when there is excessive current in a circuit.

gain The ratio of the output to the input for an amplifier.

gauss A CGS unit of flux density.

generator An energy source that produces electrical signals.

ground The common or reference point in a circuit.

ground plane A conducting surface used as a reference point for circuit returns or in antenna systems to image a radiating structure.

half-power frequency The frequency at which the output power of a filter is 50% of the maximum (the output voltage is 70.7% of maximum); another name for *critical* frequency or *cutoff* frequency.

half-splitting A troubleshooting procedure where one starts in the middle of a circuit or system and, depending on the first measurement, works toward the output or toward the input to find the fault.

Hall effect A change in current density across a conductor or semiconductor when current in the material is perpendicular to a magnetic field. The change in current density produces a small transverse voltage in the material, called the Hall voltage.

harmonics The frequencies contained in a composite waveform, which are integer multiples of the repetition frequency (fundamental).

henry (H) The unit of inductance.

hertz (Hz) The unit of frequency. One hertz equals one cycle per second.

high-pass circuit A certain type of circuit whereby higher frequencies are passed and lower frequencies are rejected.

horizontal organization A business structure that is decentralized, to allow managers to focus on a specialty and streamline decision making.

hypotenuse The longest side of a right triangle.

hysteresis A characteristic of a magnetic material whereby a change in magnetization lags the application of a magnetic force.

impedance (Z) The total opposition to sinusoidal current expressed in ohms.

impedance matching A technique used to match a load resistance to an internal source resistance in order to achieve a maximum transfer of power.

induced current (i_{ind}) A current induced in a conductor as a result of a changing magnetic field.

induced voltage (v_{ind}) Voltage produced as a result of a changing magnetic field.

inductance (L) The property of an inductor that produces an opposition to any change in current.

induction motor An ac motor that achieves excitation to the rotor by transformer action.

inductive reactance (X_L) The opposition of an inductor to sinusoidal current. The unit is the ohm.

inductive susceptance (B_L) The ability of an inductor to permit current; the reciprocal of inductive reactance. The unit is the siemens.

inductor A passive electrical component formed by a coil of wire and which exhibits the property of inductance.

input The voltage, current, or power applied to an electrical circuit to produce a desired result.

instantaneous power The value of power in a circuit at any given instant of time.

instantaneous value The voltage or current value of a waveform at a given instant in time.

insulator A material that does not allow current under normal conditions.

integrated circuit An active complex circuit consisting of resistors, transistors, and other components fabricated as a single unit that performs the function of multiple discrete components.

integrator A circuit that produces an output that approaches the mathematical integral of the input.

interface To make the output of one type of circuit compatible with the input of another so that they can operate properly together.

ion An atom or group of atoms with a net charge.

joule (J) The SI unit of energy.

kilowatt-hour (kWh) A large unit of energy used mainly by utility companies.

Kirchhoff's current law A law stating that the total current into a node equals the total current out of the node.

Kirchhoff's voltage law A law stating that (1) the sum of the voltage drops around a single closed loop (path) equals the source voltage in that loop or (2) the algebraic sum of all the voltages around a single closed path in a circuit is zero.

ladder diagram An electrical schematic of a parallel circuit which resembles a ladder. Two rails represent the two nodes that have the applied voltage; any number of "rungs", which are loads, complete the ladder.

ladder logic A graphical computer language used for programming PLCs. Programming is accomplished by placing simulated relays, contacts, switches, timers, and other control elements in "rungs" to allow the PLC to control an industrial process.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

lag To be behind; describes a condition of the phase or time relationship of waveforms in which one waveform is behind the other in phase or time.

lead To be ahead; describes a condition of the phase or time relationship of waveforms in which one waveform is ahead of the other in phase or time; also, a wire or cable connection to a device or instrument.

leading edge The first step or transition of a pulse.

LED (light-emitting diode) A type of diode that emits light when there is forward current.

Lenz's law A law that states when the current through a coil changes, the polarity of the induced voltage created by the changing magnetic field is such that it always opposes the change that caused it. The current cannot change instantaneously.

linear Characterized by a straight-line relationship.

lines of force Magnetic flux lines that exist in a magnetic field radiating from the north pole to the south pole.

load An element (resistor or other component) connected across the output terminals of a circuit that draws current from the source and upon which work is done.

load cell A transducer that uses strain gauges to convert mechanical force into an electrical signal.

load current The output current of a circuit supplied to a load.

loading The effect on a circuit when an element that draws current from the circuit is connected across the output terminals.

low-pass circuit A certain type of circuit in which lower frequencies are passed and higher frequencies are rejected.

magnetic coupling The magnetic connection between two coils as a result of the changing magnetic flux lines of one coil cutting through the second coil.

magnetic field A force field radiating from the north pole to the south pole of a magnet.

magnetic field intensity The amount of mmf per unit length of magnetic material. The unit is the ampere-turns/meter (At/m); also called *magnetizing force*.

magnetic flux The lines of force between the north and south poles of a permanent magnet or an electromagnet.

magnetic flux density The number of lines of force per unit area perpendicular to a magnetic field.

magnetomotive force The cause of a magnetic field; measured in ampere-turns.

magnitude The value of a quantity, such as the number of volts of voltage or the number of amperes of current.

maximum power transfer The condition, when the load resistance equals the source resistance, under which maximum power is transferred from source to load.

maximum power transfer theorem A theorem that states the maximum power is transferred from a source to a load when the load resistance equals the internal source resistance.

mechatronics A synergistic combination of mechanics and electronics including instrumentation and control systems.

metric prefix A symbol that is used to replace the power of ten in numbers expressed in engineering notation.

modulate To vary a characteristic (amplitude, frequency or phase) of an electromagnetic wave.

multimeter An instrument that measures voltage, current, and resistance.

mutual inductance (L_M) The inductance between two separate coils, such as in a transformer.

negative ion An atom or group of atoms with a net negative charge.

neutron An atomic particle having no electrical charge.

node A point or junction in a circuit at which two or more components are connected.

nucleus The central part of an atom containing protons and neutrons.

ohm (Ω) The unit of resistance.

ohmmeter An instrument for measuring resistance.

Ohm's law A law stating that current is directly proportional to voltage and inversely proportional to resistance.

open A circuit condition in which the current path is interrupted.

open circuit A circuit in which there is not a complete current path.

orbit The path an electron takes as it circles around the nucleus of an atom.

oscilloscope A measurement instrument that displays signal waveforms on a screen.

oscillator An electronic circuit that produces a repetitive waveform on its output with only the dc supply voltage as an input.

output The result obtained from the system after processing the inputs.

parallel The relationship between two circuit components that exists when they are connected between the same pair of nodes.

parallel resonance In a parallel *RLC* circuit, the condition where the impedance is maximum and the reactances are equal.

passband The range of frequencies passed by a filter.

passive component A component that does not require power; passive components cannot increase signal power.

peak-to-peak value The voltage or current value of a waveform measured from its minimum to its maximum points.

peak value The voltage or current value of a waveform at its maximum positive or negative points.

period (T) The time interval of one complete cycle of a given sine wave or any periodic waveform.

periodic Characterized by a repetition at fixed time intervals.

permeability The measure of ease with which a magnetic field can be established in a material.

phase The relative displacement of a time-varying waveform in terms of its occurrence with respect to a reference.

phase angle The angle between the source voltage and the total current in a reactive circuit.

phasor A representation of a sine wave in terms of both magnitude and phase angle.

photoconductive cell A type of variable resistor that is light-sensitive.

photodiode A diode whose reverse resistance changes with incident light.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

photon A discrete bundle of electromagnetic energy. In a vacuum, photons travel at the speed of light and have no rest mass.

photovoltaic effect The process whereby light energy is converted directly into electrical energy.

piezoelectric effect The property of a crystal whereby a changing mechanical stress produces a voltage across the crystal.

pole In practical terms, a single *RC* circuit in a filter or amplifier that causes the response to change at a 20 dB per decade rate above or below a certain frequency.

positive ion An atom or group of atoms with a net positive charge.

potentiometer A three-terminal variable resistor.

power The rate of energy usage. The unit is the watt (W).

power factor The relationship between volt-amperes and true power or watts. Volt-amperes multiplied by the power factor equals true power.

power of ten A numerical representation consisting of a base of 10 and an exponent; the number 10 raised to a power.

power rating The maximum amount of power that a resistor can dissipate without being damaged by excessive heat buildup.

power supply A device that converts the ac utility voltage into a constant dc.

precision A measure of the repeatability (or consistency) of a series of measurements.

primary winding The input winding of a transformer; also called *primary.*

probe An accessory used to connect a voltage to the input of an oscilloscope or other instrument.

proton A positively charged atomic particle.

pulse A type of waveform that consists of two equal and opposite steps in voltage or current separated by a time interval.

pulse repetition frequency The fundamental frequency of a repetitive pulse waveform; the rate at which the pulses repeat expressed in either hertz or pulses per second.

pulse width (t_W) The time interval between the opposite steps of an ideal pulse. For a nonideal pulse, the time between the 50% points on the leading and trailing edges.

quality factor (Q) The ratio of reactive power to true power in an inductor or a resonant circuit.

radian A unit of angular measurement. There are 2π radians in one complete revolution. One radian equals 57.3°.

Radiometric image A thermal picture in which colors represent temperature.

ramp A type of waveform characterized by a linear increase or decrease in voltage or current.

***RC* lag circuit** A phase shift circuit in which the output voltage, taken across the capacitor, lags the input voltage by a specified angle.

***RC* lead circuit** A phase shift circuit in which the output voltage, taken across the resistor, leads the input voltage by a specified angle.

***RC* time constant** A fixed time interval set by the values of *R* and *C* that determines the time response of a series *RC* circuit. It equals the product of the resistance and the capacitance.

reactive power (P_r) The rate at which energy is stored and alternately returned to the source by a capacitor or inductor. The unit is the VAR.

reference ground The metal chassis that houses the assembly or a large conducive area on a printed circuit board used as the common or reference point.

reflected load The load as it appears to the source in the primary of a transformer.

reflected resistance The resistance in the secondary circuit reflected into the primary circuit.

regulated power supply A type of power supply in which the output is automatically adjusted to keep a constant value.

relay An electromagnetically controlled mechanical device in which electrical contacts are open or closed by a magnetizing current.

reluctance The opposition to the establishment of a magnetic field in a material.

resistance Opposition to current. The unit is the ohm (Ω).

resistor An electrical component designed specifically to have a certain amount of resistance.

resolution The smallest increment of a quantity that a meter can measure.

resonant frequency The frequency at which a resonant condition occurs in a series or parallel *RLC* circuit.

retentivity The ability of a material, once magnetized, to maintain a magnetized state without the presence of a magnetizing force.

rheostat A two-terminal variable resistor.

right triangle A triangle with a 90° angle.

ripple voltage The small variation in the dc voltage on the output of a filtered rectifier caused by the slight charging and discharging action of the filter capacitor.

rise time (t_r) The time interval required for a pulse to change from 10% to 90% of its amplitude.

rising edge The positive-going transition of a pulse.

***RL* lag circuit** A phase shift circuit in which the output voltage, taken across the resistor, lags the input voltage by a specified angle.

***RL* lead circuit** A phase shift circuit in which the output voltage, taken across the inductor, leads the input voltage by a specified angle.

***RL* time constant** A fixed time interval set by the values of *R* and *L* that determines the time response of a circuit and is equal to L/R.

rms value The value of a sine wave that indicates its heating effect, also known as the *effective value.* It is equal to 0.707 times the peak value. RMS stands for root mean square.

roll-off The decrease in the response of a filter below or above a critical frequency.

rotor The moving part of a generator, alternator, or motor.

round off The process of dropping one or more digits to the right of the last significant digit in a number.

sawtooth waveform A type of electrical waveform composed of ramps; a special case of a triangular waveform in which one ramp is much shorter than the other.

schematic A symbolized diagram of an electrical or electronic circuit.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

scientific notation A system for representing any number as a number between 1 and 10 times an appropriate power of ten.

secondary winding The output winding of a transformer; also called *secondary*.

selectivity A measure of how effectively a resonant circuit passes certain frequencies and rejects others. The narrower the bandwidth, the greater the selectivity.

self-excited generator A generator in which the field windings derive their current from the output.

semiconductor A material that has a conductance value between that of a conductor and that of an insulator. Silicon and germanium are examples.

sensitivity The ohms-per-volt rating of a voltmeter.

series In an electric circuit, a relationship of components in which the components are connected such that they provide a single current path between two points.

series-aiding An arrangement of two or more series voltage sources with polarities in the same direction.

series-opposing An arrangement of two series voltage sources with polarities in the opposite direction; not a normal configuration.

series resonance In a series *RLC* circuit, the condition where the impedance is minimum and the reactances are equal.

shell An energy band in which electrons orbit the nucleus of an atom.

short A circuit condition in which there is a zero or abnormally low resistance between two points; usually an inadvertent condition.

SI Standardized international system of units used for all engineering and scientific work; abbreviation for French *Le Système International d'Unites.*

siemens The unit of conductance.

significant digit A digit known to be correct in a number.

sine wave A type of electrical waveform that is based on the mathematical sine function.

slip The difference between the synchronous speed of the stator field and the rotor speed in an induction motor.

solenoid An electromagnetically controlled device in which the mechanical movement of a shaft or plunger is activated by a magnetizing current.

solenoid valve An electrically controlled valve for control of air, water, steam, oils, refrigerants, and other fluids.

source Any device that produces energy.

speaker An electromagnetic device that converts electrical signals to sound waves.

squirrel cage An aluminum frame within the rotor of an induction motor that forms the electrical conductors for a rotating current.

stator The fixed part of a generator, alternator, or motor.

steady state The equilibrium condition of a circuit that occurs after an initial transient time.

step-down transformer A transformer in which the secondary voltage is less than the primary voltage.

step-up transformer A transformer in which the secondary voltage is greater than the primary voltage.

stiff voltage divider A voltage divider that has a load resistor that is large compared to the resistors in the divider, so that the loaded and unloaded output voltages are nearly the same.

stopband The range of frequencies between the upper and lower cutoff points.

strain gauge A variable resistor formed by a resistive material in which a lengthening or shortening due to stress produces a proportional change in resistance.

superposition A method for analyzing circuits with two or more sources by examining the effects of each source by itself and then combining the effects.

switch An electrical or electronic device for opening and closing a current path.

synchronous motor An ac motor in which the rotor moves at the same rate as the rotating magnetic field of the stator.

system A group of interrelated parts that perform a specific function.

tank circuit A parallel resonant circuit.

tapered Nonlinear, such as a tapered potentiometer.

temperature coefficient A constant specifying the amount of change in the value of a quantity for a given change in temperature.

terminal An external contact point on an electric circuit or an electronic device.

terminal equivalency A condition that occurs when two circuits produce the same load voltage and load current where the same value of load resistance is connected to either circuit.

tesla The SI unit of flux density.

thermistor A type of variable resistor in which resistance depends on temperature.

thermocouple A type of temperature transducer formed by the junction of two dissimilar metals that produces a voltage proportional to temperature.

Thevenin's theorem A circuit theorem that provides for reducing any two-terminal resistive circuit to a single equivalent voltage source in series with an equivalent resistance.

time constant A fixed-time interval set by R and C, or R and L values, that determines the time response of a circuit.

tolerance The limits of variation in the value of a component.

trailing edge The second step or transition of a pulse.

transducer A device that transforms energy from one form to another.

transfer curve A plot showing the ratio of the output to the input.

transformer An electrical device constructed of two or more coils (windings) that are electromagnetically coupled to each other to provide a transfer of power from one coil to another.

transient A temporary passing condition in a circuit; a sudden or temporary change in circuit conditions.

transient time An interval equal to approximately five time constants.

triangular waveform A type of electrical waveform that consists of two ramps.

triangular-wave oscillator An electronic oscillator that uses a comparator with two trip points and an integrator to produce triangular waves.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

trigger The activating mechanism of some electronic devices or instruments.

trimmer A small variable capacitor.

troubleshooting A systematic process of isolating, identifying, and correcting a fault in a circuit or system; the application of logical thinking combined with a thorough knowledge of circuit or system operation to find and correct a malfunction.

true power The power that is dissipated in a circuit, usually in the form of heat.

turns ratio The ratio of turns in the secondary winding to turns in the primary winding.

unbalanced bridge A bridge circuit that is in the unbalanced state as indicated by a voltage across the bridge proportional to the amount of deviation from the balanced state.

valence Related to the outer shell or orbit of an atom.

valence electron An electron that is present in the outermost shell of an atom.

VAR (volt-ampere reactive) The unit of reactive power.

varactor A diode that is used as a voltage-variable capacitor.

vertical organization A centralized business structure that has an organized top-down approach. The tendency is less toward a specialization.

volt The unit of voltage or electromotive force.

voltage The amount of energy available to move a certain number of electrons from one point to another in an electric circuit.

voltage divider A circuit consisting of series resistors across which one or more output voltages are taken.

voltage drop The decrease in the voltage across a resistor due to loss of energy.

voltage-follower A closed-loop, noninverting op-amp with a voltage gain of 1.

6238262 2013/12/10 12.27.66.5

voltage gain The ratio of output voltage to input voltage.

voltage regulation The process of maintaining an essentially constant output voltage over variations in input voltage or load.

voltage source A device that produces a constant voltage for a varying load.

voltmeter An instrument used to measure voltage.

watt (W) The unit of power.

Watt's law A law that states the relationships of power to current, voltage, and resistance.

waveform The pattern of variations of a voltage or current showing how the quantity changes with time.

weber (Wb) The SI unit of magnetic flux, which represents 10^8 lines.

Wheatstone bridge A 4-legged type of bridge circuit with which an unknown resistance can be accurately measured using the balanced state of the bridge. Deviations in resistance can be measured using the unbalanced state.

winding The loops or turns of wire in an inductor.

winding resistance The resistance of the length of wire that makes up a coil.

wiper The sliding contact in a potentiometer.

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

This page intentionally left blank

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

INDEX

A

AC (alternating current) circuits
  analysis of, 338–342
  capacitors in, 410–417
  inductors in, 508–514
  role of, 8
  voltage, ac and dc combined, 340–341
AC coupling, 419–420
Accuracy, 21, 70
AC generators. *See* Alternators (AC generators)
AC motors
  classification, 347
  induction, 347, 348–349
  rotating stator field, 347–348
  synchronous, 347, 349
Active components, 7–8
Admittance
  in parallel *RC* circuits, 450–451
  in parallel *RL* circuits, 538–540
Ah (ampere-hour) ratings of batteries, 105
Air-core transformers, 623
Alarm systems, 292–293
Alkaline-MnO_2 batteries, 43
Alternators (AC generators), 8, 321
  applications, 345–346
  defined, 342
  frequency, 343
  practical, 343–344
  rotating-armature, 344
  rotating-field, 344
  rotor current, 344–345
  simplified, 342
Ammeter, 66, 67
Ammeter shunts, 196
Ampere (amp), 47
Ampere-hour (Ah) ratings of batteries, 105
Ampere-turn (At), 287
Amplifiers
  buffer, 447–448
  transistor, 420
  tuned, 607–608
  voltage-divider bias and discrete, 232–233
Amplitude, 327
  range, 365
AM radio receivers, 609–610
Analog multimeters, 66, 70–71
Analog oscilloscopes, 358
Analog signals, 9
Angular measurement, 330–334
Anode, 41
ANSI (American National Standards Institute), 59
Antenna coupling, 608–609
APM (analysis, planning, and measuring) troubleshooting
  differentiators and integrators, 688–690
  parallel circuits and, 199–203
  RC circuits and, 470–474
  RL circuits and, 553–555
  series circuits and, 152–156
  series-parallel circuits and, 257–260
  steps, 106–108
  transformers, 646–647
Apparent power, 461, 462–464, 546–547
  rating, 637
Arbitrary function generators (AFGs), 364
Arbitrary waveform generators, 326
Arctangent, 437
Armature, 307
Armature relays, 291–292
Atomic number, 35
Atoms
  basic structure of, 35
  copper, 36
  defined, 35
  electrons, 35, 36–37
Automotive charging systems, 346
Autotransformers, 644
Average value, 329–330
AWG (American Wire Gauge), 63

B

Back emf, 309
Balanced Wheatstone bridge, 237–238
Balun transformers, 636–637
Band-pass filters, 580, 582–583, 601–605
Band-stop filters, 585–587, 605–607
Bandwidth, 468, 551, 580–581, 601
Batteries
  ampere-hour ratings of, 105
  backup, 42
  car, removing, 149
  defined, 41
  how they work, 41–43
  lead-acid, 42, 43
  replacing and possible explosions, 134
  types of, 43
Bias
  voltage-divider, 232–233
  voltages, 421
Bleeder current, 231–232
Block diagrams, 5–6
Bode plotter, 576
Bohr model, 35
Boundary, 4
Braking resistors, 98
Branches, 174, 182
  currents, 224–225
  open, 199–202
  shorted, 202
Breakdown voltage, 386
Brushes, 304
Brushless dc motor, 309
Buffer amplifiers, 447–448
Bypass capacitors, 421

C

Calibrating equipment, 21
Capacitance
  converting between units, 385–386
  defined, 382, 383–384
  formula for, 389
  meter/measurement, 394
  stray, 453
  units of, 384
  winding, 491–492, 639
Capacitive reactance
  description of, 410–411
  Ohm's law, 413
  parallel capacitors and, 412–413
  series capacitors and, 412
  symbol and unit for, 410
Capacitive sensors, 576
Capacitive susceptance, 450
Capacitive voltage divider, 414
Capacitors
  ac circuits and, 410–417
  applications, 417–422
  bypass, 421
  ceramic, 390–391
  charging, 402
  charging and discharging curves, 405–408
  charging and discharging in *RC* integrators, 658
  chip, 391
  color coding and marking, 701–704
  construction of, 383
  coupling, 420
  current and voltage during charging and discharging, 402–403
  current leads voltage by 90°, 414–415
  dc circuits and, 401–409
  decoupling, 421
  defined, 382, 383
  dielectric constant, 388–389
  dielectric strength, 386–387
  discharging, 402
  electrolytic, 391, 392
  fixed, 390–392
  labeling and color coding, 394
  leakage, 387
  mica, 390
  parallel, 399–401, 412–413
  physical characteristics of, 387–388
  plastic-film, 391–392
  plate area and separation, 388
  power in, 415–416
  RC time constant, 404–405
  sensors, 387
  series, 395–399, 412
  square wave, response to, 408–409
  storage of charge, 383, 384, 392, 403
  storage of energy, 386
  temperature coefficient, 387
  trimmers, 393
  variable, 392–394
  voltage in *RC* integrators, 659
  voltage rating, 386
  types of, 390–394
Capacity, 42
Car batteries, removing, 149
Carbon-zinc batteries, 43
Carrier, 9, 365
Cathode, 41
CE mark, 69
Center tap (CT), 641
Ceramic capacitors, 390–391

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Certification, 4
CGS (centimeter-gram-second) system, 281
Charge
  attraction and repulsion, 38
  coulomb, 38
  defined, 38
  positive and negative, 38–39
  storage in capacitors, 383, 384, 392, 403
  symbol for, 38
  unit of, 38
Charging
  capacitors, 402, 658
  resonant, 638
Chip capacitors, 391
Chokes
  longitudinal, 515
  radio frequency, 515–516
  use of term, 488, 489
Circuits
  See also AC circuits; DC circuits; Parallel circuits; *RC* circuits; *RL* circuits; *RLC* circuits; Series circuits; Series-parallel circuits
  components of, 7–8
  defined, 2, 7
  electronic, 8–9
  integrated, 3
  measurements, 66– 71
  NI Multisim for simulation of, 711–715
  schematic, 59
Circuits, electrical
  ac and dc, 8
  breakers, 61
  closed, 59
  difference between magnetic circuits and, 288
  elements of, 58–66
  ground, 65
  open, 59
  power in, calculating, 96–97
  protective devices, 61
  schematic, 59
  switches, 59–60
  wires, 63–65
Circular mil (CM), 63
Clamp meters, 328–329
Closed circuits, 59
Coaxial cable, 589
Coefficient of coupling, 620–621
Coercive force, 297
Coil, use of term, 488, 489, 619
Color codes, resistor, 52–55
Common (COM or COMM), 65
Commutator, 304
Components, 7–8
Computer memory, 422
Conductance, 50
  defined, 450
  in parallel *RC* circuits, 450
  in parallel *RL* circuits, 538–540
Conductive noise, 515
Conductors, 37
Connectors, 64
Copper
  atom, 36
  in conductors, 37
Coulomb (C), 38, 384
Coulomb's law, 38
Counter emf, 309
Coupling, coefficient of, 620–621
Coupling ac signal into dc bias circuit, 469–470
Coupling capacitors, 420
Current
  ampere, 47
  amplitudes in series *RLC* circuits, 574–577
  bleeder, 231–232
  branch, 224–225
  calculating, 87–89
  carrying conductor in magnetic field, 302–303
  defined, 46
  dividers, 193–197
  full-scale deflection, 196
  induced, 300
  in inductors, 500–502
  Kirchhoff's current law, 188–192, 338
  lags voltage by 90°, 512
  leads voltage by 90°, 414–415
  linear relationship of voltage and, 85–86
  load, 231–232
  measuring, 66, 67, 108
  measuring, and finding open branch, 200–202
  movement of electrons, 46
  relationship between voltage, resistance and, 83–85
  rotor, 344–345
  -sensing resistors, 91–92
  in series circuits, 128–129
  in series-parallel circuits, 224–225
  sources of, 47–49
  symbol and formula for, 46
  transformers, 631
  unit of, 47
  values of sine waves, 326–330
Curves, charging and discharging, 405–408
Cutoff frequency, 468, 551–552, 581, 601
Cycle, sine wave, 321

D
d'Arsonval meter movement, 293–294
DC blocking, 419–420
DC (direct current) circuits
  capacitors in, 401–409
  inductors in, 499–507
  role of, 8
  voltage, ac and dc combined, 340–341
DC generators
  equivalent circuit for, 307
  how they work, 303–307
  parts of, 304, 306
  self-excited, 306
  shunt-wound, 307
DC isolation, 628–629
DC motors
  back or counter emf, 309
  brushless, 309
  how they work, 308
  loads, 309
  maintenance, 308
  ratings, 309–310
  reversing, 310–311
  series, 311
  shunt, 127–128, 312
  torque and power, 309–310
DC resistance, 491
DC voltage
  ac and dc combined, 340–341
  sources of, 40–45
Decibel (dB) measurement, 582
Decoupling capacitors, 421
Degree, 330, 331
Delta connections, 645–646
Dielectric, 383, 388
Dielectric constant, 388–389
Dielectric strength, 386–387
Differentiator
  See also RC differentiators; *RL* differentiators
  use of term, 657, 669
Digital multimeters (DMMs), 66, 68–70
Digital oscilloscopes, 359
Digital-pattern generators (DPGs), 364
Digital signals, 9
Discharging capacitors, 402
Double-pole—double-throw (DPDT) switch, 60
Double-pole—single-throw (DPST) switch, 60
Dummy loads, 184
Duty cycle, 351–352

E
Earth ground, 65
Effective value, 327–328
Efficiency
  defined, 104
  power supply, 104–105
  transformer, 639
Electrical charge, 37–39
Electrical circuits. *See* Circuits, electrical
Electrical isolation, 620
Electrical shock, 24–25, 58, 67
Electrical systems, 5
Electrical units, 16–17
Electric field, 38
Electrolytic capacitors, 391, 392
Electromagnet, 289
Electromagnetic devices
  magnetic disks/tapes read/write head, 294–295
  magneto-optical disks, 295–296
  meter movement, 293–294
  relays, 291–293
  solenoid, 290–291
  speakers, 293
Electromagnetic field, 285
Electromagnetic induction, 44, 298–303
Electromagnetic interference (EMI) filters, 552–553
Electromagnetism
  circuit, basic, 288
  defined, 284
  left-hand rule, 285
  lines of force, 285
  magnetomotive force, 287–289
  permeability, 286
  properties, 285–289
  reluctance, 286–287
Electronic circuits, 8–9
Electronic signal generators, 321, 325–326
Electronics industry
  certification, 4
  changes to, 2
  major divisions, 2–3
Electronic systems
  defined, 5
  examples of, 4
Electrons, 35, 36–37
Elements, 35
Energy
  conversion, 101–102
  defined, 93
  levels, 36
  storage of, in capacitors, 386
  storage of, in inductors, 490
  unit of, 93–95

6238262 2013/12/10 12.27.66.5

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Engineering notation, 14–15
Error, 21
Exponent, 10
Exponential curves, 405, 407, 505–507
Exponential formulas, 505

F

Fall time, 351
Faraday, M., 278, 300
Faraday's law, 300–301, 492
Farads (F), 384, 385
FCC Commercial Operator License Examinations, 4
Feedback, 464
Ferrite-core transformers, 623
Ferromagnetic materials, 281–282, 286, 306, 490
Filters, 418–419
  band-pass, 580, 582–583, 601–605
  band-stop, 585–587, 605–607
  electromagnetic interference, 552–553
  parallel resonant, 601–607
  RC circuits as, 465–469
  RL circuits as, 550–553
  series resonant, 579–587
  signal, 421
Fixed capacitors, 390–392
Fixed inductors, 495
Flash converters, 142
Flashlights, replacing batteries in, 134
Flowcharts, 5–6
Flux, magnetic, 279
Flux density, magnetic, 279–281
Free electrons, 36
Frequency
  See also Resonant frequency
  alternator, 343
  cutoff, 468, 551–552, 581, 601
  domain plot, 356
  fundamental, 355
  half-power, 582
  pulse repetition, 351
  ratings, 638
  sine wave, 324–325
Frequency response
  RC circuits and, 467, 442–444
  RL circuits and, 531–533, 550
Fuel cells, 43–44
Full-scale deflection current, 196
Function generators, 326
  arbitrary, 364
  description of basic, 366–367
  role of, 364
Fundamental frequency, 355
Fuses, 61

G

Gain, 6
Galvanometers, 237
Gauss (G), 281
Generators, 44–45
  See also Alternators (AC) generators; DC generators; Function generators
  arbitrary waveform, 326
  electronic signal, 321, 325–326
Germanium, 37
Glass, 37
Graphene, 392
Ground
  fault, 189
  loop, 515
  measuring voltages with respect to, 149–152
  plane, 149
  positive, 149
  reference/common, 65, 149
Ground fault circuit interrupters (GFCI), 25, 26
Gyrators, 509

H

Half-power frequencies, 582
Half-splitting, 108, 153–154
Hall effect, 284
Harmonic distortion, 365
Harmonics, 321, 355–356
Hartley oscillators, 497
Henry (H), 489–490
Hertz (Hz), 324
High-pass circuit/filter, 466, 467–468, 550–551
High temperature superconductors (HTS), 347
Horizontal organization, 2
Hyperbola, 568
Hysteresis
  loss, 639
  magnetic, 296–298

I

IEEE STD 315–1975 standard, 59
Impedance
  cases of, 436
  matching, 634–637
  in parallel *RC* circuits, 449–451
  in parallel *RLC* circuits, 587–588
  in parallel *RL* circuits, 537–540
  in series *RC* circuits, 436–438
  in series *RLC* circuits, 566–567, 577–578
  in series *RL* circuits, 527–528
  triangle, 436, 527
iNARTE (interNational Association for Radio, Telecommunications, and Electromagnetics), 4
Indicator lights, 132
Induced current, 300
Induced voltage, 299–300, 489
Inductance
  defined, 489
  mutual, 620–621
  symbol for, 489
  unit of, 489–490
Induction heating system, 640–641
Induction motors, 347, 348–349
Inductive reactance
  description of, 508–509
  Ohm's law, 511
  for parallel inductors, 510–511
  for series inductors, 510
  symbol and unit for, 508
Inductive sensors, 576
Inductive susceptance, 539
Inductors
  in ac circuits, 508–514
  applications, 514–516
  current in, 500–502
  current lags voltage by 90°, 512
  in dc circuits, 499–507
  defined, 489
  energy storage, 490
  Faraday's law, 300–301, 492
  inductance, 489–490
  Lenz's law, 301, 492–493
  parallel, 498, 510–511
  physical characteristics of, 490–491
  power in, 512–513
  quality factor, 514
  RL time constant, 499
  series, 496–497, 510
  square wave, response to, 502–503
  symbols, 495
  types of, 495–496
  voltage in series *RL* circuits, 503–505
  winding capacitance, 491–492
  winding resistance, 491
Input, 4
Instantaneous power, 415–416, 513
Instantaneous value, 326–327
Insulators, 37
Integrated circuits, 3
  thermal sensors, 102
Integrator
  See also RC integrators; *RL* integrators
  use of term, 657, 658
Intermediate frequency (IF) transformers, 623, 624
Inverse tangent, 437
Ions, 36, 38–39
Iron-core transformers, 622–623

J

Joule, defined, 93

K

Kilo, 87
Kilohms, 87
Kilowatt-hour (kWh), unit of energy, 93–95
Kirchhoff's current law, 188–192, 338
  parallel *RC* circuits and, 453
  parallel *RL* circuits and, 541
Kirchhoff's voltage law, 137–140, 338
  series *RC* circuits and, 440
  series *RL* circuits and, 530

L

Label codes, resistor, 55–56
Ladder diagrams, 181
Ladder logic, 181
Lagging power factor, 548
Lead-acid batteries, 42, 43
Leakage, capacitor, 387
LEDs, 132–133
Left-hand rule, 285
Lenz's law, 301, 492–493
Linear relationship of current and voltage, 85–86
Lines of force, 278–279, 285
Liquid level-sensing system, 241
Lithium-ion batteries, 43
Lithium-MnO$_2$ batteries, 43
Load(s)
  balancing, 549
  bank, 184
  cell, 240–241
  current, 231–232
  dc motors, 309
  defined, 58
  dummy, 184
  reflected, 632–634
  resistance, 599–600
  resistive load across the secondary, 630–632
  test box, 99–100
  voltage dividers with resistive, 229–233
Loading
  defined, 21, 229
  effect of voltmeters, 234–236

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Lockout/tagout, 25
Longitudinal choke, 515
Low-pass circuit/filter, 466–467, 550

M

Magnetic circuits, difference between electric circuits and, 288
Magnetic coupling, 623
Magnetic disks/tapes read/write head, 294–295
Magnetic field intensity, 296–297
Magnetic fields, 278–284
Magnetic flux, 279
  density, 279–281
  leakage, 639
Magnetic hysteresis, 296–298
Magnetic switches, 283
Magnetism
  applications, 283–284
  attraction and repulsion, 278
  flux density of various fields, 282
  gauss (G), 281
  Hall effect, 284
  how materials become magnetized, 281–283
  lines of force, 278–279
  sensitive sensors, 283
Magnetizing force, 296
Magnetomotive force (mmf), 287–289
Magneto-optical disks, 295–296
Magnets
  electro, 289
  permanent, 283–284
  safety when working with, 282
Magnitude, 327
Maximum power transfer theorem, 249–251
Measuring data/numbers
  error, accuracy, and precision, 20–21
  rounding off, 22–23
  significant digits, 20, 21–22
Mechatronics, 3
Mega, 87
Megohms, 87
Metal detectors, 600
Meter movement, 196, 293–294
Meters
  capacitance, 394
  clamp, 328–329
  symbols, 67
Metric conversions, 18–20
Metric prefixes, 17–18
mho, 50
Mica capacitors, 390
Microfarads, 384, 385
Millman's theorem, 249, 707–709
Modulating signal, 365
Modulation, 365
Multimeter, 66
Multiple-winding transformers, 642–644
Mutual inductance, 620–621

N

National Electric Code (NEC), 65
National Institute of Standards and Technology (NIST), 21, 283
Negative ions, 36, 38–39
Neutrons, 35
Nickel-metal hydride batteries, 43
NI Multisim for circuit simulation, 711–715
Node, 188
Noise suppression, 514–515
Nonsinusoidal waveforms
  harmonics, 355–356
  pulse, 350–353
  triangular and sawtooth, 353–355
North American Board of Certified Energy Practitioners (NABCEP), 4
Norton's theorem, 242, 705–707
Nucleus, 35

O

Ohm, 50
Ohm, G. S., 82
Ohmmeter, 66, 68
Ohm's law
  ac circuits and, 338
  application of, 87–91
  capacitive reactance, 413
  description of, 82, 83–87
  for electromagnetic circuits, 288
  graphic aid for, 86
  inductor reactance, 511
  parallel circuits and, 185–188
  parallel *RC* circuits and, 452
  parallel *RL* circuits and, 540–541
  series circuits and, 130–133
  series *RC* circuits and, 438–440
  series *RL* circuits and, 529
Open branches, in parallel circuits, 199–202
Open circuits, 59
Open series, 152–154
Orbits, electron, 36
Oscillators, 8, 326
  feedback in Colpitts, 398–399
  Hartley, 497
  phase shift, 464–465
  RC circuits in, 464–465
  swept-frequency, 364
Oscilloscopes
  analog, operation of, 358
  controls, 359–360
  coupling signals into, 360–363
  defined, 357
  digital, operation of, 359
  probes, passive, 456–457
  probing, high-frequency differential, 458
  types of, 358
Output, 4
  impedance, 365
Oxidation-reduction reaction, 41

P

Parallax, 70
Parallel capacitors, 399–401, 412–413
Parallel circuits
  See also *RC* circuits, parallel; *RLC* circuits, parallel; *RL* circuits, parallel; Series-parallel circuits
  applications, 180–181
  current dividers, 193–197
  defined, 174
  Kirchhoff's current law, 188–192
  Ohm's law and, 185–188
  open branches, 199–202
  power in, 197–199
  resistors in, 174–176, 180
  shorted branches, 202
  total resistance in, 177–182
  troubleshooting, 199–203
  voltage in, 182–185
Parallel-connected cells, 42, 43
Parallel inductors, 498, 510–511
Parallel resonance. *See* Resonance, parallel
Parallel resonant filters, 601–607
Passive components, 7–8
Peak-to-peak value, 327
Peak value, 327
Period, sine wave, 322–325
Periodic waveform, 351
Permeability, 286, 490
Permittivity, 388–389
Phase, sine wave, 332–333
Phase angle
  in parallel *RC* circuits, 449–451
  in parallel *RLC* circuits, 587–588
  in parallel *RL* circuits, 537–540
  in series *RC* circuits, 436–438
  in series *RLC* circuits, 566–567, 578–579
  in series *RL* circuits, 527–528
Phase converters, 334
Phase noise, 365
Phase shift
  in motors, 548
  oscillators, 464–465
Phasor quantity, 335–336
Photoconductive cells, 57
Photovoltaic effect, 44
Picrofarads, 384, 385
Piezoelectric effect, 45
Piezoelectric sensors, 45
π (pi) function, 331
Planar transformers, 623–624
Plastic-film capacitors, 391–392
Plate area and separation, capacitance, 388
Plunger, 290
Polarity
  of induced voltage, 299–300
  of sine wave, 321–322
Polyethylene, 37
Polyphase power, 333–334
Porcelain, 37
Positive ground, 149
Positive ions, 36, 38–39
Potentiometers, 56, 57
  as adjustable voltage dividers, 144–145
Power
  apparent, 461, 462–464, 546–547
  apparent power rating, 637
  in capacitors, 415–416
  in a circuit, calculating, 96–97
  in dc motors, 309–310
  defined, 93
  factor, 461–462, 548–549
  in inductors, 512–513
  instantaneous, 415–416, 513
  loss, 104
  in parallel circuits, 197–199
  polyphase, 333–334
  rating in transformers, 637–638
  rating in watts, 94
  rating of resistors, 98–101
  in *RC* circuits, 460–464
  reactive, 416, 460, 513, 546–547
  in *RL* circuits, 546–549
  in series circuits, 147–148
  supplies, 45, 103–105, 494, 628
  triangle, 461
  true, 416, 460, 513, 546
Powers of ten, 10–14

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

6238262 2013/12/10 12.27.66.5

Precision, 21
Primary winding, 622
Programmable logic controllers (PLCs), 181
Protective devices, 61
Protons, 35, 36
Pulse generators, 364
Pulse waveforms, 350–353
  -to-dc converter, 686–687
Pulse width, 350
Push button (PB), 60

Q

Quality factor (*Q*), 514
  resonant circuits and, 583–585

R

Radian (rad), 330–331
Radiated noise, 515
Radio frequency (RF) chokes, 515–516
Radiometric image, 203
Ramps, 353–354
RC circuits
  applications, 464–470
  power in, 460–464
  troubleshooting, 470–474
RC circuits, parallel
  analysis of, 451–454
  analysis of series-parallel, 455–456
  conductance, capacitive susceptance, and admittance, 450–451
  impedance and phase angle in, 449–451
  Kirchhoff's law, 453
  measurements, 457–460
  Ohm's law, 452
  phase relationships of current and voltage, 453–454
RC circuits, series
  analysis of, 438–448
  analysis of series-parallel, 455–456
  frequency and impedance and phase angle variation, 442–444
  impedance and phase angle in, 436–438
  Kirchhoff's law, 440
  measurements, 457–460
  Ohm's law, 438–440
  phase relationships of current and voltage, 440–441
  sinusoidal response of, 435
RC differentiators
  applications, 685–688
  time response to repetitive pulses, 674–675
  time response to single pulse, 669–673
  troubleshooting, 688–690
  use of term, 657, 669
RC integrators
  applications, 685–688
  capacitor charging and discharging, 658
  capacitor voltage, 659
  steady-state time response, 665
  time constant increase, effect of, 666–668
  time response to repetitive pulses, 664–668
  time response to single pulse, 659–663
  troubleshooting, 688–690
  use of term, 657, 658
RC lag circuit, 444–446
RC lead circuit, 446–447
RC time constant, 404–405, 658, 666–668
Reactive power, 416, 513, 546–547
Receptacle
  standard, 25
  testers, 25
Rectifier, 417–418
Reed relays, 292
Reed switches, 283
Reference/common ground, 65, 149
Reflected load, 632–634
Reflected resistance, 633
Regulated power supplies, 103–105
Relative motion, 299
Relative permeability, 286
Relative permittivity, 388–389
Relays, 291–293
Reluctance, 286–287
Repetitive pulses, 351–352
  RC differentiators time response to, 674–675
  RC integrators time response to, 664–668
Repetitive waveform, analysis of, 674–675
Resistance
  calculating, 91
  defined, 50
  measuring, 66, 68, 108
  ohm, 50
  reflected, 633
  relationship between current, voltage and, 83–85
  sensors, variable, 57
  symbol for, 50
  Thevenin's theorem, 242–249
  total resistance in parallel circuits, 177–182
  total resistance in series circuits, 124–128
  total resistance in series-parallel circuits, 223–224
  unit of, 50
  Wheatstone bridge, 236–242
  winding, 491
  wire, 63, 64–65
Resistive load across the secondary, 630–632
Resistivity, 64
Resistors
  braking, 98
  color codes, 52–55
  current-sensing, 91–92
  defined, 50
  equal-value, 127, 180
  film, 51, 52
  fixed, 50–51
  label codes, 55–56
  in parallel, 174–176, 180
  power rating of, 98–101
  sense, 596
  in series, 122–124
  surge, 404
  TO style, 193
  values, table of, 699
  variable, 56–57
  wirewound, 51, 52
Resolution, 69–70
Resonance
  importance of, 565
  series, 568, 572–579
Resonance, parallel
  conditions for ideal, 593
  conditions for nonideal, 595–599
  current and phase angle, 598
  frequency, 593–595
  load resistance, 599–600
  variation of impedance with frequency, 597
Resonant circuits
  applications, 607–610
  charging, 638
Resonant filters
  parallel, 601–607
  series, 579–587
Resonant frequency
  defined, 572, 593
  parallel, 593–595
  series, 573–574
Retentivity, 297
Retriggerable multivibrator, 676
Rheostats, 56, 57
Ripple voltage, 419
Rise time, 350
RLC circuits, parallel
  current relationships, 589–590
  impedance and phase angle in, 587–588
  parallel resonant filters, 601–607
  resonant frequency, 593–595
RLC circuits, series
  analysis of, 568–571
  impedance and phase angle in, 566–567, 577–579
  resonant frequency, 573–574
  series resonant filters, 579–587
  voltage and current amplitudes in, 574–577
RL circuits
  filters, 550–553
  power in, 546–549
  sinusoidal response of, 526
  troubleshooting, 553–555
RL circuits, parallel
  analysis of, 540–542
  analysis of series-parallel, 543–546
  conductance, inductive susceptance, and admittance, 538–540
  impedance and phase angle in, 537–540
  Kirchhoff's law, 541
  Ohm's law, 540–541
  phase relationships of current and voltage, 541–542
RL circuits, series
  analysis of, 529–537
  analysis of series-parallel, 543–546
  frequency and impedance and phase angle variation, 531–533
  impedance and phase angle in, 527–528
  Kirchhoff's law, 530
  Ohm's law, 529
  phase relationships of current and voltage, 530–531
RL differentiators
  applications, 685–688
  time response to single pulse, 682–685
  troubleshooting, 688–690
  use of term, 657
RL integrators
  applications, 685–688
  time response to single pulse, 677–681
  troubleshooting, 688–690
  use of term, 657
RL lag circuit, 534–535
RL lead circuit, 536–537
RL time constant, 499
rms (root mean square) value, 327–328
Rotary switch, 60
Rotating-armature alternators, 344
Rotating-field alternators, 344

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Rotating stator field, 347–348
Rotor, 306
  current, 344–345
Rounding off, 22–23

S

Safety
  car batteries, removing, 149
  current sources, 48
  electrical shock, 24–25, 58, 67
  electrolytic capacitors and, 391
  flashlights, replacing batteries in, 134
  fuses, replacing, 61
  inductors and, 491
  jewelry/rings, wearing of, 67
  lead-acid batteries, 42
  magnets, working with, 282
  precautions, 25–26
  resistors and, 98
  storage of charge in capacitors and, 403
  utility voltages, 25
Sawtooth waveforms, 353–355
Schematic, 59
Scientific notation, 10–14
Secondary winding, 622
Seebeck effect, 45
Selectivity, 582–583
Self-excited generator, 306
Self-inductance. *See* Inductance
Semiconductors, 37
Sense resistors, 596
Sensors
  capacitive, 387, 576
  inductive, 576
  variable resistance, 57
Series-aiding arrangement, 134
Series capacitors, 395–399, 412
Series circuits
  See also *RC* circuits, series; *RLC* circuits, series; *RL* circuits, series
  current in, 128–129
  defined, 122
  equal-value resistors, 127
  Kirchhoff's voltage law and, 137–140
  Ohm's law and, 130–133
  open in, 152–154
  power in, 147–148
  resistors in, 122–124
  shorts, 154–156
  solar panels, 135–136
  total resistance in, 124–128
  troubleshooting, 152–156
  voltage dividers, 141–147
  voltage measurements, 149–152
  voltage sources in, 134–137
Series-connected cells, 42
Series dc motors, 311
Series inductors, 496–497, 510
Series-opposing, 134
Series-parallel circuits
  analysis of, 222–228
  conversion to parallel, 590–592
  current in, 224–225
  identifying, 218–222
  loading effect of voltmeters, 234–236
  maximum power transfer theorem, 249–251
  superposition theorem, 252–256
  Thevenin's theorem, 242–249
  total resistance in, 223–224
  troubleshooting, 257–260
  voltage dividers with resistive loads, 229–233
  voltage in, 225–228
  Wheatstone bridge, 236–242
Series resonance, 568, 572–579
Series resonant filters, 579–587
Shells
  electron, 36
  valence, 36
Shielded inductors, 495
Shorts
  in branches in parallel circuits, 202
  in series circuits, 154–156
Shunt
  ammeters, 196
  dc motors, 127–128, 312
  wound generator, 307
SI (Système International), units and symbols, 16–17
Siemens, 50
Signal filters, 421
Signal generators, 321, 325–326
  specifications for, 364–365
  types of, 363–364
Significant digits, 20, 21–22
Silicon, 37
Silver, in conductors, 37
Silver oxide batteries, 43
Sine (sinusoidal) waves
  See also Nonsinusoidal waveforms
  angular measurement, 330–334
  cycle, 322
  defined, 321
  electronic signal generators, 325–326
  expressions for phase-shifted, 337
  formula, 334–338
  frequency, 324–325
  period, 322–325
  phase, 332–333
  polarity of, 321–322
  polyphase power, 333–334
  radians and degrees, 330–331
  shape of, 321
  sources, 321
  voltage and current values, 326–330
Single-pole—double-throw (SPDT) switch, 59
Single-pole—single-throw (SPST) switch, 59
Single-pulse input
  RC differentiators and, 669–673
  RC integrators and, 659–663
  RL differentiators and, 682–685
  RL integrators and, 677–681
Sinusoidal response
  of *RL* circuits, 526
  of series *RC* circuits, 435
Sinusoidal voltage, symbol for, 321
Sinusoidal waveform. *See* Sine (sinusoidal) waves
Slip, 349
Solar cells, 44
Solar panels in series, 135–136
Soldering, damage from, 60
Solenoid, 290–291
Solenoid valves, 291
Speakers, 293
Spectral purity, 365
Square waves, 352
  capacitors and response to, 408–409
  inductors and response to, 502–503
Squirrel cage, 348–349
Static, 38
Stator, 306
  rotating stator field, 347–348
Steady-state time response, 665
Step-down transformers, 627
Step-up transformers, 626–627
Stiff voltage divider, 229
Stopband, 585
Strain gages, 57
  vibrating wire, 299
  Wheatstone bridge, 239–241
Superposition theorem, 252–256
Surge resistor, 404
Swept-frequency oscillators, 364
Swept-frequency test, 668
Switches
  magnetic, 283
  reed, 283
  symbols, 60
  types of, 59–60
Switch-mode power supply (SMPS), 494
Symbols
  ampere, 47
  capacitive reactance, 410–411
  CE mark, 69
  charge, 38
  circuit breakers, 61
  conductance, 50
  core, 622
  electrical quantities and SI, 16
  force, 313
  fuses, 61
  ground, 65
  inductance, 489
  inductive reactance, 508
  inductor, 495
  magnetic field intensity, 296–297, 313
  magnetic flux, 279, 313
  magnetic flux density, 279, 313
  magnetomotive force, 287–289, 313
  meter, 67
  metric prefixes and, 17
  ohm, 50
  permeability, 286, 313
  potentiometers and rheostats, 56
  relative permittivity, 388
  reluctance, 286, 313
  resistance/resistor, 50
  resistance devices, 57
  siemens, 50
  sinusoidal voltage, 321
  switches, 60
  torque, 313
  UL mark, 69
  volt, 40
Synchronous motors, 347, 349
System(s)
  block diagrams, 5–6
  defined, 4
  electrical, 5
  electronic, 5
  flowcharts, 5–6
  manufacturers, 2
  transfer/response curve, 6–7

T

Tank circuit, 595
Tapered potentiometers, 57

Username: Karl Smirni**Book:** DC/AC Fundamentals: A Systems Approach. No part of any book may be reproduced or transmitted in any form by any means without the publisher's prior written permission. Use (other than pursuant to the qualified fair use privilege) in violation of the law or these Terms of Service is prohibited. Violators will be prosecuted to the full extent of the law.

Tapped transformers, 641–642
Teflon, 37
Temperature
  coefficient, 387
  unbalanced Wheatstone bridge for measuring, 238–239
Terminal equivalency, 243
Tesla (T), 279, 281
Thermal imaging, 202
Thermistors, 57
Thermocouples, 45
Thermography, 202–203
Thevenin's theorem, 242–249
Three-phase transformers, 644–646
Time response. *See RC* differentiators; *RC* integrators; *RL* differentiators; *RL* integrators
Timing circuits, 422, 685–686
Tolerance, 51
Torque, 309–310
Transducers, 9, 236
Transfer/response curve, 6–7
Transformers
  air-core, 623
  auto-, 644
  balun, 636–637
  basic, 622–626
  characteristics, 638–640
  current, 631
  dc isolation, 628–629
  defined, 622
  efficiency, 639
  ferrite-core, 623
  impedance matching, 634–637
  intermediate frequency, 623, 624
  iron-core, 622–623
  multiple-winding, 642–644
  mutual inductance, 620–621
  planar, 623–624
  power supplies, 628
  ratings, 637–638
  reflected load, 632–634
  resistive load across the secondary, 630–632
  step-down, 627
  step-up, 626–627
  tapped, 641–642
  three-phase, 644–646
  troubleshooting, 646–647
  turns ratio, 624–625
  windings, direction of, 625
Transient time, 405, 659–660
Transistor amplifier, 420
Triangular waveforms, 353–355
Trigger pulse generator and waveshaping, 687–688
Trimmers, 393
Troubleshooting
  defined, 106
  differentiators and integrators, 688–690
  parallel circuits, 199–203
  RC circuits, 470–474
  RL circuits, 553–555
  series circuits, 152–156
  series-parallel circuits, 257–260
  steps, 106–108
  transformers, 646–647
True power, 416, 513, 546
Tuned amplifiers, 607–608
Tuned circuits, 516
Turns ratio, 624–625

U

UL (Underwriters Laboratories) mark, 69
Unbalanced Wheatstone bridge, 238–241
Units
  electrical, 16–17
  metric conversions, 18–20
Utility voltages, 25

V

Valence electrons, 36
Valence shells, 36
  Values of sine waves, voltage and current, 326–330
VAR (volt-ampere reactive), 416
Varactor, 394
Variable capacitors, 392–394
Variable inductors, 495
Vertical organization, 2
Volt, 40
Voltage
  ac and dc combined, 340–341
  amplitudes in series *RLC* circuits, 574–577
  bias, 421
  breakdown, 386
  calculating, 89–90
  capacitor, 659
  defined/formula, 40
  dividers, 141–147, 229–233, 414
  drop, 102, 137–140
  Hall, 284
  induced, 299–300, 489
  Kirchhoff's voltage law, 137–140, 338
  linear relationship of current and, 85–86
  measuring, 66, 67–68, 108
  measuring in series circuits, 149–152
  in parallel circuits, 182–185
  ramps, 353–354
  rating, 386, 638
  relationship between current, resistance and, 83–85
  ripple, 419
  in series-parallel circuits, 225–228
  sources in series circuits, 134–137
  sources of, 40–45
  unit of, 40
  values of sine waves, 326–330
  volt, 40
  working, 386
Volt-ampere reactive (VAR), 416
Voltmeters, 66, 68
  loading effect of, 234–236

W

Watchdog timer, 676
Watts
  defined, 93
  power rating in, 94
Watt's law, 96
Waveforms
See also type of
  analysis of repetitive, 674–675
  defined, 321
  modes, 365–366
  nonsinusoidal, 350–357
  sine or sinusoidal, 321–338
Wave traps, 606
Weber (Wb), 279
Wheatstone bridge, 236
  balanced, 237–238
  configuration, 236–237
  Thevenin's theorem and, 247–248
  unbalanced, 238–241
Winding(s)
  capacitance, 491–492, 639
  direction of, 625
  primary, 622
  resistance, 491, 638
  secondary, 622
  use of term, 489, 619
Wiper, 56
Wires
  gages and sizes, 63
  resistance, 63, 64–65
  splicing, 65
Working voltage, 386
Wye connections, 645–646

Z

Zinc air batteries, 43

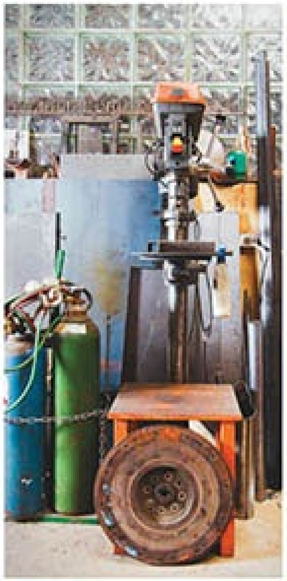

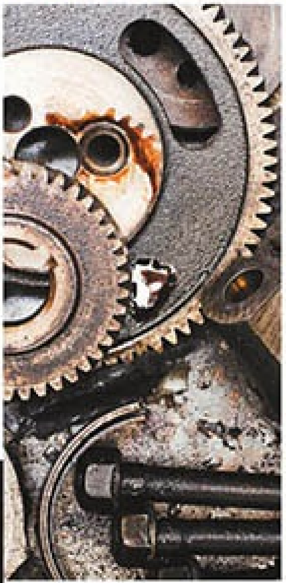

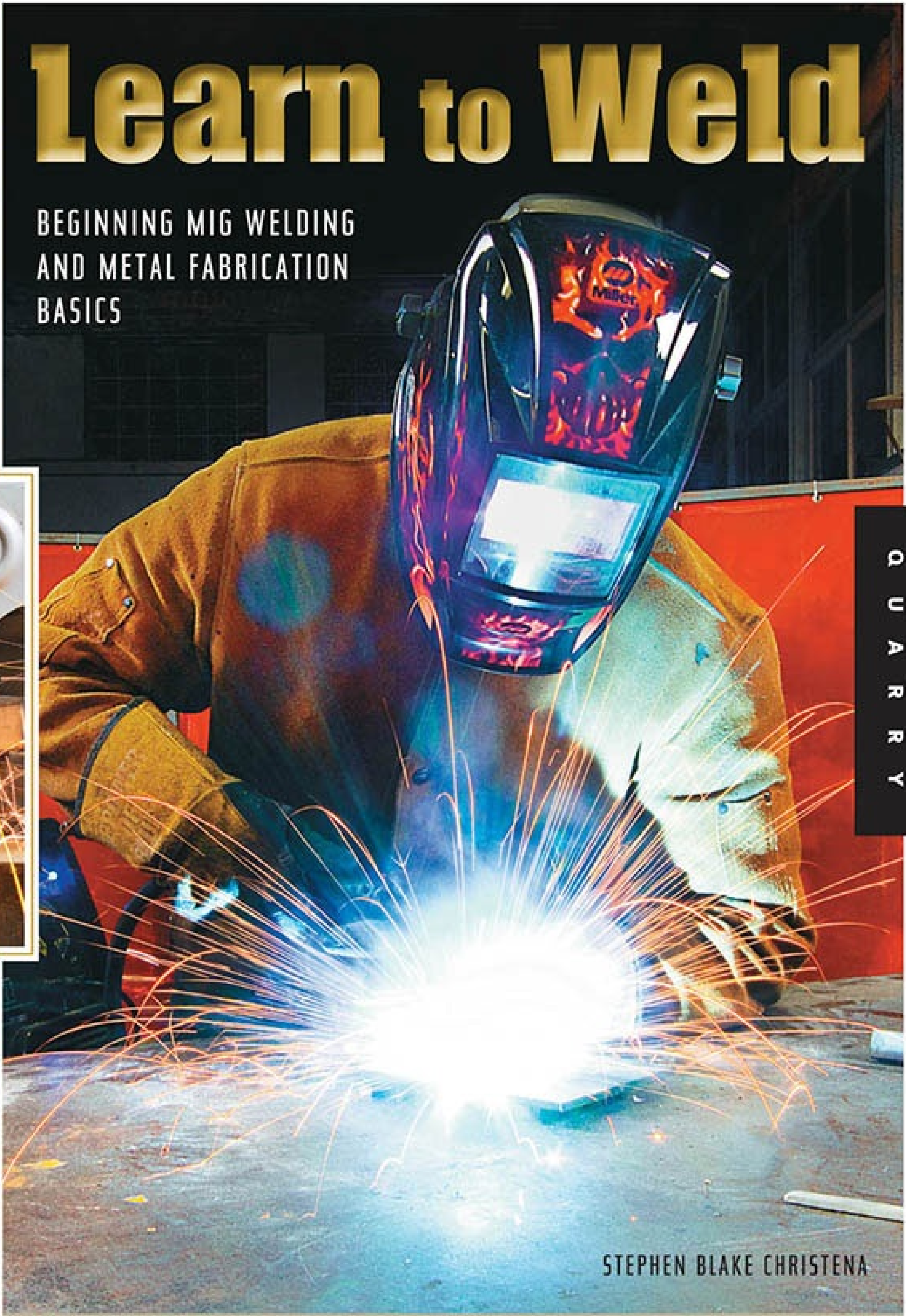

Learn to Weld

BEGINNING MIG WELDING AND METAL FABRICATION BASICS

STEPHEN BLAKE CHRISTENA

INCLUDES TECHNIQUES YOU CAN USE FOR HOME AND AUTOMOTIVE REPAIR, METAL FABRICATION PROJECTS, SCULPTURE, AND MORE

QUARRY

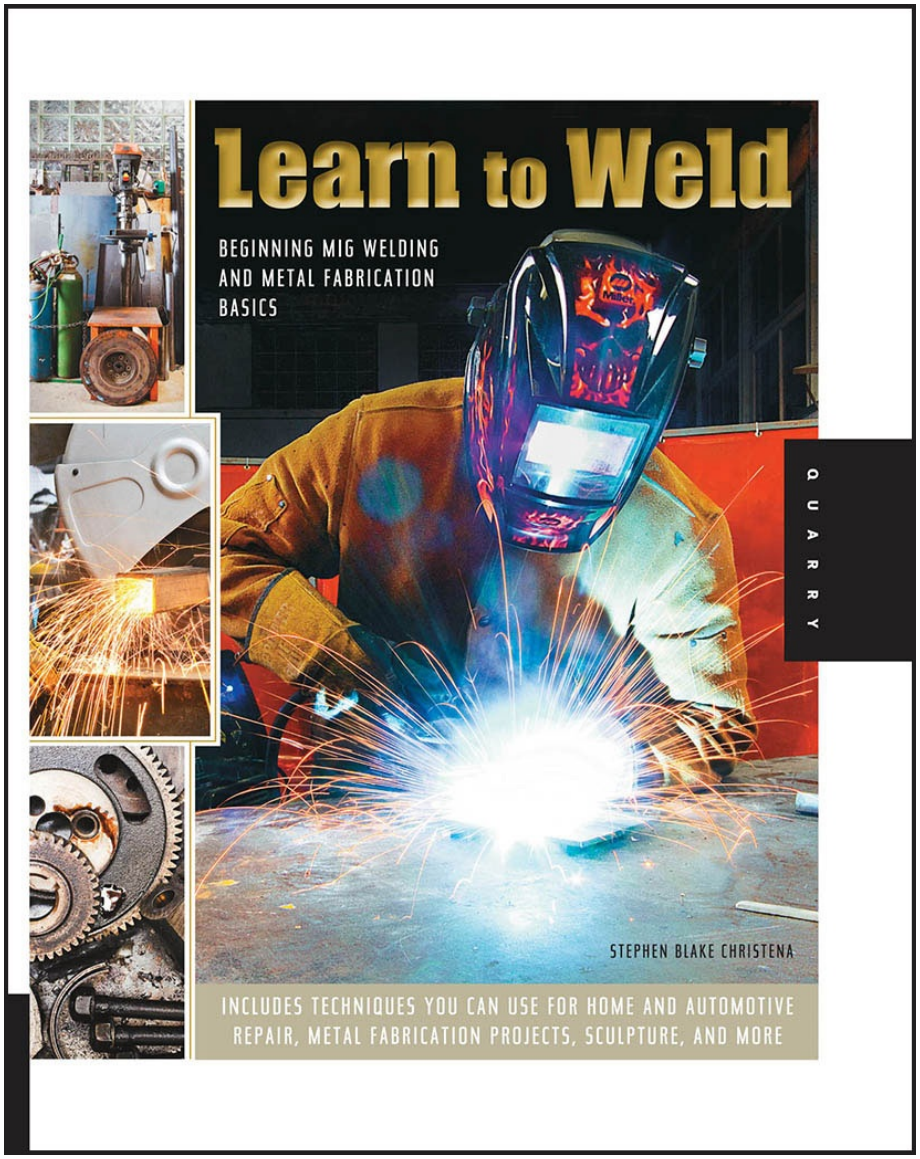

Learn to Weld

BEGINNING MIG WELDING AND METAL FABRICATION BASICS

STEPHEN BLAKE CHRISTENA

INCLUDES TECHNIQUES YOU CAN USE FOR HOME AND AUTOMOTIVE REPAIR, METAL FABRICATION PROJECTS, SCULPTURE, AND MORE

QUARRY

Learn to Weld

BEGINNING MIG WELDING AND METAL FABRICATION BASICS

STEPHEN BLAKE CHRISTENA

QUARRY
Quarry Books
100 Cummings Center, Suite 406L
Beverly, MA 01915

quarrybooks.com • craftside.typepad.com

Contents

Introduction

Part I: Welding 101

Chapter 1 Welding: Start Here

- How Welding Works
  - The Arc
  - The Filler Metal
  - Flux and Shielding Gas
  - The Ground
  - Spatter
- Why Learn MIG Welding?
  - Buying Tools and Equipment

Chapter 2 Getting Started

- Your Studio
  - Electrical
  - Ventilation and Lighting
  - Studio Layout
- Safety and Protective Gear
  - First Aid
  - Sparks Are Flying
  - Protective Gear
- Equipment and Tools
  - The Chop Saw
  - The Angle Grinder
  - The Welder
  - Other Cutting Tools

Chapter 3 Selecting, Buying, and Prepping Your Metals

- Hot Roll vs. Cold Roll
  - Filler Material
  - Decorative Options
  - Buying Metal
  - How to Order
- Preparing Your Metal
  - Let’s Prep

Chapter 4 How It Works

- Technique
  - The Puddle
  - Stance and Comfort
  - Vision
  - Hand Position
  - Angle of the Gun
- Heat vs. Thickness
- Clamps and Tacking
- Exercise 1: The Draw

Part II: Using Your MIG

Chapter 5 The Basics

- Basic Joints
- Technique
- Exercise 1: Butt Joints
- Exercise 2: Corner (or Fillet) Welds
- The Rest of the Joints
  - Corner Joints
  - Lap Joints
  - The Edge Joint

    - Exercise 3: Jigs
    - Other Welding Situations and Positions
        - Vertical Welding
        - Cross Body Welding
- Chapter 6 Projects
    - Project 1: The Cube
    - Project 2: The Welding Table
    - Project 3: The Runner
- Gallery
- Glossary
- Resources
- About the Author
- Acknowledgments
- Index

Miller

Introduction

While I was growing up in Flint, Michigan, my dad, brother, and I would make and fix all kinds of things. My dad instilled his philosophy in me at a young age: "Why pay someone else to do something you can do yourself?" We enjoyed fixing everything from cars to bicycles, dishwashers, and TVs. We re-roofed houses and rewired radios. We built wooden tool chests, skateboards, snowboards, and hockey sticks. We were always working on a wide variety of projects using everything under the sun and all kinds of cold connections like nuts and bolts, screws, nails, glue, epoxy, fiberglass, solder, pop rivets—you name it; we used it. I grew up into a confident handyman and consummate tinkerer with an appreciation for three-dimensional construction.

By the time I started college I was no stranger to creating sculptures, furniture, art, and all sorts of other things using traditional woodworking and cold connection methods. But then in my first metal sculpture class, I cut into a piece of 11 ga mild steel with an oxy-acetylene gas torch and I was hooked. Metalwork and fabrication amazed me and became my passion. The way this cold, hard, and seemingly uncooperative material was so easily cut, formed, and manipulated still astounds me. As I learned how to weld, heat, bend, and cut various metals, I could not believe how these incredibly durable materials became as easy to work with as clay. Exploring the unlimited potential of metal work became the compass for the rest of my life.

I never stopped welding. After graduation, I moved to Chicago and set up my first shop on the south side, equipped with only a small 120 V Hobart MIG welder and grinder. Over the years, I moved my shop around a few times, expanded my gear, and worked in a wide range of fab shops, learning as much as I could along the way.

By the time I started my first fabrication company, I had worked for some of the top designers in the furniture industry and for world-class architects, and I'd worked side by side with amazing metal artists, fabricators, and sculptors. Each of them influenced my perspective and expanded my knowledge of metal work, for which I am grateful.

> "Exploring the unlimited potential of metal work became the compass for the rest of my life."

Then a few years ago some motorcycle friends approached me and asked about learning to weld and chop frames for their race bikes. After a few basic lessons in welding, they were asking me to teach them how to make their own bikes lighter, how to weld and fabricate new parts, and how to build custom components. Before I knew it I was teaching community welding classes at my metal studio, Midwest Metalworks.

After a lesson recently, one of my students said to me, “You can pretty much make anything, knowing how to weld.” He had an amazed expression of bewilderment at the vast potential of this new medium that he had just learned. He had realized the fantastic new world of possibilities that I had discovered almost two decades before. Now you can too.

If you enjoy working with your hands, learning metalwork will give you a wealth of knowledge that will expand and challenge your ability to create. There are so many different techniques, styles, and methods to learn in welding, and in this book, I hope to introduce to you the enormous possibilities that come with each one.

The key place to start is learning the basics of Metal Inert Gas (MIG) welding, which we’ll learn in the following pages. You’ll learn the basic rules and principles of welding that can be applied to other processes such as Stick and Tungsten Inert Gas (TIG) welding. MIG is just the beginning.

It takes time and practice to become a good welder, but I hope through this text you will soon instinctively recognize a good weld that is laid properly. Just like an excellent musician knows when they are in tune or have played a bad note, an experienced welder understands their craft. The more arc time you log and the more projects you fabricate, the more the world of metalwork will open up to you.

—Stephen Blake Christena, Chicago 2013

I

Welding 101

1

Welding: Start Here

I teach Metal Inert Gas (MIG) welding because it has become one of the most versatile methods of welding, it's easy to learn, and it's fun! It was originally developed for welding aluminum, but was quickly adapted to other metals such as mild and stainless steel. It has become extremely popular amongst all industries from hobbies to shipyard for its ease of use, versatility, and speed.

MIG welding is the focus of this book, but before we dive into it, I wanted to give you an overview of electrode arc welding and some insight into the readily available welding processes out there.

How Welding Works

There are several different methods of welding. Today, the most common welding process is electrode arc welding. There are three types, or systems, of arc welding in this family: stick, MIG, and TIG. All three types of arc welding use the same three components. The first and most obvious is electricity, which creates the arc. Each type also uses a filler material, and the third component is flux. The differences between stick, MIG, and TIG arise in how the three of these elements are used to create a bead (a bead being the weld that is created); each type of welding has a different method of execution.

Before we talk about the differences in welders and the methods of execution, we should go over how arc welding works in general. The best way to explain this process is to clearly define the elements of welding and what they do.

The Arc

Static shock, lighting, and spark plugs firing in a vehicle’s engine are all examples of an electric arc. Not many people are aware of this, but an electric arc is actually a state of matter called plasma, which is similar to gas. The arc is created by an electrical current moving through the work piece from the welding machine’s ground clamp to the electrode. The electrode is an electrical conductor used to complete the circuit, which allows the welding machine to create an arc. The arc is created by a breakdown of a gas that discharges plasma. Certain gases such as argon, carbon dioxide, and helium have a higher conductivity and ionization that assists in creating and maintaining the arc, making them candidates for a shielding gas.

Miller
Miller

Make sure to wear proper safety gear. Skin left exposed to the arc light can get a pretty bad sun burn. I don't recommend it.

In the stick and MIG welding processes, the electrode is also the filler metal. When the filler material comes into contact with the grounded piece of base metal, it completes a circuit that creates the electric arc. The arc forms at the very end of the filler material, which melts into the weld zone, creating the joint. In the TIG welding processes, the electrode is a rod of tungsten.

The arc temperature can range, depending on the welding process used and the setting of the machine or power supply, from around 5,000°F to around 18,000°F (2,760°C to 9982°C). Carbon steel, or mild steel, the most common form of steel you will be using, liquefies around 2,600°F (1,427°C). The arc's high temperature heats the base metal to the point at which it becomes liquid. This is known as the “puddle” or “pool”.

The Filler Metal

Appropriately named, the “filler,” or “filet,” metal is just that. It is a rod or wire of metal that is fed into the puddle that forms at the point where the arc makes contact with the base metal. It adds more material to fill the weld zone, feeding the puddle and creating the bead. The filler metal is usually the same type of metal as the type being welded, but in certain cases, it can be other metals. In the stick and MIG welding processes, it also acts as the electrode. Stick and TIG welding require the operator to manually feed the filler metal into the weld zone.

Flux and Shielding Gas

The word *flux* comes from the Latin word *fluxus*, meaning to flow. That is exactly what flux does in welding. It helps stabilize the arc and keeps contaminants or oxides out of the weld zone. In stick welding, flux is a chemical coating over the filler rod that burns off, producing a shielding gas that prevents oxidation of the base and filler materials. Flux can be made from a wide variation of compounds depending on the application and material to be welded.

Flux is a component of welding with many jobs to do. The number one job of the flux is to protect the base and filler materials from oxygen and other ambient gases. It creates a barrier called the shielding gas, protecting the arc and metal from outside ambient gases that cause oxidation. It creates a purified environment for the arc to exist in. The flux keeps the liquid metal clean from impurities, also called inclusions, which would otherwise contaminate the metal. And last but not least, it controls the arc, helping the arc come up to a higher temperature to create the puddle of liquid metal.

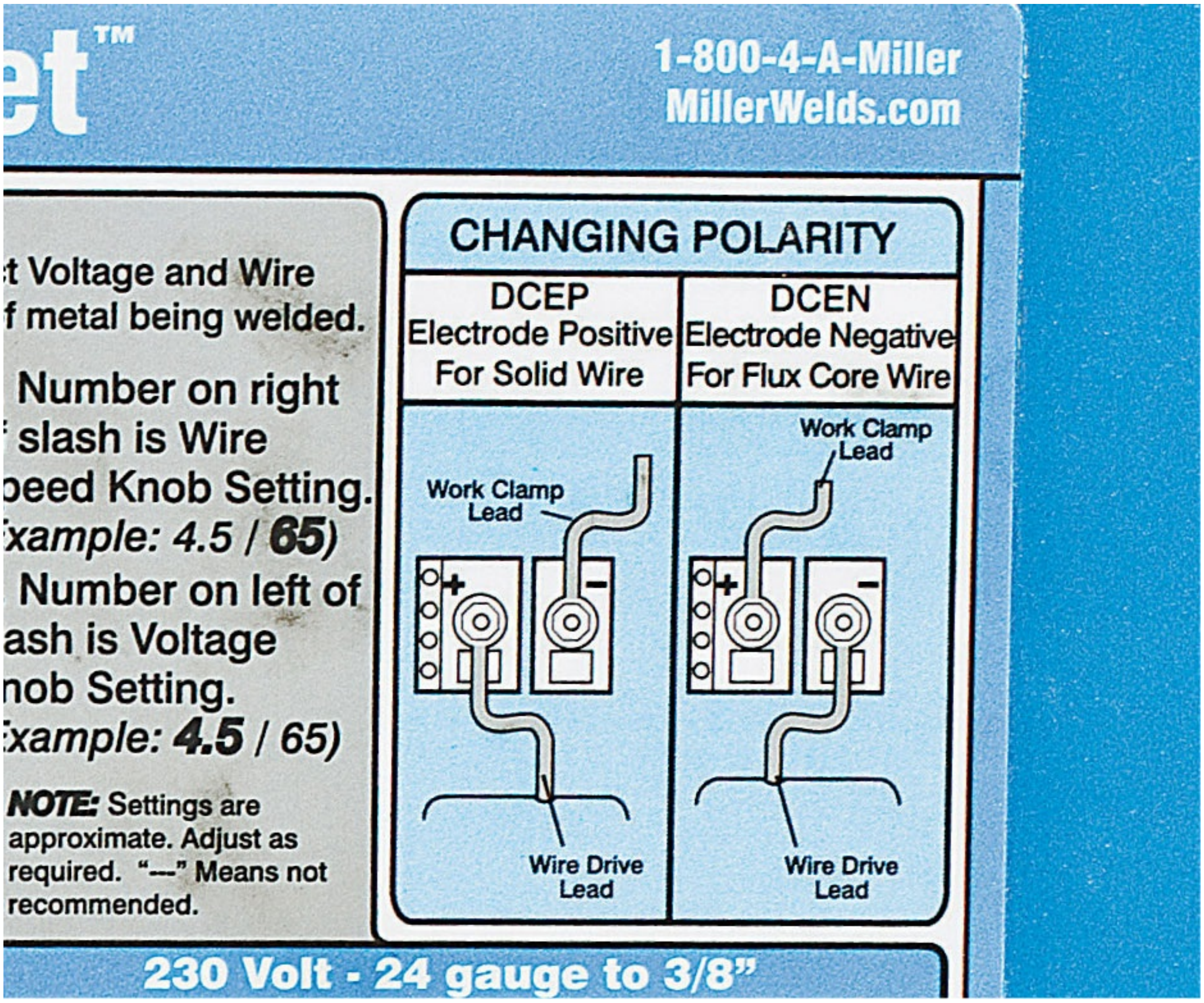

et™
1-800-4-A-Miller
MillerWelds.com
CHANGING POLARITY
DCEP
Electrode Positive
For Solid Wire
DCEN
Electrode Negative
For Flux Core Wire
Work Clamp Lead
Wire Drive Lead
Work Clamp Lead
Wire Drive Lead
+
–
+
–
230 Volt - 24 gauge to 3/8”

t Voltage and Wire f metal being welded.

Number on right slash is Wire peed Knob Setting. xample: 4.5 / **65**)

Number on left of ash is Voltage nob Setting. xample: **4.5** / 65)

NOTE: Settings are approximate. Adjust as required. “—” Means not recommended.

Be sure to change the polarity from DC positive to DC negative when you're using flux-core wire in your MIG welder.

Flux is most commonly used in stick welding, where it coats the filler metal. When the filler metal arcs to the base material, the flux is engaged. As it burns, it creates a carbon dioxide semi-inert gaseous shield that protects the arc. Once the weld is laid, the bead is covered in a byproduct of the flux called “slag.” Agents that come from oxides in the molten metal are absorbed and removed by the flux, creating slag. This hard shell is usually removed afterwards with a chipping hammer and a wire brush.

Flux is also used in autofeed machines, similar to MIG welders, which take a spool called “flux core” wire. This wire has the same properties and behavior as a flux-covered stick welder electrode. These machines are not considered MIG welders because they do not have the hookup for a shielding gas. Most MIG welders can use flux core wire, but their polarity needs to be switched over from DC+ to DC–. Some flux core processes can use a shielding gas in tandem with the wire.

Shielding gas is an inert gas that takes the place of a chemical flux. It does all the same duties as the flux. The first advantage of shield gas welding is that it has much better control over the arc. Just like blowing on the embers of a fire, the air pressure not only fuels the embers with gases, but also concentrates the direction of flow. Second, there is no slag byproduct. The shield gas process is cleaner. Third and most important, using a shielding gas makes it much easier to see the entire weld zone while welding. One drawback is that using shield gas outdoors has a tendency to not work very well. The pressure of the gas coming out of the nozzle can be easily blown away by the wind. Flux core and stick welding are the best options for outdoor welding.

The most commonly used shield gases are argon and a mixture of argon and carbon dioxide. Pure argon has great conductivity and heat transfer properties that will assist in creating a cleaner environment for the arc. And argon's ionization properties allow the arc to start and be maintained more effectively and efficiently. Pure argon is used in the MIG and TIG processes when welding on stainless steel and nonferrous metals (metals that do not contain iron) such as aluminum, titanium, or zinc. Hobbyist and beginning welders should focus on an argon/CO_2 gas mixture (75 percent argon, 25 percent CO_2). This mixture is ideal for carbon steel and will be the focus of this book.

The Ground

The ground is a crucial part of the power supply. All electric arc welders use one. In essence, you are creating a circuit, and without the ground, the circuit cannot be created. The ground clamp has to be clamped to the work piece or a conductive metal table that the work piece rests on or comes into contact with. The electrical current feeds though the ground and through the work piece, finding the shortest distance to the electrode as possible. The contact point between the ground and the machine's electrode is where the arc forms. If there is no ground, there is no arc.

Grounding to a conductive metal table allows you to manipulate the work piece easier. If you are going to be working on something large that has to remain on the floor, make sure to ground directly to the work piece.

Spatter

MIG and stick welding, as well as welding with flux core, create “spatter,” small molten balls of metal that are ejected from the weld zone at the arc point. These small molten balls can fuse themselves to the work piece and to the gun’s tip and nozzle. Using nozzle spatter protection will help the tip and nozzle last longer and prevent double arcing.

Small molten balls of steel known as "spatter" are ejected from the weld in progress.

There is a variety of spray spatter protection that can be sprayed onto the work piece. These products help to protect the work piece from needing to have dozens of little metal balls ground off of the surface. They're especially ideal for projects with moving parts or precision work; you don't want these little balls of metal inside a crank case or gear housing.

Miller
Miller
THE POWER OF BLUE
Miller

Stick welding is the best method to use when you're working outdoors or on thick plate metal. Use caution as the flux can create dangerous fumes and a byproduct known as "slag."

Why Learn MIG Welding?

Shielded Metal Arc Welding (SMAW) is what is commonly referred to as "stick" or arc welding. When most people think of electric arc welding, they think of stick. This is a manual feed process in which the filler metal electrode is coated in a chemical flux. The arc forms after striking the electrode to the grounded base metal. As the filler material melts into the puddle, creating the bead, the flux burns off, producing a shielding gas to protect the arc. A byproduct of flux called slag solidifies over the freshly made weld. Stick welding can use a direct current (DC) or an alternating current (AC) power to create the arc.

TIG welding is clean and smooth, but can be difficult to master. It's a good idea to learn MIG welding first and work your way up to TIG.

Stick welding is popular in industrial applications for its versatility and ease of use. The use of a flux makes it ideal for outdoor heavy construction. The wide variety of filler electrode sizes and metals allows quick transition to different thicknesses of material. And the high range of heat makes stick welding perfect for thicker materials.

Gas Tungsten Arc Welding (GTAW), or what is commonly referred to as Tungsten Inert Gas (TIG) Welding, is also a manual feed process: the operator is feeding the filler metal into the molten metal puddle. But in the TIG process, the filler metal is not the electrode.

TIG is a shield gas welding system in which tungsten is the electrode that produces the arc. The torch holds the tungsten rod, allowing the shield gas to pass over its tip through a nozzle. The arc is initiated when the control to the power source is activated, normally via a foot pedal controller. Much like the accelerator in a car, the pedal control allows the welder a range of amperage that creates the arc, giving more control over the heat going into the material. This control makes this method very popular for welding thinner materials and stainless steel.

MIG welding is easy to learn and fun to do. As soon as you pull and hold down the trigger you are welding. You just need to learn to control what you are doing.

TIG welding does not produce a spatter during welding, which makes it cleaner and more appealing to the automotive and motorcycle industries and to artists who would like their welds to be a part of the overall look of the piece. The operator has much more control over all the components than other systems.

Like its cousin the stick welder, TIG is a manual feed process with the ability to transition very quickly between metals. Pretty much all the operator has to do to switch from steel to stainless steel is to grab the appropriate filler rod. (Making sure you have prepped the metal and area around the piece for cross contamination is also recommended.)

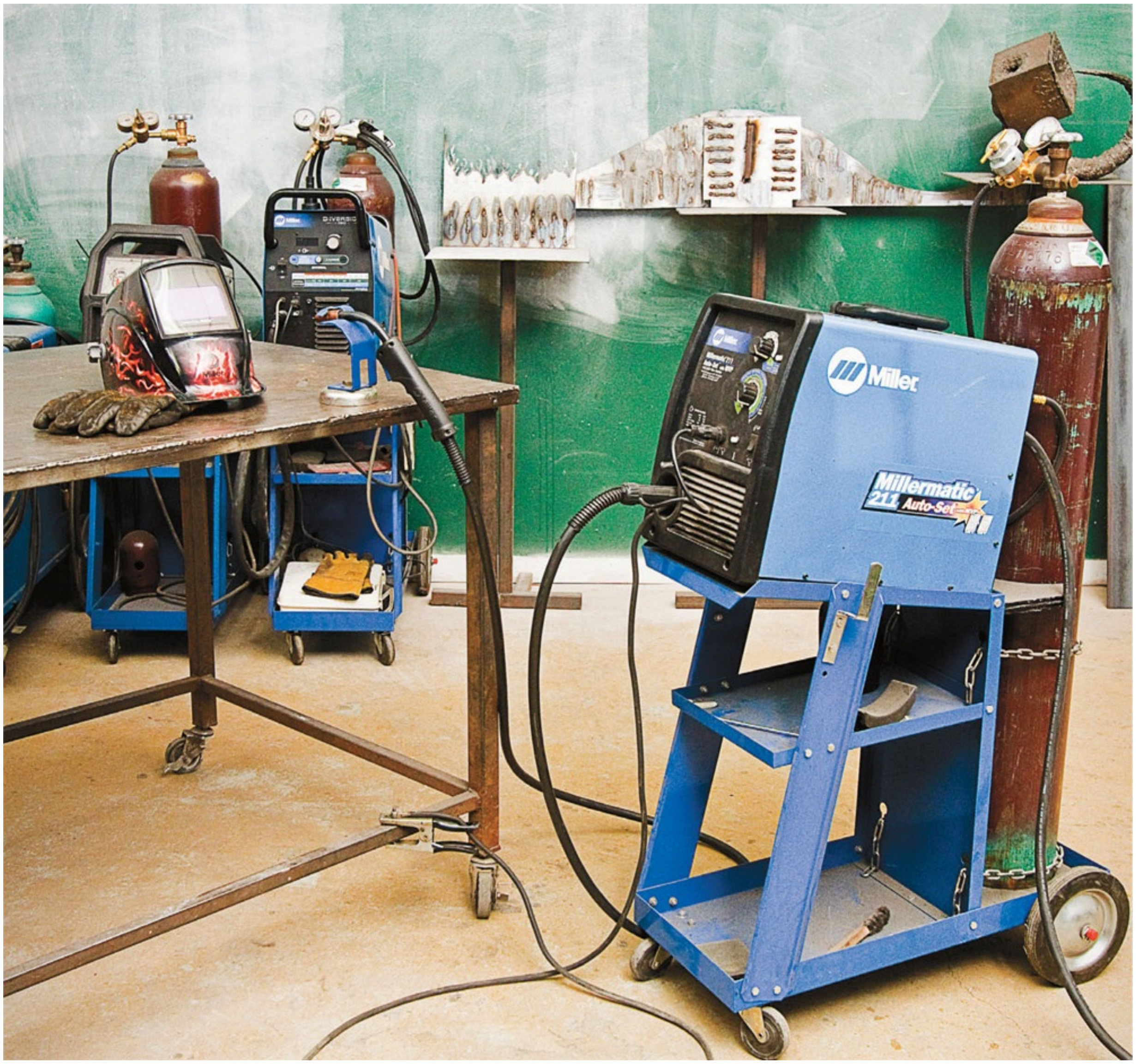

All electrode arc welding systems need a ground to complete the circuit and create the arc. The ground needs to be clamped either directly to the workpiece or to a table made from conductive metal.

Most TIG machines are equipped with DC and AC power supplies that can be changed by the push of a button. Making a quick switch from steel to aluminum is as easy as switching the output from DC to AC and grabbing the correct filler rod, which makes it popular for welding aluminum and other nonferrous metals.

The disadvantages of TIG welding include the fact that it does take a lot of time and practice to master before one can produce a good weld. It is also a slower process that takes much longer to produce a weld. And a decent TIG system has a tendency to be much more expensive than its cousins stick and MIG.

A well-trained TIG welder can make welds that are not only structurally ideal but nice to look at as well. The expressions “stack of dimes” or “nickels on the side” stem mostly from TIG welding; they refer to the look of the surface of a bead that resembles a stack of coins lying on its side. Using various control techniques, a skilled electrode arc welder can create patterns on the surface of the weld that resemble coins, ripples, waves, *etc.* This comes in handy when welds are exposed and on display, such as on motorcycle frames, sculptures, and furniture.

Gas Metal Arc Welding (GMAW), or what is commonly referred to as Metal Inert Gas Welding (MIG) and sometimes also referred to as Mechanized Inert Gas, is an automated process in which all three components initiate at the same time once the trigger on the gun is pulled.

The filler metal for a MIG welder is a wire that comes on a large spool. Just like in stick welding, the filler metal is also the electrode. The machine has a motor that feeds the wire through a hose to the gun. The power supply that creates the arc is fed through the same hose as the shield gas and wire. A direct current (DC) is most commonly used in MIG welding. A nozzle covers the gun tip where the wire comes out. Under the nozzle are a series of uniform holes that deliver the shielding gas evenly over the arc point. Once the trigger is engaged, all three elements activate simultaneously and feed continuously as long as the trigger is depressed.

The rate at which the wire is fed through the gun and the amount of electricity used is determined by the settings of the machine. You set the voltage and the wire speed in relation to the thickness of the metal to be welded. The shield gas pressure is set using the regulators that are attached to the tanks of gas. The pressure setting for the shield gas is determined by the manufacturer of the machine, normally anywhere between 20 and 30 cubic feet per hour (commonly referred to as CFH). (Note: In countries using metric measurements, the common terminology is cubic meters per hour, or CMH.)

common terminology is cubic meters per hour, or CMH.)

The wire on a MIG welder feeds through the tip of the nozzle.

Flux-core can be used in most MIG welding machines. Flux-core is an automated wire feed process just like MIG, but it does not commonly use a shielding gas. Instead, a chemical flux that produces a CO_2 gas when burned is mixed in with the core of the metal wire. Very similar to stick, it also produces a slag coating. Always work in a well-ventilated area when working with flux to avoid breathing the flux fumes.

MIG welding is one of the simpler methods of welding. It is easy to learn and has a wide variety of applications. It has also become the most common form of welding, not only in the arts and hobbies world, but in the professional metal working world in general.

> What you will need to begin is the MIG unit, a grinder, and a chop saw. All the techniques in this book will focus on these three tools.

If you want to learn to weld, this is where you start. The automated process will allow you to learn all the basic principles and techniques of welding by eliminating some of the controls that are normally affiliated with manual feed methods like stick and TIG.

Becoming a skilled MIG welder will allow you to create art and furniture or repair pretty much anything in steel. MIG welding and simple fabrication will be the focus of *Learn to Weld*. This book will provide you with all the basic information and tools you need to begin your own projects and build your interest in metal work and welding.

My approach to this book is going to be a little unconventional. It seems that most DIY books give you a laundry list of materials and equipment to buy that may or may not pertain to the beginning stages of the learning process. My approach will be to give you the proper tools and materials list that you will need for each chapter as you need them. It is not necessary to go out and purchase a full metal fabrication shop of tools and equipment right out of the gate.

Always have lots of clamps available to lock down the metal as you tack and weld.

One thing that I have learned over the years, especially if you are going to be welding a lot or turning this into a profession, is to only buy out of necessity. You don’t want to go out and buy a $2600 plasma cutter when a grinder and a chop saw will work just fine for the needs of that job. Even though plasma cutters give you much more control and options for cutting, it’s an investment for down the road. Let’s start by teaching you how to weld and fabricate a few simple things. What you will need to begin is the MIG unit, a grinder, and a chop saw. All the projects and techniques in this book will focus on these three tools.

Buying Tools and Equipment

One of the most common questions I’m asked by my students after their first

One of the most common questions I'm asked by my students after their first class is “What kind of MIG should I buy?” Well, there really isn’t a straight answer to this question. As it goes with many tools, the longer a person welds, the more preferences they develop. Welders have a tendency to be very loyal to their brands.

When it comes to welding machines, there are huge jumps in quality from brand to brand. This is the one piece of equipment that I do not compromise on. Like I said, “preferences”. Grinders burn out all the time. Some last a few years or even a decade or so. Others burn out in weeks. I actually buy cheap grinders for on-site jobs so if anything happens to it I’m not out by much. But the right welder will last a lifetime and is worth the extra few hundred bucks.

The first few things to consider when buying a welder are how often you’re going to be welding and the kind of electricity you’ll have access to in the area that you’re going to be welding in.

Both of these questions are important to any welding machine for one reason: the duty cycle. The duty cycle for any electrode arc welder is a 10 minute period of time where the welder can be running straight “arc on” and then has to rest in the ON setting to complete the full 10 minute block of time. It can be turned off if you don’t intend to use the welder for a long period of time, but in its ON setting the internal fans will allow it to cool properly. For example, my welder has a 30 percent duty cycle when plugged into 230 V. That means this machine can run arc on for 3 minutes and then has to rest while still ON for 7 minutes to complete the full duty cycle. This ensures the machine does not overheat. When welding intermittently, the time at rest will factor into the duty cycle of the machine. It’s best to monitor the usage referred to as arc time.

Other factors like ambient temperature can affect duty cycle. A pretty standard temperature duty cycle rating for most machines is between 68-82°F (20 to 28°C). If the overheated light comes on more than three or four times in the machine’s life, it can seriously damage the machine. Don’t stress about it too much, just make sure you are allowing the machine to rest.

Just a heads up, if you find that you can weld for three minutes straight, you may want to think about welding as a profession.

How does this answer our questions? If you plan on welding a lot, you will want a machine that has a higher duty cycle. Most machines with high duty cycles will normally plug into 230 V or higher. Do you have 230 V in the area where you will be welding? If yes and you are planning a major project after you learn how to weld, you may want to consider this type of machine.

The newer machines today have a lot of new features that allow quick transition between 115 V and 230 V. If you have access to 230 V in the work area, I would recommend taking a look at a few of these models. But for now, I area, I would recommend taking a look at a few of these models. But for now, I would start small. There are a lot of great 115 V and 120 V machines out there that you can do a lot with.

The next question is how thick you'll be welding for most of your projects. The 230 V machines will allow you to weld thicker gauge materials, whereas 115 V machines start to max out at around 3/16 (5 mm)- to 1/4 (6 mm)-inch-thick material.

Look out for autofeed machines. Autofeed machines are not necessarily inert gas welders. The autofeed-only machines are for flux-core wire only and do not have a shield gas hookup. Most MIG machines can also use flux-core wire.

It is not always important to buy the top of the line tool for something you won't use that often. Always research the machines and tools you need before purchasing them. If you don't already have a welder and are looking to buy used, make sure you understand how the unit was used. For example, a hobbyist welder or a "fix-it-yourselfer" is going to have a lot fewer duty hours on their machine than, say, an auto-body shop or fabrication shop. The age of the unit is not as big of a factor as the actual arc time the welding unit has on it. The ideal used MIG welder is being sold by someone who bought it five years ago to learn how to fix stuff and is now selling it because they lost interest and are more than likely just trying to clean out the garage. That welder only has a few hours of duty time on it tops, and it is going cheap.

There is always a risk when buy anything used. So before buying a used machine, read the chapters that involve the machine. This will give you a better understanding of how everything is supposed to work. Look over the machine to make sure there isn't any damage to the terminals or any of the components on the inside. Turn it on; if the machine's fan turns on right away with out any struggle that is a good sign. Run some wire through it to make sure the gun liner is not kinked and that the wire is moving freely to the gun. Check the ground clamp to make sure that the cable is secured correctly. Check the ground clamp cable for holes or cuts in the insulation. If the person selling it will let you, put your welding helmet on, put a piece of mild steel in the ground clamp, and make sure it arcs. Always check the exterior shell for damage as well.

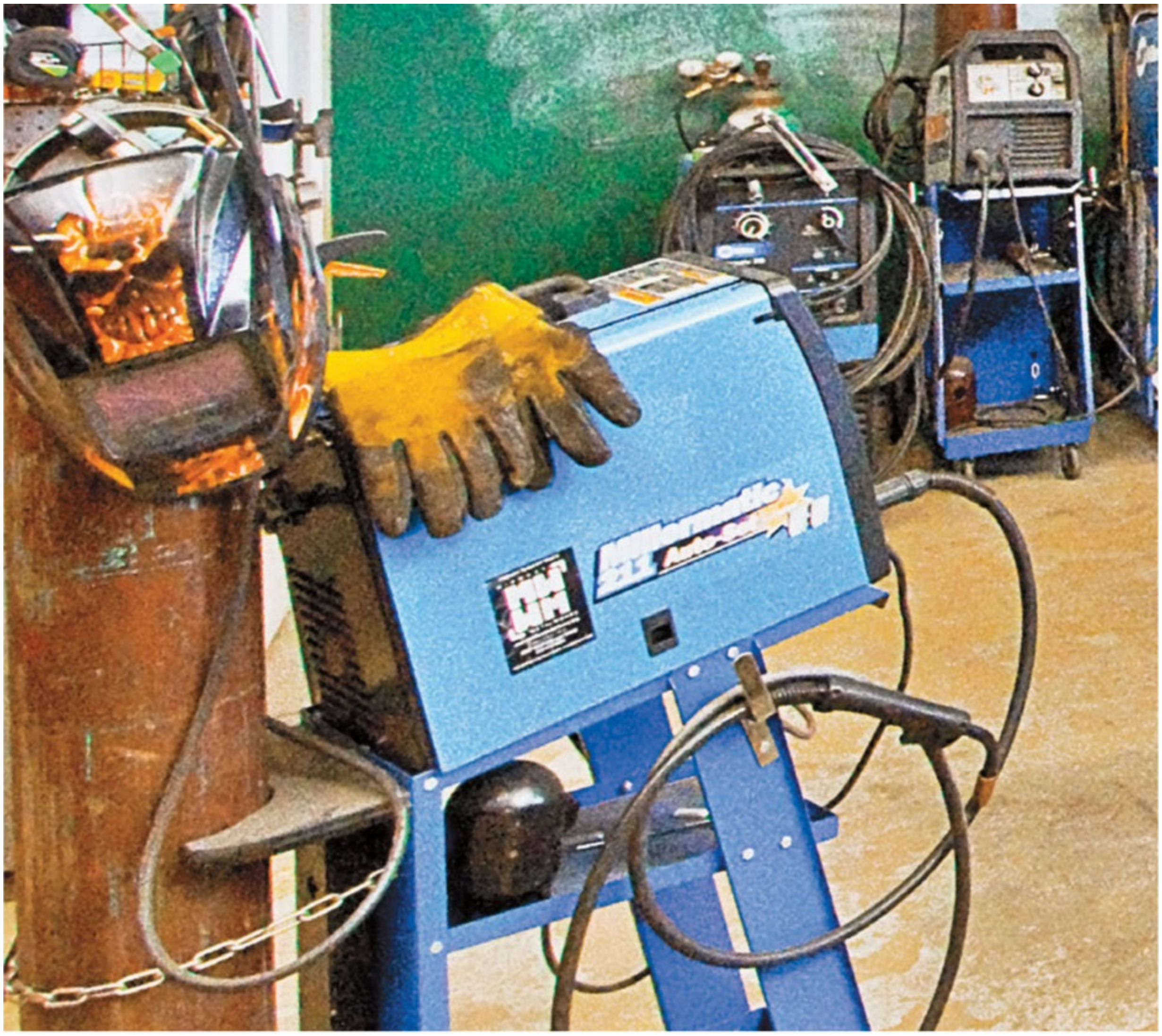

Before you buy a MIG, consider how you'll be using it. If you don't plan on using it very often, you can skip the top of the line models and choose a more affordable one.

If you are buying a new welder, make sure it is a trusted brand such as Miller, Hobart, or Lincoln. It’s a good idea to talk to someone you know or can trust in regards to the quality of a welder you’re considering. There are a lot of welding manufacturers out there that don’t have all of the safety features you need. Some of the more inexpensive wire feed units on the market today don’t have proper shut offs such as electric. When the trigger is released, the filler electrode can still arc if the nozzle touches the grounded base metal. This is extremely dangerous if you are not wearing the proper protection or if your welding helmet is up. If you find a great deal on a new welder, make sure you trust the brand.

Of course, there are the standard benefits to buying new, such as warranty. When buying new, it’s not a bad idea to build a rapport with the closest welding supply store. Normally, the people that work at these businesses have an in-supply store. Normally, the people that work at these businesses have an in-depth knowledge of welding, which is helpful if you are not sure which consumable products to purchase for your machine, such as contact tips and nozzles.

2

Getting Started

Now that you’ve decided to learn MIG welding, the first things you’ll need are a place to weld, the proper safety gear, and of course, the right equipment. Safety should always be your first consideration in choosing a location. As a welder, not only will you be working with extreme heat, but you’ll also be working with compressed gases, vapors, fumes, materials, and chemicals that are harmful to you and others. A studio absolutely must have proper ventilation, exits, and electrical access suitable for the machines you use.

Your Studio

Welding is a practice that involves high heat, lots of sparks and spatter flying around, and things that can cause fires. It’s not a hobby to start in the home. A garage, shed, or some kind of detached structure would be ideal. Masonry structures made out of concrete, brick, or cinderblock work the best as long as they have good ventilation. Sheds with concrete floors and corrugated metal walls are great places to set up a welding shop.

A garage is also a decent place to set up a small welding area. Normally, garages are already fitted with concrete floors, so the rest of your fire prevention equipment can be pretty easy to install and not very expensive. Ventilation can be provided by simply opening the garage door and directing the air out with a box fan. You also have the ability to move your welder outside on nice days that aren’t too windy. If you have an attached garage, you may want to consider using a work shed instead. But if you are set on using the garage, make sure the welding area is the furthest distance from the house as possible.

You don’t have to set up a full metal fabrication shop right out of the gate. A 6-foot (1.8 m)-square corner should provide enough room to be comfortable and get started. If you are in a garage that has exposed wood surfaces or bare studs, they should be covered with a flame retardant material such as sheet rock or, even better, sheet metal screwed right to the studs. If the floors are not cement, they should also be covered with a mortar-style compound such as a flame-resistant epoxy mortar. Welding blankets and mattes should never be used as a permanent solution to protect floors from fire. These are for temporary protection for short projects; they aren’t 100 percent resistant to flames.

Basements: These are not good! You are working with a process that produces all sorts of harmful byproducts. Basements don’t normally have the ventilation required for a safe welding environment. Also, shielding gas is roughly twice as heavy as air. This means that in small, confined areas, it will displace the air that you breathe and you will suffocate. Just like drowning in water or any other fluid, you will lose consciousness and die. And we don’t want that. To make matters worse, whoever sees you on the ground will try to come save you, and they might be met by the same fate if the gas hasn’t dissipated first. So don’t do that.

Electrical

The only drawback to a garage is that most household garages are not built with 220 V–240 V power outlets for larger machines. For the beginner, this shouldn’t be a problem because you should be starting with a smaller 115 V to 120 V machine, which most garages should be able to accommodate. There are machines on the market today that convert from 120 V to 240 V simply by changing the plug, so you can always install 240 V later on.

When making the switch from 110 V to 240 V, have a licensed electrician install and upgrade the lines. If not done properly, bad wiring can lead to a damaged machine or worse. Always make sure the breakers are a match for the machines.

The location of the outlets is a good thing to consider when setting up a welding area. It’s a good idea to start off by putting the welder and the welding table in close proximity to the outlets to be used. Multiple outlets may be needed depending on the electrical demands of the machine. Welders draw a lot of juice, so having a dedicated outlet just for the welder is recommend to avoid tripping breakers. If everything is wired properly, this shouldn’t be an issue, but it never hurts.

Ventilation and Lighting

Ventilation is key to setting up your studio. Not only will you be welding, you will also be grinding, cutting, and heating materials. These processes all produce different kinds of sparks and vapors that can be harmful, so proper ventilation is vital.

Ventilation and protection from fumes and vapor can be as easy as turning on an exhaust fan or opening a door—something that will allow fresh air to flow in and circulate the contaminated air out. This can also be achieved by air purification or welding exhaust systems on the market. These have a tendency to be pretty pricey, so they probably aren’t the route to go in right away if you have another option, but they’re something to consider down the road. A centralized exhaust extraction system can also be made with corrugated hoses hooked up to an exhaust fan.

Another method of fume and vapor protection is using respirator masks. These masks come in all types of filtration from gases and fumes to light particulates. Some manufacturers have masks that can fit under welding helmets.

You will also want the welding area to be well lit. Chain-hung halogen or florescent lights directly above the welding area should be an abundant enough light source for welding. You will appreciate a well-lit welding area when setting up to begin welding and your lid goes down. The most common complaint I hear from my students is that they can’t see anything. Good lighting is going to help you set up and get comfortable before starting a weld.

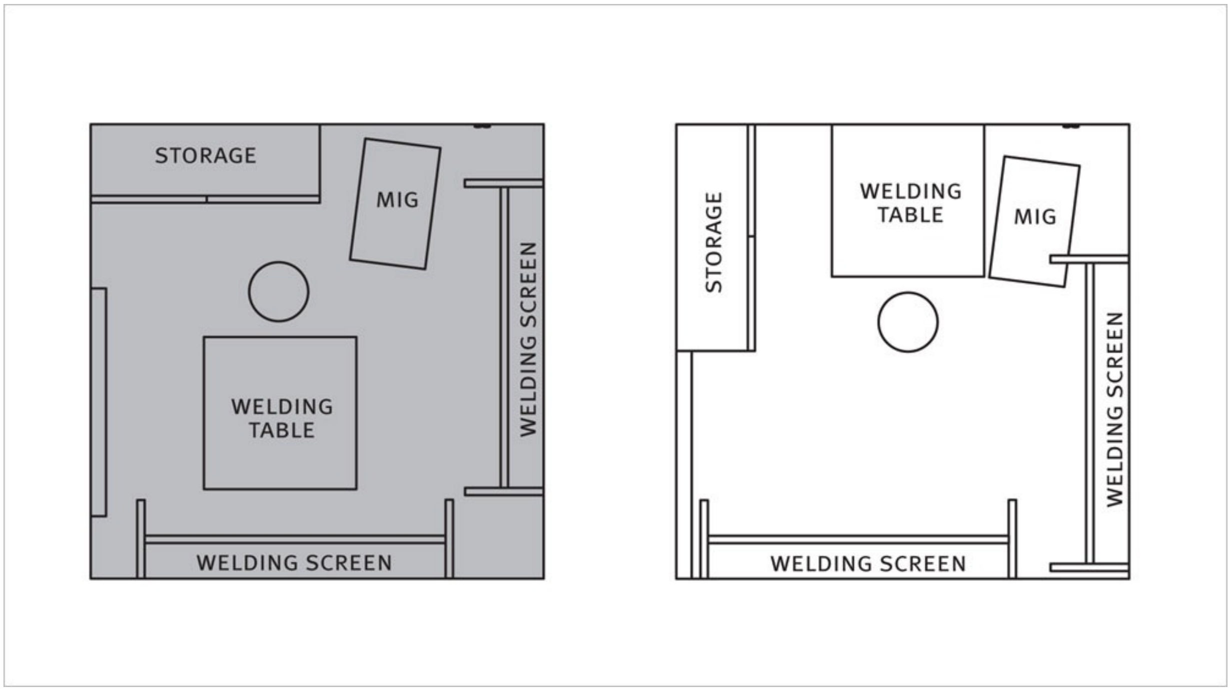

After safety, the most important considerations when setting up your work space are your comfort and access to your equipment. Here are a couple of potential layouts.

Studio Layout

For a small welding area, the layout is pretty simple. Arrange everything in a way that it is easily accessed. Arms lengths are great for grabbing a grinder or a clamp while still seated at the welding table. A wall-mounted hand tool rack is a great way to organize, conserve space, and have everything close by.

I have old steel filing cabinets for my handheld power tools such as grinders. They are durable and can normally be found for relatively cheap. One four-

They are durable and can normally be found for relatively cheap. One four drawer file cabinet can hold everything you need from protective gear to grinding abrasives to extra extension cords.

The second project in this book is building a welding table. The table I designed for you has 1 inch (25 mm) square tube rails to store clamps. You will want easy access to clamps as you will need them a lot. Last but not least, get yourself a comfortable stool. The best tip any welder can give to a beginner is to be comfortable.

The amount of space will also be determined by what you plan on building after you learn to weld. The larger the project—furniture and motorcycle frames, for example—the larger the space you will need. But to get started, a small space will be fine.

After you learn to weld, your enthusiasm will grow and so will your workspace. You’ll start to consider acquiring larger metal work machines, such as metal working lathes, benders, and drill presses. As you expand the work area, it is best to have more permanent areas for these machines due to their weight and size. These machines will also have accessories that will need to be stored. Finding or making new storage solutions will have to be determined while you decide to expand. You will find that this will happen over a slow period of time while you acquire more tools and interests.

I have moved my shop several times and worked in many different fabrication shops, and all of them are all a little different. It really all comes down to personal preference, needs, and your level of organization.

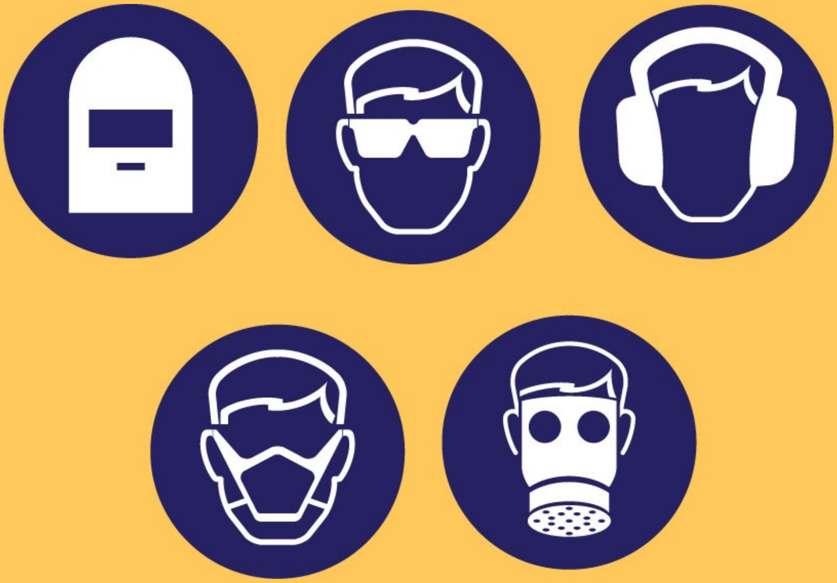

Be aware of all safety symbols provided by the manufacturer of the equipment used, such as these that represent welding helmet, safety goggles, ear protection, dust mask, and harmful vapor.

Safety and Protective Gear

Welding involves high heat, lots of very hot airborne projectiles, and intense light with concentrated rays such as ultraviolet, which can be harmful to your skin and eyes. Just like a cook in a kitchen, sometimes you will burn yourself. The longer you do metal work, the higher your probability of getting burned.

The most common way to burn yourself is to grab a fresh weld, not realizing how long you’ve let it cool (or haven’t). I’ve seen beginners do this many times. The best thing you can do to prevent burns is to be hyperaware of the temperature of the metal that you are welding. When in doubt, check the temperature before you touch it by holding your hand about 3 inches (7.6 cm) to 5 inches (12.7 cm) from the weld and slowly lowering it. If you think it’s too hot, it probably is.

First Aid

Keep a first aid kit in close proximity to the welding area. A standard first aid kit does not normally have burn treatment products. You should include some additional burn relief treatments such as burn gel, sterile bandages, and aloe. Or you can also purchase a burn kit to supplement the first aid kit. Minor burns can be treated by running room temperature or cool water over the affected area for several minutes. This will decrease the temperature of the skin and prevent further damage. Burn gel, triple action ointment, or aloe with a sterile bandage should be applied. Most importantly, if you have any doubt as to the extent of the burn, consult a physician.

Despite what you might see on TV, sunglasses are not a replacement for a welding helmet. Always wear proper safety protection for your hands, face, and body. It only takes a fraction of a second for something to go wrong and for you to get seriously hurt.

Sparks Are Flying

Spatter can reach great distances. Sparks from grinders and chop saws have been known to fly up to 40 feet (12 m) and smolder for up to 30 minutes, causing fires. Make sure that there are no flammable materials such as wood, gasoline, paint thinner, etc., in the welding area. Keeping a few buckets of water near the welding area gives you a great method of extinguishing small blazes quickly. Buckets of water are also useful for quenching hot steel.

Fire extinguishers are ideal for grease fires and other chemical-related fires and should be strategically placed around the shop. One should be centrally located, mounted to a post or wall between 4 and 5 feet (1.2 and 1.5 m) from the ground. Another should be kept within 6 feet (1.8 m) of the welding area and should also be mounted to a vertical surface. If the welding area is small, one extinguisher will do. Make sure the fire extinguishers are up to date and are checked annually.

Arc welding, which uses high amounts of electricity, can produce electric shocks, severe electrical-related injuries, or worse. In MIG welding, the wire is the electrode. Most MIG welding machines initiate all the elements of welding simultaneously when the trigger is pulled and cancel them when it’s released. Following a few basic rules will prevent injuries while using a MIG welder:

- Never get between the ground and the electrode when the arc is engaged.
- Never be grounded and pull the trigger near any body part.
- When clipping the filler wire, step away from the grounded piece.
- Hold gun handle below the trigger to avoid pressing when cutting wire.
- Always be aware of the gun in relationship to yourself and the ground.
- Always wear proper welding protective gear.
- Insulate yourself from the ground.
- Don’t weld wet, meaning you or the material.

If you don’t think you can do it yourself, have any power source and plugs for the machine(s) installed and inspected by a professional electrician to prevent electrical hazards. I’m more of a hands-on kind of guy, but even I have my limits. If you don’t know how to do something in the electrical arena, don’t. You can only hurt yourself. Use wire and breakers that are rated for the output of the machine. Always read all instructions, labels, and installation manuals before installing, operating, or servicing any welding or electrical equipment to prevent injury.

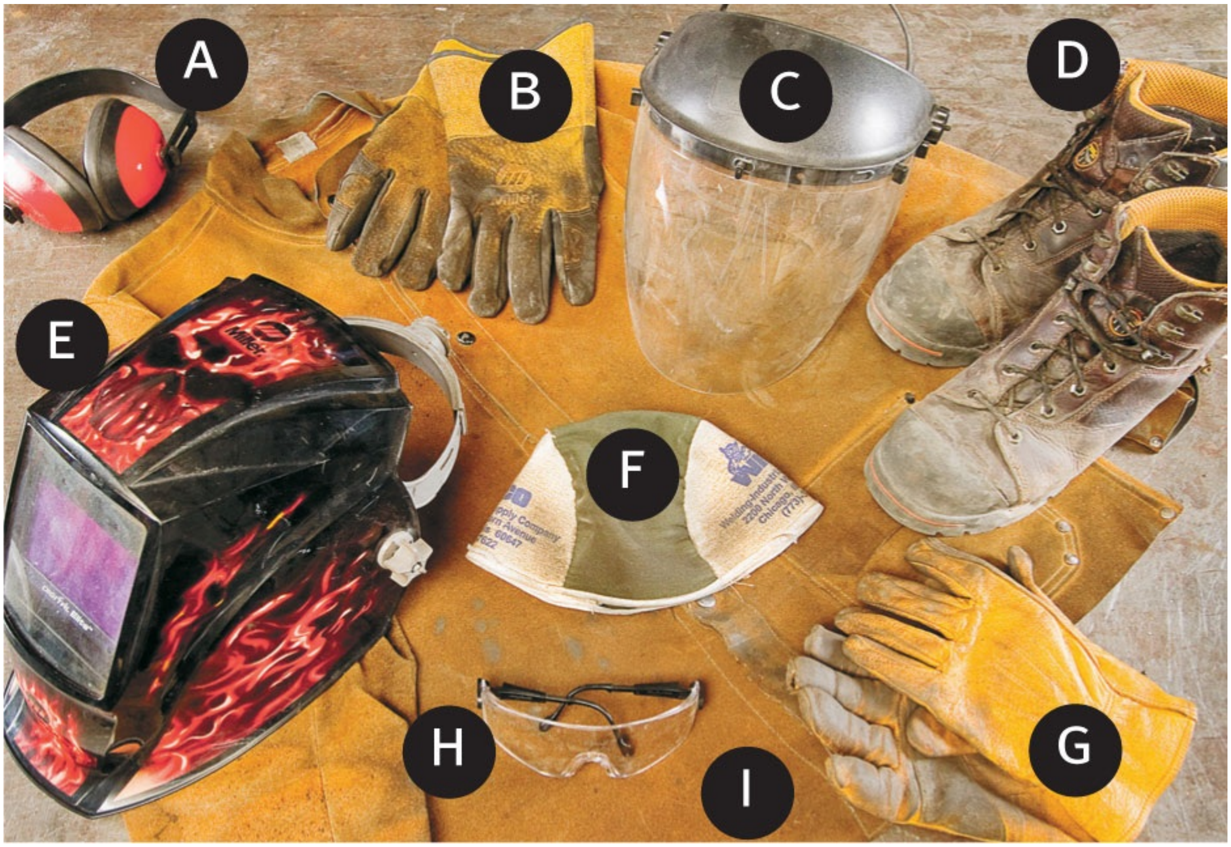

Here's the safety gear you'll need to get started: A. Muffs (hearing protection); B. MIG welding gloves; C. Clear visor ; D. Steel toe boots; E. Welding helmet (lid or cover); F. Beanie (scalp protection); G. Work gloves; H. Safety glasses or goggles; I. Welding jacket

Protective Gear

The arc light produces extensive amounts of harmful rays that can burn skin. Every 10 to 20 seconds in direct arc light is roughly the equivalent of a half hour of direct afternoon summer sun. Protect the skin with long sleeves, long pants, and MIG gloves to prevent UV burns.

MIG gloves can be bulky and cumbersome, but they provide protection from intense heat. The longer cuffs protect the forearm skin from arc rays that can burn the skin.

The welding helmet, hood, lid, or cover, no matter what you call it, should always be worn at all times during welding. The arc light is very intense and can blind you. The lid creates a shield for the entire head and neck to protect you from the arc light. It also protects the face and neck from UV burns and spatter.

When considering eye protection from arc light, one should be aware of the rating scale. The DIN is a shade rating of how dark the lens is or will become. Depending on the manufacturer or type of lid, the DIN can range from 0 to 13, sometimes higher depending on the application. Zero is the equivalent of clear safety classes or goggles, and 13 DIN is the darkest transparency to look through to prevent flashing the eyes with intense arc light.

This is a classic single shade (DIN) welding lid. The new autodarkening helmets allow you to set up with the lid down; the lens darkens as soon as the arc light hits the photo cells.

When I started welding, there were only single shade welding helmets, which

When I started welding, there were only single shade welding helmets, which were normally 13 DIN. You would have to get set up before flipping the lid down and starting to weld. The single shade helmets are perfectly good helmets, just not as versatile. The autodarkening helmets make setting up and staying on the weld zone before welding much easier.

Most welding helmets these days are autodarkening and are equipped with photocells. The photo sensitive cells signal the shield to go from the inactive shade to the active shade that the helmet is set for. For example, if the inactive DIN number is 5 and the active is 12, then when the arc light is detected by the photo cell, it changes to a 12 DIN setting in a matter of fractions of a second.

Most autodarkening helmets have a range of 5–13 DIN and have solar panels that feed the power source. The lower DIN shades are used for different applications such as plasma cutting, oxy gas cutting, and grinding. When setting the helmet, make sure to use the recommended DIN set by the manufacture. I normally have my helmet set at 12 DIN unless the voltage or amperage is dramatically different, higher or lower. I don’t recommend going lower than 12 when welding unless the manufacture of the helmet says differently for lighter duty applications.

I don’t care how many times you have seen another welder on TV weld without full-face helmet, no gloves, and no skin coverage, this is wrong. Get a good lid. Always make sure to use the proper shade shield for the job and make sure it is not damaged, cracked, or missing parts that will affect the performance or the fit. This book is not going to get into overhead welding, but additional protection may be needed for different welding positions. So, if you are a car person who intends to weld under a project car, you may want to consider additional scalp protection.

Skull caps or beanies should be worn under the welding helmet to prevent your hair and scalp from burning. Most caps are made of flame resistant material that will deflect most sparks and spatter. No matter what, scalp protection should always be worn unless you enjoy your hair catching fire.

When considering the torso, leather is normally the way to go. A good welding jacket or apron will last a long time. Modern fiber technology for welding has come a long way. There are many new protective gear options on the market that are more breathable than leather and provide more protection. Whatever you choose, make sure it is free of holes and protects the entire torso and arms. Welding is not a good way to work on your tan.

When welding, never wear nonorganic material unless it is specifically made for welding. Nonorganic materials such as nylon and polyester catch fire very quickly. They also melt into the skin and cause extensive damage. For example, most mechanics gloves do not make good metalworking gloves. Normally, the backs of these gloves are made of polyester and spandex that will melt and catch fire very rapidly if sparks or spatter hit them, just as with most tennis shoes.

I was working at a small foundry/fabrication company several years ago. One of my workers decided to wear tennis shoes to work one day against company policy. Right before lunch he was finishing up a fabrication job. As he started to walk away, he stepped in a coin-sized piece of spatter that was still molten. The spatter quickly burnt through the thin sole of his shoe. He was lucky and quickly removed his shoe before it could give him a more severe burn. You should always wear leather shoes with a good sole. Steel-toe boots are highly recommended, especially if you are going to be using heavier materials. There are a lot of work boots on the market that also come with metal strips in the soles to prevent burns and punctures to the bottom of the feet.

I shouldn’t have to mention at this point that your legs should be covered as well, but I will. Jeans and work pants, such as heavy duck cloth, are the norm. Once in a while, normally on the hotter days in the summer, people like to wear shorts. UV and spatter doesn’t care that you didn’t cover your legs or how hot it is outside. Also, do not cuff your pants. If a larger piece of spatter happens to fall into the cuff, then it’s a safe bet that your pants will catch on fire.

Never wear loose or baggy clothing when grinding. The material from loose clothing can catch on to a grinding wheel, binding into the arbor. This can be a quick way to get cut badly.

Metalwork such as welding, grinding, and cutting emits a good amount of harmful fumes and gases. Fumes are solid particles that originate from metal fabrication methods from the consumables, base metals, and any coatings present on the base metal such as zinc. Some of the airborne particulates can cause nausea, headaches, dizziness, and other health-related issues. If adequate ventilation cannot be maintained, a fume extraction machine or respirator mask may be used to reduce these harmful fumes.

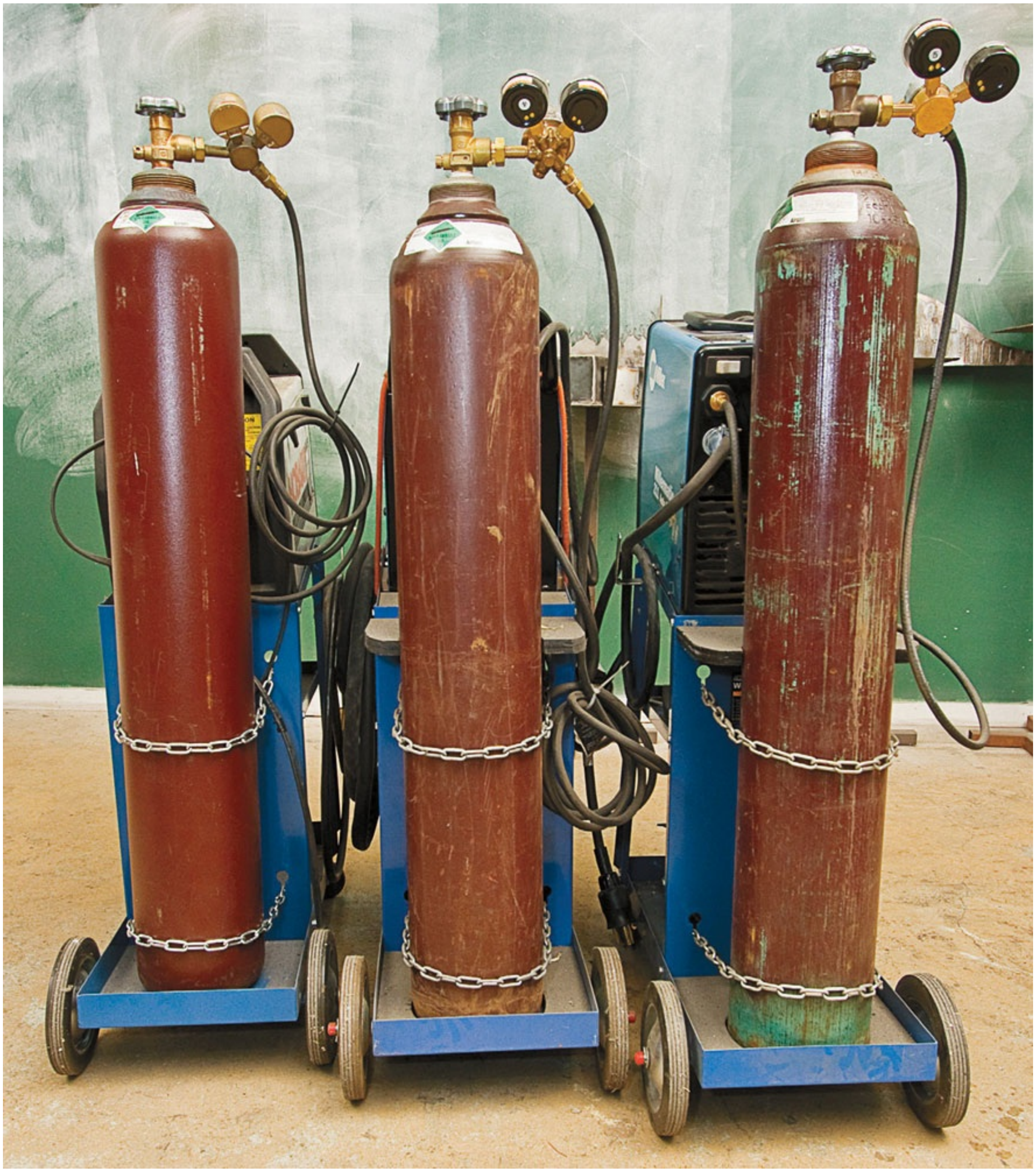

Always keep your cylinders chained to the welding cart or to a wall. You don't want a cruise missile going through your home.

Once when I was growing up, my dad took me down to the valley in Flint, Michigan, where he worked and where all the automobile manufacturing plants were. Earlier that morning there was a small industrial accident. He pointed at a large hole in a concrete wall of one of the factories and then at a hole that went through a conversion van and then a train car about 50 feet away from the van.

They happened to line up perfectly. “You know what did that, kid?” he asked. “A cylinder of compressed gas fell over and knocked off the valve.” He went on to explain how the cylinders become steel rockets if they still have pressure when a valve breaks off and they fall. The moral of this story: cylinders have so much pressure they will go through a concrete wall and only stop after they hit the other side of a train car. Or worse, they can explode.

When storing cylinders of any kind, flammable or nonflammable gas, make sure they are on level ground and secured with a chain or heavy strap to a wall, table, or cart. When not in use, always keep the cylinder caps screwed on tight. Unless you’re looking for a very interesting and dangerous way to add some ventilation, always make sure all cylinders are secured.

When a cylinder is not in use, the primary valve should be covered with its cap.

It is also a good practice to separate the flammable gases from the nonflammable gases. Chain them up in two different locations. This will prevent any confusion when changing out gases. The cylinder valve threadings are different for flammable and nonflammable gases, so you don't have to worry about using propane or acetylene as the gas for a MIG welder, but it is safer to err on the side of caution. It is also best to keep the flammable bottles further away from heat sources such as the welding area.

If you happen to have a larger shop space, welding carts are great for increased mobility. When using any welding cart, make sure that the cart can accommodate the load and has a secure way of mounting the shield gas cylinder. If working in a smaller space where the welder is going to be stationary most of the time, make sure that the shield gas cylinder is secured to a wall or heavy table.

Keep flammable gas cylinders separate from nonflammable gas cylinders.

You should always read all instructions, labels, and installation manuals before installing, operating, or servicing welding or electrical equipment to prevent injury.

So, I have covered the basics. If you have more concerns in regards to welding safety, please feel free to reference the American Welding Societies web page at www.aws.org. They have an extensive free library of safety sheets that can come in handy when in doubt.

Equipment and Tools

Basic welding and metalworking doesn't require going out and buying lots of different equipment and tools. Even though there are a wide variety of tools for metalworking, there are only three simple things that you will need to get started: a MIG welder; a 4 1/2-inch (11.4 cm) angle grinder; and a metal-cutting chop saw. That's it. Learn to use these three tools properly and you can build nearly anything. In time, as you grow your knowledge of metalworking, you can add to your arsenal of equipment as you explore different methods of building.

Fabrication always follows the same steps: Design–Measure–Cut–Prep–Tack–Weld–Finish. The design and measuring can be considered part of the cutting step. You never cut anything without knowing the length. Since the finish is subjective and is chosen by the fabricator, the mantra that will echo through this book will be "Cut–Prep–Tack–Weld." Cutting will be done for the most part by a chop saw, Prep with a grinder, and Tacking and Welding with a MIG welding unit.

A chop saw is great for making straight or angled cuts. This will be the cutting tool we will be using for the projects in this book.

When setting up or changing the disk on a chop saw, follow the recommended safety steps from the manufacturer. Always remember to unplug the saw before changing the disk or clearing debris from the guards.

The Chop Saw

Fabricating in metal is as simple a concept as cutting and pasting. It’s just on a much heavier, more industrial level. For the cutting aspect of fabrication, we will be focusing on the chop saw. A metal-cutting chop saw is specifically designed for straight and angled cuts. Unlike the miter saw (its woodworking counterpart), the chop saw is limited in its ability. Miter saws have more options for cutting angles, but should never be used for metalworking.

There are many other methods for metal cutting, sanding, grinding, bending, and shaping—methods using tools such as band saws, plasma cutters, press breaks, and shears—all of which have their place. You can spend a lot of time and money developing a full-scale metal fabrication shop. If you already have some of this equipment, that’s great, but I wouldn’t start there.

Whatever machine you decide to use, make sure it is specifically designed for metal. Cutting and sanding metal produces high heat that can actually melt components in non-metalworking equipment. The speed of the blade, the power of the motor, and the components used to build a metalworking machine are designed for friction, high heat, and metal dust. Using wood-only machines on metal can cause damage to the device and cause serious injury to you.

For our purposes, a consumable blade chop saw will work just fine. A consumable blade chop saw is just that: the blade is made from a composite material, normally aluminum oxide reinforced with fiberglass, which consumes itself as it cuts. Metal-cutting chop saws can also come with metal blades that cut cleaner and with more accuracy. They also tend to be twice as expensive as their consumable blade counter parts. Even though these machines are very nice to have in a metal shop, this is not the place I would recommend starting.

When using a chop saw, understand that these saws will not cut as quickly as a standard miter saw cuts through wood. You will want to be aware of how much pressure to put on the metal as it cuts. Not enough pressure and it will produce more heat and take much longer to cut. Too much pressure and you can overwork the motor. This can lead to stoppage of the blade or damage to the motor. When in doubt, listen to the machine. If it sounds like it’s whining and working too hard, you are probably putting too much pressure on it.

The chop saw is a great tool for cutting bars and tubes with straight or angled cuts. This is a good machine when fabricating furniture or anything that has to be framed out.

Setting up the Chop Saw

Right out of the box, the chop saw is relatively simple to set up. Most of the time, you’ll only need to assemble the fence and attach the blade. Always leave the saw unplugged before you install a blade. Pull back the lower guard and secure it in the up position with a pair of clamps. Using the provided arbor wrench, hold down the arbor lock and unscrew the arbor nut. Slip the blade under the upper plate guard in order to position the blade on the arbor. Make sure that all of the flanges and spacers are assembled properly.

CHANGING THE CHOP SAW BLADE

After you've used a blade or two, there will be a lot of debris inside the upper plate guard. When you put a new blade in, stay at arm's length from the saw and let it run for a few seconds to knock out all of the metal and old blade remnants. You should be wearing eye protection and a mask or respirator.

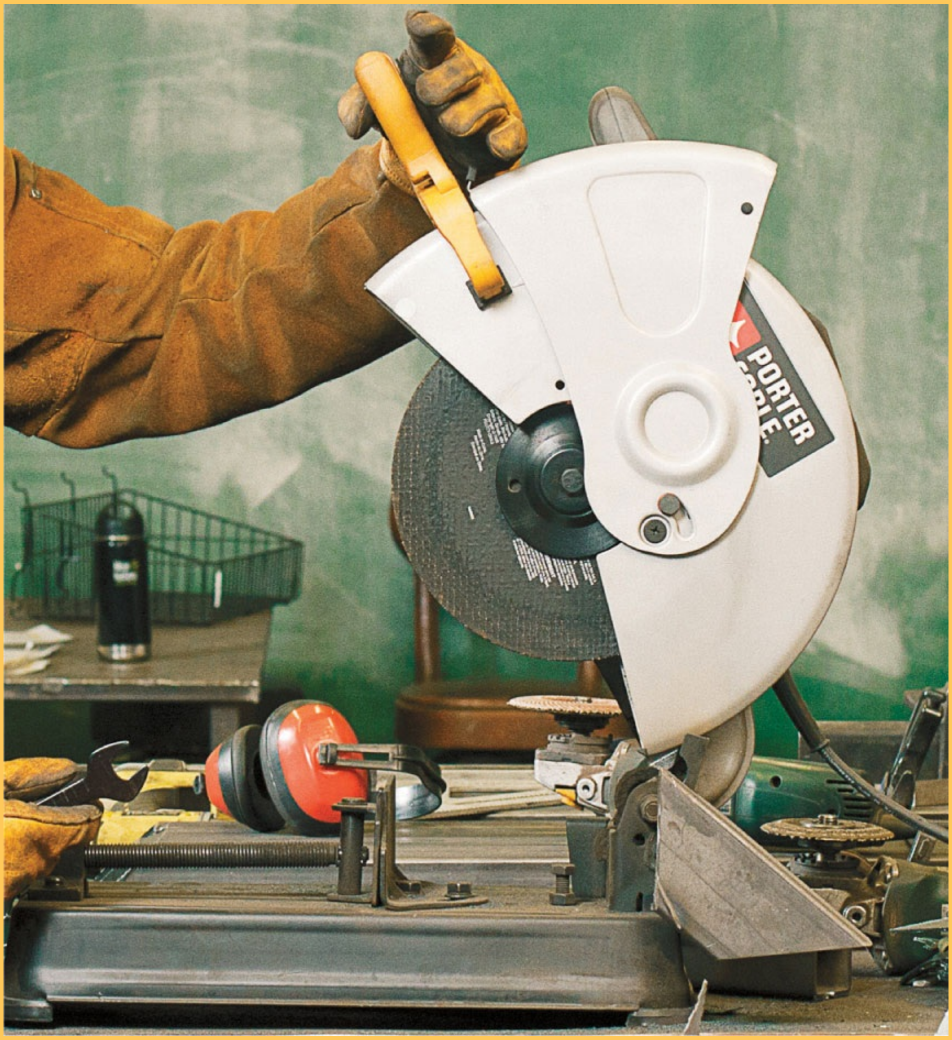

It's easier to access the arbor nut when the front guard is clamped back.

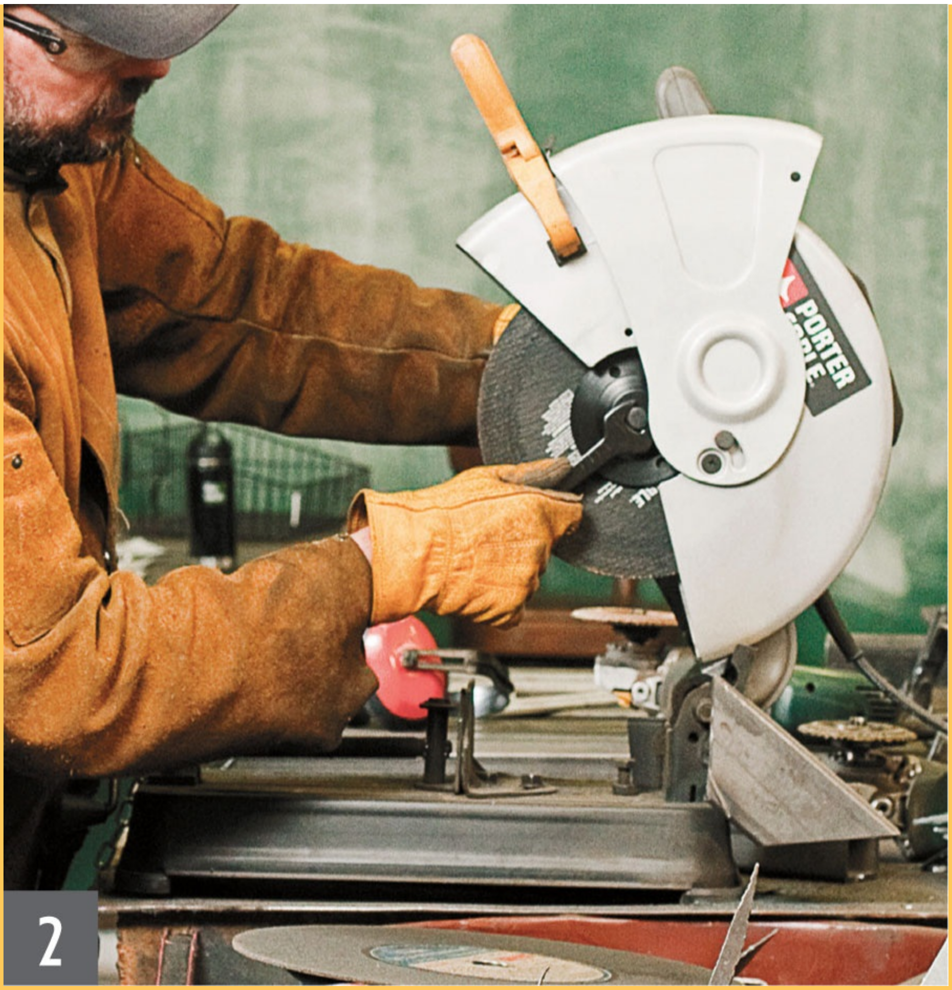

2

Hold down the arbor lock (located on the opposite side of the nut) and unscrew the nut with the wrench provided by the manufacturer.

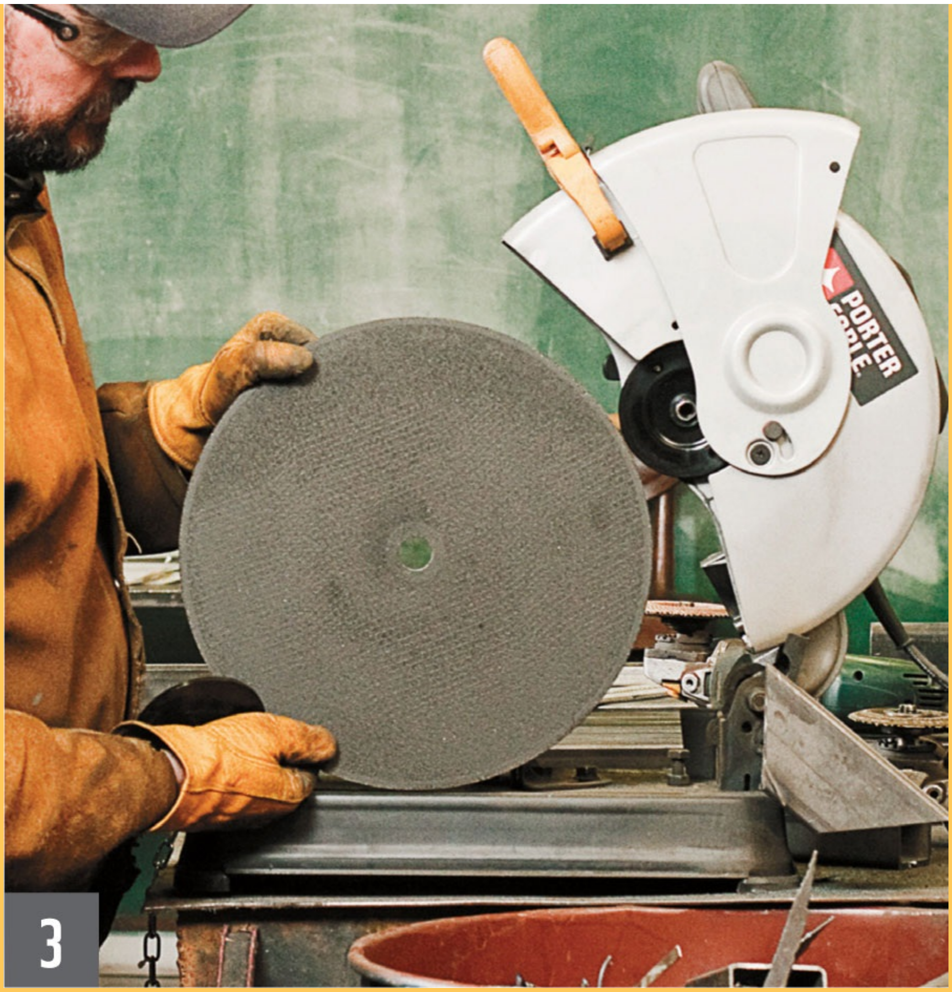
3
PORTER CABLE

Remove any metal debris or buildup from the inside of the guard before you replace the disk.

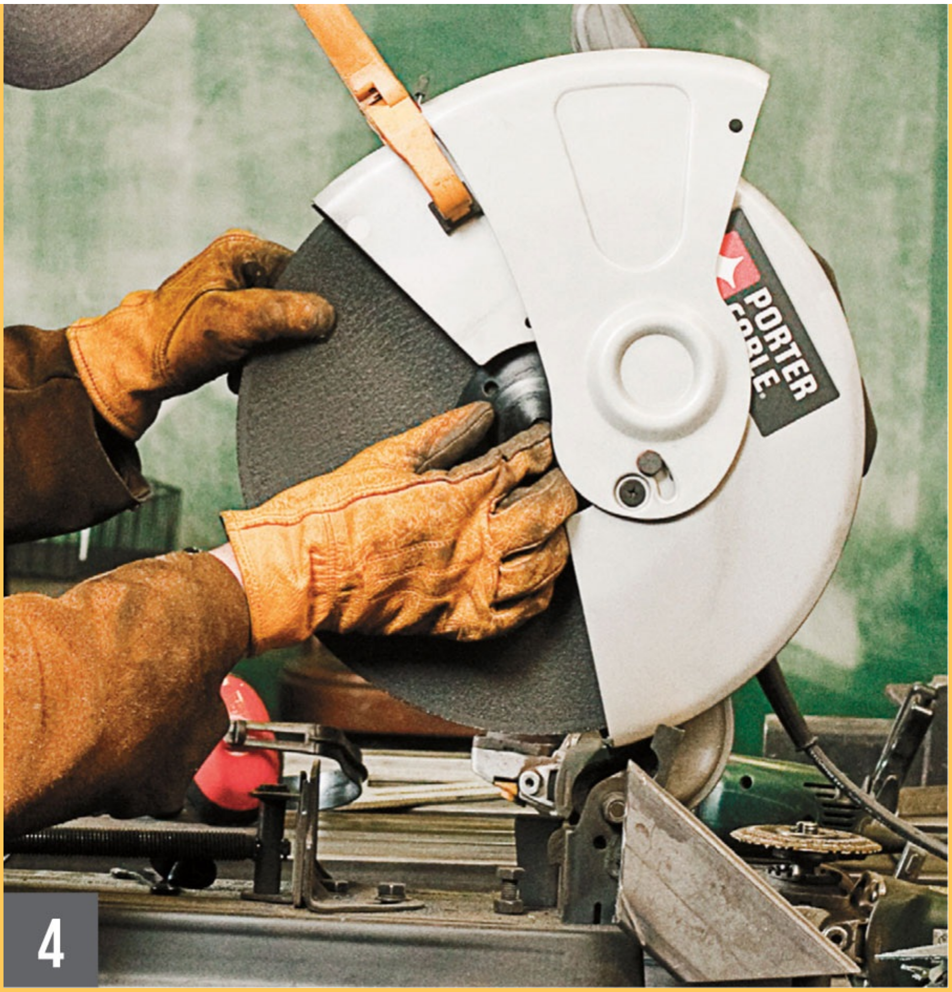
4
PORTER CABLE.

Before you replace the arbor nut, make sure the collar and flange assembly are lined up correctly with the arbor.

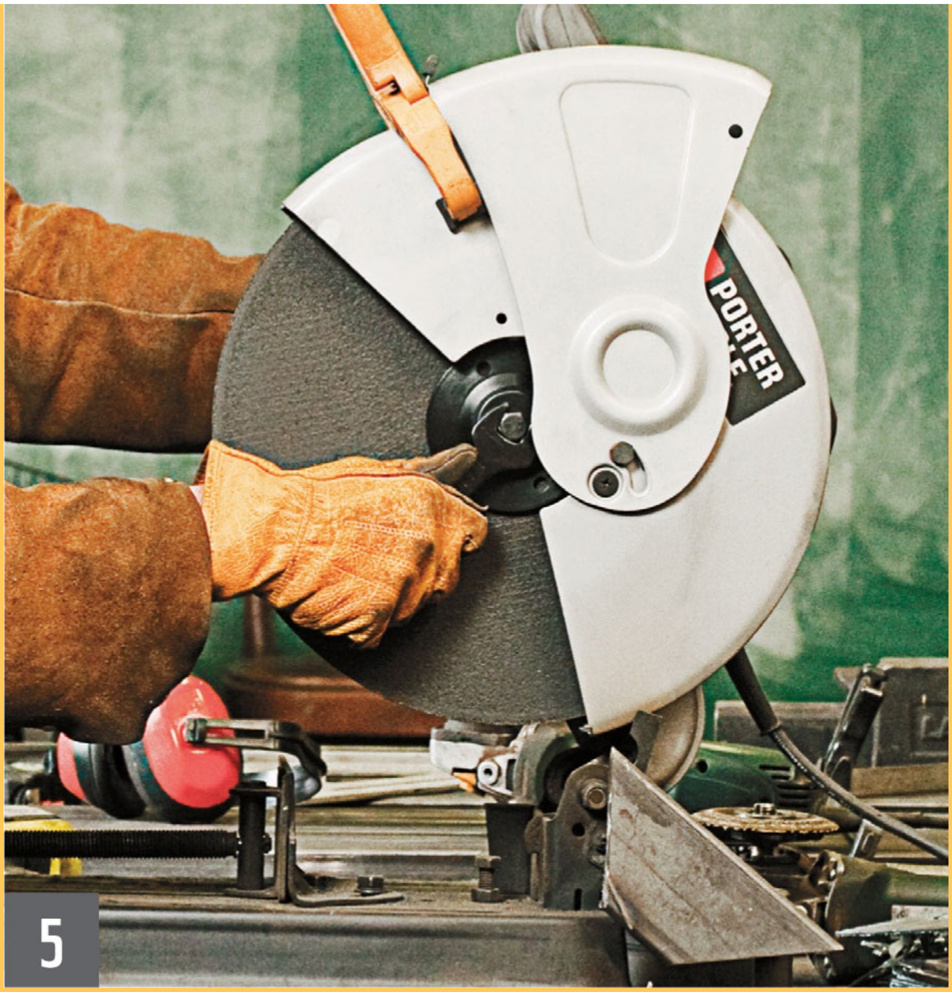

5

When tightening the nut, make sure the disk is flush to the back flange. If it is off, the disk can wobble when in motion.

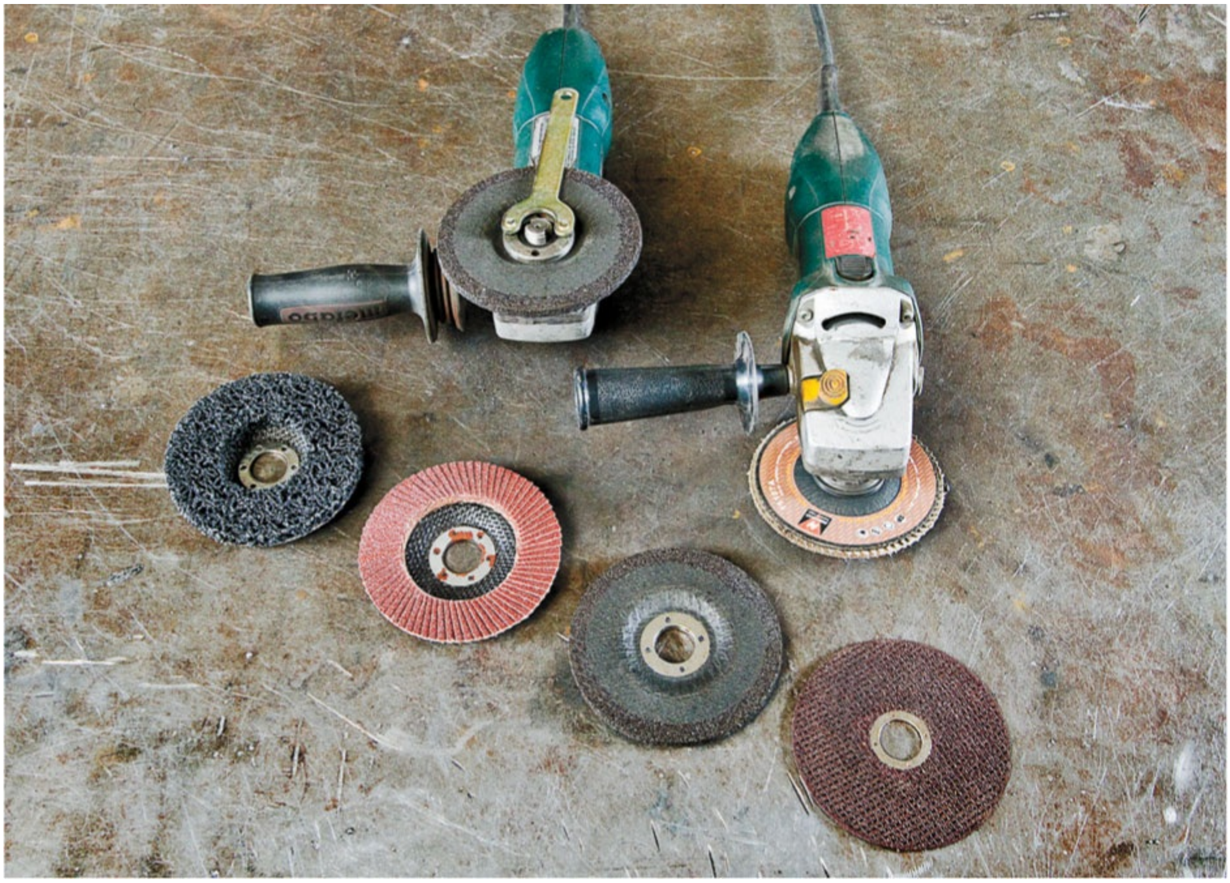

Grinders have a wide variety of disks and attachments for cutting, grinding, polishing, and finishing. Never place a grinder face down on a surface.

The Angle Grinder

The angle grinder is one of the most common tools in metalworking and one of the most diverse. The angle grinder can be used not only for grinding, but also cutting, sanding, shaping, and polishing. These versatile little machines have a wide variety of different disks for all of these purposes. They also come in many different sizes. We are going to be using a 4 1/2-inch (11.4 cm) angle grinder for cleaning and preparing our metal before welding and also to grind down some of our welds.

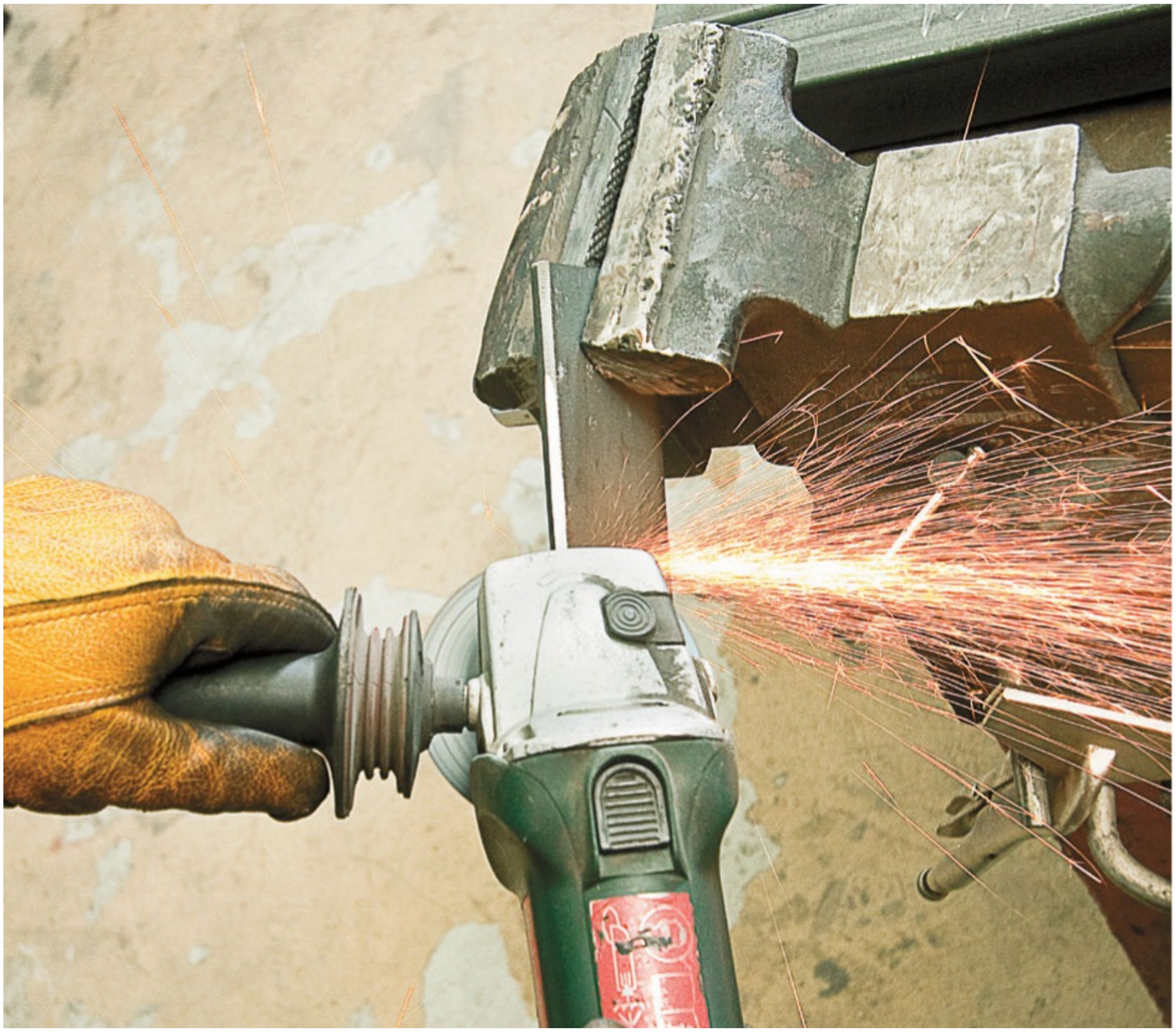

Sparks will always fly in a clockwise direction 90 degrees from the point of contact in relation to the arbor (center of the grinder head).

An angle grinder got its name for two reasons. The first is that this machine allows you to hit multiple angles in three-dimensional space, which allows for any number of angles for chamfers or bevels on a piece of base metal or to get into hard-to-reach areas. The second reason is that it is actually necessary to hold the grinder at an angle to prevent grinding on the arbor. If you are not at the proper angle to clear the arbor, you are just grinding metal-to-metal and possibly damaging the machine.

The disk on a grinder is always moving in a clockwise direction, normally indicated by an arrow at the top of the head. This is good to keep in mind, considering the majority of the sparks will fly in a clockwise direction 90 degrees from the point of contact in relation to the arbor. So, if you draw a line from the point of contact to the arbor and then draw a right angle to the right of the contact point, that will be where most of the sparks will go. A small amount of the sparks will follow the wheel. Knowing this will allow you to direct the sparks away from people or things that have the potential to catch fire.

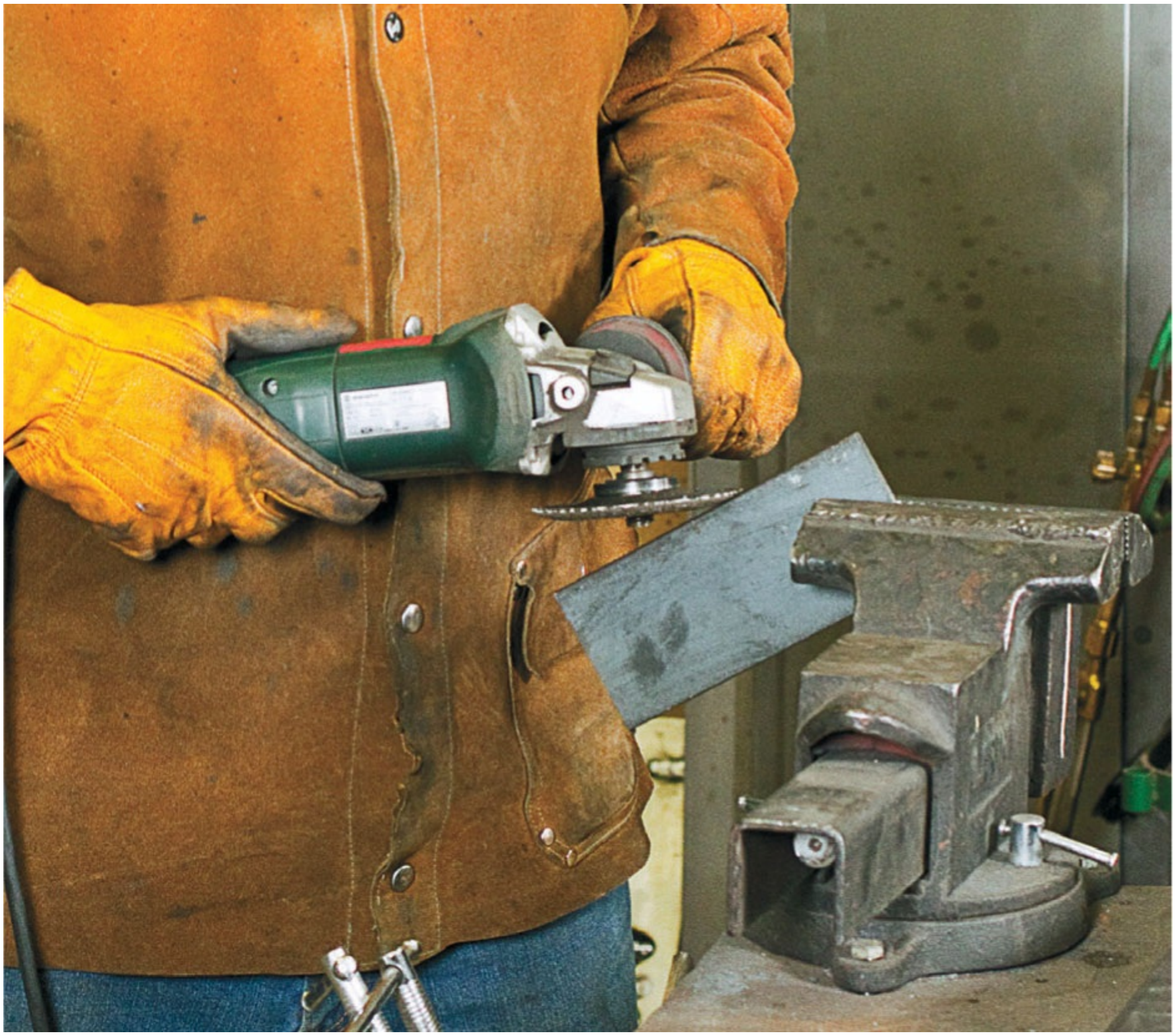

Angle grinders got their name because their exposed wheel allows you to hit pretty much any angle. The grinder needs to be held at a minimum of 15 to 30 degrees to keep from grinding on the arbor or arbor nut.

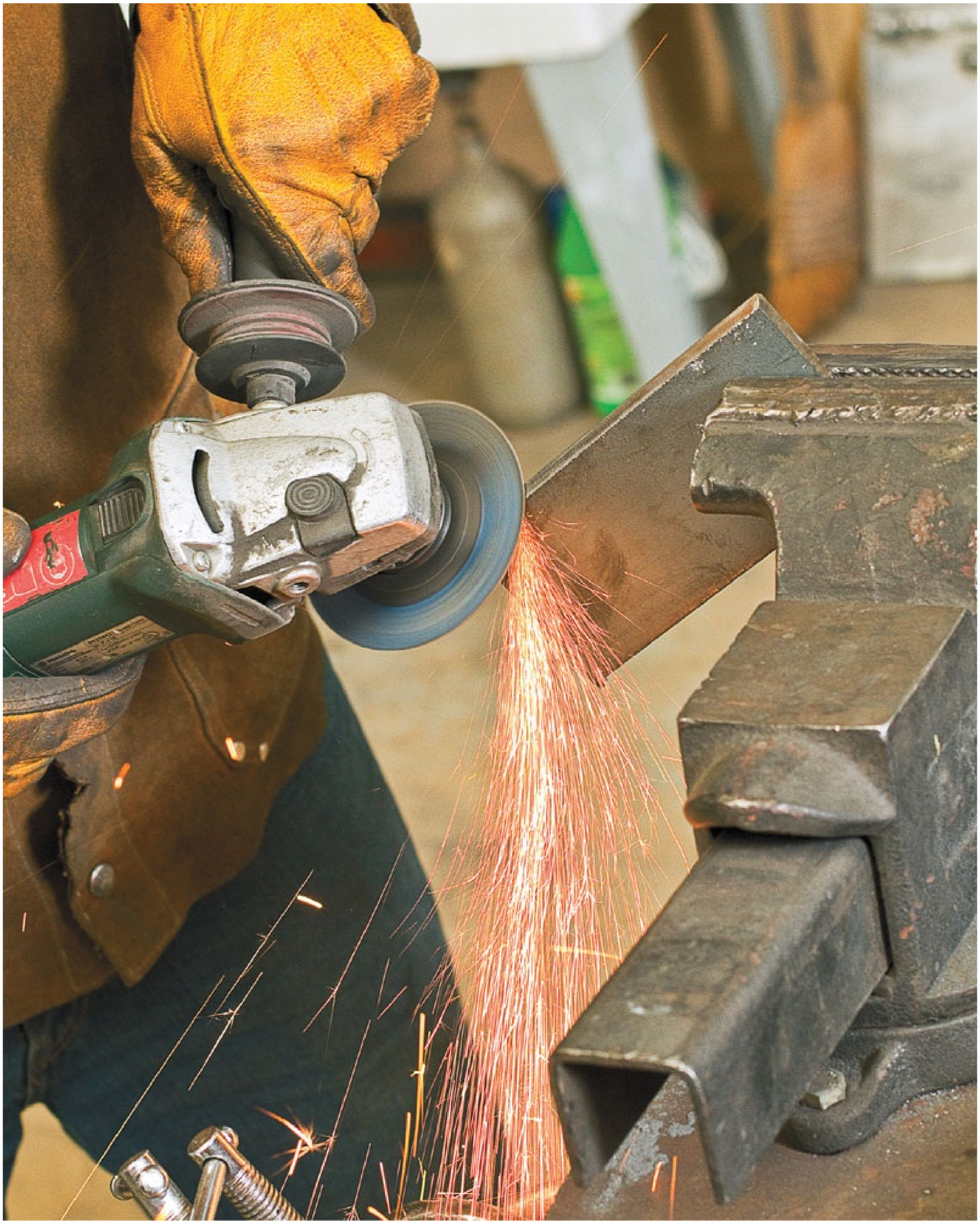

Keep a firm grip on the handle and body of the grinder. Keep your angles consistent. And as with all power tools, never take your eyes off of what you are doing.

Grinder Safety Tips

- On the top (or back) of the grinder there is a flat landing right where the head and block meet. When you are done grinding, *always* place the grinder upside down on this flat landing; this is what it is designed for.

- Never put a grinder face down on its belly. If you place it face down while the machine is still on, it can speed away and hurt someone or cause damage or speed towards you and hurt you. This can also happen if you accidentally turn the device back on while it is face down. Setting it down on its belly can also crack the disk. What happens when you turn it back on? The disk blows up, throwing shrapnel chunks of grinding disk everywhere, hurting you and anyone else in the room, and causing damage to the machine and anything else in the path of the debris.

- Always use the blade guards. Not using the guard is also more than likely against the safety guidelines in the manufacturer's owner manual. Without guards, angle grinders will spray more sparks in more directions, so you make up your own mind. You have been warned.

- While the grinder is not in use and properly at rest on its back, you can inspect the blade of a consumable disk for any damage or cracks that may lead to the shrapnel scenario mentioned above. This device is very easy to use but if you don't respect it, it can do a lot of damage. Remember to wear all safety gear and follow all the guidelines set by the manufacturer.

all safety gear and follow all the guidelines set by the manufacturer.

- Wearing gloves with anti-vibration liners helps keep your hands from going numb after long periods of grinding.

It is never a good idea to point your fingers when using an angle grinder. I know that may sound obvious, but sometimes when people re-grip the grinder they aren’t thinking of where the blade is in relation to their hands and off comes a finger. Another good practice for keeping your fingers intact is to never move your hand past the block. The block is the bulky part of the body near the head. This acts as a stopper to prevent you from putting your hand on the moving disk. If you accidentally jam a finger under this part of the grinder, it will quickly sand through your glove and burn you or remove your index finger.

Be aware of your hands and the position of the disk at all times. Keeping your fingers intact is a good thing.

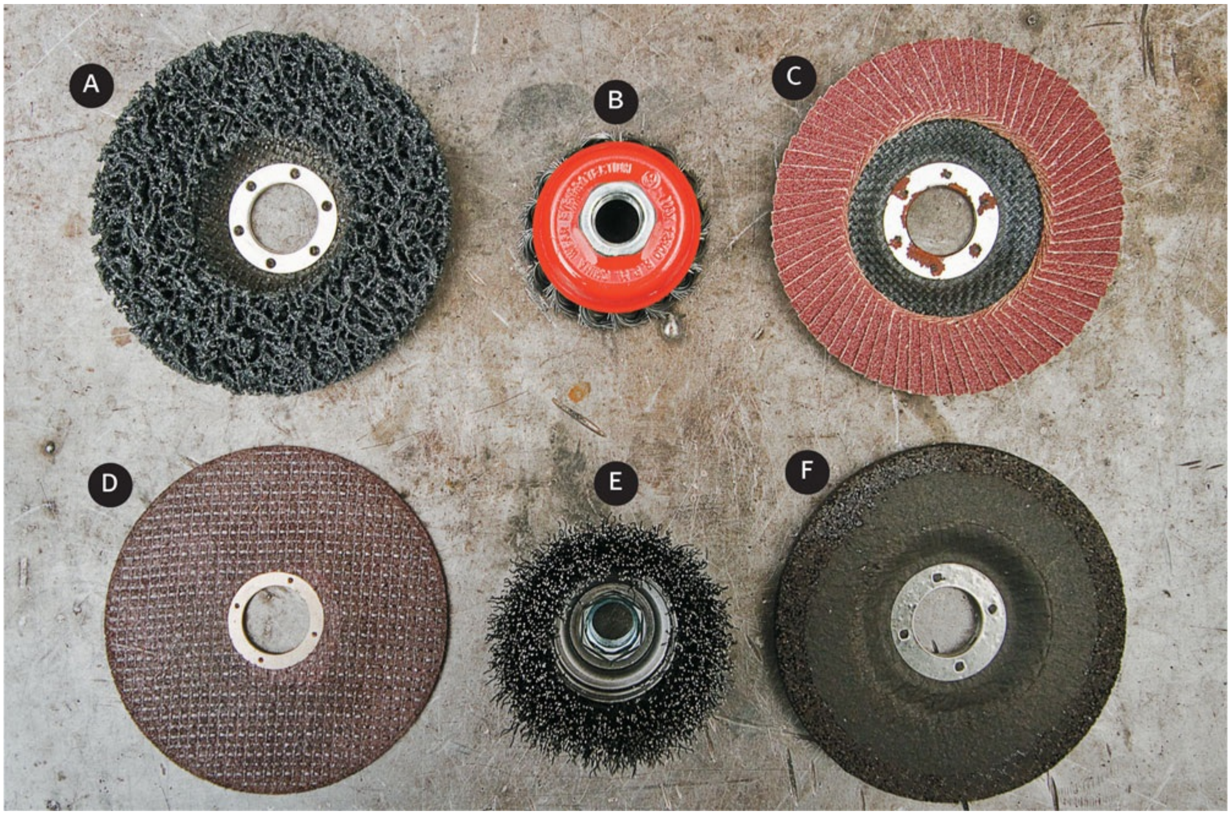

Here are some of the commonly used attachments for an angle grinder: A. Synthetic rust and paint removal disk ; B. Braided cup brush; C. Flap-disk (available in a wide range of grits); D. Aluminum oxide cutting disc; E. Fine wire brush; F. Grinding or stone disk

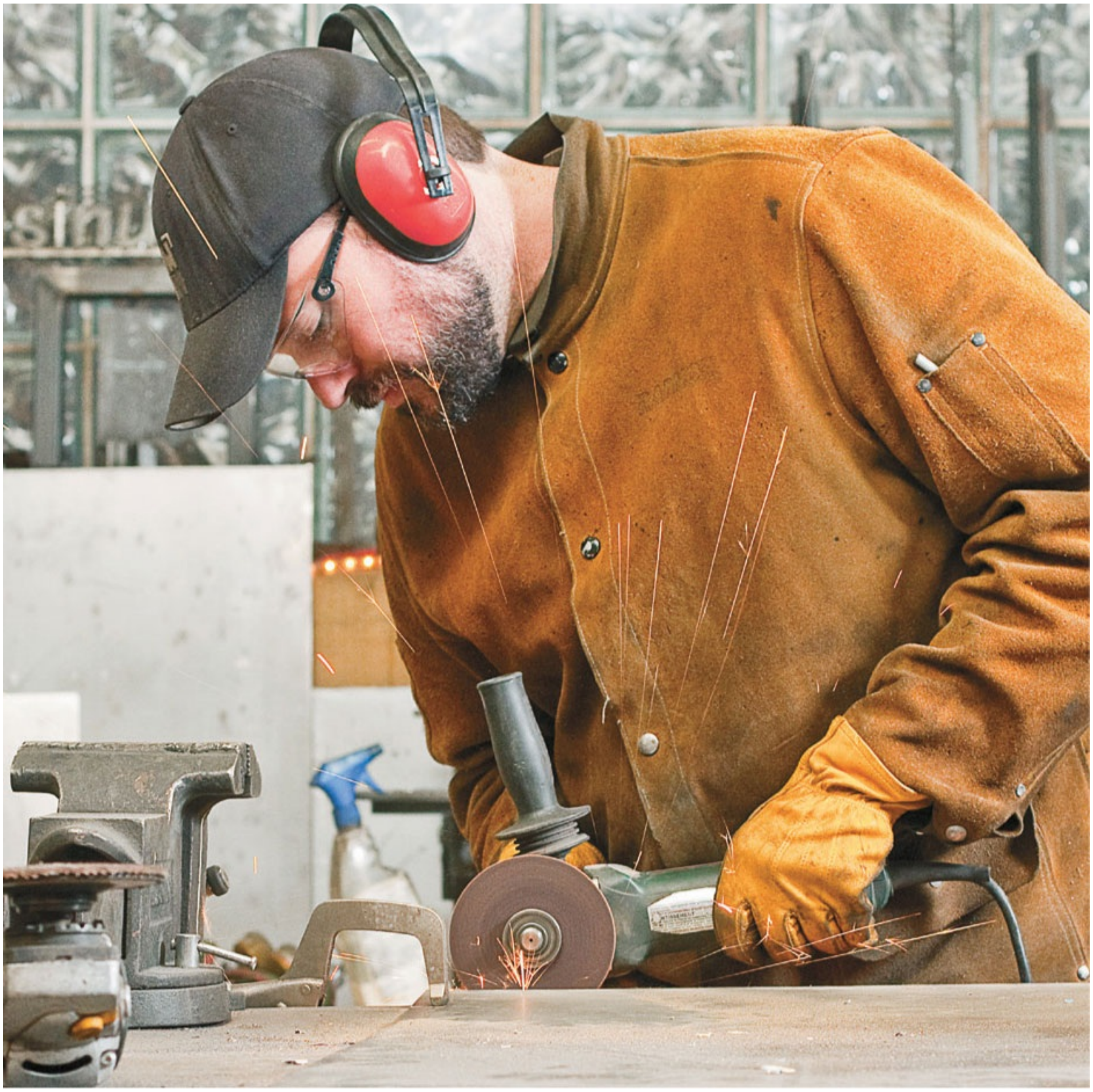

Using a grinder with a cutting disk will allow for straighter cuts. Keep the disk straight into the cut or it could bind or kick.

Don't apply too much pressure; it's best to let the grinder do the work. Keep the cut point in the bottom front quarter of the disk.

Setting up the Grinder

The setup process is very similar to the chop saw. It also has an arbor lock and wrench that you will use to release the arbor nut. Hold down the lock while using the flange wrench to remove the nut. Leaving the lower flange on the arbor, place the disk on the arbor with the correct side up. Replace and tighten down the nut.

There are many different types of disk that can be used. The three types that we will be using most throughout this book are the grinding disk, cut-off wheel, and the flap disk. There are different types of disks for different metals. For and the flap disk. There are different types of disks for different metals. For carbon steel, an abrasive should be an aluminum oxide type disk. There is no need to break the bank on zirconia, zirconia alumina, or other composites if you’ll just be grinding on mild steel.

Metal sanding and grinding grit is the exact same as the woodworking scale of abrasives: the lower the number, the larger the granules. For example, a grinding disk, or stone as it is sometimes referred to, is normally a very low number grit, around 18. This is a very large grain that will remove a greater amount of material very rapidly. The finish with lower grit abrasives will be coarse. A 120-grit flap disk will remove material more slowly but have a much smoother finish. The higher the number, the smoother and shinier the metal becomes. You can bring metals up to a mirror shine by increasing the grit number and method of grinding and polishing.

A good disk to use for metal prep is a 60-to 80-grit. This will clean off all the grease, rust, dust, dirt, oil, and mill scale from the steel to prepare it for welding.

For the projects that we are doing in this book, most of the cutting for the tubes and bars will be done with a chop saw. Sheet metal will be cut with a grinder using a cutting disk.

When cutting with a grinder, always securely clamp the material. During cutting, make sure the disk spins into the direction of the cut. Always use the bottom front quarter section of the disk; if the disk starts cutting more toward the top of the disk, it will want to kick. Let the blade do the work; it doesn’t require a lot of pressure to cut.

Have everything you need out and ready to go.

The Welder

Most MIG welders that a hobbyist or beginner welder would buy are all pretty much the same. The standard housing contains the drive motor for the wire feeder, the power source, and other components that help control the machine. The components that you will need the most understanding of are the day-to-day maintenance parts, such as the MIG gun, or weekly and monthly maintenance, such as how to load the wire spool.

The gun, or torch as it is called, is the handle piece of the equipment where all three components for welding exit to create a weld. I sometimes call it the stick. Normally, the assembly for the gun is pretty straightforward. The gun consists of the gun handle, weld trigger, contact tip, nozzle, and the shielding gas diffuser ports.

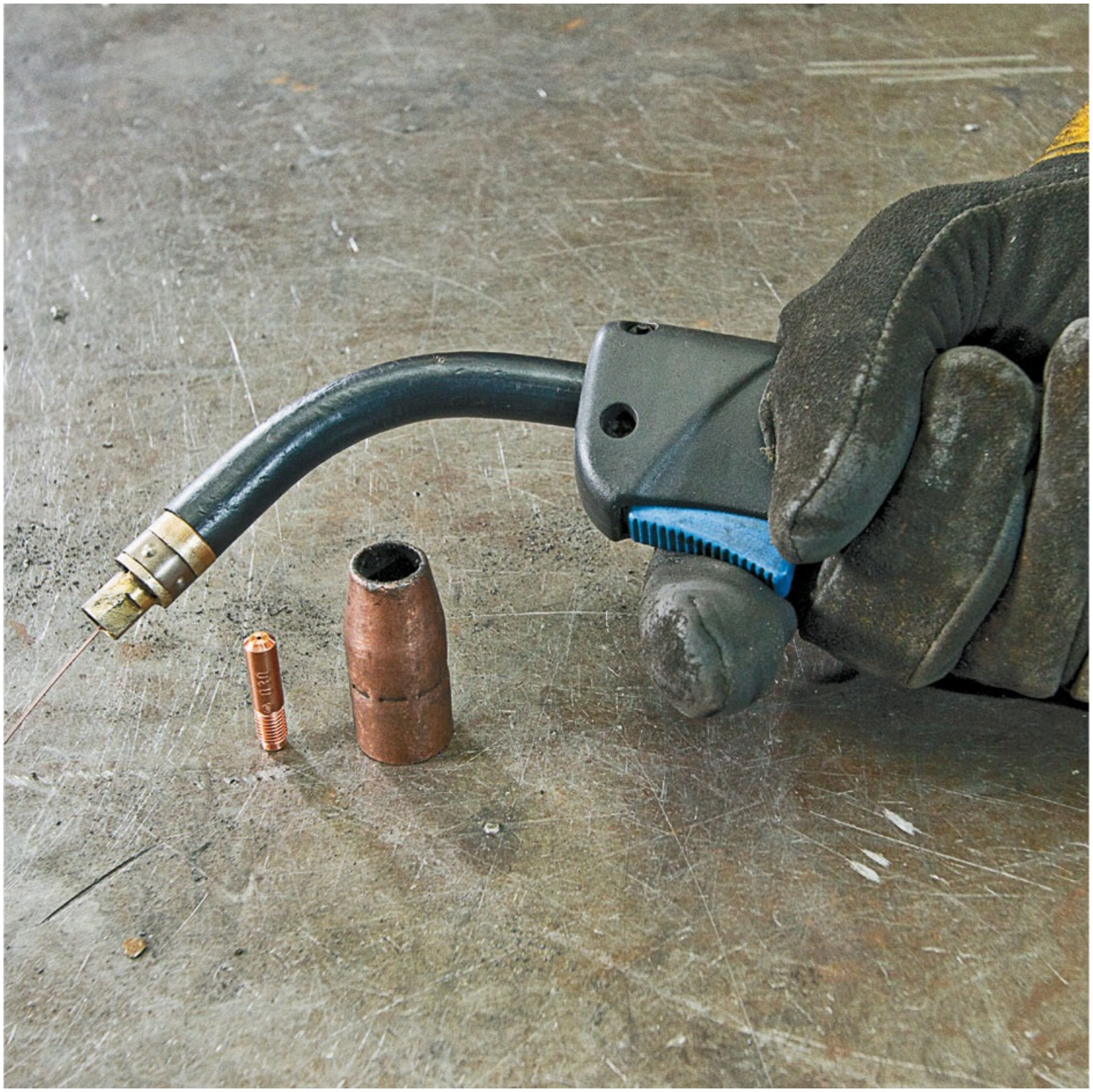

This is the contact tip and nozzle for the gun. These are consumable products and will need to be changed periodically.

All of the elements of welding initiate at the same time when the welding gun trigger is pressed and held down. The filler wire will feed through the tip, the arc will form at the tip of the filler wire (which is also the electrode), and the shielding gas will flow from the diffuser ports to create a pure environment for the arc to exist. The nozzle directs the shield gas over the tip and then over the arc.

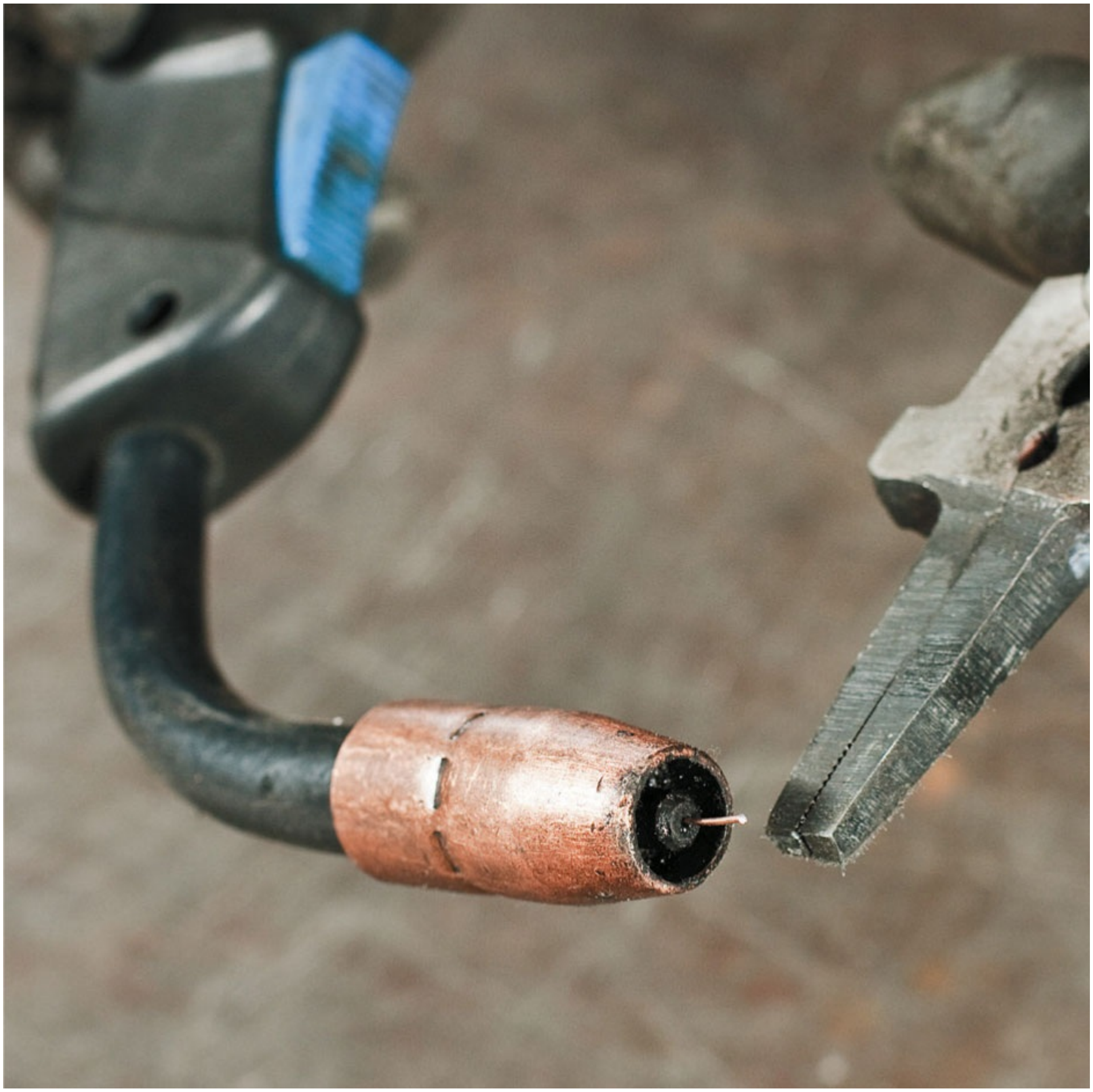

The distance the filler wire comes out from the contact tip is referred to as the “stick out.” The stick out should always be 3/8 inch (9 mm) while welding.

Right out of the box this gun is simple to set up. Most guns will only have two pieces that you will need to attach, the tip and the nozzle. Screw the tip in to the threaded adapter opening at the face of the shielding gas ports. Make sure that the tip size is the same as the welding wire used. For example: if the wire is 0.030 inches (0.762 mm) in diameter, make sure that the tip is clearly marked with a 0.030 inch (0.762 mm) stamp. You will also have to make sure the threading of the tip matches the gun (this will vary depending on the manufacture). After the tip is secure, connect the nozzle over the tip and onto the gun neck. Nozzles can be a screwed in or have a pressure fitting, also depending on the manufacturer. The tip should be even with the nozzle’s face. As you develop a better understanding of the MIG process, you may prefer to have the contact tip sticking out of the nozzle face about 1/8 inch (3 mm) or so. This is fine, but don’t go too much past that distance.

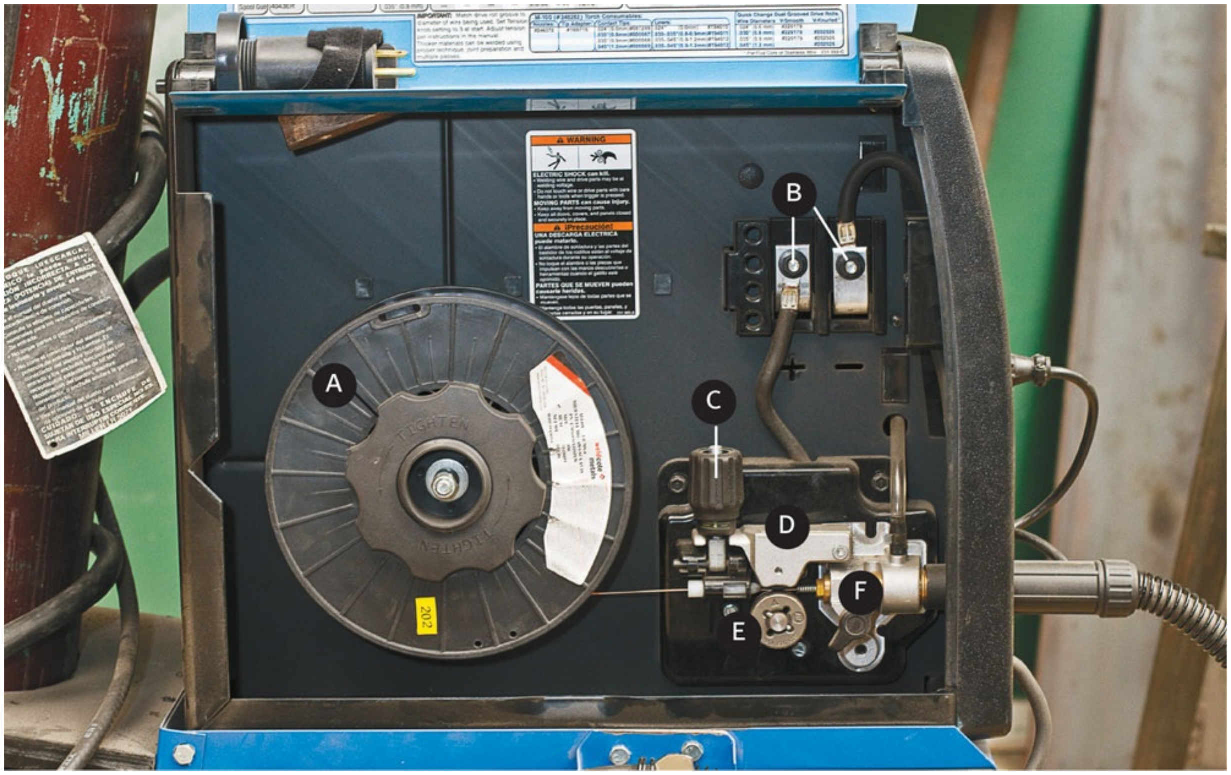

This is the inside of a fully assembled MIG welder. A. Spool of filler wire; B. Positive and negative terminals; C. Tension adjustment knob; D. Tension arm; E. Drive wheel; F. Opening to gun liner.

Now that the gun is assembled, it is time to connect the hose and ground. Connect the hose first by pushing the feed connector end of the hose into the receiving port on the housing. Make sure the fitting is all the way in. If the connector is not in the port securely, the machine will not have enough of a seal for the shield gas hose. After this has been secured, tighten the screw that fixes the connector into place.

The ground normally will come already connected to the machine. If it is not, feed the ground through the ground port receiver and secure the ground cable to the ground terminal. This is also a good time to make sure the machine is set to the correct current. Solid wire carbon steel welding requires a DC+ connection. This is an adjustment that would need to be made to the positive (+) and negative (–) terminals, normally right above the wire feed drive assembly (unless the welder is a top feed machine in which case it would be below). Sometimes, if you purchase a used MIG welder, it may be set for flux-core welding (which is DC–). This can cause problems when welding with solid wire. Other steps may need to be taken to ensure the uses of the proper voltage such as switching the positive and negative terminals. Smaller machines will normally be single-phase 120 V or 230 V; check with the manufacturer to make sure the unit is connected properly.

Before cutting the wire away from the spool, make sure to secure it with your finger and hold it down to prevent it from unraveling.

You’re almost there! It’s time to feed the wire through the wire feed assembly, gun liner, and gun. Before doing so, inspect the wire feed roller to make sure the groove is correct for the wire size. The feed roller can have multiple grooves for different diameter wires. If the groove does not correspond to the wire, it can have the wrong effect; too big or too small and the roller will not have enough bite and will slip over the wire.

There are two types of wire that are commonly used for mild-steel welding. ER70S-3 is an all-purpose type of filler wire that is great for everyday use, and ER70S-6 wire is higher in deoxidizers for welding on rusty or dirty steel. Most of the wire sold in home improvement stores is ER70S-6.

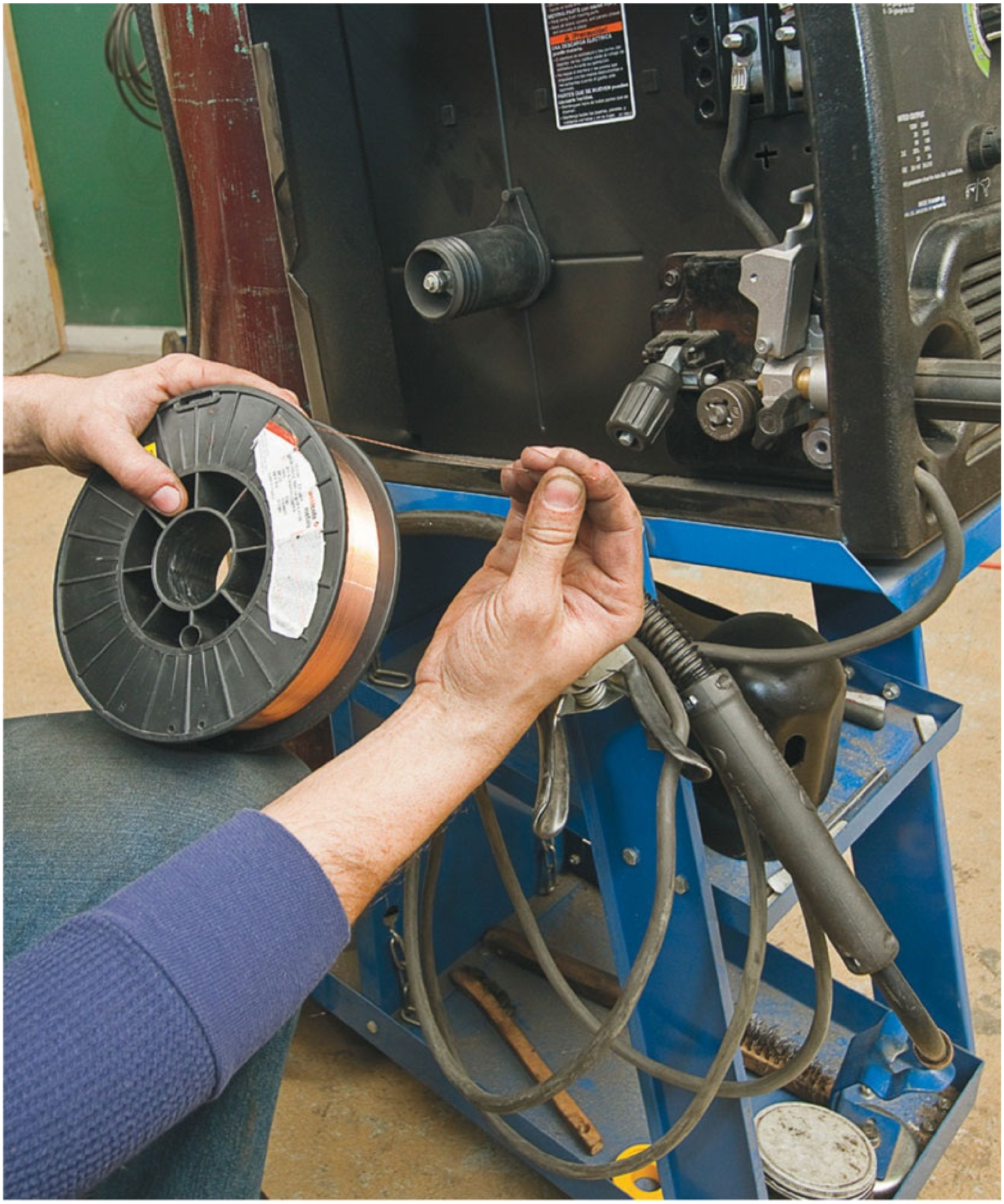

It helps to straighten the wire before you feed it.

Release the tension knob and lift the tension arm up. Now listen, this is important. This is also something most handbooks don't tell you. The spool of mild steel filler wire is essentially a spring. As soon as you cut the wire away from the spool it's going to want to unravel. Before or after putting the spool into place, make sure to hold down the wire and hold the spool in place with one hand. The wire needs to be secure at all times. If it is not, it will spring out quickly and unwind. If this does happen, you will spend a lot of time winding it back onto the spool or cutting a lot of wire away (which is wasteful).

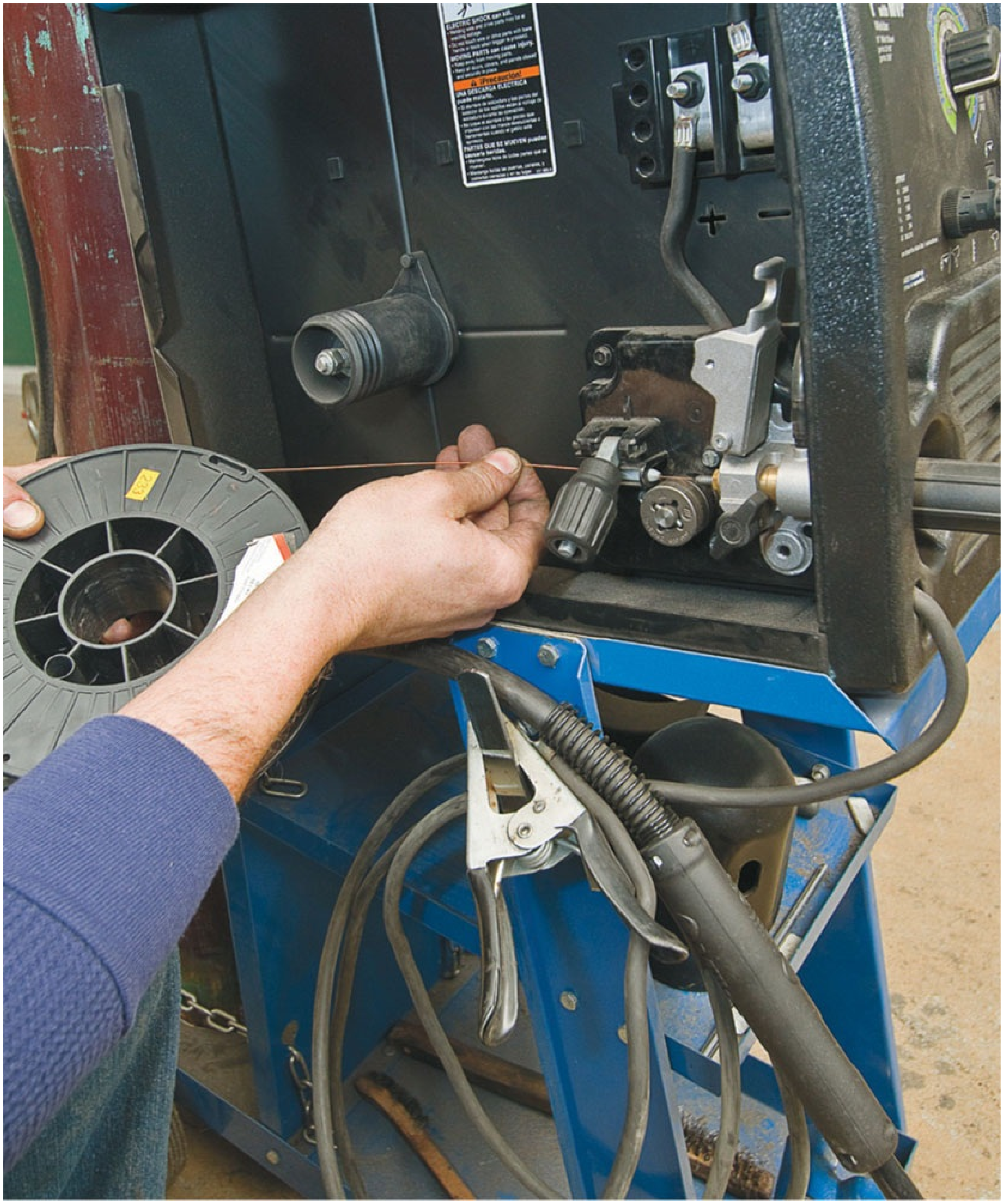

Feed the wire through the guide, over the drive wheel, and into the gun liner opening. Feed the wire a few inches into the gun liner.

Hand-cut the wire away from the spool with a pair of wire snips. Once you have cut the end away from the spool, grab the wire securely and feed the wire through the wire feed drive assembly. You may need to straighten the wire before feeding.

Feed a few inches (5 to 8 cm) into the gun liner (hose) and lock down the tension arm; then turn on the machine. Set the wire feed to about 30 or 40 percent, which is normally about a number 4 to 7 setting on the wire feed dial. Pull the gun trigger to allow the filler wire to feed all the way though the hose. Keep the gas off and the voltage turned all the way down. Make sure the gun's contact tip is away from the ground to prevent unwanted arcing.

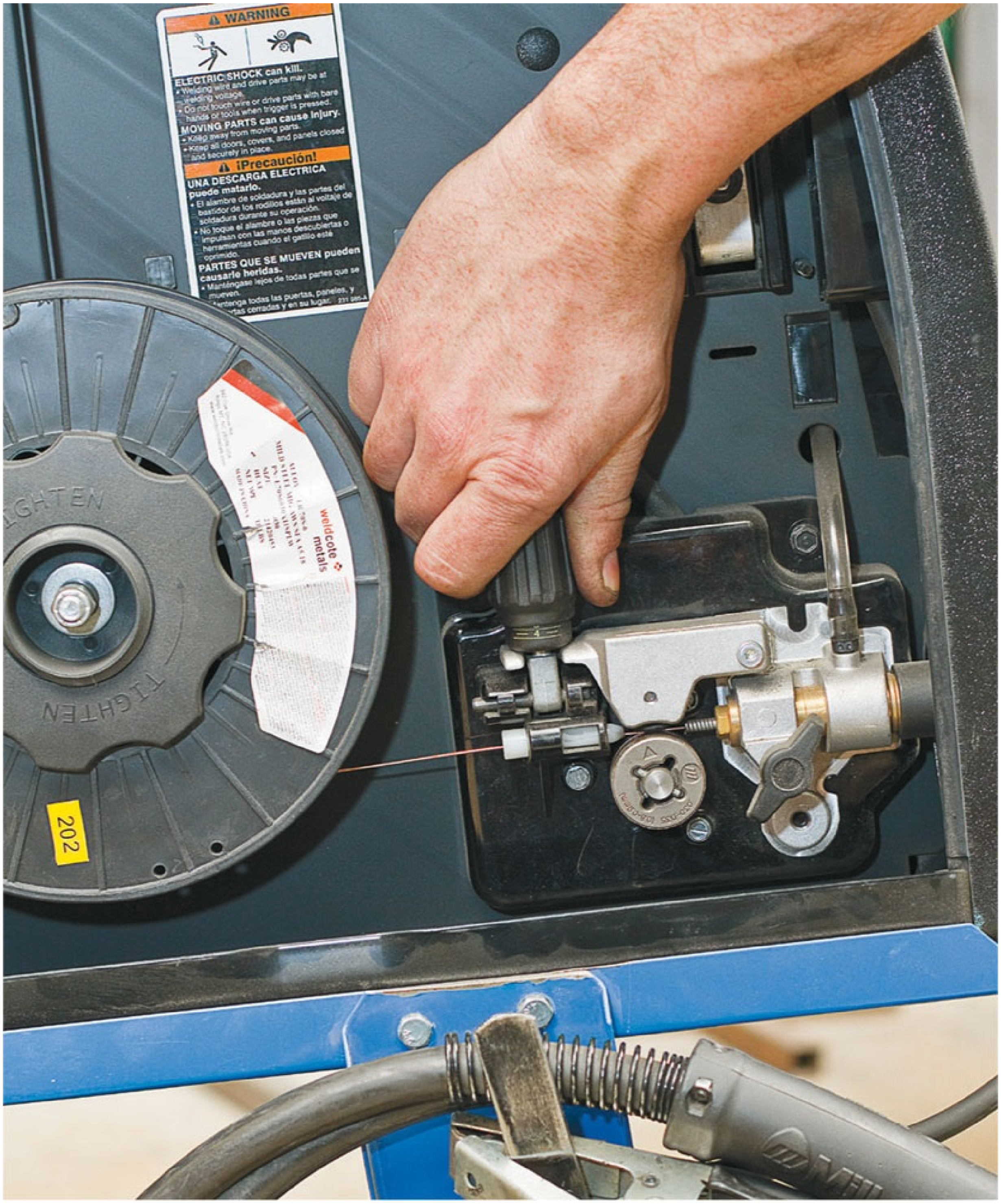
WARNING
ELECTRIC SHOCK can kill.
MOVING PARTS can cause injury.
¡Precaución!
UNA DESCARGA ELECTRICA puede matarlo.
PARTES QUE SE MUEVEN pueden causarle heridas.
TIGHTEN
TIGHTEN
weldcote metals
202

Lock down the tension arm with the tension adjustment knob. Make sure to hold down the wire at all times while putting the spool into place and tightening the screw.

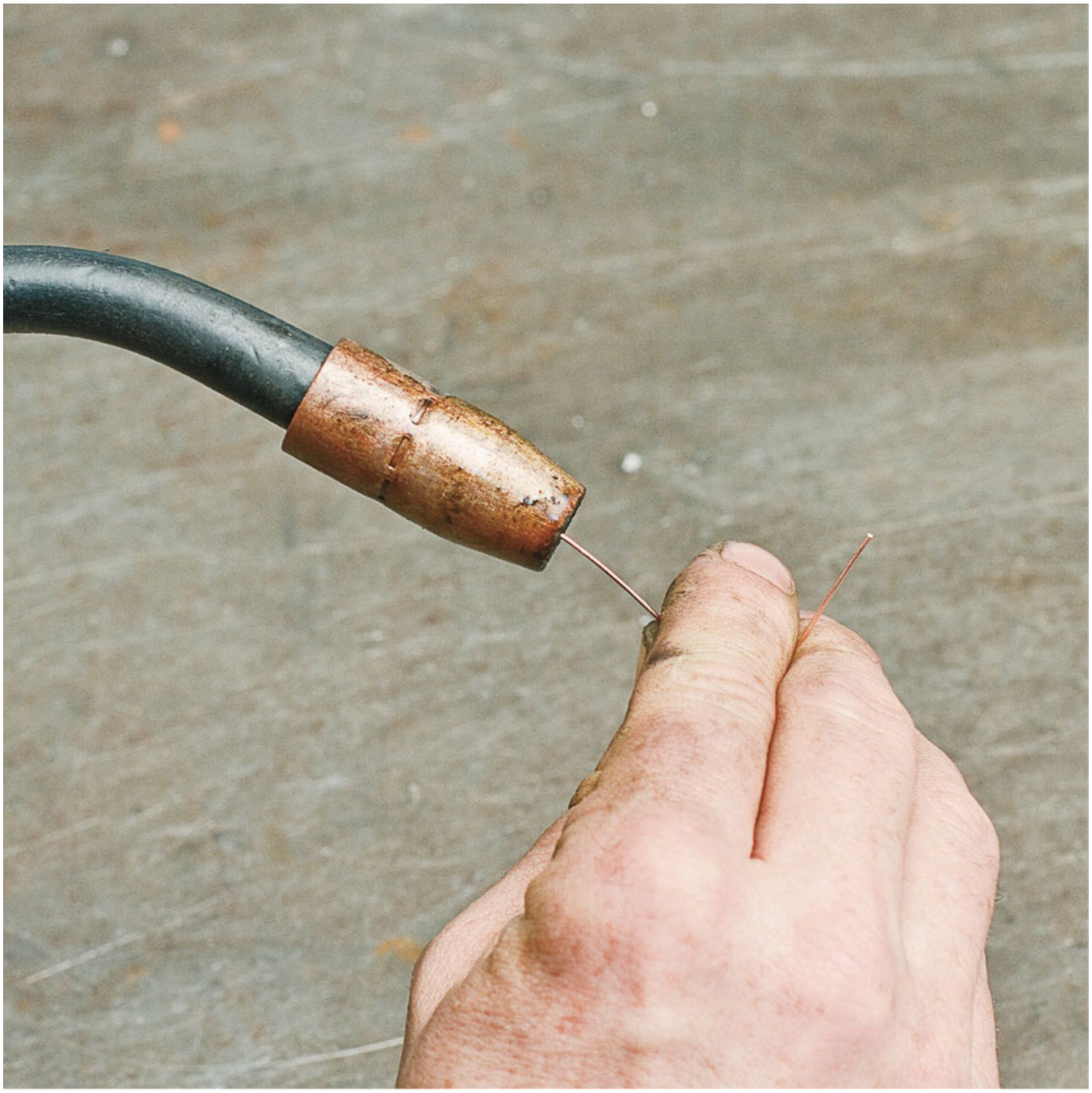

To set the tension properly, put a 90 degree bend in the wire so it points outwards and mirrors the gun neck.

Once you see the wire come through the tip, release the trigger. You may need to make an adjustment to the tension using the tension knob. Too much tension will overwork the motor, causing it to feed improperly; not enough tension will cause the wire to slip or not allow the wire to feed at all.

It’s possible that the contact tip could obstruct the wire from feeding all the way, and it may be necessary to remove the contact tip from the gun in order to feed the wire. After removing, press the trigger until the wire is exposed; then feed the tip over the wire and secure the tip back into position. Turn the welder off while re-securing the contact tip.

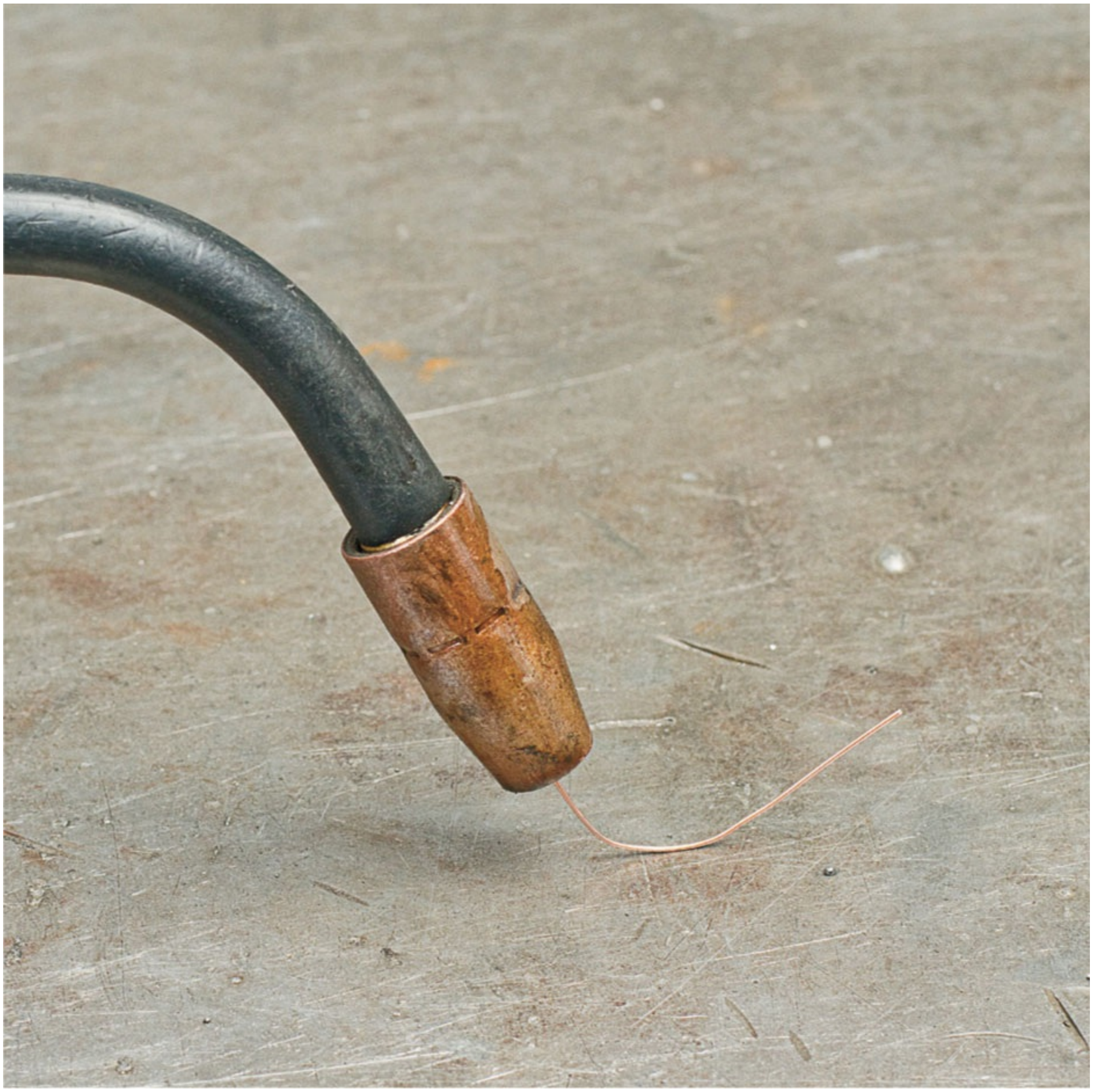

Hold the nozzle 1 inch (2.5 cm) from an ungrounded surface, with the bend in the wire touching the surface. Double check to make sure the ground clamp is not attached to your surface.

To set a good tension for the wire feeder, first set the tension knob to zero. Set the wire speed for half the capacity of the machine's wire speed output; for example, if the machine has a maximum 100-feet (30.5 m)-per-hour setting, set the wire speed for 50 feet (15.2 m) per hour. That should be a good average wire speed to properly set the welder. Feed 2 inches (5 cm) of wire out of the gun. Using your finger, put a 90-degree upward bend in the wire. At this time you do not need the gas or the voltage on. Turn the gas valve off and the voltage all the way down.

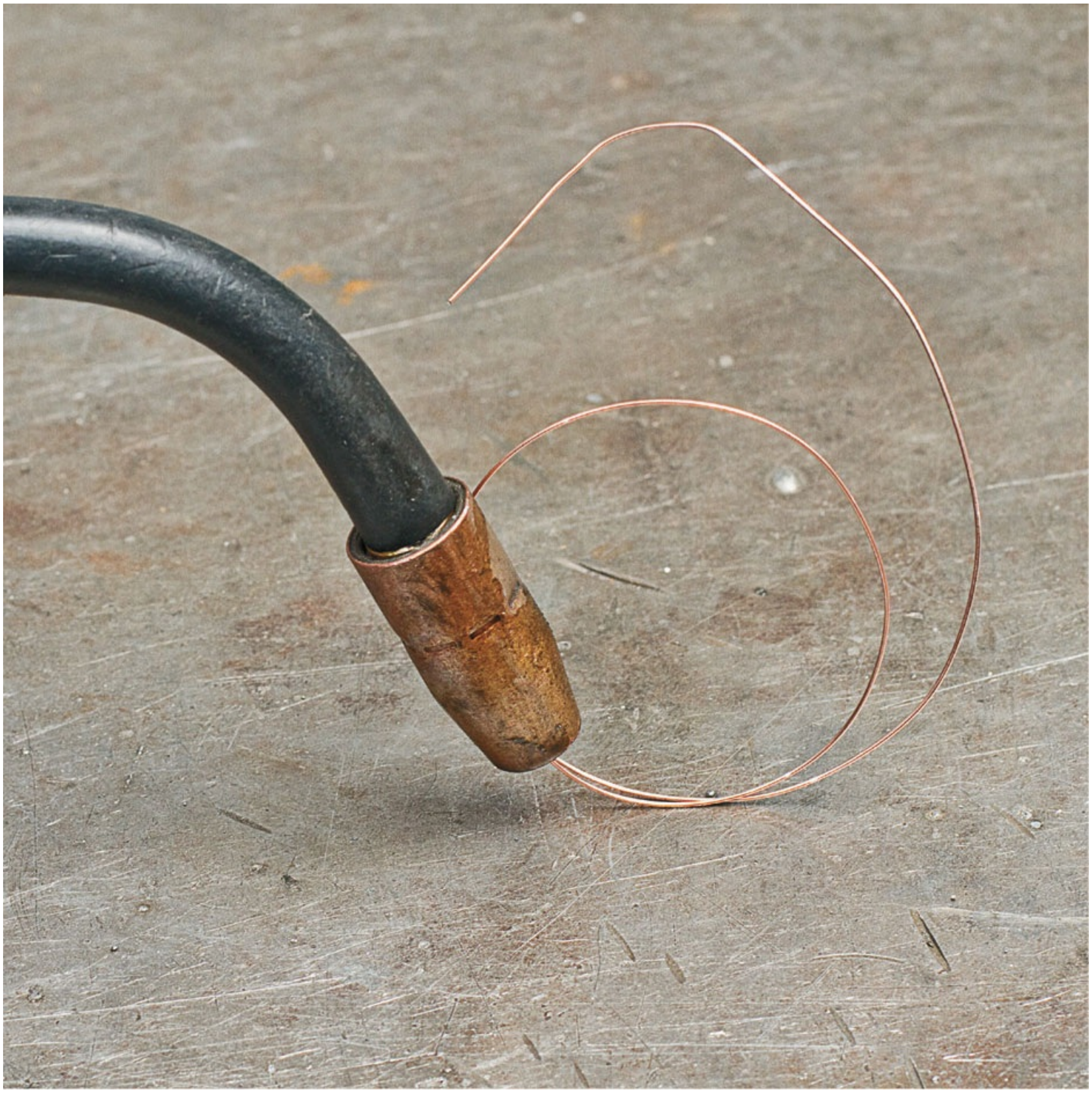

Press and hold down the gun trigger. Adjust the tension until the wire can form a 3 inch (7.6 cm) loop with no problem.

Place the bend of the wire onto an ungrounded surface, holding the gun tip 1 inch (2.5 cm) from the surface at a 30-degree angle (or at an angle where the gun handle is parallel to the surface being used). Again make sure the ground clamp is away from the gun and not clamped to the surface of the table or use a piece of wood for the surface test. We do not want an arc! Pull the trigger. If the wire does not feed, release the trigger and add tension with the tension knob. Repeat this process until the wire is feeding smoothly and not getting hung up. This may take a few times to get a good consistent tension. The wire should flow freely and also have the ability to create tight loops about 3 inches (7.6 cm) in diameter and also have the ability to create tight loops about 3 inches (7.6 cm) in diameter with the wire.

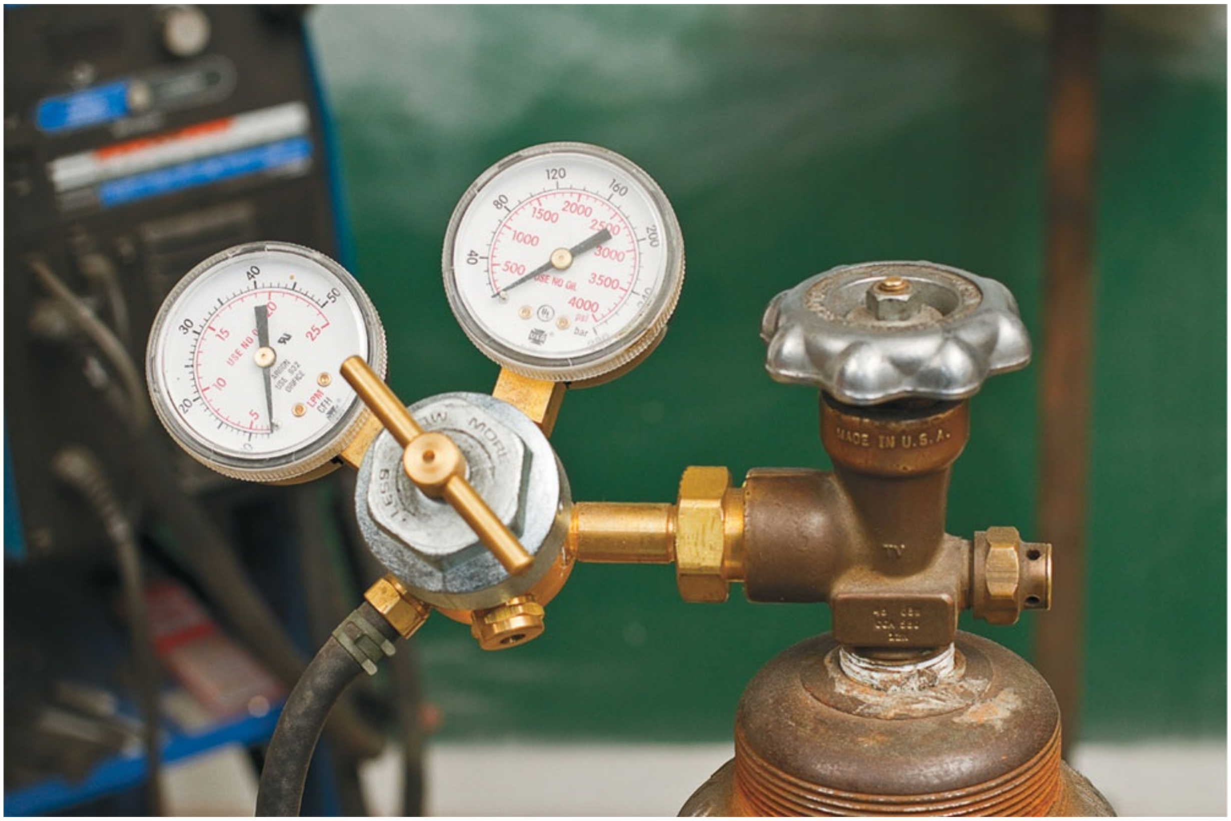

Pound-force per square inch (Lbf/in^2) is also commonly referred to as pounds per square inch (psi) or kilogram-force per square centimeter (kgf/cm^2) and is always the pressure of a compressed gas. Cubic feet per hour (CFH), or cubic meters per hour (CMH), is always the flow of the pressure going through the regulator.

Next is connecting the shield gas cylinder to the machine. Before screwing in the regulators, open the cylinder knob for a second to knock out any dust or debris that my have collected inside the fitting. Use Teflon tape on the regulator threads to seal the fitting before screwing it into place. Don’t overtighten this fitting; you want a fit snug enough to create an effective seal while also allowing you to release the fitting when the tank is spent. Repeat this process for the fitting on the machine. It should be snug but not tightened down too much. Open the cylinder knob all the way and then turn back one half to one full turn. This will ensure the shielding gas is flowing through the regulator smoothly. To check the seal, use a cup of soapy water and a small paintbrush or a spray bottle to coat over the connections. If tiny gas bubbles form in the soapy water on the connection, repeat the process until the fittings are completely sealed. I prefer to orient the regulator gauge faces at an upward angle so I have clear vision of my pressure gauges. After everything is set up, wipe off the soapy water. The machine is now ready for use.

machine is now ready for use.

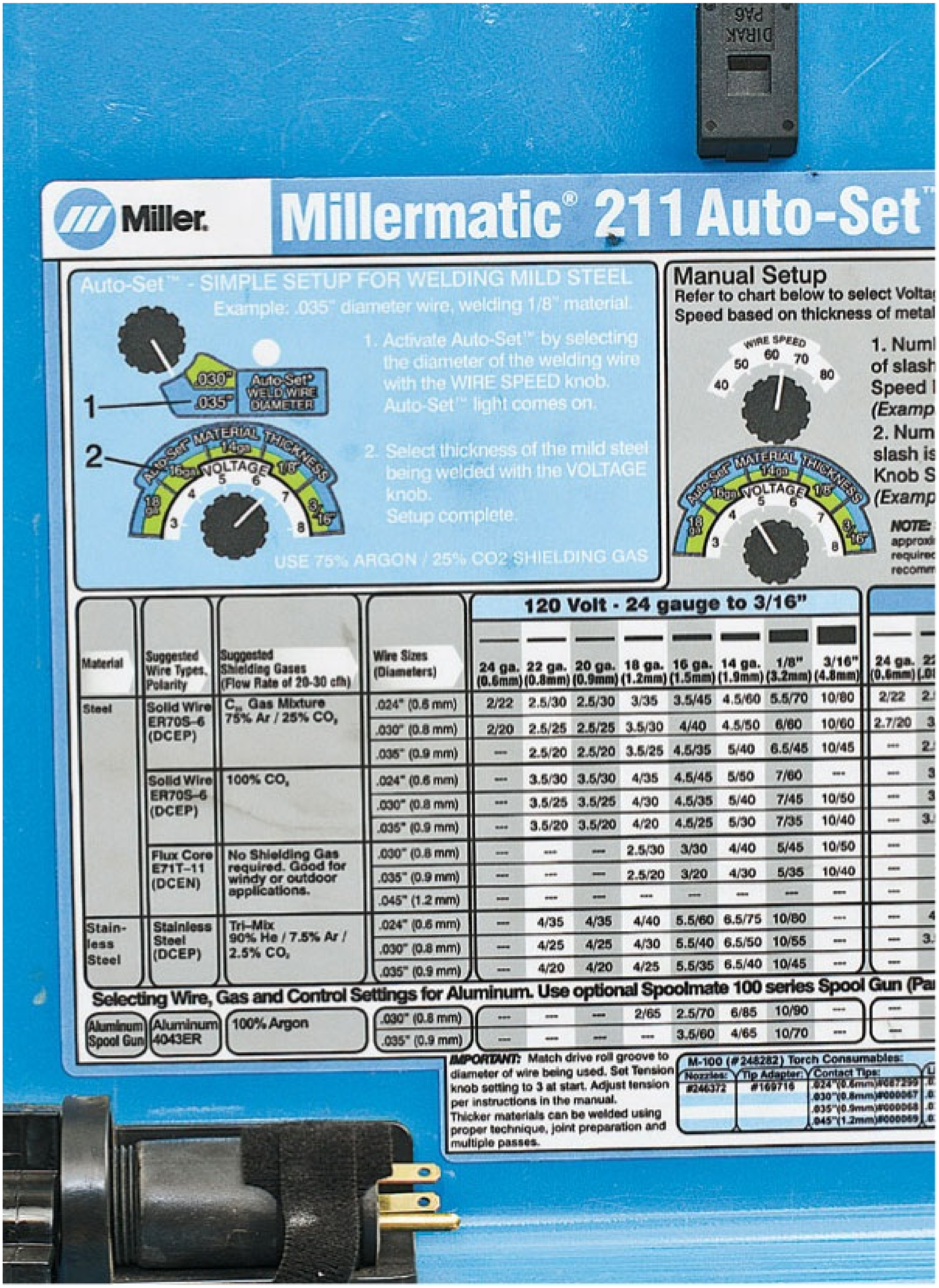

Miller. Millermatic® 211 Auto-Set

Auto-Set™ - SIMPLE SETUP FOR WELDING MILD STEEL
Example: .035" diameter wire, welding 1/8" material.

1. Activate Auto-Set™ by selecting the diameter of the welding wire with the WIRE SPEED knob. Auto-Set™ light comes on.
2. Select thickness of the mild steel being welded with the VOLTAGE knob. Setup complete.

USE 75% ARGON / 25% CO2 SHIELDING GAS

Manual Setup
Refer to chart below to select Volta
Speed based on thickness of metal

Material	Suggested Wire Types, Polarity	Suggested Shielding Gases (Flow Rate of 20-30 cfh)	Wire Sizes (Diameters)	24 ga. (0.6mm)	22 ga. (0.8mm)	20 ga. (0.9mm)	18 ga. (1.2mm)	16 ga. (1.5mm)	14 ga. (1.9mm)	1/8" (3.2mm)	3/16" (4.8mm)
				120 Volt - 24 gauge to 3/16"							
Steel	Solid Wire ER70S-6 (DCEP)	C25 Gas Mixture 75% Ar / 25% CO2	.024" (0.6 mm)	2/22	2.5/30	2.5/30	3/35	3.5/45	4.5/60	5.5/70	10/80
			.030" (0.8 mm)	2/20	2.5/25	2.5/25	3.5/30	4/40	4.5/50	6/60	10/60
			.035" (0.9 mm)	---	2.5/20	2.5/20	3.5/25	4.5/35	5/40	6.5/45	10/45
	Solid Wire ER70S-6 (DCEP)	100% CO2	.024" (0.6 mm)	---	3.5/30	3.5/30	4/35	4.5/45	5/50	7/60	---
			.030" (0.8 mm)	---	3.5/25	3.5/25	4/30	4.5/35	5/40	7/45	10/50
			.035" (0.9 mm)	---	3.5/20	3.5/20	4/20	4.5/25	5/30	7/35	10/40
	Flux Core E71T-11 (DCEN)	No Shielding Gas required. Good for windy or outdoor applications.	.030" (0.8 mm)	---	---	---	2.5/30	3/30	4/40	5/45	10/50
			.035" (0.9 mm)	---	---	---	2.5/20	3/20	4/30	5/35	10/40
			.045" (1.2 mm)	---	---	---	---	---	---	---	---
Stainless Steel	Stainless Steel (DCEP)	Tri-Mix 90% He / 7.5% Ar / 2.5% CO2	.024" (0.6 mm)	---	4/35	4/35	4/40	5.5/60	6.5/75	10/80	---
			.030" (0.8 mm)	---	4/25	4/25	4/30	5.5/40	6.5/50	10/55	---
			.035" (0.9 mm)	---	4/20	4/20	4/25	5.5/35	6.5/40	10/45	---

Selecting Wire, Gas and Control Settings for Aluminum. Use optional Spoolmate 100 series Spool Gun

Aluminum Spool Gun	Aluminum 4043ER	100% Argon	.030" (0.8 mm)	---	---	---	2/65	2.5/70	6/85	10/90	---
			.035" (0.9 mm)	---	---	---	---	3.5/60	4/65	10/70	---

IMPORTANT: Match drive roll groove to diameter of wire being used. Set Tension knob setting to 3 at start. Adjust tension per instructions in the manual.
Thicker materials can be welded using proper technique, joint preparation and multiple passes.

M-100 (#248282) Torch Consumables:

Nozzles:	Tip Adapter:	Contact Tips:
#246372	#169716	.024"(0.6mm)#087299 .030"(0.8mm)#000067 .035"(0.9mm)#000068 .045"(1.2mm)#000069

Most MIG welders will come with a chart or have the dials with indicators to help you set the voltage and wire speed correctly.

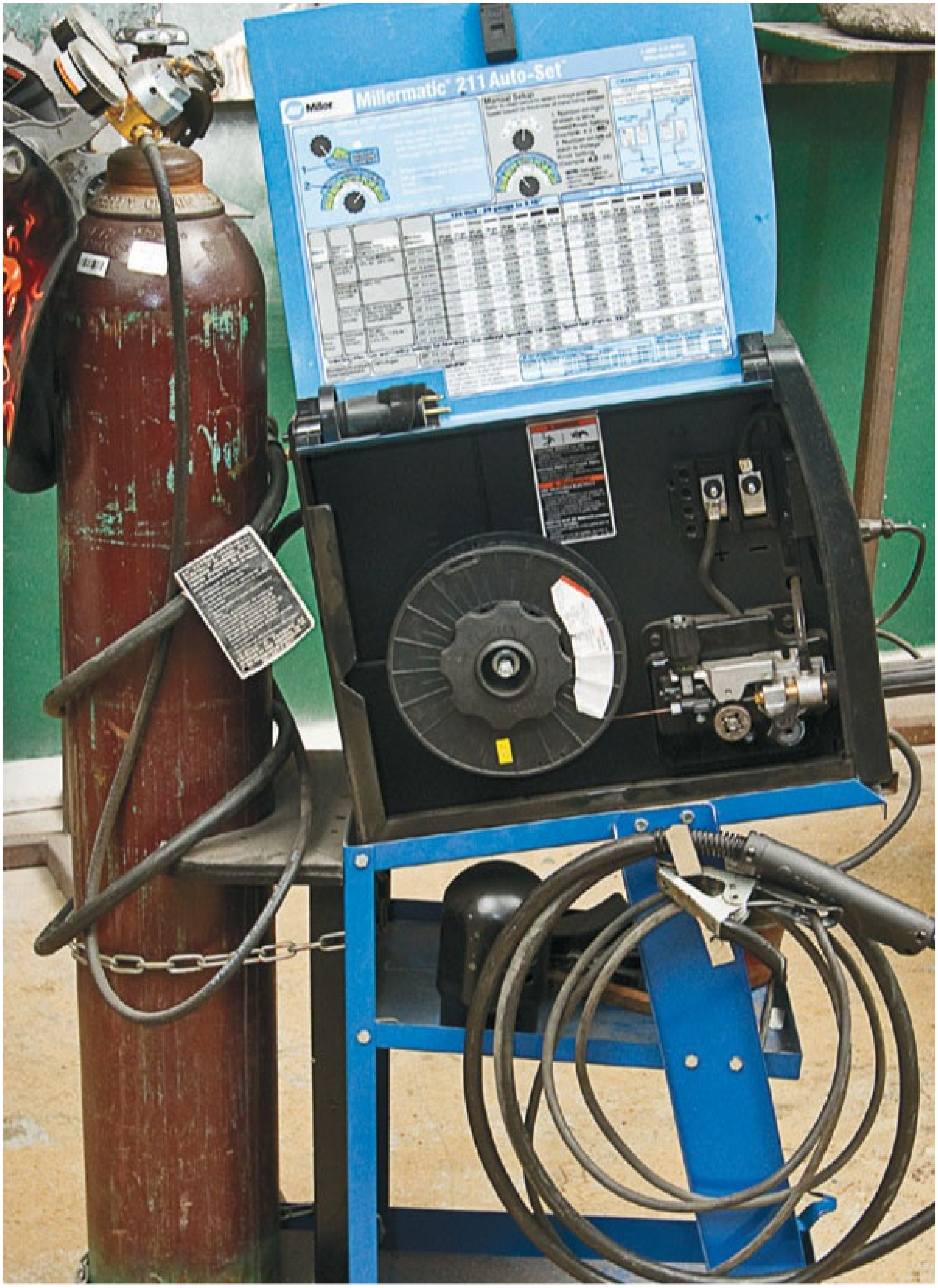
Miller
Millermatic 211 Auto-Set

Set the whole system to the thickness of the metal.

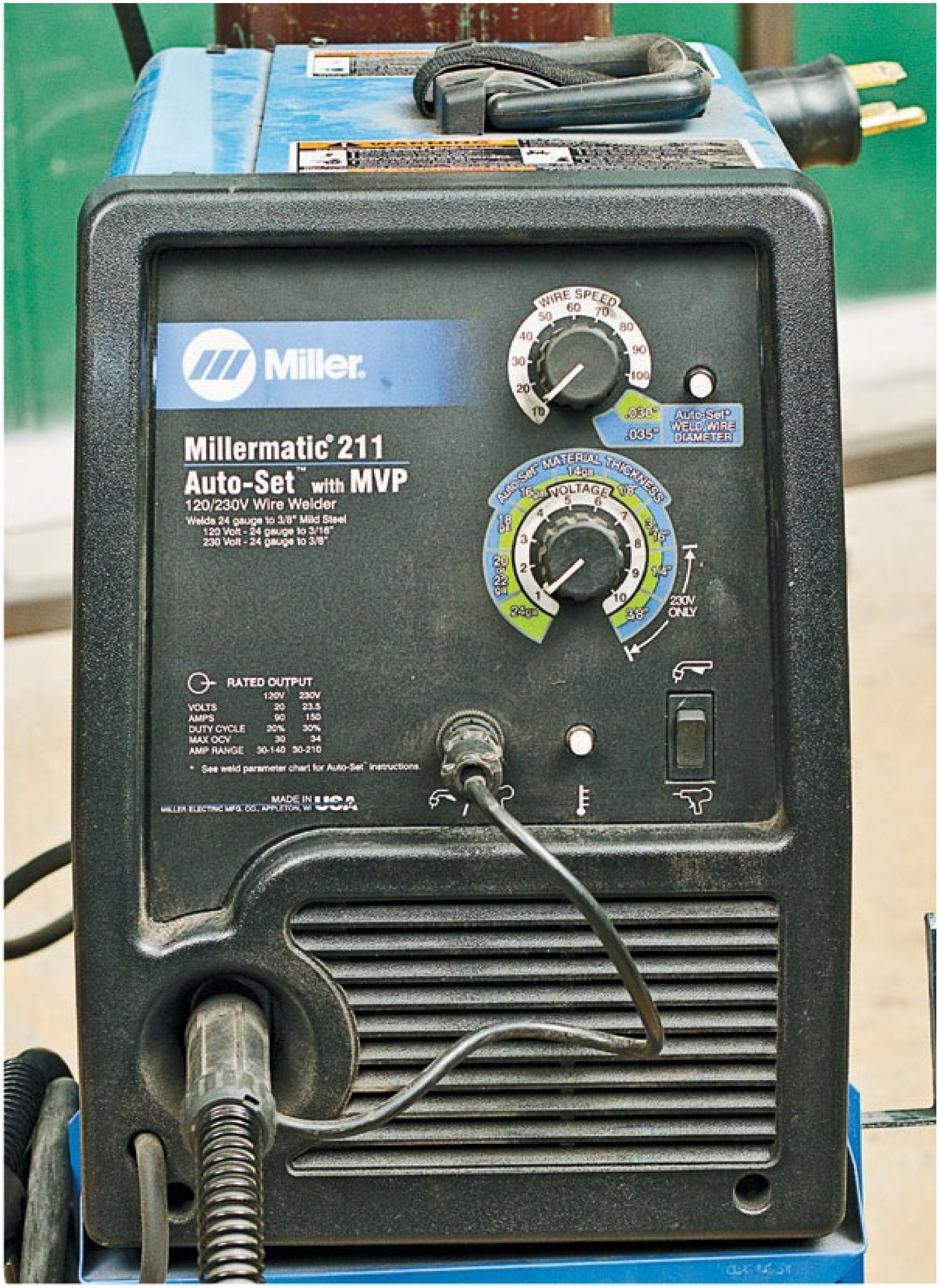
WIRE SPEED
Millermatic 211
Auto-Set with MVP
120/230V Wire Welder
VOLTAGE
Auto-Set MATERIAL THICKNESS
RATED OUTPUT
230V ONLY

Set the wire speed and voltage dials to the proper position for the thickness of the metal you are going to weld.

For welding, the machine will always be set for the thickness of the metal to be welded. A MIG will have two knobs to set. One will need to be set for the voltage, which will determine the heat or intensity of the arc. The other is the rate at which the filler material will feed from the gun tip into the weld zone. Most machines will have a chart that you will need to follow to determine the proper voltage and wire feed settings for the specific thickness or gauge (ga) of metal.

Besides the thickness of the metal, there are a few other things that will determine the settings of the machine. These components are normally constant, and you probably took notice of them while setting up the welder. This includes what the machine is plugged into, whether it is 120 V or 230 V, and the size of the filler wire, most likely 0.030 inches (0.762 mm).

On your machine, you will find a chart that shows the proper wire speed and voltage combinations to use while welding. Most charts will have a line that will begin with the type of metal to be welded on and then the gas set up, wire size, and the proper combination below the gauge to be welded.

For example, to set the machine to weld 16 ga steel using 230 V, in this chart, it indicates steel, the gas mixture, and type of wire and size. You follow the chart over to the 16 ga for 230 V, and the setting is 3.5/45. The 3.5 is the voltage setting and 45 is the wire speed setting. Now set the knobs accordingly.

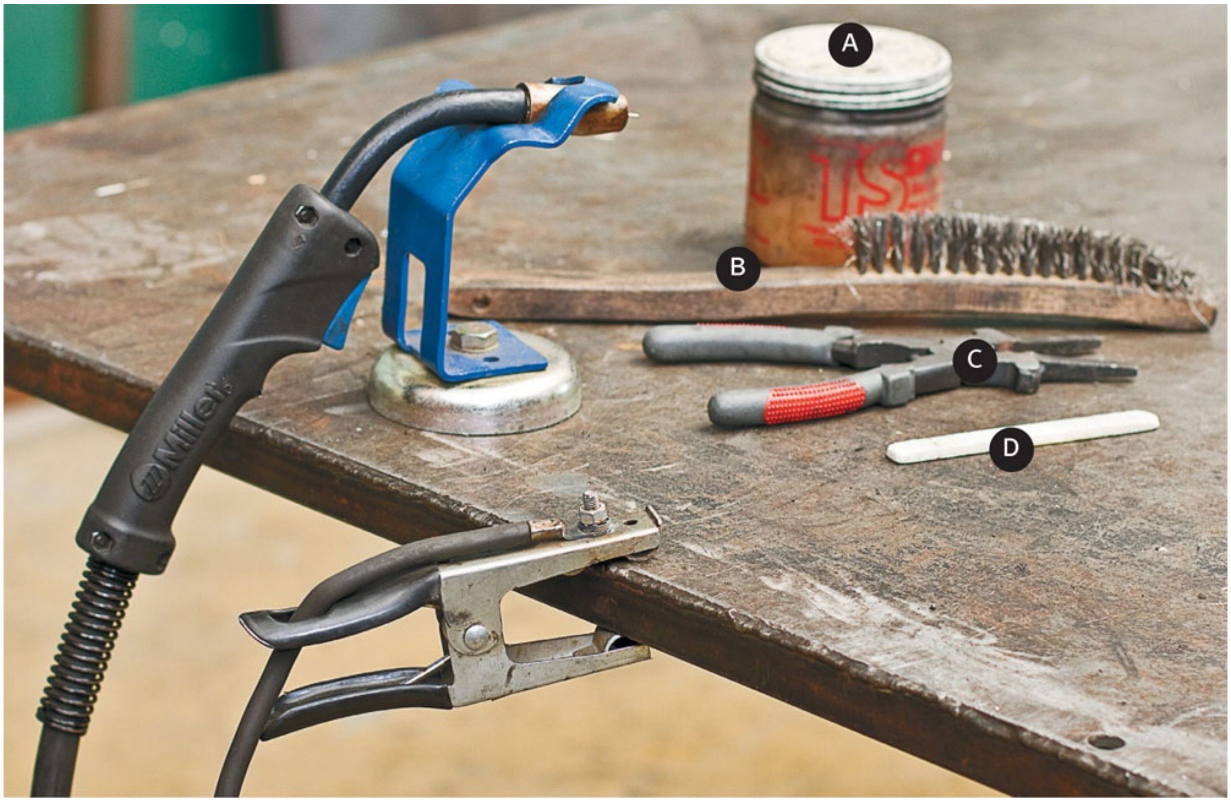

There are four items you should always have around your MIG: A. Spatter protection (tip dip); B. Stainless steel wire brush; C. MIG pliers; and D. Soapstone.

The shield gas pressure is very important. As you will notice, the regulator closest to the gas cylinder reads in pound-force per square inch (lbf/in^2) or kilogram-force per square centimeter (kgf/cm^2). This measurement is also commonly referred to as pounds per square inch (psi) and is the pressure of the cylinder indicating the amount of gas left. The outside regulator, normally the one to the left, indicates the output pressure that will go through the gun liner and out the nozzle. Just like the machine, this should be set in accordance to the material thickness, normally between 20 cubic feet per hour (CFH) (0.57 cubic meters per hour, or CMH) to 30 CFH (0.85 CMH). If you’re welding on very thin gauge material, such as an auto body, you will want the pressure to be low, around 20 CFH (0.57 CMH). Car bodies are amazingly thin metal; if the output CFH (or CMH) is too high, you can still burn holes through the steel even if the machine is set properly. Inversely, if welding on thicker materials, more gas is required. The max thickness of a machine will normally require the flow to be 30 CFH (0.85 CMH). Even though the scale of the flow regulator reads high, too much shield gas can also be problematic as it can push molten metal out of the weld zone.

To understand this concept better, think of it this way: the thinner the material, the less time and distance you can draw out a bead. For thinner material such as auto body, a bead cannot even be made. It is more of a series of tack welds in order to spread out the heat to prevent warpage and burn though in the material. The thicker the material becomes, the more time or length of the bead that can be created.

Always Within Reach

Whenever you are operating a MIG welder, keep these items close by, as you will need them for every welding task.

MIG PLIERS OR MIG SNIPS: These can be used for a lot of different reasons, but the primary purpose for them is to cut the filler wire so that the electrode stick-out is at the ideal $^3/_8$ inch (10 mm) length. MIG snips are normally $^3/_8$ inch (10 mm) thick, so when the wire is cut, the stick-out is perfect.

SPATTER PROTECTION: What I like to call tip dip or nozzle dip, this is a petroleum gel that protects the contact tip and nozzle by preventing spatter from fusing to them. Too much spatter buildup on the inside of the nozzle and on the tip can create a conductive metal bridge that will allow the nozzle to double arc, which is when there are two arcs, one at the tip of the electrode and the other at the edge of the nozzle, normally fusing the nozzle to the base metal. It also obstructs the shield gas from flowing freely through the nozzle and then over the arc.

STAINLESS STEEL WIRE BRUSH: I can't stress the importance of always welding on raw, clean metal. In between a series of tacks or between welds, the burn off should always be removed with a stainless steel brush. This is proper maintenance for welding. Brushes should always be stainless steel. NEVER use one brush for all metals. Always have a separate stainless steel brush for each metal type. It is best to label them according to the metal they are used on such as carbon, stainless, and aluminum. Cross contamination from metal to metal can cause big problems when welding.

SOAPSTONE: This is great for marking the surface of steel because it is non-inclusionary, meaning it won't get into the weld and cause any problem to its integrity.

Other Cutting Tools

There are all kinds of ways to cut metal. In this book, we are really only going cut with chop saws and angle grinders, but I thought I would let you know of a few other methods of cutting that I use in my shop.

few other methods of cutting that I use in my shop.

For the most part, the average beginner or hobbyist isn’t going to go out a buy a hydraulic sheer or industrial metal-cutting band saw. These kinds of metal-cutting machines are used when milling and fabricating on a much more industrial level such as manufacturing. Computer Controlled (or CnC) cutting machines—such as water jet, plasma, or laser—create beautiful cuts in all sorts of shapes and designs. Some CnC machines, such as CnC mills, allow the operator to create things in three dimensions. CnC mills have the ability to take a three-dimensional digital file and mill away the material, such as an aluminum billet, to create things like engines and custom parts for motorcycles, cars, and aircraft.

Machines of this size and expense really aren’t available to the individual. So how would one go about getting a straight cut on a piece of sheet metal or cutting shapes?

Oxygen-Acetylene Torch

This is a method that has been around since the beginning of the twentieth century and is still used today for cutting thicker materials. Oxy-acetylene torches use oxygen as a fuel to increase the heat created from the combustion of the compressed gas acetylene, creating a stream that exceeds 6000°F (3316°C).

When cutting with an oxyfuel torch, it is necessary to wear proper safety gear. Eye protection should have a 3 to 6 shade DIN.

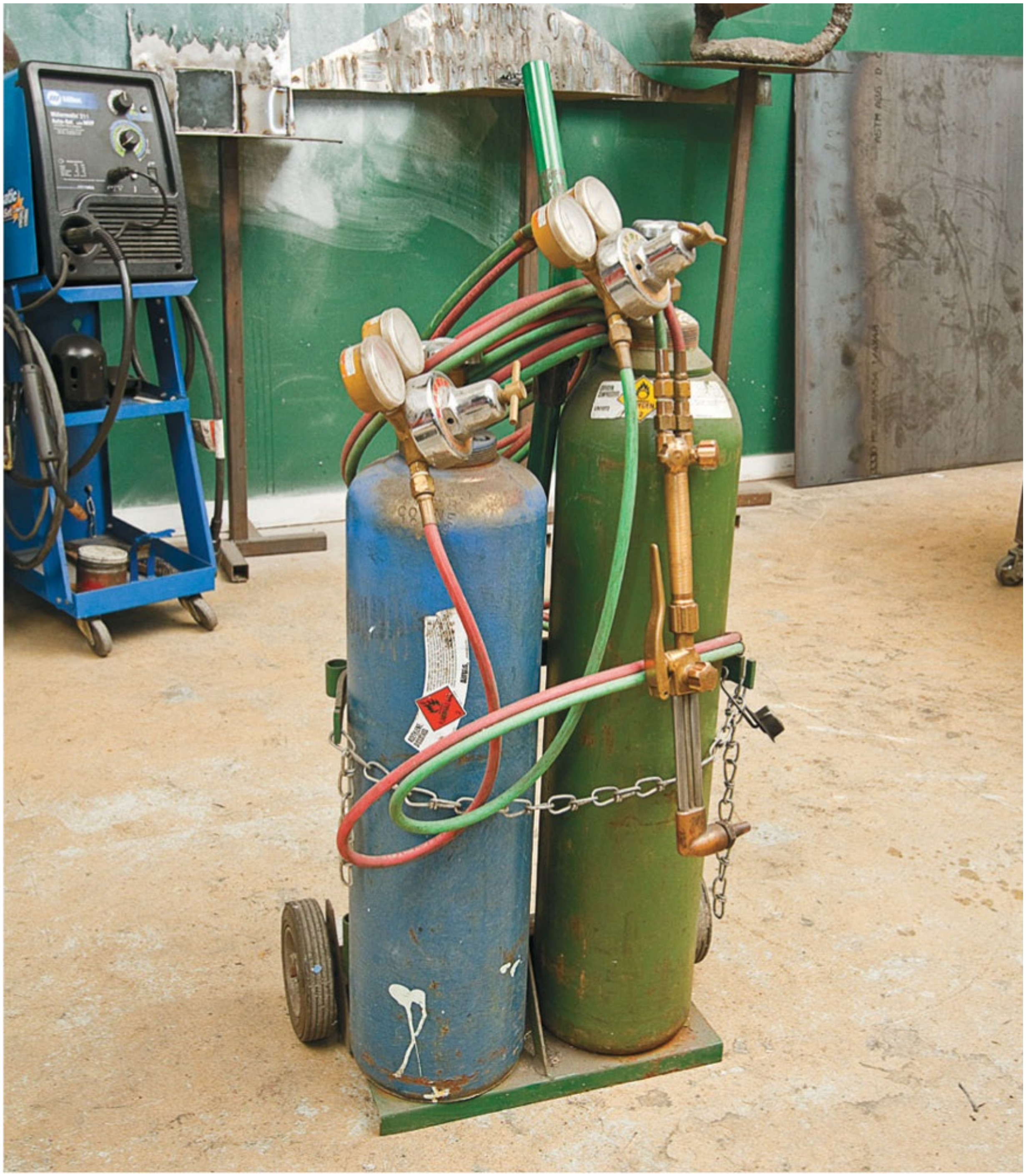

Here is an oxy-acetylene torch setup chained to hand truck. All cylinders need to be chained to something for safety. If a hand truck is not available, chain the cylinders to the wall or stationary table.

Use a paintbrush and soapy water to coat around the connections. Open the main valve. If bubbles form you have leaks and will need to reseal the connections.

Having the ability to not only cut steel but to weld, heat, solder, and braze, makes the oxy-acetylene setup very versatile. Although it is limited in its ability for welding, it makes a great cutting tool. Oxygen can be combined with other gases such as propane and MAPP gas for the same purpose. The oxy-acetylene flame burns between 5600°F and 6300°F (3093°C to 3482°C), which is the highest temperature of any of the oxyfuels; this allows it to cut through most metals.

Since the oxyfuel torch doesn't require any electricity, it's great for onsite work. It has a high cutting capacity that makes it commonplace on work sites such as bridge building.

such as bridge building.

The oxy-acetylene setup requires a cylinder of compressed oxygen and acetylene. Both tanks require regulators for a pressure reading in lbf/in^2, or psi (pounds per square inch), or kgf/cm^2 (kilogram force per square centimeter) and a flow regulator. For safety purposes and to avoid confusion, the regulators for acetylene and oxygen have opposite threading for attachment. When you are attaching regulators to cylinders, never use oil, grease, or lubricant of any kind. The mixture of oil and oxygen only takes one spark to ignite a fire in the regulator, which can lead to an explosion. This can cause you to have a very bad day.

The oxygen tank is pretty straightforward and setting up the regulators is very similar to setting up the Argon-CO^2 tank on a MIG welder. Acetylene, however, is a little more complicated. Acetylene is the combustible gas that creates the flame. It is also a very unstable gas. Acetylene alone cannot be compressed above 15 lbf/in^2 (1.05 kgf/cm^2). To have the ability for acetylene to be pressurized further, it has to be combined with a porous fiber and acetone. As it is pressurized in the tank, the fibers and acetone absorb the acetylene.

This is important safety information to know. Never use an oxy-acetylene torch when the acetylene tank is not upright. Lying on its side can cause the acetone to run through the regulators and into the hoses, thus causing damage. It can also lead to sudden combustion. You should also wait 25 minutes or more from the time the tank has been lying down to use an acetylene tank, allowing the acetone and other material to settle back to the bottom.

> Never use an oxy-acetylene torch when the acetylene tank is not upright.

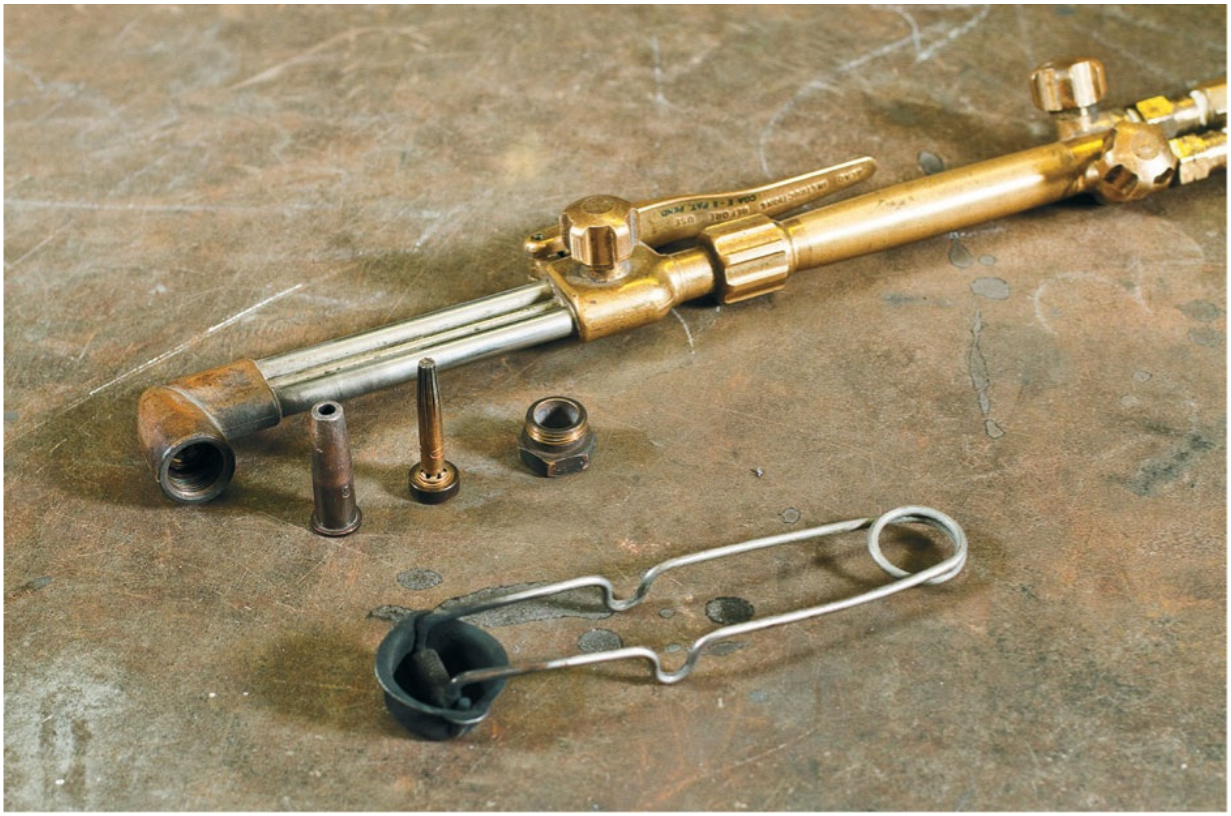

A standard torch cutting set up includes the cutting head and tip. The striker is used to ignite the acetylene gas.

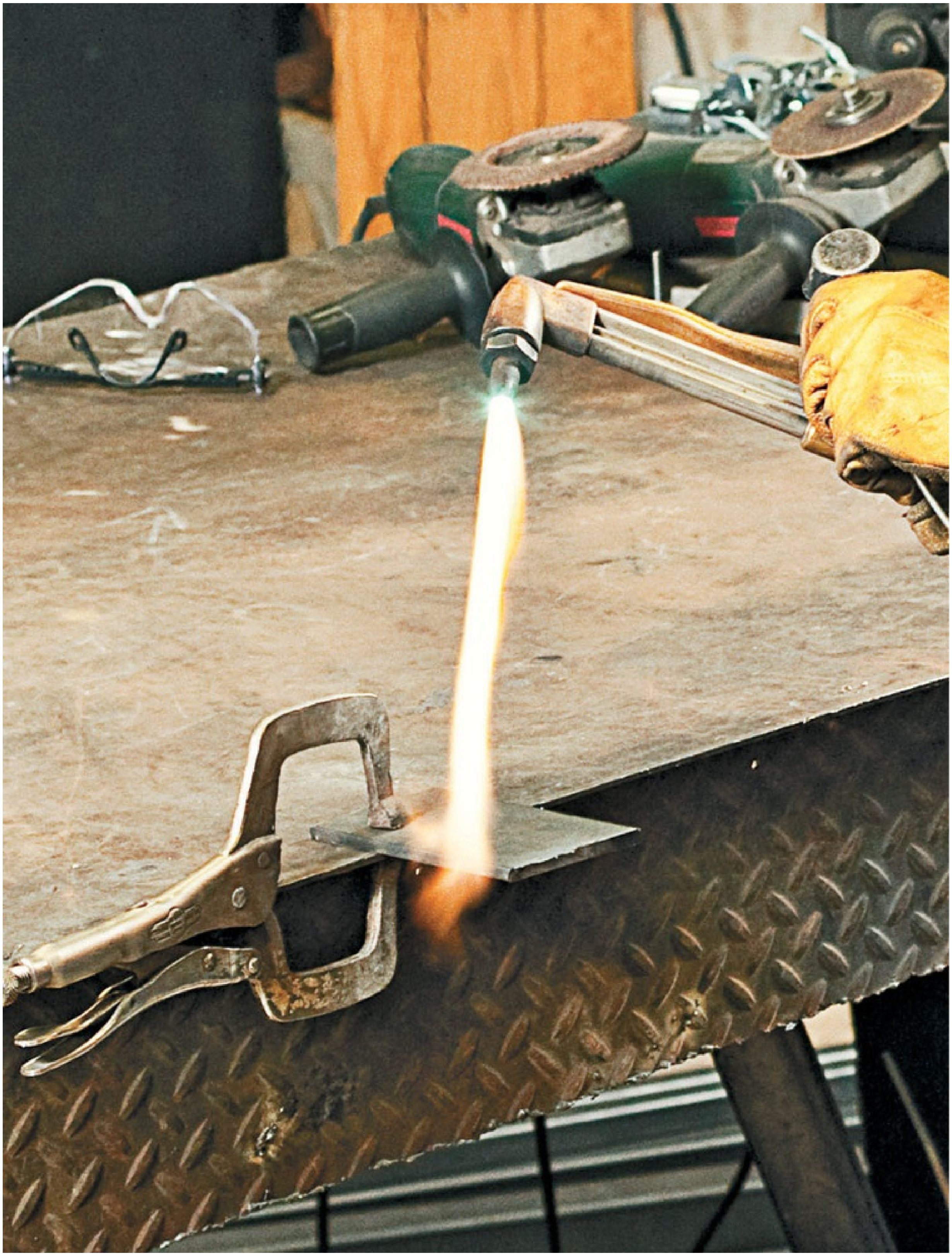

Ignite the acetylene gas first by opening the valve a little and using the striker. Once ignited, slowly add oxygen until you have the proper flame for cutting.

The acetylene tank valve should never be open more than one rotation or less than a quarter. This will allow the valve to be shut quickly in case of an emergency while also permitting enough flow for the proper flame. Acetylene is extremely unstable under pressure and the regulator should never be set over 15 lbf/in^2 (1.05 kgf/cm^2) to prevent explosions. The acetylene tanks should never be spent at more than a seventh of the bottle per day. This will prevent the acetone from leaching into the lines, which can cause damage to the hoses and even the torch.

The hoses that connect to the regulators are flexible and designed to withstand high pressures. The hoses carry the gases to the valves in the torch body. The green hose is for oxygen, and the red hose is for the acetylene. You can remember this as red for flammable.

In a cutting torch setup, the torch head will have three pipes: one pipe for the acetylene flow, one for the continual flow of oxygen, and one for the oxygen blast. The correct ratio of oxygen and acetylene creates the correct flame. The flame heats the metal until it begins to melt. The oxygen cut lever releases a higher flow of oxygen that fuels the heat further and blasts the molten metal through the cut.

The tip attaches to the torch head. The two gases stay separated until they reach the mixing chamber. The holes in the tip allow both of the gases to flow through to create the flame, allowing the metal to be preheated before the high-pressure oxygen. There are normally between 2 and 6 holes on the tip, depending on the duty it needs to perform. The tip is fitted with a center hole through which a jet of high-pressure oxygen passes. A striker is use to create a spark that ignites the gas.

The valves are used to fine-tune the amount of flow between the two gases and to generate the correct flame for cutting. The middle valve is the release and also the fine-tuning oxygen control. The spring valve connected to the lever release is the power switch for after the metal has become hot enough to cut, normally a cherry red color.

In a well-ventilated area, start the acetylene flame by cracking the valve at the base of the torch body indicated by the red hose. Hold the torch away from you and then light it by sparking the striker to ignite the acetylene. The acetylene has a very distinct odor and should never be inhaled directly. Slowly release the oxygen until the flame concentrates into a cone. Continue to add oxygen until smaller cones appear in front of the smaller preheat holes. This is the correct flame. The material and its thickness should determine the intensity of the flame. The highest temperature an oxy-acetylene torch can go is roughly 6000°F (3316°C). For thicker materials this will be the heat to use and will take more time to heat up and cut. For thinner materials, you will want to dial the heat of the torch down to create a finer cut.

To cut with an oxy-acetylene torch, position the flame to begin heating the steel by holding the torch perpendicular to the surface and edge of the steel. Hold the inner cones of flame about 1/16 inch (2 mm) to 1/8 inch (3 mm) above the plate. The proper distance away from the plate will prevent metal from sticking to the tip. Once the metal is a cherry red color, it is hot enough to start cutting. Slowly depress the oxygen lever until it is fully open. Move the torch though the cut path, maintaining the same distance and angle. For thinner materials, it is necessary to angle into the cut roughly 75 degrees. For thicker materials, decreasing this angle and slowing the movement through the cut path will be necessary.

Wait for the steel to become a cherry color before starting the cut.

The side opposite from the cut side will normally have dross, or what is

The side opposite from the cut side will normally have dross, or what is commonly referred to as slag. This is the molten metal that was blown though the cut and then fused and solidified underneath the material.

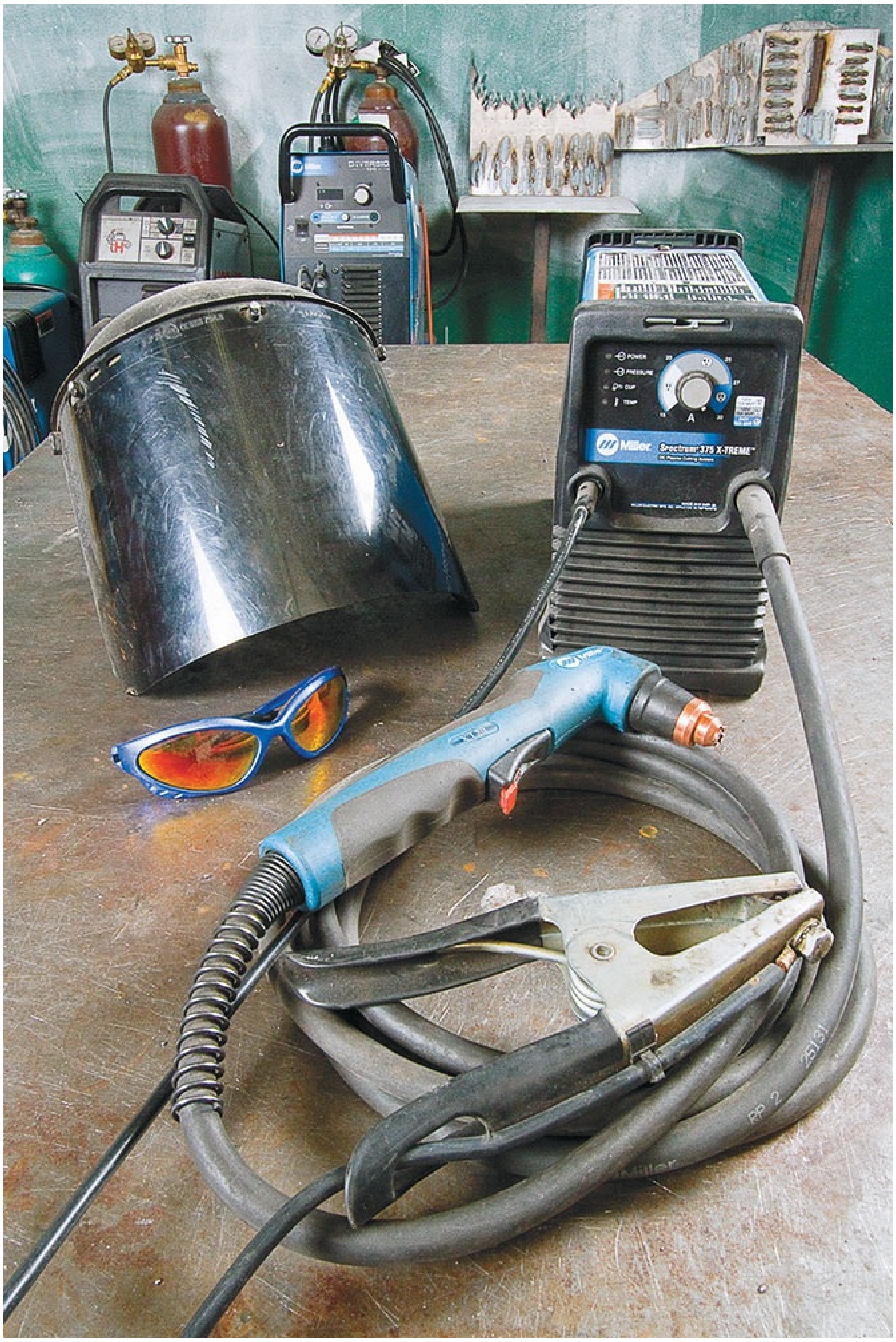

A basic plasma cutting system will include a power source, torch, and ground clamp. Most systems will require a separate air compressor. Wear proper eye protection with a shade between 3 to 6.

The Plasma Cutter

The plasma cutter is a much more elegant machine than the oxy-acetylene torch. The concentration of the plasma arc delivers a finer cut with less dross than an oxyfuel cutting method. Like the MIG welder, the plasma-cutting machine is an arc process. Plasma is a state of matter created by the ionization of gas that produces an intensely hot airborne electrical current. One of the most common examples of plasma is lightning.

The plasma cutter has a lighter, more workable torch that allows for greater control, giving the operator a more fluid motion that allows for tighter curves and more intricate designs.

In the plasma-cutting machine, this arc is produced by compressed air passing through a chamber that houses an electrode. The gas in the chamber is constricted as it ionizes, building up pressure before blowing through the tip where it is met by the grounded piece of material to form an arc. This arc is a concentrated plasma stream that can burn in excess of 40,000°F (22,204°C). The concentrated plasma stream that can burn in excess of 40,000°F (22,204°C). The high temperature of the plasma makes it possible to cut any conductive metals, whereas oxyfuel cutting systems are restricted to certain carbon metals such as mild still.

The intensity of the heat and power of the arc gives the cutter many advantages, such as the ability to move much faster through the cut. The combination of a shorter cutting time and higher temperature cuts down on the probability of thermal distortion or warpage.

One of the drawbacks of the plasma is that the flow in which the arc cuts will create a curve on one side of the cut called a kerf. The kerf will become more prominent on thicker materials and metals such as aluminum. Knowing which side the kerf will be on will help you come up with a plan to produce a cleaner edge and minimize grinding time. Practicing the speed in which to cut with plasma will minimize the kerf by producing a cleaner cut with less dross.

Sparks will be flying when you use a plasma cutter, and you will need proper ventilation.

Unfortunately, plasma cutters are initially much more expensive than oxyfuel cutting systems. Most plasma cutting machines also require air compressors to produce the necessary air pressure to create the plasma arc. If you plan on doing most of your cutting in your shop, then it might be worth the investment.

When cutting with plasma, be sure to take the necessary safety precautions. Wear the required safety gear: number 3 shade eye protection or higher, leather gloves, long-sleeve shirt, leather shoes or boots, and jeans. Cutting with plasma is just like MIG welding, so follow all of the same safety rules and never wear synthetic fibers while welding or using a plasma cutter.

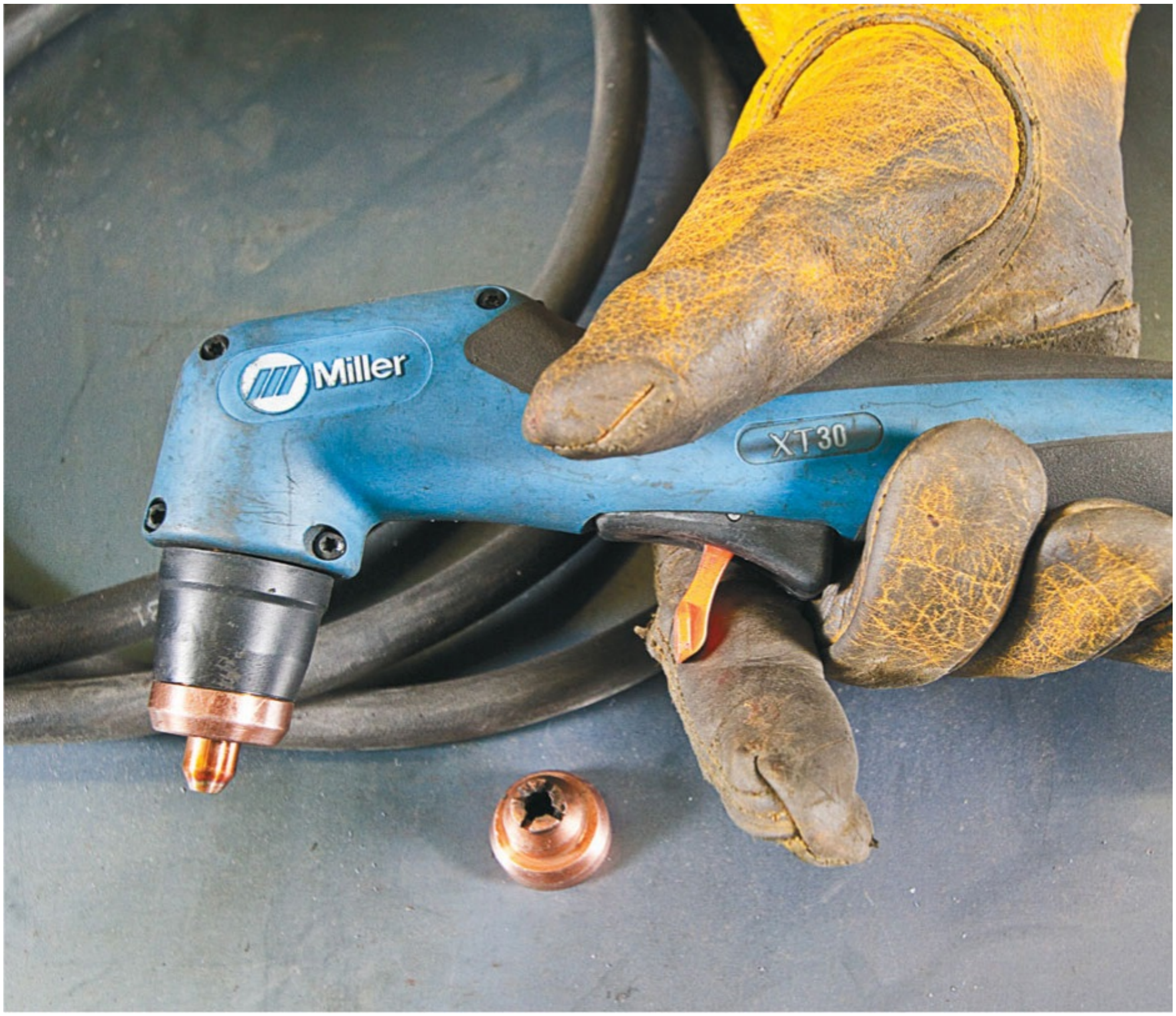

The torch of the plasma cutter is smaller and lightweight. The trigger is fitted with a safety switch to prevent accidental fire. Just like the oxyfuel torch, the plasma torch has similar consumable components. The tips and electrode will be the most frequently replaced components of the torch, but over time the other components will have to be replaced as well.

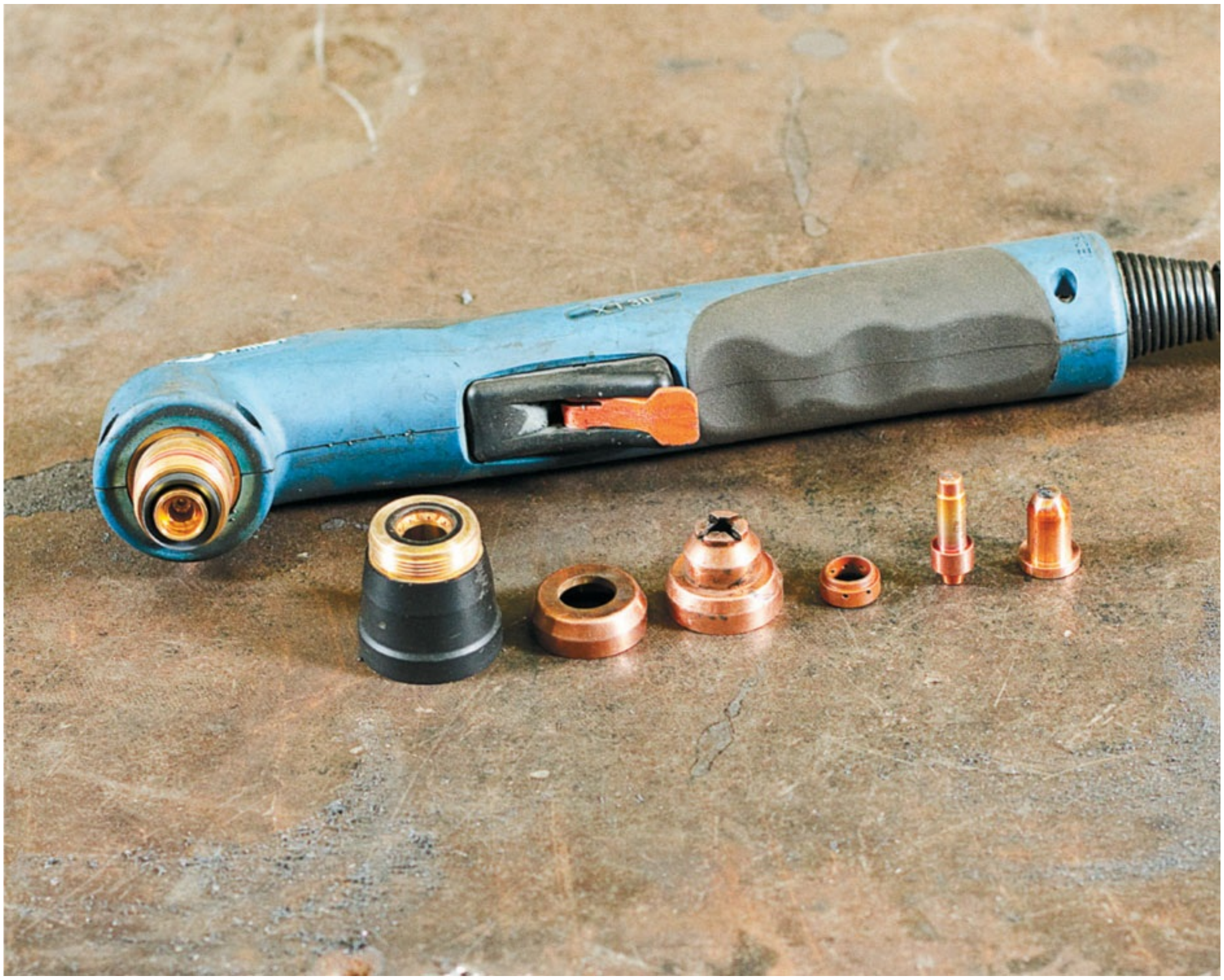

From left to right: retaining cap, deflector, drag shield, swirl ring, electrode, and tip. In this model, the drag shield and deflector are interchangeable. The drag shield encloses the tip, allowing contact to be made with the grounded base metal. When making freehand cuts, it is changed out with the deflector, exposing the tip for added vision.

Cutting with a plasma cutter is very similar to welding with a MIG. The ground has to be clamped directly to the work piece or on a conductive metal surface that the piece is resting on. Set the machine to the proper output for the thickness of the material—most plasma-cutting machines will have a thickness scale surrounding the control knob that sets the machine. Check to make sure the air hose from the air compressor is attached correctly and that the proper amount of air pressure is set. Most machines will have a pressure gauge that will tell you what pressure to use, and some even allow you to make that adjustment on the machine. Newer machines will regulate and adjust the pressure going through the machine automatically as long as there is enough pressure coming from the line.

Once the machine and work piece are set up and secured into place, you are ready to cut. Just like when you're welding, you should be comfortable. Put your body in a position where you are not tense but relaxed. Brace and support your torch hand with your opposite hand.

When cutting, the tip should always be a consistent distance away from the work piece. The closer youare, the finer the cut will be. Aim for a $^{1}/_{16}$-inch (2 mm) distance; more than that and the cut will become too wide. If you are using a drag shield (as seen in the image at the right), then the shield should make contact with the base metal.

Pull the trigger when the tip is just off the material to start the plasma arc. This will start the initial ungrounded plasma arc that will help you to engage the metal. Once the arc engages, draw the torch toward the metal. When the arc grounds as it hits the metal, it will intensify and cut into the material. You’ll need to practice to develop the right pace for a smooth cut.

After cutting with a plasma cutter, the other side of the cut will have a small amount of dross fused to edge. Simply remove the dross with anything that has a solid edge to it, such as a tack hammer or a piece of scrap. Hit the dross in a sharp downward motion to knock it off.

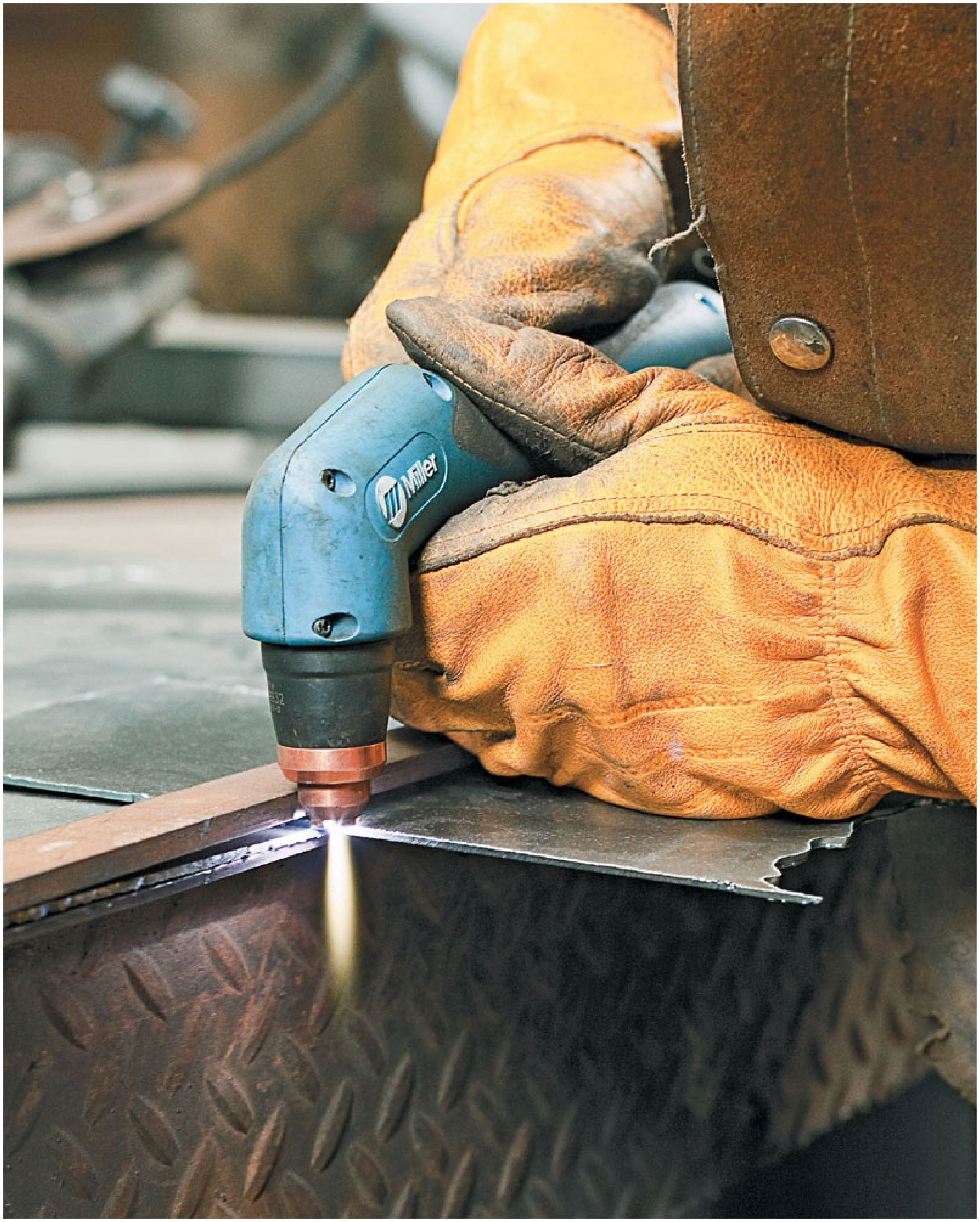

Note how I am cutting with a guide for a straight cut and am using a drag shield (the detachable tip on the plasma gun) in order to make contact with the base metal and my guide.

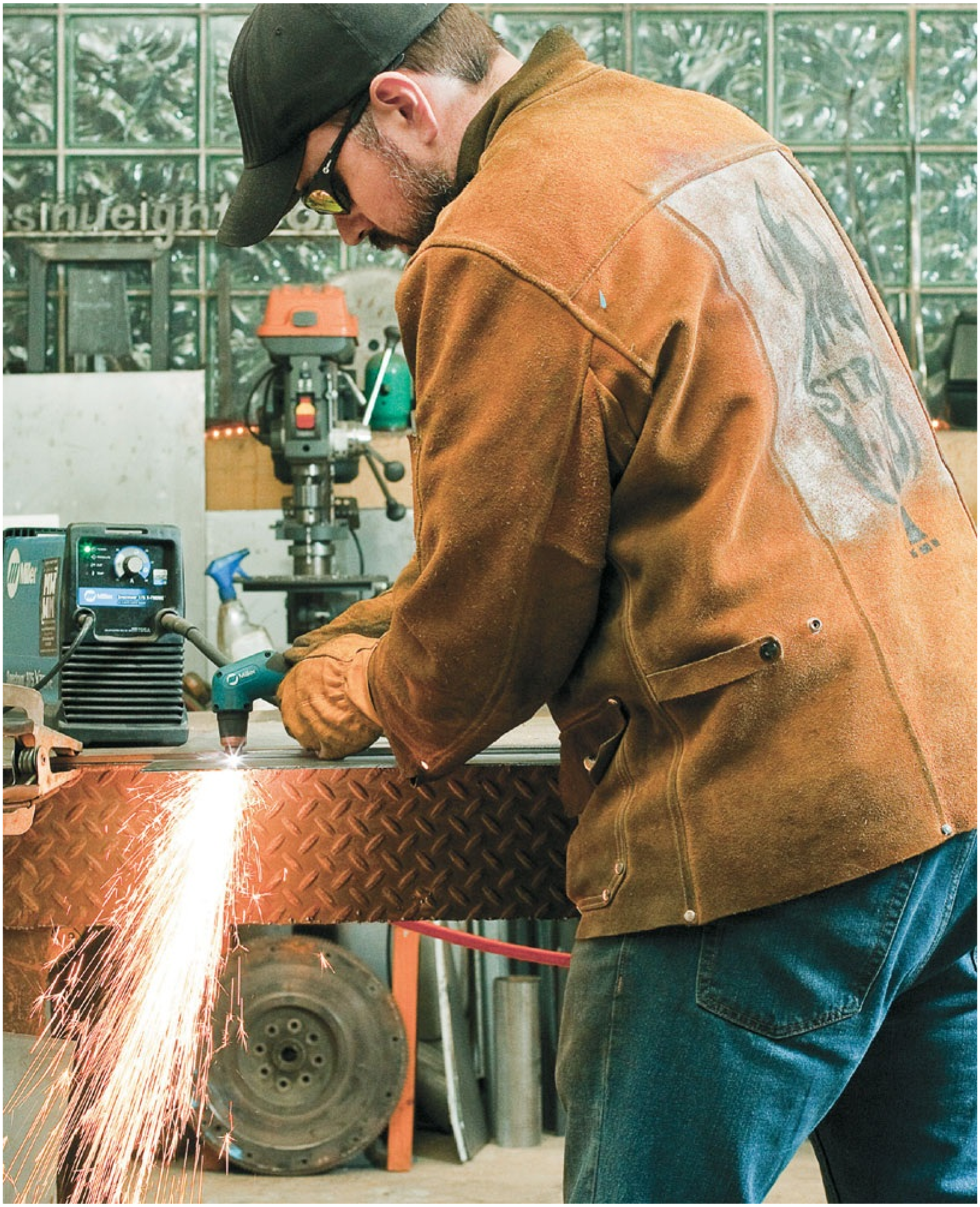

Plasma cutting is an extremely accurate method of cutting metal and provides a very clean cut. Make sure the area around and below the cut is clear of anything that may be affected by high heat.

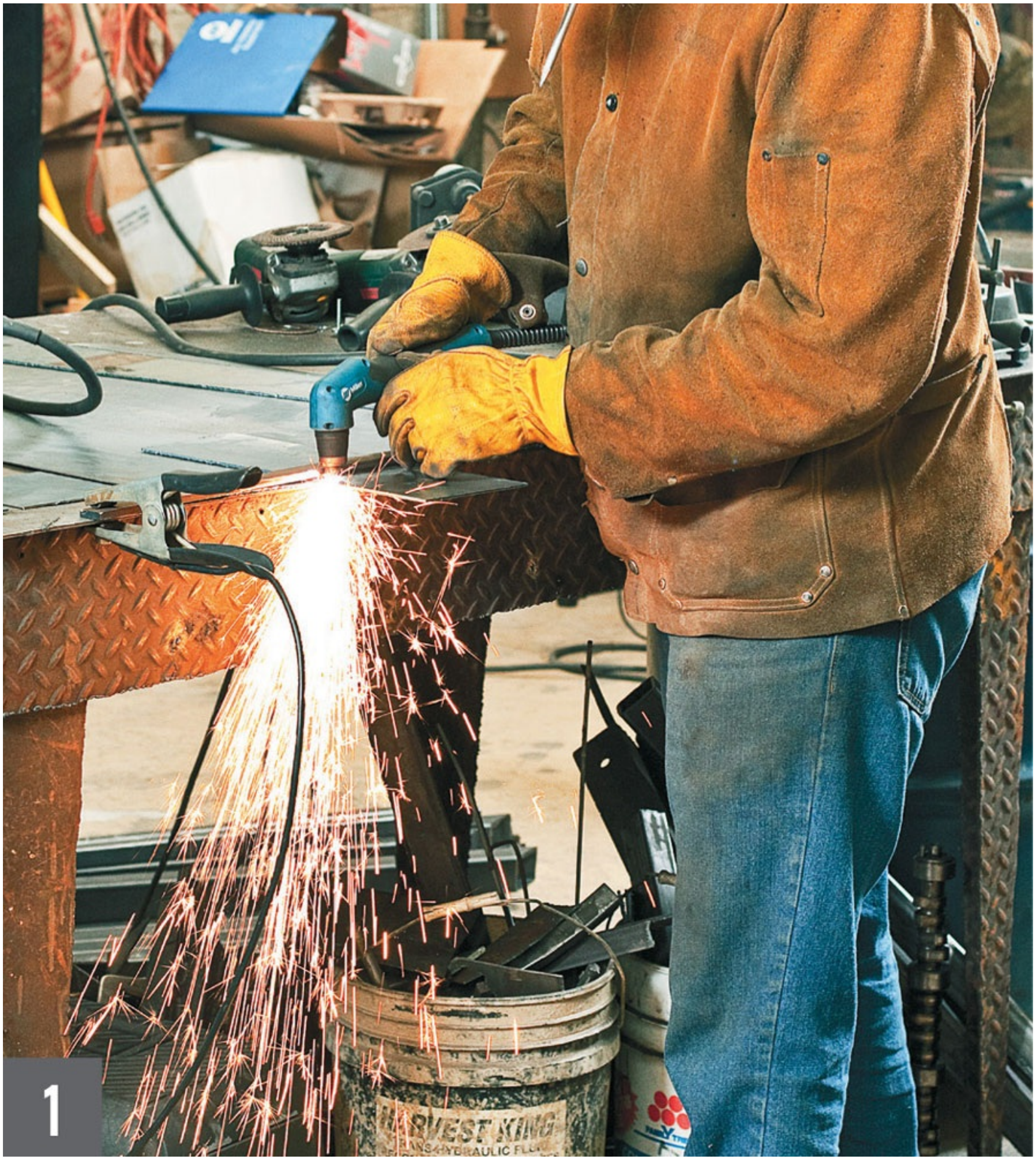
1

Start the arc off of the piece and move into the cut.

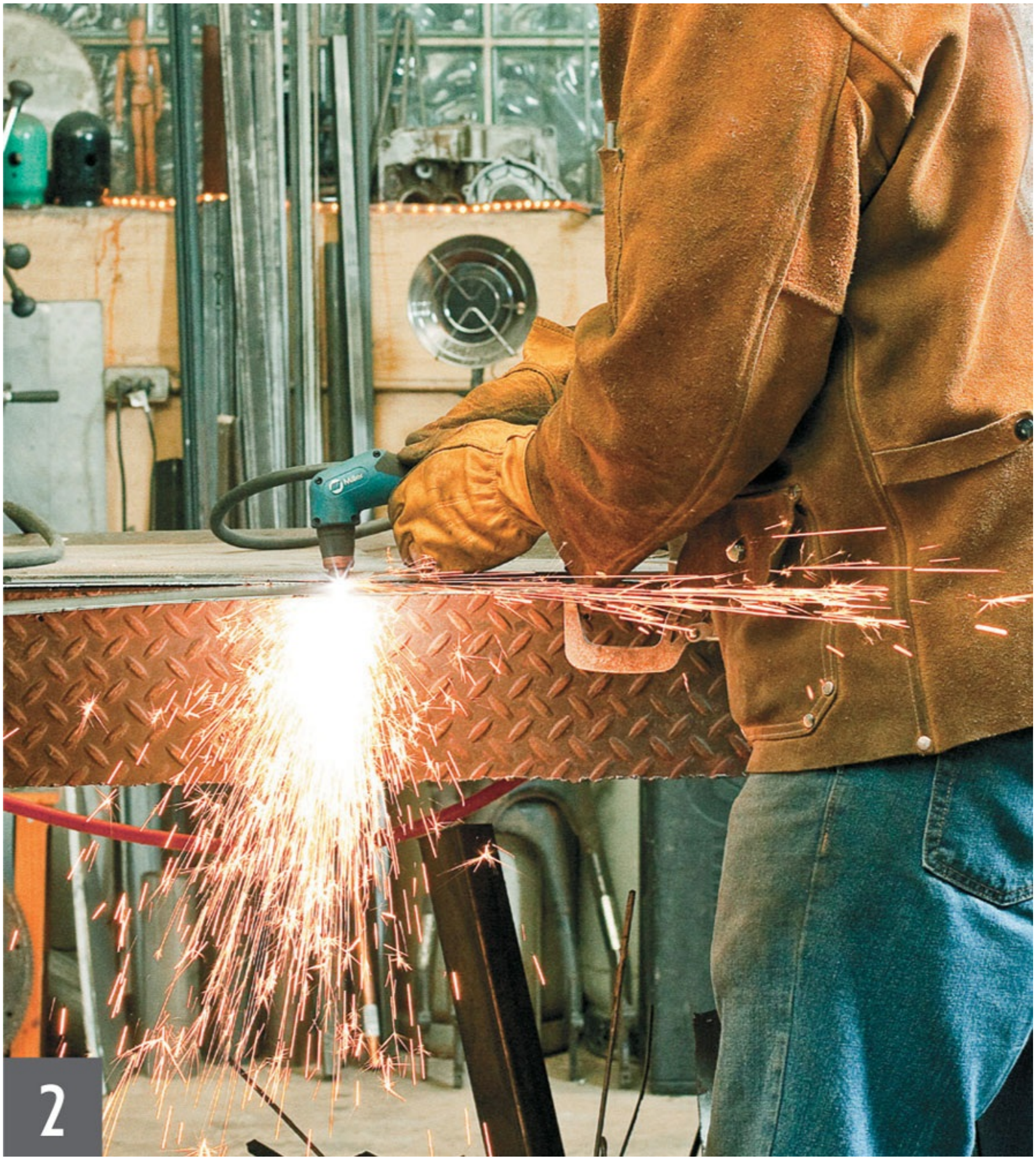

Keep the torch perpendicular to the cut.

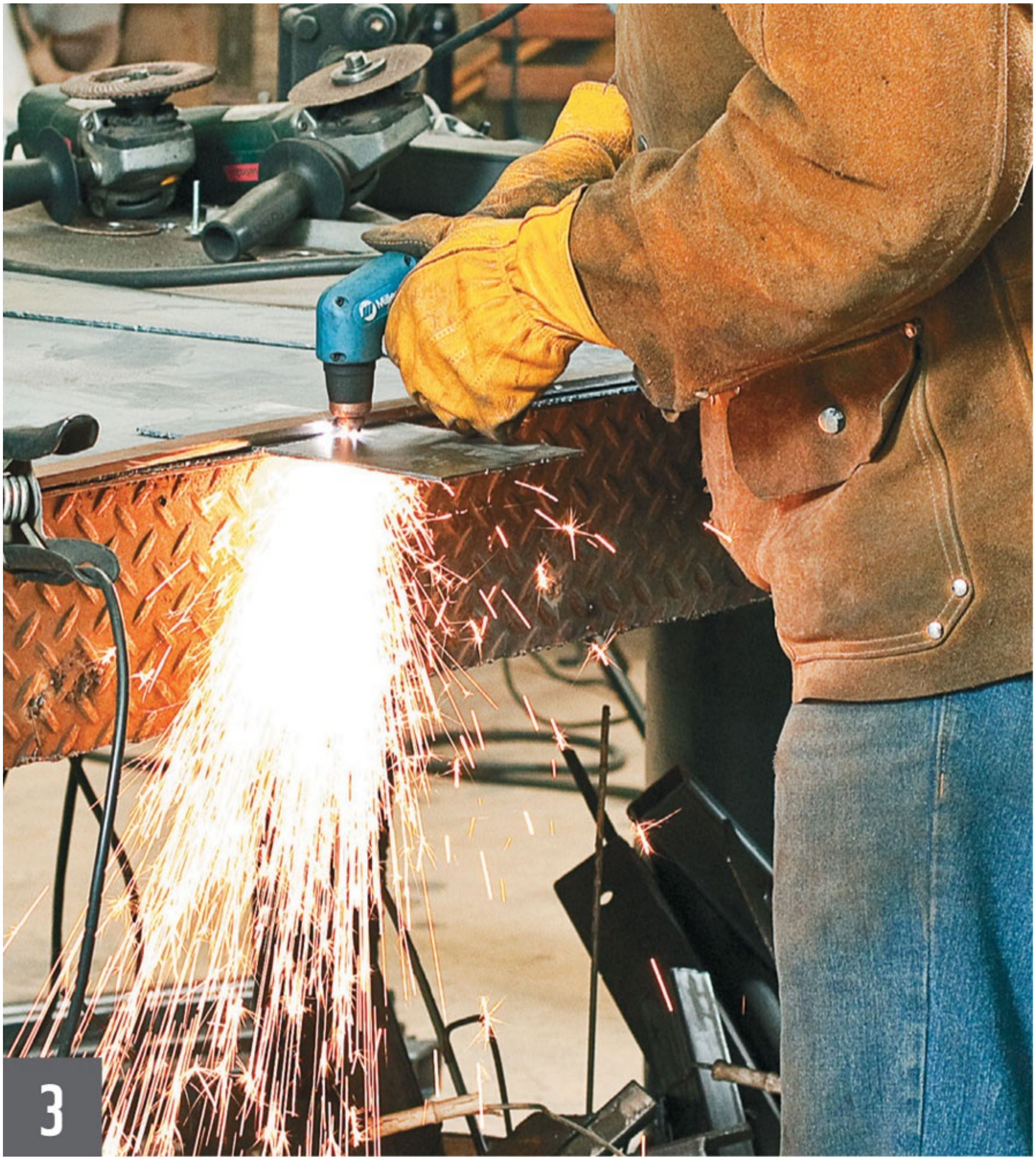

3

Movement through the cut will be determined by the thickness of the material. Find the speed that gives you the cleanest cut with the least amount of dross.

4

The torch will pick up on any jitters or minor movements in your hands. Just like when you're welding, brace and support the torch with both hands and be relaxed. It's not a good idea to have too much coffee before freehand cutting with a plasma torch.

5

When cutting without a drag shield, try to keep the tip 1/16 inch (2 mm) from the metal. Contact without a drag shield will cause double arcing.

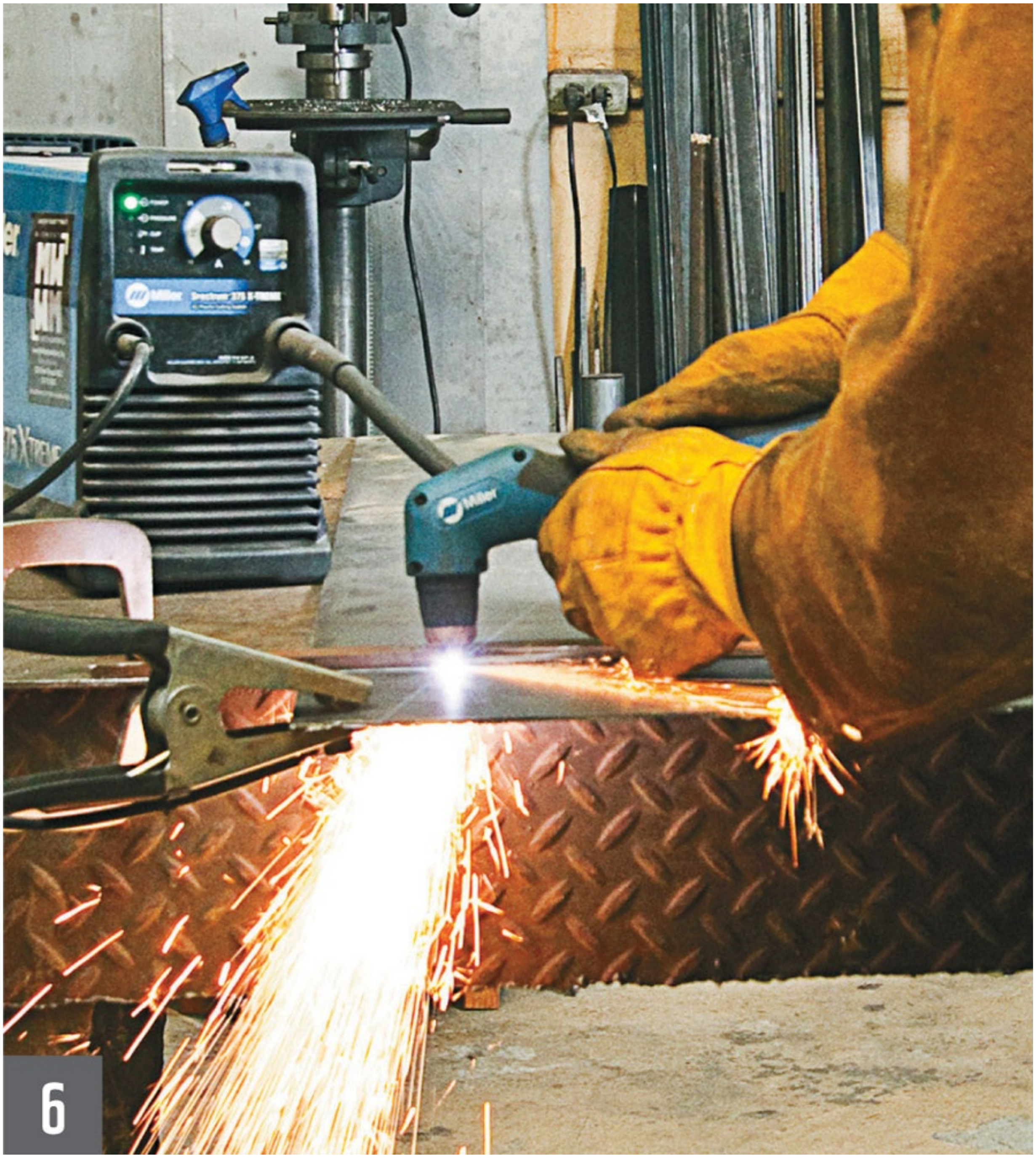
6

When cutting straight edges with a guide, make sure to measure out from where the drag shield makes contact to the guide to the center of the torch tip for a more accurate cut.

Keeping the center seam of the torch in line with the guide line will help you keep a straighter cut.

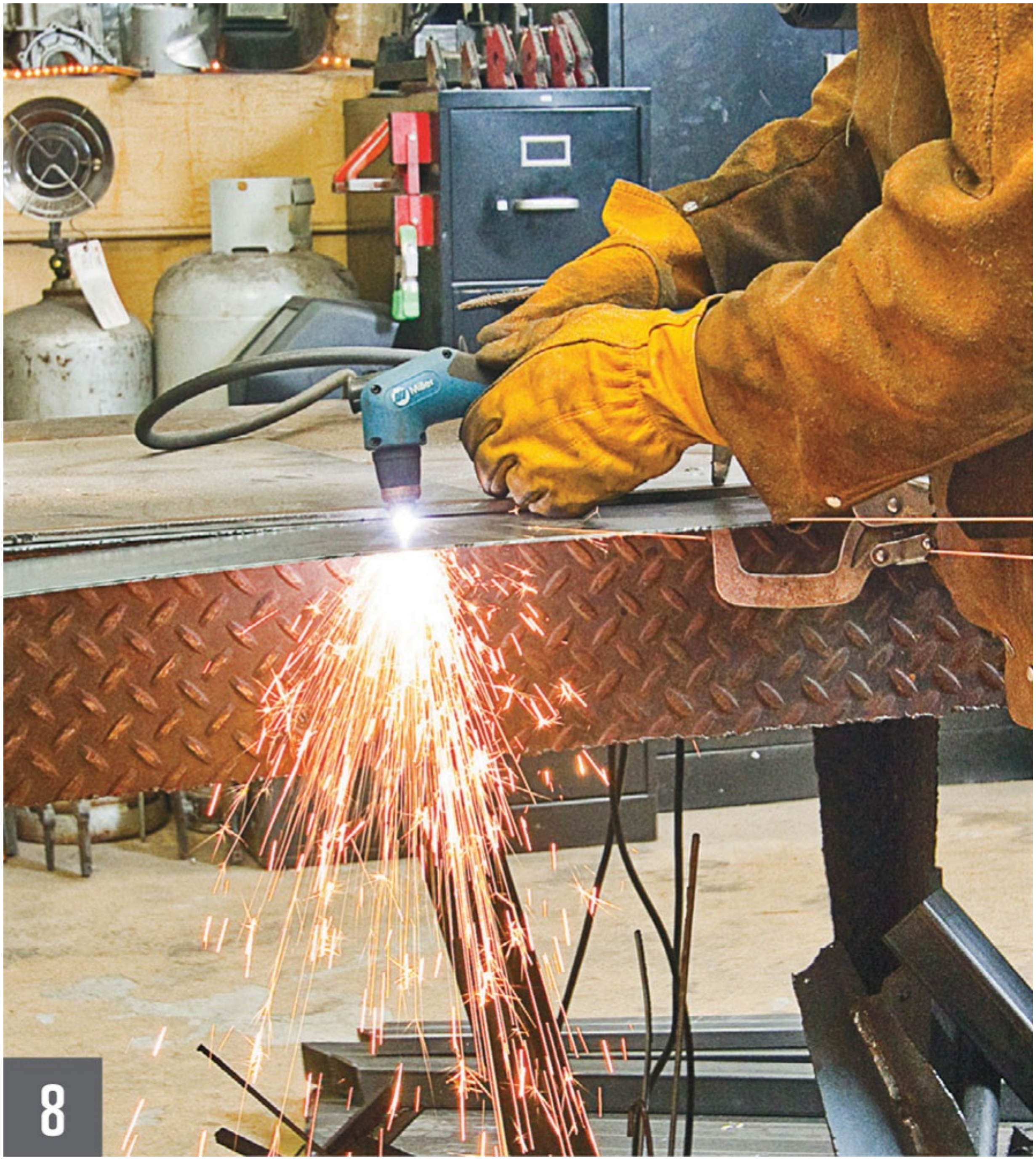

8

On long cuts, cut from both ends and meet in the middle to prevent the metal drop from bending. You can also clamp a support underneath the piece after starting the cut.

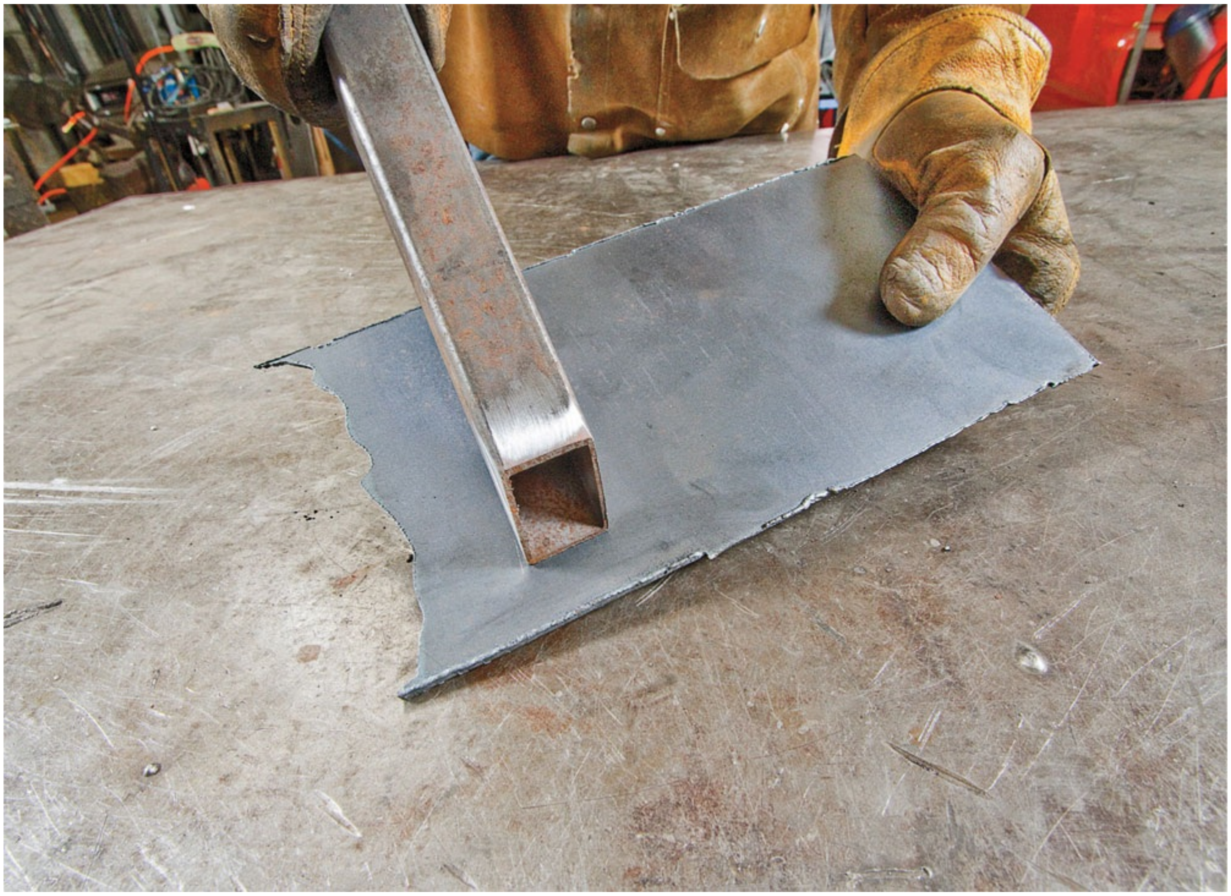

This is dross on the other side of a plasma cut. Use a piece of scrap or a chipping hammer to remove the dross.

3

Selecting, Buying, and Prepping Your Metals

Mild steel is a good place to start when learning any welding process. Mild steel is one of the most common types of metals used in fabricating. It is used in everything from machine parts, automotives, bicycles, and motorcycles to file cabinets, mailboxes, furniture, and home decoration. Mild steel also falls in the category of carbon steel along with other metals such as stainless steel, tool steel, and forged steel. Known as “ferrous” metals, they all contain iron and/or a combination of carbon or other elements to give them their unique characteristics. In general, most carbon steels will be magnetic. In the case of stainless steel, chromium, and nickel, other alloys can be added to create different grades. The grade or combination of alloy elements used to create that specific grade determines the properties of that grade. Steel can be manipulated to create a wide range of alloys by adding different elements such as chromium and nickel.

Mild steel has a low carbon content, which makes it softer and easier to work with. In contrast, adding more carbon and heat-treating the material will make the metal harder, but at the same time more brittle. Steel with more carbon is subjected to a high-heat tempering process that decreases its ductility, depleting its ability to bend under stress. These high carbon metals are known as tool steel or hardened steel and are used for nuts and bolts or creating tools for cutting, drilling, and machine parts. Carbon steel can be hardened further by adding other elements (such as manganese) to create gears, sprockets, and even stronger tools. These types of carbon steel are notorious for being difficult to weld and cut.

Plate metal refers to sheets that are ^{3}/16 inch (5 mm) or thicker. Sheets that are ^{3}/16 inch (5 mm) or thinner are called sheet metal.

Hot Roll vs. Cold Roll

There are typically two different kinds of mild steel that people will refer to when speaking or ordering steel: hot roll and cold roll. When a mill forms mild steel from stock into any number of shapes such as sheet, bar, or tube, it is pressed through a set of rollers that forms it into its desired shape through a process called rolling. The classification of rolling comes from the temperature of the metal as it passes thought the rollers.

When steel is heated to a certain temperature, it begins to crystallize. This temperature is called the recrystallization temperature. When metal is heated to this point to be formed, the process is referred to as hot rolling. When hot rolled metals are produced, this recrystallization forms a scale on the surface of the metal as it cools. This surface is a murky blue-gray colored flaky oxide that is called “mill scale.”

When metal is formed below the recrystallization temperature, the method is referred to as “cold rolling” and produces cold roll steel. This material will not have a mill scale as prominent as hot roll but rather a much more polished and shiny appearance. Cold roll steel will also have an increase of strength and hardness over its hot roll counterpart. It costs more to produce, so the price for cold roll is normally higher than hot roll.

Hot roll materials have softer and more ductile characteristics than cold roll steel, but have relatively the same characteristics when welding. These two types of steel can be welded together with a common MIG welding setup, and both should be prepped in the same way.

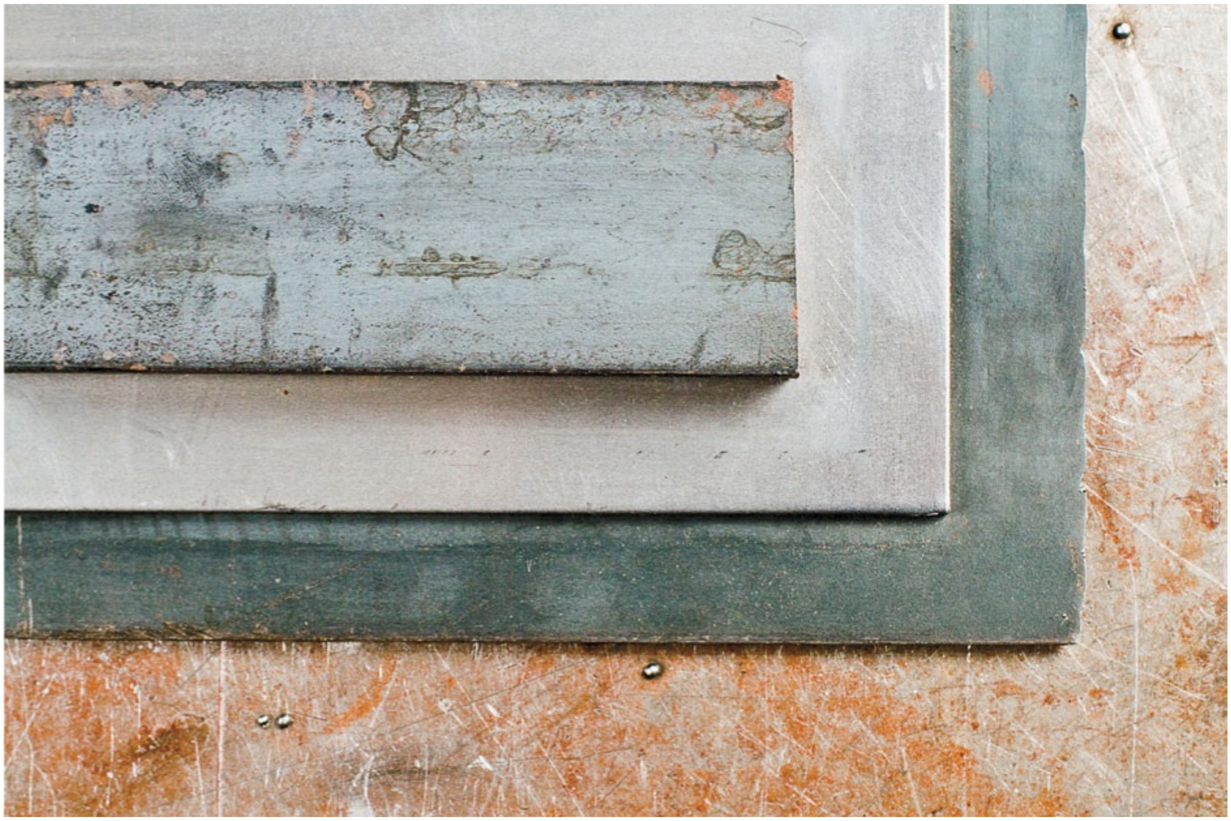

Here you see a piece of cold roll steel between two pieces of hot roll steel. Notice the difference in color. Hot roll has a murky blue/gray patina called "mill scale."

Engine parts, screws, and drill bits are all hardened steel.

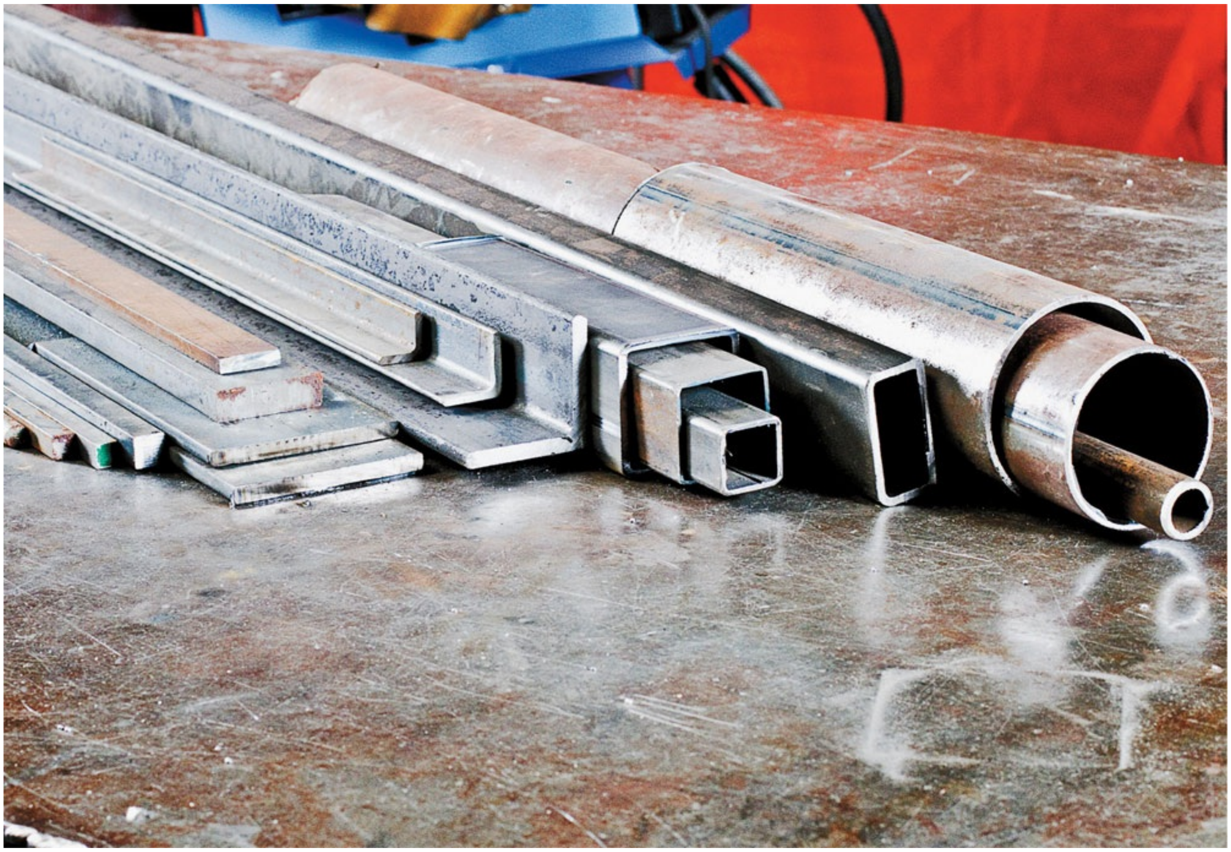

Metal is made in all different shapes and sizes to suit almost any job you can think of. You can find it in plate, bar, angles, and tubes that come in square, round, or rectangle.

Filler Material

Since welding is a process of liquefying the two pieces of base metal, the filler metal should have the same characteristics, such as the melting point. When welding mild steel, a mild steel welding wire or rod should be used. Dissimilar metals normally cannot be welded together. Nonferrous materials such as aluminum cannot be welded to a ferrous material like stainless steel. Aluminum and iron have different properties that will not allow them to be welded together, mostly due to their melting points. Steel has almost twice the melting point of aluminum. Some more advanced welding techniques require you to use different filler and base metals combinations, but for now, we’ll stick with the basics and mild steel.

Decorative Options

Metal comes in all different shapes and sizes, from bar to tube and round to rectangle. Bars, sheets, and tubes are most common for straight fabrication, but what if you are looking for something a little more decorative? Mills and foundries also create a wide variety of ornamental metals that can be used in projects. Cap rails are a good example of a decorative bar that is used for the tops of hand railings. Other decorative elements such as rosettes can be purchased to give a fence or piece of art a more attractive look.

Sheet metal also comes in a wide verity of styles and patterns. Diamond plate, expanded, and perforated sheet metal are popular for not only their decorative characteristics but their functionality as well. Diamond plate works as a tread for its strength and its durability. Expanded metal sheet and perforated sheet metal are great for threads, drainage, fencing, and industrial applications such as shelving and catwalks.

Diamond plate steel and expanded steel are just two examples of the many different styles of metals you can get. Diamond plate and perforated steel are great for steps. Expanded steel can be used for anything from steps to enclosures and decorations.

Buying Metal

No matter what the project may be, it is normally easy to find what you are looking for if you know where to look. Typical home improvement or hardware stores are not the best places to go to purchase metal. They tend to have very small selections, which they sell at a high markup. The best thing to do is to go online and search for metal distributors in your area. Be aware that the majority of them are going to be large mills or metal distribution companies that only deal in large quantities and sell to businesses such as construction. A small privately owned metal distributor that deals with local business or small manufacturers is what you are looking for. This size distributor will normally sell single pieces such as a 4 x 8 foot (1.2 x 2.4 m) full sheet, 14 ga hot roll steel, or a 24-foot (7.3 m)-long stick of 2-inch (5 cm)-square tube with a 0.065-inch (1.7 mm)-thick wall. Don’t worry, I will explain.

When buying metal, it is best to research several places to find the best price. I know this sounds elementary, but pricing differences can actually be much more drastic than you think, so shopping around could save you a lot of cash. Metal is mostly sold by the weight. The more you get, the less expensive it is. Some places may sell you an 8-foot (2.4 m) drop (a piece that is cut away from the measured purchased material) for a certain price where another place will sell you the whole stick (a complete length of milled metal) for less than the price of the other guy’s drop.

When I first moved to Chicago, I found one place with good prices and service. I have yet to find another distributor that will beat them. But it is always a good idea to find two or three places just so you know you are getting the best price. Normally, once you establish a good rapport with a distributor, they will give you a better deal.

Metal Thicknesses		
Gauge (ga)	Steel Thickness in Inches	Steel Thickness in Millimeters
3	**0.2391**	**6.073**
4	**0.2242**	**5.694**
5	**0.2092**	**5.314**
6	**0.1943**	**4.935**
7	**0.1793**	**4.554**
8	**0.1644**	**4.176**
9	**0.1495**	**3.797**
10	**0.1345**	**3.416**
11	**0.1196**	**3.038**
12	**0.1046**	**2.657**
13	**0.0897**	**2.278**
14	**0.0747**	**1.897**
15	**0.0673**	**1.709**
16	**0.0598**	**1.519**
17	**0.0538**	**1.367**
18	**0.0478**	**1.214**
19	**0.0418**	**1.062**
20	**0.0359**	**0.912**
21	**0.0329**	**0.836**
22	**0.0299**	**0.759**
23	**0.0269**	**0.683**
24	**0.0239**	**0.607**

How to Order

In welding, the thickness of the metal is very important. It’s what you use to determine the settings for your welding machine. Metal thicknesses are typically measured in gauges (abbreviated as ga). For most materials, as the gauge number increases, the material thickness decreases.

The gauge of the material corresponds to the thickness of the material in inches or centimeters. The measurement of the thickness in gauges will vary inches or centimeters. The measurement of the thickness in gauges will vary depending on the material. In this chart, you will see that a 16 ga piece of mild steel is 0.0538 inches (1.4 mm) thick, but a piece of 16 ga aluminum could range from 0.070 inches (1.8 mm) to 0.0508 inches (1.4 mm) thick, depending on the manufacturer. These numbers also can vary depending on what form of the material you order, such as sheet as opposed to tube.

The gauge is a general measuring tool that has a fairly loose range. The decimal equivalents of the gauge of the metal you are purchasing are more exact. Since the gauge for measurement can vary so much, when you’re working on something with a tight tolerance, I recommend you order by decimal, not gauge.

Unlike the filler metal, the welding wire for the MIG welding machine will always be measured in decimals for accuracy, as will the corresponding gun tips and drive wheel.

So what do you need to get started? There are a few projects in this book that will have a list of materials that you will need to get in order to complete those pieces. But for now, I want to teach you how to weld. After all, that’s what this book is all about.

Most local metal dealers will allow you to go onto the warehouse floor and take a look at what they have. Even better are the guys that will let you look through their racks of drops. These are pieces that are cut from stock material and left behind by other companies doing larger jobs. If it’s a large enough distributor, they will have a huge rack of drops with all sorts of cuts to choose from; this is what you will want for practicing. Find a sheet of $^{1}/_{8}$-inch (3 mm)-thick steel, cold or hot roll, 10-inch (25.4 cm)-square or larger. This piece of steel is what you will need for the first exercise. Next, you will need to get three or more feet of cold or hot roll $^{1}/_{2}$-inch (13 mm)-square steel bar.

If you are having trouble finding a metal distributor, try a junk or scrap steel yard. Right now we are just looking for small pieces of mild steel to get you started. Don’t break the bank for his material. A good chunk of an old car chassis will work fine for the first exercise.

It may be necessary to use a grinder to even out some cuts that may be a little off before you can prep the steel for welding.

Preparing Your Metal

The preparation of the metal is crucial to the welding process. When welding on any metal, it must be raw and free of any grease, oil, dirt, rust, paint, oxides, and other inclusionary elements that can contaminate the weld. Surfaces that are not cleaned properly can affect the weld quality. A clean weld will last much longer and will be free of included elements that may cause porosity and create a poor weld.

All of these contaminates can include themselves in the puddle and create all kinds of problems. Some particles heat to the point of off-gassing to create bubbles in the metal as it solidifies. This type of porosity is sometimes called “moon craters”. Other causes of porosity will be addressed in the next chapter.

It is common for dealers to coat carbon steel in oil to protect it from rusting in the warehouse. This oil attracts dust and dirt and over time becomes extremely grimy. It is not necessary to remove all of this grime at first; removing it from just the areas to be welded is fine. After completing the fabricated piece, you will want to remove the rest of the dirty oil with denatured alcohol or acetone. This will allow a clean and prepped surface for painting or whatever else you may decide. A wire wheel or brush can also be use for further cleaning. Wire brushes come in a wide range of bristles. Just like the grit of sandpaper or flap disk, some will add a texture to the metal’s finish. If the metal surface already has the look or texture you want, make sure to use a softer stainless steel brush.

Prepping multiple pieces at once will save you time.

Most of the cleaning and metal preparation for the weld zone is normally done with a small angle grinder. There are all kinds of grinding disks that can be used to prep the metal depending on the desired task. For some joints, it is necessary to remove a considerable amount of metal from the edges. This will require a grinding disk with a low grit such as a stone, whereas a 60-grit flap disk will work fine for cleaning the surface of cold roll steel. The amount of what needs to be removed or how dirty the metal is will normally dictate what will be used to prep that particular surface.

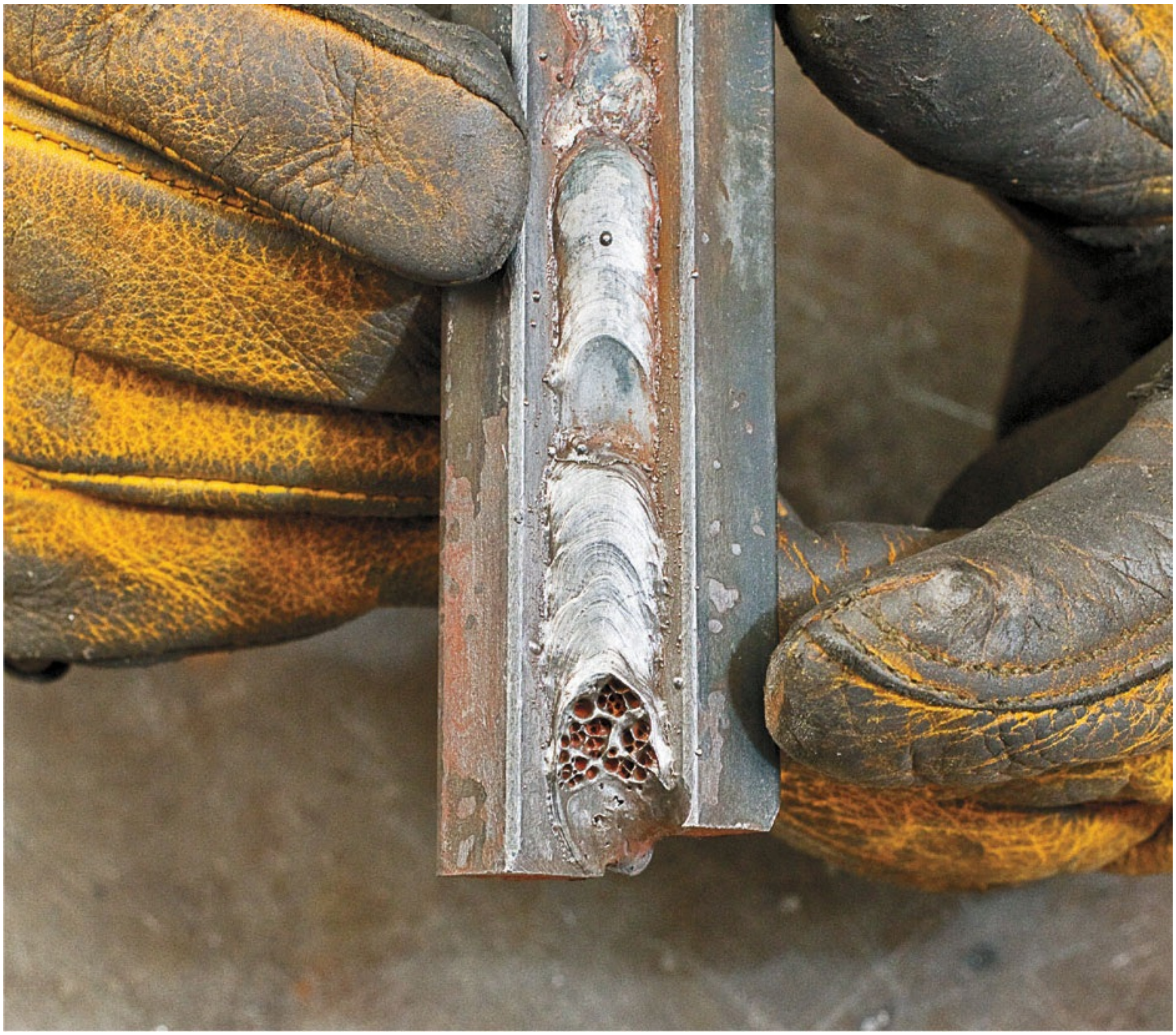

Failure to remove inclusions from the weld zone will cause poor weld quality. Dust particles can off-gas at high temperatures, creating gas pockets that lead to fracturing and poor welds. The effects of inclusions will appear as pores, pitting, and voids on the weld surface known as porosity.

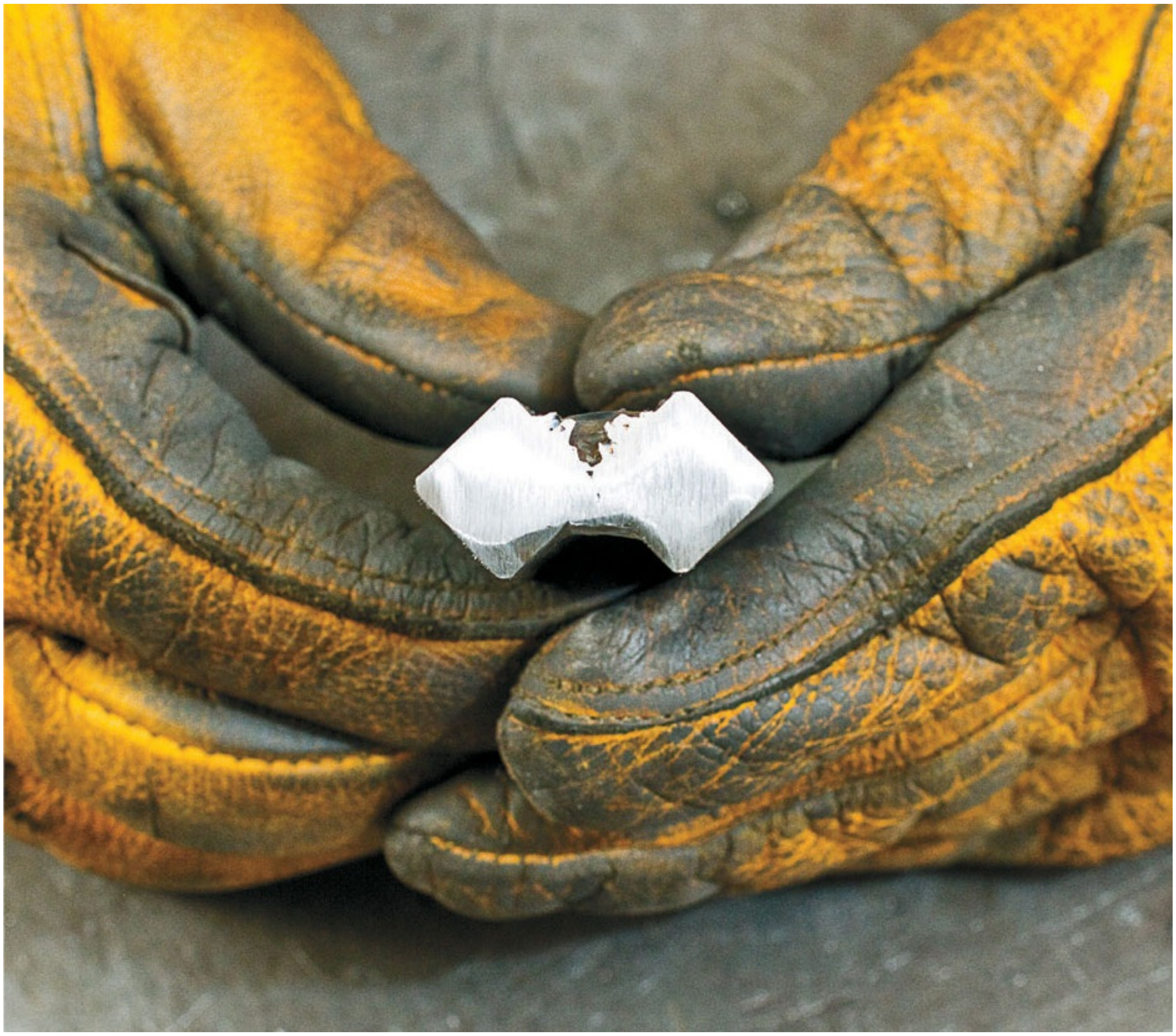

Here is a crosscut of a weld that has deep porosity. This weld is not structurally sound.

You always want to weld on clean raw steel. Prep the weld zone with a flap disk and chamfer if necessary.

There are other methods of cleaning and preparing metal for welding. Media blasting is a great way to remove debris from larger surfaces. Media blasting cabinets can be costly, so it’s not the best place to start. A simple 4 $^1/_2$-inch (11.4 cm) angle grinder will work for what we will be covering in this book. But if you plan on stripping a motorcycle frame or removing a lot of paint and rust from something that size, a media blaster is the way to go.

Manufacturers and large fabrication companies will also use a wide variety of chemicals to remove mill scale and other inclusions from large surfaces.

Chemical removal cleansing methods are extremely corrosive and can be dangerous. They are not recommended for the home hobbyist. If you do plan on using muriatic acid to remove the mill scale from hot roll steel, make sure to take the necessary safety precautions and have a lot of baking soda accessible to neutralize the acid if something does go wrong.

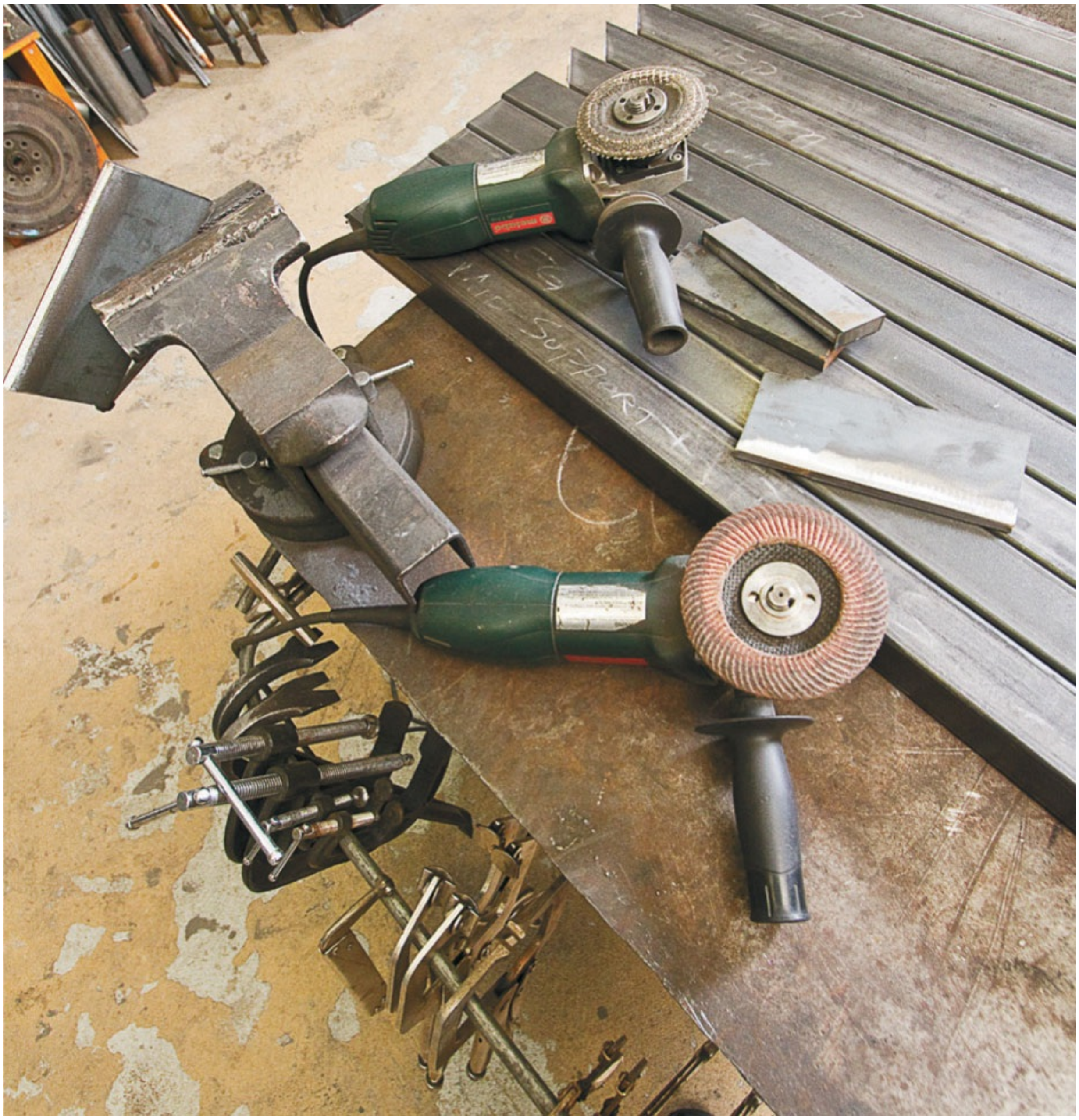

When grinders are not in use, always put them down on the back. Never place a grinder on its front.

A bench grinder with a wire wheel is a nice tool for prepping small metal pieces. It allows you to hold the piece by hand, as in the photo, or with a set of clamps in order to do proper metal preparation.

Let’s Prep

When prepping the metal with a grinder, secure the piece into place with a clamp or in a vice with the area to be prepped accessible. At a proper angle with the grinder and using medium pressure, move in a back-and-forth or up-and-down pattern until the entire area that is to be welded is clean and has a raw-metal appearance.

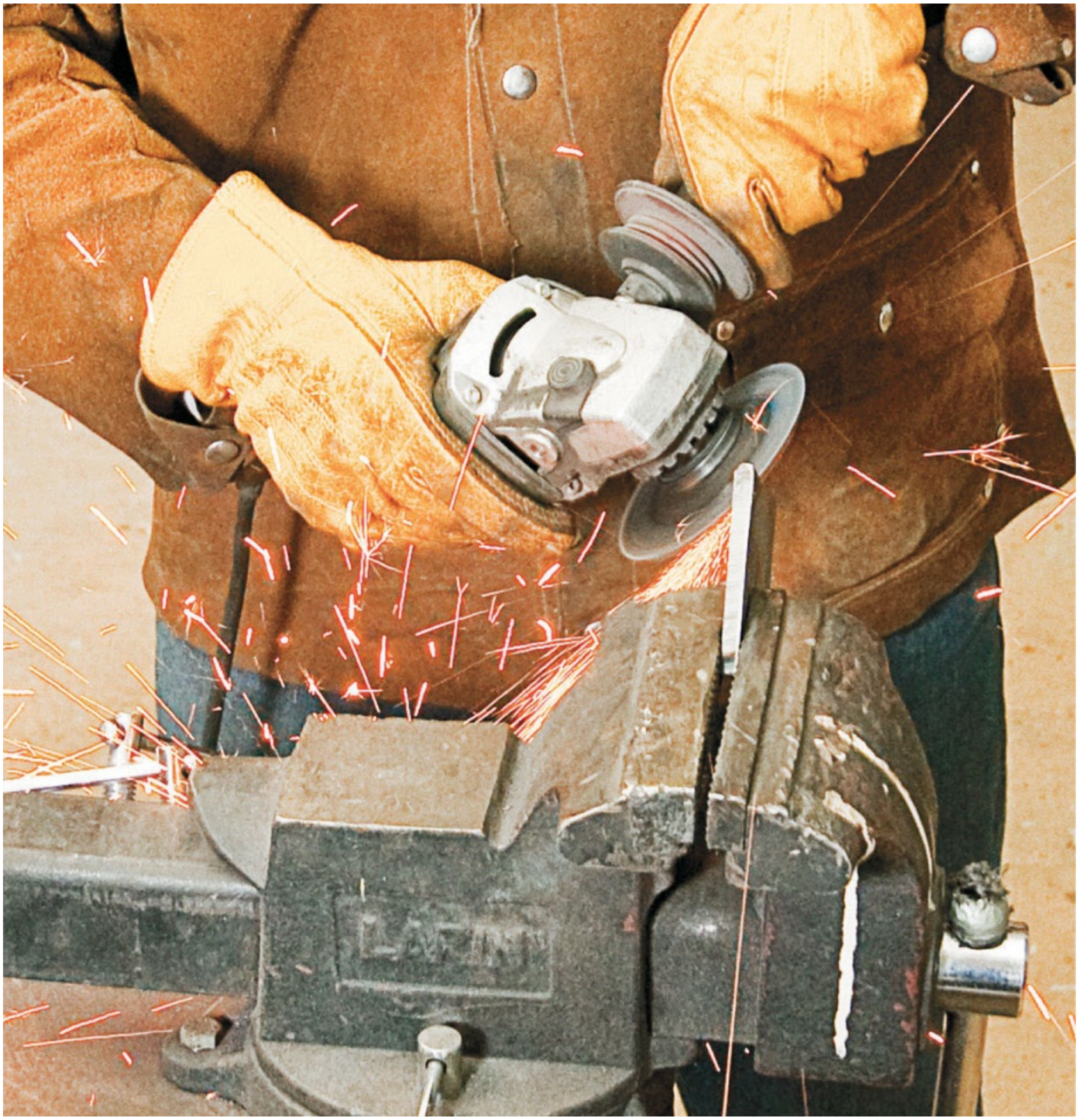

Start a chamfer by setting up the angles and pull 3 or 4 times towards you. This will create a flat surface with the grooves going in the same direction as the disk. Pushing too hard right away will cut into the piece and make a divot.

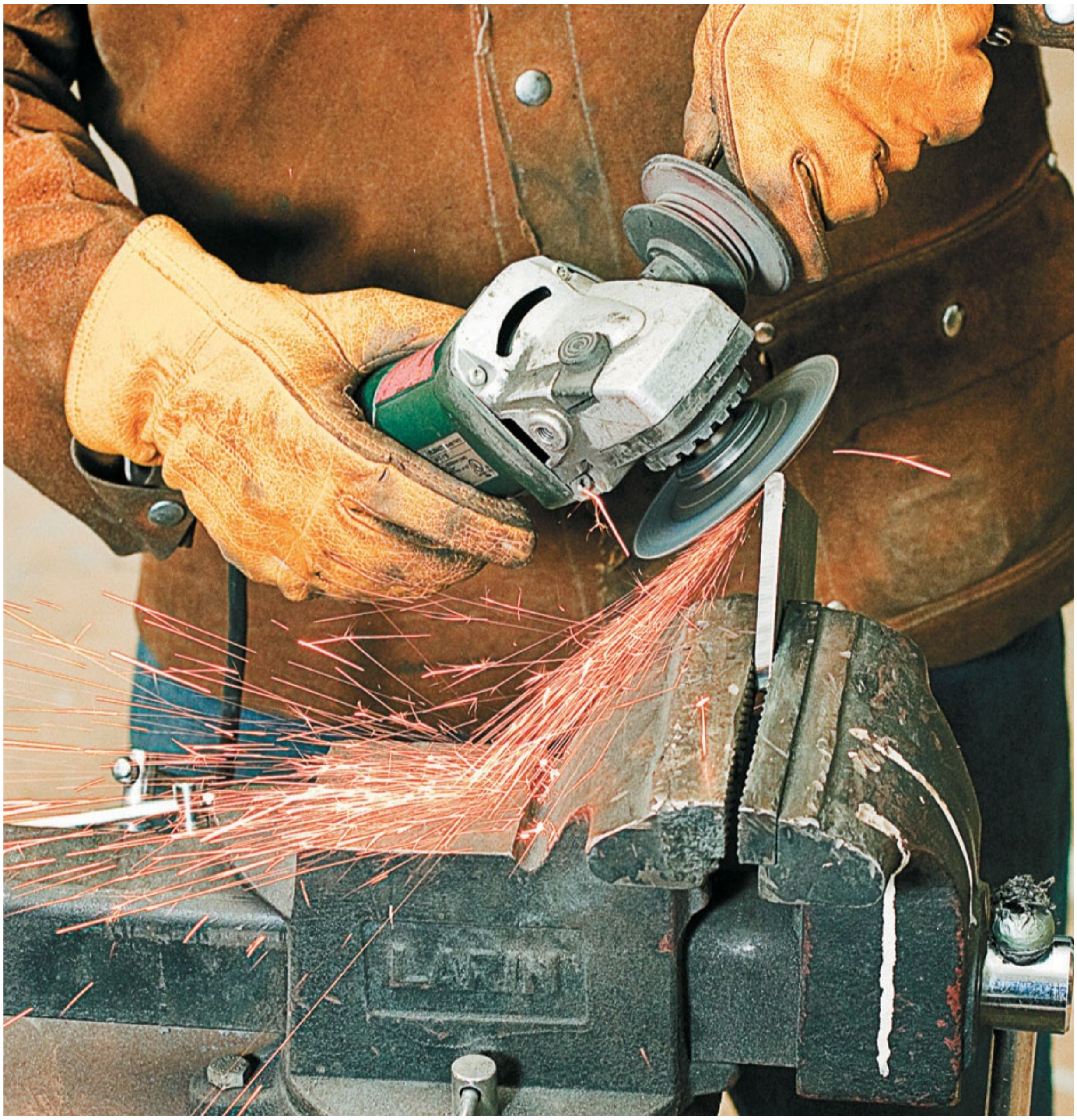

The correct angles for chamfering are 15° to clear the arbor and 45° or greater for the chamfer.

When creating a butt joint for thicker materials, it is necessary to create a chamfer. A chamfer is a bevel ground away from both pieces of base metals at a 45-degree angle to create a channel when the two pieces are put together. This does three very important things: it creates depth for penetration; it creates more surface area for the puddle to liquefy; and it forms a reservoir to receive the filler metal. The most common way to prep the metal with a chamfer is with a stone grinding disk. The stone has a low grit, meaning it will take off a lot of material very quickly as opposed to higher grits that will take material off more slowly but have a nicer finish.

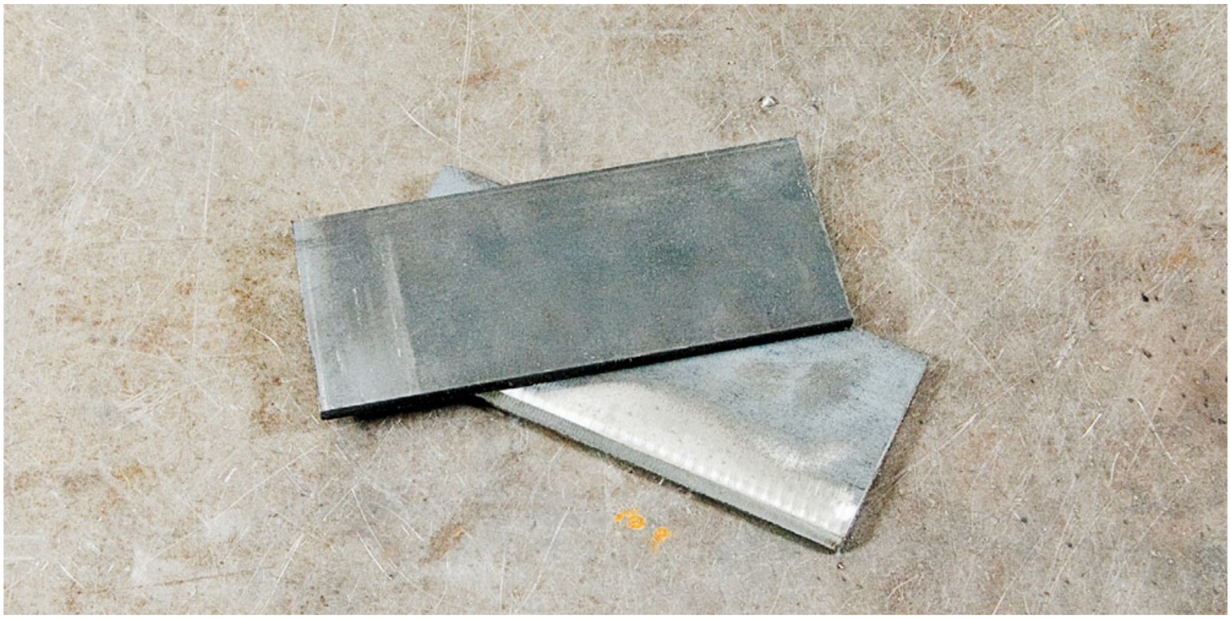
Notice how the bottom piece has been chamfered and prepped.

Just like when cleaning the material, make sure the piece is secured with a vice or clamp. To start a chamfer with a stone in an angle grinder, it is a good idea to make the first few motions a motion toward you with medium pressure. This will put a flat surface on the edge of the metal with the grain going in the direction of the spinning disk, which will make it easier to go back and forth without getting hung up. Grinding too hard right away can cause a cutting action, which will make a divot that you will then have to grind out.

When grinding, remember to be at an appropriate angle to clear the arbor and at the same time be at a 45-degree grinding angle for the chamfers. It is not necessary to grind all the way down to a razor edge, and it is recommended to leave a flat surface for the butting joint. The flat-butted portion of the material will be determined by the thickness of the material. For example, if the material is 1/4 inch (6 mm) thick, I will chamfer and leave 1/32 inch (0.7 mm) on the vertical face. Knowing what you plan to use the joint for will help you determine how the material is to be prepped. In addition to chamfering, leaving a small gap between the base metal is another good method that provides even deeper penetration.

CHANGING THE GRINDER DISK

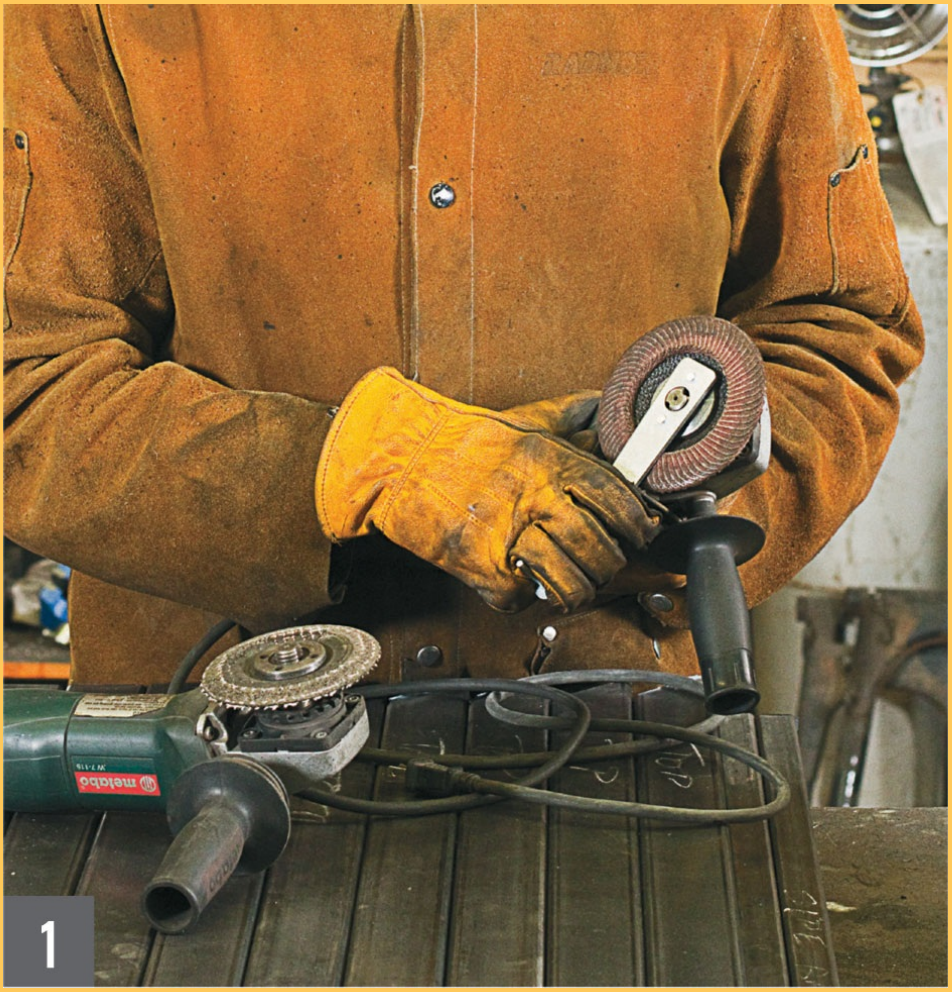
1

Using the arbor wrench, loosen and remove the arbor nut. Notice the grinder is unplugged.

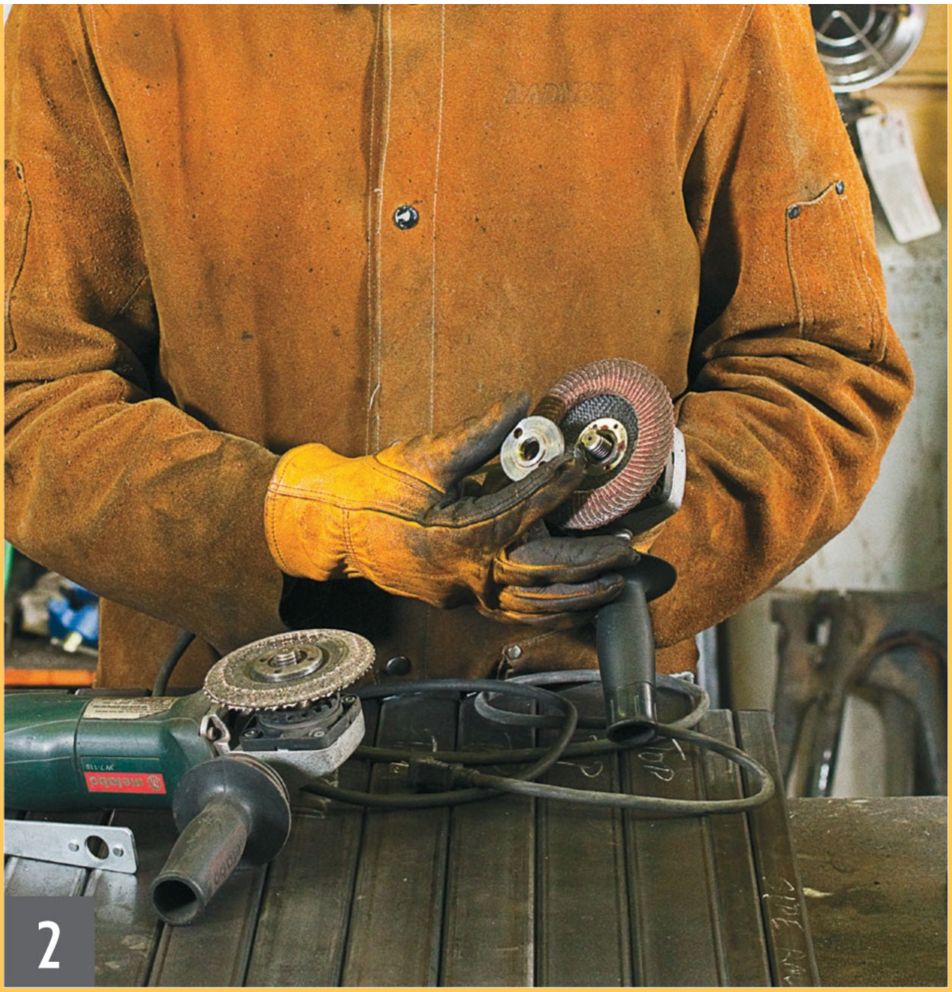

2

Switch out the disk, replace the arbor nut, and tighten by holding down the arbor lock and using the arbor wrench.

When replacing disks, it's always a good practice to unplug the grinder first. Hold down the lock located on the top of the grinder (opposite side from the disk on the grinder head). Using the arbor wrench, loosen the arbor nut and remove. Switch out the disk, replace the arbor nut, and tighten by holding down the arbor lock and using the arbor wrench.

On thinner metal, it is not required to chamfer the base metals. In fact, the thinner the material, the less chamfer or gap it needs. If the metal is extremely thin, such as 22 ga or lower, the MIG should penetrate into the metal enough to create the necessary weld if set properly. You will find that these thin gauges are the most difficult to weld. Most of the time, you will not even be creating a bead, such as in auto body repair, more of a series of stitch welds or tacks. The thinner the material, the less of a gap you need and the less heat the metal can take without warping or creating holes.

When working with a MIG welder, keep a stainless steel wire hand brush around to clean off the materials between a series of tacks or if going back to a bead that you have already created. A carbon-burn byproduct will usually be around a weld or tack that has just been created. Remove the byproduct with the brush to avoid reintroducing this material back into the puddle and weld zone.

4

How It Works

There are a lot of aspects of welding that I love. From the basics of joining two pieces of materials together to create one, working with intense heat, or simply just playing with fire, the truth is I love welding and metalworking because it gives me the ability to manipulate this seemingly unyielding material. Even deeper still is the idea that I have the ability to control the elements that are only formed in the final moments of a dying star to create anything in my imagination. It’s also pretty handy when it comes to fixing things such as cars, motorcycles, or furniture.

Technique

From the very first moment I welded, I had this overwhelming feeling of having the power of the sun in my hand. I had never before felt this sort of power of creation in any other medium that I had felt with metal work: the immense heat, the liquefaction of the steel, the solidification of the bead on the surface, to this day it all still fascinates me. It is a truly amazing process.

I hope you feel the same way. I’m excited to be the one to share this with you. Now let’s get started.

The Puddle

Most people aren’t aware of this, but the primary job of a welder is to control fluid. It happens to be a 3000°F (1649°C) or more fluid, but a fluid nonetheless. When the arc liquefies the base metals it creates a molten pool of liquid metal called the puddle. No mater what technique is used or motion created with the gun, the puddle is the primary focus of controlling how a bead is created.

As a beginning welder progresses, they will find that there are a lot of different techniques to control the puddle to create different effects in the surface of a bead. For the most part, a welder’s responsibility is to make sure the weld penetrates into the metal enough to create a proper weld. Before this can happen, a beginner needs to build a feel for movements that make even and steady motions that will produce a proper bead.

Push or Pull

There are two methods to the madness in welding when it comes to controlling the puddle: pushing and pulling. These terms are decided by the direction of movement in which the weld is created. The “push” method is the movement of traveling into the weld zone so that the weld is created behind the gun nozzle. The “pulling” method (or what I like to call the “draw”) is a dragging movement though the weld zone so that the weld is being created in front of the gun nozzle and contact tip.

Both of these methods have their good points and both should be learned as there are always going to be situations in which one will be better suited than the other. But as a beginner, the best one to start with is the pull method as it will penetrate slightly deeper into the steel than the push method and will also create a narrower bead.

Straight pull welds originated from welding butt joins. Pulling a surface weld helps beginners to practice butt joints and is also the best place to start building a speed of motion to produce a weld. What this does is help you get a better understanding of your hands’ movement in relation to the material’s thickness and the settings of the machine. It also makes you more aware of what you are trying to accomplish after completing a bead.

Developing a good pace or rhythm of the hand movements for the pull will translate to many other techniques that will help you develop as a MIG welder. As you develop your welding skills, you’ll find that there are a lot of variations in technique—the way you hold the gun, or the way you position your body in relation to the weld or angle of view—that affect the way you create a weld. For now, focus on the basics and you will discover these things and acclimate to your own style.

Stance and Comfort

Seventy-five percent of welding is being comfortable, or as comfortable as you possibly can be in any given situation. Whether it’s a welder on an oil rig, someone building a sky scraper, or a hobbyist just fixing a small part of a car, that person will always try to orient him or herself in the most comfortable position possible.

For example, a buddy of mine likes to sit low with his armpits at table level while he TIG welds on racecar parts. I, on the other hand, like to weld sitting on my favorite shop stool. I built my table to the perfect height for me and I’m nice and relaxed—to each their own.

Being comfortable is a critical part of welding. The position of the body, bracing your gun hand, and avoiding hovering with the gun are all crucial points to prevent tension and decrease fatigue. It will also allow your hands to be steadier when welding.

COMFORT TIPS

If you have to stand at the welding table, pay attention to your stance while welding. A good stance for our first exercise is to take a step back with the dominant foot roughly shoulder width. To stay balanced and relaxed, bend with your knees and waist so that your chest is lined up with the ball of the front foot.

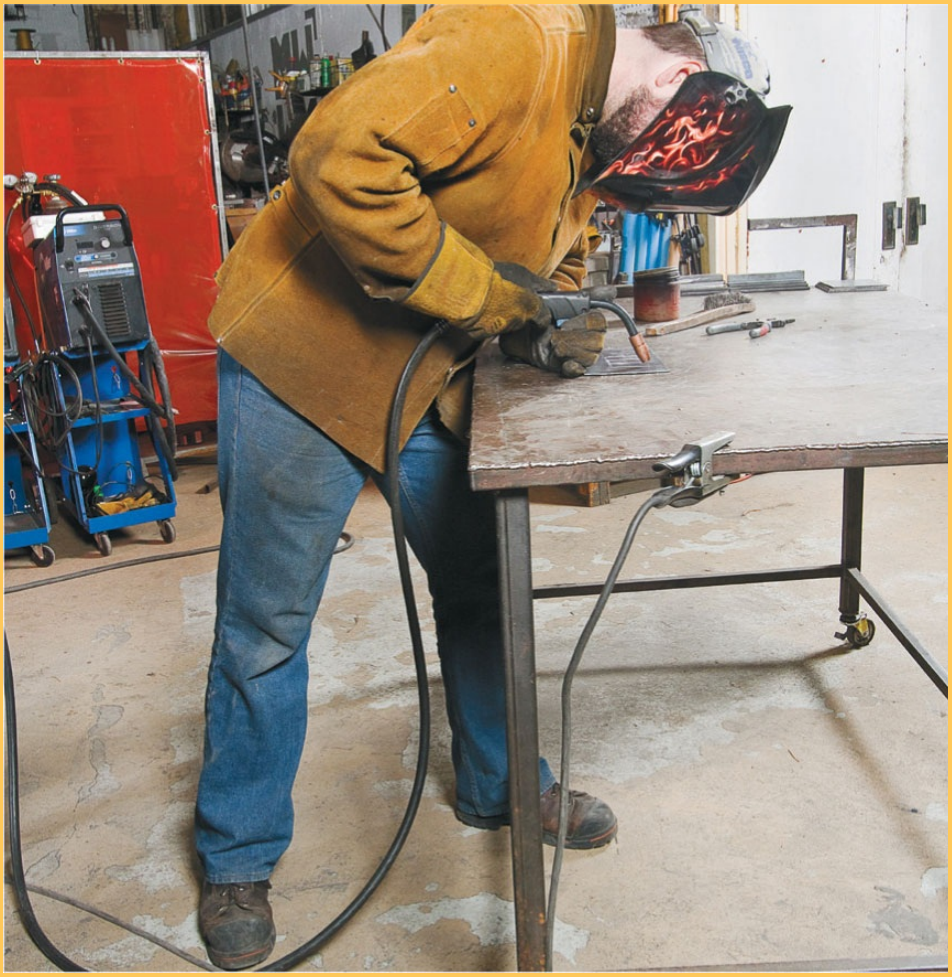

With your dominant foot, take a step back a little more than shoulder width. (If you are right handed, step back with your right foot.)

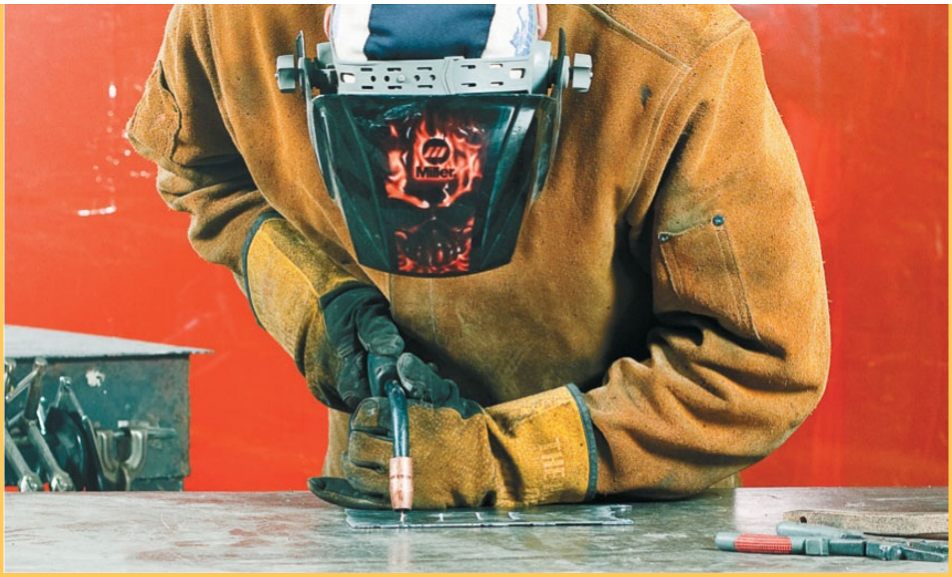

Brace and support the gun at the proper angles.

Vision

The number one complaint from all of my student or beginner welders that I hear all the time is "I can't see anything." I'm going to tell you right now that it will be hard at first. First of all, as a beginner you don't know what you are looking for. Secondly, you haven't trained your mind to really focus on what you need. Additionally, your vision has not become acclimated to adjust to what you need to see right away—yet.

But the more you weld, the more you will know what you are looking for and your mind and eyes will adjust to it faster. You need to see all of the elements that create the weld: the arc, puddle, and the weld zone (or line) to follow. All of these components are key in creating a good weld.

Every welder I have ever met has his or her own methods of seeing. Some even say they go entirely off of feel. Personally, I get real close and look way over my nozzle so I can see everything. To help me see my weld zone (my line), I move my head just a little to the side. Be careful not to go too far to the side. The natural mistake while drawing out the bead if your vision is too far off to the side is to go off the weld zone diagonally to the opposite side.

When you catch yourself leaving the weld zone, stop welding immediately. When you weld outside of the weld zone, all you're doing is creating more work for yourself because you're going to have to grind off that extra metal later if you want to make it look pretty. Stopping and starting is not that big of a deal. Just clean off the weld and start from where you left off.

Just clean off the weld and start from where you left off.

Keep Your Helmet Clean

Make sure your welding helmet is clean. When my outer clear shield gets fogged or scratched up, I use a scratch remover/polish that is formulated to work on plastic. Keep the shields clean both inside and out. And if you normally wear reading glasses, wear them under the lid. When welding on thin materials, the puddle and arc are much smaller, and the lack of arc light will make it harder to see the weld in progress.

Hand Position

The way that the gun is held and supported will become personalized over time to you and the style that you develop, and it’s one of the more important things to keep relaxed and steady. Never hold the gun with one hand or hover with both hands while welding, as the tension you’ll create in your hands translates into shaky welding. Instead, always support the trigger hand with the opposite hand for stability when welding.

The dominant hand will normally be the trigger hand. The opposite hand is going to be the brace and support hand. It will also be the hand that will drag, so it plays a lot of important rolls when drawing out a bead. Don’t support the weight of your body on your hands; just rest them on the table so that they can slide easily. You’ll be making a motion similar to wiping sand off the table.

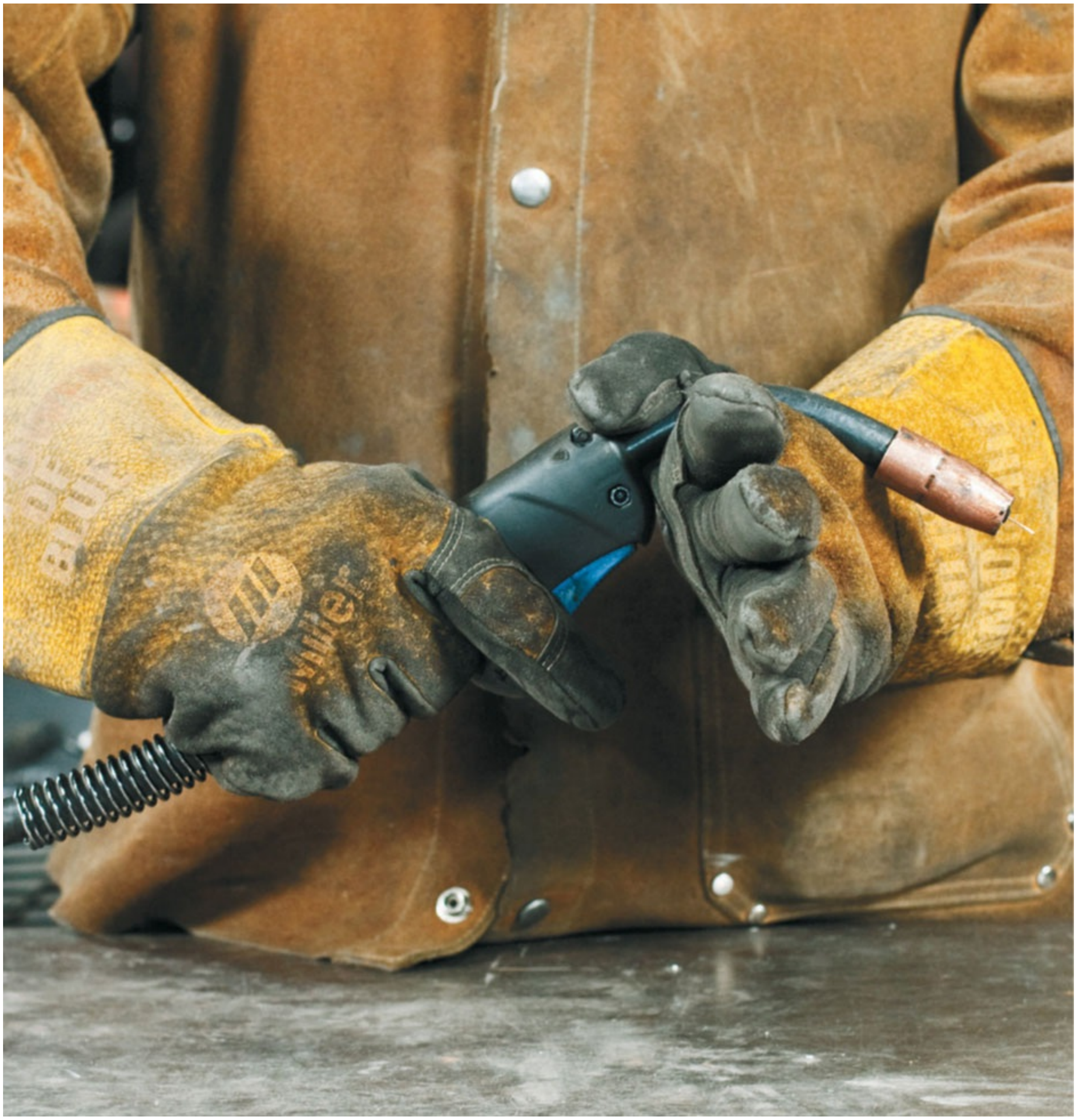

For best results, grab the gun handle with the dominant hand so that the index finger can depress the trigger. With the opposite hand, grab the neck of the gun in front of the gun body, leaving the hand open and only gripping the neck with the thumb.

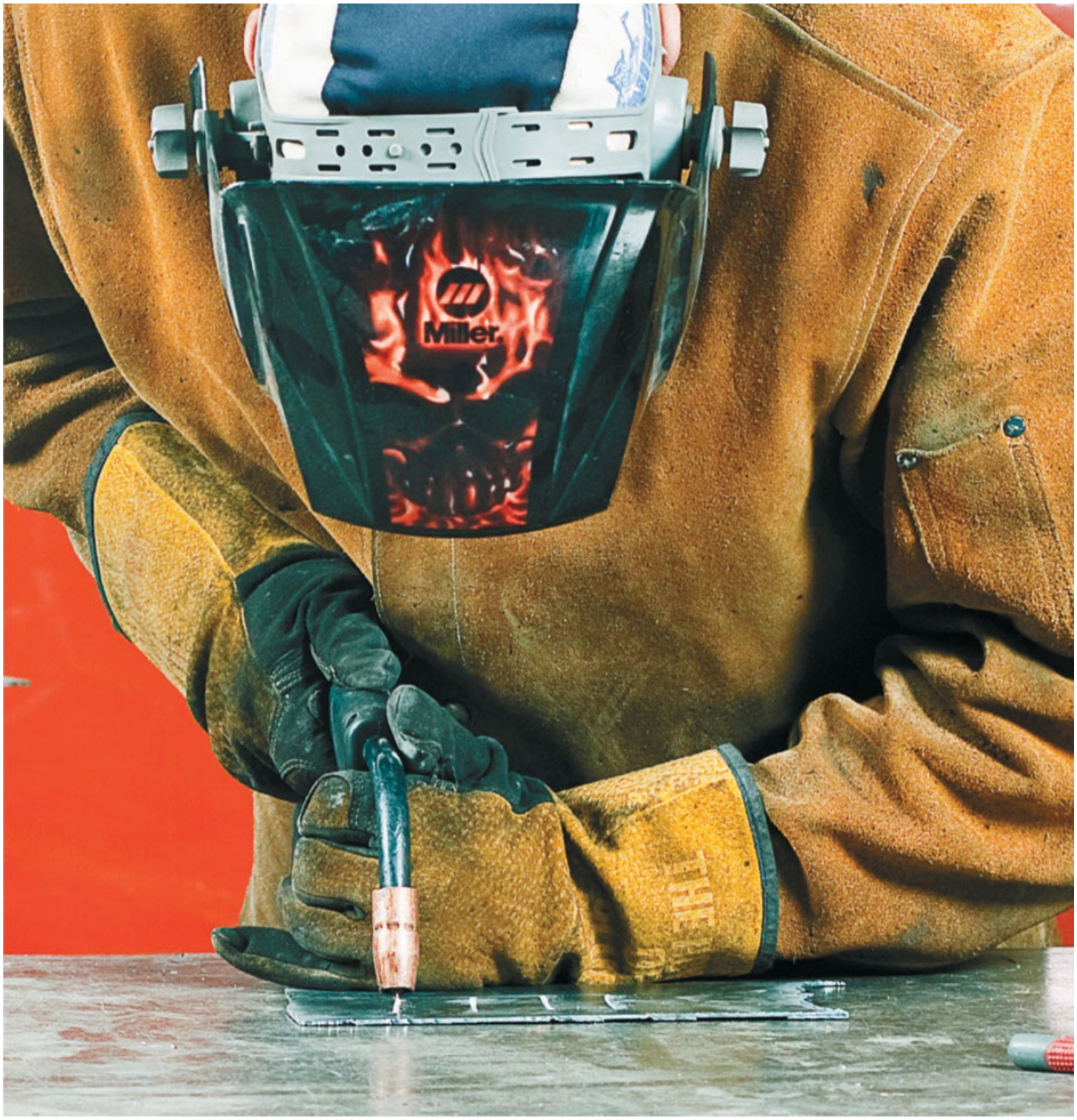

When you get into position, place the heel (or edge) of your hand on the table to support the gun. Do not lean on your hands.

Angle of the Gun

When lining up the gun, the nozzle should be straight on to the weld zone and perpendicular. This means that the nozzle of the gun should be pointing in the same direction as the weld zone with an even 90 degrees on both sides.

These two angles are very important for the proper distribution of the shield gas. If the gun is not straight and perpendicular to the weld, several things will go wrong: the shield gas will not protect the puddle properly; and the flow of the go wrong: the shield gas will not protect the puddle properly, and the flow of the gas over the puddle and weld will create opportunities for inclusions, uneven heat, and poor control. This becomes more of an issue when welding on thicker materials.

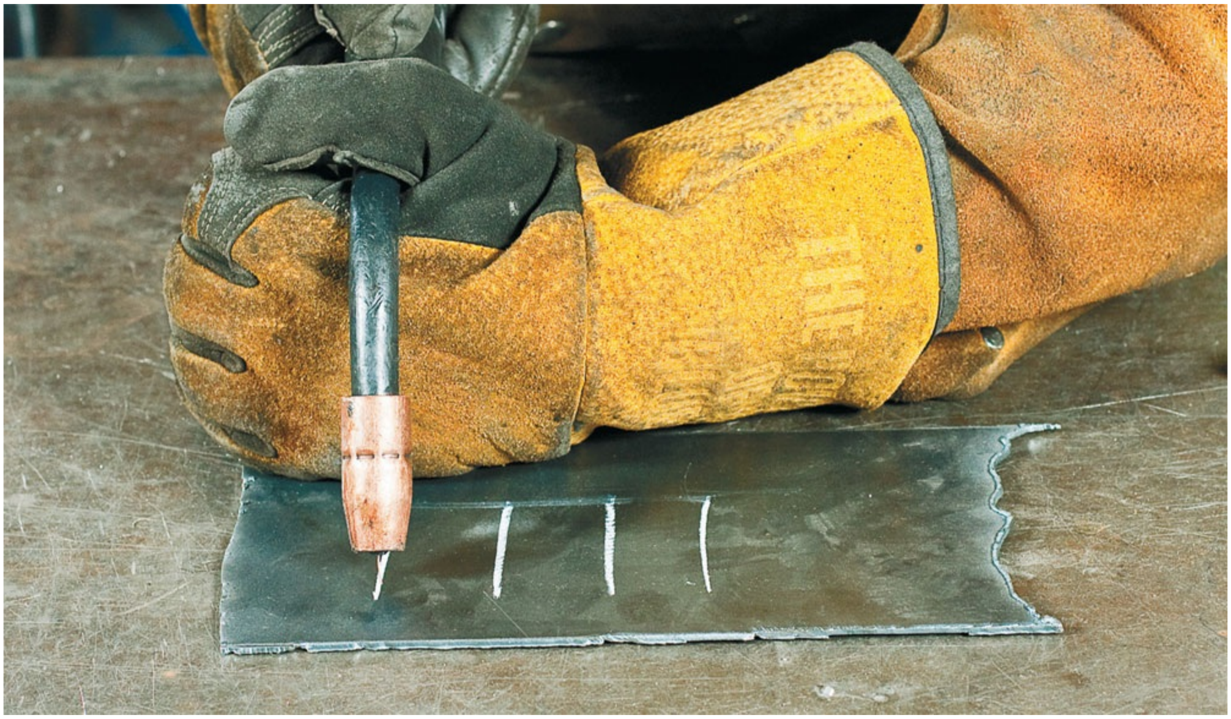

Being perpendicular with your drag angle pointing straight into the weld will allow the shield gas to hit the puddle evenly on both sides. Note: To avoid putting too much heat into the table you are welding on, it's a good idea to put risers under the sheet metal, such as two small pieces of square bar. (This is not shown in the image above.)

The third angle to pay close attention to is the travel or drag angle. When first starting out, this angle is a little bit looser, and you can play with it a little. The manufacturers of most MIG machines have put a bend in the gun neck with a decent travel angle (normally 45°), as long as the top of the gun handle is parallel with the surface that is to be welded. You will find that increasing the drag angle to 75 or even 80 degrees will give you better penetration for certain applications. But for now, let’s use the neck angle for our drag angle in order to develop vision and control.

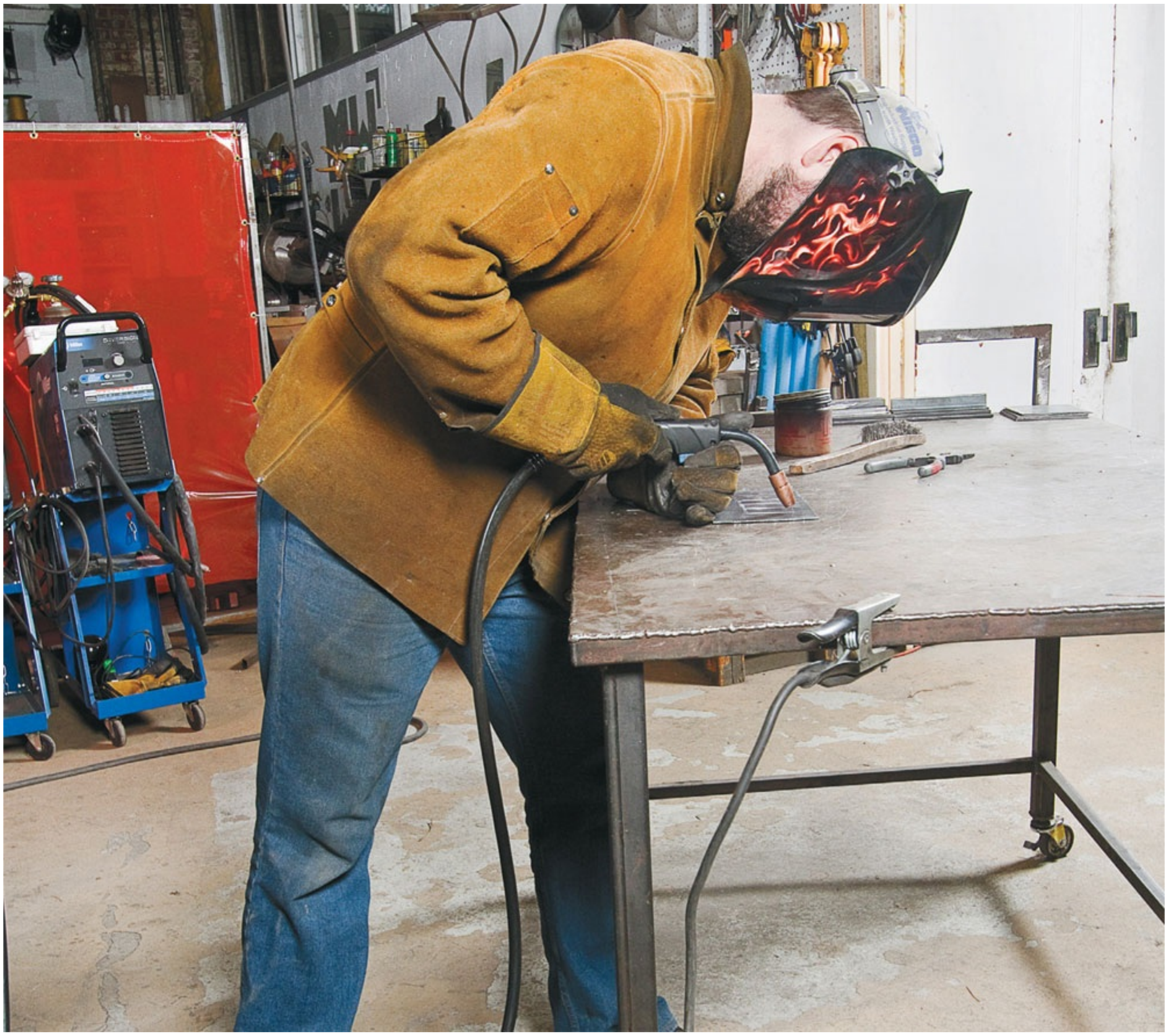

The back of the gun handle should be parallel to produce the correct drag angle. Do not drop the gun hand below the parallel point. This would direct the shield gas to push the bulk of the puddle up on to the back of the bead that was just created, instead of down and assisting with penetration.

A lot of people first starting out will set up and support themselves with their elbow. Now, I don’t have a problem with this at first, but don’t get used to doing this. It sets you up to be a short distance welder because you can only weld so far before you have to set up again and move your elbow. I would recommend getting used to dragging the heel of the support hand on the surface of the piece or table. This will help to avoid unnecessary stopping and starting.

Heat vs. Thickness

One of the jobs of a welder is to know how to control the amount of heat going into the material in relation to the material's thickness. This is done in several ways. The machine voltage and wire speed combination is normally set to the thickness of the metal. Additionally, the speed of the puddle moving through the weld zone, the technique used, the length of time in a given area, and shield gas settings are all components of control to help monitor the puddle and the heat moving through the weld.

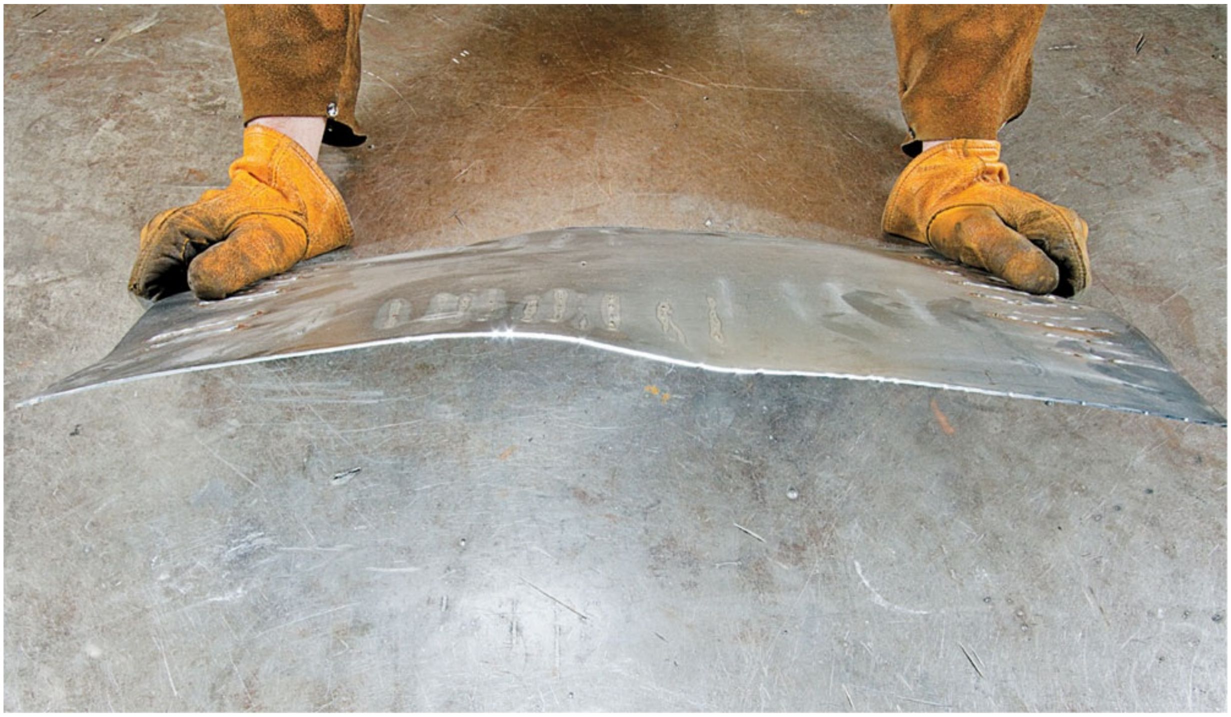

When metal is heated, it wants to expand away from the heat. As it cools, it wants to come back. Depending on where you are in the weld and how much heat is put into one given area can cause variation of heat that produces warpage, as seen in the 14 ga steel above.

The area that is heated is known as the heat-affected zone (or HAZ). There is actually a mathematical formula that involves the factors of welding such as voltage, current, heat intensity, etc., but it's not mandatory to welding that you know this formula. What you do need to know is that metal moves when it is heated. Most of the time it will begin by moving away from the heat; then as it cools it wants to come back. The amount of heat going into the steel, and the time it takes to cool, will normally determine the plus and minus of the metal's movement or what is referred to as warpage.

Steel normally starts off cold. As you begin to weld, the heat radiates through the material from the point where the weld began. As the heat moves though the weld zone it creates variation in heat. The starting point is also now starting to cool, causing another variation of temperature. The longer the weld, the more heat is produced, and the more variation in the metal, the more potential for warping.

As you can see, the sheet of 14 ga steel on page 93 is completely warped from the massive amount of heat that went into the material. This once flat piece of sheet metal has been completely deformed by random amounts of heat going into the surface in various places.

The thinner the material, the less mass it has; this decreases the material’s ability to carry the heat away from the HAZ, and it becomes more susceptible to warpage. Thinner materials will warp much faster than thicker materials. The thinner you go, the more likely the heat will burn through as well. This is a particular concern in auto body repair, as these thin-gauge sheet metals can really only be tacked together.

In contrast, thicker materials have much more mass to help spread the heat out to decrease the HAZ. This means a longer bead can be maintained. Look at it this way: you don’t want to weld much more than a 2-inch (5 cm)-long bead on sheet steel 14 ga or the danger of warping and drop through increases. As the material gets thinner, the length in which a bead can be created decreases as well. Inversely, as the material becomes thicker, the longer the length of a bead and the more heat the material can maintain.

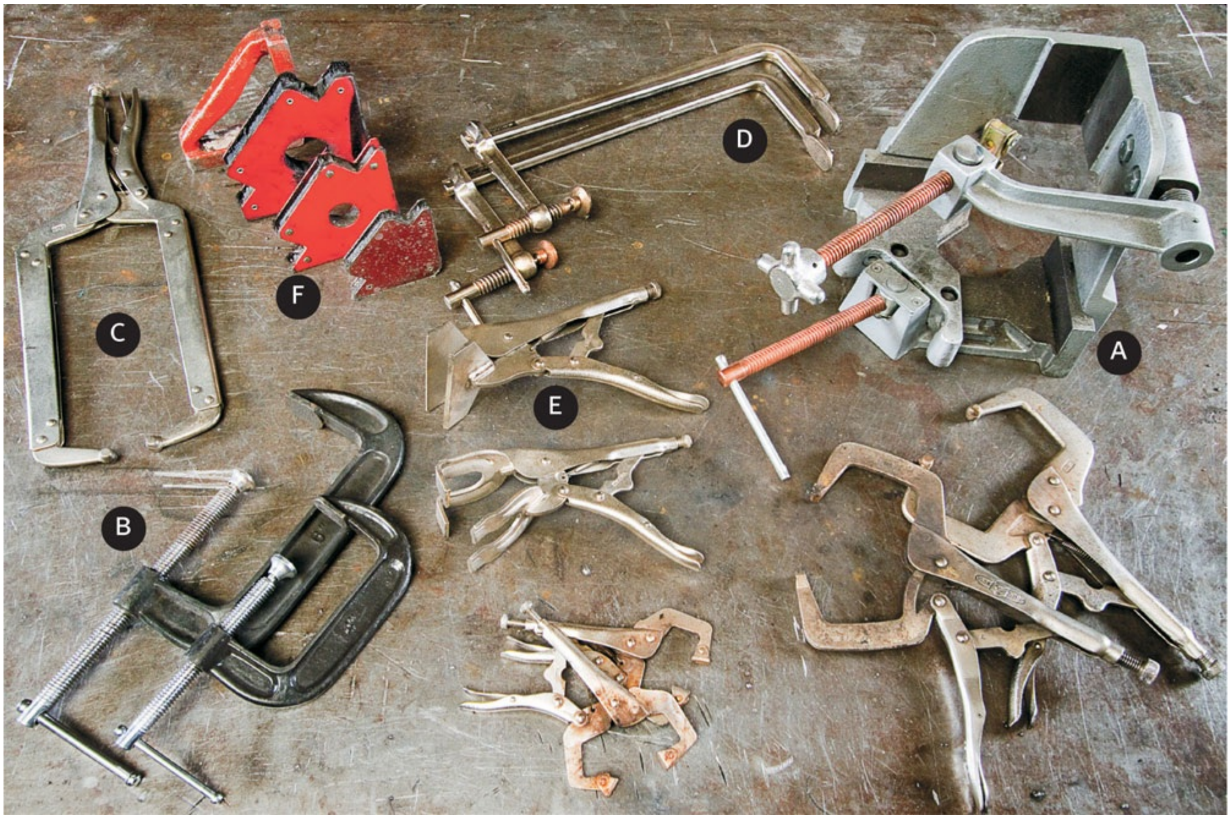

There are a wide variety of clamps, magnets, and vices that help to prevent movement during tacking and welding. 3-axis clamps (A) are great for making tables. C-clamps (B), vise grips (C), F-clamps (D), and locking sheet metal clamps (E) are great for creating jigs or securing the piece to a table. Square magnets (F) are great for keeping the angle in quick tacking situations.

Clamps and Tacking

To prevent warpage while fabricating a piece, a good welder will have a lot of different clamps handy. Clamping a metal work piece down is one of the most common methods to help prevent movement while welding the piece together. Clamps also create stability for the welder if they accidentally bump it out of placement.

There are many styles of clamps, and they all have their purpose. C-clamps, vise grips, adjustable clamps, and even magnets can help you hold the pieces in place while you tack them together. The more you can minimize movement the better.

After the pieces are clamped together, it’s never a good idea to just start welding straight away. Even when the metal is clamped, it will still move whether you like it or not. To prevent movement even further, the next step is to tack the pieces of base metal together.

If you want to completely weld the surface of a table all the way around, you must first tack the surface down. For example, a sheet of 14 ga material can only be welded for a couple of inches before it starts to warp. To prevent this, spread out the heat. Start by tacking in the middles first, and then the corners, then splitting the distance in half until the tacks are a couple of inches apart. Then weld tack to tack to spread out the heat as you go. You have a lot to weld, so don’t keep too much heat in the same area.

Tacking is a series of short welds that will help hold the individual components in place while you weld the piece. Tacking allows you to put the entire piece, or portion of the piece, together, and it acts as a clamping tool. This provides the opportunity to remeasure before welding the piece together permanently. Tacking also serves as a buffer for any mistakes you may have made in measurement. If you see after doing a few tacks that you are off by an unacceptable amount, it’s not a problem to chop through a few tacks with a grinder. But if you had a full weld on there, not only would you have the warping to deal with, but the measurement would be off as well, and you would have to cut though an entire weld and start over.

The most common way to prevent warpage, besides knowing the limitations of the steel, is to spread out the heat. This means welding in multiple areas for shorter periods of time. For example, let’s say I wanted to seal up this 14 ga sheet metal to the top of this welding table.

You wouldn’t just pick a corner and start welding. First you would clamp and tack the sheet metal down to the frame. The first series of tacks would go in a cross pattern to balance any movement from the heat. Start in the middle of a side, then straight across to the middle of the other side. Then repeat the same steps for the adjacent sides.

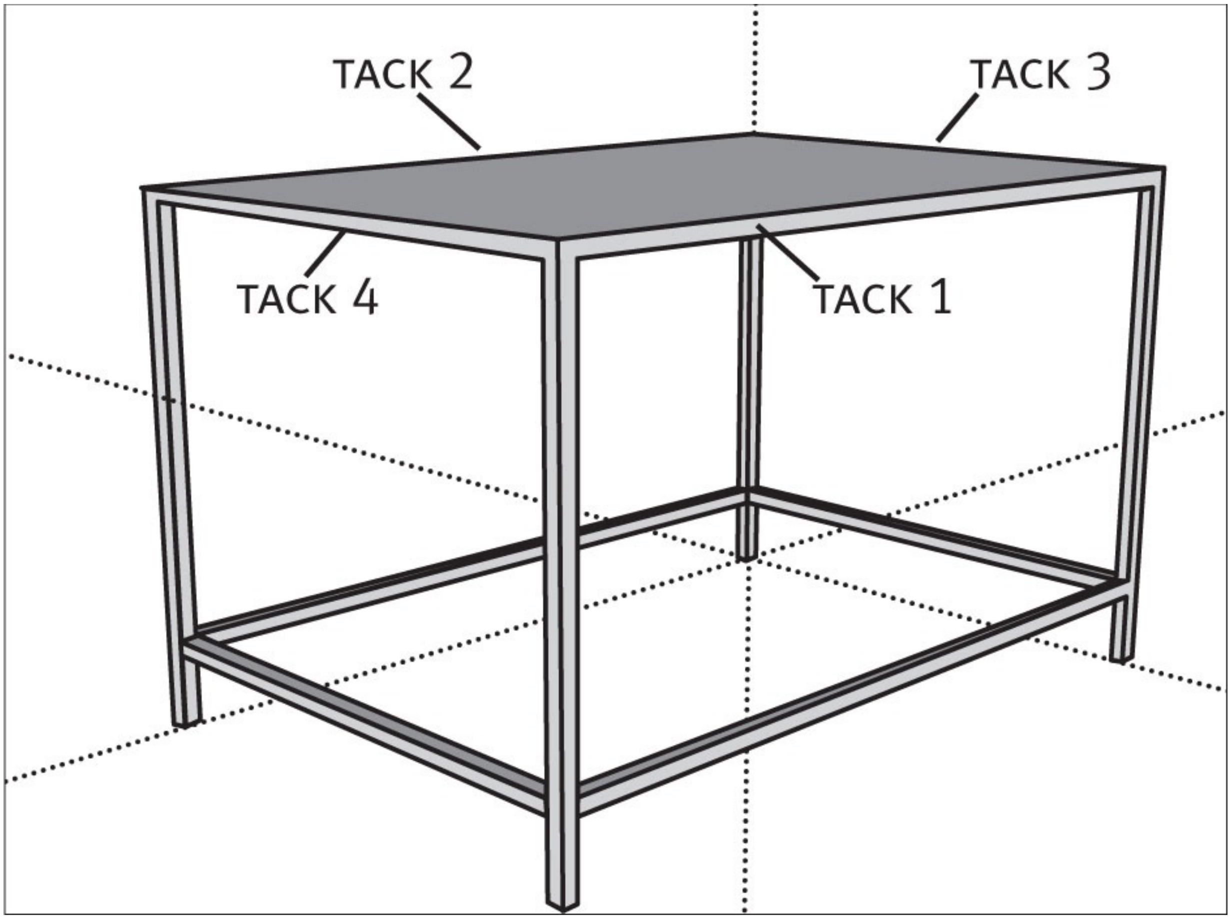

This is good tack pattern for this 14 ga top. Tack the first four corners and then check your measurements again.

After making sure the sheet metal is in the correct position, tack the corners; then continue a series of tacks by splitting the difference of the distance between the tacks by half repeatedly until the tacks all the way around the table are roughly 2 inches (5 cm) apart.

Weld the sheet metal to the frame by welding from tack to tack, spreading out the welds to prevent warping. Shown in the diagram above is a cross pattern from corner to corner and then side to side. The next welds would be in between the first series of welds in a similar pattern.

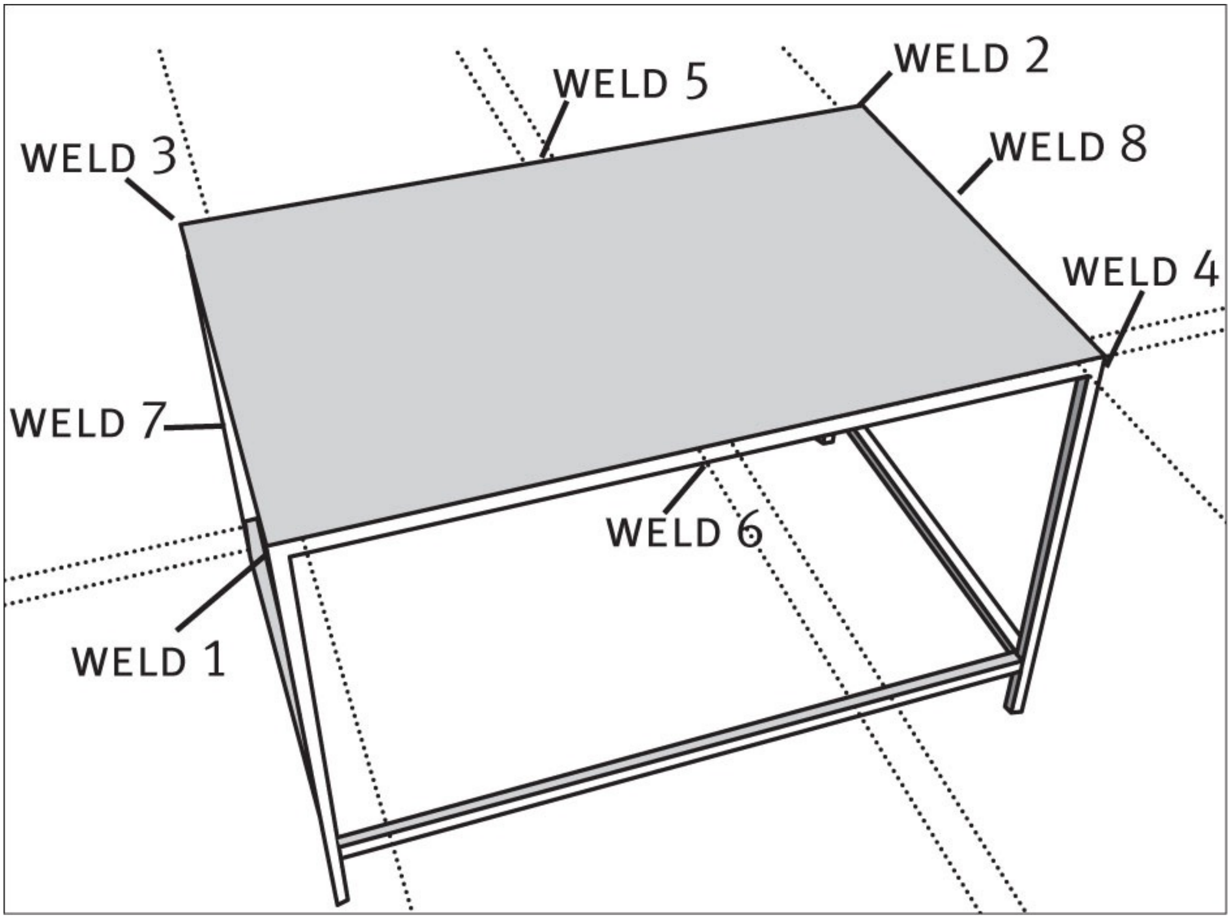

Now weld tack to tack. After welding 2 inches (5 cm), move a minimum of 8 inches (20 cm) away before you start your next weld. You have a lot of area, so the more you can spread out your welds, the less likely it is the material will warp.

Welding is just like changing the tire on a car: you never tighten the lugs by going in a circle but by going in a cross or star pattern to spread the compression out evenly. The same concept applies to welding. Start welding tack to tack at one corner and then go to the opposite corner, then the middle, spreading out the heat.

There are a lot of schools of thought when it comes to spreading the heat. Another method of spreading out heat, for example, on a sheet of 14 ga steel is to weld for 2 inches (5 cm) from tack to tack and then move 4 to 7 inches (10.2 to 17.8 cm) down the same edge and make another 2 inch (5 cm) weld. Seven to eight inches (17.8 to 20.3 cm) should provide enough distance to move past the HAZ.

The way I look at it, if you're welding on a big enough piece, then why not move further away. I like to err on the side of caution and do a star pattern to really spread out the heat. That way, by the time I get back to the original weld, it has cooled enough to continue the weld. This may be overkill, but like I said, I err on the side of caution, especially when making a table.

Tacking is one of the most important steps in fabrication. Always tack a structure or piece together before welding any part. This will ensure the piece stays as true as possible.

Basic Rules and Principles for Tacking and Welding

TACKING

- Always tack the piece together first before welding, as tacking acts as a clamping mechanism and secures the metal in place, preventing it from moving while you weld.
- Always secure the pieces to be connected in place with clamps.
- Tacking does not require arc movement, so an increased angle will help penetration.

WELDING

- The angle of the nozzle must be straight on and perpendicular to allow the shield gas to hit both sides of the puddle evenly.
- The drag angle should never be less than a 45-degree angle.
- Brace and support the gun hand with the opposite hand.
- Always be as comfortable as possible.

MATERIAL

- Steel only has limited ability to radiate the heat away from the puddle before it becomes too great and "burns though."
- The thicker the material, the further the distance a bead can be created. The thinner the material, the shorter the distance.
- For very thin metal such as sheet metal for auto body, you can't draw a bead —just use short tacks.
- Control the amount of heat you are putting into the steel by limiting the distance of your bead; then move away from the that point of heat when tacking or welding.

Exercise 1: The Draw

DIFFICULTY: BEGINNER

MATERIALS AND EQUIPMENT:

- MIG welder
- 5” x 10” (12.7 x 25.4 cm) piece of 16 ga (minimum) sheet metal
- Welding Helmet
- Welding gloves
- Long-sleeve shirt
- Leather welding jacket or apron
- MIG pliers
- Stainless steel wire brush
- Soapstone

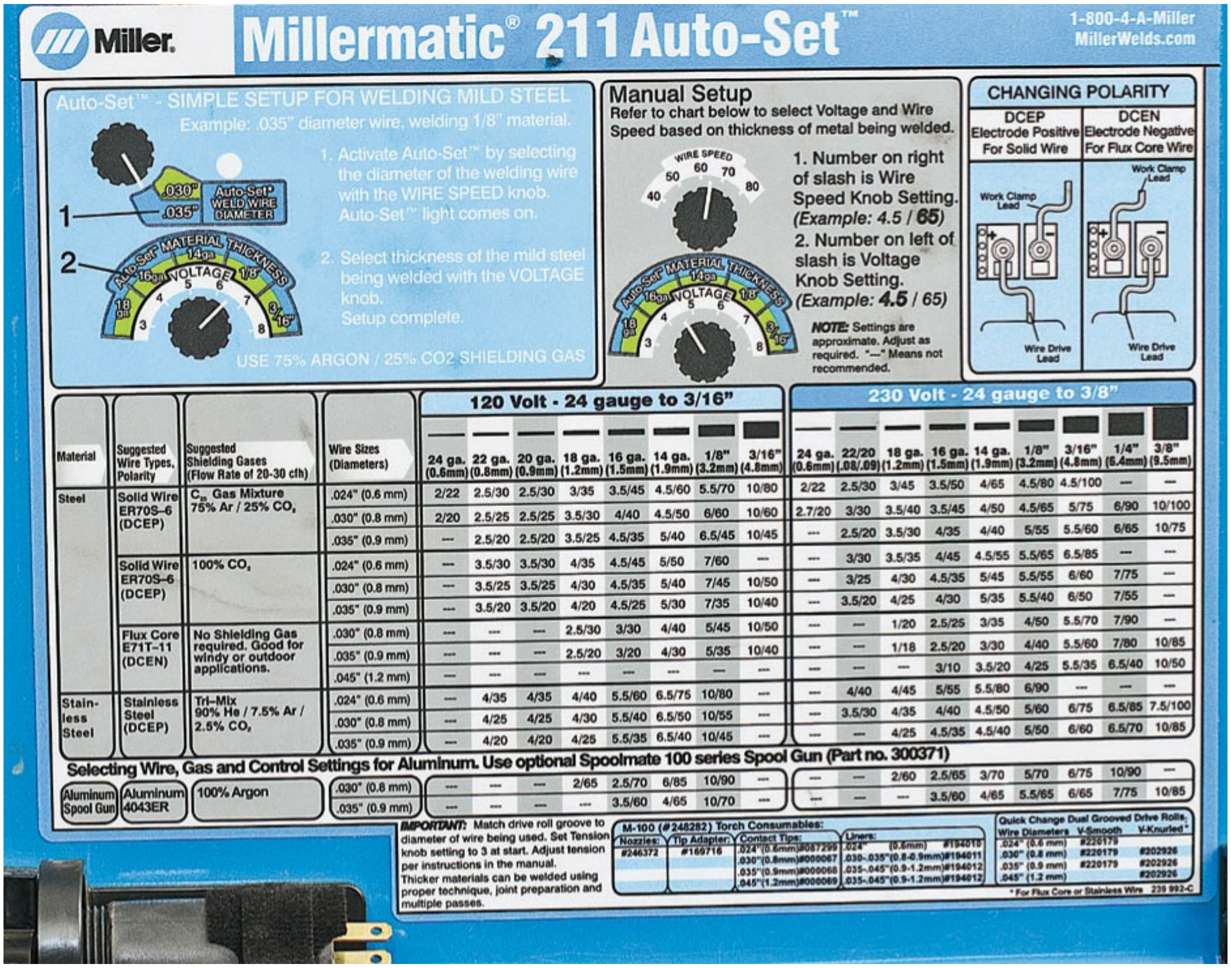

Miller Millermatic® 211 Auto-Set™

1-800-4-A-Miller
MillerWelds.com

Auto-Set™ - SIMPLE SETUP FOR WELDING MILD STEEL
Example: .035" diameter wire, welding 1/8" material.

1. Activate Auto-Set™ by selecting the diameter of the welding wire with the WIRE SPEED knob. Auto-Set™ light comes on.
2. Select thickness of the mild steel being welded with the VOLTAGE knob. Setup complete.

USE 75% ARGON / 25% CO2 SHIELDING GAS

Manual Setup
Refer to chart below to select Voltage and Wire Speed based on thickness of metal being welded.

1. Number on right of slash is Wire Speed Knob Setting. *(Example: 4.5 / **65**)*
2. Number on left of slash is Voltage Knob Setting. *(Example: **4.5** / 65)*

NOTE: Settings are approximate. Adjust as required. "---" Means not recommended.

CHANGING POLARITY

DCEP	DCEN
Electrode Positive For Solid Wire	Electrode Negative For Flux Core Wire

Material	Suggested Wire Types, Polarity	Suggested Shielding Gases (Flow Rate of 20-30 cfh)	Wire Sizes (Diameters)	120 Volt - 24 gauge to 3/16"								230 Volt - 24 gauge to 3/8"								
				24 ga. (0.6mm)	22 ga. (0.8mm)	20 ga. (0.9mm)	18 ga. (1.2mm)	16 ga. (1.5mm)	14 ga. (1.9mm)	1/8" (3.2mm)	3/16" (4.8mm)	24 ga. (0.6mm)	22/20 (.08/.09)	18 ga. (1.2mm)	16 ga. (1.5mm)	14 ga. (1.9mm)	1/8" (3.2mm)	3/16" (4.8mm)	1/4" (6.4mm)	3/8" (9.5mm)
Steel	Solid Wire ER70S–6 (DCEP)	C₂₅ Gas Mixture 75% Ar / 25% CO₂	.024" (0.6 mm)	2/22	2.5/30	2.5/30	3/35	3.5/45	4.5/60	5.5/70	10/80	2/22	2.5/30	3/45	3.5/50	4/65	4.5/80	4.5/100	---	---
			.030" (0.8 mm)	2/20	2.5/25	2.5/25	3.5/30	4/40	4.5/50	6/60	10/60	2.7/20	3/30	3.5/40	3.5/45	4/50	4.5/65	5/75	6/90	10/100
			.035" (0.9 mm)	---	2.5/20	2.5/20	3.5/25	4.5/35	5/40	6.5/45	10/45	---	2.5/20	3.5/30	4/35	4/40	5/55	5.5/60	6/65	10/75
	Solid Wire ER70S–6 (DCEP)	100% CO₂	.024" (0.6 mm)	---	3.5/30	3.5/30	4/35	4.5/45	5/50	7/60	---	---	3/30	3.5/35	4/45	4.5/55	5.5/65	6.5/85	---	---
			.030" (0.8 mm)	---	3.5/25	3.5/25	4/30	4.5/35	5/40	7/45	10/50	---	3/25	4/30	4.5/35	5/45	5.5/55	6/60	7/75	---
			.035" (0.9 mm)	---	3.5/20	3.5/20	4/20	4.5/25	5/30	7/35	10/40	---	3.5/20	4/25	4/30	5/35	5.5/40	6/50	7/55	---
	Flux Core E71T–11 (DCEN)	No Shielding Gas required. Good for windy or outdoor applications.	.030" (0.8 mm)	---	---	---	2.5/30	3/30	4/40	5/45	10/50	---	---	1/20	2.5/25	3/35	4/50	5.5/70	7/90	---
			.035" (0.9 mm)	---	---	---	2.5/20	3/20	4/30	5/35	10/40	---	---	1/18	2.5/20	3/30	4/40	5.5/60	7/80	10/85
			.045" (1.2 mm)	---	---	---	---	---	---	---	---	---	---	---	3/10	3.5/20	4/25	5.5/35	6.5/40	10/50
Stainless Steel	Stainless Steel (DCEP)	Tri–Mix 90% He / 7.5% Ar / 2.5% CO₂	.024" (0.6 mm)	---	4/35	4/35	4/40	5.5/60	6.5/75	10/80	---	---	4/40	4/45	5/55	5.5/80	6/90	---	---	---
			.030" (0.8 mm)	---	4/25	4/25	4/30	5.5/40	6.5/50	10/55	---	---	3.5/30	4/35	4/40	4.5/50	5/60	6/75	6.5/85	7.5/100
			.035" (0.9 mm)	---	4/20	4/20	4/25	5.5/35	6.5/40	10/45	---	---	---	4/25	4.5/35	4.5/40	5/50	6/60	6.5/70	10/85

Selecting Wire, Gas and Control Settings for Aluminum. Use optional Spoolmate 100 series Spool Gun (Part no. 300371)

Aluminum Spool Gun	Aluminum 4043ER	100% Argon	.030" (0.8 mm)	---	---	---	2/65	2.5/70	6/85	10/90	---	---	---	2/60	2.5/65	3/70	5/70	6/75	10/90	---
			.035" (0.9 mm)	---	---	---	---	3.5/60	4/65	10/70	---	---	---	---	3.5/60	4/65	5.5/65	6/65	7/75	10/85

IMPORTANT: Match drive roll groove to diameter of wire being used. Set Tension knob setting to 3 at start. Adjust tension per instructions in the manual.
Thicker materials can be welded using proper technique, joint preparation and multiple passes.

M-100 (#248282) Torch Consumables:

Nozzles:	Tip Adapter:	Contact Tips:		Liners:		
#246372	#169716	.024"(0.6mm)	#087299	.024"	(0.6mm)	#194010
		.030"(0.8mm)	#000067	.030-.035"	(0.8-0.9mm)	#194011
		.035"(0.9mm)	#000068	.035-.045"	(0.9-1.2mm)	#194012
		.045"(1.2mm)	#000069	.035-.045"	(0.9-1.2mm)	#194012

Quick Change Dual Grooved Drive Rolls:

Wire Diameters	V-Smooth	V-Knurled*
.024" (0.6 mm)	#220179	
.030" (0.8 mm)	#220179	#202926
.035" (0.9 mm)	#220179	#202926
.045" (1.2 mm)		#202926

* For Flux Core or Stainless Wire 239 992-C

Refer to the reference chart on your welder and set the unit for 18 ga.

For the first exercise, we will be creating a series of 2 inch (5 cm) straight draws called pull welds. These will be surfaces welds, meaning we are not going to be actually welding anything together right now. This exercise is designed to develop hand motion, vision of the weld zone, and understanding of how to pull weld. This will help to build the feel of the speed needed to move through the weld zone to create a proper weld without melting though the steel, a problem known as “burn though.”

Find a piece of scrap sheet steel about 5 x10 inches (12.7 x 25.4 cm) hot or cold roll. The surface dimension doesn’t matter too much as long as there is enough room to practice. The thickness of the piece should be at minimum 16 ga. This is a good thickness for this exercise to show and help give a better understanding of penetration.

Locate and set the voltage and wire speed combination for 18 ga based on the reference chart that’s normally found on the door of your MIG unit. Set the machine for a lower voltage and wire speed than the metal for these welds to allow practicing without blowing holes in the material.

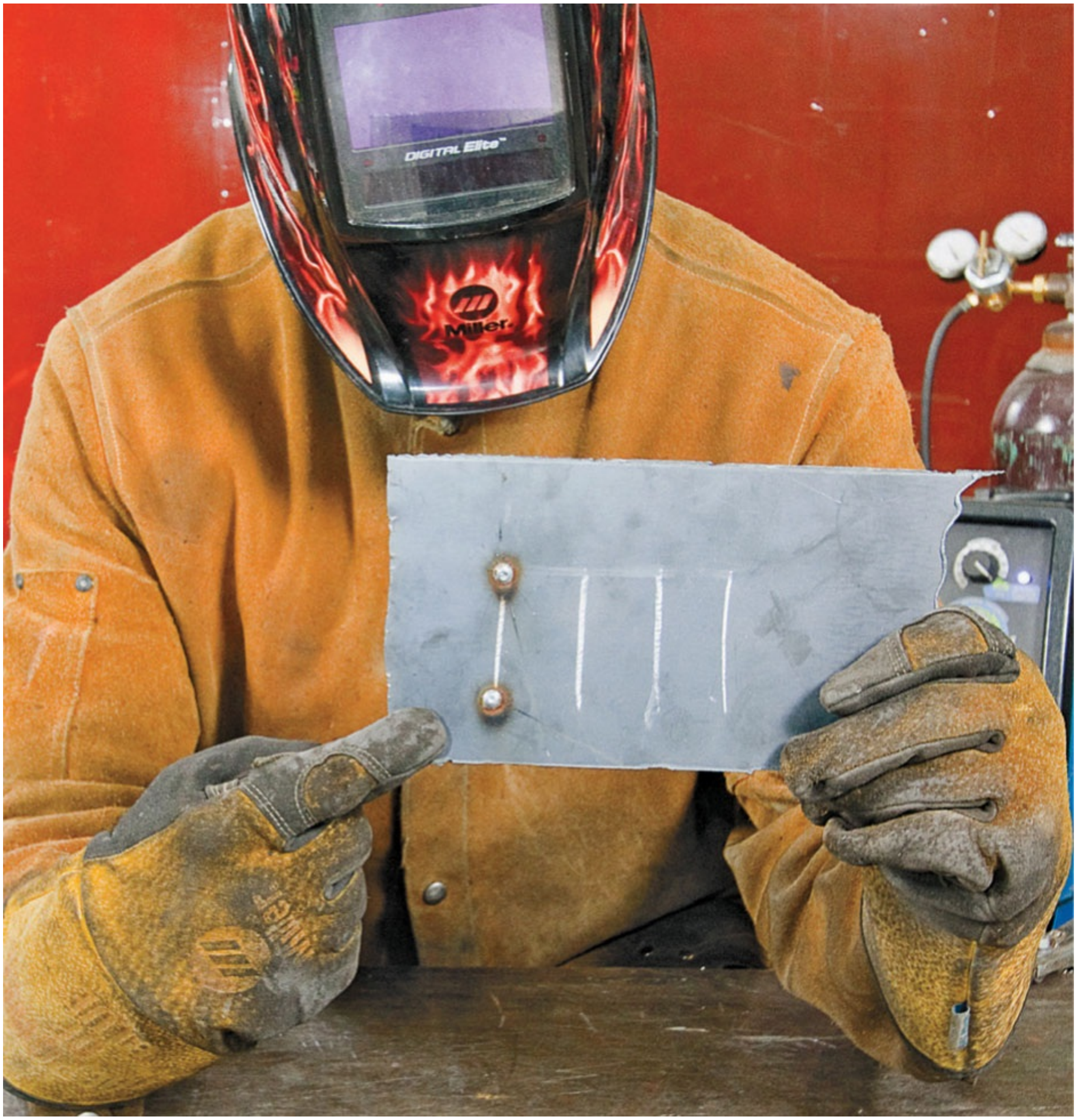

Tack at the top of the line and then the bottom. For this exercise, the tack will be a two count.

Because this is practice, it is not completely necessary to remove mill scale from the surface of the steel. Cleaning the metal with a stainless steel brush should provide enough removal of surface debris. If the piece has paint, rust, or excessive amounts of grime, then it should be prepped with some type of disk such as a 60-grit flap disk on 4 1/2-inch (11.4 cm) angle grinder. (In the future, prep all welds with a grinder.)

With the soapstone, mark six to ten 2 inch (5 cm) straight lines next to each other about 1 inch apart; these lines will represent the weld zone that the bead will be created on. Use some scrap steel bar to elevate the piece of sheet metal to protect the table surface. Make sure that the ground is clamped to the table or directly to the work piece.

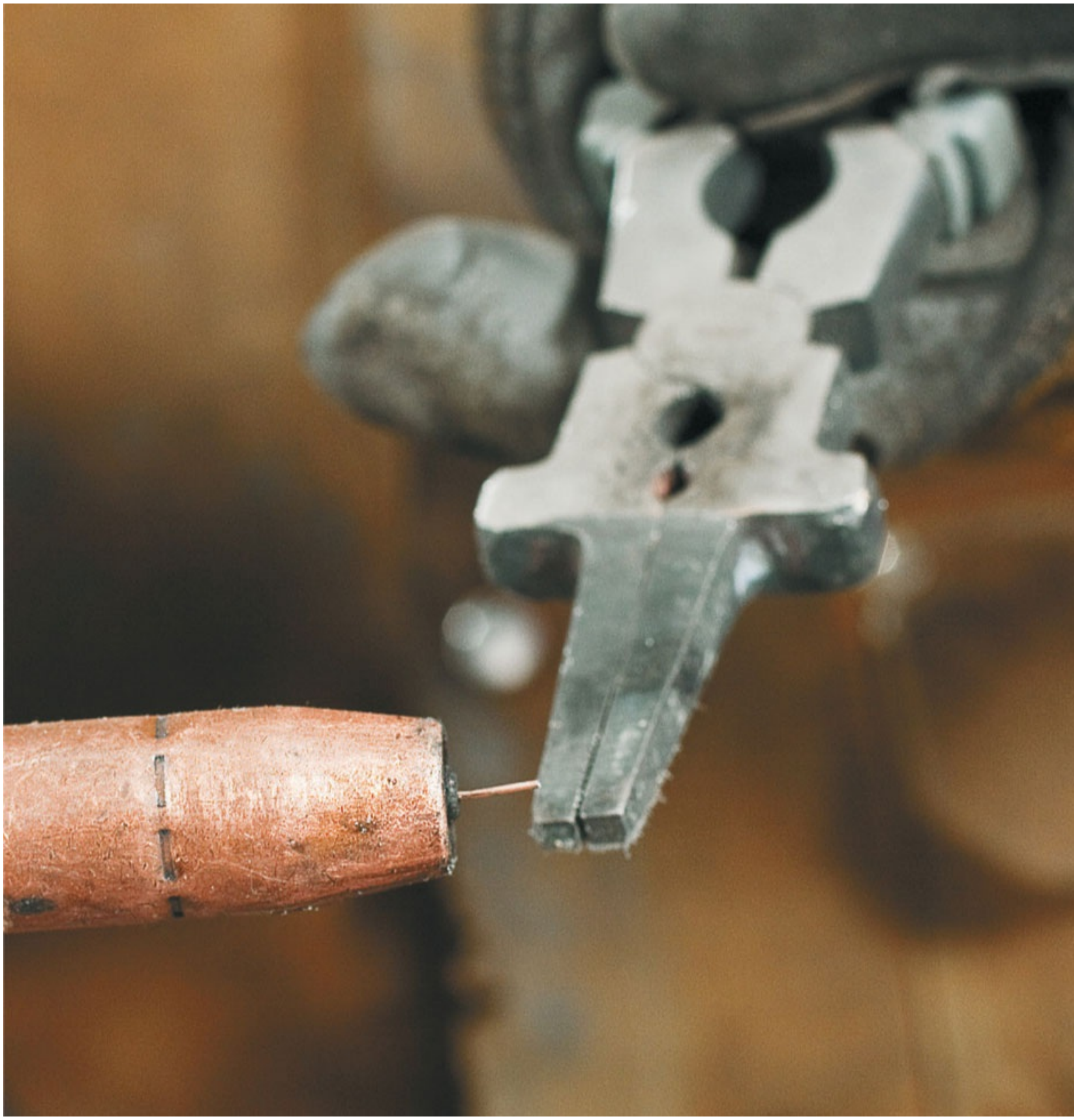

Concentrate on maintaining a 3/8-inch (10 mm) wire length (called the stick out) during tacking and welding.

If you are right handed, start with the far right line and move left. If left handed, start at the far left line and move right. This will allow you to move away from the heat as you practice.

What we are going to do is tack at the top of the line and then at the bottom. For this exercise, a tack is going to be about a two count. Not seconds, just say “one, two” in your head. This should give us a tack that is proportional to the thickness of the material and the size bead that we want to create. Then we are going to connect the two tacks with a bead.

The Unbreakable Rules

- Keep the drag angle pointing into the weld zone.
- Keep the gun perpendicular so the shield gas hits the puddle evenly.
- Brace and support the gun.

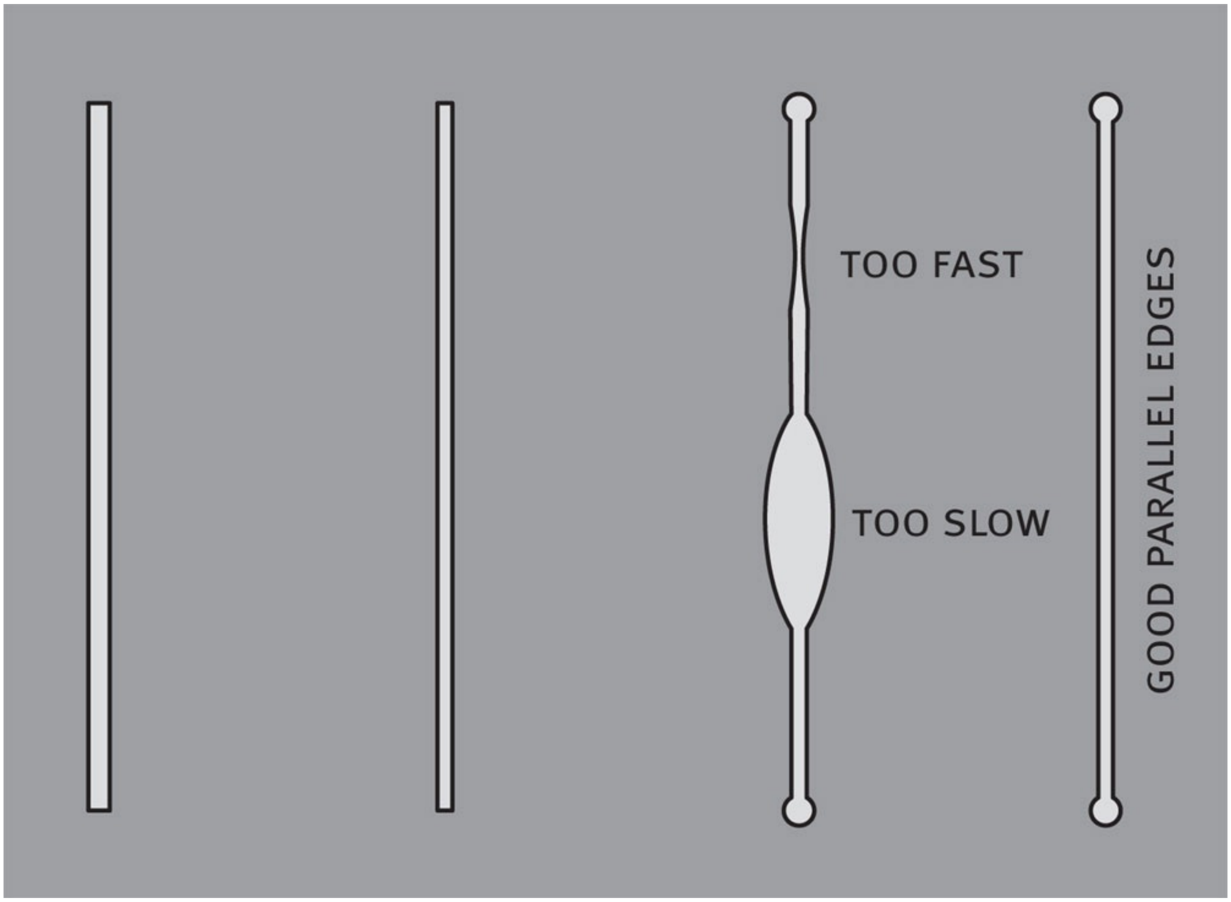

Try to be as even and consistent as possible when moving through the weld. The purpose of this exercise is to learn the correct speed of the draw through the weld zone in order to produce the correct bead. Move too fast and the edges pull in, producing smaller walls. Too slow and the heat radiates and will end up burning through.

Before tacking or welding, make sure that the stick out is at its proper 3/8-inch (10 mm) length. If it is too long, snip it with the MIG pliers; if it is too short, let some line out with the trigger (away from the ground). While conducting the draw, if you start to pull away and the stick out becomes longer than 1/2 inch (13 mm), the arc will not reach its proper heat. The shield gas also can no longer create the proper environment for the arc and keep out inclusions or oxygen, resulting in a poor weld. This is what I like to refer to as going fishing, and it will result in popping, crackling, and skipping sounds. The sound of the arc should be a consistent buzz like bacon frying.

When positioning the gun into place, touch the wire down to the metal before pulling the trigger, but do not keep pressure on the wire. The wire is the electrode and the filler material that feeds into the weld, so if pressure remains on the wire as soon as the trigger is pressed, the contact tip will plunge into the arc and burn back the wire. At higher voltages, this will sometimes cause the wire to weld to the inside of the contact tip. If this happens, you will need to unscrew and cut away the contact tip with your MIG pliers so that a new contact tip can be used. You will not feel the wire feeding into the puddle. You’ll need to learn to maintain that 3/8-inch (10 mm) distance by sight and feel.

When tacking or welding, squeeze and hold the trigger. When finished, release the trigger before pulling away. This will prevent additional wire from being fed.

Once you have the two tacks down, move to pulling a weld. Put the wire on the top tack between the top and the base of the tack right above the line. We are now going to do a pull weld.

What we want to accomplish here is a nice even and straight bead with parallel edges. Go too fast and the edges of the bead will suck in and not provide the necessary heat and filler metal to complete a good weld. Too slow and the heat will spread to the point where the base metal cannot support the heat and filler metal. If you move too slowly for too long, the metal will eventually end up dropping through and create a hole in the metal, what is referred to as burn through.

If you flip the piece over and look on the other side of the weld, you should be able to see where the weld has penetrated. If it looks as though you have welded it from this side, that means you were going too slow and the bead burned though too much. For this exercise, you should only see heat discoloration with an innermost line that mirrors the weld you just created.

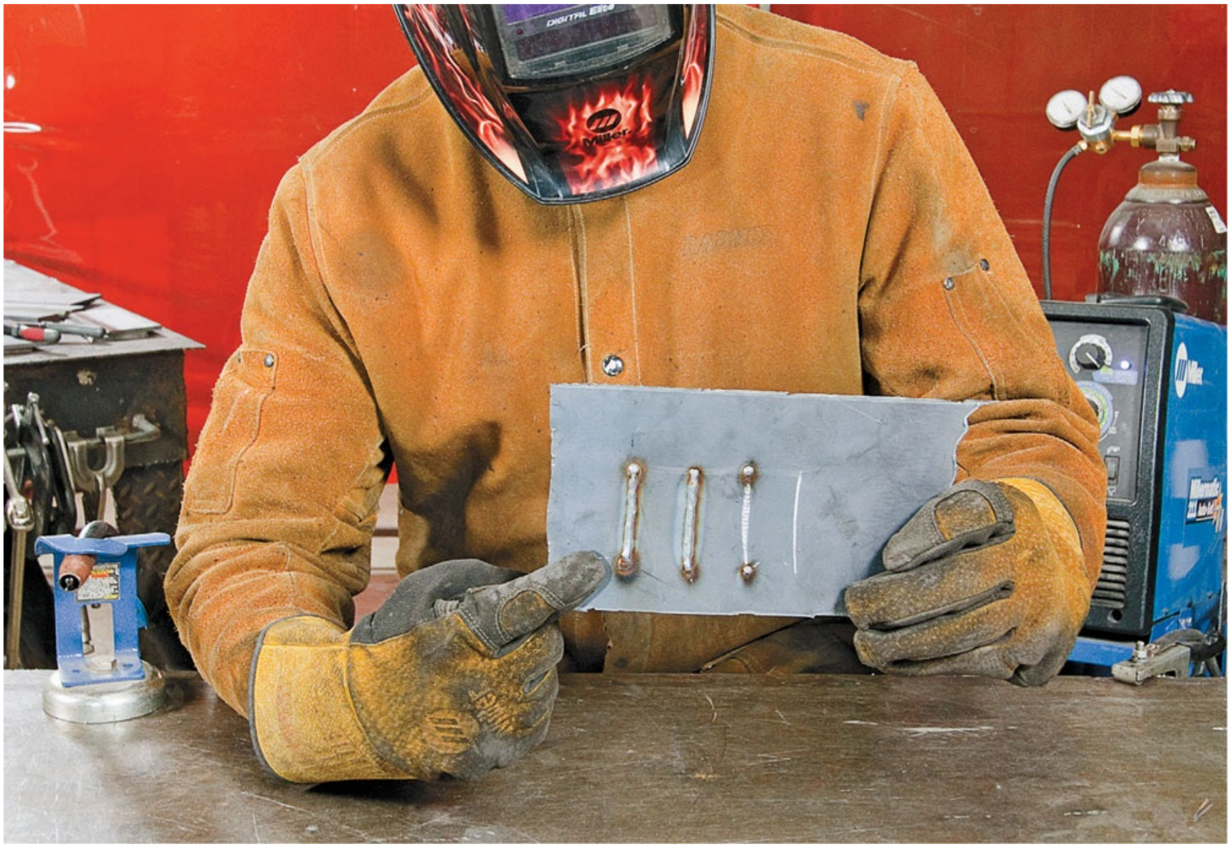

You want to end up with a bead with parallel edges. If you look at the other side, you can see if you have burned though too much. This will be an indicator that you will need to speed up your pace.

Repeating this exercise several times will help develop a consistent straight draw. You should stop and critique the welds after each draw. Focus on the rules and principles each time while tacking and welding. When you create a few straight and even beads with parallel edges with little to no burn though, you have learned the straight pull weld. By practicing this on different thicknesses of steel for different sized beads, you will find that the speed of the movement will have to be adjusted slightly.

Welding traditionally involves setting the machine to the proper combination of wire speed and voltage, as well as drawing through the weld zone to create a bead. This method can be used on thick or thin metals. As you develop your skills as a welder, you'll begin to understand when and how to adjust the settings in accordance with your technique.

Listen to the Weld

Vision is not the only sense used when welding. Remember to listen to the arc during the weld. When you weld, you want to hear a constant and steady buzzing that sounds like frying bacon. If you hear a lot of popping, cracking, and skips, chances are you are pulling too far away from the work piece.

II

Using Your MIG

5

The Basics

Whether you’re welding a lamp, a car, or a skyscraper, all projects require you know some basic joints and techniques. The more advanced the project, the more advanced the techniques you’ll use, but we’ll stick with the basics for now.

Basic Joints

Depending on who you talk to, what textbook you read or website you visit, there can be any number of different joints, from one to over four dozen. In one of the fabrication shops I worked, there was a chart of 48 different weld joints with all sorts of names. These all have their purpose for very specific tasks. Don’t feel as though you have to go out and memorize all the names for all these welding joints though. Most of these joints refer to how the weld is chamfered and prepped.

There are really only five basic welding joints that you should concentrate on right now: the butt joint; the corner (or fillet) joint; the tee joint; the lap joint; and the edge joint. These five joints are actually derived from only two joints when it comes down to how they are welded together: the corner (or fillet) joint and the butt joint.

The corner joint, the tee joint, and the lap joint are all welded with relatively the same technique and are all examples of the fillet weld. They are made by going into a corner at roughly a 45-degree angle to create the weld in a 90-degree corner.

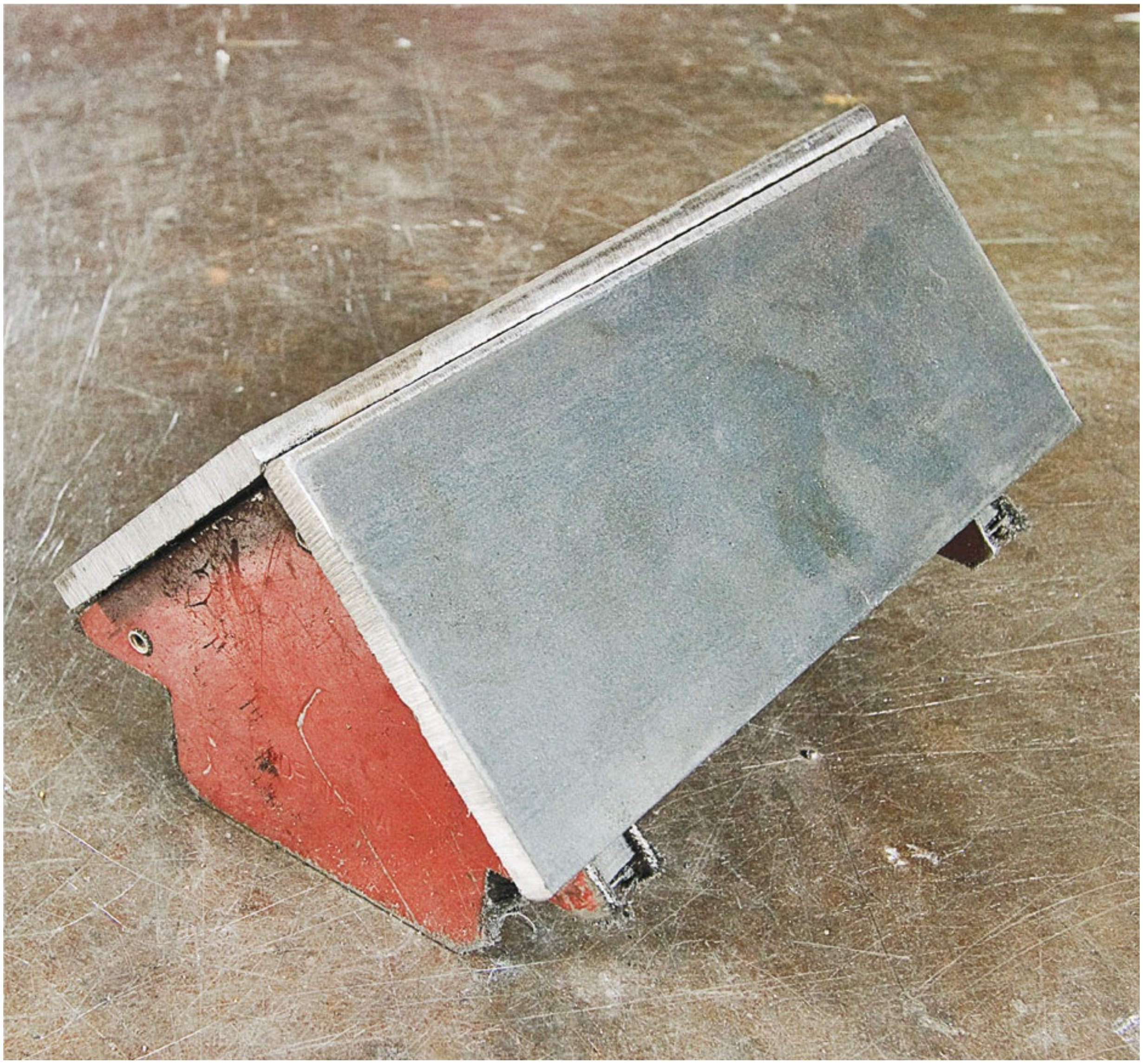

Corner Joint

The corner joint is used in many different applications. It’s the most common joint used when making furniture, working on a car, or building a flight of stairs. There are several schools of thought when it comes to corner joints and its variations. A corner joint is defined as two pieces of material that are welded to create a corner. Outer, inner, closed, and open are all examples of different names of corner joints.

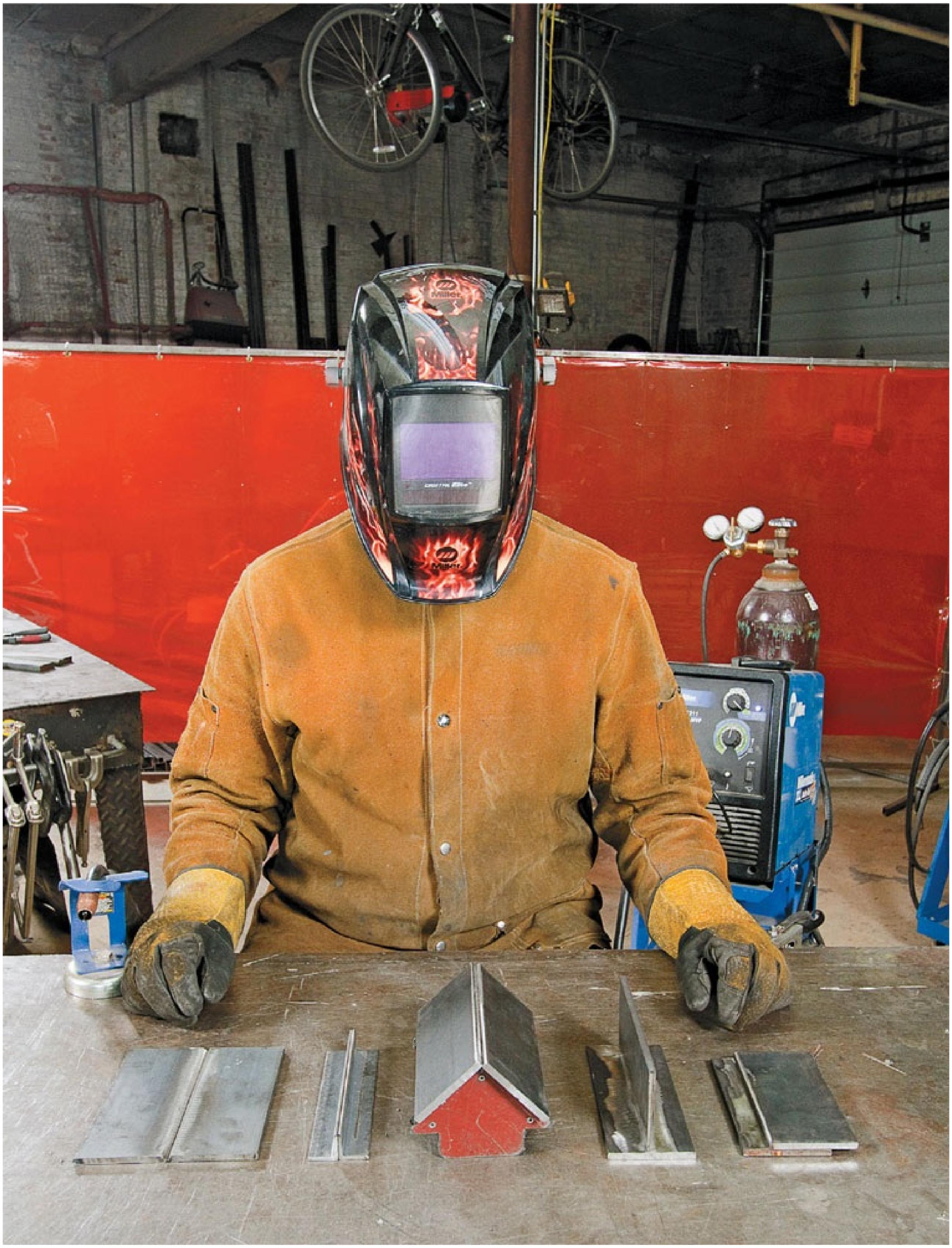

The only joint that is different from these three is the butt joint. The butt joint is just that: two pieces of base metal butting up to one another to create an even surface. Practicing the method of how to weld a butt joint is the best place to start learning how to weld.

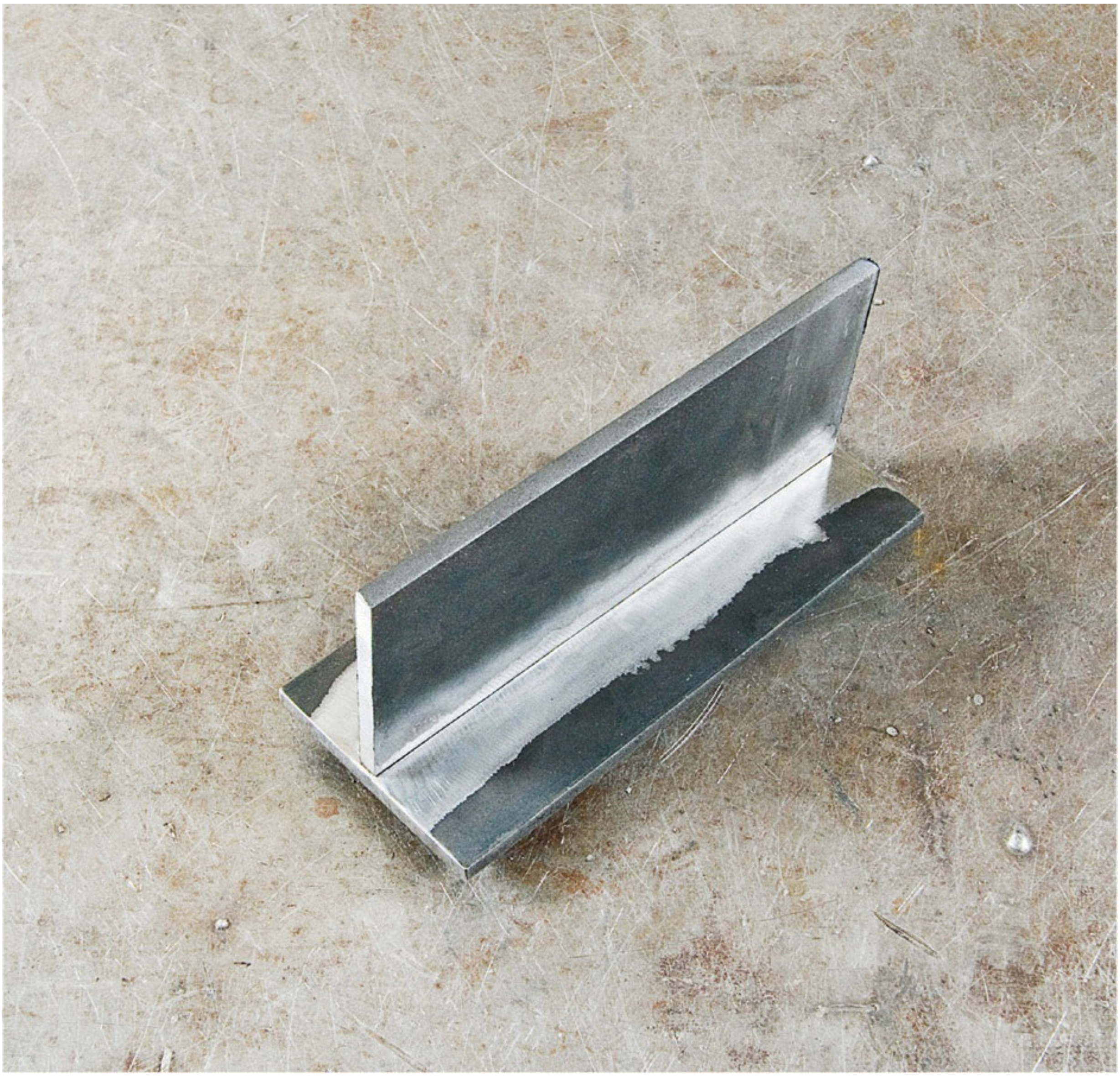

Tee Joint

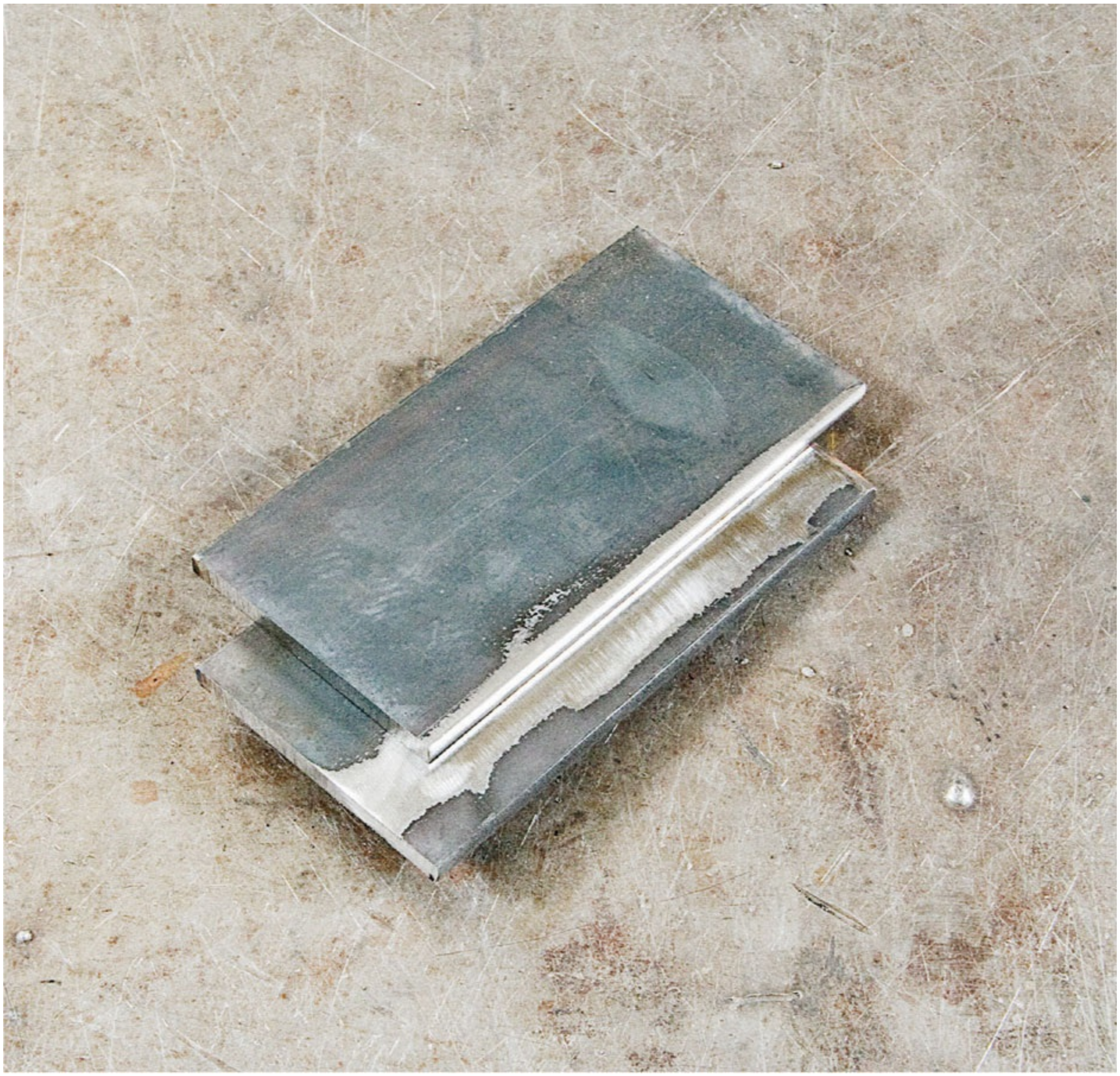

Lap Joint

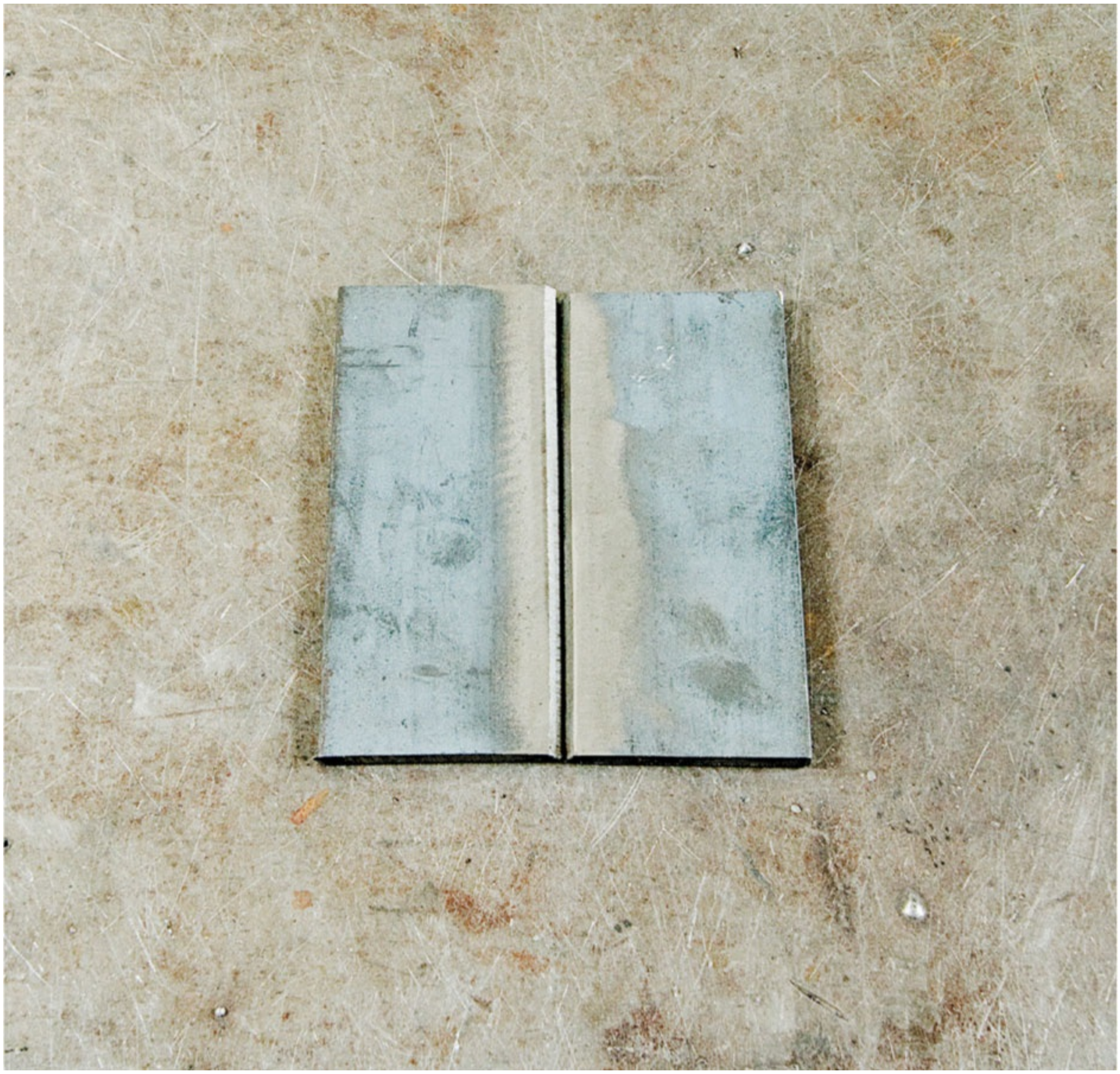
Butt Joint

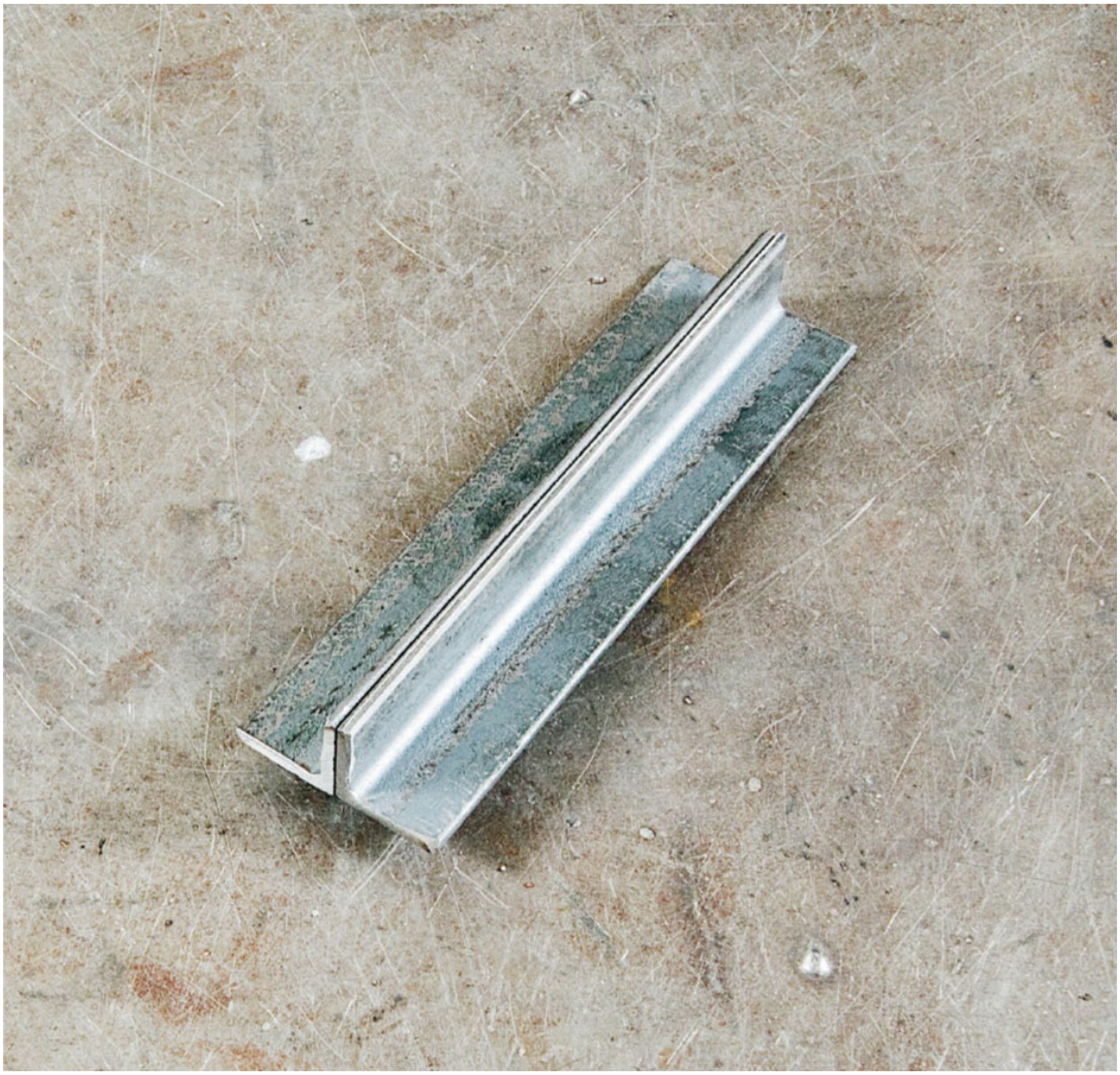

Edge Joint

As we move through the book, we will learn more about these joints and the proper way to execute a weld for each of them. For now, the focus will be on butt joints and developing the movement through a weld.

Technique

MIG welding was developed in the 1940's as an alternative process for welding aluminum. It was quickly adapted to the steel industry and is now the most widely used electrode arc process today.

In the early years of arc welding, the traditional way to conduct a weld was a straight pull or push method of controlling a weld in progress. The machine's voltage and wire speed combination in relation to the speed of the arc moving through the weld zone were the only controls for penetration and performing a weld.

Now that you have a straight pull weld down, it is time to teach you the next level of control. Remember what I said: "The number one job as a welder is to control fluid."

Gas welding—such as oxy-fuel and gas tungsten arc welding (or TIG)—are manually fed electrode arc processes where the operator has to heat the base metal with a torch in one hand and with the opposite hand has to feed the filler metal into the puddle. This style of welding allows for more control but limits the speed in which a weld can be created. Traditionally, having more control over the puddle meant having more control over the outcome and look of the weld after it was completed.

In modern welding, terms such as "dimes on the side" or "stacking nickels" have become synonymous with TIG welding. These terms refer to the welder's ability to control the outcome of a weld to make it look like a row of coins lying on their sides. This became the signature of a welder who has honed their craft. This style has become the look for welds on everything from bicycles to custom car frames to furniture. It is now the benchmark for recognizing a quality weld and welder.

The looks of these welds have become so popular that robot welders have been programmed to duplicate the weld to produce the same outcome. Even though at the atomic level the weld may not be any better of a weld than a straight weld, it is more difficult to execute. Some techniques of welding do increase penetration, but these techniques are not something that a beginner welder needs to worry about.

This outcome for a TIG weld can be mimicked with MIG. The term "MIG like TIG" was developed when MIG welders began to experiment with the control of the puddle. This control not only produced the TIG effect in the final weld, but also allowed more options for conducting welds in progress. This sparked new styles and techniques and improved methods for many aspects of sparked new styles and techniques and improved methods for many aspects of welding, such as a greater ability to control the heat going into the base material.

There are several of these techniques that will produce a wide range of patterns in a weld. The one that you are about to learn is a basic hand (or I should say arc) motion technique that offers the next level of controlling a weld. This is not the technique that will produce the MIG like TIG effect, but it will give you a better understanding of how to control the puddle and work your way up to that effect. This technique is the root of most other techniques.

Exercise 1: Butt Joints

DIFFICULTY: BEGINNER

MATERIALS AND EQUIPMENT:

- MIG welder
- Two 6 to 10" (15.2 x 25.4 cm)-long sections of 1/4" (6 mm)-thick plate
- Welding Helmet
- Welding gloves
- Long-sleeve shirt
- Leather welding jacket or apron
- MIG pliers
- Stainless steel wire brush
- Soapstone
- 4 1/2" (11.4 cm) angle grinder outfitted with an aluminum oxide grinding disk (stone)

The technique that we will start with is a pattern that moves the arc back and forth between the edges of a “V” chamfered butt joint. This will be the first exercise where you’ll actually be welding two pieces of steel together. Since techniques like this keep the heat in one place longer, they will normally only apply to thicker materials.

For this exercise, you will need to grind a 45-degree chamfer in two 6-to 7-inch (15.2 to 17.8 cm)-long pieces of 1/4-inch (6 mm)-thick plate steel. Remember the order of every project in metal fabrication: cut-prep-tack-weld.

Remove the mill scale away from the weld zone after chamfering the edges. This will ensure that if the weld rolls over the chamfered edge, it will not pull any inclusions back into the weld that may cause porosity.

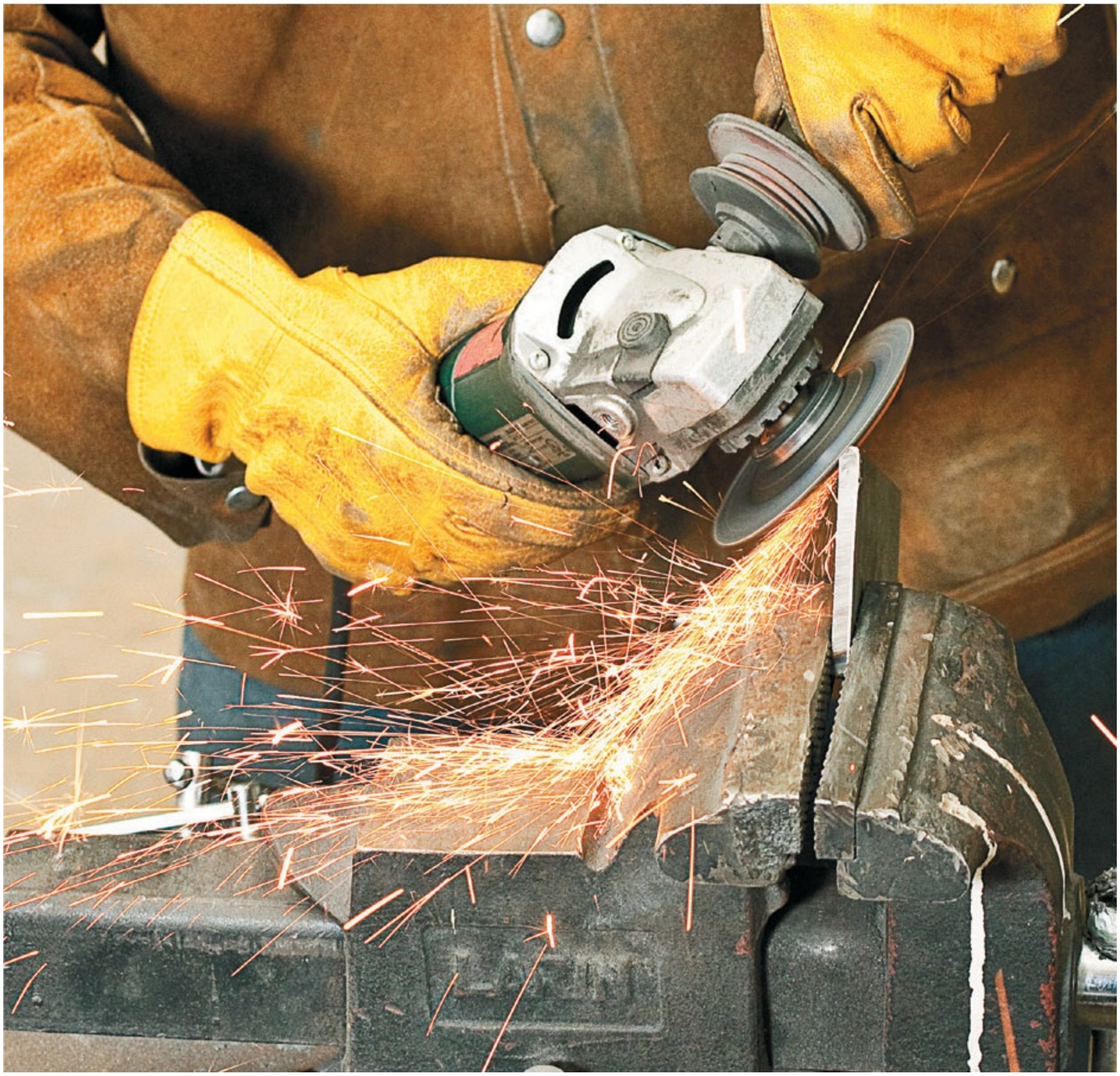

When chamfering, lock in the angles to get a nice flat surface.

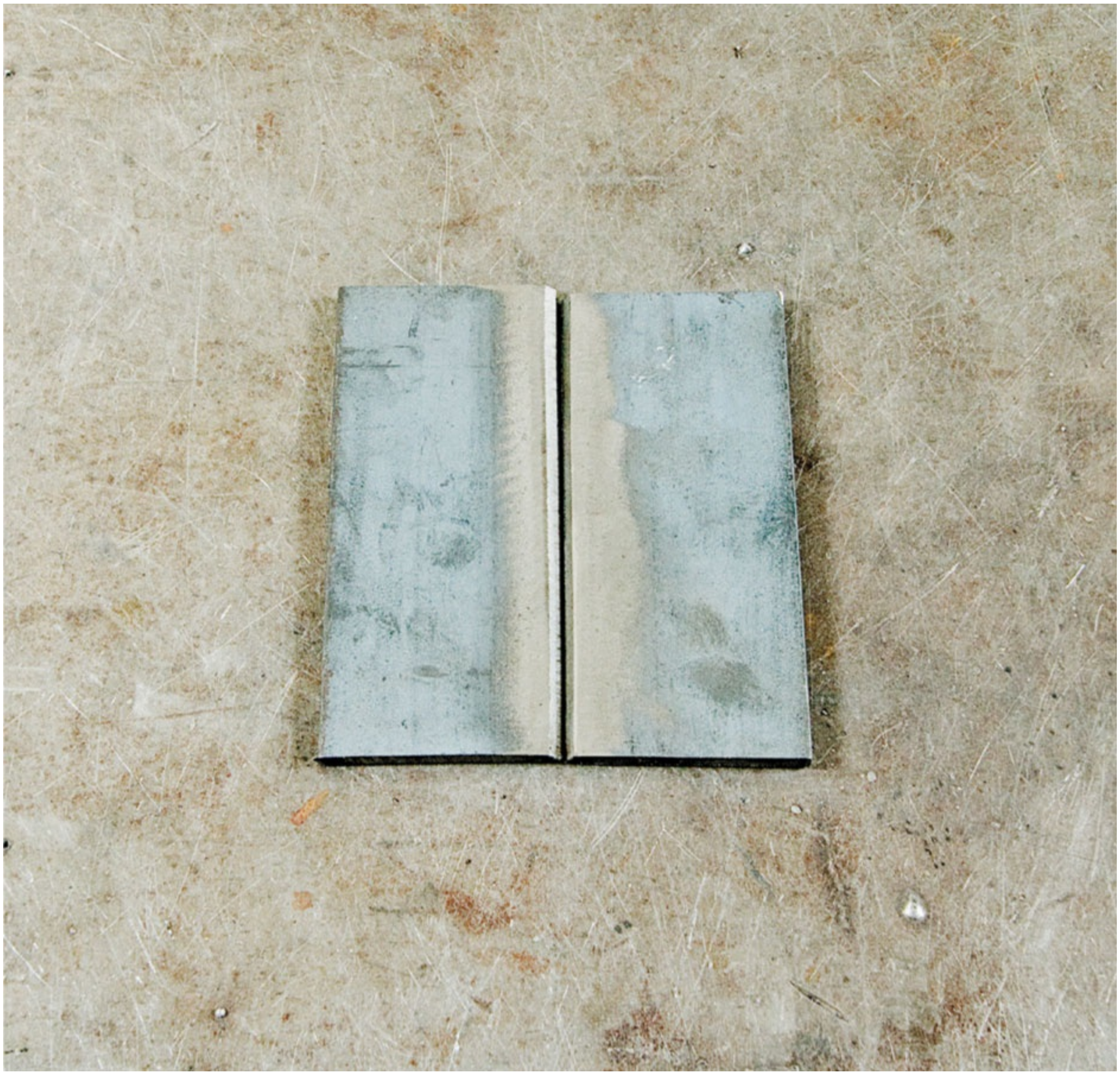
When you are done chamfering, prep the weld zone by grinding and cleaning the steel close to the weld.

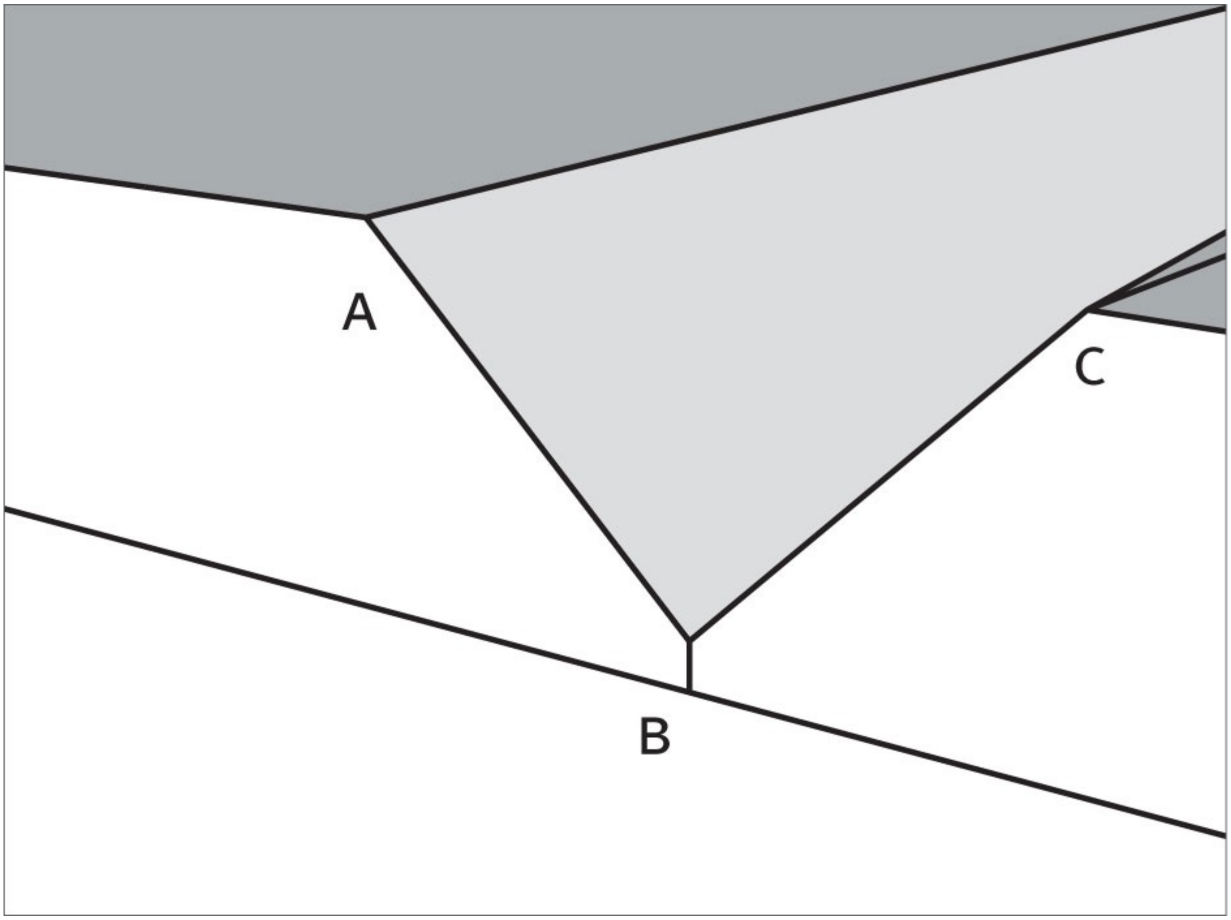

A and C are the outer lines that will complete the plain. The middle line, B, is where the metal butts together.

One of the advantages of learning a technique on v-chamfered butt joints on material this thick is that there are three distinct lines that define the weld zone: the two edges of the reservoir, “A-edge” and “C-edge,” but also the actual line of the butt joint, “B.” These lines will help in the same way three-lined paper helped people to learn cursive writing. In essence, this is the same thing, only writing with the molten puddle of steel.

Tack at the top first, then the bottom, and finally the middle.

Don't worry about a two count when you're tacking this time; instead, just worry about filling the chamfer up to the A and C edges. You may need to clamp down on the piece to prevent movement (not shown).

Tack the pieces together with three tacks—one at the top, one at the bottom, and one in the middle—to create two weld zones. Tacking for this exercise is not going to be a two count; instead, watch the puddle fill the reservoir to its edges (A and C) and then stop. For tacking, always set the machine for the thickness of the material being tacked, in this case 1/4 inch (6 mm). For this exercise, it will be necessary to increase the CFH (CMH) for the shield gas. The last exercise was a thinner piece of metal that required less shield gas. Now that it's a 1/4-inch (6 mm)-long weld, it should be set higher; 30 CFH (0.85 CMH) or the recommended flow from the machine's manufacturer should do the job.

Clean the metal after each weld and series of tacks.

This is a practice piece so it's not necessary to clamp these down to prevent

This is a practice piece so it's not necessary to clamp these down to prevent the steel from moving when heated. It is a good idea to elevate the base metals with two pieces of the same size bar to avoid putting heat into the table surface.

Welding Checklist

Before you do any welding, you should always go through a checklist. Start from the back of the tank and work your way to the table checking everything.

- Is the shielding gas tank on and set to the correct CFH (CMH) for the gauge metal?
- Are the machine's voltage and wire speed set properly?
- Is the piece grounded?

From now on, after completing a series of tacks and before each weld, clean the weld zone with a stainless steel wire brush. Notice the carbon burn around the tacks that were created? This is all of the ambient particles and debris floating around in the air that the shield gas pushed out of the weld zone but stayed close enough to the puddle to burn. Welding over this would reintroduce these potential inclusions back into the weld zone and create a poor weld. Get into the habit of scrubbing your metal now.

Now that we have everything tacked together, the next step is to weld. For these welds, we will be introducing our motion technique. When setting up for this weld, it will be necessary to turn down the voltage and wire speed on the MIG. The adjustments for the wire speed and voltage are going to be determined by the technique used and the machine.

The shield gas output should be set for the thickness of the material just like the rest of the machine. For 14 ga and thicker, the flow should be 30 CFH (0.85 CMH). This will provide a large enough environment for the larger puddle to be protected from oxygen inclusions.

The back and forth motion you'll be using will allow the heat to stay in an area for a longer period of time so less heat and filler from the machine are needed. As you develop as a welder, you will find that you can adjust the machine for a number of different approaches. If you find that you like to move a little faster with motion and travel, more heat and filler will be required. Slower movements will require less. Understand that there is a range: turning down the machine too far would result in a small puddle, less penetration, and a poor weld. Setting the machine too high will require a faster motion and more heat. Finding the appropriate balance of heat and filler for the technique used may take some time and practice.

Another aspect of heat to consider is how hot the material becomes as you weld. When welding on plate metal, you have the ability to go greater distances but the heat is also radiating through the metal. The weld zone that is now to be welded is at a greater temperature, which makes the metal liquefy faster and causes the weld to become too hot. In this case, you would need to decrease the heat. This is getting a little ahead of the game though; for right now, it’s best just to understand that you can make adjustments to the heat and flow of the material depending on the technique you use.

For this exercise, set the MIG for 1/8-inch (3 mm)-thick steel. This will provide enough heat and filler metal to move slowly through the weld zone as we execute our technique. This will also depend on the machine itself; some machines have a tendency to run a little hot and some a little cold. Remember the sizzling bacon sound.

When setting up for this, put the wire on the “B” line at the very bottom of the top tack. All three metals pieces, the two pieces of base metal and the tack, should be used to develop the puddle to begin the weld. This may take a second; build the puddle until it reaches the A and C edge.

Once the puddle has reached the edges (A and C), start to move the arc to edge C, then back to edge A, in quarter circle motions while drawing the puddle back. Basically, you’re lining the bottom quarter of the puddle, back and forth, while pulling the puddle with you. Each time, you bring the arc to just break the A and C edges with the puddle before going the other direction.

This motion is not that drastic; it is actually a small motion that only requires the movement of the fingers or a small movement of the wrists. It helps to get into a rhythm executing this “shark tooth” technique. For every movement to one side, the draw is 1/16 to 1/8 inch (2 to 3 mm). The manner in which the weld is conducted will determine the speed of the draw and the movement. Speed and movement will always be in relation to one another. Increasing the speed of the draw will require increasing the speed of the arc movement.

All of the settings and motions will be determined by the size of the puddle. The puddle should remain the same size through the entire weld, just breaking the edge as you execute the technique.

When the first weld zone is completed, take a look and evaluate the weld. Did you break the edges? Are you favoring one side? Were you too fast or too slow? The end result should be a weld with even edges and an even crown.

Practice this technique until the bead is consistent, even, the edges are parallel, and the pattern of the bead is uniform. Welding is just like learning a musical instrument; practice makes perfect.

musical instrument: practice makes perfect.

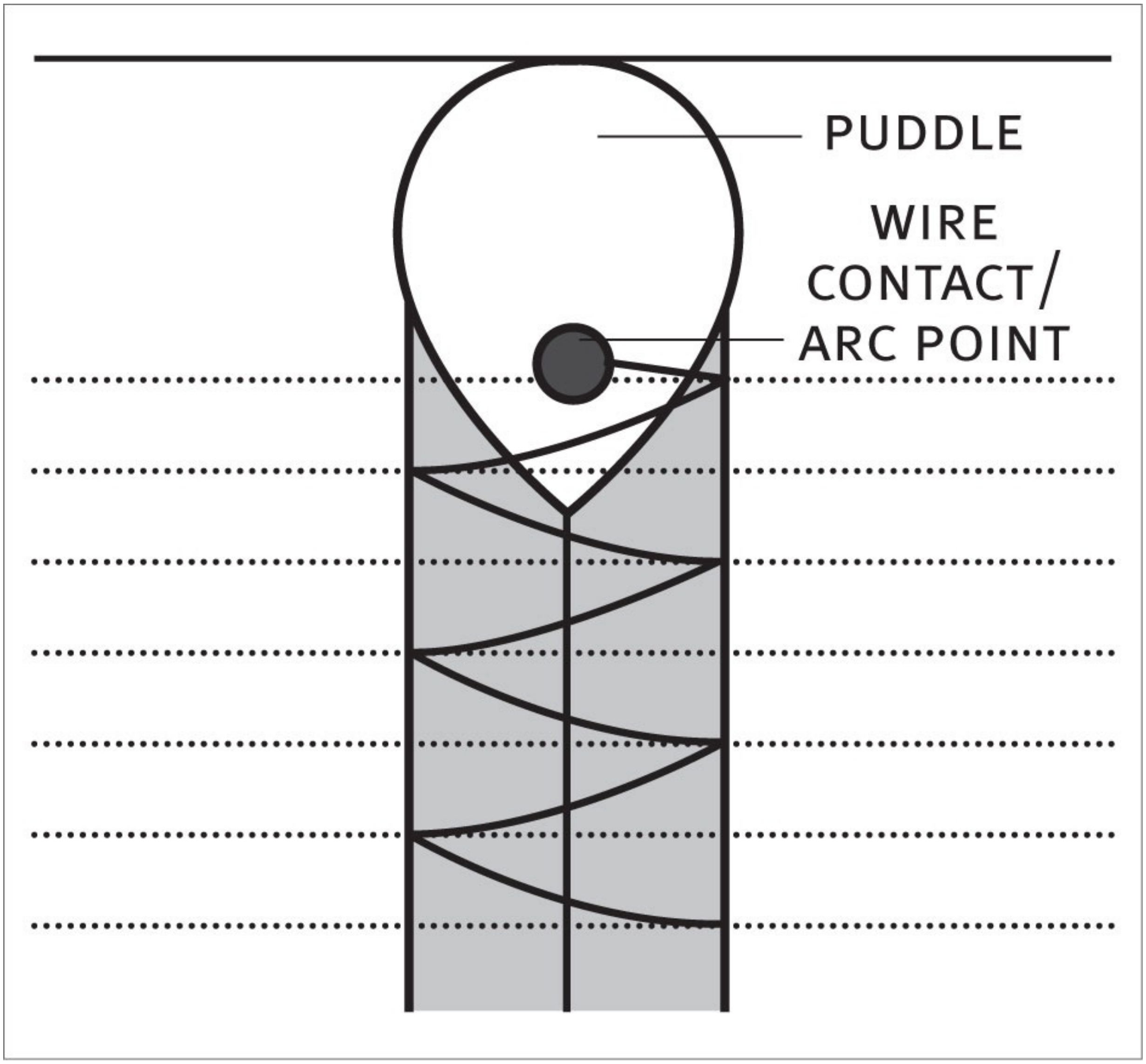
PUDDLE
WIRE CONTACT/ ARC POINT

Just break the edge to complete a motion and then go the other way.

Don't Leave the Puddle Behind

Just remember the most important thing: don't leave the puddle behind. The movement of the puddle is what creates the bead, building it with the motions of the arc. On thicker metal, the puddle does 75 percent of the melting of the base material. The arc's primary job at that point is to keep the puddle fluid. If the correct size of the puddle is not maintained, the heat and penetration are less and the bead will decrease in size, temperature, and its effectiveness to create the correct weld.

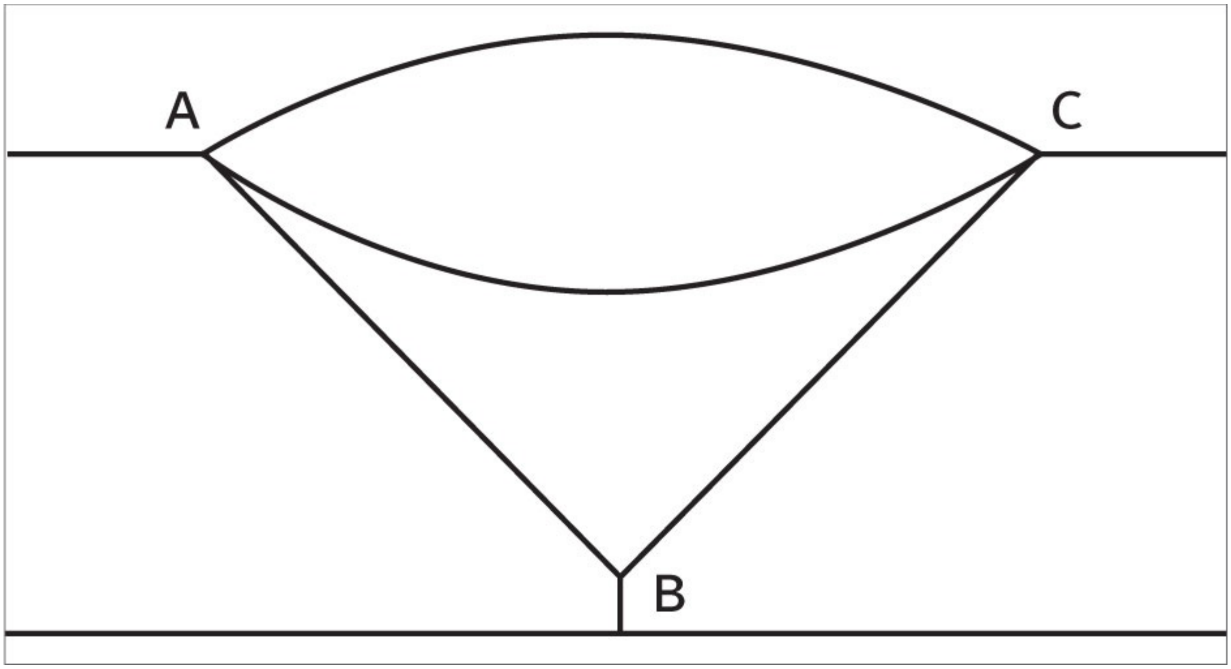

From this split view of the weld in progress, you can see there is a third dimension of movement: a slight half dip into the reservoir. Dipping too far can shorten the stick out and cause burn back.

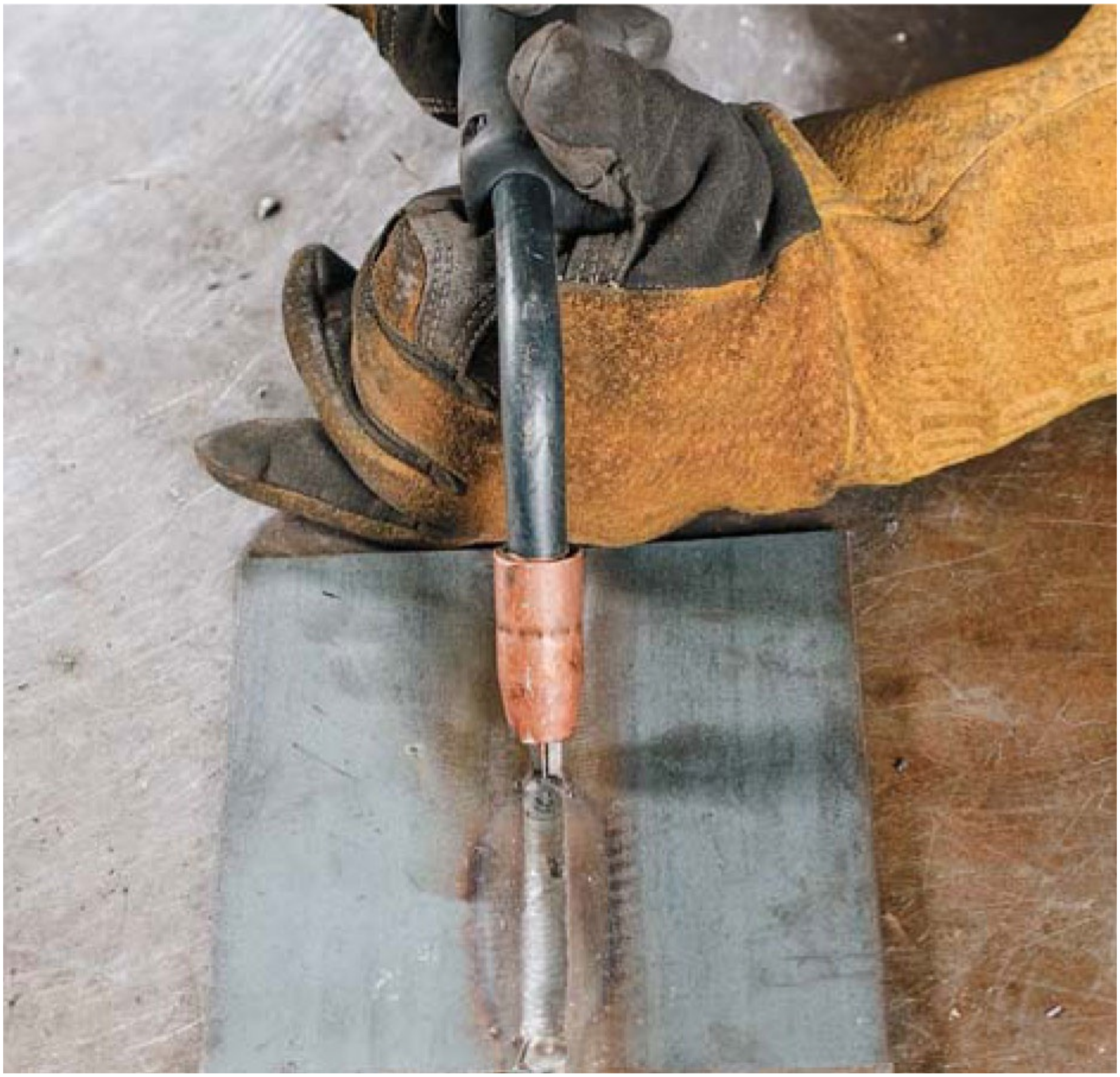

Lining the bottom section of the puddle does not allow the arc to go all the way down to the bottom of the reservoir. Adding in a little dip into the lower section of the puddle helps to maintain heat and penetration.

Remembering the Tip Dip

If the contact tip and nozzle look dry or powdery, it may be time to dip it in the spatter protection while the nozzle is warm. Don't dip the nozzle when it is cold; it could clog the nozzle and block the shielding gas.

Exercise 2: Corner (or Fillet) Welds

DIFFICULTY: BEGINNER

MATERIALS AND EQUIPMENT:

- MIG welder
- Two 6 to 10" (15.2–25.4 cm)-long sections of 1/4" (6 mm)-thick plate
- Welding Helmet
- Welding gloves
- Long-sleeve shirt
- Leather welding jacket or apron
- MIG pliers
- Stainless steel wire brush
- Soapstone
- 4 1/2" (11.4 cm) angle grinder outfitted with an aluminum oxide grinding disk or a 60-to 80-grit flap disk
- 2 right angle magnets

By now you should have a pretty good understanding of what makes a good weld. There are other factors that make a good weld but for now our focus is create them. The butt joint motion technique offers a level of control that puts the welder in a better position to create a weld and not have to rely so much on the machine. It forces the welder to concentrate on the elements creating the bead. The outcome is a cleaner and more aesthetically pleasing butt joint.

That was only one of the five basic joints. Like I said in the beginning of this chapter, there are only two joint techniques to start with that comprise the rest of the joints and will point you in a direction to other techniques and beads.

The advantages of learning the butt joint first are getting the timing and pace down, then developing a rhythm and pace for the techniques to create a bead. The next step is learning how to fight gravity. Yes, gravity.

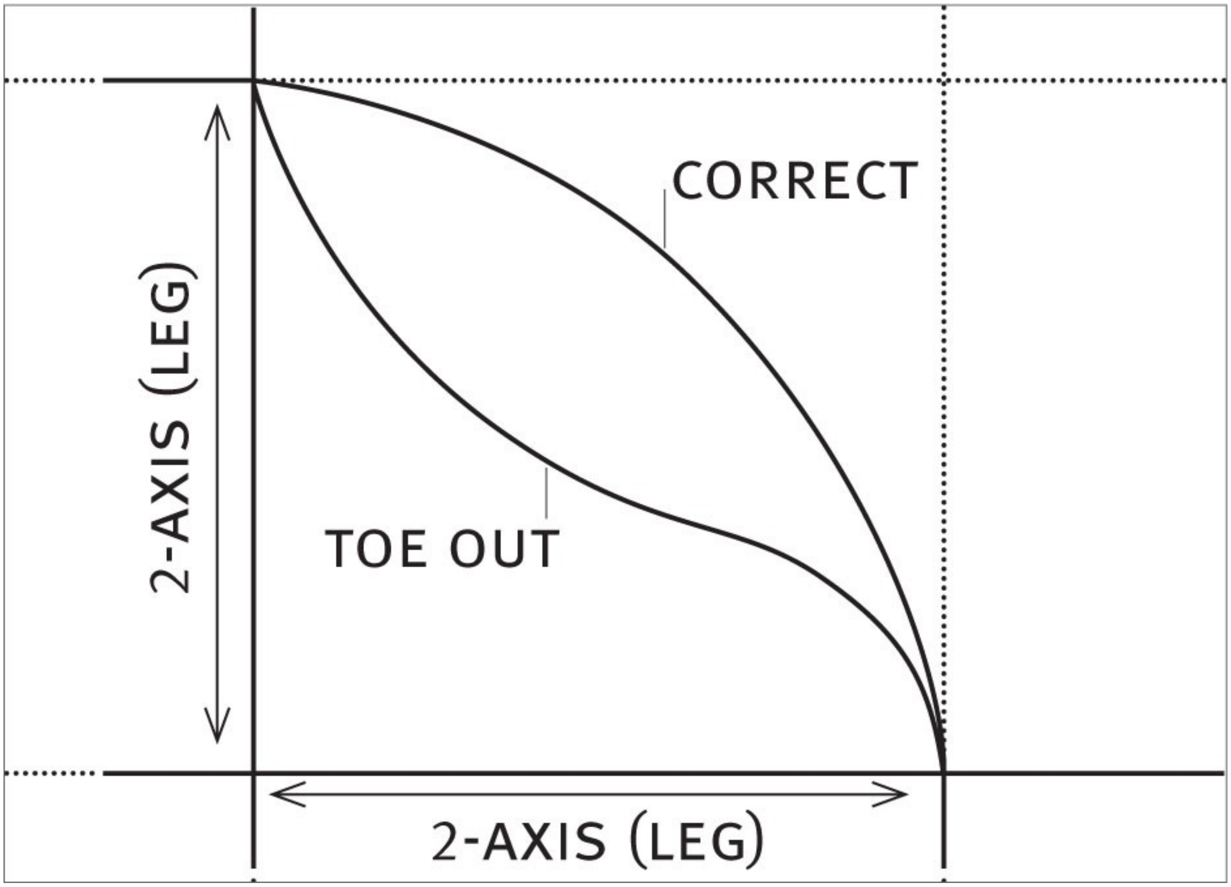
CORRECT
TOE OUT
2-AXIS (LEG)
2-AXIS (LEG)

This is the shape of crown for a fillet weld. The “A” and “C” edges of a butt joint are very useful in assisting with vision as the weld is created. When you produce a fillet weld in a corner, there are no lines to follow except for the very corner where the two base metals meet.

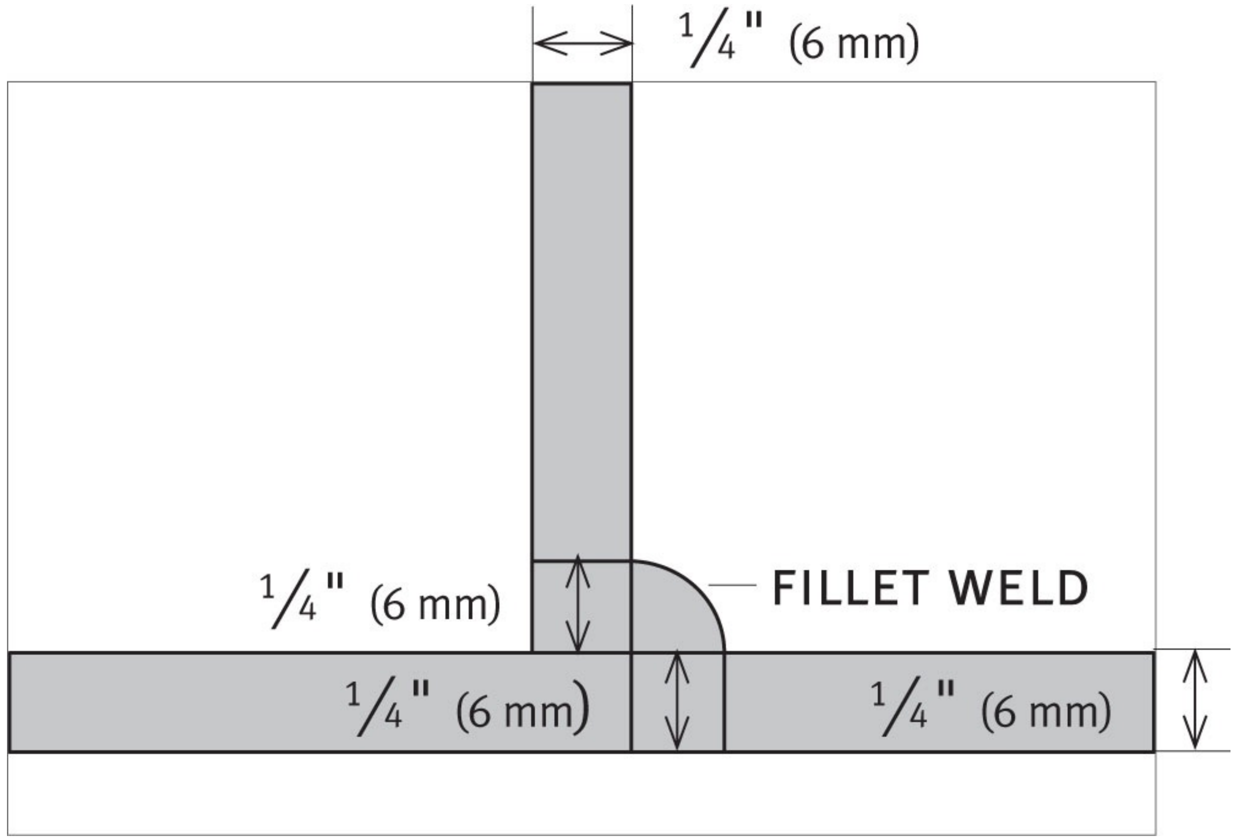

This is a cross section of a corner fillet weld.

Unlike the butt joint, there are two distinct issues that make the fillet weld harder than the butt joint. One is gravity. You now have a vertical axis to contend with. Fluid wants to naturally flow with gravity, so if the proper speed and angle of the pull weld through the weld zone is off, the puddle will want to sag and toe out. When a puddle toes out, the fluid metal flows with gravity and sags, creating a toe-like shape on the base metal.

Use a grinder with a flap disk to prep the tee joint weld zone.

Executing a traditional pull weld into a corner creates a few challenges. For the correct size weld, the bead “legs” should be as long as the material is thick; if the material in this diagram is 1/4-inch (6 mm) thick, then each of the legs on the Y and Z axes should be 1/4 inch (6 mm). To avoid the puddle from toeing out, the speed of the weld needs to increase.

If the speed is increased, then the bead gets smaller. A smaller puddle means less penetration. To counter this, the wire speed and voltage combination have to be adjusted. This will increase the heat and filler metal for better penetration and allow for the puddle to move through the weld zone faster and produce the necessary weld.

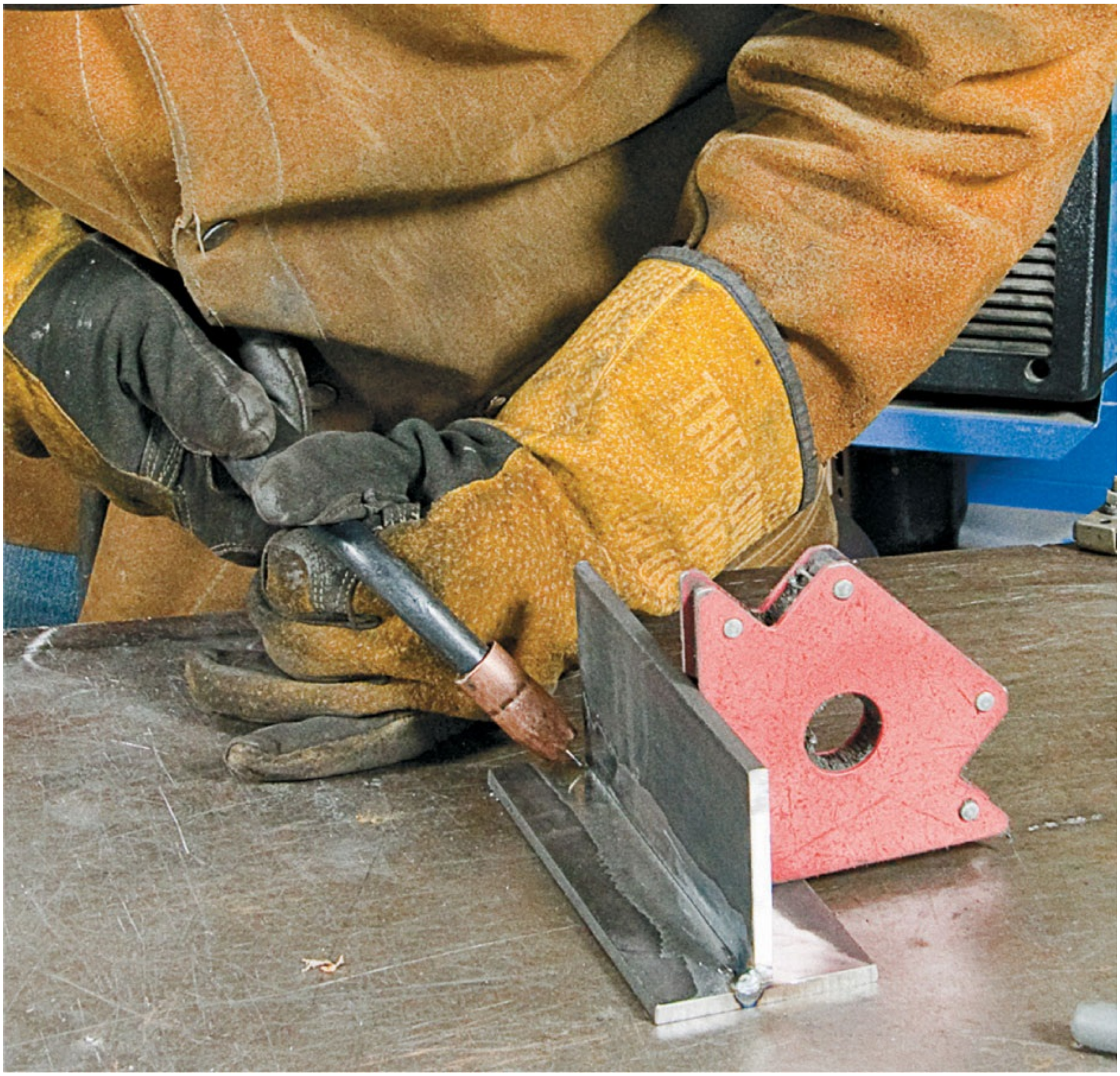

When tacking and welding, you always want to hit the puddle evenly. That means going into a 90 degree corner with the nozzle at a 45 degree angle.

The other way to produce a proper fillet weld is another motion technique that will allow for better control of the puddle. Just like the butt joint motion technique, the fillet has its own patterns and outcomes. Also just like the butt joint, the primary rolls don’t change. Be comfortable, brace the gun neck with your opposite hand, point the nozzle straight into the weld zone, and have an even angle on both sides of the stick—in this case a 45-degree angle. The stance, posture, and support should all transfer to the set up and execution of a fillet weld.

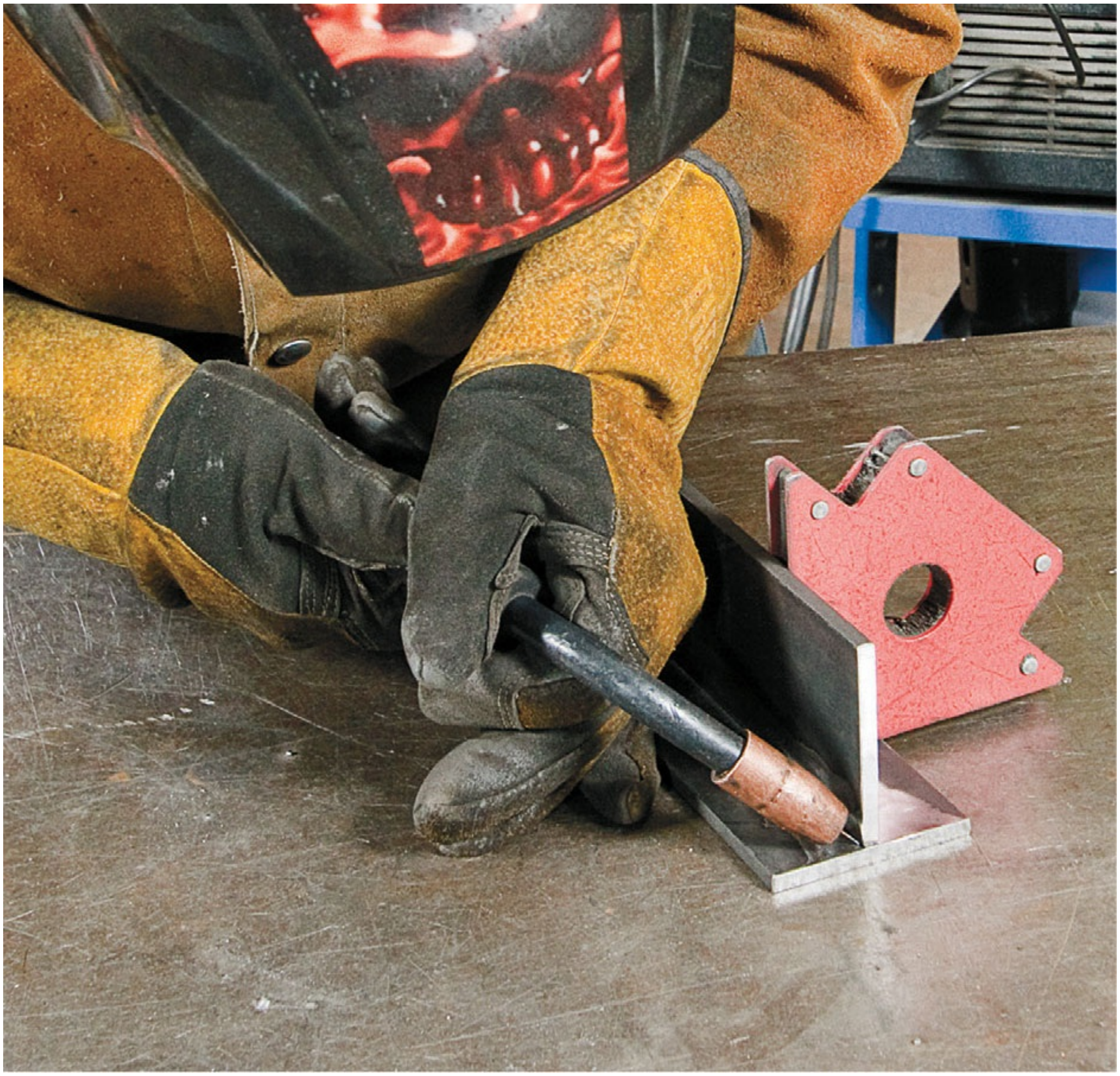
Set up first and get comfortable with the angle you will be welding at.

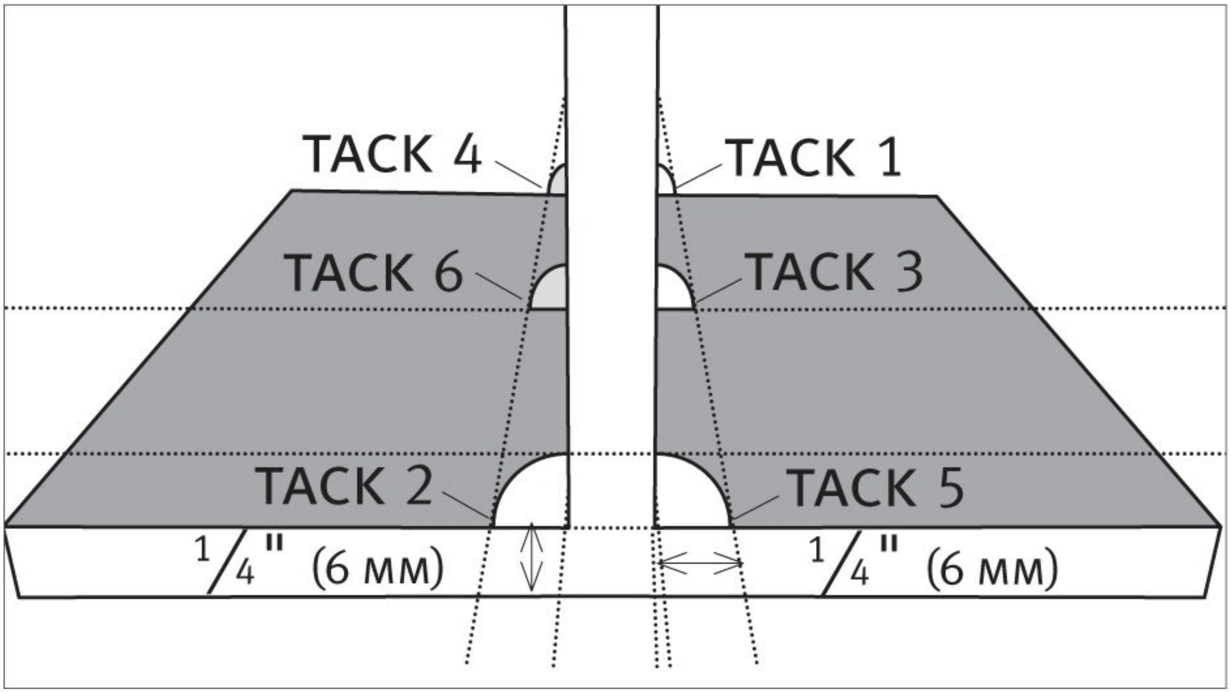
TACK 4
TACK 1
TACK 6
TACK 3
TACK 2
TACK 5
1/4" (6 MM)
1/4" (6 MM)

For tee joints, move from one corner to the next cross/opposite corner, as shown in this diagram.

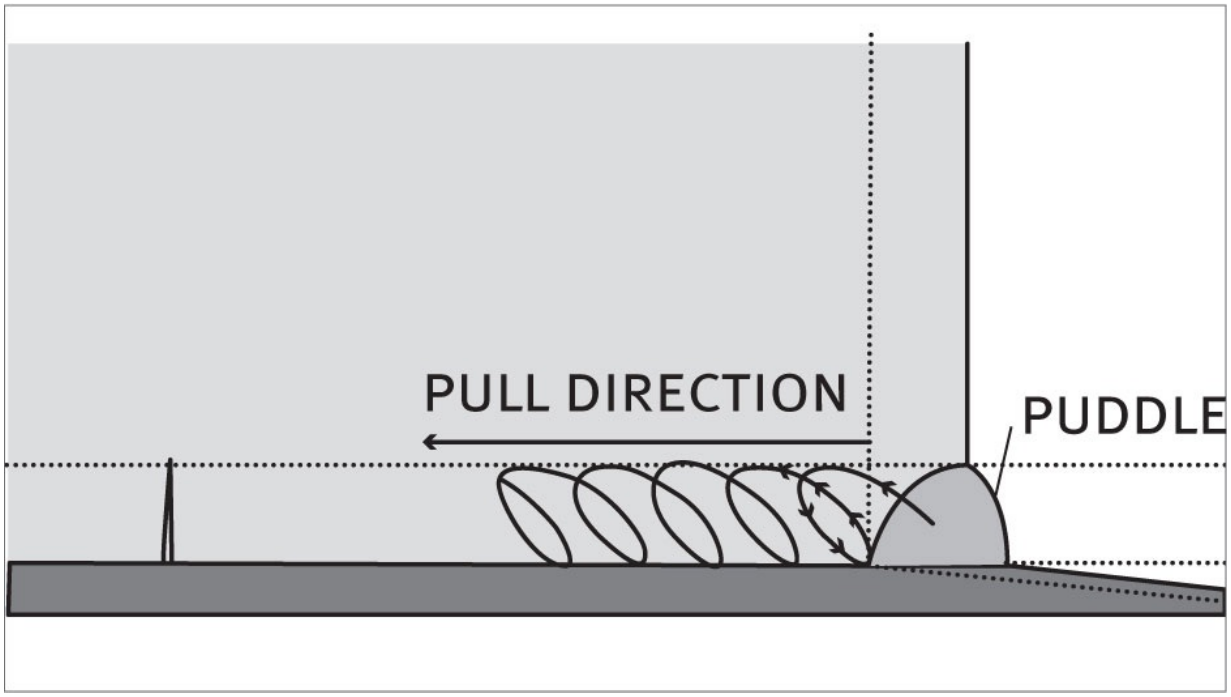
PULL DIRECTION
PUDDLE

With this looping technique, always move in an up and over direction. Think of a wheel rolling.

When tacking tee joints, it's a good idea to go from one side to the other in order to counter the movement from the heat going into the steel.

After the pieces have been prepped and secured in place, tack them together in a way that moves furthest away for the last tack. Set the MIG for 1/4-inch (6 mm)-thick material. Hold the gun at a 45-degree angle, go into the corner, and touch the wire down to tack the pieces together.

Now that the piece is tacked together, clean the tacks with the stainless steel wire brush in order to prep for welding. To assist in fighting gravity, the movement of the technique, or shape, is an inverted cursive lower case "L" or "E" pattern, as shown in the diagram on the left. Leave the machine set for 1/4-inch (6 mm)-thick stock. Set up with the gun at a 45-degree angle. This movement will begin by starting the puddle at the base of the forward tack. As the puddle develops to the correct size, move the arc up and down in this pattern while pulling through the weld zone.

These elongated circular movements will create a sort of vortex or cyclonic action that will help push the puddle back up onto the vertical piece of material, not allowing the puddle time to toe out. The speed and length of this technique will be determined by the material thickness.

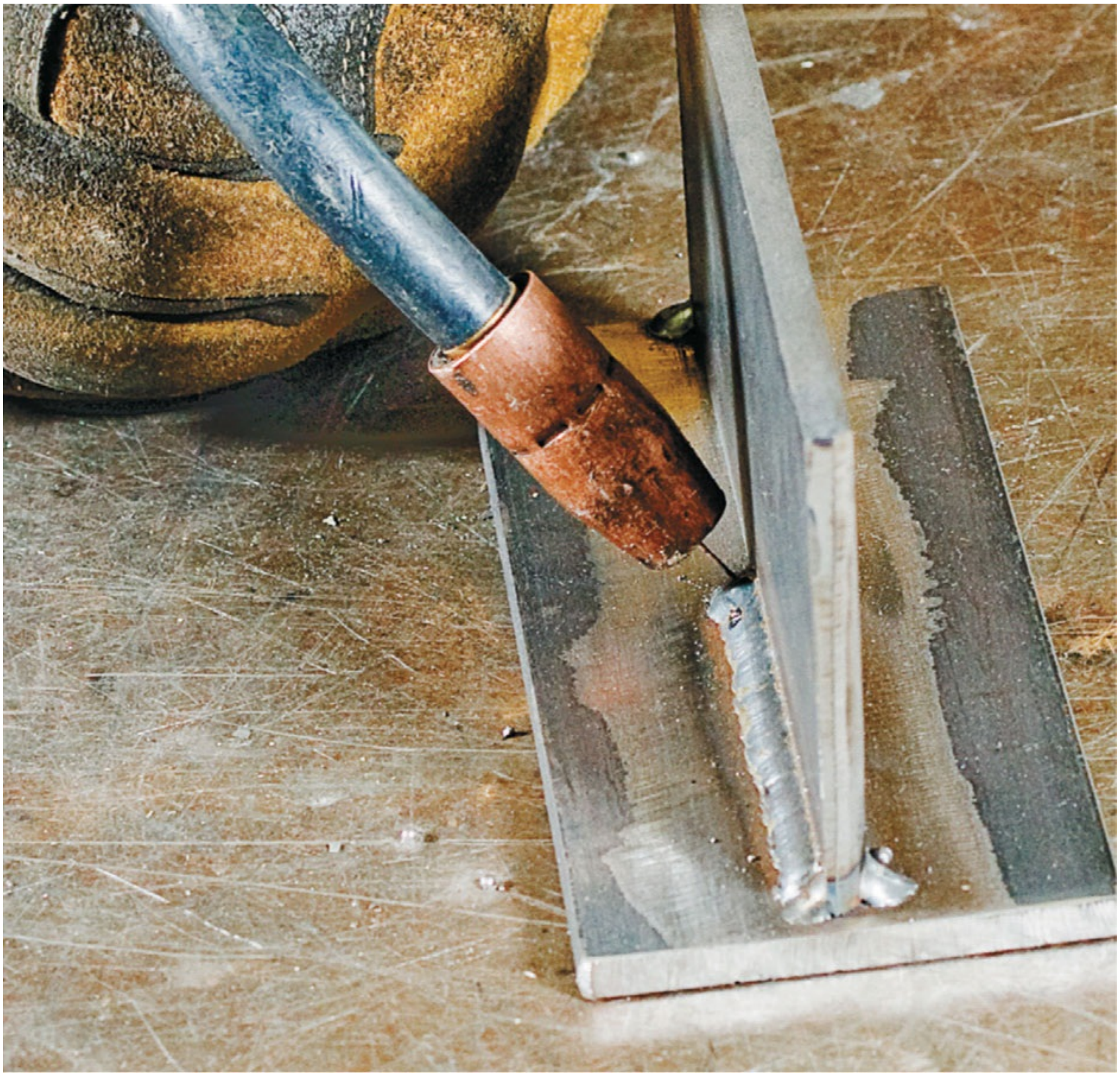

Split the angle of the corner in half so that the nozzle is 45° on both sides. Make sure the shield gas is hitting the puddle evenly.

As the material becomes thinner, the shape of the pattern shortens and becomes more of an inverted cursive lowercase “L.” As the metal increases in thickness, the more the shape elongates and becomes more of a cursive lower case “E.” This pattern can also be inverted for different effects on the surface of the puddle.

When going in to a corner from the right, I recommend using a counterclockwise motion as shown in the diagram on the right. This way, the movement of the arc is on the downward direction, then up again to pick up its own tail. Going in from the left would be opposite direction, clockwise. This means that on either side the motion is always up and over towards the direction of the weld.

Just like any other technique, this will only apply to thicker materials. The thinner the material becomes, the smaller the puddle and the better chance of burn though. Practice this technique until the edges are consistent, even, and the legs are equal to the thickness of the material.

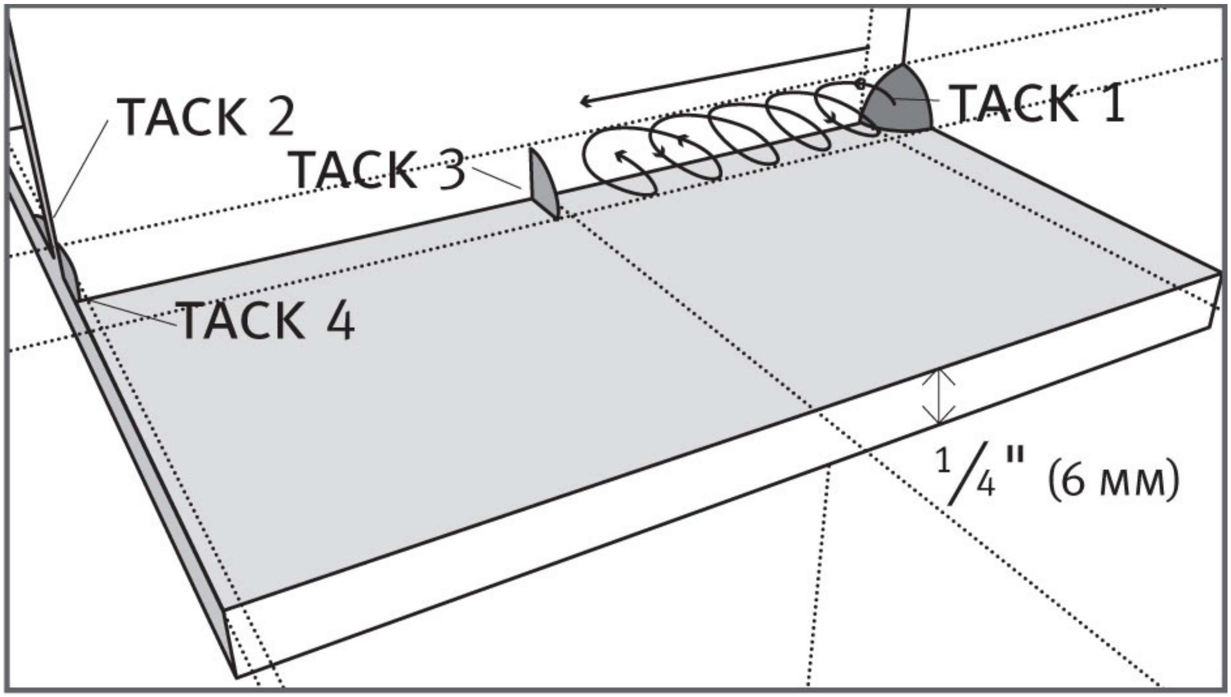

The arc motions rolls like a wheel, up and over.

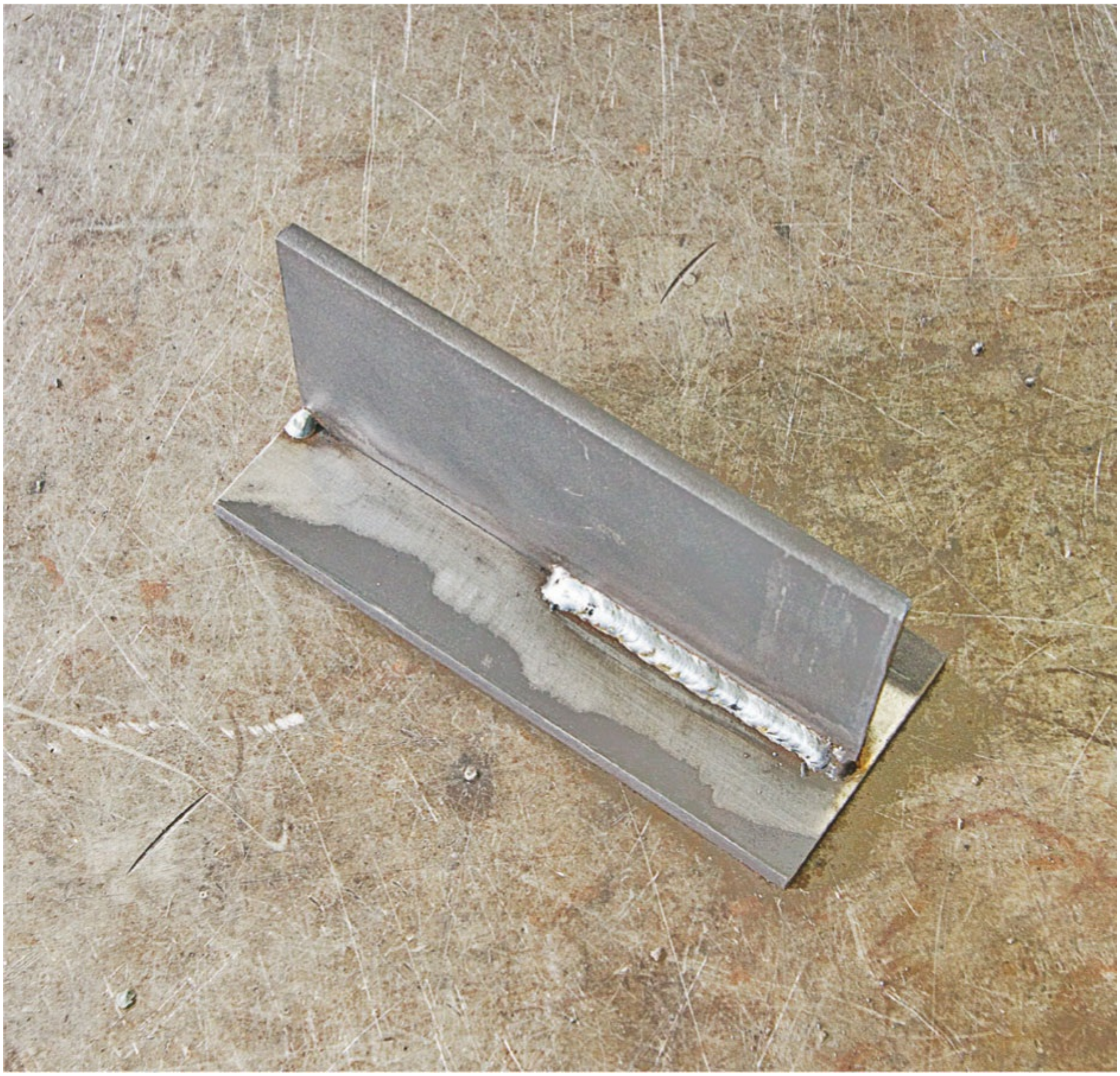

Clean the weld zone with the stainless steel wire brush and then continue the weld.

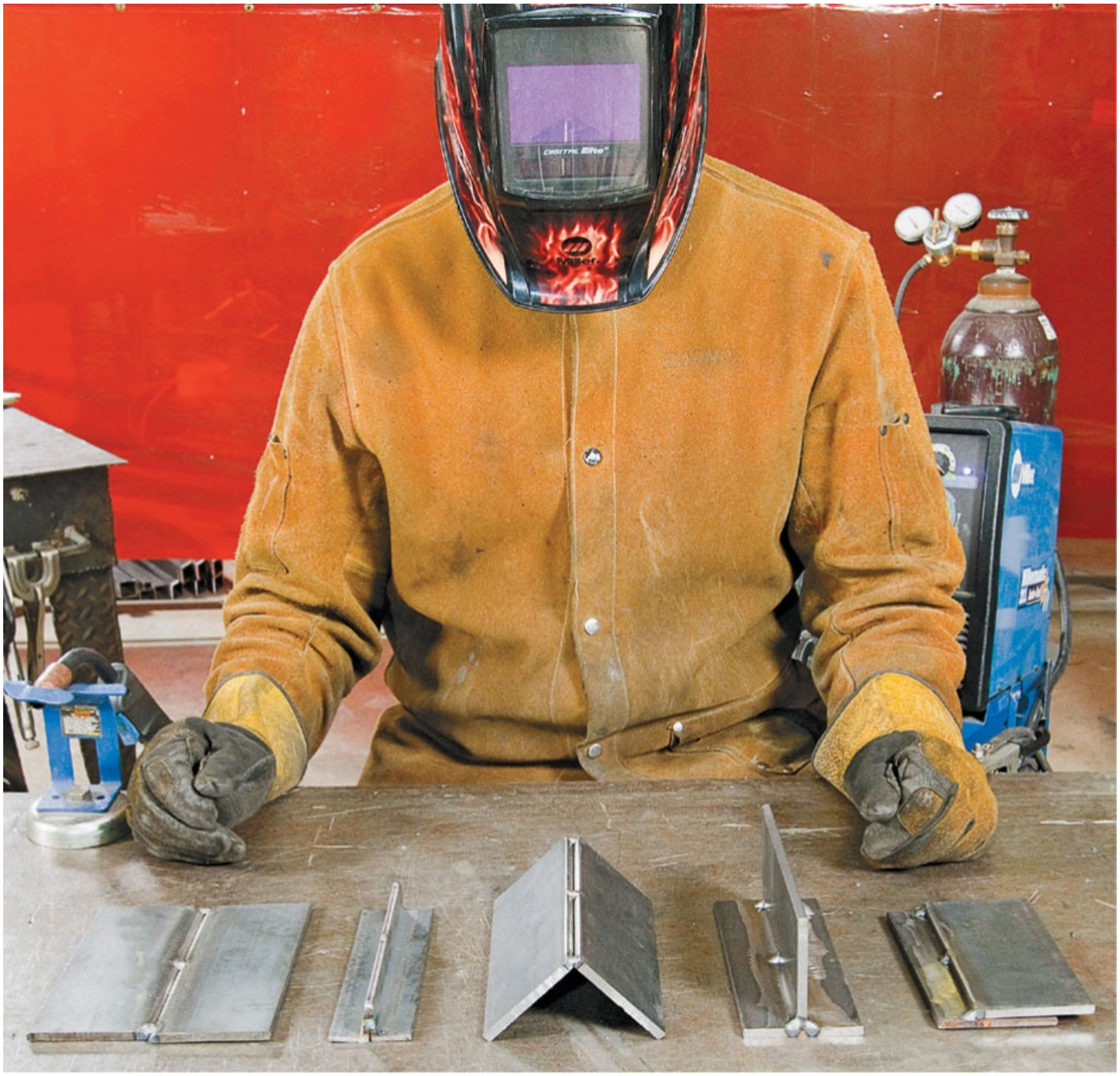

Here all the basic joints are tacked together, prepped, and ready to weld.

The Rest of the Joints

After practicing the techniques for butt joints and tee joints, the rest start to fall into place with relatively few problems. The hard part is out of the way. Let’s touch base on how the rest of the five basic joints weld together.

Corner Joints

Corner joints can use the same technique that is used for butt joints. The technique used is determined by the way the corner joint is positioned. If it’s an open outer corner joint with each leg 45 degrees from the floor, then it is better to use more of the back-and-forth shark tooth pattern, like a butt joint.

As with butt joints, the movement of the draw will always be in relation to the motion back and forth, dipping just under the A and C edges a little as you go. Because of the way the heat is distributed in a corner joint, the edges will break sooner that a butt joint. Remember that full voltage and wire speed combination may not be needed when doing a technique for this style of a corner joint.

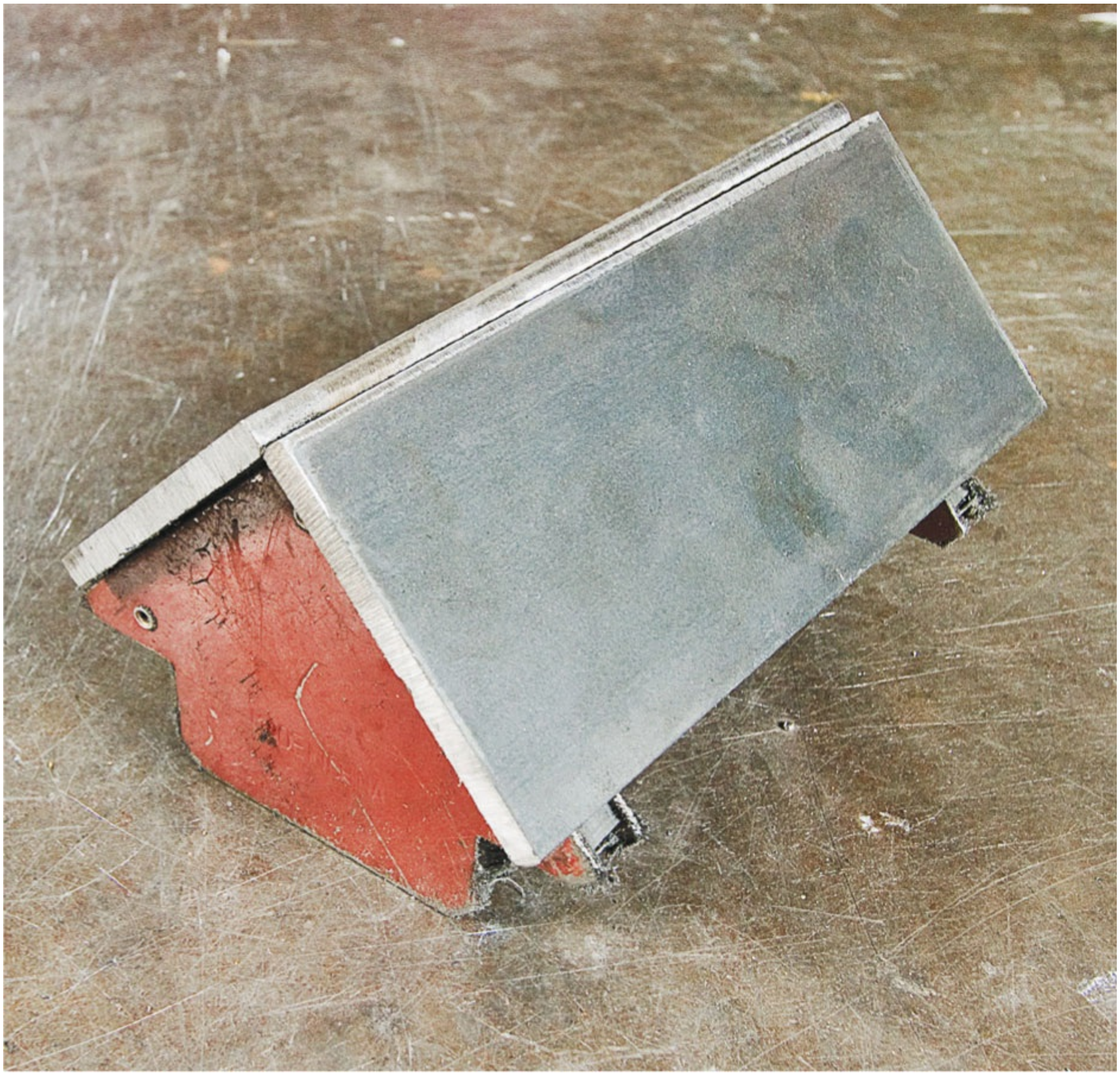

Corner joints like this will use the same shark tooth technique as a butt joint.

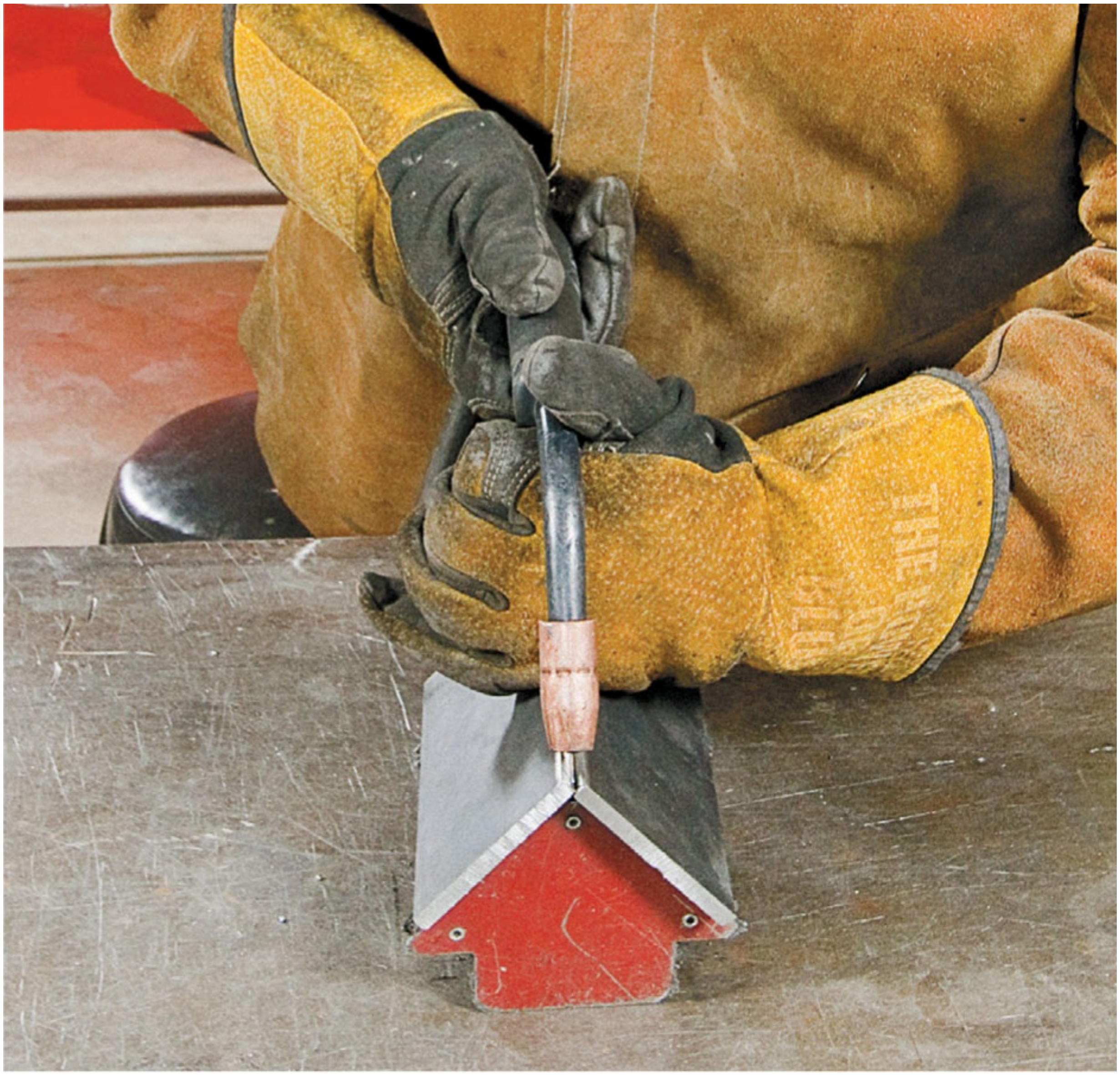

Start tacking by putting the wire down at the bottom of the joint.

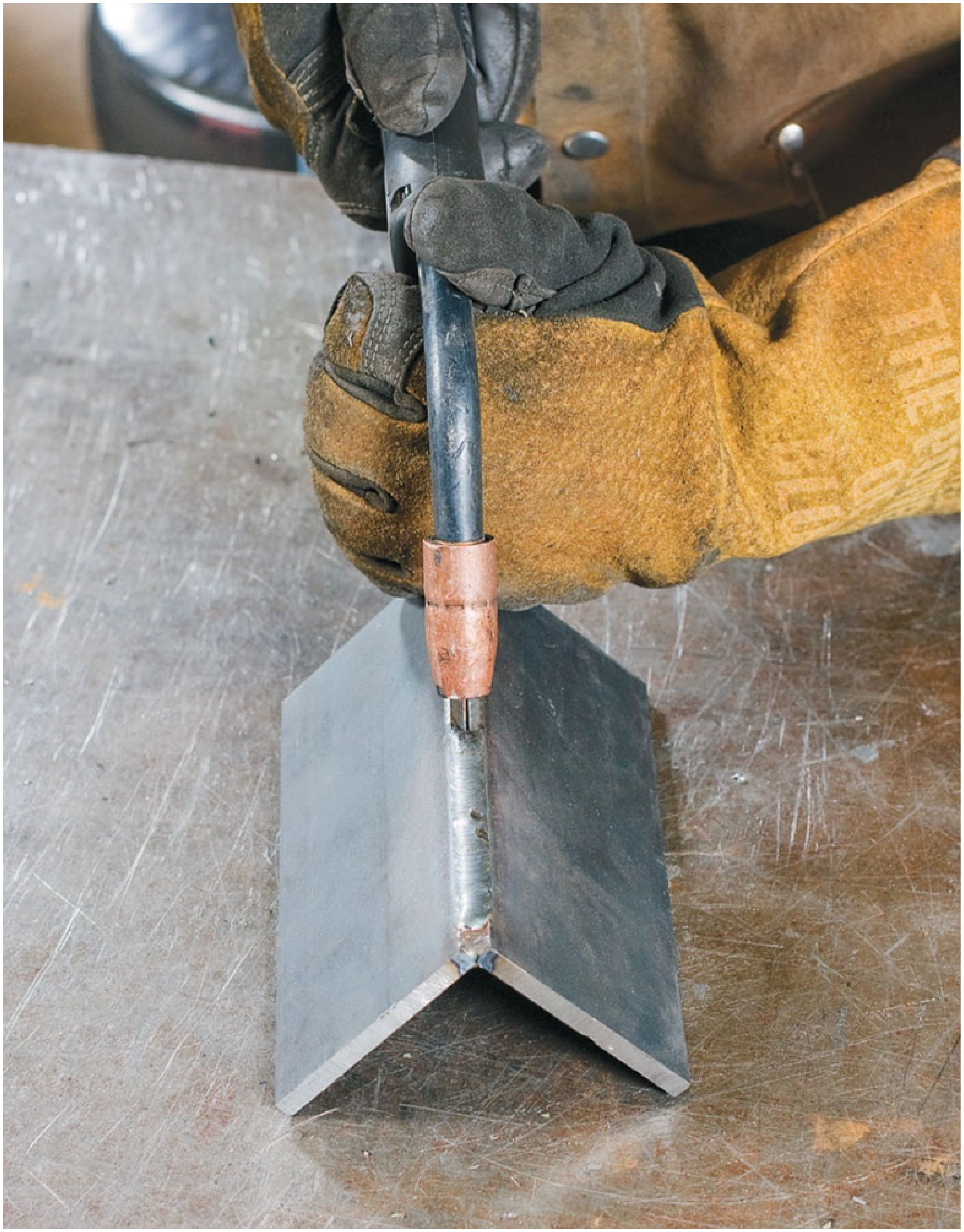

The same rules apply to all of these techniques—straight on and perpendicular so the shield gas hits the puddle evenly.

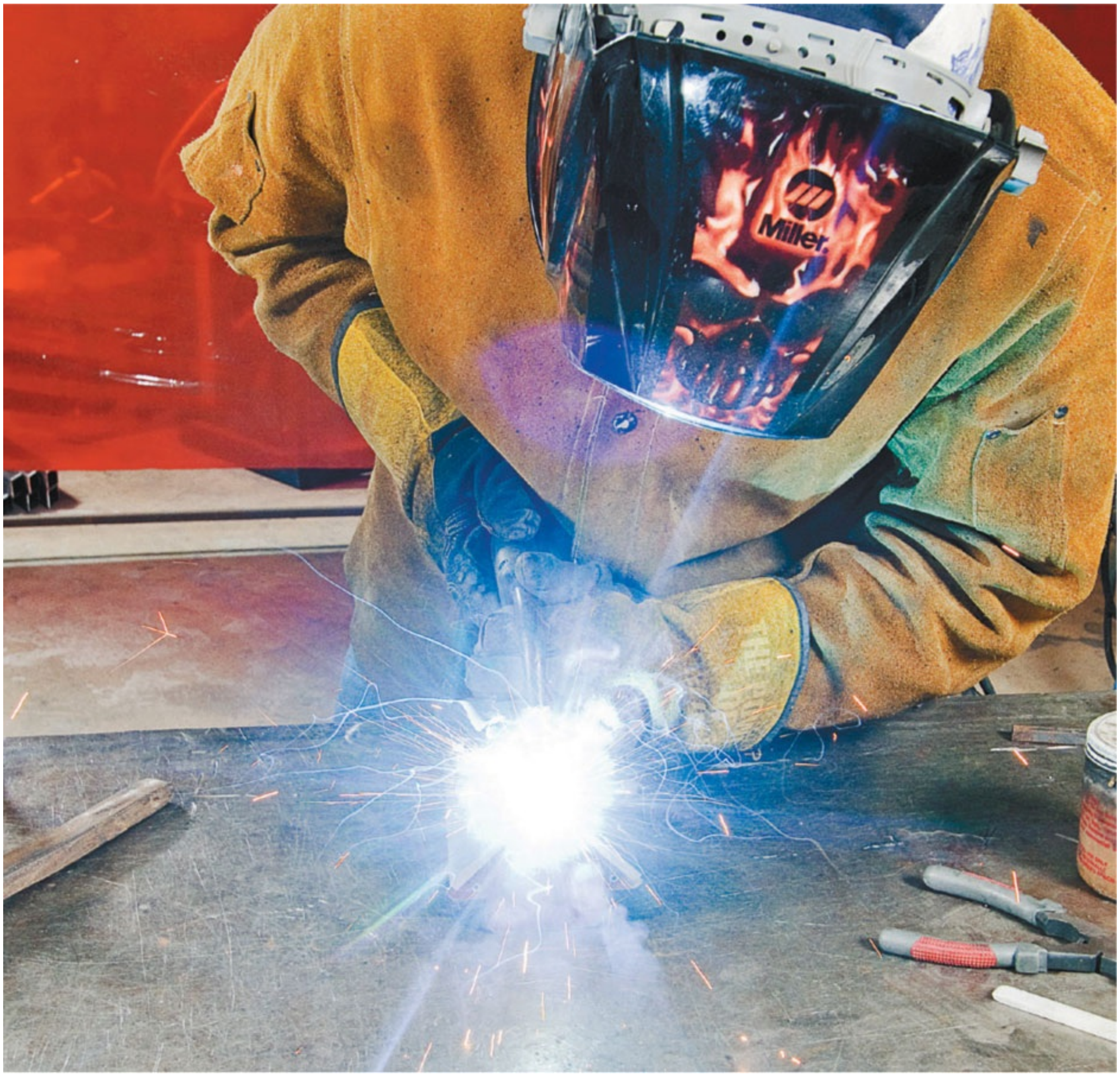

Try to remain braced and supported. If the weld is higher up, find something that height to rest your support hand on that will slide on the table.

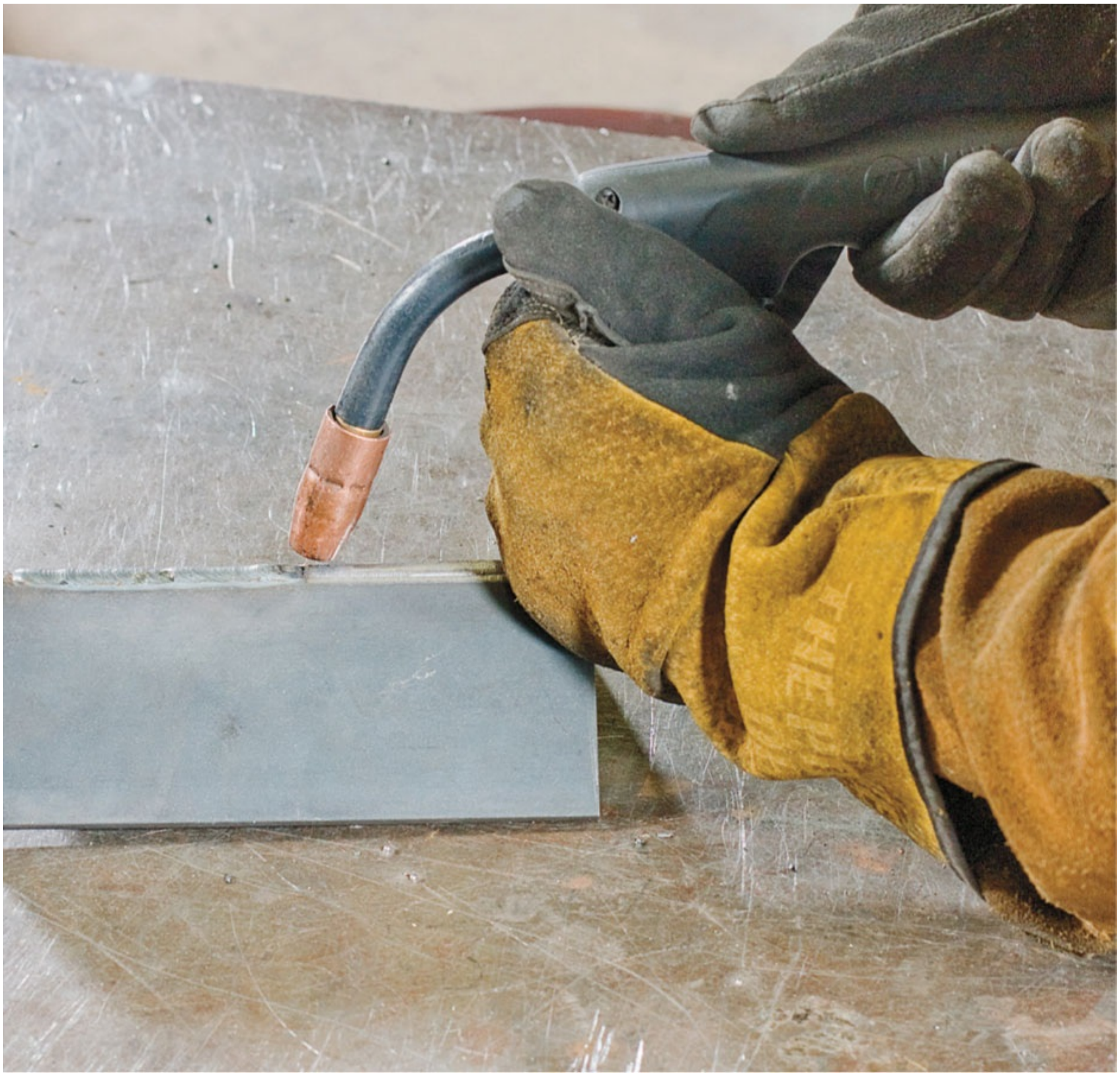

You should have a drag angle of no less than 45 degrees.

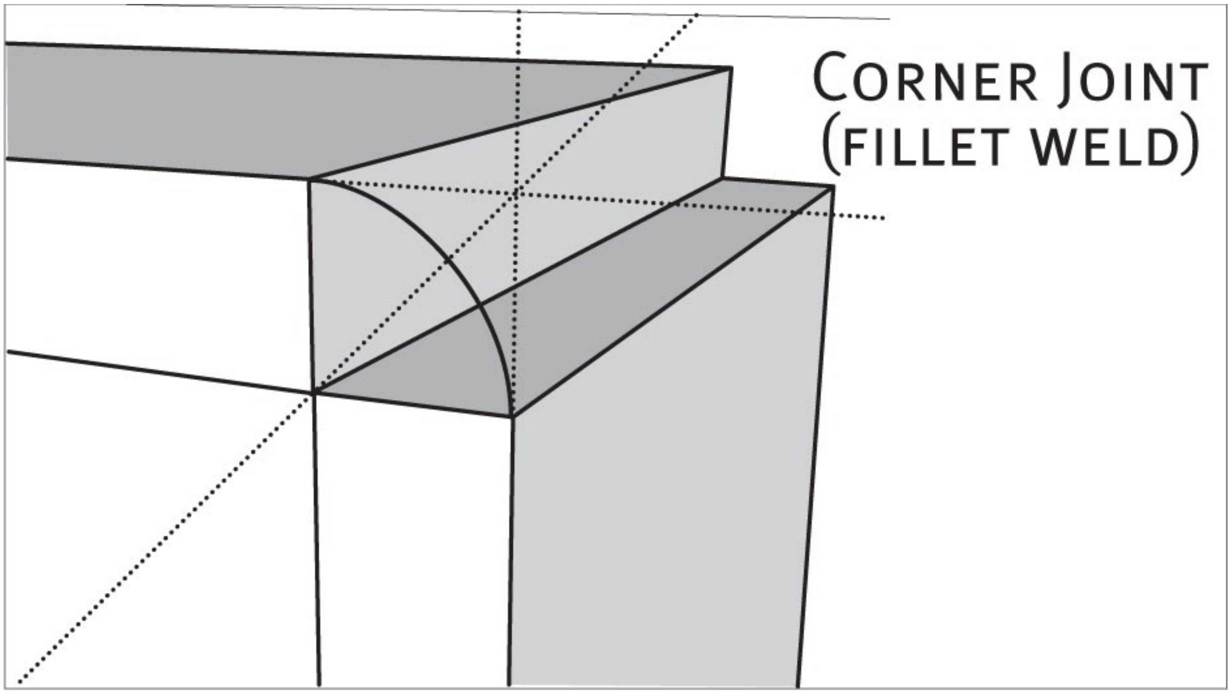
CORNER JOINT
(FILLET WELD)

For outer corner joints, it is only necessary to break the edges. This will happen faster than butt joints.

This is a great general purpose technique that can be used on all sorts of projects. The more you practice, the better you'll get.

Lap Joints

You can now begin to see where I'm coming from in regards to the welding techniques for the basic five joints. They're all similar in the way they are set up and executed, with very subtle differences in the way they are welded. Gravity, the HAZ, and the angle are all factors that determine the way in which to conduct a weld.

The lap joint, for example, is essentially a tee or corner joint. The most noticeable difference is that one of the surfaces will spread the heat out through the rest of the material, where the other surface has an edge to break.

The technique is the same as the fillet weld on page 116. Go into the corner at a 45-degree angle, create a puddle, and then draw back with the same pattern, breaking the top edge of the piece.

Edge joints radiate the heat straight down. You may not need to use a puddle control technique on them. In some cases, it is even necessary to dial down the voltage and wire speed.

The Edge Joint

Most of the time an edge joint will be made with a straight draw. The way the two pieces are put together doesn’t allow for the heat to be spread out the way it would for a butt joint or a tee joint. For an edge joint, the heat is more immediate. Since the heat is concentrated directly down, the puddle is essentially following the heat, which makes it more intense. The amount of time for the puddle to develop and to create a bead is almost split in half. On thinner metal, dialing down the machine setting is always helpful to create a bead on an edge joint.

I have to say that the edge joint is my least-used joint. I've only needed it a few times for fabrication projects that usually have something to do with angle iron and building steel staircases.

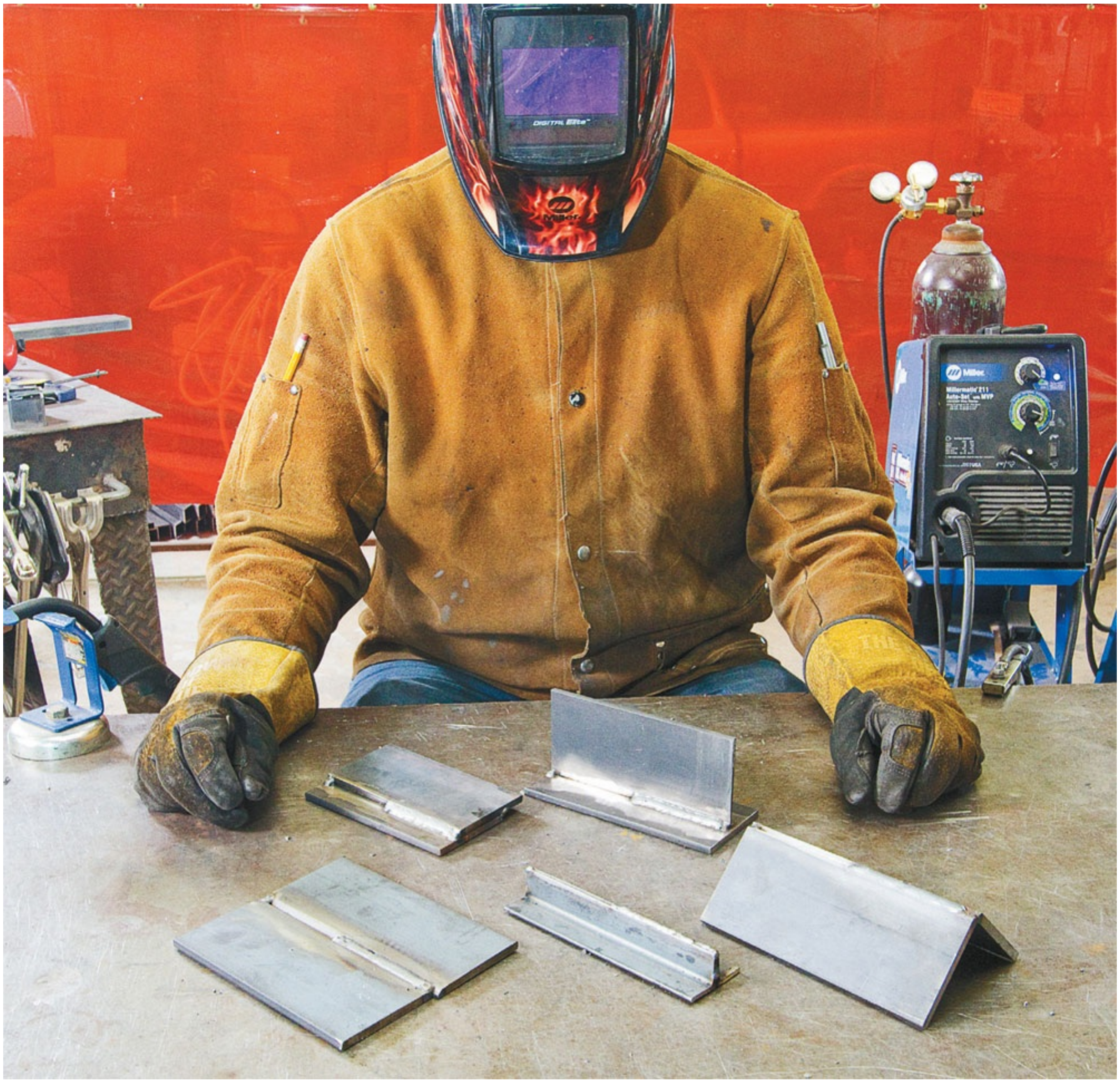

Here are all the basic weld joins. Practice arc movement and puddle control on thicker material. On thin metal, a more traditional straight draw will work best.

So there it is. All the basic weld joints welded together. Practice all of these techniques. Understand that the thinner the material gets, the less distance you techniques. Understand that the thinner the material gets, the less distance you can travel with the puddle and the less technique you can use. Techniques, such as the ones taught in this book, will only apply to thicker materials. I will use a similar technique down to about 14 ga; when the metal gets thinner than that it’s time to start doing a straight draw. Once you are comfortable and feel that you are laying down consistent beads on both thick and thin metals, it’s almost time to start making things.

Exercise 3: Jigs

DIFFICULTY: BEGINNER

MATERIALS AND EQUIPMENT:

- MIG welder
- Two 2 to 3" (5–7.6 cm)-long pieces of scrap 1 1/2"(3.8 cm)-square tube
- 10" (25.4 cm)-square section of 14 ga (or thicker) sheet metal
- Welding helmet
- Welding gloves
- Long-sleeve shirt
- Leather welding jacket or apron
- MIG pliers
- Stainless steel wire brush
- Soapstone
- Carpenter or framing square
- 4 1/2" (11.4 cm) angle grinder outfitted with an aluminum oxide grinding disk or a 60-to 80-grit flap disk

In addition to knowing basic welding techniques and joints, it’s also useful to know how to make a jig. A jig is a custom-made tool that is used to limit the motion of another tool or the motion of the piece clamped to it. This allows the project to be worked on with more accuracy.

For this jig, you’ll need a few pieces of square tube. These can be drops you have lying around or ones you cut. These should be about 2 to 3 inches (5 to 7.6 cm) in length; they need to be long enough so that a locking C-clamp can clear one of the sides to be secured. You will also need a piece of sheet metal to weld the two pieces of square tube on. The best thing about this jig is that it will line up not only one edge but up to three axes for projects like The Cube in the next chapter.

You can use almost any scrap laying around the shop to make a jig.

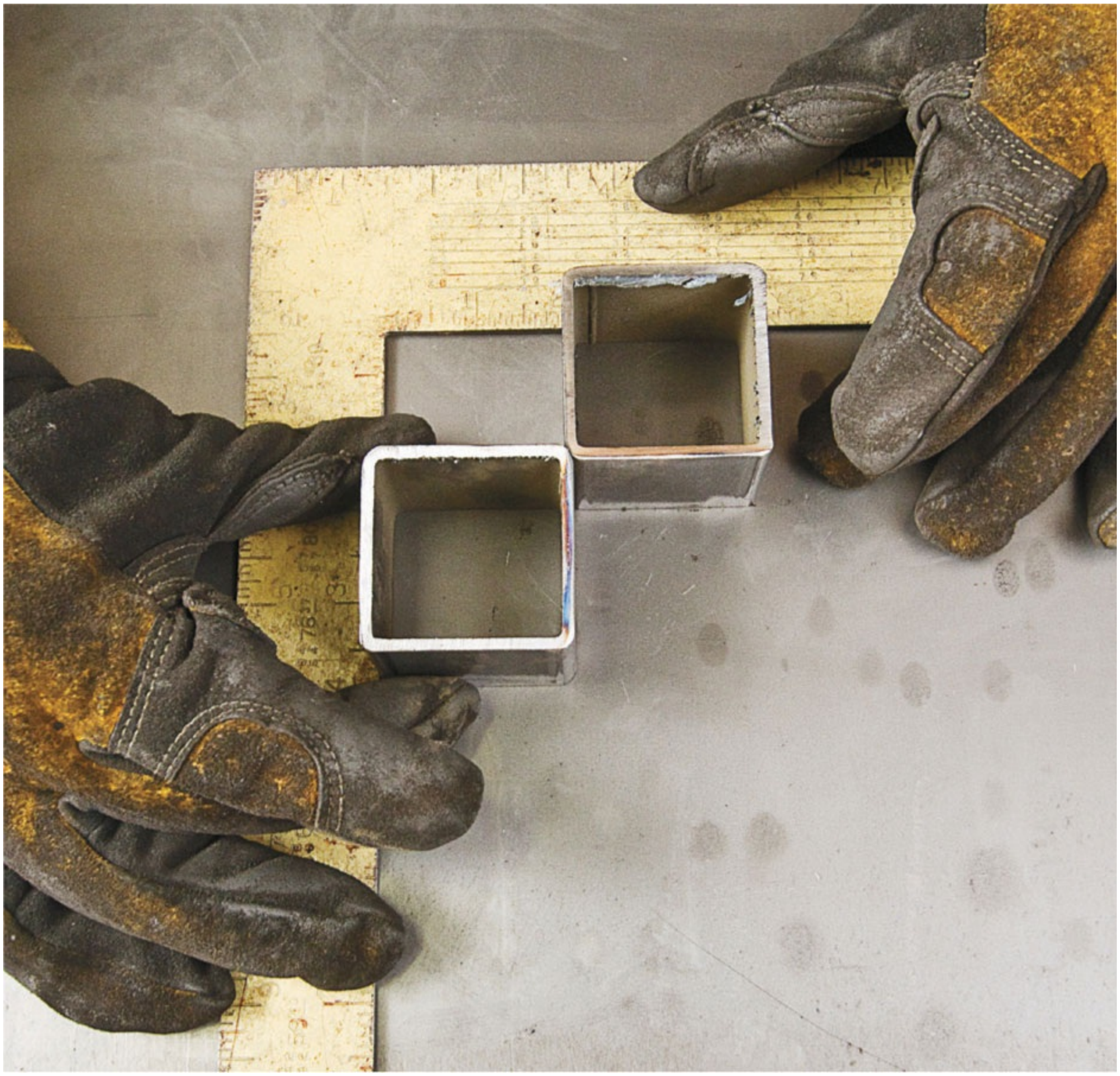

To make a right angle jig, use two pieces of square tube and a carpenter square.

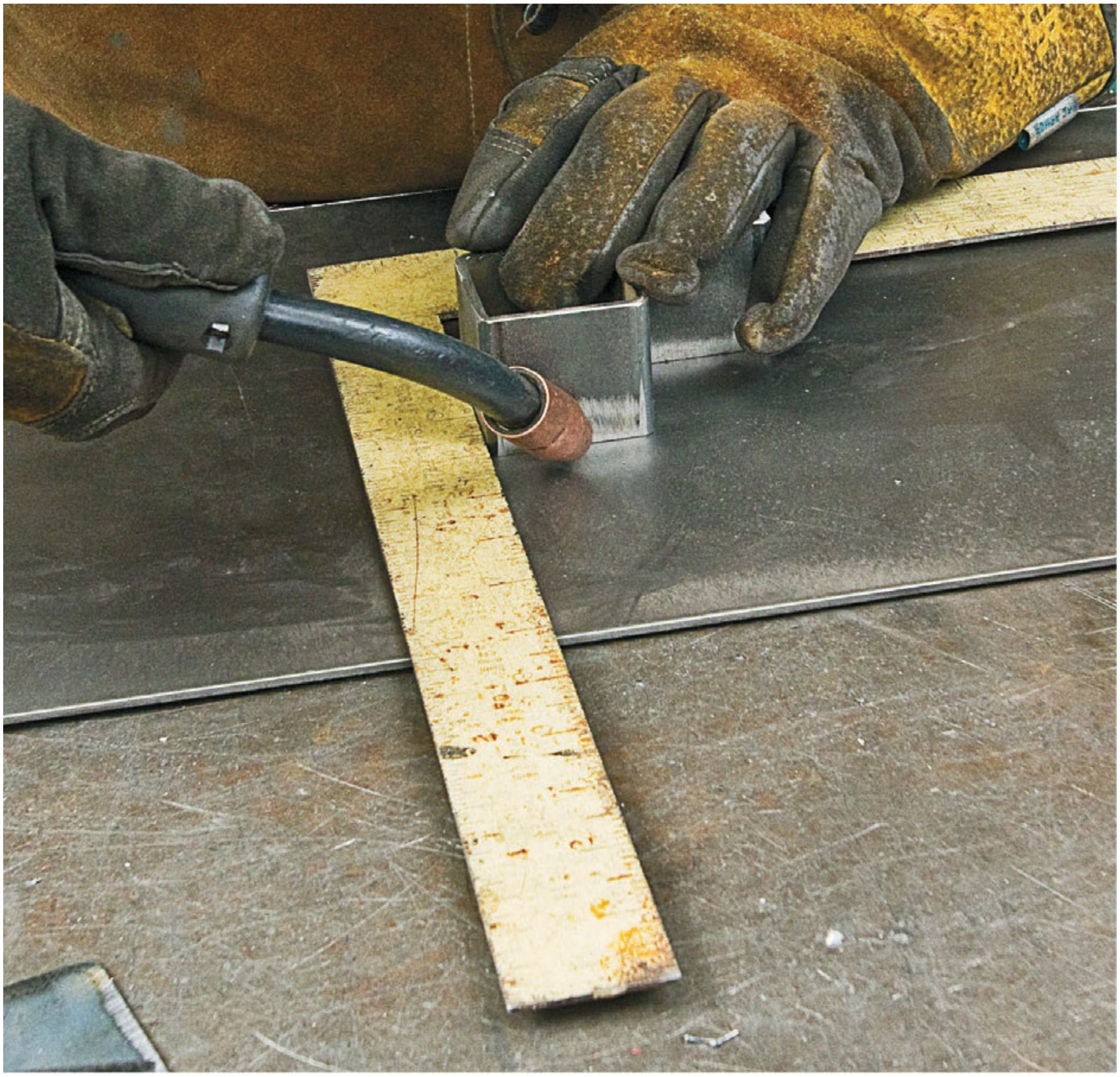

When the pieces are perfectly square, tack them into place. Avoid tacking on the side the material will be clamped to.

Cut two lengths of square tube to 2 or 3 inches (5 to 7.6 cm) in length. Line up the pieces of square tube so that the corners face each other. Using a square, line up the two outermost sides of the two pieces of tube.

Tack the tubes to the sheet metal with a fillet tack (see page 116). Make sure not to tack on the outermost edge, as this will be the edge that you'll clamp the sides of your work pieces when you use the jig. Tack the first tube into place and then check that the second tube is still square with the other. After both pieces are tacked into place, check again for square. If they are not square, cut the tacks off of one piece and re-tack until they are square. With the jig squared and tacked in place, it is ready for use.

Other Welding Situations and Positions

Before we get to building things, I want to show you a few variations of how to conduct a weld. Sooner or later there will be a situation where you don’t have the luxury of welding on a flat surface. There will always be a need to be able to weld from different angles and directions.

It’s great to work on a small piece on a nice flat work surface. Small pieces allow more freedom of movement. If you’re not comfortable welding in a certain direction, or the piece is too far away, or if the piece needs to be welded in a different location, simply pick up the piece and situate it in a way that allows you to be comfortable and have the best angle.

But what happens when you’re working on something larger than the table, like another table or maybe a vehicle? When the piece becomes too heavy or too large to make adjustments, you have to be the one to adjust. Putting yourself into situations where you have to do vertical welding or adjust the position you normally weld from will help you develop your skills as a welder.

Vertical Welding

There are two methods of vertical welding, uphill and downhill. It is much more difficult to execute an uphill bead than a downhill bead: either you have gravity with you or against you. However, despite being harder to execute, uphill welds are stronger because they give you deeper penetration.

An uphill weld is the movement of the puddle up a weld zone with a push motion. The advantage to uphill welding is that the puddle penetrates deeper into the metal from the force of gravity pushing down on it. Since gravity is against the puddle, it wants to eject itself from the weld in progress, meaning the fluid that makes up the puddle wants to fall out of the weld. This makes it very difficult to control. This style of welding is normally used in industrial situations such as building ships, bridges, and other structures that require additional penetration.

I’ll confess I don’t do much uphill welding because most of what I do does not require an uphill weld, nor do I work with material thicknesses that would require more penetration. There are also a few tricks that I have learned over the years to give me better penetration using a downhill weld.

Setting up for a vertical downhill weld is basically the same as you would for any weld: the nozzle still has to be straight on and perpendicular, you should still be as comfortable as possible, and you should still support and brace the gun so as comfortable as possible, and you should still support and brace the gun with your opposite hand. What does change is the drag angle. I exaggerate the angle by positioning it under the puddle so that the shield gas actually helps decrease the speed of the fall. Then, with a back and forth motion technique, I slowly draw the bead out. I call this pattern the falling leaf. This pattern helps you control the puddle’s fall and keep it in place longer.

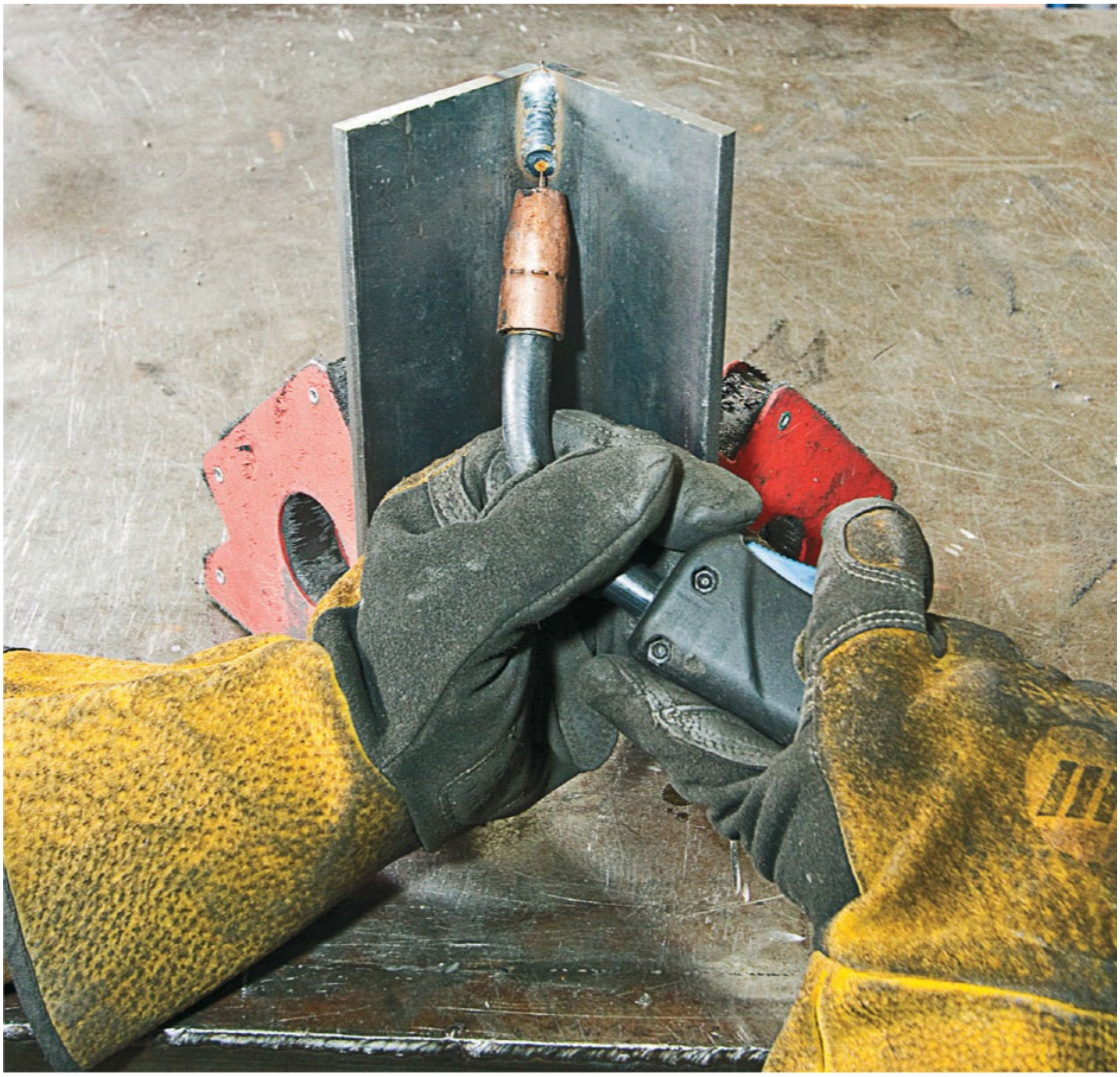

When vertical welding, always try to find a way to support your hands; if you only have the ability to use your pinky, use your pinky. Magnets come in handy as supports when vertical welding.

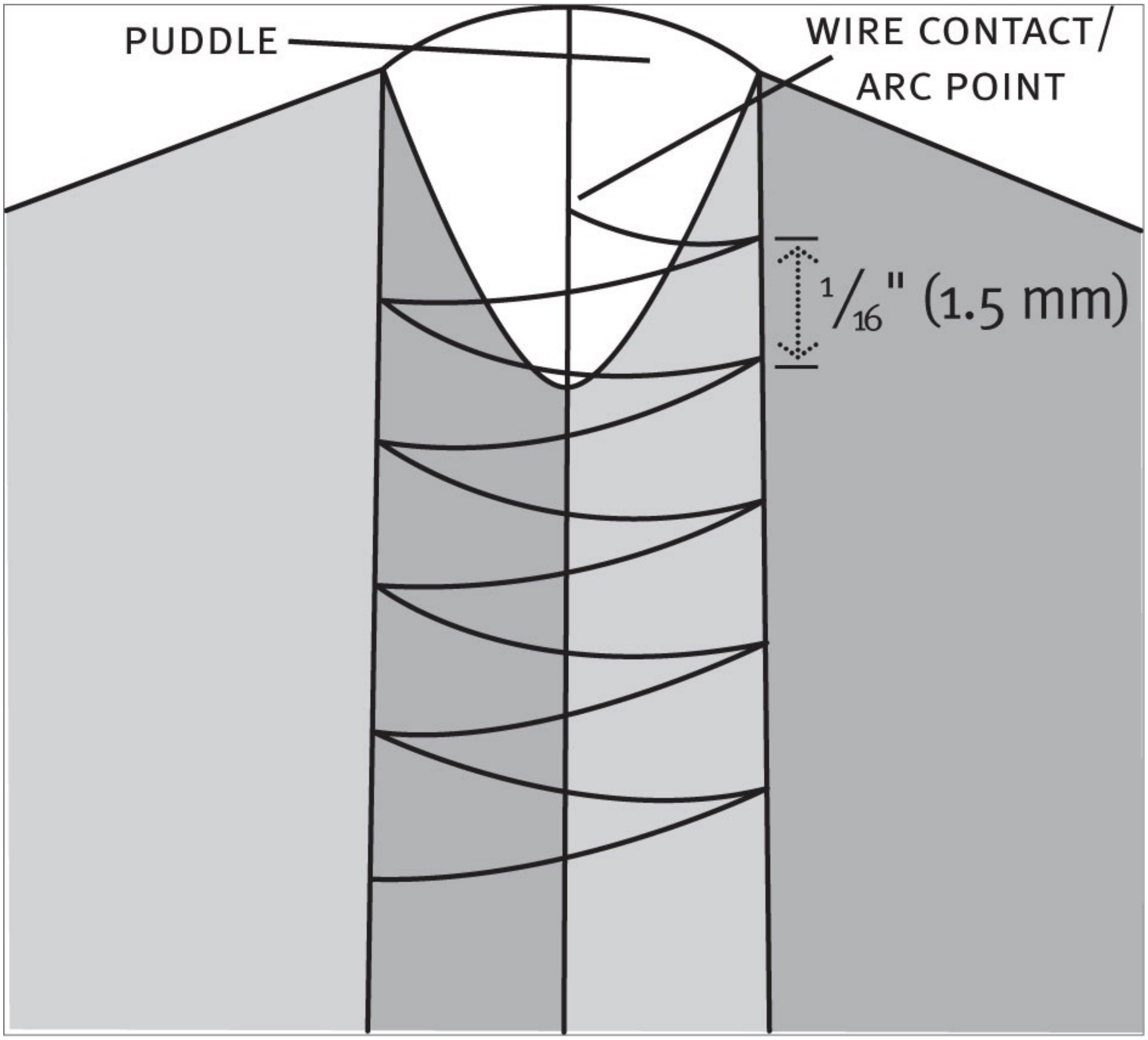
PUDDLE
WIRE CONTACT/
ARC POINT
1/16" (1.5 mm)

Like the shark tooth motion for a butt joint, build the puddle and move the arc in a falling leaf pattern. For every motion in one direction, let the puddle drop 1/16 to 1/8 inch (2 to 3 mm).

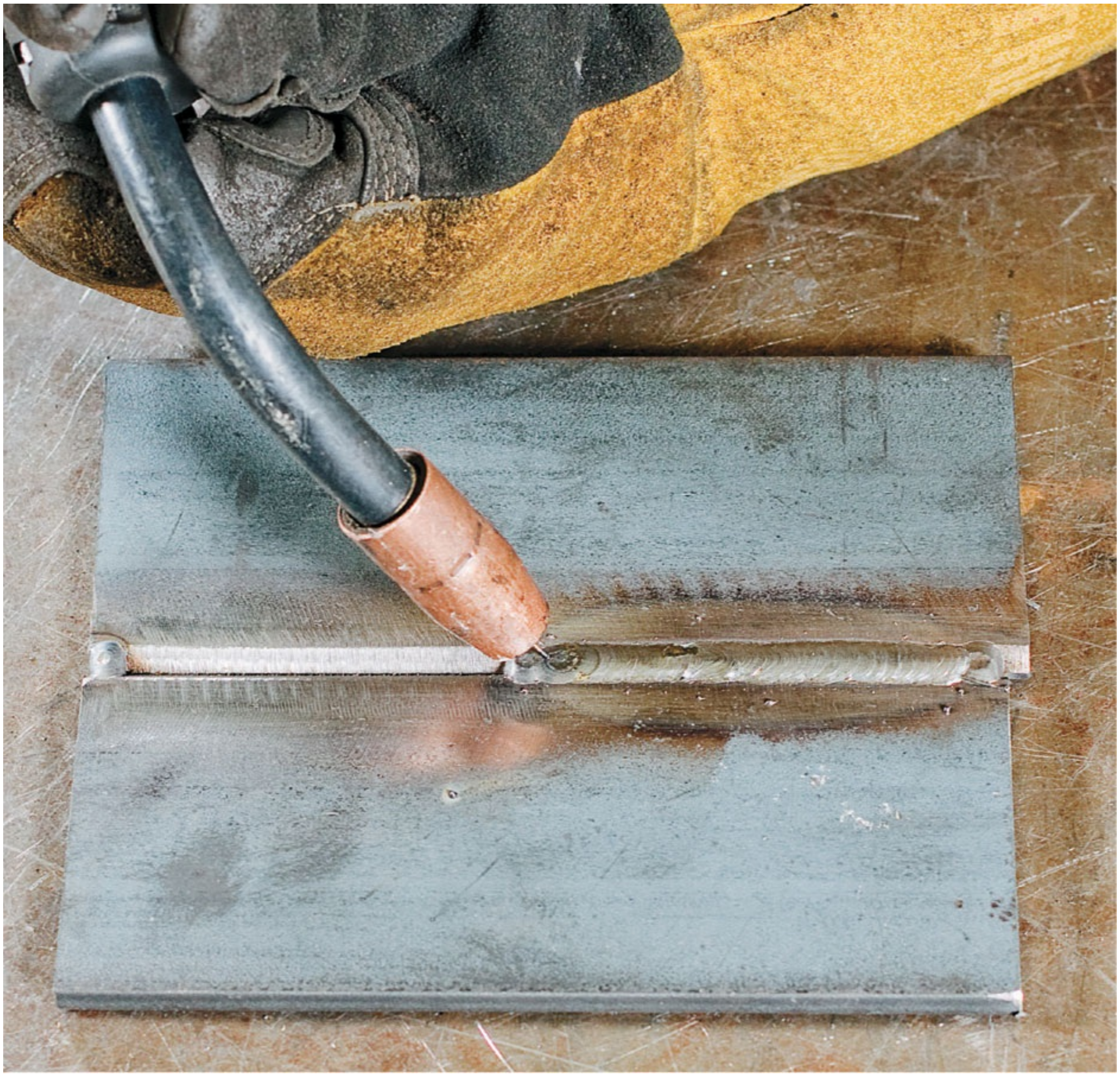

As you can see in this cross body weld, the direction of the nozzle is still straight into the weld and I am perpendicular on both sides. In order to do a cross body weld, I don't break the rules, but I do have to position my trigger hand towards me a little in order to keep those angles and adjust my drag angle.

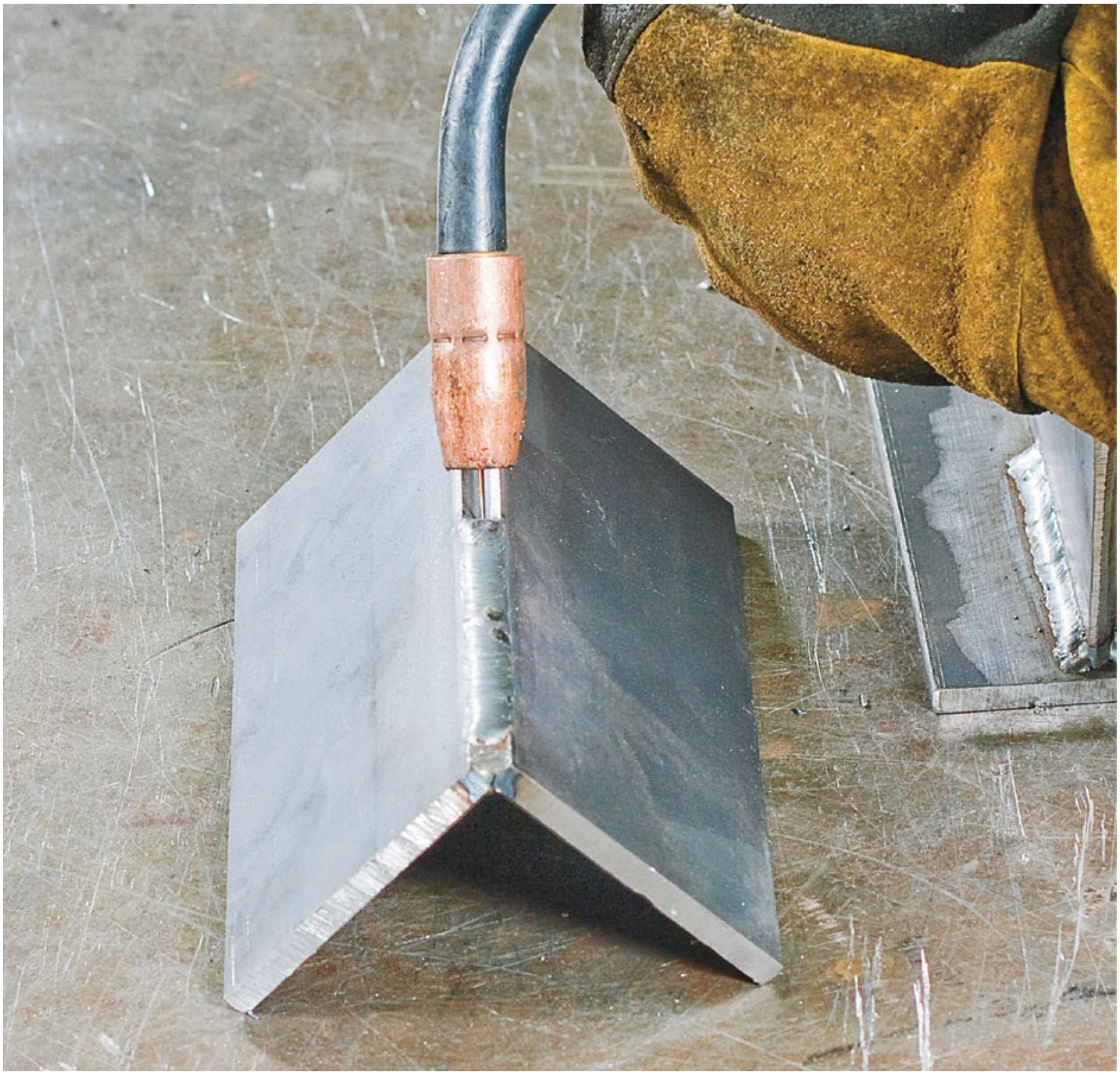

As shown in this image of welding a corner joint cross body, the nozzle is still straight on with an even angle on both sides. Again, weld straight on and perpendicular with a good drag angle.

Cross Body Welding

Another great method of welding that helps a lot is cross body welding. You will see a lot of cross body welds in production welding. Set up so that the line of the shoulders is parallel with the line to be welded. A lot of welders prefer this method because it allows for better vision of the weld in progress. Instead of pulling toward the body, this is a movement across the front of the body.

The unbreakable rules do not change. The drag angle of the nozzle still has to point into the weld, you still need even angles on both sides of the puddle, and you still need a consistant 3/8 inch (9 mm) stick out.

Now that you know a few variations of welding positions and styles, you can develop your own way you would like to set up and weld. As long as you don’t break the basic rules, with a little practice you should be able to execute a consistent bead every time.

6

Projects

By now you’ve mastered the basics and can draw a bead, tack, and weld a variety of joints and even a jig. Now we’re going to put everything together with a couple of projects that will hone your skills even more.

Don’t forget to follow the safety procedures we discussed in Chapter 2. Welding is only fun if it doesn’t include a trip to the emergency room.

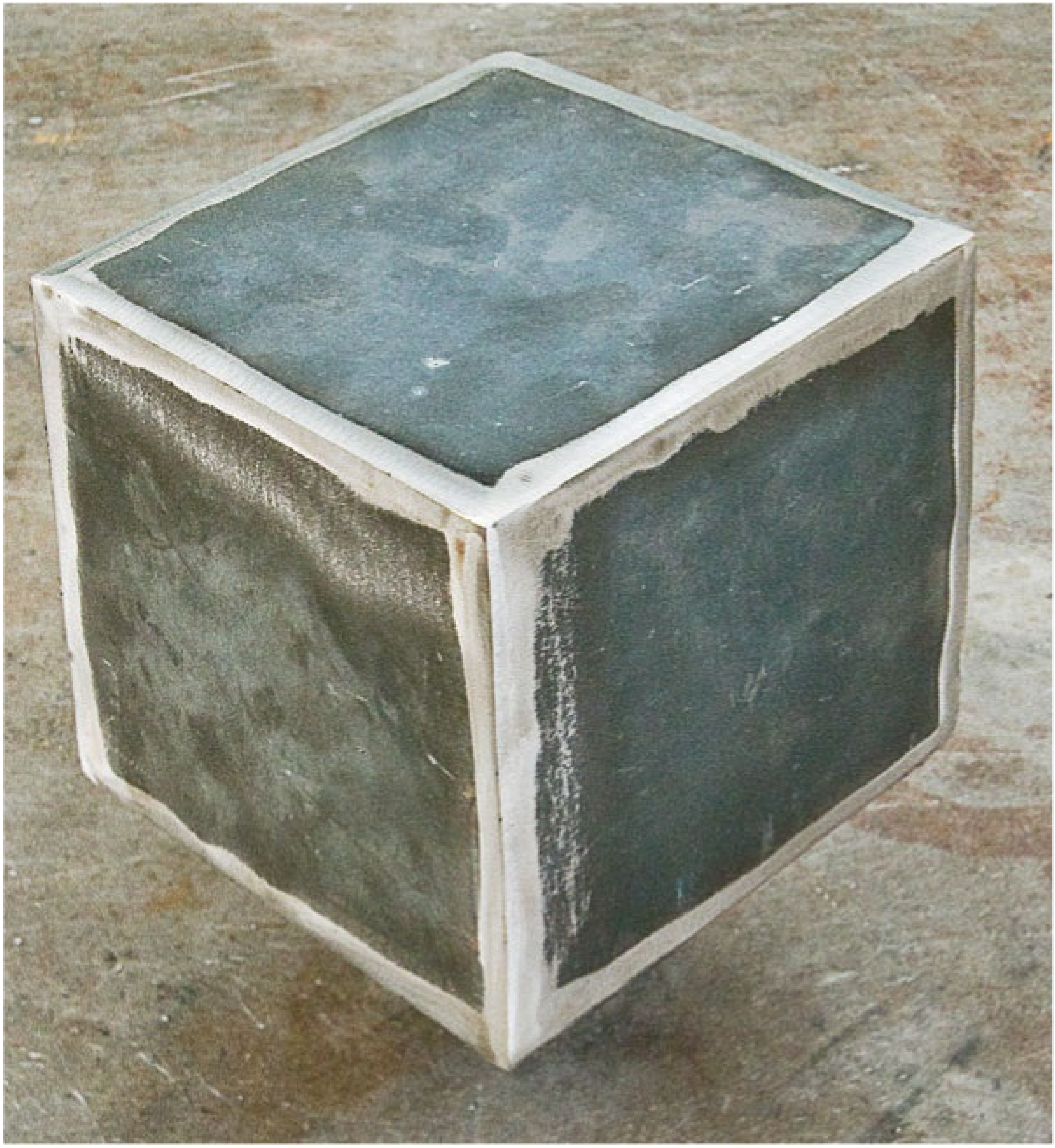

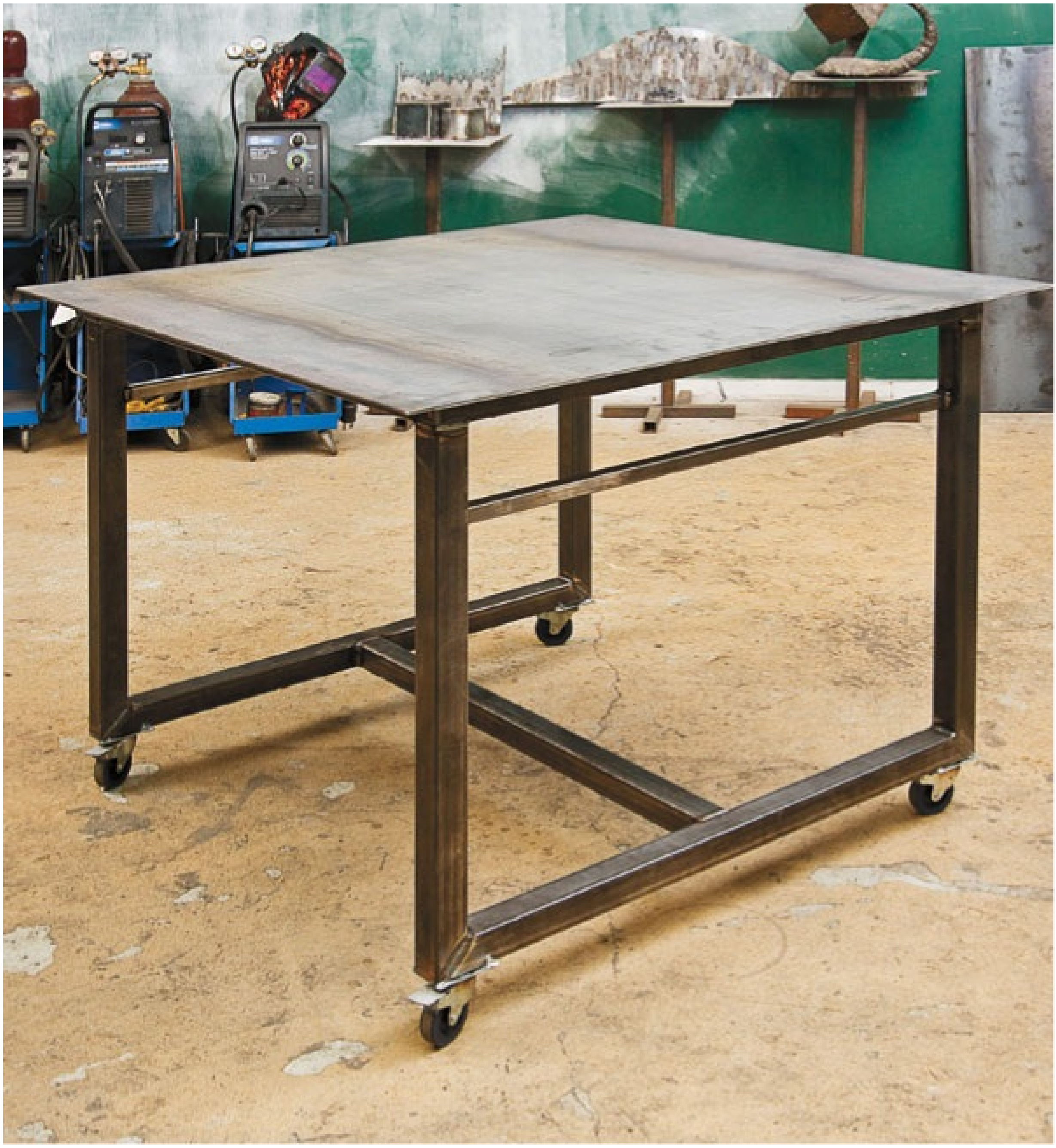

MW
WM

Project 1: The Cube

SKILL LEVEL: BEGINNER

MATERIALS AND EQUPMENT:

- MIG welder
- Two 2 to 3" (5–7.6 cm)-long pieces of scrap 1 1/2" (3.8 cm)-square tube
- 10" (25.4 cm)-square piece of 14 ga sheet metal
- Welding helmet
- Welding gloves
- Long-sleeve shirt
- Leather welding jacket or apron
- MIG pliers
- Stainless steel wire brush
- Soapstone
- 4 1/2" (11.4 cm) angle grinder outfitted with an aluminum oxide grinding disk or a 60-to 80-grit flap disk
- Cutting disks
- Corner jig
- 2 locking C-clamps

Creating a cube has a very important place in my life. It was the very first project that my professor assigned to me as a new metal arts student. I could only use a grinder, an oxy-gas torch, and of course a MIG. When finished, it had to be a perfect 6-inch (15.2 cm) cube. It taught me time, patience, and the importance of accurate measuring. Even more importantly, it taught me the reason why tacking a piece together first before welding is so crucial. That cube tought me so much that I incorporated that same assignment into what I teach today to my new welding students. That is why it will now be your first project.

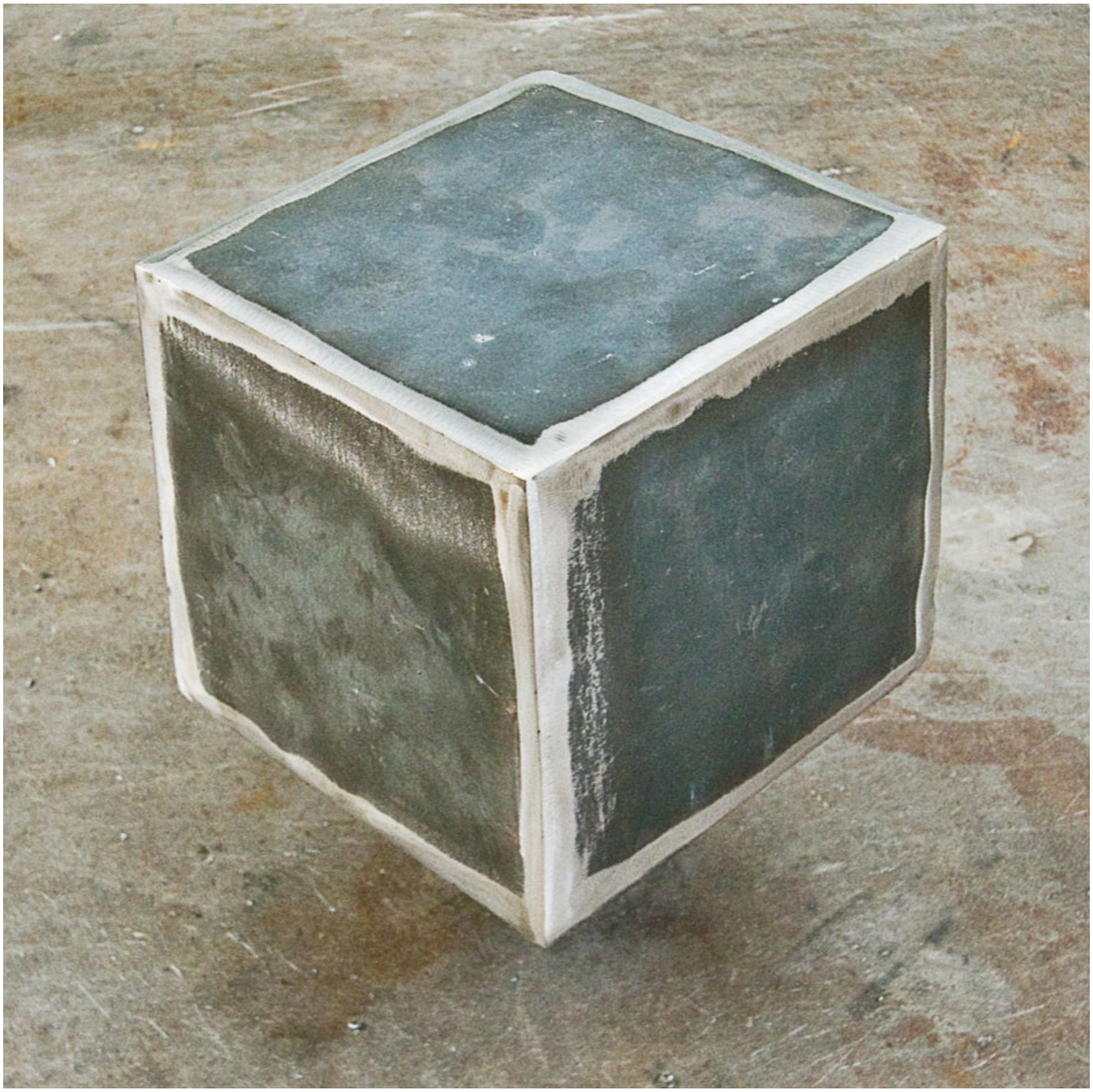

The cube that you will be fabricating will be made of 20 ga mild steel. Every weld on this cube will be outer open corner joint. How you weld them is up to you. The project will be an exercise in welding thin metal, spreading out the heat, and precision. It seems simple enough, right? By following the process of Cut-Prep-Tack-Weld, it will be.

Cut

Start where all projects start in fabrication, by coming up with a list of the parts that you need to make the piece to be fabricated, in this case a 6-inch (15.2 cm)-square cube. You will need to measure and cut six 6-inch (15.2 cm)-square

square cube. You will need to measure and cut six 6-inch (15.2 cm) square sections from a sheet of 20 ga mild steel (hot or cold roll). If you overlook anything, it will probably be the importance of size and shape of each side. Having squares that are exactly the same size will help you avoid a world of trouble.

After cutting out the squares, even them up by clamping them together so that two of the squared edges are flush.

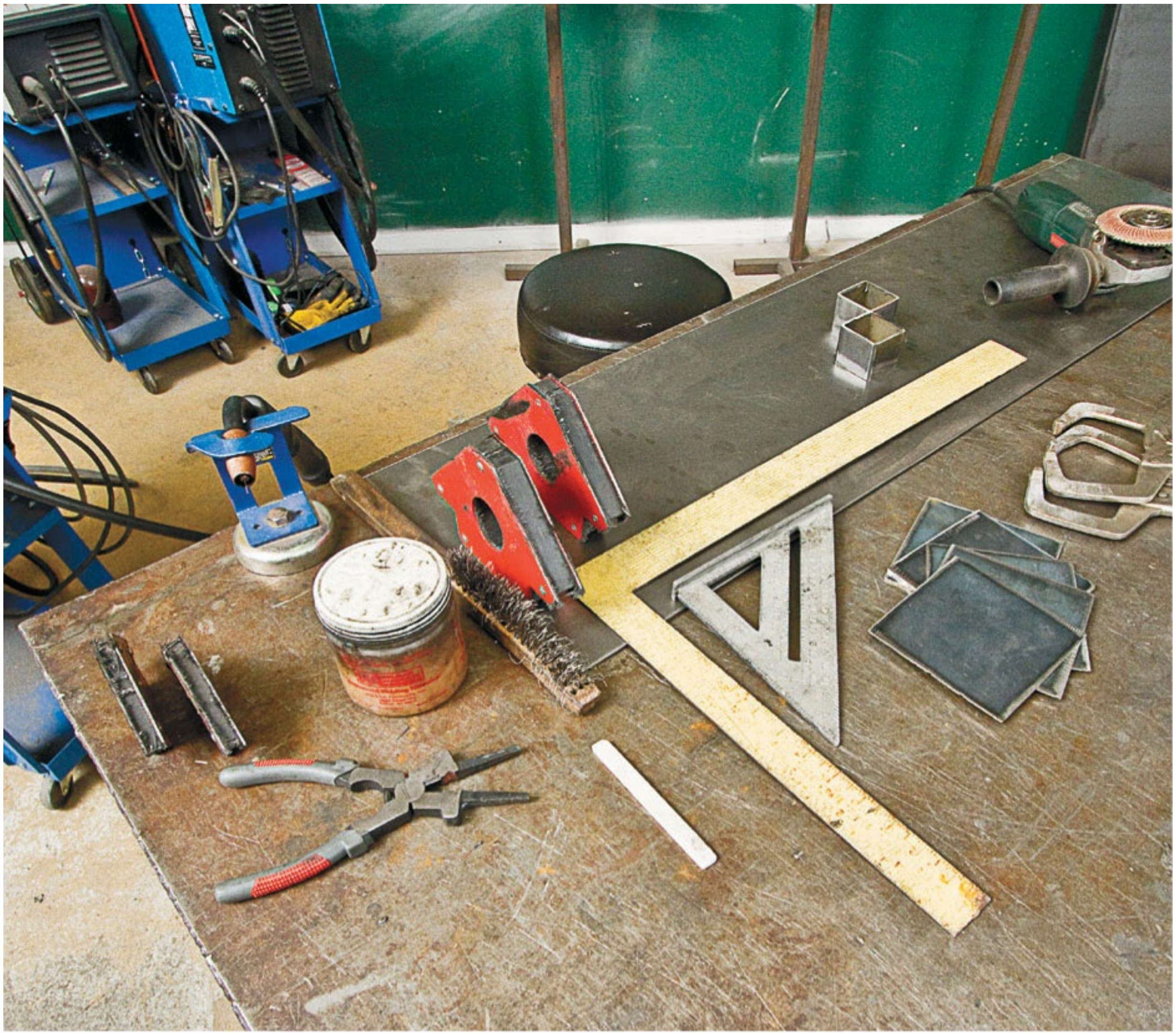

For any project, make sure everything you need is ready to go and close by.

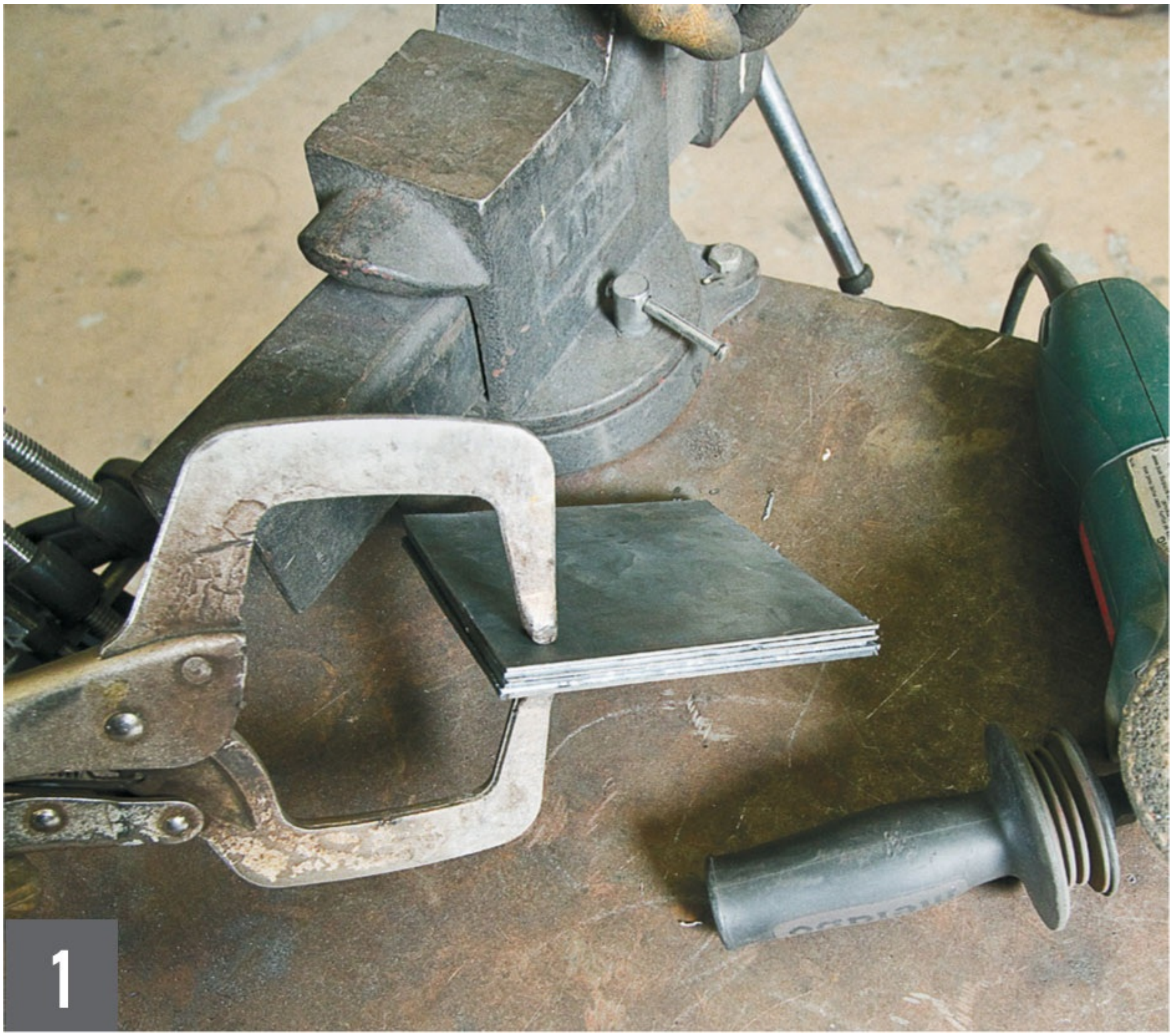

1

Your squares will not make a good cube if they are not all the same size, even, and square. Clamp them together so they are even on two edges.

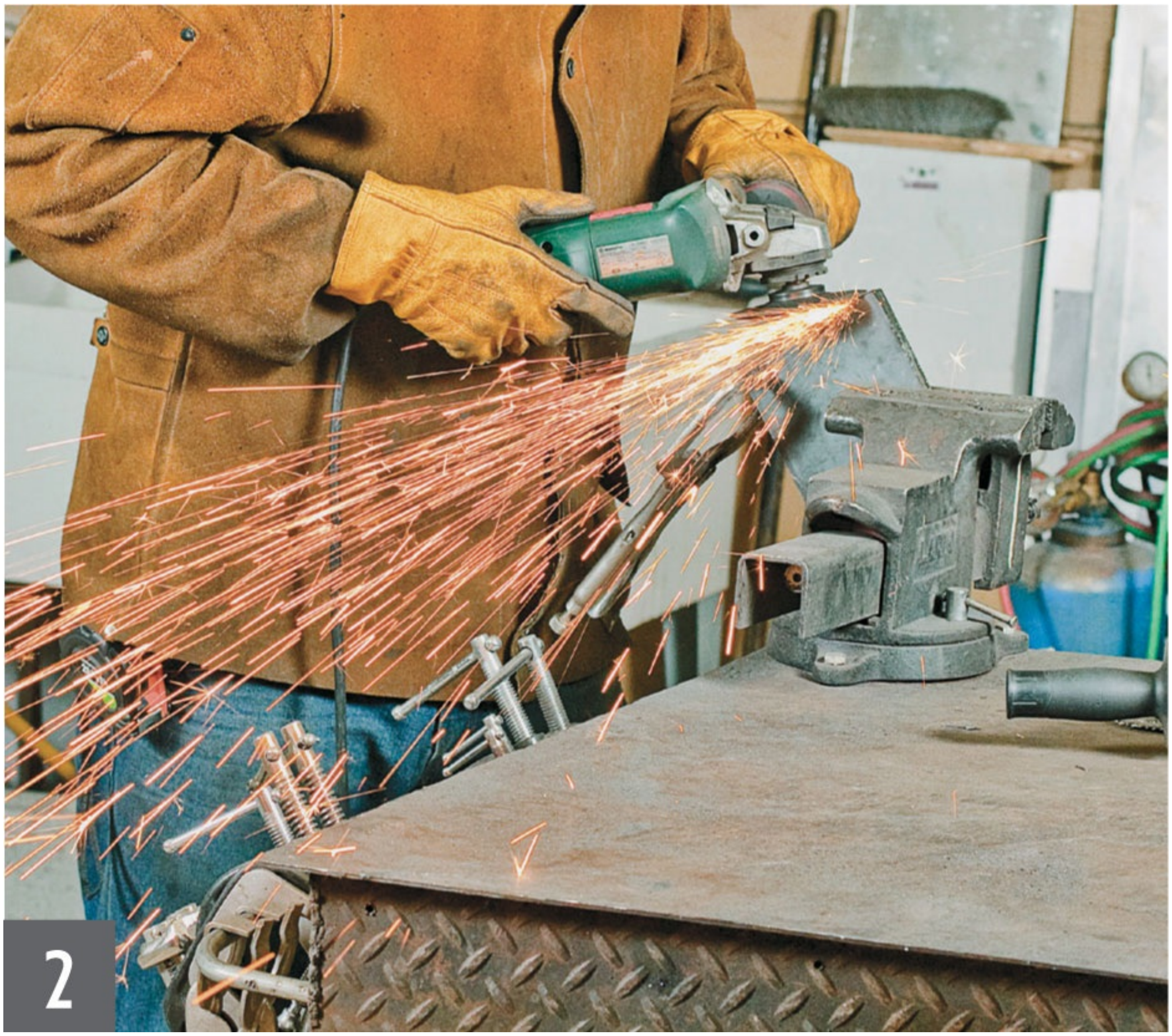
2

Grind down the two uneven sides until they are even.

Check the length of the sides as you go to make sure they stay square.

Use a grinding stone on an angle grinder to even out the other two edges. You can also use a bench grinder, just be careful not to round off the edges and leave some edges shorter than the others.

After grinding each of the two sides, make sure the pieces are still the correct length and are still square.

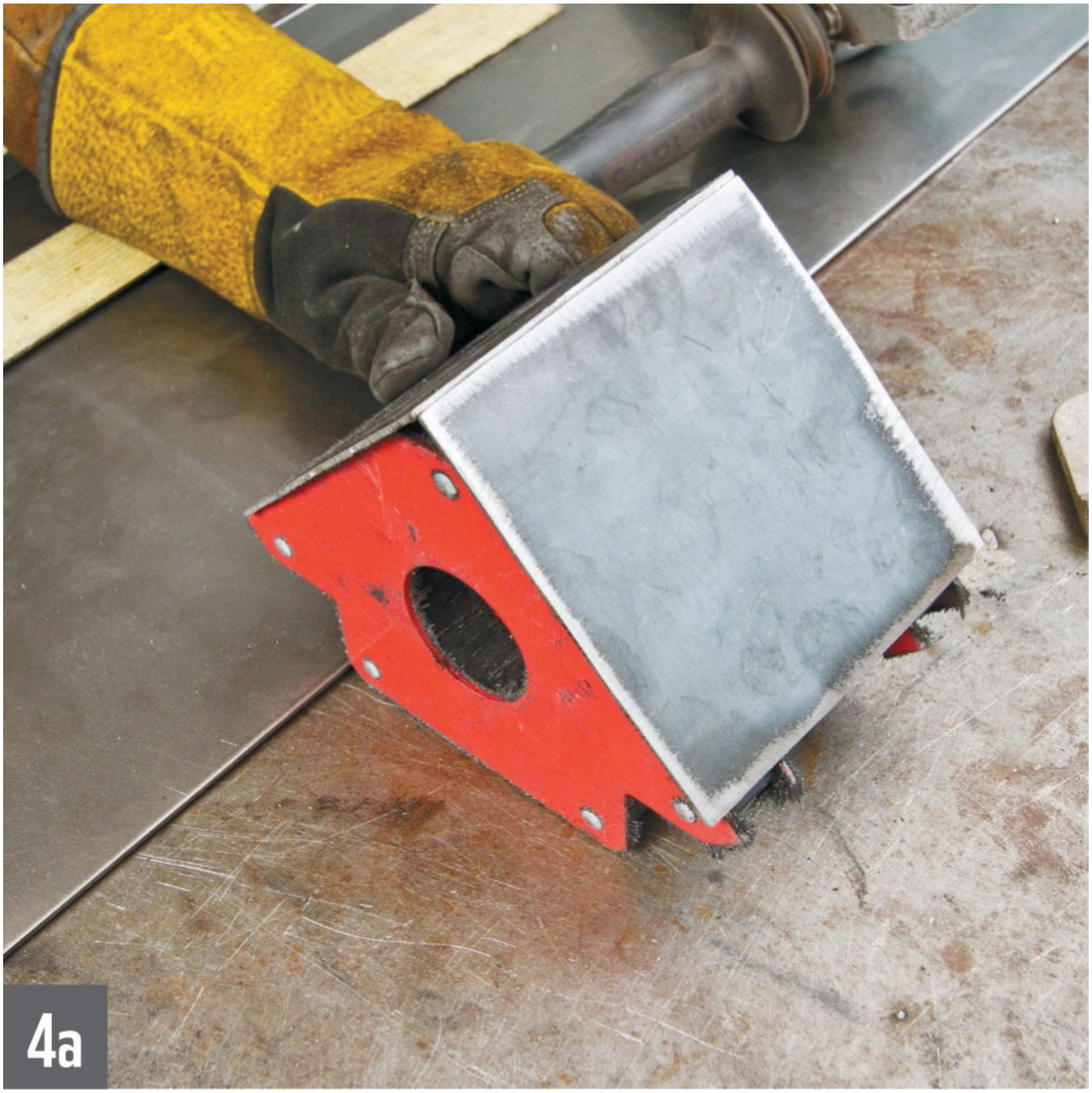
4a

After you have cut and prepped the squares so they're even, it's time to tack. Square magnets can be used to hold the sides in place.

Prep

Now that all six sides are perfect 6-inch (15.2 cm) squares, the next step is to clean and prep the metal. With an angle grinder and flap disk, make sure that all the surfaces around the weld zone are raw steel.

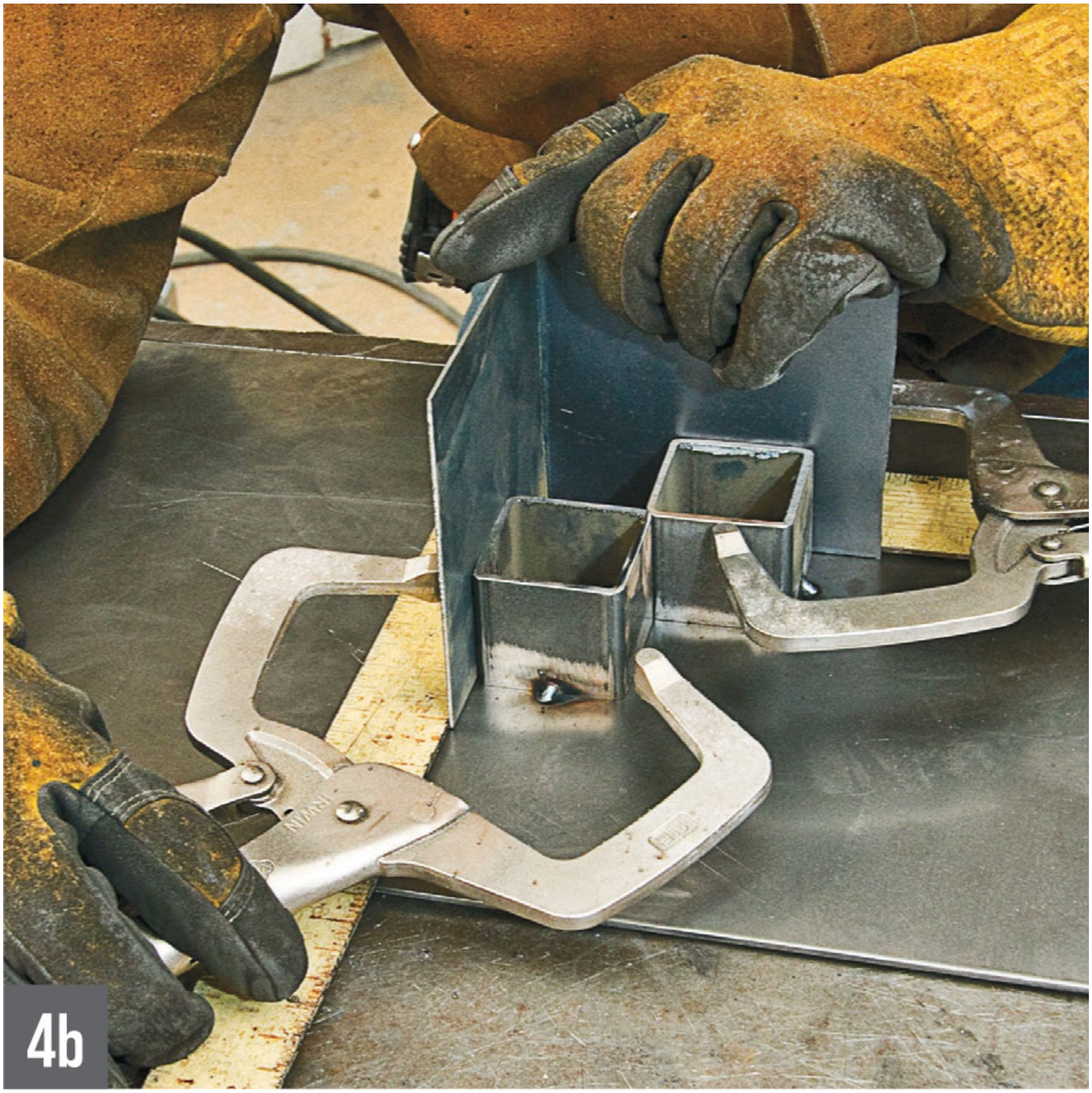

Alternatively, a jig can be used to clamp each side in place while you tack.

Tack

It’s time to tack our pieces together. For this we will need to clamp our pieces together at a perfect 90-degree angle. We could use magnets similar to the method used in Chapter 5 for a corner joint. Instead, use the jig you constructed in Chapter 5. The jig will line up not only one edge, but three of the axes needed to make the cube.

Use two locking C-clamps to secure two sides of the cube onto the jig, one on each side.

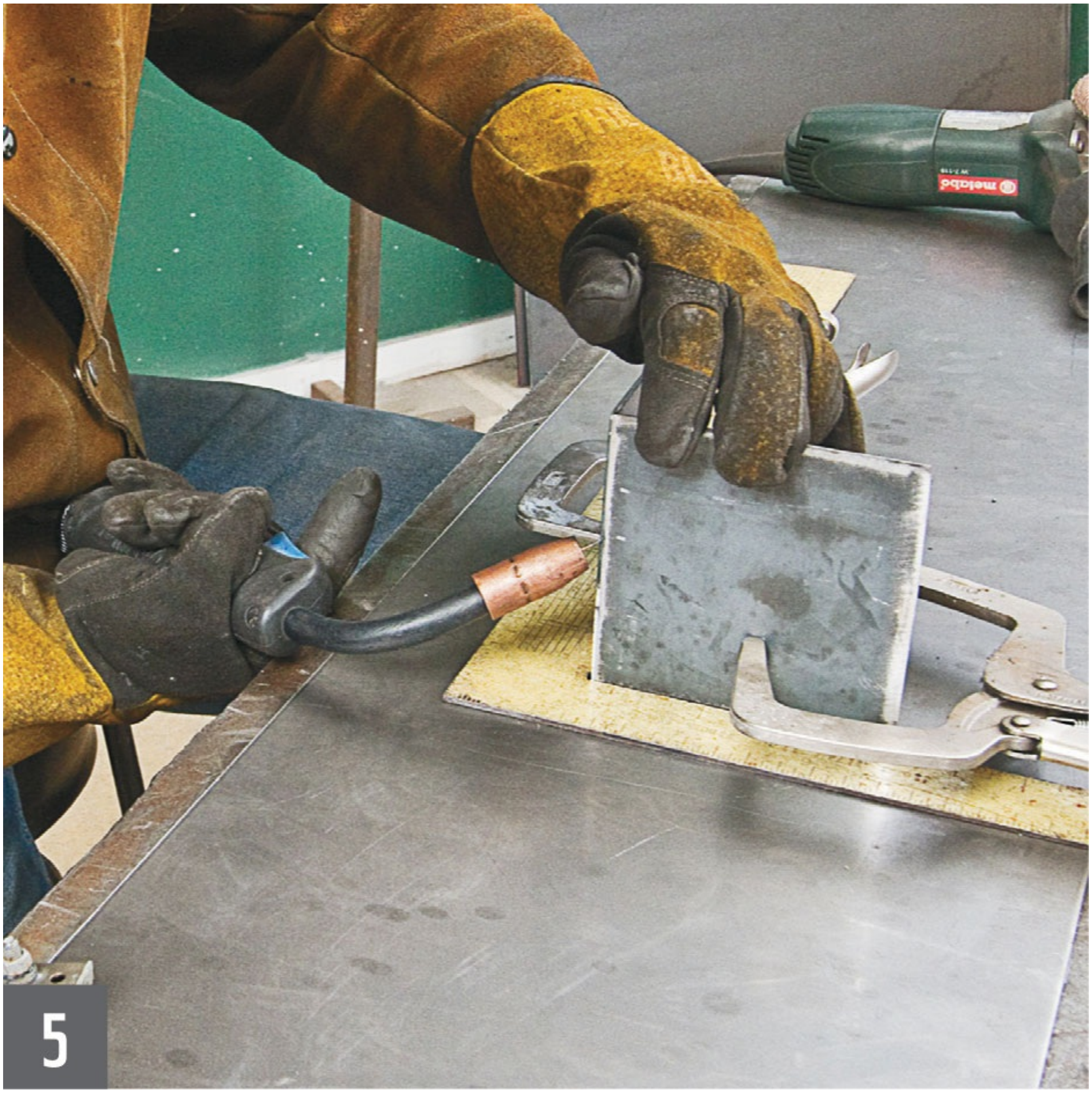

Put one tack in the middle of the two squares.

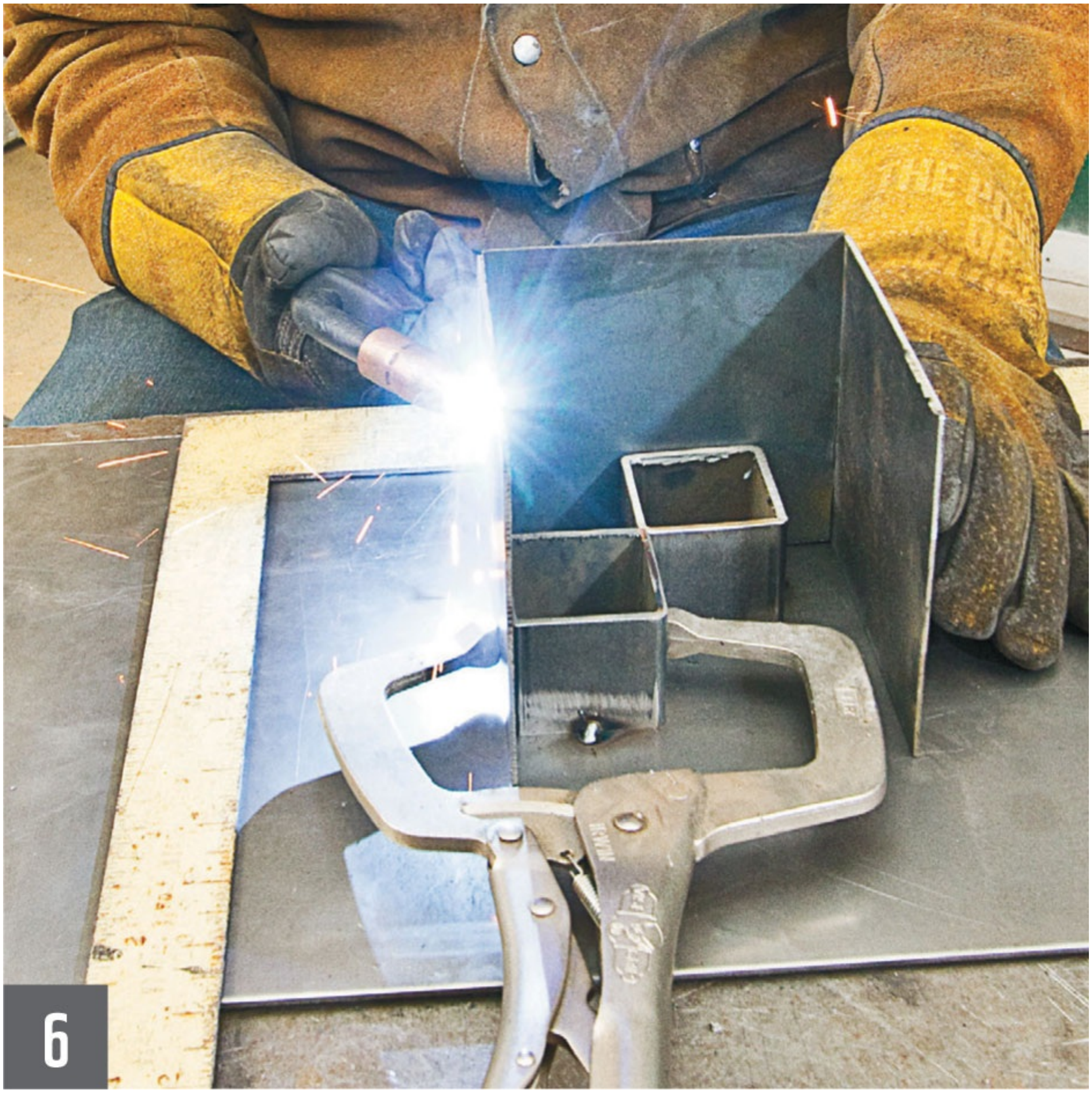

6

Once one edge is completed, add another side the same way. Clamp and tack in place.

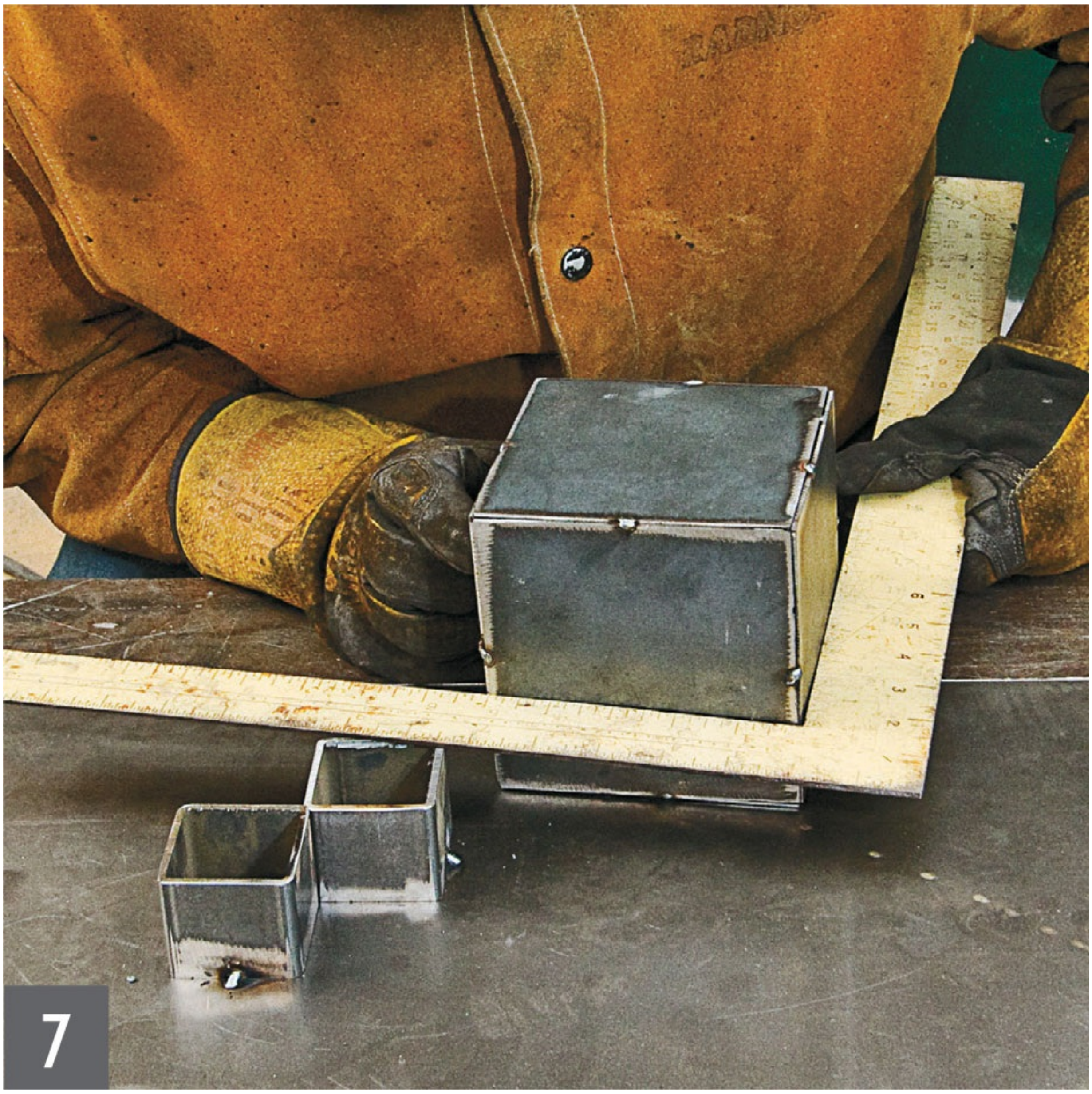

Once the cube is tacked, check to be sure everything is square. Small errors can be fixed with a hammer, but if it's off by a large margin you'll need to cut the tacks and start over.

After the edges of the two sides are lined up and everything is square, tack the two sides together with one tack right in the center.

Repeat the same steps as you rotate the box and add the rest of the sides.

If all has gone well, you should be able to fit the last two sides on the top and bottom. Tack each of them into place with four tacks, one for each edge right in the center. Once you have one tack in the middle of each edge of the cube, square it up.

Once the cube is tacked together, make adjustments as needed using a mallet. When it is completely square and in the correct alignment, tack the corners.

If the cube is not square, it may not be necessary to cut the tacks and redo the entire piece. If it appears to be out of alignment, use a hammer to tap it back into square. This method of tacking will still allow for a little movement; the tacks will bend enough for it to be realigned with a hammer. When the cube is aligned properly, secure it for welding with more tacks at all of the corners.

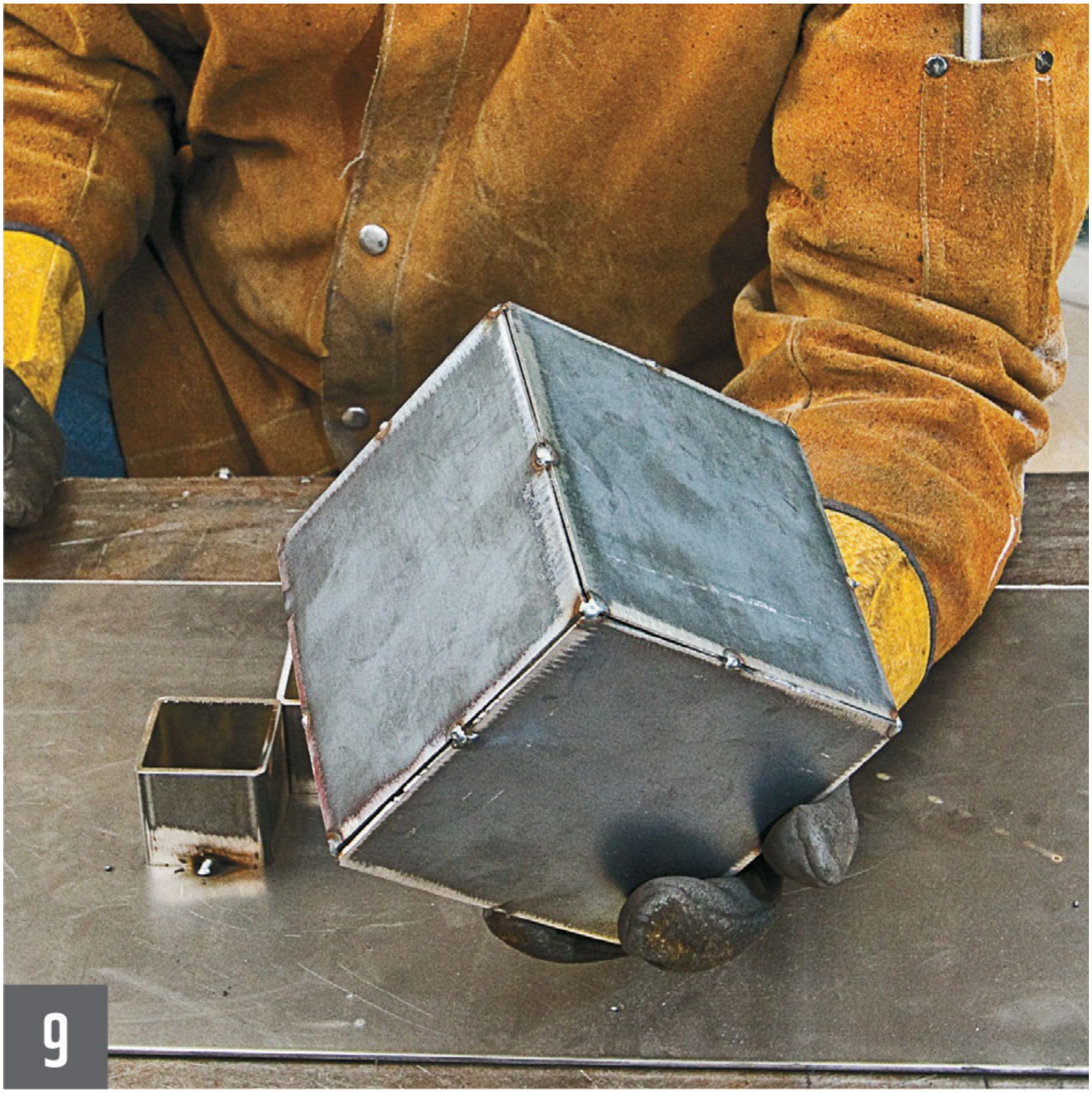

The cube should look like this when it is tacked together.

Weld

We have cut, prepped, and tacked, and now it’s time to weld. While you don’t necessarily have to use a specific technique when welding on steel this thin, on thinner materials it is best to do a push weld with an increased travel angle (drag angle but pushing), about 10 to 15 degrees with the nozzle. All of the other rules are the same. The nozzle will still need to be straight on and perpendicular, and you always want to maintain even angles on both sides of the nozzle.

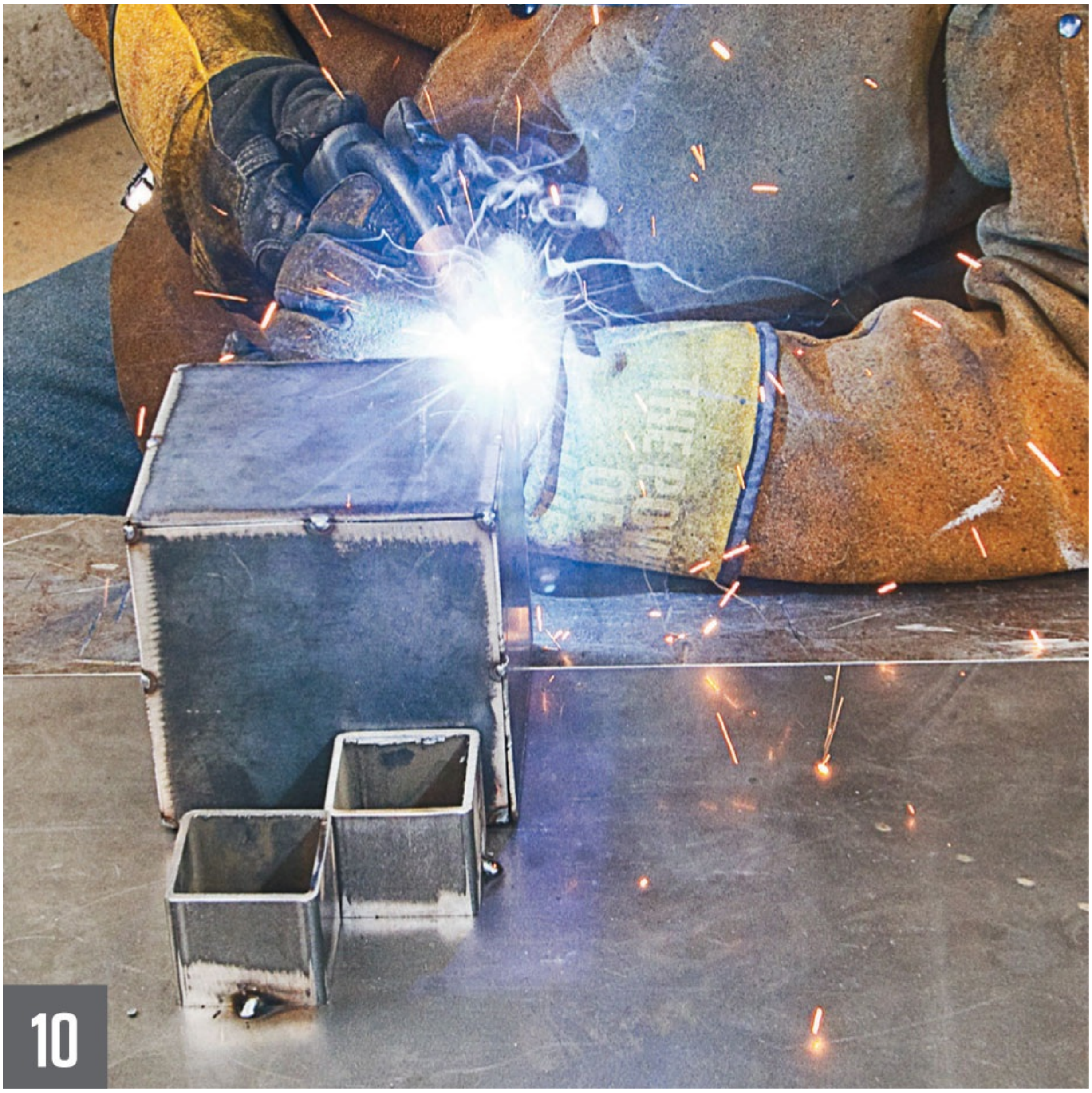

10

You will only be able to weld about a ^{1}/2- to ^{3}/4-inch (13 to 19 mm) at a time with a straight pull or push on steel this thin.

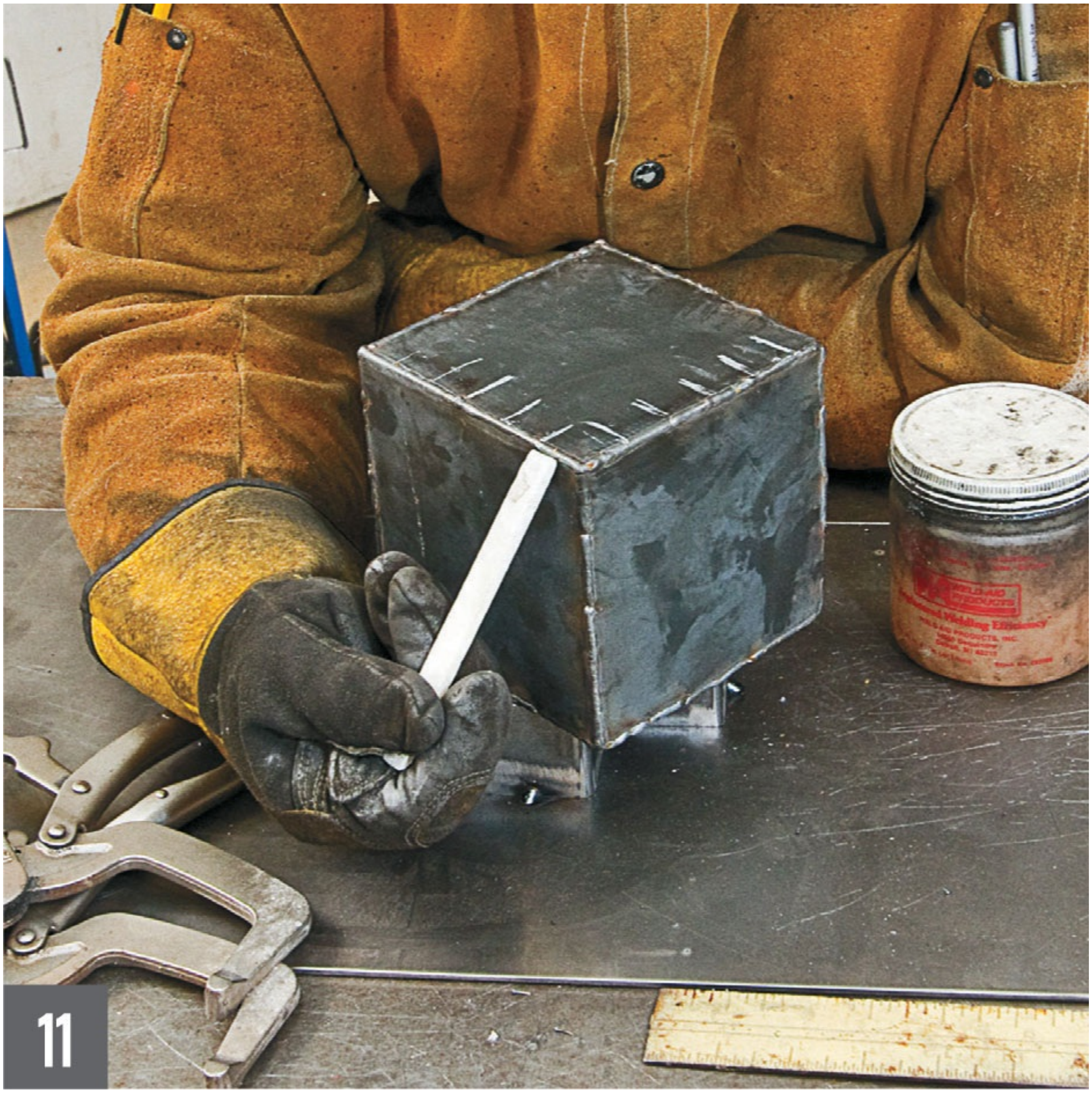
11

Mark sections on the cube in soapstone to plan your welds.

Because the metal is so thin, it will burn though and warp much faster than you may expect. You may have to adjust the setting on the MIG for a little less heat and filler. This means that while you're welding, you will need to do short welds in less time to really spread out the heat.

Use a soapstone to mark 1/2- to 3/4-inch (13 to 19 mm) sections. These marks represent the length of each of your weld zones. Start with one corner section and weld a 1/2-inch (13 mm) length; then move diagonally across to the opposite side for the next weld. Repeat this so that the heat is spread through the entire piece. After one series of welds, it may be necessary to let the cube sit for a minute to cool down and prevent warping.

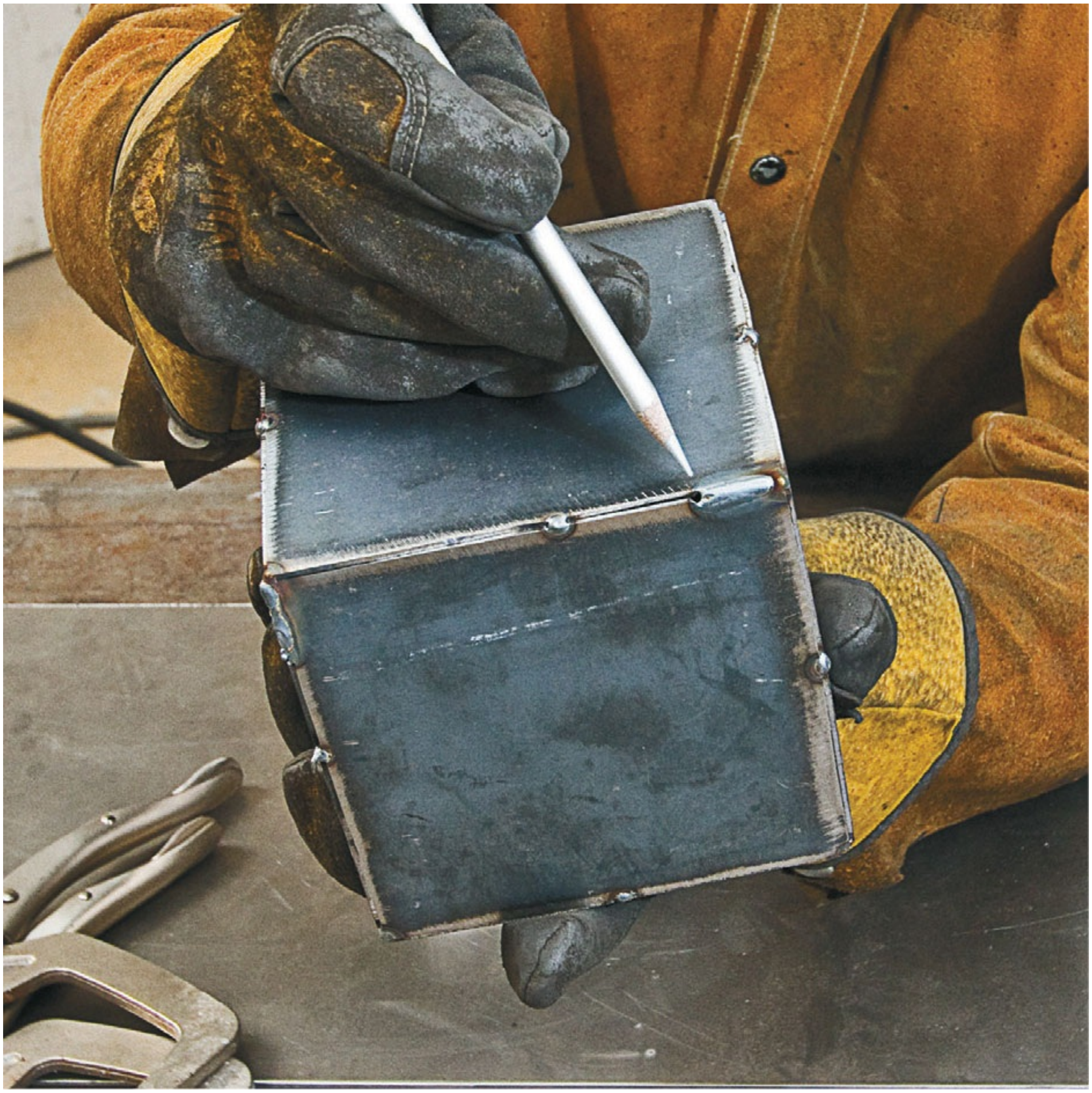

This is an example of key holing.

While welding the cube together, you may notice a slight burnthough start to form ahead of the weld. This is referred to as “key holing” and is a good indication that you should stop welding. This overheating can also lead to the piece warping or further burnthough. If it happens, quickly speed up the motion or dial down the heat and filler. It is perfectly fine if it happens right at the end of a section.

When the entire cube has been completely welded up, you’re done. The only thing left to do is a little stress test. Find an area of concrete pavement like a thing left to do is a little stress test. Find an area of concrete pavement like a sidewalk or schoolyard away from people or anything breakable. Throw the cube 10 feet (3 meters) in the air, stand back, and let it come down on the pavement. Pick it up and inspect the welds. If you don't see any cracks or fractures, good job! You have successfully fabricated a cube. If there are cracks, then do it again; at least it will be fun.

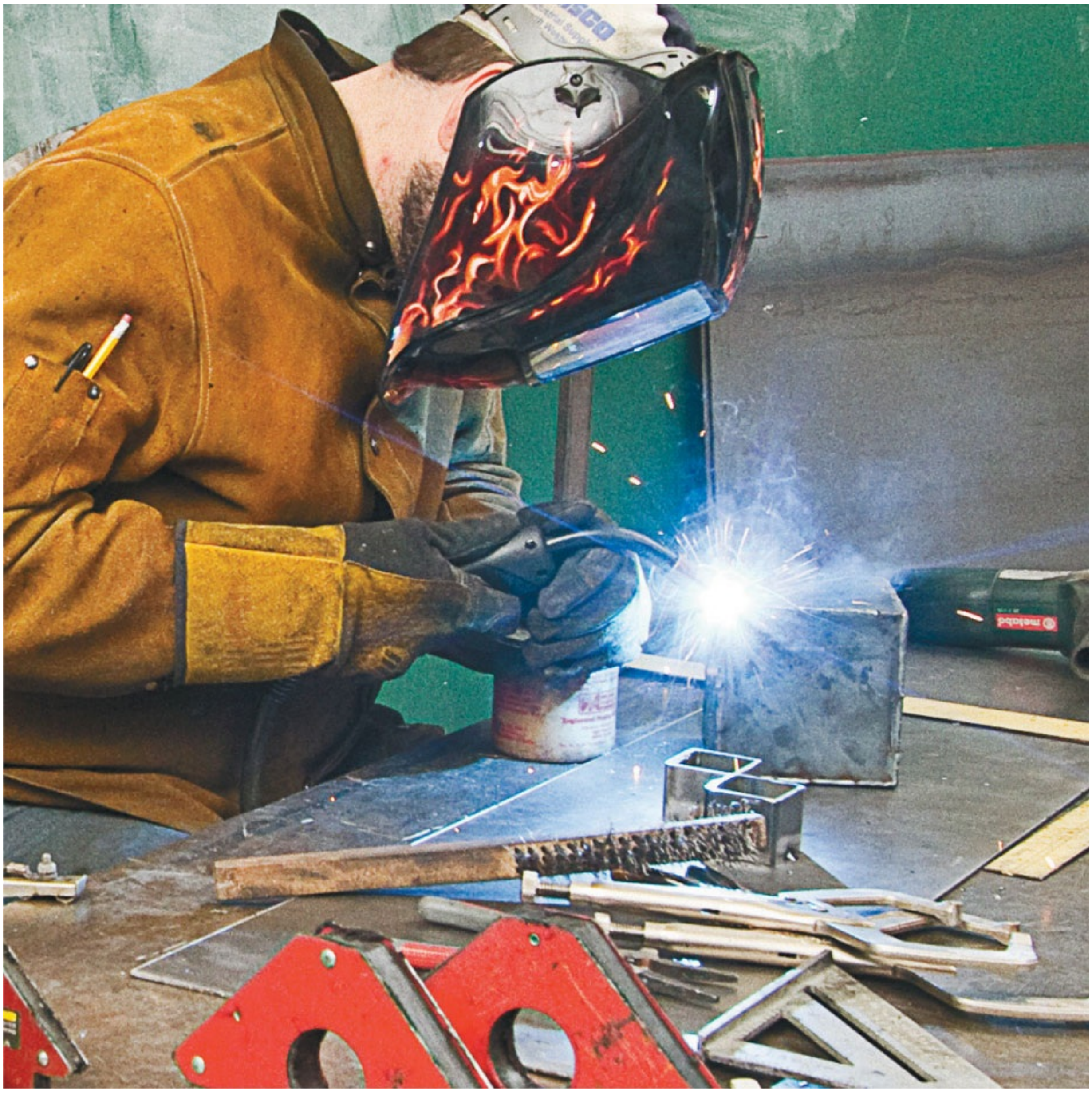

When fabricating the cube, remember to support your gun hand with your opposite hand. As you can see from this image, I'm using my jar of spatter dip to perch my hands on during welding.

There you have it—a perfect cube. Now cut it up and make something out of it.

Project 2: The Welding Table

SKILL LEVEL: Intermediate

MATERIALS AND EQUPMENT:

- MIG Welder
- Two 24’ (7.3 m)-long sticks of 1/8" (3 mm)-thick wall, hot roll 2" (5 cm)-square tube
- 24’ (7.3 m)-long stick of 14 ga, hot roll 1" (2.5 cm)-square tube
- 4’ (1.2 m)-square, 1/4" (6 mm)-thick cold or hot roll plate. (Note: This is for a very large welding table. Make adjustments to the measurements as needed to fit your space.)
- Welding helmet
- Welding gloves
- Long-sleeve shirt
- Leather welding jacket or apron
- MIG pliers
- Stainless steel wire brush
- Soapstone
- 4 1/2" (11.4 cm) angle grinder outfitted with an aluminum oxide grinding disk or a 60-to 80-grit flap disk
- Chop saw
- 3-Axis clamp, right-angle magnets, or jig
- Saw horses

As a welder, you will do a lot of projects, small and large, that require having a good welding surface. Aside from the welding machine itself, you will find that the welding table is one of the most important pieces of equipment in the shop. There are many different manufactures of welding tables that provide a wide variety of different options for accessories, sizes, and clamping systems. But starting off in welding by fabricating your own welding table is a great project. Creating a table that is true, square, level, and plum will give you a greater understanding of what to expect when building future projects such as furniture. It will also provide you with a great new tool to help build those future projects.

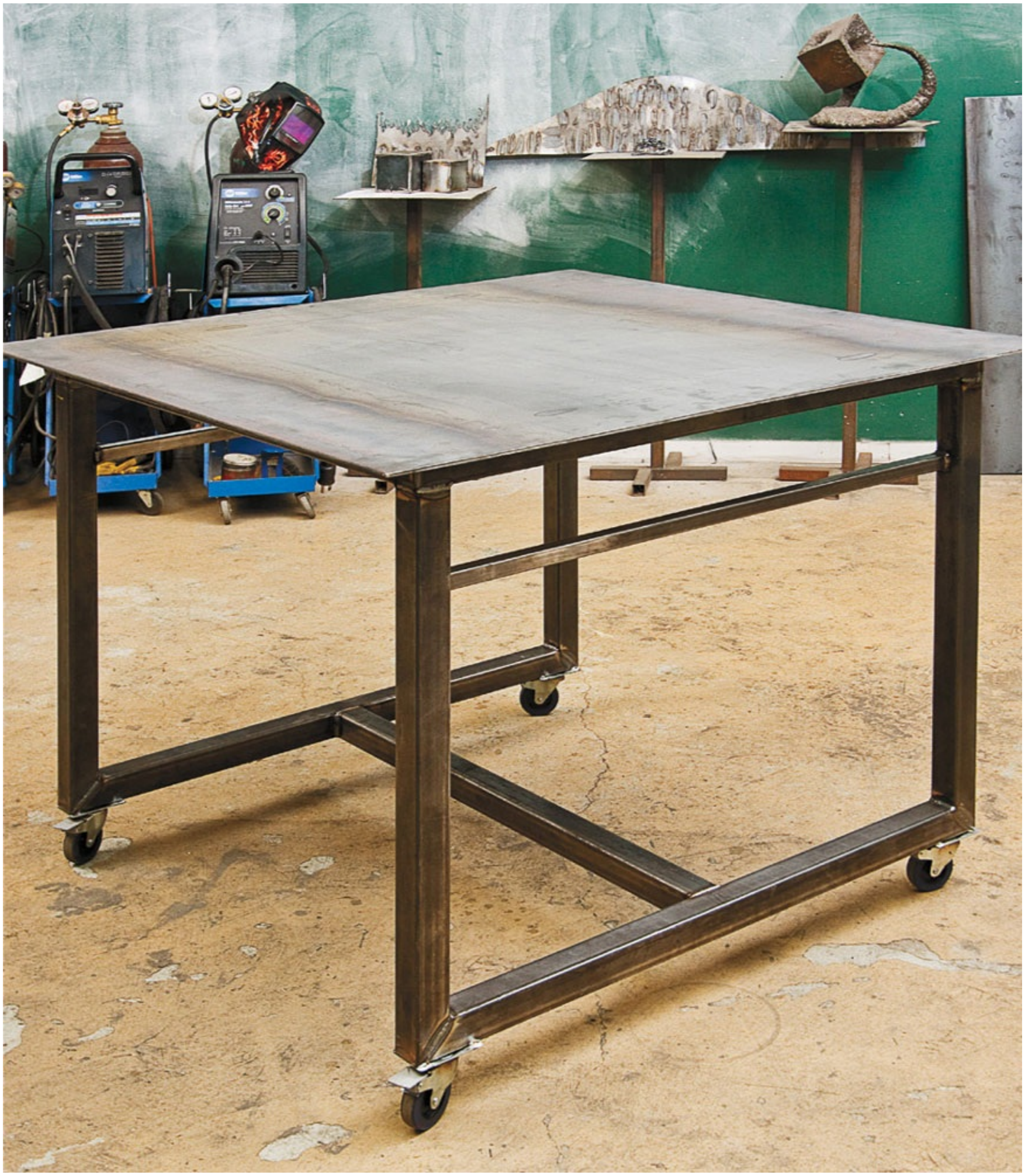

There are a lot of factors to consider when creating a good welding table. In the future as your abilities and needs grow, you will discover other things that you may want to incorporate into building the next welding table. For now, we are going to be focusing on something practical and simple.

Considerations When Making a Good Welding Table

First and foremost is conductivity. The table has to be made of a material that allows the current to flow through the entire table. If working on something small, such as a jewelry box, you can’t necessarily clamp the ground directly to the piece without it becoming awkward. You can, however, clamp the ground to a metal table that will allow the current to flow through the table and then through the work piece as you weld.

Normally, you want to make your welding table out of the same material you plan to do the most welding with. When welding high volumes of stainless steel or aluminum, it is good to have a stainless steel table to prevent contaminating the weld zone. Both of these materials are much more expensive and harder to work with, and you’re not there yet. Most common welding tables are made from hot or cold roll carbon steel. A mild steel welding table will last a long time and provide a multitude of functions.

Second is size. This will depend on the amount of space you have to work with. If working in a small garage, you obviously don’t want to build a table out of a full 4 x 8 foot (1.2 x 2.4 m) sheet of steel. Choose a size that will be enough to work on and at the same time not overpower the workspace.

This is also a great opportunity to customize the table for your needs and comfort. If you are six and a half feet (2 m) tall, consider making the legs higher so you are comfortable standing and sitting at the table. Predetermine your own height dimension by picking out a stool that is comfortable. Remember rule number one: be comfortable when you are welding. Pick a table height that will allow you to be comfortable both when you’re sitting on your stool and when you’re standing. Measure the height to your preference when you are seated and then find a surface that same height to see if you will also be comfortable standing.

The third thing to consider is mobility. Metalworking is a heavy lifting hobby. You are going to do enough of it without having to move a 300 pound (136 kg) steel table around. Heavy-duty locking swivel casters are great not only for the mobility of your table, but they also turn the table into a cart. This means that as a work piece grows heavier as you work on it, you won’t have to carry it around the shop to access other equipment.

The locking casters work great when you have to lean on the table to get more leverage. You don’t want it rolling away from you mid-weld. When you are done for the day and want to save space for the car or motorcycle, simply roll the table to the side and out of the way. Casters are also something that needs to be considered when determining the height of the legs. Most casters that I use are roughly 4 inches (10.2 cm) tall, so remember that when designing your welding table.

Next is functionality. It is always good to be able to clamp your materials down to keep them still while you weld. Having clamps accessible and close by is always a good thing. Rails made of 1-inch (2.5 cm)-square tube on the side of the table will be useful for hanging and storing clamps. This will make the work go more smoothly as you won’t have to search the work area for clamps. This will also make a great location for your ground to be clamped; some ground clamps don’t open a full 2 inches (5 cm), so having a smaller, out-of-the-way piece of steel to clamp the ground to is preferred.

Another thing to consider when determining size and clamping is the sheet of steel that will make the top of the table. Having a good 2-to 4-inch (5 to 10.2 cm) lip from the outer edge to the table’s top frame will help you clamp future projects in place. As you can imagine, this will greatly affect the size of the frame and the top surface.

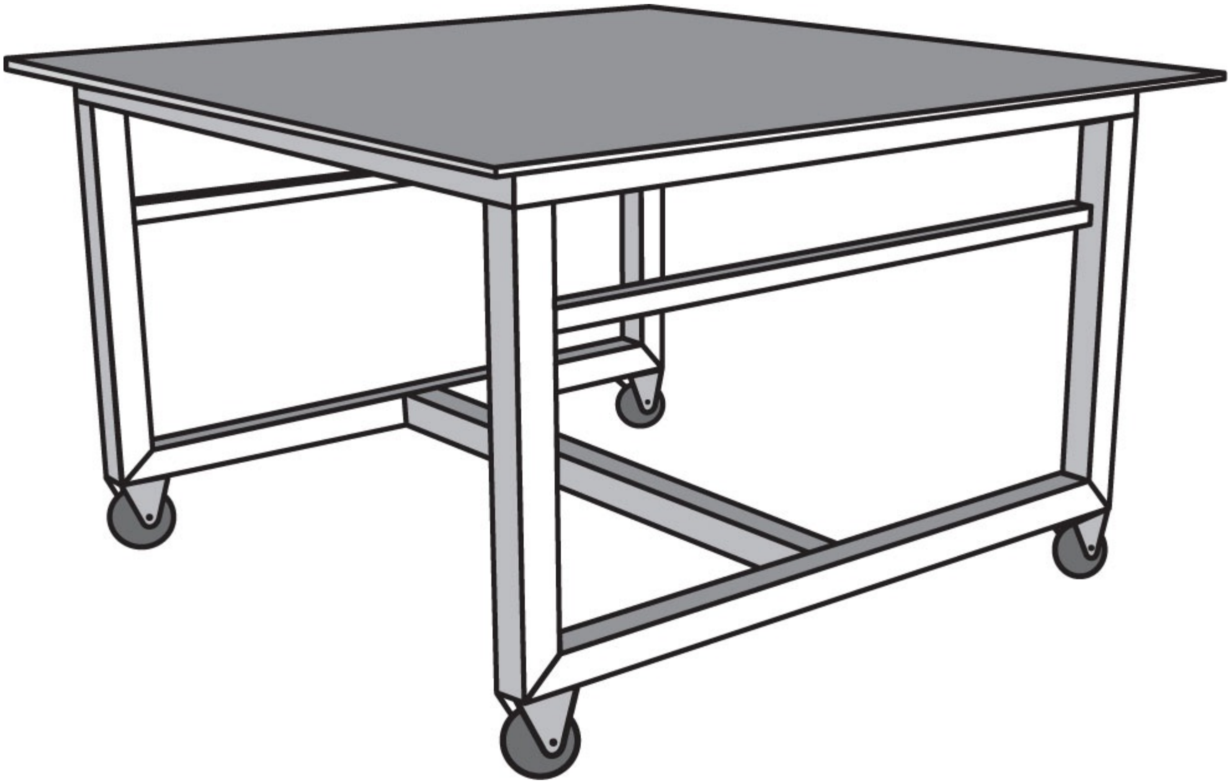

One of the most important pieces of equipment in a shop is the welding table.

Functionality is also important when making decisions on the construction

Functionality is also important when making decisions on the construction. Do you want the table to have a cutting surface? Do you need cutouts on the surface to give you more clamping area? If you want your table to double as a cutting surface, you will want to construct a portion of the top with a mesh or vertical bar stock to one side of the table so the spatter can fall though (see Tools and Equipment in Chapter 2 for methods on cutting metal). If you want more clamping area, you should have slats cut into the surface plate before welding it on. These are all things that you can consider. For now, we will have a 4-inch (10.2 cm) lip to help with clamping.

Even though I provide dimensions for this project, you should concentrate on the construction and welding. Adjust the measurements to the dimensions that fit the needs of you and your workspace. Right away you are going to notice the surface area and height are a lot more than what you might need, so make any adjustments you need. For example, instead of a 4-foot (1.2 m)-square, 1/4-inch (6 mm)-thick plate, try cutting something in between 2 or 3 feet (0.6 or 0.9 m); this may be plenty of surface area for your workspace and the projects you will be working on.

For this project, you need a nice beefy 2-inch (5 cm) square tube with an 11 ga wall. This will ensure that the structure is stable for a wide variety of projects and will be able to support their weight. This table is not going to be painted or a work of art, so to save money we are going to make the entire base from hot roll steel.

The top should be made of a piece of steel that is flat and true and will remain that way for the life of the table; the thicker it is, the better it will hold up against heat and warping. Don’t go overboard though; you don’t need a 2-inch (5 cm)-thick piece of cold roll. The top will be made from a 1/4-inch (6 mm)-thick sheet of 4-foot (1.2 m)-square cold or hot roll steel; this is half the size of a full sheet. If you are making a smaller 2-foot (0.6 m)-square table, you can use a thinner sheet of steel; I wouldn’t recommend going under 14 ga. If you do use a thinner sheet, you will want to reinforce the frame underneath the sheet with some additional square tubes or angle steel.

1

The mantra of any good craftsman is "measure twice, cut once." I normally throw one more in for good measure. (Pun intended.)

Cut

At this point, you should have all of the materials and dimensions you need to fabricate the table. Count out all the pieces based on how many you need for each dimension and how they are supposed to be cut. "Measure twice, cut once" is pretty much the mantra for everyone in the construction or fabrication industry, welding included. We are going to be making miters for the top frame, legs, and base. This will require six identical-length cuts for the top frame and base rails that will all be mitered inward on each end. The legs will only be base rails that will all be mitered inward on each end. The legs will only be mitered on one end to fit to the bottom rails.

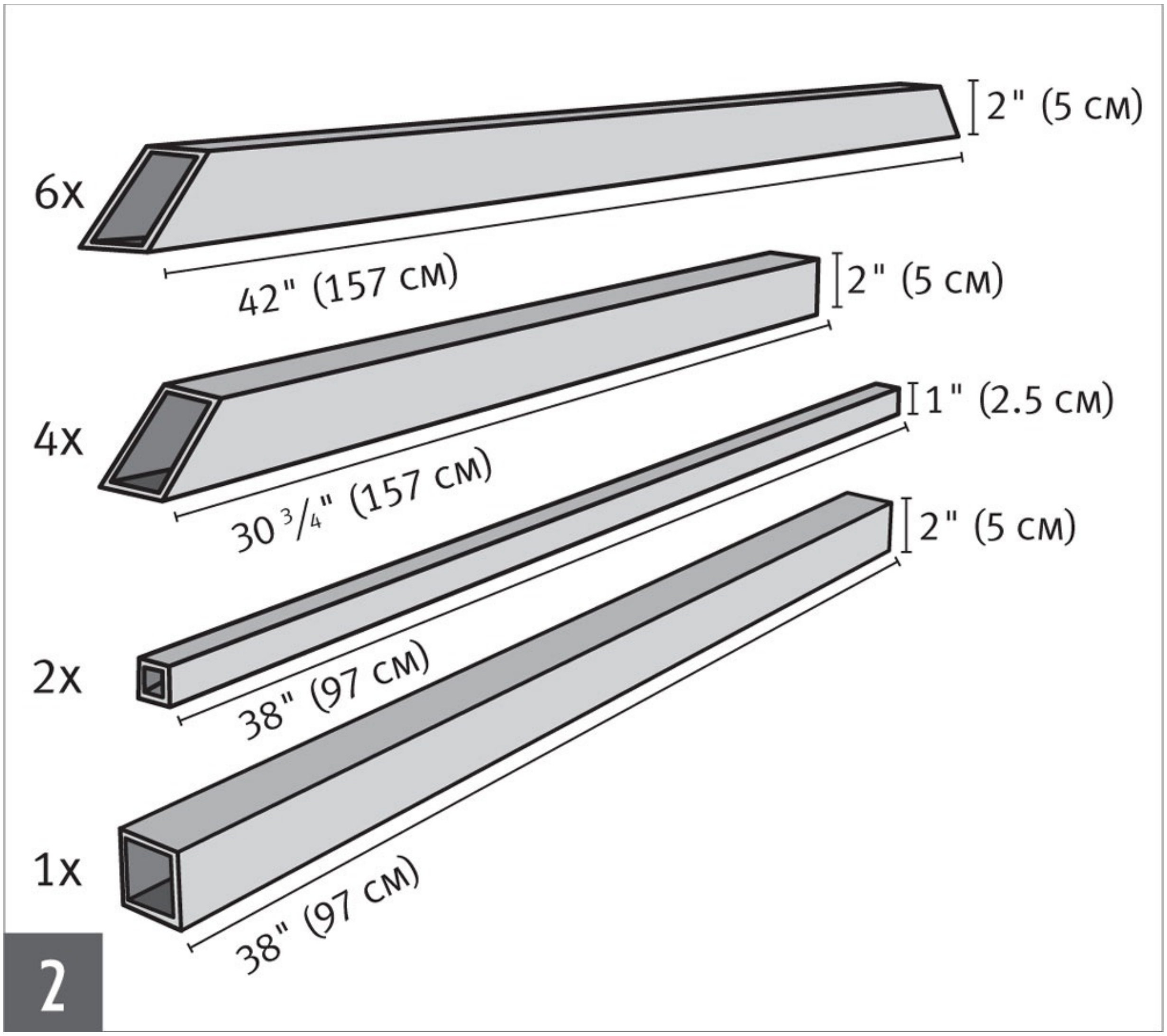

Follow your cut list to make sure you have all the right pieces in the right amounts at the right lengths.

The design and the cuts of the legs and base rails should also be determined by the casters you use. If you are going to be using a threaded stem caster, you will not need to miter the bottoms of the 2-inch (5 cm)-square legs or the tube that connects the legs. Leaving the bottom of the legs open will allow plates that you have drilled and tapped to be welded into place. I will explain that process later in this chapter.

The design I will be fabricating will use four-holed plate casters; this involves removing the coating and welding them on to the bottom of the table. In this case, I will be cutting and connecting the legs with a mitered support rail. What type of caster to use is up to you.

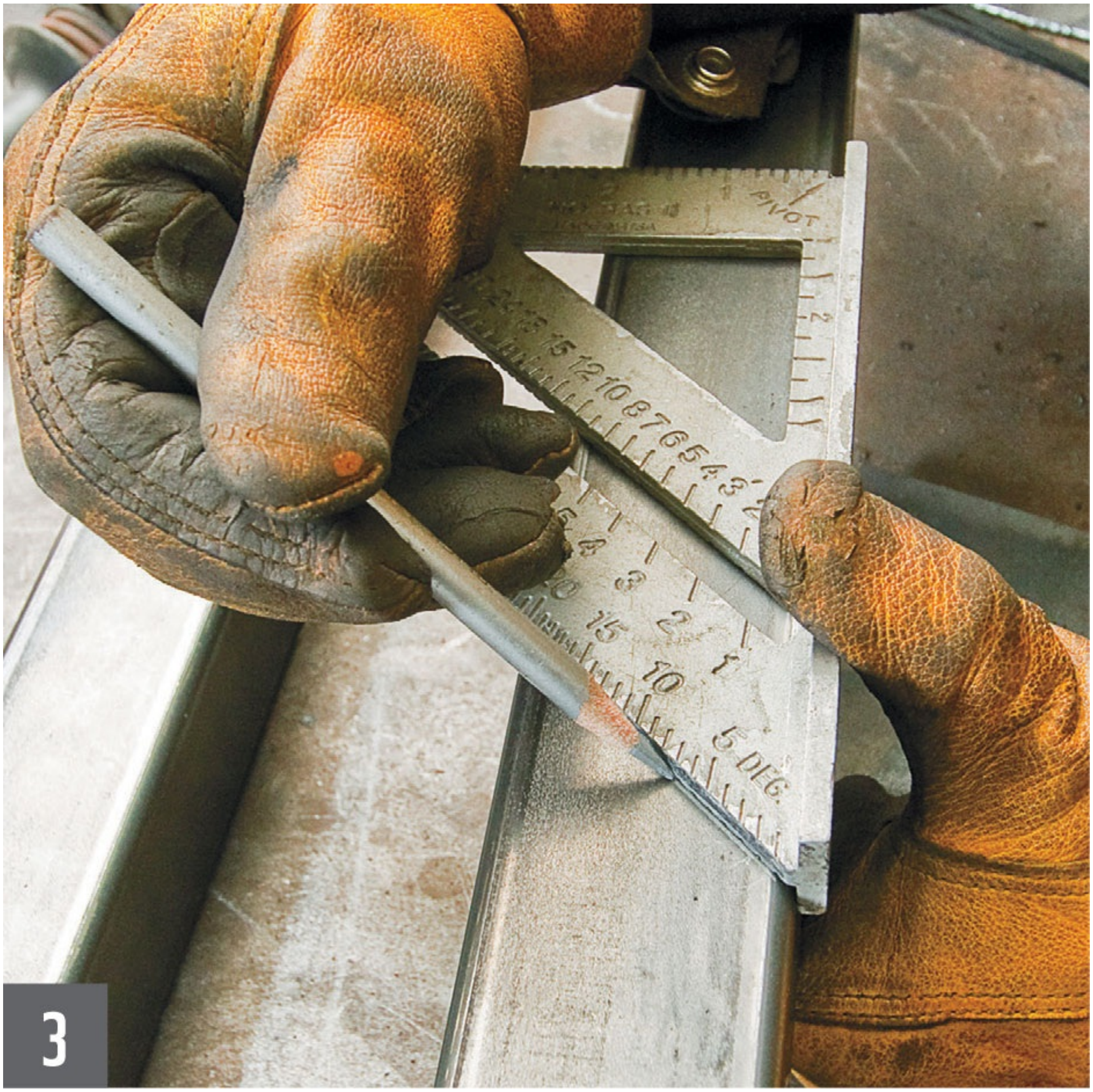

When cutting miters in steel, it's always a good practice to put the entire cut line on the metal before cutting.

Start by cutting all the longest parts that have the same measurements and angles; for example, the top frame and the two runner base pieces are all the same and they are the longest. These pieces need to be identical. There will be six total, all mitered on each end, and all outer dimensions will be 42 inches ([106.7 cm] or the measurement that you have determined).

Using a combination or framing square, mark the miter cuts needed on each end of all six pieces. Adjust the chop saw guide for a 45-degree cut. Cut all miters inward. Remember that the lower frame does not need miters if you will be using stem casters, so adjust the measurements accordingly.

PORTER CABLE
4

Try to be as accurate as possible. Cut the first one and make sure it's the correct length and angle. Then match the rest of them to it to make sure they are all the same.

Try to be as accurate as possible when making these cuts. It is best to be a little on the long side and make a rough cut. Just like any cutting method, the thickness of the blade is a factor. Make sure the blade is on the correct side of the cut line. Go slow and take your time. Put enough pressure on the saw so that it is cutting, but don’t try to bang through it. It is normal to have to even the material out to get each piece the proper measurement. Using an angle grinder with a stone grinding disk is sometimes necessary to get all the lengths and angles to match.

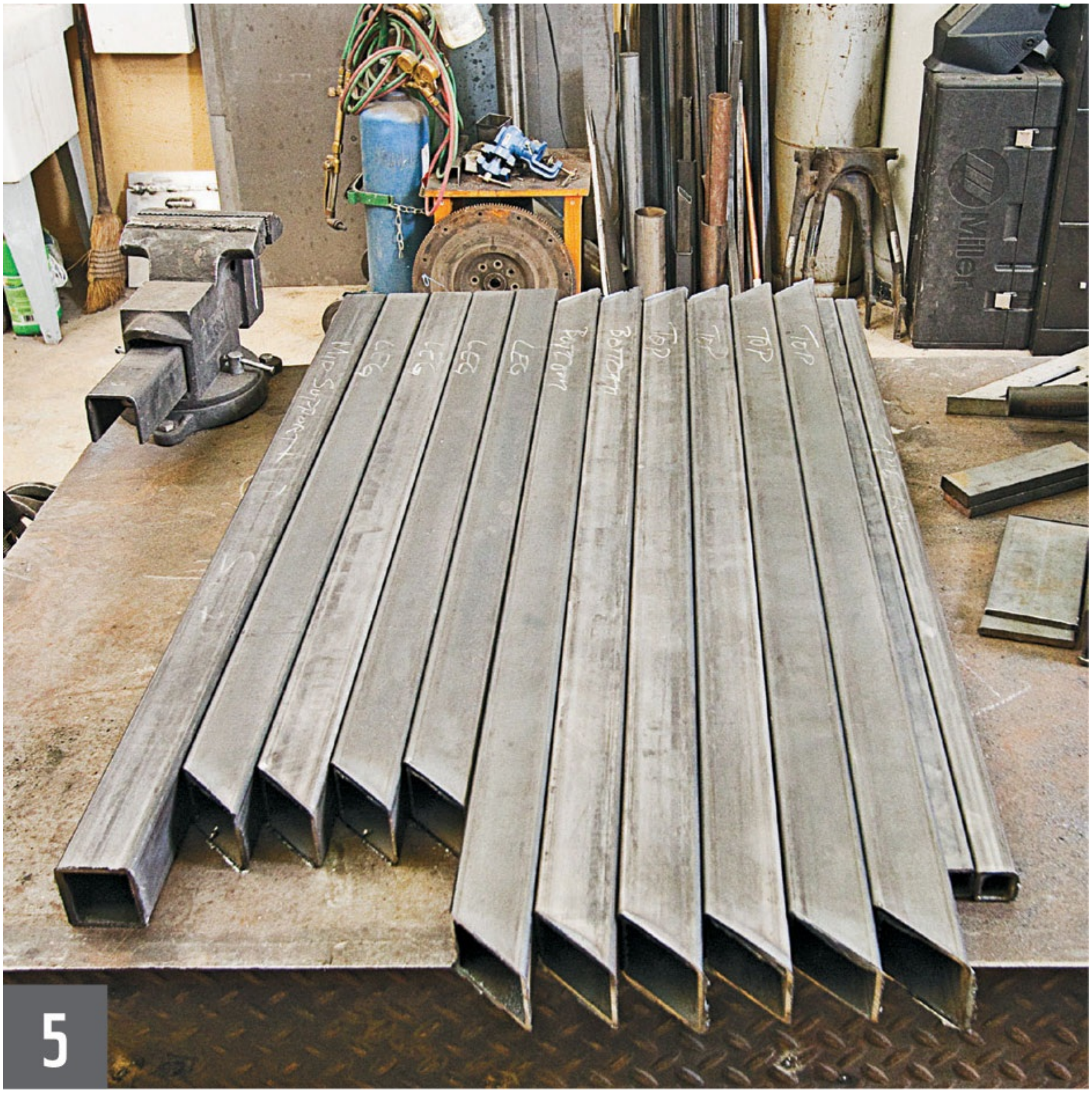
5
MID SUPPORT
LEG
LEG
LEG
BOTTOM
BOTTOM
TOP
TOP
TOP
TOP
Miller

Lay out all the pieces and take inventory. If you have all the parts you need, it's time to true them up and prep.

To help mark and keep consistent outer lengths, line up all of the material side by side and then measure the length you need. Use a carpenter square to mark a line on all of the pieces that are going to be cut the same; this line will be where you mark off your 45-degree cuts.

After all of the cuts have been made, check them for accuracy. If some are a little long or have some burs, grind them to match.

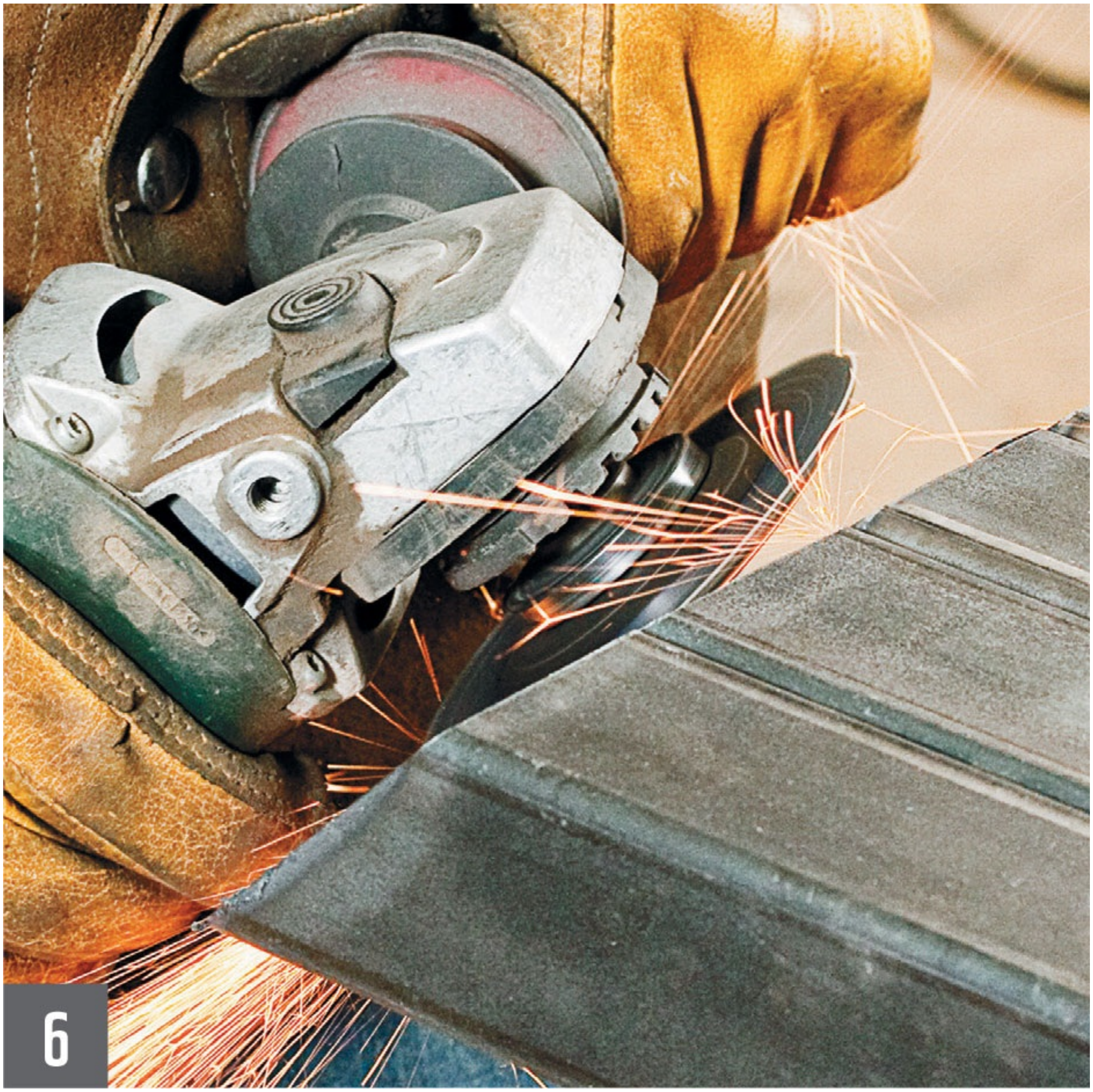

Using the most accurate piece of each series, true up the rest to match that piece.

Using Consumable Blade Chop Saws

When using a metal-cutting chop saw outfitted with a consumable blade, bear in mind that the blades are much more flexible than a wood-cutting blade. When cutting angles, be careful of how much pressure you're exerting when you push down. Too much pressure and the blade is going to want to deflect, causing a curved cut, which results in miters that are difficult to work with.

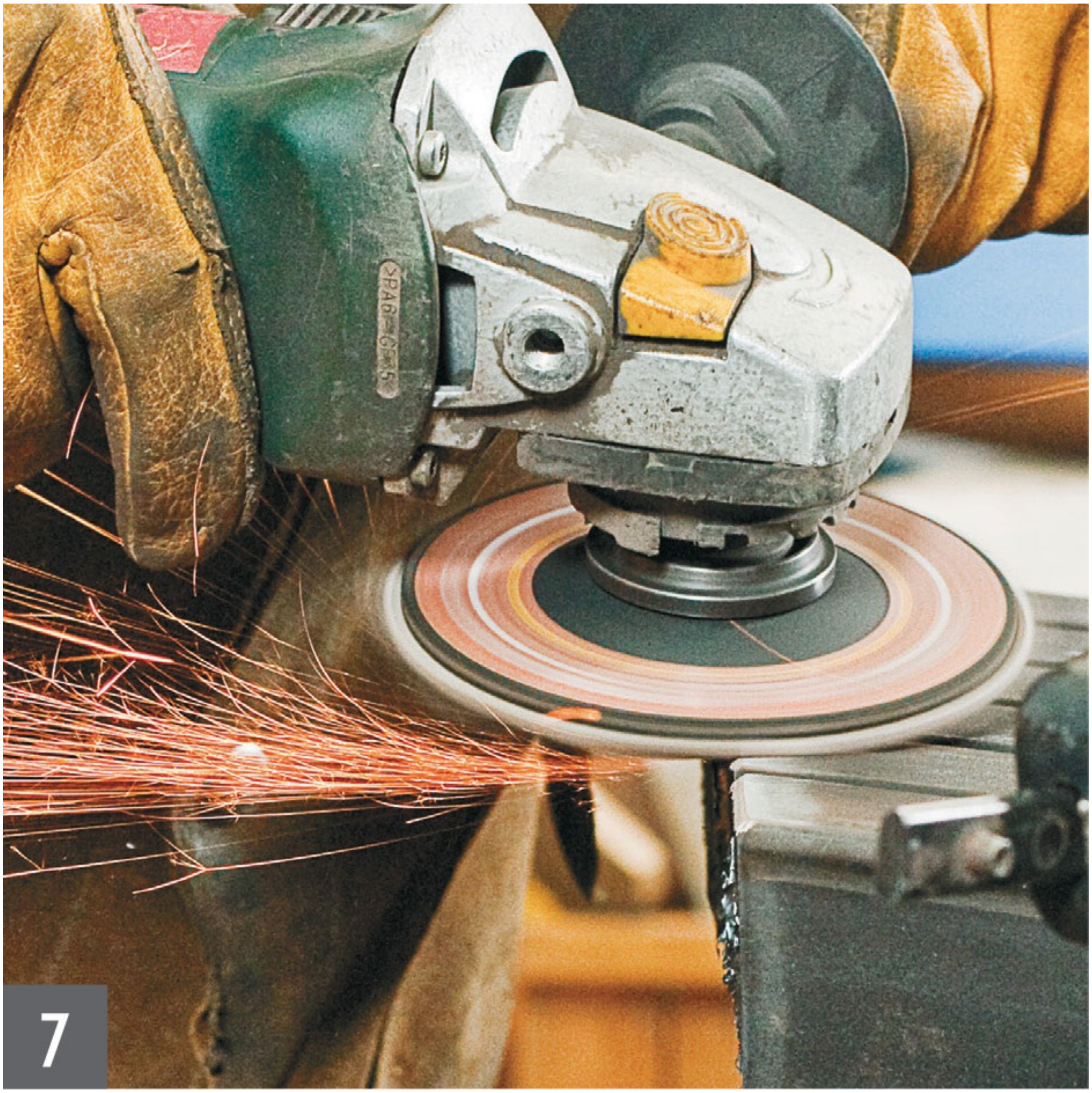

After everything is the same length, use a flap disk to prep all the pieces where they are to be welded. Chamfer the miters on the butt joints.

Prep

Prepare all of the surfaces for welding by cleaning them with denatured alcohol or acetone. With a 60-grit flap disk, remove all mill scale from the weld zones and chamfer the edges if needed.

On material this thin, I only chamfer butt joints slightly—45 degrees—for half the thickness. Then I will clean the weld zone a minimum of 1/4 inch (6 mm) from the chamfered edge. It's not necessary to chamfer the legs that connect with the top frame. The interior welds for the legs will be fillet welds. The rounded edges of the square tubes will provide a sufficient chamfer for the exterior welds.

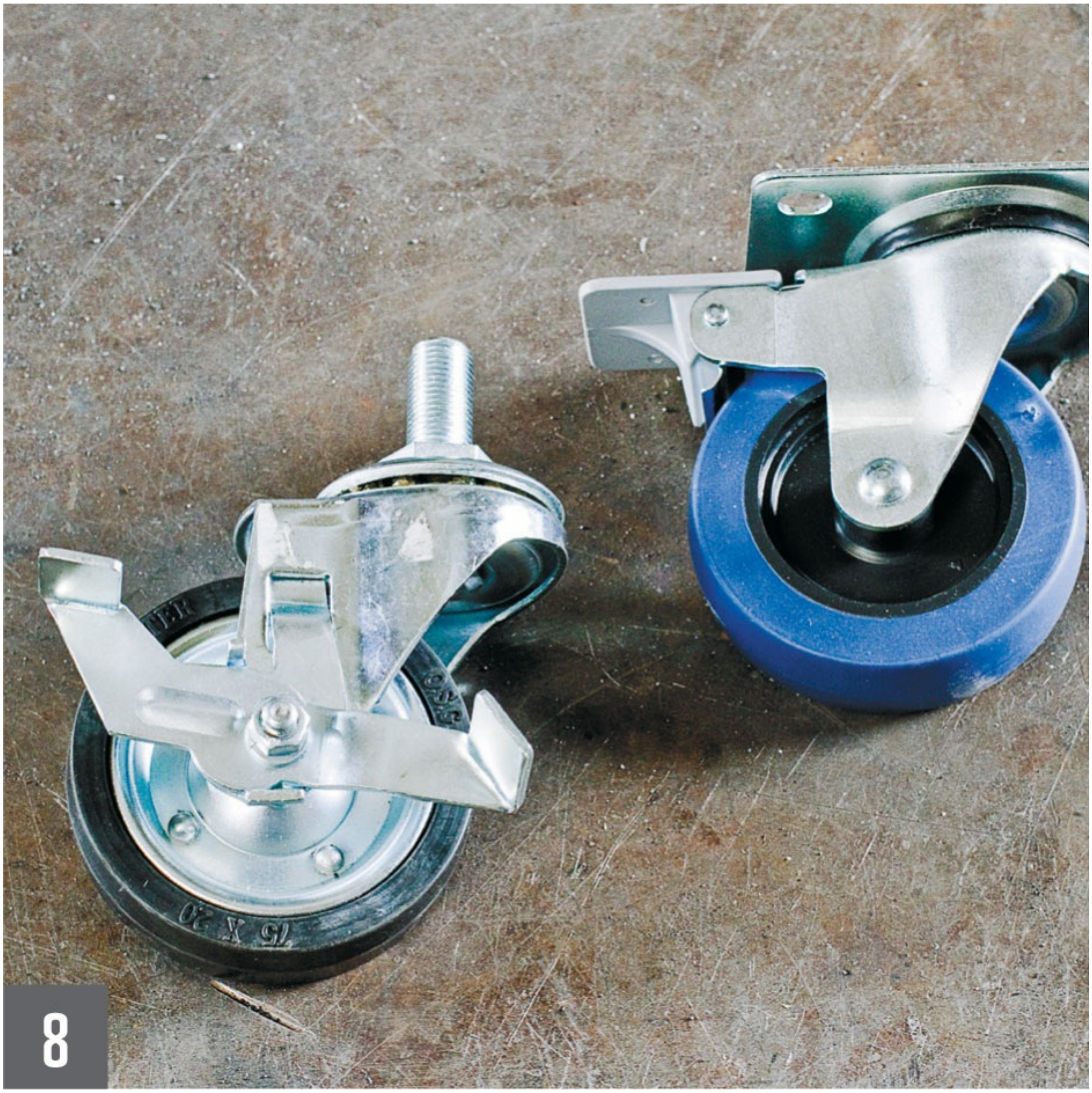
8

Get your casters ready. Here are a single-screw stem caster and a four-hole plate caster.

Mobility

The specific casters you have selected will determine the way they'll be attached to the table. A four-hole plate caster can be welded directly to a plate that is welded to the bottom of the legs or directly to the frame of the table if the frame is wide enough. There are also single screw stem casters that will require a hole is wide enough. There are also single screw stem casters that will require a hole to be drilled into a piece of plate metal to be welded to the bottom of the legs.

The point where the casters meet the legs will take a lot of abuse and a lot of torque. You don’t want to weld a caster to thinner gauge metal, as the metal can buckle under the stress of pushing the table around with something heavy on it. If the materials the table is made from are not thick enough to handle the abuse, weld plates on the bottom first. The thicker metal for the bottom will reinforce the casters. For this project, we should be using 14 ga wall or thicker so we should be okay.

It is common for most casters to be covered in some kind of plating such as zinc. When welding casters to the bottom of the table or to a plate, make sure to prep the plate on the caster thoroughly with a grinder. When grinding and welding these types of plate, always wear a respirator to prevent breathing harmful chemical fumes.

A lot of things can happen to casters. They can lose their ball bearings, bend, or the lever locks can fail. If you have a drill press, you can also consider the single screw stem casters. You can create threaded holes on the bottoms of the legs to screw in this style of caster. It makes it much easier to replace the caster if it breaks or fails.

Drilling and tapping is the method used to create a threaded hole in a plate of steel to receive a threaded screw, or in this case a stem caster. Another method is drilling an oversized hole and welding nuts on the inside of the plate before you weld the plates to the bottom of the legs. This will allow for the caster to be replaced faster when necessary.

WARNING

Most casters have zinc plating that must be removed to expose the raw steel beneath. Wear a respirator when grinding and welding on anything zinc plated.

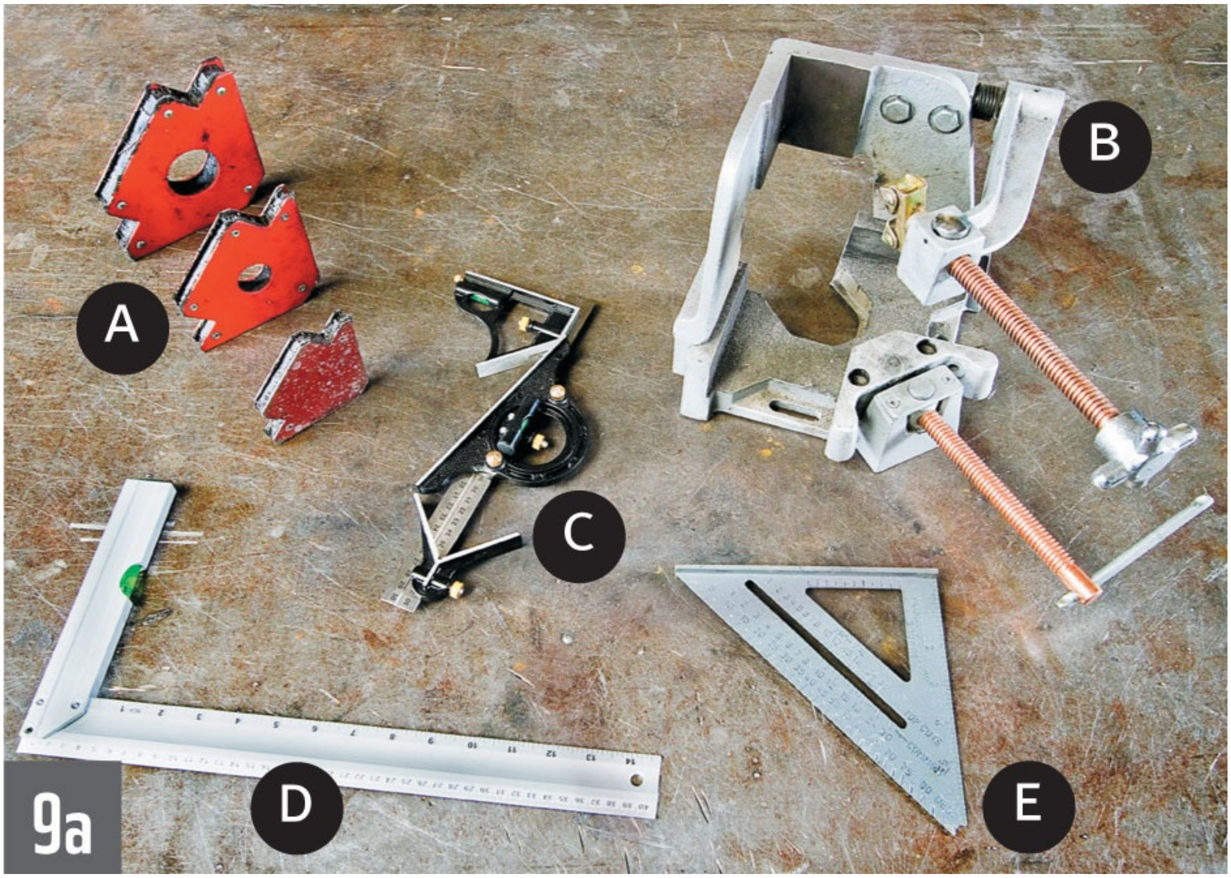

All of these tools have perfect 90 degree angles and help you make great corners. A. Magnetic squares; B. 3-axis corner clamp; C. Precision square; D. Multi-purpose precision square; and E. Rafter square

Four Ways to Create Good Corners

There are several ways to get true 90-degree corners. The first method is a corner clamp. For the most accurate corners, a three-axis corner clamp is the best. These can be expensive, so most hobbyists won't invest in this piece of equipment. However, if you plan on becoming a furniture maker or doing a lot of framing projects, this clamp is well worth the price. The three-axis corner clamp allows you to tack and weld two pieces of the frame and a leg together all at once, which saves a lot of setup time. These clamps are known for their accuracy, so you will also know that your corners are square. It's not a bad idea to check its accuracy with another square once in a while, however, to make sure it doesn't need to be tuned.

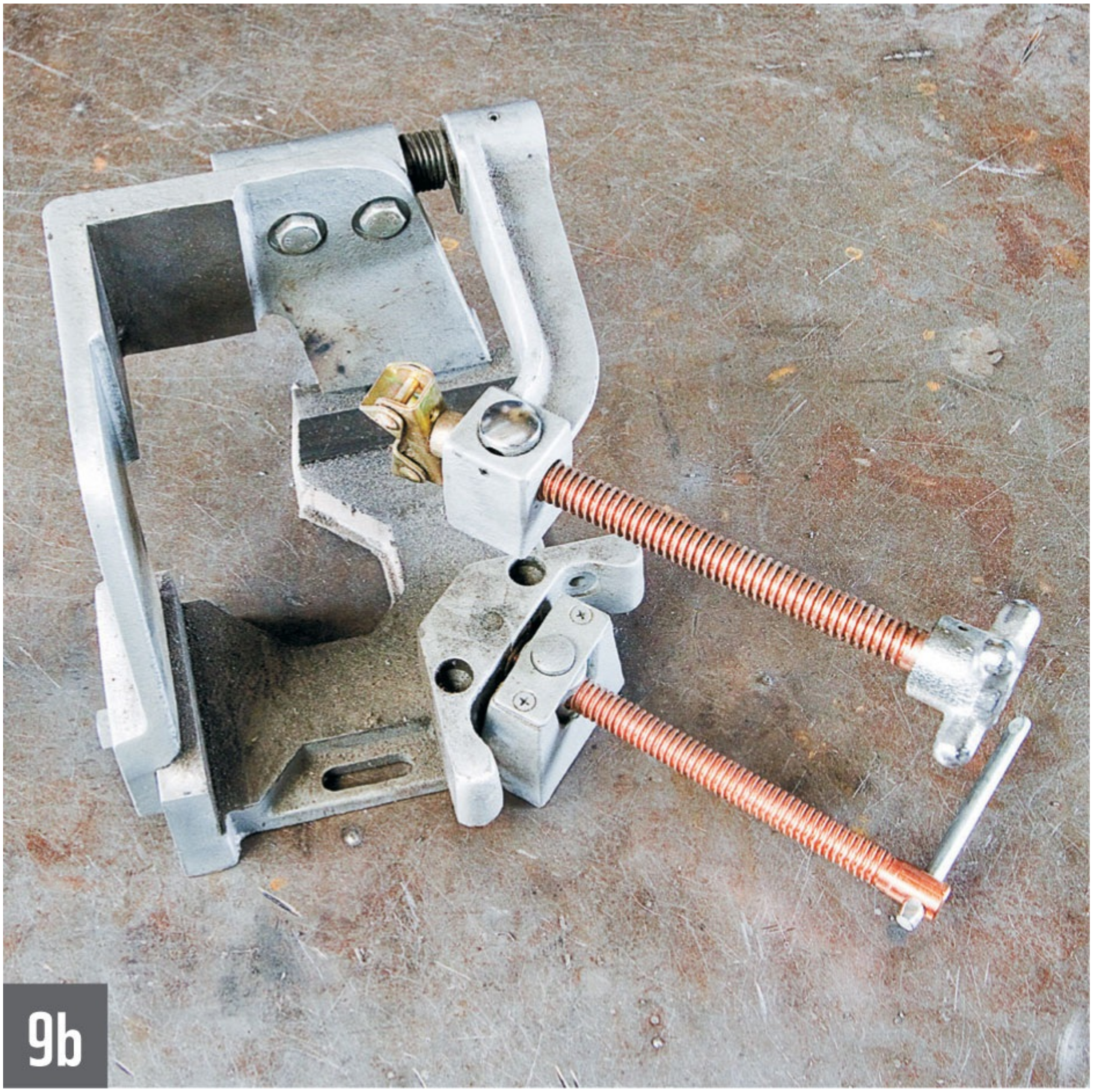

9b

This is a three-axis corner clamp.

A single corner clamp is just like the three axis corner clamp, just without the third axis. It only allows for two pieces to be tacked at a time to create a corner. These are great for creating frames. After a frame is completed, use them to corner the legs. When using this method, it is always good to have a square corner the legs. When using this method, it is always good to have a square handy to check the third angle since the clamp is only able to secure the two axes.

Magnets are also really handy as a good hand support when doing vertical welds.

The second way to create corners is with magnetic squares. Magnets are a great and inexpensive ways to help keep angles of all degree corners as the pieces are being tacked together. They come in a wide variety angles, sizes, shapes, and power. Use four of the square magnets to frame out the entire frame or just one or two corners at a time.

Magnets by themselves will not be able to prevent movement from the heat-affected zone. Use the magnets to square up the piece, but clamp it down when tacking and welding.

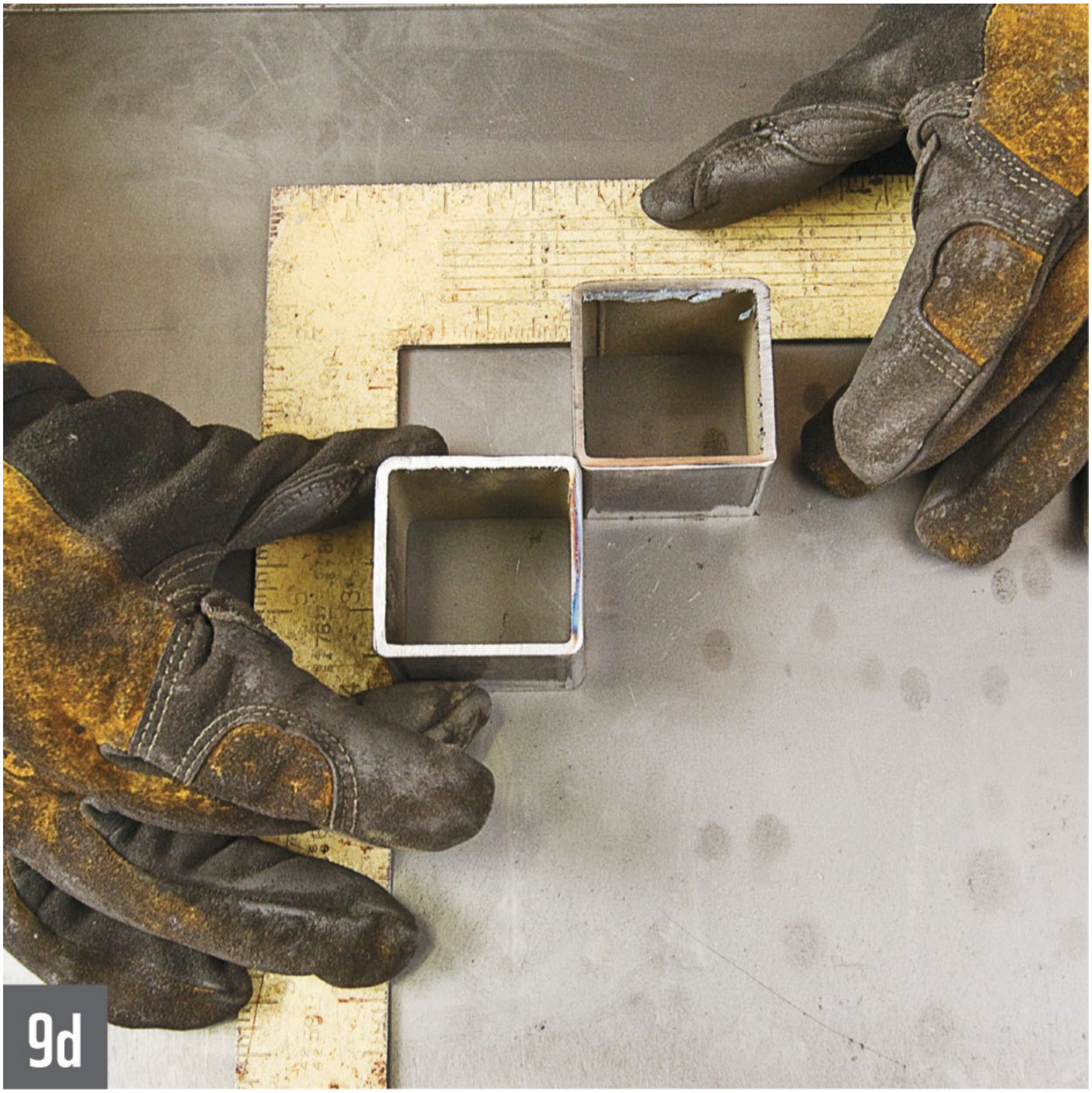
9d

Squares of all kinds can help you make a jig.

The third method is the same method we used to make our cube. Make a corner jig out of scrap wood or metal lying around the shop. Cut a few pieces of scrap that can be screwed or welded onto a flat surface. Using a square, line up the angle and clamp the pieces down and secure them in place. This will allow you to clamp the steel down at a 90-degree angle. Leave the corner open as an access point for tacking.

I recommend using steel if you have it available. It’s much more ridged than a wood jig and obviously won't catch fire, but in a pinch wood can work. The same technique can be achieved by just tacking some stock steel to a piece of flat heavy gauge steel.

It's a good idea to have more than one type of square available when you're working on a project.

The fourth and least expensive method of creating a corner is using the flattest spot on the floor and a square. This will require tacking quickly and being as careful as possible. This is not the most accurate way of creating a square corner, but in a pinch and if the tolerances for the project are pretty loose, it comes in handy. If you don't have a flat spot on the floor, the nice piece of flat sheet metal for the top of your table can be put on sawhorses. This will provide a sheet metal for the top of your table can be put on sawhorses. This will provide a temporary table on which to make the table's frame.

Once everything is cut and prepped, it's time to tack it all together. Lay your materials and equipment out and make sure you have everything you need close by.

Tack

Now that you have chosen the method of creating a corner, let's get to tacking the table together. One of the easiest ways to make any normal table is to start with the top and work down, working with the table inverted. We'll start by creating the frame for our top.

Run though the checklist of things that needs to be done. I will be using a 2

Run though the checklist of things that needs to be done. I will be using a 3-axis clamp for the demo.

Tack the top part of the frame together one corner at a time to make a square frame. For each joint, use four tacks, two on the outer corner of the joint and two for the inner. Always try to go to the opposite side from the last tack, spreading out the heat to reduce movement; the tack should have good penetration, enough to hold the structure together if it is moved.

Welder Checklist

- Is the shield gas on and at the right CFH for the job?
- Is the voltage and wire speed combination correct for the thickness of the material?
- Is the ground clamp on the piece or on the table the piece is on?

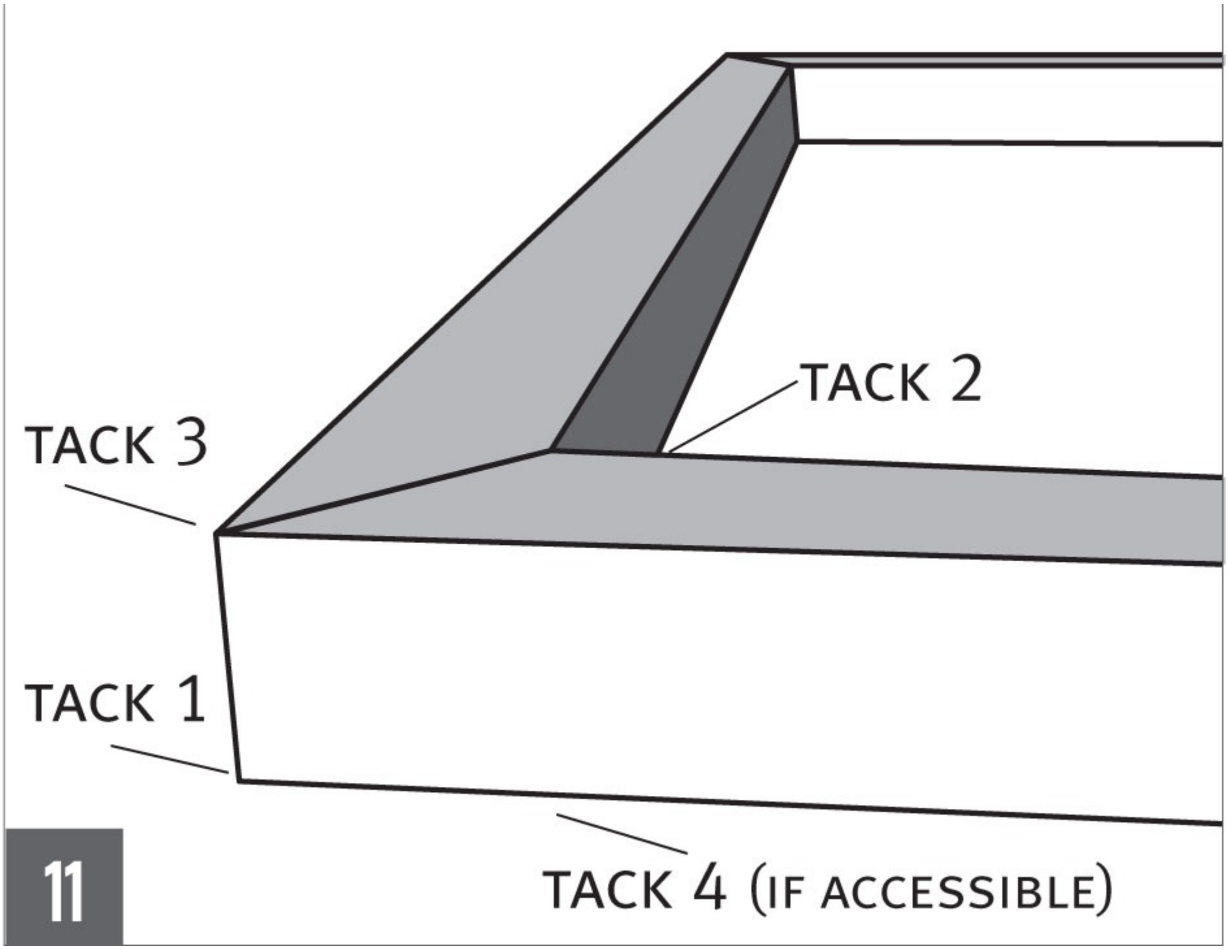

Start by tacking one corner and then move to the corner diagonally across from the one you just did. For example: Start with the outer top tack and then move to the inner lower corner tack.

Tack the top frame together first and then add the legs to the frame.

Repeat this process with all the corners. Tack the entire frame together.

Repeat these same steps to create the two legs and a base rail, matching up the miters to create the corners. This will make up the two “U” shaped legs.

Using one of the corner methods described earlier (page 145 or 146), secure one of the U-legs into place. Square up the leg and tack it into place with four tacks, one at each corner, using a cross pattern. When finished with the first U-leg, tack on the second. DO NOT weld any parts completely together at this time. These U-legs should mirror each other.

Miller
DIGITAL Elite

13

Next, add the leg cross supports and center joining support.

After all the legs are tacked into place, tack in the lower support cross braces, again using the same cross-pattern corner-tacking method. Use magnets or clamps to hold it into place as you tack.

Tack in the 1-inch (2.5 cm)-square rails that will be the clamp holders. Here's a little quick trick: If you don't have magnets, use two pieces of 2x4 lumber or a drop from the 2-inch (5 cm)-square tube to evenly space out the supports by cutting two even lengths to prop the tube up. Tack them into place. Use the same corner-to-corner tacking technique as you used on the legs.

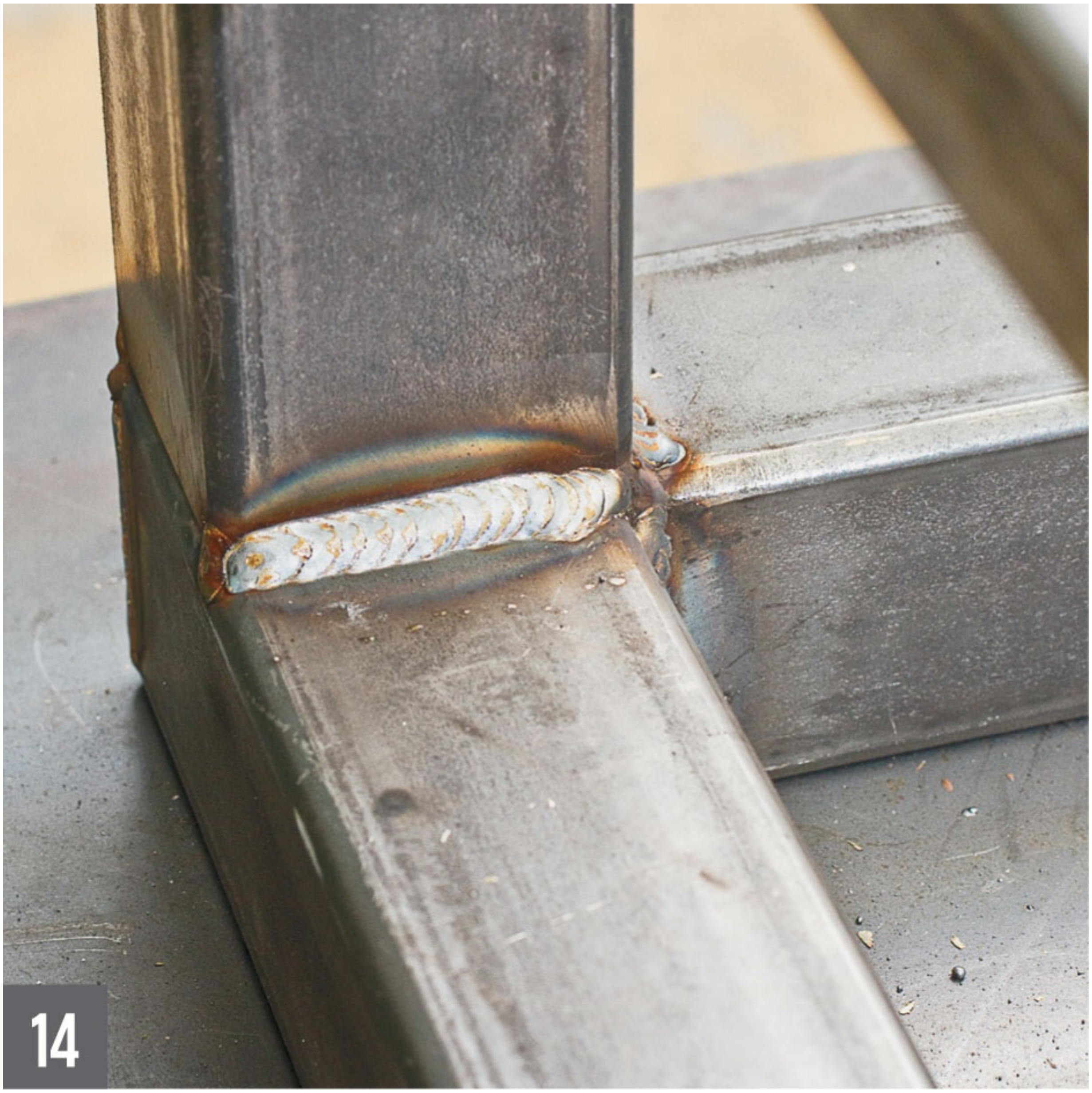

14

After it is all tacked together, the next step is to weld. Use a stainless steel wire brush to remove the burn from the weld zone left over from tacking.

Weld the Frame

Now that the structure is all tacked into place, this is a good time to make sure all the measurements are accurate and the corners are square. If every thing meets up with your tolerances, it’s time to weld.

Weld all the joints on the table so that the heat is spread evenly throughout the piece. A good pattern is to weld each joint and the opposite side, then move to another connection. For example: Start at one leg, weld the two joints facing each other, and then do the same to the leg diagonal from that leg and repeat. If the tacks were done correctly, they should have enough strength to maintain the position they are in without breaking.

Repeat this process with the supports and cross beam. Remember to spread the heat out and minimize the HAZ of each area.

WISCO
Welding Industrial Supply
2200 North West
Millermatic 211
DIVERSION

This table is going to be bigger than anything you have worked on so far. This means different positions. Just remember to be comfortable. Brace and support the gun the best you can.

Now that the frame of the table is complete, it is time to prep and weld in the casters. The method you use will depend on the type of caster you chose. I prefer stem casters; if the table is going to be stationary for a long period of time, the casters can act as leveler feet. They can also be replaced much more easily in casters can act as leveler feet. They can also be replaced much more easily in case one fails. Four-hole plate casters can be welded into place, which is a plus, and should be used if you do not have a drill press (which is required for stem casters), but they are harder to replace than the stem style. Regardless of the style chosen, all the casters should swivel and have a lock.

If you’re using swivel casters, you will need to make the footers that will seal the legs. Using the piece of 2-x 24-inch (5 x 61 cm), 3/16- or 1/4-inch (5 or 6 mm)-thick stock (length doesn’t matter; this will be chopped into 2-inch [5 cm]-square pieces after the holes are drilled), mark off 2 inch (5 cm) sections, leaving a 1/8-inch (3 mm) gap between each of the 2-inch (5 cm) sections (or whatever thickness is appropriate for the cutting method used to allow for the kerf of the blade). For example, 1/8 inch (3 mm) is the width of a chop saw blade, while the width of a band saw blade is closer to a 1/16 inch (2 mm).

WARNING

Never drill into metal without using drilling and tapping fluid.

Using a center punch, mark the centers of each of the 2-inch (5 cm)-square pieces for drilling. Choose a drill bit according to how the caster will be mounted. If the hole is to be drilled and tapped, a smaller bit will need to be used to allow for threading. If a nut is to be welded to the inside surface, a larger bit will be needed to clear the stem’s threading.

Set up the drill press using the appropriate size drill bit for the stems of the casters. Align and clamp the fence 1 inch (2.5 cm) from the point of the drill bit (the distance to the center). Drill all four holes using drilling and tapping fluid to reduce heat from friction; then cut each of the 2-inch (5 cm)-square sections away using the lines that you measured.

Unless you’re using the tapping method (in which case you can skip this step), it is time to weld on the nuts after all the holes have been drilled and caps have been cut. If the nuts are zinc coated, remove this coating with a flap disk on a grinder. Remember to always wear a respirator when welding anything with a zinc coating. Using spatter protection, spray, or motor oil, lubricate the stem and mount the caster in the hole, securing it into place with the nut. Weld the nuts into place using three heavy tacks; this is all that will be needed to secure them into place.

Unscrew the casters. Prep all of the surfaces by removing all mill scale from the weld zones and chamfering the end caps. These are now ready to be welded on.

During the fabrication of this piece so far, there may have been a small amount of movement that made the legs uneven. Here’s a little trick to fix that. More than likely the most flat and true surface in the shop is the piece of steel that you purchased for the top. Make sure all of the surfaces have been leveled. Flip over the frame of the table so it is upright and standing on that sheet of steel; then put a cap under each leg. If there is any gap, it can be eliminated by welding over it to fill it in. Check the level of the top frame. Tack the caps onto the legs by two corners across from each other on each leg. Go back and tack the rest of the corners.

Weld the caps into place. Start with the two opposite side joints on one leg and move to the next. Then go back to the first leg and weld the other two joints and move to the next. Then go back to the first leg and weld the other two joints to completely weld the caps into place.

If you are using the tapping technique to mount stem casters, flip over the frame. Using your feet to secure the table, begin tapping the hole using plenty of drilling and tapping fluid. Start slowly; once the threads start to form, it will catch. Make sure you stay level and upright. Screw the tap all the way down and then back it out smoothly. It’s a good idea to put the tap back through the freshly made threads to remove any steel burrs or debris.

If using the four-hole plate casters, make sure to prep the plates to be welded by removing the zinc coating with a grinder and a 60-to 80-grit flap disk; then cut out four steel plates the same size as the plates mounted to the casters. Flip over the frame of the table so it is upright and standing on the sheet of steel and center the plate caps under each leg.

If there are any gaps, they can now be eliminated by welding over them. Check the level of the top frame. Tack weld the caps onto the legs by two corners across from each other on each leg; then tack the rest of the corners.

When you’ve completed the tacks, weld the plates into place. Start with the two opposite side joints on one leg and move to the next; then go back to the first leg and weld the other two joints to completely weld the plates into place.

Weld the Top

Now we move on to the top of the table. Over a long period of time the plate of steel that is used for the top may become warped and distorted. When going to weld the top onto the frame it is not necessary to weld the top on completely in the square. A few 3-inch (7.6 cm)-long stitch welds will be fine to secure it into place and allow you to replace it later on if needed.

Begin by prepping all of the surfaces that are to be welded with a flap disk on a grinder.

Mark all the sides on the plate to center the frame. For example, if you decided to have a 3-inch (7.6 cm) lip, mark off the distance from the edge of the frame to the edge of the plate.

Secure the frame with clamps and tack into place. Before welding, remeasure for accuracy. Weld two 3-inch (7.6 cm)-long welds on the inside a few inches (5 to 8 cm) from the corner joints connecting the frame to the plate. Warpage won’t be as much of a factor for these welds, but it is always good practice to spread out the heat as much as possible. Start one weld and move straight across to the opposite side. Repeat until all six 3-inch (7.6 cm)-long welds are complete.

Congratulations! You have completed your welding table project, your first functional project as a welder. All that is left to do is screw in the casters (if you functional project as a welder. All that is left to do is screw in the casters (if you used that method), flip it over, and be done (unless you want to grind down the visible welds to make it look more appealing). Now you can be more comfortable when welding while saving space and time.

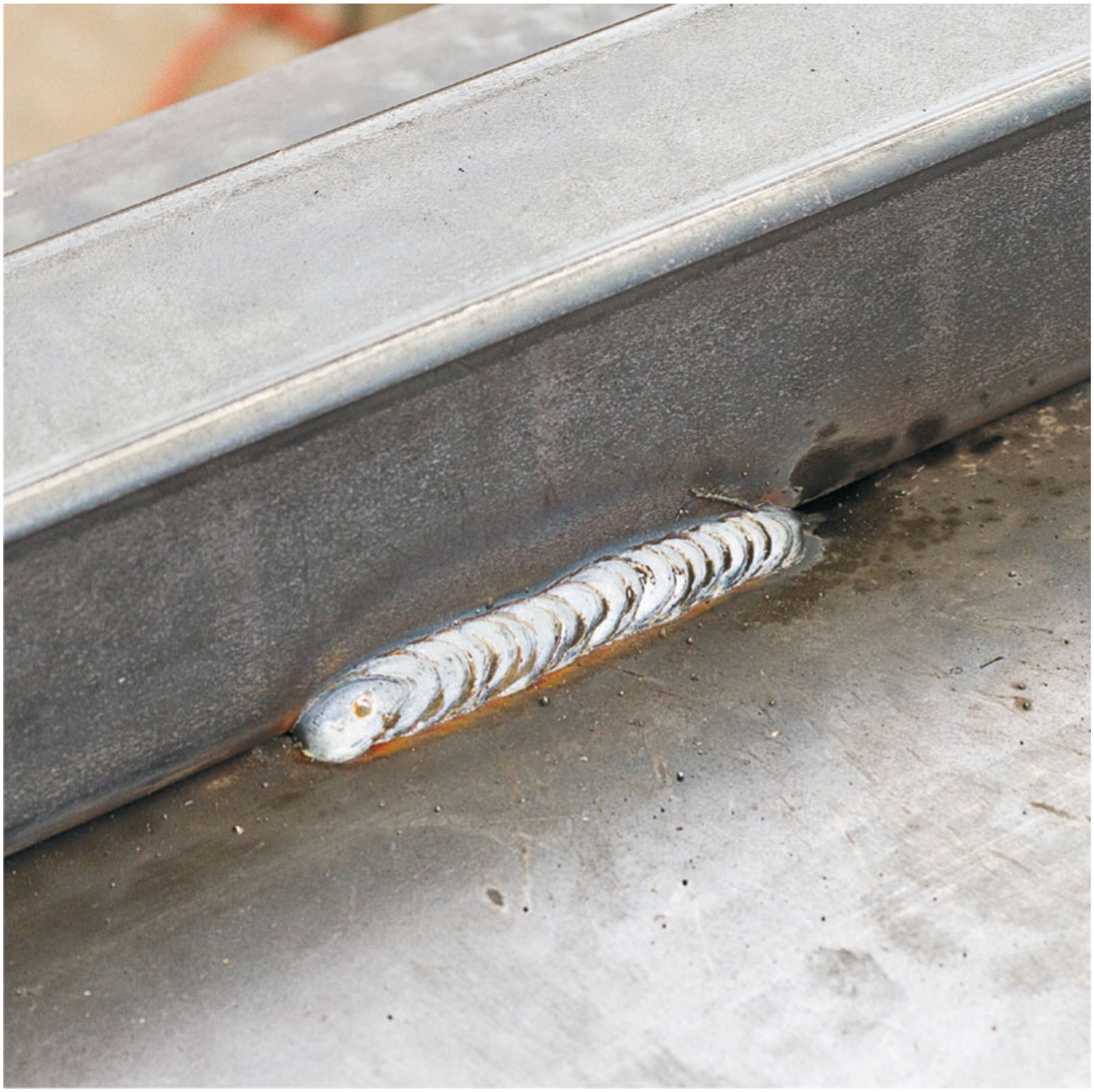

It is not necessary to weld all the way around the top of the frame you just made. Four to six 2″ to 3″ (5 to 7.6 cm) fillet welds will work fine.

Protecting Your Table

It's not necessary to paint your welding table, and in fact, it may even be a bad idea. Most paints used on metal create a layer of insulation that acts as a barrier that will not allow the current from the ground to circulate through. You want to be able to clamp the ground in multiple places on the table, so I do not recommend painting. If you would like to protect it from rust, use a rag to wipe on some oil or rust inhibitor or spray it with spatter protection spray. Over time, a lot of metal dust and particles can take a toll on the casters, so keep the bearings well-lubricated. Axle grease makes a great barrier for protecting bearings.

Project 3: The Runner

SKILL LEVEL: Intermediate

MATERIALS AND EQUPMENT:

- MIG Welder
- 24’ (7.3 m) of 11 to 16 ga cold roll 1" (2.5 cm)-square tube
- Welding helmet
- Welding gloves
- Long-sleeve shirt
- Leather welding jacket or apron
- MIG pliers
- Stainless steel wire brush
- Soapstone
- 4 1/2" (11.4 cm) angle grinder outfitted with an aluminum oxide grinding disk or a 60-to 80-grit flap disk
- Clamps
- Chop saw
- 3-Axis clamp or right-angle magnets or jig
- 3/8 x 10 x 58" (1 x 25 x 147 cm) piece of oak, or other wood of your choice

Cut, Prep, Tack, Weld

Sofa table, couch table, long table, or runner; no matter what you call it, this is a great table to fabricate. It’s a stand-alone table that doesn’t require a matching counterpart. It also doesn’t have a set purpose like a coffee table or end tables. A runner can go anywhere in the house, from the back of the couch to the hallway to the kitchen. It’s a nice and versatile table to have, plus it’s fun to fabricate.

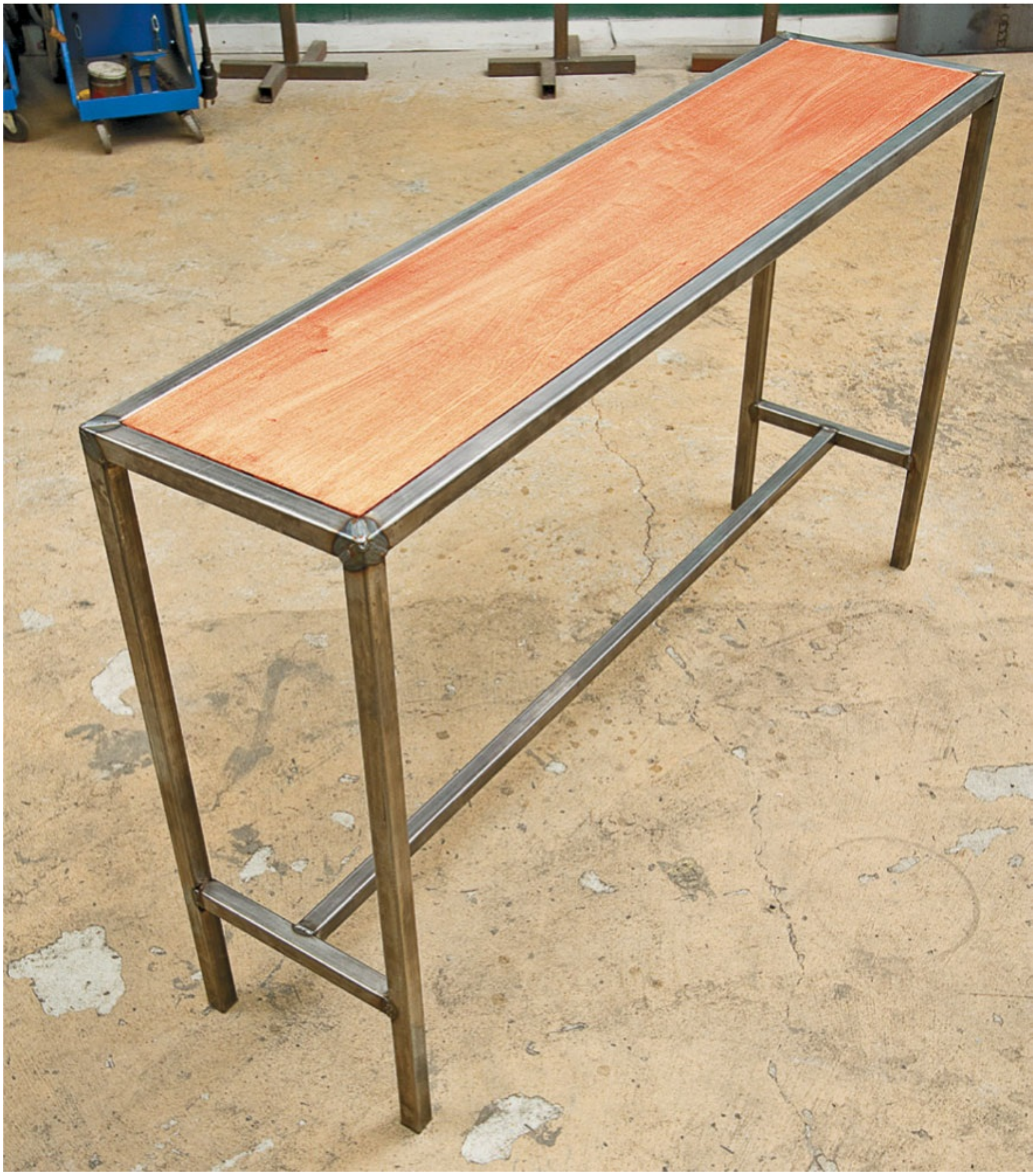

This is a runner table.

After making your own welding table, this should be a piece of cake, right? Well, I'm going to throw in a new little twist that is going to take a little more concentration when cutting and measuring. All the principles of what you have learned so far will apply to this table. At this point, I should not have to tell you to cut-prep-tack-weld, but I'm going to.

This table is something that is going to be admired for its appearance so we

This table is something that is going to be admired for its appearance so we are going to do a few things to make sure the corners, edges, and lines remain clean. This will be a practice in patience and accuracy to make the final piece have a nice overall look.

For this project, we will be using cold roll 1-inch (2.5 cm)-square tube with a 1/8-inch (3 mm)-thick wall. The top four corners of the frame will be welded to the legs with a double miter, unlike the welding table where the four corners of the top frame were single miter with the legs welded on flush.

I prefer the double miter method when making furniture. It gives the piece a cleaner and more aesthetically pleasing look with all visible corner welds ending at the same point. And after the welds are ground down, it also looks as though the framework was made from one piece of metal, as the seams are not as obvious.

Since this project will require more accuracy, I recommend something a little sharper to mark with. Silver or black oil pencils, fine point markers, or to be extra accurate, a good scribe is the best. A scribe is a hard metal point used to scratch measurements in the surface of other metals, normally made of hardened tool steel, titanium, or other metal.

Cut

As you can see from the measured drawing, cutting for this project is going to be a challenge. Each length of the top frame of the table is going to be double inward-miter cut at both ends. All of the legs of the table will have a double miter at just one end.

This means that for the top frame, the longest part is going to be the edge that holds the miter cut. Cut the lengths of all the pieces to the full size with no miter first.

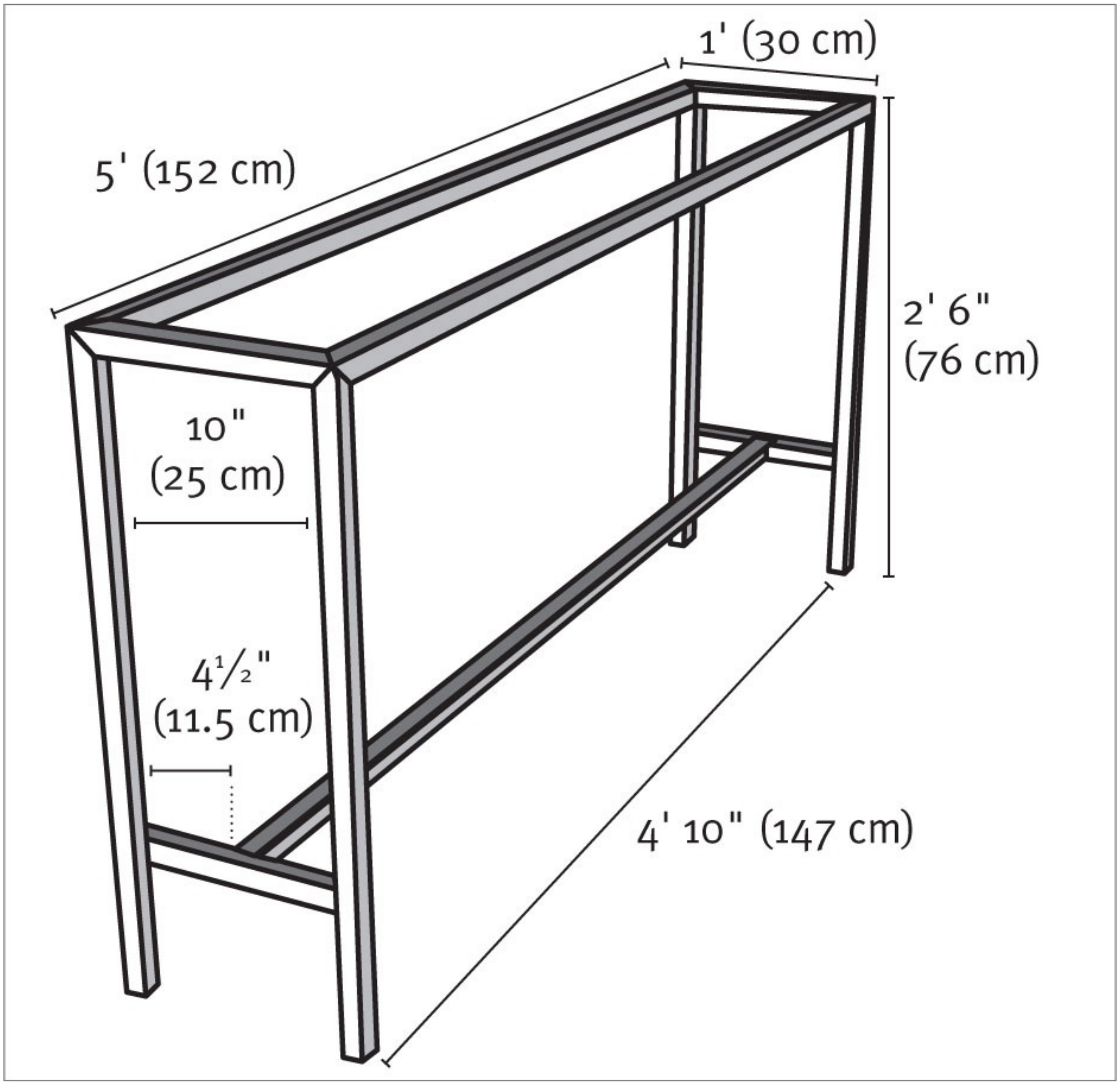
1' (30 cm)
5' (152 cm)
2' 6"
(76 cm)
10"
(25 cm)
4½"
(11.5 cm)
4' 10" (147 cm)

This is the measured drawing.

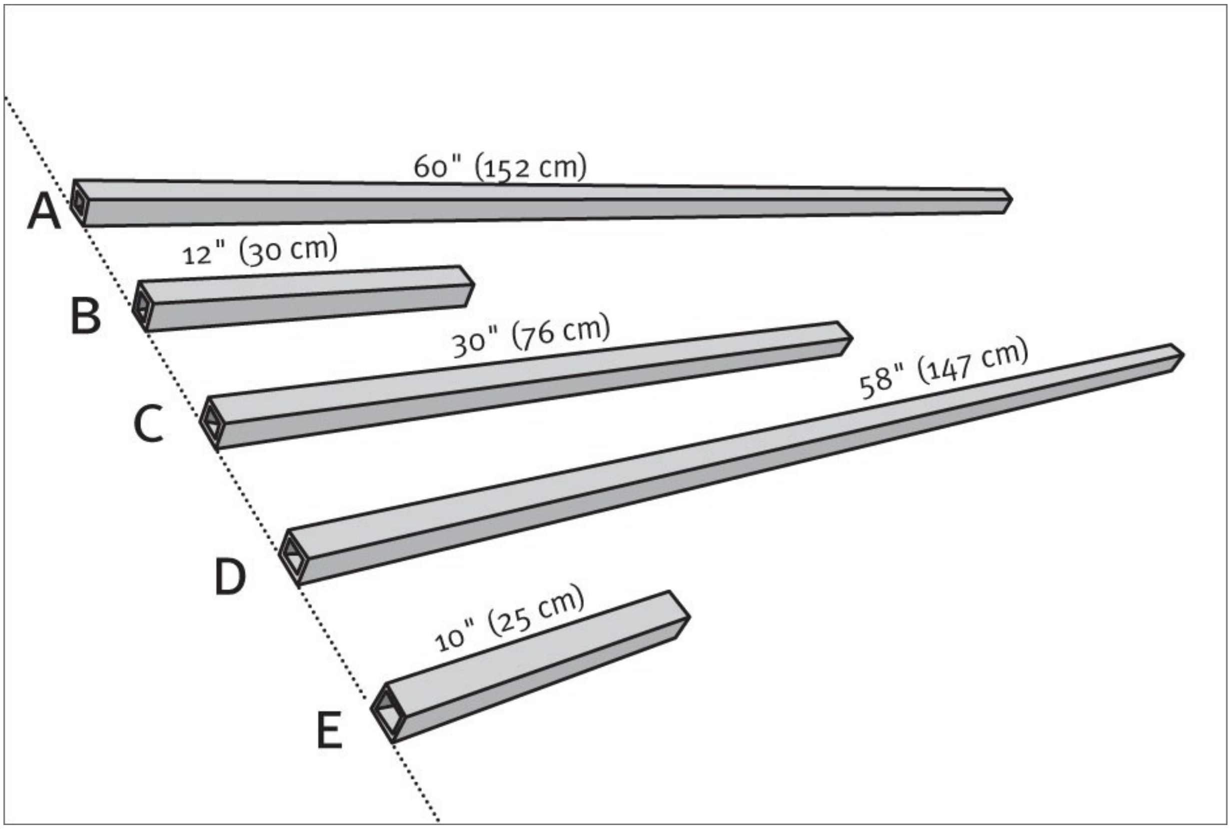
A
60" (152 cm)
B
12" (30 cm)
C
30" (76 cm)
D
58" (147 cm)
E
10" (25 cm)

The cut list: A. Two top frame sides; B. Two top frame ends; C. Four legs; D. One low cross support; E. Two lower end cross bars

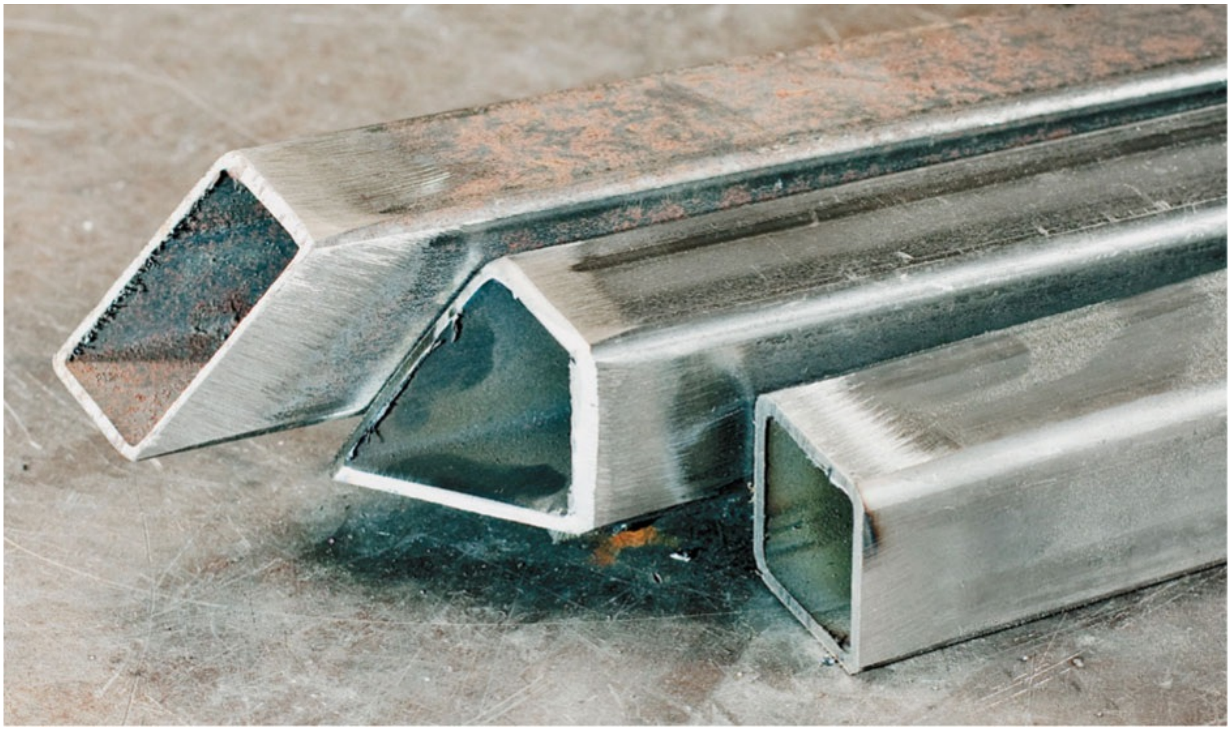

This shows a single miter, double miter, and straight cut.

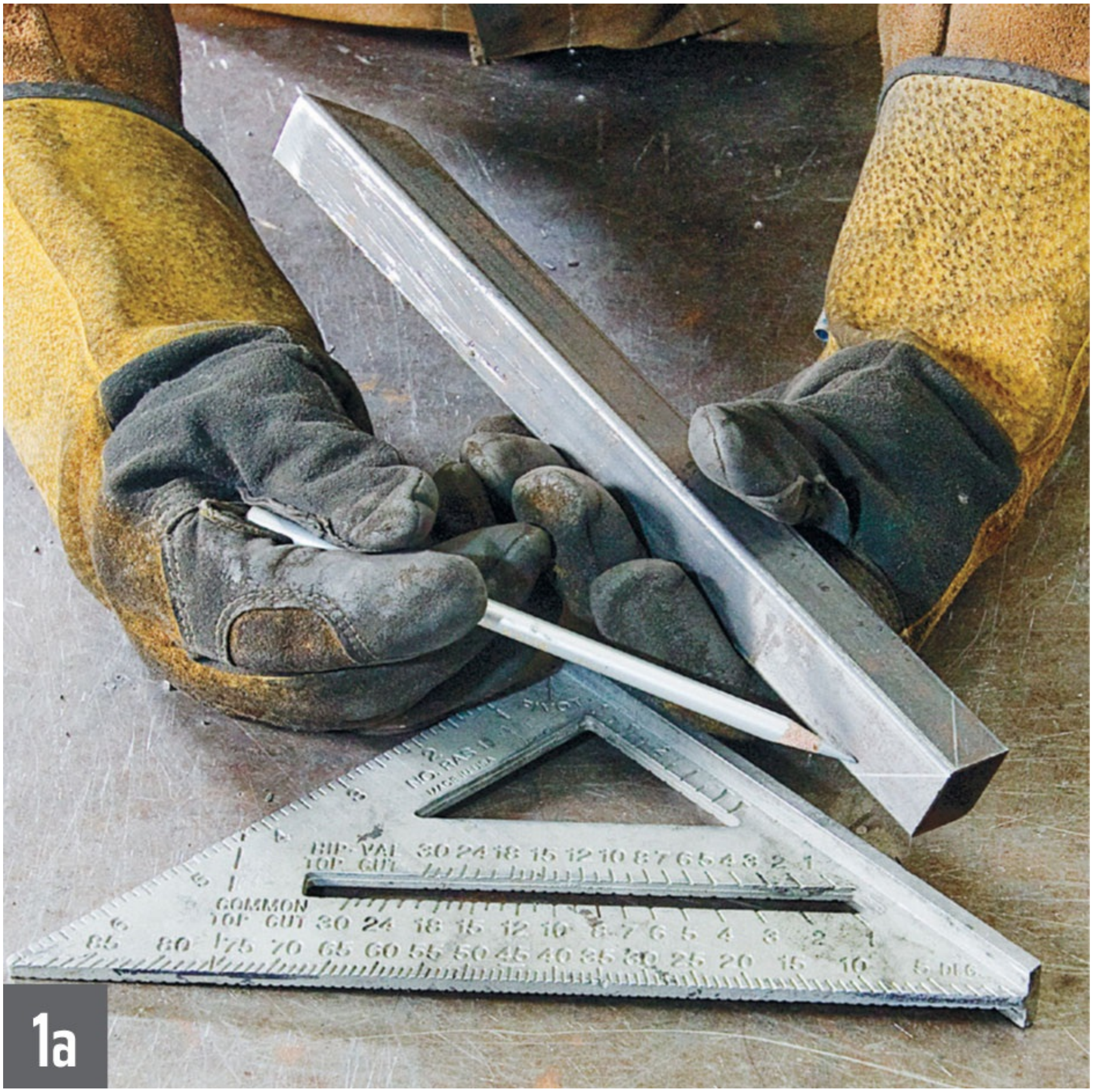
PIVOT
HIP-VAL TOP CUT
COMMON TOP CUT

1a

Always mark a double miter first.

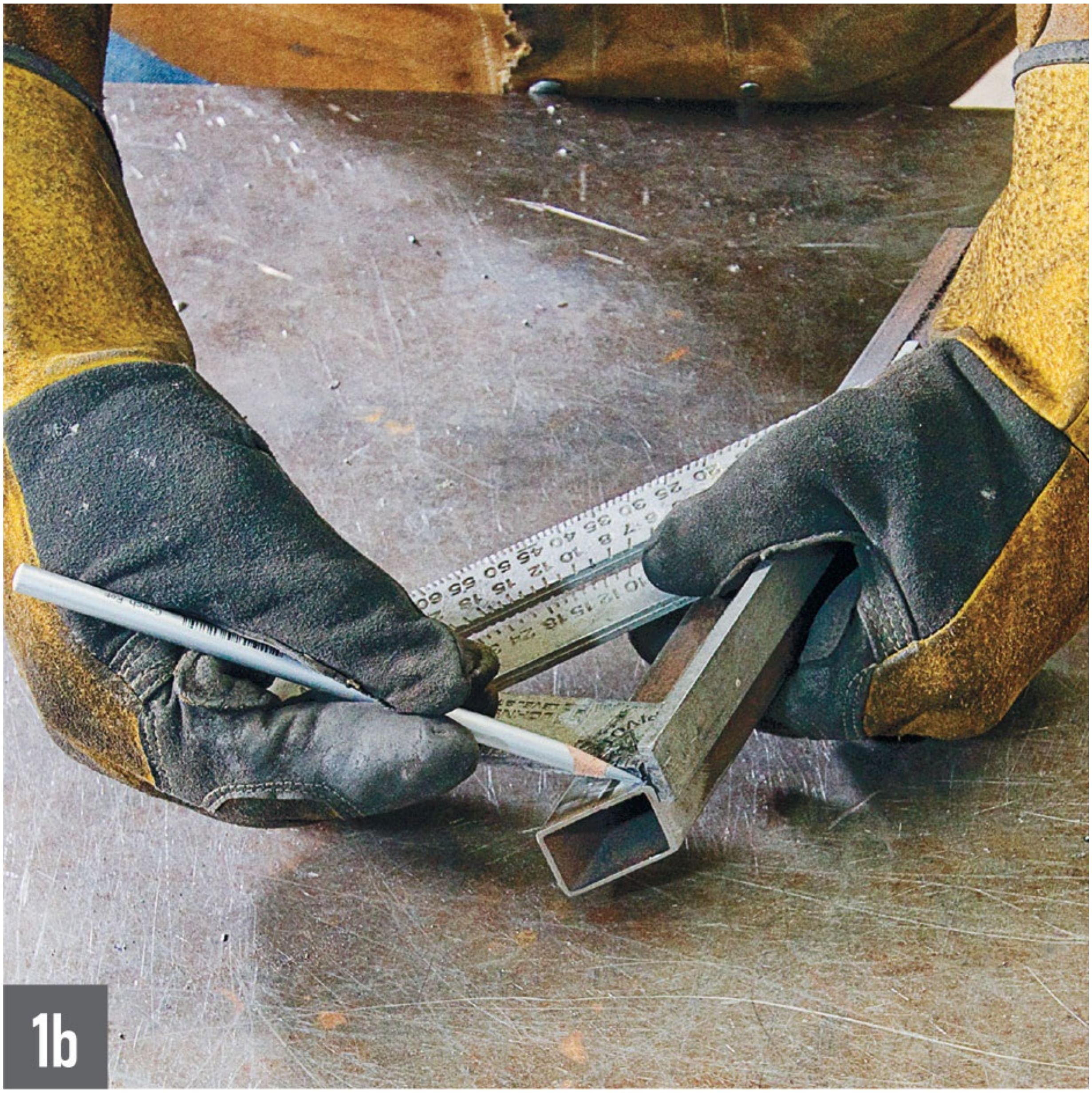

1b

Mark the next cut for the second miter to the outer most edge of the first miter.

It’s away a good idea to give yourself a little extra room (such as an extra $^{1}/_{16}$ inch [1.5 mm]) when cutting. This way, if you need to correct the length, you can grind it down.

Cut the single miters first, clearly marking all the cuts. This way you don’t accidentally cut the wrong edge.

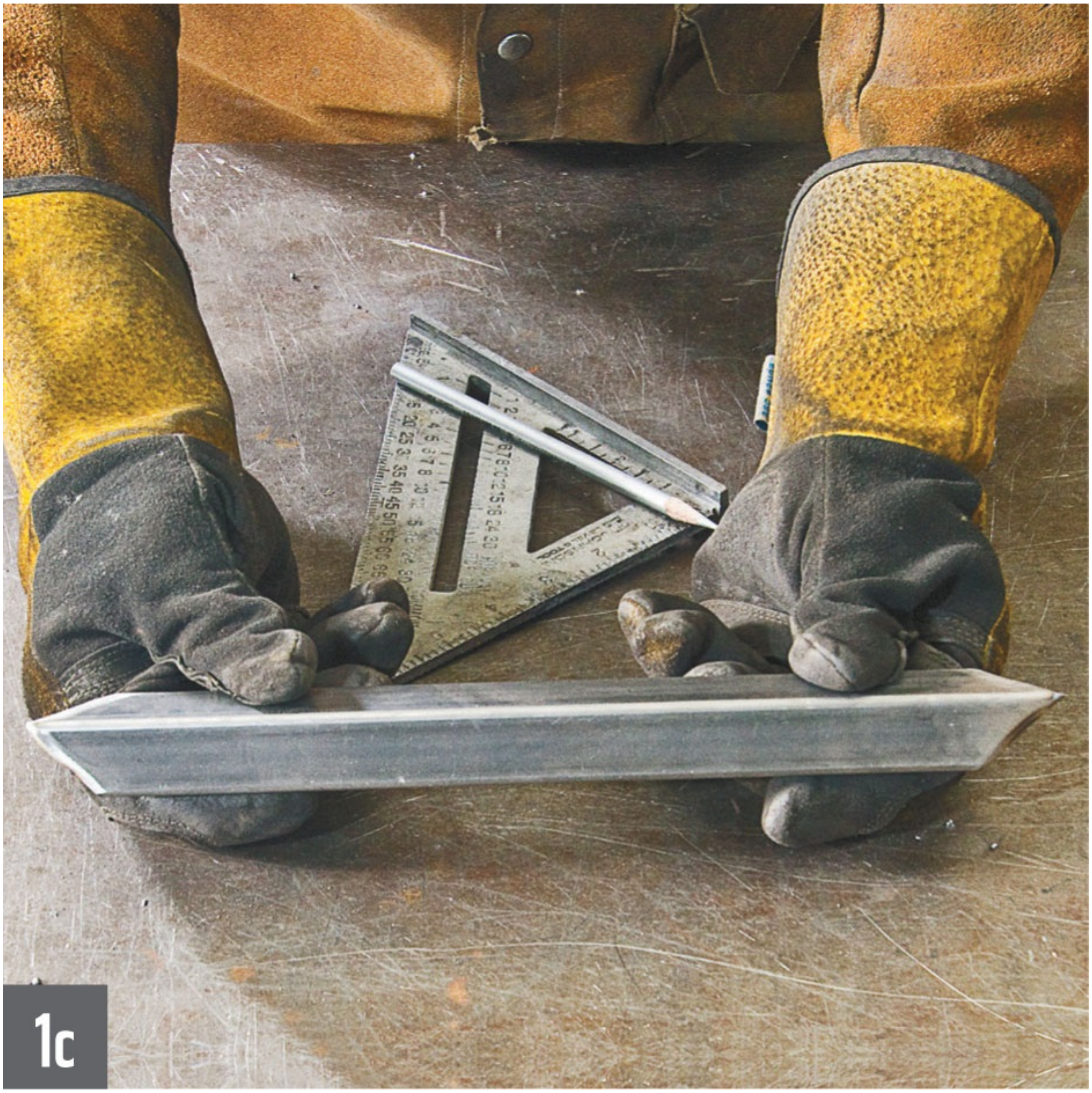

1c

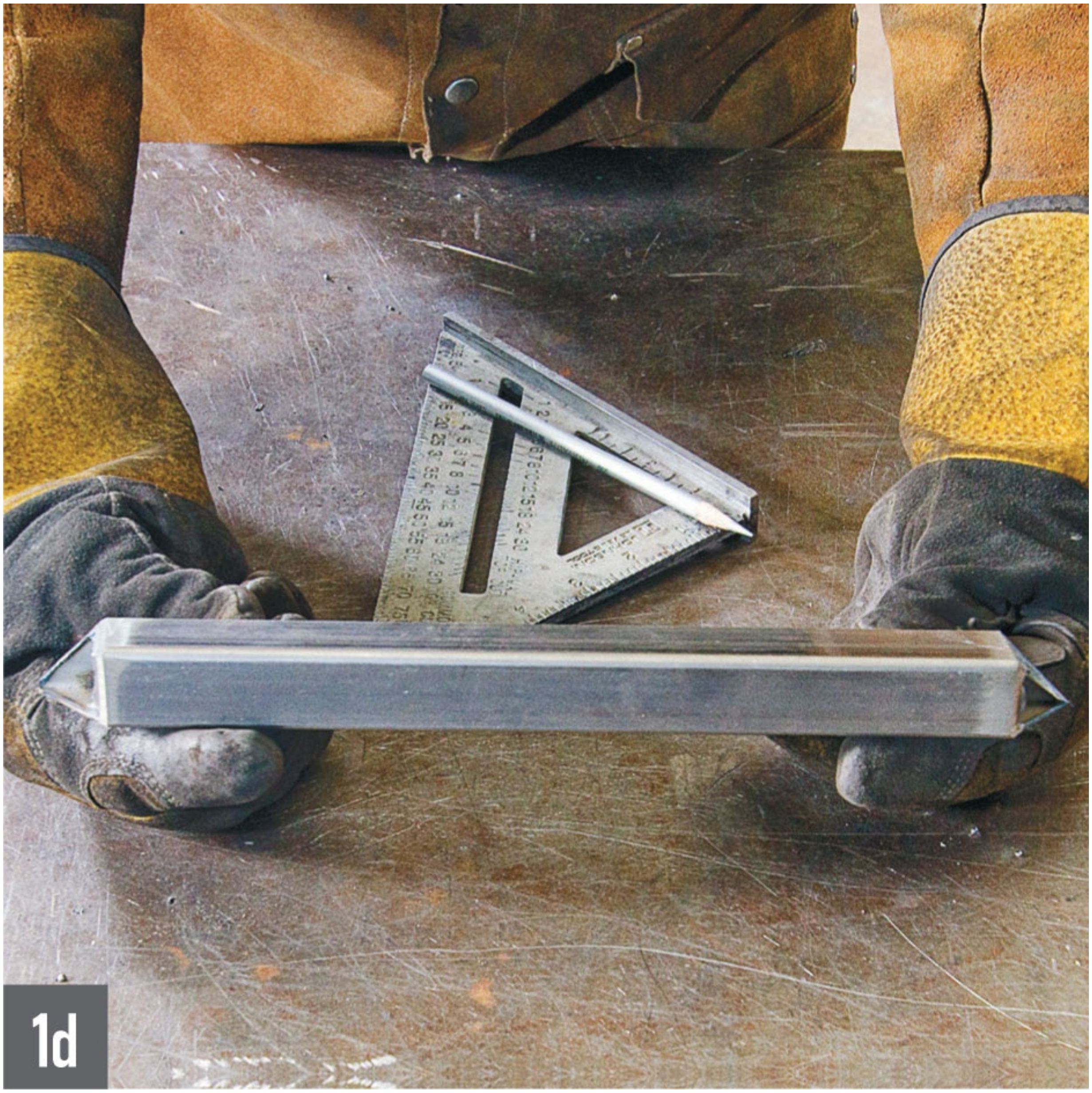

1d

The very points of the final cuts of the double miters should be on the same edge.

Since the chop saw only cuts 45 degrees in one direction, you will need to pay close attention to which way you turn or flip the material to be cut. This is why it’s also good to mark everything clearly before you cut.

Now that you have the first miter cut in all of the pieces, it’s on to the second miter. In order to give yourself a better guide, mark a cut line. Use a square to draw a line down the side of the first miter using the outermost edge of the material wall as a guide.

Lay out the cut and prepped pieces to be tacked.

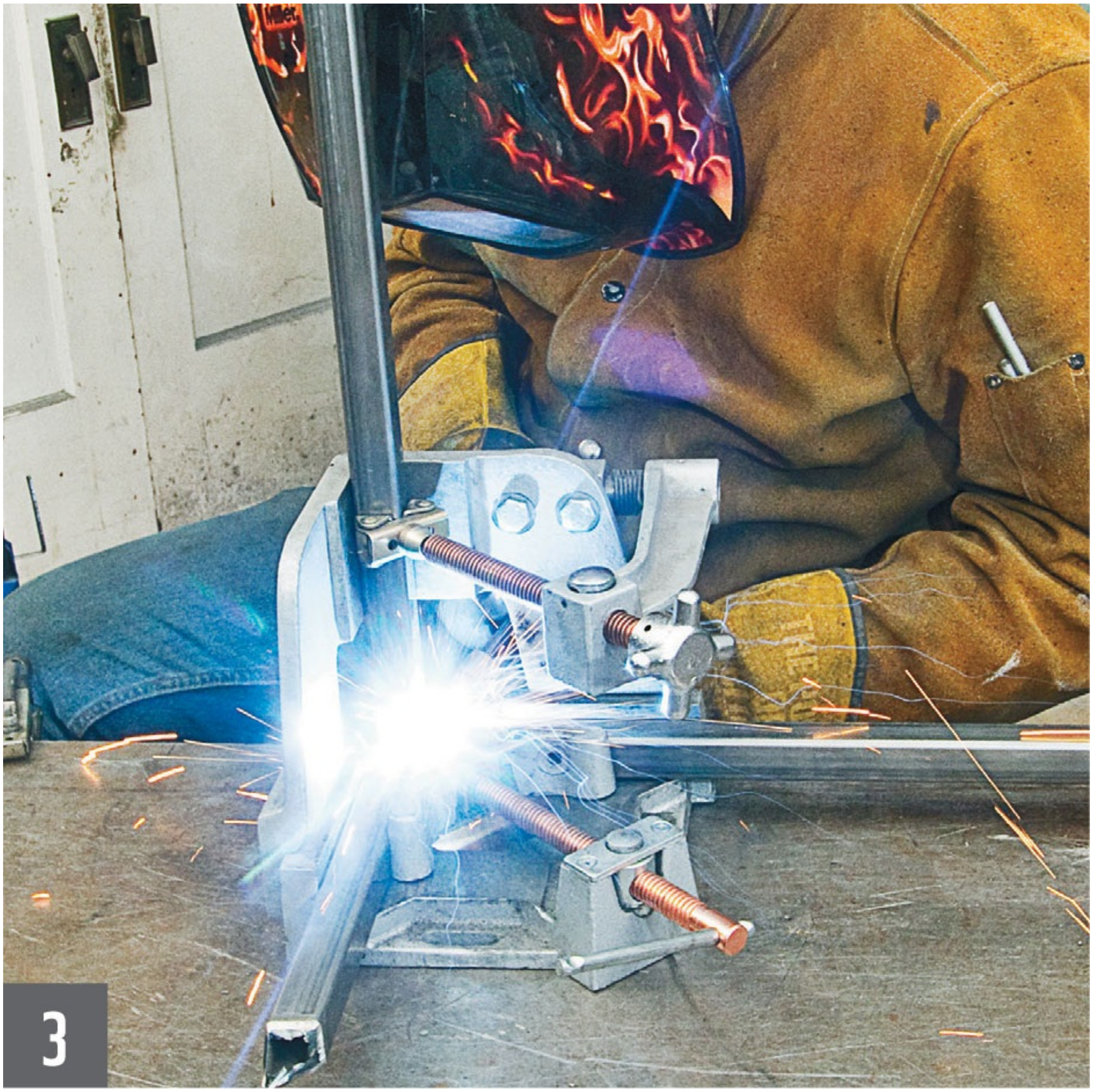

A 3-axis clamp is very useful when using a double miter.

Prep

Once you’ve finished all the cuts and checked that the lengths and angles are correct, it’s time to prep. Prep all of the weld zones with a 60-to 80-grit flap disk. The wall of the tube is 1/8-inch (3 mm) thick, so a 3/32-inch (2 mm) chamfer for our butt joints should be perfect.

Tack

Once you’re prepped and ready to tack, you should set up the MIG for 1/8-inch (3 mm) thick material. By now, you should have also selected your preferred method of joining corners (see Chapter 5).

4
NO. RAS 1
MADE IN USA
COMMON TOP CUT
HIP-VAL TOP CUT

Mark where the cross supports will be tacked into place.

You’ll need to tack all three pieces together: two lengths of the top frame and one leg. I will be using my trusty 3-axis clamp once again; this allows the three pieces to fit together snug like a puzzle and makes it easier to work on them. If you don’t have a 3-axis clamp, make a jig instead as shown in Chapter 5.

Once you have secured the three parts, the tack pattern for this setup will be the outermost corner first, then the innermost corner directly opposite;

the outermost corner first, then the innermost corner directly opposite, essentially it is the two points where all three pieces meet. Finally, tack the three points where only two pieces intersect.

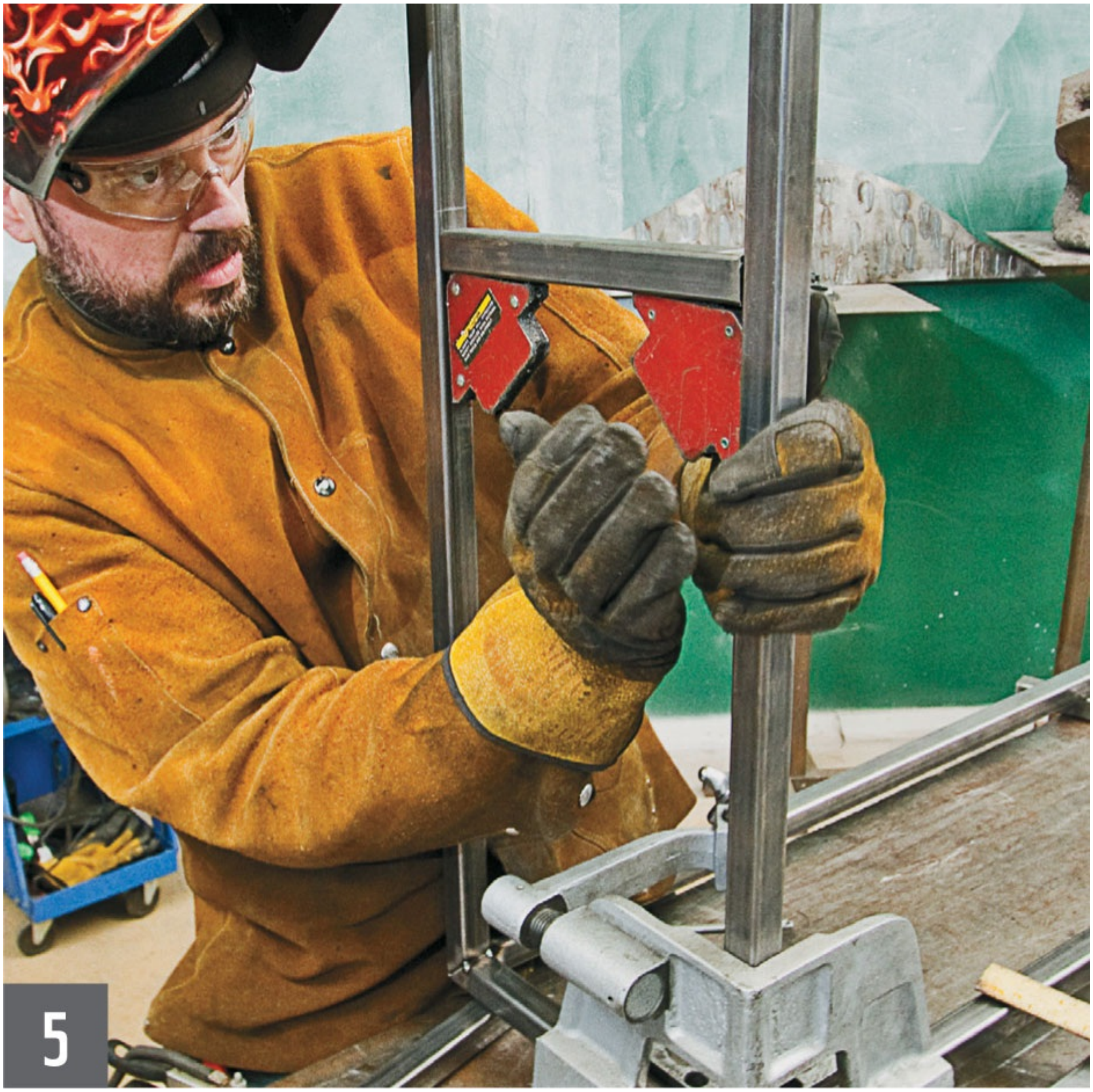

Use magnets or clamps to hold the cross supports in place while you tack them.

Tack the frame and legs together and then measure out the height of the two cross supports and mark where your welds will go.

Clamp or secure the cross supports in place with magnets, corner clamps, or cut lengths of scrap. Do not tack the cross supports in place with four tacks. This is the part where we become more meticulous about where we are going to weld and how it's going to look. We are only going to use two tacks to secure this into place: one tack per vertical butt joint right in the middle.

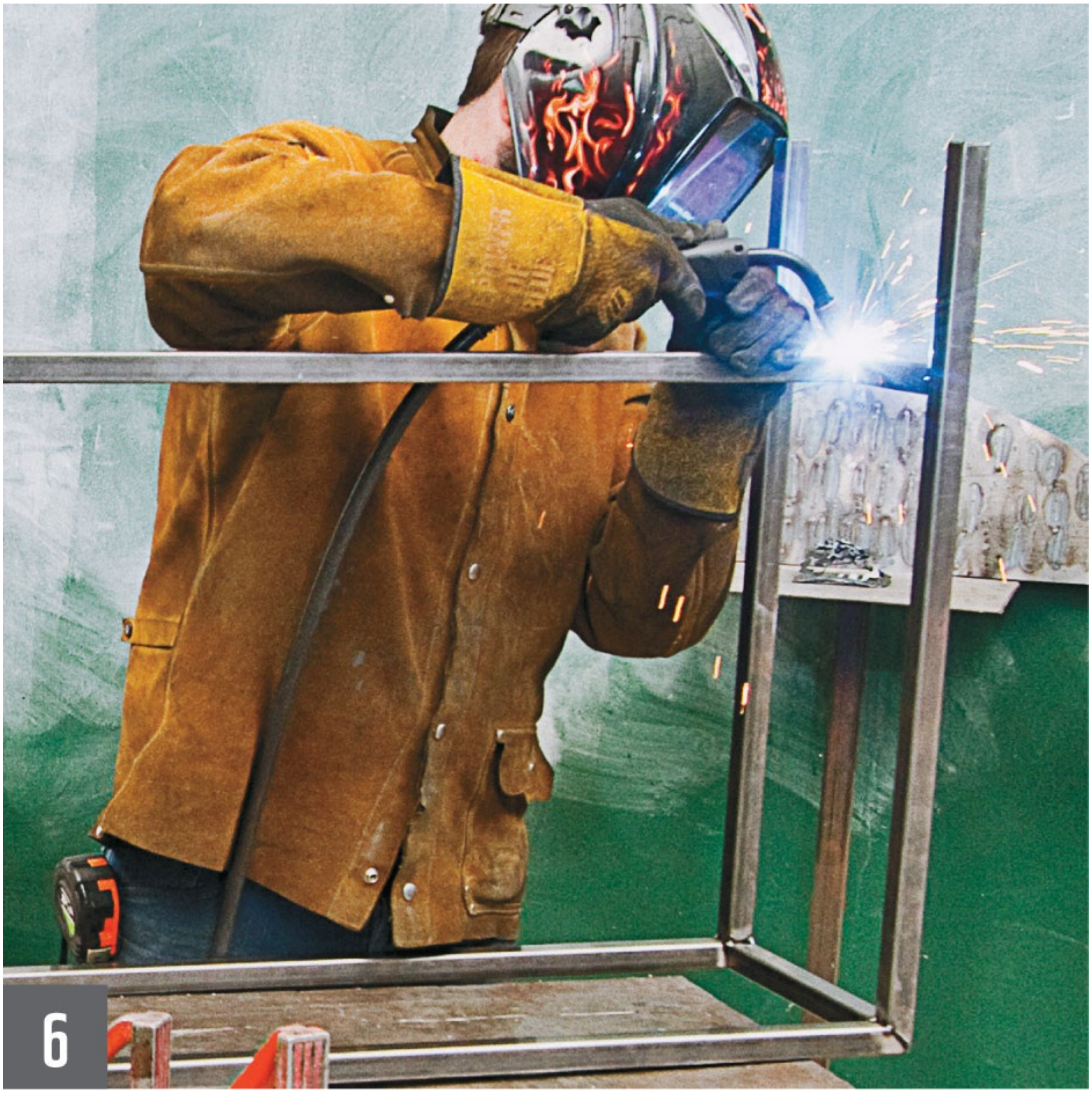

Mark and tack the center support to the leg supports.

Measure and tack the middle support piece in the center between the two cross supports. This will also only use four tacks. The four vertical inner corner joints will not be welded, so do not tack on the inside.

By only welding the vertical butt joints and not the corner fillet welds, you’ll achieve a cleaner look. The reason to only use two tacks for the supports pieces is to leave you the ability to weld over them. Leaving the corners alone will leave a perfect 90-degree corner. After grinding down the welds on the butted joints, it will have a cleaner look.

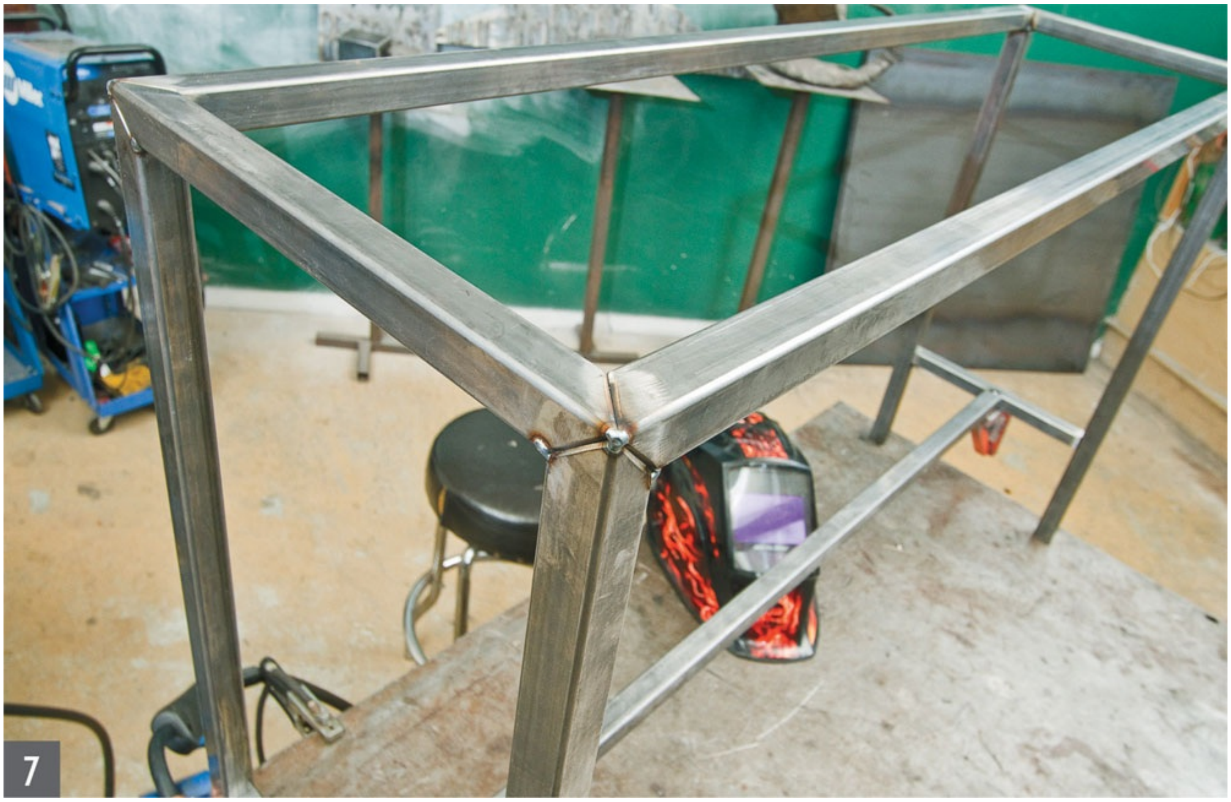

Tack the entire piece together before welding.

8a

Weld the piece together, spreading out the heat as you go.

Weld

Now that it’s all tacked together, give your project a once-over. Check all of the measurements to make sure everything falls into the guidelines of your personal tolerances. If everything looks good to you, start to weld it up, making sure to go in a pattern that spreads out the heat throughout the piece. Start with the vertical butt joints. No, you don’t have to do a vertical weld if you don’t want to, just lay the frame down so the weld is horizontal. Then move on to the corner or fillet welds.

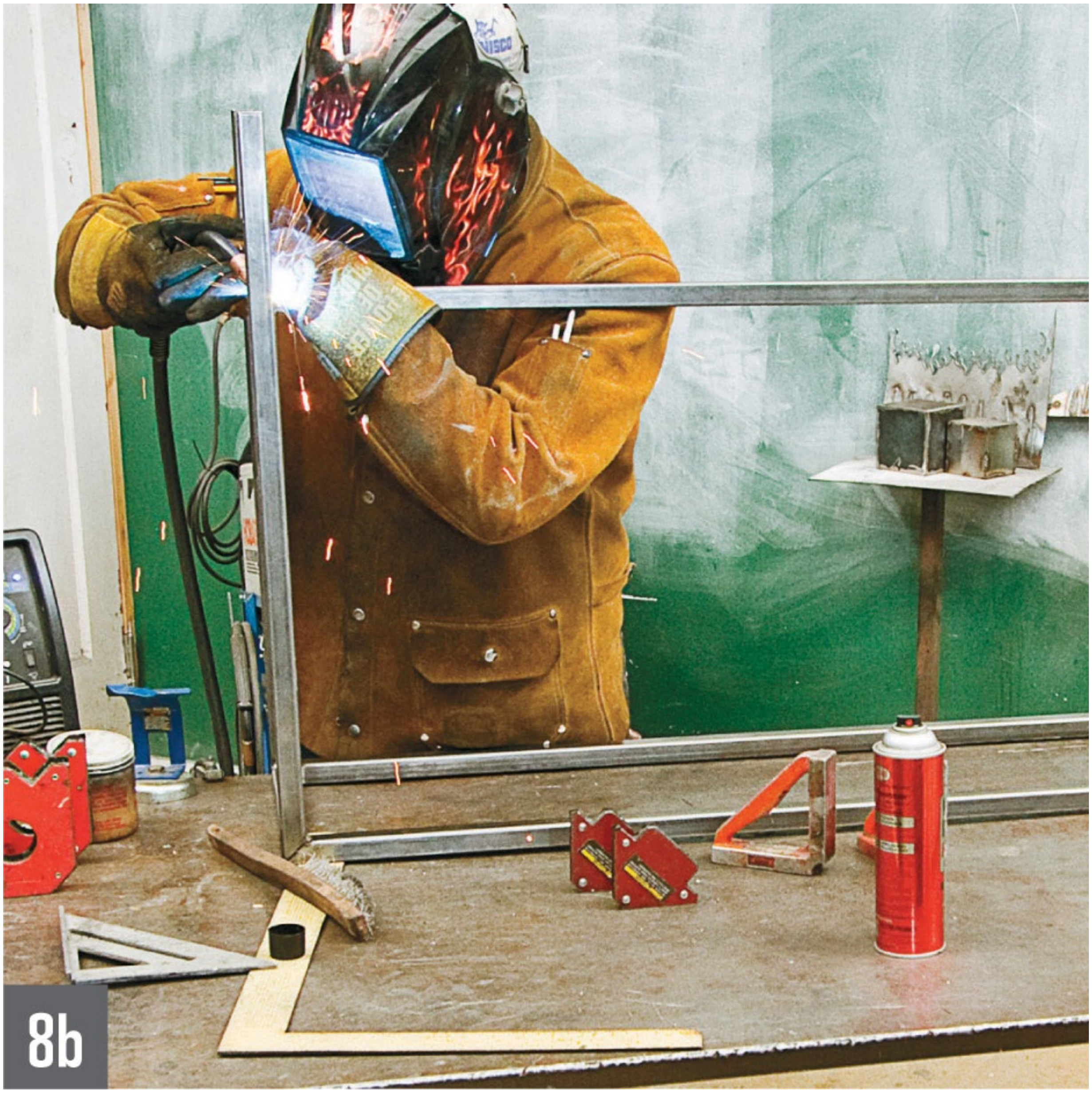

8b

With your system set for the thickness of the steel, maintain proper distance and angles while you weld.

Put yourself in the most comfortable position you can while welding large pieces.

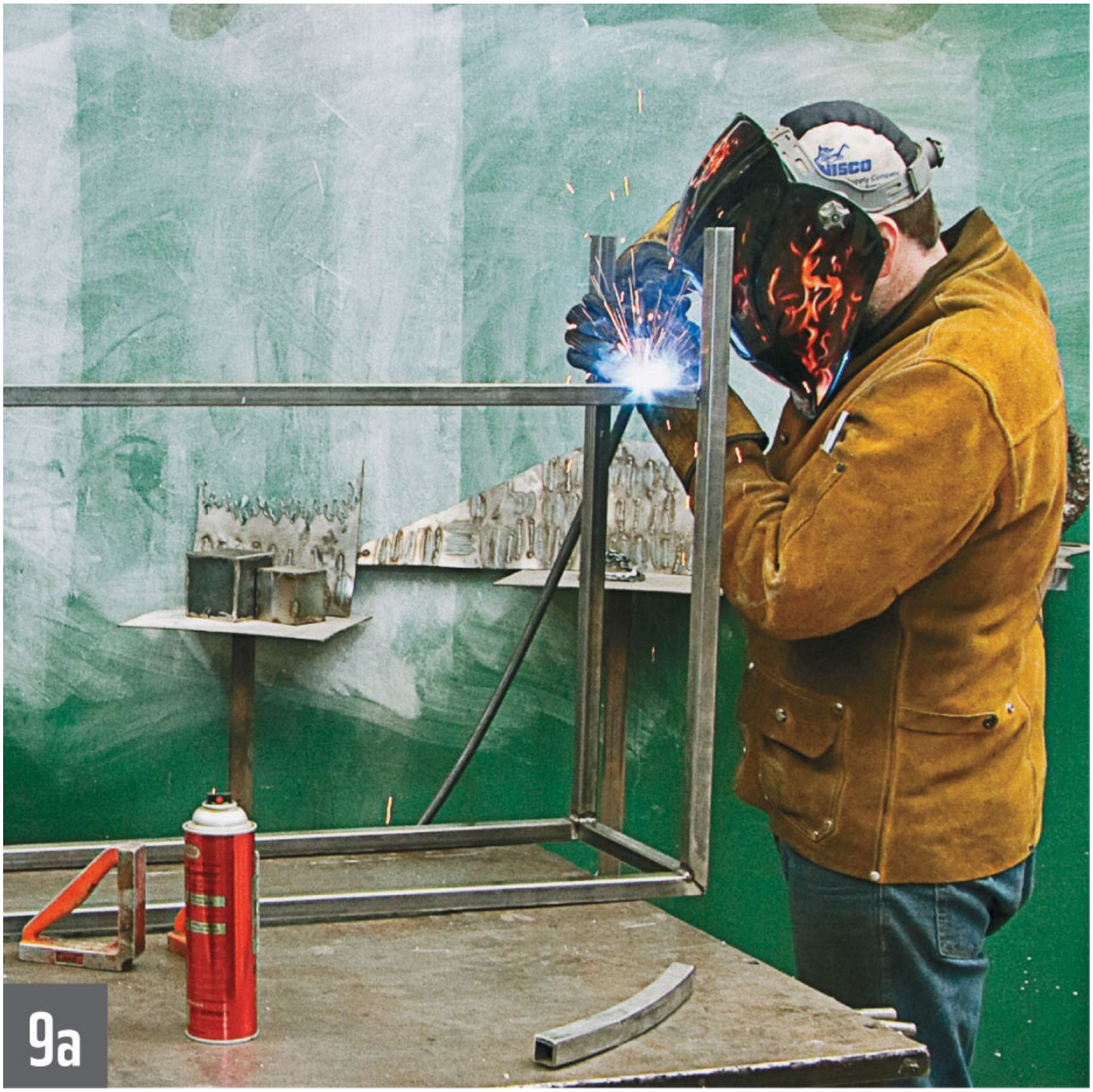
VISCO
Supply Company

9a

9b

Keep spreading out the welds as you go to keep the square tube from warping out of square.

10

This is the last weld on the frame.

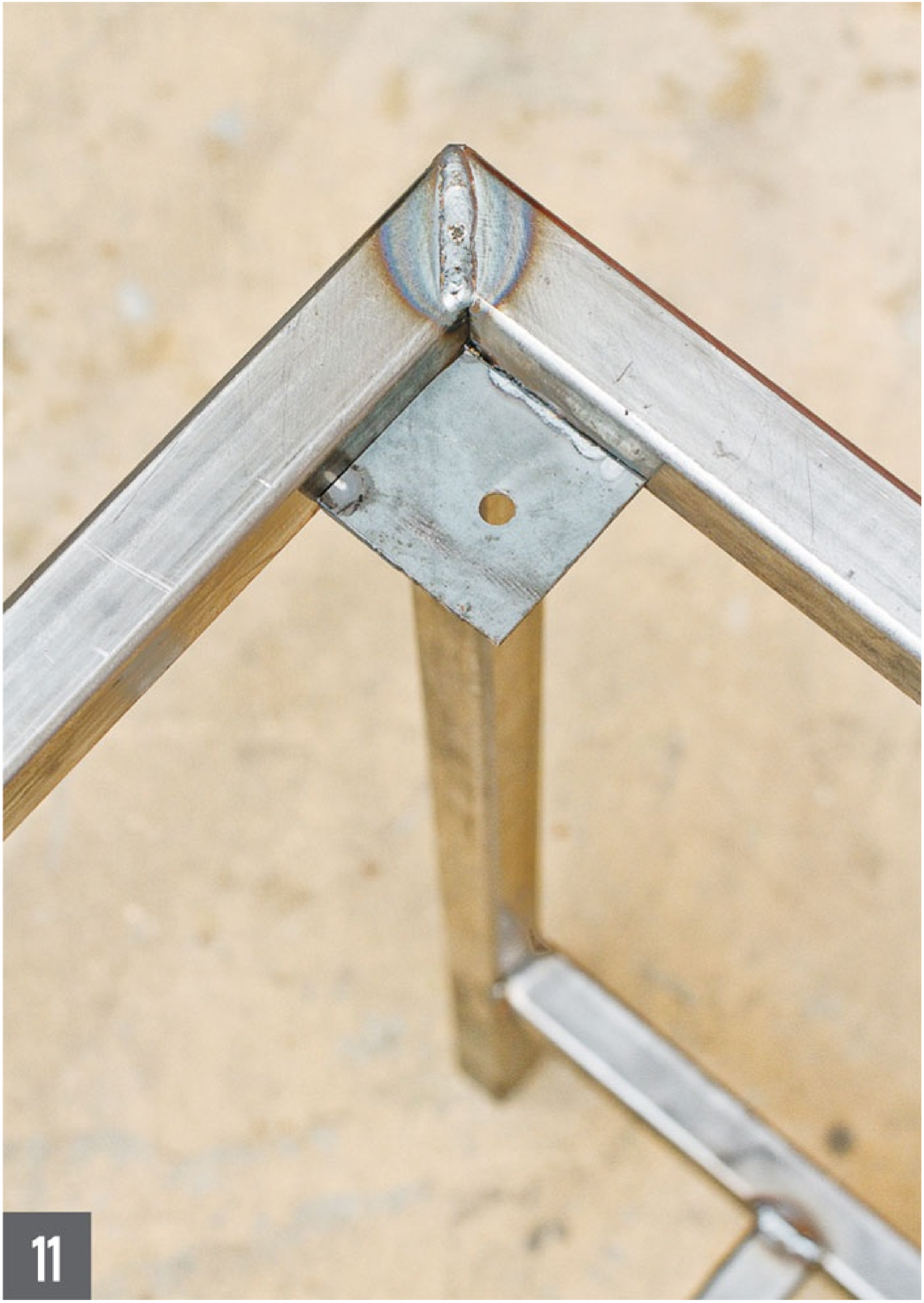

11

Tack and weld in the mounting brackets that will receive the wood top.

Now that it’s all welded up, it’s time to fit the top. You have a lot of options for the top as anything from glass to butcher block to any of a wide variety of woods will look great. What I have chosen this time around is a nice piece of birch that I will dry-fit before sanding, prepping, staining, and then clear-coating.

There are a lot of ways to mount wood in or on a table frame. The system I’m going to use for this table is six 3/8-inch-inch (10 mm)-long wood screws driven through six tabs made from 18 ga steel welded to the inside of the frame.

Make spacers the same thickness as the material for the top that you have chosen. Pre-drill and then cut six 2-inch (5 cm)-square pieces of 18 ga for the tabs. Grind down the welds on the top of the table so they are flush to the tube, creating a flat surface. Now flip the table over onto your welding table. Using your spacer, weld four of the tabs into the corners and two into the center using fillet welds.

Now that the tabs are in place, fit the wood top in the frame to see if adjustments need to be made; then stain or finish it the way you would like. I’m going to grind down the rest of the welds and clear-coat the frame with a clear metal paint such as Rust-Oleum or Krylon. Then I’ll give the birch a single coat of cherry stain and finish off with two or three coats of polyurethane.

Whatever you choose to finish with, I hope you learned a lot from this project. But most of all, I hope you have enjoyed making it.

After the brackets are welded in place, you can fit in your wood top. A few coats of clear on the frame and a few coats of stain on the wood and this table is done.

Gallery

Sculpture

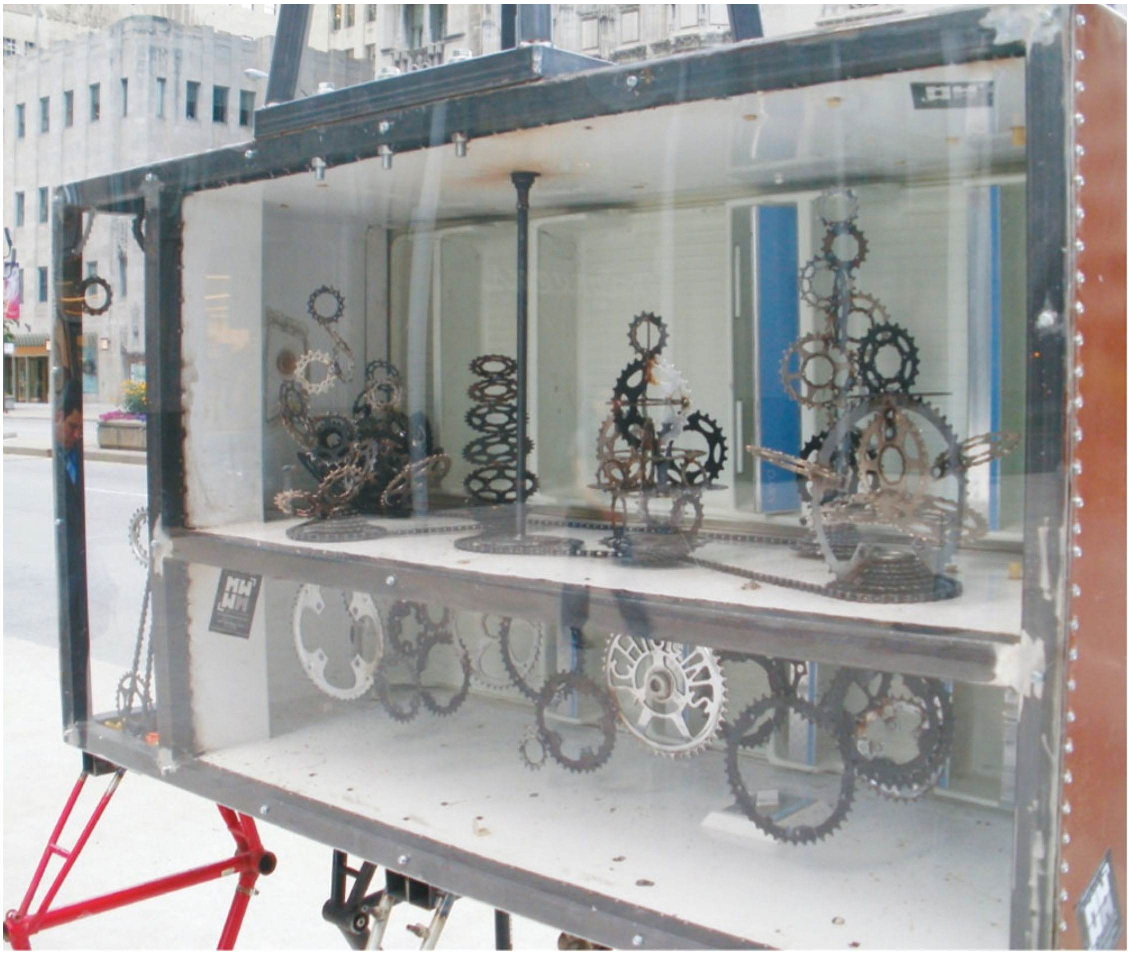

Photo by Stephen Blake Christena

Photo by Brad Cowan

Photo by Stephen Blake Christena

Furniture

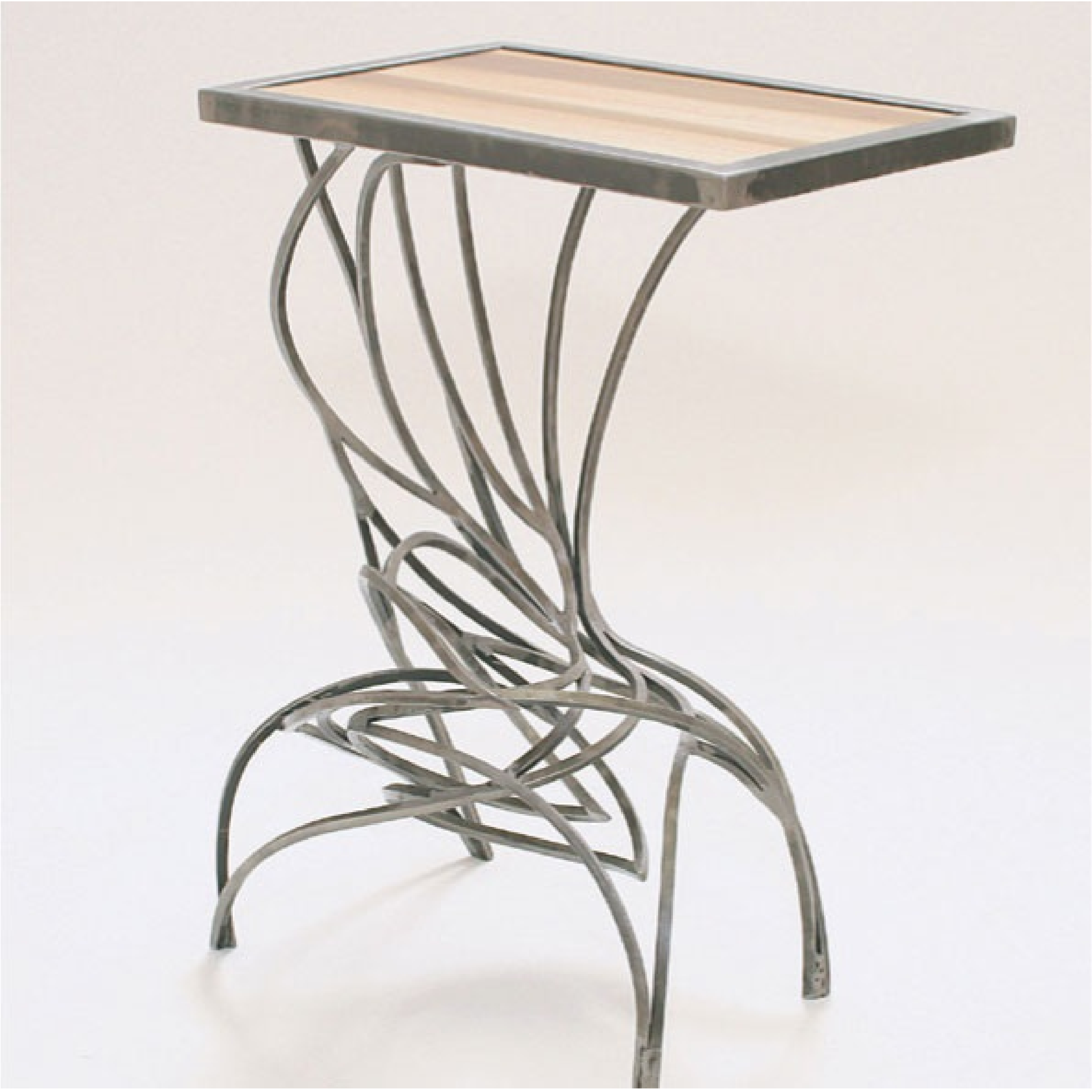

Photo by Casey Y. Spring

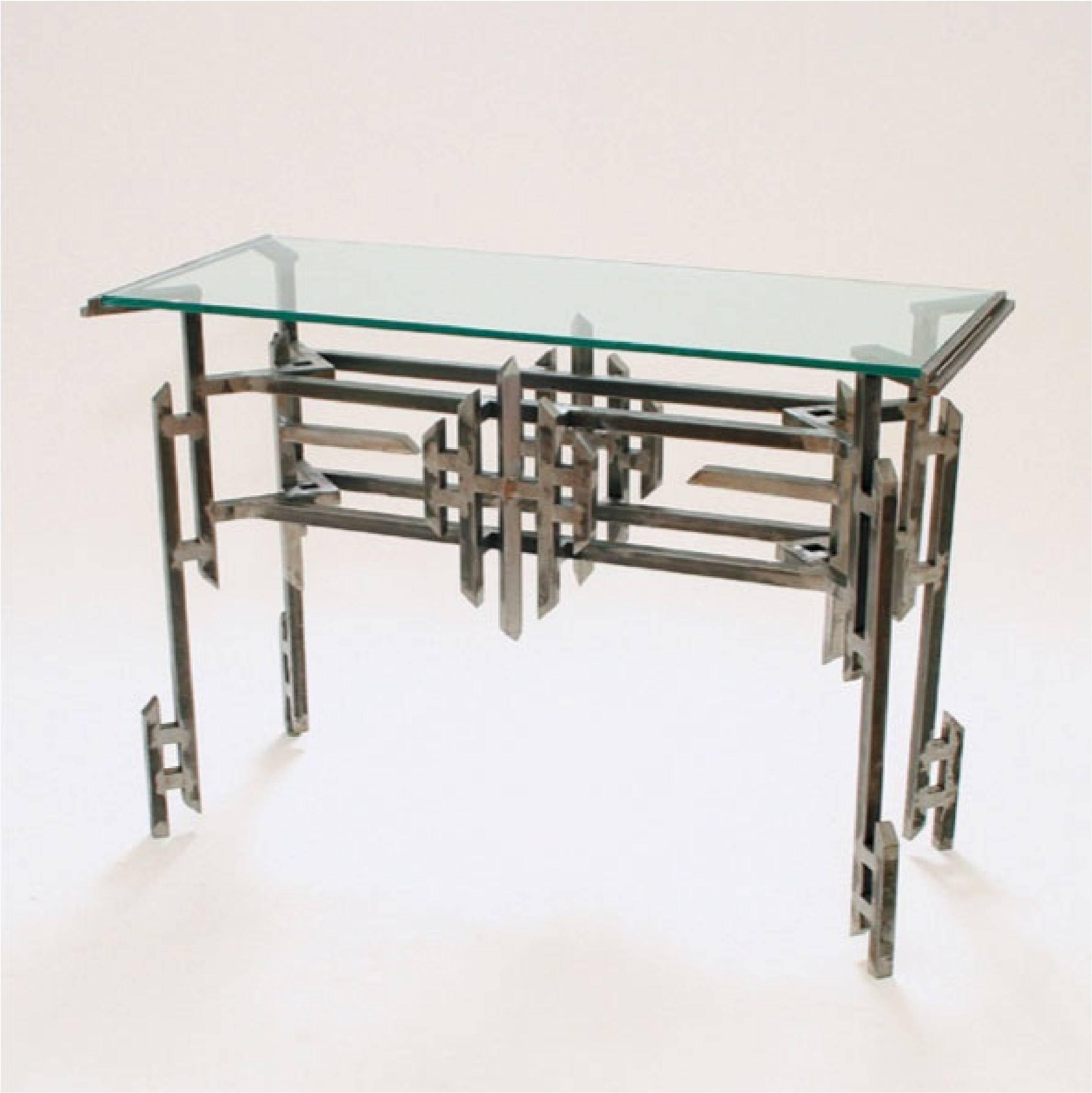

Photo by Casey Y. Spring

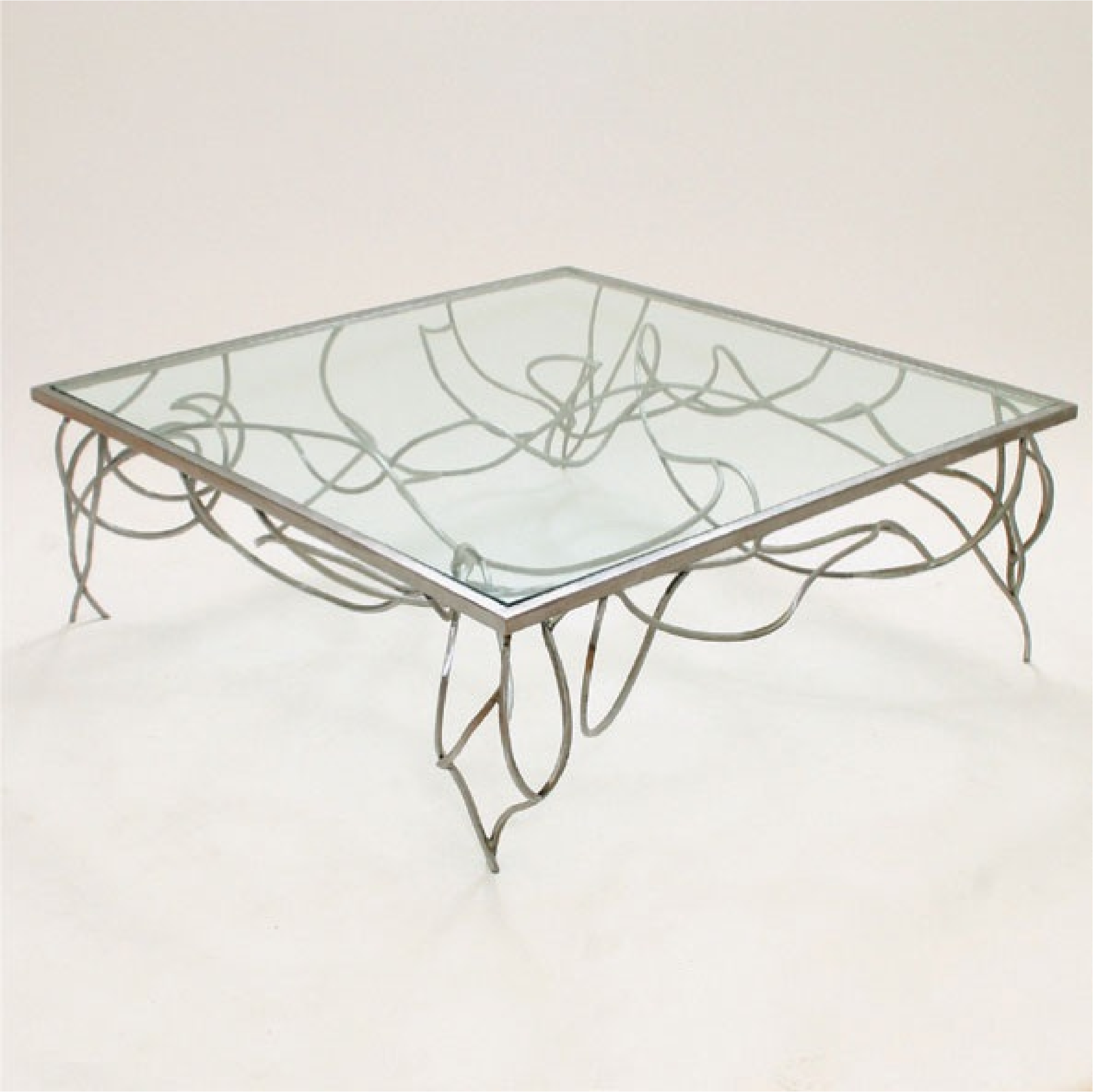

Photo by Casey Y. Spring

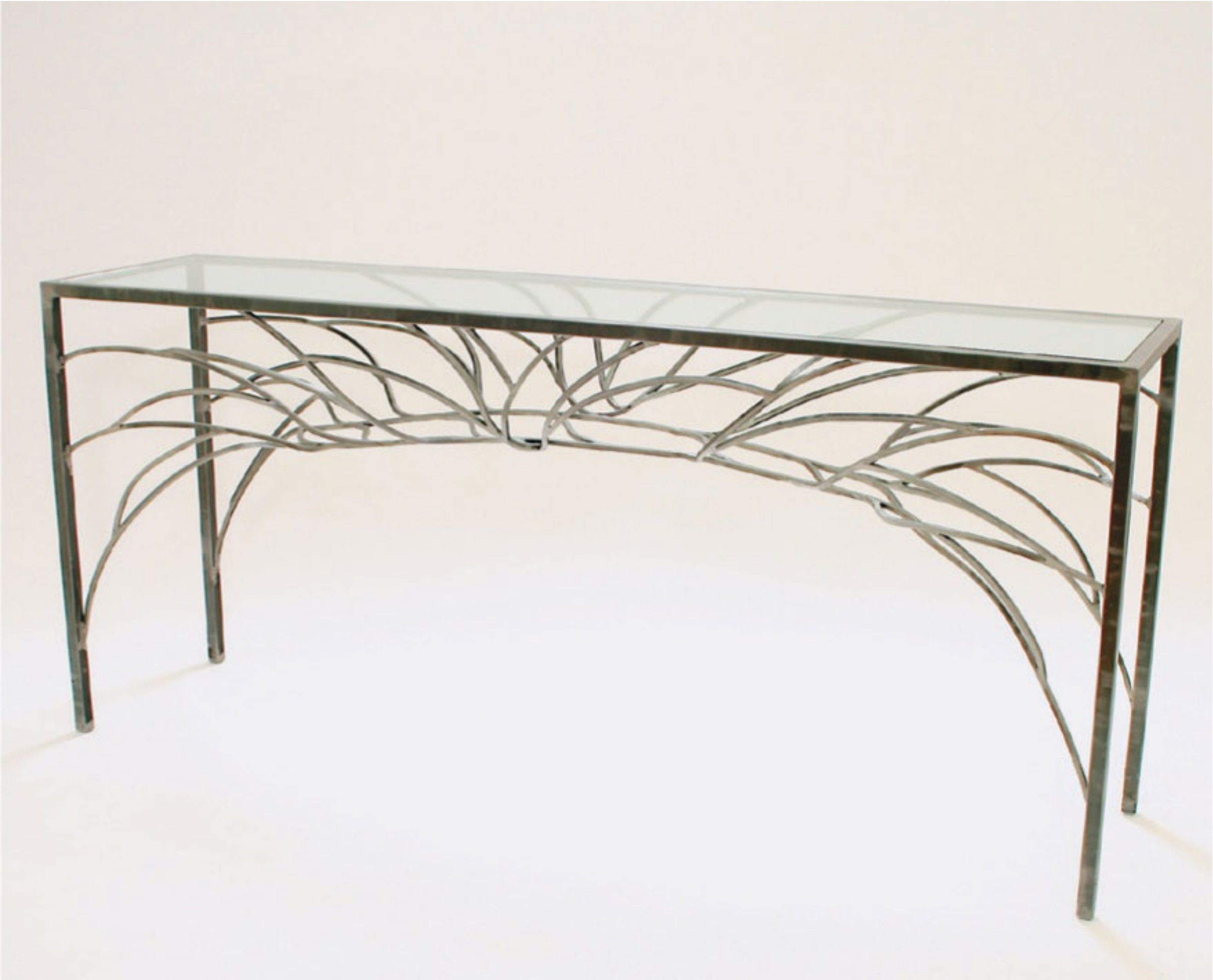

Photo by Casey Y. Spring

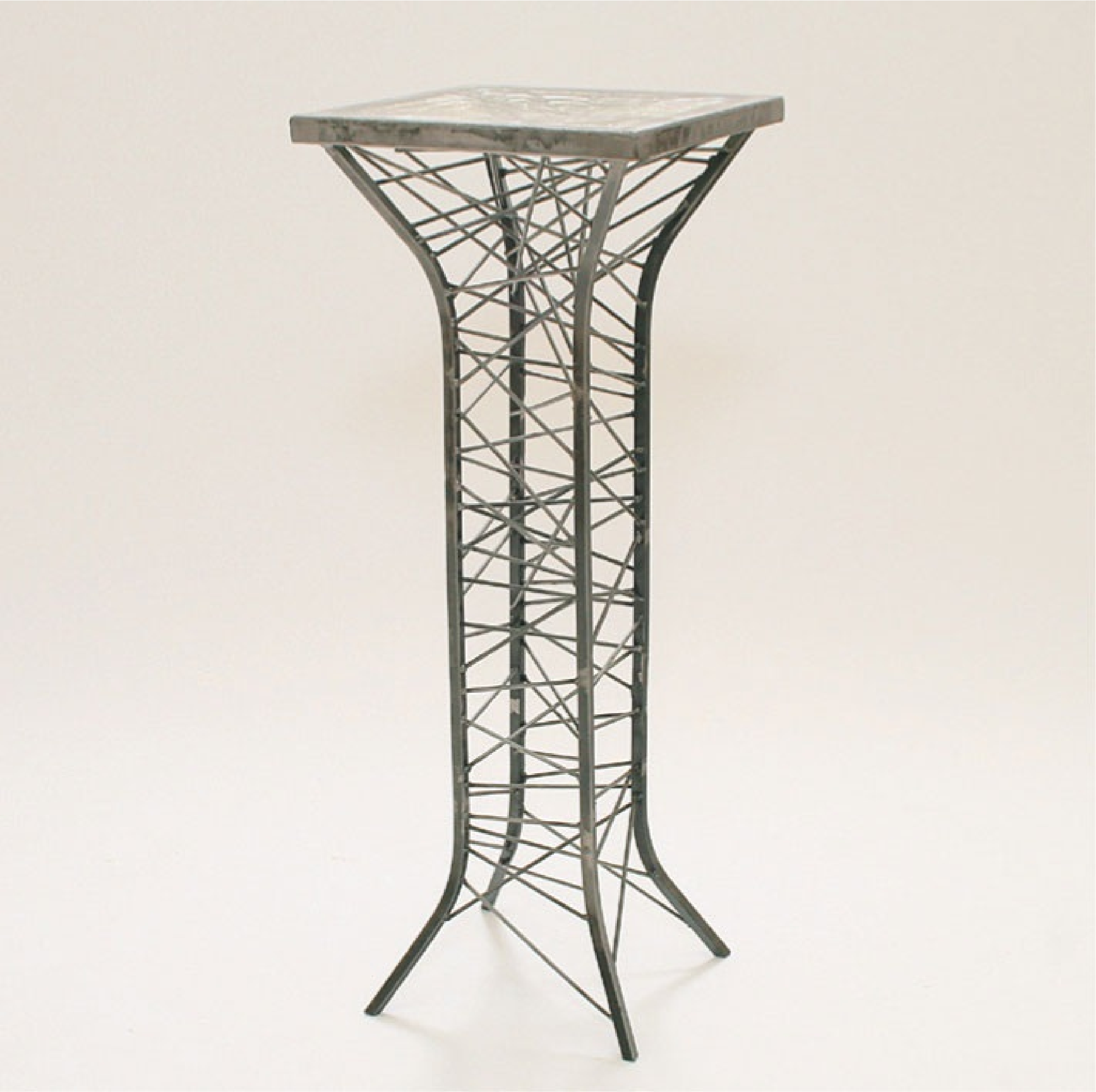

Photo by Casey Y. Spring

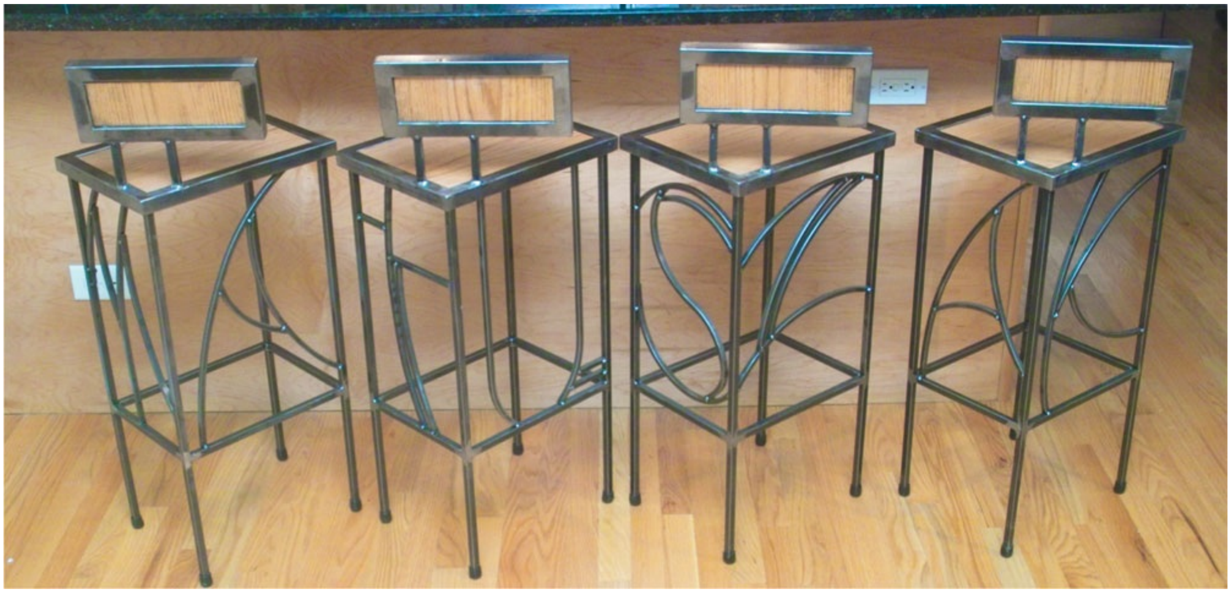

Photo by Stephen Blake Christena

Architectural Elements

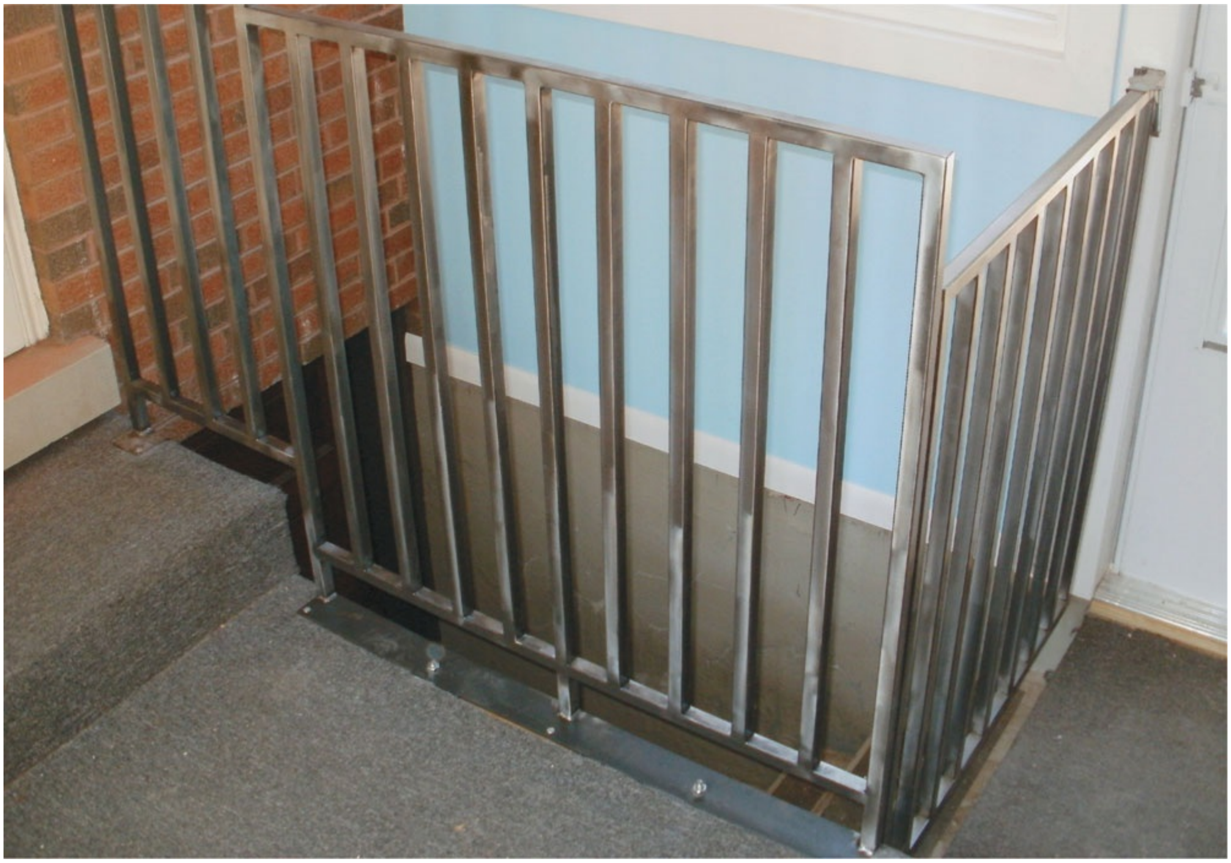

Photo by Stephen Blake Christena

Photo by Stephen Blake Christena

Photo by Stephen Blake Christena

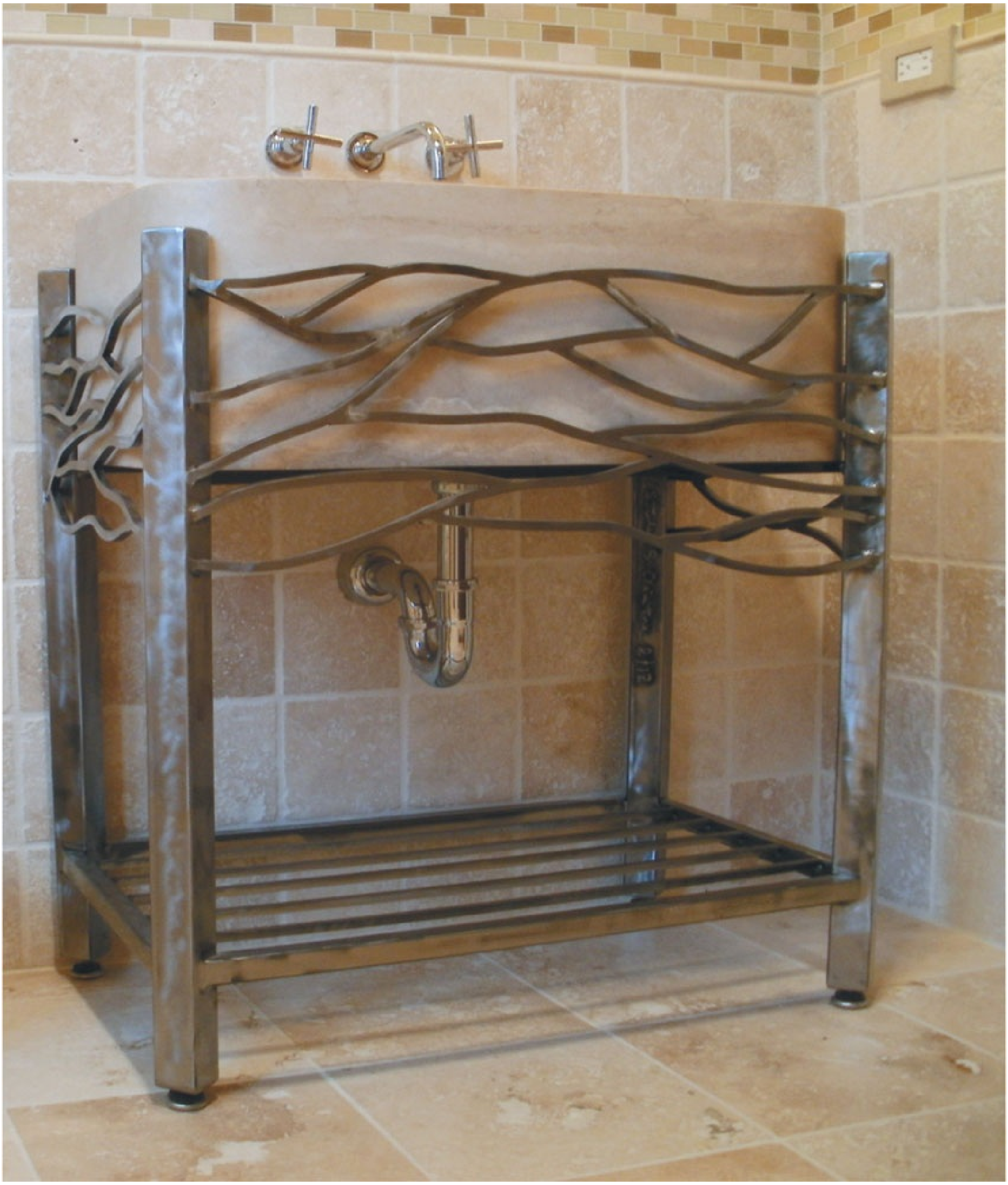

Photo by Stephen Blake Christena

Glossary

Alloy—This is a substance composed of two or more metals or a metal and other nonmetal material.

Arbor—This is the threaded shaft of a grinder or axis that holds, turns, or supports a rotating cutting or grinding wheel.

Arbor Nut— This is the threaded top piece that screws onto the arbor in order to secure a disk to the arbor. (Also see *flange*).

Arc Welding—These are the welding processes that use an electric arc to produce heat and melt and fuse the base metals. This term is often used to refer specifically to shielded metal arc welding.

Backfire and Flashback—This is the combustion taking place inside the oxyacetylene torch that creates a loud pop or explosion.

Base Metal—This is the work piece metal that is to be joined or cut.

Bead—This is the welded seam created by the puddle between work pieces that have been joined with welding.

Bevel—This means to remove edge materials so that the adjacent surfaces no longer form a right angle.

Burn Back—This is when the wire electrode filler metal becomes too short and/or overcharged and melts the wire into the contact tip of a GMAW or MIG welding gun.

Burn Though—This can happen during a weld; overheating metal to a critical point causes the material to no longer hold be able to its own weight, causing holes or an excess of material on the opposite side.

Butt Joint—This is a joint between two work pieces in the same plane.

Chamfer—This is the removal of a material’s edge, usually at a 45° angle, to the adjacent principal faces.

Consumable Electrode—This is an electrode that also serves as the filler material.

Corner Joint—This is a joint between work pieces that meet at a right angle.

Cutting Tip—This converts an oxyfuel welding torch into an oxyfuel cutting torch.

Dimes on the Side—This refers to the formation of a pattern in the bead to give the appearance of a stack of coins that has fallen onto its side. It is most commonly associated with TIG welding. It is an indication of the control of a molten metal puddle to create ripples that gives the appearance of a stack of dimes, a stack of nickels, or row of coins.

Direct Current, Electrode Negative (DC+)—This refers to direct current welding where the electrode is negative and the work piece is positive.

Direct Current, Electrode Positive (DC–)—This refers to direct current welding where the electrode is positive and the work piece is negative.

Double Arcing—This is an unintentional secondary arc separate from the main electrode arc.

Draw—This is the pace of the arc moving through the weld zone.

Drop—This is a piece of material cut away from a whole, leftover metal, scrap.

Dross—This is solidified metal byproduct that was once molten material forced through the kerf and then fused to the back edge by an oxyfuel or plasma cutting system.

Duty Cycle—This is the amount of continuous time in a 10-minute period that an arc welder can run before it needs to cool down. It is expressed as a percentage at given amperage output.

Edge Joint—This is a joint between parallel work pieces.

Electrode—This is the conductive element that makes the connection with the work piece to create an electrical arc.

Filler, Filler Metal or Fillet—This is supplied metal supplied in the form of a welding rod, sometimes flux coated, melted by an arc or flame into a joint between components to be joined.

Flange—This is a base piece fitted over the arbor for securing disk to the arbor. It is normally unthreaded.

Flashing—This is a short duration of time in which the eyes are exposed to the arc light.

Flux—This is a chemical compound that produces cleaning action and reduces the formation of oxides when heated.

Flux Core, Flux Core Wire—This is a consumable electrode wire that contains a chemical compound that produces a cleansing ionizing environment for an electrical arc to exist to reduce oxides.

Gas Metal Arc Welding (GMAW) a.k.a. Metal Inert Gas Welding (MIG)—This is a semiautomatic arc welding process using a wire electrode, which also is the filler material. An inert gas flows over the weld area to shield the molten metal from oxygen.

Gas Tungsten Arc Welding (GTAW) a.k.a. Tungsten Inert Gas Welding (TIG)—This is an arc welding process using a tungsten electrode, hand-held filler material, and an inert shielding gas.

Going Fishing—This is the natural tendency for a beginner welder to pull away from the weld during the weld in progress. Produces a poor weld.

Inclusions—These are elements and particles that can be introduced into the weld or puddle that create issues with the final weld.

Inert Gas—These are noble gases: the unreactive gaseous elements helium, argon, and carbon dioxide.

Kerf—This is the width and shape of a cut made by an instrument or tool.

Lap Joint—This is a joint between overlapping work pieces.

Media Blasting—This is to forcibly clean a surface with pressure using abrasives and forced air, also known as Sand or Abrasive blasting.

Off Gas—This is to create an exhaust by heating an element to expel gases.

Oxyacetylene Cutting—This is a cutting process using acetylene as the fuel and oxygen as the accelerant to increase flame temperatures to achieve cutting in to metals.

Oxyfuel Cutting—This is a cutting process using oxygen as an accelerant to increase the flame of a combustible compressed gas.

Plasma Cutting—This is an arc cutting process using a constricted arc flow produced through a small nozzle or orifice, from the electrode to the work piece, by compressed air or inert gas.

Plate—These are flat sheets of metal thicker than 3/16 inch (5 mm).

Puddle—This is the liquefaction of the base metal to form a molten metal pool at the point of heat. Molten pool filler metal is added to create a weld bead.

Sheet Metal—This is flat metal 3/16 inch (5 mm) or less in thickness.

Slag—This refers to oxidized impurities formed as a coating over the weld bead; it is waste material found along the bottom edge of an oxyfuel cut.

Spatter—These are small balls or droplets of molten metal ejected from the puddle during the production of a weld. It often fuses to metal.

Stick—This is a rod of filler metal covered in flux used in the arc welding process; the term is also used to describe a complete length of milled bar or tube metal.

Stick Out—This is the distance the filler metal electrode protrudes from the contact tip in GMAW or MIG.

Stick Welding—This refers to shielded metal arc welding.

Straight Draw—This is the motion of pulling the weld towards the person conducting the weld.

Tee Joint—This is a joint between work pieces meeting at right angle and forming a T shape.

Wire Feed—This is the manual or semiautomated rate in which the filler metal is pushed into a weld in progress.

Toe Out—This refers to what happens when proper speed or technique is not executed and gravity pushes down on the puddle to form a toe shape when it solidifies into a bead.

Travel Angle—This is the proper angle and orientation of the torch tip or gun nozzle moving through a cut or weld. It is also known as the Drag Angle.

Resources

For basic information on welding and welding equipment

Miller Electric
www.millerwelds.com

Strong Hand Tools and Build Pro
www.stronghandtools.com

Welding Tips and Tricks
www.weldingtipsandtricks.com

Hobart Brothers Company
www.hobartwelders.com

For continuing education and welding as a career
AWS - American Welding Society
www.aws.org

CDA Technical Institute for Commercial Dive Welding
www.commercialdivingacademy.com

The Fab School
www.thefabschool.com

Notice to Readers Read, understand, and follow safety precautions and rules when welding. In addition to general safety rules, also reference the material safety data sheets (MSDS) as provided by manufacturers, as well as the American National Standard (ANSI Z49.1) guidelines available here: www.aws.org/technical/AWS_Z49.pdf.

About the Author

Stephen Blake Christena is the owner of Midwest Metal Works in Chicago, IL. MW2 offers custom fabrication services and a learning center for hobbyists interested in MIG welding. Christena teaches small-size courses designed to cover the basics and explore students' projects and interests.

Originally from Flint, Michigan, Christena moved to Chicago in 1998 after

Originally from Flint, Michigan, Christella moved to Chicago in 1998 after graduating from Western Michigan University with a degree in art. During his time at WMU, his primary focus was directed toward metal sculpture, photography, and painting. He opened his first custom metal sculpture studio in August of 1998.

ARC ACADEMY

Learn It. Weld It. Make It.

www.arcacademy.com

MIDWEST

MW

WM

METALWORKS

Acknowledgments

In loving memory of my mother, Susan Johnson Christena, who passed before her time: you are still my inspiration for life, love, and creativity. I miss you dearly. And to my father, John M. Christena: you are and always have been my hero. Love you, dad.

I would like to thank my brother and sister, Jeff and Amy Christena, for their love and support and for being the wonderful people that they are. I would also like to thank Jerry and Irene Slusarenko: Without your generosity, support, and understanding, none of this could have been possible. Albert LaVergne, my college sculpture professor, thank you for being the most honest man in art. To all the teachers who helped to shape and guild my mind, and to all of my friends and family who make me who I am: Thank you. You know who you are.

“*I believe in teaching through example. A strong work ethic with good discipline in the pursuit of personal growth is the best way to discover your own voice. Good skills are like good grammar, they are the foundation of effective communication.*”

—Albert LaVergne

Index

acetylene gas, 58, 59, 60, 61
alloys, 70
aluminum, 16, 21, 73
American Welding Society, 37
angle grinders. *See also* equipment.
  anti-vibration gloves, 44
  blade guards, 44
  blocks, 45
  chamfer preparation with, 81–82, 83
  cut-off wheels, 46
  disk replacement, 83
  disk types, 42, 45, 46, 79
  flap disks, 46
  grinding disks, 46, 81
  grit scale, 46
  metal preparation with, 46, 79, 81
  name origin, 42
  pressure, 46, 82
  purchasing, 24
  safety, 42, 44, 45, 46
  setting up, 46
  sparks, 42
angle welds, 125
anti-vibration gloves, 44
aprons, 35
arc time, 24
arc welding
  double arcing, 17, 57, 68
  duty cycles, 24
  electricity and, 12, 16
  electrodes, 12, 14
  filler metal, 12, 14, 15

  - flux, 12, 15
  - ground clamps, 16
  - plasma, 12
  - shield gas, 15, 16
  - temperature, 14
- argon, 16
- autodarkening helmets, 34
- autofeed machines, 15, 24

- beads
  - definition of, 12
  - filler metal and, 15
  - metal thickness and, 56, 83, 94
  - puddle and, 86
  - TIG welding patterns, 22
- bench grinders, 80
- burns, 30
- butt joints
  - base metal elevation, 112
  - carbon burn, 112
  - chamfer for, 81
  - cleaning, 112
  - introduction to, 108
  - “pull” method and, 86
  - puddle control, 113–114, 115
  - “shark tooth” technique, 114
  - tacking, 111
  - weld zone, 111, 112

- cap rails, 74
- carbon burn, 83, 112
- carbon dioxide, 12, 15, 16
- carbon steel. *See* mild steel.
- chamfers, 81–82, 83
- chop saws. *See also* equipment.
  - blade replacements, 40–41

consumable blades, 40, 142
cutting angles, 39
pressure, 40, 142
setting up, 40
clamps. *See also* equipment.
grinder cutting and, 46
single corner clamp, 145
storage, 29
three-axis corner clamps, 144, 145, 157
variety of, 95
clothing, 35, 63
CnC (Computer Controlled) cutting machines, 58
cold roll steel, 72, 79, 138
comfort, 88
consumable blade chop saws, 40
corner jigs, 124–125, 146
corner joints, 106, 120
cross body welding, 127
Cube project
cutting, 130–131
grinding, 131
heat management, 135, 136
introduction to, 130
“key holing,” 136
marking, 135
measuring, 130
preparation, 132
stress test, 136
tacking, 132–134
warp prevention, 135
welding, 134–136
cut-off wheels, 46
cutting
chop saws, 39–40
(Computer Controlled) machines, 58
Cube project, 130–131

  oxy-acetylene torches, 58–61
  plasma cutters, 62–69
  Runner Table project, 153–155
  temperatures and, 40
  Welding Table project, 140–142

decorations. *See* ornamental metals.
diamond plate sheet metal, 74
DIN ratings, 34
double arcing, 17, 57, 68
downhill welding, 126
drag angle. *See* travel angle.
drag shields, 64, 68
dross. *See* slag.
duty cycles, 24
duty hours, 25

edge joints, 106, 122
electricity
  access to, 24, 28
  arc welding and, 12, 18
  duty cycle and, 24
  ground, 16
  MIG welding, 22, 32
  outlets, 28
  shocks, 32
  studio and, 28
  TIG welding, 21
  wiring, 28, 32
electrodes
  definition of, 12
  filler metal as, 14
  MIG welding, 14, 22
  stick welding, 14, 19
  TIG welding, 14
equipment. *See also* angle grinders; chop saws; clamps; oxy-acetylene torches;

plasma cutters.
  autofeed machines, 15, 24
  bench grinders, 80
  Computer Controlled (CnC) cutting machines, 58
  duty cycles, 24
  duty hours, 25
  electrical wiring and, 28, 32
  exhaust systems, 28
  filing cabinets, 29
  gas cylinders, 36–37, 54, 59
  gloves, 32, 44
  helmets, 32, 34, 89
  magnetic squares, 145
  media blasters, 80
  MIG snips, 57
  necessity considerations, 23
  organization, 29
  purchasing considerations, 23, 24
  researching, 25
  respirator masks, 28, 144
  safety symbols, 30
  soapstone, 57, 100, 135
  space requirements, 29
  stools, 88
  storage, 29
  tip dip, 57, 115
  trusted brands, 25
  usage considerations, 24
  used machines, 25
  welding carts, 37
  welding supply stores, 25
  wire brushes, 57, 79, 83, 100, 112, 118
exhaust systems, 28
expanded sheet metal, 74
eye protection, 34

fatigue, 88
ferrous metals, 70
filing cabinets, 29
filler metal
  arc and, 14
  bead and, 15
  definition of, 15
  electrode as, 14
  feeding, 15, 48
  ground clamps and, 51, 52
  melting point, 73
  metal types and, 73
  MIG welding, 14, 15, 22
  stick welding, 14, 15
  TIG welding, 20
fillet joints
  cleaning, 118
  definition of, 106
  gravity and, 116, 118
  introduction to, 106
  pattern, 119
  puddle control, 118
  tacking, 118
  uses of, 106
finger safety, 45
fire extinguishers, 31
fires, 31
first aid kits, 30
flap disks, 46
flux
  autofeed machines, 15, 24
  definition of, 15, 22
  impurities and, 15
  outdoor welding and, 16
  shield gas and, 15–16, 18
  MIG welding and, 22

  - polarity and, 49
  - slag and, 18
  - spatter and, 17–18
  - stick welding and, 15
  - ventilation and, 22
- fumes, 22, 28, 35, 144

- gases. *See also* shield gases.
  - acetylene, 58, 59, 60, 61
  - carbon dioxide, 16
  - cylinders, 36–37, 54, 59
  - ionization, 12, 16, 62
  - mixture ratios, 16
  - oxy-acetylene torches, 58–61
  - plasma, 62
  - regulators, 56
  - respirator masks for, 28
- gauges, 76, 77
- gloves, 32, 44
- GMAW (Gas Metal Arc Welding). *See* MIG welding.
- grinders. *See* angle grinders.
- ground clamps
  - arc welding and, 12, 16, 21
  - connecting to welder, 49
  - filler wire and, 51, 52
  - plasma cutters, 64
  - safety, 32, 53
  - used machines, 25
  - Welding Table project and, 138, 139
- GTAW (Gas Tungsten Arc Welding) *See* TIG welding.
- guns
  - angle of, 91–92, 98, 118
  - assembly of, 48
  - description of, 47

  - holding, 90, 98, 100, 101, 126
  - setting up, 48
  - trigger, 48

HAZ (heat-affected zone), 93, 94
helmets, 32, 34, 89
hot roll steel, 72

inclusions. *See also* slag.
  - chamfer and, 110
  - chemical removal of, 80
  - flux and, 15
  - gun angle and, 91
  - soapstone and, 57
  - shield gas flow and, 113
  - stick length and, 101
  - weld quality and, 78, 79
  - weld zone cleaning and, 112

jigs, 124–125, 146
joints
  - butt joints, 108, 110–115
  - edge joints, 106, 122
  - fillet joints, 106, 116–119
  - lap joints, 106, 122
  - corner joints, 106, 120

kerf, 62–63
“key holing,” 136

lap joints, 106, 122
lighting, 28

magnetic squares, 145
measurements
  - Cube project, 130

- metal thickness, 76, 77
- Runner Table project, 152, 153, 157–158, 159
- tacking and, 96
- Welding Table project, 137, 138, 140, 141, 149, 151

media blasting, 80

metals. *See also* mild steel.
- alloys, 70
- aluminum, 16, 21, 73, 138
- distributors, 75, 77
- ferrous, 70
- gauges, 76, 77
- iron, 70, 73
- mill scale, 46, 72, 80, 100
- oil coatings, 78
- ordering, 77
- ornamental, 74
- pricing, 75
- purchasing, 75, 77
- titanium, 16, 153
- zinc, 16, 35, 144, 150

Midwest Metalworks, 9

MIG gloves, 32

MIG snips, 57

MIG welding
- autofeed machines compared to, 15
- carbon byproducts, 83
- components of, 47
- ease of use, 23
- electricity and, 22, 32
- electrode, 14, 22
- filler metal and, 14, 15
- flux and, 22
- gas cylinder connection to, 54
- ground connection to, 49
- gun, 47–48
- hose connection to, 49

  - metal thickness settings, 55, 56
  - nozzle, 22, 48
  - polarity, 15, 49
  - safety, 32
  - setting up, 48
  - shield gas, 15, 22, 56, 101
  - spatter, 17–18
  - ventilation and, 22
  - voltage settings, 55, 99, 113
  - wire brushes for, 83
  - wire feed, 15, 22, 50–53, 55, 77, 99, 113
  - wire length, 101
  - wire tension, 52
- mild steel. *See also* metals.
  - angle grinder disks for, 46
  - carbon content, 70
  - cold roll, 72, 79, 138
  - Cube project, 130
  - distributors, 77
  - filler metal, 22, 50, 73
  - hot roll, 72
  - melting temperature, 14
  - uses of, 70
  - versatility of, 12
  - Welding Table project, 138
- mill scale, 46, 72, 80, 100
- "moon craters," 78
- muriatic acid, 80

- nozzles
  - accidental arcing with, 25
  - angle of, 91, 98, 126, 127, 134
  - connecting, 48
  - line of sight and, 89
  - MIG welding, 22, 48
  - shield gas and, 22, 48

spatter protection, 17, 57
  TIG welding, 19
  tip dip, 57, 115

ornamental metals, 74
outdoor welding
  flux core and, 16
  shield gas and, 16
  stick welding and, 16, 19
overhead welding, 34
overheating. *See* temperatures.
oxy-acetylene torches. *See also* equipment.
  acetone, 59, 60
  hoses, 60
  igniting, 61
  pressure settings, 60
  regulators, 59
  safety, 59–60
  slag, 61
  valve, 60, 61

patterns
  bead patterns, 22
  fillet joints, 119
  “nickels on the side,” 22, 109
  “shark tooth,” 114, 120
  “stack of dimes,” 22, 109
  TIG welding, 22, 109
pedal controllers, 19
perforated sheet metal, 74
plasma, 12
plasma cutters. *See also* equipment.
  cutting with, 64, 65–69
  DIN shades and, 34
  drag shields, 64, 68
  kerf, 62–63

  precision with, 23, 62
  prices of, 63
  slag, 64, 69
porosity
  chamfers and 110
  “moon craters,” 78
  preventing, 78, 79
preparation
  acetone cleaning, 78–79
  alcohol cleaning, 78–79
  angle grinders and, 79, 81
  bench grinders, 80
  benefits of, 78
  chamfers and, 81–82, 83
  chemical cleaning, 80
  Cube project, 132
  media blasting, 80
  muriatic acid, 80
  Runner Table project, 156
  small pieces, 80
  Welding Table project, 143
  wire brushes and, 79
projects
  Cube, 130–136
  Runner Table, 152–161
  Welding Table, 137–151
puddle
  butt joints, 113–114, 115
  controlling, 86, 109, 113–114, 115
  fillet joints, 118
  gravity and, 126
  heat management and, 93
  machine adjustments and, 113
  nozzle angle and, 98
  pulling, 86
  pushing, 86

- shield gases and, 91
- thin materials and, 89
- "toeing," 116, 117
- uphill welding, 126
- weld speed and, 117

recrystallization temperatures, 72
respirator masks, 28, 144
rhythm, 86
Runner Table project
- cutting, 153–155
- introduction, 152–153
- marking, 153, 154
- materials, 153
- measurements, 152, 153, 157–158, 159
- miter cuts, 153, 154
- painting, 161
- preparation, 156
- surface, 161
- tacking, 156–158
- welding, 159–161
- wood mounting, 161

safety
- angle grinders, 42, 44, 45, 46
- burns, 30
- chop saw blade replacement, 40
- clothing, 35, 63
- electric shocks, 32
- eye protection, 34
- fingers, 45
- fire extinguishers, 31
- first aid kits, 30
- flame-retardant materials, 26
- fumes, 35
- gas cylinders, 36–37

  - ground clamps, 32, 53
  - helmets, 32, 34, 89
  - metal temperatures, 30
  - MIG gloves, 32
  - MIG welding, 32
  - muriatic acid, 80
  - overhead welding, 34
  - oxy-acetylene torch, 59–60
  - plasma cutters, 63
  - respirator masks, 28, 144
  - scalp protection, 34–35
  - shield gases, 28
  - shoes, 35
  - spark fires, 31
  - spatter prevention, 57
  - spatter protection, 18
  - studio considerations, 26
  - symbols, 30
  - tip dip, 57, 115
  - trusted brands and, 25
  - UV burns, 32, 35
  - ventilation, 22, 26, 28, 35
  - water buckets, 31
  - welding blankets, 26
- saws. *See* chop saws.
- scalp protection, 34–35
- “shark tooth” technique, 114, 120
- shield gases. *See also* gases.
  - advantages of, 16
  - argon, 16
  - carbon burn and, 112
  - conductivity and, 12
  - chart for, 55
  - connecting, 54
  - flux and, 15–16
  - flux-core and, 22

- gun angle and, 91–92, 98
- ionization and, 12
- line of sight and, 16
- MIG welding and, 15, 22, 101
- mixtures, 16, 55
- nozzle and, 22, 48
- outdoor usage, 16
- pressure settings, 22, 56, 93, 111, 112, 113
- puddle and, 91
- regulators, 56
- safety, 28
- slag prevention with, 16
- spatter and, 57
- stick length and, 101
- TIG welding and, 19
- tip dip and, 57, 115
- vertical welding and, 126

shocks, 32
shoes, 35
sight. *See* vision.
single corner clamps, 145
slag. *See also* inclusions.
- removing, 15, 69
- shield gas and, 16
- stick welding and, 15, 18

SMAW (Shielded Metal Arc Welding). *See* stick welding.
soapstone, 57, 100, 135
spatter
- definition of, 17
- prevention of, 17–18, 57, 115
- safety, 32, 34–35
- stick welding and, 17–18
- table surface and, 139
- TIG welding and, 20
- tip dip and, 57, 115

stainless steel

  - alloys, 70
  - argon and, 16
  - brushes, 57, 79, 83, 100, 112, 118
  - table and, 138
  - TIG welding and, 19, 20
- stance, 88
- stick welding
  - electricity, 18
  - electrodes, 14, 19
  - filler metal, 14, 15
  - flux, 15
  - outdoor welding and, 16, 19
  - slag and, 15, 18
  - spatter and, 17–18
- studio
  - basement as, 28
  - electricity, 28
  - equipment storage, 29
  - exhaust systems, 28, 35
  - first aid kits, 30
  - flame-retardant materials in, 26
  - garage as, 26
  - layout, 29
  - lighting, 28
  - organization, 29
  - safety considerations, 26
  - shed as, 26
  - space requirements, 26, 29
  - spark fires, 31
  - ventilation, 22, 26, 28, 35

- tacking
  - butt joints, 111
  - Cube project, 132–134
  - definition of, 96
  - fillet joints, 118

  - heat control and, 98
  - measurements and, 96
  - mistake correction and, 96
  - Runner Table project, 156–158
  - warp prevention and, 96
  - Welding Table project, 147–148

- techniques
  - butt joint weld, 110–115
  - comfort, 88
  - corner joints, 120–121
  - cross body welding, 127
  - draw, 99–103
  - edge joints, 122
  - fillet weld, 116–119
  - gun angle, 91–92, 98
  - hand position, 90
  - heat control, 93–94, 98
  - jigs, 124–125
  - lap joints, 122
  - pattern techniques, 109
  - puddle control, 86
  - rhythm, 86
  - “shark tooth” technique, 114, 120
  - sound cues, 101, 103
  - stance, 88
  - tacking, 96–97, 98
  - temperature control, 93, 98
- temperatures. *See also* warping.
  - arc welding, 14
  - burns and, 30
  - cutting and, 40
  - duty cycle and, 24
  - heat control, 93–94, 98, 135, 136
  - “key holing,” 136
  - melting temperatures, 14, 73
  - metal rolling, 72

  oxy-acetylene torches, 58, 59
  plasma cutters, 62
  recrystallization, 72
  safety and, 30
  warping and, 94, 95, 96, 97, 135
textures, 79
three-axis corner clamps, 144, 145, 157
TIG welding
  bead patterns, 22
  costs of, 21
  electricity, 21
  electrode, 14
  filler metal, 20
  joint welding with, 109
  nozzle, 19
  pedal controller, 19
  shield gases, 19
  spatter, 20
  weld patterns, 109
tip dip, 57, 115
titanium, 16, 153
travel angle, 91, 126, 134

uphill welding, 126
UV burns, 32, 35

ventilation, 22, 26, 28, 35
vertical welding, 126
vision, 16, 89

warping. *See also* temperatures.
  causes of, 93, 94, 136
  clamps and, 95
  plasma cutters and, 62
  prevention of, 95, 96, 97, 135, 139, 160
  tacking and, 56, 96, 97

  thin metals and, 83, 94, 135
websites
  American Welding Society, 37
  metal distributors, 75
welding blankets, 26
welding carts, 37
Welding Table project
  casters, 138, 140, 143–144, 149–150, 151
  conductivity, 138, 139
  considerations, 138–139
  corners, 144–146
  cutting, 140–142
  frame welding, 149–151
  ground clamping to, 138, 139
  height considerations, 138
  introduction to, 137
  legs, 140
  marking, 142
  materials, 138, 139
  measurements, 137, 138, 139, 140, 141, 149, 151
  miter cuts, 140, 141
  mobility, 138, 143–144
  painting, 151
  preparation, 143
  size considerations, 138
  surface area, 138–139
  surface welding, 151
  tacking, 147–148
wire brushes, 57, 79, 83, 100, 112, 118
wire feed, 50–53, 55
wire tension, 52

zinc, 16, 35

Also Available

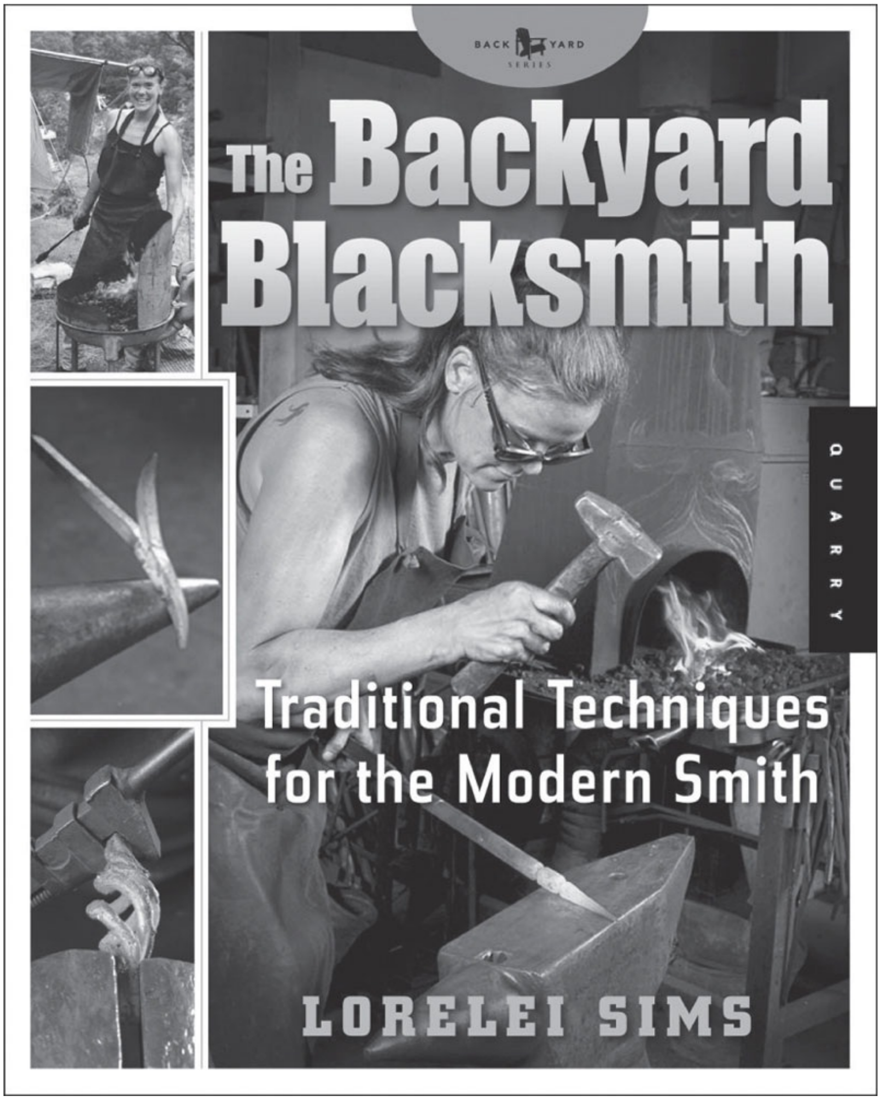
BACK YARD SERIES
The Backyard Blacksmith
Traditional Techniques for the Modern Smith
QUARRY
LORELEI SIMS

The Backyard Blacksmith
978-1-59253-251-3

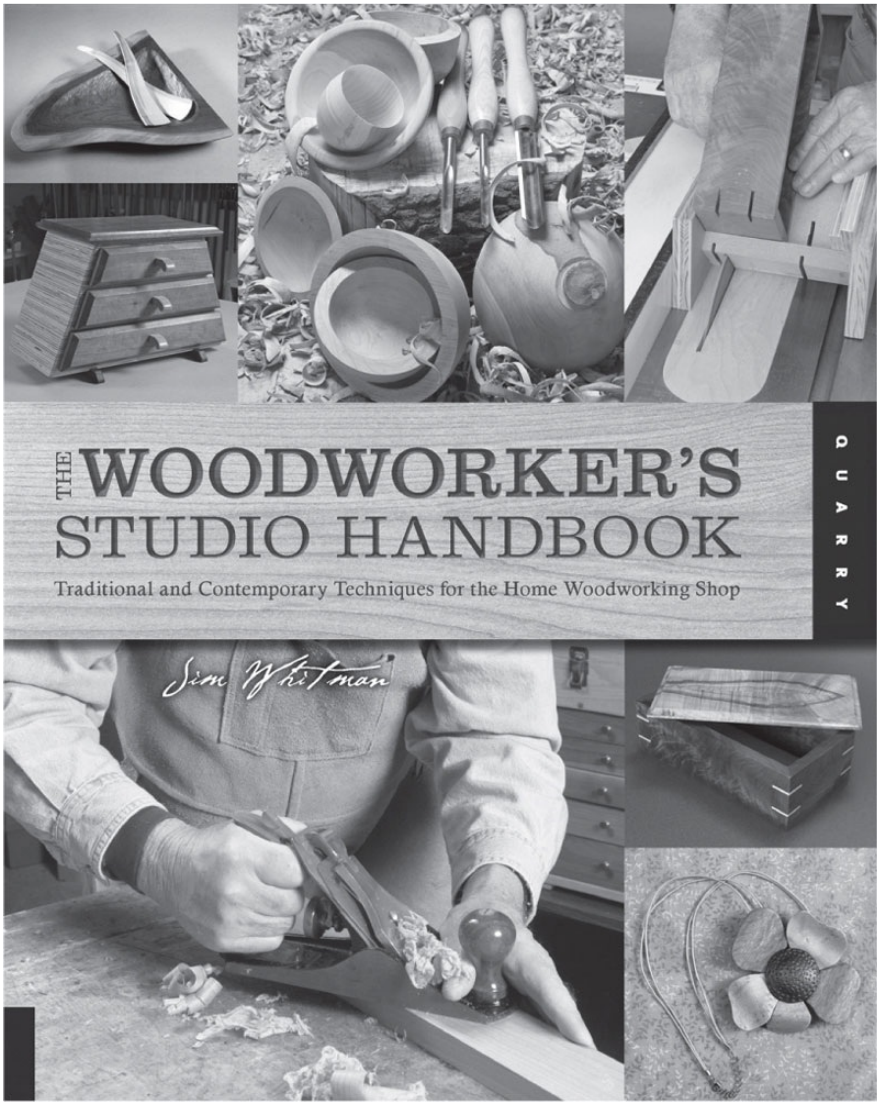

The Woodworker's Studio Handbook
978-1-59253-758-7

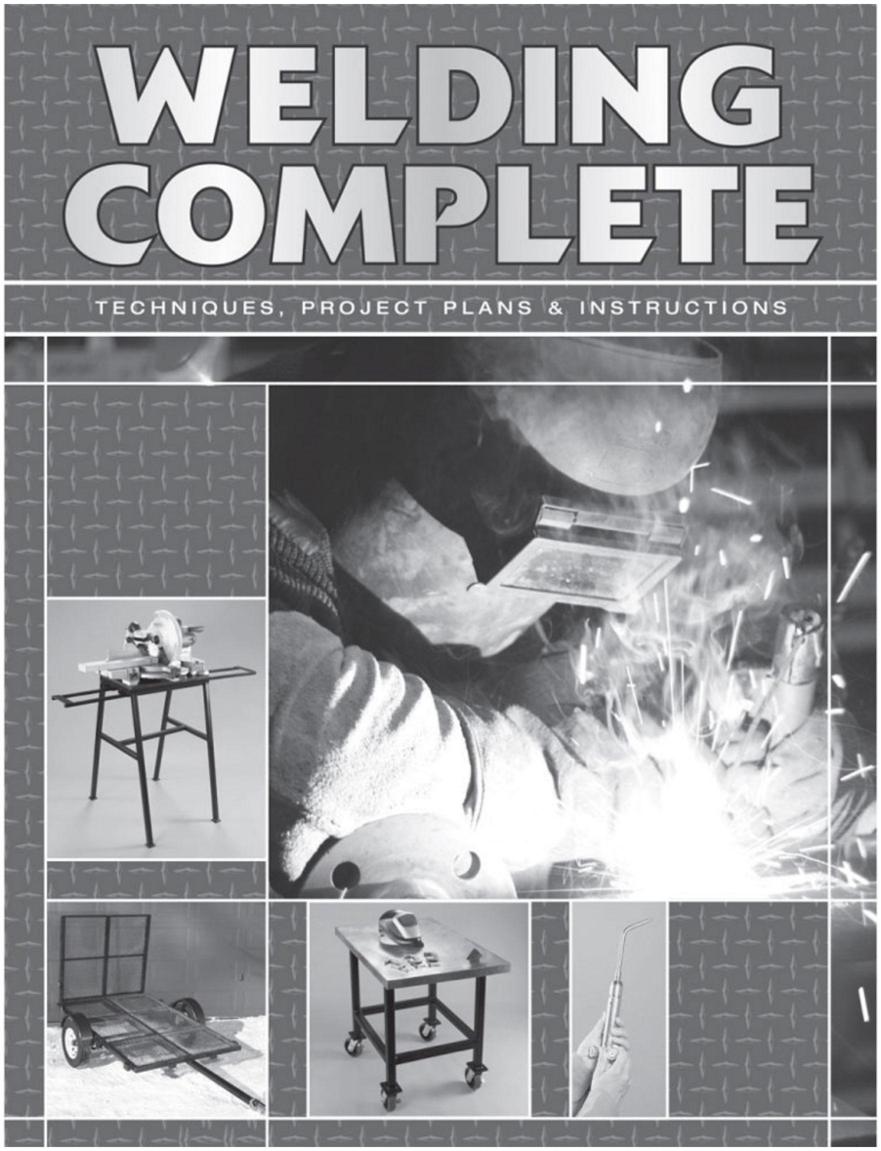
WELDING
COMPLETE
TECHNIQUES, PROJECT PLANS & INSTRUCTIONS

Welding Complete
978-1-58923-455-0

Dedicated to my parents, John and Susan Christena: I am a true cultivation of all that you have taught me, guided me, and blessed me with. You are my heroes.

Notice to Readers

Welding is a dangerous activity. Failure to follow safety procedures may result in serious injury or death. This book provides useful instruction, but we cannot anticipate all of your working conditions or the characteristics of your materials and tools. For safety, you should use caution, care, and good judgment when following the procedures described in this book. Consider your own skill level and the instructions and safety precautions associated with the various tools and materials shown. The publisher cannot assume responsibility for any damage to property or injury to persons as a result of misuse of the information provided.

Make:

Getting Started with Sensors

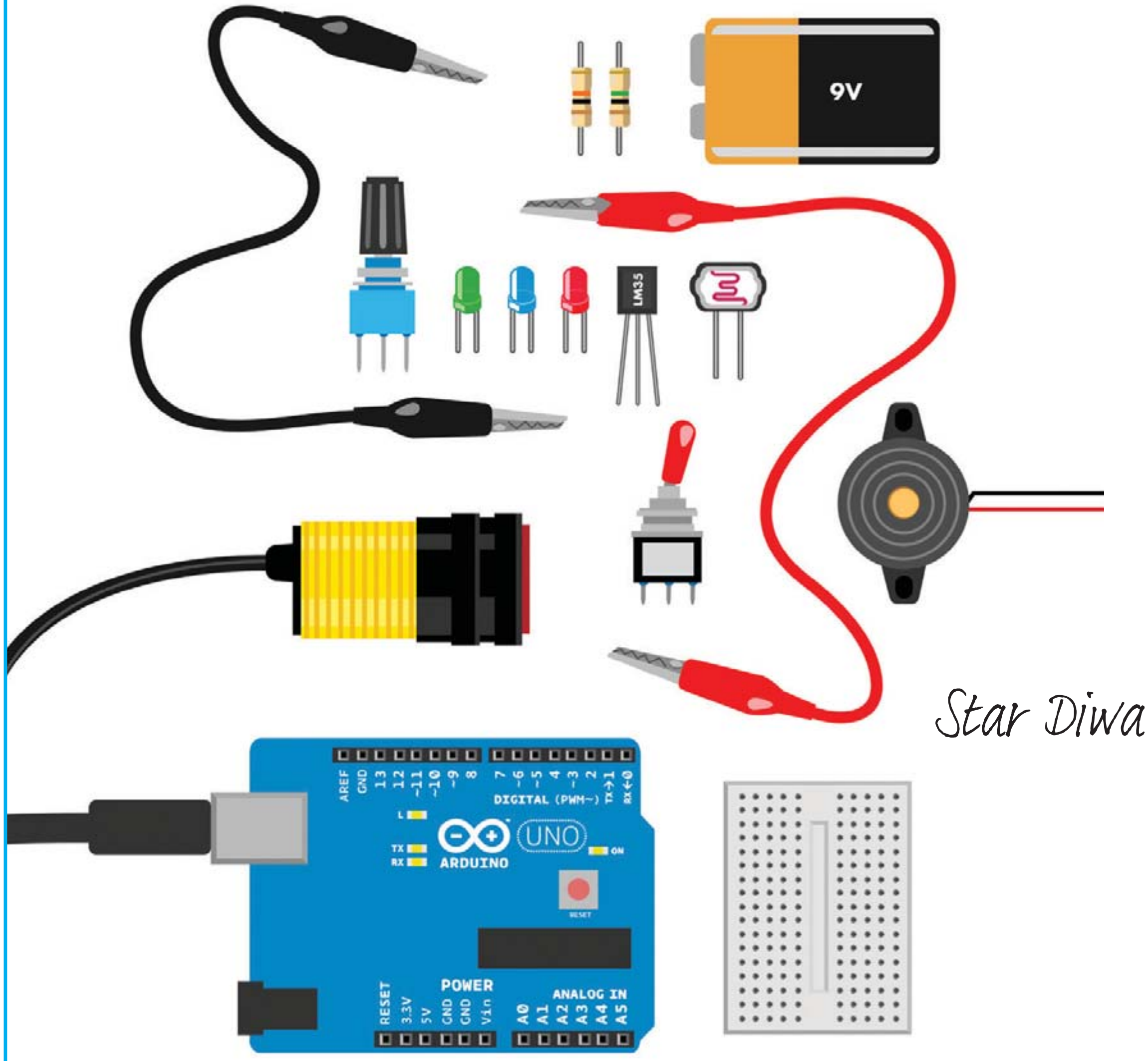
9V
LM35
DIGITAL (PWM~)
UNO
ARDUINO
ON
RESET
POWER
ANALOG IN
Star Diwa

Measure the World with Electronics, Arduino, and Raspberry Pi

Kimmo Karvinen & Tero Karvinen

Getting Started with Sensors

Kimmo Karvinen and Tero Karvinen

MAKER MEDIA™
SEBASTOPOL, CA

Getting Started with Sensors

by Kimmo Karvinen and Tero Karvinen

Copyright © 2014 Kimmo Karvinen, Tero Karvinen. All rights reserved.

Printed in the United States of America.

Published by Maker Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

Maker Media books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*http://my.safaribooksonline.com*). For more information, contact O'Reilly Media's corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

Editors: Brian Jepson, David Scheltema, and Emma Dvorak
Production Editor: Nicole Shelby
Copyeditor: Sonia Saruba
Technical Editor: Philipp Marek
Proofreader: Jasmine Kwityn
Cover Designer: Brian Jepson
Interior Designer: Nellie McKesson
Illustrator: Kimmo Karvinen
Photographer: Kimmo Karvinen
Cover Art: Kimmo Karvinen
Technical Advisor: Paavo Leinonen

August 2014: First Edition

Revision History for the First Edition:

2014-08-06: First release

See *http://oreilly.com/catalog/errata.csp?isbn=9781449367084* for release details.

Make:, Maker Shed, and Maker Faire are registered trademarks of Maker Media, Inc. The Maker Media logo is a trademark of Maker Media, Inc.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and Maker Media, Inc. was aware of a trademark claim, the designations have been printed in caps or initial caps.

While every precaution has been taken in the preparation of this book, the publisher and authors assume no responsibility for errors or omissions, or for damages resulting from the use of the information contained herein.

ISBN: 978-1-449-36708-4

[LSI]

Star Diwa

Contents

Preface vii

1/Sensors 1
- Project 1: Photoresistor to Measure Light 3
  - Parts 4
  - Build It 5
  - Discussion: Photoresistors 6
- Interactive Sensor Control 7
- Going Forward 7

2/Basic Sensors 9
- Project 2: A Simple Switch 9
  - Parts 9
  - Build It 10
  - Troubleshooting 11
- An LED Needs a Resistor 12
- Project 3: Buzzer Volume Control 12
  - Parts 13
  - Build It 13
  - Troubleshooting 14
- Project 4: Hall Effect 14
  - Parts 15
  - Build It 15
  - Troubleshooting 17
- Project 5: Firefly 17
  - Integrated Circuits 17
  - 555 Timer IC 18
  - Light Up an LED When It's Bright 19
  - Light to Darkness 21
  - Transistors 22
  - Fading an LED 24
  - 555 Fading Blink 25

Capacitors. 28
Firefly. 29

3/Sensors and Arduino 33
Project 6: Momentary Push-Button and Pull-Up Resistors. 34
Parts. 34
Build It. 35
Run the Code. 35
Pull-Up Resistors and Arduino. 37
Project 7: Infrared Proximity to Detect Objects. 40
Parts. 41
Build It. 42
Project 8: Rotation (Pot). 43
Parts. 45
Build It. 46
Run the Code. 46
Project 9: Photoresistor to Measure Light. 47
Parts. 47
Build It. 48
Run the Code. 49
Project 10: FlexiForce to Measure Pressure. 49
Parts. 50
Build It. 50
Run the Code. 51
Project 11: Measuring Temperature (LM35). 52
Parts. 53
Build It. 53
Run the Code. 55
Project 12: Ultrasonic Distance Measuring (HC-SR04). 56
Parts. 57
Build It. 57
Run the Code. 58
Conclusion. 61

4/Sensors and the Raspberry Pi 63
Project 13: Momentary Push Button. 64
Parts. 64
Build It. 64
Run the Button Code. 66
Troubleshooting. 67

Hello, Python World. 68
Project 14: Blink an LED with Python. 69
  Parts. 69
  Build the LED Blink Project. 69
  Run the Code. 70
Project 15: Adjustable Infrared Switch. 73
  Parts. 73
  Build the IR Switch Project. 73
  Run the Code. 74
  Voltage Divider. 76
Analog Resistance Sensors. 77
Project 16: Potentiometer to Measure Rotation. 77
  Parts. 78
  Build It. 78
  Install SpiDev. 78
  Allow SPI Without root. 79
  Run the Code. 80
Project 17: Photoresistor. 82
  Parts. 82
  Build It. 83
  Run the Code. 83
  Playing with Resistance Numbers. 83
Project 18: FlexiForce. 85
  Parts. 85
  Build It. 85
  Run the Code. 86
Project 19: Temperature Measurements (LM35). 86
  Parts. 86
  Build It. 86
  Run the Code. 87
Project 20: Ultrasonic Distance. 89
  Parts. 89
  Build It. 90
  Run the Code. 90
  Real Time or Fast?. 92

A/ Troubleshooting Tactics. 95

B/ Arduino IDE Setup. 97

C/ Setting Up Raspberry Pi **101**

D/ Bill of Materials **117**

Index **123**

Star Diwa

Preface

There is a world of things happening around you, most of which become knowable to you thanks to one or more of your five senses. Sensory perception happens so quickly, and so often, that it's easy to overlook how impressive a system you actually are.

Take a moment and think how many sensory events happened to you from the time you woke up to the time you began reading this book. It's likely that you can't even list all the sensory occurrences. Not only do you constantly sense the environment, but your senses also work together to compile a picture of the universe. For example, events such as people passing by, warm sun shining on your face, or observing that a cool breeze in the morning is getting warmer in the afternoon are all fine examples of your senses at work and your mind processing sensations. But how can a robot or gadget have similar input? You probably already know what makes this possible (you did buy a book on the topic): sensors.

Adding sensors to a circuit expands its capabilities just as your own senses expand your awareness and inform you about the world. Sensors provide an input for information about an environment and work much like your own senses. But sensation isn't the only issue with sensors. A component doesn't necessarily have the ability to draw conclusions when a particular event occurs. Say, for instance, that it is –5 degrees outside and you want to go for a walk; what should you wear? You know, of course, that a coat and winter clothing are in order, but a temperature sensor does not know this. It can certainly provide you with a temperature reading, but it does not make judgments or inferences about what you should wear—at least not at the component level. For sensors to matter in the same way that your own sensations *and* your reflection on these sensations matter, a level of data processing needs to occur on the sensor data. Ultimately, sensors are components that you wire so that, either through hardware or software, their data is processed—and that's what this book is about: how to wire sensors and process their data.

In the first part of this book, you'll learn how to wire up sensors to other components. The level of data processing isn't too robust at the component level, and the focus is really on just getting a sensor safely wired and teaching some of the basics. The second part of the book deals with how to process sensor data. You will learn how to easily and quickly write programs with Arduino to process sensor data, as well as how to wire and program a Raspberry Pi to support analog sensors.

In this book, you'll gain hands-on experience with some of the most useful and instructive sensors available. Among the sensors and applications in this book, you'll learn how to detect and respond to:

- Clicks and rotation with a potentiometer
- Distance with ultrasound
- Proximity with infrared sensors
- Light and dark with a photoresistor
- Temperature with a thermometer
- Relative humidity with a capacitive relative humidity sensor

What Sensors to Buy?

This book covers a number of specific sensors and components (a few are illustrated in Figure P-1). To make sourcing parts easier, we've included Appendix D, which lists a complete bill of materials for all the projects in this book.

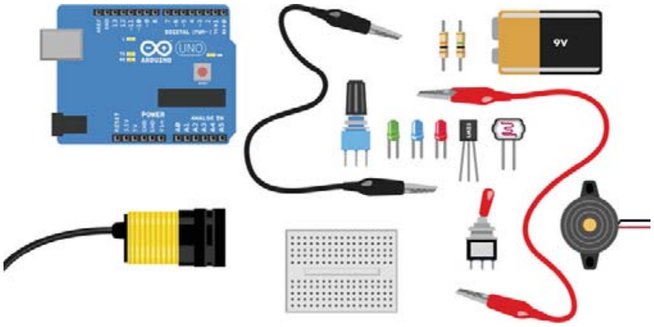

Figure P-1. *Arduino, sensors, and components*

Some well-known sellers of Arduino boards and related parts include Maker Shed, SparkFun Electronics, Parallax, and Adafruit. All four of these shops should stock most of the individual sensors used in this book, and all sell original, high-quality parts. Start with these shops.

Global electronics distributors, like Element14 and RS Components, are great places to order parts from, too. However, their product lists can be daunting for beginners to navigate. These global suppliers stock parts that differ from each other only in the pin format or voltage tolerances, which can be quite exhaustive. The parts sold from both of these global suppliers are high quality parts and very well documented.

Some online shops are very cheap, but these places typically do not sell official Arduino boards—they will say the product is "compatible." The sensors they sell may differ slightly in their pin configuration or even general appearance. At the time of writing, DealExtreme is one of the most popular shops of this sort. Even though they are based in Hong Kong and Shenzhen and offer free worldwide shipping, the quality of their parts varies a lot and delivery time can be slow. AliExpress is another popular Asian shop.

If you're ordering from abroad, research your local laws regarding custom fees. In some countries, small orders may be exempt from customs and taxes.

Conventions Used in This Book

The following typographical conventions are used in this book:

Italic
Indicates new terms, URLs, email addresses, filenames, and file extensions.

`Constant width`
Used for program listings, as well as within paragraphs to refer to program elements such as variable or function names, databases, data types, environment variables, statements, and keywords.

`Constant width bold`
Shows commands or other text that should be typed literally by the user.

`Constant width italic`
Shows text that should be replaced with user-supplied values or by values determined by context.

This element signifies a tip, suggestion, or general note.

This element indicates a warning or caution.

Using Code Examples

You can download all the source code for this book from *http://getstarted.botbook.com*.

You can extract the zip package by double-clicking it, or by right-clicking and selecting Extract from the pop-up menu.

This book is here to help you get your job done. In general, you may use the code in this book in your programs and documentation. You do not need to contact us for permission unless you're reproducing a significant portion of the code. For example, writing a program that uses several chunks of code from this book does not require permission. Selling or distributing a CD-ROM of examples from Make: books does require permission. Answering a question by citing this book and quoting example code does not require permission. Incorporating a significant amount of example code from this book into your product's documentation does require permission.

We appreciate, but do not require, attribution. An attribution usually includes the title, author, publisher, and ISBN. For example: "*Getting Started With Sensors* by Kimmo Karvinen and Tero Karvinen (Maker Media). Copyright 2014, 978-1-4493-6708-4."

If you feel your use of code examples falls outside fair use or the permission given here, feel free to contact us at *bookpermissions@makermedia.com*.

Safari® Books Online

Safari Books Online is an on-demand digital library that delivers expert content in both book and video form from the world's leading authors in technology and business.

Technology professionals, software developers, web designers, and business and creative professionals use Safari Books Online as their primary resource for research, problem solving, learning, and certification training.

Safari Books Online offers a range of product mixes and pricing programs for organizations, government agencies, and individuals. Subscribers have access to thousands of books, training videos, and prepublication manuscripts in one fully searchable database from publishers like Maker Media, O'Reilly Media, Prentice Hall Professional, Addison-Wesley Professional, Microsoft Press, Sams, Que, Peachpit Press, Focal Press, Cisco Press, John Wiley & Sons, Syngress, Morgan Kaufmann, IBM Redbooks, Packt, Adobe Press, FT Press, Apress, Manning, New Riders, McGraw-Hill, Jones & Bartlett, Course Technology, and dozens more. For more information about Safari Books Online, please visit us online.

How to Contact Us

Please address comments and questions concerning this book to the publisher:

Make:

1005 Gravenstein Highway North
Sebastopol, CA 95472
800-998-9938 (in the United States or Canada)
707-829-0515 (international or local)
707-829-0104 (fax)

Make: unites, inspires, informs, and entertains a growing community of resourceful people who undertake amazing projects in their backyards, basements, and garages. Make: celebrates your right to tweak, hack, and bend any technology to your will. The Make: audience continues to be a growing culture and community that believes in bettering ourselves, our environment, our educational system—our entire world. This is much more than an audience, it's a worldwide movement that Make: is leading—we call it the Maker Movement.

For more information about Make:, visit us online:

Make: magazine: *http://makezine.com/magazine/*
Maker Faire: *http://makerfaire.com*
Makezine.com: *http://makezine.com*
Maker Shed: *http://makershed.com/*

We have a web page for this book, where we list errata, examples, and any additional information. You can access this page at: *http://bit.ly/get-start-sensors*.

To comment or ask technical questions about this book, send email to: *bookquestions@oreilly.com*

Acknowledgments

The authors would like to thank Hipsu, Marianna, Nina, and Valtteri.

Star Diwa

1/Sensors

Sensors surround you in daily life. The world is full of them: from passive infrared sensors in motion detectors, to CO_2 detectors in air conditioning systems, and even tiny accelerometers, GPS modules, and cameras inside your smartphone and tablet—sensors are everywhere! The variety of sensor applications is remarkable.

It's safe to assume that if an electronic device is considered "smart," it's full of sensors (Figure 1-1). In fact, thanks to the proliferation of smart devices, especially phones, the price of sensors has been driven to affordability. Not only is it economically viable to add advanced sensors to your projects, but they vastly expand the kinds of projects you can make.

You'll learn about sensors in this book by making small projects and reflecting on the experience. It's more fun to build first and discuss later, but both are equally important. It's best to avoid the temptation to only build projects and skip the conceptual sections.

Getting started with sensors is easy, and only the sky is the limit. Electronics challenge some of the best brains daily and produce new innovations and dissertations. On the other hand, even a child can get started with some guidance.

If you don't know much about sensors yet, try to remember what it feels like now. After you've tackled some challenges and built a couple of gadgets, many dark mysteries of sensors will probably seem like common sense to you.

This book is suitable for anyone with an interest in sensors (see Figure 1-2). After you've built the gadgets and have read this book, you can get ideas for bigger projects from our book *Make: Arduino Bots and Gadgets* or learn more advanced sensors in *Make: Sensors*. For a wider view of the basics, see *Getting Started with Arduino, 2nd Edition* by Massimo Banzi, *Getting Started with Raspberry Pi* by Matt Richardson and Shawn Wallace, or *Make: Electronics* by Charles Platt.

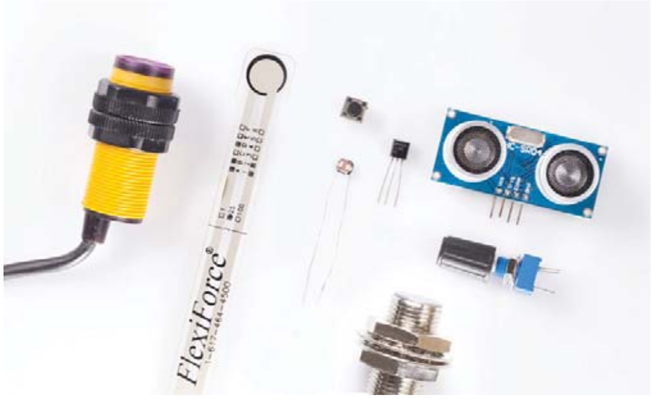

Figure 1-1. *Various sensors: infrared proximity, rotation, brightness, button, temperature, and distance*

What are sensors? Sensors are electrical components that function as input devices. Not all inputs are explicitly sensors, but almost all inputs use sensors! Consider your computer mouse or trackpad, a keyboard, or even a webcam; these are not sensors, but they definitely use sensors in their design. More abstractly, you can frame sensors as a component to measure a stimulus that is external to the system it is in (its environment). The output data is based on the measurement. For example, when you type at a keyboard, the letter that appears on your screen (the output) is based on the measurement (which switch, or key, you pressed on the keyboard). How many letters appear on screen is based on another measurement (how long you keep the key pressed).

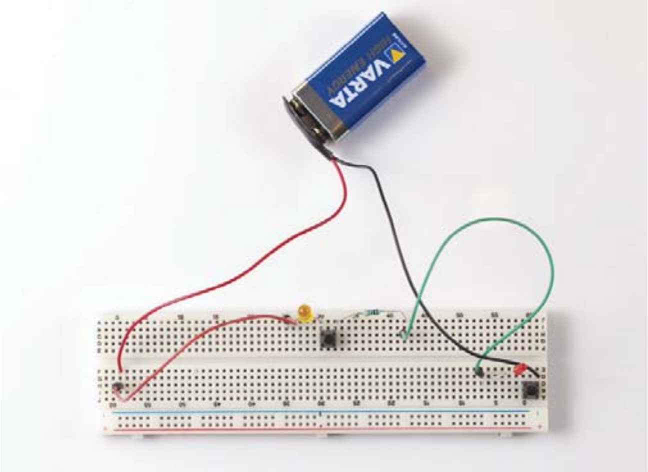

Figure 1-2. *Simple AND connection with buttons, built and designed by a four-year-old with help from an adult*

The first project uses a photoresistor to measure light. Without the photoresistor (or similar sensor), there is no way the circuit can know how bright the light is in the environment. By adding the sensor, your circuit knows something it didn't know before.

All of the projects in this book *evaluate* a particular stimulus within the environment. None of this would be possible without sensors. Let's get building so you can experience the inputs and outputs that sensors provide to projects.

Project 1: Photoresistor to Measure Light

Light in an environment is quite informative: you can determine what time of day it is based on the sun's angle, you operate a car more safely at night when its lights are on, and people who do not experience enough light in daily life can become depressed with seasonal affective disorder. As such, light influences many aspects of your life and it's fun to measure it, too.

The simplest sensor for detecting light is a *photoresistor*. It's not uncommon to also encounter another name for the exact same sensor: *light-dependent resistor* (LDR). The component works by changing its resistance based on the amount of light hitting it.

Now that you know the right sensor to use, the next question to think about is how to process the sensor's measurements. If you've ever worked with a light-emitting diode (LED) , shown in Figure 1-3, you might know that resistance is an electrically important consideration. For example, if you've ever used a larger-value resistor for the LED than a project called for, you've seen that too much resistance can restrict an LED from illuminating. This same basic observation is applicable to this project.

Figure 1-3. *LEDs*

The circuit is designed so that an LED is dependent on the photoresistor's measurement. Too much resistance and the LED simply will not turn on. Enough discussion—it's time to build! Figure 1-4 shows the finished project.

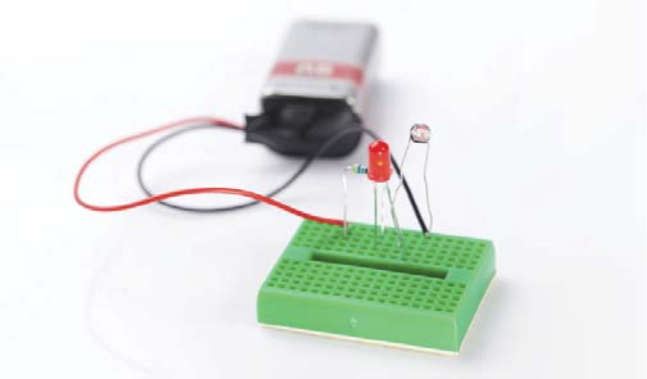

Figure 1-4. *The finished photoresistor project*

Parts

You need the following parts for this project:

- Photoresistor
- 5 mm red LED (different LEDs will work differently with this circuit; later, you'll learn a more sophisticated way to fade LEDs)
- 470 Ω resistor (four-band resistor: yellow-violet-brown; five-band resistor: yellow-violet-black-black; the last band will vary depending on the resistor's tolerance)
- Breadboard
- 9 V battery clip
- 9 V battery

All of these parts, except the 9 V battery and 470 Ω resistor, are available in the Maker Shed Mintronics: Survival Pack, part number MSTIN2. You can use two of the 220 Ω resistors in series or one 1 kΩ resistor in place of the 470 Ω resistor; both of these are available from electronics retailers such as RadioShack.

Build It

Here are the steps for building this project:

1. Orient your breadboard so that it is wider than it is tall, as shown in Figure 1-5.

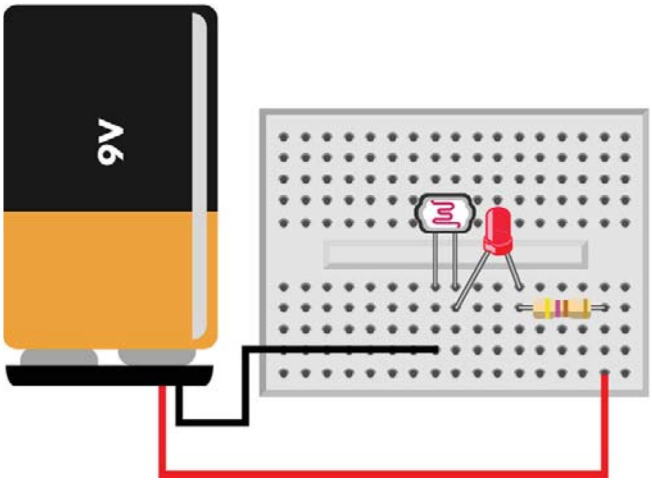

Figure 1-5. *Circuit digram for photoresistor project*

2. Look at your LED and determine which lead has a flat side above it on the colored plastic housing—this indicates the negative lead of the LED (the negative lead is also the shorter of the two), as shown in Figure 1-6. LEDs have a certain *polarity* and putting them in backward might damage them.

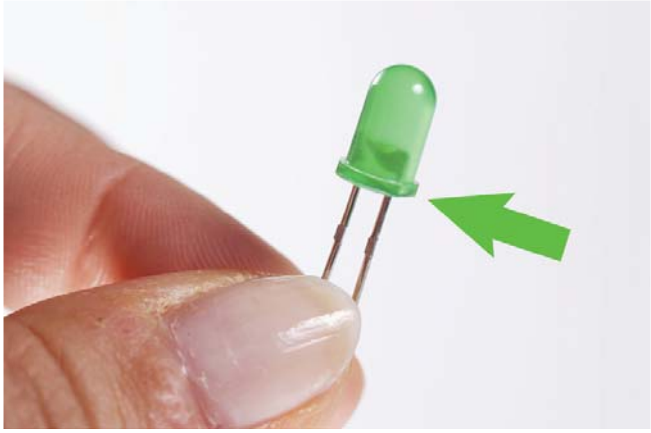

Figure 1-6. *Negative leg of the LED*

3. Insert the photoresistor so that the negative lead of the LED and one of the photoresistor leads occupy the same column. The second (positive)

LED lead should occupy its own column for now. Refer back to Figure 1-5 to see how they should be arranged.

Do you see the gap in the middle of the breadboard in Figure 1-7? That gap separates the two groups of columns, and there's no connection across it. If you want two leads in the same column to be connected, be sure they are on the same side of the gap.

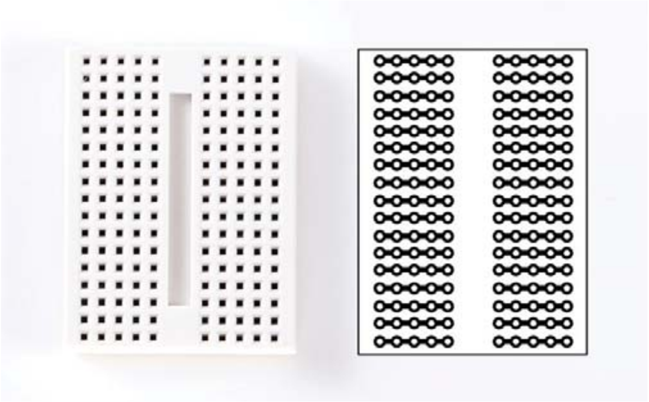

Figure 1-7. *Breadboard layout*

4. Connect the 470 Ω resistor to the column with the positive LED lead and make sure it's not the same column that already has the photoresistor and LED's negative lead in it. Make sure that the resistor's other lead is in a separate column.
5. Attach the black wire from the battery clip to the column that contains only a photoresistor lead.
6. Insert the red wire from the battery clip to the column that contains only a resistor lead.
7. Double-check the steps and if everything looks like Figure 1-5, connect the 9 V battery.

That's it. You've built your first sensor circuit. Congratulations!

Discussion: Photoresistors

It might not seem like the circuit is doing much. That's because the light levels probably have not changed much in your room. Put your finger over the photoresistor and watch the LED closely. Did anything seem to change with the LED? There should have been some change in the LED brightness. Try adding more light to the photoresistor. The opposite happened, right? Now that you've seen the photoresistor in action, how would you describe what is going on when you expose the photoresistor to more light in terms of its resistance? Is the resistance increasing or decreasing when the sensor is exposed to more light?

Here's what's going on in the circuit. The more light hits the photoresistor, the lower its resistance. If the room is quite bright, then the LED will be quite bright. If light is low in the room, the sensor resists the current flow, which is expressed by the LED getting dimmer.

The current flow through the sensor controls how bright the LED will shine. That's because the circuit is wired so that all current to the LED must pass through the photoresistor first.

The photoresistor, as you just learned, is a *resistive sensor*. There are many types of resistive sensors; this category of sensors is used to measure much more than luminosity. As you continue reading and you encounter new sensors, it's a good idea to think how the stimuli are measured and especially how the output is structured. None of the sensors will output data that is in a convenient format for end-user consumption. Instead, you will need to decide how to express or format the raw sensor data output in a way that makes sense to users.

Another sensor type is *electro-mechanical*. These sensors do not manifest changes in voltage or current, but rather by a change in their physical properties. The thermostat in your home or apartment is a great example (unless you have a digital thermostat). When the room temperature changes, a thermostat's *bimetallic coil* will expand or contract depending on whether the room's temperature increases or decreases. The sensor is actually physically expressing itself by changing shape! But even these sensors may trigger an electronic sensor (for example, a thermostat's bimetallic coil is usually connected to a tilt switch that turns the heat on or off).

Interactive Sensor Control

"Project 1: Photoresistor to Measure Light" on page 3 used a sensor in a way that didn't directly involve interaction with a human. Sure, you were the one who changed the lighting in the room to force the photoresistor to change its resistance, but it could have just as easily been the setting or rising of the sun. There are many sensors that you'll manipulate directly; you'll see these in later chapters.

Going Forward

In all the projects in this book, you'll be building small systems that collect input data by taking measurements with a sensor. The systems will do something that processes that input data, and then take action (the output of the system). At first, you'll just build things with electronic components, but later in this book, you'll use Arduino and Raspberry Pi to handle the processing.

When you use Arduino and Raspberry Pi, you'll write code that does a lot of the work for you. The benefit will become very clear because you'll find that you can change the way you respond to an input without having to rewire your circuit.

Suppose you want to challenge your friend to see who can press a force-sensing resistor down more firmly. All you would need to do is add some extra lines of code to the Arduino sketch and send the new code. How would you accomplish the same trickery if you didn't have the option to alter code? It certainly would be more challenging, and at the very least you would have to move a few things on the breadboard. And how would you display the score for the contest results? It turns out that an Arduino handles quite a lot of work for you!

But this isn't the whole story of sensors. We don't want you to think that as sensor systems increase in physical complexity, their programs become more complicated. Rather, we want you to think of this as an issue of applicability: what's the best design to accomplish your goal?

Star Diwa

2/Basic Sensors

Get ready to wire a few more sensors and learn a bit more theory. The circuits you will build in this chapter do not need to be programmed, which is to say, they will not use an Arduino or a Raspberry Pi. The sensor's data will be "interpreted" by the electrical properties of each component in the circuit design.

A sensor is a physical input to a circuit. Sensors are *transducers*. The process of converting sensed energy into another form is called *transduction*. For example, a light sensor transduces luminosity into resistance.

Another type of transducer is an *actuator*. Rather than reacting to something in the environment, an actuator makes something happen. An actuator is a physical output of a circuit. An LED, for example, transduces electrical current into light; a speaker turns it into sound.

You'll use both types of transducers (sensors and actuators) in the projects in this book. So if someone asks you how a sensor exerts control over a circuit's simply reply, "Transduction!"

Project 2: A Simple Switch

A switch can connect and disconnect the flow of electricity. It's just like a button, except that it stays where you leave it. If you switch it off, it stays off. If you switch it on, it stays on.

Parts

You need the following parts for this project:

- A switch
- A wire with alligator clips
- Two 1.5 V batteries
- Battery holder with wire leads
- 5 mm red LED

- 470 Ω resistor (four-band resistor: yellow-violet-brown; five-band resistor: yellow-violet-black-black; the last band will vary depending on the resistor's tolerance)
- Breadboard

Some toggle switches have three leads. These *single pole double throw* (SPDT) switches are used in cases where you want the switch to choose between two different paths for electricity to flow. If you only wire up two of the leads, you're just using it as a regular single pole single throw (SPST) switch.

SPDT switches are used when a simple on-off switch won't do the trick, such as the pair of light switches commonly found at the top and bottom of a staircase. Because of a multiway switching wiring scheme, flipping either switch will toggle the state of the light.

Build It

Figure 2-1 shows the circuit diagram for this project. Here's how to wire it up:

1. Orient your breadboard so that it is taller than it is wide, as shown in Figure 2-1.

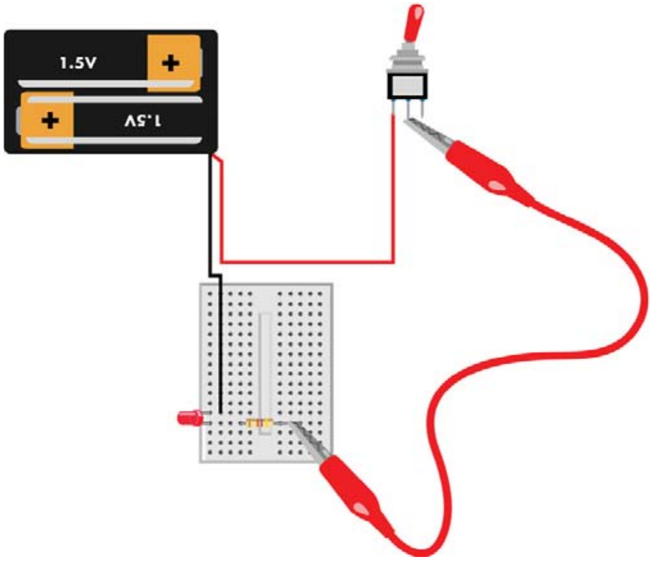

Figure 2-1. *Circuit diagram for the simple switch circuit; the short lead connects to the black wire*

2. Insert the red LED into the breadboard and keep track of which side is positive (the longer lead) and which is negative (the shorter lead).

3. Insert one lead of the 470 Ω resistor into the row containing the LED's positive lead.

   Keep in mind that the gap in the middle of the breadboard separates the two groups of rows, and there's no connection across that gap. If you want two leads in the same row to be connected, be sure they are on the same side of the gap.

4. Connect the alligator wire from the other resistor lead to the middle lead of the switch.
5. Connect the red positive wire from the battery holder to either side lead of the switch using the following technique (see Figure 2-2):
   a. Push the wire through the hole.
   b. Fold the wire into itself.
   c. Twist the folded wire to secure it.

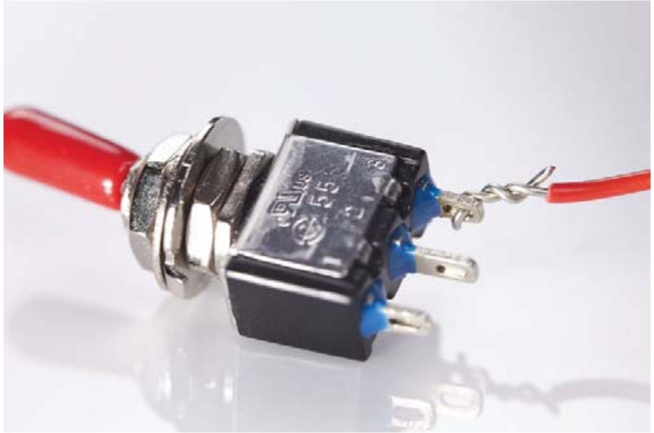

Figure 2-2. *Tying a wire to the switch leads*

6. Insert the batteries into the holder.

With everything wired up, test the switch and see if your LED turns on and off as you'd expect from a switch.

Now you can add a power switch to any of your projects. No more removing the battery when you are done playing with a prototype.

Troubleshooting

If this project doesn't work, try checking the following things:

- If your switch has three leads, did you connect the switch's center and one side lead (rather than using two side leads)?
- Make sure that the alligator clip only touches the middle lead of the switch. If the alligator clip short-circuits two leads, the LED might be lit constantly, even if you flip the switch.
- Is the battery wire connected securely? If you find the LED is never lit, try pressing on the battery wire connection and flipping the switch. For a permanent fix, twist the switch more to make the wire tighter.

An LED Needs a Resistor

All the circuits in this book work with a normal red LED and a 470 Ω resistor. In a pinch, any resistor with a brown multiplier band (third stripe) will do.

What if you have a really fancy LED, such as a blue one or a superbright LED? If you have a general idea what the LED should look like when lit, you can just try out resistors. Start with the common 470 Ω. If the LED is too dim, pick a weaker resistance.

Most LEDs don't break easily, but they have very low resistance. So if you forget the current limiting resistor, the LED might drain your battery and cause some components to overheat and break. Even if you use an online LED resistor calculator (there are many out there; just Google it or look in your phone's app store) and it gives you a zero (or less!), you should still use a very small resistor.

Third Band Trick

We use a shortcut called the *third band trick* to pick out resistors. Hold a resistor so the tolerance band is on your right—these are usually gold or silver bands. The third band in from the left is known as the multiplier band.

Some resistors have five bands, not four, although they are uncommon. But all the same, the multiplier band is the second-to-last band. For a five-band resistor, it's the fourth band.

Our rule of thumb (when working with 3 V to 5 V) for choosing a resistor by its third band: pick brown if you are protecting an LED, green if you need a pull-up resistor.

Project 3: Buzzer Volume Control

In many projects, you need the ability to slightly adjust input or output rather than toggling between on and off. A *potentiometer* will allow you to operate between the extremes of on and off, just as the photoresistor did in "Project

1: Photoresistor to Measure Light" on page 3. Potentiometers, also known as *pots*, are often used as volume controls for audio devices.

Parts

You need the following parts for this project, which are available in the Maker Shed Mintronics: Survival Pack, part number MSTIN2 (you will need to purchase the 9 V battery and piezo buzzer separately:

- DC piezo buzzer (Maker Shed part number MSPT01)
- 10 K potentiometer
- 9 V battery
- Battery clip
- Breadboard

You are using a 9 V battery in this project because common piezos typically operate with a range of 6 V to 18 V or more. Because we are going to be increasing and decreasing the current with a potentiometer, using a 9 V source voltage gives us room to increase or decrease the voltage and still produce audible output.

Build It

Here are the steps for building this project:

1. Orient your breadboard so that it is wider than it is tall, as shown in Figure 2-3.
2. Connect the negative wire of the piezo buzzer to any free column of holes on the breadboard and then insert the black (negative) battery clip wire into the same column.
3. Insert the potentiometer into the breadboard, but make sure you don't put it in the same column as the negative wires you just connected. Put the battery clip's red (positive) lead into the same column as the potentiometer's middle lead.

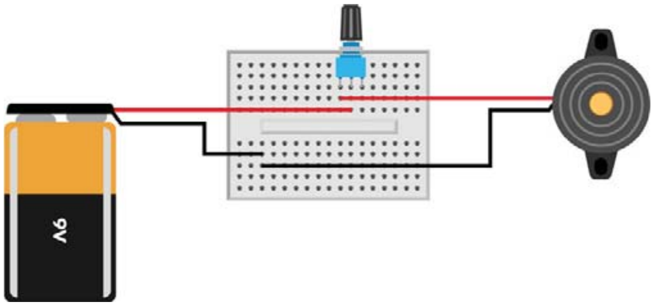

Figure 2-3. *Circuit diagram for buzzer volume control*

4. Next, connect the piezo's positive lead to one of the potentiometer's outer pins. It doesn't matter which outer pin you choose, just don't pick the middle one—that's not an outer pin.
5. Clip in the 9 V battery and turn the potentiometer left to right to change the output level of the piezo buzzer.

Now you can add loud noise to any of your projects. You also know how to use a potentiometer as an adjustable resistor. You will soon learn about many other sensors that report their value with resistance, just as a potentiometer does.

Troubleshooting

If this project doesn't work, try checking the following things:

- Does your piezo buzzer work? Connect it directly to the battery to make sure it works (the piezo black lead goes to the battery's black lead, and red goes to red).
- Did you connect the pot correctly? We're using the pot as a simple adjustable resistor, so we must connect one wire to the center lead and the other to either of the side leads.
- Check connections one by one. Start from battery positive or negative and go through the whole circuit.

Project 4: Hall Effect

A hall switch senses change in the magnetic field and detects if there is a magnet nearby. A very common use for this is a bike speedometer that identifies how often a rapidly rotating object passes by the sensor. Another typical application is a door burglar alarm like the simplified version we're going to build next.

Figure 2-4 shows the NJK-5002A Hall effect switch, which you can find from sellers on eBay or AliExpress for under $10. If you buy the similar NJK-5002C, you will have to change the way you wire the circuit (described later).

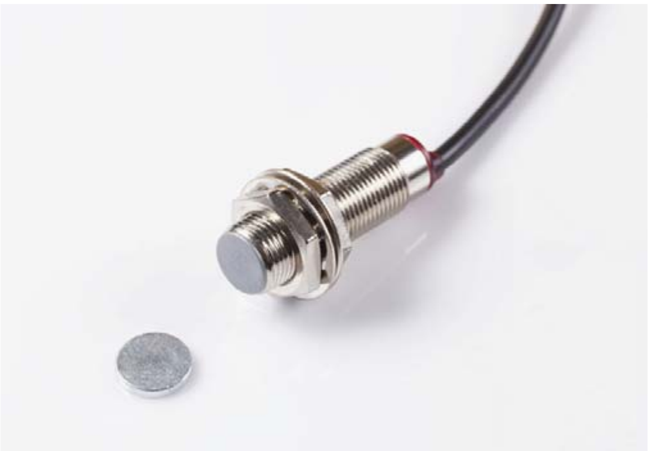

Figure 2-4. *Hall effect switch with magnet*

Parts

You need the following parts for this project:

- DC piezo buzzer
- 9 V battery
- Battery clip
- Breadboard
- NJK-5002A Hall effect switch

Build It

The NJK-5002A acts as a switch, and has connections for positive voltage (brown), negative (blue, although most components use black for negative), and a connection that outputs positive voltage (black, despite the fact that black usually indicates negative) when a magnet is held to the sensor.

If you are using the NJK-5002C, black will go negative rather than positive when a magnet is held to the sensor.

Here are the steps for building this project:

1. Orient your breadboard so that it is wider than it is tall, as shown in Figure 2-5.

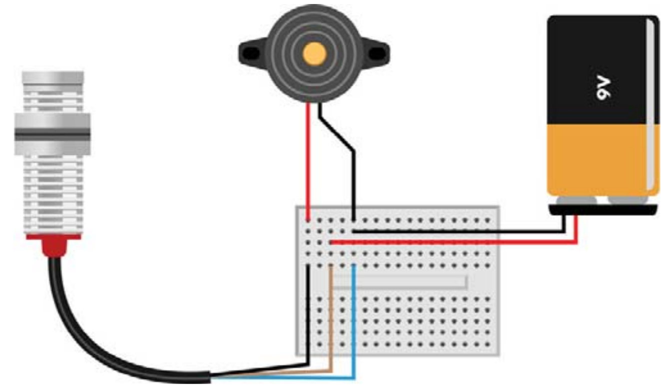

Figure 2-5. *Circuit diagram for the Hall sensor alarm*

2. Insert the black (output) Hall effect switch wire into any free column of holes on the breadboard and then insert the red positive wire of the piezo buzzer into the same column.

   If you are using the NJK-5002C instead of the NJK-5002A, connect the black negative wire from the piezo here instead of the red positive wire.
3. Insert the red positive wire of the 9 V battery clip into any free column of holes on the breadboard and then insert the brown (positive) Hall effect switch wire into the same column.

   If you are using the NJK-5002C instead of the NJK-5002A, insert the red positive wire from the piezo into this column as well.
4. Insert the blue (negative) Hall effect switch wire into any free column of holes on the breadboard and then insert the black negative wire from the battery clip into the same column.

   Insert the black negative wire from the piezo into this column. (If you are using the NJK-5002C, you don't need to make another connection to it in this step.)
5. Clip in the 9 V battery and hold a magnet to the switch to make the buzzer sound.

Now that you've learned how to work with a Hall effect switch, you can use it in your projects in many ways. You can use it to detect when two things are brought together or separated. For example, you could put a magnet in a door, and a Hall effect switch in the doorframe to know when it is opened or closed.

It can also be used to measure rotational speed: if you embed a magnet in a stationary bicycle wheel, you could use a Hall effect switch to measure how fast you are pedaling. You could use the speed as input to a video game to make your exercise sessions more interesting.

Troubleshooting

If this project doesn't work, try checking the following things:

- Try flipping the magnet over if it doesn't work. The switch won't be triggered unless you press the right pole of the magnet against it.
- Did the LED that's built into the sensor light up? If not, it's probably not wired right.
- Are you using the correct model of Hall switch? The build instructions tell you how to connect two popular models (NJK-5002A and NJK-5002C). If you're using a different model, check the datasheet for information that will help you connect it correctly.
- Make sure the piezo buzzer is working by connecting it directly to the battery. Does it make a sound?

Project 5: Firefly

Night falls, and fireflies start to glow, slowly fading in and out. In case your garden doesn't have any real fireflies, you can build your own. You'll use a light-dependent resistor to detect darkness, and a 555 timer circuit to continuously fade an LED in and out.

It's time to venture beyond basic components and leap to integrated circuits (ICs). You'll still use the components from the previous project, but also learn to use one of the most popular ICs, the 555 timer.

First, you'll make an LED light up when it's bright outside. Then you reverse this, making your LED light up in darkness. You'll make the LED fade in and out with the 555. Finally, you'll combine the circuits to make the firefly. It turns on the LED fading in darkness.

By building the 555 firefly in distinct, testable steps, you will learn a systematic approach to building analog circuits. You won't learn analog circuit design here, but at the end of this project you'll get some ideas for simple modifications. Designing completely new gadgets is easier with Arduino, which you'll see later in the book.

Integrated Circuits

ICs vary in complexity and price. The 555 timer we use here is a cheap, comparatively simple IC: it only has about 40 components inside, most of these being transistors and resistors. Still, imagine if you didn't have the luxury of using a 555 timer. Your projects would balloon in parts and the wiring would not be an easy process. If you think 40 components is a lot, you will be taken aback to learn that there are billions of components in ICs such as a processor or memory chip. The most complex IC we will be using in our sensor projects is the Broadcom BCM2835 *system-on-chip* found in the Raspberry Pi.

This book teaches you the things you need to know to use the components in this book. If you want to know more, you can search the Web for the name of the IC and the word "datasheet" (e.g., "CA555E datasheet"). Datasheets are sometimes tedious to read, but they are the authoritative source on how the IC works.

Pin numbering can be found on the datasheet. Usually, pins are numbered counter-clockwise from the notch. Hold the IC, pins down, so that you look at it from the top. Locate the polarity mark: a half-moon notch, a corner triangle, or a dot on pin 1. Pins are numbered counter-clockwise from that mark, starting from pin 1.

For the CA555E, hold the half-moon notch to the left and the pins away from you. See Figure 2-6. Pin 1 is the bottom left, and the numbers increase counter-clockwise. Thus, the last pin (8) is at top left. To make your life easier, the circuit diagrams in this book have small pin numbers on the 555 chips.

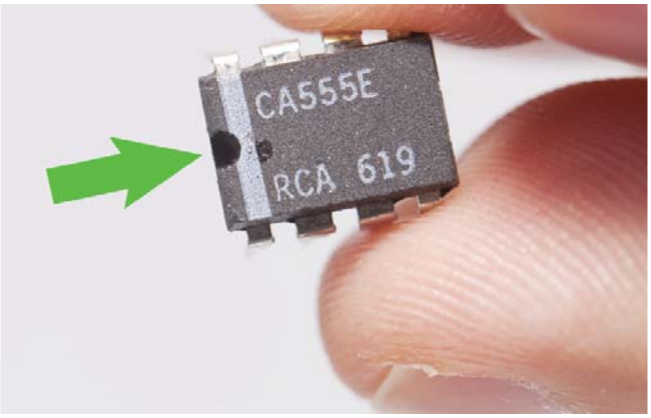

Figure 2-6. *Pins are numbered counter-clockwise from the notch*

555 Timer IC

The 555 timer IC (shown in Figure 2-7) is a multipurpose chip. You can configure it as a timed delay, an oscillator, or as a flip-flop.

Modes of 555 operation are:

- Leave the output on for a while (*monostable*).
- Toggle the output (*bistable*, also known as flip-flop).
- Blink the output without any input (*astable*).

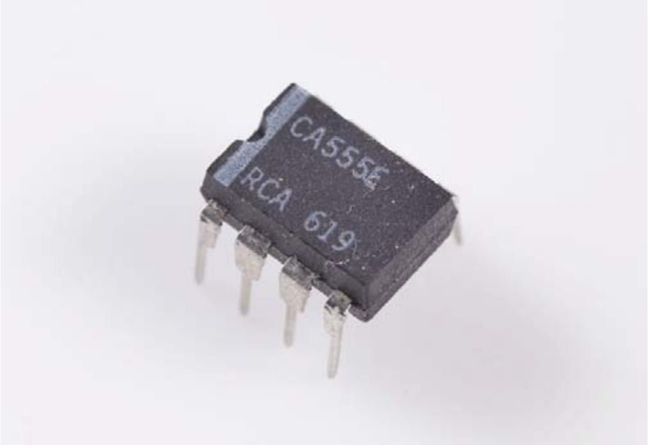

Figure 2-7. *555 timer integrated circuit*

Configuring a 555 timer is *not* analogous to the programming you will do on an Arduino or Raspberry Pi, where lines of code are compiled and executed. Instead, the 555 timer configuration is done by connecting certain components such as resistors and capacitors to the 555. Hardware sets up the 555, not software.

Building and configuring gadgets by coding with Arduino is easier than the component-by-component approach you use with the 555. Here, we have already designed the circuit for you and made the necessary calculations, making it a simple task to get familiar with the 555.

Light Up an LED When It's Bright

A light-dependent resistor (LDR) cuts off electricity when it's dark. When it's well lit, the resistance is low and current passes through. The brighter it is, the more current goes through.

How do you light up an LED when it's bright? Just connect an LDR in series with the LED. The LDR works just like any analog resistor, so you could use a potentiometer in its place.

Sound familiar? That's because you've built this circuit earlier (see Figure 2-8). Refer to the building instructions in "Project 1: Photoresistor to Measure Light" on page 3.

Put your finger over the LDR, and the LED goes dark. Aim a bright light at the LDR, and the LED is brightly lit.

Who needs light in a bright room? Can't you make it light up in the darkness?

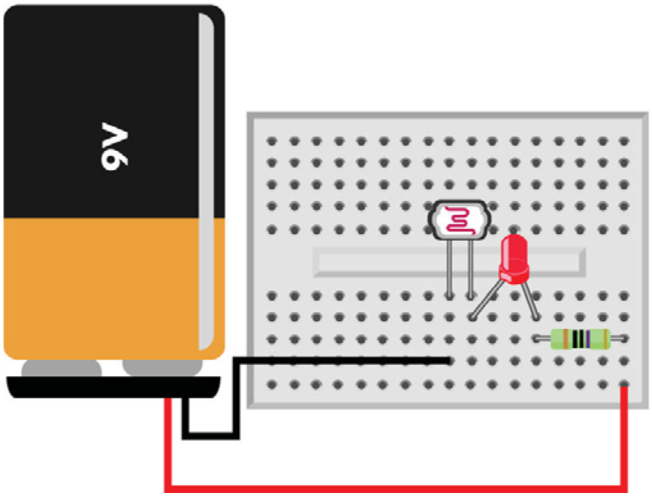

Figure 2-8. *LED/LDR circuit*

Jumper Wires

Jumper wires, also known as hookup wires, allow you to connect components to each other on a breadboard (Figure 2-9).

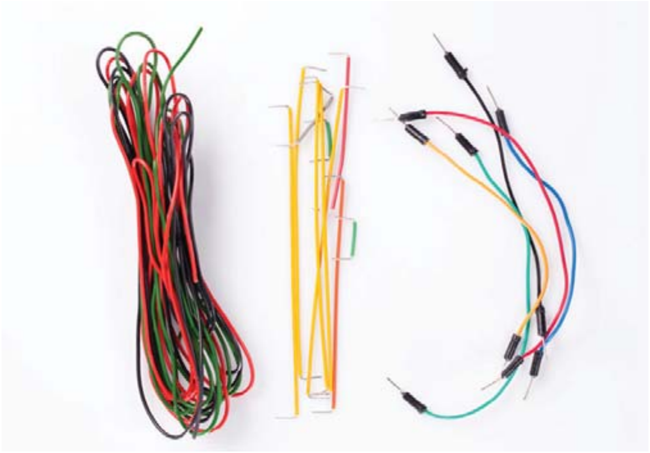

Figure 2-9. *Jumper wires*

These wires are not fancy and you can even make them yourself by stripping the ends off of insulated solid core thin wire. Just because you can make jumper wires doesn't mean it is not worth buying a pre-made collection of jumper wires. They are sold in various colors, and their ends are straight and reinforced so it's easy to push them into the breadboard.

Jumper wires are available in a number of colors, lengths, and connector housings. The core of the wire refers to the part surrounded by plastic insulation and can be stranded or solid. You want solid core wire because the breadboard clips need to latch onto something solid.

In an attempt to make life easier, jumper wires come in a variety of colors. There is nothing electrically different about these hued wires, but typically black is used for wires connected directly to a ground terminal (0 V, GND) and red is used for direct connections to a positive terminal (+5 V). We strongly recommend you reserve black for negative/ground and red for positive, but feel free to use all the other colors any way you want.

Light to Darkness

You can use a transistor and a resistor to invert the effect of the LDR.

A transistor is an amplifier (see "Transistors" on page 22). In this circuit, you use the most common transistor, a bipolar NPN transistor. The amplifier circuit here is the common emitter amplifier, the most common transistor amplifier (see "Common Emitter Amplifier" on page 23).

Controlling base current to the transistor is done with two resistors. One of the resistors is connected to the plus terminal of the battery, and pulls the voltage up. The other resistor, the LDR, is connected to ground (minus). When it's bright, LDR has very low resistance. Thus, when it's bright, the base of the transistor gets pulled to ground, so it gets no current: zero amplified is still zero and the LED stays dark.

The connection using the two resistors is called a *voltage divider*.

Parts

You need the following parts for this project:

- BC547 transistor
- Light-dependent resistor (LDR)
- 470 Ω resistor (four-band resistor: yellow-violet-brown; five-band resistor: yellow-violet-black-black; the last band will vary depending on the resistor's tolerance)
- 5 mm red LED
- 9 V battery
- 9 V battery clip
- Breadboard

All of these parts, except the 9 V battery, 470 Ω resistor, and BC547, are available in the Maker Shed Mintronics: Survival Pack, part number MSTIN2, from . You can use two of the 220 Ω resistors in series or one 1 kΩ resistor in place of the 470 Ω resistor; both of these are available from electronics retailers such as RadioShack.

Build It

Build the circuit as shown in Figure 2-10. Covering the LDR with your finger lights up the LED. Point a flashlight at the LDR and see the LED go dark.

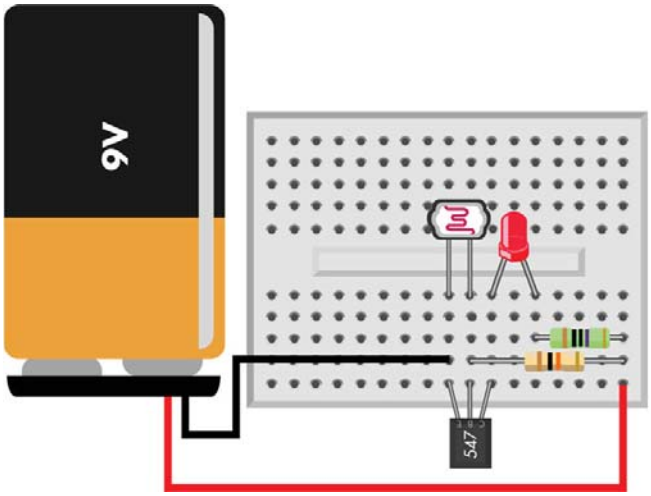

Figure 2-10. *LDR inverter circuit*

Transistors

A transistor (see Figure 2-11) is an amplifier, taking small signals and making them larger. It can also function as a switch, because a small power can control a larger power.

Transistors are actually the basis of all digital electronics. Your computer's CPU, the ATMega microcontroller in an Arduino board, and the *system-on-a-chip* on the Raspberry Pi have millions, even hundreds of millions of transistors arranged in patterns to create digital logic.

Microcontrollers, such as Arduino and Raspberry Pi, use very little power. If you want to control something that needs more power, such as a big motor, a transistor is one way to do that.

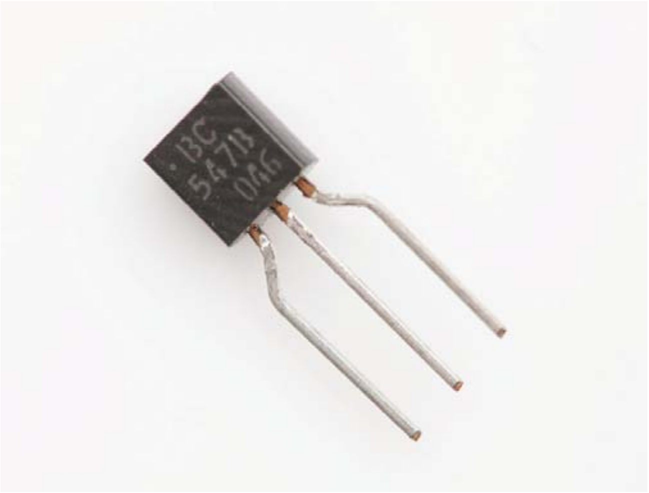

Figure 2-11. *A transistor*

In most prototyping projects, you don't need an in-depth understanding of transistor circuits. For example, you can use an Arduino and a servo (easy) instead of trying to control a motor with transistors (harder). That's why it's enough to get the high-level understanding of a transistor as an amplifier. You won't need to be able to design any transistor circuits after reading this short explanation about transistors.

The most common type of transistor is the NPN-type bipolar junction transistor (BJT). We'll only talk about these common NPN BJT transistors here, as it's much easier to learn the others once you know the NPN BJT transistor. The BC547 transistor you use in this chapter is a typical NPN BJT transistor.

An NPN BJT transistor has three leads: emitter E, base B, and collector C.

Common Emitter Amplifier

Emitter E is the common ground, the negative. The arrow points to the negative lead. As it says in the NPN mnemonic, "Not Pointing iN," the arrow in the NPN transistor symbol is pointing out to negative. The arrow is simply pointing the direction of current from positive to negative (see Figure 2-12).

The small controlling base current (BE) flows from base B to emitter E. This base current is the weak signal to be amplified.

The large collector current (CE) flows from collector C to emitter E. This collector current only flows if there is base current.

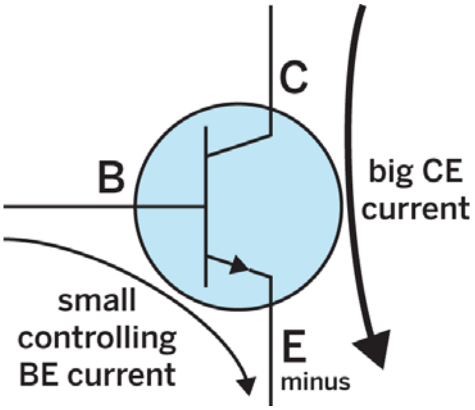

Figure 2-12. *NPN BJT transistor*

To give you an example of a transistor in action, consider a transistor where the base is at 0 V. There is no base-emitter current flowing, so the big collector-emitter current is blocked. When you apply current to the base, a small base-emitter current starts flowing. This allows the big collector-emitter current through, and could start a big motor or a bright light.

Take the transistor from this project, BC547, with leads down and the flat text side facing you. The collector is on the left, base is center, and the emitter is on the right.

Fading an LED

What could be easier than fading a light? Actually, fading an LED is tricky business. If you gradually add voltage, first it's off and then it's on. For a dimly lit LED, there is only a small voltage range available.

With microcontrollers, such as an Arduino, dimming is done by blinking the LED faster than the eye can see. Here, with analog electronics, you can use a transistor to help you control the small part of a voltage range where the LED is between on and off (see Figure 2-13).

In this circuit, the potentiometer sets the voltage of the transistor's base (B). The transistor is used in the common collector amplifier, where the small BC current controls the bigger EC current.

Build the circuit first, and test how you can use most of the potentiometer's range to control the brightness of the LED.

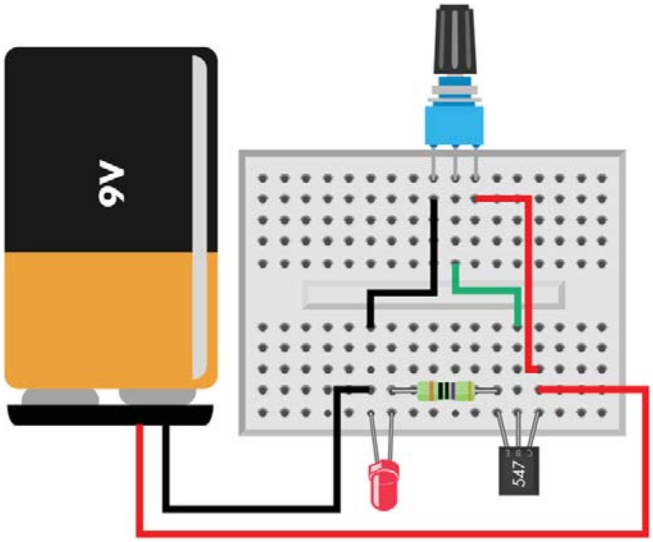

Figure 2-13. *Fading an LED with an NPN transistor*

Is there some doubt in your heart that the transistor is really needed? You can easily test it by building a worse, alternative setup (see Figure 2-14). Put a potentiometer in series with the LED. When you try out that setup, you'll see that only a small part of the potentiometer range dims the LED. How well (or poorly) the potentiometer works without the transistor depends on the LED you're using.

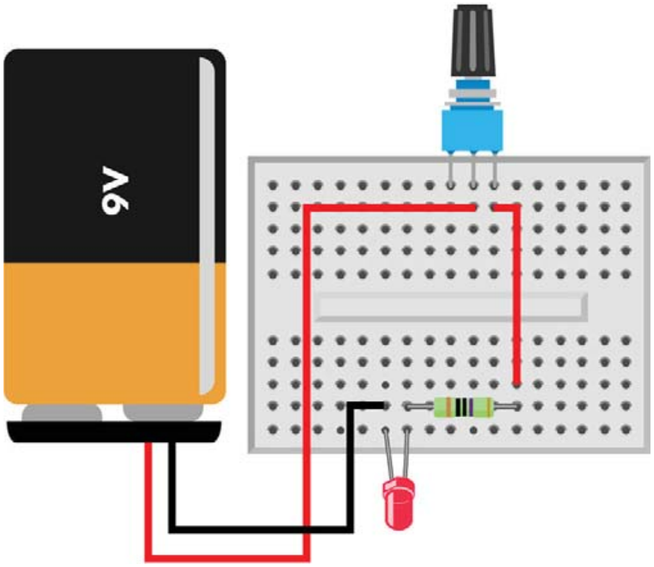

Figure 2-14. *Bad way of trying to fade an LED*

555 Fading Blink

Your firefly should continuously fade between lit and dark. You can use the 555 timer to fade the LED.

Build the circuit, then enjoy the explanation as you see your LED fading.

The previous projects can be accomplished on any type of breadboard, even tiny ones. When you use an IC, however, you need to make sure that your breadboard is split into two sections of horizontal rows. This breadboard layout allows for each lead of the IC to be isolated on its own electrically connected row.

Parts

You need the following parts for this project:

- BC547 transistor
- 470 Ω resistor (four-band resistor: yellow-violet-brown; five-band resistor: yellow-violet-black-black; the last band will vary depending on the resistor's tolerance)
- 5 mm green LED
- 9 V battery
- 9 V battery clip
- Breadboard
- CA555E, NE555, or compatible 555 timer integrated circuit
- 10 kΩ potentiometer
- 100 uF capacitor
- Jumper wires

All of these parts, except the 9 V battery, 470 Ω resistor, and BC547, are available in the Maker Shed Mintronics: Survival Pack, part number MSTIN2.

Build It

Build the circuit as shown in Figure 2-15.

In this circuit, the 555 timer is used in astable mode. The 555 output fades between 3 V and 6 V. Adjusted with a transistor amplifier, this makes the LED fade in and out.

The input pins of the 555 are 6 (THRES) and 2 (TRIG). The two pins are connected together. Pin 2 (TRIG) detects a voltage drop below 3 V and turns pin 3 (OUT) to 7 V, so that output voltage starts rising. Pin 3 (OUT) is connected to the capacitor through the potentiometer, so the 7 V voltage starts loading the capacitor. You can turn the potentiometer to choose how fast the capacitor charges.

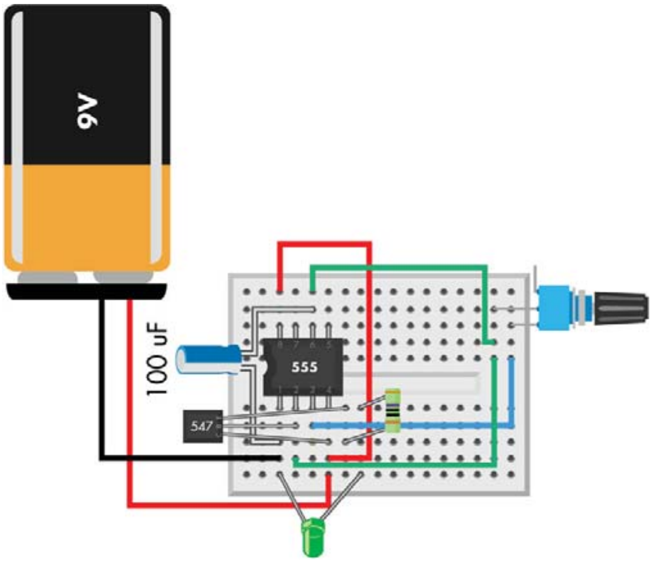

Figure 2-15. *555 fader circuit*

When the voltage on pin 6 (THRES) rises to 6 V, the 555 detects this and turns OUT low (to GND), so the voltage starts falling down. As you remember, the capacitor is connected to pin 3 (OUT) through the potentiometer, so the potentiometer chooses how fast the capacitor discharges.

As soon as the capacitor has discharged enough to bring the voltage to 3 V, pin 2 (TRIG) pin toggles pin 3 (OUT) to 7 V. The capacitor starts to charge, and the cycle starts over.

This way, the output of the 555 circuit keeps oscillating between 3 V and 6 V. Because it takes time to charge the capacitor, your circuit produces a triangle wave.

You probably want to connect an LED to the output of your new 555 oscillator. Or even better, connect the fading LED circuit with a transistor (see "Fading an LED" on page 24). So where do the two wires go? See Table 2-1. The output of the astable 555 circuit are GND (the 0V minus wire) and the point that connects to the capacitor, pin 2 (TRIG), and pin 6 (THRES).

Table 2-1. *555 pins used in this project*

Number	Name	Purpose
1	GND	Ground, 0 V, goes to the minus terminal of the battery
2	TRIG	If pin 2 (TRIG) goes below 3 V, turns pin 3 (OUT) to 7 V
3	OUT	Either 0 V or 7 V, depending on the 555 state
4	RESET	Keep this HIGH to prevent resetting timer interval

Number	Name	Purpose
6	THRES	If pin 6 (THRES) goes above 6 V, turns pin 3 (OUT) to 0 V
8	VCC	9 V, goes to the plus terminal of the battery

Capacitors

A capacitor (shown in Figure 2-17) stores electricity in an electric field. It has two conductive plates, separated by an insulator. Even the symbol for a capacitor shows this: two lines for the metal plates are separated by a gap.

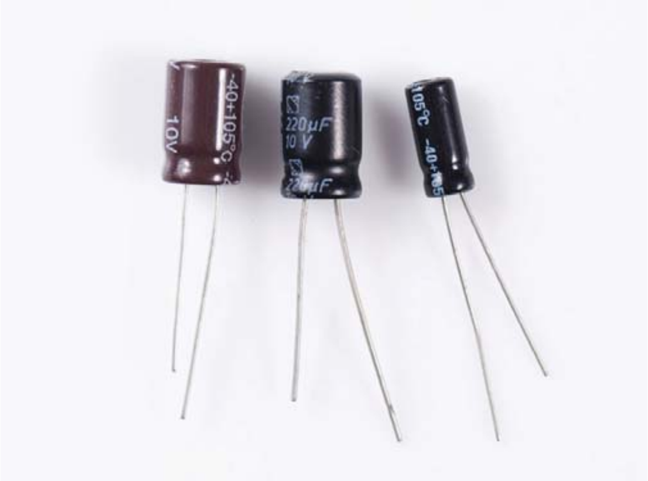

Figure 2-16. *Three capacitors*

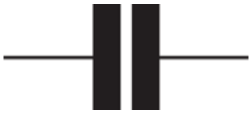

Figure 2-17. *Symbol for capacitor*

Capacitors are used in *resonant circuits*, like in radios. They can also *smooth* spikes in DC power sources (e.g., in the circuitry of an Arduino board). Yet another use is timing events, which is the usage you just worked with in the last project.

The capacitor and resistor you hook up to the 555 timer configure its mode. The values also further affect the specific expression of the configured mode and allow us to make determinations about the overall circuit.

How to Approach an Unfamiliar IC

When you see an IC, how are you supposed to know what it does? Here's something that usually works: read the tiny text on the chip and type it into a search engine. For example, some 555 timers are labeled CA555E and some are labeled NE555. The text on the chip doesn't need to make sense. In fact, the weirder the combination of letters, the more specific hits you'll get.

The datasheet tells you how a chip is used, so you can add that as a search term: "CA555E datasheet" or "NE555 datasheet." You don't need to understand the internals of an IC to use it, and datasheets of more complex chips only explain the connections of the chip. You could also try your luck for step-by-step instructions and search for "CA555E tutorial" or "NE555 tutorial."

These 555 projects wouldn't have been easy to accomplish without a 555 timer. A similar complexity simplification occurs when we move from the 555 timer to a microcontroller, and then again to a system-on-a-chip.

Firefly

Time to build the final circuit for the project. If you skipped here to just build the firefly—good, you can read the theory when it works. If you did every intermediate step, great—you can just combine them with the understanding you have built.

Build the circuit (see Figure 2-18). It simply combines the circuits you have already built: "Fading an LED" on page 24, "555 Fading Blink" on page 25, and "Light to Darkness" on page 21.

Parts

You need the following parts for this project:

- Two BC547 transistors
- 470 Ω resistor (four-band resistor: yellow-violet-brown; five-band resistor: yellow-violet-black-black; the last band will vary depending on the resistor's tolerance)
- 10 kΩ Resistor (four-band resistor: brown-black-orange; five-band resistor: brown-black-black-red; the last band will vary depending on the resistor's tolerance)
- 5 mm red LED
- 9 V battery
- 9 V battery clip
- Breadboard

- CA555E, NE555, or compatible 555 timer integrated circuit
- Potentiometer
- 100 uF capacitor
- Jumper wires
- Light-dependent resistor (LDR)

All of these parts, except the 9 V battery, 470 Ω resistor, and BC547, are available in the Maker Shed Mintronics: Survival Pack, part number MSTIN2.

Build It

Build the circuit as shown in Figure 2-18.

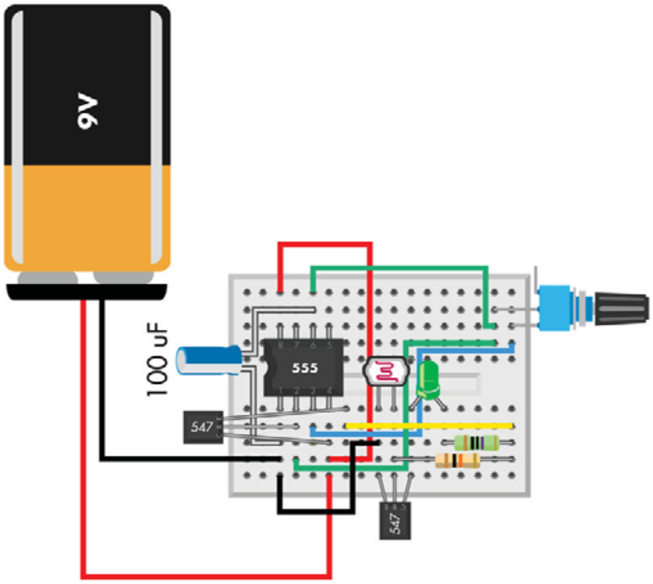

Figure 2-18. *Completed firefly*

Once the circuit is ready, try it out. Cover the LDR with your finger, and the LED gradually lights up, then dims, then lights up again. Point a flashlight at the LDR, and the LED turns off.

Troubleshooting

If this project doesn't work, try checking the following things:

- Double-check the connections.

- Test your LED and battery the way you learned at the beginning of this chapter.
- If you can't find out what's wrong, try the pulsating and photoresistor circuits separately again. This way it's easier to isolate the problem.

What's Next?

Congratulations, you have completed the chapter on component-by-component electronics! The next chapter is about Arduino, so things are just going to get easier. But before you move on, continue experimenting with the analog, component-by-component approach:

- Can you make your firefly light up in the day instead of night?
- Try replacing the potentiometer with a suitable resistor.
- Try replacing the LDR with any analog resistance sensor, so your firefly can react to being squeezed, the temperature, or the amount of alcohol on your breath.
- Try replacing the output with another output, such as a DC piezo speaker.

Enjoy your firefly!

Star Diwa

3/Sensors and Arduino

In this chapter, you will learn how to process sensor data using an Arduino, the wildly popular microcontroller development board. There are many advantages to using Arduino rather than individual components as you did in Chapter 2. One of the biggest is that you'll save time not having to wire up complex circuits to process sensor measurements (see Figure 3-1).

Figure 3-1. *Our Arduino Uno, breadboard, and laptop*

An Arduino runs *sketches*, which are programs that you write and upload to the Arduino from your computer. By writing software programs that make the decisions necessary to process sensor data, you really expand how much you can do, and, amazingly, you also cut down the time you'd have to spend working with hardware components. Sure, you have to write a few lines of code, but it's faster to write a conditional loop in code than to design it in hardware—at least for most of us.

It is also much easier to modify code than it is to modify a circuit layout. For example, suppose you want to add a delay how often a sensor takes measurements. You could accomplish this by chaining a few capacitors in series, but you'll need to do some careful planning and calculation to get the timing right. It's much easier to use a simple circuit and use the Arduino `delay()` function to control the timing. It is more efficient to type a few lines of code and upload a compiled binary program than it is to hunt for a specific component and wire it into a circuit. For example:

```
    delay(1000);  // delays code execution for 1000 milliseconds or 1 second
```

Don't worry if you've never programmed in your life; you'll learn the basics in no time, and the more you tweak your code, the more you will learn. Sketch writing will open a whole new world of opportunities in your electronics projects. Time to get building!

If this is your first time using Arduino, read Appendix B to learn how to install drivers and the Arduino integrated development environment (IDE). It's also important to try the most simple program ("Hello, world") with your new hardware.

As long as the sensors and other components are used within their electrical specifications, using an Arduino or any other programmable microcontroller board will simplify your circuit and facilitate greater flexibility. However, if you use them outside of their specifications, such as applying more voltage to an LED than you should, you risk destroying the components, or even your Arduino.

Project 6: Momentary Push-Button and Pull-Up Resistors

Buttons are not the most exotic components but surely one of the most used and common sensors. Unlike some techniques you'll see for wiring a button to an Arduino, this project does not require an external resistor. That's because you'll be using the pull-up resistor that's internal to the Arduino's general-purpose input/output (GPIO) circuitry that can actually be turned on by a single line of code. Hardware doesn't get much more programmable than this.

Parts

You need the following parts for this project:

- Momentary push button
- Arduino Uno
- Jumper wires

- Breadboard

Build It

Here are the steps for building this project:

1. Orient your breadboard so that it is wider than it is tall, as shown in Figure 3-2.
2. Insert the momentary push button anywhere on the breadboard.
3. In a breadboard column with a button lead in it, plug in a jumper wire and connect the wire's other end to GND on the Arduino.
4. Plug another jumper wire into the same row as the jumper you just connected, but in a column corresponding to the button's other pair of leads. Next, attach the wire's other end to pin 2 on the Arduino.

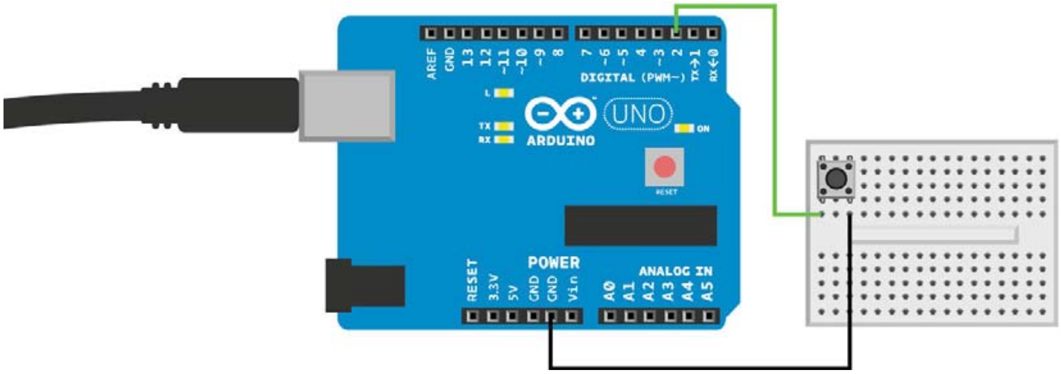
AREF
GND
13
12
~11
~10
~9
8
7
~6
~5
4
~3
2
TX→1
RX←0
DIGITAL (PWM~)
L
TX
RX
UNO
ON
ARDUINO
RESET
POWER
RESET
3.3V
5V
GND
GND
Vin
ANALOG IN
A0
A1
A2
A3
A4
A5

Figure 3-2. *Momentary push button connected to an Arduino Uno*

Run the Code

Once you're done hooking everything up, run the code shown in Example 3-1. When the button is pushed, the code lights up the built-in LED attached to pin 13.

Example 3-1. Arduino sketch for detecting a button press

```
// button.ino - light an LED by pressing button
// (c) BotBook.com - Karvinen, Karvinen, Valtokari

int buttonPin=2; // ❶
int ledPin=13;  // ❷
int buttonStatus=LOW;   // ❸

void setup()    // ❹
{
  pinMode(ledPin, OUTPUT);      // ❺
  pinMode(buttonPin, INPUT); // ❻
  digitalWrite(buttonPin, HIGH); // internal pull-up // ❼
}

void loop() // ❽
{
  buttonStatus=digitalRead(buttonPin); // ❾
  if (LOW==buttonStatus) { // ❿
    digitalWrite(ledPin, HIGH); // ⓫
  } else {      // ⓬
    digitalWrite(ledPin, LOW);  // ⓭
  }
} // ⓮
```

❶ Set up some *variables*. From now on, when you write `buttonPin` in your sketch, Arduino will put the number 2 in its place. Variables make code easier to read and modify because you can change them in a single place. All variables in this sketch are integers (`int`): whole numbers such as 1, 2, 3, or -500.

❷ Create another variable for the LED pin. The built-in LED is on digital pin 13 (`D13`).

❸ Even though the value of `buttonStatus` is modified elsewhere in the program, it's a good idea to initialize it to some value rather than leaving it undefined.

❹ The `setup()` function runs once when Arduino boots up. It contains all the one-time initialization you need in your sketch.

❺ Setting a pin to `OUTPUT` allows you to control it with the `digitalWrite()` function later.

❻ We set `buttonPin` (digital pin 2) to INPUT `mode`. This way, we can later read its value with `digitalRead()`.

❼ Enable the internal pull-up resistor. Magic (hardware being controlled by software) happens here! You'll learn some pull-up resistor theory in "Pull-Up Resistors and Arduino" on page 37. Note that you can accomplish the same thing by replacing this line and the preceding line with `pinMode(buttonPin, INPUT_PULLUP);`, but this will only work with recent versions of the Arduino software; it is not supported by older versions.

❽ Arduino runs the `loop()` function automatically after `setup()` has finished. As the name implies, `loop()` is called again and again until you reset or power down the Arduino.

⑨ Here you read the status of `buttonPin` (which you defined earlier in the sketch as digital pin 2). Then you save this answer to variable `buttonStatus`. If the button is pressed (*closed*), pin 2 is connected to ground and digitalRead returns `LOW`. If the button is not pressed (*open*), pin 2 is not connected to GND, so the internal pull-up resistor pulls D2 up to +5 V, returning HIGH.

⑩ Two equals signs (==) indicate a comparison between two expressions. Note that this is in contrast to a single equals sign (=), which is used for assigning a value to a variable. Always use two equals signs when performing a comparison in an `if` statement. By putting `LOW` on the left side of the comparison, you guard against such an error: LOW is a constant, and you cannot assign a value to it. So if you use a single equals sign, Arduino would give you an error message when you try to compile and load the sketch to your Arduino.

⑪ Here the LED is illuminated.

⑫ The `else` clause handles the case in which the `if` comparison doesn't evaluate to true. The block of code ({}) after the `else` is executed when `LOW != buttonStatus` (which is equivalent to `buttonStatus != LOW`, and means `buttonStatus` is not equal to `LOW`).

⑬ Turn off the LED.

⑭ As soon as `loop()` finishes, Arduino runs it again, starting at the top of the `loop()` function.

You can download all source code from this book's website.

Pull-Up Resistors and Arduino

The momentary push-button design works reliably because you engaged an internal *pull-up resistor* in the Arduino circuitry. You might wonder exactly what a pull-up resistor pulls: it pulls the voltage of the pin toward the 5 volts that power the Arduino.

Never trust a floating pin. If you read a pin that's not connected to anything, you get an unpredictable answer. For example, if you don't connect anything to digital pin D2 and then read it with `digitalRead()`, there are no guarantees as to what you'll get as a result. It could be `HIGH`, it could be `LOW`, it could change 10 times a second or never change. A floating pin gives a useless value (see Figure 3-3).

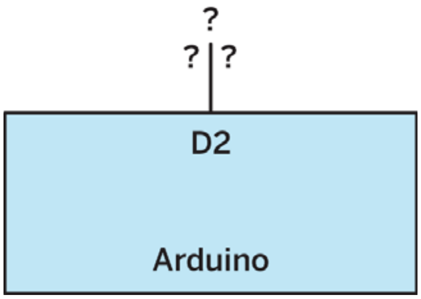

Figure 3-3. *There is no point in reading a floating pin; add a pull-up resistor*

The Arduino IDE comes with many useful example sketches. If you look closely at their button code (File → Examples → 02.Digital → Button) it actually requires an external resistor and won't work with our circuit. That's because that sketch uses a *pull-down* resistor, so a HIGH state indicates that the button is closed (pressed) and a LOW state indicates that the button is open.

If you connect the same D2 to GND (zero volts, ground, LOW), your `digitalRead()` is guaranteed to return a result of LOW (see Figure 3-4). This handles the case where the button is pressed.

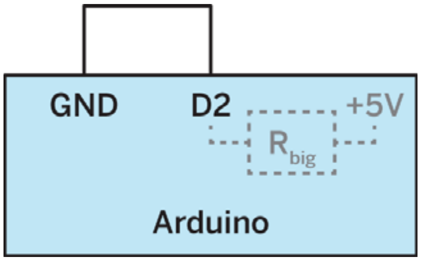

Figure 3-4. *LOW; pressing the button between D2 and GND is the same as connecting D2 to GND*

What if you *don't* press the button? We already know you have to connect a pin somewhere to read its value. But even if you connect a button to Arduino, the pin isn't really connected to anything until you press it.

Pull-up resistor to the rescue!

The pull-up resistor connects D2 to +5 V (HIGH), albeit weakly (because the resistor limits the current). If nothing else is connected to D2, the resistor will pull it up to HIGH. So when the button is open, D2 is not connected to ground, but instead to +5 V through the pull-up resistor, which means that D2 will be HIGH until you press the button (see Figure 3-5).

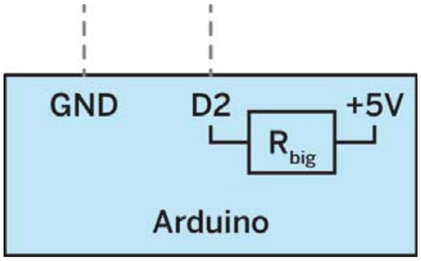

Figure 3-5. *HIGH; button up, D2 is only connected to +5 V through the pull-up resistor*

Now what happens when you press the button? It's clear that D2 and GND are now connected, but so is +5 V, isn't it? You may be wondering, "You are not telling me to connect GND and +5 V, are you? Isn't that a short circuit?"

Although GND and D2 are both connected to +5 V, that connection is through a big resistor, and because the path between GND and D2 has *less* resistance than the resistor that connects to +5 V, the current in the circuit flows through GND and D2. In other words, the pull-up resistor pulls D2 up to HIGH (+5 V) if nothing else is pulling it down.

Typical values for pull-up resistors are tens of thousands of ohms to a few million ohms (e.g., 20 kΩ to 2 MΩ).

This is how your connection works: press the button, D2 goes LOW. Release the button, and the pull-up resistor pulls it up to HIGH. How do you recognize a floating pin in your own connection? If any sensor connects to Arduino with just two leads, ask yourself whether the pin is connected in every possible state the sensor can be in. If not (as in the case with a button that's open), you probably need a pull-up resistor (see Figure 3-6) to hold it HIGH when it's not being pressed.

If you prefer to use an external resistor, it's more common to use it as a pull-down to hold the pin LOW. If you do that, remember to reverse the logic in your code (HIGH indicates a button press, LOW indicates that the button is open).

That's it for the diagrams and theory for now. As you can see, even the humble switch has a lot to teach us about how sensors need to be wired up. With that background out of the way, it's time to relax with some practical, hands-on stuff.

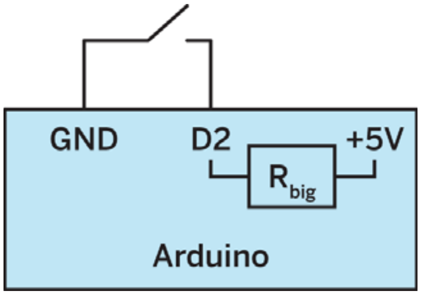

Figure 3-6. *Internal pull-up resistor; enable it with pinMode(pin, INPUT) followed by digitalWrite(pin, HIGH) or just pinMode(pin, INPUT_PULLUP)*

Project 7: Infrared Proximity to Detect Objects

Infrared sensors (such as the switch shown in Figure 3-7) consist of two parts: an infrared emitter and receiver. The emitter is actually an LED that emits light that's invisible to the human eye, as the wavelength is longer than light in the visible spectrum. An infrared LED is the same component that is used in a TV remote control.

The receiver part of the switch collects the IR light that is reflected back. Obstacles cause more light than usual to be reflected, which tells you that there is something in front of the sensor.

This sensor can become confused in strong sunlight, and some very dark materials can pass by unnoticed.

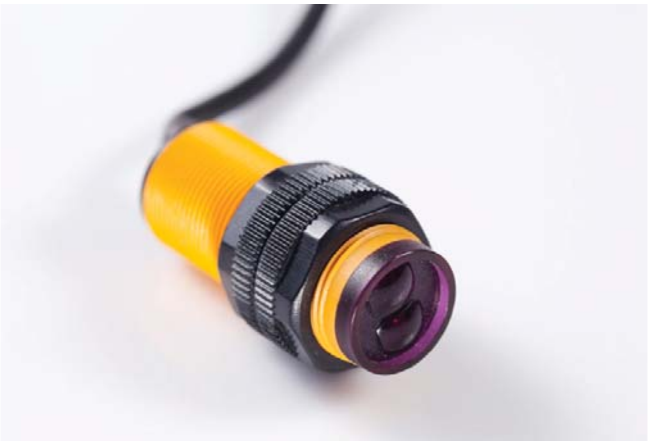

Figure 3-7. *Infrared sensor switch*

You can buy a switch like this from DFRobot (part #SEN0019), as well as from sellers on Amazon, eBay, AliExpress, and DealeXtreme. They are usually listed as "Adjustable Infrared Sensor Switch" or under the manufacturer part number E18-D80NK.

An infrared sensor switch is useful for various projects. It's especially handy when you need to know if something is close, but when you don't need to know the object's exact distance. You can set the detection range of the sensor by adjusting the trimmer potentiometer (trimpot) with a small screwdriver, as shown in Figure 3-8.

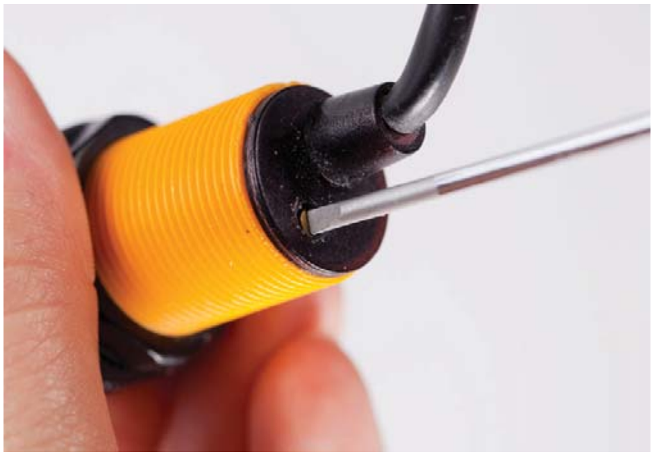

Figure 3-8. *Adjusting the infrared sensor*

Parts

You need the following parts for this project:

- Infrared sensor switch
- Arduino Uno
- Jumper wires

If your infrared switch has wires that terminate in female headers, you'll need to use jumper wires to connect each wire to the Arduino. If it terminates in male headers, you may need to build this circuit using a breadboard and jumper wires.

As long as you have a sensor that breaks out wires for ground, 5 V, and a signal, then this layout will work with your sensor. We used a sensor with separated wires, so we inserted each lead directly into the Arduino headers.

Build It

Make sure the Arduino is powered down. Connect the circuit as shown in Figure 3-9, then run the sketch shown in Example 3-2. When you put something in front of the sensor, the built-in LED attached to pin 13 is lit.

Follow these instructions for connecting the circuit:

1. Connect the yellow (signal) lead to digital pin 8 on the Arduino.
2. Connect the red (positive) lead to the +5 V pin on the Arduino.
3. Connect the green (ground) lead to a GND pin on the Arduino.

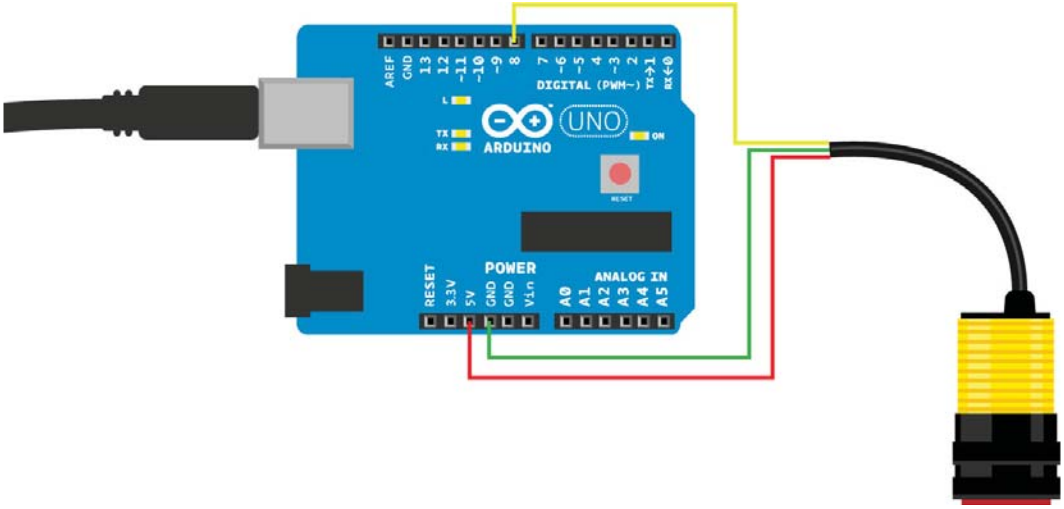
AREF
GND
13
12
~11
~10
~9
8
7
~6
~5
4
~3
2
TX→1
RX←0
DIGITAL (PWM~)
L
TX
RX
ARDUINO
UNO
ON
RESET
POWER
RESET
3.3V
5V
GND
GND
Vin
ANALOG IN
A0
A1
A2
A3
A4
A5

Figure 3-9. *Connecting the infrared sensor switch*

Example 3-2. Infrared distance switch Arduino sketch

```
// infrared_proximity.ino - light LED when object near, using Dagu IR switch
// (c) BotBook.com - Karvinen, Karvinen, Valtokari

int irPin=8; // ❶
int ledPin=13;
int objectDetected=LOW; // ❷

void setup() {
  pinMode(ledPin, OUTPUT);
  pinMode(irPin, INPUT);
  digitalWrite(irPin, HIGH); // internal pull-up
}

void loop() {
  objectDetected=digitalRead(irPin);
  if (LOW==objectDetected) {
    digitalWrite(ledPin, HIGH);
  } else {
    digitalWrite(ledPin, LOW);
  }
}
```

❶ This is similar to the code from Example 3-1, but the name `buttonPin` is changed to `irPin`.

❷ Similarly, the name `buttonStatus` is changed to `objectDetected`.

You know this code already! As you look at it, you might recognize that it's the same code you used back in Example 3-1. We changed two variable names, but that is largely cosmetic. Experiment with it yourself: can you use the infrared proximity sensor with the code from Example 3-1? Can you use a button with the code from Example 3-2?

Many sensors have identical interfaces. Now that you know how to work with two switches, which we refer to as *digital resistance sensors*, you can apply this skill to similar sensors in your own projects.

Analog Resistance Sensors and Voltage Dividers

Most sensors you'll use are analog resistance sensors. They produce their results with a gradual change in resistance. For example, the more force you exert on a force-sensing resistor like a FlexiForce, the lower its resistance.

Because you can't measure resistance directly with Arduino (or Raspberry Pi), you'll learn how to use these types of sensors with an external resistor to form a *voltage divider*. A voltage divider looks a lot like a pull-up or pull-down resistor configuration, but instead uses two resistors in series between +5 V and GND:

- The external resistor
- The sensor, whose resistance varies depending on its state

By using an analog pin to measure the voltage between these two resistors, you'll receive a voltage that varies in proportion to the state of the sensor.

Project 8: Rotation (Pot)

The potentiometer (shown in Figure 3-10) is an adjustable resistor with a knob. When you turn the knob, its resistance changes. Potentiometers are very common in everyday items such as radios and clothes irons. In RC airplanes, servomotors can determine their angular position with the help of a potentiometer.

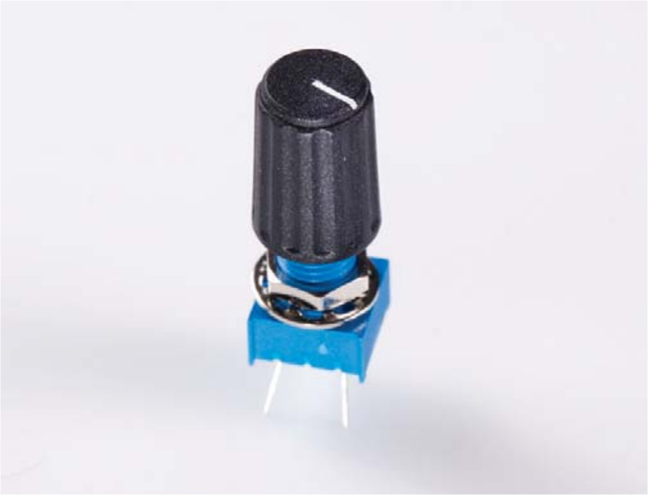

Figure 3-10. *Potentiometers can measure rotation*

To some extent, analog resistance sensors can behave like digital ones, provided they meet two criteria:

- At their lowest level of resistance, they must let enough current through to register as HIGH.
- At their highest level of resistance, they must resist enough current to register as LOW.

Try it yourself! Return to "Project 6: Momentary Push-Button and Pull-Up Resistors" on page 34, hook everything up, and then run the code. Then, replace the button with the potentiometer: one side of the potentiometer to +5 V, the other to GND, and the middle pin to digital pin 2. Now, turn the knob to max or min. If the resistance of the potentiometer is sufficient, you can turn the LED on and off.

Because the code is designed to read a button, you can only turn the LED on or off. To make use of the analog values produced by the potentiometer, you must use the `analogRead()` function. When used with analog sensors, this function gives you a range of values (from 0 to 1023), rather than just HIGH or LOW.

This project controls an LED's blinking speed with a potentiometer. Turn it all the way in one direction, and the LED blinks slowly. Turn it all the way in the other direction, and it blinks faster.

Figure 3-11 shows the wiring diagram for this project.

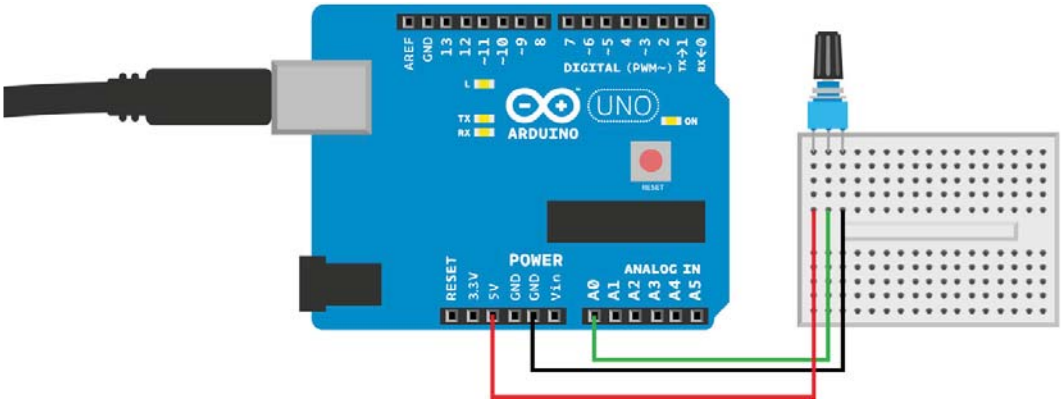

Figure 3-11. *Potentiometer wired to GND and A0 of an Arduino Uno*

The potentiometer acts as a voltage divider. One side is connected to +5 V, and the other to GND. The center pin is attached to analog input pin A0. As you turn the knob, the resistance on either side of the pin varies. For example, for a 10 kΩ potentiometer (all numbers approximate because inexpensive potentiometers are not 100% precise):

When the knob is turned all the way to one side[1]

- The resistance between the positive and center pin is zero.
- The resistance between the center and negative pin is 10 kΩ.
- Pin A0 receives a voltage of 5 V.
- `analogRead(potPin)` returns 1023.

When the knob is all the way to the other side

- The resistance between the positive and center pin is 10 kΩ.
- The resistance between the center and negative pin is zero.
- Pin A0 receives a voltage of 0 V.
- `analogRead(potPin)` returns 0.

When the knob is centered

- The resistance between the positive and center pin is 5 kΩ.
- The resistance between the center and negative pin is 5 kΩ.
- Pin A0 receives a voltage of 2.5 V.
- `analogRead(potPin)` returns approximately 512.

Parts

You need the following parts for this project:

---

[1] Which side depends on how you insert the pot into the breadboard; try it both ways!

- Potentiometer (around 10 kΩ recommended)
- Arduino Uno
- Jumper wires
- Breadboard

Build It

Wire the components up as shown in Figure 3-11.

Run the Code

Next, run the code listed in Example 3-3.

Example 3-3. Arduino sketch for reading a potentiometer

```
// pot.ino - control LED blinking speed with potentiometer
// (c) BotBook.com - Karvinen, Karvinen, Valtokari

int potPin=A0; // ❶
int ledPin=13;  // ❷
int x=0; // 0..1023 // ❸

void setup() {  // ❹
  pinMode(ledPin, OUTPUT);      // ❺
}

void loop() {   // ❻
  x=analogRead(potPin); // ❼
  digitalWrite(ledPin, HIGH);   // ❽
  delay(x/10); // ❾
  digitalWrite(ledPin, LOW);    // ❿
  delay(x/10);  // ⓫
} // ⓬
```

❶ In earlier examples, we used Arduino's digital pins, and we referred to them just by number (such as `13` for digital pin 13). The analog input pins are also numbered, but we put the letter `A` in front of the name when referring to them (`A0`, `A1`, etc.). The analog pins are on the opposite side of Arduino's digital pins.

❷ Digital pin 13 is connected to the Arduino's built-in LED.

❸ `x` holds raw values that `analogRead()` returns. Note that we use the comment to show the expected range of values. Even though you can find this information in the Arduino documentation, it's handy to have here.

❹ `setup()` performs one-time initialization.

❺ Set the pin D13 to OUTPUT, so we can control it later with `digitalWrite()`. Note that although you'll be reading A0 later, it's not necessary to explicitly configure it as an INPUT, because `analogRead()` behaves differently than `digitalRead()`.

❻ After `setup()` is finished, `loop()` is called automatically.

❼ Read the voltage from `potPin` (A0) with `analogRead()`. You'll receive a value between 0 (0 V == LOW) and 1023 (+5 V == HIGH). The code saves this value to variable `x`.

❽ Turn on the onboard LED.

❾ We'll wait for x/10 milliseconds. That's 0 when the knob is turned to the max (A0 reads 0 V). When the knob is turned the other way, the delay is 0.1 seconds (1023/10 ms).

❿ Turn off the onboard LED.

⓫ Wait again, so that the LED remains off briefly.

⓬ After `loop()` finishes, it's called automatically again.

Project 9: Photoresistor to Measure Light

How bright is it? A photoresistor lets more electricity through when it's in bright light (see Figure 3-12). Our students have used photoresistors to make robots love (or hate) light, activate lights in the dark, or create burglar alarms.

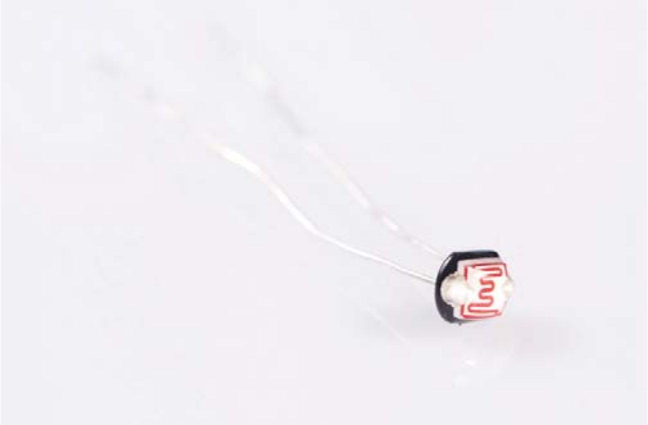

Figure 3-12. *A photoresistor is a better conductor in bright light*

Parts

You need the following parts for this project:

- Photoresistor (10 K recommended)

- 10 kΩ resistor (four-band resistor: brown-black-orange; five-band resistor: brown-black-black-red. The fourth or fifth band will vary depending on the resistor's tolerance)
- Arduino Uno
- Jumper wires
- Breadboard

Build It

Figure 3-13 shows the circuit design for this project. Build it as shown, and run the sketch listed in Example 3-4. Note the addition of the 10K resistor. This is to create a voltage divider. Back in "Project 8: Rotation (Pot)" on page 43, you didn't need a separate resistor because the potentiometer can be used as a voltage divider.

In order to produce a voltage that varies in proportion to the photoresistor's change in resistance, you need to create a voltage divider with the two resistors (the 10K resistor, and the photoresistor) and measure the voltage that's produced between them to determine how dark or light it is.

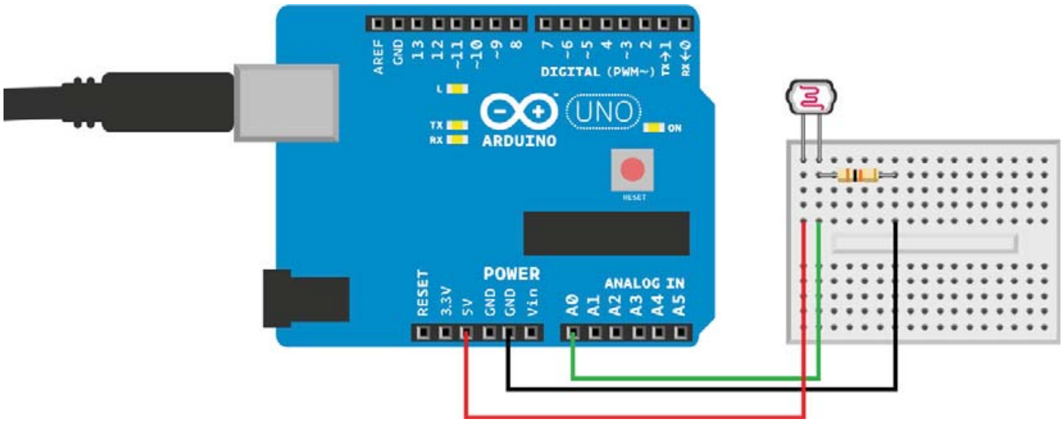
AREF
GND
13
12
~11
~10
~9
8
7
~6
~5
4
~3
2
TX→1
RX←0
DIGITAL (PWM~)
L
TX
RX
UNO
ARDUINO
ON
RESET
POWER
RESET
3.3V
5V
GND
GND
Vin
ANALOG IN
A0
A1
A2
A3
A4
A5

Figure 3-13. *Photoresistor circuit layout*

Run the Code

Example 3-4. Arduino photoresistor sketch

```
// photoresistor.ino - blink faster in dark, slower in the light
// (c) BotBook.com - Karvinen, Karvinen, Valtokari

int photoPin=A0;
int ledPin=13;
int x=-1; // 0..1023

void setup() {
  pinMode(ledPin, OUTPUT);
}

void loop() {
  x=analogRead(photoPin);
  digitalWrite(ledPin, HIGH);
  delay(x/10); //
  digitalWrite(ledPin, LOW);
  delay(x/10);
}
```

Hey, wait a minute; this looks a lot like Example 3-3! It's exactly the same code. We did replace "potPin" with "photoPin" and changed the comment on the first line. Run the program and try it out. What happens when you make a shadow or hide the photoresistor completely?

Most common sensors are analog resistance sensors, so they use a variation of this same connection. You can apply this same approach to similar sensors you meet in the future.

Project 10: FlexiForce to Measure Pressure

A force-sensitive resistor such as the FlexiForce (see Figure 3-14) is a flat sensor that measures how much pressure you exert on it. Untouched, without any pressure, it has a very high resistance. When you press the round area at the end of the sensor, the resistance drops and it becomes more conductive.

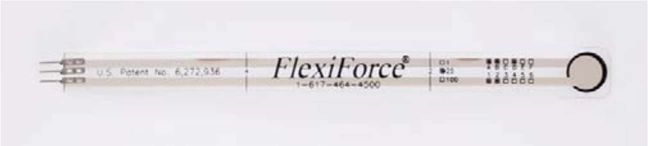

Figure 3-14. *A force-sensitive resistor measures pressure exerted on it*

Parts

You need the following parts for this project:

- Force-sensitive resistor (FSR)
- 1 MΩ resistor (four-band resistor: brown-black-green; five-band resistor: brown-black-black-yellow. The fourth or fifth band will vary depending on the resistor's tolerance)
- Arduino Uno
- Jumper wires
- Breadboard

Build It

You'll use a voltage divider here just as you did in "Project 9: Photoresistor to Measure Light" on page 47. As with that project, you need to use a resistor that's in the same range as the FlexiForce. We found that a 1 MΩ resistor produced suitable results.

In our voltage divider, analog pin A0 measures the variation in voltage ("Analog Resistance Sensors and Voltage Dividers" on page 43) as you change the amount of force exerted on the sensor, which gives you a value between 0 (voltage is near GND) to 1023 (voltage is near +5 V). Build the circuit as shown in Figure 3-15, and then run the sketch from Example 3-5.

Watching the Action Over the Serial Port

You don't need to guess what's going on inside your Arduino's brain. Arduino can talk to your computer over its built-in serial port and send text back to it (you can even send messages from the computer to Arduino), as shown in Figure 3-16. The Arduino uses its USB port to transfer serial messages.

Serial ports are a very common way for embedded devices to talk. With our book on controlling robots with your brainwaves (*Make a Mind-Controlled Arduino Robot*), we used a serial-over-USB cable to hack a commercial, off-the-shelf EEG headband. This allowed us to control the Arduino-based robot with our brainwaves! Another great example is our mobile phone controlled football robot from our book, *Make: Arduino Bots and Gadgets*. In that project, we used serial over Bluetooth, allowing the Arduino to communicate with an Android phone. But serial is not just used for communicating directions to devices. When building an early prototype of our satellite, we used serial-over-USB to debug our project. In fact, that's what we'll do in this project, too.

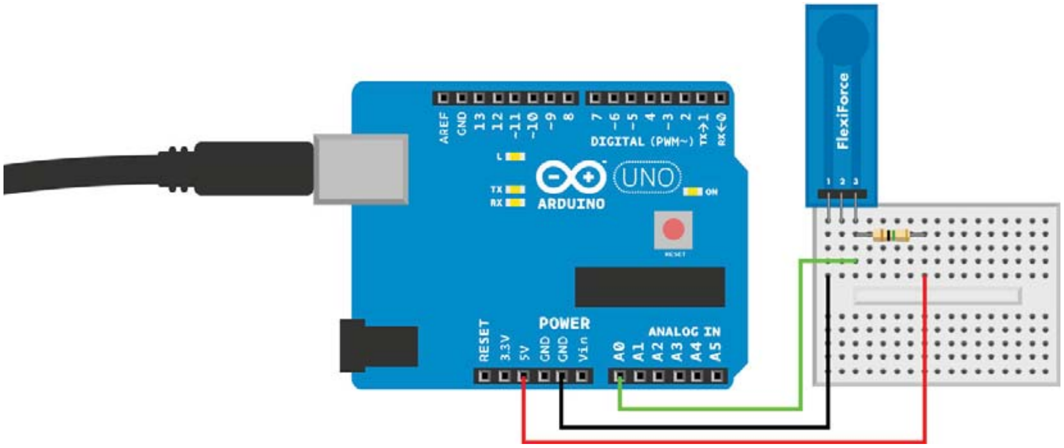
AREF
GND
13
12
~11
~10
~9
8
7
~6
~5
4
~3
2
TX→1
RX←0
DIGITAL (PWM~)
L
TX
RX
UNO
ARDUINO
ON
RESET
POWER
RESET
3.3V
5V
GND
GND
Vin
ANALOG IN
A0
A1
A2
A3
A4
A5
FlexiForce
1 2 3

Figure 3-15. *FlexiForce circuit layout*

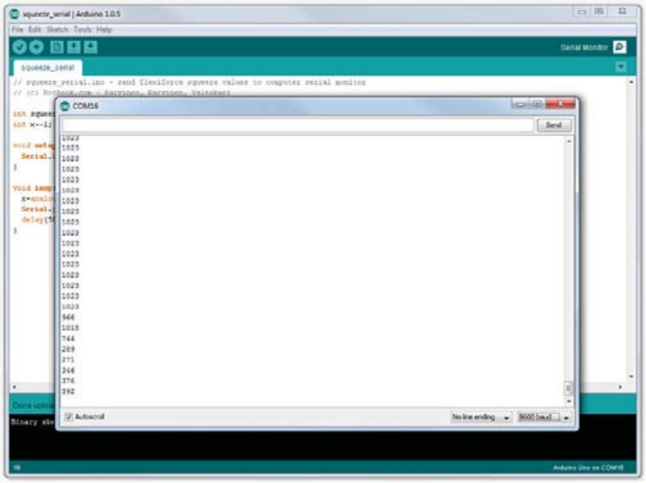
squeeze_serial | Arduino 1.0.5
File Edit Sketch Tools Help
Serial Monitor
squeeze_serial
COM16
Send
Arduino Uno on COM16

```
// squeeze_serial.ino - send flexiforce squeeze values to computer serial monitor
// (c) BotBook.com - Karvinen, Karvinen, Valtokari
```

```
1023
1023
1023
1023
1023
1023
1023
1023
1023
1023
1023
1023
1023
1023
1023
1023
966
1018
744
289
271
344
376
392
```

☑ Autoscroll

No line ending

9600 baud

Figure 3-16. *Serial Monitor*

Run the Code

Example 3-5. Arduino sketch for reading the FlexiForce sensor

```
// squeeze_serial.ino - flexiforce squeeze level to serial
// (c) BotBook.com - Karvinen, Karvinen, Valtokari

int squeezePin=A0;      // ❶
int x=-1; // 0..1023   // ❷

void setup() {  // ❸
  Serial.begin(9600); // bit/s // ❹
}

void loop() {   // ❺
  x=analogRead(squeezePin);  // ❻
  Serial.println(x);  // ❼
  delay(500); // ms     // ❽
}
```

❶ This is the pin that reads the sensor (analog pin 0).

❷ Variable `x` is global; it's declared outside any functions, so it would be available to any other functions you might add to your sketch. We initialize it to a value (-1) that you'd never get from `analogRead()` to make debugging easier: if you see this value in the results, then you would know there's a problem.

❸ The `setup()` function is called automatically; it handles one-time initialization.

❹ The serial port speed (we chose 9600 because that's what the Arduino Serial Monitor uses by default; if you've changed the setting on the Serial Monitor, you will need to change it back to 9600).

❺ The `loop()` function is called automatically (and repeatedly) after `setup()` is finished.

❻ `analogRead()` returns a value between 0 (LOW, 0 V) and 1023 (HIGH, +5 V). This value is stored to variable `x`.

❼ This prints out the value of the variable to the serial port so that it appears in the Serial Monitor. `Serial.println()` prints each value on its own line, so when you open the Serial Monitor, you'll see each value scrolling past in the window.

❽ Wait for a while: this gives you time to observe each value in the Serial Monitor, and prevents your code from consuming 100% of the Arduino's CPU time.

Experiment yourself: can you connect a potentiometer and read its current value through serial? If you need to challenge yourself, you can try converting rotation to angle in degrees or a percentage of maximum rotation.

Are you feeling really adventurous? Once you get your data to your computer, you can read it in any programming language. In our interactive painting from our book *Make: Arduino Bots and Gadgets*, we use Python's serial library to read values from the Arduino and control images on a computer.

Project 11: Measuring Temperature (LM35)

Is it hot in here? The LM35 is an inexpensive thermometer that's also easy to use (see Figure 3-17). You just need to retrieve the voltage with `analogRead()`, and then you can calculate the temperature in degrees Celsius as `voltage * 100C/V`.

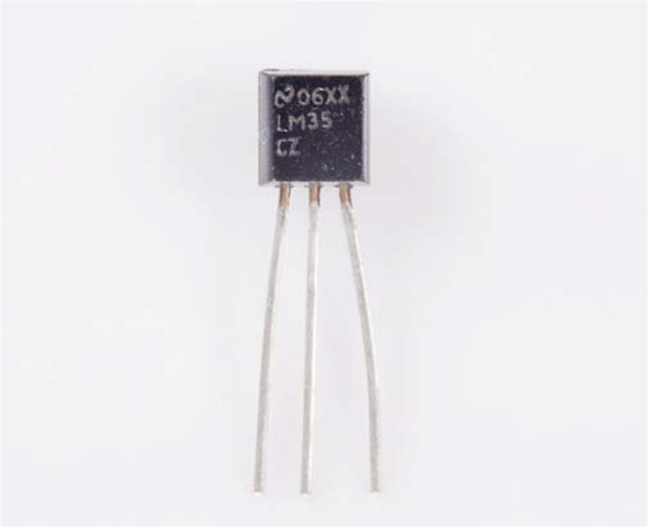

Figure 3-17. *LM35 temperature sensor*

Our students have used the LM35 to measure temperature in their summer cottages and greenhouses, for example.

Parts

You need the following parts for this project:

- LM35 temperature sensor
- Arduino Uno
- Jumper wires
- Breadboard

The LM35 differs from previous analog resistance sensors you have used. It has three leads: 5 V, ground, and signal. In this way, it's easier to wire than others because you don't need a pull-up resistor.

Build It

Figure 3-18 shows the circuit diagram. Build it as shown, and then run the code listed in Example 3-6. Open the Serial Monitor to see the current temperature. You'll need to follow these steps:

1. Connect the LM35 sensor to Arduino using a breadboard. If you're using the most common version of the LM35 (the TO-92 package, which is a common semiconductor package for transistors), hold the leads down

and orient the flat side toward you. The left lead is +5 V, middle is VOUT, and the right lead is GND.

2. Connect +5 V to Arduino's +5 V and GND to Arduino's GND. But what is VOUT? Voltage OUT indicates the temperature as an output voltage, so you need to connect it to an analog input pin on Arduino (Example 3-6 uses A0).

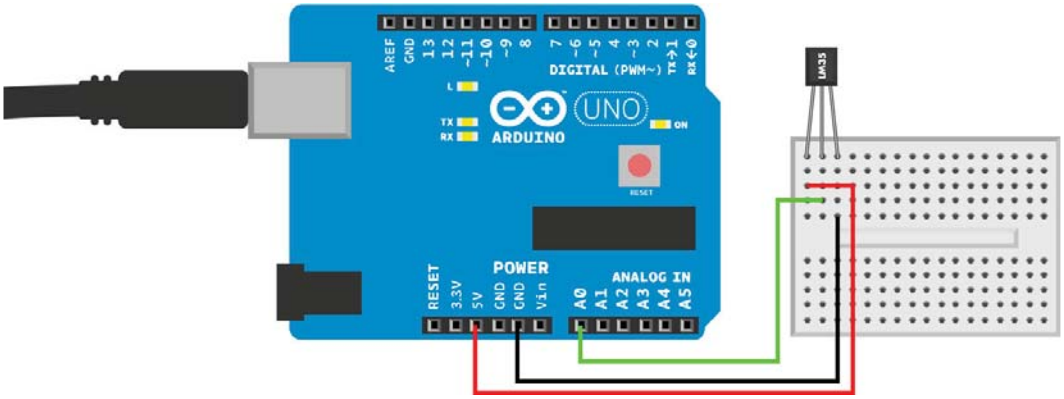
AREF
GND
13
12
~11
~10
~9
8
7
~6
~5
4
~3
2
TX→1
RX←0
DIGITAL (PWM~)
L
TX
RX
ARDUINO
UNO
ON
RESET
POWER
ANALOG IN
RESET
3.3V
5V
GND
GND
Vin
A0
A1
A2
A3
A4
A5
LM35

Figure 3-18. *Temperature sensor connected using breadboard*

Once you've measured the room temperature, you might want to test the LM35 some more. If you put it in your fridge, prepare to wait. Air is a very good insulator for heat and doesn't conduct well. When we want to quickly test cooler temperatures, we use ice cubes (see Figure 3-19). Wet ice cubes quickly conduct heat away from the LM35, so we get instant readings.

Figure 3-19. *Testing the LM35 with an ice cube*

Run the Code

Example 3-6. The LM35 Arduino circuit

```
// temperature_lm35.ino - LM35 temperature in Celsius to serial
// (c) BotBook.com - Karvinen, Karvinen, Valtokari

int lmPin = A0; // ❶

void setup()    // ❷
{
  Serial.begin(9600);   // ❸
}

float tempC()   // ❹
{
  float raw = analogRead(lmPin);  // ❺
  float percent = raw/1023.0; // ❻
  float volts = percent*5.0; // ❼
  return 100.0*volts;  // ❽
}

void loop()     // ❾
{
  Serial.println(tempC());  // ❿
  delay(200); // ms     // ⓫
}
```

❶ We use analog pin 0 (A0) to read the output of the sensor.

❷ When the Arduino boots up, `setup()` gets called automatically to perform one-time initialization.

❸ Opens the serial port. You can monitor the Arduino's serial port output by going to Arduino → Tools → Serial Monitor. Be sure to select the same speed in both this code (it's 9600 here) and in the Serial Monitor.

❹ Define a new function called `tempC()`. You'll call this function later in the code when you need to read the temperature.

❺ The function `analogRead()` returns an integer value from 0 to 1023. A value of 0 means 0 V (LOW), and a value of 1023 means +5 V (HIGH).

❻ We convert the raw reading (0-1023) to a percentage by dividing it by the maximum value. The percentage will be 0.0 for 0% and 1.0 for 100%. To get the floating-point (decimal) answer you need, you must use the floating-point divisor 1023.0 instead of integer 1023. In C and C++, 1/2==0 (integer division), but 1/2.0==0.5 (floating-point division).

❼ The `percent` variable is calculated out of a maximum of 5.0 V (HIGH). So multiplying by 5.0 gives the voltage in volts.

❽ From the LM35 datasheet (search the Web for “lm35 datasheet”), you can see that the scaling factor is 10 mV/C, with 0 V meaning 0 °C. Playing with the math, 10 mV/C = 0.010 V/C. The inverse is 1/0.010 C/V = 100 C/V. So you can get the temperature in °C by multiplying the voltage by 100.

❾ After `setup()` finishes, `loop()` is called automatically, over and over until you power down the Arduino.

❿ Call the `tempC()` function, which returns a number (e.g., 20). Print this number and a line break to the Serial Monitor. Writing `tempC()` as a function makes it easy to use it in our other projects. Now that it works, you don’t have to think about the implementation details. When you want to know the temperature, just connect the LM35 and call `tempC()`.

⓫ Pause for a while. This gives you time to read the text, keeps the serial buffer from overflowing, and prevents the program from hogging 100% of the Arduino’s CPU.

Project 12: Ultrasonic Distance Measuring (HC-SR04)

The ultrasonic distance sensor (shown in Figure 3-20) is one of our favorite sensors. It sends an ultrasonic sound and then listens for how quickly it returns, like a tiny sonar. As the name implies, the sound frequency is too high for human ears to hear. The great thing compared to the infrared sensor switch you saw earlier (“Project 7: Infrared Proximity to Detect Objects” on page 40) is that this sensor actually gives you the distance to the obstacle (you can find infrared distance sensors, such as the Sharp GP2Y0A series, that report distance as well).

Nothing is without trade-offs, so while you can get a distance reading, the ultrasonic distance sensor can miss certain objects easily. Items that are especially soft and inclined planes tend to go unnoticed. This is one reason that stealth aircraft are shaped the way they are.

There are many ultrasonic distance sensors. The Parallax PING used to be our favorite, but its cost of about $30 limited how many we could buy. Now there are ultrasonic distance sensors available for a couple of dollars each. The HC-SR04 is one of these.

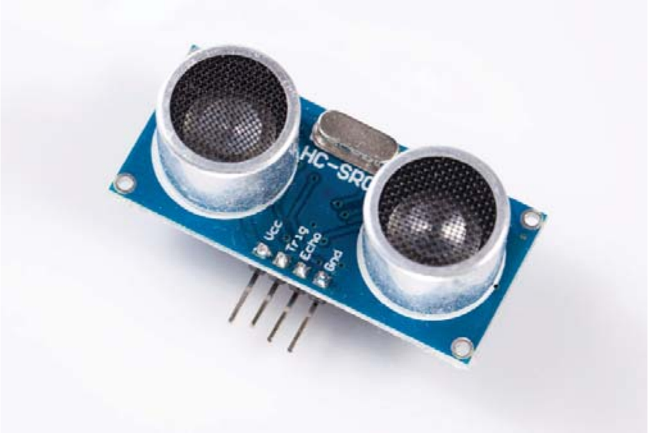
HC-SR0
Vcc
Trig
Echo
Gnd

Figure 3-20. *Ping ultrasonic distance sensor*

Parts

You need the following parts for this project:

- HC-SR04 ultrasonic distance sensor
- Arduino Uno
- Breadboard
- Jumper wires

Build It

Wire up the sensor as shown in Figure 3-21 and run the code in Example 3-7.

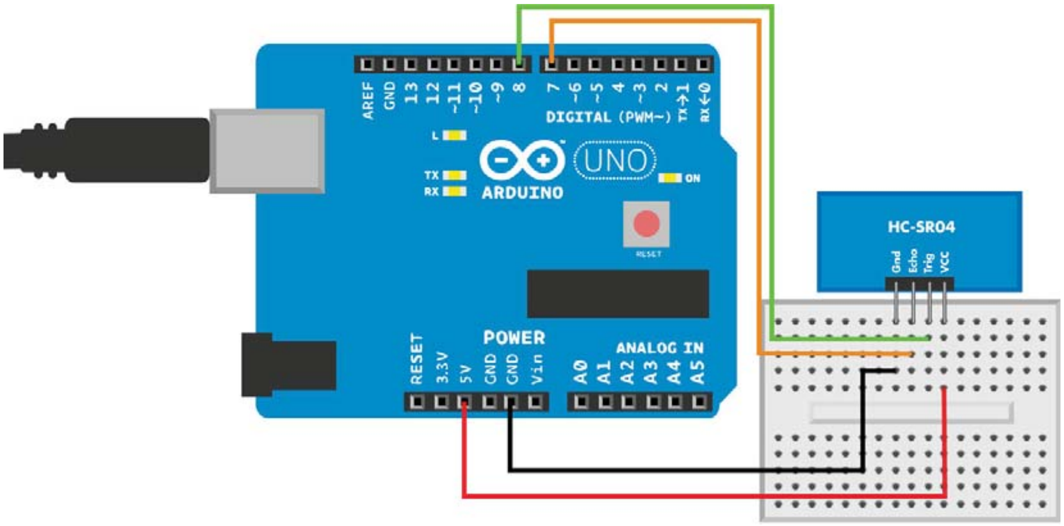
AREF
GND
13
12
~11
~10
~9
8
7
~6
~5
4
~3
2
TX→1
RX←0
DIGITAL (PWM~)
L
TX
RX
UNO
ARDUINO
ON
RESET
POWER
RESET
3.3V
5V
GND
GND
Vin
ANALOG IN
A0
A1
A2
A3
A4
A5
HC-SR04
Gnd
Echo
Trig
VCC

Figure 3-21. *HC-SR04 connected with breadboard*

Run the Code

Example 3-7. The HC-SR04 Arduino sketch

```
// hc-sr04.ino - distance using ultrasonic ping sensor
// (c) BotBook.com - Karvinen, Karvinen, Valtokari


int trigPin = 8;        // ①
int echoPin = 7;        // ②
float v=331.5+0.6*20; // m/s    // ③

void setup()    // ④
{
  Serial.begin(9600);   // ⑤
  pinMode(trigPin, OUTPUT);     // ⑥
  pinMode(echoPin, INPUT);      // ⑦
}

float distanceCm(){     // ⑧
  // send sound pulse   // ⑨
  digitalWrite(trigPin, LOW);   // ⑩
  delayMicroseconds(3); // ⑪
  digitalWrite(trigPin, HIGH);  // ⑫
  delayMicroseconds(5); // ⑬
  digitalWrite(trigPin, LOW);   // ⑭

  // listen for echo    // ⑮
  float tUs = pulseIn(echoPin, HIGH); // microseconds   // ⑯
  float t = tUs / 1000.0 / 1000.0 / 2.0; // s   // ⑰
  float d = t*v; // m   // ⑱
  return d*100; // cm   // ⑲
}

void loop()     // ⑳
{
  int d=distanceCm(); // ㉑
  Serial.println(d, DEC);       // ㉒
  delay(200); // ms     // ㉓
}
```

① You use the trigger pin to tell the sensor to transmit the sound pulse. It is marked with "Trig" on the HC-SR04 board. It's connected to Arduino digital pin D8, and when you take the pin HIGH, the pulse will be triggered.

② The echo pin goes high when the pulse returns. On HC-SR04, it's marked as "Echo." On the Arduino, it's connected to digital pin D7.

❸ The speed of sound is about 340 meters per second. It's calculated here for 20 °C air temperature (we add `20 * .6` to `331.5`). We always spell out calculations in our code for clarity. It doesn't add any extra work to Arduino, because the compiler turns the calculation into a constant before uploading your code to the Arduino.

❹ The `setup()` function is automatically run once when Arduino boots.

❺ Open the serial connection so that the result can be displayed to the Serial Monitor (Tools → Serial Monitor). Remember to set the same speed (bit/second, or "baud") in both your code and the Arduino Serial Monitor.

❻ Set the Trig pin to OUTPUT so `digitalWrite()` can later turn it on and off (HIGH or LOW).

❼ Set the Echo pin to INPUT so that `digitalRead()` can later read it to find out whether it's HIGH or LOW.

❽ Define a new function. This function `distanceCm()` returns a float, a decimal number. It takes no parameters, so there is nothing inside the parentheses after the name.

❾ There are two separate parts to reading the distance. First, we ask the HC-SR04 to send the pulse, and later, we check to see how long it took for the echo to come back.

❿ To create a clean pulse for sending the sound, the pin must be off (LOW) first. Music starts from silence.

⓫ Wait a very short time for the pin to settle. When working with different multiples of units, you should convert them to base units such as seconds to get an idea of the scale. Three microseconds is three millionths of a second, or 0.000003 seconds.

⓬ Turn the trigger pin on (HIGH, +5 V). The HC-SR04 is now transmitting the ultrasound.

⓭ Wait for a very short time, 5 ms (microseconds). Five microseconds == five millionths of a second == 0.000005 seconds. The trigger pin stays high during this time.

⓮ Turn the trigger pin off, ending the short sound pulse.

⓯ Use a comment to mark that we're in a different part of the function.

⓰ Now you need to find out how long it takes for the echo pin to go LOW. The time returned by the `pulseIn()` function is reported in microseconds. It's usually best to use plain ASCII characters in source code, so we indicate micro (millionth) with a plain *u* or *U* instead of the Greek letter µ.

⓱ Convert microseconds (millionths) to seconds by dividing. It's often practical to use SI base units, such as seconds, meters, or kilograms. To get decimal numbers and avoid premature rounding, we use a floating point in the divider (1000.0 instead of integer 1000). The time *t* is a round trip (there and back, two way, ping and pong), so we divide it by two to get one-way time.

⓲ To get distance from time, we multiply by speed. For example, if it took us 2 hours to get to Helsinki and back at 100 km/h, we'd calculate that distance with `2h/2*100km/h`, which is 100 km.

⓳ To get centimeters, we multiply meters by 100. This is a convenient unit for distances typically measured with HC-SR04.

⓴ After `setup()` finishes, `loop()` is called automatically (over and over). The `loop()` function is what your Arduino is doing most of the time.

㉑ This is all you need in your main program to use HC-SR04. Because distanceCm() is written as its own function, you can easily add it to your own projects. Even though it required some thought to write, we can forget the implementation details when we use it. To get distance, just attach the HC-SR04 sensor, include the function in your code, and call `distanceCm()`. Here, we cast (convert) the result to an integer, ignoring the decimal part. So `d` will now contain the distance in full centimeters (e.g., 23).

㉒ Print the distance to the Serial Monitor (Tools → Serial Monitor). The command `println()` prints the value of `d` as a decimal number, then a line break.

㉓ When running in an infinite loop, always wait for a while to avoid taking 100% of the Arduino's CPU time. This also gives you time to read the results in the Serial Monitor so it's not scrolling by too fast.

Experiment yourself! Can you evade the ultrasound? What surface reflects sound so that distance measuring works? What surfaces absorb or misdirect the sound, so that you become invisible to this sonar?

Looking for more of a challenge? Modify the `distanceCm()` function to take the digital pin as parameter (e.g., `distanceCm(3)` for HC-SR04 on digital pin 3). You can even connect multiple HC-SR04 sensors to the same Arduino.

Many projects use PING, HC-SR04, and similar ultrasonic distance sensors. We have used them in an interactive painting and a walking robot insect. Our students have used them in robots, a gesture mouse, burglar alarms, and interactive fluffy toys.

Conclusion

You have learned to write Arduino programs and measure the world. Your Arduino can sense pressure, temperature, light, and many other things. If you want to keep working with the Arduino to build robots and design your own devices, see our other book, *Make: Arduino Bots and Gadgets*.

Because the Arduino is simple and robust, it's a tool we always keep on hand for prototyping. But you might also want to see a platform that can run big programs (e.g., web servers) and connect big things like video projectors or televisions. Prepare to swim to the deep end, and advance to Raspberry Pi in the next chapter.

Star Diwa

4/Sensors and the Raspberry Pi

Do you wish there was a $35 Linux computer with full HD display?

In this chapter, you will process sensor data using a Raspberry Pi, an inexpensive credit card-sized computer that runs Linux. The Pi is a great board to use as a multimedia server, a safe device to try random acts of programming on, and a fun gizmo to learn the Linux command line with.

Thanks to the Pi's general-purpose input/output (GPIO) pins, you can also wire up your own custom circuits. Some of the benefits of Raspberry Pi are:

- Cheap ($35 US)
- Connects to a full HD display
- Connects to the Internet
- Runs Linux, the free operating system
- Can run full versions of Apache web server and SSH server

Unlike Arduino, the Pi does not have dedicated pins for analog input, but you'll learn how to get around that limitation and add analog input using a special integrated circuit, the MCP3200.

You'll be able to connect all the same sensors you worked with in the Arduino chapter, but the steps to wire and program them are different.

You can start from the very beginning. To install "Hello, world," learn the command-line interface, and control an LED with Raspberry Pi, see Appendix C. Once your "Hello, world" works, return here to play with a button.

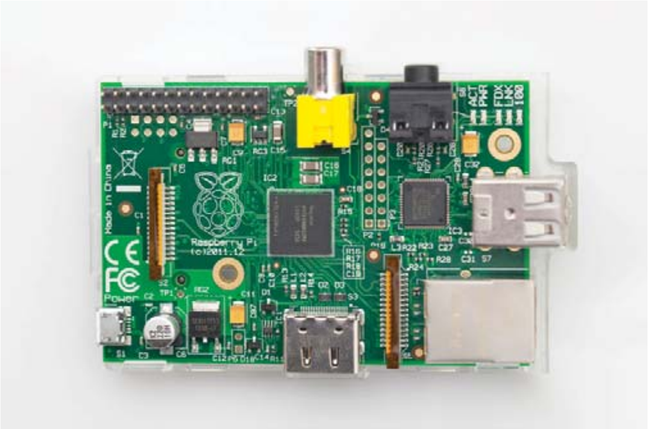

Figure 4-1. *Raspberry Pi*

Project 13: Momentary Push Button

Yes, the button strikes again! Many digital sensors behave like a button. They are on or off, 1 or 0, true or false. In this project and "Project 15: Adjustable Infrared Switch" on page 73, you'll learn to use a button and an infrared proximity sensor with Raspberry Pi.

Parts

You need the following parts for this project:

- Momentary push button
- Raspberry Pi
- Female-male jumper wires, black and green
- Breadboard

Before you build any circuit on your Raspberry Pi, you need to shut down your Raspberry Pi and disconnect it from a power source. To shut it down, click the shutdown button in the lower right of the screen, and choose the shutdown option. You can also type `sudo halt` from a terminal (LXTerminal) to begin the shutdown procedure. Wait for the shutdown process to complete before unplugging your Raspberry Pi. Double- and triple-check all connections you made before powering it up again.

Build It

Build the circuit as shown in Figure 4-2. You'll need to follow these steps:

1. Orient the breadboard as shown.
2. Insert the momentary push button anywhere on the breadboard.
3. Choose a breadboard column with a pair of button leads in it. Plug the male end of a jumper wire into that row, and connect the wire's other end to GND (pin 6 in Figure 4-2) on the Raspberry Pi.
4. Plug another jumper into a breadboard column with the other pair of button leads, and attach the wire's other end to pin 5 on the Raspberry Pi.

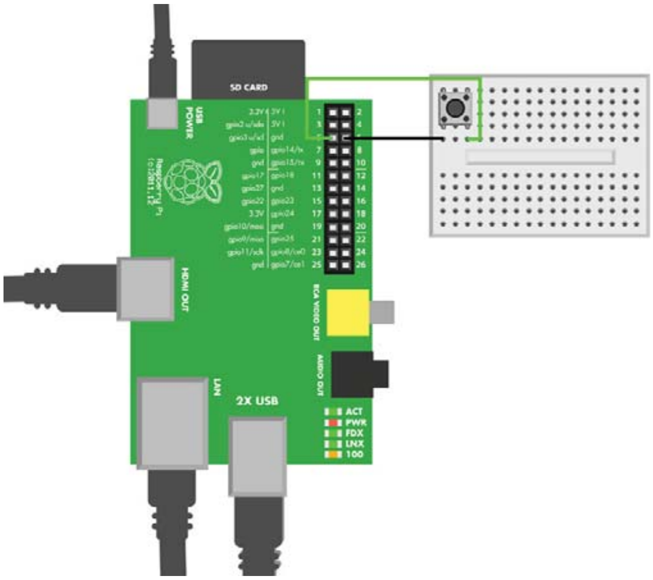

Figure 4-2. *Button on breadboard*

The button circuit has only one component: a button placed between GPIO 3 (physical pin 7) and a ground pin. When you press the button, GPIO 3 becomes connected to the ground.

On the Raspberry Pi, each GPIO pin has two numbers: one corresponding to its purpose (BCM, or Broadcom numbers), and another corresponding to its physical location. The purpose (BCM) numbers are the numbers you'll use in your programs, as shown in Table 4-1.

Table 4-1. *Pins used with button*

GPIO pin (BCM)	Board pin (place in header)
GPIO 3	5
GND	6

What happens when the button is not pressed (open)? For your program to work reliably, the pin must be connected to *something*. A pin that's not connected is *floating*, and the value from such a pin will vary with ambient electrical fields. As you might remember from "Pull-Up Resistors and Arduino" on page 37, we don't want floating pins.

There is a built-in pull-up resistor on pins GPIO 2 and GPIO 3. The internal pull-up connects the pin to +3.3 V through a 1800 Ω (1.8 kΩ) resistor. The pull-up resistor is always connected on these pins. When the button is not pressed, the pull-up resistor pulls GPIO 3 to HIGH, +3.3 V.

Be careful when connecting wires to the pins! For example, you never want to connect a GPIO to +5 V, because it could physically damage your Raspberry Pi. The pins use 3.3 V and can't handle 5 V. Before you power up your Raspberry Pi after making connections, double- and triple-check them to be sure you made them correctly.

Run the Button Code

With the button connected, you can run some code to monitor its value. Open the Terminal (LXTerminal), and follow the steps described in this section.

You should make GPIO work without root first (see Example C-1). If you want to quickly try out GPIO with root powers, use *sudo tee* instead of *tee* after the pipe. However, you really should make GPIO work as a normal user.

Export the GPIO 3 pin so that you can use it in your code (don't type the `$`; it indicates the shell prompt that you see in the terminal window):

```
$ echo 3|tee /sys/class/gpio/export
```

This created a new folder called *gpio3*. You can see it by typing *ls /sys/class/gpio/*.

To read a pin, you need to configure it to "in" mode:

```
$ echo in|tee /sys/class/gpio/gpio3/direction
```

Now you can read the value (don't hold down the button while you do this):

```
$ cat /sys/class/gpio/gpio3/value
1
```

Because you weren't pressing the button, the internal pull-up resistor pulled GPIO 3 to 3.3 V, HIGH, so you got the value `1`.

Next, hold down the button and read the value again (for a cool trick, you can recall the previous command by pressing the up arrow, so you don't have to type it again):

```
$ cat /sys/class/gpio/gpio3/value
0
```

Well done! You can now read values from digital sensors. Even though you'll soon practice with another digital sensor, the program will be very similar to this one.

Figure 4-3. *Buttons don't need to be small and black*

Troubleshooting

Before correcting any connections, be sure to power down your Raspberry Pi. Power it back up when you are sure the connections are correct. Here are some common troubleshooting issues:

Why do I always get a value of 0?
Did you put your four-legged button the right way round? If you mistakenly connected GPIO 3 and ground to leads that are always connected to each other, that's exactly the same as connecting them directly. GPIO 3 would be always connected to ground (GND, 0 V), so you'd always get a value of 0.

Why do I always get a value of 1?
This means that GPIO 3 is *never* connected to ground. A likely reason is that the button is not connected. Did you use the correct pins? Did you put all the jumpers and leads to correct rows?

Permission denied?
These commands assume that you made your GPIO work without root, as explained in "Using GPIO Without Root" on page 114. If you insist on running everything as root (which we maintain is a dangerous, less stable way), you can test it as root by putting `sudo` in front of every *tee* com-

mand. For example, you could write *echo 3|sudo tee /sys/class/gpio/export*. But of course it's much better to avoid unnecessary use of `sudo`.

Works with sudo, but not as normal user (without sudo)?
Go through the steps in "Using GPIO Without Root" on page 114 again.

In the following projects, we'll show you how to use the GPIO pins from a Python program. All of the sensor projects in this book use Python.

Python is one of the easiest languages to learn. When we code for fun, most often it's in Python. Before we get into programming GPIOs with Python, let's get you up and running with the language.

Hello, Python World

Traditionally, the first program to write in any language is "Hello, world." It lets you test to confirm that everything is working. If you are a beginner, you'll also get assurance that you know how to write a Python program!

Whenever you start programming, you should always start with a "Hello, world"—on any language, on any platform.

You'll write your Python program just like any text file. The best place to store your program files is in your home directory (*/home/pi*). If you're not already in your home directory, you can change to it with the terminal (LXTerminal) command `cd`:

```
$ cd /home/pi
```

Start editing a new file with *nano*, a command-line text editor for editing plain-text files. Its purpose is similar to other plain-text editors such as Notepad, gedit, vi, and Emacs. Type this command:

```
$ nano hello.py
```

The file needs only one line. This is Python for you: easy things are so easy, but hard things are still possible. Add this line to the file:

```
print("Hello, world!")
```

The code has just one line. `print()` is a command, but it behaves like a function. It gets one argument, the string "Hello, world!" As you probably guessed, you are asking the computer to print the string to the terminal.

Now save your work by pressing Ctrl-X (exit). When prompted about whether you want to save, type "y" for yes. Finally, confirm the name, "hello.py", by pressing Enter or Return.

Now it's time to run your Python program. This is a big moment if this is your first step in Python. Type the command shown after the $ prompt. You'll see the output as soon as you press Enter:

```
$ python hello.py
Hello, world!
```

Your Python program printed the string "Hello, world!" Now you know you can write Python, and that your Python environment works.

Project 14: Blink an LED with Python

When starting with a new feature, it's good practice to write a "Hello, world" program. It's a program that tests the feature as simply as possible.

In this case, to start working with GPIO and Python, you'll write a program that blinks an LED. Knowing how to do this in Python, it will be much easier to use this skill as a part of your robot or gadget projects.

Parts

You need the following parts for this project:

- Raspberry Pi
- Female-male jumper wires, black and green
- An LED
- 470 Ω resistor (four-band resistor: yellow-violet-brown; five-band resistor: yellow-violet-black-black; the last band will vary depending on the resistor's tolerance)
- Breadboard

Build the LED Blink Project

Connect an LED and a resistor to GPIO 27 as shown in Figure 4-4.

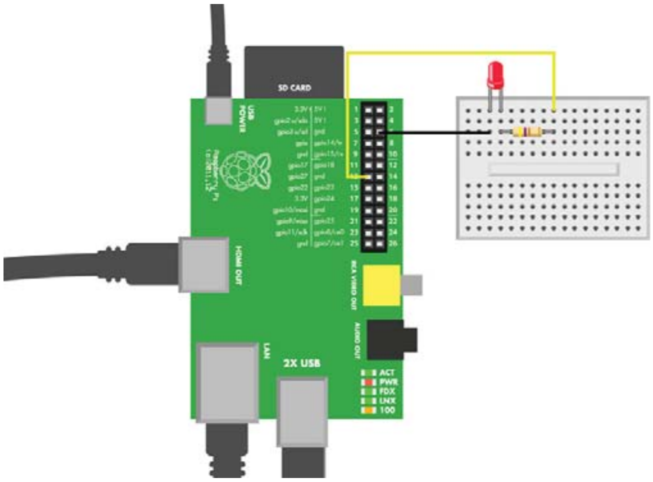

Figure 4-4. *Hello, LED*

Run the Code

Open a terminal, then change to your home directory *cd /home/pi/* if you're not already there. Create a new file with *nano led_hello.py*, type in the code from Example 4-1 (or download it from this book's website), and save it with Ctrl-X, press y, and then Enter or Return.

Run the program with:

```
$ python led_hello.py
```

The LED lights up for a moment, then dims again. Did you succeed in lighting up the LED? You can now use GPIO pins with Python.

If you get "IOError: [Errno 13] Permission denied," simply try again. If the problem persists, verify that you have enabled GPIO without root using the udev rule (see "Using GPIO Without Root" on page 114).

This code shows you the secret sauce of GPIO manipulation from Python. Later on, you'll see examples that use a *library* (a file containing code that does all the work of manipulating the GPIO for you), so those code listings will be much shorter.

As you saw earlier in "Project 13: Momentary Push Button" on page 64, you need to manipulate files in the Linux filesystem to work with GPIO pins.

Want to save typing? Download the code, unzip it, and copy it to */home/pi* on your Raspberry Pi SD card.

Example 4-1. Hello, LED Python code

```
# led_hello.py - blink external LED to test GPIO pins
# (c) BotBook.com - Karvinen, Karvinen, Valtokari

"led_hello.py - light a LED using Raspberry Pi GPIO"
# Copyright 2013 http://Botbook.com */


import time     # ①
import os       # ②

def writeFile(filename, contents):      # ③
        with open(filename, 'w') as f:  # ④
                f.write(contents)       # ⑤

# main

print "Blinking LED on GPIO 27 once..."         # ⑥

if not os.path.isfile("/sys/class/gpio/gpio27/direction"):      # ⑦
        writeFile("/sys/class/gpio/export", "27")       # ⑧

writeFile("/sys/class/gpio/gpio27/direction", "out")    # ⑨

writeFile("/sys/class/gpio/gpio27/value", "1")  # ⑩
time.sleep(2)   # seconds       # ⑪
writeFile("/sys/class/gpio/gpio27/value", "0")  # ⑫
```

① Import a library to give you more commands (functions) to work with. Here we are importing the time library, a very common library because it has the `time.sleep(seconds)` function that makes your program pause for a while.

② The OS library has commands that let you work with the underlying operating system. With `os.path.isfile(filename)`, you can test whether a file corresponding to the value in `filename` exists. Because the OS library is an abstraction of the operating system's features, the same code can work on Linux, Mac, and Windows.

③ Define a new function called `writeFile()`. Defining a new function saves you from typing the same commands over and over again. Instead of copying all of the code in this function each time you need it, you can simply call the function `writeFile("foo.txt", "Hello")`. This function takes two parameters: a filename for the name of the file to write, and the contents to put into the file. If you called `writeFile("foo.txt", "Hello")`, the values between the parentheses will be assigned to *variables* inside the function. When `writeFile()` is executed, the variable `filename` will have the value "foo.txt" and the variable `contents` will have the value "Hello".

❹ Open `filename` for writing (*w*). This defines a new file *handle* named `f` that you can use to work with the file in later lines of code. The `with` block is the easiest built-in way to manipulate text files in Python. Python will make sure that the file handle gets closed again; either at the end of the block, or if any unexpected error occurs. Of course, your newly defined function `writeFile()` makes writing files even easier because it uses `with` for you.

❺ Using the file handle `f` defined in the previous line, we write a value of variable `contents` to the file. File handle `f` is an *object*, and `write()` is a *method* of this object.

❻ Print some text to help you debug your program. For example, if you see the text appear onscreen but the LED doesn't light, you already know that your program ran but that the problem is probably in the GPIO connections. `print()` can be used in two ways. You can call it with the parameters in parentheses as in *print("foo")*, just like any function. Or, if you're feeling lazy, you can use this shorter syntax that leaves out the parentheses.

❼ Test whether the file named *direction* already exists. If not, then execute the indented block below. This test is needed because exporting the same pin twice causes an *exception* that would make your program crash.

❽ Export the pin. Exporting creates the virtual[1] files to manipulate the pin such as *direction* and *value*. The export is done by simply writing the pin number "27" to the (virtual) text file *export*. This block is indented (there's white space to the left of the line) under the `if`, so it's only executed if the `if` condition above it evaluates to true.

❾ Write the string "out" to the text file *direction*. Here you call the function `writeFile()` that we defined earlier. When you set the direction to "out," you can turn the pin on or off.

❿ Write the string "1" to the text file *value*. This turns the GPIO 27 pin to 3.3 V (HIGH). This causes current to flow through the LED and the resistor to GND (0 V). The current lights up the LED.

⓫ Wait for two seconds. All the pins stay in the state they are in, so the LED stays on. Because `sleep` was imported from the `time` library, you have to specify the library's *namespace* (`time`) before it, as in *time.sleep()*. When your program sleeps, it takes very little computing power, and other programs running in the background can have more time.

---

[1] So called because they don't really exist as files on stable storage (SD card).

⓬ Write “0” to value, turning off the GPIO pin. Both leads of the LED are now at the same voltage, so there is no current flowing, and the LED goes dark. Because this is the last command, the execution of the program ends.

When you are looking at a program, you should think through it in the order it is executed. First, libraries are imported to give you more functions to work (`import time...`).

Next, the first code in the *global scope* is run (*print “Blinking...”*). To simplify, it’s the first command that’s flush left, without any indentation (whitespace) before it.

The program execution continues line by line. The function `writeFile()` is only run when it’s called (*writeFile(“/sys/...”)*). It’s worth noting that functions are not run when they are defined (`def writeFile(filename, contents)`), even though the definition appears before the lines of code and at the same level of indentation.

Finally, the last line is executed (`writeFile("/sys/...")`). Then the program exits, and the executions stop.

Now that you know the secret sauce, the rest of the examples will use the *botbook_gpio.py* library to make the programs shorter. That library will do all of the work of manipulating files for you.

Project 15: Adjustable Infrared Switch

An adjustable infrared switch senses when something is near. You can configure the distance at which it is triggered by a nearby object. This sensor only tells you when something is near, not its actual distance.

Parts

You need the following parts for this project:

- Infrared sensor switch
- Two 1 kΩ resistors (five-band: brown-black-red-brown-any; four-band: brown-black-red-any)
- Raspberry Pi

Build the IR Switch Project

Reading any digital resistance sensor is similar to reading a button, so both the connection and the code should already look familiar to you if you’ve tried connecting a button to the Raspberry Pi.

Let's connect the infrared (IR) sensor switch to the Raspberry Pi. There are some differences from the connections for the button. The button just has a data pin and a ground connection. The IR sensor has some additional parts: a red +5 V pin and two resistors connected to the yellow data pin. This is because the IR sensor provides a new challenge: GPIO data pins can only tolerate 3.3 V, but the sensor works at 5 V.

You can lower the voltage of the data pin with a *voltage divider* (see "Voltage Divider" on page 76), which you can make with two resistors. A voltage divider is a circuit that lowers a voltage by dividing it between two resistors (one connected to a signal, the other to ground).

Make sure the Raspberry Pi is shut down and disconnected from power. Build the circuit as shown in Figure 4-5.

Usually, IR switches have black minus, yellow data, and red plus. We also have some models that have a green minus wire instead of a black one.

If your IR sensor switch has female pin headers, simply use jumper wire to connect it to the breadboard.

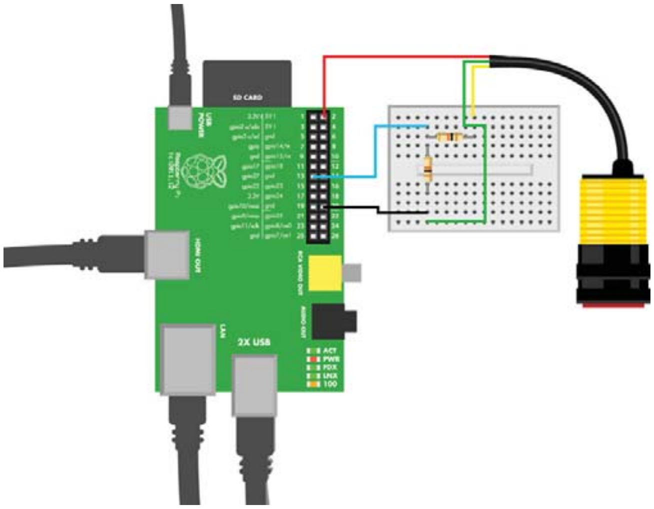

Figure 4-5. *Adjustable infrared switch connected to Raspberry Pi*

Run the Code

Power up the Raspberry Pi, then upload the code shown in Example 4-2.

Try it out:

```
$ python adjustable-infrared-sensor-switch.py
Something is inside detection range
```

Run it many times to get many measurements.

Example 4-2. Adjustable infrared switch code

```
# adjustable-infrared-sensor-switch.py - is object within predefined distance?
# (c) BotBook.com - Karvinen, Karvinen, Valtokari


import botbook_gpio as gpio     # ❶

gpio.mode(27, "in")     # ❷
x = gpio.read(27)       # ❸
if( x == 0 ):   # ❹
        print "Something is inside detection range"     # ❺
else:   # ❻
        print "There is nothing inside detection range" # ❼
```

❶ The *botbook_gpio* library contains the Python code for manipulating GPIO pins. For it to work, the *botbook_gpio.py* file must be in the same directory as this program, *adjustable-infrared-sensor-switch.py*. If you get an import error when you run it, verify that you can see both files in the same directory when you run the `ls` command. *botbook_gpio.py* simply reads and writes files under */sys/class/gpio/*, as you saw in "Project 14: Blink an LED with Python" on page 69.

❷ Setting the gpio27 pin to "in" allows its value to be read.

❸ Read the current value into new variable we call `x`. The value will be either 0 or 1.

❹ Value 0 means something has been detected. An infrared sensor switch doesn't tell the distance; it just tells whether something is nearer than the distance to which it's calibrated. An equality comparison uses two equals characters (==). The comparison operation, (==), is different from value assignment, (=).

❺ Print to the terminal that something was detected. This indented block under `if` only runs when the `if` condition is true; in this case, it runs if the value of variable `x` is zero.

❻ The indented block under `else` only runs when the `if` condition is not true. In this case, the one-line block under `else` runs if the value of variable `x` is not zero.

❼ Print to the terminal that nothing was detected. As this is the last instruction in the program, the execution of this program ends after this line. If you want to make another measurement, run your program again.

Voltage Divider

How can you convert voltage from 5 V to 3.3 V? As most Arduino and Arduino-compatible components use +5 V as HIGH, you will meet this question often when you try to use something with the Raspberry Pi.

The red wire supplies the IR sensor with +5 V, and the IR sensor outputs +5 V. This is much more than the Raspberry data pins can take, as their HIGH is +3.3 V. Connecting a GPIO pin to +5 V will likely destroy that pin.

When the IR sensor yellow wire is at zero, all is well because the Raspberry Pi ground (0 V, LOW) is exactly the same as the ground in the IR sensor. That's because the IR sensor ground is connected to the Raspberry Pi ground (black wire in Figure 4-5).

But what about the IR sensor HIGH state, +5 V? To protect the Raspberry Pi data pins, you must ramp down the voltage from 5 V to 3.3 V. This is easily done with a voltage divider.

The big 5 V is connected to ground (black wire, 0 V, GND) through two resistors. On the 5 V side, the voltage is obviously 5 V. Just as obviously, on the ground side of the two resistors, the voltage is 0 V. But in between, the voltage gets split according to the values of the resistors. You want 3.3 V for the Raspberry Pi:

```
3.3 V / 5 V = 0.66 = 66 %
```

So, you want the Raspberry Pi data pin to measure the voltage 66% on the way from ground to 5 V. Your total resistance is 2 kΩ (two resistors in series, 1000 Ω each):

```
66% * 2000 Ohm = 1320 Ohm
```

You won't find resistors in every possible size, so a resistor near that will probably do. The parts list mentions two 1 kΩ resistors. Putting the data pin between those resistors, you get 50% of the 5 V:

```
50% * 5V = 2.5 V
```

This is well above the minimum level for a HIGH signal, which is about half of the nominal 3.3 V HIGH level, 1.65 V. Because 1.65 V < 2.5 V, Raspberry considers 2.5 V HIGH.

But how did we come up with the 2 kΩ of resistance for the voltage divider? The exact amount doesn't matter. There must be enough resistance so that ground and 5 V don't short circuit. An air gap would achieve that, so the other requirement is that resistance must allow some electricity pass through. Thus, some kΩs is good.

You can halve the 5 V by building a voltage divider with two resistors. And now you know that lowering voltage doesn't require mysterious integrated circuits or breakout boards—just two cheap resistors.

Analog Resistance Sensors

Analog resistance sensors change their resistance according to how they respond to something in the physical world. For example, a light-dependent resistor will have low resistance in bright light, and high resistance in the dark.

The Raspberry Pi can't directly measure resistance, apart from two states: on and off. To measure resistance, you need an *analog-to-digital converter* (ADC). Unlike with the voltage divider, this will require a mysterious integrated circuit (that won't be mysterious to you in a moment).

It's a bit easier to measure analog resistance with Arduino. Because Arduino has a built-in ADC, there will be fewer wires and components in your connections. Raspberry Pi needs an external ADC chip to read analog resistance, which means more wires that could come loose or be connected incorrectly.

An advanced option for measuring analog resistance with the Raspberry Pi is using only sensors that communicate over industry standard protocols like I2C or SPI. Using I2C and SPI is out of the scope of this book, but you can learn about it in *Make: Sensors*.

Another option is to use Arduino for connecting to sensors and communicating with Raspberry Pi over USB serial. A similar setup using a computer rather than a Raspberry Pi is shown in *Make: Arduino Bots and Gadgets*. In this book, you'll work with the conceptually simple approach of reading analog resistance sensors with an ADC chip.

Project 16: Potentiometer to Measure Rotation

Turn up the volume on a radio—you're probably using a potentiometer. They are everywhere! Whenever you turn a knob to gradually change something, chances are that you are using a potentiometer. There are also hidden potentiometers (*trimmers*) inside many devices that technicians can adjust with a screwdriver.

The potentiometer (or *pot*) is the archetype of variable-resistance sensors. Once you learn to work with a pot, you can easily see similarities with other variable-resistance sensors.

This project teaches you how to read analog-resistance sensors with Raspberry Pi. Analog readings are interesting because you have to use an external analog-to-digital converter (ADC). In this project, you'll learn to use the MCP3002 ADC. It's controlled over SPI, an industry standard protocol.

Parts

You need the following parts for this project:

- Potentiometer
- MCP3002 (analog-to-digital integrated circuit)
- Raspberry Pi
- Jumper wires (male-to-female and male-to-male)
- Breadboard

The Raspberry Pi does not have female header connections for its GPIO pins. Instead, you will need to use a Pi cobbler or male-to-female jumper wires. Either option works, but the cobbler eliminates the concern of shorting out the GPIO pins by accidentally touching them together.

Build It

Build the potentiometer circuit shown in Figure 4-6.

Install SpiDev

All analog readings in this book use the MCP3002 analog-to-digital converter. The MCP3002 uses SPI, an industry standard protocol, to communicate with the Raspberry Pi. That's why all analog sensors in this book need the spidev library.

Enable Internet on your Raspberry Pi. Simply use a normal Ethernet cable to connect the Raspberry Pi to your wireless access point or switch. The Raspberry Pi receives an IP address automatically. You can test your Internet connectivity with the Midori web browser. Try *http://botbook.com*!

Install the requirements with `apt-get` (it uses the Internet, so if you get errors, check your connection):

```
$ sudo apt-get update
$ sudo apt-get -y install git python-dev
```

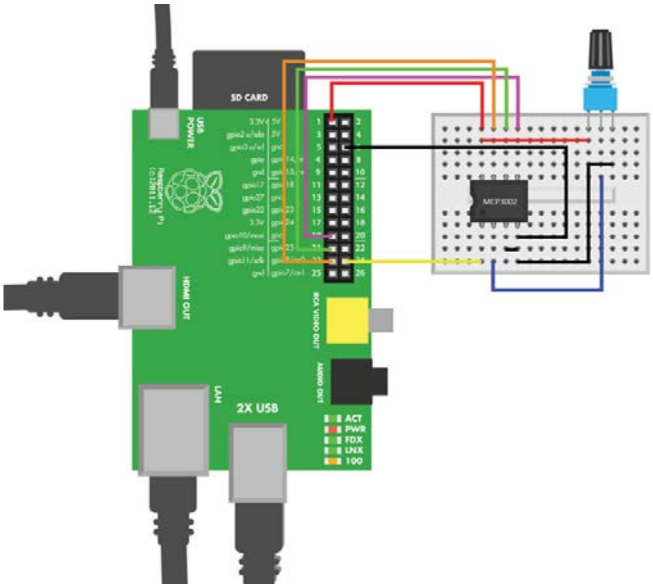

Figure 4-6. *Potentiometer connected to Raspberry Pi*

Then download and install the latest SpiDev from their version control system:

```
$ git clone https://github.com/doceme/py-spidev.git
$ cd py-spidev/
$ sudo python setup.py install
```

Allow SPI Without root

Modify udev rules, as shown in Example 4-3:

```
$ sudoedit /etc/udev/rules.d/99-spi.rules
```

Write the rule. If you edited it with nano, press Ctrl-X, then type "y," and press Enter to save.

Example 4-3. 99-spi.rules

```
# /etc/udev/rules.d/99-spi.rules - SPI without root on Raspberry Pi
# Copyright http://BotBook.com

SUBSYSTEM=="spidev", MODE="0666"
```

To avoid typos, you could alternatively copy the file with *sudo cp 99-spi.rules/etc/udev/rules.d/*.

Finally, edit the blacklist to enable the SPI device:

```
$ sudoedit /etc/modprobe.d/raspi-blacklist.conf
```

Delete the line that says “blacklist spi-bcm2708,” then press Ctrl-X to exit, type “y” then press Enter to save the file.

Reboot your Raspberry Pi. Verify that you can see some files with:

```
$ ls -l /dev/spi*
```

Can you see at least two files? Well done, now you are ready to read analog resistance sensors.

Run the Code

Once you get the circuit wired up and SpiDev installed, reading the sensor is easy with the MCP3002 library. You can download it from the website for *Make: Arduino Bots and Gadgets*.

Put *pot_once.py* (see Example 4-4) and the *MCP3002.py* library into the same directory. You can easily create the *pot_once.py* file with the *nano* editor. The *MCP3002.py* library is long, so it’s best to download it as well.

Run the code:

```
$ python pot_once.py
```

View your result. Turn the knob of the potentiometer, and run the program again. In Linux, you can press the up arrow to retrieve an earlier command so you don’t need to keep typing it over and over.

Example 4-4. Reading a potentiometer

```
# pot_once.py - measure resistance of a potentiometer, once
# (c) BotBook.com - Karvinen, Karvinen, Valtokari

import botbook_mcp3002 as mcp   # ❶

x = mcp.readAnalog(0,0) # ❷
print(x)        # ❸
```

❶ Import the MCP3002 library. The file *MCP3002.py* must be in the same directory as this program (*pot_once.py*). All the heavy lifting, like the SPI interface, is done with this library.

❷ Read the analog value from the MCP3002 chip. The `readAnalog()` function is from the MCP3002 library, as you can see from the namespace prefix “MCP3002.”. The parameters device (0) and channel (0) correspond to the circuit you built. The `readAnalog()` function returns a number between 0 and 1023. This number is saved to a new variable called `x`.

❸ Print the value of variable `x`.

Troubleshooting? If running *pot_once.py* gives you this error…

```
spi.open(0, device) .. IOError: [Errno 2] No such file or directory
```

…verify that you modified */etc/modprobe.d/raspi-blacklist.conf* as instructed in "Allow SPI Without root" on page 79, then rebooted. You should see some files if you run the command **`ls /dev/spi`**.

Repeat Readings

Most embedded devices have a never-ending loop that keeps running forever (Arduino has this: the `loop()` function runs until you shut the device down). The program will do its thing, wait for a short while, then do it again. It makes sense that an embedded device works as long as there is power: you wouldn't want your car's assisted steering or your office air conditioning to stop and require restarting all the time.

Example 4-5 shows how this is done in Python. Keep the same circuit as before, but run the new code. Now you can continuously get new readings while you adjust the potentiometer. Press Ctrl-C to kill the program and get your command prompt back.

Example 4-5. Reading a pot over and over

```
# pot_repeat.py - continuously measure resistance of a potentiometer
# (c) BotBook.com - Karvinen, Karvinen, Valtokari


import botbook_mcp3002 as mcp   # ❶
import time     # ❷

while(True):    # ❸
        x = mcp.readAnalog(0,0) # ❹
        print(x)        # ❺
        time.sleep(0.5) # seconds       # ❻
```

❶ Import the library for the analog-to-digital converter chip. Just as before, *MCP3002.py* must be in the same directory as this program, *pot_repeat.py*.

❷ Import the library that supplies the `time.sleep()` function.

❸ Repeat the block below as long as the condition is true: forever. Press Ctrl-C to kill the program.

❹ Read the analog value just like you did in *pot_once.py*.

❺ Print the value of `x`, just like in *pot_once.py*.

❻ Wait for a while. Whenever there is an infinite loop, you must make sure you give time to the operating system and other programs. Otherwise, your program could take 100% of CPU power of any one core, even one running a zillion gigahertz.

Why Does a Potentiometer Have Three Legs?

You may have noticed that you didn't use a pull-up, pull-down, or voltage divider, and you probably also noticed that a potentiometer has three leads.

One side is connected to ground (LOW, 0 V, black wire, GND, "negative"). The other side is connected to positive (HIGH, 3.3 V, +, red wire). The middle lead is connected to the data pin.

Turning the knob, you can select where the middle lead touches the resistor. Because there is a resistor on either side of the pot (between the middle lead and GND, and between the middle lead and +V), the pot itself is a voltage divider!

If the knob is turned to the minimum setting, there is nearly no resistance between the data pin (middle) and GND. All the resistance is between the data pin and +3.3 V. So the data pin gets 0 V, and `MCP3002.readAnalog(0,0)` returns 0.

When you turn the knob to the maximum, there is no resistance between the data pin and +3.3 V. All the resistance is between data and ground. So the data pin gets 3.3 V, and `MCP3002.readAnalog(0,0)` returns 1023.

When the knob is anywhere between maximum and minimum, you get a voltage between LOW (0 V) and HIGH (3.3 V). `MCP3002.readAnalog(0,0)` returns a number between 0 and 1023.

The voltage seen by the data pin is proportional to the position of the pot.

Project 17: Photoresistor

Resistor by night, wire by day. A photoresistor reduces its resistance when it's bright. It's also known as a light-dependent resistor (LDR).

The LDR changes resistance. From the Raspberry Pi's point of view, isn't this the same idea as a potentiometer?

In fact, the photoresistor can use the same code as a potentiometer. The only difference is the number of leads on the sensor. As you can see from Figure 4-7, an extra resistor is used to form a voltage divider. You may remember that a voltage divider requires two resistors; the photoresistor is the second (and variable) resistor in the divider.

Parts

You need the following parts for this project:

- Photoresistor (10 kΩ suggested)
- Resistor (10 kΩ, or choose a value that matches the photoresistor)
- MCP3002

- Raspberry Pi
- Jumper wires (male-to-female and male-to-male)
- Breadboard

The 10 kΩ resistor coloring is brown-black-orange-any (four-band) or brown-black-black-red-any (five-band).

Build It

Wire up the circuit as shown in Figure 4-7.

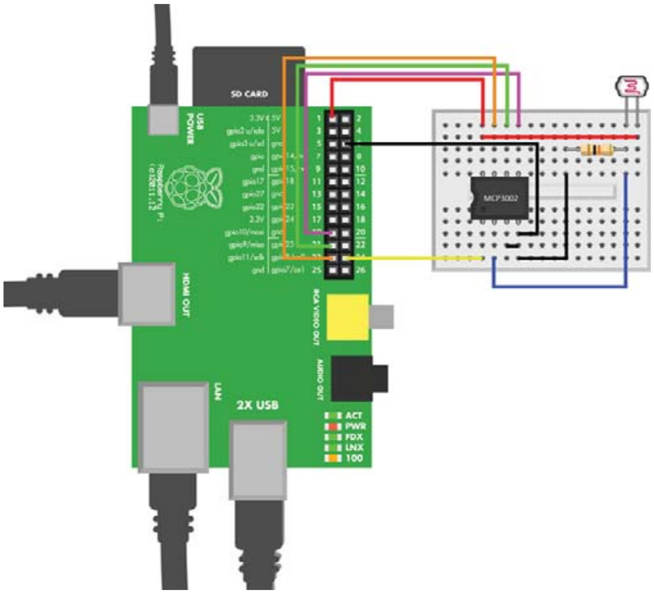
SD CARD
USB POWER
HDMI OUT
RCA VIDEO OUT
AUDIO OUT
LAN
2X USB
ACT
PWR
FDX
LNX
100
MCP3002

Figure 4-7. *Photoresistor connected to Raspberry Pi*

Run the Code

After you've built the circuit, run the pot code again:

```
$ python pot_repeat.py
```

When you see numbers appearing, try changing the value. Cover the LDR with your finger, and you see low numbers. Point a bright light at it, and the numbers increase.

Once you are done, kill the program with Ctrl-C.

Playing with Resistance Numbers

If that felt too easy, let's try playing with the measured numbers. There are three common ways to look at the measured resistance:

- Raw number, from 0 to 1023
- Voltage, from 0 V to 3.3 V
- Percentage, from 0% to 100%

You have already read the raw value with a potentiometer and light-dependent resistor. It's the value you get from this line of code:

```
MCP3002.readAnalog(0,0)
```

The raw number doesn't have an obvious unit—it's just an integer (a whole number) from 0 to 1023. This is the number you always start with when you take a sensor measurement.

It's easy to convert a raw value to a percentage, from 0% to 100%. The minimum raw value 0 is 0%. The maximum value 1023 is 100%. Any value x in-between is x/1023. For example, the raw value 348 is:

```
348/1023 = 0.34 = 34%
```

So the raw value 348 is 34% of the maximum value.

```
raw = MCP3002.readAnalog(0,0)
percent = raw/1023.0    # ❶
```

❶ To get a floating-point result, you must divide by a floating-point number. In Python (and many other languages), dividing by an integer (a whole number) gives you a counter-intuitive whole number result. In Python, 1/2 == 0, but 1/2.0 == 0.5. Notice how 2 is an integer and 2.0 is a floating-point number.

You might already know that a percent is a one hundredth part. This means that 0.34 is exactly 34%. When you speak about parts, percentages are practical. When programming, a floating-point number 0.34 is practical, as you can easily multiply a number with this.

In your own projects, you measure something to change something. You might use a potentiometer to set the volume of a beep. Or you could use an LDR to turn lights brighter when it's dark outside. In these cases, it's convenient to multiply by a percentage.

To get the voltage of the data pin, you can multiply by the percentage. The maximum voltage (HIGH) of a Raspberry Pi pin is 3.3 V (100%) and the minimum (LOW) is 0 V (0%):

```
raw = MCP3002.readAnalog(0,0)
percent = raw/1023.0
voltage = percent * 3.3
```

Project 18: FlexiForce

The FlexiForce sensor (see Figure 3-14) measures pressure. The harder you squeeze the round end of the sensor, the more electricity it lets through. Just as with a pot and LDR, you can use the same code.

Parts

You need the following parts for this project:

- Flexiforce (11 kg, 25 pound version recommended)
- 1 MΩ resistor (five-band: brown-black-black-yellow-any; four-band: brown-black-green-any)
- MCP3002
- Raspberry Pi
- Jumper wires (male-to-female and male-to-male)
- Breadboard

Build It

Build the FlexiForce circuit shown in Figure 4-8. As the FlexiForce has just two leads; this connection is similar to the light-dependent resistor (LDR). There is a second resistor to form a voltage divider

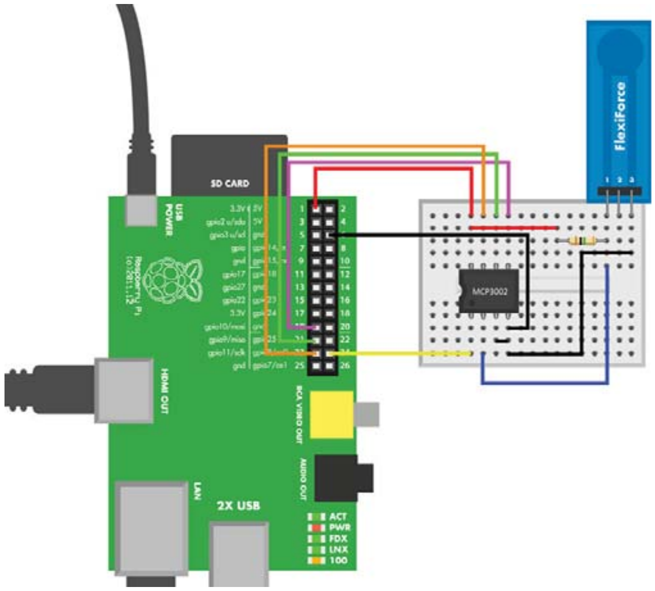
FlexiForce
SD CARD
USB POWER
Raspberry Pi
HDMI OUT
RCA VIDEO OUT
AUDIO OUT
LAN
2X USB
ACT
PWR
FDX
LNX
100
MCP3002

Figure 4-8. *FlexiForce squeeze sensor connected to Raspberry Pi*

What about the third pin? The FlexiForce has a third pin in the middle, but it's not connected to anything. It's just there to make it easier to connect the FlexiForce to a breadboard. The outer pins are live.

Run the Code

Run the potentiometer code again:

```
$ python pot_repeat.py
```

Play with the values. Try leaving the sensor untouched. Then squeeze the round end hard.

You've tried three analog resistance sensors: a pot, an LDR, and the FlexiForce. Do you recognize a pattern? All of these sensors can use the exact same code. The circuit with the analog-to-digital converter (ADC, MCP3002) is nearly identical for the different components. The only difference is the second resistor for the sensors (the first resistor in the voltage divider) that have just two leads.

Project 19: Temperature Measurements (LM35)

The LM35 reports temperature by changing its resistance. In its default configuration, the LM35 can measure temperature from 2 °C to 150 °C.

Parts

You need the following parts for this project:

- LM35 temperature sensor
- MCP3002
- Raspberry Pi
- Jumper wires (male-to-female and male-to-male)
- Breadboard

Build It

Build the circuit shown in Figure 4-9. Wires that are always in 3.3 V (same as data pin HIGH) are red in the diagram. The dangerous (to Raspberry Pi's pins) +5 V wire is yellow. Don't connect the yellow +5 V wire to data pins.

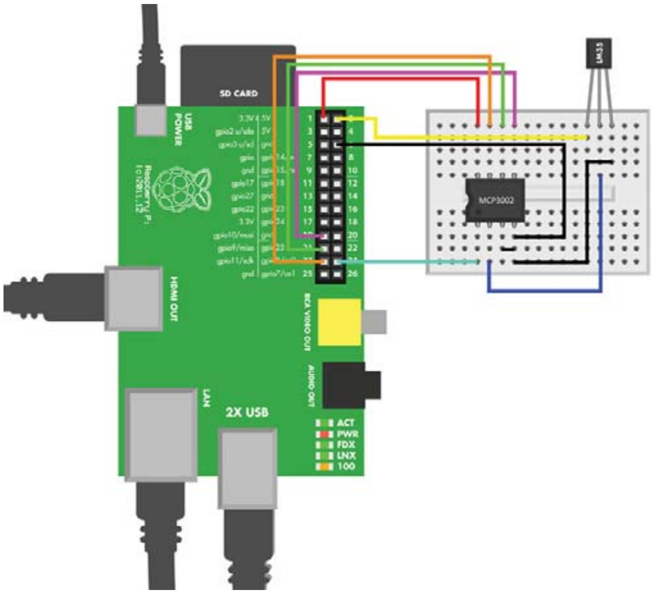

Figure 4-9. *LM35 temperature sensors connected to Raspberry Pi*

As LM35 has three leads, the connection is similar to the pot circuit. The only difference is the 5 V voltage required by the LM35.

The LM35 uses +5 V, enough to damage the Raspberry Pi's microcontroller IC. Mistakenly connecting 5 V to a data pin, even for a moment, damages the Raspberry Pi. Most likely, the damage is permanent, destroying some or all of the data pins in the GPIO header. That's why you use different colors for safe +3.3 V (red) and dangerous +5 V (yellow). Be careful where you connect +5 V. Just be careful to make correct connections and you'll be fine.

Run the Code

Every analog resistance sensor thus far (pot, LDR, and FlexiForce) has worked with the same code. You can also try pot code here.

It does work, but the raw numbers are not very useful. What good does it do to know that the temperature results in a raw value such as 815?

As you saw earlier, you can convert raw values to percentages and voltages.

All components have a datasheet that describes how they are used. Datasheets can be terse and technical, which makes books like this quite useful. You can find the datasheet for the LM35 by searching the Web for "lm35 datasheet."

The datasheet for the LM35 tells us that the output voltage changes by 0.010 V per 1 °C. The inverse of this is that a 1 V change represents 100 °C. Some sample voltage-temperature pairs are:

- 0 °C, 0.0 V (this circuit doesn't measure temperatures this cold)
- 2 °C, 0.02 V (minimum measured temperature, a cold fridge)
- 10 °C, 0.1 V
- 20 °C, 0.2 V (summer in Helsinki)
- 30 °C, 0.3 V (summer in Spain)
- 40 °C, 0.4 V
- 80 °C, 0.8 V (a sauna)
- 100 °C, 1.0 V (water boils)
- 150 °C, 1.5 V (maximum measured temperature)

This also explains why we're not worried about getting too much voltage to the Raspberry Pi data pin: to get over 3.3 V, you would need 330 °C of heat.

You must convert the raw value to voltage, and voltage to Celsius. For this, let's create a new program.

Run the code shown in Example 4-6.

Example 4-6. Reading the LM35 sensor in Python

```
# lm35.py - print temperature in Celsius
# (c) BotBook.com - Karvinen, Karvinen, Valtokari


import time     # ❶
import botbook_mcp3002 as mcp   # ❷

def readTemperature():  # ❸
        data = mcp.readAnalog(0,0)      # ❹
        percent = data / 1023.0 # ❺
        volts = percent * 3.3   # ❻
        celcius = 100.0 * volts # ❼
        return celcius  # ❽

while True:     # ❾
        t = readTemperature()   # ❿
        print("Current temperature is %i C " % t)       # ⓫
        time.sleep(0.5) # seconds       # ⓬
```

❶ The `time` library is needed for `time.sleep()` in the loop.

❷ As with previous examples, the file *MCP3002.py* must be in the same directory with this program (*lm35.py*).

❸ `readTemperature()` is a function, so it's easy to add it to your own projects.

❹ Read the raw value, from 0 to 1023. This is the same command you used with the potentiometer, LDR, and FlexiForce.

❺ Convert to a percentage of maximum value, from 0.0 (0%) to 1.0 (100%). The divider must be a floating point to get a floating-point result.

❻ Voltage is calculated from the maximum voltage.

❼ Convert voltage to Celsius, using the 0.010 V per 1 °C formula from the datasheet.

❽ Return the temperature.

❾ Keep measuring and printing temperatures until the user presses Ctrl-C to kill the program.

❿ This is all you need to do to read temperature in your own programs. Declare a new variable `t`, read the temperature, and store the value there.

⓫ Print the temperature with some helpful text. `print` uses a format string, where `%i` is replaced with an integer value, the value of the variable `t`.

⓬ Wait for a while. This allows you to read the printed temperature, and also prevents the program from taking all the CPU time.

Project 20: Ultrasonic Distance

An ultrasonic distance sensor gives you the distance to a target. For example, it can tell you there is a target 58 cm from the sensor. It's one of the most popular sensors in the courses we teach.

HC-SR04 is a cheap ultrasonic sensor, costing just a couple of dollars. This is a very nice price compared to alternatives costing much more.

Parts

You need the following parts for this project:

- HC-SR04 sensor
- Two 10 KΩ resistors (four-band: brown-black-orange-any, five-band: brown-black-black-red-any)
- Raspberry Pi
- Jumper wires (male-to-female and male-to-male)
- Breadboard

Build It

Build the circuit as shown in Figure 4-10. The sensor uses a protocol that's based on pulse length, where pulses consist of changes in voltage level from HIGH or LOW. As no analog-to-digital converter is needed, the circuit is quite simple. Be careful with the +5 V pin as connecting that to any data pin will likely break the whole GPIO header beyond repair.

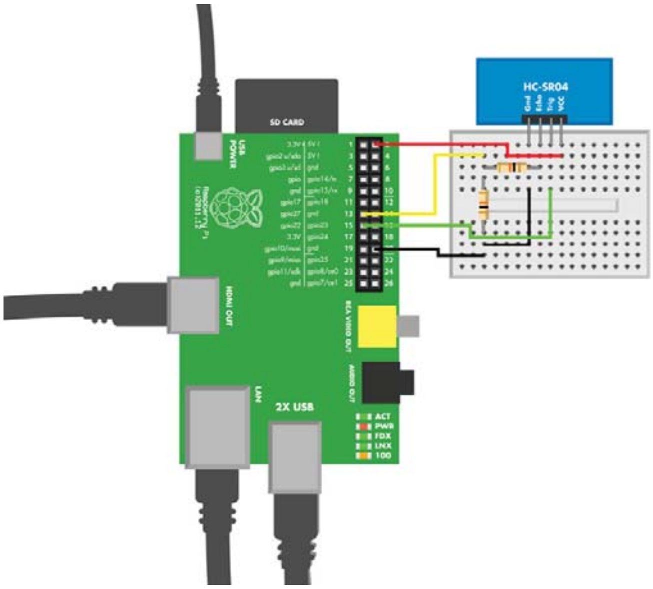
HC-SR04
Gnd
Echo
Trig
Vcc
SD CARD
USB POWER
HDMI OUT
LAN
2X USB
RCA VIDEO OUT
AUDIO OUT
ACT
PWR
FDX
LNK
100

Figure 4-10. *Cheap ultrasonic distance sensor HC-SR04 connected to Raspberry Pi*

Run the Code

Run the code shown in Example 4-7.

Example 4-7. Reading the HC-SR04 in Python

```
# hc-sr04.py - measure distance with ultrasound
# (c) BotBook.com - Karvinen, Karvinen, Valtokari


import time     # ❶
import botbook_gpio as gpio     # ❷

def readDistanceCm():   # ❸
        triggerPin = 22 # ❹
        echoPin = 27    # ❺

        v=(331.5+0.6*20) # m/s   # ❻

        gpio.mode(triggerPin,"out")     # ❼

        gpio.mode(echoPin,"in") # ❽
        gpio.interruptMode(echoPin, "both")     # ❾

        gpio.write(triggerPin, 0)       # ❿
        time.sleep(0.5) # ⓫

        gpio.write(triggerPin, 1)       # ⓬
        time.sleep(1/1000.0/1000.0)     # ⓭
        gpio.write(triggerPin, 0)       # ⓮

        t = gpio.pulseInHigh(echoPin) # s       # ⓯

        d = t*v # ⓰
        d = d/2 # ⓱
        return d*100 # cm       # ⓲

dist = readDistanceCm() # ⓳
print("Distance is %i cm" % dist)       # ⓴
```

❶ If you later improve the program measure distance continuously, you'll need `time.sleep()` in the loop.

❷ The *botbook_gpio* library controls the GPIO pins. You must have *botbook_gpio.py* in the same directory as this program, *hc-sr04.py*.

❸ The `readDistanceCm()` function is defined here. This function does all the hard work. You don't need to think how it works when you call it from your program.

❹ Define a constant for GPIO pin 22, which is connected to the HC-SR04 trigger pin. The word *trigger* was chosen because it's marked "Trig" on the sensor. Whenever `triggerPin` is mentioned in this function, it will use the number 22 in that place.

❺ The GPIO pin 27 is connected to HC-SR04 "Echo."

❻ Calculate the speed of sound, which is about 340 meters per second. The number 20 is the temperature in Celsius; feel free to change it if you happen to be in a warmer or colder place. The Raspberry Pi will do a simple calculation instantly, so there is no need to calculate it with pen and paper and put what amounts to a "magic number" into code.

❼ Set the trigger pin to "out" mode, so you can control it. The trigger pin is used for sending the sound pulse with the ultrasonic speaker. This will automatically export the pin, creating the needed *direction* and *value* files.

❽ Set the echo pin to "in" so that you can read the pulse.

❾ Enable the interrupt on `echoPin`, so that you can later measure the pulse duration with `pulseInHigh()`. The interrupt will happen "both" for rising edge (start of pulse) and falling edge (end of pulse).

❿ Make sure the trigger pin is LOW, so that you can send a short pulse later.

⑪ Wait for the pin to stabilize, so that there is a clear start for the pulse.

⑫ Bring GPIO 22 HIGH (3.3 V), starting the pulse.

⑬ Wait for a very short time. GPIO 22 stays HIGH. This wait defines the length of the pulse. One millionth of a second is one microsecond. Millionth is written as a calculation, because it's annoying to try to count the zeroes in 0.000001. To get a floating-point (decimal) number, the dividers must be floating-point numbers (1000.0 instead of integer 1000).

⑭ Turn GPIO 22 off (LOW), ending the pulse.

⑮ Measure the length of the pulse coming to GPIO 27. Save the pulse duration to a new variable, `t`. The pulse length measurement is not as precise as in Arduino (see "Real Time or Fast?" on page 92). The uncertainty of the pulse measurement limits the precision of the distance measurement. Still, `pulseInHigh()` and interrupts are much better than checking the pin state with a loop.

⑯ Calculate distance `d`, using velocity `v` and time `t`. For example, if you drive one hour at 100 km/h, you move a hundred kilometers.

⑰ One way trip is half of a two-way trip. The sound goes to target and echoes back, so the distance is half of that round trip.

⑱ Convert meters to centimeters.

⑲ Measure the distance in centimeters, and save it to new variable `d`. This is how you can measure distance in your own programs. This is the first line executed after library imports.

⑳ Print some helpful text and the distance. To get rid of extra decimals, `%i` converts distance to an integer.

Real Time or Fast?

If you tried HC-SR04 with both Arduino and Raspberry Pi, you probably noticed that Arduino gives more reliable and precise results. Arduino is more *real time* than Raspberry Pi.

Real time means that operations take the same, predictable amount of time. Real time doesn't mean fast. Raspberry Pi is much faster (more calculations per second), but it's not real time.

Arduino doesn't have an operating system, so the only thing happening on its CPU is the code you've written. Raspberry Pi has a whole operating system (Linux kernel, GNU tools, daemons, etc.). At any moment, many programs are running on a Raspberry Pi. They will affect how fast your program runs, and this makes things less predictable.

You have now gone through a long journey. You can sense the world using many different approaches. The most basic, even if tedious, way is to build a circuit component by component. Usually, the easiest way is to use Arduino. And if you need a real computer, there is Raspberry Pi.

Now it's time to make your skills serve your imagination.

Build your own ideas into prototypes. Make your own robots. Innovate devices and gadgets that have not been made before. We've written several books that can help you.

To learn advanced sensors, read *Make: Sensors*. To add an EEG sensor to your repertoire, build the bot in *Make a Mind-Controlled Arduino Robot*. *Make: Arduino Bots and Gadgets* brings you six fun projects and shows how to achieve impressive results quickly.

Go make your projects!

Star Diwa

A/Troubleshooting Tactics

No one gets everything right the first time, so it’s a good idea to be deliberate in how you approach troubleshooting. Here are some tactics we use when our projects do not behave as expected:

Verify hardware with “Hello, world”

By this, we mean get the simplest possible program working on your hardware before you introduce complexity. Traditionally, a “Hello, world” program prints a string of text onto a screen, but of course, many sensor projects don’t call for a display. Our Arduino and Raspberry Pi “Hello, world” analog is usually just a short program that will blink an onboard LED on and off at a specific intervals.

Always build and test in small steps

Racing through a project only to find it does not work is never a good experience. Our approach is to deliberately build and test as the project progresses, not wait until the end to run the first test. This tactic means we don’t end up with an extremely complex problem to troubleshoot when the build is done. Testing is time well spent! Additionally, if you do create a habit of regular testing, it is much easier to go back to a known working state. Don’t try to keep pushing on when things don’t work. If you do, you’ll end up frustrated with a pile of messy code and wiring that nobody, including you, can understand. Building in small steps does not mean that you have to understand every line of the code or inner workings of your sensor. One working component and a basic understanding of how it works is good enough.

Isolate the problem

Let’s say you are trying to make an Arduino project where a servo is controlled by moving a rotary knob (a potentiometer). Taking our advice, you built the project methodically and in small, manageable steps. But the servo is simply not moving as you expect when the knob is turned. Here is how we’d approach the issue. First, only connect the servo to your Arduino and upload the most basic code that controls servo movement. If the servo moves as expected, then it’s on to test the knob in the same way you just did for the servo. If the servo does not move as expected, you’ve isolated the problem. If both parts work independently, then you can be fairly confident the issue is in the code connecting the two components.

RTFM (read the friendly manual)

Reading documentation is often overlooked as a troubleshooting resource, but it's exactly where you should look for help! If there was no manual accompanying a sensor you bought, try checking the manufacturer's website. Another great place is online forums, but you should be careful not to blindly assume that, because it's in a forum or online, it conveys the correct answer. There are lots of flawed instructions and even nonfunctional sample programs, so it's good to have some skepticism while reading. (Another thing to be aware of: the F in RTFM isn't always considered to stand for *Friendly*. It's best not to use the acronym in an email to your best client.)

Document your work

Comment your code! No need to comment every line, but comment the code where it's tricky or where it took you a while to get it working properly. Photograph the stages of your build. It's great to have images for future use in a tutorial, but it's helpful to have them even as you are iterating through different project designs. Otherwise, you'll forget what you did in no time. If you want to kick it up a notch, publish your results online!

Star Diwa

B/Arduino IDE Setup

Here's how to set up the Arduino integrated development environment (IDE) on various platforms.

Ubuntu Linux

Connect Arduino to your computer with a USB cable. Arduino draws power directly from the USB, so no external power supply is needed. Start the terminal application.

> You can start the command-line terminal in many ways. You can open it in the main menu with Applications → Accessories → Terminal (on Xubuntu and other XFCE-based distributions such as Debian with XFCE). Super-T, also known as the *ugly key* or *Windows key*, works on many desktops. If you are using Unity in the standard Ubuntu distribution, search for "Terminal" in Dash (top-left corner).

To install the Arduino IDE, install the `arduino` package. Here's how you'd do it on Ubuntu Linux:

```
$ sudo apt-get update
$ sudo apt-get -y install arduino
```

Give yourself permission to access the serial-over-USB port (this is required by the Arduino development environment to function). The first command adds you to the dialout group, and the second command switches you into that group without you needing to log out and back in again:

```
$ sudo adduser $(whoami) dialout
$ newgrp dialout
```

Start Arduino:

```
$ arduino
```

The Arduino IDE opens.

After you have logged out and back in, you can also start Arduino IDE from the menus.

Now you're ready to test your installation. See "Hello, World" on page 99.

Windows 7 and 8

Download the latest version of the Arduino software and unzip the file you downloaded to any location that you find suitable (your *Desktop* or *Downloads* directory, for example).

Connect your Arduino Uno to your computer with a USB cable. Arduino draws power directly from the USB, so no external power supply is needed. Windows will start an automatic installation process for the Arduino drivers. It may fail after a while and display an error dialog.

If Windows fails to install the driver:

1. Open Windows Explorer, right-click Computer, and choose Manage.
2. From Computer Management, choose Device Manager on the left. Locate Arduino Uno in the device list, right-click it, and choose Update Driver Software.
3. Choose "Browse my computer for driver software." Navigate to the Arduino folder you extracted, open the *drivers* directory, choose *arduino.inf*, and click Next.
4. Windows will now install the driver.

Launch the Arduino IDE by double-clicking the Arduino icon inside the folder you unzipped.

Time to test your installation; see "Hello, World" on page 99.

OS X

Download the latest version of the Arduino software and unzip the file you downloaded. Copy it to your */Applications* folder.

Connect your Arduino Uno to your computer with a USB cable. Arduino draws power directly from the USB, so no external power supply is needed. You don't need to install a driver for OS X.

Launch the Arduino IDE by double-clicking the Arduino icon in the */Applications* folder.

Time to test your installation; see "Hello, World" on page 99.

Hello, World

Now that you have the Arduino IDE open, you can run the Arduino equivalent of "Hello, world."

First, confirm that you have the correct board selected. The Arduino Uno is the default. If you have another board, such as a Mega or a Leonardo, choose it from the Tools → Board menu.

Now you need to load the Blink test program. Choose File → Examples → 1.Basics → Blink. Click the Upload button (or choose File → Upload) to compile and upload your program to Arduino.

The first time you do this, Arduino may display an error popup: "Serial port COM1 not found." That's because you haven't chosen which serial port to use (the connection between your computer and Arduino is represented as a USB serial port). Select your serial port from the drop-down menu. On Linux, it's probably */dev/ttyACM0*; on the Mac, it may be something like */dev/usbmodem1234*; and on Windows, it's one of the COM ports.

> If you see a different error message instead of a request to choose a serial port, choose your serial port from Tools → Port. If you can't figure out which port Arduino is connected to, pay attention to the ports listed, unplug the Arduino, and make a note as to which port went away. That's the Arduino port. OS X lists each port twice (e.g., as */dev/cu.usbmodem1234* and */dev/tty.usbmodem1234*). Either one will work.

While the program is uploading, Arduino's TX and RX (transmit and receive) lights blink rapidly. Finally, when the program is running, the tiny light labeled "L" is blinking.

The L LED blinking means that everything was successfully installed, and you just got your first sketch running.

Congratulations! Remember this simple procedure: if you ever get so stuck you are wondering whether Arduino is even running your code at all, return to this "Hello, world" example. Whenever you start a new program, start with a "Hello, world" to make sure that everything is working.

Star Diwa

C/Setting Up Raspberry Pi

To make Raspberry Pi work, you need to install an operating system. Without it, you won't even get a picture on the screen.

To install an operating system, you'll need to do the following:

- Prepare a memory card: format it and extract a compressed zip archive.
- Connect cables and boot to the installer; Raspberry Pi boots automatically when you connect the power cable.
- Install Raspbian; just select it from the menu.
- Configure your Linux; it's easy using the menu that displays automatically.

Parts

You need the following parts for this project:

- Raspberry Pi Model B (the one with wired Internet)
- Micro USB cable (for power)
- 4 GB SD card (can be 8 GB too; Raspberry Pi Model B+ uses a microSD card)
- Display with HDMI port (your TV could qualify)
- HDMI cable
- USB mouse
- USB keyboard
- An SD card reader (for your regular computer; with the Raspberry Pi Model B+, you may need a microSD to SD adapter to read the card on your computer)

Optionally, it's convenient to have an Ethernet cable that connects the Raspberry Pi to the Internet.

Set It Up

The following sections describe how to set up the Raspberry Pi.

Prepare the Memory Card

To prepare the memory card, you download *NOOBS*.zip* and extract its contents to an SD card.

Did you buy an SD card with preinstalled Raspbian? You can skip this “Prepare Memory Card” section.

The SD memory card in the Raspberry Pi is the place to store data permanently, over reboots. As Raspberry Pi doesn’t have a hard disk, the memory card serves the same purpose.

Some shops also sell memory cards that have the operating system preinstalled. If you have one, you don’t have to install an operating system, and you can skip to the point where Raspberry Pi is booted from this card.

Buying the SD card is quite simple. The name SD defines both the technology and the physical size of the card. You want the big SD size (3.2 cm by 2.4 cm), not the smaller mini SD or tiny micro SD. Most cards nowadays are preformatted, and that’s convenient.

Open a web browser on your regular computer. Go to the official Raspberry Pi website and download the Quick Start Guide and the latest *NOOBS*.zip* using the “Direct download” or “Download ZIP” links. The name of the *NOOBS*.zip* varies, so we’ve marked the version number with a star. For example, the actual name could be *NOOBS_v1_2.zip*.

If you’re lucky, your memory card is preformatted when you buy it. If not, format it to FAT according to the instructions in the Quick Start Guide you just downloaded.

Put the SD memory card in the SD card reader of your regular computer. Your laptop might have an SD card reader by default. For most desktops, you need to buy a cheap external USB-connected reader.

Extract the contents of *NOOBS*.zip* to the memory card. In most modern Linux, Windows, and Mac environments, just double-click or right-click *NOOBS*.zip*. If you have trouble extracting the file in older Windows versions, download and install 7zip and extract with that.

After extraction, you should have the contents on the top level of the memory card. For example, *bootcode.bin* must be on the top level of the card and not inside a folder.

Connect Cables and Boot to Installer

Connect the cables to Raspberry Pi; they only fit in the correct holes. Connect the USB keyboard and mouse. Connect an HDMI cable between the Raspberry Pi and your television or monitor.

Start your television or monitor. Make sure it's set up to use the HDMI input to which the Raspberry Pi is connected.

Insert the SD card. The SD card only fits the slot the correct way. The SD card slot is on the short sides of the Raspberry Pi, the opposite side from the USB and Ethernet ports.

Finally, connect the micro USB cable between the Raspberry Pi and a power source. Raspberry Pi requires about 1 watt (1 W) of power, so a suitable power source is a desktop computer or a powerful USB charger.

A red PWR (power) LED lights up. Your monitor should show an image of a raspberry. This picture on the display means that Raspberry Pi is successfully running something from the SD card. Raspberry Pi doesn't have a BIOS similar to a workstation, so without software from the SD card, the screen is black.

After a moment, you see a menu of installable operating systems.

Install Raspbian

The installer menu shows you many different Linux distributions that you could install. Select Raspbian [RECOMMENDED] (see Figure C-1) and press Enter. (Or in the newest version, select "Install OS.") It's the most popular one. If you have ever used Ubuntu, Debian, or Mint, you will soon feel at home with Raspbian.

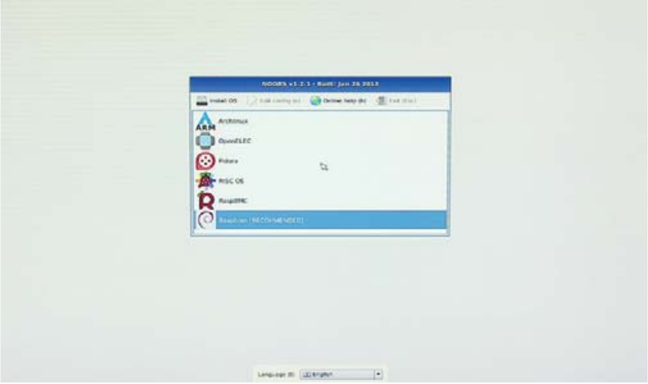

Figure C-1. *NOOBS boot screen*

Answer yes to the confirmation "All existing data on the SD card will be overwritten" by pressing Enter again. Enjoy the slideshow while Raspbian installs.

The installation takes less than 10 minutes. When installation completes, you get a pop-up "Image applied successfully." Press Enter for OK.

The Raspberry Pi reboots. You get a screen with a raspberry at the top left, and the log filling most of the screen. The log explains what's happening when Raspberry Pi boots. This can be quite entertaining to watch. Once the boot completes, you are in the "Raspberry Pi Software Configuration Tool," *raspi-config*.

Configuration with raspi-config

Raspi-config is a menu with the most common configuration tasks. You can move around with the arrow keys and Tab. To select items, press Enter or Return.

Change your password

> Move the selection to "2 Change User Password." The default username is "pi." The default password "raspberry" could be a security risk because everyone knows it, so the first task in the new system is to change the password. Choose a good password and type it at the prompt. What you type is not echoed on the screen, so you'll have to type it blindly. After typing the password twice, you get a confirmation of your successful password change. Then you return to the raspi-config menu.

Enable boot to desktop

> Even though Raspberry Pi can't replace a real workstation or a laptop, it's helpful in the beginning to have a web browser and multiple windows when working. Select "3 Enable Boot to Desktop." There is just one prompt, "Should we boot straight to desktop?" Press Enter to answer yes. Then you return to the raspi-config menu.
>
> Press Tab and the arrow keys to select <Finish> and Enter to choose it. In most keyboards, Tab is on the left side of the keyboard, above the Caps Lock key.

You can also configure your time zone, locale (you'll need to set this if you use a non-English language), keyboard, and more.

The big moment: "Would you like to reboot now?" Press Enter for <Yes>. Enjoy the log messages while Raspberry Pi shuts down.

Boot to Desktop

When Raspberry Pi boots, you briefly see four pixels filling the screen. Then you'll see the familiar Raspberry Pi logo on the top left and the boot log. After a moment, you are in a very light graphical desktop. There is a giant raspberry on the background (see Figure C-2).

Congratulations, you have now successfully installed Raspbian Linux on your Raspberry Pi!

But wait, there could be more to celebrate! If you haven't installed Linux on a computer before, congratulations again. Now you have installed Linux on a computer. (If this was your first time installing an operating system, our hat is off to you now.)

Now that you have installed Linux on the Raspberry Pi, do you want it on your desktop too? You can get a Free CD image for that from xubuntu or debian.

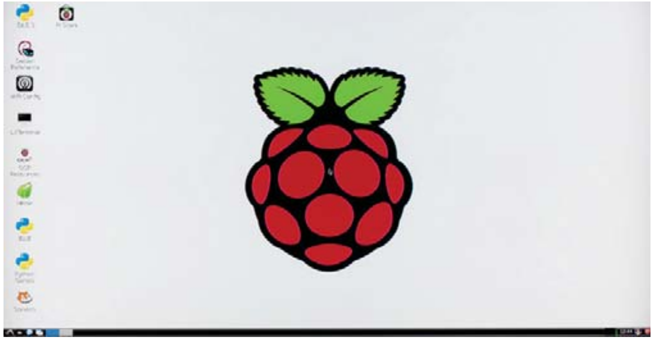

Figure C-2. *Booting to the desktop*

Using Raspberry Pi

The graphical desktop makes it easy for you to get started. And you'll get to swim in the deep end of the pool too, as you'll learn to use the famous Linux commands.

The graphical desktop is there to help learning and development. Raspberry Pi is not powerful enough to replace your laptop or desktop. When your project is ready, it's possible to run Raspberry Pi *headless*, without a display, if you want to.

The most important programs on the desktop are the terminal and web browser. You'll use LXTerminal for the command-line interface (CLI). To browse the Web, you use Midori, the light web browser.

If you connected to the wired Internet before booting up the Raspberry Pi, you can start browsing immediately. Double-click the Midori icon on the left of the desktop. When it opens, browse to *http://botbook.com* or *http://google.com*. This verifies that the Internet connection is working. You can also use the browser to copy and paste code examples from *http://botbook.com*, or to copy and paste your own code to your own blog. You do publish your successes, don't you?

To use the command-line interface, double-click the LXTerminal on the desktop. You'll get the prompt, waiting for your commands. It has some text, and at the end, a dollar sign. In the commands shown next, the dollar sign is the prompt, so you don't have to type that character.

Try writing some text. If you live outside the United States, it's likely that some of the characters are wrong. By default, you are using the US keymaps, even though your physical keyboard is different. If you didn't adjust the keyboard with raspi-config earlier, correct the keyboard now. Use your own two-letter ISO country code (fi, fr, de, etc.) in place of the default us.

```
$ setxkbmap us
```

Remember that you don't type the prompt (`$`), so you only need to type `setxkbmap us` and press Enter.

In the command line, you are always in a directory, which is your *current working directory*. You can ask the computer to print your working directory:

```
$ pwd
/home/pi/
```

As your username in this system is "pi," your home directory is */home/pi/*. This is the only place where a user can permanently store data.

You can change the directory:

```
$ cd /etc/
$ pwd
/etc/
$ cd /home/pi
```

If you ever get lost deep into the beautiful directory tree of Linux, it's easy to get back home. Return to your home directory with *cd /home/pi* or by using the shorthand `cd` all by itself:

```
$ cd
```

You can view the files in the working directory with `ls`:

```
$ ls
Desktop  ocr_pi.png  python_games
```

To create a new file, you can use a text editor:

```
$ nano foo.txt
```

When nano opens, type some text. Press Ctrl-X to save it. Then answer yes to "Save modified buffer?" by typing "Y". Accept the "File Name to Write: foo.txt" prompt by pressing Enter or Return.

You can see your new file in the file listing:

```
$ ls
Desktop  foo.txt  ocr_pi.png  python_games
```

To edit the same file again, type the same command:

```
$ nano foo.txt
```

If you want to see the contents of the file quickly, you can type:

```
$ cat foo.txt
```

When playing with hardware interfaces, you often write just one word into a text file. There is a nice shorthand way of doing that using a *pipe*. First, let's just write some text to the screen:

```
$ echo "Hello, world"
Hello, world
```

Not very surprising yet. But we can take the output of this program "Hello, world" and use it as input for another.

```
$ echo "Hello, world" | tee hello.txt
Hello, world
```

The text "Hello, world" is sent as input to the command `tee` which prints it on the screen and writes it to the specified file (*hello.txt*). The pipe symbol (|) in between the two commands is a character of its own, a tall vertical line.

How do you write the pipe character? In the US keyboard, hold Shift and press the "|\" key. The "|\" key is below the Backspace key and above Return/Enter. In the Finnish keyboard, for example, you can get the pipe character by pressing Alt-Gr and the "<>|" key near the left Shift key.

Check out the contents of the new file you created:

```
$ cat hello.txt
Hello, world
```

This trick will come in handy when you play with the file-based interface to the Raspberry Pi's hardware pins.

Python Programming

You can write programs just like any other text file. Write the first Python program, "Hello, world":

```
$ nano hello.py
```

For the contents of the *hello.py* file, type just this line:

```
print("Hello, world")
```

Press Ctrl-X, type "y," then press Enter to save it.

Now you can run your first Python program with the command *python hello.py*. It should print the text "Hello, world:"

```
$ python hello.py
Hello, world
```

Did you get the text to appear in the window? If so, you just ran some Python code.

Rootly Powers

One of the cornerstones of Linux security and stability is *user separation*. A normal user is only allowed to do things that only affect him. Just try it out:

```
$ raspi-config
Script must be run as root. Try 'sudo raspi-config'
```

You can't change system-wide settings as a normal user.

But luckily, your user "pi" is a member of the "sudo" group, so he can temporarily gain these root powers:

```
$ sudo raspi-config
```

You get the same configuration menu you used earlier to change the password and enable the graphical desktop. Use the Tab and arrow keys to select <Finish> and press Enter (or Return) to get back to command prompt.

Welcome to the Command Line

You have now played with the command-line interface. If you earn your living running Linux servers, you probably knew most of this already. Even on Linux desktops, you might have used the same commands. If not, welcome to the command-line interface!

The most important commands are those that allow you to move around (`ls`, `pwd`, `cd`) and manipulate text files (`nano`, `cat`).

As you'll soon try out yourself, all Linux configuration is in human readable text files. So by adding *sudo*, you can modify anything on the system.

The commands you just learned are surprisingly useful. Most of them work on all Linux system (Debian, Red Hat, Ubuntu, and others), all Unix systems, and even on OS X (which is based on Unix). They have stood the test of time. Many commands (`ls`, `cd`, `pwd`) have been in use before Linux, before the Web, and even before the Internet protocol suite. So you can likely keep using these commands far into the future.

Take Over an Onboard LED

Hello, hardware! Using this exercise, you will:

- Control the ACT onboard LED
- Meet the GPIO header, the main way to connect hardware to the Raspberry Pi
- Modify the system using virtual files in the */sys/* directory

Install the operating system (see Appendix C) if you haven't done so already. To use the command-line interface, open LXTerminal.

There are five LEDs near the USB connector: ACT, PWR, FDX, LNK, and 100. Could it be possible to take over one of these built-in LEDs?

Luckily, one of the LEDs can be programmatically controlled. The ACT LED is showing SD card activity by default, but we can disable this behavior. Have a look at the current setting:

```
$ cat /sys/class/leds/led0/trigger
none [mmc0]
```

The trigger file you just read shows that currently, the ACT LED (led0) is controlled by mmc0, so it's just showing SD card use. The brackets "[]" show the current selection. The other option is "none," so you will want to select that to control the LED:

```
$ echo "none"|sudo tee /sys/class/leds/led0/trigger
none
```

In Linux, all configuration is in plain-text files. All permanent, system-wide configuration is under */etc/*. And even the ever-changing state of the system can be read and modified trough virtual files in */sys/* and */proc/*. Linux gurus can usually guess the location of the configuration files for any aspect of the system, and then modify the files with just a text editor.

The first part of the command `echo "none"` just prints the word `none`. The second part writes this word to the trigger file. You can see that `none` is now selected, as it's surrounded by square brackets:

```
$ cat /sys/class/leds/led0/trigger
[none] mmc0
```

So now it's time to take control! Write "0" to the "brightness" file to turn lights out:

```
$ echo "0" |sudo tee /sys/class/leds/led0/brightness
```

Look at the Raspberry Pi board to see that the ACT LED is off. Then it's time for lights on:

```
$ echo "1" |sudo tee /sys/class/leds/led0/brightness
```

And you can see the green ACT LED light up.

The ACT light will revert to its default purpose after the next reboot. But if you want to change it back now, you can:

```
$ echo "mmc0"|sudo tee /sys/class/leds/led0/trigger
```

Did you succeed in controlling the light? Well done! You are well on your way to connecting the Raspberry Pi with the physical world.

As you gained control over the onboard LED, you:

- Manipulated the */sys/* virtual filesystem
- Had some hands-on experience with the command-line interface
- Succeeded in controlling some of the Raspberry Pi hardware

Now that you can manipulate the LED built in to the Raspberry Pi, it's time to connect your own.

Hello, GPIO: Connecting an External LED

Hello, external hardware! Soon you'll connect an external LED to your Raspberry Pi. This is the very first external hardware component you add.

Working trough this exercise, you'll control Linux trough virtual text files, and you'll get your first touch to the GPIO pin header.

This is the "Hello, world" for GPIO hardware. If you ever run into problems with any sensors, it's a good idea to repeat this exercise to solve any problems with this simple setup.

Parts

You need the following parts for this project:

- Raspberry Pi
- Female-male jumper wires, black and green

- Breadboard
- 470 Ω resistor (five-band: yellow-violet-black-black-any; four-band: yellow-violet-brown-any)
- An LED

Raspberry Pi has a GPIO pin header for connecting external components. The GPIO pin header has 26 pins (see Figure C-3). Most pins can have multiple functions, and you can choose which one you use. In this experiment, you learn the "out" mode to turn the pin on (3.3 V) and off.

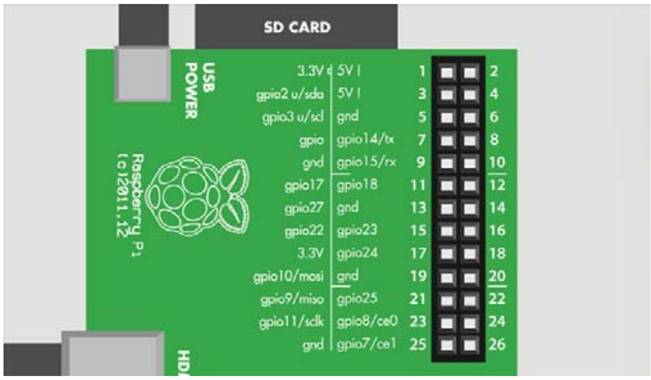
SD CARD
USB POWER
Raspberry Pi
(c)2011.12
HDI
3.3V
5V
1
2
gpio2 u/sda
5V
3
4
gpio3 u/scl
gnd
5
6
gpio
gpio14/tx
7
8
gnd
gpio15/rx
9
10
gpio17
gpio18
11
12
gpio27
gnd
13
14
gpio22
gpio23
15
16
3.3V
gpio24
17
18
gpio10/mosi
gnd
19
20
gpio9/miso
gpio25
21
22
gpio11/sclk
gpio8/ce0
23
24
gnd
gpio7/ce1
25
26

Figure C-3. *The GPIO pin header*

The Raspberry Pi Model B+ features more input/output pins, but the first 26 pins of the B+ header are compatible with earlier models. This means that you can follow the instructions in this book on the B+ without modification.

The circuit (see Figure 4-4) starts from the GPIO 27 pin that you control. When you want to turn on the LED, this pin is +3.3 V, the HIGH of Raspberry Pi. Then it has a resistor-limiting current, to protect both the LED and the more expensive Raspberry Pi. After the resistor, there is an LED in series. Finally, the minus leg of the LED is connected to ground (GND, 0 V).

Building the Circuit

Now it's time to build the circuit (see Figure C-4).

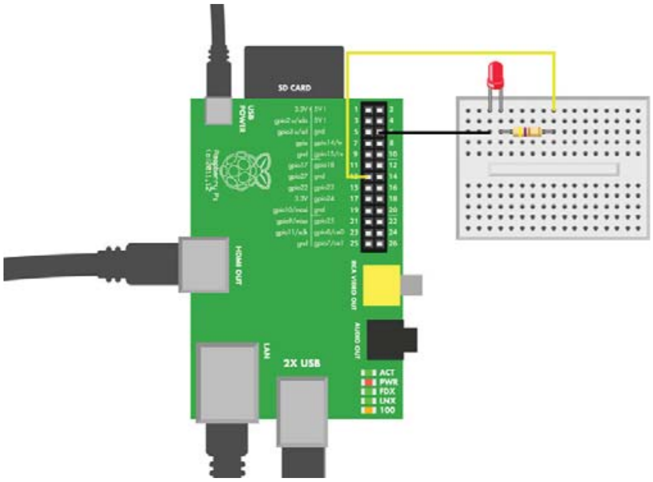

Figure C-4. *Hello, LED*

Unlike Arduino, Raspberry Pi is very sensitive to mistakes. You must be careful to only connect the pins you want to use, so you don't mistakenly short-circuit random pins together.

It is possible to build the circuit while the Raspberry Pi is on. Of course, mistakenly shorting two pins could then immediately break your Raspberry. The safer option is to shut down Raspberry Pi before connecting the wires. You can find the shutdown in the main menu, the menu on the bottom left of the desktop (Main Menu → Logout → Shutdown). After the shutdown, disconnect the powering USB cable.

Find the correct pins on the GPIO header. The pin number one is marked with a small white box on the board. The numbering diagram (see Figure C-3) shows both the human-friendly name (GPIO 27) and the running number of the pin (13). The pins are:

- GPIO 27 is pin number 13
- GND is pin number 14

Connect the female end of the green jumper wire to GPIO 27 and push the male end to a free row in the breadboard. Push one leg of the resistor to the same row with the jumper wire.

Push the free leg of the resistor to a free row on the breadboard. Push the plus leg of the LED to this row.

Push the male end of a black jumper wire into the row with the LED minus leg. Connect the female end of the black jumper wire to the GND pin on the GPIO header.

Verify that you got the same kind of connections as in Figure C-4. After you have double-checked the connections, power up the Raspberry Pi if it was off while making the connections.

Commands to Control GPIO

Open LXTerminal by double-clicking the LXTerminal icon on the desktop. Then take control of your GPIO pin.

In */sys/* you have text files that have information about the state of the system. These files are virtual, so they only exist in computer's memory.

Export the GPIO 27 pin, so we can use it:

```
$ echo 27|sudo tee /sys/class/gpio/export
```

This created a new folder *gpio27*. Turn the pin to "out" mode, so you can control it:

```
$ echo out|sudo tee /sys/class/gpio/gpio27/direction
```

Turn the pin on (+3.3 V, HIGH):

```
$ echo 1|sudo tee /sys/class/gpio/gpio27/value
```

The LED lights up. Did you get it to light up? Well done! You can now connect some hardware to your Raspberry Pi!

You can shut down the light, too:

```
$ echo 0|sudo tee /sys/class/gpio/gpio27/value
```

Troubleshooting

The LED is not lit, even though all commands seem to run. Check the LED polarity. The long plus leg goes to the plus side, the side of the GPIO 27 pin. On the minus side of the LED, there is a flat cut in the plastic. The minus side goes toward the black wire, GND. If this doesn't help, check the connections.

Hello, GPIO world! As you have learned, every project starts with a "Hello, world." If you ever run into problems with sensors, it could be a good idea to build and run this "Hello, GPIO" project again. That way, you get to solve the simple problems with a simple setup.

For us, "Hello, world" is usually the hardest project. It's not just an LED! Your "Hello, GPIO world" proves that:

- You have successfully installed Linux on your Raspberry Pi.
- You can use the command line.

- You can manipulate the state of the GPIO pins.
- You have obtained components that work with the Raspberry Pi.
- You have built a circuit that works with the Raspberry Pi (and did not fry it).

If there is a time for a break, a successful "Hello, world" might be it. Reward you brain with a cup of coffee, some jogging, or staring out of the window for a while.

Using GPIO Without Root

By default, you need to invoke root user privileges on your Raspberry Pi to access the GPIO pins in your code. This section shows you how to configure your Raspberry Pi to avoid that. You only need to go through this process once after you install your Raspberry Pi operating system.

Avoiding root privileges will make the system more secure and more stable. For example, think about a program that serves sensor data to the Web. Would you run a program that strangers can connect to as root?

In Linux, devices attached to your system are controlled by *udev*. Udev is a rule-based system that can run scripts when devices are plugged in.

Linux lets you control GPIO pins by manipulating files in */sys/class/gpio/*. By default, these files are owned by the user "root" and the group "root," which is why you can't change them without invoking root's superuser privileges. In this section, you'll see how to write a udev rule to change the group to "dialout." You'll then allow that group to read and write the files under */sys/class/gpio/*. Finally, you'll make the folders' group *sticky*, so that any newly created files and folders under it will also be owned by the "dialout" group.

All system-wide configuration in Linux is under */etc/*. Not surprisingly, udev configuration is in */etc/udev/*. Power up your Raspberry Pi, and open the terminal (LXTerminal). Next, open an editor with the `sudoedit` command so you can create a new rule file (don't type the `$`; that indicates the shell prompt that you see in the terminal window):

```
$ sudoedit /etc/udev/rules.d/88-gpio-without-root.rules
```

Add the text shown in Example C-1 to the file. Be sure to type each line as shown (don't type the numeric symbols; those are there to explain to you what is going on in this file). Udev rules are very sensitive to typos.

Want to save typing? Download the code, unzip it, and copy to */home/pi* on your Raspberry Pi SD card.

If you don't have permission to read the directory on Raspberry Pi, use `sudo chown -R pi.pi code/`.

Example C-1. 88-gpio-without-root.rules

```
# /etc/udev/rules.d/88-gpio-without-root.rules - GPIO without root on Raspberry
Pi  # ❶
# Copyright http://BotBook.com
# ❷
SUBSYSTEM=="gpio", RUN+="/bin/chown -R root.dialout /sys/class/gpio/"
SUBSYSTEM=="gpio", RUN+="/bin/chown -R root.dialout /sys/devices/virtual/gpio/"
# ❸
SUBSYSTEM=="gpio", RUN+="/bin/chmod g+s /sys/class/gpio/"
SUBSYSTEM=="gpio", RUN+="/bin/chmod g+s /sys/devices/virtual/gpio/"
# ❹
SUBSYSTEM=="gpio", RUN+="/bin/chmod -R ug+rw /sys/class/gpio/"
SUBSYSTEM=="gpio", RUN+="/bin/chmod -R ug+rw /sys/devices/virtual/gpio/"
```

❶ This comment explains the purpose of the file.

❷ Sets the owner of the two directories to be root, and the group to be dialout.

❸ Sets the *sticky bit* flag on these two directories.

❹ Configures the permissions on the directories to give members of the dialout group read and write permission.

The rules are processed in numeric order, but this is probably the only rule affecting the GPIO directories, so the number does not matter. In Morse code (CW), 88 is short for hugs and kisses. We prefer it over the often-picked number 99, which means "get lost".

To avoid typing and inevitable typos, you can download a copy of the *88-gpio-without-root.rules* file from *http://botbook.com* and copy it in place with *sudo cp 88-gpio-without-root.rules /etc/udev/rules.d/*. To edit the file, you can use `sudoedit`, which opens the file as root in the `nano` text editor:

```
$ sudoedit /etc/udev/rules.d/88-gpio-without-root.rules
```

Save the file (Ctrl-X, press y, and then press Enter).

To use your new rules, restart the udev daemon and trigger your new rule with these commands:

```
$ sudo service udev restart
$ sudo udevadm trigger --subsystem-match=gpio
```

Next, check whether the ownership on the files was set correctly:

```
$ ls -lR /sys/class/gpio/
```

The listing should mention the "dialout" group many times. The parameter `-l` means to display a long listing (with owner, group, and permissions), and `-R` means recursively list directory contents, too. If you don't see dialout in the listing, restart your Raspberry Pi and check again with the `ls -lR /sys/class/gpio/` command. If it still doesn't work, go through the steps to configure the udev rule again.

Once you have the new udev rule in place, you won't need to go through the preceding steps again unless you reinstall the operating system on your Raspberry Pi.

Star Diwa

D/Bill of Materials

This section lists the parts you'll need to complete the projects in this book.

Chapter 1

The parts for "Project 1: Photoresistor to Measure Light" on page 3 are as follows:

- Photoresistor
- 5 mm red LED (different LEDs will work differently with this circuit; later, you'll learn a more sophisticated way to fade LEDs)
- 470 Ω resistor (four-band resistor: yellow-violet-brown; five-band resistor: yellow-violet-black-black; the last band will vary depending on the resistor's tolerance)
- Breadboard
- 9 V battery clip
- 9 V battery

Chapter 2

The parts for "Project 2: A Simple Switch" on page 9 are as follows:

- A switch
- A wire with alligator clips
- Two 1.5 V batteries
- Battery holder with wire leads
- LED
- 470 Ω resistor (four-band resistor: yellow-violet-brown; five-band resistor: yellow-violet-black-black; the last band will vary depending on the resistor's tolerance)
- Breadboard

All of these except the 9 V battery and 470 Ω resistor are available in the Maker Shed Mintronics: Survival Pack, part number MSTIN2. You can use two of the 220 Ω resistors in series or one 1 kΩ resistor in place of the 470 Ω resistor; both of these are available from electronics retailers such as RadioShack.

The parts for "Project 3: Buzzer Volume Control" on page 12 are as follows:

- DC piezo buzzer (Maker Shed part number MSPT01)
- 10 K potentiometer
- 9 V battery
- Battery clip
- Breadboard

All of these except the 9 V battery and piezo buzzer are available in the Maker Shed Mintronics: Survival Pack, part number MSTIN2.

The parts for "Project 4: Hall Effect" on page 14 are as follows:

- DC piezo buzzer
- 9 V battery
- Battery clip
- Breadboard
- NJK-5002A Hall effect switch

The parts for "Project 5: Firefly" on page 17 are as follows:

- Two BC547 transistors
- Light-dependent resistor (LDR)
- 470 Ω resistor (four-band resistor: yellow-violet-brown; five-band resistor: yellow-violet-black-black; the last band will vary depending on the resistor's tolerance)
- 10 kΩ resistor (four-band resistor: brown-black-orange; five-band resistor: brown-black-black-red; the last band will vary depending on the resistor's tolerance)
- 5 mm red LED
- 9 V battery

- 9 V battery clip
- Breadboard
- CA555E timer integrated circuit
- 10 kΩ potentiometer
- 100 uF capacitor

All of these parts, except the 9 V battery, 470 Ω resistor, and BC547, are available in the Maker Shed Mintronics: Survival Pack, part number MSTIN2. You can use two of the 220 Ω resistors in series or one 1 kΩ resistor in place of the 470 Ω resistor; these are both available from electronics retailers such as RadioShack.

Chapter 3

Chapter 3 requires an Arduino to interface with your sensors. If you need to buy one, purchase an Arduino Uno or an Arduino Leonardo.

The parts for "Project 6: Momentary Push-Button and Pull-Up Resistors" on page 34 are as follows:

- Momentary push button
- Arduino Uno
- Jumper wires
- Breadboard

The parts for "Project 7: Infrared Proximity to Detect Objects" on page 40 are as follows:

- Infrared sensor switch
- Arduino Uno
- Jumper wires

The parts for "Project 8: Rotation (Pot)" on page 43 are as follows:

- Potentiometer (around 10 kΩ recommended)
- Arduino Uno
- Jumper wires
- Breadboard

The parts for "Project 9: Photoresistor to Measure Light" on page 47 are as follows:

- Photoresistor (10 K recommended)
- 10 kΩ resistor (four-band resistor: brown-black-orange; five-band resistor: brown-black-black-red; the fourth or fifth band will vary depending on the resistor's tolerance)
- Arduino Uno
- Jumper wires
- Breadboard

The parts for "Project 10: FlexiForce to Measure Pressure" on page 49 are as follows:

- Force-sensitive resistor (FSR)
- 1 MΩ resistor (four-band resistor: brown-black-green; five-band resistor: Brown-black-black-yellow; the fourth or fifth band will vary depending on the resistor's tolerance)
- Arduino Uno
- Jumper wires
- Breadboard

The parts for "Project 11: Measuring Temperature (LM35)" on page 52 are as follows:

- LM35 temperature sensor
- Arduino UNO
- Jumper wires
- Breadboard

The parts for "Project 12: Ultrasonic Distance Measuring (HC-SR04)" on page 56 are as follows:

- HC-SR04 ultrasonic distance sensor
- Arduino UNO
- Breadboard
- Jumper wires

Chapter 4

The parts for "Project 13: Momentary Push Button" on page 64 are as follows:

- Momentary Push Button
- Raspberry Pi
- Female-male jumper wires (Maker Shed part #MKKN5), black and green
- A breadboard

The parts for "Project 14: Blink an LED with Python" on page 69 are as follows:

- Raspberry Pi
- Female-male jumper wires (Maker Shed part #MKKN5), black and green
- An LED
- 470 Ω Resistor (four-band resistor: yellow-violet-brown; five-band resistor: yellow-violet-black-black; the last band will vary depending on the resistor's tolerance)
- A breadboard

The parts for "Project 15: Adjustable Infrared Switch" on page 73 are as follows:

- Infrared sensor switch
- Two 1 kΩ resistors (five band: brown-black-red-brown-any, four band: brown-black-red-any)
- Raspberry Pi

The parts for "Project 16: Potentiometer to Measure Rotation" on page 77 are as follows:

- Potentiometer
- MCP3002 (analog-to-digital integrated circuit)
- Raspberry Pi
- Jumper wires (male-to-female and male-to-male, which are Maker Shed parts #MKKN5 and #MKSEEED3)
- Breadboard

The parts for "Project 17: Photoresistor" on page 82 are as follows:

- Photoresistor (10 kΩ suggested)
- Resistor (10 kΩ, or choose a value that matches the photoresistor)
- MCP3002
- Raspberry Pi

- Jumper wires (male-to-female and male-to-male, which are Maker Shed parts #MKKN5 and #MKSEEED3)
- Breadboard

The parts for "Project 18: FlexiForce" on page 85 are as follows:

- Flexiforce (11 kg, 25 pound version recommended)
- Resistor (1 MΩ, five band: brown-black-black-yellow-any, four band: brown-black-green-any)
- MCP3002
- Raspberry Pi
- Jumper wires (male-to-female and male-to-male, which are Maker Shed parts #MKKN5 and #MKSEEED3)
- Breadboard

The parts for "Project 19: Temperature Measurements (LM35)" on page 86 are as follows:

- LM35 temperature sensor
- MCP3002
- Raspberry Pi
- Jumper Wires (male-to-female and male-to-male, which are Maker Shed parts #MKKN5 and #MKSEEED3)
- Breadboard

The parts for "Project 20: Ultrasonic Distance" on page 89 are as follows:

- HC-SR04 sensor
- Two 10 KΩ resistors (four band: brown-black-orange-any, five band: brown-black-black-red-any)
- Raspberry Pi
- Jumper wires (male-to-female and male-to-male, which are Maker Shed parts #MKKN5 and #MKSEEED3)
- Breadboard

Star Diwa

Star Diwa

Index

Symbols

- $ prompt, 66
- 555
  - astable mode, 18, 26
  - bistable mode, 18
  - input pins, 26
  - modes of operation, 18
  - monostable mode, 18
  - output, 26
- 555 timer, 17, 26

A

- actuator, 9
- ADC (see analog-to-digital converter)
- amplifier, 21
- analog resistance sensors (see sensors, analog)
- analog-to-digital converter (ADC), 63, 77, 78
- analogRead() function, 44
- apt-get command, 78
- Arduino, 8, 24, 33
  - analog input, 46
  - analogRead() function, 44
  - assigning variables, 37
  - blink test, 99
  - digital input, 36
  - digitalRead() function, 36
  - equality comparison, 37
  - HIGH, 37
  - input pins, 36
  - installation, 97–99
  - installation on Linux, 97–98
  - installation on Mac, 98
  - installation on Windows, 98
  - loop() function, 36
  - LOW, 37
  - OS X setup, 98
  - output pins, 36
  - programming, 19
  - serial port, 50
  - Serial.println() function, 52
  - setup() function, 36
  - timing, 34
  - vs Raspberry Pi, 92

B

- base, 23
- bimetallic coil, 7
- books
  - Getting Started with Arduino, 1
  - Getting Started with Raspberry Pi, 1
  - Make a Mind-Controlled Arduino Robot, 50, 93
  - Make: Arduino Bots and Gadgets, 1, 50, 52, 77, 93
  - Make: Sensors, 1, 93
- breadboard, 5, 20, 26
  - gap, 6
- Broadcom, 18, 65
- button, 9
- buying parts, viii–ix

C

- capacitor, 28–28
- cat command, 66
- cd command, 68
- cold temperatures
  - measuring, 54
- collector, 23
- command-line interface (CLI)
  - in Linux, 97
- converting raw values
  - to percentage, 55, 84
  - to voltage, 84

D

- digitalRead() function, 36
- documentation, 96

E

- echo command, 66
- emitter, 23

F
files, writing, 72
FlexiForce, 43
floating pin, 37, 66
floating point, 60
force, 49, 85

G
general-purpose input/output (GPIO), 34, 63, 65, 78
getting help, x

H
Hello, world (blink test), 99
hookup wires, 20

I
infrared, 40, 73
input devices, 2
input pin, floating, 37, 66
integrated circuits, 17, 29
  identifying, 29

J
jumper wires, 20
  color, 20

L
LDR, 19, 21 (see photoresistor)
light, 2, 7, 17, 47
light-emitting diode (LED), 3, 4, 6, 7, 9, 17, 21, 24, 26, 29, 35, 69
  choosing a resistor, 12
  polarity, 5
Linux, 63
  Arduino setup, 97
loop() function, 36
ls command, 66, 80, 116

M
magnetism, 14, 16
MCP3002, 78, 82, 85, 86
MCP3200 analog-to-digital converter, 63
movement, 43, 77

N
nano text editor, 68, 115
NJK-5002A, 15
NJK-5002C, 15

O
oscillator, 27

P
photoresistor, 3, 5, 6, 47, 82
potentiometer, 43, 46, 77, 82
pressure, 49, 85
projects
  555 firefly, 17–31
  Arduino distance measurement, 56–60, 89–92
  Arduino force sensing, 49–52
  Arduino infrared proximity, 40–43
  Arduino light measurement, 47–49
  Arduino potentiometer, 43–47
  Arduino pushbutton, 34–39
  Arduino temperature measurement, 52–56
  Hall effect, 14–17
  measure light, 3–7
  Raspberry Pi force sensing, 85–86
  Raspberry Pi infrared proximity, 73–77
  Raspberry Pi LED blink, 69–73
  Raspberry Pi light measurement, 82–83
  Raspberry Pi potentiometer, 77–82
  Raspberry Pi push button, 64–68
  Raspberry Pi temperature measurement, 86–89
  simple switch, 9–12
  volume control, 12–14
prototyping, 23
pull-up resistor, 12
Python, 68
  GPIO library, 75
  importing libraries, 71

R
Raspberry Pi, 8, 18, 63
  configuring GPIO, 114–116
  configuring SPI, 78
  installation, 101–116
  programming, 19
  reading GPIO pins, 66–67
  shutting down, 64
  vs Arduino, 92
resistance, 3, 7
resistor, 3, 4, 10–29, 48, 50, 69, 73, 82, 85
  force-sensitive, 50
  pull-down, 38
  pull-up, 12, 36–39, 66
  selecting, 12

risks
  component burnout, 12
  damaging Arduino, 34
  damaging components, 34
  damaging Raspberry Pi, 66, 87, 90
  overheating, 12

S

seasonal affective disorder, 3
senses
  human, vii
  robotic, vii
sensing
  force, 49–52, 56–60, 85–86, 89–92
  light, 3–7, 17–31, 47–49, 82–83
  magnetism, 14–17
  proximity, 40–43, 73–77
  rotation, 43–47, 77–82
  temperature, 52–56, 86–89
sensors, 9
  analog, 43, 49, 77
  connecting, 41
  digital resistance, 43
  distance, 57
  electro-mechanical, 7
  FlexiForce, 85
  infrared proximity, 41, 73
  resistive, 7
  temperature, 53, 86
serial port, 50
  speed, 52
Serial.println() function, 52
setup() function, 36
shopping for components, viii–ix
sketches, 33, 38
  example, 38
smart devices, 1
smartphone, 1
sound, 14, 16, 56, 89
speed of sound, 59
SPI, 78
spidev, 78
sudo command, 115
sudoedit command, 115
switch, 10
  closed, 37
  on/off, 9, 34, 64
  open, 37
  single pole double throw (SPDT), 10
  single pole single throw (SPST), 10

T

temperature, 52, 86
  effect on speed of sound, 59
testing, 95
timer, 555 (see 555 timer)
transducer, 9
transduction, 9
transistor, 21–22, 53
  bipolar junction, 23
  BJT, 23
  NPN, 23
twisted-wire connection technique, 11

U

Ubuntu, 97
udev, 79
ugly key, 97

V

variables, 36, 71
  global, 52
voltage divider, 21, 43, 50, 76–77, 82

W

Windows 7/8, 98
Windows key, 97

Star Diwa

About the Authors

Kimmo Karvinen works as a chief technology officer for a hardware manufacturer that specializes in integrated AV and security systems. Before that, he worked as a marketing communications project leader and as a creative director and partner in an advertisement agency. Kimmo's education includes a Master of Arts.

Tero Karvinen teaches Linux and embedded systems in Haaga-Helia University of Applied Sciences, where his work has also included curriculum development and research in wireless networking. He previously worked as a CEO of a small advertisement agency. Tero's education includes a Master of Science in Economics.

The cover and body font is Benton Sans, the heading font is Serifa, and the code font is Bitstream Vera Sans Mono.

Sensors/ DIY Projects

Getting Started with Sensors

Sensors surround us. The world is full of them: infrared sensors in motion detectors, carbon monoxide detectors in homes—even tiny accelerometers, GPS modules, and cameras inside your smartphone. Thanks to the proliferation of mobile devices, sensors are now remarkably affordable, meaning you can combine inexpensive sensors and microcontroller boards to make your own devices.

This fully-illustrated, hands-on book teaches you to connect and read sensors through a series of short projects. You'll learn to use the immensely popular Arduino and Raspberry Pi boards to process data through simple programs you develop.

In *Getting Started with Sensors*, you'll write programs and build devices that respond to:

- **Rotation with a potentiometer**
- **Distance with ultrasound**
- **Proximity with infrared**
- **Light and dark with a photoresistor**
- **Temperature with a thermometer**
- **Relative humidity with a humidity sensor**

Star Diwa

You'll also work with resistive and electromechanical sensors, as well as switches, transducers, potentiometers, buzzers, 555 timers, and voltage dividers.

There's a whole world out there full of sensory impressions you can control, track, and manage. *How will you capture it?*

For more sensor projects, check out the authors' other books: *Make: Arduino Bots and Gadgets* and *Make: Sensors*.

US $19.99 CAN $20.99
ISBN: 978-1-4493-6708-4

51999
9 781449 367084

Make:
makezine.com

spine
.562"

Technology & Engineering/Electronics

Make: Electronics

Learning by Discovery

Want to learn the fundamentals of electronics in a fun, hands-on way? With *Make: Electronics 2nd Edition*, you'll start working on real projects as soon as you crack open the book. Explore all of the key components and essential principles through a series of fascinating experiments. You'll build the circuits first, *then* learn the theory behind them!

Build working devices, from simple to complex

You'll start with the basics and then move on to more complicated projects. Go from switching circuits to integrated circuits, and from simple alarms to programmable microcontrollers.

With step-by-step instructions and more than 500 full-color photographs and illustrations, this book will help you use and understand electronics concepts and techniques.

New in the 2nd Edition!

- **» Totally new text throughout, with many new projects and others revised and updated.**
- **» Smaller, more affordable range of components.**
- **» Now includes coverage of Arduino!**

You will:

- **» Discover by breaking things: experiment with components and learn from failure.**
- **» Set up a tricked-out project space: make a work area at home, equipped with the tools and parts you'll need.**
- **» Learn about key electronic components and their functions within a circuit.**
- **» Create an intrusion alarm, holiday lights, wearable electronic jewelry, audio processors, a reflex tester, and a combination lock.**
- **» Get clear, easy-to-understand explanations of what you're doing and why.**

Charles Platt's first electronics project was a telephone answering machine, which he built when he was 15. He became a science-fiction writer (author of ***The Silicon Man***), taught classes in computer graphics, and was a senior writer at Wired, but he has retained his lifelong love for hobby electronics. He is currently a contributing editor to ***Make:*** magazine.

"This is teaching at its best!"—Hans Camenzind, inventor of the 555 timer

US $34.99 CAN $40.99
ISBN: 978-1-68045-026-2
9 781680 450262 53499

Make:
makezine.com

Make: Electronics 2nd Edition Charles Platt

Burn things out, mess things up—that's how you learn.

Make: Electronics

Second Edition

Charles Platt

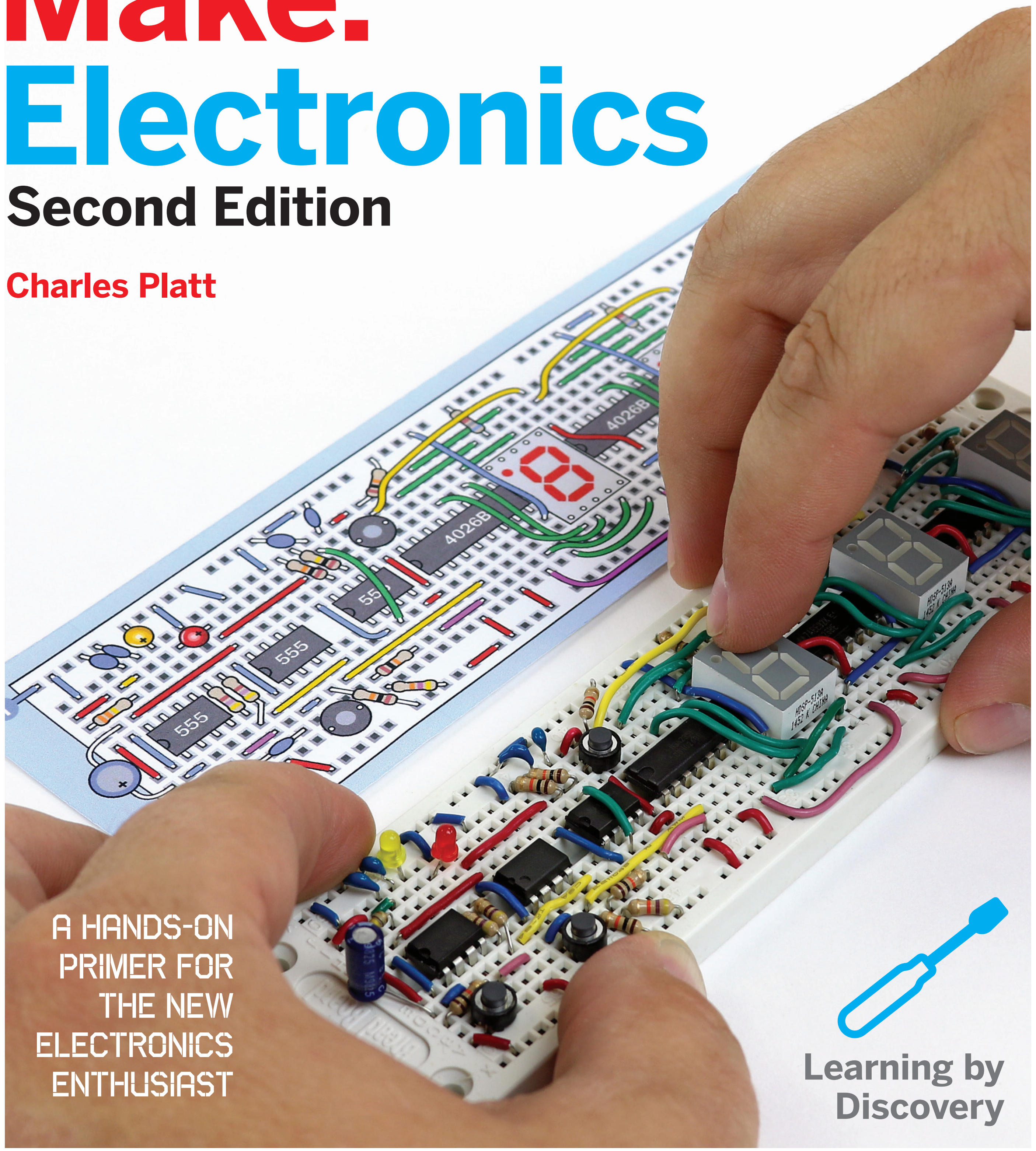

A HANDS-ON
PRIMER FOR
THE NEW
ELECTRONICS
ENTHUSIAST

Learning by
Discovery

Technology & Engineering/Electronics

Make: Electronics

Learning by Discovery

Want to learn the fundamentals of electronics in a fun, hands-on way? With *Make: Electronics 2nd Edition*, you'll start working on real projects as soon as you crack open the book. Explore all of the key components and essential principles through a series of fascinating experiments. You'll build the circuits first, *then* learn the theory behind them!

Build working devices, from simple to complex

You'll start with the basics and then move on to more complicated projects. Go from switching circuits to integrated circuits, and from simple alarms to programmable microcontrollers.

With step-by-step instructions and more than 500 full-color photographs and illustrations, this book will help you use and understand electronics concepts and techniques.

New in the 2nd Edition!

» Totally new text throughout, with many new projects and others revised and updated.

» Smaller, more affordable range of components.

» Now includes coverage of Arduino!

You will:

» Discover by breaking things: experiment with components and learn from failure.

» Set up a tricked-out project space: make a work area at home, equipped with the tools and parts you'll need.

» Learn about key electronic components and their functions within a circuit.

» Create an intrusion alarm, holiday lights, wearable electronic jewelry, audio processors, a reflex tester, and a combination lock.

» Get clear, easy-to-understand explanations of what you're doing and why.

Charles Platt's first electronics project was a telephone answering machine, which he built when he was 15. He became a science-fiction writer (author of ***The Silicon Man***), taught classes in computer graphics, and was a senior writer at Wired, but he has retained his lifelong love for hobby electronics. He is currently a contributing editor to ***Make:*** magazine.

"This is teaching at its best!"—Hans Camenzind, inventor of the 555 timer

US $34.99 CAN $40.99
ISBN: 978-1-68045-026-2

9 781680 450262 53499

Make:
makezine.com

Make: Electronics
2nd Edition
Charles Platt

Burn things out, mess things up—that's how you learn.

Make: Electronics

Second Edition

Charles Platt

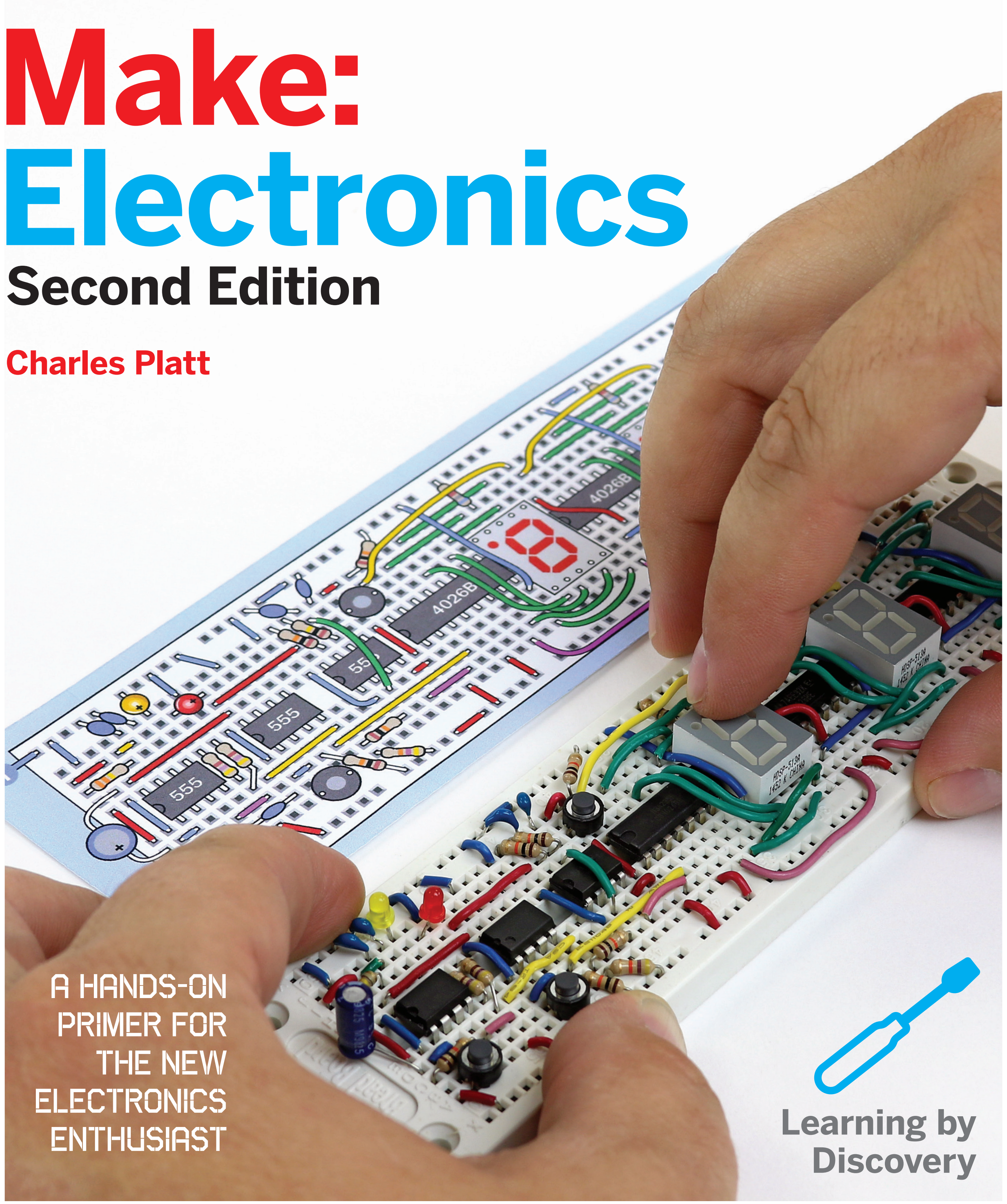

A HANDS-ON
PRIMER FOR
THE NEW
ELECTRONICS
ENTHUSIAST

Learning by
Discovery

SECOND EDITION

Make: Electronics

Charles Platt

MAKER MEDIA™
SAN FRANCISCO, CA

Make: Electronics

by Charles Platt

Copyright © 2015 Charles Platt. All rights reserved.

Printed in Canada.

Published by Maker Media, Inc., 1160 Battery Street East, Suite 125, San Francisco, CA 94111.

Maker Media books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*http://safaribooksonline.com*). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

Editor: Brian Jepson
Production Editor: Nicole Shelby
Proofreader: Charles Roumeliotis
Indexer: Charles Platt

Interior Designer: David Futato
Cover Designers: Marc de Vinck and Charles Platt
Interior Photographs and Illustrations: Charles Platt

September 2015: Second Edition
December 2009: First Edition

Revision History for the Second Edition

2015-08-07: First Release

See *http://oreilly.com/catalog/errata.csp?isbn=9781680450262* for release details.

The Make logo is a registered trademark of Maker Media, Inc. *Make: Electronics,* the cover image, and related trade dress are trademarks of Maker Media, Inc.

While the publisher and the author have used good faith efforts to ensure that the information and instructions contained in this work are accurate, the publisher and the author disclaim all responsibility for errors or omissions, including without limitation responsibility for damages resulting from the use of or reliance on this work. Use of the information and instructions contained in this work is at your own risk. If any code samples or other technology this work contains or describes is subject to open source licenses or the intellectual property rights of others, it is your responsibility to ensure that your use thereof complies with such licenses and/or rights.

978-1-680-45026-2

[TI]

Dedication

To readers of the first edition of *Make: Electronics* who contributed many ideas and suggestions for this second edition. In particular: Jeremy Frank, Russ Sprouse, Darral Teeples, Andrew Shaw, Brian Good, Behram Patel, Brian Smith, Gary White, Tom Malone, Joe Everhart, Don Girvin, Marshall Magee, Albert Qin, Vida John, Mark Jones, Chris Silva, and Warren Smith. Several of them also volunteered to review the text for errors. Feedback from my readers continues to be an amazing resource.

Acknowledgments

I discovered electronics with my school friends. We were nerds before the word existed. Patrick Fagg, Hugh Levinson, Graham Rogers, and John Witty showed me some of the possibilities.

It was Mark Frauenfelder who nudged me back into the habit of making things. Gareth Branwyn facilitated *Make: Electronics*, and Brian Jepson enabled the sequel and this new edition. They are three of the best editors I have known, and they are also three of my favorite people. Most writers are not so fortunate.

I am also grateful to Dale Dougherty for starting something that I never imagined could become so significant, and for welcoming me as a participant.

Russ Sprouse and Anthony Golin built and tested the circuits. Technical fact checking was provided by Philipp Marek, Fredrik Jansson, and Steve Conklin. Don't blame them if there are still any errors in this book. It's much easier for me to make an error than it is for someone else to find it.

Table of Contents

What's New in the Second Edition .... **vii**
Preface: How to Have Fun with This Book .... **ix**

1. **The Basics** **1**
  - Necessary Items for Chapter One 1
  - Experiment 1: Taste the Power! 7
  - Experiment 2: Let's Abuse a Battery! 12
  - Experiment 3: Your First Circuit 18
  - Experiment 4: Variable Resistance 22
  - Experiment 5: Let's Make a Battery 34

2. **Switching** **41**
  - Necessary Items for Chapter Two 41
  - Experiment 6: Very Simple Switching 50
  - Experiment 7: Investigating a Relay 60
  - Experiment 8: A Relay Oscillator 66
  - Experiment 9: Time and Capacitors 75
  - Experiment 10: Transistor Switching 84
  - Experiment 11: Light and Sound 91

3. **Getting Somewhat More Serious** **99**
  - Necessary Items for Chapter Three 99
  - Experiment 12: Joining Two Wires Together 109
  - Experiment 13: Roasting an LED 121
  - Experiment 14: A Wearable Pulsing Glow 124
  - Experiment 15: Intrusion Alarm, Part One 132

4. **Chips, Ahoy!** **141**
  - Necessary Items for Chapter Four 142
  - Experiment 16: Emitting a Pulse 145
  - Experiment 17: Set Your Tone 155

Experiment 18: Intrusion Alarm (Almost) Completed 163
Experiment 19: Reflex Tester 178
Experiment 20: Learning Logic 191
Experiment 21: A Powerful Combination 201
Experiment 22: Race to Place 210
Experiment 23: Flipping and Bouncing 218
Experiment 24: Nice Dice 223

5. What Next? 237
Tools, Equipment, Components, and Supplies 237
Customizing Your Work Area 237
Labeling 240
On the Bench 240
Reference Sources Online 242
Books 242
Experiment 25: Magnetism 243
Experiment 26: Tabletop Power Generation 247
Experiment 27: Loudspeaker Destruction 252
Experiment 28: Making a Coil React 255
Experiment 29: Filtering Frequencies 258
Experiment 30: Make It Fuzzy 267
Experiment 31: One Radio, No Solder, No Power 271
Experiment 32: Hardware Meets Software 278
Experiment 33: Checking the Real World 290
Experiment 34: Nicer Dice 297

6. Tools, Equipment, Components, and Supplies 311
Kits 311
Searching and Shopping Online 311
Checklists of Supplies and Components 316
Buying Tools and Equipment 324
Suppliers 326

Index 327

What's New in the Second Edition

All of the text from the first edition of this book has been rewritten, and most of the photographs and schematics have been replaced.

Single-bus breadboards are now used throughout (as in *Make: More Electronics*) to reduce the risk of wiring errors. This change entailed rebuilding the circuits, but I believe it was worthwhile.

Diagrams showing component placement are now used instead of photographs of breadboarded circuits. I think the diagrams are clearer.

Internal views of breadboard connections have been redrawn to match the revisions noted above.

New photographs of tools and supplies have been included. For small items, I have used a ruled background to indicate the scale.

Where possible, I have substituted components that cost less. I have also reduced the range that you need to buy.

Three experiments have been completely revised:

- The Nice Dice project that used LS-series 74xx chips in the first edition now uses 74HCxx chips, to be consistent with the rest of the book and with modern usage.
- The project using a unijunction transistor has been replaced with an astable multivibrator circuit using two bipolar transistors.
- The section on microcontrollers now recognizes that the Arduino has become the most popular choice in the Maker community.

In addition, two projects involving workshop fabrication using ABS plastic have been omitted, as many readers did not seem to find them useful.

All the page layouts have been changed to make them easily adaptable for handheld devices. The formatting is controlled by a plaintext markup language, so that future revisions will be simpler and quicker. We want the book to remain relevant and useful for many more years to come.

--Charles Platt, 2015

Preface: How to Have Fun with This Book

Everyone uses electronic devices, but most of us don't really know what goes on inside them.

You may feel that you don't need to know. You can drive a car without understanding the workings of an internal combustion engine, so why should you learn about electricity and electronics?

I think there are three reasons:

- By learning how technology works, you become better able to control your world instead of being controlled by it. When you run into problems, you can solve them instead of feeling frustrated by them.
- Learning about electronics can be fun, so long as you approach the process in the right way. It is also very affordable.
- Knowledge of electronics can enhance your value as an employee, or perhaps even lead to a whole new career.

Learning by Discovery

Most introductory guides begin by using definitions and theory to explain some fundamental concepts. Circuits are included to demonstrate what you have been told.

Science education in schools often follows a similar plan. I think of this as *learning by explanation*.

This book works the other way around. I want you to dive right in and start putting components together without necessarily knowing what to expect. As you see what happens, you will figure out what's going on. This is *Learning by Discovery*, which I believe is more fun, more interesting, and more memorable.

Working on an exploratory basis, you run the risk of making mistakes. But I don't see this as a bad thing, because mistakes are a valuable way to learn. I want you to burn things out and mess things up, to see for yourself the behavior and limitations of the parts that you are dealing with. The very low voltages used throughout this book may damage sensitive components, but will not damage you.

The key requirement of Learning by Discovery is that it has to be hands-on. You can derive some value from this book merely by reading it, but you will enjoy a much more valuable experience if you perform the experiments yourself.

Fortunately, the tools and components that you need are inexpensive. Hobby electronics should not cost significantly more than a recreation such as needlepoint, and you don't need a workshop. Everything can be done on a tabletop.

Will It Be Difficult?

I assume that you're beginning with no prior knowledge. Consequently, the first few experiments will be extremely simple, and you won't even use prototyping boards or a soldering iron.

I don't believe that the concepts will be hard to understand. Of course, if you want to study electronics more formally and do your own circuit design, that can be challenging. But in this book I have kept theory to a minimum, and the only math you'll need will be addition, subtraction, multiplication, and division. You may also find it helpful (but not absolutely necessary) if you can move decimal points from one position to another.

How This Book Is Organized

An introductory book can present information in two ways: in tutorials or in reference sections. I decided to use both of these methods.

You'll find the *tutorials* in sections headed as follows:

- Experiments
- What You Will Need
- Cautions

Experiments are the heart of the book, and they have been sequenced so that the knowledge you gain at the beginning can be applied to subsequent projects. I suggest that you perform the experiments in numerical order, skipping as few as possible.

You'll find *reference sections* under the following headings:

- Fundamentals
- Theory
- Background

I think the reference sections are important (otherwise, I would not have included them), but if you're impatient, you can dip into them at random or skip them and come back to them later.

If Something Doesn't Work

Usually there is only one way to build a circuit that works, while there are hundreds of ways to make mistakes that will prevent it from working. Therefore the odds are against you, unless you proceed in a really careful and methodical manner.

I know how frustrating it is when components just sit there doing nothing, but if you build a circuit that doesn't work, please begin by following the fault-tracing procedure that I have recommended (see "Fundamentals: Fault Tracing" on page 73). I will do my best to answer emails from readers who run into problems, but it's only fair for you to try to solve your problems first.

Writer–Reader Communication

There are three situations where you and I may want to communicate with each other.

- I may want to tell you if it turns out that the book contains a mistake which will prevent you from building a project successfully. I may also want to tell you if a parts kit, sold in association with the book, has something wrong with it. This is *me-informing-you* feedback.
- You may want to tell me if you think you found an error in the book, or in a parts kit. This is *you-informing-me* feedback.
- You may be having trouble making something work, and you don't know whether I made a mistake or you made a mistake. You would like some help. This is *you-asking-me* feedback.

I will explain how to deal with each of these situations.

Me Informing You

If you already registered with me in connection with *Make: More Electronics*, you don't need to register again for updates relating to *Make: Electronics*. But if you have not already registered, here's how it works.

I can't notify you if there's an error in the book or in a parts kit unless I have your contact information. Therefore I am asking you to send me your email address for the following purposes. Your email will not be used or abused for any other purpose.

- I will notify you if any significant errors are found in this book or in its successor, *Make: More Electronics*, and I will provide a workaround.
- I will notify you of any errors or problems relating to kits of components sold in association with this book or in *Make: More Electronics*.
- I will notify you if there is a completely new edition of this book, or of *Make: More Electronics*, or of my other books. These notifications will be very rare.

We've all seen registration cards that promise to enter you for a prize drawing. I'm going to offer you a much better deal. If you submit your email address, which may only be used for the three purposes listed above, I will send you an unpublished electronics project with complete construction plans as a two-page PDF. It will be fun, it will be unique, and it will be relatively easy. You won't be able to get this in any other way.

The reason I am encouraging you to participate is that if an error is found, and I have no way to tell you, and you discover it later on your own, you're likely to get annoyed. This will be bad for my reputation and the reputation of my work. It is very much in my interest to avoid a situation where you have a complaint.

- Simply send a blank email (or include some comments in it, if you like) to *make.electronics@gmail.com*. Please put REGISTER in the subject line.

You Informing Me

If you only want to notify me of an error that you have found, it's really better to use the "errata" system maintained by my publisher. The publisher uses the "errata" information to fix the error in updates of the book.

If you are sure that you found an error, please visit:

http://shop.oreilly.com/category/customer-service/faq-errata.do

The web page will tell you how to submit errata.

You Asking Me

My time is obviously limited, but if you attach a photograph of a project that doesn't work, I may have a suggestion. The photograph is essential.

You can use *make.electronics@gmail.com* for this purpose. Please put the word HELP in the subject line.

Going Public

There are dozens of forums online where you can discuss this book and mention any problems you are having, but please be aware of the power that you have as a reader, and use it fairly. A single negative review can create a bigger effect than you may realize. It can certainly outweigh half-a-dozen positive reviews.

The responses that I receive are generally very positive, but in a couple of cases people have been annoyed over small issues such as being unable to find a part online. I would have been happy to help these people if they had asked me.

I do read my reviews on Amazon about once each month, and will always provide a response if necessary.

Of course, if you simply don't like the way in which I have written this book, you should feel free to say so.

Going Further

After you work your way through this book, you will have grasped many of the basic principles in electronics. I like to think that if you want to know more, my own *Make: More Electronics* is the ideal next step. It is slightly more difficult, but uses the same "Learning by Discovery" method that I have used here. My intention is that you will end up with what I consider an "intermediate" understanding of electronics.

I am not qualified to write an "advanced" guide, and consequently I don't expect to create a third book with a title such as "Make: Even More Electronics."

If you want to know more electrical theory, *Practical Electronics for Inventors* by Paul Scherz is still the book that I recommend most often. You don't have to be an inventor to find it useful.

Safari® Books Online

Safari Books Online is an on-demand digital library that delivers expert content in both book and video form from the world's leading authors in technology and business.

Technology professionals, software developers, web designers, and business and creative professionals use Safari Books Online as their primary resource for research, problem solving, learning, and certification training.

Safari Books Online offers a range of plans and pricing for enterprise, government, education, and individuals.

Members have access to thousands of books, training videos, and prepublication manuscripts in one fully searchable database from publishers like O'Reilly Media, Prentice Hall Professional, Addison-Wesley Professional, Microsoft Press, Sams, Que, Peachpit Press, Focal Press, Cisco Press, John Wiley & Sons, Syngress, Morgan Kaufmann, IBM Redbooks, Packt, Adobe Press, FT Press, Apress, Manning, New Riders, McGraw-Hill, Jones & Bartlett, Course Technology, and hundreds more. For more information about Safari Books Online, please visit us online.

How to Contact Us

Please address comments and questions concerning this book to the publisher:

Make:
1160 Battery Street East, Suite 125
San Francisco, CA 94111
877-306-6253 (in the United States or Canada)
707-829-0515 (international or local)

Make: unites, inspires, informs, and entertains a growing community of resourceful people who undertake amazing projects in their backyards, basements, and garages. Make: celebrates your right to tweak, hack, and bend any technology to your will. The Make: audience continues to be a growing culture and community that believes in bettering ourselves, our environment, our educational system—our entire world. This is much more than an audience, it's a worldwide movement that Make: is leading—we call it the Maker Movement.

For more information about Make:, visit us online:

Make: magazine: *http://makezine.com/magazine/*
Maker Faire: *http://makerfaire.com*
Makezine.com: *http://makezine.com*
Maker Shed: *http://makershed.com/*

We have a web page for this book, where we list errata, examples, and any additional information. You can access this page at *http://bit.ly/make_elect_2e*.

To comment or ask technical questions about this book, send email to *bookquestions@oreilly.com*.

The Basics

1

Chapter One of this book contains Experiments 1 through 5.

In Experiment 1, I want you to get a taste for electricity—literally! You'll experience electric current and discover the nature of electrical resistance, not just in wires and components but in the world around you.

Experiments 2 through 5 will show you how to measure and understand the pressure and flow of electricity—and finally, how to generate electricity with everyday items on a tabletop.

Even if you have prior knowledge of electronics, I encourage you to try these experiments before venturing into subsequent parts of the book. They're fun, and they clarify some basic concepts.

Necessary Items for Chapter One

Each chapter of this book begins with pictures and descriptions of the tools, equipment, components, and supplies that will be required. After you have learned about them, you can flip to the back of the book where your buying options are summarized for quick reference.

- To buy tools and equipment, see "Buying Tools and Equipment" on page 324.
- For components, see "Components" on page 317.
- For supplies, see "Supplies" on page 316.
- If you prefer to get a prepackaged set of the components that you need, you have a choice of kits. See "Kits" on page 311 for more information.

I classify *tools and equipment* as items that should be useful indefinitely. They range from pliers to a multimeter. *Supplies*, such as wire and solder, will gradually be consumed in a variety of projects, but the quantities that I am recommending should be sufficient for all the experiments in the book. *Components* will be listed for individual projects, and will become part of those projects.

The Multimeter

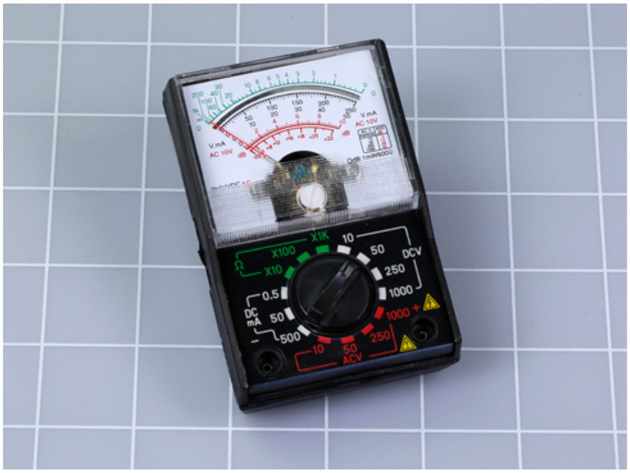

Figure 1-1 *This kind of analog meter is inadequate for your purposes. You need a digital meter.*

I'm beginning my instructional overview of tools and equipment with the multimeter, because I consider it the most essential piece of equipment. It will tell you how much voltage exists between any two points in a circuit, or how much current is passing through the circuit. It will help you to find a wiring error, and can also evaluate a component to determine its electrical resistance—or its capacitance, which is the ability to store an electrical charge.

If you're starting with little or no knowledge, these terms may seem confusing, and you may feel that a multimeter looks complicated and difficult to use. This is not the case. It makes the learning process easier, because it reveals what you cannot see.

Before I discuss which meter to buy, I can tell you what not to buy. You don't want an old-school meter with a needle that moves across a scale, as shown in Figure 1-1. That is an *analog* meter.

You want a *digital* meter that displays values numerically —and to give you an idea of the equipment available, I have selected four examples.

Figure 1-2 shows the cheapest digital meter that I could find, costing less than a paperback novel or a six-pack of soda. It cannot measure very high resistances or very low voltages, its accuracy is poor, and it does not measure capacitance at all. However, if your budget is very tight, it will probably see you through the experiments in this book.

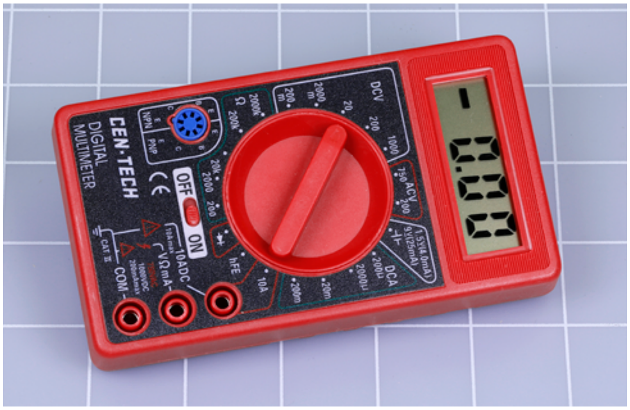

Figure 1-2 *The cheapest meter that I could find.*

The meter in Figure 1-3 offers more accuracy and more features. This meter, or one similar to it, is a good basic choice while you are learning electronics.

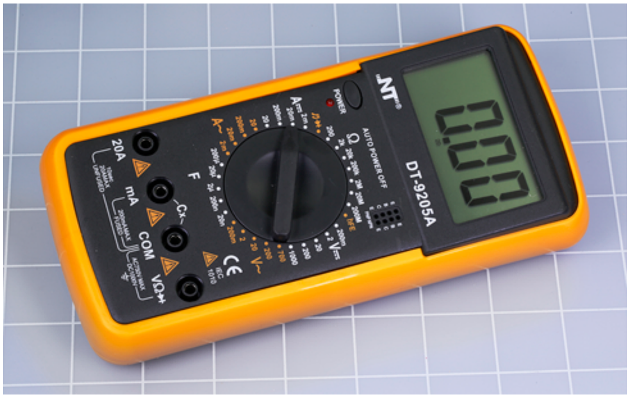

Figure 1-3 *Any meter similar to this one is a good basic choice.*

The example in Figure 1-4 is slightly more expensive but much better made. This particular model has been discontinued, but you can find many like it, probably costing two to three times as much as the NT brand in Figure 1-3. Extech is a well-established company trying to maintain its standards in the face of cut-price competitors.

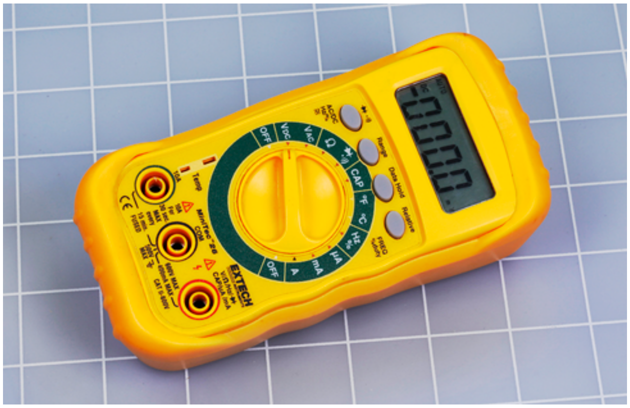

Figure 1-4 *A better-made meter at a somewhat higher price.*

Figure 1-5 shows my personal preferred meter at the time of writing. It is physically rugged, has all the features I could want, and measures a wide range of values with extremely good accuracy. However, it costs more than twenty times as much as the lowest-priced, bargain-basement product. I regard it as a long-term investment.

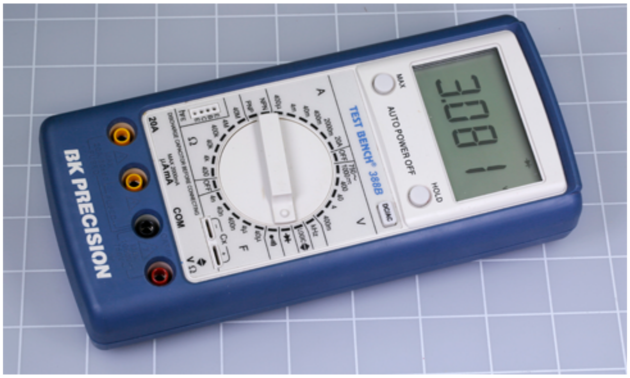

Figure 1-5 *A high-quality product.*

How do you decide which meter to buy? Well, if you were learning to drive, you wouldn't necessarily need a high-priced car. Similarly, you don't need a high-priced meter while you are learning electronics. On the other hand, the absolute cheapest meter may have some drawbacks, such as an internal fuse that is not easily replaceable, or a rotary switch with contacts that wear out quickly. So here's a rule of thumb if you want something that I would regard as inexpensive but acceptable:

- Search eBay for the absolute cheapest model you can find, then double the price, and use that as your guideline.

Regardless of how much you spend, the following attributes and capabilities are important.

Ranging

A meter can measure so many values, it has to have a way to narrow the range. Some meters have *manual ranging*, meaning that you turn a dial to choose a ballpark for the quantity that interests you. A range could be from 2 to 20 volts, for instance.

Other meters have *autoranging*, which is more convenient, because you just connect the meter and wait for it to figure everything out. The key word, however, is "wait." Every time you make a measurement with an autoranging meter, you will wait a couple of seconds while it performs an internal evaluation. Personally I tend to be impatient, so I prefer manual meters.

Another problem with autoranging is that because you have not selected a range yourself, you must pay attention to little letters in the display where the meter is telling you which units it has decided to use. For example, the difference between a "K" or an "M" when measuring electrical resistance is a factor of 1,000. This leads me to my personal recommendation:

- I suggest you use a manual-ranging meter for your initial adventures. You'll have fewer chances to make errors, and it should cost slightly less.

A vendor's description of a meter should say whether it uses manual ranging or autoranging, but if not, you can tell by looking at a photograph of its selector dial. If you don't see any numbers around the dial, it's an autoranging meter. The meter in Figure 1-4 does autoranging. The others that I pictured do not.

Values

The dial will also reveal what types of measurements are possible. At the very least, you should expect:

Volts, amps, and ohms, often abbreviated with the letter V, the letter A, and the ohm symbol, which is the Greek letter omega, shown in Figure 1-6. You may not know what these attributes mean right now, but they are fundamental.

Your meter should also be capable of measuring milliamps (abbreviated *mA*) and millivolts (abbreviated *mV*). This may not be immediately clear from the dial on the meter, but will be listed in its specification.

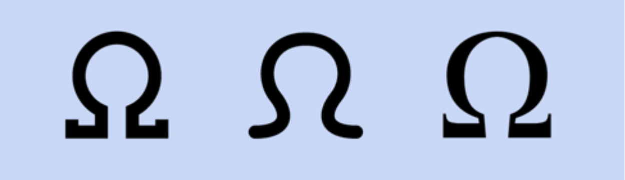

Figure 1-6 *Three samples of the Greek symbol omega, used to represent electrical resistance.*

DC/AC, meaning direct current and alternating current. These options may be selected with a DC/AC pushbutton, or they may be chosen on the main selector dial. A pushbutton is probably more convenient.

Continuity testing. This useful feature enables you to check for bad connections or breaks in an electrical circuit. Ideally it should create an audible alert, in which case it will be represented symbolically with a little dot that has semicircular lines radiating from it, as shown in Figure 1-7.

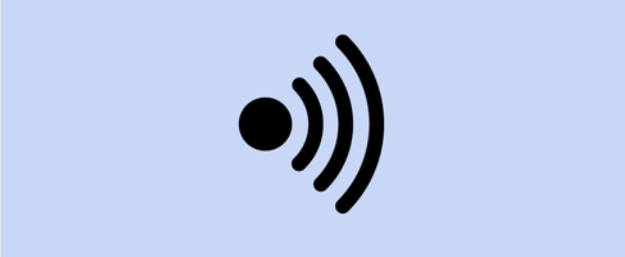

Figure 1-7 *This symbol indicates the option to test a circuit for continuity, with audible feedback. It's a very useful feature.*

For a small additional sum, you should be able to buy a meter that makes the following measurements. In order of importance:

Capacitance. Capacitors are small components that are needed in the majority of electronic circuits. Because small ones usually don't have their values printed on them, the ability to measure their values can be important, especially if some of them get mixed up or (worse) fall on the floor. Very cheap meters usually cannot measure capacitance. When the feature exists, it is usually indicated with a letter F, meaning farads, which are the units of measurement. The abbreviation CAP may also be used.

Transistor testing, indicated by little holes labeled E, B, C, and E. You plug the transistor into the holes. This enables you to verify which way up the transistor should be placed in a circuit, or if you have burned it out.

Frequency, abbreviated Hz. This is unimportant in the experiments in this book, but may be useful if you go further.

Any features beyond these are not significant.

If you still feel unsure about which meter to buy, read ahead a little to get an idea of how you will be using a meter in Experiments 1, 2, 3, and 4.

Safety Glasses

For Experiment 2, you may want to use safety glasses. The cheapest plastic type is satisfactory for this little adventure, as the risk of a battery bursting is almost nonexistent, and probably would not occur with much force.

Regular eyeglasses would be an acceptable substitute, or you could view the experiment through a little piece of transparent plastic (for instance, you can cut out a piece of a water bottle).

Batteries and Connectors

Because batteries and connectors become part of a circuit, I am categorizing them as components. See "Other Components" on page 319 for details about ordering these parts.

Almost all the experiments in this book will use a power source of 9 volts. You can obtain this from a basic nine-volt battery sold in supermarkets and convenience stores. Later I'll suggest an upgrade to an AC adapter, but you don't need that right now.

For Experiment 2, you will need a couple of 1.5-volt AA batteries. These have to be the alkaline type. You must not perform this experiment with any kind of rechargeable battery.

To transfer the power from a battery into a circuit, you need a connector for the 9-volt battery, as shown in Figure 1-8, and a carrier for a single AA battery, as shown in Figure 1-9. One carrier will be enough, but I suggest you get at least three 9-volt connectors for future use.

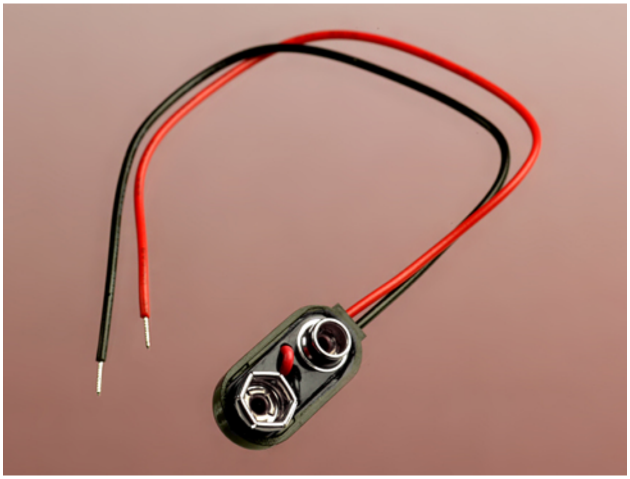

Figure 1-8 *Connector to deliver power from a 9-volt battery.*

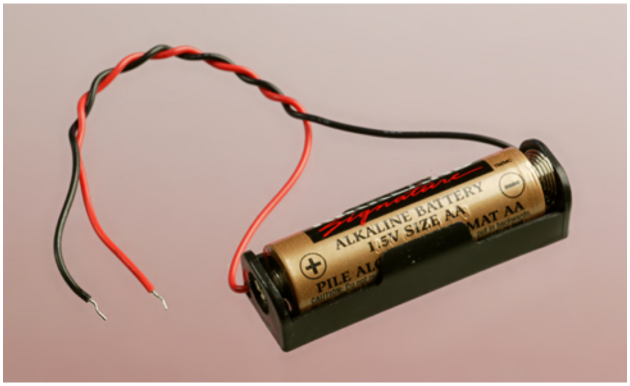

Figure 1-9 *You need a carrier like this for a single AA battery. Don't buy the type of carrier that holds two batteries (or three, or four).*

Test Leads

You will use test leads (pronounced "leeds") to connect components with each other in the first few experiments. The type of leads I mean are *double-ended*. Surely, any piece of wire has two ends, so why should it be called "double-ended"? The term usually means that each end is fitted with an *alligator clip* as shown in Figure 1-10. Each clip can make a connection by grabbing something and gripping it securely, freeing you to use your hands elsewhere.

You don't want the kind of test leads that have a plug at each end. Those are sometimes known as *jumper wires*.

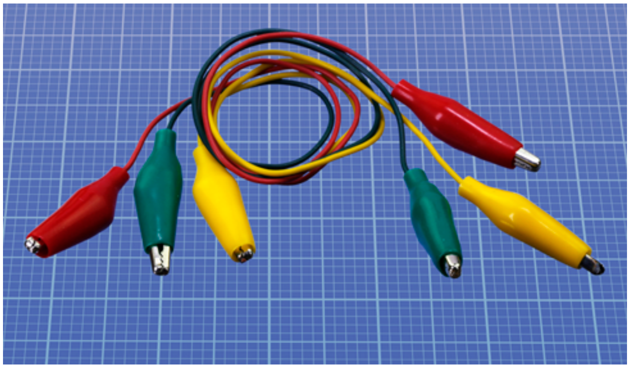

Figure 1-10 *Double-ended test leads with an alligator clip at each end.*

Test leads are classified as equipment for the purposes of this book. See "Buying Tools and Equipment" on page 324 for more information.

Potentiometer

A potentiometer functions like the volume control on an old-fashioned stereo. The kinds shown in Figure 1-11 are considered large by modern standards, but large is what you need, because you'll be gripping the terminals with the alligator clips on your test leads. A 1" diameter potentiometer is preferred. Its resistance should be listed as 1K. If you are buying your own, see "Other Components" on page 319 for details.

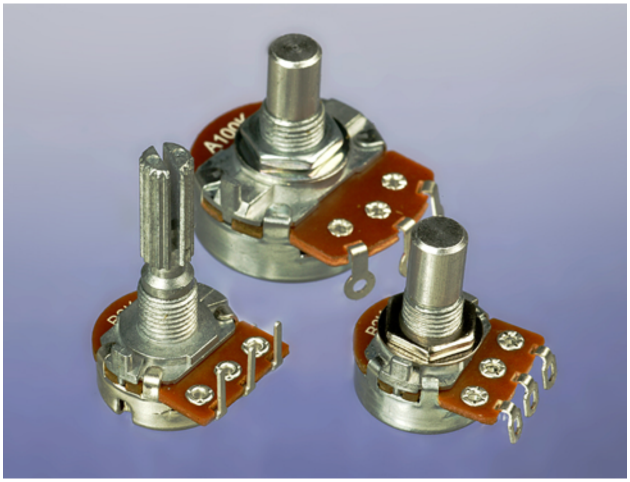

Figure 1-11 *Potentiometers of the general type required for your first experiments.*

Fuse

A fuse interrupts a circuit if too much electricity passes through it. Ideally you'll buy the type of 3-amp automotive fuse shown in Figure 1-12, which is easy to grip with test leads, and clearly reveals the element inside it. Automotive fuses are sold in a variety of physical sizes, but so long as you use one rated for 3 amps, the dimensions don't matter. Buy three to allow for destroying them intentionally or accidentally. If you don't want to use an auto parts supplier, a 2AG-size 3-amp glass cartridge fuse of the kind shown in Figure 1-13 will be available from electronics component suppliers, although it is not quite so easy to use.

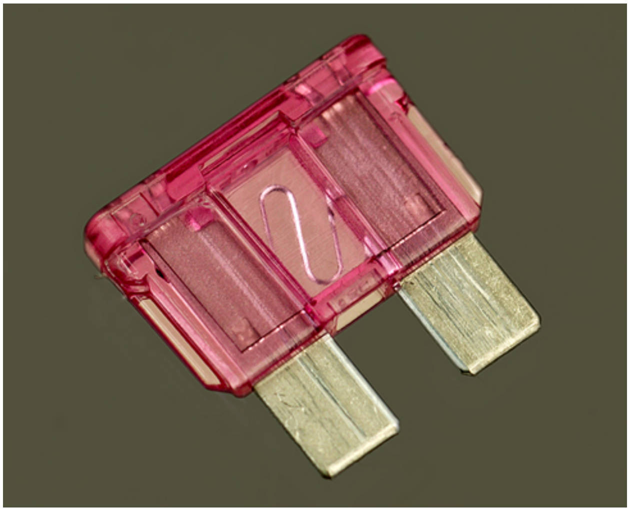

Figure 1-12 *This type of automotive fuse is easier to handle than the cartridge fuses used in electronics hardware.*

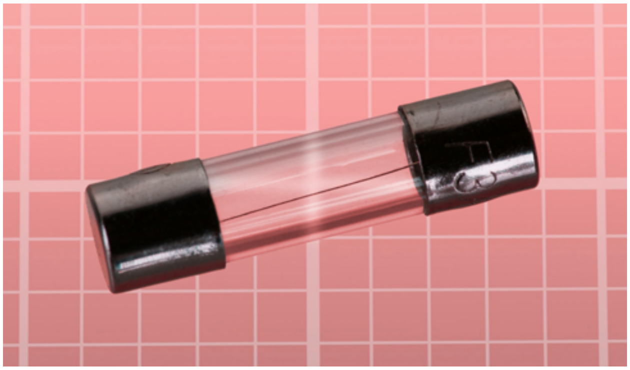

Figure 1-13 *You can use a cartridge fuse like this, although your alligator clips won't grip it so easily.*

Light-Emitting Diodes

More commonly known as *LEDs*, they come in various shapes and forms. The ones we will be using are properly known as *LED indicators*, and are often described as *standard through-hole LEDs* in catalogs. A sample in Figure 1-14 is 5mm in diameter, but 3mm is sometimes easier to fit into a circuit when space is limited. Either will do.

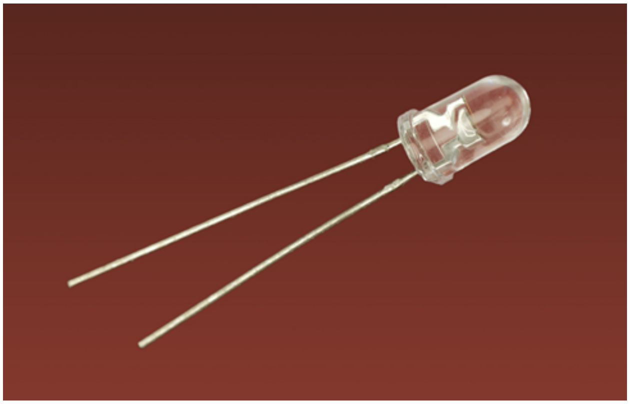

Figure 1-14 *A light-emitting diode (LED) approximately 5mm in diameter.*

Throughout this book I will refer to *generic LEDs*, by which I mean the cheapest ones that don't emit a high-intensity light and are commonly available in red, yellow, or green. They are often sold in bulk quantities, and are used in so many applications that I suggest you buy at least a dozen of each color.

Some generic LEDs are encapsulated in "water clear" plastic or resin, but emit a color when power is applied. Other LEDs are encapsulated in plastic or resin tinted with the same color that they will display. Either type is acceptable.

In a few experiments, *low-current LEDs* are preferred. They cost slightly more, but are more sensitive. For example, in Experiment 5, where you will generate a small amount of current with an improvised battery, you'll get better results with a low-current LED. See "Other Components" on page 319 for additional guidance, if you are not using components that were supplied in a kit.

Resistors

You'll need a variety of resistors to restrict the voltage and current in various parts of a circuit. They should look something like the ones in Figure 1-15. The color of the body of the resistor doesn't matter. Later I will explain how the colored stripes tell you the value of the component.

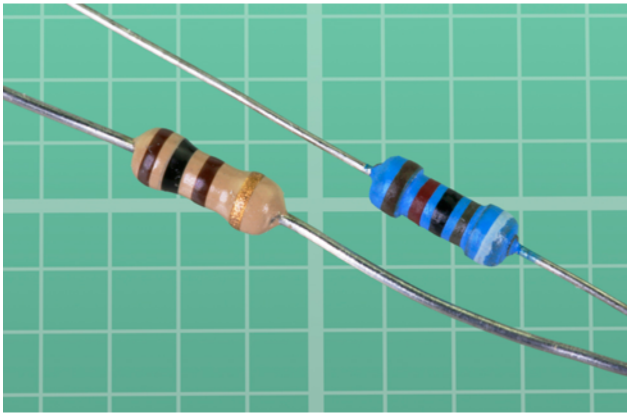

Figure 1-15 *Two resistors of the type you need, all rated for 1/4 watt.*

If you are buying your own resistors, they are so small and cheap, you would be foolish to select just the values listed in each experiment. Get a prepackaged selection in bulk from surplus or discount sources, or eBay. For more information about resistors, including a complete list of all the values used throughout this book, see "Components" on page 317.

You don't need any other components to take you through Experiments 1 through 5. So let's get started!

Experiment 1: Taste the Power!

Can you taste electricity? It feels as if you can.

What You Will Need

- 9-volt battery (1)
- Multimeter (1)

That's all!

Caution: No More than Nine Volts

This experiment should only use a 9-volt battery. *Do not* try it with a higher voltage, and *do not* use a bigger battery that can deliver more current. Also, if you have metal braces on your teeth, be careful not to touch them with the battery. Most important, never apply electric current from any size of battery through a break in your skin.

Procedure

Moisten your tongue and touch the tip of it to the metal terminals of a 9-volt battery, as shown in Figure 1-16. (Maybe your tongue isn't quite as big as the one in the picture. Mine certainly isn't. But this experiment will work regardless of how big or small your tongue may be.)

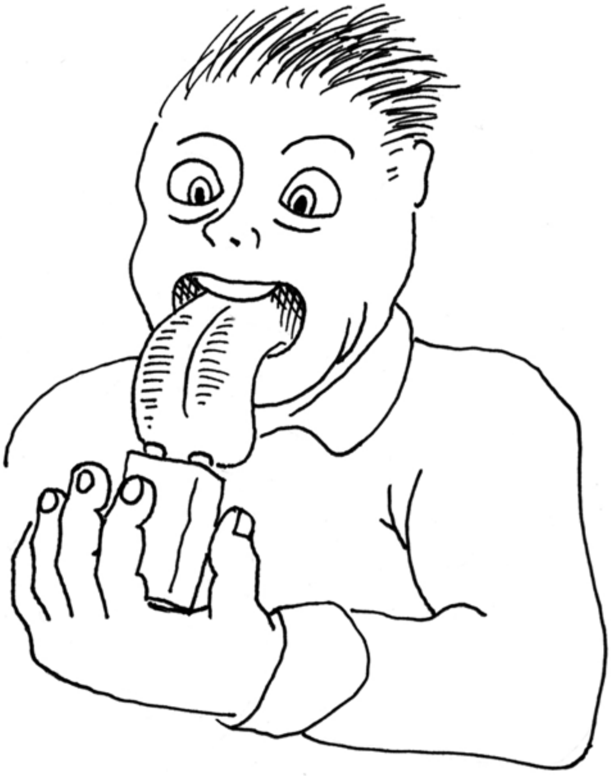

Figure 1-16 *An intrepid Maker tests the characteristics of an alkaline battery.*

Do you feel that tingle? Now set aside the battery, stick out your tongue, and dry the tip of it very thoroughly with a tissue. Touch the battery to your tongue again, and you should feel less of a tingle.

What's happening here? You can use a meter to find out.

Setting Up Your Meter

Does your meter have a battery preinstalled? Select any function with the dial, and wait to see if the display shows a number. If nothing is visible, you may have to open the meter and put in a battery before you can use it. To find out how to do this, check the instructions that came with the meter.

Meters are supplied with a red lead and a black lead. Each lead has a plug on one end, and a steel probe on the other end. You insert the plugs into the meter, then touch the probes on locations where you want to know what's going on. See Figure 1-17. The probes detect electricity; they don't emit it in significant quantities. When you are dealing with the small currents and voltages in the experiments in this book, the probes cannot hurt you (unless you poke yourself with their sharp ends).

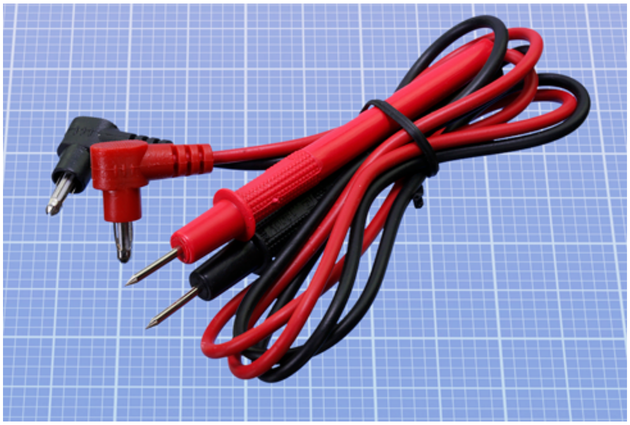

Figure 1-17 *Leads for a meter, terminating in metal probes.*

Most meters have three sockets, but some have four. See Figure 1-18, Figure 1-19, and Figure 1-20 for examples. Here are the general rules:

- One socket should be labelled COM. This is *common* to all your measurements. Plug the black lead into this socket, and leave it there.
- Another socket should be identified with the ohm (omega) symbol, and the letter V for volts. It can measure either resistance or voltage. Plug the red lead into this socket.
- The voltage/ohms socket may also be used for measuring small currents in mA (milliamps) . . . or you may see a separate socket for this, which will require you to move the red lead sometimes. We'll get to that later.
- An additional socket may be labelled 2A, 5A, 10A, 20A, or something similar, to indicate a maximum number of amps. This is used for measuring high currents. We won't be needing it for projects in this book.

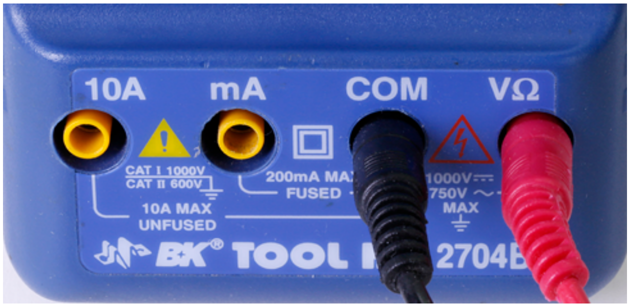
10A
mA
COM
VΩ
CAT I 1000V
CAT II 600V
10A MAX
UNFUSED
200mA MA
FUSED
1000V⎓
750V ~
MAX
BK TOOL
2704B

Figure 1-18 *Note the labeling of sockets on this meter.*

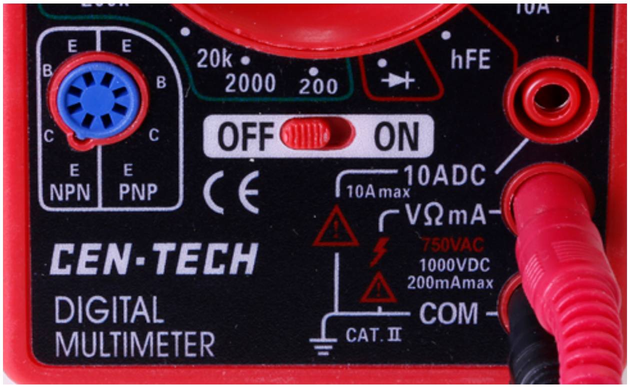
E
E
B
B
C
C
E
E
NPN
PNP
20k
2000
200
hFE
OFF
ON
CE
10ADC
10A max
VΩmA
750VAC
1000VDC
200mAmax
COM
CAT. II
CEN-TECH
DIGITAL
MULTIMETER

Figure 1-19 *Socket functions are split up differently on this meter.*

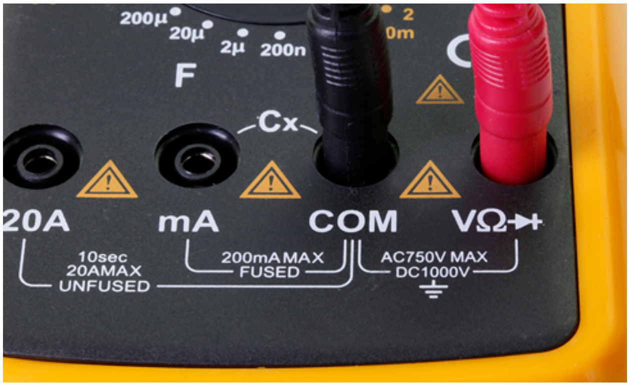
200µ
20µ
2µ
200n
F
Cx
20A
mA
COM
VΩ
10sec
20AMAX
UNFUSED
200mA MAX
FUSED
AC750V MAX
DC1000V

Figure 1-20 *Sockets on one more meter.*

Fundamentals: Ohms

You're going to evaluate the resistance of your tongue, in ohms. But what is an ohm?

We measure distance in miles or kilometers, mass in pounds or kilograms, and temperature in Fahrenheit or Centigrade. We measure electrical resistance in ohms, which is an international unit named after Georg Simon Ohm, who was an electrical pioneer.

The Greek omega symbol indicates ohms, but for resistances above 999 ohms the uppercase letter K is used, which means *kilohm*, equivalent to a thousand ohms. For example, a resistance of 1,500 ohms will be referred to as 1.5K.

Above 999,999 ohms, the uppercase letter M is used, meaning *megohm*, which is a million ohms. In everyday speech, a megohm is often referred to as a "meg." If someone is using a "two-point-two meg resistor," its value will be 2.2M.

A conversion table for ohms, kilohms, and megohms is shown in Figure 1-21.

Ohms	Kilohms	Megohms
1 Ω	0.001K	0.000001M
10 Ω	0.01K	0.00001M
100 Ω	0.1K	0.0001M
1,000 Ω	1K	0.001M
10,000 Ω	10K	0.01M
100,000 Ω	100K	0.1M
1,000,000 Ω	1,000K	1M

Figure 1-21 *Conversion table for the most common multiples of ohms.*

- In Europe, the letter R, K, or M is substituted for a decimal point, to reduce the risk of errors. Thus, 5K6 in a European circuit diagram means 5.6K, 6M8 means 6.8M, and 6R8 means 6.8 ohms. I won't be using the European style here, but you may find it in some circuit diagrams elsewhere.

A material that has very high resistance to electricity is known as an *insulator*. Most plastics, including the colored sheaths around wires, are insulators.

A material with very low resistance is a *conductor*. Metals such as copper, aluminum, silver, and gold are excellent conductors.

Measuring Your Tongue

Inspect the dial on the front of your meter, and you'll find at least one position identified with the ohm symbol. On an autoranging meter, turn the dial to point to the ohm symbol as shown in Figure 1-22, touch the probes *gently* to your tongue, and wait for the meter to choose a range automatically. Watch for the letter K in the numeric display. Never stick the probes *into* your tongue!

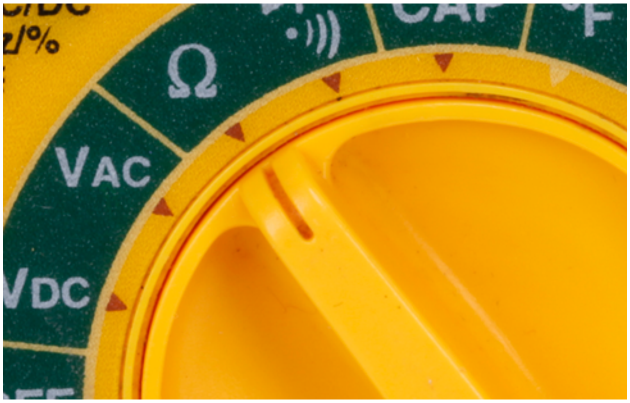

Figure 1-22 *On an autoranging meter, just turn the dial to the ohm (omega) symbol.*

On a manual meter, you must choose a range of values. For a tongue measurement, probably 200K (200,000 ohms) would be about right. Note that the numbers beside the dial are maximums, so 200K means "no more than 200,000 ohms" while 20K means "no more than 20,000 ohms." See the close-ups of the manual meters in Figure 1-23 and Figure 1-24.

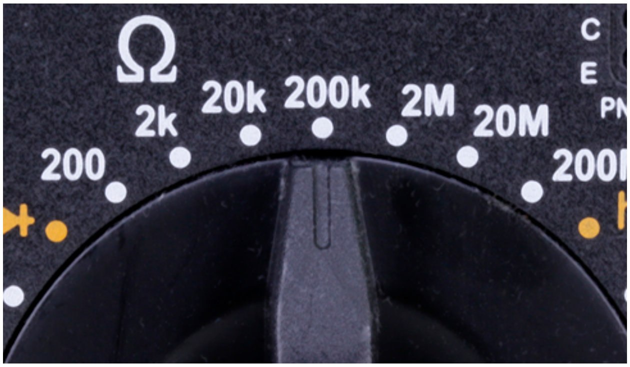

Figure 1-23 *A manual meter requires you to select the range.*

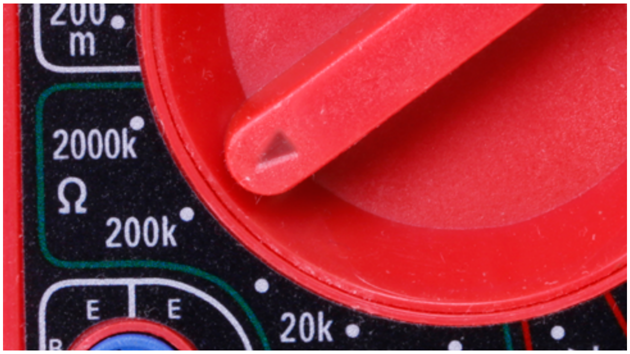

Figure 1-24 *A different manual meter dial, but the principle is the same.*

Touch the probes to your tongue about one inch apart. Note the meter reading, which should be around 50K. Put aside the probes, stick out your tongue, and use a tissue to dry it carefully and thoroughly, as you did before. Without allowing your tongue to become moist again, repeat the test, and the reading should be higher. Using a manual ranging meter, you may have to select a higher range to see a resistance value.

- When your skin is moist (for instance, if you perspire), its electrical resistance decreases. This principle is used in lie detectors, because someone who knowingly tells a lie, under conditions of stress, may tend to perspire.

Here's the conclusion that your test may suggest. A lower resistance allows more electric current to flow, and in your initial tongue test, more current created a bigger tingle.

Fundamentals: Inside a Battery

When you used a battery for the original tongue test, I didn't bother to mention how a battery works. Now is the time to rectify that omission.

A 9-volt battery contains chemicals that liberate *electrons* (particles of electricity), which want to flow from one terminal to the other as a result of a chemical reaction. Think of the cells inside a battery as being like two water tanks—one of them full, the other empty. If the tanks are connected with each other by a pipe and a valve, and you open the valve, water will flow between them until their levels are equal. Figure 1-25 may help you to visualize this. Similarly, when you open up an electrical pathway between the two sides of a battery, electrons flow between them, even if the pathway consists only of the moisture on your tongue.

Electrons flow more easily through some substances (such as a moist tongue) than others (such as a dry tongue).

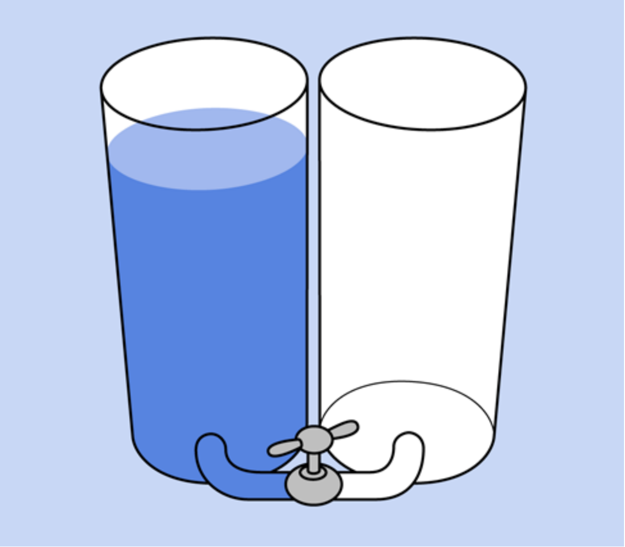

Figure 1-25 *You can think of a battery as being like a pair of interconnected water reservoirs.*

Further Investigation

The tongue test was not a very well-controlled experiment, because the distance between the probes might have varied a little between one trial and the next. Do you think that may be significant? Let's find out.

Hold the meter probes so that their tips are only 1/4″ apart. Touch them to your moist tongue. Now separate the probes by 1″ and try again. What readings do you get?

When electricity travels through a shorter distance, it encounters less resistance. As a result, the current will increase.

Try a similar experiment on your arm, as shown in Figure 1-26. You can vary the distance between the probes in fixed steps, such as 1/4″, and note the resistance shown by your meter. Do you think that doubling the distance between the probes doubles the resistance shown by the meter? How can you prove or disprove this?

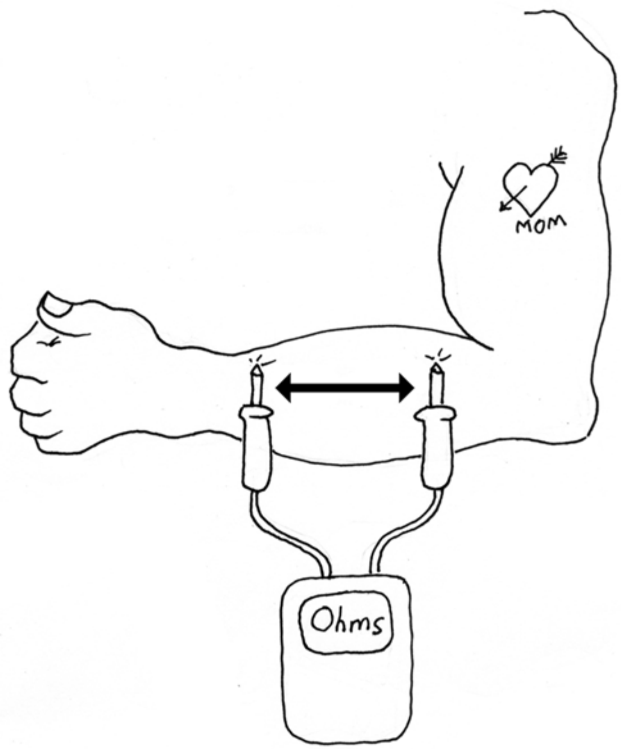

Figure 1-26 *Vary the distance between the probes, and note the reading on your meter.*

If the resistance is too high for your meter to measure, you will see an error message, such as L, instead of some numbers. Try moistening your skin, then repeat the test, and you should get a result. The only problem is, as the moisture on your skin evaporates, the resistance will change. You see how difficult it is to control all the factors in an experiment. The random factors are properly known as *uncontrolled variables*.

There is still one more variable that I haven't discussed, which is the amount of pressure between each probe and the skin. If you press harder, I suspect that the resistance will diminish. Can you prove this? How could you design an experiment to eliminate this variable?

If you're tired of measuring skin resistance, you can try dunking the probes into a glass of water. Then dissolve some salt in the water, and test it again. No doubt you've heard that water conducts electricity, but the full story is not so simple. Impurities in water play an important part.

What do you think will happen if you try to measure the resistance of water that contains no impurities at all? Your first step would be to obtain some pure water. So-called *purified water* usually has minerals added after it was purified, so, that's not what you want. Similarly, *spring water* is not totally pure. What you need is *distilled water*, also known as *deionized water*. This is often sold in supermarkets. I think you'll find that its resistance, per inch between the meter probes, is higher than the resistance of your tongue. Try it to find out.

Those are all the experiments relating to resistance that I can think of, right now. But I still have a little background information for you.

Background: The Man Who Discovered Resistance

Georg Simon Ohm, pictured in Figure 1-27, was born in Bavaria in 1787 and worked in obscurity for much of his life, studying the nature of electricity using metal wire that he had to make for himself (you couldn't truck on down to Home Depot for a spool of hookup wire back in the early 1800s).

Despite his limited resources and inadequate mathematical abilities, Ohm was able to demonstrate in 1827 that the electrical resistance of a conductor such as copper varied in inverse proportion with its area of cross-section, and the current flowing through it is proportional to the voltage applied to it, so long as temperature is held constant. Fourteen years later, the Royal Society in London finally recognized the significance of his contribution and awarded him the Copley Medal. Today, his discovery is known as Ohm's Law. I'll explain more about this in Experiment 4.

Figure 1-27 *Georg Simon Ohm, after being honored for his pioneering work, most of which he pursued in relative obscurity.*

Cleanup and Recycling

Your battery should not have been damaged or significantly discharged by this experiment. You can use it again.

Remember to switch off your meter before putting it away. Many meters will beep to remind you to switch them off if you don't use them for a while, but some don't. A meter consumes a very small amount of electricity while it is switched on, even when you are not using it to measure anything.

Experiment 2: Let's Abuse a Battery!

To get a better feeling for electrical power, you're going to do what most books tell you not to do. You're going to short out a battery. (A *short circuit* is a shortcut between two sides of a power source.)

Caution: Use a Small Battery

The experiment that I am going to describe is safe, but some short circuits can be dangerous. Never short out a power outlet in your home: there'll be a loud bang, a bright flash, and the wire or tool that you use will be partially melted, while flying particles of melted metal can burn you or blind you.

If you short out a car battery, the flow of current is so huge that the battery might even explode, drenching you in acid. Just ask the guy in Figure 1-28 (if he is able to answer).

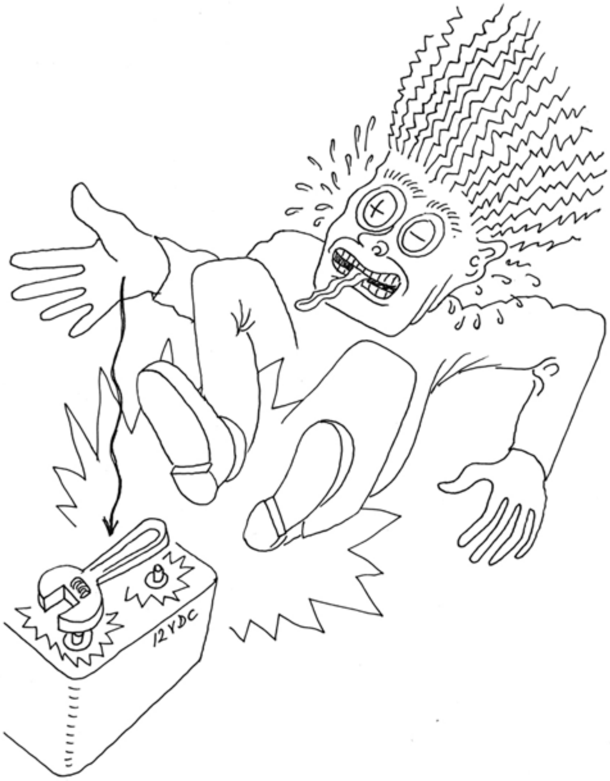

Figure 1-28 *Dropping a wrench between the terminals of a car battery will be bad for your health. Short circuits can be dramatic, even at a "mere" 12 volts, if the battery is big enough.*

Lithium batteries are often found in power tools, laptop computers, and other portable devices. Never short-circuit a lithium battery: it can catch fire and burn you. Lithium batteries have been known to catch fire even if you don't short them out, as shown in Figure 1-29. After some early laptops self-destructed, lithium battery packs were modified to prevent this kind of thing. But short-circuiting them is still a very bad idea.

Figure 1-29 *Never fool around with lithium batteries.*

Use only an alkaline battery in this experiment, and only a single AA cell. You may also want to wear safety glasses in case you happen to have a defective battery.

What You Will Need

- 1.5-volt AA battery (2)
- Battery carrier (1)
- 3-amp fuse (2)
- Safety glasses (regular eyeglasses or sunglasses will do)
- Test leads with alligator clips at each end (2)

Generating Heat with Current

Use an alkaline battery. Do not use any kind of rechargeable battery.

Put the battery into a battery carrier that terminates in two thin wires, as shown in Figure 1-9. Twist the bare ends of the wires together, as shown in Figure 1-30. At first it seems that nothing happens. But wait one minute, and you'll find that the wires are getting hot. Wait another minute, and the battery, too, will be hot.

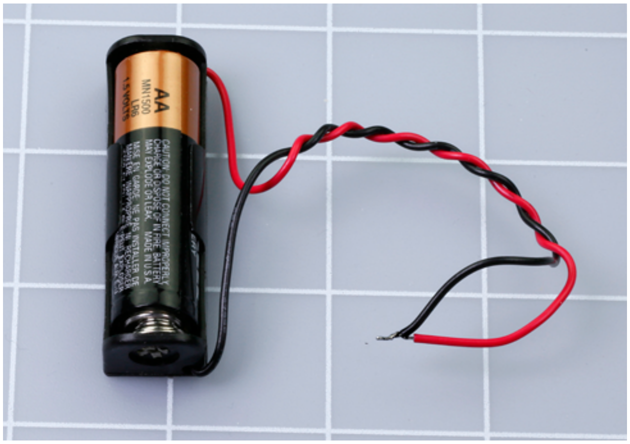

Figure 1-30 *Shorting out an alkaline battery can be safe if you follow the directions precisely.*

The heat is caused by electricity flowing through the wires and through the *electrolyte* (the conductive fluid) inside the battery. If you've ever used a hand pump to force air into a bicycle tire, you know that the pump gets warm. Electricity behaves in much the same way. You can imagine the electricity being composed of particles (electrons) that make the wire hot as they push through it. This isn't a perfect analogy, but it's close enough for our purposes.

Where do the electrons come from? Chemical reactions inside the battery liberate them, creating electrical pressure. The correct name for this pressure is *voltage*, which is measured in volts and is named after Alessandro Volta, another electrical pioneer.

Going back to the water analogy: the height of the water in a tank is proportional with the pressure of the water, and is similar to voltage. Figure 1-31 may help you to visualize this.

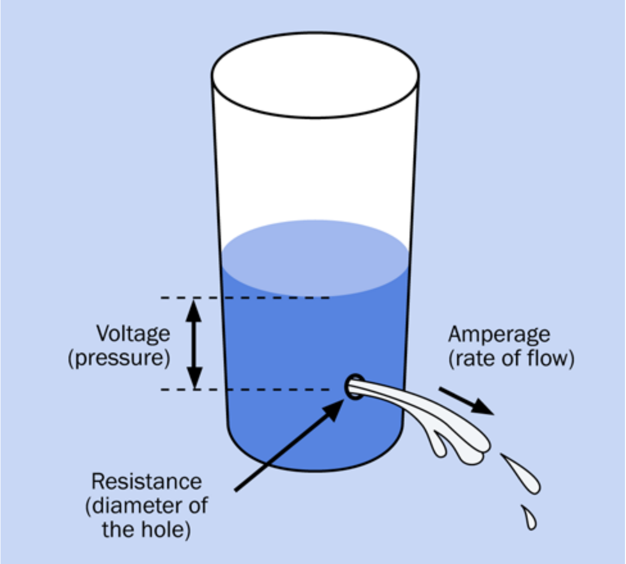

Figure 1-31 *The pressure in a source of water is analogous to the voltage in a source of electricity.*

But volts are only half of the story. When electrons flow through a wire, the amount of flow during a period of time is known as *amperage*, named after yet another electrical pioneer, André-Marie Ampère. That flow is also generally known as *current*. The current—the amperage —generates the heat.

- Think of voltage as pressure.
- Think of amperes as the rate of flow, properly known as current.

Background: Why Didn't Your Tongue Get Hot?

When you touched the 9-volt battery to your tongue, you felt a tingle, but no perceptible heat. When you shorted out a battery, you generated a noticeable amount of heat, even though you only used a 1.5-volt battery. How can you explain this?

Your meter showed you that the electrical resistance of your tongue is very high. This high resistance reduced the flow of electrons.

The resistance of a wire is very low, so if there's a wire connecting the two terminals of the battery, more current will pass through it than passed through your tongue, and it will create more heat. If all other factors remain constant:

- Lower resistance allows more current to flow.
- The heat generated by electricity is proportional with the amount of electricity (the current) passing through a conductor in a period of time. (This relationship is no longer exactly true if the resistance of the wire changes as the wire gets hot. But it remains approximately true.)

Here are some other basic concepts:

- The flow of electricity per second is measured in amperes, very commonly abbreviated as *amps*.
- The pressure of electricity that causes the flow is measured in *volts*.
- The resistance to the flow is measured in *ohms*.
- A higher resistance restricts the current.
- A higher voltage is better able to overcome resistance and increase the current.

The relationship between voltage, resistance, and amperage (pressure, resistance, and flow) is illustrated in Figure 1-32.

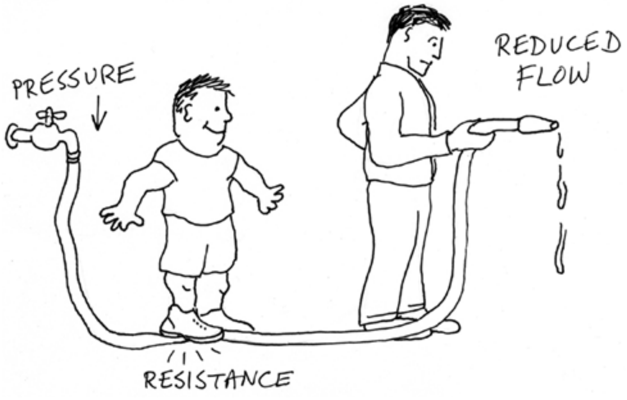

Figure 1-32 *Resistance blocks pressure and reduces flow—in water, and in electricity.*

Fundamentals: Volt Basics

The volt is an international unit, represented by an uppercase letter V. In the United States and much of Europe, AC power for domestic use is supplied at 110V, 115V, or 120V, with separate circuits for heavy-duty appliances delivering 220V, 230V, or 240V. Solid-state electronic components traditionally required DC power ranging from 5V to about 20V, although modern surface-mount devices may use less than 2V. Some components, such as a microphone, deliver voltage measured in millivolts, abbreviated mV, each millivolt being 1/1,000th of a volt. When electricity is distributed over long distances, it is measured in kilovolts, abbreviated kV. A few long-distance lines use megavolts. A conversion table for millivolts, volts, and kilovolts is shown in Figure 1-33.

Millivolts	Volts	Kilovolts
1mV	0.001V	0.000001kV
10mV	0.01V	0.00001kV
100mV	0.1V	0.0001kV
1,000mV	1V	0.001kV
10,000mV	10V	0.01kV
100,000mV	100V	0.1kV
1,000,000mV	1,000V	1kV

Figure 1-33 *Conversion table for the most common multiples of volts.*

Fundamentals: Ampere Basics

The ampere is an international unit, represented by an uppercase letter A. Household appliances may draw several amps, and a typical circuit breaker in the United States is rated for 20A. Electronic components often are rated in milliamps, abbreviated mA, each milliamp being 1/1,000th of an amp. Devices such as liquid-crystal displays may draw microamps, abbreviated μA, each microamp being 1/1,1000th of a milliamp. A conversion table for amps, milliamps, and microamps is shown in Figure 1-34.

Microamps	Milliamps	Amps
1μA	0.001mA	0.000001A
10μA	0.01mA	0.00001A
100μA	0.1mA	0.0001A
1,000μA	1mA	0.001A
10,000μA	10mA	0.01A
100,000μA	100mA	0.1A
1,000,000μA	1,000mA	1A

Figure 1-34 *Conversion table for the most common multiples of amps.*

How to Blow a Fuse

Exactly how much current flowed through the wires of the battery carrier, when you shorted out the battery? Could we have measured it?

Not easily, because if you try to use your multimeter to measure high current, you may blow the fuse inside the meter. So set aside the meter, and we'll use your 3-amp fuse, which we can sacrifice because it didn't cost very much.

First inspect the fuse, using a magnifying glass, if you have one. In an automotive fuse, you should see a tiny S-shape in the transparent window at the center. That S is a thin section of metal that melts easily. You saw it in Figure 1-12. In a glass cartridge fuse, there's a thin piece of wire that serves the same purpose.

Remove the 1.5-volt battery from its carrier. The battery is no longer useful for anything, and should be recycled if possible. Separate the two wires that you twisted together, and use two test leads to connect the carrier with a fuse, as shown in Figure 1-35 or Figure 1-36. Watch the fuse as you insert a new battery in the carrier. A break should occur in the center of the fuse element, where the metal melted. Figure 1-37 and Figure 1-38 show what I mean.

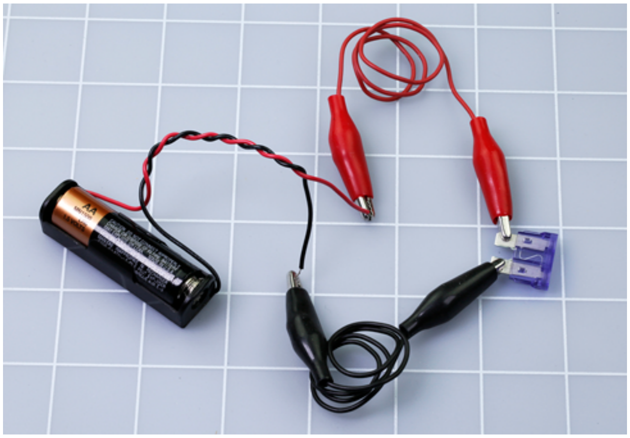

Figure 1-35 *How to short out an automotive fuse.*

Some 3-amp fuses blow more easily than others, even though they all have the same rating. If you think your fuse has not been affected, try applying the wires from the battery to it directly, instead of passing the current through the test leads. If you are not using a fresh AA battery, you may have to wait a few seconds for the fuse to respond. If all else fails, you can apply a C cell or a D cell, which have the same voltage as an AA battery but can deliver more current. But this shouldn't be necessary.

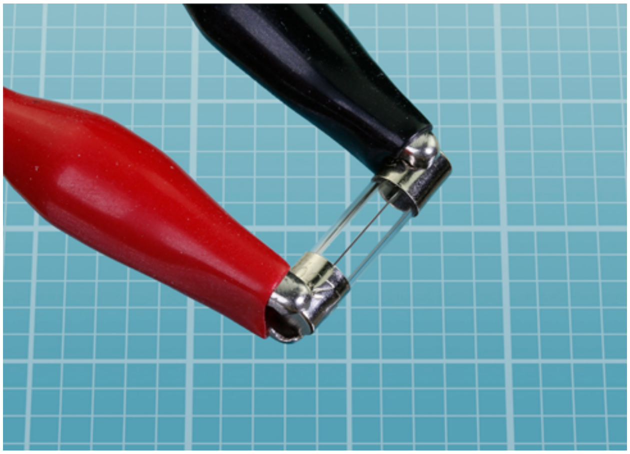

Figure 1-36 *How to apply your test leads to a small cartridge fuse.*

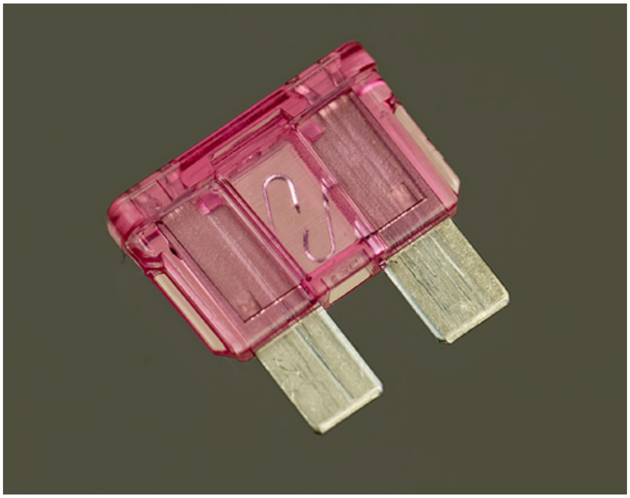

Figure 1-37 *Note the break in the element.*

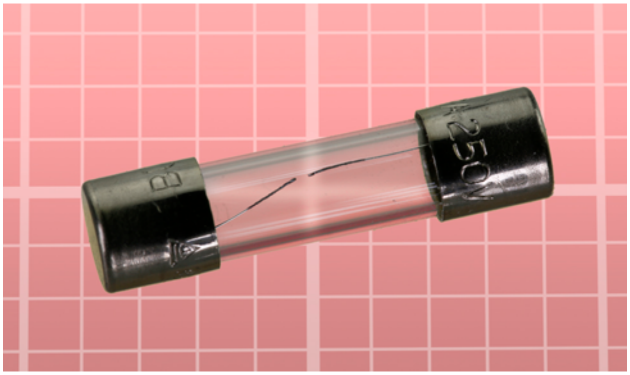

Figure 1-38 *In a shorted cartridge fuse, a similar break appears.*

This is how a fuse works: it melts to protect the rest of the circuit. That tiny break inside the fuse stops any more current from flowing.

Fundamentals: Direct and Alternating Current

The flow of current that you get from a battery is known as *direct current*, or DC. Like the flow of water from a faucet, it is a steady stream that flows in one direction.

The flow of current that you get from the power outlet in your home is very different. The "live" side of the outlet changes from positive to negative, relative to the "neutral" side, at a rate of 60 times each second (in many foreign countries, including Europe, 50 times per second). This is known as *alternating current*, or AC, which is more like the pulsing flow you get when you use a power washer to wash a car.

Alternating current is essential for some purposes, such as cranking up voltage so that electricity can be distributed over long distances. AC is also useful in motors and domestic appliances. The parts of a power outlet are shown in Figure 1-39. This style of outlet is found in North America, South America, Japan, and some other nations. European outlets look different, but the principle remains the same.

In the figure, socket A is the "hot" or "live" side of the outlet, supplying voltage that alternates between positive and negative relative to socket B, which is the "neutral" side. If an appliance develops a fault such as an internal loose wire, it should protect you by sinking the voltage through socket C, the ground.

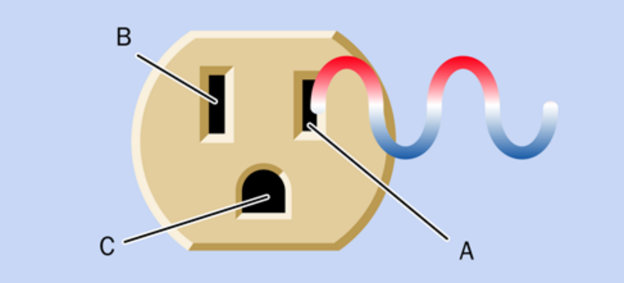

Figure 1-39 *The parts of a power outlet.*

In the United States, the outlet shown in the diagram is rated for 110 to 120 volts. Other configurations of outlet are used for higher voltages, but they still have live, neutral, and ground wires—with the exception of three-phase outlets, which are used primarily in industry.

For most of this book I'm going to be talking about DC, for two reasons: first, most simple electronic circuits are powered with DC, and second, the way it behaves is much easier to understand.

- I won't bother to mention repeatedly that I'm dealing with DC. Just assume that everything is DC unless otherwise noted.

Background: Inventor of the Battery

Alessandro Volta, shown in Figure 1-40, was born in Italy in 1745, long before science was broken up into specialties. After studying chemistry (he discovered methane in 1776) he became a professor of physics and developed an interest in the so-called galvanic response, whereby a frog's leg will twitch in response to a jolt of static electricity.

Using a wine glass full of salt water, Volta demonstrated that the chemical reaction between two electrodes (one made of copper, the other of zinc) will generate a steady electric current. In 1800, he refined his apparatus by stacking plates of copper and zinc, separated by cardboard soaked in salt and water. This "voltaic pile" was the first electric battery in Western civilization.

Figure 1-40 *Alessandro Volta discovered that chemical reactions can create electricity.*

Background: Father of Electromagnetism

Born in 1775 in France, André-Marie Ampère (shown in Figure 1-41) was a mathematical prodigy who became a science teacher, despite being largely self-educated in his father's library. His best-known work was to derive a theory of electromagnetism in 1820, describing the way that an electric current generates a magnetic field. He also built the first instrument to measure the flow of electricity (now known as a *galvanometer*), and discovered the element fluorine.

Figure 1-41 *André-Marie Ampère found that an electric current running through a wire creates a magnetic field around it. He used this principle to make the first reliable measurements of what came to be known as amperage.*

Cleanup and Recycling

You can dispose of the battery that you damaged when you shorted it. Putting batteries in the trash is not a great idea, because they contain heavy metals that should be kept out of the ecosystem. Your state or town may include batteries in a local recycling scheme. (California requires that almost all batteries be recycled.) Check your local regulations for details.

The blown fuse is of no further use, and can be thrown away.

The second battery, which was protected by the fuse, should still be OK.

The battery carrier can be reused later.

Experiment 3: Your First Circuit

Now it's time to make electricity do something more useful. To achieve this, you'll be experimenting with components known as resistors, and a light-emitting diode, or LED.

What You Will Need

- 9-volt battery (1)
- Resistors: 470 ohm (1), 1K (1), 2.2K (1)
- Generic LED (1)
- Test leads with alligator clips at each end (3)
- Multimeter (1)

Setup

It's time to get acquainted with the most fundamental component we'll be using in electronic circuits: the humble resistor. As its name implies, it resists the flow of electricity. As you might expect, its value is measured in ohms.

If you bought a bargain-basement assortment of resistors, they may be delivered to you in unmarked bags. No problem; we can determine their values easily enough. In fact, even if the packages are clearly labeled, I want you to check the resistors as we go along, because it's easy to get them mixed up. You have two alternatives:

- Use your multimeter, after setting it to measure ohms.
- Learn the color codes that are printed on most resistors. I'll explain them immediately below.

After you check the values of resistors, it's a good idea to sort them into labeled compartments in little plastic parts boxes. Personally, I like the boxes sold at the Michael's chain of crafts stores in the United States, but there are many alternatives. You can also use miniature plastic bags, which you can find by searching eBay for:

`plastic bags parts`

Fundamentals: Decoding Resistors

Some resistors have a value clearly stated on them in microscopic print that you can read with a magnifying glass, as shown in Figure 1-42.

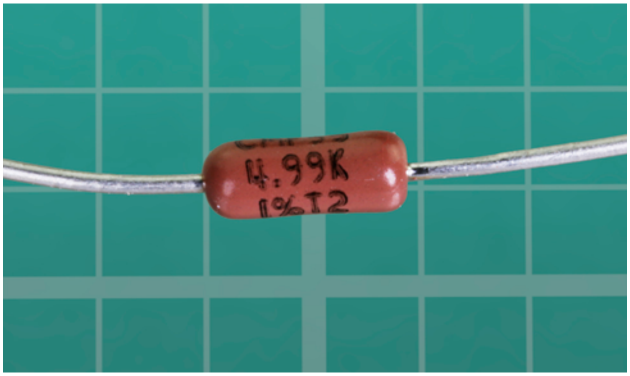

Figure 1-42 *Only a minority of resistors have their values printed on them.*

However, most resistors are color-coded with stripes. Figure 1-43 shows the coding scheme.

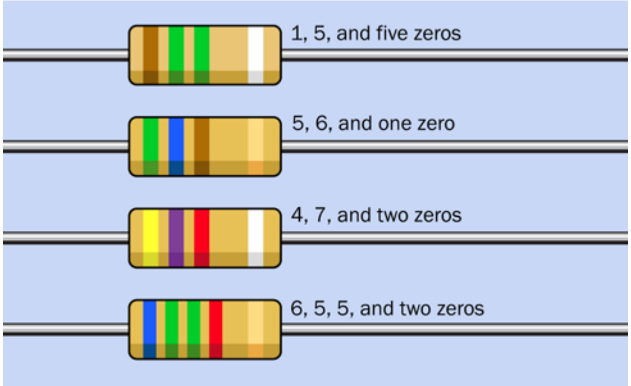
1, 5, and five zeros
5, 6, and one zero
4, 7, and two zeros
6, 5, 5, and two zeros

Figure 1-43 *The coding scheme that shows the value of a resistor. Some resistors have four stripes on the left instead of three, as explained in the text.*

Figure 1-44 shows you some examples. From top to bottom: 1,500,000 ohms (1.5M) at 10% tolerance, 560 ohms at 5%, 4,700 ohms (4.7K) at 10%, and 65,500 ohms (65.5K) at 5%.

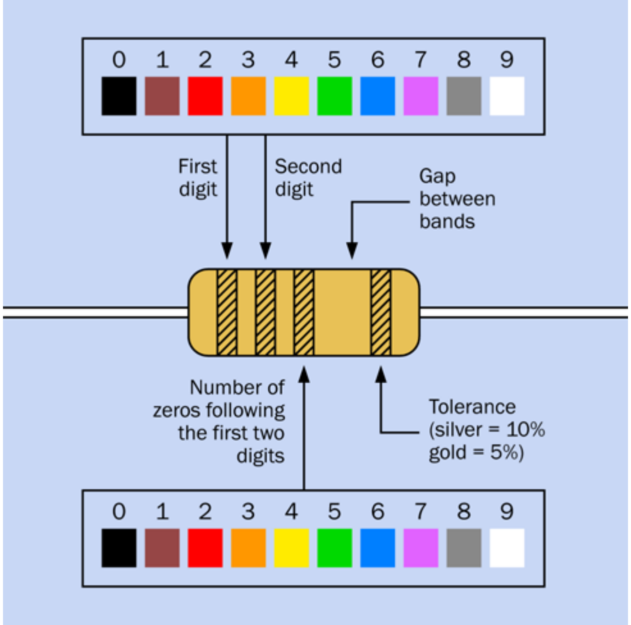
0 1 2 3 4 5 6 7 8 9
First digit
Second digit
Gap between bands
Number of zeros following the first two digits
Tolerance (silver = 10% gold = 5%)
0 1 2 3 4 5 6 7 8 9

Figure 1-44 *Four examples of color-coded resistors.*

The code can be summarized in words like this:

- Ignore the color of the body of the resistor. (The exception to this rule would be a white resistor that may be fireproof or fused, and must be replaced with the same type. But you are unlikely to encounter one of these.)
- Look for a silver or gold stripe. If you find it, turn the resistor so that the stripe is on the right-hand side. Silver means that the value of the resistor is accurate within 10%, while gold means that the value is accurate within 5%. This is known as the *tolerance* of the resistor.
- If you don't find a silver or gold stripe, turn the resistor so that the colored stripes are clustered at the left end. Usually there are three of them. If there are four, I'll deal with those in a moment.
- The colors of the first two stripes, from left to right, tell you the first two digits in the value of the resistor. The color of the third stripe from the left tells you how many zeros follow the two numbers. The color values are shown in Figure 1-43.

If you run across a resistor with four stripes instead of three, the first *three* stripes are digits and the *fourth* stripe is the number of zeros. The third numeric stripe allows a resistor to have an intermediate value.

Confusing? Well, you can still use your meter to check the values. Just be aware that the meter reading may be slightly different from the claimed value of the resistor. This can happen because your meter isn't absolutely accurate, or because the resistor is not absolutely accurate, or both. Small variations don't matter in the projects in this book.

Lighting an LED

Now take a look at one of your generic LEDs. Old-fashioned lightbulbs used to waste power by converting a lot of the power into heat. LEDs are much smarter: they convert almost all their power into light, and they last almost indefinitely—provided you treat them right!

An LED is quite fussy about the amount of power it gets, and the way it gets it. Always follow these rules:

- The *longer* wire sticking out of the LED must receive a *more positive* voltage relative to the shorter wire.
- The positive voltage difference that you apply between the long wire and the short wire must not exceed the limit stated by the manufacturer. This is known as the *forward voltage*.
- The current passing into the LED through the long wire and out through the short wire must not exceed the limit stated by the manufacturer. This is known as the *forward current*.

What happens if you break these rules? You'll see for yourself in Experiment 4.

Make sure you have a fresh 9-volt battery. You could use a connector with the battery, as shown in Figure 1-8, but I think it's easier just to clip a couple of test leads directly to the battery terminals, as in Figure 1-45.

Select a 2.2K resistor. Remember, 2.2K means 2,200 ohms. Why is it 2,200 and not a nice round amount such as 2,000? I'll explain that shortly. See "Background: Puzzling Numbers" on page 21 if you want to know right now.

The colored stripes on your 2.2K resistor should be red-red-red, meaning 2 followed by another 2 and two more zeros. You will also need a 1K resistor (brown-black-red) and a 470-ohm resistor (yellow-violet-brown), so have them ready.

Wire the 2.2K resistor into the circuit shown in Figure 1-45. Make sure you get the battery the right way around, with its positive terminal on the right.

- The "plus" symbol always means "positive."
- The "minus" symbol always means "negative."

Be sure that the long lead of the LED is on the right, and be careful that none of the alligator clips touches each other. You should see the LED glow dimly.

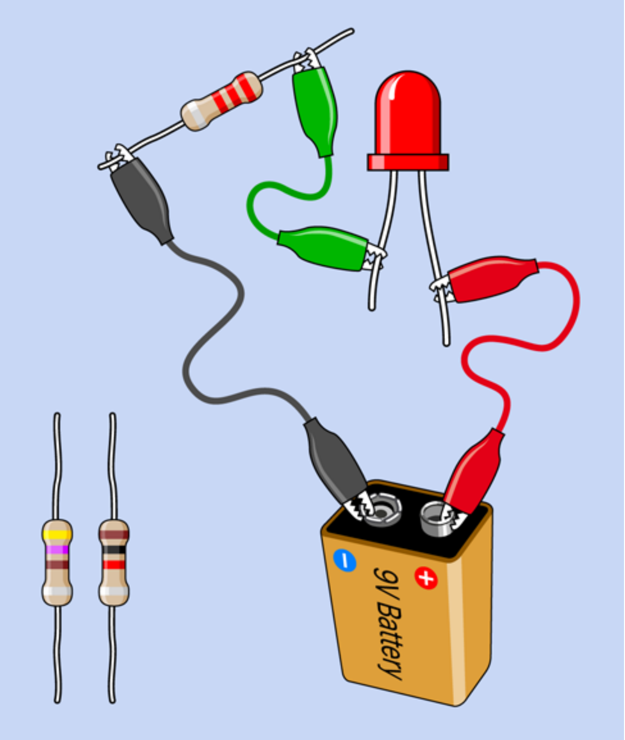

Figure 1-45 *Your first circuit, to light an LED.*

Now swap out your 2.2K resistor and substitute the 1K resistor. The LED should glow more brightly.

Swap out the 1K resistor and substitute a 470-ohm resistor, and the LED should be brighter still.

This may seem obvious, but it makes an important point. The resistor blocks a percentage of the current in the circuit. A higher-value resistor blocks more current, leaving less for the LED.

Checking a Resistor

I mentioned that you can use your meter to check the value of a resistor. This is really very easy. The procedure is shown in Figure 1-46. First, don't forget to set your meter to ohms. Disconnect the resistor from any other components, and apply the probes of your meter. If you have a manual-ranging meter, you must set the meter to a higher value than you expect to find. Otherwise, you'll get an error message.

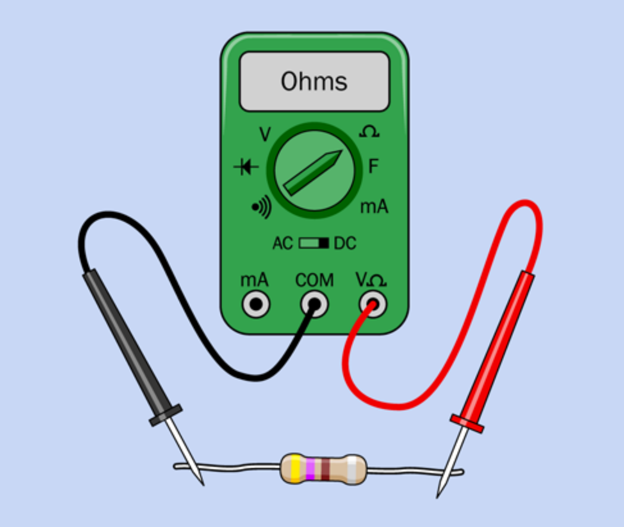

Figure 1-46 *Testing the value of a resistor.*

One thing to bear in mind is that you will get a more accurate reading if you press the probes firmly against the leads of the resistor. Don't hold the resistors and probes between your fingers—you don't want to measure the resistance of your body along with the resistance of the resistor. Place the resistor on an insulating surface, such as a nonmetallic desktop. Hold the probes by their plastic handles, and press down hard with the metal tips.

Alternatively, you can use a couple of test leads. Clip one end of each lead to each end of the resistor, and clip the other ends of the leads to the meter probes. Now you can do hands-free resistor testing, which is much easier.

Background: Puzzling Numbers

After you check a few resistors (or shop for them online) you'll notice that the same pairs of digits keep recurring. In thousands of ohms, we often find 1.0K, 1.5K, 2.2K, 3.3K, 4.7K, and 6.8K. In tens of thousands, we find 10K, 15K, 22K, 33K, 47K, and 68K.

The pairs of digits are known as *multipliers*, because you can multiply them by 1, or 1,000, or 10,000, or 100, or 10 to get basic resistor values in ohms.

There is a logical reason for this. Long ago, many resistors had an accuracy of plus-or-minus 20%, and therefore a 1.0K resistor could have an actual resistance as high as 1 + 20% = 1.2K while a 1.5K resistor could have a resistance as low as 1.5 − 20% = 1.2K. Therefore, it was pointless to have any values between 1K and 1.5K. Similarly, a 68 ohm resistor could have a value as high as 68 + 20% = just over 80 ohms, while a 100 ohm resistor could have a value as low as 100 − 20% = 80 ohms; so, it was unnecessary to have a value between 68 and 100.

In the top row of the table in Figure 1-47, the white numbers were the original multipliers for resistors. These numbers are still the most widely used today, even though modern resistor values are plus-or-minus 10% or better.

If you include the numbers in black type with the numbers in white type, you get all the possible multipliers for 10% resistors. If you then include the values in blue type, you have all the possible multipliers for 5% resistors.

1.0	1.5	2.2	3.3	4.7	6.8
1.1	1.6	2.4	3.6	5.1	7.5
1.2	1.8	2.7	3.9	5.6	8.2
1.3	2.0	3.0	4.3	6.2	9.1

Figure 1-47 *Traditional multipliers for resistor and capacitor values. See text for details.*

I have only used the original six multipliers for the projects in this book, to minimize the range of resistors that you will require. If accuracy is important (in Experiment 19, for example, where a circuit measures the speed of your reflexes) you can use a potentiometer to fine-tune the output—as I will show you in the very next experiment.

Cleanup and Recycling

You'll use the battery and the LED in the next experiment. The resistors can be reused in the future.

Experiment 4: Variable Resistance

You can vary the resistance in a circuit by inserting a *potentiometer*, which will control the current. The potentiometer in this experiment will enable you to learn more about voltage, amperage, and the relationship between them. You'll also learn how to read a manufacturer's datasheet.

What You Will Need

- 9-volt battery (1)
- Resistors: 470 ohms (1) and 1K (1)
- Generic LEDs (2)
- Test leads with alligator clips at each end (4)
- Potentiometer, 1K linear (2)
- Multimeter (1)

Look Inside Your Potentiometer

The first thing I want you to do is see for yourself how a potentiometer works, and the best way to accomplish this is to open it. This is why I asked you to acquire two potentiometers for this experiment—in case you can't put the first one back together again.

Some readers of the first edition of this book complained that it's wasteful to risk destroying a potentiometer by prying it open. But almost any learning experience consumes some resources, from pens and paper to whiteboard markers. If you really don't want to risk the future of your potentiometer, you can leave it untouched while you study the photographs that follow.

Most potentiometers are held together with little metal tabs. You need to bend the tabs upward. One way to do this is to slide in a knife and use it as a lever. Another way is to use a screwdriver—or maybe some pliers. I haven't specified any tools for this experiment, because I am hoping you already have a knife, or a screwdriver, or pliers in your home.

Figure 1-48 shows the tabs circled in red. (A fourth one is hidden behind the shaft of the component.) Figure 1-49 shows the tabs bent upward and outward.

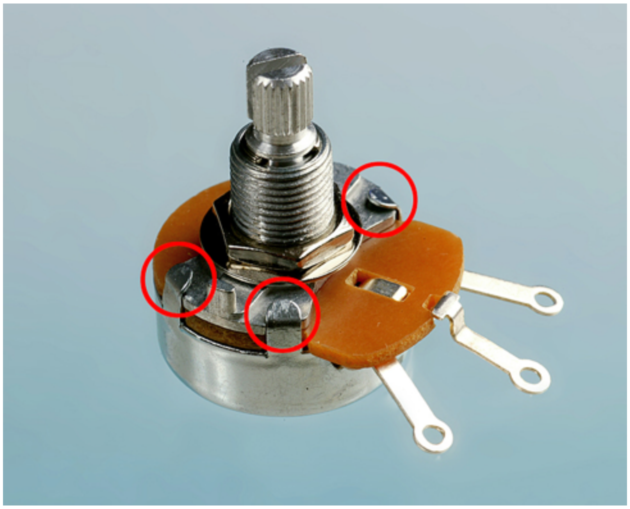

Figure 1-48 *The tabs that hold a potentiometer together.*

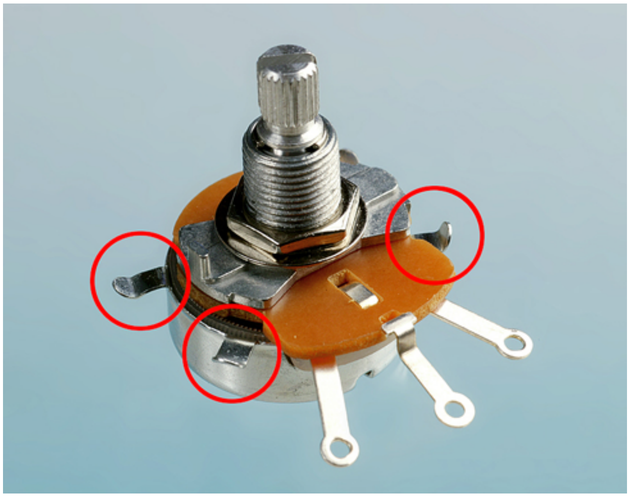

Figure 1-49 *The tabs have been bent upward and outward.*

After you have pried up the tabs, very carefully pull up on the shaft while holding the body of the potentiometer in your other hand. It should come apart as shown in Figure 1-50.

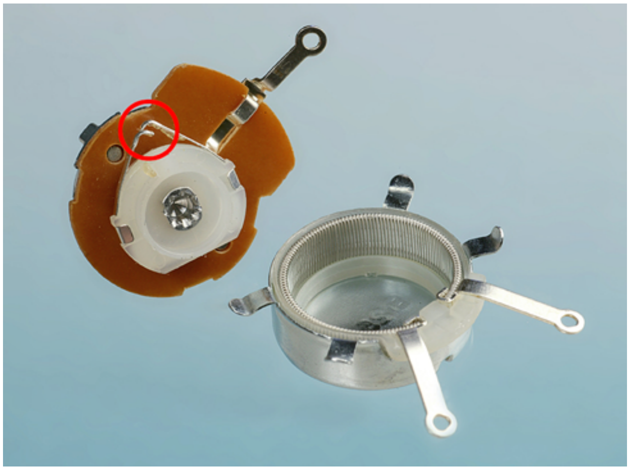

Figure 1-50 *The wiper of the potentiometer is circled.*

Inside the shell you will find a circular *track*. Depending on whether you have a really cheap potentiometer or a slightly more high-class version, the track will be made of conductive plastic or will have thin wire wrapped around it, as shown in the photograph. Either way, the principle is the same. The wire or the plastic possesses some resistance (a total of 1,000 ohms in a 1K potentiometer), and as you turn the shaft, a *wiper* rubs against the resistance, giving you a shortcut to any point from the center terminal. The wiper is circled red in Figure 1-50.

You can probably put it back together, but if necessary, use your backup potentiometer.

Testing the Potentiometer

Set your meter to measure resistance (at least 1K, on a manual-ranging meter) and touch the probes to the two adjacent terminals shown in Figure 1-51. You should find that when you turn the shaft of the potentiometer clockwise (seen from above), the resistance diminishes to almost zero. When you turn the shaft counterclockwise, the resistance increases to about 1K. Now keep the black probe where it is, and touch the red probe on the opposite terminal. The behavior of the potentiometer will be reversed.

Do you think, maybe, the middle terminal connects with the wiper inside the potentiometer? Do you think the other two terminals connect with the ends of the track?

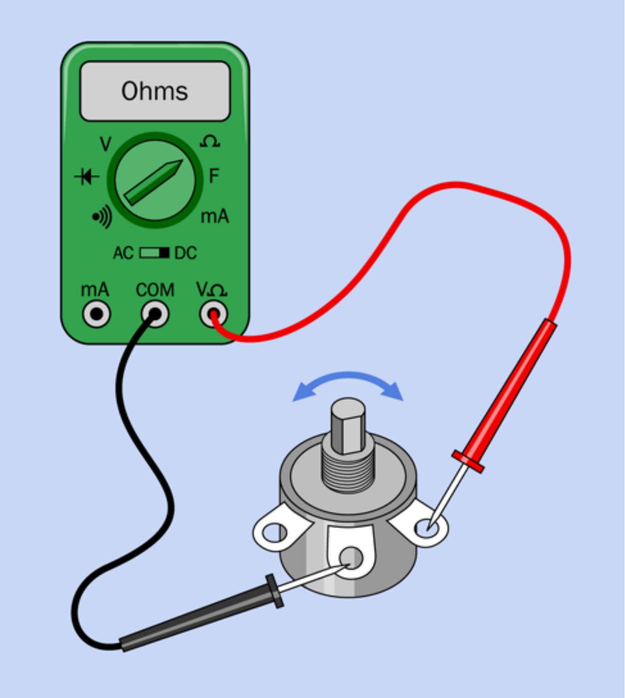

Figure 1-51 *Procedure for testing the behavior of a potentiometer.*

If you move the red probe to where the black probe is, and move the black probe to where the red probe was, the resistance between them will not change. It's the same in both directions. Unlike an LED, which has to be connected the right way around, a potentiometer has *no polarity*.

Caution: Don't Add Power

Don't apply power to a circuit while trying to measure resistance. Your meter uses a small amount of voltage from its internal battery when you are measuring resistance. You don't want that voltage to fight with voltage that you are applying from a battery.

Caution: Destructive Experiment Ahead

I have performed the next procedure many times uneventfully, but one reader reports that his LED fractured. You may wish to use safety glasses, if you want to be cautious. Regular eyeglasses will be acceptable.

Dimming Your LED

Now you can use the potentiometer to control the brightness of your LED. Connect everything exactly as shown in Figure 1-52. Make sure the two alligator clips are on the terminals shown. You are now using a variable resistance (the potentiometer) where the fixed resistor was in Experiment 3 (see Figure 1-45).

Begin with the shaft turned all the way *counterclockwise* (seen from above), otherwise you'll burn out the LED before we even get started. Now turn the shaft clockwise, very slowly, as shown by the blue arrow. You'll notice the LED glowing brighter, and brighter, and brighter—until, oops, it just went dark! You see how easy it is to destroy modern electronics? When I titled this procedure "Dimming your LED," you probably didn't realize I was talking about dimming it permanently.

Set aside that LED. I'm sorry to say, it will never glow again.

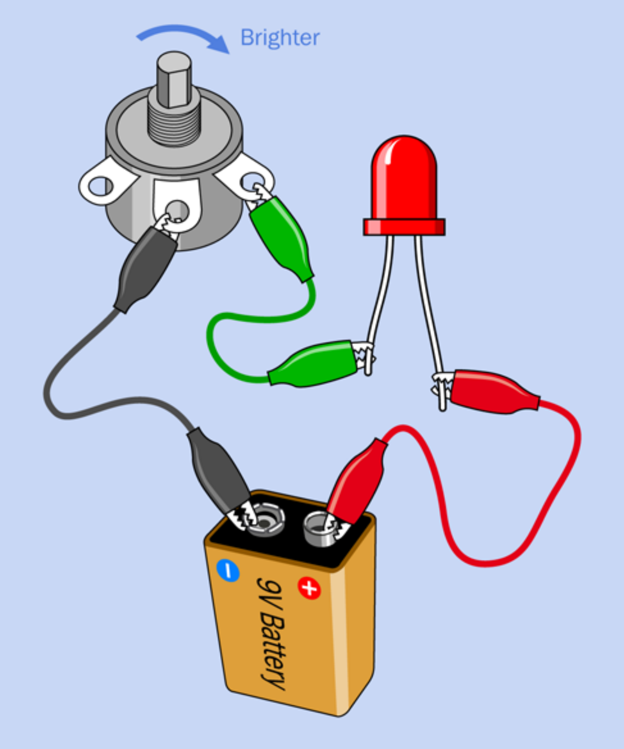

Figure 1-52 *Adjusting the brightness of an LED with a potentiometer.*

Substitute a new LED, and this time, let's protect it. Add a 470-ohm resistor, as shown in Figure 1-53. Electricity now passes through the 470-ohm resistor as well as the potentiometer, so that the LED will be protected even if the potentiometer's resistance diminishes to zero. You can turn the shaft of the potentiometer without worrying about destroying anything.

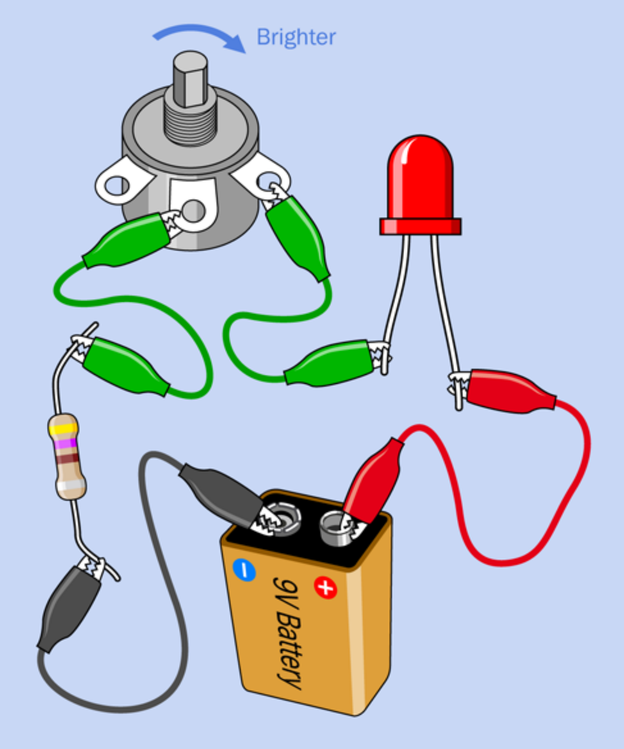

Figure 1-53 *Protecting the LED.*

The lesson that I hope you have learned is that an LED is too sensitive to be connected directly with a 9-volt battery. It must always be protected by some extra resistance in the circuit.

Could you power an LED directly from a single 1.5-volt battery? Try it. You may get a dim glow, but 1.5 volts is below the *threshold* for the LED. Let's find out how much voltage an LED needs.

Measuring Potential Difference

While the battery is connected in the circuit, set the dial of your meter to measure volts DC. You can leave the red lead plugged into the meter where it was before, because the socket for measuring volts is the same as the socket for measuring ohms.

If your meter uses manual ranging, set the voltage higher then 9 volts. Remember, the numbers beside the dial on the meter are the maximum in each range.

Now touch the probes to the terminals of the potentiometer that you are using, as shown in Figure 1-54. Try to hold the probes in place while you turn the potentiometer up a little, and down a little. You should see the voltage changing accordingly. We call this the *potential difference* between the two probes.

- "Potential difference" means the same as voltage between two points.

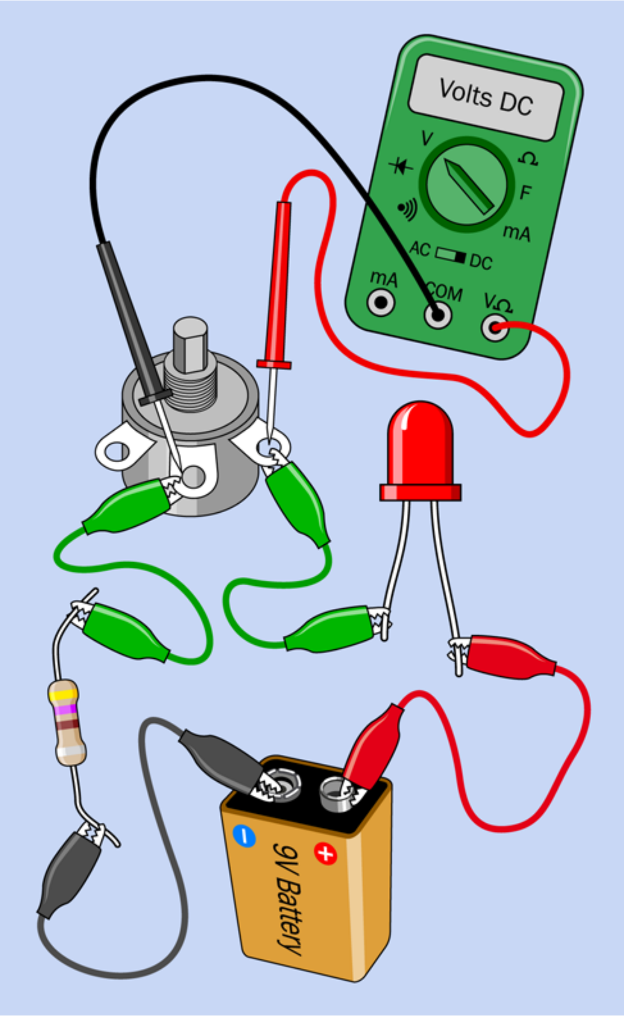

Figure 1-54 *Measuring the potential difference across an LED.*

If you measure the potential across the LED, this will also change when you adjust the potentiometer, although not as much as you might expect. An LED self-adjusts to some extent, modifying its resistance as the voltage and current fluctuate.

What if you swap the positions of the red and black probes? A minus sign should appear in the meter's display. You won't damage the meter this way, but it's less confusing if you always measure voltage with the red probe more positive than the black probe.

Lastly, touch the probes across the fixed-value resistor, and once again the potential difference will change when you adjust the potentiometer. Voltage from the battery is shared by all the components in this simple circuit. When the potentiometer reduces its share, a larger voltage difference is available for the fixed-value resistor and the LED. In addition, when the potentiometer has a lower resistance, the total resistance in the circuit is lower, allowing more current to flow.

A few things to keep in mind:

- If you add up the potential differences across all the devices in the circuit, the total will be the same as the voltage supplied by the battery.
- You measure voltage *relatively* between two points in a circuit. This is what their potential difference means.
- When measuring voltage, apply your meter like a stethoscope, without disturbing or breaking the connections in the circuit.

Checking the Flow

Now I want you to make a different measurement. I want you to check the amperage in the circuit, using your meter set to mA (milliamps). When measuring current, you must observe these rules:

- You can only measure current (amperage) when it passes *through* the meter.
- You have to insert your meter into the circuit.
- Too much current will blow the fuse inside your meter.
- You will have to use a socket on the meter labelled mA. This may be the same as the socket that you have used so far, or it may be different.

Make sure you turn the dial on your meter to measure mA, not volts, before you try this.

Caution: Meter Overload

Be careful when measuring current. For instance, if you put the probes of your meter directly against the terminals of a battery, and the meter is set to measure mA, you will create an instant overload, and the meter will blow its internal fuse. A cheap meter will not have any fuses supplied with it, so you'll have to open the case, check the value of the fuse, and search around online until you find an exact replacement. This is really annoying (I have been through it myself, more than once). A really cheap meter may not even have an easily replacable fuse.

- Always measure current when there are components in the circuit to restrict the flow.
- As a precaution, if your meter has a separate socket for measuring current, plug the red lead into that socket only while you are actually performing that task. Move the red lead back to the volts/ohms socket afterward.

Checking the Current

Insert the meter between the LED and the potentiometer, as shown in Figure 1-55. As you adjust the potentiometer up and down a little, you should find that the varying resistance in the circuit changes the flow of current—the *amperage*. The LED burned out in the previous experiment because too much current made it hot, and the heat melted it inside, just like a fuse. A higher resistance limits the amperage.

Now here's an interesting test. Turn the potentiometer all the way counterclockwise. Make a note of the current that you measure.

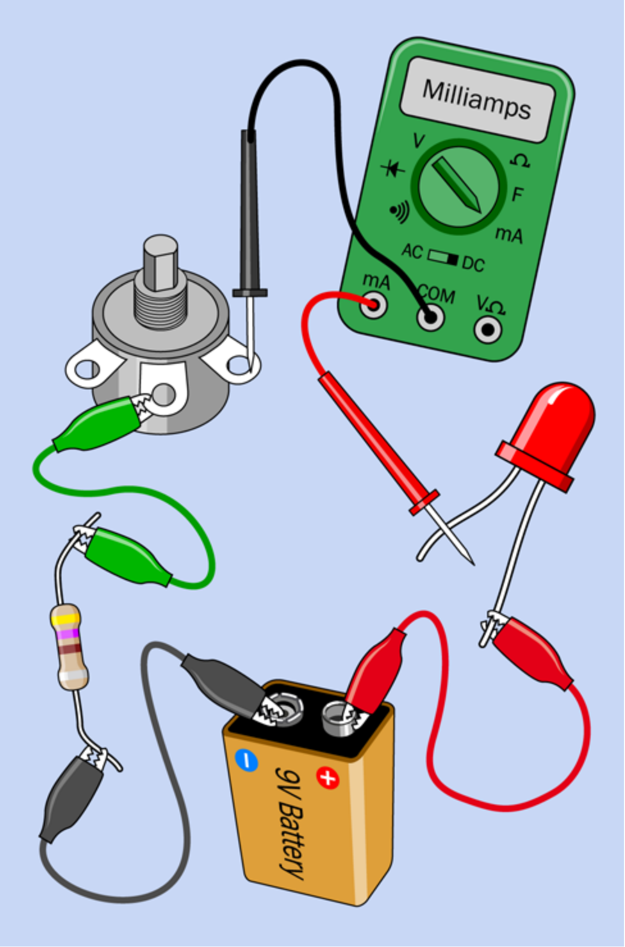

Figure 1-55 *Current passes through the meter on its way around the circuit.*

Without readjusting the potentiometer, move the meter and insert it between the battery and the LED, as shown in Figure 1-56. What's the value of the current now? It should be exactly the same as before—or very nearly the same, allowing for tiny changes in resistance that resulted from shifting the alligator clips.

- The current is the same at all points in a simple circuit. It has to be, because the flow of electrons has no place else to go.

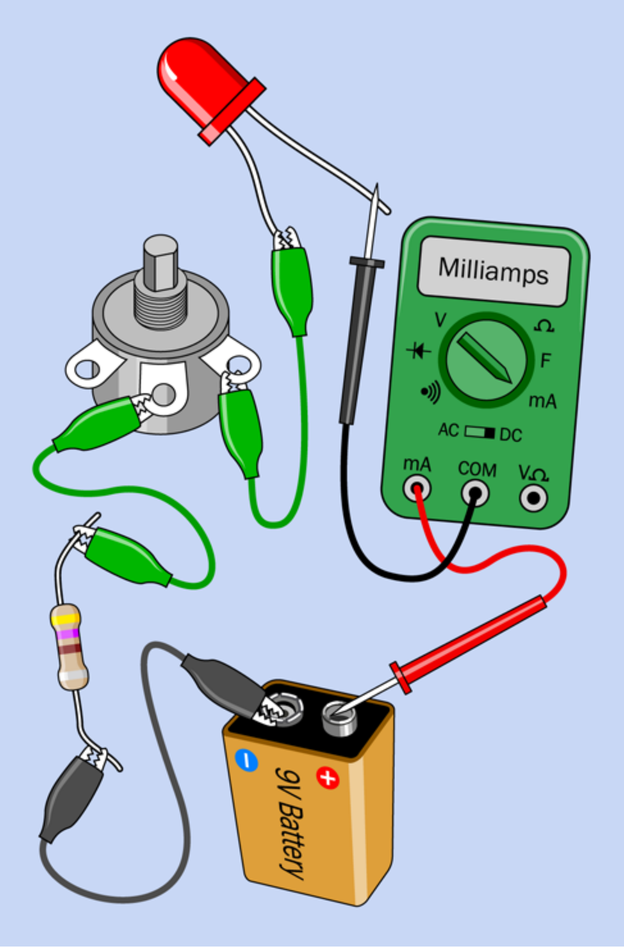

Figure 1-56 *The current flowing through a simple circuit is always the same throughout the circuit, regardless of where you measure it.*

Making Measurements

It's time now to nail this down with some numbers. This will enable you to establish the most fundamental rule in all of electronics.

Remove the LED from the circuit, and put the meter directly between the battery and the potentiometer. Remove the 470-ohm resistor, and substitute a 1K resistor (with colors brown-black-red), as shown in Figure 1-57. Now the only resistance in the circuit is provided by the 1K potentiometer plus the 1K resistor. (Your meter also has some resistance, but it's so low, we can ignore it. The wires and alligator clips also have some tiny amount of resistance, but this is even lower than the resistance of the meter.)

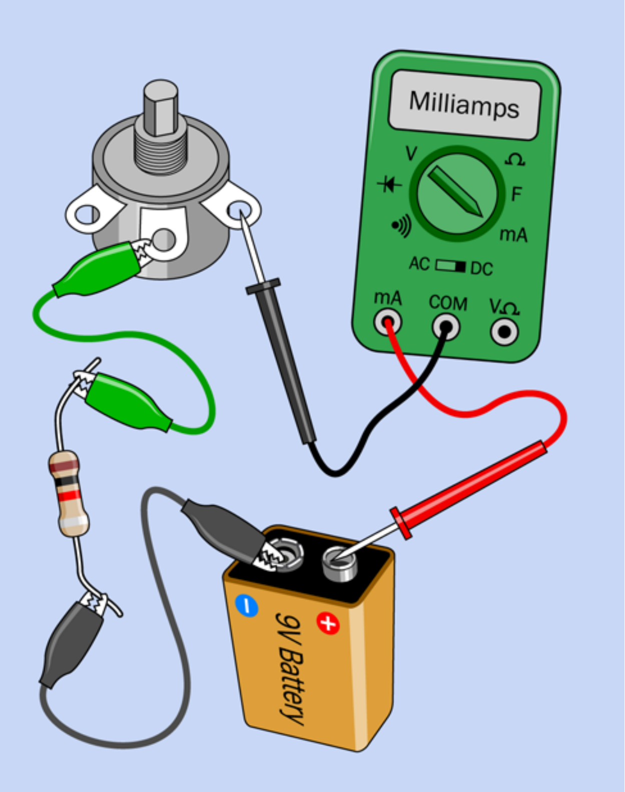

Figure 1-57 *In this final LED test, you get rid of the LED.*

Turn the potentiometer all the way clockwise, so that it provides almost-zero resistance. You now have only 1,000 ohms in the circuit, from the resistor. How much current does your meter show flowing?

Turn the potentiometer halfway, so it creates about 500 ohms resistance. The total resistance in the circuit is now 1,500 ohms, approximately. How much current does your meter show, now?

Turn the potentiometer all the way counterclockwise, so its full value is in the circuit, plus the resistor, making 2,000 ohms total. What's the amperage now?

When I tried this, I got the values shown below. Yours should be about the same.

`9mA with 1K total resistance`

`6mA with 1.5K total resistance`

`4.5mA with 2K total resistance`

Do you notice something interesting? On each line, if you multiply the number on the left by the number on the right, the result is always 9. And 9 volts just happens to be the voltage of the battery.

We only have three measurements, but if you ran a more detailed test using a series of resistors with fixed values, I'll bet that the result would be the same. I can summarize it like this:

```
battery voltage = milliamps × kilohms
```

But wait a minute: 1K is 1,000 ohms, and 1mA is 1/1,000 of an amp. Therefore, using the fundamental units of volts, amps, and ohms, our formula should really look like this:

```
voltage = (amps / 1,000) × (ohms × 1,000)
```

(I'm using the / symbol, which is often referred to as the "slash" symbol, to mean "divide by.")

The two factors of 1,000 cancel each other out, so we get this:

```
volts = amps × ohms
```

This is known as *Ohm's Law*. It is absolutely fundamental.

Fundamentals: Ohm's Law

The general way to express Ohm's Law is:

```
voltage = current × resistance
```

which is usually abbreviated like this:

```
V = I × R
```

Letter I represents the flow of current, because originally current was measured by its *inductance*, meaning the ability to induce magnetic effects. Perhaps it would be more helpful if some other letter, such as C, was used to represent current, but it's too late to persuade everyone to do that. You just have to remember that I means current.

By shifting the terms around, you get these versions of the formula:

```
I = V / R
```

```
R = V / I
```

To apply the formula, you have to make sure that the units are consistent. If V is measured in volts, and I is measured in amps, then R must be measured in ohms.

What if you have measured a current in milliamps? You must express it in amps. For instance, a current of 30mA must be written as 0.03 in the formula, because 0.03A = 30mA. If you get confused, use a calculator to divide milliamps by 1,000 to get a value in amps. Likewise, divide millivolts by 1,000 to get a value in volts.

To minimize the risk of errors, you can memorize Ohm's Law using the actual units, like this:

```
volts = amps × ohms
```

```
amps = volts / ohms
```

```
ohms = volts / amps
```

But remember:

- Volts are measured as a *difference in voltage* between two points in a simple circuit. Ohms are the resistance between the same two points. Amps are the current flowing through the circuit.

Fundamentals: Series and Parallel

In your test circuit, the resistor and the potentiometer have been connected in *series*, meaning that the electricity had to go through one before it went through the other. The alternative would have been to put them side by side, in *parallel*.

- Resistors in series are positioned so that one follows the other.
- Resistors in parallel are positioned side by side.

When you put two equal-valued resistors in series, you double the total resistance, because electricity has to pass through two barriers in succession. This is shown in Figure 1-58.

When you put two equal-valued resistors in parallel, you cut the total resistance in half, because you're giving the electricity two paths of equal resistance instead of one. This is shown in Figure 1-59.

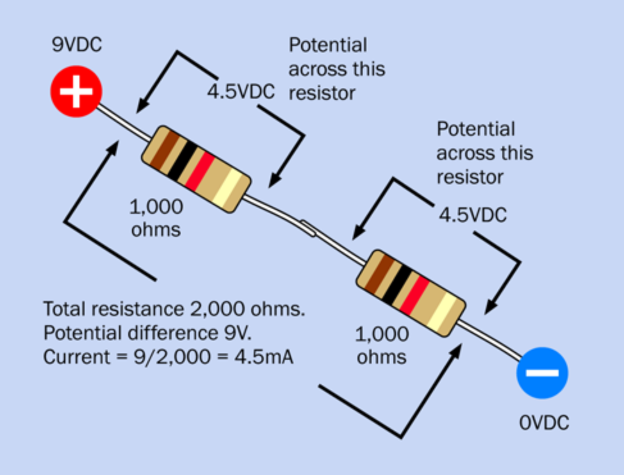
9VDC
Potential across this resistor
4.5VDC
Potential across this resistor
4.5VDC
1,000 ohms
Total resistance 2,000 ohms.
Potential difference 9V.
Current = 9/2,000 = 4.5mA
1,000 ohms
0VDC

Figure 1-58 *Two resistors of the same value in series.*

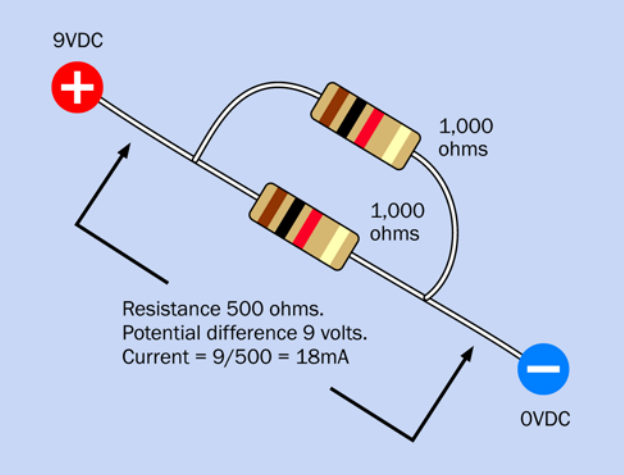

Figure 1-59 *Two resistors of the same value in parallel.*

In both figures, the current in milliamps has been calculated using Ohm's Law.

In reality we don't normally need to put resistors in parallel, but we often put other types of components in parallel. All the lightbulbs in your house, for instance, are wired in parallel across the main supply. So, it's useful to understand that resistance in a circuit goes down if you keep adding components in parallel. At the same time, as you add more pathways for electricity to follow, the total current through the circuit goes up.

Using Ohm's Law

Ohm's Law is extremely useful. For example, it can tell you precisely what resistance to put in series with an LED, to protect it adequately while generating as much light as possible.

The first step is to discover the specification for the LED established by its manufacturer. This information is easily available in a datasheet that you can locate online. Suppose you have an LED made by Vishay Semiconductors. You know that its part number is TLHR5400, because the number was printed on a label when you received a bag of LEDs in the mail, and you snipped out the label and stored it with the LEDs. (At least, that's what you should have done.)

All you have to do is Google the part number and the manufacturer's name:

```
vishay tlhr5400
```

The very first hit is the datasheet maintained by Vishay. Scroll down, and you see the information that you need. I've included the left side and right side of a screen capture in Figure 1-60. I outlined the part number of the component in red on the left, and I outlined two types of forward voltage on the right. "Typ" means typical and "Max," as I'm sure you guessed, is maximum. So, the LED typically should run with a 2V potential difference. But what does "at I_F (mA)" mean? Well, remember that the letter I is used to represent the current passing through a circuit. The letter F means "Forward." So the forward voltage in the table is measured at a forward current of 20mA, which is the recommended value for this LED.

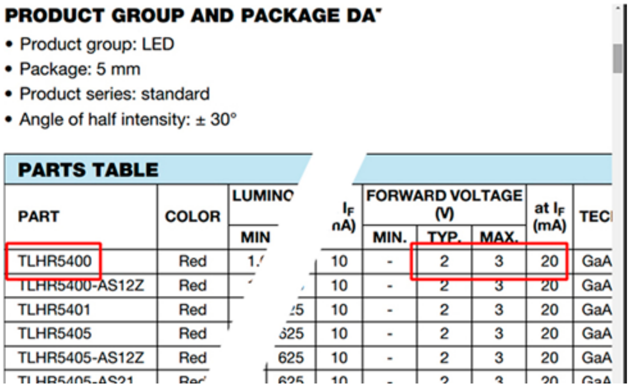

PRODUCT GROUP AND PACKAGE DA

- Product group: LED
- Package: 5 mm
- Product series: standard
- Angle of half intensity: ± 30°

PARTS TABLE

PART	COLOR	LUMINO		I_F nA)	FORWARD VOLTAGE (V)			at I_F (mA)	TEC
		MIN			MIN.	TYP.	MAX.		
TLHR5400	Red	1.		10	-	2	3	20	GaA
TLHR5400-AS12Z	Red			10	-	2	3	20	GaA
TLHR5401	Red		25	10	-	2	3	20	GaA
TLHR5405	Red		625	10	-	2	3	20	GaA
TLHR5405-AS12Z	Red		625	10	-	2	3	20	GaA
TLHR5405-AS21	Red		625	10	-	2	3	20	GaA

Figure 1-60 *Screen capture from an LED datasheet.*

How about if you have a Kingbright WP7113SGC? This time, the second hit from a Google search takes you to the appropriate datasheet, where the second page specifies a typical forward voltage of 2.2V, maximum

2.5V, and a maximum forward current of 25mA. The layout of the Kingbright datasheet is different from that of the Vishay datasheet, but the information is still easy to find.

Let's go with the Vishay LED. Now that you know that it works well with 2V and 20mA, Ohm's Law can tell you the rest.

How Big a Resistor?

In the simple circuit shown in Figure 1-61, you want to know the correct value for the resistor. Begin by recalling the rule I mentioned before:

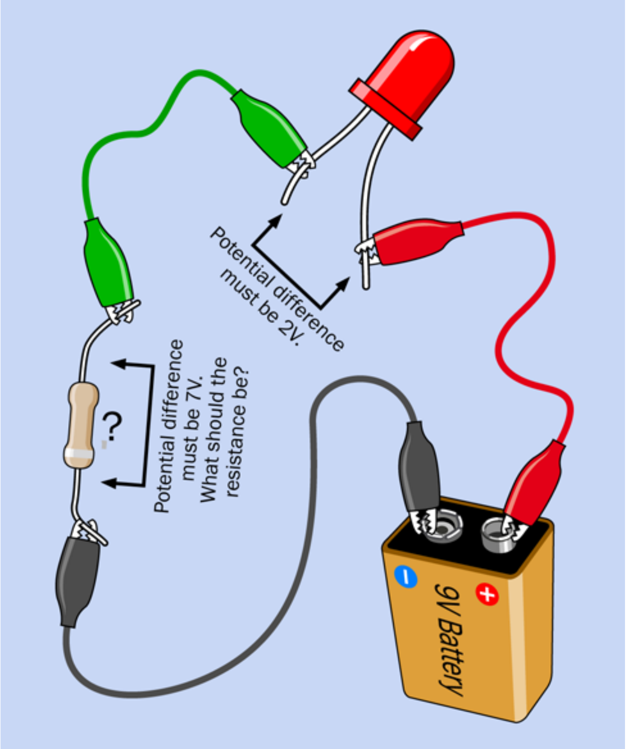

Figure 1-61 *This basic circuit enables you to calculate the value of the resistor.*

- If you add up the potential differences across all the devices in the circuit, the total will be the same as the voltage supplied by the battery.

The battery is 9V, of which we want the LED to take 2V. Therefore, the resistor must drop the voltage by 7V. What about the current? Remember another rule that I mentioned previously:

- The current in a simple circuit is the same at all locations in the circuit.

Therefore, the current through the resistor will be the same as the current through the LED. Your target is 20mA, but Ohm's Law requires you to make all the units match. If you are dealing with volts and ohms, you have to express the current in amps. Well, 20mA is 20 / 1,000 amps, which is 0.02 amps.

Now you can write down what you know, which is always the first step:

```
V = 7
```

```
I = 0.02
```

Which version of the Ohm's Law formula should you use? The one where the value that you don't know, and want to know, is on the left. That would be this one:

```
R = V / I
```

Now plug the values for V and I into the formula, like this:

```
R = 7 / 0.02
```

There's a trick that I will tell you for doing calculations involving decimal points, but to save time, use your calculator to show you the answer:

```
7 / 0.02 = 350 ohms
```

That is not a standard value for resistors, but 330 ohms is. Alternatively, just in case you use a more sensitive LED, you could move up to the next higher standard resistance value, which is 470 ohms. You may remember, I used a 470-ohm resistor in Experiment 3. Now you know why: I did the math.

Some people make the mistake of thinking that when they divide volts by amps to find the correct value for a series resistor, they should use the supply voltage (9V in this case). This is not correct, because the supply voltage is being applied to the resistor *and* the LED. To find the value for the resistor, you have to consider just the potential difference across it alone, which is 7V.

What happens if you use a different power supply? Later in the book, you'll be using a 5V supply in several experiments. How will this change the appropriate value of the resistor?

The LED still has to take 2V. The power supply is 5V, so the resistor has to drop 3V. The current should still be the same, and the calculation now looks like this:

`R = 3 / 0.02`

Therefore the resistance value is 150 ohms. But you don't need the LED to deliver its maximum light output, and you might happen to use an LED that has a lower limit than 20mA. Also, if a battery is powering the circuit, you might want to reduce the power consumption to make the battery last longer. Bearing this in mind, you could use the next higher standard value for the resistor, which would be 220 ohms.

Background: Hot Wires

I mentioned that wires have a very low resistance. Is it so low that you can always ignore it? Actually, no. If a large current is flowing through a wire, the wire will get hot, as you saw yourself when the 1.5-volt battery was shorted out in Experiment 2. And if the wire gets hot, you can be sure that some voltage is being blocked by the wire, leaving less voltage available for any device attached to the wire.

Once again, you can use Ohm's Law to establish some numbers.

Suppose that a very long piece of wire has a resistance of 0.2 ohms. You want to run 15 amps through it, to run a device that consumes a lot of power.

You begin by writing down what you know:

`R = 0.2 (resistance of the wire)`

`I = 15 (amperage through the circuit)`

You want to know V, the voltage drop between one end of the wire and the other. So you should use the version of Ohm's Law that places V on the left side:

`V = I × R`

Now plug in the values:

`V = 15 × 0.2 = 3 volts`

Three volts is not a big deal if you have a high-voltage power supply, but if you are using a 12-volt car battery, this length of wire will take one-quarter of the available voltage.

Now you know why the wiring in automobiles is relatively thick—to waste as little of the 12V power as possible.

Fundamentals: Decimals

Legendary British politician Sir Winston Churchill is famous for complaining about "those damned dots." He was referring to decimal points. Because Churchill was Chancellor of the Exchequer at the time, supervising all government expenditures, his difficulty with decimals was a bit of a problem. Still, he muddled through in time-honored British fashion, and so can you.

Suppose you have decimals in a division sum. You can make it easier by moving the decimal points in the top half and the bottom half of the division by the same number of places. So, when you wanted to know the answer to 7 / 0.02 to find the series resistor for the LED, you could move the decimals two spaces to the right:

`7 / 0.02 = 700 / 2`

This is much easier. Notice that if you move a decimal point to the right beyond a digit, you add a zero for each extra place. So, when you move the decimal in 7.0 two places to the right, you get 700.

What if you have decimals in a multiplication sum? For example, you need to multiply 0.03 by 0.002. Because you are now multiplying instead of dividing, you have to move the decimal points in opposite directions. Like this:

`0.03 × 0.002 = 3 × 0.00002`

So the answer is 0.00006. Again, if you find this too confusing, you can use a calculator. But sometimes it's quicker to use a pen and paper—or even do it in your head.

Theory: The Math on Your Tongue

I'm going back to the question I asked in the previous experiment: why didn't your tongue get hot?

Now that you know Ohm's Law, you can figure out the answer in numbers. Let's suppose the battery delivered its rated 9 volts, and your tongue had a resistance of 50K, which is 50,000 ohms. As always, begin by writing down what you know:

`V = 9`

`R = 50,000`

You want to know the current, I, so you use the version of Ohm's Law that puts this on the left:

`I = V / R`

Plug in the numbers:

`I = 9 / 50,000 = 0.00018 amps`

Move the decimal point three places to convert from amps to milliamps:

`I = 0.18 mA`

That's a very small current. It will not produce much heat.

What about when you shorted out the battery? How much current made the wires get hot? Well, suppose the wires had a resistance of 0.1 ohms (probably it's less, but I'll start with 0.1 as a guess). Write down what you know:

`V = 1.5`

`R = 0.1`

Once again I'm trying to find I, the current, so I use:

`I = V / R`

Plug in the numbers:

`I = 1.5 / 0.1 = 15 amps`

That's almost 100,000 times the current that may have passed through your tongue. It generated substantial heat in a thin wire.

A room heater or a large power tool, such as a table saw, might draw 15 amps. Maybe you're wondering if that little AA battery really could deliver as much current as that. The answer is . . . I'm not sure. I couldn't measure the current with my meter, because 15A would blow the meter's fuse, even if I plug the probe into the high-current socket labeled 10A. But I did try the experiment with a 10-amp fuse instead of a 3-amp fuse, and the 10-amp fuse survived.

Now, why was that? Ohm's Law says that the current should be 15A, but for some reason it was less. Maybe the resistance of the wire on the battery carrier was actually higher than 0.1 ohms? No, I think it was probably lower. So—what was restricting the current to a lower value than Ohm's Law predicted?

The answer is that everything in the everyday world has some electrical resistance, *even including a battery.* Always remember that a battery is an active part of a circuit.

Do you remember that when you shorted out the battery, it became hot, as well as the wires? Definitely, the battery has some *internal resistance*. You can ignore it when you are dealing with small currents in milliamps, but for large currents, the battery is actively involved.

This is why I cautioned against using a larger battery (especially a car battery). Larger batteries have a much lower internal resistance, allowing much higher currents, which can generate explosive amounts of heat. A car battery is designed to deliver literally hundreds of amps when it turns a starter motor. That's quite enough current to melt wires and cause nasty burns. In fact, you can weld metal using a car battery.

Lithium batteries also have low internal resistance, making them very dangerous when they're shorted out. Here's the take-home message:

- High current is not dangerous in the same way as high voltage. But it is still dangerous.

Background: The Watt

So far I haven't mentioned a unit that everyone is familiar with: watts.

A watt is a unit of power, and when power is applied throughout a period of time, it performs work. An engineer might say that work is done when a person, an animal, or a machine pushes something to overcome mechanical resistance. Examples would be a car cruising along a level stretch of road (overcoming friction and air resistance) or a person walking upstairs (overcoming the force of gravity).

When one watt of power is applied for one second, the work done is one *joule*, usually represented by letter J. If P is used to represent power:

`J = P × s`

Or if the formula is turned around:

`P = J / s`

When electrons push their way through a circuit, they are overcoming a kind of resistance, and so they are doing work.

The electrical definition of a watt is easy:

`watts = volts × amps`

Or, using the units customarily assigned, with W meaning watts, these three formulas all mean the same thing:

`W = V × I (watts = volts × amps)`

`V = W / I`

`I = W / V`

The terms milliwatts (mW), kilowatts (kW), and megawatts (MW) are commonly used in different situations—megawatts being usually reserved for heavy-duty equipment such as generators in power stations. Be careful not to confuse the lowercase m in the abbreviation for milliwatts with the uppercase M in the abbreviation for megawatts. A conversion table for milliwatts, watts, and kilowatts is shown in Figure 1-62.

Milliwatts	Watts	Kilowatts
1mW	0.001W	0.000001kW
10mW	0.01W	0.00001kW
100mW	0.1W	0.0001kW
1,000mW	1W	0.001kW
10,000mW	10W	0.01kW
100,000mW	100W	0.1kW
1,000,000mW	1,000W	1kW

Figure 1-62 *Conversion table for the most common multiples of watts.*

Old-style incandescent light bulbs are calibrated in watts. So are stereo systems. The watt is named after James Watt, inventor of the steam engine. Incidentally, watts can be converted to horsepower, and vice versa.

Resistors are commonly rated as being capable of dealing with 1/4 watt, 1/2 watt, 1 watt, and up. For all of the projects in this book, you can use 1/4-watt resistors. How do I know this?

Go back to the first LED circuit, using a 9V battery. Remember you wanted the resistor to drop the voltage by 7 volts, at a current of 20mA. How many watts of power would the resistor have to deal with?

Write down what you know:

`V = 7 (potential difference at the resistor)`

`I = 20mA = 0.02 amps`

You want to know W, so use this version of the formula:

`W = V × I`

Plug in the numbers:

`W = 7 × 0.02 = 0.14 watts`

This is the power being dissipated by the resistor.

Because 1/4 watt is 0.25 watts, a 1/4 watt resistor will have no trouble dealing with 0.14 watts. In fact you could almost use a 1/8 watt resistor, but in future experiments we may need resistors that can handle 1/4 watt, and there's no penalty for using a resistor that is rated for more watts than necessary. They just cost slightly more and are slightly larger.

Background: The Origins of Wattage

James Watt, shown in Figure 1-63, is known as the inventor of the steam engine. Born in 1736 in Scotland, he set up a small workshop in the University of Glasgow, where he struggled to perfect an efficient design for using steam to move a piston in a cylinder. Financial problems and the primitive state of the art of metal working delayed practical applications until 1776.

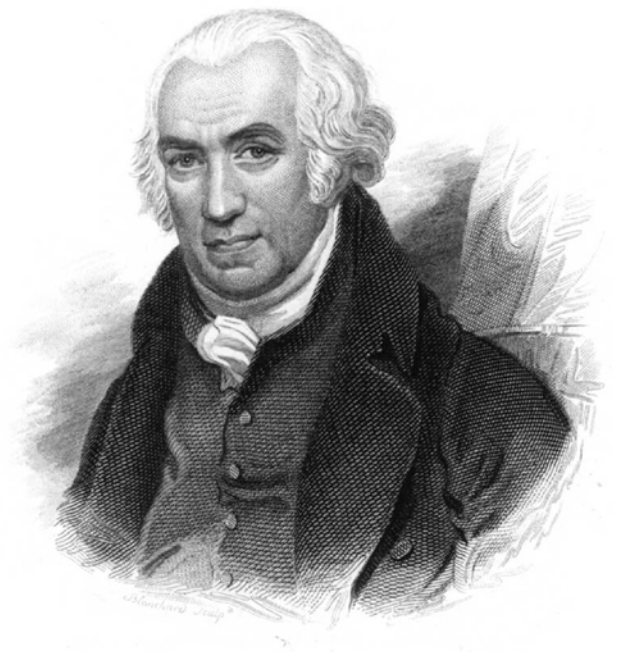

Figure 1-63 *James Watt's development of steam power enabled the industrial revolution. After his death, he was honored by having his name applied to the basic unit of power in electricity.*

Despite difficulties in obtaining patents (which could only be granted by an Act of Parliament in those times), Watt and his business partner eventually made a lot of money from his innovations. Although he predated the pioneers in electricity, in 1889 (70 years after his death) his name was assigned to the basic unit of electric power that can be defined by multiplying amperes by volts.

Cleanup and Recycling

The dead LED can be thrown away. Everything else is reusable.

Experiment 5: Let's Make a Battery

Long ago, before the Web existed, kids were so horribly deprived, they tried to amuse themselves with kitchen-table experiments such as making a battery by pushing a nail and a penny into a lemon. Hard to believe, but true!

Now that modern LEDs will emit light when just a few milliamps flow through them, the old lemon-battery experiment is more interesting. If you've never tried it, the time is right.

What You Will Need

- Lemons (2) or squeeze-bottle of pure lemon juice (1)
- Copper-plated coins, such as US pennies (4)
- One-inch (or larger) zinc-plated steel brackets from a hardware store (4)
- Test leads with alligator clips at each end (5)
- Multimeter (1)
- Low-current LED (1). (See "Light-Emitting Diodes" on page 6 for a reminder of the difference between generic and low-current LEDs)

Setup

A battery is an *electrochemical* device, meaning that chemical reactions create electricity. Naturally, this only works if you have the right chemicals working for you, and the ones I'm going to use are copper, zinc, and lemon juice.

The juice should be no problem. Lemons are cheap, or you can buy one of those little yellow plastic squeeze-bottles of concentrated juice. Either will work.

Pennies are not made of copper anymore, but they are still plated with a thin layer of copper, which is good enough. Just make sure that your pennies are new and bright. If the copper has oxidized, it will be a dull brown, and the experiment won't work as well.

Zinc is a bit more of a problem. What you need is a metal part that is *galvanized*, meaning it is coated in zinc to prevent rust. Small galvanized steel brackets should be available at your local hardware store, and they don't cost much. Brackets that measure about one inch along each side will be fine.

Lemon Test: Part One

Cut a lemon in half, and push a penny into it. As close as possible to the penny (but not touching it), push in your galvanized bracket. Now set your multimeter so that it can measure up to 2V DC, and hold one probe against the penny while you hold the other probe against the bracket. You should find that your meter detects between 0.8V and 1V.

To power a typical LED, you need more voltage. How can you get it? By putting batteries in series. In other words, more lemons! You'll use test leads to link the batteries, as shown in Figure 1-64. Notice that each lead connects a bracket to a penny. Do not connect pennies to pennies or brackets to brackets.

If you set things up carefully, keeping the pennies and brackets close to each other but making sure they don't actually touch each other, you should be able to illuminate your LED with three lemon-juice batteries in series.

Another option is to use a little parts box divided into small sections, as shown in Figure 1-65. When everything is nicely aligned, squeeze in some concentrated lemon juice. Vinegar or grapefruit juice may also work.

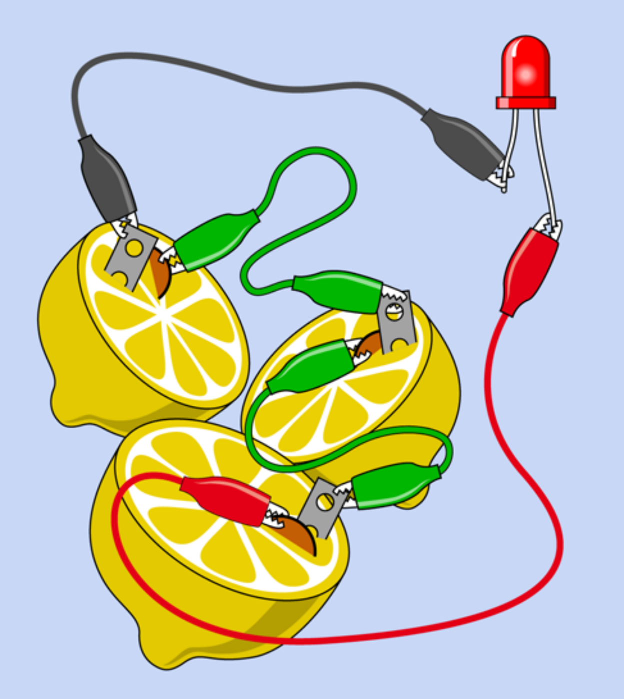

Figure 1-64 *A three-lemon battery should generate just enough voltage to drive a low-current LED.*

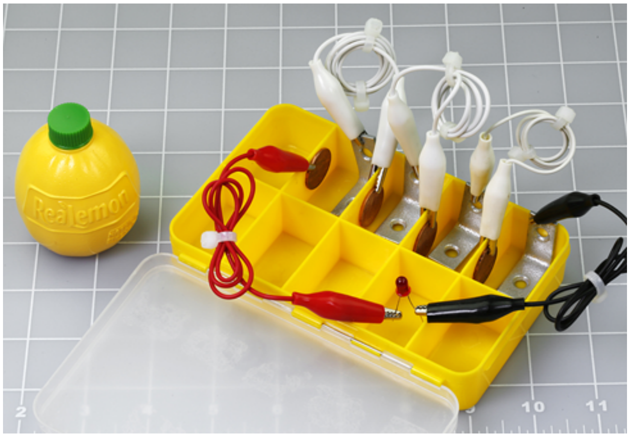

Figure 1-65 *Lemon juice, either from lemons or from a squeeze-bottle, will produce reliable results, even though the equipment doesn’t look so nice. Here a box intended for parts has been repurposed as a four-cell juice battery.*

I decided to have four cells in my juice battery, because the LED pulls the voltage down somewhat, and the battery is not capable of delivering enough current to damage the LED. The setup in the photograph worked immediately.

Theory: The Nature of Electricity

To understand why the lemon battery works, you have to start with some basic information about atoms. Each atom consists of a nucleus at the center, containing particles called protons, which have a positive charge. The nucleus is surrounded by electrons, which carry a negative charge.

Breaking up the nucleus of an atom requires a lot of energy, and can also liberate a lot of energy—as happens in a nuclear explosion. But persuading a couple of electrons to leave an atom (or join an atom) can take very little energy. For instance, when zinc reacts chemically with an acid, this can liberate electrons.

The reaction soon stops if the zinc-plated part isn’t connected with anything, as electrons accumulate with nowhere to go. They have a mutual force of repulsion, and you can imagine them like a crowd of hostile people, each one wanting the others to leave, and refusing to allow new ones to join them, as shown in Figure 1-66.

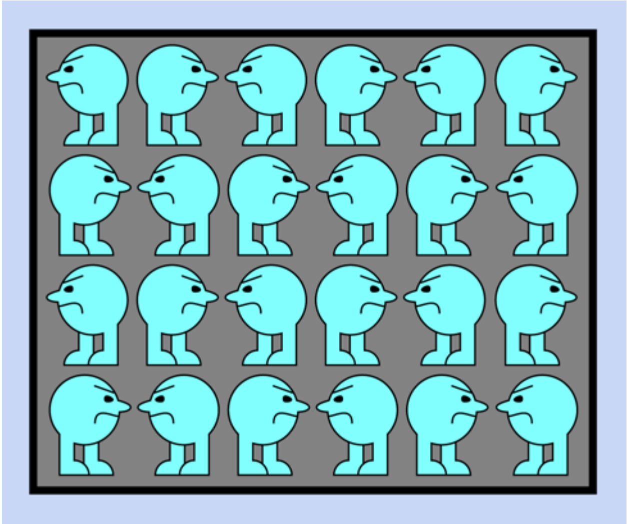

Figure 1-66 *Electrons on an electrode have a bad attitude known as mutual repulsion.*

Now consider what happens when a wire connects the zinc electrode, which has a surplus of electrons, to another electrode, made from a different material (such as copper), which contains “holes” for the electrons to occupy. The electrons can pass through the wire very easily by jumping from one atom to the next. As soon as we open up this pathway, mutual repulsion makes the electrons try to escape from each other to their new home as quickly as possible. This is how an electric current is created. See Figure 1-67.

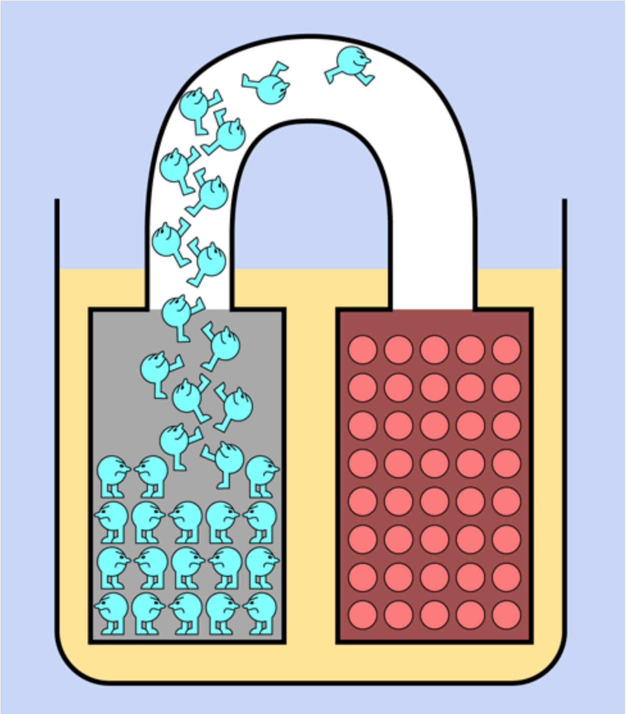

Figure 1-67 *Electrons escaping from a zinc electrode to a copper electrode.*

Now that the population of electrons on the zinc electrode has been reduced, the zinc–acid reaction can continue, replacing the missing electrons with new ones—which promptly imitate their predecessors and try to get away from each other by running away down the wire. They move with such force, we can divert them through an LED, and they will liberate some of their energy by making it light up.

The process continues until the zinc–acid reaction grinds to a halt, usually because it creates a layer of a compound such as zinc oxide, which won't react with acid and prevents the acid from reacting with the zinc underneath. (This is why your zinc electrode may have looked sooty when you pulled it out of the acidic electrolyte.)

This description applies to a *primary battery*, meaning one that is ready to generate electricity as soon as a connection between its terminals allows electrons to transfer from one electrode to the other. The amount of current that a primary battery can generate is determined by the speed at which chemical reactions inside the battery can liberate electrons. When the raw metal in the electrodes has all been used up in chemical reactions, the battery can't generate any more electricity and is dead. It cannot easily be recharged, because the chemical reactions are not easily reversible, and the electrodes may have oxidized.

In a rechargeable battery, also known as a *secondary battery*, a smarter choice of electrodes and electrolyte does allow the chemical reactions to be reversed.

Background: Positive and Negative

I've told you that electricity is a flow of electrons, which have a negative charge. In that case, why have I been talking as if electricity flows from the positive terminal to the negative terminal of a battery, in the experiments that you have performed so far?

The story started with an embarrassment in the history of research into electricity. When Benjamin Franklin was trying to understand the nature of electric current by studying phenomena such as lightning during thunderstorms, he believed he observed a flow of "electrical fluid" from positive to negative. He proposed this concept in 1747.

In fact, Franklin had made an unfortunate error that remained uncorrected until physicist J. J. Thomson announced his discovery of the electron in 1897. Electricity really is a flow of negatively charged particles, from an area of greater negative charge to some other location that is "less negative" or "more positive." In a battery, electrons originate from the negative terminal and flow to the positive terminal.

You might think that when this fact was established, everyone should have discarded Ben Franklin's idea of a flow from positive to negative. But people had been thinking in those terms for 150 years. Also, when an electron moves through a wire, you can think of an equal positive charge flowing in the opposite direction. When the electron leaves home, it takes a small negative charge with it; therefore, its home becomes a bit more positive. When the electron arrives at its destination, its negative charge makes the destination a bit less positive. This is pretty much what would happen if an imaginary positive particle traveled in the opposite direction. Moreover, all of the mathematics describing electrical behavior are still valid if you apply them to the imaginary flow of positive charges.

As a matter of tradition and convenience, Ben Franklin's erroneous concept of flow from positive to negative survived, because in the end, it makes no difference.

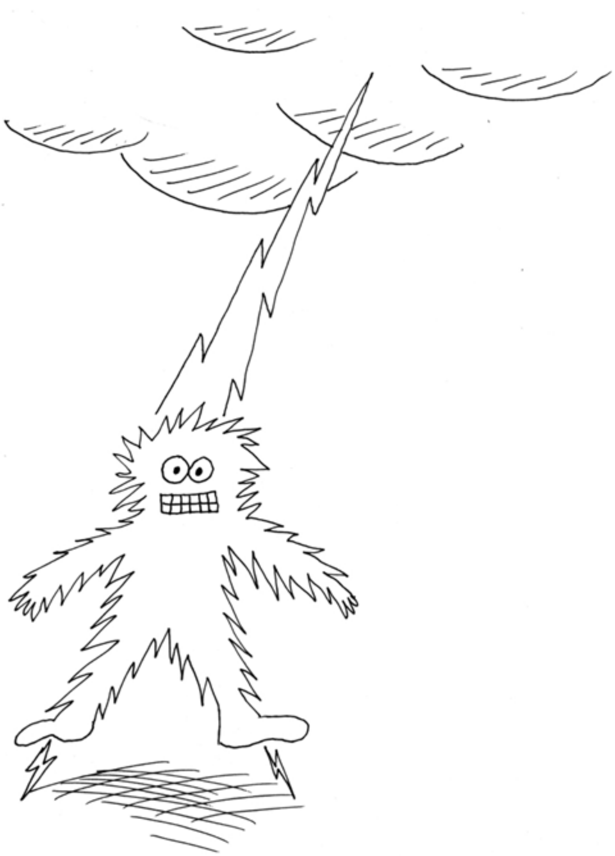

Figure 1-68 *In some weather conditions, the flow of electrons during a lightning strike can be from the ground, through your feet, out of the top of your head, and up to the clouds. Benjamin Franklin would have been surprised.*

In the symbols that represent components such as diodes and transistors, you will find arrows reminding you which way these components should be placed—and the arrows all point from positive to negative, even though that's not the way things work at all.

When Ben Franklin studied lightning, he saw this as an electric charge moving from a positive domain (the clouds in the sky) to a negative reservoir (the planet Earth). Well, it's true that the clouds are more positive, but this simply means that in reality, lightning is a transfer of electrons from the ground up to the sky. That's right: someone who is "struck by lightning" may be hurt by *emitting* electrons rather than by receiving them, as shown in Figure 1-68.

Theory: Basic Measurements

I'm going to backtrack, now, to the kinds of definitions you would normally find at the beginning of an electronics text.

Electrical potential is measured by adding up the charges on individual electrons. The basic unit is the *coulomb*, equal to the total charge on 6,241,509,629,152,650,000 electrons.

If you know how many electrons pass through a piece of wire each second, you can calculate the flow of electricity, which can be expressed in amperes. In fact:

`1 ampere = 1 coulomb/second`

`(about 6.24 quintillion electrons/second)`

Even if you could see inside a wire carrying electric current, electrons are smaller than the wavelength of visible light, so you would have no way to observe them, and there are far too many, moving much too fast. However, we have indirect ways of detecting them. For instance, the motion of an electron creates a wave of electromagnetic force. More electrons create more force, and this force can be measured. We can calculate the amperage from that. The electric meter installed at your home by the utility company functions on this principle.

The force required to push electrons through a conductor is voltage, and it creates a flow that can create heat, as you saw when you shorted out a battery. (If the wire that you used had zero resistance, the electricity running through it would not have created any heat.) We can use the heat directly, as in an electric stove, or we can use the electrical energy in other ways—to run a motor, for instance. Either way, we are taking energy from the electrons to do some work.

One volt can be defined as the amount of pressure that you need to create a flow of 1 ampere, which does 1 watt of work. As previously defined, 1 watt = 1 volt × 1 ampere, but the definition actually originated the other way around:

`1 volt = 1 watt / 1 ampere`

It's more meaningful this way, because a watt can be defined in nonelectrical terms. Just in case you're interested, we can work backward through the units of the metric system like this:

`1 watt = 1 joule / second`

`1 joule = 1 newton force acting through 1 meter`

`1 newton accelerates 1kg by 1m / sec each second`

On this basis, the electrical units can all be anchored with observations of mass, time, and the charge on electrons.

Practically Speaking

For practical purposes, I think an intuitive understanding of electricity can be more useful than the theory. I like to go back to the water analogies that have been used for decades in guides to electricity.

In Figure 1-31 I showed how the rate at which water leaks from a hole in a tank can be compared to amperage, while the height of the water in the tank creates pressure, comparable to voltage, and the size of the hole is equivalent to resistance.

Where's the wattage in this picture? Suppose you place a little water wheel where it is hit by the flow from the hole, as shown in Figure 1-69. You could attach some machinery to the water wheel. Now the flow would be doing some work. (Remember, wattage is a measurement of the rate at which work is done.)

Maybe this looks as if you would be getting something for nothing, extracting work from the flow of water without putting any energy back into the system. But remember, the water level in the tank is falling. As soon as I include some helpers hauling the waste water back up to the top of the tank, it becomes obvious that you have to put work in to get work out. See Figure 1-70.

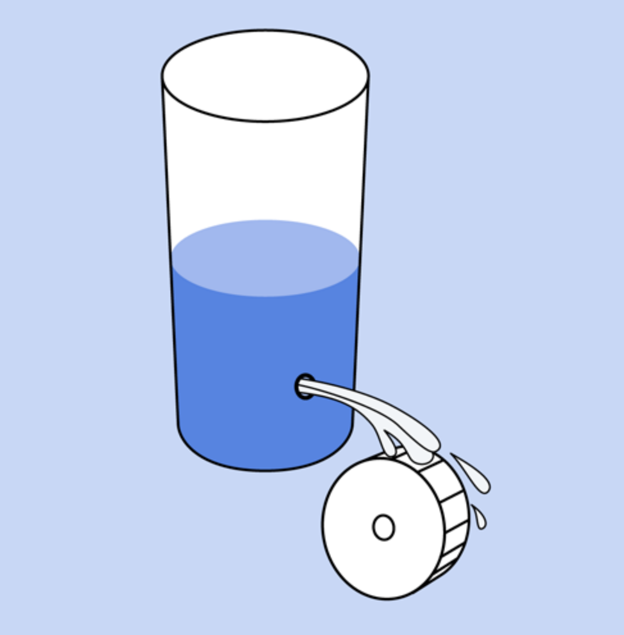

Figure 1-69 *If a wheel extracts energy from the flow of water, the flow is now doing some work, which could be measured in watts during a period of time.*

Similarly, a battery may seem to be giving power out without taking anything in, but the chemical reactions inside it are changing pure metals into metallic compounds, and the power we get out of a battery is enabled by this change of state. If it's a rechargeable battery, we have to push power back into it during the charging process, to reverse the chemical reactions.

Going back to the tank of water, suppose we can't get enough power out of it to turn the wheel. One answer could be to increase the height of the water, to create more force, as in Figure 1-71.

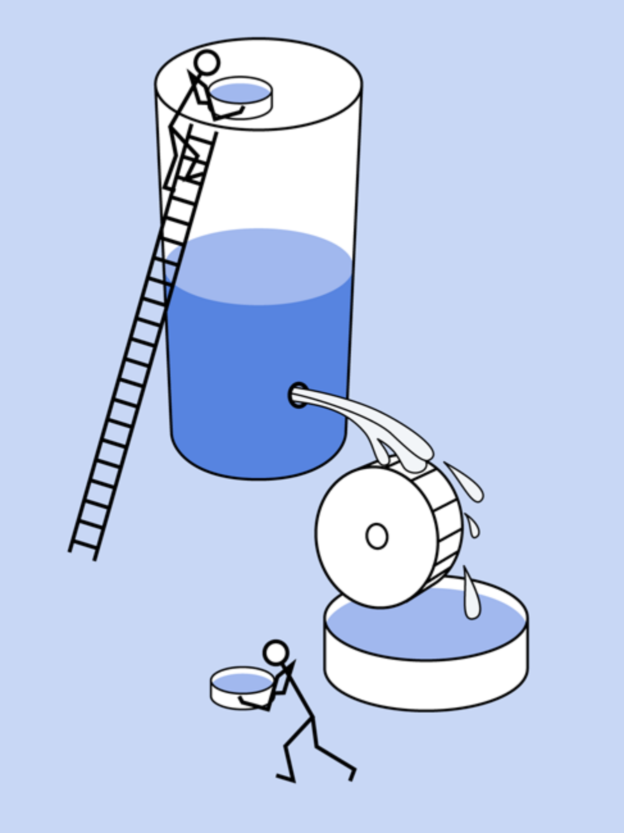

Figure 1-70 *To continue getting work out of the system, we have to put work into it.*

This would be the same as putting two batteries end to end, positive to negative, in series (as I suggested with lemons, in the lemon battery). Two batteries in series will double the voltage, as shown in Figure 1-72. As long as the resistance in the circuit remains the same, greater voltage will create more amperage, because amperage = voltage / resistance.

Thinking again about the tank analogy, what if we want to run a wheel for twice as long, and we've run out of tank capacity? Maybe we should build a second tank, and pipe their outputs through the same hole. Similarly, if you wire two batteries side by side, in parallel, you get the same voltage, but the batteries should last twice as long. Alternatively, the two batteries may be able to deliver more current than if you just used one. See Figure 1-73.

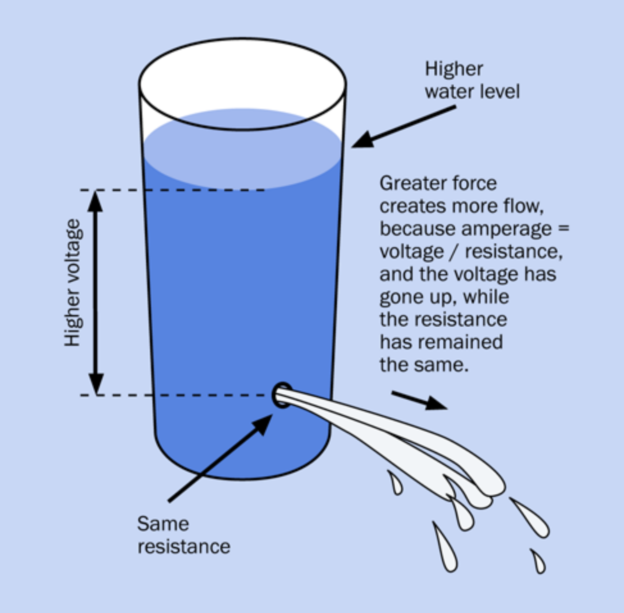
Higher
water level
Greater force creates more flow, because amperage = voltage / resistance, and the voltage has gone up, while the resistance has remained the same.
Higher voltage
Same resistance

Figure 1-71 *The amount of available work will increase with greater water pressure.*

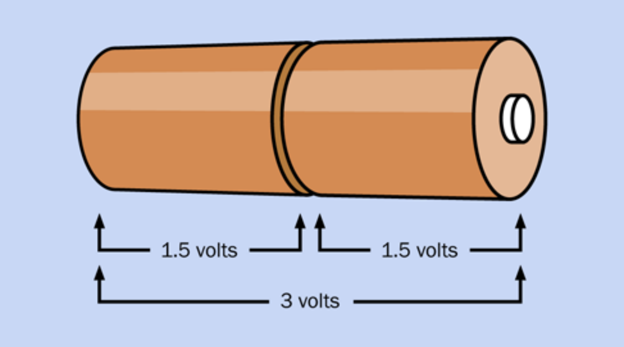
1.5 volts
1.5 volts
3 volts

Figure 1-72 *Two batteries in series provide twice the voltage of a single battery, provided they are both fully charged.*

Summing up:

- Two batteries in series deliver twice the voltage.
- Two batteries in parallel can deliver the same current for twice as long, or twice the current for the same amount of time.

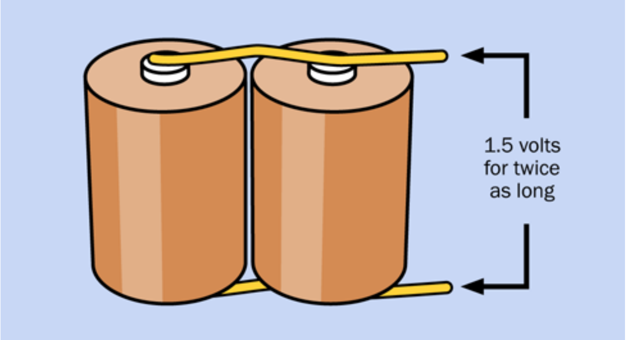

Figure 1-73 *Batteries in parallel, powering the same load as before, will run it for about twice as long. Alternatively, they can provide twice the current for the same time as a single battery.*

That's enough theory for now. In the next chapter, I'll continue with some experiments that will build on the foundations of knowledge about electricity, to take you gradually toward gadgets that can be fun and useful.

Cleanup and Recycling

The hardware that you immersed in lemons or lemon juice may be discolored, but it is reusable. Bearing in mind that some zinc ions may have been deposited in the lemons, eating them might not be such a good idea.

Switching

2

This chapter of the book contains Experiments 6 through 11, which explore the seemingly simple topic of switching—by which I mean not just manual control, but using one flow of electricity to switch, or control, another. It's such an important principle, no digital device can exist without it.

Today, switching is mostly done with transistors. I will deal with them in detail, but before that I'll back up and illustrate the concept by introducing you to relays, which are easier to understand, because you can see what's going on inside them. And before I get to relays, I'll show how manually activated switches can demonstrate some of the concepts that will follow. So, the sequence will be switches—relays—transistors.

Also in this chapter, I'll deal with capacitance, because capacitance is almost as fundamental as resistance in electronic circuits.

Necessary Items for Chapter Two

As before, when buying tools and equipment, see "Buying Tools and Equipment" on page 324 for a shopping list. If you want kits containing components and supplies, see "Kits" on page 311. If you prefer to buy your own components and supplies from online sources, see "Components" on page 317.

Essential: Small Screwdrivers

A set such as the one made by Stanley (part number 66-052) is shown in Figure 2-1. Screwdrivers that you may already have in your home will be too big for most of the little screws that you find on components.

You can buy cheaper screwdriver sets that look very similar to the one in Figure 2-1, but I think the quality of steel is better in name-brand screwdrivers.

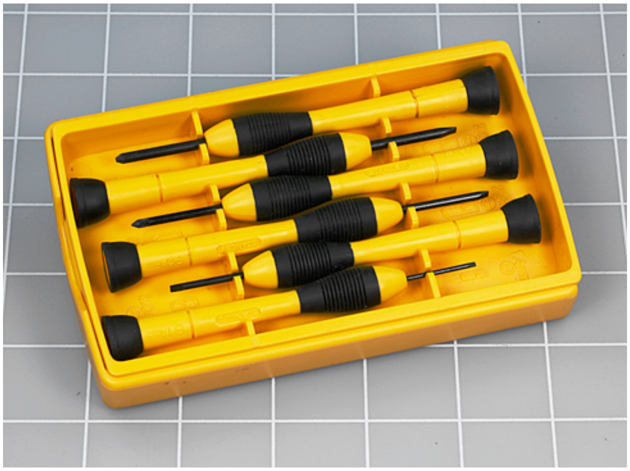

Figure 2-1 *Miniature screwdrivers, with both flat blades and Phillips blades. The white lines are spaced at intervals of 1".*

Essential: Small Long-Nosed Pliers

The type of *long-nosed pliers* that you need are no more than five inches from end to end. You'll be using them to bend wire precisely, or to pick up small parts where a finger and thumb are too big and clumsy. For these kinds of operations I don't think you benefit much from spending extra money on high-quality tools, so feel free to buy the cheapest. An example is shown in Figure 2-2. These have spring-loaded handles, which some people don't like, but you can pull out the springs—if you have a second pair of pliers to do so.

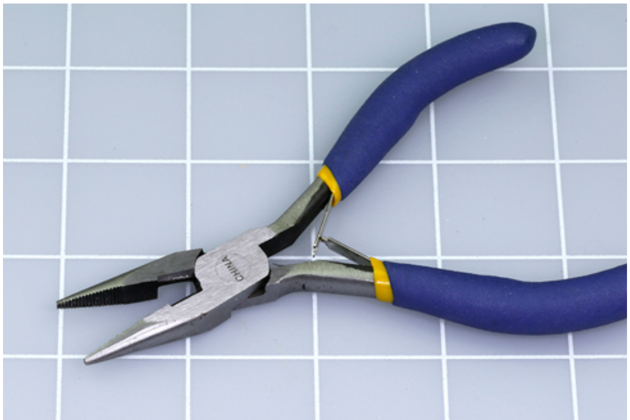

Figure 2-2 *Appropriate pliers for electronics work should be no more than 5" long.*

Optional: Sharp-Nosed Pliers

These are like small long-nosed pliers, but with very precise, pointed jaws. I like them for accessing tightly packed components on a breadboard. Your best source is a website or store that specializes in craft work such as beading. Be careful, however, that you don't buy beading pliers that have rounded noses for making loops in wire. For our purposes, the inner surfaces of the jaws should be flat, as shown in Figure 2-3.

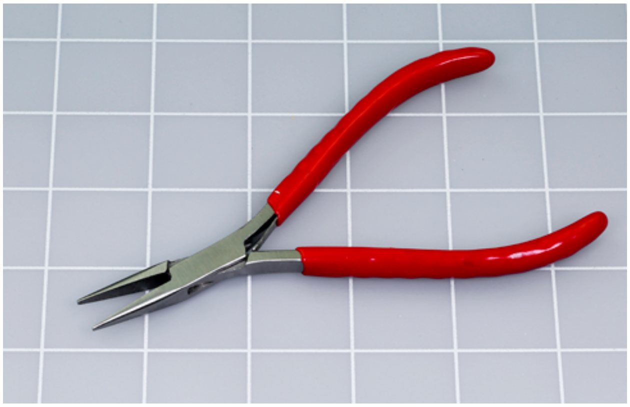

Figure 2-3 *Sharp-nosed pliers enable very precise work on a small scale.*

Essential: Wire Cutters

Pliers usually have cutting edges near the joint, which you can use for chopping wire. Often, however, a piece of wire will be attached to something else, and you won't be able to reach it with your pliers. You really need *wire cutters* such as those shown in Figure 2-4. They should be no longer than five inches. So long as you use them to cut thin, soft copper wire, they do not have to be of high quality.

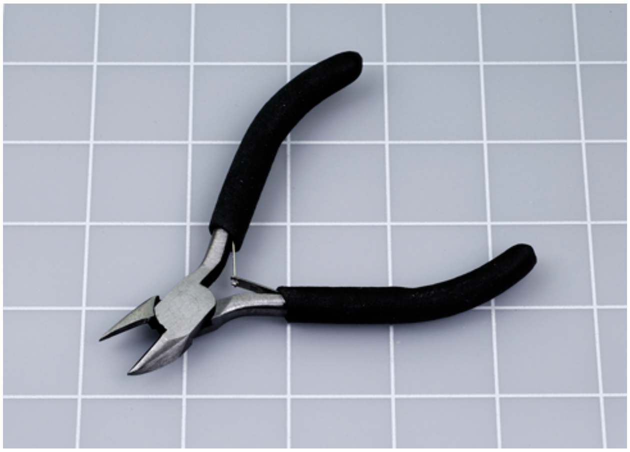

Figure 2-4 *Wire cutters should be no more than 5" long.*

Optional: Flush Cutters

Flush cutters, shown in Figure 2-5, are similar to wire cutters, and do the same job, but they are thinner, smaller, and better able to get into small spaces. However, they are less robust. Whether you use them or wire cutters is a matter of personal preference. Personally I like wire cutters.

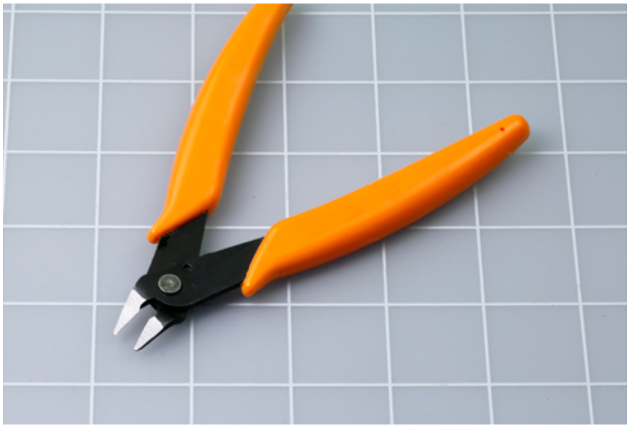

Figure 2-5 *Flush cutters can get into smaller spaces than wire cutters.*

Essential: Wire Strippers

Wire of the type that you will be using has a coating of plastic insulation. *Wire strippers* are specifically designed to remove a short section of insulation to expose the conductor inside it. Macho hobbyists may claim that they don't need any tool to do this job, but I have two corners missing from the insides of my front teeth to testify that this is a really bad idea (see Figure 2-6).

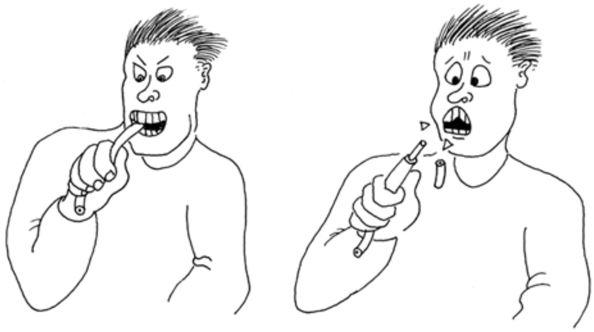

Figure 2-6 *In a hurry? Can't find your wire strippers? The temptation is obvious, but not a good idea.*

Another option is to use wire cutters, as shown in Figure 2-7. You hold the wire in one hand, close the jaws of the cutters gently with the other hand, and lever your two hands apart. However, this is a skill that requires practice. Sometimes the cutters slip off without doing anything, or they may cut the wire instead of stripping it.

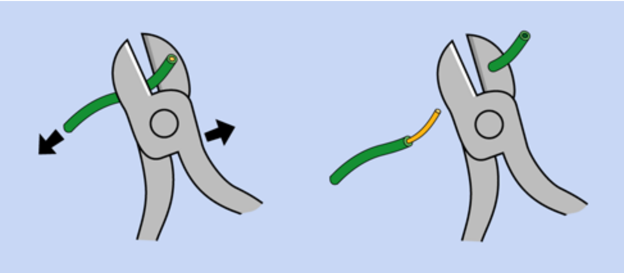

Figure 2-7 *How to use wire cutters to strip insulation. Wire strippers are easier to use.*

For a few extra dollars, a pair of wire strippers makes the job so much easier.

The first edition of this book included an option to buy so-called automatic wire strippers, which can be used with only one hand. Unfortunately they are significantly more expensive than other types of wire strippers, and many brands don't work well with the 22-gauge hookup wire required for all the circuits in this book. Therefore, I don't recommend them anymore.

The type of tool shown in Figure 2-8 is available from many manufacturers. Some brands have angled handles, some have straight handles, and some have curved handles. I don't think it matters. They all work the same way: you insert the wire in a hole of appropriate size, close the jaws, and pull away the insulation.

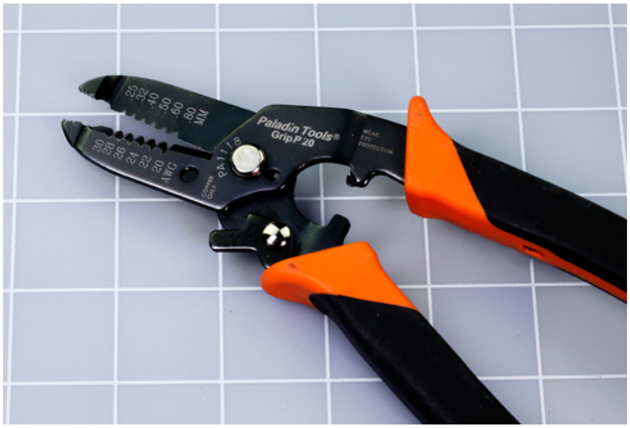

Figure 2-8 *The recommended wire strippers are designed for 20-gauge to 30-gauge wire.*

You do have to be careful, though, that they are suitable for the right size of wire.

Wire gauge is a measure of the thickness of a conductor. A higher number indicates a thinner wire. As it happens, 20-gauge wire is slightly too thick for our purposes, while 24-gauge wire is slightly too thin. The optimal thickness is 22 gauge, and you will have a much easier time if you buy a tool that is calibrated for that specific size. As you can see in Figure 2-8, a range from 20 gauge to 30 gauge includes a little hole for 22 gauge. This is the right tool for the job.

Essential: Breadboards

Breadboards are not required until Experiment 8, but I will provide a brief introduction here. The breadboard is a little plastic slab perforated with holes at intervals of 0.1". You push components and wires into the holes. Hidden beneath the plastic are conductors that make connections along the rows of holes.

A breadboard allows components to be connected together more neatly than with the test leads that you have used so far, and more easily (and reversibly) than if you join them with solder.

- A breadboard is properly known as a *solderless breadboard* and is sometimes referred to as a *prototyping board*.

The brand or source is not important, but you should be careful to buy the same configuration that I have used throughout this book. You have three options, only one of which is correct:

Breadboard Option 1: The Mini, shown in Figure 2-9. Often sold as being "suitable for Arduino," it doesn't have enough holes in it for our purposes, so, don't buy one of those.

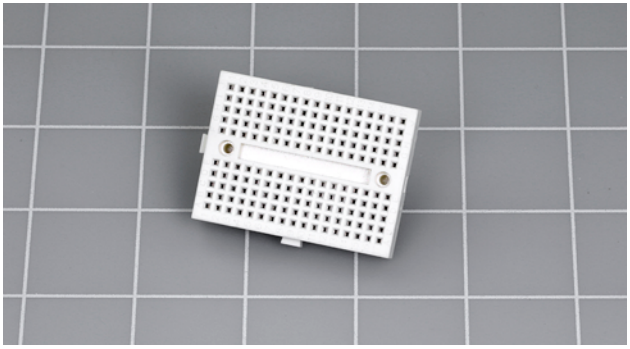

Figure 2-9 *A mini-board is not big enough for many of the projects in this book.*

Breadboard Option 2: Single-Bus, shown in Figure 2-10. The term "bus" refers to a long column of holes alongside the short, numbered rows of holes. There is a single bus on each side, outlined in red in the photograph. This is the kind of board that you want. Check a photograph of the product that you are buying, to make sure. Also, note that it should have 60 rows of holes and 700 connection points (also known as tie points). If you are buying your own, search Amazon or eBay for:

```
solderless breadboard 700
```

If you prefer, you can use a dual-bus breadboard and just ignore the extra holes.

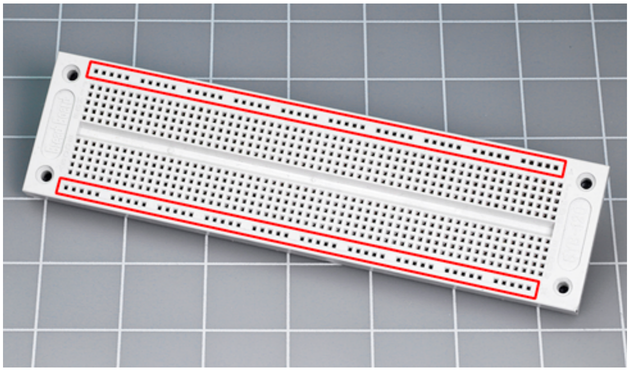

Figure 2-10 *A single-bus breadboard has a single long row of holes along each side.*

Breadboard Option 3: Dual Bus, shown in Figure 2-11. This has two long rows of holes on each side, which are the dual buses, outlined in red in the photograph. I used this type of board in the first edition of this book, because it can be more convenient. Then I saw how it encouraged wiring errors when people used it for the first time, so I don't recommend dual-bus breadboards anymore.

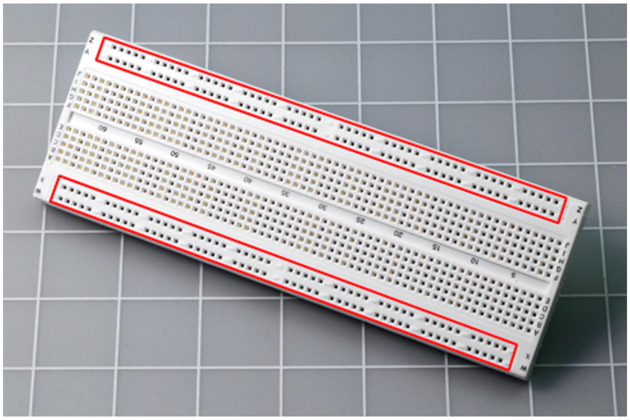

Figure 2-11 *A dual-bus breadboard has two pairs of long rows of holes, outlined in red in this photograph. This style of board is no longer recommended.*

Now that I've established the type of board that I recommend, how many of them do you need? I used to say, "Only one," bearing in mind that they are reusable, but their price has been driven down to the point where I think you should consider buying two or three. This way, you can prototype new circuits without having to disassemble old ones first.

Supplies

If you want to buy a kit containing components and supplies, see "Kits" on page 311. When buying your own supplies, see "Supplies" on page 316.

Essential: Hookup Wire

The type of wire you need for making connections on a breadboard is often called *hookup wire* although it can be found, sometimes, under the general category of *bulk wire*. Either way, it must have a solid core (not stranded), and must be 22 gauge in size.

Often it is sold in 25-foot and 100-foot lengths on plastic spools, as shown in Figure 2-12.

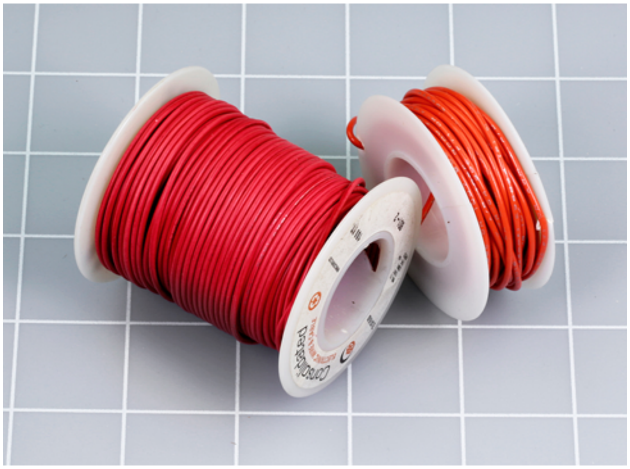

Figure 2-12 *Hookup wire is available on spools holding 25 feet and 100 feet, shown here.*

Wire is cheaper, per foot, if you buy a 100-foot spool, but I suggest you should buy smaller quantities in at least three different colors of insulation. The reason is that wire colors are helpful when you are trying to find an error in a circuit that you have built. Red and blue can be used for connections with positive and negative power, and one other color for other connections.

When the insulation is stripped away, it reveals a solid conductor, as shown in Figure 2-13. Compare this with stranded wire, shown in Figure 2-14. Stranded wire does have some uses, which I will talk about in a moment, but if you try to push it into holes in a breadboard, you are likely to become frustrated quite quickly. You really need solid wire.

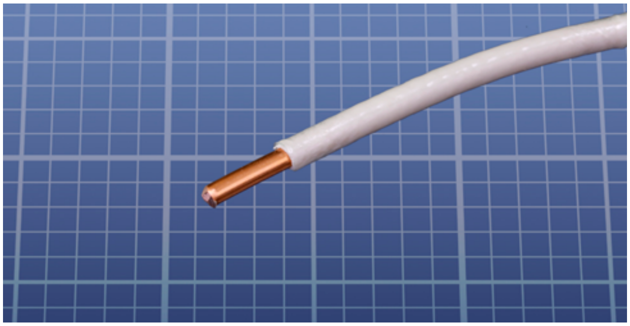

Figure 2-13 *Inside the plastic insulation, there should be a single conductor of solid wire.*

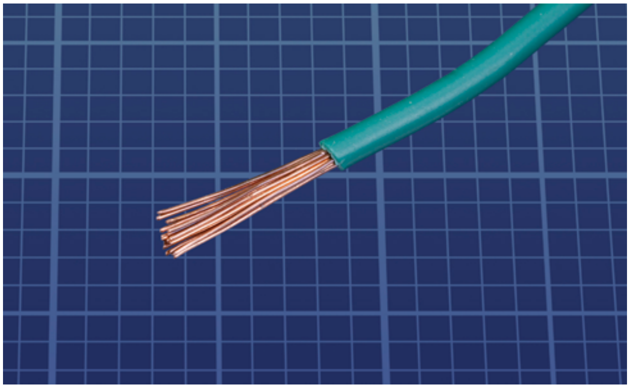

Figure 2-14 *For specific purposes (described in the text) stranded wire can be a useful option.*

Just as I emphasized the need for wire strippers that are specifically designed for 22-gauge wire, I will now emphasize that the wire should be 22-gauge, not 20-gauge, and not 24-gauge. You will find that the 24-gauge wire doesn't fit tightly in a breadboard, and may not make a reliable connection, while 20-gauge wire is just a little too thick, so that when you try to push it in, it can bend instead of sliding into place—and if you do manage to insert it, pulling it out can be difficult.

Some copper wire has a silver-colored coating, when you strip the insulation away. This is said to be "tinned." Other wire is plain copper, and I don't have any opinion about which type is better.

How much will you need? For assembling the circuits in this book, 25 feet of each color will be more than enough. However, Experiments 26, 28, and 29 require you to make coils of wire, to explore the relationship between electricity and magnetism, and to build your own crystal-set radio. If you want to pursue these projects (which I think are worthwhile), you'll need 200 feet of wire. This is your choice, as none of the kits for this book contains that much wire. See "Supplies" on page 316 for information about buying wire.

Jumpers

If you cut a section of wire, strip at least 1/4" but no more than 3/8" insulation from each end, bend the ends down, and push them into holes in a breadboard, you have just made a *jumper*, which creates a connection by jumping across some intervening holes in the board. Jumpers of this type result in a neat circuit where you can find errors relatively easily.

The trouble is, stripping insulation and making right-angle bends is tedious, even with the right tools for the job. Therefore you may be inclined to purchase *precut wire* which has been made into jumpers for you. An assortment looks like the one shown in Figure 2-15. For guidance about finding a selection like this, see "Supplies" on page 316.

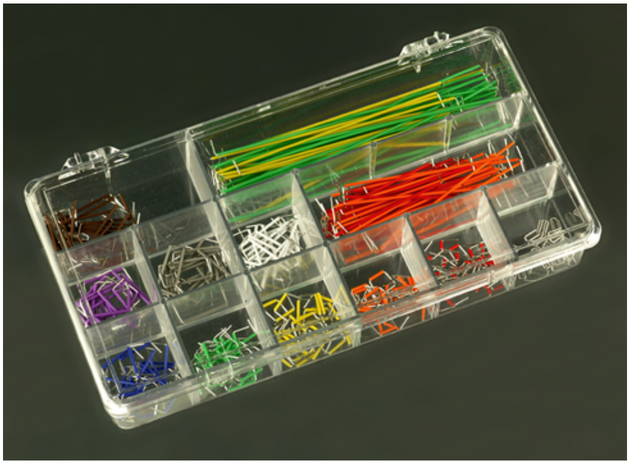

Figure 2-15 *An assortment of precut wire, stripped and bent for use in a breadboard.*

I used to use precut wire myself, but gave up on it because the wire segments are colored according to their length rather than their function. Red wires are all 0.2" long, yellow wires are all 0.3" long, and so on.

I want wires to be colored according to what they do in a circuit. Thus, red wires should always be connected to the positive side of the power supply, regardless of how long they are.

The only way to achieve this is to cut your own, which is what I do. You are welcome to use the precut wire if you so choose—but in addition to being confusingly colored, it costs more.

There is one more issue I have to clarify regarding jumpers. Many people like to use a different type of jumper wire that has a little plug at each end, just the right size to push into holes in a breadboard. These "plugged jumpers" are sold in bundles, and are probably the first thing you will find if you search for jumper wires online.

Because they are flexible, and they are usually about three inches long, you can use them to make almost any connection you are likely to need on a breadboarded circuit. They are reusable, and appear to be the simplest, quickest, cheapest option.

So far, so good; but if you make a mistake, you are going to have a problem finding it. Figure 2-16 shows a small circuit for an application outside of this book, using flexible jumpers that have plugs on each end. Figure 2-17 shows exactly the same circuit with hand-cut jumper connections using solid-core 22-gauge wire. Each of these circuits contains one wiring error. On the one using hand-cut wire, I can see it within a few seconds. On the one using flexible jumpers, I would have to dig around for quite a while, probably using a meter to search for the fault.

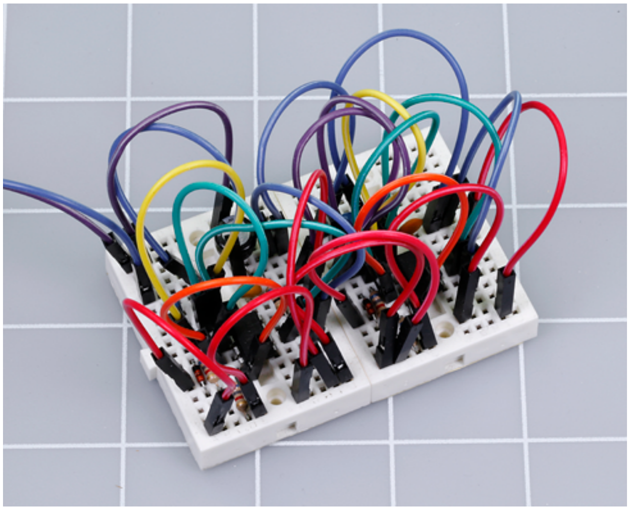

Figure 2-16 *A circuit on two mini-breadboards using flexible jumper wires with a plug at each end.*

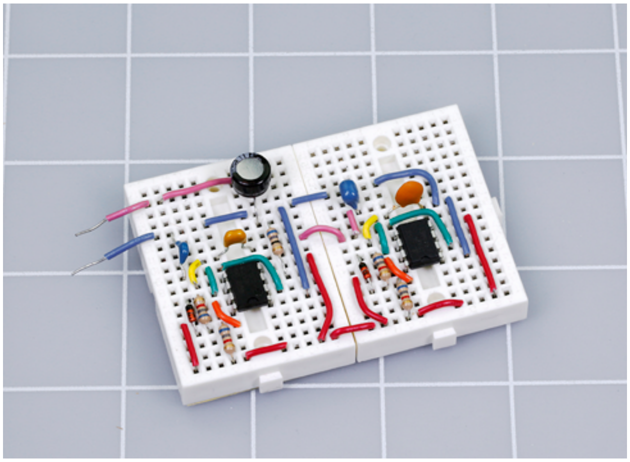

Figure 2-17 *The same circuit as in the previous figure, using hand-cut solid jumper wires.*

To make matters worse, the plugs on flexible jumpers are sometimes defective, and may contain loose connections. This can make fault-tracing almost impossible. Therefore:

- I don't recommend flexible jumpers that have plugs on each end.

Optional: Stranded Wire

Getting back to stranded hookup wire, it does have one advantage. It is much more flexible than solid wire, which is useful if you are running it out from a circuit board to a switch or a potentiometer. Flexibility can be essential if the wire is making a connection with an object that moves or vibrates.

While flexible wire is not essential for any of the projects in this book, 25 feet of 22-gauge stranded will come in handy occasionally. If you buy some, I suggest you choose a color that is different from your colors of solid wire, so that you won't confuse it with them.

Components

Again I must remind you that component kits for the projects in this book are available. See "Kits" on page 311. If you prefer to buy your own components from online sources, see "Components" on page 317.

Essential: Toggle Switch

A full-size *toggle switch* is an old-fashioned device, but useful for your switching experiments. You will need two. They should be described as SPDT, meaning single-pole double-throw. I'll explain this in detail a little later. Double-pole double-throw, described as DPDT, can also be used, but may cost slightly more.

A toggle switch with *screw terminals* will reduce the inconvenience of attaching it to hookup wire, but other types of terminals are acceptable.

A typical full-size toggle switch is shown in Figure 2-18. E-Switch ST16DD00 is an example, but cheaper generic alternatives can be found on eBay.

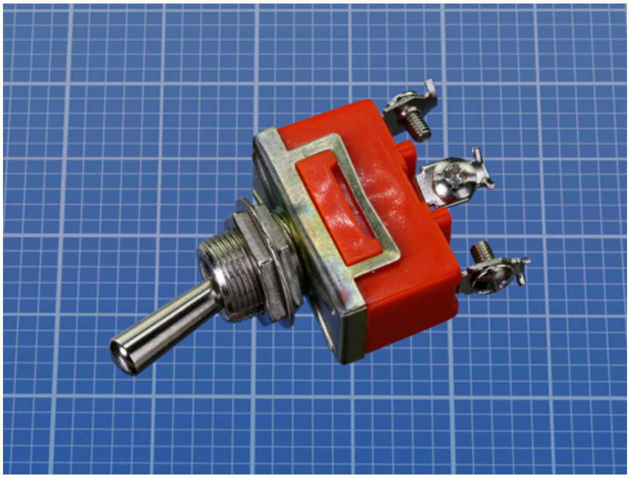

Figure 2-18 *A full-size toggle switch.*

Essential: Tactile Switch

Confusingly, a tactile switch is not what you would think of as a switch. It is a very tiny pushbutton. If you plug it into your breadboard, it provides a convenient way for a circuit to receive user input.

The most commonly used tactile switches have four little legs to push into a board, and can be annoying, because the legs often don't engage properly. The component is liable to spring out like a baby grasshopper at unexpected moments. I suggest you use a tactile switch with two pins spaced 0.2" apart. The Alps SKRGAFD-010 will be used in projects throughout this book, and is shown in Figure 2-19. Any other tactile switch with two pins spaced 0.2" apart can be substituted, such as the Panasonic EVQ-11 series.

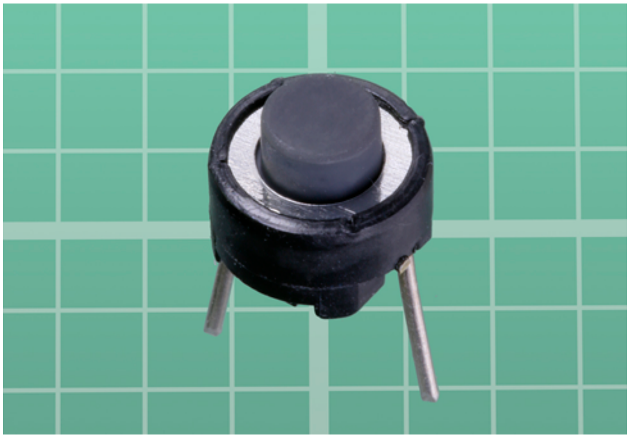

Figure 2-19 *The recommended tactile switch for breadboarded projects in this book.*

Essential: Relay

Because pin functions are not standardized among manufacturers, you have to be cautious about making substitutions when you buy a *relay*. I am recommending the Omron G5V-2-H1-DC9, shown in Figure 2-20, which should minimize confusion, because it has its pin functions printed on it.

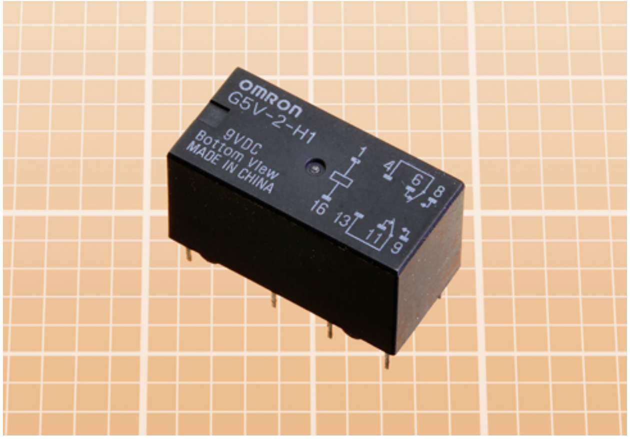

Figure 2-20 *The relay recommended for use with this book.*

Omron is a large manufacturer of relays, so I have some hope that the one I am recommending will be around for a while. You may also use the Axicom V23105-A5006-A201 or the Fujitsu RY-9W-K. All of them are are 9VDC DPDT relays, with pins spaced as shown on the left side of Figure 2-21. If the spacing is shown in millimeters, 5mm or 5.08mm are acceptable substitutes for 0.2", and 7.5mm or 7.62mm can substitute for 0.3".

If a diagram is printed on the relay, it should look like the one on the right side of Figure 2-21. Datasheets for relays almost always contain this information. You can use relays with different pin functions, but they will cause you some inconvenience, because they won't match the schematics that I will be providing.

The relays I have recommended are high-sensitivity type, meaning that they consume less current. You can substitute others, but they will draw more current. Whichever type of relay you use, it must have the same 9VDC coil voltage and pin spacing.

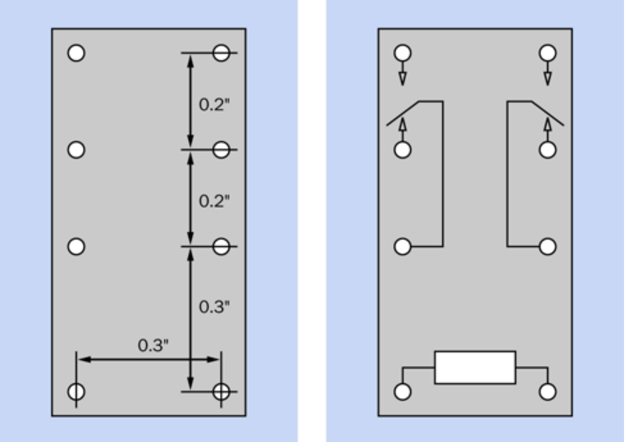

Figure 2-21 *Pin spacing and internal connections in a relay should be as shown.*

One thing to watch out for, when buying relays, is their polarity, meaning a requirement to apply current in one specified direction, because the relay won't work when current flows through its coil in the other direction. I have recommended relays that do not require polarity. Many Panasonic relays do, so read the datasheet carefully before buying one of them.

Lastly, any relay that you buy must be of the *nonlatching* type.

If this seems confusing and overly technical, you can put off buying the relay until you read Experiment 7, which describes how to use it. You will need two relays to carry out that experiment fully.

Essential: Trimmer Potentiometer

Instead of the big clunky potentiometer that you used in Experiment 4, you'll be using a *trimmer potentiometer*, which is smaller, cheaper, and plugs into a breadboard.

Examples (with various arbitrary values) are shown in Figure 2-22.

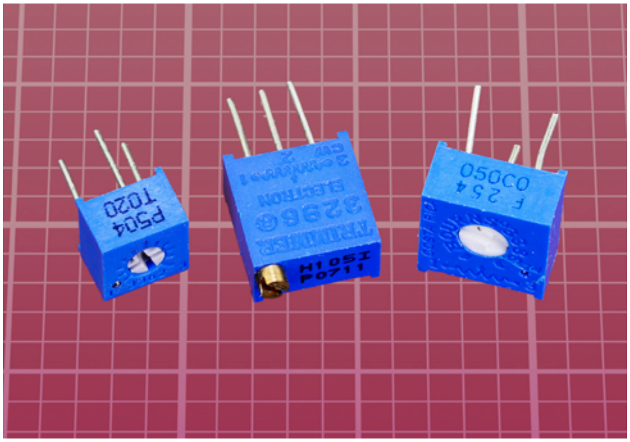

Figure 2-22 *Trimmer potentiometers.*

The trimmers at left and at right in this photograph are the type that I have chosen to use in this book. They will sit flush with a breadboard when their leads are pushed down into it. The only difference between these two samples is that one is slightly larger than the other. They are also available in variants that stand up at an angle of 90 degrees with the breadboard, but these are less accessible.

The one in the center of the photograph is a *multi-turn trimmer*, which allows much finer adjustment via a brass screw connected with a worm gear inside the component. This is less convenient, more expensive, and is not necessary for our purposes, as you won't require that degree of accuracy.

Essential: Transistors

Only one type of *transistor* is used in this book. The generic part number is 2N2222, but unfortunately, not all 2N2222 transistors are alike.

If you are using a kit, you should have no problems. If you do your own shopping, you absolutely must avoid any part that has P2N preceding the 2222 number. When the P2N2222 was introduced, the manufacturers reversed the pin functions of the 2N2222 that had been standardized for decades. (Why would they do this? I do not know.)

Here's the rule.

- Part numbers 2N2222 or PN2222 or PN2222A are okay. PN2222 has become a more common designation than 2N2222, but either will work.
- Part numbers P2N2222 or P2N2222A are *not* okay.

The trap is that if you are searching for 2N2222, you will be offered a P2N2222 because the search engine will help you out by showing parts that have extra letters preceding the number. So—shop with care! And if you have a meter that tests transistors, check each one. If the transistor has the traditional pin functions, the meter should tell you that it has an amplification ratio exceeding 200. If you have the wrong type of transistor, your meter will show an error or an amplification value lower than 50.

2N2222 transistors used to be packaged in tiny metal cans. These days, they are almost always packaged in black plastic. Examples of both types are shown in Figure 2-23. Plastic and metal packages both work equally well—so long as the transistor's part number does not begin with P2N.

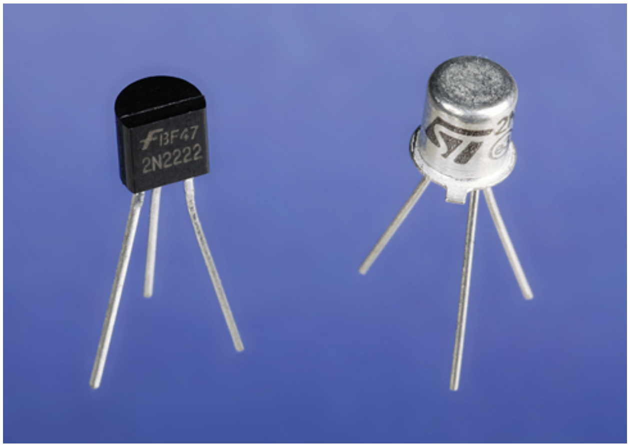

Figure 2-23 *Two 2N2222 transistors. Either can be used.*

Essential: Capacitors

Capacitors are not quite as cheap as resistors, but still cheap enough for you to consider buying an assortment of small ones in bulk. Capacitor values in the range that we will be using are most often measured in microfarads, abbreviated μF. I'll explain this in detail when you start using capacitors in your circuits.

For small values, *ceramic* capacitors are recommended. For larger values, *electrolytics* are cheaper. For additional guidance about purchasing them, see "Components" on page 317. Various capacitors are shown in Figure 2-24. The cylindrical ones are electrolytic, while the others are ceramic.

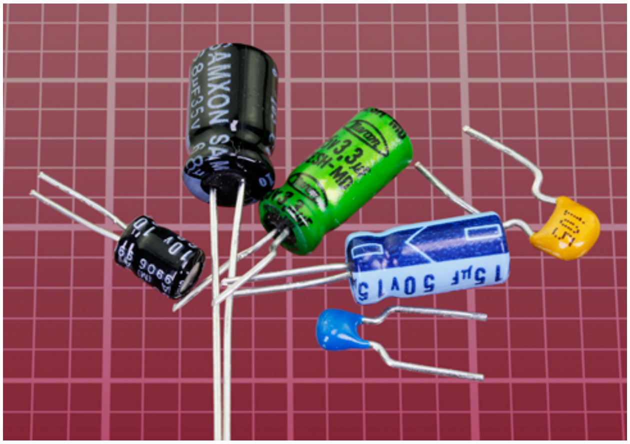

Figure 2-24 *A variety of capacitors.*

Essential: Resistors

If you are shopping for your own components, I am assuming that you bought a good selection of resistors as I suggested for Experiment 1.

Essential: Loudspeaker

Minimum size for a *loudspeaker* is 1" diameter, but 2" diameter is good. A 3" diameter is maximum. The impedance should be 8 ohms or higher. When searching for a loudspeaker, bear in mind that some suppliers call them "speakers." This may cause a search for "loudspeaker" to yield no results.

We will not be dealing with high-fidelity sound, so any cheap loudspeaker will do. A couple of samples are shown in Figure 2-25.

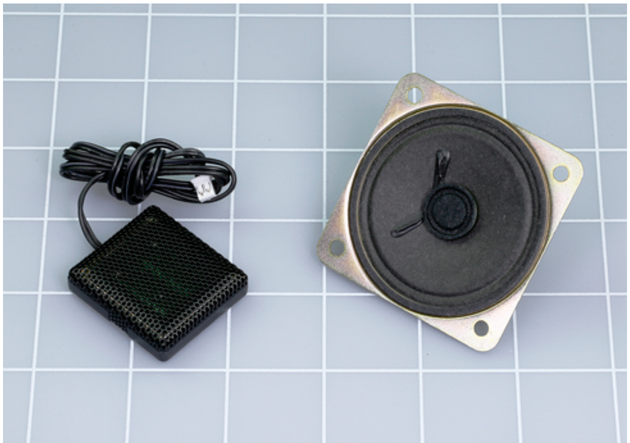

Figure 2-25 *Two speakers, one measuring 1" diameter, the other measuring 2".*

And More?

By now you may be thinking that I have specified quite a lot of components. Rest assured that almost everything that I have listed here will be reusable, and you will not need many additional parts for the remaining chapters of the book.

Experiment 6: Very Simple Switching

This experiment will acquaint you with the function of manually operated switches. You may feel that you already know how to use a switch, but when two double-throw switches are combined in a circuit, the subject becomes a little more interesting.

What You Will Need

- Screwdriver, wire cutters, wire strippers
- Hookup wire, 22-gauge, no more than 12"
- 9-volt battery (1)
- Generic LED (1)
- Toggle switches, SPDT or DPDT (2)
- 470-ohm resistor (1)
- Test leads with alligator clips at each end (2)

Assemble the parts as shown in Figure 2-26. You'll need to practice your abilities at wire stripping, here, to bare the ends of the two sections of black wire. To secure them in the screw terminals of the switches, try using your pliers to curl the end of each wire so that it looks like a letter J. Then feed it under the screw from the left, so that the screw will draw in the wire when tightened in a clockwise direction.

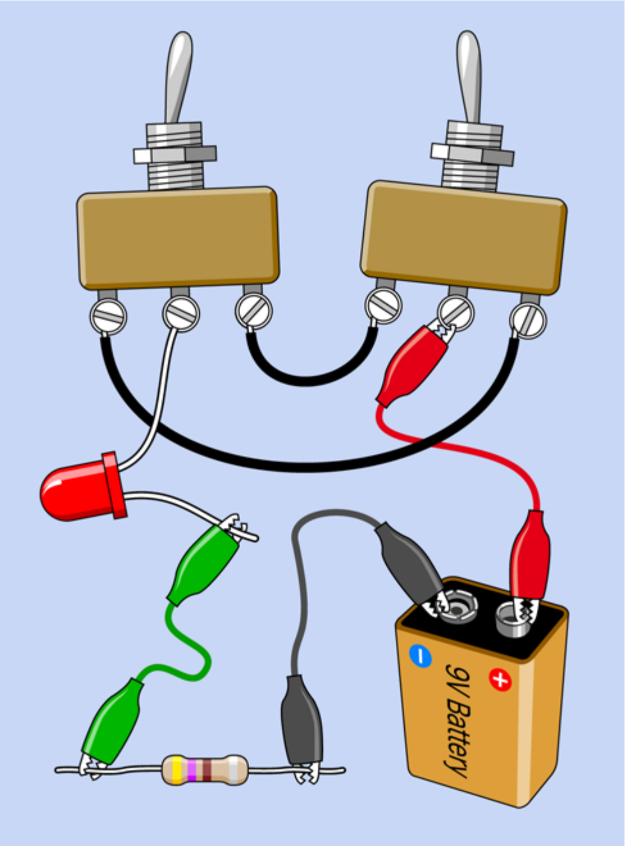

Figure 2-26 *Your very first experiment in switch connections.*

The long lead on the LED also fits one of the screw terminals. Don't use the short lead of the LED by mistake. Remember, the long lead must always be more positive than the short lead.

If your switches don't terminate in screw terminals, you will have to use a couple of alligator test leads instead of the black wires, and one more test lead to connect the LED with the center terminal of the lefthand switch.

After you connect the battery, experiment, flipping the switches. What do you find?

If the LED is on, flipping either of the switches will turn it off. If the LED is off, either of the switches will turn it on. I'll explain this interesting behavior shortly (see "Introducing Schematics" on page 55), but I have to tell you some fundamentals and background information first.

Fundamentals: All About Switches

The *toggle* in a toggle switch is the part that you flip with your finger. In the type of switch shown in Figure 2-26, flipping the toggle connects the center terminal with one of the terminals on either side of it, as shown in Figure 2-27.

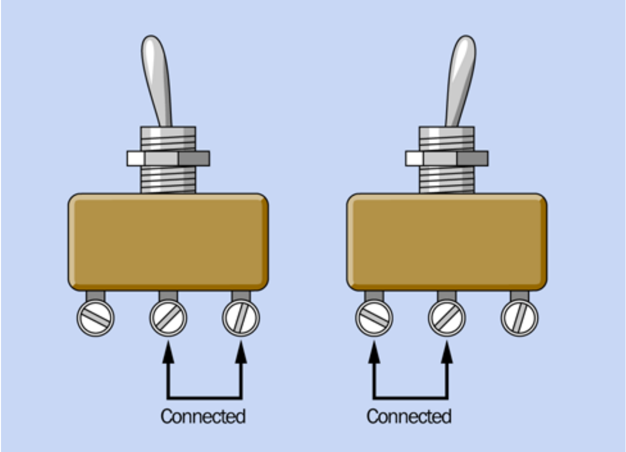

Figure 2-27 *Usually, but not absolutely always, toggle switches work like this.*

The center terminal is called the *pole* of the switch. Because you can flip, or throw, this switch to make two possible connections, it is called a *double-throw* switch, abbreviated DT (or, sometimes, 2T). A single-pole, double-throw switch is abbreviated *SPDT* (or, sometimes, 1P2T).

Some switches only have two terminals instead of three. They are on/off, meaning that if you throw them in one direction they make a contact, but in the other direction they make no contact at all. Most of the light switches in your house are like this. They are known as *single-throw* switches. A single-pole, single-throw switch is abbreviated *SPST* (or, sometimes, 1P1T).

Some switches have two entirely separate poles, so you can make two separate connections simultaneously when you flip the switch. These are called *double-pole* switches, abbreviated DP (or, sometimes, 2P). Check Figure 2-28, Figure 2-29, and Figure 2-30 for photographs of old-fashioned “knife” switches, which are still used sometimes to teach electronics to kids in school. You wouldn’t use a switch like this for any practical purpose, but they illustrate very clearly the differences between SPST, SPDT, and DPST connections.

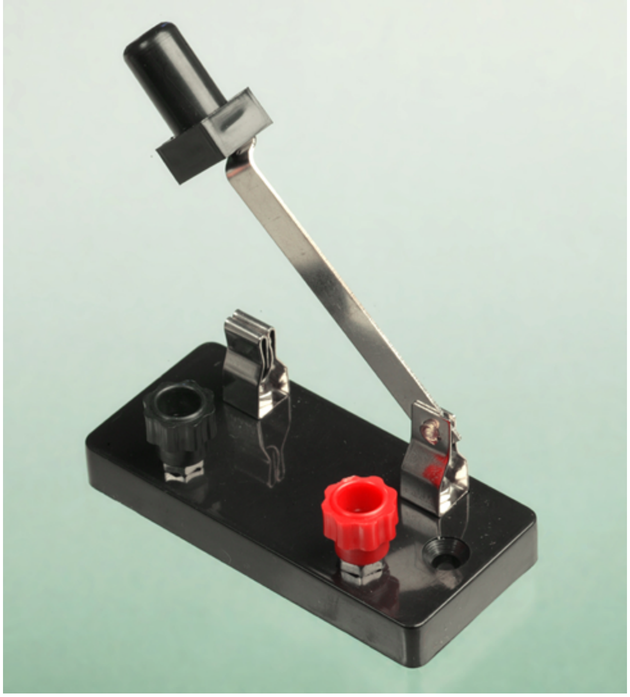

Figure 2-28 *Manufactured for educational use, this is a single-pole, single-throw (SPST) switch.*

The only place you’re likely to see a knife switch being used for serious purposes is in a horror movie. In Figure 2-31, a mad scientist is powering up his experiment with a single-pole, double-throw knife switch, conveniently mounted on the wall of his basement laboratory.

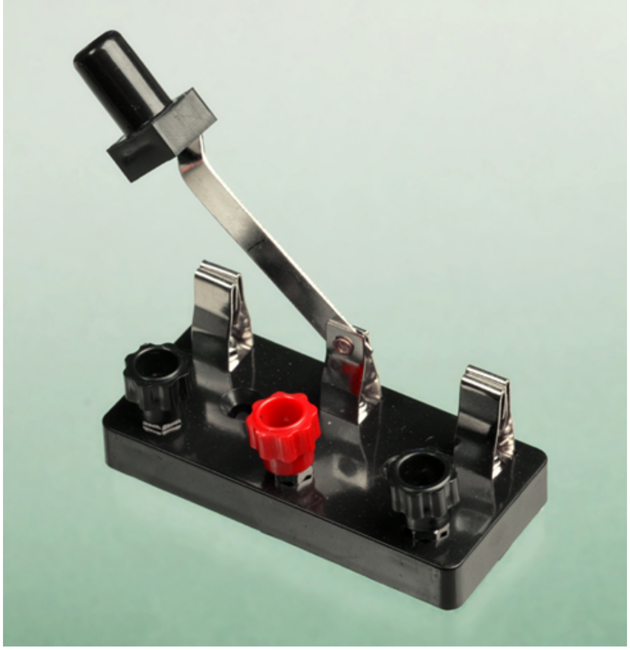

Figure 2-29 *A single-pole, double-throw (SPDT) switch connects one pole with a choice of contacts.*

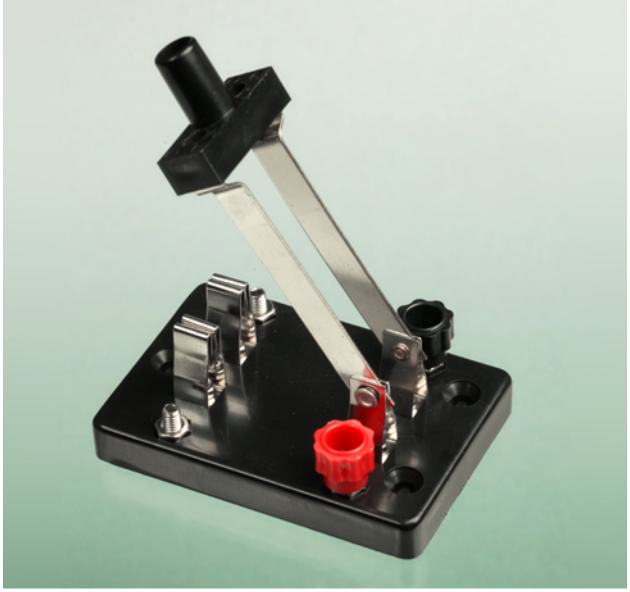

Figure 2-30 *A double-pole, single-throw (DPST) switch has two poles that are completely isolated from each other. Each pole can connect with only one contact.*

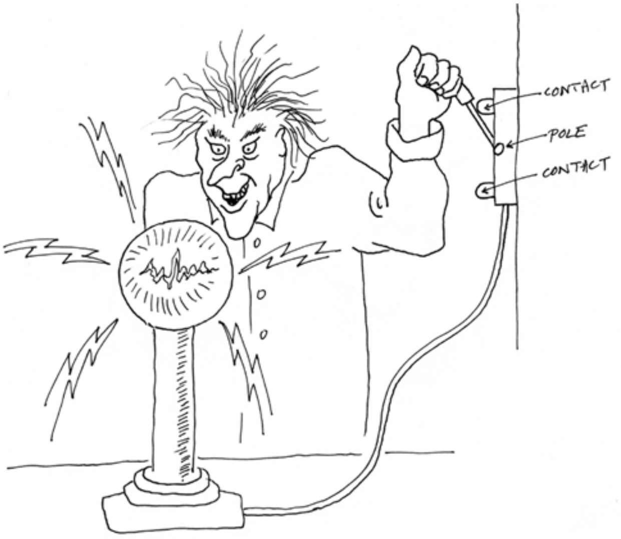

Figure 2-31 *Left: Mad scientist. Right: SPDT knife switch.*

To make things more interesting, you can buy switches that have three or four poles. (Some rotary switches have even more, but we won't be using them.) Also, some double-throw toggle switches have an additional "center off" position.

Putting all this together, I made a table showing some possible types of switches and the abbreviations that describe them. Pushbuttons use the same abbreviations. See Figure 2-32. If you're reading a parts catalog, you can check this table to remind yourself what the abbreviations mean.

	Single Pole	Double Pole	Three Pole	Four Pole
Single Throw	SPST (or 1P1T) on-off	DPST (or 2P1T) on-off	3PST (or 3P1T) on-off	4PST (or 4P1T) on-off
Double Throw	SPDT (or 1P2T) on-on	DPDT (or 2P2T) on-on	3PDT (or 3P2T) on-on	4PDT (or 4P2T) on-on
Double Throw with Center Off	SPDT (or 1P2T) on-off-on	DPDT (or 2P2T) on-off-on	3PDT (or 3P2T) on-off-on	4PDT (or 4P2T) on-off-on

Figure 2-32 *This table summarizes various options for toggle switches (and pushbuttons).*

Some switches are spring-loaded, so that they snap back to a default position when you release pressure on them. When you see ON or OFF in parentheses, you know that you have to maintain pressure on the switch to keep it in that position.

Here are some examples:

- OFF-(ON): Because the ON state is in parentheses, it's the momentary state. Therefore, this is a single-pole switch that makes contact only when you push it, and flips back to make no contact when you let it go. It is also known as a "normally open" momentary switch, abbreviated "NO." Most pushbuttons also work this way.
- ON-(OFF): The opposite kind of momentary switch. It's normally ON, but when you push it, you break the connection. So, the OFF state is momentary. It is known as a "normally closed" momentary switch, abbreviated "NC."
- (ON)-OFF-(ON): This switch has a center-off position. When you push it either way, it makes a momentary contact, and returns to the center when you let it go.

Other variations are possible, such as ON-OFF-(ON) or ON-(ON). As long as you remember that parentheses indicate the momentary state, you should be able to figure out how these switches behave.

Sparking

When you make and break an electrical connection, it tends to create a spark. Sparking is bad for switch contacts. It erodes them until the switch doesn't make a reliable connection anymore. For this reason, you must use a switch that is appropriate for the voltage and amperage that you are dealing with.

The electronic circuits in this book are low-current and low-voltage, so you can use almost any switch; but if you are switching a motor, it will tend to suck an initial surge of current that is at least double the rating of the motor when it is running constantly. For instance, you should probably use a 4-amp switch to turn a 2-amp motor on and off.

Checking Continuity

You can use your meter to check a switch. Doing this enables you to find out which contacts are connected when you turn a switch one way or the other. It's also useful if you have a pushbutton and you can't remember whether it's the type that is normally open (you press it to make a connection) or normally closed (you press it to break the connection).

It's convenient to set your meter to measure "continuity" when checking a switch. The meter will beep (or show a visual indication) if it finds a connection, and will do nothing if it doesn't. See Figure 2-33, Figure 2-34, and Figure 2-35 for examples of meters that are set to measure continuity. Remember that in Experiment 1, I showed you the symbol that is used on meters to represent continuity. See Figure 1-7.

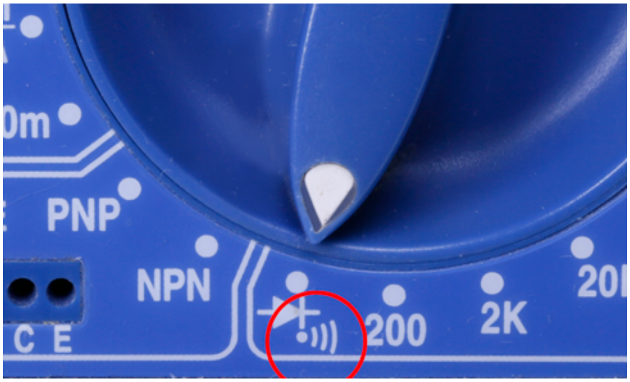

Figure 2-33 *A meter dial that has been turned to measure continuity.*

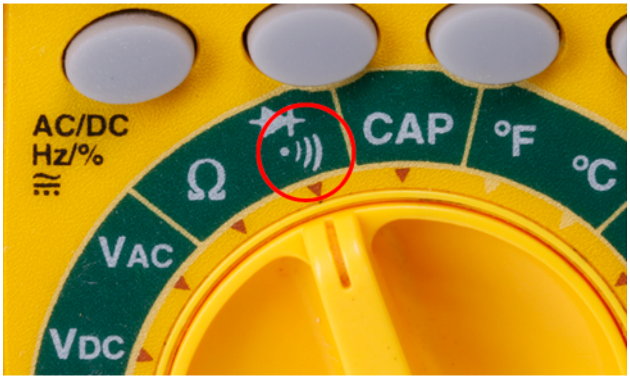

Figure 2-34 *Another meter dial set to measure continuity.*

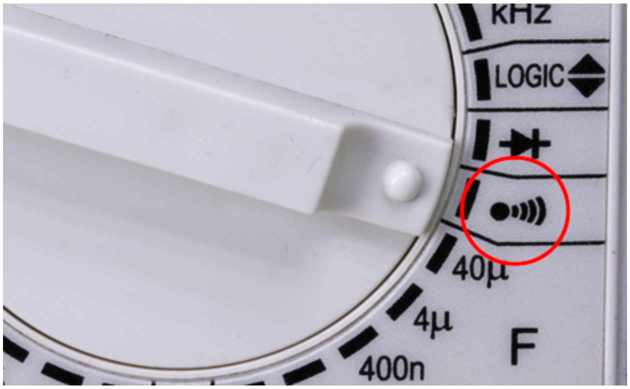

Figure 2-35 *A third meter measuring continuity.*

Background: Early Switching Systems

Switches seem to be such a fundamental feature of our world, and their concept is so simple that we can easily forget that they went through a gradual process of evolution. Primitive knife switches were quite adequate for pioneers of electricity who simply wanted to connect and disconnect some apparatus in a laboratory, but a more sophisticated approach was needed when telephone systems began to proliferate. Typically, an operator at a "switchboard" needed to connect any pair of 10,000 lines on the board. How could it be done?

In 1878, Charles E. Scribner (shown in Figure 2-36) developed the "jack-knife switch," so called because the part of it that the operator held looked like the handle of a jack knife. Protruding from it was a plug, and when the plug was pushed into a socket, it made contact inside the socket. The socket, in fact, contained the switch contacts.

Figure 2-36 *Charles E. Scribner invented the "jack-knife switch" to satisfy the switching needs of telephone systems in the late 1800s. Today's audio jacks still work on the same principle.*

Audio connectors on guitars and amplifiers still work on the same principle, and when we speak of them as being "jacks," the term dates back to Scribner's invention. Switch contacts are still mounted inside a jack socket.

Today, of course, telephone switchboards have become as rare as telephone operators. First they were replaced with relays—electrically operated switches, which I'll talk about later in this chapter. And then the relays were superceded by transistors, which made everything happen without any moving parts. In Experiment 10, you'll be switching current with transistors.

Introducing Schematics

In Figure 2-37, I've redrawn the circuit from Figure 2-26 in a simplified style known as a "schematic." From this point onward, I will be illustrating circuits with schematics, because they make the connections easier to understand. You only need to know a few symbols to interpret them.

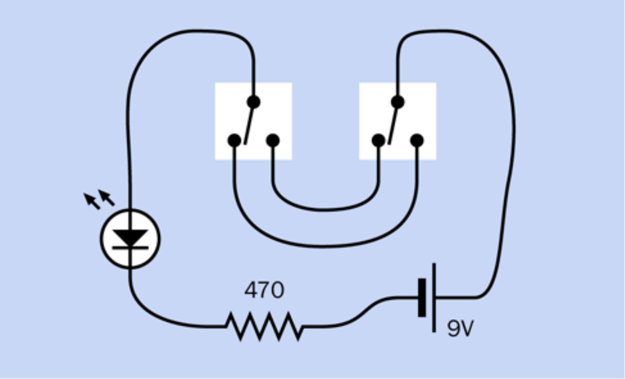

Figure 2-37 *The circuit with two switches is redrawn here as a schematic.*

Figure 2-26 and Figure 2-37 both show the same components and connections between them. In the schematic, the zigzag thing is the resistor, the symbol with two diagonal arrows is the LED, and the battery is shown as two parallel lines of unequal length.

The large triangle in the LED symbol indicates the direction of flow of *conventional current*, meaning the imaginary flow that runs from positive to negative. The pair of diagonal arrows tells you that this is the type of diode that is *light-emitting* (I'll get to other kinds of diodes later). In the battery symbol, the longer of the two lines identifies the positive side.

Trace the path that electricity can take through the circuit and imagine the switches turning one way or the other. You should see clearly, now, why either switch will reverse the state of the LED from on to off or off to on.

Figure 2-38 shows the same schematic cleaned up a bit. The lines are straight, and the power supply is now shown with the positive side at top-left, and the negative side at bottom-right. You tend to see conventional current moving from top to bottom, in schematics, while signals of some kind (such as the audio input to an amplifier) move from left to right. The top-down organization of a circuit makes it easier to understand.

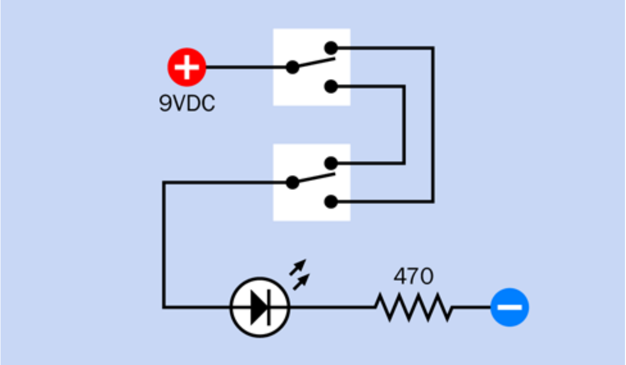

Figure 2-38 *The previous schematic has been reorganized along more conventional lines.*

The important thing to grasp is that the two schematics show exactly the same circuit, even though they look different. The type of components, and the way in which they are connected, are all that matters. The exact locations of the components are irrelevant.

- A schematic doesn't tell you where to put the components. It just tells you how to join them together.

Incidentally, you probably have an example of the circuit in Figure 2-38 in your home, especially where two light switches are placed at the top and at the bottom of a flight of stairs, and either switch can be used to turn a light on and off. This is shown in Figure 2-39, where the live and neutral wires of the AC supply enter the picture at bottom-left. The live wire is switched, while the neutral wire runs alongside it to the lightbulb (the white circle with a curly line representing the element of an old-fashioned incandescent bulb).

The only problem with schematics is that some symbols are not standardized. You may see several variants that all mean the same thing. I will be explaining them as we go along.

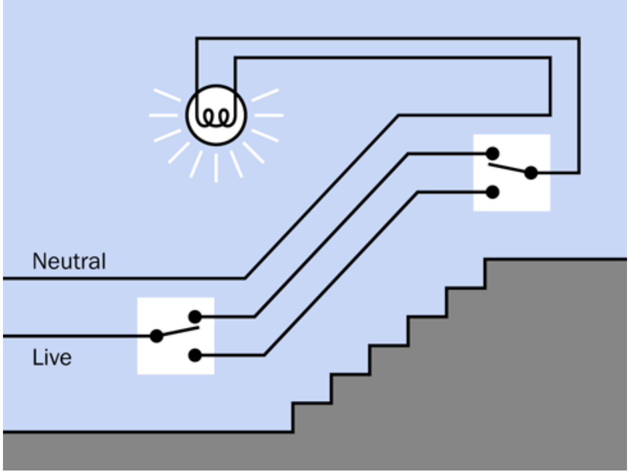

Figure 2-39 *The same circuit as shown in the previous figure is used in homes where two switches control a single light.*

Fundamentals: Basic Schematic Symbols

1. The Switch. Figure 2-40 shows five variations of that most basic component, a single-pole, single-throw switch. In each case, the pole happens to be on the right, while the contact is on the left—although with a SPST switch, this does not make a significant difference. In this book I have chosen to include a white rectangle around each switch, to emphasize that although it has two parts, they constitute one component.

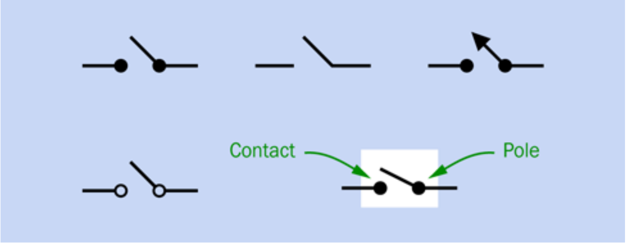

Figure 2-40 *Five variations in the schematic symbol for a SPST switch. They are all functionally identical.*

Figure 2-41 shows how things get a little more complicated when you have a double-throw, double-pole switch. A dashed line indicates that both segments of the switch move together when the switch is turned, even though each pole and set of contacts is electrically isolated from the other. The variation at center is sometimes found in large schematics where the layout makes it difficult to put the sections of a switch close to each other. Each set of contacts is identified by an abbreviation ending in A, B, C . . . and you are expected to understand that the contacts are actually contained in one switch.

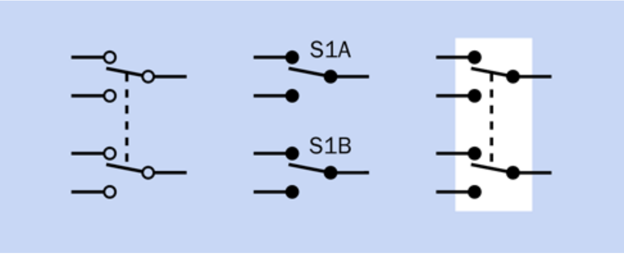

Figure 2-41 *Three variations on a DPDT switch symbol.*

2. Power Supply. The power supply for a DC (direct-current) circuit can be indicated in several ways. The top section of Figure 2-42 shows symbols for a battery. A short line indicates the negative end, while a longer line indicates the positive end. Traditionally, a single pair of lines represented a single 1.5-volt cell, two pairs indicated a 3-volt cell, and so on. But when circuits used high voltage with vacuum tubes, the person drawing the schematic would usually show a dashed line between cells instead of drawing dozens of them in a row.

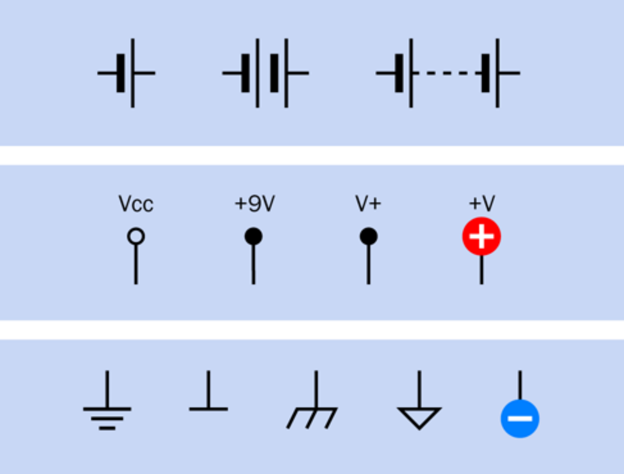

Figure 2-42 *Various ways to show positive and negative DC power in a circuit.*

A battery symbol may still be used in a simple schematic, but more often positive and negative DC power is indicated by separate symbols, shown in the center section and the bottom section of Figure 2-42. Positive is applied in one location of a circuit labelled Vcc, V_{CC}, or V+, or +V, or +V with a number added to indicate the voltage. Originally the term V_C referred to the voltage at the collector of a transistor. V_{CC} meant the supply voltage for the whole circuit, and is now used regardless of whether the circuit has any transistors in it. Many people say "vee cee cee" without knowing its derivation.

In this book, because I have the luxury of full-color reproduction, I have used a plus symbol in a red circle to indicate positive power input.

The negative side of the power supply can be shown with any of the symbols in the bottom section of Figure 2-42. It may be referred to as "negative ground" or simply "ground." Because many parts of a circuit may share a negative potential, multiple ground symbols may be found scattered around a schematic. This is more convenient than drawing lines linking them all together.

In this book, I have chosen to use the minus sign in a blue circle because it is so intuitively obvious. You will not see it used often in schematics generally.

My discussion so far has referred to battery-powered devices. In a gadget that uses AC power from a wall outlet, the situation is more complicated, because the outlet has three sockets in it for live, neutral, and ground connections. A schematic typically shows the AC source as an S shape turned on its side, as in Figure 2-43. Often the value of the power supply is shown, and in the US it is usually 110, 115, or 120 volts. Elsewhere in the circuit, the symbols shown on the righthand side of Figure 2-43 refer to the chassis of the device in which the electronics are mounted.

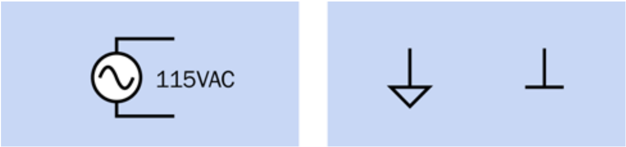

Figure 2-43 *Symbolic representation of AC power (left) and chassis in an AC device (right).*

Note that the ground pin in an AC power outlet in the home really does connect with the ground outside the building. An electronic device with a metal chassis that connects with that pin is "grounded." In a battery-powered circuit without any high voltages, grounding "in the ground" is unnecessary, but a ground symbol may still be used.

In the UK, a grounded device is sometimes referred to as being "earthed."

3. Resistor. Only two variants of the symbol for a resistor exist, shown in Figure 2-44. The symbol on the left is used in the United States, with a number beside it indicating its resistance in ohms. Alternatively the resistor may be identified as R1, R2, R3 . . . with a separate parts list showing the values. The righthand symbol in Figure 2-44 originated in Europe, and here again the number is the value of the resistor in ohms. The value of 220 ohms in the figure was selected arbitrarily.

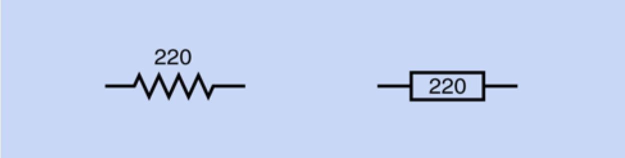

Figure 2-44 *Symbol for a resistor in the United States (left) and Europe (right).*

Remember that where a resistance value includes a decimal point, Europeans omit the point and substitute a K or M, while resistances less than 1K are shown as one or more digits followed by letter R.

4. Potentiometer. In Figure 2-45 the lefthand symbol is used in the United States, while the righthand symbol originated in Europe. In both symbols, the arrow identifies the wiper in the potentiometer. The value of 470 ohms was selected arbitrarily.

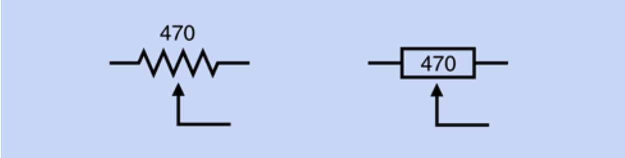

Figure 2-45 *The lefthand symbol for a potentiometer is used in the United States, while the righthand symbol originated in Europe.*

5. Pushbutton. Three possible pushbutton symbols are shown in Figure 2-46. These symbols represent the most common type of normally open pushbutton or momentary switch, where pressure closes two contacts and release of the pressure opens the circuit. In more complex pushbuttons, where a single button press closes or opens numerous contacts, the symbol for a multi-pole switch may be used.

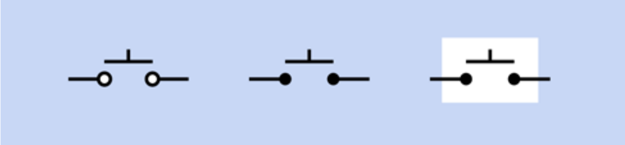

Figure 2-46 *Three variants of the pushbutton symbol. The white rectangle is added in this book for clarity, but is not used elsewhere.*

6. Light-Emitting Diode (LED). Figure 2-47 shows four variants of a symbol representing an LED. The meaning is the same, regardless of whether a circle is included, and regardless of whether the triangle is solid or open. The white highlight in the circle is added in this book for clarity, but is not used elsewhere. LED symbols may be oriented in any direction, for convenience in drawing a circuit. Arrows also may point in any direction.

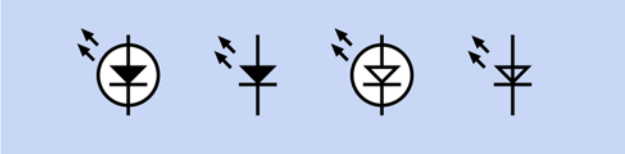

Figure 2-47 *Four ways of representing an LED. They are functionally identical.*

I'll explore other symbol variants later in the book. Meanwhile, the most important things to remember are:

- The positions of components in a schematic do not affect the functionality.
- The styles of symbols used in a schematic are not important.
- The connections between the components are extremely important.

Schematic Layout

I mentioned previously that schematics commonly show positive power at the top and negative at the bottom. This convention is a great help in understanding how the circuit works, but is not helpful at all when you actually want to build the circuit, because almost certainly you will begin by using a breadboard that imposes a completely different geometry.

Almost all the electronics books that I have seen expect you to change a circuit from the way it looks as a schematic to the way it has to be on a breadboard. This can be challenging, and may be a significant barrier to learning electronics. Therefore, the schematics in this book are all laid out in a pattern that is similar to that of a breadboard. This will make more sense after you encounter a breadboard yourself in Experiment 8.

Crossovers

The last topic I have to mention about schematics concerns the way in which they show two wires crossing each other. In the simple circuits that you have built so far, no crossovers occurred, but as circuits become more complicated, wires must pass over each other without making an electrical connection. How can a schematic illustrate this?

In the first edition of this book, I used a style in which one wire crossing another had a little semicircular "jump" in it. This is identified as the "Old Style" in Figure 2-48. I still prefer this style, because you can see so clearly that the wires are not making an electrical connection. However, a few decades ago, rendering the little jump became more troublesome as circuits were created with graphics software instead of pen and ink. At that point, jumps were used less frequently.

An alternative, identified as "Newer Style" in Figure 2-48, showed one of the wires with a break in it. This was confusing, and was not easily rendered by automated circuit-drawing software. Consequently it, too, has become rare.

The third style, labelled "Conventional," is now extremely common. In this edition of *Make: Electronics* I decided I should conform with the conventional style that is used in the rest of the world, even though I think it isn't as clear as the old style.

Perhaps you are wondering—if two lines crossing each other are not electrically connected, how do you draw them when they are connected? The answer is, you use a dot, and to avoid confusion, the dot should be large, not just a little pin-prick. The lower half of Figure 2-48 shows what I mean. This leads to the general rule:

- Two lines crossing each other do not indicate an electrical connection.
- Where lines intersect in a dot, there is an electrical connection.

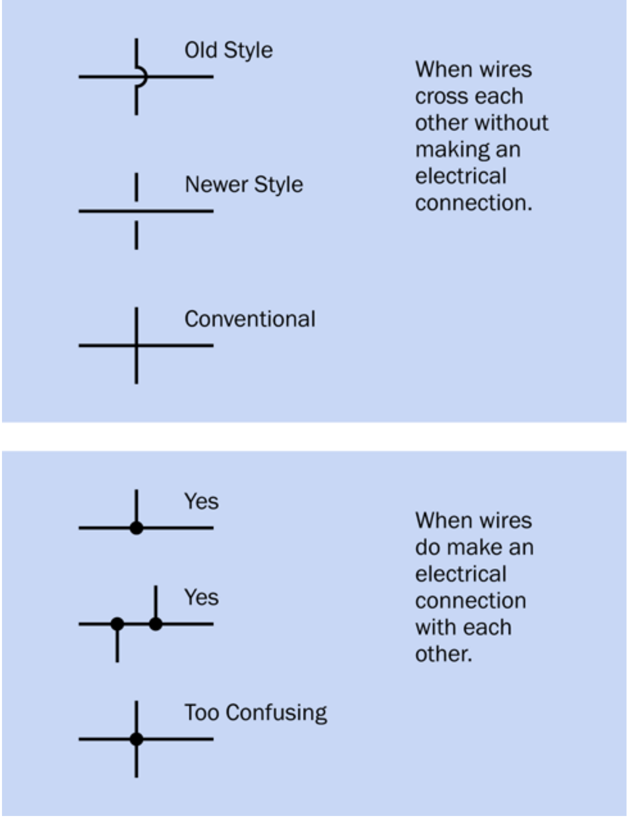
Old Style
Newer Style
Conventional
When wires cross each other without making an electrical connection.
Yes
Yes
Too Confusing
When wires do make an electrical connection with each other.

Figure 2-48 *Various styles for depicting wires that do or do not connect. See text for details.*

There's still one more note to add. In the interests of clarity, I think it's good practice to avoid the style shown at the bottom of Figure 2-48. It's too confusing. If we avoid that configuration, and use the one immediately above it instead, then we know that wires crossing each other never make a connection under any circumstances.

Colored Conductors

Did I say that crossovers would be my last topic about schematics? Actually there is one more small matter. Because I never want you to get confused between the positive and negative sides of a power supply, I'm going to be coloring all the positive conductors red in schematic drawings in the rest of the book, while the negative/ground side will be blue. Readers have told me that this was very helpful where I used it occasionally in the past, so, now I'm going to use it consistently throughout.

Black is a more commonly used color for negative/ground (as in the black wire on your meter, or the black wire from a battery connector). But blue is still sometimes used, and is more distinctive.

Just bear in mind that schematics that you encounter in the world outside of this book will not use this helpful coloring convention. All the wires will be black, and you will have to figure out which ones are connected with the power supply.

Experiment 7: Investigating a Relay

The next step in your exploration of switching is to use a remote-controlled switch. By "remote-controlled," I mean that it turns on or off in response to a signal that you send to it. This kind of switch is known as a *relay*, because it relays an instruction from one part of a circuit to another.

- Often a relay is controlled by a low voltage or small current, and switches a larger voltage or higher current.

This can be very useful. When you start your car, for instance, a relatively small, cheap ignition switch sends a small signal down a thin, inexpensive piece of wire to a relay that is near the starter motor. The relay activates the motor through a shorter, much thicker, more expensive piece of wire, capable of carrying as much as 100 amps.

Similarly, if you raise the lid on an old-fashioned top-loading washing machine during its spin cycle, you close a small switch that sends a small signal down a thin wire to a relay. The relay handles the bigger task of switching off the large motor spinning the drum full of wet clothes.

What You Will Need:

- 9-volt battery (1)
- DPDT 9VDC relays (2)
- Tactile switch, SPST (1)
- Test leads with alligator clips at each end (5)
- Utility knife (1)
- Multimeter (1)

The Relay

The type of relay that I want you to use has two pins at one end and six at the other. The six are clustered in two lines of three, as in Figure 2-49 (where the relay is upside-down with its pins in the air). If you buy two relays, you can use one for investigational purposes, meaning that you'll be cutting it open to take a look inside. If you do this very, very carefully, the relay should still be usable afterward. If not—well, you will have a spare.

Caution: Polarity Problems

Some relays are fussy about the way that you apply voltage to the coil that's hidden inside. Everything works fine with the electricity flowing one way through the coil, but if you reverse the positive and negative connections (in other words, if you reverse the polarity) the relay stops working.

This is especially annoying when the relay datasheet doesn't make it clear. The relays that I have recommended do not require a particular polarity. See "Essential: Relay" on page 48.

Procedure

Attach test leads and a tactile switch (pushbutton) as shown in Figure 2-49. (Note that the parts in these drawings are not drawn to scale.) When you press the button to apply 9 volts to the pair of pins that are separate from the others, you should hear a very faint click. Let go of the button, and you may hear another click. (If your hearing is not good, touch the relay gently with your fingertip, and you should feel a faint vibration when the click occurs.)

What's happening here? Your meter will help you to investigate. Set it to measure continuity, and verify that it's working by touching the two probes together. If it doesn't beep, you haven't set it to measure continuity, or the battery is dead, or one of the probes is plugged into the wrong socket.

Now hold the probes against the pins as shown in Figure 2-50, and press the button. The meter should beep while the button is pressed.

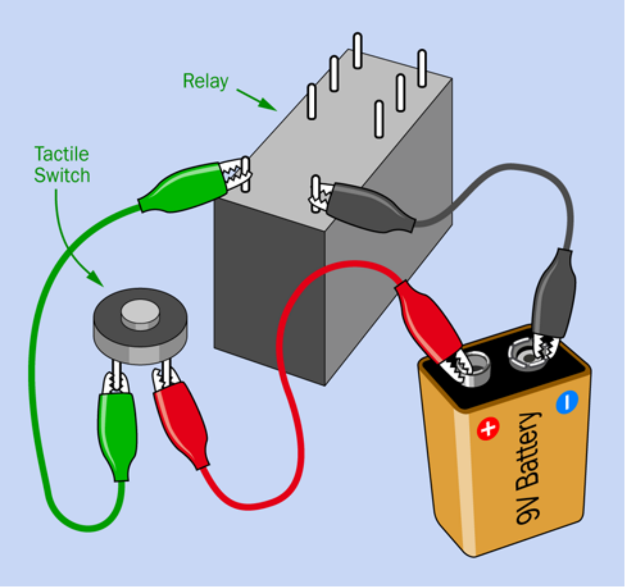
Relay
Tactile Switch
9V Battery

Figure 2-49 *The first step in figuring out what's happening inside a relay.*

This test tells you that some contacts must close inside the relay when you apply voltage to the pair of terminals at the end nearest you. But maybe you're having difficulty holding the meter probes against the pins at the same time as you press the button. In that case, try using a couple of test leads as shown in Figure 2-51. With one end of each lead clipped to a meter probe, and the other end of each lead clipped to a relay pin, you have your hands free.

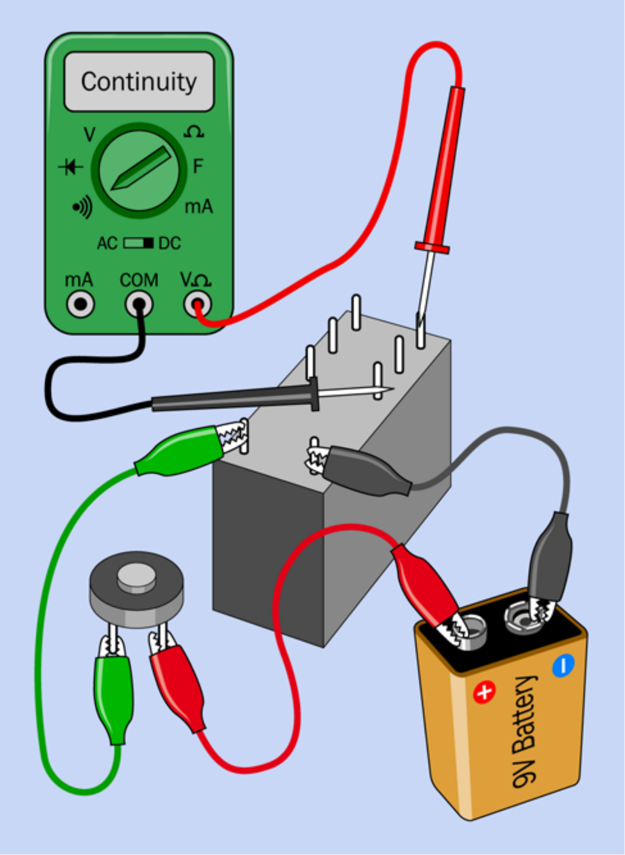
Continuity
V
Ω
F
mA
AC
DC
mA
COM
VΩ
9V Battery

Figure 2-50 *Step 2: Measuring continuity.*

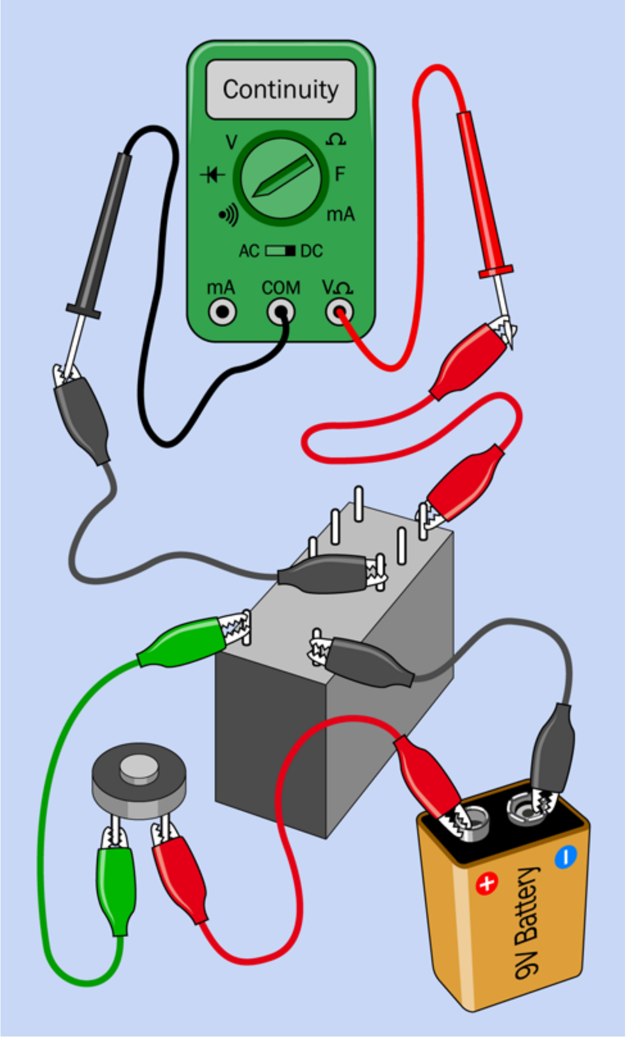

Figure 2-51 *You can extend your meter probes with test leads, to free yourself from holding the probes.*

Now try moving the red test lead from the relay pin farthest away from you to the vacant pin next to it. You should find that the behavior of the meter is reversed, so that it beeps when you don't press the button, and then stops beeping when you do press it.

What's Going On Inside

Figure 2-52 shows an x-ray view of the interior of the relay when you pressed the button. The relay contains a coil at the bottom, which generates a magnetic field, which moves a pair of internal switches. The coil moved the switch on the right to connect pins A and C internally, so the meter beeped.

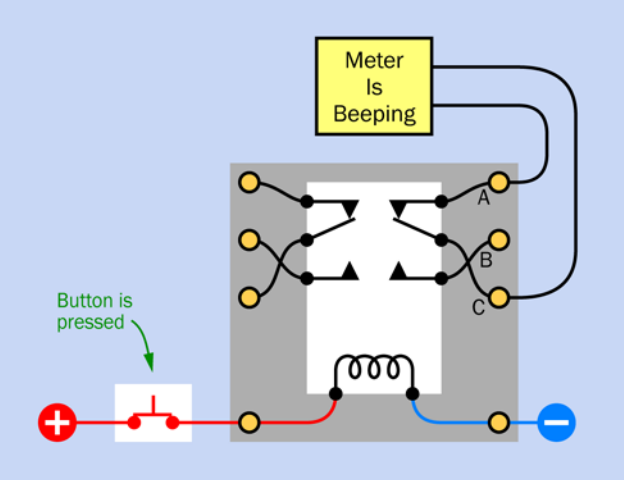

Figure 2-52 *Inside the relay, when the button is pressed, the meter starts beeping.*

You may be wondering why the coil in the relay seems to push the internal switch away from it. The reason is that there is a mechanical linkage inside the relay which converts a pulling force to a pushing force. You'll be able to inspect this when I get to the point of opening up the relay, later in this experiment.

Figure 2-53 shows what happened when you were not pressing the button. The switch contacts relaxed into their opposite state, breaking the connection between A and B and making the connection between B and C. The contacts remain in this position when no power is flowing through the relay coil.

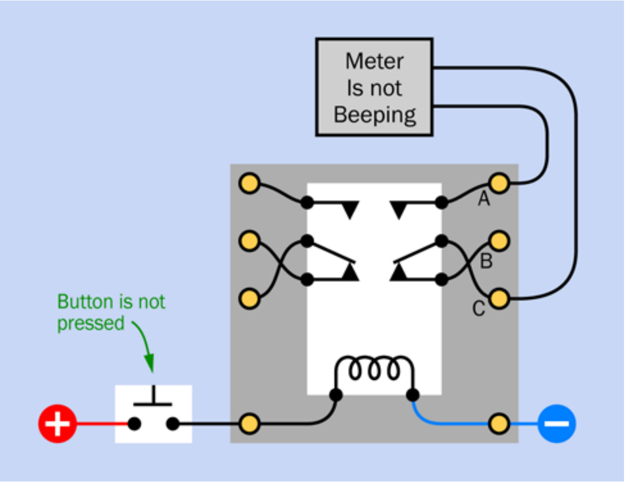

Figure 2-53 *Inside the relay, when the button is not pressed and the meter is not beeping.*

Other Relays

I believe that the pin functions that I have described are the most common for this size of relay, but I know there are some exceptions that do things differently. In fact, the first edition of this book used a relay that had different pin functions.

When you encounter a double-pole, double-throw (DPDT) relay for the first time, how can you determine what's going on inside it? You test different pairs of pins with your meter while you apply voltage to the coil. By a process of elimination, you can figure out how the pins are connected.

You can also read the manufacturer's datasheet. It should contain diagrams such as the one in Figure 2-21.

Is this all you need to know about relays? No, I've hardly scratched the surface.

- Some relays are *latching*, meaning that the internal switches remain in either position when the power is off. Latching relays usually have *two coils* to move the switches each way. I will not be using them in this book.
- Some relays have two poles, some have only one; some are double-throw, and some are single-throw.
- Some coils switch with AC, some switch with DC, and as I've mentioned previously, some DC coils require you to apply DC voltage with correct polarity.

As always, datasheets should provide the necessary information.

Figure 2-54 shows a selection of schematic symbols for various types of relays. Type A is single-pole, single-throw. Type B is single-pole, double-throw. Type C is single-pole, single-throw, drawn in the style that I like to use, with a white rectangle reminding you that the parts are enclosed in a single component. Type D is single-pole, double-throw. Type E is double-pole, double throw. Type F is single-pole, double-throw, latching.

Relay schematics are always drawn with the internal switch in its relaxed position, when power is not applied—with the exception of the latching relay, where the position of the switch is arbitrary.

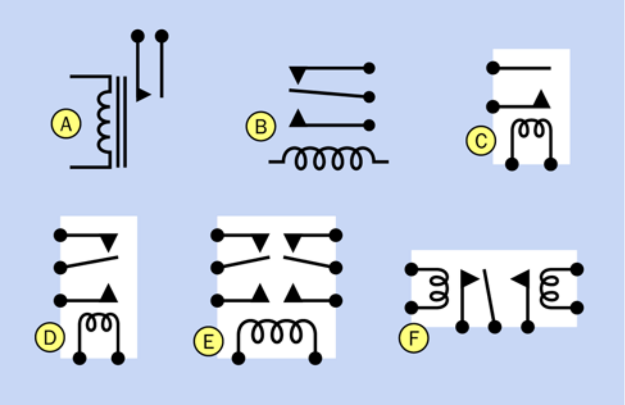

Figure 2-54 *Various schematic symbols for relays. See text for details.*

The type of relay you have been testing is a *small-signal relay*, meaning that it can't switch a lot of current. Larger relays may be capable of switching many amperes. It's important to choose a relay that is rated for the maximum current in your circuit (or higher), because overloading a relay will cause sparking that quickly erodes its contacts.

In future experiments you'll discover some practical uses for a relay—for example, in a home security system. Before you get to that, I'm going to show you how to turn a relay into an oscillator that buzzes. But first, I think it's time to take a look inside.

Opening It Up

If you have an impatient disposition, you can open your relay using methods such as those in Figure 2-55 and Figure 2-56. Generally, though, you may be better off using a most mundane piece of equipment: a box cutter or utility knife.

Figure 2-55 *Option 1 for opening a relay (probably not recommended).*

Figure 2-56 *Option 2 for opening a relay (definitely not recommended).*

Figure 2-57 and Figure 2-58 illustrate the technique that I like to use. You shave the edges of the plastic shell, beveling them until you see just a hair-thin opening. Don't go any farther; the parts inside are very, very close to your knife blade. Now pop the top off. Repeat this procedure with the remaining edges of the shell, and if you were really careful, the relay will be exposed but will still work when you energize its coil.

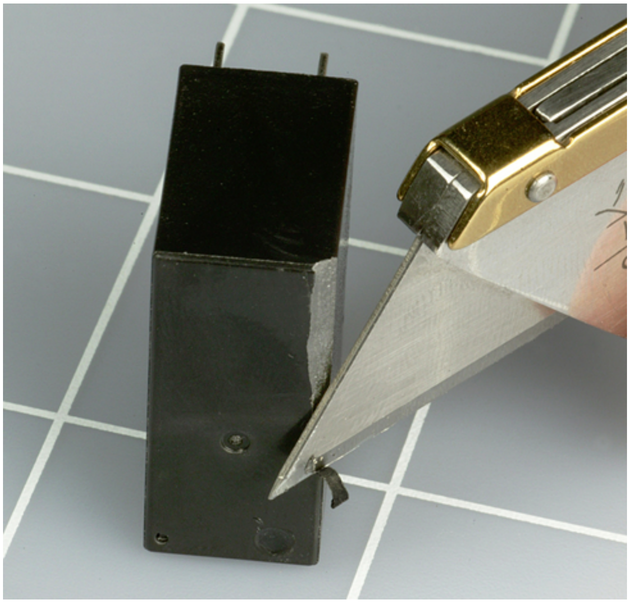

Figure 2-57 *Shaving the edges of the plastic box of a relay is a first step to opening it. Always cut away from you and downward toward your work bench.*

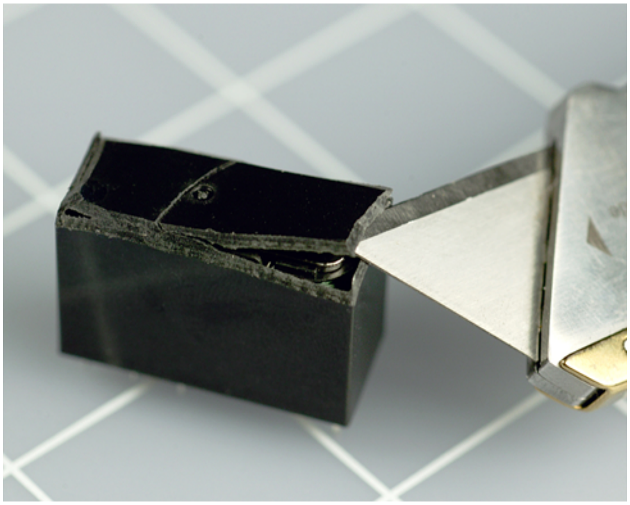

Figure 2-58 *After shaving the edges, you should be able to pry open one section of the case.*

What's Inside?

Figure 2-59 shows a simplified view of the parts in a typical relay. The coil, A, generates a magnetic attraction pulling lever B downward. A plastic extension, C, pushes outward against flexible metal strips and moves the poles of the relay, D, between the contacts. (This is a slightly different configuration from the relay recommended for experiments in this book, but the general principle is the same.)

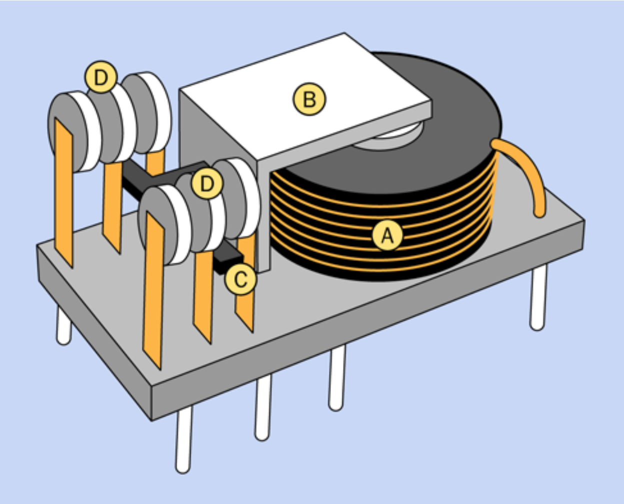

Figure 2-59 *Simplified view of the interior of a relay. See text for details.*

You can compare the diagram with an actual relay that I opened up, in Figure 2-60.

Various sizes of relays are shown with their cases removed in Figure 2-61. All of them happen to be designed for 12 volts DC. The automotive relay at far left is the simplest and easiest to understand, because it is designed without much concern for the size of the package. Smaller relays are more ingeniously designed, more complex, and more difficult to figure out. Usually, but not always, a smaller relay is designed to switch less current than a larger one.

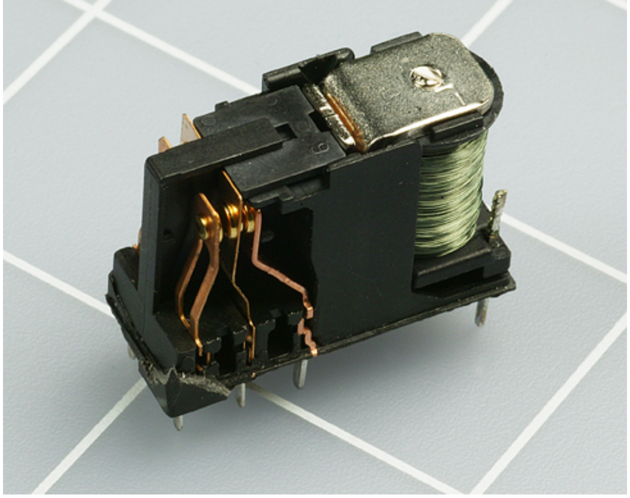

Figure 2-60 *An actual relay, exposed. The squares on the cutting mat are 1" x 1".*

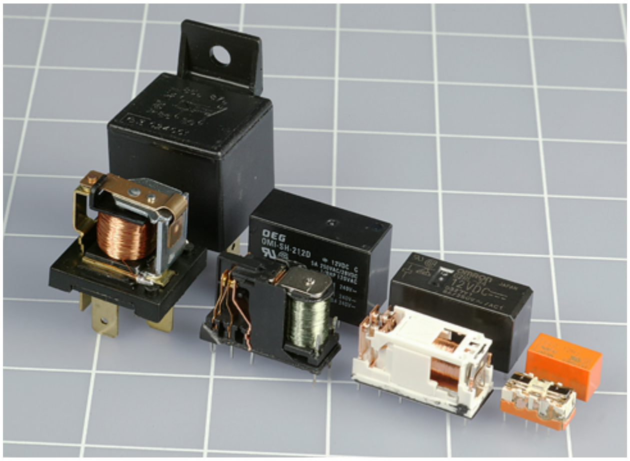

Figure 2-61 *A variety of 12-volt relays. See text for details.*

Fundamentals: Relay Terminology

Coil voltage is the voltage that the relay is supposed to receive when you energize it. This may be AC or DC.

Set voltage is the minimum that the relay will accept, to close its switch. This will be a bit less than the ideal coil voltage. In practice, a relay will probably work with even less power than the set voltage, but the set voltage tells you the minimum at which it is guaranteed to work.

Operating current is the power consumption of the coil, usually in milliamps, when the relay is energized. Sometimes the power is expressed in milliwatts.

Switching capacity is the maximum amount of current that the contacts inside the relay can switch without damage. Usually this is specified for a *resistive load*, meaning a passive device such as an incandescent lightbulb. When you use a relay to switch on a motor, it imposes an *inductive load*, which takes a big initial surge of current before it gets up to speed. Switching the motor off creates another surge. If the datasheet for a relay doesn't rate its ability to handle an inductive load, a rule of thumb is to assume that a motor may draw twice as much current when it starts, compared with when it is running.

Experiment 8: A Relay Oscillator

When you used test leads with alligator clips in previous experiments, they had two big advantages: you could assemble a circuit quickly, and you could see its connections easily.

Sooner or later, though, you have to get acquainted with a quicker, more convenient, more compact, and more versatile method for building circuits, and that time has come. I'm referring to the most widely used prototyping device: a *solderless breadboard*.

In the 1940s, circuits were built on a platform that really did look like a board that could be used for slicing bread. Wires and components were nailed, stapled, or screwed into place, because this was a whole lot easier than the alternative, which was mounting them on pieces of sheet metal. Remember, plastic barely existed back then. (A world without plastic—can you imagine it?)

Today the term "breadboard" is used for a little slab measuring about 2" by 7", and no more than 1/2" thick, as pictured in Figure 2-10. This is a wonderfully quick and easy system for assembling components. The only problem is that it creates internal connections between the components that are difficult to visualize—although I have ways to help you deal with that.

The best way to learn breadboarding is to assemble a circuit, which is exactly what you are about to do, taking the previous experiment with a relay one step further.

What You Will Need

- 9-volt battery (1)
- Battery connector (1)
- Breadboard (1)
- DPDT 9VDC relay (1)
- Generic LEDs (2)
- Tactile switch (1)
- Resistor, 470 ohms (1)
- Capacitor, 1,000µF (1)
- Pliers, wire cutters, wire strippers (1 each)
- Hookup wire, at least two colors, no more than 12" each

A Beginner's Board

Figure 2-62 shows the top part of your breadboard with the components that I would like you to plug into it.

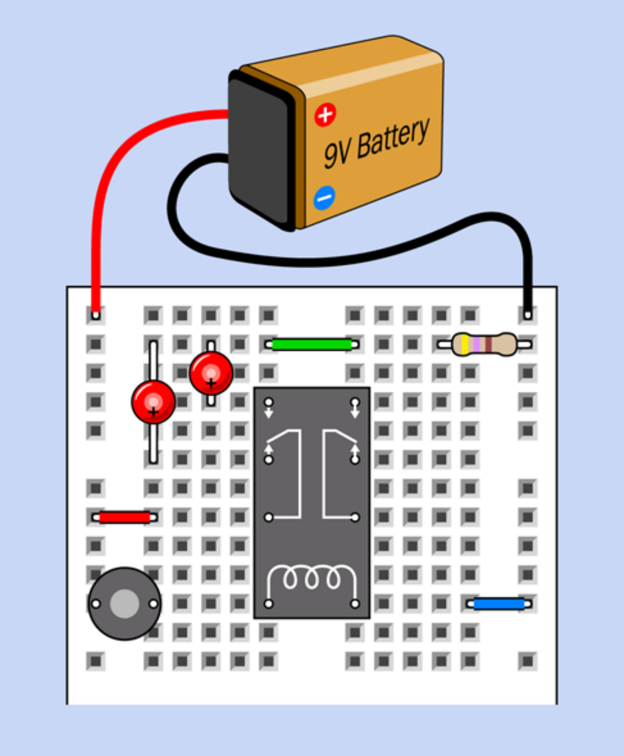

Figure 2-62 *A relay test circuit mounted in a breadboard.*

In case you're wondering exactly what some of these components are, Figure 2-63 shows all the pictorial symbols that will be used in the rest of the book in breadboard diagrams. You have not encountered most of these components yet, but you can refer back to this diagram for reference.

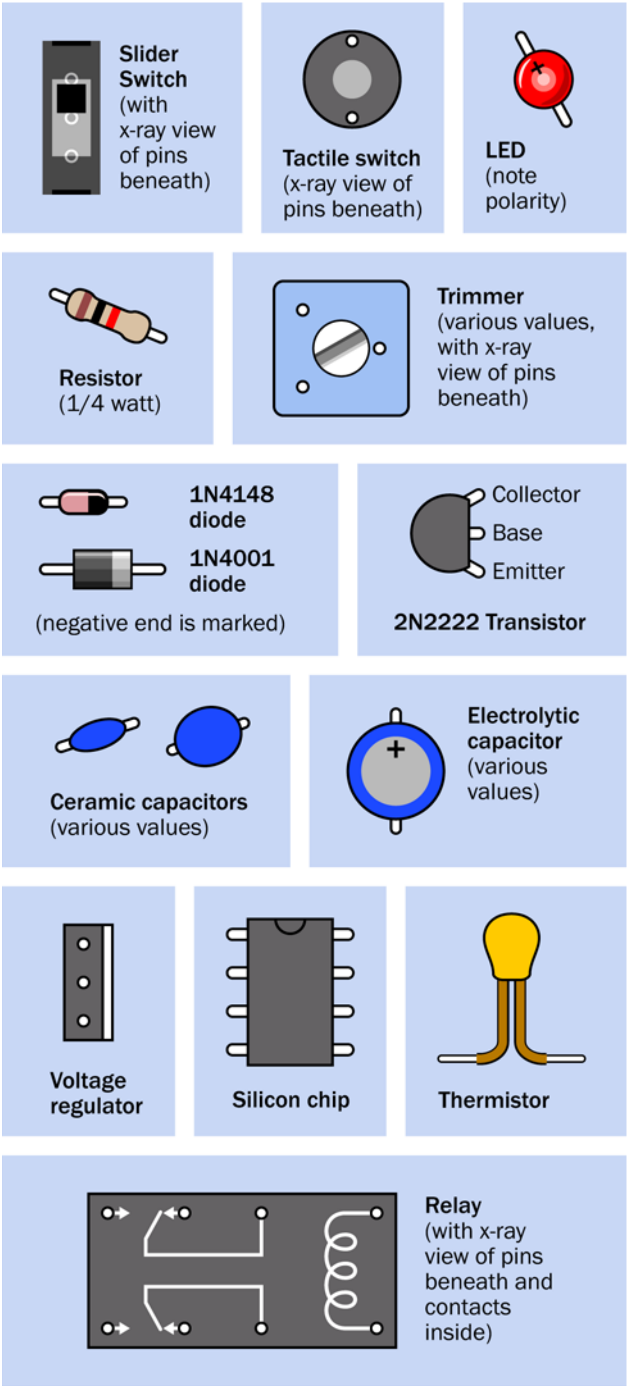
Slider Switch
(with x-ray view of pins beneath)
Tactile switch
(x-ray view of pins beneath)
LED
(note polarity)
Resistor
(1/4 watt)
Trimmer
(various values, with x-ray view of pins beneath)
1N4148 diode
1N4001 diode
(negative end is marked)
Collector
Base
Emitter
2N2222 Transistor
Ceramic capacitors
(various values)
+
Electrolytic capacitor
(various values)
Voltage regulator
Silicon chip
Thermistor
Relay
(with x-ray view of pins beneath and contacts inside)

Figure 2-63 *Representation of breadboarded components.*

In Figure 2-62 the relay from the previous experiment is in the center. You won't be able to see the pins from above, because they are inserted into the board underneath. I have shown their positions so that you know which way around the relay should be (that is, with its coil pins at the bottom). I have also shown the connections that exist inside the relay, just to remind you how they are configured. The switch is in the position where it rests when there is no power. This is the "relaxed" position.

The gray circular object is your pushbutton, more properly known as a tactile switch. I have shown an x-ray view of its pin positions, too, so you know how it should be oriented.

The two red circular objects are LEDs. Make sure the long lead of each one is on the side where the plus sign is shown.

The resistor value is 470 ohms—but you could have figured that out by looking at the colors of its stripes.

The red, green, and blue segments that look like pieces of wire, plugged into the breadboard, really are pieces of wire, plugged into the breadboard. My next task is to tell you how to make them.

Making Jumpers

If you bought an assortment of precut hookup wires, also known as jumpers, you can just go ahead and push them into the breadboard in the positions shown, although they won't be the same colors as the ones in my illustration.

As I mentioned previously, I advocate making your own. The exact procedure that I use is shown in Figure 2-64. First, strip off a couple of inches of insulation from some hookup wire. To do this, hold the wire in your left hand (or your right hand, if you are lefthanded). Hold the wire strippers in your other hand. Close the wire strippers so that they grip the wire in the hole marked "22" on the blade of the strippers. Pull the strippers away from your other hand, and they should take the insulation with them. (If you are wondering why you use the hole marked "22," it's because you are using 22-gauge wire. At least, I hope that's the size you are using.)

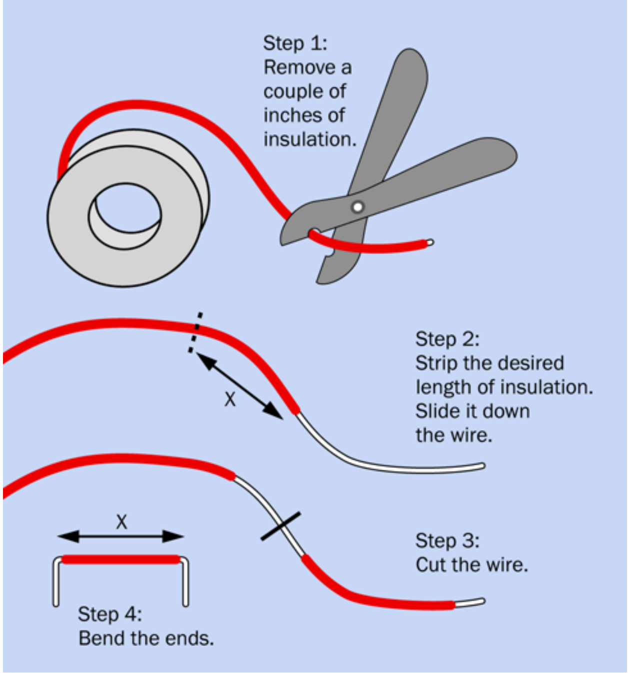
Step 1:
Remove a couple of inches of insulation.
Step 2:
Strip the desired length of insulation. Slide it down the wire.
X
Step 3:
Cut the wire.
X
Step 4:
Bend the ends.

Figure 2-64 *Procedure for making jumper wires. See text for details.*

Next assess how long the visible part of your wire is supposed to be when it is installed on the board. I will call that distance X inches. Measure X inches of the insulation that remains on your hookup wire. Drag a section of insulation that is X” long until it is about 3/8” from the end of the wire.

Using your wire cutters, or the blades built into your wire strippers, snip the wire 3/8” behind the X” section of insulation that you just dragged along the wire.

Finally, use your pliers to make a neat right-angle bend at each end, and push it into the board. Wait—does it not quite fit? If you practice a little, you’ll soon get to the point where you can make jumpers of the right length just by visual estimation.

Power Up

Lastly you need to apply power from your 9-volt battery. You should find that the wires attached to the connector terminate in little bare soldered ends that will poke into the holes in the breadboard. If you have trouble with them, try feeding them in with the tips of your pliers. If you still have trouble, you may need to strip off a couple more millimeters of insulation, using your wire strippers.

After inserting the wires into the breadboard, snap the connector onto the battery, as suggested in Figure 2-62. As soon as you have power on the breadboard, the LED on the left should light up. When you press the button, the switch inside the relay will close and the LED on the right will light up. Congratulations! You just breadboarded your first circuit.

Now, why does it work?

Inside the Board

Figure 2-65 reveals the copper strips that are hidden inside the breadboard. The little squares show where each lead of a component can poke through and make contact with the strips inside.

Each of the two long, vertical strips is known as a *bus* (plural: buses). This type of bus does not transport people, but does transport electrons, as the positive and negative sides of the power supply are typically connected with the buses.

- In this book, I am consistently placing the positive side of the supply on the left bus and negative ground on the right bus.

An important thing to notice is that each bus has two breaks in it. Not all breadboards have this feature, but many do. The purpose is to allow you to use multiple power supplies of different voltages at different locations on the board. In practice this doesn’t happen often, and the breaks in the buses are annoying, because you may tend to forget that they’re there. When you build a circuit that extends down the board, and you find a mysterious lack of power around the halfway mark, you may realize eventually that you forgot to add jumper wires bridging the gaps in the buses.

Where necessary, I will be reminding you to take care of this little detail.

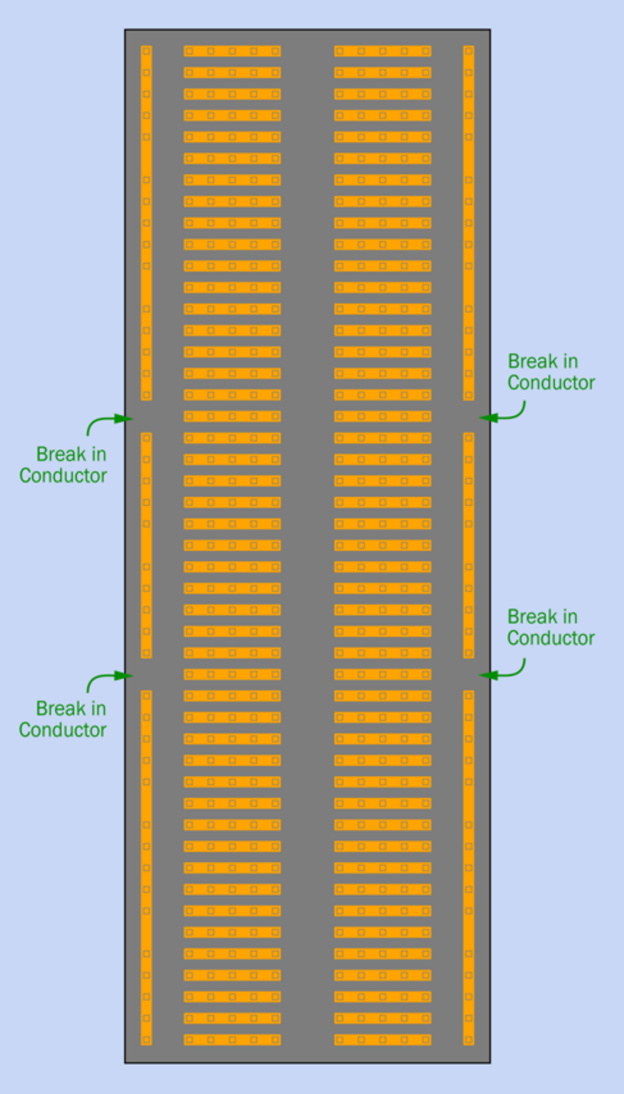

Figure 2-65 *A single-bus breadboard contains copper connecting strips in this configuration.*

Relay Circuit, Revealed

Figure 2-66 shows you the copper strips hidden inside your breadboard. They make connections between components plugged into the breadboard. The electricity takes a zigzag path, but the resistance of the copper strips is so low, the length of the path doesn't matter.

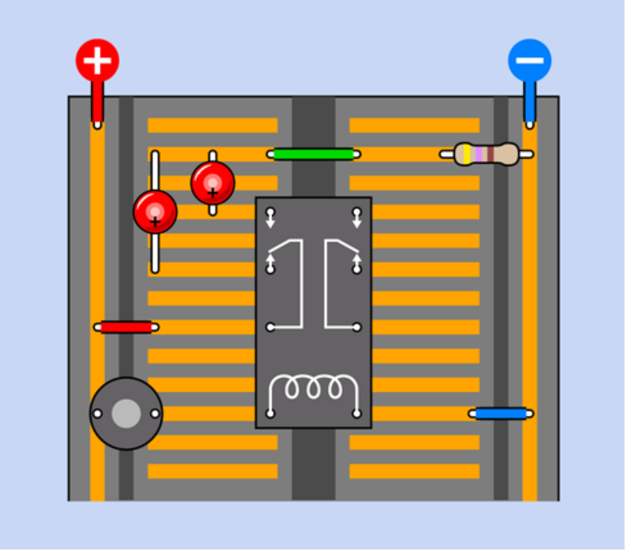

Figure 2-66 *Components on the breadboard are connected through the copper strips inside it.*

Maybe the diagram is easier to understand if I hide the copper strips that aren't doing anything, and just show the ones that are part of the circuit, as shown in Figure 2-67.

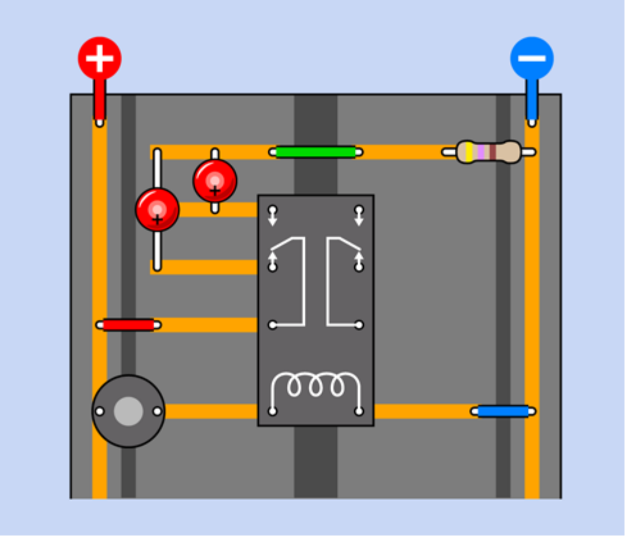

Figure 2-67 *The previous diagram is modified so that strips in the breadboard that are not an active part of the circuit have been omitted.*

Now take a look at the schematic for the same circuit, in Figure 2-68. I laid out the schematic to resemble the breadboard, to emphasize the similarity. As you continue through the book I'm going to rely on schematics more, and will expect that you can create your own breadboard layouts. But I'll take a while to get to that point.

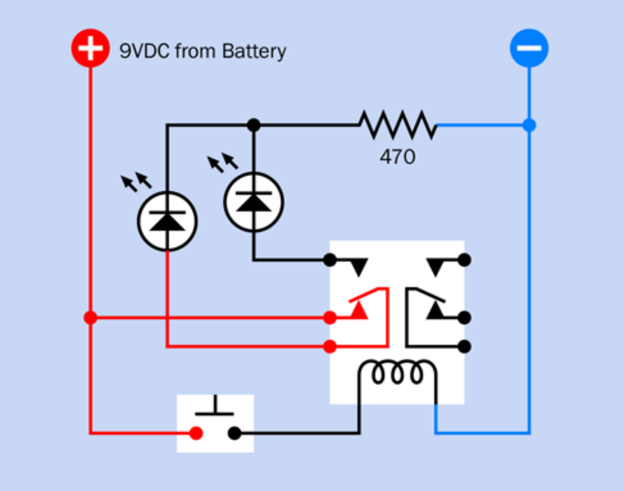

Figure 2-68 *A schematic that corresponds with the breadboard connections already shown.*

If you're wondering why there is only one 470-ohm resistor to protect two LEDs, it's because the LEDs only light up one at a time.

Making It Buzz

The next step is to modify your circuit, to make it more interesting. Look at the new schematic in Figure 2-69. Compare it with the previous version in Figure 2-68. Can you spot the difference? In the old version, the pushbutton that energized the coil received its power directly from the 9V supply. In the new version, the pushbutton gets its power through the lower contact of the relay. What effect will this have?

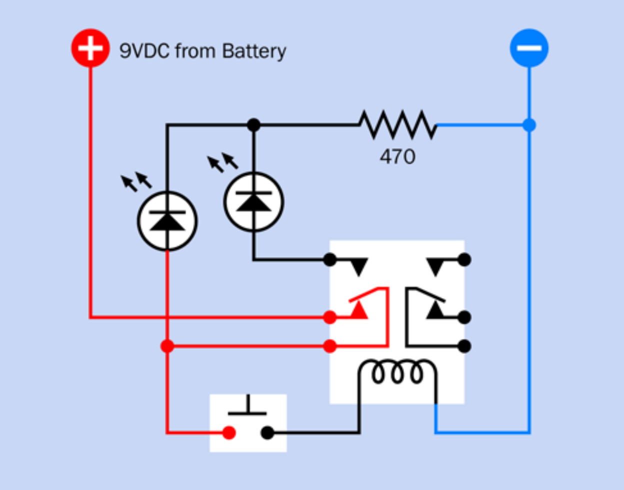

Figure 2-69 *A revised version of the previous schematic now delivers power to the pushbutton through the lower contact of the relay.*

Figure 2-70 shows how you can adapt your previous breadboarded circuit to match the new schematic. All you have to do is rotate the pushbutton by 90 degrees, and use an extra jumper wire (colored green in the figure) to link it with the same relay pin that energizes the lefthand LED.

Push the button—briefly!—and what happens? The relay makes a buzzing sound. (If your hearing is not good, touch the relay to feel it vibrating.)

Can you see what's happening, here? In its relaxed state, the switch inside the relay rests against the lower contact. This feeds positive voltage to the lefthand LED, and also to the pushbutton. Consequently, when you press the button, the power connects with the relay coil. The coil pushes the internal switch upward—but as soon as it does that, it breaks the connection feeding voltage to the coil. So, the switch falls back to its relaxed position. But this energizes the coil again, so the cycle repeats.

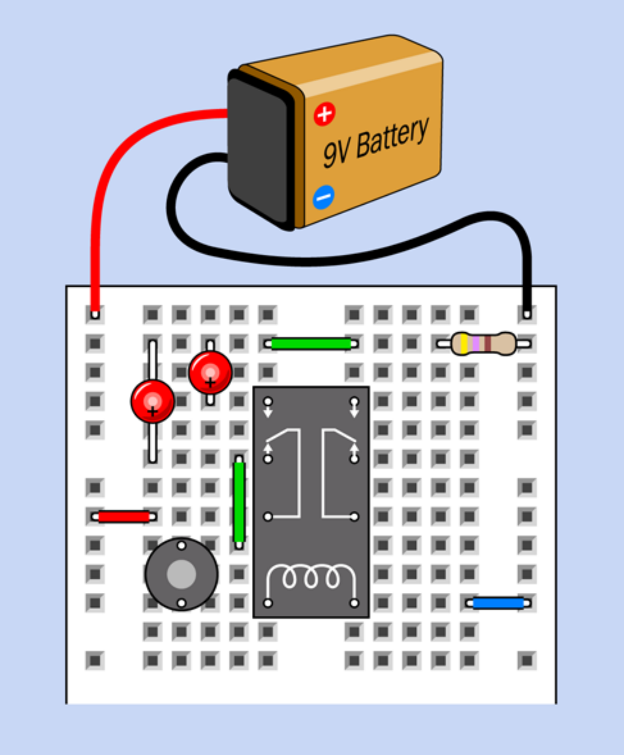

Figure 2-70 *The previous breadboarded circuit has been modified to match the revised schematic.*

The relay is *oscillating* between its two states.

Because you're using a small relay, it switches on and off quite fast. In fact, it oscillates perhaps 20 times per second (too fast for the LEDs to show what's really happening).

- When you force a relay to behave like this, it's liable to burn itself out or destroy its contacts. You are also controlling a bit more current than your tactile switch is designed to handle. So don't hold down the button for long! To make the circuit less self-destructive, we need to make everything happen more slowly. I'm going to achieve this by using a capacitor.

Adding Capacitance

Add a 1,000µF electrolytic capacitor in parallel with the coil of the relay as shown in the diagram in Figure 2-71, making sure that the capacitor's *short* wire is connected to the *negative* side of the circuit; otherwise, it won't work. In addition to the short wire, you should find a minus sign on the body of the capacitor, which is there to remind you which side should be more negative. In the diagram, I used a plus symbol, because it's more obvious than a minus symbol, and I wanted to be consistent with the style I used for LEDs.

- Electrolytic capacitors react very badly to being connected the wrong way around. They may self-destruct. Double-check the polarity.

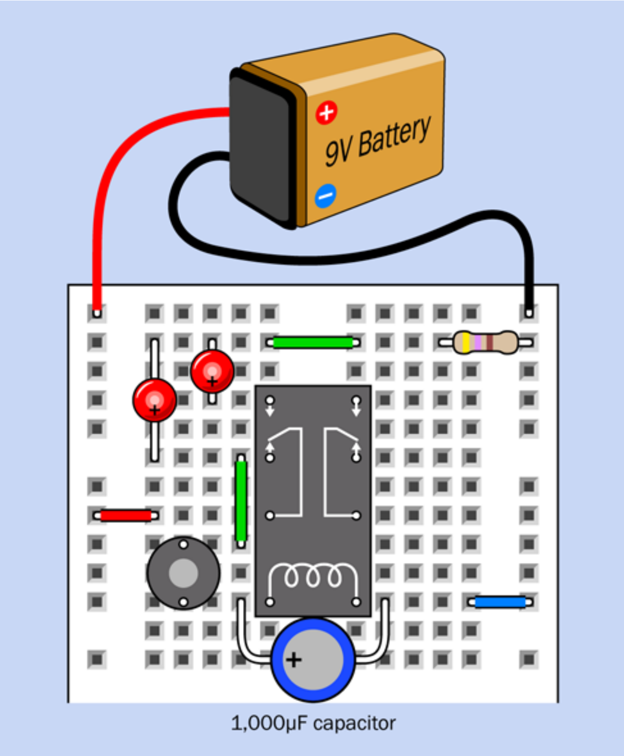

Figure 2-71 *With the addition of a large capacitor, the behavior of the circuit slows down.*

When you press the button now, the relay should click intermittently instead of buzzing.

A capacitor is like a tiny rechargeable battery. It's so small that it charges in a fraction of a second, before the relay has time to open its lower pair of contacts. Then, when the contacts are open, the capacitor releases its power to the relay (and to the lefthand LED). This keeps the coil of the relay energized for a moment. After the capacitor exhausts its power reserve, the relay relaxes and the process repeats.

During this process, the capacitor is *charging and discharging*.

Disconnect the righthand LED, and you should find that the lefthand LED pulses in a pleasing manner, gradually fading as the voltage from the capacitor diminishes.

Because the capacitor takes a big surge of current when it charges, your tactile switch may overheat if you hold it down for too long during this experiment.

Fundamentals: Farad Basics

The storage ability of a capacitor is measured in *farads*, represented by an uppercase letter F. The term is named after Michael Faraday, another in the pantheon of electrical pioneers.

The farad is a large unit, and is divided into microfarads (abbreviated µF, each being 1/1,000,000th of a farad), nanofarads (abbreviated nF, each being 1/1,000th of a microfarad), and picofarads (abbreviated pF, each being 1/1,000th of a nanofarad). In the United States, the nanofarad is used less often than in Europe. Instead, values may be expressed using picofarads and fractions of a microfarad.

A conversion table for picofarads, nanofarads, microfarads, and farads is shown in Figure 2-72.

Picofarads	Nanofarads	Microfarads	Farads
1pF	0.001nF	0.000001µF	
10pF	0.01nF	0.00001µF	
100pF	0.1nF	0.0001µF	
1,000pF	1nF	0.001µF	
10,000pF	10nF	0.01µF	
100,000pF	100nF	0.1µF	
1,000,000pF	1,000nF	1µF	0.000001F
		10µF	0.00001F
		100µF	0.0001F
		1,000µF	0.001F
		10,000µF	0.01F
		100,000µF	0.1F
		1,000,000µF	1F

Figure 2-72 *Conversion table for fractions of farads.*

Caution: Getting Zapped by Capacitors

If a large capacitor is charged with a high voltage, it can retain that voltage for minutes or even hours. Because the circuits in this book use low voltages, you don't have to be concerned about this here, but if you are reckless enough to open an old TV set and start digging around inside (which I do not recommend), you may have a nasty surprise. A large charged capacitor can kill you as easily as if you stick your finger into an electrical outlet.

Fundamentals: Capacitor Basics

No electrical connection exists inside a capacitor. Its two leads are connected internally with *plates* that are a tiny distance apart, separated by an insulator known as the *dielectric*. Consequently, DC current cannot flow through a capacitor. However, if you connect a capacitor across a battery, it will charge itself as suggested in Figure 2-73, because the charge on one plate attracts an opposite charge on the other plate.

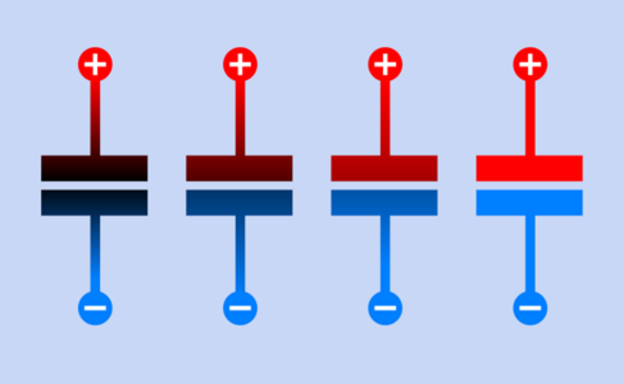

Figure 2-73 *A capacitor connected across a battery will accumulate equal and opposite charges, as suggested here.*

In modern capacitors, the plates have been reduced to strips of very thin, flexible, metallic film.

The two most common varieties of capacitors are ceramic (usually small, to store a relatively small charge) and electrolytic (which can be much larger). Electrolytics are usually shaped like miniature tin cans and may be any color, although black is most common. Older ceramics are often disc-shaped, while newer ones are little rounded blobs.

Ceramic capacitors have no polarity, meaning that you don't have to worry about which way around they are in a circuit. Electrolytics do have polarity, and won't work unless you connect them the right way around.

The schematic symbol for a capacitor contains two lines representing the two plates inside it. If both of the lines are straight, the capacitor does not have a polarity—it can be used either way around. If one of the lines is curved, that side of the capacitor must be more negative than the other. A + sign may be included to remind you of the polarity. The variants are shown in Figure 2-74.

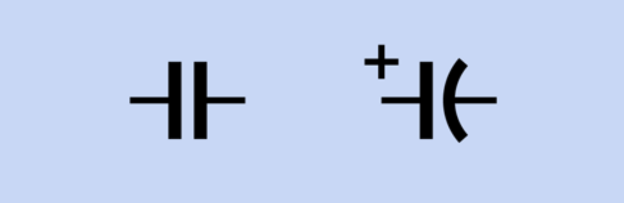

Figure 2-74 *Two variants of the symbol representing a capacitor. See text for details.*

The symbol with the curved plate is not used so much anymore. People assume that if you have an electrolytic capacitor, you'll be smart enough to connect it the right way around. Also, multilayer ceramic capacitors have become available in higher values, and may be substituted for electrolytics.

- My schematic diagrams will only use the non-polarized capacitor symbol. Whether you use an electrolytic capacitor or a ceramic capacitor is your choice.
- My breadboard diagrams will show electrolytic capacitors where I think you are most likely to use them. Still, you can substitute ceramic capacitors if you wish.

Caution: Observe Capacitor Polarity!

The most common type of electrolytic capacitor uses aluminum plates. Two other types use tantalum and niobium, respectively. All of these capacitors are fussy about polarity. In Figure 2-75, a tantalum capacitor was plugged into this breadboard, connected the wrong way to a power source capable of delivering a lot of current. After a minute or so of this abuse, the capacitor rebelled by popping open and scattering small flaming pieces, which burned their way into the breadboard. Lesson learned: observe polarity!

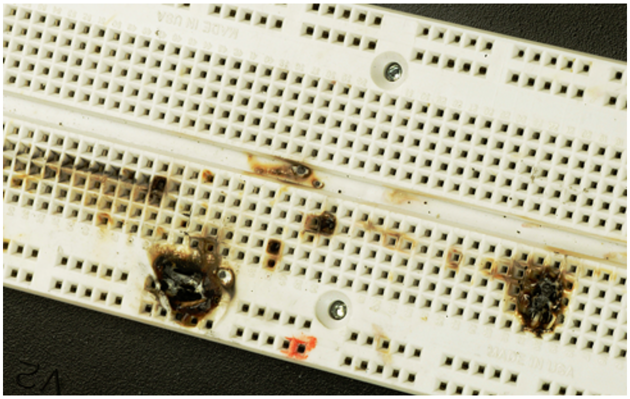

Figure 2-75 *The aftermath of an error in which a polarized tantalum capacitor was connected the wrong way around to a power source capable of supplying substantial current.*

Fundamentals: Fault Tracing

As you build more circuits on breadboards, and they become more complex, errors become more likely. No one is exempt from this unfortunate fact.

One frequent breadboarding error is to plug a wire into the wrong row on a breadboard. This is especially easy when you have a component such as a relay, where the pins are hidden. Often I pull that component out, take another look, and push it back in, just to make sure.

A more subtle error occurs when you forget the connections made by the strips hidden inside the board. Take a look at Figure 2-76. What could be simpler? Clearly, the positive source of power feeds current through the LED, through a jumper, and then through a resistor to the negative bus. But if you assemble the components as I have shown them, I absolutely, positively guarantee that they won't work.

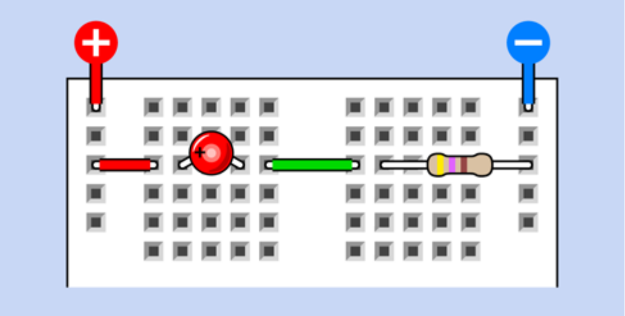

Figure 2-76 *This breadboarded circuit won't work. Can you see why?*

The situation will get much worse if you swap the positions of the resistor and the LED. Now the circuit will immediately burn out the LED.

The explanation becomes apparent when you check the x-ray view in Figure 2-77. The problem is that both leads of the LED are plugged into the same strip inside the board. The electricity has the option to run through the LED, or run through the copper strip instead—and since the resistance of the strip is a tiny fraction of the resistance of the LED, most of the electrons prefer the strip, and the LED remains dark.

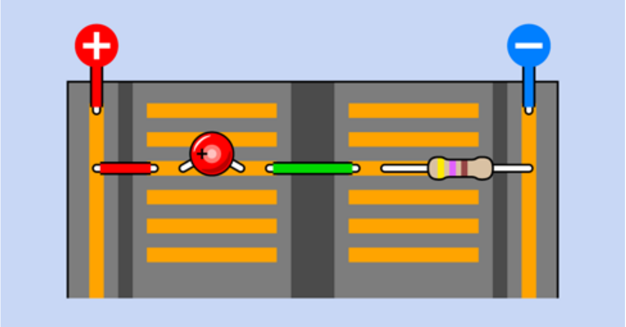

Figure 2-77 *An x-ray view helps to explain why this won't work.*

Many other kinds of errors are possible. How can you find them most quickly and efficiently? You simply have to be methodical. Try to follow these steps:

1. **Check voltages**. Attach the red lead from your meter to a connection near the top of the positive bus on your breadboard. Set the meter to measure volts (DC volts, unless an experiment suggests otherwise). Make sure the power to your circuit is switched on. Now touch the black probe from your meter to various locations down the negative bus. The meter reading should be close to the supply voltage. If you find a value close to zero, you probably forgot to include a jumper to bridge one of the gaps in the negative bus. If you find a value of a few volts, but it is significantly lower than the supply voltage, you may have a short circuit somewhere, pulling down the voltage from the battery (if you are using a battery).

   Now secure the black probe to a connection near the top of the negative bus, and check the positive bus from top to bottom.

   Finally, with the black probe still attached near the top of the negative bus, use the red probe to check voltages at random locations in your circuit. If you find a voltage close to zero, probably a connection is missing somewhere, or a component or wire is not making contact inside the breadboard.

2. **Check placement**. Make sure that all the jumper wires and component leads are exactly where they should be on the breadboard.

3. **Check component orientation.** Diodes, transistors, and capacitors that have polarity must be the right way around. When you start to use integrated circuit chips, later in this book, check that they are the right way up, and make sure that none of the pins on a chip has been bent so that it is hidden underneath the chip.

4. **Check connections.** Sometimes (seldom, but it can happen) a component may make a bad connection inside the breadboard. If you have an inexplicable intermittent fault or zero voltage, try relocating some of the components. In my experience this problem is more likely to occur if you buy very cheap breadboards. It is also more likely if you use wire that has a smaller diameter than 22 gauge. (Remember, a higher gauge number means a thinner wire.)

5. **Check component values.** Verify that all the resistor and capacitor values are correct. My standard procedure is to check each resistor with a meter before I plug it in. This is time-consuming but can actually save time in the long run.

6. **Check for damage.** Integrated circuits and transistors can be damaged by incorrect voltages, wrong polarity, or static electricity. Keep spares on hand so that you can make substitutions.

7. **Check yourself!** When all else fails, take a break. Working obsessively for long periods can create tunnel vision, which prevents you from seeing what's wrong. If you move your attention to something else for a while, then come back to your problem, the answer may suddenly appear obvious.

You may want to bookmark this list of fault-tracing procedures and come back to it later if something doesn't work.

Background: Michael Faraday and Capacitors

As previously noted, the farad is named after Michael Faraday. He was an English chemist and physicist who lived from 1791 to 1867. See Figure 2-78.

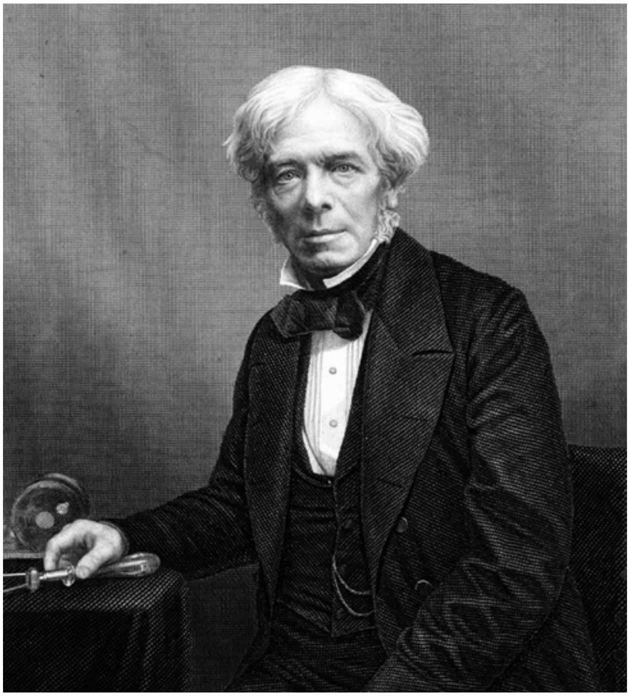

Figure 2-78 *Michael Faraday, after whom the farad is named.*

Although Faraday was relatively uneducated and had little knowledge of mathematics, he had an opportunity to read a wide variety of books while working for seven years as a bookbinder's apprentice, and thus was able to educate himself. Also, he lived at a time when relatively simple experiments could reveal fundamental properties of electricity. He made major discoveries including electromagnetic induction, which led to the development of electric motors. He also discovered that magnetism could affect rays of light.

His work earned him numerous honors, and his picture was printed on English bank notes denominated in 20 pounds sterling, from 1991 through 2001.

Experiment 9: Time and Capacitors

Electrons travel almost at the speed of light, yet we can use them to measure time in seconds, minutes, or even hours. This experiment will show you how.

What You Will Need

- Breadboard, hookup wire, wire cutters, wire strippers, test leads, multimeter
- 9-volt battery and connector (1)
- Tactile switches (2)
- Generic LED (1)
- Resistors: 470 ohms, 1K, 10K (one of each)
- Capacitors: 0.1µF, 1µF, 10µF, 100µF, 1,000µF (one of each)

Charging a Capacitor

First set your meter to measure volts DC, and measure the voltage of a 9-volt battery with your meter. If it's less than 9.2V, you need a newer battery for this particular experiment.

Install two tactile switches, a 1K resistor, and a 1,000µF capacitor on your breadboard as shown in Figure 2-79. Use a couple of test leads to connect your meter so that you can measure the voltage across the leads of the capacitor while keeping your hands free.

Snap a connector on your battery, and plug the wires into the breadboard to provide the 9VDC supply on the two buses of the breadboard, with positive on the left, as shown in the figure.

If the meter measures more than 0.1V, discharge the capacitor by pressing button B, which shorts together the two sides of the capacitor.

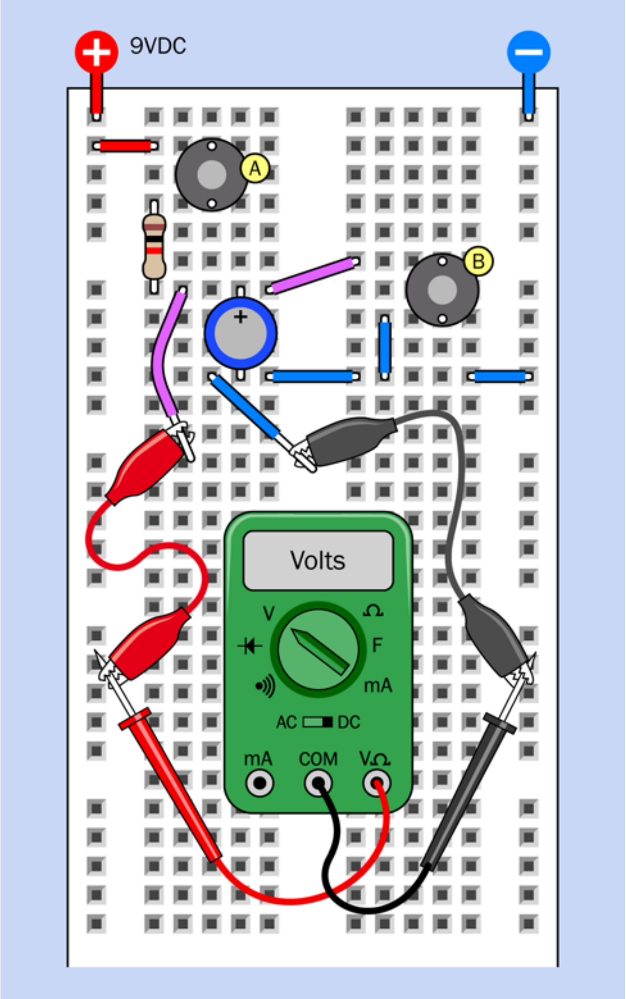

Figure 2-79 *A simple setup for timing the charging of a capacitor. The capacitor is 1,000μF and the resistor is 1K.*

A schematic in Figure 2-80 shows the same circuit, and may help you to see what's going on.

Now hold down button A while you use a watch, a clock, or a smartphone to count how many seconds the capacitor takes to charge to 9.0V. If you have an autoranging meter, it should automatically switch from measuring millivolts, at first, to measuring volts as the charge increases. When I performed this experiment, the meter took just over three seconds.

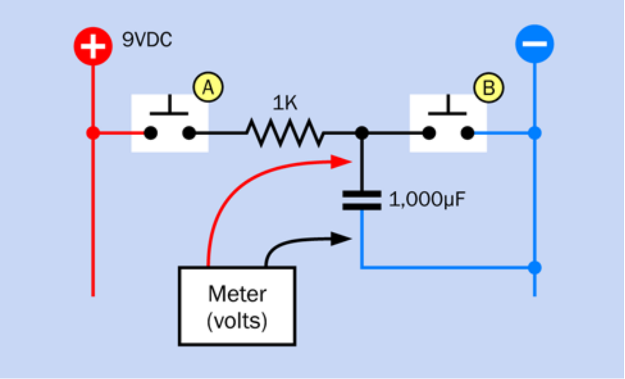

Figure 2-80 *This schematic shows the same circuit as the preceding figure, which shows the breadboarded version.*

The positive side of the capacitor has become "more positive" and the negative side has become "more negative" as electron-holes and electrons have been attracted to each other on the plates. The potential difference between the leads of the capacitor has increased, while current has not passed through it. One of the first statements you will find, when you read any introductory electronics text, is:

- A capacitor blocks DC (direct current).

So long as you apply a steady electrical potential to the capacitor, this is true.

An RC Network

Remove the 1K resistor and substitute a 10K resistor. If the meter shows that there is still some voltage across the capacitor, discharge it by pressing button B.

Now repeat the test. How long does the capacitor take to reach 9.0V, charging through a 10K resistor?

This simple combination of a capacitor and a resistor is known as an *RC network* (R for resistor, C for capacitor). It's a very important concept in electronics. Before I explain what it does, here are some questions to consider:

- Did the capacitor take exactly 10 times as long to reach 9 volts when you used the 10K resistor instead of the 1K resistor?
- Did the voltage across the capacitor rise at a steady rate, or did it increase faster at the beginning of the experiment—or toward the end?

- If you wait long enough, will the capacitor ever reach the initial value that you measured as the voltage of your battery?

Voltage, Resistance, and Capacitance

Think of the resistor as a faucet restricting the flow of water, and the capacitor as a balloon that you are trying to fill (see Figure 2-81). If you screw down the faucet until only a trickle comes through, the balloon will take longer to fill. But a slow flow of water should still fill the balloon if you wait long enough. Assuming that the balloon doesn't burst, the process will end when the pressure inside the balloon is equal to the water pressure in the pipe supplying the faucet.

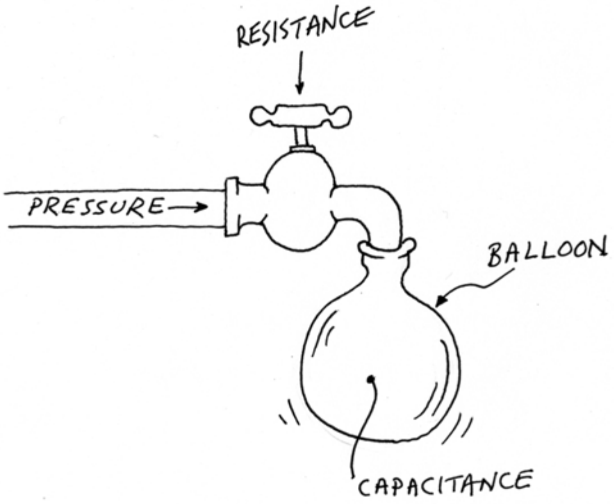

Figure 2-81 *Water flowing into a balloon can be compared with electrons flowing into a capacitor.*

But this description leaves out an important factor. As the balloon starts to fill, it stretches, exerting more pressure on its contents. As the pressure inside the balloon increases, it pushes back against the incoming flow of water. Consequently we may expect the water to flow in more slowly as the process continues.

How does this compare with electrons flowing into a capacitor? The concept is similar. Initially, the electrons go rushing in, but as they occupy more of the electron-holes, the newcomers take longer to find a resting place. The charging process gets slower, and slower. In fact, theoretically, the charge on the capacitor never quite reaches the voltage being applied to it.

Background: The Time Constant

The speed with which a capacitor charges is measured with a function known as the "time constant." The definition is very simple:

```
T = R × C
```

where T is the time constant, in seconds, if a capacitor of value C (measured in farads) is being charged through a resistor of R ohms.

Going back to the circuit you first tested, using a 1K resistor, we can put the values of the components that you used into the time-constant formula—but only if we convert the units to ohms and farads. Well, 1K is 1,000 ohms, and 1,000µF is 0.001 farads. So the calculation becomes very easy:

```
TC = 1,000 / 0.001
```

Therefore, for those values of a resistor and a capacitor, TC = 1.

But what exactly does this mean? Does it mean that the capacitor will be fully charged in one second? Sorry, it's not that simple.

- T, the time constant, is the number of seconds required for a capacitor to acquire 63% of the voltage being supplied to it, if it starts with zero volts.

What if the capacitor doesn't start from zero? If we start measuring after the capacitor has already acquired some voltage, the definition becomes a bit more complicated. If V_{DIF} is the difference between the voltage on the capacitor and the supply voltage, T is the number of seconds required for a capacitor to add 63% of V_{DIF} to its current charge.

(Why 63%? Why not 62%, or 64%, or 50%? The answer to that question is too complicated for this book, and you'll have to read about time constants elsewhere if you want to know more. Be prepared for differential equations.)

A comparison may be helpful. Figure 2-82 shows a greedy guy who is ready to eat some cake. At first he's ravenously hungry, so he takes 63% of the cake and eats it in one second, which is his time constant for eating cake. In his second bite, he takes another 63% of the cake that is left—and because he's not feeling so hungry anymore, he requires one more second (remember, this is his time constant). In his third bite, he takes 63% of what still remains, and still ingests it in one second. And so on. He is gradually filling up with cake, like a capacitor filling up with electrons. But he never quite eats all the cake, because he only takes 63% of whatever is left.

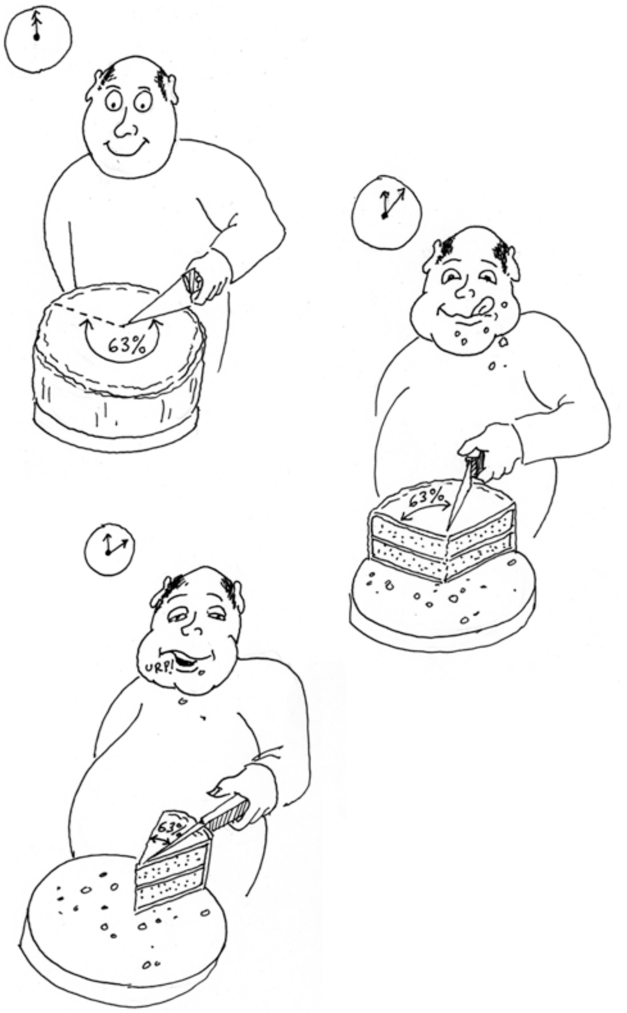

Figure 2-82 *If our gourmet always eats just 63% of the cake still on the plate, he "charges up" his stomach in the same way that a capacitor charges itself. No matter how long he keeps at it, the cake is never quite gone and his stomach is never completely filled.*

Figure 2-83 shows the process another way. After each time constant (which is one second, if we have a 1,000µF capacitor and a 1K resistor), the capacitor acquires another 63% of the difference between the charge it had and the voltage being applied.

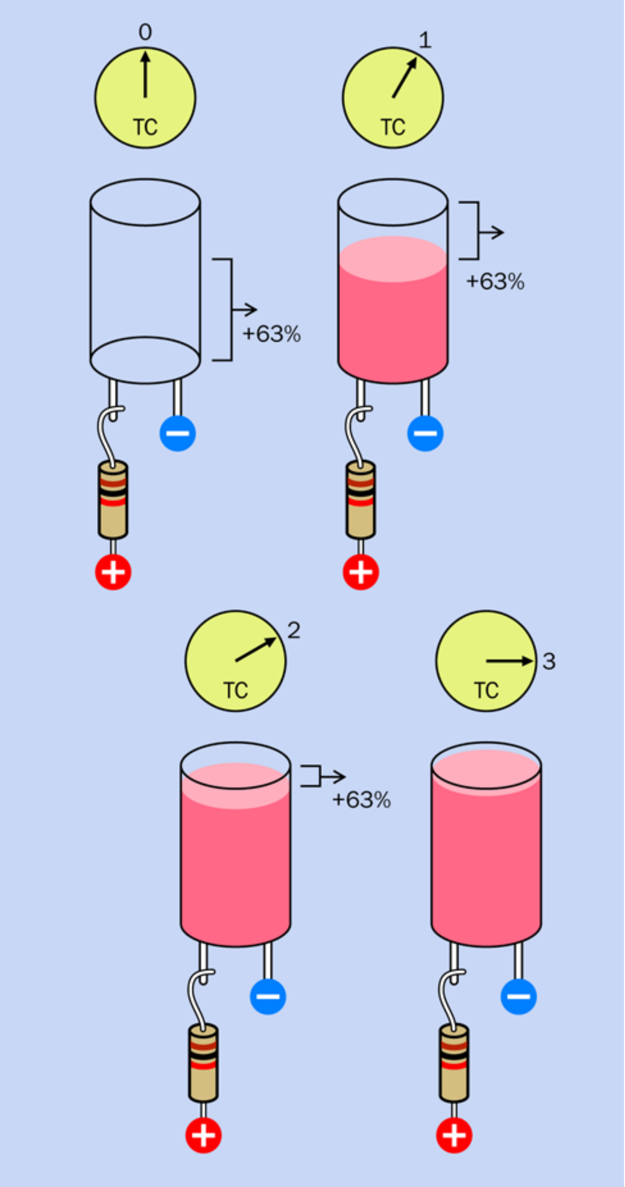

Figure 2-83 *Another way of looking at a charging capacitor.*

In a perfect world of perfect components, the charging process for a capacitor would continue for an infinite time. In the real world, we say rather arbitrarily:

- After five time constants, the charge on the capacitor will be so close to 100 percent, we can think of the process as being complete.

Background: Graphing It

I want to draw a graph showing the voltage in a capacitor as it charges. To do this, I'm going to calculate the data by using the time-constant formula.

Suppose V_{CAP} is the voltage on a capacitor right now, while V_{DIF} is the difference between that amount of charge and the battery voltage being applied (as before). The formula shown below will tell me what the new voltage on the capacitor will be after one time constant. I'll call the new voltage V_{NEW}. The formula looks like this:

$V_{NEW} = V_{CAP} + (0.63 \times V_{DIF})$

The 0.63 value is the same as 63 percent.

Suppose the battery was supplying exactly 9V and the capacitor started with exactly 0V. So, $V_{CAP} = 0$ and $V_{DIF} = 9$. Plug those values into the formula:

$V_{NEW} = 0 + (0.63 \times 9)$

My calculator tells me that 0.63 × 9 = 5.67. So after one time constant (one second, with a 1K resistor and a 1,000µF capacitor) the capacitor acquired 5.67 volts.

What about the next second? We have to repeat the calculation, using the new values. The current voltage on the capacitor, V_{CAP}, is now 5.67. The battery is still applying 9V, so V_{DIF} equals 9 minus 5.67, which is 3.33. We take those values back to the same formula:

$V_{NEW} = 5.67 + (0.63 \times 3.33)$

My calculator tells me that 0.63 times 3.33 equals 2.1, approximately. And 2.1 plus 5.67 equals 7.77. So, after two seconds, the capacitor has acquired 7.77 volts.

We can repeat this calculation any number of times, creating a sequence of numbers like this (rounded to two decimal places), showing the voltage on the capacitor at the end of each second, assuming a power supply of 9V:

```
After 1 second: 5.67 volts

After 2 seconds: 7.77 volts

After 3 seconds: 8.54 volts

After 4 seconds: 8.83 volts

After 5 seconds: 8.94 volts

After 6 seconds: 8.98 volts
```

The graph in Figure 2-84 was created by drawing a smooth curve through those values. I didn't bother to go beyond six seconds, because the values climb so close to 9V.

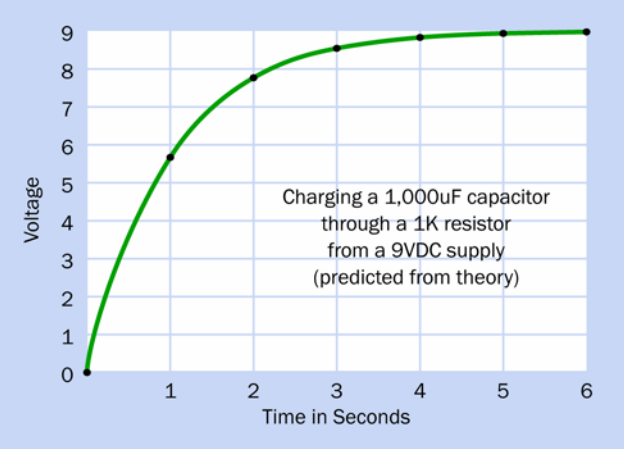

Figure 2-84 *A graph can help to show how charge accumulates on a capacitor over a period of time.*

Experimental Verification

I've told you how to calculate the charge on a capacitor in an RC network. But—how do you know I'm right? Should you just take my word for it?

Maybe you should test it yourself. In other words, you could do some *experimental verification*, which is a big part of Learning by Discovery.

Go back to the circuit that you used before, and make sure you have the 10K resistor in it, not the 1K resistor. Ask a friend to sit beside you, keeping track of the time, while you watch the display of volts on your multimeter. Every 10 seconds, your friend says "Now!" and you write down the voltage on the meter at that moment. You follow this procedure for one minute.

Because you're using a 10K resistor, not a 1K resistor, the time constant is now 10 seconds instead of 1 second. So, your readings should look like the series of voltages that I tabulated above at 1-second intervals, except that yours will be at 10-second intervals.

The voltages that you measure should be close to mine, but they won't match precisely. Why? I can think of many reasons.

- Your battery doesn't provide exactly the same voltage as mine.
- Your resistor isn't precisely 10,000 ohms.
- Your capacitor is not exactly 1,000 microfarads.
- Your meter is not totally accurate.
- You will take a few microseconds to read the meter.
- Your friend may not prompt you precisely every 10 seconds.

There are two other factors that you might not think of. First, capacitors don't store electricity perfectly. They suffer from *leakage*, in which the charge gradually leaks away. This happens even while the capacitor is gaining charge. Toward the end of the charging process, the electrons are trickling in so slowly, the leakage (the rate at which they are trickling out) becomes significant by comparison.

Additionally, your meter has an internal resistance. It's very high, but still, it's there. This means that the meter is stealing a little bit of the charge from your capacitor while you are measuring the voltage. Yes, the process of making a measurement changes the value that you are trying to measure! This is actually a common problem in physics and engineering.

I can think of ways to minimize all of these factors, but I can't imagine a way to eliminate them completely. There will always be some experimental error. This is the challenge when you use an experiment to verify a theory. Verification can be a very lengthy process, requiring a lot of patience—which is why theorists tend to have different personalities from experimentalists.

Capacitive Coupling

Now that I've told you how capacitors become charged and discharged, I have to go back to the statement I made before:

- A capacitor blocks DC (direct current).

You may remember, I also said, "So long as you apply a steady electrical potential to the capacitor, this is true."

But what if there isn't a steady electrical potential? What happens at the moment when a capacitor goes from having no charge on it to connecting suddenly to a source of voltage?

Ahhh, that's a different matter. Under those circumstances, the capacitor *allows a signal to pass through*.

How can this be? The plates inside a capacitor don't touch each other, so, how can a pulse of electricity jump from one to the other?

I'll deal with the "how" and "why" in a moment. First you need to make sure that what I'm talking about actually happens.

Take a look at the components on the breadboard in Figure 2-85. The layout is similar to the circuit in Figure 2-79, but the 10K resistor has moved from the left side to the right side, and an LED and a 470-ohm resistor have been added.

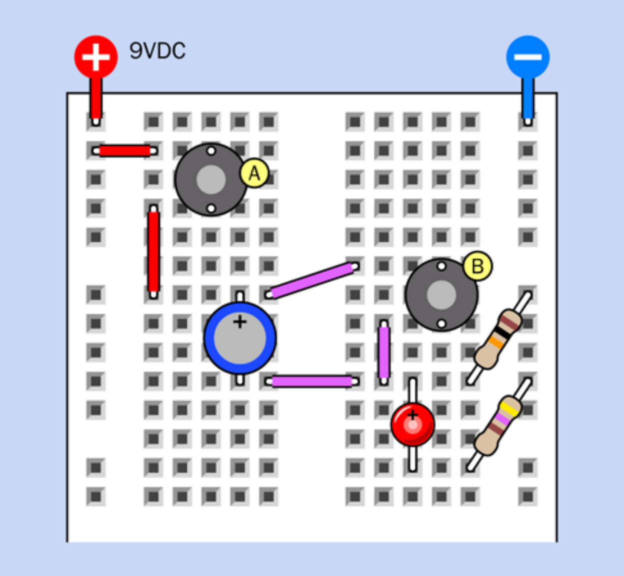

Figure 2-85 *The flashing of a red LED shows how the behavior of a capacitor changes when there is a quick change in voltage.*

The breadboarded circuit is redrawn as a schematic in Figure 2-86, which may help to clarify it.

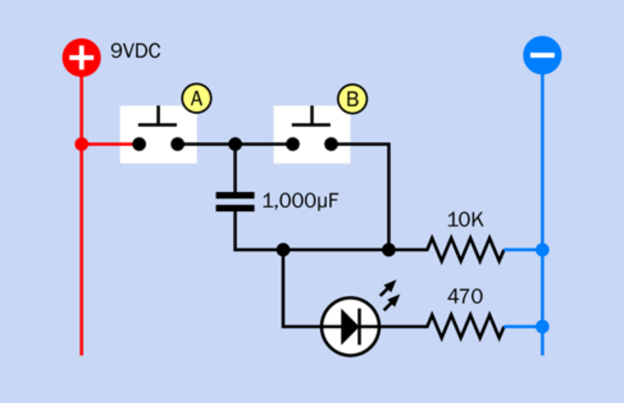

Figure 2-86 *This schematic shows the same circuit as in the preceding figure depicting a breadboarded version.*

And just to make sure there is no confusion, I'm showing the component values in Figure 2-87.

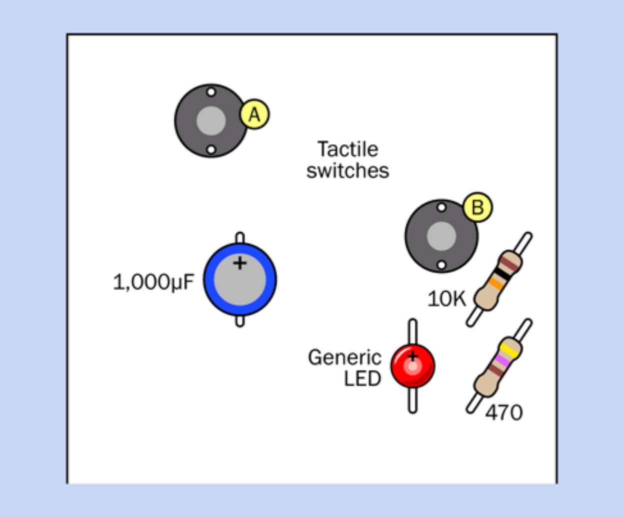

Figure 2-87 *Component values for the breadboarded circuit.*

After you assemble the circuit, first remember to press button B to discharge the capacitor. Now press button A, and—why did the LED just flash and slowly fade away?

Press button A again. This time, little or nothing happens. Evidently, the capacitor has to begin in a discharged state for this to work. So, press button B to discharge it. Now press button A again, and the LED flashes again.

We know that the capacitor started with almost no positive voltage on its lower pin, because it was connected to negative ground through the 10K resistor. We also know that the capacitor started with almost no positive voltage on its upper pin, because button B shorted both sides of the capacitor together. (That was why I asked you to discharge it.)

Then you pressed button A, which applied a sudden positive pulse, and the LED lit up on the other side of the capacitor. The current through the LED had to come from somewhere, and the only explanation is that it came through the capacitor.

Displacement Current

Let's try this again, using your meter instead of the LED and its series resistor. Figure 2-89 shows the breadboard layout, while Figure 2-88 shows the schematic. Discharge the capacitor by pressing button B, and then check the reading on your meter. It should be near 0V.

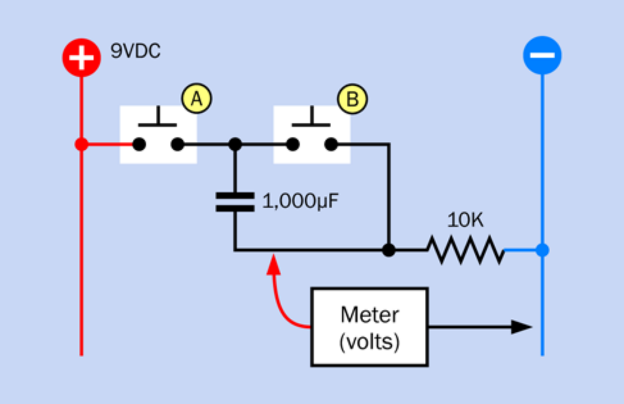

Figure 2-88 *The schematic version of the preceding breadboarded circuit.*

Watch your meter very carefully while you press button A. A digital meter does not respond very quickly, but still you will see a sudden rise in the voltage, after which the voltage gradually diminishes.

When I connected this circuit with an oscilloscope, which can measure and display very rapid changes in voltage, the trace looked look like the curve that I added at the bottom of Figure 2-89. The rise in voltage was so fast, it appeared to be instant.

The way that a capacitor allows a sudden fluctuation in voltage to pass through is well known, and is often used in electronics. But how can it happen?

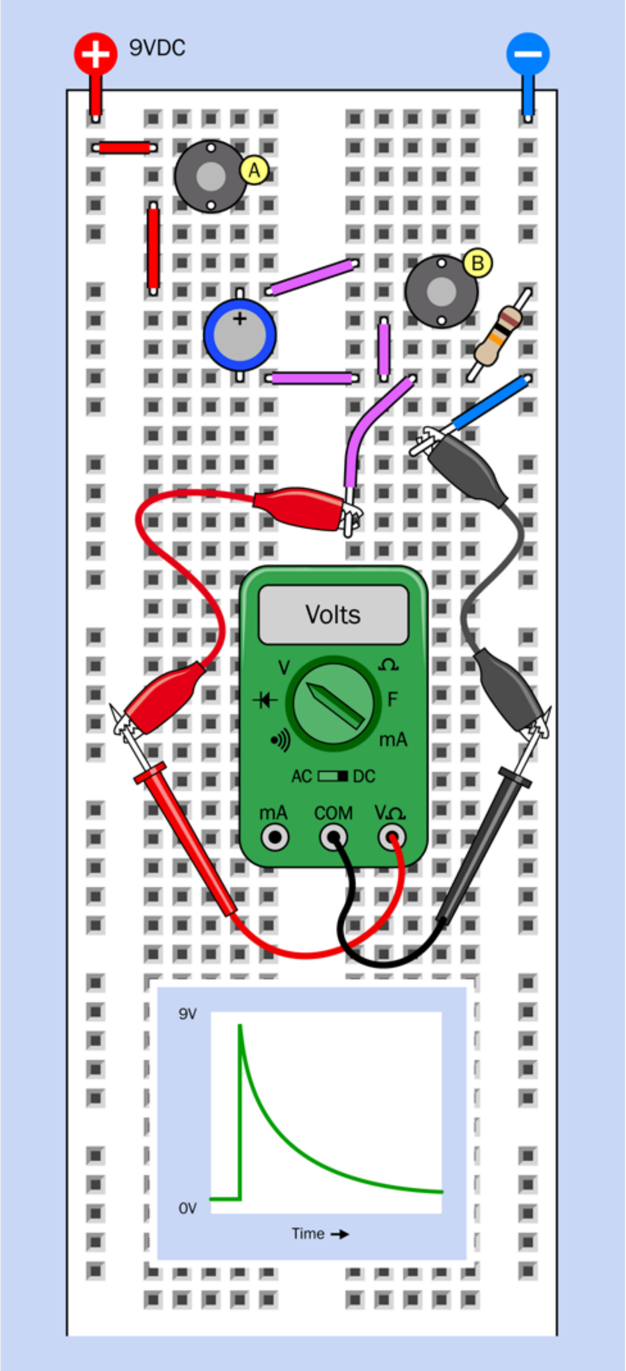

Figure 2-89 *A meter has been substituted for the LED and 470-ohm series resistor that were used in the previous version of this circuit.*

This question interested an early experimenter named James Maxwell, who felt it shouldn't happen; so he developed a theory and invented a phrase to describe what he saw. He called it *displacement current*. This conformed with some theories that he was developing at the time.

Today, there are other theories. Apparently an inrush of current creates a field effect inside the capacitor, and the field effect can induce voltage on the opposite plate. But this concept becomes complicated very quickly, and most books simply say something like, "a capacitor will block DC but will pass fluctuations in voltage."

If you substitute a smaller capacitor, you'll find that it passes a briefer pulse. Remove the meter, put the LED and its 470-ohm resistor back into the circuit, and try 100µF, 10µF, 1µF, and 0.1µF capacitors. By the end of the series, the LED barely flickers.

Alternating Current

If you reverse the circuit, it still works, although current flows in the opposite direction. Figure 2-91 shows the circuit, with the 10K resistor moved to the left, and button A moved to the right. The meter still measures the voltage from the point between the resistor and the capacitor. The schematic in Figure 2-90 shows the same revision.

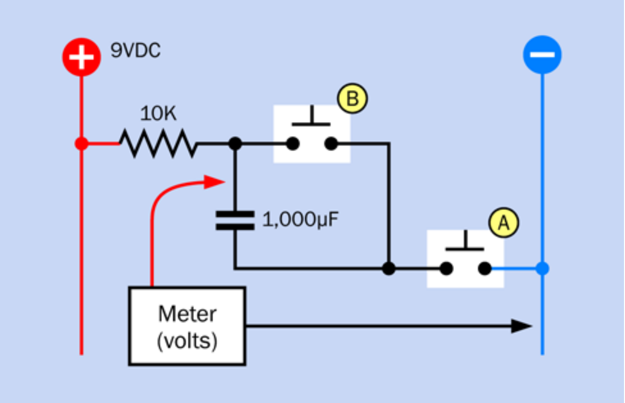

Figure 2-90 *A schematic of the preceding breadboarded circuit.*

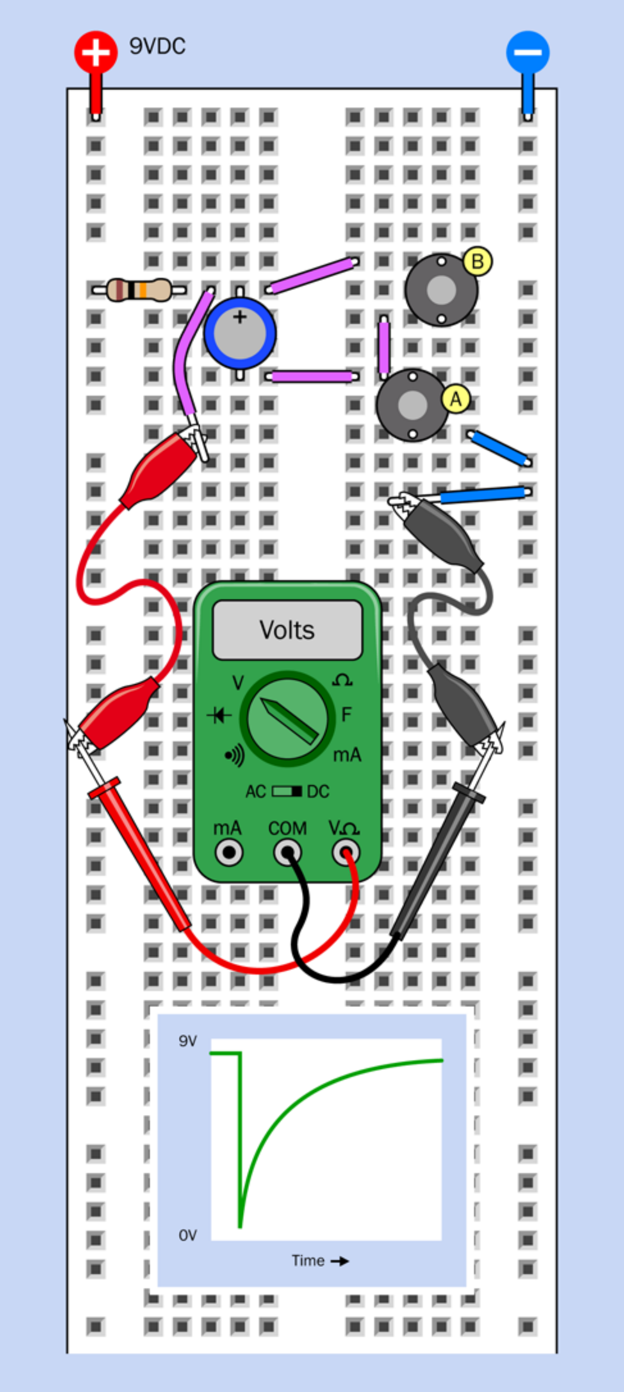
9VDC
B
A
Volts
V
Ω
F
mA
AC
DC
mA
COM
VΩ
9V
0V
Time

Figure 2-91 *The previous circuit has been modified, with the voltages reversed.*

After you press and release button B to discharge the capacitor, the meter measures approximately 9VDC, because the upper pin of the capacitor is connected with the positive bus through the 10K resistor. The capacitor is blocking DC, so it appears to have infinite resistance, and the positive charge has "nowhere to go." This is illustrated in Figure 2-92, which shows how the voltage between a pair of resistors increases when the resistance increases between that point and ground.

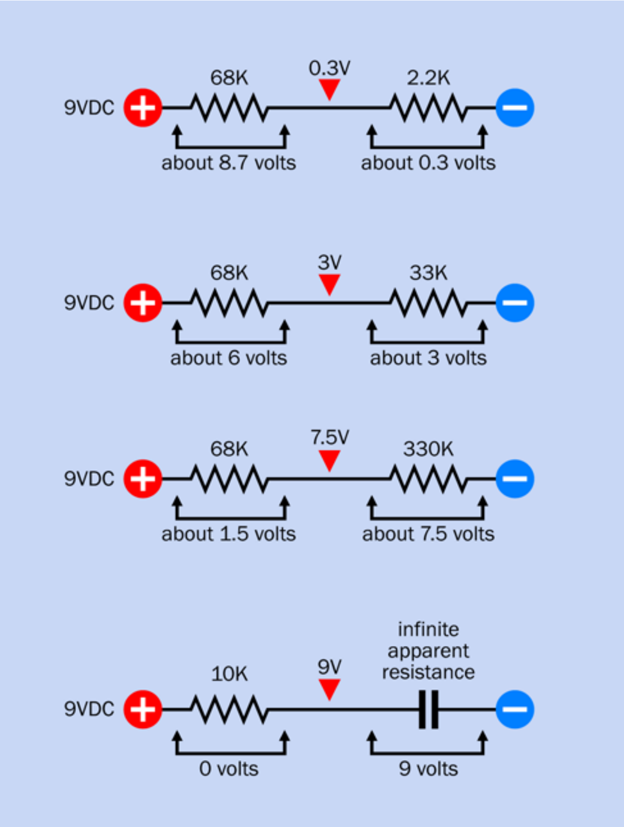
9VDC
68K
0.3V
2.2K
about 8.7 volts
about 0.3 volts
9VDC
68K
3V
33K
about 6 volts
about 3 volts
9VDC
68K
7.5V
330K
about 1.5 volts
about 7.5 volts
infinite
apparent
resistance
9VDC
10K
9V
0 volts
9 volts

Figure 2-92 *When you have a pair of resistors in series, and the left one is connected with the power supply while the right one is connected with negative ground, the voltage between the resistors increases as the value of the righthand resistor increases. A capacitor has an almost infinite effective resistance to DC current.*

However, when you press button A in your breadboarded circuit, this creates a negative pulse. The effective resistance of the capacitor disappears momentarily as the pulse passes through, causing your meter reading to fall. Then the capacitor slowly recharges, just as it did in your very first test in this experiment.

The graph in Figure 2-91 gives a rough idea of how the charge on the capacitor changes.

- A capacitor does block direct current (DC).
- The same capacitor will allow a brief fluctuation to pass through, regardless of which way the current is flowing.
- The capacitor then accumulates a charge, as I described at the beginning of this experiment.

This leads to an important conclusion. Because alternating current (AC) is a rapid series of relatively negative and relatively positive pulses, a capacitor will allow them through.

The size of the capacitor will be important. When you substituted smaller values, you saw that they would only respond briefly. A smaller capacitor will pass high-frequency fluctuations, but will block low-frequency fluctuations—and this behavior is useful in many applications, including audio. You'll see this for yourself in Experiment 29. Bear in mind that audio signals are a form of alternating current, because they fluctuate rapidly.

When a capacitor is positioned in a circuit to pass AC while blocking DC, we call it a *coupling capacitor*. It can allow a signal to travel from one part of a circuit to another, while blocking their DC voltages, which can be completely different. I'll be using this concept when we get to Experiment 11.

Experiment 10: Transistor Switching

Now that you've seen the behavior of capacitors, I'm going to move on to another fundamental component: the transistor. After you learn how that works, you'll see how capacitors and transistors can be used together.

What You Will Need

- Breadboard, hookup wire, wire cutters, wire strippers, multimeter
- Transistor, 2N2222 (1)
- 9-volt battery and connector (1)
- Resistors: 470 ohms (2), 1M (1)
- Trimmer potentiometer, 500K (1)
- Generic LED (1)

The Finger Test

I'm going to be using the 2N2222 transistor, which is the most widely used semiconductor of all time (it was introduced by Motorola in 1962 and has been in production ever since, in one form or another).

Because Motorola's patents on the 2N2222 ran out long ago, any company can manufacture their own version of it. Some versions are packaged in a little piece of black plastic, while others are enclosed in a little metal "can." I showed these two versions in Figure 2-23. For our purposes, either will do. However, note the warning that I included previously about part numbers (see "Essential: Transistors" on page 49). Some 2N2222s are not the same as others, and you need to use the right type.

Plug your transistor into your breadboard with an LED and a 470-ohm resistor, as shown in Figure 2-93. Make sure that the longer lead on your LED is facing to the left, as indicated by the + sign. Also, check that the transistor has its flat side facing to the right. In the unlikely event that you are using a transistor in a metal can, the tab sticking out from the can should point down and to the left.

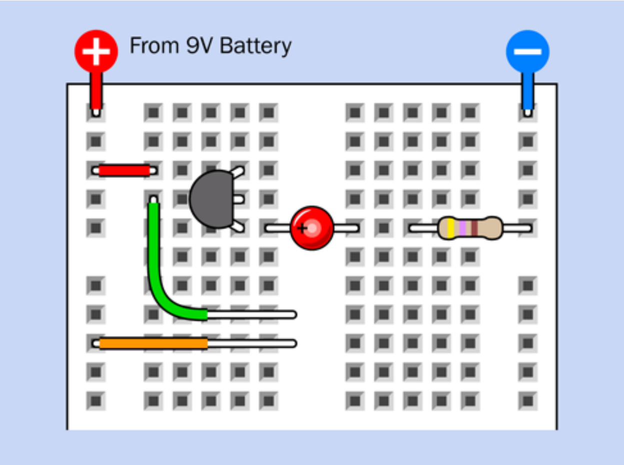

Figure 2-93 *Breadboard setup for your first transistor test.*

Notice that the wires that I have shown as green and orange have had some extra insulation removed. If you are using precut jumpers, you'll have to bend out one leg of each, so that they rest flat on the breadboard.

Now for the fun part. Press your finger to the exposed metal of the green and orange jumpers as shown in Figure 2-94, while you watch the LED. If nothing happens, moisten your finger slightly and try again. The harder you press, the brighter the LED becomes. The transistor is amplifying the tiny amount of current flowing through your finger.

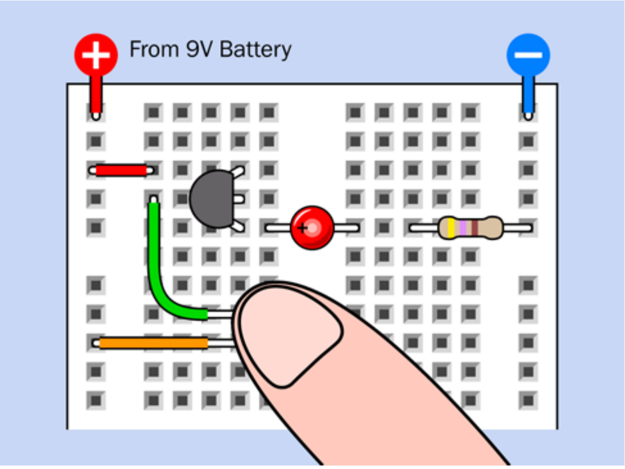

Figure 2-94 *Adding your finger makes the experiment work.*

Caution: Never Use Two Hands

The fingertip switching demo is safe if the electricity only passes through your finger. You won't even feel it, because it's 9 volts DC from a small battery. But it's not a good idea to put the finger of one hand on one wire, and the finger of your other hand on the other wire. This would allow the electricity to pass through your body. Although there is no chance of hurting yourself in this circuit because the current is so small, you should *never get into the habit of allowing electricity to run through you from one hand to the other*. Also, when touching the wires, *don't allow them to penetrate your skin*. This also means that you shouldn't apply voltage to any body ornaments that already pierce your skin.

Inside the Finger Test

Take a look at Figure 2-95, which reveals the connectors inside the breadboard, omitting the ones that are not connected in this experiment. Notice that the bottom lead of the transistor is connected through the breadboard to the LED, and then through the 470-ohm resistor to the negative bus. So, enough current flowed out of the transistor to light the LED.

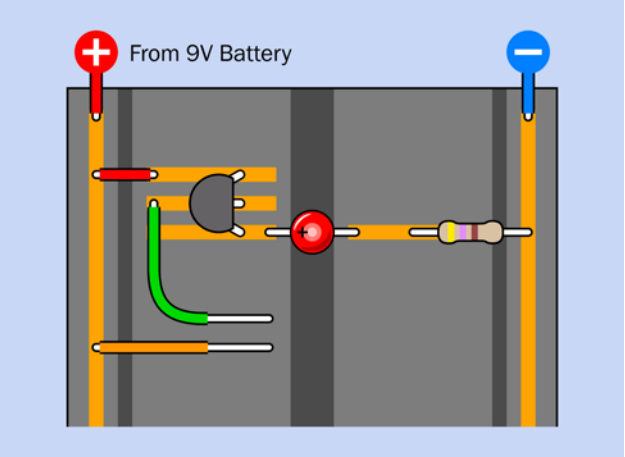

Figure 2-95 *X-ray view of the breadboard from the previous figure.*

Where did this current come from? Well, some electricity flowed in through the skin of your finger, to the center lead of the transistor. But this wasn't enough to light an LED.

There's only one other explanation. The transistor has a third lead, at the top, which is connected to the positive bus. Electricity entered the transistor through this lead. And then, somehow, this flow of current was controlled by the smaller amount of current that flowed through your finger into the center lead of the transistor.

This principle is illustrated in Figure 2-96.

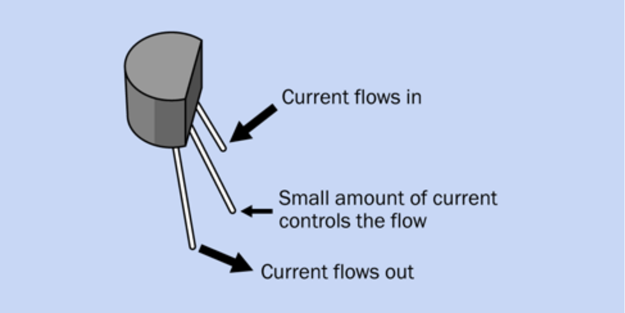

Figure 2-96 *The basic function of an NPN transistor.*

Incidentally, this phenomenon is very different from the behavior of the capacitor that you saw in the previous experiment. A capacitor just passed a quick pulse of electricity. A transistor controls a steady flow.

Fundamentals: Transistor Variants

The component that you're using in this experiment is a *bipolar transistor*. It comes in two variants named *NPN* and *PNP*. The NPN type, which you have been using, contains three layers of silicon, of which two "N" layers have a surplus of negative charge carriers. A third layer, sandwiched between the other two, is the "P" layer, with a surplus of positive charge carriers. I won't be going into a lot of detail about how the transistor works on an atomic level, because in this book I'm more interested in what a transistor does than in the theory that explains how it does it. You can find that information in any technical book, or in many sources online.

The three leads on an NPN bipolar transistor are named the collector, base, and emitter, as shown in Figure 2-97.

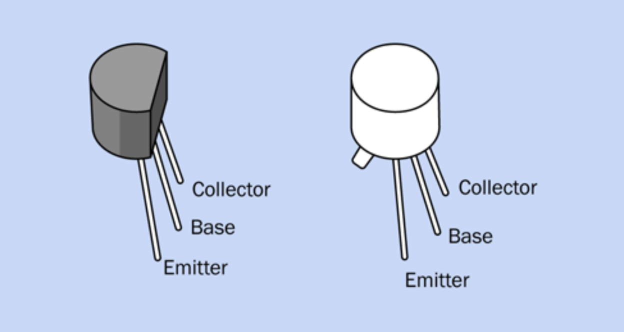

Figure 2-97 *The names of the three leads of an NPN bipolar transistor, supplied in a plastic body (left) and metal can (right).*

- When the base of an NPN transistor is slightly more positive than its emitter, the transistor allows positive current to flow in through the collector and out through the emitter.
- In this way, a very small current entering through the base of the transistor can control a larger current entering through the collector.

A PNP transistor functions oppositely to an NPN transistor. It allows negative current to flow in through the collector and out through the emitter when the base is slightly more negative than the emitter. PNP transistors are sometimes more convenient in a circuit, but are less commonly used. I will not be using them in *Make: Electronics.*

Four variants of the schematic symbol for an NPN transistor are shown in Figure 2-98. They all function the same way. The letters C, B, and E remind you that these connections are for the collector, base, and emitter.

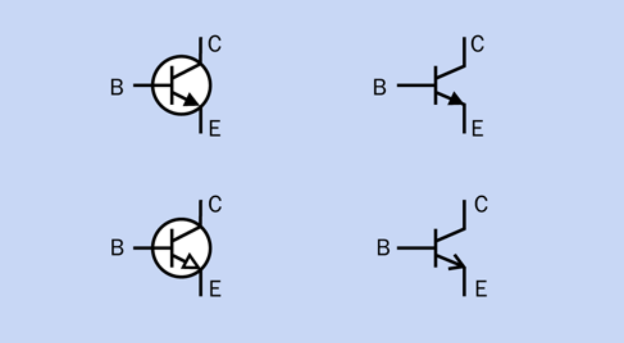

Figure 2-98 *Any of these symbols may be used to represent an NPN transistor.*

Four variants of the schematic symbol for a PNP transistor are shown in Figure 2-99. They all function the same way.

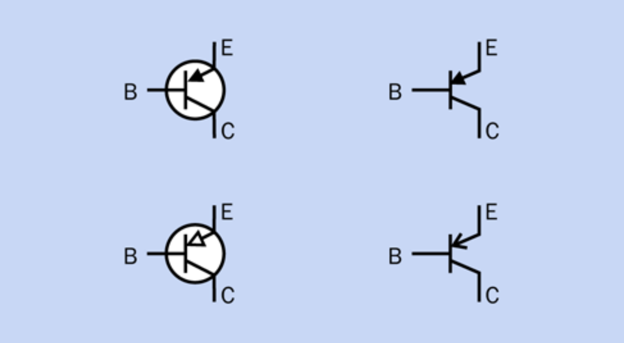

Figure 2-99 *Any of these symbols may be used to represent a PNP transistor.*

It's easy to get the symbols for PNP and NPN transistors mixed up, but there's a simple way to remind yourself which is which. The arrow in the NPN symbol points outward, never in. So, think of "NPN" as being short for "never pointing in."

Adding a Potentiometer

To learn more about the way in which a transistor works, we need a component that is a little more controllable than the tip of your finger. I think a potentiometer will do the job, but not the large type that you used previously. A *trimmer potentiometer* is what I have in mind, as pictured in Figure 2-22.

Even though they are dissimilar in shape and size, they all have three pins. These pins are functionally the same as the three tags that stick out of a large potentiometer of the kind that you used previously. The middle pin always connects with the wiper inside the trimmer, while the other two pins connect with each end of the resistive track inside it. Here is the basic rule that you must always follow:

- When you plug a trimmer into your breadboard, each pin must connect with a separate row of holes on the breadboard.

This rule is illustrated in Figure 2-100. At the top of the figure I have drawn a plan view of three types of trimmers, including the multi-turn type, because even though I am not recommending it, you may find yourself using one, some day. The pins are not visible from above, but I have shown their positions as if you can see them through the component. The positioning varies, but there are always three pins, and they should be 1/10″ apart vertically.

In the lower part of the figure, the two "Yes" examples will work because each pin connects with a different row of holes in the breadboard. The two "No" examples are unacceptable, because a pair of pins will be shorted together by a conductor inside the board.

Having dealt with the trimmer basics, I want you to add a 500K trimmer to your transistor circuit. This is shown in Figure 2-101. Connect the power and use a small screwdriver to rotate your trimmer all the way clockwise, and all the way counterclockwise. Notice that if you start with the LED completely dark, you have to turn the screw on the trimmer a little way before the LED begins to glow.

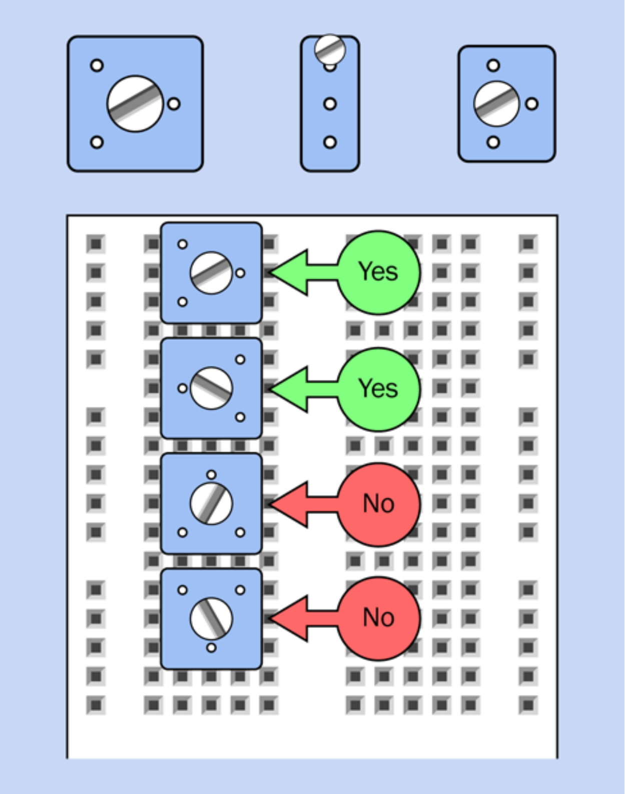

Figure 2-100 *Three types of trimmers, and the correct orientation of their pins.*

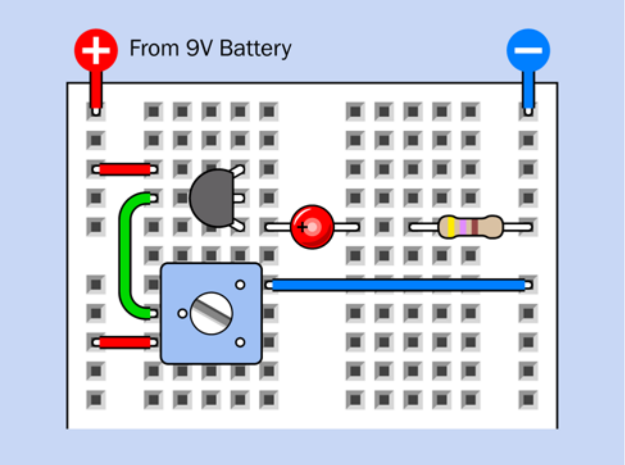

Figure 2-101 *The previous circuit now has a trimmer potentiometer so that you can control the transistor more precisely than was possible with your finger.*

Take a look at the schematic in Figure 2-102, which shows the same connections that you made on the breadboard, but in a format that is easier to understand.

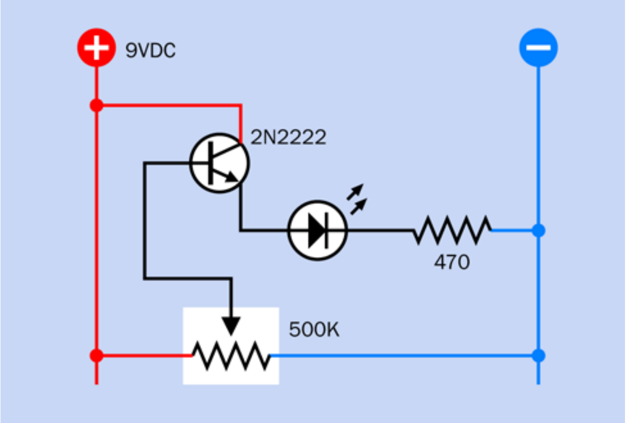

Figure 2-102 *This schematic shows the same breadboard circuit that you built using a trimmer.*

The component values are also shown in Figure 2-103.

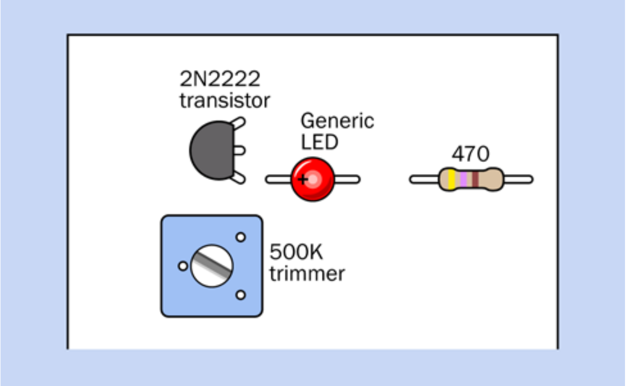

Figure 2-103 *Values for the breadboarded components.*

The potentiometer is connected between the positive bus and the negative bus. In this orientation we call it a *voltage divider*. When the wiper is at one end of the track, it connects directly with the positive side of the power supply. At the other end of the track, it connects directly with negative ground. At positions in between, it divides the voltage. Potentiometers are often used in this way, to provide a full range of voltages.

I mentioned that the LED didn't light up when you first started moving the wiper of the potentiometer from negative to positive. Is this just because the LED isn't getting enough power? Not exactly. The bipolar transistor deducts some of the power as its fee for providing a service. It won't respond unless the voltage at its base is higher than the voltage at its emitter (usually by around 0.7V). In this mode, the transistor is *positively biased*.

The general concept is illustrated in Figure 2-104.

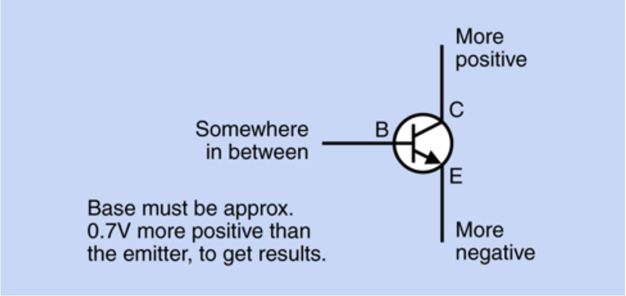

Figure 2-104 *Rule of thumb for using an NPN transistor.*

Voltage and Current

You've seen that the voltage at the base of a bipolar transistor controls the output from the transistor. Does that mean the transistor is amplifying the voltage?

You can discover this for yourself. Take your meter, set it to measure volts, and use a test lead to anchor the negative probe to the negative bus on the breadboard, as shown in Figure 2-105. Touch the red probe to the emitter pin of the transistor, make a note of the voltage, and move the probe to the base pin. I guarantee that the voltage at the emitter will be lower than the voltage on the base.

Readjust the trimmer to a different position, and try again. Regardless of how you change the voltage on the base pin, the emitter pin will always be lower.

Could this be because the 470-ohm resistor doesn't provide much resistance between the emitter of the transistor and the negative bus? Could it be pulling the voltage down?

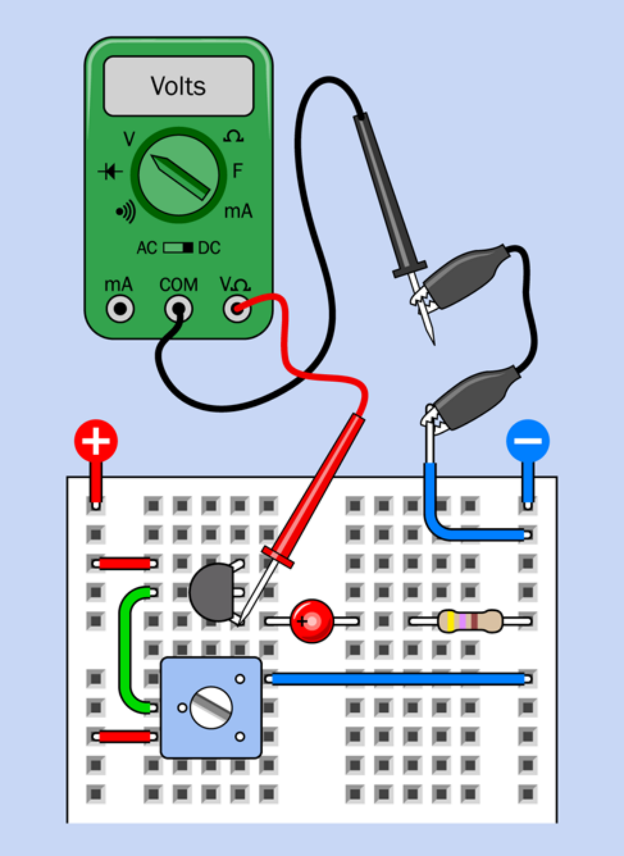

Figure 2-105 *Testing to discover if a transistor amplifies voltage.*

Let's find out. Remove the LED and the 470-ohm resistor, and substitute a 1M resistor between the emitter of the transistor and negative ground. It won't make much difference. The voltage on the emitter will still be lower than the voltage on the base.

If you have the patience to test the *current* into the base and out of the emitter, you would find something very different. This would entail setting your meter to measure milliamps and inserting it into the circuit in each location. Remember, current must pass *through* the meter to be measurable.

But I'll tell you what you would find. This particular transistor amplifies the current entering its base by a factor of more than 200:1. This is called the *beta value* of a transistor, and leads me to a fundamental fact:

- A bipolar transistor amplifies current, not voltage.

In *Make: More Electronics* you can find a lot more on this topic. I'm keeping it brief, here, because this is an introductory book.

Now I'll summarize the facts about bipolar transistors, for your future reference.

Fundamentals: All About NPN and PNP Transistors

A transistor is a *semiconductor*, midway between being a conductor and an insulator. Its effective internal resistance varies, depending on the power that you apply to its base.

All bipolar transistors have three connections: Collector, Base, and Emitter, abbreviated as C, B, and E on the manufacturer's data sheet, which will identify the pins for you.

- NPN transistors are activated by *positive* voltage on the base relative to the emitter.
- PNP transistors are activated by *negative* voltage on the base relative to the emitter.

In their passive state, both types block the flow of electricity between the collector and emitter, just like an SPST relay in which the contacts are normally open. (Actually a transistor allows a tiny bit of current known as *leakage*.)

In a schematic, the orientation of a transistor may vary. The emitter may be at the top and the collector at the bottom, or vice versa. The base may be on the left, or on the right, depending on what was most convenient for the person drawing the schematic. Be careful to look at the arrow in the transistor to see which way up it is, and whether the component is NPN type or PNP type. You can damage a transistor by connecting it incorrectly.

Transistors come in many different sizes and configurations. In a lot of them, there is no way to tell which wires connect to the emitter, the collector, or the base. You may need to check the manufacturer's datasheet to find out.

If you forget which lead is which, many multimeters have a function that will identify the emitter, collector, and base for you. Typically there are four holes labeled E, B, C, and E. When you have the emitter lead of your transistor inserted in either of the holes marked E, the base lead in B, and the collector lead in C, the meter will display the beta value of the transistor. In any other orientation, the meter reading will be unstable, or blank, or zero, or much lower than it should be (almost always below 50, and usually below 5).

Caution: Fragile Component!

Transistors are easily damaged, and the damage will be permanent.

- Never apply a power supply directly between any two pins of a transistor. You can burn it out with too much current.
- Always limit the current flowing between the collector and the emitter of a transistor by using another component such as a resistor, in the same way you would protect an LED.
- Don't apply voltage in reverse orientation. The collector of an NPN transistor should always be more positive than the base, which should be more positive than the emitter.

Background: Transistor Origins

Although some historians trace the origins of the transistor back to the invention of diodes (which allow electricity to flow in one direction while preventing reversal of the flow), the first practical and fully functional transistor was developed at Bell Laboratories in 1948 by John Bardeen, William Shockley, and Walter Brattain.

Shockley was the leader of the team, and had the foresight to see how potentially important a solid-state switch could be. Bardeen was the theorist, and Brattain actually made it work. This was a hugely productive collaboration—until it succeeded. At that point, Shockley started maneuvering to have the transistor patented exclusively under his own name. When he notified his collaborators, they were naturally unhappy about this.

A widely circulated publicity photograph didn't help, in that it showed Shockley sitting at the center in front of a microscope, as if he had done the hands-on work, while the other two stood behind him, implying that they had played a lesser role. A copy of this picture appeared on the cover of *Electronics* magazine (see Figure 2-106). In fact Shockley, as the supervisor, was seldom present in the laboratory where the real work was done.

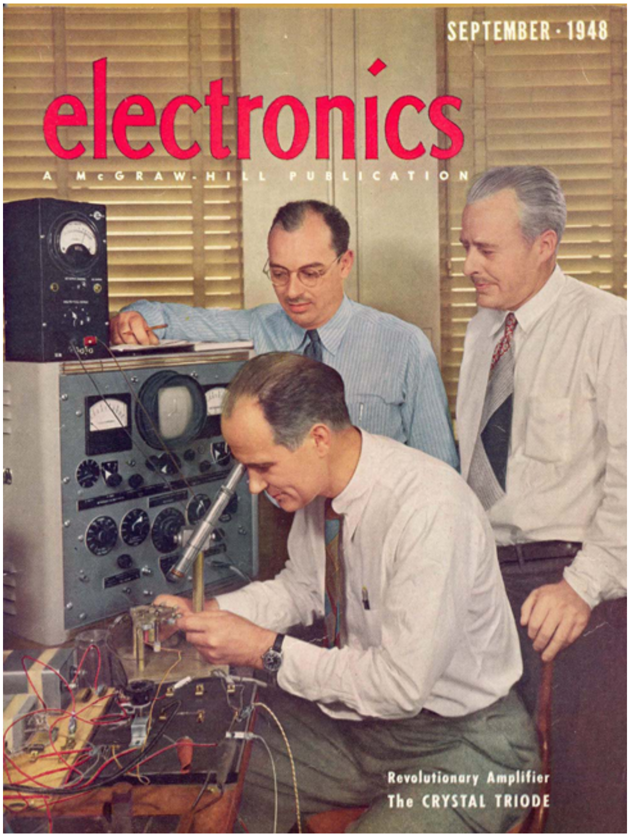

Figure 2-106 *Front, William Shockley. Rear, John Bardeen. Right, Walter Brattain. For their collaboration in development of the world's first working transistor in 1948, they shared a Nobel prize in 1956.*

The productive collaboration quickly disintegrated. Brattain asked to be transferred to a different lab at AT&T. Bardeen moved to the University of Illinois to pursue theoretical physics. Shockley eventually left Bell Labs and founded Shockley Semiconductor in what was later to become Silicon Valley, but his ambitions outstripped the capabilities of the technology of his time. His company never manufactured a profitable product.

Eight of Shockley's coworkers in his company eventually betrayed him by quitting and establishing their own business, Fairchild Semiconductor, which became hugely successful as a manufacturer of transistors and, later, integrated circuit chips.

Fundamentals: Transistors and Relays

One limitation of NPN and PNP transistors is that they require power to perform their function, unlike a relay, which can be passively on or off without any power input at all.

Relays also offer more switching options. Different versions can be normally open, normally closed, or can latch in either position. A relay can contain a double-throw switch, which gives you a choice of two “on” positions. It can also contain a double-pole switch, which makes (or breaks) two entirely separate connections. Single-transistor devices cannot provide the double-throw or double-pole features, although you can design more complex circuits that emulate this behavior.

A table comparing the attributes of transistors and relays appears in Figure 2-107.

	Transistor	Relay
Long-term reliability	Excellent	Limited
Can switch in DP or DT mode	No	Yes
Can switch large currents	Limited	Yes
Can switch alternating current	Usually not	Yes
Triggerable by alternating current	Usually not	Optional
Suitable for miniaturization	Excellent	Very limited
Able to switch at very high speed	Yes	No
Price advantage for high voltage/current	No	Yes
Price advantage for low voltage/current	Yes	No
Current leakage when nonconducting	Yes	No

Figure 2-107 *Principal attributes of relays and transistors, compared.*

The decision to use a relay or a transistor will depend on each particular application.

So much for the theory. Now what can we do with a transistor that’s fun, or useful, or both? We can do Experiment 11!

Experiment 11: Light and Sound

It is time now for your first project that has a function and a purpose. It will culminate with an ultra-simple sound synthesizer.

What You Will Need

- Breadboard, hookup wire, wire cutters, wire strippers, multimeter
- 9-volt battery and connector (1)
- Resistors: 470 (2), 1K (1), 4.7K (4), 100K (2), 220K (2), 470K (4)
- Capacitors: 0.01µF (2), 0.1µF (2), 0.33µF (2), 1µF (1), 3.3µF (2), 33µF (1), 100µF (1), 220µF (1)
- Transistors: 2N2222 (6)
- Generic LED (1)
- Loudspeaker, 8-ohm, 1” or (preferably) 2” (1)

Fluctuations

Figure 2-108 shows the new circuit that I’d like you to create on your breadboard. There’s not a lot of room among the components, so you may find that it’s easier to install them using your pliers instead of your fingers. Count the holes in the board carefully, and double-check to make sure that everything is in exactly the right place.

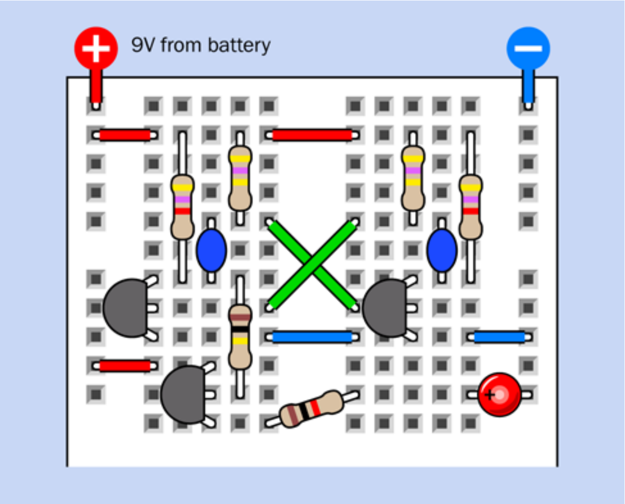

Figure 2-108 *Breadboard layout for an oscillator circuit.*

The component values are shown in Figure 2-109.

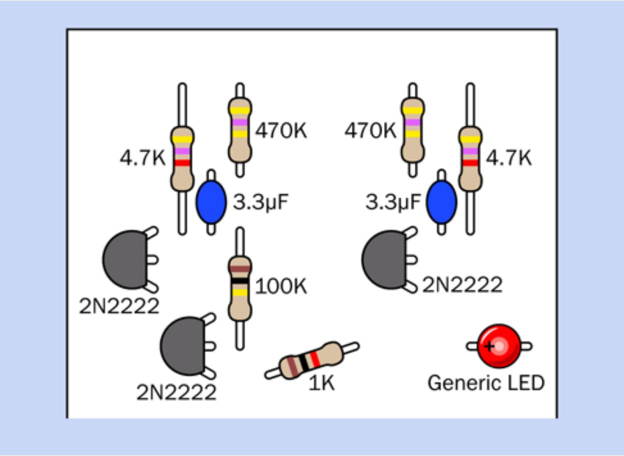

Figure 2-109 *Component values in the breadboarded circuit.*

Apply power, and the LED should flash on for approximately one second, and off for approximately one second.

Is that all? No, we've just begun. First, however, you have to understand how it works. Check the layout in Figure 2-110 if you have difficulty visualizing how the components are connected inside the board. Then look at the schematic in Figure 2-111, and you'll see that the connections between components are the same. I'm going to be using the schematic to explain what's going on.

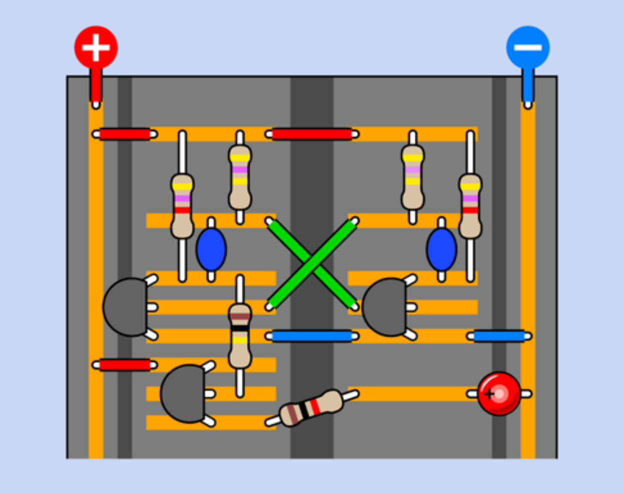

Figure 2-110 *This x-ray view may help to elucidate the interconnections between components.*

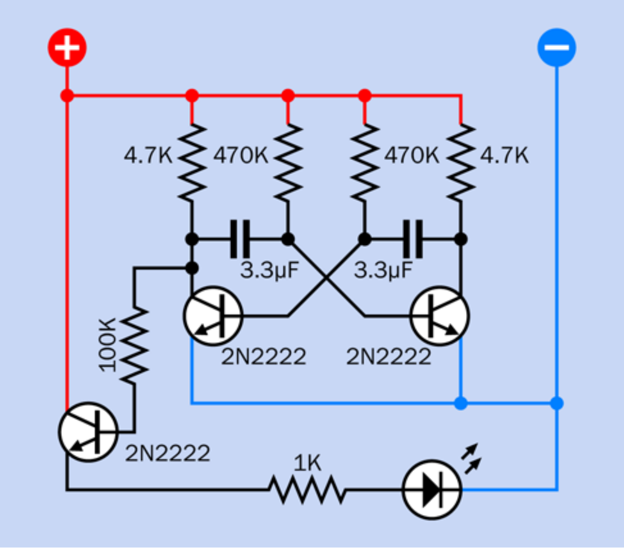

Figure 2-111 *The components in this schematic are in similar locations in the breadboard layout.*

The first thing you'll notice is that it is somewhat symmetrical. Does that mean that the left half and the right half are both doing the same thing? Yes, but not at the same time. In effect, one half turns the LED on, and the other half turns it off.

Understanding this in detail is tricky, because the voltages are constantly fluctuating, and more than one thing is happening at any given moment. However, I have created four snapshots showing the inner workings at intervals, and I'm hoping they will make everything clear.

I have omitted the third transistor and the LED from each snapshot, as they have no role in creating the oscillation in the circuit.

The first snapshot is in Figure 2-112. I have color-coded the wires like this:

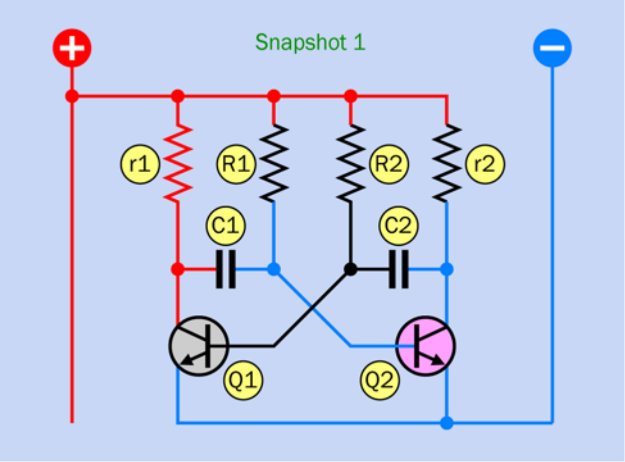

Figure 2-112 *The first of four snapshots showing voltages in the LED-flasher circuit. See text for details.*

- Black conductors and components have an unknown or indeterminate voltage.
- Blue conductors are at a near-zero voltage.
- Red conductors are rising toward positive supply voltage.
- White conductors are briefly pulled down to a more negative voltage (below negative ground), for reasons I'll get to in a moment.

As for the transistors:

- A gray transistor is not conducting electricity from collector to emitter. You can think of it as "switched off."
- A pink transistor is conducting.

The transistors are labeled Q1 and Q2, because this is a common way to identify semiconductors. The little tab sticking out of an old-fashioned metal-can transistor made it look like a letter Q when seen from above, and people got into the habit of identifying transistors this way.

To distinguish between the two sides of the circuit, r1 and R1 are on the left, and r2 and R2 are on the right. Lowercase letters identify the lower-value resistors.

One last note before I begin the explanation. Bear in mind the fundamental behavior of a transistor:

- When current into its base switches it "on," its effective internal resistance drops very low. Consequently, if its emitter is grounded at approximately 0 volts, its collector will also be near 0 volts, and so will anything connected directly to the collector. The base can also be at a relatively low voltage, so long as it is not quite as low as the emitter. You can see this happening to Q2 in Snapshot 1.
- When it switches "off," the effective internal resistance of the transistor rises to at least 5K. Consequently, any component attached to its collector isn't grounded through the transistor anymore, and can accumulate a positive charge.

Step by Step

I'm going to begin at an arbitrary moment while the circuit is already running. After I proceed through the sequence of events, I'll get to the issue of how it begins to oscillate initially.

In Snapshot 1, let's suppose that Q1 has just switched off and Q2 has just switched on. The bottom end of r1 was being grounded through Q1, but now that Q1 is off, the voltage on its collector starts to rise, and this increases the voltage on the left side of C1. The voltage on the base of Q1 also starts to rise, but not as rapidly, because R2 has a higher value. Meanwhile, because Q2 is on, it is sinking current from r2, pulling the voltage down. The base of Q2 is also sinking current through the transistor to negative ground.

That's the setup. What next?

In Snapshot 2, shown in Figure 2-113, the voltage on the base of Q1 increased enough to start it conducting. It is now sinking current from C1 and also through its base, which is why those wires are now blue. The sudden change in the voltage on the left side of C1 momentarily induces an equal drop on its righthand side, as a result of the field effect that can also be described as displacement current, described in Experiment 9. This actually pulls the voltage on the righthand side of C1 below zero, represented by the white wire. This immediately switches off Q2 by inflicting negative bias on its base.

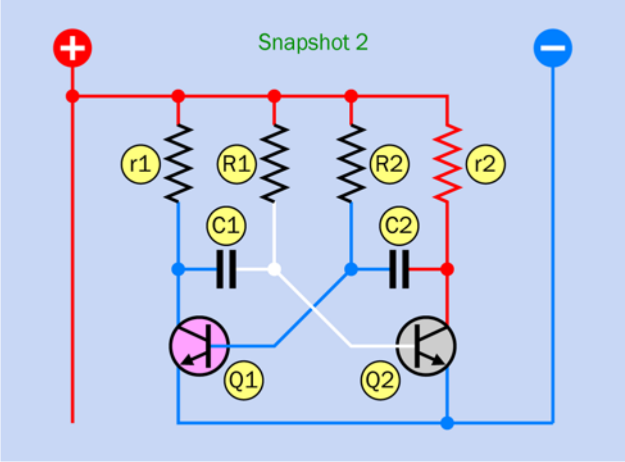

Figure 2-113 *The second snapshot.*

In Snapshot 3, shown in Figure 2-114, Q1 is still switched on while Q2 is still off. This is the mirror image condition of Snapshot 1. C1 has started to charge in the opposite direction, through R1. Gradually this is increasing the voltage on the base of Q2.

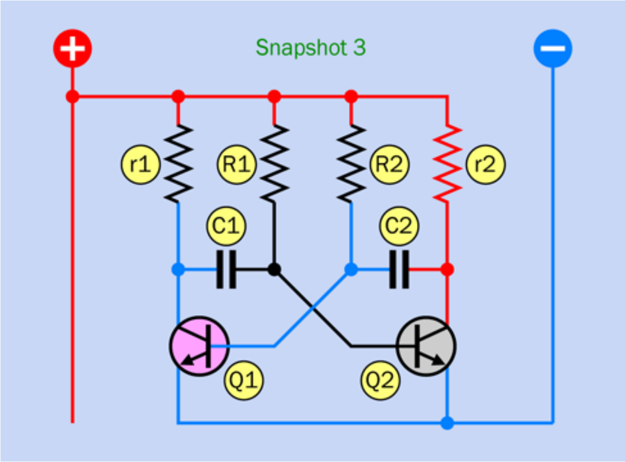

Figure 2-114 *The third snapshot.*

In Snapshot 4, shown in Figure 2-115, Q2 has started conducting, grounding the right side of C2. This transition pulls down the voltage of the left side of C2 below zero, and switches off Q1 by grounding its base. This is the mirror image of Snapshot 2.

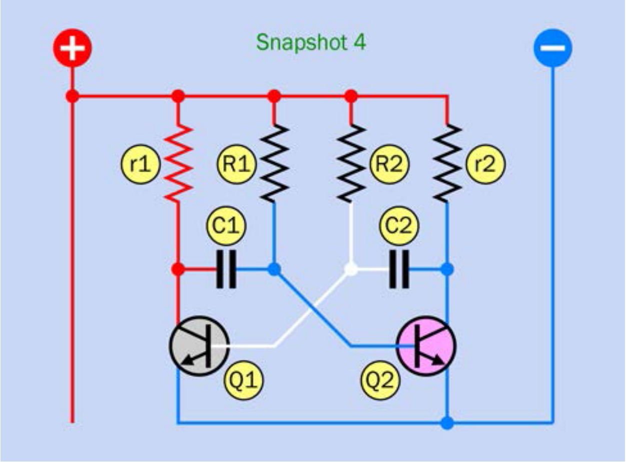

Figure 2-115 *The fourth snapshot. After this, the sequence repeats.*

After the fourth snapshot, the sequence repeats from Snapshot 1. If the extra transistor and LED are connected as in Figure 2-111, the LED should light up during snapshots 1 and 4.

Coupling Capacitor

As you can see, oscillators can be difficult to understand. This particular circuit is extremely common. In fact, if you search Google Images for "oscillator," this is probably what you'll find. Still, many people have difficulty with it.

The key is that in Snapshots 2 and 4, a sudden voltage drop on one side of a capacitor creates an equal voltage drop on the other side—the coupling effect that you saw for yourself in Experiment 9.

But How Does It Start?

Bear in mind that the circuit is basically symmetrical. When you first power it up, why shouldn't both of the transistors be on, or both of them off?

In a perfect world, where two transistors or two resistors can be absolutely identical, the circuit would initialize itself symmetrically. But in reality, there's always a tiny manufacturing difference among resistors and capacitors, causing one transistor to start conducting before the other. As soon as that happens, the circuit is out of balance, and it starts to oscillate as I described above.

Another issue to explain would be, how did I decide where to take an output from the oscillator circuit? In the original schematic, notice that r1 and r2 have much lower values than R1 and R2. This will allow the left side of C1 to charge quickly, almost to the full value of the supply voltage—and the right side of C2 behaves the same way. So, we can get a nice wide voltage range from either of those locations. I chose the one on the left, simply because it enabled an easier layout of additional components in the schematic.

If the circuit is tapped for too much current, this will slow down the process of charging the capacitor, and will affect the timing and balance of the oscillator. Therefore, I fed the signal through a 100K resistor to the base of another transistor. That transistor will draw very little current through its base but will amplify the signal so that you can do something useful with it.

Why So Complicated?

In the first edition of this book, I suggested a flashing-LED project using a component called a Programmable Unijunction Transistor (PUT). Its behavior is much easier to understand, and you only need one of them to get a result. But PUTs are not used much anymore, and some readers complained that you can't buy them easily, while others said that using a PUT was just too old-school.

Really you can still buy PUTs, but they are almost obsolete. Bipolar transistors are still widely used, so I listened to my reader feedback and abandoned the PUT. I considered various alternative oscillator schematics before settling on this one, mainly because this circuit is more common than any of the others. Also, I think all oscillator circuits can be somewhat difficult to understand.

A Processed Pulse

You've learned that two transistors can create a signal that pulses on and off, and a third transistor can amplify it to power an LED. Think back for a moment to the previous experiments. What have you learned from them that you might be able to apply to this one?

We have an output that is fluctuating quite slowly. So, we can make it more interesting by adding an RC network. (See "An RC Network" on page 76 if you need to refresh your memory regarding this concept.)

Take a look at Figure 2-116. The new RC section is at the bottom.

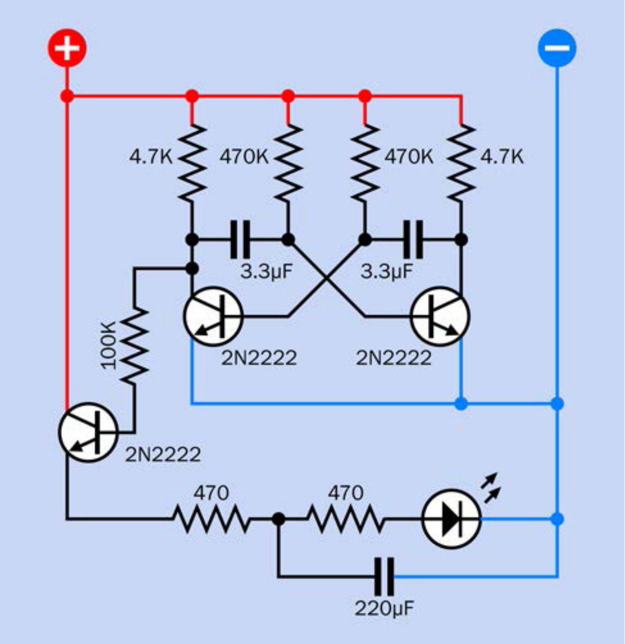

Figure 2-116 *The previous schematic has been modified with one additional resistor and a 220µF capacitor at the bottom, making an RC network.*

In Figure 2-117 the components that have been added or repositioned are in color at bottom-right, while the components that are unchanged have been grayed out.

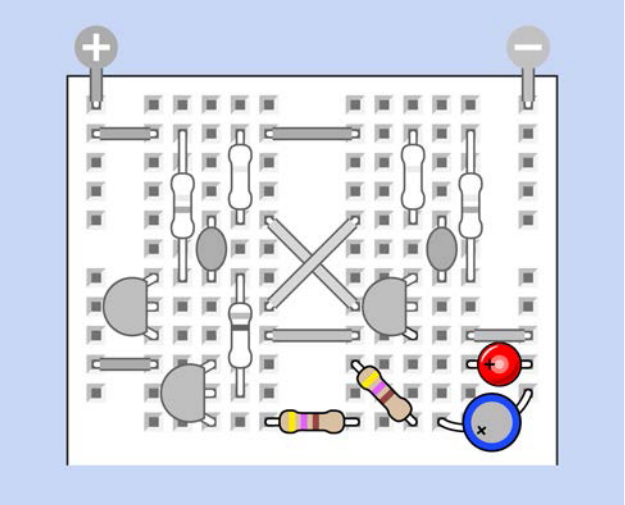

Figure 2-117 *Components in color have been added or repositioned. The new capacitor is a 220µF electrolytic.*

Now when you run the circuit, the LED pulses gently instead of flashing on and off. Can you see why? The capacitor charges through the first 470-ohm resistor, then discharges through the other. Why does this matter? Well, let's suppose you were thinking of making some electronic jewelry. Adjusting the way that it flickers or pulses could be an important aesthetic factor. On old Apple laptops, the logo used to pulse instead of flashing.

Upping the Speed

What else can you do with this circuit? You can easily adjust the speed. Remove the two 3.3µF capacitors, and substitute two 0.33µF capacitors. They should charge about 10 times faster, so the LED should flash 10 times faster, too. Is that what happens?

What if you reduce the capacitor values even more, to 0.01µF? Beyond 50 flashes per second, you have moved from a frequency that you can see to a frequency that you can hear.

How can you change the output of this circuit to be audible instead of visible? Easily! Remove the LED, the 470-ohm resistors, and the 220µF capacitor, and substitute a little loudspeaker, a 100µF coupling capacitor, and a 1K resistor as shown in Figure 2-118. The resistor grounds the emitter of the transistor, because a transistor will only work if its emitter has a defined voltage lower than the voltage at its base. The capacitor blocks the DC component of the signal, while allowing the alternating current through. In the schematic, I only included the parts that have changed. How should they be installed on the breadboard? I think you can figure that out.

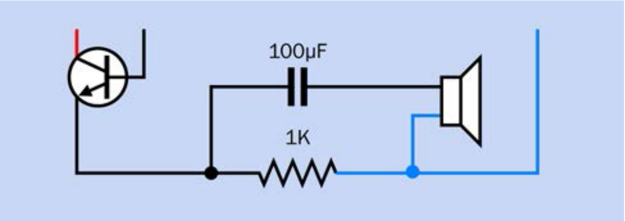

Figure 2-118 *Modifying the circuit to generate audio.*

Still More Mods

Now that you have sound, how about raising the pitch of the sound? Just substitute smaller resistors or capacitors in the oscillator circuit. You could remove the 470K resistors and substitute 220K resistors (or a value in between). Transistors can switch a signal more than a million times per second, so you certainly won't push their limits by making your oscillator run faster. A signal that oscillates 10,000 times per second sounds extremely high-pitched. If you push it to 20,000 times per second, it's above the hearing range of almost all human ears.

How about changing the character of the sound?

In the upper half of Figure 2-119 I've substituted a 1µF coupling capacitor in series with the loudspeaker, instead of the 100µF capacitor that was there previously. The lower value for the capacitor will only pass high frequencies (short pulses), and will deprive the sound of some of its lower resonance.

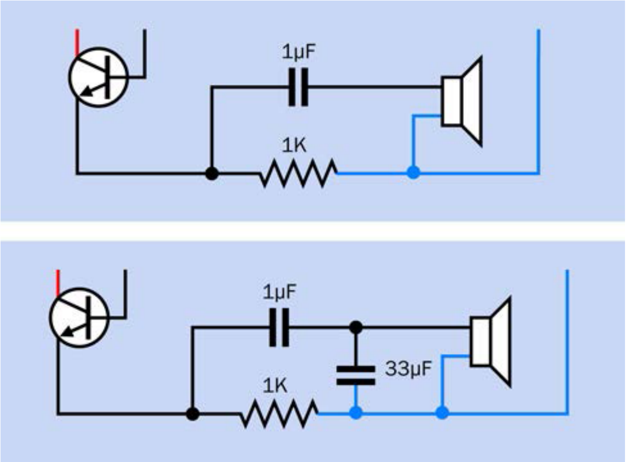

Figure 2-119 *Substituting a smaller value for the coupling capacitor blocks the lower audio frequencies, so that you only hear the higher frequencies. Placing a capacitor so that it bypasses the speaker reroutes the higher frequencies to negative ground, so that you only hear the lower frequencies.*

What if you put a capacitor across the loudspeaker, as shown in the lower part of Figure 2-119? Now the opposite effect occurs, because the capacitor is still passing higher frequencies, but routes them past the loudspeaker. In this orientation, you have a *bypass* capacitor.

These are all simple ways to modify a circuit. If you feel a little more ambitious, you could duplicate the circuit and use one section to control the other.

Restore the original component values from Figure 2-109, so that it runs at the slow original speed. Then use its output to power the duplicate section below it on the breadboard with 0.01µF capacitors to generate an audio frequency. This is shown in Figure 2-120, where the part that you built originally in this project has been grayed out, and the audio section is at the bottom.

The red piece of wire labeled A has been repositioned so that the lower section of the circuit now gets its power from the output of the upper section. The red and blue wire segments labeled B have been added to bridge the gaps that you may have in the buses of your breadboard.

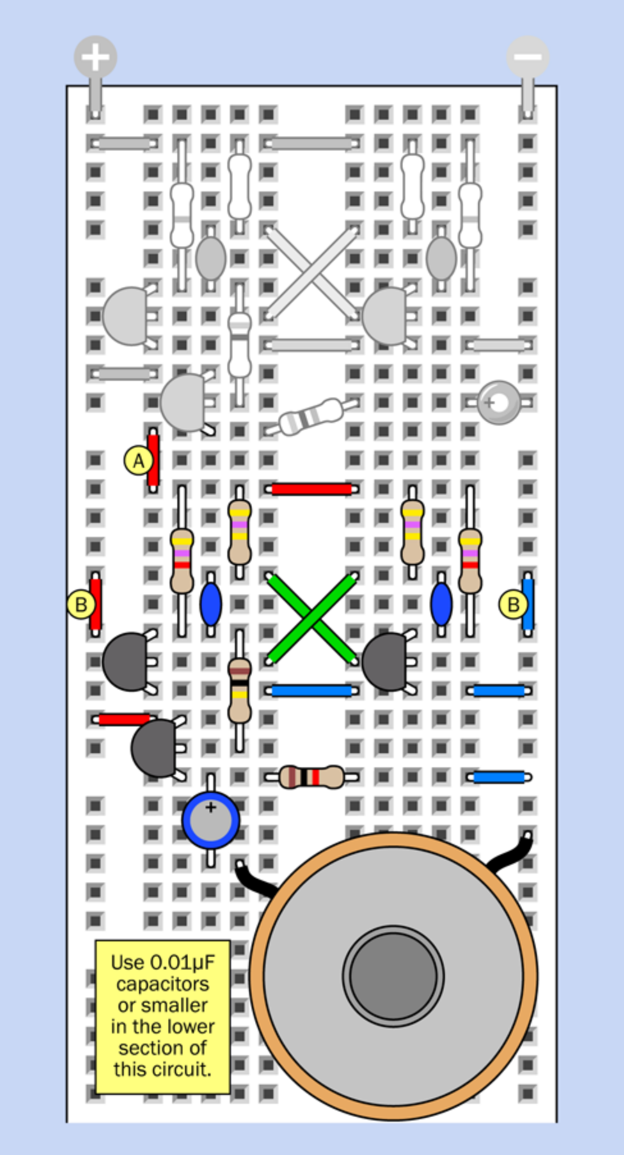
A
B
B
Use 0.01µF capacitors or smaller in the lower section of this circuit.

Figure 2-120 *Driving the audio section of the circuit by using a slower-running duplicate to provide fluctuating power.*

What happens if you change the capacitor or resistor values in the top half of the circuit, to switch the bottom half faster?

What would happen if you take a 220µF capacitor and apply it between various points (either in the top half of the circuit, or the bottom half) and negative ground? You won't damage any components, so feel free to experiment.

Another option is to go back to the "processed pulse" of light created in Figure 2-116, and change the way in which the components are physically joined together. You can remove them from the breadboard and rebuild them into a small wearable object.

I'm going to show you how to do that in Experiment 14. Of course, this will entail some soldering, but learning to solder components together is what I want you to do in Experiment 12, which is next.

Background: Mounting a Loudspeaker

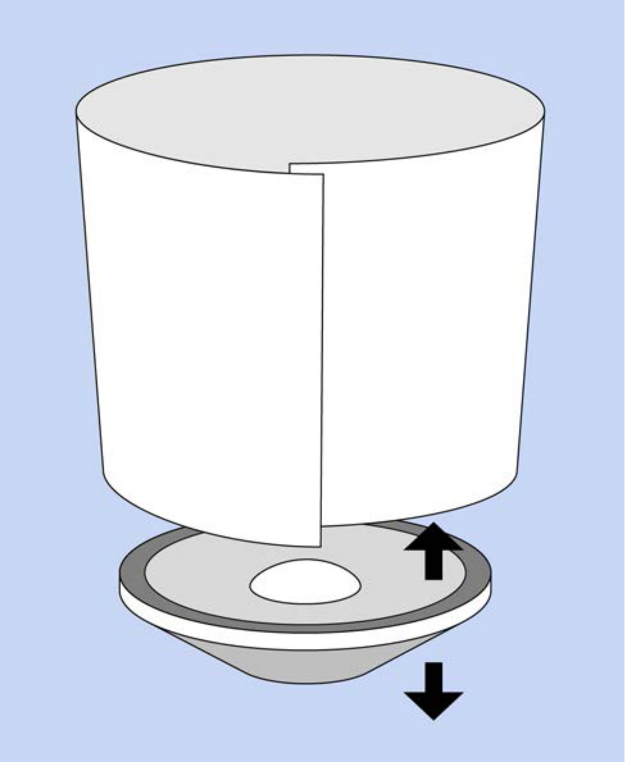

Figure 2-121 *A paper or cardboard tube increases the perceived volume from a loudspeaker.*

The *diaphragm* of a loudspeaker, also known as its *cone*, is designed to radiate sound. However, as it oscillates up and down, it emits sound from its back side as well as its front side. Because the sounds are opposite in phase, they tend to cancel each other out.

The perceived output from a loudspeaker can increase dramatically if you add a horn around it in the form of a tube that focuses the output from the front. For a miniature 1-inch loudspeaker, you can bend and tape a large file card around it. See Figure 2-121.

Better still, mount it in a box with holes drilled in it to allow sound from the front of the speaker to radiate, while the closed back side of the box absorbs sound from the rear of the speaker.

Getting Somewhat More Serious

3

This third chapter of the book applies what you have learned so far to some finished projects. I'll show you how to make a wearable version of the "pulsing glow" project from Experiment 11, and will take you through the initial development process of an intrusion alarm. Looking further ahead, in Chapter Four you'll move into the world of integrated circuit chips.

The tools, equipment, components, and supplies described below will be useful in Experiments 12 through 15, in addition to items that have been recommended previously.

Necessary Items for Chapter Three

As before, when buying tools and equipment, see "Buying Tools and Equipment" on page 324 for a shopping list. If you want kits containing components and supplies, see "Kits" on page 311. If you prefer to buy your own components from online sources, see "Components" on page 317. For supplies, see "Supplies" on page 316.

Essential: Power Supply

You could continue to use 9-volt batteries for all the projects in the book, but I am now classifying an *AC adapter* as essential, because it's so much more convenient. Also, I think it should turn out to be cheaper than buying batteries, as you start to build circuits that take more power.

To convert AC current from an outlet in your home, you have three options:

A *universal adapter* such as the one in Figure 3-1 is the most versatile option, providing a switchable range of outputs. Typically these will include 3V, 4.5V or 5V, 6V, 9V, and 12V. Universal adapters are intended to power small devices such as voice recorders, phones, and media players. They may not deliver a perfectly smooth or accurate DC output, but you should be able to smooth the power yourself with a couple of capacitors, as I will illustrate when we get to a project that uses the adapter.

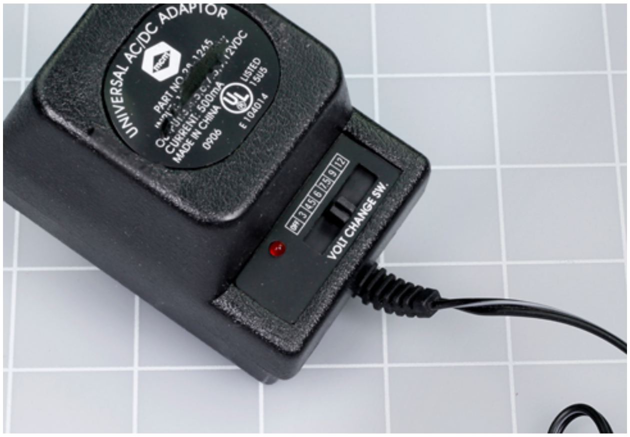

Figure 3-1 *This AC adapter plugs into the wall and allows you a choice of DC voltages by moving a small switch.*

Alternatively you can buy a *single voltage* AC adapter to deliver 9 volts DC, as shown in Figure 3-2. When you start using digital logic chips that require 5 volts, you can convert the 9 volts with a small, cheap component called a *voltage regulator*. (The regulator will also work with power from a 9-volt battery.)

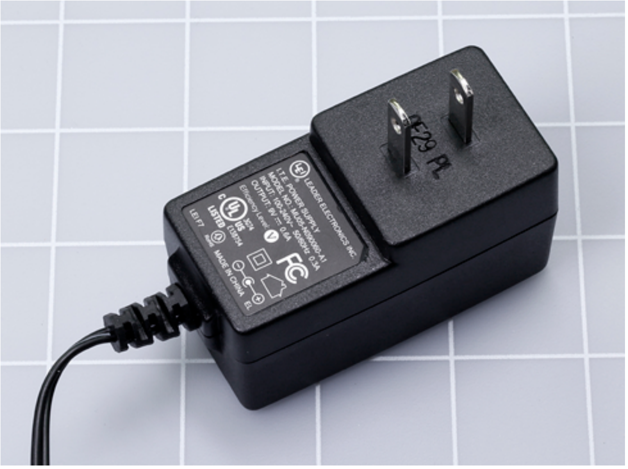

Figure 3-2 *An AC adapter that delivers a fixed 9-volt DC output.*

Your third option is to spend much more money on a proper *benchtop power supply* that provides variable DC outputs of 0V to +15V and 0V to −15V, plus a fixed 5V output. It should also have several breadboards mounted conveniently on top of its box. There is no doubt that this will be very useful if you continue into the field of electronics, but you may not feel sure about that yet.

If you decide to buy a universal adapter, see "Other Components" on page 319 and scroll down to the part referring to Chapter Three of the book, for search instructions.

Whichever kind of adapter you buy, it must have these attributes:

- The output must be DC, not AC. Almost all AC adapters give you a DC output, but there are a few exceptions.
- Outputs must be rated for at least 500mA (which may be written as 0.5A).
- It doesn't matter what kind of plug is on the end of the DC power-output wire, because you'll be cutting it off anyway.
- For the same reason, if a universal adapter offers you a range of output plug adapters, don't worry about what type they are, as you won't be using them.
- Very cheap AC adapters may become unreliable if you draw current up to their rated limit. In the United States, look for the UL symbol indicating certification from Underwriters Laboratories.

Essential: Low-Power Soldering Iron

While a breadboard is indispensable for putting together a circuit quickly to understand how it works, a *soldering iron* is necessary for making permanent electrical connections in a circuit that you want to keep. The soldering iron works by melting a thin wire composed of an alloy called *solder* until it forms a blob around the copper wires or components that you wish to join. When the solder cools, the joint becomes durable.

You don't absolutely have to have a soldering iron. You can complete all the projects in this book just by breadboarding them. However, there is a special pleasure in building something durable, and soldering is a useful skill. For these reasons, I have categorized a soldering iron as "essential."

Personally I like to have a low-power soldering iron for small parts that are vulnerable to excessive heat, and a separate general-duty soldering iron for heavier tasks (described immediately below). Some people prefer to use just one thermostatically controlled soldering iron for everything, but if it's small, I don't think it delivers the heat capacity that is sometimes required, and if it's medium-sized, it isn't so easy to use for delicate work. Also, a thermostatically controlled unit can be expensive.

A low-power soldering iron should be rated at 15W, and the smaller it is, the easier it is to handle. The tip should taper to a slender but rounded point, like a newly sharpened pencil. A plated tip is preferable, although the manufacturer may not state that the tip is plated. A well-used 15W soldering iron is shown in Figure 3-3. The discoloration is a normal consequence of heat, and does not degrade function.

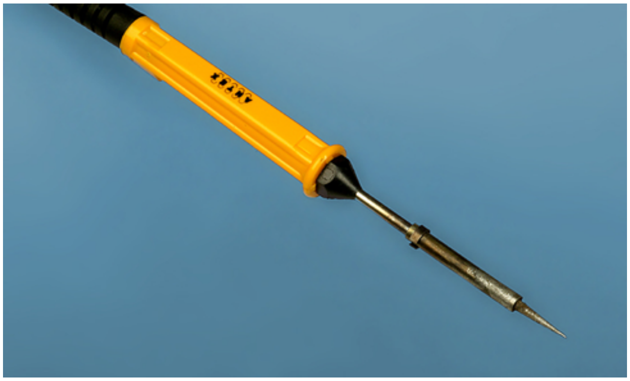

Figure 3-3 *A low-wattage soldering iron designed for precise electronics work.*

Essential: General Duty Soldering Iron

The limited heat capacity of a 15W soldering iron will be insufficient if you need to connect thicker wires, especially to components such as heavy-duty switches with terminals designed to handle substantial current. The terminals may sink heat so rapidly, the low-wattage iron will be unable to create a temperature high enough to melt solder. You may run into a similar situation when you try to attach a wire to a solder tab on a full-size potentiometer.

For these situations, you will need a soldering iron rated for 30 to 40 watts. While it is not required for most of the projects in this book, I do recommend it when you are creating your first solder joints, because the greater heat capacity will make the joints easier. A 30W iron is usually cheaper than a 15W iron, and represents a relatively small additional expense. I think a chisel-shaped tip enables better heat transfer, and since you will not be using this iron for delicate work, you don't really need a pointed tip.

Soldering Iron Terminology

Some soldering irons have a *desoldering pump* built in, to help you undo a solder joint. This is a plunger which you pull with your fingers to suck a little air in through the tip of the iron. I don't think it works very well. In any case, I have only seen it on 30W irons, which are too powerful to be used in many electronics applications.

A soldering iron may have the term "welding" in its product description. You can ignore the inaccurate use of this term, because soldering irons do not do welding in the usual sense.

A few soldering irons are packaged with a *Helping Hand* device that can hold small parts while you are working on them. This combination is worth considering, as it should cost less than if you buy the items separately. The Helping Hand device is described below.

If a soldering iron is sold with some solder included, don't use it unless it is described as electrical solder that has a *rosin core*.

Many soldering irons are described as *pencil type*. The term is not very informative, because it can be applied either to a 15W or a 30W iron.

However, a pencil-type soldering iron is different in appearance from a *pistol-grip* soldering iron, such as the Weller Therma-Boost, shown in Figure 3-4. Some people prefer the ergonomics of this kind of grip, and the Therma-Boost has a nice quick-start feature that enables it to reach working temperature in less than a minute, making it ideal for someone who tends to be impatient. However, pistol-grip irons are all rated at 30W or higher, and tend to be more expensive than the ordinary pencil-type.

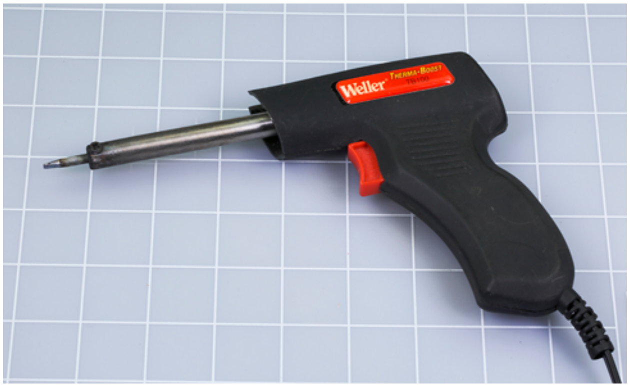

Figure 3-4 *A Weller Therma-Boost 30W soldering iron can be useful when working with heavier wire and larger components.*

Essential: Helping Hand

The so-called *Helping Hand* (sometimes known as a *third hand*) has two alligator clips that can hold components or pieces of wire precisely in position while you join them with solder. Some versions of the Helping Hand also feature a magnifying lens, a wire spiral in which you can rest your soldering iron, and a little sponge that you may use to clean the tip of your iron when it becomes dirty. These additional features are nice but not essential. See Figure 3-5.

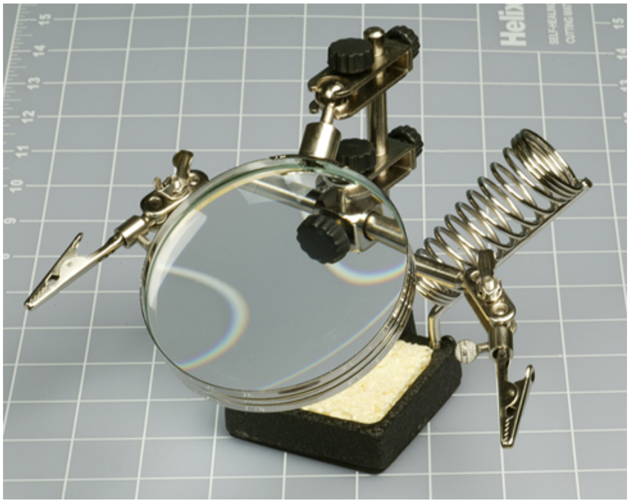

Figure 3-5 *The Helping Hand, with additional accessories.*

Essential: Magnifying Lens

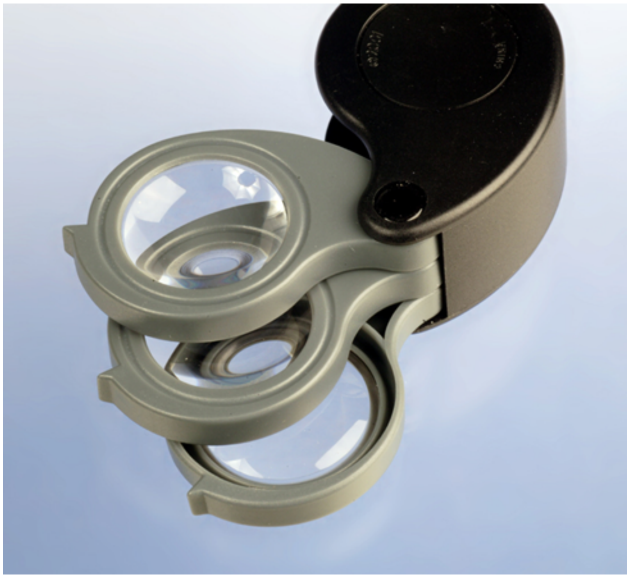

Figure 3-6 *Handheld magnification is essential for inspecting solder joints.*

No matter how good your eyes are, a small, handheld, powerful magnifying lens is essential when you are checking solder joints on perforated board. The three-lens set in Figure 3-6 is designed to be held close to your eye, and is more powerful than the large lens on a Helping Hand, which I do not find very useful. The folding lens in Figure 3-7 stands on your workbench for hands-free operation. Both are available from hobby stores or sources such as eBay and Amazon. Plastic lenses are quite acceptable if you are gentle with them.

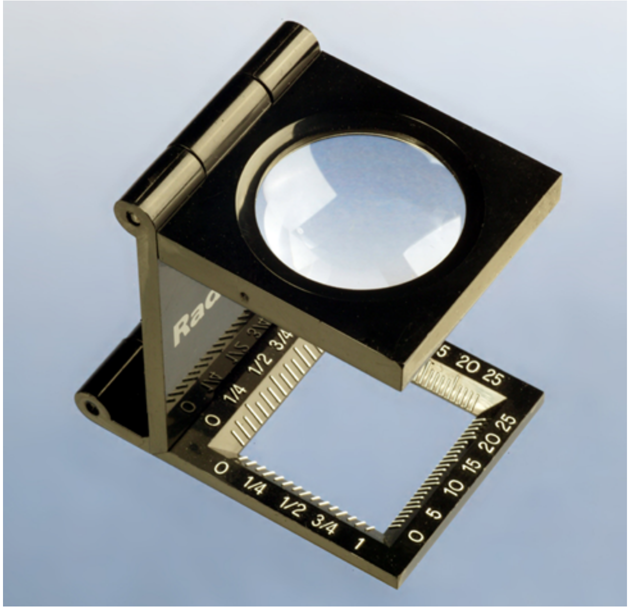

Figure 3-7 *This kind of folding magnifier can stand on your desktop.*

Optional: Clip-on Meter Test Leads

In previous experiments I have suggested that you can grab one of your meter probes with the alligator clip on a test lead, and use the alligator at the other end of the lead to grip a wire or a component.

A more elegant alternative is to buy a pair of “minigrabber” probes with little spring-loaded clips at the end. The Pomona model 6244-48-0 (shown in Figure 3-8) will do the job. However, this is a relatively expensive choice. You may prefer to look for meter leads that terminate in small alligator clips, such as those in Figure 3-9. These are usually the cheapest option. Or, you can just continue to use test leads in the manner that I suggested previously.

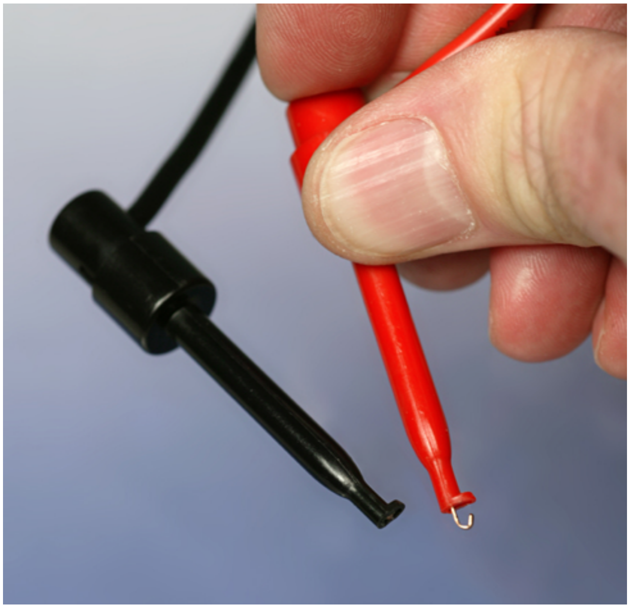

Figure 3-8 *These "minigrabbers" on meter leads will grip wires or component leads.*

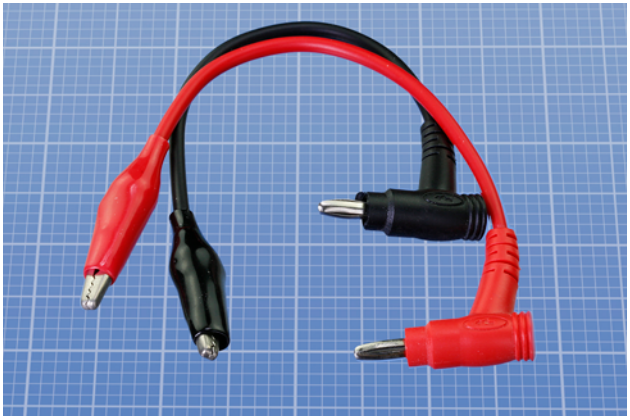

Figure 3-9 *Meter leads that terminate in miniature alligator clips.*

Optional: Heat Gun

If you join two wires with solder, you often need to insulate them. You can use electrical tape (sometimes called insulating tape), but it tends to come unstuck. A better option is heat-shrink tubing, which forms a safe, permanent sheath around a bare-metal joint. To make the tubing shrink, you use a heat gun, which is like a very powerful hair dryer. See Figure 3-10. They're available from any hardware supply source, and I suggest you buy the cheapest one you can find.

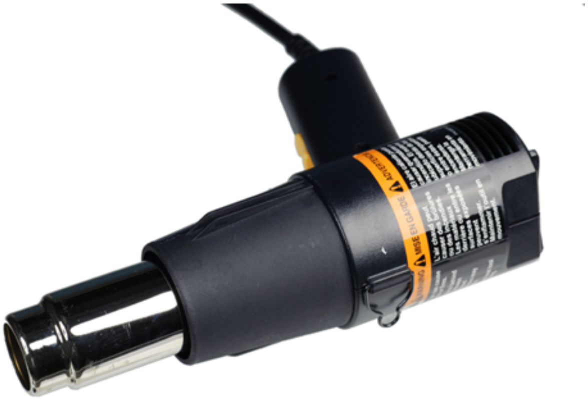

Figure 3-10 *When a heat gun is applied to heat-shrink tubing, the tubing can create a snug, insulated sheath around bare wire.*

For precise work, you may prefer to use a miniature heat gun, such as the one shown in Figure 3-11.

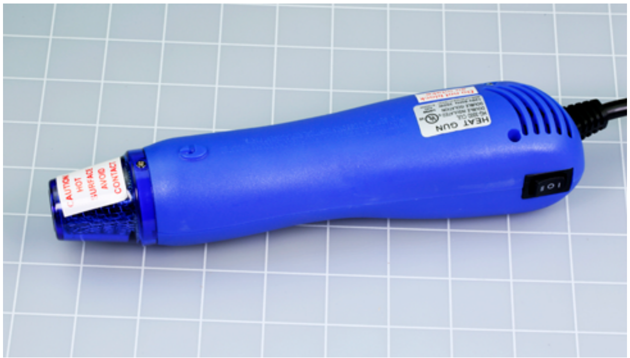

Figure 3-11 *A miniature heat gun is slightly easier to handle than the full-sized type.*

Optional: Desoldering Equipment

A *desoldering pump* is supposed to suck up hot, melted solder when you are trying to remove a solder joint that you made in the wrong place. See Figure 3-12. Some of my readers insist that this should be essential, not optional, but it's a matter of taste. Personally if I make a soldering error, I prefer to snip it out and do it over.

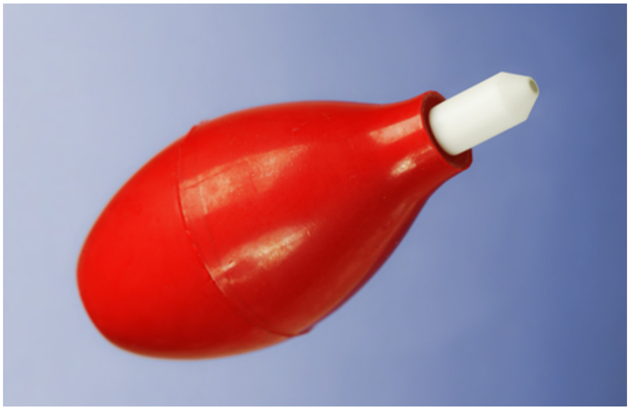

Figure 3-12 *To remove a solder joint, you can suck melted solder into this squeezable rubber bulb.*

Desoldering wick, also known as *desoldering braid*, is intended to soak up solder in conjunction with the solder pump. See Figure 3-13.

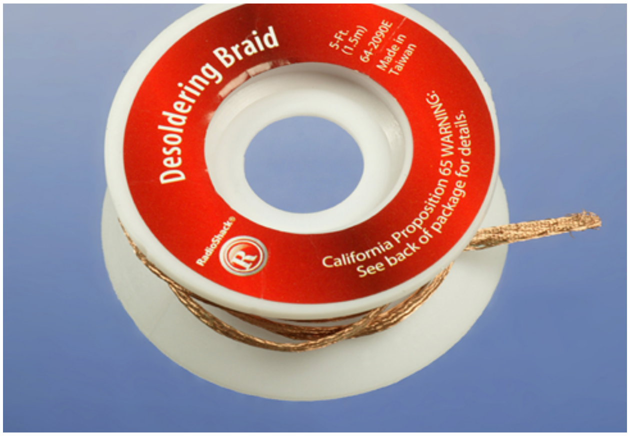
Desoldering Braid
5-Ft. (1.5m)
64-2090E
Made in Taiwan
California Proposition 65 WARNING:
See back of package for details.
RadioShack

Figure 3-13 *An additional option for removing liquid solder is to soak it up in this copper braid.*

Optional: Soldering Stand

You place your hot soldering iron into a stand when you are not using it, in the same way that you would put a kitchen knife into a rack. See Figure 3-14. If you don't want to spend the money on a stand, you can improvise a substitute, such as a piece of steel electrical conduit or even an old tin can nailed to a piece of wood. Or, you can rest the soldering iron on the edge of your work bench, and promise yourself to be veeery careful not to dislodge it. (Been there, done that.) When—not if—the soldering iron falls onto the floor, it will melt synthetic carpet or plastic floor tiles. Knowing this, you may attempt to catch it when you see it fall. If you grab it by the hot end, you will let go of it, so you might as well let it fall on the floor without the intermediate step of burning yourself.

Perhaps a soldering stand should be considered essential.

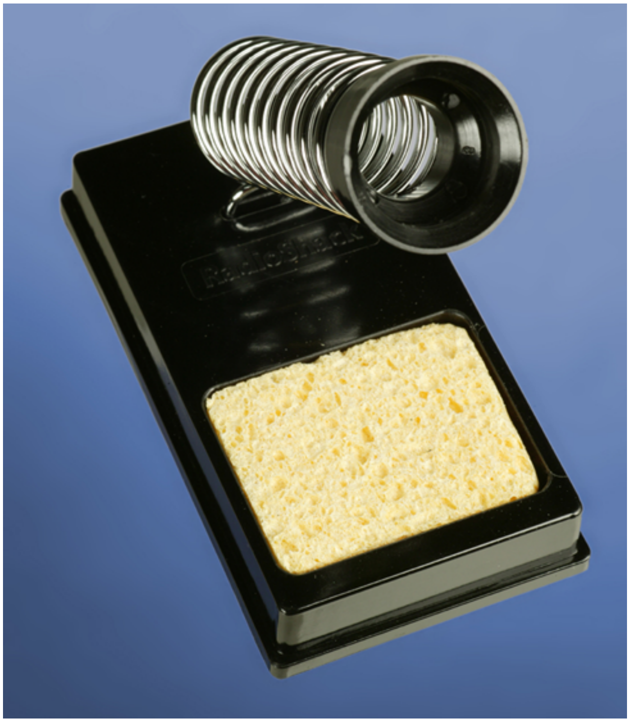

Figure 3-14 *A safe and simple stand for a hot soldering iron. The yellow sponge can be wetted with water to clean the tip of the iron.*

Optional: Miniature Hand Saw

Sooner or later, you may want to mount a finished electronics project in a decent-looking enclosure. For this purpose, you are likely to need tools to cut, shape, and trim thin plastic. For example, you may want to cut a square hole so that you can mount a square power switch in it.

Power tools are overkill for this kind of delicate work. A miniature handsaw (a.k.a. a "hobby saw") is ideal for trimming things to fit. X-Acto makes a range of tiny saw blades. See Figure 3-15.

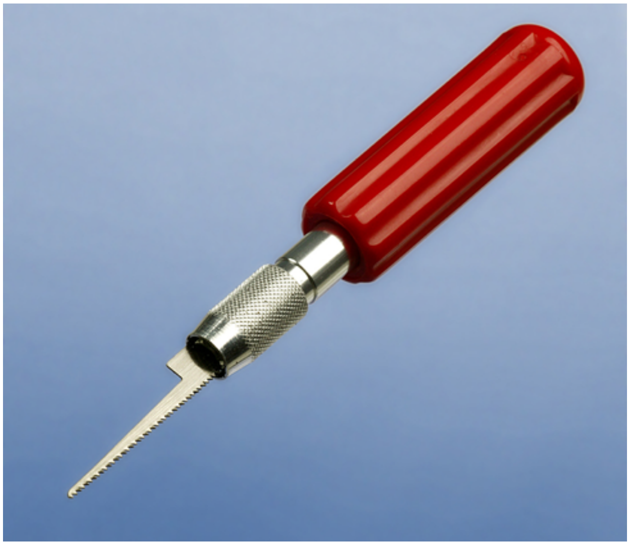

Figure 3-15 *Useful for cutting small holes to mount components in plastic boxes.*

Optional: Deburring Tool

A deburring tool instantly smooths and bevels any rough-sawn plastic or aluminum edge, and also can enlarge holes slightly. This may be necessary because some components are manufactured to metric sizes, which don't fit in the holes that you drill with American bits. See Figure 3-16.

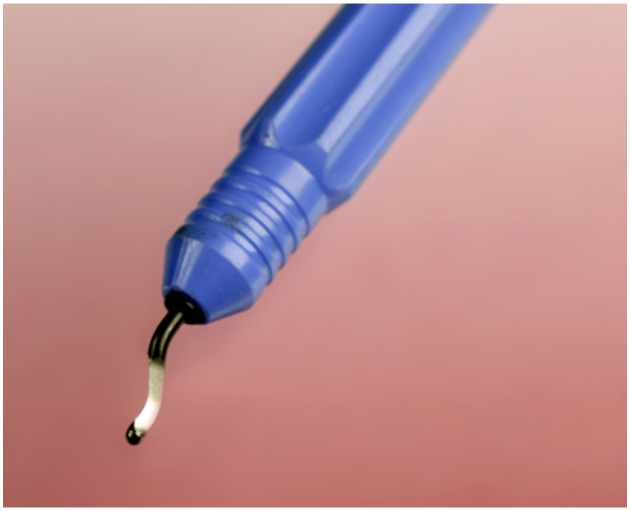

Figure 3-16 *A deburring tool.*

Optional: Calipers

These may seem like a luxury, but are very useful for measuring the external diameter of a round object (such as the screw thread on a switch or a potentiometer) or the internal diameter of a hole (into which you may want a switch or potentiometer to fit). See Figure 3-17. If you choose one that has a digital output powered by a button battery, it will be switchable between metric and inches.

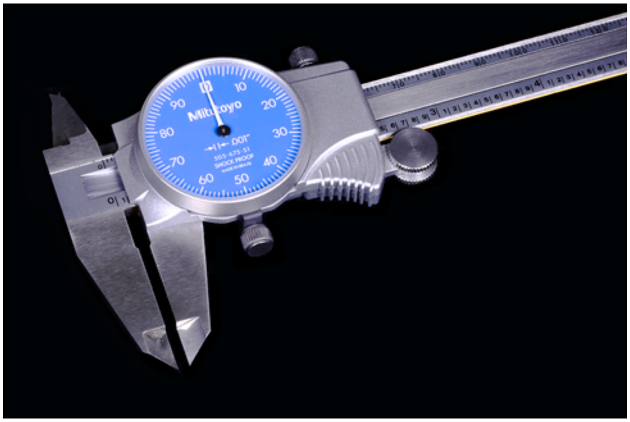

Figure 3-17 *Calipers can measure internal and external diameters.*

Supplies

While many of the tools that I listed are optional, supplies tend to be essential—unless you are absolutely positive that you never want to build a permanent version of a device. The tools and materials to build permanent projects will cost about the same as a month of cable TV. I think they're a worthwhile investment.

Essential: Solder

This is the stuff that you will melt to join components together on a permanent (we hope) basis. It's good to have very thin solder, 0.02" to 0.04" (0.5mm to 1mm) in diameter, for very small components. A range of solder thicknesses is shown in Figure 3-18. For projects in this book, a minimal amount of solder (half an ounce, or maybe three feet) will be sufficient.

Avoid buying solder that is intended for plumbers, or for craft purposes such as creating jewelry. The word "electronics" should appear in the manufacturer's description of suitable purposes for the solder.

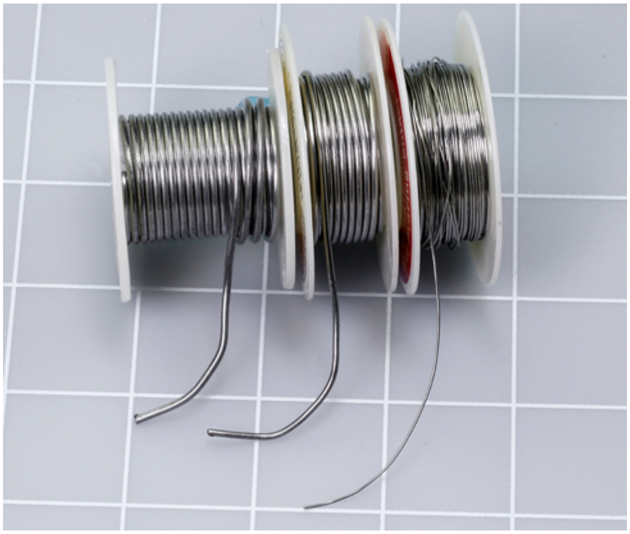

Figure 3-18 *Spools of solder in various thicknesses.*

There is some controversy about using solder that has lead in it. An experienced machinist assures me that this older type of solder makes better, easier joints at a lower temperature, and entails minimal health risk if used sparingly. He points out that lead-free solder has its own problems, as it contains more rosin, which creates more fumes. The issue has roused significant debate online, as you can see if you try this search:

```
lead tin solder safety
```

Personally, I lack the specialized knowledge to make a judgment call. I do know that if you live in the European Union, you're not supposed to use solder with lead in it, for environmental reasons.

What you do definitely need is rosin-core solder intended for electronics use. Whether it contains lead is your choice.

Optional: Heat-Shrink Tubing

For use in conjunction with your heat gun, described previously. It's good to have a range of sizes in any colors of your choice. See Figure 3-19. You slide heat-shrink tubing over a solder joint and then apply heat from a heat gun. The tube shrinks around the joint, insulating it. The diameter after shrinking is typically half of the original diameter, but some tubing has a higher shrink ratio. Different materials offer varying characteristics relating to insulation, abrasion resistance, and other factors. An amazing range of heat-shrink tubing will be found at McMaster-Carr, with details about the various properties. For our purposes, the cheapest tubing should be satisfactory, so long as it is rated for 240 volts (or higher). One bag or box containing an assortment of five or six diameters will be sufficient. You are more likely to use the small sizes than the large sizes.

Figure 3-19 *A variety of heat-shrink tubing.*

Essential: Copper Alligator Clips

These absorb heat when you are soldering delicate components. Don't be deceived by steel clips that are copper-plated; you should really get the solid copper type. Buy as few as possible, as you can reuse them indefinitely. Just two will be enough.

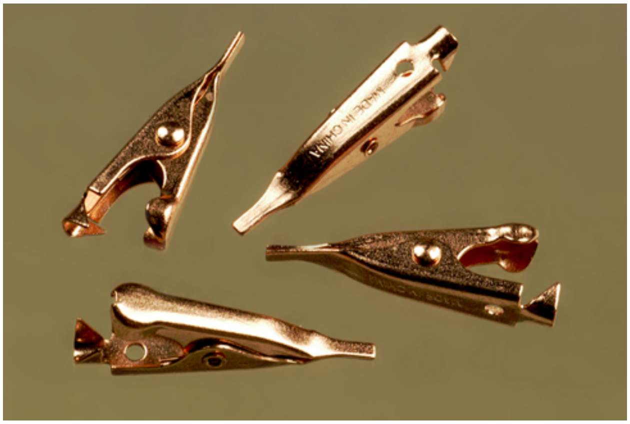

Figure 3-20 *These small copper clips absorb heat to protect components when you're soldering them.*

Optional: Perforated Board

When you're ready to move your circuit from a breadboard to a more permanent location, you'll want to solder it to a piece of perforated board, often known as "prototyping board" and also called "perf board."

The easiest type to use is plated with copper strips on the back in exactly the same layout as the conductors hidden inside a breadboard. This enables you to minimize errors by keeping the same layout of your components when they migrate to the perforated board. See Figure 3-21. Just buy one board, initially.

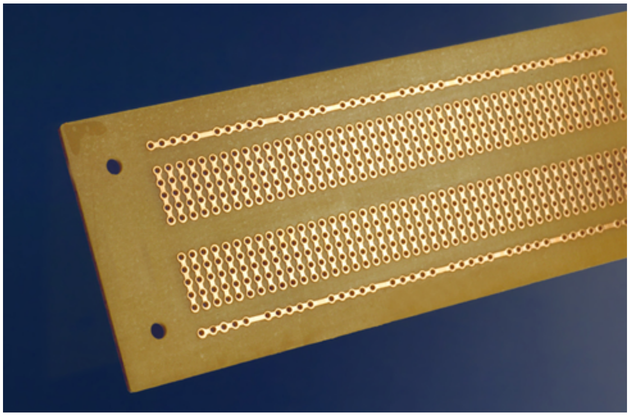

Figure 3-21 *This perforated board has a pattern of copper traces identical to the pattern of conductors inside a breadboard.*

The disadvantage of using the breadboard component layout is that it is not very space-efficient. To compress a circuit to minimal size, you can try point-to-point wiring on a plain perforated board—and I'll show you how this is done in Experiment 14. You may only require a small piece of board, but you can buy a larger piece and cut as much as you need. See Figure 3-22.

Another possibility is to use perforated board with a different pattern of copper traces. *Cut-board*, for instance, features parallel traces that you can cut with a knife where you want to break a connection. Everyone who does much soldering seems to have a favorite type of board configuration, but I think you'll need to get acquainted with the soldering process before you start exploring the options.

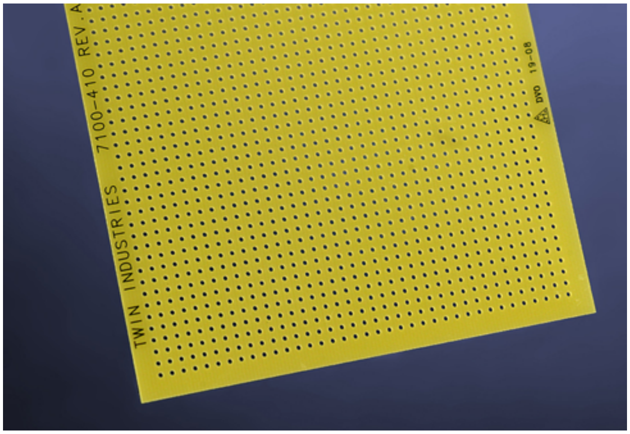

Figure 3-22 *Plain perforated board (with no copper traces) for point-to-point wiring.*

Optional: Plywood

When you use a soldering iron, hot drops of solder tend to fall onto your table or workbench. The solder solidifies almost instantly, can be difficult to remove, and will leave a scar. Consider using a two-foot square of half-inch plywood to provide disposable protection. You can buy it precut at any big-box hardware store.

Optional: Machine Screws

To mount components behind a panel, you need small *machine screws* (more commonly referred to as "bolts"). They look nice if they have flat heads that fit flush against the panel, provided you coutersink the holes. I suggest stainless-steel machine screws, #4 size, in lengths of 3/8" and 1/2", with #4 locknuts of the type that have nylon inserts, so that they won't work loose.

Essential: Project Boxes

A project box is just a small box (usually plastic) with a removable lid. As its name implies, its purpose is to hold one of your electronic projects. You mount your switches, potentiometers, and LEDs in holes that you drill in the lid, and you attach your circuit on a perforated board that goes inside the box. You can also use a project box to contain a small loudspeaker.

For the intrusion alarm project in Experiment 15, you can use a box measuring approximately 6 inches long, 3 inches wide, and 2 inches high.

Essential: Power Connectors

After you finish a project and put it in a box, you'll need a convenient way to supply it with power. You can buy the kind of low-voltage DC plug-and-socket pair pictured in Figure 3-23. They are properly known as "barrel plugs" and "barrel sockets" but you may also find them described as "6VDC plug" and "6VDC socket." They are available in various sizes, but this is unimportant, so long as your connectors are the same size as each other.

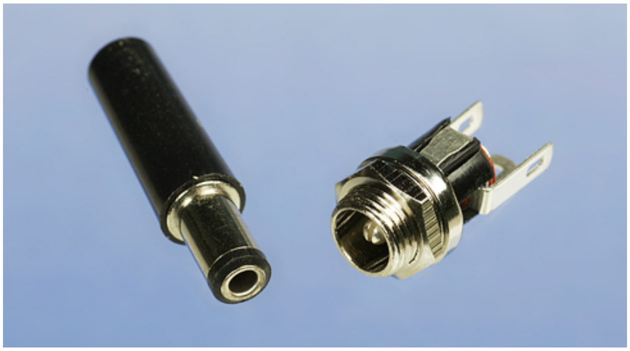

Figure 3-23 *The socket on the right can be mounted on a project box to receive power from the plug on the left, which you can attach to the wire from an AC adapter.*

Optional: Headers

After you have soldered a circuit to perforated board, you will need to connect it with separate switches or pushbuttons. An unpluggable connection is preferable, in case you ever need to fix a fault on the board.

Sometimes known as *single inline sockets and headers*, but also known as *boardmount sockets and pinstrip headers*, small connectors come in strips of 36 or more, and you can snap off as many as you need.

Figure 3-24 shows headers before and after being snapped into small sections. Make sure that the interconnects have a terminal spacing of 0.1 inch, to match your perforated board. (Some interconnects have metric spacing.)

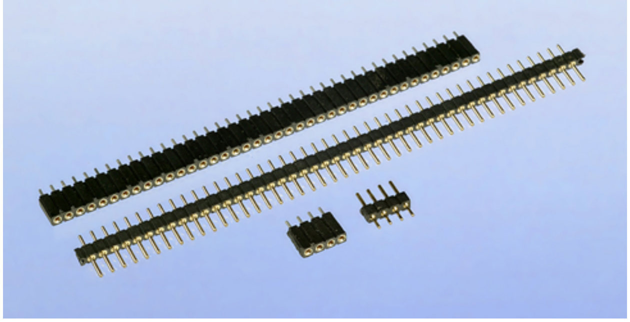

Figure 3-24 *Miniature interconnects are often referred to as "headers."*

Components

Again I must remind you that component kits are available. See "Kits" on page 311. If you prefer to buy your own components from online sources, see "Components" on page 317. In addition to the components that I described at the beginning of Chapter Two (see "Components" on page 47), you're going to need the following.

Diodes

A diode passes electric current in one direction, while blocking it in the opposite direction. The end of the diode that should be more negative is known as the *cathode*. It is marked with a line, as shown in Figure 3-25. The diode on the right, in this photograph, is a 1N4001, which is rated for slightly more current than the 1N4148 on the left. They're cheap, and likely to be useful in the future, so buy 10 of each. The manufacturer is irrelevant.

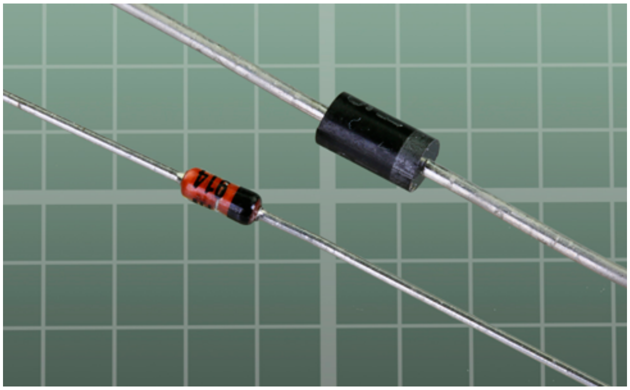

Figure 3-25 *Two diodes. The more-negative end is marked.*

Experiment 12: Joining Two Wires Together

Now you're ready to begin using everything that I have just described, and the first item has to be a soldering iron.

Your adventure into soldering begins with the prosaic task of joining one wire to another, but will lead quickly to creating a full electronic circuit on perforated board. So let's get started!

What You Will Need

- Hookup wire, wire cutters, wire strippers
- 30-watt or 40-watt soldering iron
- 15-watt soldering iron
- Thin solder
- Optional: medium solder
- "Helping Hand" to hold your work
- Optional: shrink-wrap tubing, assorted
- Optional: heat gun
- Optional: a piece of heavy cardboard or plywood to protect your work area from drops of solder

Caution: Soldering Irons Do Get Hot!

Please take these basic precautions:

Use a proper stand (such as the one incorporated in a Helping Hand) to hold your soldering iron. Don't leave the iron lying on a workbench.

If you have infants or pets, remember that they may play with, grab, or snag the wire to your soldering iron. They could injure themselves (or you).

Be careful never to rest the hot tip of the iron on the power cord that supplies electricity to the iron. It can melt the plastic in seconds and cause a dramatic short circuit.

If you drop a soldering iron, don't be a hero and try to catch it.

Most soldering irons have no warning lights to tell you that they're plugged in. As a general rule, always assume that a soldering iron is hot, even if it's unplugged. It may retain sufficient heat to burn you for longer than you expect.

Your First Solder Joint

We'll start with your general-duty soldering iron—the one rated for 30 or 40 watts. Plug it in, leave it safely in its holder, and find something else to do for five minutes. If you try to use a soldering iron without giving it time to get fully hot, you will not make good joints, because the solder may not melt completely.

Strip the insulation from the ends of two pieces of 22-gauge solid hookup wire and clamp them in your Helping Hand so that they cross each other and touch each other, as shown in Figure 3-26.

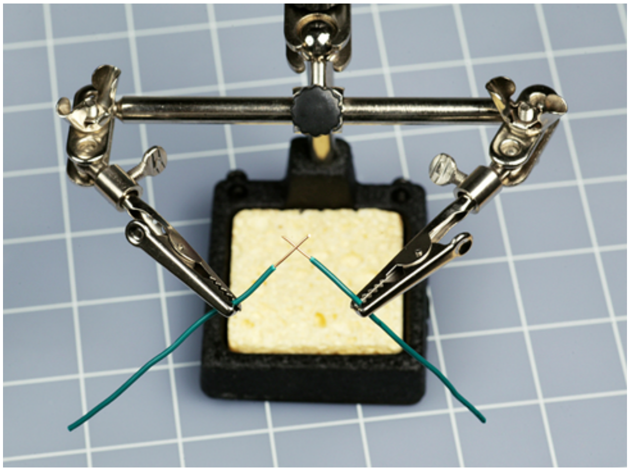

Figure 3-26 *Ready for your first soldering adventure.*

To make sure that the iron is ready, touch a thin piece of solder on the tip of it. The solder should melt instantly. If it melts slowly, the iron isn't hot enough yet.

If the tip of your soldering iron is dirty, you should clean it. The usual procedure is to wet the sponge that is built into your soldering iron stand, and rub the tip of the iron in the sponge. Personally I prefer not to do this, as I believe that getting moisture on the tip causes thermal expansion and contraction, which may open small cracks in the plating of the tip. I rub a piece of crumpled paper over the hot tip of the soldering iron, quickly enough to avoid charring the paper. Then I apply a tiny amount of solder and rub the tip again, until it is uniformly shiny.

When the tip of your soldering iron is clean. follow the steps shown in Figure 3-27, Figure 3-28, Figure 3-29, Figure 3-30, and Figure 3-31.

Step 1. Touch the hot tip of the iron against the intersection of the wires steadily for three seconds to heat them.

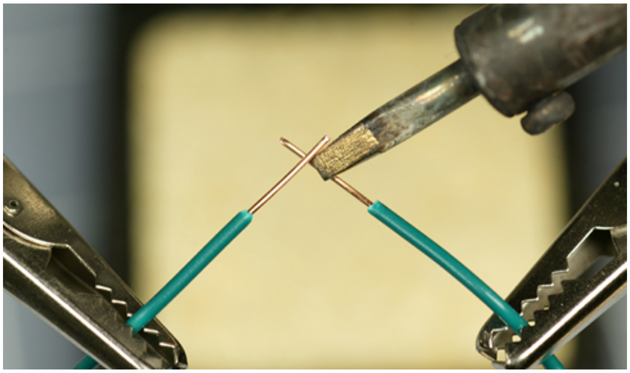

Figure 3-27 *Step 1.*

Step 2. While maintaining the iron in this position, feed a little solder to the intersection of the wires, also touching the tip of the soldering iron. Thus, the two wires, the solder, and the tip of the iron should all come together at one point.

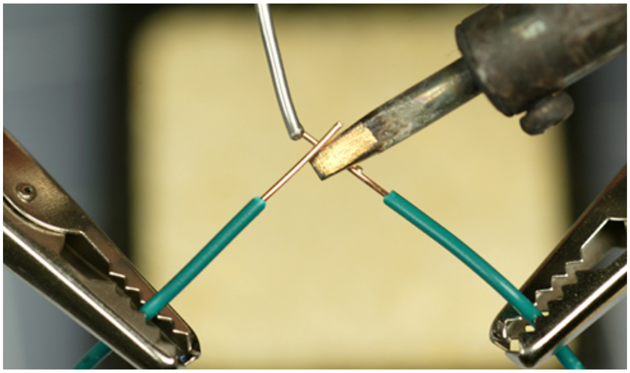

Figure 3-28 *Step 2.*

Step 3. At first, the solder may melt slowly. Be patient.

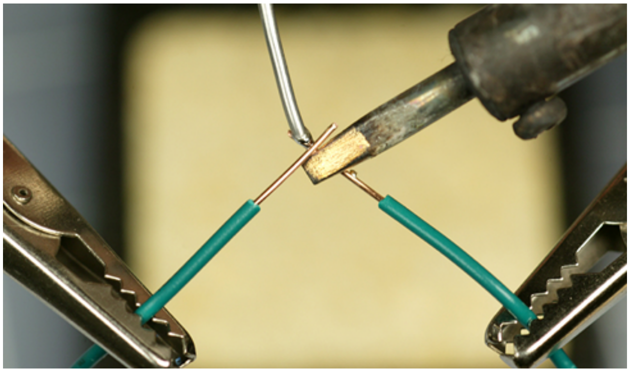

Figure 3-29 *Step 3.*

Step 4. Now you see the solder forming a nice round blob.

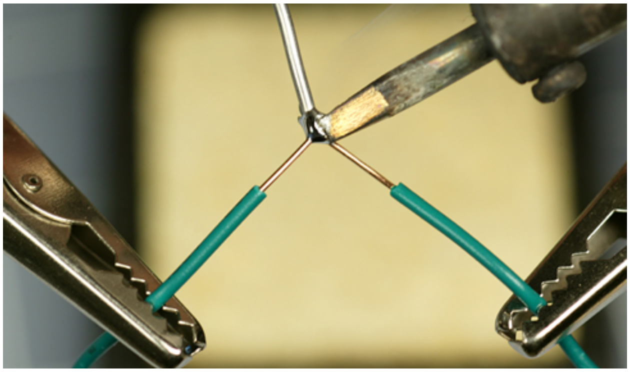

Figure 3-30 *Step 4.*

Step 5. Remove the iron and the solder. Blow on the joint to cool it. Within 10 seconds, it should be cool enough to touch. The completed joint should be shiny, uniform, and rounded in shape

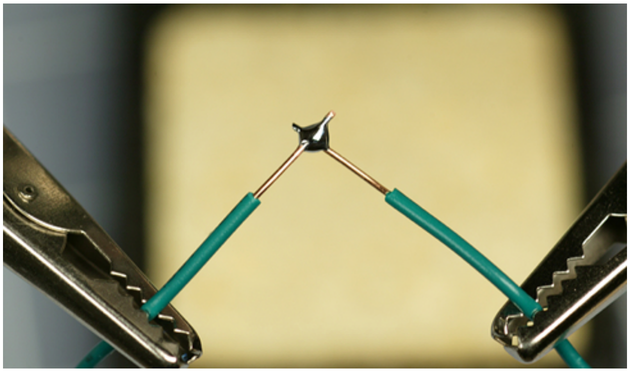

Figure 3-31 *Step 5.*

When the joint has cooled, unclamp the wires and try to tug them apart. Tug hard! If they defeat your best attempts to separate them, the wires are electrically joined and should stay joined. If you didn't make a good joint, you will be able to separate the wires relatively easily, probably because you didn't apply enough heat or enough solder. Figure 3-32 gives you the general idea.

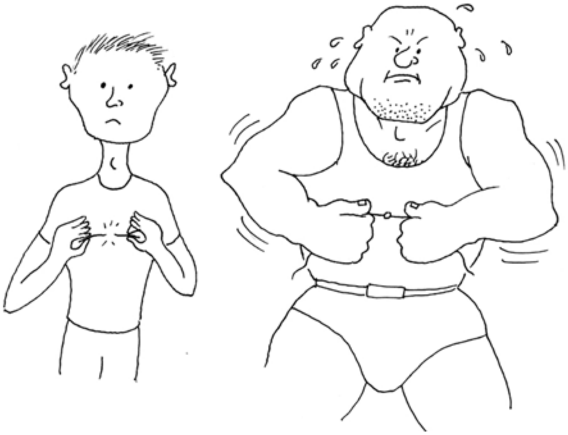

Figure 3-32 *Telling the difference between a bad solder joint (left) and a good one (right) is really not very difficult.*

The reason I asked you to begin by using the higher-powered soldering iron is that it delivers more heat, which makes it easier to use.

The soldering steps can be summarized like this: apply heat to the wires, bring in the solder while maintaining the heat, wait for the solder to start to melt, wait a moment longer for it to form a completely molten bead, and then remove the heat. The whole process should take between four and six seconds.

Background: Soldering Myths

Myth #1: Soldering is very difficult. Millions of people have learned how to do it, and they can't all be more competent than you. I have a lifelong problem with a tremor in my hands that makes it difficult for me to hold small things steadily. I also get impatient with repetitive detail work. If I can solder components, almost anyone should be able to do it.

Myth #2: Soldering endangers your health with poisonous chemicals. You should avoid inhaling the fumes, but that also applies to everyday products such as bleach and paint. You should wash your hands after handling solder, and to do a really thorough job, use a nail brush. But if soldering was a significant health hazard, we should have seen a high death rate among electronics hobbyists decades ago.

Myth #3: Soldering irons are dangerous. A soldering iron is less hazardous than the kind of iron that you might use to iron a shirt, because it delivers less heat. In fact, in my experience, soldering is safer than most activities in a typical home or basement workshop. Of course, that doesn't mean you can be careless. The iron is hot enough to burn your skin if you touch it.

Fundamentals: Eight Soldering Errors

Not enough heat. The joint looks OK, but because you didn't apply quite enough heat, the solder didn't melt sufficiently to realign its internal molecular structure. It remained granular instead of becoming a solid, uniform blob, and you end up with a *dry joint*, also known as a *cold joint*, which will come apart when you pull the wires away from each other. Reheat the joint thoroughly and apply new solder.

Carrying solder to the joint. A leading cause of underheated solder is the temptation to melt some solder onto the iron alone, and then carry the solder to the location where you want to apply it. This means that the wires will be cold when you try to make the solder stick to them. What you should do is touch the soldering iron to heat the wires first, and then apply the solder. This way, the wires are hot and help to melt the solder.

- Because this is such a universal problem, I'll repeat myself. You don't want to put hot solder on cold wires. You want to put cold solder on hot wires.

Too much heat. This may not hurt the joint, but can damage everything around it. Vinyl insulation will melt, exposing too much wire and raising the risk of short circuits. You can easily damage semiconductors, and may even melt the internal plastic parts of switches and connectors. Damaged components must be desoldered and replaced, which will take time and is a big hassle. If your attempt at soldering isn't working for some reason, pull back, pause, and allow everything to cool down a little before you try again.

Not enough solder. A thin connection between two conductors may not be strong enough. When joining two wires, always check the underside of the joint to see whether the solder penetrated completely.

Moving the joint before the solder solidifies. This may create a fracture that you won't necessarily see. It may not stop your circuit from working, but at some point in the future, as a result of vibration or thermal stresses, the fracture can separate just enough to break electrical contact. Tracking it down will then be a chore. If you clamp components before you join them, or use perforated board to hold the components steady, you can avoid this problem.

Dirt or grease. Electrical solder contains rosin that cleans the metal that you're working with, but contaminants can still prevent solder from sticking. If any component looks dirty, clean it with fine sandpaper before joining it.

Carbon on the tip of your soldering iron. The iron gradually accumulates flecks of black carbon during use, and they can act as a barrier to heat transfer. Clean the tip, as described previously.

Inappropriate materials. Electrical solder is designed for electronic components. It will not work with aluminum, stainless steel, or various other metals. You may be able to make it stick to chrome-plated items, but only with difficulty.

Failure to test the joint. Don't just assume that it's OK. Always test it by applying manual force. If you can't get a grip on the joint, slip a screwdriver blade under it and flex it just a little, or use small pliers to try to pull it apart. Don't be concerned about ruining your work. If your joint doesn't survive rough treatment, it wasn't a good joint.

Of the eight errors, dry/cold joints are by far the worst, because they are easy to make and can look OK.

Background: Soldering Alternatives

As recently as the 1950s, connections inside electronic appliances such as radio sets were still being hand-soldered by workers on production lines. But the growth of telephone exchanges created a need for a faster way to make large numbers of rapid, reliable point-to-point wiring connections, and *wire wrap* became a viable alternative.

In a wire-wrapped electronics project, components are mounted on a circuit board that has long, gold-plated, sharp-cornered square pins sticking out of the rear. Special silver-plated wire is used, with an inch of insulation stripped from its ends. A manual or power-driven wire-wrap tool twirls the end of a wire around one of the pins, applying sufficient tension to "cold-weld" the soft silver plating of the wire to the pin. The wrapping process exerts sufficient pressure to make a very reliable joint, especially as seven to nine turns of wire are applied, each turn touching all four corners of the pin.

During the 1970s and 1980s, this system was adopted by hobbyists who built their own home computers. A wire-wrapped circuit board from a hand-built computer is shown in Figure 3-33. The technique was used by NASA to wire the computer in the Apollo spacecraft that went to the Moon, but today, wire-wrapping has few commercial applications.

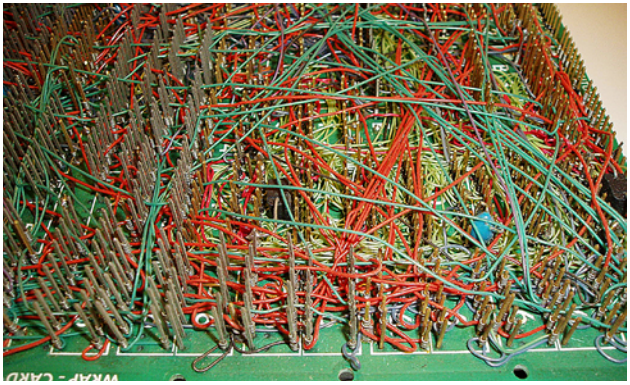

Figure 3-33 *This picture shows some of the wire-wrapping in Steve Chamberlin's custom-built, retro 8-bit CPU and computer. Connecting such a network of wires with hand soldering would have been unduly time-consuming and prone to faults. Photo credit: Steve Chamberlin.*

The widespread industrial use of "through-hole" components, such as the chips on early desktop computers, encouraged development of wave soldering, in which a wave or waterfall of molten solder is applied mechanically to the underside of a preheated circuit board where chips have been inserted. A masking technique prevents the solder from sticking where it isn't wanted.

Today, surface-mount components (which are significantly smaller than their through-hole counterparts) are glued to a circuit board with a solder paste, and the entire assembly is then heated, melting the paste to create a permanent connection.

Your Second Solder Joint

Time now to try your 15W soldering iron. Once again, you must leave it plugged in for a good five minutes to make sure it's hot enough. In the meantime, don't forget to unplug your other soldering iron, and put it somewhere safe while it cools.

Use thin solder to make this joint. It will take less heat from the less-powerful soldering iron.

This time I'd like you to align the wires parallel with each other. Joining them this way is a little more difficult than joining them when they cross each other, but it's a necessary skill. Otherwise, you won't be able to slide heat-shrink tubing over the finished joint to insulate it.

Five steps to create this joint are shown in Figure 3-34, Figure 3-35, Figure 3-36, Figure 3-37, and Figure 3-38. At the beginning, the two wires do not have to make perfect contact with each other; the solder will fill any small gaps. But as before, the wires must be hot enough for the solder to flow, and this can take an extra few seconds when you use the low-wattage iron.

Be sure to feed the solder in as shown in the pictures. Remember: don't try to carry the solder to the joint on the tip of the iron. Heat the wires first, and then touch the solder to the wires and the tip of the iron, while keeping it in contact with the wires. Wait until the solder liquifies, and you will see it running eagerly into the joint. If this doesn't happen, be more patient and apply the heat for a little longer.

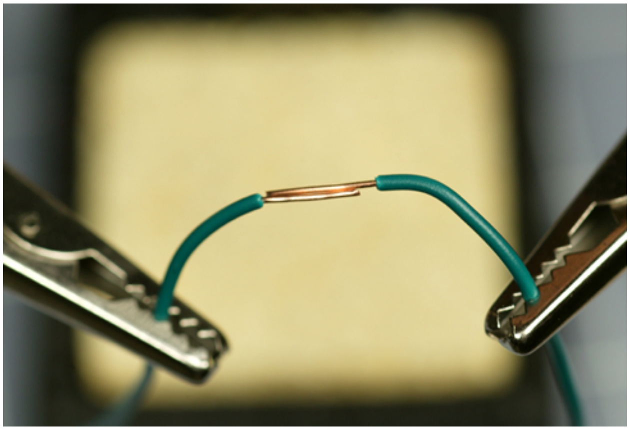

Figure 3-34 *Step 1: Align the wires.*

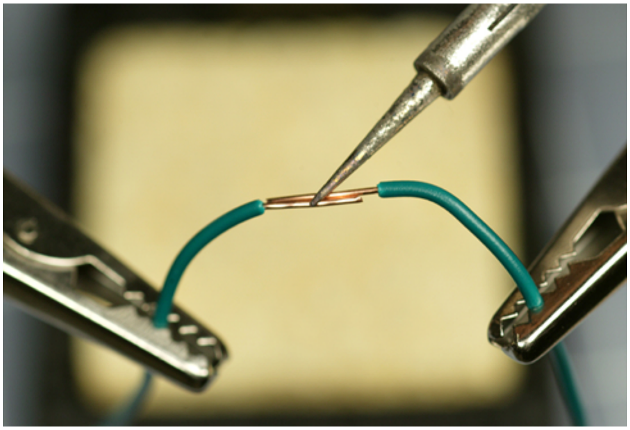

Figure 3-35 *Step 2: Heat the wires.*

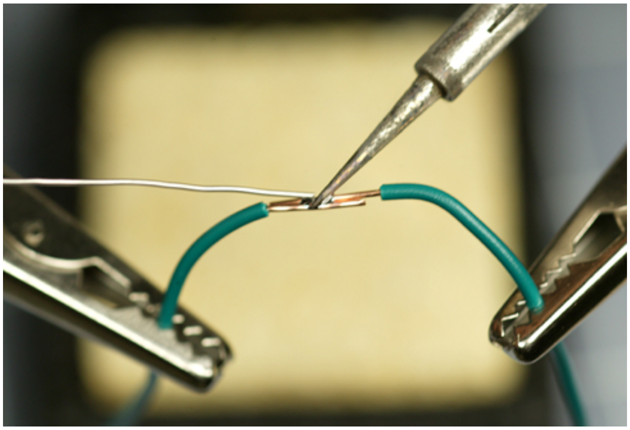

Figure 3-36 *Step 3: While continuing to heat the wires, apply the solder. Wait for it to start melting. Be patient.*

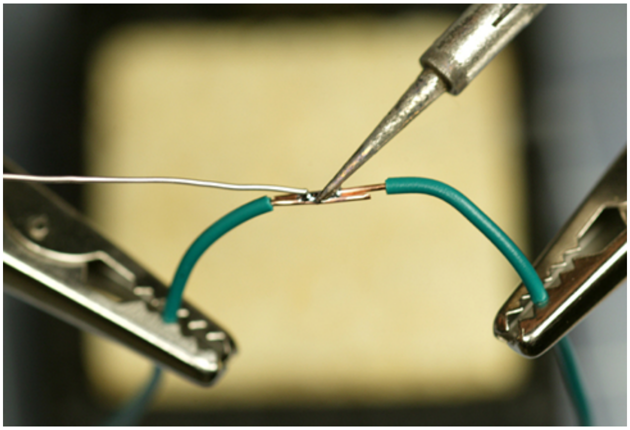

Figure 3-37 *The solder is now melting into the joint.*

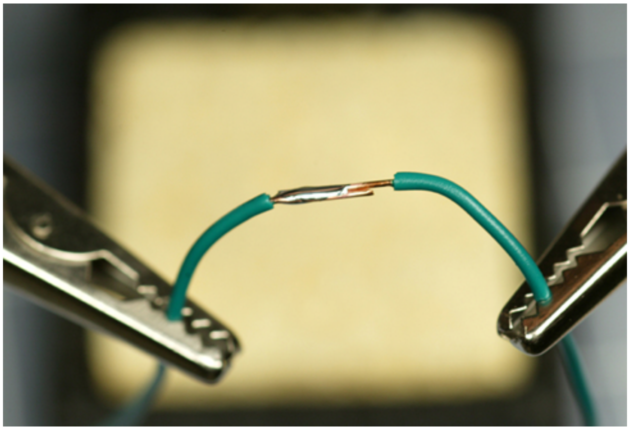

Figure 3-38 *The finished joint is shiny, and the solder has spread across the copper wires.*

The finished joint has enough solder for strength, but not so much solder that it will prevent heat-shrink tubing from sliding over it. I'll get to that in a moment.

Theory: Heat Transfer

The better you understand the process of soldering, the easier it should be for you to make good solder joints.

The tip of the soldering iron is hot, and you want to transfer that heat into the joint that you are trying to make. For this reason, you should adjust the angle of the soldering iron so that it makes the widest possible contact. See Figure 3-39 and Figure 3-40.

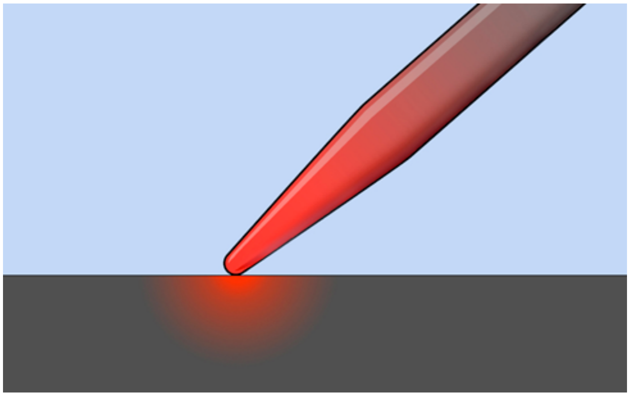

Figure 3-39 *A small contact area between the iron and the working surface allows insufficient heat transfer.*

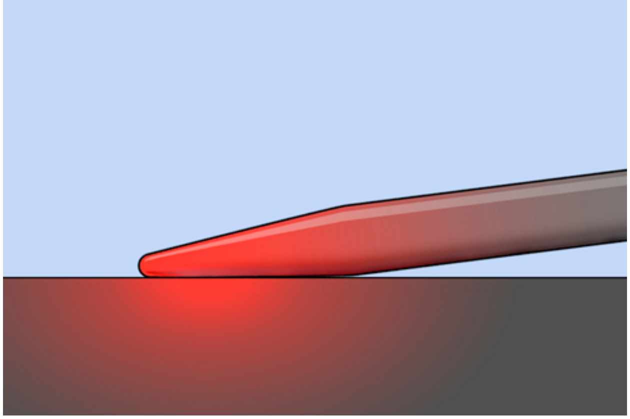

Figure 3-40 *A larger area of contact increases the heat transfer.*

Once the solder starts to melt, it broadens the area of contact, which helps to transfer more heat, so the process accelerates naturally. Initiating it is the tricky part.

The other aspect of heat flow that you should consider is that it can suck heat away from the places where you want it, and deliver it to places where you don't want it. If you're trying to solder a very heavy piece of copper wire, the joint may never get hot enough to melt the solder, because the heavy wire conducts heat away from the joint. You may find that even a 40-watt iron isn't powerful enough to overcome this problem. At the same time, while the copper doesn't get hot enough to melt the solder, it can be quite hot enough to melt insulation off the wire.

As a general rule, if you can't complete a solder joint in 10 seconds, you aren't applying enough heat.

Insulating the Solder Joint

After you've succeeded in making a good inline solder connection between two wires, it's time for the easy part. Choose some heat-shrink tubing that is just big enough to slide over the joint with a little bit of room to spare.

Of course, you do need to plan ahead. Usually you need to slip some tubing over one of the wires *before* you join them together. You'll see how this works when I get to the step-by-step procedure, below.

Assuming you have heat-shrink tubing on one of the wires, slide it along until the joint is centered under it. Hold it in front of your heat gun, and switch on the gun (keeping your fingers away from the blast of superheated air). Turn the wire so that you heat both sides. The tubing should shrink tight around the joint within half a minute. If you overheat the tubing, it may shrink so much that it splits, at which point you must remove it and start over. As soon as the tubing is tight, your job is done, and there's no point in making it any hotter. Note that while tubing mostly shrinks at right angles to its length, a little shrinkage also occurs along its length.

Figure 3-41, Figure 3-42, and Figure 3-43 show the desired result. I used white tubing because it shows up well in photographs. Different colors of heat-shrink tubing all perform the same way.

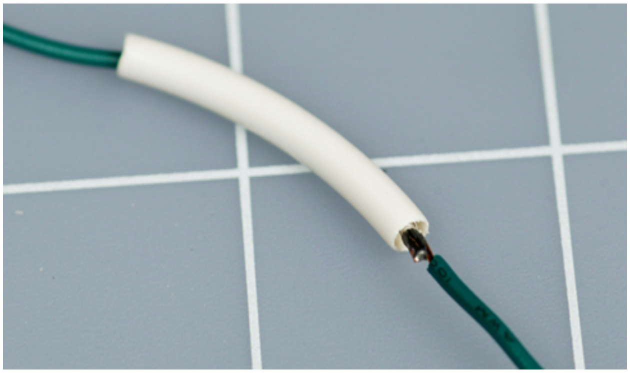

Figure 3-41 *Slip the tubing over your wire joint.*

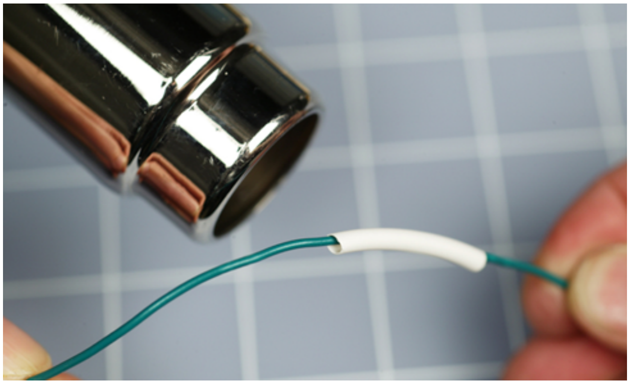

Figure 3-42 *Apply heat to the tubing.*

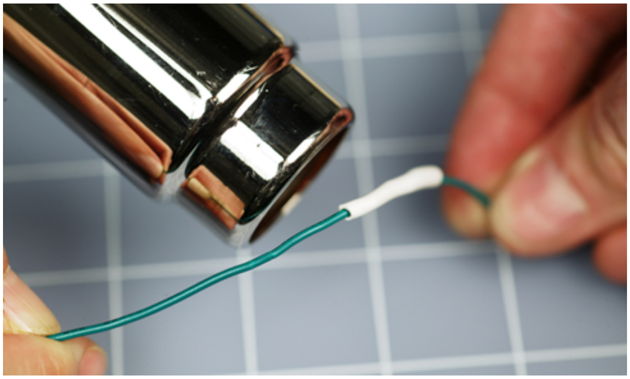

Figure 3-43 *Leave the heat on the tubing until it shrinks to firmly cover the joint.*

Caution: Heat Guns Do Get Hot!

Notice the chromed steel tube at the business end of a full-sized heat gun. Steel costs more than plastic, so the manufacturer must have put it there for a good reason —and the reason is that the air flowing through it becomes so hot that it would melt a plastic tube.

The metal tube stays hot enough to burn you for several minutes after you've used it. And, as in the case of soldering irons, other people (and pets) are vulnerable, because they won't necessarily know that the heat gun is a hazard. Most of all, make sure that no one in your home ever makes the mistake of using a heat gun as a hair dryer (see Figure 3-44).

Figure 3-44 *Members of your family should understand that a heat gun isn't quite the same as a hair dryer.*

This tool is just a little more hazardous than it appears. A mini heat gun may entail slightly less risk, but still should be treated with respect.

Wiring Your Power Source

The next application for your soldering skills will be more practical. You can add color-coded, solid-core wires to your AC adapter—or if you don't have an adapter, you can add extensions to the thin wires from a 9V battery connector. Either way, your extended 22-gauge wires will plug conveniently into a breadboard.

You can use your larger soldering iron for this, as no heat-sensitive components are involved.

If you have acquired an AC adapter, I'm assuming it is a little plastic module that plugs straight into the wall. A pair of wires emerges from it, carrying the low DC voltage that you need, and they terminate in some kind of miniature plug. The plug is suitable for a device such as a media player or a phone that has a matching socket, but for the purposes of this book, the plug is not useful, because you will want to apply the power to a breadboard.

How can this be done? I will show you.

Step One: Cut and Measure

First, let's make sure your AC adapter is doing what it's supposed to do.

Do not plug the adapter into the wall just yet. Begin by cutting the little plug off the end of the low-voltage wire, as shown in Figure 3-45. (You may notice that this photograph is of a RadioShack adapter. Ah, memories.)

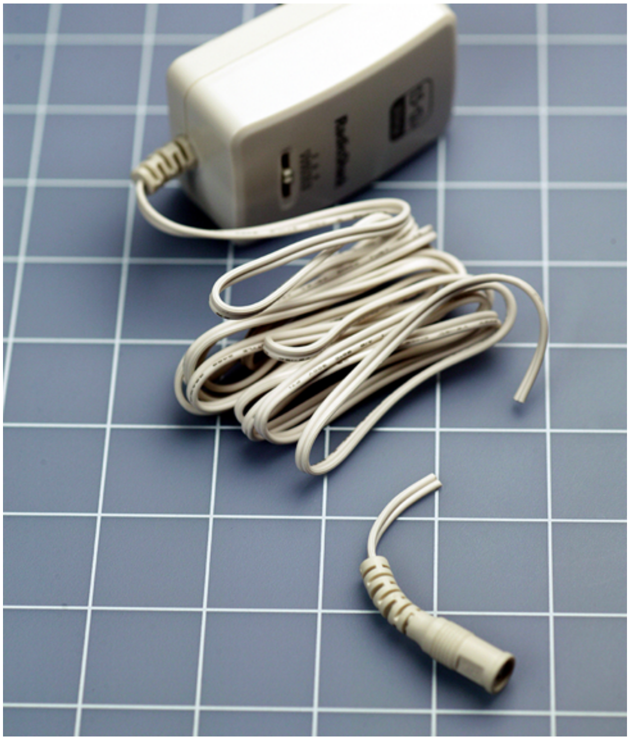

Figure 3-45 *The first step in customizing an AC adapter.*

Separate the two conductors, using wire cutters or a utility knife, and strip about 1/4" of insulation off each one. See Figure 3-46. The wires should be of unequal lengths, to reduce the risk of them touching.

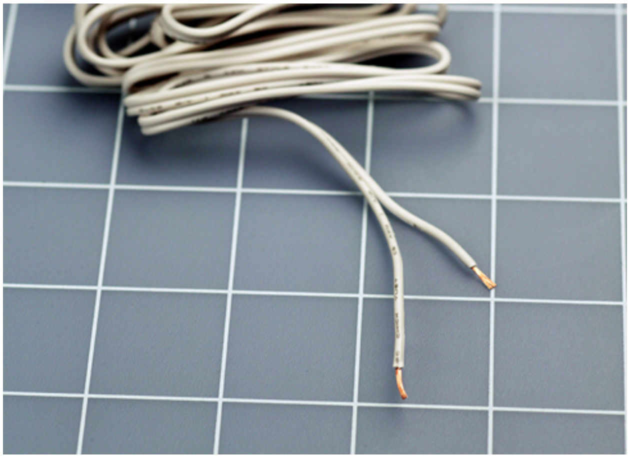

Figure 3-46 *The stripped wires.*

If the bare ends do touch each other while the adapter is plugged in, they may overload it or blow a fuse inside it. You may also get a spark that will be disconcerting, although I doubt it will hurt you. Not a big deal, but inconvenient.

Now set your meter to measure volts DC and attach it to the two wires from the AC adapter, preferably using alligator test leads to keep everything under control. After double-checking to make sure that your red meter lead is in the volts socket on the meter, not the mA socket, plug the AC adapter into the wall and measure how much voltage it is providing.

If you get a confusingly high reading, it could be because the voltage delivered by AC adapters is often greater when the adapter is not powering anything. The internal resistance of your meter is so high, the adapter will behave as if it isn't being loaded at all.

For a more meaningful test, select a resistor with a value of 680 ohms and clip it across the output from the adapter, in parallel with your meter. This will pull the voltage from the adapter down to a more appropriate level. Now you should get a value that makes sense.

It's not a good idea to use a resistor of much less than 680 ohms because the resistors on your shopping list are only rated at 1/4 watt, and if you try to push more power through them than that, they will overheat. If the 680-ohm resistor is attached to a 9-volt supply, Ohm's Law tells us that the current flowing through it will be about 13mA, and therefore the power dissipation will be about 120mW, or 0.12W, which is well within the rated maximum of a 0.25W resistor.

If you want to see how the voltage output from your AC adapter varies with a load of lower resistance, you can clip several 680-ohm resistors together in parallel. This could be an interesting test—but let's get back to the primary purpose here, which is to obtain power for your breadboard.

Step Two: Soldering

Use your meter to make absolutely sure there is no minus sign in front of the voltage that it measures on the wires from your AC adapter. If there is a minus sign, reverse the leads from your meter.

If the reading on your meter is positive, not negative, you know that the red wire from your meter is clipped to the positive side of your AC adapter. This is important, as you don't want your adapter to destroy components in your circuits by applying power to them the wrong way around.

The next steps will be the same regardless of whether you are adding solid 22-gauge wires to an AC adapter or a 9V battery connector.

Cut two pieces of solid-conductor 22-gauge wire—one of them red, the other black or blue. Each should be about 2 inches long. Strip 1/4" of insulation from both ends of each piece of wire.

Solder the 22-gauge wires to the wires from the AC adapter or the battery connector, using the technique that you practiced previously. Naturally, you attach the red wire to the positive side of your power supply.

If you have heat-shrink tubing and a heat gun, use it as you did in the practice session. The result should look like Figure 3-47. Once again the wires should be of unequal lengths, to reduce the risk of them touching each other. When the job is done, you can plug the ends of the 22-gauge wires into your breadboard.

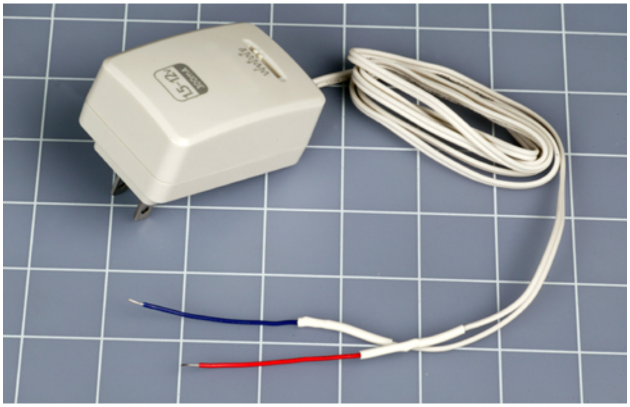

Figure 3-47 *The 22-gauge wires can be plugged into a breadboard, to supply it with power.*

Pruning a Power Cord

What else can you do with your newly acquired soldering skills? Here's a suggestion. Those of us who don't use Apple products may find ourselves with a laptop power supply that has a detachable AC power cord for plugging into the wall, in addition to the low-voltage DC wire that plugs into the computer. A typical power cord is shown in Figure 3-48.

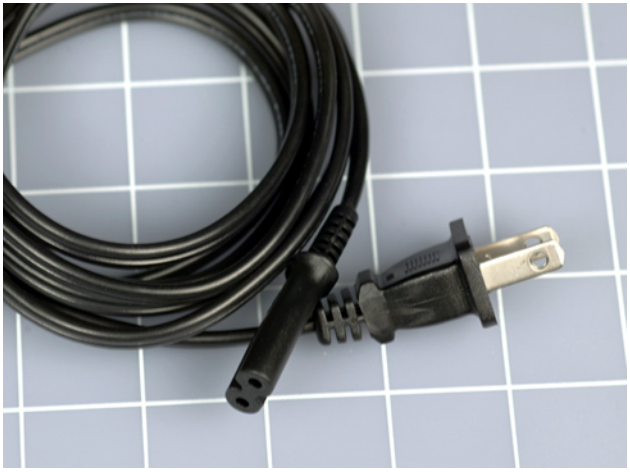

Figure 3-48 *A detachable AC power cord for a non-Apple laptop computer.*

What if you are an Apple fan? Maybe you have detachable power cords for other devices, such as a printer or a scanner. The purpose of this exercise is to reduce the length of a power cord so that it is exactly what you want, instead of lying around in a tangle. And if, like me, you have a laptop power cord that is longer than it needs to be, and you like to travel as light as possible, this exercise could be useful.

Twelve Steps to a Shorter Cord

In Figure 3-49 we see the first step, in which you boldly apply your wire cutters to chop a power cord. Needless to say, all of these steps require that the power cord should *not be plugged in* while you are working on it.

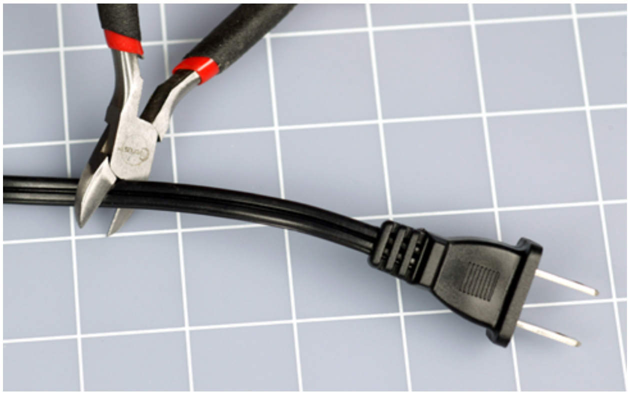

Figure 3-49 *Step 1 of 12, to shorten a power cord.*

Figure 3-50 shows the ends that you want to keep. You can store the rejected middle section of the power cord for some other purpose in the future.

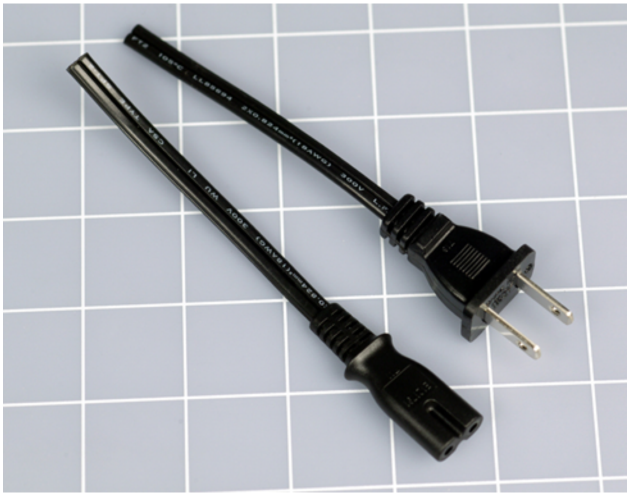

Figure 3-50 *Step 2 of 12, to shorten a power cord.*

In Figure 3-51, a utility knife on a cutting mat is an easy way to separate the two conductors in each segment of the power cord.

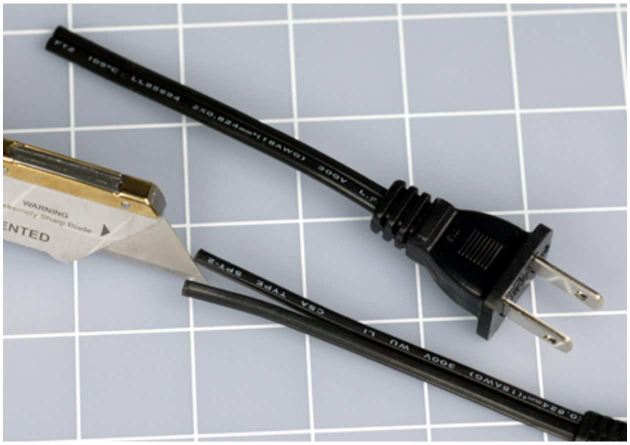

Figure 3-51 *Step 3 of 12, to shorten a power cord.*

In Figure 3-52, the segments of the power cord have been trimmed so that the conductors are of matching but unequal lengths. This way they will take up less room when you rejoin them, and there will be less risk of a short circuit if one of the joints fails for some reason.

Note that one conductor will be always marked, either with print or with molded ridges. Be sure that the marked conductors match each other when you join them.

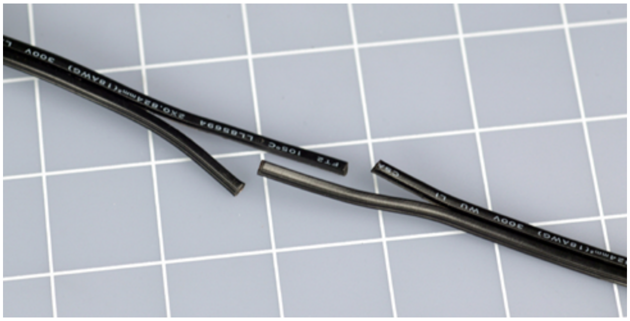

Figure 3-52 *Step 4 of 12, to shorten a power cord.*

Strip off a minimal amount of insulation. Just 1/8″ (3mm) will be enough. Then cut some pieces of heat-shrink tubing, if you have it. Each smaller piece must be just large enough to slide over each individual conductor in the power cord, while the bigger piece, about 2″ long, will cover the whole joint. See Figure 3-53.

Note that some heat-shrink tubing is rated only for low voltages. It should not be used for this project.

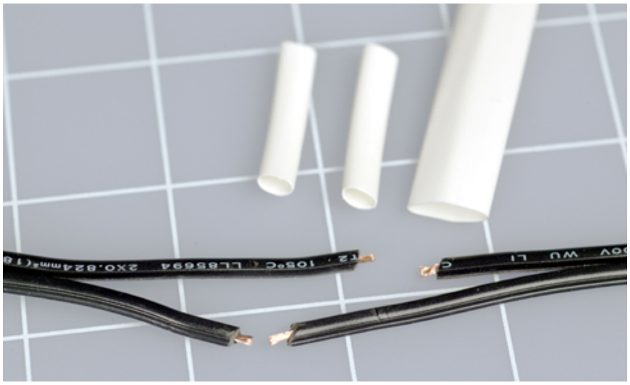

Figure 3-53 *Step 5 of 12, to shorten a power cord.*

Now for the most difficult part: activating your human memory. You have to remember to slide the tubing onto the wire *before* you make your solder joint, because the plugs on the ends of the wires will prevent you from adding any heat-shrink tubing later. If you're as impatient as I am, it's very difficult to remember to do this every time. See Figure 3-54.

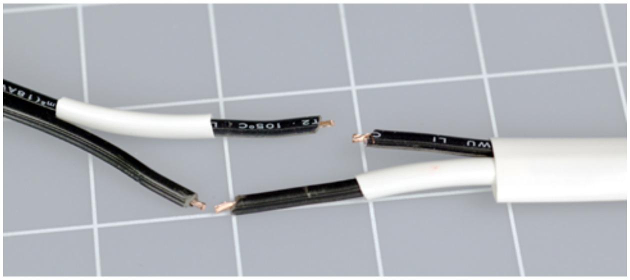

Figure 3-54 *Step 6 of 12, to shorten a power cord.*

Use your Helping Hand to align the first joint. Push the two pieces of wire together so that the strands intermingle, and then squeeze them tight between finger and thumb, so that there are no little bits sticking out. A stray strand of wire can puncture heat-shrink tubing when the tubing is hot and soft and is shrinking around the joint. See Figure 3-55.

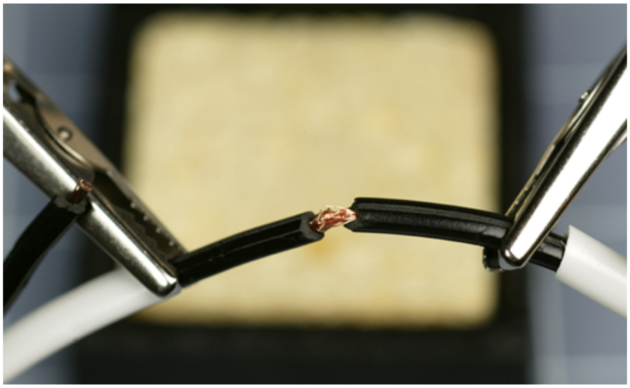

Figure 3-55 *Step 7 of 12, to shorten a power cord.*

The wire that you're joining is much heavier than the 22-gauge wire that you worked with previously, so it will suck up more heat, and you must touch the soldering iron to it for a longer time. Make sure that the solder flows all the way into the joint, and check the underside after the joint is cool. Most likely you'll find some bare copper strands there. The joint should become a nice solid, rounded, shiny blob. See Figure 3-56.

Be careful to keep the heat-shrink tubing as far away from the joint as possible while you're using the soldering iron, so that heat from the iron doesn't shrink the tubing prematurely, preventing you from sliding it over the joint later.

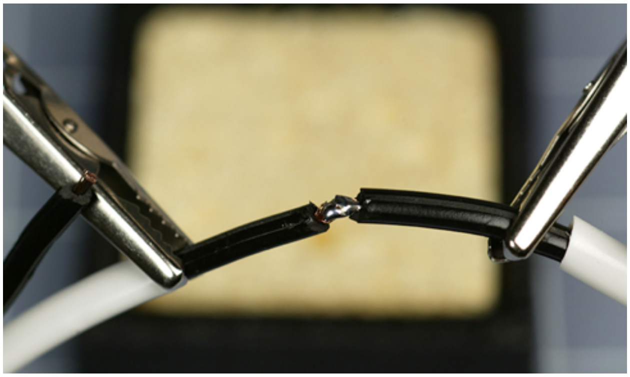

Figure 3-56 *Step 8 of 12, to shorten a power cord.*

Slide a section of heat-shrink tubing over the solder joint, and apply the heat gun, as shown in Figure 3-57. Don't allow the other pieces of heat-shrink tubing to pick up stray heat.

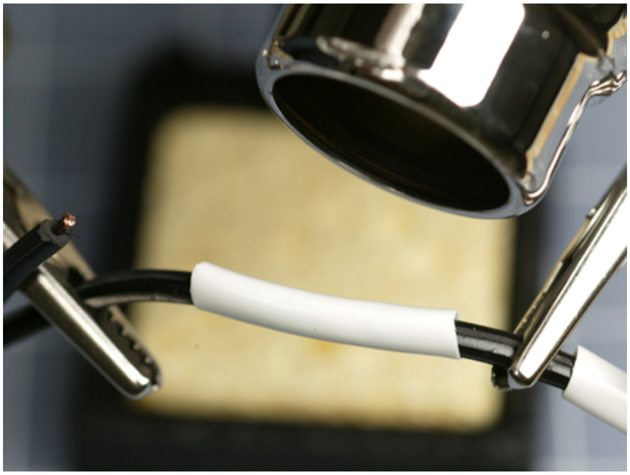

Figure 3-57 *Step 9 of 12, to shorten a power cord.*

In Figure 3-58, the heat-shrink tubing has now shrunk.

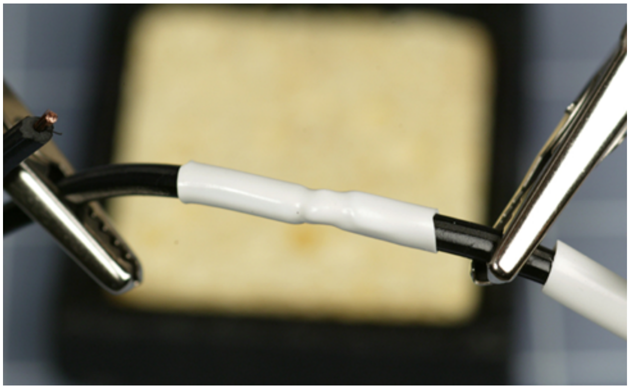

Figure 3-58 *Step 10 of 12, to shorten a power cord.*

Get ready to solder the other pair of conductors. See Figure 3-59.

Figure 3-60 shows that the second joint has been made. After you protect it with its own piece of heat-shrink tubing, you'll be ready to slide the larger section of tubing over the whole assembly. Er—you did remember to put the large tubing onto the wire at the beginning, didn't you?

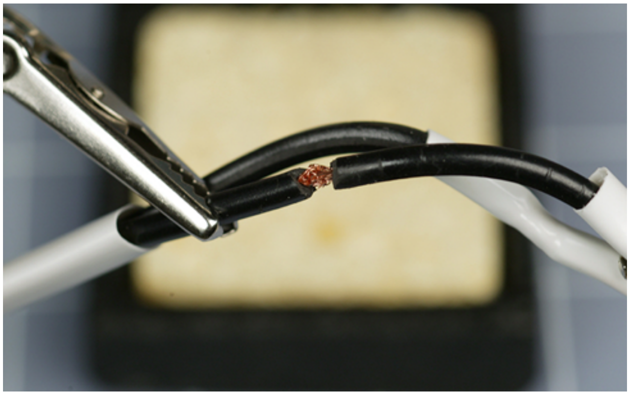

Figure 3-59 *Step 11 of 12, to shorten a power cord.*

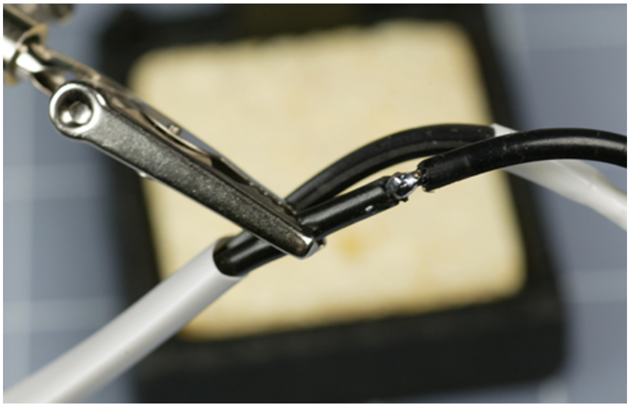

Figure 3-60 *Step 12 of 12, to shorten a power cord.*

Figure 3-60 shows the finished shortened power cord.

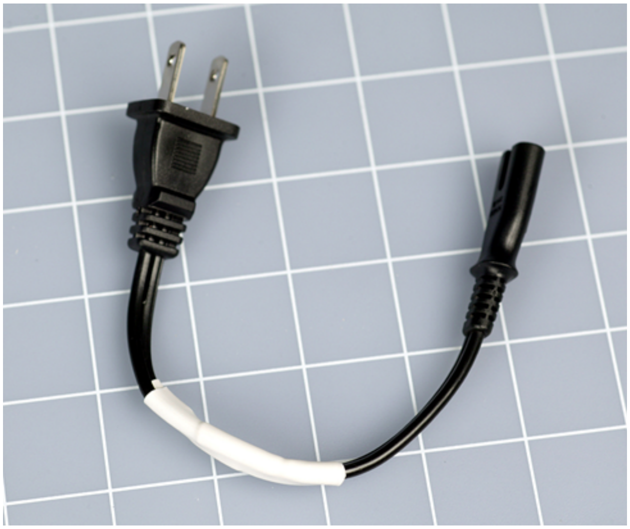

Figure 3-61 *The power cord, shortened.*

What's Next?

If you have completed the soldering exercises so far, you have sufficient basic skills to build your first soldered electronic circuit. Although—maybe first we should just run through a quick demonstration of the consequences of unintentionally excessive heat. I'd hate for you to take a lot of trouble soldering things together, only to discover that you melted a transistor or an LED. Unsoldering damaged parts is much less fun than soldering them.

Experiment 13: Roasting an LED

In Experiment 4, you discovered how easy it is to burn out an LED. What really happened, in that little adventure, is that excessive current passing through the LED created excessive heat, and the heat killed the component.

If heat caused by electricity can destroy an LED, do you think heat from a soldering iron can do the same thing? It sounds plausible, but there's only one way to make absolutely sure.

What You Will Need

- 9V battery and connector, or 9V AC-DC adapter
- Long-nosed or sharp-nosed pliers
- 30-watt or 40-watt soldering iron
- 15-watt soldering iron
- LEDs, generic (2)
- 470-ohm resistor (1)
- Helping Hand to hold your work
- One large or two small pure-copper alligator clips

The purpose of the experiment is to study the effects of heat. This means we need to know where the heat is going.

For this reason, you're not going to use a breadboard. The contacts inside the board would absorb an unknown amount of heat. I don't want you to use test leads, either, because they too will absorb heat.

Instead, please use some sharp-nosed pliers to bend each of the leads from an LED into little hooks, and do the same thing with the wires on a 470-ohm resistor. See Figure 3-62, where you will see that the wires from a 9-volt battery have been bent the same way. To make them maintain the hook shape, you may need to strip off a little extra insulation and apply a little solder.

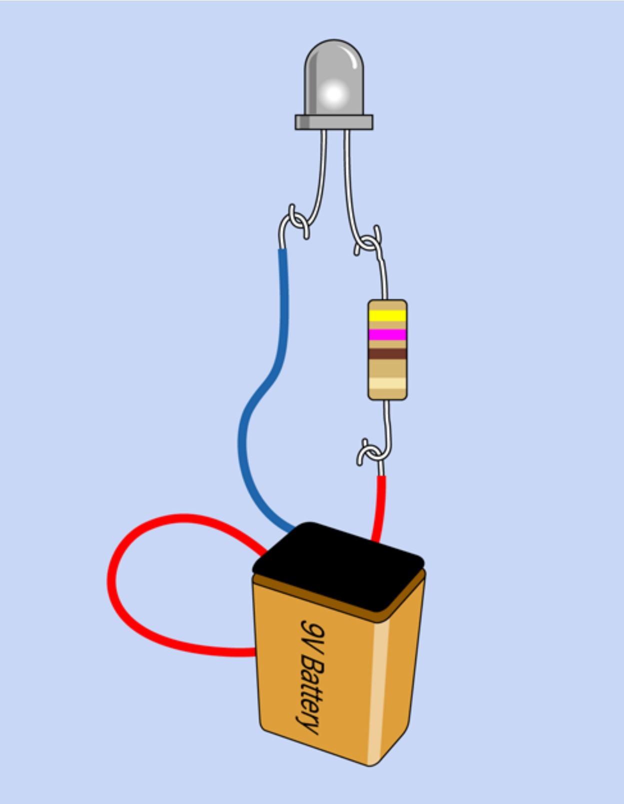

Figure 3-62 *Measuring the heat tolerance of an LED. An AC adapter can substitute for the 9-volt battery.*

To minimize heat loss through conduction, the resistor dangles from one of the leads on the LED, and the power-supply wire hangs from that, a little farther down. Gravity should be sufficient to make this work.

Grip the plastic body of the LED in your Helping Hand. Plastic is not a good thermal conductor, so the lens of the LED should not allow much heat to be conducted away through the Helping Hand.

Apply the 9 volts, and your LED should be shining brightly. I used a white LED in this experiment, because it's easier to photograph.

You'll need your low-powered 15W soldering iron, and the higher-powered one, too. Plug them in and wait at least five minutes, to make sure they are really hot. Now hold the tip of the 15W iron firmly against one of the leads on your glowing LED, while you check the time with a watch. Figure 3-63 shows the setup.

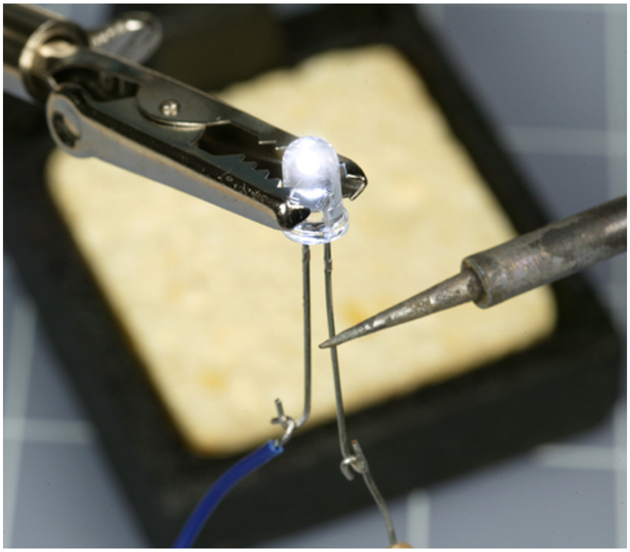

Figure 3-63 *Applying heat with a 15-watt soldering iron.*

I'm betting that you can sustain this contact for a full three minutes without burning out the LED. Now you know why a 15-watt soldering iron is recommended for delicate electronics work.

Allow your LED wire to cool, and then apply your more powerful soldering iron in the same location as before. I think you'll find that the LED will go dark after as little as 10 seconds (note that some LEDs can survive higher temperatures than others). This is why you *don't* use a 30-watt soldering iron for delicate electronics work.

The large iron doesn't necessarily reach a higher temperature than the small one. It just has a larger heat capacity. In other words, a greater quantity of heat can flow out of it, at a faster rate.

Your LED has been sacrificed to satisfy the need for knowledge. It was an honorable death. Lay it to rest in your trash, and substitute a new LED, which we will try to treat more kindly. Connect it as before, but this time add a full-size copper alligator clip (or two small clips) to one of the leads near the body of the LED, as shown in Figure 3-64. Press the tip of your 30-watt or 40-watt soldering iron against the lead just *below* the alligator clip. This time, you should be able to hold the powerful soldering iron in place for a full two minutes without burning out the LED.

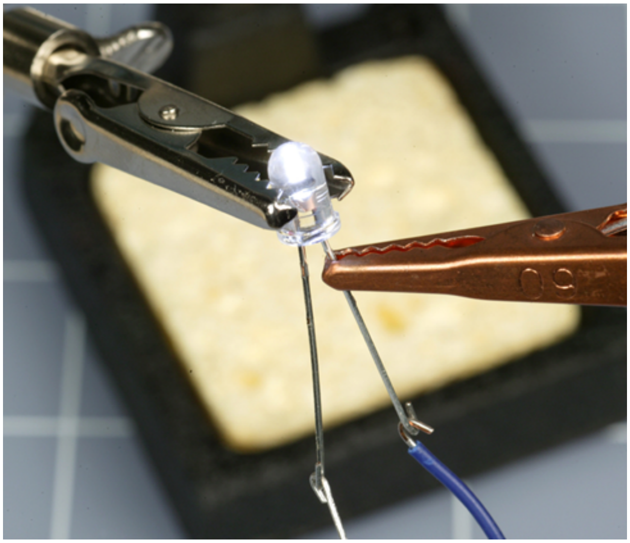

Figure 3-64 *Using a copper alligator clip as a heat sink to protect an LED.*

Where the Heat Goes

At the end of your experiment, if you touch the clip, you'll find that it's relatively hot, while the LED remains not so hot. Imagine the heat flowing out through the tip of your soldering iron, into the wire that leads to the LED —except that the heat meets the alligator clip along the way, as shown in Figure 3-65. The clip is like an empty container waiting to be filled. The heat prefers to flow into the copper clip, leaving the LED unharmed.

The alligator clip functioned as a *heat sink*. It works better than an everyday nickel-plated steel alligator clip, because copper is such a good conductor of heat.

Going back to the first part of this experiment, you saw that a 15-watt soldering iron failed to harm the LED, with no heat sink needed. Does this mean that a 15-watt iron is completely safe?

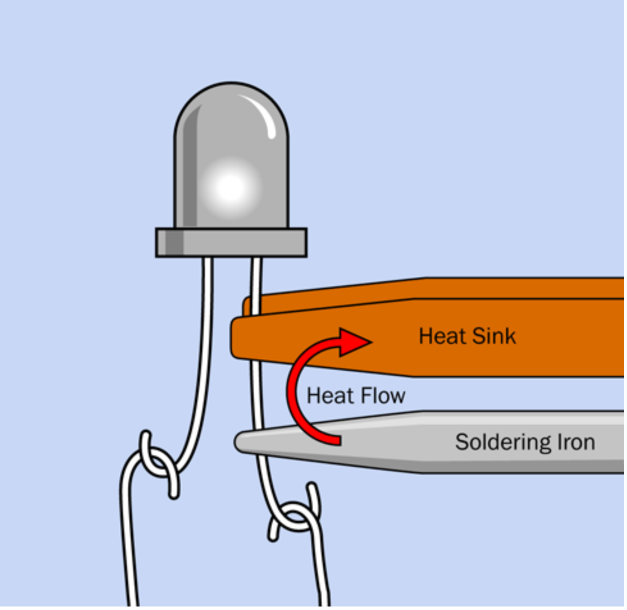

Figure 3-65 *A copper alligator clip conducts heat away from an LED.*

Well, maybe. The problem is, you don't really know whether some semiconductors may be more heat-sensitive than LEDs.

Because the consequences of burning out a component can be so annoying, I suggest you should consider playing it safe and use a heat sink in these circumstances:

- If you apply a 15-watt iron extremely close to a semiconductor for 20 seconds or more.
- If you apply a 30-watt iron within half an inch of resistors or capacitors for 10 seconds or more. (Never use it near semiconductors.)
- If you apply a 30-watt iron near anything meltable for 20 seconds or more. Meltable items include insulation on wires, plastic connectors, and plastic components inside switches.

Rules for Heat Sinking

- Full-size copper alligator clips work best, but may not fit into tight corners. Ideally you should have small ones available too.
- Clamp the alligator clip as close as possible to the component and as far as possible from the solder joint that you are trying to make. The

joint does need to get hot. Divert heat from the component, not from the joint.

- Make sure there is a metal-to-metal connection between the alligator clip and the wire to promote good heat transfer.

If you keep these points in mind, we can proceed with the fascinating challenge of point-to-point wiring.

Experiment 14: A Wearable Pulsing Glow

Until now, I've encouraged you to start putting things together without much of a theory or a plan. That's what Learning by Discovery tends to be like. Sometimes, though, a plan can be essential, and this is one of those times. I'm going to outline the requirements of this project, and then I will take you step by step through the process of building it.

What You Will Need

- 9V battery and connector, or 9V AC-DC adapter
- Hookup wire, wire cutters, wire strippers, multimeter
- 15-watt soldering iron
- Thin solder (0.022 inches)
- Plain perforated board (no copper plating necessary)
- Helping Hand
- Resistors: 470 (2), 100K (1), 4.7K (2), 470K (2)
- Capacitors: 3.3µF (2), 220µF (1)
- Transistors: 2N2222 (3)
- Generic LED (1)

Fluctuations Revisited

Please turn back to Figure 2-116 to refresh your memory of that circuit. The task, now, is to make it as small as possible, so that someone could wear it.

Imagine that the leads of the components are interconnected with rubber bands, allowing you to shuffle them around on a tabletop without them losing their connections with each other. When the rubber bands are stretched as little as possible, the circuit is as small as possible, and you can link the parts with bare wires, supporting everything with a piece of perforated board.

The only problem is that bare wires under the board cannot cross each other. The idea is that after you have verified the function of your circuit, you could send the specification to a service that etches circuit boards.

Of course, modern printed circuits are double-sided, at the very least, and many have intermediate layers, allowing multiple conductors to cross each other without making electrical contact. But it's always good to start on a simple, traditional basis, and the simplest board has components on one side and connections on the other. Components on top of the board can bridge the conductors beneath, because the insulating material of the board separates them. But conductors cannot cross each other.

My best attempt to reduce the size of this circuit is shown in Figure 3-66, on a piece of plain perforated board measuring 0.9" x 1.3". If you can come up with a design that is significantly smaller, I'd love to see it. Here are some ideas:

- Use smaller resistors rated for 1/8 watt instead of 1/4 watt.
- Mount resistors vertically.
- Thread two leads through one hole, if the holes in your board are big enough.

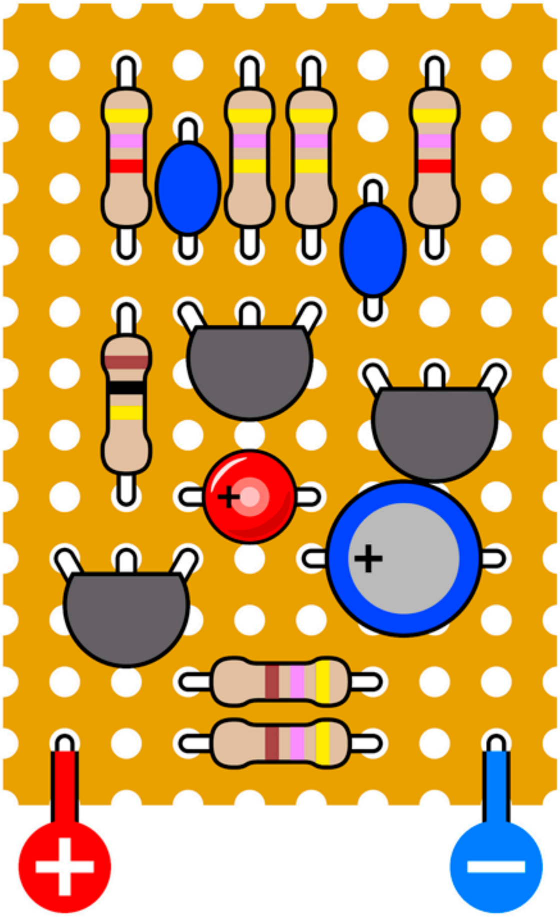

Figure 3-66 *The oscillator circuit, reduced to occupy minimal space on perforated board.*

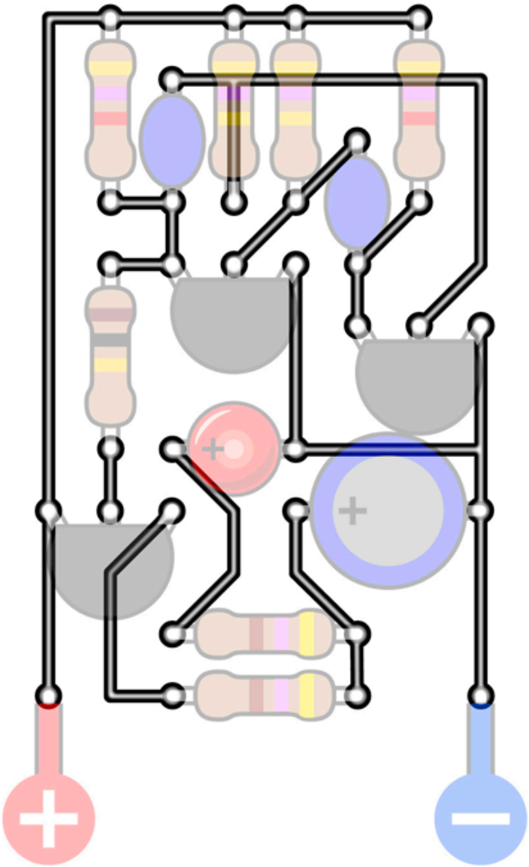

Figure 3-67 *Connections in black are wires beneath the circuit board, which is transparent in this view.*

Where are the connections between the components? They're under the board. In Figure 3-67 I grayed out the components and disappeared the board, so that you can see the wiring.

If you compare this diagram very carefully with the schematic in Figure 2-116, you should find that the connections between components are the same—unless I made an error. (I sure hope not. I don't want to have to redraw everything.)

Figure 3-68 shows yet another view, this time omitting the components while including the board, so that you can see how the connections fit the 0.1″ x 0.1″ grid of holes in the board.

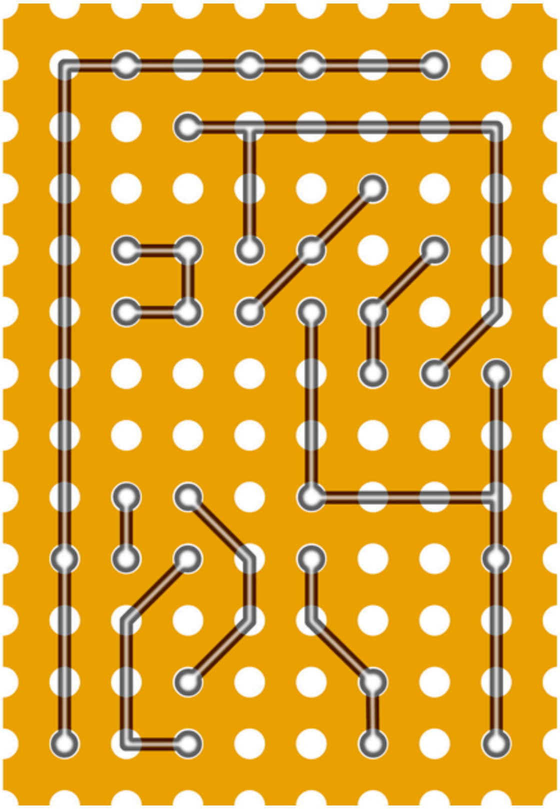

Figure 3-68 *In this view, only the board and the connections are shown. Every circular dot indicates a connection that comes through a hole in the board.*

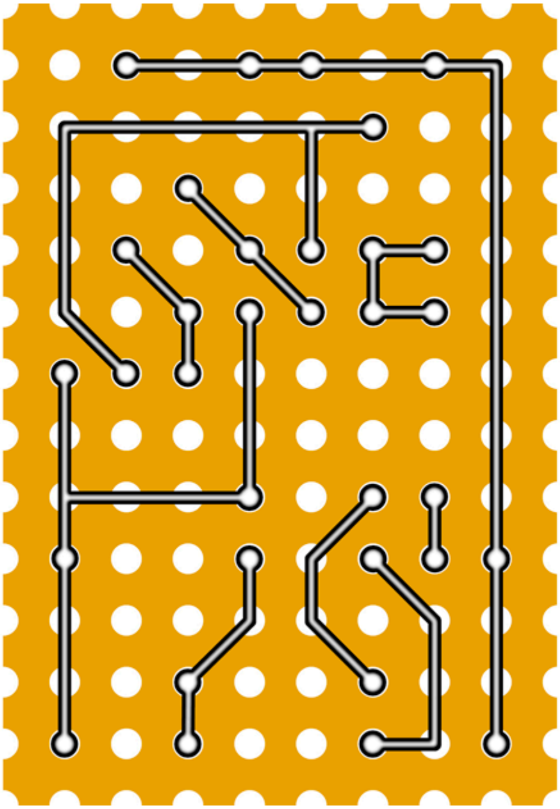

Figure 3-69 *The previously shown connections are flipped left to right, so that this is the way the board should look from the underside.*

Finally, Figure 3-69 shows the board flipped over, left to right, so that you are viewing it from behind. This will help you to make the connections when you're putting the components together. You are going to give this a try, aren't you?

Bend Wires, Add Solder

Now that you've seen the plan for this project, how are you supposed to make all those connections?

It's not so difficult. Resistors, capacitors, and transistors have wire leads that are usually at least 1/2" long. So, you can push them through the holes in the perforated board, and then bend the leads over to touch each other. While they are touching, you solder them together. Snip off any surplus, add a battery connection, and you're done.

There are three main issues to watch out for.

- Holding the board steady while you work on it requires some care and patience. Your Helping Hand will be necessary.
- The components and the solder joints that you make will be very close together. Use your copper alligator clips to provide heat protection.
- Flipping between the top side of the board and the bottom side is confusing. You can easily put a wire in the wrong place. I think this is actually the hardest part.

Perhaps you have seen perforated board on which a little circle of copper has been added around each hole. Would that be suitable for this project? The copper circles have the advantage of anchoring components securely, but they can also create a short circuit between wires that are close together. I think bare board is easier for a small project like this. A sample was shown in Figure 3-22. Some perforated board has holes that are larger, but this does not make a significant difference.

Step by Step

Here's the specific procedure for building the circuit:

Cut a piece of perfboard measuring 0.9" by 1.3" out of a plain sheet. (You don't need a ruler calibrated in tenths of an inch. Just count the rows of holes in the board.) You can use a miniature hobby saw, or you may be able to snap the board along its lines of holes, if you're careful. A hacksaw will also work. I suggest you should not use a good wood saw, because perforated board often has glass fibers in it, and they can blunt a saw.

Gather all the components and carefully insert three or four of them through holes in the board, counting the holes to make sure everything is in the right place. Flip the board over and bend the wires from the components to anchor them to the board and create the connections shown in Figure 3-69. If any of the wires isn't long enough, you'll have to supplement it with an extra piece of 22-gauge wire from your supply. Remove the insulation, which just gets in the way.

Trim the wires approximately with your wire cutters.

Make the joints with your soldering iron.

Now, the important part: check each joint using a close-up magnifying glass, and wiggle the wires with pointed-nosed pliers. If there isn't enough solder for a really secure joint, reheat it and add more. If solder has created a connection that shouldn't be there, use a utility knife to make two parallel cuts in the solder, and scrape away the little section between them.

Generally, I deal with just three or four components at a time, because I get confused if there are more than that. If I solder one component in the wrong location, undoing the error is not too difficult—unless I have already added more components to it by the time I discover the error.

Caution: Flying Wire Segments

The jaws of your wire cutters exert a powerful force that peaks and then is suddenly released when they cut through wire. This force can be translated into sudden motion of the snipped wire segment. Some wires are relatively soft, and don't pose a risk, but transistors and LEDs tend to have harder wires. Little segments can fly in unpredictable directions at high speed, creating a real hazard for your eyes when you are doing close-up work.

Everyday eyeglasses can protect you when trimming wires. If you don't use eyeglasses, plastic safety glasses are really a good idea.

Finishing the Job

I always use bright illumination. This is not a luxury; it is a necessity. Buy a desk lamp, if you don't already have one. It doesn't have to be expensive; a thrift-store item is okay.

I use a daylight-spectrum LED desk lamp, because it helps me to identify the colored bands on resistors more reliably. I stopped using a fluorescent desk lamp when I discovered that any small imperfection in the coating inside the tube can allow ultraviolet light to escape. This constitutes a hazard when you are working so close to the light.

No matter how good your vision is, you need to examine each joint with a magnifier. You'll be surprised how imperfect some of them are. Hold the magnifier as close as possible to your eye, then pick up the board and bring it closer until the joint that you are inspecting comes into focus.

Finally, you should end up with a working circuit that pulses like a heartbeat. Or does it? If you have difficulty making it work, retrace every connection and compare it with the schematic. If you don't find an error, apply power to the circuit, attach the black lead from your meter to the negative side, and then go around the circuit with the red lead, checking the presence of voltage. Every part of this circuit should show at least some voltage while it's working. If you find a dead connection, you may have made a bad solder joint, or missed one entirely.

When you're done, now what? Well, now you can stop being an electronics hobbyist and become a crafts hobbyist. You can try to figure out a way to make this thing wearable.

First you have to consider the power supply. Because of the components that I used, you really need 9 volts to make this work well. How are you going to make this 9-volt circuit wearable, with a bulky 9-volt battery?

I can think of three answers:

- You can put the battery in a pocket, and mount the flasher on the outside of the pocket, with a thin wire penetrating the fabric.
- You could mount the battery inside the crown of a baseball cap, with the flasher on the front.
- You can put three 3-volt button batteries in a stack, held in some kind of plastic clip. I'm not sure how long they'll last, though.

I have to note that the 2N2222 transistors in this project are not ideal, because they tend to use more power than field-effect transistors, also known as MOSFETS. However, I made a decision in this book that I only had space for one transistor family, and bipolar NPNs are the most fundamental type.

Regarding your choice of LED, those with a clear lens create a defined beam of light, which may not be appropriate for this project. Those that are diffused create a more pleasing glow. You can diffuse the light more by embedding the LED in a piece of transparent acrylic plastic, at least 1/4 inch thick, as shown in Figure 3-70. Roughen the front of the acrylic with fine sandpaper, ideally using an orbital sander that won't make an obvious pattern. This will make the acrylic translucent rather than transparent.

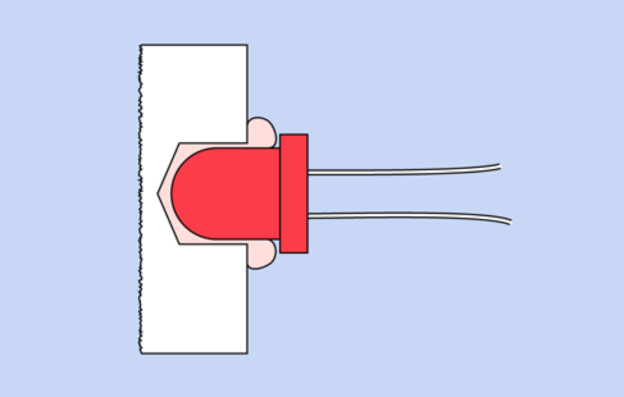

Figure 3-70 *This cross-sectional view shows a sheet of transparent acrylic in which a hole has been drilled part of the way from the back toward the front. Because a drill bit creates a hole with a conical shape at the bottom, and because the LED has rounded contours, transparent epoxy or silicone caulking can be injected into the hole before mounting the LED.*

Drill a hole slightly larger than the LED in the back of the acrylic. Don't drill all the way through the plastic. Remove all fragments and dust from the hole by blasting some compressed air into it, or by washing it if you don't have an air compressor. After the cavity is completely dry, get some transparent silicone caulking or mix some clear five-minute epoxy and put a drop in the bottom of the hole. Then insert the LED, pushing it in so that it forces the epoxy to ooze around it, making a tight seal.

Try illuminating the LED, and sand the acrylic some more if necessary. Finally, you can decide whether to mount the circuit on the back of the acrylic, or whether you want to run a wire to it elsewhere.

You can choose the resistors in the oscillator circuit to make the LED flash at about the speed of a human heart while the person is resting. This may make it look as if it's measuring your pulse, especially if you mount it on the center of your chest or in a strap around your wrist. If you enjoy hoaxing people, you can suggest that you're in such amazingly good shape, your pulse rate remains constant even when you're taking strenuous exercise.

To make a good-looking enclosure for the circuit, I can think of options ranging from embedding the whole thing in clear epoxy to finding a Victorian-style locket. I'll leave you to consider alternatives, because this is a book about electronics rather than craft projects. However, there is one craft-related issue that I want to mention, and this is a good moment to do so.

Background: Maddened by Measurement

Throughout this book, I've mostly used measurements in inches. Sometimes, though, I've ventured into the metric system, as when referring to "5mm LEDs." This isn't inconsistency on my part; it reflects the conflicted state of the electronics industry, where you'll find inches and millimeters both in daily use, often in the very same datasheet. For instance, the pin spacing of surface-mount chips tends to be measured in millimeters, but through-hole chips still have pins 0.1" apart, and probably always will.

To complicate matters further, where inches are used, there are two different systems for dividing them into fractional amounts. Drill bits, for instance, are measured in multiples of 1/64ths of an inch. Metal shims are graded in 1/1,000ths of an inch (0.001", 0.002", and so on). To make things even more confusing, the thickness of sheet metal is often measured by "gauge," as in 16-gauge steel, which happens to be about 1/16" thick.

Why doesn't the US move to the metric system, since it's so much more rational?

We can debate whether it really is rational. When it was formally introduced in 1875, the meter was defined as being 1/10,000,000 of the distance between the North Pole and the equator, along a line passing through Paris. Why Paris? Because the French came up with the idea. Since then, the meter has been redefined three times, in a series of efforts to achieve greater accuracy in scientific applications.

As for the usefulness of a 10-based system, moving a decimal point is certainly simpler than doing calculations in 1/64ths of an inch, but the only reason we count in tens is because we happen to have evolved with that number of digits on our hands. A 12-based system would really be more convenient, as numbers would be evenly divisible by 2 and 3.

This is all very hypothetical. The fact is, we're stuck with conflicts in length measurement, so I've created four charts to assist you in converting from one system to another. From these you will see that when you need to drill a hole for a 5mm LED, a 3/16-inch bit is about right. (In fact, it results in a tighter fit than if you drill an actual 5mm hole.)

Figure 3-71 will help you to convert between 1/64ths and 1/100ths of an inch. The gray column is divided into 1/64ths, the blue column is in 1/32nds, the green column is in 1/16ths, and the orange column is in 1/8ths. Customarily, if a value can be precisely expressed in larger units, we use that option; thus instead of referring to 8/64ths of an inch, we would express this as 1/8th of an inch. This causes some confusion when you're trying to figure out whether one measurement is larger than another—for instance, are 11/32nds of an inch larger than 5/8ths of an inch? Check the diagram to make sure.

Because datasheets often express dimensions using decimal fractions of an inch, a second chart in Figure 3-72 converts between decimals and 64ths. I think you're quite likely to find a measurement such as 0.375", and if you know that this is the same as 3/8", the knowledge can be useful.

Many datasheets provide measurements in both millimeters and inches, but some now use millimeters only. If you are still thinking in inches, or if you want to know if a component will fit the 1/10" hole spacing in a breadboard or perforated board, it's helpful to remember that 1/10" is equivalent to 2.54mm. Provided a component is small, pin spacing in multiples of 2.5mm is acceptable. However, when pins are 25mm or more apart, they may not fit into holes that are 25.4mm apart (that is, one inch or more).

Figure 3-73 enables conversion between millimeters, 1/100ths of an inch, and 1/64ths of an inch.

Figure 3-74 is a magnified version of the previous chart, showing tenths of millimeters and 1/1,000ths of an inch.

Some progress has been made during the past four decades toward adopting the metric system in the United States, but more decades will pass before this transition is complete. In the meantime, anyone using parts or tools manufactured or sold in the United States should be familiar with both systems. There's no way around it.

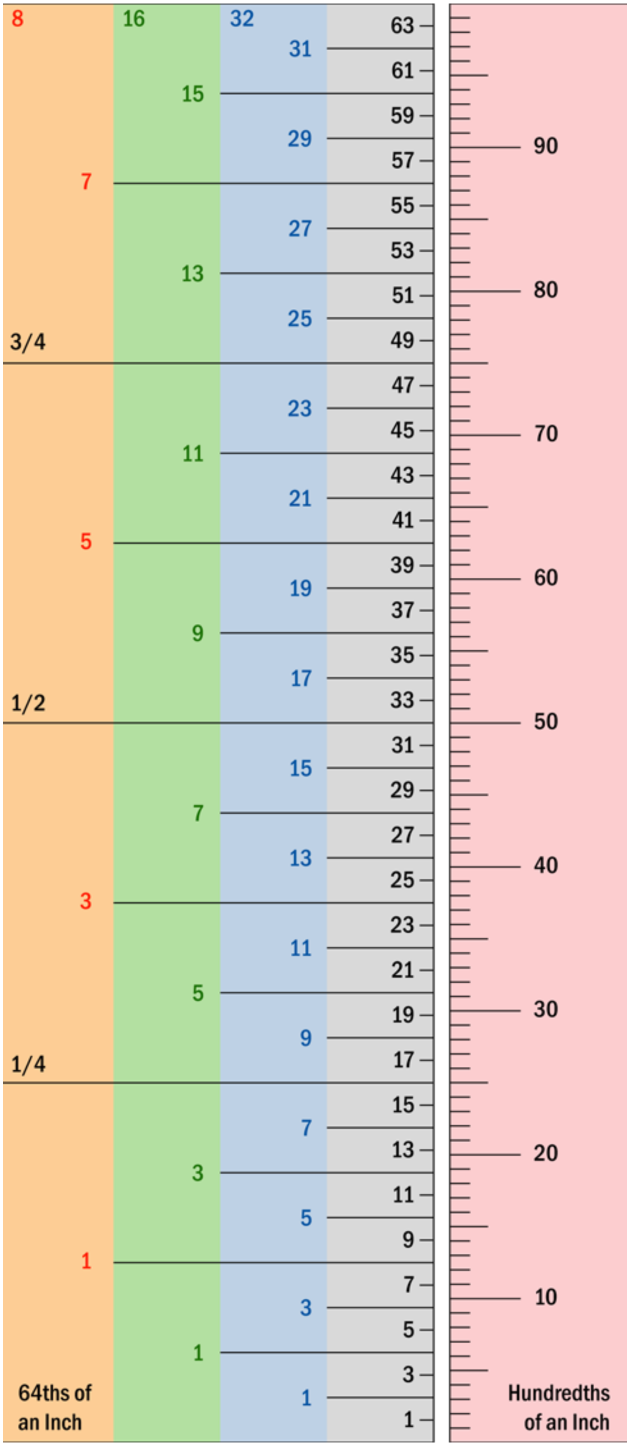
8
16
32
7
3/4
5
1/2
3
1/4
1
15
13
11
9
7
5
3
1
31
29
27
25
23
21
19
17
15
13
11
9
7
5
3
1
63
61
59
57
55
53
51
49
47
45
43
41
39
37
35
33
31
29
27
25
23
21
19
17
15
13
11
9
7
5
3
1
90
80
70
60
50
40
30
20
10
64ths of an Inch
Hundredths of an Inch

Figure 3-71 *For conversion between 64ths of inch and 100ths.*

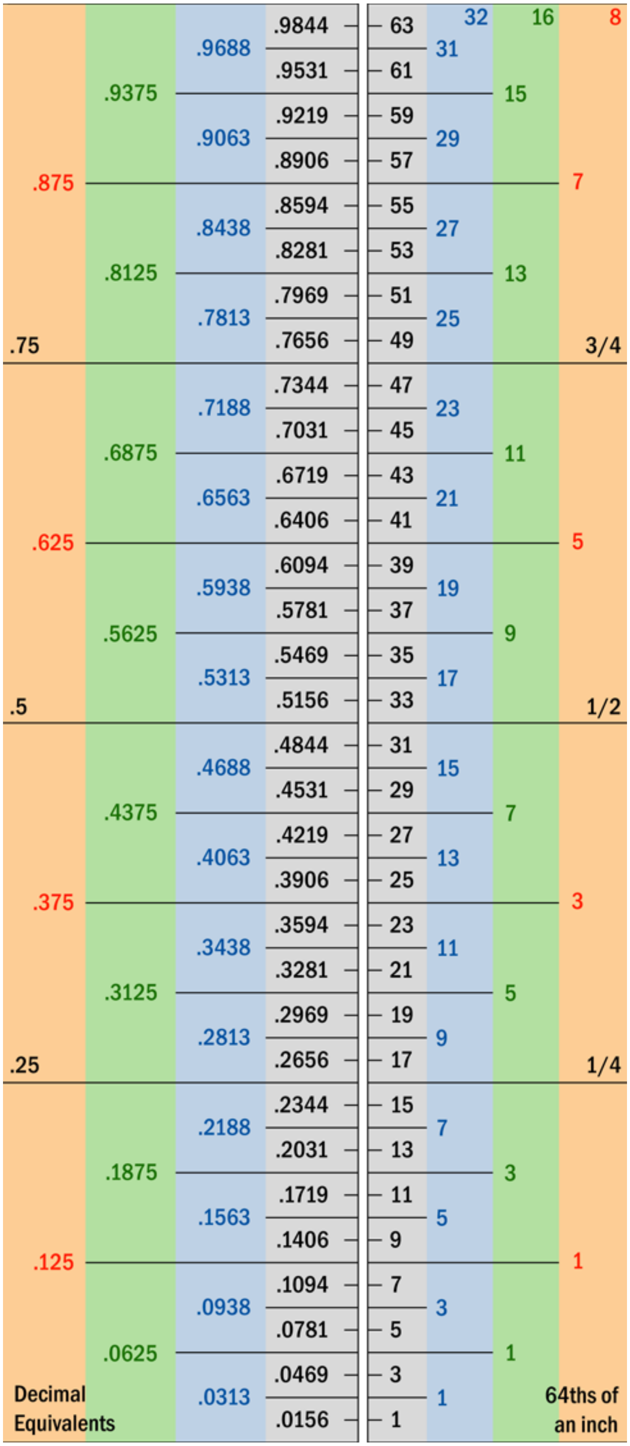
.875
.75
.625
.5
.375
.25
.125
.9375
.8125
.6875
.5625
.4375
.3125
.1875
.0625
.9688
.9063
.8438
.7813
.7188
.6563
.5938
.5313
.4688
.4063
.3438
.2813
.2188
.1563
.0938
.0313
.9844
.9531
.9219
.8906
.8594
.8281
.7969
.7656
.7344
.7031
.6719
.6406
.6094
.5781
.5469
.5156
.4844
.4531
.4219
.3906
.3594
.3281
.2969
.2656
.2344
.2031
.1719
.1406
.1094
.0781
.0469
.0156
63
61
59
57
55
53
51
49
47
45
43
41
39
37
35
33
31
29
27
25
23
21
19
17
15
13
11
9
7
5
3
1
32
31
29
27
25
23
21
19
17
15
13
11
9
7
5
3
1
16
15
13
11
9
7
5
3
1
8
7
3/4
5
1/2
3
1/4
1
Decimal Equivalents
64ths of an inch

Figure 3-72 *For conversion between decimal inch values and 64ths.*

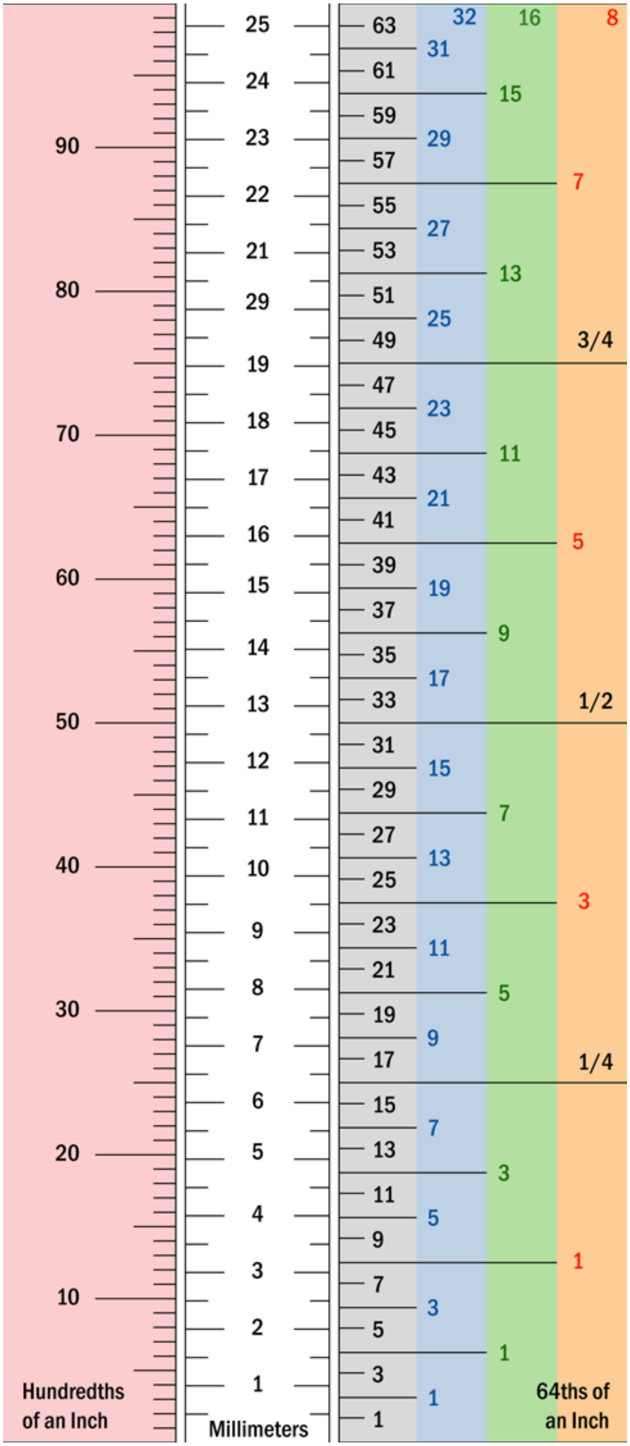

Figure 3-73 *For conversion between US measurements and metric (millimeters).*

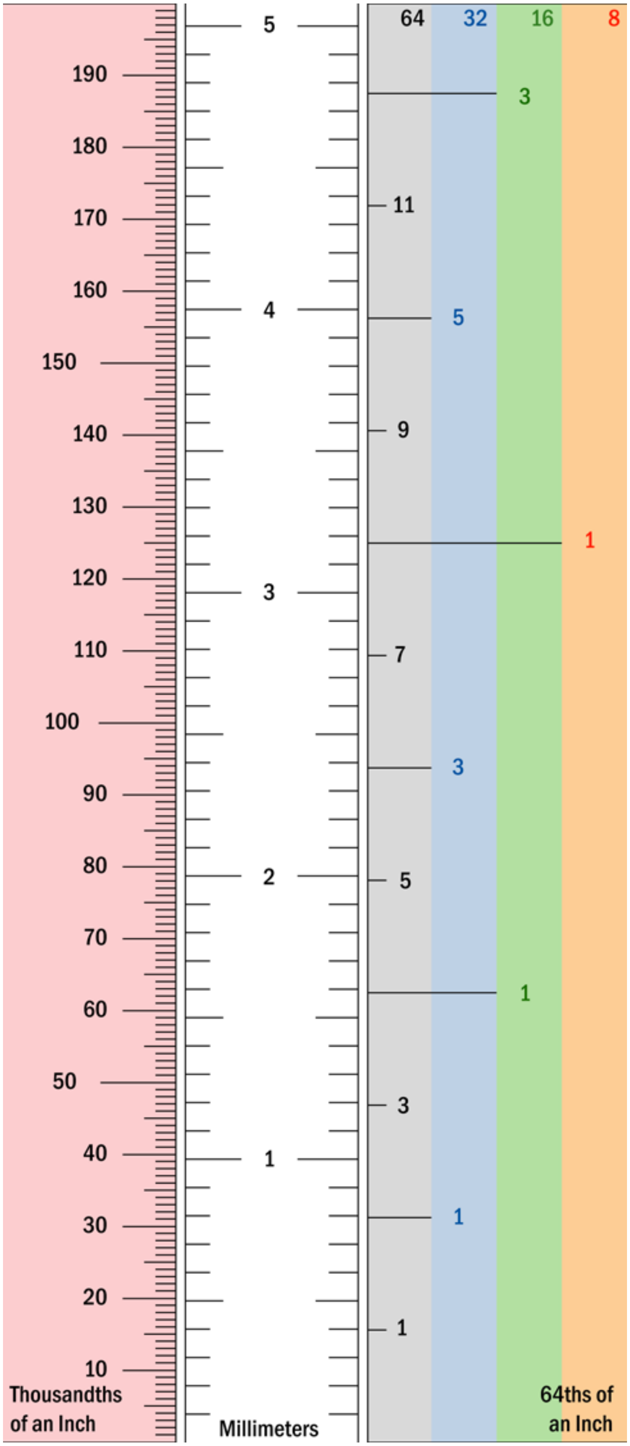

Figure 3-74 *For conversion between smaller values of US measurements and metric (tenths of millimeters).*

Experiment 15: Intrusion Alarm, Part One

Time now for an experiment that takes the knowledge you have acquired and applies it to a simplified but workable consumer product. You may not feel you need an intrusion alarm, but figuring out how to build one will be an excellent introduction to the process of creating circuits to perform tasks in the real world.

I should warn you that designing a circuit from scratch usually leads to unexpected problems and errors. It would be misleading to pretend otherwise. Consequently, in the sequence of steps described below, you'll find at least one setback and reversal—until we finally end up with a solid, workable system.

What You Will Need

- 9V battery and connector, or 9V AC-DC adapter (your choice)
- Breadboard, hookup wire, wire cutters, wire strippers, multimeter
- Generic LED (1)
- Transistor, 2N2222 (1)
- DPDT 9VDC relay (1)
- 1N4001 diode (1)
- Resistors: 470 ohms (1), 1K (1), 10K (1)

Wish List

This experiment is sufficiently complex; it requires a plan. But before I can develop a plan, I need to know what I want. This entails writing what I call a “wish list.” Along the way, I will also try to visualize how each requirement can be satisfied with the components that have been mentioned in previous experiments.

So, what does an intrusion alarm require?

1. Triggering system. The device has to detect if someone has entered the property. A sophisticated system using laser beams or ultrasound would be cool—but too difficult. Because this is a first attempt, I'll stick with widely available magnetic sensor switches for windows and doors.

2. Sound. The alarm should make some kind of distinctive, attention-getting, fluctuating sound.

3. Tamper-proof. No one should be able to kill the alarm by cutting a wire. In fact, tampering should actually make the alarm go off.

4. Sensors in series. To tamper-proof the system I can run a very small but constant current through a lot of sensor switches that are normally closed, and are wired in series. If any one switch opens, or if the wire itself is broken, this interrupts the current, which will start the alarm. I think most wired alarms are designed on this principle.

5. Off-to-on. If I use sensors in series, an “off” event, caused by opening a switch or breaking the circuit, has to turn the alarm on. Maybe a double-throw relay could do this. Current through the relay coil holds a pair of contacts open, until the flow of current stops, at which point the contacts close by default. But the relay would draw significant power while it's holding its contacts open. I want my alarm system to draw very little current while it's in “ready” mode, so that it can be powered by a battery. Alarm systems should never depend entirely on AC house current.

6. Maybe use a transistor? If I don't use a relay, a transistor could switch on the alarm when the circuit is interrupted. The base of the transistor can be held at a relatively low voltage until the circuit is broken. Then the voltage goes up, and the transistor switches on.

7. Arming the alarm. I need a little light that comes on when all the doors and windows are closed. This tells me that the alarm can be used. Then I want to press a button that starts a one-minute countdown, giving me time to leave. After a minute, the alarm is armed.

8. Self-sustaining. Once the alarm starts, I don't want it to stop easily. If someone opens a window, the alarm should continue to make a noise even if the person closes the window again. Maybe the transistor can trigger a relay, and when the relay switches on, it keeps itself powered? Or can a transistor do this?

9. Initial delay. I don't want the alarm to start whooping immediately, each time I walk into the protected area. I want it to wait for a minute, to give me time to reach it and switch it off. If I fail to deactivate it within that time window, then it can start making noise.

10. Deactivation with a code. Some kind of secret-code keypad would be good, to switch off the alarm.

Implementing the Wish List

This wish list sounds a bit ambitious, bearing in mind that the only thing you have built so far is a small oscillator using three transistors. But in fact most of the functions can be implemented fairly easily. I will leave some of the harder ones until later in the book, when I have established a broader knowledge base. In the end, I will be able to deal with everything on the list, and the components will all fit on a single breadboard (with the exception of a noisemaking circuit, which will be optional).

Magnetic Sensor Switches

Let's start with the component that triggers the alarm. A typical sensor switch consists of two modules: the magnetic module and the switch module. These are shown side by side in Figure 3-75.

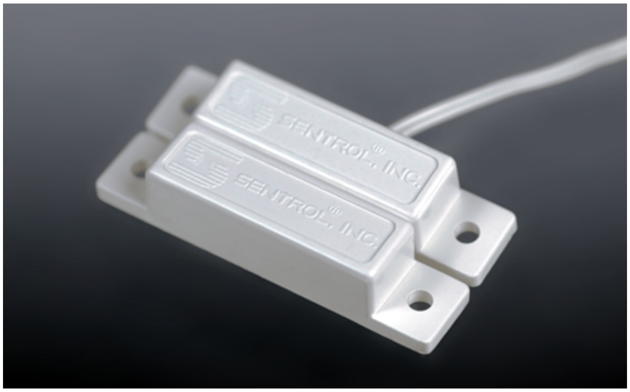

Figure 3-75 *A typical alarm sensor consists of a magnet in a plastic pod (bottom left), and a magnetically activated reed switch in a similar pod (top right).*

The magnetic module contains a permanent magnet, and nothing else. The switch module contains a *reed switch*, which makes or breaks a connection (like a contact inside a relay) under the influence of the magnet.

You attach the magnetic module to the moving part of a door or window, and attach the switch module to the window frame or door frame. When the window or door is closed, the magnetic module is almost touching the switch module. The magnet keeps the switch closed—until the door or window is opened, at which point the switch opens. See Figure 3-76 for a cutaway diagram of the magnet-switch combination.

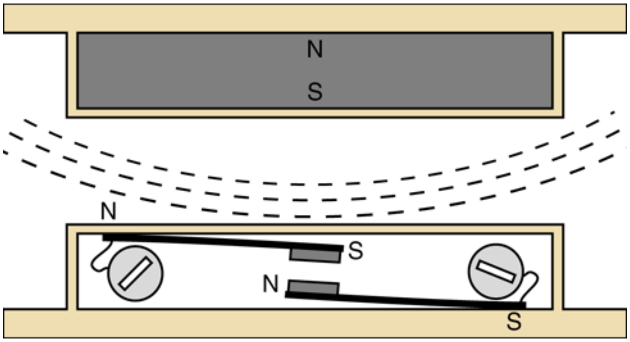

Figure 3-76 *This cutaway diagram shows the two components in a typical sensor for an alarm system: a reed switch (bottom) and the magnet that activates it (top).*

The switch consists of two flexible magnetized strips terminating in electrical contacts. Each strip connects with an external screw to which a wire can be attached.

When the magnet approaches the switch, it magnetizes the flexible strips, causing them to attract each other until the contacts close.

From my description, you can see that the reed switch is normally open (abbreviated NO), but is kept closed by the magnetic field. If you decide to buy alarm sensors, you should know that some of them contain reed switches that work the other way around. They are normally closed (abbreviated NC), but are opened by the magnetic field. Those are not the ones you want for this project.

A Break-to-Make Transistor Circuit

Now, how can we switch on the noisemaking part of the alarm? Remember, we will have a series of switches that are all normally closed, and when any one of them opens, the alarm must start.

Recall how an NPN transistor works. When the base is not so positive, the transistor blocks current between its collector and emitter. When the base is more positive, the transistor passes current.

Take a look at the schematic in Figure 3-77, which is built around our old friend the 2N2222 NPN transistor. For testing purposes, I have shown a normally closed pushbutton to represent an alarm sensor. I realize that you don't have a normally closed pushbutton in your supply of parts to build this circuit, but just use your imagination until we're ready to breadboard it.

So long as the pushbutton remains closed, it connects the base of the transistor to the negative side of the power supply through a 1K resistor. At the same time, the base is connected with the positive side of the power supply through a 10K resistor. Because of the difference in resistances, the base is closer to zero volts than nine volts, holding the transistor below its turn-on threshold. As a result, the transistor will not pass much current, and the LED should not have enough voltage to light up.

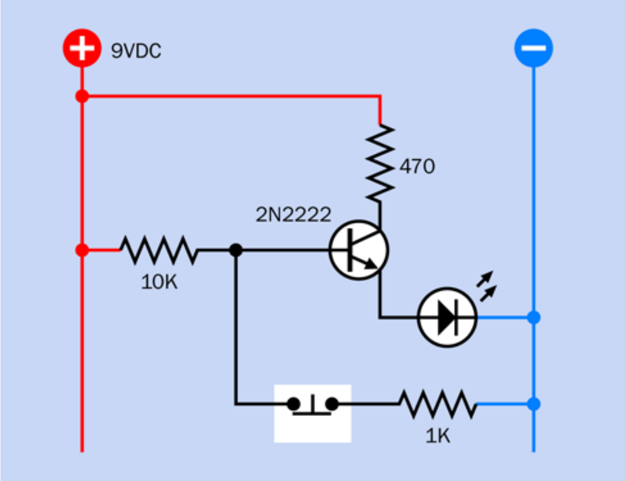

Figure 3-77 *A basic circuit in which an LED lights up when a normally closed pushbutton is opened.*

Now what happens when the pushbutton is opened? The base of the transistor loses its negative power supply and has only a positive power supply. It becomes much more positive, and tells the transistor to lower its resistance and pass more current. The LED now glows brightly. Thus, when the pushbutton breaks the connection, the LED is turned on.

This seems to be a workable system. Multiple sensors will be needed, for various doors and windows, but that's OK—we can wire as many as we like in series, as shown in Figure 3-78, where an alarm sensor can be substituted for each pushbutton. The wiring can be laid all around the house, as its total resistance should be low relative to the 10K resistor.

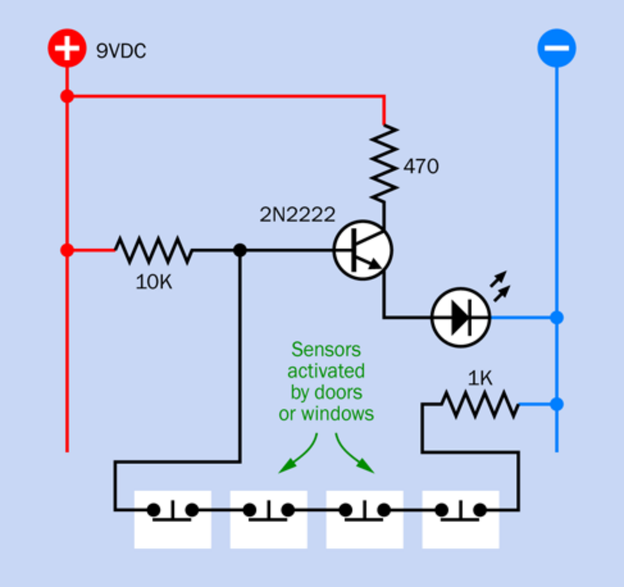

Figure 3-78 *In a network of sensors, wired in series, any one sensor will break continuity and trigger the transistor.*

While all the sensors remain closed, the transistor is drawing very little current—probably about 1mA. For development and demonstration purposes, you can run it with a 9-volt battery. For actual use, you would really want a 12-volt alarm battery that is maintained by an automatic charging system. That's outside the scope of this book, but bear in mind that alarm batteries and chargers are widely available if you ever want them.

Now suppose we swap out the LED and put a relay in there instead, as shown in Figure 3-79. (I've shown a double-pole relay, even though we have no use for the second pole right now.) So long as all the pushbuttons remain closed, the base of the transistor is held at a relatively low voltage, so the transistor does not apply power to the coil of the relay, and its contacts remain in the state shown.

When any sensor is opened, the higher voltage on the base of the transistor causes it to conduct current to the relay coil, which starts the alarm, as in Figure 3-80. (It's OK to use a relay in this mode, because the relay will not be "always on." It will normally be off, and will draw power only when the alarm is triggered.) Note that I eliminated the 470-ohm resistor from the circuit, because the relay doesn't need any protection from the power supply.

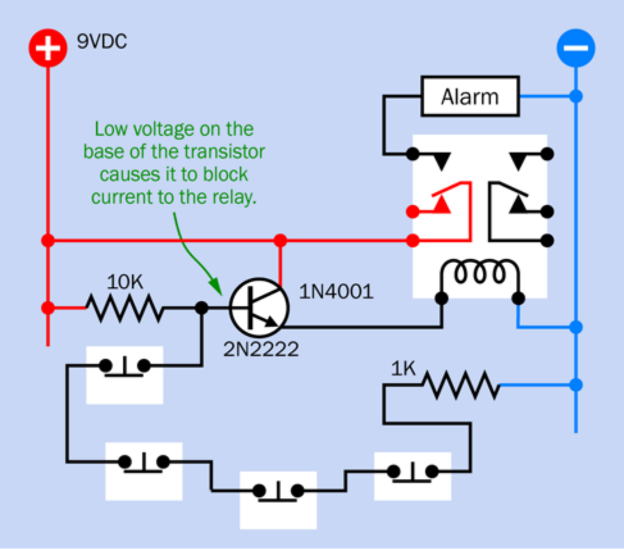
9VDC
Alarm
Low voltage on the base of the transistor causes it to block current to the relay.
10K
1N4001
2N2222
1K

Figure 3-79 *In this circuit, the relay will be activated when any switch in the sensor network is opened.*

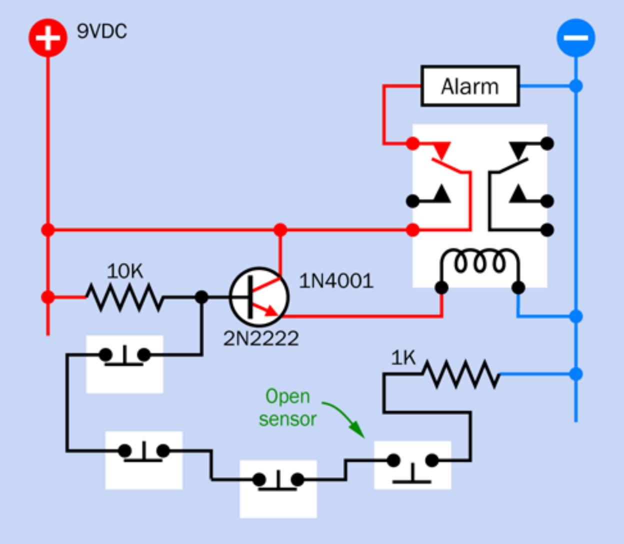

Figure 3-80 *Now that a sensor in the circuit is open, the relay is activated by the transistor.*

You can build this circuit yourself, using the same relay that you used in Experiment 7 (see "Experiment 7: Investigating a Relay" on page 60). But maybe you should wait until I develop it a little further.

A couple of things that you may want to consider:

- Will the relay overload the transistor? You can find the answer by looking at the datasheets for these two components.
- Remember that a transistor imposes a small voltage drop, even when it is "on." Will the voltage still be sufficient to activate a 9-volt relay? The datasheet for the relay will tell you the minimum operating voltage for its coil. You can verify this by testing it.

Self-Locking Relay

The circuit that I have developed so far will activate the alarm when any sensor is opened. That's good, but what happens if the sensor goes back into its closed state? The low voltage is reapplied to the base of the transistor, so it switches off the alarm. This is not good.

According to item #8 on my wish list, the alarm should be self-sustaining. It must continue making noise even after someone who has opened a door or window closes it again quickly. Therefore, the relay must lock itself on, somehow.

One way to do this would be by using a *latching relay*, which remains in either of two states, and only requires power to flip from one to the other. But a latching relay has two coils, and would require extra circuitry to unlatch it when you want to turn the alarm off. Really, it's easier to use a nonlatching relay, and I can think of a way to keep the relay switched on indefinitely, after it has received just one jolt of power.

The secret is revealed in Figure 3-81. In this view, the far-right pushbutton has closed again after being open, so the transistor has switched off—but the relay is still on, because a wire now connects its contacts back to its own coil. When the relay activates the alarm, it also activates itself.

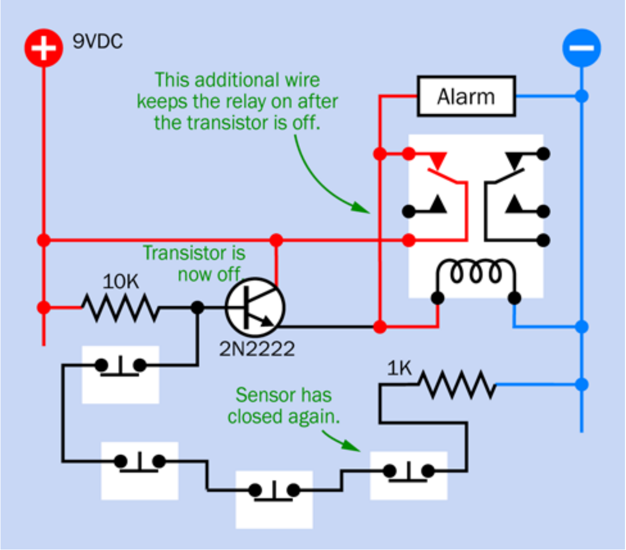
9VDC
This additional wire keeps the relay on after the transistor is off.
Alarm
Transistor is now off.
10K
2N2222
1K
Sensor has closed again.

Figure 3-81 *The sensor has closed again. The transistor is no longer active, but the alarm is locked on.*

Figure 3-82 clarifies this concept by showing the path that current can take. So long as the contacts of the relay are closed, the coil of the relay is energized via its own contacts. In this way, it keeps itself switched on.

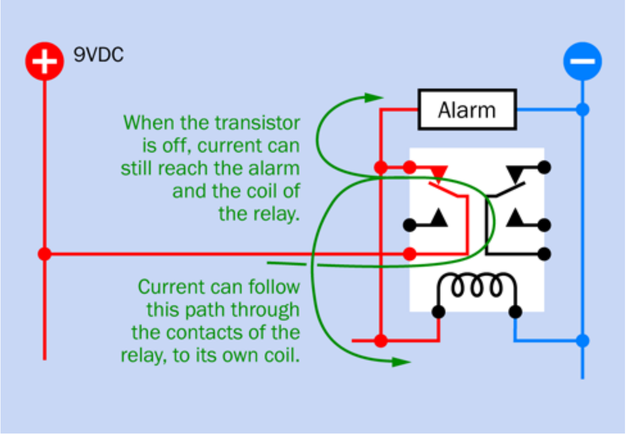
9VDC
When the transistor is off, current can still reach the alarm and the coil of the relay.
Alarm
Current can follow this path through the contacts of the relay, to its own coil.

Figure 3-82 *Close-up from the previous schematic, showing how the relay keeps itself switched on.*

Blocking Bad Voltage

This looks promising, but there's a problem. The picture in Figure 3-81 was not totally accurate. Take a look at Figure 3-83. The top part of this figure is another close-up of the relevant part of the circuit. When the alarm has locked itself on, but the transistor has turned off, current can run back from the relay coil to the emitter of the transistor. I should have colored this section of wire red, as it will be relatively positive.

Applying power backward through a transistor is not a nice thing to do. It can cause damage. What should be done about this? Maybe I can use something to block this reverse flow: a rectifier diode. This is shown in the bottom half of Figure 3-83.

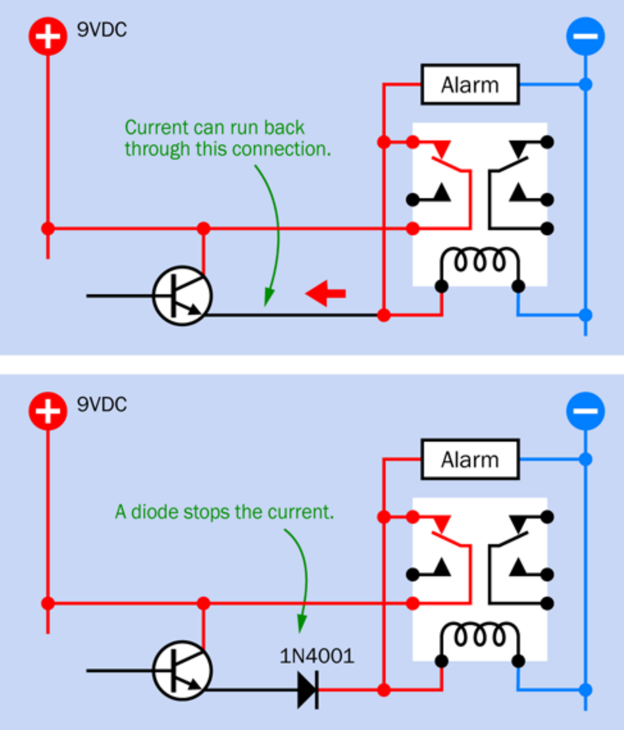
9VDC
Alarm
Current can run back through this connection.
9VDC
Alarm
A diode stops the current.
1N4001

Figure 3-83 *A diode can be added to prevent current from forcing its way back into the transistor when the alarm has locked itself on and the transistor switches off.*

A new version of the full circuit, including the diode, is shown in Figure 3-84.

But what exactly is a diode? Is it the same as a light-emitting diode (LED)? Well, yes and no.

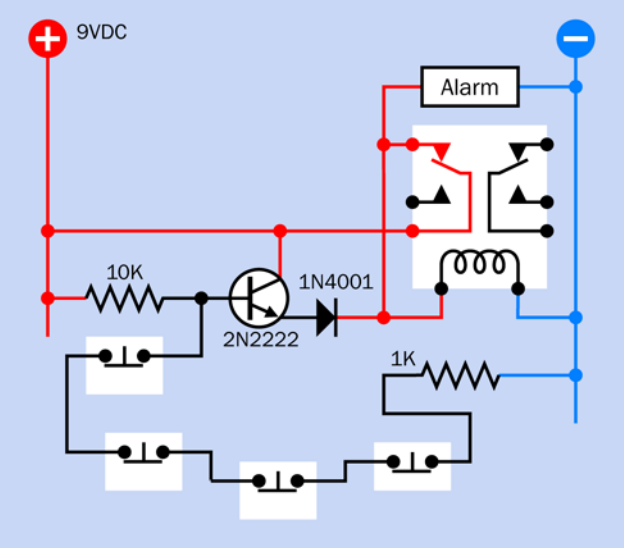

Figure 3-84 *The full circuit, now including the diode.*

Fundamentals: All About Diodes

A diode is a very early type of semiconductor. It allows electricity to flow in one direction, but blocks it in the opposite direction. Like its more recent cousin, the LED, a diode can be damaged by reversing the voltage and applying excessive power, but most diodes generally have a much greater tolerance for this than LEDs. In fact, they are designed to block reverse voltage, up to a limit specified by the manufacturer.

The end of the diode that stops positive voltage is always marked, usually with a circular band, as shown in Figure 3-25. The marked end is called the *cathode*. The other end is the *anode*, and remains unmarked. Diodes are sometimes useful in logic circuits, and can also convert alternating current (AC) into direct current (DC). If a diode isn't strong enough to withstand the current that you want it to block, you simply use a bigger diode. They come in many sizes.

It's good practice to use diodes at less than their rated capacity. Like any semiconductor, they can overheat and burn out if they are subjected to mistreatment.

The symbol for a diode looks like the heart of an LED, with the circle and the arrows removed. Three variants are shown in Figure 3-85.

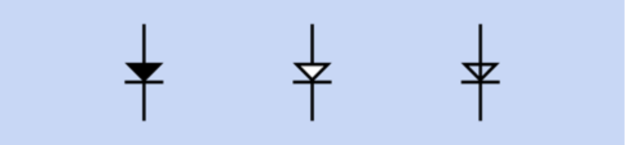

Figure 3-85 *Three schematic symbols that are used to represent a diode. They are functionally identical.*

One Problem Creates Another

Previously, I had to solve the problem of how to make the relay keep itself switched on. I solved that problem by adding an extra wire, but the wire created a new problem, in which current could flow back to the transistor. I solved that problem by adding a diode, but this creates yet another problem.

We have to pay a fee for the service that the diode provides, just as we have to pay a fee for the service that the transistor provides. In fact, because both of these components are semiconductors, the fee is very similar. It consists of a reduction in voltage.

When the relay is off, current must switch it on by passing through the transistor and then through the diode. After the relay is on, it keeps itself on, and that's not a problem. But a transistor imposes a penalty of around 0.7V, and the diode imposes an additional penalty of about 0.7V, making 1.4V total. This voltage penalty is fixed, regardless of the supply voltage.

I think 9-volt relays should work reliably at 7.6V. My Omron datasheet tells me that the G5V-2 series, which I have recommended, needs 75% of its supply voltage, which would be only 6.75V. That seems like a reasonable margin of error.

But what if someone substitutes a different relay? Some have narrower specifications than others. Or what if someone is using a battery to power the circuit, and its voltage drops below 9V? A designer should always expect the unexpected, and as a general principle, should use components as close to their rated values as possible.

A couple of readers wrote to me about the issue of voltage reduction when the circuit appeared in the first edition of the book. (Yes, I do pay attention to reader feedback.) At that time, I had specified a 12VDC power source, and felt that the 1.4V penalty imposed by the transistor and the diode would be acceptable. But in this edition, I decided that all the experiments should work with 9VDC power supplies, so that you don't have to buy an AC adapter and can just use 9V batteries if that's what you prefer. Unfortunately a deduction of 1.4V from 9V is not so acceptable.

You see how a decision leads to consequences. Now that I am using a 9VDC power supply, I think I need a better way to make the relay lock itself on.

Solving the Problem

The first step to solving a problem is to be very clear about what's happening.

The task of controlling the alarm is shared by two components: the transistor and the relay. The transistor starts the alarm. After that, the transistor doesn't do anything. It switches off, and the relay has the task of keeping itself locked on. The weakness in this system is that when a task is shared by two components, they can interfere with each other. A better plan would be to have one component in charge of everything. I should maintain the transistor in a controlling role. It should keep itself turned on, and so long as it is on, it will keep the relay on.

Ah—now I see how to fix it. All I need is to use the second pole of the relay (which is the same relay that you already used in Experiment 7). I can use the contacts in the second pole, which are normally closed, to ground the chain of sensors, as shown in Figure 3-86.

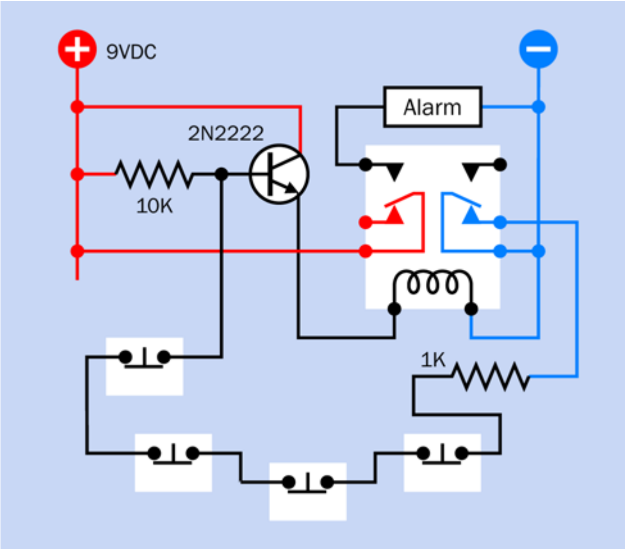

Figure 3-86 *The chain of sensors is now grounded through the right-hand contacts in the relay, which are normally closed.*

Here's how it will work:

The base of the transistor now connects to the negative side of the power supply through all the sensors, and the 1K resistor, and the right-hand relay contacts (which are normally closed). So long as this chain of connections is unbroken, the base of the transistor is at a low enough voltage to stop the transistor from passing current.

Now someone opens a sensor. The base of the transistor isn't grounded anymore, so the transistor activates the relay. The relay closes the left-hand contacts, which start the alarm. But the relay also opens the right-hand contacts.

Now if someone re-closes the sensor, it makes no difference anymore, because the right-hand contacts of the relay are open, and have cut off the connection to the negative side of the power supply. The transistor continues to pass current, and the relay remains active. This is shown in Figure 3-87.

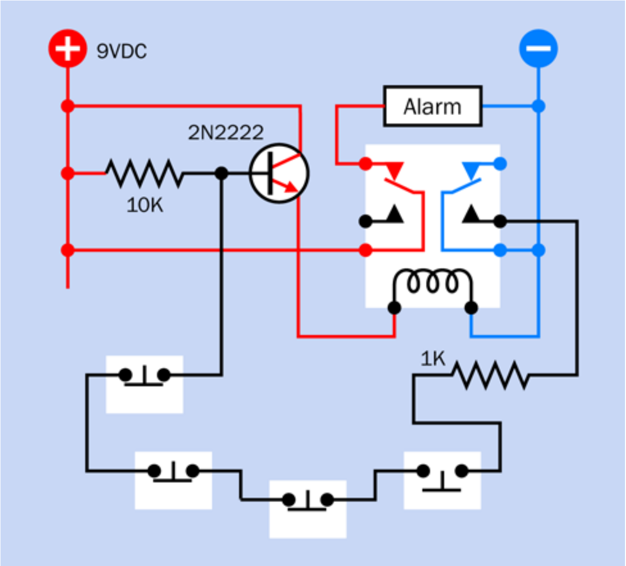

Figure 3-87 *Now that a sensor has been opened, the transistor remains energized even if the sensor is subsequently closed.*

This solves the problem.

Protection Diode

As you saw above, I eliminated the diode from the circuit. But if you take a look at Figure 3-88 (which I promise is the last version, at least for now) you will notice that the diode has found its way back in again, although it's doing something very different. It is now in parallel with the coil of the relay. What's it doing there, of all places?

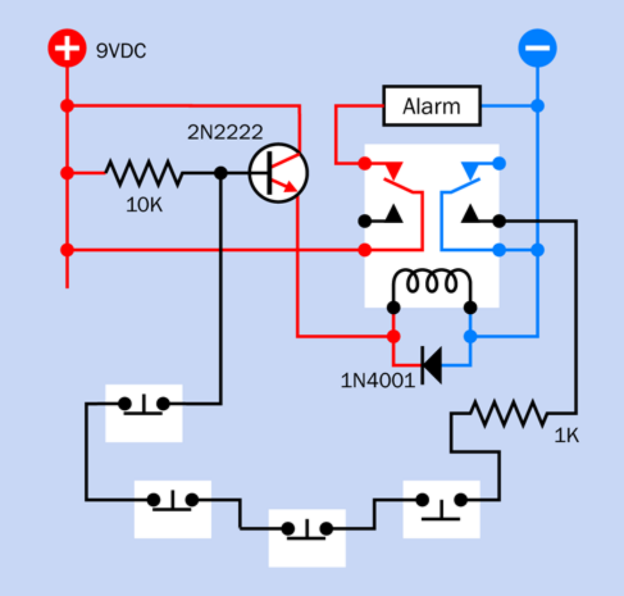

Figure 3-88 *The diode has returned, now serving a function as a protection diode.*

Much later in the book, I'll be discussing coils. One thing I can tell you right now is that a coil of wire stores energy when you apply power, and releases the energy when you disconnect the power. The release of energy creates a surge of current that can harm some types of components, especially semiconductors.

Therefore, it is standard procedure to add a *protection diode* across a relay coil. The diode is oriented so that it blocks the normal flow of current, forcing it to go through the coil, which is where we want it. But when the flow stops, and the coil tries to release its energy, the diode is there, saying to the relay, "I have a very low resistance in that direction. Why don't you shunt the current through me, instead of hassling the other components with it?"

And that is precisely what happens.

If you're only using a small relay with a small coil that doesn't take much current, you can get by without using a protection diode. Still, it's good practice, and you should always be in the habit of using one.

Time to Breadboard

I've given you a lot of explanations in this experiment, which is not what I normally like to do. But I had to show how a circuit is developed from the ground up. Now, finally, I'd like you to build it—otherwise, how will you know if it really works?

Figure 3-89 shows a breadboard layout. Instead of a noisemaker for the alarm, I've used an LED for demonstration purposes. I'll discuss noisemaking options in a moment.

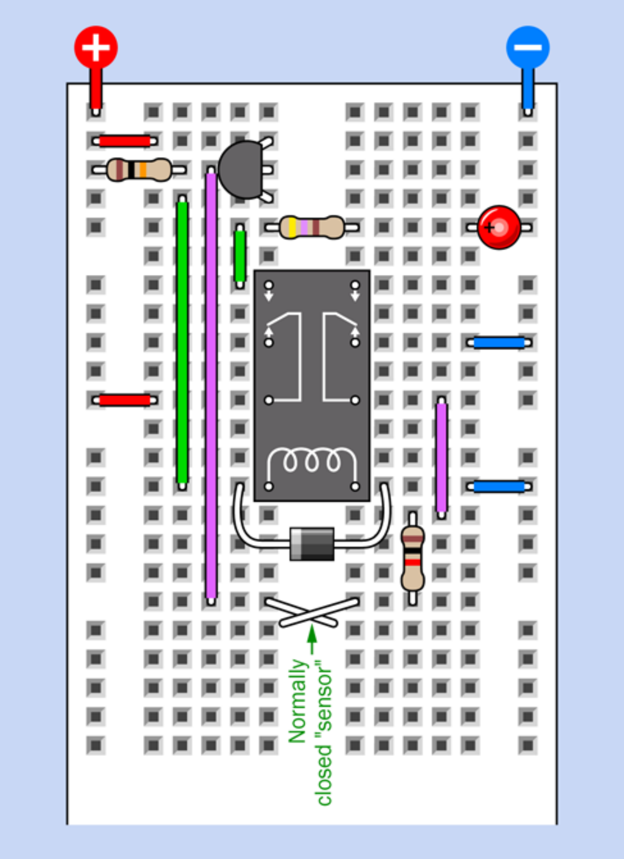

Figure 3-89 *Breadboarded alarm circuit, final version.*

Figure 3-90 shows an x-ray view of the breadboarded circuit

To simulate the alarm sensors on the breadboard, I should have used normally closed pushbuttons. But I wanted to minimize the component costs, and if you actually decide to use this alarm circuit, you'll want magnetic sensors, not pushbuttons. Consequently, as a substitute, I used two normally closed pieces of wire. This is adequate for testing. I will call them the “sensor wires.” You can see them crossing over each other below the relay.

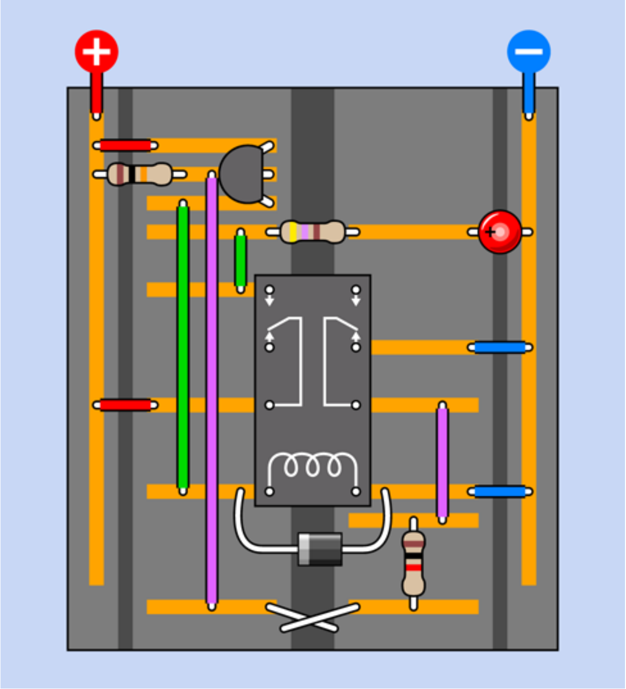

Figure 3-90 *X-ray view of the breadboarded alarm circuit.*

Make sure the wires are touching each other when you apply power to the circuit. Initially, nothing should happen.

Now disconnect the sensor wires. The LED comes on, and if you build the next version of this circuit, a noisemaker will sound, showing that the alarm has been triggered.

Now reconnect the sensor wires, imitating a situation where an intruder opened a window, heard the alarm, and quickly closed the window again. If you wired your circuit correctly, the LED will stay on.

So far so good. We have a functional circuit here. The alarm locks itself on.

But in that case—how do you ever get it to stop?

No problem. Just disconnect the power. The relay relaxes back to its default position, and the next time you apply the power, it will be in standby mode again. In a finished version of this project, you should have to enter some kind of secret code on a keypad to turn the alarm off. In Experiment 21, I’ll be suggesting a way to create a passcode-protected system. You’ll need to use logic chips, which we haven’t dealt with yet.

Adding the Sound

For the alarm sound, you could use the oscillator circuit and loudspeaker from Experiment 11. Really, though, there is a better way. A little integrated circuit chip known as a 555 timer is a better tool for the job—and it just happens to be the next thing that I want to tell you about, in Experiment 16.

The 555 timer can also satisfy items #7 and #9 on my wish list, which require a delay before the alarm starts. Therefore, I’m going to put the alarm project on hold, and we’ll finish it completely in Experiment 18.

Reference: Take-Home Messages

Even though the alarm project isn’t finished yet, it raised several important points. I’m going to summarize them here, for future reference.

- You can use a transistor to provide a high output in response to a low input, and vice versa.
- You can wire a relay to lock itself into “on” mode, simply by feeding current back to its own coil.
- A diode can stop current from flowing into places where you don’t want it to go.
- When forward current passes through a diode, the voltage is reduced by about 0.7V.
- A transistor also reduces voltage by about 0.7V.
- Voltage reduction imposed by semiconductors remains the same regardless of the supply voltage. Consequently, the reduction is more significant when the supply voltage is lower.
- A relay coil can create back-EMF (a pulse of reverse current) when it is switched off.
- A protection diode in parallel with a relay coil can suppress back-EMF. The diode should be oriented so that it blocks the normal flow of current but passes the reverse pulse created by the coil.

Chips, Ahoy!

4

Before I get into the fascinating topic of integrated circuit chips (often referred to as *ICs* or simply *chips*), I have to make a confession. Some of the things I asked you to do in previous experiments could have been done a bit more simply, if we had used chips.

Does this mean you have been wasting your time? Absolutely not! I firmly believe that by building circuits with individual components such as transistors and diodes, you acquire the best possible understanding of the principles of electronics. Still, you are going to find that chips containing dozens, hundreds, or thousands of transistor junctions will enable some shortcuts.

You may also find chips curiously addictive to play with—although you may not become quite as excited as the character in Figure 4-1.

The tools, equipment, components, and supplies described below will be useful in Experiments 16 through 24, in addition to items that have been recommended previously.

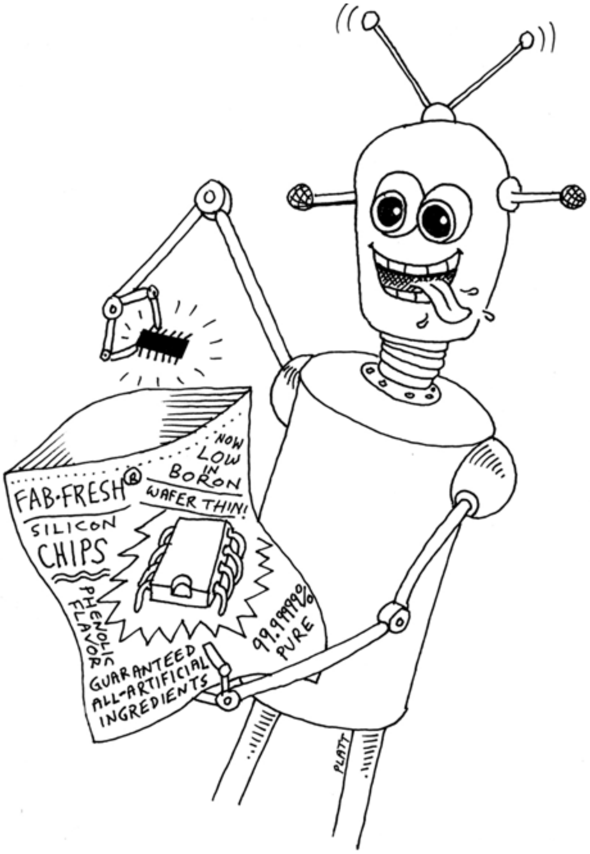
FAB-FRESH®
SILICON CHIPS
PHENOLIC FLAVOR
NOW LOW IN BORON
WAFER THIN!
99.9999% PURE
GUARANTEED ALL-ARTIFICIAL INGREDIENTS
PLATT

Figure 4-1 *My role model.*

Necessary Items for Chapter Four

The only new tool that you might consider using in conjunction with chips is a logic probe. This tells you whether a single pin on a chip has a high or low voltage, which can be helpful in figuring out what your circuit is doing. The probe has a memory function so that it will light its LED, and keep it lit, in response to a pulse that may have been too quick for the eye to see.

Some of my readers disagree with me, but I regard the logic probe as optional, not essential. Search online and buy the cheapest one you can find. I don't have any specific brand recommendations.

Components

As before, if you want kits containing components, see "Kits" on page 311. If you prefer to buy your own components from online sources, see "Components" on page 317. For supplies, see "Supplies" on page 316.

Fundamentals: Choosing Chips

Figure 4-2 shows two integrated circuit chips. The one at the top is of the old-school, *through-hole* design, with pins spaced at 1/10" so that they will fit through the holes in your breadboard or perforated board. These are the chips I will be using exclusively, because they're easy to handle. The smaller chip is a *surface-mount* design, which I will not be using, because they don't fit breadboards or perforated boards and are difficult to handle.

Many through-hole and surface-mount chips are functionally identical. The only difference is the size (although some surface-mount versions use a lower voltage).

The body of a chip is usually made of plastic or resin and is often referred to as the *package*. The traditional chip is usually sold in a *dual-inline package*, meaning that it has two (i.e., dual) rows of pins. The acronym for this package is *DIP* or (when it is made of plastic) *PDIP*.

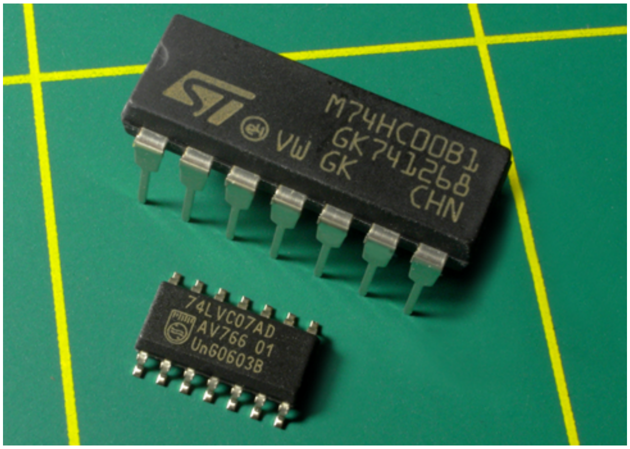

Figure 4-2 *A through-hole chip (top) and surface-mount chip (bottom).*

Surface-mount packages often are identified with acronyms beginning with letter S, as in SOIC, meaning small-outline integrated circuit. Numerous surface-mount variants exist, with different pin spacing and other specifications. They are all outside the scope of this book, and if you buy your own components, you should be careful not to select them by mistake.

Inside the package, the circuit is etched on a tiny wafer of silicon, which is where the term "chip" comes from, although the whole component is now usually referred to as a chip, and I will follow that convention here. Tiny wires inside the package link the circuit with the rows of pins that protrude on either side.

The PDIP chip in Figure 4-2 has seven pins in each row, making a total of 14. Other chips may have 4, 6, 8, 16, or more pins.

Just about every chip has a part number printed on it. Notice in the photograph that even though the chips look quite different from each other, they both have "74" in their part numbers. This is because both of them are members of the family of logic chips that were assigned part numbers from 7400 and upward when they were introduced several decades ago. This is often referred to as the 74xx family, and I'll be using these chips a lot.

Take a look at Figure 4-3. The initial letters identify the manufacturer, which you can ignore, as it really makes no difference for our purposes. (If you are wondering why "SN" identifies Texas Instruments, it's because the company used to call their chips "semiconductor networks" in the early days.)

Skip the letters until you get to the "74." After that, you find two more letters, which are important. The 7400 family has evolved through many generations, and the letter(s) inserted after the "74" tell you which generation you're dealing with. Generations have included: 74L, 74LS, 74C, 74HC, and 74AHC. There are many more.

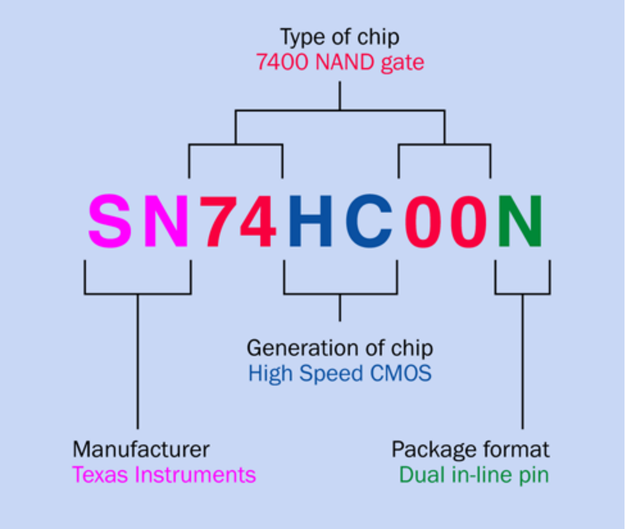

Figure 4-3 *How to decode the part number of a chip in the 74xx family.*

Generally speaking, later generations tend to be faster or more versatile than earlier generations. In this book, I am using the HC generation of the 7400 family exclusively, because almost all 7400 chips are available in it, the cost is moderate, and the chips don't use a lot of power. For our purposes, the extra speed offered by later generations is not relevant—although you can certainly use the HCT generation if you prefer.

Following the letters identifying the generation, you'll find a sequence of two, three, four, or (sometimes) five digits. These identify the specific function of the chip. Following the digits is another letter, or two letters, or more. For our purposes, those terminating letters are not important.

Looking back at Figure 4-2, the DIP chip part number, M74HC00B1, tells you that it is a chip made by STMicroelectronics, in the 74xx family, HC generation, with its function identified by numerals 00.

The purpose of this long explanation is to enable you to interpret catalog listings if you go chip shopping. You can search for "74HC00" and the search engines at online vendors are usually smart enough to show you appropriate chips from multiple manufacturers, even though there are letters preceding and following the term that you're searched for.

Just be sure that they will fit your breadboard. Limit your search results to DIP, PDIP, or through-hole packages. If the part number begins with SS, SO, or TSS, it's absolutely definitely surface-mount, and you don't want it. For a lot more information on searching and shopping, see "Searching and Shopping Online" on page 311.

All the chips needed for the experiments in this chapter of the book are listed in Figure 6-7. You will need a few other types of components, which I will list here.

Optional: IC Sockets

If you plan to immortalize any of your circuits in solder, I suggest you avoid soldering chips directly, because if you make a wiring error or damage the chip, you have to desolder multiple pins in order to remove it. This is very difficult. To avoid the problem, buy some DIP sockets, solder the sockets onto the board, and then plug the chips into the sockets. You can use the cheapest sockets you can find (you don't need gold-plated contacts for our purposes). You will need 8-pin, 14-pin, and 16-pin sockets. Quantity of each: 5 minimum. Two sockets are shown in Figure 4-4.

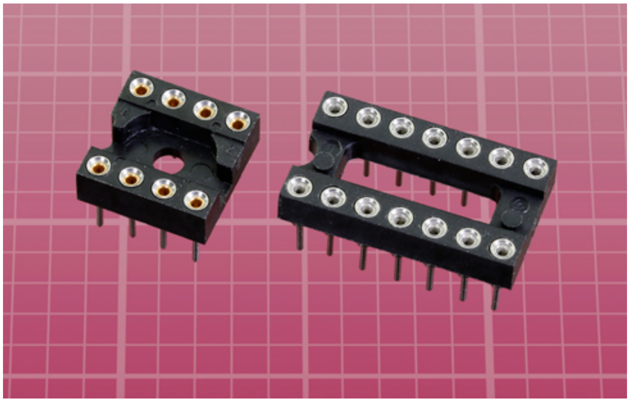

Figure 4-4 *To avoid the risk of damaging a chip by soldering it directly, it can be mounted in an IC socket after the socket has been soldered into a circuit board.*

Essential: Subminiature Slide Switch

A *slide switch* has a tiny lever that you slide to and fro with the tip of your finger, making and breaking an electrical contact inside the switch, as shown in Figure 4-5. It has three pins spaced 0.1" apart (2.54mm in metric). If you buy your own components, see "Other Components" on page 319 and go down to the subhead "Components for Chapter Four" for more information about switches.

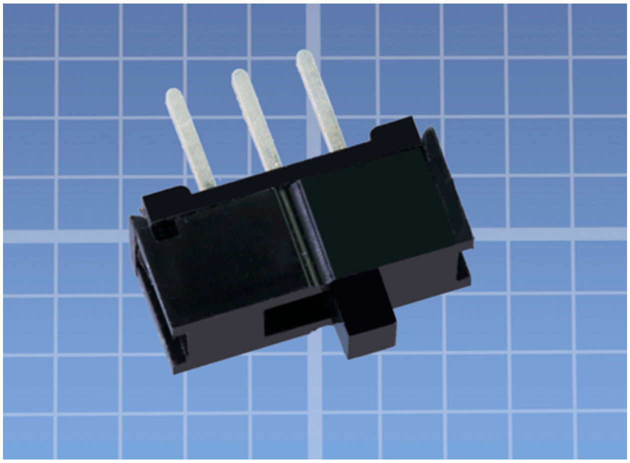

Figure 4-5 *The subminiature slide switch recommended for projects in this book.*

Caution: Switching Overload

A very small slide switch is not designed to switch significant currents or voltages. It is designed for low-powered circuits. A limit may be as low as 100mA at 12VDC. This is sufficient for our purposes. Check the manufacturer's datasheet if you want a slide switch to do more than this.

Essential: Low-Current LEDs

HC series logic chips are not designed to deliver current much beyond 5mA. You can take as much as 20mA from them to drive an LED, but this will pull down the output voltage, making it unsuitable as an input to other logic chips. I am suggesting low-current LEDs for all your experiments with logic chips.

Remember that low-current LEDs require higher-value series resistors, because they don't tolerate as much current as generic LEDs. I will mention this wherever it is important.

Essential: Numeric Displays

One of the chip projects will display its output using seven-segment numeric displays—the simple kind of digits that you still find on digital clocks and microwave ovens. See Figure 4-6. For purchasing information, see "Other Components" on page 319 and go down to the subhead "Components for Chapter Four."

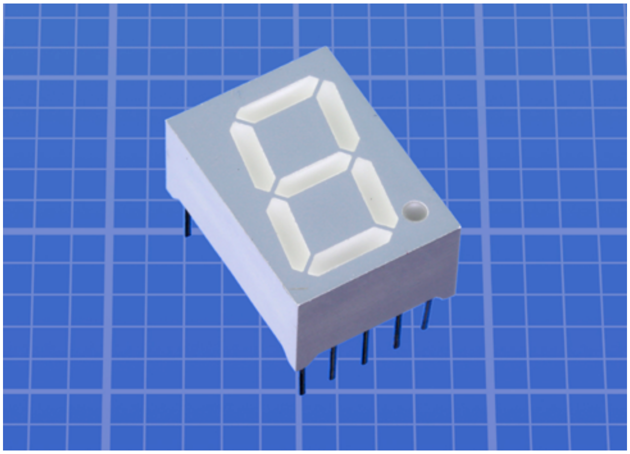

Figure 4-6 *Seven-segment displays are the cheapest way to show a numeric output and can be driven directly by some CMOS chips.*

Essential: Voltage Regulator

Because many logic chips require precisely 5 volts DC, you need a voltage regulator to guarantee this. The LM7805 does the job. The chip number will be preceded or followed with an abbreviation identifying the manufacturer and package style, as in the LM7805CT from Fairchild. Any manufacturer will do, but the regulator should look like the one in Figure 4-7. (This is known as the TO220 package style.) You need them in any logic circuit, so five would be a good number to have available.

Optional Extras

To complete the alarm system in Experiment 18, you will require magnetic sensors that you can apply to doors or windows, such as the Directed model 8601, available from dozens of sources online.

If you expect to move a project from a breadboard to a permanent enclosure, the tactile switches that you have been using will be insufficiently sturdy or accessible. For Experiment 18, you will need a full-size DPDT pushbutton switch, ON-(ON) type, with solder terminals. If you search eBay for "DPDT pushbutton" there will be no shortage of options.

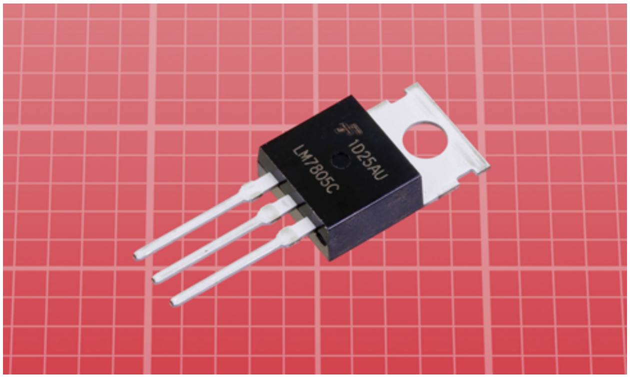

Figure 4-7 *Many integrated circuit chips require a controlled power supply of 5 volts, which can be delivered by this regulator from a supply of 7.5 to 12 volts.*

Background: How Chips Came to Be

The concept of integrating solid-state components into one little package originated with British radar scientist Geoffrey W. A. Dummer, who talked about it for years before he attempted, unsuccessfully, to build one in 1956. The first true integrated circuit wasn't fabricated until 1958 by Jack Kilby, working at Texas Instruments. Kilby's version used germanium, as this element was already in use as a semiconductor. (You'll encounter a germanium diode when I deal with crystal radios in Experiment 31.) But Robert Noyce, pictured in Figure 4-8, had a better idea.

Figure 4-8 *Robert Noyce, who patented the integrated circuit chip and cofounded Intel.*

Born in 1927 in Iowa, Noyce moved to California in the 1950s, where he found a job working for William Shockley. This was shortly after Shockley had set up a business based around the transistor, which he had coinvented at Bell Labs.

Noyce was one of eight employees who became frustrated with Shockley's management and left to establish Fairchild Semiconductor. While he was the general manager of Fairchild, Noyce invented a silicon-based integrated circuit that avoided the manufacturing problems associated with germanium. He is generally credited as the man who made integrated circuits possible.

Early applications were for military use, as Minuteman missiles required small, light components in their guidance systems. These applications consumed almost all chips produced from 1960 through 1963, during which time the unit price fell from around $1,000 to $25 each, in 1963 dollars.

In the late 1960s, MSI (medium-scale integration) chips emerged, each containing hundreds of transistors. LSI (large-scale integration) enabled tens of thousands of transistors on one chip by the mid-1970s, and today's computer chips can contain several billion transistors.

Robert Noyce eventually cofounded Intel with Gordon Moore, but died unexpectedly of a heart attack in 1990. You can learn more about the fascinating early history of chip design and fabrication at the Silicon Valley Historical Association.

Experiment 16: Emitting a Pulse

I'm going to begin our experiments with chips by introducing you to the most successful one ever made: the 555 timer. You can find numerous guides to it online, so why do I need to discuss it here? I have three reasons for doing so:

It's unavoidable. You simply have to know about this chip. Some sources estimate that more than 1 billion are still being manufactured annually. It will be used in one way or another in most of the remaining circuits in this book.

It's useful. The 555 is probably the most versatile chip that exists, with endless applications. Its relatively powerful output (rated at up to 200mA) is extremely useful, and the chip itself is hard to damage.

It's misunderstood. After reading literally dozens of guides, beginning with an early Signetics datasheet and making my way through various hobby texts, I concluded that the inner workings of the chip are seldom explained on an introductory level. I want to give you a graphic understanding of what's happening inside it, because if you don't have this, you won't be in a good position to use the chip creatively.

What You Will Need

- Breadboard, hookup wire, wire cutters, wire strippers, multimeter
- 9VDC power supply (battery or AC adapter)
- Resistors: 470 ohms (1), 10K (3)
- Capacitors: 0.01µF (1), 15µF (1)
- Trimmer potentiometers: 20K or 25K (1), 500K (1)
- 555 timer chip (1)
- Tactile switches (2)
- Generic LED (1)

Know Your Chips

The pins of a 555 timer are numbered counterclockwise (seen from above), as shown in Figure 4-9. The package has a notch, or a dimple, or both, at the end which is considered the top. The pin spacing is 1/10".

All other through-hole chips have the same specification, although they may have more pins. Usually (not always) the horizontal spacing between the two rows of pins is 3/10", which means that the chip neatly straddles the channel down the middle of a breadboard, and the conductors inside the breadboard allow you to have access to each pin of the chip. Yes, *that's* why the breadboard is designed that way.

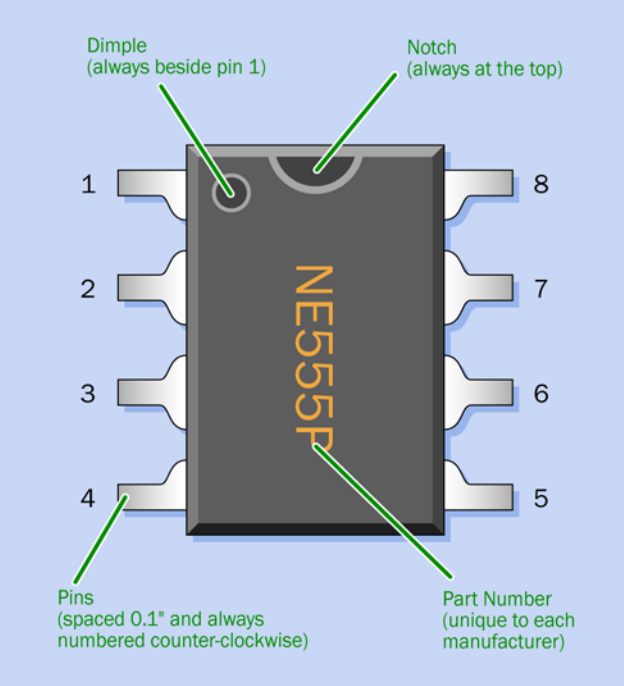
Dimple
(always beside pin 1)
Notch
(always at the top)
1
2
3
4
8
7
6
5
NE555P
Pins
(spaced 0.1" and always
numbered counter-clockwise)
Part Number
(unique to each
manufacturer)

Figure 4-9 *The package design of an eight-pin chip. While virtually all chips have a semicircular notch at the top, some do not have a dimple beside pin 1.*

Monostable Test

The pins on a 555 timer also have names, as shown in Figure 4-10. A diagram like this tells you the *pinouts* of the chip. I'll be explaining the function of each pin—but as usual, I prefer you to make a preliminary investigation of your own.

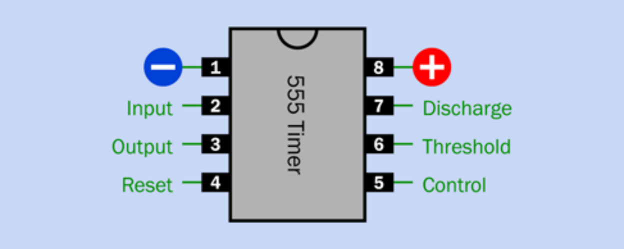
1
Input 2
Output 3
Reset 4
8
7 Discharge
6 Threshold
5 Control
555 Timer

Figure 4-10 *The pinouts of the 555 timer chip.*

The schematic of a test circuit for the timer is shown in Figure 4-11.

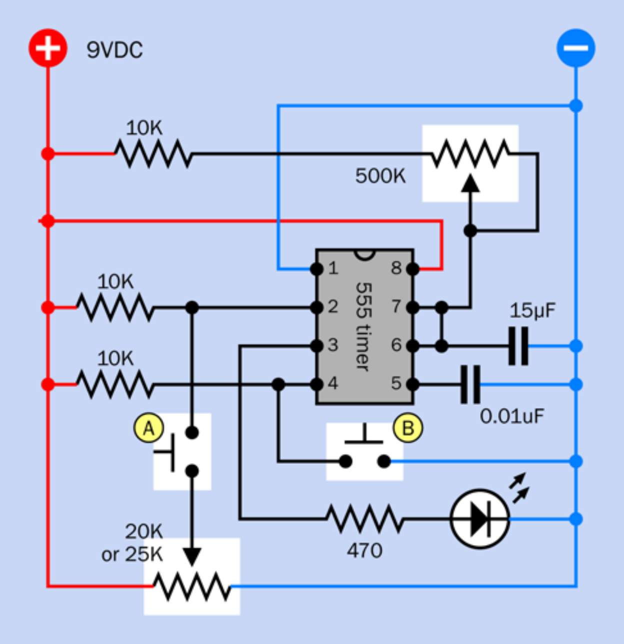

Figure 4-11 *A circuit to assist in your investigation of the 555 timer chip.*

You can set up this circuit on a breadboard as shown in Figure 4-12. Note that near the bottom-left corner there is a short red jumper connecting the top section of the positive bus to the section immediately below it. The jumper is there in case your breadboard is the type that has a break in its bus.

The component values are shown in Figure 4-13. To assist you in visualizing the connections, an x-ray version is shown in Figure 4-14.

Apply some power, and nothing happens. The timer is waiting for you to trigger it. Set it up by turning the 500K trimmer to the middle of its range.

Now rotate the 20K trimmer all the way counterclockwise, and press button A. If still nothing happens, rotate the 20K trimmer all the way clockwise, and try again. One of these settings or the other should create a pulse from the LED, depending on which way around you plugged in your trimmer. If you don't get anything, there is an error in your circuit.

Check the schematic, and you see that pin 2 of the timer—the trigger pin—is hardwired through a 10K resistor to the positive side of the power supply. But a purple wire also connects with the trigger pin, and runs down, through a tactile switch, to the trimmer. If the trimmer is rotated so that the wiper connects directly to the negative ground side of the power supply, this will allow the tactile switch to overwhelm the 10K resistor and apply low voltage to pin 2. This triggers the timer.

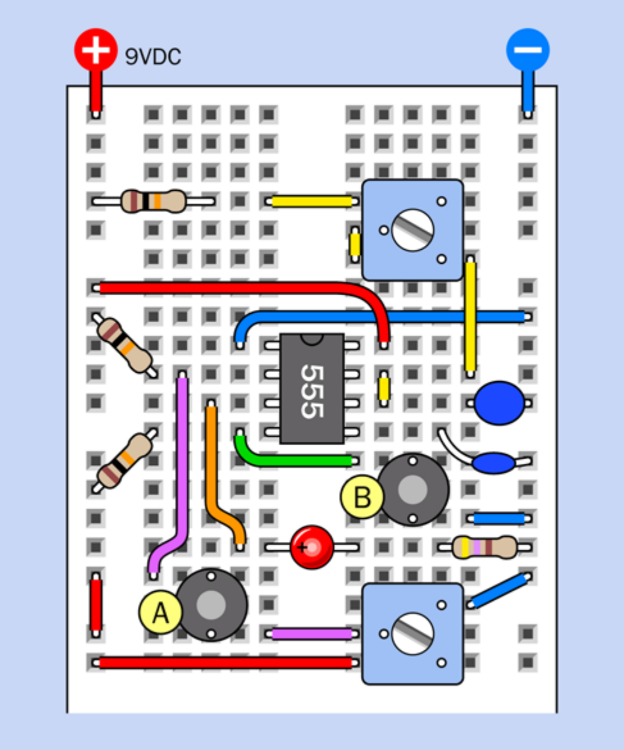

Figure 4-12 *Breadboarded version of the timer test circuit.*

If the 20K trimmer is turned all the way in the opposite direction, button A will apply positive voltage directly to pin 2, and because pin 2 already has positive voltage through the 10K resistor, an additional positive voltage through button A makes no difference.

- Positive voltage on the trigger pin is ignored by the chip.
- A drop in voltage on the trigger pin will trigger the chip.

But, how positive is positive, and how much of a drop will be low enough to act as a trigger? Let's find out.

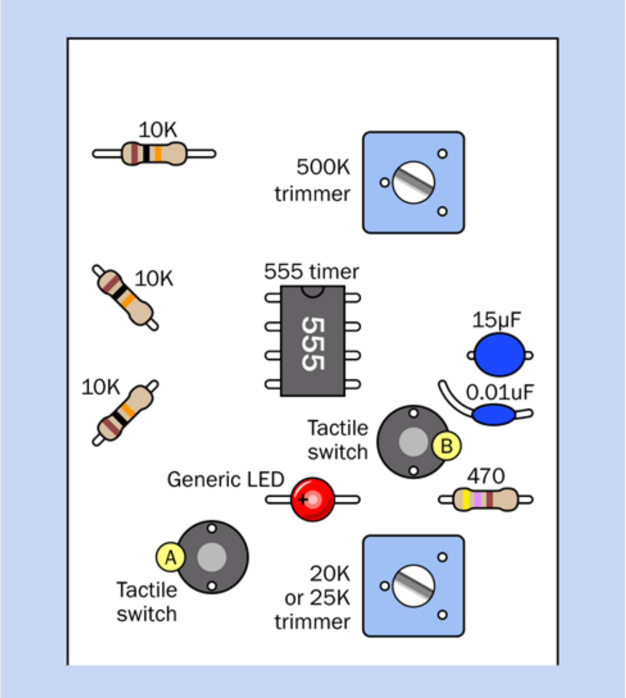

Figure 4-13 *Component values for the timer test circuit.*

Get out your meter, set it to measure DC volts, and measure the voltage between pin 2 and negative ground while you adjust the 20K trimmer to various settings and press button A. I'm betting that when you press the button to apply a voltage below 3 volts to pin 2, the timer will flash the LED. Above 3 volts, I doubt that anything will happen.

- The timer is triggered by a voltage on its trigger pin that is one-third of the supply voltage (or less).
- The LED will continue to glow after you release the button.
- You can press the button for any length of time that is less than the timer's cycle time, and the LED always emits the same length of pulse.

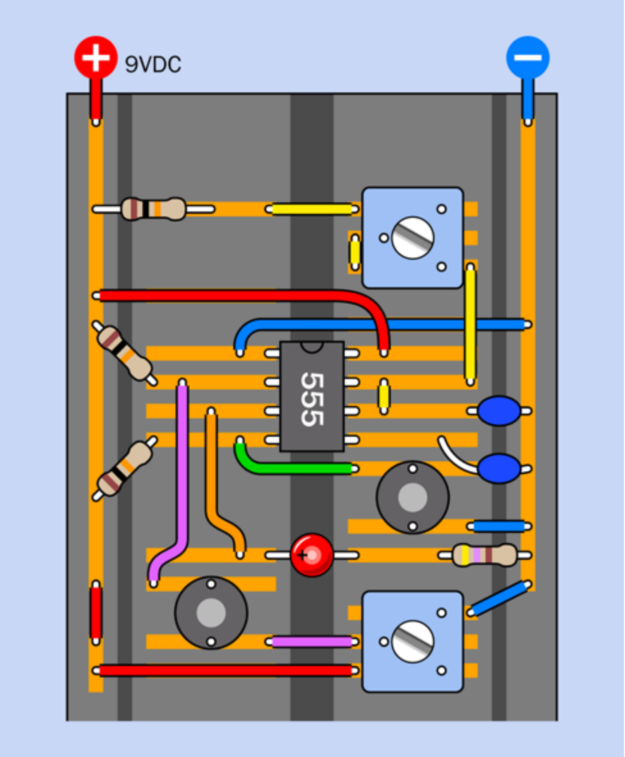

Figure 4-14 *Connections inside the breadboard wired for the timer test.*

Figure 4-15 shows how the timer is behaving, in graphical format. The 555 converts the imperfect world around it into a precise and dependable output. It doesn't switch on and off absolutely instantly, but is fast enough to *appear* instant.

Now try triggering the timer while you turn the 500K trimmer to different positions. You will find that this adjusts the length of the pulse.

- The resistance between pin 7 and the positive side of the power supply determines how long the pulse from the timer will last (in conjunction with the capacitor on pin 6).

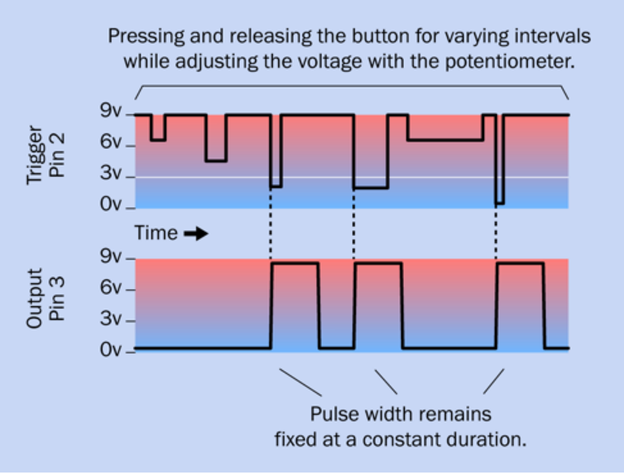

Figure 4-15 *How a 555 timer responds to different durations and voltages on its trigger pin.*

Here's another thing to try. Set the 500K trimmer to give a long pulse. Press button A, and then quickly press button B, which will stop the pulse before it completes. Hold down button B while you try to trigger the timer again with button A, and nothing happens.

- Pin 4 is the reset pin. When you ground it, you force the timer to interrupt whatever it is doing, and it is immobilized until you release pin 4 from its connection with negative ground.

Lastly, let go of button B, hold down button A, and continue to hold it down. This prolongs the pulse from the timer, until you let go of button A.

- Maintaining a low voltage on the trigger pin of the timer will *retrigger* it indefinitely.

Regarding the 10K resistors attached to pins 2 and 4—these are known as *pullup* resistors, because they hold the pins at a positive level. A more direct connection to negative ground will overwhelm the pullup resistor.

The concept of a pullup resistor is important when you are dealing with chips, because you must never allow an input pin to remain unconnected. An unconnected pin is said to be *floating*, and can cause trouble, as it may pick up stray electromagnetic fields, and we won't know what voltage is on it from one moment to the next.

Is there such a thing as a *pulldown* resistor? Absolutely. But the 555 timer needs pullup resistors, because pin 2 or pin 4 is kept in a normal state by positive voltage, and is activated by a low voltage.

- The 555 timer is triggered or reset by negative voltages on pins 2 and 4 respectively.

Timing the Pulse

If you study the schematic in Figure 4-11, you can see that positive current reaches pin 7 (the discharge pin) by passing through a 10K resistor and a 500K trimmer. (The 10K resistor is there because pin 7 should not be connected directly to the positive side of the power supply.)

You can also see that after passing through the 500K trimmer, the current can reach a 15µF capacitor. Hmmmm, a resistor followed by a capacitor—does this look like an RC network? Is the timer using the combination of the resistance and the 15µF capacitor to determine the length of the output pulse?

Yes, that's exactly what is going on. Inside the timer chip, some clever electronics are sensing the voltage on the 15µF capacitor, and the timer uses this to terminate its output pulse.

You can measure this yourself. Set the 500K trimmer to create a long pulse, and use your meter to measure the voltage on the left side of the 15µF capacitor. You should see it climbing up—until it reaches about 6 volts. The timer uses this as a signal to stop its output pulse, and the voltage quickly goes down again, because the timer is grounding it internally. This is why pin 7 is known as the discharge pin: the timer discharges the capacitor through it.

- When the voltage on the timing capacitor reaches two-thirds of the supply voltage, the timer ends its output pulse.

But why are the discharge pin and the threshold pin tied together? You'll learn about that in the next experiment, when the timer is rewired to deliver a series of pulses, instead of just one. At that time, the timer will be running in *astable* mode. Currently, you are using it in *monostable* mode.

- In monostable mode, the timer delivers only one pulse in response to a triggering event.
- In astable mode, the timer delivers a continuing series of pulses.

Finally, you may be wondering about the purpose of the 0.01µF capacitor attached to pin 5. This pin is the "control" pin, which means that if you apply a voltage to it, you can control the sensitivity of the timer. Because we are not using this function yet, it's good practice to put a capacitor on pin 5 to protect it from voltage fluctuations and prevent it from interfering with normal functioning.

Caution: Beware of Pin-Shuffling!

In all of the schematics in this book, chips are shown exactly as you would see them on a breadboard, with the pins in numerical sequence.

Other schematics that you may find, on websites or in books, do things differently. For convenience in drawing circuits, people often resequence the pin numbers. Also, there is no attempt to replicate a breadboard layout with a positive bus and a negative bus at each side. To give you an example, in Figure 4-16 the circuit is identical to that in Figure 4-11, but the pins have been shuffled to simplify connections and minimize wiring crossovers.

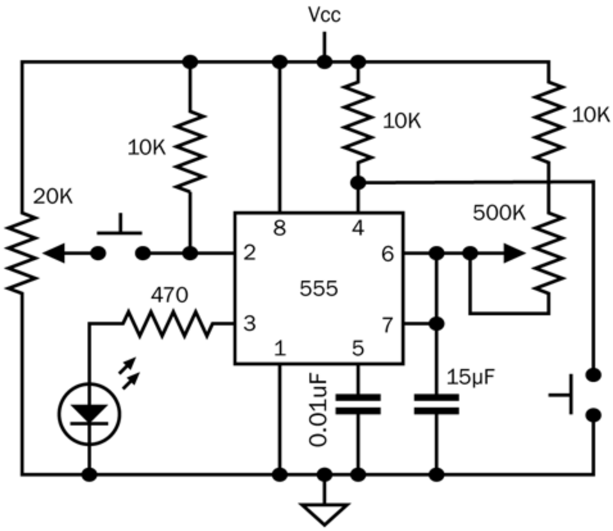

Figure 4-16 *This circuit is identical in function to the test circuit shown previously, but pins on the chip have been resequenced to simplify the schematic.*

Pin shuffling can create a circuit that is easier to understand in some ways (especially if positive power is at the top and ground at the bottom), but you have to convert the layout, often using pen and paper, before you can build it on a breadboard.

Fundamentals: Timer Duration

When you made your own RC network in Experiment 9, some annoying calculations were necessary to figure out how long a capacitor would take to reach any particular voltage. Using a 555 timer, everything is much easier. You just look up the duration of its output pulse in a table such as the one in Figure 4-17. The resistance between pin 7 and the positive side of the power supply is shown along the top of the table, the value of the timing capacitor is shown down the left side, and the numbers in the table tell you the approximate pulse duration in seconds.

	10K	22K	47K	100K	220K	470K	1M
1000µF	11	24	52	110	240	520	1100
470µF	5.2	11	24	52	110	240	520
220µF	2.4	5.2	11	24	52	110	240
100µF	1.1	2.4	5.2	11	24	52	110
47µF	0.52	1.1	2.4	5.2	11	24	52
22µF	0.24	0.53	1.1	2.4	5.3	11	24
10µF	0.11	0.24	0.52	1.1	2.4	5.2	11
4.7µF	0.052	0.11	0.24	0.52	1.1	2.4	5.2
2.2µF	0.024	0.052	0.11	0.24	0.53	1.1	2.4
1.0µF	0.011	0.024	0.052	0.11	0.24	0.52	1.1
0.47µF		0.011	0.024	0.052	0.11	0.24	0.52
0.22µF			0.011	0.024	0.052	0.11	0.24
0.1µF				0.011	0.024	0.052	0.11
0.047µF					0.011	0.024	0.052
0.022µF						0.011	0.024
0.01µF							0.011

Figure 4-17 *Pulse duration, in seconds, of a 555 timer running in monostable mode, for values of the timing resistor and timing capacitor. Times are rounded to two digits.*

- Resistor values below 1K should not be used.
- Resistor values below 10K are undesirable, as they increase power consumption.
- Capacitor values above 100µF may produce inaccurate results because leakage in the capacitor becomes comparable with its charging rate.

What if you want a time longer than 1,100 seconds or shorter then 0.01 seconds? Or what if you want a pulse duration that falls somewhere between the values in the table?

You can use this simple formula, where T is the pulse time in seconds, R is the resistance in *kilohms*, and C is the capacitance is *microfarads*.

```
T = R × C × 0.0011
```

Bear in mind that the result may not be exact, because resistor and capacitor values can be inaccurate, and because of other factors such as ambient temperature.

Theory: Inside the 555 in Monostable Mode

The plastic body of the 555 timer contains a wafer of silicon on which are etched dozens of transistor junctions in a pattern that is too complex to be explained here. However, I can summarize their function by dividing them into groups, as shown in Figure 4-18.

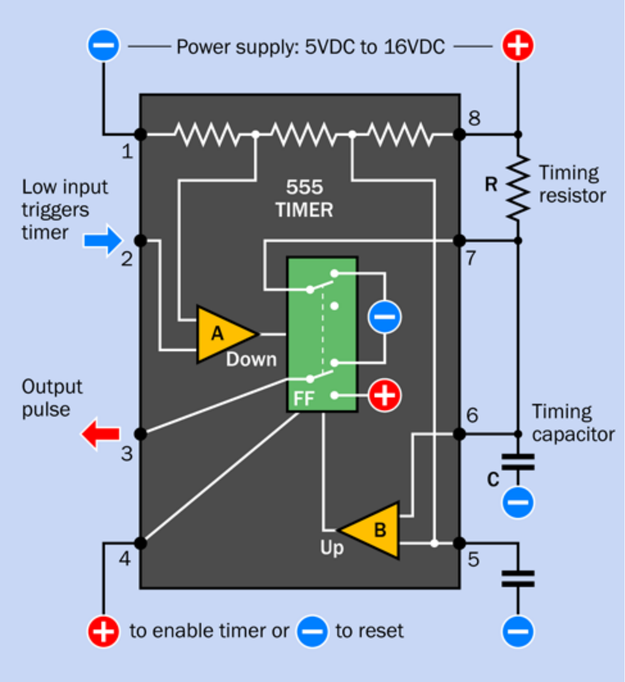

Figure 4-18 *Simplified representation of internal functions of a 555 timer, wired in monostable mode.*

The negative and positive symbols inside the chip are power sources that actually come from pins 1 and 8, respectively. I omitted the internal connections to those pins for the sake of clarity.

The two yellow triangles are *comparators*. Each comparator compares two inputs (at the base of the triangle) and delivers an output (from the apex of the triangle) depending on whether the inputs are similar or different. FF is a *flip-flop*, a logic component that can rest in one state or the other. I have depicted it as a double-throw switch, although in reality it is solid-state.

Initially when you power up the chip, the flip-flop is in its "up" position, which delivers low voltage through the output, pin 3. If the flip-flop receives a signal from comparator A, it flips to its "down" state, and flops there. When it receives a signal from comparator B, it flips back to its "up" state, and flops there. The "Up" and "Down" labels on the comparators will remind you how each one changes the switch when it is activated. Some people feel that the term "flip-flop" is derived from it having two states named "flip" and "flop." But I prefer to think of it flipping and flopping.

Notice the external wire that connects pin 7 with capacitor C. As long as the flip-flop is "up," it sinks the positive voltage coming through resistor R to pin 7, and prevents the capacitor from charging positively.

If the voltage on pin 2 drops to 1/3 of the supply, comparator A notices this, and flips the flip-flop to its "down" position. This sends a positive pulse from pin 3, the output pin, and also disconnects the negative power from pin 7. Now the capacitor can start charging through the resistor. While this is happening, the positive output from the timer continues.

As the voltage increases on the capacitor, comparator B monitors it through pin 6. When the capacitor accumulates 2/3 of the supply voltage, comparator B sends a pulse to the flip-flop, flipping it back into its original "up" state. This discharges the capacitor through pin 7. Also, the flip-flop ends the positive output through pin 3 and replaces it with a negative voltage. This way, the 555 returns to its original state.

I'll sum up this sequence of events:

- Initially, the flip-flop grounds the capacitor and grounds the output (pin 3).
- A drop in voltage on pin 2 to 1/3 of the supply voltage or less makes the output (pin 3) positive and allows capacitor C to start charging through resistor R.

- When the capacitor reaches 2/3 of the supply voltage, the chip discharges the capacitor, and the output at pin 3 goes low again.

Fundamentals: Pulse Suppression

When power is first applied to a timer that is configured in monostable mode, the timer tends to emit one pulse spontaneously before going dormant and waiting to be triggered again. This can be annoying in many circuits.

One way to prevent it is by putting a 1μF capacitor between the Reset pin and negative ground. The capacitor sinks current from the Reset pin when the power is first turned on, and holds the pin low for a fraction of a second—just long enough to stop the timer from emitting its waking-up pulse. After the capacitor is charged, it doesn't do anything more, and a 10K resistor holds the reset pin positive, so it won't interfere with the running of the timer.

I'll be using the pulse-suppression concept in subsequent experiments.

Fundamentals: Why the 555 Is Useful

In its monostable mode, the 555 will emit a single pulse of fixed (but programmable) length. Can you imagine some applications? Think in terms of the pulse from the 555 controlling some other component. A motion sensor on an outdoor light, perhaps. When an infrared detector "sees" something moving, the light comes on, but only for a specific period—which can be controlled by a 555.

Another application could be a toaster. When someone lowers a slice of bread, a switch will close, triggering the toasting cycle. To change the length of the cycle, you could use a potentiometer and attach it to the external lever that determines how dark you want your toast. At the end of the toasting cycle, the output from the 555 would pass through a power transistor, to activate a solenoid (which is like a relay, except that it has no switch contacts) to release the toast.

Intermittent windshield wipers could be controlled by a 555 timer—and on earlier models of cars, they were. The repeat rate of keys on a rather basic computer keyboard could be controlled by a 555 timer—and on the Apple II, it was.

What about the intrusion alarm in Experiment 15? One of the features on my wish list was that it should wait long enough for you to shut it off, before it starts to make a noise. The output from a 555 timer can take care of that.

The experiment that you just performed seemed trivial, but it implies a huge range of possibilities.

Fundamentals: Bistable Mode

There's another way to use the timer, known as bistable mode. This entails disabling its fundamental features. Why would you want to do this? I will explain.

Figure 4-19 shows a circuit that you can build in just a few minutes. Give it a try. The two resistors on the left are pullup resistors, 10K each. The resistor at the bottom is 470 ohms, to protect the LED. Add two tactile switches, and the timer chip itself, and you're done.

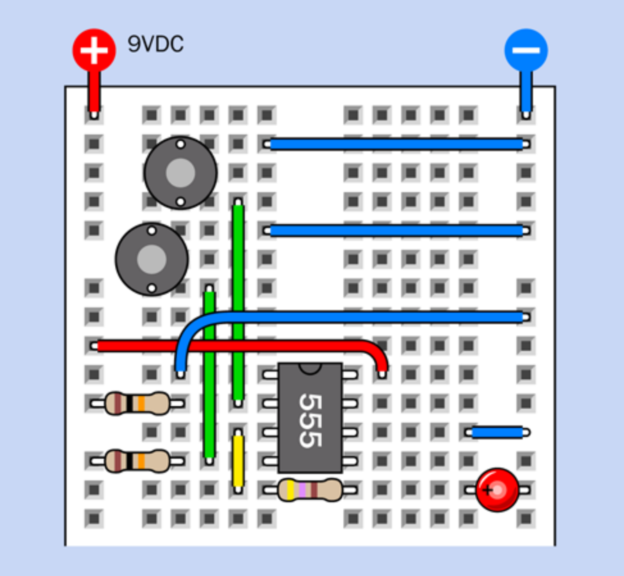

Figure 4-19 *Breadboard circuit to make the 555 timer function as a flip-flop.*

Once you have it on your breadboard, press and release the top button, and the LED lights up. For how long? For as long as you supply power to the circuit. The output from the timer continues indefinitely.

Now press and release the bottom button, and the LED goes off. For how long? For as long as you like. It won't come back on until you press the top button again.

I mentioned that there is a flip-flop inside the timer. This circuit makes the timer into one big flip-flop. It flips into its “on” state when you ground pin 2—and it flops there. It flips into its “off” state when you ground pin 4—and it flops there. Flip-flops are very important in digital circuits, as I will explain a bit later, but right now, how does this work, and why would you need it?

Take a look at the schematic in Figure 4-20. You may notice that there is no resistor or capacitor on the right-hand side. The RC network is missing. So—this timer circuit has no timing components! Normally, when you trigger the timer, its output pulse ends when the timing capacitor on pin 6 accumulates 2/3 of supply voltage. But pin 6 is grounded, so it can never reach the 2/3 value. Consequently, when you trigger the timer, the output pulse will never end.

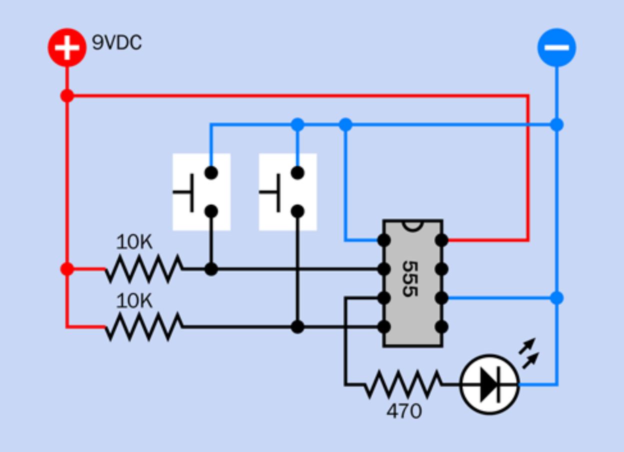

Figure 4-20 *A test circuit for the 555 timer in bistable mode.*

Of course, you can stop the output by applying a low voltage to the reset pin. But once the output stops, it will remain stopped, so long as you don’t trigger the timer again.

This configuration is called bistable because it is stable when the output is high, and it is stable when the output is low. A simple flip-flop like this can also be called a *latch*.

- A negative pulse to pin 2 turns the output positive, and latches it.
- A negative pulse to pin 4 turns the output negative, and latches it.

You do have to keep pins 2 and 4 high when you are not triggering them. That’s what the pullup resistors in the schematic are for.

It’s OK to leave pin 5 of the timer unconnected, because we’re pushing it into extreme states where any random signals from those pins will be ignored.

As for why you would need to use a timer in this way—you’ll be surprised how useful it can be. I’m going to use it in three experiments in the remainder of this book. The 555 was not really designed to function in bistable mode, but it can be convenient.

Background: How the Timer Was Born

Back in 1970, when barely a half-dozen corporate seedlings had taken root in the fertile ground of Silicon Valley, a company named Signetics bought an idea from an engineer named Hans Camenzind (pictured in Figure 4-21). It wasn’t a huge breakthrough concept—just 23 transistors and a bunch of resistors that would function as a programmable timer. The circuit would be versatile, stable, and simple, but these virtues paled in comparison with its primary selling point. Using the emerging technology of integrated circuits, Signetics could reproduce the whole thing on a silicon chip.

Figure 4-21 *Hans Camenzind, inventor, designer, and developer of the 555 timer chip for Signetics.*

This entailed some trial and error. Camenzind worked alone, building the whole thing initially on a large scale, using off-the-shelf transistors, resistors, and diodes on a breadboard. It worked, so then he started substituting slightly different values for the various components to see whether the circuit would tolerate variations during production, and other factors such as changes in temperature when the chip was in use. He made at least 10 different versions of the circuit. This took months.

Next came the crafts work. Camenzind sat at a drafting table and used a specially mounted X-Acto knife to scribe his circuit into a large sheet of plastic. Signetics then reduced this image photographically by a ratio of about 300:1. They etched it into tiny wafers, and embedded each of them in a half-inch rectangle of black plastic with the product number printed on top. Thus, the 555 timer was born.

It turned out to be the most successful chip in history, both in the number of units sold (tens of billions and counting) and the longevity of its design (not significantly changed in almost 40 years). The 555 has been used in everything from toys to spacecraft. It can make lights flash, activate alarm systems, put spaces between beeps, and create the beeps themselves.

Today, chips are designed by large teams and tested by simulating their behavior using computer software. Thus, the chips inside a computer enable the design of new chips. The heyday of solo designers such as Hans Camenzind is long gone, but his genius lives inside every 555 timer that emerges from a fabrication facility. If you'd like to know more about chip history, visit the Transistor Museum.

A personal note: in 2010, when I was writing *Make: Electronics*, I looked up Hans Camenzind online and found that he maintained his own website, which included a phone number. On impulse, I called him. This was a strange moment, to be talking to the man whose chip design I had used for more than 30 years. He was friendly (although he didn't waste words), and readily agreed to review the text of my book. Even more kindly, after he read it, he gave it a strong endorsement.

Subsequently I bought his own short history of electronics, *Much Ado About Almost Nothing*, which is still available online, and which I highly recommend. I felt honored to have had the opportunity to talk to one of the pioneers in integrated circuit design. I was sad when I heard of his death in 2012.

Fundamentals: 555 Timer Specifications

- The 555 can run from a reasonably stable voltage source ranging from 5VDC to 16VDC. The absolute maximum is 18VDC. Many datasheet specifications are measured at 15VDC. The voltage does not have to be controlled by a voltage regulator.
- Most manufacturers recommend a range from 1K to 1M for the resistor attached to pin 7, but values below 10K draw a more significant amount of current. It's a better idea to reduce the value of the capacitor than to reduce the value of the resistor.
- The capacitor value can go as high as you like, if you want to time really long intervals, but the accuracy of the timer will diminish because leakage in the capacitor becomes comparable with its charge rate.
- The timer imposes a voltage drop that is greater than the drop created by a transistor or a diode. The difference between supply voltage and output voltage will be 1V or more.
- The output is rated to source or sink 200mA, but an output current above 100mA will pull down the voltage and can affect the timing accuracy.

Caution: Not All Timers Are Equal

Everything I have said so far applies to the old original "TTL" version of the 555 timer. TTL is an acronym for *transistor–transistor logic*, which preceded modern CMOS chips that use much less power. The TTL version of the timer is also referred to as the *bipolar version*, as it contains bipolar transistors.

The advantage of the original 555 is that it is cheap and robust. You can't damage it easily, and its output is powerful enough to connect directly with a relay coil or a small loudspeaker. However, the 555 is not efficient, and tends to generate voltage spikes that sometimes interfere with the operation of other chips.

To address these disadvantages, a newer version of the 555 timer was developed using CMOS transistors, which draw less power. This chip also doesn't create voltage spikes. But its output is more limited. How much more? That depends on the particular manufacturer.

Unfortunately, there is a lack of standardization among CMOS versions of the 555 timer. Some claim to deliver 100mA while others are limited to 10mA.

Confusingly, the CMOS versions have a variety of part numbers. The 7555 is clearly identified as a CMOS chip, but others merely precede the 555 number with a different group of letters, and it's up to you to notice, and to understand what they mean.

In this book, to avoid confusion and keep things simple, I am only using the TTL version of the 555 timer, also known as the bipolar version. If you are buying your own, see "Other Components" on page 319 and go down to the subhead "Components for Chapter Four" where you will find timer buying advice.

Experiment 17: Set Your Tone

Now that you're familiar with the 555 timer in monostable mode and bistable mode, I want you to get acquainted with it in *astable* mode—so called because the output fluctuates constantly between high and low, and does not remain stable in either of those states.

This resembles the output of the transistor oscillator that you built in Experiment 11, except that it is much more versatile and easier to control, and instead of requiring two transistors, four resistors, and two capacitors to create the oscillation, you only need one chip, two resistors, and one capacitor.

What You Will Need

- Breadboard, hookup wire, wire cutters, wire strippers, multimeter
- 9-volt power supply (battery or AC adapter)
- 555 timer chips (4)
- Miniature loudspeaker (1)
- Resistors: 47 ohms (1), 470 ohms (4), 1K (2), 10K (12), 100K (1)
- Capacitors: 0.01µF (8), 0.022µF (1), 0.1µF (1), 1µF (3), 3.3µF (1), 10µF (4), 100µF (2)
- 1N4148 diode (1)
- Trimmer potentiometer, 100K (1)
- Tactile switch (1)
- Generic LEDs (4)

Astable Test

A generic astable circuit is shown in Figure 4-22. I've put a loudspeaker on the output, because the timer will be running at an audio frequency. The loudspeaker is driven through a resistor, to limit the current, and a coupling capacitor, which passes audio frequencies while blocking DC. You'll see the values for these components in the next schematic. Right now I just want you to see the general layout.

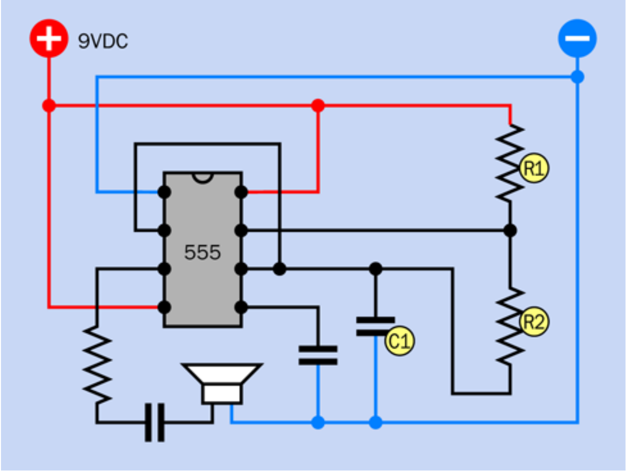

Figure 4-22 *A basic, generic circuit that runs a 555 timer in astable mode.*

The components labeled R1, R2, and C1 control the speed of the timer. These labels are always used in manufacturers' datasheets and other sources, so I'm following the same convention.

C1 does the same thing as the timing capacitor in the monostable circuit in Figure 4-11. The need for two resistors, instead of one, will be explained below.

See if you can get a sense of how this circuit may work, using the knowledge that you gained in Experiment 16. The first thing you may notice is that there's no input. Pin 2 (the trigger pin) is connected back to pin 6 (the threshold pin). Can you see how that's going to work? C1 will accumulate a charge, as it did when the timer was in monostable mode, until it reaches 2/3 of the power supply, at which point it will discharge through R2 into pin

7, and its voltage will drop. Its connection with pin 2 will mean that the trigger pin senses the drop in voltage on C1. And what does the trigger pin do when the voltage on it drops suddenly? It triggers the timer. So, in this configuration, the timer will retrigger itself.

How fast will that happen? I think you should build a test version of the circuit to find out. In Figure 4-23 I have suggested values for the components and have redrawn the schematic to include a trimmer potentiometer, so that you'll see (or, rather, hear) the effect of varying this resistance. The trimmer, plus the 10K resistor that precedes it, add up to R2. The timing capacitor, C1, is 0.022µF, and R1 is 10K.

Figure 4-24 shows the breadboard layout, while Figure 4-25 shows the component values.

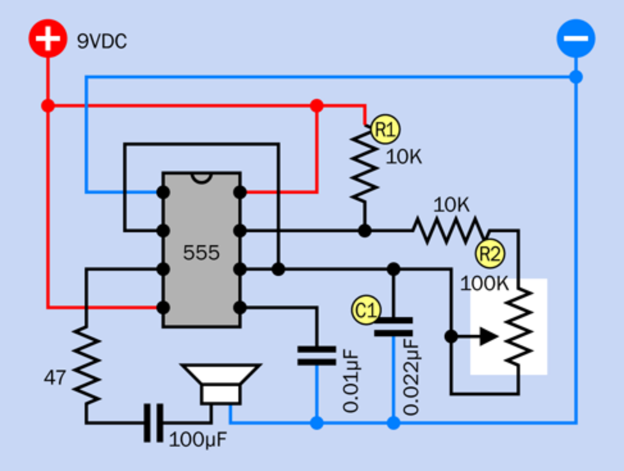

Figure 4-23 *This test circuit allows you to adjust the performance of the timer in astable mode.*

Now what happens when you apply power? Immediately, you should hear noise through the loudspeaker. If you don't hear anything, you almost certainly made a wiring error.

Notice that you don't have to activate the chip with a pushbutton anymore. The 555 timer is triggering itself, as predicted.

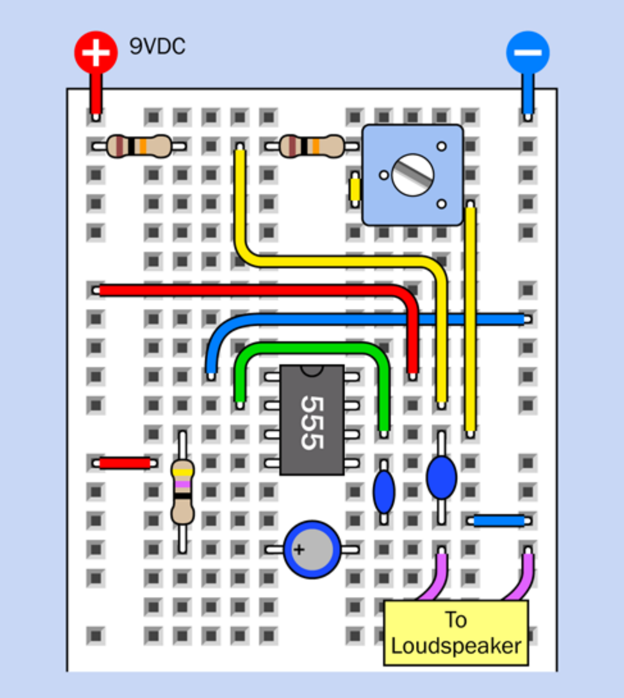

Figure 4-24 *Breadboard layout for the astable timer test.*

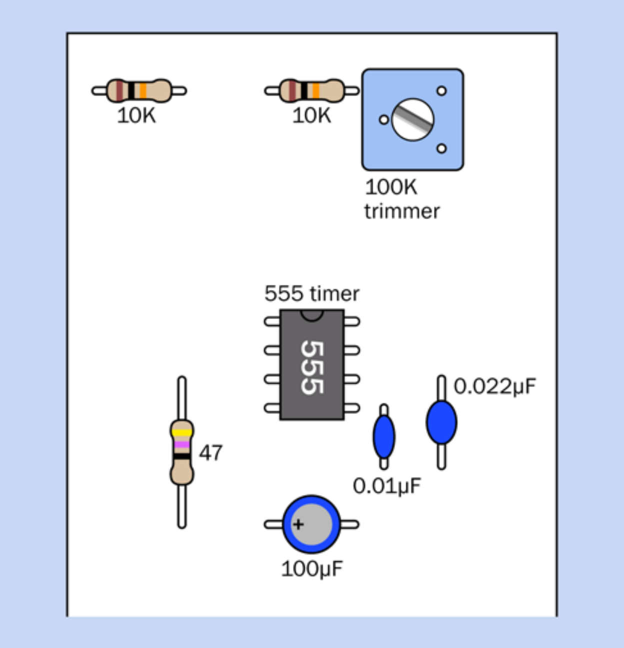

Figure 4-25 *Component values for the astable timer test.*

Rotate the screw on the trimmer potentiometer, and the pitch of the sound varies. The trimmer adjusts how quickly C1 charges and discharges, and this determines the length of one "on" cycle relative to the next "off" cycle in the audio signal. With these component values, the stream varies between approximately 300 and 1,200 pulses each second. These pulses are sent by the timer to the loudspeaker. They move its cone up and down, creating pressure waves in the air, and your ear responds to those waves, hearing them as sound.

Theory: Output Frequency

The *frequency* of a sound is its number of full cycles per second, including the high-pressure pulse and the low-pressure pulse that follows.

The term *hertz* is a unit of frequency, meaning the same thing as "cycles per second." It was introduced in Europe, named after yet another electrical pioneer, Heinrich Hertz. The abbreviation for hertz is Hz, so the output of your 555 timer in its test circuit will range approximately between 300Hz and 1,200Hz.

As in most standard units, a k can be inserted to mean "kilo," so 1,200Hz is normally written as 1.2kHz.

How do the values for the timing capacitor and the resistors determine the frequency of the timer? If R1 and R2 are measured in *kilohms*, and C1 is measured in *microfarads*, the frequency f, in hertz, is given by:

```
f = 1,440 / ( ( ( 2 × R2 ) + R1 ) × C1 )
```

Doing this calculation is a hassle, so I have provided you with a lookup table in Figure 4-26. In this table, *I am assuming that the value for the resistor labeled R1 in the schematic is fixed at 10K.* The values across the top of the table are for R2. The values on the left side of the table are for the timing capacitor, C1.

You may remember than the abbreviation pF means "picofarad," which is one-millionth of a microfarad. Nanofarads are halfway between microfarads and picofards, but the term is not used so frequently in the US, so I didn't use it in the table.

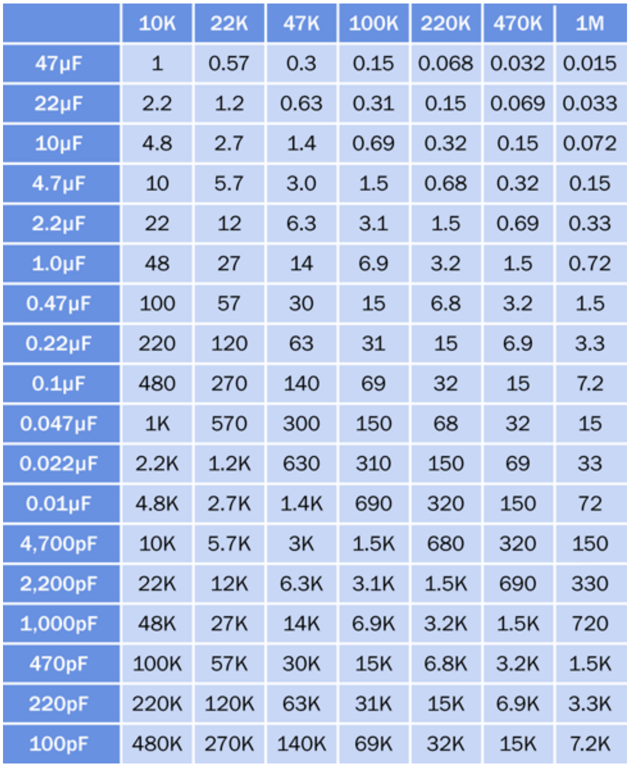

	10K	22K	47K	100K	220K	470K	1M
47µF	1	0.57	0.3	0.15	0.068	0.032	0.015
22µF	2.2	1.2	0.63	0.31	0.15	0.069	0.033
10µF	4.8	2.7	1.4	0.69	0.32	0.15	0.072
4.7µF	10	5.7	3.0	1.5	0.68	0.32	0.15
2.2µF	22	12	6.3	3.1	1.5	0.69	0.33
1.0µF	48	27	14	6.9	3.2	1.5	0.72
0.47µF	100	57	30	15	6.8	3.2	1.5
0.22µF	220	120	63	31	15	6.9	3.3
0.1µF	480	270	140	69	32	15	7.2
0.047µF	1K	570	300	150	68	32	15
0.022µF	2.2K	1.2K	630	310	150	69	33
0.01µF	4.8K	2.7K	1.4K	690	320	150	72
4,700pF	10K	5.7K	3K	1.5K	680	320	150
2,200pF	22K	12K	6.3K	3.1K	1.5K	690	330
1,000pF	48K	27K	14K	6.9K	3.2K	1.5K	720
470pF	100K	57K	30K	15K	6.8K	3.2K	1.5K
220pF	220K	120K	63K	31K	15K	6.9K	3.3K
100pF	480K	270K	140K	69K	32K	15K	7.2K

Figure 4-26 *For a 555 timer running in astable mode, values across the top refer to R2 in a standard circuit, so long as R1 has a fixed value of 10K. Numbers in the table show the frequency of the timer in Hz (cycles per second).*

Theory: Inside the 555 in Astable Mode

To gain a better understanding of what happens when the timer runs in astable mode, take a look at Figure 4-27. The internal configuration is exactly the same as in monostable mode, but the external connections are different.

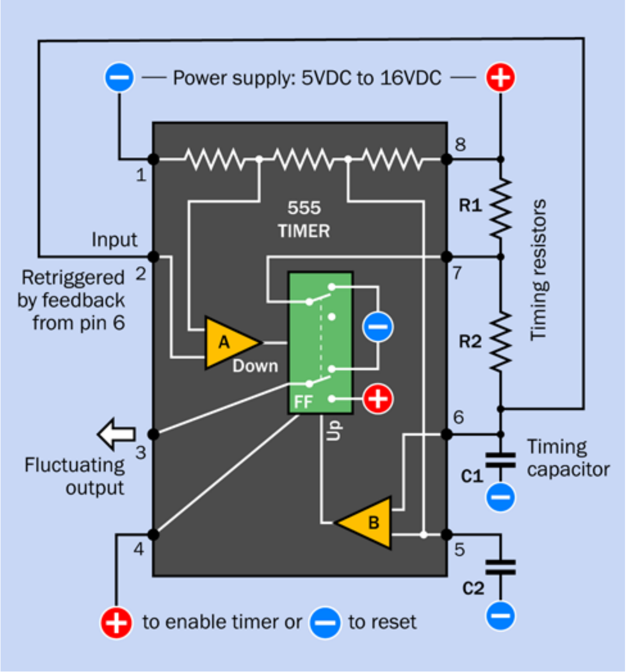

Figure 4-27 *Internal view of the 555 timer with external connections running it in astable mode.*

Initially, the flip-flop grounds C1, the timing capacitor, as before. But now the low voltage on the capacitor is connected from pin 6 to pin 2 through an external wire. The low voltage tells the chip to trigger itself. The flip-flop obediently flips to its "on" position and sends a positive pulse to the loudspeaker, while removing the negative voltage from pin 6.

Now C1 starts charging, as it did when the timer was in monostable mode, except that it is being charged through R1 + R2 in series. Because C1 has a low value, it charges quickly. When it reaches two-thirds of full voltage, comparator B takes action as before, discharging the capacitor and ending the output pulse from pin 3.

The capacitor discharges through R2 to pin 7, the discharge pin. While the capacitor is discharging, its voltage diminishes. The voltage is still linked to pin 2. When it drops to one-third of full power or less, comparator A kicks in and sends another pulse to the flip-flop, starting the process all over again.

Fundamentals: Unequal On-Off Cycles

When the timer is running in astable mode, C1 charges through R1 and R2 in series. But when C1 discharges, it dumps its voltage into the chip through R2 only. Because the capacitor charges through two resistors but discharges through only one of them, it charges more slowly than it discharges. While it is charging, the output on pin 3 is high; while it is discharging, the output on pin 3 is low. Consequently the "on" cycle is always longer than the "off" cycle. Figure 4-28 shows this as a simple graph.

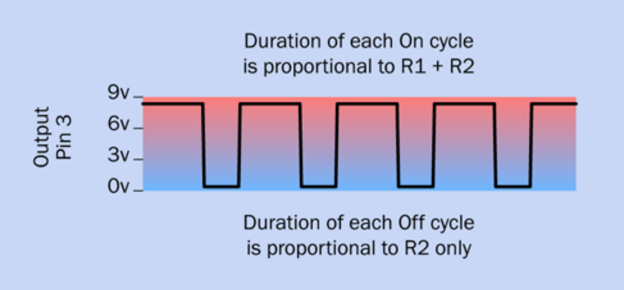

Figure 4-28 *High pulses are always longer than gaps between them in the output from a 555 timer chip, when the chip is wired in the standard way for astable operation.*

If you want the on and off cycles to be equal, or if you want to adjust the on and off cycles independently (for example, because you want to send a very brief pulse to another chip, followed by a longer gap until the next pulse), all you need to do is add a diode, as shown in Figure 4-29. (Because a diode deducts some voltage, this circuit works best with a supply voltage greater than 5VDC.)

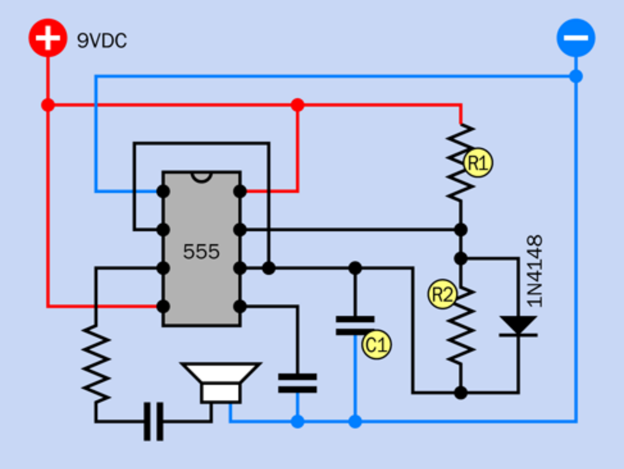

Figure 4-29 *Adding a diode to bypass R2 allows the high and low output cycles of a timer to be adjusted independently.*

Now when C1 charges, the electricity flows through R1 as before but takes a shortcut around R2, through the diode. When C1 discharges, the diode blocks the flow of electricity in that direction, and so the discharge goes back through R2.

R1 now controls the charge time on its own, while R2 controls the discharge time. The formula for calculating the frequency is now approximately:

```
Frequency = 1,440 / ( ( R1 + R2 ) × C1 )
```

where R1 and R2 are in *kilohms* and C1 is in *microfarads*. (I use the word "approximately" because the diode adds a small amount of effective resistance to the circuit, which is not factored into the formula.)

If you set R1 = R2, you should get almost equal on/off cycles.

Astable Modifications

Instead of using a potentiometer to vary the value of R2, the frequency of the timer can be changed to a limited extent by using pin 5, the control pin. This is shown in Figure 4-30.

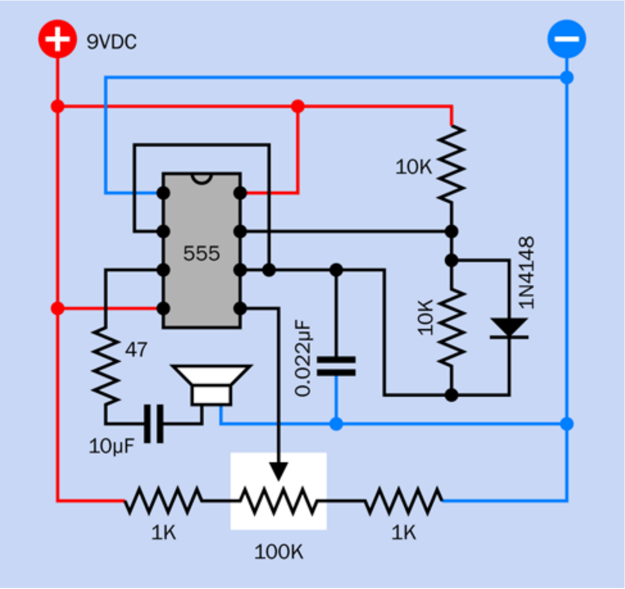

Figure 4-30 *A circuit that demonstrates the function of pin 5, the control pin, on the 555 timer.*

Disconnect the capacitor that was attached to that pin and substitute the series of resistors shown. They ensure that pin 5 always has at least 1K between it and the positive side or the negative side of the power supply. Connecting it directly to the power supply won't damage the timer, but will prevent it from generating audible tones. As you turn the potentiometer, the frequency will vary. This happens because you are changing the reference voltage on comparator B inside the chip.

Chaining Chips

Timer chips can be chained together in four possible ways. Note that these configurations are workable regardless of whether each timer is running in monostable or astable mode, except where noted.

- If you are using 9V to power a 555 timer, the output from that timer can be sufficient to power another 555 timer.
- The output from one timer can trigger the input of another timer. This only works if the second timer is running in monostable mode. In astable mode, it would be self-triggering.
- The output from one timer can unlock the reset pin of another timer.
- The output from one timer can be connected through a suitable resistor to the control pin of another timer.

These options are illustrated in Figure 4-31, Figure 4-32, Figure 4-33, and Figure 4-34.

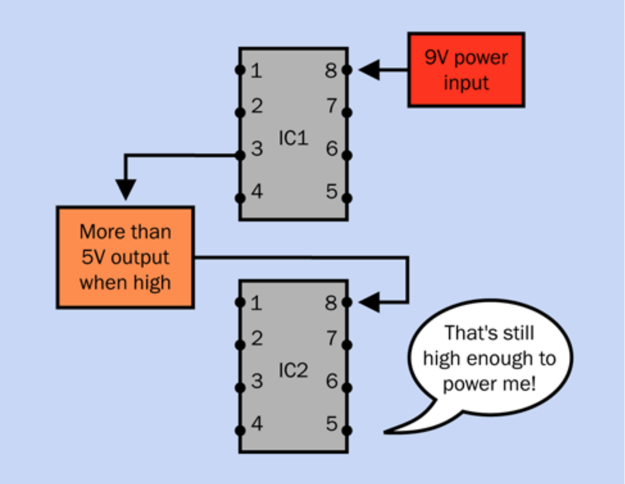
1
2
3
4
8
7
6
5
IC1
9V power input
More than 5V output when high
1
2
3
4
8
7
6
5
IC2
That's still high enough to power me!

Figure 4-31 *One timer powering another.*

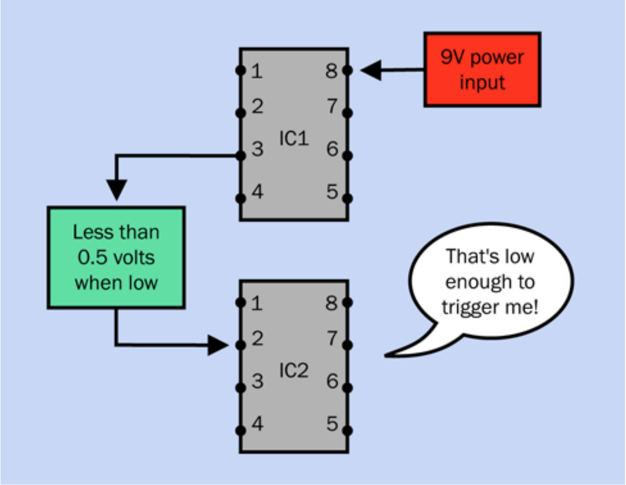
IC1
1
2
3
4
8
7
6
5
9V power input
Less than 0.5 volts when low
IC2
1
2
3
4
8
7
6
5
That's low enough to trigger me!

Figure 4-32 *One timer triggering another.*

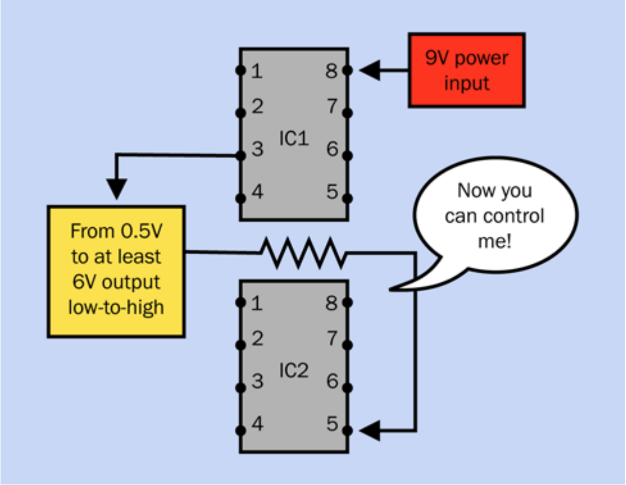
IC1
1
2
3
4
8
7
6
5
9V power input
From 0.5V to at least 6V output low-to-high
Now you can control me!
IC2
1
2
3
4
8
7
6
5

Figure 4-34 *One timer controlling the frequency of another.*

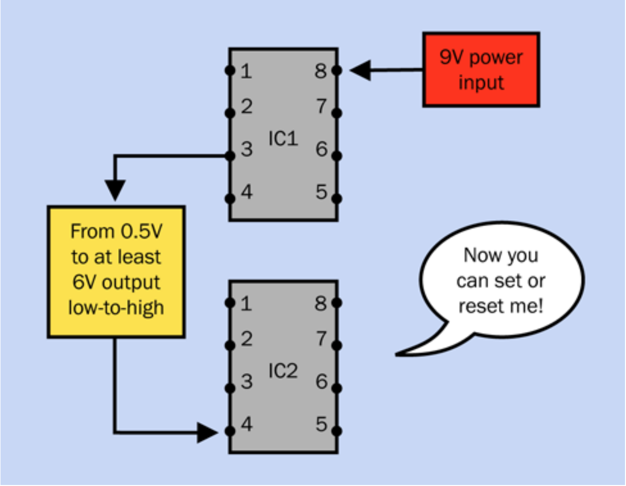
IC1
1
2
3
4
8
7
6
5
9V power input
From 0.5V to at least 6V output low-to-high
Now you can set or reset me!
IC2
1
2
3
4
8
7
6
5

Figure 4-33 *One timer setting or resetting another.*

Why would you want to chain timers together? Well, you might want to have a couple of them running in monostable mode, so that the end of a high pulse from the first one triggers the start of a high pulse in the second one, and vice versa. In fact, you could chain together as many timers as you like, with the last one feeding back and triggering the first one, and they could flash a series of LEDs in sequence, like Christmas lights.

Figure 4-35 shows four timers chained in this way. They are linked through coupling capacitors, because we only want a brief pulse from one to trigger the next. Without the capacitors, the end of the pulse from the first timer in the chain would trigger the second timer, but the output from the first timer would remain in its low state, which would continue triggering the second timer indefinitely.

In addition, each timer must have a 10K pullup resistor on its trigger pin to hold it normally high.

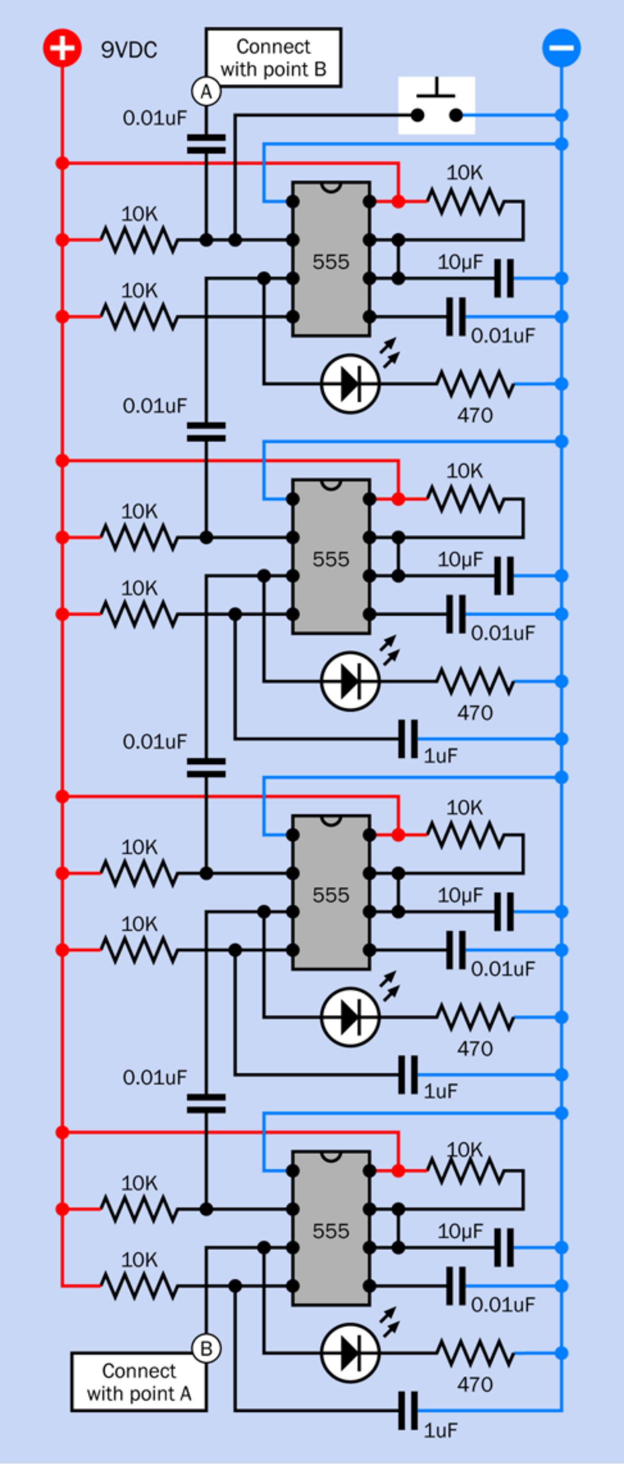

Figure 4-35 *Four timers wired to trigger each other in a recirculating sequence.*

An interesting question comes to mind when monostable timers are chained together. That is: how do they start? I mentioned in Experiment 16 that a 555 timer in monostable mode will usually emit one spontaneous pulse when it is first powered up. When multiple timers are chained together, they will all try to do this at about the same time, and because of small manufacturing differences, the outcome will be unpredictable. Sometimes they will settle down to a nice orderly sequence, but other times they will end up flashing in pairs.

The way to deal with this is to use the concept of pulse suppression that I mentioned in Experiment 16 (see "Fundamentals: Pulse Suppression" on page 152).

A 1μF capacitor between the reset pin and negative ground will hold the reset pin in a low state just long enough to suppress the timer's initial pulse. A 10K pull-up resistor sharing the reset pin will then keep it stable while the timer is running.

In my experience, this works well, although timers from a different manufacturer could conceivably behave differently, as the behavior of the reset pin is not well documented. If you have trouble with pulse suppression, try substituting a larger or smaller capacitor.

In a chain of timers, the only remaining problem is that pulse suppression works too well. You apply power, and —nothing happens, because all the timer outputs have been suppressed.

The way around this is to omit pulse suppression from just one timer. It will almost certainly emit an initial pulse when it receives power, and that will trigger the rest in sequence. This setup is illustrated in Figure 4-35.

But—wait a minute. What is this "almost certainly" phrase? Electronic circuits should always work, all the time, not "almost" all the time.

I agree. But I can't control the tendency of 555 timers to do unpredictable things when they are powered up. Therefore I also added a button, at the top of the circuit, that can be used to start the cascade if it doesn't start itself.

There is an alternative, which is for the first timer in the chain to run in astable mode. It sends a series of pulses that ripple down through the others configured in monostable mode, and the last one does not feed back to the first. In electronics terms, we say that the first one is the master and the others are the slaves.

I like this arrangement because it is totally predictable. The trouble is, you have to adjust the speed of the master timer so that it emits its next pulse exactly when the last slave in the chain has just finished emitting its pulse. Otherwise, the first timer will emit another pulse before the last pulse has ended, or there will be a gap between the last pulse and the next first pulse.

Whether this is important will depend on the application. Flashing lights won't be a problem, but if you increase the speed to drive a stepper motor, getting the timing right will be difficult.

Sounding Like a Siren

The fourth option that I listed for chaining chips, in Figure 4-34, is of special interest, because it can create a siren sound very like the noise made by a typical burglar alarm. In fact, this could be used for the audio output in the alarm project that I left unfinished in Experiment 15.

The circuit is shown in Figure 4-36. Timer 1 is wired in a basic astable circuit, which you can recognize as being similar to the circuit in Figure 4-22. The component values are larger, so the timer oscillates more slowly, around 1Hz. You can compare this circuit with the one that I suggested in Figure 2-120. The principle is similar.

Timer 2 is also wired in a basic astable circuit, running around 1kHz. The idea is that the slow fluctuations in voltage from Timer 1 are applied to the control pin of Timer 2, forcing it to modulate its sound up and down in that annoying way which we associate with alarm systems.

I encourage you to build this circuit, because you may want to use it in the final version of the intrusion alarm coming right up in Experiment 18. The breadboard layout for the siren circuit is in Figure 4-37, and the component values for the layout are shown in Figure 4-38.

Once you get it running, you may find it interesting to remove and then replace the 100µF capacitor connected between pin 6 and ground. The capacitor makes the frequency slide higher and lower instead of switching sharply up and down. I'm using it in the same way that I used a capacitor to make an LED fade smoothly in and out in Experiment 11.

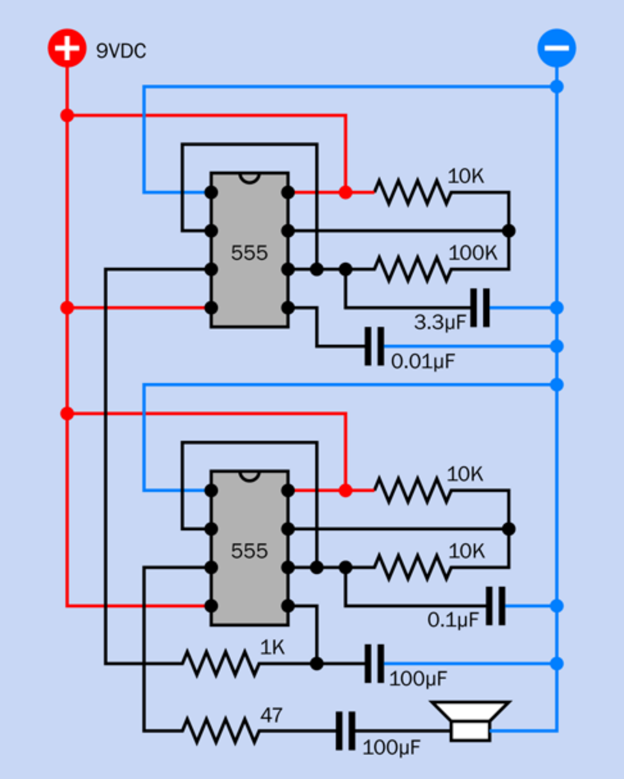

Figure 4-36 *When one timer runs relatively slowly, modulating another through its control pin (pin 5), the result is a wavering sound like an alarm siren.*

You can modify the sound in other ways. Here are some suggestions:

- Vary the 0.1µF timing capacitor to raise or lower the pitch of the basic sound.
- Double the value of the 100µF capacitor on pin 6, or divide it by two.
- Substitute a 10K potentiometer for the 1K resistor.
- Change the value of the 3.3µF capacitor.

Part of the pleasure of building things is to customize them, to make them your own. Once your siren sound satisfies you, make a note of the component values for future reference.

Incidentally, you can reduce the chip count (the number of chips) by using a 556 timer instead of two 555 timers. The 556 contains a pair of 555 timers in one package. Because you still have to make the same number of external connections (other than the power supply), I haven't bothered to use this variant.

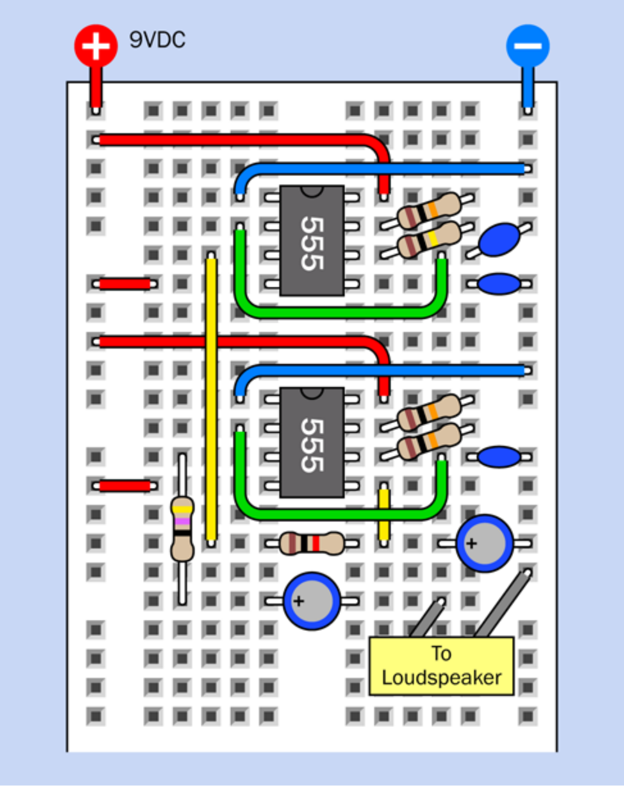

Figure 4-37 *The siren circuit, breadboarded.*

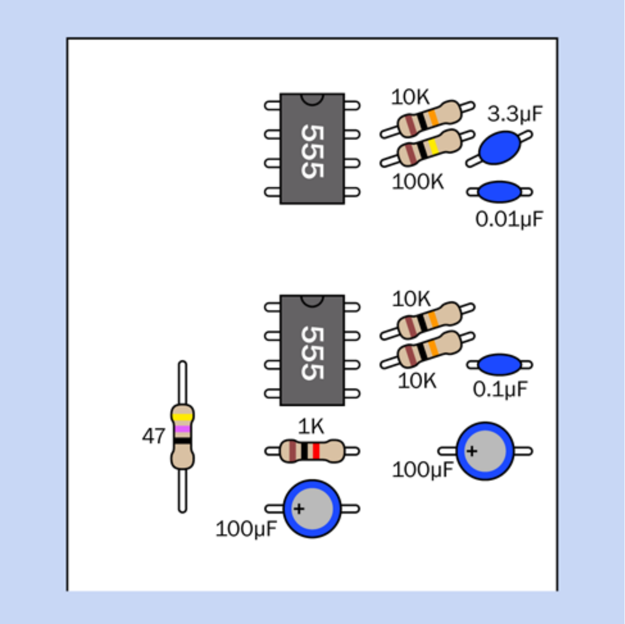

Figure 4-38 *Component values for the siren circuit.*

Experiment 18: Intrusion Alarm (Almost) Completed

Now that you've seen what the 555 timer can do, you can fulfill the remaining requirements on the intrusion-alarm wish list.

What You Will Need

- Breadboard, hookup wire, wire cutters, wire strippers, multimeter
- 9-volt power supply (battery or AC adapter).
- 555 timers (2)
- DPDT 9VDC relay (1)
- Transistors, 2N2222 (2)
- LEDs: Red, green, yellow (1 of each)
- Slide switch, SPDT, for breadboard (2)
- Tactile switch (1)
- Capacitors: 0.01µF (1), 10µF (2), 68µF (2)
- Resistors: 470 ohms (4), 10K (4), 100K (1), 1M (2)

- Diode, 1N4001 (1)

Optional (for audio output):

- Components shown in Figure 4-36

Optional (for permanent fabrication of this project):

- 15-watt soldering iron
- Thin solder
- Perforated board plated with copper in a breadboard layout
- SPDT or DPDT toggle switch (1)
- Pushbutton, SPST (1)
- Project box, at least 6″ × 3″ × 2″ (1)
- Power jack and matching power socket (1 each)
- Magnetic sensor switches, in pairs, quantity sufficient for your home
- Alarm network wiring, sufficient for your home

Three Steps to a Functional Device

This is a larger and more complicated circuit than anything you have tackled so far, but it's relatively easy to build, because you can assemble it in three parts that you can test individually. Eventually you'll end up with a breadboard looking like Figure 4-45, with component values in Figure 4-46. The equivalent schematic is shown in Figure 4-47. But we'll begin with just a small timer circuit.

Step 1

Take a careful look at Figure 4-39. Notice there are no timing components on the righthand side of the 555 timer. You may conclude that this is a version of the bistable circuit that I described in Experiment 16 (see Figure 4-20). When the timer is triggered, its output will go on indefinitely—which seems appropriate for an alarm system.

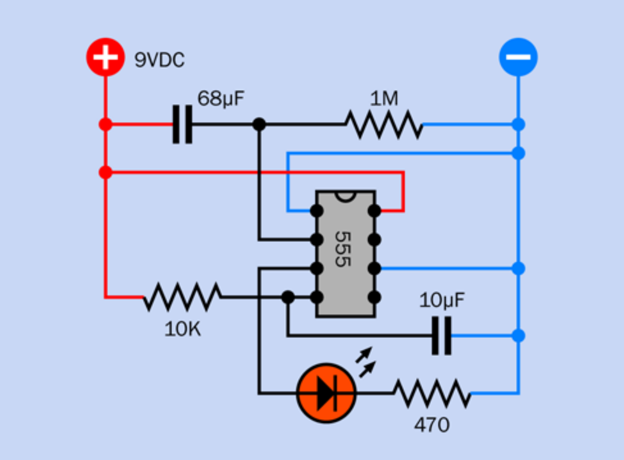

Figure 4-39 *Schematic equivalent of the bottom section of the breadboard layout.*

But there's more to it than that. This circuit also gives you a one-minute grace period in which to disable the alarm before it starts to sound, when you enter the area. (You may remember that this was #9 on the wish list of features that I compiled in Experiment 15.)

To see how it works, you can assemble the components shown in Figure 4-40. Their values are shown in Figure 4-46, and you'll see their placement at the bottom of the board in Figure 4-45.

The placement is important, because you have to leave room for the additional sections of the circuit that you'll be adding. One of those sections will be switching on the power to this one.

To make sure that you have everything in the right place, the 1M resistor on the right is on the 29th row in the board, counting down from the top. Also note that power is being applied near the components, not at the top of the board, and the positive bus is not being used, yet.

Don't power up the circuit just yet. Set your meter to measure at least 10VDC and attach it at the points shown in Figure 4-40, with the negative probe on the negative bus and the positive probe at the left end of the 1M resistor.

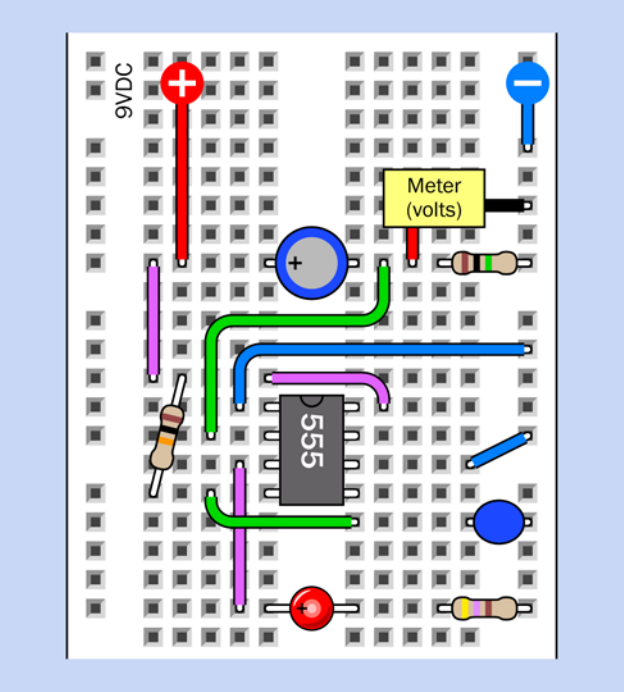

Figure 4-40 *Placement of components at the bottom of the breadboard for testing.*

Now apply power to the circuit, and you should see your meter slowly counting down from 9V. When it gets to one-third of the supply voltage, this triggers the 555 timer, and the red LED comes on. The LED is included for testing purposes; in the final version, you would substitute a noisemaking circuit.

The large 68µF capacitor delays the response of the timer. When you first apply power to the circuit, the capacitor passes the initial pulse to the point between it and the 1M resistor. The green wire runs from this point to the trigger pin of the timer. So, the pin starts off high, and you'll remember (I hope) that the timer does nothing until the trigger pin goes low.

Voltage on the righthand side of the capacitor leaks out very slowly through the 1M resistor. Eventually, it goes low enough to trigger the timer.

As for the rest of the circuit, in Experiments 16 and 17 I explained how to do "pulse suppression" to stop the timer from emitting a pulse when it is first powered up. This is why I have a 10µF capacitor and a 10K resistor applied to pin 4, the reset pin. I'm using a 10µF capacitor this time, instead of a 1µF capacitor, because this circuit reacts a bit more slowly than the one in Experiment 17.

- You can use a 555 timer with these components any time you want to delay the timer output in response to a trigger pulse.
- Use a larger or smaller value for the 68µF capacitor if you want to make the delay longer or shorter.

So far, so good. This part of the circuit will introduce a delay when it is powered up, and after the delay, it will activate the alarm for an indefinite period.

Step 2

Figure 4-41 and Figure 4-42 show the next step in building the circuit. The components that you placed previously are still there, but they have been grayed out to focus your attention on the new additions.

Don't forget to install S2, the slide switch at the bottom, and the 470-ohm resistor beside it, and the two long yellow wires. This slide switch is included for testing purposes. It represents the alarm sensors that you would use in an actual application.

The relay has the same function as in Experiment 15. In fact if you trace the connections in the circuit, you'll find that it works the same way as the one in Figure 3-88, with a couple of minor revisions. The only differences are that a 470-ohm resistor has been substituted for a 1K resistor, and I have added switch S1 at the top, with a green LED. Why? I'll get to that in a moment.

Place all the components carefully. Don't overlook the three red wires on the left and the three blue wires on the right. Make sure the pins of the relay are aligned with the wires that serve them.

Double-check that S1 is in the down position, and S2 is in the up position. For testing purposes, remove the 68µF capacitor so that the red LED will respond right away instead of waiting for a minute.

Attach power—and if you made all the connections correctly, nothing should happen. Switch S2 represents the alarm sensors, and in its up position, it simulates them being closed. Slide it down to simulate a sensor being opened, and the test LED at the bottom of the circuit should light up immediately. Slide the switch up, and the LED stays on. The circuit has locked the alarm on, regardless of a sensor being reset.

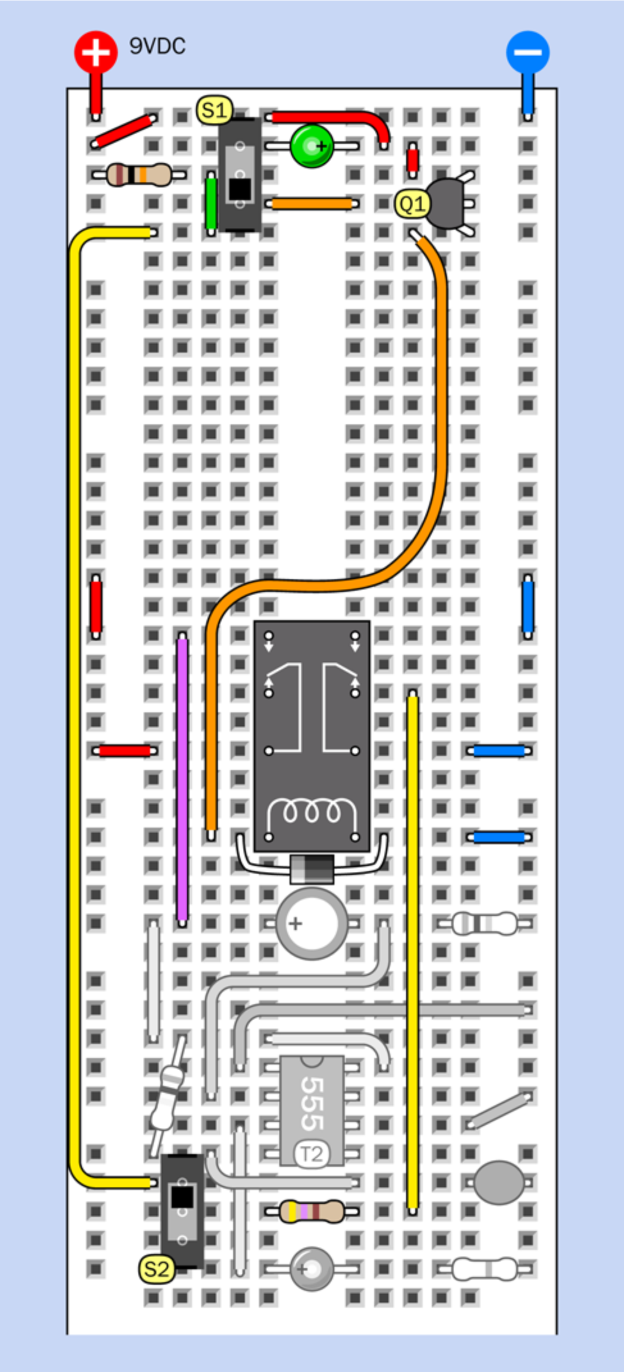

Figure 4-41 *The second step in building this circuit incorporates the relay configuration that was used in Experiment 15.*

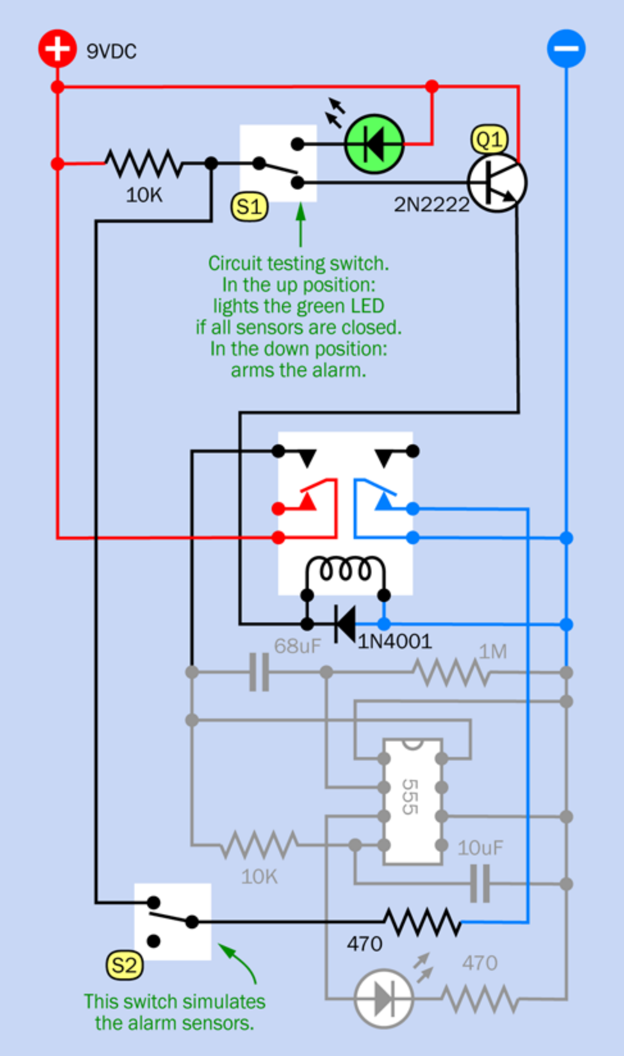

Circuit testing switch.
In the up position:
lights the green LED
if all sensors are closed.
In the down position:
arms the alarm.

This switch simulates
the alarm sensors.

Figure 4-42 *Schematic equivalent of the breadboarded second stage in the circuit.*

Disconnect the power, keep S2 in its up position (simulating the sensors being closed), and reapply the power. Now move S1, at the top, to its up position, and the green LED comes on. This is a circuit-testing feature. It checks that all the sensors are closed. When you use the alarm, you will want to conduct this test before you leave the area. This satisfies the first part of item #7 in the wish list from Experiment 15.

Keep S1 in its up position while you move S2 to its down position, to simulate an open sensor. The green LED goes out. Move S2 to its up position, and the green LED comes back on. So, the testing procedure works.

Here's how you would actually use this circuit. You leave S1 in the up (test) position. When you are ready to leave the area, you apply power to the circuit. If the green LED doesn't come on, there is a door or window open somewhere. Find the source of the problem, and correct it. When the green LED is on, you know that all the sensors are closed. Now you can arm the alarm. Move S1 down. The green LED goes out, and the alarm is now armed. When you come home, the 555 timer gives you one minute to disable the alarm, to prevent it from going off (so long as you have replaced the 68μF capacitor in the circuit). You can disable the alarm by moving S1 to its up (test) position.

Now, how and why does the circuit work?

The 10K resistor at top-left connects with the base of transistor Q1 through switch S1, when the switch is in its down position. Meanwhile, the righthand pole contact inside the relay is connected to negative ground. This connection runs through the yellow wire on the right, through a 470-ohm resistor, through switch S2 (which simulates the sensors), and back up through the other long yellow wire. It holds the base of the resistor at a low voltage (through the orange wire). So long as the base is low, the transistor doesn't conduct.

If a sensor opens, the base of the transistor isn't held low anymore, and the 10K resistor pulls it up so that the transistor starts to conduct. It triggers the relay, through the long curving orange wire. The relay will supply power to the bistable timer, which will eventually activate the alarm. At the same time, the relay breaks the negative-ground connection on the right, so now the transistor is going to continue conducting even if the sensor is closed again.

This is exactly the same circuit concept that I ended up with in Figure 3-88. The significant difference is the green LED. When you move S1 to its "test" position, it cuts off positive power to the transistor (so that the transistor cannot start the alarm). If all the sensors are closed, the LED connects through them, and the 470-ohm resistor, to negative ground, and it lights up to tell you that the system is ready.

Step 3

What else could we need in this project? Well, imagine you are using the alarm system. You want to set it before leaving the area. At this point, you suddenly realize that if you set it, and then open a door to leave, you will trigger the alarm.

The bistable timer with its 68μF capacitor added a feature to suppress the alarm for a minute, to give you time to turn it off when you arrive. Now we need another timer to suppress the alarm for a minute when you leave.

This is a little more difficult to arrange. The key is to have the extra timer pull down the voltage on transistor Q1, so that it can't trigger the relay.

The problem is, a timer output goes high, not low, during its "on" cycle. I'll have to add another transistor to convert the high output so that it pulls down the voltage on the base of transistor Q1.

Figure 4-43 and Figure 4-44 show the components that will make it happen. Once again I have grayed out the components that you placed previously.

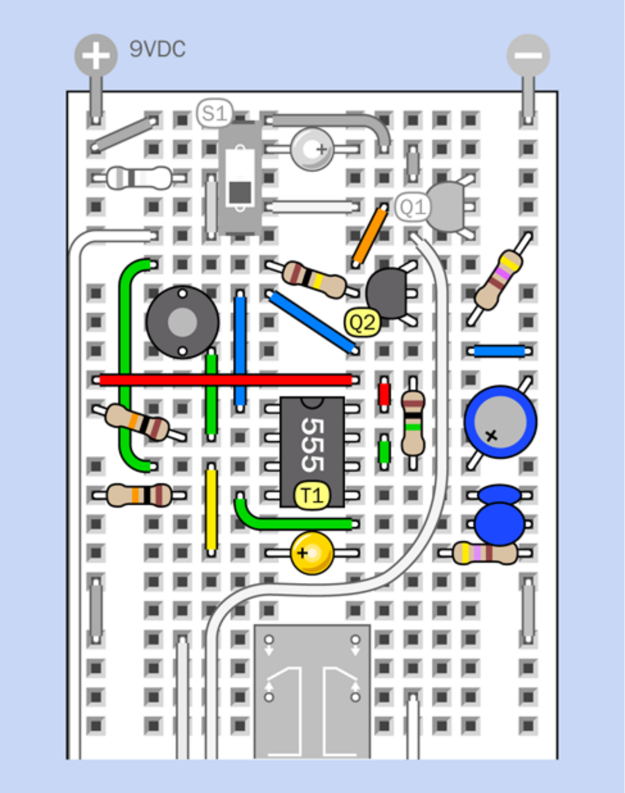

Figure 4-43 *Third and final step in building the alarm circuit.*

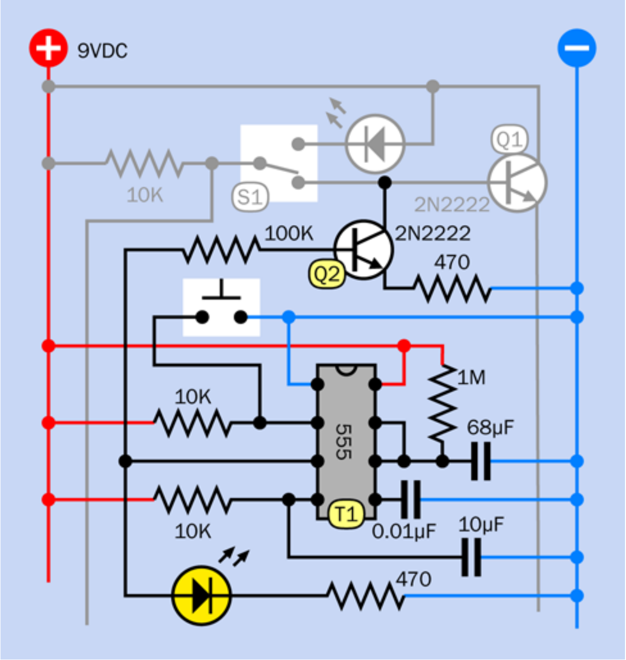

Figure 4-44 *The schematic version of the third step in building the circuit.*

The new 555 timer, labeled T1, has a pulse-suppression circuit on pin 4, its reset pin, like the other timer, so that it won't emit a pulse when you power up the circuit. You press the button to start T1. The button works by grounding the trigger pin of the timer.

While the output of the timer is high, the current flows out of pin 3 (the output pin) and lights up the yellow LED. This tells you that the alarm system is counting down to being armed. So long as you see the LED, the alarm will ignore any activity that opens a sensor switch.

Pin 3 also connects through a green wire, on the left, which is like an elongated letter C. This curls around to a 100K resistor, which is attached to the base of Q2, a second transistor. The output from the timer, through the 100K resistor, is enough to make Q2 conduct. Its emitter is grounded through a 470-ohm resistor, while its collector is attached to the base of Q1. So long as Q2 is conducting, it grounds the base of Q1, and prevents Q1 from triggering the relay and starting the alarm.

In this way, timer T1 stops the alarm from going off. When the one-minute grace period ends, T1 stops conducting, doesn't pull down the voltage on the first transistor anymore, and the alarm can go off—provided you remembered to move the switch at the top out of "test" mode, of course.

You would now use the circuit like this.

1. First put switch S1 into its "test" position and close all the doors and windows until the green LED comes on.
2. Move S1 into its lower position, so that the alarm is ready.
3. Press the button, and leave, closing the door behind you, while the yellow LED is on.

Does your version do what it is supposed to do? It should, so long as you wired it carefully. Timer T1 should light the yellow LED under any circumstances, making it easy to test. You can also touch your meter probe on the base of Q1, to verify whether the voltage is relatively high or low. So long as the voltage is relatively low, the alarm will not be triggered. When the voltage goes high, the alarm is triggered.

Don't forget to replace the 68µF capacitor in the circuit, just below the relay, to reactivate the delay timer, when your alarm is ready for prime time.

The complete breadboarded circuit is shown in Figure 4-45, the component values are in Figure 4-46, and the schematic is in Figure 4-47.

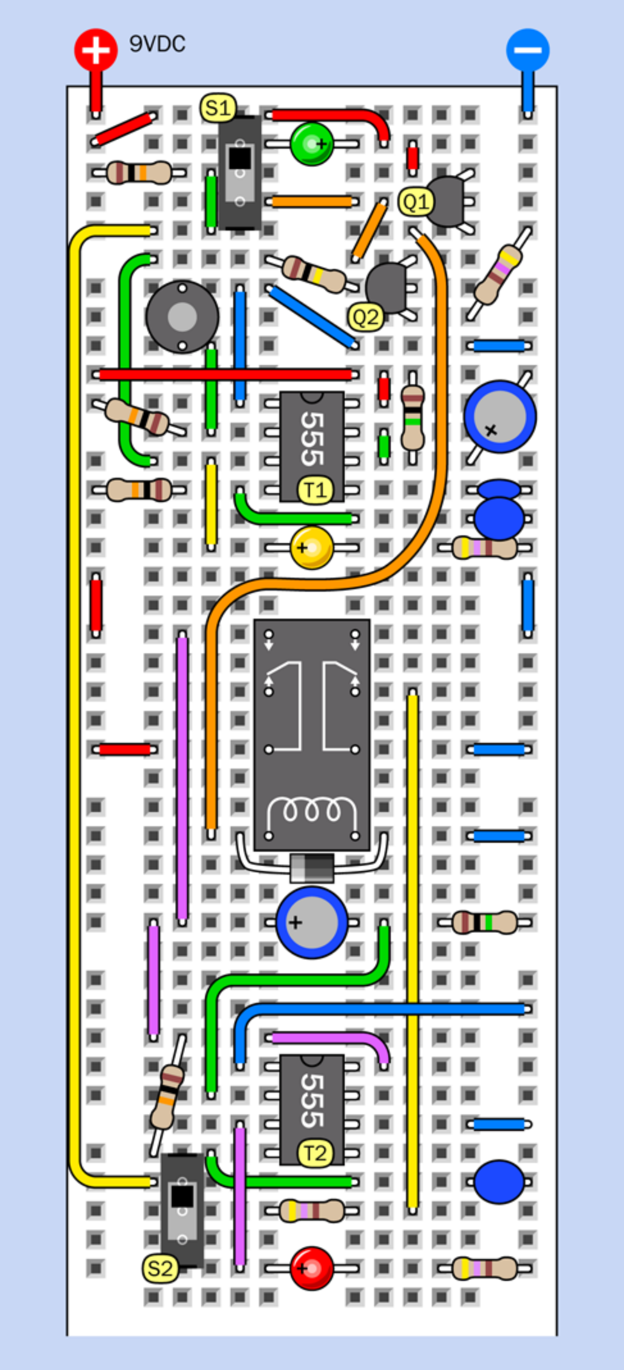
9VDC
S1
Q1
Q2
555
T1
555
T2
S2

Figure 4-45 *Breadboard layout for the complete alarm circuit.*

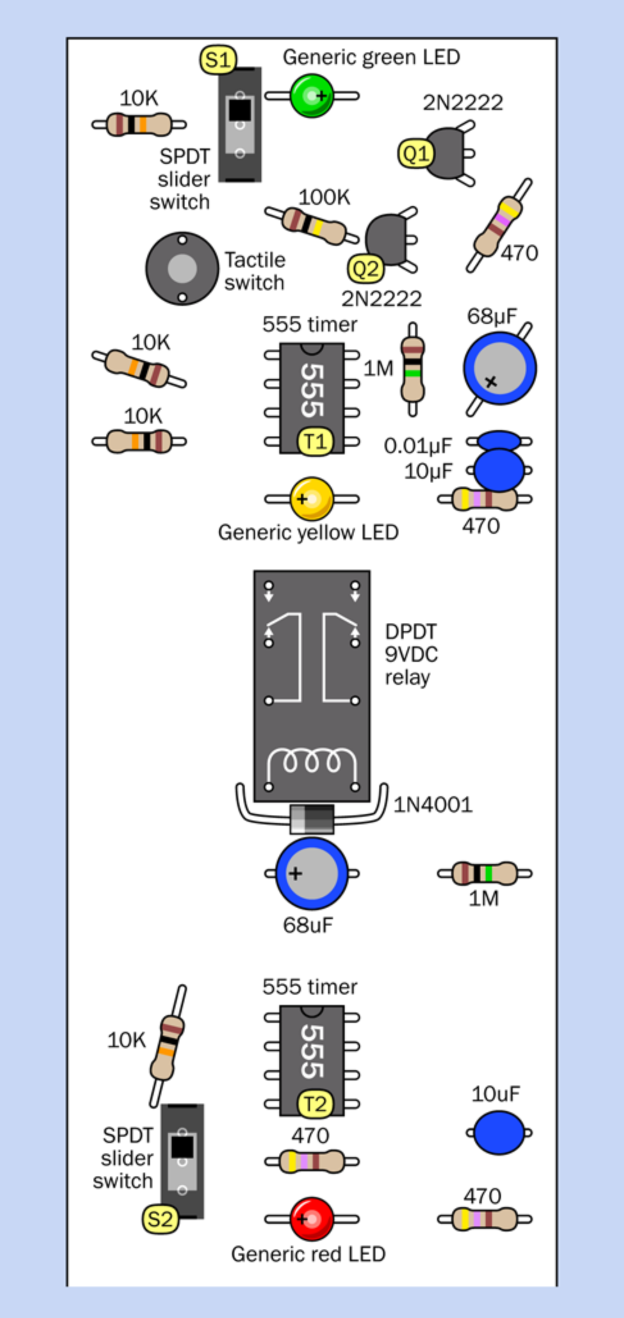
S1
Generic green LED
10K
2N2222
SPDT
slider
switch
Q1
100K
470
Tactile
switch
Q2
2N2222
10K
555 timer
68µF
555
T1
1M
10K
0.01µF
10µF
Generic yellow LED
470
DPDT
9VDC
relay
1N4001
68uF
1M
555 timer
10K
555
T2
10uF
SPDT
slider
switch
470
S2
470
Generic red LED

Figure 4-46 *Component values in the breadboard layout.*

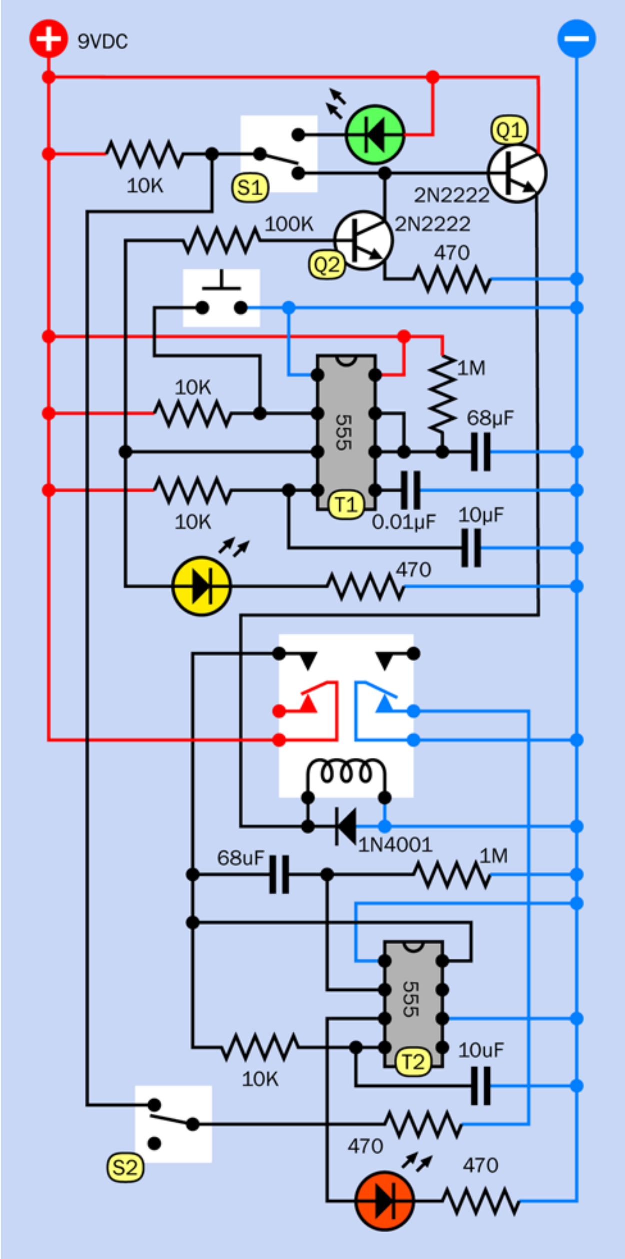

Figure 4-47 *Schematic equivalent of the breadboarded alarm circuit.*

What About the Noise?

When you want the alarm to make noise, you'll need to substitute an audio circuit or device for the red LED that you have been using for testing purposes.

The easy way to do this is to use an off-the-shelf item. Hundreds of sirens are available cheaply, ready to make an annoying sound if you simply apply power. Many of them require 12VDC, but they deliver almost as much noise at 9VDC. Just remember that timer T2 cannot deliver much more than 150mA.

If you prefer to have a sound that's all your own, you can use the circuit that I showed in Figure 4-36. Just use the output from your relay to power this circuit, and you have your own sound.

What About On and Off?

You've been testing the circuit by applying and removing power. You can add an on-off switch, but a numeric code to switch off the alarm would be more desirable.

Right now, I can't show you how to implement that, because it requires logic chips that I have not dealt with yet. But Experiment 21 will show you how it can be done.

Finalizing

Meanwhile, since the alarm circuit does actually work in its current form, I want to talk about finalizing it. By this I mean soldering it to a board, mounting the board in a box, and making everything look nice. My main concern in this book is electronics, but still, finalizing a project is an important part of the making experience, so I'll give you some suggestions.

Soldering the circuit can be easier than the procedure in Experiment 14, where I explained point-to-point wiring. You can mount the components on the type of perforated board that has copper traces on the back, in a configuration identical to the connectors inside a breadboard. Simply move each component to its comparable position, and solder it to the copper conductor underneath. No wire-to-wire soldering is necessary.

Guidance for finding and buying this kind of board is in "Supplies" on page 316 near the end of the book.

Now, how to proceed:

Carefully note the position of a component on your breadboard, then move it to the same relative position on the perforated board, poking its wires through the little holes.

Turn the perforated board upside down, make sure that it's stable, and examine the hole where the wire is poking through, as shown in Figure 4-48, which shows the *underside* of the board (the component is on the other side). A copper trace surrounds this hole and links it with others. Your task is to melt solder so that it sticks to the copper and also to the wire, forming a solid, reliable connection between the two of them.

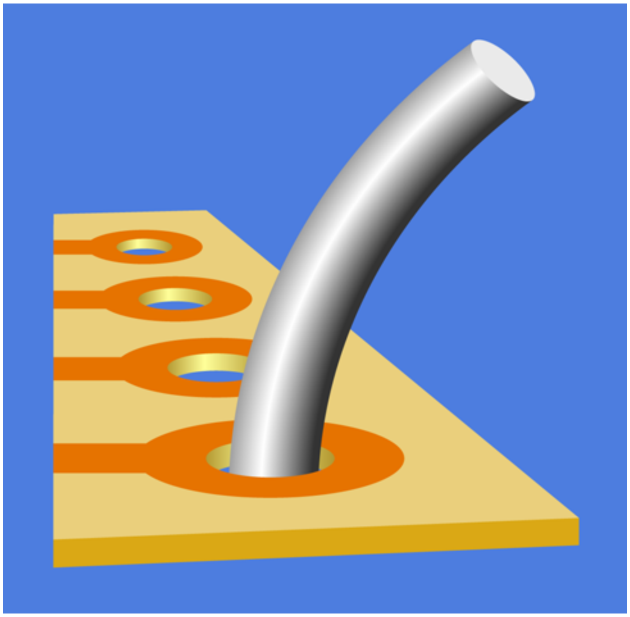

Figure 4-48 *The underside of perforated board, with a wire sticking through.*

Clamp the perforated board or rest it on a surface where it won't skid around easily. Take your low-wattage soldering iron in one hand and some solder in your other hand. Hold the tip of the iron against the wire and the copper, and feed some thin solder to their intersection. After two to four seconds, the solder should start flowing.

Allow enough solder to form a rounded bump sealing the wire and the copper, as shown in Figure 4-49. Wait for the solder to harden thoroughly, and then grab the wire with pointed-nosed pliers and wiggle it to make sure you have a strong connection. If all is well, snip the protruding wire with your cutters. See Figure 4-50.

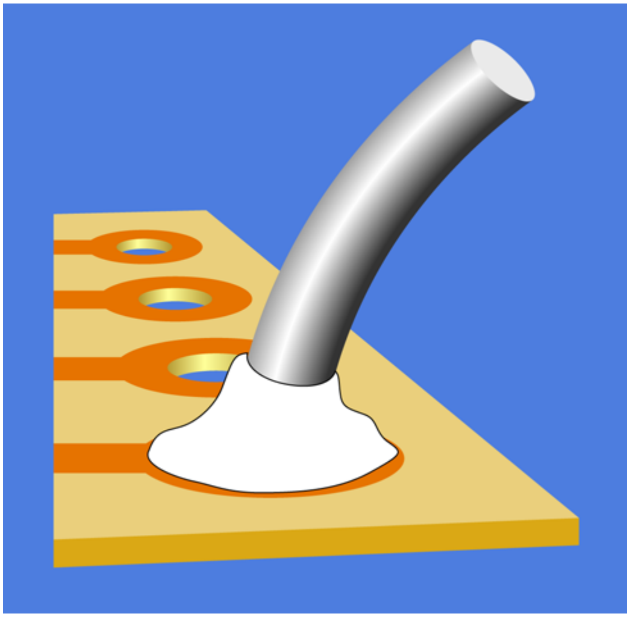

Figure 4-49 *Ideally, your solder joint should look something like this.*

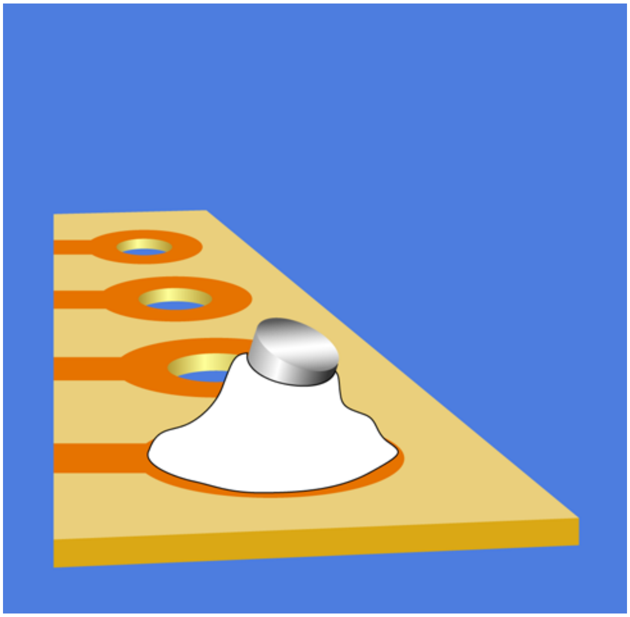

Figure 4-50 *After the solder has cooled and hardened, you snip off the projecting wire.*

Because solder joints are difficult to photograph, I'm using drawings to show the wire before and after making a reasonably good joint. The solder is shown in pure white, outlined with a black line.

The process of actually soldering components into perforated board is illustrated in Figure 4-51 and Figure 4-52.

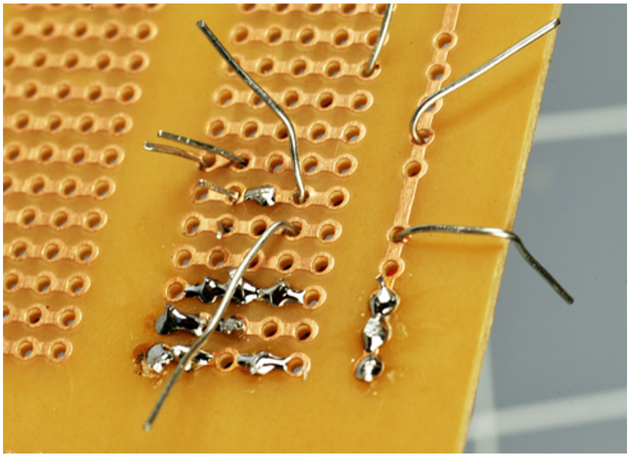

Figure 4-51 *This photograph was taken during the process of transferring components from breadboard to perforated board. Two or three components at a time are inserted from the other side of the board, and their leads are bent over to prevent them from falling out.*

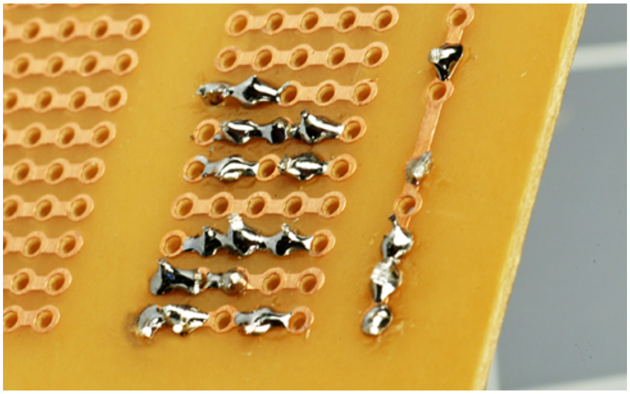

Figure 4-52 *After soldering, the leads are snipped short and the joints are inspected under a magnifying glass. Another two or three components can now be inserted, and the process can be repeated.*

Most Common Perfboarding Errors

1. Too much solder. Before you know it, solder creeps across the board, touches the next copper trace, and sticks to it, as depicted in Figure 4-53. When this happens, you can either try to suck it up with a desoldering kit, or carve it away with a knife. Personally I prefer to use a knife, because if you suck it up with a rubber bulb or solder wick, some of it will tend to remain.

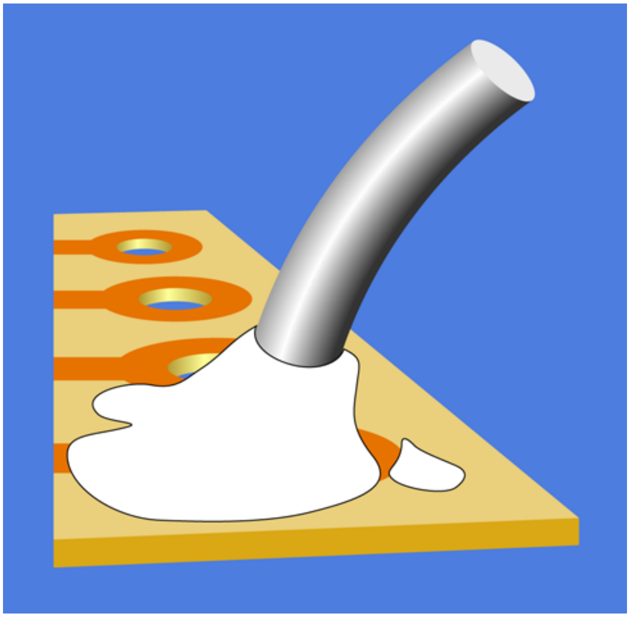

Figure 4-53 *If you use too much solder, it will tend to end up in places where you don't want it.*

Even a microscopic trace of solder is enough to create a short circuit. Check the wiring with a magnifying glass while turning the perforated board so that the light strikes it from different angles.

2. Not enough solder. If the joint is thin, the wire can break free from the solder as it cools. Even a microscopic crack is sufficient to stop the circuit from working. In extreme cases, the solder sticks to the wire, and sticks to the copper trace around the wire, yet doesn't make a solid bridge connecting the two, leaving the wire encircled by solder yet untouched by it, as shown in Figure 4-54. You may find this undetectable unless you observe it with magnification.

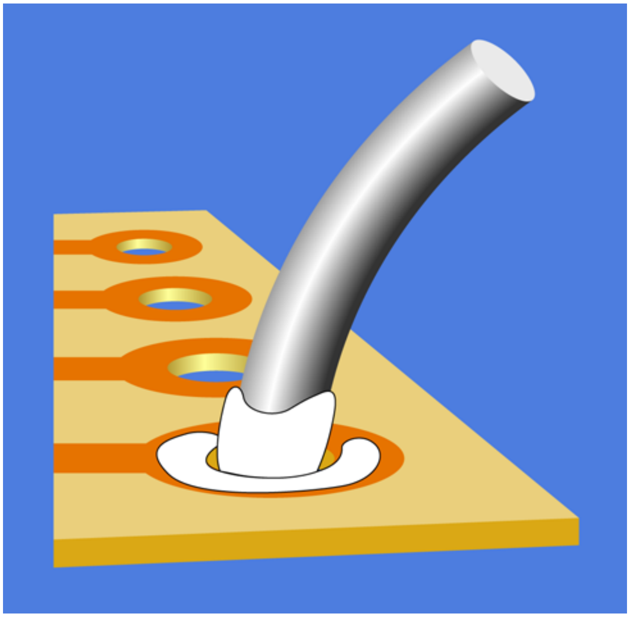

Figure 4-54 *Too little solder (or insufficient heat) can allow a soldered wire to remain separate from the soldered copper on the perforated board. Even a hair-thin gap is sufficient to prevent an electrical connection.*

You can add more solder to any joint that may have insufficient solder, but be sure to reheat the joint thoroughly.

3. Components incorrectly placed. It's very easy to put a component one hole away from the position where it should be. It's also easy to forget to make a connection.

I suggest that you print a copy of the schematic, and each time you make a connection on the perforated board, you eliminate that wire on your hardcopy, using a highlighter.

4. Debris. When you're trimming wires, the little fragments that you cut don't disappear. They start to clutter your work area, and one of them can easily get trapped under your perforated board, creating an electrical connection where you don't want it.

Clean the underside of your board with an old toothbrush before you apply power to it. Dip the toothbrush in rubbing alcohol to remove flux residues. Keep your work area as neat as possible. The more meticulous you are, the fewer problems you'll have later.

Once again, be sure to check every joint with a magnifying glass.

Fundamentals: Perforated Board Fault Tracing

If the circuit that worked on your breadboard doesn't work after you solder it to perforated board, your fault tracing procedure is a little different from that which I outlined previously.

First look at component placement, because this is the easiest thing to verify.

If all the components are placed correctly, flex the board gently while applying power. If you now get an intermittent response from the circuit, you can be virtually certain that solder didn't stick where it was supposed to stick, or a joint has a tiny crack in it.

Anchor the black lead of your meter to the negative side of the power supply, then switch on the power and go through the circuit point by point, from top to bottom, checking the voltage at each point with the red lead of the meter while you continue to flex the board. In most circuits, almost every part should show at least some voltage. If there is a dead zone, or if your meter responds intermittently, you can zero in on a joint that has something wrong with it, even though it looks good superficially.

A bright desk lamp and a magnifier are indispensable for this procedure. A gap of 1/1,000" or less is quite sufficient to stop your circuit from working. But you'll have difficulty detecting it without magnification, and even then, sometimes the light has to be exactly right.

Dirt, water, or grease can prevent solder from sticking properly to wires or copper traces. This is just another reason to be as meticulous as possible in your work habits.

The Project Box

The easiest way to enclose your perforated board is to put it in a project box. (I mentioned this item in the list of components and supplies at the beginning of Chapter 3). Hundreds of variants are available. Aluminum boxes look cool and professional, but you have to protect the circuit board from short-circuiting itself on the inside of the box. Plastic boxes are easier, and cheaper, too.

To make everything look professional, you should not just start drilling holes arbitrarily for your switches and LEDs. You should draw a layout on paper (or use a drawing program, and then print the image on paper). Just make sure there's room for the components to fit together, and try to place them similarly to the schematic, to minimize the risk of confusion.

Tape your sketch to the inside of the top panel, as shown in Figure 4-55, and then use a sharp, pointed tool (such as an awl or a needle) to press through and mark the plastic at the center of each hole. The indentations will help to center your bit when you drill the holes.

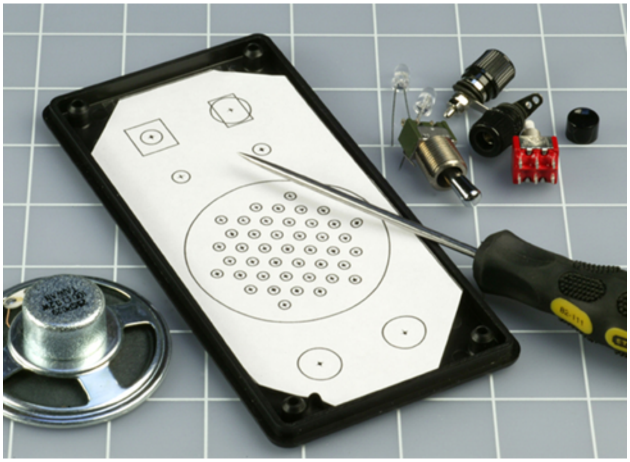

Figure 4-55 *A printed layout for the switches, LEDs, and other components has been taped to the underside of the lid of the project box. An awl is pressed through the paper to mark the center of each hole to be drilled in the lid.*

If you're using an audio circuit driving a loudspeaker (instead of a ready-made siren) you'll need to make multiple holes to vent the sound from the loudspeaker, which will be beneath the top panel of the box. The panel that I made is shown in Figure 4-56.

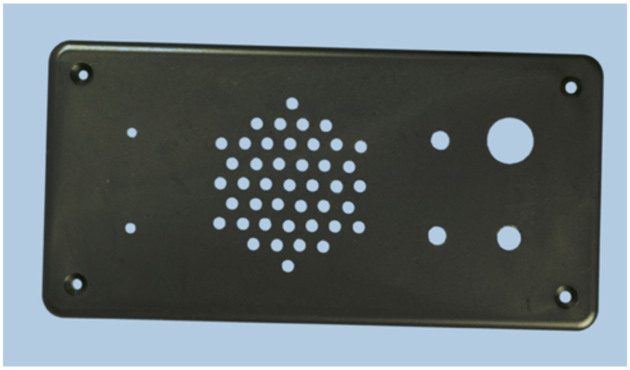

Figure 4-56 *The exterior of the panel after drilling. A small handheld cordless drill can create a neat result if the holes were marked carefully.*

I placed all the switches and LEDs on the top panel. The power input jack is positioned at one end of the box. Naturally, each hole has to be sized to fit its component, and if you have calipers, they'll be very useful for taking measurements and selecting the right drill bit. Otherwise, make your best guess, too small being better than too large. A *deburring tool* is ideal for slightly enlarging a hole so that a component fits snugly. This may be necessary if you drill 3/16-inch holes for your 5-inch mm LEDs. Fractionally enlarge each hole, and the LEDs should push in very snugly.

If your loudspeaker lacks mounting holes, you'll have to glue it in place. I used five-minute epoxy to do this. Be careful not to use too much. You don't want any of the glue to touch the speaker cone.

Drilling large holes in the thin, soft plastic of a project box can be a problem. The drill bit tends to dig in and create a mess. You can approach this problem in one of three ways:

- Use a Forstner drill bit if you have one. It creates a very clean hole. A hole saw can also be used.
- Drill a series of holes of increasing size.
- Drill a smaller hole than you need, and enlarge it with a deburring tool.

Regardless of which approach you use, you'll need to clamp or hold the top panel of the project box with its outside surface face-down on a piece of scrap wood. Then drill from the inside, so that your bit will pass through the plastic and into the wood.

Finally, mount the components in the panel, as shown in Figure 4-57, and turn your attention to the underneath part of the box.

Soldering the Switches

Your first step is to decide which way up the switch should be. Use your meter to find out which terminals are connected when the switch is flipped. You'll probably want the switch to be on when the toggle is flipped upward. The underside of my control panel is shown in Figure 4-57. I used a DPDT switch because I just happened to have one. You only need a SPST switch for the project.

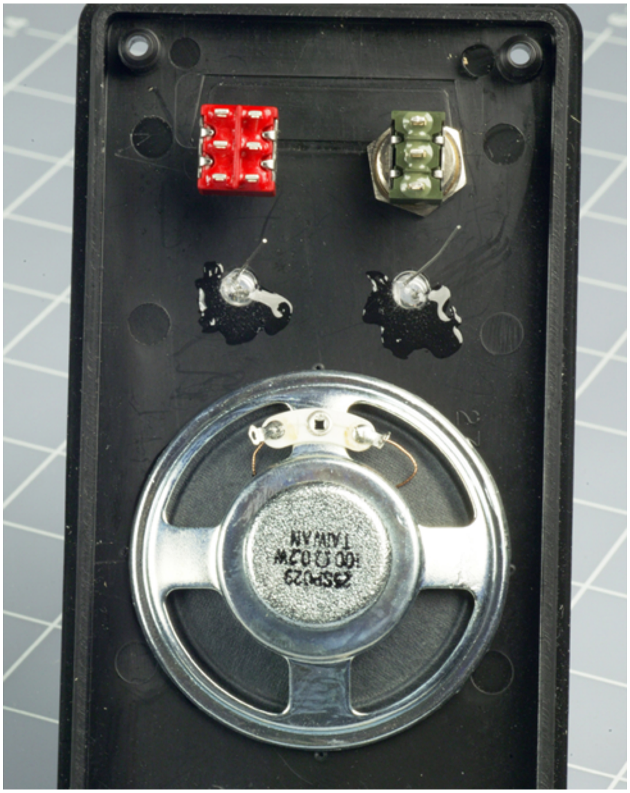

Figure 4-57 *Components have been added to the control panel of the project box (seen from the underside). The loudspeaker has been glued in place. Spare glue was dabbed onto the LEDs, just in case.*

Remember, the center terminal of any double-throw switch is almost always the pole of the switch.

Stranded wire is appropriate to connect the circuit board with the components in the top panel, because the strands flex easily and impose less stress on solder joints. Twisting each pair of wires together helps to minimize the mess.

When you connect wires or components with the lugs on the switches, your pencil-style soldering iron may not deliver enough heat to make good joints. You can use your higher-powered soldering iron in these locations, but you absolutely must apply a good heat sink to protect the LEDs when you attach them, and don't allow the iron to remain in contact with anything for more than 10 seconds. It will quickly melt insulation, and may even damage the internal parts of the switches.

In projects that are more complex than this one, it would be good practice to link the top panel with the circuit board more neatly. Multicolored ribbon cable is ideal for this purpose, with plug-and-socket connectors that attach to the board. For this introductory project, I didn't bother. The wires just straggle around, as shown in Figure 4-58.

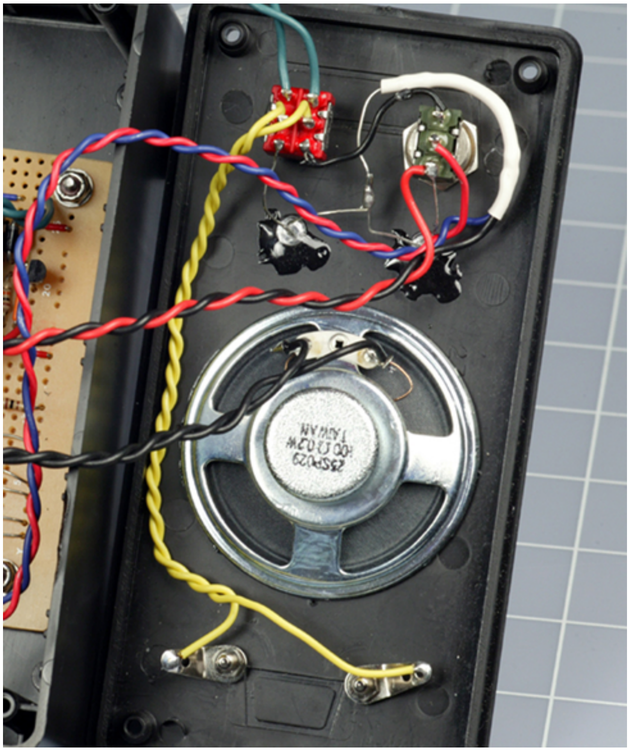

Figure 4-58 *Twisted wire pairs have been connected on a point-to-point basis, without much concern over neatness, as this is a relatively small project.*

Fixing the Board

The circuit board will sit on the bottom of the box, held in place with four #4-size machine screws (bolts) with washers and nylon-insert locknuts. I prefer using nuts and bolts, rather than glue, in case I ever need to remove the circuit to make a repair. You need to use locknuts to eliminate the risk of a nut working loose and falling among components, where it can cause a short circuit.

You'll have to cut the perforated board to fit, taking care not to damage any of the components on it. I use a band saw for cutting, but a hacksaw will work. Remember that perforated board often contains glass fibers that can blunt a wood saw.

Check the underside the board for loose fragments of copper traces after you finish cutting.

Drill bolt holes in the board, taking care again not to damage any components. Then mark through the holes to the plastic bottom of the box, and drill the box. Countersink the holes (i.e., bevel the edges of a hole so that a flat-headed screw will fit into it flush with the surrounding surface), push the little bolts up from underneath, and install the circuit board. Because you're using locknuts, which will not loosen, there's no need to make them especially tight. Indeed, excessive tightness must be avoided.

After mounting the board, test the circuit again, just in case.

Caution: Avoid Board Stress

Be extremely careful not to attach the circuit board too tightly to the project box. This can impose bending stresses, which may break a joint or a copper trace on the board.

Final Test

When you've completed the circuit, if you don't have your network of magnetic sensor switches set up yet, you can just use a piece of wire instead. I used a pair of *binding posts* on my box, just for convenience. You can equally well run a pair of wires from the circuit board out through a small hole in the lid of the box.

If everything works the way it should, it's time to screw the top of the box in place, pushing the wires inside. Because you're using a large box, you should have no risk of metal parts touching each other accidentally, but still, proceed carefully. My finished product is shown in Figure 4-59.

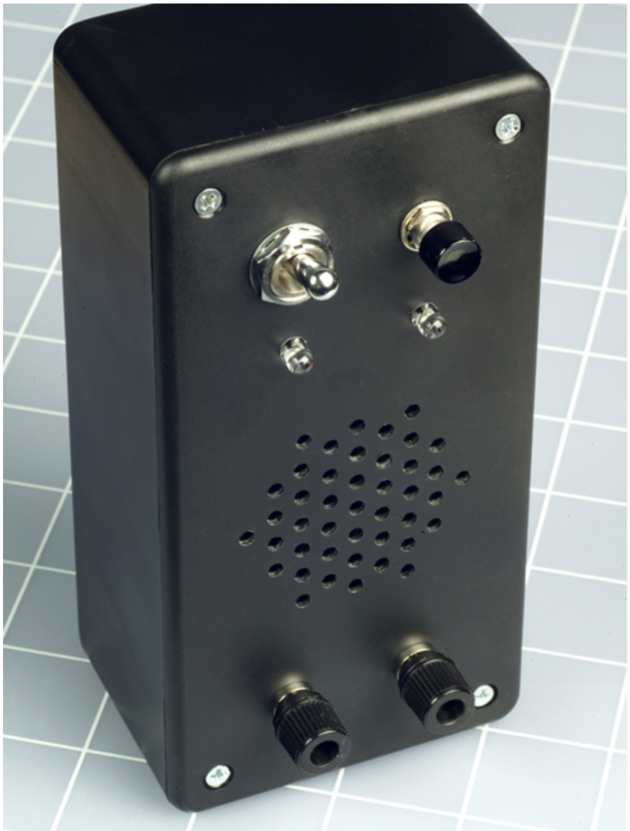

Figure 4-59 *The completed alarm box.*

Alarm Installation

If you are going to complete this project with magnetic sensor switches, you should test each one by moving the magnetic module near the switch module and then away from it, while you use your meter to test continuity between the switch terminals. The switch should close when it's next to the magnet, and open when the magnet is removed.

Now draw a sketch of how you'll wire your switches together. Always remember that they have to be in series, not in parallel! Figure 4-60 shows the concept in theory. The two terminals are the binding posts on top of your control box (which is shown in green), and the dark red rectangles are the magnetic sensor switches on windows and doors. Because the wire for this kind of installation usually has two conductors, you can lay it as I've indicated but cut and solder it to create branches. The solder joints are shown as orange dots. Note how current flows through all the switches in series before it gets back to the control box.

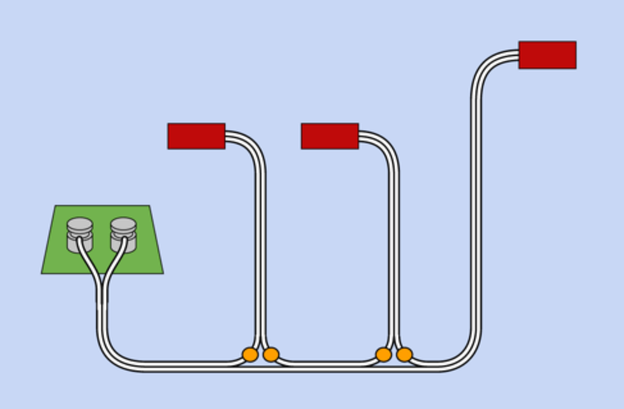

Figure 4-60 *Dual-conductor, white insulated wire can be used to connect the terminals on the alarm control box with magnetic sensors (shown in dark red). Because the sensors must be in series, the wire is cut and joined at positions marked with orange dots.*

Figure 4-61 shows the same network as you might actually install it in a situation where you have two windows and a door. The blue rectangles are the magnetic modules that activate the switch modules.

You'll need a large quantity of wire, obviously. The type of white, stranded wire that is sold for doorbells or furnace thermostats is good. Typically, it is 20-gauge or larger.

After you install all the switches, clip your meter leads to the wires that would normally attach to the alarm box. Set your meter to test continuity, and open each window or door, one at a time, to check whether you're breaking the continuity. If everything is OK, attach the alarm wires to the binding posts on your project box.

Now deal with the power supply. Use your AC adapter, with its output set to 9 volts, hooked up to your DC power plug, or attach the power plug to a 9-volt battery. The circuit will run off a 12-volt alarm battery too, but you will have to substitute a 12-volt relay for the one that I specified.

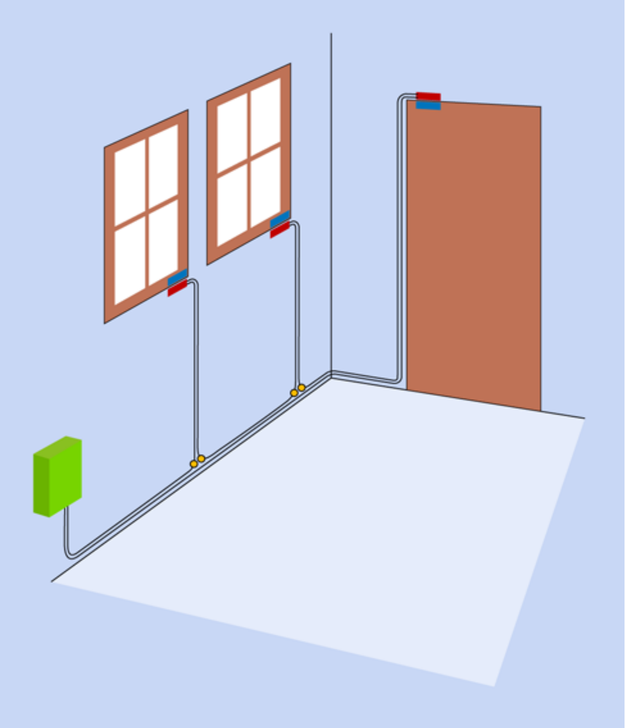

Figure 4-61 *In an installation involving two windows and a door, the magnetic components of the sensors (blue rectangles) could be placed as shown, while the switches (dark red) are located alongside them.*

The only remaining task is to label the switch, button, power socket, and binding posts on the alarm box. You know that the switch puts the alarm in and out of continuity testing mode, and the button gives you a minute to leave before the system alarms itself, but no one else knows, and you might want to allow a guest to use your alarm while you're away. For that matter, months or years from now, you may forget some details yourself.

Conclusion

The alarm project has taken you through the basic steps that you will usually follow any time you develop something:

- Make a wish list.
- Decide which types of components are appropriate.
- Draw a schematic and make sure that you understand it.

- Modify it to fit the pattern of conductors on a breadboard.
- Install components on the breadboard and test the basic functions.
- Modify or enhance the circuit, and retest.
- Transfer to perforated board, test, and trace faults if necessary.
- Add switches, buttons, a power jack, and plugs or sockets to connect the circuit with the outside world.
- Mount everything in a box (and add labeling).

Experiment 19: Reflex Tester

Because the 555 timer can run at thousands of cycles per second, you can use it to measure human reflexes. You can compete with friends to see who has the fastest response—and note how your response changes depending on your mood, the time of day, or how much sleep you got last night.

This circuit is not conceptually difficult, but requires quite a lot of wiring, and will only just fit on a breadboard that has 60 rows of holes (or more). Still, it can be tested in sections, like the circuit in Experiment 18. If you avoid making errors, the whole project can be assembled in a couple of hours.

What You Will Need

- Breadboard, hookup wire, wire cutters, wire strippers, multimeter
- 9-volt power supply (battery or AC adapter)
- 4026B chips (3)
- 555 timers (3)
- Resistors: 470 ohms (2), 680 ohms (3), 10K (6), 47K (1), 100K (1), 330K (1)
- Capacitors: 0.01µF (2), 0.047µF (1), 0.1µF (1), 3.3µF (1), 22µF (1), 100µF (1)
- Tactile switches (3)
- Generic LEDs: one red, one yellow
- Trimmer potentiometer, 20K or 25K (1)
- Single-digit numeric LED displays, height 0.56”, low-current red preferred, able to function at 2V forward voltage and 5mA forward current (3) (Avago HDSP-513A preferred, or Lite-On LTS-546AWC, or Kingbright SC56–11EWA, or similar)

Caution: Protecting Chips from Static

The 555 timer is not easily damaged, but in this experiment you will also be using a CMOS chip (the 4026B counter), which is more vulnerable to static electricity.

Whether you are likely to zap a chip by handling it depends on factors such as the humidity in your location, the type of shoes you wear, and the type of floor covering in your work area. Some people seem to accumulate a charge of static more easily than others, and I don’t have an explanation for this. Personally, I have never damaged a chip with static, but I know people who have.

If static is a risk for you, you’ll probably know about it, because you’ll suffer from sudden little jolts when you reach for a metal door handle or a steel faucet. If you really feel you need to protect chips from this kind of discharge, the most thorough precaution is to ground yourself. The best way to do this is by using an anti-static wrist strap. The conductive strap is secured to your wrist with velcro and connects through a high-value resistor (typically 1M) to an alligator clip that can be attached to any large metal object.

When you receive chips via mail order, they are usually shipped in channels of conductive plastic, or with their legs embedded in conductive foam. The plastic or the foam protects the chips by insuring that all the pins have an approximately equal electrical potential. If you want to repackage your chips, but you don’t have any conductive foam, you can poke their legs through aluminum foil.

Caution: Be Careful When Grounded

The resistor built into an anti-static wrist strap protects you from electrocuting yourself if you happen to touch a source of relatively high voltage with your other hand. This is an important feature, as an electric shock that travels from one hand to the other will pass through your chest and can stop the heart.

If you use just a plain piece of wire to ground yourself, you lose this protection. The small cost of a proper wrist strap is a sensible investment.

Now, back to the experiment.

A Quick Demo

In the previous edition of this book I suggested using a single three-digit display for this project. In this edition, I have switched to three individual digits. The cost is fractionally higher, but the wiring is much simpler, and the project can be built more easily. Also, I believe there is a better chance of single LED digits remaining available for many years to come.

I am specifying digits that are 0.56" high because this is an industry standard, with pinouts that are also standardized. If you use a smaller size, the pinouts will be different. If you use a larger size, they won't fit with the other components on the breadboard.

Let's begin by getting acquainted with one of the numerals, and the 4026B chip that will be driving it.

The first module of this circuit is shown in Figure 4-62. (If you feel that a schematic is easier to understand, see Figure 4-63, which shows the same components.)

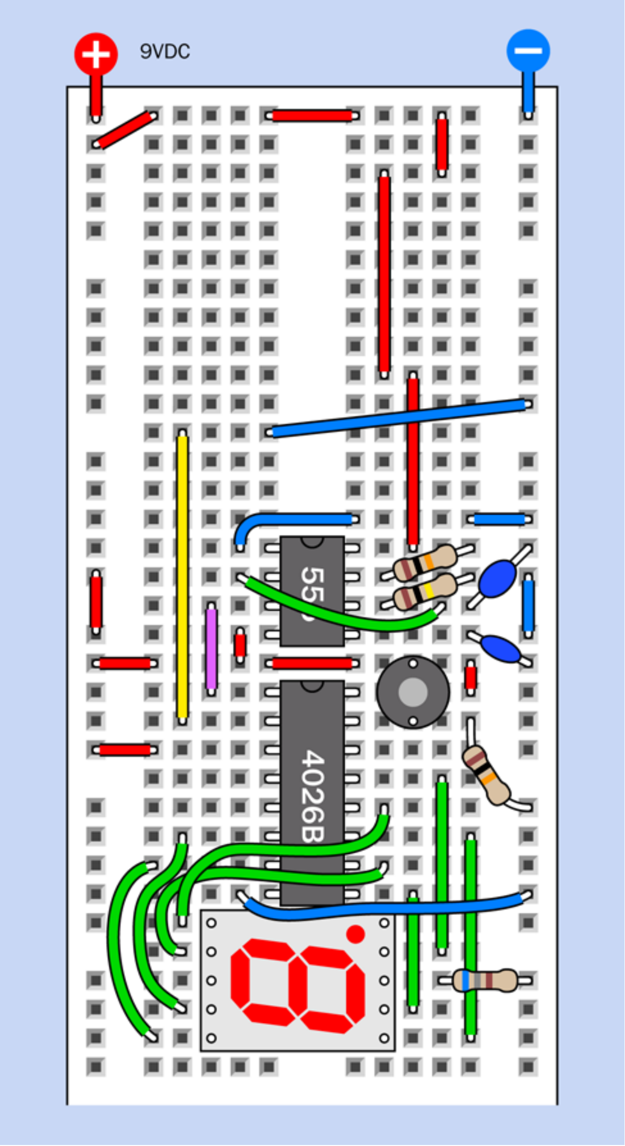

Figure 4-62 *The first module of the reflex tester demonstrates how a timer can run a counter chip that powers a single-digit LED display.*

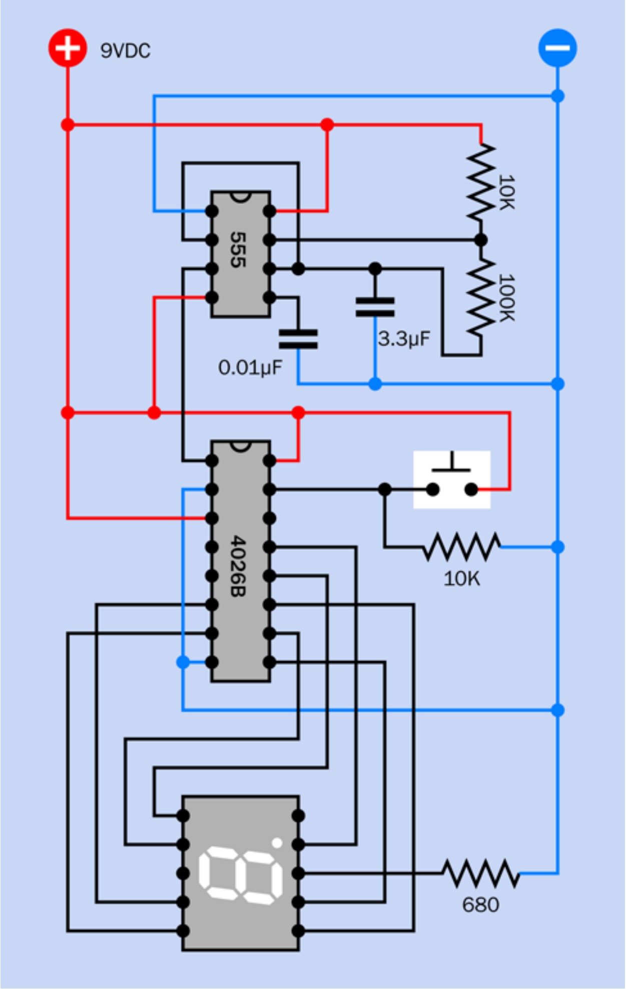
9VDC
10K
100K
555
3.3µF
0.01µF
4026B
10K
680

Figure 4-63 *The first module shown in schematic form.*

In addition, component values are shown in Figure 4-64.

You're going to be adding a lot more components to the board (in fact, it will be packed full by the time you're finished), so you need to position everything exactly as I have shown it in the figure. Count the rows of holes carefully! You'll see some wires that don't make sense right now (all those red pieces—what are they for?) but they will enable you to add and activate a couple more 555 timers as you proceed.

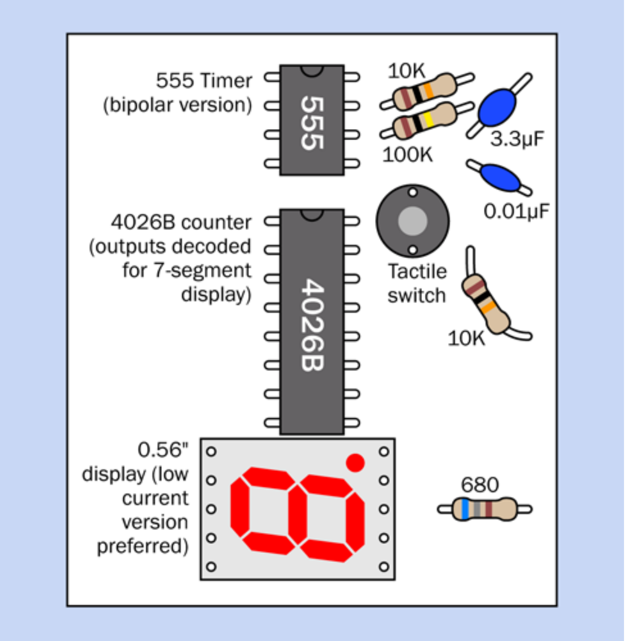
555 Timer (bipolar version)
555
10K
100K
3.3µF
0.01µF
4026B counter (outputs decoded for 7-segment display)
4026B
Tactile switch
10K
0.56" display (low current version preferred)
680

Figure 4-64 *Component values for the first module of the circuit.*

Apply power from a 9-volt battery or AC adapter, and you should see the numeric display counting repeatedly from 0 through 9.

If you don't see any numerals at all, set your meter to measure DC volts, clip the black probe to the negative side of the power supply, and use the red probe to test for voltages at key locations in the circuit, such as the power input pins of the chips. If the voltages look okay, make sure the resistor at bottom-right is 680 ohms (not 68K or 680K, which are colored similarly).

If the display shows fragments of numerals, or numerals that are not counting sequentially, you made an error in the green wires connecting with the 4026B chip.

If the display shows a 0 that does not change, you wired the 555 timer wrongly, or failed to connect the timer with the 4026B chip correctly.

Once you have the numbers counting upward, hold down the tactile switch, and notice that it forces the counter back to 0. As soon as you let go of the tactile switch, the numbers start counting again.

We have the basis for a reflex tester right here. We just need to add a couple more digits, increase the counting speed, and make some more refinements. First, however, I'll explain what's going on.

Fundamentals: LED Displays

The term "LED" is slightly confusing. The type of component that you have used in previous experiments is properly known as a *standard LED*, a *through-hole LED*, or an *LED indicator*: a little rounded component with two long leads sticking out of the base. These became so common, people started calling them simply "LEDs." But LEDs are used in other components, too, such as the glowing numeral currently plugged into your breadboard. It is properly known as an *LED display*. More precisely, it is a *seven-segment single-digit LED display*.

In Figure 4-65 the dimensions of the display are shown, along with the pin positions that are hidden underneath it. The important things to notice are that the numeral does consist of seven segments, plus a decimal point; and the pin spacings are all measured in multiples of 0.1", which is helpful for your breadboard.

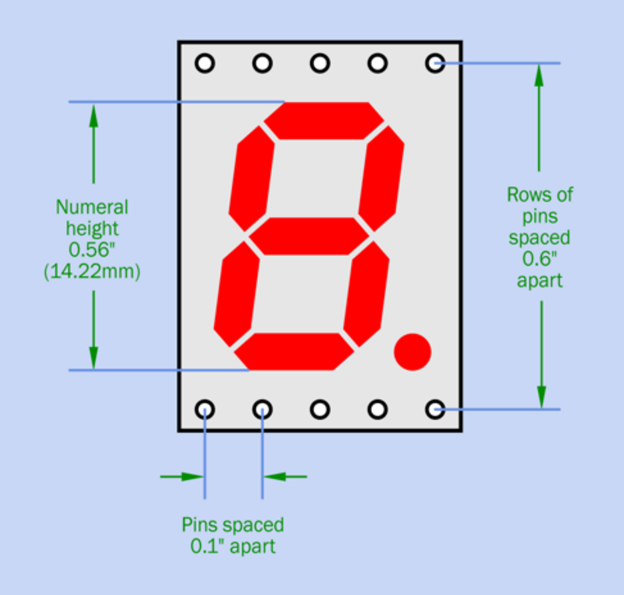
Numeral height 0.56" (14.22mm)
Rows of pins spaced 0.6" apart
Pins spaced 0.1" apart

Figure 4-65 *Dimensions and pin locations of a standard 0.56" seven-segment LED display.*

Now take a look at Figure 4-66. This shows you the internal connections between the pins and the segments of the digit. Notice that pins 3 and 8 have blue centers, indicating that they should be connected with negative ground. All the other pins are designed to receive positive power to activate the LED segments. This is called a *common cathode* type of LED, because the negative sides of the internal diodes (the cathodes) are all tied together.

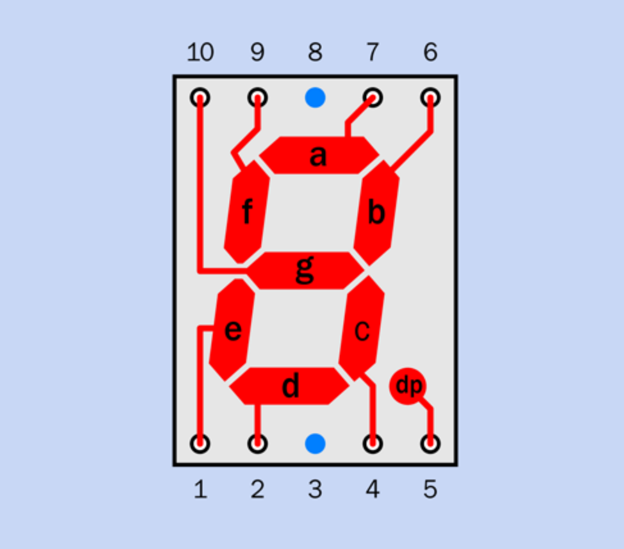

Figure 4-66 *The scheme that is used for pin numbering, and the internal connections that are hidden inside the component.*

In a *common anode* display, the situation is reversed, and you activate the segments by applying negative power to each one, while they all share a positive internal connection. You can choose whichever type of display is convenient in a circuit, but common cathode displays are more widely used.

Note that the segments are identified with lowercase letters a through g, plus dp for the decimal point. This system is common on almost all datasheets (although a few use the letter h for the decimal point).

So far so good, but I have left out a crucial piece of information: like all LEDs, the segments in the numeral must be protected with series resistors. This is a hassle, and you may wonder why the manufacturer doesn't build resistors in. The answer is that the display must be usable on a wide variety of voltages, and the values of the resistors will depend on the voltage.

Well—why can't we use just one resistor that all the segments share, perhaps between pin 3 and negative ground? Actually we can do that, but the resistor will be dropping the voltage and restricting the current for different numbers of segments, depending on which numeral is being displayed. A numeral 1 lights up only two segments, while a numeral 8 uses all seven of them. Consequently, some numerals will be brighter than others.

Does this really matter? I'm thinking that for this demo, simplicity may be more important than perfection. If you look at Figure 4-62 you'll see that I have, in fact, put just one 680-ohm resistor at bottom-right, between the LED display and the negative bus. This is not the correct procedure, but you are going to install three of the seven-segment displays in this project, and I think you'll be happy to use only three series resistors instead of 21.

Fundamentals: The Counter

The 4026B chip is known as a *decade counter,* because it counts in tens. Most counters have a *coded output*, meaning that they output the numbers in binary-coded format (which I will discuss in a later project). This counter doesn't do things this way. It has seven output pins, and it powers them in patterns that just happen to be correct for a seven-segment display. While other counters require a *driver* to convert a binary output to the seven-segment patterns, the 4026B gives you everything in one package.

This is very convenient, except that the 4026B is an old-fashioned CMOS chip that has limited power. The datasheet advises you to draw less than 5mA from any one pin when powering the chip with 9 volts.

Ideally you should pass the outputs from the counter through an array of transistors, to amplify them. You can buy a chip containing seven transistor pairs for exactly this purpose. It's called a *Darlington array*. (What if you want to display the decimal point? No problem. You can buy a different Darlington array containing eight transistor pairs.)

I could have used three Darlington-array chips to drive the three LED displays in this project, but that would have added to the complexity and expense, and I would have needed two breadboards. Therefore I decided it was acceptable to use low-current LED displays that can be driven directly by a counter. They're not as bright, but they do the job. I chose the 680-ohm resistor because it should limit the current to less than 5mA from any single pin on the counter chip, and it imposes a voltage drop of around 2V across the LEDs (which varies depending how many segments are lit up).

Now I'll give you some details about the internal workings of the 4026B. Counter chips always have several useful features built in. Take a look at Figure 4-67, which shows the pinouts of the chip. The pins with labels such as "To segment a" are easy to understand. You simply run a wire from each pin to the appropriate pin on your LED display. If you look at Figure 4-62 you'll see each of the green wires connecting the output pins from the counter with the input pins on the numeral.

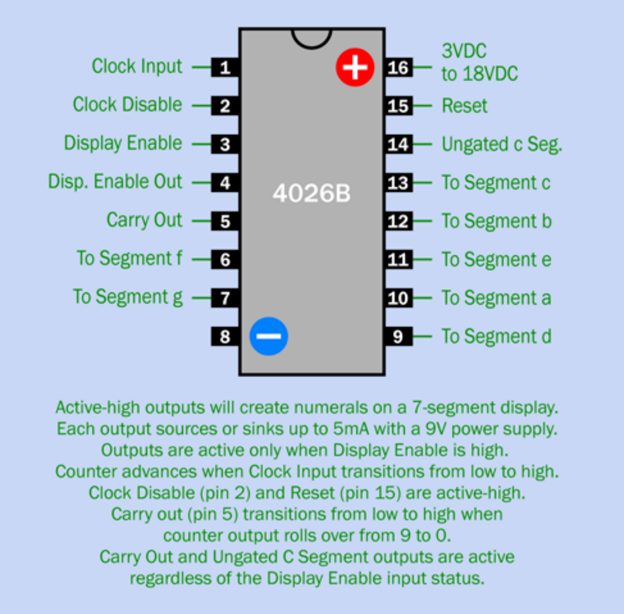
Clock Input 1
Clock Disable 2
Display Enable 3
Disp. Enable Out 4
Carry Out 5
To Segment f 6
To Segment g 7
8
4026B
16 3VDC to 18VDC
15 Reset
14 Ungated c Seg.
13 To Segment c
12 To Segment b
11 To Segment e
10 To Segment a
9 To Segment d
Active-high outputs will create numerals on a 7-segment display.
Each output sources or sinks up to 5mA with a 9V power supply.
Outputs are active only when Display Enable is high.
Counter advances when Clock Input transitions from low to high.
Clock Disable (pin 2) and Reset (pin 15) are active-high.
Carry out (pin 5) transitions from low to high when
counter output rolls over from 9 to 0.
Carry Out and Ungated C Segment outputs are active
regardless of the Display Enable input status.

Figure 4-67 *Pinouts for a 4026B counter chip, which has decoded outputs to drive a seven-segment single-digit LED display.*

Pins 8 and 16 of the chip are for negative-ground and positive power, respectively. Almost all digital chips apply power to opposite corners in this way (with the exception of the 555 timer—although, really, it is classified as an analog chip).

Because Figure 4-67 contains information that you may not need and is just included for future reference, I have also created a simplified view of the counter and the display in Figure 4-68, ignoring the pins that we won't be using, and showing the relationship of output pins with the pins on the display.

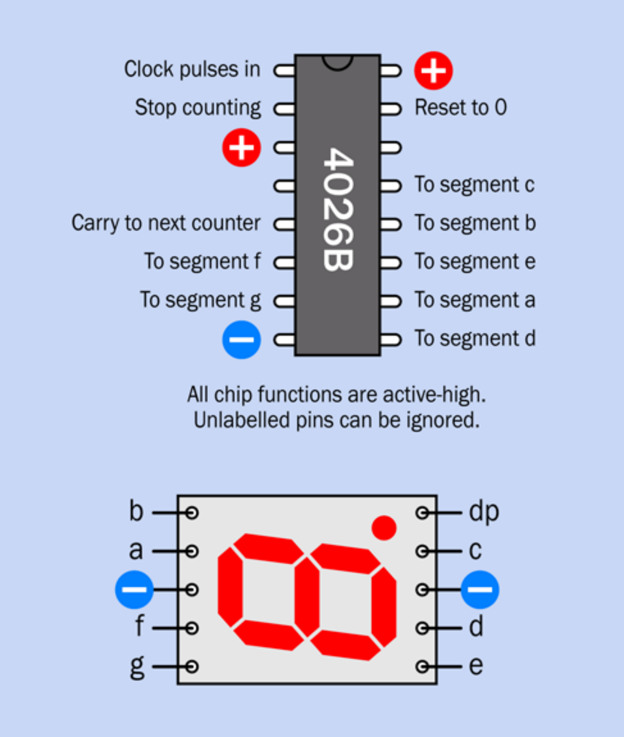

Figure 4-68 *A simplified view of the chip and the display, in the style that they will be shown on the breadboard.*

Look at pin 15, the reset pin. Now look at Figure 4-62. The pushbutton, more properly known as a tactile switch, is placed so that it will apply positive voltage to pin 15 when you press the button. (The voltage makes its way across the board to the pushbutton via the red segments of wire that I mentioned a moment ago.)

When you're not pressing the button, no positive voltage is applied to the reset pin of the counter. However, a 10K resistor permanently connects pin 15 to the negative bus on the breadboard. This is a *pulldown resistor*. It pulls the voltage on the pin to near zero—until you press the button, at which point the positive input overwhelms the negative power supplied through the resistor. Remember, if you don't apply a defined voltage to each input pin of a digital chip, you will get random, inexplicable, and confusing results. I've mentioned this before, but I have to emphasize the topic, because it's such a common cause of errors.

- To keep an input pin normally high, connect it to the positive bus through a 10K resistor (at least, for the circuits in this book). When you need to pull it low, use a switch or other device that overrides the resistor by making a more direct connection to the negative bus.
- To keep an input pin normally low, connect it to the negative bus through a 10K resistor. When you need to push it high, use a switch or other device that overrides the resistor by making a more direct connection to the positive bus.
- All the inputs on a counter chip must have some kind of connection. Never allow any floating input pins!
- Unused output pins should be left unconnected.

One more thing. Sometimes a chip has an input that we won't need at all. The 4026B, for instance, tells us that pin 3 is a display enable input. I want the display to be enabled all the time, so I connected pin 3 directly to the positive bus on a set-it-and-forget-it basis.

- If you won't be using an input pin, it must still have a defined state. You can deal with it by wiring it directly to the positive or negative side of the power supply.

Now I'll run through the remaining features of the 4026B.

The *clock input* (pin 1) accepts a stream of high and low pulses. The chip doesn't care how long the pulses are. It just responds by adding 1 to its count, each time it senses the input voltage rising from low to high.

The *clock disable* (pin 2) tells the counter to block the clock input. Like all the other pins on the chip, this one is *active-high*, meaning it performs its function when it has a high state. On your breadboard, I ran in a temporary blue wire and yellow wire to hold pin 2 low. In other words, I disabled the clock disable pin. This is confusing, so I will summarize the situation:

- When the clock disable pin is in a high state, it stops the counter from counting.
- When the clock disable pin is pulled down to negative ground, it allows the counter to count.

The *display enable* (pin 3) I already mentioned.

The *display enable out* (pin 4) will not be used here. It takes the state of pin 3 and connects it with pin 4 so that you can pass it along to other 4026B timers.

The *carry out* (pin 5) is essential if you want to count higher than 9. The pin state shifts from low to high when the counter has reached 9 and goes back to 0. If you take this output and connect it with the clock input pin of a second 4026B timer, the second timer will count in tens. You can use its carry output pin to signal a third timer, which will count in hundreds. I will be using this feature at the end of this project.

Lastly, Pin 14 can be used to restart the counter after it counts through 0, 1, and 2. This is useful in a digital clock that only counts up to 12 hours, but not relevant to us here. It is an output pin that we will not use, so it can be left unconnected.

Perhaps all the features seem confusing, but if you ever find yourself confronted with a counter chip that you've never seen before, you can figure it out (if you are patient and methodical) by looking up the manufacturer's datasheet. Then you can test it with LEDs and tactile switches, to make sure there are no misunderstandings. In fact, this is how I figured out the 4026B, myself.

Pulse Generation

Because the 555 timer accepts a power supply ranging from 5 to 15 volts, just like the 4026B, the output from the timer, on pin 3, can be connected directly to the input of the 4026B. That's the function of the purple wire in the breadboard layout. The 555 supplies pulses, and the 4026B counts them.

The rest of the wiring around the 555 timer should look familiar to you by now. You can see that it's running in astable mode. Your only question may be why it's running so slowly. We're not going to measure someone's reflexes at that speed.

True, but for this demo, I didn't want the digits to look like a featureless blur. We'll crank up the speed a bit later.

Time for a Plan

How should the reflex timer work? Here's my wish list:

1. I need a start button.
2. After the start button is pressed, there is a delay in which nothing happens. Suddenly there is a visual prompt, asking the player to respond.
3. Simultaneously, the count starts from 000, in 1/1,000ths of a second.
4. The user has to press a button to stop the counting process.
5. The count freezes, showing how much time elapsed between the prompt and the stop time. This measures the user's reflexes.
6. A reset button sets the count back to 000.

You already installed a reset button on the breadboard, which is necessary. But before that, we need a button that stops the counting process.

The counter's clock disable pin will freeze the display, but if you want to keep the display frozen, the pin has to be held in a continuing high state. In other words, it has to be latched.

Hmmm, that sounds as if we could use another 555 timer, wired in bistable mode.

A Control System

In Figure 4-69 the bistable timer has been added, with two new buttons. The diagonal blue wire in Figure 4-62 has been removed (to make room for the new timer). The other parts that you installed previously are still there, but are shown in gray.

Figure 4-70 shows the new part of the circuit in schematic form, and values of the additional parts are shown in Figure 4-71.

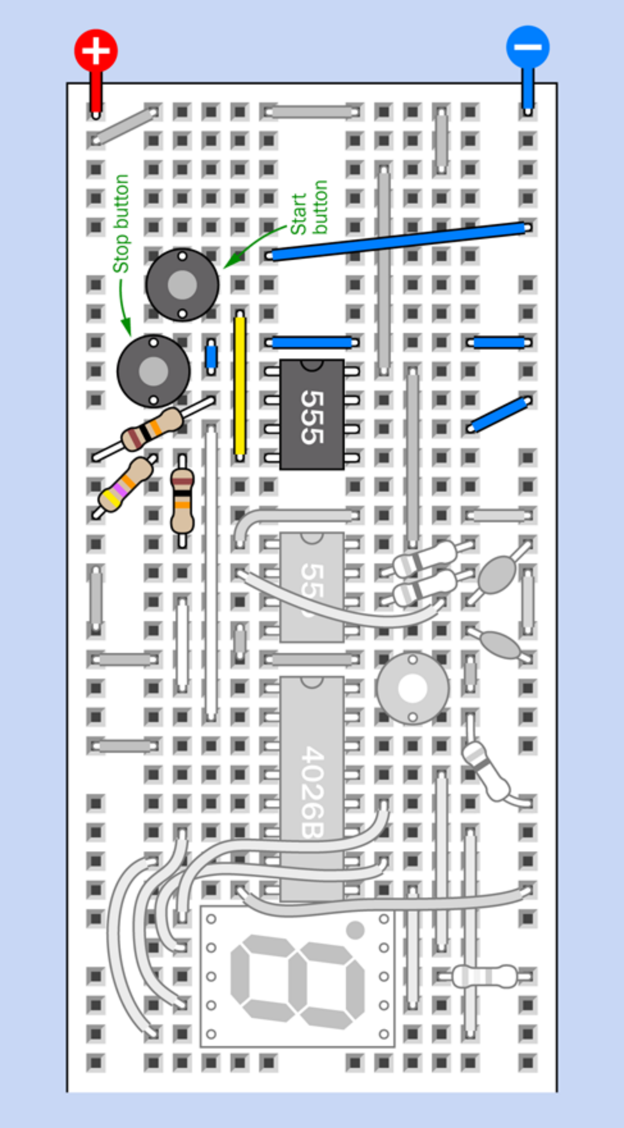
Stop button
Start button
555
555
4026B

Figure 4-69 *A bistable 555 timer has been added. Previously wired components are grayed out.*

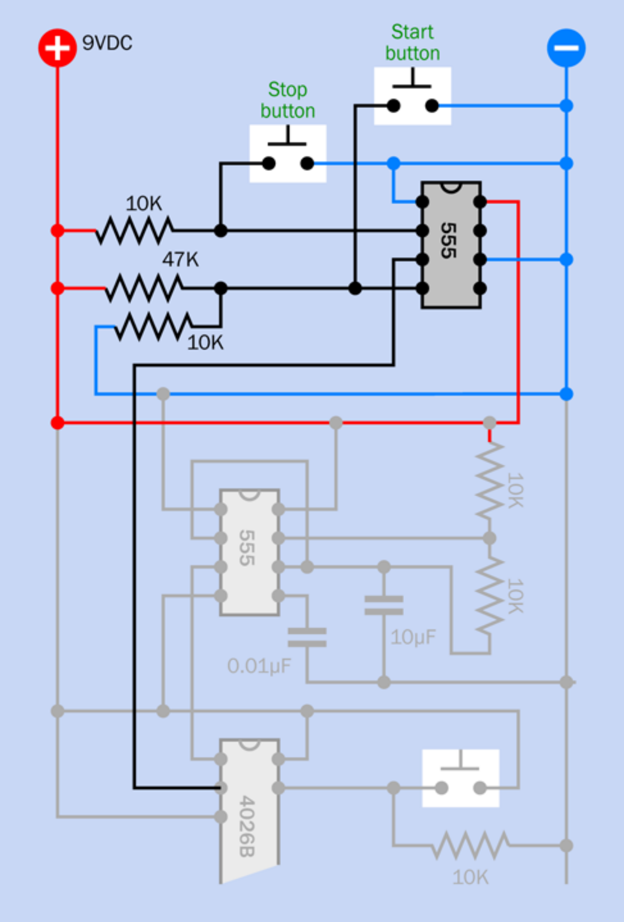
9VDC
Start button
Stop button
10K
47K
10K
555
10K
10K
555
10µF
0.01µF
4026B
10K

Figure 4-70 *This schematic shows the second timer and its associated components. Previous components are grayed out.*

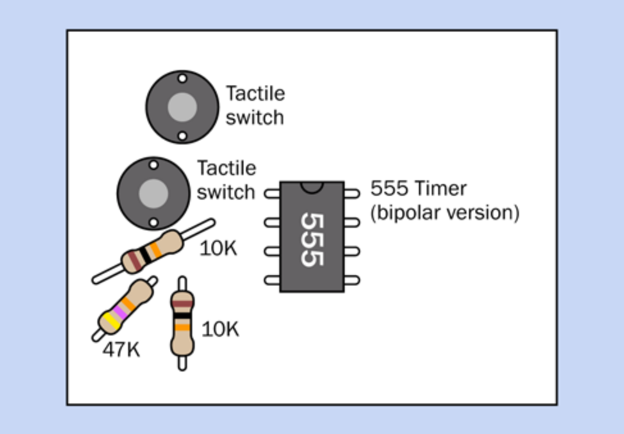
Tactile switch
Tactile switch
10K
555 Timer (bipolar version)
555
47K
10K

Figure 4-71 *Values of the parts that have been added in this module of the project.*

After you build this new section of the circuit, you can try it. You should find that the two new buttons start and stop the counting process. Do you see how they work?

Press the start button, and it grounds the reset pin of the bistable timer. The timer output on pin 3 goes low, and is connected to the clock disable pin on the counter. Remember, a *low* state on the disable pin means that the counter is *not* disabled. So, the counter starts counting. And it goes on counting, because once you trigger the bistable timer, its output is latched, and continues indefinitely.

But, you can stop it. Just press the stop button. This grounds the input pin of the bistable timer, triggering it. Consequently, the output from the timer goes high, and because the timer is working in bistable mode, the output is latched and stays high indefinitely. The high output goes to the clock disable pin, which stops the counter.

When you press the button at lower-right, which you installed initially, it still resets the timer to 000. But the timer remains latched in its disabled mode until you restart it with the start button.

The bistable 555 was exactly what we needed to run this circuit.

Progress Report

Let's see how far we've proceeded in satisfying the wishlist. It looks to me as if we're almost there. You press a button to start it, press a second button to stop it, and then once it's stopped, you press another button to set it to zero.

The only thing missing is the element of surprise. Really the person using it shouldn't know when the count will start. The idea is to measure the speed of his reflexes when he responds.

Why not add yet another timer, in monostable mode, to insert a delay before the action begins? This way, the start will be unexpected.

The Delay

First remove the start button, and the diagonal piece of blue wire connecting it with the negative bus. Leave the vertical yellow piece of wire where it is.

Now assemble some additional components as shown in Figure 4-72. The start button has been relocated to trigger the input of a third timer, which will insert a preliminary delay. The output from this timer will go high for 5 or 10 seconds, and then when it goes low it triggers the bistable timer, sending its output low, to suppress the clock disable feature of the 4026B, so that it starts counting.

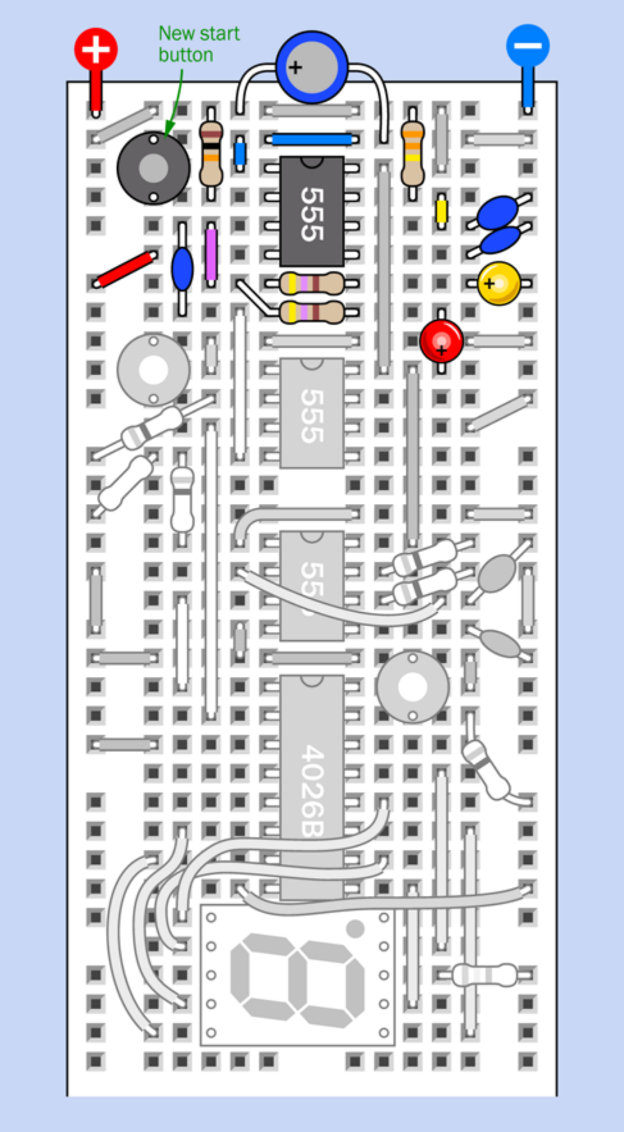

Figure 4-72 *The upper part of the reflex timer circuit is now complete.*

Be careful when installing the red and yellow LEDs. The red one is the opposite way around from what you might expect, because it is tied in to a positive source. So, its long, positive lead is at the bottom, not the top.

A schematic version of the new part of the circuit appears in Figure 4-73.

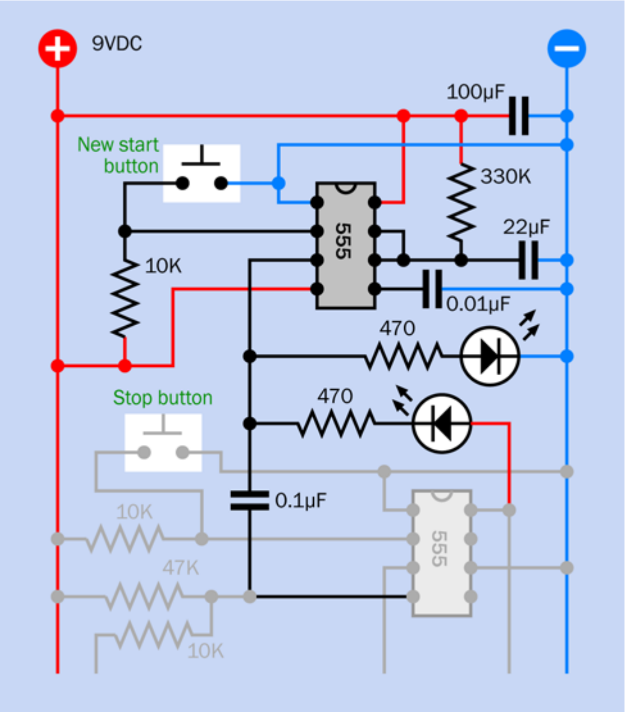

Figure 4-73 *Schematic for the new and final addition to the control circuit.*

The values of the parts that you added to the breadboard are shown in Figure 4-74.

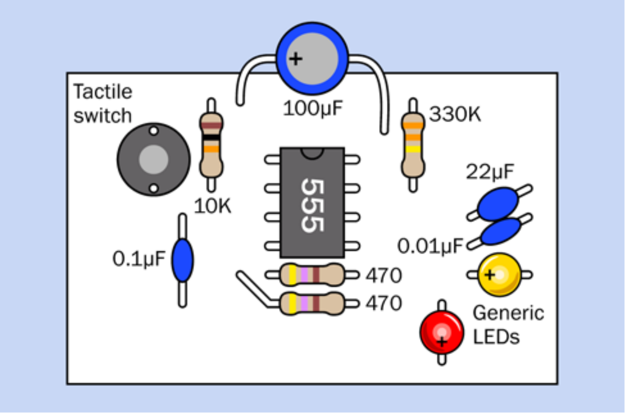

Figure 4-74 *The values of parts that have been added to the breadboard.*

Testing

When you apply power to this circuit, the counter immediately starts counting, without you asking it to. This is annoying, but easily dealt with. Press the stop button to stop the count. Press the bottom-right pushbutton to reset the timer to zero. Now you're ready for action.

Press the new start button, which creates an initial delay. During this delay, the yellow LED lights up. The delay lasts for about seven seconds, at which point the yellow LED goes out and the red LED comes on. Simultaneously, the counter starts counting—until you press the stop button.

The 100µF capacitor at the top of the breadboard looks like an afterthought, but is actually quite important. The 555 timer has a bad habit of creating voltage spikes when it switches its output, and in this circuit, the spikes can trigger the second timer without waiting for the delay. The 100µF capacitor suppresses this unfortunate tendency.

We now have all the features installed, except for the need to increase the counter speed and add a couple more counters and displays for fractions of a second.

How Does It Work?

Figure 4-75 shows you how the components are communicating with each other.

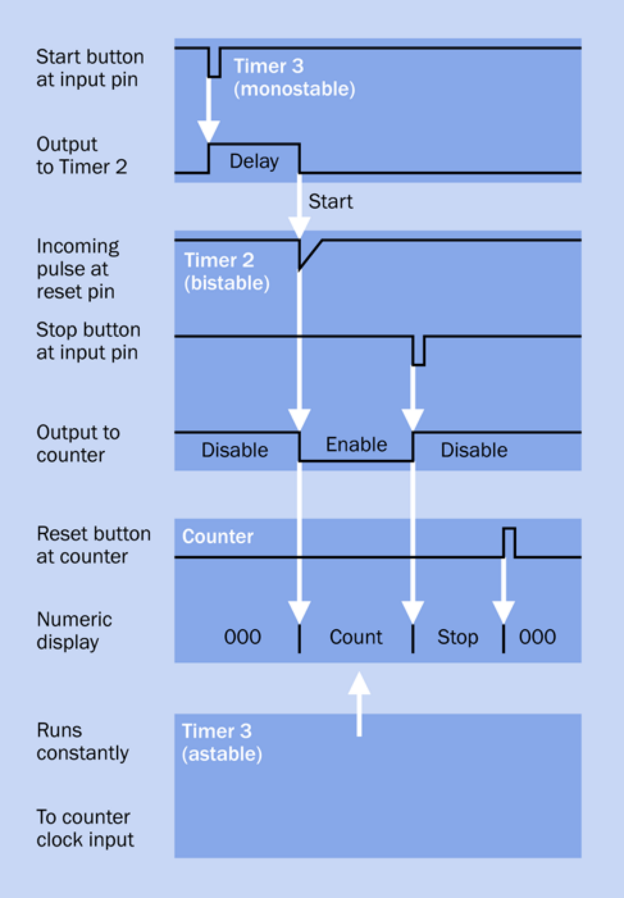
Start button
at input pin
Timer 3
(monostable)
Output
to Timer 2
Delay
Start
Incoming
pulse at
reset pin
Timer 2
(bistable)
Stop button
at input pin
Output to
counter
Disable
Enable
Disable
Reset button
at counter
Counter
Numeric
display
000
Count
Stop
000
Runs
constantly
Timer 3
(astable)
To counter
clock input

Figure 4-75 *Interaction between the components in the timer control circuit.*

I'll describe this figure from top to bottom. The start button (at the top, connected with Timer 3) pulls the input of the timer low to trigger it.

The output of Timer 3 goes high for about seven seconds. This creates your initial delay.

At the end of the delay, the output from Timer 3 drops back down. This transition passes through a 0.1µF coupling capacitor to Timer 2, which is bistable. The capacitor allows only a brief pulse to reach the reset pin of Timer 2. The pulse sends the output of Timer 2 low. This low output goes to the clock disable pin of the 4026B counter. A low state enables the counter, and the count begins.

Now we wait for the user to respond. The user presses the stop button for Timer 2, connected with its input on pin 2. The brief low input to Timer 2 sends its output high, which enables the clock disable pin on the counter, and stops the count.

Background: Development Issues

This project created a problem. When I built the original circuit several years ago, it ran fine. When some interns at Make: magazine built the circuit, it ran fine. Little did we know that the reset pin of a 555 timer behaves slightly differently on different brands. This is not documented on the datasheet.

After my book had been in print for literally years, I received notification from a reader that his version of the circuit functioned erratically, and sometimes not at all. I rebuilt the circuit, connected it with an oscilloscope, and saw that the coupling capacitor was transmitting a pulse faithfully from Timer 3 to the reset pin of Timer 2. But, sure enough, Timer 2 sometimes didn't recognize the pulse.

What was the problem? Either the pulse was too brief, or it didn't go low enough. Either way, the solution was to apply a lower pullup voltage to pin 4 of Timer 2. This is why you see two resistors attached to pin 4. They function as a voltage divider, applying slightly less than 2V to pin 4. This is enough to keep it functional, but allows the reset voltage to go down lower to enable the reset.

This now works fine—for me. We'll test the circuit again before this edition of the book goes into print. If it doesn't work for you, try applying a different voltage to pin 4 of Timer 2, using a substitute for the 47K resistor, either lower or higher. You can also try a larger coupling capacitor. And let me know. Obviously I want all the circuits in this book to work properly, every time. But I can't foresee all the manufacturing variations that may affect the outcome.

Extra Digits

Adding two more digits is pretty easy, as each will be controlled by its own 4026B counter, and all the counters and digits will be wired in basically the same way. This is shown in Figure 4-76.

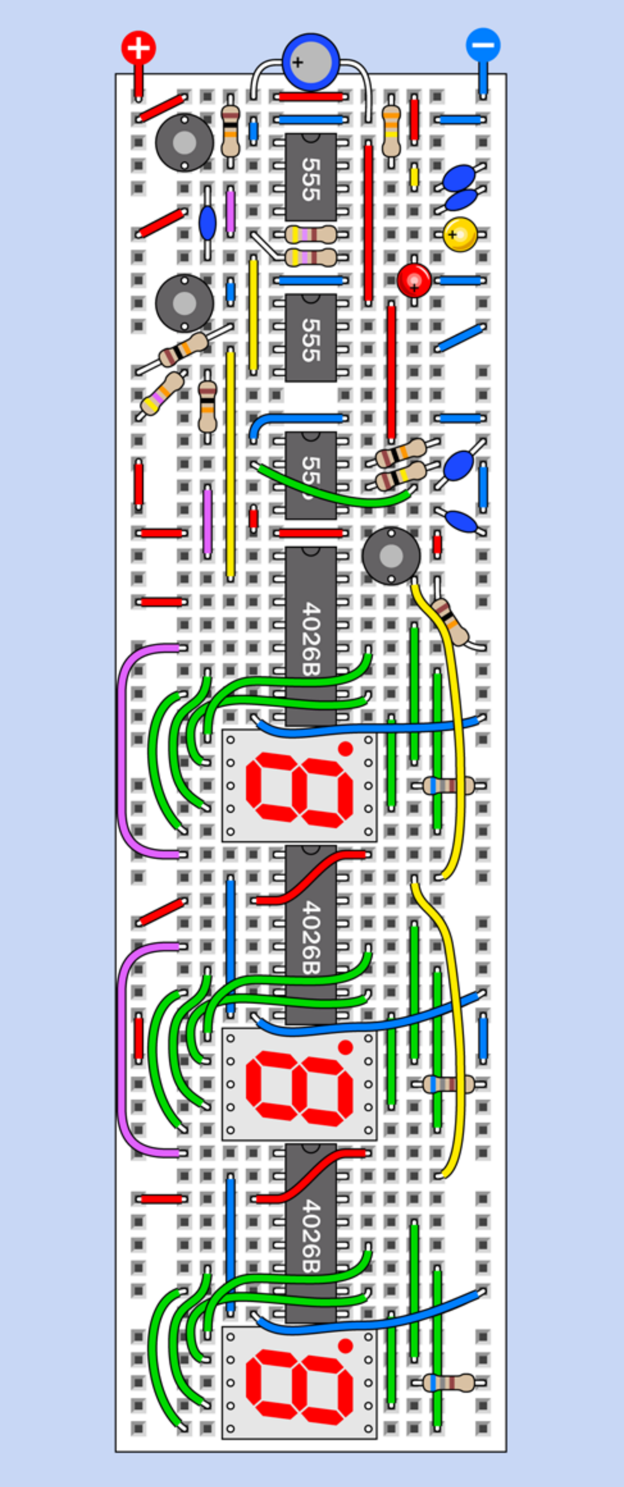

Figure 4-76 *The completed reflex timer barely fits on one 60-row breadboard.*

Notice the purple wires on the left. Each of them connects the carry output from one counter to the clock input of the next.

The yellow wires down the right side connect all the reset inputs of the counters together, so that when you reset one, you reset them all.

In the second and third counters, blue wires have been added to ground pin 2 of each. Remember, pin 2 is the clock disable pin. We never need to stop the second and third counters from counting, because they are completely under the control of the first counter. When it stops, they stop.

Don't forget to supply positive voltage to pin 16 (the power input pin) of the second counter and the third counter, by running a red wire across each chip, as shown.

Calibration

How do you make the circuit run at the right speed?

Begin by substituting a 10K resistor for the 100K resistor on the first timer that you installed, and replace the 3.3µF capacitor with a 47nF capacitor (0.047µF). In theory this should create a timer frequency of 1,023Hz, which is very close to the 1,000Hz that you want.

To fine-tune it, you'll have to substitute a trimmer potentiometer for one of the 10K resistors on the first timer that you installed. This circuit is so densely packed, there's barely enough room, but I figured out a way to squeeze in a trimmer. This is shown in Figure 4-77, which is a close-up of the immediate area around the lowest of the three timers.

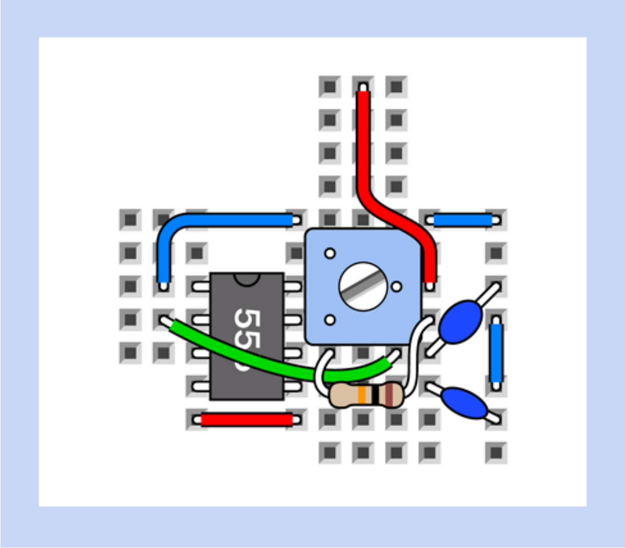

Figure 4-77 *How to squeeze in a trimmer to enable adjustment of your reflex tester.*

First move the blue wires up by one row. Snake the vertical red wire around the righthand side. Extend the lead from the remaining 10K resistor, taking care not to allow it to touch any other exposed conductors. You can now insert a trimmer, with its wiper pin making contact with the positive supply and another pin connecting with pin 7 of the timer. The third trimmer pin plugs into a vacant row on the breadboard, and can be ignored.

You should use a trimmer rated at 20K or 25K and start with it around the middle of its range. You now have three options to fine-tune the circuit so that it runs at 1kHz.

If you happen to have a multimeter with a setting on its dial to measure kHz, simply ground the black meter probe, touch the red probe to pin 3 of the first timer that you installed, and turn the trimmer until the meter shows 1kHz. Job done!

If you don't have a meter than measures frequency, perhaps you have a digital guitar tuner. They only cost a few dollars on eBay. Attach a loudspeaker to the output from the 555 timer (including a 10µF coupling capacitor and a 47-ohm series resistor), and the tuner should tell you the frequency of the note being generated by the timer.

If you don't have a suitable meter or guitar tuner, you can use any watch, clock, or phone that will display whole seconds. When the timer is running at 1kHz, the second counter will advance every 1/100th of a second, and the third counter will advance every 1/10th of a second. The third counter goes through 10 digits before repeating, which means it will display a zero once each second.

The problem is, each digit is illuminated so briefly, you'll have a hard time seeing precisely when the zero comes up. So here's what you do.

Cover all the segments of the slowest display except the segment at its bottom-right corner. This is illuminated all the time except when a 2 is displayed, at which point it blinks off. You'll have a much easier time counting the blinks of this single segment than if you try to recognize whole numbers. Adjust the trimmer potentiometer that you added, and gradually you should be able to synchronize the slowest display with a timekeeping device.

Enhancements

Whenever I finish a project, I see opportunities to improve it. Here are some ideas:

No counting at power-up. It would be nice if the circuit begins in its "ready" state, rather than already counting. I'm going to leave you to think about this.

Audible feedback when the red LED lights up. Not essential, but might be a nice feature.

A random delay interval before the count begins. Making electronic components behave randomly is very difficult, but one way to do it would be to require the user to hold his finger on a couple of metal contacts. The skin resistance of the finger would determine the delay. Because the finger pressure would not be exactly the same each time, the delay would vary.

Next?

A counter such as the 4026B is technically a logic chip. It contains *logic gates* that enable it to count. Every digital computer works on similar principles.

Because logic is so fundamental in electronics, I'm going to delve into it at much greater depth, beginning with the next experiment. The magic words AND, OR, NAND, NOR, XOR, and XNOR will open a whole new world of digital intrigue.

Experiment 20: Learning Logic

When you deal with logic gates individually, they're extremely easy to understand. When you chain them together, they become more challenging. So I'll start with them one at a time.

This experiment contains a significant amount of explanation. I don't expect you to remember it all. The purpose here is to provide a store of information that you can refer back to, later.

What You Will Need

- Breadboard, hookup wire, wire cutters, wire strippers, test leads, multimeter
- 9-volt power supply (battery or AC adapter)
- SPDT slide switch (1)
- 74HC00 quad 2-input NAND chip (1)
- 74HC08 quad 2-input AND chip (1)
- Low-current LEDs (2)
- Tactile switches (2)
- LM7805 voltage regulator (1)
- Resistors: 680 ohms (1), 2.2K (1), 10K (2)
- Capacitors: 0.1µF (1), 0.33µF (1)

The Regulator

Logic gates are much fussier than the 555 timer or the 4026B counter that you used previously. The versions that we will be using demand a precise 5VDC, with no fluctuations or "spikes" in the flow of current.

This is easy and inexpensive to achieve. Just set up your breadboard with an LM7805 voltage regulator. It delivers a well-controlled voltage of 5 volts if you supply it with a DC voltage of 7 volts or more.

For a diagram of the regulator showing its three pin functions, see Figure 4-78. For a schematic showing how to use the regulator, see Figure 4-79. For a suggestion of how to place the regulator and its two capacitors at the top of a breadboard, to occupy minimum space, see Figure 4-80. I have added a miniature on-off slide switch at top-left, and a low-current LED to show when power is connected. I think a visual indicator is useful to provide assurance that the power is on, especially when you're searching for a fault in a circuit. The high-value 2.2K resistor for the LED is chosen to consume as little current as possible, in case you are still using a 9-volt battery as your power supply.

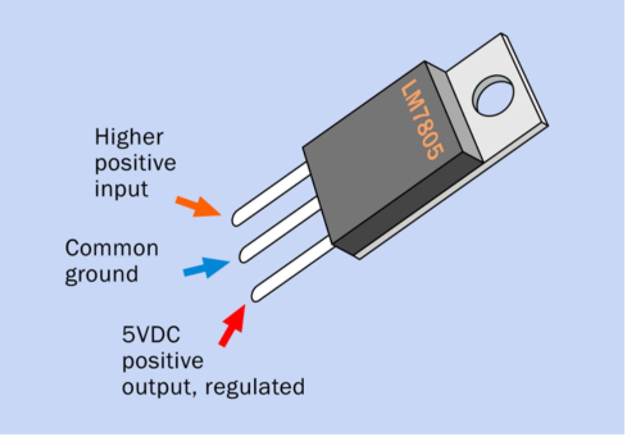

Figure 4-78 *Pin functions of the LM7805 voltage regulator, shown with the metal back facing away from you.*

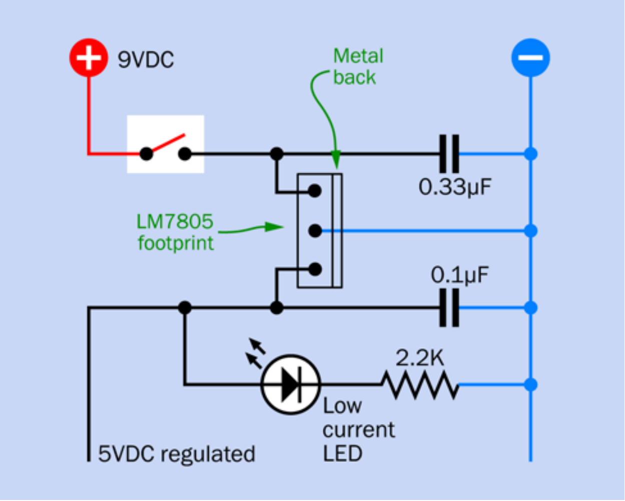

Figure 4-79 *How to use the LM7805 voltage regulator. The capacitors are mandatory.*

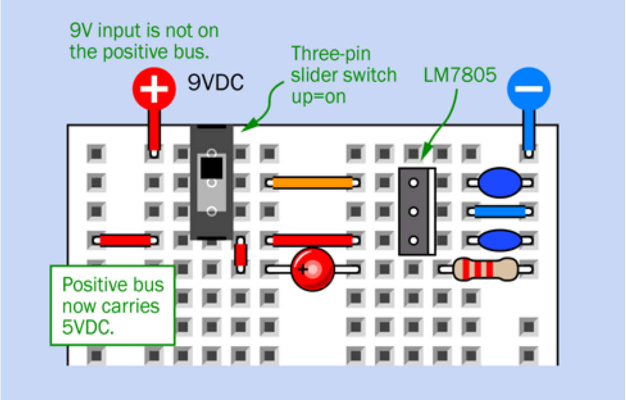
9V input is not on the positive bus.
9VDC
Three-pin slider switch up=on
LM7805
Positive bus now carries 5VDC.

Figure 4-80 *Placing a voltage regulator at the top of a breadboard, to occupy minimal room while also allowing for an on-off slide switch and a low-current LED to show that power is switched on.*

Caution: Inappropriate Inputs

DC, not AC. Remember that the LM7805 is a DC-to-DC converter. Do not confuse it with an AC adapter, which uses alternating current from an outlet in your home, and converts it to DC. Do not apply AC to the input of your voltage regulator.

Maximum Current. The LM7805 does a wonderful job of maintaining its output at an almost constant voltage, regardless of how much current you draw through it—so long as you stay within its rated range. Don't try to pull more than one amp through the voltage regulator.

Maximum Voltage. Although the voltage regulator is a solid-state device, it behaves a little like a resistor in that it radiates heat in the process of reducing a voltage. The higher the voltage you put into the regulator, and the more current that passes through it, the more heat it must get rid of. Theoretically you could use an input of 24VDC and still get a regulated output of 5VDC, but this would not be a good idea. A good input range would be 7VDC to 12VDC.

Minimum Voltage. Like all semiconductor devices, the regulator delivers a voltage that is lower than its input voltage. This is why I suggest a minimum input of 7VDC.

Heat Sinking. The purpose of the metal back with a hole in the top is to radiate heat, which it will do more effectively if you bolt it to a piece of aluminum, since aluminum conducts heat very effectively. The aluminum functions as a heat sink, and you can buy fancy ones that have multiple cooling fins. If you're not planning to draw more than 200mA through the regulator, a heat sink is unnecessary. The circuits in this book will require less than that.

Usage

When building circuits using 5-volt logic chips, you want 5VDC to be available down the positive bus of your breadboard. Note very carefully that the 9-volt input to the circuit in Figure 4-80 is *not* on the positive bus, but is applied only to the upper pin on the voltage regulator. The 5VDC output from the lower pin of the voltage regulator is connected with the positive bus.

The negative bus on the breadboard is shared by the voltage regulator and your external voltage supply. This is known as a "common ground."

After you install your voltage regulator, set your meter to measure DC volts and measure the voltage between the two buses on the breadboard, just to make sure. Logic chips will be easily damaged by incorrect or reversed voltage.

Your First Logic Gate

Now that you have your 5VDC breadboard ready, take a couple of tactile switches, two 10K resistors, a low-current LED, and a 680-ohm resistor, and set them around a 74HC00 logic chip as shown in Figure 4-81. (Because you are using a low-current LED, a 680-ohm resistor is appropriate.)

You may notice that many of the pins of the chip are shorted together and connected to the negative side of the power supply. I'll explain that in a moment.

When you connect power, the LED should light up. Press one of the tactile switches, and the LED remains illuminated. Press the other tactile switch, and again the LED stays on. Now press both switches, and the light should go out.

Pins 1 and 2 are logic inputs for the 74HC00 chip. By default, the circuit holds them at a low voltage, being connected to the negative side of the power supply through 10K pulldown resistors. But each pushbutton overrides its resistor and forces the input pin to rise near to the value of the 5V positive bus.

- When an input or output associated with a 5V logic chip is near 0VDC, we say it is *logic-low*.

- When an input or output associated with a 5V logic chip is near 5VDC, we say it is *logic-high*.

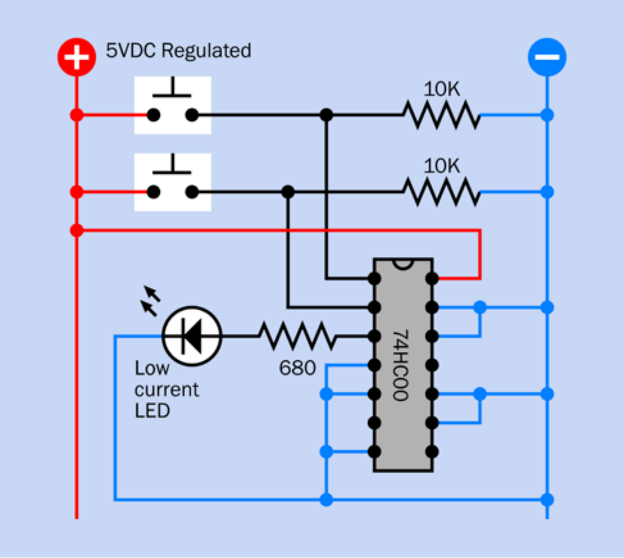

Figure 4-81 *Figuring out the logical function of the NAND gate.*

The logic output from the chip, as you saw, is normally high—but *not* if the first input *and* the second input are high. Because the chip does a “Not AND” operation, we say that it contains a *NAND logic gate*.

Logic gates can be represented with special symbols that are used in a kind of schematic known as a *logic diagram.* The logic diagram that corresponds with the circuit in Figure 4-81 is shown in Figure 4-82, where the U-shaped thing with a circle at the bottom is the logic symbol for a NAND gate. No power supply is shown for it in a logic diagram, but if you refer back to the schematic in Figure 4-81, the chip does in fact require power, at pins 7 (negative ground) and 14 (positive). This enables the chip to put out more current than its inputs take in.

- Any time you see a symbol for a logic chip, remember that it needs power to function.

The 74HC00 chip actually contains four separate NAND gates, each of which has two logical inputs and one output. They are arrayed as shown in Figure 4-83, in the diagram on the right. Because only one gate was needed for the simple test, the input pins of the unused gates were shorted to the negative side of the power supply, to stop them from floating.

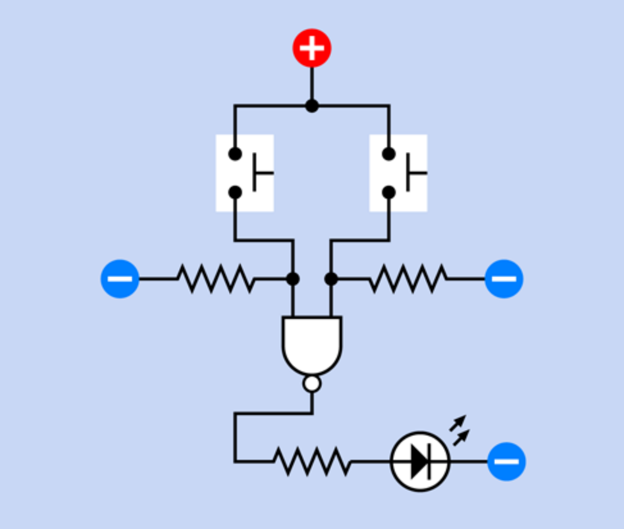

Figure 4-82 *A logic diagram can be easier to understand than a schematic showing a logic chip.*

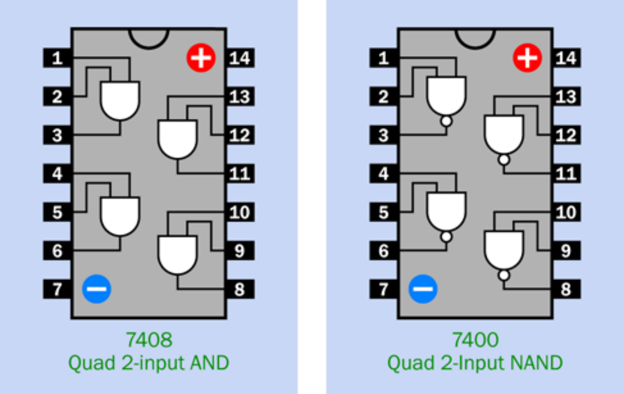

Figure 4-83 *The arrangement of gates inside two logic chips.*

Many logic chips are interchangable. In fact, let’s try that right now. First, disconnect the power. Carefully pull out the 74HC00 and put it away with its legs embedded in conductive foam (or aluminum foil, if you don’t have any foam). Substitute a 74HC08 chip, which is an AND chip. Make sure you have it the right way up, with its notch at the top. Reconnect the power and use the pushbuttons as you did before. This time, you should find that the LED comes on if the first input AND the second input are both positive, but it remains dark otherwise. Thus, the AND chip functions exactly opposite to the NAND chip. Its pinouts are shown in Figure 4-83, on the left.

You may be wondering why these things are useful. Soon you’ll see that we can put logic gates together to do things such as create an electronic combination lock, or a pair of electronic dice, or a computerized version of a TV quiz show where users compete to answer a question. If you were insanely ambitious, you could build an entire computer out of logic gates. A hobbyist named Bill Buzbee actually built a web server from vintage logic chips—see Figure 4-84.

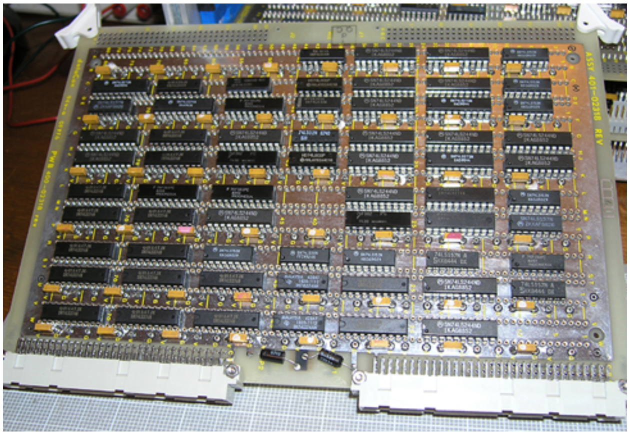

Figure 4-84 *This computer motherboard hand-made by Bill Buzbee is built from 74xx series logic chips, and functions as the hearts of a web server.*

Background: Logical Origins

George Boole was a British mathematician, born in 1815, who did something that few people are ever lucky enough or smart enough to do: he invented an entirely new branch of mathematics.

Interestingly, it was not based on numbers. Boole had a relentlessly logical mind, and he wanted to reduce the world to a series of true-or-false statements that could overlap in interesting ways.

Venn diagrams, conceived around 1880 by a man named John Venn, can be used to illustrate some logical relationships of this type. Figure 4-85 shows the simplest possible Venn diagram, where I have established one very large group (all the creatures in the world) and have defined a subgroup (consisting only of those creatures that live in the water). The Venn diagram illustrates that all creatures that live in water also live in the world, but only a subset of creatures in the world live in the water.

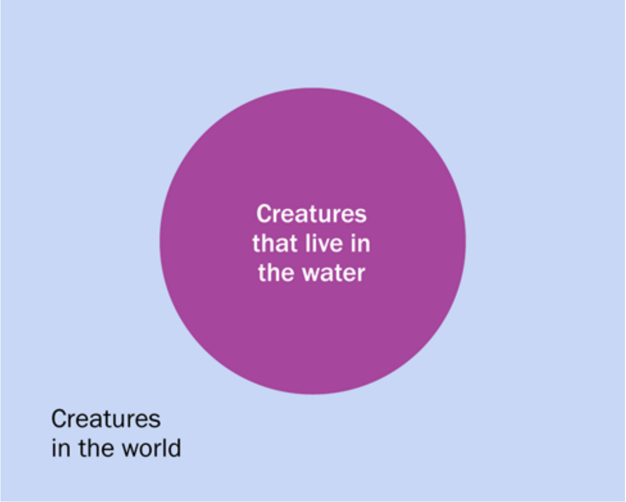

Figure 4-85 *The simplest possible relationship between a group and the larger world that contains it.*

Now I'll introduce another group: creatures that live on land. But wait—some creatures are able to live on both land and water. Frogs, for instance. These amphibians are members of both groups, and I can show this with another Venn diagram in Figure 4-86, where the groups overlap.

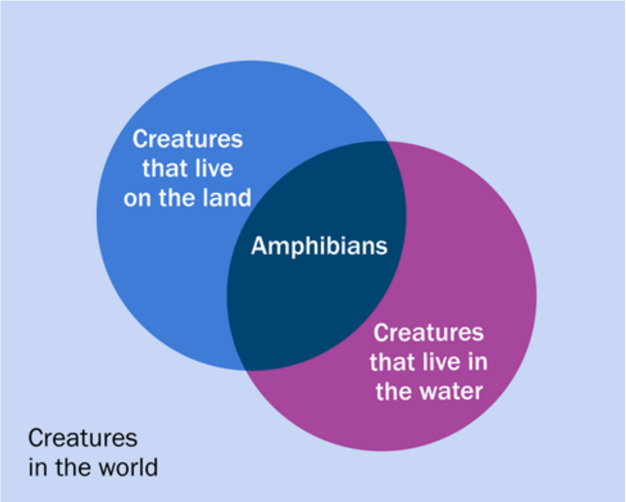

Figure 4-86 *This Venn diagram is a way of showing that some creatures in the world live on land, some live in water, and some inhabit both land and water.*

However, not all groups overlap. In Figure 4-87 I have created one group of creatures with hooves, and another group of creatures with claws. Are there any creatures that have hooves *and* claws? I don't think so. I could express this by creating a *truth table*, as in Figure 4-88.

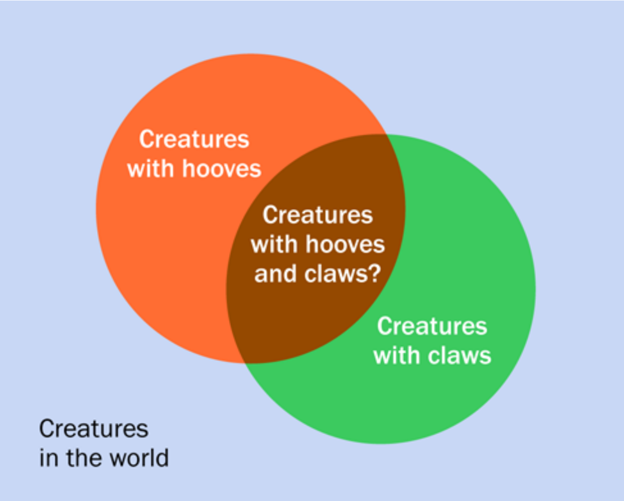

Figure 4-87 *Some subgroups do not overlap. I don't think any creatures have claws as well as hooves.*

This creature has hooves	This creature has claws	This combination can be
NO	NO	TRUE
NO	YES	TRUE
YES	NO	TRUE
YES	YES	FALSE

Figure 4-88 *The simplest form of a truth table tabulates the validity of pairs of inputs, each of which can have one of two states.*

A NAND gate could be used to represent this table, because its pattern of inputs and outputs is exactly the same, as shown in Figure 4-89.

If NAND input A is:	And if NAND input B is:	Then NAND output will be:
LOW	LOW	HIGH
LOW	HIGH	HIGH
HIGH	LOW	HIGH
HIGH	HIGH	LOW

Figure 4-89 *This truth table for a NAND gate has exactly the same pattern as the previous table.*

Beginning with these very simple concepts, Boole developed his logic language to a very high level. He published a treatise on it in 1854, long before it could be applied to electrical or electronic devices. In fact, during his lifetime, his work seemed to have no practical applications at all. But a man named Claude Shannon encountered Boolean logic while studying at MIT in the 1930s, and in 1938 he published a paper describing how Boolean analysis could be applied to circuits using relays. This had immediate practical applications, as telephone networks were growing rapidly, creating complicated switching problems.

A very simple telephone problem could be expressed like this. Long ago, it was common for two customers living in separate homes in a rural area to share one telephone line. If one of them wanted to use the line, or the other wanted to use it, or neither of them wanted to use it, there was no problem. But they could not both use it simultaneously. Once again, this is the same logical pattern as described in Figure 4-89, if we interpret "high" to mean that one person wants to use the line, and "low" meaning that the person doesn't want to use it.

But now there's an important difference. The NAND gate doesn't just illustrate the network. Because the telephone network uses electrical states, a NAND gate can *control* the network. (Actually, in the early days of networks, everything was done by relays; but an assembly of relays can still function as a logic gate.)

After Shannon's application of Boolean logic to the telephone system, the next step was to see that if you used an "on" condition to represent numeral 1 and an "off" condition to represent numeral 0, you could build a system of logic gates that could count. And if it could count, it could do arithmetic.

When vacuum tubes were substituted for relays, the first practical digital computers were built. Then transistors took the place of vacuum tubes, and integrated circuit chips replaced transistors, leading to the desktop computers that we now take for granted. But deep down, at the lowest levels of these incredibly complex devices, they still use the laws of logic discovered by George Boole.

Incidentally, when you use a search engine online, if you use the words AND and OR to refine your search, you're actually using *Boolean operators.*

Fundamentals: Logic Gate Basics

The NAND gate is the most fundamental building block of digital computers, because you can do addition sums using nothing but NANDs. If you want to learn more about this, search online for topics such as "binary arithmetic" and "half-adder." You can also find circuits that do addition sums using logical operators in my book *Make: More Electronics*.

Generally, there are seven types of logic gates:

`AND, NAND, OR, NOR, XOR, XNOR, NOT`

Their names are usually printed entirely in capital letters. Of the first six, XNOR is hardly ever used.

All of the gates have two inputs and one output, except for the NOT gate, which has only one input and one output, and is more often referred to as an *inverter*. If it has a high input, it gives a low output, and if it has a low input, it gives a high output.

Symbols that represent the seven types of gates are shown in Figure 4-90. Notice that the little circles at the bottom of some of the gates invert the output. (These circles are called *bubbles*.) Thus, the output of a NAND gate is the inverse of an AND gate.

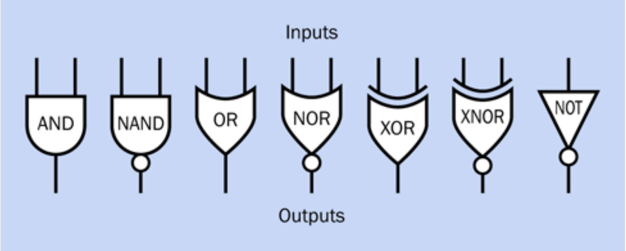

Figure 4-90 *Symbols for the six two-input logic gates, plus the NOT gate.*

What do I mean by "inverse"? This should become clear if you look at the truth tables for logic gates that I have drawn in Figure 4-91, Figure 4-92, and Figure 4-93. In each of the tables, two inputs are shown on the left followed by an output on the right, with red meaning a high-logic state and blue meaning a low-logic state. Compare the outputs of each pair of gates, and you'll see how the patterns are inverted.

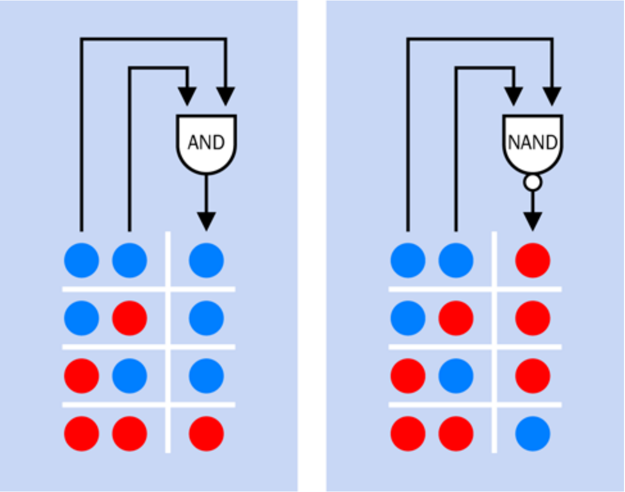

Figure 4-91 *The inputs shown on the left produce the outputs shown on the right.*

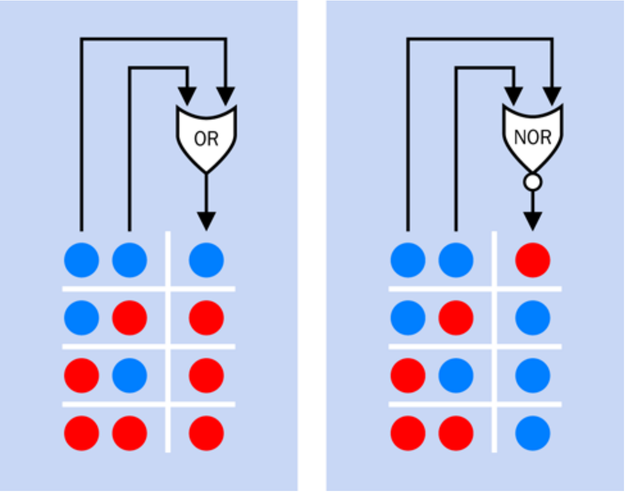

Figure 4-92 *The inputs shown on the left produce the outputs shown on the right.*

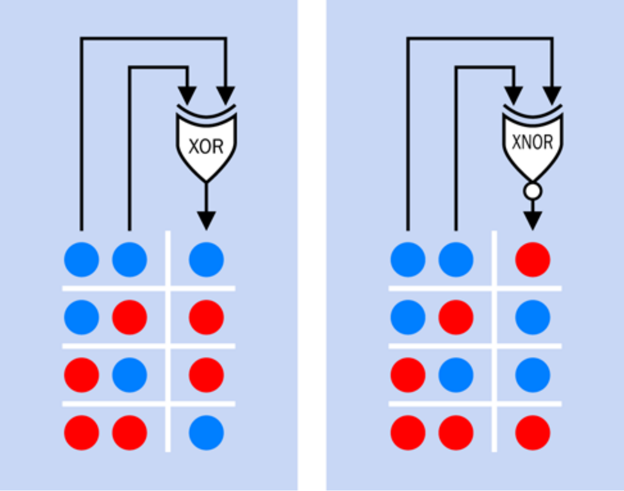

Figure 4-93 *The inputs shown on the left produce the outputs shown on the right.*

Background: The Confusing World of TTL and CMOS

Back in the 1960s, the first logic gates were built with transistor–transistor logic, abbreviated TTL, meaning that tiny bipolar transistors were etched into a single wafer of silicon. Soon, these were followed by complementary metal oxide semiconductors, abbreviated CMOS. The 4026B chip that you used in Experiment 19 is an old CMOS chip.

You may remember that bipolar transistors amplify current. Thus, TTL circuits require a significant flow of electricity to function. But CMOS chips are voltage-sensitive, enabling them to draw hardly any current while they are waiting for a signal, or while they are pausing after emitting a signal.

The table in Figure 4-94 summarizes the original advantages and disadvantages of the two types of chips. The CMOS series, with part numbers from 4000 upward, were slower and more easily damaged by static electricity, but were valuable because of their meager power consumption. The TTL series, with part numbers from 7400 upward, used much more power but were less sensitive and very fast. So, if you wanted to build a computer, you used the TTL family, but if you wanted to build a little gizmo that would run for weeks on a small battery, you used the CMOS family.

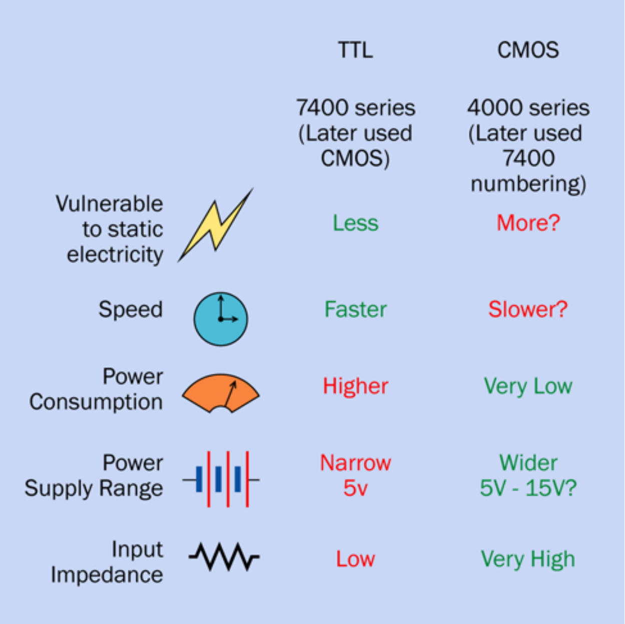

	TTL	CMOS
	7400 series (Later used CMOS)	4000 series (Later used 7400 numbering)
Vulnerable to static electricity	Less	More?
Speed	Faster	Slower?
Power Consumption	Higher	Very Low
Power Supply Range	Narrow 5v	Wider 5V - 15V?
Input Impedance	Low	Very High

Figure 4-94 *In this table comparing early CMOS and TTL chips, the CMOS attributes with a question mark were eventually reconciled with the TTL attributes.*

From this point on everything became confusing, because CMOS manufacturers wanted to grab market share by emulating the advantages of TTL chips. Newer generations of CMOS chips even changed their part numbers to begin with “74” to emphasize their compatibility, and the functions of pins on CMOS chips were swapped around to match the functions of pins on TTL chips. The voltage requirements of CMOS were also changed to match TTL.

Today, you can still find some of the old TTL chips around, especially in the LS series (with part numbers such as 74LS00 and 74LS08). However, they have become uncommon.

It’s more common to find the 4000 series CMOS chips—such as the 4026B that you used in the previous experiment. They are still being made because their wide range of acceptable power-supply voltages is useful.

Over the years, CMOS chips became faster and less vulnerable to static electricity, which explains why I added a question mark in these categories in Figure 4-94. Modern CMOS chips also mostly reduced their maximum power supply voltages to 5VDC—which is why I added a question mark in this category, too.

The situation can be summed up like this:

- Any logic chips still available in the old 4000 series will have the features I have listed in Figure 4-94. You are quite likely to find a use for the 4000 series of chips.
- You are unlikely to use old TTL chips in the 7400 series, because they have no significant advantages.

You may still find schematics that specify 74LSxx chips. You can substitute the 74HCTxx chips, which are designed to function identically.

The 74HCxx generation is by far the most popular in through-hole format. They have the high input impedance of CMOS, which is useful, and they are cheaper than some of the more modern, exotic versions. All of the logic chips in this book are of the HC type.

Now for the part numbering. Where you see a letter "x" in the list that follows, it means that various letters and numbers may appear in that location. Thus "74xx" includes the 7400 NAND gate, the 7402 NOR gate, the 74150 16-bit data selector, and so on. A combination of letters preceding the "74" identifies the chip manufacturer, while letters following the part number may identify the style of package, may indicate whether it contains heavy metals that are environmentally toxic, and other details. This is explained visually in Figure 4-3.

Here's the history of the TTL family:

- 74xx: The old original generation, now obsolete.
- 74Sxx: Higher speed "Schottky" series, now obsolete.
- 74LSxx: Lower power Schottky series, still used occasionally.

The CMOS family:

- 40xx: The old original generation, now obsolete.
- 40xxB: The 4000B series was improved but still susceptible to damage from static electricity. These chips are still commonly used, especially in hobby-electronics applications.
- 74HCxx: Higher-speed CMOS, with part numbers matching the TTL family, and pinouts matching the TTL family. I've used this generation extensively in this book, because it's widely available, and the circuits here have no need for greater speed or power.
- 74HCTxx: Like the HC series but matching the old TTL standard for maximum and minimum logic-low and logic-high voltages, respectively.
- 74xx series with other letters in the middle of the part number: More modern, faster, usually surface-mount, often designed for lower operating voltages.

What You Don't Need

Speed differences are irrelevant from our point of view, as we're not going to be building circuits running at millions of cycles per second.

Price differences between chip families are usually minor when purchased in low quantities.

Lower-voltage chips are not appropriate for our purposes, as they are almost all surface-mount format, and we would have to create a lower-voltage power supply. Because surface-mount chips are so much more difficult to deal with, and their only major advantage is miniaturization, I won't be using them. Through-hole equivalents have the same logic functions.

Fundamentals: Part Numbers and Functions

The interior connections of currently available through-hole 14-pin logic chips in the HC series are shown in Figure 4-83, Figure 4-95, Figure 4-96, Figure 4-97, Figure 4-98, Figure 4-99, Figure 4-100, and Figure 4-101.

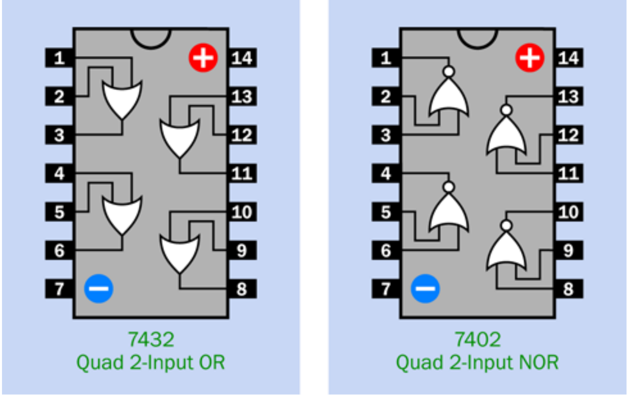
7432
Quad 2-Input OR
7402
Quad 2-Input NOR

Figure 4-95 *Standardized configuration of gates in the 74xx family of logic chips.*

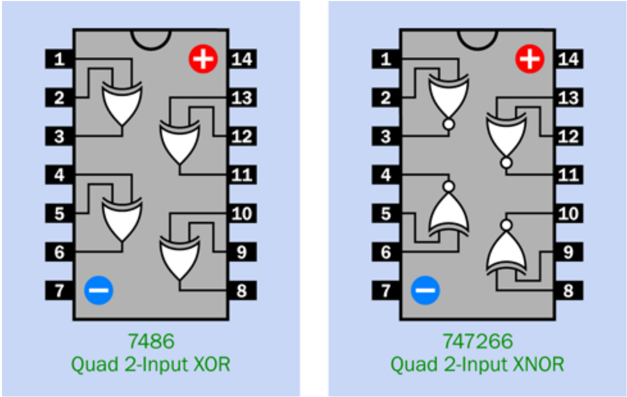
7486
Quad 2-Input XOR
747266
Quad 2-Input XNOR

Figure 4-96 *Standardized configuration of gates in the 74xx family of logic chips.*

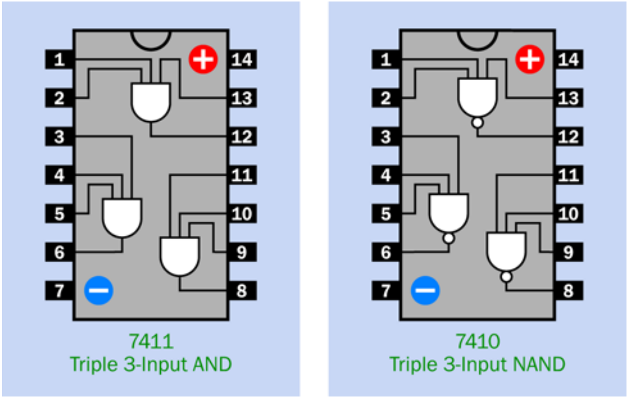
7411
Triple 3-Input AND
7410
Triple 3-Input NAND

Figure 4-97 *Standardized configuration of gates in the 74xx family of logic chips.*

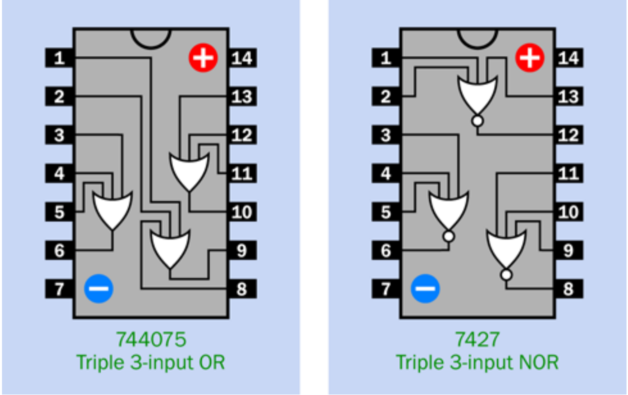
744075
Triple 3-input OR
7427
Triple 3-input NOR

Figure 4-98 *Standardized configuration of gates in the 74xx family of logic chips.*

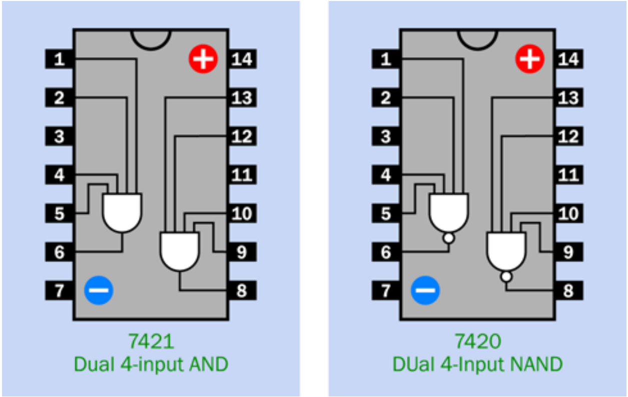
7421
Dual 4-input AND
7420
DUal 4-Input NAND

Figure 4-99 *Standardized configuration of gates in the 74xx family of logic chips.*

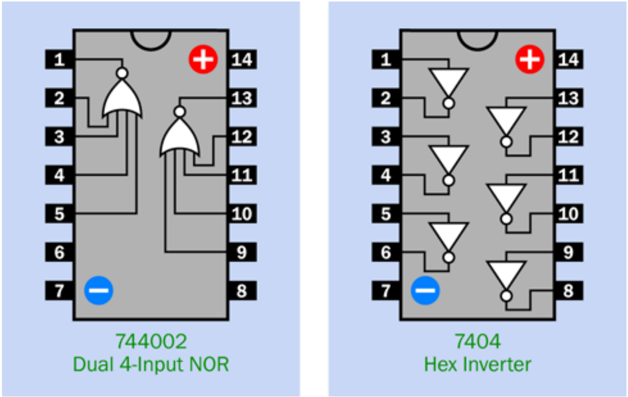
744002
Dual 4-Input NOR
7404
Hex Inverter

Figure 4-100 *Standardized configuration of gates in the 74xx family of logic chips.*

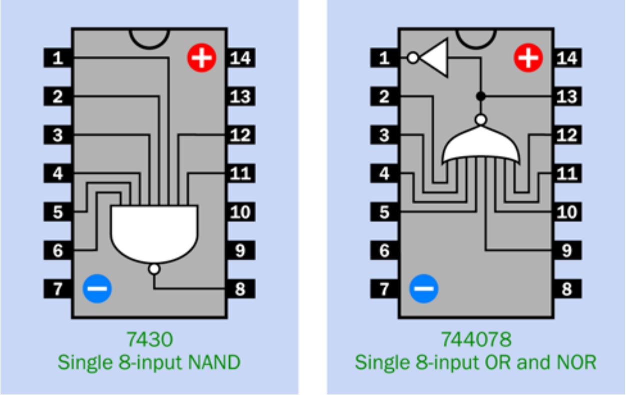
1
2
3
4
5
6
7
14
13
12
11
10
9
8
7430
Single 8-input NAND
1
2
3
4
5
6
7
14
13
12
11
10
9
8
744078
Single 8-input OR and NOR

Figure 4-101 *Standardized configuration of gates in the 74xx family of logic chips.*

All the part numbers in these chips are shown in their minimal form. Thus the 7400 chip may have actual part numbers such as 74HC00, 74HCT00, and so on, and will be preceded and followed by other letter codes; but it is generically referred to as a 7400 chip, so that's how I present it here.

It's very important to check the pin functions of logic chips against the diagrams here or a manufacturer's datasheet before using them. Interior connections may seem to follow similar patterns, but there are exceptions.

Fundamentals: Rules for Connecting Logic Gates

Permitted:

- You can connect the input of a gate directly to your regulated power supply, either positive side or negative side.
- You can connect the output from one gate directly to the input of another gate.
- The output from one gate can power the inputs of many other gates (this is known as "fanout"). The exact ratio depends on the chip, but in the 74HCxx series, you can always power at least ten inputs with one logic output.
- The output from a logic chip can drive the trigger (pin 2) of a 555 timer, so long as the timer shares the same 5VDC power supply.

- Low input doesn't have to be zero. A 74HCxx logic gate will recognize any voltage up to 1 volt as "low."
- High input doesn't have to be 5 volts. A 74HCxx logic gate will recognize any voltage above 3.5 volts as "high."

The acceptable ranges for inputs, and the minimum guarantees for outputs, are shown in Figure 4-102.

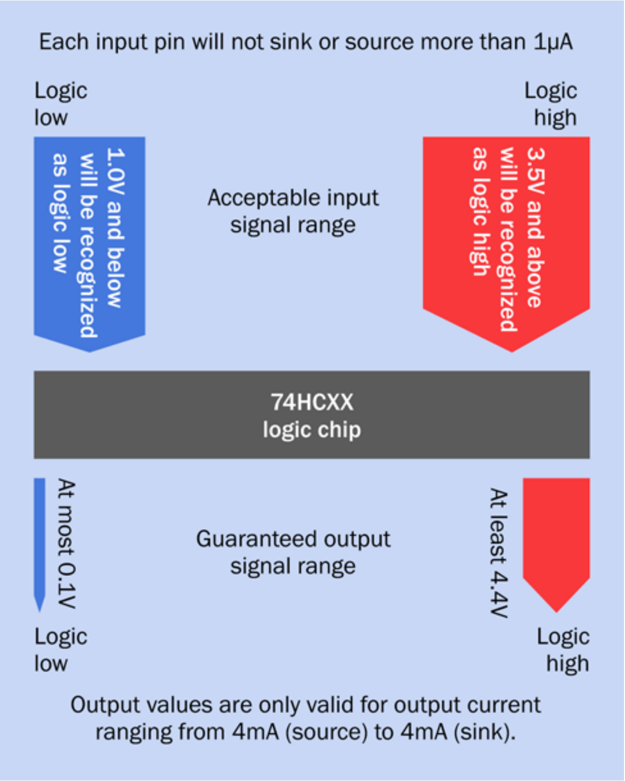
Each input pin will not sink or source more than 1μA
Logic low
Logic high
1.0V and below will be recognized as logic low
Acceptable input signal range
3.5V and above will be recognized as logic high
74HCXX logic chip
At most 0.1V
Guaranteed output signal range
At least 4.4V
Logic low
Logic high
Output values are only valid for output current ranging from 4mA (source) to 4mA (sink).

Figure 4-102 *To avoid errors, stay within the recommended input ranges for logic chips.*

Not permitted:

- No floating inputs! On CMOS chips, such as the HC family, you must always connect all input pins with a known voltage. This includes the input pins of gates that are unused on a chip.
- Any single-throw switch or pushbutton should be used with a pullup or pulldown resistor, so that when the contacts are open, the input to the chip is not floating.

- Don't use an unregulated power supply, or more than 5 volts, or less than 5 volts, to power 74HCxx logic gates.
- Be careful when using the output from a 74HCxx logic chip to power an LED. You can draw as much as 20mA from the chip, but this will pull down the output voltage. If you are also connecting that voltage to the input of a second chip, the voltage may be pulled down so much by the LED that the second chip doesn't recognize it as "high" anymore. Generally, try not to use a logic output to power an LED *at the same time as* another logic chip. Always check currents and voltages when modifying a circuit or designing a new one.
- Throughout this book, I am using low-current LEDs in conjunction with logic-chip outputs, because I think this is a good habit to acquire, in case an output powering an LED is also used as a logic input for another chip at some point in the future.
- Never apply a significant voltage or current to the output pin of a logic gate. In other words, *don't force an output into another output*.
- For this reason, don't tie the outputs together from two or more logic gates.

So much for the dos and don'ts. Now it's time for your first real logic-chip project.

Experiment 21: A Powerful Combination

Suppose you want to prevent other people from using your computer. I can think of two ways to do this: using software, or using hardware. The software would be some kind of startup program that intercepts the normal boot sequence and requests a password. It would be a little more secure than the password protection that is a standard feature of Windows and Mac operating systems.

You could certainly do it that way, but I think it would be more fun (and more relevant to this book) to do it with hardware. What I'm imagining is a numeric keypad requiring the user to enter a secret combination before the computer can be switched on. I'll call it a "combination lock," even though it won't actually lock anything. It will disable the power-on button that you normally use to switch on your computer.

Caution: The Warranty Issue

If you follow this project all the way to its conclusion, you'll open your desktop computer, cut a wire, and insert your own little circuit. You won't be going near any of the boards inside the computer, and you will only access the wire to the computer's "power on" button, but still, if you bought your computer new, you will void your warranty. I don't take this too seriously, myself, but if that makes you uneasy, here are three options:

- Breadboard the circuit for fun, and leave it at that.
- Use the circuit on some other device.
- Use it on an old computer.

What You Will Need

- Breadboard, hookup wire, wire cutters, wire strippers, multimeter
- 9-volt power supply (battery or AC adapter)
- Low-current LED (1)
- Generic LED (1)
- LM7805 voltage regulator (1)
- 74HC08 logic chip (1)
- 555 timer chip (1)
- 2N2222 transistor (1)
- DPDT 9VDC relay (1)
- Diodes: 1N4001 (1), 1N4148 (3)
- Resistors: 330 ohms (1), 470 ohms (1), 1K (1), 2.2K (1), 10K (6), 1M (1)
- Capacitors: 0.01µF (1), 0.1µF (1), 0.33µF (1), 10µF (2)
- Tactile switches (8)
- Optional: Tools to open your computer, drill four holes, and make saw cuts between the holes, to create a rectangular opening for the

keypad (if you want to take this project to its conclusion). Also, four small bolts to attach the keypad to the computer cabinet after you create the opening for it

A Three-Part Circuit

The entire breadboarded circuit is shown in Figure 4-107, but before you start to build it, let's take a look at the schematics.

The circuit is divided into three sections:

1. Power supply and three dummy pushbuttons.
2. Active pushbuttons and logic.
3. Output.

The schematic in Figure 4-103 shows part 1. This is simple enough. When you press button A, this supplies 9VDC to the voltage regulator, which delivers 5VDC down the lefthand bus. The button also supplies 9VDC down a magenta-colored wire on the right, for reasons which I'll get to in a moment.

In addition you will notice buttons B, C, and D, any one of which makes a connection with negative ground.

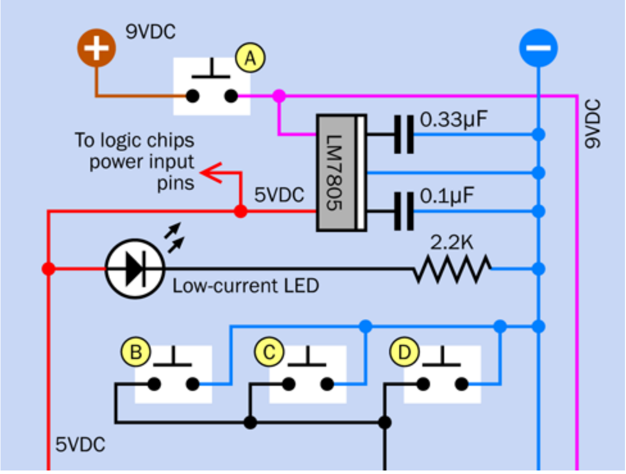

Figure 4-103 *The top section of the circuit.*

Now take a look at Figure 4-104, which shows the center section of the circuit using logic symbols. Imagine this attached to the top part of the circuit in Figure 4-103. Each of the buttons E through H can supply positive voltage to an AND gate, whose lefthand input is normally held low through a 10K pulldown resistor. The output from each gate feeds an input of the next gate.

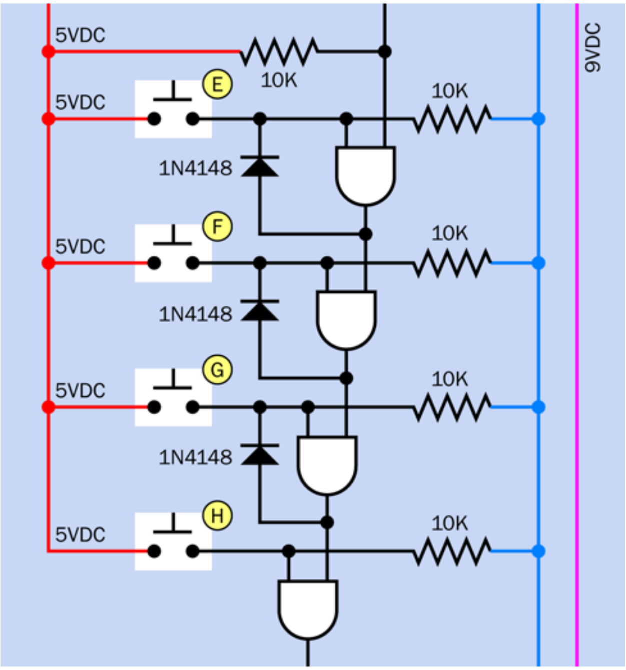

Figure 4-104 *The middle section of the circuit, using logic symbols.*

Lastly, Figure 4-105 shows the bottom section of the circuit, where the output from the last of the AND gates activates a transistor, which triggers a 555 timer. The timer controls a relay, and the relay will be used to lock and unlock your computer (or any other device that uses a simple on-off button).

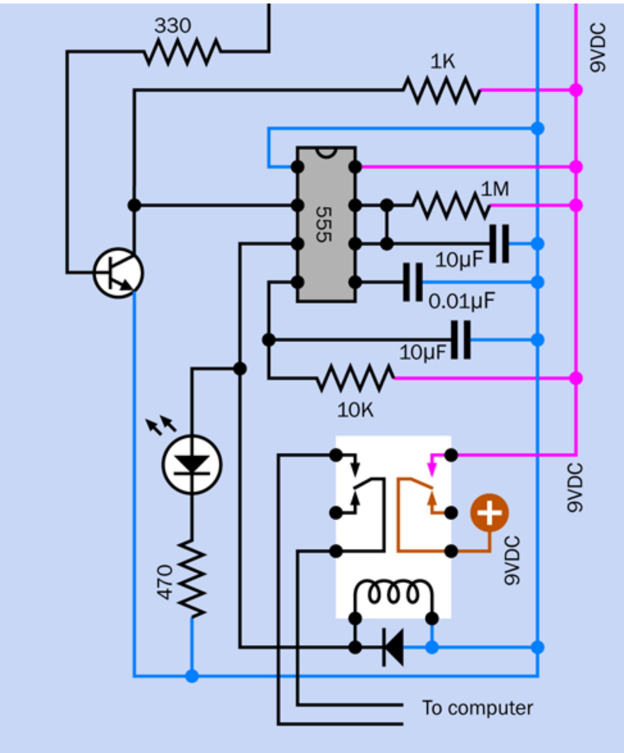

Figure 4-105 *The bottom section of the circuit.*

How It Works

I have set this up so that you have to hold down button A to activate the circuit, and you have to continue pressing it while you enter a secret sequence on the pushbuttons. This satisfies two purposes: the circuit draws no power while you're not using it, and you can't leave it switched on by mistake.

The secret sequence consists of pressing buttons E, F, G, and H in that order, while you are holding down button A. Of course if you actually install the circuit, you can scramble the positions of the buttons. I laid them out like this on the breadboard for the sake of simplicity.

Suppose you are holding down button A, and you press button E as the first in the sequence to unlock the circuit. In Figure 4-104 you can see that button E conducts 5 volts directly to the left input of the first AND gate. This overcomes the pulldown resistor, so the left input is now logic-high.

The righthand input of the AND gate is being held high through a 10K resistor. So, now both inputs of the AND gate are high, and consequently its output changes from low to high.

Current from the output circulates around through a diode, to the lefthand input. Consequently, you can let go of button E, and the output from the AND gate will maintain the lefthand input in a high state. The gate has latched itself, like the relay in Experiment 15. It is able to do this because it has its own power supply (not shown in the logic schematic) which sustains the output voltage regardless of small reductions in the input voltage.

The high output from the first AND gate also connects with the righthand input of the second AND gate. Now that the second AND gate has a high righthand input, if you press the button to pull up the lefthand input, the output from the second AND gate goes high. Note that the button wouldn't have worked before, because you needed a high output from the first AND gate to feed the second AND gate.

- After you press each button, the AND gate beside it latches itself, and you can release the button.
- You have to press the buttons in sequence. If you press four buttons out of sequence, nothing happens.
- You have to hold down button A throughout the whole procedure.

Now take a look at buttons C, D, and E. What happens if you press any of these buttons while you are entering the code to unlock the circuit? Any of the buttons will pull down the voltage on the righthand input of the first AND gate. Consequently, the output from the AND gate will go low. If the gate had latched itself, it comes unlatched. Moreover, the low output from the first AND gate will unlatch the second AND gate, and the low output from the second AND gate will unlatch the third AND gate.

Any of buttons C, D, and E will reset the whole circuit. I included them just to make entering the correct combination more difficult. Naturally I am assuming that if you install this system, all the buttons will look the same.

More than One Button?

What if someone presses more than one button simultaneously (while holding down button A)? The results will be unpredictable. If all the buttons E, F, G, and H are pressed together, the relay will be activated—except that if any of the buttons B, C, or D is also pressed, this will stop anything from happening. The possibility of simultaneous button presses can be seen as a flaw in this circuit, but the chance of someone pressing A, E, F, G, and H while not pressing B, C, or D is small. To reduce this risk further, you could add more "reset" buttons in parallel with B, C, and D.

Triggering the Relay

Suppose you enter the correct combination. The last AND gate applies about 5 volts to the base of transistor in Figure 4-105. The transistor turns on and starts conducting. This reduces the resistance between pin 2 of the 555 timer and negative ground, so the voltage on pin 2 is pulled down, and the timer is triggered.

The timer is being powered by 9 volts, which reaches it down the magenta-colored wire on the right. The timer output should be sufficient to activate the relay. Now see what the relay does: its righthand set of contacts provides an alternate source of power to the 9-volt bus.

So long as the pulse from the 555 timer continues, it holds the relay contacts closed. So long as the relay is closed, it powers the circuit—including the timer. Yes, the timer is energizing the relay, and the relay is powering the timer.

You can let go of button A now, and the relay will remain latched, so long as the pulse from the timer lasts. The pulse will end after about 30 seconds, cutting off power to the relay, so its contacts open. This turns off the timer, and also the rest of the circuit. The circuit now consumes no power at all.

The set of contacts on the left side of the relay are intended to provide power for the "on" button on your computer. So, during the brief time while the timer energizes the relay, you'll be able to switch on your computer. The rest of the time, the "on" button won't work.

The Logic Chip

Now take a look at Figure 4-106. This is the center section of the circuit, redrawn with an actual 74HC08 logic chip containing four two-input AND gates. It performs exactly the same function as the logic schematic in Figure 4-104. You can compare the two schematics to see how their functions are identical. The big difference is that the circuit showing the chip tells you how the component actually has to be installed—but, you are likely to find this diagram much harder to understand. Logic diagrams have their uses.

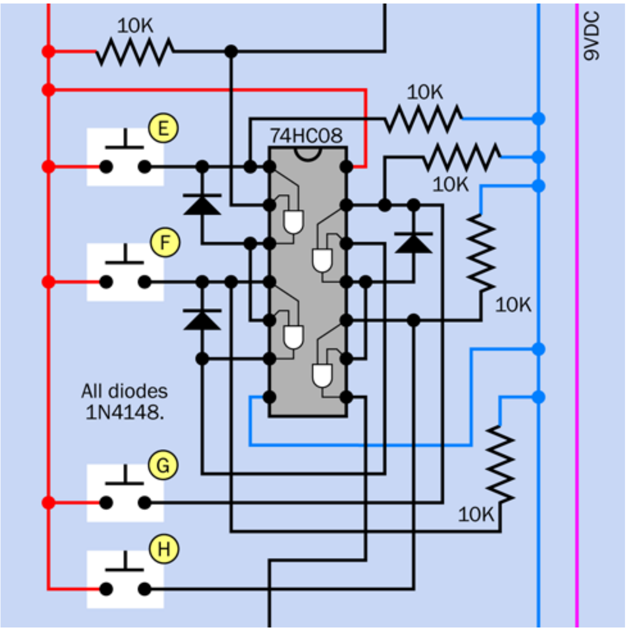

Figure 4-106 *The middle section of the circuit, showing actual components.*

Time to Build!

The fully breadboarded circuit is shown in Figure 4-107. For this project, you cannot really test it in stages; you have to build the whole thing. Component values are shown in Figure 4-108.

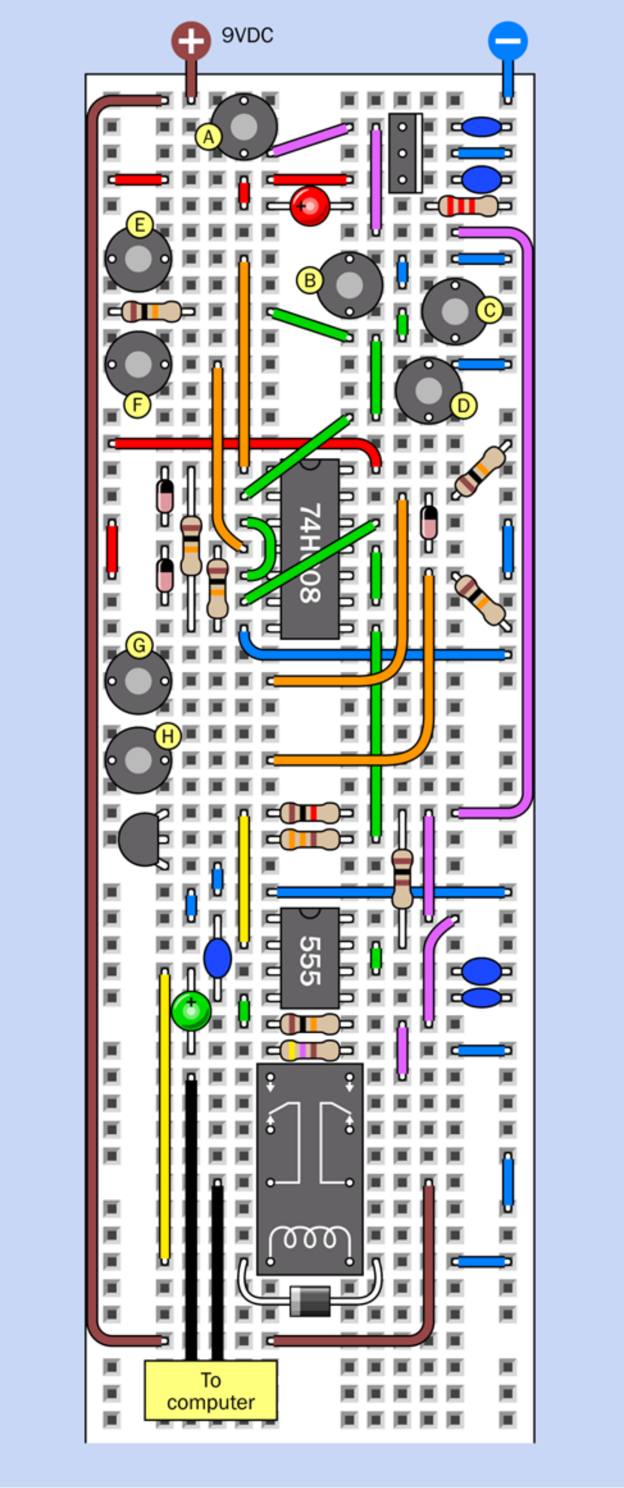
9VDC
A
B
C
D
E
F
G
H
74HC08
555
To computer

Figure 4-107 *The complete breadboarded circuit for the electronic combination lock.*

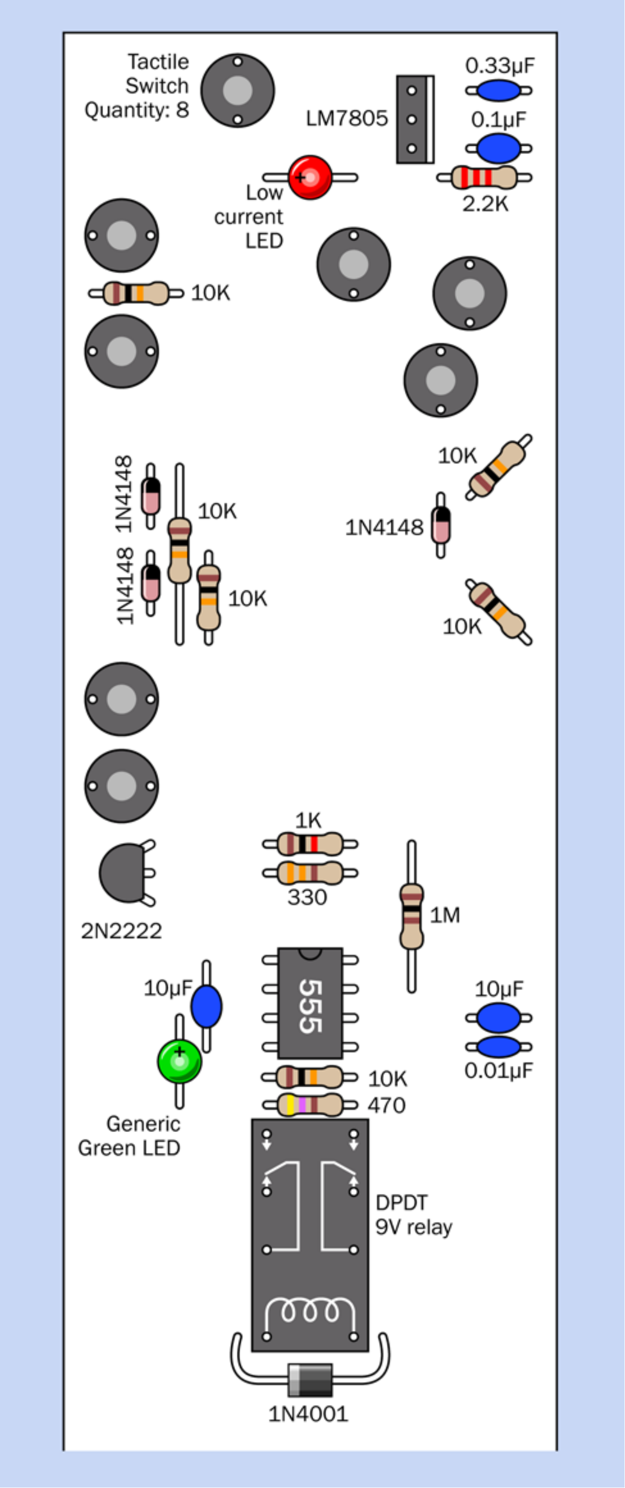
Tactile Switch Quantity: 8
0.33µF
LM7805
0.1µF
Low current LED
2.2K
10K
1N4148
10K
1N4148
10K
10K
1N4148
10K
1K
330
1M
2N2222
10µF
555
10µF
0.01µF
10K
470
Generic Green LED
DPDT 9V relay
1N4001

Figure 4-108 *The component values for the breadboard layout.*

Setup

Be careful to keep the two voltages in this circuit separate from each other. The 5VDC supply is insufficient to operate the relay, but 9VDC will burn out your logic chip. The lefthand bus on the breadboard is for the 5VDC supply. Unswitched 9VDC is taken down to the relay by the brown wire on the left of the breadboard in Figure 4-107. The purple or magenta wires on the right carry 9VDC, which is switched on either by button A or by the righthand relay contacts.

- Brown indicates the unswitched 9VDC supply from a battery or AC adapter.
- Magenta or purple is the 9VDC supply, which is switched, either by the relay or the button labelled A.
- Red is 5VDC, delivered by the voltage regulator.

After you build the circuit, attach your 9VDC power supply and hold down button A. The red LED comes on, but nothing else happens.

While you continue to hold down button A, press and release each of the buttons labeled E, F, G, and H in succession, from top to bottom. When you complete the sequence, the green LED comes on, showing that the relay has closed and you successfully unlocked the circuit.

Let go of button A, and the LED should stay on for about 30 seconds before the circuit automatically switches itself off. During this 30-second period, if this circuit is installed in your computer, you will have the opportunity to start the computer.

After the circuit switches itself off, it consumes no power at all. You can run it with a 9-volt battery, and the battery should literally last for years.

Try holding the power button again, and press the same buttons in a different sequence. Also, try including some of the buttons labeled B, C, and D. The green LED will not light up, and the relay will not be activated.

Suppose you install a finished version of this circuit. To crack the code, someone will have to know:

- You have to hold down button A while entering the correct sequence.
- If you press an incorrect button, the code must be re-entered from the start.
- Only buttons E, F, G, and H are active, and they must be pressed in that order.

That seems a very secure arrangement, to me. But if you want even more security, you can always add more buttons!

Testing

Set your meter to measure continuity, and attach its probes (using alligator test leads) to the lefthand output from the relay that is labeled “To Computer” in Figure 4-105. These two wires do not carry any voltage, so you need your meter in continuity-test mode to verify that the contacts inside the relay have closed.

Enter the correct combination of buttons, and the meter should beep. Let go of button A, and the meter should continue to beep while the 555 timer energizes the relay. At the end of the timer cycle, the relay opens and the meter stops beeping.

You can reconfigure your meter to measure current and insert it between the positive side of your battery and the 9VDC supply entry point on the breadboard. The meter should show no power consumption until button A is pressed.

Dealing with Diodes

This circuit contains two types of latching. The system for latching the relay is unusual but satisfies the requirement that the circuit should consume zero current when it is not being used. The system by which the AND gates latch themselves is another matter.

The fourth AND gate doesn’t need to latch, because only a brief pulse (from button H) is needed to start the timer. But the first three AND gates have to latch, to keep their outputs positive after you release each of the buttons E, F, and G. The diodes take care of this by feeding current from gate outputs back to their inputs.

Can you see any problem with this? Remember that a diode takes about 0.7 volts. Remember that a logic gate has to distinguish clearly between a high state and a low state on its inputs. If you start to sprinkle diodes through a logic circuit without keeping careful track of voltages, you may end up with a logic gate that doesn’t recognize an input that is supposed to be high. This is the same issue that I was concerned about in Experiment 15, where a voltage reduced by a transistor followed by a diode might have failed to trigger a relay.

When in doubt, verify voltages with your meter and check back to the input specifications illustrated in Figure 4-102.

In the combination-lock circuit that I've shown here, the output from each of the first three AND gates only circulates back through one diode to the input of the gate, so there's no reason why this should not work reliably. Just bear in mind that you have to exercise care and discretion when mixing diodes with logic chips.

Perhaps this makes you wonder—if a diode isn't the formally correct method to make a logic gate latch itself, what *is* the ideal way to do things?

One option might seem to be, replace each diode with a piece of wire, to feed the signal back to the gate input. What are the diodes really for, anyway?

They serve an important purpose. If a diode was replaced with a piece of wire, the positive voltage being applied through a pushbutton could also flow through that wire, down and around to the output of the logic gate.

- It is absolutely not a good idea to push voltage into any output of a gate.

The correct option for latching a logic state in a circuit is with a flip-flop. Previously I have used a 555 timer wired in bistable mode as a flip-flop, because we were already working with timers, and I wanted to demonstrate this application. But in this circuit, it wouldn't make sense to add four 555 timers just to serve that function. You can buy chips containing several flip-flops, and you can also make a flip-flop by combining two NAND gates or two NOR gates, as I will show you in Experiment 22.

For this little circuit creating a combination lock, I wanted to minimize the chip count and the complexity. Diodes were the simplest, easiest way to achieve this.

Questions

The output from the fourth AND gate was a single positive pulse. Why didn't I use that to activate the relay directly, instead of introducing a timer?

One reason is that the relay draws an initial surge of current exceeding the maximum 20mA that an AND gate can supply. Also of course I wanted a fixed-length pulse from the timer.

All right, but why did I add a transistor to the circuit? Because the AND gate gives a positive pulse, and the timer needs a negative transition on its trigger pin. The transistor provided a way to convert a positive to a negative. Adding a NOT gate (that is, an inverter) could have achieved the same goal, but would have increased the chip count.

In that case, why didn't I use NAND gates instead of AND gates? A NAND has a normally positive output that goes low when both of its inputs are high. That seems to be exactly what the 555 timer wants. With a NAND gate, I could have omitted the transistor.

This is true, but the preceding AND gates require positive outputs to feed back and sustain their positive inputs. So, I must retain those AND gates for the first three pushbuttons. I could only substitute a NAND gate for the last one, to create the right output for the timer. This means you would have still needed the 74HC08 chip, and you would have had to add a 74HC00 chip, just to use one of its gates. A transistor is easier and takes less room.

Here's another question. Why did I include two LEDs in the circuit? Because when you're punching buttons to unlock your computer, you need to know what's going on. The Power On LED reassures you that your battery isn't dead. The Relay Active LED tells you that the system is now unlocked, in case you are unable to hear the relay click.

Lastly, the big question: how can you actually install this circuit in a computer, assuming you are willing to give that a try? It's a lot easier than it sounds, as I will explain.

The Computer Interface

First, make sure you have wired the combination lock circuit correctly. A single wiring error can cause your circuit to deliver 9VDC through the lefthand relay contacts instead of merely closing a switch. This is important!

To make absolutely sure, change your meter to measure volts DC, and enter the correct combination on the tactile switches. If the green LED lights up but your meter measures no voltage, this is good. Anything else, and you have a wiring error.

Now let's consider how your computer normally functions when you want to switch it on.

Old computers used to have a big switch at the back, attached to the heavy metal box inside the computer that transformed house current to regulated DC voltages that a computer needs. Most modern computers are not designed this way; you leave the computer plugged in, and you touch a little button on the front of the computer (if it's not a Mac) or the keyboard (if it is a Mac). The button is connected by an internal wire to the motherboard.

This is ideal from our point of view, because we don't have to mess with high voltages. Inside your computer, don't even think of opening that metal box with the fan mounted in it, containing the computer power supply. Just look for the wire that runs from the "power up" button to the motherboard. Most often this wire will contain just two conductors, but on some computers it will be part of a ribbon cable. The key is to look at the contacts for the pushbutton, which will be attached to the conductors that you want.

First *make sure that your computer is unplugged*, ground yourself (because computers contain CMOS chips that are sensitive to static electricity), and very carefully snip just one of the two conductors from the pushbutton. Now plug in your computer and try to use the "power up" button. If nothing happens, you've probably cut the right wire. Even if you cut the wrong wire, it still prevented your computer from booting, which is what you want, so you can use it anyway.

Remember, we are not going to introduce any voltage to this wire. We're just going to use the relay as a switch to reconnect the conductor that you cut. You should have no problem if you maintain a cool and calm demeanor, and look for that single wire that starts everything. Check online for the maintenance manual for your computer if you're really concerned about making an error.

After you find the wire and cut just one of its conductors, keep your computer unplugged during the next steps.

Find where the wire attaches to the motherboard. Usually there's a small unpluggable connector. First, mark it so that you know how to plug it back in the right way around. Better still, photograph it. Then disconnect it while you follow the next couple of steps.

Strip insulation from the two ends of the wire that you cut, and solder an additional piece of two-conductor wire, as shown in Figure 4-109, with heat-shrink tubing to protect the solder joints. (This is very important!)

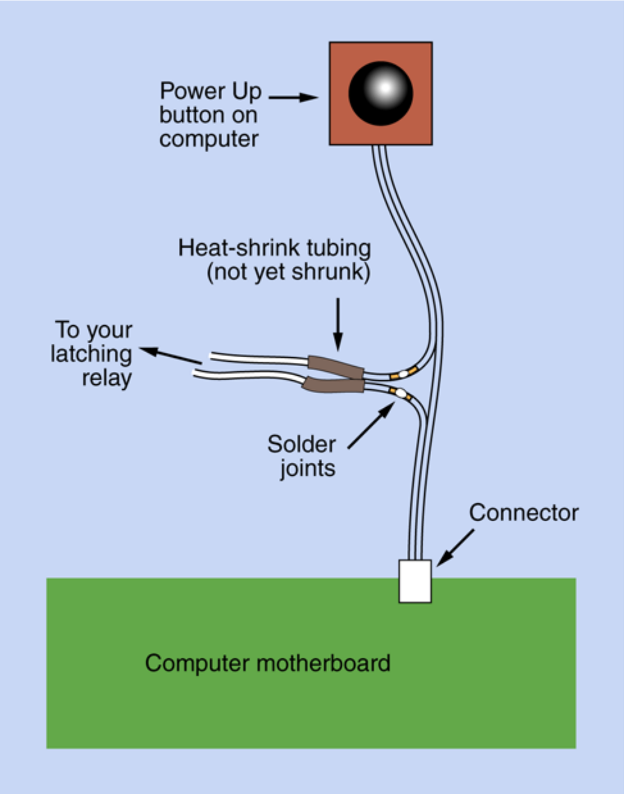

Figure 4-109 *The combination lock project can be interfaced with a typical desktop computer by cutting one conductor in the wire from the "power up" pushbutton, soldering an extension, and covering the joints with heat-shrink tubing.*

Run your new piece of wire to the relay, making sure you attach it to the pair of contacts that close, inside the relay, when it is energized by the unlocking operation. You don't want to make the mistake of unlocking your computer when you think you're locking it, and vice versa.

Reconnect the connector that you disconnected from your motherboard, plug in your computer, and press the computer's "start" button. If nothing happens, this is good! Now enter the secret combination on your keypad (while holding down the power button to provide battery power) and watch for the green LED to light up. Now try the "start" button again, and everything should work—so long as you press the button within the 30 seconds allowed by your circuit.

Having tested your circuit, the only remaining task is installation. Just remember to remove the case completely from the rest of the computer, if you are contemplating something along the lines shown in Figure 4-110.

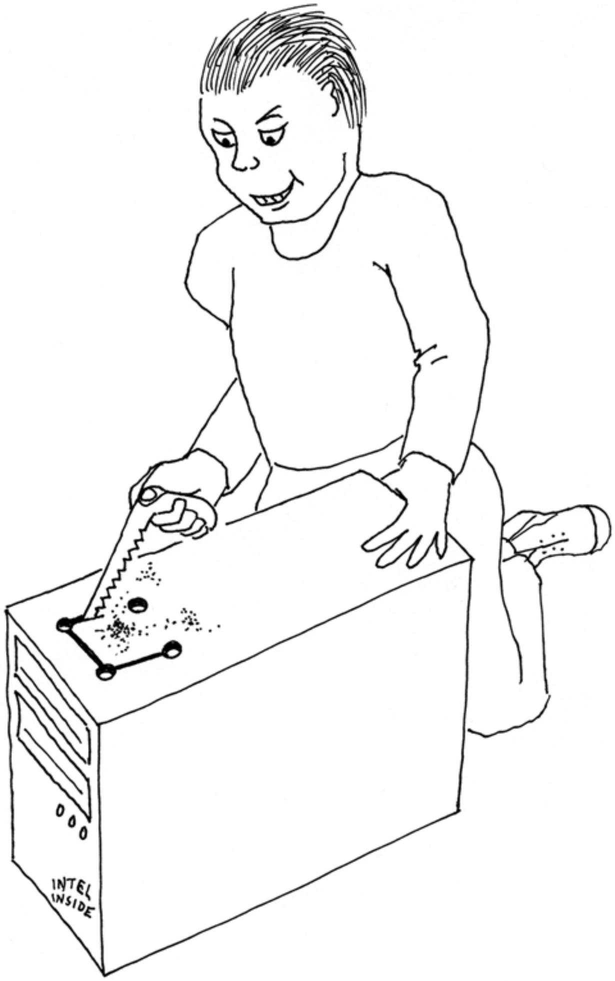

Figure 4-110 *An option for installing your keypad (not necessarily recommended).*

Enhancements

At the end of any project, there's always more you can do.

Using a Keypad. In the previous edition of this book, I suggested using a numeric keypad in this project. Some people felt that the keypad cost too much, and others had difficulty finding the right kind of keypad. After some thought, I decided just to use tactile switches this time around. They're easy to install on your breadboard, and if you decide to make a more permanent version of the circuit, you can simply mount eight pushbuttons on a square of metal or plastic. But a keypad is still an option, so long as it is *not matrix encoded*. That kind of keypad is really designed for use with a microcontroller. The kind you want will have as many pins as there are buttons, plus one.

Powering the Relay. You may be wondering if the voltage from the output of a 555 timer is sufficient to work the relay reliably. This is the same question that I discussed in Experiment 15, where I decided not to power a relay through a transistor-diode combination. The problem is, the voltage from a 555 timer varies depending on how heavily you load the output. This is why I recommended a high-sensitivity relay for this experiment. Typically it will draw less than one-third of the current of a standard type, which I felt was satisfactory for demonstration purposes. Bear in mind, I wanted to specify just one type of relay for all the experiments in this book. However, if you plan to install the circuit, and it absolutely positively has to work every time, even when your 9V battery is running low—you could consider substituting a 6VDC relay. Really? Won't the timer output overload it? Not necessarily. Some relays are designed to tolerate overvoltage. For instance, the datasheet for the Omron G5V-2-H1-DC6 6-volt relay allows a maximum voltage of 180% of the rated voltage. As always, the best recommendation is to test a circuit thoroughly, consider options, and read datasheets.

Safeguarding the Computer. To make this project more secure, you could remove the usual screws that secure the case of the computer, and replace them with tamper-proof screws. Naturally, you will also need the special tool that fits the screws, so that you can install them (or remove them, if your security system malfunctions for any reason).

Code Update. Another enhancement would be a way to facilitate changing your secret code if you feel the need. This will be difficult if you make a soldered version of the circuit, but you can install miniature plugs and sockets known as "headers" to allow you to swap the wires around.

Destructive Security. For those who are absolutely, positively, totally paranoid, you could fix things so that entering a wrong code flips a second high-amperage relay which supplies a massive power overload, melting your CPU and sending a big pulse through the hard drive. For a solid-state drive, you could consider installing a "suicide relay" that will apply a higher voltage to the 5VDC input. But I wouldn't recommend this myself.

There's no doubt about it. Messing up the hardware has major advantages compared with trying to erase data using software. It's faster, difficult to stop, and tends to be permanent. So, when the Recording Industry Association of America comes to your home and asks to switch on your computer so that they can search for illegal file sharing, just accidentally give them an incorrect unlocking code, sit back, and wait for the pungent smell of melting insulation—or a burst of gamma rays, if you go for the nuclear option (see Figure 4-111).

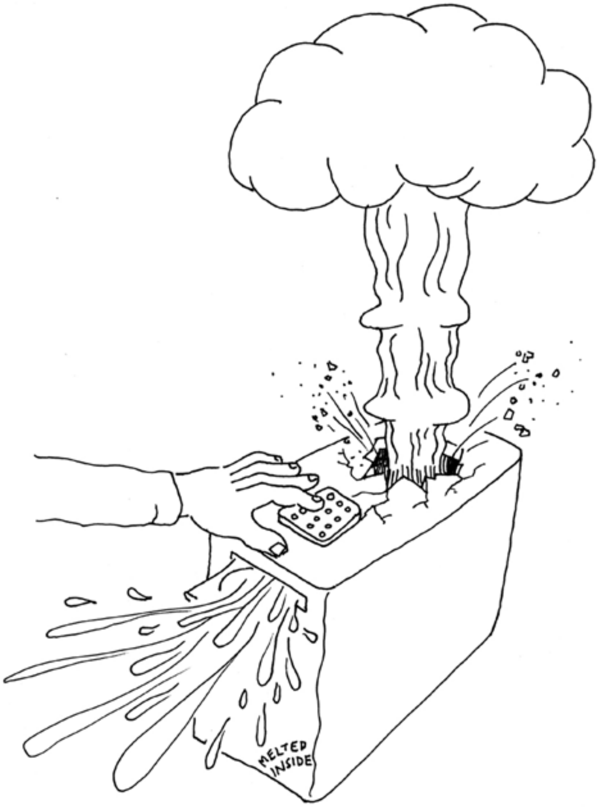

Figure 4-111 *For those who are absolutely, positively, totally paranoid: a meltdown/self-destruct system controlled by a secret key combination provides enhanced protection against data theft or intrusions by RIAA investigators asking annoying questions about file sharing.*

On a more realistic level, no system is totally secure. The value of a hardware locking device is that if someone does defeat it (for instance, by figuring out how to unscrew your tamper-proof screws, or simply ripping your keypad out of the computer case with metal shears), at least you'll know that something happened—especially if you put little dabs of paint over the screws to reveal whether they've been messed with. By comparison, if you use password-protection software and someone defeats it, you may never know that your system has been compromised.

Experiment 22: Race to Place

The next project using digital logic is going to get us into the concept of feedback, where the output is piped back to affect the input—in this case, blocking it. It's a small project, but quite subtle, and the concepts will be useful to you in the future.

What You Will Need

- Breadboard, hookup wire, wire cutters, wire strippers, multimeter
- 9-volt power supply (battery or AC adapter)
- 74HC32 logic chip (1)
- 555 timers (2)
- SPDT slide switches (2)
- Tactile switches (2)
- Resistors: 220 ohms (1), 2.2K (1), 10K (3)
- Capacitors: 0.01µF (2), 0.1µF (1), 0.33µF (1)
- LM7805 voltage regulator (1)
- Generic LEDs (2)
- Low-current LED (1)

The Goal

On quiz shows such as *Jeopardy!*, contestants race to answer each question. The first person who hits an answer button automatically locks out the other contestants, so that their buttons become inactive. How can we make a circuit that will do the same thing?

If you search online, you'll find several hobby sites where other people have suggested circuits to work this way, but they lack features that I think are necessary. The approach I'm going to use here is both simpler and more elaborate. It's simpler because it has a very low chip count, but it's more elaborate in that it incorporates "quizmaster control" to make a more realistic game.

I'll suggest some initial ideas for a two-player version. After I develop that idea, I'll show how it could be expanded to four or even more players.

A Conceptual Experiment

I want to show how this kind of project grows from an initial idea. By going through the steps of developing a circuit, I'm hoping I may inspire you to develop ideas of your own in the future, which is much more valuable than just replicating someone else's work.

First consider the basic concept: there are two people, each with a button to press, and whoever goes first locks out the other person.

Sometimes it helps me to visualize this kind of thing if I draw a sketch, so that's where I'll begin. In Figure 4-112, the signal from each button passes through an imaginary component that I'll call a "button blocker," activated by the other person's button. I'm not exactly sure what the button blocker will be or how it will work, yet, but it will be activated by one player going first, and it will block the other player.

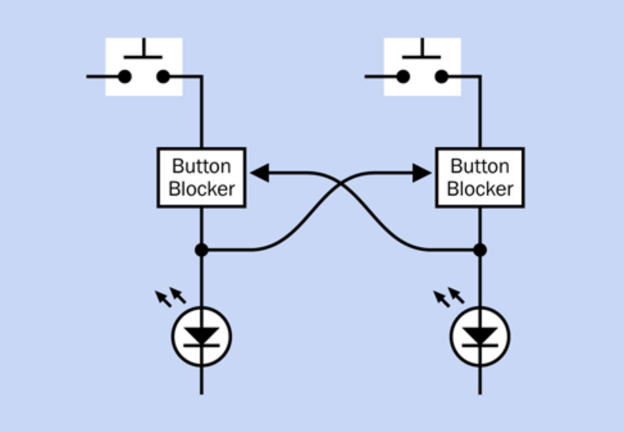

Figure 4-112 *The basic concept: whoever goes first blocks the other player.*

Now that I'm looking at it, I see a problem here. If I want to expand this to three players, it will get complicated, because each player must activate the "button blockers" of two opponents, and if there are four players, each much activate the "button blockers" of three opponents. The number of connections will become unmanageable. Figure 4-113 shows this.

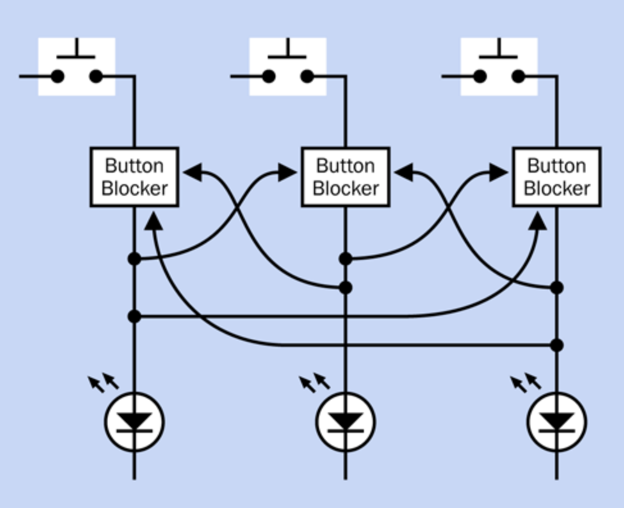

Figure 4-113 *Increasing the number of participants from two to three more than doubles the number of connections.*

Any time I see this kind of complexity, I think there has to be a better way.

Also, there's another problem. After a player lets his finger off the button, the other players' buttons will be unblocked again. This suggests to me that as in Experiments 15, 19, and 21, I need . . . a flip-flop, also known as a latch. Its purpose will be to hold the signal from the first player's button and continue to block the other players, even after the first player has released his button.

This now sounds even more complicated. But wait a minute. If the winning player's button triggers a latch, the latch now keeps the winning circuit energized, and the winner's button becomes irrelevant. So, the latch can block *all* the buttons. This makes things much simpler. I can summarize it as a sequence of events:

- First player presses his button.
- His signal is latched.
- The latched signal feeds back and blocks all the buttons.

The new sketch in Figure 4-114 shows this. Now the configuration is modular, and can be expanded to almost any number of players, just by adding more modules, without increasing in complexity.

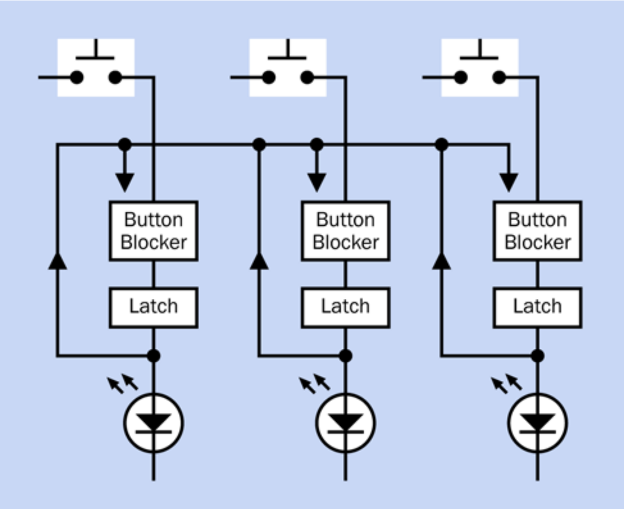
Button Blocker
Button Blocker
Button Blocker
Latch
Latch
Latch

Figure 4-114 *Any latch now blocks all the buttons.*

There's something important missing, though: a reset switch, to put the system back to its starting mode after the players have had time to press their buttons and see who won. Also, I need a way to prevent players from pressing their buttons too soon, before the quizmaster has finished asking the question. Maybe I can combine this function in just one switch, which will be under the quizmaster's control.

See Figure 4-115. In its Reset position, the quizmaster switch can reset the system and remove power to the buttons. In its Set position, the switch stops holding the system in reset mode, and provides power to the buttons. I've gone back to showing just two players, to make everything as simple as possible, but the concept is still easily expandable.

Now I have to deal with a logic problem in the diagram. The way I've drawn it, everything is joined together. I've used arrows to show the direction of signals, but I don't know how I could actually stop signals from going the wrong way. If I don't deal with this, the signal from either player will light up both LEDs. How can I stop this from happening?

I could put diodes in the "up" wires to block current from running down them. But I have a more elegant idea: I'll add an OR gate, because the inputs to an OR gate are separated from each other electrically. Figure 4-116 shows this.

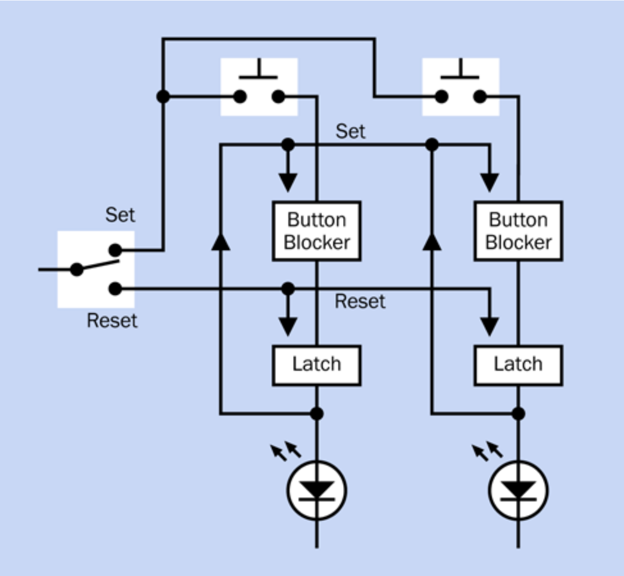
Set
Set
Reset
Reset
Button Blocker
Button Blocker
Latch
Latch

Figure 4-115 *Quizmaster control has now been added.*

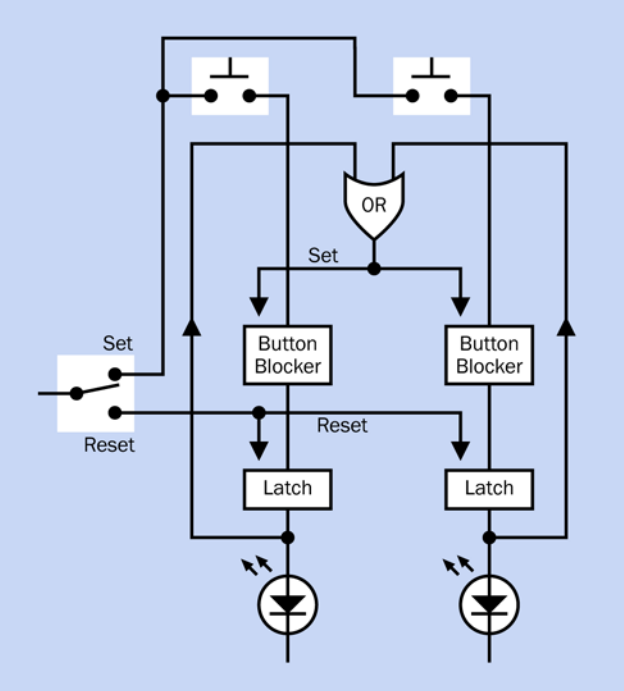
OR
Set
Set
Reset
Reset
Button Blocker
Button Blocker
Latch
Latch

Figure 4-116 *Adding an OR gate isolates one player's circuit from the other.*

The basic OR gate has only two logical inputs. Will this prevent me from adding more players? No, because you can buy an OR that has three, four, or even eight inputs. If any one of them is high, the output is high. For fewer players than a gate can handle, you tie the unused inputs to ground and ignore them.

Now I'm getting a clearer idea of what the thing I've called a "button blocker" should actually be. I think it should be another logic gate. It should say, "If there's only one input, from a button, I'll let it through. But if there are any additional inputs, I won't let them through."

Before I start choosing gates, though, I have to decide what the latch will be. I can buy an off-the-shelf flip-flop, which flips "on" if it gets one signal and "off" if it gets another, but chips containing flip-flops tend to have more features than I need for a simple circuit like this. Therefore I'm going to use 555 timers again, in bistable mode. They require very few connections, work very simply, and can deliver a good amount of current to drive bright LEDs. The only problem with them is that 555s in bistable mode require:

- Negative trigger input to create a high output
- Negative reset input to create a low output

All right, so, each player's button will have to generate a negative pulse instead of a positive pulse. That will suit the requirements of the timers.

Finally, here's a simplified schematic, in Figure 4-117. I like to show the pins of the 555 timers in their correct positions, so I had to move the components around a little to minimize wire crossovers, but you can see that logically, it's the same basic idea.

I didn't have enough room to add positive and negative symbols to show which timer pins are being held in each state, so a red circle means that the pin is being held in a high state, and a blue circle shows that it is being held in a low state. Black circles mean that the pin states may change. White circles mean that the states of these pins are not important, and the pins can be left unconnected.

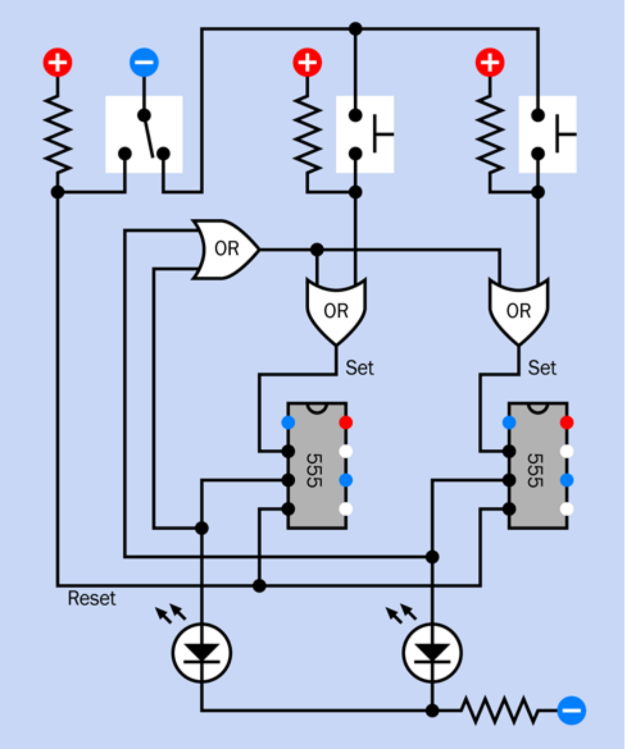

Figure 4-117 *A preliminary logic diagram. Blue pins on the timers are held in a low state, red pins are held in a high state, and white pins are not relevant.*

Before you try to build it, just run through the theory of it, because that's the final step, to make sure there are no mistakes. The important thing to bear in mind is the 555 needs a *negative* input on its trigger pin to create a *positive* output. This means that when any of the players presses a button, the button has to create a negative "flow" through the circuit.

This is a bit counterintuitive, so I'm including a four-step visualization in Figure 4-118, Figure 4-119, Figure 4-120, and Figure 4-121, showing how it will work.

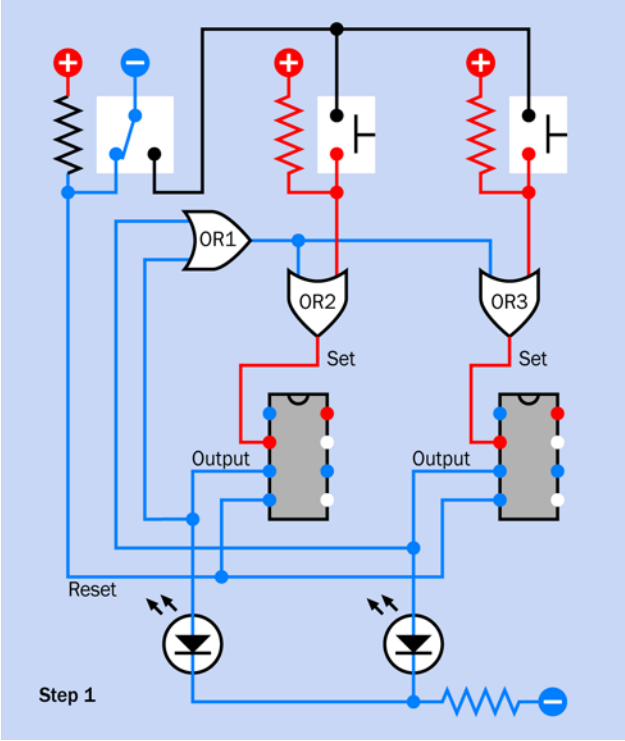

Figure 4-118 *Step 1 in the circuit visualization. Reset mode.*

In Step 1, the quizmaster's switch is in reset mode. The low voltage on the reset pins of the timers has forced them both to create negative outputs. These outputs keep the LEDs dark and also go to the OR1 gate. Because it has negative inputs, it creates a negative output, although OR2 and OR3 ignore that because one input on each of these gates is high, provided by pullup resistors beside the buttons. Remember, if either of the inputs to an OR gate is high, the output is high. And so long as the trigger pin on a timer in bistable mode is high, the timer will not be triggered. So, the circuit is stable.

In Step 2, the quizmaster has asked a question and flipped his switch to the right, to supply (negative) power to the players' buttons. However, neither of the players has responded yet, so the pullup resistors maintain the circuit in a stable state with negative outputs from the timers.

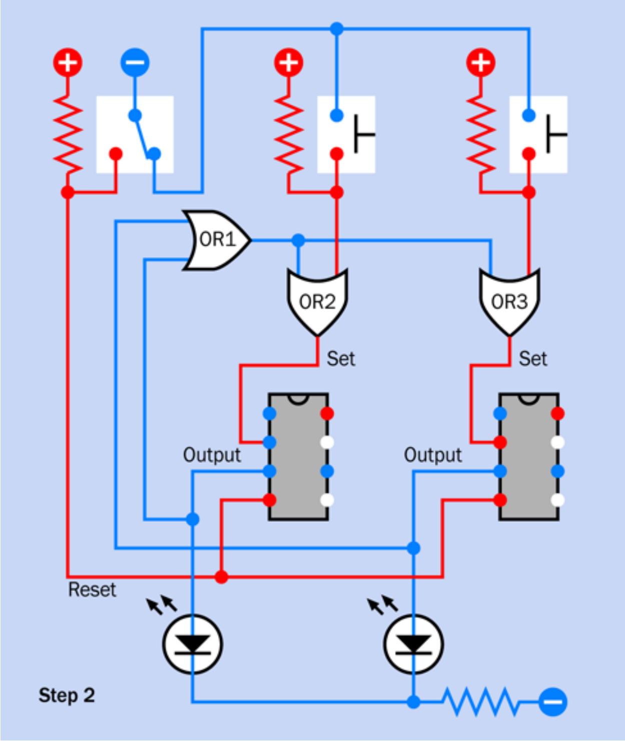

Figure 4-119 *Step 2 in the circuit visualization. Players' buttons are active, but no one has pressed a button yet.*

In Step 3, Player 1 has pressed the lefthand button. This sends a low pulse to OR2. Now that OR2 has two low inputs, its output goes low. The low pulse goes to the trigger pin of the lefthand timer. But components do not respond instantaneously, and the timer has not processed the signal yet.

In Step 4, a few microseconds later, the timer has processed the negative input signal and created a positive output pulse that lights the LED and also circulates back to OR1. Now that OR1 has a positive input, it also has a positive output. This goes to the inputs of OR2 and OR3, and so their outputs become positive. As a result, both of the timers now have positive inputs on their trigger pins. Any button press by either player will now be ignored, because OR1 continues to supply positive current in the circuit.

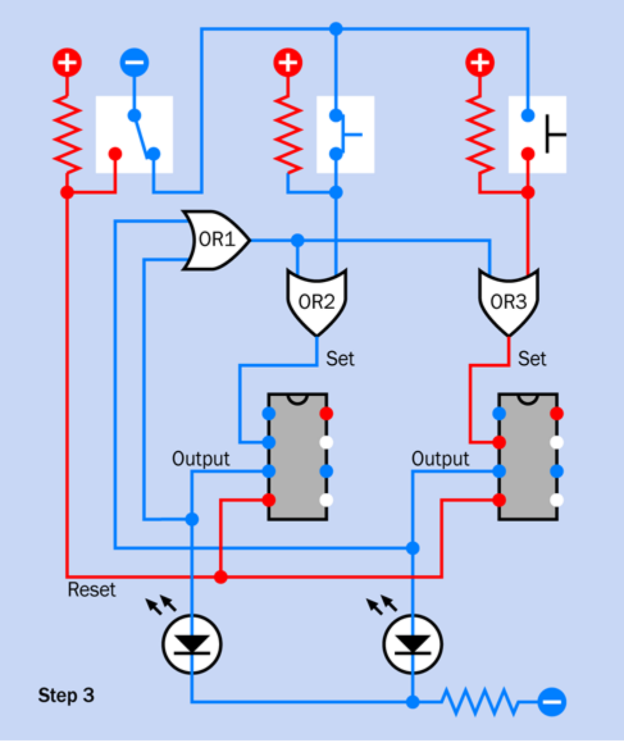

Figure 4-120 *Step 3 in the circuit visualization. The left player has pressed a button, but the 555 timer has not responded yet.*

- Remember, when a 555 is running in flip-flop mode, a low input on its trigger pin will flip it into having a high output, and the output will persist even if the trigger pin goes high again.
- The only thing that will end the 555's high output is a low state on its reset pin. That will only happen when the game master turns his switch back to reset mode.

There's only one situation that can upset this happy scenario. What if both players press their buttons absolutely simultaneously? In the world of digital electronics, this is highly unlikely. But if it somehow happens, both of the timers should react, and both of the LEDs will light up, showing that there has been a tie.

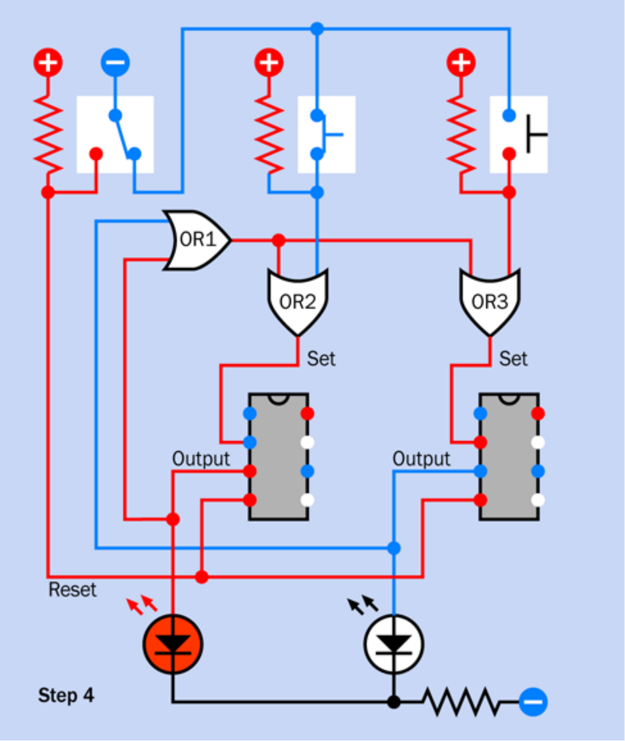

Figure 4-121 *Step 4 in the circuit visualization. The left player's action has flowed through the circuit and now blocks the right player.*

On the *Jeopardy!* TV show, you never see a tie. Absolutely never! I'm wondering, if the electronic system on the show registers a simultaneous response from two players, maybe it has a randomizing feature to select one of them. Just speculating, of course.

To demonstrate how a two-player circuit can be upgraded to handle extra players, I've included a simplified three-player schematic in Figure 4-122. The circuit could be extended outward indefinitely, the only limit being the number of inputs available on OR1.

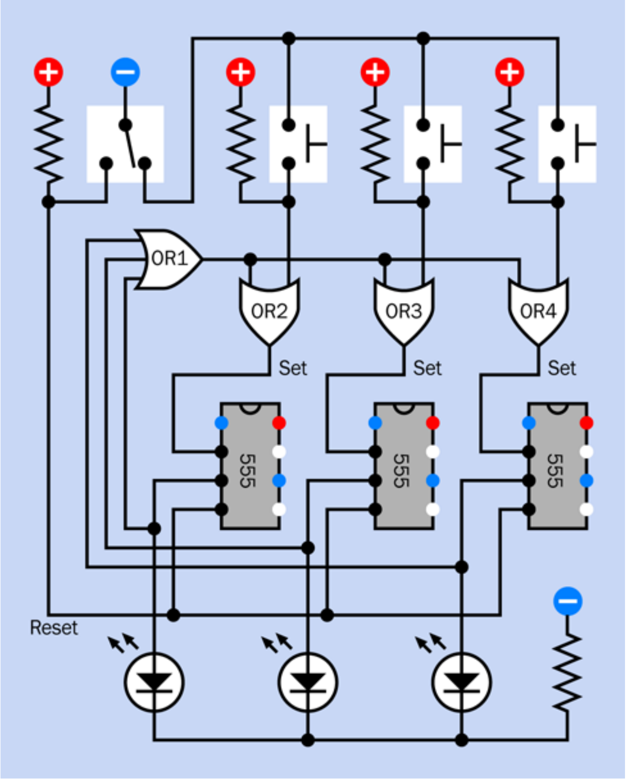
OR1
OR2
OR3
OR4
Set
Set
Set
555
555
555
Reset

Figure 4-122 *The circuit can be expanded easily for more players.*

Breadboarding It

In Figure 4-123 I have revised the schematic using an actual OR chip, in a layout as close to a breadboard configuration as possible, so that you can build this thing easily. A breadboarded version is shown in Figure 4-124, and the component values are shown in Figure 4-125.

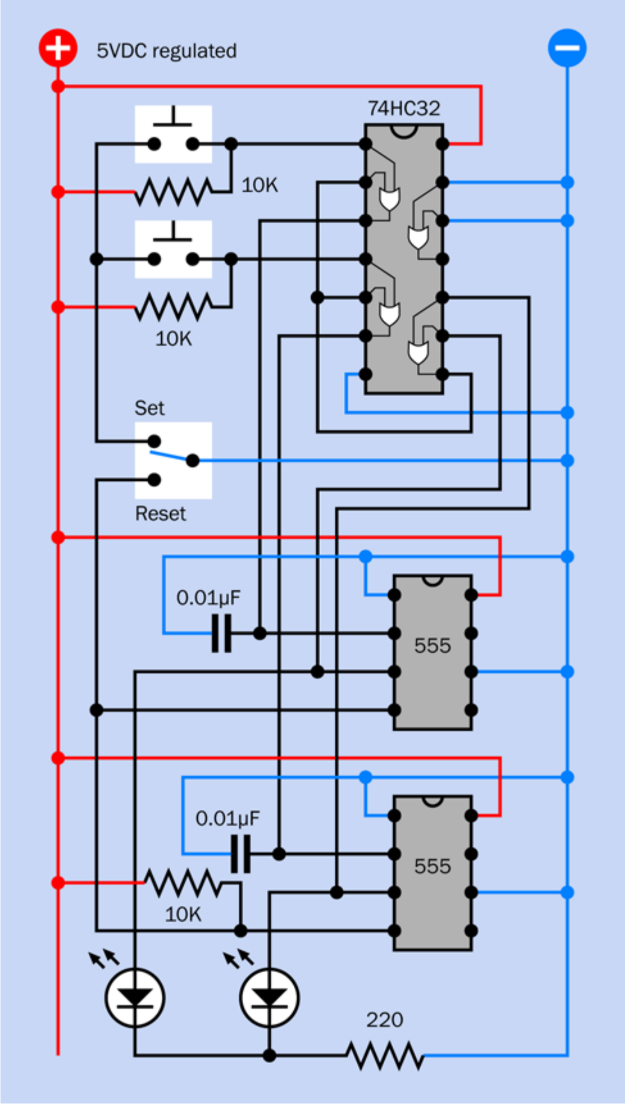
5VDC regulated
74HC32
10K
10K
Set
Reset
0.01µF
555
0.01µF
555
10K
220

Figure 4-123 *The two-player schematic has been redrawn here, using a quad two-input OR chip.*

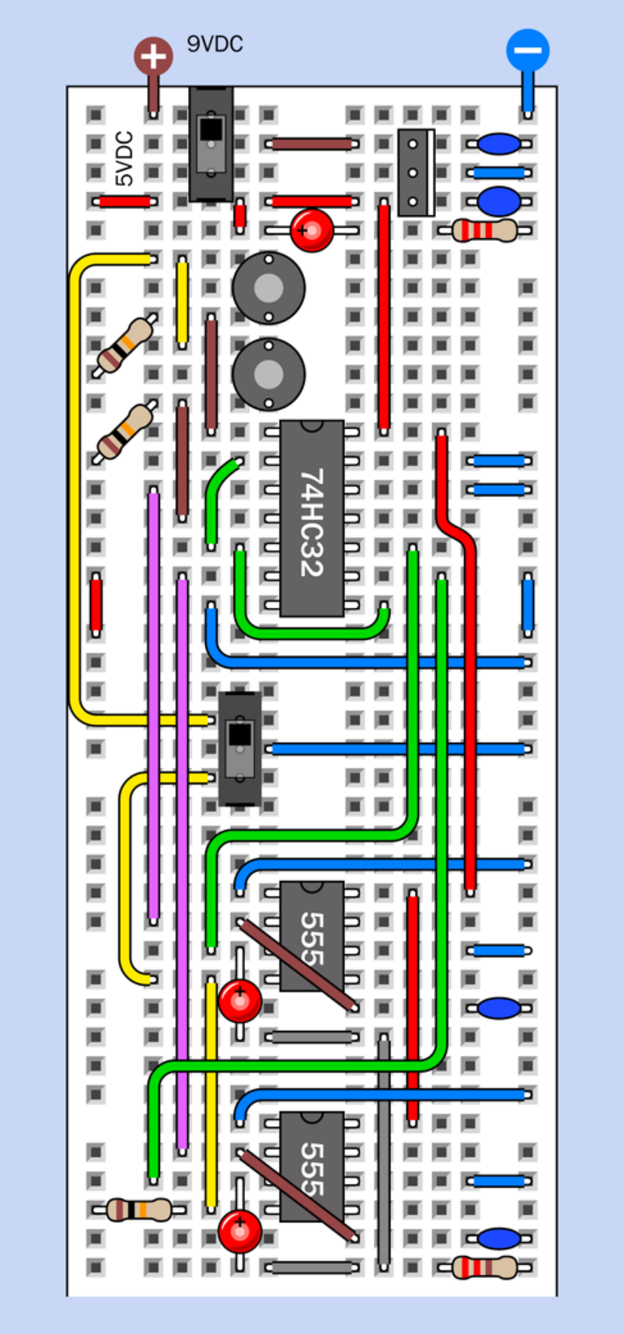
9VDC
5VDC
74HC32
555
555

Figure 4-124 *Breadboard layout, equivalent to the schematic.*

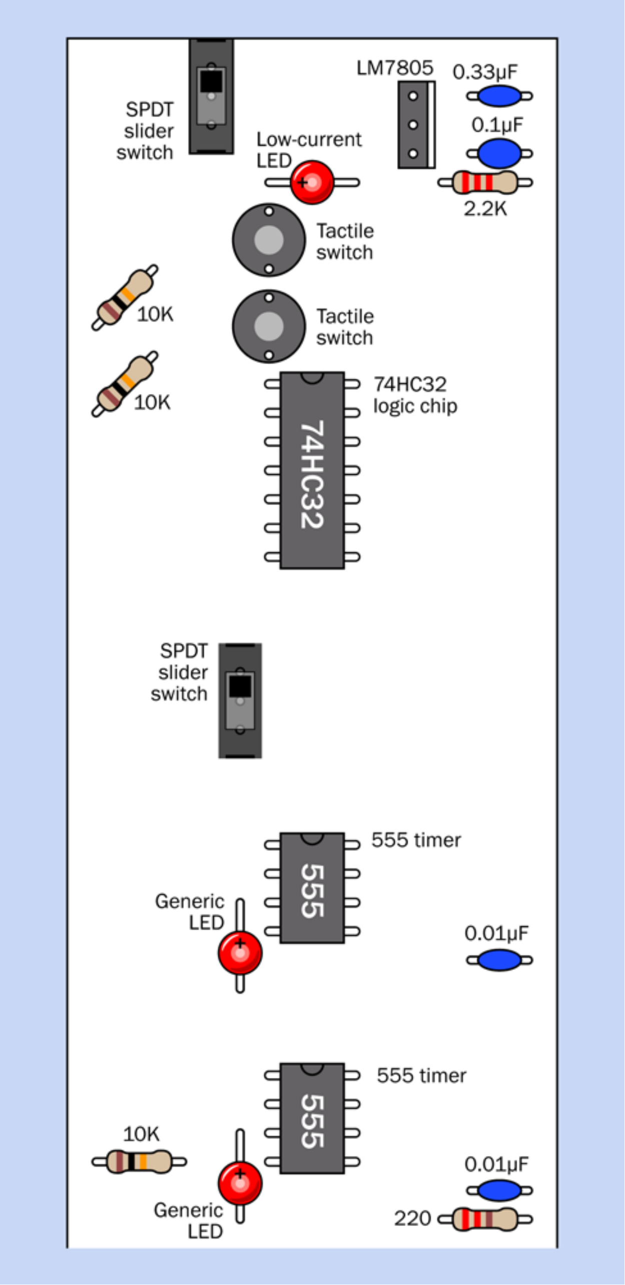
SPDT slider switch
LM7805
0.33µF
0.1µF
Low-current LED
2.2K
Tactile switch
Tactile switch
10K
10K
74HC32 logic chip
74HC32
SPDT slider switch
555 timer
555
Generic LED
0.01µF
555 timer
555
10K
0.01µF
Generic LED
220

Figure 4-125 *Component values for the breadboard layout.*

Because the only logic gates that I've used are OR gates, and only three are needed, I just need one logic chip: the 74HC32, which contains four two-input OR gates. (I've grounded the inputs to the fourth). The two OR gates on the left side of the chip have the same functions as OR2 and OR3 in my simplified schematic, and the OR gate at the bottom-right side of the chip works as OR1, receiving input from pin 3 of each 555 timer. If you have all the components, you should be able to put this together and test it quite quickly.

You may notice that I've added a 0.01μF capacitor between pin 2 of each 555 timer (the input) and negative ground. Why? Because when I tested the circuit without the capacitors, sometimes I found that one or both of the 555 timers would be triggered simply by flipping the quizmaster switch, without anyone pressing a button.

At first this puzzled me. How were the timers getting triggered, without anyone doing anything? Maybe they were responding to "bounce" in the quizmaster switch —meaning tiny and very rapid vibrations in the contacts when the switch is moved. Sure enough, this was happening, and small capacitors solved the problem. They may also slow the response of the 555 timers fractionally, but not enough to interfere with slow human reflexes.

As for the buttons, it doesn't matter if they "bounce," because each timer locks itself on at the very first impulse and ignores any hesitations that follow.

You can experiment building the circuit, disconnecting the 0.01μF capacitors, and flipping the quizmaster switch to and fro a dozen times. Because I am recommending a small, cheap slide switch, I think you will see a number of "false positives." I'm going to explain more about switch bounce, and how to get rid of it, in the next experiment.

Enhancements

After you breadboard the circuit, if you proceed to build a permanent version, I suggest that you expand it so that at least four players can participate. This will require an OR gate capable of receiving four inputs. The 74HC4078 is the obvious choice, as it allows up to eight. Just connect any unused inputs to negative ground.

Alternatively, if you already have a couple of 74HC32 chips and you don't want to bother ordering a 74HC4078, you can gang together three of the gates inside a single 74HC32 so that they function like a four-input OR. Look at the simple logic diagram in Figure 4-126 showing three ORs, and remember that the output from each OR will go high if at least one input is high.

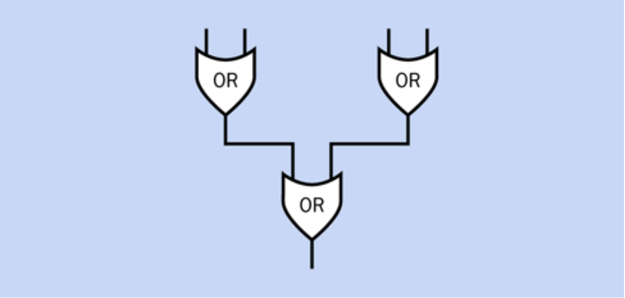

Figure 4-126 *Three two-input ORs can emulate a single four-input OR.*

And while you're thinking about this, can you figure out how to combine three two-input ANDs that will substitute for a four-input AND gate?

For a four-player game, you'll need two additional 555 timers, of course, and two more LEDs, and two more pushbuttons.

As for creating a schematic for the four-player game—I'm going to leave that to you. Begin by sketching a simplified version, just showing the logic symbols. Then convert that to a breadboard layout (which is the hard part). And here's a suggestion: pencil, paper, and an eraser can still be quicker, initially, than circuit-design software or graphic-design software, in my opinion.

Experiment 23: Flipping and Bouncing

In three experiments, now, I've used 555 timers in bistable mode. The time has come to deal with "real" flip-flops, including an explanation of how they work. I will also show how they can deal with the phenomenon that I mentioned briefly in the previous experiment: *switch bounce*.

When a switch is flipped from one position to another, its contacts vibrate very briefly. This is the "bounce" of which I speak, and it can be a problem in circuits where digital components respond so quickly, they interpret every tiny vibration as a separate input. If you connect a pushbutton to the input of a counter chip, for instance, the counter may register 10 or more input pulses from a single press of the button. A sample of actual switch bounce is shown in Figure 4-127.

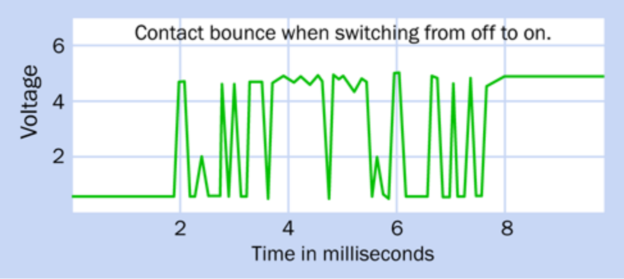

Figure 4-127 *Fluctuations created by vibrating contacts when a switch is closed. (Derived from a datasheet at Maxim Integrated corporation.)*

There are many techniques for debouncing a switch, but using a flip-flop is probably the most fundamental.

What You Will Need

- Breadboard, hookup wire, wire cutters, wire strippers, multimeter
- 9-volt power supply (battery or AC adapter)
- 74HC02 logic chip (1), 74HC00 logic chip (1)
- SPDT slide switches (2)
- Low-current LEDs (3)
- Resistors: 680 ohms (2), 10K (2), 2.2K (1)
- Capacitors: 0.1μF (1), 0.33μF (1)
- LM7805 voltage regulator (1)

Assemble the components on your breadboard, as shown in Figure 4-128. The same circuit is shown as a schematic in Figure 4-129, and the component values are shown in Figure 4-130. When you apply power, one of the LEDs at the bottom should be lit.

Now I want you to do something odd. Please disconnect the wire labeled A in Figure 4-128. Just pull it out of the board. If you refer to the schematic in Figure 4-129, you'll see that you have disconnected power to the pole of the switch, leaving the two NOR gates connected only with their pulldown resistors.

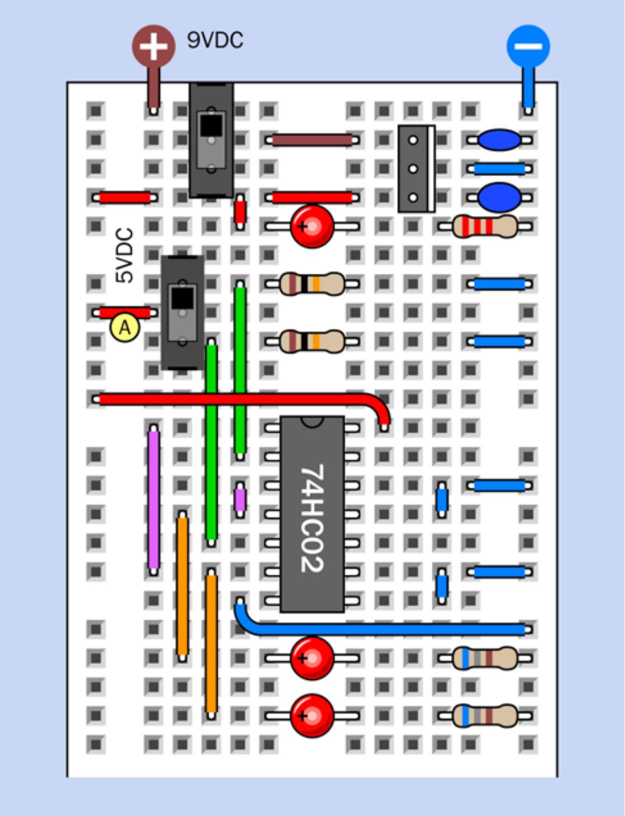

Figure 4-128 *Breadboarded flip-flop circuit using NOR gates.*

You may be surprised to find that the LED remains lit.

Push the wire back into the board, slide the switch to its opposite position, and the first LED should go out, while the other LED should come on. Once again, pull out the wire, and once again, the LED should remain lit.

Here's the take-home message:

- A flip-flop requires only an initial input pulse—for example, from a switch.
- After that, it ignores that input.

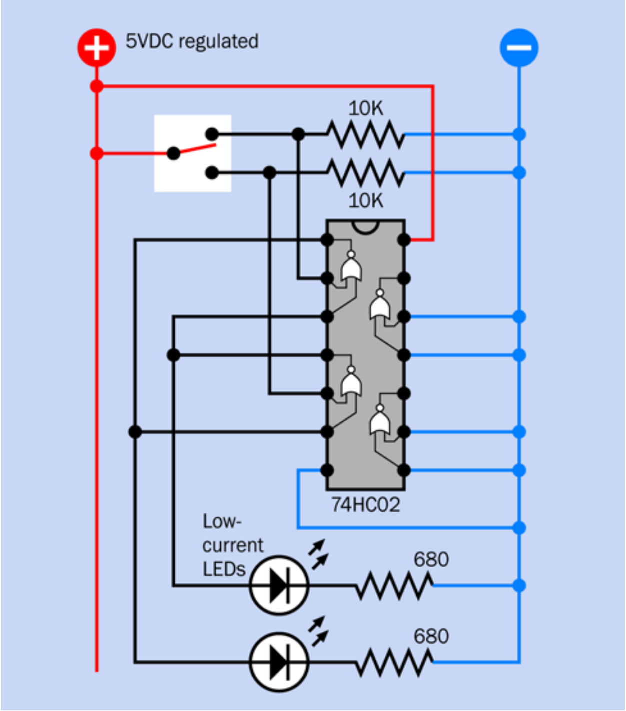

Figure 4-129 *Flip-flop schematic using NOR gates.*

How It Works

Two NOR gates or two NAND gates can function as a flip-flop.

- Use NOR gates when a double-throw switch delivers a positive input.
- Use NAND gates when a double-throw switch delivers a negative input.

Either way, you have to use a double-throw switch.

I've mentioned the double-throw switch three times (actually, four times if you count this sentence) because for some strange reason, most introductory books fail to emphasize this point. When I first started learning electronics, I went crazy trying to understand how two NORs or two NANDs could debounce a simple SPST pushbutton—until finally I realized that they can't. The reason is that when you power up the circuit, the NOR gates (or NAND gates) need to be told in which state they should begin. Their initial orientation comes from the switch being in one state or the other. A SPST pushbutton cannot do that, when it is not being pushed. So you have to use a double-throw switch. (Now I've mentioned it five times.)

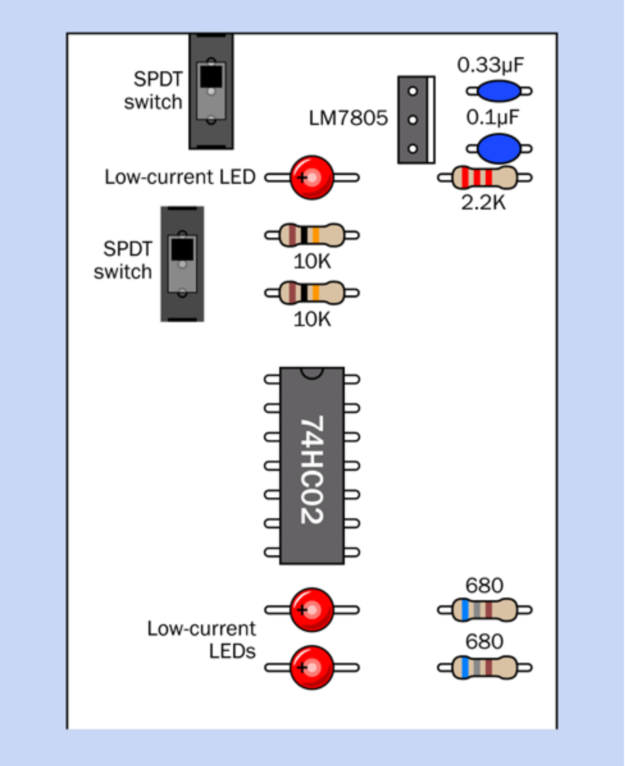

Figure 4-130 *Component values for the breadboarded NOR-based flip-flop.*

Debouncing with NORs

I've created a multiple-step schematic in Figure 4-131 and Figure 4-132 to show the changes that occur as a switch flips to and fro with two NOR gates. To refresh your memory, I've also included a truth table in Figure 4-133 showing the logical outputs from NOR gates for each combination of inputs.

Referring initially to Figure 4-131, in Step 1, the switch is supplying positive current to the lefthand side of the circuit, overwhelming the negative supply from the pulldown resistor, so we can be sure that the NOR gate on the left has one positive logical input. Because any positive logical input will make the NOR give a negative output (as shown in the truth table in Figure 4-133), the negative output crosses over to the righthand NOR, so that it now has two negative inputs, which make it give a positive output. This crosses back to the lefthand NOR gate. So, in this configuration everything is stable.

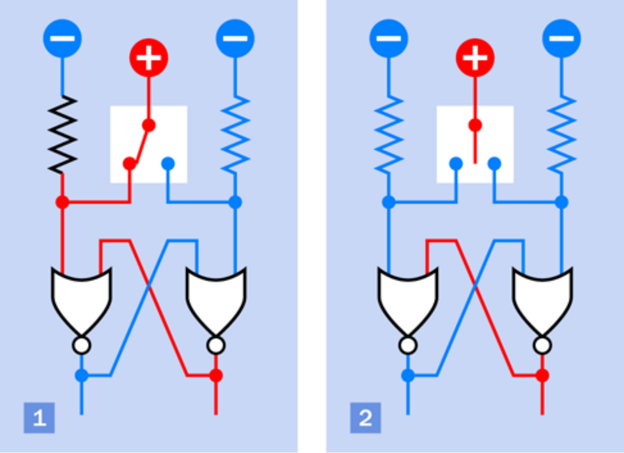

Figure 4-131 *When the switch moves to a neutral center position, the status of the NOR gates remains unchanged.*

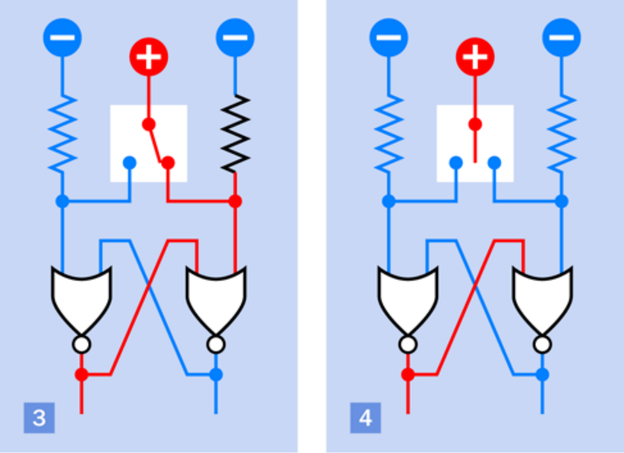

Figure 4-132 *After the states of the NOR gates are reversed, they remain that way when the switch moves back to its neutral center position.*

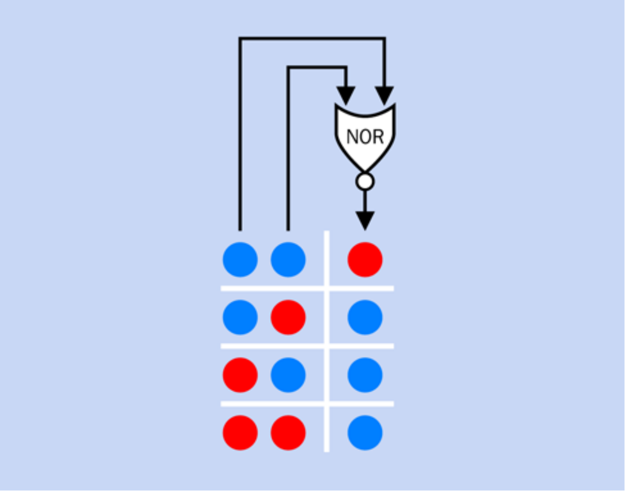

Figure 4-133 *A reminder of the truth table for a NOR gate.*

Now comes the clever part. In Step 2, suppose that you move the switch so that it doesn't touch either of its contacts. (Or suppose that the switch contacts are bouncing, and failing to make a good contact. Or suppose you disconnect the switch entirely.) Without a positive supply from the switch, the lefthand input of the left NOR gate goes from positive to negative, as a result of the pulldown resistor. But the righthand input of this gate is still positive, and one positive is all it takes to make the NOR maintain its negative output, so nothing changes. In other words, the circuit has "flopped" in this state, regardless of whether the switch is disconnected.

Referring to Figure 4-133, if the switch turns fully to the right and supplies positive power to the righthand pin of the right NOR gate, that NOR recognizes that it now has a positive logical input, so it changes its logical output to negative. That goes across to the other NOR gate, which now has two negative inputs, so its output goes positive, and runs back to the right NOR.

In this way, the output states of the two NOR gates change places. They flip, and then flop there, even if the switch breaks contact or is disconnected again, as in Step 4.

If a switch bounces so severely that the pole connection fluctuates all the way between one contact and the other, this circuit won't work. It only works if the output alternates between one connection, and no connection at all. But this is generally the case with a SPDT switch.

Debouncing with NANDs

The drawings in Figure 4-134 and Figure 4-135 show a similar sequence of events if you use a negatively powered switch with two NAND gates. To refresh your memory of NAND behavior, I am including Figure 4-136.

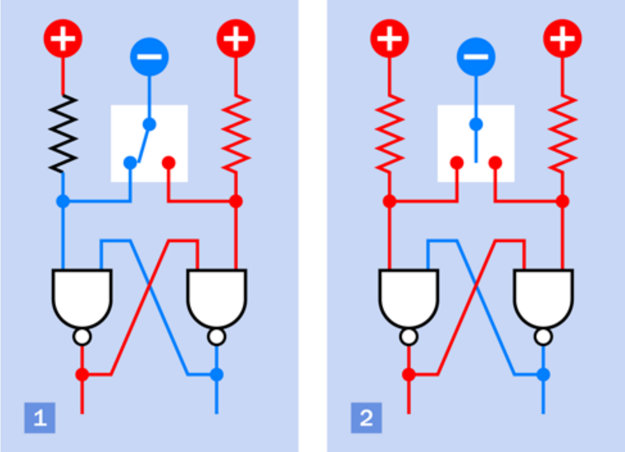

Figure 4-134 *Two NAND gates can be used as a flip-flop with pullup resistors and a switch providing negative power.*

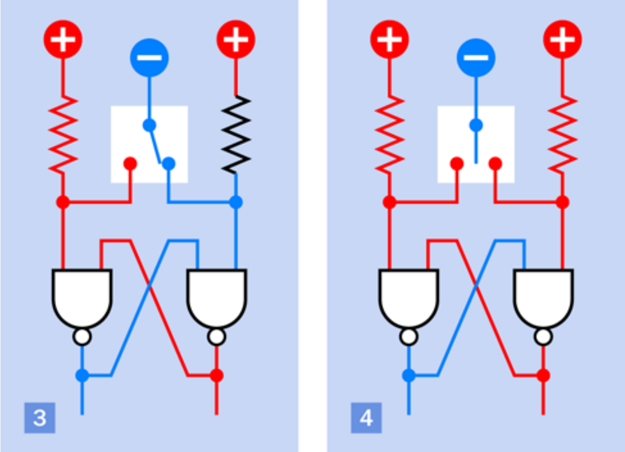

Figure 4-135 *Once again, the gate states remain unchanged when the switch is unconnected with either of them.*

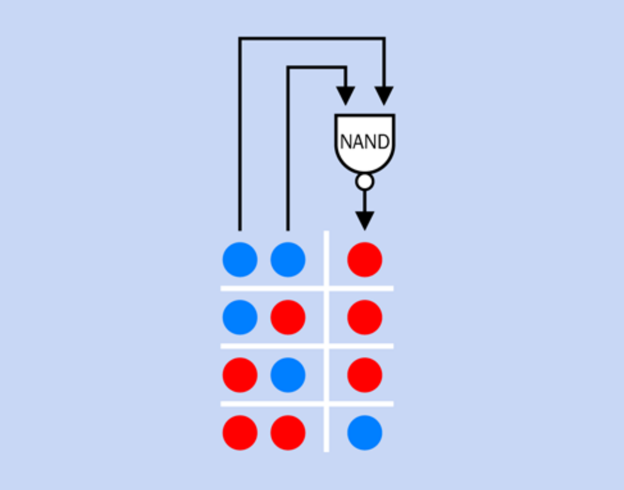

Figure 4-136 *A reminder of the truth table for a NAND gate.*

If you want to verify the function of the NAND circuit, you can use your 74HC00 chip, specified in the parts list for this experiment, to test it yourself. Be careful, though: the gates inside the NOR chip are upside-down compared with the gates inside the NAND chip. You will have to move some wires around on your breadboard, because the two chips are not swappable. Refer back to Figure 4-83 and Figure 4-95 for the specifications.

Jamming Versus Clocking

The NOR and the NAND circuits are examples of a *jam-type flip-flop*, so called because the switch forces it to respond immediately, and jams it into that state. You can use this circuit anytime you need to debounce a switch (as long as it's a double-throw switch).

A more sophisticated version is a *clocked flip-flop*, which requires you to set the state of each input first and then supply a clock pulse to make the flip-flop respond. The pulse has to be clean and precise, which means that if you supply it from a switch, the switch must be debounced—probably by using another jam-type flip-flop! Considerations of this type have made me reluctant to use clocked flip-flops in this book. They add a layer of complexity that I prefer to avoid in an introductory text. If you want to know more about flip-flops, I explore them in greater detail in *Make: More Electronics*. It's not a simple topic.

What if you want to debounce a single-throw button or switch? Well, you have a problem! One solution is to buy a special-purpose chip such as the 4490 "bounce eliminator," which contains digital delay circuitry. A specific part number is the MC14490 from On Semiconductor. This contains six circuits for six separate inputs, each with an internal pullup resistor. It's relatively expensive, however—more than 10 times the price of a 74HC02 containing NOR gates. Really, you will have an easier time if you avoid single-throw switches and use double-throw switches (or pushbuttons) that are more easily debounced.

Or, you could use a 555 timer wired in flip-flop mode. My preference for that option now seems to make more sense.

Experiment 24: Nice Dice

Electronic circuits to simulate the throw of one or two dice have been around for decades. However, new configurations still exist, and this project provides an opportunity to learn more about logic while ending up with something useful. I especially want to introduce you to binary code, the universal language among digital chips.

What You Will Need

- Breadboard, hookup wire, wire cutters, wire strippers, multimeter
- 9-volt power supply (battery or AC adapter)
- 555 timer (1)
- 74HC08 logic chip (1), 74HC27 logic chip (1), 74HC32 logic chip (1)
- 74HC393 binary counter (1)
- Tactile switch (1)
- SPDT slide switches (2)
- Resistors: 100 ohms (6), 150 ohms (6), 220 ohms (7), 330 ohms (2), 680 ohms (4), 2.2K (1), 10K (2), 1M (1)
- Capacitors: 0.01µF (2), 0.1µF (2), 0.33µF (1), 1µF (1), 22µF (1)
- LM7805 voltage regulator (1)
- Low-current LEDs (15)
- Generic LED (1)

A Binary Counter

At the heart of every electronic dice circuit that I have ever seen is a counter chip of some kind. Often it's a *decade counter* with 10 "decoded" output pins that are energized one at a time, in sequence. A die only has six faces, but if you tie the seventh pin of the counter back to its reset pin, the counter will restart after it reaches six. (Note that "dice" is really a plural word that should only be used for two dice or more, while "die" is the singular word, although many people don't realize this.)

I always like to do things differently, so I decided not to use a decade counter, partly because I wanted a binary counter to satisfy my desire to demonstrate binary code. This added a little complexity to the circuit, but will enrich the learning process—and when all is said and done, you'll have a circuit that runs two dice (not just one) with a modest chip count, while still fitting on a breadboard.

The counter chip that I've chosen is widely used: the 74HC393. It actually contains two counters, but the second one can be ignored for the time being. The pinouts are shown in Figure 4-137.

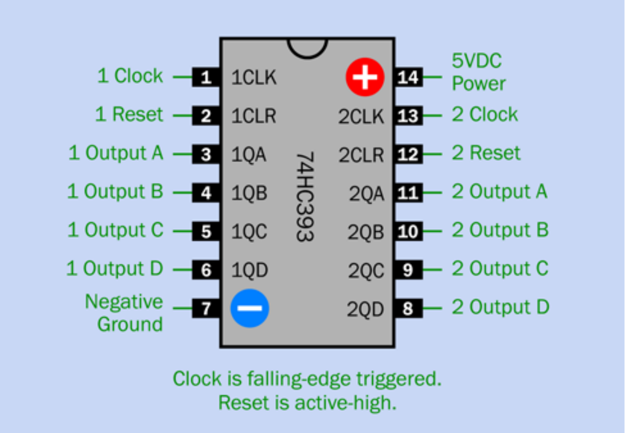

Figure 4-137 *Pin functions of the 74HC393 binary counter.*

Manufacturers have an odd habit of identifying the pin functions of digital chips using as few letters as possible. These cryptic abbreviations can be difficult to understand. To give you an example, in Figure 4-137 the pin labels inside the outline of the chip are the ones I found in a datasheet from Texas Instruments. (To make things even more confusing, other manufacturers use different abbreviations of their own. There is no standardization.)

Outside the outline of the counter I have restated the pin functions in relatively plain English, in green lettering. The number preceding each pin function refers to counter #1 or counter #2, packaged separately in the chip.

Counter Testing

The best way to understand this chip is to bench-test it. Figure 4-138 shows the schematic, and Figure 4-139 shows it on a breadboard. Figure 4-140 shows the breadboarded component values.

Remember:

- This is a 5V logic chip. Don't leave out the voltage regulator.
- Note there is a 0.1µF capacitor between the power-supply pin of the timer, and ground. This is to suppress little voltage spikes that the timer tends to generate. They can confuse the counter if they are uncontrolled.

The capacitor and resistors that I have specified with the timer will run at about 0.75Hz. In other words, the beginning of one pulse, and the beginning of the next pulse, will be slightly more than one second apart. You can see this by watching the yellow LED on the timer output. (If the yellow LED does not behave this way, you made a wiring error somewhere.)

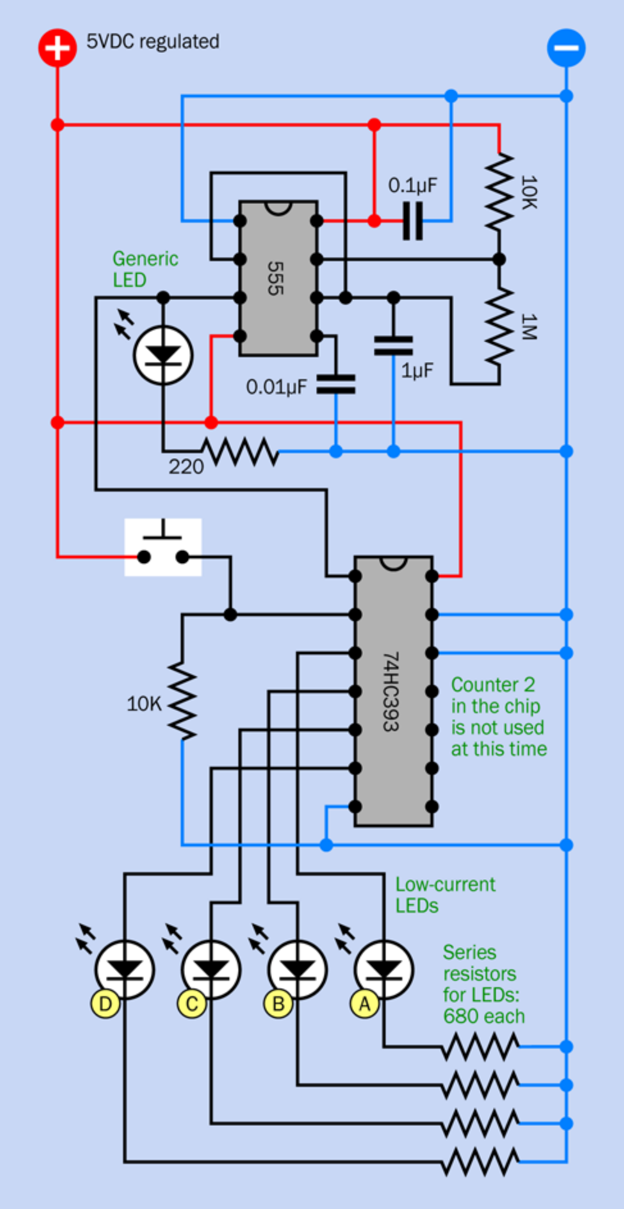
5VDC regulated
0.1µF
10K
Generic LED
555
1M
1µF
0.01µF
220
10K
74HC393
Counter 2 in the chip is not used at this time
Low-current LEDs
Series resistors for LEDs: 680 each
D
C
B
A

Figure 4-138 *Schematic for observing the output and the reset function of the 74HC393 decade counter.*

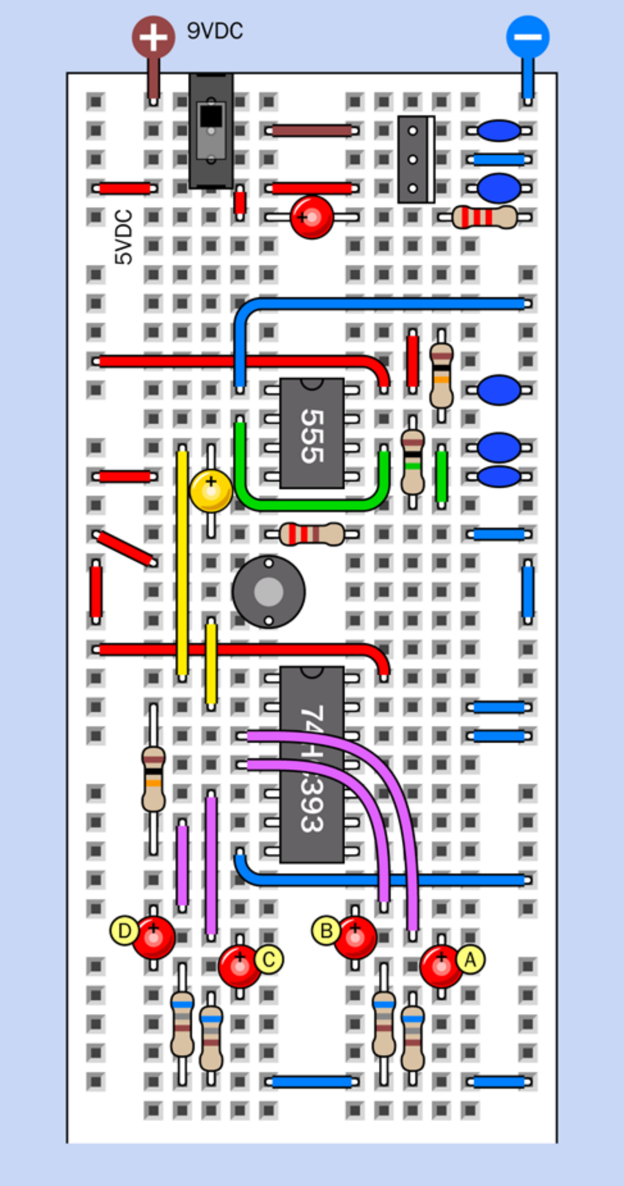
9VDC
5VDC
555
74HC393
D
C
B
A

Figure 4-139 *The breadboarded test circuit.*

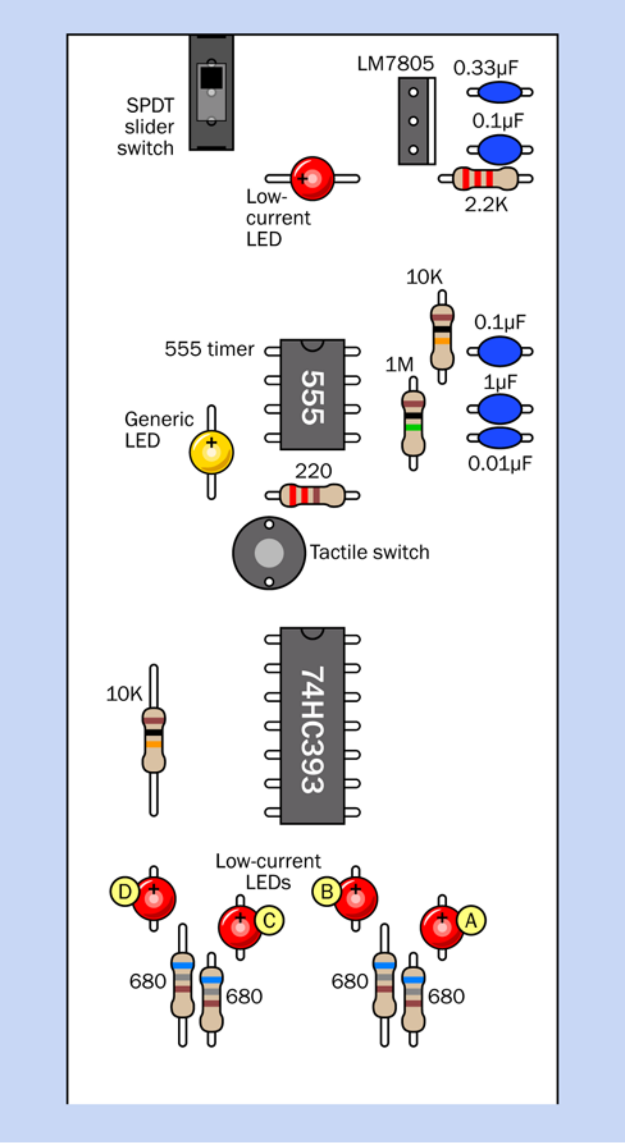
SPDT slider switch
LM7805
0.33µF
0.1µF
Low-current LED
2.2K
10K
0.1µF
555 timer
1M
1µF
555
Generic LED
0.01µF
220
Tactile switch
10K
74HC393
Low-current LEDs
D
B
C
A
680
680
680
680

Figure 4-140 *Component values in the breadboarded circuit.*

The four red LEDs labeled A, B, C, and D will display the output states of the counter. If your connections are correct, they will run through the sequence shown in Figure 4-141, where a black circle indicates that an LED is not lit, and a red circle indicates that it is lit.

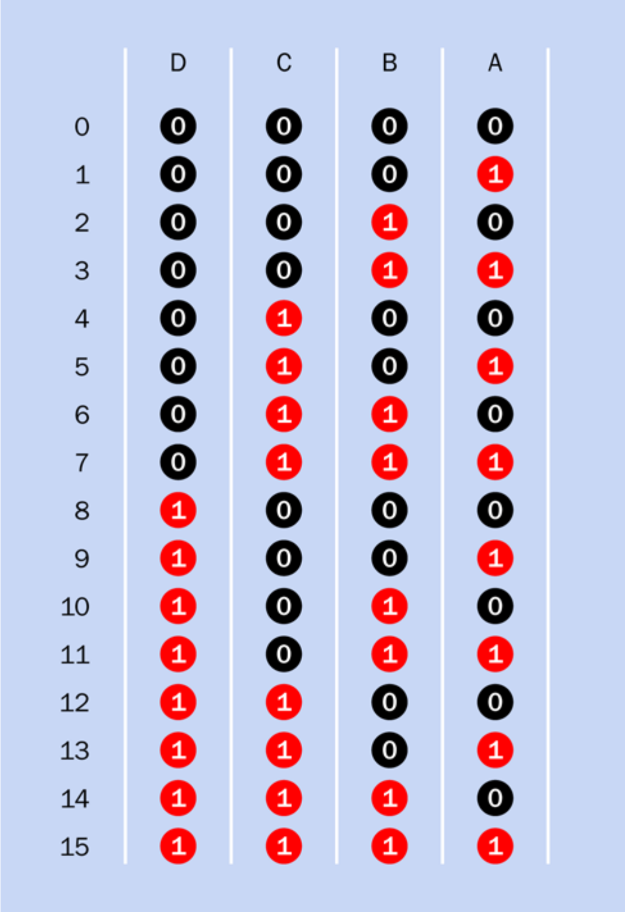

	D	C	B	A
0	0	0	0	0
1	0	0	0	1
2	0	0	1	0
3	0	0	1	1
4	0	1	0	0
5	0	1	0	1
6	0	1	1	0
7	0	1	1	1
8	1	0	0	0
9	1	0	0	1
10	1	0	1	0
11	1	0	1	1
12	1	1	0	0
13	1	1	0	1
14	1	1	1	0
15	1	1	1	1

Figure 4-141 *The full sequence of outputs from a binary counter.*

I'm going to tell you some more, now, about binary and decimal arithmetic. Do you really need to know this? Yes, it's useful. A variety of chips such as decoders, encoders, multiplexers, and shift registers use binary arithmetic, and of course it is absolutely fundamental in almost every digital computer that has ever been made.

Fundamentals: Binary Code

As you can see in Figure 4-141, every time the LED in column A goes out, the LED in column B reverses its state—from on to off, or off to on. Every time the LED in column B goes out, it reverses the LED in column C, and so on. One consequence of this rule is that each LED flashes twice as fast as the one to its left.

The row of LEDs represents a *binary number*, meaning a number written in only two digits: 0 and 1, as shown in the white font in Figure 4-141. The equivalent decimal number is shown in the black font on the left.

The LEDs can be considered *binary digits*, commonly known as *bits.*

The rule for counting in binary is very simple. In the rightmost column, start with 0, then add 1—and then, because you can only count in ones and zeroes, the next time you want to add 1, you have to revert to 0 and carry 1 to the next column to the left.

What if the numeral in the next column to the left is already a 1? Change it back to 0 and carry 1 to the next column after that. And so on.

The rightmost LED represents the *least significant bit* of a four-bit binary number. The leftmost LED is showing us the *most significant bit*.

Rising Edge, Falling Edge

When you run the test, notice that each transition of the rightmost red LED (either from on to off, or off to on) always occurs when the yellow LED goes out. Why is this?

Most counters are *edge triggered*, meaning that the rising edge or the falling edge of a high pulse nudges the counter along to the next value in its series, when the pulse is applied to the clock input pin. The behavior of the LEDs clearly shows you that the 74HC393 is falling-edge triggered. In Experiment 19, we used a counter that was rising-edge triggered. The type that you use depends on your application.

The 74HC393 counter also has a reset pin, just like the 4026B chip from Experiment 19.

- Some datasheets describe a reset pin as a "master reset" pin, which may be abbreviated MR.
- Some manufacturers call a "reset" pin a "clear" pin, which may be abbreviated CLR on a datasheet.

Whatever it's called, the reset pin will always have the same end result. It forces all the outputs of the counter to go low—which in this case means 0000 binary.

A reset pin requires a separate pulse. But does the reset occur when the pulse begins, or when it ends?

Let's find out. If you built the circuit carefully, the reset pin is held in a low state through a 10K resistor. But there is also a tactile switch that can connect the reset pin directly to the positive bus. This overwhelms the 10K resistor, and forces the reset pin to go high.

As soon as you press the tactile switch, all the outputs go dark, and they stay dark until you let go of the tactile switch. Evidently, the reset function of the 74HC393 is triggered and held by a high state.

The Modulus

Switch off the power, disconnect the pullup resistor and the tactile switch from the reset pin (pin 2), and substitute a wire as shown in Figure 4-142. All the previous connections have been grayed out. The new wire, in solid black, connects the fourth digit, from output D, to the reset pin. Figure 4-143 shows the revision on the breadboard; the new connection is colored green.

What do you think will happen?

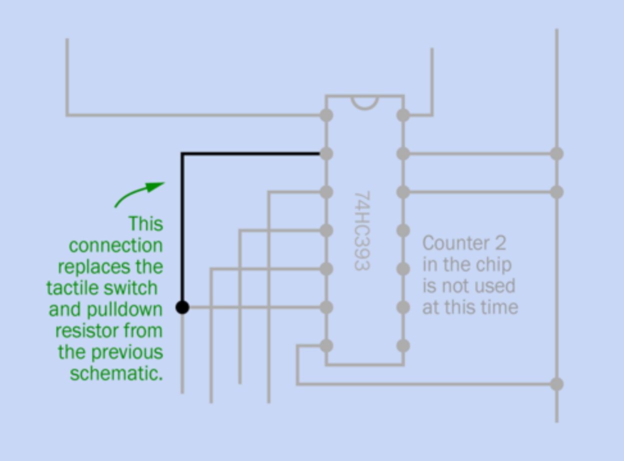
This connection replaces the tactile switch and pulldown resistor from the previous schematic.
74HC393
Counter 2 in the chip is not used at this time

Figure 4-142 *Adding an automatic reset to the timer.*

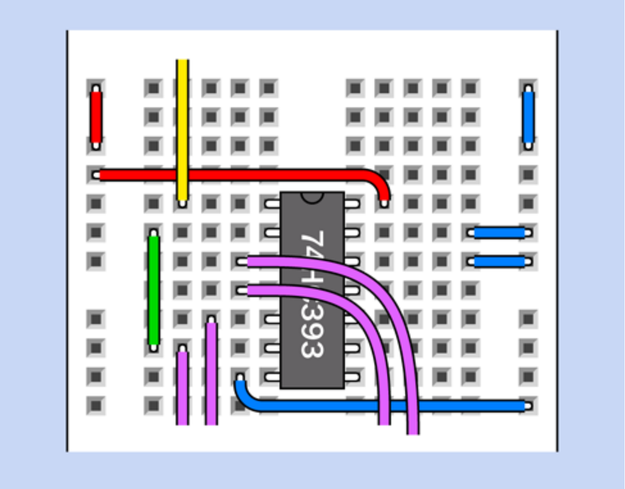

Figure 4-143 *Close-up of the revision on the breadboard. The pulldown resistor, the tactile switch, and associated connections have been removed. The green wire has been inserted.*

Run the counter again. It counts from 0000 up to 0111. The very next binary output should be 1000, but as soon as the fourth digit transitions from 0 to 1, the high state is sensed by the reset pin, forcing the counter back to 0000.

Can you see the leftmost LED flicker, before the counter resets? I doubt it, because the counter responds in less than a millionth of a second.

The counter now runs from 0000 through 0111 before automatically repeating itself. Because counting from 0000 through 0111 binary is equivalent to counting from 0 to 7 in decimal, we now have a *divide-by-8* counter. (Previously, it was a divide-by-16 counter.)

Suppose you move your reset wire from the fourth digit to the third digit. Now you have a divide-by-4 counter.

- You can easily wire almost any 4-bit binary counter so that it resets after 2, 4, or 8 input pulses.

The number of states in the output of a counter, before it repeats, is known as the *modulus*, often abbreviated as "mod." A mod-8 counter repeats after eight pulses (which are numbered from 0 through 7).

Converting to Modulus 6

What about the project that we are supposed to be working on here, to generate electronic dice patterns?

I'm getting to it. Because a die has six sides, I have a feeling that we may need to rewire the counter so that it repeats after six states.

In binary code, the output sequence will look like this: 000, 001, 010, 011, 100, 101. (We can ignore the most significant bit, in column D, because we don't need it for just six states.) I need to make the counter reset after its output of 5 decimal, which is 101 binary.

(Why 5 decimal, and not 6 decimal? Because we are counting upward from 0. It would be more convenient in this project if the counter would start from 1, but it doesn't do that.)

What's the next output after 101 binary? The answer is 110 binary.

Is there something distinctive about 110? If you study the sequence, you'll see that 110 is the first in the series that begins with two high bits.

How can we tell the counter, "When you have a 1 in column B, and a 1 in column C, reset to 0000?" The word "and" in the last sentence should give you a clue. An AND gate has a high output when, and only when, its two inputs are high. Just what we need.

Can we drop it right in? Absolutely, because all the members of the 74HCxx chip family are designed to talk to each other. In Figure 4-144 you will see that I have added an AND gate. Of course, to use it on your breadboard, you'll have to add an appropriate chip, which is a 74HC08. It contains four AND gates, of which we only need one. Therefore, in addition to power connections, its unused inputs must be grounded. This is a bit of a hassle, but I'll show you how to do it, as soon as I make a few more additions and modifications. (The unused outputs must be left unconnected.)

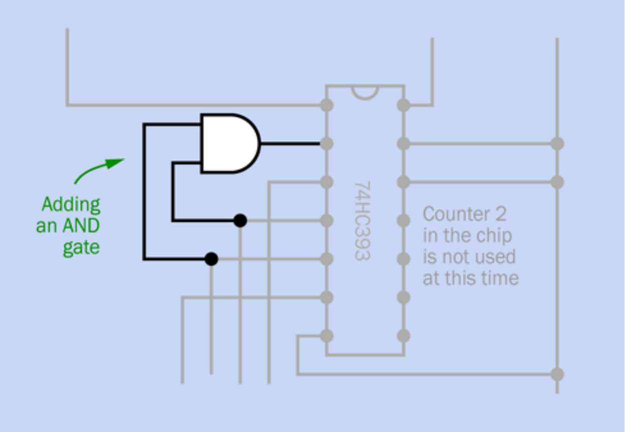

Figure 4-144 *An AND gate has been added, to adjust the counter so that it only cycles through six output states instead of its usual 16.*

Meanwhile, please remember this take-home message:

- You can use logic chip(s) with a counter to change the modulus of the counter by looking for a distinctive pattern in the output states, and feeding back a signal to the reset pin.

Not Using a Seven-Segment Display

For the dice display, I could just use a seven-segment numeral that counts from 1 to 6. But this is a problem, because the counter runs from 0 through 5. I don't see an easy way to convert 000 binary to a seven-segment numeral 1, 001 binary to a seven-segment numeral 2, and so on.

Could we make the counter skip 000 binary somehow? Er, maybe, but I'm not sure how. Perhaps by using a three-input OR gate that would feed back to the clock input to advance the counter to the next state—but then this would conflict with the usual clock signals, and it all sounds like a lot of trouble to me.

In any case, I'm not very excited by using a seven-segment numeral in this project, because it's not visually appealing. Why not use LEDs that emulate the pattern of spots on an actual die? The sequence is shown in Figure 4-145.

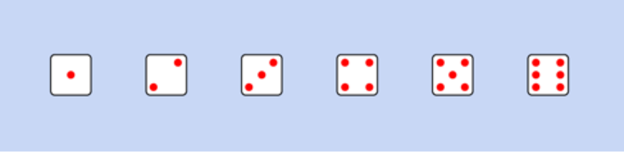

Figure 4-145 *The series of spot patterns to be reproduced by LEDs.*

Can you figure out a way to convert the binary output from the counter to illuminate the LEDs in those patterns?

Choosing the Gates

I'll start with the easiest option. If I connect output A from the counter (see Figure 4-138) to the LED that represents the center spot of the die, that will work out well, because the center spot is only illuminated for patterns 1, 3, and 5, and goes out for patterns 2, 4, and 6. This is exactly the way that output A behaves.

Then things get a little tricky. I need to illuminate a diagonal pair of spots for patterns 4, 5, and 6, and the other diagonal pair for 2, 3, 4, 5, and 6. But how?

Figure 4-146 shows my answer to the problem. You'll see I have added a couple more logic gates: a three-input NOR, and a two-input OR. Alongside I have shown the sequence of binary numbers, and the spot patterns that each number creates on the die.

To make things work, I had to begin with the die showing a 6 pattern when the counter starts at 000 binary. The sequence of patterns really doesn't matter, so long as they are all represented. They're going to be selected at random anyway.

Figure 4-147 shows how the counter outputs light up the different patterns of spots. And just in case this still isn't entirely clear, I have created a snapshot sequence showing high and low states in the circuit as the counter increments from 000 through 101. The snapshots are shown in Figure 4-148, Figure 4-149, and Figure 4-150.

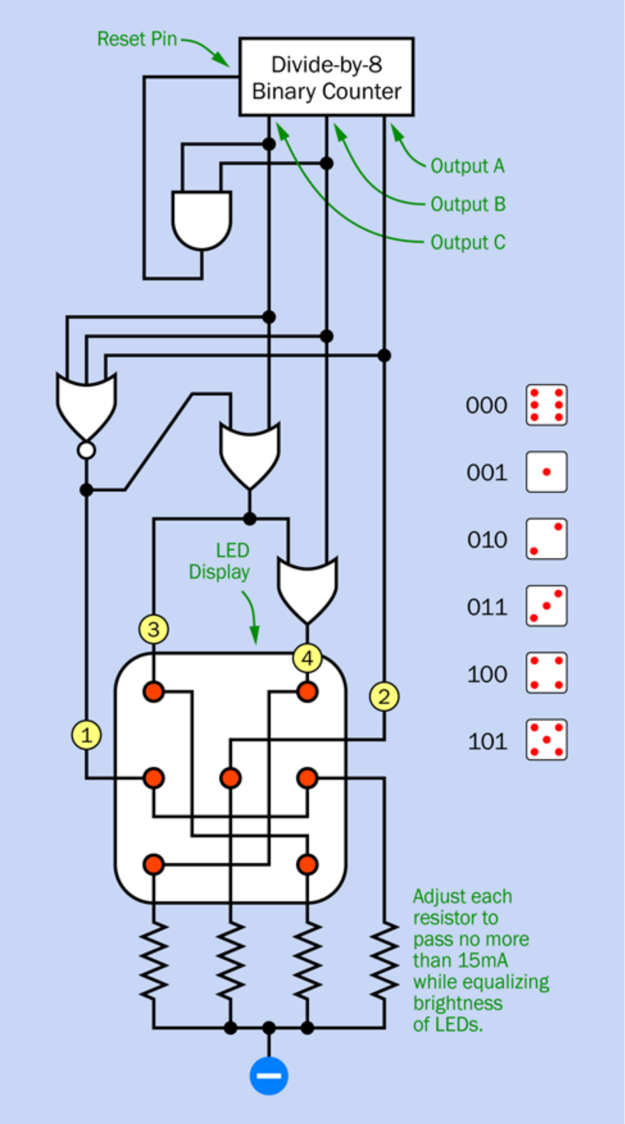
Reset Pin
Divide-by-8 Binary Counter
Output A
Output B
Output C
000
001
010
011
100
101
LED Display
3
4
1
2
Adjust each resistor to pass no more than 15mA while equalizing brightness of LEDs.

Figure 4-146 *A logic network to create the sequence of spots on a die.*

I squeezed the snapshots so that I could fit two into the width of a column, and I omitted the AND gate, because it doesn't do anything during the counting range of 000 through 101. It responds only when the counter tries to advance to 110—at which point, the AND sets it back to 000.

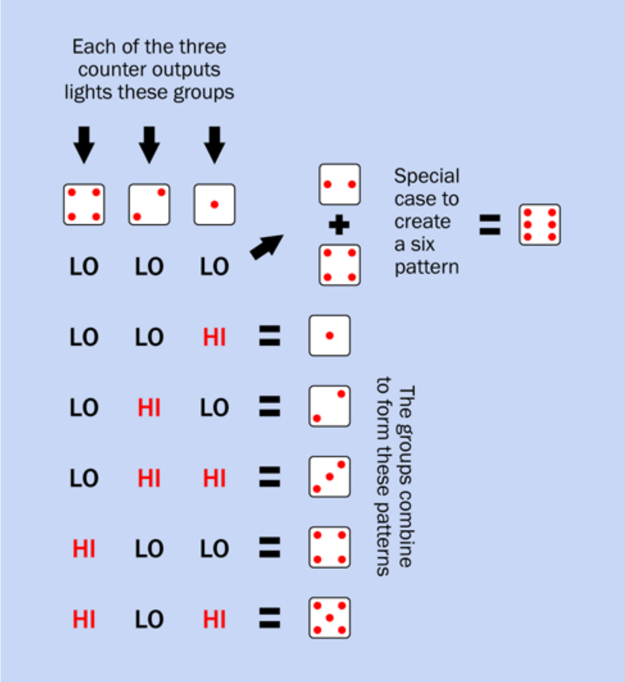
Each of the three counter outputs lights these groups
LO LO LO
Special case to create a six pattern
LO LO HI
LO HI LO
LO HI HI
HI LO LO
HI LO HI
The groups combine to form these patterns

Figure 4-147 *How the outputs from the binary counter are used to light the patterns of spots.*

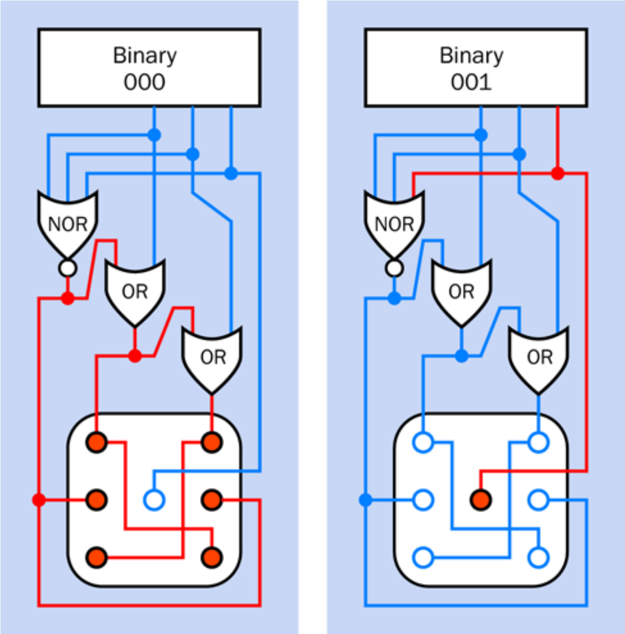
Binary 000
NOR
OR
OR
Binary 001
NOR
OR
OR

Figure 4-148 *The logic for generating patterns 6 and 1.*

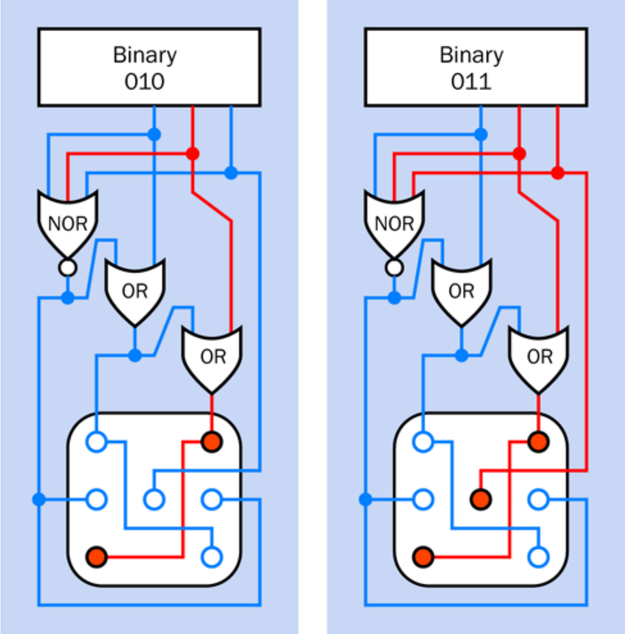
Binary 010
NOR
OR
OR
Binary 011
NOR
OR
OR

Figure 4-149 *The logic for generating patterns 2 and 3.*

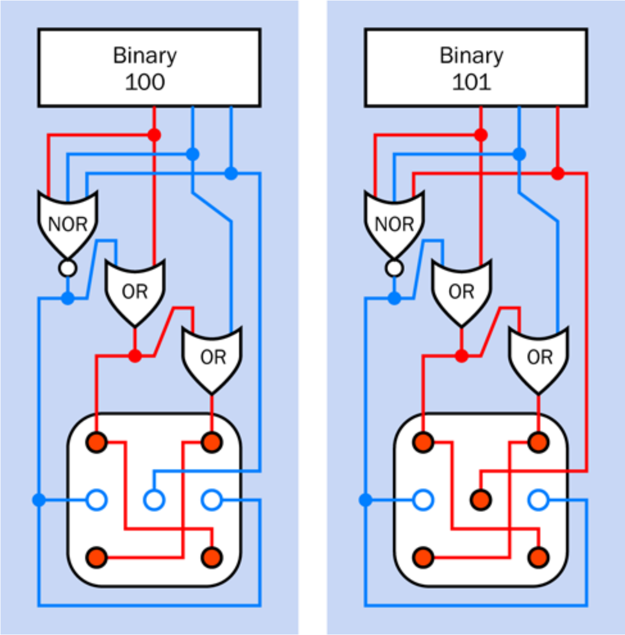
Binary 100
NOR
OR
OR
Binary 101
NOR
OR
OR

Figure 4-150 *The logic for generating patterns 4 and 5.*

If you're wondering how I came up with this selection of logic gates to translate the counter output into the dice patterns, I'm not sure I know how to tell you. There's a certain amount of trial and error, and some intuitive guesses involved in creating this kind of logic diagram. At least, that's the way it is for me. There are more rigorous and formal ways of doing it, but personally I don't find them so easy.

The Finished Circuit

The schematic in Figure 4-151 was derived from the logic diagram in Figure 4-146. The breadboarded version is shown in Figure 4-152.

The component values are shown in Figure 4-153. Note that I changed the timing resistor and timing capacitor for the 555 timer, so that it now runs around 5kHz. The idea of this circuit is that you will stop the timer at an arbitrary moment, after it has gone through hundreds of cycles. This way, you should end up with a random number.

I have added a switchable 22µF capacitor that can make the timer run slowly (about 2Hz) if you want to demonstrate how the counting works, to anyone who is skeptical.

I haven't bothered to show the values for the lower half of the breadboard, because the only components there are the chips. That's a nice aspect of building circuits based on logic: you don't have to worry about squeezing in resistors and capacitors. Chips and wires do most of the work.

The numbered outputs at the bottom of the circuit in Figure 4-151 and Figure 4-152 correspond with the inputs to the pattern of LEDs shown in Figure 4-154. There isn't room on the breadboard to add the LEDs, so you will either need a second breadboard, or you can drill some holes to mount the LEDs in a piece of plywood or plastic.

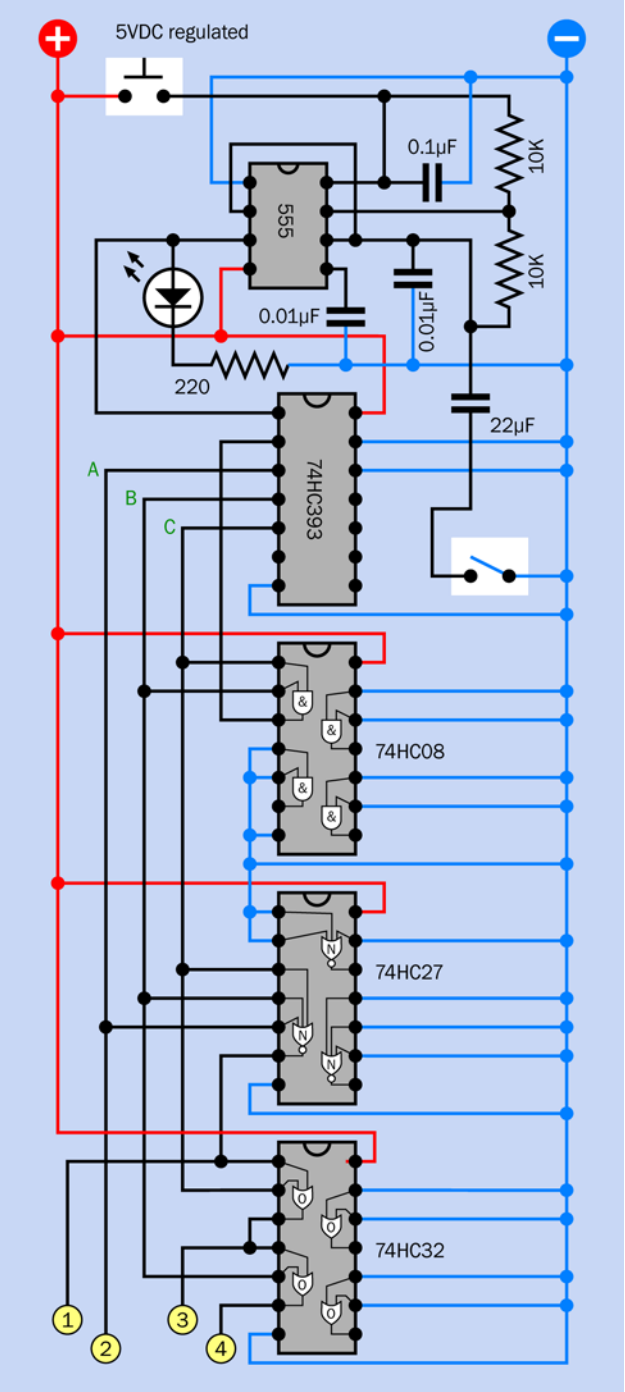

Figure 4-151 *A complete circuit for emulating the roll of a single die.*

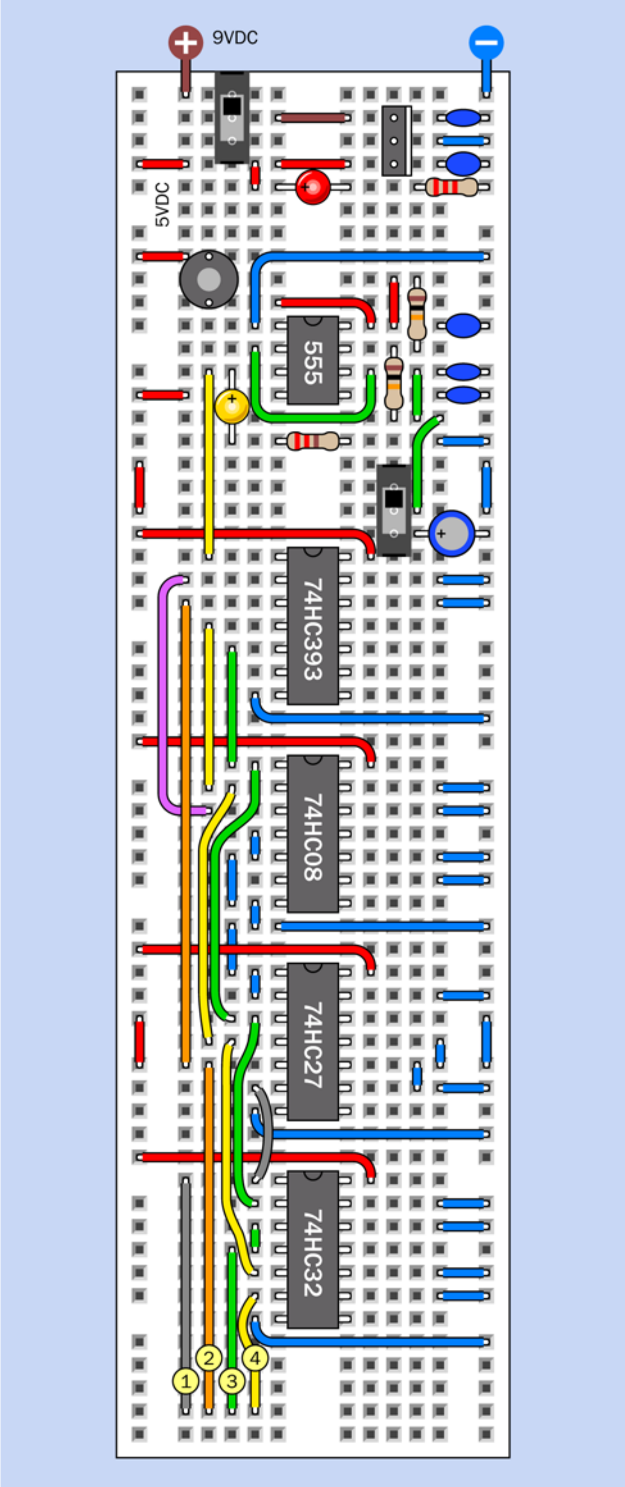
9VDC
5VDC
555
74HC393
74HC08
74HC27
74HC32
1
2
3
4

Figure 4-152 *Breadboarded version of the single-die circuit.*

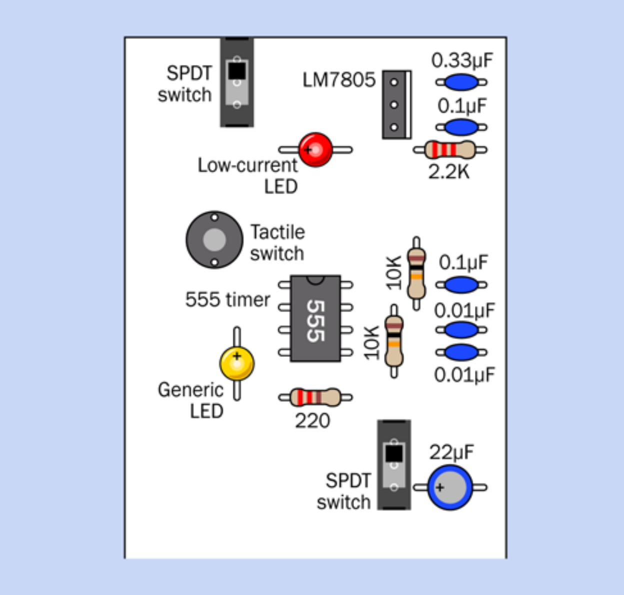
SPDT switch
LM7805
0.33µF
0.1µF
Low-current LED
2.2K
Tactile switch
10K
0.1µF
555 timer
555
0.01µF
10K
0.01µF
Generic LED
220
22µF
SPDT switch

Figure 4-153 *Values for the components in the control section of the dice simulation.*

Three pairs of the LEDs are wired in series, because a logic chip isn't powerful enough to drive a pair of LEDs in parallel. Putting them in series will require you to use a lower-value resistor than usual. The way to do this is to apply 5VDC to one of the pairs of LEDs through your meter, set to measure milliamps. Try a 220-ohm series resistor, and see how much current you measure. If you aim for a maximum of 15mA, that will be within the specifications of HC chip outputs. You may need a resistor with a value of 150 ohms or 100 ohms, depending on the characteristics of the LEDs that you are using.

Finally apply 5VDC through a 330-ohm resistor to the center LED, and compare its brightness to the LEDs wired in pairs. You may have to increase the resistor value to make the center LED visually equivalent to the others.

Hook up the LEDs to the logic circuit, hold down the button, release it, and you have your dice value (or, to be semantically correct, your die value).

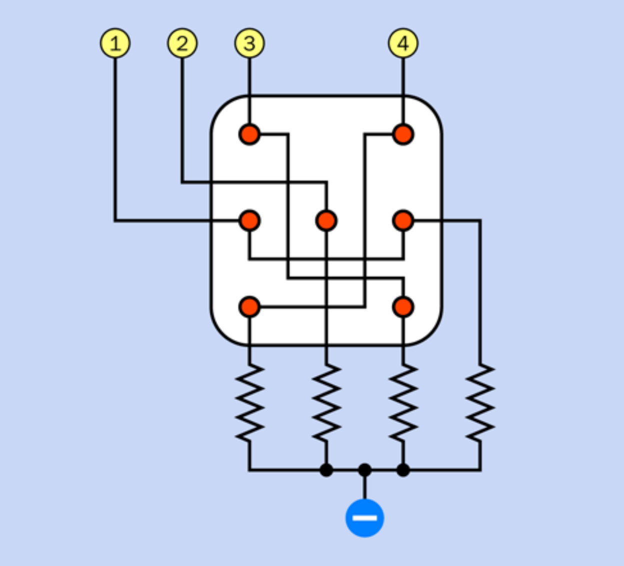

Figure 4-154 *Wiring seven LEDs (six of them in pairs, in series) to display the spot patterns of a single die.*

How do you know that it gives you a really random result? Really, the only way to be sure is to use it repeatedly and note how many times each number comes up. You might need to run it maybe 1,000 times to get decent verification. Because the circuit relies on the behavior of a human being pressing a button, there is no way to automate the testing process. All I can say is, the result really *should* be random.

Good News

There are more chips in this circuit than you've used in previous circuits in the book, but in the immortal words of Professor Farnsworth, in one of my favorite TV shows, *Futurama*: "Good news, everyone!"

The good news is that you can modify this circuit to simulate two dice instead of one, simply by adding some more wires and LEDs. You don't need any more chips at all.

We have a lot of unused logic gates in the AND, NOR, and OR chips. Three ANDs are left over, two NORs, and two ORs. In addition, there is also an entirely separate counter in the 74HC393 chip. This is exactly what you need.

The question is, how to create a second series of random numbers, different from the first. Maybe add another 555 timer, running at a different speed?

I don't like this idea, because the two timers will go in and out of phase with each other, and some pairs of values may come up more than others. I think it would be a better idea if the first counter runs from 000 through 101 binary, and then triggers the second counter to advance from 000 to 001. The first counter goes from 000 through 101 again, and triggers the second counter to move on to 010. And so on.

The second counter will run at one-sixth the speed of the first, but if you drive them fast enough, the patterns will still be too fast to see. The great advantage of this arrangement is that every possible combination of values will be displayed an equal number of times, so they should all have an almost-equal chance of coming up, just as if you were using two real dice.

Why do I say "almost equal"? Because, bear in mind there is a tiny delay when a counter resets itself from 101 binary to 000 binary. But if the lefthand counter is running at around 5kHz, a delay of less than one-millionth of a second seems trivial.

Chained Counters

The last remaining question is how the first counter can advance the second counter when it reaches 101 and flips back to 000.

This is easy. Consider what happens when the output from the first counter changes from 011 to 101 to 110, the last value lasting only for an instant before it resets to 000. After the C output reaches a high state, it drops low.

What does the clock input of the second counter need, to make it advance by one count? You already know that. It needs a high state that drops low. All you have to do is to connect the C output, from the first timer, to the clock input of the second timer. Indeed, the chip is designed to work this way, so that the falling state from one timer acts as a "carry" signal to advance the next timer.

The schematic in Figure 4-155 shows the circuit for two dice. I'm not including another breadboard picture, because you should be able to add the new wiring yourself. It is almost exactly a mirror image of the wiring that you already installed, but don't forget to move it down by one row of holes on the breadboard, to allow room for the positive power supply to each chip.

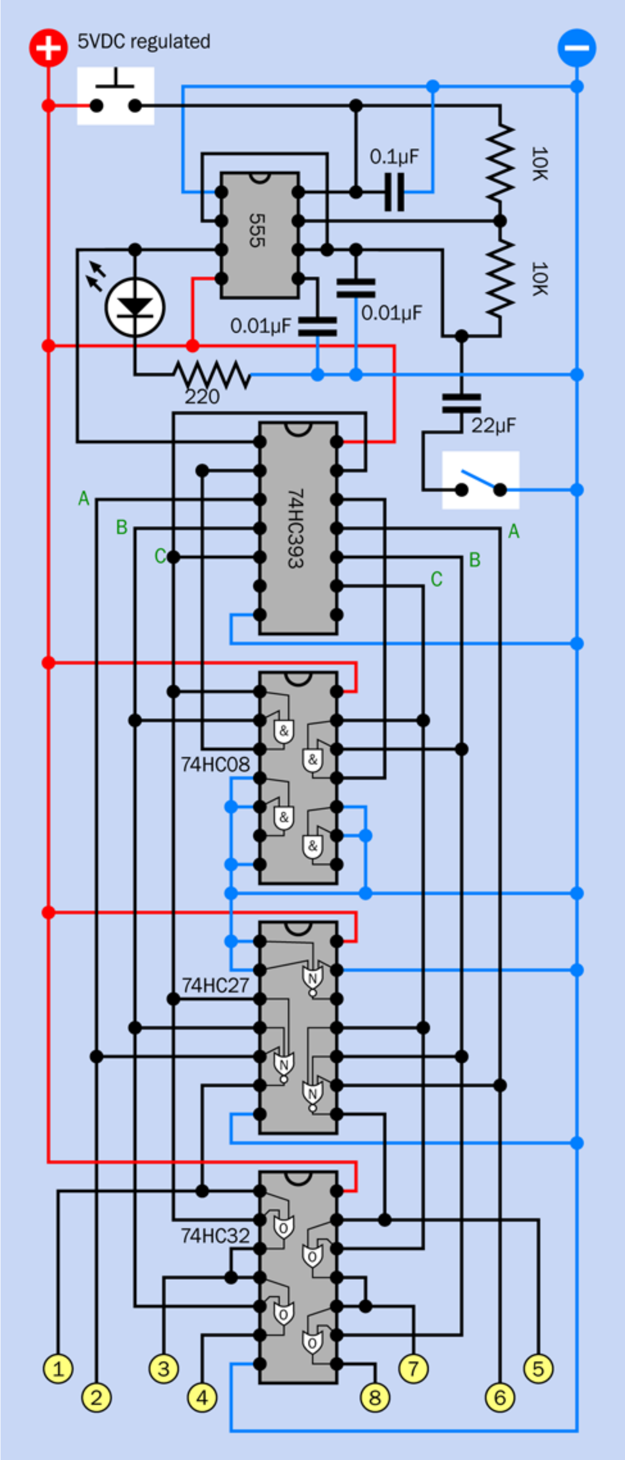

Figure 4-155 *Complete circuit to run two LED dice.*

Going Further

Could the circuit be simplified? As I mentioned at the beginning, a decade counter would require simpler logic than a binary counter. You wouldn't need an AND gate to make it count with a modulus of six, because just the seventh output pin on the decade counter could be connected back to the reset.

However, if you want to run two dice, you would need two decade counters, and that would mean two separate chips. Also, you would still need two chips to handle the logic for the two displays. To see why, search online for digital dice. At this point, you should be able to understand the schematics that will be displayed in Google Images.

The only simplification I can see, in the circuit that I have described, would be to substitute two diodes for each OR gate. The circuits that you'll find online often do this, but you will end up with a signal passing through two diodes in succession, which will drop the voltage below a level that I feel is acceptable.

The Slowdown Problem

The version of this project that I included in the first edition of *Make: Electronics* included a nice extra feature. When you took your finger off the "run" button, the die pattern display gradually slowed before it stopped. This increased the suspense of waiting to see what the final number would be.

The feature was enabled by splitting the power supply to the 555 timer. The timer was "always on" but voltage to its RC network was shut off when the player stopped pressing the "run" button. At that point, a large capacitor slowly discharged into the network, and the timer slowed as the voltage diminished.

A reader named Jasmin Patry sent email telling me that when he used the circuit, the value 1 came up disproportionately often, and he suspected that this had something to do with the slowdown feature.

Jasmin turned out to be a video game designer who understands much more about randomicity than I do. He had the polite, patient style of a man who really knew what he was talking about, and he seemed interested in helping to fix the problem that he had identified.

After he sent me graphs showing the relative frequency of each number in the simulation, I had to agree that the problem existed. I suggested many possible explanations, all of which turned out to be wrong. In the end Jasmin successfully proved that the lower power consumption of a single LED, relative to the higher power consumption of six LEDs, allowed the timer to keep running for a little longer when the voltage was marginal. This increased the odds that it would stop during that period.

Eventually Jasmin suggested a substitute circuit, in which a second 555 timer was added, and the outputs from the two timers were merged through an XOR gate. He successfully proved that this eliminated the bias toward one number. I was delighted that one of my readers had learned so much from reading my book, he was able to identify and fix a mistake that he found in it.

In this new edition, I have omitted the slowdown capacitor that caused the trouble in the first edition. But I haven't adopted Jasmin's circuit, because it was quite complicated. A single die would have needed its own pair of 555 timers, plus the XOR gate that he wanted. He also used diodes that I would have replaced with OR gates, and there wasn't nearly enough room on the breadboard.

With his permission, I will send his circuit as a freebie to anyone who registers with me (using the procedure described in the Preface—see "Me Informing You" on page 10). I cannot easily reprint it here, because I would have to redraw it completely to fit the two-column format.

Slowdown Alternatives

You'd think there must be a simpler way to make the display slow down, without affecting the randomicity. When I looked online, I found that someone had used an NPN transistor with its emitter connected to pin 7 of the timer, and a capacitor between its base and its collector, so that when power was disconnected, the transistor output would gradually diminish. Several other people had done the same thing in their dice circuits. However, I suspect that this configuration may be subject to the same problems that Jasmin found.

I have also seen circuits using exactly the same slowdown capacitor configuration that I used (for example, on the Doctronics website). I think they are almost certainly susceptible to the problem that I have described.

My final answer on this topic may be dissatisfying: I don't know how to achieve a slowing effect without adding more components that will greatly complicate the circuit.

However—just before the text of this book was finalized, my friend and fact-checker Fredrik Jansson suggested powering the 555 timer from a separate voltage regulator, to isolate it from power fluctuations in the rest of the circuit. I like this idea, but there wasn't time to try it before the book went into production.

I built a completely different dice circuit using a PICAXE microcontroller, but discovered that this had its own randomicity issues because of the imperfect random-number generator built into the chip.

In Experiment 34 (the last one in this book), you'll find that I have created yet another dice simulation, using an Arduino. But here again, I had to depend on the built-in random-number generator, and I am skeptical that it creates an evenly distributed range of numbers.

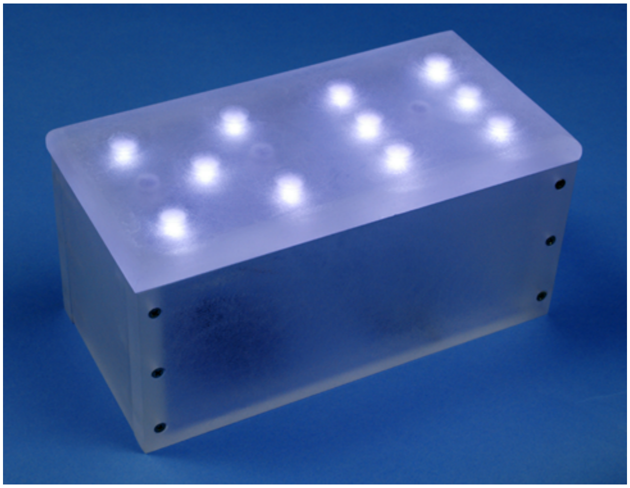

Figure 4-156 *This electronic dice display uses 10mm LEDs embedded in a box of sanded polycarbonate plastic.*

The issue of randomness is really not simple at all. I became so interested in it after my emails with Jasmin Patry, I explored it at length in *Make: More Electronics*, and also wrote a column about it in *Make:* magazine (volume 45) in collaboration with Aaron Logue, who runs his own little website describing projects that he builds. He introduced me to the concept of using a reverse-biased transistor to generate random noise, which is then processed with a clever algorithm attributed to the great computer scientist, John von Neumann. This, I think, is as close to a perfect random-number generator as you can get—but the chip count is not trivial.

All of these enhancements go beyond the range of an introductory-level book. If any readers have a *really* simple enhancement to the dice circuit presented here, to add the slowdown effect, my email inbox is always open, and yes, I do read the messages.

Meanwhile, I'm including a couple of photographs of finished electronic-dice projects. The one in Figure 4-156 was included in the first edition of this book, in 2009. Figure 4-157 shows one that I built around 1975, after Don Lancaster's amazing *TTL Cookbook* told me how I could use 74xx logic chips. Forty years later, the LEDs still light up randomly. (At least, I think they're random.)

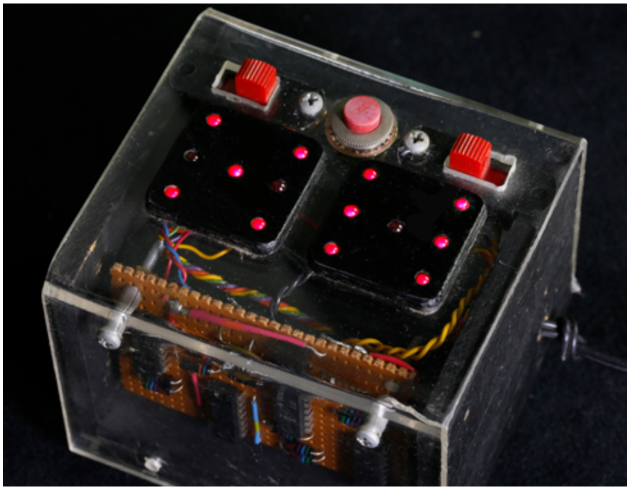

Figure 4-157 *Electronic dice designed and built around 1975, in a box of Lucite and black-painted plywood.*

What Next?

5

At this point, we can branch out in numerous directions. Here are some possibilities:

Audio: This is a large field including hobby projects such as amplifiers and "stomp boxes," to modify guitar sound.

Electromagnetism: This is a topic that I haven't even mentioned yet, but it has some fascinating applications.

Radio-frequency devices: Anything that receives or transmits radio waves, from an ultra-simple AM radio onward.

Programmable microcontrollers: These are tiny computers on a single chip. You write a little program on your desktop computer, and load it into the chip. The program tells the chip to follow a sequence of procedures, such as receiving input from a sensor, waiting for a fixed period, and sending output to a motor. Popular controllers include the Arduino, PICAXE, BASIC Stamp, and many more.

I don't have space to develop all of these topics fully, so what I'm going to do is introduce you to them by describing just a few projects in each category. You can decide which interests you the most, and then proceed beyond this book by reading other guides that specialize in that interest.

I'm also going to make some suggestions about setting up a productive work area, reading relevant books, catalogs, and other printed sources, and generally proceeding further into hobby electronics.

Tools, Equipment, Components, and Supplies

No additional tools or equipment are needed for this final chapter of the book. For a summary of all the components, see Figure 6-8. See "Supplies" on page 316 for a list of additional supplies (primarily wire, for the coils in Experiments 25, 26, 28, 29, and 31).

Customizing Your Work Area

At this point, if you're getting hooked on the fun of creating hardware but haven't allocated a permanent corner to your new hobby, I have some suggestions. Having tried many different options over the years, my main piece of advice is this: don't build a workbench!

Many hobby electronics books want you to go shopping for 2 × 4s and plywood, as if a workbench has to be custom-fabricated to satisfy strict criteria about size and shape. I find this puzzling. To me, the size and shape is not very important. I think the most important issue is storage.

I want tools and parts to be easily accessible, whether they're tiny transistors or big spools of wire. I certainly don't want to go digging around on shelves that require me to get up and walk across the room.

This leads me to two conclusions:

- You need storage around the workbench.
- You need storage below the workbench.

Many DIY workbench projects allow little or no storage underneath. Or, they suggest open shelves, which will be vulnerable to dust. My minimum configuration would be a pair of two-drawer file cabinets with a slab of 3/4-inch plywood or a Formica-clad kitchen countertop placed across them. File cabinets are ideal for storing all kinds of objects, not just files, and you can often find cheap ones at yard sales and thrift stores.

Of all the workbenches I've used, the one I liked best was an old-fashioned steel office desk—the kind of monster that dates back to the 1950s. They're difficult to move (because of their weight) and don't look beautiful, but you can buy them cheaply from used office furniture dealers, they're generous in size, they withstand abuse, and they last forever. The drawers are deep and usually slide in and out smoothly, like good file-cabinet drawers. Best of all, the desk has so much steel in it that you can use it to ground yourself before touching components that are sensitive to static electricity. If you use an anti-static wrist strap, you can simply attach it to a sheet-metal screw that you drive into one corner of the desk.

What will you put in the deep drawers of your desk or file cabinets? Some paperwork may be useful, perhaps including the following documents:

- Product datasheets
- Parts catalogs
- Sketches and plans that you draw yourself

The remaining capacity of each drawer can be filled with plastic storage boxes. The boxes can contain tools that you don't use so often (such as a heat gun or a high-capacity soldering iron), and larger-sized components (such as loudspeakers, AC adapters, project boxes, and circuit boards). You should look for storage boxes that measure around 11 inches long, 8 inches wide, and 5 inches deep, with straight sides. Boxes that you can buy at Walmart will be cheaper, but they often have tapering sides, which are not space-efficient.

The boxes that I like best are Akro-Grids, made by Akro-Mils (see Figure 5-1 and Figure 5-2). These are very rugged, with optional transparent snap-on lids. Perspective in the photographs make the boxes look as if they taper downward, but they don't. You can download the full Akro-Mills catalog online and then search online for retail suppliers. You'll find that Akro-Mils also sells an incredible variety of parts bins, but I don't like open bins because their contents are vulnerable to dust and dirt.

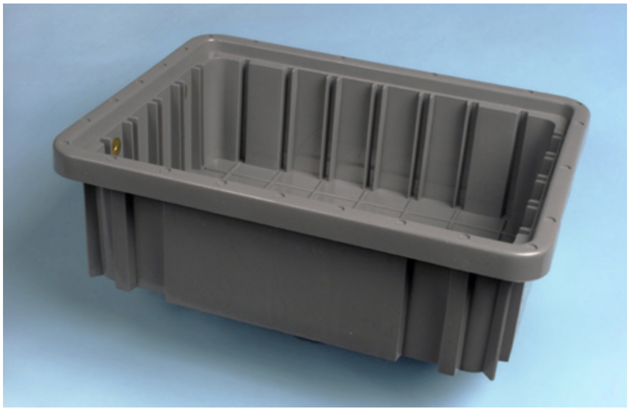

Figure 5-1 *Akro-Grid boxes contain grooves allowing them to be partitioned into numerous compartments for convenient parts storage. The height of the box in this photograph allows three to be stacked in a typical file-cabinet drawer.*

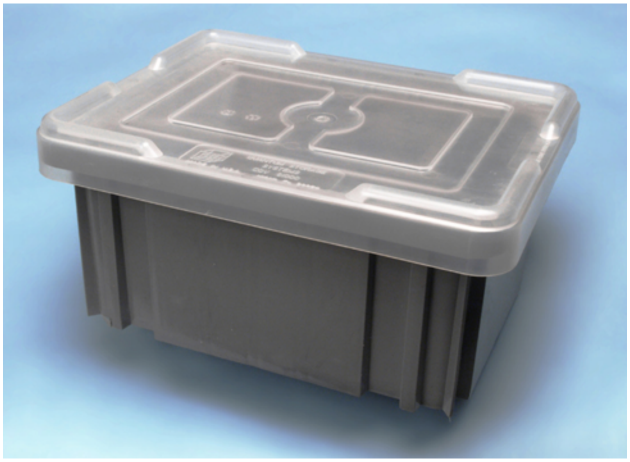

Figure 5-2 *Lids are sold separately for Akro-Grid boxes to keep the contents dust-free. The taller box shown here allows two to be stacked in a file-cabinet drawer.*

For medium-size components, such as potentiometers, power connectors, control knobs, and toggle switches, I like storage containers measuring about 11 inches long, 8 inches wide, and 2 inches deep, divided into four to six sections. You can buy these from Michaels (the craft store), but I prefer to shop online for the Plano brand, as they seem more durably constructed. The Plano products that are most suitable for medium-size electronic parts are classified as fishing-tackle boxes.

For undivided, flat-format storage boxes, the Prolatch 23600-00 is ideally sized to fit a file-cabinet drawer, and the latches are sufficiently secure that you could stack a series of them on their long edges. See Figure 5-3.

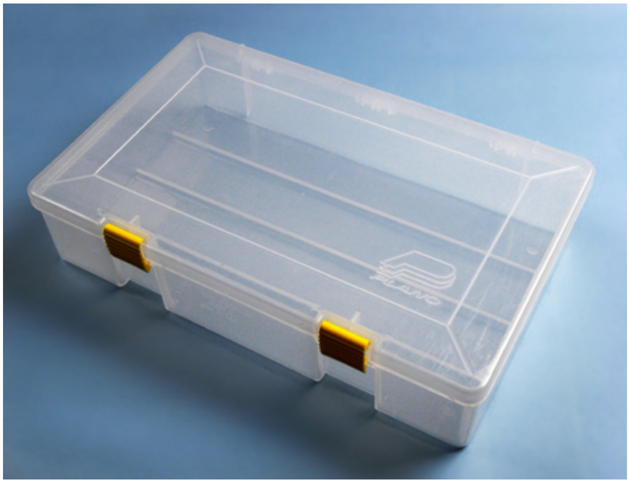

Figure 5-3 *This Plano brand box is undivided, making it useful for storing spools of wire or medium-size tools. When stacked upright on its long edge, three will fit precisely in a file-cabinet drawer.*

Plano also sells some really nicely designed toolboxes, one of which you can place on your desktop. It will have small drawers for easy access to screwdrivers, pliers, and other basics. Because you need a work area that's only about three feet square for most electronics projects, surrendering some desk space to a toolbox is not a big sacrifice.

If you have a steel desk with relatively shallow drawers, one of them can be allocated for printed catalogs. Don't underrate the usefulness of hardcopy, just because you can buy everything online. The Mouser catalog, for instance, has an index, which is more useful in some respects than their online search feature, and the catalog is divided into helpful categories. Many times I've found useful parts that I never knew existed, just by browsing, which is much quicker than flipping through PDF pages online, even with a broadband connection. Currently, Mouser is still quite generous about sending out their catalogs, which contain over 2,000 pages. McMaster-Carr will also send you a catalog, but only after you've ordered from them, and only once a year. It is probably the most comprehensive, most wonderful catalog of tools and hardware in the world.

Now, the big question: how to store all the dinky little parts, such as resistors, capacitors, and chips? I've tried various solutions to this problem. The most obvious is to buy a case of small drawers, each of which is removable, so you can place it on your desk while you access its contents. But I don't like this system, for two reasons. First, for very small components, you need to subdivide the drawers, and the dividers are never secure. And second, the removability of the drawers creates the risk of accidentally emptying the contents on the floor. Maybe you're too careful to allow this to happen, but I'm not. In fact, on one occasion I tipped the entire case of drawers onto the floor.

My personal preference is to use Darice Mini-Storage boxes, shown in Figure 5-4. You can find these at Michaels in small quantities, or buy them more economically in bulk online—just search for:

```
darice mini storage box
```

The blue boxes are subdivided into five compartments that are exactly the right size and shape for resistors. The yellow boxes are subdivided into 10 compartments, which are ideal for semiconductors. The purple boxes aren't divided at all, and the red boxes have a mix of divisions. All the boxes share the same stock number: 2505-12.

The dividers are molded into the boxes, so you don't have the annoyance associated with removable dividers that slip out of position, allowing components to mix together. The box lids fit tightly, so that even if you drop one of the boxes, it probably won't open. The lids have metal hinges, and a ridge around the edge that makes the boxes securely stackable.

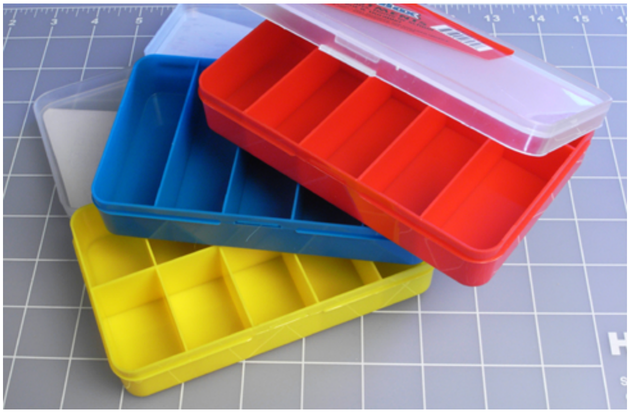

Figure 5-4 *Darice Mini-Storage boxes are ideal for components such as resistors, capacitors, and semiconductors. The boxes can be stacked stably, stored on shelves, or grouped in larger boxes. The brand sticker is easily removed after being warmed with a heat gun.*

After considerable searching, I found cheap plastic bins, with lids, measuring about 8 inches by 13 inches by 5 inches deep. Each bin will hold nine Darice parts boxes. The bins can then be categorized and stored on shelves.

Labeling

No matter which way you choose to store your parts, labeling them is essential. Any ink-jet printer will produce neat-looking labels, and if you use peelable (nonpermanent) labels, you'll be able to reorganize your parts in the future, as always seems to become necessary. I use color-coded labels for my collection of resistors, so that I can compare the stripes on a resistor with the code on the label, and see immediately if the resistor has been put in the wrong place. See Figure 5-5.

Even more important: you need to place a second (nonadhesive) label inside each compartment with the components. This label tells you the manufacturer's part number and the source, so that reordering is easy. I buy a lot of items from Mouser, and whenever I open their little plastic bags of parts, I snip out the section of the bag that has the identifying label on it, and slide it into the compartment of my parts box before I put the parts on top of it. This saves frustration later.

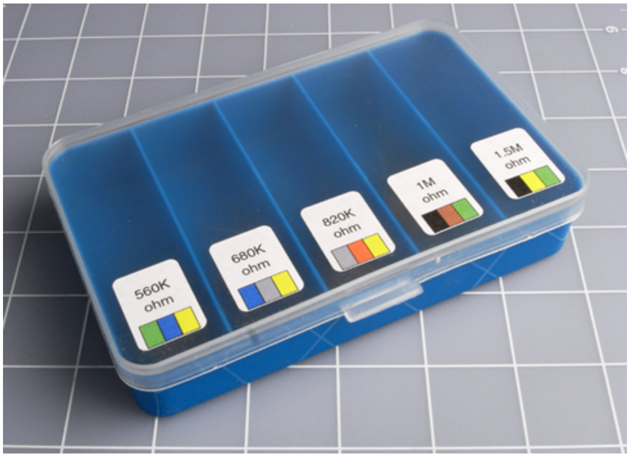

Figure 5-5 *To check that resistors are not placed in the wrong compartments, print the color code on each label.*

If I were *really* well organized, I would also keep a database on my computer listing everything that I buy, including the date, the source, the type of component, and the quantity. But I'm not that well organized.

On the Bench

Some items are so essential that they should sit on the bench or desktop on a permanent basis. These include your soldering iron(s), Helping Hand with magnifier, desk lamp, breadboard, power strip, and power supply. For a desk lamp, I prefer to use one that contains an LED bulb, for reasons explained in Experiment 14.

A power supply for your projects is a matter of personal preference. If you're serious about electronics, you can buy a unit that delivers properly smoothed current at a variety of properly regulated and calibrated voltages. Your little wall-plug AC adapter cannot do this, and its output may vary depending on how heavily you load it. Still, as you've seen, it is sufficient for basic experiments, and when you're working with logic chips, you need to mount a 5-volt regulator on your breadboard anyway. Overall, I consider a good power supply optional.

Another optional item is an oscilloscope. This will show you, graphically, the electrical fluctuations inside your wires and components, and by applying probes at different points, you can track down errors in your circuit. It's a neat gadget to own, but it will cost a few hundred dollars, and for our tasks so far, it has not been necessary. If you plan to get seriously into audio circuits, an oscilloscope becomes far more important, because you'll want to see the shapes of the waveforms that you generate.

You can try to economize on an oscilloscope by buying a unit that plugs into the USB port of your computer and uses your computer monitor to display the signal. I have tried one of these, and was not entirely happy with the results. It worked, but did not seem accurate or reliable for low-frequency signals. Maybe I was unlucky, but I decided not to try any other brands.

The surface of your desk or workbench will undoubtedly become scarred by random scuffs, cut marks, and drops of molten solder. I use a piece of half-inch plywood, two feet square, to protect my primary work area, and I clamp a miniature vise to its edge. In the past I used to cover the plywood with a square of conductive foam, to reduce the risk of static discharge from me to sensitive components. Over the years, though, I realized that my particular combination of carpet, chair, and shoes does not cause me to suffer from static. This is a matter for you to determine by experience. If you see a tiny little spark sometimes when you touch a metal object, and you feel a little zap of electricity, you need to consider grounding yourself and perhaps using anti-static foam (or a piece of metal) on your working surface.

Inevitably, during your work you'll create a mess. Little pieces of bent wire, stray screws, fasteners, and fragments of stripped insulation tend to accumulate, and can be a liability. If metal parts or fragments get into a project that you're building, they can cause short circuits. So you need a trash container. But it has to be easy to use. I use a full-size garbage pail, because it's so big that I can't miss it when I throw something toward it, and I can never forget that it's there.

Last, but most essential: a computer. Now that all datasheets are available online, and all components can be ordered online, and many sample circuits are placed online by hobbyists and educators, I don't think anyone can work efficiently without quick Internet access. To avoid wasting space, you can place a tower computer on the floor and mount its monitor on the wall—or use a tablet, or a small, cheap laptop that has a minimal footprint.

A possible workbench configuration, using a steel desk, is shown in Figure 5-6. A more space-efficient configuration is suggested in Figure 5-7.

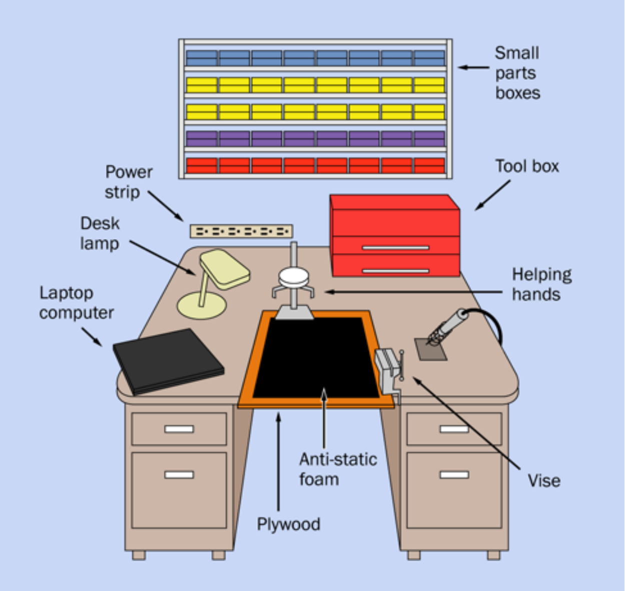

Figure 5-6 *An old steel office desk can be as good as, if not better than, a conventional workbench when building small electronics projects. It provides a large work area and ample storage, and has sufficient mass for you to ground yourself when dealing with components that are sensitive to static electricity.*

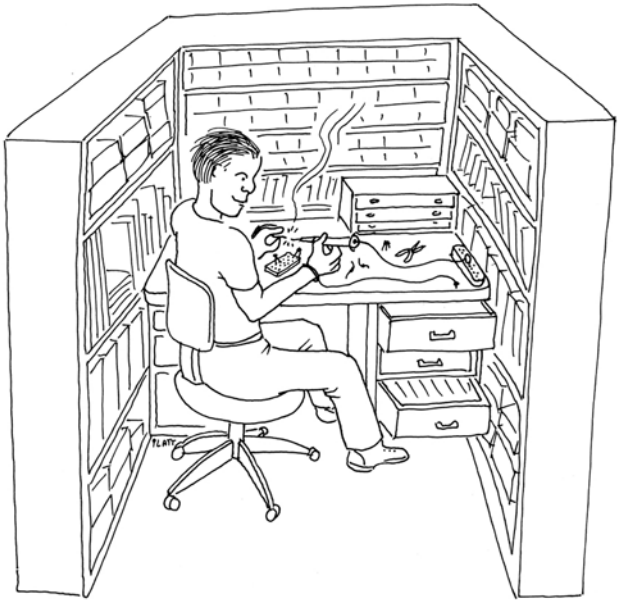

Figure 5-7 *For maximum utilization of available space, consider walling yourself in.*

Reference Sources Online

When people ask me to recommend a website offering basic information at entry level, I recommend Doctronics.

I like the way they draw their schematics, and I like the way they include many illustrations of circuits on breadboards, as I do myself. They also sell kits, if you're willing to pay and wait for shipping from the UK.

My next favorite hobby site is also British-based: the Electronics Club. It's not as comprehensive as Doctronics, but very friendly and easy to understand.

For a more theory-based approach, try ElectronicsTutorials.

This will go a little farther than the theory sections I've included here.

For an idiosyncratic selection of electronics topics, try Don Lancaster's Guru's Lair.

Lancaster wrote *The TTL Cookbook* more than 30 years ago, which opened up electronics to at least two generations of hobbyists and experimenters. He knows what he's talking about, and isn't afraid of getting into some fairly ambitious areas such as writing his own PostScript printer drivers and creating his own serial-port connections. You'll find a lot of ideas there.

Books

Yes, you do need books. A few of the ones I use are stacked up in Figure 5-8.

As you're already reading this one, I won't recommend other beginners' guides. Instead, I'll suggest some titles that will take you farther in various directions, and can be used for reference.

Make: More Electronics is the sequel that I wrote to this book, which includes all the topics (such as op-amps) for which I did not have space here. Some of the circuits are more ambitious. If you read that book, and this book, you will cover most aspects of electronics that are accessible to an individual on a moderate budget.

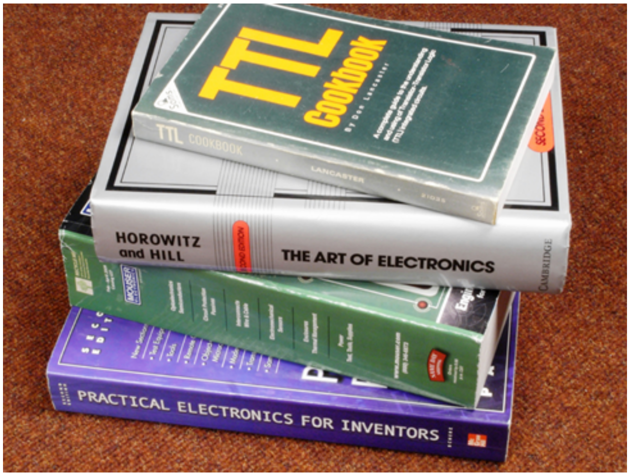

Figure 5-8 *A sun-damaged copy of Don Lancaster's classic guide to TTL chips remains at the top of my stack of reference books. It opened up a whole new era of hobby electronics more than 40 years ago. Much of its information is still useful, and copies are available from secondhand sources, including Amazon.*

The Encyclopedia of Electronic Components is a project that I began before I realized quite how demanding it would be. Consequently, its three volumes were repeatedly delayed. Volumes 1 and 2 are in print as I write this. By the time you read this, maybe Volume 3 will be in print too. The idea of these books is that they are ideal for quick reference. They can remind you of what you forgot, and they go into a lot of detail. By comparison, *Make: Electronics* is a teaching guide full of hands-on tutorials, in which I try not to get bogged down in detail.

Now let me list the books that I consider most important, written by other people:

Practical Electronics for Inventors, by Paul Scherz with Simon Monk (McGraw-Hill, Second Edition, 2013): This is a massive, comprehensive book, well worth the $40 cover price. Despite its title, you won't need to invent anything to find it useful. It's my primary reference source, covering a wide range of concepts, from the basic properties of resistors and capacitors all the way to some fairly highend math.

Getting Started with Arduino, by Massimo Banzi and Michael Shiloh (Make, 2014): This is the simplest introduction around, and will help to familiarize you with the processing language used in Arduino (similar to the C language, if you know anything about that).

Making Things Talk, by Tom Igoe (Make: Books, 2011): This ambitious and comprehensive volume shows how to make the most of the Arduino's ability to communicate with its environment, even getting it to access sites on the Internet.

TTL Cookbook, by Don Lancaster (Howard W. Sams & Co, 1974): The 1974 copyright date is not a misprint! You may be able to find some later editions, but whichever one you buy, it will be secondhand. Lancaster wrote his guide before the 74xx series of chips was emulated on a pin-for-pin basis by CMOS versions such as the 74HCxx series, but it's still a good reference, because the concepts and part numbers haven't changed, and his writing is so accurate and concise. Just bear in mind that his information about high and low logic voltages is no longer accurate.

CMOS Sourcebook, by Newton C. Braga (Sams Technical Publishing, 2001): This book is entirely devoted to the 4000 series of CMOS chips, not the 74HCxx series that I've dealt with primarily here. The 4000 series is older and must be handled more carefully, because it's more vulnerable to static electricity than the generations that came later. Still, the chips remain widely available, and their great advantage is their willingness to tolerate a wide voltage range, typically from 5 to 15 volts. This means you can set up a 12-volt circuit that drives a 555 timer, and use output from the timer to go straight into CMOS chips (for example). The book is well organized in three sections: CMOS basics, functional diagrams (showing pinouts for all the main chips), and simple circuits showing how to make the chips perform basic functions.

The Encyclopedia of Electronic Circuits, by Rudolf F. Graf (Tab Books, 1985): A totally miscellaneous collection of schematics, with minimal explanations. This is a useful book to have around if you have an idea and want to see how someone else approached the problem. Examples are often more valuable than general explanations, and this book is a massive compendium of examples. Many additional volumes in the series have been published, but start with this one, and you may find it has everything you need.

The Circuit Designer's Companion, by Tim Williams (Newnes, Second Edition, 2005): Much useful information about making things work in practical applications, but the style is dry and fairly technical. May be useful if you're interested in moving your electronics projects into the real world.

The Art of Electronics, by Paul Horowitz and Winfield Hill (Cambridge University Press, Second Edition, 1989): The fact that this book has been through 20 printings tells you two things: (1) many people regard it as a fundamental resource; (2) secondhand copies should be widely available, which is an important consideration, as the list price is over $100. It's written by two academics, and has a more technical approach than *Practical Electronics for Inventors*, but I find it useful when I'm looking for backup information.

Getting Started in Electronics, by Forrest M. Mims III (Master Publishing, Fourth Edition, 2007): Although the original dates back to 1983, this is still a fun book to have. I think I have covered many of its topics here, but you may benefit by reading explanations and advice from a completely different source, and it goes a little farther than I have into some electrical theory, on an easy-to-understand basis, with cute drawings. Be warned that it's a brief book with eclectic coverage. Don't expect it to have all the answers.

Experiment 25: Magnetism

Now that I have surveyed your future options, let me deal with a very important topic that has been waiting in the background: the relationship between electricity and magnetism. Quickly this will lead us into audio reproduction and radio, and I'll describe the fundamentals of self-inductance, which is the third and final basic property of passive components (resistance and capacitance being the other two). I left self-inductance until last because it has limited application to DC circuits. But as soon as we start dealing with analog signals that fluctuate, it becomes fundamental.

Fundamentals: A Two-Way Relationship

Electricity can create magnetism:

- When electricity flows through a wire, the electricity creates a magnetic force around the wire.

This principle is used in almost every electric motor in the world.

Magnetism can create electricity:

- When a wire moves through a magnetic field, the field creates a flow of electricity in the wire.

This principle is used in power generation. A diesel engine, or a water-powered turbine, or a windmill, or some other source of energy can turn coils of wire through a powerful magnetic field. Electricity is induced in the coils. With the exception of solar panels, all practical sources of electric power use magnets and coils of wire.

In the next experiment, you'll see a dramatic mini-demo of this effect. It should be a part of any school science class, but even if you've done it in the past, I suggest that you do it again, because setting it up takes only a matter of moments.

What You Will Need

- Large screwdriver (1)
- 22-gauge wire, or thinner (no more than 6 feet)
- 9-volt battery (1)
- Paper clip (1)

Procedure

This couldn't be simpler. Wind the wire around the shaft of the screwdriver, near its tip. The turns should be neat and tight and closely spaced, and you'll need to make 100 of them, within a distance of no more than 2 inches. To fit them into this space, you'll have to make turns on top of previous turns. If the final turn tends to unwind itself, secure it with a piece of tape.

Now apply the 9-volt battery. At first sight, this looks like a very bad idea, because you're going to short out your battery just as you did in Experiment 2. But when you pass current through a wire that's coiled instead of straight, the flow of current is inhibited (in a way that I will explain shortly), and the current does some work (such as, it can move a paper clip).

Put a small paper clip near the screwdriver blade, as shown in Figure 5-9.

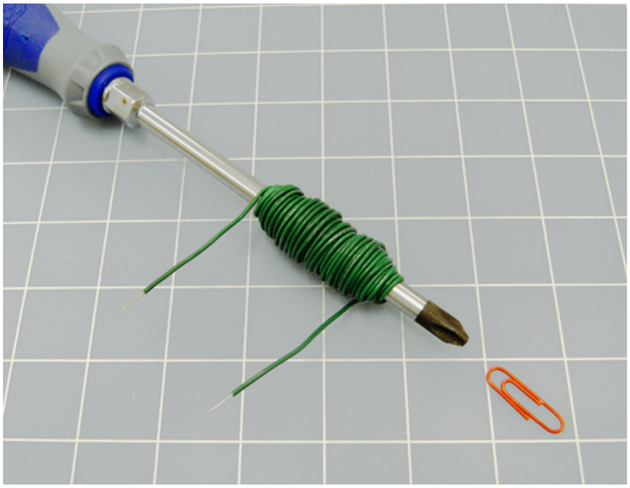

Figure 5-9 *This most basic electromagnet is just strong enough to attract a paper clip.*

The surface should be smooth, so that the paper clip can slide across it easily. Because many screwdrivers are already magnetic, you may find that the paper clip is naturally attracted to the tip of the blade. If this happens, move the clip just outside the range of attraction. Now apply the 9 volts to the circuit, and the clip should jump to the tip of the screwdriver.

Congratulations: you just made an electromagnet. The schematic is shown in Figure 5-10.

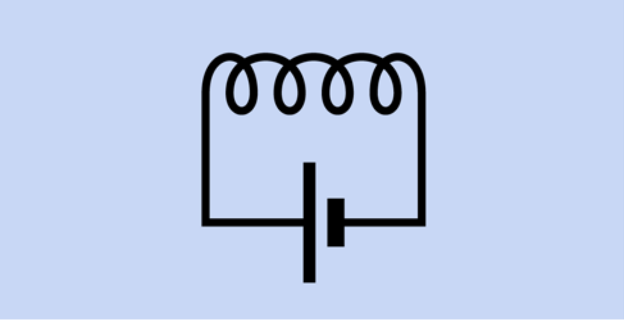

Figure 5-10 *A schematic can't get much simpler than this.*

Theory: Inductance

When electricity flows through a wire, it creates a magnetic field around the wire. Because the electricity "induces" this effect, it is known as *inductance*. This is illustrated in Figure 5-11.

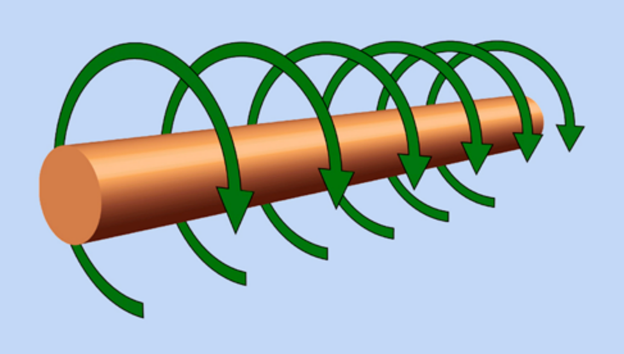

Figure 5-11 *When the flow of electricity is from left to right along this conductor, it induces a magnetic force shown by the green arrows.*

The field around a straight wire is very weak, but if we bend the wire into a circle, the magnetic force starts to accumulate, pointing through the center of the circle, as shown in Figure 5-12. If we add more circles, to form a coil, the force accumulates even more. And if we put a steel or iron object (such as a screwdriver) in the center of the coil, the effectiveness increases further.

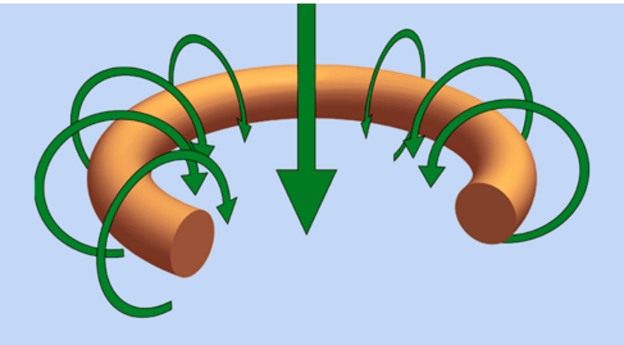

Figure 5-12 *When the conductor is bent to form a circle, the cumulative magnetic force acts through the center of the circle, as shown by the large arrow.*

Figure 5-13 shows this graphically, along with a formula known as "Wheeler's approximation," which allows you to calculate the inductance of a coil approximately, assuming you know the inner radius, the outer radius, the width, and the number of turns. (The dimensions must be in inches, not metric.) The basic unit of inductance is the henry, named after American electrical pioneer Joseph Henry. Because this is a large unit (like the farad), the formula expresses inductance in microhenries.

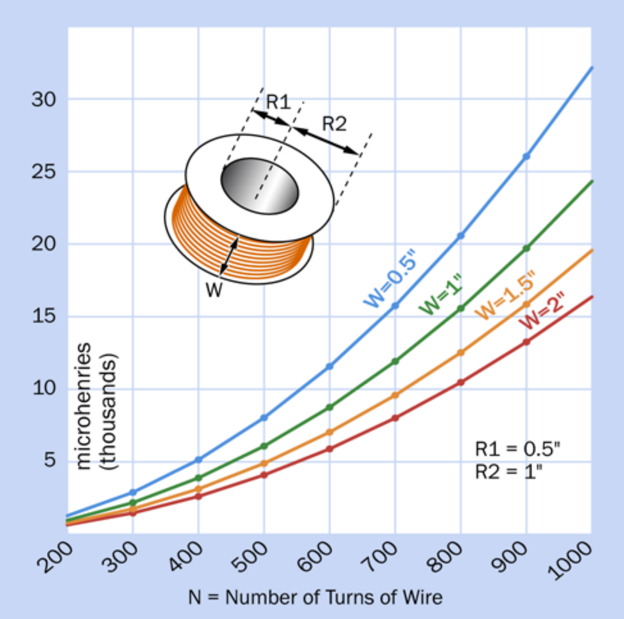

From Wheeler's approximation formula:

$$\text{Microhenries} = \frac{0.8 \times A \times A \times N \times N}{(6 \times A) + (9 \times W) + (10 \times D)}$$

Where A (average winding radius) = $\frac{R1 + R2}{2}$

D (radius difference) = R2 - R1

Figure 5-13 *A graphical representation showing how the dimensions and number of turns in a coil affect its inductance, calculated approximately with a simple formula.*

You'll see from the graphs that if you keep the basic size of a coil the same, and double the number of turns (by using thinner wire or wire with thinner insulation), the reactance of the coil increases by a factor of four. This is because the formula includes the factor N x N at the top. Here are some take-home messages:

- Inductance increases with the diameter of the coil.
- Inductance increases approximately with the square of the number of turns. (In other words, three times as many turns create nine times the inductance.)
- If the number of turns remains the same, inductance is lower if you wind the coil so that it's

slender and long, but is higher if you wind it so that it's fat and short.

Fundamentals: Coil Schematics and Basics

Check the schematic symbols for coils in Figure 5-14. Moving from left to right, either of the first two symbols represents a coil with an air core (the first symbol is older than the second). The third and fourth symbols indicate that the coil is wound around a solid iron core, or a core composed of iron particles or ferrite, respectively.

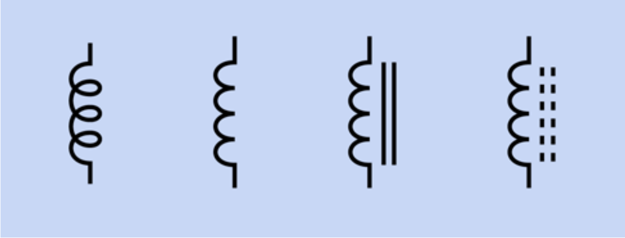

Figure 5-14 *Schematic symbols to represent coils. See text for details.*

An iron core will add to the inductance of a coil, because it increases the magnetic effect.

If you measure the magnetic field created by a coil with a positive power source at one end and negative ground at the other end, the field will reverse if you reverse the polarity of the power supply.

Perhaps the most widespread application of coils is in transformers, where alternating current in one coil induces alternating current in another, often sharing the same iron core. If the primary (input) coil has half as many turns as the secondary (output) coil, the voltage will be doubled, at half the current—assuming hypothetically that the transformer is 100% efficient.

Background: Joseph Henry

Born in 1797, Joseph Henry was the first to develop and demonstrate powerful electromagnets. He also originated the concept of "self-inductance," meaning the "electrical inertia" that is a property of a coil of wire.

Henry started out as the son of a day laborer in Albany, New York. He worked in a general store before being apprenticed to a watchmaker, and was interested in becoming an actor. Friends persuaded him to enroll at the Albany Academy, where he turned out to have an aptitude for science. In 1826, he was appointed Professor of Mathematics and Natural Philosophy at the Academy, even though he was not a college graduate and described himself as being "principally self-educated." Michael Faraday was doing similar work in England, but Henry was unaware of it.

Henry was appointed to Princeton in 1832, where he received $1,000 per year and a free house. When Morse attempted to patent the telegraph, Henry testified that he was already aware of its concept, and indeed had rigged a system on similar principles to signal his wife, at home, when he was working in his laboratory at the Philosophical Hall.

Henry taught chemistry, astronomy, and architecture, in addition to physical science, and because science was not divided into strict specialties as it is now, he investigated phenomena such as phosphorescence, sound, capillary action, and ballistics. In 1846, he headed the newly founded Smithsonian Institution as its secretary. His photograph appears in Figure 5-15.

Figure 5-15 *Joseph Henry was an American experimenter who pioneered the investigation of electromagnetism. This photograph is archived in Wikimedia Commons.*

Experiment 26: Tabletop Power Generation

In Experiment 5, you saw that chemical reactions can generate electricity. Now it's time to see electricity generated by a magnet.

What You Will Need

- Wire cutters, wire strippers, test leads, multimeter
- Cylindrical neodymium magnet, 3/16" diameter by 1.5" long, axially magnetized (1)
- Hookup wire, 26-gauge, 24-gauge, or 22-gauge, total 200 feet
- Low-current LED (1)
- Capacitor, 1,000µF (1)
- Switching diode, 1N4001 or similar (1)

Optional extras:

- Cylindrical neodymium magnet, 3/4" diameter by 1" long, axially magnetized (1)
- Half-inch diameter wooden dowel, 6" long (minimum)
- Steel screw, #6 size with flat head
- PVC water pipe, 3/4" internal diameter, 6" long (minimum)
- Two pieces of 1/4" plywood, each about 4" x 4" (you will need a 1" hole saw or Forstner bit to drill a hole through the plywood)
- Spool of magnet wire, quarter-pound, 26-gauge, about 350 feet (1)

Procedure

First, you need a magnet. Neodymium magnets are the strongest available, and are fairly cheap if you choose the small cylindrical type. A magnet of just 3/16" diameter and 1.5" long will be sufficient. Wrap about 10 turns of 22-gauge wire tightly around it, as shown in Figure 5-16. Now allow the wire to loosen slightly, so that the magnet can slide through the coils.

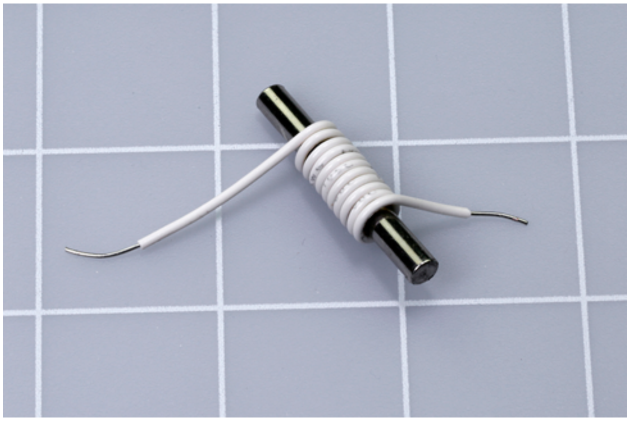

Figure 5-16 *Just ten turns of wire can be sufficient to create a small electrical potential when a magnet moves through them.*

Set your meter to measure millivolts AC (not DC, because we're going to be dealing with alternating pulses of electricity). Strip a little insulation from each end of the coil, and use alligator test leads to attach the meter. Grasp the magnet between finger and thumb, and shuttle it quickly to and fro inside the coil. I'm guessing you'll see a value of 3mV to 5mV on your meter. Yes, this small magnet, and 10 turns of wire, can generate a few millivolts.

Try winding a bigger coil, with the layers overlapping, as shown in Figure 5-17. Move the magnet quickly again. You should find that you are generating more voltage.

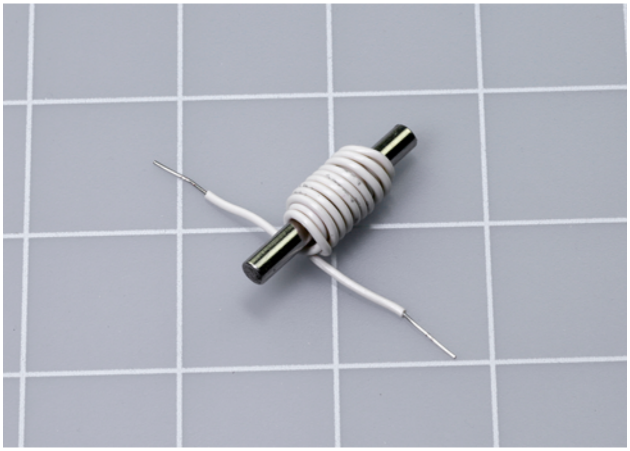

Figure 5-17 *Adding more turns of wire will increase the measured voltage when the magnet moves through them.*

Remember the formula from the previous experiment, in which I showed how passing electricity through more coils of wire would induce a stronger magnetic field. It works both ways:

- More coils of wire will generally induce a higher voltage, when a magnet moves through the coils.

This leads me to wonder—if we had a bigger, stronger magnet and a *lot* of turns of wire, could we generate enough electricity to power something, such as, maybe, an LED?

Lighting an LED

I'm going to use 22-gauge wire, because you already have it for the other experiments. The trouble is, it is relatively thick and has thick insulation. Two hundred turns of this wire really begin to bulk up. This is why we should be using *magnet wire*, which is pure copper wire with an ultra-thin coating of insulation made from shellac or plastic film. Magnet wire is designed to pack as densely as possible.

However, you may not feel like spending the money for a spool of magnet wire, bearing in mind that you are unlikely to find any other use for it. So, I decided to see if 22-gauge hookup wire would be workable in this experiment. The answer is yes, but only just.

You do need 200 feet. That will cost a bit of money, but you can always reuse the wire for normal purposes, such as creating jumpers for breadboards.

You can join two or more pieces of wire together when you are winding a spool, and so long as you twist the stripped ends tightly, you don't need to solder them.

You also need a more powerful magnet. The smallest that worked for me is cylindrical, measures 1″ long and 3/4″ in diameter, and is axially magnetized, meaning that the north and south poles are at opposite ends of its axis. (The axis is an imaginary line that runs through the center of the cylinder, parallel with its curved sides. You can imagine the cylinder like a shaft rotating around its axis.)

The equipment that I ended up with is shown in Figure 5-18. The magnet is at the right-hand side. The spool I made from 1/4″ plywood, and it is just over four inches in diameter. A piece of 3/4″ PVC water pipe runs through the center, and its inside diameter is just a fraction wider than the diameter of the magnet, so that the magnet can slide through it freely.

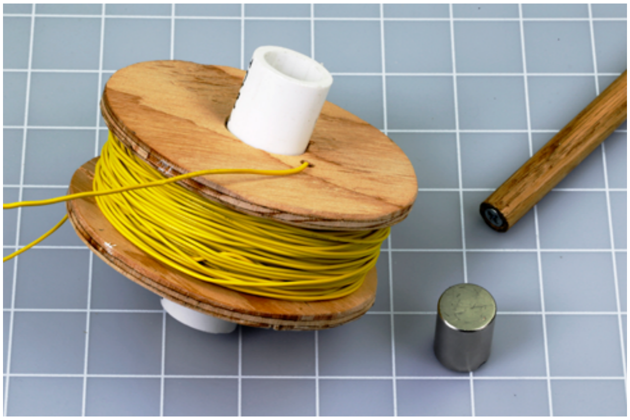

Figure 5-18 *Two hundred turns of 22-gauge wire on a homemade spool, with a magnet that attaches itself to a screw in a wooden dowel.*

Push-fit the plywood circles onto the pipe, to create a spool. Now you need to wind 200 feet of wire onto the spool—taking care to leave access to the inside end of the wire. I drilled a small hole in one of the plywood circles, near the center, and poked the wire out through the hole.

The width of the coil that you are going to wind should be about the same as the length of the magnet, and the magnet inside the tube should be able to emerge completely on either side of the coil. The cross-section of the spool in Figure 5-19 shows what I mean.

To hold the magnet conveniently, I drilled a hole in one end of a piece of 1/2″ wooden dowel, and inserted a #6 1″ flat-headed screw. I was then able to hold the dowel like a handle, while the magnet attracted itself firmly to the screw.

Now for the big moment. Use a couple of alligator test leads to attach the ends of your coil to the inputs of your meter, and set the meter to AC volts, as you did before. This time, though, set it to measure up to 2 volts.

With the magnet attached to the dowel, you can push it as quickly as possible in and out of the PVC pipe. Alternatively, remove the magnet from the dowel, drop it into the pipe, and rattle it up and down with your finger and thumb over the ends of the pipe. If you really work hard, your meter should show a voltage of around 0.8V.

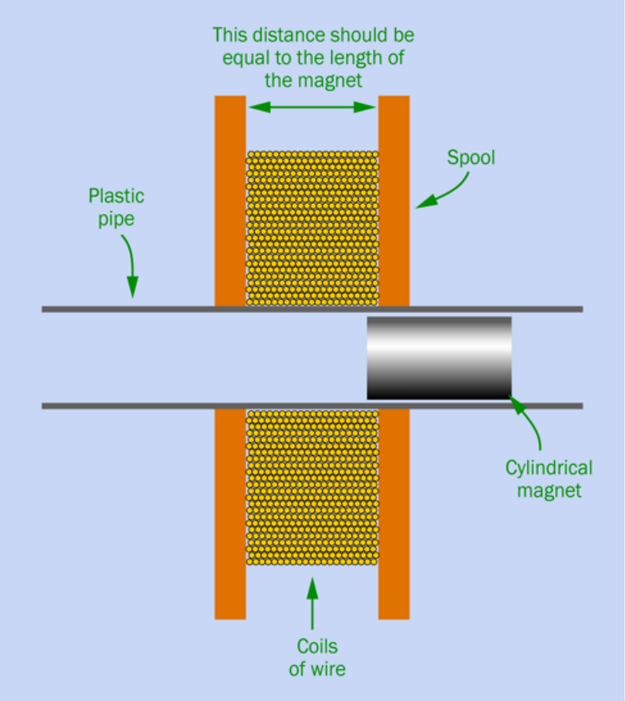

Figure 5-19 *Setup for generating just enough power to light an LED.*

You took all that trouble, and you got less than a volt?

Ah, but your meter is *averaging* the current. Each pulse is probably peaking at a higher voltage.

Disconnect your test leads from the meter, and attach them to a low-current LED. Clamp the LED so that it doesn't jostle around. Now when you move the magnet vigorously, I think you will see the LED flicker. If it doesn't, reverse the orientation of the magnet in the tube and try again. You really need a low-current LED to make this work.

Optional Extensions

If you are willing to spend a little more money, you can get much more impressive results.

First, use a bigger magnet. I get excellent results from one that is 2″ long and 5/8″ in diameter. Of course, you'll need a larger diameter of PVC pipe to accommodate the magnet.

Second, buy a spool of proper magnet wire. I used about 500 feet of 26-gauge wire. It's easy to buy online; there are dozens of suppliers.

If you're fortunate, your magnet wire will be supplied on a plastic spool with a hole in the middle just a little bit bigger than the diameter of your magnet. And, better still, the spool of magnet wire will allow you access to the "tail" of the wire sticking out in the center of the spool, as shown circled in red in Figure 5-20.

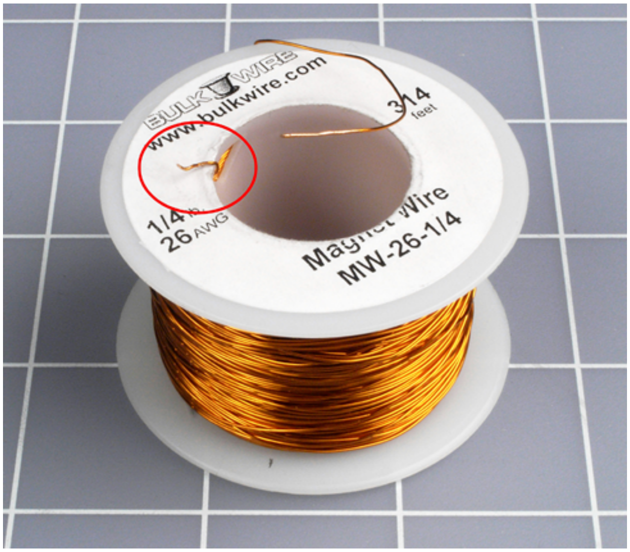

Figure 5-20 *A spool of magnet wire with the inside end accessible, circled in red.*

To remove the thin film of insulation from the ends of the magnet wire, you can scrape them very gently with a knife blade, or rub them with fine sandpaper. Check with a magnifying glass to make sure that some insulation has been removed. You can also apply your meter to check the resistance, which should be less than 100 ohms.

Now you can attach an LED to each end of the magnet wire on the spool, and generate voltage by pushing your magnet in and out of the center of the spool, as shown in Figure 5-21.

If the spool is the wrong size, or if the tail of the wire isn't accessible, you'll just have to rewind the wire from one spool to another. Suppose you have 500 feet; that will entail rewinding about 2,000 turns. If you can make four turns per second, you'll require 500 seconds—a little less than 10 minutes, which I think is tolerable.

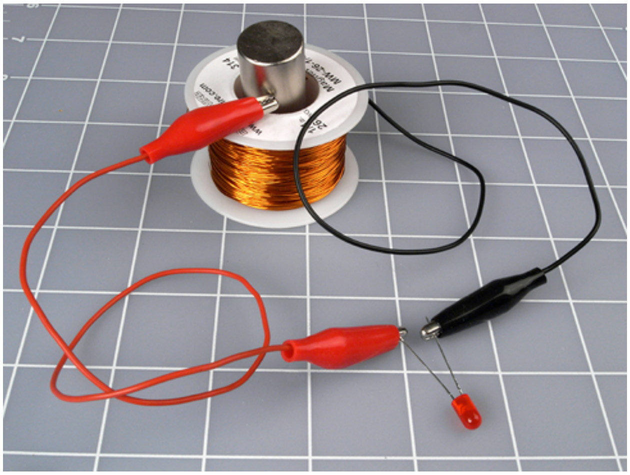

Figure 5-21 *Ready for power generation, on a rather small scale.*

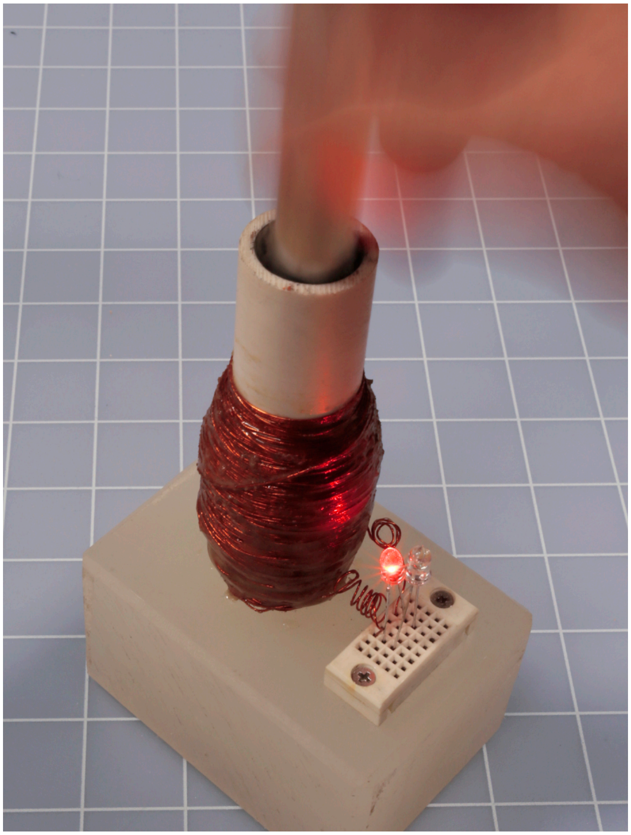

Figure 5-22 *Demonstration device capable of dazzling results.*

Figure 5-22 shows a larger-scale device that I built for demonstration purposes. The coil of magnet wire is coated with epoxy glue, so that it won't unravel, and I mounted the pipe in a block of plastic that holds it securely. My neodymium magnet attaches itself to a steel screw in to the end of an aluminum rod, also visible in the photograph.

I added two high-intensity LEDs to the coil, with their polarity in opposite directions. When the magnet shuttles up and down, the LEDs light up the whole room. Also, their opposite polarity shows that voltage travels through the coil in one direction on the upstroke, and in the other direction on the downstroke. See Figure 5-23.

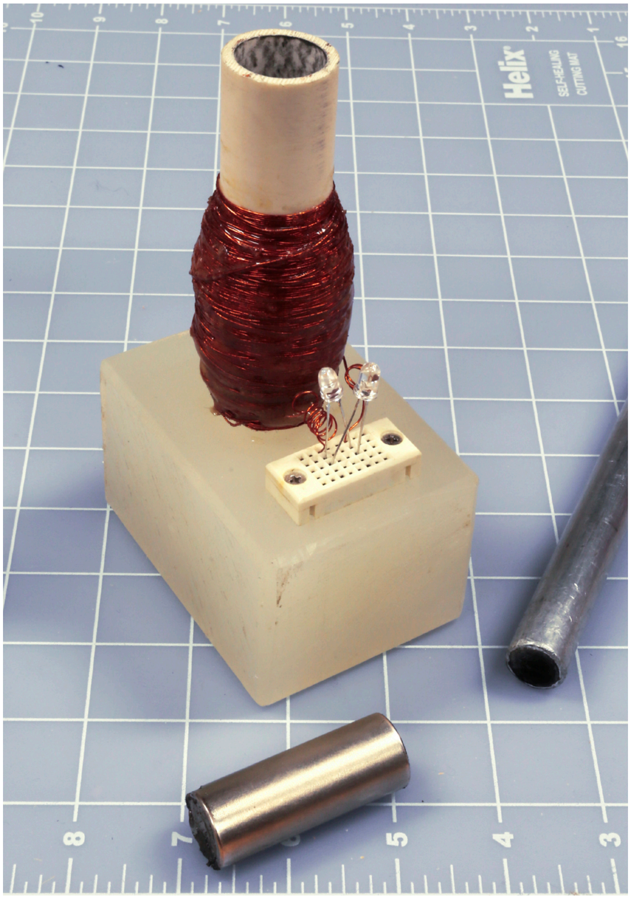

Figure 5-23 *The LED generator in action.*

Caution: Blood Blisters and Dead Media

Beware of the spooky capabilities of neodymium.

Neodymium magnets are breakable. They're brittle and can shatter if they slam against a piece of magnetic metal (or another magnet). For this reason, many manufacturers advise you to wear eye protection.

You can easily pinch your skin and get *blood blisters* (or worse). Because a magnet pulls with increasing force as the distance between it and another object gets smaller, it closes the final gap very suddenly and powerfully. Ouch!

Magnets never sleep. In the world of electronics, we tend to assume that if something is switched off, we don't have to worry about it. Magnets don't work that way. They are always sensing the world around them, and if they notice a magnetic object, they want it, *now*. Results may be unpleasant, especially if the object has sharp edges and your hands are in the way. When using a magnet, create a clear area on a nonmagnetic surface, and watch out for magnetic objects underneath the surface. For example, my magnet sensed a steel screw embedded in the underside of a kitchen countertop, and slammed itself into contact with the countertop unexpectedly.

It's hard to take this seriously until it happens to you. But, seriously, neodymium magnets don't fool around. Proceed with caution.

Also, remember that *magnets create magnets.* When a magnetic field passes across an iron or steel object, the object picks up some magnetism of its own. If you wear a watch, be careful not to magnetize it. If you use a smartphone, keep it away from magnets. Likewise, any computer or disk drive is vulnerable. The magnetic stripe on a credit card is easily erased. Also keep magnets well away from TV screens and video monitors (especially cathode-ray tubes). Last but not least, powerful magnets can interfere with the normal operation of cardiac pacemakers.

Charging a Capacitor

Here's another thing to try. Disconnect the LED from whatever coil of wire you created, and connect a 1,000µF electrolytic capacitor in series with a 1N4001 signal diode, as shown in Figure 5-24. Attach your meter, measuring DC volts (not AC, this time), across the capacitor.

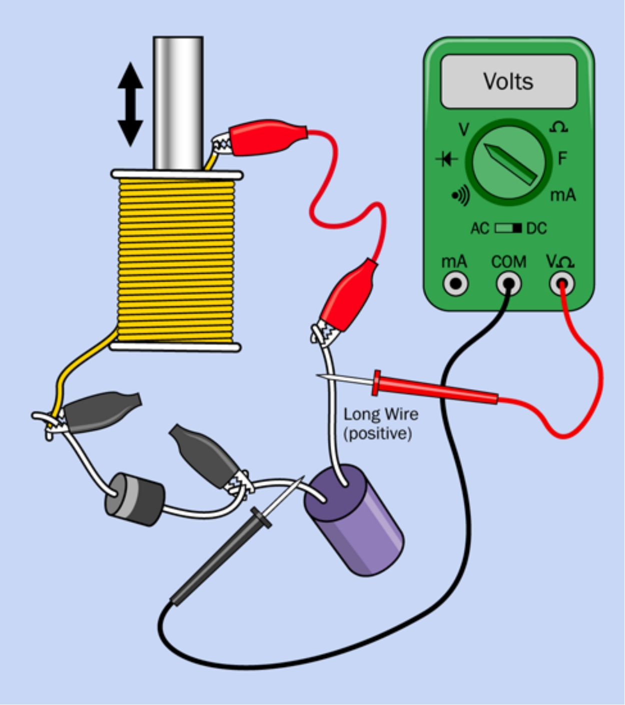

Figure 5-24 *A diode enables you to accumulate voltage from your coil in a capacitor.*

If your meter has a manual setting for its range, set it to at least 2VDC. Make sure the positive (unmarked) side of the diode is attached to the negative (marked) side of the capacitor, so that positive voltage will pass through the capacitor and then through the diode.

Now move the magnet vigorously up and down in the coil. The meter should show that the capacitor is accumulating charge. When you stop moving the magnet, the voltage reading may decline very slowly, mostly because the capacitor discharges itself through the internal resistance of your meter.

This experiment is more important than it looks. Bear in mind that when you push the magnet into the coil, it induces current in one direction, and when you pull it back out again, it induces current in the opposite direction. You are actually generating alternating current.

The diode only allows current to flow one way through the circuit. It blocks the opposite flow, which is how the capacitor accumulates its charge. If you jump to the conclusion that diodes can be used to convert alternating current to direct current, you're absolutely correct. We say that the diode is "rectifying" the AC power.

Next: Audio

Experiment 25 showed that voltage can create a magnet. Experiment 26 has shown that a magnet can create voltage. We're now ready to apply these concepts to the detection and reproduction of sound.

Experiment 27: Loudspeaker Destruction

You saw that electricity running through a coil can create enough magnetic force to pull a small metallic object toward it. What if the coil is very light, and the object is heavier? In that case, the coil can be pulled toward the object. This principle is at the heart of a loudspeaker.

To understand how a loudspeaker works, there's really no better way than to disassemble it. Maybe you'd prefer not to spend a few dollars on this destructive but educational process—in which case, you might consider picking up a piece of nonfunctional audio equipment at a yard sale, and pulling a loudspeaker out of that. Or simply take a look at my photographs illustrating the process step by step.

What You Will Need

- Cheapest possible loudspeaker, 2" minimum (1)
- Utility knife (1)

Procedure

Figure 5-25 shows a small loudspeaker seen from the rear. A magnet is hidden in the sealed cylindrical section.

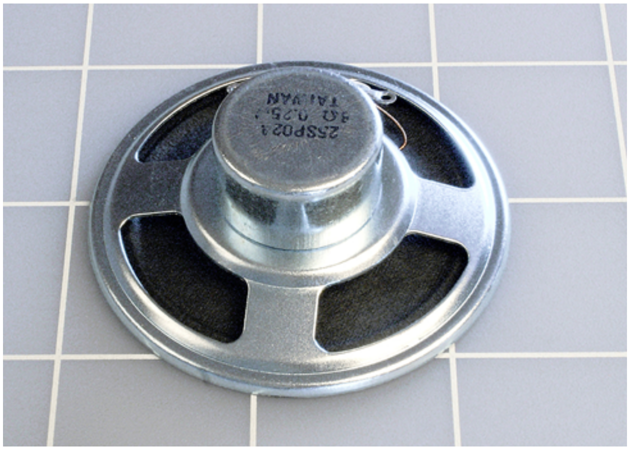

Figure 5-25 *The back of a small loudspeaker.*

Turn the loudspeaker face-up, as shown in Figure 5-26. Cut around the perimeter of its cone with a sharp utility knife or X-Acto blade. Then cut around the circular center and remove the O-shaped circle of black paper that you've created.

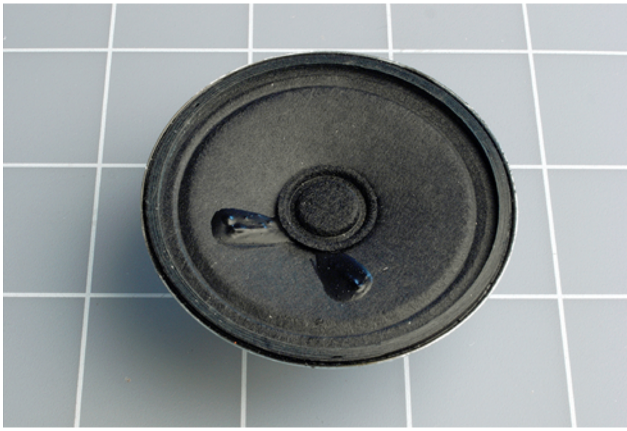

Figure 5-26 *A two-inch loudspeaker ready for its fate.*

The unconed speaker is shown in Figure 5-27. The yellow weave at the center is the flexible section that normally allows the cone to move in and out, while preventing it from deviating from side to side.

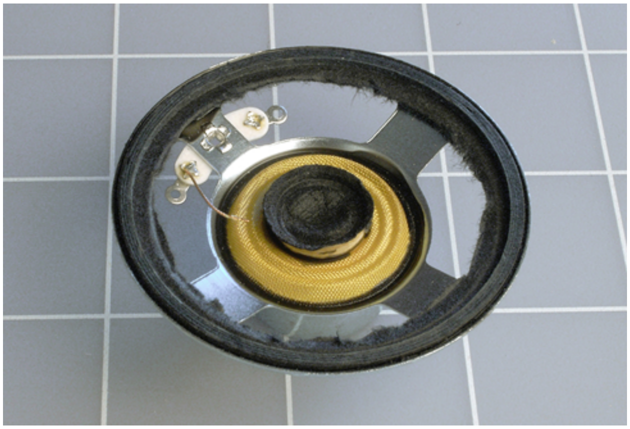

Figure 5-27 *The loudspeaker with its cone removed.*

Cut around the outside edge of the yellow weave, and you should be able to pull up a hidden paper cylinder, which has a copper coil wound around it, as shown in Figure 5-28. In the photograph, I've turned it over so that it is easily visible.

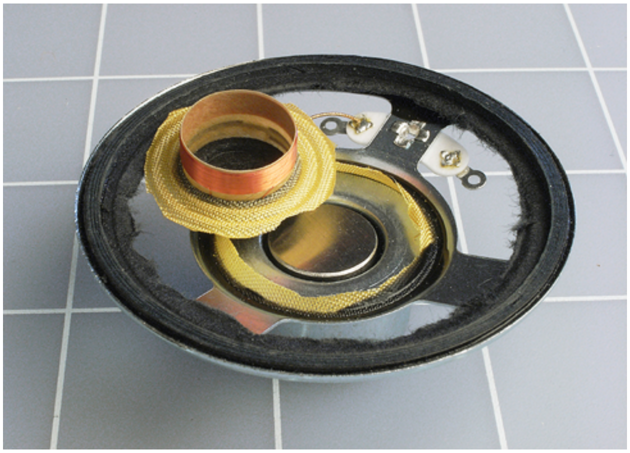

Figure 5-28 *The copper coil is normally hidden inside the groove of the magnet, below.*

The two ends of this copper coil normally receive power through flexible wires from two terminals at the back of the speaker. When the coil sits in the groove visible in the magnet, the coil reacts to voltage fluctuations by exerting an up-and-down force in reaction to the magnetic field. This vibrates the cone of the loudspeaker and creates sound waves.

Large loudspeakers in your stereo system work exactly the same way. They just have bigger magnets and coils that can handle more power (typically, as much as 100 watts).

Whenever I open up a small component like this, I'm impressed by the precision and delicacy of its parts, and the way it can be mass-produced for such a low cost. I imagine how astonished Faraday, Henry and the other pioneers of electrical research would be, if they could see the components that we take for granted today. Henry spent days winding coils by hand to create electromagnets that were far less efficient than this cheap little loudspeaker.

Background: Origins of Loudspeakers

As I mentioned at the beginning of this experiment, a coil will move if its magnetic field interacts with a heavy or fixed object. If the object is a permanent magnet, the coil will interact with it more strongly, creating more vigorous motion. This is how a loudspeaker works.

The idea was introduced in 1874 by Ernst Siemens, a prolific German inventor. (He also built the world's first electrically powered elevator in 1880.) Today, Siemens AG is one of the largest electronics companies in the world.

When Alexander Graham Bell patented the telephone in 1876, he used Siemens' concept to create audible frequencies in the earpiece. From that point on, sound-reproduction devices gradually increased in quality and power, until Chester Rice and Edward Kellogg at General Electric published a paper in 1925 establishing basic principles that are still used in loudspeaker design today.

At sites such as Radiola Guy, you'll find photographs of very beautiful early loudspeakers, which used a horn design to maximize efficiency, as shown in Figure 5-29. As sound amplifiers became more powerful, speaker efficiency became less important compared with quality reproduction and low manufacturing costs. Today's loudspeakers convert only about 1% of electrical energy into acoustical energy.

Amplion small horn
radio speaker
model AR-114
RadiolaGuy.com
AMPLION

Figure 5-29 *This beautiful Amplion AR-114x illustrates the efforts of early designers to maximize efficiency in an era when the power of audio amplifiers was very limited. Photos by "Sonny, the RadiolaGuy." Many early speakers are illustrated at http:// www.radiolaguy.com. Some are for sale.*

Theory: Sound, Slectricity, and Sound

Time now to establish a more specific idea of how sound is transformed into electricity and back into sound again.

Suppose someone bangs a gong with a stick, as shown in Figure 5-30. The flat metal face of the gong vibrates in and out, creating pressure waves that the human ear perceives as sound. Each wave of high air pressure is followed by a trough of lower air pressure, and the wavelength of the sound is the distance (usually ranging from meters to millimeters) between one peak of pressure and the next.

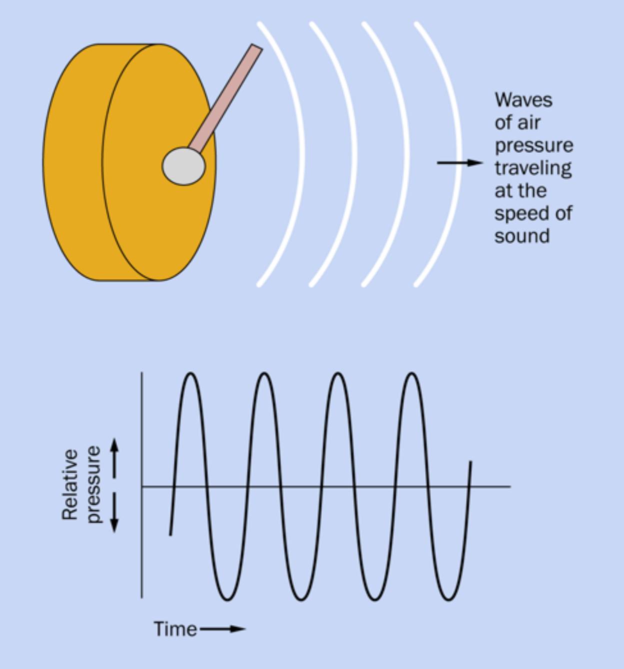

Figure 5-30 *Striking a going makes its flat surface vibrate. The vibrations create waves of pressure in the air.*

The frequency of the sound is the number of waves per second, usually expressed as hertz.

Suppose we put a very sensitive little membrane of thin plastic in the path of the pressure waves. The plastic will flutter in response to the waves, like a leaf fluttering in the wind. Suppose we attach a tiny coil of very thin wire to the back of the membrane so that it moves with the membrane. And let's position a stationary magnet inside the coil of wire. This configuration is like a tiny, ultra-sensitive loudspeaker, except that instead of electricity producing sound, the sound will produce electricity. Pressure waves make the membrane oscillate along the axis of the magnet, and the magnetic field creates a fluctuating voltage in the wire. The principle is illustrated in Figure 5-31.

This is known as a *moving-coil* microphone. There are other ways to build a microphone, but this is the configuration that is easiest to understand. Of course, the voltage that it generates is very small, but we can amplify it using a transistor, or a series of transistors, as suggested in Figure 5-32.

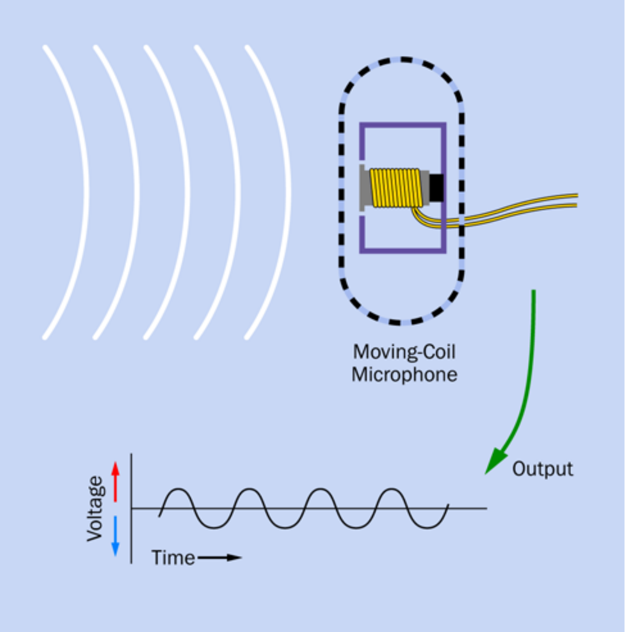
Moving-Coil Microphone
Output
Voltage
Time

Figure 5-31 *Sound waves entering a moving-coil microphone make a membrane vibrate. The membrane is attached to a coil on a sleeve around a magnet. Motion of the coil induces small currents.*

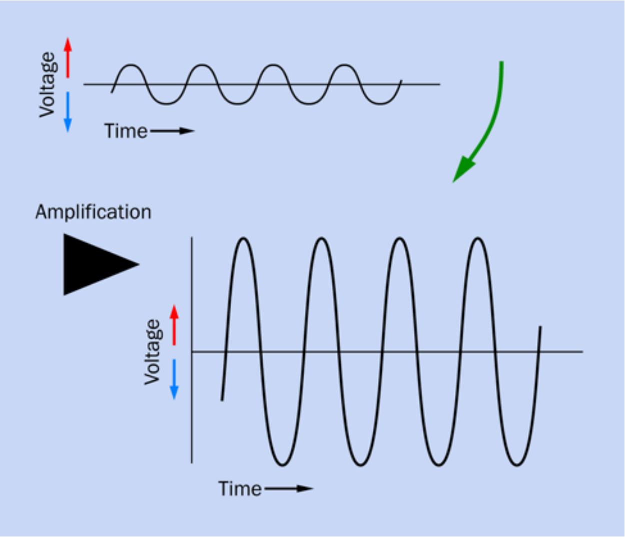
Voltage
Time
Amplification
Voltage
Time

Figure 5-32 *Tiny signals from the microphone pass through an amplifier, which enlarges their amplitude while retaining their frequency and the shape of their waveform.*

Then we can feed the output through the coil around the neck of a loudspeaker, and the loudspeaker will recreate the pressure waves in the air, as shown in Figure 5-33.

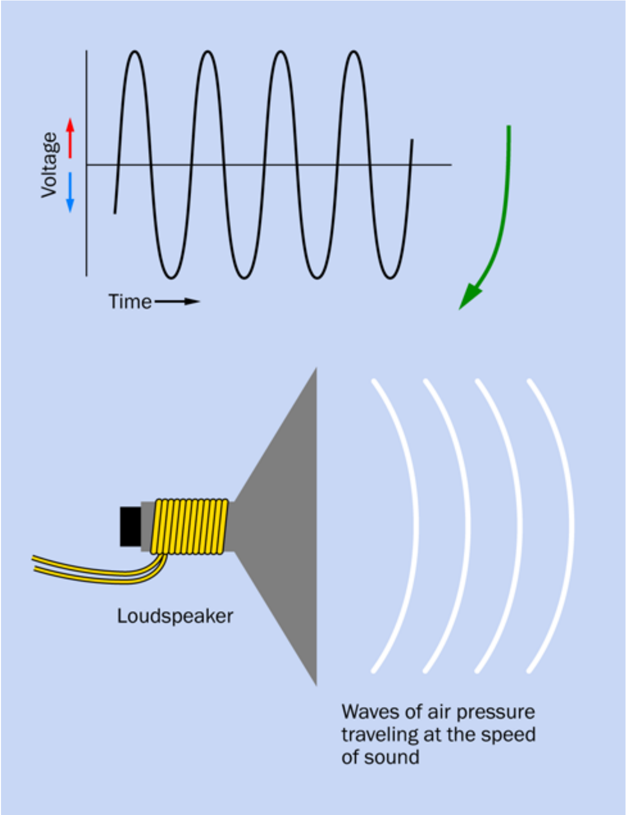
Voltage
Time
Loudspeaker
Waves of air pressure traveling at the speed of sound

Figure 5-33 *The amplified electrical signal is passed through a coil around the neck of a loudspeaker cone. The magnetic field induced by the current causes the cone to vibrate, reproducing the original sound.*

Somewhere along the way, we may want to record the sound and then replay it. But the principle remains the same. The hard part is designing the microphone, the amplifier, and the loudspeaker so that they reproduce the waveforms *accurately* at each step. It's a significant challenge, which is why accurate sound reproduction can be elusive.

Experiment 28: Making a Coil React

You've seen that when you pass current through a coil, the current creates a magnetic field. When you disconnect the current, what happens to the field that it created?

The energy in the field is converted back into a brief pulse of electricity. We say that this happens when the field *collapses*.

This experiment will enable you to see it for yourself.

What You Will Need

- Breadboard, hookup wire, wire cutters, wire strippers, multimeter
- Low-current LEDs (2)
- Hookup wire, 22 gauge (26 gauge preferred), 100 feet (1 spool)
- Resistor, 47 ohms (1)
- Capacitor, 1,000 µF or larger (1)
- Tactile switch (1)

Procedure

Take a look at the schematic in Figure 5-34. The breadboarded version is shown in Figure 5-35. For the coil, you can use a spool of 100 feet of 22-gauge hookup wire. Alternatively, if you created your own coil of 200 feet of wire in Experiment 26, you can use that; and if you splurged on a spool of magnet wire, that will be even better.

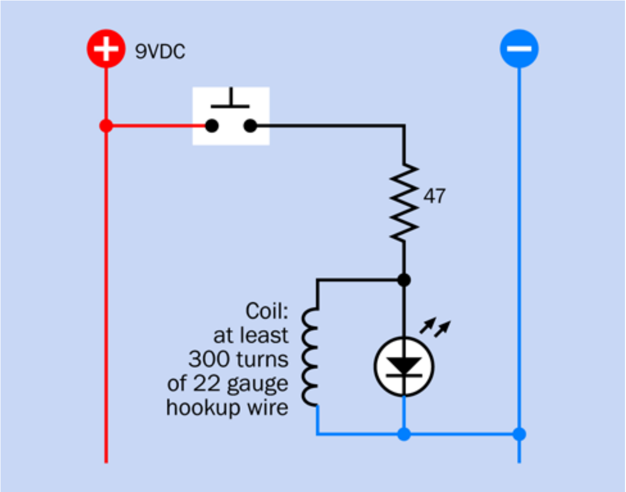

Figure 5-34 *A simple circuit to demonstrate the self-inductance of a coil.*

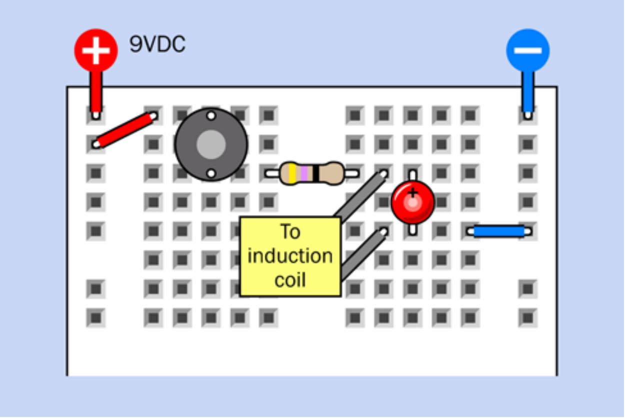

Figure 5-35 *Breadboarded version of the self-inductance experiment.*

When you look at the schematic, it doesn't seem to make much sense. The 47-ohm resistor seems too small to protect the LED—but why should the LED light up at all, when the electricity can go around it through the coil?

Now test the circuit, and I think you'll be surprised. Each time you press the button, the LED blinks briefly. Can you imagine why that should be?

Try adding a second LED, the other way up, as in Figure 5-36 and Figure 5-37. Press the button again, and the first LED flashes, as before. But now when you release the button, the second LED flashes.

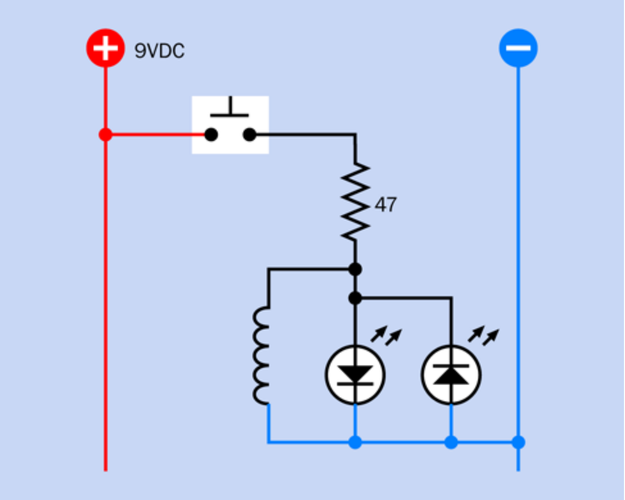

Figure 5-36 *One LED flashes when a magnetic field is created; the other flashes when the field collapses.*

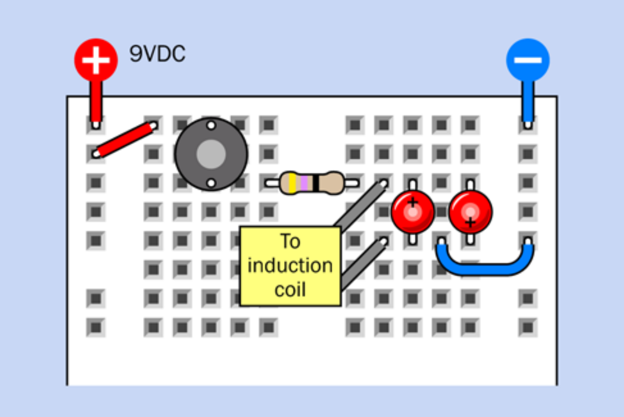

Figure 5-37 *Breadboarded version of the two-LED demonstration circuit.*

A Collapsing Field

Here's what happened during this experiment. At first, the coil required a brief amount of time to build up a magnetic field. This took a moment, and during that moment the coil blocked some of the flow of current. As a result, some of the current detoured through the first LED. Once the magnetic field was established, current flowed through the coil more normally.

This response of the coil is known as *self-inductance.* Sometimes people use the term *inductive reactance*, or just *reactance*, but since self-inductance is the correct term, I'll be using it here.

When you disconnected the power, the magnetic field collapsed, and the energy from the field was converted back into electricity in a short, brief pulse. This caused the second LED to flash when you let go of the button.

Naturally, different sizes of coil store and release different amounts of energy.

Perhaps you remember in Experiment 15, I advised you to tie a diode across the coil of a relay to absorb the surge that occurs when a relay coil is switched on and off. You have now seen this effect for yourself.

Resistors, Capacitors, and Coils

The three primary types of passive components in electronics are resistors, capacitors, and coils. We can now list and compare their properties.

A *resistor* constrains current flow, and drops voltage.

A *capacitor* allows a pulse of current to flow initially, but blocks direct current.

A *coil* (often referred to as an *inductor*) blocks DC current initially, but allows a continuing flow of direct current.

In the circuit that I just showed you, I didn't use a higher-value resistor because I knew the coil would allow only a very brief pulse. The blinking LEDs would have been less easily visible if I had used a more usual 330-ohm or 470-ohm resistor.

Don't try to run the circuit without the coil of wire included. You will quickly burn out one or both of the LEDs. The coil may look as if it isn't doing anything, but it is.

Here's one last variation on this experiment, to test your memory and understanding of electrical fundamentals. Build a new circuit shown in Figure 5-38 and Figure 5-39 using a 1,000µF capacitor instead of a coil (be careful to get its polarity the right way around, with the positive lead at the top.) Also, use a 470-ohm resistor, because the coil isn't there to block and divert current anymore.

First hold down button B for a second or two to make sure that the capacitor is discharged. Now, what will you see when you press button A? Maybe you can guess. Remember, a capacitor will pass an initial pulse of electricity. Consequently, the bottom LED lights up—and then gradually fades out, because the capacitor accumulates a positive charge on its upper plate and a negative charge on its lower plate. As this occurs, the potential across the lower LED diminishes to zero.

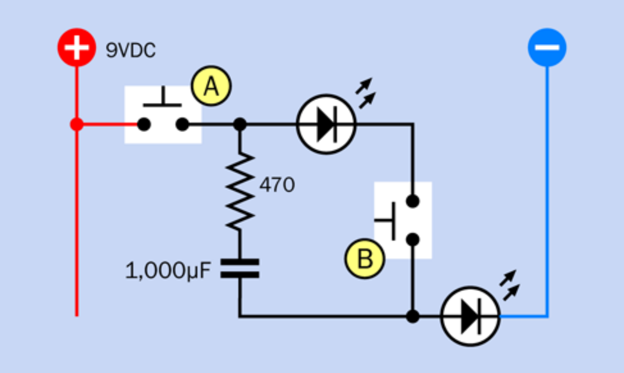

Figure 5-38 *In many ways, the behavior of a capacitor is opposite to the behavior of a coil.*

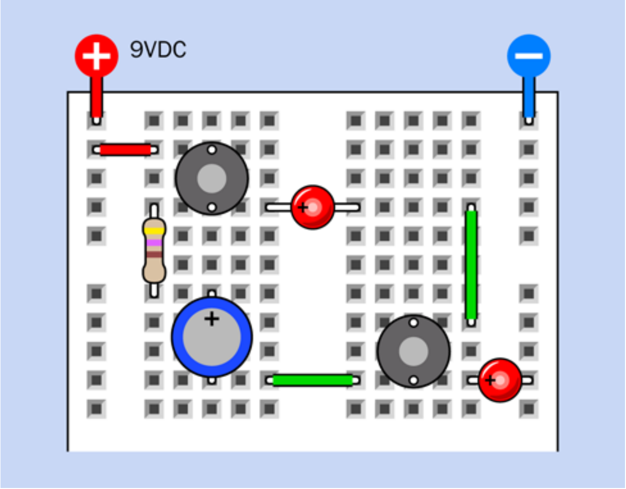

Figure 5-39 *Breadboarded version of the capacitor demo.*

The capacitor is now charged. Press the right-hand button, and the capacitor discharges through the upper LED. You can see this as the equivalent of the experiment in Figure 5-37, but using a capacitor instead of a coil.

Capacitors and inductors both store power. You were able to see this more obviously with the capacitor, because a high-value capacitor is much smaller than a high-value coil.

Theory: Alternating Current Concepts

Here's a simple thought experiment. Suppose you set up a 555 timer to send a stream of pulses through a coil. This will be a primitive form of alternating current.

Will the self-inductance of the coil interfere with the stream of pulses? That will depend on how long each pulse is, and how much inductance the coil has. If the frequency of pulses is just right, the self-inductance of the coil will last just long enough to block each pulse. Then the coil will recover in time to block the next one. In conjunction with a resistor (or just the resistance of the loudspeaker) a coil can suppress some frequencies while allowing others to pass through.

If you have a stereo system that uses a small speaker to reproduce high frequencies and a large speaker to reproduce low frequencies, almost certainly there is a coil somewhere in the speaker cabinet, stopping the higher frequencies from reaching the large speaker.

What happens if you substitute a capacitor for a coil? If the AC pulses are long relative to the value of the capacitor, it will tend to block them. But if the pulses are shorter, the capacitor can charge and discharge in rhythm with the pulses, and will allow them through.

I don't have space in this book to get deeply into alternating current. It's a vast and complicated field where electricity behaves in strange and wonderful ways, and the mathematics that describe it can become quite challenging, involving differential equations and imaginary numbers. (What is an imaginary number? The most obvious example is the square root of minus-one. How can that exist? Well—it can't, which is why we say that it's imaginary. Yet it crops up in electrical theory. If that sounds interesting, you may want to check it out.)

But I haven't finished with coils yet. The next experiment will demonstrate the audio effects that I just described above.

Experiment 29: Filtering Frequencies

In this experiment, you'll change the sound of sound. Using coils and capacitors, you can filter sections of the audible spectrum to create a rich variety of effects.

What You Will Need

- Breadboard, hookup wire, wire cutters, wire strippers, test leads, multimeter
- 9-volt power supply (battery or AC adapter)
- Loudspeaker, 8 ohms impedance, minimum 4 inches diameter (1)
- Audio amplifier chip, LM386 (1)
- 22-gauge hookup wire, 100 feet
- Small plastic storage bin as a loudspeaker enclosure (1)
- 555 timer (1)
- Resistors, 10K (2)
- Capacitors: 0.01µF (3), 2.2µF (1), 100µF (1), 220µF (3)
- Trimmer potentiometers: 10K (1), 1M (1)

- SPDT slide switches (4)
- Tactile switch (1)

A Home for Your Speaker

The little speaker that I recommended for previous projects was adequate when all you needed was a few beeps, but small speakers have a limited capability to reproduce bass notes. Since I want you to be able to hear how electronic components can affect those notes, it's time to consider a larger speaker, such as the one in Figure 5-40, which has a cone four inches in diameter.

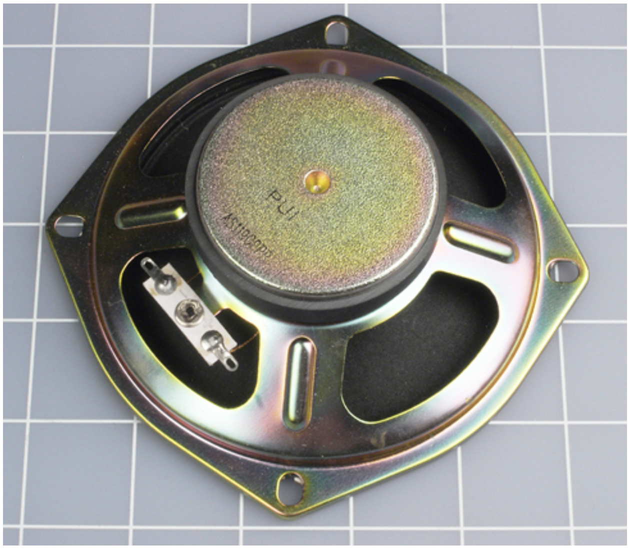

Figure 5-40 *A loudspeaker that would be suitable for this project.*

Bearing in mind my previous comments on the need to suppress out-of-phase sound waves from the back of the speaker, you'll need a box to contain it. The box will also boost the sound by resonating, in the same way that the body of an acoustic guitar resonates with the vibrations of its strings.

If you have time to make a plywood box, that would be ideal, but the simplest and cheapest enclosure is probably a plastic storage bin with a snap-on lid. Figure 5-41 shows the speaker bolted into the bottom of a bin. Drilling holes neatly in thin plastic is quite a challenge, and —well, I didn't try too hard.

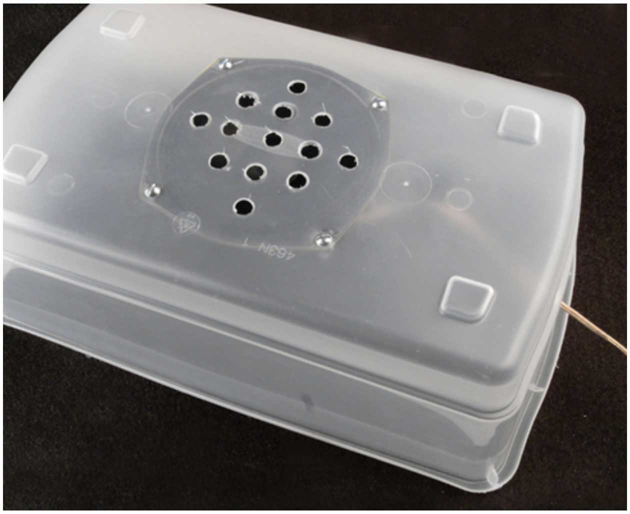

Figure 5-41 *A resonant enclosure is necessary if you want to hear some bass (lower frequencies) from your speaker. A cheap plastic storage bin is sufficient for demo purposes.*

To improve the attributes of the plastic bin, you can put some soft, heavy fabric inside it before you snap the lid on. A hand towel or some socks should be sufficient to absorb some of the vibration.

A Single Chip

Back in the 1950s, you needed vacuum tubes, transformers, and other power-hungry, heavyweight components to build an audio amplifier. Today, you can buy a chip for about $1 that will do the job, if you add a few capacitors around it, and a volume control.

One of the simplest, cheapest, and easiest to use is the LM386, which is available from multiple manufacturers, each of which prefaces or appends some extra identifying letters and numbers to it. The LM386N-1, LM386N, and LM386M-1 are all basically the same for our purposes. Just make sure you buy the through-hole version, not the surface-mount version. Pinouts for this amplifier are shown in Figure 5-42.

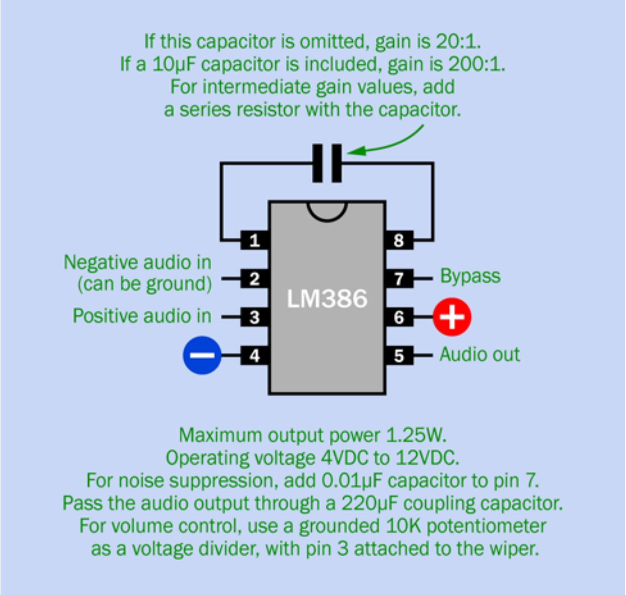
If this capacitor is omitted, gain is 20:1.
If a 10µF capacitor is included, gain is 200:1.
For intermediate gain values, add
a series resistor with the capacitor.
Negative audio in
(can be ground)
Positive audio in
1
2
3
4
8
7
6
5
LM386
Bypass
Audio out
Maximum output power 1.25W.
Operating voltage 4VDC to 12VDC.
For noise suppression, add 0.01µF capacitor to pin 7.
Pass the audio output through a 220µF coupling capacitor.
For volume control, use a grounded 10K potentiometer
as a voltage divider, with pin 3 attached to the wiper.

Figure 5-42 *Pinouts for the LM386 single-chip amplifier.*

This little chip works with a power supply ranging from 4VDC to 12VDC, and although it is rated for just 1.25 watts, you'll be surprised how loud it can sound. It has a nominal amplification ratio of 20:1.

Test, 1-2-3

For testing purposes, I want a source of frequencies covering a large range of the audible spectrum. A simple way to achieve this is with a 555 timer. The schematic in Figure 5-43 shows the timer at the top, with component values that can deliver a range from around 70Hz to 5KHz when you twiddle the 1M trimmer potentiometer. Unfortunately you won't hear this as a linear response, by which I mean that a small rotation of the trimmer will have a much greater audible effect on high frequencies than on low frequences. But, it will be good enough for demo purposes, and the lower frequencies provide a more dramatic demo of audio filtering anyway.

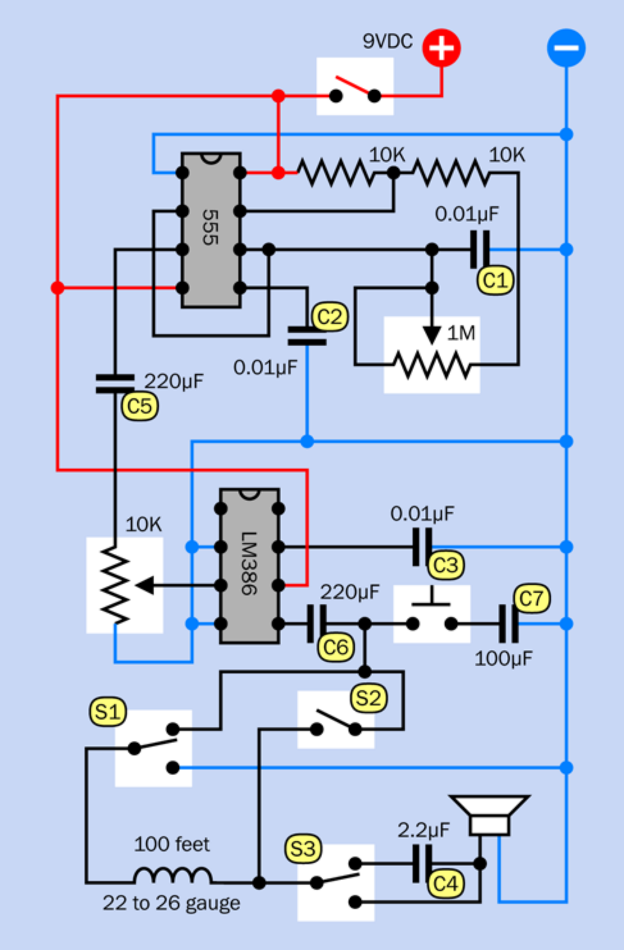
9VDC
10K
10K
555
0.01µF
C1
C2
1M
0.01µF
220µF
C5
10K
LM386
0.01µF
C3
220µF
C7
C6
100µF
S1
S2
100 feet
22 to 26 gauge
S3
2.2µF
C4

Figure 5-43 *A basic audio experimentation circuit.*

The breadboarded version of the circuit is shown in Figure 5-44 and the component values are shown in Figure 5-45.

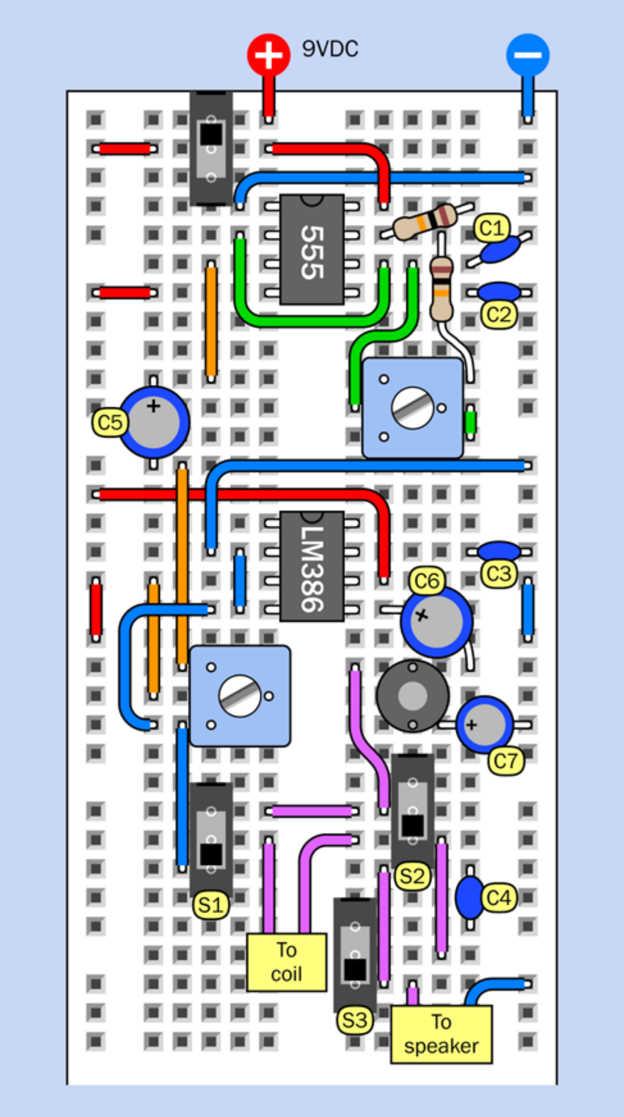
9VDC
555
C1
C2
C5
LM386
C6
C3
C7
S1
S2
C4
To coil
S3
To speaker

Figure 5-44 *Breadboarded version of the audio experimentation circuit.*

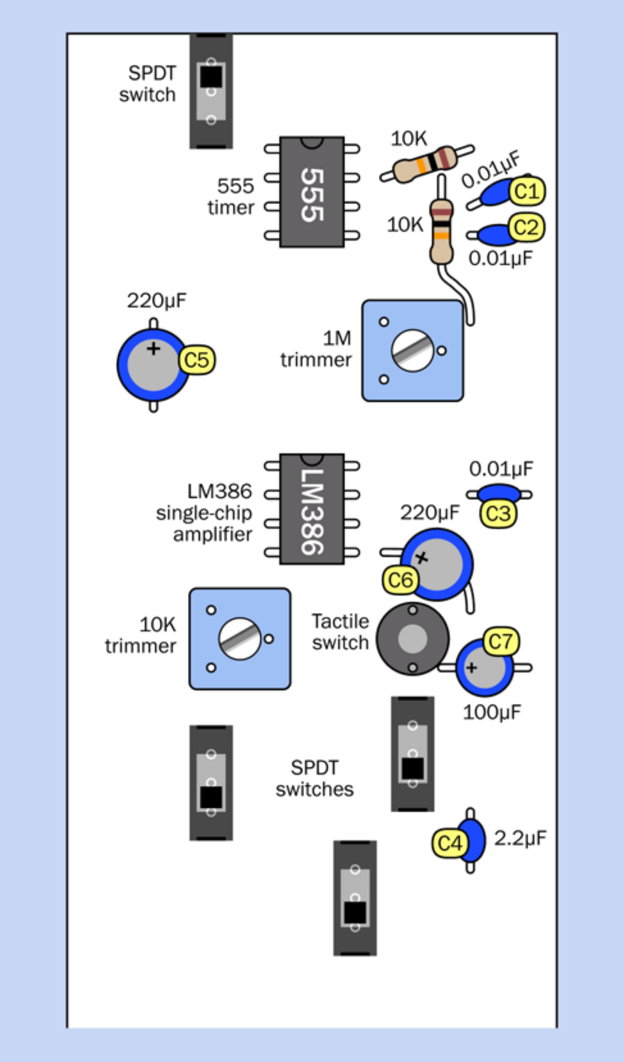
SPDT switch
555 timer
555
10K
0.01µF
C1
10K
C2
0.01µF
220µF
C5
1M trimmer
LM386 single-chip amplifier
LM386
0.01µF
220µF
C3
C6
10K trimmer
Tactile switch
C7
100µF
SPDT switches
C4
2.2µF

Figure 5-45 *Component values in the audio experimentation circuit.*

When you build this circuit, I have to warn you that amplifiers are sensitive to all electrical fluctuations, not just the ones you want to hear. Any electrical interference will be reproduced as a mess of scratchy, buzzing sounds, and this problem will be much worse if you use unnecessarily long pieces of wire to connect components.

The kind of jumper wires with little plugs on the end are especially undesirable in an amplifier circuit, as they behave like radio antennas. I have tried to limit the lengths of all the wires in the breadboard layout shown in Figure 5-44, and I encourage you to do the same. The only locations where wire length doesn't matter too much are on the power-output side of the chip, where you need to attach wires to your loudspeaker and to a coil.

For the coil, magnet wire of 22 gauge or thinner is ideal, but you will get some audible results from a 100-foot spool of 22-gauge hookup wire, and the 200 feet of hookup wire that I suggested in the previous experiment would be better.

Now, before powering up your breadboard, please pay attention to the three slide switches near the bottom of the circuit, and make sure that all of them are in the "down" position. In other words, slide them toward the bottom of the breadboard. Also turn the two trimmer potentiometers about halfway through their range.

You can power this project with an AC adapter or a 9-volt battery, with no regulation necessary. However, if you use an adapter, it may introduce some hum into the circuit. You can reduce this by placing a capacitor of 1,000µF or more between the two buses of the breadboard. If you use a battery, the power consumption of the amplifier will limit battery life to two or three hours, and some of the sound filters will pull down the voltage slightly, affecting the audio frequency created by the 555 timer.

As soon as you switch on the power, you should hear a tone. If you don't, your first troubleshooting strategy should be to disconnect the upper lead of the 220µF capacitor from the output pin of the 555 timer, and touch your speaker wires very briefly between that pin and the negative bus. If you don't hear anything, you made a wiring error around the timer. If you do hear something, then your error is related to the LM386 amplifier chip.

Make sure you connected power to the correct pins of the LM386. The positive and negative supply pins are not in the same positions as on logic chips.

Still no sound? Detach the top end of the short vertical piece of blue wire above the 10K trimmer. Touch the end of this piece of wire with your finger, and you should hear some whistling and buzzing sounds, because this is attached to the input pin of the amplifier (pin 4). *Still* nothing? Try attaching your speaker between the negative side of capacitor C6, and the negative power supply bus. C6 is a coupling capacitor connecting you directly with the output pin of the LM386.

If none of these attempts is successful, you'll have to go around the circuit with your meter, checking voltages.

Adventures in Audio

Assuming that your circuit is now up and running, I'll explain the functions of the components before I suggest some things to try. I'm going to refer to the labels that I applied to components in the breadboard layout shown in Figure 5-44.

Capacitor C1 sets the frequency of the timer, in conjunction with the 1M trimmer. Just in case you want to hear a sound higher in pitch than 5KHz, you can substitute a 0.0068µF (6.8nF) capacitor.

C5 is a coupling capacitor. It has a large value so that it will be transparent to a wide range of frequencies. Its purpose is to block DC from the 555 timer, because you only want to amplify the fluctuations, not the basic voltage.

Capacitor C6 is another coupling capacitor, protecting your loudspeaker from the DC coming out of the amplifier.

Capacitor C7 couples the amplifier output to negative ground when you press the button beside it. The value of C7 is chosen so that it takes away the higher frequencies, shunting them to ground. Without those frequencies, the sound that goes to the loudspeaker sounds more mellow.

Capacitor C4 is switched in and out of the circuit by slide switch S3. When the slide is up, sound from the 555 passes through C4 on its way to the amplifier. Because C4 has a small value, it blocks low frequencies, leaving you with a thin, tinny sound.

The complicated part of the circuit relates to the coil. I wanted you to hear the difference when a coil is connected in parallel with the loudspeaker, and when it is in series with the loudspeaker. Switches S1 and S2 give you those options, as shown in Figure 5-46 and Figure 5-47. When the coil is in parallel with the speaker, this is sometimes described as bypassing the speaker.

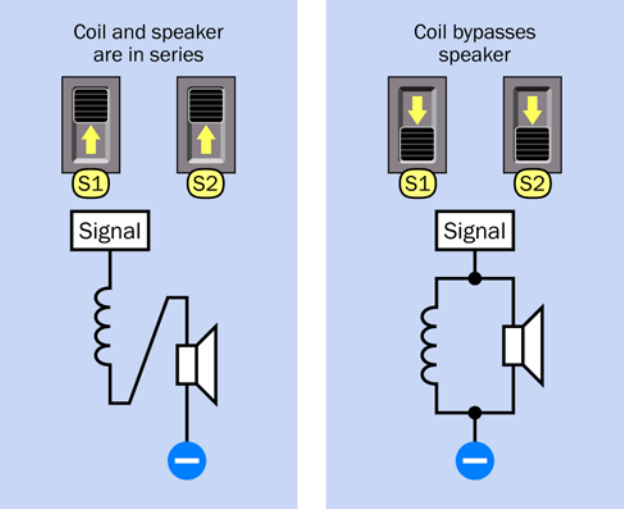

Figure 5-46 *Switches S1 and S2 (identified in the breadboard diagram for the circuit) allow you to feed audio to the speaker in parallel with an external coil, or in series with it.*

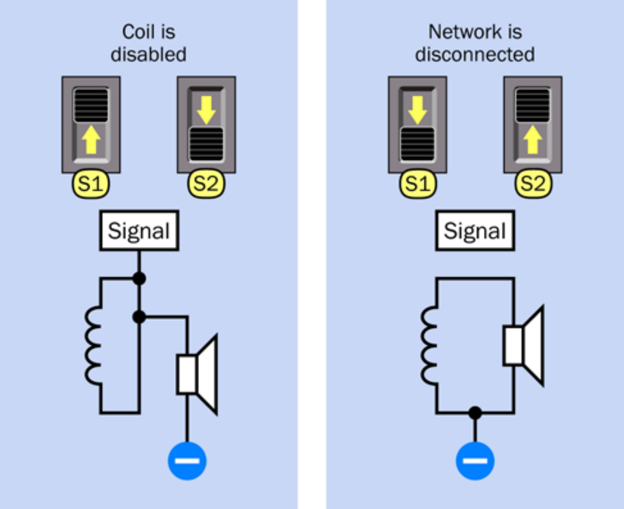

Figure 5-47 *Two other configurations of S1 and S2 allow you either to bypass the coil, or mute the output from the amplifier.*

You have quite a lot to play with here, especially bearing in mind that you can adjust the frequency and volume of the sound while you test the various filters. You can also test the effect of using two filters simultaneously. For instance, press the button to activate the bypass capacitor C7, which cuts the high frequencies, and switch C4 into the circuit at the same time, to cut the low frequencies. Now you have a *bandpass filter*, so called because you are passing just a narrow band of frequencies in the midrange.

The trimmer at bottom-left functions as a volume control, but you'll find that it only works properly in the middle part of its range. If you push it too high or too low, the circuit starts oscillating. This is a problem with amplifier circuits. The solution tends to involve adding small and large capacitors in various locations. I decided not to bother, because the midrange of the trimmer is usable.

The capacitors and coils in this circuit are all operating on a *passive* basis. They block some frequencies, but they don't boost any frequencies. A more sophisticated audio filtering system uses transistors to provide *active* filtering, but requires a lot more electronics.

Theory: Waveforms

If you blow across the top of a bottle, the mellow sound that you hear is caused by the air vibrating inside the bottle. If you could make a graph of the pressure waves, they would have a rounded profile.

If you could slow down time and draw a graph of the alternating voltage in any power outlet in your house, it would have the same profile.

If you could measure the speed of a pendulum swinging slowly to and fro in a vacuum, and draw a graph of the speed relative to time, once again it would have the same profile.

That profile is a *sine wave*, so called because you can derive it from basic trigonometry. In a right-angled triangle, suppose one of the sides adjacent to the right angle is called "a." If you divide the length of "a" by the length of the sloping side of the triangle (the hypoteneuse), the result is the *sine* of the angle opposite side "a."

To make this simpler, imagine a ball on a string rotating around a center point, as shown in Figure 5-48. Ignore the force of gravity, the resistance of air, and other annoying variables. Just measure the vertical height of the ball and divide it by the length of the string, at regular instants of time, as the ball moves around the circular path at a constant speed. Plot the result as a graph, and there's your sine wave, shown in Figure 5-49. Note that when the ball circles below its horizontal starting line, we consider its distance negative, so the sine wave becomes negative, too.

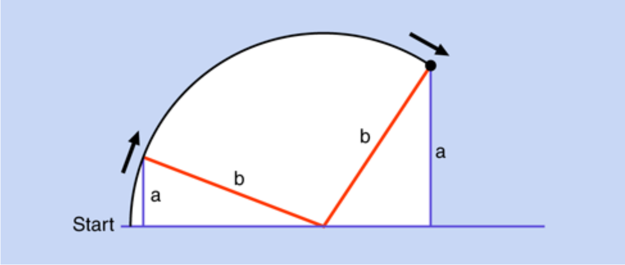

Figure 5-48 *You can draw a sine wave by beginning with simple geometry.*

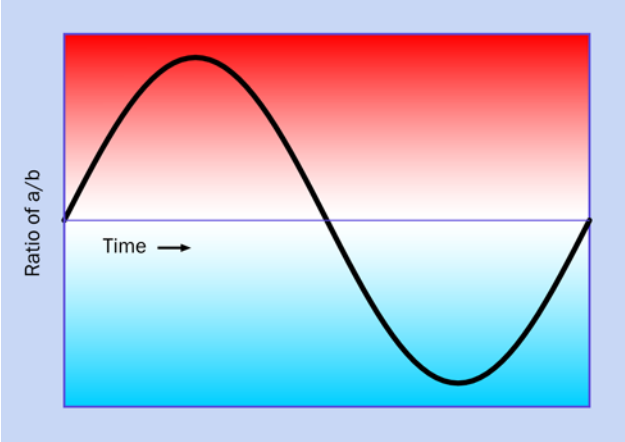

Figure 5-49 *An audio sine wave is generated by any instrument that uses a vibrating column of air—such as a flute. It is a gentle, harmonious sound.*

Why should this particular curve turn up in so many places and so many ways in nature? There are reasons for this rooted in physics, but I'll leave you to dig into that topic if it interests you. Getting back to the subject of audio reproduction, what matters is this:

- The static pressure in the air around you is called *ambient pressure*. It results from air being pulled down by the force of gravity. (Yes, air does have weight.)
- Almost any sound consists of a wave that is higher than ambient pressure, followed by a wave that is lower than ambient pressure—just like waves in the ocean.
- We can represent the higher and lower waves of pressure by voltages that are relatively high and relatively low, which is why I used the red and blue background in Figure 5-49.
- Any sound can be broken down into a mixture of sine waves of varying frequency and amplitude.

Or, conversely:

- If you put together the right mix of audio sine waves, you can create *any sound at all*.

Suppose that there are two sounds playing simultaneously. Figure 5-50 shows one sound as a purple curve and the other as a green curve. When the two sounds travel either as pressure waves through air or as alternating electric currents through a wire, their amplitudes are added together to make the more complex curve, which is shown in black. Now try to imagine dozens or even hundreds of different frequencies being added together, and you have an idea of the complex waveform of a piece of music.

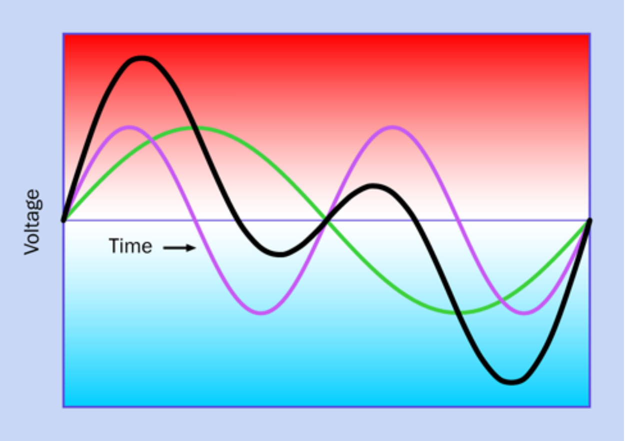

Figure 5-50 *When two sine waves are generated at the same time (for instance, by two musicians, each playing a flute), the combined sound creates a compound curve. The purple sine wave is twice the frequency of the green sine wave. The compound curve (black line) is the sum of the distances of the sine waves from the baseline of the graph.*

An astable 555 timer circuit creates a *square wave*. This is because the output from the timer switches abruptly from low to high and back again. The result is shown in Figure 5-51. A sine wave sounds gentle and melodious, as it varies smoothly. A square wave tends to sound harsh, and has a "buzz" to it. That buzz is really composed of *harmonics*, meaning frequencies that are two or more times the basic frequency.

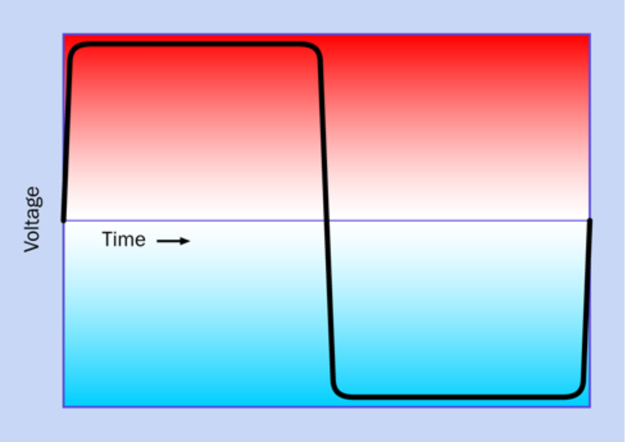

Figure 5-51 *The square wave that you might obtain from a source such as a 555 timer, which switches its output abruptly on and off.*

Because a square wave contains high-frequency harmonics, it's a good choice for testing audio filters. A low-pass filter, which only allows lower frequencies, will remove the buzz by rounding off the corners of the square wave.

Mangling Some Music

You may be wondering, if the LM386 is an audio amplifier, can it amplify music? Yes, and in fact, that's what it is designed to do. You can test this for yourself using any audio device with a headphone output.

Bear in mind that the LM386 is only a mono amplifier, so you won't be able to hear both audio channels from your music player. To obtain just one of them, use a cable with a miniature audio jack at each end. Cut off one of them, strip the insulation from the wire, and you will find a mesh of fine wires that are the shielding in the cable, to be connected with negative ground. Inside the shielding will be two conductors, carrying the signals for the left channel and the right channel. Snip off one of them (it doesn't matter which one), and throw it away—but don't allow the conductor in the residual stub of wire to short-circuit with the shielding.

Strip insulation from the remaining conductor. The wires inside are very thin, and you will be able to deal with them more easily in this experiment if you add a little solder. The desired result is shown in Figure 5-52.

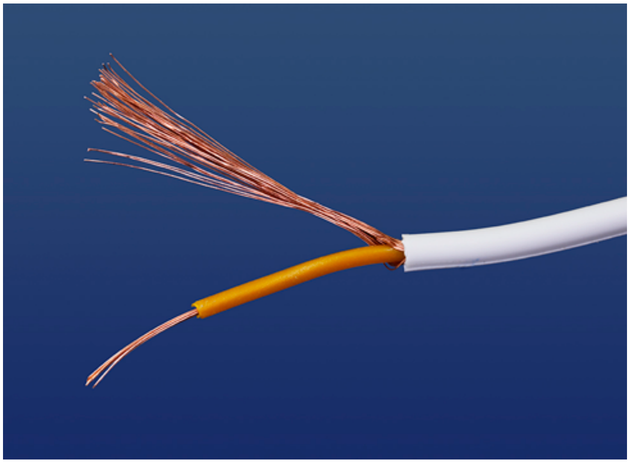

Figure 5-52 *An audio cable stripped to expose its shielding and one conductor. The shielding connects with negative ground.*

Make sure the power to your amplifier circuit is off, and push all the slide switches to the down positions. Remove the orange piece of wire that connects pin 3 of the 555 timer with the 220µF capacitor below it. You are taking the 555 timer out of the circuit, and using the positive end of capacitor C6 as your input point.

Use one of your alligator test leads to grab the positive capacitor lead, and attach the other end of the test lead to the audio conductor in your cable. Use another test lead to connect the shielding from the cable to negative ground in your circuit. It's essential that your music player must share negative ground with your amplifier circuit.

Switch on your circuit, then switch on your music player, and you should hear music. If it's too loud and distorted, you may need to insert a 1K or 10K resistor between the audio wire from the music player and the positive end of the capacitor.

Once you have the volume right, you can play with your high-pass and low-pass filters to see how they affect the music. They won't make it sound good, but they will make it sound different.

Background: Crossover Networks

In a traditional audio system, each speaker cabinet contains two drivers—one of them a small speaker called a *tweeter*, which reproduces high frequencies, the other a large speaker known as a *woofer*, which reproduces low frequencies. (Modern systems often remove the woofer and place it in a separate box of its own that can be positioned almost anywhere, because the human ear has difficulty sensing the direction of low-frequency sounds. In this system, the woofer may be referred to as a *subwoofer*, because it is capable of reproducing very low frequencies.)

Audio frequencies are divided between a tweeter and a woofer by filtering them, so that the tweeter doesn't try to deal with any low frequencies, and the woofer is protected from high frequencies. The circuit that takes care of this is called a "crossover network," and truly hardcore audiophiles have been known to make their own (especially for use in car systems) to go with speakers of their choice in cabinets that they design and build themselves.

If you want to make a crossover network, you should use high-quality polyester capacitors (which have no polarity, last longer than electrolytics, and are better made) and a coil that has the right number of turns of wire and is the right size, to cut high frequencies at the appropriate point. Figure 5-53 shows a polyester capacitor, while Figure 5-54 shows an audio crossover coil that I bought on eBay for $6. I was curious to find out what was inside it, so I took it apart.

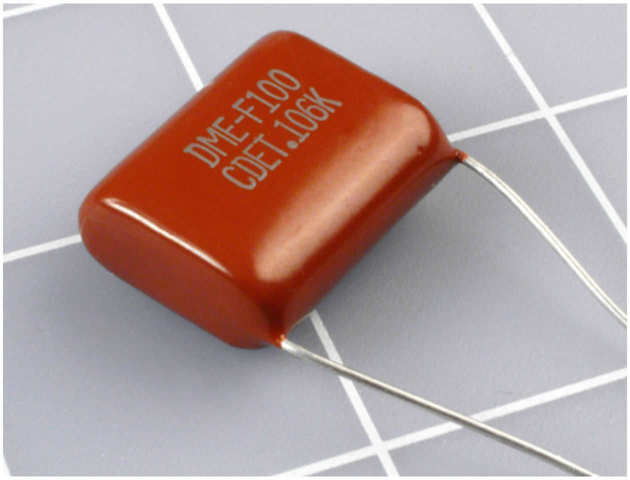

Figure 5-53 *Some nonelectrolytic capacitors have no polarity, such as this high-quality polyester film capacitor.*

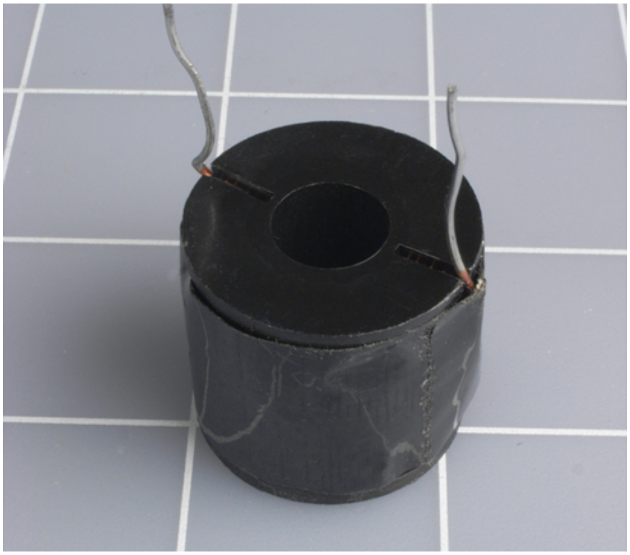

Figure 5-54 *What exotic components may we find inside this high-end audio component?*

First I peeled away the black vinyl tape that enclosed the coil. Inside was some typical magnet wire—copper wire thinly coated with shellac or semitransparent plastic, as shown in Figure 5-55. I unwound the wire, and as I did so, I counted the number of turns.

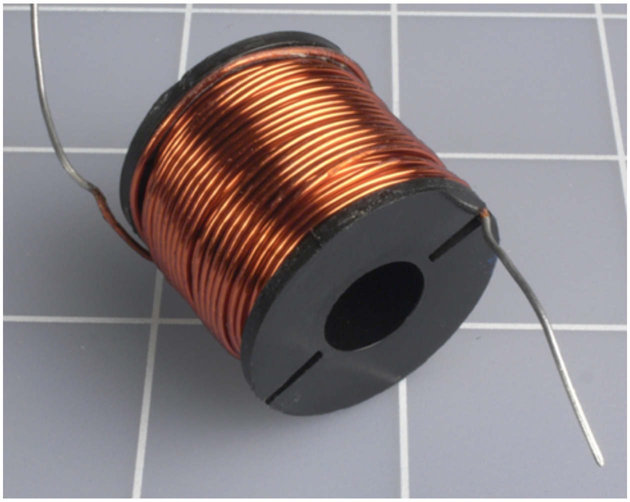

Figure 5-55 *The black tape is removed, revealing a coil of magnet wire.*

Figure 5-56 shows the wire and the spool that it was wound around.

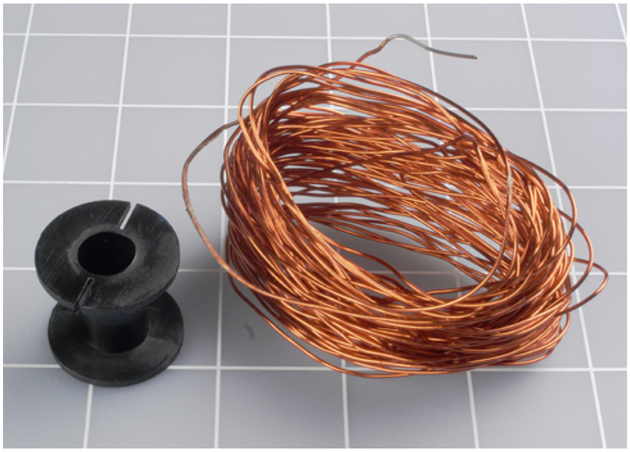

Figure 5-56 *The audio crossover coil consists of a plastic spool and some wire. Nothing more.*

So here's the specification for this particular coil in an audio crossover network. Forty feet of 20-gauge copper magnet wire, wrapped in 200 turns around a small plastic spool.

Conclusion: there's a lot of mystique attached to audio components. They are frequently overpriced, and you can make your own coil if you start with these parameters and adjust them to suit yourself.

Suppose you want to put some thumping bass speakers into your car. Could you build your own filter so that they only reproduce the low frequencies? Absolutely—you just need to wind a coil, adding more turns until it cuts as much of the high frequencies as you choose. Just make sure the wire is heavy enough so that it won't overheat when you push 100 or more audio watts through it.

Here's another project to think about: a color organ. You can tap into the output from your stereo and use filters to divide audio frequencies into three sections, each of which drives a separate set of colored LEDs. The red LEDs will flash in response to bass tones, yellow LEDs in response to the midrange, and green LEDs in response to high frequencies (or whatever colors you prefer). You can put signal diodes in series with the LEDs to rectify the alternating current, and series resistors to limit the voltage across the LEDs to, say, 2.5 volts (when the music volume is turned all the way up). You'll use your meter to check the current passing through each resistor, and multiply that number by the voltage drop across the resistor, to find the wattage that it's handling, to make sure the resistor is capable of dissipating that much power without burning out.

Audio is a field offering all kinds of possibilities if you enjoy designing and building your own electronics.

Experiment 30: Make It Fuzzy

Let's try one more variation on the circuit in Experiment 29. This will demonstrate another fundamental audio attribute: distortion.

What You Will Need

- The breadboarded circuit from Experiment 29, plus:
- 2N2222 transistor (1)
- Resistors: 330 ohms (1), 10K (1)
- Capacitors: 1µF (2) and 10µF (1)

Making the Mods

The modifications to the circuit are very minor. You need to add a transistor, two resistors, and three capacitors. Figure 5-57 shows the new components at the top end of the breadboard, with the preexisting components grayed out.

Figure 5-58 shows the same components, and their values, in the relevant section of the schematic, with other components omitted.

The 2N2222 transistor overloads the input of the LM386, while the 1µF capacitors, C8 and C9, limit low frequencies to emphasize the fuzzy effect.

The purpose of C10 is to boost the LM386. This is a feature of the chip: if you add a capacitor between pin 1 and pin 8, the power of the amplifier can increase from 20:1 to 200:1.

Thus, in two separate ways this unfortunate little amplifier chip is being forced to do more than its designers intended. Naturally enough, it will complain about this cruel treatment.

Make the modifications, and switch on the power. The output previously had some buzz in it, because it was basically a square wave. But now if you experiment with the 10K trimmer and the 1M trimmer, you can make the output scream, Hendrix-style.

If the result is too extreme, you can remove the 330-ohm resistor and substitute one with a slightly higher value. And what exactly is going on here?

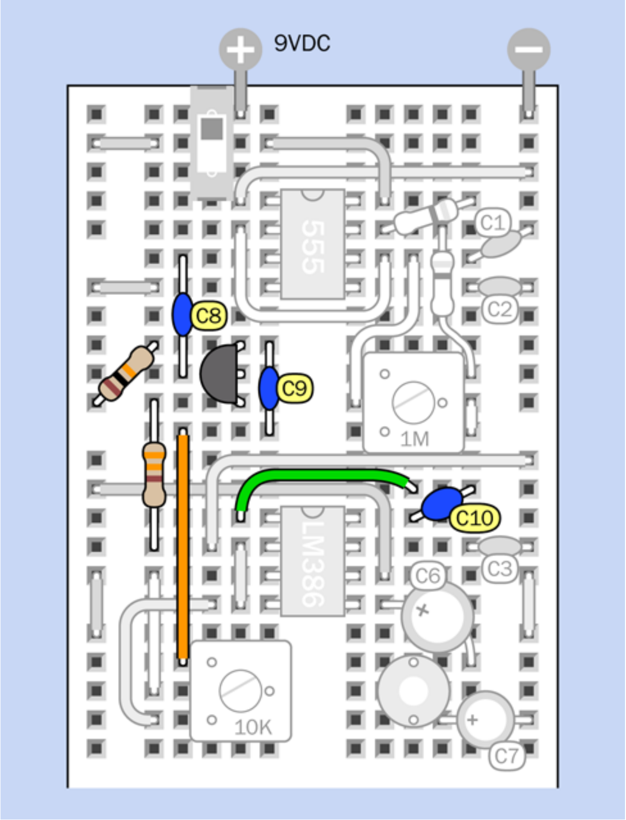

Figure 5-57 *Modifications to the circuit from Experiment 29, to add more distortion.*

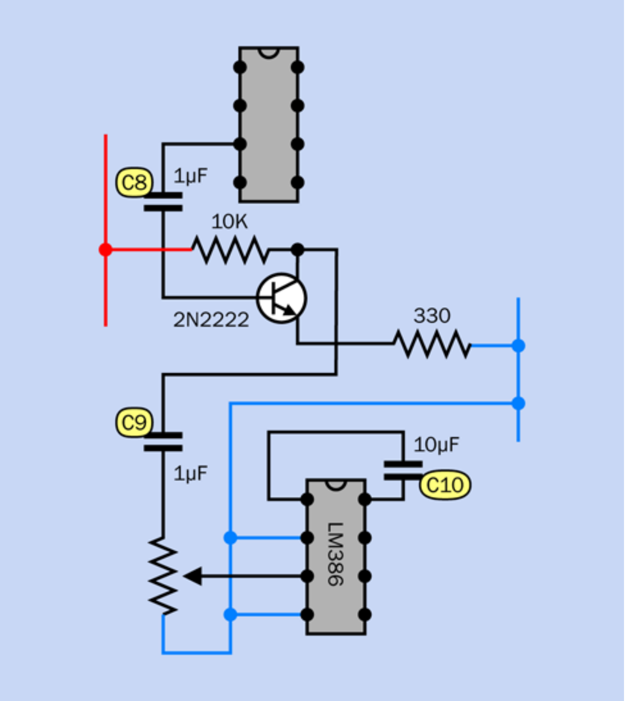

Figure 5-58 *The additional components, with values.*

Background: Clipping

In the early days of "hi-fi" (high-fidelity) sound, engineers labored mightily to perfect the process of audio reproduction. They wanted the waveform at the output end of the amplifier to look identical with the waveform at the input end, the only difference being that it should be bigger, so that it would be powerful enough to drive loudspeakers. The slightest distortion of the waveform was unacceptable.

Little did they realize that their beautifully designed tube amplifiers would be abused by a new generation of rock guitarists whose intention was to create as much distortion as possible.

If you push a vacuum tube—or a transistor—to amplify a sine wave beyond the component's capabilities, it runs out of power and "clips" the top and bottom of the curve. This makes it look more like a square wave, and as I explained Experiment 29, a square wave has a harsh, buzzing quality. For rock guitarists trying to add an edge to their music, the harshness is a desirable feature.

The sequence in Figure 5-59 shows what happens. So long as the output stays within the voltage limits of the amplifier, the signal can be faithfully reproduced. But in the second frame of the sequence, the input to the amplifier has increased to the point where the output would exceed the limits (suggested by the gray sections of the curve). Because the amplifier only has so much power available, it clips the signal, as shown in the third frame in the sequence.

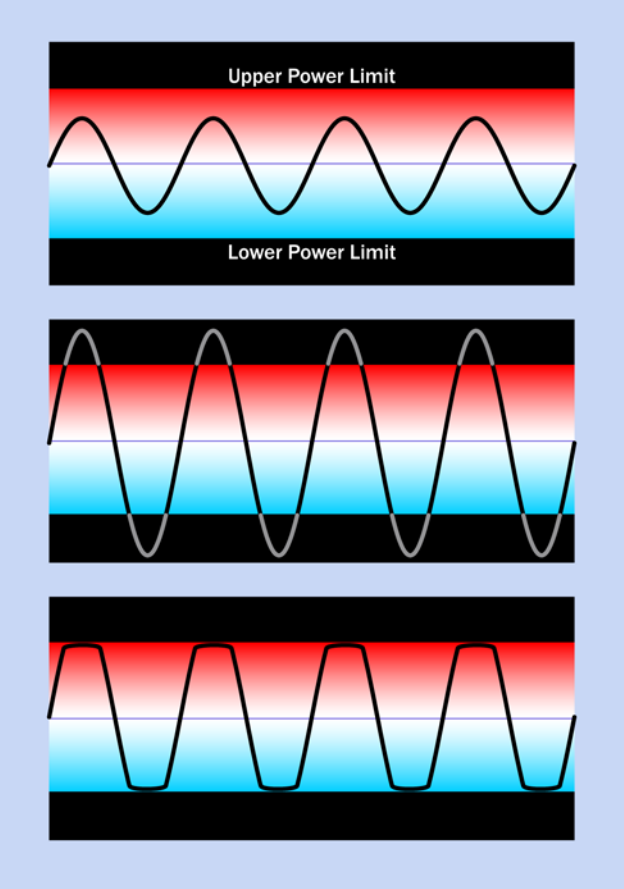

Figure 5-59 *When a sine wave (top) is passed through an amplifier that is turned up beyond the limit of its components, the amplifier chops the wave (bottom).*

For rock guitarists, clipping sounded good, and "stomp boxes" were introduced to create the effect. A very early example is shown in Figure 5-60

Figure 5-60 *This Vox Wow-Fuzz pedal was one of the early stomp boxes, which deliberately induced the kind of distortion that audio engineers had been trying to get rid of for decades.*

Background: Stomp-Box Origins

The Ventures recorded the first single to use a fuzz box, titled "The 2,000 Pound Bee," in 1962. Truly one of the most awful instrumentals ever made, it used distortion as a gimmick and must have made other musicians conclude that this was a sound destined for oblivion.

Then Ray Davies of the Kinks started experimenting with distortion, initially by plugging the output from one amp into the input of another, supposedly during the recording of his hit "You Really Got Me." This overloaded the input and created clipping that sounded more musically acceptable. From there it was a short step to Keith Richards using a Gibson Maestro Fuzz-Tone when the Rolling Stones recorded "(I Can't Get No) Satisfaction" in 1965.

Today, you can find thousands of advocates promoting as many different mythologies about "ideal" distortion. In Figure 5-61, I've included a schematic from Flavio Dellepiane, a circuit designer in Italy who gives away his work (with a little help from Google AdSense).

Flavio is a self-taught Maker, having gained much of his knowledge from electronics magazines such as the old British publication, *Wireless World*. In the fuzz circuit that I'm including here, he uses a very high-gain amplifier consisting of three field-effect transistors (FETs), which closely imitate the rounded square wave typical of an overdriven tube amp.

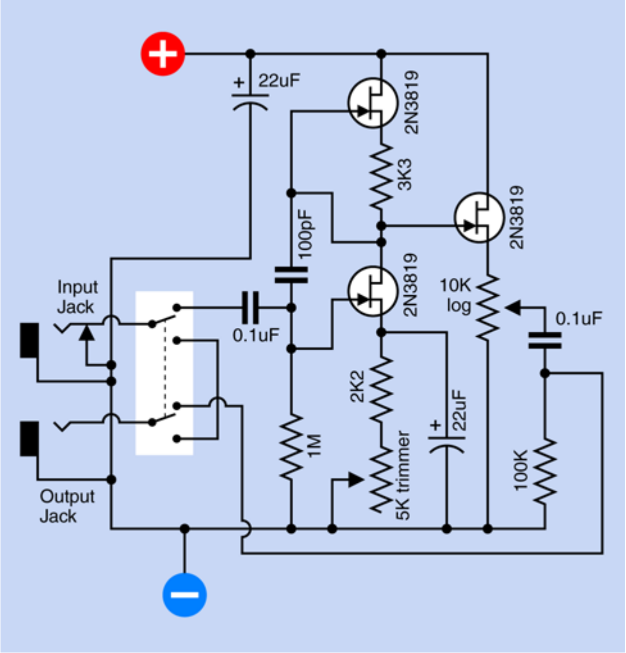

Figure 5-61 *This circuit designed by Flavio Dellepiane uses three transistors to simulate the kind of distortion that used to be created by overloading the input of a tube amplifier.*

Flavio offers dozens more schematics on his site, developed and tested with a dual-trace oscilloscope, low-distortion sinewave oscillator (so that he can give audio devices a "clean" input before he abuses it), distortion meter, and precision audio voltmeter. This last item, and the oscillator, were built from his own designs, and he gives away their schematics, too. Thus his site provides one-stop shopping for home-audio electronics hobbyists in search of a self-administered education.

Before fuzz, there was tremolo. A lot of people confuse this with vibrato, so let's clarify that distinction right now:

- Vibrato applied to a note makes the frequency waver up and down, as if a guitarist is bending the strings.
- Tremolo applied to a note makes its volume fluctuate, as if someone is turning the volume control of an electric guitar up and down very quickly.

Harry DeArmond sold the first tremolo box, which he named the Trem-Trol. It looked like an antique portable radio, with two dials on the front and a carrying handle on top. Perhaps in an effort to cut costs, DeArmond didn't use any electronic components. His steam-punkish Trem-Trol contained a motor fitted with a tapered shaft, with a rubber wheel pressing against it. The speed of the wheel varied when you turned a knob to reposition the wheel up and down the shaft. The wheel, in turn, cranked a little capsule of "hydro-fluid," in which two wires were immersed, carrying the audio signal. As the capsule rocked to and fro, the fluid sloshed from side to side, and the resistance between the electrodes fluctuated. This modulated the audio output.

Today, Trem-Trols are an antique collectible. Johann Burkard has posted an MP3 of his DeArmond Trem-Trol online so you can actually hear it.

The idea of using a mechanical source for electronic sound mods didn't end there. The original Hammond organs derived their unique, rich sound from a set of toothed wheels turned by a motor. Each wheel created a fluctuating inductance in a sensor like the record head from a cassette player.

It's fun to think of other possibilities for motor-driven stomp boxes. Going back to tremolo: imagine a transparent disc masked with black paint, except for a circular stripe that tapers at each end. While the disc rotates, if you shine a bright LED through the transparent stripe toward a phototransistor, you have the basis for a tremolo device. You could even create never-before-heard tremolo effects by keeping a library of discs with different stripe patterns. Figure 5-62 shows what I have in mind, while Figure 5-63 suggests some disc patterns. For a real fabrication challenge, how about an automatic disc changer?

In the world of solid-state electronics, today's guitarists can choose from a smorgasbord of effects, all of which can be home-built using plans available online. For reference, try these special-interest books:

Analog Man's Guide to Vintage Effects by Tom Hughes (For Musicians Only Publishing, 2004). This is a guide to every vintage stomp box and pedal you can imagine.

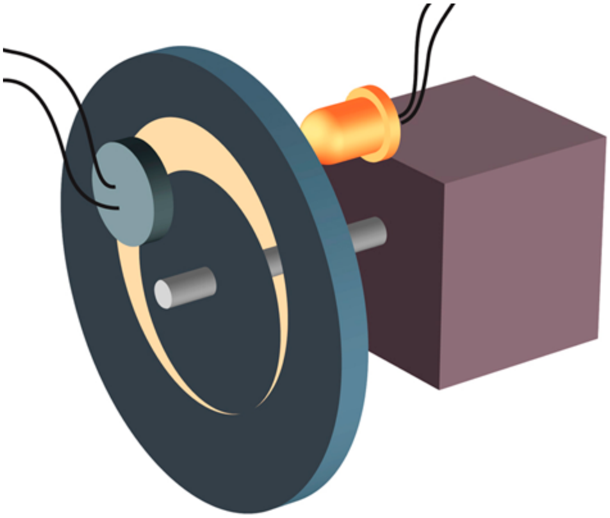

Figure 5-62 *A hypothetical neo-electromechanical tremolo generator.*

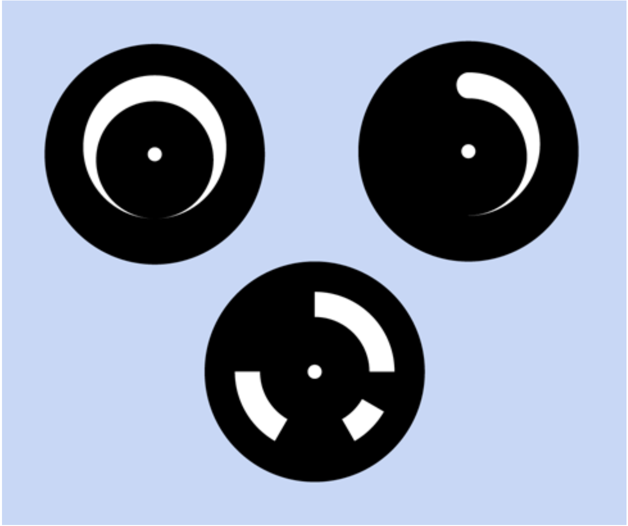

Figure 5-63 *Different stripe patterns could be used to create various tremolo effects.*

How to Modify Effect Pedals for Guitar and Bass by Brian Wampler (Custom Books Publishing, 2007). This is an extremely detailed guide for beginners with little or no prior knowledge. Currently it is available only by download, from sites such as Open Library.

but you may be able to find the previously printed edition from secondhand sellers, if you search for the title and the author.

Of course, you can always take a shortcut by laying down a couple hundred dollars for off-the-shelf stomp boxes that use digital processing to emulate distortion, metal, fuzz, chorus, phaser, flanger, tremolo, delay, reverb, and several more, all in one convenient package. Purists, of course, will claim that it "doesn't quite sound the same," but maybe that's not the point. Some of us simply can't get no satisfaction until we build our own stomp box and then tweak it, in search of a sound that doesn't come off-the-shelf and is wholly our own.

Experiment 31: One Radio, No Solder, No Power

Getting back to the principle of inductance, I want to show you how it can enable a simple circuit that receives AM radio signals without a power supply. This is often known as a crystal radio, because the earliest examples used a natural mineral crystal that functioned as a semiconductor. The idea originated at the dawn of telecommunications, but if you've never tried it, you've missed an experience that is truly magical.

What You Will Need

- Rigid cylindrical object, about three inches in diameter, such as a vitamin bottle or water bottle (1)
- Hookup wire, 22 gauge, 60 feet minimum
- Heavier wire, 16 gauge preferred, 50 to 100 feet (this wire may be stranded, and you can try a thinner gauge to reduce the cost, although your radio may not pull in so many stations)
- Polypropylene rope ("poly rope") or nylon rope, 10 feet
- Germanium diode (1)
- High-impedance earphone (1)
- Test lead (1)
- Alligator clips (3) or use extra test leads

Optional:

- 9-volt power supply (battery or AC adapter)
- LM386 single-chip amplifier

- Small loudspeaker (2" acceptable)

The diode and headphone can be ordered from the Scitoys Catalog. A high-impedance earphone is also available from amazon.com.

Step 1: The Coil

You need to create a coil that will resonate with radio transmissions in the AM waveband. The coil will consist of 65 turns of 22-gauge hookup wire, measuring approximately 60 feet.

You can wind the coil around any empty glass or plastic container, so long as it has parallel sides providing a constant diameter close to 3". A water bottle will do, if it isn't the type made of extremely thin plastic that will be easily squashed or deformed under pressure.

I just happened to have a vitamin bottle that was exactly the right size. In the photographs, you'll notice that it has no label. I softened its adhesive with a heat gun (lightly, to avoid melting the bottle) and then peeled it off. Some remaining adhesive residue was removed with a little xylene.

After you prepare a clean, rigid bottle, use a sharp object such as an awl or a nail to punch two pairs of holes in it, as shown in Figure 5-64. The holes will be used to anchor the ends of the coil.

Strip some insulation from the end of your hookup wire, and anchor it in one pair of holes, as shown in Figure 5-65. Now wrap five turns of wire around the bottle, and keep it from unwinding itself by applying a small, temporary piece of tape. Duct tape is ideal, or regular Scotch tape will do. "Magic" tape isn't strong enough and will be difficult to remove.

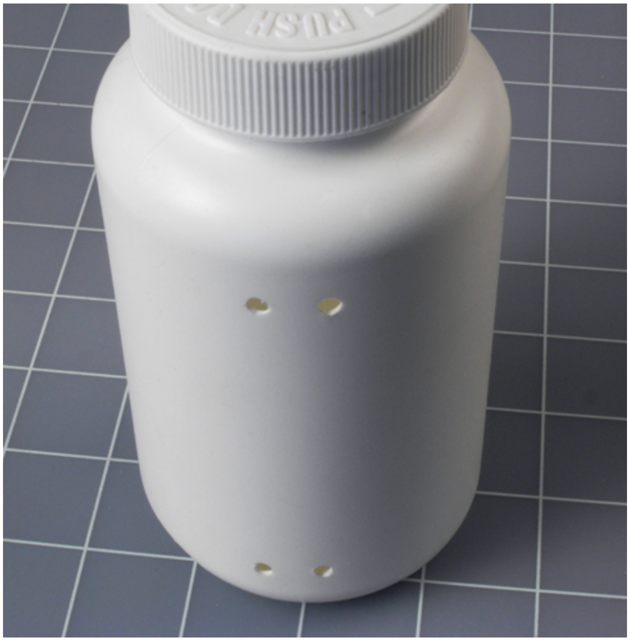

Figure 5-64 *The holes will anchor wire wrapped around the bottle.*

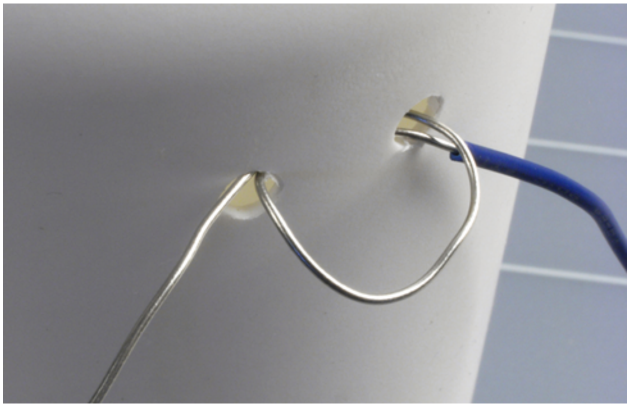

Figure 5-65 *Anchor one end of your wire in a pair of holes.*

Now you need to strip away about half an inch of insulation from the wire. The idea is that you should be able to tap into the coil at this point. Using your wire strippers, make an incision in the insulation and then pull the plastic coating away from your incision. See Figure 5-66.

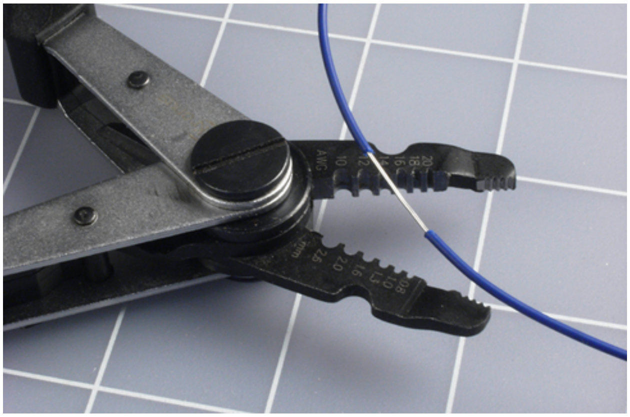

Figure 5-66 *Use your wire strippers and your thumbnails to pull back about half an inch of insulation.*

The next step is to twist the exposed wire into a loop, to make it easily accessible and prevent the insulation from closing up. See Figure 5-67.

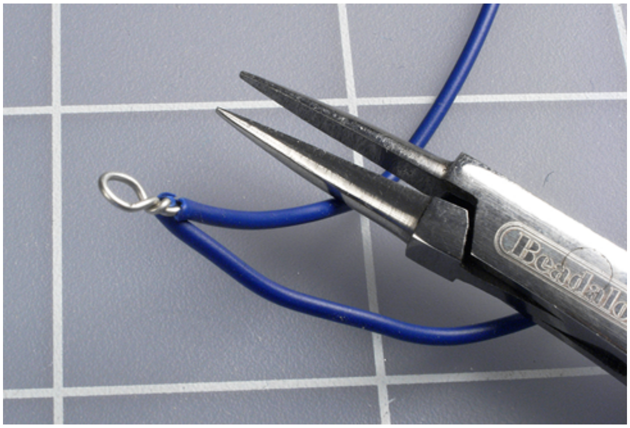

Figure 5-67 *Create a loop in the section of wire that you exposed.*

You just created a *tap* on your coil. Remove the piece of tape that you used to hold your first five turns temporarily, and wind another five turns around the bottle. Apply the tape again, and create another tap. You'll need a total of 12 of them, altogether. It doesn't matter if they don't line up with each other precisely. When you have made the last tap, wind five more turns around the bottle and then cut the wire. Bend the end into a U shape about a half-inch in diameter, so that you can hook it through the pair of holes that you drilled at the far end of the bottle. Pull the wire through, then loop it around again to make a secure anchor point.

My coil wrapped around a vitamin bottle is shown in Figure 5-68.

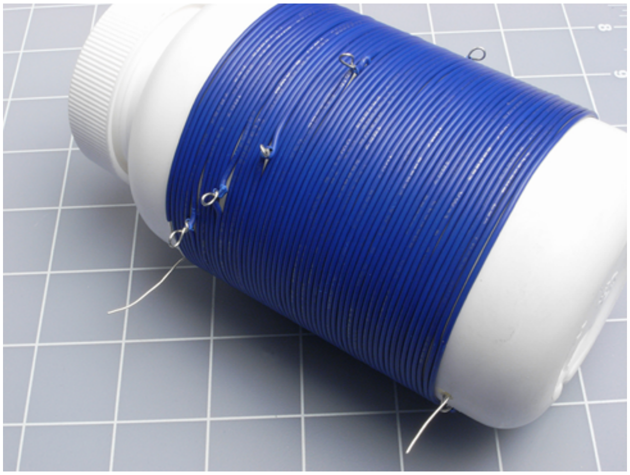

Figure 5-68 *The completed coil, wrapped tightly around the bottle.*

Your next step is to set up an antenna, which will be a section of wire that is as thick as possible and as long as possible. If you live in a house with a yard outside, this is easy: just open a window, toss out a reel of 16-gauge wire while holding the free end, then go outside and string up your antenna by using polypropylene rope ("poly rope") or nylon rope, available from any hardware store, to hang the wire from any available trees, gutters, or poles. The total length of the wire should be 50 to 100 feet. Where it comes in through the window, suspend it on another length of poly rope. The idea is to keep your antenna wire as far away from the ground or from any grounded objects as possible.

If you don't have an accessible yard, you can string up your antenna indoors, hanging it with poly rope or nylon rope from window treatments, door knobs, or anything else that will keep it off the floor. The antenna doesn't have to be in a straight line; in fact you can run it all around the room.

Caution: High Voltage!

The world around us is full of electricity. Normally we're unaware of it, but a thunderstorm is a sudden reminder that there's a huge electrical potential between the ground below and the clouds above.

If you put up an outdoor antenna, never use it if there is any chance of a lightning strike. This can be extremely dangerous. Disconnect the indoor end of your antenna, drag it outside, and push the end of the wire into the ground to make it safe.

Antenna and Ground

Use an alligator test lead to connect the end of your antenna wire with the top end of the coil that you made.

Next you need to establish a ground wire. This literally has to connect with the ground outside. Ideally, you should bury a couple of feet of bare wire in soft, moist earth—although this may be problematic if you live in a desert area, as I do. If you use a grounding stake of the type sold by wholesale electrical supply houses to ground welding equipment, be careful where you hammer it into the ground. You don't want to hit any hidden conduits.

A cold-water pipe is often suggested as a good connection with the ground, but (duh!) this will work only if the pipe is made of metal. Even if your home is plumbed with copper pipes, a section may have been repaired and replaced with plastic at some time in the past.

Probably the most reliable option is to attach the wire to the screw in the cover plate of an electrical outlet, as the electrical system in your house will be ultimately grounded. But be sure to anchor the wire securely, so there is absolutely no risk of it touching the sockets in the outlet. I would prefer not to insert the ground wire in the ground socket of the outlet, because there is the risk of poking it into the live socket by mistake.

Now you need a couple of slightly hard-to-find items: a germanium diode, which functions like a silicon-based diode but is better suited to the tiny voltages and currents that you'll be dealing with, and a high-impedance earphone. The kind of earphones or ear buds that you use with a media player will not work here; this has to be an old-school item, like the one shown in Figure 5-69. If it has a plug on the end, you'll have to snip it off and then carefully strip insulation from the tip of each wire.

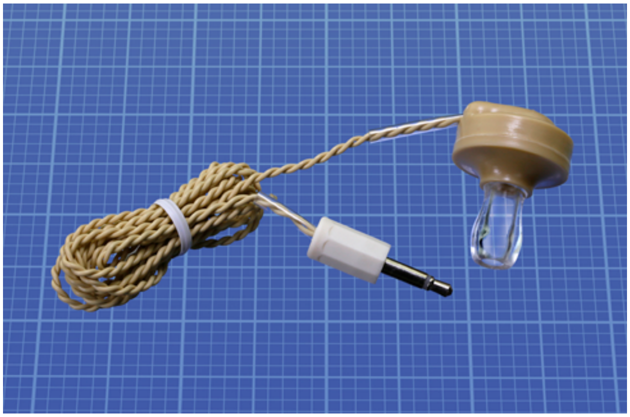

Figure 5-69 *This is the type of earphone you need for your no-power radio.*

The parts are assembled with test leads and alligator clips, as shown in Figure 5-70. The real-world version that I built isn't as neat as the diagram, but the connections are still the same, as shown in Figure 5-71. Notice that the test lead at the bottom can latch on to any of the taps on your coil. This is how you will be tuning your radio.

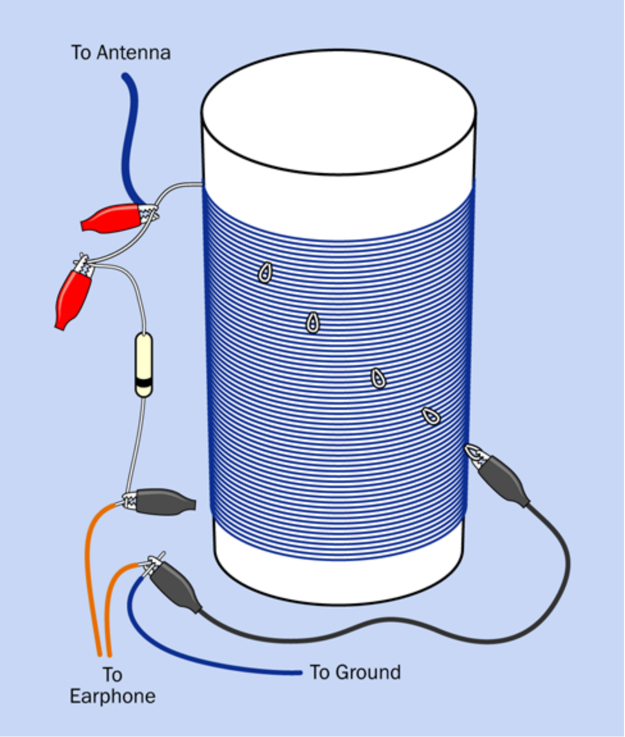

Figure 5-70 *The assembled components.*

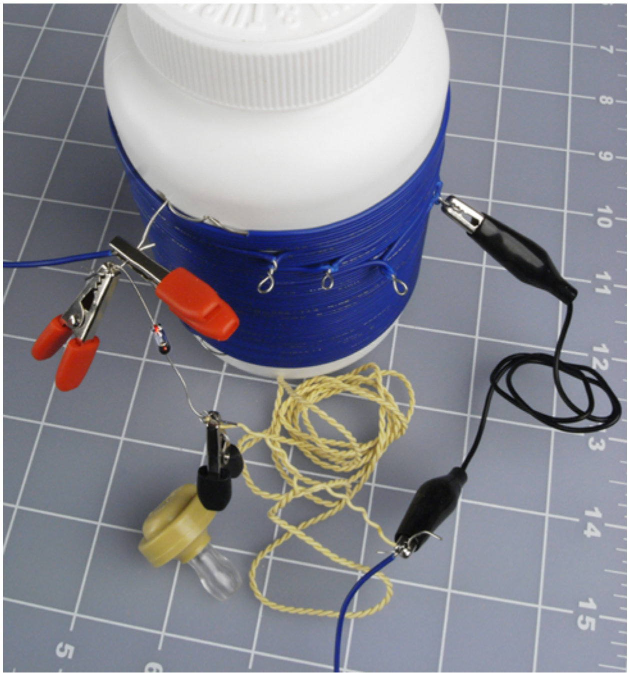

Figure 5-71 *The real-world version.*

If you followed the instructions, and you live within 20 or 30 miles of an AM radio station, and your hearing is reasonably good, you will be able to listen to the faint sounds of radio on your earphone—even though you are not applying any power to the circuit that you built. This project is many decades old, but can still be a source of surprise and wonder. (See Figure 5-72.)

If you live too far from a radio station, or you can't put up a very long antenna, or your ground connection isn't very good, you may not hear anything. Don't give up; wait till sunset. AM radio reception changes radically when the sun is no longer exciting the atmosphere with its radiation.

To choose among radio stations, move the alligator clip at the end of your test lead from one tap to another on your coil. Depending on where you live, you may pick up just one station, or several, playing individually or simultaneously.

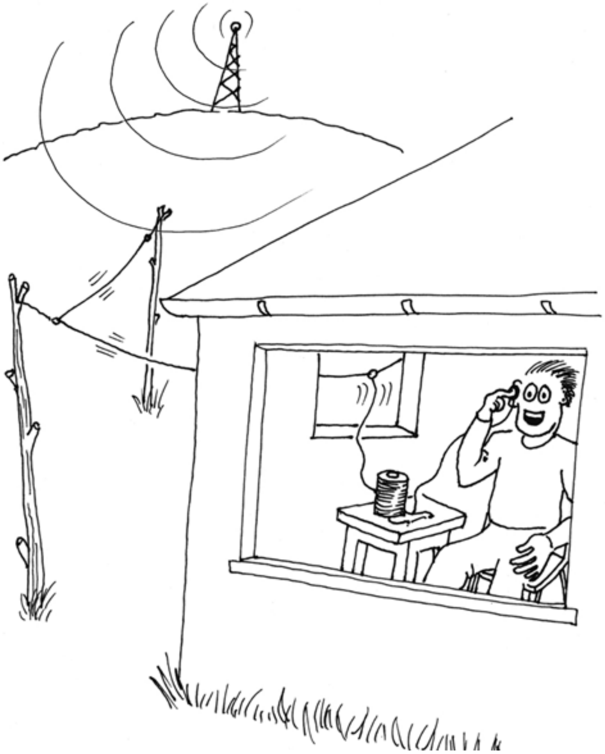

Figure 5-72 *The simple pleasure of picking up a radio signal with ultra-simple components and no additional power.*

It may seem that you're getting something for nothing here, but really you are taking energy from a source of power—the transmitter located at a radio station. A transmitter pumps power into a broadcasting tower, modulating a fixed frequency. When the combination of your coil and antenna resonates with that frequency, you're sucking in just enough voltage and current to energize a high-impedance headphone.

The reason you had to make a good ground connection is that power will only flow through your coil if it has somewhere to go. You can think of the ground as being an almost infinite power sink, with a reference voltage of zero. The transmitter at an AM radio station is also likely to have a potential relative to ground. See Figure 5-73.

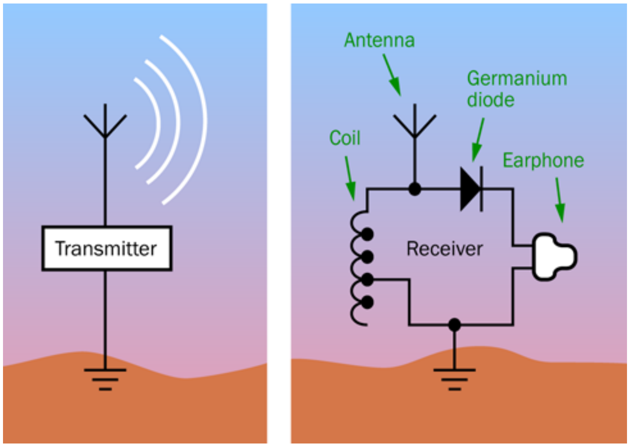

Figure 5-73 *Your no-power radio takes just enough energy from a distant transmitter to create a barely audible sound in your earphone.*

Enhancements

If you have difficulty hearing anything through your earphone, try substituting a piezoelectric transducer, also known as a piezo beeper. You need the type that does not have an oscillator built in, and functions passively, like a loudspeaker. Press it tightly against your ear, and you may find that it works as well as an earphone, or better.

You can also try amplifying the signal. Ideally you should use an op-amp for the first stage, because it has a very high impedance. However, I decided to put op-amps in *Make: More Electronics*, where I had room to explore the topic more thoroughly. As a substitute, you can feed the signal directly into the same LM386 single-chip amplifier from Experiment 29.

Figure 5-74 shows how simple the circuit can be. The germanium diode can connect directly with the LM386 input, as I don't think you'll need a volume control. Be sure to include the 10µF capacitor between pins 1 and 8, to increase the amplification of the chip to its maximum value. Even where I live, about 120 miles from Phoenix, Arizona, I was able to pick up a station broadcasting from the Phoenix area.

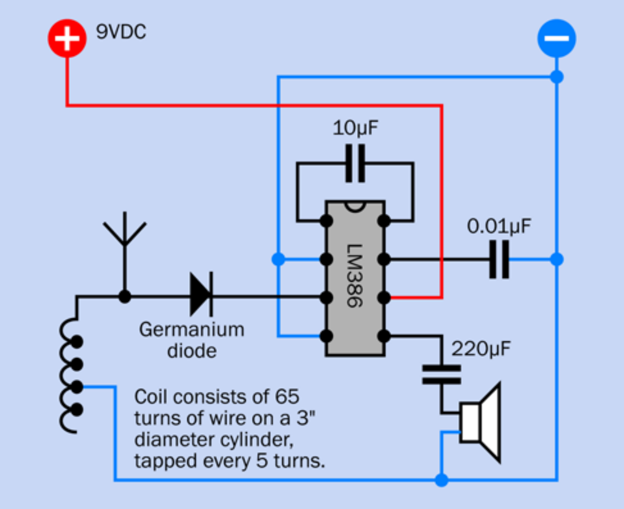
9VDC
10µF
0.01µF
LM386
Germanium diode
220µF
Coil consists of 65 turns of wire on a 3" diameter cylinder, tapped every 5 turns.

Figure 5-74 *The LM386 single-chip amplifier can make your crystal-set radio audible through a loudspeaker.*

If you want to improve the selectivity of your radio, you can add a variable capacitor to tune the resonance of your circuit more precisely. Variable capacitors are uncommon today, but you can find one at the same specialty source that I recommended for the earphone and the germanium diode: the Scitoys Catalog.

This source is run by a smart man named Simon Quellan Field, whose site suggests many fun projects that you can pursue at home. One of his clever ideas is to remove the germanium diode from your radio circuit and substitute a low-current LED in series with a 1.5-volt battery. This didn't work for me, because I live in a remote location; but if you're close to a transmitter, you may be able to see the LED varying in intensity as the broadcast power runs through it.

Theory: How Radio Works

High-frequency electromagnetic radiation can travel for miles. To make a radio transmitter, I could use a 555 timer chip running at, say, 850 kHz (850,000 cycles per second), and would pass this stream of pulses through an extremely powerful amplifier to a transmission tower—or maybe just a long piece of wire. If you had some way to block out all the other electromagnetic activity in the air, you could detect my signal and amplify it.

This was more or less what Guglielmo Marconi did when he performed a groundbreaking experiment in 1901, except that he had to use a primitive spark gap, rather than a 555 timer, to create the oscillations. His transmissions were of limited use, because they had only two states: on or off. You could send Morse code messages, and that was all.

Marconi is pictured in Figure 5-75.

Figure 5-75 *Guglielmo Marconi, the great pioneer of radio (photograph from Wikimedia Commons).*

Five years later, the first true audio signal was transmitted by imposing lower audio frequencies on the high-frequency carrier wave. In other words, the audio signal was "added" to the carrier frequency, so that the power of the carrier varied with the peaks and valleys of the audio. This is shown in Figure 5-76.

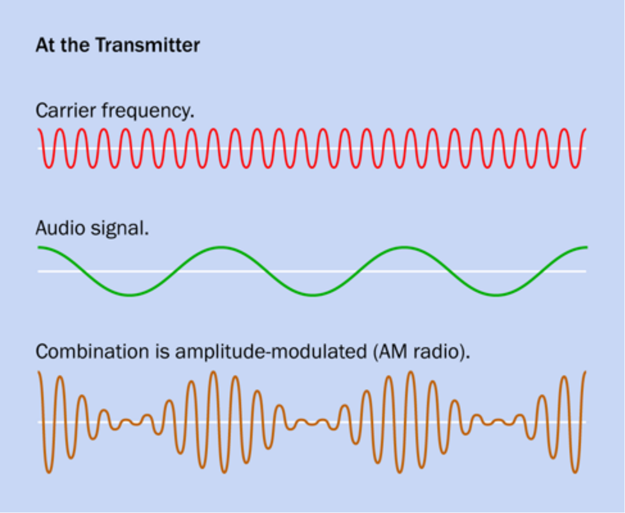

Figure 5-76 *Using a carrier wave of fixed frequency to transmit an audio signal.*

At the receiving end, a very simple combination of a capacitor and a coil detected the carrier frequency out of all the other noise in the electromagnetic spectrum. The values of the capacitor and the coil were chosen so that their circuit would resonate at the same frequency as the carrier wave. The basic circuit is shown in Figure 5-77, where the variable capacitor is represented by a capacitor symbol with an arrow through it.

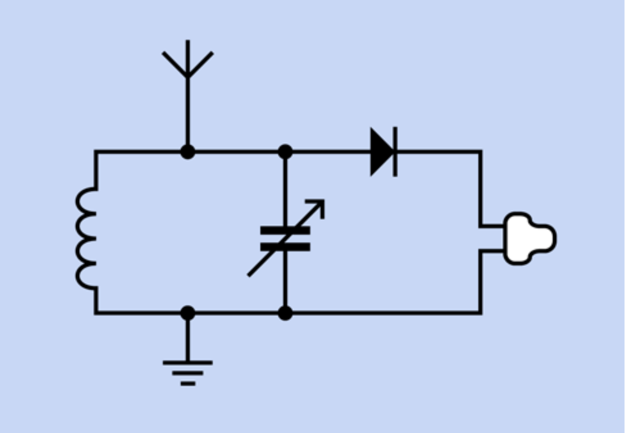

Figure 5-77 *When a variable capacitor is added to the previous circuit, it enables better discrimination among different signals sharing the spectrum.*

The carrier wave fluctuates up and down so rapidly, an earphone cannot possibly keep up with the positive-negative variations. It will remain hesitating at the midpoint between the highs and lows, producing no sound at all. A diode solves this problem by blocking the lower half of the signal, leaving just the positive voltage spikes. Although these are still very small and rapid, they are now all pushing the diaphragm of the earphone in the same direction, so that it averages them out, approximately reconstructing the original sound wave. This is shown in Figure 5-78.

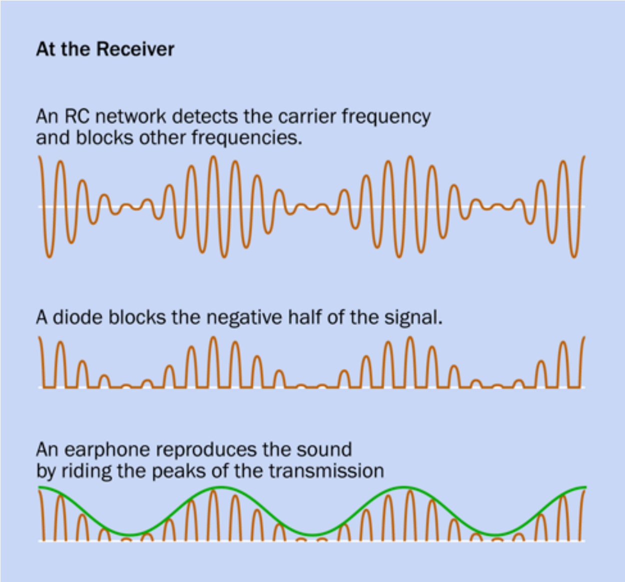
At the Receiver
An RC network detects the carrier frequency and blocks other frequencies.
A diode blocks the negative half of the signal.
An earphone reproduces the sound by riding the peaks of the transmission

Figure 5-78 *How a simple AM radio receiver decodes the signal and reproduces it on an earphone.*

When a capacitor is added to a receiving circuit, an incoming pulse from the transmitter is initially blocked by the self-inductance of the coil, while it charges the capacitor. If an equally negative pulse is received after an interval that is properly synchronized with the values of the coil and the capacitor, it coincides with the capacitor discharging and the coil conducting. In this way, the right frequency of carrier wave makes the circuit resonate in sympathy. At the same time, audio-frequency fluctuations in the strength of the signal are translated into fluctuations in voltage in the circuit.

If you are wondering what happens to other frequencies pulled in by the antenna, the lower ones pass through the coil to ground, while the higher ones pass through the capacitor to ground. They are just “thrown away.”

The waveband allocated to AM radio ranges from 300kHz to 3MHz in carrier frequencies. Many other frequencies are allocated for special purposes, such as ham radio. It’s not so difficult to pass the ham radio exam, and with appropriate equipment and a well-situated antenna, you can talk directly with people in widely scattered locations—without relying on any communications network to connect you.

Experiment 32: Hardware Meets Software

I’m guessing that many readers of this book—perhaps most readers—have heard of the Arduino. This experiment, and the next two, will show you how you can set up an Arduino and then write programs for it, instead of just downloading applications that you find online.

What You Will Need

- Arduino Uno board, or compatible clone (1)
- USB cable with A-type and B-type connector at opposite ends (1)
- A desktop or laptop computer with an available USB port (1)
- Generic LED (1)

Definitions

A microcontroller is a chip that works like a little computer. You write a program consisting of instructions that the microcontroller can understand, and copy them into some memory in the chip. The memory is nonvolatile, meaning that it will preserve its contents even when the power is switched off.

Normally I would suggest jumping right in and writing a program, but learning to use a microcontroller entails a bigger investment of time and mental energy than was required by the components that I have dealt with previously. How do you know if you want to get into this, until you know some more details? Therefore, I must begin with an explanation and orientation. After that, this experiment will take you through the process of setting up an Arduino and performing the most basic test. Experiments 33 and 34 will take you further into programming the Arduino and using other components in conjunction with it.

The setup-and-test procedure may take an hour or two. You should allocate the time when you will be able to follow the instructions without distraction. Once you have gone through this initial process, everything will be much easier.

Real-World Applications

A typical microcontroller application might run like this:

- Receive an input from a rotational encoder that is functioning as a volume control on a car stereo.
- Figure out which way the encoder is rotating.
- Count the pulses from the encoder.
- Tell a programmable resistor how many equivalent steps it should take, to adjust the volume of the stereo up or down.
- Wait for more inputs.

A microcontroller could also handle a much larger application, such as all the inputs, outputs, and decisions associated with the intrusion alarm system in Experiment 15. It would scan sensors, activate a siren through a relay after a delay period, receive and verify a keypad sequence when you want to shut off the alarm—and much more.

All modern cars contain microcontrollers, which deal with complex tasks such as timing the ignition in the engine—and simple matters such as sounding a chime if you don't fasten your seat belt.

A microcontroller can handle small but important tasks that I discussed in previous experiments, such as debouncing a pushbutton or generating an audio frequency.

Bearing in mind that one little chip can do so many different things, why don't we use it for everything?

The Right Tool for the Job?

A microcontroller is versatile and powerful, but is more suitable in some situations than others. It is ideal for applying logical operations along the lines, "If this happens, do that, but if that happens, do something else," but it adds cost and complexity to a project, and of course it entails a major learning process: you have to master the computer language that tells the microcontroller what to do.

If you don't want to learn the language, you can download and use programs that other people have written. Many Makers use this option, because it provides immediate results. You can find thousands of Arduino programs in libraries online, and they cost nothing.

But a program probably won't do exactly what you want. Inevitably, you will need to modify it—and you will be back in a situation where you need to understand the language in order to get full use out of the chip.

Writing a program for the Arduino can be relatively simple, depending on the application. However, it is not a one-step operation. The code has to be tested, and the processes of revision and debugging can be time-consuming. One little error will create unexpected results, or can stop anything from happening at all. You have to reread your code, find the mistakes, and try again.

After you have everything working, the results can be impressive. For this reason, personally I think that programming a microcontroller is worthwhile, so long as your expectations are realistic.

You will have to try it for yourself to find out if it's something you want to pursue actively.

One Board, Many Chips

I'll begin with the most basic question. What is an Arduino?

If you think that it's a chip, you are not quite correct. Each product with the Arduino name on it consists of a small circuit board designed by Arduino, containing a microcontroller chip made by a completely different company. The Arduino Uno board uses an Atmel ATmega 328P-PU microcontroller. The board also includes a voltage regulator, some sockets that enable you to plug in wires or LEDs, a crystal oscillator, a power-supply connector, and a USB adapter that enables your computer to communicate with the board. See Figure 5-79 for a photograph of the board with some of the components identified.

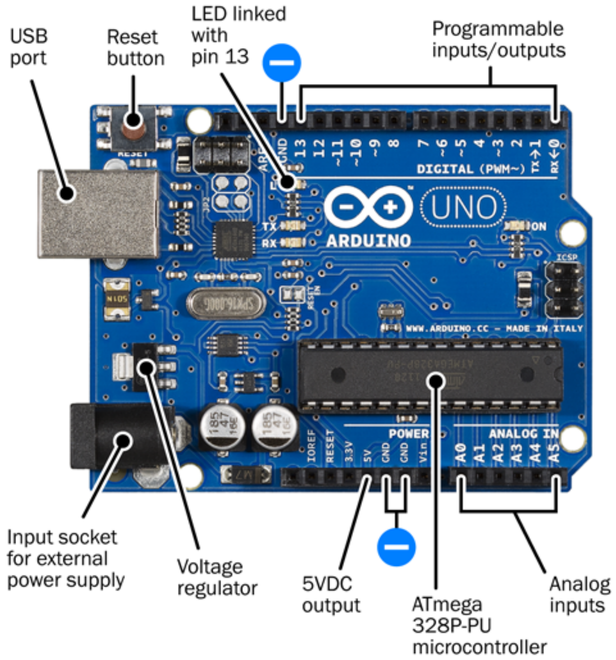
USB port
Reset button
LED linked with pin 13
Programmable inputs/outputs
Input socket for external power supply
Voltage regulator
5VDC output
ATmega 328P-PU microcontroller
Analog inputs

Figure 5-79 *An Arduino Uno board, featuring the ATmega 328P-PU microcontroller made by Atmel.*

If you buy an ATmega 328P-PU on its own from a component supplier, it will cost you less than one-sixth the retail price of an Arduino board containing the chip. Why should you have to pay so much more, just to get that little circuit board? The answer is that development of the board and the clever software that enables you to use it was not a trivial matter.

The clever software is called an IDE, which is an acronym for *integrated development environment*. After you install it on your computer, the IDE is a user-friendly place in which to write a program and then *compile* it, meaning that it converts the C language instructions (which human beings can understand) into machine code (which the Atmel chip will understand). You then copy it to the ATmega chip.

In case this is confusing, I'll summarize the story so far.

- An Arduino is a circuit board into which is plugged an Atmel microcontroller.
- IDE software written by the Arduino company enables you to write a program on your computer.
- After you write your program, the IDE software compiles it, to create code that the chip understands.
- The IDE software sends the code to the Atmel chip, which stores it.

Once your code is in the chip, the chip doesn't really need the Arduino board anymore. Theoretically, you could unplug the ATmega328 and use it somewhere else —in a breadboard, or in a circuit that you solder around it. The chip will still do what you programmed it to do, because the code remains stored in the chip.

In reality there are some little snags associated with this, but you can learn about them by reading a very good book titled *Make: AVR Programming* by Elliot Williams. It tells you exactly how to transplant the ATmega chip.

If you learn how to do this, the implications are significant. You will only need one Arduino board, while you can buy a lot of Atmel chips very cheaply. Put one chip into the board, program it, remove it, and use it in a standalone project. Put another chip into the board, transfer a different program into it, and use it in a different project.

This is relatively easy if you have the version of the Arduino Uno where the microcontroller is a through-hole chip mounted in a socket. You can pry it out with a miniature screwdriver, and substitute another chip by pushing it in with your finger and thumb. (Another version of the Uno comes with a surface-mounted microcontroller, soldered in. This version does not allow you to relocate the chip.)

Beware of Imitations?

Now that I have explained the basics, I'll explain how to set everything up.

While various models of Arduino exist, the one I'm going to deal with here is the Arduino Uno, and my instructions will apply to version R3 or later.

You can buy the Arduino board from many sources, because it was designed and marketed as an "open source" product, allowing anyone to create a copy of it, just as any manufacturer can make a 555 timer (although for slightly different reasons).

Mouser, Digikey, Maker Shed, Sparkfun, and Adafruit all sell genuine Arduino products. On eBay, however, you can find *unlicensed* copies of the Arduino board for one-third of the price. You can see that they are unlicensed, because they don't have the Arduino logo printed on them. To help you in distinguishing the real boards from the imitations, you can see the logo at Figure 5-80.

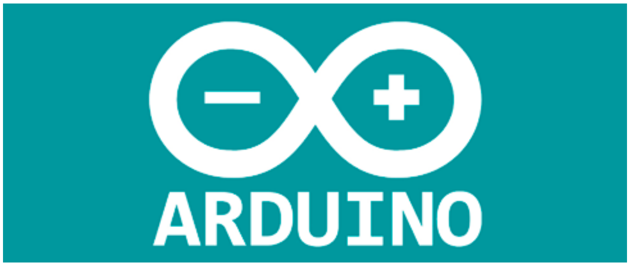

Figure 5-80 *Only the boards manufactured or licensed by Arduino are supposed to have this logo on them.*

The unlicensed boards are completely legal. This is not like buying pirated software or music. The only feature that Arduino chose to control is its trademark, which may not be used by other manufacturers. (Actually, some scammers use the logo illegally, but you'll know they are not genuine Arduino boards, because they're so cheap.) Just to make things more complicated, because of a dispute between Arduino and its former manufacturer, authentic Arduino boards are marketed under the name Genuino outside of the United States.

If you buy an imitation board, will it be reliable? I would trust the ones from Adafruit, Sparkfun, Solarbotics, Evil Mad Scientist, and a few others. I can't buy and test them all, so you have to make up your own mind about this, based on feedback from other buyers and your general impression of a supplier. But remember, you only have to buy one board, after which you can use it to program multiple Atmel chips, using the plan I described above. So maybe it's not such a big deal to pay a little extra for the genuine Arduino product. By doing so, you help the company to continue its development of new products in the future.

Personally, I bought a genuine Arduino board.

Setup

I will assume, now, that you have acquired an Arduino Uno, or an imitation that you believe is trustworthy. You will also need a standard USB cable with a type A plug on one end and a type B plug on the other end, as shown in Figure 5-81. This is not usually supplied when you buy the board on its own. If you don't have a spare cable, you can borrow one from another device while you do your setup and initial test. Cables are cheaply available online from sites such as eBay.

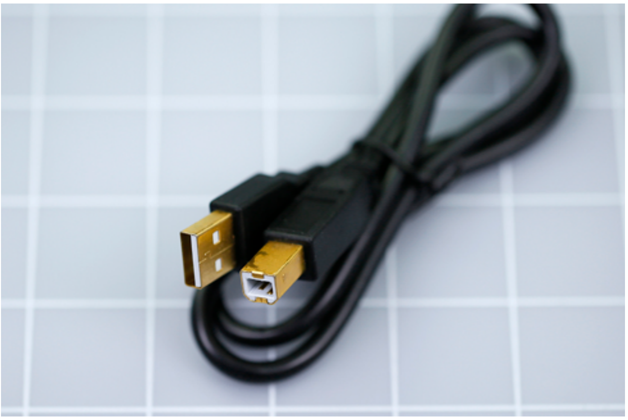

Figure 5-81 *You will need this type of USB cable to attach your Arduino board to a USB port on your computer.*

You have the board and the cable, so now you need the IDE software. Go to the Arduino website and click the "download" tab. Then choose the IDE software that is appropriate for your computer. It is currently available in variants for Mac OS, or Linux, or Windows. I will be using version 1.6.3, but my instructions should also apply to later versions. You can download the IDE software for free.

Note that you need a computer running Windows XP or later, or running Mac OS X 10.7 or later, or Linux 32-bit, or Linux 64-bit. (These requirements are valid at the time of writing. Arduino may change its requirements in the future.)

The procedure for setup under the three different operating systems is described below, primarily based on instructions in the nice little introductory guide titled *Getting Started with Arduino* by Massimo Banzi and Michael Shiloh. You can also find installation instructions on sites such as SparkFun and Adafruit. Lastly, there are instructions on the Arduino website as well.

Unfortunately, all these instructions are slightly different. For instance, Arduino's website tells me to plug in my board before running the installer, but the *Getting Started* book tells me to run the installer before plugging in my board. This makes my task difficult, because both the Arduino website and the *Getting Started* book were written in collaboration with Arduino developers.

Below I will give you my best guess about what will work on each system.

Linux Installation

This is likely to be the most challenging, as there are so many variants of the operating system. I will have to refer you to the Arduino site for guidance.

I regret that I can't help you, myself, with Linux.

Windows Installation

I am inclined to use the procedure recommended in *Getting Started with Arduino*. This is also recommended at the website maintained by Sparkfun.

Do not plug in the board yet. First, identify the IDE installation program that you downloaded. Its name may be something like *arduino-1.6.3-windows.exe*, although by the time you read this, the version numbers almost certainly will have changed. The *.exe* at the end of the file name may not be visible, depending on the system settings on your computer.

- Some guides refer to the installation download as being a zipped file, which has to be unzipped. So far as I can determine, file zipping has been discontinued by Arduino. You can run it as is.

Double-click the icon, and you should see an installation sequence that is familiar to you from installing other software from other vendors.

You have to agree to the terms of a licensing agreement (if you don't agree, you can't run the software).

You are asked if you want shortcuts on your desktop and in your Start menu. Allow the shortcut to be added to your desktop. The Start menu is up to you.

You will be asked to approve the installation folder for the IDE software, and you should accept the default that is suggested.

If you are a dinosaur like me, still using Windows XP (and actually, there are quite a few million of us), you may see a warning such as the one shown in Figure 5-82. The exact look of this will depend on your version of Windows. Ignore the warning by choosing to "continue anyway." If you are asked permission to install device drivers, say "yes."

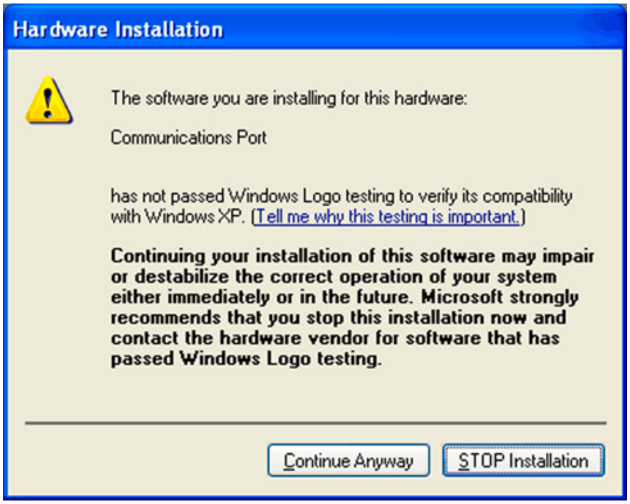

The software you are installing for this hardware:

Communications Port

has not passed Windows Logo testing to verify its compatibility with Windows XP. (Tell me why this testing is important.)

Continuing your installation of this software may impair or destabilize the correct operation of your system either immediately or in the future. Microsoft strongly recommends that you stop this installation now and contact the hardware vendor for software that has passed Windows Logo testing.

Figure 5-82 *Users of the old WinXP operating system can ignore this warning.*

- On Windows 8, a security feature prevents you from installing unsigned device drivers. This should not be an issue with modern versions of the Arduino IDE installer, but if you somehow obtain older versions from somewhere else, you can Google:

  `sparkfun disable driver signing`

  which should take you to the section of the Sparkfun site that contains helpful advice on the topic.

After installing the IDE software, plug your Arduino into your computer with the USB cable.

You don't need to use the circular power input jack on the Arduino board, so long as the board is connected with a USB port on your computer. The board receives power through the USB cable. Bear in mind that a shorter, thicker cable will minimize voltage drop. Also, if you are using a laptop computer, especially if it is an older one, it may limit the power that you can draw from a USB port to 250mA. Even on a desktop computer, which is supposed to deliver 500mA via USB, that power may be shared across three or four USB ports. A device such as an external hard drive can take a significant amount.

Watch the board while it identifies itself to the computer. You should see a green LED that lights up steadily, and a yellow LED that blinks. A couple more LEDs nearby on the board, labeled TX and RX, should flicker briefly. They show that data is being transmitted and received.

On your computer, go to the shortcut for the IDE software. It is simply named "Arduino," and the installer placed it on your desktop. If you don't want it on the desktop, drag it to a different location. Double-click it to launch the Arduino IDE software.

In the window that opens, pull down the Tools menu, go to the Ports submenu (which is currently named Serial Ports in the Mac version), and you should see a list of serial ports on your computer. They will be named COM1, COM2, and upward.

What is a serial port? In the early days of Windows (and before that, MS-DOS) computers didn't have USB connectors. They used "serial protocol" through D-shaped connectors, and the computer kept track of the connector that it was using by assigning it a "port number." This system is still buried in Windows, even though decades have passed since it was established and the protocol has become rare for domestic applications.

All you need to know is if the Arduino IDE software and Windows have agreed on the port number that has been assigned to your Uno board. Ideally, when you choose Tools > Ports in the IDE software, you will see the Uno in the list with a check mark beside it, and everything is fine. If so, skip the troubleshooting section that follows, and go directly to the Old Arduino Blink Test, below (see "The Old Arduino Blink Test" on page 284).

Windows Troubleshooting

There are two bad possibilities that can affect your port assignment.

- In the Port submenu of the Arduino IDE software, you may see the Arduino Uno listed, but it is not checkmarked. Another port may be checkmarked instead. Try checkmarking the correct port. You may see a warning if the IDE software doesn't approve of the Uno board that you are using. Ignore the warning, click the box saying "Don't show me this again," and proceed to the Old Arduino Blink Test (see "The Old Arduino Blink Test" on page 284).
- You may see no port in the list labeled Arduino Uno. In that case, make a note of the COM ports that are listed. Close the IDE menus. Unplug the Uno board. Wait five seconds. Open the IDE Port submenu again, and see which COM port has disappeared. Close the submenu. Plug the Uno back in. Reopen the submenu. Click the port that has reappeared, to checkmark it. Proceed to the Old Arduino Blink Test (see "The Old Arduino Blink Test" on page 284).

Windows allows you an option to verify your port settings. Click your Start menu and choose the Help and Support service. In the window that opens, type the words "Device Manager" as your search term. It should be the first hit in the list of search results. Open Device Manager. If you are using Windows XP, Device Manager shows you a Ports list. On versions of Windows later than XP, when you go to Device Manager you may have to select View > Show Hidden Devices to reveal the Ports list.

You should see your Arduino Uno listed there. If there is a yellow circle beside it, or an exclamation point, right-click it to find out what's wrong.

If Windows complains that it cannot find the device driver for your board, tell it to look in the Arduino folder containing all the files that were extracted by the installer.

One known issue with the Arduino Uno and Windows ports is that the IDE software can get confused if you have more than nine ports already assigned. This is unusual, but if you have this problem, try unassigning some ports, or manually assign an unused port that has a single-digit number.

If you are still having difficulties, go to "If All Else Fails" on page 284 below.

Mac Installation

After completing the IDE installer download, locate the icon that the computer has created, double-click it, and you'll see a disk image that contains the Arduino IDE software. You can drag this into your Applications folder.

Now plug your Arduino into your computer with the USB cable.

- You don't need to use the circular power input jack on the Arduino board, so long as the board is connected with your computer. The board takes its power through the USB cable.

Watch the board while it identifies itself to the computer, and you should see a green LED that lights up steadily, and a yellow LED that blinks. A couple more LEDs nearby on the board, labelled TX and RX, should flicker briefly. They show that data is being transmitted and received.

If a window opens telling you that a "new network interface" has been detected, click Network Preferences, and then Apply. It doesn't matter if the Uno is described as "not configured." Close the window.

Double-click the program icon for the Arduino IDE that you dragged into your Applications folder. You need to select the correct port for communication with the Uno board. In the Tools menu of the IDE software, click the Serial Port option, and select `/dev/cu.usbmodemfa141` (or a similarly named port) from the list that pops up.

If everything has proceeded as described here, you can continue with the Old Arduino Blink Test (see "The Old Arduino Blink Test" on page 284).

If All Else Fails

This book may stay in print for a while. At least, I hope it does! Software, on the other hand, changes frequently. My installation instructions for the Arduino IDE may be out of date by the time you read this.

I will try to revise my instructions in each new printing and each new ebook version, so that they will be as accurate as possible. But of course you may happen to be reading an older printing or an older ebook.

What to do? Your best bet is to go to the Arduino site, or the Sparkfun site, and follow installation procedures listed there. A website is more easily and quickly updated than a book.

The Old Arduino Blink Test

I am assuming that you have launched the IDE software. Its main window should look something like the screenshot in Figure 5-83, although subsequent versions may change somewhat.

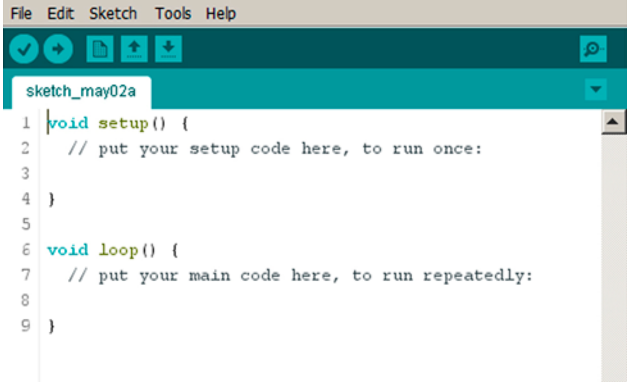
File
Edit
Sketch
Tools
Help
sketch_may02a

```
void setup() {
  // put your setup code here, to run once:

}

void loop() {
  // put your main code here, to run repeatedly:

}
```

Figure 5-83 *Default window that opens when the Arduino IDE is launched.*

Before you can make your Arduino do something, you have to check that the IDE software has correctly identified the version of the board that you connected with your computer.

In the IDE main window, pull down the Tools menu, open the Boards submenu, and verify that the Arduino Uno has a bullet point beside it, as in Figure 5-84. If it doesn't, click it to select it.

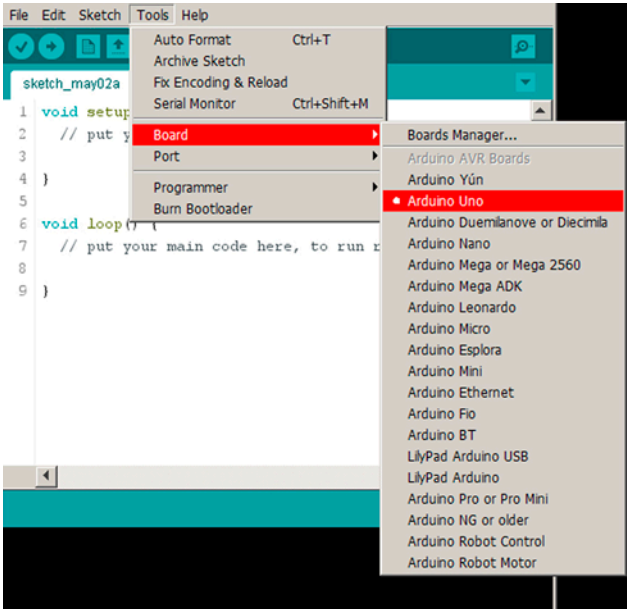
File
Edit
Sketch
Tools
Help
sketch_may02a
Auto Format Ctrl+T
Archive Sketch
Fix Encoding & Reload
Serial Monitor Ctrl+Shift+M
Board
Port
Programmer
Burn Bootloader
Boards Manager...
Arduino AVR Boards
Arduino Yún
• Arduino Uno
Arduino Duemilanove or Diecimila
Arduino Nano
Arduino Mega or Mega 2560
Arduino Mega ADK
Arduino Leonardo
Arduino Micro
Arduino Esplora
Arduino Mini
Arduino Ethernet
Arduino Fio
Arduino BT
LilyPad Arduino USB
LilyPad Arduino
Arduino Pro or Pro Mini
Arduino NG or older
Arduino Robot Control
Arduino Robot Motor

```
void setup
  // put y
}

void loop()
  // put your main code here, to run r

}
```

Figure 5-84 *When you are using the Arduino Uno, it should have a bullet point beside it in the Boards Manager submenu.*

Now you are ready to give your Arduino some instructions. At the top of the workspace in the main IDE window, you'll see the word "sketch" followed by today's date and the letter "a." What is this "sketch"? Is it a picture that you are going to draw?

No, in the Arduino world, "sketch" means the same thing as "program." Perhaps this is because the developers didn't want to make people feel intimidated by the idea that they were going to program a computer. Similarly, while he was still alive, Steve Jobs seemed to feel that users of handheld devices would be more comfortable if he referred to programs as "apps." Jobs was probably right, but where Makers are concerned, I don't think they are so easily intimidated. In fact I think they *want* to program computers. Otherwise, why would you be reading this?

"Sketch" means "program" for the Arduino, but I will continue to use the word "program," because that is what it really is, and I feel silly calling it a "sketch." When you read source materials online, people use the word "program" at least as often as they use the word "sketch."

Now I need to remind you of the sequence of events that we are going to follow. First, you *write a program* in the IDE window. Then you *compile* it by selecting a menu option, to convert it into instructions that a microcontroller can understand. Then you *upload* it to the Arduino board, and then the board automatically *runs the program*.

The IDE window on my computer contains some default text shown in Figure 5-83. Future versions of the IDE may do things a little differently, but the principle will be the same. You will see some lines beginning with two slash marks, like this:

```
// put your setup code here, to run once.
```

This is known as a *comment line*. It is for human interest, to explain what's going on.

- When the program that you write is compiled for the microcontroller, the compiler will ignore all lines that begin with // marks.

The next line reads:

```
void setup() {
```

This is a line of *program code*, for the compiler and the microcontroller to understand. But you need to know what it means, because the setup routine has to be at the beginning of every Arduino program, and I'm hoping you will start to write your own programs in the future.

The word `void` tells the compiler that this procedure won't generate any numerical result or output.

The term `setup()` says that the following procedure is something that has to be done once only, right at the beginning.

Notice that there is a { mark following `setup()`.

- Every complete function in C language should be contained within a { mark and a } mark.

Because a `{` mark must always be followed by a `}` mark, there must be a `}` mark somewhere on the opening screen. Yes, it's a couple of lines farther down. There's nothing between the `{` mark and the `}` mark, so there are no instructions in this procedure. This is because you are going to write them.

- It doesn't matter if { and } are on separate lines. The compiler for the Arduino ignores line breaks and all whitespace larger than a single space between words.
- The { and } are properly known as *braces*.

It's time to type something on the empty line beneath "put your setup code here." Try this:

```
pinMode(13, OUTPUT);
```

You have to type it exactly. The compiler won't tolerate typographical errors. Also, because the C language is *case sensitive*, you must distinguish between uppercase and lowercase letters. `pinMode` must be `pinMode`, not `pin mode` or `Pinmode`. `OUTPUT` must be `OUTPUT`, not `output` or `Output`.

The word `pinMode` is a command to the Uno, telling it how to use one of its pins. The pin can receive data as an input, or can send data as an output. `13` is the pin number, and if you check your board, you'll find one of the little connectors is identified as 13, right beside the yellow LED. I have selected 13 as an arbitrary choice.

A semicolon marks the end of the instruction.

- A semicolon must be included at the end of each instruction. Always. Don't forget!

Now move on down to the empty line below the message that says:

```
// put your main code here, to run repeatedly.
```

You can tell from the double slash marks that this is another comment line for your information. The compiler will ignore it. Below it, carefully type these instructions:

```
void loop() {
digitalWrite(13, HIGH);
delay(100);
digitalWrite(13, LOW);
delay(100);
}
```

If you have any prior familiarity with an Arduino, you'll be groaning, now, as you think to yourself, "Oh no, it's the old blink test!" Yes indeed, and that's why I subtitled this section, "The Old Arduino Blink Test." This is the program that almost everyone uses as a preliminary test (although I have changed the delay values, for reasons that will become apparent). Please humor me by typing the program into the IDE. I will be getting to some more challenging projects soon enough.

Also, you may be just a little sketchy on what some of the statements actually mean.

`void` means the same as before.

`loop()` is an instruction telling the Arduino to do something over and over again. What does it have to do? It has to obey the procedure between the braces.

`digitalWrite` is a command to send something out of a pin. Which pin? I am specifying `13` because its mode was defined previously.

- You can't use a digital pin unless you have previously specified what its mode will be.

What should the pin do? Go to a HIGH state.

Don't forget the semicolon at the end of the instruction.

`delay` tells the Arduino to wait for a while. How long? `100` means 100 milliseconds. There are 1,000 milliseconds in a second, so the Arduino is going to wait for one-tenth of a second. During that time, pin 13 will stay high.

I think you can figure out what the next two lines mean.

In a moment, you can activate the program. But first, go to your board and insert the leads of an LED between connector 13 and the connector labelled GND, right beside it.

- Make sure the short lead of the LED is in the GND connector.
- The LED does not need a series resistor, as this is built in to connector 13.

The little yellow LED on my board was already blinking, by default, as soon as I plugged in the board. The LED that I just inserted also starts blinking, because the yellow surface-mount LED built into the board is wired in parallel with pin 13.

In previous versions of the Uno, the onboard LED did not start blinking as soon as you plugged in the board. In future versions, Arduino may disable this "blinking by default." Either way, it doesn't matter, because your program is going to change the blink rate.

Verify and Compile

Next you must check if you made any typographical errors. Pull down the Sketch menu and choose Verify/Compile, as shown in Figure 5-85. The IDE examines your code, and if it sees any problems, it complains about them.

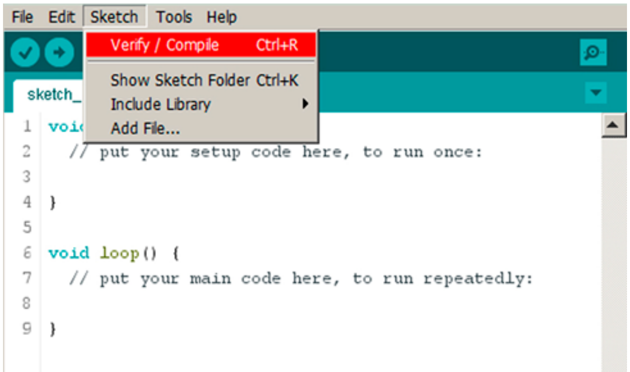

```
voi
  // put your setup code here, to run once:

}

void loop() {
  // put your main code here, to run repeatedly:

}
```

Figure 5-85 *Choose the Verify/Compile option before sending your program to the Arduino.*

You can test this. In your program listing, change `pinMode` to `piMode`, then Verify/Compile, and see what happens.

You get an error message in the black space at the bottom of the IDE window. You can stretch this black space by dragging its upper border with your mouse, so you can see more than two lines at a time, without scrolling them. The error message that I get says, "*piMode* was not declared in this scope."

You see, there are *reserved words* and *defined functions* in the C language, which have special meanings. You have used a couple of them already, such as `digitalWrite` and `delay`.

But `piMode` doesn't exist as a reserved word or defined function, so the compiler complains that you have not declared it, to say what it is.

Fix your program text till you can Compile/Verify without any errors.

Upload and Run

Now pull down the File menu, and choose Upload. Personally I always imagine that I am downloading from my big computer into the little Arduino, but everyone calls it uploading, so I guess that's what it is.

If the upload is successful, you'll see the message "Done Uploading" just above the black error window.

If it is not successful, and the uploading process never quite ends—this is not good. It means you still have some communications problems, probably because the COM port assignments don't match. Go back to the troubleshooting section for your type of computer, above. But save your program first. Pull down the File menu, choose Save, and give your program a name. After you fix your COM issues, you will be able to reload the program, if necessary, and try again.

If everything worked as planned, the onboard yellow LED, and your LED, are now blinking rapidly—on for 1/10th of a second, and off for 1/10th of a second, in accordance with the instructions in your program.

You may feel that this is a small achievement after taking so many steps, but we have to start somewhere, and a blinking LED is usually where microcontroller programming begins. In the next experiment, you will write a new program that will do something much more useful.

Here's a summary of what you may have learned so far, and what you generally have to do, to program the Arduino:

- Start a new program (or "sketch," as Arduino prefers to call it).
- Select the New option from the File menu if necessary.
- Every program must begin with a `setup` function, which will run once only.
- You must declare the number of a digital pin, and its mode, by using the `pinMode` command, before you can do something with the pin later.
- The mode of a pin can be INPUT or OUTPUT.
- Some pin numbers are not valid. Look at your Uno board to see the numbering system that is used.
- A pair of braces must enclose every function or block in a program. But they can be on separate lines.
- The compiler ignores line breaks and extra whitespace.
- Every instruction in a function or block must end in a semicolon.
- Every Arduino program must contain a `loop` function (after the `setup` function), which will run repeatedly.
- `digitalWrite` is a command to make a pin that is set for output have a state that is specified as HIGH or LOW.
- `delay` makes the Arduino do nothing for a specified number of milliseconds (thousandths of a second).
- The numbers in parentheses after a command are *parameters* telling the Arduino how the command must be applied.
- Use the Verify/Compile option in the Sketch menu to check your program before you try to upload it to the Arduino.
- You must fix any errors found in the Verify/Compile operation.
- Reserved words are a vocabulary of commands that the Arduino understands. You have to spell them correctly. Uppercase is considered different from lowercase.
- After you upload your program, it will start running automatically, and will continue to do so

until you disconnect power from the board or upload a new program.

- There is a Reset button (a tactile switch) beside the USB connector on the Uno board. When you press it, the Arduino restarts your program at the beginning, with all the values reset.

Caution: Lost Code

If you modify your program and upload it to the microcontroller, the new version will *overwrite* the old version. In other words, the old version will be erased. If you didn't save it on your computer under a different filename, it may be gone forever. Be very careful when uploading revised programs. Saving each version on your computer, under a new name, is a sensible precaution.

After program instructions have been uploaded into the microcontroller, there is no way to read them back out again.

Programming Entails Detail

I don't know if you noticed, but the summary of points to remember, in this experiment, has been much longer than the summaries in any experiments where you used individual components. Writing a program entails a lot of detail, and you have to get everything exactly right. Personally I enjoy this, because once you get it right, it will always be right, and it will always work the same way. Programs never wear out. If you save them in a suitable medium, they can last forever. The software that I wrote in the 1980s will still run 30 years later in a DOS window on my desktop computer.

Some people don't enjoy detail work, or they tend to make typing errors, or they don't like the way in which a computer language makes inflexible demands (such as *always* insisting that you begin a program with a `Setup` function, even if you don't have a setup). Different kinds of people enjoy different aspects of electronics, and that's the way it should be. If everyone wanted to write programs, and no one ever wanted to touch a piece of hardware, we wouldn't have any computers. It's up to you to decide what activity suits you.

Personally, I'm going to continue with another experiment that uses the Arduino in a more interesting way. I want to show you how a microcontroller can do things more easily than individual components in some cases.

Before I end this experiment, though, you may be wondering what happens if you disconnect the Arduino from your computer.

- The Arduino needs power to *run* your program.
- The Arduino does not need power to *store* your program. The program is stored automatically in the microcontroller, like data in a flash drive.
- If you want to run your program when the board is not connected with your computer, you have to provide power in the round black socket next to the USB socket on the board.
- The power supply can range from 7VDC to 12VDC. It does not have to be regulated, because the Arduino board contains its own regulator, which changes your power input to 5VDC on the board. (Some Arduinos use 3.3VDC, but not the Uno.)
- The power supply jack is 2.1mm in diameter, with center pin positive. You can buy a 9V AC-to-DC adapter with that kind of plug on its output wire.
- If you connect external power while the Arduino is also connected with a USB cable, the Arduino automatically uses the external power.
- You can disconnect your Arduino from the serial cable anytime, without bothering to use the "Safely Remove Hardware" option that exists in some versions of Windows.

Background: Origins and Options Among Programmable Chips

In factories and laboratories, many procedures are repetitive. A flow sensor may have to control a heating element. A motion sensor may have to adjust the speed of a motor. Microcontrollers are perfect for this kind of routine task.

A company named General Instrument introduced an early line of microcontrollers in 1976, and called them PICs, meaning Programmable Intelligent Computer—or Programmable Interface Controller, depending on which historical source you believe. General Instrument sold the PIC brand to another company named Microchip Technology, which owns it today.

Arduino uses Atmel microcontrollers, but PICs are still an alternative, and are used as the basis for a hobby-education version licensed by a British company named Revolution Education Ltd. They call their range of chips the PICAXE, for no apparent rational reason other than they must think that it sounds cool. (I'm not so sure that they're right about that.)

The PICAXE comes with its own IDE, which uses a different computer language named BASIC. In some ways this is a simpler language than C. Another range of microcontrollers, the BASIC Stamp, also uses BASIC, with additional, more powerful commands.

If you search for PICAXE on Wikipedia, you'll find an excellent introduction to all the various features. In fact, I think it's a clearer overview than you'll get from the PICAXE website.

Unlike the Arduino, you don't need to buy a special board to program PICAXE chips. A customized USB cable is all you need—in addition to the appropriate IDE software, which you can download for free.

The first edition of this book contained some introductory information about PICAXE products. You can probably find that edition secondhand, if you're interested.

Fundamentals: Advantages and Disadvantages

Now that you have learned some basics, I need to discuss issues that may affect your decision about whether to use a microcontroller in a project.

Longevity

The flash memory that stores a program in an ATmega328 is guaranteed by the manufacturer for 10,000 read-write operations, with provision for automatically locking out memory locations that go bad. This would seem to be ample, and we may hope that a microcontroller will last almost indefinitely. However, we don't know yet if it will really have the same longevity as an old-school logic chip, some of which are still working 40 years after they were manufactured. Does this matter? You have to make up your own mind about that.

Obsolescence

Microcontrollers are maturing rapidly as a technology. When I wrote the first edition of this book, the Arduino was relatively new, and its future was uncertain. The Arduino now dominates the hobby-electronics field, but what will the situation be another five years from now? No one knows. A product such as the Raspberry Pi is an entire computer on a chip. No one can predict whether it, or something else like it, will displace the Arduino.

Even if the Arduino remains the microcontroller system of choice, we have already seen new versions of the hardware, and updates to the IDE software that must be used when you program the chip. One way or another, you may have to keep yourself informed about developments in the field as they occur, and you may even have to abandon one brand of microcontroller and switch to a different brand.

By comparison, in most cases, individual components in through-hole format have reached the end of their development cycle. Some relatively recent innovations have been introduced, such as rotational encoders or small dot-matrix LED and LCD displays. However, most of these new products are designed for use with microcontrollers. In the simple world of transistors, diodes, capacitors, logic chips, and single-chip amplifiers, the knowledge that you acquire today should be valid 10 years from now.

Hybrid Circuits

Last, and perhaps most important, microcontrollers cannot be used alone. They always require some other component, even if it's just a switch, a resistor, or an LED, and the other components must be properly compatible with the inputs and outputs of the microcontroller.

Therefore, to make practical use of a microcontroller, you still have to be familiar with electronics generally. You need to understand basic concepts such as voltage, current, resistance, capacitance, and inductance. You should probably know about transistors, diodes, alphanumeric displays, Boolean logic, and other topics that I have covered so far in this book. And if you are going to build prototypes, you will still need to know how to use a breadboard or make solder joints.

Bearing all this in mind, I can summarize the pros and cons.

Individual Components: Advantages

Simplicity.

Instant results.

No programming language needed.

Cheap, for small circuits.

Today's knowledge will be valid tomorrow.

Better for analog applications such as audio.

Still necessary when using microcontrollers.

Individual Components: Disadvantages

Capable of performing one function only.

Circuit design is challenging for applications involving digital logic.

Not easily scalable. Large circuits are difficult to build.

Revisions to a circuit may be difficult or even impossible.

More components in a circuit generally require more power.

Microcontrollers: Advantages

Extremely versatile, able to perform many functions.

Additions or revisions to a circuit can be easy (just rewrite the program code).

Huge online libraries of applications, freely available.

Ideal for applications involving complex logic.

Microcontrollers: Disadvantages

Relatively expensive for small circuits.

Significant programming skills required.

Time-consuming development process: write code, install code, test, revise-and-debug, reinstall—in addition to troubleshooting the circuit hardware.

Rapidly evolving technology requires a continuing learning process.

Each microcontroller has individual quirks and features, requiring study and memorization.

Greater complexity means more things that can go wrong.

Requires a desktop or laptop computer, and data storage for programs. Data may be lost accidentally.

Requires a regulated power supply (usually 5VDC or 3.3VDC), like a logic chip. Limited output of 40mA per pin, or less. Cannot drive a relay or a loudspeaker, as a 555 timer can. You must buy a separate driver chip if you need to deliver more power.

Summing Up

Now I am ready to answer the question, "Should I use microcontrollers or individual components?"

My answer is that you need both. That is why I am including microcontrollers in a book that is primarily about individual components.

In the next experiment, I'll show how a sensor and a microcontroller can work together.

Experiment 33: Checking the Real World

A switch is either "on" or "off," but most of the inputs we receive from the real world tend to vary between those extremes. A thermistor, for instance, is a sensor that varies its electrical resistance over a wide range, depending on its temperature.

A microcontroller would be very useful if it could process that kind of input. For instance, it could receive the input from the thermistor, and then it could function like a thermostat, turning on a heater if the temperature falls below a minimum value, and turning off the heater when the room is warm enough.

The ATmega328 that is used in the Arduino Uno can do this, because six of its pins are classified as "analog inputs," meaning that they don't just evaluate an input as "logic-high" or "logic-low" on a digital basis. They convert it internally using what is called an *analog-to-digital converter*, or ADC.

On the 5-volt version of the Arduino, an analog input must range from 0VDC to 5VDC. (Actually, the upper limit can be modified, but that introduces some complexity that I will leave until later.) A thermistor doesn't generate any voltage; it just varies its resistance. So, I'll have to think of a way in which a change in resistance can deliver a change in voltage.

Once that problem has been dealt with, an ADC inside the microcontroller will convert the voltage on the analog pin into a digital value ranging from 0 through 1023. Why that numeric range? Because it can be expressed in

10 binary digits, and the ADC isn't accurate enough to justify a wider range with smaller increments.

After the ADC has supplied a number, your program can compare it with a target value, and can take appropriate action—such as changing the state of an output pin, which can supply voltage to a solid-state relay, which can activate a room heater.

The sequence beginning with the thermistor and ending with the digital value is shown visually in Figure 5-86.

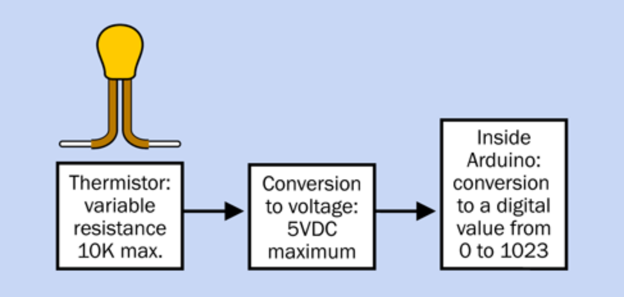
Thermistor: variable resistance 10K max.
Conversion to voltage: 5VDC maximum
Inside Arduino: conversion to a digital value from 0 to 1023

Figure 5-86 *A simplified view of the plan for processing the status of a thermistor.*

The following experiment will show you how to do this.

What You Will Need

- Breadboard, hookup wire, wire cutters, wire strippers, test leads, multimeter
- Thermistor, 10K, 1% or 5% accuracy (1) (This must be the NTC type, meaning that its resistance drops as the temperature increases. A PTC thermistor behaves oppositely)
- Arduino Uno board (1)
- Laptop or desktop computer with an available USB port (1)
- USB cable with A-type and B-type connector at opposite ends (1)
- 6.8K resistor (1)

Using a Thermistor

The first step is to get to know your thermistor. It has very thin leads, because they must not conduct heat into or out of the tip where the temperature-measuring junction is located. The leads are probably too thin to plug into your breadboard reliably, so I suggest you grab them with a couple of alligator test leads, and use the leads to grab the probes of your meter, as shown in Figure 5-87.

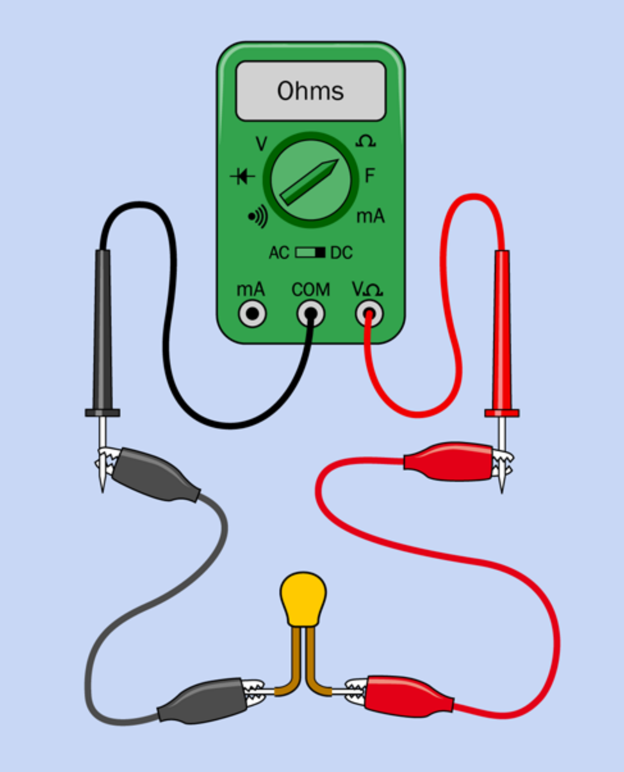

Figure 5-87 *Testing a thermistor.*

The thermistor that I am recommending is rated at 10K. That's its maximum resistance when it gets really cold. Its resistance doesn't change much as the temperature rises until around 25 degrees Celsius (77 degrees Fahrenheit). After that, the resistance declines more rapidly.

You can test it with your meter. At room temperature, the thermistor should have a resistance around 9.5K. Now grip it between your finger and thumb. As it absorbs your body heat, its resistance goes down. At body temperature (arbitrarily agreed to be 37 degrees Celsius or 98.6 degrees Fahrenheit) its resistance is around 6.5K.

How can you convert this resistance range to the 0V to 5V required by the microcontroller?

First bear in mind that the maximum value corresponding with room temperature should actually be lower than 5V. The real world is an unpredictable place. What if your thermistor gets much hotter than you expected, for some surprising reason? Maybe you place your soldering iron beside it, or maybe you rest it on a warm piece of electronic equipment.

Here we have the first lesson in analog–digital conversion: allow for unexpected, extreme values when measuring the everyday world.

Range Conversion

The simplest way to convert the thermistor's resistance into a voltage value is to choose a resistor approximately equal to the average resistance of the thermistor in the temperature range that interests us. Put the resistor and the thermistor in series to create a voltage divider, apply 5VDC at one end and 0VDC at the other end, then take the voltage at the midpoint between the components, as shown in Figure 5-88.

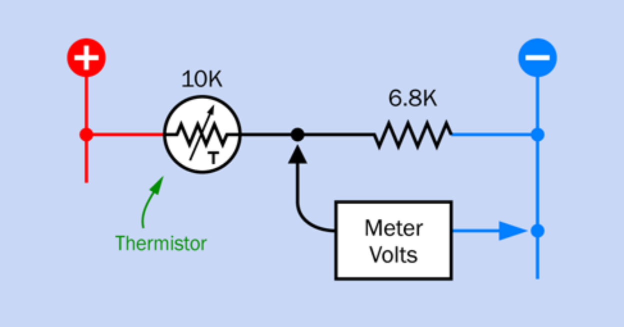

Figure 5-88 *The simplest circuit for deriving a voltage from the changing resistance of a thermistor.*

Normally, to set up this circuit, you would need to install a voltage regulator to provide the 5VDC. However, the Arduino has its own voltage regulator, and conveniently provides a 5VDC output (see Figure 5-79). You can tap into this output and take it across to your breadboard with a jumper wire. You'll also need to tap one of the ground outputs from the Arduino, and take that to your breadboard as well.

When I tried that, and varied the thermistor temperature from around 25 to 37 Celsius, my meter measured a voltage from 2.1V to 2.5V. You should try that yourself, to check my numbers.

Obviously we run no risk of endangering the microcontroller with those voltages. But now I see a different problem: the range isn't wide enough to achieve optimal accuracy.

Figure 5-89 shows the conversion between input voltage and internal digital equivalents. The range from 2.1V to 2.5V is defined by the darker blue rectangle. It will be converted into digital values from around 430 to 512, which is a spread of 82—only a small fraction of the complete range from 0 through 1,023.

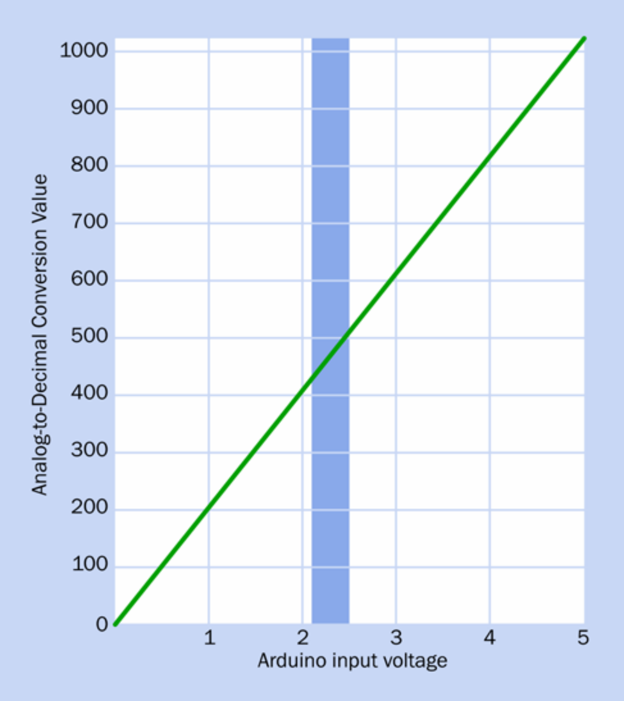

Figure 5-89 *Conversion graph from Arduino input voltage to ADC values. The blue rectangle identifies approximate voltages derived from a 10K thermistor in series with a 6.8K resistor, for temperatures between 75 and 95 degrees Fahrenheit.*

Using this limited range will be like using a small number of pixels from a high-resolution photograph. Inevitably, there will be a lack of detail. Wouldn't it be nice if we could somehow convert our voltages into a digital range covering maybe 500 values instead of 82 values?

One way to do that would be by amplifying the voltage, but that will entail using an additional component, such as an op-amp. This is possible, but then we will need resistors to control feedback, and everything will get complicated. The whole idea of the microcontroller was to keep it simple!

There is another option, which is to use a feature of the Arduino that sets a lower maximum voltage for the range. But this requires me to supply a sample of the new maximum voltage to one of the pins. To create that voltage, I would need to use another voltage divider, and then calculate a new conversion from voltage inputs to ADC values. Really, I want to avoid that kind of thing, at least until I have a simple program running.

After thinking about this a bit more—maybe I can live with a range of 82 values to represent a range from approximately 75 to 95 degrees Fahrenheit. That's an accuracy of about 1/4 degree for each digital step created by the ADC. It's not good enough for a clinical thermometer, but perfectly adequate for room temperature.

Connections

So let's give it a try. But wait—are we still going to use a separate breadboard with that funky-looking Uno board on which the microcontroller is mounted?

Yes, that's the plan. There are three ways to hook everything up:

- You can buy a gadget called a *protoshield*, which is like a mini-breadboard that sits on top of the Uno board and plugs into the connectors. I am not a fan of this device, because it leads us away from an eventual finished circuit on a regular breadboard.
- You can pry the microcontroller off the Uno board and plug it into a breadboard, where you can connect components with its pins in a normal manner. But if you do that, you have no way to load a program into the microcontroller, and you will need a crystal oscillator to make the microcontroller run at the same speed as when it was on the Arduino board.
- You can mount your thermistor and resistor on a regular breadboard, and then run the voltage across from the thermistor circuit to the Uno board on a jumper wire, in the same way that you are already supplying positive voltage and ground from the Arduino to the breadboard. It's messy, but it is what most people seem to do. If you finish a program and mount it permanently on the microcontroller, then maybe you can move the chip to a more convenient location.

Figure 5-90 shows the arrangement. Figure 5-91 shows how it looks in a photograph. I have to admit, this is one occasion where those little wires with plugs at each end are convenient, although I still don't entirely trust them.

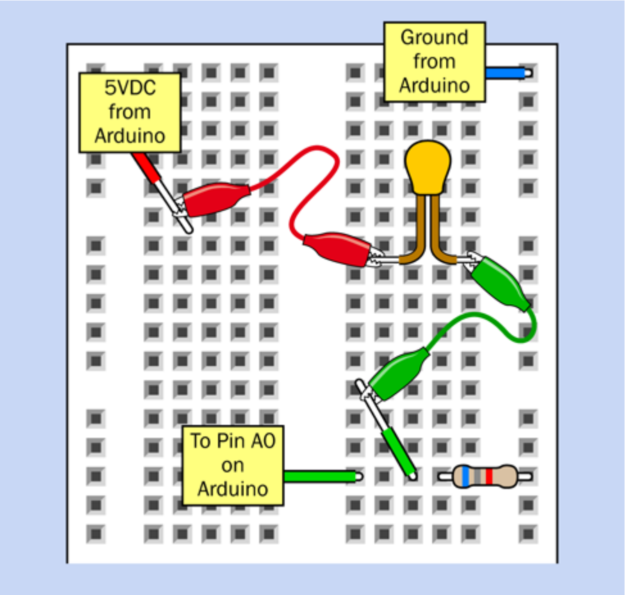

Figure 5-90 *Wiring the thermistor circuit to the Arduino.*

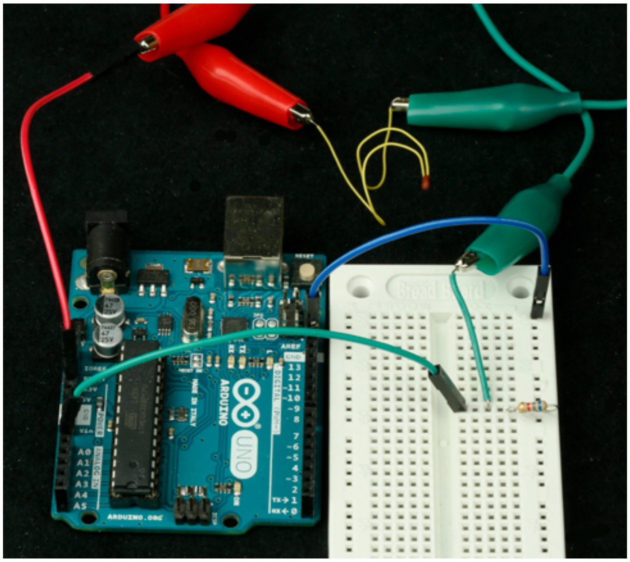

Figure 5-91 *Using jumper wires to connect the thermistor circuit with the Arduino.*

What, No Output?

Now you're all set to convert an analog input to an internal digital value. But wait—something is missing, here. You don't have an output!

In an ideal world, the Uno board would be sold with a nice little alphanumeric display, so that you could use it like a real computer. In fact, you can obtain displays like that which will work with the Uno, but once again this would entail some complexity. Like almost everything in the microcontroller world, it is not plug-and-play. The microcontroller has to be programmed to send text to the display.

So, I will keep things simple. I will use the little yellow LED on the Uno board as an indicator. I will pretend that the indicator represents a room heater, which comes on if the temperature is low and switches off when the temperature is high.

Hysteresis

Suppose we're heating a greenhouse where we want the temperature to be a toasty 85 degrees Fahrenheit. Suppose the voltage from the resistor-thermistor combination is 2.3V at that temperature. Look it up on the graph in Figure 5-89 and you'll see that the ADC inside the microcontroller will convert this to a digital value of around 470.

So, 470 is our threshold. If the number goes down to 469, we put on the heat (or simulate it, by lighting the LED). If the number goes up to 471, we turn off the heat.

But wait. Does this make sense? It means a fractional increase in temperature sensed by the thermistor will trigger the LED, and a fractional decrease will switch it off. The system will be forever fluctuating.

A real thermostat doesn't respond to every little temperature variation when someone opens or closes a door. When it switches on, it stays on, until the temperature is a bit above the target. Then when the heat is off, it stays off, until the temperature is a bit below the target.

This behavior is known as *hysteresis*, and I discuss it in more detail, in relation to a component called a *comparator*, in my sequel to this book, *Make: More Electronics*.

How can we implement hysteresis in a microcontroller program? We need a broader range of values than 469 through 471. The program could say something like, "If the LED is on, keep it on until the temperature value exceeds 490. Then switch if off." And, "If the LED is off, keep it off until the temperature value drops below 460. Then switch it on."

Can we do this? Yes, quite easily. The program listed in Figure 5-92 uses this logic. I derived it from a screen capture of the Arduino IDE, so I have good reason to think that it works.

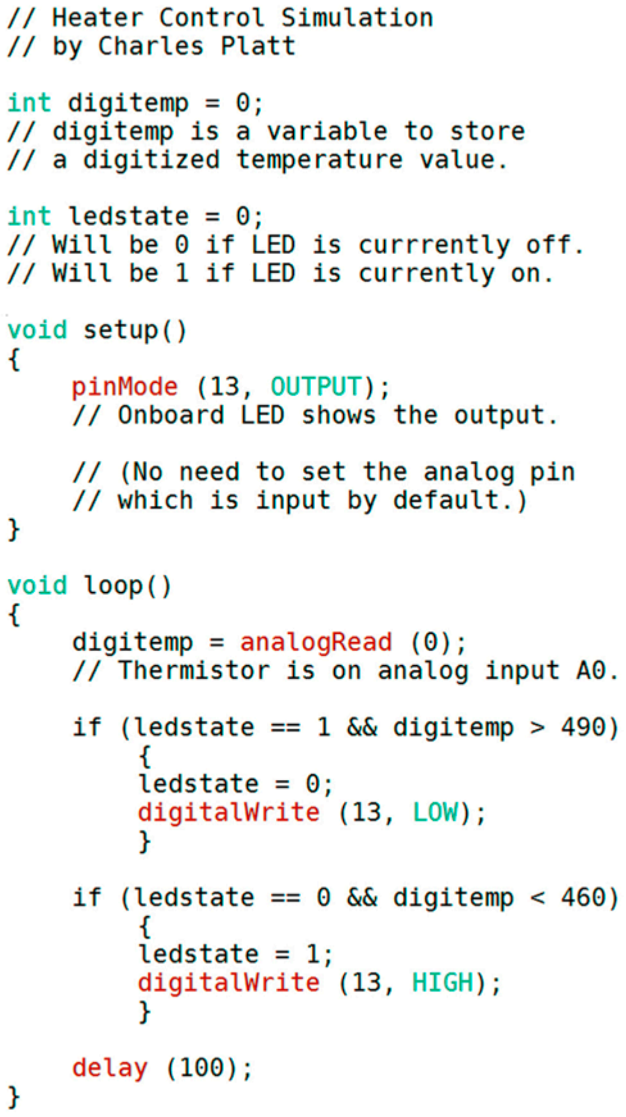

```
// Heater Control Simulation
// by Charles Platt

int digitemp = 0;
// digitemp is a variable to store
// a digitized temperature value.

int ledstate = 0;
// Will be 0 if LED is currrently off.
// Will be 1 if LED is currently on.

void setup()
{
     pinMode (13, OUTPUT);
     // Onboard LED shows the output.

     // (No need to set the analog pin
     // which is input by default.)
}

void loop()
{
     digitemp = analogRead (0);
     // Thermistor is on analog input A0.

     if (ledstate == 1 && digitemp > 490)
          {
          ledstate = 0;
          digitalWrite (13, LOW);
          }

     if (ledstate == 0 && digitemp < 460)
          {
          ledstate = 1;
          digitalWrite (13, HIGH);
          }

     delay (100);
}
```

Figure 5-92 *Program listing for control of an imaginary heating device.*

The program also introduces some new concepts—but first, type it into the IDE. You don't need to include all the comment lines, which I just added as explanations. You can type the much shorter version in Figure 5-93 where the comment lines have been omitted.

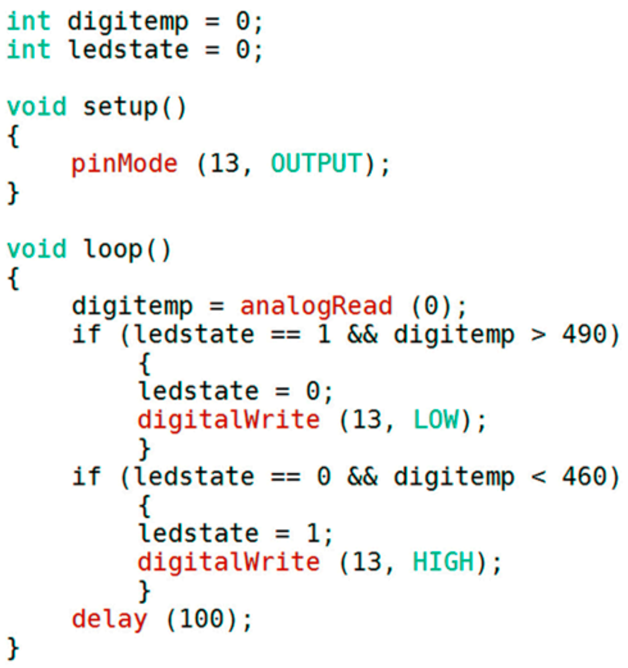

```
int digitemp = 0;
int ledstate = 0;

void setup()
{
     pinMode (13, OUTPUT);
}

void loop()
{
     digitemp = analogRead (0);
     if (ledstate == 1 && digitemp > 490)
          {
          ledstate = 0;
          digitalWrite (13, LOW);
          }
     if (ledstate == 0 && digitemp < 460)
          {
          ledstate = 1;
          digitalWrite (13, HIGH);
          }
     delay (100);
}
```

Figure 5-93 *The same program as before, with the comment lines omitted.*

Verify/Compile your program, and fix any typing errors (probably there will be a semicolon missing in one place or another—that's the most common error).

Plug in your Arduino, upload the program, and if the temperature of your thermistor is below 85 degrees Fahrenheit, the yellow LED should switch on.

Hold the thermistor between your finger and thumb to make the thermistor think that the room temperature has increased. After a few seconds, the LED goes out. Now let go of the thermistor, and it cools down—but the LED stays on for a while, because the hysteresis in the system is telling it to wait until the temperature is sufficiently low. Eventually, the LED switches on again. Success!

But how does the program work?

Line by Line

This program introduces the concept of a *variable*. This is a little space in the microcontroller's memory where a digital value can be stored. You can think of it as a "memory box." On the outside of the box is a label that has the name of the variable. On the inside is a numeric value.

`int digitemp = 0;` means I have invented a variable named `digitemp`. It is an *integer* (a whole number), and it starts with a value of zero.

`int ledstate = 0;` means I have invented another integer variable, to keep track of whether the LED on the board is on or off. There's no easy way I can get the microcontroller to look at the LED and tell me what its state is, so I will have to keep track of it myself.

The `setup` of the program just tells the microcontroller to use pin 13 as an output. I don't need to tell it to use pin A0 as an input, because analog pins are inputs by default.

Now comes the heart of the program, in the loop. First, I use the command `analogRead` to tell the microcontroller to read the state of an analog port. Which one? I specify `0`, which means analog port A0. This is where the connection from my breadboard is plugged in.

What do I do with the information from the ADC after it reads the port? There is only one sensible place to put it: in the variable named `digitemp` that I created for this purpose.

Now that `digitemp` contains a value, I can examine it. First I want to say, if the heat is on (the LED is illuminated) and the temperature value is greater than 490, it's time to turn the heat off. The "if" part is tested like this:

```
if (ledstate == 1 && digitemp > 490)
```

The double `==` means, "make a comparison and see if these two items are the same." A single equals sign means, "assign this value to a variable," which is a different operation.

The double `&&` is a "logical AND." Yes, we have Boolean logic here, just as if we were using an AND logic gate. But instead of wiring up a chip, we just type a line of code.

The > symbol means, "is greater than."

The "if" test is contained in parentheses. If the statement in the parentheses is true, the microcontroller performs the following procedure, between the braces. In that procedure, `ledstate = 0` records the fact that the LED is going to be switched off. `digitalWrite (13, LOW);` actually switches it off.

The second "if" test is very similar, except it only applies if the LED is off, and the temperature has fallen much lower. Then, we switch the LED on.

Finally there is a delay of 1/10th of a second because we don't need to check the temperature more often than that.

And that's it.

Additional Details

I have thrown in some pieces of syntax, such as the "if" test, and the double equals sign, and the && logical operator, without giving you a whole list of all the syntax that you can use in C language. You can find that kind of list online. I don't have space for it here.

Notice a couple of things about the program:

- Lines are indented to clarify the logical structure. The compiler ignores extra white space, so you can feel free to add as much of it as you like.
- The IDE uses color to help you see if you made any typing errors.
- When you make up a name for a variable, you can use any combination of letters, numerals, and underscore characters—so long as the combination is not the same as a word that already has a special meaning in the C language. For instance, you can't have a variable named `void`.
- Some people like to begin variable names with a capital letter, and some don't. It's your choice.
- Each variable should be declared at the beginning of the program, so that the compiler knows what to expect.
- An integer (declared with the `int` term) can have a value ranging from -32,768 to +32,767. The C language on this microcontroller allows you other types of variables that have a wider range of values, or that can have fractional values. But I won't need to use a larger numeric value until Experiment 34.

For a beginning language reference, go to the main Arduino website, click the Learning tab, and choose Reference from the menu that drops down. You can also pull down the Help menu from the Arduino IDE and choose the Reference submenu.

Enhancements

This program does the job that I set out to do, but it's very limited. The biggest limitation is that it uses specific numbers for the minimum and maximum temperature values. That's like having a thermostat that is glued into just one position and can never be adjusted. How could the program be enhanced so that the user can adjust the threshold temperature for the heat to go on and off?

I think the way to do it would be to add a potentiometer. The ends of the potentiometer's track would be connected to 5V and 0V, and the wiper would be connected to another analog input on the microcontroller. This way, the potentiometer functions as a voltage divider and supplies the full range from 0VDC to 5VDC.

Then I would add another procedure in the loop where the microcontroller would check the potentiometer setting, and digitize it.

The result would be a number in the full range from 0 through 1,023. I would have to convert this to a number compatible with the likely range of values for the `digitemp` variable. Then I would put the result in a new variable named, perhaps, `usertemp`. Then I would see if the actual room temperature, measured by the thermistor, is significantly above or below `usertemp`.

Notice that I skipped over one little detail: exactly how I would convert the potentiometer input to the range suitable for `usertemp`. All right, I'll deal with that now.

If the likely range of values for the thermistor is 430 through 512 as I estimated previously, that range can be thought of as having a middle value of 471, plus or minus 41. The potentiometer has a middle value of 512, plus or minus 512 for its full range. Therefore:

```
usertemp = 471 + ( (potentiometer - 512) * .08)
```

where `potentiometer` is the value from the potentiometer input, and the asterisk symbol is used by the C language as a multiplication sign. That would be close enough.

Yes, arithmetic does tend to be involved in programming, sooner or later, somewhere or other. There is no way to pretend otherwise. But it seldom goes beyond high-school math.

In the enhanced version of the program, I would still have to take care of hysteresis. The first "if" statement would have to be converted so that it says something like,

```
if (ledstate == 1 && digitemp > (usertemp + 10) )
```

then switch off the LED. But

```
if (ledstate == 0 && digitemp < (usertemp - 10) )
```

then switch on the LED. That would give me a hysteresis range of plus-or-minus 10, using the values from the ADC.

Now that I've described this modification, maybe you can consider making it yourself. Just remember to declare each new variable before you use it in the body of the program.

Experiment 34: Nicer Dice

In this last experiment, I'm going to revisit Experiment 24, which used logic chips to create dice patterns. Instead of logic chips, we can now use "if" statements with logical operators in a microcontroller program. We can replace several pieces of hardware with a dozen lines of computer code, and instead of a 555 timer, a counter, and three logic chips, we will need just one microcontroller. This is a great example of an appropriate application. (Of course, it does still require some LEDs and series resistors.)

What You Will Need:

- Breadboard, hookup wire, wire cutters, wire strippers, test leads, multimeter.
- Generic LED (7).
- Series resistor, 330 ohms (7).
- Arduino Uno board (1).
- Laptop or desktop computer with an available USB port (1).
- USB cable with A-type and B-type connector at opposite ends (1).

The Limits to Learning by Discovery

Learning by discovery works well when you're getting to know an electronic component. You can put it on a breadboard, apply power, and see what happens. Even when you're designing a circuit, you can use some trial and error and make modifications as you go along.

Writing programs is different. You have to be disciplined and logical; otherwise, you will tend to write buggy program code that doesn't work reliably. Also, you have to plan ahead; otherwise, you will waste a lot of time redoing previous work or throwing it away entirely.

I don't enjoy planning, but I dislike wasting time even more. So, I do plan, and in this final project, I'm going to describe the planning process. I'm sorry that you won't get some immediate pleasure from just putting some parts together and seeing what happens, but if I don't explain the software development process, I will be creating a misleading impression by making programming seem simpler than it really is.

Randomicity

The first question seems obvious: "What do I really want this program to do?" The question is necessary because if you don't define your goals clearly, there is no way the microcontroller can figure them out for you. This is similar to the process that I described in Experiment 15, of making a "wish list" for the intrusion alarm; but for a microcontroller, greater detail is required.

The basic requirement is very simple. I want a program that will choose a random number and display it in some LEDs that resemble the spots on a die (the word "die" being the singular of "dice," as I mentioned once before).

Because choosing a random number is fundamental to this program, you need to be properly informed on the subject. So let's check the Arduino website, where there is a language reference section. It is not as comprehensive as I would like, but is a good starting point.

To find it, go to the Arduino home page, click the Learning tab, and select Reference, and you will find a section titled Random Numbers. When you click that, you find there is a function specifically created for the Arduino, named `random()`.

This is not entirely surprising, because almost all high-level computer languages have some kind of random function built in, and it always works by using mathematical trickery to generate a stream of numbers that are unpredictable to a human observer and continue for a very long time before the sequence repeats. The only problem is that because they are created mathematically, the sequence will start from the same place each time you run the program.

What if you want the sequence to start at a different place? There is another Arduino function named `randomSeed()` that initializes the number generator by looking at the state of a pin on the microcontroller that is not connected with anything. As I have mentioned before, a floating logic pin picks up any electromagnetic radiation that may be around, and you never know what to expect from it. So `randomSeed()` can be genuinely random, and sounds like a good idea—although we would have to remember not to use that unconnected, floating pin for anything else.

Setting aside the issue of seeding the random number generator for the time being, let's suppose I use the Arduino's `random()` function to choose a number for the output of the dice program. How should this actually work?

I think the player will press a button, at which point the randomly selected pattern of dice spots can be displayed. Job done! Then if you need to "throw the die" a second time, you just press the button again, and another randomly selected pattern of dice spots will be created.

That sounds very convenient—but I don't think it will look very interesting. It may not seem very plausible, either. People may wonder if that number is really random. I think the problem is that control of the procedure has been taken away from the user.

Going back to the hardware version of this project, I liked the way it displayed a rapid blur of patterns, and I liked that the player could press a button to stop the sequence arbitrarily.

Maybe the program should emulate this, instead of using the `random()` function. It can count through numbers 1 through 6 over and over again, very fast—like the counter chip in the hardware version of Nice Dice.

But this worries me in a different way. When the program counts from 1 to 6, and then repeats, I think the microcontroller will take an extra few microseconds to go back to the start of the loop. So, a 6 will always be displayed for a fraction longer than the other numbers.

Maybe I can combine the two concepts. I can use the random number generator to create a series of numbers, and I will display them in very quick succession, until the player presses a button to stop at an arbitrary moment.

I like that plan. But then what? I could have another button to restart the rapid number display. But, no, that's not necessary: the same button can do it. Press to stop, press to restart.

You see, now I am getting a clearer idea of what I want this program to do. This will help me to take the next step, figuring out the instructions to get the microcontroller to do it.

Pseudocode

I like to write *pseudocode,* which is a series of statements in English that will be easily converted into computer language. Here is my pseudocode plan for the program that I am naming Nicer Dice. Bear in mind that these instructions will be executed extremely fast, so that the numbers are a blur.

Main loop:

- Step 1. Choose a random number.
- Step 2. Convert it to a pattern of dice spots, and light up the appropriate LEDs.
- Step 3. Check to see if a button has been pressed.
- Step 4. If a button has not been pressed, go back to Step 1 and choose another random number, so that the sequence will repeat rapidly. Otherwise…
- Step 5. Freeze the display.
- Step 6. Wait for the player to press the button a second time. Then go back to Step 1 and repeat.

Do you see any problems with this sequence? Try to visualize it from the microcontroller's point of view. If you were receiving the instructions in the program, would you have everything you need to get the job done?

No, you wouldn't, because some instructions are missing. Step 2 says, "light up the appropriate LEDs," but—there is no instruction anywhere to switch them off!

You always have to remember:

- The computer *only* does what you tell it to do.

If you want the illuminated LEDs to be switched off before a new number is displayed, you have to include an instruction to do that.

Where should I put it? Well, I need to blank the display immediately before each new number is chosen and displayed. Therefore, the right place to zero the display is at the beginning of the main loop. I'll include it as Step 0.

- Step 0. Switch off all the LEDs.

But wait. Depending on the number that was displayed in the previous cycle, some of the LEDs in the pattern will be on, and some will be off. If we switch off all the LEDs to clear the display, this will include some of the LEDs that were off already. The microcontroller won't care, but it will waste some time executing that instruction. Maybe it would be more efficient just to switch off the LEDs that were previously on, and ignore the ones that are already off.

That will entail more complexity in the programming, and it may not be necessary. In the very early days of computing, people had to *optimize* a program to save processor cycles, but I think even a microcontroller is fast enough, now, for us not to care about the time that will be wasted by switching off two or three LEDs that were off already. I'll use an all-purpose routine to switch off all the LEDs, regardless of their current states.

Button Inputs

What else is missing from the list of pseudocode instructions?

There is a button issue.

Once again, I need to visualize what I want the program to do. The rapid display is cycling through numbers very quickly. The player presses a button to stop it. The display freezes, showing the current number. In Step 6 the microcontroller waits indefinitely until the player presses the button again, to create a new rapid display.

Wait a minute. How can the player press the button "again" without releasing it first?

With the pseudocode in its current form, this is what the microcontroller will actually do, bearing in mind that it performs tasks very, very quickly:

- Program tells the microcontroller to check the button.
- Microcontroller finds that the button is pressed.
- The display freezes. Microcontroller waits for the button to be pressed again.
- But it finds that the button is still being pressed, because the player hasn't had time to stop pressing it.
- Microcontroller says, "Oh, the button is pressed, so I should resume the rapid display."

Consequently the frozen display will only last for an instant.

Here's the solution to the problem. An extra step in the sequence:

- Step 5A. Wait for the player to release the button.

This will stop the computer from running ahead and displaying more numbers, until the player is ready.

Is this it? Are we done, yet?

No, I'm afraid not. You may think this is getting a bit laborious, but in that case I have to say, I'm sorry, but this is what programming is like. If someone tells you that you can just throw some instructions together and see them work, I'm afraid this is often not the case.

One more button problem still exists. Step 6 says, wait for the button to be pressed again to restart the rapid display. OK, the player presses the button, the display resumes—but the microcontroller is so fast, it will zip through the process of zeroing the current display and displaying a new die pattern, and once again it will recheck the button before the player has time to stop pressing it. Consequently, when the microcontroller gets to Step 4, it will find that the button is still being pressed, so it will freeze the display again.

Now what? Maybe I should add a new Step 7, telling the microcontroller to wait for the button to be released before resuming the rapid display.

That's counter-intuitive. I don't think anyone will understand that you have to press the button, *and release it*, before the rapid display resumes. We could just say, "Oh, you have to do that, because the program requires it." But that is a *very bad way of thinking*.

- A program has to do what the user expects. We should never force the user to do things to satisfy the program.

In any case, the idea of waiting for the button to be released before the rapid display resumes still won't work. There's another issue: contact bounce. This happens when the button is pressed, *and* when the button is released. Consequently, if someone lets go of a button and the process resumes, and the program checks the button again a millisecond later, the contacts may still be vibrating, and they may appear to be open or closed, on an unpredictable basis.

This is the kind of thing we run into when a microcontroller interacts with the physical world. The microcontroller wants everything to be precise and stable, but the physical world is imprecise and unstable.

I had to think carefully about this particular problem before I decided how I wanted to solve it.

One solution would be to go back to having two buttons, one to start the rapid display, and one to stop it. This way, as soon as the "start" button is pressed, the microcontroller can ignore its status and its contact bounce, and just wait for the "stop" button to be pressed.

But from the player's point of view, I like the simplicity of just having one button. Surely, there's a way to make it work?

I went back to describing what I wanted the program to do, as clearly as possible. I said to myself: "I want the program to resume showing the rapid sequence as soon as the button is pressed for a second time. But after that, the program should ignore the button until after it is released, *and* until the contacts have stopped bouncing."

Why not simply lock out the button for a second or two? Actually this will be a good idea, because the random display should run for a little while before the player can stop it again. This will make the display look "more random," as it runs through all those numbers.

Suppose I lock out the button for, say, two seconds after the rapid display begins. Step 4 should be rewritten like this:

- Step 4. If a button has not been pressed, OR the rapid display has been running for less than two seconds, go back to the top and choose another random number. Otherwise…

Notice the OR word. Those Boolean operators really come in handy.

The System Clock

I think I have resolved all the button problems, but now I have a new problem. I need to measure two seconds.

Is there a system clock inside the microcontroller? Probably there is. Maybe the C language can access it, and ask it to measure a time interval.

Check the language reference. Yes, there is a function named `millis()` that measures milliseconds. It runs like a clock, starting from zero each time a program begins. This function is capable of reaching such a high number, it doesn't reach its limit and restart from zero until about 50 days have passed. That should certainly be enough.

But—no, there is one more little snag. The Arduino doesn't allow my program to reset the system clock on demand. The clock starts counting like a stopwatch when the program starts running, but unlike a stopwatch, I can't stop it.

How can I deal with this? I will have to use the system clock in the same way that I would use the clock on my kitchen wall, in the real world. When I want to cook a hard-boiled egg, I make a mental note of the time when the water starts boiling. Let's suppose it is 5:02pm and I want to boil the egg for seven minutes. I say to myself, "5:02 plus seven minutes is 5:09, so I'll take the egg out at 5:09."

What I am doing, inside my head, is comparing the clock, which keeps running, with my memorized time limit of 5:09. I am saying to myself, "Has the clock reached 5:09 yet?" When the time on the clock is equal to 5:09, or greater, my egg is cooked.

The way to do this in the dice program is to invent a variable that will work like my memory of the time at the start of the egg-boiling process. Immediately before the rapid display begins, I store the current value of the system clock in the variable, plus two seconds. Then I can get the program to say, "Has the system clock reached the value of my variable yet?" until it finally gets there.

Suppose I name the variable "ignore," because it will tell me at what time in the future the program should stop ignoring the button. Then Step 4 can ask the microcontroller, "Has the system clock exceeded the *ignore* variable yet?" and if it has, the program can resume paying attention to the button.

I cannot reset the system clock, but I can reset the "ignore" variable to match the current value of `millis()`, plus two seconds, each time a new rapid display session begins.

Final Draft of Pseudocode

Bearing all these issues in mind, here is the revised and, I hope, final sequence of events for the program:

- Before the loop begins, establish input and output of the logic pins, and set the "ignore" variable to the current time, plus two seconds.
- Step 0. Switch off all the LEDs.
- Step 1. Choose a random number.
- Step 2. Convert it to a pattern of dice spots, and light up the appropriate LEDs.
- Step 3. Check to see if the button has been pressed.
- Step 4. Check to see if the system clock has caught up with the "ignore" variable yet.
- Step 4a. If the button has not been pressed, OR the system clock has not caught up with the "ignore" variable yet, go back to Step 0. Otherwise…
- Step 5. Freeze the display.
- Step 5A. Wait for the player to release the button.
- Step 6. Wait as long as it takes for the player to press the button once more, to restart the display.
- Step 7. Reset the "ignore" variable to the system clock, plus two seconds.
- Go back to Step 0.

Do you think it will work? Let's find out.

Hardware Setup

Figure 5-94 shows seven LEDs wired on a breadboard to display the spots on a die. The concept is the same as in Figure 4-146, except that the Arduino can deliver 40mA from each output pin, and therefore I don't have to drive pairs of LEDs in series. A single output pin can easily drive a pair of LEDs in parallel, and for each generic LED, a 330-ohm resistor is ample.

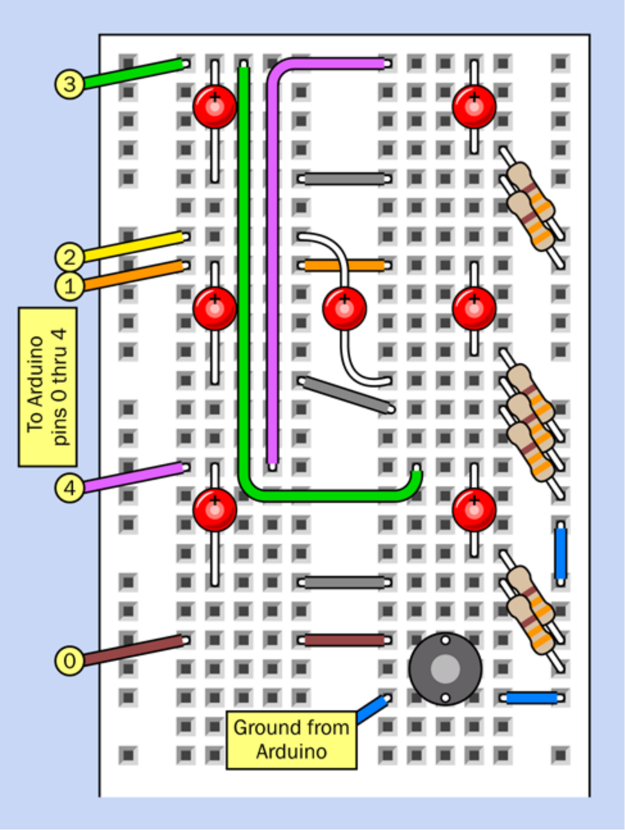

Figure 5-94 *Seven LEDs wired on a breadboard to display the spot patterns of a die.*

The numbers on the wires are the same as the numbering system in Figure 4-146. They have nothing to do with the die values. It's just an arbitrary way to identify each wire. Also, I can plug the wires numbered 1 through 4 into digital outputs numbered 1 through 4 on the Uno board. This will help to make everything clear.

I'll use digital connection 0, on the Uno board, as an input that checks the status of the pushbutton. However, note that the Uno uses digital pins 0 and 1 when receiving USB data. If you have any problem uploading your program, disconnect the wire from digital input 0 temporarily.

Don't connect the ground wire from the breadboard to the Uno board just yet. It's safer to upload the program first, because the program will tell the microcontroller which pins are outputs and which pins are inputs. A previous program may have configured them differently, and so long as the Arduino board is plugged in, it will want to run whatever program is in its memory. This may not be safe for the Arduino outputs, because:

- You must be very careful not to apply any voltage to a digital pin that is configured as an output.

Now, the Program

Figure 5-95 shows the program that I wrote to match the pseudocode. The same program is listed in Figure 5-96 with the comments removed, so that you'll be able to copy-type it more quickly. Please type it in the IDE editing window.

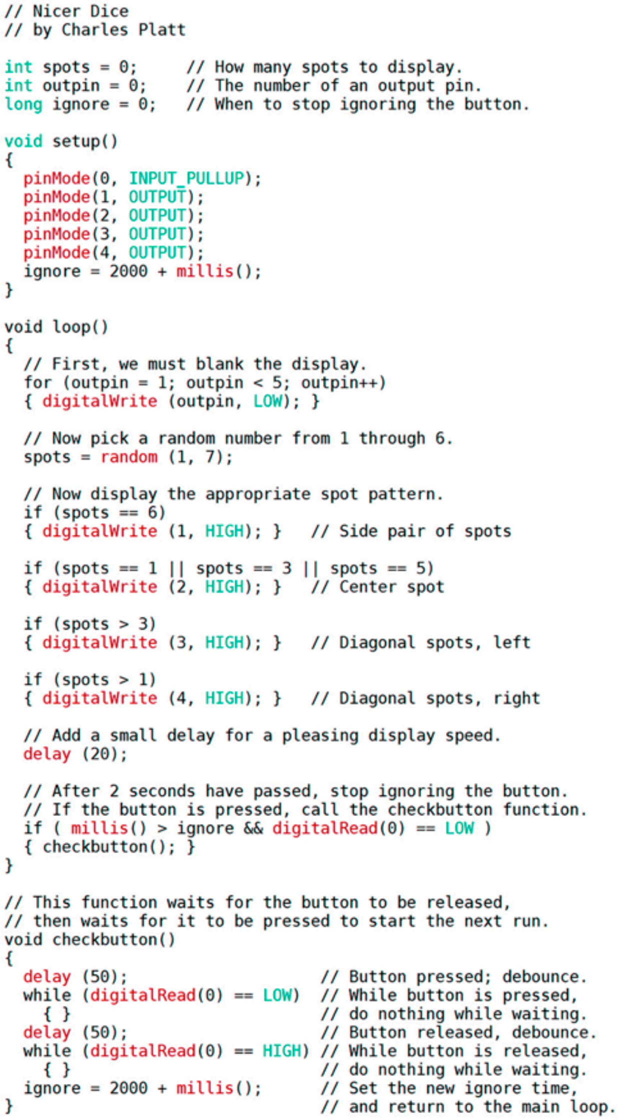

```
// Nicer Dice
// by Charles Platt

int spots = 0;     // How many spots to display.
int outpin = 0;    // The number of an output pin.
long ignore = 0;   // When to stop ignoring the button.

void setup()
{
  pinMode(0, INPUT_PULLUP);
  pinMode(1, OUTPUT);
  pinMode(2, OUTPUT);
  pinMode(3, OUTPUT);
  pinMode(4, OUTPUT);
  ignore = 2000 + millis();
}

void loop()
{
  // First, we must blank the display.
  for (outpin = 1; outpin < 5; outpin++)
  { digitalWrite (outpin, LOW); }

  // Now pick a random number from 1 through 6.
  spots = random (1, 7);

  // Now display the appropriate spot pattern.
  if (spots == 6)
  { digitalWrite (1, HIGH); }   // Side pair of spots

  if (spots == 1 || spots == 3 || spots == 5)
  { digitalWrite (2, HIGH); }   // Center spot

  if (spots > 3)
  { digitalWrite (3, HIGH); }   // Diagonal spots, left

  if (spots > 1)
  { digitalWrite (4, HIGH); }   // Diagonal spots, right

  // Add a small delay for a pleasing display speed.
  delay (20);

  // After 2 seconds have passed, stop ignoring the button.
  // If the button is pressed, call the checkbutton function.
  if ( millis() > ignore && digitalRead(0) == LOW )
  { checkbutton(); }
}

// This function waits for the button to be released,
// then waits for it to be pressed to start the next run.
void checkbutton()
{
  delay (50);                    // Button pressed; debounce.
  while (digitalRead(0) == LOW)  // While button is pressed,
    { }                          // do nothing while waiting.
  delay (50);                    // Button released, debounce.
  while (digitalRead(0) == HIGH) // While button is released,
    { }                          // do nothing while waiting.
  ignore = 2000 + millis();      // Set the new ignore time,
}                                // and return to the main loop.
```

Figure 5-95 *The Nicer Dice program listing.*

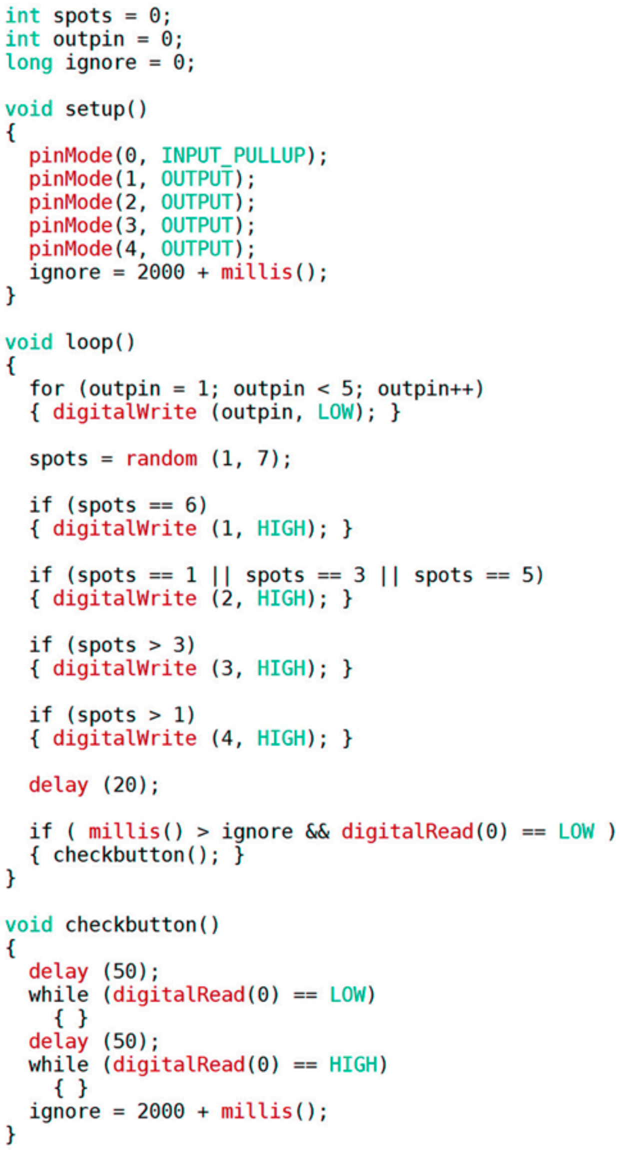

```
int spots = 0;
int outpin = 0;
long ignore = 0;

void setup()
{
  pinMode(0, INPUT_PULLUP);
  pinMode(1, OUTPUT);
  pinMode(2, OUTPUT);
  pinMode(3, OUTPUT);
  pinMode(4, OUTPUT);
  ignore = 2000 + millis();
}

void loop()
{
  for (outpin = 1; outpin < 5; outpin++)
  { digitalWrite (outpin, LOW); }

  spots = random (1, 7);

  if (spots == 6)
  { digitalWrite (1, HIGH); }

  if (spots == 1 || spots == 3 || spots == 5)
  { digitalWrite (2, HIGH); }

  if (spots > 3)
  { digitalWrite (3, HIGH); }

  if (spots > 1)
  { digitalWrite (4, HIGH); }

  delay (20);

  if ( millis() > ignore && digitalRead(0) == LOW )
  { checkbutton(); }
}

void checkbutton()
{
  delay (50);
  while (digitalRead(0) == LOW)
    { }
  delay (50);
  while (digitalRead(0) == HIGH)
    { }
  ignore = 2000 + millis();
}
```

Figure 5-96 *The same program listing with comments removed.*

While you are typing it, you will find that the second "if" statement contains a character that you haven't seen before. In fact, you may have never typed it, ever, on your keyboard. It is a vertical line, sometimes called the "pipe" symbol. On a Windows keyboard, you are likely to find it above the Enter key. You create it by pressing Shift and the backslash \ character. The listing uses two pairs of pipe symbols in the second "if" statement, and I will explain them when I go through the program line by line.

When you finish, choose the Sketch > Verify/Compile option in the IDE to see if you made any errors.

Some of the error messages can be difficult to understand, and they refer to line numbers. But line numbers are not displayed with the program! This seems like a cruel joke, telling you which line your error is on, but not showing you the number. Maybe there is a way to switch on line numbering? If you check the Help, and search for "line numbers," I don't think you'll find anything. Check the Arduino forums, and you will find a lot of people complaining that they can't display line numbers.

Ah, but the forums show the oldest posts first. If you scroll all the way down to the newest posts, you find that the problem finally was solved. Arduino just didn't get around to documenting it. Go to File > Preferences, and you will see an option to switch on the line numbers.

Of course the error messages can be a bit difficult to understand, but here are the most common causes of errors, which you should check before you try to fix anything.

- You may have forgotten to put a semicolon at the end of an instruction.
- You may have left out a closing brace. Remember, braces { must always } be in pairs.
- Although a command word often contains a mix of uppercase and lowercase letters, as in `pinMode`, you may have typed it in all lowercase letters. The IDE should display the command words in red, when they are correctly typed. If you see one that is in black, it contains a typing error.
- You may have omitted parentheses from a function name such as `void loop()`.
- You may have used a single = sign where you should have two == signs. Remember, = means "assign a value" while == means "compare a value."
- You may have typed only one | symbol or only one & symbol where they should be used in pairs.

When your Verify/Compile operation finds no additional error messages, upload the program. Now plug in the ground wire connecting your breadboard with the Uno board, and the LEDs should start flashing. Wait a couple of seconds, then press the button, and the display stops, displaying a random spot pattern. Press the button again, and the display resumes. Hold down the button, and after the two-second "ignore" period, the display stops again.

The pseudocode has been successfully implemented!

Now, how does the program work?

Short and Long Integers

The program includes some words that you haven't seen before, and one really important new concept.

One new word is `long`. Until now, you have typed `int` (meaning "integer") before each variable name. But the value of `int` is limited between -32,768 to +32,767. When you need to store a bigger number than that, a long integer can be used, allowing a range of values from -2,147,483,648 to 2,147,483,647.

Why not use long integers for everything? Then we don't have to worry about the limit for a regular integer. True, but a long integer takes twice as long to process (or more), and takes twice as much memory. We don't have a whole lot of memory on the Atmel microcontroller.

The system clock uses the `millis()` function to count milliseconds. If we only allowed it to count up to 32,767, that would be enough for just over half a minute. We may need more than that, so the function stores its value in a long integer. (How do I know this? I read it in the language reference. You have to read the documentation to use a computer language.)

When I created my "ignore" variable to memorize the current value of the system clock, the variable had to be defined so that it is compatible with the clock; so, it has to be defined as a long integer, by using the word `long`.

What happens if you try to store a number outside of the permitted range in an integer (or long integer) variable? Your program will produce unexpected results. It's up to you to make sure that this never happens.

Setup

The setup section of the program is fairly straightforward. You haven't used those `pinMode()` instructions before, but they are easy to understand.

The first one has a parameter, `INPUT_PULLUP`, which is very useful: it activates a pullup resistor built into the microcontroller, so you don't have to add a pullup resistor yourself. But, bear in mind, it is a *pullup* resistor, not a *pulldown* resistor. Therefore, the input state of the pin is normally high, and when you use a button, it must ground the pin of the chip, to make it low. Remember:

- When the button is pressed, the `digitalRead()` function returns a `LOW` value.
- When the button is released, the `digitalRead()` function returns a `HIGH` value.

The "for" Loop

At the beginning of the `void loop()` function, there is a different kind of loop. It is called a "for" loop because it begins with the word `for`. This is a very basic and convenient way to make the microcontroller count through a series of numbers, storing each new number in a variable, and throwing out the previous value. The syntax works like this:

- The reserved word `for` is followed by three parameters in parentheses.
- Each parameter is separated from the next by a semicolon.
- The first parameter is the first value that will be stored in the specified variable. (It is properly known as the initialization code.) In this program, the first value is 1, stored in a variable which I created, named `outpin`.
- The second parameter is the value where the loop stops counting (properly known as the stop condition). Because the loop stops at that point, the final value in the variable will actually be one less than the limit. In this program, the limit is `< 5`, meaning "less than 5." So the loop will count from 1 through 4, using the `outpin` variable.
- The third parameter is the amount that the loop adds to the variable in each cycle (properly known as the iteration expression). In this case, we are counting in ones, and the C language allows me to specify that by using two `++` symbols. So, `outpin++` means, "add 1 to the outpin variable in each cycle."

"For" loops allow you to specify all kinds of conditions. They are extremely flexible. You should read up on them in the language reference. This "for" loop just counts from 1 through 4, but it could count as easily from 100 to 400, or to any range you like, limited by the type of the integer used in the loop (`int` or `long`).

During each cycle of the loop, the microcontroller is told to do something. The procedure it performs is listed in braces after the loop is defined. Like any procedure, it can contain numerous operations, each ending in a semicolon. There is only one operation in this procedure: write a `LOW` state to the pin specified by the variable `outpin`. Because `outpin` is going to count from 1 through 4, the "for" loop is going to create a low output on digital pins 1 through 4.

Ohhhh, *now* you can see what this is all about. The loop is switching off all the LEDs.

Isn't there a simpler way to do this? Sure, you could use these four commands:

```
digitalWrite (1, LOW);
digitalWrite (2, LOW);
digitalWrite (3, LOW);
digitalWrite (4, LOW);
```

But I wanted to introduce you to the concept of a "for" loop, because it's basic and important. Also, what if you wanted to turn off nine LEDs? Or what if you wanted the microcontroller to flash an LED 100 times? A "for" loop is often the best way to make a procedure efficient when repetition is involved.

The Random Function

After the "for" loop has zeroed the die display, we get to the `random()` function, which chooses a number between the limits in the parentheses. We want a die value from 1 through 6, so why is the range listed from 1 through 7? Because actually the function is choosing fractional values from something like 1.00000001 to 6.99999999 and throwing away the numbers which follow the decimal point. So, 7 is a limit that is never actually reached, and the output will be from 1 through 6.

Whatever the random number is, it is stored in another variable that I made up, named `spots`, meaning the number of spots on the face of a die.

The "if" Statement

Now it's time to see what the value of `spots` is, this time around, and light up the appropriate LEDs.

The first "if" statement is simple enough. If we have 6 spots, this will be the only occasion where we write a high value through output pin 1, which is connected with the left and right LEDs.

Why don't we switch on all the diagonal LEDs, too? The answer is, they will be switched on for other die values also, and it's more efficient to minimize the number of "if" tests. You'll soon see how this works.

The next "if" uses the pipe symbol that I mentioned previously. A pair of `||` symbols means OR in the C language. So, the function says that if we have a die value of 1, OR 3, OR 5, we light up the center LED, by putting a high state on pin 2.

The third "if" says that if the `spots` value is greater than 3, we need to light two of the diagonally placed LEDs. These will be required to display the patterns for a 4, 5, or 6.

The last "if" says that if the `spots` value is greater than 1, the other diagonally placed LEDs must be illuminated, too.

You can test the logic of these "if" functions by looking back at the spot patterns in Figure 4-146. The logic gates in that figure were chosen to fit the binary output from the counter chip, so they're different from the logical operations in the "if" functions of the program. Still, the LEDs are paired in the same way.

Flash Speed

After the "if" functions, I inserted a delay of 20 milliseconds, because I think it makes the display more interesting. Without this delay, the LEDs flash so rapidly, they are just a blur. With the delay, you can see them flashing, but they're still too fast for you to stop at a number that you want—although you can try!

You may wish to adjust the delay value to a number higher or lower than 20.

Creating a New Function

Now comes the important part. In the pseudocode that I wrote, we have reached Steps 3, 4, and 4a. To refresh your memory:

- Step 3. Check to see if the button has been pressed.
- Step 4. Check to see if the system clock has caught up with the “ignore” variable yet.
- Step 4a. If the button has not been pressed, OR the system clock has not caught up with the “ignore” variable yet, go back to Step 0. Otherwise…

These steps can be combined in one “if” function. In pseudocode, it would be like this:

- If (the button is not pressed OR the system clock is less than the “ignore” value), go back to step 0.

But there is a problem with this. The term “go back to…” suggests that I want to direct the microcontroller to a specific part of the program. This may seem a natural thing to do, but when you write in C, you should try to avoid transferring control from one section of a program to another.

The reason is that a lot of “go here” and “go there” instructions make the program difficult to understand—not just for other people, but for you, too, when you take another look at it six months from now, and you can’t remember what you had in mind.

The concept of C is that each part of a program is contained in a separate block, and the program runs by *calling* them when you want them. Think of each block of instructions as an obedient servant who only does one thing, such as washing the dishes or taking the garbage out. When you need that task to be executed, you just call the servant by name.

The blocks are properly known as *functions*, which is confusing, because we have been dealing with functions such as `setup()` and `loop()` already. But in fact you can write your own function, which works in basically the same way.

I decided that the correct way to write this program is to split off the button-checking function into—well, a function. I have called it `checkbutton()`, but I could have called it anything at all, so long as the word wasn’t already being used for some other purpose.

You see the `checkbutton()` function at the bottom of the listing, preceded by the word `void`, because this function doesn’t send any value back to the rest of the program.

`void checkbutton()` is the *header* for the function, after which the procedure is contained within braces, as usual. All this function does is:

- Wait 50ms for the contact to stop bouncing.
- Wait for the button to be released.
- Wait another 50ms for the end of the contact bounce created by releasing the button.
- Wait for the button to be pressed again (in other words, wait for the released state to end).
- Reset the `ignore` variable.

When the microcontroller gets to the end of the function, where does it go? Simple: back to the line immediately following the one that called the function. Where is that? Immediately below the “if” function, above. That is how you call a function: you just state its name (including the parentheses, which sometimes have parameters inside them, although not in this case).

You can, and should, create as many functions as you like in a program, using each one to perform a separate task. To learn about this, I suggest you read any general reference on the C language. The Arduino documentation doesn’t go into much detail on functions, because —well, they’re a little difficult to understand, if they start to pass values to and fro. Still, they are at the heart of the C language.

Structure

The line that begins `if ( millis() > ignore` has the same purpose as Step 4 in my pseudocode, except that it now works the other way around. Instead of deciding whether to send the microcontroller back to the beginning, it determines whether to call my `checkbutton()` function. Previously I summarized the logic as, “if (the button is not pressed OR the system clock is less than the “ignore” value), go back to step 0.” The revised version says, “if the button-ignore period is over, AND the button is pressed, make a detour to the checkbutton() function.”

After the microcontroller does that and returns, it reaches the end of the main `loop` function, and the `loop` function always repeats automatically.

Really, this program does only one thing. It selects random numbers and displays them as spot patterns, over and over again. If the button is pressed, it pauses and waits, but when the button is pressed again, the program resumes doing what it was doing before. The button-checking routine is just a momentary interruption.

Therefore, the natural structure for this program is to have a main loop that just selects and displays numbers, and then if the button is pressed, the microcontroller is sent on a brief detour to the `checkbutton()` function and back again.

The Arduino documentation doesn't say anything about structure, because it wants to get you started in making things happen as quickly as possible. So the Arduino simply forces you to use the mandatory `setup` function, followed by the `loop` function, and that's it.

But as soon as a program begins to grow in size, you really need to divide it into your own functions, to keep it from becoming a complicated mess. A standard C-language tutorial will explain this in more detail.

Of course, if you just want to use the Arduino to do one simple thing, such as switch on a heater when a room gets cold, you can put all the procedures in the main `loop` function, and that will be all you need. But this is a waste of the microcontroller's capabilities. It can do so much more. The trouble is, when you try to do something more ambitious—such as simulate the throw of a die—the instructions accumulate. Structuring them helps to keep everything clear.

There is another advantage to dividing a program into functions. You can save the functions separately, and reuse them in other programs later. The `checkbutton()` function could be reused in any game where you want to stop the action by pressing a button and restart it by pressing the button a second time.

Likewise, you can use other people's functions in your own programs, provided the authors don't restrict you from doing so by controlling their copyright. A vast number of functions in the C language are available freely online, many of them specifically written for the Arduino. Functions exist to control almost any alphanumeric display, for instance. This leads to a very important but often ignored recommendation for programmers:

- Don't reinvent the wheel.

You don't need to waste time writing your own function if someone else will let you use theirs.

This is another reason why the concept of functions is so important in C.

But Is It Too Difficult?

The more you write programs, the easier it gets. The learning curve is steep at the beginning, but after some practice you'll write a "for" loop without thinking much about it. Everything will seem obvious.

That's what programmers like to say. Is it true?

Sometimes it is, and sometimes it isn't. In the maker movement we tend to assume that anyone can take control of the techno-world around us. In fact I subscribe to this belief myself—but computer programming pushes this philosophy to the limit.

I used to teach an introductory programming class, and I noticed a very wide range of aptitudes among the students. Some of them found programming a very natural thought process, while others found it extremely difficult, and this didn't always have much to do with intelligence.

At one end of the scale, at the end of my 12-week, 36-hour programming course, one student wrote an entire simulation of a slot machine, including graphics showing the spinning wheels, and money that came tumbling out.

At the other end of the scale, I had a student who was a pharmacist. He was a very smart, well-educated man, but no matter how hard he tried, he couldn't get the syntax right, even in simple "if" statements. "This is really annoying me," he said, "because it makes me feel stupid. And I know I'm not stupid."

He was right, he wasn't stupid, but I came to the conclusion that I couldn't help him, because I had learned a fundamental fact.

- To be good at writing programs, you must be able to think like a computer.

For whatever reason, the pharmacist couldn't do that. His brain just didn't work that way. He could describe to me the pharmacology of a medication, its molecular structure, and a lot more; but that was of no help to him in writing programs.

When the Arduino was marketed, evangelists described it as a device for creative people and others who didn't think of themselves as programmers. Supposedly, it would be so simple, everyone could use it.

The trouble is, I'm old enough to remember when HTML was introduced with the same idea—that it would be so easy, everyone would be coding their own web pages. Well, a few people did, but not "everyone." Today, only a tiny minority will hand-code HTML (I am one of them, but I'm eccentric for doing so).

Going back even farther, to the dawn of computing as we know it, the BASIC computer language was created with the idea that "everyone" could use it. In the 1980s, with the advent of desktop computers, evangelists predicted that people would be writing little programs in BASIC to balance their checkbooks or store recipes. Well, a lot of people gave it a try, but how many people today still do so?

My purpose in emphasizing this is to reassure you that if you are one of the people who finds it difficult, there's no stigma attached to this. I am sure you have other skills that you can pursue instead. In fact, building things with individual components could be one of them, as I think it requires different thought processes. Personally, I find writing programs much easier than designing circuits, but for someone else, the reverse could be equally true.

Upgrading the Nicer Dice Program

As in the version of this program in Experiment 24, the obvious upgrade is to add a second die display. This can be done very easily with the Arduino board, because it has additional digital outputs that can drive a second set of LEDs. You simply need to duplicate the section of the program that begins with zeroing the display and ends with the `delay(20);` function. Substitute the new pin numbers for your additional LEDs in the `digitalWrite()` functions, and job done!

Other Microcontrollers

I already mentioned the PICAXE. Its documentation is good, the tech support is excellent, and the language is easier to learn than C. So why didn't the PICAXE capture everyone's imagination? I don't know; maybe because it has a wacky name. I think you should check it out. Start by looking at its entry in Wikipedia.

The BASIC Stamp has a larger vocabulary of commands than the PICAXE, and a bigger range of add-on devices (including displays with graphical capability, and a little keyboard that is specifically designed for use with the controller). You can buy it in the form of surface-mount components squeezed onto a tiny board that will plug into a breadboard, as shown in Figure 5-97. A very nice design.

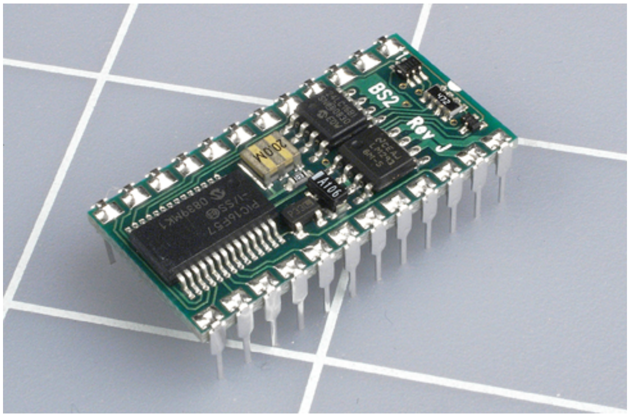

Figure 5-97 *The BASIC Stamp microcontroller consists of surface-mounted components on a platform that has pins spaced at 1/10-inch intervals, for insertion in a breadboard or perforated board.*

On the downside, you'll find that everything associated with the BASIC Stamp is a bit more expensive than in the PICAXE world, and the download procedure isn't quite as simple.

New products such as the Raspberry Pi extend the functionality of a microcontroller to the point where it becomes a real computer. By the time you read this, still more alternatives will be emerging in this turbulent field. Before you commit yourself to learning one of them in detail, I think it's a good idea to spend a day or two studying the online documentation and forums.

When I am thinking about learning something new, I do Google searches such as:

```
microcontroller problems OR difficulties
```

(I would substitute the name of an actual product for "microcontroller" in this search phrase.)

This is not because I am negative by nature. I just don't want to spend a lot of time on a product that turns out to have unresolved issues.

Unexplored Territory

It's time, now, for me to do some general summing up.

If you've taken the time to complete most of the projects in this book with your own hands, you have gained a very rapid introduction to the most fundamental areas of electronics.

What have you missed along the way? Here are some topics that remain wide open for you to explore. Naturally you should search online if they interest you.

The informal approach of Learning by Discovery that I have used in this book tends to be light on theory. I've avoided most of the math that you'd be expected to learn in a more rigorous course on the subject. If you have mathematical aptitude, you can use it to gain a much deeper insight into the way in which circuits work.

We didn't go very far into binary code, and you didn't build a half-adder, which is a great way to learn how computers function on the most fundamental level. But in *Make: More Electronics*, I show you how to do this.

I avoided going deeply into the fascinating and mysterious properties of alternating current. Here again, some math is involved, and just the behavior of current at high frequencies is an interesting topic in itself.

For reasons already stated, I avoided surface-mount components—but you can go into this area yourself for a relatively small investment, if you like the idea of creating fascinatingly tiny devices.

Vacuum tubes were not mentioned, because at this point, they are mainly of historical interest. But there's something very special and beautiful about tubes, especially if you can enclose them in fancy cabinetwork. In the hands of a skilled craftsperson, tube amplifiers and radios become art objects.

I didn't show you how to etch your own printed circuit boards. This is a task that appeals only to certain people, and the preparation for it requires you to make drawings or use computer software for that purpose. If you happen to have those resources, you might want to do your own etching. It could be a first step toward mass-producing your own products.

I didn't cover static electricity at all. High-voltage sparks don't have any practical applications, and they entail some safety issues—but they are stunningly impressive, and you can easily obtain the necessary information to build the equipment. Maybe you should try.

Op-amps and higher-level digital logic are other topics that I haven't touched here. However, they are included in *Make: More Electronics*.

In Closing

I believe that the purpose of an introductory book is to give you a taste of a wide range of possibilities, leaving you to decide for yourself what you want to explore next. Electronics is ideal for those of us who like to do things ourselves, because almost any application—from robotics, to radio-controlled aircraft, to telecommunications, to computing hardware—can be pursued by just one person, working at home, with limited resources.

As you delve deeper into the areas of electronics that interest you most, I trust you'll have a satisfying learning experience. But most of all, I hope you have a lot of fun.

Tools, Equipment, Components, and Supplies

5

This chapter is divided into five parts.

Kits. Various kits have been prepared, containing components and supplies that you can use to complete the projects in the book. See "Kits" on page 311 just below, for details.

Searching and Shopping Online. Instead of buying kits, you may prefer to do your own shopping. I have compiled some tips to help you. See "Searching and Shopping Online" on page 311 just below, for details.

In **Checklists of Supplies and Components** you will find an itemization of everything that you need. The lists of supplies begin at "Supplies" on page 316, and the lists of components begin at "Components" on page 317.

Buying Tools and Equipment. I have listed all the tools that are discussed at the beginning of each chapter of the book, and I have some suggestions about where to find them. See "Buying Tools and Equipment" on page 324.

Suppliers is a list of sources. Abbreviations in this list are used in the buying guides. See "Suppliers" on page 326.

Kits

Kits of components for the experiments in this book are still being finalized as the book goes to press. One kit should contain all the parts that you need for Chapters One, Two, and Three of the book. An additional kit should offer parts for Chapter Four. A separate soldering kit may also be available.

For information, please visit *www.plattkits.com*.

The page will be updated as more options become available. Please note that kits may be offered by independent suppliers who are not affiliated with Maker Media in any way.

Searching and Shopping Online

I'm including some general advice about searching for parts because many readers of the first edition of this book seemed to have difficulty getting the results they were looking for. I'll start with the most basic considerations, and work up. Even some seasoned shoppers may find a couple of tips here that are useful.

For a comprehensive list of all suggested suppliers, see "Suppliers" on page 326. Here are the ones that I consider primary:

Electronic components are available from large retail vendors online, most of which don't impose minimum quantities. Mouser, Digikey, and Newark are the obvious choices, maintaining huge inventories. Find them at:

Mouser Electronics ships from Texas

Digi-Key ships from Minnesota

Newark element 14 ships from Arizona

In addition, don't forget eBay, where prices are often lower than from other sources, especially if you use Asian sellers. eBay is not so useful for parts for which there is less demand, such as logic chips.

Tools and equipment are available from eBay, Amazon, and Sears, but if you want a truly monumental selection, McMaster-Carr is unbeatable.

They also have excellent tutorials—for instance, on the properties of different types of plastic, or the relative advantages of different drill bits.

The Art of Searching

The easiest search is for a specific part number, if you have one. You can enter it into the search field on a site such as mouser.com, where the algorithm is smart enough to allow some flexibility. For example, suppose you want a 7402 logic chip. Mouser will helpfully suggest that a SN7402N from Texas Instruments may be what you want, bearing in mind that Texas Instruments adds SN to the front and N to the end of the basic chip type.

However, the search will not be helpful if there is an additional code in the middle of a part number. When you search for a 7402, Mouser won't think of showing you any chips in the 74HC02 family, because HC is added in the middle.

Try the Chat Option

Suppose you only have an incomplete part number, or you don't know if the part is obsolete, or you generally need a bit of help. Don't overlook the option to make a voice call. A large distributor will have sales representatives who can assist you. It doesn't matter that you are an individual buying small quantities.

Better still, open a chat window. This will allow you to copy-paste a part number into the window and get a fairly quick answer advising you of similar options if the part is unavailable.

Parts on Google

If you want to do comparison shopping, use a generic search engine. I'll assume that Google is your default, because I feel it's the most appropriate for our purposes.

If a part number is long and complicated, you will have a better chance of finding what you want, and not being offered things that you don't want. Searching for a 7402 on Google will generate results including a Pantone ink color and an Institutes of Health standard. Searching for 74HC02 will narrow the results to logic chips.

Unfortunately, you are now likely to get hits from a lot of datasheet resellers. These companies harvest datasheets from electronics manufacturers and repackage them for you with ads that pay for this "service." It wouldn't matter, except that the reseller often shows you just one datasheet page at a time, because each page will have a new set of ads, all of which make money for the reseller. Waiting for each page to display is a waste of time, so I often use a hyphen as a minus sign to block datasheets when I search for components on Google, like this:

```
74HC02 -datasheet
```

Note that when you specify a part number, the search engine is less likely to compensate for any little errors that you make. Google understands that if you type "compoments" you probably meant "components," but it will not know that an 84HC02 chip should really be a 74HC02.

Datasheets

What if you actually want to see a datasheet, because you need to check the specification of a component before buying it? Go to one of the big distributors, find the part that you want, and you will see the option to click a datasheet icon. This will link you with a printable multipage document (almost always in PDF format) maintained by the component manufacturer itself. For me, this is a lot quicker than dealing with datasheet resellers on Google.

General Search Techniques

If you are looking for a component type, a search term that is brief and vague is usually inadequate. Suppose you search for:

```
switch
```

In my location, the first hit from this search was for light switches, the second hit was for a local wine bar, and I was then offered a variety of network switches (which are like routers). I also found a company named Switch that helps people to find a new job. How can you avoid these irrelevant hits?

As a first step, add a word to define your area of interest. For example, this may help:

```
switch electronic
```

Better still, if you want a DPDT toggle switch rated for 1 amp, just say so:

```
"toggle switch" dpdt 1a
```

Note the use of quote marks to nail down a specific phrase, discouraging Google from showing near-miss search results that are not quite what you asked for. Also note that search terms are not case-sensitive; there's no advantage in putting a term such as dpdt in caps.

You can narrow your search even further by naming a source, such as:

```
"toggle switch" dpdt 1a amazon
```

Why mention Amazon, if you can go to amazon.com and do your searching there? Because the search capability at amazon.com has fewer features than Google. In this example, it would not recognize the use of quote marks.

Fortunately Amazon allows Google to crawl all over its site and index everything, so that a search from Google can jump you straight to an Amazon list of toggle switches.

Exclusions

Use the minus option to avoid items that you don't want. For instance, if you are only interested in a full-size toggle switch, you could try this:

```
"toggle switch" dpdt 1a amazon -miniature
```

Note that the minus sign is another piece of syntax that Amazon's search feature doesn't understand.

Alternatives

Don't forget the AND and OR logical operators. If a single-pole, double-throw switch will work just as well for you as a double-pole, double-throw switch, you could try this on Google:

```
"toggle switch" dpdt OR spdt 1a -miniature
```

But even this may cause some problems, because naming conventions in electronics can be inconsistent. Some people refer to a DPDT switch as a 2P2T switch. Some call a SPDT switch a 1P2T switch. You'll need a lot of ORs to cover these alternatives.

Too Much Typing?

Personally I find that a carefully constructed, detailed search will save time by avoiding subsequent searches. However, if you don't want the chore of typing an elaborate search term, you have other options. One is to click the word "Images" which Google displays immediately above each set of search results, adjacent to the word "Web." Google Images will show you pictures of every conceivable kind of switch, and because our brains are well equipped to recognize images quickly, scrolling through a lot of pictures can be a more efficient way to find what you want than scrolling through a lot of text.

Alternatively, you can click the "Shop" option above Google's search results. This will give you the ability to list items in order of price from dozens or hundreds of different vendors. Some vendors will not be included, however.

Vendor Categories

Another option is to go to a vendor's site and use their system of categories. At mouser.com, digikey.com, and newark.com, if you search for "switch" you'll be shown a list of different kinds of switches. Click the type you want, and you'll be offered additional options, to narrow your search one step at a time.

Ultimately on mouser.com and other large-vendor sites you are likely to see little windows listing attributes such as voltages, amperages, and other values. This can be frustrating, because the lists are not intelligently managed. Some switches rated for half an amp, for instance, are grouped under 0.5A, while others are grouped separately as 500mA. These ratings are identical, but the people who create the listings just seem to copy the specifications from datasheets, some of which use amps while others use milliamps.

What to do? Use your Control-click option (Command-click on a Mac). Holding down the Ctrl key (or Command key) while you click the additional selection allows you to select the 0.5A switches *and* the 500mA switches *and* any others that might be suitable—including 1A switches, as a higher current rating will work fine at lower currents.

Which to Click First?

When using categories on a vendor site, it's useful to begin by choosing the attributes that you absolutely, positively need. For instance, if you are shopping for logic chips, begin by selecting through-hole versions, because you definitely won't want the tiny surface-mount versions. But note that a "DIP" package (meaning, dual-in-line pin) is almost the same thing as a "PDIP" (plastic DIP) package, which is the same as a "through-hole" version.

Conversely, any chip format identified by an acronym beginning with S is almost always a surface-mount version, which you do not want. SMT, in particular, means surface-mount.

A Real-Life Search

Here's an example of an actual search that I performed for a part that I used in this book. I knew what I wanted it to do, but I didn't know the part number.

I wanted a counter with a 3-bit output for use in the "Nice Dice" circuit (see "Experiment 23: Flipping and Bouncing" on page 218). So I went to Mouser Electronics and I started by searching for:

```
counter
```

While I was typing my search term, Mouser suggested an autocomplete:

```
Counter ICs
```

An IC is an integrated circuit, which is the same thing as a chip. So I clicked the autocomplete suggestion, which took me to a page suggesting 821 matches. Little scrolling windows would allow me to narrow the search by manufacturer, counter type, logic family, and much more. How should I proceed?

I scrolled horizontally to the window allowing me to choose the mounting style. Only two options: SMD/SMT (which are surface mount chips) and through-hole (the chips that plug into breadboards, not requiring a magnifying glass). I clicked the through-hole option and then clicked the Apply Filters button. That gave me 177 matches.

All the logic chips in this book are HC type in the 7400 family, so I went to the Logic Family window, and clicked 74HC. But, not so fast! I know that Mouser often lists the same thing under different names, so I scrolled through the other options. Sure enough I found HC listed separately from 74HC. I Control-clicked it to select both terms.

Now I had 52 options to choose from. As Counter Type, I selected Binary, because I wanted a binary output. This left me with 33 remaining matches.

There were no 3-bit chips, but I could use a 4-bit chip and ignore the highest bit. I saw two options in Number of Bits: 4 and 4-bit. I Control-clicked to select them both. Counting sequence could be Up, or Up/Down. I only wanted Up, so I clicked that. Now only nine matches left! Time to inspect the results. I wanted to use the most commonly available chip, which I determined by seeing how many of each one were in stock. I saw more than 7,000 of the SN74HC393N by Texas Instruments.

I clicked the datasheet link to make sure it would do what I wanted. A 14-pin chip providing maximum continuous output current of plus-or-minus 25mA with a nominal 5-volt supply ("nominal" means "typically used"). Yes, this was a standard logic chip in the 74HCxx family. Actually it contained two 4-bit counters, and I only needed one, but I wasn't going to quibble over that, and in fact I realized I could make use of the second counter in the chip, if I enlarged the scope of my project.

The 74HC393 would cost me about 50 cents. Might as well put six of them in the shopping cart. That's only $3, so maybe I should look for something else, reasonably small and light, so that I could add it without paying any additional shipping charge. But first I printed the datasheet for the 74HC393 and added it to my paper-based file-folder system.

You can see that this process entailed a lot of clicking. But it took me less than 10 minutes, and I found exactly what I wanted.

I could have followed a different path. Because I knew I wanted a chip in the 74xx family, I could have gone to this URL, which I keep bookmarked for easy reference:

www.wikipedia.org/wiki/List_of_7400_series_integrated_circuits

This includes all the 74xx logic chips that have ever been made. If you go to this page, you can press Control-F to search the text, and then type in:

```
4-bit binary counter
```

It has to be an exact match, which means you must type 4-bit, not 4 bit. The search yields 13 hits, and you can compare the features of the chips. After you choose one, you can copy its number and paste it into the search field of a site like Mouser, which takes you to that one component.

The only problem is, the Wikipedia page doesn't tell me which chips are old and almost out of production, and which are still popular. For my purposes, writing a book that I hope will be around for a while, I have to stick with the most popular components. This may be a good idea from your point of view, too, because if you build a circuit around an old chip, you are locking yourself into the past.

I could have used yet another approach, doing a Google search for people discussing and advising each other about counter chips. But you get the general idea. You don't need a part number, to find what you want.

eBay Options

I buy a lot of parts through eBay because I find bargains there, and also because most companies selling through eBay are extremely quick and reliable. To minimize your time and trouble, you need to know a few search basics that are specific to eBay.

First, don't hesitate to click the little "Advanced" option just to the right of the Search button on eBay's home page. This will allow you to specify attributes such as the country of origin (if you want to avoid overseas suppliers), and can limit your search to Buy It Now items. You can also specify a minimum price, which can be useful to eliminate stuff that is too cheap to be any good. Then, before starting the actual search, I usually click the display option for Price + Shipping: Lowest First.

Once you find what you want, it's time to check the seller's feedback. For sellers within the United States, I want 99.8% or better. I've never had a problem with sellers rated 99.9%, but I have been disappointed sometimes with service from sellers rated 99.7%.

If a supplier is in an Asian nation such as China, Hong Kong, Thailand, or others, you can be less fussy about feedback, because a lot of buyers give bad feedback when they don't receive something as quickly as they expect. Overseas sellers will warn you that a small packet will take 10 to 14 days in transit, but buyers complain anyway, and this drags down the feedback rating. In reality, in my experience, every item that I have ordered from overseas sources has always turned up, and has always been what I wanted. You just need to exercise a little patience.

After you find what you want on eBay, you may want to click the Add to Cart button, rather than the Buy it Now button, because you can look for additional items from the same seller, and you'll save time by grouping them into one shipment. This should also reduce shipping costs.

Click the Visit Store option in the Seller Information window, or if the seller doesn't have an eBay store, click to See Other Items. You then have the option to search within that seller's list of products. After you add as many as you want to your cart, it's checkout time.

You can make direct contact with overseas suppliers, instead of finding them through eBay. Tayda Electronics in Thailand (abbreviated tay in my list—see "Suppliers" on page 326) is a popular source.

Amazon

I don't think amazon.com is very useful for components, but it can be a good source for tools and for supplies such as wire or solder. The only problem I have is that Amazon doesn't like to show me the cheapest stuff first. You have to choose that option repeatedly after every search, and if the products are scattered among different store categories, the option to resequence the results will not exist. Even when you can prioritize the list by cheapness, Amazon (unlike eBay) is not smart enough to factor in the shipping cost. Pliers with a price of $4.95 and $6 shipping will be listed as being cheaper than pliers selling for $5.50 with $3 shipping. On the other hand, Amazon ships fast and if you buy a bunch of items at one time, all of them warehoused by Amazon themselves, you may be able to spend enough to get free shipping.

Killing Autocomplete

One last tip regarding Google. The default mode of the search engine will prompt you with a pop-up list of similar terms while you are still trying to type your search string. I get really annoyed with the interference of this autocomplete option, so I have disabled it—and you can, too.

In the address bar of your browser, use this URL to launch Google:

http://www.google.com/webhp?complete=0

Now save it as one of your favorites, and when you click the favorite, Google won't try to tell you what it thinks you are looking for. It will wait quietly for you to finish typing.

You can also use the URL as the default page that will open every time you launch your browser.

Is Searching Worth the Trouble?

You may feel that you don't want to memorize all these searching techniques. OK, that's why Maker Shed and myself are offering kits to go with this book. Buy a kit, and you should have all the components that you need, with no searching necessary.

But what are you going to do if you get interested in projects outside of this book? Suppose you see a circuit online—or suppose you want to modify a circuit, or design your own. At that point, I think you'll have to buy your own parts, and even if you try to obtain them all from one source, search techniques can be valuable.

Checklists of Supplies and Components

Photographs and general information are provided at the beginning of each chapter of this book. See "Necessary Items for Chapter One" on page 1, "Necessary Items for Chapter Two" on page 41, "Necessary Items for Chapter Three" on page 99, and "Necessary Items for Chapter Four" on page 142.

Below you will find listings of all the components and supplies. But I need to clarify the distinction between these two words.

Supplies are items such as solder or wire that I suggest you buy in a one-time purchase, sufficient for all the experiments. It doesn't make sense to consider how many inches of wire you will need for each project.

Components become an integral part of a project. You may be able to reuse these items, but only if you remove them from a previous project. Therefore, to take one example, a breadboard is included with components.

Supplies

The following supplies will be sufficient for all the projects. See "Suppliers" on page 326 for a list of sources from which you can buy these supplies, and abbreviations that I will use to refer to them.

Hookup Wire

You need 22 gauge, solid conductor, in at least two colors (red and blue), and preferably two more colors (your choice). Automotive wire is acceptable, so long as it is solid-core. Search eBay or Google for

```
solid wire 22 gauge OR awg
```

or check discount suppliers such as all, elg, and jam, or hobby suppliers such as ada and spk. (AWG is an acronym for American Wire Gauge.)

Quantities? If you want to do the experiments in Experiments 26, 28, 29, 30, and 31 that explore the world of inductance, you really need 200 feet of wire. Different colors can be joined together temporarily when winding coils. The wire can be unwound afterward and reused for other purposes.

If you are willing to skip the inductance experiments, I suggest you buy three spools of 25 feet each. You can find smaller lengths than 25 feet, but the price per foot goes up rapidly.

Jumpers

Personally I prefer not to use precut jumper wires, but if you choose them, one box should be sufficient. In addition you will need 25 feet of raw hookup wire, to make connections that are longer than the longest precut jumper. To find precut jumpers, you have to use the right search term. On Google, search for:

```
jumper wire box
```

The word "box" is the key to finding what you want. It automatically eliminates the type of undesirable flexible jumpers with plugs at each end, which are generally sold in bundles, not boxes. I don't think they are a good idea.

Stranded Wire

This is an optional addition, for situations where flexibility is important. One 25-foot spool will be enough.

Solder

This is usually sold by weight. See "Essential: Solder" on page 105 for the pros and cons of solder containing lead. Either way, be sure you buy *electronic solder* with a *rosin core*. The thickness may range from 0.02" to 0.04" (0.5mm to 1mm). If you only want to solder a couple of projects, three feet of solder will be sufficient, and some sources on eBay will sell very small quantities. Otherwise, try all, elg, jam, ada, amz, and spk.

Heat-Shrink Tubing

This is optional, but useful. One assortment of three or four (small) sizes will be enough. Because it has automotive applications, you can find it from hardware suppliers such as hom, har, and nor as well as hobbyist sources.

Perforated Board (unplated)

Only required in Experiment 14, although you can use it to build a permanent version of any projects in the book if you are willing to solder point-to-point wiring. A small piece, perhaps 4″ x 8″, will be enough for three average projects. Finding unplated board can be difficult, as most of it has copper or nickel solder pads. I think these are undesirable when doing point-to-point wiring, as they increase the risk of short circuits. Search for:

```
perforated board bare -copper
```

Also try searching for "prototyping board," or "proto board," or "phenolic board." Note that unplated board is also known as "unclad" in some places. At the time of writing, Keystone Electronics makes very small, cheap pieces of unplated perforated board, available through mou and dgk. You can also find unplated board at jam.

Perforated Board (plated)

This type of board is used for the finished version of Experiment 18, but of course you can use it for other projects where you want to make a permanent version. For convenience, use the type that has copper traces in the same pattern as connections inside a breadboard. This can be hard to find, because there are many different patterns, and the pattern that you want does not have a generally agreed name.

The BusBoard SB830 describes it as a "solderable breadboard" and it is currently available through amz. At ada, you find something similar named "Perma-Proto." GC Electronics 22-508 is another option, available through jam.

The Schmartboard 201-0016-31 (available through mou) is a two-part package consisting of a breadboard and a matching perforated board. The manufacturer suggests placing the perforated board over the breadboard and inserting components through both of them while developing and testing the circuit. Then lift up the perforated board, and the components are already in position, ready for soldering. Unfortunately this may not work for components with very short leads.

Machine Screws (bolts)

These and nuts with nylon inserts are available from hardware stores, but probably not in the small sizes necessary for attaching perforated board to the inside of a project box, or similar tasks. I suggest you buy #4 size flat-headed bolts in lengths of 3/8″ and 1/2″. My favorite source for this kind of hardware is McMaster-Carr.

Project Boxes

These vary a lot in price. Those made of ABS plastic are usually the cheapest. Try discount suppliers such as all, elg, and jam, or hobby suppliers such as ada and spk.

Components

Quantities and specifications for resistors, capacitors, and other components are listed below. See "Suppliers" on page 326 for a list of sources, and abbreviations that I will use to refer to them. The biggest suppliers are dgk, eby, mou, and nwk. You may find lower prices at all, elg, jam, and spk, but the selections will be smaller, and you should compare the shipping costs of buying from multiple suppliers with the cost of buying all your components in one shipment from a source where the prices are slightly higher.

Resistors

Any manufacturer is acceptable. Lead length is usually unimportant. All projects in this book can use a quarter-watt power rating (the most common value). A tolerance of 10% is acceptable, and the color bands on 10% resistors are easier to read than the bands on 5% or 1% resistors. However, you can buy 5% or 1% resistors if you wish.

The total number of resistors used in each section of the book is shown in Figure 6-1, but because resistors and capacitors are cheaply available in quantity, I don't think it makes sense to buy specific numbers for individual experiments. You will save time and money by buying packaged assortments.

- To buy sufficient resistors for *all* the projects in this book (with some to spare), get at least 10 of each of these values: 47 ohms, 220 ohms, 330 ohms, 1K, 2.2K, 4.7K, 6.8K, 10K, 47K, 100K, 220K, 330K, 470K, 680K, 1M. Also, get 20 of 470 ohms. Prepackaged assortments are your best bet. The quantities I have specified assume you will reuse some resistors after they serve their purpose in simple demonstration experiments.

Resistors	Book Chapters					Total
	1	2	3	4	5	
47 ohms				2	1	3
100 ohms				6		6
150 ohms				6		6
220 ohms				8		8
330 ohms				3	8	11
470 ohms	2	6	4	12		24
680 ohms				10		10
1K	2	2	1	4		9
2.2K	1			5		6
4.7K		4	2			6
6.8K					1	1
10K		1	1	41	4	47
47K				1		1
100K		2	1	4		7
220K		2				2
330K				1		1
470K		4	2			6
1M		1		4		5

Figure 6-1 *The number of resistors used in experiments in each chapter of the book.*

Capacitors

Find them at the same sources listed for resistors, above. Any manufacturer is acceptable. Radial leads are preferred, meaning that both leads emerge from the same end of the capacitor, instead of a lead at each end. A working voltage of at least 16VDC is recommended for power supplies up to 12VDC. You can substitute capacitors with higher working voltages, but the components will be physically larger. Other ratings such as temperature and impedance are not important for our purposes.

Ceramic capacitors are likely to last for many decades, while the longevity of electrolytic capacitors is a subject of some debate. For larger values, you have to use electrolytics, because the cost of ceramics becomes prohibitive. Personally I would use ceramics for values below 10µF and electrolytics for 10µF and upward, but you'll probably save money if you use electrolytics for values of 1µF and upward.

If you want to know the exact number of capacitors required in each chapter of the book, see Figure 6-2.

- To buy sufficient capacitors for *all* the projects in this book (with some to spare), get at least five of each of these values: 0.022µF, 0.047µF, 0.33µF, 1µF, 2.2µF, 3.3µF, 10µF, 100µF, 220µF. Also get at least 10 of values 0.01µF and 10µF. You only need two of each of these values: 15µF, 22µF, 68µF, 1,000µF. The quantities I have specified assume you will reuse some capacitors after they serve their purpose in simple demonstration experiments.

Capacitors	Book Chapters					Total
	1	2	3	4	5	
0.01µF		2		18	3	23
0.022µF				1		1
0.047µF				1		1
0.1µF		3		9		12
0.33µF		2		5		7
1µF		2		4	2	8
2.2µF					1	1
3.3µF		2	2	3		7
10µF		1		8	1	10
15µF				1		1
22µF				2		2
33µF		1				1
68µF				2		2
100µF		2		5	1	8
220µF		1	1		3	5
1,000µF		2			2	4

Figure 6-2 *The total number of capacitors used in experiments in each chapter of the book.*

Components other than Resistors or Capacitors	Book Chapters 1, 2, and 3	Extra for Chapter 4
LED (generic)	4	2
LED (low current)	1	15
Battery 9V	1	
Battery 9V connector	1	
Battery 1.5V	2	
Battery 1.5V carrier	1	
Breadboard	1	
Trimmer 500K	1	
Trimmer 100K		1
Trimmer 20K or 25K		1
Transistor 2N2222	6	
Speaker (small)	1	
Toggle switch	2	
Tactile switch	2	6
SPDT slide switch		2
Relay 9VDC DPDT	2	
AC-DC Adapter	1	
Diode 1N4001	1	
Perforated board 3" x 6"	1	
Fuse 3A	2	
Potentiometer 1K	2	
Lemons (or lemon juice)	2	
Galvanized brackets 1"	4	
Diode 1N4148		3
Timer 555 TTL type		4
7-seg LED display		3
4026B counter		3
74HC00 2-input NAND		1
74HC08 2-input AND		1
LM7805 regulator		1
74HC32 2-input OR		1
74HC02 2-input NOR		1
74HC27 3-input NOR		1
74HC393 counter		1

Figure 6-3 *Minimum numbers of components, assuming you will reuse items from each experiment in subsequent experiments. Items for Chapter Four are additional to items listed for Chapters One, Two, and Three.*

Other Components

For components other than resistors and capacitors, the minimum necessary quantities to build all the projects in Chapter One, Chapter Two, and Chapter Three of the book are shown in Figure 6-3. These quantities assume you will reuse all the components from each experiment in subsequent experiments. Components for Chapter Four are additional to those for previous sections of the book. Components for Chapter Five are not listed here, as the experiments are so diverse; see the beginning of each experiment in Chapter Five for a summary of options.

If you are concerned about burning out chips or transistors, which are vulnerable to damage, please add at least 1 to each quantity in Figure 6-3.

What if you may want to keep some of the projects that you build, instead of reusing the components for subsequent projects? In that case, please consult the tables for individual experiments that follow, and add the numbers of components for those experiment(s) which interest you.

The information that you need for finding and buying components is also provided below.

See "Suppliers" on page 326 for a list of suppliers that are referred to by abbreviations of their names. For most electronic components, go to all, eby, elg, jam, and spk for special deals, or dgk, mou, and nwk for one-stop shopping where everything will be available.

Components for Chapter One

Components for Chapter One, other than resistors and capacitors, are listed in Figure 6-4.

Components for Chapter One	Experiments					Total
	1	2	3	4	5	
LED (generic)			1	2		3
LED (low current)					1	1
Battery 9V	1		1	1		3
Battery 1.5V		2				2
Battery 1.5V carrier		1				1
Fuse 3A		2				2
Lemons (or lemon juice)					2	2
Galvanized brackets 1"					4	4
Potentiometer 10K				2		2
Deionized water (1 glass)					1	1

Figure 6-4 *Components other than resistors and capacitors, used in the first chapter of this book.*

Generic LEDs

The Lumex SLX-LX5093ID or Lite-On LTL-10223W are examples, but generic LEDs can be from any manufacturer. Probably 5mm LEDs are easier to handle, but 3mm LEDs can fit more easily into a crowded breadboard.

A typical forward current would be 20mA, typical forward voltage around 2VDC (blue and white LEDs will require a higher voltage). If you find a bunch of LEDs bundled for a low price on a site such as eBay, they can be considered generic.

Low-Current LEDs

These should be rated for 3.5mA forward current or less. The Kingbright WP710A10LID is an example, although the manufacturer, physical size, and color are unimportant. You could actually use this type of LED in all the experiments, but if you do, you should double the values of all the series resistors to protect it, as its maximum rating may be as low as 6mA.

Batteries

Nine-volt batteries can be the everyday alkaline type, available from supermarkets and convenience stores. Rechargable 9-volt batteries are an acceptable alternative.

The AA-size 1.5-volt batteries used in Experiment 2 must be alkaline. Do not use any type of rechargeable battery in this experiment.

Battery Connectors and Carriers

Just one carrier for a 1.5-volt battery will be necessary and sufficient. Note that a battery carrier may also be described as a *battery holder* or *battery receiver*. Make sure you get the type that holds only one AA battery (not two, three, or four). The Eagle 12BH311A-GR is an example.

You should buy at least three connectors for 9-volt batteries, because you may want to leave them attached to circuits that you build. Nine-volt connectors are sometimes described as *snap connectors* or *battery snaps*. Typical examples are the Keystone model 235 or Jameco Reliapro BC6-R. Buy whatever is cheapest, but make sure it terminates in wire leads.

Fuse

The 3A fuse in Experiment 2 ideally should be the automotive type, as its blades are easy to grip with alligator clips. Any automotive parts source will stock this type of fuse. The physical size is unimportant. Alternatively buy a 2AG size cartridge fuse, which is the smallest cartridge size, from an electronics supplier. It should be the fast-blow type, not a delay-fuse or "slow-blo" type. The voltage rating is not important. Littelfuse 0208003.MXP is an example.

Potentiometer

The full-size 1K potentiometer required for Experiment 4 should ideally be 1" in diameter, but sizes as small as half an inch are acceptable. Power rating, voltage rating, tolerance, shaft type, shaft diameter, and shaft length are unimportant. Select a potentiometer that has linear taper, and makes one turn, with panel mounting style, and solder lug terminals. Buy two. The Alpha RV24AF-10-15R1-B1K-3 and Bourns PDB181-E420K-102B are examples.

Juice and Brackets

If you use lemon juice in a squeeze bottle for Experiment 5, make sure it is undiluted and unsweetened. Vinegar is an acceptable substitute.

The 1" brackets for Experiment 5 must be galvanized. Pipe straps and hanger straps, to mount conduits and pipes, are an acceptable substitute. Any hardware source will have them cheaply available.

Deionized Water

This is often known as distilled water. Your local supermarket should have this, but make sure it is not "purified" and is not "spring water." It must have zero mineral content.

Components for Chapter Two

Components for Chapter Two, other than resistors and capacitors, are listed in Figure 6-5.

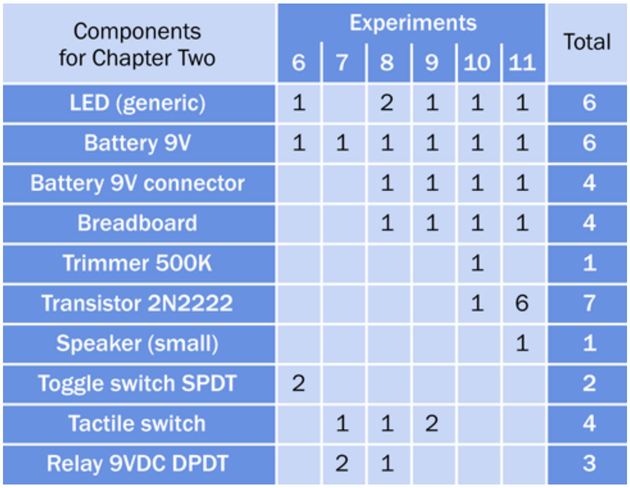

Components for Chapter Two	Experiments						Total
	6	7	8	9	10	11	
LED (generic)	1		2	1	1	1	6
Battery 9V	1	1	1	1	1	1	6
Battery 9V connector			1	1	1	1	4
Breadboard			1	1	1	1	4
Trimmer 500K					1		1
Transistor 2N2222					1	6	7
Speaker (small)						1	1
Toggle switch SPDT	2						2
Tactile switch		1	1	2			4
Relay 9VDC DPDT		2	1				3

Figure 6-5 *Components other than resistors and capacitors used in the second chapter of this book.*

Breadboards

A breadboard is classified here as a component, because it cannot be separated from the circuit; on the contrary, it is the foundation for the circuit. You have to decide how many circuits you may want to keep on their breadboards, and how many you are likely to disassemble so that the breadboard can be reused. Ideally, each breadboard should have a single bus on each side, and 700 connection points, as shown in Figure 2-10. Search Google or eBay for

```
solderless breadboard 700
```

However, if you prefer, you can use a dual-bus breadboard and ignore the extra lines of holes.

Trimmer Potentiometer

Trimmers of the recommended type are shown at left and at right in Figure 2-22, and a discussion of other types accompanies that photograph. The power rating is not important. The preferred type is single-turn, and it must terminate in pins that are spaced in multiples of 0.1" (2.54mm or 2.5mm). The Vishay T73YP504KT20 is a low-cost 500K trimmer.

Transistors

Before purchasing any 2N2222 transistors, see "Essential: Transistors" on page 49 for an important cautionary note.

Toggle Switch

This should be panel-mount type, ideally with screw terminals, although pins or solder lugs will be acceptable. It can be SPDT or DPDT. Voltage and current ratings are unimportant for the experiments in this book. The NKK S302T is an example, but you can find cheaper switches on eBay.

Tactile Switch

The type of tactile switch shown in Figure 2-19 is very strongly recommended, with two pins 0.2" apart, ideal for insertion in a breadboard. Avoid buying the more common tactile switches that have four pins or leads. The Alps SKRGAFD-010 is preferred (currently available from Mouser). Any tactile switch with 2 pins spaced 0.2" can be substituted, such as the Panasonic EVQ–11 series.

Relay

See "Essential: Relay" on page 48 for information about the recommended type of 9VDC, DPDT relay. The Omron G5V-2-H1-DC9, Axicom V23105-A5006-A201, and Fujitsu RY-9W-K have all been tested for suitability.

Components for Chapter Three

Components for Chapter Three, other than resistors and capacitors, are listed in Figure 6-6.

Many of the components for projects in Chapter Three have already been mentioned for Chapters One and Two; see above.

Components for Chapter Three	Experiments			Total
	13	14	15	
LED (generic)	2	1	1	4
Power source 9VDC	1	1	1	3
Breadboard			1	1
Transistor 2N2222		3	1	4
Diode 1N4001			1	1
Relay 9VDC DPDT			1	1

Figure 6-6 *Components other than resistors and capacitors used in the third chapter of this book.*

AC Adapter

This must have an output of 9VDC. It may have additional outputs providing different voltages. See "Essential: Power Supply" on page 99 for a discussion of the options. Minimum output should be 500mA (0.5A) DC.

If you want a multi-voltage adapter, finding one can be tricky, because if you search for "ac adapter" you will find hundreds or even thousands of single-voltage units. Your answer is to search a source such as eBay for:

```
ac adapter 6v 9v
```

This should provide you with several affordable multi-voltage options. Make sure the photograph of the unit shows a little switch to select the various voltages.

Diode

The 1N4001 switching diode is cheap and generic. Buy 8 or 10, and buy a similar quantity of 1N4148 signal diodes at the same time.

Headers

These miniature plugs and sockets are an optional item. Examples are Mill-Max part numbers 800-10-064-10001000 and 801-93-050-10-001000, or 3M part numbers 929974-01-36RK and 929834-01-36-RK.

Components for Chapter Four

Components for Chapter Four, other than resistors and capacitors, are listed in Figure 6-7.

Slide Switch

The recommended slide switch is SPDT with three pins spaced 0.1" apart, and is shown in Figure 4-5. I suggest the EG1218 made by E-switch. If you buy an alternative, it must terminate in solder pins for insertion in a breadboard. An example is the NKK CS12ANW03, but if you search eBay for

```
slide switch breadboard
```

you will find some that are much cheaper. The type of contact plating, voltage rating, and current rating are unimportant for the projects in this book.

Components for Chapter Four	Experiments									Total
	16	17	18	19	20	21	22	23	24	
LED (generic)	1	4	3	2		1	2		1	14
LED (low-current)					2	1	1	3	15	22
Power source 9VDC	1	1	1	1	1	1	1	1	1	9
Breadboard	1	1	1	1	1	1	1	1	1	9
Trimmer 20K or 25K	1			1						2
Trimmer 100K		1								1
Trimmer 500K	1									1
Tactile switch	2	1	1	3	2	8	2		1	20
SPDT Slide switch			2		1		2	2	2	9
Diode 1N4001			1			1				2
Diode 1N4148		1				3				4
Timer 555 TTL type	1	4	4	3		1	2		1	16
Speaker (small)		1	1							2
7-seg LED display				3						3
4026B counter				3						3
74HC00 2-input NAND					1			1		2
74HC08 2-input AND					1	1			1	3
LM7805 regulator					1	1	1	1	1	5
Relay 9V DPDT			1			1				2
Transistor 2N2222			2			1				3
74HC32 2-input OR							1		1	2
74HC02 2-input NOR								1		1
74HC27 3-input NOR									1	1
74HC393 counter									1	1

Figure 6-7 *Components other than resistors and capacitors used in the fourth chapter of this book.*

Integrated Circuit Chips

See "Fundamentals: Choosing Chips" on page 142 for a discussion of chips. While all the chips you will need are listed in Figure 6-7 (with the exception of one more 555 timer required for Experiment 29), it's a good idea to buy an extra chip of each type, as they are easily damaged by incorrect voltage, reversed polarity, overloaded outputs, or static electricity.

Any manufacturer is acceptable. The "package" of a chip refers to its physical size, and this attribute should be checked carefully when ordering. All logic chips must be in a DIP package (meaning a dual-inline package with two rows of pins that have 0.1" spacing). This may also be referred to as PDIP (meaning a plastic dual-inline package). They are also described as "through hole." The DIP and PDIP descriptors may be appended with the number of pins, as in DIP-14 or PDIP-16. This number can be ignored.

Surface-mount chips will have packaging descriptors beginning with S, as in SOT or SSOP. Do not buy any chips with "S" type packages.

The chip family used exclusively in this book is HC (high-speed CMOS), as in 74HC00, 74HC08, and similar generic identifiers. These numbers will have additional letters or numbers added by individual manufacturers as prefixes or suffixes, as in SN74HC00DBR (a Texas Instruments chip) or MC74HC00ADG (from On Semiconductor). These versions are functionally identical. Look carefully, and you will see the 74HC00 generic number embedded in each proprietary number.

Old TTL logic chips, such as the 74LS00 series, have compatibility issues. They are not used or recommended for any of the projects in this book.

555 Timer

Unlike the logic chips, you do want the TTL version of the timer (also known as the bipolar version), not the CMOS version. Here are some guidelines:

The TTL version (which you want) often states "TTL" or "bipolar" in its datasheet, specifies a minimum power supply of 4.5V or 5V, specifies an inactive current consumption of at least 3mA, and will source or sink 200mA. Part numbers often begin with LM555, NA555, NE555, SA555, or SE555. If you search by price, the TTL versions of the 555 timer are the cheapest.

The CMOS versions (which you don't want) always state "CMOS" on the first page of their datasheets, allow a minimum power supply of 2V in most cases, claim an inactive current consumption in microamps (not milliamps), and will not source or sink more than 100mA. Part numbers include TLC555, ICM7555, and ALD7555. If you search by price, the cheapest CMOS version of the 555 timer still costs almost twice as much as the cheapest TTL version.

Seven-Segment Display

The display used in Experiment 19 must be an LED device, height 0.56", low-current red preferred, able to function at 2V forward voltage and 5mA forward current. The Avago HDSP-513A is preferred, or Lite-On LTS-546AWC, or Kingbright SC56–11EWA, or similar.

Components for Chapter Five

Components for Chapter Five, other than resistors and capacitors, are listed in Figure 6-8.

Neodymium Magnets

I suggest K&J Magnetics as a source of supply, as the site maintains a very informative primer on magnets.

In Europe, supermagnete.de is a popular source.

16-Gauge Wire

This is only required for the antenna in Experiment 31. If the cost is prohibitive, try 50 or 100 feet of 22-gauge wire. If you live relatively close to an AM radio station, it should be adequate.

High-Impedance Earphone

Only required for Experiment 31, this can be ordered from the the Scitoys catalog.

You may also find them on Amazon. On eBay, search for:

```
crystal radio earphone
```

If you search eBay for crystal radio headphones instead of earphones, you'll find some antique items from the early days of radio.

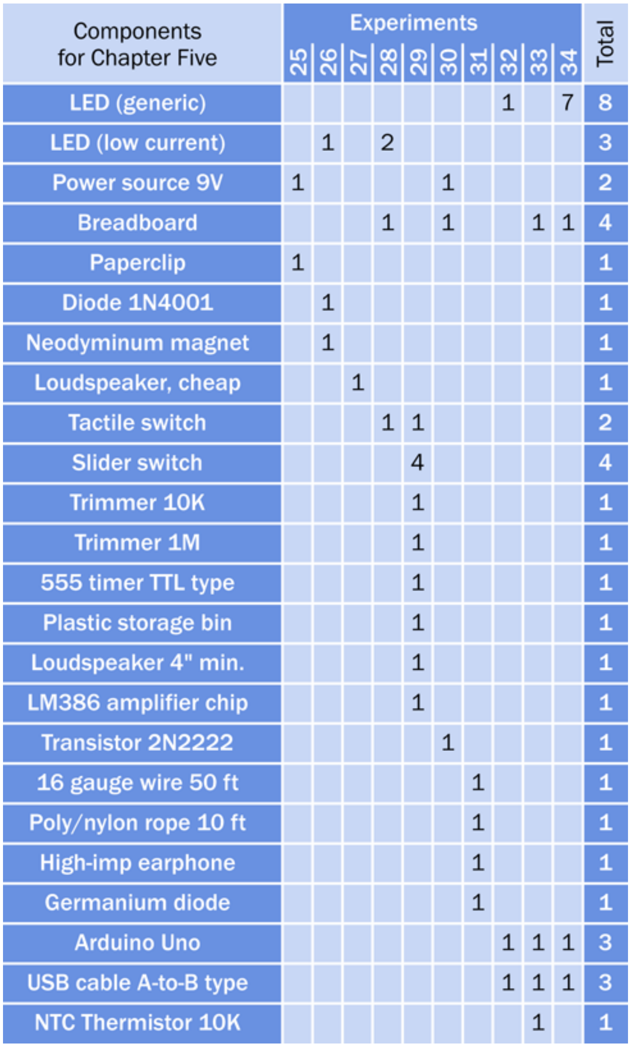

Components for Chapter Five	Experiments 25	26	27	28	29	30	31	32	33	34	Total
LED (generic)								1		7	8
LED (low current)		1		2							3
Power source 9V	1					1					2
Breadboard				1		1			1	1	4
Paperclip	1										1
Diode 1N4001		1									1
Neodyminum magnet		1									1
Loudspeaker, cheap			1								1
Tactile switch				1	1						2
Slider switch					4						4
Trimmer 10K					1						1
Trimmer 1M					1						1
555 timer TTL type					1						1
Plastic storage bin					1						1
Loudspeaker 4" min.					1						1
LM386 amplifier chip					1						1
Transistor 2N2222						1					1
16 gauge wire 50 ft							1				1
Poly/nylon rope 10 ft							1				1
High-imp earphone							1				1
Germanium diode							1				1
Arduino Uno								1	1	1	3
USB cable A-to-B type								1	1	1	3
NTC Thermistor 10K									1		1

Figure 6-8 *Components other than resistors and capacitors used in the fifth chapter of this book.*

Germanium Diode

Available from the same sources as the high-impedance earphone, above. Some may also be available from dgk, mou, or nwk.

Arduino Uno Board

For a discussion of sources, see "Beware of Imitations?" on page 280

Thermistor

The recommended thermistor for Experiment 33 is the Vishay 01-T-1002-FP. If you make a substitution, use a 10K NTC-type thermistor rated at 1% or 5% accuracy with wire leads.

Buying Tools and Equipment

See "Components" on page 317 for a list of components and see "Supplies" on page 316 for a list of supplies.

Photographs and general information about tools and equipment are provided at the beginning of each chapter of this book. See "Necessary Items for Chapter One" on page 1, "Necessary Items for Chapter Two" on page 41, and "Necessary Items for Chapter Three" on page 99. No additional tools are required for Chapters Four and Five.

Because products come and go, I have not included stock numbers or names of manufacturers of tools and equipment. The specifications and photographs at the beginning of each chapter should provide you with sufficient guidance, and if you restrict your search to large sites such as amazon.com or ebay.com, you may find everything you need fairly quickly, all in one place.

While it's true that expensive tools may be manufactured with greater precision and durability, the cheapest products should be satisfactory for the purposes of this book.

See "Suppliers" on page 326 for URLs that are referenced here with three-letter abbreviations

Tools and Equipment for Chapter One

For photographs and discussion of these items, see Chapter 1.

Only one of each item is required, unless otherwise specified.

Multimeter

For a discussion of multimeter features, see "The Multimeter" on page 1. Good sources include all, amz, eby, and jam.

Test Leads

Double-ended test leads should terminate in alligator clips approximately 1" long. The wire connecting them should measure 12" to 15" (not longer). You need at least three red and three black. Additional colors are useful.

You don't want the kind of test leads that have a plug at each end. Those are sometimes known as "jumper wires." Just search a site such as eBay for

```
test leads double ended alligator
```

and you should find what you want. Buy 10. Sources include all, eby, jam, or spk.

Safety Glasses

Try amz, eby, har, hom, or wal. Ideally, look for glasses with ANSI Z87 rating (you can use this as a search term). Avoid tinted glasses.

Tools and Equipment for Chapter Two

For photographs and discussion of these items, see Chapter 2.

Long-Nosed Pliers

They should measure approximately 5" end-to-end, with a flat inside jaw, not round. From amz, eby, mcm, mic.

Wire Cutters

Also known as "side cutters," and should measure approximately 5" end-to-end. From amz, eby, har, hom, nor, or mcm.

Flush Cutters

These are optional. From amz, eby, har, hom, nor, or mcm.

Wire Strippers

You need the type with specifically sized holes for numbered wire gauges, but the most common range (10 to 20 gauge) is not suitable.

I think you should buy wire strippers that have a specific provision for 22-gauge wire, because there's no point in making your task more difficult than it needs to be. Search online for:

```
wire strippers 20 30
```

This should find tools with holes for 20–, 22–, 24–, 26–, 28–, and 30–gauge wire. Alternatively, look on amz, eby, elg, jam, and spk.

Tools and Equipment for Chapter Three

For photographs and discussion of these items, see Chapter 3.

Low-Power Soldering Iron

This should be rated for 15W with a plated, slender, conical tip. Try all, amz, eby, jam, and mcm.

General-Duty Soldering Iron

Should be rated at 30W or 40W. Try amz, eby, har, hom, mcm, nor, or srs.

Helping Hand

Can be found at ada, amz, eby, jam, or spk.

For a small, close-up *magnifying lens* try amz, eby, or wal. It may be listed as a *magnifier* or a *loupe*.

Minigrabbers

The Pomona model 6244-48-0 is available from amz, dgk, mou, and nwk. For cheaper alternatives try eby, which is also your first choice for meter probes terminating in alligator clips.

Heat Gun

Usually sold as a general-purpose tool, and therefore is available from hardware stores. Try amz, har, hom, or nor. For a miniature heat gun, try eby.

Desoldering Equipment

Various options are available from amz, elg, jam, spk, and eby.

Soldering Iron Stand

This will be found in the same places that sell soldering irons.

Miniature Saw

My personal favorite is the #15 X-Acto blade. You also need the handle that it fits in. It is available online from Tower Hobbies, Hobbylinc, ArtCity, and many other arts/crafts sources. Also look for the larger X-Acto saw blade, #234 or #239, which you can use for cutting perforated board.

Deburring Tool

If your local hardware store does not stock this item, it is inexpensively available from amz, eby, mcm, nor, srs, and some specialty sources. The standard blade in this tool is intended for right-handed use. Left-handed blades are made, but can be hard to find. Some blades are harder than others; an E300 means that it is intended for soft metals and most plastics.

Calipers

I like Mitutoyo calipers, although many cheaper brands exist and will be sufficient for everyday use. The Mitutoyo website will show you all their available models, after which you can Google "Mitutoyo" to find retail sources. Many people prefer calipers with a digital display, switchable between metric and inches. I prefer calipers that do not require a battery.

Copper Alligator Clips

Available cheaply and in small quantities from the big general electronics suppliers such as dgk, mou, or nwk.

Suppliers

The three-letter abbreviations preceding each supplier are used throughout the text to suggest appropriate sources.

ada: Adafruit

all: All Electronics

amz: Amazon

dgk: Digi-Key

eby: eBay

elg: Electronic Goldmine

evl: Evil Mad Scientist

har: Harbor Freight

hom: Home Depot

ins: Instructables

jam: Jameco

mcm: McMaster-Carr

mic: Michaels crafts stores

mou: Mouser Electronics

nwk: Newark Electronics

nor: Northern Tool

plx: Parallax

spk: Sparkfun

srs: Sears

tay: Tayda Electronics

Many of these sites also host extensive tutorials and other helpful information. You can learn a lot by browsing through their pages.

Index

Numbers & Symbols

1N4001 diode, 108
1N4148 diode, 108
2N2222 transistor, 49, 84
4026B decade counter, 182, 184
  (see also counter)
555 timer
  applications, 152
  astable, 149, 155, 264
  audio frequency source, 260
  audio test, 156
  basics, 145
  bipolar version, 154
  bistable, 152, 184, 213
  buying guide, 323
  calibration, 189
  capacitor values, 150
  combinations, 159
  connected with counter, 184
  control pin, 150, 159
  delayed initial response, 165
  diode, 158
  discharge pin function, 148
  frequency, 157, 159
  history, 153
  internal function, 151, 157
  monostable, 149, 186
  pin numbering, 146
  pinouts, 146
  pulse duration, 149, 150
  pulse suppression, 152, 161, 168
  reset pin function, 149
  resistor values, 150
  retriggering, 149
  siren sound, 162
  specifications, 154
  test circuit, 147, 155
  trigger pin function, 147
  TTL and CMOS versions, 154
  unequal cycles, 158
556 timer, 163
7402 quad 2-input OR, 198
7404 hex inverter NOT, 198
7410 triple 3-input NAND, 198
7411 triple 3-input AND, 198
7420 dual 4-input NAND, 198
7421 dual 4-input AND, 198
7427 triple 3-input NOR, 198
7432 quad 2-input OR, 198, 217
744002 dual 4-input NOR, 198
744075 triple 3-input OR, 198
747266 quad 2-input XNOR, 198
7486 quad 2-input XOR, 198
74HC00 quad 2-input NAND, 192, 222
74HC08 quad 2-input AND, 204
74HC393 binary counter, 223
74xx family (see chip, 74xx family)

A

AC (alternating current)
  basics, 16
  symbol, 57
  through a coil, 258
  with capacitors, 84
AC adapter
  basics, 99
  buying guide, 322
  modification, 116
  single voltage type, 99
  universal type, 99
alarm system project (see intrusion alarm)
alligator clip
  copper, 106, 122, 123
  test leads, 5, 102
alternating current (see AC)
ambient pressure, 264
amp (see ampere)
amperage (see current)
ampere, 14
  (see also current)
  basics, 14, 15
  conversion table, 15
  measurement, 25, 37
Ampère, André-Marie, 14, 17
amplifier chip LM386, 259, 276
AND
  7408 quad 2-input, 204
  7411 triple 3-input, 198

- 7421 dual 4-input, 198
- 74HC08 quad 2-input, 193
- demo, 193
- symbol, 196
- truth table, 196

anode, 137
anti-static wrist strap (see wrist strap)
Arduino
- advantages and disadvantages, 289
- analog input, 290
- blink test, 284
- clock, 300
- comment line, 285
- dice simulation program, 297, 302
- errors, 303
- IDE, 280, 281
- installation, 282
- integers, 304
- introduction, 278
- language reference, 297
- Mac installation, 283
- protoshield, 293
- pseudocode, 298
- random function, 297, 305
- setup, 280
- sketch, 285
- summary, 287
- thermostat simulation program, 294
- unlicensed copies, 281
- Windows troubleshooting, 283

Atmel, 279, 289
atomic structure, 35
audio (see sound)
autoranging in multimeter (see multimeter, autoranging)

B

Bardeen, John, 90
base of a transistor, 86
BASIC Stamp, 308
battery
- attaching to breadboard, 68
- basics, 4
- buying guide, 320
- car, 12, 32
- carrier, 4, 13
- connector, 4
- electrolyte, 13
- how it works, 10, 13
- internal resistance, 32
- lemon juice, 34
- lithium, 12, 32
- parallel, 39
- primary, 36
- secondary, 36
- series, 39
- short circuit, 12, 13, 32
- symbol, 57

Bell, Alexander Graham, 253
bias, 88
binary code, 226
binding posts, 176
board
- cut-board, 107
- perforated, 107, 126, 170
- perforated errors, 172
- perforated, buying guide, 317

bolts, 107
bolts buying guide, 317
books recommended, 242
Boole, George, 194
boolean logic, 195
box, project, 107, 173
Brattain, Walter, 90
breadboard
- basics, 43, 66
- bus, 44, 68
- buying guide, 321
- component depiction, 66
- dual-bus, 44
- fault tracing, 73
- interior connections, 68
- mini, 44
- single-bus, 44

break-to-make circuit, 133
bubble, 196
button blocker
- breadboarded, 216
- phase 1, 211
- phase 2, 211
- phase 3, 212
- phase 4, 212
- phase 5, 212
- schematic, 213
- sequence, 213

C

C language, 285
calipers, 105
Camenzind, Hans, 153
capacitance, 4
capacitor
- AC application, 84
- basics, 49, 72
- buying guide, 318
- bypass, 96
- ceramic, 50, 72
- charge-discharge demo, 71, 76
- coupling, 80, 96
- DC blocking, 76
- displacement current, 82
- electrolytic, 50, 72
- filtering audio frequencies, 262
- measuring the value, 4
- polarity, 71, 73
- polyester, 266
- RC network, 76
- shock hazard, 72
- symbol, 73
- tantalum, 73
- test circuit, 75
- time constant, 77
- variable, 276

car battery, 12, 32
carrier frequency, 277
cathode, 137
ceramic capacitor (see capacitor, ceramic)
checklists of supplies and components, 316
chip
- 4000 series, 197
- 74xx family, 142
- buying guide, 323
- dual-inline package, 142
- HC, 143
- history, 145, 197
- LS series, 197
- package, 142

  part number, 142, 198
  pin numbering, 146
  size, 146
  socket, 143
  static discharge damage, 178
  surface-mount, 142, 143
  through-hole, 142
  TTL and CMOS, 197
clipping, 268
clocked flip-flop, 222
CMOS output current, 182
coil
  AM radio, 272, 277
  audio filter, 262, 262, 266
  compared with capacitor, 257
  inductance, 245
  inductive reactance, 257
  iron core, 246
  loudspeaker, 253
  power generation, 247
  symbol, 246
  tap, 273
  transformer, 246
  voltage spike, 139
collector of a transistor, 86
combination lock
  breadboarded, 202, 204
  computer interface, 207
  enhancements, 209
  installation, 207
  introduction, 201
  logic diagram, 202
  schematics, 202
  testing, 206
common anode, 181
common cathode, 181
comparator, 151
components
  checklist, 317
  vendors, 311
conductor definition, 9
connector, 108
contacting the author, x
continuity testing, 3, 53
conventional current (see current, conventional)
coulomb, 37
counter
  4026B, 182
  74HC393, 223
  binary, 223
  carry out, 184
  clock disable, 183
  clock input, 183
  coded outputs, 182
  decade, 182, 223
  edge-triggered, 226
  modulus, 227
  reset pin, 227
crossover network, 266
current
  basics, 14, 16
  conventional, 55
  danger, 32
  displacement, 82
  flow direction, 36
  measurement, 25
  resistance, relationship, 14
  uniform in a simple circuit, 26

D

Darlington array, 182
datasheet, LED, 29
DC (direct current)
  basics, 16
  plug-and-socket, 108
DeArmond, Harry, 270
debouncing (see switch, bounce)
deburring tool, 105
decimal points, 31
desk lamp, 127
desoldering braid, 104
desoldering pump, 101, 103
desoldering wick, 104
dice simulation
  introduction, 223
  logic diagram, 229
  randomicity, 234
  schematic, 231
  slowdown effect, 235
  two dice, 233
diode
  1N4001, 108
  1N4148, 108
  555 timer, 158
  anode, 137
  basics, 108, 137
  buying guide, 322
  cathode, 137
  germanium, 274
  germanium buying guide, 324
  protection, 139
  radio, 278
  rectifier, 136, 251
  symbol, 137
  voltage drop, 137
  with logic chip, 206
DIP (see chip, dual-inline package)
direct current (see DC)
displacement current, 82
display, numeric
  seven-segment display, 144
distortion, 268
double-throw switch (see switch, configurations)
DPDT (see switch, configurations)
drilling holes in plastic, 174
Dummer, W. A., 145

E

earphone, 274
earphone buying guide, 323
edge-triggered counter, 226
electricity and magnetism, 243
electrolyte, 13
electrolytic capacitor (see capacitor, electrolytic)
electromagnet, 244
electron holes, 35
electrons, 10, 13, 32, 35, 36, 37
emitter of a transistor, 86
Encyclopedia of Electronic Components, 242
errata reports, xi
error notifications, x

F

farad
  basics, 72
  conversion table, 72
Faraday, Michael, 72, 75, 246
fault tracing, x, 73, 173

feedback
  from author to reader, x
  from reader to author, x, xi
filtering audio frequencies, 262
flip-flop, 151, 153, 207, 213, 219, 222
floating pin, 149, 183
flush cutters, 42
forward current, 20
forward voltage, 20
Franklin, Benjamin, 36
fuse
  automotive, 5, 15
  buying guide, 320
  cartridge, 5, 15
  how to blow, 15
  suitable types, 5

G

gauge of wire (see wire, gauge)
germanium, 145, 274
glasses, safety (see safety glasses)
ground
  grounding yourself, 178
  power outlet, 16
  radio, 274, 275
  symbol, 57

H

Hammond organ, 270
harmonics, 264
HC generation (see chip, HC)
headers, 108
headers buying guide, 322
heat gun, 103, 115
heat sink, 123
heat transfer, 114, 122, 123
heat-shrink tubing, 106, 114, 117, 119, 120
heat-shrink tubing buying guide, 317
helping hand, 101, 109, 119, 122
henry (unit of inductance), 245
Henry, Joseph, 245, 246
hertz, 157
Hertz, Heinrich, 157
holes (positive charges), 35
hookup wire (see wire, hookup)
hysteresis, 294

I

IC (see chip)
inches
  conversion of fractions, 129
  conversion to metric, 129
inductance
  basics, 28
  created by a coil, 245
inductive reactance, 257
insulator definition, 9
integers, 304
integrated circuit (see chip)
internal resistance of a battery, 32
intrusion alarm
  arming, 167
  breadboard layout, 139
  complete circuit, 168
  installation, 176
  part one, 132
  part two, 163
  soldering, 170
  usage, 167
  wish list, 132
inverter (see NOT)

J

jack plug, 54
jam-type flip-flop, 222
Jeopardy game, 210
joule, 32
jumper wire
  basics, 46
  how to make, 67
  precut, 46
  precut buying guide, 316
  with plugs, 5, 46, 261

K

keypad encoding, 209
Kilby, Jack, 145
kilohm, 9
kilovolt, 15
kilowatt, 33
knife switch (see switch, knife)

L

latch, 153
leads, test (see test leads)
learning methods, ix
LED
  basics, 6, 20
  buying guide, 320
  datasheet, 29
  display, 181
  forward current, 20
  forward voltage, 20
  generic, 6
  heat damage, 121
  indicator, 181
  installation, 128
  low-current, 6, 144
  low-current buying guide, 320
  polarity, 20
  series resistor, 24, 30
  symbol, 55, 58
  test circuit, 20
  threshold voltage, 24
  through-hole, 181
lemon battery, 34
light-emitting diode (see LED)
lightning, 37
lithium battery (see battery, lithium)
LM386 amplifier chip, 259, 276
LM7805 voltage regulator, 144, 191
logic chip (see chip)
logic diagram
  combination lock, 202
  dice simulation, 229
  NAND demo, 193
logic gate, 192
  (see also AND, NAND, OR, NOR, XOR, XNOR, and NOT)
  rules, 200
  voltages, 200
logic probe, 142

logic symbols, 196
logic-high, 193
logic-low, 192
loudspeaker
  basics, 50
  coil, 253
  disassembly, 252
  enclosure, 259
  history, 253
  mounting in a tube, 98
loupe, 102

M

machine screws, 107
machine screws buying guide, 317
magnet
  buying guide, 323
  cautions, 250
  charging a capacitor, 251
  power generation, 247
  wire, 248
magnet wire, 267
magnetic field, 257
magnetic sensor (see switch, magnetic sensor)
magnetism and electricity, 243
magnifying lens, 102, 127
Make: More Electronics, xi, 222, 235, 242, 294
Marconi, Guglielmo, 277
measurement units, 129
megawatt, 33
megohm, 9
meter (see multimeter)
metric system
  conversion to inches, 129
  origins, 129
microamp, 15
microcontroller
  applications, 279
  ATmega, 279
  history, 288
  options, 308
microfarad, 72
microphone, 254
milliamp, 15
millivolt, 15
milliwatt, 33
minigrabbers, 102
modulus of a counter, 227
Moore, Gordon, 145
multimeter
  autoranging, 3
  basics, 3, 4, 8
  best, 2
  better, 2
  BK, 2
  buying options, 2
  cheapest, 2
  continuity testing, 3
  current measurement, 26
  Extech, 2
  probes, 8
  purpose, 2
  ranging, 3
  recommended, 2
  setup, 7
  sockets, 8

N

NAND
  7410 triple 3-input, 198
  7420 dual 4-input, 198
  74HC00 quad 2-input, 192
  demo, 192
  gate, 193
  switch debouncing, 220, 222
  symbol, 196
  truth table, 195, 196
nanofarad, 72
newton, 38
nicerdice Arduino program, 302
NOR
  7402 quad 2-input, 198
  7427 triple 3-input, 198
  744002 dual 4-input, 198
  switch debouncing, 220
  symbol, 196
  truth table, 196
NOT
  7404 hex inverter, 198
  inverter, 196
  symbol, 196
Noyce, Robert, 145
numeric display (see seven-segment display)

O

ohm
  conversion table, 9
  definition, 8
  European notation, 9
  named after, 9, 11
Ohm's law
  applications, 29, 30, 31
  basics, 28
  units, 28
Ohm, Georg Simon, 9, 11
omega symbol, 9
OR
  7432 quad 2-input, 198, 217
  744075 triple 3-input, 198
  symbol, 196
  three or more inputs, 213
  truth table, 196
oscillator, two-transistor, 94

P

P2N2222 transistor, 49
part numbers for chips (see chip, part number)
PDIP (see chip, dual-inline package)
perforated board (see board, perforated)
PICAXE, 289, 308
picofarad, 72
pinouts
  4026B counter, 182
  555 timer, 146
  7402 quad 2-input NOR, 198
  7404 hex inverter NOT, 198
  7410 triple 3-input NAND, 198
  7411 triple 3-input AND, 198
  7420 dual 4-input NAND, 198
  7421 dual 4-input AND, 198
  7427 triple 3-input NOR, 198
  7432 quad 2-input OR, 198, 217
  744002 dual 4-input NOR, 198
  7440750 triple 3-input OR, 198

  747266 quad 2-input XNOR, 198
  7486 quad 2-input XOR, 198
  74HC00 quad 2-input NAND, 193
  74HC08 quad two-input AND, 193
  74HC393 binary counter, 223
  LM386 amplifier chip, 259
  LM7805 voltage regulator, 191
  seven-segment display, 181
pliers (see tools)
plug and socket, 108
plywood, 107
point-to-point wiring, 126
pole of switch (see switch, pole)
potential difference
  definition, 25
  measuring, 25, 28
potentiometer
  basics, 5
  buying guide, 320
  internal parts, 22
  symbol, 58
  test circuit, 22, 23
  track, 23
  trimmer, 49, 86
  wiper, 23
power
  basic requirements, 4
  connector, 108
  generation with magnet and coil, 247
  outlet, 16
  supply, 99, 100, 128
  symbol, 57
power cord, shortening, 118
power, watts, 32
printed circuit, 124
program
  Arduino IDE environment, 285
  code, 285
  compile, 285, 286
  pseudocode, 298
  storage, 288
  upload, 285, 287
programmable unijunction transistor (see PUT)
project box, 107, 173
project box buying guide, 317
protection diode, 139
prototyping board (see breadboard)
pseudocode, 298
pulldown resistor (see resistor, pulldown)
pullup resistor (see resistor, pullup)
pulsing glow project, 124
pushbutton, 47
  (see also switch, tactile)
  switch, 144
  symbol, 58
PUT, 95

R

radio
  AM, 271, 275, 277
  amplifier, 276
  antenna, 273
  coil, 272, 277
  crystal, 271
  enhancements, 276
  ground, 274
  history, 277
randomicity, 234, 298, 305
ranging in multimeter (see multimeter, ranging)
RC network, 76, 95, 149
reactance, 257
reed switch (see switch, reed)
reference books, 242
reference sources online, 242
reflex tester
  basics, 178
  calibration, 189
  circuit malfunction, 188
  enhancements, 190
  operation, 187
  wish list, 184
registering for updates, x
relay
  basics, 48
  buying guide, 321
  buzzer circuit, 70
  coil voltage, 65
  combination lock, 204
  datasheets, 63
  inductive load, 66
  internal parts, 62, 65
  intrusion alarm, 134, 165
  latching, 48, 63, 135
  LED flasher, 67
  opening it up, 63
  operating current, 65
  oscillator, 66, 70
  pin functions, 63
  polarity of coil, 48, 60
  power requirements, 209
  resistive load, 66
  self-locking, 135
  set voltage, 65
  small-signal type, 63
  switching capacity, 66
  symbols, 63
  terminology, 65
  test circuit, 60
  transistor comparison, 91
resistance
  basics, 14
  current, relationship, 14
  parallel, 28
  series, 28, 30
  skin, 10
  tongue, 9, 10
  water, 11
  wires and heat, 31
resistor
  basics, 6
  buying guide, 317
  color codes, 19
  power rating, 33
  pulldown, 149, 183
  pullup, 149
  standard multipliers, 21
  storage, 18
  symbol, 58
  value measurement, 18, 21
rosin core, 101

S

safety glasses, 4
saw, miniature, 104

schematic
  basics, 55, 56
  color-coded, 59
  crossovers, 59
  layout, 58, 150
  pin-shuffling, 150
screwdrivers, 41
Scribner, Charles E., 54
search techniques, 312
self-inductance, 257, 258
semiconductor, 89
seven-segment display, 144, 180, 181
seven-segment display buying guide, 323
Shannon, Claude, 195
Shockley, William, 90, 145
shopping online, 311
short-circuit a battery, 12, 32
Siemens, Ernst, 253
Signetics, 153
sine wave, 263
single-throw switch (see switch, configurations)
skin resistance, 10
slide switch (see switch, slide)
socket for IC (see chip, socket)
solder, 100, 101, 105, 109
solder buying guide, 316
soldering
  alternatives, 112
  errors, 111
  myths, 111
  technique, 109, 113, 116, 120, 171
  wave-soldering, 112
soldering iron
  basics, 100
  general duty, 101, 109, 122
  low-power, 100, 113, 122
  pencil type, 101
  pistol grip, 101
  stand, 104
  Weller, 101
solderless breadboard (see breadboard)
sound
  and electricity, 254
  distortion, 268
  filtering frequencies, 262
  frequency, 157
  microphone, 254
  siren, 162
  waves, 254
sound synthesizer, 91
SPDT (see switch, configurations)
speaker (see loudspeaker)
SPST (see switch, configurations)
square wave, 264, 268
static electricity (see chip, static discharge damage)
stomp box, 269
storage boxes, 238
subwoofer, 266
suppliers, 326
supplies checklists, 316
switch
  bounce, 218
  configurations, 51
  continuity testing, 53
  domestic lighting, 56
  jack-knife, 54
  knife, 51
  magnetic sensor, 133, 144, 177
  momentary, 53
  pole, 51, 175
  reed, 133
  screw terminals, 47
  slide, 144
  slider buying guide, 322
  soldering, 175
  sparking prevention, 53
  symbol, 56
  table of configurations, 53
  tactile, 47
  tactile buying guide, 321
  tactile symbol, 58
  telephone systems, 54
  test circuit, 50
  toggle, 47, 51
  toggle buying guide, 321

T

tactile switch (see switch, tactile)
test leads, 5, 102
Texas Instruments, 145
thermistor, 290, 291
thermistor buying guide, 324
Thomson, J. J., 36
time constant, 77
timer, 555 (see 555 timer)
toggle switch (see switch, toggle)
tongue
  resistance, 9, 10
  test with 9V battery, 7
tools
  buying guide, 324
  flush cutters, 42
  pliers, long-nosed, 41
  pliers, sharp-nosed, 42, 122
  screwdrivers, 41
  vendors, 312
  wire cutters, 42
  wire strippers, 43
transformer, 246
transistor
  2N2222 substitutions, 49, 84
  amplifying current, 88
  base, 86
  basics, 49
  bipolar, 86
  collector, 86
  damage, 90
  emitter, 86
  fingertip demo, 85
  history, 90
  measuring the value, 4, 90
  NPN vs. PNP, 86
  P2N2222, 49
  principle, 85, 88
  relay comparison, 91
  summary, 89
tremolo, 270
trimmer (see potentiometer, trimmer)
truth table, 195, 195, 196
tweeter, 266

U

uncontrolled variable, 11, 80
units of measurement, 129

V

vendors
    components, 311
    tools, equipment, 312
    URLs, 326
Venn diagram, 194
vibrato, 270
volt
    conversion table, 15
    named after, 13
Volta, Alessandro, 17
voltage
    basics, 13, 14, 15, 37
    drop, 137
    regulator, 99, 144, 191

W

wafer, 142
water, 11
    resistance, 11
watt
    basics, 32, 38
    conversion table, 33
    formula, 32
    power, 32
Watt, James, 33
waveform, 263, 264, 268
Wheeler's approximation, 245
wire
    bulk, 45
    color-coded, 46, 59
    gauge, 43, 45, 67
    hookup, 45, 109
    hookup buying guide, 316
    joining two pieces, 117
    magnet, 248, 267
    solid, 45
    spools, 45
    stranded, 45, 47, 175
    stranded buying guide, 316
wire cutters, 42, 43
wire strippers, 43
wire, jumper (see jumper wire)
wire-wrap, 112
woofer, 266
work area layout, 237
work bench configuration, 238
wrist strap, 178

X

XNOR
    747266 quad 2-input, 198
    symbol, 196
    truth table, 196
XOR
    7486 quad 2-input, 198
    symbol, 196
    truth table, 196

About the Author

Charles Platt became interested in computers when he acquired an Ohio Scientific C4P in 1979. After writing and selling software by mail order, he taught classes in BASIC programming, MS-DOS, and eventually in Adobe Illustrator and Photoshop. He wrote five computer books during the 1980s.

He has also written science-fiction novels such as *The Silicon Man* (published originally by Bantam and later by Wired Books) and *Protektor* (from Avon Books). He stopped writing science fiction when he started contributing to *Wired* in 1993, and became one of its three senior writers a couple of years later.

Charles began contributing to Make: Magazine in its third issue and is currently a contributing editor. Currently he is designing and building prototypes of medical equipment in his workshop in the northern Arizona wilderness.

Colophon

The cover photograph is by Marc de Vinck.

The heading and cover font are BentonSans, the text font is Myriad Pro, and the constant-width font is TheSansMonoCondensed.

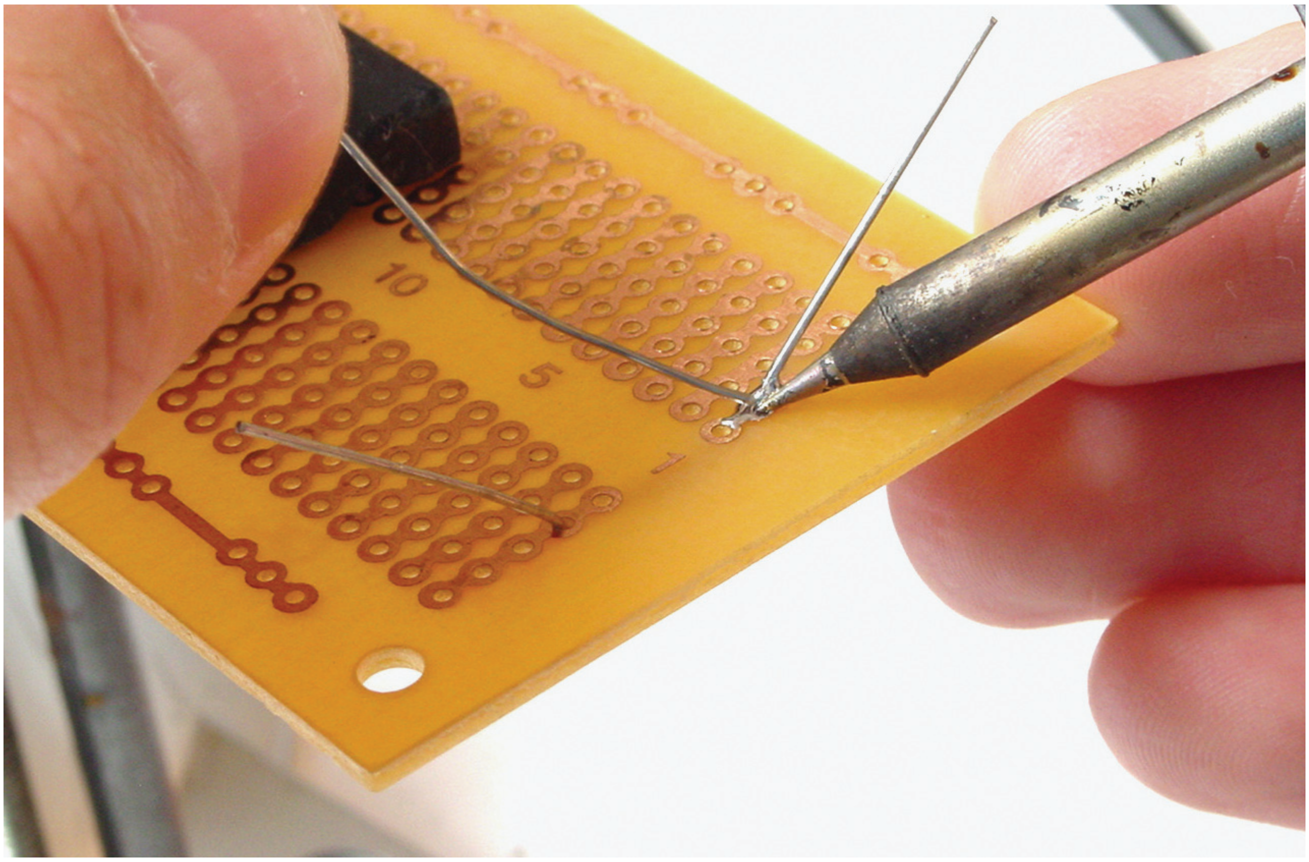

Soldering and Desoldering

Step-by-step instructions for making (and unmaking) the perfect solder joint.

By Joe Grand

The two key parts of soldering are good heat distribution and cleanliness of the soldering surface and component. With practice, you'll become comfortable and experienced with the process.

In this primer, I'll explain how to solder a component onto a printed circuit board (PCB). I'll also provide desoldering tips and show you how to remove a surface-mount component from a printed circuit board using a Chip Quik kit. And I'll show you how to remove a component by removing the solder in a way that won't damage the components or the circuit board.

Photography by Joe Grand

Reprinted with permission from *Hardware Hacking*, copyright 2004, Syngress Publishing, ISBN: 1-932266-83-6, pp. 34–40.

Tools of the Trade

Soldering iron You could pay as little as $10 or as much as $1,000 for a soldering iron. I recommend a fine-tip, 700°F, 50W soldering stick iron. A good general-purpose iron for hardware hacking is the Weller W60P Controlled-Output Soldering Iron, which sells for under $70.

Solder Should be thin gauge (0.032" or 0.025" diameter) 60/40 rosin core.

Desoldering tool (aka solder sucker) A manual vacuum device that pulls up hot solder, useful for removing components from circuit boards. I like the one RadioShack sells (#64-2098, $10).

IC extraction tool Helps lift integrated circuits from the board during removal/desoldering.

Chip Quik SMD Removal Kit Allows you to remove surface mount components quickly and easily. Chipquik.com offers the kit for $16.

Sandpaper A very fine-grit sandpaper is useful for removing oxidation from component and circuit board surfaces.

Desoldering braid Woven metal material used to wick up melted solder.

Small, flat-tip screwdriver Comes in handy for removing some types of components.

Needlenose pliers, wire cutters, and vise These common tools will make your job easier.

The Chip Quik SMD Removal Kit

The Chip Quik SMD Removal Kit allows you to quickly and easily remove surface-mount components such as PLCC, SOIC, TSOP, QFP, and discrete packages. The main component of the kit is a low-melting-temperature solder (requiring less than 300°F) that reduces the overall melting temperature of the solder on the SMD pads. Essentially, this enables you to just lift the part right off the PCB.

CHIPQUIK
CURITY
Alcohol Prep
2 Ply - Medium
CHIP QUIK
No-Clean Paste Flux

INCLUDES

- Alcohol pads for cleaning the circuit board after device removal
- A special low-melting-temperature alloy
- Standard no-clean flux
- Application syringe

Desoldering Tips

FOR STANDARD THROUGH-HOLE COMPONENTS

- First grasp the component with a pair of needlenose pliers. Heat the pad beneath the lead you intend to extract and pull gently. The lead should come out. Repeat for the other lead.
- If solder fills in behind the lead as you extract it, use a spring-loaded solder sucker to remove the excess solder.

FOR THROUGH-HOLE ICs OR MULTI-PIN PARTS

- Use a solder sucker or desoldering braid to remove excess from the hole before attempting to extract the part.
- You can use a small, flat-tip screwdriver or IC extraction tool to help loosen the device from the holes.
- Be careful to not overheat components, since they can become damaged and may fail during operation.

BEFORE YOU START

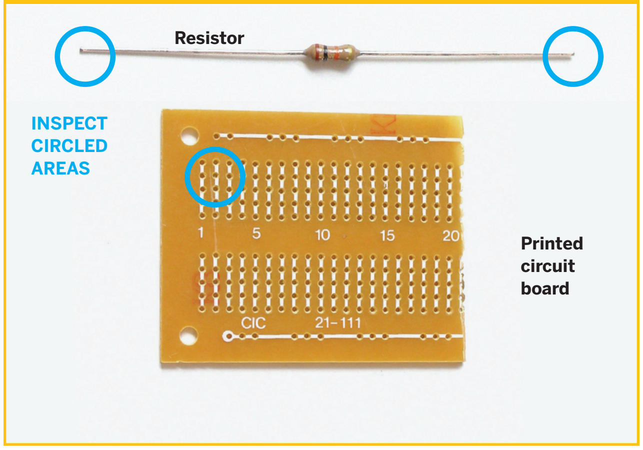

Joe Grand is the president of Grand Idea Studio, Inc. (joe@grandideastudio.com), a product-development and intellectual-property licensing firm. He specializes in embedded system design, computer security research, and inventing new concepts and technologies. He is also a host on The Discovery Channel's *Prototype This*.

SOLDERING A RESISTOR TO A CIRCUIT BOARD

BEFORE YOU START

Inspect the leads or pins for oxidation. If the metal surface is dull, sand with fine sandpaper until shiny. In addition, use the sandpaper to clean the oxidation and excess solder from the soldering iron tip to ensure maximum heat transfer.

This simple example shows the step-by-step process to solder a through-hole component to a printed circuit board (PCB). I used a piece of prototype PCB and a single resistor.

DANGER: It's important to consider safety precautions. Improper handling of the soldering iron can lead to burns or other physical injuries. Wear safety goggles and other protective clothing when working with solder tools. With temperatures hovering around 700°F, the tip of the soldering iron, molten solder, and flux can quickly sear through clothing and skin. Keep all soldering equipment away from flammable materials and objects. Be sure to turn off the iron when it's not in use and store it properly in its stand.

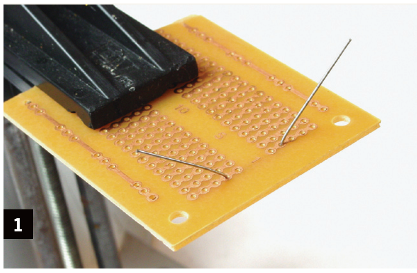
1

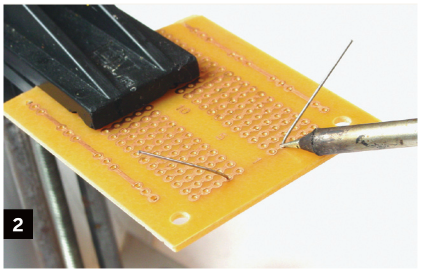
2

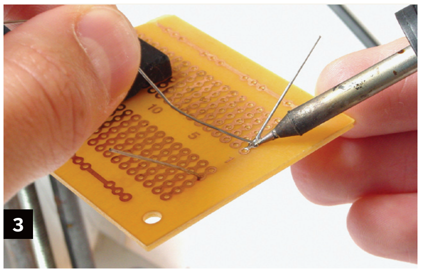
3

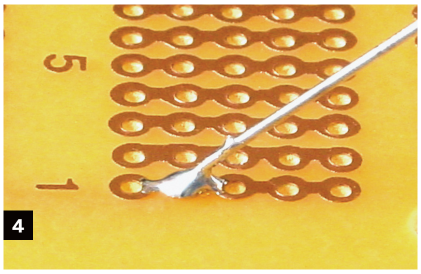
4

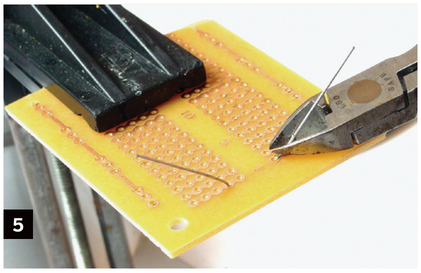
5

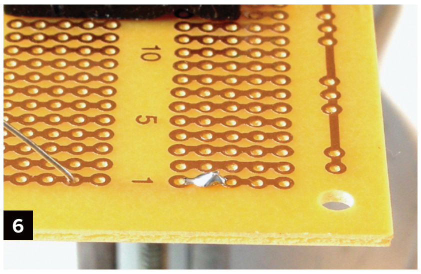
6

1. Bend and insert the component leads into the desired holes on the PCB. Flip the board to the other side. Slightly bend the lead you'll be soldering to prevent the component from falling out when the board is turned upside down.

2. To begin the actual soldering process, allow the tip of your iron to contact both the component lead and the pad on the circuit board for about 1 second before feeding solder to the connection. This will allow the surface to become hot enough for solder to flow smoothly.

3. Next, apply solder sparingly and hold the iron in place until solder has evenly coated the surface. Ensure that the solder flows all around the 2 pieces (component lead and PCB pad) that you're fastening together.

Don't put solder directly onto the hot iron tip before it has made contact with the lead or pad; doing so can cause a cold-solder joint (a common mistake that can prevent your hack from working properly). Soldering is a function of heat, and if the pieces aren't heated uniformly, solder may not spread as desired. A cold-solder joint will loosen over time and can build up corrosion.

4. When it appears that the solder has flowed properly, remove the iron from the area and wait a few seconds for the solder to cool and harden. Do not attempt to move the component during this time.

The solder joint should appear smooth and shiny, resembling the image above. If your solder joint has a dull finish, reheat the connection and add more solder.

5. Once the solder joint is in place, snip the lead to the desired length. Usually, you'll simply cut the remaining portion of the lead that isn't part of the actual solder joint. This prevents any risk of short circuits between leftover component leads on the board.

6. Here's a completed soldering example.

p050-53_SIP_SolderPrimer_F2.indd 52 10/13/10 6:33:28 AM

DESOLDERING

SMD REMOVAL WITH CHIP QUIK

BEFORE YOU START

Use a rubbing alcohol pad to remove any residue from the solder pads. Verify that the solder pads are clean and free of cuts or solder jumps before proceeding. Desoldering, or removing a soldered component from a circuit board, is typically trickier than soldering because you can easily damage the device, the circuit board, or surrounding components. For surface mount devices (SMDs) with more than a few pins, the easiest method to remove the part is the Chip Quik SMD Removal Kit, as shown in the following step-by-step example.

PLEASE READ through this example completely before attempting SMD removal on an actual device. When removing the device, be careful not to scratch or damage any of the surrounding components or pull up any PCB traces.

THREE PRIMARY FUNCTIONS OF FLUX

- **Cleans metal surfaces to assist the flow of *filler metals* (solder) over *base metals* (device pins).**
- **Assists with heat transfer from heat source (soldering iron) to metal surface (device pins).**
- **Helps in the removal of *surface metal oxides* (created by oxygen in the air when the metal reaches high temperatures).**

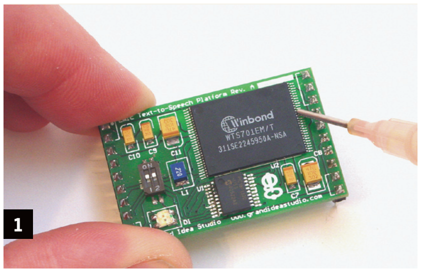
1

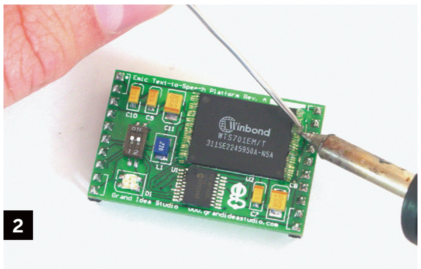
2

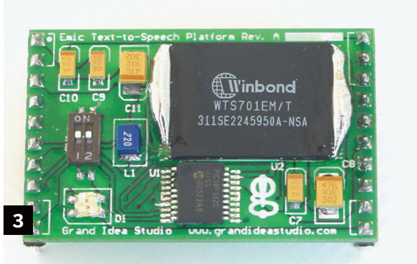
3

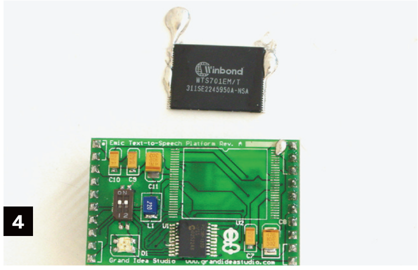
4

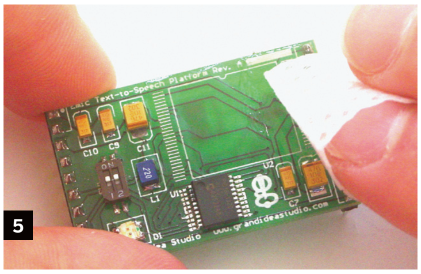
5

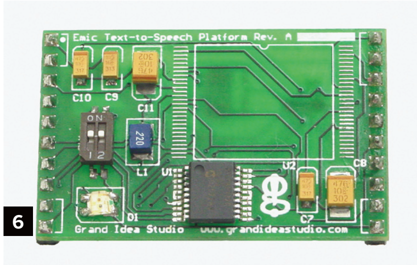
6

1. The first step is to assemble the syringe, which contains the no-clean flux. Simply insert the plunger into the syringe and push down to dispense the compound. The flux should be applied evenly across all the pins on the package you'll be removing. (Flux is a chemical compound used to assist in the soldering or removal of electronic components or other metals.)

2. Once the flux is evenly spread over the pins of the target device, the next step is to apply the special Chip Quik alloy to the device. This step is just like soldering: apply heat to the pins of the device and the alloy at the same time. The alloy has a melting point of approximately 300°F, which is quite low. You shouldn't have to heat the alloy with the soldering iron for very long before it begins to melt. The molten alloy should flow around and under the device pins. Starting at one end of the device, simply heat and apply the alloy. Repeat for the other side(s) of the device.

3. Flux will help ensure a nice flow of the alloy onto the device pins. Make sure the alloy has come in contact with every single pin by gently moving the soldering iron around the edges of the device. Avoid touching nearby components on the PCB with the soldering iron.

4. Now that the alloy has been properly applied to all pins of the device, it's time to remove the device from the board. After making sure that the alloy is still molten by reheating all of it with the soldering iron, gently slide the component off the board. You can use a small, jeweler's flat-tip screwdriver to help with the task. If the device is stuck, reheat the alloy and wiggle the part back and forth to help the alloy flow underneath the pads of the device and loosen the connections.

5. The final step in the desoldering process is to clean the circuit board. This step is important because it will remove any impurities left behind from the Chip Quik kit and get you ready for the next step.

First, use the soldering iron to remove any stray alloy left on the device pads or anywhere else on the circuit board. Next, apply a thin, even layer of flux to all of the pads that the device was just soldered to. Use the included alcohol swab or a flux-remover spray to remove the flux and clean the area.

6. The desoldering process is now complete. The surface-mount device has been removed and the circuit board cleaned.

Mathematical Astrology

CHAPTER 1
ASTROLOGICAL TERMINOLOGY -1

Under this topic 'Astrological Terminology' we propose to discuss and acquaint our students with various terms and their meaning, definition, etc. commonly used in astrology, particularly those used in mathematical Astrology. In addition, certain astronomical terminology will also be discussed in these lessons, to the extent these are used in mathematical astrology. The various terminology with which the students are expected to be familiar are as follows

1. The solar system.
2. The earth
3. The equator of the earth
4. Northern hemisphere, and southern hemisphere.
5. Geographical longitudes (Rekhansha) & Geographical latitudes (Akshansha)
6. Meridian of Greenwich as reference point at the earth's equator
7. Celestial sphere or the cosmic sphere
8. Celestial poles
9. Celestial Equator
10. Ecliptic or the Ravi Marga
11. Zodiac
12. Celestial longitude (Sphuta)
13. Celestial latitude (Vikshepa)
14. Declination (Kranti)
15. Right ascension (Dhruva)
16. Oblique ascension or Rashimaan
17. Equinoctial points
18. Precession of the equinoxes and Ayanamsha
19. Moveable and fixed zodiacs
20. The Sayana and Nirayana system
21. The Table of Ascendants
22. The Table of Houses
23. The Ephemeris

We will now take the above mentioned terms and discuss these one by one so as to make these terms clear to the students. It may however be mentioned here· that a large number of the above mentioned terms are quite simple and self explanatory. Most of the students, particularly those who have studied the geography as a subject during their school education, would be familiar with the terms mentioned above. Nevertheless we. will discuss and explain all the above mentioned terms in a systematic manner so that the very concept of these terms is understood by the students.

1. The Solar System

Our Solar System is centered round the Sun. Nine planets viz. Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus (or Herschel), Neptune and Pluto alongwith belt of Asteroids revolves in elliptical orbits around the Sun.

In Hindu Astrology, the last three planet i. e. Uranus (or Herschel), Neptune and Pluto have no place. On the other hand the classical Hindu Astrology recognizes the Moon and the two shadowy planets i. e. Rahu and Ketu (or the Moon's Nodes) as equivalent to planets. Rahu and Ketu are not physical bodies but are mathematically calculated sensitive points of intersection of the orbits of the Moon and the Sun (or in fact that of the Earth but which appears to be that of the Sun).

1.1 The planets Mercury and Venus are situated in the space between the Sun and the Earth. These planets are therefore known as 'Inner Planets'. These are also known as 'Inferior Planets'.

Figure 1

1.2 The other three planets namely Mars, Jupiter and Saturn are so situated in the space that their orbits are on the outer side of the EaI1h. These Planets are therefore known as 'Outer Planets' or 'Superior Planets'.

2. The Earth

The Earth is ever spinning on its axis. In addition to its spinning, the earth is also revolving round the sun. It is therefore always in a state of motion in the space at a speed of nearly 30 kms per second or 1,600 kms a minute or 96,60,00,000 kms per year.

2.1 In early times, the earth was believed to be the centre of universe or our solar system. It was thought that the Sun, the Moon and other planets (stars) actually revolved around the earth, as they appear to do. But now we know that the earth is a globe, that it rotates or spins on its axis and the Sun and stars appear to revolve around it from east to west, because the earth is revolving around its axis from west to east. This phenomenon can be best explained with the example of a moving train. When we look out side the window from a running train, the trees at a distance or the telephone poles, electric posts and other similar objects appear to be moving in the opposite direction to that of the train, which actually is not correct. We know that the trees, the telephone poles, the electric posts etc., all are fixed at a place. Similarly living on the earth's surface, we also keep on moving in the space with the same speed as that of the earth. The sun which is actually stationary would appear to us to be moving in the opposite direction to that of the earth. As the earth is moving from west to east, the Sun and other stars in the space will appear to be moving in the opposite direction i. e. from east to west. This is what we actually observe also.

2.2 The axis of the Earth slants at an angle of about 231;i° from the perpendicular to the plane of its orbit. If the plane of orbit of earth is treated as horizontal, then perpendicular to this lane will be known as vertical and then the axis of earth can be stated to be slanting at an angle of about 231;2^{0} (23 degrees 28 minutes to be precise) to the vertical.

Figure 2

The axis of the earth is so inclined that the northern end of the axis always points to the Pole Star or commonly known as Dhruv Tara. Where the northern and southern end of the axis of earth meet the surface of the earth, those points are known as North and South Poles of the earth, respectively.

3. The Equator of the Earth

If we draw a plane passing through the centre of the earth and perpendicular to the earth's axis, it will cut the surface of the earth in a circle. This great circle on the surface of the earth is known as the Earth's Equator or terrestrial Equator.

Figure 3

4. Northern and Southern Hemisphere

We know that the globe of the earth is not a perfect sphere like a ball. In fact the earth's diameter along the equator is larger than its diameter along the axis due to the fact that the earth is slightly flattened at the poles where as it is slightly bulging out at the equator. The shape of the Earth is comparatively more similar to that of an orange or a melon rather than that of a perfect sphere. Even then, for easy comprehension/calculations and understanding the various phenomenon, we consider the earth's globe to be a perfect sphere, though it is actually not so. In Para 1.6 above we have seen that the imaginary plane cuts the earth's surface in a great circle known as earth's equator. However, if the same plane was to cut the earth's globe (or the sphere) into two parts, each part will be exactly half of the sphere and will therefore be known as the Hemisphere. The hemisphere towards the North end of the axis of the earth is known as Northern Hemisphere. Similarly the hemisphere towards the south end of the axis of the earth is known as Southern Hemisphere. Figure 4 illustrates the above phenomenon clearly where in the two halves of the earth's globe have been shown separated at the plane of the earth's equator.

Figure 4

5. Geographical Longitudes (Rekhansha) and Geographical Latitudes (Akshansha)

In order to fix the position of an object or a point on a plane, we have to divide the plane by drawing two sets of parallel lines at equal intervals perpendicular to each other. A graph paper which all of us would have used in our school days, is a good example to understand this phenomenon. In the adjacent figure 5 we have two sets of parallel lines which are at equal intervals and at the same time are perpendicular to each other, i. e. to say that all lines in N-S direction are perpendicular to all the Lines in W-E direction. Similarly all lines in W-E direction are parallel to each other but perpendicular to the lines in N-S direction. 'With the help of these equidistant parallel and perpendicular lines, we can correctly find the coordinates of any given point viz. A, B, C or D with reference to any given point of reference (say '0').
For example:
For 'A' we can say 7 units in E direction and 7 units in N direction.
For 'B' we can say 6 units in W direction and 5 units in N direction.
For 'C' we can say 8 units in W direction and 8 units in S direction. and similarly
For 'D' we can say 4 units in E direction and 5 units in S direction.

Figure 5

Alternatively, if the coordinates of any point are known, we can locate the point exactly on the plane by counting the number of units indicted by the coordinates, in the appropriate direction. The same concept is applied to the earth's surface also with slight modifications as the surface of the earth is not a perfect plane but is having curvature, the earth's globe being a sphere for all practical purposes.

5.1 The surface of the earth's sphere is imagined to be cut by several planes each one of them passing through the centre of the earth and perpendicular to the plane of Earth's equator. These planes will describe imaginary circles on the surface of the earth so that each one of these imaginary circles will be passing through the North as well as the South pole of the earth and will have the same centre as that of the earth. The distance measured along the surface of the earth between any two such consecutive circles will be zero at both the poles (as all the circles will be passing through the poles) and will be maximum at the equator. These circles are known as the 'Meridians of Longitude'. These have been explained in the figure 6.

Figure 6

5.2 Again let us imagine the surface of the earth to be cut by imaginary planes which are all parallel to the plane of earth's equator. These planes will also describe circles on the surface of the earth and the centres of all such circles will be falling on the axis of the earth and each one of these circles will be parallel to each other as well as parallel to the earth's equator. These circles are known as parallels of Latitude.

5.3 Students will recall that the 'meridians of longitudes' are nothing but concentric circles on the surface of the earth whose planes are all perpendicular to the plane of equator. Similarly, the 'parallels of latitudes' are again circles on the earth's surface but with their planes parallel to the plane of earth's equator. It is therefore self evident that at any given point on the surface of earth, the meridian of longitude and the parallel of latitude will be mutually perpendicular to each other and will therefore intersect each other at right angles or 90°.

5.4 Students are advised to re-read Para 1.8.1 to 1.8.3 above so that the application of the concept of 2 sets of equidistant parallel lines, each set being mutually perpendicular to the other set (Para 1.8) could be properly understood by them to locate or identify any place or city on the surface of the earth.

Figure 7

5.5 We have already seen that the earth's equator is a circle. As any circle comprises of 3 60° of arc so the earth's equator will also have 360°. For easy comprehension, we may imagine that there are 180 numbers of concentric circles drawn on the surface of earth in such a way that their planes. are perpendicular to the plane of earth's equator. These 180 circles will describe 360 lines on the surface of earth (each circle will give two lines i.e. one in the front and the other at the back) which as we already know (Para 1.8.1) are known as meridians of longitudes. Each of these 360 meridians of longitude will pass from both the poles of the earth and at equator will be 1 ° apart. The distance between any two consecutive lines measured along the surface of earth will be maximum at earth's equator which will go on decreasing as we proceed along these lines either towards North Pole or towards the South Pole where it will become 'Zero'.

5.6 We may also consider for easy comprehension that the circles which are known as the Parallels of Latitude are also 180 in numbers i. e. 90 circles in the Northern Hemisphere and the remaining 90 circles in the southern Hemisphere so that the angular distance (angle substanded at the centre of earth) between any two consecutive circle is 1^0 again as in the case of Meridians of the longitudes. We will therefore, have a set of parallel lines at 1^0 angular distance apart running from E to W or W to E around the earth's globe all of which will be perpendicular to the Meridians of longitude (para 1.8.3).

5.7 We can now super-impose the figures 6 and 7 and see that the new figure formed by merging or superimposing the two figures will have a graph like appearance drawn on the surface of the earth which by and large will be somewhat similar to figure 5. The only exception will be that the lines in N-S direction or the Meridians of longitudes will not be eactly parallel to each other in the true sense. However as the students may be aware that earth's globe has a circumference of about 40,232 kms or 25,000 miles (approx.), the space of earth's surface covered between two consecutive lines of 1^0 angular distance in N-S as well as E-W directions will be roughly of the order of 110 kms x 110 kms or 69 miles x 69 miles. Hence we may consider them to be parallel for the place or city under consideration.

Figure 8

5.8 From the figure 8 above though it is clear that the meridians of longitudes are never exactly parallel in the strict sense, but as explained in para 1.2.5.7 for the limited spaces marked as 'A', 'B', and 'C' on the earth's surface these meridians (shown by dotted arrows in the figure) are considered as parallel. Therefore the conditions of figure 5 in para 1.2.5.8 above are considered to have been fulfilled.

5.9 Having drawn 2 sets of parallel lines at equal distance which are mutually perpendicular also, we are now set to locate any place on the surface of earth. We now only need to know its coordinates from a given reference point. In the context of earth's globe these coordinates are known as 'geographical longitudes' which are measured along the earth's equator either towards 'East' or 'West' from the reference point/line. The other coordinate being the geographical latitudes which are measured in perpendicular direction from earth's equator either towards 'North' or 'South' from the reference point or line. For the purpose of longitudes, the reference line or the reference meridian has been chosen as the meridian passing through Greenwich near London. This meridian i.e. the meridian passing through Greenwich is considered as 0° longitude and the longitudes of all other places on earth is measured with reference to this meridian only either towards East or towards West. Hence all places, cities etc. on the surface of earth are located within either 0^0E to 180^0E longitude or 0^0W to 180^0W longitude. Similarly for the purposes of latitudes the reference line or parallel of latitude is the equator itself. The latitudes of all places, cities etc. situated on the surface of the earth are measured from the equator whose latitude is 0°, either towards North or South depending on whether the place is in Northern or Southern Hemisphere. Hence the latitudes vary from 0^0N to 90^0N for places in Northern Hemisphere and from 0^0S to 90^0S for places situated in Southern Hemisphere. Thus the point of intersection of 0° longitude i.e. the Meridian of Greenwich with the Earth's Equator is considered as the reference point 'O'

5.10 Students would have seen that the explanation for Geographical longitudes and latitudes have been dealt with in much greater detail and is quite exhaustive in its content. If the phenomenon is clear with reference to the earth's globe, students will find it easy to understand when the same is applied to the space and the planets, which is of our primary concern while talking about the Astrology.

CHAPTER 2

ASTROLOGICAL TERMINOLOGY II

2.1 In the previous chapter we have seen how to locate or define a place on the earth's surface. We will now apply the similar principles to the space and see how to locate or define the position of various planets situated in the space. For this purpose, we will have to imagine that the entire space around our planet earth is a huge sphere with infinite diameter which extends far beyond the farthest of the planets with which we are concerned in Astrology. So living on this planet earth, the other planets in the space including the sun and the Moon would appear to us to be situated on the imaginary surface of this imaginary sphere.

2.2 Celestial Sphere or the Cosmic Sphere

The imaginary sphere in the space surrounding our entire Solar system, mentioned in Para 2.1 above, is known as the celestial sphere or the cosmic sphere.

2.3 Celestial Poles

If the Earth's axis is extended infinitely towards North and South, it will meet the imaginary surface of the cosmic sphere or the celestial sphere at some point. These points on the surface of cosmic sphere are known as the Celestial Poles and the extended axis becomes the imaginary axis of the celestial sphere.

2.4 Celestial Equator

The projection of earth's equator or the terrestria equator on the imaginary surface of the celestial sphere is known as the Celestial Equator.

2.4.1 As the earth's equator divides the earth's globe into two halves, similarly the celestial equator divides celestial or cosmic sphee into two equal halves or hemispheres. These are known as Northern celestial hemisphere and the Southern celestial hemisphere.

2.5 Ecliptic (Ravi Marg)

The apparent path of the Sun in the space along which it seems to move around the earth is known as Ecliptic. This is also known as Ravi Marg. The Ecliptic or the Ravi Marg, like the orbits of other planets is not a circle but is elliptical or oval in shape. Ecliptic can also be defined as a projection of Earth's orbit around the Sun on to the surface of cosmic sphere. The plane of Ecliptic is inclined to the plane of celestial equator at an angle of about 23~ ^{0} due to the slant! inclination of the earth's axis to the vertical. Figure 9 given below will clarify the position.

Figure 9 of book

2.6 Zodiac

If one observes the movement of planets, it is seen that they also move in their own orbits along with the Sun's path, but their path deflects north-south also. However the planets never proceed more than 9° either north or south of the ecliptic. Hence if a parallel line on either side of the ecliptic is drawn at an angular distance of about 9° then the ecliptic will come in the middle and either side will be a broad band/path way in which all planets can be located. This imaginary belt/band stretching about 9° north and 9° south of the ecliptic within which the planets and the Moon remain in course of their movement in the heavens, is known as Zodiac. In astrology we refer to this broad band of 18 ° instead of referring to the entire sky.

2.7 Celestial Longitude (Sphuta)

This is the arc of the ecliptic intercepted between the first point of Aries (Nirayana) and a perpendicular arc to the ecliptic drawn through the body (planet) and the poles of the ecliptic. In other words it can also be defined as the angular distance of any heavenly body (viz. planets etc.) measured in degrees along the ecliptic, in one direction from the origin (or the reference point - **first point of Aries of the zodiacal sign or the vernal Equinox).** The first point of Aries is different in **Sayana** and **Nirayana** system. Students will recall that in the case of Geographical longitudes, the measurement was along he terrestrial equator and it was either towards east or west from the Greenwich or the reference point or 0° longitude so that the maximum longitude of any place on the surface of earth could be either 180^0E or 180^0W. However in the case of **Zodiac or to say the celestial sphere, the measurement of celestial longitude of any planet is in one direction** only from the origin or the reference point. As such the celestial longitudes of various planets will be from 0° to 360° and in this case there is nothing like measuring towards east or west.

2.8 Celestial Latitude (Jlikshepa)

It is the angular distance across the celestial sphere measured north or south from the ecliptic along the great circle passing through the poles of the ecliptic and the object.

2.9 Declination (Kranti)

It is the angular distance on the celestial sphere north or south of the celestial equator. It is measured along the hour circle passing through the celestial object.

2.10 Right Ascension (Dhruva)

It is the angular distance on the celestial sphere measured eastward along the celestial equator from the vernal equinox to the hour circle passing through the celestial object.

2.11 It is considered necessary here to clarify the position to the students with regard to Para 2.7 to 2.10 above. While dealing with places on the surface of the earth we had only one system of coordinates viz. geographical longitudes in East-West direction and geographical latitudes in North-South direction. However in the case of heavenly bodies like planets etc., we have two different systems of coordinates. The first system of coordinates i.e. celestial longitudes and latitudes is normally followed by the Astrologers where as the other system i.e. Right Ascension and Declination is followed by the Astronomers. In spite of there being two systems, the methodology adopted is by and large the same as is applicable to geographical longitudes and latitudes except that in the two systems mentioned above in context of celestial sphere the measurements along the Ecliptic as well as the celestial equator are unidirectional unlike towards east or west in the case of earth's equator.

2.12 Secondly in the first system which is followed b . the Astrologers (celestial longitudes and celestial latitudes the measurements are along and perpendicular to the Ecliptic where as in the second system adopted by Astronomers i. e. Right Ascension and declination, the measurements are along and perpendicular to the celestial equator. The figure give below will make the point clear to the students.

Figure 10 of book

Students may please see the figure 10 carefully. It depi a cosmic sphere or a celestial sphere on the surface of which a heavenly body e.g. Planet ‘P’ is situated. The other constituents of the figure are:

(a) WOE	Celestial equator
(b) W_1OE_1	Ecliptic
(c) N and S	North and South poles of t celestial sphere or of the celestial equator.
(d) N_1 and S_1	North and South Poles of Ecliptic.
(e) NPP_2S	One side of the great circle (or the hour circle or circle of declination) passing

	through the planet P, and Poles N and S, perpendicular to the celestial equator and· intersecting it at point P_2
(f) $N_1PP_1S_1$	One side of the great circle passing through the Planet P Poles $N_1$and S_1 of the Ecliptic and perpendicular to it (Ecliptic) and intersecting the ecliptic at P_1
(f) OP_1	Is the angular distance measured along the ecliptic or we can call this as Celestial longitude of planet P.
(h) PP_1	is the angular distance measured perpendicular to the ecliptic. It is the celestial latitude of Planet P. As P is situated above the ecliptic in the Northern hemisphere the celestial latitude will be North.
(i) OP_2	it is the angular distance measured along the celestial equator. It is therefore right ascension of planet P.
(j) PP_2	It is the angular distance measured perpendicular to the celestial equator. It is therefore declination of Planet P. As the planet P is situated In the northern hemisphere, the declination 'of the planet will be North.

2.13 Oblique Ascension (Rashimaan)

Rashimaan means the rising periods of each of the twelve rashis or signs of the Zodiac. It is the time required taken by each Rashi to rise completely through its 30 degrees on the eastern horizon of any place on earth. The Rashimaans vary from Akshamsha to Akshamsha. These are always given in Sayana system i. e. to say that the time of oblique ascension is computed for the signs of moveable Zodiac. For the present studens may remember the definition only. We will revert back to the subject when dealing with the traditional method of casting the horoscope.

CHAPTER 3

ASTROLOGICAL TERMINOLOGY III

3.1 Equinoctial Points

In the celestial sphere the ecliptic intersects the equator at two points because the Sun crosses the celestial equator twice a year. These two points are known as Equinoctial points or Equinoxes, because when Sun is at either of these two points, the duration of day and night on earth will be equal.

3.1.1 When the sun crosses the celestial equator on its way from southern hemisphere to northern hemisphere that equinoctial point is known as the Spring or Vernal equinox. At that time the 'Sayana' longitude of the sun is 0°. Similarly when the Sun, continuing its sojourn of the heavens again crosses the celestial Equator on its way from northern hemisphere to southern hemisphere, that Equinoctial point is known as Autumnal Equinox. At that moment the Sayana longitude of the sun is 180°. Now-a-days the Sun is at Vernal Equinox on or around 21st March and at Autumnal Equinox on 23rd September, each year.

3.2 Precession of the Equinoxes

It has been observed and can be proved mathematically that the vernal equinox or the first point of Aries from where the longitudes (Sayana or Tropical) of all the planets are measured along the ecliptic is not a fixed point. Each year when the Sun reaches the vernal equinox, the position of earth with reference to some fixed star shifts by 50*1/3*” of arc west wards compared to position at equinoctial moment of the previous year. Therefore the vernal (Equinox (VE.) point is receding back along the ecliptic at the rate of about 50*1/3*” per year. This slight retrograde motion of the equinoxes is known as the precession of the equinoxes.

3.3 Moveable and fixed Zodiacs

Due to precession of equinoxes, the V.E. slips backwards from its original position (recognised as star Revati by the Hindus). The Zodiac which reckons the first degree of Aries (Mesha) from the VE. (which has a precession every year) is known as the Moveable Zodiac, while in the case of the Fixed Zodiac, the first degree of Aries (Mesha) is reckoned from a particular star in the Revati group of stars, which is fixed.

3.4 Ayanamsha

The angular distance measured along the Ecliptic, between the first point from where the fixed Zodiac commences and the VE. point, at an Epoch is known as Ayanamsha. The exact period when the point of beginning of both the Zodiacs i. e. the Moveable Zodiac and the Fixed Zodiac coincided, is not known. Accordingly the Ayanamsha or the precessional distance varies from 19^0 to 24^0. A number of dates are given as the year of coincidence. However, we follow the year as given by N.C. Lahiri in his Ephemeries i.e. 285 AD. which is same as adopted by the astronomical observatories all over India.

3.5 The Sayana and Nirayana Systems

The system of Astronomy which recognises the Moveable Zodiac, belongs to Sayana school while that which considers the fixed zodiac is termed as the Nirayana system. The Sayana system is used by the western astrologers where as the Hindu Astrologers use the Nirayana system. There are different names prevalent for the two types of Zodiacs which are as follows:

(a) Moveable Zodiac is also known as Tropical Zodiac, Sayana Zodiac and the Zodiac of Signs.

(b) Fixed Zodiac is also known as Sidereal Zodiac, Nirayana Zodiac and the Zodiac of Constellations.

3.6 Determination of Approximate Ayanamsha

Though we must always use the Ayanamsha for any Epoch as given in the Ephemeries of N.C. Lahiri, we can roughly work out the value of Ayanamsha by following the method given below:

(a) Subtract 285 from the year of the birth or the given year (A.D.)
(b) Multiply the remainder by 50 X " and reduce the product into degrees, minutes and seconds.

Example: Ayanamsha for the year 2002 can be worked out as follows :
2002 - 285 = 1717
Multiply by 50 1/3" = 1717 x 50 1/3" = or 86422.33"
Approximate Ayanamsha = 24°-0'-22".33

3.7 Sign

The zodiac consists of 360 degrees. This is divided into 12 equal sectors each of 3 0° and each sector is called a 'Sign' or a 'Rashi'. The twelve signs/rashis of the zodiac their Names, Lords, etc. are as follows:

No.	Sign	Rashi	Extent	Lord of sign
1.	ARIES	MESHA	0° to 30°	MARS
2.	TAURUS	VRISHA	30° to 60°	VENUS
3.	GEMINI	MITHUNA	60° to 90°	MERCURY
4.	CANCER	KARKA	90° to 120°	MOON
5.	LEO	SIMHA	120° to 150°	SUN
6.	VIRGO	KANYA	150° to 180°	MERCURY
7.	LIBRA	TULA	180° to 210°	VENUS
8.	SCORPIO	VRISCHIKA	210° to 240°	MARS
9.	SAGITTARIUS	DHANU	240° to 270°	JUPITER
10.	CAPRICORN	MAKARA	270° to 300°	SATURN
11.	AQUARIUS	KUMBHA	300° to 330°	SATURN
12.	PISCES	MEENA	330° to 360°	JUPITER

3.8 Nakshatras or Stars/Constellations

The Hindus have yet another division of the Zodiac. In this system the entire zodiac is divided into 27 equal parts of i3°20' each. These divisions are called Nakshatras or stars or Constellations or Asterisms. Their names, extension in the zodiac, their lords etc. are as follows :

S.No.	Name of Nakshatra	Extent (Longitude)	Extent Sign/Rashi	Lord of Nakshatra / Constellations	No. of Years in Vimshottari Dasha
1.	Ashwini	0° to 13°20'	Mesha 0° to Mesha 13°20'	KETU	7
2.	Bharani	13°20' to 26°40'	Mesha 13°20' to Mesha 26°40'	VENUS	20
3.	Krittika	26°40' to 40°	Mesha 26°40' to Vrisha 10°	SUN	6
4.	Rohini	40° to 53°20'	Vrisha 10° to Vrisha 23°20'	MOON	10
5.	Mriga~hira	53°20' to 66°40'	Vrisha 23°20' to M:ithuna 6°40'	MARS	7
6.	Ardra	66°40' to 80°	Mithuna 6°40' to Mithuna 20°	RAHU	18
7.	Punarvasu	80° to 93°20'	Mithuna 20° to	JUPITER	16

			Karka 3°20'		
8.	Pushya	93°20' to 106°40'	Karka 3°20' to Karka 16°40'	SATURN	19
9.	Ashlesha	106°40' to 120°	Karka 16°40' to Karka 30° or Simha0o	MERCURY	17
				TOTAL	120
10.	Magha	120° to 133°20'	Simha 0° to Simha 13°20'	KETU	7
11.	Poorva Phalguni	133°20'10 146°40'	Simha 13°20' to Simha 26°40'	VENUS	20
12.	Unra Phalguni	146°40' to 160°	Simha 26°40' to Kanya 10°	SUN	6
13.	Hasta	160° to 173°20'	Kanya 10° to Kanya 23°20'	MOON	10
14.	Chitra	173°20'10 186°40'	Kanya 23°20' to Tula 6°40'	MARS	7
15.	Swati	186°40' to 200°	Tula 6°40' to Tula 20°	RAHU	18
16.	Vishakha	200° to 213 °20'	Tula 20° to Vishchika 3°20'	JUPITER	16
17.	Anuradha	213°20' 10 226°40'	Vrishchika3°20'to Vrishchika 16°40'	SATURN	19
18.	Jyeshtha	226° 40' to 240°	Vrishchika 16°40' to Vrishchika 30° or Dhanu 0°	MERCURY	17
				TOTAL	120
19.	Moola	240° to 253°20'	Dhanu 0° to Dhanu 13°20'	KETU	7
20.	Poorvashadha	253°20' to 266°40'	Dhanu 13°20'to Dhanu 26°40'	VENUS	20
21.	Unrashadha	266°40' to 280°	Dhanu 26°40' to Makara 10°	SUN	6
22.	Shravana	280° to 293°20'	Makara 10° to Makara 23°20'	MOON	10
23.	Dhanishtha	293°20' to 306°40'	Makara 23°20' to Kumbha 6°40'	MARS	7
24.	Shatahhisha	306°40' to 320°	Kumbha 6°40' to Kumbha 20°	RAHU	18
25.	Poorva Bhadra	320° to 333°20'	Kumbha 20° to Meena 3°20'	JUPITER	16
26.	Uttra Bhadra	333°20' to 346°40'	Meena 3°20' to Meena 16°40'	SATURN	19

27.	Revati	346°40' to 360°	Meena 16°40' to Meena 30° or Mesha 0°	MERCURY	17
				TOTAL	120

3.9 Ascendant or Lagna

The ascendant or the lagna point is the point of intersection of the ecliptic at the given time with the horizon ofthe place. In astrology it is the first house of the horoscope. This point of intersection is very important as it is considered to be the commencing point of the horoscope. The earth rotates on its axis from West to East in about 24 hours. Due to this rotatory motion the whole sky (Zodiac) appears to come up from below the horizon gradually. The Ascendant or the Lagna is the 'Rising sign' in the eastern horizon. The period of each lagna is not equal like the rashi or the sign division. As all the 12 rashis or signs must rise one after the other in a day (due to rotation of earth on its axis once a day) each rashi/sign becomes the lagna one after the other consecutively, with the passage of time. The names of the lagnas or the Ascendants are the same as that of the rashi/sign rising at any given time.

3.10 The Tenth House or M.C.

The point of intersection of the ecliptic with the meridian of the place, at any time, is the 'tenth house' for that moment. It is also known as the Mid-heaven or Medium Coeli (M.C) and also known as meridian. The longitude of this point is, as usual, measured along the ecliptic from the first point of Aries. The right ascension of the Me is the sidereal time of the moment which is often called R.A.M. C

3.11 The Table of Ascendants

This is a small book containing several tables for use in casting a horoscope. One such book commonly used by the students of astrology is by N. C Lahiri. This gives the sidereal time at Noon for the central station of India or for the standard Meridian for India (longitude 82° 30'E) for the year 1900 and for each day of the year. It also contains tables for correction to be applied for different years and different places so that we can find out the sidereal time for any day of any year for all the places on earth. Then this book gives the tables for the different Ascendants (the degree and sign of the ecliptic) rising for every 4 minutes interval of Sidereal Time for different latitudes on the earth (0° North to 60° North). In our lessons we will be making extensive use of this book. We shall therefore advise our students to purchase a copy of this book (Tables of Ascendants) for their use.

3.12 The Table of Houses

This also is a small book like Tables of Ascendants. This is available in Nirayana as well as Sayana systems. It gives the longitudes of the 10th, 11th, 12th, 1st (Ascendant), 2nd and 3rd cusps (Midpoint of a house according to Hindu system and beginning of a house according to the western system of astrology) for different latitudes and sidereal time. With the help of this book one can directly note down the longitudes of the aforesaid six houses and by adding 180° or 6 signs to those longitude, the longitudes of the remaining six houses i.e. 4th, 5th , 6th, 7th, 8th and 9 th house can be found out. However since we follow the method of house divisions as given by sage Parashara, the tables of ascendants mentioned in para 3.11 will meet the requirements and there would be no need to buy the table of houses.

3.13 The Ephemeris

This is nothing but a modern Panchang. It is available in book form. It tabulates the positions of celestial objects (planets etc.) in an orderly sequence for one complete year (for the current years) . condensed Ephemeris are available for the past years. In these lessons we shall be following the Indian ephemeris of N.C. Lahiri, published by Astro Research Bureau, Calcutta. Students are therefore advised to have with them complete set of these ephemeris.

3.14 We have so far discussed about the Astrological Terminology commonly used in Mathematical Astrology. However in addition to the Terms discussed so far we will come across some more terms which will be discussed at the appropriate place, as we proceed further with our lessons.

CHEPTER -4

Time Measurement

4.1 The time is measured in hours, minutes, seconds and fractions of second in the western system which is now commonly used in India and other countries of the world. However the division of time is peculiar to 'Hindus'. It begins with a Tatpara and ends in a kalpa. A kalpa is equal to 4,320,000,000 sidereal years. The Hindu day (an apparent solar day) begins from Sunrise and ends with the next sunrise when the next day begins.

4.2 The Hindu division of time is as under:
60 Tatpara = 1 Para
60 Para = 1 Vilipta
60 Vilipta = 1 Lipta (or Vilipta)
60 Lipta = 1 Vighati (or Pala)
60 Vighati = 1 Ghati = 24 minutes
60 Ghatis = 1 Day = 24 hours

4.3 Again a Hindu measure of time is in terms of 'Yuga'.
The details are as under :

1st Yuga Sat Yuga = 17,28,000 sidereal years
2nd Yuga Treta Yuga = 12,96,000 sidereal years
3rd Yuga Dwapar Yuga = 8,64,000 sidereal years
4th Yuga Kali Yuga = 4,32,000 sidereal years
Total of 4 Yugas or 1 Mahayuga = 43,20,000 sidereal years
1 Kalpa = 1,000 Mahayuga
= 43,20,000 x 1000
= 43,20,000,000 Sidereal years. 2 Kalpa
1 day of Brahma = 2 kalpa (i.e. 1 Kalpa day and 1 kalpa night)
1 years of Brahma = 360 x 2 Kalpas
1 Year of Brahma = 720 Kalpa
Aayu of Brahma = 1,000 Brahma Varsha.

4.4 Sidereal Day

This is the time taken by the earth to rotate once on its axis with reference to any fixed star. This is known as Nakshatra Dina among the Hindus and is equal to 23 hrs. 56 min. (approx.) of 'Mean solar day'.

4.5 Sidereal Year

The Sidereal year is the mean period of revolution of the earth in its orbit with respect to the background stars (from fixed star to fixed star).

4.6 Apparent Solar Day

This is also known as 'Savana Day'. This is longer than the 'Sidereal Day' by about 4 (four) minutes. According to Surya-Sidhanta Savana day is reckoned from sunrise to next sunnse.

4.7 Mean Solar Day

This is reckoned by considering the average length of all the days in a year.

4.8 Months

There are two types of months in vogue in Hindus which are as follows :

(a) Lunar Month or Chandra Maan : It has 30 lunar days or Tithis and is measured from New Moon to next new Moon. At some other places it is measured from Full Moon to next Full Moon.
(h) Solar Month or Saur Maan : It is the time the Sun takes to move in one sign and is measured from one Sankranti to the next Sankranti.

4.9 Years

In Hindus there are three types of different years in vogue which are as follows.
(a) The Savana year: It has 360 mean solar days
. (h) The Lunar year : It has 354 mean solar days
(c) The Nakshatra year: It has 324 mean solar days

4.10 Tropical Year

The Tropical Year or the year of seasons, is the time a the passage of the sun from one Vernal Equinox to the next Vernal Equinox. The VE. point slips to the west at the rate 0 50 1/3 per year.

4.11 Anomalistic Year

The anomalistic year is the mean interval between successive passages of the earth through perihelion. Perihelion is the point on a planetary orbit (in this case earth) when it ; at the least distance from the Sun.
4.12 The lengths of different years mentioned in pa: 4.5,4.10 and 4.11 above, according to modern calculation given by Dr. B. V Raman in his book A Manual of Hindu Astrology) are as follows:

Year	Length			
	D	H	M	S
The Tropical year	365	5	48	45.6
The Sidereal Year	365	6	9	9.7
The Anomalistic Year	365	6	13	48

CHAPTERS- 5

TIME DIFFERENCES

5.1 Students are aware that the Sun is the creator of time, day and night and the seasons. A Hindu day commences from the sunrise and remains in force till the next sunrise, when the next day commences. When the sun is exactly overhead it is called Mid day or Local noon. At the moment of sunrise for any place, the local time for that place is Zero hour (or Ghati) as per traditional Hindu system of reckoning the time. However as the earth is not a flat body but spherical and also rotating on its axis, the Sun rises at different times at different places. As the rotation of the earth on its own axis is from west to east, it is evident that the eastern part of the earth will see the Sun first, and due to the rotation of the earth, further western parts of the earth moves towards east gradually and see the Sun. This process goes on and on. In other words, as we live on this planet earth we do not see or feel the rotation of the earth from west to east, but we see that the Sun rises in the east and gradually comes over head and then sets in the west.

5.2 Local Time

We have seen above that the eastern parts of the earth will see the Sun first and subsequently as more and more western parts move to east due to rotation of earth, those parts will also gradually see the Sun. In other words it means that the Sun will rise later at a particular place as compared to a place towards east of the earlier place. It therefore implies that Zero hour of the day will commence earlier at a place which is in the east of another place where the Zero hour of the day will commence later. Similarly the Noon time or the Mid day will occur earlier in the eastern part Of the earth as compared to any place towards west of the earlier place. We know that earth complete one full rotation (360°) on, its axis in about 24 hours or 24 x 60 = 1440 minutes. It simply means that Earth will take about 1440/360 = 4 min. to rotate by 1 ° on its axis. We can therefore conclude that Zero hour at a place' B' which is 1 ° towards west of place 'A' will commence later by 4 minutes as compared to place 'A'. So the local time differs from place to place 'A". Strictly speaking as neither the earth is a perfect sphere nor its orbit around the Sun is a perfect circle and as also the axis of earth is inclined by about 231/2 ° to the perpendicular to the plane of earth's orbit, even the duration of time or the rate of elapsing of local time is not uniform for the same place. In order to have a uniform rate of time lapse' and also to avoid complex mathematical computations, a more convenient term has been adopted for Astrological purposes which is known as 'Local Mean Time' (or a particular place. The local time or more accurately the local meantime (LMT) which is created by the gradual rising of the Sun and the roundness and rotation of the earth is the real or natural time of a place. This differs from place to place and is dependent on the longitude and latitude of the place. In Astrology we reduce every given time into Local Mean Time first and then proceed further.

5.3 Standard time

As explained above, the local time differs from place to place. This becomes quite inconvenient when we have to refer. to time at a broader perspective say National or International level. With the advent of the postal department and later the railways etc., this difficulty increased in india as well as elsewhere. A new version of time was therefore decided upon and was called the 'Standard Time' to have the uniformity of time which was the L.M.T. of a centrally located longitude in that country. In case of India, it was in the year 1906 AD when it was decided that the terrestrial longitude 82°30'E will be taken as the Standard Meridian of India for this purpose and the Local Mean Time (L.M.T.) at this Meridian (82°30' East longitude) will be the 'Indian Standard Time'. By this arrangement, the watches began to show a uniform time through out the country.

5.4 In the case of bigger countries like U.S.A., the entire country was divided into 4 or 5 convenient zones and for each zone a standard meridian representing that zone was chosen as standard meridian for that zone. The L.M. T. at that particular meridian was considered to be the Zonal Standard Time (ZST) applicable for that zone only. Students are advised to see the last page of their Tables of Ascendants by N.C. Lahiri, where in the 5 different time zones have been indicated for the U. S.A.

5.5 Greenwich Mean Time

As in the case of a country, the necessity to have a uniform time, generated the concept of 'Standard Time' for a country, in the international affairs, it was considered necessary to have a standard time which could be referred to by all the countries of the world whenever needed. Students will recall that the meridian passing through the Greenwich (near London) has been choosen as the 0° longitude or the reference point for reckoning the terrestrial longitudes of all places on earth. Therefore the local time at the meridian of Greenwich at any epoch (moment) is known as Greenwich Meantime (G.M.T.).

5.6 Conversion of Time

Students will recall that while discussing the local time vide Para 5.2 above, we worked out that a difference of 1 ° in the geographical longitudes of two adjacent places on earth will make a difference of 4 minutes in their local time. Therefore by using this principle, we can convert any given time in one form to the other two forms viz if the time given is in 1ST, we can convert it to the LMT or GMT as desired or required, if the geographical (terrestrial) longitude of the place is known. Similarly, we can always convert any meridian of longitude from degrees to time and find out the time zone for any country if the longitude of the standard meridian for that country is known.

5.7 Time Zone

Time Zone is nothing but the standard meridian for any country (or any zone within a big country) expressed in terms of hours and minutes difference with reference to the G.M.T. In other words the difference between standard Time of any country or zonal standard time of any zone and the GMT is known as the time zone for that country or zone. The time zone is (+) or (-) according to the longitude of standard meridian for the country or zone being towards the east or west of Greenwich.

Example 1 : The geographical longitude of the standard meridian for India for IST =.82°30' East
Multiplying this by 4 we get (as 1° is equal to 4 minutes, 1' will be equal to 4 sec.)
= 82° x 4 min : 30' x 4 sec
=328 min : 120 secs
= 330 min. =5 hrs 30 minutes
Since India is towards east of Greenwich the time zone for India will be = (+) 5h 30m which means that IST will always be ahead of GMT by 5^h30^m So if the GMT is 6:00 AM, the Equivalent IST will be $6^h:00^m + 5^h:30^m = 11\ h\ 30^m$ AM.

5.8 Similarly, if the time zone of a country is known, we can work out the longitude of the Standard Meridian for that country.

Example 2 : Find out the longitude of the Standard Meridian for Mexico if it falls in Time Zone (-) 6 hrs.
(a) Since the Time Zone is (-) hence the longitude of the place will be towards "west" of Greenwich
(b) Time Zone = 6 hours or 6 x 60
    6 x 60 = 360 mins (As 4 mins = 1 °)
    So 360 mm= 360/4 = 90°
Therefore, longitude of Standard Meridian for Mexico = 90°W

5.8 L.M.T. Correction

This is duration of time or a measure of time to be applied to find L.M.T. of a place from the standard time of that country or zone shown by the watch. This is obtained by multiplying the difference of longitude of the place from the Standard Meridian by 4 minutes per degree. The L.M.T correction is either (+) or (-) depending on whether the place is towards East (E) or West (W) from the standard meridian for that country. In order to find out the L.M.T. correction, one should proceed as follows :

Step 1: Find out the longitude of Standard Meridian for the country or zone as the case may be. If it is not readily available, note down the time zone from the list of cities given by N.C.Lahiri in his hook Tables of Ascendants or Indian Ephemeris and work out the longitude as explained in Examples 2 above.

Step 2: Note down the longitude of the place. for which the L.M. T. or L.M. T. correction is desired.
Step 3: Also note whether the longitude of the place is towards east or west of the Standard Meridian.
Step 4: Find out the difference in degrees (or the angular difference between the two long).
Step 5: Multiply the difference (in step 4 above) by 4'. The product will give you the L.M.T. correction. (Apply correction at the rate of 4' per degree to find out LMT).
Step 6: Prefix the sign (+) or (-) to the L.M. T. correction depending whether the place is towards 'E' or 'W' of the Standard meridian.
Step 7 : Apply the L.M.T. correction to the ST (Standard Time) given to get the L.M.T.

5.9 If the L.M.T. is given and it is desired to find the Standard Time, then reverse the sign (+ or -) prefixed to the L.M.T. correction worked out in step 5 and 6 above and then apply the correction. The Examples given below will clarify the above steps.

Example 3 : If the time of birth of a native born at Delhi is 11:30 AM (IST) find out the Local Mean Time of birth of native.

Solution:

Step 1: Long. of standard meridian of India
For IST = 82°30' E
Step 2: Long. of place of Birth
i. e. Delhi = 77° 13' E
Step 3: By persual of the two long. it is clear that Delhi is towards west of the standard meridian of India.
Step 4: Difference in the two long. = 82°30'-77°13' = 5°17'
Step 5: Multiplying by 4 we get = (5x4)+ (17^{x}4) = 20^{min} 68^{sec}

Step 6: Keeping in view step 3, the sign to be prefixed *will* be (-).
Hence L.M.T. corr = (-) 21^{m} 8^{sec}

Step 7: Applying the L.M.T. correction we get the L.M.T. of birth
11 hrs 30 mins (-) 21 mins 8 secs. = 11^{h} 8^{m} 52^{s}

Example 4 : Find out the L.M. T. correction in example 2 given earlier.

Example 2:
Long. of standard meridian = 90^{0}W
Long. of Mexico (from the tables)= 99°01'W

Hence Mexico is further west of standard meridian
Difference in the two long. = 99°01'-90° = 9° 1'
Multiplying by 4 we get = 36^{m} 4^{sec}
Hence L.M.T. correction = (-) 36^{m} 4^{sec}

5.10 However we have given the above examples so that the students can practice and understand the principle thoroughly. At times the students may come across a city or place which does not find a place in the aforesaid lists given by N. C.Lahiri. In such cases though the geographical longitude and latitude can be known from any standard Atlas, unless the students are well conversant with the underlying principles, they may find it difficult to work out the L.M.T. corrections.

Example 5 : If a child is born at 11:30 AM (LMT) at New Delhi, what will be the:
(a) ZST of Birth
(b) GMT of Birth
(c) IST of Birth
Solution (a) ZST of Birth
Step 1: Time Zone of New Delhi = (+) 5 hrs 30 min.
or = 5 x 60+30 = 330^{min}
Divide by 4 we get = $82^0 30'$
Since it is(+), so the Standard Meridian has to be towards ‘E' of Greenwich. Therefore Long. of Standard Meridian = $82^0 30'$ E
Step 2: Long. of New Delhi (from the tables) = 77°13' E

Step 3: Hence New Delhi is towards west of its Standard Meridian. So the LMT correction will be (-).

Step 4: Difference in the two long. = $82^0 30'$ - 77°13' = 5° 17'

Step 5: Multiply by 4, we get = $20^{min}\ 68^{sec}$
or = $21^{min}\ 08^{sec}$
Step 6: Hence L.M.T. correction= (-) $21^{m}\ 08^{sec}$

Now since LMT of Birth is given and we have to find the ZST of Birth, so in accordance with Para 5.9 above, we have to reverse the sign and then apply the correction. Hence,

ZST of Birth = 11:30 (AM)(+) $21^{min}\ 08^{sec}$
or Time of Birth = 11^{hrs}: 51^{m}: 08^{S} (AM) ZST

(b) GMT of Birth : As the time zone for Los Angels is (+) 5 hrs 30 min , hence the ZST will be more (advance) by 5 hrs 30 min from GMT .

= 11hrs:51m:08sec (-) 5 hrs 30 min ^{s}
= 06^{hrs}:21^{m}: 08^{sec}
or 6hrs:21m:08sec (AM)

(c) IST of Birth: As the Time Zone of India is (+) 5 Hours 30 Min. the IST will be ahead (more) of GMT by 5 Hrs. 30 Min. the IST of Birth will ahead (more) of GMT by 5 Hrs. 30 Min.
So IST of Birth = GMT by (+) $5^{h}\ 30^{m}$
= 6h:21m:08sec (+) 5^{h}:30^{m}
= 11 h:51^{m}:08^{sec} (AM)

CHAPTER 6

SIDEREAL TIME

6.1 We have seen in the previous lesson the three different systems of measuring time viz Local Time or Local Mean Time (LMT), Standard Time for any country or Zone *(i.e.* 1St, ZST etc.,) and the Greenwich Mean Time (GMT). There is yet another system/measure of time which is called 'Sidereal Time'. **The Sidereal Time system is derived from the earth's rotation with respect to the stars.** The students will recall that while discussing the Time Measures vide Lesson 4, we defined the **Sidereal day as the time taken by the earth to rotate once on its axis with reference to any fixed star.** The duration of this sidereal day is equal to 23 hrs 56 min (approximately) or 23 hrs 56 min 4.091 sec. more precisely, of mean solar day. **In astronomical terminology,** the sidereal time at any instant is defined to be the **west hour angle of the Vernal Equinoctial (VE) point** or the **first point of Sayana Aries (Mesha) from the upper meridian** of the place. However for the purpose of its application to mathematical astrology, it will suffice to define the Sidereal Time as the Local Time reckoned according to the apparent rotation of the celestial sphere. **In** other words, whenever the time is reckoned with reference to the sidereal day, it is called Sidereal Time. The Sidereal Time is 'Zero' hour when the first ponit of Aries or Mesha (in sayana system) *i.e.* vernal or the spring equinox crosses the observer's meridian (which is the great circle on the celestial sphere, passing through the zenith and both the celestial poles).

6.2 Necessity to have the Sidereal Time System

Students may be aware that for any astrological delineation, the horoscope prepared for a particular epoch (moment) is not only a necessity but the only astrological equipment available to the astrologer based on which he analyses the shape of things to come in the future. The horoscope which is a map of heavens at the given moment, contains 12 houses and the commencement of the horoscope is the 'first house' or the 'lagna' or the 'ascendant'. It is therefore most important to calculate the correct lagna or the ascendant without which no horoscope can be prepared. Students may now recall that while discussing about the ascendant or lagna vide Para 3.9 of Lesson 3 it was stated that due to the rotatory motion of the earth from west to east on its axis, the whole of sky (or the zodiac with which an astrologer is concerned) appears to come up (or rising) from below the horizon gradually and the sign or rashi (and more particularly the exact degree of the zodiac or that sign) rising in the eastern horizon, is known as the 'lagna' or 'ascendant'. As the lagna or ascendant or the sign of zodiac rising on the eastern horizon of a place at any time, is dependent on the rotation of earth on its axis due to which the time system known as the 'Sidereal Time' is also created, so it becomes evident that the rising sign or the lagna, in turn, is dependent on the sidereal time of the place at the given moment or epoch. It therefore transpires that in order to know the lagna or the rising sign for a particular moment or epoch (be it a birth of a child or birth of a question, incident or accident etc.) it is necessary to first calculate the sidereal time of the moment at that place where the birth of a child or a question or incident has taken place. Students may please refer to the Tables of Ascendants by N.C. lahiri and see themselves that the Ascendants for the different latitudes are given with reference to the sidereal time only. We therefore now proceed to discuss the method to calculate the sidereal time of a given moment or epoch.

6.3 How to calculate the Sidereal Time of a given moment or Epoch

Students are advised to refer to the Tables of Ascendants by N.C. Lahiri (all references in this lesson pertain to the seventh edition of the book published in 1985) and proceed as follows:

Step 1: Note down the sidereal time at 12h noon local mean time for 82°30'E longitude for 1900 AD for the day and month of the given moment from Table I at page 2.

Step 2: Note the correction for the given year from Table II given on pages 3 and 4 o~ the book and apply to sidereal time in step 1.

Step 3 Note the correction for the different localities from Table III given on page 5. A detailed list of principal cities of India has been given on pages 100 to 107. The last column of the table indicates the correction to the 'Indian Sidereal Time'. Similarly the table for the foreign cities has been given on page 109 to 111 of the book and the last column of the table again indicates the correction to the Indian Sidereal Time.

Step 4: The correction for the year (step 2) and the correction for the place (step 3) should be applied to the sidereal time noted in step 1 according to the sign (+) or (-) prefixed to the correction as shown in the respective table. Having applied these corrections, the result obtained (let us call it' A') will represent the Sidereal time for the given date, year and place but will be for the local noon *i.e.* 12 hrs, as we have not yet applied the correction for the hour and minutes before or after the local noon, as the case may be, for the give moment.

Step 5 Convert the given time of epoch into LMT by applying the LMT correction. This has been discussed elaborately in great detail and explained with the help of examples also vide para 5.8 of the preceding chapter. However the quantum and the sign (+, -) of the correction to be applied to the 1ST or ZST, as the case may be, has also been indicated in the tables at pages 100 to 107 for principal cities of India under column LMT from 1ST and, at pages 109 to 111 for foreign cities under column LMT from ZST.

Step 6: As the Sidereal time noted in the step 1 pertains to the local noon, we have to find out as to how many hours before or after the local noon, is the given time of the moment or Epoch. In other words we have to find out the "Time Interval" between the Local Mean Noon (LMN) and the LMT of the given moment. So, in case the LMT of the given moment is before noon, subtract it from 12:00 hours. In case the LMT of the given moment is in the afternoon, the LMT itself becomes the Time Interval (T.I.) also because after 12 noon our watches show 1:00 PM and not 13 :00 which means that 12 hours have already been deducted.

Step 7: The Time Interval (T.I.) worked out in step 6 above is to be increased by applying the correction given in table IV which gives the correction for hours and minutes of the T.I. By applying this correction we get the Increased T.I. Let us call it (B).

Step 8: The 'Increased T.I.' (B) is added to the corrected Sidereal Time (A) in step 4 above in the case of PM (afternoon) births or epoch and, subtracted from the (A) in the case of AM (before noon) births or epoch, as the case may be. The result thus obtained is the Sidereal Time of the birth or epoch or the given moment. The above mentioned eight steps can be explained with the help of a practical example or illustration.

Example 1 : Find out the Sidereal Time of birth of a native born at Delhi on Saturday the 01st Nov. 2008 at 09:30 AM (IST)
Solution: Use Tables of Ascendants by N.C. Lahiri

Step 1 : Sidereal Time at 12h noon
on 01st November 1900 (page 3) = 14^h : 40^m : 21'
Step 2 : Correction for the year 2008
(from page 4) = (+) 03^m : 20'
Step 3 : Correction for Place (Delhi)
(page 5 as well as page 102) = (+) 0^m : 03'
Step 4 : Sid. Time on 25th Oct. 2010
at Delhi, at noon (A) = 14^h : 14^m : 14'

Step 5 : IST of birth (given) = 09 : 30 : 00
LMT correction (page 102) = (-) 21 : 08
Therefore LMT of Birth = 09 : 08 :52

Step 6: TI from noon (subtract the
LMT from 12h being AM birth)
(12 hrs - 9 hrs 8 min 52 sec) = 02 : 51 : 08
Step 7 : Correction to increase the TI from
table IV (page 5)
for 2 hrs = 00 : 20
for 51 min 8 sec = 00 : 09

Therefore Increased TI (B) = 02 : 51 : 37

Step 8 : Being AM Birth (A)-(B) = 14^h : 14^m : 14'
(-) 02^h : 51^m : 37'

Sidereal Time of Birth = 11^h : 22^m : 37'

6.4 Caution
We hope that by now the students would have understood the methodology to work out the Sidereal Time very clearly. However before we end this topic, we will like to caution our students to note carefully the few points mentioned below :

6.4.1 Unlike the civil time (LMT or GMT or 1ST or 2ST) the Sidereal Time is never expressed in terms of AM or PM. It is always starting at 'O' hour and goes upto 24 hour after which it again starts as 0^{hour} .

6.4.2 WAR TIME: From **1st Sept. 1942 to 14th Oct. 1945, the Indian Standard Time (1ST) was advanced by one hour all over India** including modern Bangla Desh and Pakistan for purposes of daylight saving during the war period and was thus ahead of GMT by 6 H 30 min. Therefore any recorded time during this period (Both days inclusive) must be reduced by 1 hour to get the corrected IST before LMT correction is applied to obtain the LMT of birth. (Provided the same correction is not made while noting down the time on the record.)

***6.4.3 SUMMER TIME* :** Students are advised to refer to page 112 of their *Tables of Ascendants* and read carefully each and every word thereof in order to acquaint themselves with the summer timings being observed in Britain, USA, Canada, Mexico, USSR and other European countries mentioned therein. The recorded time falling on the dates/ period of summer timings indicated in page 112, must therefore be corrected first as applicable, *before* it is converted to local Mean time of epoch.

CHAPTER 7

SUNRISE AND SUNSET

7.1 In the previous lesson we have seen the methodology for working out the Sidereal Time of birth or of an epoch. With this Sidereal Time we enter the relevant Table of Ascendants for the latitude of the place of birth to find out the Ascendant. However, before we proceed on to find the ascendant or lagna or the rising sign, we deal with the subject of sunrise and sunset in this lesson. The time of sunrise, sunset etc is very useful in astrological calculation to find out the dinmaan, ratrimaan (i. e. the duration of· day and night), Ishtakala or Ishtaghati which forms the basis to calculate the lagna rising by the traditional method,. Kaal haras, Kaal velaas, Hara lagna, Mandi, Rahu kaalam etc., which have great significance in the Hindu Astrology.

7.2 Sunrise

The exact moment at which the sun first appears at the eastern horizon of a place is time of sunrise. As the Sun has a definite diameter, the solar disc takes some time *i. e.* about 5 to 6 minutes to rise. Therefore, from the first visibility cif the upper limb of the solar disc to the time when the bottom limb of the solar disc is just above the horizon of the place, there will be a time diference of about 5 to 6 minutes. It has, therefore, been acknowledged that for astrological purpose we may take the moment at which the centre or the middle of the solar disc is at the eastern horizon of the place as the sunrise time for that place.

7.3 Sunset

Similarly the sunset for a particular place is the exact moment at which the centre or the middle of the solar disc is at the western horizon of the place.

7.4 Apparent Noon

This is marked when the centre of the sun or the middle of the Solar Disc is exactly on the meridian of the place. The apparent noon is almost the same for all places.

7.5 Ahas and Ratri

Ahas is the duration of day i. e. the duration of time from sunrise to sunset. Ratri is the duration of time from sunset to sunrise. On the equator, the Ahas and Ratri are always 30 ghatis or 12 hours each, while on other latitudes the sum of Ahas and Ratri will be 24 hours or 60 ghatis.

7.6 Calculation of time of Sunrise and Sunset

In this lesson we propose to calculate the time of Sunrise and Sunset by the method of 'interpolation' from the given data in the Ephemeris. However there is a proper method to calculate the time of sunrise and sunset without making any reference to the given data in the Ephemeris. We don't propose to discuss that method through this lesson' as the same is not only cumbersome but involves too much mathematical calculation needing enormous time which is not warranted being beyond the scope and purview of these lessons.

7.7 Calculation of time of Sunrise and Sunset by Method of Interpolation

Step l: As the time of sunrise or sunset differs from latitude to latitude we must first of all note the latitude for the place where the time of sunrise etc., is desired.

Step 2: Refer to page 93 and 94 of Lahiri's Indian Ephemeris for the year 2002 and select two such consecutive dates that the date for which the sunrise time is desired falls in between the two selected dates. Similarly select two such consecutive latitudes from the table at page 78 so that the latitude of our desired place falls in between the two latitudes so selected.

Step 3 : Note down the timings of sunrise or the sunset as the case may be, for the above selected dates and latitudes as given in the table.

Step 4 : Find the time of sunrise and/or sunset by interpolation (simple ratio and proportion method). The time so obtained will be the Local mean Time (LMT)· of the time of visibility of the upper limb of the solar Disc. Add 3 minutes to the time of sunrise and deduct 3 minutes from the time of sunset to get the LMT of coincidence of the centre of the solar disc with the horizon.

Step 5 : In case the time is required in terms of IST or ZST, apply LMT correction as applicable by reversing the (+) or (-) sign prefixed to the LMT correction as given in the list of table of Ascendants from Page 100 to 111.

7.8 The above method has also been indicated at page 99 of Lahiri's Indian Ephemeris for the year 2008 and students are advised to follow the same with advantage. However we also give below the illustration to explain the steps mentioned above more clearly to our students.

Example 1 : Desired IST of Sunrise and Sunset at Delhi on Oct 05.
Solution: Use page 99 of Lahiri's *Indian Ephemeris* for 2008 .
Step 1 : Latitude of Delhi (Page 150 of Ephemeris for 2008) = 28°39'N or 28.65°N
Step 2 : Dates selected are Oct 03 and Oct 07, Latitudes selected are 20^0N and 35°N
Step 3 : The data given for the above mentioned dates and latitudes at page 94 of the Ephemeris is as follows:

	Sunrise (LMT)		Sunset (LMT)	
Date	Latitudes		Latitude	
	20^0N	35°N	20^0N	35°N
Oct 03	5:58	6: 13	5:31	5: 16
Oct 31	6:01	6:20	5:26	5:07

Step 4 : We can now obtain the values for the Oct 27 by simple interpolation which are as follows :
5:11 (-)17^m
".c ()17 17
vanatlOn lor 8.65° = + $15^{X8.65}$ (-)15 x8.65
(Delhi's Lat (-) 20°); (28.65°-20°=8.65°)

9:80 min
or say = (+) 10 min
Therefore LMT of
upperlirnb visibility= 6: 09 AM
LMT for centre of

solar disc = (+) 0:03
6.12 AM
= 9:80 min. = (-) 10 min
(-)0:03 5.15 PM

Step 5 : Students may now compare this with the Time of sunrise and sunset (upper Limb) for Delhi given on page 91 of Ephemeris which is as follows for 27 Oct.

6:30 AM 1ST of Sunrise (Upper limb)
Deducting 21^m 0:21

6:09 AM LMT of Sunrise . (Upper limb)
5:40 PM 1ST of Sunset (Upper limb)
0:21
5:19 PM LMT of Sunset (Upper limb)

Step 6 : Solar Disc
correction (+) 0 : 03 (-) 0 : 03
LMT for Center 6 : 12 AM 5 : 16 PM
of Solar Disc
which agrees with that worked out in Step 4.

CHAPTER 8 **CASTING OF HOROSCOPE I MODERN AND TRADITIONAL METHOD**

8.1 The horoscope is a map of heavens for a given moment at a particular place. It indicates the sign of Zodiac rising on the eastern horizon of the place at the given moment which is known as the lagna or the Ascendant. It is also known as the first house and the successive Rashis/ signs becomes the successive houses or Bhavas (as called in Hindu Astrology). Apart from the lagna or the Ascendant this map also indicates position of various Rashis and Planets at the given moment/epoch.

8.2 Forms of Horoscope

There are many types/forms presently in vogue in different parts of India as well as in the European countries. For the reference of students we give here some of the most commonly used formats by Astrologers in India and abroad. Students are advised to make themselves familiar with these 'Formats', though they may follow anyone of these appealing to be the most convenient :

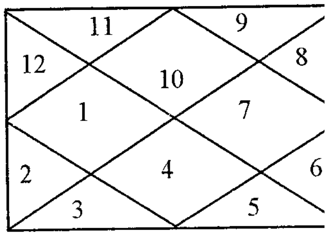

Pisces	Aries	Taurus	Gemini
Meena	Mesha	Vrisha	Milbuna

8.2.1 *TYPE 1* : This is the format which is commonly used in NorthINorth-west part of India. The top middle portion is always treated as the lagna or Asc or the I House and the number of the Rashi/sign rising at the moment of birth on the eastern horizon of the place is indicated here e.g. 10 for makar or Capricorn. Then the counting of houses is done anti-clockwise. So the II house will have the sign! Rashi next to Capricon (Makar) *i.e.* Kumbha or Aquarius written there as NO.1I. The number of the successive Rashil sign is then written consecutively one after the other in the succeeding houses, anticlockwise. Then the position of the planets at the moment is worked out and posted in the horoscope in the respective rashi/sign occupied by them in the Zodiac.

8.2.2 *TYPE II* : This type of format of Horoscope is commonly used in the Southern part of India. In this type the counting of houses is in clockwise direction. Here the position of Rashis/sign are fixed for all the horoscopes, e.g. the top left hand square in the chart represent the sign Pisces (Meena) and succeeding Squares in clockwise direction will represent Aries (Mesha), Taurus (Vrisha), Gemini (Mithuna), and so on. As this sequence of sign/Rashis is fixed for all the horoscopes, these are never written in chart. The sign/Rashis rising on the eastern horizon of the place or the lagna or Asc is marked in the appropriate sign in the chart as shown and word lagna or Asc is written in that sign. Afterwards the planets according to their position in the Zodiac at the moment are posted in the respective sign in the chart to make the map or the horoscope complete.

8.2.3 *TYPE III* : This type of chart is commonly used in Bengal and Neighbouring area. In a way it combines the two charts discussed earlier i. e. Type I and Type II in as much as the counting of houses is done anticlockwise (like Type I) but the position of Rashis/signs is fixed for all the horoscope (as in the case of Type II). The other aspects like posting the position of planets etc., are similar to other charts. The lagna is written in the appropriate sign in the chart.

Type **III** Type IV

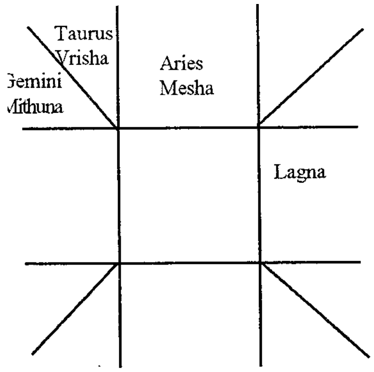

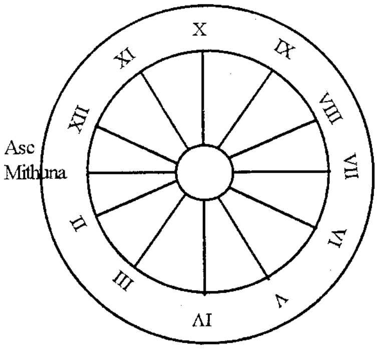

8.2.4 TYPE IV: This chart is commonly used by Western Astrologers. Nowadays some Astrologers in India particularly in Maharashtra have also started using this type of chart for the horoscopes. This is a circular chart as shown in the figure, the twelve Bhavas or houses are marked in chart and symbol along with the degree is also indicated on each bhava. In Indian (Hindu) Astrology, the cusps are treated as bhava Madhaya or the middle point of the houses where as in Western Astrology, the Asc cusp means the beginning of the first house, the II cusp means end of 1st house and beginning of the II house and so on. The planets are also shown with their symbols only and the degrees of the zodiac acquired by a planet is also written along with the Planet in the chart.

8.3 Casting of Horoscope

The process of casting of horoscope involves two main activities. Firstly we have to find by calculation the exact degree of longitude of the Ascendant or the lagna. Secondly we have to calculate the longitudes of all the nine planets or grahas mentioned earlier in chapter 1.

8.3.1 There are mainly two important methods to find out the lagna and the planetary position at the time of birth of a child or a question, event, or incident/accident. The first method is called the *modern method* by using the table of Ascendants and ephemeris. The other method is *traditional method* adopted by the Hindu astrologers where the horoscopes are prepared with the help of traditional Panchangas (almanacs, a kind of traditional ephemeris). Now a days with the advent of calculators, log tables, computers etc. comparatively more accurtae horoscopes can be prepared by using modern method. In these lessons, therefore, our emphasis will be more on to the modern method. However for the academic interest of the students we will discuss the traditional method also at the appropriate time and place. But for the present let us proceed with the modern method of casting horoscope.

8.4 Modern Method of Casting Horoscope

As already metnioned in para 8.3 above it involves or consists of two stages, viz:

(a) calculation of longitude of lagna/Asc *(b)* calculation of longitudes of planets

We will therefore take up the above two stages one by one.

8.4.1 *CALCULATION OF LONGITUDE OF LAGNA:*

We have already discussed in earlier lessons that the long. of lagna or the Ascendant is calculated by using the Tables of Ascendants which gives the Ascendants rising at different latitudes for each 4 minutes interval of Sidereal time. Accordingly the Sidereal time of Birth/epoch is very important to know the lagna/ Ascendant. In lesson 7 we have discussed at length how to find out the sidereal time of birth/epoch and we hope that by now our students are well conversant with the calculation of sidereal time of the epoch. We will now advise our students to proceed as follows to calculate the longitude of the lagna or the Ascendant:

Step 1: Calculate the Sidereal time of birth/epoch by following the 8 steps given in chapter 7.

Step 2: In the book Table (i.e. Ascendants by N.C.Lahiri, locate the page where Ascendants for the appropriate latitude i. e. the latitude of the place of the Birth are given. In case table for exact latitude is not available, then the other table for the latitude which is nearest to the latitude of the place of birth could be made use of. In case a more precise work is needed, the students may find out/calculate the Ascendant at two consecutive latitudes falling either side of the given latitude & then find out the exact longitude by interpolation of the two Ascendants. However we feel that in most of cases the calculation of Ascendant for the nearest latitude may serve the purposes and the interpolation may not be necessary.

Step 3: Calculate the Ascendant/lagna with the help of the appropriate Table.

Step 4: As the table of Ascendants by N. C. Lahiri gives the Nirayana longitudes of Ascendants for the year 1938, it is necessary to apply the Ayanamsha correction as given at Page 6 of the book to get the correct lagna. The above steps can be best explained with the help of an example.

Example 1 : Calculate the long of Ascendant/lagna for the Native of Example 1 in Para 6.3.

Solution: Referring to the example 1 of chapter 6 we get:

Step 1: Sidereal Time of Birth = 11 hrs 22mts 22secs.

<\sc for Delhi have been given at page 48. So vVe use the table given at page 48. (Also available it page 134-35 of Indian Ephemeris for 2002).

fhe Ascendant/lagna is calculated as follows : :Refer Page 134 of Ephemeris for 2002)

Sidereal Time	Long. of Ascendant
11 h 22^m oosec	7^5 16° 30'
0 0 22^{see}	0 0 04'

F or additional 22 sec of Sidereal Time increase will be = 12 -;- 60 x 22 = 4'

... for 11hrs 22mts 22secs= 7^s 16° 34'.

Ayanamsha correction for the year 2002 (Refer Page 135 of Ephemeris for 2002) = (-) 0° 54'

Therefore correct lagna/ Asc = 7^s-15° -40'

or Scorpio 15°-40'

(As 7 signs *i.e.* upto Libra already passed)

IIe 2 : Calculate the Asc of lagna for the native NO.2 in para 6.3.

Sid. Time of Birth = 11-49-15

The latitude of New York is 40° 43'N

(This can be noted from the table given at page 111). An appropriate table giving the longitude (nearest latitude 41 °_0' North) is given at page 62. So we use this table to calculate the Asc.

Calculate the lagna or Asc as follows :

Sidereal Time	Ascendant/lagna
11 h 48^m osee	7^s 15° 28'
11^h 52^m osee	7^s 16° 17'

variation in 4 minute = 49'

Imts 15secs (or 75 sees) = - x75 240

=15' (Appx.)

= 7^5 15° 28' + 15'

Step 4: Ayanamsha correction
for the year 2002 = -54'

Therefore correct Lagna = 7^5 14° 49' or Ascendant is Scorpio 14° 49'

Question : Students may please choose the places situated in northern Hemisphere out of the 10 places given in Question of Exercise 6 and work out the longitude of lagna/ Asc in all those cases.

CHAPTER 9

CASTING OF HOROSCOPE II

9.1 Calculation of Ascendant for places situated in Southern Hemisphere (or the Southern Latitudes)

The methodology for calculation of lagna/ Ascendant for places located in Southern hemisphere/southern latitude is exactly similar as for Northern latitude, if we have with us Tables of Ascendants for Southern Latitudes. The Lahiri's tables available to us are for Northern latitude. If the same tables are to be used for calculating the Lagna rising in places situated in the Southern latitudes, it is but obvious that some modification is definitely called for. As such for calculating the Lagna in Southern Latitude with the help of Tables for Northern Latitude, we have to proceed as follows:

Step 1: Find out the Sidereal Time of Birth by following the eight steps, 1 to 8 given in chapter : 6 as done in the case of Northern Latitude.

Step 2: Add 12 hours to the Sidereal time worked out in step 1. If the total Sidereal Time after adding 12 Hours exceeds 24 hrs., then subtract 24 Hours from it, and retain the remainder. The Sidereal Time so modified will be called as modified Sidereal Time.

Step 3 Locate the appropriate table for the Latitude of the place of birth in the Tables of Ascendants for Northern Latitudes.

Step 4: By using the Modified Sidereal Time worked out in step 2 above, calculate the Ascendant in the similar way as in chapter 8 using the Table located in step 3.

Step 5: Apply Ayanamsha correction (Page-6) for the appropriate year, i.e. the year of birth.

::,tep 6: Add 6 Signs to the Ascendant Calculated/worked out in step 5 to get the correct Lagna. If the Asc. exceeds 12 signs then subtract 12 signs from it.

9.2 Students may please note that modification incoported above is applied only for the places in Southern Latitudes if the *Tables of Ascendants* used is for Northern Latitudes and vice-versa. If the *Tables of Ascendants* are available for the same hemisphere in which the birth has taken place, no modification is necessary. Students are also ~dvised to read the Example 3 given in the Tables of Ascendants for Northern Latitudes by N.C. Lahiri, at page - *(viii)* in the beginning of the book. We will now explain the abOve mentioned 6 steps with the help of an example.

Example 1 : Calculate the lagna for the native of example no. 3 in chapter 6. (DaB 17-8-2002 TaB 15-25 hrs. 2ST)

Solution: Place Sydney, Latitude: 33° 52' South Step 1 : Sid. Time of Birth = 13^h 11^m 32^{5ec}

Step 2 : By adding 12h we get the modified Sidereal Time as 25^h 11 m 32^{5ec}. As it exceeds 24^h, deduct 24 hrs. Therefore, Modified Sidereal Time = I^h 11^m 32^{5ec}.

Step 3 : Latitude of Place of Birth is 33°52'S. Hence use the Table for 34°0'N (Page-55)

Step 4 : The Lagna is calculated as under :

Sidereal Time	Lagna @
$I^h 8^m 0^5$	3^5 6° 18'
$I^h 12^m 0^5$	3^5 7° 8'

@ To be corrected

Variation in 4 Mins. = 50'

(or in 240 Secs)

Therefore, variation in 212 Sec. = 50+240x 212 = 44.17'

or Say = 44'

Hence Lagna for I^h llm 32^{5ec} = $3^5$6°18' + 44'

= $3^5$7°02'

Step.5 Apply Ayanamsha Correction for 2002 = (--)0°54' . Corrected Lagna in North Latitude $3^5$7°02'(-)54' = $3^5$6°08'

Step 6 : Add 6 signs to get the lagna in Southern Latitude = $+6^5$

= $9^5$6°8'

Therefore, Lagna or Asc Capricon 6°8'

Example 2 : Calculate the lagna for the native born on 14th November 2001 at 4hrs 48mts (ZST) in Lima (Peru).

Solution: Refer N.C. Lahiri's Table of Ascendant at page llO and note birth place *i.e.* Lima (Peru) and latitudes, longitudes, time corrections etc.

Time Zone (-) 5 hours. Latitude Longitude 77°-02' west L.M.T. -08 sec.

12°-02' south fromZST(-)8 min

I.S.T. from Z.S.T. + 1 Oh-3 om, correction to Indian Sidereal Time (+) Imt. 45 secs.

Step 5: ZST of the birth of native
Step 6: LMf Local Tmie Correction (page-110) Step 7: L.M.T. of birth
Step 8: As it is fore noon birth
1.1. from noon (12 hours(-) $4^{h}39^{m}52^{sec}$) = 7: 20 : 08
Correction to increase the 1.1. (Page 5)=(+) 01 : 12
Hence the increase 1.1. (B) 7 : 21 :20
Being A.M. birth (A)-(B) 8 : 14 : 10
(15 : 35 : 30 (-) 7 : 21 : 20) = 8: 14 : 10
Step 12: The Sidereal time of birth 8 : 14 : 10
Step 13: The Latitude indicates the birth place is in southern
Hemisphere. But the Lahiri's Table of Acendant is for Northern latitudes. Therefore the method prescribed in para 9.1 is to be used *i.e.* add 12 hours to the sidereal time available at step 12.
Step 14: Modified Sidereal Time 12hrs + 8^{hrs} 14^{mts} lOsee
Step 3 :
Step 4 :
Step 9 :
Step 10:
Step 11:
Sidereal Time of 14 Noy. 1900
12 noon at Longitude 82°30'East (Page 3) =
Correction for the)'ear 2001
(page 4)
Correction to 1. Sid. Time (P-ll 0) Sid. Time of 14th Nov. 2001 of Peru
at 12 noon (A)
=(+) 0 : 2 : 08 =(+) 0 : 01 : 45
=(-) 0 : 8 : 08 := **4: 39 :.52**
Step 15: Calculate Ascendant on the basis of Latitude 12°_02' North (P-19 of Table of Ascendant), the table is for 12° North, which is nearest. The use is as under:

Hence Ascendant is 0115°-30' + 36' or Mesha
Aynamsha correction (P-6)
Correct Ascendant

Sidereal Time			Ascendant		
Hrs	Mts	Secs	Rasi	degree	mts
20	16	00	0	16	36
20	12	00	0	15	30
0	04	00	0	1	06'

Modified S.T. of birth = 20: 14: 10, which is more by 0:2 10 (20 : 14 : 10 (-) 20 : 12 : 00) or 130 secs Variation is 4 mins or 240 secs = 66'
'U .,. d 66 . 858 3 5 varIatIOn m 130 secon s = 240x 130 = 24 =.5.7
or =36' or *0/16°-06'* = 16° 06' 0° -,-53'
Step 16: 12° 2' is Southern Latitude, hence
add 6 signs to the above *W* + 0115°-13') = *6^{s}/15°-13' i.e.* Tula Ascendant of 15° 13'
Step 17: Hence the native born with Tula 15 ° 13' Ascendant.

EXERCISE - 9

Calculate the Ascendants for the data given below :

(a) Jakarta	21-4-1943	5:25 AM (1ST)
(b) Mombasa	11-7-1923	10:30 PM (IST)
(c) Narobi	17-8-1986	6:24 PM (2ST)
(d) Canbera	23-4-1972	4:40 PM (2ST)
(e) Sydney	15-9-1936	3:25 AM (1ST)

CHAPTER 10 CASTING OF HOROSCOPE III MODERN METHOD

10.1 Calculation of Longitudes of *Planets/Planetary Position at Birth* or *Graha Spashta*

We have already advised our students to purchase and have with them a complete set of Lahiri's Indian Ephemeris (Please refer Para 3.13 in Chapter 3). A perusal of these Ephemeris reveals that :

(a) In the yearly Ephemeris e.g. for the year 2001, 2002, 2003, the daily position of all planets including Moon has been given at 5:30 AM (IST).

(b) In the condensed Ephemeris for the year 1941-51, 1951-61, 196L-71, 1971-81, 1981-85, 1986-1990, 1991~1995 & 1996-2000 etc., daily position of Moon has been given for 5:30 AM (IST) where as for the remaining Planets except RahulKetu, the position has been given at 5:30 AM (IST) for every alternate day. Rahu's position has been given for 1 st . of each month for true as well as mean Rahu.

(c) In the Ephemeris (condensed) for the years 1900 to 1941, the position have been given for 5:30 PM (IST) daily for Moon, twice in a week *i. e.* for Sundays and Wednesdays for Mercury and weekly position *i. e.* for every Sunday in respect of Saturn, Jupiter, Mars, Sun, and Venus. Rahu's position has been given monthly *i. e.* for 1 st of each month. The Rahu's position in this Ephemeris is for 'Mean' Rahu

only and not for 'True' Rahu. *True* Rahu is by considering the actual *ova/elliptical* shape of orbit of Moon and the ecliptic, where as *Mean* Rahu is calculated by considering their orbits as perfect circle. As the later is not factually correct we prefer to have only *True* Rahu in our calculations as far as possible.

10.2 Keeping in view the above three different types of data available in the Ephemeris, we propose to discuss the calculation of planetary positions in three different parts. Accordingly we will first of all take the Ephemeris for the year 2002 and calculate the planetary position for the given time of Birth of a native. It should be noed that the Lahiri's Indian Ephemeris gives the position of Planets either for 5 :30 AM (IST) or 5:30 PM (IST). *Accordingly any time of Birth whether it is given in LMT or ZST or GMT must be first converted to IST* so *as to use these Ephemeris.*

10.3 Calculation of Planetary Position by using Yearly Ephemeris

The calculation of Planetary position is best explained with the help of an example. However before we take up an example it is necessary to advise the students that while selecting the two consecutive dates from the ephemeris for obtaining the reference position of planets, care must be taken to see that the dates should be such so that our date and time of birth falls in between the two for convenience in interpolation.

Example 1 : Calculate the planetary position at the time of birth of a native at Delhi on Sunday the' 25 Oct. 2002 at 09:30 am (IST) [Para 6.3 and para 8.4].

Solution: Use Lahiri's *Indian Ephemeris* for 2002 we will calculate the moon's position first.

MOON (Page 32)

Position at 5:30 AM (IST) on *25110/02* = 1'18°3' 58" Position at 5 :30 AM (IST) on *2611* 0/02 = 2^s (XPI9' 52"

Motion in 24 hours OSI2°15'54"

Time elapsed from 5:30 AM to 9:30 AM 4 hours

Therefore motion in 4 hours (1/6th of 24 hourly motion)

OS2°2'39"

Add position at 5:30 AM of 25-10-2002 1'18°3'58"

Position at Birth 1'20°6'37"

or Vrish 20°6'37"

or rounding off we can say position of Moon = Vrish20°7'

Note: The position at 5:30 AM is indicated in the 7th column. The first column gives the dates of the month.

10,4 The above method of finding out the proportionate motion in 4 hours is by simple arithmetic or by using an electronic calculator. We can also find out this motion by using the log tables given at page (156-57) of the Ephemeris. liowever, as these log tables are proportionate and are meant for use with 24 hourly motion.

Therefore motion of Moon in 24 hours =12° 15' 54" Log of motion *(i.e.* 12° 15' 54" or 12° 16')

(page 156) =0.2915

Log of 4 hours =0.7781

(T11lle Interval from 5:30 AM to **Birth** T11lle)

Total = 1. 0696

By taking antilog of this we will get the desired motion in 4 hours. Since there is no separate table, we have to locate the nearest figure to 1.0696 in the table and then read the degrees and minutes. We find that antilog of (nearest)

1.0720 is 2° 2'
and 1.0685 is 2° 3'
Therefore variation of 35 is equal to I' or 60"
So variation of 1.0696 (-) 1.0685 is 63~x 11 = 18.857" · ☽
Deducting 19" from 2° 3' we get the motion of Moon in 4 hours Adding this to position at 5:30 AM on 25-10-2002
Position of the Moon at Birth
or say =

10.5 By looking at the above calculation student may feel that using logarithms is rather a cumbersome process. Actually it is not so. In the above calculations we have tried to show to the students that if more precision is required we can work out the longitudes of planets upto seconds (") of arc by log table also. However in most of the cases, calculation of longitudes of planets upto nearest Minute (') of arc will suffice or meet our requirement. Therefore we need not interpolate the figures while working out the Antilog and only the nearest figure will do. In the context of Moon, while taking antilog the nearest figure is 1.0685 for which antilog is 2°03' and this will meet our requirement. More over in the instant example the time interval from 5:30 AM to time of birth *i.e.* 9:30 AM is 4 hours which is a round figure and students can easily make 1/6 (of 24 hours motion) to get the 4 hours motion. However more often than not, the time interval may be like 7 Hrs 21 Min., **11** Hrs 39 Min. and so on. In such cases the use of logarithms will be easier and quicker. Students may threfore decide for themselves as
1'18°3'58" I^{s}2006'39" 1 s2001'

P 15° 11' = +6s
to which method *i.e.* the calculator method or the logarithm method appears to be the easier one and may adopt the same. The whole idea is only to get the proportionate motion of planets during the time interval from the given reference position to the time of birth.

10.6 With the above background we can now proceed to find out the longitudes of other planets at the time of birth. It may further be mentioned here that unless the planetary positions are required correct upto seconds (") of arc, we may round off the same to the nearest minute (') of arc by neglecting 30" or less and by adopting next higher minute for 31" and above. In the case of remaining planets, we have 24^h position for each. Our time interval 4^h is also fixed for all the planets. So we can find out/calculate their planetary position simultaneously in one operation in a tabular form. (See next page)

As Ketu is always opposite to RAHU or 6 sign away from RAHU, its longitude is calculated by adding 6 sign to the longitude of Rahu. Accordingly :

Longitude of True Rahu Add 6 sign
Longitude of Ketu 7^5 15° II'
(If it exceeds 12s, deduct 12 signs
but this is not the case here,
Therefore Longitude of Ketu $7^5$15° II'

Students will recall that we had calculated the longitude of lagna for this native *vide* Example 1 of Chapter 8 (Para 8.4.1) as 7^5 15° 40'. We can now draw the chart as follows:

Calculation of Planetary Position at 09:30 **AM (1ST) 25-10-2002**

Position of Planets at 1ST on (Page	Sun	l\ '1ercury	-\'cnus (R)	!dars	Junite	Satur (R)	Rahu (D)
26-10-2002 25-10-2002	6'8'30'1 6$7°30'2	Y26°12' 5'24°33'	6'17'14' 6'17°46	5'12°4 5'12°0	3~2P 3'21°4	2$5°0' 2'5°0	1$15°1 1'15°
Motion in 24	59'47"	1°39'	(_)0°32	39'	7'	(-)2	I
Log of in 24 Hrs. (P- Log 61' Time i.e. 4 Hrs.	1.3802 0.7781	I.1627 0.7781	1.6532 0.7781	1.5673 0.7781	2.313 0.778	2.857 07781	3.158 0.778
Total	2.1583	1.9408	24313	2:3454	3.091	3.635	3.936
Nearest in the Log	2.1584	1.9279	24594	2:3133	3.158	_	_
By Taking of above we get Motion in 4 Add position 25-10-2002	0°10' 6^{5}7°30'	0°17' 5'24°33'	(-)0°5' 6'17"46	0°7' 5'12°0	'. 0°1' 3'21°4	(-)0°0 2'5°0	(+)0° 1'15°
Position at on 25-10-	6'7°40'	5'24°50'	6'17°41	5'12°8'	3'21°4	2'5°2'	PI5°

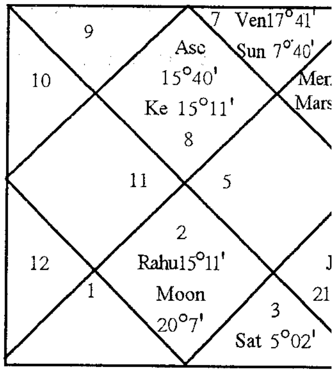
9
7 Ven17°41'
Asc
Sun 7°40'
10
15°40'
Mer
Ke 15°11'
Mars
8
11
5
2
12
Rahu15°11'
J
1
Moon
21
3
20°7'
Sat 5°02'

	RahlSO Moon 20°7'	Sat
	25-10-2002 09: 30 AM (IST) DELHI	Jup 21 °42'
	YABc Sun 7° 15°40' Yen ~*Kej* 17°41'	MaI2 Mer 24°50'

10.7 Students may please note that the pOSItiOn of Planets in the heavens is dependent on the date & time only and is independent of the place of Birth. The place of birth is important for calculating the Rising Sign or Lagna or

Ascendant. Before we close this discussion we will take up another example to work out the longitude/position of planet.

Example 2 : Calculate the longitude of Planets for a native born at 10:24 PM (IST) on 11 July, 2002.

Solution : As the place of birth has not been given we can not calculate the lagna. As such only the longitudes of planets are required to be calculated. This has been worked out in the tabular form in the next page which is self explanatory.

Question 1 : Calculate the Planetary Position
(longitudes of Planets) for following date md time :

(a) 26-1-2003 10:20 AM (1ST)
(b) 25-12-2003 7:30 PM (ZST) London
(c) 15-08-2003 7:30 AM (IST)
(d) 25-04-2003 00:45 AM (ZST) New York

Note : For (b) & (d) students may refer to Para 10.2.

Position at	Sun	Moon	Mere	Venus	Mars	Jup	Sat	Ra (R)
5.30 12-7-2002 5.30 AM 11-7-2002	2^{s}25'36'22" 2^{5}24'39'08"	3^{s}W2'16" 3^{s}1'57'23"	2^{s}15'02' 2^{s}13'Ol'	4"7'22' 4^{s}6'15'	3^{s}5'03' 3^{s}4'25'	3^{s}1'30' 3^{s}1"16'	1'28'42' 1^{5}28'35'	1'23'40' 1'23'44'
Motion in 24 Hl's Log Motion in 24 Ill's. Log of Time interval* (W'54^{m})	57'-14" 1.4025 0.1523	14'04'53" 0.2315 0.1523	2'-01 1.0756 0.1523	1'-07' 1.3323 0.1523	0'-38' 1.5786 0.1523	0'-14' 2.0122 0.1.523	0'-07' 2.3133 0.1523	H 2.5563 0.1523
Total Nearest figure given in table	1.5548 1.5563	0.3838 0.3838	1.2279 1.2289	1.4846 1.4863	1.7309 1.7270	2.1645 2.1584	2.4656 2.4594	2.7086 2.6812
Taking anti-log we get the motion till time of birth Add reference position	0°40' 2"24'39'08"	9°55' 3"1'57'23"	J025' 2^{s}13'01'	0°47' 4^{s}6'15'	0°27' 3^{5}4'25'	0°10' 3^{s}1'16'	0°5' 1^{s}28'35'	(-)0°3' 1^{S}23'44'
Position at birth	2^{s} 25'19'08"	3^{s} 11 °52'23"	2^{s} 14°26'	4^{s} 7°2'	3^{s}4°52'	3^{s} OJ026'	1" 28°40'	1'

*Time Interval for all planets from 5 : 30 A.M. to 10 : 24 P.M. = 16 Hrs 54 Min. (R) Means Retrograde i.e. the planet appears to be moving backwards.

Note: Ketu's position would be six signs away from Rahu and hence not calculated separately.

Positio	Sun	Moon	Mer	Ven	Mars	Jup	Sat	Ra (R)
	I S ,	S ,	S~ ,	S = ,	S ,	S ~,	S	S ,
5.30 15-2-2001	10 2 30	70 51	9 28 1	11 15 3	7 6 13	1 8 02	1 0 36	2 20 5
5.30 AM 14-2-2001	10 1 29	6 17 54	9 29 20	11 14 5	7 5 42	1 7 58	1 0 34	2 20 5
24 hrs motion	11	12 57	(-) 1 09	0 41	0 31	(4	0 2	(-) 0 1
Log 0[24 hrs Motion	1.3730	0.2679	1.319	1.54	1.6670	2.5563	2.8573	3.1584
+Iog of Time interval*	0.2213	0.2213	0.2213	0.2213	0.2213	0.2213	0.2213	0.2213
Total	1.5943	0.4892	1.540	1.76	1.8883	2.7776	3.0786	3.3797
Nearest tigure given in table	1.5902	0.4890	1.545	1.76	1.8796	2.8573	3.1584	3.1584
Taking anti-log we get the motion till time of birth	0°37'	7°47'	(-) 0°41'	0°25	0°19'	0°2'	0°01'	(-)0°01
Add reference position	10 1 29	6 17 54	9 29 20	11 14 54	7 5 42	1 7 58	1 0 34	2 20 51
	10s 2° 06'	6s 25° 41'	9" 28° 39'	11"	7" 6° 01'	1" 8° 00'	IS 0° 35'	2s 20°

Queen Mary
University of London

Mathematical writing

for undergraduate students

Franco Vivaldi
School of Mathematical Sciences

© The University of London, 2013

Last updated: August 15, 2013

To Mary-Carmen.

Preface

This book has evolved from a course in mathematical writing offered to second year undergraduate students at Queen Mary, University of London.

Instructions on writing mathematics are normally given to postgraduate students, because they must write research papers and a thesis. However, there are compelling reasons for providing a similar training at undergraduate level, and, more generally, for raising the profile of writing in a mathematics degree[1].

A researcher knows that writing an article, presenting a result in a seminar, or simply explaining ideas to a colleague are decisive tests of one's understanding of a topic. If a sketched argument has flaws, these flaws will surface as soon as one tries to convince someone else that the argument is correct. The act of exposition is inextricably linked to thinking, understanding, and self-evaluation.

For this reason, undergraduate students should be encouraged to elucidate their thinking in writing, and to assume greater responsibility for the quality of the exposition of their ideas. Their first written submissions tend to be cryptic collections of symbols, which easily hide from view learning inadequacies and fragility of knowledge. It is quite possible to perform a correct calculation having only limited understanding of the subject matter, but it is not possible to write about it. A good writing assignment exposes bad studying habits (approaching formal concepts informally, or treating them as mere processes —see [1]), and provides a most effective tool for correcting them.

This course's declared objective is to teach the students how to develop and present mathematical arguments, in preparation for writing a thesis in their final year. For those who will not write a thesis, this course represents an indispensable minimal alternative, which is also more manageable in terms of teaching resources. The writing material is taken mostly from introductory courses in calculus and algebra. This suffices to challenge even the most capable students, who commented on the 'unexpected depth' required in their thinking, once forced to offer verbal explanations. The students are asked to use words and symbols with the same clarity, precision, and conciseness found in books and lecture notes. This demanding exercise encourages logical accuracy, attention to structure, and economy of thought —the attributes of a mathematical mind. It also forces us to understand better the mathematics we are supposed to know.

---

[1] The poor quality of student writing in Higher Education has raised broad concern [42].

The development of writing techniques proceeds from the particular to the general, from the small to the large: words, phrases, sentences, paragraphs, to end with short compositions. These may represent the introduction of a concept, the proof of a theorem, the summary of a section of a book, the first few slides of a presentation.

The first chapter is a warm-up, listing dos and don'ts of writing mathematics. An essential dictionary on sets, functions, sequences, and equations is presented in chapter 2; these words are then used extensively in simple phrases and sentences. The analysis of mathematical sentences begins in chapter 3, where we develop some constructs of elementary logic (predicates, quantifiers). This material underpins the expansion of the mathematical dictionary in chapter 4, where basic attributes of real functions are introduced: ordering, symmetry, boundedness, continuity. Mathematical arguments are studied in detail in the second part of the book. Chapters 6 and 7 are devoted to basic proof techniques, while chapter 8 deals with existence statements and definitions. Some chapters are dedicated explicitly to writing: chapter 1 gives basic guidelines; chapter 5 is concerned with mathematical notation and quality of exposition; chapter 9 is about writing a thesis. Solutions and hints to selected exercises are given in appendix A.

The symbol **[**$\notin$**]** appears often in exercises. It indicates that the written material should contain *no mathematical symbols*. (The allied symbol **[**$\notin$, n**]** specifies an approximate word length n of the assignment.) In an appropriate context, having to express mathematics without symbols is a most useful exercise. It brings about the discipline needed to use symbols effectively, and is invaluable for learning how to communicate to an audience of non-experts. Consider the following question: **[**$\notin$, 100**]**

> *I have a circle and a point outside it, and I must find the lines through this point which are tangent to the circle. What shall I do?*

The mathematics is elementary; yet answering the question requires a clear understanding of the structure of the problem, and a fair deal of organisation.

> *Write down the equation of a line passing through the point. This equation depends on one parameter, the line's slope, which is the quantity to be determined.*

> *Adjoin the equation of the line to that of the circle, and eliminate one of the unknowns. After a substitution, you'll end up with a quadratic equation in one unknown, whose coefficients still depend on the parameter.*
>
> *Equate the discriminant of the quadratic equation to zero, to obtain an equation —also quadratic— for the slope. Its two solutions are the desired slopes of the tangent lines. Any geometrical configuration involving vertical lines (infinite slope) will require some care.*

The most challenging exercise of this kind is the MICRO-ESSAY, where the synthesis of a mathematical topic has to be performed in a couple of paragraphs, without using any symbols at all. This exercise prepares the students for writing abstracts, a notoriously difficult task.

The available literature on mathematical writing is almost entirely targeted to post-graduate students and researchers. An exception is *How to think like a mathematician*, by K. Houston [19], written for students entering university, which devotes two early chapters to mathematical writing. The advanced texts include *Mathematical writing*, by D. E. Knuth, T. L. Larrabee, and P. M. Roberts [22], *Handbook of writing for the mathematical sciences* by N. Higham [18], and *A primer of mathematical writing* by S. G. Krantz [20]. Equally valuable is the concise classic text *Writing mathematics well*, by L. Gillman [12]. Unfortunately, this 50-page booklet is out of print, and used copies may command high prices.

The timeless, concise book *The elements of style*, by W. Strunk Jr and E. B. White [36] is an ideal complement to the present textbook. Anyone interested in writing should study it carefully.

This book was inspired by the lecture notes of a course in logic given by Wilfrid Hodges at Queen Mary in 2005-2006. This course used writing as an essential tool, in an innovative way. Wilfrid has been an ideal companion during a decade-long effort to bring writing to centre stage in our mathematics degree, and I am much indebted to him.

The development of the first draft of this book was made possible by a grant from the Thinking Writing Secondment Scheme, at Queen Mary, University of London; their support is gratefully acknowledged. I also thank Will Clavering and Sally Mitchell for their advice and encouragement throughout this project, and Ivan Tomašić for the discussions on logic over coffee. The students of the Mathematical

Writing course at Queen Mary endured the many early versions of this work; I thank them for spotting errors and mis-prints. I also thank Lara Alcock, for her thoughtful, detailed feedback on an earlier version of the manuscript, which prompted me to re-consider the structure of some chapters.

Finally, I owe a special debt to my daughter Giulia who had the patience to read the final draft, suggesting improvements in hundreds of sentences and removing hundreds of commas.

Franco Vivaldi
London, 2013.

Contents

1 Some writing tips **1**
- 1.1 Grammar . . . 2
- 1.2 Numbers and symbols . . . 4
- 1.3 Style . . . 6
- 1.4 Preparation and structure . . . 7

2 Essential dictionary **11**
- 2.1 Sets . . . 12
  - 2.1.1 Defining sets . . . 16
  - 2.1.2 Sets of numbers . . . 18
  - 2.1.3 Writing about sets . . . 23
- 2.2 Functions . . . 25
- 2.3 Representations of sets . . . 29
- 2.4 Sequences . . . 31
  - 2.4.1 Some constructs involving sequences . . . 33
- 2.5 Equations . . . 36
- 2.6 Expressions . . . 39
  - 2.6.1 Levels of description . . . 40
  - 2.6.2 Characterising expressions . . . 43

3 Mathematical sentences **51**
- 3.1 Relational operators . . . 52
- 3.2 Logical operators . . . 53
- 3.3 Predicates . . . 57
- 3.4 Quantifiers . . . 60
  - 3.4.1 Quantifiers and functions . . . 65
  - 3.4.2 Existence statements . . . 68
- 3.5 Negating logical expressions . . . 69

3.6 Relations 72

4 Describing functions 79
4.1 Ordering properties 79
4.2 Symmetries 81
4.3 Boundedness 83
4.4 Neighbourhoods 84
4.5 Continuity 86
4.6 Other properties 87
4.7 Describing real sequences 90

5 Writing well 97
5.1 Choosing words 97
5.2 Choosing symbols 100
5.3 The sigma-notation 106
5.4 Improving formulae 109
5.5 Writing definitions 112
5.6 Introducing a concept 115
5.7 Writing a short description 117

6 Forms of argument 123
6.1 Anatomy of a proof 123
6.2 Proof by cases 127
6.3 Implications 129
  6.3.1 Direct proof 130
  6.3.2 Proof by contrapositive 131
6.4 Proving conjunctions 133
  6.4.1 Loops of implications 135
6.5 Proof by contradiction 136
6.6 Counterexamples and conjectures 137
6.7 Wrong arguments 141
  6.7.1 Examples vs. proofs 141
  6.7.2 Wrong implications 142
  6.7.3 Mishandling functions 144
6.8 Writing a good proof 145

7 Induction 155
7.1 The well-ordering principle 156
7.2 The infinite descent method 157
7.3 Peano’s induction principle 159
7.4 Strong induction 162
7.5 Good manners with induction proofs 164

8 Existence and definitions 169
8.1 Proofs of existence 169
8.2 Unique existence 173
8.3 Definitions 176
8.4 Recursive definitions 178
8.5 Wrong definitions 180

9 Writing a thesis 185
9.1 Theses and other publications 186
9.2 Title 189
9.3 Abstract 191
9.4 Citations and bibliography 196
    9.4.1 Avoiding plagiarism 199
9.5 TEX and LATEX 200

A Solutions to exercises 205

Chapter 1

Some writing tips

The following short mathematical sentences are poorly formed in one way or another. Can you identify all the errors and would you know how to fix them? Compare your answers with those given at page 205.

Exercise 1.1.

1. a is positive.
2. Two is the only even prime.
3. If $x > 0 \;\; g(x) \neq 0$.
4. We minus the equation.
5. $x^2 + 1$ has no real solution.
6. When you times it by negative x, $\leqslant$ becomes $\geqslant$.
7. The set of solutions are all odd.
8. $\sin(\pi x) = 0 \;\Rightarrow x$ is integer.
9. An invertible matrics is when the determinant is non-zero.
10. This infinete sequence has less negative terms.

In this chapter you will learn how to recognise and correct common mistakes, the first step towards writing mathematics well. By the time you'll reach the exercises at the end of the chapter you should already feel a sense of progress. You should return to this chapter repeatedly, to monitor the assimilation of good practice.

1.1 Grammar

You are advised to use an English dictionary, e.g.[1], [32], and to recall the basic terminology of grammar (adjective, adverb, noun, pronoun, verb, etc.[2]) —see, for instance, [36, pp. 89–95].

- Write in complete sentences. Every sentence should begin with a capital letter, end with a full stop, and contain a subject and a verb. The expression 'A cubic polynomial' is not a sentence, because it doesn't have a verb. It would be appropriate as a caption, or a title, but you can't simply insert it in the middle of a paragraph.

- Make sure that the nouns match the verbs grammatically.

Bad: The set of primes are infinite.
Good: The set of primes is infinite.

(The verb refers to 'the set', which is singular.)

- Make a pronoun agree with its antecedent.

Bad: Each function should be greater than their derivative.
Good: Each function should be greater than its derivative.

(The pronoun 'its' refers to 'function', which is singular.)

- If possible, do not split infinitives.

Bad: We have to thoroughly examine this proof.
Good: We have to examine this proof thoroughly.

Bad: I was taught to always simplify fractions.
Good: I was taught always to simplify fractions.

(The infinitives are 'to examine' and 'to simplify'.) In some cases a split infinitive may be acceptable, even desirable.

Good: This is a sure way to more than double the length of the manuscript.

---

[1] Abbreviation for the Latin *exempli gratia*, meaning 'for the sake of example'.

[2] Abbreviation for the Latin *et cetera*, meaning 'and the rest'.

- Check the spelling. No point in crafting a document carefully, if you then spoil it with spelling mistakes. If you use a word processor, take advantage of a spell checker. These are some frequently misspelled words:

  BAD: auxillary, catagory, consistant, correspondance, impliment, indispensible, ocurrence, preceeding, refering, seperate.

  These are misspelled mathematical words that I found in mathematics examination papers:

  BAD: arithmatic, arithmatric, derivitive, divisable, falls (false), infinaty, matrics, orthoganal, orthoginal, othogonal, reciprical, scalor, theorom.

- Be careful about distinctions in meaning.
  Do not confuse *it's* (abbreviation for *it is*) with *its* (possessive pronoun).

  BAD: Its an equilateral triangle: it's sides all have the same length.

  Do not confuse the noun *principle* (general law, primary element) with the adjective *principal* (main, first in rank of importance).

  BAD: the principal of induction
  BAD: the principle branch of the logarithm

  Do not use *less* (of smaller amount, quantity) when you should be using *fewer* (not as many as).

  BAD: There are less primes between 100 and 200 than between 1 and 100.

- Do not use *where* inappropriately. As a relative adverb, *where* stands for *in which* or *to which*; it does not stand for *of which*.

  BAD: We consider the logarithmic function, where the derivative is positive.
  GOOD: We consider the logarithmic function, whose derivative is positive.

  The adverb *when* is subject to similar misuse.

  BAD: A prime number is when there are no proper divisors.
  GOOD: A prime number is an integer with no proper divisors.

- Do not use *which* when you should be using *that*. Even when both words are correct, they have different meanings. The pronoun *that* is defining, it is used to identify an object uniquely, while *which* is non-defining, it adds information to an object already identified.

> The argument that was used above is based on induction.
> [*Specifies which argument.*]
> The following argument, which will be used in subsequent proofs, is based on induction.
> [*Adds a fact about the argument in question.*]

A simple rule is to use *which* only when it is preceded by a comma or by a preposition, or when it is used interrogatively.

- In presence of parentheses, the punctuation follows strict rules. The punctuation outside parentheses should be correct if the statement in parentheses is removed; the punctuation within parentheses should be correct independently of the outside.

Bad: This is bad. (Superficially, it looks good).
Good: This is good. (Superficially, it looks like the BAD one.)

Bad: This is bad, (on two accounts.)
Good: This is good (as you would expect).

1.2 Numbers and symbols

Effectively combining numbers, symbols, and words is a main theme in this course. We begin to look at some basic conventions.

- A sentence containing numbers and symbols must still be a correct English sentence, including punctuation.

Bad: $a < b \;\; a \neq 0$
Good: We have $a < b$ and $a \neq 0$.
Good: We find that $a < b$ and $a \neq 0$.
Good: Let $a < b$, with $a \neq 0$.

Bad: $x^2 - 7^2 = 0$. $x = \pm 7$.
Good: Let $x^2 - 7^2 = 0$; then $x = \pm 7$.
Good: The equation $x^2 - 7^2 = 0$ has two solutions: $x = \pm 7$.

- Omit unnecessary symbols.

Bad: Every differentiable real function f is continuous.
Good: Every differentiable real function is continuous.

- If you use small numbers for counting, write them out in full; if you refer to specific numbers, use numerals.

BAD: The equation has 4 solutions.
GOOD: The equation has four solutions.
GOOD: The equation has 127 solutions.

BAD: Both three and five are prime numbers.
GOOD: Both 3 and 5 are prime numbers.

- If at all possible, do not begin a sentence with a numeral or a symbol.

BAD: ρ is a rational number with odd denominator.
GOOD: The rational number ρ has odd denominator.

- Do not combine operators ($+$, $\neq$, $<$, etc.) with words.

BAD: The difference $b-a$ is <0
GOOD: The difference $b-a$ is negative.

- Do not misuse the implication operator $\Rightarrow$ or the symbol $\therefore$. The former is employed only in symbolic sentences (section 3.2); the latter is not used in higher mathematics.

BAD: a is an integer $\Rightarrow$ a is a rational number.
GOOD: If a is an integer, then a is a rational number.

BAD: $\Rightarrow\ x=3$.
BAD: $\therefore\ x=3$.
GOOD: hence $x=3$.
GOOD: and therefore $x=3$.

- Within a sentence, adjacent formulae or symbols must be separated by words.

BAD: Consider $A_n, n<5$.
GOOD: Consider A_n, where $n<5$.

BAD: Add p k times to c.
BAD: Add p to c k times.
GOOD: Add p to c, repeating this process k times.

For displayed equations the rules are a bit different, because the spacing between symbols becomes a syntactic element. Thus an expression of the type

$$A_n = B_n, \quad n < 5$$

is quite acceptable (see section 5.4).

1.3 Style

A good sentence needs a lot more than grammatical correctness.

- Give priority to clarity over fancy language. Avoid long and involved sentences; break long sentences into shorter ones.

Bad: We note the fact that the polynomial $2x^2 - x - 1$ has the coefficient of the x^2 term positive.
Good: The leading coefficient of the polynomial $2x^2 - x - 1$ is positive.

Bad: The inverse of the matrix A requires the determinant of A to be non-zero in order to exist, but the matrix A has zero determinant, and so its inverse does not exist.
Good: The matrix A has zero determinant, hence it has no inverse.

- Place important words in a prominent position within the sentence. Suppose you are introducing the logarithm:

Bad: An important example of a transcendental function is the logarithm.

In this classic bad opening '*an example of something is something else*', the focus of attention is the transcendental functions, not the logarithm.

Good: The logarithm is an important example of a transcendental function.
Good: Let us now define a key transcendental function: the logarithm.

Suppose you wish to emphasise the scalar product:

Bad: A commonly used method to check the orthogonality of two vectors is to verify that their scalar product is zero.
Good: If the scalar product of two vectors is zero, then the vectors are orthogonal.

- Prefer the active to the passive voice.

Bad: The convergence of the above series will now be established.
Good: We establish the convergence of the above series.

- Vary the choice of words to avoid repetition and monotony[3].

Bad: The function defined above is a function of both x and y.

---

[3] If necessary, consult a thesaurus.

Good: The function defined above depends on both x and y.

- Do not use unfamiliar words unless you know their exact meaning.

Bad: A simplistic argument shows that our polynomial is irreducible.
Good: A simple argument shows that our polynomial is irreducible.

- Do not use vague, general statements to lend credibility to your writing. Avoid emphatic statements.

Bad: Differential equations are extremely important in modern mathematics.

Bad: The proof is very easy, as it makes an elementary use of the triangle inequality.
Good: The proof uses the triangle inequality.

- Do not use jargon, or informal abbreviations: it looks immature rather than ‘cool’.

Bad: Spse U subs x into T eq. Wot R T soltns?

- Enclose side remarks within commas, which is very effective, or parentheses (getting them out of the way). To isolate a phrase, use hyphenation —it really sticks out— or, if you have a word processor, *change font* (**but** `don't` overdo *it*).

- Take punctuation seriously. To improve it, read [39].

1.4 Preparation and structure

There are things one must keep in mind when preparing any document.

- Begin by writing your document in draft form, or at least write down a list of key points. Few people are able to produce good writing at the first attempt.

- Consider the background of your readers; are they familiar with the meaning of the words you use? It’s easy to write a mathematics text that’s too difficult; it’s almost impossible to write one that’s too easy.

- Form each sentence in your head before writing it down. Then read carefully what you have written. Read it aloud: how does it sound? Have you written what you intended to write? Is it clear? Don’t hesitate to rewrite.

- Split the text into paragraphs. Each paragraph should be about one ‘idea’, and it should be clear how you are moving from one idea to the next. Be prepared to re-arrange paragraphs. The first idea you thought of may not have been the best one; the sequence of arguments you have chosen may not be optimal.
- When you finish writing, consider the opening and closing sentences of your document. The former should motivate the readers to keep reading, the latter should mark a resting place, like the final bars in a piece of music.
- Word processing has changed the way we write, and often a document is the end-product of several successive approximations. After prolonged editing, one stops seeing things. If you have time, leave your document to rest for a day or two, and then read it again.

Exercises

Exercise 1.2. Improve the writing, following the guidelines given in this chapter.

1. There are 3 special cases.
2. X is a finite set.
3. It does not tend to infinaty.
4. It follows $x-1=y^4$.
5. $\therefore\ c^{-1}$ is undefined.
6. The product of 2 negatives is positive.
7. We square the equation.
8. We have less solutions than we had before.
9. $x^2=y^2$ are two othogonal lines.
10. Let us device a strategy for a proof.
11. This set of matrix are all invertible.
12. If the integral $=0$ the function is undefined.
13. Purely imaginary is when the real part is zero.

14. Construct the set of vertex of triangles.
15. From the fact that $x = 0$, I can't divide by x.
16. A circle is when major and minor axis are the same.
17. The function f is not discontinuous.
18. Plug-in that expression in the other equation.
19. I found less solutions than I expected.
20. When the discriminant is < 0, you get complex.
21. We prove Euler theorom.
22. The definate integral is where you don't have integration limits.
23. The asyntotes of this hiperbola are othogonal.
24. A quadratic function has 1 stationery point.
25. The solution is not independent of s.
26. a is negative $\therefore$ $\sqrt{a}$ is complex.
27. Thus $x = a$. (We assume that a is positive).
28. Each value is greater than their reciprical.
29. Remember to always check the sign.
30. Differentiate f n times.

Chapter 2

Essential dictionary

In writing mathematics we use words and symbols to describe facts. We need to explain the meanings of words and symbols, and to state and prove the facts. In this chapter we consider a mathematical dictionary; we'll be concerned with facts later.

Ideally, we ought to explain the meaning of every word and symbol that we use. But this is impossible: we would need to explain the words used in the explanation, and so on. Instead, we should only explain a word or symbol if our explanation will make it clearer than it was before. Accordingly, we shall call a word or symbol **primitive**[1] if it's suitable to use without an explanation of its meaning.

An ordinary English word like 'thousand' is obviously primitive, but for more specialised words we must consider the context. When we communicate to the general public, the term **multiplication** can safely be regarded as primitive. Likewise, there should be no need to explain to a mathematician what an **eigenvalue** is, while a number theorist will be familiar with **conductor.** Then there are extremes of specialisation: only a handful of people on this planet will know the meaning of **Hsia kernel.** Finally, terms such as **exceptional set** mean different things in different contexts. Understanding what constitutes an appropriate set of primitives is an essential pre-requisite for the communication of complex knowledge. Getting it right is tricky.

This chapter describes some two hundred mathematical words, highlighted in boldface, and provided with accompanying notation. This is our essential mathematical dictionary, built around few basic terms: **set**, **function**, **equation**, **sequence**. As we introduce new words, we use them in short phrases and sentences.

---

[1] This terminology is due to Blaise Pascal (French: 1623-1662).

2.1 Sets

A **set** is a collection of *well-defined*, *unordered*, *distinct* objects. (This is the so-called 'naive definition' of a set, due to Cantor[2].) These objects are called the **elements** of a set, and a set is determined by its elements. We may write

> *The set of all odd integers*
> *The set of vertices of a pentagon*
> *The set of differentiable real functions*

In simple cases, a set can be defined by listing its elements, separated by commas, enclosed within curly brackets. The expression

$$\{1,2,3\}$$

denotes the set whose elements are the integers 1, 2 and 3. Two sets are equal if they have the same elements:

$$\{1,2,3\}=\{3,2,1\}.$$

(By definition, the order in which the elements of a set are listed is irrelevant.)

It is customary to ignore repeated set elements: $\{2,1,3,1,3\}=\{2,1,3\}$. This convention, adopted by computer algebra systems, simplifies the definition of sets. If repeated elements are allowed and not collapsed, then we speak of a **multiset**: $\{2,1,3,1,3\}$. The **multiplicity** of an element of a multiset is the number of times the element occurs. Reference to multiplicity usually signals that there is a multiset in the background:

> *Every quadratic equation has two complex solutions, counting multiplicities.*

Multisets are not as common as sets.

The set $\{\}$ with no elements is called the **empty set**, denoted by the symbol $\emptyset$. The empty set is distinct from 'nothing', it is more like an empty container. For example, the statements

> *This equation has no solutions.*
> *The solution set of this equation is empty.*

---

[2] Georg Cantor (German: 1845–1918).

have the same meaning.

To assign a symbol to a mathematical object, we use an **assignment statement** (or **definition**), which has the following syntax

$$A := \{1,2,3\}. \tag{2.1}$$

This expression assigns the symbolic name A to the set $\{1,2,3\}$, and now we may use the former in place of the latter. The symbol '$:=$' denotes the **assignment operator**. It reads '*becomes*', or '*is defined to be*', rather than '*is equal to*', to underline the difference between assignment and equality (in computer algebra, the symbols = and := are not interchangeable at all!). So we can't write $\{1,2,3\} := A$, because the left operand of an assignment operator must be a symbol or a symbolic expression.

The right-hand side of an assignment statement such as (2.1) is a collection of symbols or words that pick out a unique thing, which logicians call the *definiens* (Latin for 'thing that defines'). The left-hand side is a symbol that will be used to stand for this unique thing, which is called the *definiendum* (Latin for 'thing to be defined'). These terms are rather heavy, but they are widely used [35, chapter 8]. The definiendum may also be a symbolic expression —see below.

While it's very common to use the equal sign '$=$' also for an assignment, the specialised notation $:=$ improves clarity. There are other symbols for the assignment operator, namely

$$\overset{\text{def}}{=} \qquad \overset{\nabla}{=}, \tag{2.2}$$

which make an even stronger point.

To indicate that x is an element of a set A, we write

$$x \in A \qquad x \textit{ is an element of } A \qquad x \textit{ belongs to } A.$$

The symbol $\notin$ is used to negate membership. Thus

$$\{7,5\} \in \{5,\{5,7\}\} \qquad 7 \notin \{5,\{5,7\}\}.$$

(Think about it.)

A **subset** B of a set A is a set whose elements all belong to A. We write

$$B \subset A \qquad B \textit{ is a subset of } A \qquad B \textit{ is contained in } A$$

and we use $\not\subset$ to negate set inclusion. For example

$$\{3,1\} \subset \{1,2,3\} \qquad \emptyset \subset \{1\} \qquad \{2,3\} \not\subset \{2,\{2,3\}\}.$$

Every set has at least two subsets: itself and the empty set. Sometimes these are referred to as the **trivial** subsets. Every other subset —if any— is called a **proper subset.** Motivated by an analogy with $\leqslant$ and $<$, some authors write $\subseteq$ in place of $\subset$, reserving the latter for proper inclusion: $\mathbb{R} \subseteq \mathbb{R}$, $\mathbb{Q} \subset \mathbb{R}$. Proper inclusion is occasionally expressed with the pedantic notation $\underset{\neq}{\subset}$.

The **cardinality** of a set is the number of its elements, denoted by the prefix #.

$$\#\{7,-1,0\} = 3 \qquad\qquad \#A = n.$$

The absolute value symbol $|\cdot|$ is also used to denote cardinality: $|A| = n$. Common sense will tell when this choice is sensible. A set is **finite** if its cardinality is an integer, and **infinite** otherwise (see section 2.3). To indicate that the set A is finite, without disclosing its cardinality, we write

$$\#A < \infty. \tag{2.3}$$

Next we consider the words associated with operations between sets. We write $A \cap B$ for the **intersection** of the sets A and B: this is the set comprising elements that belong to both A and B. If $A \cap B = \emptyset$, we say that A and B are **disjoint,** or have **empty intersection.** The sets $A_1, A_2, \ldots$ are **pairwise disjoint** if $A_i \cap A_j = \emptyset$ whenever $i \neq j$.

We write $A \cup B$ for the **union** of A and B, which is the set comprising elements that belong to A or to B (or to both A and B).

We write $A \smallsetminus B$ for the (**set**) **difference** of A and B, which is the collection of the elements of A that do not belong to B. The **symmetric difference** of A and B, denoted by $A \Delta B$, is defined as

$$A \Delta B \overset{\text{def}}{=} (A \smallsetminus B) \cup (B \smallsetminus A).$$

The assignment operator '$\overset{\text{def}}{=}$' (cf. (2.2)) makes it clear that this is a definition. This notation establishes the meaning of $A\Delta B$, which is a symbolic expression rather than an individual symbol. The following examples illustrate the action of set operators.

$$\begin{aligned}
\{1,2,3\} \cap \{3,4,5\} &= \{3\} \\
\{1,2,3\} \cup \{3,4,5\} &= \{1,2,3,4,5\} \\
\{1,2,3\} \smallsetminus \{3,4,5\} &= \{1,2\} \\
\{1,2,3\}\, \Delta\, \{3,4,5\} &= \{1,2,4,5\}.
\end{aligned}$$

The above **set operators** are **binary**; they have two sets as **operands.** The identities

$$A \cap B = B \cap A \qquad (A \cap B) \cap C = A \cap (B \cap C)$$

express the **commutative** and **associative** properties of the intersection operator. Union and symmetric difference enjoy the same properties, but set difference doesn't.

Let A be a subset of a set X. The **complement** of A (in X) is the set $X \smallsetminus A$, denoted by A' or by A^c. The complement of a set is defined with respect to an **ambient set** X. Reference to the ambient set may be omitted if there is no ambiguity. So we write

The odd integers is the complement of the even integers

since it's clear that the ambient set is the integers.

With set operators we can construct new sets from old ones, although, in a sense, we are re-cycling things we already have. To create genuinely new sets, we introduce the notion of **ordered pair**. This is an expression of the type (a,b), with a and b arbitrary quantities. Ordered pairs are defined by the property

$$(a,b) = (c,d) \qquad \text{if} \qquad a = c \quad \text{and} \quad b = d. \tag{2.4}$$

The ordered pair (a,b) should not be confused with the set $\{a,b\}$, since for pairs order is essential and repetition is allowed. (Ordered pairs may be defined solely in terms of sets —see exercise 2.15.) Let A and B be sets. We consider the set of all ordered pairs (a,b), with a in A and b in B. This set is called the **cartesian product** of A and B, and is written as

$$A \times B.$$

Note that A and B need not be distinct; one may write A^2 for $A \times A$, A^3 for $A \times A \times A$, etc. Because the cartesian product is **associative**, the product of more than two sets is defined unambiguously. Also note that the explicit presence of the multiplication operator '$\times$' is needed here, because the expression AB has a different meaning: it's the **algebraic product** of two sets, to be defined in section 2.1.2 (see equation (2.13)).

A **partition** of a set A is a collection of pairwise disjoint non-empty subsets of A, whose union is A. A partition may be described as a **decomposition** of a set into **classes**. For instance, the set $\{\{2\},\{1,3\}\}$ is a partition of $\{1,2,3\}$, the even and odd integers form a partition of the integers, and the plane may be partitioned into concentric circles.

The **power set** $\mathbf{P}(A)$ of a set A is the set of all subsets of A. Thus if $A = \{1,2,3\}$, then

$$\mathbf{P}(A) = \{\{\},\{1\},\{2\},\{3\},\{1,2\},\{1,3\},\{2,3\},\{1,2,3\}\}\,.$$

To construct a subset of A, we consider each element of A, and we decide whether to include it or to leave it out. Because any sequence of choices is allowed, if A has n elements, then $\mathbf{P}(A)$ has 2^n elements. (A formal proof requires induction, see exercise 7.3, p. 165.)

2.1.1 Defining sets

Defining a set by listing its elements is inadequate for all but the simplest situations. How do we define large or infinite sets? A simple device is to use the **ellipsis** '$\ldots$', which indicates the deliberate omission of certain elements, the identity of which is made clear by the context. For example, the set $\mathbb{N}$ of **natural numbers** is defined as

$$\mathbb{N} := \{1,2,3,\ldots\}.$$

Here the ellipsis represents all the integers greater than 3. Some authors regard 0 as a natural number, so the definition

$$\mathbb{N} := \{0,1,2,3,\ldots\}$$

is also found in the literature. Both definitions have merits and drawbacks; mathematicians occasionally argue about it, but this issue will never be resolved. So, when using the symbol $\mathbb{N}$, one may need to clarify which version of this set is employed[3]. The set of **integers,** denoted by $\mathbb{Z}$ (from the German *Zahlen*, meaning numbers), can also be defined using ellipses

$$\mathbb{Z} := \{\ldots,-2,-1,0,1,2,\ldots\} \qquad \text{or} \qquad \mathbb{Z} := \{0,\pm 1,\pm 2,\ldots\}.$$

To define general sets we need more powerful constructs. A **standard definition** of a set is an expression of the type

$$\{x : x \text{ has } \mathscr{P}\} \tag{2.5}$$

where $\mathscr{P}$ is some unambiguous property that things either have or don't have. This expression identifies the set of all objects x that have property $\mathscr{P}$. The colon ':'

---

[3] Some authors denote the second version by the symbol $\mathbb{N}_0$.

separates out the object's symbolic name from its defining properties. The vertical bar '|' or the semicolon ';' may be used for the same purpose.

Thus the empty set may be defined symbolically as

$$\emptyset \overset{\text{def}}{=} \{x : x \neq x\}.$$

The property $\mathscr{P}$ is 'x is not equal to x', which is not satisfied by any x. Likewise, the cartesian product $A \times B$ of two sets (see section 2.1) may be specified as

$$\{x : x = (a,b) \text{ for some } a \in A \text{ and } b \in B\}.$$

The rule 'x has property $\mathscr{P}$' now reads: 'x is of the form (a,b) with $a \in A$ and $b \in B$'. The same set may be defined more concisely as

$$\{(a,b) : a \in A \text{ and } b \in B\}.$$

This is a variant of the standard definition (2.5), where the type of object being considered (ordered pair) is specified at the outset. This form of standard definition can be very effective.

The set $\mathbb{Q}$ of **rational numbers** —ratios of integers with non-zero denominator— is defined as follows

$$\mathbb{Q} := \left\{\frac{a}{b} : \ a \in \mathbb{Z},\ b \in \mathbb{N},\ \gcd(a,b) = 1)\right\}. \tag{2.6}$$

The property $\mathscr{P}$ is phrased in such a way as to avoid repetition of elements. This is the so-called **reduced form** of rational numbers. The rational numbers may also be defined abstractly, as infinite sets of equivalent fractions —see section 3.6.

One might think that in the expression for a set we could choose any property $\mathscr{P}$. Unfortunately this doesn't work for a reason known as the *Russell-Zermelo paradox*[4] (1901). Let $\mathscr{P}$ be the property of being a set that is not a member of itself. Thus the quantity

$$x = \{3, \{3, \{3, \{3\}\}\}\}$$

has property $\mathscr{P}$, whereas

$$x = \{3, x\} \qquad \text{or} \qquad x = \{3, \{3, \{3, \{3, \ldots\}\}\}\}$$

does not have it. (In the above expression, the nested parentheses must match, so the notation $\{3, \{3, \{3, \{3, \ldots\}$ is incorrect.) Let now W be the set of all sets with property $\mathscr{P}$. Then a set S is a member of W if S is not a member of S. Apply this to W: W is a member of W precisely if W is not a member of W. Impossible!

---

[4] Bertrand Russell, British philosopher and logician (1872–1970); Ernst Zermelo (German: 1871–1953).

Fortunately, we can define a set in such a way that the definition guarantees the existence of the set. A **Zermelo definition** identifies a set W by describing it as

The set of members of X that have property $\mathscr{P}$

where the **ambient set** X is given beforehand, and $\mathscr{P}$ is a property that the members of X either have or do not have. In symbols, this is written as

$$W := \{x \in X : x \text{ has } \mathscr{P}\}. \tag{2.7}$$

For example, the expression

The set of real numbers strictly between 0 and 1

is a Zermelo definition: the ambient set is the set of real numbers, and we form our set by choosing from it the elements that have the stated property.

Zermelo definitions work because it's a basic principle of mathematics (the so-called *subset axiom*) that for any set X of objects and any property $\mathscr{P}$, there is exactly one set consisting of the objects that are in X and have property $\mathscr{P}$. In section 3.3 we shall see that the definiens of a Zermelo definition —a sentence with a variable x in it— is just a special type of function, called a **predicate.**

Both styles of definitions, standard and Zermelo, are widely used in mathematical writing.

2.1.2 Sets of numbers

The 'open face' symbols $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ were introduced in the previous section to represent the natural numbers, the integers, and the rationals, respectively. Likewise, we denote by $\mathbb{R}$ the set of **real** numbers (its symbolic definition is left as exercise 2.14), while the set of **complex** numbers is denoted by $\mathbb{C}$. The set $\mathbb{C}$ may be written as

$$\mathbb{C} \overset{\text{def}}{=} \left\{x + iy : i^2 = -1,\ x, y \in \mathbb{R}\right\}.$$

The symbol i is called the **imaginary unit,** while x and y are, respectively, the **real part** $\mathrm{Re}(z)$ and the **imaginary part** $\mathrm{Im}(z)$ of the complex number $z = x + iy$. The sets $\mathbb{R}$ and $\mathbb{C}$ are represented geometrically as the **real line** and the **complex plane** (or **Argand plane**), respectively. A plot of complex numbers in the Argand plane is called an **Argand diagram**. We have the chain of proper inclusions

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

We turn to operations with numbers. The **sum** and **difference** of two numbers x and y are always written $x + y$ and $x - y$, respectively. By contrast, their **product** may be written in several equivalent ways:

$$xy \qquad x \cdot y \qquad x \times y, \tag{2.8}$$

and so may their **quotient**:

$$\frac{x}{y} \qquad x/y \qquad x : y.$$

(The notation $x : y$ is used mostly in elementary texts.) Do not confuse the product dot '$\cdot$' with the **decimal point** '.', e.g., $3 \cdot 4 = 12$ and $3.4 = 17/5$.

The **reciprocal** of x, defined for $x \neq 0$, is also written in several ways:

$$\frac{1}{x} \qquad 1/x \qquad x^{-1}.$$

The notation for exponentiation is x^y, where x is the **base** and y the **exponent**. If the exponent is a positive integer, then exponentiation is defined as repeated multiplication [5], which may be written symbolically as follows:

$$x^n \overset{\text{def}}{=} \underbrace{x \cdots x}_{n} \qquad n \geqslant 1.$$

The assignment operator $\overset{\text{def}}{=}$ indicates that this is a definition. The use of the underbrace is necessary to specify the number of terms in the product, because all terms are identical. Also note the use of the **raised ellipsis** '$\cdots$' to represent repeated multiplication (or repeated applications of any operator), to be compared with the ordinary ellipsis '$\ldots$', used for sets and sequences (see section 2.4). Thus

$$\underbrace{x \cdots x}_{4} = x \cdot x \cdot x \cdot x \qquad \underbrace{x, \ldots, x}_{4} = x, x, x, x$$

---

[5] Defining exponentiation for a general exponent requires the logarithmic and exponential functions.

whereas the notation $x \ldots x$ is incorrect.

In arithmetic, the symbol '$|$' is used for **divisibility.**

$$3|x \qquad 3 \textit{ divides } x \qquad x \textit{ is a multiple of } 3.$$

EXAMPLE. Turn symbols into words.

$$\{x \in \mathbb{Z} : x \geqslant 0, 2 \mid x\}$$

BAD: The set of integers that are greater than or equal to zero, and such that 2 divides them. [*Robotic.*]

GOOD: The set of non-negative even integers.

A positive divisor of an integer n, which is not 1 or n, is called a **proper divisor**. A **prime** is an integer greater than 1 that has no proper divisors. The acronyms **gcd** and **lcm** are used for **greatest common divisor** and **least common multiple.** (The expression **highest common factor** (hcf) —a variant of gcd which is popular in schools— is seldom used in higher mathematics.) Some authors use (a,b) for $\gcd(a,b)$; this is to be avoided, since this notation is already overloaded. Two integers are **co-prime** (or **relatively prime**) if their greatest common divisor is 1.

We now construct new sets from the sets of numbers introduced above. An **interval** is a subset of $\mathbb{R}$ of the type

$$[a,b] := \{x \in \mathbb{R} : a \leqslant x \leqslant b\}$$

where a,b are real numbers, with $a < b$. This interval is **closed,** that is, it contains its end points. (A point is sometimes regarded as a degenerate closed interval, by allowing $a = b$ in the definition.) We also have **open** intervals

$$(a,b) := \{x \in \mathbb{R} : a < x < b\}$$

as well as **half-open** intervals

$$[a,b) \qquad (a,b].$$

The notational clash between an open interval $(a,b) \subset \mathbb{R}$ and an ordered pair $(a,b) \in \mathbb{R}^2$ is unfortunate but unavoidable, since both notations are firmly established. For (half) open intervals, there is the following alternative — and very logical— notation

$$]a,b[\qquad [a,b[\qquad]a,b],$$

which, for some reason, is not so common.

The interval with end-points $a = 0$ and $b = 1$ is the (open, closed, half-open) **unit interval.** A semi-infinite interval

$$\{x \in \mathbb{R} : a < x\} \qquad \{x \in \mathbb{R} : x \leqslant b\}$$

is called a **ray.** The rays consisting of all positive real or rational numbers are particularly important, and have a dedicated notation

$$\mathbb{R}^+ := \{x \in \mathbb{R},\ x > 0\} \qquad \mathbb{Q}^+ := \{x \in \mathbb{Q},\ x > 0\} \tag{2.9}$$

whereas $\mathbb{Z}^+$ is just $\mathbb{N}$.

Some authors extend the meaning of interval to include also rays and lines, and use expressions such as

$$(-\infty, \infty) \qquad [a, \infty) \qquad (-\infty, b]. \tag{2.10}$$

As infinity does not belong to the set of real numbers, the notation $[1, \infty]$ is incorrect.

A variant of (2.9) is used to denote non-zero real and rational numbers

$$\mathbb{R}^* := \{x \in \mathbb{R},\ x \neq 0\} \qquad \mathbb{Q}^* := \{x \in \mathbb{Q},\ x \neq 0\}. \tag{2.11}$$

This notation is common but not universally recognised; before using these symbols, a clarifying comment may be appropriate (see section 5.2).

The set $\mathbb{R}^2$ of all ordered pairs of real numbers is called the **cartesian plane,** which is the cartesian product of the real line with itself. If $(x, y) \in \mathbb{R}^2$, then the first component x is called the **abscissa** and the second component y the **ordinate**.

The set $\mathbb{Q}^2 \subset \mathbb{R}^2$, the collection of points of the plane having rational coordinates, is called the set of **rational points** in $\mathbb{R}^2$. The set $[0, 1]^2 \subset \mathbb{R}^2$ is called the **unit square.** In $\mathbb{R}^3$ we have the **unit cube** $[0, 1]^3$, and for $n > 3$ we have the **unit hypercube** $[0, 1]^n \subset \mathbb{R}^n$. The following subsets of the cartesian plane are related to the geometrical figure of the circle:

$$\begin{array}{ll} \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\} & \textbf{unit circle} \\ \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leqslant 1\} & \textbf{closed unit disc} \\ \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 < 1\} & \textbf{open unit disc.} \end{array} \tag{2.12}$$

Thus the closed unit disc is the union of the open unit disc and the unit circle. The (unit) circle is denoted by the symbol $\mathbb{S}^1$.

Let X and Y be sets of numbers. The **algebraic sum** $X+Y$ and **product** XY (also known as **Minkowski**[6] **sum** (**product**)), are defined as follows:

$$X+Y \overset{\text{def}}{=} \{x+y : x\in X, y\in Y\} \qquad XY \overset{\text{def}}{=} \{xy : x\in X, y\in Y\} \tag{2.13}$$

with the stipulation that repeated elements are to be ignored. For example, if $X=\{1,3\}$ and $Y=\{2,4\}$, then

$$X+Y=\{3,5,7\} \qquad XY=\{2,4,6,12\}.$$

The expression 'sum of sets' is always understood as an algebraic sum. In the case of product, it is advisable to use the full expression to avoid confusion with the cartesian product.

If $X=\{x\}$ consists of a single element, then we use the shorthand notation $x+Y$ and xY in place of $\{x\}+Y$ and $\{x\}Y$, respectively. For example

$$\frac{1}{2}+\mathbb{N}=\left\{\frac{1}{2},\frac{3}{2},\frac{5}{2},\dots\right\} \qquad 3\mathbb{Z}=\{\dots,-6,-3,0,3,6,\dots\}.$$

This notation is economical and effective; it leads to concise statements such as

$$m\mathbb{Z}+n\mathbb{Z}=\gcd(m,n)\mathbb{Z}$$

(see exercise 2.5). Elementary —but significant— applications of this notation are found in **modular arithmetic**. Let m be a positive integer. We say that two integers x and y are **congruent modulo** m if m divides $x-y$. This relation is denoted by[7]

$$x\equiv y \pmod{m}.$$

Thus

$$-3\equiv 7 \pmod{5} \qquad 1\not\equiv 12 \pmod{7}.$$

The integer m is called the **modulus**. The set of integers congruent to a given integer is called a **congruence** (or **residue**) **class**. One verifies that the congruence class of x modulo m is the infinite set $x+m\mathbb{Z}$ (involving the algebraic sum and product of sets), which is given explicitly as

$$x+m\mathbb{Z}=\{x,x\pm m,x\pm 2m,x\pm 3m,\dots\}.$$

---

[6] Hermann Minkowski (Polish: 1864–1909)

[7] This notation is due to Carl Friedrich Gauss (German: 1777–1855).

For example, the odd integers are the congruence class $1+2\mathbb{Z}$. The congruence class of x modulo m is also denoted by $[x]_m$, $x \pmod{m}$, or, if the modulus is understood, by $[x]$ or $\bar{x}$.

The set of congruence classes modulo m is denoted by $\mathbb{Z}/m\mathbb{Z}$. If $m = p$ is a prime number, the notation $\mathbb{F}_p$ (meaning 'the field with p elements') may be used in place of $\mathbb{Z}/p\mathbb{Z}$. The set $\mathbb{Z}/m\mathbb{Z}$ contains m elements:

$$\mathbb{Z}/m\mathbb{Z} = \{m\mathbb{Z}, 1+m\mathbb{Z}, 2+m\mathbb{Z}, \dots, (m-1)+m\mathbb{Z}\}.$$

Variants of this notation are used extensively in algebra, where one defines the sum/product of more general sets, such as groups and rings.

2.1.3 Writing about sets

The vocabulary on sets developed so far is sufficient for our purpose. We begin to use it in short phrases which define sets.

1. *The set of ordered pairs of complex numbers.*
2. *The set of rational points on the unit circle.*
3. *The set of prime numbers with fifty decimal digits.*
4. *The set of lines in the cartesian plane, passing through the origin.*

Note that we haven't used any symbols. The set in item 1 is $\mathbb{C}^2$. In item 2, among the infinitely many points of the unit circle, we consider those having rational coordinates. There is no difficulty in writing this set symbolically:

$$\{(x,y) \in \mathbb{Q}^2 : x^2 + y^2 = 1\}$$

although its properties are not obvious from the definition. This set is non-empty (the points $(0,\pm 1)$, $(\pm 1, 0)$ belong to it), but is it infinite? This example illustrates the power of a verbal definition. Item 3, which defines a subset of $\mathbb{N}$, makes an even stronger point. This set must be extremely large, but can we even show that it is non-empty? In item 4, each line counts as a single element, rather than an infinite collection of points (otherwise our set of lines would be the whole plane). The symbolic definition of this set is awkward; to simplify it, in section 2.3 we'll consider suitable **representations** of this set.

It is possible to specify a *type* of set, without revealing its precise identity. In each of the following sets there is at least one unspecified quantity.

The set of fractions representing a rational number.

The set of divisors of an odd integer.

A proper infinite subset of the unit circle.

The sum of two finite sets of real numbers.

A finite set of consecutive integers.

The set of all partitions of a set.

A set of partitions of the natural numbers.

Next we define sets in two ways, first with words and symbols, and then with words only. One should consider the relative merits of the two formulations.

Let $X = \{3\}$.
The set whose only element is the integer 3.

Let $X = \{m\}$, *with* $m \in \mathbb{Z}$.
A set whose only element is an integer.

Let $m \in \mathbb{Z}$, *and let* X *be a set such that* $m \in X$.
A set which contains a given integer.

Let X *be a set such that* $X \cap \mathbb{Z} \neq \emptyset$
A set which contains at least one integer.

Let X *be a set such that* $\#(X \cap \mathbb{Z}) = 1$
A set which contains precisely one integer.

In the first two examples the combination of ‘let’ and ‘$=$’ replaces an assignment operator. An expression such as ‘*Let* $X \overset{\nabla}{=} \{3\}$’ would be overloaded.

The distinction between definite and indefinite articles is essential, the former describing a unique object, the latter a class of objects. In the following phrases, a change in one article, highlighted in boldface, has resulted in a logical mistake.

BAD: A proper infinite subset of **a** unit circle.

BAD: **A** set whose only element is the integer 3.

BAD: **The** set whose only element is an integer.

BAD: **The** set which contains precisely one integer.

As a final exercise, we express some geometric facts using set terminology.

The intersection of a line and a conic section has at most two points.

The set of rational points in an open interval is infinite.

A cylinder is the cartesian product of a segment and a circle.

There is no finite partition of a triangle into squares.

The reader should re-visit familiar mathematics and describe it in the language of sets.

2.2 Functions

Functions are everywhere. Whenever a process transforms a mathematical object into another object, there is a function in the background. 'Function' is arguably the most used word in mathematics.

A **function** consists of two sets together with a rule[8] that assigns to *each* element of the first set a *unique* element of the second set. The first set is called the **domain** of the function and the second set is called the **co-domain**. A function whose domain is a set A may be called a function **over** A or a function **defined on** A. The terms **map** or **mapping** are synonymous with function. The term **operator** is used to describe certain types of functions (see below).

A function is usually denoted by a single letter or symbol, such as f. If x is an element of the domain of a function f, then the **value of** f **at** x, denoted by $f(x)$ is the unique element of the co-domain that the rule defining f assigns to x. The notation

$$f : A \to B \qquad x \mapsto f(x) \tag{2.14}$$

indicates that f is a function with domain A and co-domain B that **maps** $x \in A$ **to** $f(x) \in B$. The symbol x is the **variable** or (the **argument**) of the function. The symbols $\to$ and $\mapsto$ have different meanings, and should not be confused. The function

$$I_A : A \to A \qquad x \mapsto x$$

---

[8] Below, we'll replace the term 'rule' with something more rigorous.

is called the **identity** (**function**) on A. When explicit reference to the set A is unnecessary, the identity is also denoted by Id or $\mathbb{1}$.

In definition (2.14) the symbols used for the function's name and variable are inessential; the two expressions

$$f:\mathbb{R}\smallsetminus\{0\}\to\mathbb{R}\quad x\mapsto\frac{1}{x}\qquad\qquad x:\mathbb{R}\smallsetminus\{0\}\to\mathbb{R}\quad f\mapsto\frac{1}{f}$$

define exactly the same function (even though the rightmost expression breaks just about every rule concerning mathematical notation —see section 5.2).

Functions of several variables are defined over cartesian products of sets. For example, the function

$$f:\mathbb{Z}\times\mathbb{Z}\to\mathbb{N}\qquad\qquad (x,y)\mapsto\gcd(x,y)$$

depends on two integer arguments, and hence is defined over the cartesian product of two copies of the integers. This definition requires a value for $\gcd(0,0)$, which normally is taken to be zero.

Let $f:A\to B$ be a function. The set

$$\{(x,f(x))\in A\times B\,:\,x\in A\}\tag{2.15}$$

is called the **graph** of f. A function is completely specified by three sets: domain, co-domain and graph. We can now reformulate the definition of a function, replacing the imprecise term 'rule' with the precise term 'graph'. We write a formal definition.

> DEFINITION. *A* **function** *f is a* **triple** *(X,Y,G) of non-empty sets. The sets X and Y are arbitrary, while G is a subset of $X\times Y$ with the property that for every $x\in X$ there is a unique pair $(x,y)\in G$. The quantity y is called the* **value of the function at** *x, denoted by $f(x)$.*

We see that, besides sets, the definition of a function requires the constructs of ordered pair and triple. It turns out that these quantities can be defined solely in terms of sets (see exercise 2.15). So, to define functions, all we need are sets after all.

Given a function $f:A\to B$, and and a subset $X\subset A$, the set

$$f(X)\overset{\text{def}}{=}\{f(x)\,:\,x\in X\}\tag{2.16}$$

is called the **image of X under f**. The assignment operator gives meaning to the symbolic expression $f(X)$, which otherwise would be meaningless, since we stipulated that the argument of a function is an element of the domain, not a subset of it. Thus $\sin(\mathbb{R})$ is the closed interval $[-1,1]$.

Clearly, $f(X) \subset B$, and $f(A)$ is the smallest set that can serve as co-domain for f. The set $f(A)$ is often called the **image** or the **range** of the function f. This term is sometimes used to mean co-domain, which should be avoided. A **constant** is a function whose image consists of a single point.

The notation (2.16) is suggestive and widely used. However, in computer algebra, the quantities $f(x)$ and $f(X)$ (with x an element and X a subset of the domain, respectively) are written with a different syntax, e.g., `f(x)` and `map(f,X)` with Maple.

A function is said to be **injective** (or **one-to-one**) if distinct points of the domain map to distinct points of the co-domain. A function is **surjective** (or **onto**) if $f(A) = B$, that is, if the image coincides with the co-domain. A function that is both injective and surjective is said to be **bijective.**

For any non-empty subset X of the domain A, we define the **restriction of f to X** as

$$f|_X : X \to B \qquad\qquad x \mapsto f(x).$$

Given two functions $f : A \to B$ and $g : B \to C$, the **composition** of f and g is the function

$$g \circ f : A \to C \qquad\qquad x \mapsto g(f(x)).$$

The notation $g \circ f$ reminds us that f acts before g. The image $g(f(x))$ of x under $g \circ f$ is denoted by $(g \circ f)(x)$, where the parentheses isolate $g \circ f$ as the function's symbolic name. The hybrid notation $g \circ f(x)$ should be avoided.

If $f : A \to B$ is a bijective function, then the **inverse** of f is the function $f^{-1} : B \to A$ such that

$$f^{-1} \circ f = I_A \qquad\qquad f \circ f^{-1} = I_B$$

where $I_{A,B}$ are the identities in the respective sets. A function is said to be **invertible** if its inverse exists. If $f : A \to B$ is injective, then we can always define the inverse of f by restricting its domain to $f(A)$ if necessary. Let $f : A \to B$ be a function, and let C be a subset of B. The set of points

$$f^{-1}(C) \overset{\text{def}}{=} \{x \in A : f(x) \in C\} \tag{2.17}$$

is called the **inverse image** of the set C.

Since the definition of inverse image does not involve the inverse function, the inverse image exists even if the inverse function does not. These two concepts must be carefully distinguished. When the reciprocal of a function comes into play, things get very confusing, since we now have three unrelated objects represented by closely related notation:

$$f^{-1}(x) \qquad f^{-1}(\{x\}) \qquad f(x)^{-1}.$$

The first expression is well-defined if x belongs to the image of f, and f is invertible there. In the second expression there is no condition on f, and x need only be an element of the co-domain. In the third expression the point x must belong to the domain of f, and $f(x)$ must be non-zero. Thus

$$\sin^{-1}(1) = \frac{\pi}{2} \qquad \sin^{-1}(\{1\}) = \frac{\pi}{2} + 2\pi\mathbb{Z} \qquad \sin(1)^{-1} = \csc(1).$$

In the first expression we tacitly assume that $\sin^{-1} = \arcsin : [-1,1] \to [-\pi/2, \pi/2]$. In the third expression the symbol csc denotes the cosecant ($\csc(x) = 1/\sin(x)$), defined in the domain $\mathbb{R} \smallsetminus \pi\mathbb{Z}$. The hybrid expression

$$\sin^{-1}(1) = \frac{\pi}{2} + 2\pi\mathbb{Z},$$

is, strictly speaking, incorrect, but it is also unambiguous and simpler than the correct expression. Some matematicians will reject it with outrage, others will consider it an acceptable transgression.

Let us use the word ‘function’ in short expressions. These are function definitions:

1. *The integer function that squares its argument.*
2. *The function that returns 1 if its argument is rational, and 0 otherwise.*
3. *The function that counts the number of primes smaller than a given real number.*
4. *The function that gives the distance between two points on the unit circle, measured along the circumference.*

We surmise that the function in item 2 is defined over the real numbers. Item 3 is a much-studied function in number theory. The image of the function in item 4 is the closed interval $[0, \pi]$.

With a judicious use of definite and indefinite articles, we can specify a function’s type without committing ourselves to a specific object.

1. *The inverse of a trigonometric function.*
2. *The composition of a function with itself.*
3. *An integer-valued bijective function.*
4. *A function which coincides with its own inverse.*

In item 2, we infer that the function maps its domain into itself. Functions of type 3 will be considered in the next section to define cardinality of sets. Functions of type 4 are called **involutions.**

Writing about **real functions** is considered in chapter 4.

2.3 Representations of sets

To be able to work with **abstract sets**, we need concrete **representations** of them. Representing a set means identifying its elements with a collection of familiar objects, such as vectors or matrices. This identification gives a description of a set in terms of another set, presumably easier to handle. For instance, a representation may provide the data structure for computer implementation.

Two sets A and B are said to be **equivalent** (written $A \sim B$) if there is a bi-unique correspondence between the elements of A and the elements of B, namely, if there exists a bijective function $f : A \to B$. Equivalent sets have the same cardinality, and vice-versa. The cardinality of a set is defined using this equivalence. A set equivalent to $\{1, 2, \ldots n\}$ is said to have **cardinality** n, and a set equivalent to $\mathbb{N}$ is said to be **countable** or **countably infinite**. The set $\mathbb{Z}$ is countable, and so is $m\mathbb{Z}$, for any $m \in \mathbb{N}$. A set X is **uncountable** if it contains a countably infinite subset Y, but X is not equivalent to Y. The set $\mathbb{R}$ is uncountable. (We see that characterising the cardinality of infinite sets requires a more sophisticated approach than mere ‘counting’.)

A **representation** of a set A is any set B which is equivalent to A. (This is the most general acceptation of the term representation. In algebra, representations are based on a stronger notion of equivalence than the one given above.)

For instance, the open unit interval and the real line are equivalent, as established by the bijective function

$$f : \mathbb{R} \to (0, 1) \qquad x \mapsto \frac{1}{\pi}\arctan(x) + \frac{1}{2}. \tag{2.18}$$

Likewise, the exponential function establishes the equivalence $\mathbb{R} \sim \mathbb{R}^+$.

We consider some representation problems. Let L be the set of lines in the plane passing through a point (a,b). This set is infinite, and its meaning is easily grasped. Each element λ of L is an infinite subset set of $\mathbb{R}^2$, which we write symbolically as

$$\lambda = \big\{(x,y) \in \mathbb{R}^2 \,:\, y = b + s(x-a)\big\}$$

where s is a real number representing the line's slope. The line $x = a$ is not of this form, and must be treated separately. Collecting all the lines together, we obtain a symbolic description of L

$$L = \Big\{\big\{(x,y) \in \mathbb{R}^2 \,:\, y = b + s(x-a)\big\} \,:\, s \in R\Big\} \cup \Big\{\big\{(a,y) \in \mathbb{R}^2 \,:\, y \in \mathbb{R}\big\}\Big\}.$$

The simple verbal definition of L seems to have drowned in a sea of symbols!

We look for a set equivalent to L with a more legible structure. An obvious simplification results from representing L as a set of **cartesian equations**:

$$L \sim \{y = b + s(x-a) \,:\, s \in \mathbb{R}\} \cup \{x = a\}.$$

We have merely replaced the set of solutions of an equation with the equation itself (cf. section 2.5). This identification provides the desired bi-unique correspondence between the two sets.

We can simplify further. Because a and b are given, hence fixed, there is no need to specify them explicitly; it suffices to give the (possibly infinite) value of the slope. Alternatively, we could identify a line by an angle θ between 0 and π, measured with respect to some reference axis passing through the point (a,b). Because the angles 0 and π correspond to the same line, only one of them is to be included, resulting in the half-open interval $[0,\pi)$. The equivalence between $\mathbb{R} \cup \{\infty\}$ and $[0,\,\pi)$ may be achieved with a transformation of the type (2.18), where the included end-point 0 corresponds to the point at infinity.

Finally, any half-open interval may be identified with the **circle** $\mathbb{S}^1$, by gluing together the end-points of the interval. In our case this is achieved with the function $\theta \mapsto (\cos(2\theta), \sin(2\theta))$. The essence of our set is now clear:

$$L \sim \mathbb{R} \cup \{\infty\} \sim [0,\,\pi) \sim \mathbb{S}^1.$$

As a second example, let us consider the set Γ of open segments in the plane. Each segment is identified by its end-points, and each end-point is specified by a pair of real numbers. It would seem that $\Gamma \sim \mathbb{R}^4$, but the correspondence between these sets is not bi-unique, because a segment does not change if we interchange the end-points. Furthermore, if the end-points are the same, we obtain the empty set, not a segment.

Rather than removing from $\mathbb{R}^4$ the unwanted duplications, we change representation. We identify a segment via its mid-point z (a pair of real numbers), length r (a positive real number), and orientation θ (an angle between 0 and π). We see that

$$\Gamma \sim \tilde{\Gamma} \qquad \text{where} \qquad \tilde{\Gamma} := \mathbb{R}^2 \times \mathbb{R}^+ \times [0, \pi).$$

Consider now the subset U of Γ consisting of all segments of unit length. Using our representation, we have

$$U \sim \{(z, r, \theta) \in \tilde{\Gamma} : r = 1\} \sim \mathbb{R}^2 \times [0, \pi).$$

In the last equivalence, we have removed the idle variable r, because its value is fixed.

2.4 Sequences

A **sequence** is an ordered list of objects, not necessarily distinct, called the **terms** (or the **elements**) of the sequence. The terms of a sequence are represented by a common symbol, and each term is identified by an integer **subscript**:

$$(a_1, a_2, \ldots, a_n) \qquad\qquad (a_1, a_2, \ldots). \tag{2.19}$$

Here the common symbol is a, and the integer values assumed by the subscript begin from 1. The quantity a_1 reads *"a sub 1"*, etc. Subscripts may also begin from 0, or from anywhere. The expression on the left denotes a finite sequence, the one on the right suggests that the sequence is infinite.

The **length** of a sequence is the number of its elements. Two sequences are equal if they have the same length, and if the corresponding terms are equal. If k is an unspecified integer, then a_k is called the **general term** of the sequence.

For example, the sequence of primes

$$(p_1, p_2, p_3, \ldots) = (2, 3, 5, \ldots)$$

is infinite. The general term p_k is the kth prime number.

There are other notations for sequences; they display the general term alongside information about the subscript range:

$$(a_k)_{k=1}^n \qquad (a_k)_1^n \qquad (a_k)_{k=1}^\infty \qquad (a_k)_{k\geqslant 1} \qquad (a_k). \tag{2.20}$$

There are also the **doubly-infinite sequences,** where the subscript runs through all the integers:

$$(a_k)_{k=-\infty}^{\infty} = (\ldots, a_{-1}, a_0, a_1, \ldots).$$

In section 5.2 we shall discuss the usage of these notations.

The ellipsis is ubiquitous in sequence notation. When we use it, we must make sure that the missing terms are defined unambiguously. Thus the general term of the sequence of monomials

$$(2x, 2x^2, 2x^3, \ldots)$$

is clearly equal to $2x^k$. However, the expression

$$(3, 5, 7, \ldots)$$

is ambiguous, because there are several plausible alternatives for the identity of the omitted terms, such as $(9, 11, 13, \ldots)$ or $(11, 13, 17, \ldots)$. In the former case, we resolve the ambiguity by displaying the general term

$$(3, 5, \ldots, 2k+1, \ldots).$$

In the latter, we need an accompanying sentence.

A **subsequence** of a sequence (a_k) is any sequence obtained from (a_k) by deleting terms. For instance, the primes that give remainder 1 upon division by 4 form a subsequence of the sequence of primes

$$(5, 13, 17, 29, \ldots).$$

Some types of finite sequences have a dedicated terminology. A two-element sequence is an (**ordered**) **pair**, and a three-element sequence a **triple.** Occasionally one sees the terms **quadruple** or **quintuple** (I wouldn't go much beyond that), while an n-element sequence may be called an n**-uple**. A finite sequence of numbers may be called a **vector**, in which case we speak of **dimension** rather than length.

An infinite sequence $(a_1, a_2, \ldots)$ represents a **function** defined over the natural numbers. If the elements of the sequence belong to a set A, then such a function is defined as

$$a : \mathbb{N} \to A \qquad\qquad k \mapsto a_k.$$

We see that in the expression a_k, the symbol a is the function's name, the subscript k is an element of the domain, and $a_k \in A$ is the value $a(k)$ of the function at k. This interpretation clarifies the meaning of expressions such as a_{k^2}: it is the composition of two functions, much like $\sin(x^2)$.

Finally, we consider **sets of sequences** and associated notation. Let A be any set. We form the set of all *finite* sequences of elements of A with a given number n of elements. This set is the cartesian product A^n of n copies of the set A: the first element a_1 of a sequence is chosen from the first copy of A, the second element from the second copy of A, and so on. For instance, the set $\{0,1\}^n$ is the set of all binary sequences with n-digits, while $\mathbb{Q}^n$ is the set of all n-uples of rational numbers. It follows that the infinite union

$$\bigcup_{n\geqslant 1} \mathbb{Z}^n$$

represents the set of all finite integer sequences.

The set of all infinite sequences of elements of A has the structure of a cartesian product with infinitely many terms. It would seem natural to denote it by A^∞. However, the idiomatic notation $A^{\mathbb{N}}$ is more common, due to its greater flexibility. Thus $A^{\mathbb{Z}}$ denotes the set of doubly-infinite sequences of elements of A, and one even finds $A^{\mathbb{Z}^2}$ for the set of two-dimensional arrays: $(a_{i,j}) \in A^{\mathbb{Z}^2}$.

2.4.1 Some constructs involving sequences

Let (A_k) be a sequence of sets, which may be finite or infinite. The binary set operations of union and intersection generalise to an arbitrary number of operands as follows:

$$\bigcup_k A_k = A_1 \cup A_2 \cup \cdots \qquad \bigcap_k A_k = A_1 \cap A_2 \cap \cdots .$$

These expressions denote the set of elements belonging to at least one of the sets A_k, and to every one of the sets A_k, respectively. As with any binary operator, the implied unions and intersections are represented by the raised ellipsis. A sequence of sets is **descending** (or **nested**) if

$$A_1 \supset A_2 \supset A_3 \supset \cdots$$

and **ascending** if

$$A_1 \subset A_2 \subset A_3 \subset \cdots .$$

Example. The following symbols

$$m\mathbb{Z} \supset m^2\mathbb{Z} \supset \cdots \supset m^k\mathbb{Z} \supset \cdots \qquad\qquad \bigcap_{k\geqslant 1} m^k\mathbb{Z} = \begin{cases} \{0\} & \text{if } |m| \neq 1 \\ \mathbb{Z} & \text{if } |m| = 1 \end{cases}$$

tell a story. Say it with words **[$\not\in$]**.

> *The set of multiples of the power of an integer contains the set of multiples of a higher power of the same integer. By considering increasing powers we obtain an infinite nested sequence of sets. Apart from trivial cases, the only element common to all these sets is 0.*

Given a finite sequence of numbers $(a_1, \dots, a_n)$, we form the sum and the product of its elements

$$\sum_{k=1}^{n} a_k = a_1 + a_2 + \cdots + a_n \qquad \prod_{k=1}^{n} a_k = a_1 \times a_2 \times \cdots \times a_n. \tag{2.21}$$

Again, the raised ellipses represent repeated additions and multiplications, respectively. The symbol $\sum$ is called the **summation symbol.** The subscript k is the **index** of summation, while 1 and n are, respectively, the **lower limit** and **upper limit** of summation. The quantity a_k is the **general term** of the sum. The integer sequence $(1, 2, \dots, n)$, specifying the values assumed by the index of summation, is called the **range** of summation. The symbol $\prod$ is called the **product symbol,** and all terminology introduced for sums extends with obvious modifications to products.

EXAMPLE. For $n \geqslant 0$, the n**-dimensional unit sphere** $\mathbb{S}^n$ is defined as follows

$$\mathbb{S}^n = \{(x_0, \dots, x_n) \in \mathbb{R}^{n+1} : x_0^2 + \cdots + x_n^2 = 1\}$$

This Zermelo definition, to be compared with the definition (2.12) of the unit circle $\mathbb{S}^1$, employs finite sequences notation, with a combination of ordinary and raised ellipses. For $n = 0$, we have $\mathbb{S}^0 = \{-1, 1\}$.

A sum with infinitely many summands is called a **series**. If the limit of finite sums of type (2.21)

$$\lim_{n \to \infty} \sum_{k=1}^{n} a_k$$

exists, then such a limit is called the **sum of the series**, and the series is said to **converge**. Otherwise the series **diverges**. If a series has non-negative terms, then convergence is sometimes expressed with the suggestive notation (cf. (2.3), p. 14)

$$\sum_{k \geqslant 0} a_k < \infty. \tag{2.22}$$

An infinite product is called just that, and its value is defined as the limit of a sequence of finite products. Infinite products are often written in the form

$$\prod_{k \geqslant 0} (1 + a_k)$$

because convergence requires that $a_k \to 0$. Sum and product notation will be considered again in section 5.3.

Let A be a set of numbers, and let $(a_0, \ldots, a_n)$ be a finite sequence of elements of A with $a_n \neq 0$. A **polynomial** over A in the **indeterminate** x is an expression of the type

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n. \tag{2.23}$$

The elements of the sequence are called the **coefficients** of the polynomial, and the integer n is its **degree.** The coefficients a_0 and a_n are called, respectively, the **constant** and the **leading coefficient.** Each addendum in a polynomial is called a **monomial**, and a polynomial with two terms is a **binomial**. A polynomial of degree two is said to be **quadratic**; then we have **cubic**, **quartic**, and **quintic** polynomials.

The set of all polynomials over the set A with indeterminate x is denoted by $A[x]$. For example

$$x^2 - x - 1 \in \mathbb{Z}[x] \qquad\qquad \frac{1}{2} - y^3 \in \mathbb{Q}[y].$$

A **rational function** is the ratio of two polynomials

$$\frac{a_0 + a_1 x + \cdots + a_n x^n}{b_0 + b_1 x + \cdots + b_m x^m}. \tag{2.24}$$

Its degree is the largest of m and n (assuming that $a_n b_m \neq 0$). The set of all rational functions with coefficients in a set A and indeterminate x is denoted by $A(x)$.

A **multivariate polynomial** is a polynomial in more than one indeterminate. (The term **univariate** is used to differentiate from multivariate.)

$$x^2 y^2 - \frac{1}{2} x^4 - x y^3 \in \mathbb{Q}[x, y].$$

The **total degree** of each monomial is the sum of the degrees of the indeterminates, and the degree of a polynomial is the largest total degree among the monomials with non-zero coefficient. A multivariate polynomial is **homogeneous** if all monomials have the same total degree. The expression above may be described as

> *A homogeneous quartic polynomial in two indeterminates with rational coefficients.*

If we replace the finite sum (2.23) with an infinite sum, we obtain a (**formal**) **power series**.

$$\sum_{k \geqslant 0} a_k x^k \tag{2.25}$$

This is a 'polynomial of infinite degree'. The attribute 'formal' is used if we are not concerned with assigning specific values to the indeterminate x. In this case the power series, like a polynomial, is an algebraic object. For the values of x for which the series (2.25) converges, the power series represents a function.

EXAMPLE. Explain what is a polynomial. **[$\not\epsilon$]**

> *A polynomial is a finite sum. Each term, called a monomial, is the product of a coefficient (typically, a real or complex number) and one or more indeterminates, each raised to some non-negative integer power.*

2.5 Equations

Let f and g be functions with the same domain X and co-domain Y. In the most general setting, an **equation** (**on** or **over** X) is an expression of the type

$$f(x) = g(x). \tag{2.26}$$

The quantity x is the equation's **unknown**. The expression (2.26) defines a property that each point $x \in X$ either has or doesn't have[9]. This prompts the definition of the **solution set** of equation (2.26), given by

$$\{x \in X \,:\, f(x) = g(x)\}.$$

For instance, the expression $x^2 - 3x + 1 = -(1 + 3x)$ is an equation. Over the complex numbers, the solution set is $\{\sqrt{-2}, -\sqrt{-2}\}$, while over $\mathbb{R}$ the solution set is empty. We see that the solution set of an equation depends on the ambient set.

If the co-domain Y of f and g is a set of numbers (e.g., $Y = \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$), then, by replacing f by $f - g$, we can write equation (2.26) in the simpler form

$$f(x) = 0. \tag{2.27}$$

An element x of the solution set of this equation is called a **zero** of f, but if $f(x)$ is a polynomial, we speak of a **root** of f. We also say that f **vanishes at** x. A function f **vanishes identically** on a set if it vanishes at every point of this set. For example, the real function $x \mapsto \sin(\pi x)$ vanishes identically on $\mathbb{Z}$.

More generally, an equation may be reduced to the form (2.27) if the co-domain of f and g is a **group** with respect to addition. In this case the zero on the right-hand side is the zero element of the group. This is not necessarily the number 0, but it could be a zero matrix, a zero polynomial, a zero function, etc.

---

[9] In chapter 3 we shall see that an equation is a special type of **predicate**, which in turn is a special type of **function**.

An equation whose unknown is a set is called a **set equation**. In the expression

$$(x+x)\cap x=\emptyset \tag{2.28}$$

the quantity $x+x$ can only be an algebraic sum of sets, so this is a set equation. It may be defined over any set of number sets, say, $\mathbf{P}(\mathbb{Z})$, the power set of the integers. A solution of this equation is called a **sum-free set**; thus the set of odd integers is sum-free. If we interpret the empty set as 'zero', then equation (2.28) is again of the form (2.27). There exists a procedure to transform any set equation to this form (see exercise 2.16).

An equation whose unknown is a function is called a **functional equation**

$$f(x)=f(x+1) \qquad\qquad f(f(x))=x. \tag{2.29}$$

Here the unknown is f (not x), which belongs to a suitable **set of functions** (such as the set $C^0(\mathbb{R})$ of continuous real functions or the set $C^1(\mathbb{R})$ of differentiable real functions). To make it evident that f is the unknown, let $\mathbb{1}$ be the identity function (over $\mathbb{R}$, say), and let τ be the function that increases its argument by 1: $\tau(x)=x+1$. Equations (2.29) may be rewritten without reference to the argument of f as follows:

$$f=f\circ\tau \qquad\qquad f\circ f=\mathbb{1}. \tag{2.30}$$

If a functional equation is in the standard form (2.27), then the zero on the right-hand side is the **zero function**, namely the constant function that assumes the value zero everywhere.

A **differential equation** is a functional equation that involves derivatives of the unknown:

$$a)\quad \frac{dy}{dx}=y \qquad\qquad b)\quad x^2\frac{dy^2}{dx^2}+x\frac{dy}{dx}+(x^2-1)y=0. \tag{2.31}$$

Here the unknown is y (not x), and $y=y(x)$ is a function in a suitable set of functions. The **order** of a differential equation is the order of the highest derivative appearing in the equation. Thus the order of equation $a)$ is 1 and that of $b)$ is 2. The standard form of a differential equation of order n is the following variant of (2.27):

$$f(x,y,y^{(1)},y^{(2)},\dots,y^{(n)})=0 \tag{2.32}$$

The quantity $y^{(k)}$ denotes the k-derivative of y. The only unknown is y, since its derivatives clearly depend on it. The argument x of y is listed if it appears explicitly, as in equation $b)$. The differential equation (2.32) is defined over a set of functions that are differentiable n times. Again, the symbol 0 on the right-hand side of the equation represents the zero function.

An equation whose solution set is equal to the ambient set is called an **identity,** or an **indeterminate equation**. After simplification, every identity reduces to the standard form $0 = 0$. This doesn't mean that identities are trivial, far from it; identities express the equivalence of functions. However, they are ephemeral quantities, which disappear if they are simplified.

EXAMPLE. The identity

$$x^{2^n} - y^{2^n} = (x-y)\prod_{k=0}^{n-1}(x^{2^k} + y^{2^k})$$

gives the full factorisation of the difference of two monomials whose degree is a power of 2, into the product of polynomials with integer coefficients.

EXAMPLE. Over the set $\mathbb{R}^2$, the expression $x + y = y + x$ is an identity, representing the commutativity of the addition of real numbers. The similar expression $x + y = 1 - y$ is an equation, whose solution set is a line in $\mathbb{R}^2$.

By restricting the ambient set to the solution set, every equation becomes an identity. Clearly, $\sin(\pi x) = 0$ is an equation over $\mathbb{R}$ and an identity over $\mathbb{Z}$. For a more meaningful example, consider the equation $f(x) = x^5 - x = 0$. The factorisation $x^5 - x = x(x-1)(x+1)(x^2+1)$ shows that the solution set over $\mathbb{C}$ is $\{0, \pm 1, \pm\sqrt{-1}\}$. Consider now the set

$$X = \mathbb{Z}/5\mathbb{Z} = \{0 + 5\mathbb{Z},\, 1 + 5\mathbb{Z},\, 2 + 5\mathbb{Z},\, 3 + 5\mathbb{Z},\, 4 + 5\mathbb{Z}\}$$

of congruence classes modulo 5. Let us evaluate our function f at all points of $\mathbb{Z}/5\mathbb{Z}$, writing k for $k + 5\mathbb{Z}$:

$$\begin{aligned}
f(0) &= 0 \equiv 0 \pmod 5\\
f(1) &= 0 \equiv 0 \pmod 5\\
f(2) &= 30 \equiv 0 \pmod 5\\
f(3) &= 240 \equiv 0 \pmod 5\\
f(4) &= 1020 \equiv 0 \pmod 5.
\end{aligned}$$

The function f vanishes identically over $\mathbb{Z}/5\mathbb{Z}$, and hence the expression $x^5 = x$ is an identity!

An expression of the type

$$\begin{cases} f_1(x) = 0 \\ f_2(x) = 0 \\ \quad\vdots \\ f_n(x) = 0 \end{cases} \qquad x = (x_1, \dots, x_m) \tag{2.33}$$

where all functions have the same domain and co-domain, is called a **system of** n **simultaneous equations in** m **unknowns.** The solution set of a system of equations is the intersection of the solution sets of the individual equations (see example 3.3, p. 59).

EXAMPLE. Explain what is an equation, and its solutions. **[$\not\mathcal{E}$]**

BAD: *An equation is when we equate two functions. The solution is when the functions are the same.*

The inappropriate use of 'when' is easily spotted (see section 1.1), but there is a more serious flaw. The expression 'equating two functions' means that we seek conditions under which the two functions become the same function. That's not what we had in mind. The operands of the equal sign in expression (2.26) are not functions, but rather values of functions.

GOOD: *An equation is an expression that identifies the value of two functions at a generic point of their common domain. The solutions of an equation are the points at which the two functions assume the same value.*

(The expression 'equating two functions' may be appropriate for **functional equations**, see (2.29).)

2.6 Expressions

The generic term **expression** indicates the symbolic encoding of a mathematical object. For instance, the string of symbols '$2+3$' is a valid expression, and so is '$x \mapsto f(x)$', while '$2+\times 3$' is incorrect and does not represent an object.

It would seem that any correct expression should have —in principle, at least— a unique **value**, representing some agreed 'fully simplified' form of the expression. For instance, it could be argued that the value of $13/91$ is $1/7$, and that of $\sqrt{2187}$ is $27\sqrt{3}$. The simplified value is more concise and informative, and would enable us —among other things— to recognise when two expressions represent the same object.

This is not so simple. For example, the two expressions

$$\sqrt{3}-\sqrt{2} \qquad\qquad \sqrt{5-2\sqrt{6}}$$

have the same value, yet there is no compelling reason for choosing one over the other. The same object may be viewed from different angles, and our choice of representation will depend on the context.

The following well-known identity makes an even stronger point:

$$\frac{1-x^n}{1-x} = 1+x+x^2+\cdots+x^{n-1}.$$

The right-hand side —the sum of n monomials— is the ‘fully simplified’ version, while the ‘unsimplified’ left-hand side has only four terms. While simplification reveals the true nature of the object —a polynomial as opposed to a rational function— it’s the unsimplified value which provides information about the sum.

Given that agreeing on a unique value of an expression proves difficult, we shift our attention to the **type** of value (set, number, function, etc.), which assigns the expression to a certain class of objects. This class provides the broadest characterisation of the object in question. Again, we must exercise some judgement. The expressions

$$1+1 \qquad\qquad \int_0^{\pi} \sin(x)dx$$

have the same value, but their structure is so different, that the coincidence of their values seems secondary. Whereas the expression on the left is unquestionably ‘*a number*’, or ‘*a positive integer*’, that on the right is ‘*a definite integral.*’ On the other hand, there may be circumstances in which the reductionist description of the integral as a number is appropriate, for instance when discussing integrability of functions.

Keeping in mind the difficulties mentioned above, we now turn to the description of mathematical expressions.

2.6.1 Levels of description

We develop the idea of successive refinements in the description of an expression, from the general to the particular. The appropriate level of details to be included will vary, depending on the situation. We treat in parallel verbal and symbolic descriptions, as far as it is reasonable to do so.

Let us consider the definition of a set. The coarsest level of description is

$\{\dots\}$ *A set*

where the object's type is identified by the curly brackets. The use of the indeterminate article —'a' set rather than 'the' set— reflects our incomplete knowledge.

The next level in specialisation identifies the ambient set:

$\{(x,y) \in \mathbb{Z}^2 \,:\, \dots\}$ *A set of integer pairs*

Now we begin to build the defining properties of our set:

$\{(x,y) \in \mathbb{Z}^2 \,:\, \gcd(x,y) = 1, \dots\}$ *A set of pairs of co-prime integers*

The final step completes the definition:

$\{(x,y) \in \mathbb{Z}^2 \,:\, \gcd(x,y) = 1,\ 2|xy\}$ *The set of pairs of co-prime integers, with exactly one even component*

Accordingly, the indefinite article has been replaced by the definite article. Now both words and symbols describe one and the same object, and one should consider the relative merits of the two presentations. The term 'exactly' is, strictly speaking, redundant, but it helps the reader realise that x and y cannot both be even. A robotic translation of symbols into words

> *The set of elements of the cartesian product of the integers with themselves, whose components have greatest common divisor equal to 1, and such that 2 divides the product of the components*

lacks the synthesis that comes with understanding.

Expression may be **nested**, like boxes within boxes. We begin with two expressions

$\left(\cdots\right)^2$ *A square*

$\sum\cdots$ *A sum*

We only see the outer structure of these objects. We compose them in two different ways

$\left(\sum\cdots\right)^2$	*The square of a sum*
$\sum\left(\cdots\right)^2$	*A sum of squares*

The first term in each expression identifies the object's outer layer. There is still one indefinite article, reflecting a degree of generality. We specialise further:

$\displaystyle\sum_{n=1}^{\infty}\left(\frac{1}{n}\right)^2$	*The sum of the square of the reciprocal of the natural numbers.*

Words or symbols now define a unique object, with three levels of nesting. By contrast, in the nested expression

$\displaystyle\left(\sum_{n=1}^{\infty}a_n\right)^2$	$a_n \in \mathbb{Q}$	*The square of the sum of the elements of a rational sequence,*

the innermost object —a rational sequence— is still generic.

In these examples words and symbols are interchangeable; in actual writing, some concepts are best expressed with words, others with symbols, while most of them require both. For instance, the symbolic expression

$$(1-x, 2x+x^2, \ldots, nx^{n-1}+(-x)^n, \ldots) \tag{2.34}$$

defines an infinite sequence succinctly and unambiguously. Using words, we could begin to describe it as follows:

A sequence
An infinite sequence
An infinite sequence of binomials

Increasing further the accuracy of the verbal description is pointless, since the symbolic expression (2.34) is clearly superior in delivering exact information. On the other hand, words can place this expression *in a context,* which is something symbols cannot do. To illustrate this point, we supplement the description given above with additional information which emphasises a particular property.

An infinite sequence of binomials

with integer coefficients
with unbounded coefficients
with increasing degree
whose leading term alternates in sign.

2.6.2 Characterising expressions

We expand our dictionary with terms describing broad attributes of expressions.

An expression involving numbers, the four arithmetical operations, and raising to an integer or fractional power (extraction of roots), is called an **arithmetical expression.** The value of an arithmetical expression is a number. A combination of rational numbers and square roots of rational numbers is called a **quadratic irrational** or a **quadratic surd**. The following expressions are arithmetical.

$191861^2 - 3 \cdot 110771^2 = -2$	*An arithmetic identity, with a surprising cancellation.*
$\dfrac{3+2\sqrt{2}}{8-3\sqrt{7}}$	*The ratio of two quadratic surds having distinct radicands.*

If indeterminates are present, we speak of an **algebraic expression**.

$\dfrac{\sqrt[6]{ab-(ab)^{-1}}}{\sqrt[3]{a^2b^2+ab+1}}$	*An algebraic expression with two indeterminates and higher-index roots.*

Polynomials and rational functions are algebraic expressions (see section 2.4.1). They may also be characterised as **rational expressions**, since they don't involve fractional powers of the indeterminates.

$\dfrac{1}{x+1} + \dfrac{1}{x^2+1} + \cdots + \dfrac{1}{x^n+1}$	*The sum of finitely many rational functions, with increasing degree.*
$\dfrac{\big((x^2+1)^2+1\big)^2+1}{(x^2+1)^2+1}$	*A rational expression, involving repeated compositions of a polynomial function with itself.*

The following mathematical Russian doll

$$\sqrt{x+\sqrt{2x+\sqrt{3x+\cdots+\sqrt{nx}}}} \qquad n \in \mathbb{N}$$

could be described as

An algebraic expression in one indeterminate, involving a finite number of nested square roots.

The functions sine, cosine, tangent, secant, etc., are called **trigonometric functions** (or **circular functions**). A **trigonometric expression** is an expression containing trigonometric functions

$8\cos(z)^4 - 8\cos(z)^2 + 1$	*A quartic trigonometric polynomial.*

Trigonometric functions belong to the larger class of **transcendental functions**, which are functions not definable by an algebraic expression (the exponential, the logarithm, etc.).

An equation defined by algebraic expressions is called an **algebraic equation.** Likewise, we speak of **trigonometric** and **transcendental** equations.

$x^n - x - 1 = 0$	*An algebraic equation.*
$\cos(x) = \sin(x)$	*A trigonometric equation.*
$\log(1+x) = -x$	*A transcendental equation.*

The term **analytical expression** is used in the presence of infinite processes.

$\sqrt{1+x} = 1 + \frac{1}{3}x - \frac{1}{9}x^2 + \frac{5}{81}x^3 + \cdots$	*The first few terms of the series expansion of an algebraic function*
$\lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n = e$	*Napier's constant as the limit of a sequence of rational numbers*
$2\prod_{k=1}^{\infty}\frac{(2k)^2}{(2k)^2 - 1} = \pi$	*An infinite product formula for Archimedes' constant.*

An **integral expression** is an expression involving integrals

$\ln(x) = \int_0^x \frac{1}{t}dt$	*An integral expression for the natural logarithm.*

The term **combinatorial** is appropriate for expressions involving counting functions, such as the **factorial** function, or the **binomial coefficient**.

$\frac{1}{2^{2k}}\frac{(2k)!}{(k!)^2}$	*A rational combination of exponentials and factorials.*
$\sum_{k=0}^{n}\binom{n+k}{k}$	*A finite sum of binomial coefficients.*

In chapter 3 we deal with the expressions found in logic: the **boolean** expressions.

Exercises

Exercise 2.1. For each of the following topics:

prime numbers, fractions, complex numbers,

i) write five short sentences; *ii*) ask five questions. The sentences should give a definition or state a fact; the questions should have mathematical significance, and preferably possess a certain degree of generality. **[∉]**[10]

Exercise 2.2. Define five interesting finite sets. **[∉]**

Exercise 2.3. The following expressions define sets. Turn words into symbols, using standard or Zermelo definitions.
[*Represent geometrical objects, e.g., planar curves, by their cartesian equations.*]

1. *The set of negative odd integers.*
2. *The set of natural numbers with three decimal digits.*
3. *The set of rational numbers which are the ratio of consecutive integers.*
4. *The set of rational points in the closed unit cube.*
5. *The complement of the open unit disc in the complex plane.*
6. *The set of vectors of unit length in three-dimensional euclidean space.*
7. *The set of circles in the plane, passing through the origin.*
8. *The set of hyperbolae in the plane, whose asymptotes are the coordinate axes.*
9. *The set of lines tangent to the unit circle.*

Exercise 2.4. The following expressions define sets. Turn symbols into words. **[∉]**

1. $\{x \in \mathbb{Q} \,:\, 0 < x < 1\}$
2. $\{1/(2n+1) \,:\, n \in \mathbb{Z}\}$
3. $\{m2^{-k} \,:\, m \in 1 + 2\mathbb{Z},\ k \in \mathbb{N}\}$
4. $\{x \in \mathbb{R} \smallsetminus \mathbb{Z} \,:\, x^2 \in \mathbb{Z}\}$
5. $\{z \in \mathbb{C} \smallsetminus \mathbb{R} \,:\, z^2 \in \mathbb{R}\}$

---

[10] This exercise must be completed without using any mathematical symbol.

6. $\{z \in \mathbb{C} : |\mathrm{Re}(z)| + |\mathrm{Im}(z)| \leqslant 1\}$
7. $\{(m,n) \in \mathbb{Z}^2 : m \,|\, n\}$
8. $\{(x,y,z) \in \mathbb{R}^3 : xyz = 0\}$
9. $\{(x_1, \dots, x_n) \in \mathbb{R}^n : \sum_k x_k = 0\}$
10. $\{x \in \mathbb{R} : \sin(2\pi x) = 0\}$
11. $\{(x,y) \in \mathbb{R}^2 : \sin(\pi x)\sin(\pi y) = 0\}$.

Exercise 2.5. Turn symbols into words. **[**$\not\varepsilon$**]**

1. f^{-1}
2. $f^{-1}(x)$
3. $f(x^{-1})$
4. $f(x)^{-1}$
5. $f(x)^2$
6. $f \circ f$
7. $f(\mathbb{R} \smallsetminus \mathbb{Q})$
8. $f(A) \cap f(B)$
9. $f^{-1}(\{0\})$
10. $f(\mathbb{R}) \cap \mathbb{Q}$.

Exercise 2.6. For each item, provide two levels of description: **[**$\not\varepsilon$**]**

i) a coarse description, which only identifies the object's type (set, function, polynomial, etc);
ii) a finer description, which defines the object in question, or characterises its structure.

1. $3^3 + 4^3 + 5^3 = 6^3$
2. $\sqrt{3}/(\sqrt[3]{3}/(\sqrt[4]{3}))$
3. $\frac{7}{5} < \sqrt{2} < \frac{17}{12}$
4. $x^3 - x - 1$
5. $xy > 0$
6. $(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3$
7. $y - x^2 - x = 0$
8. $\dfrac{d^2}{dt^2}x - 3\dfrac{d}{dt}x - 2 = 0$

9. $\sin(x-y) = \sin(x)\cos(y) - \cos(x)\sin(y)$
10. $A \cup B = B \cup A$
11. $X = \{X\}$.

Exercise 2.7. Same as in previous exercise.

1. $2Z \setminus 4\mathbb{Z}$
2. $\{\{\emptyset\}\}$
3. $(\mathbb{Q} \setminus \mathbb{Z})^2$
4. $\mathbb{R}^2 \setminus \mathbb{Q}^2$
5. $(\mathbb{R} \times \mathbb{Z}) \cup (\mathbb{Z} \times \mathbb{R})$
6. $1 + 2(1 + 2(1 + 2\mathbb{Z}))$
7. $(\{a\}, \{a\}, \{a\}, \dots)$
8. $(1, -3, 5, \dots, (-1)^n(2n+1), \dots)$
9. $(1, 2^{2k}, 3^{2k}, \dots, n^{2k})$
10. $((x_1), (x_1, x_2), (x_1, x_2, x_3), \dots)$
11. $\underbrace{[0,1] + [0,1] + \cdots + [0,1]}_{n}$.

Exercise 2.8. Same as in previous exercise.

1. $f : \mathbb{R} \to \mathbb{R}, \quad x \mapsto x + 1$
2. $\int_0^\infty \frac{f(x)}{g(x)} dx$
3. $\frac{d}{dx} f(x)g(x) = \frac{df(x)}{dx} g(x) + \frac{dg(x)}{dx} f(x)$
4. $f(x) = \int_0^x \frac{df(y)}{dy} dy$
5. $\int F(x,y) dx$

6. $\int_0^1 dx \int_0^1 H(x,y)dy$

7. $\cos(x) = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!}$

8. $\frac{\partial F(x,y,z)}{\partial x} + \frac{\partial F(x,y,z)}{\partial y} + \frac{\partial F(x,y,z)}{\partial z}$

9. $\prod_{k=1}^{n} \frac{\partial \mathscr{L}(\theta_1,\ldots,\theta_n)}{\partial \theta_k}.$

Exercise 2.9. Explain clearly and plainly. **[**$\notin$, 50**]**

1. How do I divide two fractions?
2. I have a positive integer. How do I check if it's prime?
3. I have a positive integer. How do I check if it's a cube?
4. I have two vectors on the plane. How do I check if they are linearly independent?
5. I have a cartesian equation of a circle, and a point. How do I check if the point lies inside the circle?
6. I have two lines in three-dimensional space. How do I check if they intersect?
7. I have two vectors on the plane. How do I check if they are linearly independent?
8. I have a list of quadratic polynomial functions and I must select the functions that assume both positive and negative values. What shall I do?
9. I have two real functions. From a sketch, it seems that their graphs intersect, and do so at a right angle. How do I verify that this is indeed the case?
10. I have three points on the plane. How do I compute the centre of the circle passing through them?
11. I have four points on the plane. How do I check if the points are vertices of a square?

Exercise 2.10. Explain concisely. **[**$\not\epsilon$, 30**]**

1. What is the difference between an ordered pair and a set?
2. What is the difference between an equation and an identity?
3. What is the difference between a function and its graph?

Exercise 2.11. I have two finite sets and a function between them. I am able to compute the value of the function at each point of the domain, and to count and compare the elements of these sets. I need explicit instructions for answering the following questions. **[**$\not\epsilon$, 50**]**

1. How do I check that my function is surjective?
2. How do I check that my function is injective?

Exercise 2.12. Answer the questions as clearly as you can.

1. Let A and B be sets. Why are the sets $A^2 \smallsetminus B^2$ and $(A \smallsetminus B)^2$ not necessarily equal? Under what conditions are they equal?
2. Let $f : X \longrightarrow Y$ be a function, and let A be a subset of X. Why are the sets A and $f^{-1}(f(A))$ not necessarily equal? Under what conditions are they equal?
3. Let A and B be subsets of the domain of a function f. Why are $f(A) \cap f(B)$ and $f(A \cap B)$ not necessarily equal? Under what conditions are they equal?
4. Why are the sets $\emptyset$ and $\{\emptyset\}$ distinct? What are the elements of the set $\mathbf{P}(\mathbf{P}(\emptyset))$? (The symbol $\mathbf{P}(X)$ denotes the power set of a set X.)

Exercise 2.13. Consider the function that performs the prime factorization of a natural number greater than 1. What would you choose for co-domain? Explain, discussing possible representations.

Exercise 2.14.

1. Represent the polynomial set $\mathbb{Z}[x]$ as a set of integer sequences.
2. Represent the real numbers $\mathbb{R}$ as as a set of decimal expansions.

Exercise 2.15. Prove that the definition

$$(a,b) \overset{\text{def}}{=} \{\{a\},\{a,b\}\}$$

satisfies (2.4). (This shows that an ordered pair can be defined in terms of a set, so there's no need to introduce a new object.) Hence define an ordered triple in terms of sets.

Exercise 2.16. Let X,Y be sets, and let $f,g : X \to \mathbf{P}(Y)$. Prove that the set equation $f(x) = g(x)$ is equivalent to the equation

$$f(x)\,\Delta\,g(x) = \emptyset$$

in the sense that they have the same solution set. (Thus every set equation may be reduced to the form $F(x) = \emptyset$, which is a generalisation of 2.27.)

Chapter 3

Mathematical sentences

Consider the following theorem of analysis:

$$\textit{Let } f:\mathbb{R}\to\mathbb{R} \textit{ be a differentiable function. Then } f \textit{ is continuous.} \tag{3.1}$$

This statement comprises two sentences. The first sentence does not state a fact: it's an assumption. We aren't told which function f is, so there is no question of this statement about f being true or false. But still we use it as a basis for the rest of the argument. The second sentence does just that: it's a deduction of a new fact from the assumption. The symbol f is an internal variable of the statement, and we can make it disappear:

Every differentiable real function is continuous.

This statement is equivalent to (3.1).

In this chapter we begin to analyse mathematical sentences, and for this purpose we need some elements of logic. The chief attribute of a mathematical statement is its truth or falsehood. Accordingly, we define the two **logical** (or **boolean**) **constants** TRUE and FALSE, abbreviated T, F or $1, 0$, respectively. The simplest sentences are the **relational expressions**, formed by combining pairs of mathematical objects via **relational operators**. The value of a relational expression is a logical constant. To form more complex sentences we combine relational expressions using **logical operators**, much like combining numbers using addition and multiplication. The formulation of high-level statements such as (3.1) requires logical functions —called **predicates**— and **quantifiers**, which resemble integrals.

These constructs have accompanying symbols, and the symbols of logic are cryptic and alluring. While they help us understand the structure of sentences, they don't necessarily help us write good sentences. Excessive use of logical symbols clutters the exposition, obscuring meaning. A recurrent theme of this chapter is that, quite often, concepts of great logical depth are better expressed with words.

3.1 Relational operators

We begin with a simple symbolic sentence:

$$0 < 1. \tag{3.2}$$

This is a **relational expression**. The **relational operator** '$<$' converts two numbers into the logical constant TRUE. Any significant mathematical sentence that evaluates to TRUE is called a **theorem,** and expression (3.2) is one of the first theorems of analysis; it underpins all inequalities among real numbers.

Relational operators comprise very familiar objects:

$$= \qquad \neq \qquad < \qquad \leqslant \qquad > \qquad \geqslant . \tag{3.3}$$

These operators are **binary**, they act on two operands. The first two act on elements of any set; the others act on real numbers (more generally, on elements of an **ordered set** —see section 3.6).

Interesting things can be said with simple relational expressions:

$$9^3+10^3=1^3+12^3, \qquad\qquad \frac{355}{113}-\pi<3\times10^{-7}. \tag{3.4}$$

These expressions also feature prominently in programming languages:

```
if x<1 then
   x:=-x+1
else
   x:=x-1
fi:
```

Here the logical value of the expression '`x<1`' determines which assignment statement is executed.

At the most basic level we have the relational operators associated with sets, namely the **membership** and **inclusion operators** (section 2.1) and their negation:

$$\in \qquad \notin \qquad \subset \qquad \not\subset \qquad \supset \qquad \not\supset . \tag{3.5}$$

For example, the relational expression

$$\sqrt{2} \notin \mathbb{Q} \tag{3.6}$$

is TRUE, that is, the square root of 2 is not a rational number. We shall prove this fact in section 6.1.

There are countless relational operators in mathematics: the **divisibility operator** '$|$' and the **congruence operator** '$\equiv$' in arithmetic (section 2.1.2), the **isomorphism operator** '$\cong$' in algebra, the **orthogonality operator** '$\perp$' in geometry, etc. By contrast, the symbol '$\approx$', used in mathematical physics, does not represent a relational operator because expressions such as $\pi \approx 3.14$ cannot be assigned unequivocally a value TRUE or FALSE.

3.2 Logical operators

The following sentence is TRUE:

$$\textit{The integer } 2^{57885161}-1 \textit{ is prime.} \tag{3.7}$$

The proof requires a sophisticated theory and a lot of computer time[1]. However, it is not difficult to see that this expression is decomposable into **finitely many** relational expressions. Indeed the primality of $p = 2^{57885161}-1$ could in principle be established by verifying that p is not divisible by any prime up to $\sqrt{p}$. (The locution 'in principle' is cheeky; no computer present or future will ever complete such a brute-force computation within the lifetime of our Universe.)

This example motivates the definition of the **logical** (or **boolean**) **operators** NOT and AND (to be defined below), with which we rewrite (3.7) as follows:

$$\text{NOT}(2|p) \quad \text{AND} \quad \text{NOT}(3|p) \quad \text{AND} \quad \text{NOT}(5|p) \quad \text{AND} \quad \cdots.$$

We have produced a complex symbolic sentence by combining relational expressions with logical operators. This process is analogous to the construction of complex arithmetical expressions from arithmetical constants (numbers) and operators ($+$, $-$, etc.).

The operator NOT (represented by the symbol $\neg$) is **unary** —it takes one boolean operand and produces a boolean value. In this respect NOT resembles the unary arithmetical operator '$-$', which changes the sign of a number. The operator AND (represented by the symbol $\wedge$) is **binary**: think of it as a kind of multiplication.

---

[1] This expression refers to the largest known prime (as of February 2013) with 17425170 digits: see [5].

The following **truth tables** define the operators NOT and AND by specifying their action on all possible choices of operands:

P	$\neg P$
T	F
F	T

P	Q	$P \wedge Q$
T	T	T
T	F	F
F	T	F
F	F	F

(3.8)

The expression $\neg P$ is called the **negation** of P. Negating relational expressions is straightforward, as we already have all the relevant symbols —see (3.3) and (3.5). Thus

$$\neg(x < y) = (x \geqslant y), \qquad \neg(x \in A) = (x \notin A), \qquad \neg(A \not\subset B) = A \subset B.$$

The expression $P \wedge Q$ is called a **conjunction** or **compound expression**. In a compound expression the operator AND may appear implicitly. For instance, the expression $0 < x < 1$ is compound: $(0 < x) \wedge (x < 1)$.

Next we define the operator OR (represented by the symbol $\vee$) and the **implication operator** (represented by the symbol $\Rightarrow$), as follows:

P	Q	$P \vee Q$
T	T	T
T	F	T
F	T	T
F	F	F

P	Q	$P \Rightarrow Q$
T	T	T
T	F	F
F	T	T
F	F	T

(3.9)

These operators could also be expressed in terms of $\neg$ and $\wedge$, see exercise 3.1. The operator OR is **inclusive**, namely $\mathrm{T} \vee \mathrm{T} = \mathrm{T}$, which is not the meaning usually attributed to the conjunction 'or' in English usage[2].

The expression

$$P \Rightarrow Q \tag{3.10}$$

is called an **implication,** and it reads

P implies Q *Q follows from P* *if P, then Q*
P only if Q *P is sufficient for Q.*

---

[2] There is also an **exclusive** version of OR, called XOR ($\veebar$), for which $\mathrm{T} \veebar \mathrm{T} = \mathrm{F}$.

The expression P is the **hypothesis,** and Q is the **conclusion.** Many mathematical statements have this form, theorem (3.1), for example.

According to the truth table (3.9), if P is false, then the expression $P \Rightarrow Q$ is true for any Q. This convention, which is perhaps unexpected, is in fact in agreement with the common usage of an implication:

> *If 4 is a prime number, then I am a Martian.*
> *If I win the lottery, then I'll buy you a Ferrari.*

The meaning of these sentences is clear. In particular, if I don't win the lottery, then I am free do what I want about the Ferrari.

From (3.9) we see that the expressions $P \Rightarrow Q$ and $Q \Rightarrow P$ are different. Accordingly, we introduce the operator $\Leftarrow$ as follows:

$$P \Leftarrow Q \overset{\text{def}}{=} Q \Rightarrow P.$$

The expression on the left is called the **converse** of the implication (3.10). It reads

P is implied by Q	*P follows from Q*
P if Q	*P is necessary for Q.*

The value of an implication and its converse are unrelated. In the following example the former is false and the latter is true:

$$(x^2 = 25) \Rightarrow (x = -5) \qquad\qquad (x^2 = 25) \Leftarrow (x = -5). \tag{3.11}$$

Inappropriate reversal of an implication is a common mistake in proofs —see section 6.7.2.

Formulae (3.11) combine relational and boolean operators, and the parentheses suggest the order in which these operators are evaluated. In this case the parentheses are redundant, since relational operators must necessarily take precedence over boolean operators. Thus the expression $x^2 = 25 \Rightarrow x = -5$ could not be interpreted as $x^2 = (25 \Rightarrow x) = -5$. However, when two or more boolean operators are present, parentheses may be inserted to change the meaning of an expression. For instance, the expressions

$$(\text{F} \wedge \text{T}) \Rightarrow \text{T} \qquad\qquad \text{F} \wedge (\text{T} \Rightarrow \text{T})$$

evaluate to TRUE and FALSE respectively. As with arithmetical operators, there is an agreed order with which boolean operators are evaluated: first $\neg$, then $\Rightarrow$, and then $\wedge$ and $\vee$. The last two operators have the same precedence and are evaluated left to right if both appear. Thus the expression $P \vee \neg Q \Rightarrow R \wedge S$ is evaluated as $(P \vee (\neg Q)) \Rightarrow (R \wedge S)$; we note that the redundant parentheses add clarity to the sentence.

The **equivalence operator** $\Leftrightarrow$ is the conjunction of the direct and converse implications:

$$P \Leftrightarrow Q \overset{\text{def}}{=} (P \Rightarrow Q) \wedge (P \Leftarrow Q) \tag{3.12}$$

and it corresponds to the following truth table:

P	Q	$P \Leftrightarrow Q$
T	T	T
T	F	F
F	T	F
F	F	T

The expression $P \Leftrightarrow Q$ is read aloud as

P implies and is implied by Q	*P is equivalent to Q*
P if and only if Q	*P is necessary and sufficient for Q*

The awkward expression *'if and only if'* is quite common; the abbreviated form 'iff' is found in formulae as an alternative to '$\Leftrightarrow$'. So the expressions

$$A \subset B \Leftrightarrow A \setminus B = \emptyset \qquad\qquad A \subset B \text{ iff } A \setminus B = \emptyset$$

have the same meaning; they are read aloud as

A is a subset of B if and only if the set difference of A and B is empty.

If —as in this example— the value of $P \Leftrightarrow Q$ is TRUE, then the statements P and Q are **logically equivalent**, meaning that one is is just a rewording of the other.

The **contrapositive** of the implication (3.10) is the implication

$$\neg P \Leftarrow \neg Q. \tag{3.13}$$

The contrapositive is constructed by reversing the operator and negating the operands. Great care must be exercised in distinguishing between direct, converse, and contrapositive implications.

DIRECT :	*If x is a multiple of 4, then x is even.*	(TRUE)
CONVERSE :	*If x is even, then x is a multiple of 4.*	(FALSE)
CONTRAPOSITIVE :	*If x is odd, then x is not a multiple of 4.*	(TRUE)

While the value of an implication and of its converse are unrelated, every implication is equivalent to its contrapositive. This is identity (*iv*) of the following theorem.

Theorem. *For all* $P, Q \in \{\mathrm{T}, \mathrm{F}\}$*, the following holds:*

$$\begin{array}{rccc} (i) & \neg(P \vee Q) & \Leftrightarrow & (\neg P \wedge \neg Q) \\ (ii) & \neg(P \wedge Q) & \Leftrightarrow & (\neg P \vee \neg Q) \\ (iii) & (P \Rightarrow Q) & \Leftrightarrow & (\neg P \vee Q) \\ (iv) & (P \Rightarrow Q) & \Leftrightarrow & (\neg P \Leftarrow \neg Q). \end{array}$$

PROOF. We prove (*iii*), leaving the other cases as an exercise. We will show that the two sides of the equivalence (*iii*) have the same truth table. The truth table of the left-hand side is given in (3.9); that of the right-hand is computed as follows:

P	Q	$\neg P$	$\neg P \vee Q$	$P \Rightarrow Q$
T	T	F	T	T
T	F	F	F	F
F	T	T	T	T
F	F	T	T	T

We see that the two truth tables are equal. □

The statements (*i*) and (*ii*) are known as *De Morgan's laws*[3]. Using theorem 3.2, one can express the operators $\vee$, $\Rightarrow$, $\Leftrightarrow$ in terms of $\vee$ and $\wedge$ (exercise 3.1).

All items in the theorem have the form $A(P,Q) \Leftrightarrow B(P,Q)$, where the statements A and B have the same value for any choice of P and Q and hence are equivalent. A statement of this kind is called a **tautology**. In logic a distinction is made between a tautology and a syntactically correct —but not necessarily true— double implication, by introducing the symbol $\leftrightarrow$ for the latter circumstance. So one would write $(P \vee Q) \leftrightarrow (P \wedge Q)$ rather than $(P \vee Q) \Leftrightarrow (P \wedge Q)$. A similar distinction exists between $\Rightarrow$ and $\rightarrow$. For our purpose these additional symbols are unnecessary.

3.3 Predicates

The sentence

The integer x *is divisible by 7*

---

[3] Augustus De Morgan (British: 1806–1871).

contains an unspecified symbol, and thus it assumes a logical value only if the symbol is assigned an integer value. We are prompted to introduce the following function:

$$\mathscr{P} : \mathbb{Z} \to \{\mathrm{T}, \mathrm{F}\} \qquad x \mapsto 7 \mid x. \tag{3.14}$$

We see that $\mathscr{P}(x)$ is the symbolic translation of our sentence, and that $\mathscr{P}(22) = \mathrm{F}$, $\mathscr{P}(-91) = \mathrm{T}$.

Using a function to encode a sentence is a simple idea with far-reaching consequences. Let us thus define a **predicate** (or **boolean function**) to be any function that assumes boolean values. If the domain of a predicate $\mathscr{P}$ is X, then we speak of a predicate **on** (or **over**) X. So (3.14) is a predicate on the integers.

Let $\mathscr{P} : X \to \{\mathrm{T}, \mathrm{F}\}$ be a predicate on X. There is a distinguished subset A of X, determined by $\mathscr{P}$ via the following Zermelo definition:

$$A := \{x \in X \,:\, \mathscr{P}(x)\}. \tag{3.15}$$

With reference to equation (2.7) and the discussion that follows, we see that the expression $\mathscr{P}(x)$, which means ‘x has property P’, is the definiens of the set A. So the Zermelo definition (3.15) may be expressed in the language of functions using an inverse image

$$\{x \in X \,:\, \mathscr{P}(x)\} \overset{\text{def}}{=} \mathscr{P}^{-1}(\{\mathrm{T}\})$$

(cf. definition (2.17), p. 27).

Conversely, let X be a set and let A be a subset of X. The predicate

$$\mathscr{P}_A : X \to \{\mathrm{T}, \mathrm{F}\} \qquad x \mapsto x \in A \tag{3.16}$$

is called the **characteristic function** of A (in X). Explicitly, $\mathscr{P}(x)$ is equal to T if x belongs to A and to F otherwise. For example, the function (3.14) is the characteristic function of the set $7\mathbb{Z}$ of integer multiples of 7.

So to every predicate on a set X we associate a unique subset of X, and vice-versa. We say that there is a **bi-unique correspondence** between these two classes of objects, established by expressions (3.16) and (3.15).

Sets are manipulated using set operators. Predicates are manipulated using boolean operators in a natural way by defining, say,

$$(\mathscr{P} \wedge \mathscr{Q})(x) \overset{\text{def}}{=} \mathscr{P}(x) \wedge \mathscr{Q}(x)$$

and similarly for all other operators. The following theorem establishes correspondences between these two classes of objects.

Theorem. *Let* X *be a set, let* $A, B \subseteq X$*, and let* $\mathscr{P}_A$*,* $\mathscr{P}_B$ *be the corresponding characteristic functions. The following holds (the prime denotes taking the complement)*:

$$\begin{array}{lrcl} (i) & \neg\,\mathscr{P}_A & = & \mathscr{P}_{A'} \\ (ii) & \mathscr{P}_A \wedge \mathscr{P}_B & = & \mathscr{P}_{A\cap B} \\ (iii) & \mathscr{P}_A \vee \mathscr{P}_B & = & \mathscr{P}_{A\cup B} \\ (iv) & \mathscr{P}_A \Rightarrow \mathscr{P}_B & = & \mathscr{P}_{(A\smallsetminus B)'} \\ (v) & \mathscr{P}_A \Leftrightarrow \mathscr{P}_B & = & \mathscr{P}_{(A\cap B)\,\cup\,(A\cup B)'}. \end{array}$$

(This baroque display of symbols is unappealing. In section 5.7 we'll make it more digestible by writing a short essay about it without symbols.)

Proof. To prove (i) we note that the function $x \mapsto \neg\mathscr{P}_A(x)$ evaluates to TRUE if $x \notin A$ and to FALSE otherwise. However, from the definition of the complement of a set, we have $x \notin A \Leftrightarrow x \in A'$.

Next we prove (iv). We'll prove instead

$$\neg(\mathscr{P}_A \Rightarrow \mathscr{P}_B) = \mathscr{P}_{(A\smallsetminus B)}$$

which, together with (i), gives us (iv). Let $\mathscr{P}_A := (x \in A)$ and let $\mathscr{P}_B := (x \in B)$. From the truth table (3.9) of the operator $\Rightarrow$, we find that $\neg(\mathscr{P}_A \Rightarrow \mathscr{P}_B)(x)$ is TRUE precisely when $\mathscr{P}_A(x)$ is TRUE and $\mathscr{P}_B(x)$ is FALSE. This means that

$$(x \in A) \wedge (x \notin B),$$

but this is just the definition of the characteristic function of the set $A \smallsetminus B$, as desired.

The proof of (ii), (iii), (v) is left as an exercise. □

We illustrate the significance of this theorem with examples.

Example. Let a and b be real numbers. The predicates $x \mapsto (x \geqslant a)$ and $x \mapsto (x < b)$ (over $\mathbb{R}$) are the characteristic functions of two rays, the latter without end-point. According to theorem 3.3 part (ii), the predicate $x \mapsto ((x \geqslant a) \wedge (x < b))$ is the characteristic function of the intersection of these rays. Depending on whether $a < b$, or $a \geqslant b$, this intersection is the half-open interval $[a, b)$ or the empty set.

Example. Let X, Y be sets, and let $f, g : X \to Y$ be functions. An **equation** (on X) is a predicate of the type

$$\mathscr{P} : X \to \{\mathrm{T}, \mathrm{F}\} \qquad x \mapsto (f(x) = g(x)).$$

In the language of predicates, an equation is the characteristic function of its own solution set $\{x \in X \,:\, \mathscr{P}(x)\}$ (section 2.5). The system of two equations

$$f_1(x) = g_1(x) \qquad f_2(x) = g_2(x)$$

defined over the same set corresponds to the predicate $x \mapsto \mathscr{P}_1(x) \wedge \mathscr{P}_2(x)$. From theorem 3.3, part (*ii*), it follows that the solution set of a system of two equations is the intersection of the solution sets of the individual equations.

Example. If $A \subset B$, then $A \smallsetminus B$ is empty, and from part (*iv*) of the theorem we obtain

$$(\mathscr{P}_A \Rightarrow \mathscr{P}_B) \,=\, \mathscr{P}_{\emptyset'} = \mathscr{P}_X.$$

The subset relation has been translated into an identity of functions. For example, the functional identity

$$(\mathscr{P}_{4\mathbb{Z}} \Rightarrow \mathscr{P}_{2\mathbb{Z}}) \,=\, \mathscr{P}_{\mathbb{Z}}$$

says that every multiple of four is even.

3.4 Quantifiers

Interesting mathematical statements —such as (3.1)— refer to families of objects rather than to individual objects. The formulation of general statements requires two special symbols, the **universal quantifier** $\forall$ and the **existential quantifier** $\exists$. These symbols translate into words in several equivalent ways:

$\forall$:	*for all*	*given any*	*for any choice of*
$\exists$:	*for some*	*there exists*	*we can find*

Expressions with quantifiers have the following syntax:

$$\begin{array}{ll} \forall x \in X, \mathscr{P}(x) & \textit{For all } x \textit{ in } X,\ \mathscr{P}(x). \\ \exists x \in X, \mathscr{P}(x) & \textit{There exists } x \textit{ in } X\textit{, such that } \mathscr{P}(x). \end{array} \tag{3.17}$$

where X is a set, and $\mathscr{P}$ is a predicate over X. The quantifier is followed by the symbol being quantified, the membership operator, a set, and a predicate. The acronym ‘s.t.’ (for ‘such that’) is sometimes inserted in the second symbolic expression (3.17) before the predicate:

$$\exists x \in X \quad \text{s.t. } \mathscr{P}(x).$$

Without predicate the expressions $\forall x \in X$ is incomplete and meaningless. The expression $\exists x \in X$ is meaningful but not used; one instead writes $X \neq \emptyset$.

The meaning of the expressions (3.17) should be clear; in particular, each expression has a logical value. For example, the following sentences are TRUE:

$\forall x \in \mathbb{R},\ x^2 \geqslant 0$ *The square of every real number is non-negative.*

$\exists x \in \mathbb{Q},\ 0 < x < 1$ *There is a rational number in the open unit interval.*

In translating the symbolic expressions we haven't merely converted symbols into words following (3.17). We have also synthesised meaning, and in doing so the quantified variable x has become silent —no explicit reference to it remains in the sentence. This is an intrinsic phenomenon, as we shall see.

If the ambient set is clear from the context, then reference to it may be omitted. This is invariably the case if the quantifier appears in conjunction with an inequality that restricts the variable's range. Of the following equivalent expressions

$$\forall n > 3,\ n! > 2^n \qquad \forall n \in \mathbb{N} \setminus \{1,2,3\},\ n! > 2^n \qquad \forall n \in \mathbb{N},\ n > 3 \Rightarrow n! > 2^n$$

the first one is preferable. The presence of the factorial makes it clear that we are dealing with natural numbers.

EXAMPLE. Using quantifiers, it is possible to define relational operators in terms of other operators, thereby reducing the number of primitive operators. The inclusion operator can be defined in terms of the membership operator:

$$A \subset B \overset{\text{def}}{\Leftrightarrow} \forall x \in A,\ x \in B. \tag{3.18}$$

The divisibility operator is defined in terms of multiplication:

$$m|n \overset{\text{def}}{\Leftrightarrow} \exists k \in \mathbb{Z},\ mk = n. \tag{3.19}$$

EXAMPLE. The symbolic sentences (3.17) may be expressed in the language of functions:

$$\begin{aligned} \forall x \in X, \mathscr{P}(x) &\Leftrightarrow \mathscr{P}(X)=\{\text{T}\} \\ \exists x \in X, \mathscr{P}(x) &\Leftrightarrow \mathscr{P}(X)\neq\{\text{F}\}. \end{aligned} \tag{3.20}$$

These equivalences may give the impression that we have managed to bypass quantifiers altogether. This is not the case. The standard definition of image of a set under a function (2.16) has a hidden quantifier in it, which is revealed when the definition is converted to Zermelo form. Thus if $f : A \to B$ is a function, then

$$f(A) = \{f(x) : x \in A\} = \{y \in B : \exists x \in A,\ y = f(x)\}.$$

Of the two definitions, the first one is more intuitive.

Several quantifiers may appear within the same sentence. For illustration, let us consider the following two-variable predicate:

$$\mathscr{L}(x,y) := (\text{‘}x \text{ loves } y\text{’}).$$

This function is defined over the cartesian product $G \times G$, where G is a set of people. So the lovers and the loved ones belong to the same set.

Now choose $g \in G$, say, $g =$ George. Then ‘everybody loves George’ is a boolean expression, which is spelled out as

For all elements x of G, the value of $\mathscr{L}(x,g)$ is TRUE.

This expression translates into symbols as follows:

$$\forall x \in G, \;\; \mathscr{L}(x,g).$$

This sentence may be true or false (depending on how charming George is). The following sentences with one or two quantifiers exemplify several possible constructs.

	words	symbols
1	*Everybody loves George*	$\forall x \in G, \;\; \mathscr{L}(x,g)$
2	*Somebody loves George*	$\exists x \in G, \;\; \mathscr{L}(x,g)$
3	*Everybody loves himself*	$\forall x \in G, \;\; \mathscr{L}(x,x)$
4	*George is in love*	$\exists x \in G \setminus \{g\}, \;\; \mathscr{L}(g,x)$
5	*George is selfish*	$\mathscr{L}(g,g) \wedge (\forall x \in G \setminus \{g\}, \;\; \neg \mathscr{L}(g,x))$
6	*Everybody is in love*	$\forall x \in G, \;\; \exists y \in G, \;\; \mathscr{L}(x,y)$
7	*Somebody is in love*	$\exists x \in G, \;\; \exists y \in G, \;\; \mathscr{L}(x,y)$
8	*Somebody loves everybody*	$\exists x \in G, \;\; \forall y \in G, \;\; \mathscr{L}(x,y)$
9	*Somebody is loved by everybody*	$\exists y \in G, \;\; \forall x \in G, \;\; \mathscr{L}(x,y)$

In the sentences 4 and 5, the indeterminate x belongs to the set $G \setminus \{g\}$, to ensure that x is distinct from g. It is possible to transfer this constraint to the predicate, e.g., in item 4:

$$\exists x \in G, \;\; (x \neq g) \wedge \mathscr{L}(g,x).$$

Two-quantifier sentences have implied parentheses, which establish the order of evaluation of the sub-expressions. Thus expression 6 is written in full as

$$\forall x \in G,\ \ (\exists y \in G,\ \ \mathscr{L}(x,y)).$$

According to (3.17), the quantity in parentheses must be equal to $\mathscr{Q}(x)$ for some predicate $\mathscr{Q}$ over G; indeed we see that $\mathscr{Q}(x) =$'x loves somebody'. From the sentence's symbolic structure and meaning, it is clear that the choice of y depends on x, in general. Therefore, exchanging the order of the quantifiers alters the meaning of the expression, as confirmed by comparing sentences 6 and 9.

Quantifiers are **operators** which act on boolean functions producing new functions. Their effect is to reduce the number of variables by one, a process that calls to mind definite integration. We should think of the two expressions

$$\forall x \in X \qquad\qquad \int_X dx$$

as being structurally similar. The operator on the left acts on predicates over the set X, namely on functions $\mathscr{P} : X \longrightarrow \{\mathrm{T},\mathrm{F}\}$. The operator on the right acts, say, on functions $F : X \longrightarrow \mathbb{R}$, which are integrable over a subset X of the real line. (To fix ideas, assume that X is an interval.) Inserting the appropriate functions in each expression

$$\forall x \in X,\ \ \mathscr{P}(x) \qquad\qquad \int_X dx F(x)$$

we obtain a boolean constant (TRUE or FALSE) on the left, and a numerical constant (a number) on the right. Now suppose that both $\mathscr{P}$ and F are functions of two variables x and y, defined over the cartesian product $X \times Y$ of two sets. Then each expression

$$\forall x \in X,\ \ \mathscr{P}(x,y) \qquad\qquad \int_X dx F(x,y)$$

produces a function of y, the variable that is not quantified in one case, and not integrated over in the other. If we quantify/integrate with respect to both variables,

$$\forall x \in X,\ \exists y \in Y,\ \ \mathscr{P}(x,y) \qquad\qquad \int_X dx \int_Y dy F(x,y),$$

once again we obtain constants.

Returning to the examples above, we note that $\mathscr{L}(x,g)$ is a predicate in one variable (g is fixed), and so is $\mathscr{L}(x,x)$. Hence $\exists x, \mathscr{L}(x,g)$ and $\forall x, \mathscr{L}(x,x)$ are constants. On the other hand, any expression in which the number of quantifiers is smaller than the number of arguments in the predicate is a predicate with fewer arguments, as the following examples illustrate.

words	symbols
x is in love	$\exists y \in G \setminus \{x\},\ \mathscr{L}(x,y)$
x is loved	$\exists y \in G \setminus \{x\},\ \mathscr{L}(y,x)$
x is a hippy	$\forall y \in G,\ \mathscr{L}(x,y)$
x is selfish	$\mathscr{L}(x,x) \wedge (\forall y \in G \setminus \{x\},\ \neg\mathscr{L}(x,y))$
x is a lover of George	$x \neq g \wedge \mathscr{L}(x,g) \wedge \mathscr{L}(g,x)$

Now, each predicate is the characteristic function of a subset of G, which we construct with a Zermelo definition.

words	symbols
the people in love	$\{x \in G : \exists y \in G \setminus \{x\},\ \mathscr{L}(x,y)\}$
the loved ones	$\{y \in G : \exists x \in G \setminus \{y\},\ \mathscr{L}(x,y)\}$
the hippies	$\{x \in G : \forall y \in G,\ \mathscr{L}(x,y)\}$
the selfish people	$\{x \in G : \mathscr{L}(x,x) \wedge (\forall y \in G \setminus \{x\},\ \neg\mathscr{L}(x,y))\}$
George's lovers	$\{x \in G : x \neq g \wedge \mathscr{L}(x,g) \wedge \mathscr{L}(g,x)\}.$

The sentence '*everybody loves somebody*' is of the form

$$\forall x \in X,\ \exists y \in Y,\ \mathscr{P}(x,y) \tag{3.21}$$

where X and Y are sets and $\mathscr{P}$ is a predicate over $X \times Y$. Letting $X = Y = \mathbb{N}$ and $\mathscr{P}(x,y) = (x < y)$, we obtain a new statement which has the same formal structure:

> *Given any integer, one can find a larger integer.*

or better,

> *There is no greatest integer.*

It's useful to think of the interplay between universal and existential quantifiers as a representation of an adversarial system —like a court of law— based on the following rules of engagement:

$\forall$	*given any*	(my opponent's move)
$\exists$	*I can find*	(my move).

The quantifiers create a tension between the two contenders, and I should expect to be challenged by my opponent. Accordingly, we rewrite our statement more emphatically:

> *Given any integer, no matter how large, one can always find a larger integer.*

The expressions 'no matter how large' and 'always' are inessential to the claim, but they expose the dynamics of the process. This aspect of predicate calculus will be considered again in chapter 6, when we deal with proofs.

EXAMPLE. The **Archimedean property** of the real numbers states that

> *The integer multiples of any positive quantity can be made arbitrarily large.*

The symbolic version of this statement requires three quantifiers.

$$\forall x \in \mathbb{R}^+,\ \ \forall y \in \mathbb{R},\ \ \exists n \in \mathbb{N},\ \ nx > y.$$

In this expression x is the positive quantity in question, y is the (large) quantity we want to exceed, and n is the multiple of x needed to achieve this. Note that $n = n(x,y)$. Clearly, the Archimedean property is best expressed with words than with symbols, and the large number of quantifiers hidden within this sentence may be taken as an indication of the concept's logical depth.

3.4.1 Quantifiers and functions

Invariably, statement about functions will refer to the elements of the domain or co-domain, and quantifiers may be made to act on these elements. For instance, a function is **injective** if it maps distinct points to distinct points (section 2.2). This concise statement unfolds as follows:

> *Given any two points in the domain of the function, if they are distinct, then so are their images under the function.*

This sentence has a universal quantifier ('given any') and an implication operator ('if … then'). Let $f : A \to B$. A literal translation into symbols is

$$\forall x,y \in A,\ \ (x \neq y) \Rightarrow (f(x) \neq f(y))$$

where $\forall x, y \in A$ is a shorthand for $\forall x \in A,\ \forall y \in A$. Replacing the implication by its contrapositive (theorem 3.2 (*iv*), p. 57) we obtain a neater expression:

$$\forall x, y \in A,\ \ (f(x) = f(y)) \Rightarrow (x = y).$$

The original definition ('distinct points have distinct images') restricts the ambient set to pairs of distinct points. This can be done symbolically:

$$\forall x \in A,\ \ \forall y \in A \smallsetminus \{x\},\ \ f(x) \neq f(y).$$

Now the predicate is simpler but the set is more complicated. It is always possible to transform an implication by absorbing the hypothesis into the ambient set (exercise 3.1.3).

A function f is **surjective** if image and co-domain coincide: $f(A) = B$. This high-level expression hides two quantifiers; we first spell it out:

> *Given any point in the co-domain, we can find a point in the domain which maps to it.*

and then we rewrite it symbolically:

$$\forall y \in B,\ \ \exists x \in A,\ \ f(x) = y.$$

This expression is of the form (3.21).

In section 2.4 we noted that a sequence of elements of a set A may be interpreted as a function

$$a : \mathbb{N} \to A \qquad k \mapsto a_k.$$

Characterising sequences is then analogous to characterising functions.

For example, consider the set $\mathbb{Z}^{\mathbb{N}}$ of all integer sequences (this notation was developed in section 2.4). To isolate sequences with certain properties —a subset of $\mathbb{Z}^{\mathbb{N}}$— we use a Zermelo definition. For example, the set

$$\{a \in \mathbb{Z}^{\mathbb{N}} : a_1 \in 2\mathbb{Z}\}$$

consists of all integer sequences whose first term is even. By applying the quantifier $\forall$ to the subscript k (which is the variable in our functions), we can deal with all terms of the sequences at once. So the set of integer sequences with only even terms is given by

$$\{a \in \mathbb{Z}^{\mathbb{N}} : \forall k \in \mathbb{N},\ a_k \in 2\mathbb{Z}\} = (2\mathbb{Z})^{\mathbb{N}}. \tag{3.22}$$

There seems to be something wrong here. The quantifier 'integrates out' the indeterminate k: does this mean that the predicate is just a constant? No, because the Zermelo variable is actually $a = (a_k)$, an element of the ambient set $\mathbb{Z}^{\mathbb{N}}$. If we write (3.22) as $\{a \in \mathbb{Z}^{\mathbb{N}} : \mathscr{P}(a)\}$, then we see the structure of the predicate:

$$\mathscr{P} : \mathbb{Z}^{\mathbb{N}} \to \{\mathrm{T}, \mathrm{F}\} \qquad a \mapsto (\forall k \in \mathbb{N},\ a_k \in 2\mathbb{Z}).$$

Likewise, the set

$$\{a \in \mathbb{Z}^{\mathbb{N}} : \exists k \in \mathbb{N},\ a_k \in 2\mathbb{Z}\}$$

is

> *The set of integer sequences with an even term.*

The above expression reflects the very literal interpretation of what's written, which is what mathematicians —and lawyers— are notorious for. A mathematical geek would be entertained by the fact that the statement '*In London there is an underground station*', is true; a normal person would instead perceive this as a puzzling understatement ('Where do they go from there?'). So a characterisation of the type

> *The set of integer sequences with at least one even term.*

is preferable to the one given before. The qualifier 'at least' is superfluous, but helps the reader note an essential point.

The set

$$\{a \in \mathbb{Z}^{\mathbb{N}} : \forall n \in \mathbb{N},\ n > 2 \Rightarrow 2|a_n\}$$

is described as

> *The set of integer sequences whose terms, after the second, are even.*

No information is given about the first two terms, yet this statement could be misinterpreted as meaning that these terms are odd. A more eloquent description of this set is

> *The set of integer sequences whose terms are all even, with the possible exception of the first two terms.*

These examples show how much can be done to improve the clarity of a definition; this will be our concern in chapter 5.

3.4.2 Existence statements

An **existence statement** is a logical expression with a leading existential quantifier $\exists$. In this section we consider some examples; we shall take a deeper look at the question of existence in chapter 8.

The standard symbolic form of an existence statement is (3.17), but in a mathematical sentence the quantifier is often hidden:

The number $\cos(\pi/3)$ *is rational.*

To make the quantifier visible, we write

For some rational number r*, we have* $\cos(\pi/3) = r$.

or, in symbols,

$$\exists r \in \mathbb{Q},\ \ r = \cos(\pi/3).$$

This statement is weaker than the identity $\cos(\pi/3) = 1/2$, because it does not require us to reveal which rational number our expression is equal to.

Equation (3.19) shows that an existential quantifier is hidden in the definition of the divisibility of integers. Therefore the sentence 'n is even', or, more generally, 'm divides n' are existence statements.

The sentence '*n is not prime*' is an existence statement, because it means '*there is a proper divisor of n*'. We write it symbolically as

$$\begin{aligned}&\exists m \in \mathbb{N},\ \ (m|n) \wedge (m \neq 1) \wedge (m \neq n)\\ &\qquad \Leftrightarrow\ \ \exists m \in \mathbb{N},\ \ (\exists k \in \mathbb{N},\ \ mk = n) \wedge (m \neq 1) \wedge (m \neq n)\end{aligned}$$

where we have been careful in excluding trivial divisors.

The following well-known theorem in arithmetic is an existence statement.

Theorem (Lagrange[4] 1770). *Every natural number is the sum of four squares.*

Let us we analyse it in some detail. First, we consider two instances of the theorem

$$5 = 2^2 + 1^2 + 0^2 + 0^2 \qquad\qquad 7 = 2^2 + 1^2 + 1^2 + 1^2.$$

The first identity shows that some integers are the sum of fewer than four squares, while one checks that 7 requires all four squares. We restate the second identity without disclosing the details:

$$\exists\, a, b, c, d \in \mathbb{Z},\ \ 7 = a^2 + b^2 + c^2 + d^2.$$

---

[4] Joseph-Louis Lagrange, born Giuseppe Lodovico Lagrangia (Italian: 1736–1813).

By replacing 5 or 7 with an unspecified natural number n, we obtain a predicate $\mathscr{L}$ over $\mathbb{N}$:

$$\mathscr{L}(n) := \big(\exists\, a,b,c,d \in \mathbb{Z},\ \ n = a^2+b^2+c^2+d^2\big) \tag{3.23}$$

or

> *n is a sum of four squares.*

To state Lagrange's theorem with symbols, we quantify the remaining variable n

$$\forall n \in \mathbb{N},\ \ \exists a,b,c,d \in \mathbb{Z},\ \ n = a^2+b^2+c^2+d^2. \tag{3.24}$$

Five quantifiers are necessary to turn the predicate $n = a^2+b^2+c^2+d^2$ (a function of five variables) into a boolean constant. In the original formulation of this theorem, all five variables have fallen silent.

Having buried all meaning inside the predicate (3.24), Lagrange's theorem now disappears into a set identity:

$$\mathbb{N} = \{n \in \mathbb{N} : \mathscr{L}(n)\}.$$

Any theorem on natural numbers can be put in this form for an appropriate predicate $\mathscr{L}$. (Think about it.)

Negating expressions with a leading universal quantifier always results in existence statements, as we shall see in the next section.

3.5 Negating logical expressions

If $\mathscr{L}$ is a logical expression, then its negation is $\neg\mathscr{L}$, which is false if $\mathscr{L}$ is true and vice-versa. Relational expressions are negated by crossing out the operator with a forward slash —see (3.3) and (3.5), section 3.1. The only exceptions are the inequalities, whose negations have dedicated symbols.

Theorem 3.2 provides the negation formulae for the main compound expressions

$$\begin{aligned} \neg(P\wedge Q) &\ \Leftrightarrow\ \neg P \vee \neg Q \\ \neg(P\vee Q) &\ \Leftrightarrow\ \neg P \wedge \neg Q \\ \neg(P\Rightarrow Q) &\ \Leftrightarrow\ P \wedge (\neg Q). \end{aligned} \tag{3.25}$$

The first two formulae are items (*i*) and (*ii*) of the theorem (de Morgan's laws); the third follows immediately from (*iii*) and (*i*). Note that the negation of an implication does not have the form of an implication.

If $\mathscr{L}$ is an expression with quantifiers and boolean operators, then the expression $\neg\mathscr{L}$ unfolds into an equivalent expression which we seek to determine. The following sentences contain a negation and a universal quantifier:

> *Not all primes are odd.*
> *A square integral matrix is not necessarily invertible.*
> *Not all continuous real functions are differentiable.*

We recognise that these are disguised existence statements:

> *There is an even prime.*
> *There is a square integral matrix which is not invertible.*
> *There is a continuous real function which is not differentiable.*

The following theorem confirms this observation.

Theorem. *Let $\mathscr{P}$ be a predicate on a set A, and let*

$$\mathscr{L} := \forall x \in A,\, \mathscr{P}(x) \qquad \mathscr{M} := \exists x \in A,\, \mathscr{P}(x).$$

Then

$$\neg\mathscr{L} \Leftrightarrow \exists x \in A, \neg\mathscr{P}(x) \qquad \neg\mathscr{M} \Leftrightarrow \forall x \in A, \neg\mathscr{P}(x). \tag{3.26}$$

Proof. From (3.20), we have the boolean equivalence

$$\mathscr{L} \Leftrightarrow \mathscr{P}(A) = \{\mathrm{T}\}$$

where $\mathscr{P}(A)$ is the image of A under $\mathscr{P}$ (see section 2.2). But then

$$\neg\mathscr{L} \Leftrightarrow \mathscr{P}(A) \neq \{\mathrm{T}\} \Leftrightarrow \neg\mathscr{P}(A) \neq \{\mathrm{F}\}.$$

Using again (3.20), we obtain

$$\neg\mathscr{L} \Leftrightarrow \exists x \in A,\ \neg\mathscr{P}(x)$$

as claimed.

The second formula is proved similarly. From $\mathscr{M} \Leftrightarrow \mathscr{P}(A) \neq \{F\}$, we obtain

$$\neg\mathscr{M} \Leftrightarrow \mathscr{P}(A) = \{\mathrm{F}\} \Leftrightarrow \neg\mathscr{P}(A) = \{\mathrm{T}\}$$

and hence

$$\neg\mathscr{M} \Leftrightarrow \forall x \in A,\ \neg\mathscr{P}(x).$$

□

It's now easy to deduce the rule for negating expressions with two or more quantifiers. Let $\mathscr{P}$ be a predicate on $A \times B$, and let us consider the expression

$$\mathscr{L} := \forall x \in A,\ \exists y \in B,\ \mathscr{P}(x,y).$$

Repeated applications of theorem 3.5 give

$$\begin{aligned} \neg\mathscr{L} &\Leftrightarrow \neg\big(\forall x \in A,\ (\exists y \in B,\ \mathscr{P}(x,y))\big) \\ &\Leftrightarrow \exists x \in A,\ \neg(\exists y \in B,\ \mathscr{P}(x,y)) \\ &\Leftrightarrow \exists x \in A,\ \forall y \in B,\ \neg\mathscr{P}(x,y). \end{aligned}$$

It should be clear how similar formulae may be derived, having any combination of two quantifiers.

Finally, we consider the case of an arbitrary sequence of quantifiers, e.g.,

$$\mathscr{L} = \forall x_1 \in X_1,\ \exists x_2 \in X_2,\ \dots\ , \forall x_n \in X_n,\ \mathscr{P}(x_1,\dots,x_n)$$

where the X_i are sets, and $\mathscr{P}$ is a predicate on the cartesian product $X_1 \times X_2 \times \cdots \times X_n$. Parentheses make it clear that this is a nested array of predicates:

$$\mathscr{L} = \forall x_1 \in X_1,\ (\exists x_2 \in X_2,\ (\dots\ , (\forall x_n \in X_n,\ \mathscr{P}(x_1,\dots,x_n)))).$$

Repeatedly applying theorem 3.5, from the outside to the inside, we obtain

$$\neg\mathscr{L} \Leftrightarrow \exists x_1 \in X_1,\ \forall x_2 \in X_2,\ \dots\ , \exists x_n \in X_n,\ \neg\mathscr{P}(x_1,\dots,x_n).$$

Namely, the negation of $\mathscr{L}$ is obtained by replacing each $\forall$ with $\exists$, and vice-versa, *without changing their order,* and then replacing $\mathscr{P}$ with $\neg\mathscr{P}$. A formal proof requires the principle of induction (see chapter 7).

EXAMPLE. The statement

$$\forall n,a,b \in \mathbb{Z},\ (n|ab) \Rightarrow (n|a \vee n|b)$$

is false (why?). We negate it using theorem 3.5, obtaining a true statement:

$$\exists n,a,b \in \mathbb{Z},\ (n|ab) \wedge (n \not| a \wedge n \not| b).$$

EXAMPLE. Lagrange's theorem, expressed symbolically in equation (3.24), would be false if four squares were replaced by three squares. This fact is written symbolically as

$$\exists n \in \mathbb{N},\ \forall a,b,c \in \mathbb{Z},\ n \neq a^2 + b^2 + c^2.$$

In words:

> *There is a natural number which cannot be written as the sum of three squares.*

3.6 Relations

We introduce some general terms whose definition requires predicates and quantifiers. These terms are ubiquitous in higher mathematics.

Let X and Y be sets. A **relation** $\mathscr{R}$ on $X \times Y$ is a predicate over $X \times Y$. If $X = Y$, we speak of a **relation on** X. It is customary to write $x\mathscr{R}y$ to mean $\mathscr{R}(x,y)$, and the expression $x\mathscr{R}y$ is called a **relational expression**. All relational expressions introduced at the beginning of this chapter are of this form. Thus the membership operator $\in$ defines a relation on $X \times \mathbf{P}(X)$, where X is some ambient set.

A relation on a set X is sometimes defined as a **subset** of X^2, rather than a predicate over X^2. In this sense, the set $\{(1,1),(2,1)\}$ is a relation on $\{1,2\}$. The correspondence between sets and predicates described in section 3.3 clarifies the connection between the two constructs.

A relation $\mathscr{R}$ on a set X is called an **equivalence relation** if it satisfies the following properties:

$$\begin{array}{rll} \forall x \in X, & x\mathscr{R}x & \textbf{reflexivity} \\ \forall x,y \in X, & x\mathscr{R}y \Rightarrow y\mathscr{R}x & \textbf{symmetry} \\ \forall x,y,z \in X, & (x\mathscr{R}y \wedge y\mathscr{R}z) \Rightarrow x\mathscr{R}z & \textbf{transitivity.} \end{array}$$

For an equivalence relation, the expression '$x \sim y$' is a common alternative to $x\mathscr{R}y$.

EXAMPLE. The relational operator '$=$' defines an equivalence relation on any set. This is the **trivial** equivalence.

EXAMPLE. For any natural number m, the congruence relation $x \equiv y \pmod{m}$ defined in section 2.1.2 is an equivalence relation on $\mathbb{Z}$.

EXAMPLE. The relation '$\sim$' on $\mathbb{N} \times \mathbb{N}$, defined by $(m,n) \sim (j,k)$ if $m+k = n+j$, is an equivalence relation. By interpreting the pair (m,n) as the quantity $z = m-n$, we see that equivalent pairs correspond to the same value of z. With this device, one can construct integers from pairs of natural numbers. More precisely, every integer is an infinite set of equivalent natural numbers. This is an **abstract definition** of the integers. The virtue of this construction is that it requires only addition in $\mathbb{N}$, not subtraction.

EXAMPLE. The relation '$\sim$' on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$, defined by $(m,n) \sim (j,k)$ if $mk = nj$, is an equivalence relation. By interpreting the pair (m,n) as $r = m/n$, we see that equivalent pairs correspond to the same value of r. With this device, one can define a rational number as an infinite collection of equivalent pairs of integers. This abstract definition of the rational numbers requires only multiplication in $\mathbb{Z}$, not division.

A relation $\mathscr{R}$ on a set X is called a **partial ordering** if it satisfies the following properties:

$$\begin{array}{rll} \forall x \in X, & x\mathscr{R}x & \textbf{reflexivity} \\ \forall x,y \in X, & (x\mathscr{R}y \wedge y\mathscr{R}x) \Rightarrow x = y & \textbf{anti-symmetry} \\ \forall x,y,z \in X, & (x\mathscr{R}y \wedge y\mathscr{R}z) \Rightarrow x\mathscr{R}z & \textbf{transitivity.} \end{array}$$

A set is **partially ordered** if a partial ordering is defined on it. A partial ordering is usually denoted by the symbol '$\leqslant$'. So we write $x \leqslant y$ instead of $x\mathscr{R}y$.

EXAMPLE. The relational operator $\leqslant$ defines a partial ordering in $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$, but not in $\mathbb{C}$.

EXAMPLE. The set $\mathbf{P}(X)$ of all subsets of a set X is partially ordered by set inclusion, whereby $\leqslant$ means $\subset$.

A partially ordered set X is said to be **ordered**, if all pairs of elements of X are **comparable**, meaning that we either have $x \leqslant y$ or $y \leqslant x$. The real line is an ordered set; the power set $\mathbf{P}(X)$ of a set X, which is partially ordered by set inclusion, is not ordered.

An ordered set X is said to be **well-ordered** if any non-empty subset $A \subset X$ has a smallest element. The symbolic definition requires three quantifiers.

$$\forall A \in \mathbf{P}(X) \smallsetminus \emptyset, \ \exists a \in A, \ \forall x \in A, \ a \leqslant x.$$

EXAMPLE. Any finite ordered set is well-ordered. The closed unit interval $[0,1]$ is ordered but not well-ordered, because the subset $(0,1]$ has no smallest element. The natural numbers are well-ordered. This property forms the basis of the principle of induction, which we consider in chapter 7.

Exercises

Exercise 3.1. Some straightforward proofs.

1. Show that the operators $\vee, \Rightarrow, \Leftrightarrow$ may be defined in terms of $\neg$ and $\wedge$.
2. Complete the proof of theorems 3.2 and 3.3.

3. Show that given any predicates $\mathscr{P},\mathscr{Q}$ over a set X, there is a subset A of X such that
$$(\forall x\in X,\ \mathscr{P}(x)\Rightarrow\mathscr{Q}(x))\ \Leftrightarrow\ (\forall x\in A,\ \mathscr{Q}(x)).$$
This equivalence says that the hypothesis of an implication may be absorbed into a suitable ambient set.

Exercise 3.2. Rewrite each expression with symbols, using the quantifier $\exists$.

1. $Y\not\supset X$
2. $X\cap Y\neq\emptyset$
3. $z\in X+Y$
4. $\#X>1$
5. $\sin\circ\cos\neq\cos\circ\sin$
6. $\#X>\#Y$.

Exercise 3.3. Write each statement with symbols, using at least one quantifier.

1. *The integer n is a cube.*
2. *The fraction a/b is not reduced.*
3. *The equation $f(x)=0$ has a rational solution.*
4. *The unit circle has a rational point.*
5. *The polynomial $p(x)$ has no integer roots.*
6. *The function $f:A\to B$ is not injective.*
7. *The function $f:A\to B$ is not surjective.*
8. *The function $f:A\to B$ is constant.*
9. *The function $f:A\to B$ is not constant.*
10. *Every cubic equation with integer coefficients has a real solution.*
11. *The integer n has a square divisor*[5].
12. *The integer n is not divisible by 3.*

---

[5] The square divisor 1^2 is excluded.

13. *There are integers that can be written in two different ways as the sum of two positive cubes.*[6]
14. *Any open interval contains a rational number.*
15. *No open interval contains a smallest element.*
16. *The equation* $f(x) = 0$ *has infinitely many integer solutions.*
17. *The equation* $f(x) = 0$ *has finitely many rational solutions.*

Exercise 3.4. Consider the following statements.

1. *Every prime greater than 2 is odd.*
2. *Two integers which are co-prime are also prime.*
3. *The reciprocal of a positive reduced fraction is reduced.*
4. *The inverse of an invertible matrix is invertible.*
5. *Every integer that divides the product of two integers also divides one of the factors.*
6. *Every subset of an infinite set is infinite.*
7. *Three consecutive odd integers greater than 3 cannot all be prime.*
8. *Every integer which is the product of two primes has four distinct divisors.*[7]

In each case:

i) Decide if the statement is true or false, hence rewrite it as an explicit implication (*For all . . . x, if x . . . , then . . .*).

ii) State the contrapositive.

iii) State the converse, and decide whether it's true or false.

iv) State the negation.

---

[6] See (3.4), p. 52.

[7] You may assume that all relevant integers are positive.

Exercise 3.5. (W. Hodges.) We wish to to build up a set of predicates to describe family relations. You are given the two predicates

x is a son of y *x is a daughter of y*.

Your task is to write definitions of the following predicates, in some appropriate order such that the later definitions use only the given predicates and earlier definitions. (The order below is just alphabetical.)

x is an aunt of y
x is a brother of y
x is a child of y
x is the father of y
x is female
x is a grandchild of y
x is a half-brother of y
x is male
x is the mother of y
x is a nephew of y
x is a parent of y
x is a sister of y

Use the symbols x,y,z, etc., to denote people, words for everything else, and parentheses to specify the order of evaluation of logical operators.

Exercise 3.6. Let x and y be natural numbers, and let $\mathscr{P}(x,y)$ mean: 'x is a proper divisor of y' (that is, $x|y$ and $x \neq 1, y$). Thus $\mathscr{P}$ is a predicate over $\mathbb{N} \times \mathbb{N}$.

(a) Write each statement with words. **[**$\not\mathscr{E}$**]**

1. $\exists x \in \mathbb{N},\ \ \mathscr{P}(x,5)$
2. $\exists x \in \mathbb{N},\ \ \mathscr{P}(5,x)$
3. $\forall x \in \mathbb{N},\ \ \mathscr{P}(x,x^2)$
4. $\exists x \in \mathbb{N},\ \forall y \in \mathbb{N},\ \ \mathscr{P}(x,y)$
5. $\forall x \in \mathbb{N} \smallsetminus \{1\},\ \exists y \in \mathbb{N},\ \ \mathscr{P}(x,y)$

(b) Write each predicate $\mathscr{Q}$ with words, apart from the predicate's argument. **[**$\not\mathscr{E}$**]**

1. $\mathscr{Q}(y) = \exists x \in \mathbb{N},\ \ \mathscr{P}(x,y)$

2. $\mathscr{Q}(y) = \forall x \in \mathbb{N},\ \neg\mathscr{P}(x,y)$
3. $\mathscr{Q}(x) = \exists y \in \mathbb{N},\ \mathscr{P}(x,y)$

(c) Define each set with words. **[$\not\varepsilon$]**

1. $\{y \in \mathbb{N} : \exists x \in \mathbb{N},\ \mathscr{P}(x,y)\}$
2. $\{y \in \mathbb{N} : \forall x \in \mathbb{N},\ \neg\mathscr{P}(x,y)\}$
3. $\{x \in \mathbb{N} : \exists y \in \mathbb{N},\ \mathscr{P}(x,y)\}$.

Exercise 3.7. The following expressions define sets; turn symbols into words. **[$\not\varepsilon$]**

1. $\{x \in \mathbb{N} : \forall m,n \in \mathbb{N},\ (x|mn) \Rightarrow (x|m \vee x|n)\}$
2. $\{x \in \mathbb{N} : \forall m \in \mathbb{N},\ x|m^2 \Rightarrow x|m\}$
3. $\{x \in \mathbb{N} : \forall m \in \mathbb{N},\ x|m^3 \Rightarrow x|m^2\}$.

Exercise 3.8. Decide if each sentence is true or false, hence write its negation with symbols. Then rewrite both with words. **[$\not\varepsilon$]**

1. $\forall n \in \mathbb{N},\ 1/n \notin \mathbb{N}$
2. $\forall n \in \mathbb{N},\ \sqrt{n} \in \mathbb{R} \setminus \mathbb{Q}$
3. $\forall x,y \in \mathbb{R},\ xy = yx$
4. $\forall n \in \mathbb{Z},\ 2 \,|\, n(n+1)$
5. $\forall m,n \in \mathbb{Z},\ (m+n \in 2\mathbb{Z}) \Rightarrow (m \notin 2\mathbb{Z} \vee n \notin 2\mathbb{Z})$
6. $\forall n \in \mathbb{N},\ \exists r \in \mathbb{Q},\ n < r^2 < n+1$
7. $\forall y \in \mathbb{R},\ \exists x \in \mathbb{R}^+,\ \log(x) = y$
8. $\forall x,y > 1,\ \exists n \in \mathbb{N},\ x^n > y$
9. $\forall \varepsilon > 0,\ \exists r \in \mathbb{Q},\ |r - \sqrt{2}| < \varepsilon$.

Exercise 3.9. Consider the sentences

1. *Everybody loves George*
2. *Everybody loves somebody.*
3. *Somebody loves everybody.*

We wish to determine the computational effort required to find out if a sentence is true or false. With the notation of section 3.4, let the set G have n elements. First suppose that the sentence is true. How many evaluations of the predicate 'x loves y' are needed to verify this? Give the minimum and maximum number of evaluations, with an explanation. Do the same assuming that the sentence is false.

Exercise 3.10. Define some interesting predicates on the power set of $\mathbb{Z}$. Do the same for the power set of $\mathbb{Z}[x]$.

Chapter 4

Describing functions

We know only a handful of words describing structural properties of functions:

> *injective, surjective, bijective (invertible), constant.*

In this chapter we consider attributes of **real functions** $f : \mathbb{R} \to \mathbb{R}$. We expand our vocabulary substantially, and introduce important terms that are found in more general settings (*periodic, bounded, continuous*). Then we export this terminology to the **real sequences**, which are functions defined over $\mathbb{N}$.

4.1 Ordering properties

The real line $\mathbb{R}$ is **ordered**, meaning that for any pair (x, y) of real numbers, precisely one of the three relational expressions

$$x < y \qquad\qquad x = y \qquad\qquad x > y$$

is true, and the other two are false. We begin with properties that are formulated in terms of ordering.

A simple attribute of a real function is the sign of the values it assumes.

$$\forall x \in \mathbb{R},\ \ f(x) > 0 \qquad\qquad \textbf{positive} \tag{4.1}$$

$$\forall x \in \mathbb{R},\ \ f(x) < 0 \qquad\qquad \textbf{negative} \tag{4.2}$$

If (4.1) is formulated with the non-strict inequality $f(x) \geqslant 0$, then we say that the function is **non-negative**. For the inequality $f(x) \leqslant 0$, the term ‘non-positive’ is uncommon, and one would normally use **negative or zero**. For example, the exponential function is positive, the absolute value function is non-negative, and the sine function is neither positive nor negative. Because inequalities are reversed under sign change, if a function f has any of the stated properties, then $-f$ has the complementary property (e.g, if f is positive, then $-f$ is negative).

Note that the terms 'non-negative' and 'not negative' have different meaning, the latter being the logical negation of negative. This distinction is very clear in the symbolic definitions.

$\forall x \in \mathbb{R},\ f(x) \geqslant 0$	**non-negative**
$\exists x \in \mathbb{R},\ f(x) \geqslant 0$	**not negative**

(The second symbolic expression is obtained by negating (4.2) according to theorem 3.5.) So a non-negative function is also not negative, but not vice-versa. Two similar expressions with different meanings can easily lead to confusion, and one must remain vigilant.

Next we consider how the action of a function affects the ordering of the real line; the order may be preserved, reversed, or a bit of both. There are two competing terminologies, labelled *I* and *II* in the table below. Each has advantages and disadvantages.

	I	*II*
$\forall x,y \in \mathbb{R},\ x > y \Rightarrow f(x) > f(y)$	**increasing**	**strictly increasing**
$\forall x,y \in \mathbb{R},\ x > y \Rightarrow f(x) \geqslant f(y)$	**non-decreasing**	**increasing**
$\forall x,y \in \mathbb{R},\ x > y \Rightarrow f(x) < f(y)$	**decreasing**	**strictly decreasing**
$\forall x,y \in \mathbb{R},\ x > y \Rightarrow f(x) \leqslant f(y)$	**non-increasing**	**decreasing**

A function that is either increasing or decreasing (strictly or otherwise) is said to be **monotonic.**

Thus the arc-tangent is increasing in *I* and strictly increasing in *II*. In *I* no function can be both increasing and decreasing, and most functions are neither, for instance a constant or the cosine. The constants are the functions that are both non-decreasing and non-increasing.

The disadvantage of *I* is that, as we did above for 'non-negative', we must differentiate between 'non-increasing' and the logical negation of increasing ($\exists x,y \in \mathbb{R},\ (x > y) \wedge (f(x) \leqslant f(y))$).

The terminology *II* eliminates the annoying distinction between the prefixes 'non-' and 'not', but introduces a new problem. Now a constant is both increasing and decreasing, clashing with common usage. (After years without a pay rise, I wouldn't say: 'my salary is increasing'.)

To be safe, use 'strictly' to mean strict inequality in any case.

4.2 Symmetries

Real functions may have symmetries, expressing invariance with respect to changes of the argument. A function f is **even** or **odd**, respectively, if

$$\forall x \in \mathbb{R},\ \ f(-x) = f(x) \qquad \text{or} \qquad \forall x \in \mathbb{R},\ \ f(-x) = -f(x).$$

So the cosine is even, the sine is odd, the exponential is neither even nor odd, and the zero function is both even and odd. The property of being even or odd has a geometrical meaning: graphs of even functions have a mirror symmetry with respect to the ordinate axis, while those of odd functions are symmetrical with respect to the origin (see figure 4.2). There is an easy way of constructing even/odd functions.

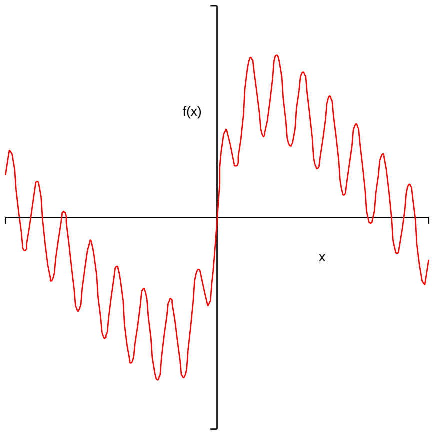

Figure 4.1: An odd function

For any real function g, the function $x \mapsto g(x)+g(-x)$ is even, and so is the function $g(f(x))$ for any even function f. So the function $x \mapsto g(|x|)$ is even. To construct odd functions, we first define the **sign function**

$$\mathrm{sign}(x) = \begin{cases} +1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases} \qquad x \in \mathbb{R}. \tag{4.3}$$

The sign function is odd: $\mathrm{sign}(-x) = -\mathrm{sign}(x)$. Then, for any real function g, the function $x \mapsto \mathrm{sign}(x)\,g(|x|)$ is odd, as easily verified. This construct ensures that our function vanishes at the origin, which is a property of all odd functions.

Next we turn to translational symmetry, which is called **periodicity**. A function f is **periodic with period** T if

$$\forall x \in \mathbb{R}, \;\; f(x+T) = f(x) \tag{4.4}$$

for some non-zero real number T (see figure 4.2).

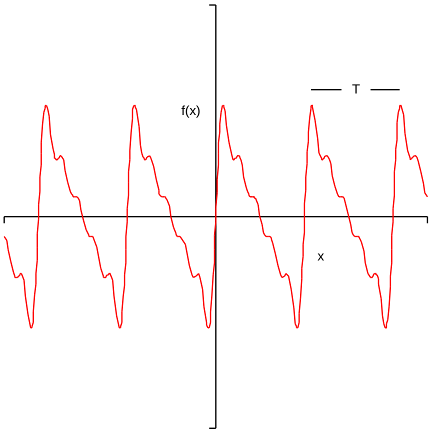

Figure 4.2: A periodic function

For instance, the sine function is periodic with period $T = 2\pi$. If a function is periodic with period T, then it is also periodic with period $2T$, $3T$, etc. For this reason one normally requires the period T to be the smallest positive real number for which (4.4) is satisfied. To emphasise this point we use the terms **minimal** or **fundamental period**.

If we say that a function f is **periodic** —without reference to a specific period— then the existence of the period must be required explicitly:

$$\exists T \in \mathbb{R}^+, \;\; \forall x \in \mathbb{R}, \;\; f(x+T) = f(x) \tag{4.5}$$

where $\mathbb{R}^+$ is the set of of positive real numbers —see (2.9). Note that the period T must be non-zero (lest this definition say nothing) and **without loss of generality**[1] we shall require the period to be positive. (The case of negative period is dealt with the substitution $x \mapsto x - T$ in (4.4).)

The presence of symmetries reduces the amount of information needed to specify a function. If a function is even or odd, knowledge of the function for non-negative values of the argument suffices to characterise it completely. Likewise, if a periodic function is known over any interval of length equal to the period, then the function is specified completely.

One should be aware that symmetries are special; a function chosen 'at random' will have no symmetry.

4.3 Boundedness

A set $X \subset \mathbb{R}$ is **bounded** if there is an interval containing it[2], namely if

$$\exists a, b \in \mathbb{R}, \ \forall x \in X, \ a < x < b \tag{4.6}$$

or

$$\exists b \in \mathbb{R}, \ \forall x \in X, \ |x| < b. \tag{4.7}$$

The two definitions are equivalent (exercise 4.11). The numbers a and b in (4.6) are an **upper** and a **lower bound** for X.

A real function f is **bounded** if its image $f(\mathbb{R})$ is a bounded set. In symbols:

$$\exists b \in \mathbb{R}, \ \forall x \in \mathbb{R}, \ |f(x)| < b.$$

For example, the sine function is bounded and the exponential is not. The periodic function displayed in figure 4.2 is bounded.

A function f is **bounded away from zero** if its reciprocal is bounded. This means that for some positive constant c we have $|f(x)| > c$ for all values of x. In symbols:

$$\exists c \in \mathbb{R}^+, \ \forall x \in \mathbb{R}, \ |f(x)| > c.$$

The hyperbolic cosine is bounded away from zero (what could be a value of c in this case?) but the exponential function is not.

---

[1] This expression indicates that an inessential restriction or simplification is being introduced. See section 6.8 for another example.

[2] The term 'interval' is intended in the proper sense —rays are excluded, cf. (2.10).

4.4 Neighbourhoods

A **neighbourhood** of a point $x \in \mathbb{R}$ is any *open* interval containing x. The interval is required to be open to ensure that x is not an end-point. Plainly, any real number has infinitely many neighbourhoods. Although this definition makes no reference to the size of the interval, this construct comes to life only when very small intervals come into play. Then the neighbourhood concept characterises 'proximity' in a concise manner that does not rely on quantitative information. A skillful use of this term leads to terse and incisive statements. For instance, the sentence

> *The function f is bounded in a neighbourhood of x_0.*

means that there is an open interval containing x_0 whose image under f is a bounded set. If we write this statement in symbols

$$\exists a, b \in \mathbb{R},\ \exists c \in \mathbb{R}^+,\ x_0 \in (a,b) \wedge (\forall x \in (a,b),\ |f(x)| < c)$$

we realise just how much information is packed into it.

The function f has a **maximum** at x if the value of f at x is greater than the value at all other points, namely if

$$\forall y \in \mathbb{R},\ f(x) > f(y). \tag{4.8}$$

The concept of **minimum** is defined similarly. With a non-strict inequality in (4.8) we obtain a weaker variant of this concept. The function f has a **local maximum** at x if (4.8) holds in some neighbourhood of x, but not necessarily in the whole domain of f. Thus the arc-tangent and the exponential have no maximum or minimum, the hyperbolic cosine has a minimum but no maximum, and the function $x \mapsto \arctan(x) + \sin(x)$ has infinitely many local maxima and minima, but no maximum or minimum.

The statement

> *Every neighbourhood of x contains a point of A.*

which is trivial if $x \in A$, has considerable depth if $x \notin A$, because it says that there are infinitely many open intervals, as small as we please, each containing x as well as a point in A. In fact every neighbourhood of x must contain *infinitely many points* of A; equivalently, points in A get **arbitrarily close to** x. (Think about it.) For instance, let A be one of

$$(0,1) \qquad \{1/n : n \in \mathbb{N}\} \qquad \sin(\mathbb{N}).$$

We see that in all cases 0 does not belong to A (why is $0 \notin \sin(\mathbb{N})$?), and the above applies.

By a **neighbourhood of infinity** we mean a ray $\{x \in \mathbb{R} : x > a\}$, where a is a real number. A neighbourhood of $-\infty$ is defined similarly, and both points at infinity $\pm\infty$ are handled at once with the construct $\{x \in \mathbb{R} : |x| > a\}$. The sentence

The function f is constant for all sufficiently large x.

written symbolically as

$$\exists a \in \mathbb{R},\ \forall x \in \mathbb{R}^+,\ \ f(x+a) = f(a).$$

or as

$$\exists a \in \mathbb{R},\ \forall x > a,\ \ f(x) = f(a)$$

says that f is constant in some neighbourhood of infinity.

Let f be defined in a neighbourhood of a point x, excluding, possibly, the point x itself. If f is unbounded in every neighbourhood of x, then we say that f is **singular at** x, and the point x is called a **singularity** (or a **singular point**) of the function. A function that is not singular is said to be **well-behaved** or **regular**, the latter term used mostly for function of a complex variable. Singularities are common; we collect some ideas in a paragraph.

A polynomial has no singularities; the same holds for the sine, the co-sine, and the exponential. The logarithm is singular at the origin and regular everywhere else. A rational function has finitely many singularities; they are the roots of the polynomial at the denominator. The tangent, secant, and co-secant functions have infinitely many singularities, evenly spaced along the real line.

We denote the set of all neighbourhoods of x by $\mathscr{N}_x$. This is an **abstract set**, and we look for a concrete **representation** of it (see section 2.1.3). We identify an open interval with an ordered pair (a,b) of real numbers representing the interval's end-points[3]. Collecting all these pairs together, we obtain a subset N_x of $\mathbb{R}^2$:

$$N_x = \{(a,b) \in \mathbb{R}^2 : a < x < b\} \sim \mathscr{N}_x. \tag{4.9}$$

The set N_x is the (open) north-west quadrant of the plane with respect to the origin (x,x). The symbol '$\sim$' in (4.9) indicates that the sets $\mathscr{N}_x$ and N_x are **equivalent** rather than **equal**, meaning that there is a bijective function from $\mathscr{N}_x$ to N_x.

---

[3] By a Freudian coincidence, the ordered pair (a,b) here represents the open interval (a,b).

4.5 Continuity

Continuity is a fundamental concept in the theory of functions. Loosely speaking, a function is continuous if it has ‘no jumps’, but this naive definition is only adequate for simple situations. We begin by considering continuity at a specific point of the domain of a function. An informal —yet accurate— characterisation of continuity is to say that a function is continuous at a point if its value there can be inferred unequivocally from its values at neighbouring points. Thus neighbourhoods come into play.

For example, the value of the sign function (4.3) at the origin cannot be inferred from the surrounding environment; even if we defined, say, $\mathrm{sign}(0) = 1$, the ambiguity would remain. Things can get a lot worse. Consider the real function

$$x \mapsto \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q}. \end{cases} \tag{4.10}$$

Given that in any neighbourhood of any real number there are both rational and irrational numbers, there is no way of inferring the value of this function at any point by considering how the function behaves in the surrounding region.

We say that f is **continuous at the point** a if

$$\forall J \in \mathscr{N}_{f(a)},\ \exists I \in \mathscr{N}_a,\ \ f(I) \subset J. \tag{4.11}$$

Let’s go though this concise definition in slow motion. The set $\mathscr{N}_a$ consists of all open intervals containing the point a where continuity is being tested. The set $\mathscr{N}_{f(a)}$ consists of all neighbourhoods of $f(a)$, the value of the function at a. An arbitrary interval J containing $f(a)$ is given, typically very small. We must now find an interval I containing a whose image fits inside J. The choice of I will depend on J, and if we succeed in all cases, then the function is continuous at a. We don’t necessarily know what the set $f(I)$ looks like; all we need to know is that if I becomes small, so does $f(I)$.

A function that is not continuous at a point of the domain is said to be **discontinuous**. A function is **discontinuous everywhere** if it has no points of continuity, like the function (4.10).

Continuity may be defined without reference to neighbourhoods, but in this case one must supply qualitative information. The full notation is considerably more involved than (4.11):

$$\forall \varepsilon \in \mathbb{R}^+,\ \exists \delta \in \mathbb{R}^+,\ \forall x \in \mathbb{R},\ \ |x-a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon$$

and it remains so even if we omit all references to the ambient set $\mathbb{R}$:

$$\forall\varepsilon > 0,\ \exists\delta > 0,\ \forall x,\ |x-a| < \delta \Rightarrow |f(x)-f(a)| < \varepsilon.$$

A function is **continuous** if it is continuous at all points of the domain. An additional quantifier is required in definition (4.11). If A is the domain of f, then we write

$$\forall a \in A,\ \forall J \in \mathscr{N}_{f(a)},\ \exists I \in \mathscr{N}_a,\ f(I) \subset J. \tag{4.12}$$

A function f is **differentiable at** a if the limit

$$\lim_{x\to a} F(x) \qquad \text{where} \qquad F(x) = \frac{f(x)-f(a)}{x-a} \qquad x \neq a$$

exists. The function F is called the **incremental ratio** of f at a, which is not defined at $x = a$. However, if f is differentiable at a, then by letting $F(a) \overset{\text{def}}{=} \lim_{x\to a} F(x)$, the function F becomes **continuous** at a. Let's sum this up in a sentence.

> *A function is differentiable at a point if its incremental ratio is continuous at that point; in this case the value of the incremental ratio is the derivative of the function.*

A function is said to be **differentiable** if it is differentiable at all points of its domain. A function that is differentiable **sufficiently often** (all derivatives up to a sufficiently high order exist) is said to be **smooth**. The expression 'sufficiently often' is deliberately vague; its precise meaning will depend on the context.

Many elementary real functions are continuous in their respective domains: the polynomials, the trigonometric functions, and the exponential and logarithmic functions, etc. These functions are also differentiable infinitely often.

4.6 Other properties

A function of the form $f(x) = ax$, where a is a real number, is said to be **linear**. The term linear originates from 'line', and functions of the type $f(x) = ax + b$, are sometimes referred to as being 'linear' because their graph is a line (the correct term is **affine**). Likewise, a function given by a polynomial will be characterised, in the first instance, by the degree of the polynomial. So we'll speak of a **quadratic, cubic, quartic** function, etc.

Consider the absolute value function, defined as follows:

$$|x| = \begin{cases} x & \text{if } x \geqslant 0 \\ -x & \text{if } x < 0. \end{cases} \tag{4.13}$$

This function is not linear, but it is made of two linear pieces, glued together at the origin. Functions made of linear or affine pieces are said to be **piecewise linear** or **piecewise affine** (see equation (4.3) and figure 4.6). A function is **piecewise defined** if its domain is partitioned into disjoint intervals or rays, with the function being specified independently over each interval. The properties of a piecewise-defined function may fail at the end-points of the intervals of definition. In this case terms such as **piecewise increasing**, **piecewise continuous**, **piecewise differentiable, piecewise smooth** may be used —see figure 4.4. A piecewise constant function is called a **step-function**.

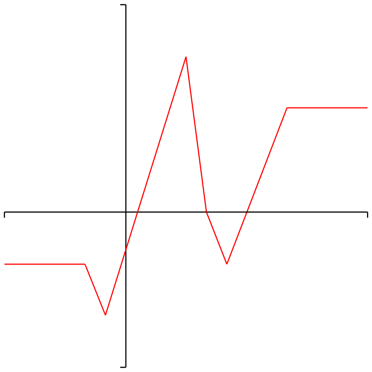

Figure 4.3: A continuous piecewise affine function.

Example. If x is a real number, we define the **floor** of x, denoted by $\lfloor x \rfloor$, to be the largest integer not exceeding x. Similarly, the **ceiling** function $\lceil x \rceil$ represents the smallest integer not smaller than x. Floor and ceiling are the prototype step functions. Closely connected to the floor is the **fractional part** of a real number x, denoted by $\{x\}$. (The notational clash with the set having x as its only element is one of the most spectacular in mathematics!) The fractional part is defined via the equation

$$x = \lfloor x \rfloor + \{x\}$$

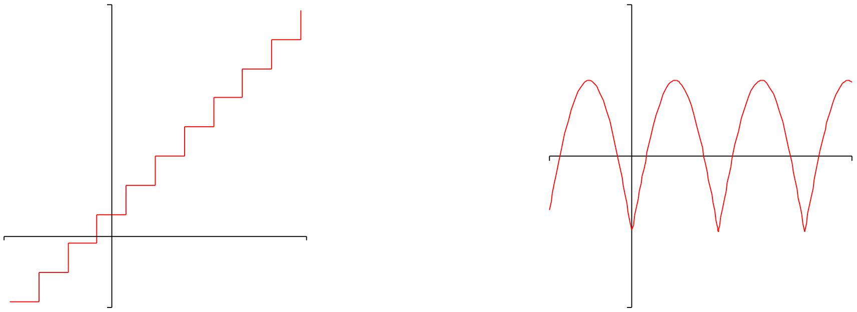

Figure 4.4: Left: a step function. (The vertical segments are just a guide to the eye; they are not part of the graph of the function.) Right: a periodic continuous function which is piecewise differentiable.

from which it follows that $0 \leqslant \{x\} < 1$. This function is discontinuous and piecewise affine.

Example. Describe the following function: **[**$\not\varepsilon$**]**

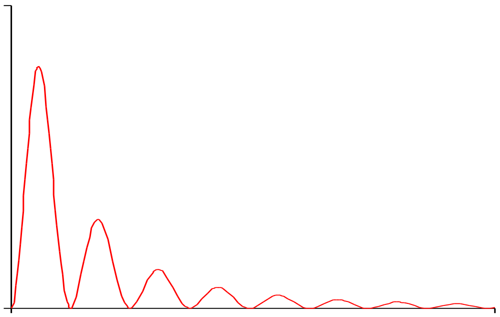

> *This is a smooth function, which is bounded and non-negative. It features an infinite sequence of evenly spaced local maxima, whose height decreases monotonically to zero for large arguments. There is one zero of the function between any two consecutive local maxima.*

We rewrite it, borrowing some terminology from physics.

> *This function displays regular oscillations of constant period, with amplitude decreasing monotonically to zero.*

The following functions have a behaviour that is qualitatively similar to that displayed above.

$$f(x) = \sin^2(x)e^{-x} \qquad\qquad f(x) = \frac{1+\sin(x)}{1+x^2}.$$

4.7 Describing real sequences

A sequence may be thought of as a function defined over the natural numbers, or, more generally, over a subset of the integers (section 2.4). Indeed in number theory the real (or complex) sequences are called **arithmetical functions.** Using this analogy, some terminology introduced for real functions translates literally to real sequences. So the terms

> *positive, negative, increasing, decreasing, monotonic,*
> *constant, periodic, bounded, bounded away from zero*

have the same meaning for sequences as they have for functions. We write

> *The sequence of primes is positive, increasing, and unbounded.*

Other terms require amendments or are simply not relevant to sequences. For instance, injectivity is not used —we just say that the terms of a sequence are distinct— while surjectivity is rarely significant. Invertibility is used in a different sense than for functions —see section 8.4. The terms ‘even’ and ‘odd’ are still applicable to doubly-infinite sequences.

A sequence which settles down to a periodic pattern from a certain point on is said to be **eventually periodic**. More precisely, a sequence $(x_1, x_2, \dots)$ is eventually periodic if the set $\{x_1, x_2, \dots\}$ is finite. (Think about it.) If the periodic pattern consists of a single term, then the sequence is said to be **eventually constant**.

The idea of continuity does not apply to sequences because the neighbourhoods of an integer are trivial (small neighbourhoods contain a single point). Except at infinity, that is. The **neighbourhoods of infinity** are the sets of the form $I_K = \{k \in \mathbb{N} : k > K\}$, for some integer K. They form a sequence of nested infinite sets, which is appropriate for defining continuity.

Accordingly, we say that a real sequence **converges** if, as a function over $\mathbb{N}$, it's **continuous at infinity**. Adapting definition (4.11) to the present situation, we say that the real sequence (a_k) converges to the limit c if

$$\forall J \in \mathscr{N}_c, \ \exists K \in \mathbb{N}, \ \forall k > K, \ a_k \in J.$$

This expression says that given any neighbourhood J of c, we have $a_k \in J$ **for all sufficiently large** k, namely for all $k \in I_K$.

Exercises

Exercise 4.1. Write each symbolic sentence without symbols, apart from f.

1. $f(0) = 0$
2. $f(\mathbb{R}) = \mathbb{R}$
3. $\# f(\mathbb{R}) = 1$
4. $f(\mathbb{Z}) = \{0\}$
5. $0 \in f(\mathbb{Z})$
6. $f^{-1}(\mathbb{Z}) = \emptyset$
7. $f(\mathbb{R}) \subset \mathbb{Z}$
8. $f(\mathbb{R}) \supset \mathbb{Z}$
9. $f = f^{-1}$
10. $f(\mathbb{Q}) \cap \mathbb{Q} = \emptyset$.

Exercise 4.2. Same as in previous exercise.

1. $\forall x \in \mathbb{Z},\ f(x) = 0$
2. $\forall x \in \mathbb{R},\ f(f(x)) = x$
3. $\forall x \in \mathbb{R}^+,\ f(-x) = 0$
4. $\forall x \in [0,1],\ \ f(x) \neq 0$
5. $\forall x \in \mathbb{Q} \smallsetminus \{0\},\ f(x) \neq 0$
6. $\forall x \in \mathbb{Q},\ \ f(x) \notin \mathbb{Q}$
7. $\forall x \in \mathbb{Z},\ \ \exists y \in \mathbb{N},\ \ f(x+y) = 0$
8. $\forall x \in \mathbb{R},\ \ \exists y \in \mathbb{R}^+,\ \ |f(x+y)| < |f(x)|$.

Exercise 4.3. Turn words into symbols.

1. *The function f has no zero.*
2. *The function f vanishes at the origin.*
3. *The function f does not vanish except, possibly, at the origin.*
4. *The function f vanishes identically outside the open unit interval.*
5. *The zeros of f include all even integers.*
6. *The function f may vanish only at the rationals.*
7. *The function f has infinitely many zeros in the unit interval.*

Exercise 4.4. Consider the following implications, where f is a real function.

1. *If f is decreasing, then $-f$ is increasing.*
2. *If f is decreasing, then $|f|$ is increasing.*
3. *If $|f|$ is increasing, then f is monotonic.*
4. *If f is even, then $|f|$ is also even.*
5. *If f is unbounded, then f is surjective.*
6. *If f is continuous, then $|f|$ is continuous.*
7. *If f is differentiable, then $|f|$ is differentiable.*
8. *If f is periodic, then f^2 is periodic.*

Of each implication:

(a) state the converse, and decide whether it's true or false;
(b) state the contrapositive, and decide whether it's true or false;
(c) state the negation.

Exercise 4.5. As in exercise 3.4, p. 75.

1. *Every differentiable real function is continuous.*
2. *A bounded real function cannot be injective.*
3. *The sum of two odd functions is an odd function.*
4. *No polynomial function is bounded.*
5. *The composition of two periodic functions is periodic.*

Exercise 4.6. Consider the definition (4.5) of a periodic function. What happens if we place the quantifiers in reverse order?

$$\forall x \in \mathbb{R},\ \exists T \in \mathbb{R}^*,\ \ f(x+T) = f(x) \tag{4.14}$$
$$\forall x \in \mathbb{R},\ \exists T \in \mathbb{R}^+,\ \ f(x+T) = f(x). \tag{4.15}$$

1. Find a function that satisfies (4.14) but not (4.5).
2. Find a function that satisfies (4.15) but not (4.5).
3. Characterize the set of functions that satisfy (4.14). Do the same for (4.15).

Exercise 4.7. Describe the behaviour of the following functions. **[**¢, 30**]**

1)

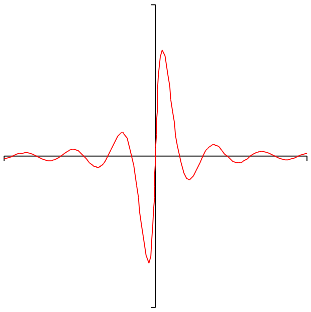

2)

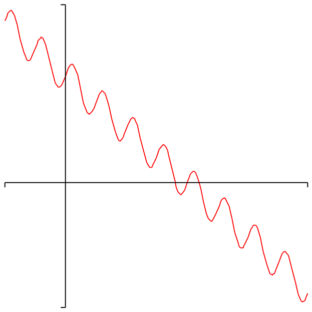

3)

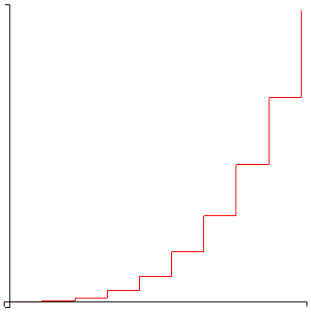

4)

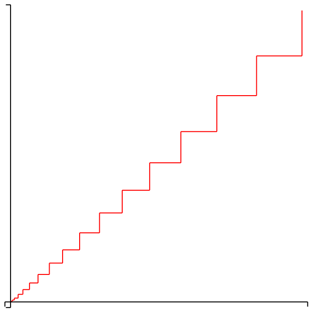

5)

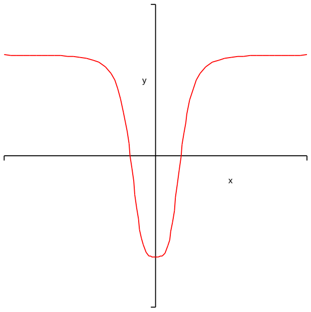

6)

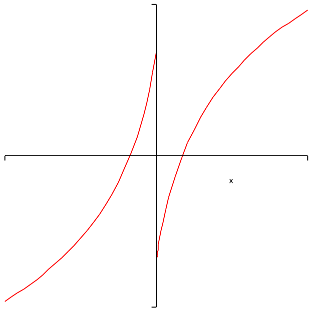

7)

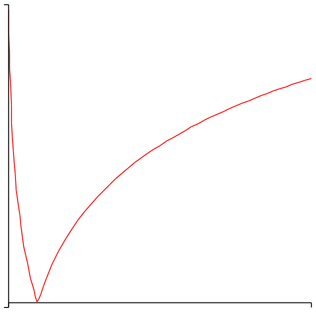

8)

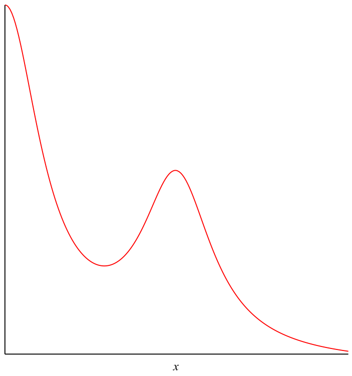

9)

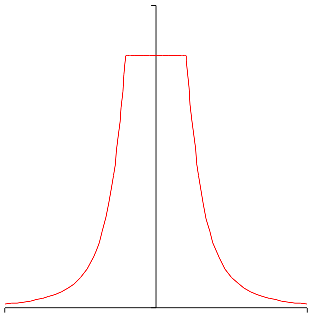

10)

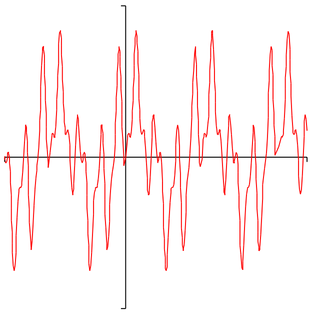

Exercise 4.8. By experimenting on a computer, if necessary, define real functions whose behaviour is qualitatively similar to those of the previous problem.

Exercise 4.9. Give an example of a real function with the stated properties.

1. Increasing, with decreasing derivative.
2. Unbounded, with unbounded derivative.
3. Unbounded, with bounded derivative.
4. Discontinuous, with continuous absolute value.
5. With infinitely many zeros, but not periodic.
6. Unbounded and vanishing infinitely often.
7. The image is a half-open interval.
8. The image is the set of natural numbers.

Exercise 4.10. Express each statement with symbols.

1. *The sequences* (a_k) *and* (b_k) *are distinct.*
2. *The sequence* (a_k) *is eventually constant.*
3. *The sequence* (a_k) *is not periodic.*
4. *The sequence* (a_k) *is eventually periodic.*
5. *The sequence* (a_k) *has infinitely many negative terms.*
6. *Eventually, all terms of the sequence* (a_k) *become negative.*
7. *The terms of the sequence* (a_k) *get arbitrarily close to zero.*
8. *Each term of the sequence* (a_k) *appears infinitely often.*
9. *Each term of the sequence* (a_k) *appears at least twice.*
10. *Each natural number appears infinitely often in the sequence* (a_k)*.*

Exercise 4.11. Prove that the definitions of boundedness (4.6) and (4.7) are equivalent.

Chapter 5

Writing well

In this chapter we consider some techniques for writing mathematics. We deal with small-scale features: choosing an appropriate terminology and notation, writing clear formulae, mixing words and symbols, writing definitions, introducing a new concept. These are pre-requisites for the more substantial task of structuring and delivering a mathematical argument, to be tackled in later chapters.

5.1 Choosing words

Precision must be the primary concern of anyone who writes mathematics. We list some common mistakes and inaccuracies that result from a poor choice of words.

Bad: the equation $x-3 \leqslant 0$
Good: the inequality $x-3 \leqslant 0$

Bad: the equation $x^2-1=(x-1)(x+1)$
Good: the identity $x^2-1=(x-1)(x+1)$

Bad: the identity $x=\sqrt{x^2}$
Good: the equation $x=\sqrt{x^2}$

Bad: the solution of $x^k=x$
Good: a solution of $x^k=x$

Bad: the minima of a quadratic function
Good: the minimum of a quadratic function

Bad: the function $\sin(x)$
Good: the sine function

Bad: the function $f(A)$ of the set A
Good: the image of the set A under f

Bad: the interval $[1,\infty)$
Good: the ray $[1,\infty)$ (the infinite interval $[1,\infty)$)

Bad: the set $\mathbb{Z}$ minus $k\mathbb{Z}$
Good: the set difference of $\mathbb{Z}$ and $k\mathbb{Z}$

Bad: the area of the unit circle
Good: the area of the unit disc

Bad: the absolute value is positive
Good: the absolute value is non-negative

Bad: the coordinates of a complex number
Good: the real and imaginary parts of a complex number

Bad: the function crosses the vertical axis at a positive point
Good: the graph of the function intersects the ordinate axis at a positive point

Once our writing is accurate, we consider refining the choice of words to differentiate meaning, improve legibility, or avoid repetition.

A set is often called a set, although **collection** or **family** are pleasing alternatives. A set becomes a **space** if it has an added structure, like a **metric space** (a set with a distance) or a **vector space** (a set with an addition and a scalar multiplication).

The word **element** expresses a special kind of subsidiary relationship. If A is a set and $a \in A$, then we say that a is an element of A. The term **member** is a variant, and so is **point**, which is the default choice for geometrical sets. So if $A \subset \mathbb{R}^n$, then x is a **point** of A. For sequences, we may replace 'element' by **term**. This is necessary if the elements of a sequence are added (multiplied) together, in which case the operands are the terms of the sum (product), not the elements. However, if $V = (v_k)$ is a vector, then v_k is a **component** of V, not a term (even though a vector is just a finite sequence). But if $M = (m_{i,j})$ is a matrix, then $m_{i,j}$ is an element or an **entry** of M, not a component (even though a matrix is just a sequence of vectors).

The word **variable** is used in connection with functions and equations. In the former case it has the same meaning as **argument**, and it refers to the function's input data. In the latter case it means **unknown** —a quantity whose value is to be found. Polynomials and rational functions may represent functions or algebraic objects; for the latter, the term **indeterminate** is preferable to variable.

The term **parameter** is used to identify a variable which is assigned a value that remains fixed in the subsequent discussion. For example, the indefinite integral of a function is a function which depends on a parameter, the integration constant:

$$\int g(x)dx = f(x) + c$$

The two symbols x and c play very different roles here, so we have a **one-parameter family** of functions, rather than a function of two variables.

A function is not always called a function. This may be true for real functions, but in more general settings the terms **map**, **mapping**, and **transformation** are at least as common. So a complex function $\mathbb{C} \to \mathbb{C}$ may be called a mapping, and a function between euclidean spaces $\mathbb{R}^n \to \mathbb{R}^m$ is a mapping or a transformation. Some functions are called **operators**. These include basic functions, like addition of numbers or intersection of sets, represented by specialised symbols ($+, \cap$) and syntax ($x+y$ rather than $+(x,y)$). The **arithmetical operators**, the **set operators**, the **relational operators**, and the **logical operators** are functions of this type. The attribute **binary** means that there are two operands, while **unary** is used for single operands, like changing the sign of a number. The term **operator** is also used to characterise functions acting on functions which produce other functions as a result. The **differentiation operator** is a familiar example. A real-valued function acting on functions is called a **functional**. So definite integration is a functional.

In logic, a function is called a **predicate,** a **boolean function,** or a **characteristic function**. These terms represent the same thing, but they are not completely interchangeable. 'Characteristic function' should be used if there is explicit reference to the associated set, and 'predicate' (or 'boolean function') otherwise. If the arguments of a function are also boolean, then there is a strong case for using 'boolean function'. Some examples:

> *The negation of a predicate is a predicate.*
>
> *Let χ be the characteristic function of the prime numbers.*
>
> *Let us consider the boolean function $(x,y) \mapsto \neg x \wedge y$.*
>
> *The predicate '$n \mapsto (3 \mid n(n+1))$' is the characteristic function of a proper subset of the integers.*
>
> *A relational operator is a boolean function of two variables.*

In mathematics it is acceptable to change the meaning of established words. The following re-definitions of universal geometrical terms were taken from mathematics research literature:

> *By a* **triangle** *we mean a metric space of cardinality three.*
>
> *By a* **segment** *we mean a maximal subpath of P that contains only light or only heavy edges.*
>
> *By a* **circle** *we mean an affinoid isomorphic to* $\max \mathbf{C}_p(T, T-1)$.

Re-defining such common words requires some self-confidence, but in an appropriate context this provocative device may be quite effective. Clearly the new meaning is meant to remain confined to the document in which it's introduced.

5.2 Choosing symbols

Choosing mathematical notation is difficult. Mathematicians are notoriously reluctant to accept standardisation of notation, to a degree unknown in other disciplines. Indeed, the ability to adjust quickly to new notation is regarded as one of the skills of the trade. The reality is somewhat different: absorbing new notation requires effort, and most people would gladly avoid it. So if we intend to communicate mathematics without confusing or alienating our audience, the notation must be simple, logical, and consistent.

Two golden rules should inform the use of symbols:

- DO NOT INTRODUCE UNNECESSARY SYMBOLS.

- DEFINE EVERY SYMBOL BEFORE USING IT.

Once defined, a symbol should be used consistently. Never use the same symbol for two different things or two symbols for the same thing, even in instances appearing far apart in the document. Don't write 'A_j, for $1 \leqslant j < n$' in one place and 'A_k, for $1 \leqslant k < n$' in another, unless there is a good reason to do so. Such small inconsistencies produce some 'notational pollution'. As the pollution piles up, reading becomes tiresome.

In this section we offer guidelines on how to choose symbols. These are not laws, and may be adapted to one's taste or rejected. What's important is to develop awareness of notation, and to make conscious decisions about it.

SETS. Represent sets by capital letters, Roman or Greek, such as

$$S \qquad \mathscr{A} \qquad \Omega.$$

The large variety of fonts available in modern typesetting systems increases our choice. When dealing with generic sets, then A,B,C or X,Y,Z are good symbols. For specific sets, choose a symbol that will remind the reader of the nature of the set. So, for an **alphabet** $\{a,b,c,\dots\}$, the symbols $A,\mathscr{A}$ are obvious choices; likewise, F is appropriate for a set of functions, etc.

Lower-case symbols like x,y represent the elements of a set. So $x \in A$ is a good notation, $X \in A$ is bad, and $X \in a$ is very bad. If more than one set is involved, consider using matching symbols. Thus

$$a \in A \qquad b \in B \qquad c \in C$$

is more coherent than

$$x \in A \qquad y \in B \qquad z \in C.$$

Some thought may be required for sets of sets. Consider the expressions

$$f(A \cap B) \qquad\qquad f(x \cap y).$$

The left expression will be interpreted as the image of the intersection of two sets under a function. In this case $A \cap B$ represents a subset of the domain of f. However, suppose that the domain of f is a set of sets (e.g., a power set, see section 2.1). Then the argument of f is a set, and this expression becomes dangerously ambiguous. The notation in the right expression removes this ambiguity. The combination of standard symbols x,y for variables and a set operator makes it clear that the argument $x \cap y$ of the function is an element, rather than a subset, of the domain.

INTEGERS. When choosing a symbol for an integer, begin from the middle region of the Roman alphabet

$$i,j,k,l,m,n \tag{5.1}$$

particularly if an integer is used as subscript or superscript[1]. However, use p for a prime number, and q if there is a prime different from p. The list of adjacent letters (5.1) cannot be extended; the preceding symbols, f,g,h are typical function names (see below), while the symbol that follows, o, is rarely used, not only for its resemblance to 0 (zero), but also because it has an established meaning in asymptotic analysis (it appears in expressions of the form $o(\log x)$). A capital letter in the list (5.1) may be used to represent a large integer, or combined with lower-case letters to denote an integer range: $n = 1, \ldots, N$. We'll return to this point in section 5.3, in connection with sums and products.

---

[1] Physicists use Greek letters for subscripts and superscripts.

RATIONALS. For rational numbers, use lower-case Roman letters in the ranges *a–e*, or *p–z*. The notation

$$r = \frac{m}{n}$$

is good, because 'r' reminds us of 'rational', while numerator and denominator conform to the convention for integers. If there is more than one rational, use adjacent symbols, s, t in this case.

REAL NUMBERS. For real numbers, use the same part of the Roman alphabet as for rationals, or the Greek alphabet:

$$\alpha, \beta, \gamma, \ldots$$

If there are both rational and real numbers and if the distinction between them is important, then use Roman for the rationals and Greek for the reals.

Some Greek symbols have preferential meaning: small quantities are usually represented by ε, δ, while for angles one uses ϕ, φ, θ. Famous real constants have dedicated symbols:

$$\begin{aligned} \pi =& 3.141592653\ldots && \textbf{Archimedes' constant } (\text{or } \textbf{Pi}) \\ e =& 2.718281828\ldots && \textbf{Napier's constant}^2 \\ \gamma =& 0.333177924\ldots && \textbf{Euler-Mascheroni constant}^3\textbf{.} \end{aligned} \qquad (5.2)$$

COMPLEX NUMBERS. Complex numbers tend to occupy the end of the Roman alphabet, and your first choice should be z or w. On the complex plane, we write $z = x + iy$, where x and y are the real and imaginary parts of z, and i is **the imaginary unit**. (However, number theorists use $\sqrt{-1}$, not i.) In polar coordinates, the standard notation is $z = \rho e^{i\theta}$. Be careful not to use i for any other purpose, such as a summation index.

UNKNOWNS. The quintessential symbol for an equation's unknown is x, invariably followed by y and z if there are more unknowns. For a large number of unknowns it is necessary to use sequence notation $x_1, \ldots, x_n$. This notation is also appropriate for the indeterminates of a polynomial or the arguments of a function.

COMPOSITE OBJECTS. Composite objects (groups, graphs, matrices) are best represented with capital letters, Roman or Greek. So use G or Γ for a group or a graph, and M for a matrix. If you have two groups, use adjacent symbols, like G and H. As with sets, for these objects' components consider using matching symbols, e.g., $g \in G$. A notable exception are graphs, where v and e are invariably used for vertices and edges, respectively.

FUNCTIONS. The default choice for a function's name is, of course, f, and if there is more than one function, use the adjacent symbols g, h. Lower-case symbols work well with any number of variables: $f(x)$, $f(x,y,z)$, $f(x_1,\ldots,x_n)$. But if the co-domain of a function is a cartesian product, then the function is a composite object —a vector, in fact— and capital letters are preferable. So a real function of two variables may be specified as

$$F:\mathbb{R}^2 \to \mathbb{R}^2 \qquad (x,y) \mapsto (f_1(x,y), f_2(x,y)).$$

Greek symbols, either capital or lower-case, are also commonly used for functions' names. The contrast between Roman and Greek symbols may be exploited to separate out the symbols' roles, as in $\mu(x)$ or $f(\lambda)$.

Some famous functions are named after, and represented by, a symbol (often a Greek one), thereby creating a strong bond between object and notation. The best known are **Euler's gamma function**[4] Γ

$$\Gamma(s) = \int_0^\infty x^{s-1}e^{-x}dx$$

(the extension of the factorial function to complex arguments), and **Riemann's zeta-function**[5] ζ

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}. \tag{5.3}$$

There is a peculiar notation for this function: its complex argument s is commonly written as $s = \sigma + i\tau$, with σ and τ real numbers. Other functions with dedicated notation are Euler's φ-function (see section 5.3), Dedekind's η-function, Kroneker's δ-function, Weierstrass' $\mathscr{P}$-function, Lambert's $\mathscr{W}$-function, etc.

SEQUENCES AND VECTORS. Sequences pose specific notational problems due to the presence of indices. Consider the various possibilities listed in (2.20), p. 31:

---

[4] Leonhard Euler (Swiss: 1707–1783). An accessible account of Euler's mathematics is found in [10].

[5] Berhard Riemann (German: 1826–1866)

which one should we choose? Be guided by the principle of economy: a symbol should be introduced only if it's strictly necessary. So the notation (a_k) is quite adequate for a generic sequence, or if the specific properties of the sequence (the initial value of the index, its finiteness) are not relevant. When more information is needed, the notation $(a_k)_{k\geqslant 1}$ is more economical than $(a_k)_{k=1}^{\infty}$, but the latter may be a better choice if it is to be contrasted with $(a_k)_{k=1}^{n}$. In turn, the latter is not as friendly as $(a_1,\dots,a_n)$, although it is more concise.

If a sequence is referred to often, even the stripped down notation (a_k) could become heavy, and it may be advisable to allocate a symbol for the sequence

$$a=(a_1,a_2,\dots) \qquad\qquad \mathbf{v}=(v_1,\dots,v_n).$$

As usual, we have employed matching symbols, using, respectively, a minimalist lower-case Roman character and a lower-case boldface character which is common for vectors. When using ellipses, two or three terms of the sequence usually suffice, but there are circumstances where more terms or a different arrangement of terms is needed.

For example, in the expression

$$(1+x,1+x^2,\dots,1+x^{2^k},\dots)$$

the insertion of the general term removes any ambiguity, while the ellipsis on the right suggests that the sequence is infinite —cf. (2.19). The notation

$$(a_1,\dots,a_{k-1},a_{k+1},\dots,a_n)$$

denotes a subsequence of a finite sequence obtained by deleting the k-th term, for an unspecified value of $k\neq 1,n$.

Things get complicated with sequences of sequences. This situation is not at all unusual; for instance, we may have a sequence of vectors whose components must be referred to explicitly. We write

$$V=(V_1,V_2,\dots) \qquad \text{or} \qquad \mathbf{v}=(\mathbf{v}_1,\mathbf{v}_2,\dots).$$

Let V_k (or $\mathbf{v}_k$) be the general term of our sequence. How are we to represent its components? As usual, we choose the matching symbol v, with a subscript indicating the component. However, the integer k has to appear somewhere, and its range must be specified. It is advisable to keep k out of the way as much as possible:

$$V_k=(v_1^{(k)},\dots,v_n^{(k)}) \qquad k=1,2,\dots.$$

In this expression we have used ellipses to specify the two ranges of indices; we could have used inequalities as well:

$$V_k = (v_j^{(k)}) \qquad 1 \leqslant j \leqslant n, \quad k \geqslant 1.$$

The parentheses are obviously needed for the superscript, for otherwise v_i^k would be interpreted as v_i raised to the k-th power. However, it may just happen that we need to raise the vector components to some power. Clearly we can't use adjacent superscripts $v_3^{(2)\,4}$, so parentheses are needed, but the straightforward notation $(v_3^{(2)})^4$ is awkward. For a more elegant solution, we represent k as an additional subscript, adopting, in effect, matrix notation:

$$V_k = (v_{1,k}, \dots, v_{n,k}) \qquad k \geqslant 1.$$

As a side note, one should keep in mind that with vectors the multiplication symbols '$\cdot$' and '$\times$' are reserved for the scalar and vector products, respectively —cf. (2.8), p. 19. Hence for scalar multiplication we must use juxtaposition:

$$a(bV \cdot cW) \qquad\qquad x\mathbf{v} \times y\mathbf{u}.$$

DERIVED SYMBOLS. Closely related objects require closely related notation. Proximity in the alphabet, e.g., x, y, z may be used for this purpose. For a stronger bond, the meaning of a symbol may be modified using subscripts, superscripts and other decorations:

$$A^* \qquad \overline{\eta} \qquad n^+ \qquad \underline{h} \qquad \tilde{e} \qquad \Omega_- \qquad Z_r.$$

The derived symbol $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ is often found in the literature. Many symbols derived from $\mathbb{R}$ are in use:

$$\begin{aligned} &\mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\} &\qquad &\mathbb{R}_{\geqslant 0} = \{x \in \mathbb{R} : x \geqslant 0\} \\ &\mathbb{R}^* = \mathbb{R} \smallsetminus \{0\} & &\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}. \end{aligned}$$

These sentences illustrate the use of derived symbols:

> *Let $f : X \to X$ be a function, and let $x^* = f(x^*)$ be a fixed point of f.*
>
> *We consider the endpoints x_- and x_+ of an interval containing x.*
>
> *Let f be a real function, and let*
>
> $$f^+ : \mathbb{R} \to \mathbb{R} \qquad x \mapsto \begin{cases} f(x) & \textit{if } f(x) \geqslant 0 \\ 0 & \textit{if } f(x) < 0. \end{cases}$$

It must be noted that there is no general agreement on the meaning of decorations. Thus for a set the overbar denotes the so-called **closure** —adjoining to a set all its limit points (the transformation from $\mathbb{R}$ to $\overline{\mathbb{R}}$ is essentially a closure operation). However, for complex numbers the overbar denotes complex conjugation. If f is a function, then f' is the derivative of f, but for sets a prime indicates taking the complement.

EXAMPLE. We illustrate notational problems raised by the coexistence of variables and parameters. For a fixed value of z, the bivariate polynomial

$$f(x,z) = -z^2 + xz + 1$$

becomes a polynomial in x, and we wish to adopt a notation that reflects the different role played by the symbols x and z. We replace z with a, to keep it far apart from x in the alphabet, and we rewrite the expression above as

$$f_a(x) = ax + 1 - a^2 \qquad a \in \mathbb{R}. \tag{5.4}$$

We now have a **one-parameter family** of linear polynomials in x. For fixed a, the equation $y = f_a(x)$ is the cartesian equation of a line, so we also have a one-parameter family of lines on the plane. Plotting some of these lines reveals a hidden structure: they form the **envelope** of a parabola (figure 5.2). Likewise, if we fix $x = a$ we obtain a one-parameter family of quadratic polynomials $g_a(z) = -z^2 + az + 1$.

5.3 The sigma-notation

The notation for sums

$$\sum_{k=1}^{n} a_k = a_1 + a_2 + \cdots + a_n \qquad\qquad \sum_{k=1}^{\infty} a_k = a_1 + a_2 + \cdots, \tag{5.5}$$

was introduced by Fourier[6]. It is called the (delimited) **sigma-notation**, as it makes use of the capital Greek letter with that name. The index of summation is a **dummy variable**. This a variable used for internal book-keeping (like an integration variable), and its identity does not affect the value of the sum:

$$\sum_{k=1}^{n} k^2 = \sum_{j=1}^{n} j^2 = 1^2 + 2^2 + \cdots + n^2.$$

---

[6] Jean Baptiste Joseph Fourier (French: 1768–1830).

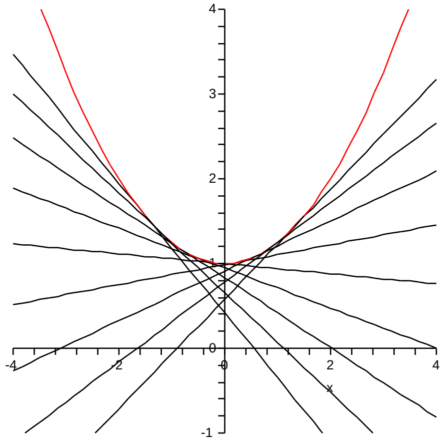

Figure 5.1: The one-parameter family of lines $y = f_a(x)$, where f_a is given by (5.4). These lines are tangent to the parabola $y = x^2/4 + 1$.

The summation index is invariably one of the six lower-case roman letters listed in (5.1). I suggest that you adopt a summation symbol and stick to it, unless there is a good reason to change it.

A **double sum** is defined as follows:

$$\begin{aligned}\sum_{j=1}^{J}\sum_{k=1}^{K} a_{j,k} &\overset{\text{def}}{=} \sum_{j=1}^{J}\left(\sum_{k=1}^{K} a_{j,k}\right)\\ &= \sum_{k=1}^{K} a_{1,k} + \sum_{k=1}^{K} a_{2,k} + \cdots + \sum_{k=1}^{K} a_{J,k}.\end{aligned}$$

The sum in parentheses is a function of the outer summation index j; this sum is performed repeatedly, each time with a different value of j. The use of matching symbols for the index and the upper limit of summation is particularly appropriate here. If the two ranges of summations are independent, inner and outer sums can be swapped. The commutative and associative laws of addition ensure that the value of the sum will not change. We illustrate this process with an example.

$$\sum_{j=0}^{1}\sum_{k=1}^{3} a_{j,k} = (a_{0,1} + a_{0,2} + a_{0,3}) + (a_{1,1} + a_{1,2} + a_{1,3})$$

$$
\begin{aligned}
&= (a_{0,1}+a_{1,1})+(a_{0,2}+a_{1,2})+(a_{0,3}+a_{1,3}) \\
&= \sum_{k=1}^{3}\sum_{j=0}^{1} a_{j,k}.
\end{aligned}
$$

This chain of equalities adopts a standard layout and alignment. If the indices in a double sum have the same range, then they may be grouped together. The following expressions represent the same sum:

$$\sum_{i=1}^{N}\sum_{j=1}^{N} a_{i,j} \qquad\qquad \sum_{i,j=1}^{N} a_{i,j}. \tag{5.6}$$

There are variants to the delimited sigma-notation (5.5). In an unrestricted sum, range information may be omitted altogether:

$$\sum_{k}\binom{n}{k} = 2^n \quad n \geqslant 0.$$

(This sum has in fact only finitely many non-zero terms.) The summation range may also be specified by inequalities placed below the summation symbols:

$$\sum_{1\leqslant k\leqslant n} a_k \qquad\qquad \sum_{k\geqslant 1} a_k \qquad\qquad \sum_{1\leqslant j,k\leqslant N} a_{j,k}. \tag{5.7}$$

The benefits of this notation become evident if we need to change summation index. Consider the following manipulation:

$$\sum_{-2\leqslant k\leqslant n-3} 2^{k+2} = \sum_{0\leqslant k+2\leqslant n-1} 2^{k+2} = \sum_{0\leqslant k\leqslant n-1} 2^{k} = 2^n - 1. \tag{5.8}$$

After adding 2 to each term in the inequalities, we have simply replaced $k+2$ with k, obtaining the sum of a geometric progression. The latter is then evaluated explicitly. With this notation, the change of summation index is unproblematic.

The greatest generality is achieved by the **standard form** of the sigma-notation

$$\sum_{\mathscr{P}(k)} a_k \tag{5.9}$$

where $\mathscr{P}$ is any predicate over $\mathbb{Z}$. The range of summation consists of those values of k for which $\mathscr{P}(k)$ is TRUE. Expressions (5.7) are of this type and the unrestricted summation becomes

$$\sum_{k\in\mathbb{Z}} a_k.$$

The advantage of this notation is that the summation index is no longer restricted to a sequence of consecutive integers, and the range of summation may be altered by adding conditions:

$$\begin{aligned}
\sum_{0<|k|\leqslant 2} a_k &= a_{-2}+a_{-1}+a_1+a_2\\
\sum_{\substack{1\leqslant k\leqslant 12\\ \gcd(k,12)=1}} a_k &= a_1+a_5+a_7+a_{11}\\
\sum_{\substack{1\leqslant k\leqslant 10\\ k \text{ prime}}} a_k &= a_2+a_3+a_5+a_7.
\end{aligned}$$

Example. For any natural number n, let $\varphi(n)$ be the number of positive integers smaller than n and relatively prime to it (with $\varphi(1)=1$). Thus $\varphi(12)=4$. This is **Euler's φ-function** of number theory, which is defined in symbols as follows:

$$\varphi(1)=1 \qquad \varphi(n)=\sum_{\substack{1\leqslant k<n\\ \gcd(k,n)=1}} 1, \qquad n>1.$$

Simply by letting $a_k=1$ in (5.9), we have turned a summation into an algorithm that counts the elements of the set of which $\mathscr{P}$ is the characteristic function. This neat device illustrates the flexibility of the sigma-notation. Alternatively, Euler's function may be defined as the cardinality of a set, using the cardinality symbol '#' (see section 2.1)

$$\varphi(1)=1 \qquad \varphi(n)=\#\{k\in\mathbb{N}\,:\,k<n,\ \gcd(k,n)=1\} \qquad n>1.$$

The constructs introduced above for sums are also applicable to products as well as to combinations of sums and products.

5.4 Improving formulae

When the physicist Stephen Hawking was writing his book *A Brief History of Time*, an editor warned him that every new equation in the book would halve the readership. So he included a single equation. When writing for the general public, one should expect diffidence —even hostility— towards formulae. Even though mathematicians are trained to deal with formulae, it is still safe to assume that no one likes to struggle with too many symbols. In this section we explore techniques to improve the clarity of formulae, through presentation, notation, and layout.

Formulae may be embedded in the text or displayed, and in a document one usually finds both arrangements. In either case a formula must obey standard punctuation rules. The following passage features embedded formulae with appropriate punctuation.

> *For each $x \in X$ we have the decomposition $x = \xi + \lambda$, with $\xi \in \Xi$ and $\lambda \in \Lambda$; accordingly, we define the function $P : X \to \Xi$, $x \mapsto \xi$, which extracts the first component of x.*

The punctuation generates rests as it would in an ordinary English sentence. The two items comprising the function definition are separated by a comma, which would not be necessary in a displayed formula, see (2.14) p. 25. The formula defining P is echoed by a short sentence, for added clarity.

This is a fully punctuated displayed formula:

$$a_1 = 1; \quad a_{k+1} = \begin{cases} a_k^2 - 1, & 1 \leqslant k < 10; \\ a_k^2, & k \geqslant 10. \end{cases}$$

If full punctuation seems heavy (it may do so here), then we may replace some punctuation by increased spacing or by words:

$$a_1 = 1 \qquad a_{k+1} = \begin{cases} a_k^2 - 1 & \text{if } 1 \leqslant k < 10 \\ a_k^2 & \text{if } k \geqslant 10. \end{cases}$$

A displayed formula is normally embedded within a sentence, and the punctuation at the end of a formula must be appropriate to the structure of the sentence. In particular, if a sentence terminates at a formula —as in the example above— then the full stop at the end of the formula must always be present.

EXAMPLE. The following untidy formula

$$f(x) = \frac{14x - 2x^3 - 2x^2 + 14}{-2x - 4}$$

could represent a typical unprocessed output of a computer algebra system. It contains redundant information (a common factor between numerator and denominator), the monomials at numerator are not ordered, and there are too many negative signs. The properties of the rational function $f(x)$ are not evident from it. There are several ways to improve the formula's layout:

$$\begin{aligned} f(x) &= \frac{x^3+x^2-7x-7}{x+2} \\ f(x) &= \frac{(x+1)(x^2-7)}{x+2} \\ f(x) &= x^2-x-5+\frac{3}{x+2}. \end{aligned}$$

If the degree or the coefficients of f are important, then the first version is appropriate; the second version makes it easy to solve the equation $f(x)=0$; the third version is a preparation for integrating f.

EXAMPLE. The following subset of the rationals

$$A = \left\{ y \in \mathbb{Q} : y = \frac{x}{x^2+1},\ \ x \in \mathbb{Z},\ \ x < 0 \right\}$$

is defined with a disproportionate volume of notation. To save symbols we switch from the Zermelo to the standard set definition, and consider only elements of the required form. We also remove the inequality, and replace it with a negative sign. Finally, we use a more appropriate symbol for the natural numbers:

$$A = \left\{ \frac{-n}{n^2+1} : n \in \mathbb{N} \right\}.$$

Now the formula is transparent.

EXAMPLE. To simplify a complex formula, we apply the 'divide and conquer' method. In the cluttered expression

$$R = x(ad-bc) - y(ad-cb)^2 + z(ad-cb)^3$$

the sub-expression $ad-bc$ appears as a unit. We exploit this fact to improve the layout:

$$R = x\kappa - y\kappa^2 + z\kappa^3 \qquad\qquad \kappa = ad-bc.$$

The formula is tidier, and the structure of R is clearer.

EXAMPLE. Our next challenge is to improve an intricate double sum:

$$z(y_1,y_2,\ldots)=\sum_{i=1}^{\infty}\sum_{y=0}^{y_i-1}(y+1)x^{i-1}. \tag{5.10}$$

The meaning of z is not at all evident. This quantity depends on x, but its dependence does not appear explicitly; the poor choice of symbols obscures matters further. We note that the parameters y_i are integers, since they are the upper limit of the inner summation; accordingly, we replace the symbol y with n —see (5.1). Then we adopt the 'divide and conquer' method, splitting up the sum as follows:

$$z(x,n_1,n_2,\ldots)=\sum_{i=1}^{\infty}d_ix^{i-1} \qquad d_i=\sum_{k=0}^{n_i-1}(k+1).$$

We now see that z is a power series in x; its coefficients are finite sums, determined by the elements of an integer sequence. These are sums of arithmetic progressions, which can be evaluated explicitly.

$$\sum_{0\leqslant k\leqslant n-1}(k+1)=\sum_{1\leqslant k\leqslant n}k=\frac{n(n+1)}{2}.$$

(In this passage we have dropped the subscript i, since the association $n\leftrightarrow n_i$ is clear, and we have switched to the standard sigma-notation to change variable — cf. equation (5.8).) Our original double sum (5.10) can now be written as

$$z(x,\mathbf{n})\;=\;\frac{1}{2}\sum_{i=1}^{\infty}n_i(n_i+1)x^{i-1} \qquad \mathbf{n}=(n_1,n_2,\ldots)$$

where we have used again the 'divide and conquer' method. The dependence of z on the variable x and the sequence $\mathbf{n}$ is now clear.

5.5 Writing definitions

A definition requires a pause, to give the reader time to absorb it. This may be achieved by giving the definition twice, first with words then with symbols (or vice-versa), by using two different formulations, or by supporting the definition with an example. Let us consider some examples of definitions.

EXAMPLE.

> *Let $\mathscr{P}$ be a predicate over the rational numbers, that is, a function of the type*
> $$\mathscr{P} : \mathbb{Q} \to \{T, F\}.$$

The definition of the symbol $\mathscr{P}$ uses some jargon (predicate over a set) so the second part of the sentence recalls its meaning. A previous exposure to this concept is tacitly assumed; for a first encounter we would need a more considerate style, such as in the opening paragraph of section 3.3, p. 57.

EXAMPLE.

> *For given $\varepsilon > 0$, we define the set*
> $$S_\varepsilon = \{(x,y) \in \mathbb{R}^2 \,:\, |x-y| \leqslant \varepsilon/\sqrt{2}\},$$
> *that is, S_ε is a strip of width ε symmetrical with respect to the main diagonal in the cartesian plane.*

This time the symbolic (Zermelo) definition appears first, while a verbal explanation clarifies the geometric meaning of S_ε, which is not obvious from the formula. The quantity ε appears as a subscript, indicating that it is a parameter. Zermelo definitions should be used sparingly, and should not be used at all if we write for a non-mathematical audience. (Lars Ahlfors in his beautifully written text *Complex analysis* deliberately avoids them [2].)

EXAMPLE.

> *For every real number λ, let $\Pi(\lambda)$ be the plane in three-dimensional euclidean space orthogonal to the vector $v(\lambda) = (1, \lambda, \lambda^2)$. Thus $\Pi(\lambda)$ consists of the points $z = (x_1, x_2, x_3) \in \mathbb{R}^3$ for which the scalar product $z \cdot v = x_1 + x_2\lambda + x_3\lambda^2$ is equal to zero.*

In this definition the quantity λ appears as an argument of both Π and v, and thus the latter represent functions. The second sentence restates the definition in a form suitable to computation, recalling the connection between orthogonality and scalar product, and introducing further notation.

EXAMPLE.

> *Let $\mathscr{N}$ be the set of sequences of natural numbers, such that every natural number is listed infinitely often. For example, the sequence*
>
> $$(1,1,2,1,2,3,1,2,3,4,\ldots)$$
>
> *belongs to $\mathscr{N}$.*

A non-experienced reader will have no idea that such a construction is at all possible, so we give an example.

The definition of a symbol should appear as near as possible to where the symbol is first used; defining a symbol immediately after its first appearance is also acceptable, provided that the definition is given *within the same sentence.*

EXAMPLE. We give the same definition three times, articulating the changes in emphasis that accompany each version.

> *Consider the power series*
>
> $$h(x) = \sum_{n=1}^{\infty} a_n x^n,$$
>
> *where the coefficient a_n is the square of the n-th triangular number.*

The definition of a_n immediately follows its first appearance. The displayed formula represents a general power series, and its specific nature is revealed only by reading the entire sentence. For this reason, this format may not be ideal if the formula is to be referred to from elsewhere in the text. This definition puts some burden on readers who are unfamiliar with the term **triangular number**.

In our second version we change both text and formula.

> *Let t_n be the n-th triangular number. We consider the power series*
>
> $$h(x) = \sum_{n=1}^{\infty} t_n^2 x^n.$$

Now t_n is defined before being used, with the symbol t chosen so as to remind us of ‘triangular’. The formula is clearer. Just glancing at it makes us want to find out what t_n is, and to do this one would begin to scan the text preceding, rather than following, the formula.

Our third version combines verbal and symbolic definitions.

> *We consider the power series $h(x)$ whose coefficients are the square of the triangular numbers, namely,*
>
> $$h(x) = \sum_{n=1}^{\infty} t_n^2 x^n \qquad t_n = \frac{n(n+1)}{2}.$$

The formula is more complex because everything is defined within it. But it is also self-contained, so it'll be easy to refer to.

5.6 Introducing a concept

Opening the exposition with a formal definition is rarely a good way to introduce a new concept, particularly if we write for non-experts. The following examples show how to set the scene for a gentle introduction to a new idea. In each case we ask a question —a simple rhetorical device to engage and prepare an audience.

EXAMPLE. Introducing recursive sequences.

> *Let n be a natural number. How is the power 2^n defined? We could use repeated multiplication*
>
> $$2^n := \underbrace{2 \times 2 \times \cdots \times 2}_{n},$$
>
> *but we could also write*
>
> $$2^1 := 2 \qquad \textit{and} \qquad 2^n := 2 \times 2^{n-1} \quad n > 1.$$
>
> *The second formula is an example of a* **recursive definition** *of a sequence (a_n). When $n = 1$, the first term $a_1 = 2$ is defined explicitly; then, assuming that a_{n-1} has been defined, we define a_n in terms of it.*
>
> [*The recursive definition of a general sequence follows (see section 8.4)*].

The opening question leads to the familiar definition of exponentiation in terms of multiplication. The less familiar recursive definition comes after, illustrated by a simple example and supported by an appropriate notation. After this preamble, the reader should be ready to confront an abstract definition.

EXAMPLE. Introducing the exponential function.

> *The process of differentiation turns a real function into another real function. For example, differentiation turns the sine into the cosine. Are there functions that are not changed at all by differentiation?*
>
> [*The definition of the exponential function follows.*]

We recall a structural property of differentiation and present a familiar example. The question that follows suggests how the argument will develop.

EXAMPLE. Introducing the rational numbers.

> *A rational number is represented by a pair of integers, the numerator and the denominator. As these integers need not be co-prime, we may choose them in infinitely many ways. How are we to construct a single rational number from an infinite set of pairs of integers? How do we define the set* $\mathbb{Q}$ *from the cartesian product* $\mathbb{Z}\times(\mathbb{Z}\smallsetminus\{0\})$*?*
>
> [*The definition of the set of rational numbers follows.*]

The first two sentences recall elementary facts. A question then invites the reader to think about this problem more carefully. A second question, which echoes the first, uses proper terminology and notation, in preparation for a formal construction.

EXAMPLE. We design an exercise structured as a list of questions. The topic is the number of relations on a finite set.

> *What is a relation on a set?*
>
> *Can a relation be defined on the empty set?*
>
> *How many relations can one define on a two-element set?*
>
> *Let* n *be the number of relations on a set. What values can* n *assume?*

The first question checks background knowledge; the other questions gently explore the problem, guiding the reader from the specific to the general. Textbook exercises are sometimes structured in this way, to encourage independence in learning.

Formulating questions is not just a method to grab attention or structure exercises. Asking the right questions —those that chart the boundaries of our knowledge— is the essence of research.

5.7 Writing a short description

Writing a synopsis of a mathematical topic is a common task. It could be the abstract of a presentation, the summary of a chapter of a book, an informal explanation of a theorem. It could be the closing paragraph of a large document, which distills the essence of an entire subject.

Writing a short essay is difficult —the shorter the essay, the greater the difficulty— so this task is appropriate for the final stages of learning. We shall adopt the format of a MICRO-ESSAY: 100-150 words (one or two paragraphs) and no mathematical symbols. Within such a confined space one is forced to make difficult decisions on what to say and what to leave out; the lack of symbols gives further prominence to the concepts. Our command of the syntax will be put to the test.

Our first MICRO-ESSAY is a summary of section 2.1 on sets **[**$\notin$, 150**]**. We have access to all relevant material, and the main difficulty is to decide what are the highlights of that section. We select two ideas: how to define a set, and how to construct new sets from old ones.

> *A set is a collection —finite or infinite— of distinct mathematical objects, whose ordering is unimportant. A small set may be defined by listing all its elements explicitly; a large set is instead defined by specifying the characteristic properties of the elements.*
>
> *Combining numbers with arithmetical operators gives new numbers. Likewise, combining sets via set operators (union, intersection, difference) gives new sets. But this is a bit like recycling what we already have. A class of brand-new sets are the so-called* cartesian products*, which are constructed by pairing together elements of existing sets. A well-known example is the cartesian plane, which is made of pairs of elements of the real line, each representing a coordinate.*

The description of set operators exploits an analogy with arithmetical operators, and for this reason the first two sentences of the second paragraph have the same structure. We avoid a precise definition of the term ‘cartesian product’ (not enough space!), opting instead for an informal description and an illustrative example.

Next we write a MICRO-ESSAY on prime numbers, a synthesis of our knowledge of this topic. **[**$\notin$, 100**]**

> *A prime is a positive integer divisible only by itself and unity. (However, 1 is not considered prime.) The importance of primes in arithmetic*

> *stems from the fact that every integer admits a unique decomposition into primes. The infinitude of primes —known from antiquity— and their unpredictability make them object of great mathematical interest.*
>
> *Many properties of an integer follow at once from its prime factorization. For instance, looking at the exponents alone, one can determine the number of divisors, or decide if an integer is a power (i.e.,*[7] *a square or a cube).*
>
> *Primality testing and prime decomposition are computationally challenging problems with applications in digital data processing.*

This essay begins with a definition. The technical point concerning the primality of 1 has been confined within parentheses, to avoid cluttering the opening sentence. The following two sentences deliver core information in a casual —yet precise— way. We state two important theorems (the Fundamental Theorem of Arithmetic, and Euclid's theorem on the infinitude of the primes) without mentioning the word 'theorem'. We also give a hint of why mathematicians are so fascinated by primes. The second paragraph elaborates on the importance of unique prime factorisation, by mentioning two applications without unnecessary details. The short closing paragraph, like the last sentence of the first paragraph, is an advertisement of the subject, meant to encourage the reader to learn more.

Now a real challenge: write a MICRO-ESSAY on theorem 3.3, p. 59, which we reproduce here for convenience.

Theorem. *Let* X *be a set, let* $A, B \subseteq X$, *and let* $\mathscr{P}_A$, $\mathscr{P}_B$ *be the corresponding characteristic functions. The following holds (the prime denotes taking complement)*:

$$\begin{array}{crcl}
(i) & \neg\,\mathscr{P}_A & = & \mathscr{P}_{A'} \\
(ii) & \mathscr{P}_A \wedge \mathscr{P}_B & = & \mathscr{P}_{A\cap B} \\
(iii) & \mathscr{P}_A \vee \mathscr{P}_B & = & \mathscr{P}_{A\cup B} \\
(iv) & \mathscr{P}_A \Rightarrow \mathscr{P}_B & = & \mathscr{P}_{(A\smallsetminus B)'} \\
(v) & \mathscr{P}_A \Leftrightarrow \mathscr{P}_B & = & \mathscr{P}_{(A\cap B)\cup(A\cup B)'}.
\end{array}$$

We must extract a theme from a daunting list of inscrutable symbols. We inspect the material of section 3.3 leading to the statement of this theorem: it deals with characteristic functions, and the main idea is to link characteristic functions to sets. We can see such a link in every formula: on the left there are logical operators, on the right set operators. Given the specialised nature of this theorem, some jargon is unavoidable, so we write for a mathematically mature audience.

---

[7] Abbreviation for the Latin *id est*, meaning 'that is'.

> *The action of a logical operator on characteristic functions produces a new characteristic function. Now, every characteristic function corresponds to a set, and so the new function corresponds to a new set. Since we could have obtained the new set directly from the old sets by means of set operators, we now have two different ways to do the same thing. This theorem establishes a correspondence between these two classes of operators. Under this correspondence, the logical operator of negation is represented by the set operation of taking the complement. More precisely, the negation of the characteristic function of a set is the characteristic function of the complement of this set. Analogous results are established with respect to the main logical operators.*

Here words are better than symbols to describe the structure, but they can't compete with symbols for the details. So, to avoid tedious repetitions, we have chosen to explain just one formula carefully (the easiest one!), mentioning the other formulae under the generic heading 'analogous results'.

Exercises

Exercise 5.1. Consider the following question:

> *Why is it that when the price of petrol goes up by 10% and then comes down 10%, it doesn't finish up where it started?*

1. Write an explanation for the general public. Do not use mathematical symbols, as most people find them difficult to understand. **[**$\not\varepsilon$**]**

2. Write an explanation for mathematicians, combining words and symbols for maximum clarity. You should deal with the more general problem of two opposite percentage variations of an arbitrary non-negative quantity, not specialised to petrol, or to 10% variation, or to the fact that the decrease followed the increase and not the other way around.

Exercise 5.2. Consider the following question:

> *I drive ten miles at 30 miles an hour, and then another ten miles at 50 miles an hour. It seems to me my average speed over the journey should be 40 miles an hour, but it doesn't work out that way. Why not?*

Write an explanation for the general public, clarifying why such a confusion may arise. You may perform some basic arithmetic, but do not use symbols as most people find them difficult to understand. **[**$\not\varepsilon$**]**

Exercise 5.3. Consider the following question:

> *I tossed a coin four times, and got heads four times. It seems to me that if I toss it again I am much more likely to get tails than heads, but it doesn't work out that way. Why not?*

Write an explanation for the general public, clarifying why such a confusion may arise. You may use symbols such as H and T for heads-tails outcomes, but avoid using other symbols. **[**$\not\varepsilon$**]**

Exercise 5.4. Consider the following question:

> *In a game of chance there are three boxes: two are empty, one contains money. I am asked to choose a box by placing my hand over it; if the money is in that box, I win it. Once I have made my choice, the presenter —who knows where the money is— opens an empty box and then gives me the option to reconsider. I can change box if I wish. It seems to me that changing box would make no difference to my chances of winning, but it does not work out that way. Why not?*

Write for the general public, explaining what is the best winning strategy. **[**$\not\varepsilon$**]**

Exercise 5.5. Formulate a list of questions to help a student approach the following problem:

> *Prove that for all integers* m,n*, we have* $m\mathbb{Z}+n\mathbb{Z}=\gcd(m,n)\mathbb{Z}$.

The assignment should consist of four of five questions of increasing difficulty, the last of which deals with the statement above.

Exercise 5.6. Write a MICRO-ESSAY on each topic. **[**$\not\varepsilon$, 150**]**

1. *Quadratic equations and complex numbers.*
2. *Definite vs. indefinite integration.*
3. *What are matrices useful for?*
4. *Images and inverse images of sets.*
5. *Rational numbers vs. fractions.*
6. *How to get the integers from pairs of natural numbers.*

Exercise 5.7. Write a two-page essay on each topic.

1. *The many ways of defining the logarithm.*
2. *All fractions in a very small interval have very large denominators (with at most one exception, that is).*
3. *In probability, a random variable is not random and is not a variable: it's a function!*
4. *Sequences racing to infinity: who gets there first?*

Chapter 6

Forms of argument

A mathematical theory begins with a collection of **axioms** or **postulates**. These are statements that are assumed to be true —no verification or justification is required. The axioms are the building blocks of a theory. From them we deduce other true statements called **theorems**, and a **proof** is the process of deduction that establishes a theorem. As more and more theorems are proved, the theory is enriched by a growing list of true statements.

When we develop a mathematical argument, we need to organise sentences so that every sentence is either an axiom or a true statement derived from axioms or earlier statements. In practice only significant statements will be called theorems. In the proof of a theorem many true statements may be derived, but these fragments are not usually assigned any formal label. Sometimes a proof rests on statements which, if substantial, are not of independent significance: they are called **lemmas**. Finally, a **proposition** is a statement which deserves attention, but which is not sufficiently general or significant to be called a theorem.

In this chapter we survey some methods to give shape to a mathematical argument. Our survey will continue in chapter 7, devoted to induction techniques.

6.1 Anatomy of a proof

A first analysis course begins with a list of axioms defining the real number system. Then one may be asked to prove that

$$\forall x \in \mathbb{R},\ \ -(-x) = x. \tag{6.1}$$

But isn't this statement obvious? In recognising the truth of this assertion we make implicit use of knowledge derived from experience, which is not part of the axioms. A proof from the axioms requires erasing all previous knowledge.

We now introduce some axioms, state a theorem equivalent to (6.1), and then prove it from the axioms. We put the proof under X-ray. Our purpose is to dissect a mathematical argument, articulating every step. We use the language mechanically, to facilitate the identification of the logical elements of the proof. To avoid making implicit assumptions, we represent familiar objects with an unfamiliar notation.

We are given a set Ω *and a binary operator '*$\odot$*' on* Ω*, with the following properties:*

G1: $\forall x, y, \in \Omega,\ \ x \odot y \in \Omega$

G2: $\forall x, y, z \in \Omega,\ \ (x \odot y) \odot z = x \odot (y \odot z)$

G3: $\exists \diamondsuit \in \Omega,\ \ \forall x \in \Omega,\ \ x \odot \diamondsuit = \diamondsuit \odot x = x$

G4: $\forall x \in \Omega,\ \ \exists x' \in \Omega,\ \ x \odot x' = \diamondsuit$

(These axioms define a general object: a **group**.) Setting $\Omega = \mathbb{R}$, and '$\odot$' = '$+$', one recognises that $\diamondsuit$ represents 0 and x' represents $-x$. We are ready to state and prove our theorem.

1. **Theorem.**
2. $\forall x \in \Omega,\ \ x'' = x$.
3. Proof.
4. Let $x \in \Omega$ be given;
5. then (by G4) $x'' \in \Omega$;
6. let $a := (x \odot x') \odot x'',\ \ b := x \odot (x' \odot x'')$;
7. hence (by G1 and 5) $a, b, \in \Omega$, and (by G2) $a = b$;
8. then (by G3 and G4) $a = \diamondsuit \odot x'' = x''$;
9. then (by G3 and G4) $b = x \odot \diamondsuit = x$;
10. hence (by 7, 8, 9) $x'' = x$.

11. □

Items 1,3,11 are **logical tags** —expressions or symbols that say things about the text. They announce the statement of the theorem, and the beginning and the end of the proof, respectively. The symbol '□' in item 11 may be replaced by the acronym Q.E.D.[1].

Items $5,8,9$ begin with the logical tag 'then', used here to mean that what comes after is deduced from the axioms. Formally, item 5 is our first **theorem**, being a true statement deduced from the axioms. Items 7 and 10 are deductions of a slightly different nature, and we have flagged them with the tag 'hence'. These items use not only axioms but also facts assembled from previous statements.

Items 4 and 6 are not statements but **instructions**. Item 4 instructs us to assume that x is an arbitrary element of a set. This standard opening sentence mirrors the expression $\forall x \in \Omega$ in the theorem. Item 6 instructs us to define two new quantitites. Many different types of instruction may be found in proofs, for instance instructions to draw a picture.

In item 10 we use implicitly the **transitivity** of the equality operator (if $x = y$ and $y = z$, then $x = z$), which depends on the fact that equality is an equivalence relation (see section 3.6).

Our next proof establishes the irrationality of $\sqrt{2}$. We use the same style and conventions as in the previous theorem. However, this proof is quite removed from the axioms of arithmetic, in that it uses some definitions and facts without justification. In this respect this proof is more typical than the previous one.

1. **Theorem**.
2. $\sqrt{2}$ is irrational.
3. PROOF.
4. Suppose $\sqrt{2}$ is rational;
5. then $\exists m, n \in \mathbb{Z}$ such that $\sqrt{2} = m/n$.
6. Choose m, n in 5 so that they are co-prime.
7. Since $\sqrt{2} = m/n$,
8. then $2 = m^2/n^2$;

---

[1] Abbreviation for the Latin *quod erat demonstrandum*, meaning 'which had to be demonstrated'.

9. then $2n^2 = m^2$;
10. then m^2 is even;
11. hence (by 10 and theorem A) m is even;
12. then $m = 2h$ for some integer h;
13. hence (by 9 and 12) $2n^2 = 4h^2$;
14. then $n^2 = 2h^2$;
15. then n^2 is even;
16. hence (by 15 and theorem A) n is even;
17. hence (by 11 and 16) m and n are not co-prime;
18. hence (by 6 and 17) assumption 4 is false;
19. then its negation is true.
20. □

A fair amount of mathematics is assumed in the proof:

the definition of the rational numbers and of $\sqrt{2}$ (items 2,4,5);

the definition of co-primality (item 6);

some properties of equations (items 8,9);

the definition of an even integer (items 12, 17);

a theorem of arithmetic, referred to as theorem A[2] (items 11,16).

The argument contains novel elements. The core of the proof is item 4, the **assumption** that $\sqrt{2}$ is rational. This is puzzling: how can we assume that $\sqrt{2}$ is rational when we know it isn't? How can we let $\sqrt{2} = m/n$ when we haven't been told what numbers m and n are? This is not so strange; assumptions belong very much to common reasoning.

---

[2] If a prime divides the product of two integers, then it divides one of the factors.

> *Suppose that I meet Einstein on top of Mount Everest.*
>
> *Suppose that people walk laterally.*

We are clearly free to explore the logical consequences of these assumptions. In mathematics, assumptions are handled formally via the implication operator $\Rightarrow$, developed in chapter 3.

The assumption that $\sqrt{2}$ is rational eventually leads to a **contradiction**: items 6 and 17 are conflicting statements. From this fact we deduce that the assumption in item 4 is false (item 18). Here we abandon the assumption 4. But if the statement $\sqrt{2} \in \mathbb{Q}$ is false, then its negation is true. This is item 19, which is what we wanted to prove.

We now abandon the robotic, over-detailed proof style adopted in this section, and analyse various methods of proof in more typical settings.

6.2 Proof by cases

Some mathematical arguments are made tidier by breaking them into a number of cases, of which precisely one holds, while all lead to the desired conclusion.

EXAMPLE. Consider the following statement:

> *The solution set of the inequality* $2|x| \geqslant |x-1|$ *is the complement of the open interval* $(-1, 1/3)$.

PROOF. Let x be a real number. There are three cases.
Case I: If $x < 0$, then the inequality is $-2x \geqslant 1-x$, giving $x \leqslant -1$.
Case II: If $0 \leqslant x < 1$, then the inequality is $2x \geqslant 1-x$, giving $1/3 \leqslant x < 1$.
Case III: If $x \geqslant 1$, then the inequality is $2x \geqslant x-1$, which is always verified, giving $x \geqslant 1$.
So the required solution set is the set of x such that $x \leqslant -1$, or $1/3 \leqslant x < 1$, or $x \geqslant 1$, which is the union of two rays

$$(-\infty, -1] \cup [1/3, 1) \cup [1, \infty) = (-\infty, -1] \cup [1/3, \infty).$$

This set is obtained from the real line by removing the open interval $(-1, 1/3)$, as desired. $\square$

The opening sentence 'Let x be a real number' acknowledges that all real values of x will have to be considered. The second sentence announces that the proof will branch into three cases, determined by the sign of the expressions within absolute value. Note the careful distinction between strict and non-strict inequalities, to avoid missed or repeated values of x.

The structure of this proof is shaped by the structure of the absolute value function, which is piecewise defined —see equation (4.13) and section 4.6. A proof by cases should be expected with any statement involving piecewise-defined functions.

EXAMPLE. A proof by cases on divisibility.

Let n be an integer; then $n^5 - n$ is divisible by 30.

PROOF. We factor the integer 30 and the polynomial $n^5 - n$:

$$30 = 2 \times 3 \times 5 \qquad n^5 - n = n(n-1)(n+1)(n^2+1).$$

For each prime $p = 2, 3, 5$ we will show that, for any integer n, at least one factor of $n^5 - n$ is divisible by p.

1. $p = 2$. Let $n = 2k + j$, for some $k \in \mathbb{Z}$ and $j = 0, 1$.
   If $j = 0$, then n is divisible by 2.
   If $j = 1$, then $n - 1$ is divisible by 2.

2. $p = 3$. Let $n = 3k + j$, for some $k \in \mathbb{Z}$ and $j = 0, \pm 1$.
   If $j = 0$, then n is divisible by 3;
   if $j = \pm 1$, then $n \mp 1$ is divisible by 3.

3. $p = 5$. Let $n = 5k + j$, for some $k \in \mathbb{Z}$ and $j = 0, \pm 1, \pm 2$.
   If $j = 0$, then n is divisible by 5.
   If $j = \pm 1$, then $n \mp 1$ is divisible by 5.
   If $j = \pm 2$, then $n^2 + 1 = 25k^2 \mp 20k + 5 = 5(5k^2 \mp 4k + 1)$ is divisible by 5.

The proof is complete. □

After the factorisations we declare our intentions. There are three cases, one for each prime divisor p of 30. Each prime in turn leads to p cases, corresponding to the possible values of the remainder j of division by p. To simplify book-keeping, we consider also negative remainders, and then pair together the remainders which differ by a sign. Note the presence of the symbols ± and ∓ within the same sentence. The positive sign in the first expression matches the negative sign in the second expression, and vice-versa.

6.3 Implications

Many statements take the form of implications:

> *If P then Q.*

(See section 3.2.) For example:

> *If p is an odd prime, then $2^{p-1}-1$ is divisible by p.*
> *If the sequence (a_k) is periodic, then (a_{2k}) is also periodic.*

Implications may appear in disguise, without the ‘if. . . then’ construct or the operator ‘$\Rightarrow$’:

> *A is a subset of B.*
> *Every repeating decimal is rational.*
> *The determinant of an invertible matrix is non-zero.*

When we rewrite them explicitly as implications, we note that it is necessary to introduce a new quantity:

> *If $x \in A$, then $x \in B$.*
> *If r is a repeating decimal, then r is rational.*
> *If A is an invertible matrix, then $\det(A) \neq 0$.*

This is due to the presence of a hidden universal quantifier. So in these examples the symbolic form of the implication is not $P \Rightarrow Q$, but rather

$$\forall x \in X, \quad \mathscr{P}(x) \Rightarrow \mathscr{Q}(x) \tag{6.2}$$

where $\mathscr{P}$ and $\mathscr{Q}$ are predicates over X. We expand one of our statements, to tell the full story:

> *For all real numbers r, if the decimal digits of r are repeating, then r is rational.*

Spelling out an implication in this way will help us organise the proof.

6.3.1 Direct proof

From the truth table (3.9) we see that if P is false, then $P \Rightarrow Q$ is true regardless of the value of Q, so there is nothing to prove. If P is true, then the implication is true provided that Q is true. So a **direct proof** of the implication consists in **assuming** P and **deducing** Q. The assumption of P lasts only until we have proved Q; after that P is discharged and may no longer be used.

Every direct proof of *'If P then Q'* must contain a section during which P is assumed. The beginning of the section is announced by a sentence such as

Assume P. *Suppose P.* *Let P.*

The task of the rest of the proof is to prove not the original theorem, but Q. We make this clear by writing

RTP: Q

where RTP stands for *Remains To Prove* (or *Required To Prove*). The block ends when the proof of Q is completed. This is announced with a closing sentence such as

We have proved Q. *The proof of Q is complete.*

The above considerations suggest what should be the opening sentence of a direct proof of an implication:

> **Theorem.** *If $p > 3$ is a prime and $p+2$ is also prime, then $p+4$ is composite.*
> PROOF. Suppose p is a prime number greater than 3, such that $p+2$ is prime. RTP: $p+4$ is composite.

All three assumptions ($p > 3$, p prime, $p+2$ prime) must be used in the proof. If not, either the proof is wrong, or the theorem contains redundant information.

Often you can work out how the proof of an implication must start, even if you haven't the faintest idea of what the mathematics is about. We illustrate this point with some real life examples:

> **Theorem.** *A closed subset of a compact set is compact.*
> PROOF. Let X be a compact set, and let C be a subset of X. Assume that C is closed. RTP: C is compact.

Theorem. *If $\lambda \in \mathbb{C}$ is a root of a monic polynomial whose coefficients are algebraic integers, then λ is an algebraic integer.*
PROOF. Let p be a monic polynomial whose coefficients are algebraic integers, and let $\lambda \in \mathbb{C}$ be a root of p. RTP: λ is an algebraic integer.

Theorem. *Every basis of a finite-dimensional linear space has the same number of elements.*
PROOF. Let V be a finite-dimensional linear space, and let B_1 and B_2 be two bases for V. Suppose B_1 has n_1 elements and B_2 has n_2 elements. RTP: $n_1 = n_2$.

Establishing a good notation is often decisive. In all examples above the proof begins by giving names to things. Some authors make this unnecessary by including names in the theorem; others obscure the statement of a theorem by putting too many names in it. Let us rewrite the last theorem in such a way as to establish some notation within the statement itself.

Theorem. *Let V be a finite-dimensional linear space. Then every basis for V has the same number of elements.*

Theorem 1. *Let V be a finite-dimensional linear space, and let B_1 and B_2 be two bases for V. Then $\#B_1 = \#B_2$.*

Let us compare the three formulations of this theorem. The first is plain and effective. The second contains a minimum of notation, which does no harm but is unnecessary. The last version establishes some useful notation. Normally the notation should be introduced within the proof, unless one plans to use it at a later stage. For instance, one could find the sentence

> *Let V, B_1, B_2 be as in theorem 1*

in the text following the theorem.

6.3.2 Proof by contrapositive

In section 3.2 we have seen that any implication $P \Rightarrow Q$ is equivalent to its **contrapositive** $\neg Q \Rightarrow \neg P$: they are both true or both false for any value of P and Q. This equivalence gives us a method for proving implications, called **proof by contrapositive**, which is a useful alternative to a direct proof. To prove by contrapositive that

If P then Q,

we prove instead that

If not Q then not P,

that is, we assume that Q is false and then deduce that P is false.

A proof by contrapositive is structurally identical to a direct proof, only the predicates are different. Thus in place of (6.2) we consider the statement

$$\forall x \in X,\ \neg\mathscr{Q}(x) \Rightarrow \neg\mathscr{P}(x). \tag{6.3}$$

In forming the contrapositive we haven't altered the quantifier; we have merely replaced a boolean function with an equivalent function, much like replacing $\sin^2(x)$ with $1 - \cos^2(x)$.

How do we decide between a direct proof and a proof by contrapositive? We must compare the assumptions P and $\neg Q$, and decide which of the two is easier to handle. Sometimes it's necessary to try both approaches to find out. The following examples show the reasoning behind such decisions.

Example. Consider the statement

$$\forall n \in \mathbb{N},\ (2^n < n!) \Rightarrow (n > 3) \tag{6.4}$$

The assumption $\mathscr{P}(n) = (2^n < n!)$ is problematic, because its value is not easily computable. By contrast, $\neg\mathscr{Q}(n) = (n \leqslant 3)$ is straightforward, and the contrapositive implication

$$\forall n \in \mathbb{N},\ n \leqslant 3 \Rightarrow\ (2^n \geqslant n!) \tag{6.5}$$

involves checking that the boolean expression $(2^n \geqslant n!)$ is true for only three values of n.

Proof. We only have to check three cases:

$$\begin{aligned} n = 1&: \quad 2 = 2^1 \geqslant 1! = 1 \\ n = 2&: \quad 4 = 2^2 \geqslant 2! = 2 \\ n = 3&: \quad 8 = 2^3 \geqslant 3! = 6. \end{aligned}$$

Thus the expression (6.5) is true, and the proof is complete. □

Example. Consider the statement

> *If the average of four distinct integers is equal to 10, then one of the integers is greater than 11.*

The direct implication involves an assumption on an average value, which entails loss of information; the contrapositive implication involves four integers of bounded size. We opt for the latter, which seems easier.

> *Given four distinct integers not greater than 11, their average is not equal to 10.*

PROOF. Let four distinct integers be given. If none of them exceeds 11, then the largest value their sum can assume is $11+10+9+8=38$. So the largest possible average is

$$\frac{38}{4}=\frac{19}{2}<10$$

as desired. □

6.4 Proving conjunctions

A **conjunction** is a statement of the type

> *P and Q.*

The statements P and Q are called the **conjuncts**. As with implications, conjunctions are not necessary presented explicitly.

A direct proof of a conjunction consists of the separate proofs of the two conjuncts, which must be differentiated clearly. It's common practice to put the conjuncts in an ordered list, say $i)$ and $ii)$. (If there are more than two conjuncts, the list may be extended using lower-case Roman numerals: $iii)$, $iv)$, etc.) Then the proof itself should use the matching labels $i)$ and $ii)$, and in each case we should begin by stating what we intend to prove.

PROOF.

$i)$ We prove P. …

$ii)$ We prove Q. …

Equality of sets is a hidden conjunction.

> *The sets A and B are equal.*

What are the conjuncts? Two sets are equal if they have the same elements, namely every element of A is an element of B, and vice-versa.

$$(A=B) \;\Leftrightarrow\; (A\subset B)\wedge(B\subset A).$$

The structure of the proof is clear once we recall the definition (3.18) of a subset.

PROOF.

$i)$ We prove that $A\subset B$. Let $x\in A$ be given. RTP: $x\in B$.
$ii)$ We prove that $B\subset A$. Let $x\in B$ be given. RTP: $x\in A$.

The equivalence of two statements is the canonical conjunction, since the equivalence operator $\Leftrightarrow$ is the conjunct of an implication and its converse —see (3.12).

P if and only if Q.

The structure of the proof is predictable.

PROOF.

$i)$ We prove $P\Rightarrow Q$. $\dots$
$ii)$ We prove $Q\Rightarrow P$. $\dots$

We could replace either part by a proof of the contrapositive implication.

The following theorem, which provides an alternative characterisation of primality, is of the above type.

Theorem[3]. *A natural number $n>1$ is prime if and only if n divides $(n-1)!+1$.*

This theorem says $P\wedge Q$, where

$$\begin{aligned} P &= \forall n>1,\ (n \text{ is prime}) \;\Rightarrow\; (n\mid(n-1)!+1)\\ Q &= \forall n>1,\ (n\mid(n-1)!+1) \;\Rightarrow\; (n \text{ is prime}). \end{aligned}$$

The outline of the proof is now clear.

PROOF.

$i)$ Let n be a prime number.
RTP: n divides $(n-1)!+1$.

---

[3] This result, known as Wilson's theorem, was first formulated by Bhaskara (a 7th Century Indian mathematician) and first proved by Lagrange.

ii) Let n be a natural number greater than 1 which divides $(n-1)!+1$.
RTP: n is prime.

The term ‘precisely’ may be used to turn a one-sided implication into an equivalence, hence a conjunction. Thus the proof of the statement

The set $\mathbb{Z}/n\mathbb{Z}$ *is a field precisely if* n *is prime.*

will have to be carried out in two stages.

Proof. Let $n \in \mathbb{N}$ be given.

i) Assume that $\mathbb{Z}/n\mathbb{Z}$ is a field. RTP: n is prime.

ii) Assume that n is prime. RTP: $\mathbb{Z}/n\mathbb{Z}$ is a field.

6.4.1 Loops of implications

The equivalence of several statements is sometimes expressed by a chains of implications in which the last statement in the chain coincides with the first one. This argument combines conjunctions with implications.

The equivalence of two statements P_1 and P_2, may be viewed as a loop:

$$P_1 \Rightarrow P_2 \Rightarrow P_1.$$

More generally, we consider loops of n implications:

$$P_1 \Rightarrow P_2 \Rightarrow \cdots \Rightarrow P_{n-1} \Rightarrow P_n \Rightarrow P_1.$$

Proving all implications in the loops amounts to proving that all statements are equivalent, namely that $P_i \Leftrightarrow P_j$, for all $i, j = 1 \dots, n$. This follows from the transitivity of the implication operator.

For example, let G be a group and H a subgroup of G. A **left coset** of H in G is a set $gH = \{gh : h \in H\}$ where g is an element of G. (This is a variant of the algebraic product of sets (2.13)). A result of group theory states that if x, y are elements of G, then the following statements are equivalent:

(a) $xH = yH$
(b) $xH \subseteq yH$
(c) $xH \cap yH \neq \emptyset$
(d) $y^{-1}x \in H$.

These are usually proved in a circle, for example $(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (d) \Rightarrow (a)$, but other arrangements are possible, e.g., $(a) \Leftrightarrow (b) \Rightarrow (c) \Rightarrow (d) \Rightarrow (b)$.

In chapter 7 we will use a loop of implications to show the equivalence of four formulations of the principle of induction.

6.5 Proof by contradiction

A proof by contradiction of a proposition P consists of assuming $\neg P$ and deducing a false statement:

$$\neg P \Rightarrow \text{FALSE}. \tag{6.6}$$

The false statement could be anything, including any assertion that contradicts the assumption $\neg P$. Contradiction works because if (6.6) holds, then we have the boolean equation

$$(\neg P \Rightarrow \mathrm{F}) = \mathrm{T}$$

to be solved for P. This equation has the unique solution $P = \mathrm{T}$, from the truth tables (3.8) and (3.9).

For example, in the proof by contradiction of the irrationality of $\sqrt{2}$, presented in section 6.1, we assumed that $\sqrt{2} \in \mathbb{Q}$, and deduced that

$$\exists m, n \in \mathbb{N},\ \ (\gcd(m,n) = 1)\ \Rightarrow\ (m, n \in 2\mathbb{Z})$$

which is clearly false.

A classic proof by contradiction is Euclid's proof of the infinitude of primes.

Theorem. *The number of primes is infinite.*

PROOF. Assume there are only finitely many primes, $p_1, \dots, p_n$. Consider the integer

$$N = 1 + \prod_{k=1}^{n} p_k. \tag{6.7}$$

Then N is greater than all the primes; moreover, if we divide N by p_k we get remainder 1, and therefore N is not divisible by any of the primes. Now, every integer greater than 1 is divisible by some prime. It follows that N must be divisible by a prime not in the given list, a contradiction. ☐

The statement that every integer greater than 1 is divisible by some prime, which is essential to the proof, needs some justification (see section 7.1).

To apply the method of contradiction to an implication $P \Rightarrow Q$, we assume the negation of the implication, namely $\neg(P \Rightarrow Q)$. By theorem 3.2, p. 57 this is equivalent to $P \wedge \neg Q$, so our proof amounts to establishing that

$$(P \wedge \neg Q) \Rightarrow \text{FALSE}.$$

This form of proof by contradiction is called the **both ends method**: To prove that *if P then Q*, we assume both P and not-Q, and we deduce something impossible. The appealing feature of this method is that it gives us two assumptions to exploit, namely P and $\neg Q$. This is often how mathematicians work, although this strategy may not be made explicit. It also easily causes confusion in writing, as the following example (from a textbook) shows.

Theorem. *For all real numbers x, if $x > 0$, then $1/x > 0$.*

BAD PROOF: If $1/x < 0$ then $(-1)/x > 0$, so $x((-1)/x) > 0$, i.e., $-1 > 0$, a contradiction. Therefore $1/x \geqslant 0$. Since $1/x$ can't be 0, we conclude that $1/x > 0$. □

In this proof, the assumption of P is hidden, and the assumption of $\neg Q$ is confused. Not helpful writing!

PROOF. We prove it by contradiction. Let us assume that $x > 0$ and $1/x \leqslant 0$. Multiplying the second inequality by -1, we obtain $-1/x \geqslant 0$, and since x is positive, multiplication by x yields $x(-1/x) \geqslant 0$. Simplification gives $-1 \geqslant 0$, which is false. □

The final deduction rests on the inequality $0 < 1$, which may be derived from the axioms of the real number system.

6.6 Counterexamples and conjectures

Suppose that a statement concerning all elements of a set is false. To prove this, it is sufficient to exhibit a single element of that set for which the proposition fails. This construct is called a **counterexample.** In symbols, we show that the statement

$$\forall x \in X, \ \mathscr{P}(x)$$

is false by proving its negation, namely

$$\exists x \in X, \ \neg\mathscr{P}(x).$$

To explore the idea of a counterexample, let us consider a famous polynomial proposed by Euler: $p(n) = n^2 + n + 41$. We evaluate $p(n)$ at integer values of the indeterminate: $n = 0, 1, \ldots$.

n	0	1	2	3	4	5	6	7	8	9	10	$\cdots$	20	$\cdots$	30
$p(n)$	41	43	47	53	61	71	83	97	131	151	181	$\cdots$	461	$\cdots$	971

All these values of $p(n)$ are prime! Moreover, $p(-n) = p(n-1)$, so it seems plausible —if a bit daring— to put forward the following conjecture:

Conjecture 1. *For all integers n, $p(n)$ is prime.*

A conjecture is a statement that we wish were a theorem. Three things may happen to a conjecture: *i*) someone proves it, and the conjecture becomes a theorem; *ii*) someone produces a counterexample, and the conjecture is proved false; *iii*) none of the above, and the conjecture remains a conjecture.

Our case is *ii*). The negation of conjecture 1 is

> *There is an integer n such that $p(n)$ is composite.*

To prove it, we exhibit such a value of n. For $n = 40$, we find

$$p(40) = 40^2 + 40 + 41 = 40(40+1) + 41 = 41(40+1) = 41^2.$$

So $p(40)$ is not prime, and conjecture 1 is false.

For a mathematician, the formulation of a conjecture is accompanied by the fear of a counterexample. Let us return to conjecture 1: we know it's false, but can it be modified into a weaker statement? It can be verified that $n = 40$ is the smallest value for n for which $p(n)$ is not prime[4]. Are there other such values? If these values were exceptional, then a meaningful weaker version of conjecture 1 could be

Conjecture 2. *For all but finitely many integers n, $p(n)$ is prime.*

Even if $p(n)$ were composite for, say, a billion values of n, the conjecture would still hold. Unfortunately, conjecture 2 is not true either. Its negation

> *There are infinitely many integers n for which $p(n)$ is composite*

is established by the simple, devastating counterexample:

$$p(41k) = (41k)^2 + 41k + 41 = 41(41k^2 + k + 1) \qquad k = 1, 2, \ldots \tag{6.8}$$

which says that 41 is a proper divisor of $p(41k)$ for infinitely many values of k.

Let's make one final attempt to salvage something from our original claim.

Conjecture 3. *There are infinitely many integers n such that $p(n)$ is prime.*

This is a very meaningful conjecture. Nobody has ever been able to prove that any quadratic polynomial with integer coefficients assumes infinitely many prime values. On the other hand, if we perform a numerical experiment we discover that, as n increases, $p(n)$ continues to supply prime values, even though the composite values slowly become dominant (see exercise 9.3.1). So this conjecture is very likely to be true, and it makes sense to put it forward.

---

[4] This phenomenon has deep roots, see [8, p. 155].

Arguably, the most famous conjecture in mathematics is the **Riemann's hypothesis** (RH), formulated in 1859. This conjecture (essentially) says that the zeta-function defined in equation (5.3), p. 103 vanishes only at points whose real part is equal to $1/2$. The depth of this statement is not at all apparent from its idiosyncratic formulation.

In a sense, the mathematics community cannot afford to wait for this important matter to be settled, and several theorems about the RH have been proved. Some of these theorems formulate the RH in terms of equivalent statements, establishing the importance of the RH in many areas of mathematics. There are also proofs of **conditional theorems**, which assume the validity of the RH, and deduce some of its consequences.

Other famous conjectures include the **twin-prime conjecture,** attributed to Euclid:

There are infinitely many primes p such that $p+2$ is also prime.

and the **Goldbach conjecture**[5]:

Every even integer greater than 2 can be written as a sum of two primes.

Both conjectures are easy to formulate and widely believed to be true, but a proof seems hopelessly difficult. This situation is common in the theory of numbers, where assessing the difficulty of a problem is not always easy. To illustrate this phenomenon, let us arrange the natural numbers into four columns, with the primes highlighted in boldface:

---

[5] Christian Goldbach (German: 1690–1764)

1	**2**	**3**	4
5	6	**7**	8
9	10	**11**	12
13	14	15	16
17	18	**19**	20
21	22	**23**	24
25	26	27	28
29	39	**31**	32
33	34	35	36
37	38	39	40
41	42	**43**	44
45	46	**47**	48
49	50	51	52
53	54	55	56
57	58	**59**	60
61	62	63	64
65	66	**67**	68
69	70	**71**	72
73	74	75	76
⋮	⋮	⋮	⋮

Now we ask some questions about this table, listed in order of increasing difficulty.

1. Why is there just one prime in the second column and no primes at all in the fourth?
2. The first two rows combined contain four primes; are there other pairs of adjacent rows with the same property?
3. Does the table contain infinitely many primes?
4. There are rows without primes. Are there infinitely many of them?
5. Are there adjacent rows without primes? What about sequences of three or more adjacent rows without primes?
6. Are there infinitely many primes in the first column?
7. Are 'half' of the primes in column 1 and 'half' in column 3?
8. Are there infinitely many rows containing two primes?

At the end of primary school, a pupil should be able to answer question 1, while in secondary school one could see why the answer to question 2 must be negative (one of three consecutive odd integers is divisible by 3). The answer to questions 3 to 7 is affirmative. Euclid's theorem (p. 136), which answers question 3, is normally taught in a first year university course, although we have seen that the proof does not require advanced ideas. Answering questions 4 and 5 will involve some ingenuity, but no major difficulty.

The mathematics needed to answer 6 and 7 becomes difficult. The proof that column 1 contains infinitely many primes, known since the 17th Century, is given today in an undergraduate course in number theory. The formulation of question 7 can be made precise, and the affirmative answer was given at the end of the 19th Century. Understanding the proof requires a solid background in analysis.

Many mathematicians would conjecture that the answer to question 8 is again positive, but at present nobody knows how to prove it. This is a variant of the twin primes conjecture stated above.

An accessible introduction to this beautiful part of mathematics is found in [37].

6.7 Wrong arguments

In constructing a mathematical argument it's easy to make mistakes. In this section we identify some common faulty arguments: confusing examples with proofs, assuming what we are supposed to prove, mishandling functions. Other mistakes will be examined in chapter 8 and in the exercises. Awareness of these problems should help us in avoiding them.

6.7.1 Examples vs. proofs

The verification of a statement in specific cases does not constitute a form of proof. Our study of Euler's polynomial in section 6.6 shows how misleading examples can be. This state of affairs is peculiar to mathematics, and indeed to modern mathematics. Ancient Greek, Medieval Arab, and Renaissance western mathematicians, instead of giving general arguments tended to offer examples to be copied. In other scientific disciplines, such as biology, a proof of a statement consists of careful experimental verifications of its validity.

In our first example the fault is easy to spot.

Theorem. *For all primes p, the integer $2^p - 2$ is divisible by p.*

Wrong Proof.

$$2^2-2=2\cdot 1,\quad 2^3-2=3\cdot 2,\quad 2^5-2=5\cdot 6,\quad 2^7-2=7\cdot 18,\quad \text{etc.}\ \square$$

The theorem has been proved only for $p=2,3,5,7$.

The next example is similar, but not so clear-cut [35, p. 138f].

Theorem. *For all* x, y *and* z, *if* $x+y<x+z$ *then* $y<z$.

Wrong Proof. Suppose $x+y<x+z$. Take $x=0$. Then

$$y=0+y<0+z=z.$$

□

The mistake here is that we took x to be 0, which is a special value of x. This assumption is wholly unjustified, since the quantities x,y,z are controlled by an existential quantifier, and hence no condition may be imposed on them. By adding the assumption that $x=0$, we have in fact proved the

Weaker Theorem. *For all* y *and* z, *if* $0+y<0+z$, *then* $y<z$.

This is not what we claimed to prove.

6.7.2 Wrong implications

Inappropriate handling of implications is a common source of mistakes. Often the problem originates from confusing an implication with its converse, that is, from $P\Rightarrow Q$ and Q, incorrectly deducing P. Since both implications $\mathrm{T}\Rightarrow\mathrm{T}$ and $\mathrm{F}\Rightarrow\mathrm{T}$ are true (see the truth table (3.10)), the statement P could be true or false. These faulty deductions are sometimes called circular arguments, or *non sequiturs*[6].

A common instance of this fault is assuming what one is meant to prove, and deducing from it a true statement.

Prove that $\sqrt{2}+\sqrt{6}<\sqrt{15}$

Wrong Proof.

$$\begin{aligned}\sqrt{2}+\sqrt{6}<\sqrt{15}\quad &\Rightarrow\quad (\sqrt{2}+\sqrt{6})^2<15\\ &\Rightarrow\quad 8+2\sqrt{12}<15\\ &\Rightarrow\quad 2\sqrt{12}<7\\ &\Rightarrow\quad 48<49\end{aligned}$$

---

[6] Latin for ‘it does not follow’.

The last statement is true, which completes the proof. □

We were supposed to prove P, where $P = (\sqrt{2}+\sqrt{6} < \sqrt{15})$. Instead we have assumed P, and correctly deduced from it the true statement $Q = (48 < 49)$. Therefore we still don't know if the original statement P is true or false. Indeed, had we started from the false statement $\sqrt{2}+\sqrt{6} < -\sqrt{15}$, we would have reached exactly the same conclusion.

There are two methods for fixing problems of this kind.

First method: retracing the steps.

We regard the chain of deductions displayed above as 'rough work'; then we start from the end and attempt to prove the chain of converse implications.

PROOF.

$$\begin{aligned} 48 < 49 \quad &\Rightarrow \quad \sqrt{48} < \sqrt{49} \\ &\Rightarrow \quad 2\sqrt{12} < 7 \\ &\Rightarrow \quad 8+2\sqrt{12} < 15 \\ &\Rightarrow \quad (\sqrt{2}+\sqrt{6})^2 < 15 \\ &\Rightarrow \quad \sqrt{2}+\sqrt{6} < \sqrt{15} \end{aligned}$$

where the first and the last implications are justified by the fact that we have taken the positive square root of each side. We have proved the proposition TRUE $\Rightarrow P$, from which we deduce that P is TRUE. □

Clearly, we could not have come up with such a proof without having done the 'rough work' first.

Second method: contradiction.

PROOF. Let us assume $\neg P$:

$$\begin{aligned} \sqrt{2}+\sqrt{6} \geqslant \sqrt{15} \quad &\Rightarrow \quad (\sqrt{2}+\sqrt{6})^2 \geqslant 15 \\ &\Rightarrow \quad 8+2\sqrt{12} \geqslant 15 \\ &\Rightarrow \quad 2\sqrt{12} \geqslant 7 \\ &\Rightarrow \quad 48 \geqslant 49. \end{aligned}$$

We have proved that $\neg P \Rightarrow$ FALSE, which implies that $\neg P$ is FALSE, that is, P is TRUE. □

The advantage of this method is that it doesn't require any 'rough work'.

6.7.3 Mishandling functions

Many things can go wrong when we deal with functions. For instance, we may apply a function to arguments outside its domain (e.g., forming ln(0)), or invert a function incorrectly (e.g., taking the wrong sign of a square root).

Let us begin with a proof from calculus which isn't quite correct, even if it captures the essence of the argument. The chain rule for differentiating composite functions states that if f and g are differentiable, then

$$(g \circ f)'(x) = g'(f(x))f'(x).$$

Sloppy proof. Let ε be a real number. We compute

$$\frac{g(f(x+\varepsilon)) - g(f(x))}{\varepsilon} = \frac{g(f(x+\varepsilon)) - g(f(x))}{f(x+\varepsilon) - f(x)} \times \frac{f(x+\varepsilon) - f(x)}{\varepsilon}.$$

Letting ε tend to zero gives the desired result. □

The right-hand side is obtained by multiplying and dividing by $f(x+\varepsilon) - f(x)$. But this quantity may well be zero for non-zero ε, for instance, if f is constant, which invalidates the argument. A valid proof requires a bit more delicacy.

Proof. Let $y = f(x)$. Since g is differentiable at y, we write

$$g(y+\delta) - g(y) = g'(y)\delta + \sigma(y,\delta)\delta$$

where σ is a continuous function with $\sigma(y,0) = 0$. Letting $\delta = f(x+\varepsilon) - f(x)$, we have

$$\begin{aligned}\frac{g(f(x+\varepsilon)) - g(f(x))}{\varepsilon} &= g'(f(x))\,\frac{f(x+\varepsilon) - f(x)}{\varepsilon} \\ &\quad + \sigma(y, f(x+\varepsilon) - f(x))\,\frac{f(x+\varepsilon) - f(x)}{\varepsilon}.\end{aligned}$$

If we now let ε tend to zero, then the last term tends to zero, being the product of a function that tends to zero and a function that tends to a finite limit, because f is differentiable. This gives the desired result. □

Mishandling a non-invertible function may have serious consequences:

Wrong theorem. *Every negative real number is positive.*

WRONG PROOF. Let α be a negative real number. Then

$$\begin{aligned}\alpha &= -1|\alpha| = (-1)^1|\alpha| = (-1)^{2\cdot\frac{1}{2}}|\alpha| = [(-1)^2]^{\frac{1}{2}}|\alpha| \\ &= 1^{\frac{1}{2}}|\alpha| = |\alpha| > 0\end{aligned}$$

as required. □

What's wrong with this proof? The function $x \mapsto x^2$ is not one-to-one, hence not invertible, and to invert it we must restrict its domain. If we restrict the domain to the range $x \geqslant 0$ and then take the positive sign for the square root, then $x \mapsto x^2$ and $x \mapsto \sqrt{x}$ are the inverse of each other: $x^{2\cdot\frac{1}{2}} = \sqrt{x^2} = x$. However, if x is negative, to ensure that squaring and square root are still inverse of each other, one must take the negative sign of the square root.

The mistake occurs in the last equality

$$1^{\frac{1}{2}}|\alpha| = |\alpha|$$

where the original negative argument $x = -1$ is paired with the positive sign of the square root. Had we chosen the correct negative sign, we would have recovered the value -1 that we started with. Exercise 6.3 deals with a more general instance of the same phenomenon.

6.8 Writing a good proof

Proofs come in all shapes and sizes. Correctness is, of course, imperative, but a good proof requires more: in a proof we must resolve the conflicting requirements of clarity and conciseness. On the one hand, reading a proof is demanding, and any assistance will be welcome. On the other hand, a proof is a rigorous exercise in logic and the exposition must remain succinct and essential. The level of exposition and the amount of detail will depend on the mathematical maturity of the target audience.

There is one universal rule. At all key junctures in a proof:

- SAY WHAT YOU PLAN TO DO.
- WHEN YOU'VE DONE IT, SAY SO.

We now examine statements of theorems and proofs that are less than optimal, and we seek to improve them. Our proofs are very detailed, and suitable for first-year university students. Occasionally, we also provide a concise version of a proof for a mathematically mature audience.

Our first theorem is taken from a textbook

Bad Theorem. *If $x^2 \neq 0$, then $x^2 > 0$.*

Bad proof. If $x > 0$ then $x^2 = xx > 0$. If $x < 0$ then $-x > 0$, so $(-x)(-x) > 0$, i.e., $x^2 > 0$. $\square$

The statement of the theorem is incomplete, in that there is no information about the quantity x. The theorem is an implication, but in the proof what has happened to the assumption $x^2 \neq 0$? If the assumption is obvious, one may leave it out, but one must judge whether the readers are sophisticated enough to see what it must be. Furthermore, this is a proof by cases (see section 6.2), and it would be helpful if this were made explicit.

Theorem. *For all real numbers x, if $x^2 \neq 0$, then $x^2 > 0$.*

Proof. Let x be a real number such that $x^2 \neq 0$. Then $x \neq 0$, and have two cases:
$i)$ $x < 0$. Then $-x > 0$, so $(-x)(-x) > 0$, that is, $x^2 > 0$.
$ii)$ $x > 0$. Then $x^2 = xx > 0$. $\square$

The following statement establishes a simple relation between rational and irrational numbers.

Bad Theorem. $\forall a, b \in \mathbb{R},\ (a \in \mathbb{Q} \wedge b \notin \mathbb{Q}) \Rightarrow a + b \notin \mathbb{Q}$.

Bad Proof: Suppose $a + b \in \mathbb{Q}$. Then $a + b = m/n$.
If $a \in \mathbb{Q}$, then $a = p/q$, and $b = m/n - p/q \in \mathbb{Q}$. $\square$

The symbolic formulation obscures the simple content of the theorem. The proof is simple, but it introduces unnecessary symbols, while clarity could be improved by inserting some short comments.

Theorem. *The sum of a rational number and an irrational number is irrational.*

Proof. Let a and z be a rational and an irrational number, respectively. Consider the identity $z = (a + z) - a$. If $a + z$ were rational, then z would also be rational, being the difference of two rational numbers. This contradiction shows that $a + z$ must be irrational. $\square$

The choice of symbols keeps the two numbers well-separated in our mind.

Next we consider an arithmetical statement.

Bad Theorem. $\forall n \in \mathbb{Z},\ 2 \nmid n \Rightarrow 8 | n^2 - 1$.

Bad proof.
$2 \nmid k \Rightarrow \exists j \in \mathbb{Z}\ k = 2j + 1$;

$k^2 - 1 = 4j(j+1)$;
$\forall j \in \mathbb{Z},\ 2 \mid j(j+1)$. □

Both statement and proof are exceedingly cryptic, as they use only symbols. A mixture of words and symbols is necessary.

THEOREM. *The square of an odd integer is of the form* $8n+1$.

DETAILED PROOF. Let k be an odd integer. We will show that $k^2 = 8n+1$, for some $n \in \mathbb{Z}$.

Since k is odd, we have $k = 2j+1$, for some integer j. We find

$$k^2 = (2j+1)^2 = 4j^2 + 4j + 1 = 4j(j+1) + 1. \tag{1}$$

Now, one of j or $j+1$ is even, and therefore their product is even. Thus we have $j(j+1) = 2n$, for some n. Inserting this expression in (1) we find $k^2 = 8n+1$, as desired. □

This proof is quite detailed, and can safely be shortened.

PROOF. An odd integer k is of the form $k = 2j+1$, for some integer j. A straightforward manipulation gives $k^2 = 4j(j+1)+1$. Our claim now follows from the observation that the product $j(j+1)$ is necessarily even. □

The adverb ‘necessarily’ is inserted to signal that the conclusion (the product $j(j+1)$ is even) requires a moment’s thought. The expression ‘straightforward manipulation’ is used to omit some steps in the derivation, while warning the reader that some calculations are required. In such a circumstance, the adjective ‘straightforward’ is preferable to ‘easy’, which is subjective, or ‘trivial’, which carries a hint of arrogance. So it’s appropriate to claim that the arithmetical identity

$$\begin{aligned} 29 \;=\; & \left(2 + \frac{\sqrt{-1}\,(1+\sqrt{5})}{2}\right)\left(2 - \frac{\sqrt{-1}\,(1+\sqrt{5})}{2}\right) \\ & \times \left(2 + \frac{\sqrt{-1}\,(1-\sqrt{5})}{2}\right)\left(2 - \frac{\sqrt{-1}\,(1-\sqrt{5})}{2}\right) \end{aligned}$$

can be verified with a ‘straightforward calculation’.

In the following example everything needs re-writing: definition, statement of theorem, proof.

BAD DEFINITION. Let b_j be decimal digits. We define

$$\overline{b_1 \cdots b_n} = b_1 \cdots b_n b_1 \cdots$$

BAD THEOREM. *Let* $m \in \mathbb{Z}$, *and let* $x = m + 0.a_1 \cdots a_k\overline{b_1 \cdots b_n}$. *Then* $x \in \mathbb{Q}$.

BAD PROOF: Let $A = 0.a_1 \cdots a_k$, $B = 0.b_1 \cdots b_n$. Then $A, B \in \mathbb{Q}$, and

$$0.\overline{b_1 b_2 \ldots b_n} = B\left(1 + \frac{1}{10^n} + \frac{1}{10^{2n}} + \cdots\right) = B\frac{10^n}{10^n - 1} =: B' \in \mathbb{Q}.$$

Thus $x = m + A + B'10^{-k} \in \mathbb{Q}$. □

In the definition, the periodicity of the sequence of digits is not as clear as it could be. In the statement of the theorem, the digits a_j are undefined. In addition, a hybrid notation is used for x, which appears as the sum of a decimal number and an integer expressed symbolically.

As we did before, we state the theorem with words, which is more effective. We introduce the notation *after* the statement of the theorem but *before* the beginning of the proof. This arrangement should be considered if the notation is needed outside the confines of the proof, or if the proof is heavy and needs lightening up, or simply as a variation from the rigid definition-theorem-proof format. We number the various steps in the argument for future reference.

THEOREM. *Every number whose decimal digits eventually repeat is rational.*

1. Before proving the theorem, we establish some notation. Let $a_1 a_2 \cdots$ and $b_1 b_2 \cdots$ be strings of decimal digits: $a_i, b_i \in \{0, 1, \ldots, 9\}$. We use the over-bar to denote a string of digits consisting of a pattern repeating indefinitely:

$$\overline{b_1 \cdots b_n} = \underbrace{b_1 \cdots b_n}_{n}\underbrace{b_1 \cdots b_n}_{n} \cdots. \tag{6.9}$$

DETAILED PROOF.
2. Let x be a real number with eventually repeating decimal digits. Then x has a decimal representation of the type

$$x = a_0 \cdots a_j . a_{j+1} \cdots a_m \overline{b_1 \cdots b_n}$$

for some integers m, n, j, with $m \geqslant 0$, $n \geqslant 1$, and $0 \leqslant j \leqslant m$.
3. We shall compute the value of x in terms of two integers A and B, defined as follows:

$$A = a_0 \cdots a_m = \sum_{k=0}^{m} a_k 10^{m-k} \qquad B = b_1 \cdots b_n = \sum_{k=1}^{n} b_k 10^{n-k}.$$

First, we get rid of the aperiodic part; we shift it to the left of the decimal point by multiplying by a suitable power of 10, and then we subtract it off, to get

$$10^{m-j}x - A = 0.\overline{b_1 \cdots b_n}. \tag{6.10}$$

4. Next we compute, using (6.9)

$$\begin{aligned}
0.\overline{b_1 \cdots b_n} &= \frac{B}{10^n} + \frac{B}{10^{2n}} + \frac{B}{10^{3n}} + \cdots \\
&= B\left(\frac{1}{10^n} + \frac{1}{10^{2n}} + \frac{1}{10^{3n}} + \cdots\right) \\
&= B\sum_{k=1}^{\infty}\left(\frac{1}{10^n}\right)^k = \frac{B}{10^n - 1}.
\end{aligned}$$

5. In the last step we have used the formula of the sum of the geometric series

$$\sum_{k=1}^{\infty} q^k = \frac{q}{1-q} \qquad |q| < 1.$$

From equation (6.10) and the result above, we obtain

$$10^{m-j}x - A = \frac{B}{10^n - 1}$$

and hence

$$x = 10^{j-m}\left(A + \frac{B}{10^n - 1}\right).$$

The right-hand side of this equation consists of sums and products of rational numbers, and is therefore rational. □

Let us examine the main steps in the proof:

1. Since we don't begin the proof straight away, we say so. The key notation (the over-bar) is established first in words, then in symbols. The under-brace highlights the structure of the expression.
2. In the theorem there is a hidden quantifier, and the proof begins accordingly. Then we introduce the notation for x, and we say why.
3. At every opportunity, we declare our intentions.
4. In this passage, one must decide what constitutes an appropriate amount of details: we have been cautious here.
5. A reminder of a well-known summation formula.

We give a concise version of the same proof.

PROOF. Let x be a real number with eventually repeating decimal digits b_k. Without loss of generality, we may assume that the integer part of x is zero, and that the fractional part is purely periodic:

$$x = 0.\overline{b_1 \cdots b_n}. \tag{1}$$

(Any real number may be reduced to this form by first multiplying by a power of 10, and then subtracting an integer, and neither operation affects the property of having repeated digits.) Defining the decimal integer $B = b_1 \cdots b_n$, we find, from (1)

$$\begin{aligned} x &= \frac{B}{10^n} + \frac{B}{10^{2n}} + \frac{B}{10^{3n}} + \cdots = B \sum_{k=1}^{\infty} \left(\frac{1}{10^n} \right)^k \\ &= \frac{B}{10^n - 1}. \end{aligned}$$

We see that x is rational. □

The expression 'without loss of generality' indicates that the restriction being introduced does not weaken the argument in any way (see also section 4.2). A brief parenthetical remark is added for clarification: it too could be omitted.

Exercises

Exercise 6.1. You are given mathematical statements with cryptic proofs. Rewrite the proofs in a style appropriate for first-year university students, who may have forgotten basic facts.

(a) *Prove that the line through the point* $(4,5,1) \in \mathbb{R}^3$ *parallel to the vector* $(1,1,1)^T$ *and the line through the point* $(5,-4,0)$ *parallel to the vector* $(2,-3,1)^T$ *intersect at the point* $(1,2,-2)$. (The symbol T denotes the transpose.)

BAD PROOF.

$$\begin{aligned} &\mathbf{v} = (4,5,1)^T + \lambda\,(1,1,1)^T = (4+\lambda, 5+\lambda, 1+\lambda)^T \\ &\quad = (5,-4,0)^T + \mu\,(2,-3,1)^T = (5+2\mu, -4-3\mu, \mu)^T; \\ &4+\lambda = 5+2\mu;\ 5+\lambda = -4-3\mu;\ 1+\lambda = \mu; \\ &4+\lambda = 5+2(1-\lambda) \ \Rightarrow\ \lambda = -3,\ \mu = -2; \\ &\text{this gives } \mathbf{v} = (1,2,-2)^T.\ \square \end{aligned}$$

(b) *Prove that the real function* $x \mapsto 3x^4 + 4x^3 + 6x^2 + 1$ *is positive.*

BAD PROOF.

$f'(x) = 12x(x^2+x+1) = 0 \Leftrightarrow x = 0;$
$f''(0) > 0,\ f(0) > 0.$ □

(c) *Prove that for all real values of a, the line*

$$y = ax - \left(\frac{a-1}{2}\right)^2$$

is tangent to the parabola $y = x^2 + x$.

BAD PROOF.

$ax - (a-1)^2/4 = x^2 + x;$
$x^2 + x(1-a) + (a-1)^2/4 = 0;$
$x = (a-1)/2.$
For this x, dy/dx is the same. □

(d) *Prove that the vector $(1,0,1)$ does not belong to the subspace of $\mathbb{R}^3$ generated by the vectors $(3,1,2)$ and $(-2,1,0)$.*

BAD PROOF:

$3t - 2s = 1;\ t + s = 0;\ 2t = 1.$
$\Rightarrow t = 1/2 \Rightarrow s = -1/2 \Rightarrow 3/2 = 0.$ □

Exercise 6.2. The following theorem and proof have several faults. (a) Explain what they are; (b) write an appropriate revision.
WRONG THEOREM. *For all numbers x and y,*

$$\frac{x^2+y^2}{|xy|} > 2.$$

WRONG PROOF. For

$$\begin{aligned} \frac{x^2+y^2}{|xy|} > 2 \quad &\Rightarrow \quad x^2+y^2 > 2xy \\ &\Rightarrow \quad x^2 - 2xy + y^2 > 0 \\ &\Rightarrow \quad (x-y)^2 > 0. \end{aligned}$$

The last inequality is trivially true, which proves it. □

Exercise 6.3. Identify the flaw in the proof of the following statement.

WRONG THEOREM. *The sine and cosine functions are constant over* $\mathbb{C}$.

WRONG PROOF. We shall be using the formulae

$$\cos(\theta)=\frac{e^{i\theta}+e^{-i\theta}}{2}\qquad\qquad \sin(\theta)=\frac{e^{i\theta}-e^{-i\theta}}{2i}$$

where θ is an arbitrary complex number. We find

$$e^{i\theta} \;=\; e^{i\theta\frac{2\pi}{2\pi}}=e^{2\pi i\frac{\theta}{2\pi}}=\left(e^{2\pi i}\right)^{\frac{\theta}{2\pi}}=1^{\frac{\theta}{2\pi}}=1$$

and hence $\cos(\theta)=1$ and $\sin(\theta)=0$, as desired. □

Exercise 6.4. Write the first few sentences of the proof of each statement, introducing all relevant notation in an appropriate order, and identifying the RTP. (For this task, understanding the meaning of the statements is unimportant.)

1. *On a compact set every continuous function is uniformly continuous.*
2. *Every valuation of a field with prime characteristic is non-archimedean.*
3. *No polynomial with integer coefficients, not a constant, can be prime for all integer values of the indeterminate.*
4. *Any hyperbolic matrix in* $\mathrm{SL}_2(\mathbb{Z})$ *is conjugate to a matrix with non-negative coefficients.*
5. *The only bi-unique analytic mapping of the interior of the unit disc into itself is a linear fractional transformation.*
6. *In a field with a non-archimedean absolute value, the connected component of any point is the set consisting of only that point.*
7. *The characteristic polynomial of the incidence matrix of a Pisot substitution is irreducible over* $\mathbb{Q}$.
8. *The first Betti number of a compact and orientable Riemannian manifold of positive Ricci curvature is zero.*
9. *Every differential of the second or third kind differs from some normal differential by a differential of the first kind.*
10. *Any two antiderivatives of a generalised function differ only by a constant.*
11. *A real function continuous in a closed interval attains all values between its upper and lower bounds.*

12. *A subset of a metric space is open if and only if its complement is closed.*
13. *A stable set of a graph is maximum if and only if there exists no maximal alternating sequence of odd length.*
14. *A bi-infinite sequence over the alphabet* $\{0, 1\}$ *is Sturmian if and only if it is balanced and not eventually periodic.*
15. *The curvature of a hermitian manifold vanishes if and only if it is possible to choose a parallel field of orthonormal holomorphic frames in a neighbourhood of each point of the manifold.*

Chapter 7

Induction

Induction is a fundamental method of proof, used in every corner of mathematics. It is applicable to statements of the form

$$\forall n \in \mathbb{N}, \;\; \mathscr{P}(n) \tag{7.1}$$

where $\mathscr{P}$ is a predicate over $\mathbb{N}$. This sentence says that all natural numbers have property $\mathscr{P}$.

Induction is first met in the proof of summation identities and inequalities:

1)	$\displaystyle\sum_{k=0}^{n} x^k = \frac{1-x^{n+1}}{1-x}$	$x \neq 1$	Sum of geometric progression
2)	$\displaystyle(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$		Binomial theorem
3)	$(1+x)^n \geqslant 1+nx$	$x > -1$	Bernoulli inequality[1]
4)	$\displaystyle\prod_{k=1}^{n} x_k \leqslant \left(\frac{1}{n}\sum_{k=1}^{n} x_k\right)^n$	$x_k > 0$	Arithmetico-geometric mean (AGM) inequality
5)	$\displaystyle\left\lvert\sum_{k=1}^{n} x_k y_k\right\rvert^2 \leqslant \sum_{k=1}^{n} \lvert x_k\rvert^2 \sum_{k=1}^{n} \lvert y_k\rvert^2$		Cauchy-Schwarz inequality[2]

(The variables are complex numbers in 1,2, and real numbers in 3,4,5.)

The expression (7.1) unfolds into an infinite sequence of boolean expressions:

$$\mathscr{P}(1),\, \mathscr{P}(2),\, \mathscr{P}(3),\, \ldots$$

and to prove it we must show that all expressions in the sequence are true:

$$\mathrm{T},\ \mathrm{T},\ \mathrm{T},\ \ldots.$$

For instance, the Bernoulli inequality unfolds as follows:

$$1+x \geqslant 1+x, \quad (1+x)^2 \geqslant 1+2x, \quad (1+x)^3 \geqslant 1+3x, \quad \ldots$$

each inequality being valid for all real numbers $x > -1$.

A special form of mathematical argument is needed for this purpose, called **mathematical induction**. It takes several forms, and we will examine four:

(A) The well-ordering principle.
(B) The infinite descent method.
(C) The induction principle.
(D) The strong induction principle.

We will show that these principles are equivalent in the sense that any one of them implies all the others. The different forms of the principle are appropriate for different situations.

The term mathematical induction was coined by De Morgan around 1840. The attribute ‘mathematical’ is used to differentiate this concept from the homonymous concept in philosophy, which has a different meaning[3]. The induction principle — in any of the above forms— is one of the five **Peano axioms**[4] which characterise the structure of the set $\mathbb{N}$ of natural numbers.

7.1 The well-ordering principle

The **well-ordering** (or **least counterexample**) **principle** states that

(A) *Every non-empty set of natural numbers contains a least element.*

This principle is clearly false for subsets of $\mathbb{Z}, \mathbb{Q}$ or $\mathbb{R}$, and, less trivially so, for subsets of $\mathbb{Q}^+$ or $\mathbb{R}^+$, or of any real or rational interval. A possible strategy for proving (7.1) is to combine well-ordering with contradiction. Such a proof will begin as follows:

---

[3] In philosophy, an inductive argument is one that uses very strong premises to support the *probable* truth of a conclusion.

[4] Giuseppe Peano (Italian 1858–1932).

PROOF. Suppose (7.1) is false. Let k be the least natural number n for which $\mathscr{P}(n)$ is false. ...

Then with this k we try to deduce a contradiction, using the knowledge that $\mathscr{P}(n)$ is true for $n = 1, \ldots, k-1$. The integer k defined in the proof exists by virtue of the well-ordering principle: it is the least element of the set $\{n \in \mathbb{N} : \neg\mathscr{P}(n)\}$, which is a subset of $\mathbb{N}$. By assumption, this subset is non-empty.

Let us apply this strategy to some specific problems.

Proposition. *If $n \geqslant 7$, then $n! > 3^n$.*

PROOF. Suppose this statement is false, and let k be the least integer $n \geqslant 7$ such that $n! \leqslant 3^n$. We know that $k > 7$, since

$$7! = 5040 > 2187 = 3^7.$$

Put $j = k-1$. Then $j \geqslant 7$, and hence $j! > 3^j$ by choice of k. So

$$k! = (j+1)j! > (j+1)3^j > 3 \cdot 3^j = 3^k$$

contradicting the choice of k. □

The initial verification that $k > 7$ creates enough room for the principle to work. We note that $6! = 720$ and $3^6 = 729$, so the inequality $n \geqslant 7$ is the best possible one. Under these circumstances, we say that $n \geqslant 7$ is a **strict bound**. Our next statement is more substantial:

Theorem. *Every integer greater than 1 is a product of one or more prime numbers.*

PROOF. Suppose the statement is false, and let k be the smallest integer greater than 1 which is not a product of primes. Then k is not prime, so $k = mn$ for two smaller integers $m, n \geqslant 2$. Since m, n are smaller than k and greater than 1, they are products of prime numbers. So k is a product of primes too, because $k = mn$. This contradicts the choice of k. □

This form of mathematical induction can be very effective. The assumption that k is a smallest counterexample puts the maximum amount of information on the table; it gives us the feeling that we are examining a concrete object.

7.2 The infinite descent method

This is a second form of mathematical induction, developed in the 17th Century by Fermat[5]. It exploits the following variant of the well-ordering principle:

---

[5] Pierre de Fermat, French lawyer and amateur mathematician (1601 or 1607/8–1665)

(B) *There is no infinite decreasing sequence of natural numbers.*

This principle follows from the well-ordering principle ($(A) \Rightarrow (B)$): if there were an infinite decreasing sequence of natural numbers

$$n_1 > n_2 > n_3 > \cdots$$

then the set $\{n_1, n_2, n_3, \ldots\}$, which is a non-empty subset of $\mathbb{N}$, would have no smallest element. Conversely, well-ordering follows from descent ($(B) \Rightarrow (A)$): if there were a non-empty set $S \subset \mathbb{N}$ without a smallest element, then choosing an arbitrary element $n_1 \in S$, we could find $n_2 \in S$ with $n_2 < n_1$. Repeating this process we would generate an infinite decreasing sequence of natural numbers.

Fermat applied this method extensively to arithmetical problems. He pointed out that infinite descent is a method for proving that certain properties or relations for whole numbers are *impossible*, and for this purpose he combined descent with contradiction —as we did for well-ordering. To prove that no positive integer can have a certain set of properties, we assume that a positive integer does have these properties, and we try to deduce that there is a smaller positive integer with the same set of properties. If we succeed, then, by the same argument, these properties would hold for some smaller positive integer, and so forth *ad infinitum*. Because a sequence of natural numbers cannot decrease indefinitely, no positive integer can have the stated properties.

Infinite descent was brought to prominence by Euler, who proved with it that it is impossible to find natural numbers x, y, x such that

$$x^3 + y^3 = z^3.$$

This is a special case of the celebrated **Fermat's last theorem**, in which the exponent 3 is replaced by an arbitrary integer n greater than 2. In keeping with this tradition, we now apply the method of infinite descent to prove the non-existence of integer solutions of a polynomial equation.

Theorem. *For every prime p the equation $px^2 = y^2$ has no solutions in the natural numbers.*

This statement is equivalent to the irrationality of $\sqrt{p}$, as one sees by dividing each term of the equation by x^2, and then taking the positive square root. The argument presented here is a variant and a generalisation of that given in section 6.1 for the irrationality of $\sqrt{2}$.

PROOF. Suppose that there are natural numbers n_0, n_1, such that $pn_1^2 = n_0^2$. So pn_1^2 is the square of a natural number. Since p divides $n_0^2 = n_0 n_0$, by the basic property of prime numbers p must divide n_0, and we have $n_0 = pn_2$ for some natural number n_2. Then $pn_1^2 = p^2 n_2^2$, and hence $n_1^2 = pn_2^2$. So $n_2 < n_1$ and pn_2^2 is the square of a natural number.

Repeating the argument, there is $n_3 < n_2$ such that pn_3^2 is the square of a natural number. This continues for ever, which is impossible. Hence the equation $px^2 = y^2$ has no solution, as stated. $\Box$

7.3 Peano's induction principle

This is a third form of mathematical induction; it was proposed by Dedekind[6] and formalised by Peano. The **principle of induction** states that:

(C) *If* $\mathscr{P}$ *is a predicate over* $\mathbb{N}$ *such that*

$i)$ $\mathscr{P}(1)$ *is true*

$ii)$ *for all* $k \in \mathbb{N}$, $\mathscr{P}(k) \Rightarrow \mathscr{P}(k+1)$

then $\mathscr{P}(n)$ *is true for all* $n \in \mathbb{N}$.

The first condition is called the **basis** for the induction (or the **base case**); the second the **inductive step.** The bi-unique correspondence between predicates $\mathscr{P}$ over $\mathbb{N}$ and subsets S of $\mathbb{N}$, given by (see equations (3.15) and (3.15))

$$S = \{n \in \mathbb{N} : \mathscr{P}_S(n)\}, \qquad \mathscr{P}_S(x) = (x \in S)$$

allows us to reformulate the induction principle in the language of sets.

(C') *Let* S *be a subset of* $\mathbb{N}$. *If*

$$i) \quad 1 \in S \qquad ii) \quad \forall k \in \mathbb{N},\ k \in S \Rightarrow (k+1) \in S$$

then $S = \mathbb{N}$.

The principle of induction follows from well-ordering $((A) \Rightarrow (C))$. To show this, we define the set

$$S' = \mathbb{N} \setminus S = \{n \in \mathbb{N} : \neg\mathscr{P}(n)\}.$$

---

[6] Richard Dedekind (German: 1831–1916).

We prove that S' is empty by contradiction. If S' is non-empty, then, by the well-ordering axiom, S' has a least element k. Since $1 \in S$, we have $k \neq 1$, and since k is the smallest element of S', it follows that $k-1 \in S$. Then, putting $j = k-1$, we find from condition $ii)$ that $j+1 = k \in S$, which contradicts the fact that $k \in S'$.

The base case for Peano's axiom could be any integer, not just 1. Indeed if $\mathscr{P}(n)$ is valid for all $n \geqslant k$, then by letting $m = n-k+1$, the range $n \geqslant k$ becomes $m \geqslant 1$.

Peano's induction provides straightforward proofs of finite sums and products formulae:

$$\forall n \in \mathbb{N}, \qquad \sum_{k=1}^{n} a_k = F(n) \tag{7.2}$$

$$\forall n \in \mathbb{N}, \qquad \prod_{k=1}^{n} a_k = G(n) \tag{7.3}$$

where (a_k) is a sequence of numbers (or more generally of elements of a ring), and F and G are explicit functions of n, hopefully easier to compute than the original sum or product.

In an inductive proof of (7.2), the base case consists of verifying that $a_1 = F(1)$. For the inductive step, we assume that the formula holds for some $n = k$, and using the induction hypothesis we obtain:

$$\sum_{j=1}^{k+1} a_j = \sum_{j=1}^{k} a_j + a_{k+1} = F(k) + a_{k+1}.$$

Thus the proof reduces to the verification of the statements

$$a_1 = F(1), \qquad F(k) + a_{k+1} = F(k+1) \quad k \geqslant 1 \tag{7.4}$$

which no longer involve summation. For example, to prove the formula

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6} \qquad n = 1, 2, \dots$$

we must verify that

$$1^1 = \frac{1 \cdot 2 \cdot 3}{6}, \qquad \frac{k(k+1)(2k+1)}{6} + (k+1)^2 = \frac{(k+1)(k+2)(2k+3)}{6}.$$

For a product formula, the expressions (7.4) are replaced by

$$a_1 = G(1), \qquad G(k) \times a_{k+1} = G(k+1) \quad k \geqslant 1. \tag{7.5}$$

In some cases the identities (7.4) and (7.5) may be established using a computer algebra system. This is an instance of **computer-assisted proof**[7] —see exercise 9.1.

A similar arrangement applies to the inductive proof of inequalities. However, even in the simplest cases, such proofs are not as mechanical as those of summation and product formulae, as the following example illustrates.

Proposition. *If $n > 3$, then $2^n \geqslant n^2$.*
PROOF. We prove it by induction on n. The base case $n = 4$ is immediate: $2^4 \geqslant 4^2$. Assume now that for some $k \geqslant 4$ we have $2^k \geqslant k^2$. Then

$$2^{k+1} = 2 \cdot 2^k \geqslant 2 \cdot k^2$$

where the inequality follows from the induction hypothesis. To complete the proof, we must show that for all $k \geqslant 4$ we have $2 \cdot k^2 \geqslant (k+1)^2$. Now the polynomial

$$p(k) = 2 \cdot k^2 - (k+1)^2 = k^2 - 2k - 1$$

has roots $1 \pm \sqrt{2}$, with

$$-1 < 1 - \sqrt{2} < 1 + \sqrt{2} < 3.$$

Thus, for $k \geqslant 3$, we have $p(k) > 0$, or $2k^2 > (k+1)^2$. Hence $2^{k+1} \geqslant (k+1)^2$, completing the induction. □

The next example warns us that a straightforward inductive strategy doesn't always work.

Proposition. *For all integers $n \geqslant 0$ and real numbers $x > -1$, we have*

$$(1+x)^n > nx. \tag{7.6}$$

We attempt to prove this by induction on n.

Base case: $n = 0$. Since $x > -1$, we have $(1+x)^0 = 1 > 0 = 0x$, as desired.
Inductive step: Assume (7.6) for some $n = k \geq 0$. Then

$$\begin{aligned} (1+x)^{k+1} &= (1+x)(1+x)^k \\ &> (1+x)kx \quad \text{(by induction hypothesis)} \\ &= kx + kx^2 \ \text{(??)} \end{aligned}$$

---

[7] This concept raises controversy among mathematicians.

We don't seem to be getting anywhere.

PROOF. We prove the *stronger* inequality $(1+x)^n \geqslant 1+nx$. This is the **Bernoulli inequality** —see p. 155.
Base case: $n=0$. Since $x>-1$, we have $(1+x)^0 = 1 \geqslant 1+0x$.
Inductive step: Assume the inequality for $n=k\geqslant 0$. Then

$$\begin{aligned} (1+x)^{k+1} &= (1+x)(1+x)^k \\ &\geqslant (1+x)(1+kx) \quad \text{(by induction hypothesis)} \\ &= 1+(k+1)x+kx^2 \geqslant 1+(k+1)x. \end{aligned}$$

The inductive step is complete. □

7.4 Strong induction

In an induction proof, the successful completion of the inductive steps rests on the assumption that to prove $\mathscr{P}(k+1)$ we only need $\mathscr{P}(k)$. Should we instead (or in addition) need $\mathscr{P}(k-1)$, then this method would collapse.

This situation is not unusual. Suppose we want to prove by induction that every integer is a product of prime numbers. To prove that this statement is true for, say, $k+1=12$, the preceding case $k=11$ is of no use. We would need instead $k=3$ and $k=4$.

The rescue is a form of induction called **strong induction** (or **complete induction**), which is formulated as follows:

(D) *Let $\mathscr{P}$ be a predicate on $\mathbb{N}$. If*

- $i)$ *$\mathscr{P}(1)$ is true;*
- $ii)$ *for any $k\in\mathbb{N}$, if $\mathscr{P}(1), \mathscr{P}(2), \ldots, \mathscr{P}(k)$ are true, so is $\mathscr{P}(k+1)$;*

then $\mathscr{P}(n)$ is true for all $n\in\mathbb{N}$.

As with Peano's induction, the base case for strong induction can be any integer, and there is an equivalent formulation in terms of sets:

(D') *Let S be a subset of $\mathbb{N}$. If*

- $i)$ *$1\in S$;*
- $ii)$ *for any $k\in\mathbb{N}$, $\{1,2,\ldots,k\}\subset S \Rightarrow k+1\in S$*

then $S = \mathbb{N}$.

Strong induction follows from Peano's induction ($(C) \Rightarrow (D)$), since the latter has fewer conditions. (One could say that induction is *stronger* than strong induction!)

Let us prove by strong induction that every integer greater than 1 is a product of one or more prime numbers. Having established this result earlier by well-ordering, it's instructive to compare the two arguments.

PROOF. We use strong induction. The integer 2 is prime, which establishes the base case. Suppose that k is an integer greater than 1 and that every integer j with $2 \leqslant j < k$ is a product of primes. We have two cases:

Case I: k is a prime. Then k is a product of primes.
Case II: k is not a prime. Then there are $i, j > 1$ with $ij = k$. By the induction hypothesis, i and j are products of primes, and hence so is k. $\square$

With the help of strong induction, Euclid's elegant proof of the infinitude of primes (p. 136) yields an **upper bound** (or an **estimate**) for the nth prime number.

Theorem. *For all* $n \geqslant 1$*, the* n*th prime* p_n *satisfies the upper bound* $p_n \leqslant 2^{2^{n-1}}$*.*

PROOF. We use strong induction. The base case is clear: $p_1 = 2 \leqslant 2^{2^0} = 2$. Assuming that for some $k \geqslant 1$ we have $p_j \leqslant 2^{2^{j-1}}$ for $j = 1, 2, \ldots k$, we have

$$\begin{aligned} p_{k+1} &\leqslant 1 + \prod_{j=1}^{k} p_j \leqslant 1 + \prod_{j=1}^{k} 2^{2^{j-1}} \\ &= 1 + 2^{\sum_{j=1}^{k} 2^{j-1}} = 1 + 2^{2^k - 1} < 2^{2^k} \end{aligned}$$

where the first inequality follows from Euclid's construction (6.7), p. 136, while the second inequality follows from the induction hypothesis. The proof is complete. $\square$

To complete the proof that the four formulations of the induction principle are equivalent, it remains to show that well-ordering follows from strong induction ($(D) \Rightarrow (A)$). When this is done, we will have proved that

$$(A) \Leftrightarrow (B) \qquad \text{and} \qquad (A) \Rightarrow (C) \Rightarrow (D) \Rightarrow (A)$$

and equivalence follows from a loop of implications (cf. section 6.4.1).

We use the both ends method. Assume that strong induction holds but that there is a non-empty subset S of $\mathbb{N}$ without a least element. Let $S' = \mathbb{N} \smallsetminus S$. Then $1 \in S'$, for otherwise 1 would be the least element of S. Likewise, if for some $k \geqslant 1$ we have $\{1, \dots, k\} \subset S'$, then we must also have $k+1 \in S'$, because otherwise $k+1$ would be the smallest element of S. Hence, from strong induction, we have that $S' = \mathbb{N}$, and hence $S = \emptyset$, contradicting the hypothesis $S \neq \emptyset$.

7.5 Good manners with induction proofs

Induction proofs are easy to structure, as there is a clear procedure to follow. But even if the steps in a proof are predictable, they should be spelled out clearly.

1. At the beginning of the proof, say with respect to what variable induction is performed, unless this is obvious.

> *We proceed by induction on the degree d of the polynomial.*

2. If the induction variable is n, then in the inductive step it is considered good practice to go from $n = k$ to $n = k+1$ rather than from n to $n+1$.

3. In the inductive step, say at what point the induction hypothesis is used.

> *. . . where the last inequality follows from the induction hypothesis.*

4. Say when the inductive step is complete.

> *This completes the induction.*

If an induction argument is easy, then it may be tempting to skip it altogether.

> *Let $x_1 = 2$, and let $x_{n+1} = x_n^2$, for $n \geqslant 1$. Then $x_n = 2^{2^{n-1}}$.*

The last sentence is a bit blunt, and it could be replaced by more considerate sentences, such as:

> *A straightforward induction argument shows that $x_n = 2^{2^{n-1}}$.*
>
> *Then $x_2 = 2^2$, $x_3 = 2^{2^2}$, and, in general, $x_n = 2^{2^{n-1}}$.*

Exercises

Exercise 7.1. Prove the following formulae by induction, first analytically, then with a computer-assisted proof.

1. $$\sum_{k=1}^{n}(2k-1)=n^2$$
2. $$\sum_{k=1}^{n}k!\cdot k=(n+1)!-1$$
3. $$\sum_{k=1}^{n}k^3=\left(\sum_{k=1}^{n}k\right)^2$$
4. $$\sum_{k=1}^{n}\frac{1}{k(k+1)}=\frac{n}{n+1}$$
5. $$\sum_{k=2}^{n}\frac{1}{k^2-1}=\frac{3}{4}-\frac{2n+1}{2n(n+1)}$$
6. $$\prod_{k=0}^{n-1}(1+x^{2^k})=\frac{1-x^{2^n}}{1-x} \qquad x\neq 1.$$

Exercise 7.2. Prove by induction all identites and inequalities listed on page 155. (The Bernoulli inequality is proved in section 7.3.)

Exercise 7.3. Prove the following statements by induction.

1. The sum of the cubes of three consecutive natural numbers is a multiple of 9.
2. The sum of the internal angles of a polygon with n sides is $(n-2)\pi$.
3. The negation of a boolean expression with n leading quantifiers is obtained by interchanging quantifiers and negating the predicate.
4. The regions in which the plane is subdivided by n lines may be coloured with just two colours in such a way that no two neighbouring regions have same colour.

Exercise 7.4. Find each formula, and then prove it by induction.

1. The formula for the cardinality of the power set of a finite set.
2. The formula for the higher-order derivatives of the product of two functions.
3. The formula for the derivative of the repeated composition of a function with itself.
4. The formula for the maximum number of regions defined by n lines in the plane.

Exercise 7.5. The proof of the following false statements is faulty. Explain in one sentence what the flaw is.

1. WRONG THEOREM. *For any natural number n, the following holds:*

$$2+4+\cdots+2n=(n-1)(n+2).$$

WRONG PROOF. Assume that $2+4+\cdots+2k=(k-1)(k+2)$ for some $k\in\mathbb{N}$. Then

$$\begin{aligned}2+4+\cdots+2(k+1) &= (2+4+\cdots+2k)+2(k+1)\\ &= (k-1)(k+2)+2(k+1) \quad \text{(by the induction hypothesis)}\\ &= k(k+3)=((k+1)-1)((k+1)+2)\end{aligned}$$

which completes the induction. □

2. WRONG THEOREM. *For any natural number n, the following holds:*

$$\forall i,j\in\mathbb{N},\ (\max(i,j)=n)\Rightarrow(i=j).$$

(Thus any two natural numbers are equal.)

WRONG PROOF. The statement is true for $n=1$: if $\max(i,j)=1$, then, necessarily, $i=j=1$. Now assume the statement to be true for some $k\geqslant 1$, and let i,j be natural numbers such that $\max(i,j)=k+1$. Then $\max(i-1,j-1)=k$, and by the induction hypothesis, we have that $i-1=j-1$. But then $i=j$, and therefore our statement is true for $n=k+1$. This completes the induction. □

3. Wrong theorem. (Pólya[8]) *In any group of n horses, all horses have the same colour.*

Wrong proof. The statement is clearly true for $n = 1$. Assume now that the statement is true for some $n = k \geqslant 1$, and consider an arbitrary collection of $k+1$ horses

$$\underbrace{\bullet\,\bullet\;\cdots\;\bullet\,\bullet}_{k+1}$$

By the induction hypothesis, the first k horses have the same colour, and so do the last k horses.

$$\rlap{$\underbrace{}_{k}$}\bullet\,\overbrace{\bullet\;\cdots\;\bullet\,\bullet}^{k}$$

But then all $k+1$ horses have the same colour as the $k-1$ horses common to the two sets.

$$\bullet\,\underbrace{\bullet\;\cdots\;\bullet}_{k-1}\,\bullet$$

This completes the inductive step. □

---

[8] George Pólya (Hungarian: 1887–1985).

Chapter 8

Existence and definitions

In this chapter we take a closer look at existence statements. These statements assert that there exists a quantity x that has a certain property $\mathscr{P}$. In symbols:

$$\exists x \in X, \quad \mathscr{P}(x)$$

where X is a set and $\mathscr{P}$ is a predicate over X. For example, the statement

The integer 10^9 *is the sum of two primes*

asserts the existence of a pair of prime numbers with a given sum (what are X and $\mathscr{P}$ in this case?).

We shall discuss proofs of existence and the connection between existence and definitions.

8.1 Proofs of existence

To prove that something exists it is not necessary that the object in question be constructed explicitly. For example, Euclid's proof of the infinitude of primes (p. 136) establishes the existence of infinitely many things without producing any of them. By contrast, we proved that Euler's polynomial n^2+n+41 is composite for infinitely many values of n by exhibiting such values explicitly in equation (6.8).

Accordingly, existence proofs are said to be **constructive** if an explicit construction is given, and **non-constructive** if no explicit construction is given. The constructive method may be described as a two-stage process:

(WHAT?) Identify an element x of X.
(WHY?) Show that $\mathscr{P}(x)$ is TRUE.

The format of the proof should make clear which part is the WHAT and which part is the WHY.

We begin with a constructive existence proof:

Theorem. *There is a real number α such that $\alpha^2 = 2$.*

PROOF [29, p. 13]. Draw a square of side 1.
(WHAT?) Let α be the length of a diagonal of the square.
(WHY?) Then by Pythagoras, $\alpha^2 = 2$. $\square$

This minimalist proof takes for granted that the length of a diagonal of the unit square is a real number. A justification of this step requires the tools of analysis.

Let us re-visit the counterexample (6.8) concerning Euler's polynomial.

Theorem. *There are infinitely many integers n for which $p(n) = n^2 + n + 41$ is composite.*

PROOF.
(WHAT?) Let $n = 41k$, for $k \in \mathbb{N}$.
(WHY?) Then $p(41k) = 41(41k^2 + k + 1)$, not a prime. $\square$

In this proof the WHAT part identifies an infinite sequence of integers.

Theorem. *If two integers are the sum of two squares, then so is their product.*

This statement is an implication; the existence statement in the deduction is conditional to another existence statement in the hypothesis. We give a concise constructive proof, where an eloquent formula supplies at once the WHATs and the WHYs.

PROOF. This statement follows from the algebraic identity

$$(j^2 + k^2)(m^2 + n^2) = (jm + kn)^2 + (km - jn)^2. \qquad \square$$

Many non-constructive existence proofs employ contradiction. One assumes that the object being defined does not exist, and derives a false statement. It is not surprising that an argument of this kind could fail to provide information about the object in question.

However, not all non-constructive existence proofs use contradiction.

Theorem. *There is an irrational number a such that $a^{\sqrt{2}}$ is rational.*

PROOF. Consider $\sqrt{2}^{\sqrt{2}}$. We have two cases:
CASE I: $\sqrt{2}^{\sqrt{2}}$ is rational. Then $a = \sqrt{2}$ is as required.

CASE II: $\sqrt{2}^{\sqrt{2}}$ is irrational. Then

$$\left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = \sqrt{2}^{\sqrt{2}\sqrt{2}} = \sqrt{2}^{2} = 2,$$

so $a = \sqrt{2}^{\sqrt{2}}$ is as required. □

This is a proof by cases —see section 6.2. We note that we aren't given a WHAT. There are two different WHATs, and we have no idea which of them works for the WHY.

Our next example is a well-known existence statement, **Dirichlet's box principle** (or the **pigeon-hole principle**).

Theorem. *Given n boxes and m objects in them, if $m > n$, then at least one box contains more than one object.*

The proof of this rather self-evident implication is necessarily non-constructive. We use contradiction, namely the **both ends method** described in section 6.5. It's easy to predict where contradiction will lead us.

PROOF. We use contradiction. Let m and n be integers, and let m_j be the number of objects in the jth box. If $m > n$ and $m_j \leqslant 1$, for $j = 1, \dots, n$, then

$$m = \sum_{j=1}^{n} m_j \leqslant \sum_{j=1}^{n} 1 = n$$

giving $m \leqslant n$, contrary to the assumption. □

In this proof we are not given the WHAT, and hence there cannot be a WHY step either. The proof begins by introducing the symbol m_j; this simple but important step makes the rest immediate.

Dirichlet's principle may seem informal, as it deals with 'objects in boxes' or even 'pigeon-holes'. It is in fact a crisp mathematical statement. For a geometric interpretation, let $\mathbb{N}_0^n = (\mathbb{N} \cup \{0\})^n$ be the set of n-dimensional vectors with non-negative integer components. Dirichlet's principle says that if $m > n$, then every point of the set

$$\Omega_m := \{(m_1, \dots, m_n) \in \mathbb{N}_0^n : \sum_j m_j = m\}$$

lies outside the n-dimensional unit cube in $\mathbb{N}_0^n$.

The non-constructive proof of the following statement employs Dirichlet's box principle.

Theorem. *In any set of integers with more than n elements there must be two integers whose difference is divisible by n.*

PROOF. Let n be given, and let S be a set of m integers, with $n < m$. We divide each element of S by n, obtaining m integer remainders. These remainders can assume at most n distinct values, and therefore two elements of S, say x_i and x_j, must have the same remainder, by Dirichlet's box principle. But then $x_i - x_j$ gives reminder zero when divided by n, as desired. □

Solving equations is a quintessential mathematical task. In some cases we may be looking for a particular number, which means that we are after a constructive existence proof. In other cases we may be satisfied by a proof that a solution exists at all, or that it exists in a specified range of values of the argument. Let us examine a constructive and a non-constructive existence proof of the same statement.

Proposition. *The equation $x^4 - 10x^2 + 1 = 0$ has four distinct real solutions.*

CONSTRUCTIVE PROOF. Let $f(x) = x^4 - 10x^2 + 1$.
(WHAT?) Let $x_\pm = \sqrt{3} \pm \sqrt{2}$; then $x_\pm$ are real numbers.
(WHY?) We compute, using the binomial theorem

$$\begin{aligned} f(x_\pm) &= (\sqrt{3} \pm \sqrt{2})^4 - 10(\sqrt{3} \pm \sqrt{2})^2 + 1 \\ &= 9 \pm 12\sqrt{6} + 36 \pm 8\sqrt{6} + 4 - 30 \mp 20\sqrt{6} - 20 + 1 \\ &= 0 \end{aligned}$$

where the top and bottom signs in ± and ∓ match. The above calculation shows that $x_\pm$ are roots of f. Since the real numbers $x_\pm$ are positive and distinct, and f is an even function, we conclude that $-x_\pm$ are also roots of f. □

This constructive existence proof involved 'guessing' the solutions, and then verifying that the guess was correct. (There was no guess here: I started from x_+, and then determined the equation of which x_+ is a solution.)

NON-CONSTRUCTIVE PROOF. Let $f(x) = x^4 - 10x^2 + 1$. The computation

$$f(0) = 1 \qquad f(1) = -8 \qquad f(4) = 97,$$

shows that f changes sign twice in the interval $(0,4)$. Because f is continuous, there exist two distinct real numbers x_-, x_+ in the open intervals $(0,1)$ and $(1,4)$, respectively, at which the function f vanishes. The real numbers $-x_{\pm}$ are also roots of f, because f is even. The four numbers $\pm x_{\pm}$ are clearly distinct. $\square$

The key argument in the proof was inferring the existence of a root from a change of sign of a continuous function. This is the **intermediate value theorem** from real analysis, which is a non-constructive existence theorem [24, theorem 4.4]:

Theorem. *Any real-valued continuous function f on a closed interval $[a,b]$ assumes every value between $f(a)$ and $f(b)$.*

Our non-constructive proof provides some information about the solutions, in the form of bounds. These bounds can be sharpened by doing some extra work; e.g., with f as above we have

$$f\left(\frac{7}{22}\right) = \frac{-503}{234256} \qquad f\left(\frac{6}{19}\right) = \frac{1657}{130321}$$

and hence

$$\frac{6}{19} < x_{-} < \frac{7}{22}.$$

Considering that $7/22 - 6/19 < 3 \times 10^{-3}$, our knowledge of the solution x_{-} has increased markedly. We see that the distinction between a constructive and a non-constructive proof is blurred.

A non-constructive existence statement that provides some information about an object —typically a number— in the form of bounds, is said to be **effective**[1]. The above non-constructive proof is effective. Likewise, the estimate of the size of the *n*th prime number given by theorem 7.4 (p. 163) may be characterised as an effective version of Euclid's theorem.

In the mathematics literature the meaning of 'constructive proof' is sometimes extended to include methods that, while not providing an actual object, provide an algorithm for constructing the object.

8.2 Unique existence

Unique existence is a stronger version of existence:

> *There is exactly one element of X having property $\mathscr{P}$.*

---

[1] Some authors use 'effective' to mean 'constructive'.

where, as usual, X is a set and $\mathscr{P}$ is a predicate over X. For example:

> *The equation* $f(x) = 0$ *has exactly one solution.*
> *The sets A and B have only one point in common.*
> *The function g has precisely one stationary point.*
> *The matrix M has a unique real eigenvalue greater than 1.*

The expressions 'exactly', 'only', 'precisely' differentiate unique existence from existence.

Unique existence is a hidden conjunction; the conjuncts are

> EXISTENCE: *There is one element of* X *with property* $\mathscr{P}$.
> UNIQUENESS: *If* $x, y \in X$ *have property* $\mathscr{P}$*, then* $x = y$.

The corresponding symbolic expression is overloaded:

$$(\exists x \in X,\ \mathscr{P}(x)) \wedge (\forall x, y \in X,\ (\mathscr{P}(x) \wedge \mathscr{P}(y)) \Rightarrow (x = y)) \tag{8.1}$$

and to simplify it we introduce a new quantifier $\exists!$ denoting unique existence:

$$\exists!\, x \in X,\ \mathscr{P}(x).$$

For example, given a function $f : X \to Y$, the concise expression

$$\forall y \in Y,\ \exists!\, x \in X,\ f(x) = y$$

states that f is bijective.

In chapter 7 we proved (twice) that any integer n greater than 1 is a product of primes. For each n, this statement asserts the existence of an unspecified number of primes. The **fundamental theorem of arithmetic** upgrades existence to unique existence.

Theorem. *Every natural number greater than 1 may be written as a product of prime numbers. This representation is unique, apart from re-arrangements of the factors.*

The expression 'apart from re-arrangements of the factors' gives the impression of lack of uniqueness. This is not the case: the theorem states that there exists a unique **multiset** of prime numbers (see section 2.1) with the property that the product of its elements is the given natural number. Alternatively, there is a unique **set** of pairwise co-prime prime-powers with the same property.

In a proof of unique existence the conjuncts are normally proved separately, in either order.

Theorem. *The identity element of a group is unique.*

The group axioms were given in section 6.1. The existence of an identity element does not require a proof, being an axiom (axiom G3). We only have to prove uniqueness.

Proof. Let us assume that a group has two identity elements, $\diamondsuit_1$ and $\diamondsuit_2$. Applying the axiom G3 to each identity element, we obtain

$$\diamondsuit_1 \odot \diamondsuit_2 = \diamondsuit_1 \qquad \diamondsuit_1 \odot \diamondsuit_2 = \diamondsuit_2$$

and hence $\diamondsuit_1 = \diamondsuit_2$. $\Box$

Theorem. *For every set X and subset Y, there is a unique set Z such that $Y \cup Z = X$ and $Y \cap Z = \emptyset$.*

Proof. Let $Z = X \smallsetminus Y$. By construction, $Y \cap Z = \emptyset$. Since $Y \cup Z$ is a subset of X, and every element of X belongs to either Z or Y, we have $Y \cup Z = X$. So a set Z with the required properties exists.

To prove uniqueness, assume that Z is a set with the stated properties. Then, since $Y \cup Z = X$, Z must be a subset of X, and it must contain all elements of X that aren't in Y. So Z must contain $X \smallsetminus Y$.

Since $Y \cap Z = \emptyset$, the set Z can't contain any elements of X that are in Y. So Z must be $X \smallsetminus Y$. $\Box$

The existence part of this proof is constructive. For this reason, the proof of uniqueness does not establish the second conjunct in (8.1) explicitly. Rather than considering two objects with the stated properties and showing that they are the same, we consider a single object and show that it is the same as the one defined in the existence part.

Theorem (Fermat). *The only natural numbers x for which $1+x+x^2+x^3$ is a square are $x = 1$ and $x = 7$.*

Proving existence is immediate: $1+1+1^2+1^3 = 2^2$, $1+7+7^2+7^3 = 20^2$. Uniqueness is much harder, see [11, pp. 38, 382].

8.3 Definitions

Definitions are closely related to unique existence. When we define a symbol or a name, we must ensure that this quantity actually exists, and that it has the stated properties. A definition may identify a unique quantity (*Let a be the length of a diagonal of a regular pentagon with unit area*), or specify membership to a unique non-empty set (*Let p be a prime such that $p+2$ is also prime*). If these conditions are met, then our object is **well-defined**; otherwise it is **ill-defined**.

Let us begin with definitions of sets. Here there is a distinctive safety net: if the conditions imposed on a set are too restrictive, then this set will be empty rather than ill-defined. Nonetheless, we should be alert to this possibility, because the consequences of an empty definition could be more serious than formally correct nonsense. For example, consider the following definition:

> *Let $\mathscr{M}$ be the set of 2 by 2 integral matrices with odd entries and unit determinant.*

The set $\mathscr{M}$ is empty because the determinant of a matrix with odd entries is an even integer, and so it cannot be 1. It's easy to run into trouble now: *Let $M \in \mathscr{M}$.*

Less trivially, a set defined by several conditions could be the solution set of a system of simultaneous equations, which is the intersection of the solution sets of the individual equations. If an empty intersection is possible, then the definition should flag it:

> *Let $A_1, \ldots, A_n$ be sets, and let A denote their (possibly empty) common intersection*
>
> $$A = \bigcap_{k=1}^{n} A_k = A_1 \cap A_2 \cap \cdots \cap A_n. \tag{8.2}$$

For $n > 2$, the set A is well-defined, because the intersection operator is associative. For $n = 1$, the meaning of formula (8.2) is unspecified, but an experienced reader will interpret it as meaning $A = A_1$. So a separate treatment is not necessary in this case. If $n = 0$ the only possible meaning is $A = \emptyset$. The style of this definition suggests that n must be positive; otherwise a clarifying remark would become necessary.

Next we consider an innocent-looking integer sequence whose existence is more delicate than expected.

> *Let q_n be the smallest prime number with n decimal digits.*

Does such a prime exist for all n? Let's look at the first few terms of the sequence:

$$(q_n)_1^\infty = (2, 11, 101, 1009, 10007, \ldots). \tag{8.3}$$

In all cases, the prime q_n lies just above 10^{n-1}, and it seems unavoidable that there is at least one prime between 10^{n-1} and 10^n.

However, given that there are arbitrarily large gaps between consecutive primes[2], an argument is needed to rule out the possibility that a large gap could include all integers with n digits, for some n. Such an argument is not elementary, and we shouldn't keep the readers pondering on this. So the definition of the sequence (q_n) should incorporate a remark or a footnote of the type:

> *This sequence is well-defined due to* Bertrand's postulate[3]*: for all* $n \geqslant 1$ *there is at least one prime* p *such that* $n < p \leqslant 2n$ *[14, p. 343].*

A function definition $f : A \to B$ requires specifying two sets, the domain A and the co-domain B, as well as a rule that associates to every point $x \in A$ a unique point of $f(x) \in B$. For the function f to be well-defined, we must ensure that the specification of A and of the rule $x \mapsto f(x)$ do not contain any ambiguity.

By contrast, there is flexibility in the specification of the co-domain B, in the sense that any set containing $f(A)$ may serve as a co-domain. Formally, different choices of B correspond to different functions, although such distinctions are often unimportant. But then why don't we always choose $f(A)$ as co-domain? This would have the advantage of making every function surjective. The problem is that we may not know what $f(A)$ is, or the description of $f(A)$ may be exceedingly complicated.

For instance, let us return to the sequence q_n given in (8.3). Writing

$$q_n = 10^{n-1} + a_n \tag{8.4}$$

we find

$$(a_n)_1^\infty = (1, 1, 1, 9, 7, \ldots).$$

Let us now define the function

$$f : \mathbb{N} \to \mathbb{N} \qquad f(n) = a_n.$$

This function is well-defined but we have limited knowledge of the image $f(\mathbb{N})$. It can be shown that $f(\mathbb{N})$ cannot intersect any of the sets $2\mathbb{N}$, $(2+3\mathbb{N})$, and $5\mathbb{N}$, but we don't know if there are other constraints —see exercise 9.3.3 p. 202.

---

[2] For $n > 1$, the integer $1 + n!$ is followed by $n - 1$ composite integers. (Why?)

[3] Conjectured by Joseph Bertrand (French: 1822–1900) in 1845. Proved by Pafnuty Chebyshev (Russian: 1821–1894) in 1850.

8.4 Recursive definitions

Recursion is a method of defining sequences which is connected to the principle of induction. This construct can deliver great complexity from minimal ingredients.

We have pointed out that a sequence (a_k) corresponds to a function over $\mathbb{N}$, whereby a_k represents the value of the function at the point $k \in \mathbb{N}$. The existence of such a function does not necessarily imply that its values are easily recoverable from the definition of the sequence. For example, we don't know any 'useful' way to express the kth prime number as an explicit function of k. In absence of an explicit formula for the elements of a sequence, we seek to represent a_{k+1} in terms of a_k. These are **recursive sequences**, which are defined by two data: the first term of the sequence, called the **initial condition**, and the rule which determines a term from the previous one.

For instance, the **factorial sequence** $n! = 1 \cdot 2 \cdot 3 \cdots n$, can be defined recursively as follows:

$$0! = 1 \qquad (k+1)! = (k+1) \cdot k! \qquad k \geqslant 0. \tag{8.5}$$

This definition, in effect, specifies the order in which the product is computed. Likewise, the nth derivative $g^{(n)}$ of a function g is defined by the recursive rule

$$g^{(0)}(x) = g(x) \qquad g^{(k+1)}(x) = \frac{d}{dx} g^{(k)}(x) \qquad k \geqslant 0.$$

More generally, given any set X and any function $f : X \to X$, we define a recursive sequence (a_k) of elements of X as follows:

$$a_1 \in X \quad \text{given}; \qquad a_{k+1} = f(a_k), \quad k \geqslant 1. \tag{8.6}$$

The first term of the sequence (the initial condition) is given. Because f is a function, if a_k is well-defined for some $k \geqslant 1$, then so is a_{k+1}. The induction principle then ensures that the whole sequence is well-defined. Since each initial condition in X gives rise to a sequence, we get lots of sequences from just one function!

The term a_{k+1} of a recursive sequence may depend on $d \geqslant 1$ preceding terms, not just the last one:

$$a_{k+1} = f(a_k, a_{k-1}, \ldots, a_{k-d+1}). \tag{8.7}$$

In this case we have $f : X^d \to X$, and d is called the **order** of the sequence. In a recursive sequence of order d, the first d terms of the sequence must be supplied explicitly because they don't have the required number of preceding terms. These sequences have d initial conditions and they are well-defined by virtue of **strong induction**.

The **Fibonacci sequence**[4] is a second-order sequence:

$$a_0 = 1; \qquad a_1 = 1; \qquad a_{k+1} = f(a_k, a_{k-1}) = a_k + a_{k-1}, \quad k \geqslant 1.$$

We find

$$(1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \ldots).$$

Because the term $a_{k-1} = a_{k+1} - a_k$ is a well-defined function of the following two terms, we can extend the Fibonacci sequence backwards to obtain a doubly-infinite sequence:

$$(\ldots, -21, 13, -8, 5, -3, 2, -1, 1, 0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots).$$

A recursive sequence that can be extended backwards is said to be **invertible**. Thus the Fibonacci sequence is invertible.

The behaviour of a recursive sequence may be quite unpredictable, as the following example illustrates. We define the integer function

$$f : \mathbb{N} \to \mathbb{N} \qquad\qquad x \mapsto \begin{cases} x/2 & \text{if } x \text{ is even} \\ 3x+1 & \text{if } x \text{ is odd} \end{cases}$$

and then we consider the associated first-order recursive sequence:

$$x_0 = n \qquad x_{t+1} = f(x_t), \quad t \geqslant 0. \tag{8.8}$$

The initial condition $x_0 = 1$ leads to a **periodic sequence**

$$(1, 4, 2, 1, 4, 2, 1, \ldots) \tag{8.9}$$

consisting of indefinite repetitions of the pattern $(1, 4, 2)$. If instead we start with $x_0 = 7$, we find

$$(7, 22, 11, 34, 17, 52, 26, 13, 40, 20, 10, 5, 16, 8, 4, 2, 1, 4, 2, 1, \ldots)$$

so, after an irregular initial excursion, the sequence settles down to the same periodic pattern as the previous sequence. The analysis of the sequences (8.8) for general initial condition n presents formidable difficulties, which are distilled into a famous unsolved problem, the so-called **'$3x+1$' conjecture** [41]:

Conjecture. *For any n, the sequence (8.8) contains 1.*

This conjecture states that all these sequences are **eventually periodic** and that they reach the same periodic pattern.

We remark that any open conjecture leads to objects whose existence is at present undecidable. We explore this issue in exercises 8.4 and 9.3.

---

[4] Leonardo Pisano, known as Fibonacci (Italian: 1170–ca.1240).

8.5 Wrong definitions

As we did in section 6.7 for logical arguments, we consider here some common mistakes made in definitions. A faulty definition may imply that the object being defined does not exist at all, or that it is not the one we had in mind, or that there is more than one object that fits the description.

We begin with an incorrect function definition which can be put right in several ways.

WRONG DEFINITION. *Let f be given by:*

$$f : \mathbb{Q} \to \mathbb{Q} \qquad f(x) = x^2y + 1.$$

The mistake is plain but fatal: the expression $f(x)$ involves the unspecified quantity y. As defined, f is not a function. There are many legitimate interpretations of what this formula could mean, each resulting in a different function.

1. We re-define the domain, supplying the missing information via a second argument.

$$f : \mathbb{Q}^2 \to \mathbb{Q} \qquad f(x,y) = x^2y + 1.$$

In this definition y is also rational; clearly there are other possibilities.

2. The missing argument is regarded as a **parameter** (see section 5.1).

$$f_\lambda : \mathbb{Q} \to \mathbb{Q} \qquad f_\lambda(x) = \lambda x^2 + 1 \qquad \lambda \in \mathbb{Q}.$$

We have highlighted the change in status of the variable y by switching to the Greek alphabet, and turning it into a subscript. For every value of λ, we have a different function.

3. The missing argument is regarded as an **indeterminate.** In this case $f(x)$ is a polynomial in y, and we must re-define the co-domain of f accordingly.

$$f : \mathbb{Q} \to \mathbb{Q}[y] \qquad f(r) = r^2y + 1.$$

The symbol $\mathbb{Q}[y]$ denotes the set of all polynomials in the indeterminate y with rational coefficients (see section 2.4.1). To avoid confusion, we have replaced x with a symbol that normally is not used for an indeterminate, and which reminds us that this variable is rational.

In the second example, the existence of a function requires more conditions than those given. The definition relies on a hidden assumption, so that using it adds extra information (we say that the definition is *creative*).

WRONG DEFINITION. *Let A and B be sets, and let $f,g : A \to B$ be functions. We define the function $h = f + g$ as follows:*

$$h : A \to B \qquad x \mapsto f(x) + g(x).$$

The hidden assumption is that the elements of the co-domain can be added together. However, B may be a set where addition is not defined (e.g., f and g are predicates) or which is not closed under addition (e.g., B is an interval). We give a more restrictive definition without hidden assumptions.

DEFINITION. *Let A be a set, let B be a set closed under addition, and let $f,g : A \to B$ be functions. . . .*

In the next example we attempt to extend to modular arithmetic the concept of the reciprocal of an element, but our definition has a hidden assumption.

WRONG DEFINITION. *Let m be a natural number, and let a be an integer. If $[a]_m \neq [0]_m$, then we define the multiplicative inverse $[a]_m^{-1}$ as follows:*

$$[a]_m^{-1} \overset{\text{def}}{=} [b]_m \quad \text{where} \quad [a]_m[b]_m = [1]_m.$$

Take $m = 6$ and $a = 3$. We have $[3]_6 \neq [0]_6$, but the equation $[3]_6[x]_6 = [1]_6$ has no solution. What went wrong? We have assumed that the implication

$$\text{if } a \neq 0 \text{ and } ab = ac, \text{ then } b = c,$$

which is valid for real or complex numbers, is also valid for congruence classes. Requiring that $[a]_m \neq [0]_m$, namely that a is not divisible by the modulus is not enough; for this implication to hold, we must assume that a is co-prime to the modulus.

DEFINITION. *Let m be a natural number and let a be an integer co-prime to m. We define the multiplicative inverse $[a]_m^{-1}$ of $[a]_m$ as follows. . .*

In the following definition both existence and uniqueness are problematic.

WRONG DEFINITION. *Let $p \in \mathbb{Z}[x]$, and let z_p be the root of p having smallest modulus.*

If p has degree zero, then z_p does not exist, so this definition relies on the hidden assumption that p has positive degree. If z_p exists, then it may still not be unique (e.g., $p = x^2 + 1$), but the use of the definite article ('the root') implies that it is. Requiring that p be non-constant solves the existence problem. The way we address uniqueness depends on the context. If we merely require that no root of p has smaller modulus than z_p, then we don't need uniqueness.

DEFINITION. *Let $p \in \mathbb{Z}[x]$ with $\deg(p) > 0$, and let z_p be a root of p with smallest modulus.*

This definition does not identify z_p; it tells us that z_p is a member of a well-defined non-empty set, but this set may have more than one element. If, on the other hand, we require a specific complex number, then we must supply additional information. In the following definition we impose a particular condition on the argument of z_p, making sure that there is no ambiguity.

DEFINITION. *Let p be a non-constant polynomial with integer coefficients, and let z_p be the root of p with smallest modulus; if there is more that one root with this property, then let z_p be the one with smallest argument (with $2\pi < \arg(z) \leqslant 0$).*

In our final example, the definition hides a subtle lack of uniqueness.

WRONG DEFINITION. *Let $x \in [0,1]$, and let f be the function that associates to x its binary digits sequence*

$$f : [0,1] \to \{0,1\}^{\mathbb{N}} \qquad x \mapsto (c_1, c_2, \ldots),$$

where

$$x = \sum_{k=1}^{\infty} \frac{c_k}{2^k} \qquad c_k \in \{0,1\}. \tag{8.10}$$

(The co-domain of f is the set of all infinite binary sequences, see section 5.2.) What's wrong with this definition? Consider the rational numbers

$$x_n = \frac{1}{2^n} = \sum_{k=1}^{\infty} \frac{1}{2^{k+n}} \qquad n \geqslant 1.$$

If we write x_n in the form (8.10), then the above identity shows that there are two binary representations, namely

$$(\underbrace{0, \cdots, 0}_{n-1}, 1, 0, 0, 0, \ldots) \qquad \text{and} \qquad (\underbrace{0, \cdots, 0}_{n-1}, 0, 1, 1, 1, \ldots). \tag{8.11}$$

So the function f is not uniquely defined at these rationals and, more generally, at all rationals of the form

$$\sum_{k=1}^{n-1} \frac{c_k}{2^k} + x_n$$

where $c_1, \ldots, c_{n-1}$ are arbitrary binary digits.

We resolve this ambiguity by choosing consistently the first of the two digit sequences in (8.11).

DEFINITION. *Let $x \in [0,1)$ be given by*

$$x = \sum_{k=1}^{\infty} \frac{c_k}{2^k} \qquad c_k \in \{0,1\}.$$

Without loss of generality, we assume that all sequences (c_k) contain infinitely many zeros. Let now f be the function that associates to x its binary digit sequence

$$f : [0,1) \to \{0,1\}^{\mathbb{N}} \qquad x \mapsto (c_1, c_2, \ldots).$$

Exercises

Exercise 8.1. Use Dirichlet's box principle to prove the following statements.

1. Given any five points in the unit square, there are two points whose distance apart is at most $1/\sqrt{2}$.
2. In any finite group of people, there are two people with the same number of friends.
3. Every recursive sequence of elements of a finite set is eventually periodic.

Exercise 8.2. Each function definition contains an error. Explain what is the error, and how it should be corrected.

1. $f : \mathbb{R} \to \mathbb{R} \qquad x \mapsto \dfrac{1}{x^2 + x - 1}$
2. $f : \mathbb{R} \to \mathbb{R} \qquad x \mapsto \sqrt{x^2 - 1}$
3. $f : \mathbb{Q} \to \mathbb{Z} \qquad \dfrac{a}{b} \mapsto |a - b|$
4. $f : \mathbb{Z} \to \mathbb{Z} \qquad n \mapsto n\mathbb{Z} \cap (n^2 + 1)\mathbb{Z}$.

Exercise 8.3. The following definitions have several flaws. Explain what they are, hence write a correct, clearer definition.

1. *Let $a = (b_1, b_2, \ldots)$ be a given sequence of elements of B. We define the function*
$$f : \mathbb{Z} \to B \qquad f(m) = \sum_{k=1}^{m} b_k^{-1}.$$

2. *Let X be a subset of $\mathbb{R}$, and let $f(X)$ be the number of integers in X, namely*
$$f : \mathbb{R} \to \mathbb{N} \qquad\qquad f(X) = \#(x \in X \cap x \in \mathbb{Z}).$$

Exercise 8.4. Use each conjecture to define a function whose existence cannot be decided at present.

1. The twin-primes conjecture (section 6.6).
2. The Goldbach conjecture (section 6.6).
3. The $3x+1$ conjecture (section 8.4).

Chapter 9

Writing a thesis

The highlight of undergraduate mathematical writing is the report on a final year project, called a **thesis** or **dissertation**. This document has a distinctive structure, in between a short book and a research paper. A thesis surveys a body of advanced literature and presents original work. At undergraduate level, usually the former outweighs the latter.

Writing a thesis is an irreplaceable experience in university education. This document can reveal a great deal about the author's knowledge, understanding, curiosity, and enthusiasm. Examination scripts do not convey so much information.

One doesn't learn how to write a thesis by reading an instruction manual. The thesis will take shape gradually, as a result of regular interactions between student and supervisor. The document's high-level organisation —number of chapters, outline of their content— is normally considered at a relatively early stage of the project. Planning will also suggest appropriate headings and will aid the sequencing of the arguments. Periodic reviews of work in progress may alter priorities, or even re-direct the research. The writing process culminates with the exercise of proof-reading, in itself a valuable, if painstaking, experience.

In this chapter I give an overview of the structure of a thesis, and then discuss selected aspects of writing —choosing the title, writing the abstract, compiling the bibliography— where analysing examples may be helpful. I also include a brief introduction to LaTeX, the software adopted by the mathematical publishing industry. Most theses are now typeset in LaTeX.

9.1 Theses and other publications

A thesis is a substantial document. Its length varies considerably, depending on level, topic, and institutional requirements. An undergraduate thesis could be some fifty pages long. There are also theses at master and doctoral level, called MSc and PhD theses respectively[1].

A good undergraduate thesis may rise to the level of a master thesis. A doctoral dissertation is longer, has greater depth, and more original research. The brief guidelines given in this chapter apply to any thesis. For further reading, see Higham's book [18], which devotes two chapters to PhD theses and has an extended bibliography.

A thesis is subdivided into heading, body, and closing matters. The heading supplies essential information: title, abstract, table of contents, acknowledgements. One may also find lists of mathematical symbols, figures, and tables, which some institutions require explicitly. Title and abstract define the subject matter; we'll consider them in some detail in sections 9.2 and 9.3. LaTeX will take care of the menial task of compiling the table of contents. The acknowledgements is a paragraph where the author thanks people for their guidance and support (the supervisor, typically), for useful discussions, for pointing out things, or for allowing the author to use their text in the thesis.

The body contains the bulk of the material, organised into chapters, sections, and, if appropriate, sub-sections. The first chapter is the introduction, written so as to make the thesis self-contained. Here we find motivations, background material and key references, enough details to be able to grasp the main results, and a description of the content of the rest of the thesis.

The length of individual chapters may vary, but each chapter must be a substantial component of the whole, with a clear identity. A brief closing chapter is desirable if not compulsory. This is an opportunity to re-visit ideas and results from an informed perspective, discuss limitations of the work, identify open problems, and chart directions of future research.

Any accessory material which should not interfere with the main text (data tables, computer programs, tedious proofs, etc.) is confined to appendices. The last item in a thesis is the bibliography, listing the sources cited in the text; it'll be considered in section 9.4.

There is a symbiotic relationship between a thesis and a research paper; the higher the level of the thesis, the stronger the bond. At doctoral level, research papers lead to a thesis, or are extracted from it after the thesis is completed. At undergraduate level this connection is necessarily more remote, yet a research element must be present in every respectable thesis.

---

[1] Acronyms for Master of Science and Philosophy Doctor.

But a thesis is not a research paper. A research paper communicates new results to specialists, invariably within tight space constraints. It is seldom self-contained: sketchy motivations, unexplained jargon, and terse citations are commonplace. A thesis and a paper also differ in format. A paper is shorter, divided into sections, not chapters, while its simpler navigation renders a table of contents unnecessary.

A thesis is not a book either. A book covers thoroughly a body of established material. It is longer than a thesis, and we expect it to be wholly self-contained. A book begins with a preface, which is neither an abstract nor an introduction, but a bit of both. It is rare to find a book without exercises in each chapter, and even rarer one without a subject index at the end. (Notably, the celebrated book [14] has neither.) A textbook, being conceived as teaching material, may offer solutions or hints to exercises. This is a modern trend: hardly any of the classic, timeless mathematics textbooks provided solutions or hints to exercises —a fact worth pondering on. Some advanced publications resemble theses: the *review articles*, published in dedicated research journals, and the *research monographs*, published as books, usually within a collection. These are a sort of advanced theses, which collect together previously published works. They usually appear when a new research area has reached maturity.

Most scholarly publications are specialised; they are not intended for the general public. So how narrow should the target audience of a thesis be? This is largely a question of personal taste. Conscious decisions must be made from the very beginning, because the title, the abstract, and the first paragraphs of the introduction will set the tone for the rest of the document.

Let us inspect the opening sentences of some theses and research papers, to see how strongly the character of the work is established with just few words. Later on, we shall analyse titles and abstracts from the same perspective.

Our first example is a master thesis (from [33], with minor editing).

> Let $GL(d,q)$ denote the general linear group of invertible $d \times d$-matrices over the finite field $\mathbb{F}_q$ with q elements. In computational group theory, it is of interest to calculate the order n of an element $A \in GL(d,q)$, i.e.,
>
> $$n = \min\{j \in \mathbb{N} : A^j = I\}$$
>
> where I is the $d \times d$ identity matrix. The integer d can be large, e.g., $d > 100$.

The style is formal, with definitions and symbols appearing straight away. The jargon (general linear group, finite field, order of a matrix, etc.) and the corresponding notation ($GL(d,q)$, $\mathbb{F}_q$, etc.) belong to standard introductory algebra. The author still provides brief explanations, which convey an impression of helpful, considerate writing. This document promises to be accessible to any mathematically mature reader, a desirable feature of any undergraduate thesis.

Our next example, from a thesis in geometry [15], illustrates a completely different expository strategy.

> In 1984, Schechtman *et. al.* [SGBC84] announced that the symmetry group of an aluminium-manganese alloy crystal, produced by rapid cooling, was that of the icosahedron. Such a symmetry is not possible for a periodic structure in three dimensions. This discovery brought down a long-held assumption in crystallography, that the only structures with some sense of long-range order were periodic.

The thesis begins with an account of the surprising outcome of a physics experiment. (We shall learn about literature citation in section 9.4.) The author tells a story, skillfully building some drama. The near complete absence of mathematical terms —formal definitions and symbols are given much later— makes this excerpt very inviting. An engaging, accessible style is welcome in any publication, at any level.

A thesis can be highly specialised. One could use the established dictionary of a particular research area, without explanations or apologies, from the very beginning. This choice undoubtedly simplifies the writer's task, but it creates problems for the readers, thereby reducing the readership basis. This is legitimate, as long as the jargon is used wisely to sharpen the exposition, not gratuitously to impress the reader.

The following opening passage is taken from a PhD thesis in the area of *ergodic theory* [3].

> Ergodic optimisation is a branch of the ergodic theory of topological dynamics which is concerned with the study of T-invariant probability measures, and ergodic averages of real-valued functions f defined on the phase space of a dynamical systems $T : X \to X$, where X is a compact metric space.

This sentence explains the meaning of the expression 'ergodic optimisation', which features in the thesis' title. To achieve this within a limited space, the author relies on the reader's familiarity with half a dozen advanced concepts. Three important symbols are also defined in the same sentence. Undeniably, this thesis is for specialists.

In research papers, extremes of specialisation are accepted, or at least tolerated. Some editors insist —wisely, in my view— that the author provide clear motivations and a minimum of background, but this requirement is by no means universal. Papers that begin with a dry, unmotivated list of definitions are not uncommon. The following opening paragraph of standard definitions appears unchanged in three related publications. [26, 27, 28].

> Let K be an algebraic extension of $\mathbb{Q}_p$, and let $\mathscr{O}$ be its integer ring with maximal ideal $\mathscr{M}$, and residue field k. If $\overline{K}$ is and algebraic closure of K, we denote by $\overline{\mathscr{O}}$ and $\overline{\mathscr{M}}$ be the integral closure of $\mathscr{O}$ in K and the maximal ideal of $\overline{\mathscr{O}}$, respectively.

In the following extreme example, the author disposes of definitions altogether [6].

> 1. **Introduction.** The notation used is that of [2]; in particular ...

One cannot proceed without the cited publication (by the same author). Only a committed reader will accept such a blunt treatment.

9.2 Title

The title provides an essential description of the content of a document. The titles of research papers compete for attention with other titles in journals, bibliographical lists, databases, etc. They help the reader decide whether to read on. A thesis has a captive readership —the examiners— so this problem does not arise immediately. Nonetheless, a good title can do a lot to give the thesis its unique identity, so it's worth spending some time thinking about it.

A title is a short phrase without symbols. To get an idea of the wide range of possibilities, we now examine titles of various types of publications, and assess the suitability of the title for a thesis, possibly after amendments. Alongside each title we display the bibliographical data —see section 9.4.

Introductory real analysis [A. N. Kolmogorov and S. V. Fomin, Dover Publications, New York (1970).] The term 'introductory' immediately calls to mind a textbook, so this term is out of place in a thesis' title. The shorter title *Real analysis* would be much too general for a thesis, as it defines a vast area of mathematics.

Making transcendence transparent. (Subtitle: *An intuitive approach to classical transcendental number theory*.) [E. Burger and R. Tubbs, Springer-Verlag, New York (2004).] Subtitles can be quite effective. In this example, the title has an appealing alliteration, while the subtitle spells out the title's promise. However, the title's pedagogical slant and broad take are again the signature of a textbook, not a thesis.

Algebraic aspects of cryptography [N. Koblitz, Springer-Verlag, Berlin (1998).] The subject area covered by this textbook appears to be narrower than in the previous examples, whereas in fact it is still rather broad. If this title were used in a thesis, an expert would probably expect a literature survey. A thesis with an original research component would require a more specific title.

The arithmetic-geometric mean of Gauss. [T. Gilmore, MSc thesis, Queen Mary, University of London (2012).] As the reference to Gauss suggests, this topic is classical, and an expert will know that the arithmetic-geometric mean is connected to many deep mathematical theories. So this is a good title for a thesis, even if there is no indication of what might be the original contribution.

On the zeros of polynomials with restricted coefficients [P. Borwein and T. Erdelyi, *Illinois J. Math.* **41** (1997) 667–675.] This is a rare instance of a research problem with a concise and non-technical description, which makes a good title. The preposition 'on' lends an authoritative tone to it, promising some general result.

Toward a theory of Pollard's ρ-method [Eric Bach, *J. Information and Computation* **90** (1991) 139–155.] A very eloquent title. It suggests that the method being considered was not supported by a theory, and that some progress has been made. This would be an ideal title for a thesis. The use of the symbol ρ is justified, because the aforementioned method is invariably associated to it.

A complete determination of the complex quadratic fields of class number one [H. M. Stark, *Michigan Math. J.* **14** (1967) 1–27.] The forceful, unambiguous title of this well-known publication announces the solution of a long-standing problem. Such an assertive style suits well a research paper with a strong result; the trimmed version *Quadratic fields of class number one* could be the title of a thesis.

Diophantine integrability [R. G. Halburd, *J. Phys A: Math. Gen.* **38** (2005) 1–7.] This title is short and effective. Each term has an established meaning in a specific area of mathematics, and the unexpected juxtaposition of the two terms suggests interdisciplinarity. This title would be perfect for a thesis.

Outer billiards on kites [R. E. Schwartz, volume 171 of *Annals of Mathematics Studies*, Princeton University Press, Princeton and Oxford (2009).] The title of this research monograph cleverly exploits double meanings. Both terms 'Outer billiards' and 'kites' are mathematical, but not many readers will be familiar with them. The unlikely significance that this phrase would have in ordinary English is bound to raise interest among a general mathematical audience.

Euclidean algorithms are Gaussian [V. Baladi and B. Vallée, *Journal of Number Theory* **110** (2005) 331–386.] Summarising a result with a sentence makes an effective title for a research article, but perhaps less so for a thesis. The variant *Probabilistic aspects of Euclidean algorithms* is not so specific, and would work better for a thesis.

Metric number theory: the good and the bad. [R. E. Thorn, PhD thesis, Queen Mary, University of London (2004).] An excellent thesis title, enlivened by a colon. After announcing the research area (metric number theory), a clever reference to an epic Western provides a hint to the specific topic (badly approximable numbers).

To persist or not to persist? [J. Holfbauer and S J Schreiber, *Nonlinearity* **17** (2004) 1393–1406.] This paper is about 'uniformly persistent vector fields', and the author has transformed the subject matter into a Shakespearean question. Such a self-conscious title must be supported by strong content.

It is easy to determine whether a given integer is prime, [Andrew Granville *Bulletin of the American Math. Soc.* **42** (2005) 3–38.] Another title-sentence, with the added bonus of being provocative, because primality testing is a notoriously difficult problem. A fair deal of self-confidence is needed to choose a title of this kind. (This paper won the author the Chauvenet prize, awarded by the Mathematical Association of America to the author of an outstanding expository article on a mathematical topic.)

9.3 Abstract

An abstract is a short summary of a document. If well-written, it complements the title and sustains the reader's attention. The abstract of a research paper is placed immediately after the title, on the front page. In a thesis, where space constraints are not so severe, the abstract usually appears on a separate page.

It is difficult to compose a convincing abstract before the work has been completed, understood, and written out. So the abstract —the first thing one reads— is typically the last thing to be written. The main results must be in evidence, expressed with precision and clarity; at the same time, specialised jargon should be reduced to a minimum, so as not to alienate potential readers. A delicate balance must be achieved between these conflicting requirements.

Abstracts are short, rarely exceeding 200 words. The use of symbols is to be reduced to a minimum, the space is confined and every word counts. The exercise of writing MICRO-ESSAYS (see section 5.7) is good training for writing abstracts, as both tasks require similar rigour and discipline.

We now examine in detail abstracts taken from mathematical literature. They are less than optimal, and we seek to improve them. (The original text has been edited, mainly to protect anonymity.) We don't intend to produce templates of well-written abstracts —this would be futile, as no two abstracts are the same. Rather, we want to sharpen our analytical skills, learn how to spot problems and find ways to solve them. For this purpose, understanding the mathematics is not essential. Indeed, dealing with some unknown words and symbols has a certain advantage, as one must pay attention to the writing's internal structure. We have met a similar situation in chapter 6, when we learnt how to organise the beginning of a proof.

Our first example, from a research paper, exemplifies a rather common problem: the use of unnecessary symbols.

> ABSTRACT. Let F be a rational map of degree $n \geqslant 2$ of the Riemann sphere $\overline{\mathbb{C}}$. We develop a theory of equilibrium states for the class of Hölder continuous functions f for which the pressure is larger than $\sup f$. We show that there exists a unique conformal measure (reference measure) and a unique equilibrium state, which is equivalent to the conformal measure with a positive continuous density.

The fault is plain: the symbols F, n, and $\overline{\mathbb{C}}$ are introduced, but not used again in the abstract, breaking the first golden rule of symbol usage —see section 5.2. Unused symbols distract rather than help. Even if these symbols are used again in the main text, the abstract is not the place where the notation is to be established, and these definitions should be deferred. On this account, the symbol f in the second sentence is used appropriately. Embedded in the second sentence we find the abbreviation 'sup', for 'supremum'. This abbreviation (much like 'max', for maximum, or 'lim', for limit) is meant for formulae, and should not be mixed with words. So we replace it by the full expression.

Re-writing the first sentence without symbols and amending the second sentence do not require any understanding of the mathematics. We leave the third sentence as it is.

> ABSTRACT. We consider rational maps of the Riemann sphere, of degree greater than one. We develop a theory of equilibrium states for the class of Hölder continuous functions f for which the pressure is larger than the supremum of f. We show that . . .

The symbol f in the second sentence is not strictly necessary, and can easily be disposed of.

> We develop a theory of equilibrium states for the class of Hölder continuous functions for which the pressure is larger than the supremum of the function.

The next abstract belongs to a thesis. The author presents an accessible problem using limited jargon.

> ABSTRACT. The logistic map is a well-studied map of the unit interval into itself. However, if we treat x as a discrete variable, as is done in any computer, then every orbit is eventually periodic. Thus the aperiodic behaviour that the continuous map displays for some value of the parameter r cannot be obtained from computer simulation. We investigated the differences and the similarities between the dynamics of a continuous map and its discrete approximation. We found that the limit cycles of a discrete map follow the unstable periodic orbits of the corresponding continuous map.

There are unknown words (logistic map, orbit, limit cycle), but also sufficiently many familiar terms (map, unit interval, eventually periodic, etc.) to help us discern the subject matter. The first sentence says that this work is about the 'logistic map', presented as a well-known object of investigation. The second and third sentences motivate the study. The author introduces the symbol x, but we are not told what it is. We guess that it must represent the argument of the logistic map, namely a point in the unit interval. Furthermore, this symbol is not used again in the abstract, so it is unnecessary. The symbol r suffers from similar problems. It's a parameter, but we are not told which quantity depends on it. We guess that this is the logistic map, which is referred to as the 'continuous map', presumably meaning that its argument assumes a continuum, rather than a discrete set, of values. The symbol r is not used again either.

With two important ideas not in sharp focus, the main message of the thesis —if we discretise the domain of the logistic map, then we obtain new phenomena worth studying— gets a bit lost.

The closing sentence, which summarises the main findings, is vague. What does 'we found' mean? Presumably this is not a proof, otherwise the author would have stated that clearly. What does 'follow' mean? Presumably it indicates some form of convergence. Even if in the abstract there isn't enough room for a complete statement of the results, we must find a way to render these vague statements acceptable.

Finally, the author employs the past tense (we investigated, we found). In an abstract, the present tense is more common. Based on the above considerations, we rewrite the abstract as follows.

> ABSTRACT. We consider the logistic map, a well-studied map of the unit interval which depends on a parameter. If the domain of this map is discretised, as happens in any computer simulation, then, necessarily, all orbits become eventually periodic. Thus the aperiodic orbits observed for certain parameter values no longer exist. We investigate differences and similarities between the original map and its discrete approximations. We provide evidence that the limit cycles of the discrete map converge to the unstable orbits of the original map, in a sense to be made precise.

The last expression states that the thesis contains experimental data or heuristic arguments that support a clearly formulated notion of convergence. (We hope that this is the case in the present thesis!)

Next we consider the opening sentences of an abstract of a research paper. The problems are more subtle here: there are no superfluous symbols, and we can't rely on an understanding of the mathematics!

> ABSTRACT. In this work we investigate properties of minimal solutions of multi-dimensional discrete periodic variational problems. A well-known one-dimensional example of such a problem is the Frenkel-Korontova model. We select a family . . .

The opening sentence begins with the expression 'in this work'; this is redundant and should simply be deleted. Expressions of this type serve a purpose only if the present work is being juxtaposed to other works: 'In 1964, Milnor proved an estimate. . . In this work, we prove. . . '

The object of the investigation is described by a long string of attributes: 'multi-dimensional discrete periodic variational problem'. Such a flat arrangement of words requires a pause or a highlight. At this stage however, it's not clear how this should be done, so we read on.

The second sentence begins with an expression of the type 'an example of this is that'. We have already criticised this format in section 5.6. As written, the emphasis is placed on the word 'example', but the terms 'one-dimensional' or 'Frenkel-Korontova model' surely are more significant. The author's intentions are now clearer: the term 'multi-dimensional' in the first sentence is contrasted with ‘one-dimensional’ in the second sentence, and a known example is given for clarification. To get across the message that this work generalises ‘one-dimensional’ to ’multi-dimensional’, we shall isolate the latter in the first sentence and emphasise the former in the second.

> ABSTRACT. We investigate properties of minimal solutions of discrete periodic variational problems, in the multi-dimensional case. These generalise one-dimensional problems, such as the well-known Frenkel-Korontova model. We select a family . . .

In our final abstract some symbols are necessary.

> ABSTRACT. Let $f : X \to X, X = [0,1)$, be an IET (interval exchange transformation) ergodic with respect to the Lebesgue measure on X. Let $f_t : X_t \to X_t$ be the IET obtained by inducing f to $X_t = [0,t)$, $0 < t < 1$. We show that
>
> $$X_{wm} = \{0 < t < 1 \,:\, f_t \ \text{ is weakly mixing}\}$$
>
> is a residual subset of X of full Lebesgue measure. The result is proved by establishing a Diophantine sufficient condition on t for f_t to be weakly mixing.

The overall content of the abstract is unclear, although we recognise several words introduced in chapters 2 and 3. These fragments of understanding are sufficient to clarify the logic of the argument.

The set X is the unit interval, and we have a function f of this set into itself. This function belongs to a certain class of functions, identified by the acronym IET. The author assumes that f has certain properties, specified by some jargon. Then a parameter t is introduced, and the same construction is repeated for a one-parameter family of functions f_t over sub-intervals X_t. The functions f_t, constructed from f via a process we need not worry about, are also of type IET. (Evidently the author believes that this assertion does not require justification.)

Using a Zermelo definition, the author introduces a subset X_{wm} of the open unit interval, determined by a property of f_t, called ‘weak mixing’. The subscript *wm* in the symbol X_{wm} must be an abbreviation for this expression. The paper establishes some properties of the set X_{wm}. An indication of the argument used in the proof is then given.

Our first task is to remove some symbols, including the displayed equation. This isn't too difficult; for example, the symbol X_{wm} is never used. The first two sentences are formal definitions, which need to be made more readable. The acronym IET is useful, since the long expression it represents is used twice. However, the primary object is 'interval exchange transformation', not IET, so the latter should follow the former, not precede it. Finally, in the expression 'Diophantine sufficient condition' the two adjectives compete with each other, so we separate them out.

> ABSTRACT. Let f be an interval exchange transformation (IET) of the unit interval, ergodic with respect to the Lebesgue measure, and let f_t be the IET obtained by inducing f on the sub-interval $[0,t)$, with $0 < t < 1$. We show that the set of values of t for which f_t is weakly mixing is a residual subset of full Lebesgue measure. The result is proved by establishing a Diophantine condition on t, which is sufficient for weak mixing.

More examples of abstracts in need of improvement are given as exercises.

9.4 Citations and bibliography

Scholarly writing invariably requires reference to other sources. This is carried out in two stages. First, the source is mentioned in the text in an abbreviated manner, so as not to interfere with the reading. Second, a full ordered list of all cited sources (books, journal papers, web pages, etc.) is presented at the end of the document, in a section headed 'Bibliography', or 'References'.

There are some conventions for structuring citations and for writing the bibliography. We begin with a passage involving references to books.

> We briefly review some properties of p-adic numbers, introduced by Hensel in 1908 [17]. Background material may be found in Gouvea's (2000) [13]; the arithmetical side of the theory is developed in Hasse (1980) [16, chapter 7].

Each item enclosed in square brackets is a citation, which refers to a bibliographical entry. Citations should not intrude upon the text, and a sentence in which all citations have been removed should still read well. The citations [13] and [17] refer to books as a whole, whereas [16] refers to a specific chapter in a book. The targeted item within a book may also be a theorem ([14, theorem 191]), a page ([14, p 156]), a range of pages ([14, pp 156–158], [14, 156*ff*][2]), etc. The bibliographical data corresponding to the text above are organised as follows.

> [13] F. Q. Gouvêa, *p-adic numbers*, Springer-Verlag, Berlin (2000).
> ⋮
> [16] H. Hasse, *Number theory*, Springer-Verlag, Berlin (1980).
> [17] K. Hensel, *Theorie der Algebraischen Zhalen*, Teubner, Leipzig and Berlin (1908).

The entries are usually ordered alphabetically by the author's surname. Each item contains the following information:

> label, author(s), title, publisher, publisher's home town, year of publication.

In our case the labels are numbers (there are other possibilities, see below), and the book's title is typeset in italic font. In [16] the publisher is 'Springer-Verlag', located in Berlin. As the publishing industry becomes global, the publisher's address is becoming increasingly irrelevant; publishers with multiple addresses are commonplace. The book identifier ISBN (International Standard Book Number), often given, is helpful for searches in libraries, catalogues, and databases. (For a reference including the ISBN, see [22].)

There are other ways to cite books, as in the following shorter variant of our paragraph.

> We briefly review some properties of p-adic numbers, introduced in [17]. Background material may be found in [13]: the arithmetical side of the theory is developed in [16, chapter 7].

To simplify the citations, we have removed the author's name and the year of publication from the text. However, reading has become more laborious because the reference number conveys no information about the source. The following labelling method offers a compromise, providing essential information in abbreviated form.

> We briefly review some properties of p-adic numbers, introduced in [Hen08]. Background material may be found in [Gou80]; the arithmetical side of the theory is developed in [Has80, chapter 7].

Citations to journal papers follow the same rules given for books. The bibliographical information is slightly different. Here is one example:

---

[2] The suffix *ff* to a page number means 'and the following pages'.

[30] P. Morton and J. H. Silverman, Periodic points, multiplicities, and dynamical units, *J. Reine Angew. Math.*, **461** (1995) 81–122.

The data to be supplied comprises

> label, author, title of paper, name of journal, volume number, year of publication, page numbers.

The title is in Roman font, whereas the name of the journal appears in italics, often in a standard abbreviated form. Thus *J. Reine Angew. Math.* stands for *Journal für die reine und angewandte Mathematik*[3]. The volume number is in boldface. Page numbers are normally given in full, but the style 361–7 (rather than 361–367) is allowed. As for books, there is an identifier for journals, called ISSN (International Standard Serial Number).

Other sources for citations are articles in electronic journals [9], articles in conference proceedings [31], books in collections [4], reprints of books [8], theses [3], etc. There are great variations in the style of a bibliography. The best thing is to look around for examples of bibliographical entries of the various types by different authors and publishers.

For unpublished documents on the web there is as yet no agreed format. The main information is the URL (Uniform Resource Locator), which identifies a web page uniquely. These are two examples of entries for web-based material.

[7] P. Cvitanović, R. Artuso, R. Mainieri, G. Tanner, G. Vattay, N. Whelan, and A. Wirzba, *the Chaos Book*, `http://chaosbook.dk/` (Sept 24, 2010).

[34] L. Pottmeyer, Heights and totally real numbers, preprint (2012) `arXiv:1206.2456v1`, 12 pages.

Item [7] is a web-book. In addition to authors, title, and URL, the date is provided. Web pages don't always carry dates, and even when they do, there is no way of deciding whether the displayed date corresponds to the last revision. The best one can do in this situation is to state when the web source was last accessed. Reference [7] has a large number of authors. If the authors need to be mentioned in the text, an abbreviation is required. In this case we would write 'In Cvitanović *et al.*[4]. [7]. . . '. The web-book [7] has been available for several years, and it is gradually evolving. This document structure would not be possible within the confines of traditional publishing.

---

[3] This is the oldest mathematics journal still in existence, founded in 1826.

[4] Abbreviation for the Latin *et alii*, meaning 'and others'.

The item [34] is an unpublished paper. It is stored in a large, open-access archive of preprints of scientific papers, called `arXiv`, supported by Cornell University Library. Each preprint in this archive has a unique 8-digit identifier, plus a version number (`v1`, in this example) which reflects the fact the the papers may be updated. Most preprints in the `arXiv` will eventually become available in a more permanent form. So, when proof-reading bibliographical entries, one should check that the papers in the `arXiv` haven't been published. Some editors accept only citations to permanent and accessible material.

A word of warning. When quoting from the web, be aware that there is an awful lot of rubbish there, and most web pages have no guarantee of correctness. Also keep in mind that anybody can type keywords in a search engine and access the corresponding entries in Wikipedia. Citations to an online encyclopedia rarely deserve a place in a respectable bibliography.

9.4.1 Avoiding plagiarism

There is no copyright in mathematical ideas, the law of intellectual property being primarily commercial. But there is copyright in using other people's words, pictures, and software. The act of publishing material produced by others as if it were your own is a serious academic offence, called **plagiarism**.

To respect copyright when using other people's words one must name the author, cite the source, and make clear which are the quoted words (either by quotation marks or by display). For extensive use of someone else's text (more than a paragraph, say), one must obtain the explicit permission of the copyright holder, and then insert a suitable remark in the acknowledgements section: 'I thank XXX for permission to quote YYY'.

The instant availability of free material on the web has increased considerably the risk of plagiarism caused by carelessness and naïvety rather than dishonesty. When cutting and pasting from other people's text, it is essential to keep a record of the source. A text stored without a note of its source attached could easily be used as our own text at a later time. As articles, books, and theses are now submitted in electronic form, it is possible to detect this type of plagiarism using specialised software.

9.5 TEX and LATEX

The programming language TEX, developed by D. E. Knuth [21], and its user-friendly offspring LATEX, written by L. Lamport [23], are the standard software in mathematical and scientific publishing. This book was typeset in LATEX.

This typesetting system is intended for text containing mathematical formulae. It also supports the high-level features needed to manage a complex document: generating tables of contents, numbering of chapters and equations, cross referencing, handling bibliographical data, etc.

Let us inspect a very short LATEX source file[5]:

```
\documentclass[12pt]{article}
\begin{document}
In 1735, Leonhard Euler proved the following remarkable formula:
\begin{equation}
     \sum_{n=1}^\infty\frac{1}{n^2}=1+\frac{1}{4}
          +\frac{1}{9}+\frac{1}{16}+\cdots=\frac{\pi^2}{6}.
\end{equation}
\end{document}
```

The information is stored in plain text. We must type it in a file, and then run the LATEX programme, which generates a new file suitable for display, typically in `PDF` format[6]. The end result is the following text, placed at the top of an empty page with the page number 1 at the bottom:

In 1735, Leonhard Euler proved the following remarkable formula:

$$\sum_{n=1}^\infty\frac{1}{n^2}=1+\frac{1}{4}+\frac{1}{9}+\frac{1}{16}+\cdots=\frac{\pi^2}{6}. \qquad (1)$$

The contrast between the awkwardness of the LATEX source and the neatness of the output may be off-putting at first. On closer acquaintance, it will instead become clear why this logical and powerful language has become so widespread.

In the source file we note frequent occurrences of the distinctive backslash symbol '\', followed by a word. These are LATEX commands, which instruct the programme to do something. We see that, after an initial preamble, a pair of commands delimit the beginning and the end of the document. Within the document, another pair delimits a displayed equation, and within the equation various commands generate mathematical symbols and structures. It appears that certain things occur automatically, such as the centering and numbering of the displayed equation. Also, there is some freedom in the way the source code is layed out, since the formula is split between two lines in the source code, but it appears on a single line in the output.

---

[5] The source file for the *Mathematical Writing* book contains over half a million characters.

[6] This is an established format for document representation.

You may already guess the shape of commands that will typeset a matrix or a table. Of course, there is a lot more than typesetting. For instance LATEX may be expanded by incorporating auxiliary *libraries* of commands for specific tasks, such as BIBTEX, which manages bibliographical information. Similarly, the appearance of a document may be altered by means of *macros*. These are libraries transparent to the user, which reset the layout parameters (font type and size, line and paragraph spacing, style of bibliographical list, etc.). Publishers use this facility to convert standard LATEX source files into documents with the desired appearance. The present book used macros.

Learning LATEX takes some time, but if you plan to write a thesis this is a worthwhile investment. There is no shortage of introductory material: the Cambridge University resources are a good place to start [25].

Exercises

Exercise 9.1. The following abstracts are poorly written. Identify the problems, hence write an improved version.

1. In this project the author examines some properties of continued fractions (CF). In the beginning, definitions, notations, and basic results and theorems are shown. Periodic continued fractions and best approximations are examined subsequently in depth. We examined a number of applications to mathematics and astronomy.

2. The most representative example of a 2-dimensional area-preserving twist map is the standard map, which is studied. Orbits for which the momentum p grows linearly (plus a periodic function) are shown to exist, classified and determined numerically. These orbits are the accelerator modes. The linear stability of these orbits is determined. The range in the parameter values for which they exist is also determined.

3. We consider equal parameter generalized quadrangles, $GQ(s,s)$. All GQs of order 2 and 3, $GQ(2,2)$ and $GQ(3,3)$, are known. It is conjectured all the GQ

of order five are known as well. The known GQ of order 5 is the symplectic GQ, $W(5)$. $W(5)$ along with its dual are conjectured to be the only GQs of order 5. The construction of a symplectic GQ is given and then used to construct the known GQ of order 5. Information about GQ was gathered, including some basic combinatorics, affine GQ, and incidence matrices in an attempt to prove the above conjecture.

4. Under the boundary condition on the initial value $T_k(0)$ $(T_k(0) > 0; k = 1, 2, \ldots)$ that $T_k(0) \to 0$ $(k \to \infty)$, we integrate the semi-infinite system of nonlinear differential equations $\dot{T}_k = 2T_k(T_{k+1} - T_{k-1})$, $(k = 1, 2, \ldots; T_0 = 0)$ to obtain their general solution. We further investigate the asymptotic time behaviours of this general solution as $t \to \pm\infty$.

Exercise 9.2. This is the longest opening sentence of a mathematical abstract I managed to find in the literature [38].

> By Morava's point of view on the stable homotopy category, the quotient in some sense associated to the filtration related to the height of formal group laws is studied by the category of modules over the function ring of the deformation space of the Honda group law of height n with the lift of the action of the automorphism group on the closed fibre through the Adams-Novikov spectral sequence.

1. Can you break it into shorter sentences?

2. Can you find a longer opening sentence in any mathematical publication?

Exercise 9.3. This exercise proposes some MICRO-PROJECTS in experimental mathematics. You are given a sequence to study; your task is to generate numerically a sufficient number of terms, analyse the data, and then write a succinct report — three/four pages, say.

In the report you must define the problem, give a concise account of your findings, and synthesise them in a precise mathematical conjecture. Invariably, some plots will be necessary. You should provide at least one bibliographical item, and include the numerical code in an appendix.

For an introduction to experimental mathematics, see [40].

1. EULER'S POLYNOMIAL. Let $p(n) = n^2 + n + 41$ be Euler's polynomial (see section 6.6), and let a_n be the number of prime values assumed by p over the first n natural numbers, that is,

   $$a_n = \#\{k \in \mathbb{N} : k \leqslant n,\ \ p(k) \text{ is prime}\} \qquad n \geq 1.$$

   Refine conjecture 3 of section 6.6 by identifying non-trivial bounds of the form

   $$f(n) \leqslant a_n \leqslant g(n) \tag{9.1}$$

   to hold for all sufficiently large n. The functions f and g must be chosen so as to minimise the growth rate of the difference $g - f$.

2. EXPONENTIAL SUMS. Given a real sequence (a_k) we define a sequence (z_n) of complex numbers via the sum

   $$z_n = \sum_{k=1}^{n} e^{2\pi i a_k} \qquad n \geq 1.$$

   The number z_n may be interpreted as the end-point of a walk on the complex plane, consisting of n steps of unit length. Study (z_n) for one of the following sequences (a_k):

   (a) $a_k = k^2/50$
   (b) $a_k = k^2/(50 + 1/100)$
   (c) $a_k = k \log(k)$.

   Formulate a conjecture concerning the behaviour of z_n or $|z_n|$ as n tends to infinity.

3. PRIMES AND POWERS OF TEN. Let a_n be as in (8.4), p. 177.

   (a) Prove that $f(\mathbb{N}) \subset A$, where

   $$A = \mathbb{N} \smallsetminus (2\mathbb{N} \cup (2 + 3\mathbb{N}) \cup 5\mathbb{N}\}). \tag{9.2}$$

   (b) Provide experimental evidence that $f(\mathbb{N})$ is unbounded.
   (c) Provide experimental evidence that $f(\mathbb{N}) = A$.

Appendix A

Solutions to exercises

Exercise 1.1 (p. 1):

1. The number a is positive.
2. The only even prime is 2.
   (The integer 2 is the only even prime.)
3. If $x > 0$, then $g(x) \neq 0$.
4. We change sign to both sides of the equation.
5. The polynomial $x^2 + 1$ has not real roots.
   (The equation $x^2 + 1 = 0$ has no real solutions.)
   (The solution set consists of odd integers.)
6. Multiplying both sides by a negative value of x, the inequality is reversed.
7. All solutions are odd.
8. If $\sin(\pi x) = 0$, then x is an integer.
   (The condition $\sin(\pi x) = 0$ implies that x is an integer.)
9. A matrix is invertible if its determinant is non-zero.
10. This infinite sequence has fewer negative terms.

Exercise 1.2 (p. 8):

1. There are three special cases.
3. It does not tend to infinity.
5. Therefore c^{-1} is undefined.
7. We square both sides of the equation.
9. The equation $x^2 = y^2$ describes two orthogonal lines.
11. All matrices in this set are invertible.
13. Purely imaginary means that the real part is zero.
15. Since x is zero, I can't divide by x.
17. The function f is continuous.
19. I found fewer solutions than I expected.
21. We prove Euler's theorem.
23. The asymptotes of this hyperbola are orthogonal.
25. The solution depends on s.
27. Thus $x = a$. (We assume that a is positive.)
29. Always remember to check the sign.

Exercise 2.1 (p. 45):

Bad: Is 39 a prime number? [*Specific and insignificant.*]
Good: Why is 1 not a prime number?

Bad: What is $1/2 + 1/2^2 + 1/2^3$?
Good: Is a fraction the same as a rational number?

Bad: What is the real part of $2 + 3i$?
Good: What is the real part of i^i?

Exercise 2.2 (p. 45):

BAD: The set of natural number less than 10.
GOOD: The power set of a finite set.

Exercise 2.3 (p. 45):

1. $\{1-2n : n \in \mathbb{N}\}$
   ($\{n \in \mathbb{Z} : n < 0, 2 \nmid n\}$)

3. $\{(n+1)/n : n \in \mathbb{Z} \setminus \{0\}\}$

5. $\{z \in \mathbb{C} : |z| \geqslant 1\}$.

7. Let $F(x,y) = 0$ be the equation of a circle with the stated property. The function F must be such that $F(0,0) = 0$.

9. No such a line goes through the origin, so we represent lines as $ax + by + 1 = 0$. The tangency condition results in a relation between a and b.

Exercise 2.4 (p. 45):

1. The set of rational points in the open unit interval.

3. The set of rationals whose denominator is a power of 2.

5. The imaginary axis in the complex plane, excluding the origin.

7. The set of integer pairs whose first component divides the second.
   (The set of integer points on the plane lying on lines with integer slope.)

9. The set of points in a finite-dimensional Euclidean space whose coordinates have zero sum.

11. A regular array of orthogonal lines on the plane, each parallel to one coordinate axis and lying at integer distance from it.

Exercise 2.5 (p. 46):

1. The inverse of a function.

3. The value of a function at the reciprocal of the argument.

5. The square of the value of a function at a point.

7. The image of the irrational numbers under a function.

9. The pre-image of the origin under a function.

Exercise 2.6 (p. 46):

1. *i*) An identity.
   (*ii*) An arithmetical identity, expressing a cube as the sum of three cubes.

3. *i*) A chain of inequalities.
   ii) Upper and lower rational bounds for the square root of 2.

5. *i*) An inequality.
   (*ii*) An algebraic inequality in two unknowns.
   (The inequality defining the first and third open quadrants in the cartesian plane.)

7. *i*) An equation.
   ii) The cartesian equation of a parabola passing through the origin.

9. *i*) An identity.
   ii) The trigonometric formula for the sine of the difference of two angles.

11. *i*) An equation.
    ii) A set equation, whose solutions are the sets having themselves as the only element.

Exercise 2.7 (p. 47):

1. *i*) A set.
   ii) The set of even integers that are not divisible by 4.

3. *i*) A set.
   ii) The set of rational points on the plane, with non-integer components.
   (The set of pairs of rational numbers that are not integers.)

5. See exercise 2.4.11.

7. *i*) A sequence.
   ii) An infinite sequence of sets, each containing one and the same element.

9. *i*) A finite sequence.
   ii) A finite sequence, obtained by raising consecutive natural numbers to the same even power.

11. *i*) A set. (A set of real numbers.)
    ii) The algebraic sum of a finite numbers of copies of the closed unit interval.

Exercise 2.8 (p. 47):

1. *i*) A function.
   ii) The real function that adds 1 to its argument.
   (The function that performs the right unit translation on the real line.)

3. *i*) An identity. (A functional identity.)
   ii) The formula for the derivative of the product of two functions.

5. *i*) An integral.
   ii) The indefinite integral of a function of two variables, performed with respect to the first variable.

7. *i*) An identity. (A definition.)
   ii) The power series of the cosine.

9. *i*) A finite product of functions. (An analytic expression.)
   ii) The product of all the partial derivatives of a function of several variables.

Exercise 2.9 (p. 48):

1. The quotient of two fractions is a fraction. Its numerator (denominator) is the product of the numerator (denominator) of the first fraction and the denominator (numerator) of the second fraction.

3. Compute the cube of the natural numbers: 1, 8, 27, etc. Stop when either *i*) you obtain your integer (which is then a cube); or *ii*) you obtain a larger integer, in which case your integer is not a cube.

   If your integer is 1, then it's a cube; otherwise, perform its prime factorization. If the exponent of each prime factor is divisible by 3, then your integer is a cube.

5. Transform the equation of the circle so that one side is zero, and the other side has positive quadratic coefficients. Then substitute the coordinates of the point into the equation. If you get a negative value, then the point is inside the circle; otherwise it's outside (or on the boundary).

7. Form the matrix that has your vectors as rows (or columns). Then compute the determinant of the matrix, and verify that it is non-zero.

9. Two things must happen: $i)$ there is a point at which the value of the two functions is the same; $ii)$ at this point, the product of the derivatives of the two functions is equal to -1.

11. Represent these points as complex numbers. Compute their arithmetic mean, and subtract it from each point. Now choose one of the translated points, and multiply it by increasing powers of the imaginary unit. If you get all the other translated points, in some order, then the original points form the vertices of a square.

Exercise 2.11 (p. 49):

1. Compute the value of the function at each point of the domain, collecting together the resulting values. Eliminate all duplicates, so as to turn this list into a set. Your function is surjective precisely when the cardinality of this set is the same as that of the co-domain.

   Compute the value of the function at each point of the domain, collecting together the resulting values. Now verify that each element of the co-domain also belongs to your list of values. If this is true, then your function is surjective; otherwise it isn't.

Exercise 2.12 (p. 49):

1. Let $X = (A^2 \smallsetminus B^2)$ and $Y = (A \smallsetminus B)^2$. If $A \subset B$, then, necessarily, $A^2 \subset B^2$, and hence $X = Y = \emptyset$. If A and B are disjoint, then so are A^2 and B^2, and hence $X = Y = A^2$. We will show that in all other cases Y is a proper subset of X.

   We begin with an example. Let $A = \{1,2\}$ and $B = \{1\}$. The pair $(2,1)$ belongs to $A^2 \smallsetminus B^2 = \{(1,2),(2,1),(2,2)\}$ but not to $(A \smallsetminus B)^2 = \{(2,2)\}$, as desired. Let us generalise this argument. If A is not a subset of B, then there is an element a in $A \smallsetminus B$, and if A and B are not disjoint, then we can find $b \in A \cap B$. Now, the pair $z = (a,b)$ belongs to X (because $z \in A^2$ and $z \notin B^2$) but doesn't belong to Y (because the second component of z is not in $A \smallsetminus B$). We have shown that Y is a proper subset of X, as desired.

Exercise 2.13 (p. 49): Consider *infinite* sequences of non-negative integers, where the kth term represents the power of the kth prime in the prime factorisation of an integer.

Exercise 2.14 (p. 49):

1. Consider infinite integer sequences.
2. Consider sums of the type
$$\sum_{k=n}^{\infty} d_k 10^{-k}$$
where n is an integer, and the coefficients d_k are decimal digits.

Exercise 3.1 (p. 73):

1. Use theorem 3.2, page 57.
3. Consider the set of which $\mathscr{P}$ is the characteristic function.

Exercise 3.2 (p. 74):

1. $\exists x \in X,\ x \notin Y$
3. $\exists x \in X,\ \exists y \in Y,\ z = x + y$
5. $\exists x \in \mathbb{R},\ \sin(\cos(x)) \neq \cos(\sin(x))$.

Exercise 3.3 (p. 74):

1. $\exists k \in \mathbb{Z},\ n = k^3$
3. $\exists x \in \mathbb{Q},\ f(x) = 0$
5. $\forall x \in \mathbb{Z},\ p(x) \neq 0$
7. $\exists y \in B,\ \forall x \in A,\ f(x) \neq y$
9. $\exists x, y \in A,\ f(x) \neq f(y)$
11. $\exists k \in \mathbb{N} \setminus \{1\},\ \exists m \in \mathbb{Z},\ mk^2 = n$
13. $\exists a, b, c, d \in \mathbb{N},\ (a \neq c) \wedge (a \neq d) \wedge (a^3 + b^3 = c^3 + d^3)$
15. $\forall a, b, x \in \mathbb{R},\ (a < x < b) \Rightarrow (\exists y \in \mathbb{R},\ a < y < x)$
17. $\exists n \in \mathbb{N},\ \#\{x \in \mathbb{Q} : f(x) = 0\} = n$.

Exercise 3.4 (p. 75): In part *i*) there is freedom in the choice of the ambient set: the smaller the set, the simpler the predicate. (But if the set becomes too small, then the implication disappears, see exercise 3.1.3.)

1. *i*) For all primes p, if $p > 2$, then p is odd. (TRUE)

   ii) For all primes p, if p is even, then $p \leqslant 2$.

   iii) For all primes p, if p is odd, then $p > 2$. (TRUE)

   iv) There is a prime p such that $p > 2$ and p is even.

3. *i*) For all positive fractions a/b, if a/b is reduced, then b/a is reduced. (TRUE)
   (For all fractions a/b, if a/b is positive and reduced, then b/a is reduced.)

   ii) For all positive fractions a/b, if b/a is not reduced, then a/b is not reduced.

   iii) For all positive fractions a/b, if b/a is reduced, then a/b is reduced. (TRUE)

   iv) There is a positive fraction a/b such that a/b is reduced but b/a is not reduced.

5. See example 3.5, p. 71.

7. *i*) For all odd integers n, if n is greater than 3, then one of n, $n+2$, $n+4$ is not prime. (TRUE)
   (For all integers n, if n is odd and greater than 3, then one of n, $n+2$, $n+4$ is not prime.)

   ii) For all odd integers n, if n, $n+2$ and $n+4$ are all prime, then $n \leqslant 3$.

   iii) For all odd integers n, if one of n, $n+2$ and $n+4$ is not prime, then $n > 3$. (FALSE)

   iv) There is an odd integer n such that $n > 3$, and n, $n+2$, $n+4$ are all prime.

Exercise 3.5 (p. 76): This is the definition of nephew:

x is a nephew of y := *There is z such that x is a son of z and*
(z is a brother of y or z is a sister of y).

Thus to define nephew we must first define of brother and sister; in turn these will require other definitions.

Exercise 3.6 (p. 76):

(a.1) The integer 5 has a proper divisor. (FALSE)

(a.3) Every natural number properly divides its square. (FALSE)

(a.5) Every integer greater than 1 is a proper divisor of some integer. (TRUE)

(b.1) The integer y has a proper divisor.

(b.3) The integer x is a proper divisor of some integer.

(c) Consider part (b).

Exercise 3.7 (p. 77):

1. The set of prime numbers, together with the integer 1.

Exercise 3.8 (p. 77): Use theorem 3.5, p. 70.

1. (FALSE). $\exists n \in \mathbb{N},\ 1/n \in \mathbb{N}$.
   The reciprocal of a natural number is not a natural number.
   There is a natural number whose reciprocal is also a natural number.

3. (TRUE). $\exists x, y \in \mathbb{R},\ xy \neq yx$.
   The product of real numbers is commutative.
   The product of real numbers is not commutative.

5. (FALSE). $\exists m, n \in \mathbb{Z},\ (m+n \in 2\mathbb{Z}) \wedge m \in 2\mathbb{Z} \wedge n \in 2\mathbb{Z}$.
   If the sum of two integers is even, then at least one summand is odd.
   There are two even integers whose sum is even.

7. (TRUE). $\exists y \in \mathbb{R},\ \forall x \in \mathbb{R}^+,\ \log(x) \neq y$.
   The logarithmic function is surjective.
   The logarithmic function is not surjective.

9. (TRUE). $\exists \varepsilon > 0,\ \forall r \in \mathbb{Q},\ |r - \sqrt{2}| \geqslant \varepsilon$
   There are rational numbers arbitrarily close to the square root of 2.
   There is a neighbourhood of the square root of 2 without any rational number.

Exercise 3.9 (p. 78): Let $G = \{1, \dots, n\}$. Think of the computation of the predicate as the evaluation of the entries of an $n \times n$ matrix, where the (i, j)-entry is the value of 'i loves j'. Then fix the order in which the entries are evaluated.

1. Let $g = n$, say. If the sentence is true, then for all $k = 1, \dots, n$ we must verify that $\mathscr{L}(k, g) = \mathtt{T}$ (George is included), so in any case we need n function evaluations.

   If the sentence is false, then the minimum is 1 (we find $\mathscr{L}(1, g) = \mathtt{F}$ at the first evaluation). The maximum is n (we find $\mathscr{L}(k, g) = \mathtt{T}$ for all $k \neq n$, and $\mathscr{L}(g, g) = \mathtt{F}$).

Exercise 4.1 (p. 91): First translate symbols into words literally; then synthesise the meaning of the literal sentence. (It may be helpful to draw the graph of a function that has the stated property, and a function that hasn't.)

1. The function f vanishes at the origin.
3. The function f is constant (over the reals).
5. The function f vanishes at some integer.
7. The function f assumes only integer values (over the reals).
9. The function f coincides with its own inverse.
   (The function f is an involution.)

Exercise 4.2 (p. 91):

1. The zeros of f include all the integers.
3. The function f is identically zero for negative values of the argument.
5. The function is non-zero at all rationals, with the possible exception of 0.
7. The function f has infinitely many integer roots.

Exercise 4.3 (p. 91):

1. $\forall x \in \mathbb{R},\ \ f(x) \neq 0.$
3. $\forall x \in \mathbb{R},\ \ f(x) = 0 \Rightarrow x = 0.$
5. $\forall x \in 2\mathbb{Z},\ \ f(x) = 0.$
7. $\forall x \in (0,1),\ \ \exists y \in (0,1),\ \ f(x+y) = 0.$

Exercise 4.4 (p. 92):

1. (a) If $-f$ is increasing, then f is decreasing. (TRUE)
   (b) If $-f$ is not increasing, then f is not decreasing. (TRUE)
   (c) There is a decreasing function f such that $-f$ is not increasing.

3. (a) If f is monotonic, then $|f|$ is increasing.
   (FALSE, e.g., $f(x)=x$)
   (b) If f is not monotonic, then $|f|$ is not increasing.
   (FALSE, e.g., $f(x)=e^x$ for $x<0$ and $-e^x$ for $x \geqslant 0$.)
   (c) There is a non-monotonic function whose absolute value is increasing.

5. (a) If f is surjective, then f is unbounded. (TRUE)
   (b) If f is not surjective, then f is bounded.
   (FALSE, e.g., $f(x)=\lfloor x \rfloor$.)
   (c) There is an unbounded function which is not surjective.

Exercise 4.5 (p. 92):

1. *i*) For all real functions f, if f is differentiable, then f is continuous. (TRUE)

   ii) For all real functions f, if f is discontinuous, then f is not differentiable.

   iii) Every continuous real function is differentiable. (FALSE)

   iv) There is a real function f which is differentiable but not continuous.

3. *i*) For all real functions f and g, if f and g are odd, then $f+g$ is odd. (TRUE)

   ii) For all real functions f and g, if $f+g$ is not odd, then one of f or g is not odd.

   iii) For all real functions f and g, if $f+g$ is odd, then f and g are odd. (FALSE)

   iv) There are odd real functions f and g such that $f+g$ is not odd.

Exercise 4.7 (p. 93):

1. A smooth bounded odd function, which vanishes infinitely often and approaches zero alternating sign as the argument tends to infinity.

3. This is a step function defined only for positive arguments. The steps have unit length, while their heights increases monotonically.

5. A differentiable even function, with a global minimum at the origin and no maximum. For positive arguments the function increases monotonically, approaching a positive limit.

Exercise 4.9 (p. 94):

1. The logarithm.

3. The identity.

5. The sine of the square of the argument.

7. The square of the arc-tangent.

Exercise 4.10 (p. 95):

1. $\exists k \in \mathbb{N},\ a_k \neq b_k$

3. $\forall n \in \mathbb{N},\ \exists k \in \mathbb{N},\ a_k \neq a_{k+n}$

5. $\forall k \in \mathbb{N},\ \exists j \in \mathbb{N},\ a_{k+j} < 0$

7. $\forall \varepsilon \in \mathbb{R}^+,\ \exists k \in \mathbb{N},\ |a_k| < \varepsilon$

9. $\forall n \in \mathbb{N},\ \exists k \in \mathbb{Z} \smallsetminus \{0\},\ a_{n+k} = a_k$.

Exercise 5.1 (p. 119):

1. When we say that the price of petrol, say, on Thursday, is ten percent higher that the price on Wednesday, we mean that the Thursday price is equal to the Wednesday price plus one tenth of the Wednesday price, so it's $(1+1/10) = 11/10 = 1.1$ times the Wednesday price. Likewise, if we say that the Friday price is 10 percent lower than the Thursday price, that means it's the Thursday price minus one tenth of the Thursday price (*not* one tenth of the Wednesday price!); so the Friday price is $(1-1/10) = 9/10 = 0.9$ times the Thursday price. In all, the Friday price is the Wednesday price times 11/10 times 9/10, which is 99/100=0.99 times the Wednesday price —nearly the Wednesday price but not quite.

2. Let $\mathbb{R}^+$ be the set of non-negative real numbers, representing the values of an observable (e.g., the price of a commodity). We consider the function

   $$f_p : \mathbb{R}^+ \to \mathbb{R}^+$$

   where $f_p(x)$ is the result of varying x by a given percentage amount p, followed by a variation by $-p$. The parameter p is a real number, which we require to be in the range $-100 \leqslant p \leqslant 100$, to ensure that the process does not result in negative quantities, i.e., that $f_p(x) \in \mathbb{R}^+$.

   A single change of x by p% corresponds to the mapping

   $$x \mapsto x\left(1+\frac{p}{100}\right).$$

   If this transformation is followed by a percentage change with the opposite sign, we obtain

   $$f_p(x) = x\left(1+\frac{p}{100}\right)\left(1-\frac{p}{100}\right) = x\left[1-\left(\frac{p}{100}\right)^2\right].$$

   For fixed p, this is a linear function of x, whose coefficient depends on $|p|$. If we exclude the trivial case $p=0$, we have $f_p(x) < x$, which means that at the end of the process the value of x has become smaller. The decrease is maximal if $p = \pm 100$, in which case the final value is zero, irrespective of the initial value.

Exercise 5.4 (p. 120): Imagine that there are 100 boxes, instead of three; the presenter opens 98 boxes, all of them empty.

Exercise 5.5 (p. 120): Consider the extended Euclid's algorithm.
Exercise 5.7 (p. 121):

1. You should identify at least three ways.

2. Construct a sequence of intervals with given mid-point and length converging to zero.

3. The domain of a 'random variable' is the space of events, such as the set of outcomes of n coin tosses. For each outcome, the 'random variable' gives us a real number.

4. Begin with the following lineup: $n!, n^{10}$, 2^n. Consider $n^{\log(n)}$ against $\log(n)^n$.

Exercise 6.1 (p. 150):

(a) A point on the first line is represented by the position vector

$$\mathbf{l}(\lambda) = (4,5,1)^T + \lambda\,(1,1,1)^T = (4+\lambda, 5+\lambda, 1+\lambda)^T$$

for some real number λ. Similarly, a point on the second line has position vector

$$\mathbf{m}(\mu) = (5,-4,0)^T + \mu\,(2,-3,1)^T = (5+2\mu, -4-3\mu, \mu)^T$$

for some real μ. For the lines to intersect, there must exist values of λ and μ for which the two position vectors are the same, namely

$$\mathbf{l}(\lambda) = \mathbf{m}(\mu). \tag{A.1}$$

The above vector equation corresponds to three scalar equations:

$$\begin{aligned} 4+\lambda &= 5+2\mu && \text{(A.2)} \\ 5+\lambda &= -4-3\mu && \text{(A.3)} \\ 1+\lambda &= \mu. && \text{(A.4)} \end{aligned}$$

Eliminating μ from equations (A.2) and (A.4), we obtain

$$4+\lambda = 5+2(1+\lambda)$$

which yields the solution $\lambda = -3$, $\mu = -2$. This solution also satisfies equation (A.3), and hence it satisfies all three equations. Substituting these values in equation (A.1) we obtain

$$\mathbf{l}(-3) = \mathbf{m}(-2) = (1,2,-2)^T$$

which is the position vector of the common point of the two lines. □

(c) Let a be a real number, and let $y = L_a(x)$ be the equation of the line with parameter a. Likewise, let $y = P(x)$, be the equation of the parabola. The points of intersection of the two curves are the solutions of the equation $L_a(x) = P(x)$, which reads

$$ax - \frac{(a-1)^2}{4} = x^2 + x.$$

A straightforward rearrangement yields

$$x^2 + x(1-a) + \frac{(a-1)^2}{4} = \left(x - \frac{a-1}{2}\right)^2 = 0, \tag{1}$$

and we find two identical solutions:

$$x_+ = x_- = \frac{a-1}{2}.$$

It remains to verify that, at the point x_+ at which L_a and P intersect, the two curves have the same slope, that is, their derivatives are the same. Now, the derivative of L_a is equal to a at every point; for P we find

$$P'(x) = 2x + 1 \qquad P'(x_+) = P'((a-1)/2) = 2\frac{a-1}{2} + 1 = a,$$

as desired. □

Exercise 6.2 (p. 151):

(a) The statement of the theorem is imprecise in several respects.

The nature of the numbers x and y is not specified (the inequality would be meaningless for complex numbers).

The case in which one of x or y is zero should be excluded, since in this case the left-hand side of the inequality is undefined.

The statement is false unless the inequality is made non-strict; indeed the equality holds for infinitely many values of x and y.

There are several flaws in the proof.

The basic deduction is carried out in the wrong direction, which proves nothing. This is the main flaw. (Proving that $P \Rightarrow$ TRUE gives no information about P.)

> The assertion 'the last equation is trivially true' is, in fact, false for $x = y$.
>
> The writing is inadequate, without sufficient explanations, and also imprecise (the expression $(x-y)^2 > 0$ is an inequality, not an equation).

(b) Revised statement:

THEOREM: *For all nonzero real numbers x and y, the following holds*

$$\frac{x^2+y^2}{|xy|} \geqslant 2.$$

PROOF: Let x and y be real numbers, with $xy \neq 0$. Since the square of a real number is non-negative, we have $(x \pm y)^2 \geqslant 0$. We shall deduce our result from this statement. (Both signs will be used.)

We begin by noting the chain of implications

$$(x \pm y)^2 \geqslant 0 \;\Rightarrow\; x^2 \pm 2xy + y^2 \geq 0 \;\Rightarrow\; x^2 + y^2 \geq \mp xy.$$

Now, since xy is non-zero, we may divide both sides of the rightmost inequality by the positive quantity $|xy|$, to obtain

$$\frac{x^2+y^2}{|xy|} \geqslant \mp 2 \frac{xy}{|xy|}.$$

To complete the proof, we must consider the value of $xy/|xy|$. We have two cases:

Case I: If xy is positive, then $xy = |xy|$, and by choosing the positive sign on the right-hand side our result is proved.

Case II: If xy is negative, then $-xy = |xy|$, and by choosing the negative sign on the right-hand-side our result is proved. ☐

Exercise 6.3 (p. 152): The flaw is subtle: it's related to the incorrect inversion of the logarithmic function. In section 6.7.3 we dealt with the same problem in a simpler setting.

Exercise 6.4 (p. 152):

1. Let X be a compact set, and let f be a function over X. Assume that f is continuous: RTP: f is uniformly continuous.

3. Let $f(x)$ be a polynomial with integer coefficients, and assume the f is not a constant. RTP: There is an integer n such that $f(n)$ is not prime.

5. Let D be the interior of the unit disc, and let f be a bi-unique analytic mapping of D into itself. RTP: f is a linear fractional transformation.

7. Let P be a Pisot substitution, let M be its incidence matrix, and let f be the characteristic polynomial of M. RTP: f is irreducible over $\mathbb{Q}$.

9. Let D be a differential of the second or third kind. RTP: There is a normal differential N such that $D-N$ is a differential of the first kind.

11. Let f be a real continuous function over a closed interval I, and let m and M be the upper and lower bounds of f, respectively. RTP: For all $m \leqslant y \leqslant M$, there is an $x \in I$ such that $f(x) = y$.

Exercise 7.1 (p. 165): The difficulty in a computer-assisted proof consists in converting the two members of each identity in (7.4) to the same form. In the following Maple codes the functions `expand` and `normal` are used for this purpose. In each case, the output of the last expression is the boolean constant TRUE.

1.
```
a:=k->2*k-1:
F:=n->n^2:
a(1)=F(1) and F(k)+a(k+1)=expand(F(k+1));
```

3.
```
a(1)=F(1) and normal(F(k)+a(k+1))=F(k+1);
```

5. The base case is $n = 2$.

```
a:=k->1/(k^2-1):
F:=n->3/4-(2*n+1)/(2*n*(n+1)):
a(2)=F(2) and normal(F(k)+a(k+1))=normal(F(k+1));
```

Exercise 7.4 (p. 166):

1. The formula is given in section 2.1.

3. Define $f^n = \underbrace{f \circ f \circ \cdots \circ f}_{n}$, and $x_n = f^n(x)$, with $x_0 = x$. Then

$$(f^n(x))' = \prod_{k=0}^{n-1} f'(x_k).$$

Exercise 7.5 (p. 166):

1. The base case is not proved.
3. The inductive step fails for $k = 2$.

Exercise 8.1 (p. 183):

1. Divide the square into four suitable regions.
2. Consider the function giving the number of friends of a person.
3. The boxes are the elements of the set X. What are the objects?

Exercise 8.2 (p. 183):

1. The denominator of $f(x)$ vanishes at $x = (-1 \pm \sqrt{5})/2$. Hence the domain of the function should be restricted to $\mathbb{R} \smallsetminus \{(-1 \pm \sqrt{5})/2\}$.
3. The value of the function at a/b is not well-defined, because a and b are not necessarily co-prime. Such a value should be changed to $|a-b|/\gcd(a,b)$. Alternatively, one could change the domain to the set of reduced rationals, or to the set of co-prime integer pairs in $\mathbb{Z} \times \mathbb{N}$.

Exercise 8.3 (p. 184):

1. FAULTS:

   The integer m must be positive, otherwise the summation is undefined. Alternatively, the sum should be defined to be 0 if $m < 1$.

   The set B is undefined. In particular, we cannot conclude that the reciprocal of the terms of the sequence is well-defined.

   The co-domain of f is not necessarily B, even if the sum is well-defined.

   The symbol a is introduced but not used.

   To improve clarity, the dependence of f on the sequence should be made explicit, as a parameter or a second argument. For coherence, this quantity should be called b, not a. Stating that the sequence is infinite would also be helpful.

   REVISED STATEMENT:

Let $b = (b_1, b_2, \ldots)$ be a fixed infinite sequence of non-zero complex numbers. We define the function f_b as follows

$$f_b : \mathbb{N} \to \mathbb{C} \qquad m \mapsto \sum_{k=1}^{m} b_k^{-1}.$$

(The set $\mathbb{C}$ could be replaced by any field.)

Exercise 8.4 (p. 184):

1. Consider the number of twin primes smaller than a given bound.
2. Is the representation of an even integer as a sum of two primes unique?
3. Consider the smallest non-negative integer t for which the sequence with initial condition $x_0 = n$ has $x_t = 1$.

Exercise 9.1 (p. 201):

1. The writing is clumsy; remove reference to project and author in the first sentence; the acronym CF is not used; the verb ‘examine’ is over-used.
2. Bad first sentence: ‘an example of this is that’; the numeral ‘2’ is inappropriate; the topic of the thesis (accelerator modes) should be given more prominence; a symbol is introduced but not used; the last two sentences could be merged into one.
3. Inappropriate and heavy use of symbols obscures the main message; introduce a concise symbol for to $GQ(s,s)$; the conjecture is formulated twice; the last sentence is clumsy.
4. The first sentence should be re-arranged; some symbols are unnecessary, e.g., those specifying the range of k-values.

Exercise 9.3 (p. 202):

1. The following Maple code generates the sequence (a_n):

   ```
   N:=10000:
   a:=array(1..N,[1]):
   for n from 2 to N do
      if isprime(n^2+n+41) then
        a[n]:=a[n-1]+1
      else
        a[n]:=a[n-1]
      end if
   end do:
   ```

   We start with the lower bound $f(n) = \sqrt{n}$, while the counterexample (6.8), p. 138 shows that $g(n) \leqslant 40n/41$ for $n \geqslant 41$. The following code displays the ratios $f(n)/a_n$ and $g(n)/a_n$ within the same plot:

   ```
   with(plots):
   f,g:=n->sqrt(n),n->n*40/41:
   display({plot([seq([n,f(n)/a[n]],n=1..N)]),
            plot([seq([n,g(n)/a[n]],n=1..N)],color=blue)});
   ```

   Sharpen these bounds by experimenting with fractional powers and logarithms. (In particular, choose g so that $g(n)/n \to 0$.) A pertinent reference is Golbach's theorem (1752), which states that no non-constant polynomial with integer coefficient can be prime for all (or all sufficiently large) values of the indeterminate, see [14, theorem 21].

2. This code generates and plots the sequence (z_n):

   ```
   a:=n->n^2/50:
   N:=5000:
   TwoPiI:=evalf(2*Pi*I):
   z:=array(1..N):
   z[1]:=evalf(exp(TwoPiI*a(1))):
   for n from 2 to N do
     z[n]:=z[n-1]+evalf(exp(TwoPiI*a(n))):
   end do:
   plot([seq([Re(z[n]),Im(z[n])],n=1..N)]):
   ```

To increase speed, the exponentials are converted to floating-point with `evalf`. These sequences produce beautiful patterns, but the latter will become apparent only if several thousand terms are plotted.

Gauss first studied these sums in the case $a_n = n^2/p$, $N = p$, with p prime. This is a **quadratic Gauss' sum**, which can be the starting point for a bibliographical search.

3. (a) If $a_n \in \mathbb{N} \smallsetminus A$, then q_n is divisible by 2,3, or 5.

   (b-c) The following code generates the sequence (a_n):

```
N:=500:
a:=[seq(nextprime(10^(n-1))-10^(n-1),n=1..N)]:
```

   This computation is very time-consuming, as it deals with huge integers.

   Define the sequences of sets

$$A_n = \{a_k : k \leqslant n\}, \qquad W_n = A \smallsetminus A_n, \qquad n \geqslant 1,$$

   where A is given in (9.2). Let a_n^+ be the largest element of A_n and let a_n^- be the smallest element of W_n. Both sequences are non-decreasing. The set $f(\mathbb{N})$ is unbounded if and only if $a_n^+ \to \infty$, whereas the condition $a_n^- \to \infty$ is seen to be equivalent to $f(\mathbb{N}) = A$.

   The following code generates the sequences $(a_n^\pm)$:

```
A:={$1..N} minus
    select(n->igcd(n,10)>1 or (n mod 3)=2, {$1..N}):
ap,am:=array(1..N),array(1..N):
for n to N do
   {op(a[1..n])}:
   A minus %:
   ap[n]:=max(%%):
   am[n]:=min(%%)
end do:
```

   You may consider replacing a_n^+ with the less volatile arithmetical mean of the elements of A_n, generated with the command `add(k,k=a[1..n])/n`. A plot of few hundred terms of this sequence should convince you that $f(\mathbb{N})$ is unbounded, but you must decide if your data provide enough support to the conjecture that $f(\mathbb{N}) = A$.

The bibliography could refer to expository works on the randomness of prime numbers, e.g., [37, chapter 3].

Bibliography

[1] L. Alcock and A. Simpson, Ideas from mathematics education: an introduction for mathematicians, The Higher Education Academy, Maths, Stats and OR Network, School of Mathematics, The University of Birmingham, Birmingham UK (2009) [ISBN:978-0-9555914-3-3].

[2] L. V. Ahlfors, *Complex Analysis*, third edition, McGraw-Hill, New York (1979).

[3] V. Anagnostopoulou, Sturmian measures and stochastic dominance in ergodic optimisation, PhD Thesis, Queen Mary, University of London (2008).

[4] V. G. Berkovich, *Spectral theory and analytic geometry over non-archimedean fields*, volume 33 of *Mathematical surveys and monographs*, Americal Mathematical Society, Providence, RI, USA (1990).

[5] C. Caldwell, The Prime Pages, The University of Tennessee at Martin `http://primes.utm.edu/` (January 2013).

[6] L. Carlitz, Finite sums and interpolation formulas over $GF[p^n,x]$, *Duke Math. J.* **15** (1948) 1001–1012.

[7] P. Cvitanović, R. Artuso, R. Mainieri, G. Tanner, G. Vattay, N. Whelan, and A. Wirzba, *the Chaos Book*, `http://chaosbook.dk/` (Sept 24, 2010).

[8] H. Cohn, *Advanced Number Theory*, Dover Publications, New York (1980). (Originally published as *A second course in number theory*, Wiley, New York (1962).)

[9] D. Denisov and V. Wachtel, Conditional limit theorems for ordered random walks, *Electron. J. Probab.* **15** (2010) 292–322. MR2609589.

[10] W. Dunham, *Euler, The Master of Us All*, volume 22 of *The Dolciani mathematical expositions*, The Mathematical Association of America, Washington, DC (1999).

[11] H. M. Edwards, *Fermat's Last Theorem*, Springer-Verlag, New York (1977).

[12] L. Gillman, *Writing Mathematics Well,* The Mathematical Association of America, Washington, DC (1987).

[13] F. Q. Gouvêa, *p-adic Numbers*, second edition, Springer-Verlag, Berlin (2000).

[14] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, fifth edition, Oxford University Press, Oxford (1979).

[15] E. O. Harriss, On canonical substitution tilings, PhD thesis, Imperial College, London (2004).

[16] H. Hasse, *Number Theory*, reprint of the 1980 edition, Springer-Verlag, Berlin (2002).

[17] K. Hensel, *Theorie der Algebraishen Zahlen*, Teubner, Leipzig and Berlin (1908).

[18] N. J. Higham, *Handbook of Writing for the Mathematical Sciences* second edition, SIAM, Philadelphia (1998).

[19] K. Houston, *How to Think Like a Mathematician*, Cambridge University Press, Cambridge (2009).

[20] S. G. Krantz, *A Primer of Mathematical Writing*, American Mathematical Society, Providence, RI, USA (1997).

[21] D. E. Knuth, *The TeXbook*, Addison-Wesley, Reading, Mass., USA (1984).

[22] D. E. Knuth, T. L. Larrabee, and P. M. Roberts, *Mathematical Writing*, The Mathematical Association of America, Washington, DC (1989).

[23] L. Lamport, *LaTeX: A Document Preparation System*, second edition, Addison-Wesley, Reading, Mass., USA (1994).

[24] S. Lange, *Undergraduate Analysis*, second edition, Springer-Verlag, New York (1997).

[25] Text processing using LaTeX, Dept. of Engineering, University of Cambridge `http://www.eng.cam.ac.uk/help/tpl/textprocessing/` (January 2013).

[26] H-C. Li, p-adic dynamical systems and formal groups, *Compos. Math.* **104** (1996) 41–54.

[27] H-C. Li, When is a p-adic power series an endomorphism of a formal group?, *Proc. Amer. Math. Soc.*, **124** (1996) 2325–2329.

[28] H-C. Li, Isogenies between dynamics of formal groups, *J. Number Theory* **62** (1997) 284–297.

[29] M. Liebeck, *A Concise Introduction to Pure Mathematics*, Chapman & Hall, Boca Raton, Florida, USA (2000).

[30] P. Morton and J. H. Silverman, Periodic points, multiplicities, and dynamical units, *J. Reine Angew. Math.* **461** (1995) 81–122.

[31] W. Narkiewicz, Finite polynomial orbits. A survey. In *Algebraic Number Theory and Diophantine Analysis*, F. Halter-Koch and R. F. Tichy, editors, proceeding of the international conference, Graz, Austria, August 30–September 5, 1998, Walter de Gruyter, Berlin (2000) 331–338.

[32] *Concise Oxford English Dictionary*, 12th edition, Oxford University Press, Oxford (2011).

[33] R. H. Pichler, Calculating the order of a matrix, MSc thesis, Queen Mary, University of London (1992).

[34] L. Pottmeyer, Heights and totally real numbers, preprint (2012). `arXiv:1206.2456v1`, 12 pages.

[35] P. Suppes, *Introduction to Logic*, Dover Publications, New York (1999). (Originally published by D. Van Nostrand Company, New York (1957).)

[36] W. Strunk and E. B. White, *The Elements of Style,* fourth edition, Longman, New York (2000).

[37] G. Tenenbaum and M. Mendès France, *The Prime Numbers and their Distribution*, volume 6 of *Student mathematical library*, American Mathematical Society, Providence, RI, USA (2000).

[38] T. Torii, One-dimensional formal group law of height n and $n-1$, PhD Thesis, Johns Hopkins University (2001).

[39] L. Truss, *Eats, Shoots and Leaves: the Zero-Tolerance Approach to Punctuation*, Profile Books (London) (2003).

[40] F. Vivaldi, *Experimental Mathematics with Maple*, Chapman & Hall, Boca Raton, Florida, USA (2001).

[41] G. J. Wirsching, *The Dynamical System Generated by the* $3x+1$ *Function*, volume 1681 of *Lecture notes in mathematics*, Springer-Verlag, Berlin (1998).

[42] *Writing matters: the Royal Literary Fund report on student writing in higher education*, S. Davis, D. Swinburne, and G. Williams, editors, The Royal Literary Fund, London (2006). (Electronic copy available from `http://www.rlf.org.uk/fellowshipscheme/research.cfm` (January 2013).)

Index

$3x+1$ conjecture, 177, 182
e, see Napier's constant
TeX,LaTeX, 198
[∉], 6, 31, 37, 43, 44, 46, 47, 74, 75, 87, 91, 115, 117–119

abscissa, 19
affine, 85
AGM inequality, 153
algebraic product, 20, 21
algebraic sum, 20, 21, 23, 35
alphabet, 99
ambient set, 13, 16, 39, 70, 72
Archimedes
  Archimedean property, 63
  constant, 42, 100
Argand plane, *see* complex plane
argument, 23, 96
arithmetic progression, 110
ascending, 31
assignment statement, 11
associative, 13, 174
axiom, 121
  subset, 16

base, 17
Bernoulli inequality, 153, 160
Bertrand's postulate, 175
bi-unique correspondence, 27, 56
bijective, 25, 83
binary, 13, 97
binomial, 33
  coefficient, 42
  theorem, 153
boolean function, 97
boolean operator
  AND ($\wedge$), 52
  NOT ($\neg$), 52
  OR ($\vee$), 52
  XOR ($\veebar$), 52
  if ($\Leftarrow$), 53
  if and only if ($\Leftrightarrow$), 54
  only if ($\Rightarrow$), 52, 125
both ends method, 135, 162, 169
bound, 171, 201
  lower, 81
  strict, 155
  upper, 81, 161
bounded
  function, 81
  set, 81
bounded away from zero, 81

Cantor, 10
cardinality, 12, 27
cartesian plane, 19, 115
cartesian product, 13, 15, 19, 23, 24, 31, 60, 97, 101, 115
Cauchy-Schwarz inequality, 153
ceiling, 86
chain rule of differentiation, 142
characteristic function, 56, 57, 62, 97

co-domain, 23, 175
co-prime, 18
coefficient, 33
commutative, 13, 36
complement, 13
complex
    number, 17, 100
    plane, 17
component, 96
composition of functions, 25, 30, 41
computer-assisted proof, 159
congruence
    class, 20
    operator, 51
congruent, 20
conjecture, 136
conjunction, 52, 131, 132, 172
    conjuncts, 131
continuous, 84
contradiction, *see* proof
contrapositive, 54, 129–131
convergence, 88
converse, 53, 54, 141
coset, 133
countable, 27
counterexample, 135

De Morgan, 154
    laws, 55, 64, 67
decimal point, 17, 147
Dedekind, 101, 157
definition, 11, 174
    recursive, 113
degree, 33
    total, 33
descending, 31
difference, 17
dimension, 30
Dirichlet's box principle, 169, 170, 181
discontinuous, 84, 87
disjoint, 12
divide and conquer method, 109, 110
divisibility, 18
    operator, 51
domain, 23, 175
double sum, 105
dummy variable, 104

eigenvalue, 9
element, 10, 29, 96
ellipsis, 14, 15, 32, 39
    raised, 17, 31, 32
empty set, 10, 15, 174
entry, 96
envelope, 104
equality, 50
equation, 34, 57
    algebraic, 41
    difference, 176
    differential, 35
    functional, 35, 37
    polynomial, 156
    simultaneous, 37
    transcendental, 41
equivalence relation, 70
equivalent
    equations, 48
    sets, 27, 28, 83
    statements, 54, 55
estimate, *see* bound
Euclid, 137, 161
    theorem, 134, 167
Euler, 101, 135, 156, 198, 201
    φ-function, 107
    gamma-function, 101
Euler-Mascheroni constant, 100

eventually periodic, 88, 146, 177, 191
exceptional set, 9
existence proof, 167
  constructive, 167
  effective, 171
  non-constructive, 167, 168
existence statement, 167
exponent, exponentiation, 17
expression, 37
  algebraic, 41
  arithmetical, 41
  combinatorial, 42
  compound, 52
  integral, 42
  nested, 42
  relational, 49, 50, 77
  trigonometric, 42

factorial, 42, 59, 176
Fermat, 156, 173
Fibonacci sequence, 177
finite, 12
floor, 86
Fourier, 104
fractional part, 86
function, 24, 101
  arithmetical, 88
  bounded, 81
  constant, 25
  continuous, 84, 85
  decreasing, 78
  differentiable, 85
  discontinuous, 84
  even, 79, 170
  increasing, 78
  invertible, 25
  linear, 85
  monotonic, 78
  negative, 77
  odd, 79
  periodic, 80
  positive, 77
  regular, 83
  singular, 83
  smooth, 85
functional, 97
fundamental theorem of arithmetic, 172

Gauss, 20, 188, 222
gcd, *see* greatest common divisor
general term, 29, 30, 32, 102
geometric progression, 106
Goldbach
  conjecture, 137, 182
  theorem, 222
graph, 24, 46, 96, 101
greatest common divisor, 18, 20, 24, 39
group, 34, 35, 101, 133
  axioms, 122, 173
  identity, 173

hcf, *see* greatest common divisor

identity, 36, 38, 58, 66, 145, 198
identity function, 24, 35
if, iff, *see* boolean operator
ill-defined, 174
image, 25, 68, 81
imaginary
  part, 17
  unit, 17, 100
implication, *see* boolean operator, only if
inclusion operator, 50
incremental ratio, 85
indeterminate, 33, 97, 178
indeterminate equation, *see* identity
induction, 154, 176

    basis of, 157
    inductive step, 157
    strong, 154, 160, 176
inequality, 50
infinite, 12
infinite descent, 154, 156
infinity, 19
initial condition, 176
injective, 25, 63, 64
integer, 14, 99
intermediate-value theorem, 171
intersection, 12, 23
interval, 57
    closed, 18
    half-open, 18, 28
    open, 18
inverse, 25
inverse image, 25, 56
invertible, 25
involution, 27
ISBN, 195
ISSN, 196

Kroneker, 101

Lambert, 101
Latin
    Q.E.D., 123
    definiendum, 11
    definiens, 11
    e.g., 2
    et al., 196
    etc., 2
    i.e., 116
    non sequitur, 140
lcm, *see* least common multiple
least common multiple, 18
least counterexample, 154
lemma, 121
length, 29
linear, 85
logical
    constant, 49
    equivalence, 54
    expression, 51
    operator, 49, 51
    tag, 123

map, mapping, 23, 97
mathematical induction, *see* induction
maximum, 82
    local, 82
member, 11, 96
membership operator, 50
metric space, 96
micro-essay, 7, 115, 116, 119, 190
micro-project, 201, 221
minimal period, 80
minimum, 82
Minkowski, 20
    sum (product), *see* algebraic sum (product)
modular arithmetic, 20, 36, 179
modulus, 20
monomial, 33
multiplicity, 10
multiset, 10, 172
multivariate, 33

Napier's constant, 42, 100
natural number, 14
necessary and sufficient, 132
negation of logical expression, 52, 67
neighbourhood, 82
    of infinity, 83, 88
nested

  expression, 39, 42, 69
  parentheses, 15
  sequence of sets, 31, 32, 88
number
  complex, 17, 100
  integer, 14
  natural, 14
  rational, 15, 100
  real, 16, 100

one-to-one, *see* injective
onto, *see* surjective
operand, 13
operator, 23, 61
  arithmetical, 97
  assignment, 11
  binary, 50, 52
  boolean, 13
  differential, 97
  divisibility, 59
  equivalence, 54
  inclusion, 59
  logical, 49, 97
  membership, 59
  relational, 97
  unary, 52
order, 35, 176
ordered pair, 13
ordering, 78
  partial, 71
ordinate, 19

pairwise disjoint, 12
parameter, 6, 97, 104, 178, 191, 193, 214
partition, 13
Peano, 157
  axioms, 154
periodic, 80
Pi, π, *see* Archimedes' constant
piecewise, 86, 87, 126
pigeon-hole principle, *see* Dirichlet's box principle
point, 96
polynomial, 33, 41
  homogeneous, 33
  multivariate, 33
  univariate, 33
postulate, *see* axiom
power series, 34, 112
power set, 14, 35, 47, 97
predicate, 16, 57, 60, 62, 97, 111, 172
prime, 18, 99, 115, 134, 137, 138, 167, 172
primitive, 9
product, 17
  infinite, 33
  symbol, 32
product of sets, *see* algebraic product
proof, 121
  both ends method, 135, 162, 169
  by cases, 125, 168
  by contradiction, 134, 141, 154, 156, 158, 169
  by contrapositive, 129–131
  by induction, 153
  direct, 128
  of conjunction, 131
proper
  divisor, 18
  subset, 12
proposition, 121
punctuation, 4, 108

quadratic irrational, 41
quantifier, 49
  existential, 58, 62, 172

  universal, 58, 62, 127
quotient, 17

range, *see* image
rational
  function, 33, 41
  number, 15, 100
  point, 19
ray, 19
real
  function, 77
  line, 17
  number, 16, 100
  part, 17
  sequence, 77
reciprocal, 17, 26, 40, 81
recursive
  definition, 113
  sequence, 113
reduced form, 15
regular, 83
relation, 70
  anti-symmetric, 71
  equivalence, 70
  reflexive, 70
  symmetric, 70
  transitive, 70
relational
  expression, 49, 50, 70
  operator, 49, 50
representation, 27, 28, 83
restriction, 25
Riemann
  hypothesis, 137
  zeta-function, 101
root of polynomial, 34, 83, 170
Russell-Zermelo paradox, 15

sequence, 29, 101
  doubly-infinite, 30, 177
  finite, 29–31
  infinite, 29, 31
  invertible, 177
  periodic, 177
  recursive, 176
series, 32
  convergent, 32
  divergent, 32
  geometric, 147
  sum of, 32
set, 10, 39, 99
  abstract, 27, 83
  bounded, 81
  finite, 12
  infinite, 12
  ordered, 71
  partially ordered, 71
set difference, 12, 173
set equation, 35, 48
set operator, 13, 97
sigma-notation, 104
sign function, 80
singular, singularity, 83
solution set, 34
space, 96
square root, 41, 123, 143
standard definition, 14, 15
step function, 86
subscript, 29, 30, 99
subsequence, 30, 102
subset, 11, 59
  operator, *see* inclusion operator
sum, 17
sum of sets, *see* algebraic sum
sum-free set, 35
summation, 104
  index, 32

  limits, 32
  range, 32
  standard form, 107
  symbol, 32
surjective, 25, 64
symmetric difference, 12

tautology, 55
term, 29, 96
theorem, 50, 121
  conditional, 137
thesis, 183
transitive, 123
triangular number, 112
triple, 24, 30, 47
trivial, 145
  equivalence, 70
  subset, 12
truth table, 52, 54
twin-primes conjecture, 137, 139, 182

unary, 97
uncountable, 27
union, 12
unique existence, 171
unit
  circle, 19
  cube, 19, 169
  disc, 19
  hypercube, 19
  interval, 19
  sphere, 32
  square, 19
univariate, 33
unknown, 34, 96, 100
URL, 196

value
  of a function, 23, 24
  of an expression, 37
vanishing, 34
  identically, 34
variable, 23, 96
  dummy, 104
  random, 119
vector, 30, 101, 102
  dimension, 30
vector space, 96

Weierstrass, 101
well-
  behaved, 83
  defined, 174
  ordered, 71
  ordering principle, 154
Wilson's theorem, 132
without loss of generality, 81, 148, 181

Zermelo definition, 16, 34, 56, 62, 109
zero
  function, 35
  of equation, 34

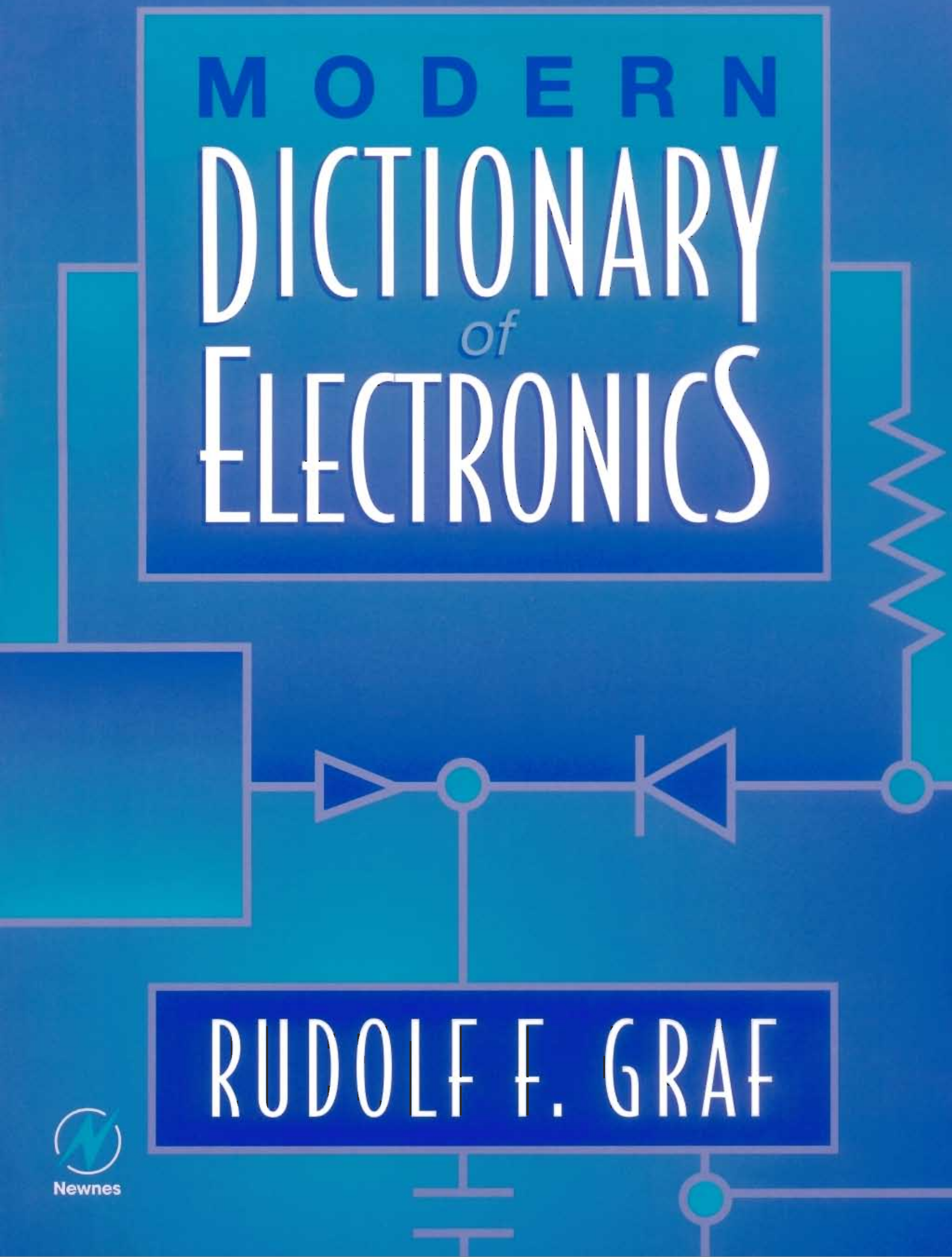

SEVENTH EDITION

MODERN DICTIONARY of ELECTRONICS

RUDOLF F. GRAF

Newnes

MODERN
DICTIONARY
of
ELECTRONICS

SEVENTH EDITION

REVISED AND UPDATED

Rudolf F. Graf is an author whose name is well-known to engineers, technicians, and hobbyists around the world. He graduated as an electronics engineer from Brooklyn Polytechnic Institute and did his graduate work New York University. Mr. Graf has been active in the electronics industry for more than fifty years in capacities ranging from design and consulting engineer, chief engineer, chief instructor at electronics and television schools, and consulting editor. He also held various sales and marketing positions. Mr. Graf is the author or co-author of more than 150 technical articles published by major magazines. He has written about 50 books on electricity and electronics, with more than 2 million copies in print, including the best-selling *Video Scrambling & Descrambling for Satellite & Cable TV, Second Edition* and the *Circuits* series of books, both published by Newnes. A number of his books have been translated into several European languages as well as Chinese, Japanese, and Russian.

MODERN
DICTIONARY
of
ELECTRONICS

SEVENTH EDITION

REVISED AND UPDATED

Rudolf F. Graf

Newnes

Boston Oxford Auckland Johannesburg Melbourne New Delhi

Newnes is an imprint of Butterworth-Heinemann.

Copyright © 1999 by Rudolf F. Graf

A member of the Reed Elsevier Group.

All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.

Recognizing the importance of preserving what has been written, Butterworth-Heinemann prints its books on acid-free paper whenever possible.

Butterworth-Heinemann supports the efforts of American Forests and the Global ReLeaf program in its campaign for the betterment of trees, forests, and our environment.

Library of Congress Cataloging-in-Publication Data

Graf, Rudolf F.
Modern dictionary of electronics / Rudolf F. Graf. — 7th ed., revised and updated.
p. cm.
ISBN 0-7506-9866-7 (alk. paper)
1. Electronics — Dictionaries. I. Title
TK7804.G67 1999
621.381'03 — dc21 99-17889
CIP

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

The publisher offers special discounts on bulk orders of this book.
For information, please contact:
Manager of Special Sales
Butterworth-Heinemann
225 Wildwood Avenue
Woburn, MA 01801-2041
Tel: 781-904-2500
Fax: 781-904-2620

For information on all Butterworth-Heinemann publications available, contact our World Wide Web home page at: http://www.bh.com

10 9 8 7 6 5 4 3 2 1

Typeset by Laser Words, Madras, India
Printed in the United States of America

It gives me great pleasure to dedicate this edition to
Allison, Sheryl, Daniel, David, Russell and Scott,
the loveliest children this side of heaven.

Preface

When the first edition of this dictionary was published in 1961, today's everyday items like color TVs, VCRs, CD players, computers, FAX machines, ATMs, cordless and cell phones, pagers, tape recorders, digital watches, pocket calculators, lasers, and many others too numerous to mention, were non-existent or mere laboratory curiosities. Since then, electronics has undergone significant changes based primarily on the meteoric expansion of integrated circuits and their apparently limitless applications. Vacuum tubes were replaced by semiconductors, and numerous technologies like ferrite core or bubble memories were relegated to the electronics graveyard. No other industry has ever grown so much and matured so fast, paced by technological advances that occur at a feverish pace. The first edition of this dictionary contained about 10,000 definitions of then current terms. And now, a scant 38 years later, this seventh edition contains approximately 25,000 terms—a clear indication of the phenomenal growth of our industry.

As technologies evolve and fresh products and concepts are introduced, suitable terminology must be developed to be able to communicate. The originators of the new words give them their initial meaning, but their exact definitions change with technological advances and through actual use by others. The contents of this dictionary is thus an analysis of words and their meanings as determined by common usage, written in a modern and popular style to provide clear and concise explanations of each entry. Continual updating of a work such as this is vital, so that those involved in the world of electronics have the power to communicate with those about them and to grasp new concepts as they emerge.

All entries are allowed as much space as is necessary for complete and meaningful definitions. Terms are explained clearly and precisely without excessive technical jargon. Original entries from the previous edition have been reviewed and many were revised to keep pace with current usage. Where more than one definition exists for a term, they are arranged numerically. This method, however, does not necessarily imply a preferred order of meanings. Important words from formative technologies that are no longer in use are retained in this edition for their historical interest.

My thanks go out to Ms. Tara Troxler Thomas and to Charles Thomas whose dedication to this project and skill at the word processor made it possible to deliver the manuscript for this work to the publisher in a timely fashion.

Industry and technical sources—notably the IEEE and the ASA—generously aided in defining many terms during the preparation of earlier editions of this work.

While this volume is as up-to-date as possible at the time of writing, the field of electronics is expanding so rapidly that new terms are constantly being developed and older terms take on broader or more specialized meanings. It is the intention of the publisher to periodically issue revised editions of this dictionary; thus suggestions for new terms and definitions are always welcome.

Rudolf F. Graf
February 1999

A

A — 1. Abbreviation for angstrom unit, used in expressing wavelength of light. Its length is 10^{-8} centimeter. 2. Chemical symbol for argon, an inert gas used in some electron tubes. 3. Letter symbol for area of a plane surface. 4. Letter symbol for ampere.

a — Letter symbol for atto- (10^{-18}).

A0 — The Federal Communications Commission (FCC) designation for radio emission consisting solely of an unmodulated carrier.

A1 — The FCC designation for radio emission consisting of a continuous-wave carrier keyed by telegraphy.

A-1 or **A.1** — The atomic time scale maintained by the U.S. Naval Observatory; presently it is based on weighted averages of frequencies from cesium-beam devices operated at a number of laboratories.

A2 — The FCC designation for radio emission consisting of a tone-modulated continuous wave.

A3 — The FCC designation for radio emission consisting of amplitude-modulated speech transmission.

A4 — The FCC designation for radio emission consisting of amplitude-modulated facsimile signals.

A5 — The FCC designation for radio emission consisting of amplitude-modulated television video signals.

A− (A-minus or **A-negative)** — Sometimes called F−. Negative terminal of an A battery or negative polarity of other sources of filament voltage. Denotes the terminal to which the negative side of the filament-voltage source should be connected.

A+ (A-plus or **A-positive)** — Sometimes called F+. Positive terminal of an A battery or positive polarity of other sources of filament voltage. The terminal to which the positive side of the filament voltage source should be connected.

ab- — The prefix attached to names of practical electric units to indicate the corresponding unit in the cgs (centimeter-gram-second) electromagnetic system, e.g., abampere, abvolt, abcoulomb.

abac — *See* alignment chart.

abampere — Centimeter-gram-second electromagnetic unit of current. The current that, when flowing through a wire 1 centimeter long bent into an arc with a radius of 1 centimeter, produces a magnetic field intensity of 1 oersted. One abampere is equal to 10 amperes.

A battery — Source of energy that heats the filaments of vacuum tubes in battery-operated equipment.

abbreviated dialing — 1. A system using special-grade circuits that require fewer than the usual number of dial pulses to connect two or more subscribers. 2. Ability of a phone system to require only two to four digits, while the network dials the balance of the seven to fourteen digits required.

abc — Also ABC. *See* automatic bass compensation.

abcoulomb — Centimeter-gram-second electromagnetic unit of electrical quantity. The quantity of electricity passing any point in an electrical circuit in 1 second when the current is 1 abampere. One abcoulomb is equal to 10 coulombs.

aberration — 1. In lenses, a defect that produces inexact focusing. Aberration may also occur in electron optical systems, causing a halo around the light spot. 2. In a cathode-ray tube, a defect in which the electron "lens" does not bring the electron beam to the same point of sharp focus at all points on the screen. 3. Failure of an optical lens to produce exact point-to-point correspondence between an object and its image. 4. Blurred focusing of light rays due to the difference in bending (refraction) imparted on different light frequencies (colors) as they pass through a lens.

abfarad — Centimeter-gram-second electromagnetic unit of capacitance. The capacitance of a capacitor when a charge of 1 abcoulomb produces a difference of potential of 1 abvolt between its plates. One abfarad is equal to 10^9 farads.

abhenry — Centimeter-gram-second electromagnetic unit of inductance. The inductance in a circuit in which an electromotive force of 1 abvolt is induced by a current changing at the rate of 1 abampere per second. One abhenry is equal to 10^{-9} henry.

abmho — Centimeter-gram-second electromagnetic unit of conductance. A conductor or circuit has a conductance of 1 abmho when a difference of potential of 1 abvolt between its terminals will cause a current of 1 abampere to flow through the conductor. One abmho is equal to 10^9 mho. Preferred term: absiemens.

abnormal glow — In a glow tube, a current discharge of such magnitude that the cathode area is entirely surrounded by a glow. A further increase in current results in a rise in its density and a drop in voltage.

abnormal propagation — The phenomenon of unstable or changing atmospheric and/or ionospheric conditions acting on transmitted radio waves. Such waves are prevented from following their normal path through space, causing difficulties and disruptions of communications.

abnormal reflections — *See* sporadic reflections.

abnormal termination — The shutdown of a computer program run or other process by the detection of an error by the associated hardware, indicating that some ongoing series of actions cannot be executed correctly.

abohm — Centimeter-gram-second electromagnetic unit of resistance. The resistance of a conductor when, with an unvarying current of 1 abampere flowing through it, the potential difference between the ends of the conductor is 1 abvolt. One abohm is equal to 10^{-9} ohm.

abort — 1. To cut short or break off (an action, operation, or procedure) with an aircraft, guided missile, or the like, especially because of equipment failure. An abort may occur at any point from start of countdown or takeoff to the destination. An abort can be caused by human technical or meteorological errors, miscalculation, or malfunctions. 2. The process of halting a computer program in an orderly fashion and returning control to the operator or operating system. 3. Abnormal termination of a computer program, caused by hardware or software malfunction or operator cancellation.

AB power pack—Assembly in a single unit of the A and B batteries of a battery-operated circuit. Also, a unit that supplies the necessary A and B voltages from an ac source of power.

abrasion machine—A laboratory device for determining the abrasive resistance of wire or cable. The two standard types of machines are the squirrel cage with square steel bars and the abrasive grit types.

abrasion resistance—A measure of the ability of a wire or wire covering to resist damage due to mechanical causes. Usually expressed as inches of abrasive tape travel.

abrasion soldering—Soldering difficult metals by abrading the surface oxide film beneath a pool of molten solder.

abrasive trimming—Trimming a ceramic capacitor or a film resistor to its nominal value by notching the surface with a finely adjusted stream of abrasive material such as aluminum oxide.

abscissa—Horizontal, or *x*, axis on a chart or graph.

absence-of-ground searching selector—In dial telephone systems, an automatic switch that rotates, or rises vertically and rotates, in search of an ungrounded contact.

absolute accuracy—1. The tolerance of the full-scale set point referred to as the absolute voltage standard. 2. Parameter for a d/a converter. It is the overall accuracy of the converter, in which all levels are compared with an absolute standard. Absolute accuracy includes the combination of all nonlinearity and end-point errors.

absolute address—1. An address used to specify the location in storage of a word in a computer program, not its position in the program. 2. A binary number assigned permanently as the address of a storage location in a computer. 3. A fixed location in the memory of the CPU, as opposed to a relative address, which is specified according to its distance from another location.

absolute altimeter—1. Electronic instrument that furnishes altitude data with regard to the surface of the earth or any other surface immediately below the instrument, as distinguished from an aneroid altimeter, the readings of which depend on air pressure. 2. An altimeter that employs transmitted and reflected radio waves for its operation and thus does not depend on barometric pressure for its altitude indication.

absolute code—A code using absolute addresses and absolute operation codes; that is, a code that indicates the exact location where the reference operand is to be found or stored.

absolute coding—Coding written in machine language. It can be understood by the computer without processing.

absolute delay—The time interval between the transmission of two synchronized radio, loran, or radar signals from the same or different stations.

absolute digital position transducer—A digital position transducer, the output signal of which is indicative of absolute position. Also called encoder.

absolute efficiency—Ratio of the actual output of a transducer to that of a corresponding ideal transducer under similar conditions.

absolute error—1. The amounts of error expressed in the same units as the quantity containing the error. 2. Loosely, the absolute value of the error, that is, the magnitude of the error without regard to its algebraic sign.

absolute gain of an antenna—The gain in a given direction when the reference antenna is an isotropic antenna isolated in space.

absolute instruction—A computer instruction that explicitly states, and causes the execution of, a specific operation.

absolute language—The language in which instructions must be given to the computer. The absolute language is determined when the computer is designed. Synonyms: machine language, machine code.

absolute loader—Program to load a computer program at specified numerical addresses.

absolute maximum rating—Limiting values of operating and environmental conditions, applicable to any electron device of a specified type as defined by its published data and not to be exceeded under the worst probable conditions. Those ratings beyond which the life and reliability of a device can be expected to decline.

absolute maximum supply voltage—The maximum supply voltage that may be applied without the danger of causing a permanent change in the characteristics of a circuit.

absolute minimum resistance—The resistance between the wiper and the termination of a potentiometer, when the wiper is adjusted to minimize that resistance.

absolute Peltier coefficient—The product of the absolute temperature and the absolute Seebeck coefficient of a material.

absolute power—Power level expressed in absolute units (e.g., watts or dBm).

absolute pressure transducer—1. A pressure transducer that accepts two independent pressure sources simultaneously, and the output of which is proportional to the pressure difference between the sources. 2. A transducer that senses a range of pressures, which are referenced to a fixed pressure. The fixed pressure is normally total vacuum.

absolute scale—*See* Kelvin scale.

absolute Seebeck coefficient—The integral from absolute zero to the given temperature of the quotient of the Thomson coefficient of a material divided by its absolute temperature.

absolute spectral response—Output or response of a device, in terms of absolute power levels, as a function of wavelength.

absolute system of units—Also called coherent system of units. A system of units in which a small number of units is chosen as fundamental, e.g., units of mass, length, time, and charge. Such units are termed absolute units. All other units are derived from them by taking a definite proportional factor in each of those laws chosen as the basic laws for expressing the relationships between the physical quantities. The proportional factor is generally taken as unity.

absolute temperature—Temperature measured from absolute zero, a theoretical temperature level variously defined as −273.2°C, −459.7°F, or 0 K.

absolute temperature scale—Thermodynamic temperature scale, named for Lord Kelvin (1848), in which temperatures are given in kelvins (K). (In the SI system the degree sign and the word *degree* are not used for Kelvin temperatures.) The absolute zero of temperature is 0 K, −273.2°C, or −459.7°F. The kelvin is the same size as the Celsius degree.

absolute tolerance—Also called accuracy. The maximum deviation from the nominal resistance (or capacitance) value, usually given as a percentage of the nominal value.

absolute units—A system of units based on physical principles, in which a small number of units are chosen as fundamental and all other units are derived from them; e.g., abohm, abcoulomb, abhenry, etc.

absolute value—The numerical value of a number or symbol without reference to its algebraic sign. Thus, 3 is the absolute value of $|3|$ or $|-3|$. An absolute value is signified by placing vertical lines around the number or symbol.

absolute value device—A computing element that produces an output equal to the magnitude of the input signal, but always of one polarity.

absolute zero—Lowest possible point on the scale of absolute temperature; the point at which all molecular activity ceases. Absolute zero is variously defined as −273.2°C, −459.7°F, or 0 K.

absorbed wave—A radio wave that becomes lost in the ionosphere due to molecular agitation and the accompanying energy loss it undergoes there. Absorption is most pronounced at low frequencies.

absorber—1. In a nuclear reactor, a substance that absorbs neutrons without reproducing them. Such a substance may be useful in control of a reactor or, if unavoidably present, may impair the neutron economy. 2. Any material or device that absorbs and dissipates radiated energy. 3. In microwave terminology, a material or device that takes up and dissipates radiated energy. It may be used for shielding, to prevent reflection, or to transmit one or more radiation components selectively.

absorption—1. Dissipation of the energy of a radio or sound wave into other forms as a result of its interaction with matter. 2. The process by which the number of particles or photons entering a body of matter is reduced by interaction of the particle or radiation with matter. Similarly, the reduction of the energy of a particle while traversing a body of matter. This term is sometimes erroneously used for capture. 3. Penetration of a substance into the body of another. 4. Conversion of radiant energy into other forms by passage through, or reflection from, matter. 5. The adhesion of a fluid in extremely thin layers to the surfaces of a solid. 6. Reduction in strength of an electromagnetic wave propagating through a medium, determined by dielectric properties of the material.

absorption attenuation—Loss in an optical fiber due primarily to impurities, including metals, such as cobalt, chromium, and iron, as well as OH ions.

absorption circuit—A tuned circuit that dissipates energy taken from another circuit or from a signal source. This effect is especially evident in a resonant circuit such as a wavemeter or wave trap.

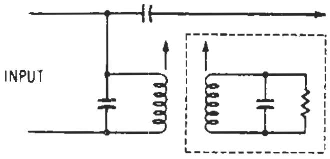

Absorption circuit.

absorption coefficient—1. Measure of sound-absorbing characteristics of a unit area of a given material compared with the sound-absorbing characteristics of an open space (total absorption) having the same area. 2. Ratio of loss of intensity caused by absorption to the total original intensity of radiation.

absorption current—The current flowing into a capacitor following its initial charge, due to a gradual penetration of the electric stress into the dielectric. Also, the current that flows out of a capacitor following its initial discharge.

absorption dynamometer—An instrument for measuring power, in which the energy of a revolving wheel or shaft is absorbed by the friction of a brake.

absorption fading—A slow type of fading, primarily caused by variations in the absorption rate along the radio path.

absorption frequency meter—*See* absorption wavemeter.

absorption loss—1. That part of transmission loss due to dissipation or conversion of electrical energy into other forms (e.g., heat), either within the medium or attendant upon a reflection. 2. The loss of optical flux or energy caused by impurities in the transmission medium as well as intrinsic material absorption. Expressed in decibels per kilometer.

absorption marker—1. A sharp dip on a frequency-response curve due to the absorption of energy by a circuit sharply tuned to the frequency at which the dip occurs. 2. A small pip or blank introduced on an oscilloscope trace to indicate a frequency point. It is so called because it is produced by a frequency-calibrated tuned trap similar to an absorption wavemeter.

absorption modulation—Also called loss modulation. A system for amplitude-modulating the output of a radio transmitter by means of a variable-impedance device (such as a microphone semiconductor or vacuum-tube circuit) inserted into or coupled to the output circuit.

absorption trap—A parallel-tuned circuit coupled either magnetically or capacitively to absorb and attenuate interfering signals.

absorption wavemeter—Also called absorption frequency meter. An instrument for measuring frequency. Its operation depends on the use of a tuned electrical circuit or cavity loosely coupled inductively to the source. Maximum energy will be absorbed at the resonant frequency, as indicated by a meter or other device. Frequency can then be determined by reference to a calibrated dial or chart.

absorptivity—A measure of the portion of incident radiation or sound energy absorbed by a material.

abstraction—A simplified description or specification of a system that emphasizes some of the system's details or properties while suppressing others. A high level of abstraction or a highly abstract machine is one in which very few machine details are apparent to a programmer, who sees only a broad set of machine concepts. Abstract machines are created by surrounding a primitive machine with layers of operating systems. To converse with higher levels of abstract machines, the user needs higher levels of languages.

A/B switch—A switch that selects one of two inputs (A or B) for routing to a common output while providing adequate isolating between the two signals.

A-B test—1. Direct comparison of two sounds by playing first one and then the other. May be done with two tape recorders playing identical tapes (or the same tape), two speakers playing alternately from the same tape recorder, or two amplifiers playing alternately through one speaker, etc. 2. An audio comparison test for evaluating the relative performance of two or more components or systems by quickly changing from one to the other. The left- and right-hand channels or the record and replay sound signals are often designated A and B. A and B test facilities are installed at most high-fidelity dealers.

abvolt—Centimeter-gram-second electromagnetic unit of potential difference. The potential difference between two points when 1 erg of work is required to transfer 1 abcoulomb of positive electricity from a lower to a higher potential. An abvolt is equal to 10^{-8} volt.

ac—Abbreviation for alternating current.

ac bias—The alternating current, usually of a frequency several times higher than the highest signal frequency, that is fed to a record head in addition to the signal current. The ac bias serves to linearize the recording process.

accelerated aging—A test in which certain parameters, such as voltage and temperature, are increased above normal operating values to obtain observable deterioration in a relatively short period. The plotted results give expected service life under normal conditions. Also called accelerated life test.

accelerated graphics port—Abbreviated AGP. A slot inside PCs for high speed video to be used instead of the standard slot type, called PCI.

accelerated life test—Test conditions used to bring about, in a short time, the deteriorating effect obtained under normal service conditions.

accelerated service test—A service or bench test in which some service condition, such as speed, temperature, or continuity of operation, is exaggerated to obtain a result in a shorter time than that which elapses in normal service.

accelerating conductor or relay—A conductor or relay that causes the operation of a succeeding device to begin in the starting sequence after the proper conditions have been established.

accelerating electrode—An electrode in a cathode-ray or other electronic tube to which a positive potential is applied to increase the velocity of electrons or ions toward the anode. A klystron tube does not have an anode but does have accelerating electrodes.

accelerating time—The time required for a motor to reach full speed from a standstill (zero speed) position.

accelerating voltage—A high positive voltage applied to the accelerating electrode of a cathode-ray tube to increase the velocity of electrons in the beam.

acceleration—1. The rate of change in velocity. Often expressed as a multiple of the acceleration of gravity ($g = 32.2\ \text{ft/s}^2$). 2. The rate of change in velocity of a stepping motor measured in rad/s; it is the result of rotor torque divided by rotor and load inertia. 3. A vector quantity that specifies rate of change of velocity.

acceleration at stall—The value of servomotor angular acceleration calculated from the stall torque of the motor and the moment of inertia of the rotor. Also called torque-to-inertia ratio.

acceleration time—In a computer, the elapsed time between the interpretation of instructions to read or write on tape and the possibility of information transfer from the tape to the internal storage, or vice versa.

acceleration torque—Numerical difference between motor torque produced and load torque demanded at any given speed during the acceleration period. It is this net torque that is available to change the speed of the driven load.

acceleration voltage—Potential between a cathode and anode or other accelerating element in a vacuum tube. Its value determines the average velocity of the electrons.

accelerator—A device for imparting a very high velocity to charged particles such as electrons or protons. Fast-moving particles of this type are used in research or in studying the structure of the atom itself. 2. A circuit that speeds up a computer or monitor. Typically a circuit card with an extra processing chip and/or additional RAM.

accelerator board—An adapter with a microprocessor that makes a computer run faster.

accelerator dynamic test—A test performed on an accelerometer by means of which information is gathered pertaining to the overall behavior frequency response and/or natural frequency of the device.

accelerometer—1. An instrument or device, often mounted in an aircraft, guided missile, or the like, used to sense accelerative forces and convert them into corresponding electrical quantities, usually for measuring, indicating, or recording purposes. It does not measure velocity or distance, only changes in velocity. 2. A transducer that measures acceleration and/or gravitational forces capable of imparting acceleration. 3. A sensor whose electrical output is proportional to acceleration.

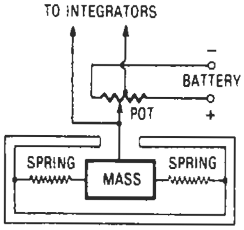

Accelerometer.

accentuation—Also called preemphasis. The emphasizing of any certain band of frequencies, to the exclusion of all others, in an amplifier or electronic device. Applied particularly to the higher audio frequencies in frequency-modulated (FM) transmitters.

accentuator—1. Network or circuit used for preemphasis, that is, an increase in amplitude of a given band of usually audible frequencies. 2. A circuit or device, such as a filter, tone control, or equalizer, used to emphasize a band of frequencies, usually in the audio-frequency spectrum.

acceptable-environmental-range test—A test to determine the range of environmental conditions for which an apparatus maintains at least the minimum required reliability.

acceptable quality level—Abbreviated AQL. The maximum percentage of defective components considered to be acceptable as an average for a process or the lowest quality a supplier is permitted to present continually for acceptance. *Also see* AQL.

acceptance angle—1. The solid angle within which all incident light rays will enter the core of an optical fiber. Expressed in degrees. 2. In fiber optics, a measure of the maximum angle within which light may be coupled from a source or emitter. It is measured relative to the fiber's axis. 3. The critical angle, measured from the core centerline, above which light will not enter an optical fiber. It is equal to the half-angle of the acceptance cone. 4. The maximum angle within which light will be accepted by an element, such as a detector.

acceptance cone—1. A parameter that defines acceptable light-launching angles. Only light launched at angles within this cone will be waveguided (fiber optics). 2. A cone with an included angle twice that of the acceptance angle.

acceptance pattern—In fiber optics, a curve of total transmitted power plotted against the launch angle.

acceptance sampling plan—A plan for the inspection of samples as a basis for acceptance or rejection of a lot.

acceptor—Also called acceptor impurity. An impurity lacking sufficient valence electrons to complete the bonding arrangement in the crystal structure. When added to a semiconductor crystal, it accepts an electron from a neighboring atom and thus creates a hole in the lattice structure of the crystal, making a p-type semiconductor. 2. An impurity from column III of the periodic table, which adds a mobile hole to silicon, thereby making it more p-type and accepting of electrons. Boron is the primary acceptor used to dope silicon (compare with *donor*).

acceptor circuit — 1. A circuit that offers minimum opposition to a given signal. 2. A circuit tuned to respond to a single frequency.

acceptor impurity — *See* acceptor.

acceptor-type semiconductor — A p-type semiconductor.

access — 1. To gain access to a computer's memory location in which binary information is already stored or can be stored. 2. To open up a set of connections to allow reading from or writing into this location.

access arm — In a computer storage unit, a mechanical device that positions the reading and writing mechanism.

access code — 1. The preliminary digit or digits that a telephone user must dial to be connected to a particular outgoing trunk group. 2. A group of characters or numbers that identifies a user to a computer or any other secure system. 3. One or more numbers and/or symbols that are keyed into the repeater with a telephone tone pad to activate a repeater function, such as an autopatch.

access control — 1. The control of pedestrian and vehicular traffic through entrances and exits of a protected area or premises. 2. The process of limiting access to resources of a system to only authorized users, programs, processes, or other systems.

access grant — Multiprocessor system response that satisfies a previous service request.

access hole — A hole drilled through successive layers of a multilayer board to gain access to a land or pad location on one of the inside layers.

access method — 1. A data-management technique available for use in transferring data between the main storage and an input/output device. 2. A software component of a computer operating system that controls the flow of data between application programs and either local or remote peripheral devices.

access mode — 1. A technique used in COBOL to obtain a specific logic record from, or to place it into, a file assigned to a mass storage device. 2. The operation of an alarm system such that no alarm signal is given when the protected area is entered; however, a signal may be given if the sensor, annunciator, or control unit is tampered with or opened.

accessory card — An additional circuit card that can be mounted inside a personal computer and connected to the system bus.

access protocol — A defined set of procedures that function as an interface between a computer user and a network, enabling the user to employ the services of that network.

access provider — Telecommunications company that links businesses and individuals to the Internet using modem devices, high-speed ISDN lines, or dedicated links.

access time — Also called waiting time. 1. The time interval (called read time) between the instant of calling for data from a storage device and the instant of completion of delivery. 2. In a memory system, the time delay, at specified thresholds, from the presentation of an enable or address input pulse until the arrival of the memory data output. 3. The time required for a computer to move data between its memory section and its CPU. 4. A time interval that is characteristic of a storage device. Essentially, it is a measure of the time required to communicate with that device, or, more specifically, it is the time between the application of a specified input pulse (assuming that other necessary inputs are also present) and the availability of valid data signals at an output. The access time can be defined only with reference to an output signal. 5. The time required by a computer to begin delivering information after the memory or storage has been interrogated. 6. The time it takes a computer to retrieve a piece of information. With hard disks or compact discs, maximum access time is measured as the time it takes to move from one end of the disk to the other, find a piece of information, and transfer that information to RAM.

accidental jamming — Jamming caused by transmission from friendly equipment.

ac circuit breaker — A device that is used to close and interrupt an ac power circuit under normal conditions or to interrupt this circuit under faulty or emergency conditions.

accompanying audio (sound) channel — Also known as co-channel sound frequency. The rf carrier frequency that supplies the sound to accompany a television picture.

ac component — In a complex wave (i.e., one containing both ac and dc), the alternating, fluctuating, or pulsating member of the combination.

accordion — A type of contact used in some printed-circuit connectors. The contact spring is given a *z* shape to permit high deflection without excessive stress.

ac-coupled flip-flop — A flip-flop that changes state when triggered by the rise or fall of a clock pulse. There is a maximum allowable rise or fall time for proper triggering.

ac coupling — Coupling of one circuit to another circuit through a capacitor or other device that passes the varying portion but not the static (dc) characteristics of an electrical signal.

accumulation key — In a calculator, it automatically accumulates products and totals of successive calculations.

accumulator — 1. In an electronic computer, a device which stores a number and which, on receipt of another number, adds the two and stores the sum. An accumulator may have properties such as shifting, sensing signals, clearing, complementing, etc. 2. A chemical cell able to store electrical energy (British). Also called secondary cell. 3. The "scratch pad" section of the computer, in which arithmetic operations are carried out. 4. A register and related circuitry that hold an operand for arithmetic and logic operations. 5. A register or latch internal to the MPU where data is stored temporarily before being sent to another location internal or external to the MPU chip.

accuracy — 1. The maximum error in the measurement of a physical quantity in terms of the output of an instrument when referred to the individual instrument calibration. Usually given as a percentage of full scale. 2. The quality of freedom from mistake or error in an electronic computer, that is, of conformity to truth or to a rule. 3. The closeness with which a measured quantity approaches the true value of that quantity. (*See* true value.) 4. The degree to which a measured or calculated value conforms to the accepted standard or rule. 5. The measure of a meter's ability to indicate a value corresponding to the absolute value of electrical energy applied. Accuracy is expressed as a percentage of the meter's rated full-scale value. To be meaningful, accuracy specifications must always consider the effects of time, temperature, and humidity. 6. Confidence in the correlation between measurements in one location and another, or between a measurement and a recognized standard. 7. The correctness or certainty of position when the rotor of a stepping motor comes to rest. It is usually expressed as a percentage of the step angle, but can also be specified in degrees or minutes of arc. In steppers, the error is not cumulative, but occurs only at the completion of the last step. 8. The degree of freedom from error, that is, the degree of conformity to some standard. Accuracy is contrasted with *precision*. For example, four-place numbers are less precise than six-place numbers; however, a properly computed four-place number might be more accurate than an improperly computed six-place number. 9. As applied to an adc, the term describes the difference between the actual input voltage and the full-scale weighted equivalent of the binary code.

accuracy rating of an instrument — The limit, usually expressed as a percentage of full-scale value, not exceeded by errors when the instrument is used under reference conditions.

ACD — Abbreviation for automatic call distributor. A switching system that automatically distributes incoming calls to a centralized group of receivers in the sequence in which the calls are received. It holds calls until a receiver is available.

ac/dc — Electronic equipment capable of operation from either an ac or dc primary power source. Abbreviation for alternating current/direct current.

ac/dc receiver — A radio receiver designed to operate directly from either an ac or a dc source.

ac/dc ringing — A method of telephone ringing in which alternating current is used to operate a ringing device, and direct current is used to aid the action of a relay that stops the ringing when the called party answers.

ac directional overcurrent relay — A device that functions on a desired value of ac overcurrent flowing in a predetermined direction.

ac dump — The intentional, accidental, or conditional removal of all alternating-current power from a system or component. An ac dump usually results in the removal of all power, since direct current is usually supplied through a rectifier or converter.

ac erasing head — In magnetic recording, a device using alternating current to produce the magnetic field necessary for removal of previously recorded information.

acetate — A basic chemical compound in the mixture used to coat recording discs.

acetate base — The transparent plastic film that forms the tough backing for acetate magnetic recording tape.

acetate disc — A mechanical recording disc, either solid or laminated, made mostly from cellulose nitrate lacquer plus a lubricant.

acetate tape — A sound-recording tape with a smooth, transparent acetate backing. One side is coated with an oxide capable of being magnetized.

ac generator — 1. A rotating electrical machine that converts mechanical power into alternating current. Also known as an alternator. 2. A device, usually an oscillator, designed for the purpose of producing alternating current.

A channel — One of two stereo channels, usually the left.

achieved reliability — Reliability determined on the basis of actual performance of nominally identical items under equivalent environmental conditions. Also called operational reliability.

achromatic — 1. In color television, a term meaning a shade of gray from black to white, or the absence of color (without color). 2. Black-and-white television, as distinguished from color television. 3. Literally, color free. In an optical system, the term is used when chromatic aberration is corrected for at least two wavelengths. A color that is defined as being achromatic is often referred to as gray. 4. Having no color; being a neutral such as black, white, or gray.

achromatic lens — A lens that has been corrected for chromatic aberration. Such a lens is capable of bringing all colors of light rays to approximately the same point of focus by combining a concave lens of flint glass with a convex lens of crown glass. A lens that transmits light without separating it into its constituent colors.

achromatic locus — Also called achromatic region. On a chromaticity diagram, an area that contains all points representing acceptable reference white standards.

achromatic region — *See* achromatic locus.

acicular — Needle-shaped; descriptive of the shape of the magnetizable particles composing the coating of a recording tape. Modern tapes are premagnetized during the coating process to line the "needles" up with the direction of the tape, thus providing maximum sensitivity from the oxide.

acid — A chemical compound that dissociates and forms hydrogen ions when in aqueous solution.

acid depolarizer — An acid, such as nitric acid, sometimes introduced into a primary cell to prevent polarization.

acid fluxes — Fluxes consisting of inorganic acids and salts, which are used when a surface to be soldered is below the ideal for rapid wetting. Also called corrosive fluxes.

acknowledge — A control signal used to complete a handshaking sequence in telecommunications. The acknowledge signal indicates that the information has been accepted by the receiving computer.

ac line — A power line delivering alternating current only.

ac line filter — A filter designed to dissipate or bypass to ground any extraneous signals or electrical noise on an ac power line, while causing virtually no reduction of the power-line voltage or power. Used to keep unwanted signals and noise out of sensitive equipment.

aclinic line — Also called isoclinic line. On a magnetic map, an imaginary line that connects points of equal magnetic inclination or dip.

ac magnetic biasing — In magnetic recording, the method used to remove random noise and/or previously recorded material from the wire or tape. This is done by introducing an alternating magnetic field at a substantially higher frequency than the highest frequency to be recorded.

ac noise — Noise that displays a rate of change that is fast relative to the response capability of the device.

ac noise immunity — A measure of a logic circuit's ability to maintain the prescribed logic state in the presence of such noise. It is defined in terms of the amplitude and pulse width of an input noise signal to which the element will not respond.

acorn tube — A button- or acorn-shaped vacuum tube with no base, designed for UHF applications. Electrodes are brought out through the glass envelope on the side, top, and bottom.

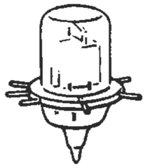

Acorn tube.

acoustic—Also acoustical. Pertaining to sound or the science of sound.

acoustic absorption loss—The energy lost by conversion into heat or other forms when sound passes through or is reflected by a medium.

acoustic absorptivity—The ratio of sound energy absorbed by a surface to the sound energy arriving at the surface. Equal to 1 minus the reflectivity of the surface.

acoustical—*See* acoustic.

acoustical attenuation constant—The real part of the acoustical propagation constant. The commonly used unit is the neper per section or per unit distance.

acoustical coupler—A device for connecting a telephone handset to a computer input port.

acoustical-electrical transducer—A device designed to transform sound energy into electrical energy and vice versa.

acoustical material—Any material considered in terms of its acoustical properties; especially, a material designed to absorb sound.

acoustical mode—A mode of crystal-lattice vibration that does not produce an oscillating dipole.

acoustical ohm—A measure of acoustic resistance, reactance, or impedance. One acoustical ohm is equal to a volume velocity of 1 cubic centimeter per second when produced by a sound pressure of 1 microbar.

acoustical phase constant—The imaginary part of the acoustical propagation constant. The commonly used unit is the radian per section or per unit distance.

acoustical reflectivity—*See* sound-reflection coefficient.

acoustical transmittivity—*See* sound-transmission coefficient.

acoustic burglar alarm—Also called acoustic intrusion detector. A burglar alarm that is responsive to sounds produced by an intruder. Concealed microphones connected to an audio amplifier trip an alarm when sounds within a predetermined range of frequencies exceed a predetermined normal level.

acoustic capacitance—In a sound medium, a measure of volume displacement per dyne per square centimeter. The unit is centimeter to the fifth power per dyne.

acoustic clarifier—A system of cones loosely attached to the baffle of a speaker and designed to vibrate and absorb energy during sudden loud sounds, thereby suppressing them.

acoustic compliance—1. The measure of volume displacement of a sound medium when subjected to sound waves. 2. That type of acoustic reactance which corresponds to capacitive reactance in an electrical circuit.

acoustic coupler—1. A device that converts digital signals into audio signals, enabling data to be transmitted over the telephone lines via a conventional telephone. 2. A modem device that connects a terminal or computer to the handset of a telephone.

acoustic coupling—Coupling resonator elements by mechanical means through the use of wires, rods, or nonelectroded sections of quartz or ceramic. The terms *acoustic* and *mechanical* can be used interchangeably.

acoustic delay line—A device that retards one or more signal vibrations by causing them to pass through a solid or liquid.

acoustic dispersion—The change of the speed of sound with frequency.

acoustic elasticity—1. The compressibility of the air in a speaker enclosure as the cone moves backward. 2. The compressibility of any material through which sound is passed.

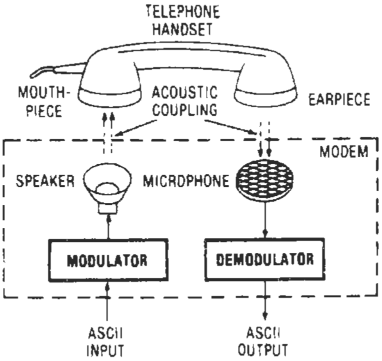

Acoustic coupler.

acoustic feedback—1. Also called acoustic regeneration. The mechanical coupling of a portion of the sound waves from the output of an audio-amplifying system to a preceding part or input circuit (such as the microphone) of the system. When excessive, acoustic feedback will produce a howling sound in the speaker. 2. The pickup, by a turntable, of vibrations from the speaker. If these vibrations reach the cartridge, they will be reamplified, causing noise (usually a rumble, but in extreme cases a howl) and/or distortion. Also, feedback resulting from such sound waves setting some part of an amplifier circuit into vibration and thus modulating the currents in the circuit. Acoustic feedback usually causes howling or whistling.

acoustic filter—1. A sound-absorbing device that selectively suppresses certain audio frequencies while allowing others to pass. 2. Any sound-absorbing or sound-transmitting arrangement, or combination of the two, that passes sound waves of desired frequency while attenuating or eliminating others.

acoustic frequency response—The voltage-attenuation frequency measured into a resistive load, producing a bandwidth approaching sufficiently close to the maximum.

acoustic generator—A transducer, such as a speaker, headphones, or a bell, that converts electrical, mechanical, or other forms of energy into sound.

acoustic homing system—1. A system that uses a sound signal for guidance purposes. 2. A guidance method in which a missile homes in on noise generated by a target.

acoustic horn—Also called horn. 1. A tube of varying cross section having different terminal areas that change the acoustic impedance to control the directivity of the sound pattern. 2. A tapered tube (round or rectangular, but generally funnel shaped) that directs sound and, to some extent, amplifies it.

acoustic impedance—1. Total opposition of a medium to sound waves. Equal to the force per unit area on the surface of the medium divided by the flux (volume velocity or linear velocity multiplied by area) through that surface. Expressed in ohms and equal to the mechanical impedance divided by the square of the surface area. One unit of acoustic impedance is equal to a volume velocity of 1 cubic centimeter per second produced by a pressure of 1 microbar. Acoustic impedance contains both acoustic resistance and acoustic reactance. 2. The degree of resistance to transmitting sound imparted by the characteristic elasticity of a given substance.

acoustic inertance—A type of acoustic reactance that corresponds to inductive reactance in an electrical circuit. (The resistance to movement or reactance offered by the sound medium because of the inertia of the effective mass of the medium.) Measured in acoustic ohms.

acoustic intensity — The limit approached by the quotient of acoustical power being transmitted at a given time through a given area divided by the area as the area approaches zero.

acoustic interferometer — An instrument for measuring the velocity or frequency of sound waves in a liquid or gas. This is done by observing the variations of sound pressure in a standing wave, established in the medium between a sound source and a reflector, as the reflector is moved or the frequency is varied.

acoustic intrusion detector — *See* acoustic burglar alarm.

acoustic labyrinth — A loudspeaker enclosure in which the rear of the loudspeaker is coupled to a tube which, at the resonant frequency of the loudspeaker, is one quarter of a wavelength long. The tube, folded upon itself in order to save space, gives the appearance of a labyrinth.

acoustic lens — 1. An array of obstacles that refract sound waves in the same way that an optical lens refracts light waves. The dimensions of these obstacles are small compared with the wavelengths of the sounds being focused. 2. A device that produces convergence or divergence of moving sound waves. When used with a loudspeaker, the acoustic lens widens the beam of the higher-frequency sound waves.

acoustic line — Mechanical equivalent of an electrical transmission line. Baffles, labyrinths, or resonators are placed at the rear of a speaker to help reproduce the very low audio frequencies.

acoustic memory — A computer memory using an acoustic delay line. The line employs a train of pulses in a medium such as mercury or quartz.

acoustic mine — Also called sonic mine. An underwater mine that is detonated by sound waves, such as those from a ship's propeller or engines.

acoustic mirage — The distortion of a sound wavefront by a large temperature gradient in air or water. This creates the illusion of two sound sources.

acoustic ohm — The unit of acoustic resistance, reactance, or impedance. One acoustic ohm is present when a sound pressure of 1 dyne per square centimeter produces a volume velocity of 1 cubic centimeter per second.

acoustic phase constant — The imaginary part of the acoustic propagation constant. The commonly used unit is the radian per section or per unit distance.

acoustic phonograph — A mechanical record player (now obsolete) in which the needle sets a thin diaphragm into vibration. The diaphragm in turn causes the air in a horn to vibrate, thus reproducing the recorded sound.

acoustic pickup — 1. In nonelectrical phonographs, the method of reproducing the material on a record by linking the needle directly to a flexible diaphragm. 2. In an acoustic phonograph, a pickup consisting of a needle, needle holder, and vibrating diaphragm.

acoustic radiator — In an electroacoustic transducer, the part that initiates the radiation of sound vibration. A speaker cone or headphone diaphragm is an example.

acoustic radiometer — An instrument for measuring sound intensity by determining the unidirectional steady-state pressure caused by the reflection or absorption of a sound wave at a boundary.

acoustic reactance — That part of acoustic impedance due to the effective mass of the medium, that is, to the inertia and elasticity of the medium through which the sound travels. The imaginary component of acoustic impedance; expressed in acoustic ohms.

acoustic reflectivity — The ratio of the rate of flow of sound energy reflected from the surface on the side of incidence to the incident rate of flow.

acoustic refraction — A bending of sound waves when passing obliquely from one medium to another in which the velocity of sound is different.

acoustic regeneration — *See* acoustic feedback.

acoustic resistance — That component of acoustic impedance responsible for the dissipation of energy due to friction between molecules of the air or other medium through which sound travels. Measured in acoustic ohms and analogous to electrical resistance.

acoustic resonance — An increase in sound intensity as reflected waves and direct waves that are in phase combine. May also be due to the natural vibration of air columns or solid bodies at a particular sound frequency.

acoustic resonator — An enclosure that intensifies those audio frequencies at which the enclosed air is set into natural vibration.

acoustics — 1. Science of production, transmission, reception, and effects of sound. 2. In a room or other location, those characteristics that control reflections of sound waves and thus the sound reception in it.

acoustic scattering — The irregular reflection, refraction, or diffraction of a sound wave in many directions.

acoustic shock — Physical pain, dizziness, and sometimes nausea brought on by hearing a loud, sudden sound. (The threshold of pain is about 120 dBm.)

acoustic surface-wave component — A passive electroacoustic device that has metallized interdigital transducer elements deposited on the surface of a piezoelectric substrate. The device allows acoustic energy to be generated, manipulated, and detected on the substrate surface. Most of the acoustic energy is confined to a region within one wavelength of the surface of the substrate. When the metallization is subjected to an alternating voltage, a strain develops between the interdigital fingers and also at the frequency of excitation. This alternating strain on the crystal surface launches a Rayleigh surface-wave front that travels in both directions and that originates from the center of the transducer. The wave exists as an electroacoustic vibration.

acoustic suspension — 1. A loudspeaker system in which the moving cone is held by an overcompliant suspension, the stiffness required for proper operation being supplied by air that is trapped behind the cone in a sealed enclosure. While relatively inefficient, such a system permits good bass reproduction in a unit of moderate size. 2. A speaker enclosure design in which the speaker cone is suspended in an airtight box. This enables the acoustic pressure of the air enclosed therein to provide the principal restoring force for the diaphragm of the speaker. It needs somewhat more power from the amplifier than a free speaker but has better low-frequency performance.

acoustic system — Arrangement of components in devices designed to reproduce audio frequencies in a specified manner.

acoustic telemetry — The utilization of sound energy for the transmission of information. It differs from other telemetry methods in that information derived from the received signal is encoded by the transmitting source.

acoustic transformer — A device that transmits power along a glass or ceramic rod and isolates the power supply from the signal input.

acoustic transmission — Direct transmission of sound energy without the intermediary of electric currents.

acoustic transmission system — An assembly of elements adapted for the transmission of sound.

acoustic treatment — Use of certain sound-absorbing materials to control the amount of reverberation in a room, hall, or other enclosure; that is, to make the room less live.

acoustic wave — A traveling vibration by which sound energy is transmitted in air, water, or the earth. The characteristics of these waves may be described in terms of change of pressure, particle displacement, or density.

acoustic wave filter — A device designed to separate sound waves of different frequencies. (Through electroacoustic transducers, such a filter may be associated with electric circuits.)

acoustoelectric effect — Generation of an electric current in a crystal by a traveling longitudinal sound wave.

acousto-optic Bragg cell — A modulation device that impresses analog information on light beams. This transducer is composed of two sets of interleaved electrodes of alternating polarities deposited on an optical waveguide. An electrical signal applied to each pair of adjacent electrodes buckles the film between them. This distortion changes the refractive index of the waveguide and creates physical waves in the film, commonly called surface waves. These waves are generated at a rate equal to that of the applied electrical signal.

acousto-optics — The study of the interactions between sound waves and light in a solid medium. Sound waves can be made to modulate, deflect, and focus light waves — an important characteristic in laser and holographic applications.

ac plate resistance — Also called dynamic plate resistance. Internal resistance of a vacuum tube to the flow of alternating current. Expressed in ohms, the ratio of a small change in plate voltage to the resultant change in plate current, other voltages being held constant.

ac power supply — A power supply that provides one or more ac output voltages, e.g., ac generator, dynamotor, inverter, or transformer.

acquisition — 1. The process of pointing an antenna or telescope so that it is properly oriented to allow gathering of tracking or telemetry data from a satellite or space probe. 2. In radar, the process between the initial location of a target and the final alignment of the tracking equipment on the target. 3. The gathering of data from transducers or a computer.

acquisition and tracking radar — A radar set that locks onto a strong signal and tracks the object emitting or reflecting the signal. May be airborne or on the ground. Tracking radars use a dish-type antenna reflector to produce a searchlight-type beam.

acquisition radar — A radar set that detects an approaching target and feeds approximate position data to a fire-control or missile-guidance radar, which then takes over the function of tracking the target.

acquisition range — Also called capture range. The range of input frequency about f_0 under which a phase-locked loop, which is initially unlocked, will become locked. This range is narrower than the normal tracking range and is a function of the loop-filter characteristics and the input amplitude.

acquisition time — 1. Time delay between request for data conversion and the holding of the analog value by a sample-and-hold amplifier. 2. In a sample-and-hold circuit, how long it takes after the sample command is given for the hold capacitor to be charged to a full-scale voltage change and to remain within a specified error band around its final value. 3. The time it takes for the output of a sample-and-hold circuit to change from its previous value to a new value when the circuit is switched from the hold mode to the sample mode. It includes the slew time and settling time to within a certain error band of the final value and is usually specified for a full-scale change.

ac receiver — A radio receiver designed to operate from an ac source only.

ac reclosing relay — A device that controls the automatic reclosing and locking out of an ac circuit interrupter.

ac relay — A relay designed to operate from an alternating-current source.

ac resistance — Total resistance of a device in an ac circuit. *See also* high-frequency resistance.

acronym — A word formed from the first letter or letters of the words describing some item, e.g., FORTRAN from *for*mula *tran*slation.

across-the-line starting — Connection of a motor directly to the supply line for starting. Also called full-voltage starting.

ac signaling — Using ac signals or tones to transmit data and/or control signals.

ACTCRBS — Abbreviation for air traffic control radar beacon system. A control system in use worldwide. Air separation information exchanged between plane and air traffic controller must be sent by radio.

ac time overcurrent relay — A device that has either a definite or an inverse time characteristic and functions when the current in an ac circuit exceeds a predetermined value.

actinic — In radiation, the property of producing a chemical change, such as the photographic action of light.

actinium — A radioactive element discovered in pitchblende by the French chemist Debierne in 1889. Its atomic number is 89, its atomic weight 227, and its symbol Ac.

actinodielectric — A photoconductive dielectric.

actinoelectric — Exhibiting a temporary rise in electrical conductivity during exposure to light.

actinoelectric effect — 1. The property of some special materials whereby when an electric current is impressed on them, their resistance changes with light. 2. The property of certain materials (such as selenium, cadmium sulfide, germanium, and silicon) that causes them to change their electrical resistance or generate a voltage on exposure to light.

actinoelectricity — Electricity produced by the action of radiant energy on crystals.

actinometer — An instrument that measures the intensity of radiation by determining the amount of fluorescence produced by that radiation.

action area — In the rectifying junction of a metallic rectifier, that portion which carries the forward current.

action current — A brief and very small electric current that flows in a nerve during a nervous impulse.

action potential — 1. The instantaneous value of the voltage between excited and resting portions of an excitable living structure. 2. The voltage variations in a nerve or muscle cell when it is excited or fired by an appropriate stimulus. After a short time, the cell recovers its normal resting potential, typically about 80 millivolts. The interior of the cell is negative relative to the outside.

activate — To start an operation, usually by application of an appropriate enabling signal.

activating — 1. Chemically treating a basic metal to remove oxides and other passive films to make it more receptive to electroplating. 2. A treatment that renders nonconductive material receptive to electroless deposition. (Nonpreferred synonyms: seeding, catalyzing, and sensitizing.)

activation — 1. Making a substance artificially radioactive by placing it in an accelerator such as a cyclotron or by bombarding it with neutrons. 2. To treat the cathode or target of an electron tube in order to create or increase its emission. 3. The process of adding electrolytes to a cell to make it ready for operation. 4. Causing the acceleration of a chemical reaction.

activation time—In a cell or battery, the time interval from the moment activation is initiated to the moment the desired operating voltage is obtained.

activator—An additive that improves the action of an accelerator.

active—1. Controlling power from a separate supply. 2. Requiring a power supply separate from the controls. 3. Containing, or connected to and using, a source of energy.

active area—The portion of the rectifying junction of a metallic rectifier that carries forward current.

active balance—In operation of a telephone repeater, the summation of all return currents at a terminal network balanced against the local circuit or drop impedance.

active circuit—A circuit that contains active elements such as transistors, diodes, or integrated circuits.

active communications satellite—A communications satellite in which on-board receivers and transmitters receive signals beamed at them from a ground terminal, amplify them greatly, and retransmit them to another ground terminal. Less sensitive receivers and less powerful transmitters can be used on the ground than are needed for passive satellites. Also called active comsat.

active component—1. Those components in a circuit that have gain, or direct current flow, such as SCRs, transistors, thyristors, or tunnel diodes. They change the basic character of an applied electrical signal by rectification, amplification, switching, and so forth. (Passive elements like inductors, capacitors, and resistors have no gain characteristics.) 2. A device, the output of which is dependent on a source of power other than the main input signal. 3. A device capable of some dynamic function (such as amplification, oscillation, signal control) and which usually requires an external power supply for its operation. 4. Broadly, any device (including electromechanical relays) that can switch (or amplify) by application of low-level signals.

active computer—The one of two or more computers in an installation that is online and processing data.

active comsat—*See* active communications satellite.

active current—In an alternating current, a component in phase with the voltage. The working component as distinguished from the idle or wattless component.

active decoder—A device that is associated with a ground station and automatically indicates the radar beacon reply code that is received in terms of its number or letter designation.

active delay line—A digital delay module that incorporates a passive delay line and a series of logical gate circuits. These modules are used specifically with digital or logic signals. Also called digital delay line, digital delay unit, digital delay module, and digital programmable delay line.

active device—*See* active component.

active display—A display, such as a cathode-ray tube, electroluminescent display, or plasma panel, that presents information by emitting light.

active ECM—*See* jamming.

active electric network—An electric network containing one or more sources of energy.

active element—1. The driven or self-excited element in a multielement antenna or antenna array. 2. Also known as the responsive element. That part of a detector on which the infrared energy is projected and which, when radiation falls on it, undergoes a physical change that results in an electrical signal. *See* active component.

active equalizer—An equalizer designed to correct deficiencies in a speaker system's response. Such equalizers, which are designed to precisely match specific speaker systems, usually connect between the amplifier and preamplifier, or in one of the amplifier's tape-monitor circuits.

active filter—1. A device employing passive network elements and amplifiers. It is used for transmitting or rejecting signals in certain frequency ranges or for controlling the relative output of signals as a function of frequency. 2. A high-pass, low-pass, bandpass, or band-elimination filter that uses an active element, such as an operational amplifier, and relatively small capacitors, rather than the larger inductors and capacitors that would be required in a conventional passive filter. 3. A circuit whose gain depends on the frequency of the input signal. 4. A filter, consisting of an amplifier and suitable tuning elements, usually inserted in a feedback path. 5. A filter that uses active devices such as operational amplifiers to synthesize the filter response function. This technique has an advantage at high speeds because the need for inductors (with their poor high-frequency characteristics) is eliminated.

active guidance—*See* active homing.

active homing—A system whereby a missile homes in on a target by means of a radar aboard the missile. Also called active guidance.

active infrared detection—An infrared detection system in which a beam of infrared rays is transmitted toward one or more possible targets, and the rays reflected from the target are detected.

active infrared system—A system in which the object is irradiated by a source of infrared energy, which, in turn, is reflected by the object onto a detector. A snooperscope is an active infrared system.

active intrusion sensor—An active sensor that detects the presence of an intruder within the range of the sensor. Examples are an ultrasonic motion detector, a radio-frequency motion detector, and a photoelectric alarm system. *See also* passive intrusion sensor.

active jamming—1. Intentional radiation or reradiation of electromagnetic waves to impair the use of a specific portion of the electromagnetic-wave spectrum. 2. Transmission or retransmission of signals for the express purpose of disrupting communications.

active junction—In a semiconductor, a change in n-type to p-type doping, or vice versa, by a diffusion step. On discrete transistors there are two active junctions, the collector-base junction and the emitter-base junction.

active leg—Within a transducer, an electrical element that changes its electrical characteristics as a function of the applied stimulus.

active line—In a U.S. television picture, one of the lines (approximately 488) that make up the picture. The remaining 37 of the 525 available lines are blanked; they are called inactive lines.

active maintenance downtime—The time during which work is actually being done on an item, from the recognition of an occurrence of failure to the time of restoration to normal operation. This includes both preventive and corrective maintenance.

active material—1. In the plates of a storage battery, lead oxide or some other active substance that reacts chemically to produce electrical energy. 2. The fluorescent material, such as calcium tungstate, used on the screen of a cathode-ray tube.

active matrix—A display matrix with a transistor at each pixel location to individually store its state (on or off). Pixels in active matrix panels only need to be addressed when they are being turned on or off.

active mixer and modulator — A device requiring a source of electrical power and using nonlinear network elements to heterodyne or combine two or more electrical signals.

active network — 1. A network containing passive and active (gain) elements. 2. An electrical network that includes a source of energy.

active pressure — In an ac circuit, the pressure that produces a current, as distinguished from the voltage impressed on the circuit.

active probe — A test probe, generally used with an oscilloscope, that is so named because of the active components used within probe circuitry. These components consist of one or all of the following: transistors, diodes, integrated circuits, or FETs. If FETs are used, these probes are often referred to as FET probes.

active pull-up — An arrangement in which a transistor is used to replace the pull-up resistor in an integrated circuit in order to provide low output impedance without high power consumption.

active *RC* network — A network formed by resistors, capacitors, and active elements.

active redundancy — That redundancy wherein all redundant items are operating simultaneously rather than being switched on when needed.

active repair time — That portion of corrective maintenance downtime during which repair work is being done on the item, including preparation, fault location, part replacement, adjustment and recalibration, and final test time. It may also include part procurement time under shipboard or field conditions.

active satellite — A satellite that receives, regenerates, and retransmits signals between stations. *See also* communications satellite.

active sensor — A sensor that detects the disturbance of a radiation field that is generated by the sensor. *See also* passive sensor.

active sonar — *See* sonar.

active splitter — *See* line splitter.

active substrate — 1. A substrate in which active elements are formed to provide discrete or integrated devices. Examples of active substrates are single crystals of semiconductor materials within which are transistors, resistors, and diodes, or combinations of these elements. Another example is ferrite substrates within which electromagnetic fields are used to perform logical, gating, or memory functions. 2. A substrate for an integrated component in which parts display transistance. 3. A working part of the electronic circuit, which it supports physically. 4. In an integrated circuit, a substrate consisting of single-crystal semiconductor material into which the various IC components are formed; it acts as some or all of the components. This is in contrast to a substrate consisting of a dielectric, on whose surface the various components are deposited.

active swept-frequency interferometer radar — A dual radar system for air surveillance. It provides angle and range information of high precision for pinpointing target locations by trigonometric techniques.

active systems — In radio and radar, systems that require transmitting equipment, such as a beacon or transponder, to be carried in the vehicle.

active tracking system — Usually, a system that requires the addition of a transponder or responder on board the vehicle to repeat or retransmit information to the tracking equipment; e.g., dovap, secor, azusa.

active transducer — 1. A type of transducer in which its output waves depend on one or more sources of power, apart from the actuating waves. 2. A transducer that requires energy from local sources in addition to that which is received.

active trim — Trimming of a circuit element (usually resistors) in a circuit that is electrically activated and operating to obtain a specified functional output for the circuit. *See* functional trimming.

active wire — The wire of an armature winding that produces useful voltage. That portion of the winding in which induction takes place.

activity — 1. In a piezoelectric crystal, the magnitude of oscillation relative to the exciting voltage. 2. The intensity of a radioactive source. 3. Operations that result in the use or modification of the information in a computer file.

activity curve — A graph showing how the activity of a radioactive source varies with time.

activity ratio — The ratio of the number of records in a computer file that have activity to the total number of records in the file.

ac transducer — A transducer that, for proper operation, must be excited with alternating currents only. Also a device, the output of which appears in the form of an alternating current.

actual height — The highest altitude at which refraction of radio waves actually occurs.

actual power — The average of values of instantaneous power taken over one cycle.

actuating device — A mechanical or electrical device, either manual or automatic, that operates electrical contacts to bring about signal transmission.

actuating system — 1. In a device or vehicle, a system that supplies and transmits energy for the operation of a mechanism or other device. 2. A manually or automatically operated mechanical or electrical device that operates electrical contacts to effect signal transmission.

actuating time — The time at which a specified contact functions.

actuator — 1. In a servo system, the device that moves the load. 2. The part of a relay that converts electrical energy into mechanical motion. 3. Switch part to which an external force is applied to operate the switch. 4. A manual or automatic switch or sensor, such as a holdup button, magnetic switch, or thermostat, that causes a system to transmit an alarm signal when manually activated or when the device automatically senses an intruder or other unwanted condition. 5. A motorized arm that moves a satellite dish into position under the control of a receiver.

ACU — Abbreviation for automatic calling unit.

ac voltage — *See* alternating voltage.

acyclic machine — A direct-current machine in which the voltage generated in the active conductors maintains the same direction with respect to those conductors at all times.

a/d — Abbreviation for analog-to-digital. Also a-d, A-D, or A/D.

adapter — 1. A fitting designed to change the terminal arrangement of a jack, plug, socket, or other receptacle, so that other than the original electrical connections are possible. 2. An intermediate device that permits attachment of special accessories or provides special means for mounting. 3. A device for connecting two parts of an apparatus that would not be directly connectable because of incompatible dimensions, terminations, currents, voltages, frequencies, etc.

adaptive communication — A method in which automatic changes in the communications system allow for changing inputs or changing characteristics of the device or process being controlled. Also called self-adjusting communication or self-optimizing communication.

adaptive control — 1. A control method that uses sensors for real-time measurement of process variables with calculation and adjustment of control parameters as a method of achieving near-optimum process performance. 2. A method of control in which actions are continuously adjusted in response to feedback.

adaptive control system — A device whose parameters are automatically adjusted to compensate for changes in the dynamics of the process to be controlled. An AFC circuit utilizing temperature-compensating capacitors to correct for temperature changes is an example.

adaptive telemetry — Telemetry having the ability to select certain vital information or any change in a given signal.

adaptor — A device that locates and supports products to be tested. Generally, it is made of an insulating material with locator pins mounted to precisely position the product to a spring contact probe test pattern. Also, an adaptor serves as an intermediate between the circuit verifier and the interchangeable test head that contains the test pattern.

adc — Abbreviation for analog-to-digital converter. Also ADC.

Adcock antenna — A pair of vertical antennas separated by one-half wavelength or less and connected in phase opposition to produce a figure-8 directional pattern.

Adcock direction finder — A radio direction finder using one or more pairs of Adcock antennas for directional reception of vertically polarized radio waves.

Adcock radio range — A type of radio range utilizing four vertical antennas (Adcock antennas) placed at the corners of a square, with a fifth antenna in the center.

a/d converter — Abbreviation for analog-to-digital converter. Also a-d, A-D, or A/D converter. 1. A unit or device that converts an analog signal, that is, a signal in the form of a continuously variable voltage or current, to a digital signal. 2. A circuit that accepts information in a continuously varying ac or dc current or voltage and whose output is the same information in digital form. 3. A circuit or device for producing a set of digital output signals representing the magnitude of a voltage applied to its input.

add-and-subtract relay — A stepping relay capable of being operated so as to rotate the movable contact arm in either direction.

addend — A quantity that, when added to another quantity (called the augend), produces a result called the sum.

adder — 1. A device that forms the sum of two or more numbers or quantities impressed on it. 2. In a color TV receiver, a circuit that amplifies the receiver primary signal coming from the matrix. Usually there is one adder circuit for each receiver primary channel. 3. An arrangement of logic gates that adds two binary digits and produces sum and carry outputs.

add-in — Components (expansion boards, cartridges, or chips) that can increase a computer's capabilities, such as memory, graphics, and communications. Add-ins usually refer to an entire circuit board. *See also* add-on.

add-in memory — Additional computer memory that is added to a computer system within the computer's physical housing. Typically the add-in memory is inserted, in board form, into an available card slot on the assembly, the connections for which have already been placed on the existing computer. Additional memory may be in the form of semiconductor RAM, CCD, bubble memory, disk, or tape.

additional station — Any amateur radio station licensed to an amateur radio operator, normally for a specific land location other than the primary station.

addition record — A new record created during the processing of a file in a computer.

additive — Sometimes referred to as the key. A number, series of numbers, or alphabetical intervals added to a code to put it in a cipher.

additive color — A system that combines two colored lights to form a third.

additive primaries — Primary colors that can be mixed to form other colors, but which cannot themselves be produced by mixing other primaries. Red, green, and blue are the primaries in television because, when added in various proportions, they produce a wide range of other colors.

additive process — A printed-circuit manufacturing process in which a conductive pattern is formed on an insulating base by electrolytic chemical deposition.

additive synthesis — A technique for creating musical notes whereby sine waves are added together to create new waveforms. Frequently used in electric organs rather than in synthesizers.

additron — An electrostatically focused, beam-switching tube used as a binary adder in high-speed digital computers. (No longer used.)

add mode — Allows entry of numbers in a calculator to two decimal places without the need to enter the decimal point.

add-on — Circuitry or system that can be attached to a computer to increase memory or performance. *See also* add-in.

add-on component or **add-on device** — A discrete or integrated prepackaged or chip component that is attached to a film circuit to complete the circuit functions.

add-on memory — Additional computer memory that is added externally and is plug compatible with the computer system. The add-on memory is connected with an external connector cable to the computer, where provision has been made for memory expansion. Such a memory device is also available in its own housing, in which case it is physically placed beside the computer's main cabinet. Additional memory may be in the form of semiconductor RAM, CCD, bubble memory, disk, or tape.

address — 1. An expression, usually numerical, that designates a specific location in a storage or memory device or other source or destination of information in a computer. 2. An identification, as represented by a name, label, or number, for a register, location in storage, or any other data source or destination, such as the location of a station in a communications network. 3. Loosely, any part of an instruction that specifies the location of an operand for the instruction. 4. To select the location of a stored information set for access. 5. In computer technology, a number used by the central processing unit (CPU) to specify a location in memory. 6. Element(s) of a packet frame that identifies the source and/or destination stations by means of an agreed bit pattern. 7. A unique sequence of letters or numbers for the location of data or the identity of an intelligent device.

address bus — 1. A unidirectional bus over which digital information appears to identify either a particular memory location or a particular device. 2. The set of output pins from a microprocessor chip and the associated circuitry linking them to other devices for the purpose of addressing those chips or parts of them. *See also* bus system.

address characters — Blocks of alphanumeric characters that identify users or stations uniquely.

address comparator — In a computer, a device that ensures that an address being read is the right one.

address computation — The process by which the address part of an instruction in a digital computer is produced or modified.

address constant — *See* base address.

addressed memory—In a computer, memory sections containing each individual register.

address field—The portion of an instruction that specifies the location of a particular piece of information in a computer memory.

addressing mode—An addressing method. One of several different addressing methods possible in microprocessors.

address modification—In a computer, a change in the address portion of an instruction or command such that, if the routine which contains that instruction or command is repeated, the computer will go to a new address or location for data or instructions.

address part—In an electronic computer instruction, a portion of an expression designating location. *See also* instruction code.

address-routing indicator—Group of characters contained in a message heading that designates the destination of the message.

add-subtract time—The time required by a digital computer to perform addition or subtraction. It does not include the time required to obtain the quantities from storage and put the result back into storage.

add time—The time required in a digital computer to perform addition. It does not include the time required to obtain the quantities from storage and put the result back into storage.

a/d encoder—Analog-to-digital encoder. A device that changes an analog quantity into an equivalent digital representation. Also referred to as an a-d, A-D, or A/D encoder.

adf—*See* automatic direction finder. Also referred to as ADF.

adiabatic damping—A reduction in the size of an accelerator beam as the energy of the beam is increased.

adiabatic demagnetization—A technique used to obtain temperatures within thousandths of a degree of absolute zero. It consists of applying a magnetic field to a substance at a low temperature and in good thermal contact with its surroundings, insulating the substance thermally, and then removing the magnetic field.

A-display—Also called A-scan. A radarscope presentation in which time (distance or range) is one coordinate (usually horizontal) and the target appears displaced perpendicular to the time base.

adjacency—In character recognition, a condition in which the character-spacing reference lines of two characters printed consecutively on the same line are less than a specified distance apart.

adjacent- and alternate-channel selectivity—A measure of the ability of a receiver to differentiate between a desired signal and signals that differ in frequency from the desired signal by the width of one channel or two channels, respectively.

adjacent audio (sound) channel—The rf carrier frequency that contains the sound modulation associated with the next-lower-frequency television channel.

adjacent channel—That frequency band immediately above or below the one being considered.

adjacent-channel attenuation—*See* selectance.

adjacent-channel interference—Undesired signals received on one communication channel from a transmitter operating on a channel immediately above or below.

adjacent-channel selectivity—The ability of a receiver to reject signals on channels adjacent to the channel of the desired station.

adjacent conductor—Any conductor next to another conductor, either in the same multiconductor cable layer or in adjacent layers.

adjacent sound channel—In television, the rf channel containing the sound signal modulation of the next lower channel.

adjacent video carrier—The rf carrier that carries the picture modulation for the television channel immediately above the channel to which the viewer is tuned.

adjustable component—Any circuit component whose electrical value may be varied at will, e.g., adjustable capacitor, inductor, resistor, or load.

adjustable resistor—1. A resistor that has the resistance wire partly exposed to enable the amount of resistance in use to be adjusted occasionally by the user. Adjustment requires the loosening of a screw, the subsequent moving of the lug, and retightening of the screw. 2. A fixed resistor with a movable contact (or tap) that can be positioned along the length of the resistive path.

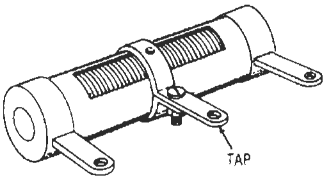

Adjustable resistor.

adjustable voltage divider—A wirewound resistor with one or more movable terminals that can be slid along the length of the exposed resistance wire until the desired voltage values are obtained.

adjusted circuit—Also called bolted-fault level. In a circuit, the current measured under short-circuit conditions with the leads that are normally connected to the circuit breaker bolted together.

adjusted decibels—An expression of the ratio of the noise level to a reference noise at any point in a transmission system, when the noise meter has been adjusted to allow for the interfering effect under specified conditions.

admittance—1. The ease with which an alternating current flows in a circuit. The reciprocal of impedance; usually expressed in siemens. Symbol is Y or y. 2. The (sinusoidal) current in a circuit divided by the terminal voltage. 3. The vector sum of a resistive component of conductance and a reactive component of susceptance.

ADP—Abbreviation for automatic data processing.

ADSL—Abbreviation for Asymmetric Digital Subscriber Line. A video display terminal distribution video system delivering video over existing (i.e., copper) telephone lines.

adsorption—The deposition of a thin layer of gas or vapor particles onto the surface of a solid. The process is known as chemisorption if the deposited material is bound to the surface by a simple chemical bond.

ADU—Abbreviation for automatic dialing unit.

advance ball—In mechanical recording, a rounded support (often sapphire) that is attached to a cutter and rides on the surface of the recording medium. Its purpose is to maintain a uniform mean depth of cut and to correct for small irregularities on the surface of the disc.

advance calling—A telecommunications feature in which voice messages can be spoken into the telephone for automatic delivery at a prearranged time to any other telephone or telephones.

advanced license—A license issued by the FCC to amateur radio operators who are capable of sending and receiving Morse code at the rate of 13 words per minute and are familiar with general and intermediate radio theory and practice. Its privileges include exclusive use of certain frequencies.

advance wire—An alloy of copper and nickel, used in the manufacture of electric heating units and some wirewound resistors.

aeolight—A glow lamp that uses a cold cathode and a mixture of inert gases. Its illumination can be regulated with an applied signal voltage and it is often used as a modulation indicator for motion picture sound recording.

aerial—*See* antenna.

aerial cable—A cable installed on a pole line or similar overhead structure.

aerodiscone antenna—An aircraft antenna that is aerodynamically shaped and is physically small compared with other antennas having similar electrical characteristics. Its radiation pattern is omnidirectional and linearly polarized.

aerodrome control radio station—A radio station providing communications between an airport control tower and aircraft or mobile aeronautical radio stations.

aerodynamics—The science of the motion of air and other gases. Also, the forces acting on bodies such as aircraft when they move through such gases, or when such gases move against or around the bodies.

aeromagnetic—Pertaining to the magnetic field of the earth as surveyed from the air.

aeronautical advisory station—A station used for civil defense and advisory communications with private aircraft stations.

aeronautical broadcasting service—The broadcasting service intended for the transmission of information related to air navigation.

aeronautical broadcast station—A station that broadcasts information regarding air navigation and meteorological data pertinent to aircraft operation.

aeronautical fixed service—A fixed service intended for the transmission of information relating to air navigation and preparation for and safety of flight.

aeronautical fixed station—A station operating in the aeronautical fixed service.

aeronautical ground station—A radio station operated for the purpose of providing air-to-ground communications in connection with the operation of aircraft.

aeronautical marker-beacon station—A land station operating in the aeronautical radionavigation service and providing a signal to designate a small area above the station.

aeronautical mobile service—A radio service between aircraft and land stations or between aircraft stations.

aeronautical radio-beacon station—An aeronautical radionavigation land station transmitting signals that are used by aircraft and other vehicles to determine their position bearing or position in relation to the aeronautical radio beacon station.

aeronautical radionavigation service—A radionavigation service intended for use in the operation of aircraft.

aeronautical radio service—1. Service carried on between aircraft stations and/or land stations. 2. Special radio for air navigation. 3. Service that includes aircraft-to-aircraft, aircraft-to-ground, and ground-to-aircraft communications important to the operation of aircraft.

aeronautical station—A land station (or in certain instances a shipboard station) in the aeronautical mobile service that carries on communications with aircraft stations.

aeronautical telecommunication—Electronic and nonelectronic communications used in the aeronautical service.

aeronautical telecommunication agency—An agency to which is assigned the responsibility for operating a station or stations in the aeronautical telecommunication service.

aeronautical telecommunication log—A record of the activities of an aeronautical telecommunication station.

aeronautical telecommunications—Any telegraph or telephone communications of signals, writing, images, and sounds of any nature by wire, radio, or other system or process of signaling, used in the aeronautical service.

aeronautical telecommunication service—Telecommunication service provided for aeronautical purposes.

aeronautical telecommunication station—A station in the aeronautical telecommunication service.

aeronautical utility land station—A land station located at an airport control tower and used for communications connected with the control of ground vehicles and aircraft on the ground.

aeronautical utility mobile station—A mobile station used at an airport for communications with aeronautical utility land stations, ground vehicles, and aircraft on the ground.

aerophare—*See* radio beacon.

AES—Abbreviation for Audio Engineering Society. A professional group; the official association of technical personnel, scientists, engineers, and executives in the audio field.

AF—*See* audio frequency.

AFC—*See* automatic frequency control.

afocal—An optical system with one set of object and image points at infinity. Literally, "without a focal length." An afocal system receives its input image from infinity and projects its output image to infinity.

AFSK—Abbreviation for audio-frequency shift keying. With this method of modulation, two tones (mark = 2125 Hz, space = 2295 Hz) are fed directly into the microphone jack of the transmitter.

afterglow—Also called phosphorescence. 1. The light that remains in a gas-discharge tube after the voltage has been removed, or on the phosphorescent screen of a cathode-ray tube after the exciting electron beam has been removed. 2. The luminosity that remains in a rarefied gas after an electrodeless discharge has traversed the gas.

afterpulse—In a photomultiplier, a spurious pulse induced by a preceding pulse.

AGC—*See* automatic gain control.

age—To maintain an electrical component in a specified environment, as with respect to pressure, temperature, applied voltage, etc., until its characteristics stabilize.

aggregate function—A command that performs calculations based on a set of values rather than on a single value.

agile receiver—A satellite receiver that can be tuned to any desired channel.

aging—1. Storing a permanent magnet, capacitor, semiconductor, meter, or other device, sometimes with voltage applied, until its desired characteristics become essentially constant. 2. The change of a component or a material with time under defined environmental conditions, leading to improvement or deterioration of properties.

agonic line—An imaginary line on the earth's surface, all points of which have zero magnetic declination.

AGP—Abbreviation for accelerated graphics port.

AGREE—Advisory Group on Reliability of Electronics Equipment.

AI—*See* artificial intelligence.

aided tracking—A system of tracking a target signal in bearing, elevation, or range (or any combination of these variables) in which manual correction of the tracking error automatically corrects the rate at which the tracking mechanism moves.

AIEE—Abbreviation for American Institute of Electrical Engineers. Now merged with IRE to form IEEE.

air bearing—A means of supporting magnetic tape on an air film rather than by means of a sliding or rolling contact. Usually, an air bearing is a perforated cylinder; pressurized air flows through the perforations and forms a film that prevents the tape from contacting the cylinder.

airborne intercept radar—Short-range airborne radar employed by fighter and interceptor planes to track down their targets.

airborne long-range input—Airborne equipment designed to extend air-surveillance coverage seaward so that long-range interceptors may be used.

airborne moving-target indicator—A type of airborne-radar display that does not present essentially stationary objects.

airborne noise—Undesired sound in the form of fluctuations of air pressure about the atmospheric pressure as a mean.

airborne radar platform—Airborne surveillance and height-finding radar for early warning and control.

air capacitor—A capacitor in which air is the only dielectric material between its plates.

aircarrier aircraft station—A radio station aboard an aircraft that is engaged in or essential to the transportation of passengers or cargo for hire.

air cell—A primary cell in which depolarization at the positive electrode is accomplished chemically by reduction of the oxygen in the air.

air column—The air space within a horn of an acoustic chamber.

air condenser—*See* air capacitor.

air-cooled tube—An electron tube in which the generated heat is dissipated to the surrounding air directly, through metal heat-radiating fins, or with the aid of channels or chimneys that increase the air flow.

air-core cable—A telephone cable in which the interstices in the cable core are not filled with a moisture barrier.

air-core coil—A number of turns of spiral wire in which no metal is used in the center.

air-core transformer—A transformer (usually rf) having two or more coils wound around a nonmetallic core. Transformers wound around a solid insulating substance or on an insulating coil form are included in this category.

aircraft bonding—Electrically connecting together all the metal structure of the aircraft, including the engine and metal covering of the wiring.

aircraft flutter—Flickering (repetitive fading and intensifying) in a TV picture as the signal is reflected from flying aircraft. The reflected signal arrives in or out of phase with the normal signal and thus strengthens or weakens the latter.

aircraft station—A radio station installed on aircraft and continuously subject to human control.

air defense control center—Principal information, communications, and operations center from which all aircraft, antiaircraft operations, air-defense artillery, guided missiles, and air-raid warning functions of a specific area of air defense responsibility are supervised and coordinated.

air defense identification zone—Airspace of defined dimensions within which the ready identification, location, and control of aircraft is required.

air dielectric capacitor—A capacitor with a dielectric consisting of air.

airdrome control station—A station used for communication between an airport control tower and aircraft.

air environment—In communications electronics, all airborne equipment that is part of the communications-electronics system, as distinguished from the equipment on the ground, which belongs to the ground environment.

air gap—1. A nonmagnetic discontinuity in a ferromagnetic circuit. For example, the space between the poles of a magnet—although filled with brass, wood, or any other nonmagnetic material—is nevertheless called an air gap. This gap increases magnetic reluctance and prevents saturation of the core. 2. The air space between two magnetically or electrically related objects.

air/ground control radio station—An aeronautical telecommunication station with the primary responsibility of handling communications related to the operation and control of aircraft in a given area.

air/ground liaison code—Set of symbols for a limited number of words, phrases, and sentences used for communications between air and ground forces.

air-motion transformer—A type of speaker in which the air is not pushed into vibration by a piston, but rather squeezed by the contractions of a folded diaphragm.

air navigation radio aids—Aeronautical ground stations, radio beacons, direction finders, and similar facilities.

airplane flutter rejection—The measure of a receiver's immunity to the effects of wavering signals produced by aircraft in the reception path.

airport beacon—A beacon (light or radio) to indicate the location of an airport.

airport control station—A station that furnishes communications between an airport control tower and aircraft in the immediate vicinity; messages are limited to those related to actual aviation needs.

airport radar control—The surveillance-radar portion of radar approach control.

airport runway beacon—A radio-range beacon that defines one or more approaches to an airport.

airport surface detection equipment—Abbreviated ASDE. 1. Radar that shows the movement of aircraft and other vehicles on the ground at an airport. Valuable tool at night and during low visibility. 2. A digital radar system used to track planes and vehicles on airport runways and up to 200 feet in altitude. Unlike previous surface radar systems, ASDE provides clear images in bad weather.

airport surveillance radar—1. Abbreviated ASR. A short-range radar system that maintains constant surveillance over aircraft at the lower levels of flight. Distinct from air route surveillance radar (ARSR), which is long-range radar—150-mile (241-km) radius—to control traffic between terminals. 2. An air-traffic-control radar that scans the airspace 30 to 60 miles (48 to 98 km) around an airport and displays the location of all aircraft below a certain altitude and all obstructions near the control tower.

air-position indicator—Airborne computing system that presents a continuous indication of aircraft position on the basis of aircraft heading, air speed, and elapsed time.

air-spaced coax — A coaxial cable in which air is basically the dielectric material. The conductor may be centered by means of a spirally wound synthetic filament, by beads, or by braided filaments. This construction is also referred to as an air dielectric.

air surveillance — Systematic observation of airspace by electronic, visual, or other means, primarily to identify and determine the movements of aircraft and missiles, friendly and enemy, in the airspace under observation.

airtime — Time spent on a cellular phone, which is usually billed to the subscriber on a per-minute basis.

air-to-ground communication — Transmission of radio signals from an aircraft to stations or other locations on the earth's surface, as differentiated from ground-to-air, air-to-air, or ground-to-ground communications.

air-to-ground radio frequency — The frequency or band of frequencies agreed upon for transmission from an aircraft to an aeronautical ground station.

air-to-surface missile — A missile designed to be dropped from an aircraft. An internal homing device or the aircraft's radio guides it to a surface target.

airwaves — Slang expression for radio waves used in radio and television broadcasting.

alacritized switch — 1. A mercury switch treated to yield a low adhesional force between the rolling surface and mercury pool, resulting in a decreased differential angle. 2. A mercury switch in which the tendency of the mercury to stick to the mating parts has been reduced.

alarm — A device that signals the existence of an abnormal condition by means of an audible or visible discrete change, or both, intended to attract attention. An alarm circuit produces or transmits an alarm signal.

alarm condition — A threatening condition, such as an intrusion, fire, or holdup, sensed by a detector.

alarm device — A device that signals a warning in response to an alarm condition, such as a bell, siren, or annunciator.

alarm discrimination — The ability of an alarm system to distinguish between those stimuli caused by an intrusion and those which are a part of the environment.

alarm hold — A means of holding an alarm once sensed. The typical magnetic trap does not hold or latch, and thus the reclosing of a trapped door resets the typical magnetic trap. A hold circuit applied to such a device indicates that the door has been opened and continues to so indicate until reset.

alarm line — A wired electrical circuit used for the transmission of alarm signals from the protected premises to a monitoring station.

alarm relay — A relay, other than an annunciator, used to operate, or to operate in connection with, a visual or audible alarm.

alarm state — The condition of a detector that causes a control unit in the secure mode to transmit an alarm signal.

alarm system — An assembly of equipment and devices designated and arranged to signal the presence of an alarm condition requiring urgent attention, such as unauthorized entry, fire, temperature rise, etc. The system may sound a local warning or alert the police, a central station, or a proprietary service.

albedo — The reflecting ability of an object. It is the ratio of the amount of light reflected compared with the amount received.

ALC — Abbreviation for automatic level (volume) control. 1. A special compressor circuit included in some tape recorders for automatically maintaining the recording volume within the required limits regardless of changes in the volume of the sound. 2. A circuit that automatically maintains recording levels within permissible limits, so that, no matter how loud or soft the sound being recorded, the signal on the tape will not get strong enough to overmodulate and distort or soft enough to be lost in noise. Also known as automatic volume control (AVC).

Alexanderson alternator — An early mechanical generator used as a source of low-frequency power for transmission or induction heating. It is capable of generating frequencies as high as 200,000 hertz.

Alexanderson antenna — A vlf antenna consisting of a horizontal wire connected to ground at equally spaced points by vertical wires with base-loading coils; the transmitter is coupled to an end coil.

Alford antenna — A square loop antenna comprising four linear sides with their ends bent inward so that capacitive loading is provided to equalize the current around the loop.

algebraic adder — In a computer, an adder that provides the algebraic rather than arithmetic sum of the entered quantities.

algebraic logic — A calculator mode that permits all calculations to be done in the order in which they are written.

algebraic sum — The sum of two or more quantities combined according to their signs. (Compare with *arithmetic sum.*)

ALGOL — 1. An international problem language designed for the concise, efficient expression of arithmetic and logical processes and the control (iterative, etc.) of these processes. From *algo*rithmic *l*anguage. 2. A high-level language that has a context-free structure.

algorithm — 1. A set of rules or processes for solving a problem in a finite number of steps (for example, a full statement of an arithmetic procedure for finding the value of sin x with a stated precision). *See also* procedure. 2. A series of equations, some of which may state inequalities, that cause decisions to be made and the computational process to be altered based on these decisions. 3. A set of rules or directions for getting a specific output from a specific input. The distinguishing feature of an algorithm is that all vagueness must be eliminated; the rules must describe operations that are so simple and well defined they can be executed by a machine. Furthermore, an algorithm must always terminate after a finite number of steps. 4. An ordered sequence of mathematical steps that always produces the correct answer to a problem, though the solution may be more lengthy than necessary. 5. A set of well-defined procedures for the solution of a problem in a limited number of steps. Algorithms are implemented in a computer by a programmed sequence of instructions.

algorithmically generated pattern — An array of digital data automatically generated by a predetermined software routine or program.

algorithmic language — An arithmetic language by which a numerical procedure may be presented to a computer precisely and in a standard form.

algorithmic pattern generation — Real-time generation of input test patterns during test execution according to specified procedures, formulas, or algorithms. Also refers to procedures or algorithms used in automatic-test-generation software for specific fault sets.

alias — An alternate label. For instance, a label and one or more aliases may be used to identify the same data element or point in a computer program.

aliasing — 1. The introduction of error into the Fourier analysis of a discrete sampling of continuous data when components with frequencies too great to be analyzed with the sampling interval being used contribute to the amplitudes of lower-frequency components. 2. A phenomenon arising as a result of the sampling process in which high-frequency components of the original analog signal (whether information or noise) appear as lower frequencies in the sampled signal. Aliasing occurs when the sampling rate is less than twice the highest frequency existing in the original analog signal. 3. Undesirable distortion component that can arise in digital audio equipment when the input signal's frequency exceeds one-half of the digital circuitry's sampling rate. 4. The mistaking of some object or situation for another, especially because of the way data is examined. Examples: movies of propellers and wagon wheels that seem to turn backward, musical notes that are wrongly analyzed by sequential measurement (for instance, thought to be an octave too low), and, especially, "jaggies." 5. Undesirable stair-step distortions in computer-generated images caused by improper sampling techniques. The most common effect is a jagged edge along object boundaries.

aliasing noise—A distortion component that will be created if a sampled signal bandwidth is effectively greater than one-half the sampling rate.

align—1. To adjust the tuned circuits of a receiver or transmitter for maximum signal response. 2. To put into proper relative position, agreement, or coordination when placing parts of a photomask together or placing a photomask over an etched pattern in the oxide on a semiconductor wafer. 3. To adjust the tuning of a multistage device so that all stages are adjusted to the same frequency or so that they work together properly.

aligned bundle—*See* coherent bundle.

aligned-grid tube—A multigrid vacuum tube in which at least two of the grids are aligned one behind the other to give such effects as beam formation and noise suppression.

alignment—1. The process of adjusting components of a system for proper interrelationship. The term is applied especially to (*a*) the adjustment of tuned circuits in a receiver to obtain the desired frequency response and (*b*) the synchronization of components in a system. 2. In a tape recorder, the physical positioning of a tape head relative to the tape itself. Alignment in all respects must conform to rigid requirements in order for a recorder to function properly. 3. The accuracy or proper relative position of an image on a photomask with respect to an existing image on a substrate, as in a photoresist coating, or etched in the oxide of an oxidized silicon wafer. 4. A technique in the fabrication process of semiconductors by which a series of six to eight masks are successively registered to build up the various layers of a monolithic device. Each mask pattern must be accurately referenced to or aligned to all preceding mask patterns. 5. The accuracy of coordination or relative position of images on a semiconductor oxide coating and on the photomask, or any other images placed in relation to those.

alignment chart—Also called nomograph, nomogram, or abac. Chart or diagram consisting of two or more lines on which equations can be solved graphically. This is done by laying a straightedge on the two known values and reading the answer at the point where the straightedge intersects the scale for the value sought.

alignment pin—1. A pin in the center of the base of a tube. A projecting rib on the pin ensures that the tube is correctly inserted into its socket. 2. Any pin or device that will ensure the correct mating of two components designed to be connected.

alignment protractor—An instrument that indicates error in a pickup's lateral alignment. It fits on the center spindle of the turntable, and the pickup stylus fits into a small hole on the device. The correct indication is shown when the angle of lateral movement of the pickup head is at 90° to the tangent of the groove at any point, although minimal tracking error is expected with most pickup arms.

alignment tool—A special screwdriver or socket wrench used for adjusting trimmer or padder capacitors or cores in tuning inductances. It is usually constructed partly or entirely of nonmagnetic material. *See also* neutralizing tool.

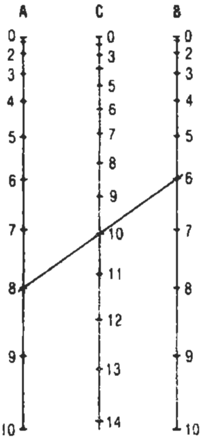

Alignment chart.

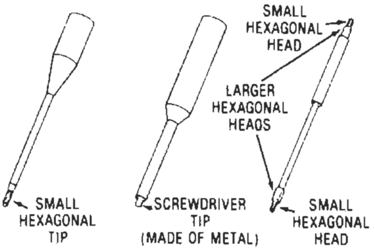

Alignment tools.

alive—1. Electrically connected to a source of potential difference, or electrically charged to have a potential different from that of the earth. 2. Energized. 3. Reverberant, as a room in which sound reflects and echoes.

alive circuit—A circuit that is energized.

alkali—A compound that forms hydroxyl ions when in aqueous solution. Also called a base.

alkaline cell—1. A primary cell, similar to the zinc-carbon cell, in which the negative electrode is granular zinc mixed with a potassium hydroxide (alkaline) electrolyte; the positive electrode is a polarizer in electrical contact with the outer metal can of the cell. A porous separator divides the electrodes. This type of cell delivers a terminal potential of 1.5 volts and has a 50 percent to 100 percent higher capacity than does a 1.5-volt zinc-carbon cell. Also called an alkaline-manganese cell. 2. A primary dry cell that has a very low internal resistance and high

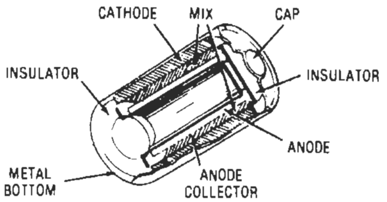

Alkaline cell.

service capacity. It is characterized by a relatively flat discharge curve under load.

all-channel tuning—Ability of a television set to receive all assigned channels, VHF and UHF, channels 2 through 83, as well as cable channels.

all-diffused monolithic integrated circuit—Also called compatible monolithic integrated circuit. A microcircuit consisting of a silicon substrate into which all the circuit parts (both active and passive elements) are fabricated by diffusion and related processes.

Allen screw—A screw having a hexagonal hole or socket in its head. Often used as a setscrew.

Allen wrench—A straight or bent hexagonal rod used to turn an Allen screw.

alligator clip—A spring-loaded metal clip with long, narrow meshing jaws similar to the jaws of an alligator; it is used for making temporary electrical connections, generally at the end of a test lead on interconnection wire.

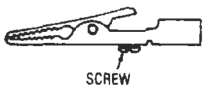

Alligator clip.

allocate—In a computer, to assign storage locations to main routines and subroutines, thus fixing the absolute values of symbolic addresses.

allocated channel—A channel assigned to a specific user.

allocated frequency band—A segment of the radio-frequency spectrum established by a competent authority that designates the use that may be made of the frequencies contained therein.

allocated-use circuit—1. A circuit in which one or more channels have been allocated for the exclusive use of one or more services by a proprietary service; may be a unilateral or joint circuit. 2. Communication link specifically assigned to user(s) warranting such facilities.

allocation technique—The method of providing a process access to a shared resource.

allochromatic—Exhibiting photoelectric effects due to the inclusion of microscopic impurities or as a result of exposure to various types of radiation.

allophone—A variation in the pronunciation of a phoneme. An allophone can be regarded as the sound that results when a phoneme is placed in its environment.

allotter—In a telephone system, a distributor, associated with the finder control group relay assembly, that allots an idle linefinder in preparation for an additional call.

allotter relay—In a telephone system, a relay of the linefinder circuit, the functions of which are to preallot an idle linefinder to the next incoming call from the line and to guard relays.

allowable deviation—The permissible difference between any range of conditions and a reference condition.

alloy—1. A composition of two or more elements, of which at least one is a metal. It may be a solid solution, a heterogeneous mixture, or a combination of both. 2. Method of making pn junctions by melting a metallic dopant so that it dissolves some of the semiconductor material and then hardens to produce a doped alloy.

alloy deposition—The process of depositing an alloy on a substrate during manufacturing.

alloy-diffused transistor—A transistor in which the base is diffused and the emitter is alloyed. The collector is the semiconductor substrate into which alloying and diffusion are effected.

alloyed contact—An ohmic contact formed by an alloy process.

alloy junction—Also called fused junction. A junction produced by alloying one or more impurity metals to a semiconductor. A small button of impurity metal is placed at each desired location on the semiconductor wafer, heated to its melting point, and cooled rapidly. The impurity metal alloys with the semiconductor material to form a p or n region, depending on the impurity used.

alloy-junction photocell—A photodiode in which an alloy junction is produced by alloying (mixing) an indium disc with a thin wafer of n-type germanium.

alloy-junction transistor—Also called fused-junction transistor. A semiconductor wafer of p- or n-type impurities fused, or alloyed, into opposite sides of the wafer to provide emitter and base junctions. The base region comprises the original semiconductor wafer.

alloy process—A fabrication technique in which a small part of the semiconductor material is melted together with the desired metal and allowed to recrystallize. The alloy developed is usually intended to form a pn junction or an ohmic contact.

alloy transistor—A transistor in which the emitter and collector junctions are both alloy junctions.

all-pass filter—A network designed to produce a delay (phase shift) and an attenuation that is the same at all frequencies; a lumped-parameter delay line. Also called all-pass network.

all-pass network—A network designed to introduce phase shift or delay but not appreciable attenuation at any frequency.

all-relay central office—An automatic central-office dial switchboard in which relay circuits are used to make the line interconnections.

all-wave antenna—A receiving antenna suitable for use over a wide range of frequencies.

all-wave receiver—A receiver capable of receiving stations on all the commonly used wavelengths in shortwave bands as well as in the broadcast band.

alnico—An alloy consisting mainly of *al*uminum, *ni*ckel, and *co*balt plus iron. Various subscripts and combinations of letters are available. Material can be found both in cast and sintered form, including isotropic and anisotropic alloys. Capable of very high flux density and magnetic retentivity, the alloy is used in permanent magnets for speakers, magnetrons, etc.

alpha—1. Emitter-to-collector current gain of a transistor connected as a common-base amplifier. For a junction transistor, alpha is less than unity, or 1. Alpha is usually defined as the ratio of a small change in collector current to the corresponding change in emitter current, when the collector-base voltage is kept constant. 2. Brain wave signals whose frequency is approximately 8 to 12 Hz. The associated mental state is relaxation, heightened awareness, elation, and in some cases, dreamlike.

alphabet—An ordered set of all the letters and associated marks used in a language, for example, the Morse code alphabet, the 128 characters of the U.S. ASCII alphabet.

alphabetic coding—A system of abbreviation used in preparing information for input into a computer. Information may then be reported in the form of letters and words as well as in numbers.

alphabetic-numeric—Having to do with the alphabetic letters, numerical digits, and special characters used in electronic data processing work.

alphabetic string—A character string containing only letters and special characters.

alpha cutoff frequency—The frequency at which the current gain of a common-base transistor stage has decreased to 0.707 of its low-frequency value. Gives a rough indication of the useful frequency range of the device.

alphameric (alphanumeric)—Generic term for alphabetic letters, numerical digits, and special characters that are machine processable.

alphameric characters—1. A character set that mixes alphabetic characters, numeric characters, and usually punctuation characters. The alphabetic characters may be uppercase and/or lowercase or even in Japanese or Arabic script. 2. Consisting of letters and numbers. Also called alphameric or alphanumeric.

alphanumeric—1. A generic term for alphabetic letters, numerical digits, and special ASCII characters that can be processed by a computer. A character set containing any combination of the above. 2. Consisting of letters and numbers. 3. All letters in the alphabet, the numbers 0 through 9, and special characters—such as -, /, *, $, (), +, and #—that are machine processable.

alphanumeric code—In computer practice or in communications, a code in which the letters of the alphabet are represented by numbers.

alphanumeric display—Device consisting of a typewriter-style keyboard and a display (CRT) screen on which text is viewed.

alphanumeric keys—Keys on a data entry device that resemble those on a standard keyboard. Usually they are used to manually input or edit text for the display system, although they can also be used in a function key mode.

alphanumeric reader—An instrument that reads alphabetic, numeric, and special characters by means of a photosensor that measures the varying intensity of the characters reflected from a light source.

alphanumeric readout—A type of digital readout that displays both letters and numerals.

alpha particle—A small, electrically charged particle thrown off at very high velocity by many radioactive materials, including uranium and radium. Identical to the nucleus of a helium atom, it is made up of two neutrons and two protons. Its electrical charge is positive and is equal in magnitude to twice that of an electron.

alpha ray—A stream of fast-moving alpha particles that produce intense ionization in gases through which they pass, are easily absorbed by matter, and produce a glow on a fluorescent screen. The lowest-frequency radioactive emissions.

alpha system—A signaling system in which the signaling code to be used is designated by alphabetic characters.

alpha-wave detector—A device that detects and displays alpha-wave segments of brain wave output. Used in biofeedback. Also called alpha-wave meter or sensor.

alpha-wave meter—*See* alpha-wave detector.

ALS—Abbreviation for advanced low-power Schottky (Texas Instruments). A low-power, high-speed transistor-transistor logic (TTL) family.

alterable memory—A storage medium that may be written into.

alteration switch—A manual switch on a computer console or a program-simulated switch that can be set on or off to control coded machine instructions.

alternate channel—A channel located two channels above or below the reference channel.

alternate-channel interference—Interference caused in one communication channel by a transmitter operating in the channel after an adjacent channel. *See also* second-channel interference.

alternate facility—A communications-electronics facility that is established for the purpose of replacing or supplementing another facility or facilities under real or simulated emergency conditions.

alternate frequency—The frequency assigned for use at a certain time, or for a certain purpose, to replace or supplement the frequency normally used.

alternate mode—A means of displaying on an oscilloscope the output signals of two or more channels by switching the channels, in sequence, after each sweep.

alternate route or routing—A secondary or backup communications path to be used if the normal (primary) routing is not possible.

alternate voice/data operation—Modem operations coordinated by voice over the same line that accommodates transmission. The modem is patched out of the circuit to allow this. A special switch, called an exclusion key, converts the line from voice to data.

alternating-charge characteristic—The function relating, under steady-state conditions, the instantaneous values of the alternating component of transferred charge to the corresponding instantaneous values of a specified periodic voltage applied to a nonlinear capacitor.

alternating current—Abbreviated ac. 1. A flow of electricity that reaches maximum in one direction, decreases to zero, then reverses itself and reaches maximum in the opposite direction. The cycle is repeated continuously. The number of such cycles per second is the frequency. The average value of voltage during any cycle is zero. 2. Any signal that varies with time. It usually means that the current actually changes polarity with time. The plot of current versus time usually is a sine wave that comprises a succession of instantaneous values, the greatest of which is the amplitude or peak value. The time taken by one complete cyclic repetition is the period, and the number of periods in one second is the frequency.

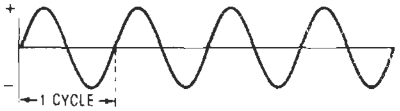

Alternating current.

alternating current/direct current—A term applied to electronic equipment indicating it is capable of operation from either an alternating-current or direct-current primary power source.

alternating-current erasing head—An erasing head used in magnetic recording, in which alternating current produces the magnetic field necessary for erasing. Alternating-current erasing is achieved by subjecting the medium to a number of cycles of a magnetic field of a decreasing magnitude. The medium is, therefore, essentially magnetically neutralized.

alternating-current generator—A rotary machine that generates alternating current when its rotor, which may be either the armature or the field, is rotated by an engine or a motor. Also called an alternator.

alternating-current pulse—An alternating-current wave of brief duration.

alternating-current transmission—In television, that form of transmission in which a fixed setting of the controls makes any instantaneous value of signal correspond to the same value of brightness only for a short time.

alternating flasher—A control that provides voltage first to one load and then to another load. This cycle repeats normally at a fixed rate per minute.

alternating quantity—A periodic quantity that has alternately positive and negative values, the average value of which is zero over a complete cycle.

alternating voltage—Also called ac voltage. Voltage that is continually varying in value and reverses its direction at regular intervals, such as that generated by an alternator or developed across a resistance or impedance through which alternating current is flowing.

alternation—One-half of a cycle—either when an alternating current goes positive and returns to zero, or when it goes negative and returns to zero. Two alternations make one cycle. The complete rise and fall of a current traveling in one direction, or one-half of an alternating-current cycle.

alternator—A device for converting mechanical energy into electrical energy in the form of an alternating current.

alternator transmitter—A radio transmitter that generates power by means of a radio-frequency alternator.

altimeter—An instrument that indicates the altitude of an aircraft above a specific reference level, usually sea level or the ground below the aircraft. It may be similar to an aneroid barometer, which utilizes the change of atmospheric pressure with altitude, or it may be electronic.

altimeter station—An airborne transmitter, the emissions from which are used to determine the altitude of an aircraft above the surface of the earth.

altitude delay—The synchronization delay introduced between the time of transmission of the radar pulse and the start of the trace on the indicator. This is done to eliminate the altitude circle on the plan-position-indicator display.

ALU—Abbreviation for arithmetic and logic unit. 1. A device that performs the basic mathematical operations such as addition, subtraction, multiplication, and division of numbers (usually binary) presented to its inputs and provides an output that is an appropriate function of the inputs. 2. The arithmetic and logic unit internal to the microprocessor chip. This register handles all arithmetic and logical operations carried out as part of a microprocessor instruction. 3. The part of a CPU that executes adds, subtracts, shifts, AND logic operations, OR logic operations, etc. 4. A complex array of gates that can be used to perform binary arithmetic, logic operations, shifts and rotates, and complementing. 5. One of the three essential components of a microprocessor, the other two being data registers and control. The ALU performs addition and subtraction, logic operations, masking, and shifting (multiplication and division).

alumina—1. A ceramic used for insulators in electron tubes or substrates in thin-film circuits. It can withstand continuously high temperatures and has a low dielectric loss over a wide frequency range. Aluminum oxide (AL_2O_3). 2. The substrate material on which are deposited thin conductive and resistive layers for thin-film microwave integrated circuits.

aluminized-screen picture tube—A cathode-ray picture tube that has a thin layer of aluminum deposited on the back of its fluorescent surface to improve the brilliance of the image and also prevent ion-spot formation.

aluminizing—The process of applying a film of aluminum to a surface, usually by evaporation in a vacuum.

aluminum-electrolytic capacitor—A capacitor with two aluminum electrodes (the anode has the oxide film) separated by layers of absorbent paper saturated with the operating electrolyte. The aluminum-oxide film or dielectric is repairable in the presence of an operating electrolyte.

aluminum-steel conductor—A composite conductor made up of a combination of aluminum and steel wires. In the usual construction, the aluminum wires surround the steel.

alumoweld—A thin coating of aluminum fused to a steel core. Used in line wire and cable messengers.

AM—*See* amplitude modulation.

amateur—Also called a ham. 1. A person licensed to operate radio transmitters as a hobby. Any amateur radio operator. 2. A nonprofessional, usually noncommercial, devotee of any technology (as a hobby).

amateur band—Any one of several radio frequency bands assigned for noncommercial use by licensed radio amateurs. In the United States, there are twelve such bands between 1.80 MHz and 1.3 GHz. Assignments are made by the Federal Communications Commission.

amateur call letters—Call letters and numbers assigned to amateur stations by the licensing authority. Call-letter combinations consist of a letter prefix denoting the country in which the station is situated, plus a number designating the location within the country, and two or more letters identifying the particular station. Example: K2ABC (K or W = United States, 2 = New York, and ABC = identification of individual licensee issued alphabetically except for special circumstances).

amateur extra license—A license issued by the FCC to amateur radio operators who are able to send and receive Morse code at the rate of 20 words per minute and who are familiar with general, intermediate, and advanced radio theory and practice. Its privileges include all authorized amateur rights and the exclusive rights to operate on certain frequencies.

amateur radio—The practice of operating electronic communications equipment as a hobby in the amateur service. Also refers to the equipment used for this purpose.

amateur radiocommunication—Noncommercial radiocommunication by or among radio stations solely with a personal aim and without pecuniary or business interest.

amateur radio license—The instrument of authorization issued by the Federal Communications Commission, comprised of a station license and, in the case of the primary station, incorporating an operator license.

amateur radio operation—Radiocommunication conducted by an amateur radio operator from an amateur radio station.

amateur radio operator—1. A person interested in radio technique solely with a personal aim and without pecuniary interest, and holding a valid Federal Communications Commission license to operate amateur radio stations. 2. A private citizen who operates electronic communications equipment as a hobby.

amateur radio service—A radiocommunication service of self-training, intercommunication, and technical investigation carried on by amateur radio operators.

amateur service—A radiocommunication service that licensed operators with no pecuniary interest use for self-training, communication, and technical investigations.

amateur station—A radio transmitting station operated by one or more licensed amateur operators.

amateur-station call letters—*See* amateur call letters.

ambience—1. Reverberant or reflected sound that reaches a listener's ear from all directions as sound waves "bounce" successively off the various surfaces of a listening area—the walls, ceiling, etc. The term is usually reserved for large areas such as auditoriums and concert halls, though home listening-rooms have their own ambience effects. 2. The indirect sounds heard in a concert hall or other large listening area that contribute to the overall auditory effect obtained when listening to live performances.

ambient—Surrounding. The surrounding environment coming into contact with the system or component in question. *See also* ambient noise; ambient temperature.

ambient level—The level of interference emanating from sources other than the test sample, such as inherent noise of the measuring device and extraneous radiated fields.

ambient light—Normal room light. Light existing in a room or other location that is characteristic of the environment.

ambient-light filter—A filter used in front of a television picture-tube screen to reduce the amount of ambient light reaching the screen and to minimize the reflections of light from the glass face of the tube.

ambient lighting—Lighting designed to provide a substantially uniform level of illumination throughout an area, exclusive of any provision for special local requirements.

ambient noise—1. Acoustic noise in a room or other location. Usually measured with a sound-level meter. The term *room noise* commonly designates ambient noise at a telephone station. 2. Unwanted background noise picked up by a microphone, that is, any extraneous clatter in a room. Also any acoustic coloration that influences sounds, brought about by the acoustic properties of a room in which a recording is being made or replayed. 3. Interference present (in a communication line) at all times. 4. Background electrical noise in electrical measurements and operation.

ambient operating temperature—The temperature of the air surrounding an object, neglecting small localized variations.

ambient pressure—The general surrounding atmospheric pressure.

ambient temperature—1. Temperature of air or liquid surrounding any electrical part or device. Usually refers to the effect of such temperature in aiding or retarding removal of heat by radiation and convection from the part or device in question. 2. The prevailing temperature in the immediate vicinity of an object; the temperature of its environment. 3. A temperature within a given volume, e.g., a room or building.

ambient temperature range—The range of environmental temperatures in the vicinity of a component or device over which it may be operated safely and within specifications. For forced-air cooled operation, the ambient temperature is measured at the air intake.

ambiguity—1. An undesirable tendency of a synchro or servo system to seek a false null position in addition to the proper null position. 2. Inherent error resulting from multiple-bit changes in a polystropic code. (Proper logic design prevents such errors.)

ambiguous count—A count on an electronic scaler that is obviously impossible.

ambisonic reproduction—The recreation of the ambience of an original recording situation with associated directionality. Sound from every direction is picked up by a tetrahedral microphone array and is then encoded onto two channels, which, upon decoding, produce sound through several speakers in a continuous range of directions around the listener, thus approximating the original. It can be subdivided into periphonic and pantophonic systems, the former concerning a complete sphere of information, the latter relating to a horizontal circle. Pantophonic reproduction does not distinguish vertical directionality, but still achieves remarkable realism.

AM broadcast channel—Any of the 10-kHz wide bands of radio frequencies, which extend from 530 to 1710 kHz and are used for standard amplitude-modulated radio broadcasts.

American Institute of Electrical Engineers (AIEE)—Now merged with IRE to form the IEEE.

American Morse code—A system of dot-and-dash signals originated by Samuel F. B. Morse and still used to a limited extent for wire telegraphy in North America. It differs from the international Morse code used in radiotelegraph transmission.

American National Standards Institute, Inc.—Abbreviated ANSI. An independent, industry-wide association that establishes standards for the purpose of promoting consistency and interchangeability among the products of different manufacturers. Formerly United States of America Standards Institute (USASI) and American Standards Association (ASA).

American Radio Relay League (ARRL)—An organization of amateur radio operators.

American Standards Association—Abbreviated ASA. *See* American National Standards Institute, Inc.

American wire gage (AWG)—The standard system used for designating wire diameter. Gage sizes range from No. 40, the smallest diameter wire, to No. 4/0, the largest. AWG sizes are used for specifying both solid and stranded wire. Gage numbers have an inverse relationship to size, i.e., larger numbers have smaller diameter.

American wire gage (AWG)

AWG	Diameter (mm)	AWG	Diameter (mm)
1	7.35	21	0.723
2	6.54	22	0.644
3	5.83	23	0.573
4	5.19	24	0.511
5	4.62	25	0.455
6	4.12	26	0.405
7	3.67	27	0.361
8	3.26	28	0.321
9	2.91	29	0.286
10	2.59	30	0.255
11	2.31	31	0.227
12	2.05	32	0.202
13	1.83	33	0.180
14	1.63	34	0.160
15	1.45	35	0.143
16	1.29	36	0.127
17	1.15	37	0.113
18	1.02	38	0.101
19	0.912	39	0.090
20	0.812	40	0.080

AM/FM receiver—A device capable of converting either amplitude- or frequency-modulated signals into audio frequencies.

AM/FM tuner—A device capable of converting either amplitude- or frequency-modulated signals into low-level audio frequencies.

AML—Abbreviation for automatic modulation limiting. A circuit that uses an agc (automatic gain control) effect to prevent overmodulation. As a stronger voice signal is applied, this stage reduces the gain of the audio amplifier(s), keeping the modulation level below 100 percent.

ammeter — An instrument for measuring either direct or alternating electric current. Its scale is usually graduated in amperes, milliamperes, microamperes, or kiloamperes.

ammeter shunt — A low-resistance conductor placed in parallel with the meter movement so that most of the current flows through this conductor and only a small part passes through the movement itself. This arrangement extends the usable range of the meter.

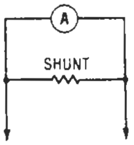

Ammeter shunt.

amorphous — A characteristic, particularly of a crystal, determining that it has no regular structure.

amorphous silicon — A noncrystalline form of silicon used to fabricate transistors on large-area flat displays. Although it is not as good a semiconductor as crystalline silicon, amorphous silicon is much easier to lay down.

amorphous silicon cell — A photovoltaic cell made of silicon and hydrogen atoms deposited in an irregular atomic structure on substrate.

amortisseur winding — *See* damper winding.

amp — Abbreviation for ampere.

ampacity — The maximum current an insulated wire or cable can safely carry without exceeding either the insulation or jacket material limitations. Expressed in amperes. *See also* current-carrying capacity.

amperage — The number of amperes flowing in an electrical conductor or circuit.

ampere — Letter symbol: A. 1. A unit of electrical current or rate of flow of electrons. One volt across 1 ohm of resistance causes a current flow of 1 ampere. A flow of 1 coulomb per second equals 1 ampere. An unvarying current is passed through a solution of silver nitrate of standard concentration at a fixed temperature. A current that deposits silver at the rate of 0.001118 gram per second is equal to 1 ampere, or 6.25×10^{18} electrons per second passing a given point in a circuit. 2. The constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular sections, and placed 1 meter apart in a vacuum will produce between these conductors a force equal to 2×10^{-7} newtons per meter of length.

ampere-hour — A current of 1 ampere flowing for 1 hour. Multiplying the current in amperes by the time of flow in hours gives the total number of ampere-hours. Used mostly to indicate the amount of energy a storage battery can deliver before it needs recharging, or the energy a primary battery can deliver before it needs replacing. One ampere-hour equals 3600 coulombs.

ampere-hour capacity — The amount of current a battery can deliver in a specified length of time under specified conditions. For example, a 100-ampere-hour battery can supply 20 amperes for 5 hours.

ampere-hour efficiency — The number of ampere-hours obtained from a storage battery divided by the number of ampere-hours required to recharge the storage battery to its original condition.

ampere-hour meter — An electrical meter that measures and registers the amount or the integral, with respect to time, of the current that passes through it and is consumed in the circuit.

Ampère's rule — Current in a certain direction is equivalent to the motion of positive charges in that direction. The magnetic flux generated by a current in a wire encircles the current in the counterclockwise direction when the current is approaching the observer.

ampere-turn — A measure of magnetomotive force, especially as developed by an electric current, defined as the magnetomotive force developed by a coil of one turn through which a current of 1 ampere flows; that is, 1.26 gilberts.

amp-hr — Abbreviation for ampere-hour or ampere-hours.

amplidyne — A special direct-current generator used extensively in servo systems as a power amplifier. The response of its output voltage to changes in field excitation is very rapid, and its amplification factor is high.

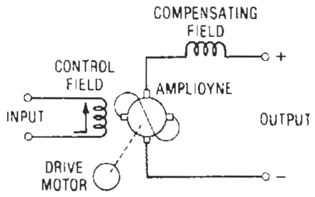

Amplidyne.

amplification — 1. Increase in size of a medium in its transmission from one point to another. May be expressed as a ratio or, by extension of the term, in decibels. 2. An increase in the magnitude of a signal brought about by passing through an amplifier.

amplification factor (λ) — 1. In a vacuum tube, the ratio of a small change in plate voltage to a small change in grid voltage required to produce the same change in plate current (all other electrode voltages and currents being held constant). 2. In any device, the ratio of output magnitude to input magnitude.

amplified AGC — An automatic gain-control (AGC) circuit in which the control voltage is amplified before being applied to the tube or transistor, the gain of which is to be controlled in accordance with the strength of the incoming signal.

amplified back bias — Degenerative voltage developed across a fast time-constant circuit within a stage of an amplifier and fed back into a preceding stage.

amplifier — 1. A device that draws power from a source other than the input signal and that produces as an output an enlarged reproduction of the essential features of its input. The amplifying element may be an electron tube, transistor, magnetic circuit, or any of various devices. 2. A device for increasing the magnitude of a signal by means of a varying control voltage, maintaining the signal's characteristic form as closely as possible to the original. 3. An electronic device for magnifying (and usually controlling) electrical signals. High-fidelity amplifiers consist of a preamplifier equalizer section, plus a power or basic amplifier section. In an integrated amplifier, both sections are built on one chassis and made available as a single unit. Alternately, the two sections are available as separate units. 4. Device for increasing power associated with a signal (voltage or current). Basic types include dc, ac, audio, linear, radio, video, differential, pulse, logarithmic.

amplifier noise — All spurious or unwanted signals, random or otherwise, that can be observed in a completely isolated amplifier in the absence of a genuine input signal.

amplifier nonlinearity — 1. The inability of an amplifier to produce an output at all times proportionate to its input. 2. Gain deviation from a straight line on a plot of amplifier output versus input (the transfer curve).

amplify — To increase in magnitude or strength, usually said of a current or voltage.

amplifying delay line — A delay line used in pulse-compression systems to amplify delayed superhigh-frequency signals.

amplistat — A self-saturating type of magnetic amplifier.

Amplitron — (Raytheon) A broadband crossed-field amplifier with a reentrant electron stream. The electron stream interacts with the backward wave of a nonreentrant rf structure.

amplitude — 1. The magnitude of variation in a changing quantity from its zero value. The word must be modified with an adjective such as peak, rms, maximum, etc., which designates the specific amplitude in question. 2. The level of an audio or other signal in voltage or current terms. 3. The extent to which an alternating or pulsating current or voltage swings from zero or from a mean value.

amplitude-controlled rectifier — A rectifier circuit in which a thyratron is the rectifying element.

amplitude density distribution — A function that gives the fraction of time that a voltage is within a narrow range.

amplitude distortion — Distortion that is present in an amplifier when the amplitude of the output signal fails to follow exactly any increase or decrease in the amplitude of the input signal. It results from nonlinearity of the transfer function and gives rise to harmonic and intermodulation distortion. No amplifier is completely free from the effect because its transfer function is slightly curved. The nature of the curvature determines the order of the distortion produced, but negative feedback and other circuit configurations help minimize the curvature within the dynamic range and hence keep the distortion at a very low level.

amplitude distribution function — A function that gives the fraction of time that a time-varying voltage is below a given level.

amplitude fading — Fading in which the amplitudes of all frequency components of a modulated carrier wave are uniformly attenuated.

amplitude-frequency distortion — The distortion that occurs when the various frequency components of a complex wave are not amplified, attenuated, or transmitted equally well.

amplitude-frequency response — The variation of gain, loss, amplification, or attenuation of a device or system as a function of frequency. Usually measured in the region where the transfer characteristic is essentially linear.

amplitude gate — *See* slicer.

amplitude-level selection — The choice of the voltage level at which an oscilloscope sweep is triggered.

amplitude limiter — A circuit or stage that automatically reduces the amplification to prevent signal peaks from exceeding a predetermined level.

amplitude-modulated transmitter — A transmitter in which the amplitude of its radio-frequency wave is varied at a low frequency rate — usually in the audio or video range. This low frequency is the intelligence (information) to be conveyed.

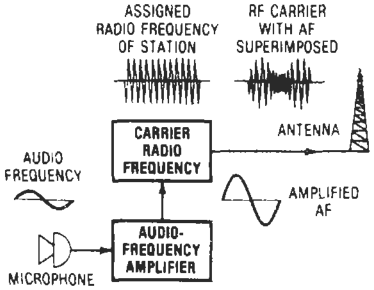
ASSIGNED RADIO FREQUENCY OF STATION
RF CARRIER WITH AF SUPERIMPOSED
ANTENNA
CARRIER RADIO FREQUENCY
AUDIO FREQUENCY
AMPLIFIED AF
AUDIO-FREQUENCY AMPLIFIER
MICROPHONE

Amplitude-modulated transmitter.

amplitude-modulated wave — A constant-frequency waveform in which the amplitude varies in step with the frequency of an impressed signal.

amplitude modulation — Abbreviated AM. 1. Modulation in which the amplitude of a wave is the characteristic subject to variation. Those systems of modulation in which each component frequency (f) of the transmitted intelligence produces a pair of sideband frequencies at carrier frequency plus f and carrier frequency minus f. In special cases, the carrier may be suppressed; either the lower or upper sets of sideband frequencies may be suppressed; the lower set of sideband frequencies may be produced by one or more channels of information. The carrier may be transmitted without intelligence-carrying sideband frequencies. The resulting emission bandwidth is proportional to the highest frequency component of the intelligence transmitted. 2. A process in which the program information is imposed on a carrier signal of constant frequency by varying its amplitude in proportion to program level. Used on the standard broadcast band (530 to 1710 kHz) and on long-wave and shortwave bands.

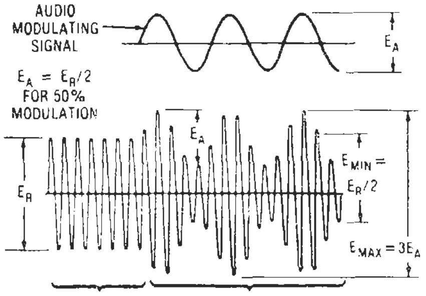

Amplitude modulation.

amplitude-modulation noise level — Undesired amplitude variations of a constant radio-frequency signal, especially in the absence of any intended modulation.

amplitude noise — The effect on radar accuracy of the fluctuations in amplitude of the signal returned by the target. These fluctuations are caused by any change in aspect if the target is not a point source.

amplitude of noise — When impulse-type noise is of random occurrence and so closely spaced that the individual waveshapes are not separated by the receiving equipment, then the noise has the waveshape and characteristics of random noise. Random-noise amplitude is proportional to the square root of the bandwidth. If the impulses are separated, the noise no longer has the waveshape of random noise and its amplitude is directly proportional to the bandwidth of the transmission system.

amplitude permeability—The relative permeability at a stated value of field strength and understated conditions, the field strength varying periodically with time and no direct magnetic-field component being present.

amplitude range—The ratio, usually expressed in decibels, between the upper and lower limits of program amplitudes that contain all significant energy contributions.

amplitude resonance—The condition that exists when any change in the period or frequency of the periodic agency (but not its amplitude) decreases the amplitude of the oscillation or vibration of the system.

amplitude response—The maximum output amplitude that can be obtained at various points over the frequency range of an instrument operated under rated conditions.

amplitude selection—The process of selecting that portion of a waveform which lies above or below a given value or between two given values.

amplitude separator—A television-receiver circuit that separates the control impulses from the video signal.

amplitude-shift keying—Abbreviated ask. The modulation of digital information on a carrier by changing the amplitude of the carrier.

amplitude-suppression ratio—In frequency modulation, the ratio of the magnitude of the undesired output to the magnitude of the desired output of an FM receiver when the applied signal is simultaneously amplitude and frequency modulated. Generally measured with an applied signal that is amplitude modulated 30 percent at a 400-hertz rate and is frequency modulated 30 percent of the maximum system deviation at a 1000-hertz rate.

amplitude versus frequency distortion—Distortion caused by the nonuniform attenuation or gain of the system, with respect to frequency under specified terminal conditions.

AM rejection ratio—The ratio of the recovered audio output produced by a desired FM signal with specified modulation, amplitude, and frequency to that produced by an AM signal, on the same carrier, with specified modulation index.

AM suppression—The ability of an FM tuner to reject AM signals. Expressed in decibels, it is the ratio between the tuner output with a 100-percent modulation FM signal to its output with a 30-percent modulated AM signal.

AM tuner—A device capable of converting amplitude-modulated signals into low-level audio frequencies.

amu—Abbreviation for atomic mass unit.

analog—1. In electronic computers, a physical system in which the performance of measurements yields information concerning a class of mathematical problems. 2. Of or pertaining to the general class of devices or circuits in which the output varies as a continuous function of the input. 3. The representation of numerical quantities by means of physical variables, e.g., translation, rotation, voltage, resistance; contrasted with *digital*. 4. A continuous representation of phenomena in terms of points along a scale, each point merging imperceptibly into the next. An analog voltage, for example, may take any value. Real-world phenomena, such as heat and pressure, are analog (compare with *digital*).

analog adder—An analog circuit or device that receives two or more inputs and delivers an output that is equal to their sum.

analog amplifier—A device whose output is continuously proportional to the input stimulus.

analog channel—A computer channel in which the transmitted information can have any value between the defined limits of the channel.

analog circuit—A circuit in which the output varies as a continuous function of the input, as contrasted with digital circuits.

analog communications—A system of telecommunications employing a nominally continuous electrical signal that varies in frequency, amplitude, etc., in some direct correlation to nonelectrical information (sound, light, etc.) impressed on a transducer.

analog computer—1. A computer operating on the principle of creating a physical (often electrical) analogy of the mathematical problem to be solved. Variables such as temperature, light, pressure, distance, angle, shaft speed, or flow are represented by the magnitude of a physical phenomenon such as voltage or current. The computer manipulates these variables in accordance with the mathematical formulas "analogued" on it. 2. A computer system in which both the input and output are continuously varying signals. 3. A computing machine that works on the principle of measuring, as distinguished from counting. 4. A computer that solves problems by setting up equivalent electric circuits and making measurements as the variables are changed in accordance with the corresponding physical phenomena. An analog computer gives approximate solutions, whereas a digital computer gives exact solutions. 5. A nondigital computer that manipulates linear (continuous) data to measure the effect of a change in one variable on all other variables in a particular problem. (Compare: *digital computer.*)

analog computing—Computing system in which continuous signals represent mechanical (or other) parameters.

analog data—1. A physical representation of information such that the representation bears an exact relationship to the original information. The electrical signals on a telephone channel are an analog data representation of the original voice. 2. Data represented in a continuous form, as contrasted with digital data represented in a discrete (discontinuous) form. Analog data is usually represented by physical variables, such as voltage, resistance, rotation, etc.

analog input module—An I/O rack module that converts an analog signal from a user device to a digital signal that may be processed by the processor.

analog meter—An indicating instrument that employs a movable coil and pointer arrangement (or equivalent) to display values along a graduated scale.

analog multiplexer—1. Circuit used for time-sharing of analog-to-digital converters between a number of different analog information channels. Consists of a group of analog switches arranged with inputs connected to the individual analog channels and outputs connected in common. 2. Two or more analog switches with separate inputs and a common output, with each gate separately controllable. Multiplexing is performed by sequentially turning on each switch one at a time, switching each individual input to a common output. 3. A device that selects one of several analog signals according to a digital code. Analog multiplexers (amux) are available in many forms; their chief application is as a front end in data-acquisition systems, enabling a single analog-to-digital converter to monitor more than one information channel.

analog network—A circuit or circuits that represent physical variables in such a manner as to permit the expression and solution of mathematical relationships between the variables, or to permit the solution directly by electric or electronic means.

analog output — 1. A signal (voltage) whose amplitude is continuously proportionate to the stimulus, the proportionality being limited by the resolution of the device. 2. An output quantity that varies smoothly over a continuous range of values rather than in discrete steps.

analog panel meter — *See* APM.

analog recording — A method of recording in which some characteristic of the record current, such as amplitude or frequency, is continuously varied in a manner analogous to the time variations of the original signal.

analog representation — A representation that does not have discrete values, but is continuously variable.

analog signal — 1. An electrical signal that varies continuously in both time and amplitude, as obtained from temperature or pressure, or speed transducers. A voltage level that changes in proportion to the change in a physical variable. 2. A signal representing a variable that may be continuously observed and continuously represented.

analog switch — 1. A device that either transmits an analog signal without distortion or completely blocks it. 2. Any solid-state device, with or without a driver, capable of bilaterally switching voltages or current. It has an input terminal, output terminal, and, ideally, no offset voltage, low *on* resistance, and extreme isolation between the signal being gated and control signals. 3. A means to interconnect two or more circuits whose information is represented in analog form using a network that may or may not be time divided and may or may not consist of linear elements.

analog-to-digital conversion — 1. The process of converting a continuously variable (analog) signal to a digital signal (binary code) that is a close approximation of the original signal. 2. The process of quantizing a continuous function.

analog-to-digital converter — Abbreviated a-d converter, adc, or ADC. 1. A circuit that changes a continuously varying voltage or current (analog) into a digital output. The input may be ac or dc, and the output may be serial or parallel, binary or decimal. 2. Device that translates analog signals (voltages, pressures, etc.) from sensors into numerical digital form (binary, decimal, etc.).

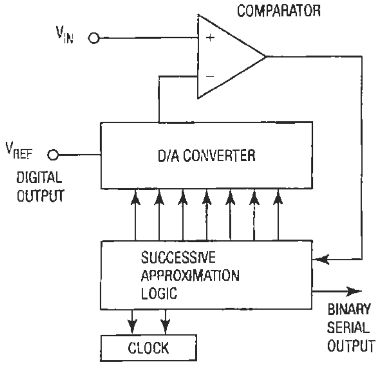

Analog-to-digital converter.

analog transmission — Transmission of a continuously variable signal as opposed to a discretely variable one.

analog value — A continuously variable value, such as a current or voltage.

Analytical Engine — An early form of general-purpose digital computer invented in 1833 by Charles Babbage.

analyzer — 1. An instrument or other device designed to examine the functions of components, circuits, or systems and their relations to each other, as contrasted with an instrument designed to measure some specific parameter of such a system or circuit. 2. Of computers, a routine the purpose of which is to analyze a program written for the same or a different computer. This analysis may consist of summarizing instruction references to storage and tracing sequences of jumps. 3. An instrument that evaluates and/or measures one or more specific parameters (e.g., voltage, current, frequency, logic level, bit time, distortion). 4. A test assembly that checks the performance of, or locates trouble in, electronic equipment. Also called test set and tester.

anastigmat — A lens system designed so as to be free from the aberration called astigmatism.

anchor — An object, such as a metal rod, set into the ground to hold the end of a guy wire.

ancillary equipment — Equipment not directly employed in the operation of a system but necessary for logistic support, preparation for flight, or assessment of target damage; e.g., test equipment, vehicle transport.

AND circuit — Synonym for AND gate.

AND device — A device that has its output in the logic 1 state if and only if all the control signals are in the logic 1 state.

Anderson bridge — A bridge normally used for the comparison of self-inductance with capacitance. It is a six-branch network in which an outer loop of four arms is formed by four nonreactive resistors and the unknown inductor. An inner loop of three arms is formed by a capacitor and a fifth resistor in series with each other and in parallel with the arm opposite the unknown inductor. The detector is connected between the junction of the capacitor and the fifth resistor and at that end of the unknown inductor separated from a terminal of the capacitor by only one resistor. The source is connected to the other end of the unknown inductor and to the junction of the capacitor with two resistors of the outer loop. The balance is independent of frequency.

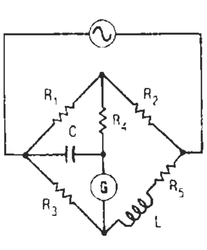

Anderson bridge.

AND gate — 1. In an electronic computer, a gate circuit with more than one control (input) terminal. No output signal will be produced unless a pulse is applied to all inputs simultaneously. 2. A binary circuit, with two or more inputs and a single output, in which the output is logic 1 only when all inputs are logic 1, and the output is logic 0 if any one of the inputs is logic 0.

AND/NOR gate — A single logic element that performs the operation of two AND gates with outputs feeding a NOR gate. No access to the internal logic elements is provided (i.e., no connection is available at the outputs of the AND gates).

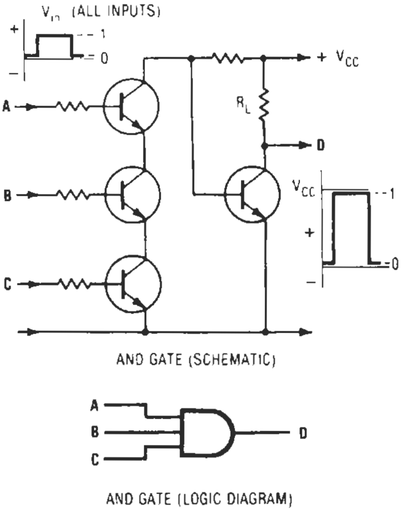

AND gate with three inputs.

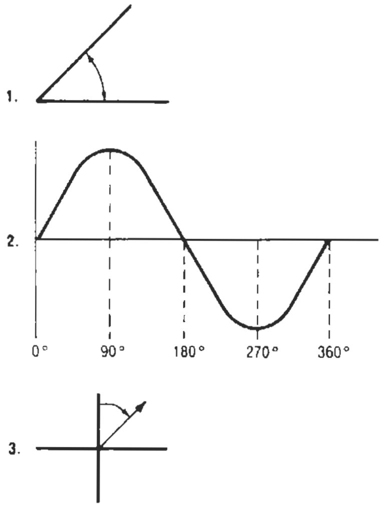

Angle.

AND/OR circuit—A gating circuit that produces a prescribed output condition when several possible combinations of input signals are applied. It exhibits the characteristics of the AND gate and the OR gate.

android—1. A mobile mechanism possessing the ability to manipulate objects external to itself under the constant control of its own resident intelligence, operating within guidelines initially established and occasionally updated by a human being, a computer, or some other external intelligence. 2. Automaton of manlike form.

anechoic—Nonreflective, producing no echoes.

anechoic chamber—1. A room or chamber specially designed to absorb all sound within, thus preventing sound reflections or reverberation. Such rooms are used for evaluation of microphones and speakers. 2. A room lined with material that traps sound waves so the sound is perfectly absorbed and the room is acoustically dead. Such a chamber is used for testing microphones and speakers. 3. A derived term for a room or enclosure that is designed to be echo free over a specified frequency range. Any sound reflections within this frequency range must be less than 10 percent of the source sound pressure.

anechoic enclosure—A special echo-free enclosure used for testing audio transducers, in which all wall surfaces have been covered with acoustically absorbent materials so that reflections of the sound waves are eliminated. Also known as a dead room or an anechoic room.

anechoic room—A room whose walls have been treated so as to make them absorb a particular kind of radiation almost completely; used for testing components of sound systems, radar systems, etc., in an environment free of reflections.

anelectronic—*See* anelectrotonus.

anelectrotonus—The reduced sensitivity produced in a nerve or muscle in the region of contact with the anode when an electric current is passed through it.

anemometer—An instrument used for measuring the force or speed of wind.

angels—Short-duration radar reflections in the lower atmosphere. Most often caused by birds, insects, organic particles, tropospheric layers, or water vapor.

angle—1. A fundamental mathematical concept formed when two straight lines meet at a point. The lines are the sides of the angle, and the point of intersection is the vertex. 2. A measure of the distance along a wave or part of a cycle, measured in degrees. 3. The distance through which a rotating vector has progressed.

angle jamming—An electronic countermeasures technique in which azimuth and elevation information present in the modulation components of the returning echo pulse of a scanning fire-control radar is jammed by transmitting a pulse similar to the radar pulse but with angle information of erroneous phase.

angle modulation—Modulation in which the angle of a sine-wave carrier is the characteristic varied from its normal value by modulation. Phase and frequency modulation are particular forms of angle modulation.

angle noise—Tracking error introduced into radar by variations in the apparent angle of arrival of the echo from a target due to finite target size. (This effect is caused by variations in the phase front of the radiation from a multiple-point target as the target changes its aspect with respect to the observer.)

angle of arrival—Angle made between the line of propagation of a radio wave and the earth's surface at the receiving antenna.

angle of azimuth—The angle measured clockwise in a horizontal plane, usually from the north. The north used may be true north, Y-north, or magnetic north.

angle of beam—The angle that encloses most of the transmitted energy from a directional-antenna system.

angle of convergence—Angle formed by the lines of sight of both eyes when focusing on an object.

angle of deflection—The angle formed between the new position of the electron beam in a cathode-ray tube and the normal position before deflection.

angle of departure—The angle of the line of propagation of a radio wave with respect to a horizontal plane at the transmitting antenna.

angle of divergence—In cathode-ray tubes, a measure of its spread as the electron beam travels from the cathode to the screen. The angle formed by an imaginary center line and the border line of the electron beam. In good tubes, this angle is less than 2°.

angle of elevation—The angle between the horizontal plane and the line ascending to the object.

angle of incidence—The angle between a wave or beam striking a surface and a line perpendicular to that surface.

angle of lag—The angular phase difference between one sinusoidal function and a second having the same frequency. Expressed in degrees, the amount the second function must be retarded to coincide with the first.

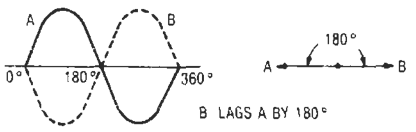

Angle of lag.

angle of lead—1. The time or angle by which one alternating electrical quantity leads another of the same cyclic period. 2. The angle through which the commutator brushes of a generator or motor must be moved from the normal position to prevent sparking.

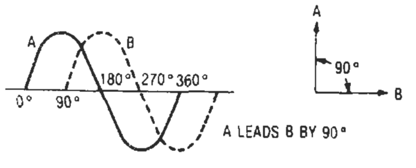

Angle of lead, 1.

angle of radiation—The angle between the surface of the earth and the center of the beam of energy radiated upward into the sky from a transmitting antenna.

angle of reflection—The angle between a wave or beam reflected from a surface and a line perpendicular to that surface. This angle lies in a common plane with the angle of incidence and is equal to it.

angle of refraction—The angle between a wave or beam as it passes through a medium and a line perpendicular to the surface of that medium. This angle lies in a common plane with the angle of incidence.

angle tracking noise—Any deviation of the tracking axis from the center of reflectivity of a target. The resultant of servo noise, receiver noise, angle noise, and amplitude noise.

angstrom unit—A unit of measurement of wavelength of light and other radiation. Equal to one ten-thousandth (10^{-4}) of a micrometer or one hundred-millionth of a centimeter (10^{-8} cm). The visible spectrum extends from about 4000 to 8000 angstrom units. Blue light has a wavelength in the region of 4700 angstroms; yellow, 5800; and red, 6500. A measure of wavelength equal to 10^{-10} meter, or 0.1 nanometer, the preferred term.

angular acceleration—The rate at which angular velocity changes with respect to time, generally expressed in radians per second.

angular accelerometer—A device capable of measuring the magnitude of, and/or variations in, angular acceleration.

angular aperture—The largest angular extent of wave surface that an objective can transmit.

angular deviation loss—The ratio of the response of a microphone or speaker on its principal axis to the response at a specified angle from the principal axis (expressed in decibels).

angular distance—The angle subtended by two bodies at the point of observation. It is equal to the distance of wavelengths multiplied by 2π radians or by 360°.

angular frequency—Frequency expressed in radians per second. It is equal to the number of hertz (cycles per second) multiplied by 2π.

angular length—Length expressed in radians or equivalent angular measure equal to 2π radians, or 360°, multiplied by the length in wavelengths.

angular momentum—The momentum that a body has by virtue of its rotational movement.

angular phase difference—Phase difference between two sinusoidal functions expressed as an angle.

angular rate—The rate of change of bearing.

angular resolution—The ability of a radar to distinguish between two targets solely on the basis of angular separation.

angular velocity—The rate at which an angle changes. Expressed in radians per second, the angular velocity of a periodic quantity is the frequency multiplied by 2π. If the periodic quantity results from uniform rotation of a vector, the angular velocity is the number of radians per second passed over by the rotating vector. Generally designated by the Greek letter omega (ω).

anharmonic oscillator—An oscillating system in which the restoring force is a nonlinear function of the displacement from equilibrium.

anhysteresis—The process whereby a material is magnetized by applying a unidirectional field upon which is superimposed an alternating field of gradually decreasing amplitude.

ANIK—The Canadian domestic satellite system used to transmit the network television feeds of the Canadian Broadcasting Corporation. All ANIK satellites are operated by TeleSat Canada of Ottawa. ANIK satellites have both 4-GHz C-band and 12-GHz Ku-band transponders. ANIK means *brother* in Inuit (Eskimo).

animation—A moving on-screen representation of the activities taking place in a simulation.

anion—1. A negatively charged ion which, during electrolysis, is attracted toward the anode. A corresponding positive ion is called a cation. 2. A negative ion that moves toward the anode in a discharge tube, electrolytic cell, or similar device.

anisotropic—1. Describing a substance that exhibits different magnetic, electrical, optical, and other physical properties when measured along axes in different directions. 2. A material that has characteristics such as wave propagation constant, magnetic permeability, conductivity, etc., that vary with direction; that is, not isotropic.

anisotropic body—A body in which the value of any given property depends on the direction of measurement, as opposed to a body that is isotropic.

anisotropic magnet—A magnetic material having a better magnetic characteristic along the preferred axis than along any other.

anisotropic material—A material having preferred orientation so that the magnetic characteristics are superior along a particular axis. This may be as a result of rolling, heat treatment in a magnetic field, or, in the case of some of the sintered magnets, the direction of press.

anisotropy—Directional dependence of magnetic properties, leading to the existence of easy or preferred directions of magnetization. Anisotropy of a particle may be related to its shape, to its crystalline structure, or to the existence of strains within it.

anneal—1. To heat a metal to a predetermined temperature and then let it cool slowly. This process prevents brittleness and often stabilizes electrical characteristics.

2. To heat and then gradually cool in order to relieve mechanical stresses. Annealing copper makes it softer and less brittle.

annealed laminations — Laminations that have been annealed for transformers or choke coils.

annealed wire — Wire that has been softened by heating and gradual cooling to remove mechanical stresses.

annotation — An added descriptive comment or explanatory note.

annular — Ringed; ring-shaped.

annular conductor — A conductor consisting of a number of wires stranded in three reversed concentric layers surrounding a saturated hemp core. The core is usually made wholly or mostly of nonconducting material. This construction has the advantage of lower total ac resistance for a given cross-sectional area of conducting material by eliminating the greater skin effect at the center.

annular transistor — A mesa transistor in which the semiconductor regions are arranged in concentric circles about the emitter.

annulling network — An arrangement of impedance elements connected in parallel with filters to annul or cancel capacitive or inductive impedance at the extremes of the passband of a filter.

annunciation relay — 1. An electromagnetically operated signaling apparatus that indicates whether a current is flowing or has flowed in one or more circuits. 2. A nonautomatic reset device that gives a number of separate visual indications upon the functioning of protective devices, and which may also be arranged to perform a lockout function.

annunciator — 1. A visual device consisting of a number of pilot lights or drops. Each light or drop indicates the condition that exists or has existed in an associated circuit and is labeled accordingly. 2. A device for sounding an alarm or attracting attention. The indication is usually aural, but occasionally may be visual or both aural and visual. 3. An alarm-monitoring device that consists of a number of visible signals, such as flags or lamps indicating the status of the detectors in an alarm system or systems. Each circuit in the device is usually labeled to identify the location and condition being monitored. In addition to the visible signal, an audible signal is usually associated with the device. When an alarm condition is reported, a signal is indicated — visible, audible, or both. The visible signal is generally maintained until reset either manually or automatically.

anode — 1. The positive electrode, such as the plate of a vacuum tube; the element to which the principal stream of electrons flows. 2. In a cathode-ray tube, the electrodes connected to a source of positive potential. These anodes are used to concentrate and accelerate the electron beam for focusing. 3. The less noble and/or higher-potential electrode of an electrolytic cell, at which corrosion occurs. This may be an area on the surface of a metal or alloy, the more active metal in a cell composed of two dissimilar metals, or the positive electrode of an impressed-current system.

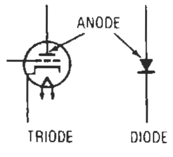

Anode.

anode-balancing coil — A set of mutually coupled windings used to maintain approximately equal currents in anodes operating in parallel from the same transformer terminal.

anode breakdown voltage — The potential required to cause conduction across the main gap of a gas tube when the starter gap is not conducting and all other tube elements are held at cathode potential.

anode-bypass capacitor — Also called plate-bypass capacitor. A capacitor connected between the anode and ground in an electron-tube circuit. Its purpose is to bypass high-frequency currents and keep them out of the load.

anode characteristic curve — A graph that shows how the anode current of an electron tube is affected by changes in the anode voltage.

anode circuit breaker — A device used in the anode circuits of a power rectifier for the primary purpose of interrupting the rectifier circuit if an arcback should occur.

anode current — The electron flow in the element designated as the anode. Usually signifies plate current.

anode dark space — In a gas tube, a narrow, dark zone next to the surface of the anode.

anode dissipation — The power dissipated as heat in the anode of an electron tube because of the bombardment by electrons and ions.

anode efficiency — *See* plate efficiency.

anode-load impedance — *See* plate-load impedance.

anode modulation — *See* plate modulation.

anode neutralization — Also called plate neutralization. A method of neutralization in which a portion of the anode-cathode ac voltage is shifted 180° and applied to the grid-cathode circuit through a neutralizing capacitor.

anode power input — *See* plate power input.

anode power supply — The means for supplying power to the plate of an electron tube at a more positive voltage than that of the cathode. Also called plate power supply.

anode pulse modulation — *See* plate pulse modulation.

anode rays — Positive ions coming from the anode of an electron tube; these ions are generally due to impurities in the metal of the anode.

anode saturation — *See* plate saturation.

anode sheath — A layer of electrons surrounding the anode in mercury-pool arc tubes.

anode strap — A metallic connector between selected anode segments of a multicavity magnetron, used principally for mode separation.

anode supply — Also called plate supply. The dc voltage source used in an electron-tube circuit to place the anode at a high positive potential with respect to the cathode.

anode terminal — 1. In a diode (semiconductor or tube), that terminal to which a positive dc voltage must be applied to forward-bias the diode. Compare with *cathode terminal.* 2. In a diode (semiconductor or tube), that terminal at which a negative dc voltage appears when the diode is employed as an ac rectifier (blocking). 3. That terminal which is internally connected to the anodic element of any device.

anode voltage — The potential difference existing between the anode and cathode.

anode voltage drop (of a glow-discharge, cold cathode tube) — Difference in potential between cathode and anode during conduction, caused by the electron flow through the tube resistance (*IR* drop)

anodic protection — Corrosion inhibition based on the electrolytic formation of a protective passive film on metals by applying to them a positive (anodic) potential; e.g., aluminum is anodized (oxidized) by a positive charge in a sulfuric acid solution.

anodic silver — A precious metal used in plating; fine silver in different configurations, such as shot, cones, bars, etc., is sacrificed during the silver-plating process.

anodization — The formation of an insulating oxide over certain elements, usually metals, by electrolytic action. The most commonly anodized materials are tantalum, aluminum, titanium, and niobium. Anodization is particularly useful where protection of a conductor is required. The base metal can form the conductor and the anodized surface layer can form the insulator.

anodize — To deposit a protective coating of oxide on a metal by means of an electrolytic process in which it is used as the anode.

anodizing — An electrochemical oxidation process used to improve the corrosion resistance or to enhance the appearance of a metal surface. Aluminum and magnesium parts are frequently anodized.

anomalous displacement current — Also called dielectric absorption. The current in addition to the normal leakage current in a circuit containing a capacitor with an imperfect dielectric after the normal charging or discharging current has become negligibly small.

anomalous photoconductivity — A spectral phenomenon in which the degree of the photoresponse of an illuminated semiconductor is determined by the wavelength composition of the incident light.

anomalous propagation — 1. Propagation that is unusual or abnormal. 2. The conduction of UHF signals through atmospheric ducts or layers in a manner similar to that of a waveguide. These atmospheric ducts carry the signals with less than normal attenuation over distances far beyond the optical path taken by UHF signals. Also called superrefraction. 3. In sonar, pronounced and rapid variations in the strength of the echo due to large, rapid focal fluctuations in propagation conditions.

anonymous FTP (File Transfer Protocol) — The procedure of connecting to a remote computer as an anonymous or guest user in order to transfer public files back to a local computer. *See also* FTP; protocol.

A-N radio range — A navigational aid that provides four equisignal zones for aircraft guidance. Deviation from the assigned course is indicated aurally by the Morse code letters A (·—) or N (—·). On-course position is indicated by an audible merging of the A and N code signals into a continuous tone.

ANSI — American National Standards Institute. The U.S. government organization with responsibility for the development and promulgation of (among others) data processing standards.

A-N signal — A radio-range, quadrant-designation signal that indicates to the pilot whether he or she is on course or to the right or left.

ANSI keyboard — Abbreviation for American National Standards Institute keyboard. A typewriter standard unit that offers a choice of uppercase characters only or uppercase and lowercase combined.

ANSI standards — A series of standards recommended by the American National Standards Institute.

answerback — 1. The response of a terminal to remote-control signals. 2. A signal sent by a data receiver to a data transmitter indicating that it is ready to receive data or is acknowledging the receipt of data. *See also* handshaking. 3. A reply message from a terminal, manually or automatically initiated, to verify that the right terminal has been accessed and is in operation.

answerback (W-R-U) system — A system capable of being remotely controlled by another station. When tripped by a unique access code, a short predetermined message is broadcast.

answerback unit — An electromechanical device used with a teletypewriter set to transmit a predetermined message of not more than 21 characters in response to a request signal. It can transmit either a five-level 7.42 unit code or an eight-level 11.0 unit code at speeds of up to 100 words per minute.

answering cord — The cord nearest the face of a telephone switchboard. It is used for answering subscriber's calls and calls on incoming trunks.

answering service — A business that contracts with subscribers to answer incoming telephone calls after a specified delay or when scheduled to do so. It may also provide other services, such as relaying fire or intrusion alarm signals to proper authorities.

answer lamp — In a telephone switchboard, a lamp that lights when an answer cord is plugged into a line jack; it extinguishes when the telephone answers and lights when the call is complete.

answer tone — Tone signal, with a frequency between 2025 and 2225 Hz and a duration of at least 1.5 s, used by an answering modem to indicate its ready condition to an originating modem.

antenna — Also called aerial. 1. That portion, usually wires or rods, of a radio transmitter or receiver station used for radiating waves into or receiving them from space. It changes electrical currents into electromagnetic radio waves, and vice versa. 2. A section of wire or a metallic device designed to intercept radio waves in the air and convert them to an electrical signal for feeding to a receiver. Under relatively difficult reception conditions, such as created by location, terrain, obstructions, etc., an antenna becomes fairly critical and should be one especially designed for its intended purposes. 3. A device for transmitting and receiving radio waves. Depending on their use and operating frequency, antennas can take the form of a single piece of wire, a dipole, a grid such as a yagi array, a horn, a helix, a sophisticated parabolic-shaped dish, or a phase array of active electronic elements of virtually any flat or convoluted surface. 4. A device that collects and focuses electromagnetic energy, i.e., contributes an energy gain. Gain is proportional to surface area for a microwave dish.

antenna array — 1. A combination of antennas assembled to obtain a desired pickup or rejection pattern. 2. An arrangement of two or more directional antennas, spaced and connected so that they are in phase and their effects are electrically additive.

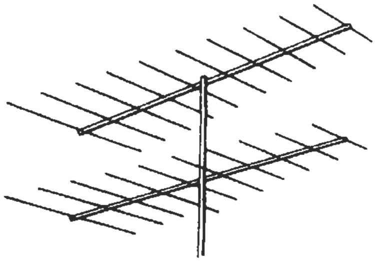

Antenna array.

antenna bandwidth — 1. The range of frequencies over which the impedance characteristics of the antenna are sufficiently uniform that the quality of the radiated signal is not significantly impaired. 2. The frequency range over which a certain antenna characteristic falls within acceptable limits. For instance, an antenna may have a bandwidth of 1 MHz over which the standing-wave ratio is 2 : 1 or less. 3. The frequency range throughout which an antenna will operate at a specified efficiency without the need for alteration or adjustment.

antenna beam width—The angle, in degrees, between two opposite half-power points of an antenna beam.

antenna coil—In a radio receiver or transmitter, the inductance through which antenna current flows.

antenna coincidence—That instance when two rotating, highly directional antennas are pointed toward each other.

antenna-conducted interference—Any signal that is generated within a transmitter or receiver and appears as an undesired signal at the antenna terminals of the device, e.g., harmonics of a transmitter signal, or the local-oscillator signal of the receiver.

antenna cores—Ferrite cores of various cross sections for use in radio antennas.

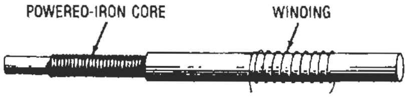

Antenna core.

antenna coupler—1. A radio-frequency transformer used to connect an antenna to a transmission line or to connect a transmission line to a radio receiver. 2. A radio-frequency transformer, link circuit, or tuned line used to transfer radio-frequency energy from the final plate-tank circuit of a transmitter to the transmission line feeding the antenna.

antenna crosstalk—A measure of undesired power transfer through space from one antenna to another. Usually expressed in decibels, the ratio of power received by one antenna to the power transmitted by the other.

antenna current—The radio-frequency current that flows in an antenna.

antenna detector—A device consisting of an antenna and electronic equipment to warn aircraft crew members of their being observed by radar sets. The device is usually located in the nose or tail of the aircraft and illuminates a light on one or more panels when radar signals are detected.

antenna diplexer—A coupling device that permits several transmitters to share one antenna without troublesome interaction.

antenna-directivity diagram—A curve representing, in polar or Cartesian coordinates, a quantity proportional to the gain of an antenna in the various directions in a particular plane or cone.

antenna disconnect switch—A safety switch or interlock plug used to remove driving power from the antenna to prevent rotation while work is being performed.

antenna duplexer—A circuit or device that permits one antenna to be shared by two transmitters without undesirable interaction.

antenna effect—1. Cause of error in a loop antenna due to the capacitance to ground. 2. In a navigational system, any undesirable output signal that results when a directional antenna acts as a nondirectional antenna. 3. The tendency of wires or metallic bodies to act as antennas, i.e., to radiate or pick up radio signals.

antenna effective area—In any specified direction, the square of the wavelength multiplied by the power gain (or directional gain) in that direction, and divided by 4π. (When power gain is used, the effective area is that for power reception; when directive gain is used, the effective area is that for directivity.)

antenna efficiency—The relative ability of an antenna to convert rf energy from a transmitter into electromagnetic waves. If the gain rating of a directional antenna is 10 dB, for example, it is often assumed that the effective radiated power will be 10 times greater than the rf power fed to it. However, if the antenna efficiency is, say, 50 percent, a loss of 3 dB, the true gain will be only 7 dB ($10 - 3 = 7$ dB).

antenna elevation—The physical height of an antenna above the earth.

antenna factor—The value of decibel that must be added to a two-terminal voltmeter reading to obtain the actual induced antenna open-circuit voltage or the electric-field strength.

antenna farm—A large plot of ground (5 to 2000 acres) surrounding a radio transmitting or receiving station that provides space and adequate clearance for the installation of several large antennas, such as rhombic antennas.

antenna field—1. The region defined by a group of antennas. 2. A group of antennas placed in a geometric configuration that is specific for a particular trajectory measuring system. 3. The effective free-space energy distribution produced by an antenna or group of antennas.

antennafier—An integrated low-profile antenna and amplifier for use with compact, portable communications systems.

antenna front-to-back ratio—The ratio of field strength in front of a directional antenna (i.e., directly forward in the line of maximum directivity) to the field strength in back of the antenna (i.e., 180° from the front). Measured at a fixed distance from the radiator.

antenna gain—1. The effectiveness of a directional antenna in a particular direction, compared against a standard (usually an isotropic antenna). The ratio of standard antenna power to the directional antenna power that will produce the same field strength in the desired direction. 2. The increase in signal level at the antenna terminals with reference to the level at the terminals of a half-wave dipole antenna, expressed in decibels. 3. For a given antenna, the ratio of signal strength (received or transmitted) to that obtained with a simple dipole antenna.

antenna ground system—That portion of an antenna closely associated with the earth and including an extensive conducting surface, which may be the earth itself.

antenna height—The average height above the terrain from 2 to 10 miles (3.2 to 16 km) from the antenna. In general, the antenna height will be different in each direction from the antenna. The average of these various heights is considered the antenna height above average terrain.

antenna height above average terrain—The height of the center of radiation of an antenna above an averaged value of the elevation above sea level for the surrounding terrain.

antenna illumination—Describes how a feedhorn "sees" the surface of a dish as well as the surrounding terrain.

antenna impedance—The impedance an antenna presents to a transmitter or receiver at the attachment point of the transmission line or feeder. It varies from about 50 to 600 ohms, depending on antenna type and installation.

antenna induced microvolts—The voltage that exists across the open-circuited antenna terminals, as calculated from a measurement.

antenna lens—An arrangement of metal vanes or dielectric material used to focus a microwave beam in a manner similar to an optical lens.

antenna lobe—*See* lobe.

antenna matching—Selection of components to make the impedance of an antenna equal to the characteristic impedance of its transmission line.

antennamitter—An integrated low-profile antenna and oscillator for use with compact, portable communications systems.

antenna pair—Two antennas located on a base line of accurately surveyed length. The signals received by these antennas are used to determine quantities related to a target position.

antenna pattern—Also called antenna polar diagram. A plot of angle versus free-space field intensity at a fixed distance in the horizontal plane passing through the center of the antenna.

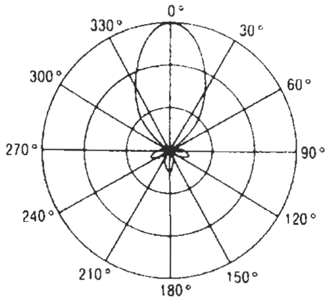

Antenna pattern.

antenna-pattern measuring equipment—Devices used to measure the relative field strength or intensity existing at any point or points in the space immediately surrounding an antenna.

antenna pedestal—A structure that supports an antenna assembly (motors, gears, synchros, rotating joints, etc.).

antenna polar diagram—*See* antenna pattern.

antenna polarization—The position of an antenna, with respect to the surface of the earth, that determines the wave polarization for which the antenna is most efficient. A vertical antenna radiates and receives vertically polarized waves; a horizontal antenna radiates and receives horizontally polarized waves broadside to itself and vertically polarized waves at high angles off its ends.

antenna power—The square of the antenna current of a transmitter, multiplied by the antenna resistance at the point where the current is measured.

antenna power gain—The power gain of an antenna in a given direction is four times the ratio of the radiation intensity in that direction to the total power delivered to the antenna. (The term is also applied to receiving antennas.)

antenna preamplifier—A low-noise rf amplifier, usually mast-mounted near the terminals of the receiving antennas, used to compensate for transmission-line loss and thereby improve the overall noise figure.

antenna reflector—In a directional-antenna array, an element that modifies the field pattern in order to reduce the field intensity behind the array and increase it in front. In a receiving antenna, the reflector reduces interference from stations behind the antenna.

antenna relay—A relay used in radio stations to automatically switch the antenna to the receiver or transmitter and thus protect the receiver circuits from the rf power of the transmitter.

antenna resistance—The total resistance of a transmitting antenna system at the operating frequency. The power supplied to the entire antenna circuit, divided by the square of the effective antenna current referred to the feed point. Antenna resistance is made up of such components as radiation resistance, ground resistance, radio-frequency resistance of conductors in the antenna circuit, and equivalent resistance due to corona, eddy currents, insulator leakage, and dielectric power loss.

antenna resonant frequency—The frequency (or frequencies) at which an antenna appears to be a pure resistance.

antenna stabilization—A system for holding a radar beam steady despite the roll and pitch of a ship or airplane.

antenna structure—A structure that includes the radiating system, its supporting structures, and appurtenances mounted thereon.

antenna switch—Switch used for connecting an antenna to or disconnecting it from a circuit.

antenna system—An assembly consisting of the antenna and the necessary electrical and mechanical devices for insulating, supporting, and/or rotating it.

antenna terminals—On an antenna, the points to which the lead-in (transmission line) is attached.

antenna tilt error—The angular difference between the antenna tilt angle shown on the mechanical indicator and the electrical center of the radar beam.

antennaverter—A receiving antenna and converter combined in a single unit that feeds directly into the receiver IF amplifier.

antenna wire—A wire, usually of high tensile strength, such as copperweld, bronze, etc., with or without insulation, used as an antenna for radio and electronic equipment.

antiaircraft missile—A guided missile launched from the surface against an airborne target.

anti-aliasing—The smoothing or removal of diagonal lines in digitized images at low resolutions that appear as stair-steps in order to recreate smoother diagonal lines.

anti-aliasing filter—A filter (normally low pass) that band-limits the input signal before sampling to less than half the sampling rate to prevent aliasing noise.

anticapacitance switch—A switch with widely separated legs, designed to keep capacitance at a minimum in the circuits being switched.

anticathode—Also called target. The target of an X-ray tube on which the stream of electrons from the cathode is focused and from which the X-rays are radiated.

anticlutter circuit—In a radar receiver, an auxiliary circuit that reduces undesired reflection in order to permit the detection of targets that otherwise would be obscured by such reflections.

anticlutter gain control—A device that automatically and gradually increases the gain of a radar receiver from low to maximum within a specified period after each transmitter pulse. In this way, short-range echoes producing clutter are amplified less than long-range echoes.

anticoincidence—A nonsimultaneous occurrence of two or more events (usually, ionizing events).

anticoincidence circuit—1. A counter circuit that produces an output pulse when either of two input circuits receives a pulse, but not when the two inputs receive pulses simultaneously. 2. A circuit that provides an output only when all inputs are absent; a NAND circuit.

anticollision radar—A radar system used in an aircraft or ship to warn of possible collision.

antiferroelectricity—The property of a class of crystals that also undergo phase transitions from a higher to a lower symmetry. They differ from the ferroelectrics in having no electric dipole moment.

antiferroelectric materials—Those materials in which spontaneous electric polarization occurs in lines of ions; adjacent lines are polarized in an antiparallel arrangement.

antiferromagnetic materials—Those materials in which spontaneous magnetic polarization occurs in equivalent sublattices; the polarization in one sublattice is aligned antiparallel to the other.

antiferromagnetic resonance—The absorption of energy from an oscillating electromagnetic field by a system of processing spins located on two sublattices, with the spins on one sublattice going in one direction and the spins on the other sublattice in the opposite direction.

antiferromagnetism—A phenomenon of magnetism characterized by the elimination of magnetic moments and decrease in magnetic susceptibility with a decrease in temperature due to the equal power of atomic magnets.

antihunt—A stabilizing signal or equalizing circuit used in a closed-loop feedback system of a servomechanism to prevent the system from hunting, or oscillating. Special types of antihunt circuits are the anticipator, derivative, velocity feedback, and damper.

antihunt circuit—A circuit used to prevent excessive correction in a control system.

antihunt device—A device used in positioning systems to prevent hunting, or oscillation, of the load around an ordered position. The device may be mechanical or electrical. It usually involves some from of feedback.

antijamming—1. Minimizing the effect of enemy electronic countermeasures to permit echoes from targets detected by radar to be visible on the indicator. 2. Controls or circuit features incorporated to minimize jamming.

antijamming radar data processing—Use of data from one or more radar sources to determine target range in the presence of jamming.

antilogarithm—The number from which a given logarithm is derived. For example, the logarithm of 4261 is 3.6295. Therefore the antilogarithm of 3.6295 is 4261.

antimagnetic—Made of alloys that will not remain in a magnetized state.

antimicrophonic—Specifically designed to prevent microphonics. Possessing the characteristic of not introducing undesirable noise or howling into a system.

antimissile missile—A missile that is launched to intercept and destroy another missile in flight.

antinode—The point on a transmission line at which the current is maximum and the voltage is minimum.

antinodes—Also called loops. The points of maximum displacement in a series of standing waves. Two similar and equal wave trains traveling at the same velocity in opposite directions along a straight line result in alternate antinodes and nodes along the line. Antinodes are separated from their adjacent nodes by half the wavelength of the wave motion.

antinoise carrier-operated device—A device commonly used to mute the audio output of a receiver during standby or no-carrier periods. Usually the automatic volume control voltage is used to control a squelch tube which, in turn, controls the bias applied to the first audio tube so that it is permitted to operate only when a carrier is present at the receiver input. Thus, the receiver output is heard when a signal is received, and is muted when no signal is present.

antinoise microphone—A microphone that discriminates against acoustic noise. A lip or throat microphone is an example.

antiphase—Two identical signals disposed in 180° phase opposition. When superimposed, they tend to cancel each other because their waveform patterns are of equal magnitude but opposite polarity.

antiproton—An elementary atomic particle that has the same mass as a proton but is negatively charged.

antirad—A material that inhibits damage caused by radiation.

antiresonance—A type of resonance in which a system offers maximum impedance at its resonant frequency.

antiresonant circuit—A parallel resonant circuit offering maximum impedance to the series passage of the resonant frequency.

antiresonant frequency—1. The frequency at which the impedance of a system is very high. 2. Of a crystal unit, the frequency for a particular mode of vibration at which, neglecting dissipation, the effective impedance of the crystal unit is infinite.

antisidetone—1. In a telephone circuit, special circuits and equipment that are so arranged that only a negligible amount of the power generated in the transmitter reaches the associated receiver. 2. Pertaining to the reduction or elimination of interference in telephone circuits between the microphone and earphone of the same telephone.

antisidetone circuit—A telephone circuit that prevents sound, introduced in the local transmitter, from being reproduced in the local receiver. (Reduces sidetones.)

antisidetone induction coil—An induction coil designed for use in an antisidetone telephone set.

antisidetone telephone set—A telephone set with an antisidetone circuit.

antiskating bias—A bias force applied to a pivoted pickup arm to counteract the inward force (toward the center of the record) resulting from the drag of the stylus in the groove and the offset angle of the head.

antiskating device—A mechanism found on modern phonograph pickups that provides a small outward force on a pickup arm. This counteracts the arm's tendency to move toward the turntable center (inward) due to offset geometry, and reduces stylus/groove friction.

antistatic agents—Methods employed to minimize static electricity in plastic materials. Such agents are of two basic types. Metallic devices that come into contact with the plastics and conduct the static to earth give complete neutralization initially, but because it is not modified, the surface of the material can become prone to further static accumulation during subsequent handling. Chemical additives, which are mixed with the compound during processing, give a reasonable degree of protection to the finished products.

antistatic cleaner—Substance used on phonograph records that helps to prevent the buildup of a static charge that attracts dust.

antistatic coating—1. An electrically conductive layer for carrying off static charges that could accumulate on a surface. 2. A conductive coating applied to a TV or monitor screen (or on a glass panel immediately in front of the screen) that conducts away any static charge and prevents dust from adhering to the surface of the television.

antistatic sprays—Chemical agents which, when applied to circuits and plastic surfaces, leave a conductive coating that acts to repel dust and dirt and changes surface characteristics. Good antistatic sprays will leave a resistivity reading of 20 to 100 megohms per square inch ($3-15\ \mathrm{M\Omega/cm^2}$) on plastics and 100 megohms or more per square inch on glass surfaces. *See* static eliminators.

antistickoff voltage—A small voltage, usually applied to the rotor winding of the coarse synchro control transformer in a two-speed system. The antistickoff voltage acts to eliminate the possibility of ambiguous behavior in the system.

antitransmit-receive box—A second transmit-receive switch used in a radar antenna system to minimize absorption of the echo signal in the transmitter circuit during the interval between transmitted pulses.

antitransmit-receive switch—Abbreviated atr switch. An automatic device employed in a radar system to prevent received energy from being absorbed in the transmitter.

antitransmit-receive tube—*See* atr tube.

antivoice-operated transmission—A method of radiocommunication in which a voice-activated circuit prevents the operation of the transmitter during reception of messages on an associated receiver.

aperiodic—1. Having no fixed resonant frequency or repetitive characteristics or no tendency to vibrate. A circuit that will not resonate within its tuning range is often called aperiodic. 2. Not characterized by predictable periods or steps.

aperiodic antenna—An antenna designed to have a constant impedance over a wide frequency range (for example, a terminated rhombic antenna) due to the suppression of reflections within the antenna system.

aperiodic damping—Also called overdamping. The condition of a system when the amount of damping is so large that when the system is subjected to a single disturbance, either constant or instantaneous, the system comes to a position of rest without passing through that position. Although an aperiodically damped system is not strictly an oscillating system, it has such properties that it would become an oscillating system if the damping were sufficiently reduced.

aperiodic function—A function having no repetitive characteristics and not repeatable within a specified period.

aperiodic waveform—A nonrepeating, random, one-shot waveform.

aperture—1. In a unidirectional antenna, that portion of the plane surface which is perpendicular to the direction of maximum radiation and through which the major part of the radiation passes. 2. In an opaque disc, the hole or window placed on either side of a lens to control the amount of light passing through. 3. Also called aperture time. The amount of certainty about the exact time when the encoder input was at the value represented by a given output code. In general, the aperture is equal to the conversion time; it may be reduced by the use of sample-and-hold circuits. 4. In an electron gun, the opening that determines the size of, and has an effect on, the shape of the electron beam. In television optics, it is the effective diameter of the lens that controls the amount of light reaching the photoconductive or photoemitting image pickup tube. 5. An opening that will pass light, electrons, or other forms of radiation.

aperture antenna—A type of antenna whose beam width is determined by the dimensions of a horn, lens, or reflector.

aperture compensation—Reduction of aperture distortion by boosting the high-frequency response of a television-camera video amplifier.

aperture correction—Compensation for the loss in sharpness of detail because of the finite dimensions of a scanning beam in the horizontal dimension.

aperture delay time—The time elapsed from the application of the hold (or encode) command until the sampling switch in a sample-and-hold circuit opens fully and the device actually takes the sample. Aperture delay time is a fixed delay time and is normally not an error source since the hold clock edge can be advanced to compensate for it.

aperture distortion—In a television signal, the distortion due to the finite dimension of the camera-tube scanning beam. The beam covers several mosaic globules simultaneously, resulting in a loss of picture detail.

aperture illumination—The field distribution in amplitude and phase through the aperture.

aperture jitter—Also called aperture uncertainty time. 1. In a sample-and-hold circuit, the time variation or uncertainty with which the switch opens, or the time variation in aperture delay. 2. A source of error in a sampling system, which determines the maximum slew rate limitation of the sampled analog input signal for a given system resolution.

aperture mask—Also called shadow mask. A thin sheet of perforated material placed directly behind the viewing screen in a three-gun color picture tube to prevent the excitation of any one color phosphor by either of the two electron beams not associated with that color.

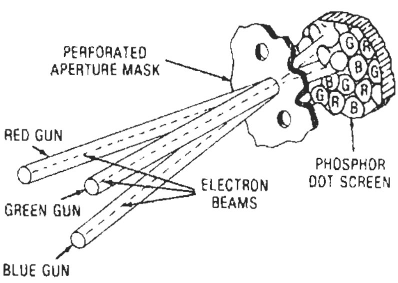

Aperture mask.

aperture plate—A ferrite memory plate containing a large number of uniformly spaced holes arranged in parallel rows and interconnected by plated conductors to provide a magnetic memory plate.

aperture time—1. In a sample-and-hold circuit, the averaging time of a sample-hold during the sample-to-hold transition. 2. The time required by a sample-and-hold device to go from the sample mode into the hold mode, once the hold command has been received. The aperture time is generally a few nanoseconds, measured from the 50-percent point of the mode-control transition to the time when the output stops tracking the input. *See* aperture, 3.

aperture-time uncertainty—The possible variation in aperture time from one sample-to-hold transition to the next.

APL—Abbreviation for average picture level. The average luminance level of the part of a television line between blanking pulses.

APM—Abbreviation for analog panel meter. A scale-and-pointer meter capable of indicating a continuous, rather than incremental, range of values from zero to the rated full-scale value.

apogee—The point in an elliptical satellite orbit that is farthest from the surface of the earth. Geosynchronous satellites that maintain circular orbits around the earth are first launched into highly elliptical orbits with apogees of 22,237 miles. When the communication satellite reaches the appropriate apogee, a rocket motor is fired to place the satellite into its permanent circular orbit of 22,237 miles. *Also see* perigee.

A positive (A+ or A plus) — 1. Positive terminal of a battery or positive polarity of any other sources of voltage. 2. The terminal to which the positive side of the filament-voltage source of a vacuum tube should be connected.

A power supply — A power supply used as a source of heating current for the cathode or filament of a vacuum tube.

apparatus — 1. Any complex device. 2. Equipment or instruments used for a specific purpose.

apparatus wire and cable — Insulated wire and cable used in connecting electrical apparatus to a power source, also including wire and cable used in the apparatus itself.

apparent bearing — The direction from which the signal arrives with respect to some reference direction.

apparent power — In an ac circuit, the power value obtained by simple multiplication of current by voltage with no consideration of the effect of phase angle. (Compare with *true power.*)

apparent power loss — For voltage-measuring instruments, the product of nominal end-scale voltage and the resulting current. For current-measuring instruments, the product of the nominal end-scale current and the resulting voltage. For other types of instruments (for example, wattmeters), the apparent power loss is expressed for a stated value of current or voltage. Also called volt-ampere loss.

apparent source — *See* effective acoustic center.

Applegate diagram — A graphical representation of electron bunching in a velocity-modulated tube, showing their positions along the drift space. This bunching is plotted on the vertical coordinate, against time along the horizontal axis.

applet — A small computer program that performs a simple task.

AppleTalk — A networking protocol developed by Apple Computer for communication between Apple Computer products and other computers. This protocol is independent of what network it is layered on.

Appleton layer — In the ionosphere, a region of highly ionized air capable of reflecting or refracting radio waves back to earth. It is made up of the F_1 and F_2 layers.

apple tube — A color-television picture tube in which the three colors of phosphors are laid in fine vertical strips along the screen. The intensity of the electron beam is modulated as its sweeps over them so that each color is produced with appropriate brightness.

appliance — Any electrical equipment used in the home and capable of being operated by a nontechnical person. Included are units that perform some task that could be accomplished by other, more difficult means, but usually not those used for entertainment (radios, TVs, hi-fi sets, etc.).

appliance wire and cable — A classification of Underwriters' Laboratories, Inc., covering insulated wire and cable intended for internal wiring of appliances and equipment. Each construction satisfies the requirements for use in particular applications.

application — 1. The use of a computer for a specific purpose, e.g., designing a brochure or writing a letter. 2. System or problem to which a computer is applied. An application may be of the computational type, in which arithmetic computations predominate, or of the data-processing type, in which data-handling operations predominate. *See also* application program.

application factor — A modifier of the failure rate. It is based on deviations from rated operating stress (usually temperature and one electrical parameter).

application-oriented language — 1. A programming language that is primarily useful in some specialized area. 2. A problem-oriented programming language whose statements resemble or contain the terminology of the computer user.

application program — 1. A computer program intended to solve a problem or do a job, as distinct from systems programs, which control the operations of the computer system. 2. A computer program that performs a data-processing function rather than a control operation. 3. A program used to perform some logical or computational task that is important to the user rather than some internal computer function. 4. Software designed for a specific purpose, such as accounts payable, inventory, payroll, and word processing. 5. A computer program that accomplishes specific tasks, such as word processing.

application schematic diagram — Pictorial representation using symbols and lines to illustrate the interrelation of a number of circuits.

application-specific integrated circuit — *See* ASIC.

applications software — 1. A program that depends on the specific end application and is used to do the real work or apparent work that is visible to the user. Generally this is the software that is used for dedicated computer-based systems (systems designed to perform a single or specific set of functions). Typical applications include food and chemical processing, production control, automotive electronics, computer-controlled sewing machines, photographic equipment (both for computer-controlled cameras and for darkroom computerized processing), energy distribution systems, word processing, mailing lists, payrolls, and inventory. 2. Computer programs that perform specific tasks, such as word processing or database management.

applicators (applicator electrodes) — 1. In dielectric heating, the electrodes between which the dielectric item is placed and the electrostatic field developed. 2. Appropriately shaped conducting surfaces between which an alternating electric field is established for the purpose of producing dielectric heating. 3. In medical electronics, the electrodes applied to a patient undergoing diathermy or ultrasonic therapy.

applied voltage — 1. The potential between a terminal and a reference point in any circuit or device. 2. The voltage obtained when measuring between two given points in a circuit with voltage applied to the complete circuit. 3. The voltage presented to a circuit point or system input, as opposed to the voltage drop resulting from current through an element that results from the applied voltage.

applique circuit — A special circuit provided to modify existing equipment in order to allow for some special usage.

approach-control radar — Any radar set or system used in a ground-controlled approach system, e.g., an airport-surveillance radar, precision approach radar, etc.

approach path — In radio aircraft navigation, that portion of the flight path in the immediate vicinity of a landing area where such a flight path terminates at the touchdown point.

approved circuit — *See* protected wireline distribution system.

APT — Abbreviation for automatically programmed tool. A high-level or simplified programming language.

AQL — Abbreviation for acceptable quality level. A statistically defined quality level, in terms of percent defective accepted on an average of 95 percent of the time. In other words, a sampling plan with 1 percent AQL passes (accepts) lots 1 percent defective 95 percent of the time.

aquadag layer—Trademark of Acheson Industries, Inc. A conductive graphite coating on the inner side walls of some cathode-ray tubes. It serves as an electrostatic shield or as a postdeflection and an accelerating anode. Also applied to outer walls and grounded; here it serves, with the inner coating, as a capacitor to filter the applied high voltage.

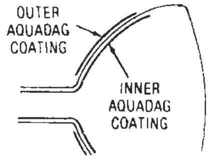

Aquadag coating.

arbiter—The section of a multiprocessor network's operating system that resolves simultaneous system-resource requests.

arbitrary function fitter—A circuit having an output voltage or current that is a presettable, adjustable, usually nonlinear function of the input voltage(s) or current(s) fed to it.

arbitrary waveform generator—*See* waveform generator.

arbor—*See* mandrel.

arc—1. A luminous discharge of electricity through a gas. Characterized by a change in space potential in the immediate vicinity of the cathode; this change is approximately equal to the ionization potential of the gas. 2. A prolonged electrical discharge, or series of prolonged discharges, between two electrodes. (Both produce a bright-colored flame, as contrasted with a dim corona-glow discharge). 3. To form an arc. 4. The electric current in a flamelike stream of incandescent gas particles. 5. An electric current through air or across the surface of an insulator associated with high voltage; usually occurs when a contact is opened, or when deenergizing an inductive load.

arcback—Also called backfire. Failure of the rectifying action in a tube, resulting from the flow of a principal electron stream in the reverse direction due to the formation of a cathode spot on the anode. This action limits the peak inverse voltage that may be applied to a particular rectifier tube.

arc converter—A form of oscillator utilizing an electric arc to generate an alternating or pulsating current.

arc discharge—1. A discharge between electrodes in gas or vapor. Characterized by a relatively low voltage drop and a high current density. 2. The sustained, luminous thermionic discharge between anode and cathode in a gas-filled tube.

arc-discharge tube—A gas-filled or mercury-vapor tube that utilizes ionic phenomena for switching, voltage regulation, or rectification.

arc drop—The voltage drop between the anode and cathode of a gas rectifier tube during conduction.

arc-drop loss—In a gas tube, the product of the instantaneous values of arc-drop voltage and current averaged over a complete cycle of operation.

arc-drop voltage—The voltage drop between the anode and cathode of a gas rectifier tube during conduction.

arc failure—1. A flashover in the air near an insulation surface. 2. An electrical failure in the surface heated by a flashover arc. 3. An electrical failure in the surface damaged by the flashover arc.

arc function—An inverse trigonometric function.

arc furnace—An electric furnace heated by arcs between two or more electrodes.

architecture—1. Organizational structure of a computing system, mainly referring to the CPU or microprocessor. 2. The manner in which the basic computer functions are organized and partitioned on the silicon chips. 3. The manner in which a system (such as a network or a computer) or program is structured.

archival—Pertaining to long-term storage of data.

archival backup—Backing up only files that have been changed since the last backup.

archive—1. A procedure for transferring information from an online storage diskette or memory area to an offline storage medium. 2. To copy computer programs and data onto an auxiliary storage medium, such as a disk or tape, for long-term retention.

arcing—The production of an arc, e.g., at the brushes of a motor or at the contact of a switch.

arcing contacts—Special contacts on which the arc is drawn after the main contacts of a switch or circuit breaker have opened.

arcing time—1. The interval between the parting, in a switch or circuit breaker, of the arcing contacts and the extension of the arc. 2. The time elapsing, in a fuse, from the severance of the fuse link to the final interruption of the circuit under the specified condition.

arc lamp—Source of brilliant artificial light obtained by an electric arc passing between two carbon rods. The arc is struck by bringing the two rods together and then rapidly separating them. As the arc burns, the carbon rods are vaporized away. A mechanism is employed to keep the space between the two rods constant. This type of lamp is used extensively in motion picture projectors and spotlights. The illumination of the arc lamp is derived from the incandescence of the positive electrode and from the heated, luminous, ionized gases or vapor that surround the arc.

arc oscillator—A negative-resistance oscillator comprising a sustained dc arc and a resonant circuit.

arcover—The (usually abrupt) creation of an arc between electrodes, contacts, or plates of a capacitor.

arcover resistance—The resistance of a material to the effects of a high-voltage, low-current arc (under prescribed conditions) passing across the surface of the material. The resistance is stated as a measure of total elapsed time required to form a conductive path on the surface (material carbonized by the arc).

arcover voltage—Under specified conditions, the minimum voltage required to create an arc between electrodes separated by a gas or liquid insulation.

arc percussive welding—A type of welding in which the materials to be welded are separated by a gap, across which an arc is struck; the arc melts the surfaces of the materials, and the materials are simultaneously brought together. *See also* pulse arc welding.

arc resistance—The length of time that a material can resist the formation of a conductive path by an arc adjacent to the surface of the material. Also called tracking resistance.

arc suppressor—A device, or combination of devices, used for arc suppression. *See* spark suppressor.

arc-through—In a gas tube, a loss of control with the result that a principal electron stream flows in the normal direction during what should be a nonconducting period.

area code — A three-digit number code identifying one of the geographic areas of the United States, Canada, and Mexico to permit direct distance dialing on the telephone system. The area code precedes the central office code in the complete 10-digit telephone number, and must be used when the called telephone is in a numbering plan area different from that of the calling telephone. The first digit of the area code is never a 1 or 0. *See also* direct distance dialing.

area protection — Protection of the inner space or volume of a secured area by means of a volumetric sensor.

area redistribution — A method of measuring the duration of irregularly shaped pulses. A rectangle is drawn having the same peak amplitude and the same area as the original pulse under consideration. Because the same time units are used in measuring the original and the new pulse, the width of the rectangle is considered the duration of the pulse.

area sensor — A sensor with a detection zone that approximates an area, such as a wall surface or the exterior of a safe.

A register — The accumulator for all arithmetical operations in a computer. Also called A accumulator.

argon — An inert gas used in discharge tubes and some electric lamps. It gives off a purple glow when ionized; its symbol is Ar.

argon glow lamp — A glow lamp containing argon gas that produces a pale blue violet light.

argument — 1. A variable upon which the value of a function depends. The arguments of a function are listed in parentheses after the function name. The computations specified by the function definition are made with the variables specified as arguments. 2. The number that a function works on to produce its results. 3. The independent variable of a function. Arguments can be passed as part of a subroutine call where they would be used in that subroutine.

arithmetic and logic unit — Computer element that can perform the basic data manipulations in the central processor. Usually it can add, subtract, complement, negate, rotate, AND, and OR. Abbreviated ALU.

arithmetic capability — The ability to do addition, subtraction, and in some cases multiplication and division.

arithmetic check — A check of a computation making use of the arithmetical properties of the computation.

arithmetic element — Synonym for arithmetic unit.

arithmetic mean — 1. Usually, the same as *average*. It is obtained by first adding quantities together and then dividing by the number of quantities involved. 2. A figure midway between two extremes and is found by adding the minimum and maximum together and dividing by two.

arithmetic operation — 1. In an electronic computer, the operations in which numerical quantities form the elements of the calculation, including the fundamental operations of arithmetic (addition, subtraction, multiplication, comparison, and division). 2. Adding, subtracting, incrementing, or decrementing data in registers or memory.

arithmetic organ — *See* arithmetic unit.

arithmetic shift — In a digital computer, the multiplication or division of a quantity by a power of the base used in the notation.

arithmetic statement — 1. An expression and a variable separated by an equals sign. The expression is evaluated and the resulting value is assigned to the variable. 2. Instruction specifying an arithmetic operation.

arithmetic sum — The sum of two or more quantities regardless of their signs. Compare with *algebraic sum.*

arithmetic symmetry — Filter response showing mirror-image symmetry about the center frequency when frequency is displayed on an arithmetic scale. Constant envelope delay in bandpass filters usually is accompanied by arithmetic symmetry in the phase and amplitude responses and generally requires a computer design. *See also* geometric symmetry.

arithmetic unit — Also called arithmetic element or arithmetic organ. In an automatic digital computer, that portion in which arithmetical and logical operations are performed on elements of information.

armature — The moving element in an electromechanical device, such as the rotating part of a generator or motor, the movable part of a relay, or the spring-mounted, iron portion of a bell or buzzer.

armature contacts — 1. Contacts mounted directly on the armature. 2. Sometimes used for movable contacts.

armature control of speed — The varying of voltage applied to the armature of a shunt-wound motor to control the motor's speed over the basic speed range.

armature core — An assembly of laminations forming the magnetic circuit of an armature.

armature gap — The space between the armature and pole face.

armature hesitation — A delay or momentary reversal of the motion of the armature.

armature-hesitation contact chatter — Chatter caused by delay or momentary reversal in direction of the armature motion of a relay during either the operate or the release stroke.

armature-impact contact chatter — Chatter caused by impact of the armature of a relay on the pole piece in operation, or on the backstop in release.

armature overtravel — That portion of the available stroke occurring after the contacts of a relay have touched.

armature reaction — In an armature, the reaction of the magnetic field produced by the current on the magnetic lines of force produced by the field coil of an electric motor or generator.

armature rebound — Return motion of a relay armature after striking the backstop.

armature-rebound contact chatter — Chatter caused by the partial return of the armature of a relay to its operated position as a result of rebound from the backstop in release.

armature relay — A relay operated by an electromagnet that, when energized, causes an armature to be attracted to a fixed pole or poles.

armature slot — In the core of an armature, a slot or groove into which the coils or windings are placed.

armature stud — In a relay, an insulating member that transmits the motion of the armature to an adjacent contact member.

armature travel — The distance traveled during operation by a specified point on the armature of a relay.

armature voltage control — A means of controlling the speed of a motor by changing the voltage applied to its armature windings.

armature wire — Stranded annealed copper wire, straight lay, soft loose white cotton braid. It is used for low-voltage, high-current rotor winding motors and generators. Straight lay permits forming in armature slots, and compressibility.

armchair copy — Amateur term for clear, static-free signals.

armed sweep — *See* single sweep.

arming the oscilloscope sweep — Closing a switch that enables the oscilloscope to trigger on the next pulse.

armor — A braid or wrapping of metal, usually steel, placed over the insulation of wire or cable to protect it from abrasion or crushing.

armor clamp—A fitting for gripping the armor of a cable at the point where the armor terminates or where the cable enters a junction box.

armored cable—Two or more insulated wires collectively provided with a metallic covering, primarily to protect the insulated wires from damage.

Armstrong frequency-modulation system—A phase-shift modulation system originally proposed by E. H. Armstrong.

Armstrong oscillator—An inductive feedback oscillator that consists of a tuned gate circuit and an untuned tickler coil in the drain circuit. Feedback is controlled by varying the coupling between the tickler and the gate circuit.

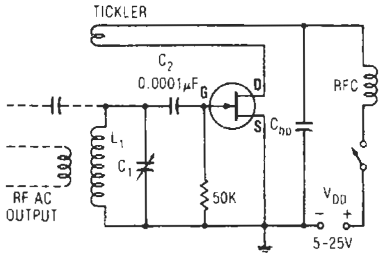

Armstrong oscillator.

arr—*See* automatic repeat request.

array—1. In an antenna, a group of elements arranged to provide the desired directional characteristics. These elements may be antennas, reflectors, directors, etc. 2. A series of items, not necessarily arranged in a meaningful pattern. 3. The group of patterns on a wafer or in the artwork or photomask for semiconductor processing. *See* random-access memory.

array antenna—An antenna comprising a number of radiating elements, generally similar, arranged and excited to obtain directional effects.

array device—A group of many similar, basic, complex, or integrated devices without separate enclosures. Each has at least one of its electrodes connected to a common conductor, or all are connected in series.

array noise—Unwanted disturbance in a memory integrated circuit generated by the normal movement of data within the array.

array processor—1. A computer optimized in architecture and instruction set to handle programs involving computations on large batches of data, such as fast Fourier transforms and large matrix computations. An array processor takes blocks of data and instructions from a host mini or large computer and performs the computations at speeds many times as high as those that are possible through the host computer alone. The host may be considered the data-organizing front end; the array processor is the processing unit. 2. A computer dedicated by its design to performing repetitive arithmetical calculations on large arrays of data with high precision, wide dynamic range, and high throughput. Usually most input/output operations and file management chores are left to the host computer in order to free the peripheral array processor to concentrate on its calculations. 3. A single computer that operates on one piece of data at a time. 4. A processor in a computer that performs matrix arithmetic much faster than is done in a standard computer. Capable of performing operations on all the elements in large matrices at one time. Also called a vector processor.

arrester—Also called a lightning arrester. 1. A protective device used to provide a bypass path directly to ground for lightning discharges that strike an antenna or other conductor. 2. A power-line device capable of reducing the voltage of a surge applied to its terminals, interrupting current, if present, and restoring itself to original operating conditions. 3. Device that diverts high voltages to ground and away from the equipment it protects.

ARRL—Abbreviation for American Radio Relay League.

arrowhead—A linearly polarized, frequency-independent, log-periodic antenna.

ARSR—Abbreviation for air route surveillance radar.

ARTCC—Abbreviation for air route traffic control center. A complex data-handling facility designed by Burroughs, IBM, and Raytheon to computerize as much in-route air traffic control as possible.

articulation—Sometimes called intelligibility. 1. In a communications system, the percentage of speech units understood by a listener. The word *articulation* is customarily used when the contextual relationships among the units of speech material are thought to play an unimportant role; the word *intelligibility* is used when the context is thought to play an important role in determining the listener's perception. 2. A quantitative measurement of the intelligibility of human speech, where 100 percent is completely understandable. For the typical sound reinforcement or other communications system, no more than a 15-percent articulation loss is acceptable. 3. The ability of a mechanism to pivot, grasp, or extend.

articulation equivalent—The articulation of speech reproduced over a complete telephone connection, expressed numerically in terms of the trunk loss of a working reference system that is adjusted to give equal articulation.

artificial antenna—Also called dummy antenna. A device that simulates a real antenna in its essential impedance characteristics and has the necessary power-handling capabilities, but which does not radiate or receive radio waves. Used mainly for testing and adjusting transmitters.

artificial ear—A microphone-equipped device for measuring the sound pressures developed by an earphone. To the earphone it presents an acoustic impedance equivalent to the impedance presented by the human ear.

artificial echo—1. Received reflections of a transmitted pulse from an artificial target, such as an echo box, corner reflector, or other metallic reflecting surface. 2. A delayed signal from a pulsed radio-frequency signal generator.

artificial horizon—A gyroscopically operated instrument that shows, within limited degrees, the pitching and banking of an aircraft with respect to the horizon. Lines or marks on the face of the instrument represent the aircraft and the horizon. The relative positions of the two are then easily discernible.

artificial intelligence—Abbreviated AI 1. The design of computer and other data-processing machinery to perform increasingly higher-level cybernetic functions. 2. The capability of a device to perform functions that are normally associated with human intelligence, such as reasoning, learning, and self-improvement. Related to machine learning. 3. The imitation by artificial systems of characteristics described as intelligent when observed in humans. Artificial intelligence embraces concepts and theories from many different disciplines, including mathematics, cybernetics, computer science, psychology, biology, and others. 4. Overlapping subsets called expert systems, knowledge representations, inference schemes, program synthesis, scene analysis, and robotics. 5. The ability of a machine to perform certain complex functions normally associated with human intelligence, such as judgment, pattern recognition, understanding, learning, planning, and problem solving. 6. Computer programs developed to mimic human intelligence, such as reasoning, learning, problem solving, and making decisions. Artificial intelligence programs enable computers to perform tasks such as playing chess, proving mathematical theorems, etc. 7. An area of computer science dedicated to the development of machines that can learn, understand, interpret, and arrive at conclusions in a manner that would be considered intelligent if a person were doing it.

artificial ionization—Introduction of an artificial reflecting or scattering layer into the atmosphere to permit beyond-the-horizon communications.

artificial language—In computer terminology, a language designed for ease of communication in a particular area of activity, but one that is not yet natural to that area (as contrasted with a natural language evolved through long usage).

artificial line—A lumped-constant network designed to simulate some or all the characteristics of a transmission line over a desired frequency range.

artificial line duct—A balancing network simulating the impedance of the real line and distant terminal apparatus. It is employed in a duplex circuit to make the receiving device unresponsive to outgoing signal currents.

artificial load—Also called dummy load. A dissipative but essentially nonradiating device having the impedance characteristics of an antenna, transmission line, or other practical utilization circuit. Energy is dissipated in the form of heat. Used to test radio transmitters, engine generators, etc. Permits testing under load conditions without the creation of any standing waves or radiating a signal.

artificial radioactivity—Radioactivity induced in stable elements under controlled conditions by bombarding them with neutrons or high-energy, charged particles. Artificially radioactive elements emit beta and/or gamma rays.

artificial voice—A small speaker mounted in a specially shaped baffle that is proportioned to simulate the acoustical constants of the human head. It is used for calibrating and testing close-talking microphones.

Artos stripper—A machine that, when properly adjusted, will automatically measure to a predetermined length, cut, strip, count, and tie wire in bundles.

ARTS—Abbreviation for Automated Radar Terminal System. A multiprocessor computing system used at terminal radar approach controls (tracons) and airport towers, based on Sperry Univac 1140 computers. Various configurations, differing in size of memory and number of processors, exist at different facilities; the largest, ARTS IIIE, is installed at the New York tracon.

artwork—1. A topological pattern of an integrated circuit, made with accurate dimensions so that it can be used in mask making. Generally, it is a large multiple of the final mask size, and final reduction is accomplished through the use of a step-and-repeat camera. 2. Detailed, original drawing (often developed with the aid of a computer) showing layout of an integrated circuit. 3. The images formed by drawing, scribing, or by cutting and stripping on a film or glass support, which are reduced, contact-printed, or stepped and repeated to make a photomask or intermediate. 4. Layouts and photographic films created to produce thick-film screens and thin-film masks.

ASA—Abbreviation for American Standards Association. *See* American National Standards Institute.

ASA code—A code that was recommended by the American Standards Association for industry-wide use in the transmission of information. Now ANSI code.

asbestos—A nonflammable material generally used for heat insulation, such as in a line-cord resistor. No longer used on new equipment.

A-scan—Also called A-display. On a cathode-ray indicator, a presentation in which time (range or distance) is one coordinate (horizontal) and signals appear as perpendicular deflections to the time scale (vertical).

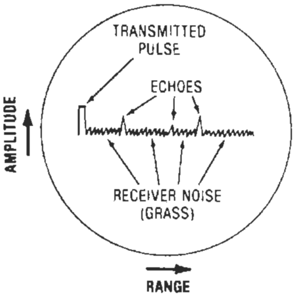

A-scan on a CRT.

ASCII—Acronym for American Standard Code for Information Interchange (pronounced "ask-ee"). 1. A standard code used extensively in data transmission, in which 128 numerals, letters, symbols, and special control codes are each represented by a seven-bit binary number 0000000 through 1111111. For example, numeral 5 is represented by 011 0101, letter K by 100 1011, percent symbol (%) by 010 0101, and start of text (STX) control code by 000 0010. 2. A standard code used by many computers, video-display terminals, teleprinters, and computer peripherals. A full eight-bit ASCII word may be transmitted in parallel or serial form, with the eighth bit often providing parity information. Keyboard encoders convert a single switch closure into an ASCII data word; character generators convert stored ASCII data words (and some timing commands) into groups of dots suitable for raster-scan display.

A-scope—An oscilloscope that uses an A-scan to present the range of a target as the distance along a horizontal line from the transmitted pulse pip to the target, or echo pip. Signals appear as vertical excursions of the horizontal line, or trace.

ASDE—*See* airport surface detection equipment.

as-fired—Values of thick-film resistors or smoothness of ceramic substrates as they come out of the firing furnace prior to trimming and polishing, respectively (if required).

ASI—An abbreviation for standards published by the American Standards Institute. Now American National Standards Institute.

ASIC—Abbreviation for application-specific integrated circuit. 1. Semiconductor circuits specifically designed to suit a customer's particular requirement, as opposed to general-purpose parts that can be used in many different systems or applications. 2. An integrated circuit designed to fill the specific requirement of a unique application.

ask—*See* amplitude-shift keying.

aspect ratio—1. Ratio of frame width to frame height. 2. The ratio of an object's height to its width. In graphics this ratio usually pertains to the face of a rectangular CRT or to the characters or symbols drawn by the character generator. 3. The ratio between the length of a film resistor and its width: equal to the number of squares of the resistor. 4. The ratio of the width of a television picture tube to its height. In the United States, the television standard is 4 : 3.

asperities—Local microscopic points on an electrode surface at which there is considerable field enhancement. They lead to a dependence of electric strength on electrode area (area effect).

aspheric— 1. Not spherical; an optical element having one or more surfaces that are not spherical. 2. A mirror or lens surface that varies slightly from a true spherical surface. This is done to reduce lens aberrations.

ASR—Abbreviation for automatic send and receive and airport surveillance radar. 1. A terminal equipped with recording devices, usually a paper-tape reader and punch, which is capable of answering a call, recording a message, or sending data loaded in its tape reader without the need for an operator in attendance at the time of the call. Also used to specify terminals that have paper-type equipment used by the operator. 2. A teletypewriter that contains a keyboard, page printer, paper-tape transmitter, and paper-tape punch. Paper tape can be prepared offline, which can take place while hard copy is being received from the line or while other paper tape is being transmitted.

ASRA—Abbreviation for automatic stereophonic recording amplifier. An instrument developed by Columbia Broadcasting System for stereo recording. Compression of the vertical component of the stereo recording signal is automatically decreased or increased as required by the recording conditions.

assemble— 1. To collect, interpret, and coordinate the data required for a computer program, translate the data into computer language, and project it into the master routine for the computer to follow. 2. To translate from a symbolic program to a binary program by substituting binary operation codes for symbolic operation codes and replacing symbolic addresses with absolute or relocatable addresses.

assembler— 1. A program that prepares a program in machine language from a program in symbolic language by substituting absolute operation codes for symbolic operation codes and absolute or relocatable addresses for symbolic addresses. 2. A unit that converts the assembly language of a computer program into the machine language of the computer, accepting mnemonics and symbolic addresses instead of actual binary values for addresses, instructions, and data. 3. A program that accepts instructions, addresses, and data in symbolic form (character strings that represent machine instructions, addresses, data, among others). Then it automatically translates symbols into their corresponding numerical values. It permits symbolic addressing by assigning values to labels used to indicate program-jump locations. 4. A simple programming language that allows the programmer to define labels and fixed values and to then use these labels with a mnemonic instruction set to produce a machine code program. 5. Program that converts source-code (mnemonic) input into op-code (binary) machine language instructions. If such a program were not available, the programmer would have to enter all instructions in ones and zeros, a much more tedious and error-prone procedure. 6. A computer program that converts a higher-level (Englishlike) programming language into machine-readable instructions.

assembler program—Software, usually supplied by the computer manufacturer, to convert an assembly language application program into machine language.

assembly— 1. A complete operating unit, such as a radio receiver, made up of subassemblies, such as an amplifier and various components. 2. Process in which instructions written in symbolic form by the programmer are changed to machine language by the computer.

assembly language— 1. A computer language that has one-to-one correspondence with an assembly program. The assembly program directs a computer to operate on a program in symbolic language to produce a program in machine language. *See also* high-order language; machine language, 3; and source language. 2. Grouped alphabet characters, called mnemonics, that replace the numeric instructions of machine language. These mnemonics are easier to remember than machine instructions and hence easier to develop into a working program. 3. A machine-oriented language based primarily on a one-to-one relationship between machine instructions and user-supplied source code. 4. Microprocessor commands written in mnemonic form. Typically, three-letter abbreviations, called mnemonics, are used to represent each instruction, and each mnemonic can usually be equated to one machine-code instruction. 5. A human-oriented symbolic-mnemonic source language that is used by the programmer to encode programs and associated databases. Assembly language programs are read by the assembler and converted to executable machine language programs during the assembly process. Assembly language is easier to remember and manipulate than machine language. 6. Human-oriented varieties of machine languages. Precisely the same final program code can be produced from an assembler as by hand-coding machine language. However, assembly languages prove more convenient for people than the numeric-only machine languages.

assembly-language programming—*See* symbolic-language programming.

assembly-output language—An optional symbolic assembly language listing of the object-code output from a high-level language compiler. Can be quite helpful as a debugging tool because it shows exact machine code in a readable format.

assembly program—A program that enables a computer to assemble mnemonic language into machine language; for example, a FORTRAN assembly program. Also called assembly routine.

assembly robot—A computerized robot, probably a sensory model, designed specifically for assembly-line jobs. For light, batch-manufacturing applications, the arm's design may be fairly anthropomorphic.

assembly routine—*See* assembly program.

assertion checking—Evaluating a program by embedding statements that should always hold true.

assignable cause—A definitely identified factor contributing to a quality variation.

assigned frequency—The center of the frequency band assigned to a station.

assigned frequency band—The frequency band, the center of which coincides with the frequency assigned to the station, and the width of which equals the necessary bandwidth plus twice the absolute value of the frequency tolerance.

associative memory—A computer memory in which the data are stored and indexed by content, as in a dictionary, in contrast with the storage of a random-access memory. 2. A memory in which the storage locations are identified by their contents rather than by their addresses. Enables faster interrogation to retrieve a particular data element.

associative storage—Computer storage in which locations may be identified by specification of part or all of their contents. Also called parallel-search storage or content-addressed storage.

astable— 1. Pertaining to a device that has two temporary states: the device alternates between these states with a period and duty cycle determined by circuit time constants. *See also* bistable. 2. Refers to a device that has two temporary states. The device oscillates between the two states with a period and duty cycle predetermined by time constants.

astable circuit—A circuit that continuously alternates between its two unstable states at a frequency determined by the circuit constants. It can be readily synchronized by applying a repetitive input signal of slightly higher frequency. A blocking oscillator is an example of an astable circuit.

astable multivibrator (free-running)—A circuit having two momentarily stable states, between which it continuously alternates, remaining in each for a period controlled by the circuit parameters and switching rapidly from one to the other.

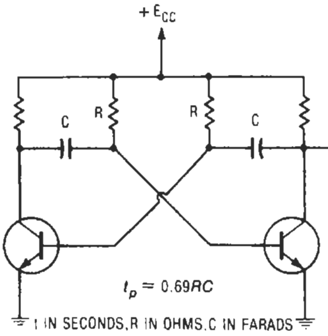

Circuit.

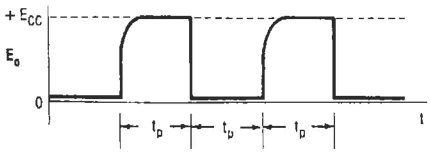

Output.

Astable multivibrator.

astatic—1. Having no particular orientation or directional characteristics; such as a vertical antenna. 2. Being in neutral equilibrium; having no tendency toward any change of position.

astatic galvanometer—A sensitive galvanometer used for detecting small currents. Consists of two small magnetized needles of equal size and strength arranged in parallel and with their north and south poles adjacent, suspended inside the galvanometer coil. Since the resultant magnetic moment is zero, the earth's magnetic field does not affect the system.

A station—One of a pair of transmitting stations in a loran system. The A-station signal always occurs less than half a repetition period after the immediately preceding signal of the other station of the pair and more than half a repetition period before the next succeeding signal of the other station.

astigmatism—A type of spherical aberration in which the rays from a single point of an object do not converge on the image, thereby causing a blurred image. Astigmatism in an electron-beam tube is a focus defect in which electrons in different axial planes come to focus at different points.

astrionics—Electronics as involved with astronautics.

astrocompass—An instrument for determining direction relative to the stars. It is unaffected by the errors to which magnetic or gyrocompasses are subject.

astrodome—A rigid hemispherical structure used to cover large tracking instruments to protect them from the elements. It is usually constructed so that the dome rotates with the instrument.

astronautics—The science and art of operating space vehicles.

astrotracker—A device for tracking stars.

A-supply—The A battery, transformer filament winding, or other voltage source that supplies power for heating the filaments of vacuum tubes.

asychronous motor—An ac motor whose speed is not proportional to the frequency of the supply voltage.

asymmetrical cell—A cell, such as a photoelectric cell, in which the impedance to the flow of current is greater in one direction than in the other direction.

asymmetrical distortion—Distortion affecting a two-condition or binary modulation or restitution, in which all the significant intervals corresponding to one of the two significant conditions have longer or shorter durations than the corresponding theoretical durations of the excitation. If this particular requirement is not met, distortion is present.

asymmetrical SCR—A fast silicon-controlled rectifier (SCR) with low reverse blocking, which is voltage that causes conduction without an input trigger at the gate input of the SCR.

asymmetric sideband transmission—*See* vestigial sideband transmission.

asymmetry control—In pH meters, an adjustment sometimes provided to compensate for differences in the electrodes.

asymptote—A line that comes nearer and nearer to a given curve but never touches it.

asymptotic breakdown voltage—A voltage that will break down insulation if applied over a long period.

asynchronous—1. A communication method in which data is sent when it is ready without being referenced to a timing clock, rather than waiting until the receiver signals that it is ready to receive. 2. Transmission in which each data byte is preceded by a start bit and followed by one or more stop bits. Data transmission is intermittent, with an irregular time interval between data bytes. 3. Lacking a regular time relationship; not related through repeating time patterns. Hence, as applied to computer program execution, unexpected or unpredictable with respect to the instruction sequence. 4. Modems, terminals, and transmissions in which each character of information is individually framed (synchronized), usually by start and stop elements. The interval between characters is not fixed. 5. An external interface that can be started and stopped by a microprocessor or other equipment. The opposite is *synchronous*, which means that the data is randomly available. 6. Having no set pattern, cycle, or speed of transmission. 7. Not synchronized by a clocking signal; in code sets, character codes containing start and stop bits. 8. A mode of data transmission in which time intervals between transmitted characters may be of unequal length. Transmission is independently controlled by start and stop elements at the beginning and end of each character. 9. Pertaining to a mode of data communications that provides a variable time interval between characters during transmission.

asynchronous communication—1. A method of transferring data in which the timing of character placement on connecting communication lines is not critical. Each transferred character is preceded by a start bit and followed by a stop bit, permitting the interval between characters to vary. 2. A relatively simple and cheap system for moving data between machines at speeds up to 1200 or 2400 baud. 3. Data communication of the start/stop type. Each character is sent individually without regular or predictable time relationships with other characters. 4. Method of communications in which data is sent as soon as it is ready.

asynchronous computer — 1. An automatic digital computer in which an operation is started by a signal denoting that the previous operation has been completed. 2. A computer in which each operation starts as a result of a signal generated by the completion of the previous operation or by the availability of the equipment required for the next operation.

asynchronous device — A device in which the speed of operation is not related to any frequency in the system to which it is connected.

asynchronous input/output — The ability to accept input data while simultaneously delivering output data.

asynchronous inputs — The terminals that affect the output state of a flip-flop independently of the clock terminals. Called set, preset, reset, or clear; sometimes referred to as dc inputs.

asynchronous logic — Logic networks whose operational speed depends only on the signal propagation through the network, rather than on clock pulses as in synchronous logic.

asynchronous machine — 1. Any machine in which its speed of operation is not proportionate to the frequency of the system to which the machine is connected. 2. A multiprocessor system whose processes occur as needed by, and whose operations follow, input data instead of an autonomous clock.

asynchronous operation — 1. Generally, an operation that is started by a signal at the completion of a previous operation. It proceeds at the maximum speed of the circuits until it is finished and then generates its own completion signal. 2. A mode in which entry of data into a flip-flop does not require a gating or clock pulse. 3. Operation of a switching network by a free-running signal that triggers successive instructions. The completion of one instruction triggers the next. 4. A computer operation that does not proceed in step with some external timing.

asynchronous shift register — A shift register that does not require a clock. Register segments are loaded and shifted only at data entry.

asynchronous transmission — Transmission in which each character of the information is synchronized individually (usually by the use of start and stop elements).

ATARS — Acronym for Automated Traffic Advisory and Resolution Service. A ground-based automatic collision-avoidance system being developed for use at air terminals.

ATC — 1. Abbreviation for automatic temperature control (General Motors). A means of automatically maintaining desired passenger compartment temperature in a vehicle. Temperature is sensed by a thermistor. Control valves are actuated by a vacuum motor to adjust the proportion of heated or cooled air. 2. Also abbreviation for automated technical control. A computer system used to maintain control of a data-communication network.

AT-cut crystal — A quartz-crystal slab cut at a 35° angle with respect to the optical, or Z-axis, of the crystal. It has practically a zero temperature coefficient and is used at frequencies of about 0.8 to 250 MHz.

ATE — Abbreviation for automatic test equipment.

ATG — *See* automatic test generation.

atmosphere — 1. The body of air surrounding the earth. 2. A unit of pressure defined as the pressure of 760 mm of mercury at 0°C. Approximately 14.7 pounds per square inch.

atmospheric absorption — The energy lost in the transmission of radio waves due to dissipation in the atmosphere.

atmospheric absorption noise — The dominant noise factor, at frequencies above 1000 MHz, caused by the absorption of energy from radio waves by oxygen and water vapor in the atmosphere.

atmospheric duct — Within the troposphere, a condition in which the variation of refractive index is such that the propagation of an abnormally large proportion of any radiation of sufficiently high frequency is confined within the limits of a stratum. This effect is most noticeable above 3000 MHz.

atmospheric electricity — Static electricity between clouds, or between clouds and the earth.

atmospheric noise — Also called atmospherics. 1. The noise heard during radio reception because of atmospheric interference. 2. A product of the discharging of lightning and other phenomena in the atmosphere.

atmospheric pressure — 1. The barometric pressure of air at a particular location on the earth's surface. The nominal, or standard, value of atmospheric pressure is 760 mm of mercury (14.7 pounds per square inch) at sea level. Atmospheric pressure decreases at higher altitudes. 2. Pressure exerted by the atmosphere on all things exposed to it. Although it varies constantly, it is considered as a standard to be normal when it is 14.7 lb/in^2 (1.033×10^4 kg/m^2) at sea level.

atmospheric radio wave — A radio wave that is propagated by reflections in the atmosphere. May include the ionospheric wave, the tropospheric wave, or both.

atmospheric radio window — That portion of the frequency spectrum that will allow radio-frequency waves to pass through the earth's atmosphere (approximately 10 to 10,000 MHz).

atmospheric refraction — The bending of the path of electromagnetic radiation from a distant point as the radiation passes obliquely through varying air densities.

atmospherics — Also referred to as static, atmospheric noise, and strays. In a radio tuner or receiver, noise due to natural weather phenomena and electrical charges existing in the atmosphere.

atom — 1. The smallest portion of an element that exhibits all properties of the element. It is pictured as composed of a positively charged nucleus containing almost all the mass of the atom, surrounded by one or more electrons. In the neutral atom, the number of electrons is such that their total charge (negative) exactly equals the positive charge in the nucleus. 2. The basic unit of a chemical element, consisting of a positively charged nucleus surrounded by a number of electrons sufficient to counterbalance the charge of the nucleus. The identity of an element, in a chemical sense, depends on the number of positive charges in the nucleus of its atom. The nucleus also contains particles that contribute mass but no charge. The stability of a nucleus depends on its ratio of charge to mass.

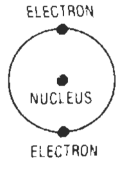

Atom.

atomic battery—*See* nuclear battery.

atomic charge—The electronic charge of an ion, equal to the number of ionization multiplied by the charge on one electron.

atomic fission—*See* fission.

atomic frequency—The natural vibration frequency of an atom.

atomic fuel—A fissionable material, i.e., one in which the atomic nucleus may be split to release energy.

atomic mass unit (unified)—One-twelfth of the mass of an atom of the ^{12}C nuclide. Use of the old atomic mass unit (amu), defined by reference oxygen, is deprecated.

atomic migration—The progressive transfer of a valence electron from one atom to another within the same molecule.

atomic number—The number of protons (positively charged particles) in the nucleus of an atom. All elements have different atomic numbers, which determine their positions in the periodic table. For example, the atomic number of hydrogen is 1, that of oxygen is 8, iron 26, lead 82, and uranium 92.

atomic ratio—The ratio of quantities of different substances to the number of atoms of each.

atomic theory—A generally accepted theory concerning the structure and composition of substances and compounds. It states that everything is composed of various combinations of ultimate particles called atoms.

atomic time—Time scales based on molecular or atomic resonance effects, which are apparently constant and equivalent (or nearly equivalent) to ephemeris time.

atomic weight—The approximate weight of the number of protons and neutrons in the nucleus of an atom. The atomic weight of oxygen, for example, is approximately 16 (actually it is 16.0044)—it contains 8 neutrons and 8 protons. Aluminum is 27 and contains 14 neutrons and 13 protons. If expressed in grams, these weights are called gram atomic weights.

atr switch—*See* antitransmit-receive switch.

atr tube—Abbreviation for antitransmit-receive tube. A gas-filled, radio-frequency switching tube used to isolate the transmitter while a pulse is being received.

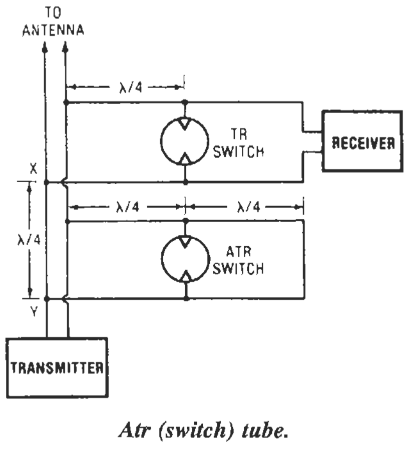

Atr (switch) tube.

ATS—Abbreviation for automotive anti-theft system (General Motors). A security system that causes the horn to blow, headlights and parking lights to flash, and the dome light to turn on if the hood, trunk, or doors are opened without use of a key.

attached foreign material—In a semiconductor, a foreign substance that cannot be removed when subjected to a nominal gas flow. Lint, silicon dust, etc., are not considered attached since they can be removed after die mounts.

attachment cap—*See* attachment plug.

attachment cord—*See* patch cord.

attachment plug—An assembly consisting of two or more blades projecting from a small insulating base, with provision for connecting the plug to a cord. Also called attachment cap.

attack—1. The length of time it takes for a tone in an organ to reach full intensity after a key is depressed. On most organs this effect is adjustable by either a switch or potentiometer. 2. The action of a control system in response to a sudden error condition. 3. The responsiveness of an amplifier to signals with a fast rise-time, such as produced by percussive sounds of a transient nature. 4. The beginning of a sound or the initial transient of a musical note.

attack time—The interval required for an input signal, after suddenly increasing in amplitude, to attain a specified percentage (usually 63 percent) of the ultimate change in amplification or attenuation due to this increase.

attendant's switchboard—A switchboard, of one or more positions, that permits an operator in the central office to receive, transmit, or cut in on a call to or from one of the lines serviced by the office.

attended operation—Data-set applications in which individuals are required at both stations to establish the connection and transfer the data sets from talk (voice) mode to data mode. Compare with *unattended operation.*

attention display—A computer-generated tabular or vector message placed on the display tubes of a control facility to draw attention to a particular situation.

attenuate—To obtain a fractional part of or reduce in amplitude an action or signal.

attenuating—Decreasing electrical current, voltage, or power in a communicating channel. Refers to audio, radio, or carrier frequencies.

attenuation—1. The decrease in amplitude of a signal during its transmission from one point to another. It may be expressed as a ratio or, by extension of the term, in decibels. 2. *See* insertion loss. 3. The decrease in amplitude of a signal at a specified frequency due to its transmission through a filter. This is expressed as a function of the amplitude ratio V_1/V_2, which is the reciprocal of the magnitude of the transmission function. 4. Optical power loss per unit length, usually expressed in decibels per kilometer. 5. Loss of electrical power in a length of cable. Amount of power leaving a length of cable as compared with the amount introduced. Attenuation is measured in decibels per 100 feet. 6. Applied to coaxial cables, the power drop or signal loss in a circuit, expressed in decibels. It is also the decrease in amplitude of a wave with distance in the direction of wave propagation when the amplitude at any given place is constant in time, or the decrease in amplitude with time at a given place. Attenuation is generally expressed in decibels per unit, usually 100 feet, and is indicative of the power loss.

attenuation constant—1. The real component of the propagation constant. 2. For a traveling plane wave at a given frequency, the rate at which the amplitude of a field component (or the voltage or current) decreases exponentially in the direction of propagation, in nepers or decibels per unit length.

attenuation distortion—1. In a circuit or system, its departure from uniform amplification or attenuation over the frequency range required for transmission. 2. Distortion that causes a decrease in the amplitude of a field component (voltage or current) in the direction of propagation.

attenuation equalizer—A corrective network designed to make the absolute value of the transfer impedance of two chosen pairs of terminals substantially constant for all frequencies within a desired range.

attenuation-frequency distortion—A form of wave distortion in which the relative magnitudes of the different frequency components of the wave are changed.

attenuation network—1. A network providing relatively slight phase shift and substantially constant attenuation over a range of frequencies. 2. The arrangement of circuit elements, usually impedance elements, inserted in circuitry to introduce a known loss or to reduce the impedance level without reflections.

attenuation ratio—The magnitude of the propagation ratio, which indicates the relative decrease in energy.

attenuator—1. A resistive network that provides reduction of the amplitude of an electrical signal without introducing appreciable phase or frequency distortion. 2. A distributed network that absorbs part of a signal and transmits the remainder with a minimum of distortion or delay. 3. Network for reducing signal level. Sometimes necessary in the input circuit of a tuner to avoid overloading by strong local signals. 4. An electronic transducer, either fixed or adjustable, that reduces the amplitude of a wave without causing significant distortion.

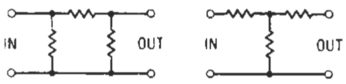

Attenuators, 1.

attenuator tube—A gas-filled, radio-frequency switching tube in which a gas discharge, initiated and regulated independently from the radio-frequency power, is used to control this rf power by reflection or absorption.

atto-—Prefix meaning 10^{-18}. Letter symbol is a.

attraction—The force that exists between two unlike magnetic poles (N and S), or between two unlike static charges (+ and −), or between two masses.

attribute—The manner in which a variable is handled by a computer.

audibility—1. The ability to be heard, usually construed as being heard by the human ear. 2. The ratio of the strength of a specific sound to the strength of a sound that can barely be heard. Usually expressed in decibels.

audibility threshold—The minimum sound intensity that the average human ear can hear, normally considered to be 0 dB sound pressure level at 1000 Hz.

audible—Capable of being heard; in most contexts, by the average human ear.

audible ringing tone—That tone received by the calling telephone, indicating that the called telephone is being rung.

audible tones—Sounds composed of frequencies that the average human can detect.

audio—Pertaining to frequencies corresponding to a normally audible sound wave. These frequencies range roughly from 15 to 20,000 hertz.

audio amplifier—*See* audio-frequency amplifier.

audio band—The range of audio frequencies passed by an amplifier, receiver, transmitter, etc. *See also* audio frequency.

audio carrier—In the NTSC system, a carrier located 4.5 MHz above the video carrier; it extends to within 0.25 MHz of the assigned channel frequency. For channel 2, its active frequency is 59.75 MHz.

audio-channel wire—A small-diameter wire, shielded and jacketed, used primarily in radio and television for wiring consoles, panels, etc.

audio component—That portion of any wave or signal which contains frequencies in the audible range (between 15 and 20,000 hertz).

audio frequency—Abbreviated AF. Any frequency corresponding to a normally audible sound wave. Audio frequencies range roughly from 15 to 20,000 hertz.

audio-frequency amplification—An increase in voltage, current, or power of a signal at an audio frequency.

audio-frequency amplifier—Also called audio amplifier. A device that contains one or more electron tubes or transistors (or both) and is designed to amplify signals within a frequency range of about 15 to 20,000 hertz.

audio-frequency choke—An inductance used to impede the flow of audio-frequency currents.

audio-frequency noise—In the audio-frequency range, any electrical disturbance introduced from a source extraneous to the signal.

audio-frequency oscillator—An oscillator circuit using an electron tube, transistor, or other nonrotating device capable of producing audio signals.

audio-frequency peak limiter—A circuit generally used in the audio system of a radio transmitter to prevent overmodulation. It keeps the signal amplitude from exceeding a predetermined value.

audio-frequency shift keying—*See* AFSK.

audio-frequency shift modulator—A system of facsimile transmission over radio, in which the frequency shift required is applied through an 800-hertz shift of an audio signal rather than by shifting the transmitter frequency. The radio signal is modulated by the shifting audio signal, usually at 1500 to 2300 hertz.

audio-frequency spectrum—The full range of sounds we can hear. In a person with good hearing, the range of sounds in the audio spectrum is usually between 15 and 20,000 Hz. In older people it is usually 50 to 10,000 Hz.

audio-frequency transformer—Also called audio transformer. An iron-core transformer for use with audio-frequency currents to transfer signals from one circuit to another. Used for impedance matching or to permit maximum transfer of power.

audiogram—Also called threshold audiogram. A graph showing hearing loss, percentage of hearing loss, or percentage of hearing as a function of frequency.

audio-level meter—An instrument that measures audio-frequency power with reference to a predetermined level. Usually calibrated in decibels.

audiometer—An electronic instrument for measuring hearing acuity. In simple units, the listener is provided (usually through earphones) with an audio signal (commonly a pure tone) of known intensity and frequency. More complex instruments provide a variety of signals (pure tones, white noise, and speech) through a variety of output transducers (such as earphones, bone vibrators, or loudspeakers).

audio mixer—An amplifier circuit used for blending two or more audio signals, such as those delivered by microphones and record players.

audion—A three-electrode vacuum tube introduced by Dr. Lee de Forest.

audio oscillator—*See* audio-frequency oscillator.

audio output—The output signal from any audio equipment. It is generally measured in volts or watts, rms.

audio patch bay—Specific patch panels provided for termination of all audio circuits and equipment used in the channel and technical-control facility. This equipment can also be found in transmitting and receiving stations.

audio peak limiter—*See* audio-frequency peak limiter.

audiophile—A person who is interested in good musical reproduction for his or her own personal listening and who uses the latest audio equipment and techniques.

audio rectification—The phenomenon of rf signals being picked up, rectified, and amplified by audio circuits—notably by high-gain preamplifiers.

audio response—1. A computer output technique that is formatted from stored words previously recorded in the computer. Programmed instructions often communicate with a student through this technique. 2. The fidelity with which audio-frequency equipment reproduces its input signal.

audio signal—An electrical signal whose frequency is within the audio range.

audio spectrum—1. The continuous range of audio frequencies, extending from the lowest to the highest (from about 15 to 20,000 hertz). 2. The range of frequencies that can be detected as sound by the human ear.

audio subcarrier—Subcarriers of satellite video signals that are modulated by audio signals. The frequency range can be from 5 to 8 MHz but is usually 6.2 or 6.8 MHz.

audio taper—Semilogarithmic change of resistance. Used on tone controls in audio amplifiers to compensate for the lower sensitivity of the human ear when listening to low-volume sounds.

audio transformer—*See* audio-frequency transformer.

audiovisual—Involving both sight and sound (e.g., audiovisual education uses films, slides, phonograph records, and the like to supplement instruction).

audiovisual system—A system of communications that simultaneously transmits pictorial and audio signals.

auditing—Examination of software for consistency and traceability.

augend—In arithmetic addition, the number increased by having another number (called the addend) added to it.

augmented operation code—In a computer, an operation code that is further defined by information contained in another portion of an instruction.

aural—Pertaining to the ear or to the sense of hearing. This term is often used to distinguish between sound that is actually heard and sound represented by audio-frequency currents. *See also* audio.

aural radio range—A radio range whose courses are normally followed by interpretation of an aural signal.

aural signal—The signal corresponding to the sound portion of a television program. In general, the audible component of a signal.

aural transmitter—The equipment used to transmit the aural (sound) signals from a television broadcast station.

aurora—Sheets, streamers, or streaks of pale light often seen in the skies of the northern and southern hemispheres. The aurora borealis and aurora australis.

auroral absorption—Absorption of radio waves due to auroral activity. *See also* aurora.

auroral absorption index—A factor that relates the average auroral absorption with the geographic location of the points of reflection from the ionosphere.

authentication—Security measure designed to protect a communications system against fraudulent transmissions and establish the authenticity of a message by an authenticator within the transmission derived from certain predetermined elements of the message itself.

authenticator—A group of letters or numerals, or both, inserted at a predetermined point in a transmission or message for the purpose of attesting to the authenticity of the message or transmission.

authoring—The process of planning, designing, and producing multimedia applications.

authorized access switch—A device used to make an alarm system or some portion or zone of a system inoperative in order to permit authorized access through a protected port. A shunt is an example of such a device.

authorized carrier frequency—A specific carrier frequency authorized for use, from which the actual carrier frequency is permitted to deviate, solely because of frequency instability, by an amount not to exceed the frequency tolerance.

autoalarm—Also called automatic alarm receiver. 1. A device that is tuned to the international distress frequency of 500 kHz and that automatically actuates an alarm if any signal is received. 2. A circuit or device operated from a radio receiver to alert a radio operator that an incoming message is addressed to him or her. 3. Complete receiving, selecting, and warning device capable of being actuated automatically by intercepted radio-frequency signals forming the international automatic-alarm signal.

auto balance—A system for detecting errors in color balance in the white and black areas of the picture and automatically adjusting the white and black levels of both the red and blue signals as needed for correction.

auto call—An alerting device that sounds a preset code of signals in a building to page those persons whose code is being sounded.

autocondensation—A method of introducing high-frequency alternating current into living tissue for therapeutic purposes. The patient is connected as one plate of a capacitor to which the current is applied.

autoconduction—A method of introducing high-frequency alternating currents into living tissues for therapeutic purposes. The patient is placed inside a coil and acts essentially as the secondary of a transformer.

autocorrelation—1. The correlation of a waveform with itself. It gives the Fourier transform of the power spectrum of the waveform (the power-density spectrum in the case of random signals). 2. A mathematical technique to measure the degree of rhythmic activity in physical phenomena that vary in a complex manner as a function of time.

autocorrelation function—A measure of the similarity between time-delayed and undelayed versions of the same signal, expressed as a function of delay.

autocorrelator—A circuit that distinguishes coherent programs (music or speech) from random noise (hiss) and operates filters that attenuate noise without an audible loss of program frequencies.

autodyne circuit—A vacuum-tube circuit that serves simultaneously as an oscillator and as a heterodyne detector.

autodyne reception—A type of radio reception employed in regenerative receivers for the reception of cw code signals. In this system the incoming signal beats with the signal from an oscillating detector to produce an audible beat frequency.

AUTOEXEC.BAT—Abbreviation for automatic execute batch file. The main file that tells a computer what to do when it is turned on. It literally automatically executes a batch of files.

auto iris control—An accessory unit that measures the video level of a TV camera and opens and closes the iris of the lens to compensate for light changes.

auto light range—The range of light (e.g., sunlight to moonlight) over which a TV camera is capable of automatically operating at specified output.

autoluminescence—The luminescence of a substance that is produced by energy within it (e.g., radioactive material).

auto-man—A locking switch that controls the method of operation (i.e., automatic or manual).

automata—A plural form of automaton.

automated communications—Combination of techniques and facilities by which intelligence is conveyed from one point to another without human effort.

automatic—1. Self-regulating or self-acting; capable of producing a desired response to certain predetermined conditions. 2. Self-acting and self-regulating; operating without human intervention; often implying the presence of a feedback control system. 3. Pertaining to a process or device that, under specific conditions, performs its functions without intervention by a human operator.

automatic-alarm receiver—Complete receiving, selecting, and warning device capable of being actuated automatically by intercepted radio-frequency signals forming the international automatic-alarm signal.

automatic-alarm-signal keying device—A device that automatically keys the radio-telegraph transmitter on board a vessel to transmit the international automatic-alarm signal.

automatic answer—A feature by which a communicating word processor, twx, etc., may receive text without an operator in attendance.

automatic back bias—A radar-receiver technique that consists of one or more automatic gain-control loops to prevent large signals from overloading a receiver, whether by jamming or by actual echoes.

automatic bass compensation—Abbreviated abc. 1. A circuit used in a receiver or audio amplifier to make the bass notes sound more natural at low-volume settings. The circuit, which usually consists of resistors and capacitors connected to taps on the volume control, automatically compensates for the poor response of the human ear to weak sounds. 2. A circuit used in some audio equipment to increase the amplitude of the bass notes to make them appear more natural at low volume settings.

automatic bias—*See* self-bias.

automatic brightness control—A circuit used in television receivers to keep the average brightness of the reproduced image essentially constant. Its action is similar to that of an automatic volume-control circuit.

automatic call—A communications feature that allows a transmission control unit to automatically establish a connection with one or more message recipients.

automatic call distributor—*See* ACD.

automatic calling unit—Abbreviated ACU. A dialing device, supplied by the communication common carrier, that permits a business machine to dial calls automatically over the communication networks.

automatic carriage—A control mechanism for a typewriter or other listing device that can automatically control the feeding, spacing, skipping, and ejecting of paper or preprinted forms.

automatic check—An operation performed by equipment built into an electronic computer to automatically verify proper operation.

automatic chrominance control—A color-television circuit that automatically controls the gain of the chrominance bandpass amplifier by varying the bias.

automatic circuit breaker—A device that automatically opens a circuit, usually by electromagnetic means, when the current exceeds a safe value. Unlike a fuse, which must be replaced once it blows, the circuit breaker can be reset manually when the current is again within safe limits.

automatic coding—A technique by which a digital computer is programmed to perform a significant portion of the coding of a problem.

automatic color purifier—*See* automatic degausser.

automatic computer—A computer capable of processing a specified volume of work without a need for human intervention other than program changes.

automatic connections—Connections between users made by electronic switching equipment without human intervention.

automatic constant—In a calculator, a provision that allows the user to multiply or divide a series of numbers by the same divider or multiplier without reentering each time.

automatic contrast control—A television circuit that automatically changes the gain of the video intermediate-frequency and radio-frequency stages to maintain proper contrast in the television picture.

automatic controller—A device or instrument for measuring and regulating that operates by receiving a signal from a sensing device, comparing this signal with a desired value, and issuing signals for corrective action.

automatic crossover—1. A type of current-limiting circuit on a power supply provided with an adjustment for setting the short-circuit current to an adjustable maximum value. 2. A term applied to bimodal power supplies (constant voltage/constant current) that describes the transferral from one operating mode to the other at a predetermined value of load resistance. Usually the crossover point is preset by means of front panel controls.

automatic current limiting—An overload-protection mechanism designed to limit the maximum output current of a power supply to a preset value. Usually it automatically restores the output when the overload is removed.

automatic cutout—A device, operated by electromagnetism or centrifugal force, to automatically disconnect some parts of an equipment after a predetermined operating limit has been reached.

automatic data processing—Abbreviated ADP. The processing of digital information by automatic computers and other machines. Also called integrated data processing.

automatic data-processing system—A system that includes electronic data-processing equipment together with auxiliary and connecting communications equipment.

automatic degausser—Also called automatic color purifier and degausser. An arrangement of degaussing coils mounted around a color-television picture tube. These coils are energized only for a short while after the set is turned on. They serve to demagnetize any parts of the picture tube that may have been affected by the earth's magnetic field or the magnetic field of any nearby home appliance.

automatic dial—A communications feature whereby a calling unit has the ability to automatically establish a connection with one or more message recipients.

automatic dialer—1. Device that will automatically dial any of a group of preselected telephone numbers. 2. A device that automatically dials telephone numbers on a network.

automatic dialing unit—Abbreviated ADU. A device capable of generating dialing digits automatically.

automatic digital network—Automatic communications network for end-to-end message-switched digital data communication.

automatic direction finder—Abbreviated ADF. Also called an automatic radio compass. An electronic device, usually for marine or aviation application, which provides a radio bearing to any transmitter whose frequency is known but whose direction and location are not.

automatic electronic data-switching center—Communications center designed specifically for the automatic electronic transmission, reception, relay, and switching of digitalized data.

automatic error correction—A technique, usually requiring the use of special codes and/or automatic retransmission, that detects and corrects errors occurring in transmission. The degree of correction depends on coding and equipment configuration.

automatic exchange—A telephone exchange in which connections are made between subscribers by means of devices set in operation by the originating subscriber's instrument without the intervention of an operator.

automatic fine tuning—A circuit in a receiver that automatically maintains the correct tuner oscillator frequency and compensates for drift and for moderate amounts of inaccurate tuning. Similar to automatic frequency control.

automatic focusing—A method of electrostatically focusing a television picture tube; the focusing anode is internally connected through a resistor to the cathode and thus requires no external focusing voltage.

automatic frequency control—Abbreviated AFC. 1. A system that produces an error voltage in proportion to the amount by which an oscillator drifts away from its correct frequency, the error voltage acting to reverse the drift. 2. A control circuit in a receiver or tuner that compensates for small variations in the carrier signal frequency to provide a stable audio output. A circuit function in a tuner or receiver that keeps the unit accurately tuned to the desired station, eliminating any tendency to drift.

automatic frequency correction—*See* automatic frequency control.

automatic function key correction—When the wrong function key is depressed in a calculator, pressing the correct function key automatically replaces it.

automatic gain control—1. A means of controlling gain of a receiver through feedback to suit the strength of the incoming signal. Ideally, the output level at the speaker should remain constant over a wide range of input signals. For low-level inputs the feedback signal is small and gain is high. For stronger signals the feedback loop cuts gain to prevent overload. 2. A self-acting compensation device that maintains the output of a transmission system constant within narrow limits in the face of wide variations in the attenuation of the system. 3. A radar circuit that prevents saturation of the radar receiver by long blocks of received signals or by a carrier modulated at low frequency. 4. A process by which gain is automatically adjusted as a function of input or other specified parameter.

automatic gain stabilization—A circuit, used in certain identification friend-or-foe equipment and radar beacon systems, which serves to maintain optimum sensitivity in a superregenerative stage by keeping the noise-pulse load constant. The system prevents random noises from triggering the automatic transmitter associated with the receiver.

automatic grid bias—Grid-bias voltage provided by the difference in potential across a resistance (or resistances) in the grid or cathode circuit due to grid or cathode current or both.

automatic intercept—Automatic recording of messages a caller may wish to leave when the called party is away from his telephone.

automatic level compensation—A system that automatically compensates for variations in the circuit. *See also* automatic volume control.

automatic level control—*See* ALC.

automatic light control—The process by which the illumination incident on the face of a television pickup device is automatically adjusted as a function of scene brightness.

automatic message accounting—*See* automatic toll ticketing.

automatic message-switching center—A center in which messages are routed automatically according to information they contain.

automatic modulation control—A transmitter circuit that reduces the gain for excessively strong audio input signals without affecting the strength of normal signals, thus permitting higher average modulation without overmodulation; this is equivalent to an increase in the carrier-frequency power output.

automatic noise-limiter—A circuit that automatically clips off all noise peaks above the highest peak of the desired signal being received. This circuit prevents strong atmospheric or human-made interference from being troublesome.

automatic numbering equipment—Equipment used in association with tape transmitters to transmit a channel number.

automatic pedestal control—The process by which the pedestal height of a television signal is automatically adjusted as a function of the input or other specified parameter.

automatic phase control—A circuit used in color television receivers to synchronize the burst signal with the 3.58-MHz color oscillator.

automatic pilot—*See* autopilot.

automatic programming—Any technique designed to simplify the writing and execution of programs in a computer. Examples are assembly programs that translate from the programmer's symbolic language to the machine language, those which assign absolute addresses to instruction and data words, and those which integrate subroutines into the main routine.

automatic quality control—A technique in which the quality of a product being processed is evaluated in terms of a predetermined standard, and proper corrective action is taken automatically if the quality falls below the standard.

automatic radio compass—A radio direction finder that automatically rotates the loop antenna to the correct position. A bearing can then be secured from the indicator dials without mechanical adjustment or calculation. *See also* automatic direction finder.

automatic record changer—An electrically operated mechanism that automatically feeds, plays, and rejects a number of records in a preset sequence. It consists of a motor, turntable, pickup arm, and changer. Some changers are designed to play automatically 16⅔-, 33⅓-, 45-, and 78-rpm records.

automatic regulation—In a power supply, the automatic holding of the output voltage or current to a constant value in spite of variations in the input voltage of load resistance.

automatic relay—A means of selective switching that causes automatic equipment to record and retransmit communications.

automatic repeater station—A station that receives signals and simultaneously retransmits them, but usually on a different frequency.

automatic repeat request—Abbreviated arr. A system of error checking in which an error-detecting code is included with the transmitted data. When the receiver is unable to verify the message, it initiates a retransmission.

automatic reset—The automatic reversion of a timer to its ready state after it has completed a timing cycle or after the input circuit has been interrupted.

automatic reset relay—Also called automatic reset. 1. A stepping relay that returns to its home position either when it reaches a predetermined contact position or when a pulsing circuit fails to energize the driving coil within a given time. It may either pulse forward or be spring-reset to the home position. 2. An overload relay that restores the circuit as soon as an overcurrent situation is corrected.

automatic retransmission—Retransmission of signals by a radio station whereby the retransmitting station is actuated solely by the presence of a received signal through electrical or electromechanical means, i.e., without any direct, positive action by the control operator.

automatic reverse—Ability of some four-track stereo tape recorders to play the second pair of stereo tracks automatically, in the reverse direction, without need to interchange the empty and full reels after the first pair of stereo tracks has been played.

automatic scanning—A variable-speed sweep of the entire frequency range of a radio-frequency interference meter. It may also include the scanning of a portion of this frequency range over a predetermined sector.

automatic scanning receiver—A receiver that can automatically and continuously sweep across a preselected frequency, either to stop when a signal is found or to plot signal occupancy within the frequency spectrum being swept.

automatic secure voice communications—A network that provides cryptographically secure voice communications through the use of a combination of wideband and narrow-band voice-digitizing techniques.

automatic selective control (or transfer) relay—A device that operates to select automatically between certain sources or conditions in an equipment, or performs a transfer operation automatically.

automatic send—A communications capability whereby a communicating system has the ability to automatically send out a message in an unattended mode.

automatic send/receive—A teletype-writer unit that includes a keyboard, printer, paper tape, reader/transmitter, and paper-tape punch. This combination of facilities may be used online or offline, and, in some cases, online and offline simultaneously.

automatic sensitivity control—1. A circuit used for automatically maintaining receiver sensitivity at a predetermined level. It is similar to an automatic gain control, but it affects the receiver constantly rather than during the brief interval selected by the range gate. 2. The self-acting mechanism that varies the system sensitivity as a function of the specified control parameters. This may include automatic target control, automatic light control, or any combination thereof.

automatic sequencing—The ability of a computer to perform successive operations without additional instructions from a human being.

automatic short-circuiter—A device used in some forms of single-phase commutator motors to short-circuit the commutator bars automatically.

automatic short-circuit protection—An automatic current-limiting system that enables a power supply to continue operating at a limited current, and without damage, into any output overload, including a short circuit. The output voltage is restored to normal when the overload is removed, as distinguished from a fuse or circuit-breaker system that opens with overload and must be replaced or reclosed manually to restore power.

automatic shutoff—In a tape recorder, a switching arrangement that automatically shuts the recorder off when the tape breaks or runs out. Also, a switching arrangement that stops a record changer after the last record.

automatic starter—1. A device which, after being given the initial impulse by means of a push button or similar device, starts a system or motor automatically in the proper sequence. 2. A self-acting starter that is completely controlled by master or pilot switches or some other sensing device.

automatic switchboard—Telephone switchboard in which the connections are made by the operation of remotely controlled switches.

automatic switch center—A switch center in which messages originating at any subscriber terminal are relayed automatically through one or more switching centers to their destinations.

automatic target control—The self-acting mechanism that controls the vidicon target potential as a function of the scene brightness.

automatic telegraph transmission—A form of telegraphy in which signals are transmitted mechanically from a perforated tape.

automatic telegraphy—A form of telegraphy in which signals are transmitted and/or received automatically.

automatic telephone dialer—A device that, when activated, automatically dials one or more preprogrammed telephone numbers (e.g., police, fire department) and relays a recorded voice or coded message giving the location and nature of the alarm; used with intrusion alarms and security systems.

automatic test generation—Abbreviated ATG. Automatic test-pattern generation (ATPG). Calculation of a specific set of input test patterns with a computer program providing algorithmic and heuristic routines.

automatic threshold variation—A constant false-alarm rate scheme that is an open-loop type of automatic gain control in which the decision threshold is varied continuously in proportion to the incoming intermediate frequency and video noise level.

automatic time switch—A combination of a switch with an electric or spring-wound clock arranged to turn an apparatus on and off at predetermined times.

automatic toll ticketing—System whereby toll calls are automatically recorded, timed, and toll tickets printed, under control of the calling telephone's dial pulses and without the intervention of an operator. Also called automatic message accounting.

automatic tracking—In radar, the process whereby a mechanism, actuated by the echo, automatically keeps the radar beam locked on the target; may also determine the range simultaneously.

automatic transfer equipment—Equipment that automatically transfers a load so that a source of power may be selected from one of several incoming lines.

automatic tuning—An electrical, mechanical, or electrical/mechanical system that automatically tunes a circuit to a predetermined frequency when a button or other control is operated.

automatic turntable—A record player whose tone arm is positioned for playing records when a control is operated, and which shuts off automatically at the end of play. *See* record changer.

automatic video-noise leveling—A constant false-alarm rate scheme in which the video-noise level at the output of the receiver is sampled at the end of each range sweep and the receiver gain is readjusted accordingly to maintain a constant video-noise level at the output.

automatic voltage regulator—A device or circuit that maintains a constant voltage, regardless of any variation in input voltage or load.

automatic volume compression—*See* volume compression.

automatic volume control—Abbreviated AVC. 1. A self-acting compensation device that maintains the output of a transmission system constant within narrow limits in the face of wide variations in attenuation in that system. 2. A self-acting device that maintains the output of a radio receiver or amplifier substantially constant within relatively narrow limits while the input voltage varies over a wide range. 3. *See* automatic level compensation.

automatic volume expansion—Also called volume expansion. An audio-frequency circuit that automatically increases the volume range by making loud portions louder and weak ones weaker. This is done to make radio reception sound more like the actual program, because the volume range of programs is generally compressed at the point of broadcast.

automatic zero and full-scale calibration correction—A system of zero and sensitivity stabilization in which electronic servos are used that compare demodulated "zero" and "full-scale" signals with reference voltages.

automation—1. The method or act of making a manufacturing or processing system partially or fully automatic. 2. The entire field of investigation, design, development, application, and methods of rendering or making processes or machines self-acting or self-moving; rendering automatic. 3. Automatically controlled operation of an apparatus process or system by mechanical or electronic devices that take the place of observation, effort, and decision by a human operator.

automaton—1. A device that automatically follows predetermined operations or responds to encoded instructions. 2. Any communication-linked set of elements. 3. A machine that exhibits living properties. 4. A mechanism, fixed or mobile, possessing the ability to manipulate objects external to itself under the constant control of a programming routine previously supplied by an external intelligence. 5. A machine that is designed to simulate the operations of living things.

automonitor—1. To instruct an automatic digital computer to produce a record of its information-handling operations. 2. A program or routine for this purpose.

automotive analyzer—An instrument containing numerous automotive test features combined into one portable unit.

automotive electronics—The branch of engineering science that deals with the generation, control, conversion, and application of electricity in self-propelled vehicles.

automotive primary wire—Low-voltage single or multiconductor wire for automotive applications. Resistant to oil and weather.

autopatch—1. A repeater-to-telephone connection that makes it possible to place a telephone call from a mobile transmitter that has DTMF capability. An access code is frequently required to establish the connection to the phone line. 2. A remotely controllable device frequently used at a repeater location that patches a radiocommunications system into a telephone land-line network. 3. A component that allows telephone calls to be placed through a repeater.

autopilot—Also called automatic pilot, gyropilot, or robot pilot. A device containing amplifiers, gyroscopes, and servomotors that automatically control and guide the flight of an aircraft or guided missile. The autopilot detects any deviation from the planned flight and automatically applies the necessary corrections to keep the aircraft or missile on course.

autopolarity—A feature of a digital voltmeter or digital multimeter wherein the correct polarity (either negative or positive) for a measured quantity is automatically indicated on the display.

auto radio—A radio receiver designed to be installed in an automobile and powered by the storage battery of the automobile.

autoregulation induction heater—An induction heater in which a desired control is effected by the change in characteristics of a magnetic charge as it is heated at or near its Curie point.

auto reverse (cassette deck)—Automatically reverses tape direction at the end of each side of an audiocassette, for extended unattended listening.

auto single play—A turntable in which records are played one at a time and the arm performs a playing cycle, lifting itself from the record at the end (and in some cases shutting off the motor).

autostarter—*See* autotransformer starter.

Autosyn—A trade name of the Bendix Corp. for a remote-indicating instrument or system based on the synchronous-motor principle, in which the angular position of the rotor of a motor at the measuring source is duplicated by the rotor of the indicator motor.

auto tracking—Also called automatic tracking operation. A master/slave connection of two or more power supplies, each of which has one of its output terminals in common with one of the output terminals of all of the other supplies, such connections being characterized by one-knob control and proportional output voltage from all supplies. Useful where simultaneous turn-up, turn-down, or proportional control of all power supplies in a system is required.

autotransformer—1. A transformer with a single winding (electrically) in which the whole winding acts as the primary winding, and only part of the winding acts as the secondary (step-down); or part of the winding acts as the primary, and the whole winding acts as the secondary (step-up). 2. A voltage, current, or impedance transforming device in which parts of one winding are common to both the primary and secondary circuits.

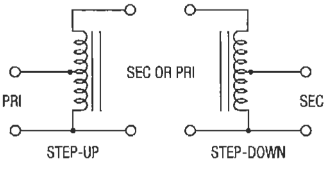

Autotransformers, 1.

autotransformer starter—Also called a compensator or autostarter. A motor starter having an autotransformer to furnish reduced voltage for starting. Includes the necessary switching mechanism.

autozero—Displaying all zeros in a digital multimeter or digital voltmeter when no measurement is being made.

aux—Abbreviation for auxiliary. Often applied to amplifier inputs, where it refers to an extra input facility as distinct from "mic," "tuner," "pickup," etc.

auxiliary actuator—A mechanism that may be attached to a switch to modify its characteristics.

auxiliary bass radiator—A parasitic (nonelectrically driven) unit resembling a bass speaker unit located in a loudspeaker enclosure as if it were an ordinary unit, to increase the movement of air and hence enhance the bass performance at a given enclosure volume.

auxiliary circuit—Any circuit other than the main circuit.

auxiliary contacts—In a switching device, contacts, in addition to main circuit contacts, that function with the movement of the latter.

auxiliary electrodes—Metallic electrodes pushed or driven into the earth to provide electrical contact for the purpose of performing measurements on grounding electrodes of grounded grid systems.

auxiliary equipment—Equipment not directly controlled by the central processing unit of a computer.

auxiliary function—In automatic control of machine tools, a machine function other than the control of the motion of a workpiece or cutter, e.g., control of machine lubrication and cooling.

auxiliary-link station—A station, other than a repeater station, at a specific land location, licensed only for the purpose of automatically relaying radio signals from that location to another specific land location.

auxiliary memory—*See* auxiliary storage.

auxiliary operation—An operation performed by equipment not under continuous control of the computer central processing unit.

auxiliary relay—1. A relay that responds to the opening or closing of its operating circuit to assist another relay or device in the performance of a function. 2. A relay actuated by another relay and used to control secondary circuit functions such as signals, lights, or other devices. Also called slave relay.

auxiliary-station line filter—A line filter for use at repeater points to separate frequencies of different carrier systems using the same line pair. For example, such a filter might be used at a high-frequency carrier-system repeater to bypass the low-frequency carrier system and voice frequencies around the repeater.

auxiliary storage—1. In a computer, storage that supplements the main storage, such as hard disks, floppy disks, magnetic tapes, or optical discs. 2. Storage capacity, such as magnetic tape, disk, or drum, in addition to the main memory of a computer. Also called auxiliary memory.

auxiliary switch—A switch actuated by some device such as a circuit breaker, for signaling, interlocking, or other purpose.

auxiliary transmitter—A transmitter held in readiness in case the main transmitter of a broadcasting station fails.

availability—1. The ratio, expressed as a percent, of the time during a given period that an equipment is correctly operating to the total time in that period. Also called operating ratio. 2. The probability that a system is operating satisfactorily at any point in time when used under stated conditions, where the total time considered includes operating time, active repair time, administrative time, and logistic time.

available conversion gain—Ratio of available output-frequency power from the output terminals of a transducer to the available input-frequency power from the driving generator, with terminating conditions specified for all frequencies that may affect the result. Applies to outputs of such magnitude that the conversion transducer is operating in a substantially linear condition.

available gain—The ratio of the available power at the output terminals of the network to the available power at the input terminals of the network.

available line—In a facsimile system, that portion of a scanning line that can be used for picture signals. Expressed as a percentage of the length of the scanning line.

available machine time—Time after the application of power during which a computer is operating correctly.

available power—1. The mean square of the open-circuit terminal voltage of a linear source, divided by four times the resistive component of the source impedance. 2. Of a network, the power that would be delivered to a conjugately matched load. It is the maximum power that a network can deliver. Available power, though defined in terms of an output load impedance, is independent of that impedance.

available power gain—Sometime called completely matched power gain. Ratio of the available power from the output terminals of a linear transducer, under specified input-termination conditions, to the available power from the driving generator. The available power gain of an electrical transducer is maximum when the input-termination admittance is the conjugate of the driving-point admittance at the input terminals of the transducer.

available signal-to-noise ratio—Ratio of the available signal power at a point in a circuit to the available random-noise power.

avalanche—1. Rapid generation of a current flow with reverse-bias conditions as electrons sweep across a semiconductor junction with enough energy to ionize other bonds and create electron-hole pairs, making the action regenerative. 2. One of the mechanisms responsible for voltage breakdown of semiconductor junctions and devices. When avalanche occurs, carriers moving through the crystal lattice have achieved sufficient kinetic energy to knock further carriers from the lattice, producing a snowballing increase in current level. Provided the current increase is limited externally, avalanche breakdown causes no permanent damage to the device.

avalanche breakdown—1. In a semiconductor diode, a nondestructive breakdown caused by the cumulative multiplication of carriers through field-induced impact ionization. 2. In a reverse-biased semiconductor, the sudden, marked increase of reverse current at the bias voltage at which avalanche begins. The action resembles a breakdown but it is nondestructive when the current is limited by external means. 3. A breakdown that is caused by the action of a strong electric field which causes some free carriers to gain enough energy to liberate new hole-electron pairs by ionization.

avalanche conduction—A form of conduction in a semiconductor in which charged-particle collisions create additional hole-electron pairs.

avalanche current—The high current through a semiconductor junction in response to an avalanche voltage.

avalanche diode—1. Also called breakdown diode. A silicon diode that has a high ratio of reverse-to-forward resistance until avalanche breakdown occurs. After breakdown the voltage drop across the diode is essentially constant and is independent of the current. Used for voltage regulating and voltage limiting. Originally called zener diode, before it was found that the zener effect had no significant role in the operation of diodes of this type. 2. A silicon diode in which avalanche breakdown occurs across the diode's pn junction. *See* IMPATT diode; TRAPATT diode.

avalanche impedance—*See* breakdown impedance.

avalanche noise—1. A phenomenon in a semiconductor junction in which carriers in a high-voltage gradient develop sufficient energy to dislodge additional carriers through physical impact. 2. Electrical noise generated in a junction diode operated at the point at which avalanche just begins.

avalanche photodiode— 1. A photodiode that takes advantage of the avalanche multiplication of photocurrent. It is particularly suited to low-noise and/or high-speed applications. 2. A device that utilizes avalanche multiplication of photocurrent by means of hole electrons created by absorbed photons. When the device's reverse-bias voltage nears breakdown level, the hole-electron pairs collide with substrate atoms to produce multiple hole-electron pairs.

avalanche transistor—A transistor that, when operated at a high reverse-bias voltage, supplies a chain generation of electron-hole pairs.

avalanche voltage—The applied voltage at which avalanche breakdown occurs.

avalanching—A process resulting from high fields in a semiconductor device, in which an electron is accelerated by the field, hits an atom, and releases more electrons, which continue the sequence.

AVC—*See* automatic volume control.

average—*See* arithmetic mean.

average absolute pulse amplitude—The average of the absolute value of instantaneous amplitude taken over the pulse duration. Absolute value means the arithmetic value regardless of algebraic sign.

average brightness—The average illumination in a television picture.

average calculation operation—A typical computer calculating operation longer than an addition and shorter than a multiplication, often taken as the mean of nine additions and one multiplication.

average current—The arithmetic mean of the instantaneous currents of a complex wave, averaged over one half cycle.

average electrode current—The value obtained by integrating the instantaneous electrode current over an averaging time and dividing by the average time.

average life—*See* mean life, 1.

average noise factor—*See* average noise figure.

average noise figure—Also called average noise factor. In a transducer, the ratio of total output noise power to the portion attributable to thermal noise in the input termination, with the total noise being summed over frequencies from zero to infinity and the noise temperature of the input termination being standard (290 K).

average outgoing quality—The ultimate average quality of products shipped to the customer that results from composite sampling and screening techniques.

average power output of an amplitude-modulated transmitter—The radio-frequency power delivered to the transmitter output terminals, averaged over a modulation cycle.

average pulse amplitude—The average of the instantaneous amplitudes taken over the pulse duration.

average rate of transmission—Effective speed of transmission.

average value—1. The value obtained by dividing the sum of a number of quantities by the number of quantities. The average value of a sine wave is 0.637 times the peak value. 2. The dc voltage of current amplitude that will transfer the same electrical charge to a capacitor as the ac waveform during a half period. Mathematically, it is the average of the absolute value of all the instantaneous amplitudes.

average voltage—The sum of the instantaneous voltages in a half-cycle waveshape, divided by the number of instantaneous voltages. In a sine wave, the average voltage is equal to 0.637 times the peak voltage.

aviation channels—A band of frequencies, below and above the standard broadcast band, assigned exclusively for aircraft and aviation applications.

aviation services—The aeronautical mobile and radionavigational services.

avionics—1. An acronym designating the field of *avi*ation electr*onics*. 2. The branch of electronics that is concerned with aviation applications. 3. The design, production, and application of electronic devices and systems for use in aviation and astronautics.

Avogadro's number—The actual number of molecules in one gram-molecule, or of atoms in one gram-atom, of an element or any pure substance (6.023×10^{23} molecules/mole).

AWG—1. American wire gage. A means of specifying wire diameter. The higher the number, the smaller the diameter. 2. A scale of gage sizes which, with the exception of the largest sizes, 4/0 through 1/0, increases with the descending values of wire diameter. For example, a 1 AWG wire has a diameter of 0.289 inch (7.34 mm) and a wire of 40 AWG has a diameter of 0.0031 inch (78.74 μm). AWG is applied to stranded as well as solid conductors.

axial leads—Leads coming out the ends and along the axes of a resistor, capacitor, or other axial part, rather than out the side.

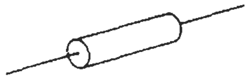

Axial leads.

axial ratio—Ratio of the major axis to the minor axis of the polarization ellipse of a waveguide. This term is preferred over *ellipticity* because, mathematically, ellipticity is 1 minus the reciprocal of the axial ratio.

axis—The straight line, either real or imaginary, passing through a body around which the body revolves or around which parts of a body are symmetrically arranged.

Ayrton-Perry winding—1. Two conductors connected in parallel so that the current flows in opposite directions in each conductor and thus neutralizes the inductance between the two. 2. A noninductive winding with two inductors that conduct current in opposite directions, with the opposing currents canceling the magnetic field.

Ayrton shunt—Also called universal shunt. A high-resistance parallel connection used to increase the range of a galvanometer without changing the damping.

azel display—A modified type of plan-position indicator presentation showing two separate radar displays on one cathode-ray screen. One display presents bearing information, and the other shows elevation.

az-el mount—An antenna mount that tracks satellites by moving in two directions: the azimuth in the horizontal plain, and elevation up from the horizon.

azimuth—1. The angular measurement in a horizontal plane and in a clockwise direction. 2. In a tape recorder, the angle that recording and playback head gaps make with the line along which the tape moves. The head is oriented until this angle is 90°. 3. The vertical setting (alignment) of the head in a tape recorder. 4. Compass direction from due north measured in degrees clockwise. (True north can be found by sighting the star Polaris at

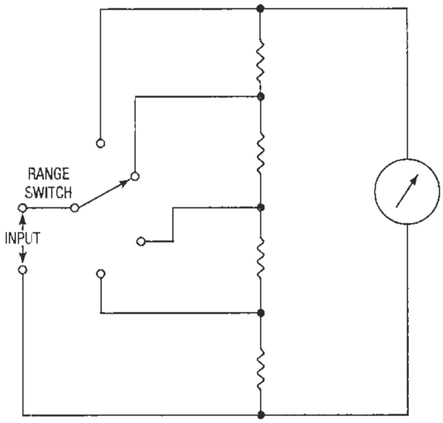

Ayrton shunt.

night, or by applying a local correction for magnetic deviation to a compass reading.)

azimuth alignment—Alignment of the recording and reproducing gaps so that their center lines lie parallel with one another. Misalignment of the gaps causes a loss in output at short wavelengths. *See* head alignment.

azimuth blanking—Blanking of the CRT screen in a radar receiver as the antenna scans a selected azimuth region.

azimuth-elevation mount—A movable dish-antenna mount and aiming system in which one pivot allows rotation in the horizontal plane about the azimuth angle from due north. The other pivot is the elevation above the horizon. (This mount can be more difficult to aim than a polar mount.)

azimuth gain reduction—A technique that allows control of the radar receiver system throughout any azimuth sectors.

azimuth loss—High-frequency losses that are caused by recording head misalignment.

azimuth rate—The rate of change of true bearing.

azimuth resolution—The angle or distance by which two targets must be separated in azimuth to be distinguished by a radar set when the targets are at the same range.

azimuth stabilization—The presentation of indications on a radar display so that north, or any specific reference line of direction, is always at the top of the screen.

azimuth-stabilized plan-position indicator—A plan-position indictator scope on which the reference bearing (usually true or magnetic north) remains fixed with respect to the indicator, regardless of the vehicle orientation.

azimuth versus amplitude—An electronic counter-countermeasures receiver with plan-position indicator-type display attached to the main antenna, used to display strobes due to jamming aircraft. It is useful in making passive position fixes when two or more radar sites can operate together.

azusa—A short base-line continuous-wave phase-comparison electronic tracking system operating on the C-band, in which a single station provides two direction cosines and slant range.

B

B—1. Symbol for the base of a transistor. 2. Symbol for magnetic flux. 3. Abbreviation for photometric brightness. 4. B or b. Abbreviation for susceptance.

B− (B-minus or B-negative)—Negative terminal of a B battery or the negative polarity of other sources of anode voltage. Denotes the terminal to which the negative side of the anode-voltage source should be connected.

B+ (B-plus or B-positive)—Positive terminal of a B battery or the positive polarity of other sources of anode voltage. The terminal to which the positive side of the anode voltage source should be connected.

babble—1. The aggregate crosstalk from a large number of disturbing channels. 2. In a carrier or other multiple-channel system, the unwanted disturbing sounds that result from the aggregate crosstalk or mutual interference from other channels. 3. Crosstalk from a large number of channels in a system.

babble signal—A type of electronic deception signal used to confuse enemy receivers. Generally, it has characteristics of energy transmission signals. It can be composed by superimposing incoming signals on previously recorded intercepted signals; this composite signal can then be radiated as a jamming signal.

BABS—Abbreviation for blind approach beacon system. A pulse-type ground-based navigation beacon used for runway approach. The BABS ground beacon is installed beyond the far end of the runway on the extended center line. When interrogated by an aircraft, it retransmits two diverging beams, one of short- and the other of long-duration pulses. The beams are transmitted alternately, but because of the first switching, the aircraft receives what appears to be a continuous transmission of both beams. The cathode-ray tube in the aircraft displays both long and short pulses superimposed on each other. When the aircraft is properly aligned with the runway, the pulses will be of equal amplitude.

back bias—1. A degenerative or regenerative voltage that is fed back to circuits before its originating point. Usually applied to a control anode of a tube. 2. A voltage applied to a grid of a tube (or tubes) to restore a condition that has been upset by some external cause. 3. *See also* reverse bias.

back bonding—Bonding active chips to a substrate using the back of the chip, leaving the face with its circuitry uppermost. The opposite of back bonding is face-down bonding.

backbone—A high-voltage, high-capacity transmission line or group of lines having a limited number of large-capacity connections between loads and points of generation.

backbone network—1. A network that links several smaller networks. 2. A transmission facility designed to interconnect low-speed distribution channels or clusters of dispersed user devices.

back contact—Relay, key, jack, or other contact designed to close a circuit and permit current to flow when, in the case of a relay, the armature has released or fallen back, or, in other cases, when the equipment is inoperative.

back current—Also called reverse current. The current that flows when reverse bias is applied to a semiconductor junction.

back diode—A tunnel diode that is usually chosen for its reverse-conduction characteristics.

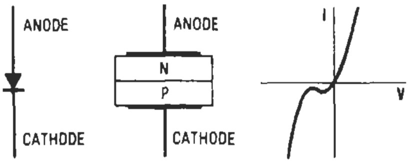

Symbol. Junction. Characteristic.

Back diode.

back echo—An echo due to the back lobe of an antenna.

back emission—Also called reverse emission. Emission from an electrode occurring only when the electrode has the opposite polarity from that required for normal conduction. A form of primary emission common to rectifiers during the inverse portion of their cycles.

back end—In semiconductor manufacturing, the package assembly and test stages of production. Includes burn-in and environmental test functions.

backfill—Filling an evacuated (hybrid) circuit package with dry inert gas prior to hermetic sealing of the package.

backfire—*See* arcback.

back focal length—Distance from the center of a lens to its principal focus on the side of the lens away from the object.

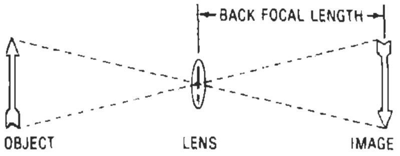

Back focal length.

back focus—The distance from the rear vertex of a lens to the focal plane, with the subject at infinite distance.

background—1. The picture white of the copy being scanned when the picture is black and white only. Also undesired printing in the recorded copy of the picture being transmitted, resulting in shading of the background area. 2. Noise heard during a radio program; this noise is caused by atmospheric interference or operation of the receiver at such high gain that inherent tube and circuit noises become noticeable. 3. A low-priority job that a computer works on when it isn't occupied by more pressing matters.

background control—In color television, a potentiometer used as a means of controlling the dc level of a color signal at one input of a tricolor picture tube. The setting of this control determines the average (or background) illumination produced by the associated color phosphor.

background count—Count caused by radiation from sources other than the one being measured.

background noise—1. The total system noise independent of the presence or absence of a signal. The signal is not included as part of the noise. 2. In a receiver, the noise in the absence of signal modulation in the carrier. 3. Any unwanted sound that intrudes upon program material, such as sounds produced as a result of surface imperfections of a disc record.

background processing—The automatic execution of lower-priority programs in a computer when the system resources are not being used for higher-priority (foreground) programs. For example, background program processing is temporarily suspended to service interrupt requests from I/O devices that require foreground processing.

background program—1. Low-priority program in a multiprogramming environment, which operates when the processor is not doing anything else. 2. That portion of the resident computer program that is run when no immediate pressing needs exist in the system.

background radiation—Radiation due to the presence of radioactive material in the vicinity of the measuring instrument.

background response—In radiation detectors, response caused by ionizing radiation from sources other than that to be measured.

backhaul—A terrestrial communications channel linking an earth station to a local switching network or population center.

back-haul—Use of excess circuit mileage by routing through switching centers that are not in a direct facility path between an originating office and a terminating office.

backing—Flexible material (usually cellulose acetate or polyester) on which is deposited the magnetic-oxide coat that "records" the signal on magnetic recording tape. Also known as base.

backing up—Making copies of files to prevent loss of their contents if the originals are damaged.

backlash—In a potentiometer, the maximum difference that occurs in shaft position when the shaft is moved to the same actual output-ratio point from opposite directions. Resolution and contact-width effects must be excluded from this measure.

backloaded horn—A speaker enclosure arranged so the sound from the front of the cone feeds directly into the room, and the sound from the rear feeds into the room via a folded horn.

back loading—A form of horn loading particularly applicable to low-frequency speakers; the rear radiating surface of the speaker feeds the horn, and the front part of the speaker is directly exposed to the room.

back lobe—1. In the radiation pattern of a directional antenna, that part which extends backward from the main lobe. 2. The three-dimensional petal-shaped pattern representing antenna directional response that is pointing away from the intended direction.

back metallization—Metal applied to the side of the transistor wafer opposite the active areas. Provides the collector contact in bipolar transistors and permits the transistor chips to be bonded to the package or thin-film circuit substrate.

back pitch—The winding pitch of the back end of the armature, that is, the end opposite the commutator.

backplane—1. Area of a computer or other equipment where various logic and control elements are interconnected. Often takes the form of a rat's nest of wires interconnecting printed-circuit cards in the back of computer racks or cabinets. 2. The physical area where printed circuit boards in a system plug in. Usually contains the buses of the system either in printed-circuit or wire-wrap form. Also called a motherboard. 3. A printed circuit card located in the back of a chassis. It has sockets into which specific modules fit for interconnection. 4. A printed circuit motherboard with connectors placed at intervals to allow connection and communication between daughterboards.

backplate—In a camera tube, the electrode to which the stored-charge image is capacitively coupled.

back porch—In a composite picture signal, that portion which lies between the trailing edge of a horizontal-sync pulse and the trailing edge of the corresponding blanking pulse. A color burst, if present, is not considered part of the back porch.

back-porch effect—The continuation of collector current in a transistor for a short time after the input signal has dropped to zero. The effect is due to storage of minority carriers in the base region. It also occurs in junction diodes.

back-porch tilt—The slope of the back porch from its normal horizontal position. Positive and negative refer, respectively, to upward and downward tilt to the right.

backscattering—1. Radiation of unwanted energy to the rear of an antenna. 2. The reflected radiation of energy from a target toward the illuminating radar.

back-shunt keying—A method of keying a transmitter, in which the radio-frequency energy is fed to the antenna when the telegraph key is closed and to an artificial load when the key is open.

backside illumination—A charge-coupled device fabrication technique employing thinned silicon, in which the image is impressed on the side opposite the MOS electrodes.

backspace—In a computer, an operation whose function is to move backward in a sequential file one record at a time.

backstop—That part of the relay that limits the movement of the armature away from the pole face or core. In some relays, a normally closed contact may serve as a backstop.

backswing—The amplitude of the first maximum excursion in the negative direction after the trailing edge of a pulse, expressed as a percentage of the 100-percent amplitude.

backtalk—Transfer of information to the active computer from a standby computer.

back-to-back circuit—Two tubes or semiconductor devices connected in parallel but in opposite directions so that they can be used to control current without introducing rectification. Also called inverse-parallel connection.

backup—1. A duplicate copy of a file or program. 2. A system, device, file or facility that can be used as an alternative in case of a malfunction or loss of data. 3. An item kept available to replace an item that fails to perform satisfactorily. 4. An item under development intended to perform the same general function performed by another item also under development. 5. The hardware and software resources available to recover after a degradation or failure of one or more system components. 6. A system or element, such as a circuit component, that is used to replace a similar system or component in case of failure of the latter.

backup control—*See* redundancy.

backup copy—A copy of a file or data set to be kept for reference in case the original file or data set is destroyed or lost.

backup facility—A communications-electronics facility that is established for the purpose of replacing or supplementing another facility or facilities under real or simulated emergency conditions.

backup item—An additional item to perform the general functions of another item. It may be secondary to an identified primary item or a parallel development to improve the probability of success in performing the general function.

backwall—The plate in a pot core that connects the center post to the sleeve.

backwall photovoltaic cell—A cell in which light must pass through the front electrode and a semiconductor layer before reaching the barrier layer.

backward-acting regulator—A transmission regulator in which the adjustment made by the regulator affects the quantity that caused the adjustment.

backward diode—A highly doped, alloyed germanium junction that operates on the principle of quantum-mechanical tunneling. The diode is "backward" because its easy current direction is in the negative-voltage rather than the positive-voltage region of the I/V curve. The backward diode has a negative-resistance region, but the resultant valley of its I/V curve is much less pronounced than in tunnel diodes.

backward wave—In a traveling-wave tube, a wave having a group velocity opposite the direction of electron-stream motion.

backward-wave oscillator—An oscillator employing a special vacuum tube in which oscillatory currents are produced by using an oscillatory electromagnetic field to bunch the electrons as they flow from cathode to anode.

backward-wave tube—A traveling-wave tube in which the electrons travel in a direction opposite to that in which the wave is propagated.

back wave—*See* spacing wave.

baffle—1. In acoustics, a shielding structure or partition used to increase the effective length of the external transmission path between two points (e.g., between the front and back of an electroacoustic transducer). A baffle is often used in a speaker to increase the acoustic loading of the diaphragm. (Although this term sometimes is used to designate the entire cabinet, or enclosure, that houses a loudspeaker, strictly speaking it refers to the panel on which the speaker is mounted, usually the front panel of such an enclosure. The term derives from its original use in preventing, or baffling, the speaker's rear sound waves from interfering with its front waves.) 2. In a gas tube, an auxiliary member placed in the arc path and having no separate external connection. 3. A device for deflecting oil or gas in a circuit breaker. 4. A single shielding device designed to reduce the effect of ambient light on the operation of an optical transmission link. 5. An opaque shielding device designed to reduce the effect of stray light on an optical system.

baffle plate—A metal plate inserted into a waveguide to reduce the cross-sectional area for wave-conversion purposes.

bail—A loop of wire used to prevent permanent separation of two or more parts assembled together (e.g., the bail holding dust caps on round connectors).

Bakelite—A trademark of the Bakelite Corp. for its line of plastic and resins. Formerly, the term applied only to its phenolic compound used as an insulating material in the construction of radio parts.

bake-out—Subjecting an unsealed (hybrid) circuit package to an elevated temperature to bake out moisture and unwanted gases prior to final sealing.

balance—1. The effect of blending the volume of various sounds coming over different microphones in order to present them in correct proportion. 2. The maintenance of equal average volume from both speaker systems of a stereo installation. 3. Relative volume, as between different voices or instruments, bass and treble, or left and right stereo channels. 4. Either a condition of symmetry in an electrical circuit, such as a Wheatstone bridge, or the condition of zero output from a device when properly energized. In the latter sense, depending on the nature of the excitation, two general categories of balance may be encountered: for dc excitation, resistive balance; for ac excitation, resistive and/or reactive balance.

balance control—1. On a stereo amplifier, a differential gain control used to vary the volume of one speaker system relative to the other without affecting the overall volume level. As the volume of one speaker increases and the other decreases, the sound appears to shift from left to center to right, or vice versa. 2. A variable resistor used to compensate for any slight loss of signal in the right or left channel of a stereo amplifier. To some extent, this control can compensate for unbalanced speakers and be used for adjustment when the listener is not in an equidistant position between the two loudspeakers. 3. A variable component, such as a potentiometer or variable capacitor, used to balance bridges, null circuits, or phase speakers.

balanced—1. Electrically alike and symmetrical with respect to ground. 2. Arranged to provide balance between certain sets of terminals. 3. A type of line in which both wires are electrically equal.

balanced amplifier—1. An amplifier circuit with two identical signal branches connected to operate in phase opposition and with their input and output connections each balanced to ground; for example, a push-pull amplifier. 2. A transistor amplifier stage in which two identical transistors are used and the input signal and output power are equally divided between them.

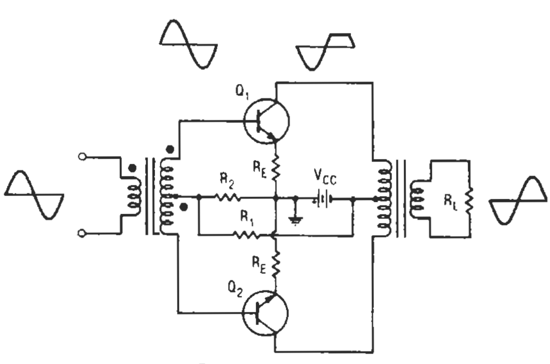

Balanced push-pull amplifier.

This technique produces approximately twice the output power of a single transistor stage, with generally improved dynamic range and reduced voltage standing-wave ratio.

balanced armature—An armature that is approximately in equilibrium with respect to both static and dynamic forces.

balanced-armature unit—The driving unit used in magnetic speakers, consisting of an iron armature pivoted between the poles of a permanent magnet and surrounded by coils carrying the audio-frequency current. Variations in the audio-frequency current cause corresponding changes in the armature magnetism and corresponding movements of the armature with respect to the poles of the permanent magnet.

balanced bridge—A bridge circuit with its components adjusted so that it has an output voltage of zero.

balanced circuit—1. A circuit with two sides electrically alike and symmetrical to a common reference point, usually ground. 2. A circuit terminated by a network that has infinite impedance losses. 3. A circuit terminated by a network whose impedance balances that of the line, resulting in negligible return losses. 4. A circuit whose electrical midpoint is grounded, as opposed to the single-ended circuit, which has one side grounded. 5. A nulled bridge circuit. 6. Telephone circuit in which the two conductors are electrically balanced to each other and to ground. 7. A circuit so arranged that the impressed voltages on each conductor of the pair are equal in magnitude but opposite in polarity with respect to ground.

balanced converter—*See* balun.

balanced currents—Also called push-pull currents. In the two conductors of a balanced line, currents that are equal in value and opposite in direction at every point along the line.

balanced detector—A demodulator for frequency-modulation systems. In one form, the output consists of the rectified difference of the two voltages produced across two resonant circuits, one circuit being tuned slightly above the carrier frequency and the other slightly below.

balanced line—A line or circuit utilizing two identical conductors. Each conductor is operated so that the voltages on them at any transverse plane are equal in magnitude and opposite in polarity with respect to ground. Thus, the currents on the line are equal in magnitude and opposite in direction. A balanced line is preferred where minimum noise and crosstalk are desired.

balanced-line system—A system consisting of a generator, balanced line, and load adjusted so that the voltages of the two conductors at all transverse planes are equal in magnitude and opposite in polarity with respect to ground.

balanced low-pass filter—A low-pass filter designed to be used with a balanced line.

balanced magnetic switch—A magnetic switch that is operated by a balanced magnetic field in such a manner as to resist defeat with an external magnet. It signals an alarm when it detects either an increase or decrease in magnetic field strength.

balanced method—A method of measurement in which the reading is taken at zero. It may be a visual or audible reading; in the latter case, the null is the no-sound setting.

balanced modulator—An amplitude modulator in which the control grids of two tubes are connected for parallel operation, and the screen grids and plates for push-pull operation. After modulation, the output contains the two sidebands without the carrier.

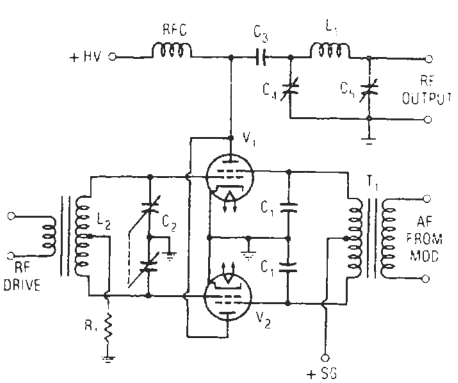

Balanced modulator.

balanced network—1. A hybrid network in which the impedances of the opposite branches are equal. 2. A network in which the corresponding series impedance elements are identical and symmetrical with respect to ground.

balanced oscillator—Any oscillator in which (*a*) the impedance centers of the tank circuits are at ground potential and (*b*) the voltages between either end and the centers are equal in magnitude and opposite in phase.

balanced output—A three-conductor output (as from a microphone) in which the signal voltage alternates above and below a third neutral circuit. This symmetrical arrangement tends to cancel any hum picked up by long lengths of interconnecting cable.

balanced output transformer—1. A push-pull output transformer with a center-tapped primary winding. 2. An output transformer with a grounded center tap on its secondary winding.

balanced probe—A probe, used with an electronic voltmeter or oscilloscope, that has a balanced input and (usually) a single-ended output.

balanced telephone line—A telephone line that is floated with respect to ground so that the impedance measured from either side of the line to ground is equal to that of the other side to ground.

balanced termination—For a system or network having two output terminals, a load presenting the same impedance to ground for each output terminal.

balanced transmission line—A transmission line having equal conductor resistances per unit length and equal impedances from each conductor to earth and to other electrical circuits.

balanced voltages—Also called push-pull voltages. On the two conductors of a balanced line, voltages (relative to ground) that are equal in magnitude and opposite in polarity at every point along the line.

balanced-wire circuit—A circuit with two sides electrically alike and symmetrical to ground and other conductors. Commonly refers to a circuit the two sides of which differ only by chance.

balancer—In a direction finder, that portion used for improving the sharpness of the direction indication. It balances out the capacitance effect between the loop and ground.

balance stripe—A magnetic sound stripe placed on the edge of a motion-picture film opposite the main stripe: it provides mechanical balance for the film.

balance-to-unbalance transformer—A device for matching a pair of lines, balanced with respect to earth, to a pair of lines not balanced with respect to earth. *See also* balun.

balancing network—1. An electrical network designed for use in a circuit in such a way that two branches of the circuit are made substantially conjugate (i.e., such that an electromotive force inserted into one branch produces no current in the other). 2. Electronic circuitry used to match two-wire to four-wire facilities, sometimes called a hybrid. Balancing is necessary to maximize power transfer and minimize echo. 3. Another name for a hybrid, a circuit that connects a two-wire line to a four-wire line and maximizes power transfer while minimizing echo.

balancing unit—An antenna-matching device used to permit efficient coupling of a transmitter or receiver having an unbalanced output circuit to an antenna having a balanced transmission line. 2. A device for converting balanced to unbalanced transmission lines, and vice versa, by placing suitable discontinuities at the junction between the lines instead of using lumped components.

ball—In face bonding, a method of providing chips with contact.

ballast—A device used with an electronic-discharge lamp to obtain the necessary circuit conditions (voltage, current, and waveform) for starting and operating.

ballasting—An integrated circuit design technique that prevents current hogging.

ballast lamp—A lamp that maintains a nearly constant current by increasing its resistance as the current increases.

ballast resistor—A special type of resistor used to compensate for fluctuations in alternating-current power-line voltage. It is usually connected in series with the power supply to a receiver or amplifier. The resistance of a ballast resistor increases rapidly with increases in current through it, thereby tending to maintain an essentially constant current despite variations in the line voltage.

ballast tube—A current-controlling resistance device designed to maintain a substantially constant current over a specified range of variations in the applied voltage to a series circuit.

ball bond—1. A type of thermocompression bond in which a gold wire is flame-cut to produce a ball-shaped end that is then bonded to a metal pad by pressure and heat. 2. A bond formed when a ball-shaped end interconnecting wire is deformed by thermocompression against a metallized pad. The bond is also called a nail-head bond from the appearance of the flattened ball.

ball bonding—A bonding technique that uses a capillary tube to feed the bonding wire. The end of the wire is heated and melts, thus forming a large ball. The capillary and ball are then positioned on the contact area and the capillary is lowered. This forms a large bond. The capillary is then removed and a flame is applied, severing the wire and forming a new ball.

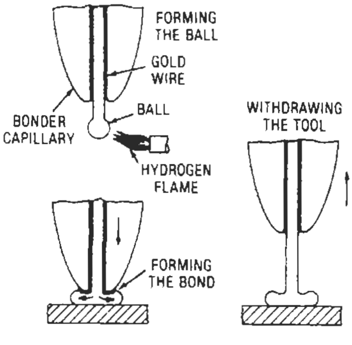

Ball bonding.

ballistic galvanometer—An instrument that indicates the effect of a sudden rush of electrical energy, such as the discharge current of a capacitor.

ballistic-missile early-warning system—An electronic system for providing detection and early warning of attack by enemy intercontinental ballistic missiles. Abbreviated BMEWS.

ballistics—A general term used to describe the dynamic characteristics of a meter movement—most notably, response time, damping, and overshoot.

ballistic trajectory—In the trajectory of a missile, the curve traced after the propulsive force is cut off and the body of the missile is acted upon only by gravity, aerodynamic drag, and wind.

balop—Contraction of balopticon, an apparatus for the projection of opaque images in conjuction with a television camera.

balopticon—*See* balop.

balun (*bal*anced *un*balanced)—1. Also called balanced converter or bazooka. An acronym from *bal*anced to *un*balanced. A device used for matching an unbalanced coaxial transmission line to a balanced two-wire system. 2. Usually, a transformer designed to accept 75-ohm unbalanced input (coaxial cable) and deliver the signal at 300-ohm balanced (twin lead). Usable in the converse sense, and sometimes necessary for matching a tuner with 300-ohm balanced antenna terminals to a 75-ohm coaxial line. 3. An impedance-matching transformer device used to connect balanced twisted-pair cabling with unbalanced coaxial or other cabling systems.

banana jack—A jack that accepts a banana plug. Generally designed for panel mounting.

banana plug—A plug with a banana-shaped spring-metal tip and with elongated springs to provide a low-resistance compression fit.

band—1. Any range of frequencies lying between two defined limits that is used for a specified purpose. 2. A group of radio channels assigned by the FCC to a particular type of radio service.

very low freq. (VLF)	10–30 kHz
low freq. (LF)	30–300 kHz
medium freq. (MF)	300–3000 kHz
high freq. (HF)	3–30 MHz
very high freq. (VHF)	30–300 MHz
ultrahigh freq. (UHF)	300–3000 MHz
superhigh freq. (SHF)	3000–30,000 MHz
extremely high freq. (EHF)	30–300 GHz

3. A group of tracks or channels on a magnetic drum in an electronic computer. *See also* track, 2. 4. In instrumentation, a range of values that represents the scope of operation of an instrument.

bandage—Rubber ribbon, about 4 inches wide, used as a temporary moisture protection for a splice in telephone or coaxial cable.

band center—The geometric mean between the limits of a band of frequencies.

band compression—Reduction of the frequency band needed to transmit a message while maintaining acceptable quality of the message.

banded cable—Two or more cables banded together by stainless-steel strapping.

band-elimination filter—1. Also called band-stop filter. A wave filter with a single attenuation band, neither of the cutoff frequencies being zero or infinite. The filter passes frequencies on either side of this band. 2. A filter that attenuates frequencies within its rejection band, but passes frequencies above and below this band.

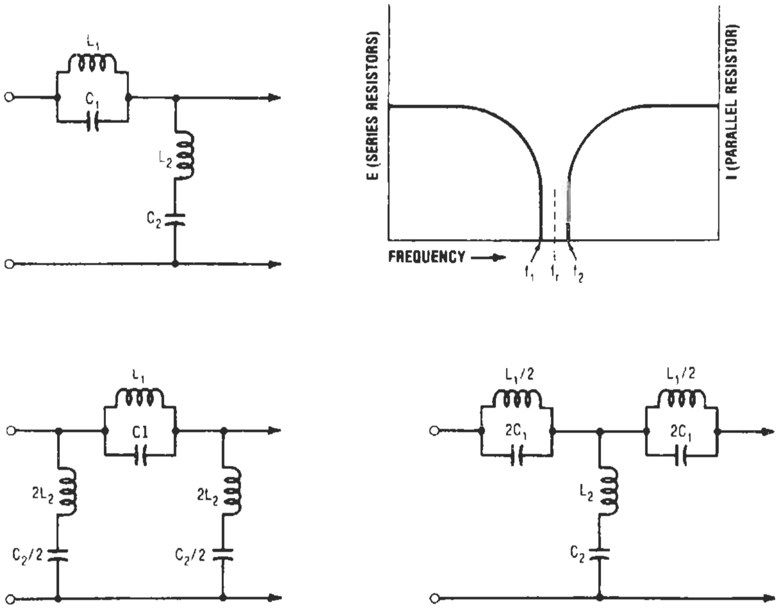
L1
C1
L2
C2
E (SERIES RESISTORS)
I (PARALLEL RESISTOR)
FREQUENCY
f1
fr
f2
L1
C1
2L2
2L2
C2/2
C2/2
L1/2
L1/2
2C1
2C1
L2
C2

Band-elimination filters.

bandgap—1. The energy difference between the conduction band and the valence band in a material. 2. The difference in the energy of an electron in a stable state and an electron in an excited state after absorbing energy. 3. The minimum energy required for a valence electron in a semiconductor to make a transition into the conduction band where the electron can move more freely throughout the crystal.

bandgap energy—The difference in energy between the conduction band and the valence band.

band marking—A continuous circumferential band applied to a conductor at regular intervals for identification.

band opening—A condition that results in greater-than-normal communication range on VHF and UHF amateur bands.

bandpass—1. A specific range of frequencies that will be passed through a device. 2. The number of hertz expressing the difference between the limiting frequencies at which the desired fraction (usually half-power) of the maximum output is obtained.

bandpass amplifier—An amplifier that is tuned to pass only selected frequencies between preset limits.

bandpass amplifier circuit—A stage designed to uniformly amplify signals of certain frequencies only.

bandpass coupling—A coupling circuit with an essentially flat-topped frequency response so that a band of frequencies rather than a single frequency is coupled to a following circuit.

bandpass filter—A wave filter with a single transmission band, neither of the cutoff frequencies being zero

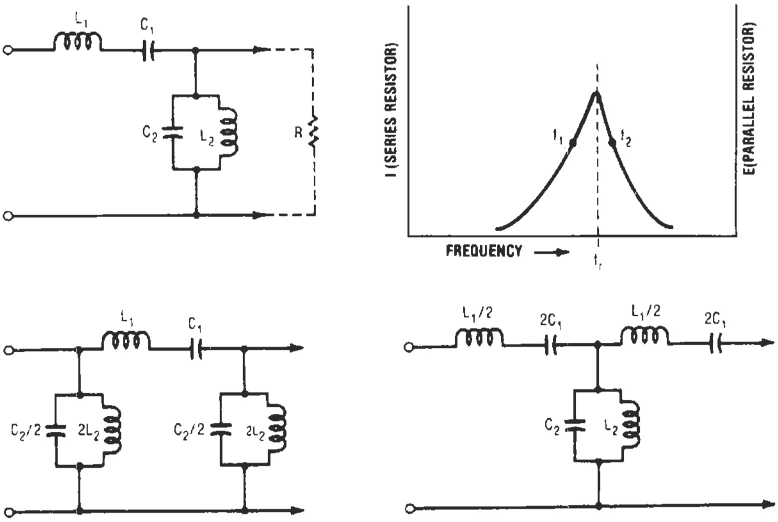
L1
C1
C2
L2
R
I (SERIES RESISTOR)
E(PARALLEL RESISTOR)
f1
f2
FREQUENCY
fr
L1
C1
C2/2
2L2
C2/2
2L2
L1/2
2C1
L1/2
2C1
C2
L2

Bandpass filters.

or infinite. The filter attenuates frequencies on either side of this band.

bandpass flatness—The variations in gain in the bandpass of a filter or tuned circuit.

bandpass response—Also called flat-top response. The response characteristic in which a definite band of frequencies is transmitted uniformly.

band plan—A voluntary system of frequency allocations for each amateur radio band.

band-reject filter—A filter that does not pass a band of frequencies but passes both higher and lower frequencies. Sometimes called a notch filter.

band selector—Also called bandswitch. A switch used to select any one of the bands in which an electronic apparatus such as a receiver, signal generator, or transmitter is designed to operate.

B & S gage—Brown and Sharpe wire gage, in which the conductor sizes rise in geometrical progression. Adopted as the American wire gage (AWG) standard.

bandsplitting—Also called split-streaming. A technique for combining several data channels onto one transmission facility by interleaving the data on a bit-by-bit basis. *See also* multiplexing.

bandspreading—1. The spreading of tuning indications over a wider range to facilitate tuning in a crowded band of frequencies. 2. The method of double-sideband transmission in which the frequency band of the modulating wave is shifted upward so that the sidebands produced by modulation are separated from the carrier by a frequency at least equal to the bandwidth of the original modulating wave. In this way, second-order distortion products may be filtered out of the demodulator output.

bandspread tuning control—A separate tuning control provided on some shortwave receivers to spread the stations in a single band of frequencies over an entire tuning dial.

band-stop filter—*See* band-elimination filter.

bandswitch—*See* band selector.

bandswitching—In a receiver, transmitter, or test instrument, the process of selecting one of two or more self-contained tuned circuits to change from one frequency spectrum to another within the frequency range of the device's intended operation.

bandwidth—1. The range within the limits of a band. The width of a bandpass filter is generally taken as the limits between which its attenuation is not more than 3.0 decibels greater than its average attenuation throughout its passband. Also used in connection with receiver selectivity, transmitted frequency spectrum occupancy, etc. 2. In a given facsimile system, the difference in hertz between the highest and lowest frequency components required for the adequate transmission of the facsimile signals. 3. The least frequency interval of a wave, outside of which the power spectrum of a time-varying quantity is everywhere less than some specified fraction of its value at a reference frequency. 4. The range of frequencies of a device, within which its performance, with respect to some characteristic, conforms to a specified standard. 5. The range of audio frequencies over which an amplifier or receiver will respond and provide a useful output. 6. The complete range of frequencies over which a particular information system can function. Because it varies with length in optical fibers, bandwidth is typically expressed as a frequency-distance (megahertz-kilometer) product. 7. A range of frequencies available for signaling. In data transmission, the greater the bandwidth, the greater the capacity to transmit data bits. *See* hertz. 8. The number of hertz expressing the difference between the lower and upper limiting frequencies of a frequency band; also, the width of a band of frequencies. 9. The information-carrying capability of a communication line or channel. 10. In data communications, the difference between the highest and lowest frequencies of a band. A measure of the capacity of a communication channel, expressed in bits per second or bauds in digital communications channels, and in cycles per second or hertz in an analog system. 11. The difference between an analog signal's lowest frequency component and its highest signal component as measured in hertz (Hz). 12. The speed of a digital communications circuit in bits per second.

bandwidth limited gain control—A control that adjusts the gain of an amplifier while varying the bandwidth. An increase in gain reduces the bandwidth.

bang-bang controller—A discontinuity-type nonlinear system that contains time delay, dead space, and hysteresis.

bank—An aggregation of similar devices (e.g., transformers, lamps) connected and used together. In automatic switching, a bank is an assemblage of fixed contacts over which one or more wipers or brushes move to establish electric connections.

bank-and-wiper switch—A switch in which the electromagnetic ratchets or other mechanisms are used, first, to move the wipers to a desired group of terminals, and second, to move the wipers over the terminals of this group to the desired bank contacts.

bank winding—Also called banked winding. A compact multilayer form of coil winding used for reducing distributed capacitance. Single turns are wound successively in two or more layers, the entire winding proceeding from one end of the coil to the other without being returned.

bantam tube—A compact tube having a standard octal base but a considerably smaller glass envelope than the standard glass octal tube has.

bar—*See* microbar. 1. A subdivision of a crystal slab. 2. A vertical or horizontal line on a television screen, used for testing. 3. A symbol, placed over a letter, used to indicate the inverse, or complement, of a function. For example, inversion of *A* is $\bar{A}$, read "*A* bar" or "not *A*".

bar code—1. A self-contained message with information encoded in the physical widths of bars and spaces in a printed pattern. 2. Coding of consumer products that uses combinations of bars of varying thicknesses. Designed to be read by an optical wand or bar code reader. 3. A printed bar-and-space representation of digital data configured to represent numeric or alphanumeric information. In a more restricted sense, a bar code is a sequence of binary ones and zeros that represents a character, or a set of sequences that represents a set of characters. The machine-readable pattern of bars and spaces is called the symbol. Therefore the term *bar code*, comprising the symbol and the code, represents messages with data encoded in the widths of printed bars and (in some cases) the spaces between the bars. 4. A code made up of a series of variable-width vertical lines that can be read by an optical bar reader.

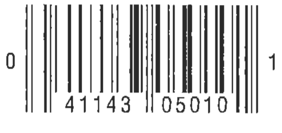

Bar code.

bar code reader—A photoelectric scanner that reads bar codes by means of reflected light.

bar code scanner—An optical scanning device designed to read information printed in the form of bars of different size by detection and processing of the varying reflectivity of light in the bar code.

bare conductor—A conductor not covered with any insulating material.

bar generator—A generator of pulses or repeating waves that are equally separated in time. These pulses are synchronized by the synchronizing pulses of a television system so that they produce a stationary bar pattern on a television screen.

bar-graph display—A display presenting an illuminated line or bar whose length varies in proportion to some parameter being measured.

bar-graph monitoring oscilloscope—An oscilloscope for observation of commutated signals appearing as a series of bars with lengths proportional to channel modulation. The same oscilloscope is commonly used for setup and troubleshooting observations.

barium—An element, the oxide of which is used in the cathode coating of vacuum tubes.

barium titanate—A ceramic that has electric properties and is capable of withstanding much higher temperatures than Rochelle salt crystals. Used in crystal pickups, sonar transducers, and capacitors.

Barkhausen effect—A succession of abrupt changes that occur when the magnetizing force acting on a piece of iron or other magnetic material is varied.

Barkhausen interference—Interference caused by Barkhausen oscillations.

Barkhausen-Kurz oscillator—Circuit for generating ultrahigh frequencies. Its operation depends on the variation in the electrical field around the positive grid and less positive plate of a triode; the variation is caused by oscillatory electrons in the interelectrode spaces.

Barkhausen oscillation—A form of parasitic oscillation in the horizontal-output tube of a television receiver; it results in one or more narrow, dark, ragged vertical lines near the left side of the picture or raster.

Barkhausen oscillator—*See* Barkhausen-Kurz oscillator.

Barkhausen tube—*See* positive-grid oscillator tube.

bar magnet—A bar of metal that has been so strongly magnetized that it holds its magnetism and thereby serves as a permanent magnet.

barn—A unit of measure of nuclear cross sections. Equal to 10^{-24} square centimeter.

Barnett effect—The magnetization resulting from the rotation of a magnetic specimen. The rotation of a ferromagnet produces the same effect as placing the ferromagnet in a magnetic field directed along the axis of rotation. On the macroscopic model, the domains of a ferromagnet can be considered a group of electron systems, each acting as an independent gyroscope or gyrostat.

barometer—An instrument for measuring atmospheric pressure. There are two types of barometers commonly used in meteorology: the mercury barometer and the aneroid barometer.

barometric pressure—The weight of the atmosphere per unit of surface. The standard barometer reading at sea level and 59°F (15°C) is 29.92 inches (760 mm) of mercury absolute.

bar pattern—A pattern of repeating lines or bars on a television screen. When such a pattern is produced by pulses that are equally separated in time, the spacing between the bars on the television screen can be used to measure the linearity of the horizontal or vertical scanning systems.

bar quad—*See* B-quad.

barrage jamming—Simultaneous jamming of a number of adjacent channels or frequencies.

barrel—1. The cylindrical portion of a solderless terminal, splice, or contact, in which the conductor is accommodated. 2. The portion(s) of a terminal or contact that is (are) crimped. When designed to receive the conductor, it is called the wire barrel. When designed to support or grip the insulation, it is called the insulation barrel.

bar relay—A relay in which a bar actuates several contacts simultaneously.

barrel distortion—1. Negative distortion that causes a grid pattern to be imaged as barrel-shaped. 2. In television, distortion that makes the televised image appear to bulge outward on all sides like a barrel. 3. In camera or image tubes, the distortion that results in a monotonic decrease in radial magnification in the reproduced image away from the axis of symmetry of the electron optical system. 4. A distortion of a video image that results in the image resembling a barrel rather than a perfect rectangle. 5. Image distortion in an optical or video system, characterized by bending of the edges of a displayed grid into a barrel shape with convex sides and compressed corners. Can be corrected by image processing techniques.

barrel effect—The boomy or hollow voice quality obtained when the voice is transmitted from a reverberant environment; usually accompanied by a loss of intelligibility and a sense of loss of privacy for far-end users.

barretter—1. A voltage-regulator tube consisting of an iron-wire filament in a hydrogen-filled envelope. The filament is connected in a series with the circuit to be regulated and maintains a constant current over a given voltage variation. 2. A positive coefficient resistor whose resistance increases as temperature increases.

barretter mount—A waveguide mount in which a barretter can be inserted to measure electromagnetic power.

barricade shield—A type of movable shield for protection from radiation.

barrier—1. A partition for the insulation or isolation of electric circuits or electric arcs. 2. In a semiconductor, the electric field between the acceptor ions and the donor ions at a junction. *See* depletion layer.

barrier capacitance—*See* depletion-layer capacitance.

barrier-film rectifier—A rectifier in which a film having unilateral (single-direction) conductivity is in contact with metal or other normally conducting plates.

barrier grid—A grid close to, or in contact with, a storage surface of a charge storage tube. This grid establishes an equilibrium voltage for secondary-emission charging, and it serves to minimize redistribution.

barrier height—In a semiconductor, the difference in potential from one side of a barrier to the other.

barrier layer—*See* depletion layer.

barrier-layer cell—A type of photovoltaic cell in which light acting on the surface of the contact between layers of copper and cuprous oxide causes an electromotive force to be produced. *See* photovoltaic cell.

barrier-layer rectification—*See* depletion-layer rectification.

barrier plate—A layer of slow-diffusing metal (usually palladium or nickel) placed between two fast-diffusing materials to slow or prevent their interdiffusion.

barrier region—*See* depletion region.

barrier shield—A wall or enclosure shielding the operator from an area where radioactive material is being used or processed by remote-control equipment.

barrier strip—1. A terminal strip with protective barriers between adjacent terminals. 2. A continuous section of dielectric material that insulates electrical circuits from each other or from ground.

barrier voltage—The voltage necessary to cause electrical conduction in a junction of two dissimilar materials, such as a pn junction diode.

bar test pattern—Special test pattern for adjusting color TV receivers or color encoders. The upper portion consists of vertical bars of saturated colors and white. The lower horizontal bars have black-and-white areas and I and Q signals.

base—1. The region between the emitter and collector of a transistor which receives minority carriers injected from the emitter. It is the element that corresponds to the control grid of an electron tube. 2. In a vacuum tube, the insulated portion through which the electrodes are connected to the pins. 3. On a printed circuit board, the portion that supports the printed pattern. 4. A thin, strong, and flexible material, usually a polyester or acetate film, on which is deposited a magnetic formulation to make recording tape. *See also* alkali; backing; positional notation; radix. 5. One of the three regions that form a bipolar transistor. It physically separates the emitter and collector regions. Minority carriers are injected from the emitter into the base, where they subsequently either recombine or diffuse into the collector.

base address—A given address from which an absolute address is obtained by combination with a relative address. Also called address constant.

baseband—1. The frequency band occupied by the aggregate of the transmitted signals used to modulate a carrier, before they combine with a carrier in the modulation process. 2. In CD-4 records, the left- or right-channel's band containing musical information for the front and back channels, recorded at listening frequencies in the standard sound spectrum (from 30 to 15,000 hertz). 3. The output signal of a video camera, videotape recorder, or satellite TV receiver before remodulation so that it can be viewed on an ordinary TV set. A signal in a satellite TV receiver goes from 4 GHz through the down-converter to become intermediate frequency and then through an FM modulator to become baseband. The American NTSC TV bandwidth is 4.2 MHz at baseband. 4. The basic direct output signal from a television camera, satellite television receiver, or videotape recorder. Baseband signals can be viewed only on monitors. To display the baseband signal on a conventional television set, a "modulator" is required to convert the baseband signal to one of the VHF or UHF television channels that the television set can be tuned to receive.

baseband frequency response—Response characteristics over the frequency band occupied by all of the signals that modulate a transmitted carrier.

baseband signal—A signal that is not modulated onto a carrier.

baseband signaling—A form of transmission that uses discrete pulses, without modulation. Also called baseband transmission.

baseband transmission—A method of using low-frequency transmission of signals across coaxial cables for short distances. *See also* baseband signaling.

baseband video—Same as composite video (CVS or CVBS). A composite video signal contains video picture information for color, brightness, and synchronization (horizontal and vertical).

base-coupled logic—Abbreviated bcl. A circuit configuration designed for subnanosecond propagation delays and rise and fall times; it can be used for a bit rate of more than 1G bit/s. Base-coupled logic circuits consist of a current-mode switch and emitter followers. Switching is done by means of a base-coupled current-mode switch.

base electrode—An ohmic or majority-carrier contact to the base region of a transistor.

base film—The plastic substrate that supports the coating of magnetic recording tape. The base film of most instrumentation and computer tapes is made of polyester. For less critical uses, cellulose acetate and polyvinyl chloride are employed.

basegroup—1. Designation for a number of carrier channels in a particular frequency range that forms a basic unit (channel bank) for further modulation to a final frequency band. 2. Twelve communication-set paths capable of carrying the human voice on a telephone set; a unit of a frequency-division-multiplexing system's bandwidth allocation.

base insulator—Heavy-duty insulator used to support the weight of an antenna mast and to insulate the mast from the ground or some other surface.

base line—1. In radar displays, the visual line representing the track of the radar scanning beam. 2. In graphical presentations, the horizontal scale, often representing time, bias, or some other variable.

base-line break—In radar, a technique that uses the characteristic break in the base line on an A-scope display due to a pulse signal of significant strength in noise jamming.

base load—In a dc converter, the current that must be taken from the base to maintain a saturated state.

base-loaded antenna—A vertical antenna the electrical height of which is increased by adding inductance in series at the base.

base material—1. An insulating material (usually a copper-clad laminate) used to support a conductive pattern. 2. Rigid or flexible insulating substrate that supports conductive patterns or interconnections. *See* substrate.

base metal—Metal from which connectors, contacts, or other metal accessories are made and on which one or more metals or coatings may be deposited. Sometimes called basis metal.

base number—The radix of a number system (10 is the base number, or radix, for the decimal system; 2 for the binary system).

base-One peak voltage—The peak voltage measured across a resistor in series with base-One when a unijunction transistor is operated as a relaxation oscillator in a specified circuit.

base pin—One of the metal prongs on the base of an electron tube, which makes contact with springs in a tube socket.

base point—*See* radix point.

base region—In a transistor, the interelectrode region, between the emitter and collector junctions, into which minority carriers are injected.

base register—A high-speed computer storage location that contains the address of the first data word or instruction of a segment. All addresses within the segment are referenced to the start of the segment; during execution, the physical address of each memory reference is calculated by simply adding the contents of the base register to the relative addresses contained within the segment. Relocation is easily accomplished by simply moving a segment in memory and placing the new address in the base register.

base resistance—Resistance in series with the base lead in the common-T equivalent circuit of a transistor.

base resistor—The (external) resistor connected to the base of a bipolar transistor. In a common-emitter circuit, the base resistor is analogous to the grid resistor of a vacuum-tube amplifier or the gate resistor of a field-effect transistor amplifier.

base ring—Ohmic contact to the base region of power transistors; so called because it is ring-shaped.

base spreading resistance—In a transistor, the resistance of the base region caused by the resistance of the bulk material of the base region.

base station—1. A fixed station used by a dispatcher to communicate with mobile units. 2. The central transmitter in a communications system that acts as the cell hub for communicating with handsets and/or mobile units. 3. A land station, in the land mobile service, carrying on a service with land mobile stations. (A base station may secondarily communicate with other base stations incident to communications with land mobile stations.) Sometimes defined as a station in a land mobile system that remains in a fixed location and communicates with mobile stations.

base-timing sequencing—Sharing of a transponder on a time basis between several ground transmitters through the use of coded timing signals.

base voltage—The voltage between the base terminals of a transistor and the reference point.

BASIC—1. A simplified computer language intended for use in engineering applications. 2. Acronym for beginners all-purpose symbolic introduction code. An interpreter language that is one of the easier languages to learn and use. 3. An easy-to-learn and easy-to-use programming language developed at Dartmouth College. BASIC can be used to solve business problems as well as in scientific applications.

basic access method—A method of computer access in which each input/output statement results in a corresponding machine input/output operation.

basic frequency—In any wave, the frequency that is considered the most important. In a driven system, it would in general be the driving frequency, whereas in most periodic waves it would correspond to the fundamental frequency.

basic linkage—In a computer, a linkage that is used repeatedly in one routine, program, or system and that follows the same set of rules each time it is used.

basic processing unit—The principal section for control and data processing within a communications system.

basic protection—Fundamental lightning protection measures and/or devices, such as the use of gas tubes or carbon-block protectors, which are applied directly to transmission media at apparatus locations to provide initial voltage limitation.

basic Q—*See* nonloaded *Q*.

basic radio pager—A simple one-way communication device that emits audio sounds when it detects the proper sequence of selective tone signals sent by a dispatcher. The recipient must then telephone the signal originator to get a message. A radio pager is capable of receiving and decoding signals radiated from a call-service transmitter but has no transmitting facilities of its own.

basic rectifier—A metallic rectifier in which each rectifying element consists of a single metallic rectifier cell.

basic speed range—The range over which a motor and control are capable of delivering full load torque without overheating or clogging. It is obtained by armature voltage control.

basic television service—In any of the television-delivery services besides broadcasting (which is free), the monthly fee to subscribe to the lowest tier or set of services and programs, as distinguished from the premium services.

basis metal—*See* base metal.

basket winding—A coil winding in which adjacent turns are separated except at the points of crossing.

bass—Sounds in the low audio-frequency range. On the standard piano keyboard, all notes below middle C (261.63 hertz).

bass boost—A deliberate adjustment of the amplitude-frequency response of a system or component to accentuate the lower audio frequencies.

bass-boosting circuit—A circuit that attenuates the higher audio frequencies in order that low or bass frequencies will be emphasized by comparison.

bass compensation—Emphasizing the low-frequency response of an audio amplifier at low volume levels to compensate for the lowered sensitivity of the human ear to weak low frequencies.

bass control—A manual tone control that has the effect of changing the level of bass frequencies reproduced by an audio amplifier.

bass half-loudness points—The low frequency, in hertz, at which the sound power of a speaker rolls off to become half as loud as the rest of the tonal spectrum; a relative measure of a speaker's ability to reproduce low bass sounds.

bass reflex—A ported loudspeaker configuration using an acoustically tuned vent, through which low-frequency sound from inside the enclosure passes to strengthen and extend bass response.

bass-reflex enclosure—A type of speaker enclosure in which the rear wave from the speaker emerges through an auxiliary opening or port of critical dimensions to reinforce the bass tones. *See also* vented baffle.

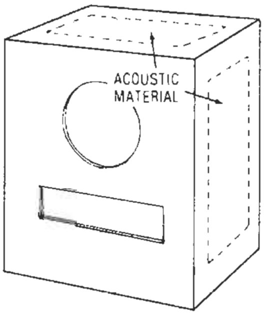

Bass-reflex enclosure.

bass response—1. The extent to which a speaker or audio-frequency amplifier handles low audio frequencies. 2. The ability of any device to pick up or reproduce low audio frequencies.

bassy—Term applied to sound reproduction in which the low-frequency notes are overemphasized.

BAT—Abbreviation for battery.

batch—1. A group of documents to be processed; an arbitrary subdivision of a job by the supervisor into smaller, more manageable parts. A batch is the smallest group of such documents accessible by name (number) for data entry, data verify, peripheral device transfers, etc. 2. A type of computer operating system in which jobs are processed one at a time and a job must complete its execution before the next is begun. 3. Computer processing mode in which a program is submitted and the result is delivered back. No interactive communication between program and user is possible. 4. The group of programs considered as a single unit for processing on a computer.

batch control sample—A representative batch extracted either at random or at specific intervals from a process or product for quality-control purposes. Results of equivalent tests of the batches are averaged to interpolate the quality of the total process.

batch environment—A situation in which a computer receives instructions and program(s) from a terminal or other peripheral device, then executes the requested operations at its own convenience.

batch file—A text file that contains operating system and program commands, used for standard procedures performed repetitiously.

batching—*See* batch processing.

batch mode—Operational mode in a computer in which input to or output from a process is transmitted as a single set of successive messages. Contrast with *interactive mode*.

batch process—A method of fabricating monolithic resistors, capacitors, and diodes with the same process at the same time.

batch processing—1. In a computer, a method of processing in which a number of similar input items are grouped for processing during the same machine run. 2. Pertaining to the technique of executing a set of computer programs such that each is completed before the next program of the set is started. Loosely, the execution of computer programs serially. 3. A technique by which items to be processed must be coded and collected into groups prior to processing. 4. A data-processing technique in which input data is accumulated offline and processed in batches.

bat handle—Standard form of a toggle-switch lever, having a shape similar to that of a baseball bat.

bathtub capacitor—A type of capacitor enclosed in a metal housing having broadly rounded corners like those on a bathtub.

bathyconductorgraph—A device used from a moving ship to measure the electrical conductivity of seawater at various depths.

bathythermograph—A device that automatically plots a graph showing temperature as a function of depth when lowered into the sea.

Batten system—A method developed by W. E. Batten for coordinating single words in a computer to identify a document. Sometimes called peek-a-boo system.

battery—Abbreviated BAT. 1. A dc voltage source consisting of two or more cells that converts chemical, nuclear, solar, or thermal energy into electrical energy. 2. In communications, a source (not necessarily a storage device) of direct current or the current itself. 3. Two or more cells coupled together in series or parallel. In the former configuration the arrangement gives a greater voltage (two cells give twice the voltage, three cells give three times the voltage, and *n* cells give *n* times the voltage); the latter arrangement gives the same voltage as the individual cell but a greater current.

battery acid—A solution that serves as the electrolyte in a storage battery. In the common lead-acid storage battery, the electrolyte is diluted sulfuric acid.

battery cable—A single conductor cable, either insulated or uninsulated, used for carrying current from batteries to the point where power is needed. May also be used for grounding.

battery capacity—The amount of energy obtainable from a storage battery, usually expressed in ampere-hours.

battery charger—Device used to convert alternating current into a pulsating direct current that can be used for charging a storage battery.

battery clip—A metal clip with a terminal to which a connecting wire can be attached, and with spring jaws that can be quickly snapped onto a battery terminal or other point to which a temporary connection is desired.

battery life—The number of times that a battery can be charged and discharged. One complete charge and one complete discharge is called a cycle. The number of complete cycles a battery will give depends on construction of the battery, charging procedure, maintenance, and operation.

battery post adapter—A device connected to a battery post and used for connecting ammeter leads and simulating a charged battery.

battery pulses—Negative potential pulses from the central office battery that are applied to a telephone circuit.

battery receiver—A radio receiver that obtains its operating power from one or more batteries.

battery separator—An insulator that separates the positive and negative plates in a storage battery.

battle short—A switch for short-circuiting safety interlocks and lighting a red warning light.

bat wing—An element on an FM or TV transmitting or receiving antenna, so called because of its shape.

baud—1. A unit of signaling speed derived from the duration of the shortest code element. 2. A unit of signaling speed equal to the number of discrete conditions or signal events per second. For example, in Morse code 1 baud equals one-half dot cycle per second; in a train of binary signals, 1 baud is 1 bit per second; and in a train of signals, each of which can assume one of eight different states, 1 baud is one 3-bit value per second. 3. A measurement of communication channel capacity as a function of time. For example, a 110-baud line is divided into 110 equal parts. Within each of these parts a certain amount of data can be placed, typically, one bit. This means that a speed of 110 baud is 110 bits per second. 4. The unit of modulation rate (or signaling speed). The reciprocal of the duration of the minimum signaling element (pulse width). 5. The number of times per second the line condition changes. If the line condition represents the presence or absence of a single bit (as in two-state signaling), then the signaling speed in bauds is the same as bits per second. However, if the signaling is not two-state, then bauds are not equal to bits per second. The latter condition exists, for instance, in "dibit" or four-state signaling, in which the baud rate is equivalent to the number of bits per second times two. 6. In an equal length code, 1 baud corresponds to a rate of one signal element per second. Thus, with a signal element duration of 20 ms, the modulation rate is 50 baud (per second). The term is both singular and plural. 7. A data-communication-rate unit used similarly to bits per second (bps) for low-speed data; the number of signal-level changes per second (regardless of the information the signals contain). 8. A measure of serial data flow between a computer and/or communication devices. One baud is equal to 1 bit per second (1 bps). 9. Unit of signaling speed. The speed in bauds is the number of line changes (in frequency, amplitude, etc.) or events per second. At low speeds, each event represents only one bit condition, and baud rate equals bps. As speed increases, each event represents more than one bit, and baud rate does not truly equal bps. But in common usage, baud rate and bps are often used interchangeably. 10. A variable unit of data transmission speed (as one bit per second). Often confused with bits per second (bps). 11. Signaling rate unit for analog communications. One baud is equal to one change of state per second. Where there are two possible states (e.g., two tone frequencies), 1 baud equals 1 bps. Where there are 2^n possible states (e.g., four possible phase shifts in a sinusoidal wave), 1 baud equals *n* bps. In low-speed transmission, in which modems with frequency-shift keying are used, the terms *baud* and *bps* may be used interchangeably. Otherwise, the two terms are never interchangeable.

Baudot code — 1. A data-transmission code in which one character is represented by five equal-length bits. This code is used in most dc teletypewriter machines, in which one start element and 1.42 stop elements are added. 2. Five-level code (plus one start and one stop bit) used primarily by amateurs in rtty communication. Only code allowed by the FCC without special waiver. 3. A code of 32 numbers used for alphabetic and symbolic communication. Invented in 1880 by the Frenchman J. M. E. Baudot, the code bears his name; the term *baud* is also derived from it. 4. Data-transmission code in which five bits represent one character. Use of shift letters/figures enables 64 alphanumeric characters to be represented. Baudot is used in many teleprinter systems with one start bit and 1.5 stop bits added.

baud rate — 1. The number of signal events per second occurring on a communications channel. Although not technically accurate, baud rate is commonly used to mean bit rate. *See* bps. 2. The speed at which data is transmitted, measured in symbols per second. This is not the same as bits per second, since each symbol can carry several bits of information. 3. The number of code elements transferred in one second based on the length of the shortest element. If each element consists of one bit, the baud rate would equal the number of bits per second. 4. Binary speed through a serial interface, traditionally defined as the number of signal elements per second. When each element is one bit, the baud rate equals the number of bits per second. 5. The number of bits transmitted per second in a serial data transmission system. The number of bits per second may also include control bits as well as data bits.

baud-rate generator — Oscillator, usually adjustable, that provides clock signals for connection of a peripheral. Typical rates are 110, 300, 9600 bauds and higher.

bay — 1. A portion of an antenna array. 2. A vertical compartment in which a radio transmitter or other equipment is housed.

bayonet base — A base having two projecting pins on opposite sides of a smooth cylindrical base; the pins engage corresponding slots in a bayonet socket and hold the base firmly in the socket.

bayonet coupling — 1. A quick-coupling device. Connection is accomplished by rotating two parts under pressure. Pins on the side of the male connector engage slots on the side of the female connector. 2. A quick-coupling device for plug and receptacle connectors, accomplished by rotation of a cam operating device designed over a short cylindrical stud or studs to bring the connector halves together.

bayonet socket — A socket for bayonet-base tubes or lamps; it has slots on opposite sides and one or more contact buttons at the bottom.

bazooka — *See* balun.

B battery — The battery that furnishes the required dc voltages to the plate and screen-grid electrodes of the vacuum tubes in a battery-operated circuit.

BBS — Abbreviation for bulletin board system. A system by which a group of users with common interests (such as a business, club, or professional society) can share information by posting it to an electronic bulletin board. Some subscriber-based BBSs offer limited Internet services. The system allows people to carry on discussions, upload and download files, and make announcements without the people being connected to the computer at the same time. *See also* bulletin board.

BC — Abbreviation for bare copper or bell cord.

BCAS — Abbreviation for beacon collision avoidance system. An airborne automatic collision-avoidance system.

BCD — Abbreviation for binary-coded decimal. 1. A system of representing numerical, alphabetic, and special characters in which individual decimal digits are represented by some binary code. For example, in an 8-4-2-1 BCD notation, 16 might be represented as 001 (for 1) and 0110 (for 6). In pure binary notation, 16 is 10000; the number 23 is represented by 0010 0011 in the BCD notation. 2. A code defined with 6 bits per character and capable, therefore, of representing up to 64 unique values, each representing one of the 26 letters of the English alphabet (uppercase only), a digit (0 to 9), or a special or punctuation character. BCD is no longer used. *See* EBCDIC.

BCD code — An eight-bit code (of which one bit is used for odd parity check) used by IBM. It provides 88 characters, plus 16 functions.

BCD counter — A counter in which each section consists of four flip-flops or stages, each section of which counts to nine (binary 1001) and then resets to zero (binary 0000). The outputs are in BCD form.

B channel — One of two stereo channels, usually the right, together with the microphone, speakers, or other equipment associated with this channel.

bci — Abbreviation for broadcast interference, a term denoting interference by transmitters with reception of broadcast signals on standard broadcast receivers.

B-display — On a radarscope, a type of presentation in which the target appears as a bright spot. Its bearing is indicated by the horizontal coordinate and its range by the vertical coordinate.

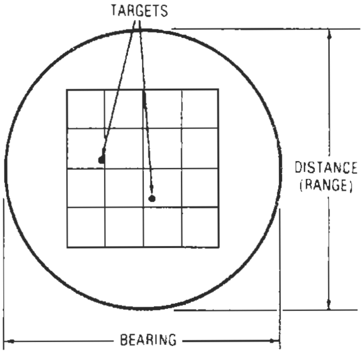

B-display.

beacon — 1. A device that emits a signal for use as a guidance or warning aid. Radar beacons aid the radar set in locating and identifying special targets that may otherwise be difficult or impossible to sense. 2. Low-power carrier transmitted by a satellite that supplies the controlling engineers on the ground with a means of monitoring telemetry data, tracking the satellite, or conducting propagation experiments. This tracking beacon is usually a horn or omni antenna.

beacon delay — The amount of inherent delay within the beacon, i.e., the time between the arrival of a signal and the response of the beacon. In a pulse beacon, delay ordinarily is measured between the leading 3-dB points of the triggering pulse and the reply pulse.

beacon receiver—A radio receiver for converting into perceptible signals the waves emanating from a radio beacon.

beacon skipping—A term used to describe a condition in which beacon return pulses are missing at the interrogating radar. Beacon skipping can be caused by interference, overinterrogation of the beacon, antenna nulls, or pattern minima.

beacon station—1. A station broadcasting beacon signals for direction finding or navigation. 2. A radar transmitting station.

beacon stealing—The loss of beacon tracking by a (desired) radar due to (interfering) interrogation signals from another radar.

beacon time-sharing—A technique by which two or more radars may interrogate and track a long-recovery type of beacon without exceeding the duty cycle of the beacon. This technique is accomplished by the proper sequencing of the various radar interrogations. It is necessary to ensure that the total of all interrogations does not exceed the beacon duty cycle and that enough time is allowed for the modulator section of the beacon to recover before it receives the next interrogation.

beacon transmitter—A transmitter specially adapted for the transmission of beacon signals.

beaded coax—A coaxial cable in which the dielectric consists of beads made of various insulating materials.

beaded support—Ceramic and plastic beads used to support the inner conductor in coaxial transmission lines.

beaded transmission line—A line using beads to support the inner conductor in coaxial transmission lines.

bead thermistor—A thermistor consisting of a small bead of semiconducting material, such as germanium, placed between two wire leads. Used for microwave power measurement, temperature measurement, and as a protective device. The resistance decreases as the temperature increases.

beam—1. A flow of electromagnetic radiation concentrated in a parallel, converging, or diverging pattern. 2. The unidirectional or approximately unidirectional flow of radiated energy or particles. 3. A shaft or column of light, a bundle of rays that may or may not consist of parallel, converging, or diverging rays.

beam-addressable technology—The applications of reversible writing with a laser beam on particular storage materials. In one method, an amorphous film is heated and then crystallized for writing. Bubble writing involves the formation of bubbles on the film-glass interface of a glass substrate by the intense light beam.

beam alignment—The adjustment of the electron beam in a camera tube (on tubes employing low-velocity scanning) to cause the beam to be perpendicular to the target surface.

beam angle—The angle between the directions, on either side of the axis, at which the intensity of the radio-frequency field drops to one-half the value it has on the axis.

beam antenna—1. An antenna that concentrates its radiation into a narrow beam in a definite direction. 2. A directional antenna that radiates or intercepts more energy in one direction than in others.

beam bender—*See* ion trap.

beam bending—Deflection of the scanning beam of a camera tube by the electrostatic field of the charges stored on the target.

beam blanking—Interruption of the electron beam in a cathode-ray tube by the application of a pulse to the control grid or cathode.

beam breaker—A level-measuring device using a light or electrical beam or pneumatic jet between a source and a detector or reflector.

beam candlepower—The candlepower of a bare source that, if located at the same distance as the beam, would produce the same illumination as the beam.

beam convergence—The converging of the three electron beams of a three-gun color picture tube at a shadow-mask opening.

beam-coupling coefficient—In a microwave tube, the ratio of the amplitude, expressed in volts, of the velocity modulation produced by a gap to the radio-frequency gap voltage.

beam crossover—The point of overlap of a beam from an antenna that is nutated or rotated about the center line of the antenna radiation direction. The crossover point is normally at the half-power point. The received energy, when commutated into four quadrants, provides the necessary information for the servoamplifier error signal used to align the antenna to a target.

beam current—The current carried by the electron stream that forms the beam in a cathode-ray tube.

beam cutoff—In a television picture tube or cathode-ray tube, the condition in which the control-grid potential is so negative with respect to the cathode that electrons cannot flow and thereby form the beam.

beam-deflection tube—An electron-beam tube in which current to an output electrode is controlled by the transverse movement of an electron beam.

beam droop—A form of distortion of the normal rectilinear fan-shaped radiation pattern of a detection radar in which a portion of the fan is at a lower elevation than the rest of the fan.

beam-forming electrode—Electron-beam focusing elements in power tetrodes and cathode-ray tubes.

beam hole—An opening through a reactor shield and, generally, through the reactor reflector that permits a beam of radioactive particles or radiation to be used for experiments outside the reactor.

beam-index color tube—A color picture tube in which the signal generated by an electron beam after deflection is fed back to a control device or element in such a way that an image in color is provided.

beam lead—1. A metal beam deposited directly onto the surface of the die as part of the wafer processing cycle in the fabrication of an integrated circuit. Upon separation of the individual die (normally by chemical etching instead of the conventional scribe-and-break technique), the cantilevered beam is left protruding from the edge of the chip and can be bonded directly to interconnecting pads on the circuit substrate without the need for individual wire interconnections. 2. A long structural member not supported everywhere along its length and subject to the forces of flexure, one end of which is permanently attached to a chip device and the other end intended to be bonded to another material, providing an electrical interconnection or mechanical support or both.

beam-lead bonding—1. A free-bonding technique in which thick gold extensions of the thin-film terminals of semiconductor devices and circuits are electroformed so they extend beyond the edges of the chips. 2. A method of interconnecting ICs in a circuit by bonding beam leads located on the IC chip's back surface to the circuit's conducting paths. 3. A hybrid bonding technique that provides for multiple bonding simultaneously.

beam-lead device—An active or passive chip component possessing beam leads as its primary interconnection and means of mechanical attachment to a substrate.

beam-lead isolation—The method in which electrical isolation between IC elements is produced by interconnecting the elements with thick gold leads and selectively etching the silicon from between elements without affecting the gold leads. This process leaves the elements as separate units supported by the gold leads.

beam leads—1. A generic term describing a system in which flat metallic leads extend from the edges of a chip component much as wooden beams extend from a roof overhang. These are then used to interconnect the component to film circuitry. 2. Techniques for attachment of lead frames to silicon chips, including vacuum and chemical deposition, diffusion thermal-compression techniques, welding, etc.

beam-lobe switching—A method of determining the direction of a remote object by comparison of the signals corresponding to two or more successive beam angles at directions slightly different from that of the object.

beam modulation—*See* z-axis modulation.

beam optics—A discipline within the broad study of optics that is specifically oriented toward the investigation of waves with small angular divergence.

beam parametric amplifier—A parametric amplifier in which a modulated electron beam provides a variable reactance.

beam-positioning magnet—A magnet used with a tricolor picture tube to influence the direction of one of the electron beams so that it will have the proper spatial relationship with the other two beams.

beam-power tube—An electron-beam tube in which directed electron beams are used to contribute substantially to its power-handling capability, and in which the control and screen grids essentially are aligned.

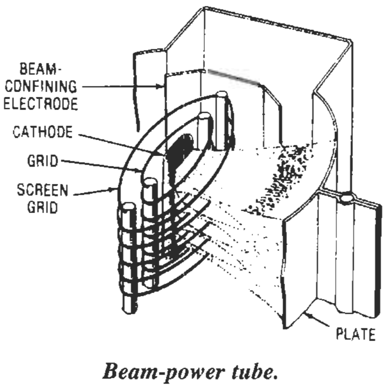

Beam-power tube.

beam relaxor—A type of sawtooth scanning-oscillator circuit that generates but does not amplify the current wave required for magnetic deflection in a single-beam-power pentode.

beam-rider control system—A system whereby the control station sends a beam to the target, and the missile follows this beam until it collides with the target.

beam-rider guidance—A form of missile guidance wherein a missile, through a self-contained mechanism, automatically guides itself along a beam transmitted by a radar.

beams—CB radio term for any type of directional antenna.

beam splitter—A device used for dividing a light beam (as by a transparent mirror) into two components, one transmitted and the other reflected. 2. An optical device used to divide the beam of light into two beams. Either prisms or partial mirrors.

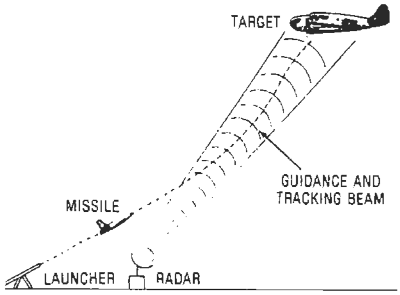

Beam-rider guidance.

beam splitting—A process for increasing the accuracy of locating targets by radar. By noting the azimuths at which one radar scan first discloses a target and at which radar data from it ceases, beam splitting calculates the mean azimuth for the target.

beam-splitting mirror—In an oscilloscope camera system, a tilted, transparent mirror that allows rays to pass horizontally from the oscilloscope screen to the camera and also to be reflected vertically to the viewer's eye.

beam spreader—An optical element the purpose of which is to impart a small angular divergence to a colliminated incident beam.

beam switching—A method of obtaining the bearing and/or elevation of an object more accurately by comparing the signals received when the beam is in a direction differing slightly in bearing and/or elevation. When these signals are equal, the object lies midway between the beam axes.

beam-switching tube—A multiposition, high-vacuum, constant-current distributor. The beam-switching tube consists of many identical arrays around a central cathode. Each array comprises a spade that automatically forms and locks the electron beam, a target-output electrode that gives the beam current its constant characteristics, and a high-impedance switching grid that switches the beam from target to target. A small cylindrical magnet, permanently attached to the glass envelope, provides a magnetic field. This field, in conjunction with an applied electric field, comprises the crossed fields necessary for operation of this tube. It is used in electronic switching and in distributing, such as counting, timing, sampling, frequency dividing, coding, matrixing, telemetering, and controlling.

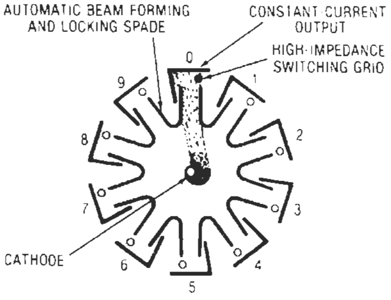
AUTOMATIC BEAM FORMING AND LOCKING SPADE
CONSTANT CURRENT OUTPUT
HIGH-IMPEDANCE SWITCHING GRID
0
1
2
3
4
5
6
7
8
9
CATHODE

Beam-switching tube.

beam width—1. The angular width of a radio, radar, or other beam measured between two reference lines. 2. The width of a radar beam measured between lines of half-power intensity. 3. Of a dish antenna, the angle of sky that can be illuminated (picked up or sent out) by the dish. Large dishes have narrow beam widths, which reduce noise from their sides. Small dishes have wider beam widths and are noisier but easier to aim.

bearing—1. The horizontal direction of an object, or point, usually measured clockwise from a reference line or direction through 360°. 2. Support for a rotating shaft. 3. The horizontal angle at a given point, measured from a specific reference datum, to a second point relative to another as measured from a specific reference datum.

bearing cursor—A mechanical bearing line of a plan-position indicator type of display for reading the target bearing.

bearing loss—The loss of power through friction in the bearings of an electric motor (brushes removed and no current in the windings).

bearing resolution—The minimum angular separation in a horizontal plane between two targets at the same range that will allow an operator to obtain data on either individual target.

beat—Periodic variations that result from the superimposition of waves having different frequencies. The term is applied both to the linear addition of two waves, resulting in a periodic variation of amplitude, and to the nonlinear addition of two waves, resulting in new frequencies, of which the most important usually are the sum and difference of the original frequencies.

beat frequency—One of the two additional frequencies produced when two different frequencies are combined. One beat frequency is the sum of the two original frequencies; the other is the difference between them.

beat-frequency oscillator—Abbreviated BFO. An oscillator that produces a signal which mixes with another signal to provide frequencies equal to the sum and difference of the combined frequencies.

beating—The combining of two or more frequencies to produce sum and difference frequencies called beats.

beating-in—Interconnecting two transmitter oscillators and adjusting one until no beat frequency is heard in a connected receiver. The oscillators are then at the same frequency.

beating oscillator—*See* local oscillator.

beat marker—A marker pip resulting from the beat not between the sweep-generator signal and the signal from a marker oscillator; a marker is visible on an oscilloscope during visual alignment of a tuned circuit.

beat note—The difference frequency produced when two sinusoidal waves of different frequencies are applied to a nonlinear device.

beat reception—*See* heterodyne reception.

beats—1. Beat notes that are generally at a sufficiently low audio frequency that they can be heard or counted. 2. The signal formed when two signals of different frequencies are simultaneously present in a nonlinear device. The frequency of the beat is equal to the difference in frequency of the two primary signals. For example, beats are produced in superheterodyne receivers, where the beat is between the incoming signal and the local oscillator in the receiver. 3. Periodic variations of amplitude that result when two periodic waves having different frequencies are superimposed.

beat tone—Musical tone due to beats, produced by the heterodyning of two high-frequency wave trains.

beaver tail—A fan-shaped radar beam, wide in the horizontal plane and narrow in the vertical plane. The beaver tail is swept up and down for height finding.

bed-of-nail tester—*See* in-circuit-tester.

bedspring—A broadside antenna array with a flat reflector.

before start—The interval before the starting circuit to a timer has been operated. The timer is fully reset and all contacts are in the precycle position.

bel—The fundamental unit in a logarithmic scale for expressing the ratio of two amounts of power. The number of bels is equal to the $\log_{10} P_1/P_2$, where P_1 is the power level being considered and P_2 is an arbitrary reference level. The decibel, equal to 1/10 bel, is a more commonly used unit.

B eliminator—A power pack that changes the ac power-line voltage to the dc source required by the vacuum tubes. In this way, batteries can be eliminated.

bell—1. An electrical device consisting of a hammer vibrated by an electromagnet. The hammer strikes the sides of the bell and emits a ringing noise. The electromagnet attracts an armature or piece of soft iron forming part of the hammer lever. A contact breaker then opens the circuit and cuts off the attraction. A spring draws the hammer back to its original position, closing the circuit and repeating the action. 2. An electromechanical device in which an electrically vibrated clapper repeatedly strikes one or two gongs, which give out a musical tone.

Bellini-Tosi antenna—A direction-finding antenna comprising two vertical orthogonal triangular loops installed with their bases over ground and used with a goniometer.

Bellini-Tosi direction finder—An early radio direction-finder system consisting of two loop antennas at right angles to each other and connected to a goniometer.

bellows—1. A pressure-sensing element consisting of a ridged metal cylinder closed at one end. A pressure difference between its outside and inside will cause the cylinder to expand or contract along its axis. 2. A mechanical pressure-sensing element consisting of a metallic bellows with a plate on one end. Pressure applied to the open end causes the plate to move. The amount of the movement becomes a measure of the applied pressure.

bellows contact—A contact in which a multileaf spring is folded. This type provides a more uniform spring rate over the full tolerance range of the mating unit.

bell-shaped curve—A statistical curve (so called from its characteristic shape) that exhibits a normal distribution of data. The curve typically describes the distribution of errors of measurement around the real value.

bell transformer—A small iron-core transformer; its primary coil is connected to an ac primary line, and its secondary coil delivers 10 to 20 volts for operation of a doorbell, buzzer, or chimes.

bell wire—Cotton-covered copper wire, usually No. 18, used for doorbell and thermostat connections in homes and for similar low-voltage work.

belt drive—A drive system used to rotate a turntable, in which the motor pulley drives the platter with a belt.

benchmark—1. In connection with microprocessors, a frequently used routine or program selected for the purpose of comparing different makes of microprocessors. A flowchart in assembly language is written out for each microprocessor, and the execution of the benchmark by each unit is evaluated on paper. (It is not necessary to use hardware to measure capability by benchmark.) 2. A point of reference from which measurements can be made. 3. A test standard for measuring product performance. 4. A test point for facilitating measurement of a product run on several computers for the purposes of comparing speed, throughput, and ease of conversion. 5. The act of determining a benchmark. 6. Standard measure used to test performance.

benchmark problem—A problem used in the evaluation of the performance of computers relative to each other.

benchmark program—1. A sample program used to evaluate and compare computers. In general, two computers will not use the same number of instructions, memory words, or cycles to solve the same problems. 2. A specific program written to measure the speed of a computer in a well-defined situation, such as serial transfer and 8-bit by 8-bit multiplication. 3. A set of standards used in testing a software or hardware product or system, from which a measurement can be made. Benchmarks are often run on a system to verify that it performs according to specifications.

bench test—A test in which service conditions are approximated, but the equipment is conventional laboratory equipment and not necessarily identical with that in which the product will be employed in normal service.

bend—A change in the direction of the longitudinal axis of a waveguide.

bend loss—1. A form of attenuation that is caused by bending an optical fiber at a restrictive radius. The term also applies to losses due to minute distortions in the fiber itself caused by bending it. 2. A form of increased attenuation caused by allowing high-order modes to radiate from the side of an optical fiber. The two common types of bend losses are (*a*) those occurring when the fiber is curved around a restrictive radius of curvature and (*b*) microbend caused by small distortions of the fiber imposed by externally induced perturbations, such as poor cabling techniques.

bend radius—The minimum radius an optical fiber can bend before breaking.

bend waveguide—A section of waveguide in which the direction of the longitudinal axis is changed.

Benito—A cw navigational system in which the distance to an aircraft is determined on the ground by measuring the phase difference of an audio signal transmitted from the ground and retransmitted by the aircraft. Bearing information is obtained by ground direction finding of the aircraft signals.

bent gun—A TV picture-tube neck arrangement with an electron gun that is slanted to direct the undesired ion beam toward a positive electrode and still allow the electron beam to pass to the screen.

beryllia—1. Beryllium-oxide (BeO) ceramics that have high thermal conductivity characteristics. Used as substrates in hybrid circuit manufacturing and for thick-film substrates in high-power applications. 2. Beryllium oxide is used in various forms as an insulator and structural element (as in resistor cores).

beryllium—An elemental metal whose atomic number is 4. Beryllium is present in various dielectrics and alloys used in electronics.

beryllium oxide (BeO)—A ceramic material having very high heat conductivity, good thermal shock resistance, and high strength. Used in metal/ceramic packages for higher-power microwave transistors and as substrates in some MIC power amplifiers.

Bessel function—A mathematical function used in the design of a filter for maximally constant time delay with little consideration for amplitude response. This function is very close to a Gaussian function.

best fit—An algorithm for computer memory allocation that searches the memory-free list for the unused memory block that is closest in size to that needed by the requesting task.

beta—Symbolized by the Greek letter *beta* (β). Also called current-transfer ratio. 1. The current gain of a transistor connected as a grounded-emitter amplifier; it is the ratio of a small change in collector current to the corresponding change in base current, with the collector voltage constant. 2. A parameter used to express the current gain of a bipolar transistor. There are many versions of beta, but all relate a change in collector current to the corresponding change in base current, with the collector-emitter voltage kept constant. 3. A symbol used to denote B quartz. 4. Brainwave signals whose frequency is approximately 13 to 28 Hz. The associated mental state is irritation, anger, jitteriness, frustration, worry, tension, etc. 5. A prerelease version of software, distributed to a selected group of users to test. By the end of a beta test, all major bugs should have been discovered and repaired.

beta circuit—In a feedback amplifier, the circuit that transmits a portion of the amplifier output back to the input.

beta cutoff frequency—The frequency at which the beta of a transistor is 3 decibels below the low-frequency value.

beta particle—A small electrically charged particle thrown off by many radioactive materials. It is identical with the electron and possesses the smallest negative electrical charge found in nature. Beta particles emerge from radioactive material at high speeds, sometimes close to the speed of light.

beta ray—1. A stream of beta particles. 2. Electrons or positrons given off by a radioactive nucleus in the process of decay.

beta test—Testing computer software before it is released commercially.

betatron—A large doughnut-shaped accelerator that produces artificial beta radiation. Electrons (beta particles) are whirled through a changing magnetic field. They gain speed with each trip and emerge with high energies (on the order of 100 million electron volts in some instances).

bev—A billion electron volts. An electron possessing this much energy travels at a speed close to that of light—186,000 miles a second (3×10^8 m/s).

bevatron—A very large circular accelerator in which protons are whirled between the poles of a huge magnet to produce energies in excess of one billion electron volts.

Beverage antenna—*See* wave antenna.

beyond-the-horizon propagation—*See* scatter propagation.

bezel—1. A holder designed to receive and position the edges of a lens, meter, window, or dial glass. 2. The flange or cover used for holding an external graticule or CRT cover in front of the CRT in an oscilloscope. May also be used for mounting a trace-recording camera or other accessory item.

BFO—Abbreviation for beat-frequency oscillator.

***B-H* curve**—Curve plotted on a graph to show successive states during magnetization of a ferromagnetic material. A normal magnetization curve is a portion of a symmetrical hysteresis loop. A virgin magnetization curve shows what happens the first time the material is magnetized.

***B-H* meter**— A device for measuring the intrinsic hysteresis loop of a sample of magnetic material.

biamplification—The technique of splitting the audio-frequency spectrum into two sections and using individual power amplifiers to drive a separate woofer and tweeter. Crossover frequencies for the amplifiers usually vary between 500 and 1600 Hz. Biamplification has the advantages of allowing smaller-power amps to produce a given sound pressure level and reducing distortion effects produced by overdrive in one part of the frequency spectrum affecting the other part.

bias—1. The electrical, mechanical, or magnetic force applied to a relay, semiconductor, vacuum tube, or other device for the purpose of establishing an electrical or mechanical reference level for the operation of the device. 2. Direct-current potential applied to the control grid of a vacuum tube. 3. Bias derived from a direct current, used on signaling or telegraph relays or electromagnets to secure the desired time spacing of transitions from marking to spacing. 4. A method of restraining a relay armature, by means of spring tension, to secure a desired time spacing of transitions from marking to spacing. 5. The average direct-current voltage between the control grid and cathode of a vacuum tube. 6. The effect on teletypewriter signals produced by the electrical characteristics of the line and the equipment. 7. Energy applied to a relay to hold it in a given position. 8. A high-frequency signal applied to the audio signal at the tape recording head to minimize distortion and noise and increase frequency response and efficiency. Although sometimes dc (fixed magnetic polarity) is used, the bias signal is usually above 40 kHz to avoid audible intermodulation distortion. Every tape formulation has slightly different bias requirements. 9. The sideways thrust of a pickup arm. 10. Communication-signal distortion related to bit timing. 11. The departure from a reference value of the average of a set of values.

bias cell—A dry cell used in the grid circuit of a vacuum type to provide the necessary C-bias voltage.

bias compensator—A device that counteracts the inward bias of a pickup arm as it tracks the record. The compensator exerts an outward force on the arm and generally can be adjusted to have a definite relationship to the playing weight of the pickup.

bias current—The current through the base-emitter junction of a transistor. It is adjusted to set the operating point of the transistor.

bias distortion—1. Distortion resulting from operation on a nonlinear portion of the characteristic curve of a vacuum tube, semiconductor, or other device, due to improper biasing. 2. In teletype circuits, the uniform shifting of mark pulses from their proper position in relationship to the start pulses.

biased induction—Symbolized by B_b. The biased induction at a point in a magnetic material that is subjected simultaneously to a periodically varying magnetizing force and biasing magnetizing force is the algebraic mean of the maximum and minimum values of the magnetic induction at the point.

biased ringer—A polarized telephone bell whose armature is held at one end of its travel by a small biasing spring so it responds only to pulsating current of one polarity. Current pulses of the opposite polarity will merely attract the armature more strongly to the pole piece of the electromagnet but the bell will not ring.

bias-induced noise—The difference between bulk-erased and zero-modulation noise.

biasing magnetizing force—Symbolized by H_b. A biasing magnetizing force at a point in a magnetic material that is subjected simultaneously to a periodically varying magnetizing force and a constant magnetizing force is the algebraic mean of the maximum and minimum values of the combined magnetizing forces.

bias meter—A meter used in teletypewriter work for determining signal bias directly in percent. A positive reading indicates a marking signal bias; a negative reading indicates a spacing signal bias.

bias oscillator—An oscillator used in magnetic recorders to generate an ac signal in the range of 40 to 80 kHz for the purpose of magnetic biasing to obtain a linear recording characteristic. Usually the bias oscillator also serves as the erase oscillator.

bias port—In a fluidic device, the port at which a biasing signal is applied.

bias resistor—A resistance connected into a self-biasing vacuum-tube or semiconductor circuit to produce the voltage drop necessary to provide a desired biasing voltage.

bias-set frequency—In direct magnetic tape recording, a specified recording frequency employed during the adjustment of bias level for optimum record performance (not the frequency of the bias).

bias telegraph distortion—Distortion in which all mark pulses are lengthened (positive bias) or shortened (negative bias). It can be measured with a steady stream of unbiased reversals (square waves having equal-length mark and space pulses). The average lengthening or shortening does not give true bias distortion unless other types of distortion are negligible.

bias voltage—1. A voltage, usually dc, used to set the operating point of a circuit above or below a reference voltage. 2. The base or grid voltage that establishes a semiconductor's or vacuum tube's desired dc operating voltage. 3. A steady voltage that presets the operating threshold or operating point of a circuit or device, such as a transistor or vacuum tube.

bias windings—Control windings of a saturable reactor, by means of which the operating condition is translated by an arbitrary amount.

biax—Two-hole, orthogonal, cubical ferrite computer memory elements.

BiCMOS—Abbreviation for bipolar complementary metal oxide semiconductor. An IC technology combining the linearity and speed advantages of bipolar and the low-power advantages of CMOS on a single IC. BiCMOS can operate at either ECL (emitter-coupled-logic) or TTL (transistor-transistor-logic) levels and is ideal for mixed-signal devices. (BiCMOS may eclipse CMOS, just as CMOS edged out MOS and bipolar circuits.)

biconical antenna—An antenna that is formed by two conical conductors, having a common axis and vertex, and excited at the vertex. When the vertex angle of one of the cones is 180°, the antenna is called a discone.

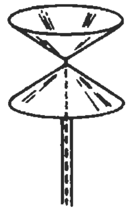

Biconical antenna.

bidirectional—1. Responsive in opposite directions. An ordinary loop antenna is bidirectional because it has maximum response from the opposite directions in the plane of the loop. 2. Refers to a type of computer bus structure in which a single conductor is used to transmit data or signals in either direction between a peripheral device and a central processor or memory. 3. In open-reel or cassette recorders, the ability to play (and, in some cases, record) both stereo track pairs on a tape by reversing the tape's direction of motion without removing and replacing the tape reels or cassette. 4. In microphones, a figure-8 pickup pattern.

bidirectional antenna—An antenna having two directions of maximum response.

bidirectional bus—1. In computers, a data path over which both input and output signals are routed. 2. In a computer, a bus that carries signals in either direction. The bus also carries special signals that tell the devices connected to it which way data is passing. 3. A bus used by any individual device for two-way transmission of messages, that is, both input and output.

bidirectional bus driver — A signal-driving device in a microcomputer that permits direct connection of a buffer-to-buffer arrangement on one end (the interface to I/O memories, etc.) and data inputs and outputs on the other. This permits bidirectional signals to pass and provides drive capability in both directions.

bidirectional current — A current that is both positive and negative.

bidirectional data bus — A data bus in which digital information can be transferred in either direction.

bidirectional diode thyristor — A two-terminal thyristor having substantially the same switching behavior in the first and third quadrants of the principal voltage-current characteristic.

bidirectional lines — Links between devices in a system that may carry information in either direction, but not both simultaneously.

bidirectional loudspeaker — A speaker that delivers sound waves to the front and rear.

bidirectional microphone — 1. A microphone in which the response predominates for sound incidences of 0° and 180°. 2. A microphone that is equally sensitive to sounds arriving from in front or in back, but discriminates against sounds arriving from the sides.

bidirectional printer — A printer that prints from left to right as well as from right to left, avoiding delays caused by carriage returns.

bidirectional pulses — Pulses, some of which rise in one direction and the remainder in the other direction.

bidirectional pulse train — A pulse train in which some pulses rise in one direction and the remainder in the other direction.

bidirectional thyristor — A thyristor that can be made conductive at any instant when the voltage between the main terminals is either positive or negative.

bidirectional transducer — *See* bilateral transducer.

bidirectional transistor — A transistor that is specified with parameter limits in both the normal and inverted configuration and has substantially the same electrical characteristics when the terminals normally designed as emitter and collector are interchanged. (Bidirectional transistors are sometimes called symmetrical transistors. The term, however, is deprecated because it might give the incorrect impression of an ideally symmetrical transistor.)

bidirectional triode thyristor — A three-terminal thyristor having substantially the same switching behavior in the first and third quadrants of the principal voltage-current characteristic.

bifet — Linear circuit that combines bipolar transistors with junction field-effect transistors on the same silicon chip and provides broader bandwidth, faster slewing, and higher impedance than standard bipolar devices when incorporated into monolithic operational amplifiers.

bifilar — A winding made noninductive by winding two wires carrying current in opposite directions together, side by side, as one wire.

bifilar resistor — A resistor wound with a wire doubled back on itself to reduce the inductance.

bifilar suspension — A type of galvanometer movement that is highly resistant to overloads, in which a D'Arsonval moving coil is supported at each end by two taut wires. The elmination of the pivot, with its attendant friction, results in superior sensitivity and precision.

bifilar transformer — A transformer in which the turns of the primary and secondary windings are wound together side by side and in the same direction. This type of winding results in near-unity coupling, so that there is a very efficient transfer of energy from primary to secondary.

bifilar winding — 1. A method of winding noninductive resistors in which the wire is folded back on itself and then wound double, with the winding starting from the point at which the wire is folded. 2. A winding consisting of two insulated conductors side by side to produce (*a*) two balanced windings, (*b*) a resistor with minimum inductance, and (*c*) maximum coupling between two windings.

bifurcate — Describes lengthwise slotting of a flat spring contact, as used in a printed-circuit connector, to provide additional independently operating points of contact. Example: bifurcated contact.

bifurcated — Usually fork-shaped. Refers to physical construction of a contact whereby two mating portions make physical contact. Yet, if one tip section of the contact fails, the remaining section maintains the physical and electrical connection.

bifurcated connector — A hermaphroditic connector containing fork-shaped mating contacts.

bifurcated contact — 1. A movable contact that is forked (divided) to provide two contact-mating surfaces in parallel for a more reliable contact. 2. A connector contact (usually a flat spring) that is slotted lengthwise to provide additional, independently operating points of contact.

bilateral — Having a voltage-current characteristic curve that is symmetrical with respect to the origin, that is, being such that if a positive voltage produces a positive current magnitude, an equal negative voltage produces a negative current of the same magnitude.

bilateral amplifier — An amplifier capable of receiving as well as transmitting signals; it is used primarily in transceivers.

bilateral antenna — An antenna, such as a loop, having maximum response in exactly opposite directions (180° apart).

bilateral bearing — A bearing that indicates two possible directions of wave arrival. One of these is the true bearing, and the other is a bearing displaced 180° from the true bearing.

bilateral circuit — A circuit wherein equipment at opposite ends is managed, operated, and maintained by different services.

bilateral element — A two-terminal element, the voltage-current characteristic of which has odd symmetry around the origin.

bilateral network — 1. A network in which a given current flow in either direction results in the same voltage drop. 2. A network that passes current and signals equally well in both directions.

bilateral transducer — 1. Also called bidirectional transducer. A transducer capable of transmission simultaneously in both directions between at least two terminations. 2. A device capable of measuring stimuli in both a positive and a negative direction from a reference zero or rest position.

billboard antenna — An antenna array consisting of several bays of staked dipoles spaced 1/4 to 3/4 wavelength apart, with a large reflector placed behind the entire assembly. The required spacing of the dipoles tends to make the array inconveniently large at frequencies below the VHF range.

bimag — *See* tape-wound core.

bimetal — A union of two dissimilar metals (especially those having a different temperature coefficient of expansion), usually welded together over their entire surface.

bimetal cold-junction compensation — Automatic mechanical correction for ambient temperature change at the cold junction of a thermocouple, which would normally cause erroneous readings.

bimetallic strip — A strip formed of two dissimilar metals welded together. Because the metals have different temperature coefficients of expansion, the strip bends or curls when the temperature changes.

bimetallic switch — A temperature-sensitive switch that uses a bimetallic element.

bimetallic thermometer — 1. A device containing a bimetallic strip that expands or contracts as the temperature changes. A calibrated scale indicates the amount of change in temperature. 2. A strip of two metals having different coefficients of expansion bonded together in the form of a spiral or helix. Movement caused by a temperature change becomes a measure of temperature.

bimetallic wire — Any wire formed of two different metals joined together (not alloyed). It can include wire with a steel core for high-strength clad wire, or plated or coated wire.

bimetal mask — A mask formed by chemically etching openings in a metal film or plate where it is not protected by photoresist or other chemically resistant material.

bimorph cell — Two crystal elements (usually Rochelle salt) in rigid combination, arranged to act as a mechanical transforcer in headphones, microphones, pickups, and speakers.

BiMOS — Abbreviation for bipolar metal oxide semiconductor. A general term to refer to bipolar and metal oxide semiconductors on one chip. Sometimes used interchangeably with *BiCMOS.*

binary — 1. A numbering system using a base number, or radix, of 2. There are two digits (1 and 0) in the binary system instead of 10 as in the decimal. 2. Pertaining to a characteristic or property involving a selection, choice, or condition in which there are two possibilities or alternatives. 3. A bistable multivibrator. 4. Two-valued logic using only the values true and false. Represented in a computer circuit by the presence of current (equivalent to 1) or its absence (equivalent to 0). All computer programs are executed in binary form. 5. A counting system in which the value of any digit can only be 1 or 0. As with the decimal system, the right-hand digit denotes the number of units of the next value, etc. In decimal the units can be 0 to 9; in binary, the units are 0 and 1.

binary arithmetic — Mathematical operations performed with only the digits 0 and 1.

binary card — A card that contains data in column binary or row binary form.

binary cell — In an electronic computer, an elementary unit of storage that can be placed in either of two stable states.

binary chain — A series of binary circuits, each of which can exist in either one of two states, arranged so each circuit can affect or modify the condition of the next circuit.

binary channel — A transmission facility limited to the use of two symbols.

binary code — A method of representing numbers in a scale of 2 (on or off, high level or low level, one or zero, presence or absence of a signal) rather than the more familiar scale of 10 used in normal arithmetic. Electronic circuits designed to work in two defined states are much simpler and more reliable than those working in ten such states.

binary-coded — Expressed by a series of binary codes (0s and 1s).

binary-coded character — A decimal digit, alphabetic letter, punctuation mark, etc., represented by a fixed number of consecutive binary digits.

binary-coded decimal — Abbreviated BCD. A coding system in which each decimal digit from 0 to 9 is represented by four binary digits:

Decimal Digit	*Binary Code*
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001

2. The representation of decimal numbers in binary form. It is useful in adc systems intended to drive decimal displays. Its advantage over decimal is that only four lines are needed to represent 10 digits. The disadvantage of coding digital-to-analog converters or analog-to-digital converters in BCD is that a full 4 bits could represent 16 digits, whereas only 10 digits are represented in BCD. 3. A numbering system used in many computers, in which the basic binary system is used to represent decimal numbers.

binary-coded digit — One element of a notation system for representing a decimal digit by a fixed number of binary positions.

binary-coded octal system — An octal numbering system in which each octal digit is represented by a three-place binary number.

binary counter — *See* binary scaler.

binary digit — 1. A character that represents one of the two digits in the number system that has a radix of 2. Also called bit. 2. Either of the digits, 0 or 1, that may be used to represent the binary conditions on or off. 3. A whole number in the binary scale of notation; this digit may be only 0 (zero) or 1 (one). It may be equivalent to an on or off condition, yes or no, etc.

binary file — A file stored in a binary format containing data or program instructions in a computer-readable format. Special software is required to display such a file.

binary incremental representation — Incremental representation in which the value of an increment is rounded to plus or minus one quantum and is represented by one binary digit.

binary magnetic core — A ring-shaped magnetic material that can be made to take either of two stable states of magnetic polarization.

binary notation — *See* binary number system.

binary number system — A number system employed in computers and digital systems, in which successive digits are coefficients of powers of the base 2, rather than the base 10. For example, the decimal number 13 is represented by the binary number $1101(1 \times 2 + 1 \times 2 + 0 \times 2 + 1 \times 2)$. Since the only values in the binary system are 0 and 1, quantities, or *bits* (binary digits), are represented electronically with either of two conditions, typically a high voltage representing a 1 and a low voltage representing a 0. Also called binary notation.

binary numeral — The binary representation of a number; for example, 101 is the binary numeral and V is the Roman numeral of the number of fingers on one hand.

binary phase-shift keying—Abbreviated BPSK. A modulation scheme that uses two phases to represent data; one phase represents a mark, and the other phase represents a space.

binary point—1. The point that marks the place between integral powers of 2 and fractional powers of 2 in a binary number. 2. The radix point in a mixed binary numeral, such as 110.011, separating the fractional part from the integer part. In the binary numeral 110.011, the binary point is between the two 0s.

binary pulse-code modulation—A form of pulse-code modulation in which the code for each element of information consists of one of two distinct kinds, e.g., pulses and spaces.

binary raster data—Computer data in an on/off format (ones and zeros) that is formed into scan lines that can be used to control a light source, such as a laser, for exposure of light-sensitive material.

binary scaler—Also called binary counter. 1. A counter that produces one output pulse for every two input pulses. 2. A counting circuit, each stage of which has two distinguishable states. 3. A flip-flop having a single input (called a T flip-flop). Each time a pulse appears at the input, the flip-flop changes state.

binary search—Also called dichotomizing search. 1.A search in which a set of items is divided into two parts; one part is rejected, and the process is repeated on the accepted part until those items with the desired property are found. 2. A search that starts in the middle of a database, first determining if the desired record is above or below the midpoint, then proceeding to the middle of the remaining records, and so on.

binary signal—A voltage or current that carries information in the form of changes between two possible values.

binary signaling—A communications mode in which information is passed by the presence and absence, or plus and minus variations, of one parameter of the signaling medium.

binary system—A system of mathematical computation based on powers of 2.

binary-to-decimal conversion—The process of converting a number written in binary notation to the equivalent number written in the ordinary decimal notation.

binary-to-hexadecimal conversion—The process of converting a numeral written in base 2 to the equivalent numeral written in base 16.

binary-to-octal conversion—The process of converting a numeral written in base 2 to the equivalent numeral written in base 8.

binary word—A related grouping of ones and zeros that has a meaning assigned by definition, or that has a weighted numerical value in the natural binary number system.

binaural—Two-channel sound in which each channel recorded is heard only through one ear. In recording, microphones are spaced to approximate the distance between a person's own ears. To hear the recording binaurally, the listener must use headphones. Compare with *stereo*.

binaural disc—A stereo record with two separate signals recorded in its grooves. Stereophonic sound is obtained by feeding each signal into its own speaker or headphone.

binaural effect—The effect that makes it possible for a person to distinguish the difference in arrival time or intensity of sound at his or her ears and thereby determine the direction from which a sound is arriving.

binaural recorder—A tape recorder that employs two separate recording channels, or systems, each with its own microphone, amplifier, recording and playback heads, and earphones. Recordings using both channels are made simultaneously on a single magnetic tape having two parallel tracks. During playback, the original sound is reproduced with depth and realism. For a true binaural effect, headphones are necessary.

binaural sound—Sound recorded or transmitted by pairs of equipment so as to give the listener the effect of having heard the original sound.

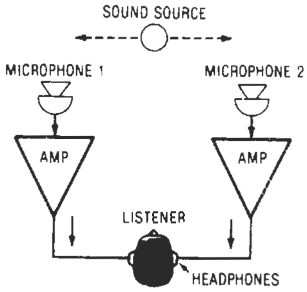

Binaural sound-reproducing system.

binder—1. A substance, such as cement, used to hold particles together and thus provide mechanical strength in, for example, carbon resistors and phonograph records. 2. Materials or substances added to thick-film compositions and unfired substrates to give sufficient strength for prefire handling.

binding—The act of assigning absolute addresses to a program.

binding energy—The minimum energy required to dissociate a nucleus into its component neutrons and protons. Neutron or proton binding energies are those energies required to remove a neutron or a proton, respectively, from a nucleus. Electron binding energy is that energy required to remove an electron from an atom or a molecule.

binding post—1. A bolt-and-nut terminal for making temporary electrical connections. 2. A device for clamping or holding electrical conductors in a rigid position.

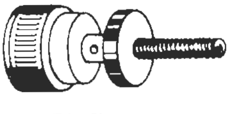

Binding post.

binistor—A four-terminal controlled rectifier semiconductor that provides bistable negative-resistance characteristics.

binomial array—A directional-antenna array used for reducing minor lobes and providing maximum response in opposite directions.

biochemical fuel cell—An electrochemical generator of electrical power in which bi-organic matter is used as the fuel source. In the usual electrochemical reaction, air serves as the oxidant at the cathode and microorganisms are used to catalyze the oxidation of the bi-organic matter at the anode.

bioelectricity — Electric currents and potential differences that occur in living tissues. Muscle and nerve tissue, for example, are generators of bioelectricity, although the potential registered may be less than 1 millivolt in some cases.

bioelectric potential — *See* bioelectricity.

bioelectrogenesis — The practical application of electricity drawn directly from the bodies of animals, including humans, to power electronic devices and appliances.

bioelectronics — 1. The application of electronic theories and techniques to the problems of biology. 2. The integrated, long-term electronic control of various impaired physiologic systems by means of small, low-power electrical and electromechanical devices. (The pacemaker is therefore a bioelectronic instrument.)

bioengineering — *See* bionics.

biogalvanic battery — A device that makes use of reactions between metals and the oxygen and fluids in the body to generate electricity.

biologic energy — Energy that is produced by bodily processes and that can be used to supply electrical energy for implanted devices such as electronic cardiac pacemakers, bladder stimulators, etc. The biologic energy can result from muscle movement (such as that of the diaphragm), temperature differences, pressure differences, expansion of the aorta, oxidation of materials within the gastrointestinal tract, or other processes.

biomedical oscilloscope — An oscilloscope designed or modified to be used in medical applications. Such oscilloscopes have slow sweep rates and long-persistence screens because of the low frequencies of many biological signals.

bionics — Also called bioengineering. 1. The study of living systems so that their characteristics and functions can be related to the development of mechanical and electronic hardware. 2. The reduction of various life processes to mathematical terms to make possible duplication or simulation with systems hardware. 3. The art that treats electronic simulation of biological phenomena. 4. The emulation of biological components, "body parts," with electromechanical ones. 5. The application of observed operational processes of sophisticated living organisms to mechanical and electrical systems in order to analogize capabilities or efficiency.

BIOS — Abbreviation for basic input-output system. The software that translates operating system instructions into commands to and from the hardware components in a computer system.

biotelemetry — The process of remote measurement or recording of such biological variables as pulse rate, temperature, etc. Typically, the information is transmitted between the patient and the receiving equipment by a radio link.

biotelescanner — A device that can analyze and radio data on life forms during space exploration.

biphase suppressed carrier — A digital radio modulation scheme that uses one phase of a carrier to signify a 0 and the opposite phase to denote a 1. If the two phases are less than 180° apart, some residual carrier will be transmitted.

bipolar — 1. Having two poles. 2. Having to do with a device in which both majority and minority carriers are present. In connection with integrated circuits, the term describes a specific type of construction; bipolar and MOS are the two most common types of IC construction. 3. The semiconductor technology employing two-junction transistors. 4. A transistor structure whose electrical properties are determined within the silicon material. Memories using this technology are characteristically high-speed devices. 5. General name for npn and pnp transistors, since working current passes through semiconductor material of both polarities (p and n). Current in the collector is controlled by a current between base and collector. Also applied to integrated circuits that use bipolar transistors. 6. In bipolar transistors, the working current consists of both positive and negative electrical charges. The first transistor and the first ICs were bipolar types. (Most present-day discrete transistors are also bipolar.) Bipolar IC transistors operate faster than unipolar (MOS) transistors and consume more power. They take up more space on a chip and cost more to manufacture. 7. In meters, the capability of measuring voltages or currents of either polarity — positive or negative — with respect to a reference point or ground. 8. One of several fundamental processes for fabricating ICs. A bipolar IC is made up of layers of silicon with differing electrical characteristics. Current flows between the layers when a voltage is applied to the junction or boundary between the layers. 9. A technology of IC fabrication that uses transistor switching elements based on majority carriers for switching and amplification. 10. Technology using transistors with both negative and positive charge carriers allowing current flow in only one direction (unlike MOS circuits). Bipolar memory is characterized by very fast access times and very high power consumption.

bipolar complementary metal oxide semiconductor — *See* BiCMOS.

bipolar device — 1. A semiconductor device in which there are both majority and minority carriers. (This is the case in all npn and pnp transistors.) 2. A current-driven electronic device with two poles. Operation relies on the flow of both electrons and holes.

bipolar electrode — An electrode without metallic connection with the current supply, one face of which acts as an anode surface and the opposite face as a cathode surface when an electric current is passed through a cell.

bipolar electrolytic capacitors — An electrolytic capacitor designed to withstand an alternating voltage and/or a reversal of the applied direct voltage. Also called nonpolar capacitor.

bipolar ferreed — A device comprising two sealed reed switches, a permanent magnet, and a winding with a semipermanent magnet. When the winding is pulsed with current in one direction, the reed contacts close; when the winding is pulsed in the opposite direction, the reed contacts open.

bipolar magnetic driving unit — A headphone or speaker unit having two magnetic poles acting directly on a flexible iron diaphragm.

bipolar memory cell — In a computer, a system comprising a storage latch, a pair of control gates, and an output gate. The control and output gates need not be part of the storage cell, but they usually are included because each latch requires both control and output gating.

bipolar pulse — 1. A pulse that has appreciable amplitude in both directions from the reference axis. 2. A current or voltage pulse that may be either positive or negative.

bipolar transistor — 1. A transistor that uses both negative and positive charge carriers. 2. An active semiconductor device formed by two pn junctions whose function is amplification of an electric current. Bipolar transistors are of two types: npn and pnp, depending on the manner in which the two pn junctions are combined. Bipolar transistors have three sections: emitter, base, and collector. Operation of a bipolar transistor depends on the migration of both electrons and holes, in contrast to field-effect transistors, in which only one polarity carrier predominates.

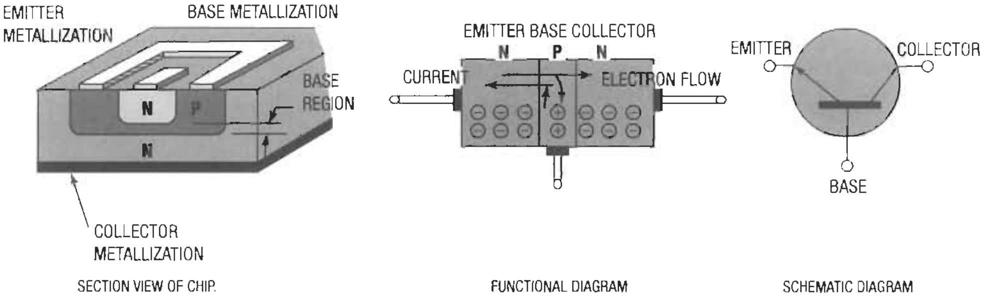

Bipolar (npn) transistor.

biquadratic filter — A filter transfer function that contains complete quadratic equations in both the numerator and denominator and provides the basis for implementing high-pass, low-pass, and single-frequency notch characteristics as well as band-reject realizations.

biquinary code — A mixed-radix notation in which each decimal digit to be represented is considered to be the sum of two digits, the first of which is 0 or 1 with significance 5 and the second of which is 0, 1, 2, 3, or 4 with significance 1.

biradial — Having an elliptical cross section. A term used with reference to phonograph styli. *See* elliptical stylus.

biradial stylus — Also called elliptical stylus. A stylus tip that has a small radius where it touches the walls of the record grooves, as distinguished from a conventional stylus, which has a hemispherical top used with lightweight pickup arms to reduce tracking distortion.

bird — Jargon or nickname for communication satellites.

birdcage — A defect in stranded wire in which the strand in the stripped portion between the covering of an insulated wire and a soldered connection (or an end-time lead) has separated from the normal lay of the strands.

bird-dogging — *See* hunting, 1.

birdnesting — Clumping together of chaff dipoles after they have been dropped from an aircraft.

Birmingham wire gage — Abbreviated BWG. The Birmingham wire gage was used extensively in Great Britain and the United States for many years, but is now obsolete. Its uses have persisted, however, for certain purposes, including galvanized steel wire for cable armor.

biscuit — *See* preform.

B-ISDN — Abbreviation for Broadband Integrated Services Digital Network. An ISDN service requiring a broadband channel operating at speeds greater than a single primary rate interface — 34 to as much as 6000 Mbps.

bislope triggering — Allows positive- and negative-going signal polarities to initiate waveform storage in both analog and digital oscilloscopes. A triggering threshold or level control on the front panel determines the trigger characteristics.

bistable — 1. A circuit element with two stable operating states, e.g., a flip-flop in which one transistor is saturated while the other is turned off. It changes state for each input pulse or trigger. 2. A device capable of assuming either one of two stable states. 3. Of or pertaining to the general class of devices that operate in either of two possible states in the presence or absence of the setting input. 4. Element that has two output possibilities and that will hold a given condition until switched.

bistable contacts — A contact combination in which the movable contact remains in its last operated position until the magnetic polarity of the coil is reversed.

bistable device — 1. Any device, such as a flip-flop, that has two stable states and may be readily switched from one state to the other. 2. A device with only two stable states, such as on and off.

bistable display — A matrix-controlled display that has information storage at the display surface and that requires that an element be addressed only once to ensure that it is on or off.

bistable latch — A rudimentary flip-flop that can be enabled to store a logical 1 or a logical 0. One bistable latch device is commonly used in memory and register circuits for the storage of each bit.

bistable multivibrator (flip-flop) — A circuit having two stable states; it will stay in either one indefinitely until appropriately triggered, after which it immediately switches to the other state.

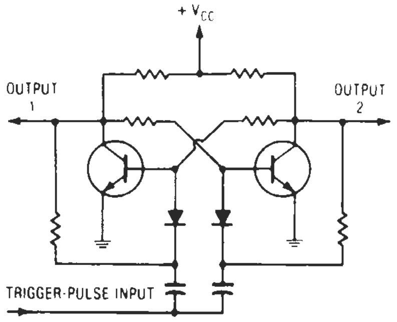

Bistable multivibrator.

bistable relay — A relay that requires two pulses to complete one cycle composed of two conditions of operation. Also called locked, interlocked, and latching relay.

bistatic radar — Radar system in which the receiver and transmitter have separate antennas and are some distance apart.

biswitch — A two-terminal integrated device that basically performs the function of two pnpn switches interconnected so as to provide bilateral switching.

bisynchronous — 1. Data transmission in which synchronization of characters is controlled by timing signals generated at both the sending and receiving stations. 2. Method of transmitting computerized data that allows for multiple error detection.

bit — Abbreviation for binary digit. 1. A unit of information equal to one binary decision, or the designation of one of two possible and equally like values or states (such as 1 or 0) of anything used to store or convey information. It may also mean "yes" or "no." 2. The smallest part of information in a binary notation system. A bit is either a one (1) or a zero (0). In a BCD system, four bits represent one decimal digit. 3. The smallest unit of information, representing either a mark or a space (1 or 0). In data transmission, the common unit of speed is bits per second (bps). 4. In a computer, one bit is enough to tell the difference between yes or no, up or down, on or off, 1 or 0, in short, any two opposites. Computers represent information in the form of bits because their circuits can have only two states: on or off. 5. The bit can take the form of a magnetized spot, an electronic impulse, a positively charged magnetic core, etc. A number of bits together are used to represent a character in a computer (byte, word).

bit copier — A program that copies individual data bits on a disk without regard to their information content; it is not affected by improperly formatted data.

bit density — The number of bits of information contained in a given area, such as the number of bits written along an inch of magnetic tape.

bit diddling — A method of increasing storage efficiency by packing extra information into unused parts of a computer word.

bit interleave — A technique in time-division multiplexing in which bits of data are transmitted in one frame.

bit map — 1. A graphics image consisting of rows and columns of dots. 2. A screen display in which each pixel location corresponds to a unique main memory location accessible by the CPU. Also refers to images intended for display on this type of digital system.

bit-mapped display — A method of CRT display that uses a separate area of computer memory to specify the locations of individual pixels, resulting in high-quality images.

bit-mapped font — A font with characters formed by a pattern of dots.

bit parallel — 1. A method of simultaneously moving or transferring all bits in a contiguous set of bits over separate wires, one wire for each bit in the set. 2. Refers to a set of concurrent data bits present on a like number of signal lines used to carry information. Bit-parallel data bits may be acted on concurrently as a group (byte) or independently as individual data bits.

bit plane — The circuitry required to make a bit-map display, especially when combinable or stackable to allow more than one bit per pixel. Thus, two bit planes give each pixel four possible colors to select from; five planes yield 32.

bit rate — 1. The number of binary bits transmitted per unit time; for example, a bit rate of 80 means that 80 binary bits are transmitted per second. 2. The rate at which binary digits, or pulses representing them, pass a given point in a communication line. 3. The rate at which data bits (digital information) are transmitted over a communication path, normally expressed in bits per second (bps). Not to be confused with the data signaling rate (baud), which measures the rate of signal changes transmitted. 4. Informal term for data rate, when it is measured in bits per second. 5. The rate of transfer of information necessary to ensure acceptable reproduction of the information at the receiver.

bit-rate generators — Devices that provide the reference frequencies required by serial interfaces (TTY, UART, cassette, modem) and also furnish adjustment-free crystal stability with easily changed, multiple frequencies.

bit-rate-length product — The product of the bit rate that a fiber or cable is able to handle and the length for tolerable dispersion at the bit rate, with the product usually stated in units of megabits per kilometer/second. Typical bit-rate-length products for graded index fibers with a numerical aperture of 0.2 is 1000 MB per km/s for research fibers and 200 MB per km/s for production fibers. The product is a good measure of fiber performance in terms of transmission capability.

bit serial — A method of sequentially moving or transferring a contiguous set of bits one at a time over a single wire, according to a fixed sequence.

bit slice — 1. As in a 4-bit slice and a 2-bit slice. A multichip microprocessor in which the control section is contained on one chip, and one or more identical arithmetic and logic unit (ALU) sections and register sections are contained on separate chips called slices. For example, three 4-bit slices connected in parallel with the control section produce a 12-bit word microprocessor. 2. An enhanced subsection of a microprocessor's ALU, embodying an architecture that permits the cascading or stacking of devices to increase word bit size. (Support chips are required to construct a functional MPU.) 3. The arithmetic and logic unit (ALU) within the central processor unit (CPU) of a microcomputer. Circuits capable of handling a few bits — normally four but sometimes less — are stacked in parallel to provide an ALU and registers capable of handling some specific word length. For example, to get the 16-bit word length normally used for numerical control, 4-bit slices can be combined, with each slice on a separate chip. The slice excludes the control section, which must be implemented by external devices. 4. A microprocessor in which the CPU is partitioned into two or more silicon chips, each containing 2- or 4-bit register logic. 5. A microcomputer chip that is a quarter or an eighth of an entire processor. When a CPU is too complex or would dissipate too much heat to be put on a single chip it is sliced into 2- or 4-bit chunks that are then wired together on circuit cards.

bit-slice processor — A microprocessor whose word (or byte) capacity is achieved though the use of interrelated smaller-capacity processors, e.g., a 16-bit unit derived from eight 2-bit slices.

bits per second — Abbreviated bps. 1. The unit of information rate. It expresses the number of binary digits passed though a channel per second. 2. The speed at which modems send and receive data. For example, a 14.4 modem moves 14,400 bits per second; a 28.8 modem, also known as a V.34, is twice as fast. An earlier term, *baud*, is considered archaic.

bit stream — 1. A binary signal without regard to grouping according to character. 2. A continuous series of bits transmitted on a communications link.

bit string — A string of binary digits in which each bit position is considered an independent unit.

bit time — 1. In a serial binary computer, the time during which each bit appears. 2. The amount of time that one bit of information in a digital pattern remains in its 1 or 0 state.

black — A signal produced at any point in a facsimile system by the scanning of a selected area of subject copy having maximum density.

black and white — *See* monochrome.

black-and-white transmission — *See* monochrome transmission.

black area — An area with only encrypted signal present.

blackbody — 1. An idealized emitter for which total radiated energy and the spectral distribution of the energy are accurately known functions of temperature. 2. A solid that radiates or absorbs energy with no internal reflection of the energy at any wavelength. Physically it may be a hollow sphere coated on the side with lampblack and with an opening through which energy may enter or leave. 3. An ideal body that absorbs all incident light and therefore appears perfectly black at all wavelengths. The radiation emitted from such a body when it is hot is called blackbody radiation. The spectral energy density of blackbody radiation is the theoretical maximum for a body in thermal equilibrium. 4. An ideal body that would absorb all radiation incident on it. When heated by external means, the spectral energy distribution of radiated energy would follow curves shown on optical spectrum charts. The ideal blackbody is a perfectly absorbing body. It reflects none of the energy that may be incident on it. It radiates (perfectly) at a rate expressed by the Stefan-Boltzmann law, and the spectral distribution of radiation is expressed by Planck's radiation formula. When in thermal equilibrium, an ideal blackbody absorbs perfectly and radiates perfectly at the same rate. The radiation will be just equal to absorption if thermal equilibrium is to be maintained.

blackbody luminous efficiency—The efficiency of an incandescent blackbody as a source of visible light. It is a function of temperature.

blackbody radiation—*See* blackbody.

black box—1. A term used loosely to refer to any subcomponent that is equipped with connects and disconnects so that it can be readily inserted into or removed from a specified place in a larger system (e.g., the complete missile or some major subdivision) without benefit of knowledge of its detailed internal structure. 2. A term pertaining to either the functional transformation that acts upon a specified input to give a particular output or to the apparatus for accomplishiug this transformation (without regard to the detailed circuitry used). 3. A useful mathematical approach to an electronic circuit which concerns itself only with the input and output and ignores the interior elements, discrete or integrated. 4. An equipment specified only in terms of its performance.

black compression—Also called black saturation. The reduction in the gain of a television picture signal at those levels corresponding to dark areas in the picture with respect to the gain at that level corresponding to the midrange light value in the picture. The overall effect of black compression is to reduce contrast in the low lights of the picture.

blacker-than-black—1. The amplitude region of the composite video signal below the reference black level in the direction of the synchronizing pulses. 2. That portion of the standard television signal devoted to the synchronizing signal.

blacker-than-black level—A voltage value used in an electronic television system for control impulses. It is greater than the value representing the black portions of the image.

black level—That level of the picture signal corresponding to the maximum limit of black peaks.

black light—1. Invisible light radiation. May be either ultraviolet or infrared radiation, both of which are invisible. 2. A lamp that produces a principal portion of its radiation in the ultraviolet region.

black-light emitter—A source of electromagnetic radiation in the ultraviolet or infrared region, just outside the visible spectrum.

black matrix—Picture tube in which the color phosphors are surrounded by black for increased contrast.

black negative—The television picture signal in which the polarity of the voltage corresponding to black is negative with respect to that which corresponds to the white area of the picture signal.

black noise—In a spectrum of electromagnetic wave frequencies, a frequency spectrum of predominantly zero power level at all frequencies except for a few narrow bands or spikes, such as might be obtained when scannning a black area in facsimile transmission systems on which there are a few white spots or speckles on the surface.

blackout—1. Interruption of radiocommunication due to excess absorption caused by solar flares. During severe blackouts, all frequencies above approximately 1500 kHz are absorbed excessively in the daylight zone. 2. Passive defense that consists of interrupting all forms of communication or identification. 3. A sudden, unexpected loss of all electrical power, typically lasting for many minutes or even hours.

black peak—A peak excursion of the picture signal in the black direction.

black-peak clipping—Limiting the amplitude of the picture signal to a preselected maximum black level, usually at blanking level.

black reference—The blanking level of pulses in a TV signal beyond which the sync pulse is in the blacker-than-black region.

black saturation—*See* black compression.

black scope—Cathode-ray tube being operated at the threshold of luminescence with no video signals applied.

black signal—Also called picture black. A signal produced at any point in a facsimile system by the scanning of a maximum density area of the subject copy.

black transmission—1. In an amplitude-modulation facsimile system, a form of transmission in which the maximum transmitted power corresponds to the maximum density of the copy. 2. In a frequency-modulation system, a form of transmission in which the lowest transmitted frequency corresponds to the maximum density of the copy.

blade contact—A flat male contact designed to mate with a tuning fork or a flat-formed female contact. It is used in multiple-contact connectors.

blank—1. The result of the final operation on a crystal. 2. To cut off the electron beam of a cathode-ray tube. 3. A code character to denote the presence of no information rather than the absence of information. In the Baudot code, it is composed of all spacing pulses. In paper tape, it is represented by a feed hole without intelligence holes. 4. A no-information condition in a data-recording medium or storage location. This vacancy can be represented by all spaces or all zeros, depending on the medium.

blank coil—Tape for perforation in which only the feed holes have been punched.

blank deleter—A device used to eliminate the receiving of blanks in perforated paper tape.

blanked picture signal—The signal resulting from adding blanking to a picture signal. Adding the sync signal to the blanked picture signal forms the composite picture signal.

blanketing—The overriding of a signal by a more powerful one or by interference, so that a receiver is unable to receive the desired signal.

blank groove—*See* unmodulated groove.

blanking—1. The process of making a channel or device noneffective for the desired interval. In television, blanking is the substitution, during prescribed intervals, of the picture signal by a signal whose instantaneous amplitude is such as to make the return trace invisible. *See also* gating. 2. The process whereby the beam in an image pickup tube or CRT is cut off during the retrace period.

blanking level—Also called pedestal level. 1. In a composite picture signal, the level that separates the range of the composite picture signal containing picture information from the range containing synchronizing information. 2. Usually referred to as the front porch or back porch. At 0 IRE units, it is the level that will shut off the picture tube, resulting in the blackest possible picture.

blanking pulse—A square wave (positive or negative) used to switch off electronically a part of a television or radar set for a predetermined length of time.

blanking signal—A wave made up of recurrent pulses related in time to the scanning process and used to effect blanking. In television, this signal is composed of pulses at line and field frequencies, which usually originate in a central sync generator and are combined with the picture signal at the pickup equipment in order to form the blanked picture signal. The addition of a sync signal completes the composite picture signal.

blanking time—The length of time the electron beam of a cathode-ray tube is cut off.

blanking zone—*See* blanking pulse.

blank instruction—*See* no-operation instruction.

blank record—A recording disk on which no material has been recorded.

blank tape—Also called raw tape or virgin tape. 1. Tape on which nothing has been recorded. 2. Magnetic tape that has never been subjected to the recording process and is therefore substantially free from noise.

blast filter—Also called a pop filter. A dense mesh screen on a microphone, which minimizes overload caused by loud, close sounds.

blasting—1. Overloading of an amplifier or speaker, resulting in severe distortion of loud sounds. 2. Severe audible distortion due to overloading of sound-reproducing equipment.

bleeder—A resistor connected across a power source to improve voltage regulation, provide a current path under no-load conditions, or dissipate stored energy on shut-off.

bleeder current—The current drawn continuously from a power supply by a resistor. Used to improve the voltage regulation of the power supply. (A technology no longer in use.)

bleeder resistor—1. A resistor used to draw a fixed current. Also used, as a safety measure, to discharge a filter capacitor after the circuit is deenergized. 2. A resistor placed in the power supply of a radio receiver or other electronic device to stabilize the voltage supply.

bleeding—1. Migration of plasticizers, waxes, or similar materials to the surface to form a film or bead. 2. In photomasking, poor edge definition or acuity caused by spread of image onto adjacent areas. 3. A condition in which a plated hole discharges process material or solution from crevices or voids. 5. During hybrid circuit manufacturing, the lateral spreading or diffusion of a printed film into adjacent areas beyond the geometric dimensions of the printing screen. This may occur during drying or firing.

bleeding whites—An overloading condition in which white areas in a television picture appear to flow into the black areas.

bleedout—The tendency of absorbed electrolytes, impurities, base materials, and preplates to diffuse to the surface of gold plating.

blemish—1. On the storage surface of a charge-storage tube, an imperfection that produces a spurious output. 2. An area in a fiber or fiber bundle that has a reduced light transmission capability, i.e., increased attenuation, due to defective or broken fibers, foreign substances, or other spoilage.

blended data—Q-point that results from the combination of scanning data and tracking data to form a vector.

blending—A means of obtaining intermediate viscosities from materials of the same type but different viscosities. This term is also applied to resistive inks that can be blended with each other to achieve intermediate resistivities.

blind approach—An aircraft landing approach when visibility is poor, usually made with the aid of instruments and radiocommunication.

blind approach beacon system—*See* BABS

blind landing—Landing an aircraft entirely by means of instruments and electronic communications.

blind zone—An area from which echoes cannot be received; generally, an area shielded from the transmitter by some natural obstruction and therefore from which there can be no return.

blinking—1. An ECM technique by which two aircraft separated a short distance and within the same azimuth resolution appear as one target to a tracking radar. The two aircraft alternately spot jam, causing the radar system to oscillate from one target to the other and making it impossible to obtain an accurate solution of the fire-control problem. 2. In pulse systems, a method of providing information in which the signal is modified at its source so that the presentation on the display scope alternately appears and disappears. In loran, this indicates that a station is malfunctioning.

blip—Sometimes referred to as pip. 1. On a cathode-ray display, a spot of light or a base-line irregularity representing the radar reflection from an object. 2. A discontinuity in the insulation of a wire.

blip-frame ratio—The ratio of the number of computer frames during which radar data was obtained to the total number of computer frames.

blip-scan ratio—The ratio between a single recognizable blip on a radarscope and the number of scans necessary to produce it. The blip-scan ratio of any given radar set varies with the range, antenna tilt, level of operator and set performance, target aspect, wind, etc.

blister—1. The enclosure housing an airborne radar antenna. 2. A lump or raised section of a conductor or resistor caused by outgassing of the binder or vehicle during firing.

blistering—The development, during firing, of enclosed or broken macroscopic vesicles or bubbles in a body or in a glaze or other coating.

blivet—An excess of coating material, such as a lump around a dust particle on a wire or a surface. *See also* land, 2.

Bloch wall—The transition layer separating adjacent ferromagnetic domains.

block—1. A group of computer words considered as a unit because they are in successive storage locations. 2. The set of locations or tape positions in which a block of words is stored. 3. A circuit assemblage that functions as a unit, such as a circuit building block of standard design or the logic block in a sequential circuit. 4. A set of contiguous bits and/or bytes that make up a definable quantity of information. 5. A section of information recorded on magnetic tape or disk. One block may consist of several records, that is, collections of information consisting of one or more related items; or a record may extend over several blocks, depending on the characteristics of the device and the needs of the programmer. 6. A group of consecutive words, characters, or bits that is handled as a single unit, particularly with respect to input/output operations. 7. A group of characters that is written or read as a physical unit as distinct from a logical unit (*see* record). A block may contain one or more complete records, or part of a record. 8. A string of data elements that is recorded or transmitted as a unit. 9. In word processing, a selected section of characters. In data management, a group of records. In communications, a fixed batch of data that is transferred together.

block address — A method of identifying words through use of an address that specifies the format and meaning of the words in the block of information.

block cancel character — A character used to signify that the preceding portion of the block is to be disregarded. Also called block ignore character.

block code — A special code or character used to separate blocks of data. A block code is used typically on paper tape and generally occurs at both the beginning and end of a block. Thus, the information on a paper tape containing a number of blocks would be started by a block code, there would be a block code between adjacent blocks, and the data would be ended by a block code.

block diagram — 1. A diagram in which the essential systems units are drawn as blocks, and their relationship to each other is indicated by appropriately connected lines. The path of the signal or energy may be indicated by lines or arrows. 2. In computer programming, a graphical representation of the data-processing procedures within the system. It is used by programmers as an aid to program development. 3. A diagram in which a system or computer program is represented by annotated boxes and interconnecting lines. Synonym: flowchart. 4. A chart that graphically depicts the functional relationships of hardware making up a system. The block diagram serves to indicate the various data and control signal paths between functional units of the system hardware. 5. A drawing in which circuit functions are represented as blocks of various geometries.

block downconversion — The process of lowering an entire band of frequencies in one step to some intermediate range to be processed by a receiver. Multiple block downconversion receivers are capable of independently selecting channels because each can process the entire block of signals.

block downconverter — 1. A device that converts an entire band (e.g., the 3.7–4.2 GHz C-band) down to a lower band of frequencies. 2. A type of downconverter that changes the microwave signal into an IF frequency that contains all the transponder frequencies (channels) of the satellite. The block downconverter allows inexpensive multiple receivers to tune all the channels simultaneously using one central downconverter — an advantage when using multiple receivers with a single antenna.

blocked impedance — The input impedance of a transducer when its output is connected to a load of infinite impedance.

blocked resistance — Resistance of an audio-frequency transducer when its moving elements are restrained so they cannot move; it represents the resistance due only to electrical loss.

blockette — In digital computer programming, a subgroup, or subdivision, of a group of consecutive machine words transferred as a unit.

block gap — 1. An area used to indicate the end of a block or record on a data medium. 2. An absence of data along a specified length of magnetic tape between adjacent blocks of data.

block-grid keying — A method of keying a continuous-wave transmitter by operating the amplifier stage as an electronic switch. During the spacing interval when the key is open, the bias on the control grid becomes highly negative and prevents the flow of plate current so that the tube has no output. During the marking interval when the key is closed, this bias is removed and full plate current flows.

block ignore character — *See* block cancel character.

blocking — 1. Application of an extremely high-bias voltage to a transistor, vacuum tube, or metallic rectifier to prevent current from flowing in the forward direction. 2. Combining two or more records into one block. 3. A condition in a switching system in which no paths or circuits are available to complete a call, resulting in a busy tone returned to the calling party. A denial or busy condition.

blocking capacitor — 1. A capacitor that introduces a comparatively high series impedance for limiting the flow of low-frequency alternating and direct current without materially affecting the flow of high-frequency alternating current. 2. A capacitor used to block direct current while allowing an alternating current of certain frequencies to pass.

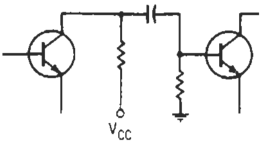

Blocking capacitor.

blocking layer — *See* depletion layer.

blocking oscillator — Also called squegging oscillator. 1. An electron-tube oscillator that operates intermittently as its grid bias increases during oscillation to a point

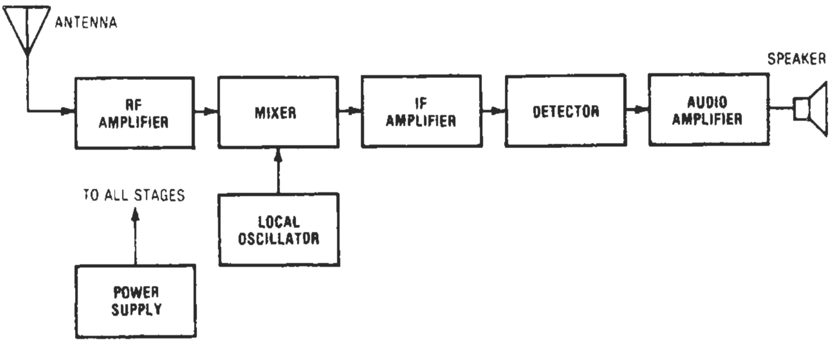

Block diagram, 1.

at which the oscillations stop, and then decreases until oscillation resumes. 2. A relaxation oscillator consisting of an amplifier (usually a single stage) whose output is coupled back to the input by means that include capacitance, resistance, and mutual inductance. 3. A relaxation-type oscillator that conducts for a short period of time and is cut off for a relatively long period.

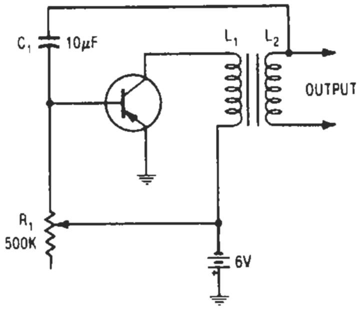

Blocking oscillator.

blocking-oscillator driver — A circuit that develops a square pulse used to drive the modulator tubes, and which usually contains a line-controlled blocking oscillator that shapes the pulse into the square wave.

blocking relay — A device that initiates a pilot signal to block tripping on external faults in a transmission line or in other apparatus under predetermined conditions, or which cooperates with other devices to block tripping or to block reclosing on an out-of-step condition or on power swings.

block length — In a computer, the total number of records, words, or characters contained in one block.

block loading — In a computer, a form of fetch in which the control sections of a load module are brought into continuous positions of main storage.

block mark — A method of indicating the end of one block of data and the start of another on tape or in data transmission. On magnetic tape, the block mark is a block gap; on paper tape it is a block code; in data transmission it is typically a pause or a code.

block-multiplexer channel — A computer-peripheral multiplexer channel that interleaves blocks of data.

block protector — A rectangular piece of carbon, Bakelite with a metal insert, or porcelain with a carbon insert that, in combination with each other, make one element of a protector. They form a gap that will break down and provide a path to ground for excessive voltages.

block size — The number of data elements in a block.

block sort — A computer sorting technique in which the file is first divided according to the most significant character of the key, and the separate portions are then sorted one at a time. It is used particularly for large files.

block transfer — In a computer, the process of transmitting one or more blocks of data.

blooming — 1. An increase in the size of the scanning spot on a cathode-ray tube, caused by defocusing when the brightness control is set too high. The result is expansion and consequent distortion of the image. May also be caused by insufficient high voltage. 2. The defocusing of regions of the picture where the brightness is at an excessive level, due to enlargement of spot size and halation of the fluorescent screen of the cathode-ray picture tube.

blooper — A radio receiver that is oscillating and radiating an undesired signal.

blow — The opening of a circuit because of excessive current, particularly when the current is heavy and a melting or breakdown point is reached and a fuse blows.

blower — 1. An electric fan used to supply moving air for cooling purposes. *See* fan. 2. A high-pressure device in which a rotating impeller moves an air mass in a spiral direction around the shaft. There are single-stage, dual-stage, and vacuum types.

blown jacket — The common term given to an outer covering of insulation of a cable that was applied by the controlled inflation of the cured jacket tube and the pulling of the cable through it.

blowout coil — An electromagnetic device used to establish a magnetic field in the space where an electrical circuit is broken and thus displace and extinguish the arc.

blowout magnet — A strong permanent magnet or electromagnet used for reducing or deflecting the arc between electrodes or contacts.

blue-beam magnet — A small permanent magnet used to adjust the static convergence of the electron beam for blue phosphor dots in a three-gun color picture tube.

blue glow — The glow normally seen in vacuum tubes containing mercury vapor; it is due to ionization of the molecules of mercury vapor.

blue gun — In a three-gun color picture tube, the electron gun whose beam strikes the phosphor dots emitting the blue primary color.

blue noise — In a spectrum of electromagnetic wave frequencies, a region in which the spectral density is proportional to the frequency (sloped) rather than independent of frequency (flat), as in white noise that is more of a uniformly distributed constant-amplitude frequency spectrum.

blue restorer — The dc restorer in the blue channel of a three-gun color-television picture-tube circuit.

blue video voltage — The signal voltage that controls the grid of the blue gun in a three-gun picture tube. This signal is a reproduction of the blue output signal of the color-television camera.

B-MAC — A method of transmitting and scrambling television signals. In such transmissions, MAC (multiplexed analog component) signals are time-multiplexed with a digital burst containing digitized sound, video synchronizing, authorization, and information.

BMEWS — *See* ballistic-missile early-warning system.

BNC — A bayonet-locking connector for miniature coax; BNC is said to be short for bayonet-neil-connector. Contrast with *TNC*.

BNC coax connector — A twist-lock connector for various types of RG-type coaxial cables.

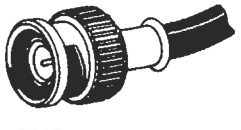

BNC coax connector.

board — 1. A telephone or audio mixer panel containing patch jacks. 2. A circuit board.

board computer — A computer in which all electronic components are on a single circuit board.

board-test simulation — A testing technique in which the circuit to be tested is modeled, component by component and node by node, in the test system computer. From this model the system can calculate the correct response to any input pattern, plus predict failure modes and their responses. This allows only those patterns that identify faults to be used as the test pattern stimulus.

boat — 1. A container for materials to be evaporated or fired. 2. A wafer holder used in a diffusion furnace.

BOB — Abbreviation for break-out box.

bobbin — 1. A small insulated spool that serves as a support for a coil or wirewound resistor. 2. Spool used for taking up drawn wire and subsequently used for pay-out packages in cabling and stranding equipment.

bobbin core — *See* tape-wound core.

body capacitance — The capacitance to ground that is introduced into an electric circuit by the proximity of, or contact with, the human body.

body effect — Characteristic shift in threshold voltage resulting from bias applied to a semiconductor device's substrate.

body electrodes — Electrodes placed on or in the body to couple electrical impulses from the body to an external measuring or recording device.

boffle — A speaker enclosure, developed by H. A. Hartley, containing a group of stretched, resilient and sound-absorbing screens.

bogey — 1. The average, or published, value for a tube characteristic. A bogey tube would be one having all characteristics of a bogey value. 2. An average, published, or nominal value for some characteristic of a device.

bogey electron device — An electron device whose characteristics have the published nominal values for the type. A bogey electronic device for any particular application can be obtained by considering only those characteristics that are directly related to the application.

boilerplate — 1. A full-size model that simulates the weight, size, and shape, but not all of the functional features, of the actual item. 2. That part of the specifications of a component, piece of equipment, system, or the like that defines and describes the set of conditions of the sale.

boiling point — The temperature at which a liquid vaporizes when heated. The exact point depends on the absolute pressure at the liquid-vapor surface.

bolometer — 1. A radiation detector that converts incident radiation into heat, which, in turn, causes a temperature change in the material used in the detector. This change is then measured to give an indication of the amount of incident radiant energy. 2. A very sensitive thermometric instrument used for the detection and measurement of radiant energy. Its essential component is a short, narrow strip that is covered with a dead-black absorbing coating. It is mounted at the lower end of a long, cylindrical tube having a stop across it to exclude unwanted radiation. The electrical resistance of the strip changes with changes in temperature that arise from absorbing varying amounts of radiant energy.

bolted fault level — *See* adjusted circuit.

Boltzmann's constant — 1.38×10^{-16} J/K. Relates the average energy of a molecule to its absolute temperature.

bomb — 1. A computer program that fails spectacularly. 2. A programmer can bomb a computer system by deliberately writing a program that will disrupt the system. 3. In a computer, to fail or to crash.

bombardment — 1. The directing of high-speed electrons at an electrode, causing secondary emission of electrons, fluorescence, disintegration, or the production of X-rays. 2. The process of directing high-speed particles at atoms to cause ionization or transmutation.

bond — 1. Electrical interconnection made with a low-resistance material between a chassis, metal shield cans, or cable shielding braid, in order to eliminate undesirable interaction and interference resulting from high-impedance paths between them. 2. To make an electrical bond, an interconnection that performs a permanent electrical and/or mechanical function. 3. A low-resistance electrical connection between two ground connections or between similar parts of two circuits. *See* valence bond.

bondability — Those surface characteristics and conditions of cleanliness of a bonding area that must exist in order to provide a capability for successfully bonding an interconnection material by one of several methods, such as ultrasonic or thermocompression wire bonding.

bondable wire — An insulated wire whose surface has been treated to facilitate adherence to other materials, such as potting compounds. The term also could be applied to magnetic wires used in making coils where bonding the turns together is desirable.

bond deformation — The change in the form of the lead produced by the bonding tool, causing plastic flow, in making the bond.

bonded assembly — An assembly whose supporting frame and metallic noncircuit elements are connected so as to be electrically shorted together.

bonded-barrier transistor — A transistor made by alloying the base with the alloying material on the end of a wire.

bonded cables — Cables consisting of preinsulated conductors or multiconductor components that are laid in parallel and bonded into a flat cable.

bonded nr diode — An n-junction semiconductor device in which the negative resistance arises from a combination of avalanche breakdown and conductivity modulation due to the current through the junction.

bonded pickup — *See* bonded transducer.

bonded strain gage — A pressure transducer that uses a pressure-sensing system consisting of strain-gage elements firmly bonded to a pressure-responsive member. Thermal stability and insensitivity to shock and vibration are improved by means of this bonded construction.

bonded transducer — Also called bonded pickup. A transducer that employs the bonded strain-gage principle of transduction.

bonding — 1. Soldering or welding together various elements, shields, or housings of a device to prevent potential differences and possible interference. 2. A method used to produce good electrical contact between metallic parts of any device. Used extensively in automobiles and aircraft to prevent static buildup. Also refers to the connectors and straps used to bond equipment. 3. The means employed to obtain an electromagnetically homogenous mass having an equipotential surface. 4. The attachment of wire to a circuit. 5. The permanent joining of metallic parts to form an electrically conductive part. *See* ball bonding; die bonding; stitch bond; thermal compression bonding; wedge bonding; wire bond; wobble bond. 6. The process of connecting wires from the package leads to the chip (or die) bonding pads. Part of the assembly process. Alternately, the process of securing a semiconductor die to a lead frame or package.

bonding area — The area, defined by the extent of a metallization land or the top surface of the terminal, to which a lead is or is to be bonded. Also called bond site.

bonding conductor — 1. A conductor that serves to connect exposed metal surfaces together. 2. Device used to connect exposed metal to ground. It normally carries no current but is used to eliminate shock or spark hazards and ensures the operation of circuit protective devices in cases of insulation breakdown.

bonding island — *See* bonding pad.

bonding pad — Also called bonding island. 1. A relatively large metallic area at the edge of an integrated circuit chip; this area is connected through a thin metallic strip to some specific circuit point to which an external connection is to be made. 2. A metallized area at the end of a thin metallic strip to which a connection is to be made. *See also* beam lead.

bonding wire — Fine gold or aluminum wire for making electrical connections to a semiconductor die or a hybrid circuit between various bonding pads on the semiconductor device substrate and device terminals or substrate lands.

bond liftoff — The failure mode whereby the bonded lead separates from the surface to which it was bonded.

bond schedule — The values of the bonding machine parameters used when adjusting for bonding. For example, in ultrasonic bonding, the values of the bonding force, time, and ultrasonic power.

bond site — *See* bonding area.

bond strength — A measure of the amount of stress required to separate a layer of material from the base to which it is bonded. Peel strength, measured in pounds per inch of width, is obtained by peeling the layer; pull strength, measured in pounds per square inch, is obtained by a perpendicular pull applied to a surface of the layer.

bond-to-bond distance — The distance, measured from the bonding site on the die to the bond impression on the post, substrate land, or fingers, that must be bridged by a bonding wire or ribbon.

bond-to-chip distance — In beam-lead bonding, the distance from the heel of the bond to the component.

bone conduction — The process by which sound is conducted to the inner ear through the cranial bones.

book capacitor — A two-plate trimmer capacitor that has its plates hinged together like the pages of a book. The capacitance is varied by changing the angle between the plates.

Boolean algebra — 1. A system of mathematical logic dealing with classes, propositions, on-off circuit elements, etc., associated by operators such as AND, OR, NOT, EXCEPT, IF ... THEN, etc., thereby permitting computations and demonstration as in any mathematical system. Named after George Boole, famous English mathematician and logician, who introduced it in 1847. 2. Algebraic rules for manipulating logic equations. 3. Shorthand notation for expressing logic functions.

Boolean calculus — Boolean algebra modified to include time.

Boolean equation — Expression of relations between logic functions.

Boolean function — A mathematical function in Boolean algebra.

Boolean operator — The symbol or word used to specify the inclusion or exclusion of criteria such as AND, OR, NOT, etc. Same as logical operator.

boom — A mechanical support for a microphone, used in a television studio to suspend the microphone within range of the actors' voices but out of camera range.

boost capacitor — A capacitor used in the damper circuit of a television receiver to supply a boosted B voltage.

boost charge — The partial charge of a storage battery, usually at a high current rate for a short period.

boosted B voltage — In television receivers, the voltage resulting from the combination of the B-plus voltage from the power supply and the average value of voltage pulses coming through the damper tube from the horizontal deflection-coil circuit. The pulses are partially or wholly smoothed by filtering. This boosted voltage may be several hundred volts higher than the B-plus voltage.

booster — 1. A carrier-frequency amplifier, usually a self-contained unit, connected between the antenna or transmission line and a television or radio receiver. 2. An intermediate radio or TV station that retransmits signals from one fixed station to another. 3. A small, self-contained transformer designed to be connected to a cathode-ray tube socket to increase the filament voltage and thereby extend the life of the tube. 4. A device inserted into a line (or cable) to increase the voltage. Boosting generators are also used to raise the level of a dc line. Transformers are usually employed to boost ac voltages. The term booster also is applied to antenna preamplifiers.

booster amplifier — A circuit used to increase the output current of the voltage capabilities of an operational amplifier circuit without loss of accuracy (ideally) or inversion of polarity. Usually employed inside the loop for accuracy.

booster voltage — The additional voltage supplied by the damper tube of a television receiver to the horizontal-output, horizontal-oscillator, and vertical-output tubes, resulting in a greater sawtooth sweep output.

boot — 1. A form placed around the wire termination of a multiple-contact connector for the purpose of containing the liquid potting compound until it hardens. 2. A housing, usually made from a resilient material, used to protect connector or other terminals from moisture. 3. An accessory, usually of a flexible material, designed to be placed around the terminals of a component as a protective housing. 4. To make a computer operate by loading in a program. 5. The initial starting-up of a PC. The operating system is brought into main memory and takes over control. 6. The automatic routine that clears the memory, loads the operating system, and prepares a computer for use. A *cold boot* occurs when the power is first switched on. A *warm boot* refers to restarting a computer that is already turned on (clearing the memory and reloading the operating system) without first switching it off. Also called boot up.

booting — Loading a computer's memory with the necessary information so it can function. A cold boot occurs when a computer's power is turned on and there is nothing in memory.

boot loader — Also called bootstrap. A program in a minicomputer that usually works on a simple data format called core image. The data format, in this context, is the organization of the data as it appears on the input device from which the program is being loaded. Core image data is binary, bit-for-bit identical to what will appear in memory after loadings. The boot leader is used to bring simple programs into memory and run them immediately.

bootstrap — 1. A technique or device that brings itself into a desired state through its own action; for example, a routine whose first few instructions are sufficient to cause the rest of the routine to be brought into the computer from an input device. 2. A feedback technique that tends to improve linearity and input impedance of circuits operating over a wide range of input signals. *See* boot loader.

bootstrap circuit — A single-stage amplifier in which the output load is connected between the negative end of the plate supply and the cathode, the signal voltage being applied between the grid and cathode. The name *bootstrap* arises from the fact that the change in grid voltage also changes the potential of the input source (with respect to ground) by an amount equal to the output signal.

bootstrap driver — An electronic circuit used to generate a square pulse to drive a modulator tube.

bootstrap loader — Device for loading first instructions (usually only a few words) of a routine into memory, then using these instructions to bring in the rest of the routine.

bootstrapping — A feedback technique that tends to improve the linearity of circuits that operate over a wide input-signal range.

boresight — The direction along the principle axis of either a transmitting or a receiving antenna.

boresight error — The angular deviation of the electrical boresight of an antenna from its reference boresight.

boresighting — The initial alignment of a directional microwave or radar antenna system through use of an optical procedure or a fixed, known target.

boresight point — The area of maximum signal strength of a down-link signal from a satellite. The center of the transponder footprint.

BORSCHT — Abbreviation for Battery, Overvoltage, Ringing, Supervision, Coding, Hybrid, and Test. 1. An acronym for the function that must be performed in the central office of a telephone system when digital voice transmission occurs. 2. The seven functions (battery feed, overvoltage protection, ringing, supervision (or signaling), coding, hybrid, and testing) that are used in telecommunications and performed by a subscriber line interface circuit (SLIC).

boss — *See* land, 2.

BOT — Abbreviation for beginning of tape. *See* load point.

BOT/EOT markers — The reflective markers on the back (nonoxide) side of a magnetic tape, used to locate the beginning of the data (BOT) and provide an early warning of the end of the tape (EOT). Tape-drive sensors optically detect these markers.

bottom — To reach a point on an operating or characteristic curve at which a negative change in the independent variable, as, for example, in input, no longer produces a constant change in the dependent variable, as, for example, output.

bottoming — 1. A condition in which a stylus reaches the bottom of a record groove because its tip radius is smaller than optimum for the groove. Also the opposite of the pinch effect. 2. Excessive movement of the diaphragm of a headphone or the cone of a speaker so that the magnet or supporting structure is struck by the moving coil-piston assembly.

bottom metallization — The metallization that may be provided over the back portion of an uncased IC chip, facilitating its face-up attachment.

bounce — 1. An unnatural, sudden variation in the brightness of a television picture. 2. Sudden variations in a video picture presentation (brightness, size, etc.) independent of scene illumination.

bounce buffer — The electronic circuitry used to eliminate the effects of bounce of a mechanical switch.

boundary — An interface between p and n material at which donor and acceptor concentrations are equal.

boundary defect — In a crystal, the boundary area between two adjacent perfect crystal regions that are tilted slightly with respect to each other.

boundary-element method — A procedure for solving electromagnetic field problems by breaking the domains into smaller segments and by using integral equations in which the solution variables are confined within the boundaries.

boundary-layer photocell — *See* photovoltaic cell.

boundary marker — A transmitting device installed near the approach end of an airport runway and approximately on the localizer course line.

bound charge — On a conductor, the charge which, owing to the inductive action of a neighboring charge, will not escape to ground; residual charge.

bound circuit — A circuit designed to limit the excursion of a signal. The limit value it establishes may be nominal when used for protection, or highly precise when used operationally.

bound electron — An electron bound to the nucleus of an atom by electrostatic attraction.

bow-tie antenna — An antenna generally used for UHF reception. It consists of two triangles in the same plane, usually with a reflector behind them. The transmission line is connected to the points, which form a gap.

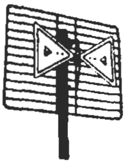

Bow-tie antenna.

boxcar — One of a series of pulses having long duration in comparison with the spaces between them.

boxcar circuit — A circuit used in radar to sample voltage waveforms and store the latest value sampled.

boxcar detector — A signal recovery instrument that is used either to retrieve the waveform of a repetitive signal from noise or to measure the amplitude of a repetitive pulse buried in noise. The detector has two modes of operation: scan and single-point. The former is used for waveform retrieval and the latter for pulse measurement.

boxcar integrator — A signal processor that uses a narrow filter to reduce the noise with little or no effect on the signal bandwidth. A simple integrator consists of a gated switch and a low-pass filter. During the time when the gate is closed, the repetitive input signal is applied to the low-pass filter, which acts as an integrator.

boxcar lengthener — A circuit that lengthens a series of pulses without changing their heights.

box pattern — A pin arrangement for plug-in packages in which the pins form four rows in either a square or rectangle.

bpi/cpi — Bytes per inch (bits per inch per track), the number of bytes per inch written on a magnetic tape, and cpi (characters per inch) are used interchangeably.

B-plus — 1. Symbol, B+. The positive dc voltage required for certain electrodes of tubes, transistors, etc. 2. The positive terminal of a B power supply.

B-plus boost — The positive voltage that is added to the low dc B+ voltage in a TV receiver by the action of the damper tube.

B power supply — A power supply that provides the plate and screen voltages applied to a vacuum tube.

bps — Abbreviation for bits per second. 1. In serial data transmission, the instantaneous bit speed within one character as transmitted. 2. A measurement of the speed at which data is transmitted and received by a modem. The larger the number, the faster the data is sent and received. Typical rates are 2400, 14,400, 28,800, and 56,000 bps. Often confused with *baud*, although the terms are not interchangeable.

BPSK — *See* binary phase-shift keying.

B-quad — A quad arrangement similar to the S-quad except for a short between the junction of the two sets of series elements. Also called bridge quad or bar quad.

Bragg's law—An expression of the conditions under which a system of parallel atomic layers in a crystal will reflect an X-ray beam with maximum intensity.

braid—1. A weave of organic or inorganic fiber used as a covering for a conductor or group of conductors. 2. A woven metal tube used as shielding around a conductor or group of conductors. When flattened, it is used as a grounding strap.

braided wire—1. A flexible wire made up of small strands woven together. 2. Woven bare or tinned copper wire used as shielding for wires and cables and as ground wire for batteries or heavy industrial equipment. There are many different types of construction.

brain waves—The patterns of lines produced on the moving chart of an electroencephalograph as the result of electrical potentials produced by the brain, picked up by electrodes, and amplified in the machine.

brake—An electromechanical friction device to stop and hold a load.

brake wire—Wires used in automotive and truck trailers to supply current to the electrical braking system.

braking magnet—*See* retarding magnet.

branch—1. In an electronic network, a section between two adjacent branch points. 2. A portion of a network consisting of one or more two-terminal elements in series. 3. An instruction to a computer to follow one of several courses of action, depending on the nature of control events that occur later.

branch circuit—1. That portion of the wiring system between the final overcurrent device protecting the circuit and the outlet. 2. (As applied to appliances) A circuit designed for the sole purpose of supplying an appliance or appliances; nothing else can be connected to this circuit, including lighting. 3. (General purpose) A circuit that supplies lighting and appliances. 4. (Individual) A circuit that supplies just one piece of equipment, such as a motor, an air conditioner, or a furnace. 5. (Multiwire) A circuit consisting of two or more underground conductors with a potential difference between them, and a grounded conductor with an equal potential between it and any one ungrounded conductor. (This may be a 120/240-volt three-wire circuit or a wye-connected circuit with two or more phase wires and a neutral.) It is not a circuit using two or more wires, connected to the same phase, and the neutral.

branch current—The current in the branch of a network.

branch impedance—In a passive branch, the impedance obtained by assuming a driving force across and a corresponding response in the branch, no other branch being electrically connected to the one under consideration.

branching—1. In a computer, a method of selecting, on the basis of the computer results, the next operation to execute while the program is in progress. 2. The function of a computer that alters the logic path depending on some detected condition or data status. For example, the program would branch to a recorder routine when the projected available balance goes negative.

branching instructions—In a computer, conditional and unconditional jumping, calling subroutines, and return-looping operations.

branch order—An instruction used to link subroutines into the main program of a computer.

branch point—1. In an electric network, the junction of more than two conductors. *See also* node, 1. 2. In a computer, a point in the routine where one of two or more choices is selected under control of a routine.

branch voltage—The voltage across a branch of a network.

branch windings—Forked polyphase transformer windings.

brass pounder—Amateur term for a Morse code operator, especially one who spends long hours handling traffic.

braze bonding—The joining of similar or dissimilar metals by introducing a braze filler metal at the joint and establishing a conventional brazed joint, and then diffusing the braze filler into the base metals by subsequent heat treating. Melting occurs in the braze filler independent of the base metals. Base-metal fusion is not required.

brazing—1. A group of welding processes in which the filler is a nonferrous metal or alloy with a melting point greater than 1000°F (537°C) but lower than that of the metals or alloy to be joined. Brazing is sometimes referred to as hard soldering. 2. Joining two metal parts (usually made of iron or steel) together with a suitable melted copper-alloy metal.

breadboard—1. Developmental or prototype version of a circuit. Solderless sockets and bus strips in modular form are often used to create expandable matrices for placement of ICs, capacitors, resistors, and so forth. 2. Perforated substrates that facilitate trial positioning of circuit components and wiring arrangements leading to final circuit construction and packaging. They are used in design, construction, and assembly.

breadboard circuit—A circuit simulation using discrete components or partially integrated components to prove the feasibility of a circuit.

breadboard construction—An arrangement in which electronic components are fastened temporarily to a board for experimental work.

breadboard model—1. An assembly in rough form to prove the feasibility of a circuit, device, system, or principle. 2. An experimental model of a circuit in which the components are fastened temporarily to a chassis or board and electrically tested.

break—1. In a communication circuit, the taking control of the circuit by a receiving operator or listening operator. The term is used in connection with half-duplex telegraph circuits and two-way telephone circuits equipped with voice-operated devices. 2. In a circuit-opening device, the minimum distance between the stationary and movable contacts when these contacts are open. 3. An open circuit or on-hook condition as determined by the dial of a telephone set. 4. A signal sent over a backward (secondary) channel by a receiving start-stop terminal on a half-duplex circuit (or, in some cases, over the transmit channel on a full-duplex circuit), usually to indicate a requirement to transmit. 5. The word used to interrupt a conversation on a repeater to indicate there is an emergency.

break alarm—1. An alarm condition signaled by the opening or breaking of an electrical circuit. 2. The signal produced by a break alarm condition (sometimes referred to as an open-circuit alarm or trouble signal), designed to indicate possible system failure.

breakaway panels—Hardboards held together in a grouping by using breakaway tabs. Breakaway panels make handling easier for automatic insertion and wave soldering. When necessary, boards can be separated by snapping them apart much like a soda cracker.

break-before-make—1. The action of opening a switching circuit before closing another associated circuit. 2. Movable contact that breaks one circuit before making the next circuit.

break-before-make contacts—Contacts that interrupt one circuit before establishing another.

break contact—1. In a switching device, the contact that opens a circuit upon operation of the device (normally closed contact). 2. Contacts that open when a key or relay is operated.

break distance—The effective open-gap distance between the stationary and movable contacts.

breakdown—1. An electric discharge through an insulator, insulation on wire, or other circuit separator, or between electrodes in a vacuum or gas-filled tube. 2. Phenomenon occurring in a reverse-biased semiconductor diode, the initiation of which is observed as a transition from a region of high dynamic resistance to a region of substantially lower dynamic resistance for increasing magnitude of reverse voltage. Also collector-emitter breakdown, punch-through, secondary breakdown, etc. 3. Initiation of a spark discharge between two electrodes. 4. A disruptive current (discharge) through insulation. 5. Failure of insulation for any reason. 6. A state of a circuit device in which the electric field exceeds a maximum allowed value. A sharp current increase results, which may destroy the device.

breakdown diode—*See* avalanche diode.

breakdown impedance—Also called avalanche impedance. The small-signal impedance at a specified direct current in the breakdown region of a semiconductor diode.

breakdown region—The entire region of the volt-ampere characteristic beyond the initiation of breakdown due to reverse current in a semiconductor-diode characteristic curve.

breakdown strength—*See* dielectric strength.

breakdown torque—The maximum torque a motor will develop, without an abrupt drop in speed, as the rated voltage is applied at the rated frequency.

breakdown voltage—1. That voltage at which an insulator or dielectric ruptures, or at which ionization and conduction take place in a gas or a vapor. 2. The voltage measured at a specified current in the breakdown region of a semiconductor diode at which there is a substantial change in characteristics from those at lower voltages. Also called zener voltage. 3. The voltage required to jump an air gap. 4. The reverse-bias voltage applied to a pn junction for which large currents are drawn for relatively small increases in voltage. 5. The voltage at which a disruptive discharge takes place, either through or over the surface of insulation. 6. The voltage at which the insulation between two conductors will break down.

break elongation—The relative elongation of a specimen of recording tape or base film at the instant it breaks after having been stretched at a given rate.

breakerless ignition—A semiconductor electronic ignition system that does not use mechanical breaker contacts for timing or triggering purposes, but retains the distributor mechanism for distribution of the secondary voltage.

breaker points—1. The low-voltage contacts that interrupt the current in the primary circuit of the ignition system of a gasoline engine. 2. A pair of movable points that are opened and closed to break and make the primary circuit.

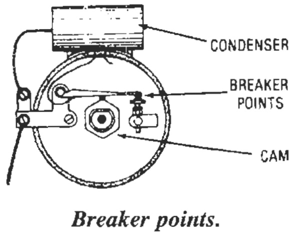

Breaker points.

break frequency—In a plot of log gain (attenuation) versus log frequency, the frequency at which the asymptotes of two adjacent linear slope segments meet.

break-in keying—In the operation of a radiotelegraph communication system, a method by which the receiver is capable of receiving signals during transmission spacing intervals.

break-in operation—A method of radiocommunication involving break-in keying that allows the receiving operator to interrupt the transmission.

break-in relay—A relay used for break-in operation.

break-make contact—Also called transfer contact. A contact form in which one contact opens its connection to another contact and then closes its connection to a third contact.

breakout—The exit point of a conductor or number of conductors along various points of a main cable of which the conductors are a part. This point is usually harnessed or sealed with some synthetic-rubber compound.

breakout box—Abbreviated BOB. 1. A testing device that permits the user to monitor the status of the various signals between two communicating devices, and to cross and tie interface leads, using jumper wires. 2. A device for monitoring and manipulating the interface signals between a DTE (data terminal equipment) and a DCE (data circuit-terminating equipment) or a pair of DTEs. 3. A device used to separately test each electrical line in a communications cable.

breakover—In a silicon-controlled rectifier or related device, a transition into forward conduction caused by the application of an excessively high anode voltage. In some cases this is destructive to the device.

breakover voltage—The value of positive anode voltage at which a silicon-controlled rectifier with the gate circuit open switches into the conductive state.

break period—The time interval during which the circuit contacts of a telephone dial are open.

breakpoint—1. A place in a computer routine, specified by an instruction, instruction digit, or other condition, where the routine may be interrupted by external intervention or by a monitor routine. 2. A special instruction that may be inserted in a program to break off the normal program control and return control to a debug-type program. When a breakpoint is executed, the debug program indicates what the computer was doing at that point. 3. A hardware or software condition (bit pattern) that stops program execution; e.g., specific addresses or control signals. The user can then examine registers and memory locations, alter data generated by the program, and even request that execution resume from that point. 4. A point in a program where the program flow is interrupted for the purpose of testing the logic of the program up to that point. Usually a breakpoint will transfer control to a routine that will display test data.

breakpoint instruction—In the programming of a digital computer, an instruction that, together with a manual control, causes the computer to stop.

breakpoint switch—A manually operated switch that controls conditioned operation at breakpoints; it is used primarily in debugging.

breakup—*See* color breakup.

breathing—1. Amplitude variations similar to bounce, but at a slow, regular rate. 2. Slow, rhythmic pulsations of a quantity, such as current, voltage, brightness, beat note, etc., readily noticeable on a CRT display.

breezeway—In the NTSC color system, that portion of the back porch between the trailing edge of the sync pulse and the start of the color burst.

B register—A computer register that stores a word which will change an instruction before the computer carries out that instruction.

bremsstrahlung—Electromagnetic radiation emitted by a fast-moving charged particle (usually an electron) when it is slowed down (or accelerated) and deflected by the electric field surrounding a positively charged atomic nucleus. X-rays produced in ordinary X-ray machines are bremsstrahlung. (In German, the term means braking radiation.)

brevity code—A code that has as its sole purpose the shortening of messages rather than concealment of content.

Brewster angle—The angle of incidence at which the reflection of parallel-polarized electromagnetic radiation at the interface between two dielectric media equals zero.

bridge—1. In a measuring system, an instrument in which part or all of a bridge circuit is used to measure one or more electrical quantities. 2. In a fully electronic stringed instrument, the part that converts the mechanical vibrations produced by the strings into electrical signals. 3. Equipment and techniques used to match circuits to each other, ensuring minimum transmission impairment. Bridging is normally required on multipoint data channels where several local loops or channels interconnect. 4. A device used singly, or with other like devices, to connect two or more similar local area networks (LANs) in order to allow workstations on one network to communicate with resources on the other network or networks as though they resided on the same network. 5. Equipment that connects two or more LANs that use the same protocol, allowing communication between devices on separate LANs.

bridge circuit—A network arranged so that, when an electromotive force is present in one branch, the response of a suitable detecting device in another branch may be zeroed by suitable adjustment of the electrical constants of still other branches.

bridged connection—A connection of a circuit across, or in parallel with, another circuit.

bridged T-network—A T-network in which the two series impedances of the T are bridged by a fourth impedance.

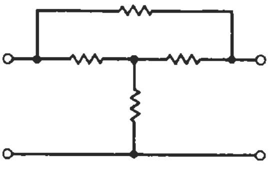

Bridged T-network.

bridge duplex system—A duplex system based on the Wheatstone-bridge principle in which a substantial neutrality of the receiving apparatus to the transmitted currents is obtained by an impedance balance. Received currents pass through the receiving relay, which is bridged between the points that are equipotential for the transmitted currents.

bridge hybrid—*See* hybrid junction, 2.

bridge quad—*See* B-quad.

bridge rectifier—A full-wave rectifier with four elements connected in the form of a bridge circuit so that dc voltage is obtained from one pair of opposite junctions when an alternating voltage is applied to the other pair of junctions.

bridge tap—An unterminated length of line attached (bridged) at some point between the extremities of a communication line; bridge taps are undesirable.

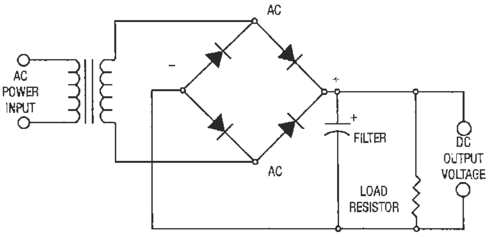

Bridge rectifier.

bridge transformer—*See* bridging transformer.

bridging—1. The shunting of one electrical circuit by another. 2. Connecting two electrical circuits in parallel. Usually the input impedances are large enough so as not to affect the signal level.

bridging amplifier—An amplifier with an input impedance sufficiently high (at least 10 times) that its input may be bridged across a circuit without substantially affecting the signal level of the circuit.

bridging connection—A parallel connection by means of which some of the signal energy in a circuit may be withdrawn with imperceptible effect on the normal operation of the circuit.

bridging contacts—1. A set of contacts in which some of the signal energy in a circuit may be withdrawn with imperceptible effect on the normal operation of the circuit. 2. A set of contacts in which the moving contact touches two stationary contacts simultaneously during transfer.

bridging gain—Ratio between the power a transducer delivers to a specified load impedance under specified operating conditions and the power dissipated in the reference impedance across which the transducer input is bridged. Usually expressed in decibels.

bridging loss—1. Ratio between the power dissipated in the reference impedance across which the input of a transducer is bridged and the power the transducer delivers to a specified load impedance under specified operating conditions. Usually expressed in decibels. 2. The loss resulting from bridging an impedance across a transmission circuit. Unless the impedance is a pure resistance, the loss will vary with frequency.

bridging transformer—Also called bridge transformer and hybrid coil. 1. A transformer designed to couple two circuits having at least nominal ohmic isolation and operating at different impedance levels, without introducing significant frequency or phase distortion or significant phase shift. 2. A transformer whose primary impedance is at least 10 times the source impedance of the device that feeds it. The secondary can be lower or higher than the primary impedance.

brightness—1. The attribute of visual perception in accordance with which an area appears to emit more or less light. Used with cathode-ray tubes. (*Luminance* is the recommended name for the photometric quantity that has also been called brightness.) 2. *See* luminance. 3. A surface measurement of light intensity per unit projected area. Usually expressed in footlamberts.

brightness control—1. In a television receiver, the control that varies the average brightness of the reproduced image. 2. The manual bias control of a cathode-ray tube that determines both the average brightness and the contrast of a picture.

brightness of image—The apparent luminance of an image observed by the eye.

brightness scale—A graduated range of stimuli perceived as having equivalent differences of brightness.

brightness signal—*See* luminance signal.

brilliance—1. The degree of brightness and clarity in a reproduced cathode-ray tube. 2. The degree to which the higher audio frequencies sound like the original when reproduced by a receiver or public-address amplifier.

brilliance control—A potentiometer used in a three-way speaker system to adjust the output level of the tweeter for proper relative volume between the treble and the lower audio frequencies produced by the complete speaker system.

British Standard Wire Gage—A modification of the Birmingham wire gage; the legal standard of Great Britain for all wires. It is variously known as Standard Wire Gage (SWG), New British Standard (NBS), English Legal Standard, and Imperial Wire Gage.

British thermal unit—The energy required to raise the temperature of 1 pound of water 1 degree Fahrenheit.

broadband—1. As applied to data transmission, the term denotes transmission facilities capable of handling frequencies greater than those required for high-grade voice communications (higher than 3 to 4 kilohertz). 2. Having an essentially uniform response over a wide range of frequencies. To design or adjust (an amplifier) for bandwidth. 3. A communications channel in which bandwidth is greater than 64 kb/s and that can provide higher-speed data communications than a standard telephone circuit. Also called wideband.

broadband amplifier—1. An amplifier that has an essentially flat response over a wide frequency range. 2. An amplifier capable of amplifying a wide band of frequencies with minimal distortion.

broadband antenna—1. An antenna that is capable of receiving a wide range of frequencies. 2. An antenna that is not sharply resonant. It provides adequate gain across a relatively wide band of frequencies. Examples: television receiving and log-periodic antennas.

broadband electrical noise—Also called random noise. A signal that contains a wide range of frequencies and has a randomly varying instantaneous amplitude.

broadband interference—Interference occupying a frequency range that is much greater than the bandwidth of the equipment being used to measure it.

broadband klystron—A klystron in which three or more externally loaded, stagger-tuned resonant cavities are used to broaden the bandwidth.

broadband random vibration—Single-frequency component vibrations that are random in both phase and amplitude at minute increments of frequency throughout a specified bandwidth. Typically it occurs when intense, high-power-level noise impinges on structures.

broadband tube (tr and **pre-tr tubes)**—A gas-filled, fixed-tuned tube incorporating a bandpass filter suitable for radio-frequency switching.

broadcast—1. Radio or television transmission intended for public reception, for which receiving stations make no receipt. 2. To send messages or communicate simultaneously with many or all points in a circuit.

broadcast band—Radio frequency range between 530 and 1710 kHz in which all commercial AM broadcasting stations are assigned.

broadcasting—The transmitting of speech, music, or visual programs for commercial or public-service motives to a relatively large audience (as opposed to two-way radio, for example, which is utilitarian and is directed toward a limited audience).

broadcasting-satellite service—Abbreviated BSS. 1. International designation for direct-to-home satellite transmissions. In the Western Hemisphere, BSS operates in the 12.2- to 12.7-GHz frequency band. 2. A radiocommunication service in which signals transmitted or retransmitted by satellites are used for direct reception by the general public.

broadcasting service—A radiocommunication service in which the transmissions are intended for direct reception by the general public. This service may include sound transmissions, television transmissions, or other types of transmissions.

broadcasting station—A station in the broadcasting service.

broadcast receiver—A receiver intended primarily for picking up standard broadcast stations.

broadcast teletext—A one-way system in which textual and graphic information in digital form is placed on unused portions of the television signal and retrieved by the user via a special terminal built into or attached to a TV set. The information is broadcast in a continuous cycle or loop and is retrieved by the user as it passes.

broadside array—1. An antenna array whose direction of maximum radiation is perpendicular to the line or plane of the array (depending on whether the elements lie on a line or a plane). A uniform broadside array is a linear array whose elements contribute fields of equal amplitude and phase. 2. A unidirectional antenna array whose individual radiating elements are all in the same plane.

broad tuning—A tuned circuit or circuits that respond to frequencies within one band or channel, as well as to a considerable range of frequencies on each side.

Brownian motion noise—Also identified as thermal or Johnson noise. A random movement of microscopic particles, in organic or inorganic fluid suspension, caused by collision with surrounding molecules.

brownout—1. Refers to low line voltage, which can cause misoperation and possible damage of equipment. For example, a motor trying to start at low voltage can actually be in a lock-rotor condition and overheat. 2. Deliberate lowering of the line voltage by a power company to reduce load demands. 3. A period of low-voltage electrical power due to unusually heavy power demands. Brownouts can cause computers to operate erratically or to crash. 4. A sudden, unexpected reduction in electrical power, usually lasting just a few seconds, but long enough to cause computer equipment to fail from insufficient power levels.

browse—To view data.

browser—1. A software application that permits viewing and possibly searching of content in an information database, typically text, static images, or graphics, in a random or leisurely fashion at the user's discretion. 2. An application used to view information from the Internet. Browsers provide a user-friendly interface for navigating through and accessing the vast amount of information on the Internet.

browsing—1. Exploring an online area, usually on the World Wide Web. 2. To look at files or computer listings in search of something interesting.

Bruce antenna—Original name of the rhombic end-fire antenna, which consists of a diamond-shaped arrangement of four wires with the feed line at one end and a resistive termination at the opposite end of the longer diagonal. Bruce's name also is given to a series-fed array of vertically polarized resonant rectangular loops (or half-loops over ground) with one-quarter-wave width and one-half-wave spacing.

brush—1. A piece of conductive material, usually carbon or graphite, that rides on the commutator of a motor and forms the electrical connection between it and the power source. 2. A carbon block or metal leaf spring used to make sliding contact with a rotating contact, such as the commutator of a generator.

brush discharge—An intermittent discharge of electricity that starts from a conductor when its potential exceeds a certain value but is too low for the formation of an actual spark. It is generally accompanied by a whistling or crackling noise.

brush-discharge resistance—*See* corona resistance.

brush fluxing—A specialized wave-solder technique. A 360° bristled brush rotates in a foaming flux head to transfer the flux to the board. *See also* spray fluxing; wave fluxing; foam fluxing.

brush rocker—A movable rocker, or yoke, on which the brush holders of a dynamo or motor are fixed so that the position of the brushes on the commutator can be adjusted.

brush station—In a computer, a position where the holes of a punched card are sensed, particularly when this is done by a row of brushes sweeping electrical contacts.

brute force—The use of seemingly inefficient design in order to achieve a desired result. Sometimes this is done in order to avoid involved design procedures, critical adjustments, or the like, but often it is the only possible approach. For example, the miniaturization of low-frequency speakers requires brute force in the form of greatly increased amplifier power.

brute-force filter—A type of power-pack filter that depends on large values of capacitance and inductance to smooth out pulsations, rather than on the resonant effects of a tuned filter.

brute supply—A power supply that is completely unregulated. It employs no circuitry to maintain the output voltage constant with a changing input line or load variations. The output voltage varies in the same percentage and in the same direction as the input line voltage. If the load current changes, the output voltage changes inversely—an increase in load current drops the output voltage.

BSC—Abbreviation for binary synchronous communications. A line-control procedure for communicating, expressed in eight-bit EBCDIC, seven-bit U.S. ASCII, or six-bit transcode. The data code selected must include the required line-control characters, used according to specified rules.

B-scope—A type of radarscope that presents the range of an object by a vertical deflection of the signal on the screen, and the bearing by a horizontal deflection.

"B" service—FAA service pertaining to transmission and reception by teletype or radio of messages containing requests for and approval to conduct an aircraft flight, flight plans, in-flight progress reports, and aircraft arrival reports.

B-spline—A mathematical representation of a smooth curve.

B-spline surface—The mathematical description of a three-dimensional surface that passes through a set of B-splines.

BSS—*See* broadcasting-satellite service.

B supply—A source for supplying a positive voltage to the anodes and other positive electrodes of electron tubes. Sometimes called B+ supply.

BT-cut crystal—A piezoelectric plate that is cut from a quartz crystal at an angle of rotation (about the x-axis) of $-49°$. At approximately 25°C it has a zero temperature coefficient of frequency.

bubble—*See* dot, 2.

bubble logic—An obsolete form of magnetic logic.

bubble memory—1. A bubble-memory chip surrounded by the two orthogonal coils, sandwiched between two permanent-bias magnets, which provide nonvolatility and stable domains. The coils create a rotating magnetic field that moves bubbles or cylindrical magnetic domains through a film on the chip. Films may consist of magnetic metals or synthetic ferrite or garnet crystals. In addition, control functions are located in aluminum-copper elements on the magnetic film. This entire grouping is then enclosed in a magnetic shield to prevent external magnetic fields (up to about 20 oersteds) from affecting the data. 2. A device that stores data as the absence or presence of tiny right-cylindrical domains of magnetization in a magnetic medium. Bubble memories operate as shift registers with complex control signals for the various magnetic fields required. 3. A nonvolatile storage technique that uses magnetic fields to create regions of magnetization. A pulsed field breaks the regions into isolated bubbles, free to move along the surface of the crystal sheet that contains the regions. The presence or absence of a bubble represents digital (bit, not bit) information. External electromagnetic fields manipulate the bubbles (information) past read/write locations within the memory. 4. Serial-access nonvolatile memory storing data in magnetic domains (bubbles) of opposite polarity from the rest of the cell. It is characterized by very low access times and high cost, but is able to store data after power is removed. This nonvolatility makes it well suited for portable applications in which disk storage is impractical. (Note: Bubble memories are no longer used.)

bubble sort—A sorting algorithm for putting the elements of a file in order.

buck—To oppose, as one voltage bucking another, or the magnetic fields of two coils bucking each other.

bucket—A general term for a specific reference in storage in a computer, such as a section of storage, the location of a word, a storage cell, etc.

bucket brigade—1. A shift register that transfers information from stage to stage in response to timing signals. 2. A technique of building a photosensing array using field-effect transistors (FETs). Two FET amplifiers are at each sensing area and are interconnected as an element of a shift register. The charges sensed by the FETs are successively clocked from element to element. The electrical charges representing bits of stored information are transferred from one element to the next by means of clock pulses that raise the level of each element in the correct sequence, in the same way as a firefighting brigade passes a bucket of water down the line.

bucket curve—A graphical presentation of noise within a microwave system, displayed in a form that permits separation and identification of the individual idle and intermodulation (IM or cross-modulation) impairments.

bucking—Counteracting one quantity (such as a current or voltage) by opposing it with a like quantity of equal magnitude but opposite polarity.

bucking coil—A coil connected and positioned in such a way that its magnetic field opposes the magnetic field of another coil. The hum-bucking coil of an electrodynamic speaker is an example.

bucking voltage—A voltage that is opposite in polarity to another voltage in the circuit and hence bucks, or opposes, the latter voltage.

buckling—The warping of the plates of a battery due to an excessively high rate of charge or discharge.

buffer—1. A circuit or component that isolates one electrical circuit from another. Usually refers to electron-tube amplifiers used for this purpose. 2. A vacuum-tube stage used chiefly to prevent undesirable interaction between two other stages. In a transmitter, it generally follows the master-oscillator stage. 3. An isolating circuit used in an electronic computer to avoid reaction of a driven circuit. 4. A storage device used to compensate for a difference in the rate of flow of information or the time or occurrence of events when transmitting information from one device to another. 5. A computer circuit having an output and a multiplicity of inputs and designed so that the output is energized. Thus, a buffer performs the circuit function equivalent to the logical OR. 6. A noninverting digital circuit element that may be used to handle a large fanout or to invert input and output levels. Normally, a buffer is an emitter follower. 7. A storage device in which data is temporarily stored during data transfers between peripherals and computers. A means of temporarily storing data must be used when a peripheral device inputs data faster than the computer can process it, or when the computer sends data faster than a peripheral can receive it. 8. A soft fiber-optic material that mechanically isolates individual fibers in a fiber-optic cable or bundle from small geometrical irregularities, distortions, or roughness of adjacent surfaces. The buffer has no optical function. 9. A high-speed area of storage in a computer that is temporarily reserved for use in performing the input/output operation, into which data is read or from which data is written. 10. A storage area in a computer for data that is used to compensate for a speed difference when transferring data from one device to another. Usually refers to an area reserved for I/O operations, into which data is read, or from which data is written. 11. A temporary-storage device used to compensate for a difference in data rate and data flow between two devices (typically a computer and a printer). *See also* spooler.

buffer amplifier — 1. An amplifier designed to isolate a preceding circuit from the effects of a following circuit. 2. An amplifier stage that is used to isolate a frequency-sensitive circuit from variations in the load presented by following stages. 3. A unity gain amplifier used to isolate the loading effect of one circuit from another.

buffer capacitor — A capacitor connected across the secondary of a vibrator transformer, or between the anode and cathode of a cold-cathode rectifier tube, to suppress voltage surges that might otherwise damage other parts in the circuit.

buffer circuitry — Circuitry necessary to adapt signals between two systems, e.g., between a test system and a board under test.

buffer computer — A computing system provided with a storage device so that input and output data may be stored temporarily to match the slower speed of input/output devices with the higher speed of the computer.

buffer/driver — A device used to increase the fanout or drive capability of a digital circuit.

buffered terminal — A terminal that contains storage equipment so that the rate at which it sends or receives data over its line does not need to agree exactly with the rate at which the data is entered or printed.

buffer element — A low-impedance inverting driver circuit. Because of its very low source impedance the element can supply substantially more output current than the basic circuit. As a consequence, the buffer element is valuable in driving heavily loaded circuits or minimizing rise-time deterioration due to capacitive loading.

buffer gate — A logic gate with a high output-drive capability, or fanout; a buffer gate is used when it is necessary to drive a large number of gate inputs from one gate function.

buffering — 1. Storing data between transfer operations. Data read from a disk is buffered before transfer to a system memory, and data to be written is buffered after transfer from a system memory. 2. The process of temporarily storing data in a software program or in RAM to allow transmission devices to accommodate differences in data transmission rates.

buffer storage — 1. A synchronizing element used between two different forms of storage (usually internal and external). 2. An input device in which information from external or secondary storage is assembled and stored ready for transfer to internal storage. 3. An output device into which information from internal storage is copied and held for transfer to secondary or external storage. Computation continues during transfers from buffer storage to secondary or internal storage or vice versa. 4. A device for the purpose of storing information temporarily during data transfers. 5. *See* cache.

buffer storage unit — A computer unit used for temporary storage of data before transmission to another destination. The unit often is able to accept and give back data at widely varying rates so that data transfer takes place efficiently for each device connected to it; that is, a high-speed device is not slowed by the presence of a slow-speed device in the same system unless the data transfer occurs directly between the two devices.

buffer stripper — A device that removes flat cable insulation from conductors; a unit of motorized buffing wheels that scrapes the insulation and brushes it away. Also called an abrasion stripper.

bug — 1. A semiautomatic telegraph sending key consisting of a lever that is moved to one side to produce a series of correctly spaced dots, and to the other side to produce a dash. 2. A circuit fault due to improper design or construction. 3. An electronic listening device, generally small and concealed, used for commercial or military espionage. 4. To plant a microphone or other sound sensor or to tap a communication line for the purpose of surreptitious listening or audio monitoring; loosely, to install a sensor in a specified location. 5. The microphone or other sensor used for the purpose of surreptitious listening. 6. A device or system used for surreptitiously sensing and removing audio information from a target area. In the sense in which it is usually used, a bug transmits information from the target area to the listening post via radio; however, the information may be transmitted over dedicated wire by varying dc, as in an ordinary telephone circuit, or over power wiring by carrier current, as in a wireless intercom. The sensor in this system (transducer that changes sound into electrical impulses) is normally a microphone, but might be something that has another primary purpose, such as a telephone ringer. 7. A term used to denote a mistake in a computer program or a malfunction in a computer hardware component. 8. A defect in hardware or software that causes a program to malfunction.

bugging height — The distance between the hybrid substrate and the lower surface of the beam-lead device that occurs because of deformation of beam leads during beam-lead bonding.

build — The increase of diameter due to insulation.

building-out — Addition to an electric structure of an element or elements electrically similar to an element or elements of the structure to bring a certain property or characteristic to a desired value.

building-out circuit — A short section of transmission line, or a network that is shunted across a transmission line, for the purpose of impedance matching.

building-out network — A network designed to be connected to a basic network so that the combination will simulate the sending-end impedance, neglecting dissipation, of a line having a termination other than that for which the basic network was designed.

building-out section — A short section of transmission line, either open or short-circuited at the far end, shunted across another transmission line for use on an impedance-matching transformer.

building wire—Insulated wires used in buildings for light and power, 60 volts or less. Usually not exposed to outdoor environment.

bulb—The glass envelope that encloses an incandescent lamp or an electronic tube.

bulb-temperature pickup—A temperature transducer in which the sensing element is enclosed in a metal tube or sheath to protect it against corrosive liquids or other contaminants.

bulge—The difference between the actual characteristic and a linear characteristic of the attenuation-frequency characteristic of a transmission line.

bulk degausser—*See* bulk eraser.

bulk effect—An effect, such as current, resistance, or resistivity, observed in the overall body of a sample of material, as opposed to a region within the material or on its surface.

bulk-erased noise—The noise arising when a bulk-erased tape is reproduced with the erase and record heads completely deenergized. Ideally, this noise is governed by the number of magnetic particles that pass by the head in unit time.

bulk eraser—Also called bulk degausser or degausser. 1. Equipment for erasing a roll of tape. The roll is usually rotated while a 60-hertz ac erasing field is decreased, either by withdrawing the roll from an electromagnet or by reducing the ac supply to an electromagnet. 2. A device used to erase an entire tape at one time. Bulk erasers are usually more effective than a recorder's erase heads.

bulk noise—*See* excess noise.

bulk resistance—The portion of the contact resistance that is due to the length, cross section, and material.

bulk resistivity—Resistance measured between opposite faces of a cube of homogeneous material.

bulk resistor—A resistor made by providing ohmic contacts between two points of a homogeneous, uniformly doped crystal of silicon material.

bulk storage—An external device, such as a tape recorder, on which a computer's contents can be recorded to be reloaded into the computer when needed.

bulletin board—A computer system operated so that people can access it via telephone lines and leave messages for others to see.

bump—A means of providing connections to terminal areas of a device. A small mound is formed on the device (or substrate) pads and is utilized as a contact for face-down bonding.

bump chip—A chip that has on its termination pads a bump of solder or other bonding material that is used to bond the chip to external contacts. It allows for simultaneous bonding of all leads, rather than one at a time as in wire bonding.

bump contacts—Small amounts of material formed on the chip substrate to register with terminal pads, as when the chip is employed in flip-chip circuits.

bunched pair—A group of pairs tied together or otherwise associated for identification.

buncher—1. The input resonant cavity in a conventional klystron oscillator. 2. In a velocity-modulated tube, the electrode that concentrates the electrons in the constant-current electron beam into bunches.

buncher gap—*See* input gap.

buncher resonator—The input cavity resonator in a velocity-modulated tube. It serves to modify the velocity of the electrons in the beam.

bunching—1. Grouping pairs together for identification and testing. 2. In a velocity-modulated electron stream, the action that produces an alternating convection-current component as a direct result of the differences of electron transit time produced by the velocity modulation.

bunching parameter—One-half of the product of the bunching angle in the absence of velocity modulation and the depth of velocity modulation.

bunching time—The time in the armature motion of a relay during which all three contacts of a bridging-contact combination are electrically connected.

bunching voltage—The radio-frequency voltage between the grids of the buncher resonator in a velocity-modulated tube such as a klystron. Generally, the term implies the peak value of this oscillating voltage.

bunch stranding—1. A method in which a number of wires are twisted together in a common direction and with a uniform pitch to form a finished, stranded wire. 2. A group of wires of the same diameter twisted together without a predetermined pattern.

bundle—1. A number of optical fibers grouped together, usually carrying a common signal. 2. A collection of glass or plastic fibers that transmit data in the form of optical energy. 3. A group of optical fibers contained within a single jacket.

bundle connector—Fiber-optic connector that joins fiber bundle to fiber bundle. It is used when fiber-optic cables penetrate bulkheads, when new cables are spliced into transmission networks, and for repair of breaks. *See* source connector; detector connector.

bundled cable—Individual insulated wires laced together to form a bundle to facilitate handling.

bundled software—Software that comes with a computer.

bundling—*See* lacing and harnessing.

buried cable—Also called direct burial cable. 1. A cable installed underground and not removable except by disturbing the soil. 2. A cable installed directly in the earth without use of underground conduit.

buried channel—Because charge trapping can occur at the surface of the $Si-SiO_2$ interface, a thin doped layer can be introduced in the silicon just below the oxide (typically by ion implantation) to prevent trapping of charges. (MIS technology term.)

buried layer—1. A layer of very low resistivity, usually of n+ material, between the high-resistivity n-type collector region and p-type substrate of an integrated-circuit transistor. The buried layer tends to reduce the series collector resistance of the transistor without having an adverse effect on the breakdown voltage. 2. An underlying layer of a silicon IC formed by introducing impurities into the silicon, then covering it with additionally grown silicon.

buried resistors—Terminating film resistors deposited on inner layers of multilayer boards in order to reduce conductor lengths.

burn—*See* burned-in image.

burned-in image—Also called burn. An image that remains in a fixed position in the output signal of a camera tube after the camera has been turned to a different scene.

burn-in—1. Operation of a device to stabilize its failure rate. 2. The operation of items prior to their ultimate application, intended to stabilize their characteristics and to identify early failures. 3. Phase of component testing in which infant mortality or early failures are screened out by running the circuit for a specified length of time (such as 168 hours). 4. Subjecting components to a high temperature, normally 125°C, for a specified period under an electrical power stress (normally 80 percent of rated power). The process is designed to accelerate the aging of a device beyond the infant mortality life stage. After an appropriate burn-in, the devices should have a very low failure rate, normally defined by military standards. 5. Operation of a component, module, or system under some increased stress, typically elevated temperature, so as to cause failure at the vendor's rather than at the user's plant. 6. A test of computer components that runs the system for a day or two to detect defective chips. 7. A test in which a product, such as a chip, is operated prior to its use to detect failures that would otherwise occur in about the first six months of the product's life.

burning — Programming a read-only memory chip.

burn-in period — The time during which components are operated at predetermined stress conditions prior to their installation in the user's equipment. *See* early-failure period.

burst — 1. A sudden increase in the strength of a signal. 2. The cosmic-ray effect on matter, causing a sudden intense ionization that gives rise to great numbers of ion pairs at once. 3. A group of events occurring together in time. 4. A number of events that occur within a short period of time. 5. One of a series of successions of bits, frames, or other elements of data, occurring at regular intervals. 6. A pulsed rf envelope, generally 2.3 μs (8 cycles) in duration, that appears on the back porch of the horizontal synch pulse. This 3.579545-MHz frequency is used to synchronize the television receiver's regenerated subcarrier color oscillator. 7. *See* color burst.

burst error — A series of consecutive errors in data transmission.

burst noise — An unwanted signal that is characterized by an excessively large interfering effect that is extended over a relatively short but finite time interval.

burst pedestal — A rectangular pulselike television signal that may be part of the color burst. The amplitude of the color-burst pedestal is measured from the alternating-current axis of the sine-wave portion of the horizontal pedestal.

burst pressure — The maximum pressure to which a device can be subjected without rupturing.

burst rate — The bit rate during the time of transmission from a terminal that transmits TDMA bursts.

burst separator output — The amplitude of the chroma reference burst at the output of the gated burst amplifier.

burst sequence — An arrangement of color-burst signals in which the polarity of the burst signal is the same at the start of each field so that the stability of color synchronization is improved.

burst transmission — Radio transmission in which messages are stored and then released at 10 to 100 or more times the normal speed. The received signals are recorded and returned to the normal rate for the user.

bus — 1. The term used to specify an uninsulated conductor (a bar or wire); may be solid, hollow, square, or round. 2. Sometimes used to specify a bus bar. 3. The communications path between two switching points. 4. Wire used to connect two terminals inside an electrical unit. A common point for electrical circuits to return. Can be bare, tinned, or insulated. 5. A power line that provides power to a large number of circuits. In computing, a bus is a group of wires that conveys information to a large number of devices. The information may be data, commands, or addresses, or all three in sequence. All the devices in the system are connected to the bus. Each device continually listens for a command addressed to it. Only one device is allowed to transmit over the bus at once. Bus-oriented systems are flexible and easy to expand. 6. A circuit over which data or power is transmitted or received. 7. A power or signal line used in common by several parts of a computer. 8. A signal path to which a number of inputs may be connected for feed to one or more outputs. 9. A group of conductors considered as a single entity that interconnects various parts of a system, e.g., data bus or address bus. 10. Usually a group of wires over which digital information is transmitted from one of several sources to one of several destinations. Commonly found in digital systems. 11. A group of wires that allows memory, CPU, and I/O devices to exchange words. 12. A connective link between multiple processing sites (colocated only), where any of the processing sites can transmit to any other, but only one way at a time. 13. A mechanical, logical, and electrical interconnection scheme for modular microcomputer circuit elements; provides a common path for data, address, and control information between those modular elements. 14. A group of conductors that provide time-shared communication paths for the transmission of information between equipment units. 15. A common pathway or channel between hardware devices. Can be serial (information travels one bit at a time) or parallel (information travels in groups of bits moving simultaneously along multiple parallel paths). 16. A data path shared by many devices (e.g., multipoint line) with one or more conductors for transmitting signals, data, or power. 17. The physical channel over which electric signals are transferred between the components of a system, along with the protocol rules governing the transfer.

bus analyzer — 1. A tool that captures the bus signals at a given instant for later analysis. 2. An instrument that analyzes digital data on the bus lines of a microprocessor.

bus architecture — A data-communications structure that consists of a common connection among a number of printer/plotter modules.

busback — The connection, by a common carrier, of the output portion of a circuit back to the input portion of a circuit. *See also* loopback test.

bus bar — 1. A heavy copper strap or bar used to carry high currents or to make a common connection between several circuits. 2. A heavy copper (or other metal such as aluminum) strip or bar used on switchboards and in power plants to carry heavy currents. 3. Interconnection device that distributes power from remote power supplies to cabinets or drawers, across connector backpanels, and on PC boards. These flat, rectangular components frequently consist of two or more conductor layers electrically insulated from one another and from other components by thin dielectric layers. Usually, conductor layers are bonded or laminated together to provide a rigid, mechanically and electrically stable package that is mounted with screws, clamps, or foam adhesive. Bus bars mounted on PC boards are typically soldered in place along with other components that are to be mounted on the boards.

bus driver — 1. An integrated circuit that is added to the data bus system in a computer to facilitate proper drive to the CPU when several memories are tied to the data bus line. Drivers are necessary because of capacitive loading, which slows down the data rate and prevents proper time sequencing of microprocessor operation. 2. A buffering device that increases the driving capability of a microprocessor that itself may be capable of driving no more than a single TTL load. 3. A circuit that amplifies a bus data or control signal sufficiently to ensure valid reception of that signal at the destination.

bushing — A mechanical device used as a lining for an opening to prevent abrasion to wire and cable. Also used as a low-cost method of insulating, anchoring, cushioning, and positioning. Usually a nonmetallic material is preferred.

business data processing — 1. Automatic data processing used in accounting or management. 2. Data processing for business purposes, such as recording and summarizing the financial transactions of a business.

business machine — Customer-provided equipment that is connected to the communications services of a common carrier for the purpose of data movement.

busing — The joining of two or more circuits together.

bus master—In a bus structure, in which control of data transfers on the bus is shared between the central processor and associated peripheral devices, the term bus master refers to the device controlling the current bus transaction.

bus organization—The manner in which many circuits are connected to common input and output lines (buses).

bus reactor—A current-limiting reactor connected between two different buses or two sections of the same bus for the purpose of limiting and localizing the disturbance due to a fault on either bus.

bus slave—In a bus structure, in which control of data transfers on the bus is shared between the central processor and associated peripheral devices, the device currently receiving or transmitting data from or to the bus master is referred to as the bus slave.

bus system—A network of paths inside a microprocessor that facilitates data flow. The important buses in a microprocessor are identified as data bus, address bus, and control bus.

bust this—Phrase used instead of a normal message ending to indicate that the entire message, including heading, is to be disregarded.

busy test—In telephony, a test to find out whether certain facilities that may be desired, such as a subscriber line or a trunk, are available for use.

busy tone—Interrupted low tone returned to the calling party to indicate that the called line is busy.

Butler antenna—An array antenna in which hybrid junctions are incorporated into the feed system to obtain a plurality of independent beams.

Butler oscillator—A two-tube (or transistor) crystal-controlled oscillator in which the crystal forms the positive feedback path when excited in its series-resonant mode.

butt connector—*See* butt contact.

butt contact—A hemispherically shaped contact designed to mate end to end without overlap, with axes aligned, against a similarly shaped contact. When properly aligned, the two convex surfaces form a reasonably good surface-to-surface contact, usually under spring pressure, with the ends designed to provide optimum surface contact.

butterfly capacitor—*See* butterfly resonator.

butterfly circuit—Frequency-determining element having no sliding contacts and providing simultaneous change of both inductance and capacitance. It is used to replace conventional tuning capacitors and coils in ultrahigh-frequency oscillator circuits. The rotor of the device resembles the opened wings of a butterfly.

butterfly resonator—Also called butterfly capacitor. A tuning device that combines both inductance and capacitance in such a manner that it exhibits resonant properties at very high and ultrahigh frequencies (characterized by a high tuning ratio and Q). So called because the shape of the rotor resembles the opened wings of a butterfly.

Butterworth filter—A filter network that exhibits the flattest possible response in the passband. The response is monotonic, rolling off smoothly into the stopband, where it approaches a constant slope of 6 dB/octave.

Butterworth function—A mathematical function used in designing a filter for maximally constant amplitude response with little consideration for time delay or phase response.

butt joint—1. A splice or other connection formed by placing the ends of two conductors together and joining them by welding, brazing, or soldering. 2. A connection between two waveguides that maintains electrical continuity by providing physical contact between the ends.

button—1. The metal container in which the carbon granules of a carbon microphone are held. 2. Also called dot. A piece of metal used for alloying onto the base wafer in making alloy transistors.

button capacitor—A fixed button-shaped ceramic or silvered-mica capacitor that, because of its disk shape and mode of terminal connection, offers very low internal inductance.

button-hook contact—A contact with a curved, hooklike termination, often located at the rear of hermetic headers to facilitate soldering or desoldering of leads.

buttonhook feed—A rod shaped like a question mark supporting the feedhorn and LNA. A buttonhook feed for use with commercial-grade antennas is often a hollow waveguide that directs signals from a feedhorn to an LNA behind the antenna.

buttons—In a computer, objects that, when clicked once, cause something to happen.

button silver-mica capacitor—A stack of silvered-mica sheets encased in a silver-plated brass housing. The high-potential terminal is connected through the center of the stack. The other capacitor terminal is formed by the metal shell, which connects at all points around the outer edge of the electrodes. This design permits the current to fan out in a 360° pattern from the center terminal, providing the shortest possible electrical path between the center terminal and chassis. The internal series inductance is thus kept small.

button stem—*See also* pressed stem. In a vacuum tube, the glass base onto which the mount structure is assembled. The pins may be sealed into the glass; if so, no base is needed. In some large tubes, the stiff wires are passed directly into the base pins to give added strength.

button up—To close or completely seal any operating device.

butt splice—A device for joining two conductors placed end to end with their axes in line (that is, conductors not overlapping).

buzzer—A signaling device in which an armature vibrates to produce a raucous, nonresonant sound.

BV_{EBO}—The reverse-breakdown voltage of the emitter-to-base junction of a transistor with the collector open-circuited.

BWG—Abbreviation for Birmingham wire gage.

BX cable—Insulated wires enclosed in flexible metal tubing or flexible spiral metal armor used in electrical wiring.

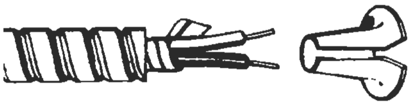

BX cable.

bypass—1. A shunt (parallel) path around one or more elements of a circuit. 2. A secondary channel that permits routing of data in a computer sample around the data compressor, regardless of the value of the sample, at intervals determined by the operator.

bypass capacitor—A capacitor used for providing a comparatively low-impedance ac path around a circuit element.

bypass filter—A filter providing a low-attenuation path around some other circuit or equipment.

bypassing—Reducing high-frequency current in a high-impedance path by shunting that path with a bypass element (usually a capacitor).

B-Y signal — One of the three color-difference signals in color television. The B-Y signal forms a blue primary signal for the picture tube when combined — either inside or outside the picture tube — with a luminance, or Y, signal.

byte — 1. A single group of bits processed together (in parallel), usually 8 bits. 2. A sequence of adjacent binary digits, usually shorter than a word, operated on as a unit. 3. The number of bits that a computer processes as a unit. This may be equal to or less than the number of bits in a word. For example, both an 8-bit and a 16-bit length computer may process data in 8-bit bytes. 4. The smallest addressable unit of main storage in a computer system. The byte consists of 8 data bits. 5. A byte is universally used to represent a character. Microcomputer instructions require one, two, or three bytes. A word may be one byte long. One byte has two nibbles. Computers and microprocessors work with words of 4, 8, 12, 16, 24, or 32 bits. 6. A set of bits that represents a single character. Usually there are 8 or 10 bits in a byte, depending on how the measurement is being made. 7. A unit of data equal to 8 bits, and hence capable of storing any one of $2^8 = 256$ distinct values.

byte-multiplexer channel — A channel that interleaves bytes of data from different sources. Contrast with *selector channel.*

byte-serial — A sequence of bit-parallel data bytes used to carry information over a common bus.

C

C—1. Symbol for capacitor, capacitance, carbon, coulomb, centigrade or Celsius, transistor collector, and (when lowercase) velocity of light. 2. A general-purpose programming language developed in the 1970s by Dennis Ritchie of AT&T, Bell Labs. Its generality, machine independence, and efficiency have made C popular for many application areas. The UNIX operating system is written in C, and the close linking of UNIX and C has made C the de facto standard language in engineering software development.

C– (C minus)—The negative terminal of a C battery, or the negative polarity of other sources of grid-bias voltage. Used to denote the terminal to which the negative side of a grid-bias voltage source should be connected.

C+ (C plus)—Positive terminal of a C battery, or the positive polarity of other sources of grid-bias voltage. The terminal to which the positive side of the grid-bias voltage source should be connected.

C^2 system—Abbreviation for command and control system.

C^3—Abbreviation for command, control, and communications.

C^3I—Abbreviation for command, control, communications, and intelligence. Includes computers, displays, communications equipment, and other supporting devices that provide the means for communications as well as intelligence-gathering capability at all levels of military command.

C^3L—Abbreviation for complementary contact-current logic.

C^4—Abbreviation for command, control, communications, and computer.

cabinet—1. A protective housing for electrical or electronic equipment. 2. An enclosure for mounting equipment. If fabricated of steel, the cabinet helps attenuate electrostatic and electromagnetic radiation noise interference.

cable—1. An assembly of one or more conductors, usually within a protective sheath, so arranged that the conductors can be used separately or in groups. 2. A stranded conductor (single-conductor cable) or a combination of conductors insulated from one another (multiple-conductor cable). 3. Jacketed combination of fiber bundles with cladding and strength-reinforcing components. 4. An assembly of insulated conductors into a compact form that is covered by a flexible, waterproof protective covering.

cable armor—In cable construction, a layer of steel wire or tape, or other extra-strength material, used to reinforce the lead wall.

cable assembly—A cable with plugs or connectors on each end for a specific purpose. It may be formed in various configurations.

cable attenuation—Reduction of signal intensity along a cable, usually expressed in decibels per foot, hundred feet, kilometer, mile, etc.

cable clamp—A device used to give mechanical support to the wire bundle or cable at the rear of a plug or receptacle.

cable complement—A group of cable pairs that have some common distinguishing characteristic.

cable core—That portion of an insulated cable lying under the protective covering or jacket.

cable coupler—A device used to join lengths of similar or dissimilar cable having the same electrical characteristics.

cable fill—The ratio of the number of pairs in use to the total number of pairs in a cable.

cable filler—Material used in multiple-conductor cables to occupy the spaces between the insulated conductors.

cable messenger—A stranded cable supported at intervals by poles or other structures and employed to furnish frequency points of support for conductors or cables.

cable modem—A device that enables a user to connect a computer to existing cable TV networks at Ethernet speeds and access Internet and/or online services. This device can be much faster than a telephone modem, operating at 14.4 kbs, 28.8 kbs, or higher.

cable Morse code—A three-element code used mainly in submarine-cable telegraphy. Dots and dashes are represented by positive and negative current impulses of equal length, and a space by the absence of current.

cable-ready television—A television receiver that can receive unscrambled cable television without the use of a converter.

cable run—The path occupied by a cable on cable racks or other support from one termination to another.

cable sheath—A protective covering of rubber, neoprene, resin, or lead over a wire or cable core.

cable splice—A connection between two or more separate lengths of cable. The conductors in one length are individually connected to conductors in the other length, and the protecting sheaths are so connected that protection is extended over the joint.

cable terminal—A means of electrically connecting a predetermined number of cable conductors in such a way that they can be individually selected and extended by conductors outside the cable.

cable TV—1. Television system in which programs are received by a local central antenna and distributed by cable to individual homes. The term does not apply to a system serving fewer than 50 subscribers or serving only subscribers in one or more multiple-unit buildings under common ownership, control, or management. 2. Linking a TV set via cable to a system operator that, for a monthly fee, provides the viewer with typically 30 to 75 or more channels of programming. 3. A system for distributing television programming by a cable network rather than by electromagnetic radiation.

cable vault — A vault in which the outside plant cables are spliced to the tipping cables.

cabling — 1. The assembly of wire bundles extended from one physical structure to another to interconnect the circuits within structures. Cabling differs from wire jumpers in that it is understood to be external to the physical structures and may include tubing sheaths, zipper tubing, or rubber jackets. 2. Twisting together two or more insulated conductors by machine to form a cable. This also is a term loosely applied to bundling wires together, such as in the forming of wire harnesses. 3. In fiber optics, a method by which a group of fibers or bundle of fibers is mechanically assembled.

cache — 1. A small, fast memory built into a processor to give faster access to the data and instructions that a program uses repeatedly (also called cache memory, buffer storage). 2. A small, fast storage area interposed between the main memory and CPU to improve memory-transfer rates and processor efficiency. The name is derived from the fact that the memory is hidden from or transparent to the programmer. 3. A storage area for frequently accessed information. Retrieval of the information is faster from the cache than the original source. There are many types of cache, including RAM cache, secondary cache, disk cache, and cache memory, to name a few. 4. In a computer system, a small but very high-speed memory buffer situated between the processor and main memory. It operates on the principle that certain memory locations tend to be accessed very often (normally for reads). Thus, when a main memory location is read, it is stored in the cache at the same time.

cache memory — 1. A high-speed, low-capacity computer memory similar to a scratch-pad memory except that it has a larger capacity. 2. The fastest portion of the overall memory, which stores only the data that the computer may need in the immediate future. 3. A high-speed buffer memory used between the central processor and main memory. The cache is filled at medium speed from the main memory. Instructions and programs can operate at higher speed if found in the cache. If not found, a new segment is loaded. The cache contains the instruction or sequence of instructions most likely to be executed next. 4. A special, extra-fast part of RAM in which frequently accessed information is stored. Same as memory cache and RAM cache. 5. *See* cache. 6. A fast random-access memory system designed to store the most frequently accessed data in RAM.

cactus needle — A phonograph needle made from the thorn of a cactus plant.

CAD — Abbreviation for computer-aided design. 1. Use of a computer to aid in the design of complex MSI or LSI circuit arrays. CAD is especially useful for custom IC fabrication. 2. Man/computer interactions for the design and testing of customer MSI/LSI arrays and other complex engineering designs in a reasonable time frame. 3. Any system or process using a computer to aid in the creation or modification of a design. 4. High-performance design workstations that enable designers to manipulate parts diagrams and simulate operations, among other things. Can be linked to computer-aided manufacturing systems.

CADAM — *C*omputer-*a*ugmented *d*esign *a*nd *m*anufacturing system for design and analysis functions, developed by Lockheed for the aerospace industry and licensed by IBM.

CAD/CAM — Computer-aided design/computer-aided manufacturing. Two highly specialized technical applications of a computer to improve the productivity of the engineer.

cadmium — A metallic element widely used for plating steel hardware or chassis to improve its appearance and solderability and to prevent corrosion. It is also used in the manufacture of photocells.

cadmium cell — A standard cell used as a voltage reference; at 20°C its voltage is 1.0186 volts.

cadmium selenide photoconductive cell — A photoconductive cell that uses cadmium selenide as the semiconductor material. It has a fast response time and high sensitivity to longer light wavelengths, such as those emitted by incandescent lamps and some infrared light sources.

cadmium sulfide cell — *See* cadmium sulfide photoconductive cell.

cadmium sulfide photoconductive cell — A photoconductive cell in which a small wafer of cadmium sulfide is used to provide an extremely high dark-light resistance ratio. Some of the cells can be used directly as a light-controlled switch operated from the 120-volt ac power line.

CAE — Abbreviation for computer-aided engineering. The use of general-purpose computers and special application software to automate routine, iterative hardware engineering tasks.

cage — 1. A completely shielded enclosure. 2. A screened room that is covered with a grounded fine-mesh conductive screen on all sides to allow measurements within the cage without any influence or interference from extraneous signals.

cage antenna — An antenna comprising a number of wires connected in parallel and arranged in the form of a cage. This arrangement reduces the copper losses and increases the effective capacity.

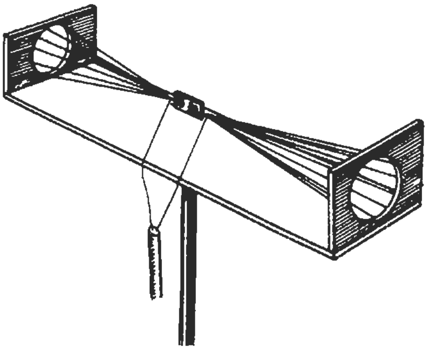

Cage antenna.

calculating — Computing a result by multiplication, division, addition, or subtraction or by a combination of these operations. A data-processing function.

calculating punch — A punched-card machine that reads data from a group of cards and punches new data in the same or other cards.

calculator — 1. A device capable of performing arithmetic. 2. A calculator as in definition 1 that requires frequent manual intervention. 3. Generally and historically, a device for carrying out logical and arithmetical digital operations of any kind.

calculator mode — Also called fast answerback. An interactive computer system that has a mode which allows the terminal to be used like a desk calculator. The user types an expression, and the computer evaluates it and returns the answer immediately.

calendar age—Age of an item or object measured in terms of time elapsed since it was manufactured.

calendar life—That period of time, expressed in days, months, or years, during which an item may remain installed and in operation, but at the end of which the item should be removed and returned for repair, overhaul, or other maintenance.

calibrate—To ascertain, by measurement or by comparison with a standard, any variations in the readings of another instrument, or to correct the readings.

calibrated triggered sweep—In a cathode-ray oscilloscope, a sweep that occurs only when initiated by a pulse and that moves horizontally at a known rate.

calibration—1. The process of comparing an instrument or device with a standard to determine its accuracy or to devise a corrected scale. 2. Taking measurements of various parts of electronic equipment to determine the performance level of the equipment and whether it conforms to technical order specifications. 3. Comparison of the performance of an item of test and measuring equipment with a reference standard traceable to the National Bureau of Standards or some other authoritative source or specification. 4. A test during which known values of measure are applied to a transducer, and its corresponding output readings are recorded.

calibration accuracy—Finite degree to which a device can be calibrated (influenced by sensitivity, resolution, and reproducibility of the device itself and the calibrating equipment). Expressed as a percentage of full scale.

calibration curve—A smooth curve connecting a series of calibration points.

calibration marker—On the screen of a radar indicator, the markings that divide the range scale into accurate intervals for range determination or checking against mechanical indicating dials, scales, or counters.

call—1. A transmission made for the purpose of identifying the transmitting station and the station for which the transmission is intended. 2. To transfer control to a specified closed subroutine. 3. In communication, the action performed by the calling party, or the operations necessary in making a call, or the effective use made of a connection between two stations.

call accounting—Daily and monthly accounting reports for individual callers and for hierarchical groupings of callers in a communication system.

call announcer—Device for accepting pulses from an automatic telephone office and reproducing the corresponding number with speechlike sounds.

callback device—A device on the receiving end of a communications network that assures the authenticity of the sender by calling the sender back.

call circuit—A communication circuit between switching points used by traffic forces for transmitting switching instructions.

call diverter—A device that intercepts a call directed to a telephone subscriber and diverts it to another number.

caller ID—A telephone company service that provides the name and number of the caller from the phone company's information network. A caller ID device translates that data into usable form by displaying it on a screen. It works on standard telephones only.

call forwarding—A service available in some dial offices whereby a subscriber can have calls to his or her number forwarded to another phone by dialing the forwarding number to the equipment. All incoming calls are then forwarded automatically. The forwarding can be discontinued at will, by command from the base telephone.

calligraphic display—Also known as a vector stroke or vector refresh. A refresh graphic system that draws the picture on the screen vector by vector. The monitor is driven by X and Y signals that control the beam's position for drawing. A vector list is continually processed and converted to the analog X and Y signals. The rate at which the picture is refreshed need not be synchronized to the ac line. Calligraphic displays draw smooth and very high-resolution lines. Since the display is refreshed from a vector list, any changes in this list immediately appear on the screen, creating dynamic displays. The main disadvantage is that increasingly complex pictures take longer to refresh, causing flicker.

call in—To transfer control of a digital computer temporarily from a main routine to a subroutine, which is inserted in the sequence of calculating operations to fulfill a subsidiary purpose.

call indicator—Device for accepting pulses from an automatic switching system and displaying the corresponding called number before an operator.

calling device—Apparatus that generates the pulses used to control the establishment of connections in an automatic telephone switching system.

calling mode—The ability to originate and/or answer on the dial-up network. A central-site modem normally has auto-call capability, meaning it can automatically dial a specific number through an automatic calling unit.

calling party control—*See* CPC.

calling sequence—In a computer, a sequence of instructions required to enter a subroutine. It may contain information required by the subroutine.

call letters—A series of government-assigned letters, or letters and numbers, that identify a transmitting station.

call number—In computer operations, a set of characters that identifies a subroutine and contains information with respect to parameters to be inserted in the subroutine or information related to the operands.

call-setup time—The overall length of time required to establish a switched call between pieces of data-terminal equipment.

call sign—1. Any combination of characters or pronounceable words that identifies a communication facility, a command, an authority, an activity, or a unit; used primarily for establishing and maintaining communications. 2. The station identification assigned to a licensee by the FCC.

call word—A call number that is exactly the size of one machine word.

calomel electrode—An electrode consisting of mercury in contact with a solution of potassium chloride saturated with mercurous chloride (calomel). *See also* glass electrode.

calorimeter—An apparatus for measuring quantities of heat. Used to measure microwave power in terms of heat generated.

calorimeter system—A precision rf watt-meter as well as an efficient dummy load able to absorb energy at any frequency band. It can absorb and measure any level of microwave energy, and functions by circulating a known amount of liquid through a suitably designed low VSWR load. The load is located in a waveguide or coaxial section that mates to the terminal of the rf energy source being measured or absorbed.

CAM—Abbreviation for content addressable memory. 1. A computer memory in which information is retrieved by addressing the content (the data actually stored in the memory) rather than by selecting a physical location. *See also* associative storage. 2. Abbreviation for computer-aided manufacturing. The use of computer technology in manufacturing processes. It includes numerical and programmable controls, factory information management, robot controls, materials handling, and storage-and-retrieval systems.

cam actuator—An electromechanical device in which a switch is closed when the high spot of a rotating cam, or eccentric, is in a certain position.

camera—*See* television camera.

camera cable—A cable or group of wires that carry the picture signal from the television camera to the control room.

camera chain—A television camera, associated control units, power supplies, monitor, and connecting cables necessary to deliver a picture for broadcasting.

camera signal—The video-output signal of a television camera.

camera tube—1. An electron-beam tube in which an electron current or charge-density image is formed from an optical image and scanned in a predetermined sequence to provide an electric signal. 2. The electron-beam tube of a television camera that converts an optical image into a pattern of electrostatic charges and then scans the pattern to produce a corresponding electric signal for transmission.

camera tube target—The storage surface of an electron-beam tube that is scanned by an electron beam to generate an output signal current corresponding to the charge-density pattern stored.

camp-on—Also called clamp-on. A method of holding a call for a line that is already in use and of signaling when the line becomes free.

can—1. A metal shield placed around a tube, coil, or transformer to prevent electromagnetic or electrostatic interaction. 2. A metal package for enclosing a device, as opposed to a plastic or ceramic package.

Canadian Standards Association—Abbreviated CSA. In Canada, a body that issues standards and specifications prepared by various voluntary committees of government and industry.

canal ray—Also called positive ray. Streams of positive ions that flow from the anode to the cathode in an evacuated tube.

candela—Abbreviated cd. 1. Formerly candle. The unit of luminous intensity. The luminous intensity of 1/60th of 1 square centimeter of projected area of a blackbody radiator operating at the temperature of solidification of platinum (2046 K). Values for standards having other spectral distributions are derived by the use of accepted spectral luminous efficiency data for photopic vision. 2. International unit of luminous intensity, also called new candle. Prior to 1948, the standard was a specific type of candle and was termed candle, candle power, or international candle.

candela/cm^2—Luminance unit called a stilb.

candle—The unit of luminous intensity. One candle is defined as the luminous intensity of 1/60th square centimeter of a blackbody radiator operating at the solidification temperature of platinum.

candlepower—1. Luminous intensity expressed in terms of standard candles. 2. A measure of the intensity of light produced by a source. This standard of measurement is used in France, Britain, and the United States. One candlepower corresponds approximately to the light produced in the horizontal direction by an ordinary sperm candle weighing six to the pound and burning at the rate of 120 gr/hr. 3. The luminous intensity of a source of light expressed in candelas.

candoluminescence—1. A phenomenon that produces white light without need for very high temperatures. 2. The luminescence of an incandescent material.

canned cycle—The use of preparatory functions on a punched tape to initiate a complete machining sequence; the need for much repetitive information in the program is thereby eliminated.

cannibalization—A method of maintenance or modification in which the required parts are removed from one system or assembly for installation on a similar system or assembly.

cantilever—The rod, or tube, that supports the stylus of a phonograph cartridge at its free end, is pivoted at or near its other end, and that transfers the stylus motion to the generating elements of the cartridge. Usually made of aluminum, but beryllium is used in some recent cartridges. Also known as the shank.

cantilevered contact—A spring contact in which the contact force is provided by one or more cantilevered springs. It permits more uniform contact pressure and is used almost exclusively in printed circuit board connectors.

capacitance—Abbreviated *C*. 1. Also called capacity. In a capacitor or a system of conductors and dielectrics, that property which permits the storage of electrically separated charges when potential differences exist between the conductors. The capacitance of a capacitor is defined as the ratio between the electric charge that has been transferred from one electrode to the other and the resultant difference in potential between the electrodes. The value of this ratio is dependent on the magnitude of the transferred charge. 2. Capacitance opposes any change in circuit voltage. A voltage change is delayed until the stored charges can be altered through current. The unit of capacitance is the farad. 3. The property of an electric system—comprised of conductors and associated dielectrics—that determines (*a*) the displacement currents in the system for a given rate of potential difference change between the conductors and (*b*) how much electrical charge will be stored in the dielectric for a given potential difference between the conductors.

capacitance alarm system—An alarm system in which a protected object is electrically connected as a capacitance sensor. The approach of an intruder causes sufficient change in capacitance to upset the balance of the system and initiate an alarm signal. Also called a proximity alarm system.

capacitance between two conductors—The ratio between the charge transferred from one conductor to the other and the resultant difference in the potentials of the two conductors when insulated from each other and from all other conductors.

capacitance bridge—A four-arm ac bridge for measuring capacitance by comparison against a standard capacitor.

capacitance detector—*See* capacitance sensor.

capacitance divider—A circuit made up of capacitors and used for measuring the value of a high-voltage pulse by making available only a small, known fraction of the total pulse voltage for measurement.

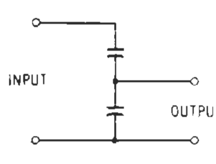

Capacitance divider.

capacitance level detector—A device with single or multiple probes that operates based on the fact that a change in capacitance level causes a change in probe capacitance.

capacitance meter—An instrument for measuring capacitance. If the scale is graduated in microfarads, the instrument is usually designated a microfaradmeter.

capacitance-operated intrusion detector—A boundary alarm system in which the approach of an intruder to an antenna wire encircling the protected area (a few feet above ground) changes the antenna-to-ground capacitance and thereby sets off the alarm.

capacitance ratio—The ratio of maximum to minimum capacitance, as determined from a capacitance characteristic, over a specified voltage range.

capacitance relay—An electronic circuit incorporating a relay that responds to a small change in capacitance, such as that created by bringing the hand or body near a pickup wire or plate.

capacitance sensor—A sensor that responds to a change in capacitance in a field containing a protected object or in a field within a protected area. Also called capacitance detector.

capacitance switch—A keyboard switch in which two pads on the circuit board under each keyswitch serve as capacitor plates connected to the drive and sense circuits. Depression of the key causes an increase in the series capacitance, coupling the two elements and creating an analog signal in the sense circuit.

capacitance tolerance—The maximum percentage deviation from the specified nominal value (at standard or stated environmental conditions) specified by the manufacturer.

capacitive coupling—Also called electrostatic coupling. The association of two or more circuits with one another by means of mutual capacitance between them. For example, between stages of an amplifier, that type of interconnection that employs a capacitor in the circuit, between the plate of one tube and the grid of the following tube or the collector of one transistor and the base of the following transistor.

capacitive diaphragm—A resonant window placed in a waveguide to provide the effect of capacitive reactance at the frequency being transmitted.

capacitive-discharge ignition—Also called capacitor-discharge ignition. An electronic ignition system used on internal combustion engines to provide nearly constant high voltage regardless of engine speed. A dc-to-dc step-up converter charges a capacitor when the distributor breaker points are closed; when they are open, the capacitor discharges through the ignition coil, thereby generating the ignition voltage

capacitive divider—Two or more capacitors placed in series across a source, making available a portion of the source voltage across each capacitor. The voltage across each capacitor will be inversely proportional to its capacitance.

capacitive feedback—The process of returning part of the energy in the plate or output circuit of a vacuum tube to the grid, or input, circuit by means of a capacitance common to both circuits.

capacitive load—A predominantly capacitive load, that is, one in which the current leads the voltage.

capacitive post—A metal post or screw extending at right angles to the E field in a waveguide. It provides capacitive susceptance in parallel with the waveguide for purposes of tuning or matching.

capacitive reactance—Symbolized by X_C. The impedance a capacitor offers to ac or pulsating dc. Measured in ohms and equal to $1/2\pi fC$, where f is in hertz and C is in farads.

capacitive speaker—*See* electrostatic speaker.

capacitive storage welding—A particular type of resistance welding whereby the energy is stored in banks of capacitors, which are then discharged through the primary of the welding transformer. The secondary current generates enough heat to produce the weld.

capacitive transduction—Conversion of the measurand into a change in capacitance.

capacitive tuning—Tuning by means of a variable capacitor.

capacitive voltage divider—A combination of capacitors connected in series to form a capacitive voltage dividing device for application with ac voltages.

capacitive welding—An electronic welding system in which energy stored in a capacitor is discharged through the joint to be welded. The resulting current develops the heat necessary for the operation.

capacitive window—A conductive diaphragm extended into a waveguide from one or both sidewalls to introduce the effect of capacitive susceptance in parallel with the waveguide.

capacitivity—*See* dielectric constant.

capacitor—1. A device consisting essentially of two conducting surfaces separated by an insulating material or dielectric such as air, paper, mica, glass, plastic film, or oil. A capacitor stores electrical energy, blocks the flow of direct current, and permits the flow of alternating current to a degree dependent essentially on the capacitance and the frequency. 2. An electrical energy storage device used in the electronics industry for varied applications, notably as elements of resonant circuits, in coupling and bypass application, blockage of dc current, as frequency determining and timing elements, as filters and delay-line components, and in voltage transient suppression.

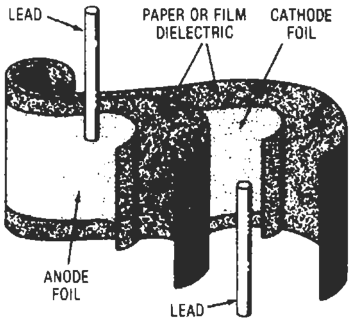

Capacitor (internal construction).

capacitor antenna—Also called condenser antenna. An antenna that consists of two conductors or systems of conductors and the essential characteristic of which is its capacitance.

capacitor bank—A number of capacitors connected together in series, parallel, or in series-parallel.

capacitor braking—A means of stopping an induction motor. The capacitor or capacitors can be applied to the winding after shut-off.

capacitor color code—Color dots or bands placed on capacitors to indicate one or more of the following: capacitance, capacitance tolerance, voltage rating, temperature coefficient, and the outside foil (on paper or film capacitors).

capacitor-discharge ignition—*See* capacitive-discharge ignition.

capacitor discharge system — An ignition system that stores its primary energy in a capacitor.

capacitor electrolyte — A current-conducting material (nonsolid or solid) serving as the cathode in an electrolytic capacitor.

capacitor filtering — A method for improving the form factor of a direct current by means of a parallel capacitor. Also, a means for increasing the magnitude of a rectified voltage.

capacitor-input filter — A power-supply filter in which a capacitor is connected directly across, or in parallel with, the rectifier output.

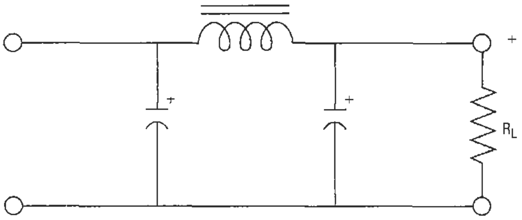

Capacitor-input filter.

capacitor losses — The active power dissipated by a capacitor.

capacitor microphone — *See* electrostatic microphone.

capacitor motor — A single-phase induction motor with the mean winding arranged for direct connection to the power source, together with an auxiliary winding connected in series with a capacitor.

capacitor pickup — A phonograph pickup that depends for its operation on the variation of its electrical capacitance.

capacitor-run motor — A single-phase induction motor using a capacitor in series with a second primary winding displaced 90° from the first. This winding remains in the circuit full time, converting the machine to polyphase operation.

capacitor series resistance — An equivalent resistance in series with a pure capacitance that gives the same resultant losses as the actual capacitor. The equivalent circuit does not represent the variation in capacitor losses with frequency.

capacitor speaker — *See* electrostatic speaker.

capacitor-start motor — 1. An ac split-phase induction motor in which a capacitor is connected in series with an auxiliary winding to provide a means of starting. The auxiliary circuit opens when the motor reaches a predetermined speed. 2. A type of motor that provides greater starting torque for a split-phase induction motor by connecting a capacitor in series with the auxiliary starting winding to provide a leading current vector of 90°, which momentarily converts the motor to a polyphase induction motor.

capacitor voltage — The voltage across the terminals of a.

capacity — 1. The current-output capability of a cell or battery a period of time. Usually expressed in ampere-hours (amp-hr). 2. Capacitance. 3. The limits, both upper and lower, of the items or numbers that may be processed in a computer register, in the accumulator. When quantities exceed the capacity, a computer interrupt develops and requires special handling. 4. The total quantity of data that a part of a computer can hold or handle, usually expressed as words per unit of time. 5. The capability of a specific system to store data, accept transactions, process data, and generate reports. 6. In a calculator, the maximum number of digits that can be entered as one factor or obtained in a result. In most machines, the capacity is equivalent to the number of digits in the display. In a few machines, it is larger than the number of digits in the display and the flip-flop key is used to show the full result. 7. The maximum number of bits that a system can process, or transmit, per second; e.g., $C = H/T$, where T is time required for the processing or transmission and H is information content in bits.

capillary tool — A tool used in bonding; the wire is fed to the bonding surface of the tool through a bore located along the long axis of the tool.

capstan — 1. The driven spindle or shaft in a tape recorder — sometimes the motor shaft itself — that rotates against the tape (which is backed up by a rubber pressure or pinch roller), pulling it through the machine at a constant speed during recording and playback modes of operation. The rotational speed and diameter of the capstan thus determine the tape speed. 2. A revolving shaft or flangeless pulley that drives the tape by squeezing it against a pinch roller, and that controls the rate at which tape passes over the heads of the tape recorder or deck. 3. A rotating shaft that is connected to the motor in a tape recorder. It moves the tape at constant speed across the heads. The tape is pressed against the capstan by a pinch roller.

capstan idler — *See* pressure roller.

captive screw — Screw-type fastener that is retained when unscrewed and cannot easily be separated from the part it secured.

capture area — The area of the antenna elements that intercept radio signals.

capture bandwidth — The frequency range over which an unlocked, free-running oscillator can be brought into lock by either phase- or injection-locking techniques.

capture effect — 1. The selection of the stronger of two frequency-modulated signals of the same frequency, with the complete rejection of the weaker signal. If both signals are of equal strength, both may be accepted and no intelligible signal will result. 2. An effect occurring in a transducer (usually a demodulator) whereby the input wave having the largest magnitude controls the output.

capture range — The range of frequencies over which a phase-locked loop can detect a signal on the input and respond to it. This is sometimes called the lock-in range. *See* acquisition range.

capture ratio — 1. The ability of a tuner to reject unwanted FM stations and interference on the same frequency as a desired one, measured in dB. The lower the figure, the better the performance of the tuner. 2. The power ratio of two signals in the same channel required to keep the signal/interference ratio to a value of 30 dB referred to 100-percent modulation and 1-mV input signal level. The ratio of the powers of the two input signals is expressed in decibels; the smaller the dB number the better the capture ratio. Topflight tuners have a value as low as 1 dB, but 4.5 dB is usually sufficient. 3. A measure of an FM tuner's ability to discriminate against weaker signals arriving on the same frequency as the desired one.

carbon — One of the elements, consisting of a nonmetallic conductive material occurring as graphite, lampblack, diamond, etc. Its resistance is fifty to several hundred times that of copper and decreases as the temperature increases.

carbon arc — An electric discharge between two carbon rods that are touched together to start the arc and then separated slightly. The light comes from the heated carbon vapor. High-intensity arcs use cored carbons, the core being filled with the oxides of thorium and cerium, which radiate brightly when heated.

carbon brush—A current-carrying brush made of carbon, carbon and graphite, or carbon and copper.

carbon-composition resistor—Hot- or cold-molded fixed resistor made from mixtures of granulated carbon and ceramic binder. In some versions the composition forms a monolithic structure; in others the composition is thickly applied to a ceramic core. Hot-molded carbon-composition resistors are specified where low-cost, reliable resistors with tolerances of ±5 and ±10 percent are acceptable.

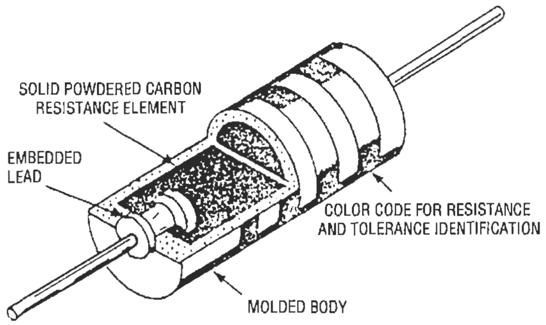

Carbon-composition resistor.

carbon-contact pickup—A phonograph pickup that depends for its operation on the variation in resistance of carbon contacts.

carbon-film resistor—1. A resistor formed by vacuum-depositing a thin carbon film on a ceramic form. 2. Carbon-film resistors are general-purpose, low-cost types with axial leads. Specification characteristics generally match those of carbon-composition resistors but at a lower cost.

carbonize—To coat with carbon.

carbonized filament—A thoriated-tungsten filament treated with carbon. A layer of tungsten carbide formed on the surface slows down the evaporation of the active emitting thorium and thus permits higher operating temperatures and much greater electron emission.

carbonized plate—An electron-tube anode that has been blackened with carbon to increase its heat dissipation.

carbon microphone—A microphone that depends for its operation on the variation in resistance of carbon contacts.

carbon-pile regulator—An arrangement of carbon discs whose series resistance decreases as more pressure or compression is applied.

carbon resistor—Also called composition resistor. A resistor consisting of carbon particles that are mixed with a binder molded into a cylindrical shape, and then baked. Terminal leads are attached to opposite ends. The resistance of a carbon resistor decreases as the temperature increases.

carbon transfer recording—A type of facsimile recording in which carbon particles are deposited on the record sheet in response to the received signal.

Carborundum—A compound of carbon and silicon used in crystals to rectify or detect radio waves.

carcinotron—A voltage-tuned, backward-wave oscillator tube used to generate frequencies ranging from UHF up to 100 GHz or more.

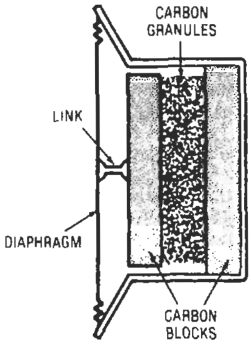

Carbon microphone.

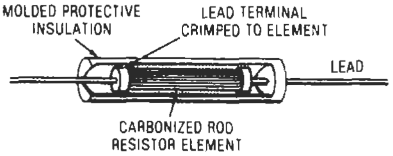

Composition type.

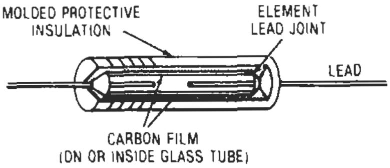

Deposited-film type.

Carbon resistor.

card—Nonpreferred term for printed circuit board. *See* printed circuit board.

card bed—A mechanical device for holding punch cards to be transported past the punching and reading stations.

card code—An arbitrary code in which holes punched in a card are assigned numeric or alphabetic values.

card column—One of 20 to 90 single-digit columns in a tabulating card. When punched, a column contains only one digit, one letter, or one special code.

card-edge connector—Also called edgeboard connector. A connector that mates with printed wiring leads running to the edge of a printed circuit board.

Cardew voltmeter—The earliest type of hot-wire instrument. It consisted of a small-diameter platinum-silver wire sufficiently long to give a resistance high enough to be connected directly across the circuit being measured. The wire was looped over pulleys and it expanded as current flowed, causing the pointer to rotate.

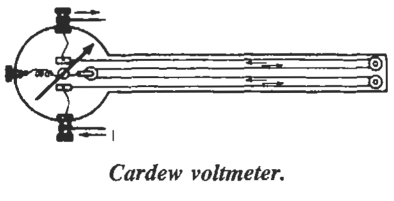

Cardew voltmeter.

card face—The printed side of a punched card if only one side is printed.

card feed—A mechanism that moves punch cards, one at a time, into a machine.

card field—On a punch card, the fixed columns in which the same type of information is routinely entered.

card hopper—*See* card stacker.

cardiac monitor—An instrument that usually has an oscilloscope display of the heart wave, and combines the features of several cardiac instruments, such as an electrocardiograph, cardiotachometer, etc. May also allow upper and lower limits to be set and trigger audible and/or visual alarms when these limits are exceeded.

cardiac pacemaker—1. A device that controls the frequency of cardiac contractions. 2. A device that stimulates the heart and controls its rhythm by means of electrodes placed on the chest wall or implanted under the skin.

cardiac stimulator—Also called a pacemaker. An electronic device (sometimes implanted in the patient) that supplies electric pulses to stimulate regular heart action.

card image—1. A representation in storage of the holes punched in a card, in such a manner that the holes are represented by one binary digit and the unpunched spaces are represented by the other binary digit. 2. In machine language, a duplication of the data contained in a punch card.

cardiograph—An instrument (or recording of instruments) for measuring the form or force of heart motion.

cardioid—1. A heart-shaped polar response pattern, with strong rejection to signals arriving from the rear. 2. The quasi-heart-shaped sensitivity pattern of most unidirectional microphones. Hypercardioid and supercardioid microphones have basically similar patterns, but with longer, narrower areas of sensitivity at the front and slightly increased rear sensitivity. 3. A heart-shaped pattern that is typical of some directional antennas or microphones.

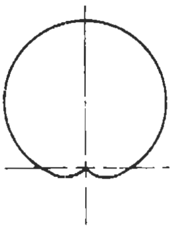

Cardioid diagram.

cardioid microphone—A microphone that has a heart-shaped response pattern that gives nearly uniform response for a range of about 180° in one direction and a minimum response in the opposite direction.

cardiostimulator—A device used to stimulate the heart and/or regulate its beat. *See also* pacemaker; defibrillator.

cardiotachometer—A measuring instrument that provides a meter reading proportional to the rate at which the heart beats.

card jam—A pile-up of cards in a machine.

card machine—A machine used to transfer information from or to punched cards.

card programmed—The capability of being programmed by punched cards.

card punch—1. A device to record information in cards by punching holes in the cards to represent letters, digits, and special characters. 2. Device used in data handling systems to enter data on cards according to a desired code.

card reader—1. A device designed to read information from punched cards and convert each hole into an electrical impulse for use in a computer system. 2. Device used in data handling systems to sense data on punched cards via a mechanical or photoelectric technique. 3. Equipment that takes a stack of punched cards and reads the information on them into the computer's memory or onto magnetic tape or disk for future reference. 4. A system that generally uses a photodetector to read punched cards for information by sensing the light transmitted through the punched holes.

card row—On a punched card, one of the horizontal lines of punching positions.

card sensing—The process of sensing or reading the information in punched cards and converting this information, usually into electrical pulses.

card stacker—A mechanism that stacks cards in a pocket or bin after they have passed through a computer. Also called card hopper.

card-to-tape—Having to do with equipment that transfers information directly from punched cards to punched tape or magnetic tape.

carillon—A bell tower designed to play from a keyboard. In an organ, this may be achieved by tube synthesis of bell-like tones struck with felt hammers, or completely electronically.

Carnot theorem—A thermodynamic principle that states that a cycle continuously operating between a low temperature and a high temperature can be no more efficient than a reversible cycle operating between the same temperatures.

carriage tape—*See* control tape.

carrier—1. A wave of constant amplitude, frequency, and phase that can be modulated by changing amplitude, frequency, or pulse. 2. An entity capable of carrying an electric charge through a solid (e.g., holes and conduction electrons in semiconductors). 3. A wave that has at least one characteristic that can be varied from a known reference value by a modulation process. 4. That part of the modulated wave that corresponds to the unmodulated wave in a specified way. 5. A frequency on which a second, information-carrying signal is impressed. 6. Holder for electronic parts and devices that facilitates handling during processing, production, imprinting, or testing operations and protects such parts under transport. 7. An analog signal at a fixed amplitude and frequency that combines with an information-bearing signal in the modulation process to produce an output signal suitable for transmission. 8. A continuous signal modulated with a second, information-carrying signal.

carrier-amplitude regulation—The change in amplitude of the carrier wave in an amplitude-modulated transmitter when symmetrical modulation is applied.

carrier band—In CD-4. discs, the left- or right-channel's band that contains musical information recorded at very high frequencies (in the 20–45 kilohertz range). In playback, the demodulator recovers those frequencies, which have been frequency modulated.

carrier beat—In facsimile transmission, an undesirable heterodyne of signals, each synchronous with a different stable oscillator, causing a pattern in received copy. When one or more of the oscillators is fork controlled, this is called fork beat.

carrier chrominance signal—In color television, sidebands of a modulated chrominance subcarrier, plus any unsuppressed subcarrier, added to the monochrome signal to convey color information.

carrier color signal — In color television, sidebands of a modulated chrominance subcarrier, plus the chrominance subcarrier, if not suppressed, that are added to the monochrome signal to convey color information.

carrier concentration — The number of carriers in a cubic centimeter of semiconductor material.

carrier control — A control by the presence or absence of an rf carrier.

carrier current — 1. The current associated with a carrier wave. 2. High-frequency alternating current superimposed on ordinary telephone, telegraph, or power-line frequencies. The carrier may be tone modulated to operate switching relays or transmit data.

carrier-current communication — The superimposing of a high-frequency alternating current on ordinary telephone, telegraph, and power-line frequencies for telephone communication and control.

carrier-current control — Remote control in which the receiver and transmitter are coupled together through power lines.

carrier-current transmitter — A device that transmits signals via the standard ac power lines.

carrier frequency — 1. The frequency (hertz) of the wave modulated by the intelligence wave; usually a radio frequency (rf). 2. The reciprocal of the period of a periodic carrier. 3. The frequency of the unmodulated fundamental output from a radio transmitter.

carrier-frequency interconnection — In the formation of carrier networks, the transfer of groups of channels between terminals of wire-line cable or radio carrier systems at carrier frequencies.

carrier-frequency peak-pulse power — The power averaged over that carrier-frequency cycle that occurs at the maximum pulse of power (usually half the maximum instantaneous power).

carrier-frequency pulse — A carrier that is amplitude modulated by a pulse. The amplitude of the modulated carrier is zero before and after the pulse.

carrier-frequency range — The continuous range of frequencies within which a transmitter may normally operate. A transmitter may have more than one carrier-frequency range.

carrier-frequency stereo disc — A stereo disc with two laterally cut channels. One channel is cut in the usual manner. The second channel is employed to frequency modulate a supersonic carrier frequency. The playback cartridge delivers the signal for one channel plus the carrier frequency containing the other channel. The latter must then be demodulated to obtain the second channel.

carrier injection — The process whereby light is emitted at the junction of n- and p-type semiconductors when an external electric source is applied to drive the electrons and the holes into the junction.

carrier-isolating choke coil — An inductor inserted in series with a line on which carrier energy is applied to impede the flow of carrier energy beyond that point.

carrier leak — The carrier-frequency signal remaining after suppression in a suppressed carrier system.

carrier level — The strength, expressed in decibels, of an unmodulated carrier signal at a particular point in a system.

carrier lifetime — The time required for excess carriers doped into a semiconductor to recombine with other carriers of the opposite sign.

carrier line — A transmission line used for multiple-channel carrier communications.

carrier loading — The insertion of additional lump inductance in a cable section of a transmission line utilized for carrier transmission up to about 35 kHz. Loading serves to minimize impedance mismatch between cable and open wire and to reduce the cable attenuation.

carrier mobility — The average drift velocity of carriers per unit electric field in a homogeneous semiconductor. The mobility of electrons is usually different from that of holes.

carrier noise — Undesired variation of a radio-frequency carrier signal in the absence of intended modulation. Also called residual modulation.

carrier noise level — Also called residual modulation. The noise level produced by undesired variations of a radio-frequency signal in the absence of any intended modulation.

carrier on microwave — A means of transmitting many voice messages on one microwave radio channel. Transmission is point to point by microwave antennas mounted on towers or tall buildings.

carrier on wire — A means widely used by the telephone companies to transmit many voice messages on a single pair of wires. Circuits involving one or more carrier links never evidence dc continuity.

carrier-operated anti-noise device — A device whose purpose is to mute the audio output of a receiver during standby or intervals of no carrier.

carrier power — The rf power output of an AM transmitter when not modulated.

carrier power output rating — The power available at the output terminals of a transmitter when the output terminals are connected to the normal load circuit or to a circuit equivalent thereto.

carrier repeater — An assembly, including an amplifier (or amplifiers), filters, equalizers, level controls, etc., used to raise the carrier signal level to a value suitable for traversing a succeeding line section while maintaining an adequate signal-to-noise ratio.

carriers — Entities that carry an electrical charge and are also able to move relatively freely through a crystal lattice. The two most commonly encountered carriers are conduction band electrons, which are negatively charged, and valence band holes, which are positively charged.

carrier shift — 1. The transmission of radio teletypewriter messages by shifting the carrier frequency in one direction for a marking signal and in the opposite direction for a spacing signal. 2. The condition resulting from imperfect modulation, whereby the positive and negative excursion of the envelope pattern are unequal, thus effecting a change in the power associated with the carrier.

carrier signaling — In a telephone system, the method by which ringing, busy signals, or dial-signaling relays are operated by the transmission of a carrier-frequency tone.

carrier storage time (of a switching transistor) — The time interval between the beginning of the fall of the pulse applied to the input terminals and the beginning of the fall of the pulse generated by charge carriers at the output terminals. (The time is generally measured between the 90-percent values of the two pulse amplitudes.)

carrier suppression — The method of operation in which the carrier wave is not transmitted.

carrier swing — The total deviation of a frequency- or phase-modulated wave from the lowest to the highest instantaneous frequency.

carrier system — A means of obtaining a number of channels over a single wideband communication path by modulating each channel on a different carrier frequency at the originating end and demodulating at the receiving point to restore the signals to their original form.

carrier tap choke coil — A carrier-isolating choke coil inserted in series with a line tap.

carrier tap transmission choke coil—An inductor inserted in series with a line tap to control the amount of carrier energy flowing into the tap.

carrier telegraphy—The form of telegraphy in which the transmitted signal is formed by modulating the alternating current, under control of the transmitting apparatus, before supplying it to the line.

carrier telephony—Ordinarily applied only to wire telephony. That form of telephony in which carrier transmission is used, the modulating wave being at an audio frequency.

carrier terminal—Apparatus at one end of a carrier transmission system, whereby the processes of modulation, demodulation, filtering, amplification, and associated functions are effected.

carrier-to-noise ratio (C/N)—1. Ratio of the magnitude of the carrier to the magnitude of the noise after selection and before any nonlinear process such as amplitude limiting and detection. This ratio is expressed in many different ways—for example, in terms of peak values in the case of impulse noise, and in terms of root-mean-square values in the case of random noise. 2. The ratio of the received carrier power to the noise power in a given bandwidth. The C/N is an indicator of how well an earth receiving station will perform in a particular location and is calculated from satellite power levels, antenna gain, and the combined antenna and LNA noise temperature.

carrier-transfer filter—A group of filters arranged to form a carrier-frequency crossover or bridge between two transmission circuits.

carrier transmission—1. That form of electrical transmission in which a single-frequency wave is modulated by another wave containing the information. 2. A system for transmitting many voice channels over a common telephone circuit.

carrier-type dc amplifier—An amplifier system that converts a dc input to modulated ac, amplifies, and synchronously detects to provide amplifier dc output.

carrier wave—1. The single-frequency transmitted wave that is modulated by another wave containing the information. 2. The basic frequency or pulse repetition rate of a signal, bearing no intrinsic intelligence until it is modulated by another signal that does bear intelligence. A carrier may be amplitude, phase, or frequency modulated.

carry—1. A signal or expression produced in an electronic computer by an arithmetic operation on a one-digit place of two or more numbers expressed in positional notation and transferred to the next higher place for processing there. 2. A signal or expression—as defined in (1)—that arises when the sum of two digits in the same digit place equals or exceeds the base of the number system in use. If a carry into a digit place will result in a carry out of the same digit place and the normal adding circuit is bypassed when this new carry is generated, the result is called a high-speed or stand-on-nines carry. If the normal adding circuit is used, the result is called a cascade carry. If a carry resulting from the addition of carries is not allowed to propagate, the process is called a partial carry; if it is allowed to propagate, it is called a complete carry. A carry generated in the most significant digit place and sent directly to the least significant place is called an end-around carry. 3. In direct subtraction, a signal or expression—as defined in (1)—that arises when the difference between the digits is less than zero. Such a carry is frequently called a borrow. 4. The action of forwarding a carry. 5. The command directing a carry to be forwarded.

carry look-ahead—A circuit that predicts the final carry from propagate and generate signals supplied by partial adders. Used to speed up binary addition significantly by eliminating the carry propagation (or ripple) delay.

carry time—The time required for a computer to transfer a carry digit to the next higher column and add it there.

cartridge—1. The electromechanical transducer of a phonograph pickup head that converts stylus vibrations to an electrical signal. It is generally detachable and fits into the head shell of a pickup. Most cartridges are either magnetic or piezoelectric types. 2. Generally, any enclosed package containing a length of magnetic tape and its basic winding receptacles, designed to eliminate the need for handling or threading the tape. Specifically, the word *cartridge* is used to describe that variety of package that contains a continuous (endless) loop of tape on a single reel. 3. A film or tape magazine containing only one spool. 4. A plastic container that holds recording tape for easy loading into a matching recorder or player, especially the eight-track cartridge.

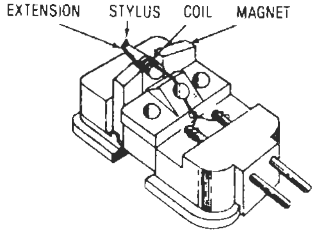

Cartridge.

cartridge fuse—1. A tubular fuse whose end caps are enclosed in a glass or composition insulating tube to confine the arc or vapor when the fuse blows. 2. A short tube of fiber containing a fusible link or wire that is connected to metallic ferrules at the ends of the tube. Serves to interrupt excessive currents by melting of the fusible link.

cascadable counter—A logic counting block that has available the necessary connections to permit more than one counter to be operated in series, thus increasing the modulus of the counter subsystem.

cascade—Also called tandem. An arrangement of two or more similar circuits or amplifying stages in which the output of one circuit provides the input of the next.

cascade amplification—In a series of amplifiers, amplification by each of the preceding output.

cascade amplifier—1. A multiple-stage amplifier in which the output of each stage is connected to the input of the next stage. 2. A multistage amplifier whose stages are forward coupled in succession.

cascade-amplifier klystron—A klystron that has three resonant cavities to provide increased power amplification and output. The extra resonator is located between the input and output resonators and is excited by the bunched beam energizing from the first resonator gap, thereby producing further bunching of the beam.

cascade connection—Two or more similar component devices arranged in tandem, with the output of one connected to the input of the next.

cascade control—Also called piggyback control. An automatic control system in which the control units, linked in chain fashion, feed into one another in succession. Each unit thus regulates the operation of the next in line.

cascaded carry—In a computer a system of executing the carry process in which carry information cannot

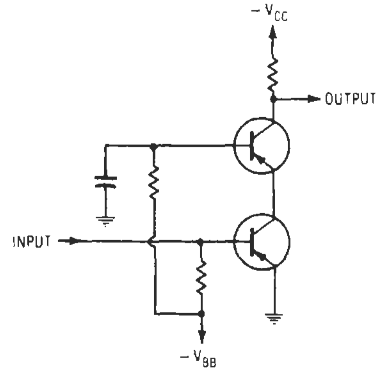

Cascade amplifier.

be passed on to place $(N + 1)$ unless the Nth place has received carry information or produced a carry.

cascaded feedback canceler — Also called velocity-shaped canceler. A sophisticated moving-target indicator canceler that provides clutter and chaff rejection.

cascaded systems — Multistorage operations; the input to each stage is the output of a preceding stage, thereby causing interdependence among the stages.

cascaded thermoelectric device — A thermoelectric device having two or more stages that are arranged thermally in series.

cascade image tube — An image tube that functions in low-light-level conditions by virtue of its series of stacked sections wherein the output of one section becomes the input for the next.

cascading — The connecting of two or more circuits in series so that the output from one provides the input to the next.

cascode amplifier — An amplifier using a neutralized grounded-cathode input stage followed by a grounded-grid output stage. The circuit has high gain, high input impedance, and low noise.

CASE — Abbreviation for computer-aided software engineering. A working environment consisting of programs and development tools that help automate the design and implementation of programs and procedures for business, engineering, and scientific computer systems.

case — In a computer, a set of data for use in a particular program.

case pressure — The total differential pressure in the internal cavity of a transducer and the ambient pressure. The term is commonly used to summarize the limiting combined differential and/or line-pressure capabilities of differential transducers.

case temperature — The temperature on the surface of the case at a designated point.

Cassegrain antenna — 1. An antenna whose feed is positioned near the vertex of the reflector, with a small subreflector placed near the focal point. The feed illuminates the subreflector, and the subreflector redirects the waves toward the main reflector, which then forms the radiated beam (called the secondary beam or pattern). The shapes of the subreflector and main reflector are so chosen that the secondary rays will emerge parallel to the main-reflector axis. 2. An antenna that utilizes a subreflector at the focal point that reflects energy to or from a feed located at the apex (center) of the main reflector.

Cassegrain feed — 1. A method of feeding a reflector antenna in which a waveguide located in the center of the main reflector feeds energy to a small reflector, which reflects it in turn to the main reflector. 2. An antenna feed design that includes a primary reflector, the dish, and a secondary reflector that redirects microwaves via a waveguide to a low-noise amplifier.

cassette — 1. A thin, flat, rectangular enclosure that contains a length of narrow magnetic recording tape permanently affixed to two flangeless floating reels that wind and unwind the tape while it passes an external recording and/or playback head. 2. A flat enclosure that contains two flangeless reels that link a narrow magnetic tape. 3. A sealed instant-load cartridge containing a length of tape and separate supply and takeup reels or hubs. Cassettes, unlike continuous-loop cartridges, can be rewound as well as fast-forwarded. 4. A film or tape magazine containing two spools. 5. Most commonly applied to the compact cassette developed by Philips, but also to a variety of micro and mini cassette systems that are not mutually compatible. 6. Recently applied to mass storage requirements of microcomputers and minicomputers. A digital cassette is certified for digital recording, which differs from the audio recording requirements. 7. A miniature reel-to-reel tape system retaining the flexibility and freedom of back-and-forth tape movement provided by a reel system; it eliminates the inconvenience of tape threading. 8. An open metal or plastic carrier used on IC production lines for moving groups of wafers.

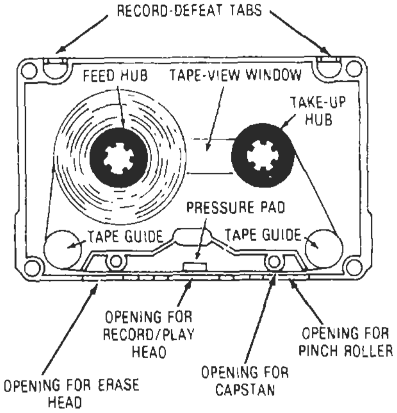

Cassette.

cassette recorder — A magnetic tape recording and playback device for entering or storing programs.

casting-out-nines check — A partial verification of an arithmetical operation on two or more numbers. It involves casting out nines from the number and from the results.

CAT — Abbreviation for computer-aided tomography, computer-assisted tomography, and computerized axial tomography. Literally, section graphics — the graphic display of a cross section of a piece of material or of the human body. Unlike ordinary X-ray photographs, which are produced directly by X-ray radiation on a photosensitive film (resulting in shadows on the film) computer-aided tomography displays are created by a computer after it processes the sensed X-ray signals. The big advantage of CAT is that a given display is not influenced by material in front of or behind the plane of interest, as is the case with ordinary X-ray photos. In fact, a three-dimensional picture can be obtained from multiple CAT displays, each displaced a small, finite distance from the adjacent one.

catadioptric—Optical process using both reflection and refraction of light.

catalog—An ordered compilation of descriptions of items, including sufficient information to afford access to the items. *See also* directory.

catalog search—A computerized data retrieval technique involving the search of keywords stored as key nouns or characteristics of the product. A display shows all items from the database with the specified keywords.

catalytic converter—A device that enhances certain chemical reactions that help to reduce the levels of undesirable exhaust gases.

cataphoresis—The migration toward the cathode of particles suspended in a liquid; movement of the particles is caused by the influence of an electrostatic field.

catastrophic failure—1. A sudden failure without warning, as opposed to degradation failure. 2. A failure the occurrence of which can prevent the satisfactory performance of an entire assembly or system.

catcher—In a velocity-modulated vacuum tube, an electrode on which the spaced electron groups induce a signal. The output of the tube is taken from this element.

catching diode—A diode connected to act as a short circuit when its anode becomes more positive than its cathode; the diode then tends to prevent the voltage of a circuit terminal from rising above the voltage at the cathode.

categorization—The process by which multiple addressed messages are separated to form individual messages for single addresses.

catena—A chain or connected series.

catenate—*See* concatenate.

cathamplifier—A push-pull vacuum-tube amplifier in which the push-pull transformer is in the cathode circuit.

cathode—1. In an electron tube, the electrode through which a primary source of electrons enters the interelectrode space. 2. General name for any negative electrode. 3. The lower-potential electrode of a corrosion cell, in which the action of the corrosion current may reduce or eliminate corrosion, or the negatively charged metallic parts of an impressed current system. 4. When a semiconductor diode is biased in the forward direction, that terminal of the diode that is negative with respect to the other terminal. 5. The negative electrode of a polar capacitor. 6. The negative pole of a plating apparatus at which positively charged ions leave the plating solution. A metal is deposited on the cathode. *See* anode. 7. In electrolytic plating, the workpiece being plated.

cathode activity—Measure of the efficiency of an emitter. The mathematical relationship between two values of emission current measured under two conditions of cathode temperature.

cathode bias—A method of biasing a vacuum tube by placing the biasing resistor in the common cathode-return circuit, thereby making the cathode more positive—rather than the grid more negative—with respect to ground.

cathode-coupled amplifier—A cascade amplifier in which the coupling between two stages is accomplished by a common cathode resistor.

cathode coupling—The use of an input or output element in the cathode circuit for coupling energy to another stage.

cathode current—*See* electrode current.

cathode-current density—The current per square centimeter of cathode area, expressed as amperes or milliamperes per centimeter squared.

cathode dark space—Also called Crookes' dark space. The relatively nonluminous region between the cathode and negative glow in a glow-discharge cold-cathode tube.

cathode emission—The process whereby electrons are emitted from the cathode structure.

cathode follower—Also called grounded-plate amplifier. A vacuum-tube circuit in which the input signal is applied to the control grid, and the output is taken from the cathode. Electrically, such a circuit possesses a high input impedance and a low output impedance characteristic and a gain of less than unity. The equivalent circuit using a transistor is called an emitter follower.

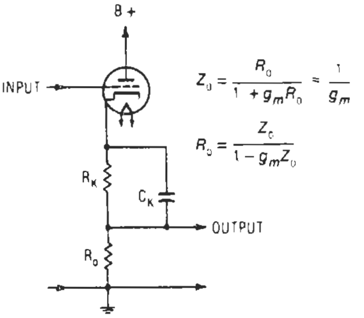
B +
INPUT
$Z_0 = \frac{R_0}{1 + g_m R_0} \approx \frac{1}{g_m}$
$R_0 = \frac{Z_c}{1 - g_m Z_0}$
R_K
C_K
OUTPUT
R_0

Cathode follower.

cathode glow—The apparent luminosity or glow that immediately envelops the cathode in a gas-discharge tube when operating at low pressures. The glow increases as the pressure decreases.

cathode guide—The element of a glow tube used in switching the neon glow from one indicated number to the next.

cathode heating time—The time required for the cathode to attain a specified condition, for example, a specified value of emission or a specified rate of change in emission.

cathode interface—A resistive and capacitive layer formed between the nickel sleeve and oxide coating of an indirectly heated cathode. Raising the cathode temperature will largely nullify the layer.

cathode keying—A method of keying a radiotelegraph transmitter by opening the plate return lead to the cathode or filament center tap.

cathode luminous sensitivity (of a multiplier phototube)—The photocathode current divided by the incident luminous flux.

cathode modulation—A form of amplitude modulation in which the modulating voltage is applied to the cathode circuit.

cathode pulse modulation—Modulation produced in an amplifier or oscillator by applying externally generated pulses to the cathode circuit.

cathode radiant sensitivity—The current leaving the photocathode divided by the incident radiant power of a given wavelength.

cathode ray—A stream of electrons emitted, under the influence of an electric field, from the cathode of an evacuated tube or from the ionized region nearby.

cathode-ray charge storage tube—A charge storage tube in which the desired information is written by means of a cathode-ray beam.

cathode-ray instrument—*See* electron-beam instrument.

cathode-ray oscillograph—An apparatus capable of producing, from a cathode-ray tube, a permanent record of the value of an electrical quantity as a function of time.

cathode-ray oscilloscope—Abbreviated CRO. 1. A test instrument that, when properly adjusted, makes possible the visual inspection of alternating-current signals. It consists of an amplifier, time-base generating circuits, and a cathode-ray tube for transformation of electrical energy into light energy. 2. A system wherein a supplied signal causes deflection of a CRT electron beam, thus forming a visible trace on the phosphor CRT screen that allows examination of electrical quantities.

cathode-ray output—A term used in data processing to describe a cathode-ray tube that displays graphic or character data.

cathode-ray storage—An electrostatic data storage device in which a cathode-ray beam provides access to the data.

cathode-ray tube—Abbreviated CRT. 1. A vacuum tube in which its electron beam can be focused to a small cross section on a luminescent screen and can be varied in position and intensity to produce a visible pattern. 2. A vacuum tube with an electron gun at one end and a fluorescent screen at the other. Electrons emitted from a heated cathode are accelerated by a series of annular anodes at progressively higher positive voltages. The electron beam is then deflected by two pairs of electrostatically charged plates between the gun and the screen. Electromagnets are often used in place of the deflector plates.

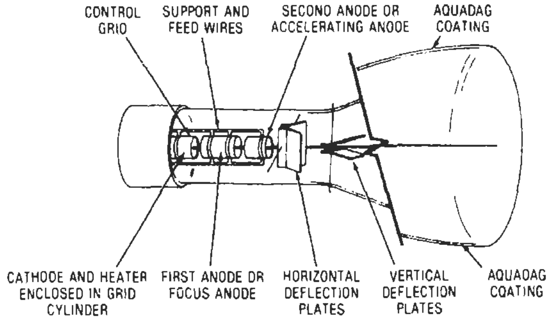
CONTROL GRID
SUPPORT AND FEED WIRES
SECOND ANODE OR ACCELERATING ANODE
AQUADAG COATING
CATHODE AND HEATER ENCLOSED IN GRID CYLINDER
FIRST ANODE OR FOCUS ANODE
HORIZONTAL DEFLECTION PLATES
VERTICAL DEFLECTION PLATES
AQUADAG COATING

Cathode-ray tube.

cathode-ray-tube display—Abbreviated CRT display. 1. A device in which controlled electron beams are used to present data in visual form. 2. The data presentation produced by such a device. 3. In a calculator, a type of display resembling a small television tube. Usually, two to four rows of digits can be displayed simultaneously. 4. A high-speed device, similar to a television picture tube, that provides a visual nonpermanent display of system input/output data, such as instructions as they are being developed and data in storage. 5. A device, used to present data (alphanumeric, graphical, or a combination of the two), that incorporates a cathode-ray tube as the presenting element.

cathode-ray tuning indicator—Commonly called magic eye. A small-diameter cathode-ray tube that visually indicates whether an apparatus such as a radio receiver is tuned precisely to a station.

cathode resistor—A resistance connected in the cathode circuit of a tube so that the voltage drop across it will supply the proper cathode-biasing voltage.

cathode spot—On the cathode of an arc, the area from which electrons are emitted at a current density of thousands of amperes per square centimeter and where the temperature of the electrode is too low to account for such currents by thermionic emission.

cathode sputtering—The method of disintegrating the substance of the cathode by bombarding it with ions and then depositing it on another electrode or electron-tube envelope. *See also* sputtering, 1.

cathodic protection—Corrosion inhibition employed in protecting underground (or underwater) metal from electrochemical corrosion; it involves making the metal cathodic (negative) to a buried (or immersed) anode. Two methods are available for the required polarization: a dc power supply applying a positive polarity to a stainless steel or platinum plated titanium anode, or a dc self-generating sacrificial anode of zinc, magnesium, or aluminum alloy.

cathodofluorescence—Fluorescence that results from a material's exposure to cathode rays.

cathodoluminescence—Luminescence produced by the bombardment with high-energy electrons of a metal in a vacuum. Small amounts of the metal are vaporized in an excited state by the bombardment and emit radiation characteristic of the metal.

cation—1. A positive ion that moves toward the cathode in a discharge tube, electrolytic cell, or similar equipment. The corresponding negative ion is called an anion. 2. An atom with a deficiency of electrons and therefore having a positive charge.

CATV—Abbreviation for community antenna television or cable television. A system of distributing TV signals to homes by cable.

catwhisker—A small, pointed wire used to make contact with a sensitive area on the surface of a crystal or semiconductor.

Cauer-elliptic filter—Also known as an elliptic or elliptical filter. A transfer function having an extremely high rate of attenuation (a very fast rolloff) near the corner frequency. Exhibits passband ripple and a renewal of gain beyond the stopband.

cavitation—The production of gas-filled cavities in a liquid when the pressure is reduced below a certain critical value with no change in the temperature. Ordinarily this is a destructive effect because the high pressures produced when these cavities collapse often damage mechanical components of hydraulic systems, but the effect is turned to advantage in ultrasonic cleaning.

cavitation noise—The noise produced in a liquid by the collapse of the bubbles created by cavitation.

cavity—A metallic enclosure that can be made to resonate at a desired microwave frequency. Primarily used to describe a cavity filter, which is a highly selective tuning element at microwave frequencies that may be used as the frequency-determining element of an oscillator or as a low-pass, bandpass, or highpass filter. Generally of fixed frequency, or may be mechanically tunable over a very limited frequency range.

cavity filter—A selective tuned device having the proper coupling means for insertion into a transmission line to produce attenuation of unwanted off-frequency signals.

cavity impedance—The impedance that appears across the gap of the cavity of a microwave tube.

cavity magnetron—A magnetron having a number of resonant cavities forming the anode; used as a microwave-transmitting oscillator.

cavity oscillator—Abbreviated CO. An oscillator in which the primary frequency-determining element is either a waveguide or coaxial cavity. Oscillator frequency can be mechanically tuned and voltage tuned (via a tuning varactor) over a relatively narrow band.

cavity radiation—The radiation (heat) emerging from a small hole leading to a constant-temperature enclosure. Such radiation is identical with blackbody radiation at the same temperature, no matter what the nature of the inner surface of the enclosure.

cavity resonator—1. A space that is normally bounded by an electrically conducting surface and in which oscillating electromagnetic energy is stored; the resonant frequency is determined by the geometry of the enclosure. 2. A section of coaxial line or waveguide completely enclosed by conducting walls; it is often made variable for use as a wavemeter.

cavity resonator frequency meter—A cavity resonator used for determining the frequency of an electromagnetic wave.

cavity-tuned, absorption-type frequency meter—A device used for measuring frequency. Its operation depends on the use of an enclosure with a conductive inner wall; the resonant frequency of the wall is determined by its internal dimensions.

cavity-tuned, heterodyne-type frequency meter—A device for measuring frequency. Its operation depends on the use of an enclosure, the resonant frequency of which is determined by its internal dimensions.

cavity-tuned, transmission-type frequency meter—A device for measuring frequency. Its operation depends on the use of an enclosure with a conductive inner wall; the resonant frequency of the wall is determined by its internal dimensions.

C-band—1. Microwave band in which the wavelengths are at or near 5.6 cm. It includes the top two sidebands of the S-band and the bottom three sidebands of the X-band. 2. The band of frequencies between 4 and 8 GHz, with the 6- and 4-GHz band being used for satellite communications. Specifically, the 3.7 to 4.2 GHz satellite communication band is used as the downlink frequencies in tandem with the 5.925 to 6.425 GHz band that serves as the uplink.

CBASIC—A version of BASIC that runs on the CP/M operating system. It is a structured language often preferred by programmers working in BASIC.

C battery—Also called grid battery. The energy source that supplies the voltage for biasing the grid of a vacuum tube.

CBC—Canadian Broadcasting Corporation. The government radio and television organization of Canada.

C-bias—*See* grid bias.

CCD—Abbreviation for charged-coupled device. 1. A semiconductor storage device in which an electrical charge is moved across the surface of a semiconductor by electrical control signals. Zeros or ones are represented by the absence or presence of a charge. A charge transfer system in which charges created by either an input diode or by an impinging photon are contained in MOS (metal-oxide semiconductor) or MIS (metal-insulated semiconductor) capacitors fabricated on a single crystal wafer. By varying electrode voltages successively, charge packets are moved from capacitor to capacitor to a single output amplifier. 2. Large buffer memory for minicomputer systems, interfacing between the magnetic storage disks or tape drives and the RAM, or as plug-compatible disk replacements. 3. A shift register formed by a string of closely spaced MOS capacitors. A CCD can store and transfer analog-charge signals—either electrons or holes—that may be introduced electrically or optically. The storing and transferring of charge occurs between potential wells at or near a silicon-silicon dioxide interface. The MOS capacitors, pulsed by a multiphase clock voltage, form these wells. For a three-phase, *n*-channel CCD, the charges transferred between potential wells are electrons. 4. Functionally, a shift-register memory for either analog or digital information in which the data is represented as stored charges. The charges in a CCD are stored in a linear array of potential wells, with the potential of each well controlled by a voltage applied to an isolated metal-oxide semiconductor (MOS) capacitor above the well. By applying a traveling voltage wave to this linear array of capacitors, any charges within the wells are pushed along from well to well. For a digital memory, the potential wells are either uncharged or fully charged to represent 0 and 1. To implement an analog memory, the charge is varied linearly in proportion to the sampled input voltage. Thus, analog delay lines as well as digital memories are implementable with CCD technology. Also known as a bucket brigade device because of the way charge is transferred from one cell to another in a recirculating fashion. 5. For imaging devices, a self-scanning semiconductor array that utilizes MOS technology, surface storage, and information transfer by digital shift register techniques.

CCIF—Abbreviation for International Telephone Consultative Committee.

CCIR—Abbreviation for International Radio Consultative Committee.

CCITT—1. Abbreviation for Comité Consulatif International Télégraphique et Téléphohique. The original French name of the committee that published international communications standards. Replaced by ITU. 2. Abbreviation for Consultative Committee for International Telephone and Telegraph. An international standards group that is a part of the International Telecommunications Union (ITU).

CCS—Abbreviation for continuous commercial service. Refers to the power rating of transformers, tubes, resistors, etc. Used for rating components in broadcasting stations and some industrial applications.

CCTV—Abbreviation for closed-circuit television.

CCTV camera—That part of a closed-circuit TV system that captures the picture and produces the video (picture) signal.

CCTV monitor—That part of a closed-circuit TV system that receives the picture from the CCTV camera and displays it on the picture tube.

ccw—Abbreviation for counterclockwise.

CD—Abbreviation for compact disc.

CD-4—1. A record-playback system for discrete discs. Invented by the Victor Co. of Japan (JVC) and developed by JVC and RCA Records, the system needs a demodulator and special cartridge with a special stylus for discrete four-channel playback. The system is not compatible with matrix quad discs and is not used for FM broadcasting. Also called quadradisc. 2. A recording and playback system similar in some respect to FM multiplex stereo broadcast and reception. Each wall of the record groove carries a single channel of information—left front plus left rear on the inner wall and right front plus right rear on the outer wall of the groove. In addition, each groove wall carries a 30-kHz FM subcarrier that is modulated by the front-minus-back difference signals that are needed to decode or demodulate the quadraphonic signal into four discrete channels. 3. A discrete four-channel disc recording and playback system, using a frequency-modulated 30-kHz carrier to convey additional information that can be combined with the audio output of the cartridge to produce four essentially independent program channels. Requires a cartridge frequency response to at least 45 kHz and a special demodulator.

CD-4 capability—The ability of a cartridge to reproduce the ultrasonic signals necessary for discrete four-channel disc reproduction using a CD-4 demodulator.

cdi—Abbreviation for collector-diffusion isolation.

CD-I—Abbreviation for compact disc-interactive. A home entertainment system based on a player that connects to any TV and stereo system, with information stored digitally on compact disc. It was introduced by Philips Consumer Electronics in 1991.

C-display—A type of radar display in which the signal is a bright spot, with the bearing as the horizontal coordinate and the elevation angle as the vertical coordinate.

CDMA—Abbreviation for code-division multiple access. Digital cellular system multiple-access standard that allows for higher capacity and greater security. Stations use spread-spectrum modulations and orthogonal codes to avoid interfering with one another. *See also* code-division multiple access.

CDPD—Abbreviation for Cellular Digital Packet Data. A standards-based technology for wireless communication of data.

CD-R—Abbreviation for compact disc-recordable (recordable CD). Same as CD-WO. CD format is compatible with CD-ROM and can be written to once and read many times.

CD-ROM—Abbreviation for compact disc-read-only memory. 1. A compact disc with a format for storing different types of information digitally, which can be played on a CD-ROM drive connected to a personal computer. 2. A compact 5-1/4-inch optical disc, typically used to store text, images, audio, video, and programs that can run on suitably equipped computers. CD-ROM drives come in single, double, and quad speeds (150 kbs, 300 Kbs, and 600 kbs, respectively). 3. A read-only optical disc capable of storing large amounts (up to 250,000 pages) of data.

CD-ROM drive—A peripheral device attached to a computer that allows it to read/play a CD-ROM disc. All CD-ROM players can also play back audio CDs, but require external headphones or speakers to hear them.

CD-WO—Compact disc-write once. Recordable compact disc. Same as CD-R.

ceiling—The maximum voltage that may be attained by an exciter under specified conditions.

celestial guidance—A system of guidance in which star sightings that are automatically taken during the flight of a missile provide position information used by the guidance equipment.

cell—1. A single unit that produces a direct voltage by converting chemical energy into electrical energy. 2. A single unit that produces a direct voltage by converting radiant energy into electrical energy; for example, a solar or photovoltaic cell. 3. A single unit that produces a varying voltage drop because its resistance varies with illumination. 4. Elementary unit of storage. 5. In corrosion processes, a source of electric potential that is responsible for corrosion. It consists of an anode and a cathode immersed in an electrolyte and electrically bonded together. The anode and cathode may be separate metals or dissimilar areas on the same metal. The different metals will develop a difference in potential that is accompanied by corrosion of the anode. When this cell involves an electrolyte, as it does in corrosion processes, it is referred to as an electrolytic cell. 6. The geographic area served by a single low-power transmitter/receiver. A cellular system's service area is divided into multiple cells.

cell counter—An electronic instrument used to count white or red blood cells or other very small particles.

cell-type enclosure—A prefabricated basic shielded enclosure of double-walled copper-mesh construction. The original screen-room design.

cell-type tube (tr, atr, and pre-tr tubes)—A gas-filled radio-frequency switching tube that operates in an external resonant circuit. A tuning mechanism may be incorporated into the external resonant circuit or the tube.

cellular—Type of mobile telephone service in which the geographic serving area is divided into subregions (cells), each with its own antenna and switching node.

cellular radio—The terrestrial mobile telephone technology that increases the number of available channels by dividing an area into cells; each cell may use the same frequencies as other cells, except that adjacent cells may not use the same frequencies.

cellular system—1. A mobile telephone system that divides large service areas into small cells, each with its own low-power transmitter. A telephone call is switched by computers from one transmitter to the next without interrupting the signal as a vehicle moves from cell to cell. Calls can be divided and frequencies reused over shorter intervals. 2. Method of mobile telephone service that divides radio communication service areas into small cells, or districts. The cellular approach

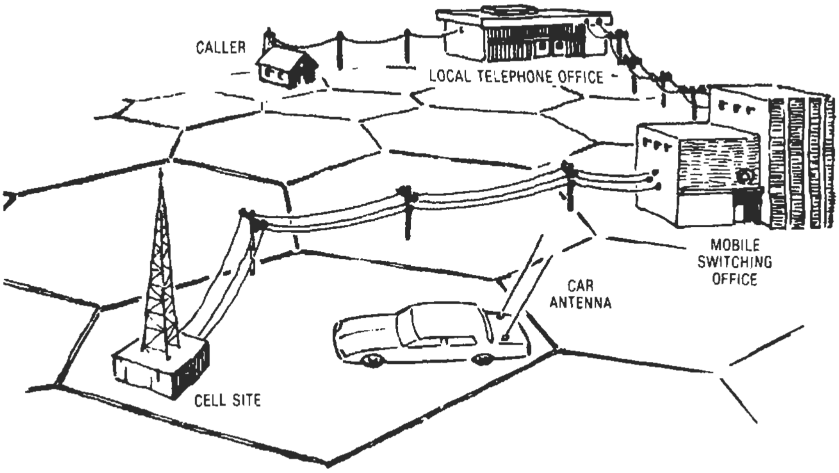

Cellular telephone system.

utilizes low-power transmitters that allow uninterrupted communications through sophisticated switching equipment linking the cells. This system, by reusing frequencies and standardizing service and equipment, has substantially improved mobile telephone service.

cellulose acetate — An inexpensive transparent plastic film used as the backing material for many recording tapes.

cellulose-nitrate disc — *See* lacquer disc.

Celsius temperature scale — Also called centigrade temperature scale. A temperature scale based on the freezing point of water defined as 0°C and the boiling point defined as 100°C, both under conditions of normal atmospheric pressure (760 mm of mercury).

cent — A measure of frequency, defined as equal to 100th of a semitone.

center-fed antenna — An antenna in which the feeder wires are connected to the center of the radiator.

center feed — 1. Attaching feeder wires or a transmission line to the center of the radiator of an antenna. 2. Connection of signal input terminals to the center of a coil.

center frequency — Also called resting frequency. 1. The average frequency of the emitted wave when modulated by a symmetrical signal. 2. The frequency at the center of a spectrum display (for linear frequency scanning). It is usually tunable. 3. Also called free-running frequency. The frequency at which a phase-locked loop operates when not locked onto an incoming (input) signal.

centering control — One of two controls used to shift the position of the entire image on the screen of a cathode-ray tube. The horizontal-centering control moves the image to the right or left, and the vertical-centering control moves it up or down. *See also* framing control.

centering diode — A clamping circuit used in some types of plan-position indicators.

centering magnet — A magnet that centers the televised picture on the face of the tube. Also called framing magnet.

center of gravity — A point inside or outside a body and around which all parts of the body balance each other.

center of mass — On a line between two bodies, the point around which the two bodies would revolve freely as a system.

center poise — Scale of viscosity for insulating varnishes.

center ring — The part that supports the stator in an induction-motor housing. The motor end shields are attached to the ends of the center ring.

center tap — A connection at the electrical center of a winding, or midway between the electrical ends of a resistor or other portion at a circuit.

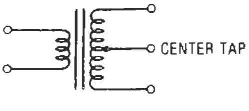

Center tap.

center-tapped inductor — An inductor that has a tap at half the total inductance.

center wire — A fine loop of wire used in proportional counters as an anode. A high voltage is applied to it to set the conditions for radiation measurement.

center-zero meter — A dc meter that has its zero point at the center of the scale, e.g., a dc galvanometer.

centi- — One hundredth (10^{-2}) of a specific quantity or dimension.

centigrade temperature scale — The older name for a Celsius temperature scale in the English-speaking countries. Officially abandoned by international agreement in 1948, but still in common usage.

centimeter waves — Microwave frequencies between 3 and 30 GHz, corresponding to wavelengths of 10 to 1 centimeters.

central battery exchange — Manual telephone exchange in which a battery situated at the exchange is the source of current for operating supervisory signals, for subscribers' calling signals, and for the current required to enable a subscriber to speak over his or her line.

central office — The facility at which a communications common carrier terminates customer lines and locates the equipment for interconnecting those lines.

central-office equipment — Apparatus used in a telephone central office to furnish communication services.

central-office line — *See* subscriber line.

central processing unit — Also called central processor; abbreviated CPU. 1. The part of a computer system that contains the main storage, arithmetic unit, and special register groups. Performs arithmetic operations, controls instruction processing, and provides timing signals and other housekeeping operations. 2. A group of registers and logic that form the arithmetic/logic unit plus another group of registers with associated decoding logic that form the control unit. Most MOS devices are single-chip CPUs, in that the registers hold as many bits as the word length of the unit. With bit-slice devices, however, central processing units of any bit width can be assembled essentially by connecting the bit-slice parts in parallel. Externally, a bit-slice device will appear to be a coherent single CPU capable of handling words of the desired bit length. 3. That part of a computer system that controls the interpretation and execution of instructions. In general, the CPU contains the following elements: arithmetic and logic unit (ALU), timing and control, accumulator, scratch-pad memory, program counter and address stack, instruction register, and I/O control logic. 4. That unit of a computing system that fetches, decodes, and executes programmed instructions and maintains the status of results as the program is executed.

central processor — *See* central processing unit; CPU.

central station — A control center to which alarm systems in a subscriber's premises are connected, where circuits are supervised and where personnel are maintained continuously to record and investigate alarm or trouble signals. Facilities are provided for the reporting of alarms to police and fire departments or to other outside agencies.

central station alarm system — An alarm system, or group of systems, the activities of which are transmitted to, recorded in, maintained by, and supervised from a central station. This differs from proprietary alarm systems in that the central station is owned and operated independently of the subscriber.

centrex — 1. A service offered by telephone companies. It uses central-office equipment to provide features comparable with those provided by a PBX. 2. An improved PBX system that also provides direct inward dialing (DID) and automatic number identification (ANI) of the calling PBX station.

centrifugal force — The force that acts on a rotating body and that tends to throw the body farther from the axis of its rotation.

centripetal force — The force that compels a rotating body to move inward toward the center of rotation.

Ceracircuits—A trademark of the Sprague Electric Company for hybrid thick-thin film integrated circuits that consist of discrete passive and semiconductor active elements attached to precision resistor substrates to form functional electronic modules.

ceramic—1. A claylike material, consisting primarily of magnesium and aluminum oxides, that after molding and firing is used as an insulating material. It withstands high temperatures and is less fragile than glass. When glazed, it is called porcelain. 2. Pertaining to or made of clay or other silicates. 3. Piezoelectric part of a pickup, speaker, or microphone that acts as a transducer. It has characteristics that are similar to a crystal transducer, but it is more robust. 4. Nonmetallic and inorganic material, e.g., alumina, beryllia, or steatite, formed through heat processing, used in microelectronic substrates and component parts.

ceramic amplifier—An amplifier that makes use of the piezoelectric properties of ceramics such as barium titanate and the piezoresistive properties of semiconductors such as silicon. An ac signal applied through electrodes to a barium titanate bar produces deformation of the bar and the attached silicon strip, thereby producing a corresponding variation in resistance. This resistance change causes the load current to vary. The device is essentially a current amplifier with extremely high input impedance and low output impedance.

ceramic-based microcircuit—A microminiature circuit printed on a ceramic substrate. Usually consists of combinations of resistive, capacitive, or conductive elements fired on a waferlike piece of ceramic.

ceramic capacitor—A capacitor whose dielectric is a ceramic material such as steatite or barium titanate, the composition of which can be varied to give a wide range of temperature coefficients. The electrodes are usually silver coatings, fired on opposite sides of the ceramic disc or slab, or fired on the inside and outside of a ceramic tube. After connecting leads are soldered to the electrodes, the unit is usually given a protective insulating coating.

ceramic dielectric—1. One of a great variety of ceramic materials used as a dielectric in capacitors: some typical materials are barium titanate, barium strontium titanate, and titanium dioxide. Different ceramic dielectrics provide the desired temperature coefficient of capacitance and medium-to-dielectric constants. 2. A ceramic such as isolantite, porcelain, or steatite, used as an insulator.

ceramic filter—Electrically coupled, two-terminal piezoelectric ceramic resonators in ladder and lattice configurations. Monolithic filters with ceramic substrates are also called ceramic filters.

ceramic microphone—1. A microphone with a ceramic cartridge. 2. A transducer that uses a piezoelectric ceramic pickup element (barium titanate) to convert sound to electrical energy.

ceramic permanent magnet—A permanent, nonmetallic magnet made from pressed and sintered mixtures of metallic-oxide powders, usually oxides of barium and iron.

ceramic pickup—1. A phonograph pickup with a ceramic cartridge. 2. A pickup whose generator system is based on piezoelectricity produced by the stressing of natural and man-polarized crystals.

ceramic transducer—*See* piezoelectric transducer.

CERDIP—1. Abbreviation for *cer*amic *d*ual-*i*nline *p*ackage. A package assembled with the leadframe sandwiched between two ceramic layers and sealed by firing a glass frit. 2. A ceramic dual-inline package for ICs.

Cerenkov counter—An instrument that detects high-energy charged particles by an analysis of the Cerenkov radiation emitted.

Cerenkov radiation—1. Light emitted when charged particles pass through a transparent material at a velocity greater than that of light in that material. It can be seen, for example, as a blue glow in the water around the fuel elements of pool reactors. P. A. Cerenkov was the Russian scientist who first explained the origin of this light. 2. The radiation produced when a charged particle traverses a medium that has a refraction index considerably greater than unity.

Cerenkov rebatron radiator—A device in which a tightly bunched velocity-modulated electron beam is passed through a hole in a dielectric. The reaction between the higher velocity of the electrons passing through the hole and the slower velocity of the electromagnetic energy passing through the dielectric results in radiation at some frequency higher than the frequency of modulation of the electron beam.

cermet—1. A metal-dielectric mixture used in making thick-film resistive elements. The first half of the term is derived from *cer*amic and the second half from *met*al. 2. A solid homogeneous material usually consisting of a finely divided admixture of a metal and ceramic in intimate contact. Cermet thick films are normally combinations of dielectric materials and metals.

cermet potentiometer—A potentiometer in which the resistive element is made by combining very fine particles of ceramic or glass with precious metals.

CERT—Abbreviation for character error-rate testing. Checking a data line with test characters.

certification—Verification that specified testing has been performed, and required parameter values have been attained.

certified magnetic tape—Magnetic tape that has been tested and is certified to be free from error over its entire recording surface.

cesium—A chemical element having a low work function. Used as a getter in vacuum tubes and in cesium-oxygen-silver photocell cathodes.

cesium-oxide cell—A photoemissive detector sensitive to wavelengths up to 1 micrometer. It has one sharp maximum of sensitivity at 350 nanometers and a broad maximum at 800 nanometers.

cesium-vapor lamp—A low-voltage arc lamp for producing infrared radiation.

cev—*See* corona extinction voltage.

CGA—Abbreviation for color graphics adapter. A color graphics system for IBM PCs and compatibles. Provides less resolution than EGA or VGA monitors.

C_{gk}—Symbol for grid-cathode capacitance in a vacuum tube.

CGM—Abbreviation for computer graphics metafile. A standard file format that stores object-oriented graphics in device-independent form, enabling them to work in different systems and programs.

C_{gp}—Symbol for grid-plate capacitance in a vacuum tube.

cgs—Abbreviation for centimeter-gram-second. These quantities of space, mass, and time are the basis of absolute units.

cgs electromagnetic system of units—A coherent system of units for expressing the magnitude of electrical and magnetic quantities. The most common fundamental units of these quantities are the centimeter, gram, and second. Their unit of current (abampere) is of such a magnitude that if maintained constant in two straight parallel conductors having an infinite length and negligible circular sections and placed 1 centimeter apart in a vacuum, a force equal to 2 dynes per centimeter of length will be produced.

cgs electrostatic system of units — A coherent system of units for expressing the magnitude of electrical and magnetic quantities. The most common fundamental units of these quantities are the centimeter, gram, and second. Their unit of electrical charge (stacoulomb) is of such a magnitude that two equal unit point charges 1 centimeter apart in a vacuum will repel each other with a force of 1 dyne.

chad — The piece of material removed when a hole or notch is formed in a storage medium such as punched tape or punched cards.

chadless — Pertaining to tape in which the data holes are deliberately not punched through and a flap of material remains attached to the tape.

chadless tape — A type of punched paper tape in which each chad is left fastened by about a quarter of the circumference of the hole. Chadless punched paper tape must be sensed by mechanical fingers, because chad interferes with reliable electrical or photoelectrical reading.

chad tape — Tape used in printing telegraph or teletypewriter operation, in which the perforations are severed from the tape to form holes that represent characters. Normally, the characters are not printed on chad tape.

chaff — A general name applied to radar-confusion reflectors that consist of thin, narrow, metallic strips of various length and frequency responses used to reflect radar echoes.

chain — A series of processing locations through which information must pass on a store-and-forward basis to reach a subsequent location.

chain calculations — In a calculator, series of continued operations in a single mode. Example: 118 × 94 × 116 × 395.

chained list — A list in which the items may be dispersed but in which each item contains an identifier for locating the next item to be considered.

chaining — 1. In a computer, a system of storing records such that each record belongs to a list or group of records and has a linking field for tracing the chain. 2. If a computer program is too big to fit into the memory, it can be written in a series of segments. The computer works on one segment at a time and continues operating this way until the program is finished. 3. The ability of an executing program to call another program that resides on a disk.

chaining search — A search technique in which each item contains an identifier for locating the next item to be considered.

chain printer — In a computer, a high-speed printer having type slugs carried on the links of a revolving chain.

chain radar beacon — A radar beacon with a very fast recovery time, so that simultaneous interrogation and tracking of the beacon by a number of radars is possible.

chain radar system — A radar system comprising a number of radars or radar stations located at various sites along a missile flight path. These radar stations are linked together by data and communication lines for target acquisition, target positioning, and/or data-recording purposes. The target-acquisition link makes it possible for any radar to position any other radar on target.

challenger — *See* interrogator.

chance failure — *See* random failure.

changer — A device that plays several phonograph records in sequence automatically. It consists of a turntable, an arm, and a record stacking and dropping mechanism.

channel — 1. A portion of the spectrum assigned for the operation of a specific carrier and the minimum number of sidebands necessary to convey intelligence. 2. A single path for transmitting electric signals. (Note: The word *path* includes separation by frequency division or time division. *Channel* may signify either a one-way or two-way path, providing communication in either one direction only or in two directions.) *See also* alternate channel. 3. In electronic computers, that portion of a storage medium which is accessible to a given reading station. 4. The path along which information, particularly a series of digits or characters, may flow. 5. In computer circulating storage, one recirculating path containing a fixed number of words stored serially by word. 6. An area, under the silicon dioxide of a planar surface, that has been changed from one type of conductivity to the opposite type. A channel is the conductive path between the source and drain in an IGFET. Generally, channels are undesirable in other instances. Thick oxides or heavily doped regions called channel stoppers are used to prevent channels. 7. A complete sound path. A single-channel, or monophonic, system, has one channel. A stereophonic system has at least two full channels designated as left (A) and right (B). Monophonic material may be played through a stereo system; both channels will carry the same signal. Stereo material, if played on a monophonic system, will mix and emerge as monophonic sound. 8. The conducting charge layer between source and drain induced by the applied gate voltage. The charge layer is holes in a p-type device, and electrons in n-types. 9. An independent signal path. Stereo recorders have two such channels, quadraphonic ones have four. 10. A thin semiconductor layer between the source region and the drain region, whose conductance is controlled by the gate voltage. 11. The band of frequencies (including the assigned carrier frequency) within which a radio system must operate in order to prevent interference with stations on adjacent channels. 12. The part of a communications system that connects a message source to a message link; a path for electrical transmission between two or more points.

channel balance — Equal response on both left and right channels.

channel bank — 1. The part of a carrier multiplex terminal in which are performed the first step of modulation of the voice frequencies into a higher-frequency band and the final step in the demodulation of the received higher-frequency band into voice frequencies. 2. Communication equipment that multiplexes, typically used for multiplexing voice-grade channels.

channel capacity — 1. The maximum number of elementary digits that can be handled per unit time in a particular channel. 2. The maximum possible rate of information transmission through a channel at a specified error rate. It may be measured in bits per second or bauds. 3. The total number of individual channels in a system.

channel designator — A number assigned for reference purposes to a channel, tributary, or trunk. Also called channel sequence number.

channel diffusion stops — A narrow doped region beside each sensing channel in a CCD that prevents excess charges generated within a particular light-sensing site from spreading sideways.

channel effect — Current leakage over a surface path between the collector and emitter of some types of transistors.

channel frequency — The band of frequencies that must be handled by a carrier system to transmit a specific quantity of information.

channeling — The utilization of a modulation-frequency band for the simultaneous transmission from two or more communication channels in which the channel separation is accomplished by the use of carriers or subcarriers, each in a different discrete frequency band forming a subdivision of the main band. This term covers a special case of multiplex transmission.

channel interval — The time allocated to a channel, including on and off time.

channelization — The assignment of circuits to channels, and the arrangement of those channels into separate groups and supergroups.

channelizing — The process of subdividing wideband transmission facilities for the purpose of putting many different circuits requiring comparatively narrow bandwidths on a single wideband facility.

channel pulse — A telemetering pulse that, by its time or modulation characteristics, represents intelligence on a channel.

channel pulse synchronization — Synchronization of a local-channel rate oscillator by comparison and phase lock with separate channel-synchronizing pulses.

channel reliability — The percentage of time that the channel meets the arbitrary standards established by the user.

channel reversal — Shifting the outputs of a stereo system so the channel formerly heard from the left speaker now comes from the right and vice versa.

channel-reversing switch — A switch that reverses the connections of two speakers in a stereo system with respect to the channels, so that the channel heard previously from the right speaker is heard from the left and vice versa.

channel sampling rate — The number of times per second that individual channels are sampled. This is different from comutation rate, since it is possible for more than one channel to be applied to a given commutator input (with subcommutation).

channel selector — A switch or dial used for selecting a desired channel.

channel separation — 1. In stereo, the electrical or acoustical difference between the left and right channels. Inadequate separation can lessen the stereo effect; excessive separation can exaggerate it beyond natural proportions. 2. The degree to which the two signals in a stereo system are electrically isolated. Usually expressed as a ratio in decibels. 3. The amount of stereo program material from one channel appearing in the cartridge output for the other channel. Expressed in decibels relative to the desired channel output, with values of 20 to 30 dB (the higher figure being preferable) through most of the audible frequency range being typical of good cartridges. 4. The degree to which the signal in one amplifier is kept separate from an adjacent undriven amplifier. Channel separation for FM stereo decoders is typically 40 dB, whereas phono cartridge channel separation is typically between 20 and 30 dB.

channel sequence number — *See* channel designator.

channel shift — The interchange of communications channels; for example, the shift from a calling frequency to a working frequency.

channel shifter — A radiotelephone carrier circuit by means of which one or two voice-frequency channels are shifted from normal channels to higher voice-frequency channels as a means of reducing crosstalk between channels. At the receiving end, the channels are shifted back by a similar circuit.

channel stop — A barrier put in place during the manufacture of a semiconductor to prevent small leakage paths going to the outside of the chip.

channel strip — An amplifier or other device having a sufficiently wide bandpass to amplify one television channel.

channel subcarrier — The channel required to convey telemetric data involving a subcarrier band.

channel synchronizing pulse separator — A device for separating channel synchronizing pulses from commutated signals.

channel-to-channel connection — A device for rapid data transfer between two computers. A channel adapter is available that permits the connections between any two channels on any two systems. Data is transferred at the rate of the slower channel.

channel-utilization index — In a computer, the ratio between the information rate (per second) through a channel and the channel capacity (per second).

channel wave — Any elastic wave propagated in a sound channel because of a low-velocity layer in the solid earth, the sea, or the atmosphere.

character — 1. In electronic computers, one of a set of elementary symbols that may be collectively arranged in order to express information. These symbols may include the decimal digits 0 through 9, the letters A through Z, punctuation and typewriter symbols, and any other single symbol that a computer may read, store, or write. 2. One of a set of symbols used to present information on a display tube. 3. Part of a computer word that has a meaning in itself. For example, six bits recorded across a magnetic tape make up a character and signify a number or letter symbol. 4. A combination of holes punched in a line. 5. A letter, digit, or other symbol that is used as part of the organization, control, or representation of data. A character is often in the form of a spatial arrangement of adjacent or connected strokes. 6. A symbol, mark, or event that a data-processing machine can read, write, or store. It is used to represent data to a machine. 7. An alphabetic, numeric, or special graphic symbol. Each character is represented in its set by a unique binary code. 8. A letter, number, or sign made up of a specific number of bits. Nonprinting characters may be used for control functions. 9. A language unit consisting of bits. 10. A letter, digit, or other symbol that is the representation of data. A connected sequence of characters is called a character string.

character boundary — In character recognition, the largest rectangle having a side parallel to the document reference edge, each of the sides of which is tangent to a given character outline.

character check — Verification of the observance of rules for character formation.

character code — A special way of using a group of bits to represent a character. Different codes may be used in different equipment according to the internal design.

character crowding — Also called packing. The effect of reducing the time interval between subsequent characters read from tape. It is caused by a combination of mechanical skew, gap scatter, jitter, amplitude variation, etc.

character density — A measure of the number of recorded characters per unit of length or area.

character display tube — A form of cathode-ray tube in which the cathode-ray beam can be shaped, either by electrostatic or electromagnetic deflection or by passing the beam through a mask, into symbols or letters.

character emitter — In a computer, an electromechanical device that puts out coded pulses.

character generator — 1. A unit that accepts input in the form of one of the alphanumeric codes and prepares the electrical signals necessary for its display in the proper position on a dot matrix, TV system, or CRT. 2. That part of the display controller that draws alphanumeric characters and special symbols for the screen. A character is automatically drawn and spaced every time a character code is interpreted. 3. A hardware or software device that provides the means for formulating a character font and that also may provide some controlling function during printing. 4. A circuit that generates the letters or numbers on a display or printer. 5. A device that superimposes text over a video image, such as program credits at the end of a program.

character-generator cathode-ray tube — A cathode-ray tube that generates symbols for use in other displays. Basically, the tube operates by scanning specific characters on the target and generating them as video signals to other cathode-ray systems.

character-graphics color system — A system that can address color only to a block of pixels (character cell); all pixels within this block have the same color.

character interleave — Also byte interleave. A technique in time-division multiplexing in which bytes of data are transmitted in one frame.

characteristic — 1. An inherent and measurable property of a device. Such a property may be electrical, mechanical, thermal, hydraulic, electromagnetic, or nuclear; it can be expressed as a value for stated or recognized conditions. A characteristic may also be a set of related values (usually in graphical form). 2. The integral part of a logarithm to the base 10; also, the power of 10 by which the significant digits of a floating-point number are multiplied.

characteristic curve — 1. A graph plotted to show the relationship between changing values. 2. A plot of how any value changes with respect to another in a component or circuit: for example, a plot of how current changes as voltage is changed on a component or circuit.

characteristic distortion — 1. Displacement of signal transitions due to the persistence of transients caused by preceding transitions. 2. Repetitive displacement or disruption peculiar to specific parts of a teletypewriter signal. The two types of characteristic distortion are line and equipment.

characteristic frequency — The frequency that can be easily identified and measured in a given emission.

characteristic impedance — Also called surge impedance. 1. The driving-point impedance of a line if it were of infinite length. 2. In a delay line, the value of terminating resistance that provides minimum reflection to the network input and output. 3. The ratio of voltage to current at every point along a transmission line on which there are no standing waves. 4. The square root of the product of the open-and short-circuit impedance of the line. 5. The ratio of the applied voltage to the steady-state current that flows when a transmission line is terminated in a pure resistance that is equal to the characteristic impedance value of the cable. 6. That value of pure resistance which, when connected to the output terminals of a transmission line, makes the cable appear infinitely long; i.e., no signal is reflected back up the cable toward the source. 7. The ratio of voltage to current in a propagating wave, i.e., the impedance that is offered to this wave at any point of the line. In printed wiring, its value depends on the width of the conductor, the distance from the conductor to the ground plane(s) and the dielectric constant of the media between them. 8. A property of antenna transmission lines that is determined primarily by the diameter of the conductors and the spacing between them.

characteristic impedance of free space — The relationship between the electric and magnetic intensities of space due to the expansion of the impedance concept to electromagnetic fields.

characteristic spread — The range between the minimum and maximum values for a given characteristic that is considered normal in any large group of tubes or other devices.

characteristic telegraph distortion — Distortion that does not affect all signal pulses alike. Rather, the effect on each transition depends on the signal previously sent, because remnants of previous transitions or transients persist for one or more pulse lengths.

characteristic wave impedance — The ratio of the transverse electric vector to the transverse magnetic vector at the point it is crossed by an electromagnetic wave.

character parity — Adding an overhead bit to a character code to provide error-checking capability.

character reader — A computer input device that can directly recognize printed or written characters; they need not first be converted into punched holes in cards or paper or into polarized magnetic spots.

character read-out systems — Photoelectrically controlled, alphanumeric reading devices that convert characters to audible or sorting signals that can be fed to a computer, electric typewriter, tape punch, or other machine.

character recognition — The automatic identification of graphic, phonic, or other characters. *See also* magnetic-ink character recognition; optical character recognition.

character sensing — To detect the presence of characters optically, magnetically, electrostatically, etc.

character set — 1. An ordered group of unique representations called characters, such as the 26 letters of the English alphabet, 0 and 1 of the Boolean alphabet, the signals in the Morse code alphabet, the 128 characters of the U.S. ASCII alphabet, etc. 2. All the letters, numbers, and symbols used by a device or language.

character string — Two or more alphanumeric characters or special symbols (math, Greek, etc.) aligned in a textual format on the screen.

character subset — A selection from a character set of all characters having a specified common feature: for example, in the definition of a character set, the digits 0 through 9 are a character subset.

Charactron — Trade name of General Dynamics/ Electronics for a specially constructed cathode-ray tube used to display alphanumeric characters and other special symbols directly on its screen.

charge — 1. The electrical energy stored in a capacitor or battery or held in an insulated object. 2. The quantity of electrical energy in (1) above. 3. In electrostatics, the amount of electricity present on any substance that has accumulated electric energy.

charge amplifier — 1. An operational amplifier with capacitive feedback. The output voltage that results from a charge signal input is returned to the input circuit through the feedback capacitor in the direction necessary to maintain the input circuit voltage at or near zero. Thus, the net charge input is stored in the feedback capacitor, producing a potential difference across it equal to the value of charge divided by the value of capacitance. This potential difference determines the relationship of the output voltage signal magnitude to the input charge signal magnitude. The transfer characteristic of the amplifier depends on the value of the feedback capacitor. 2. An amplifier whose output voltage is proportional to the input charge from a piezoelectric transducer.

charge carrier — A mobile hole or conduction electron in a semiconductor.

charge-coupled device — Abbreviated CCD. Semiconductor device arrayed so that the electric charge at the output end produces a stimulus to the next device.

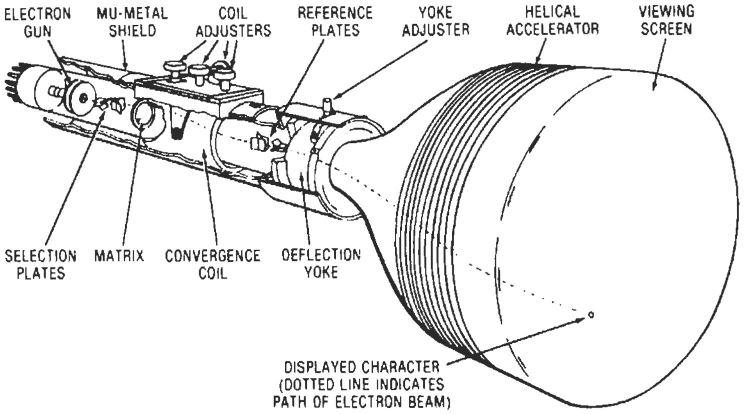

Charactron.

charge density—The charge per unit area on a surface, or charge per unit volume in space.

charged particle—An ion, an elementary particle that carries a positive or negative electric charge.

charge injection device—*See* CID.

charge injection imaging device—*See* CID.

charge offset—During the sample-to-hold transition of a sample/hold circuit, the charge transferred to the holding capacitor because of the switching process. It is usually expressed in millivolts.

charger—A device used to convert alternating current into a pulsating direct current that can be used for charging a storage battery.

charge retention—The ability of a battery to hold its energy once it has been charged.

charge-storage tube—A storage tube that retains information on its surface in the form of electric charges.

charge transfer—The process in which an ion takes an electron from a neutral atom of the same type, with a resultant transfer of electronic charge.

charging—1. The process of converting electrical energy to stored chemical energy. 2. The process of storing electrical energy in a capacitor.

charging current—The current produced when a dc voltage is first applied to a capacitor. This current decreases exponentially with time.

charging rate—1. The rate of current flow used in charging a battery. 2. The rate at which charging current flows into a capacitor or capacitance-resistance circuit. Expressed in amperes, milliamperes, or microamperes.

chart recorder—A data recorder that provides a record of the values of a physical parameter, in the form of a graph on a piece of chart paper, either with respect to time or to some other variable. The recording system consists of essentially three elements: a transducer to convert the variable to be measured (temperature, pressure, rpm, etc.) into an electrical signal, a signal conditioner to process the signal into a form such that it may be recorded, and the recording device.

chaser—1. An array of elements similar to a ring except that as each successive element is switched to the "on" condition, the others remain on as well; when all stages are on, the next pulse turns them all off and the process generally repeats. 2. A repeat-cycle flasher with three or more outputs each operating in sequence to the other. Normally used on signs or displays to create a moving effect.

chassis—1. A sheet-metal box, frame, or simple plate on which electronic components and their associated circuitry can be mounted. 2. The entire equipment (less cabinet) when so assembled. *See also* printed circuit board.

chassis ground—A connection to the metal structure that supports the electrical components that make up the unit or system.

chat rooms—Areas on an online service, BBS, or the Internet that allow real-time, typed-in communication with other people.

chatter—1. A sustained rapid opening and closing of contacts due to variations in the coil current. 2. The vibration of a cutting stylus in a direction other than the direction in which it is driven.

chattering—Rapid audible cyclic action within an electromechanical device.

chatter time—The interval of time from initial actuation of a contact to the end of chatter.

cheater cord—An extension cord used to conduct power to a piece of equipment (especially a TV) by temporarily bypassing the safety interlock connector.

check—The partial or complete verification of the correctness of equipment operations, the existence of certain prescribed conditions, and/or the correctness of results.

check bit—1. A binary check digit. 2. The bit that is automatically added by the computer to an item of data when it is necessary to make it either even or odd parity. Synonym: parity bit.

check character—A character used to perform a check.

check digit—A digit added to each number in a coding system that allows for detection of errors in the recording of the code numbers. Through the use of the check digit and a predetermined mathematical formula, recording errors, such as digit reversal, can be noted.

checkerboard—*See* worst-case noise pattern.

checking code—Machine instructions that read part of a diskette to determine whether it has been copied.

checkout—A series of operations and calibration tests used to determine the condition and status of a system or element of the system.

checkpoint—1. In a computer routine, a point at which it is possible to store sufficient information to permit restarting the computation from that point. 2. The status of a long-running program is often recorded at frequent intervals called checkpoints. If something goes wrong, the program can be restarted at its last checkpoint instead of from the beginning.

checkpoint routine—A computer routine in which information for a checkpoint is stored.

check problem — A problem that, when incorrectly solved, indicates an error in the programming or operation of a computer.

check register — A special register provided in some computers to temporarily store transferred information for comparison with a second transfer of the same information in order to verify that the information transferred each time agrees precisely.

check routine — A program whose purpose is to determine whether a computer or a program is operating correctly.

checksum — 1. In a computer, a summation of digits or bits summed according to an arbitrary set of rules and primarily used for checking purposes. 2. A value that is the arithmetic sum of all the bytes in a program or program segment. As the program is loaded, the loader computes the sum of all bytes and compares the result with the checksum. If the two values are equal, it is assumed that the program segment was loaded without error. 3. A character added after a block of *n* words that contains the truncated binary sum of preceding nibbles. Used to verify the integrity of data in a ROM or on a tape. 4. Error detection data in a diskette sector based on the sum of all data bits in a block; if the sums as written and as calculated do not agree, an error is reported. 5. A simple error-checking technique used in data communications.

cheese antenna — An antenna with a cylindrical parabolic reflector enclosed by two plates perpendicular to the cylinder and so spaced that more than one mode can be propagated in the desired direction of polarization. It is fed on the focal line.

chelate — A molecule in which a central inorganic ion is covalently bonded to one or more organic molecules, with at least two bonds to each molecule; used as a laser dopant.

chemical deposition — The process of depositing a substance on a surface by means of the chemical reduction of a solution.

chemically deposited printed circuit — A printed circuit formed on a base by the reaction of chemicals alone. Dielectric, magnetic, and conductive circuits can be applied.

chemically reduced printed circuit — 1. A printed circuit formed by chemically reducing a metallic compound. 2. A printed circuit formed by the chemical erosion (etching) of portions of the metallic surface of a metal-clad insulative material.

chemical vapor deposition — Abbreviated CVD. A gaseous process that deposits insulating films or metal onto a wafer at elevated temperature. Often, reduced pressure is used to promote the chemical reaction.

chemisorption — *See* adsorption.

CHIL — Abbreviation for current-hogging injection logic.

Child's law — Also known as the three-halves power equation. It states that the current in a thermionic diode varies directly with the three-halves power of the anode voltage and inversely with the square of the distance between electrodes.

chip — 1. The uncased and normally leadless form of an electronic component part, either passive or active, discrete or integrated. 2. A single substrate on which all the active and passive circuit elements have been fabricated using one or all of the semiconductor techniques of diffusion, passivation, masking, photoresist, and epitaxial growth. A chip is not ready for use until packaged and provided with external connectors. The term is also applied to discrete capacitors and resistors that are small enough to be bonded to substrates by hybrid techniques. 3. A tiny piece of semiconductor material, scribed or etched from a semiconductor slice, on which one or more electronic components are formed. The total number of usable chips obtained from a wafer is the yield. 4. An unpackaged semiconductor device; a die incorporating an integrated circuit cut from a silicon wafer. By extension, every LSI package is commonly called a chip. 5. An electronic circuit element prior to having terminal connections added and prior to being encased for physical protection. 6. A single piece of semiconductor material (silicon, sapphire, germanium, etc.) containing one or more circuits, usually packaged as a unit. 7. Also called thread. In mechanical recording, the material removed from the recording medium by the recording stylus as it cuts the groove. 8. In punched cards, a piece of cardboard removed in the punching process.

chip and wire — A hybrid technology employing face-up bonded chip devices interconnected to the substrate conventionally, i.e., by flying wires.

chip architecture — The design or structure of an IC chip, incorporating arithmetic and logic unit, registers, and control-bus pathway configuration.

chip capacitor — 1. A capacitor of which the small dimensions or the nature and configuration of its terminations render it suitable for use mainly in hybrid circuits. 2. Discrete device that introduces capacitance into an electronic circuit, made in tiny wedge or rectangular shapes to be soldered onto hybrid circuits. 3. A subminiature capacitor (usually ceramic or solid tantalum) in chip form.

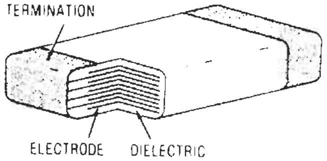

Chip ceramic (multilayer) capacitor.

chip carrier — 1. A low-profile component package, usually square, whose active chip cavity or mounting area is a large fraction of the package size and whose external connections are usually on the four sides of the package. 2. A physical package into which a die is mounted.

chip component — An unpackaged circuit element (active or passive) for use in hybrid microelectronics. Besides ICs, the term includes diodes, transistors, resistors, inductors, and capacitors.

chip-in-tape — An automated hybrid bonding technique that provides multiple, simultaneous bonding of leads by means of thermal and mechanical energy transmission through a deformable prepunched tape.

chip-level integration — The combination of two or more integrated-circuit functions and/or technologies on one IC to achieve miniaturization, reduce systems cost, and make new applications possible. Particularly important for signal processing and power control solutions.

chip-outs — Semiconductor die defects where fragments of silicon on the face have been chipped off in processing, leaving an active junction exposed.

chip resistor — A subminiature resistor formed on a small insulating substrate.

chip set — A set of integrated circuits that supplies all or most of the circuitry needed to build a functional item of electronic equipment. Most modems and computers are built from chip sets.

chip sets — The microprocessor chip in addition to RAMs, ROMs, and interface I/O devices. The chip sets mounted on a board are also referred to as the CPU portion of the microcomputer. Also called microcontroller.

Chireix antenna—Also called Chireix-Mesny antenna. Resonant series-fed array of square loops with half-wave sides. The loops feed each other in cascade, corner to corner, and the antenna resembles a double zigzag.

chirp—1. An all-encompassing term for the various techniques of pulse expansion and pulse compression applied to pulse radar. A technique to expand narrow pulses to wide pulses for transmission, and to compress wide received pulses to the original narrow pulse width and waveshape. This improves the signal-to-noise ratio without degradation to the range resolution and range discrimination. 2. A colloquial expression for a coded pulse. In coding the pulse, the carrier frequency is increased in a linear manner for the duration of the pulse, and when the pulse is translated to an audio frequency, it sounds like a chirp. 3. A change in the pitch of code signals, generally due to poor regulation of the transmitter power supply. 4. A pulsed frequency-modulation scheme in which a carrier is swept over a wide frequency band during a given pulse interval.

chirp modulation—Swept-frequency modulation used in some radar and sonar equipment to increase the on-target energy and improve range resolution by making full use of the average power capability of the transmitter.

chirp radar—Radar in which a swept-frequency signal is transmitted, received after being returned from a target, and compressed in time to give a final narrow pulse called the chirp signal. This type of radar has high immunity to jamming and provides inherent rejection of random noise signals.

choke—1. An inductance used to impede the flow of pulsating direct current or alternating current by means of its self-inductance. 2. An inductance used in a circuit to present a high impedance to frequencies appreciably limiting the flow of direct current. Also called choke coil. 3. A groove or other discontinuity in a waveguide surface so shaped and dimensioned as to impede the passage of guided waves within a limited frequency range.

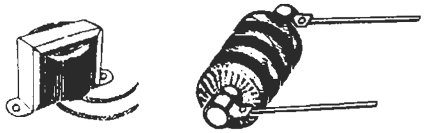

Chokes.

choke coil—Also called impedance coil. An inductor (reactor) used to limit or suppress the flow of alternating current without appreciable effect on the flow of direct current.

choke flange—A waveguide flange with a grooved surface; the groove is so dimensioned that the flange forms part of a choke joint.

choke-input filter—A power-supply filter in which a choke is the first element in series with the input current from the rectifier.

choke joint—1. A connector between two sections of transmission line in which the gap between sections to be connected is built out to form a series-branching transmission line carrying a standing wave, in which actual contact falls at or near a current minimum. 2. A joint for connecting two sections of waveguide together. Permits efficient energy transfer without the necessity of an electrical contact at the inside surface of the guide.

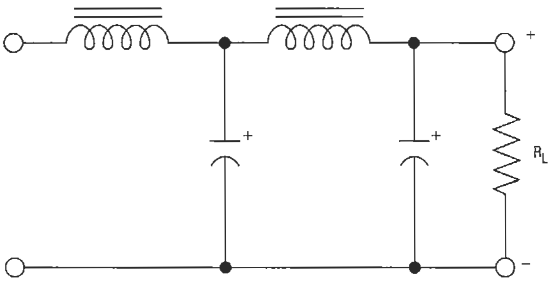

Choke-input filter.

cholesteric phase—An arrangement of liquid crystal molecules that occurs only in optically active substances and is considered to be a twisted nematic phase with a helical structure. It consists of layers resembling the smectic phase, but each layer has an order characteristic of a nematic phase.

chopped mode—A time-sharing method of displaying with a single CRT gun the output signals of two or more channels in sequence at a rate not referenced to the sweep.

chopper—1. A device for interrupting a current or a light beam at regular intervals. Choppers are frequently used to facilitate amplification. 2. An electromechanical switch for the production of modified square waves. The waves are of the same frequency as a driving sine wave and bear a definite relationship to it. 3. An electromechanical or electronic device used to interrupt a dc or low-frequency ac signal at regular intervals to permit amplification of the signal by an ac amplifier. It may also be used as a demodulator to convert an ac signal to dc. 4. A rotating shutter for interrupting an otherwise continuous stream of particles. Choppers can release short bursts of neutrons with known energies. Used to measure nuclear cross sections.

chopper amplifier—A circuit that amplifies a low-level signal after it has gone through a chopper.

chopper stabilization—1. The addition of a chopper amplifier to the regulator input circuitry of a regulated power supply in order to reduce output drift. 2. A method of improving the dc drift of an amplifier by utilization of chopper circuits.

chopper-stabilized amplifier—1. An amplifier configuration utilizing a carrier-type dc amplifier to reduce the effect of input offset and drift of a direct-coupled amplifier. 2. An amplifier in which the dc input is chopped, simulating ac in order to overcome amplifier dc drift.

chopping—Removal by electronic means of one or both extremities of a wave at a predetermined level.

chopping frequency—The frequency at which a chopper interrupts a signal.

chord—A harmonious combination of tones sounded together through the use of one or more fingers on either or both hands. On chord organs, a full chord is selected by depressing a single chord button.

chord organ—An organ with provision for playing a variety of chords, each produced by means of a single button or key.

chorus—A natural electromagnetic phenomenon in the VLF range. Probably originates in the exosphere. Also called dawn chorus because it sounds like birds at dawn. It generally consists of a multitude of rising tones, each tone rising from 1–2 kHz to 3–4 kHz and usually lasting 0.1 to 0.5 second.

Christiansen antenna—A radiotelescope composed of two interferometer arrays placed at right angles. It resembles a Mills cross antenna.

Christmas-tree pattern—1. *See* optical pattern. 2. A pattern resembling a Christmas tree, sometimes produced on the screen of a television receiver when the horizontal oscillator falls out of sync.

chroma—1. That quality which characterizes a color without reference to its brightness; that quality which embraces hue and saturation. White, black, and gray have no chroma. 2. The quality of light perception that includes color and its purity. The purity of a color varies inversely with the amount of noise, that is, white light, mixed with it. Thus pink, red, and deep-red describe chromas for the visible wavelength in the vicinity of 650 nm. In parlance more relevant to optics, chroma pertains to the response of the eye to the combined effects of hue and saturation.

chroma-clear raster—Also called white raster. Looks like a clear raster, but each of the three guns in the CRT is operating under the influence of a color level determined by a white video signal. In this case, all TV set chroma circuits are working as though the TV set were receiving a color transmission of a completely white scene.

chroma control—A variable resistor that controls saturation by varying the level of chrominance signal fed to the demodulators of a color television receiver.

chroma detector—A circuit that detects the absence of chrominance information in a color encoder input. The chroma detector automatically deletes the color burst from the color encoder output when the absence of chrominance is detected.

chromatic aberration—1. An effect that causes refracted white light to produce an image with colored fringes due to the various colors being bent at different angles. 2. An optical lens defect that causes light-color separation because the optical material focuses different light colors at different points. A lens without this defect is said to be achromatic.

chromaticity—1. The combination of the hue and saturation attributes of color. 2. A term quantitatively descriptive of a color, and dependent on both hue and saturation, but without reference to brilliance.

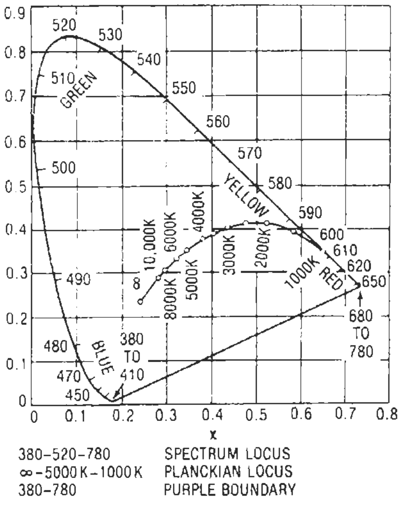

Chromaticity diagram.

chromaticity coordinates—Proportions of standard primaries (tristimulus values) required for a color match; ratios of each tristimulus value of a color to their sum. In the CIE calorimetric system, designated *X*, *Y*, and *Z*.

chromaticity diagram—A plane diagram formed when any one of the three chromaticity coordinates is plotted against another.

chromaticity flicker—The flicker that results from fluctuation of the chromaticity only.

chromatron—A color kinescope that has a single electron gun and whose color phosphors are laid out in parallel lines on its screen. The electron beam is directed to the correct phosphor by a deflection grid or wire grill near the face of the tube.

chrominance—1. Calorimetric difference between any color and a reference color having a specified chromaticity. In standard color-television transmission, the specified chromaticity is that of the zero subcarrier. 2. Pertaining to chroma, that is, to the mix of color and white light. Thus, pale pink may be either more or less bright than its pure constituent, red. An example of this is seen in a TV receiver, wherein the color signal is processed in the chrominance channel, but the brightness information is handled separately in the luminance channel. 3. A color term defining the hue and saturation of a color. Does not refer to brightness. 4. The hue and saturation of a color. The chrominance signal is modulated onto a 4.43 MHz carrier in the PAL television system and a 3.58 MHz carrier in the NTSC television system.

chrominance amplifier—The amplifier that separates the chrominance signal from the total video signal.

chrominance cancellation—A cancellation of the brightness variations produced by the chrominance signal on the screen of a monochrome picture tube.

chrominance-carrier reference—A continuous signal having the same frequency as the chrominance subcarrier, and a fixed phase with the color burst. The phase reference of carrier chrominance signals for modulation or demodulation.

chrominance channel—In color television, a combination of circuits designed to pass only those signals having to do with the reproduction of color.

chrominance component—Either of the I and Q signals that add to produce the complete chrominance signal in NTSC systems.

chrominance contrast—The color contrast between two adjacent surfaces of identical area, shape, texture, and luminance. The human eye is more sensitive to differences in color than it is to differences in brightness.

chrominance demodulator—A demodulator used in color-television reception for deriving video-frequency chrominance components from the chrominance signal and a sine wave of the chrominance subcarrier frequency.

chrominance gain control—In red, green, and blue matrix channels, variable resistors that individually adjust the primary-signal levels. Used in color television.

chrominance modulator—A modulator used in color-television transmission for generating the chrominance signal from the video-frequency chrominance components and the chrominance subcarrier.

chrominance primary—One of two transmission primaries, the amounts of which determine the chrominance of a color. Chrominance primaries have zero luminance and are not physical.

chrominance signal—1. The chrominance-subcarrier sidebands added to a monochrome television signal to convey color information. The components of the chrominance signal represent hue and saturation but do not include luminance or brightness. 2. That portion of

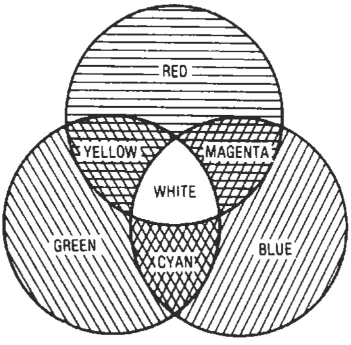

Chrominance primaries.

the total video signal that contains the color information. Without the chrominance signals, the received TV picture would be in black and white.

chrominance subcarrier—Also called color carrier. An rf signal that has a specific frequency of 3.579545 MHz and that is used as a carrier for the I and Q signals.

chrominance-subcarrier oscillator—In a color TV receiver, a crystal-controlled oscillator that generates the subcarrier signal for use in the chrominance demodulators.

chrominance video signals—Output voltages from the red, green, and blue sections of a color-television camera or receiver matrix.

chromium dioxide—A type of recording-tape coating that produces very good quality at low recording speeds. Because of its magnetic properties, it requires a higher value of bias current in the recorder. The high performance inherent in chromium dioxide tape can only be realized in a tape machine having provision for a CrO_2 bias setting. On a standard recorder, the chromium dioxide tape will appear to have high-frequency emphasis and may likely be difficult to erase. Chromium dioxide has a good dynamic range and a low noise level. Used in the cassette format with a suitable machine equipped with the Dolby system, it can make recordings that meet the best high-fidelity standards.

chronistor—A subminiature elapsed-time indicator that uses electroplating principles to totalize the operating time of equipment up to several thousand hours.

chronograph—An instrument for producing a graphical record of time as shown by a clock or other device.

chronoscope—An instrument for measuring very small intervals of time.

CID—Abbreviation for charge injection device or charge injection imaging device. 1. A memory in which the charge is stored in an X-Y addressable array of potential cells. For image arrays, the charge for each of the cells is generated by an associated photodioide. 2. A solid-state imaging device utilizing an image sensor composed of a two-dimensional array of coupled MOS charge-storage capacitors and designed to convert near infrared energy to electrical signals, providing broad gray shade or tonal rendition. The sensor collects minority carrier charge, generated by photon energy in the substrate near the charge-storage capacitors, and stores it in the surface inversion region. By injecting the stored charge into the substrate, and by monitoring the current, signal readout is achieved.

CIE—Initials of the Commission Internationale de l'Éclairage, or International Commission on Illumination.

CIE source—Standard light source representative of the quality of specified natural or artificial illumination.

CIE standard chromaticity diagram—A chromaticity diagram in which the *X* and *Y* chromaticity coordinates are plotted in rectangular coordinates.

CIM—Abbreviation for computer-integrated manufacturing. Applying information technology to production processes and organizational structure to streamline operations. Often focused on integrating systems and processes distributed across a company, such as order entry, scheduling, and production.

cinching—Longitudinal slippage between the layers in a tape pack as a result of acceleration or deceleration of the roll.

cipher—Cryptographic system in which arbitrary symbols or groups of symbols represent units of plain text of regular length, usually single letters, or in which units of plain text are rearranged, or both, according to certain predetermined rules.

cipher machine—Mechanical and/or electrical apparatus for enciphering and deciphering.

cipher telephony—A technique by which mechanical and/or electrical equipment is used for scrambling or unscrambling, or enciphering or decoding, radio or voice messages.

ciphertext or cryptogram—A secret form of a message.

ciphony—*See* cipher telephony.

circle cutter—A tool consisting of a center drill with an adjustable extension-arm cutter, used to cut holes in panels and chassis.

circle-dot mode—A method of storage of binary digits in a cathode-ray tube in which one kind of digit is represented by a small circle on the screen, and the other kind is represented by a similar circle with a concentric dot.

circle of confusion—The circular image of a point source due to the inherent aberrations in an optical system.

circlotron amplifier—A one-port, nonlinear cross-field high-power microwave amplifier that uses a magnetron as a negative-resistance element, much as a maser uses an active material.

circuit—1. Path through which electrical signals flow. 2. An electronic path between two or more points capable of providing a number of channels. 3. A number of conductors connected together for the purpose of carrying an electrical current. 4. The interconnection of a number of devices in one or more closed paths to perform a desired electrical or electronic function. Examples of simple circuits are high- or low-pass filters, multivibrators, oscillators, and amplifiers. 5. A complete path of electron flow from a negative terminal of voltage source through a conductor and back to the positive terminal. 6. An electrical system using two or more wires in which the current flows from the source to one or more electrical devices and back again to the source of supply. 7. A complete, closed path. Confusion between circuit and network is common. *Circuit* refers to a closed path within a network. 8. An array of elements interconnected to perform functions beyond the range of single-element capability. *See* channel, 2.

circuit analysis—Careful determination of the nature and behavior of a circuit and its various parts. The analysis may be theoretical, practical, or both.

circuit analyzer—Also called multimeter. Several instruments or instrument circuits combined in a single enclosure and used in measuring two or more electrical quantities in a circuit.

circuit bonding jumper—The connection between portions of a conductor in a circuit to maintain required ampacity of the circuit.

circuit breaker—1. An automatic device that, under abnormal conditions, will open a current-carrying circuit

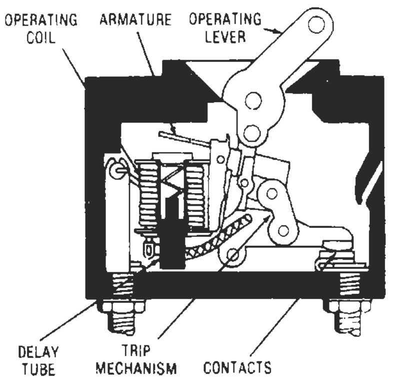

On position.

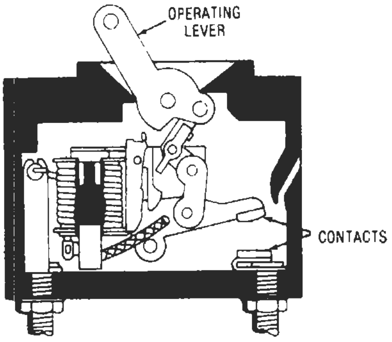

Trip position.

Circuit breaker.

without damaging itself (unlike a fuse, which must be replaced when it blows). 2. A device for interrupting a circuit under normal or abnormal conditions by means of separable contacts. 3. An electromagnetic device that opens a circuit automatically when the current exceeds a predetermined value. It can be reset by operating a lever or by other means.

circuit-breaker cascade system—A system wherein the protective devices are arranged in order of ratings such that those in series will coordinate and provide the required protection.

circuit capacity—The number of communication channels that can be handled by a given circuit at the same time.

circuit card—A printed circuit board containing electronic components.

circuit commutated turn-off time—The time interval between the instant when the principal current has decreased to zero after external switching of the principal voltage circuit and the instant when a thyristor is capable of supporting a specified rate of rise of on-state voltage without turning on.

circuit component—An element of a circuit, such as a resistor, capacitor, diode, inductor transformer, integrated circuit, or transistor.

circuit density—The number of equivalent transistors per unit area of a IC chip.

circuit diagram—A drawing in which symbols and lines represent the components and wiring of an electronic circuit. Also called circuit schematic. *See* schematic diagram.

circuit dropout—A momentary interruption of a transmission because of the complete failure of a circuit.

circuit efficiency (of the output circuit of electron tubes)—Ratio of the power, at the desired frequency, delivered to a load at the output-circuit terminals of an oscillator or amplifier to the power, at the desired frequency, delivered to the output circuit by the electron stream.

circuit element—1. Any basic constituent of a circuit except the interconnections. 2. A discrete unit of resistance, inductance, or capacitance that, when two or more are interconnected, forms an electric circuit.

circuit hole—On a printed circuit board, a hole that lies partially or completely within the conductive area.

circuit noise—The noise brought to the receiver electrically from a telephone system, but not the noise picked up acoustically by the telephone transmitters.

circuit noise level—At any point in a transmission system, the ratio of the circuit noise at that point to some arbitrary amount of circuit noise chosen as a reference. This ratio is usually expressed in decibels above reference noise, abbreviated dBrn, signifying the reading of a circuit-noise meter; or in adjusted decibels, abbreviated dBa, signifying the circuit-noise meter reading adjusted to represent interfering effect under specified conditions.

circuit-noise meter—Also called noise-measuring set. An instrument for measuring the circuit noise level. Through the use of a suitable frequency-weighting network and other characteristics, the instrument gives equal readings for noises of approximately equal interference. The readings are expressed in decibels above the reference noise.

circuit parameters—The values of the physical quantities associated with circuit elements; for example, the resistance (parameter) of a resistor (element), or the inductance per unit length (parameter) of a transmission line (element).

circuit protection—Automatic protection of a consequence-limiting nature used to minimize the danger of fire or smoke, as well as the disturbance to the rest of the system, that may result from electrical faults or prolonged electrical overloads.

circuit Q—The quality factor of a circuit, equal to the ratio of the circuit reactance to the circuit resistance. In a tuned resonant circuit, it determines the selectivity of the circuit; the higher the Q, the more selective the circuit.

circuit reentrancy—*See* reentrancy, 1.

circuit reliability—The percentage of time the circuit meets arbitrary standards set by the user.

Circuitron—A combination of active and passive components mounted in a single tube-type envelope and functioning as one or more complete operating stages.

circuit switching—A communication method in which an electrical connection between calling and called stations is established on demand for exclusive circuit use until the connection is released.

circuit synthesis—The development of a circuit by the use of theoretical or practical knowledge of basic electronics principles and component parameters. Compare *circuit analysis*.

circular antenna—A horizontally polarized antenna derived essentially from a half-wave antenna but having its elements bent into a circle.

circularly polarized loop vee—An airborne communications antenna that provides an omnidirectional radiation pattern for use in obtaining optimum near-horizon communications coverage.

circularly polarized wave—Applied usually to transverse waves. An electromagnetic wave for which the electric and/or magnetic field vector at a point describes a circle.

circular magnetic wave—A wave with circular magnetic lines of force.

circular mil—A unit of area equal to the area of a circle whose diameter is 1 mil (0.001 in or 25.4 µm); equal to square mil × 0.78540. Used chiefly in specifying cross-sectional areas of round conductors.

circular mil area—The square of the diameter of a round conductor measured in thousandths of an inch. The circular mil area of a braid is the sum of the circular mil area of each of the wires that make up the braid.

circular polarization—1. Polarization such that the vector representing the wave has a constant magnitude and rotates continuously about a point. 2. Simultaneous transmission of vertically and horizontally polarized radio waves. 3. A method of transmitting signals in a rotating corkscrew-like pattern. Both right-hand rotating and left-hand rotating signals can be transmitted simultaneously on the same frequency, thereby doubling the capacity of a satellite to carry communications channels. 4. Polarization of electromagnetic waves whose electric field rotates uniformly along the signal path. Broadcasts used by Intelsat and other international satellites use circular polarization, not horizontally or vertically polarized waves as are common in North American and European transmissions.

circular scanning—Scanning in which the direction of maximum radiation generates a plane or a right circular cone with a vertex angle close to 180°.

circular trace—A CRO time base produced by applying sine waves of the same frequency and amplitude, but 90° out of phase, to the horizontal- and vertical-deflection plates of a cathode-ray tube. This results in a circular trace, and signals then give inward or outward radial deflections from the circle.

circular waveguide—A waveguide having a circular cross-sectional area.

circulating memory—1. *See* circulating register. 2. A type of memory in which a data stream circulates in a loop. One example is a string of shift-register stages with the last output connected to the first input. At every clock pulse, a particular bit would be accessed as it passed a certain point in the circuit. Circulating memories also use other delay techniques, including electrical and acoustical delay lines.

circulating register—Also called circulating memory. A register (or memory) consisting of a means for delaying the information and a means for regenerating and reinserting it into the delaying means. This is accomplished as the information moves around a loop and returns to its starting place after a fixed delay.

circulating storage—A device using a delay line to store information in a train or pattern of pulses. The pulses at the output end are sensed, amplified, reshaped, and reinserted into the input end of the delay line.

circulator—1. A microwave coupling device having a number of terminals so arranged that energy entering one terminal is transmitted to the next adjacent terminal in a particular direction. 2. An arrangement of phase shifters and waveguide or coax that distributes incoming signals among selected outputs. For example, a four-port circulator will transfer a signal entering from port 1 to port 2. In turn, a signal entering port 2 will leave only by port 3. A signal entering port 3 will leave by port 4, and a port-4 entering signal leaves by port 1. Can be used to isolate a transmitter and receiver when both are connected to the same antenna.

circumferential crimp—The type of crimp in which symmetrical indentations are formed in a barrel by crimping dies that completely surround the barrel.

circumnaural—A headphone in which the earpiece completely surrounds the wearer's ear and is sealed to the head to provide tight bass coupling.

CISC—Abbreviation for complex instruction set computer. A type of computer architecture that has long/complex instructions that are general purpose and powerful, but considered to be slower than RISC computers. Intel × 86 chips are CISC technology.

Citizens band—Abbreviated CB. A band of radio frequencies allocated to the Citizens Radio Service.

Citizens Radio Service—A radiocommunications service of fixed, land, and mobile stations intended for short-distance personal or business radiocommunications, radio signaling, and control of remote objects or devices by radio; all to the extent that these uses are not specifically prohibited by the FCC rules and regulations.

cladding—1. A covering for the core of an optical fiber that provides optical insulation and protection. Generally fused to the fiber, it has a low index of refraction. 2. In fiber optics, a sheathing of a lower

Citizens band channels (Class D)

Channel No.	Frequency (MHz)
1	26.965
2	26.975
3	26.985
4	27.005
5	27.015
6	27.025
7	27.035
8	27.055
9	27.065
10	27.075
11	27.085
12	27.105
13	27.115
14	27.125
15	27.135
16	27.155
17	27.165
18	27.175
19	27.185
20	27.205
21	27.215
22	27.225
23*	27.255
24	27.235
25	27.245
26	27.265
27	27.275
28	27.285
29	27.295
30	27.305
31	27.315
32	27.325
33	27.335
34	27.345
35	27.355
36	27.365
37	27.375
38	27.385
39	27.395
40	27.405

*Shared with Class C radio control

refractive index material, intimately in contact with the core of a higher refractive index material, that serves to provide optical insulation and protection to the reflection interface. 3. A method of applying a layer of metal over another metal whereby the junction of the two metals is continuously welded. 4. A relatively thin layer or sheet of metal foil that is bonded to a laminate core to form the base material for printed circuits.

clamp—*See* clamping circuit.

clamper—*See* clamping circuit.

clamping—The process that establishes a fixed level for the picture signal at the beginning of each scanning line.

clamping circuit—A circuit that adds a fixed bias to a wave at each occurrence of some predetermined feature of the wave. This is done to hold the voltage or current of the feature at (clamp it to) a specified fixed or variable level. *See also* dc restorer.

clamping diode—A diode used to fix a voltage level at some point in a circuit.

clamp-on—*See* camp-on.

clamp-on ammeter—An ac ammeter with a built-in current transformer whose core can be clamped around the conductor in which current is to be measured.

clapper—1. A hinged or pivoted armature. 2. In a bell, the ball or hammer that strikes the bell; in an electric bell, it is attached to the vibrating armature.

Clapp oscillator—A Colpitts-type oscillator using a series-resonant tank circuit for improved stability.

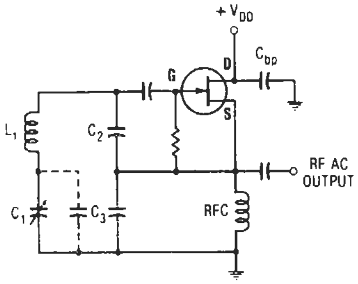

Clapp oscillator.

clarifier—A control on an SSB transceiver that enables adjustment of frequency so that the frequencies of the recovered audio signal will be essentially the same as the frequencies of the modulating signal fed to a distant transmitter. It is carefully adjusted so the received signal sounds natural. Its effective range is usually ±600 to ±1500 Hz.

Clark cell—An early standard cell that used an anode of mercury, a cathode of zinc amalgam, and an electrolyte containing zinc sulfate and mercurous sulfate. Its voltage is 1.433 at a temperature of 15°C.

Clarke belt—The circular orbital belt at 22,238 miles (35,863 kilometers) above the equator, named after the writer Arthur C. Clarke, in which satellites travel at the same speed as the earth's rotation. Also called the geosynchronous or geostationary orbit

Clarke orbit—That circular orbit in space 22,238 miles from the surface of the earth at which geosynchronous satellites are placed. This orbit was first postulated by the science fiction writer Arthur C. Clarke in *Wireless World* magazine in 1945. Satellites placed in these orbits, although traveling around the earth at thousands of miles an hour, appear to be stationary when viewed from a point on earth because the earth is rotating upon its axis at the same angular rate that the satellite is traveling around the earth. *See also* geostationary.

class A amplifier—1. A class of service of a power-amplifer stage in which both bias and drive signal (to be amplified) are adjusted to allow continuous output current at all times. Current through the load resistor generates the output voltage. Efficiency is low—theoretically reaching 50 percent for a pure sine wave, but in practice generally hovering between 20 and 30 percent. This class of service is generally limited to selected audio applications, amateur single-sideband linear rf service, and some CB. 2. An amplifier in which the grid bias and alternating grid voltage are such that plate current flows at all times. To denote that no grid current flows during any part of the input cycle, the suffix "1" is sometimes added to the letter or letters of the class identification. The suffix "2" denotes that grid current flows during part of the input cycle. 3. An amplifier in which the output transistors or tubes are operating permanently on linear portions of their transfer characteristics. Efficiency is low, but a constant current is drawn from the power supply whatever the signal level. Usually recognized by the use of a single transistor or tube driving the loudspeaker. 4. Operation that implies biasing the tubes or transistors to the middle parts of their transfer characteristics so that the device is driven upward on one half-cycle and downward on the other half-cycle. In a push-pull power amplifier, one of the device pair is driven upward on negative half-cycles while its partner is driven downward, the mode reversing on the positive half-cycles. The stage thus draws a constant current at all drive levels within the dynamic range of the amplifier. A class A power amplifier has an efficiency of almost 50 percent. 5. For a transistor amplifier, an amplifier with a single output device that has collector current for the full 360° of the input cycle.

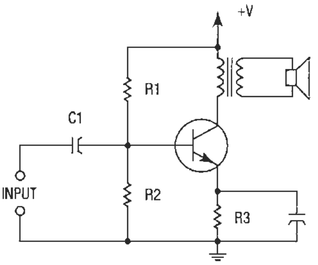

Class A amplifier.

class AB amplifier—1. An amplifier in which the grid bias and alternating grid voltage are such that plate current flows for more than half but less than the entire electrical cycle. To denote that no grid current flows during any part of the input cycle, the suffix "1" is sometimes added to the letter or letters of the class indentification. The suffix "2" denotes that grid current flows during part of the cycle. 2. One type of power amplifier engineered so that at low drive level the stage operates at class A, while at increasing drive level the mode changes to class B.

class A computing device—A computing device for use in a commercial, industrial, or business environment, exclusive of any device marketed for use by the general public or intended to be used in the home.

class A0 emission—The incidental radiation of an unmodulated carrier wave from a station.

class A1 emission—A carrier wave (unmodulated by an audio frequency) keyed normally for telegraphy to transmit intelligence in the International Morse code at a speed not exceeding 40 words per minute (the average word is composed of five letters).

class A2 emission—A carrier wave that is amplitude modulated at audio frequencies not exceeding 1250 hertz. The modulated carrier wave is keyed normally for telegraphy to transmit intelligence in the International Morse code at a speed not exceeding 40 words per minute, the average word being composed of five letters.

class A3 emission—A carrier wave that is amplitude modulated at audio frequencies corresponding to those necessary for intelligible speed transmitted at the speed of conversation.

class A GFCI—A ground fault circuit interrupter that will trip when a fault current to ground is 5 milliamperes or more.

class A insulating material—A material or combination of materials, such as cotton, silk, and paper, suitably impregnated, coated, or immersed in a dielectric liquid such as oil. Other materials or combinations of materials may be included if shown to be capable of satisfactory operation at 105°C.

class A modulator—A class A amplifier used for supplying the signal power needed to modulate the carrier.

class A operation—Operation of a vacuum tube with grid bias such that plate current flows throughout 360° of the input cycle.

Class A signal area—A strong TV signal area, defined by the FCC as receiving a signal strength equal to or greater than approximately 2500 microvolts per meter for channels 2 through 6, 3500 microvolts per meter for channels 7 through 13, and 5000 microvolts per meter for channels 14 through 69.

Class A station—A station in the Citizens Radio Service licensed to be operated on an assigned frequency in the 460- to 470-MHz band with input power of 60 watts or less.

class A transistor amplifier—1. An amplifier in which the input electrode and alternating input signal are biased so that output current flows at all times. 2. An amplifier with a single output device that has collector current for the full 360° of the input cycle.

class B amplifier—1. A power-amplifier stage that offers improved efficiency with linear performance. Class B most commonly is configured in the push-pull arrangement popular in both highly linear audio and rf amplifiers. Adjustments in bias and drive effect a 50-percent output-current duty cycle for each active element. When the collector, drain, or plate voltage is at its maximum value, the current is zero. The fact that only one active element is on at any time creates the equivalent of a continuous current source for the load. Output power is the same as for class A, but input power is proportional to average load current. Efficiency can reach 78 percent, though usually limited to 50 or 60 percent. Amplitude of the output is independent of the supply voltage (provided the amplifier is not saturated). 2. An amplifier in which the grid bias is approximately equal to the cutoff value so that, when no exciting grid voltage is applied, the plate current will be approximately zero and will flow for approximately half of each cycle when an alternating grid voltage is applied. To denote that no grid current flows during any part of the input cycle, the suffix "1" is sometimes added to the letter or letters of the class identification. The suffix "2" denotes that grid current flows during part of the cycle. 3. An amplifier in which two transistors or tubes operate on positive and negative half-cycles of the signal waveform. Each operates from a low initial current, but this rises as the signal level increases. Usually recognized by the use of two transistors or tubes operating in antiphase to drive the loudspeaker. 4. Transistor or tube (power) amplifier whose biasing is adjusted such that the push-pull transistors (or tubes) operate at a low no-drive current (called quiescent current). When drive is applied, the current in one of the pair rises while the partner is pushed into cutoff on one half-cycle, the mode reversing on the other half-cycle.

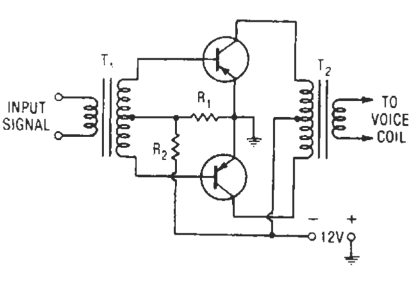

Class B audio amplifier.

class B computing device—A computing device marketed for use in a residential environment, notwithstanding use to commercial, business, and industrial environment. Examples of such devices include electronic games, personal computers, calculators, and similar electronic devices marketed for use by the general public.

class B GFCI—A ground-fault circuit interrupter that will trip when a fault current to ground is 20 milliamperes or more.

class B insulating material—A material or combination of materials such as mica, glass fiber, asbestos, etc., suitably bonded. Other materials or combinations, not necessarily inorganic, may be included if shown to be capable of satisfactory operation at 130°C.

class B modulator—A class B amplifier used specifically for supplying the signal power needed to modulate a carrier.

class B operation—Operation of a vacuum tube with the triad bias set at or very near cutoff, so that plate current flows for approximately the positive half of each cycle of the input signal.

Class B station—A station in the Citizens Radio Service licensed to be operated on an authorized frequency in the 460- to 470-MHz band with input power of 5 watts or less.

class B transistor amplifier—1. An amplifier in which the input electrode is biased so that when no alternating input signal is applied, the output current is approximately zero, and when an alternating input signal is applied, the output current flows for approximately half a cycle. 2. The most common type of audio amplifier, which basically consists of two output devices each of which conducts for 180° of the input cycle.

class C amplifier—1. A power-amplifier stage in which bias is adjusted so that drive-signal voltge produces output current for less than half of the total cycle. Class C amplifiers normally operate in a saturated condition and are insensitive to drive variations. They can be modulated by variations of the power-supply voltage because in saturation they perform as voltage sources. Efficiency can theoretically approach 100 percent, but 60 to 80 percent is typical. Because this class is highly nonlinear, it is restricted to rf applications and enjoys wide popularity in broadcast amplifier applications. 2. An amplifier in which the grid bias is appreciably beyond the cutoff point, so that plate current is zero when no alternating grid voltage is applied, and plate current flows for appreciably less than half of each cycle when an alternating grid voltage is applied. To denote that no grid current flows during any part of the input cycle, the suffix "1" is sometimes added to the letter or letters of the class identification. The suffix "2" denotes that grid current flows during part of the cycle.

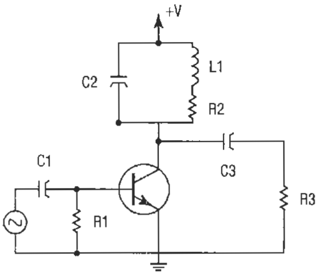

Class C amplifier.

class C insulating material — Insulation consisting entirely of mica, porcelain, glass, quartz, or similar inorganic materials. Other materials or combinations of materials may be included if shown to be capable of satisfactory operation at temperatures over 220°C.

class C operation — Operation of a vacuum tube with grid bias considerably greater than cutoff. The plate current is zero with no input signal to the grid and flows for appreciably less than one-half of each cycle of the input signal.

Class C station — A station in the Citizens Radio Service licensed to be operated on an authorized frequency in the 26.96- to 27.23-MHz band, or on the frequency 27.255 MHz, for the control of remote objects or devices by radio or for the remote actuation of devices that are used solely as a means of attracting attention, or on an authorized frequency in the 72- to 76-MHz band for the control of model aircraft only.

class C transistor amplifier — An amplifier in which the collector current flows for less than 180° of each input cycle. Although highly efficient, high distortion results and the load is frequently tuned to minimize this distortion (primarily used in rf power amplifiers).

class D amplifier — 1. A switching or sampling amplifier with extremely high efficiency (approaching 100 percent). The output devices are used as switches, voltage appearing across them only while they are off and current flowing only when they are saturated. 2. Also called pulse width modulator amplifier. In this amplifier a very high-frequency series of pulses are modulated in their width by the audio signal. The output stages need to conduct for a short interval only to amplify the tips of these pulses; when they do conduct, they are highly efficient — conducting as much as 90 to 95 percent. The high-frequency pulses associated with class D amplifiers (500 kHz or more) present special problems in the transistor switching speed and suitable transistor availability. 3. A power-amplifier stage in which push-pull configuration is driven as a two-position switch alternately connecting the output between V_{DD} and ground. The class D power amplifier delivers a square wave to a tuned circuit that passes only the fundamental (switching) frequency to the load. Extreme linearity and very high efficiency (approaching 100 percent) are possible. Applications consist mainly of audio and low-frequency broadcast equipment.

class D auxiliary power — An uninterruptible (no-break) power unit that makes use of stored energy to provide continuous power within specified tolerances for voltage and frequency.

Class D station — A station in the Citizens Radio Service licensed to be operated on an authorized frequency in the 26.965- to 27.405-MHz band, with input power of 5 watts or less, and to be used for radio telephony only.

class D telephone — A telephone restricted to use in special classes of service, such as fire alarm, guard alarm, and watchman services.

class E amplifier — Another switch-mode design amplifier (*see* class D amplifier) that uses only one active element. It combines the switching action of that element with the transient response of the tuned-load network to achieve efficiencies approaching 100 percent in practice. Use has been generally limited to medium-frequency rf designs.

class F amplifier — A power-amplifier stage very similar to both class D and class E. Class F amplifiers differ in that the tuned output circuit introduces a third-harmonic component properly phased to improve output-power capability. This configuration achieves very high efficiencies, approaching 90 percent in rf applications.

class F insulating material — A material or combination of materials such as mica, glass fiber, asbestos, etc., suitably bonded. Other materials or combinations of materials, not necessarily inorganic, may be included if shown to be capable of satisfactory operation at 155°C.

class G amplifier — 1. A power-amplifier stage in which two class B amplifiers with different supply voltages are combined. Small-amplitude signals are boosted by the one with the lower supply voltage, resulting in much higher average efficiency for speech and music. 2. An amplifier that uses a minimum of two pairs of output transistors. One pair is powered by a lower voltage supply than the other. When signal levels are relatively low, only the low-powered pair of transistors does the amplifying. When signals exceed the low-voltage supply amplitude, the other transistor pair, which operates from the higher voltage supply, takes over, while the first pair is simultaneously turned off. In this way each pair of transistors is always operating over its most efficient range, and overall amplifier efficiency is greater than with a class B design. Thus, less massive heat sinks are needed for the output transistors, and the complete amplifier or receiver is lighter in weight.

class H amplifier — 1. An amplifier somewhat similar to class G operation that uses only one set of output transistors, but these transistors are connected to two different power-supply voltages. The lower voltage powers the output devices for low-level signals, while the higher voltage takes over when the input signal amplitudes exceed the limits of the low-voltage supply. As in class G, this approach results in a more efficient use of the output transistors, and the audio signals themselves do not have to be switched from one device to another during the process. 2. A power-amplifier stage in which the supply voltage of the class B amplifier is varied by an efficient class S amplifier so it remains just above the minimum value required to prevent saturation. This configuration also achieves much higher average efficiency for speech and music signals.

class H insulating material — A material or combination of materials, such as silicone elastomer, mica, glass fiber, asbestos, etc., suitably bonded. Other materials or combinations of materials may be included if shown to be capable of satisfactory operation at 180°C.

class J oscilloscope — *See* J-scope.

class O insulating material—An unimpregnated material or combination of materials, such as cotton, silk, or paper. Other materials or combinations of materials may be included if shown to be capable of satisfactory operation at 90°C.

class S amplifier—A pulse-width-modulated audio amplifier in which the active elements are switched by a control frequency several times higher than the signal frequency being amplified. Class S offers an ideal efficiency of 90 percent.

clavier—Any keyboard, either hand or foot operated.

clean room—A confined area in which the humidity, temperature, and particulate matter are precisely controlled within specified units. The class designation of the clean room defines the maximum number of particles of 0.3-micron size or larger that may exist in one cubic foot of space anywhere in the designated area. For example, in a Class 1 clean room, only one particle of any kind may exist in one cubic foot of space. Newer clean rooms are typically Class 1 to 10, and are needed for manufacturing ICs with feature size close to 1 micron.

clear—1. Also called reset. To restore a storage or memory device to a prescribed or nonprogrammed state, usually to zero or off (empty). 2. Remove all components of a calculation in a calculator. 3. In a calculator, to erase the contents of a display, memory, or storage register. 4. As used in security work, the term *clear* is synonymous with *reset*, meaning that a latched circuit is restored to normal state. 5. Signal to reset or set all signals to an initial known state (usually zero). 6. The process of setting the contents of a register, flag, or memory location to zero. 7. To erase the contents of a display or a memory or storage register.

clearance—The shortest distance through space between two live parts, between live parts and supports or other objects, or between any live part and grounded part.

clear channel—In the standard broadcast band, a channel such that the station assigned to it is free of objectionable interference through all of its primary service area and most of its secondary service area.

clear entry—Remove only the last number, not the entire calculation, in a calculator.

clear entry/clear all—In a calculator, a key used to clear the last entry or to clear the machine completely.

clearing—1. Removal of a flaw or weak spot in the dielectric of a metallized capacitor by the electrical vaporization of the metallized electrode at the flaw. 2. The ability of a lightning protector to interrupt follow current before the operation of circuit fuses or breakers. In the case of a simple gap, clearing frequently requires some external assistance.

clearing ends—The operation of removing the sheath from the end of a cable, eliminating all moisture, and checking for crosses, shorts, and grounds in preparation for testing.

clearing-out drop—A drop signal associated with a cord or trunk circuit and operated by ringing current to attract the operator's attention.

clear input—An asynchronous input to a flip-flop used to set the Q output to logic zero.

clear memory key—Removes what is stored in a memory register of a calculator.

clear raster—A raster free of snow such as would be obtained in the absence of a video signal on either the cathodes or the grids of the three guns in the color CRT (mostly a function of bias conditions).

clear terminal—*See* reset terminal.

clear to send—*See* CTS.

click—To point a mouse pointer at a word or icon on a monitor, press a mouse button, and then release it quickly. Clicking is usually performed to select or deselect an item or to activate a program or a program feature.

click and pop suppressor—An audio-signal-processing accessory. It removes or greatly reduces the audible transient sounds resulting from scratches and blemishes on the surface of a phonograph record.

click filter—A capacitor and resistor connected across the contacts of a switch or relay to prevent a surge from being introduced into an adjacent circuit. *See also* key-click filter.

click-noise modulation—A clipping action performed to increase the bandwidth of a jamming signal. Results in more energy in the sidebands, correspondingly less energy in the carrier, and an increase in the ratio of average power to peak power.

client—A software program or computer that requests information from another computer.

client-server network—A network that uses a central computer (server) to store data that is accessed from other computers on the network (clients).

clipboard—A temporary storage place in a computer where text or graphics are stored.

clipper—A device whose output is zero or a fixed value for instantaneous input amplitudes up to a certain value, but is a function of the input for amplitudes exceeding the critical value.

clipper amplifier—An amplifier designed to limit the instantaneous value of its output to a predetermined maximum.

clipper-limiter—Also called slicer. A device whose output is a function of the instantaneous input amplitude for a range of values lying between two predetermined limits, but is approximately constant at another level for input values above the range.

clipping—1. The loss of initial or final parts of words or syllables due to less than ideal operation of voice-operated devices. 2. Term used to express the clipping of the peaks of a waveform when an amplifier is driven beyond its power capacity. The flattening of the tips of the sine wave due to clipping. 3. Severe distortion caused by overloading the input of an amplifier. A sine-wave signal waveform has a flat top and bottom at the peaks when clipping occurs. 4. The deforming and distortion of speech signals due to limiting the maximum amplitude of the signals. 5. The shearing off of the peaks of a signal. For a picture signal, this may affect either the positive (white) or negative (black) peaks. For a composite video signal, the sync signal may be affected. 6. Removing parts of display elements that lie outside defined bounds. Also called scissoring. 7. The loss of one or more bits at the beginning of a transmission, typically caused by a delay in line turnaround or echo suppression. (May also occur in voice communication, with the loss of the beginning of an initial syllable.)

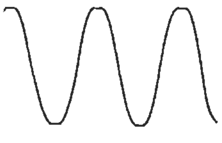

Clipping, 3.

clipping level—The signal level at which clipping (distortion) just begins to occur.

clock—1. A pulse generator or signal waveform used to achieve synchronization of the timing of switching circuits and the memory in a digital computer system. It determines the speed of the CPU. 2. A timing device in a system: usually it provides a continuous series of timing pulses. 3. An electronic circuit that generates timing pulses to synchronize the operation of a computer as well as keep time. 4. A strobe signal that activates a certain sequence of operations. 5. An electronic circuit or device for producing precisely timed, repetitive voltage pulses of fixed frequency and amplitude.

clock cable — Cable of specific impedance and electrical characteristics used to distribute the clock (master) frequency where needed in digital computers.

clocked — Pertaining to the type of operation in which gating is added to a basic flip-flop to permit the flip-flop to change state only when there is a change in the clocking input or an enabling level of the clocking input is present.

clocked flip-flop — A flip-flop circuit designed so that it is triggered only if trigger and clock pulses are present at the same time.

clocked R-S flip-flop — A flip-flop in which two conditioning inputs control the state the flip-flop assumes upon arrival of the clock pulse. If the S (set) input is enabled, the flip-flop assumes the logic 1 condition when clocked; if the R (reset) input is enabled, the flip-flop assumes the logic 0 condition when clocked. A clock pulse must be applied to change the state of the flip-flop.

clock frequency — In digital computers, the master frequency of periodic pulses that are used to schedule the operation of the computer.

clock generator — A test-signal generator that supplies a chain of pulses identical with those supplied by the clock circuit of a digital computer.

clocking — Time-synchronizing communication information.

clock input — That flip-flop terminal whose condition or change of conditions controls the admission of data through the synchronous inputs and thereby controls the output state of the flip-flop. The clock signal permits data signals to enter the flip-flop and, after entry, directs the flip-flop to change state accordingly.

clock pulse — 1. The synchronization signal produced by a clock. 2. A pulse used to gate information into a flip-flop operated in the synchronous mode. (In JK flip-flops, the clock pulse causes counting if the data inputs are both held in logic 1.)

clock rate — 1. The rate at which a word or characters of a word (bits) are transferred from one internal computer element to another. Clock rate is expressed in cycles (in a parallel-operation machine, in words; in a serial-operation machine, in bits) per second. 2. The speed (frequency) at which a processor operates, as determined by the rate at which words or bits are transferred through internal logic sequences. 3. The minimum or maximum pulse rate at which adc counters may be driven. There is a fixed relationship between the minimum conversion rate and the clock rate, depending on the converter accuracy and type. All factors that affect conversion rate of an adc limit the clock rate. 4. The time rate at which pulses are emitted from a clock.

clock skew — 1. Phase shift in a single clock distribution system in a digital circuit. It results from different delays in clock driving elements and/or different distribution paths. 2. Unintentional time difference between clock edges; can exist between clock phases or between clock signals in different parts of a circuit.

clock slips — The relative shift of a system clock with respect to data in synchronous systems. Clock slips can cause modems to lose synchronization.

clock stagger — 1. Time separation of clock pulses in a multiphase clock system. 2. Voltage separation between the clock thresholds in a flip-flop.

clockwise-polarized wave — *See* right-handed polarized wave.

clone — 1. A PC designed to duplicate the behavior and performance of another personal computer, usually an IBM PC. 2. A copy that performs the same as the hardware, software, cellular phone, or computer on which it was based.

close-captioned TV — A text service for the hard-of-hearing TV audience that decodes a text subcarrier and displays it at the bottom of the TV frame on the accompanying video picture. It does not interfere with the standard audio FM subcarrier.

close coupling — 1. Coupling between two circuits so that (*a*) most of the power flowing in one is transferred to the other and (*b*) impedance changes in one circuit greatly affect the other. 2. Also called tight coupling. Any degree of coupling greater than critical coupling.

closed architecture — 1. A system whose characteristics are proprietary and therefore cannot be readily connected with other systems (compare with *open architecture*). 2. Equipment designed to work only with peripherals and accessories made by the same company.

closed array — An array that cannot be extended at either end.

closed circuit — 1. A complete electric circuit through which current may flow when a voltage is applied. 2. A program source, audio or video, that is not broadcast for general consumption, but is fed to remote monitored units by wire.

closed-circuit communication systems — Certain communication systems that are entirely self-contained and do not exchange intelligence with other facilities and systems.

closed-circuit jack — A jack that has its through circuits normally closed. Circuits are opened by inserting a mating plug.

closed-circuit signaling — Signaling in which current flows in the idle conditions and a signal is initiated by increasing or decreasing the current.

closed-circuit system — An intrusion alarm system in which the sensors of each zone are connected in series so that the same current exists in each sensor. When an activated sensor breaks the circuit or the connecting wire is cut, an alarm is transmitted for that zone.

closed-circuit television — Abbreviated CCTV. 1. A television system in which the television signals are not broadcast, but are transmitted over a closed circuit and received only by interconnected receivers. 2. Transmission and reception of video signals via wire carriers.

closed entry — A design that places a limit on the size of a mating part.

closed-entry contact — A female contact designed to prevent the entry of a device that has a cross-sectional dimension greater than that of a mating pin.

close-differential relay — A relay whose dropout value is specified close to its pickup value.

closed loop — 1. A circuit in which the output is continuously fed back to the input for constant comparison. 2. In a computer, a group of indefinitely repeated instructions. 3. A system with feedback control in which the output is used to control the input. 4. An automatic control system in which feedback is used to link a controlled process back to the original command signal. The feedback mechanism compares the actual controlled value with the desired value; if there is any difference, an error signal is created that helps correct the variation. In automation, feedback closes the loop. 5. A control arrangement in which data from the process or device being controlled is fed to the computer to affect the control operation, i.e., the computer can perform all control functions without intervention of an operator. *See also* open loop.

closed-loop bandwidth—The frequency at which the closed-loop gain drops 3 dB from its midband or dc value.

closed-loop drive—1. A tape transport mechanism in which the tape's speed and tension are controlled by contact with a capstan at each end of the head assembly. 2. A tape transport system that drives both incoming and outgoing tape in order to control the portion of the tape contacting the heads and isolate it from the reels or cassette hubs. There are several closed-loop geometries regularly used with open-reel recorders, but dual-capstan drive is the most popular for both open-reel and cassette tapes.

closed-loop feedback—An automatic means of sensing speed variations and correcting to maintain close speed regulation.

closed-loop gain—1. The response of a feedback circuit to a voltage inserted in series with the amplifier input. 2. The overall gain of an amplifier with an external negative-feedback loop.

closed-loop input impedance—The impedance looking into the input port of an amplifier with feedback.

closed-loop output impedance—The impedance looking into the output port of an amplifier with feedback.

closed-loop system—Automatic control equipment in which the system output is fed back for comparison with the input for the purpose of reducing any difference between input command and output response.

closed-loop voltage gain—The voltage gain of an amplifier with feedback.

closed magnetic circuit—A circuit in which the magnetic flux is conducted continually around a closed path through low-reluctance ferromagnetic materials, for example, a steel ring or a toroid core.

closed repeater—A repeater whose access is limited to a select group.

closed routine—In a computer, a routine that is entered by basic linkage from the main routine other than being inserted as a block of instructions within a main routine.

closed subroutine—In a computer, a subroutine not stored in the normal program sequence. Transfer is made from the program to the storage location of the subroutine, and then, following execution of the subroutine, control is returned to the main program.

close memory—Part of a directly addressable computer memory that provides fast cycle time and is usually employed for frequently used accesses.

close-talking microphone—Also called noise-canceling microphone. A microphone designed to be held close to the mouth of the speaker so that ambient noise will not degrade the speech.

closing rating—In a relay, conditions under which the contact must close, with a prescribed duty cycle and contact life.

cloud absorption—Absorption of electromagnetic radiation as a result of water drops and water vapor in a cloud.

cloud attenuation—Reduction in microwave radiation intensity due largely to scattering, rather than absorption, by clouds.

cloverleaf antenna—A nondirectional VHF transmitting antenna that consists of a number of horizontal four-element radiators arranged much like a four-leaf clover, stacked a half-wave apart vertically. These horizontal units are energized to give maximum radiation in the horizontal plane.

club station—A separate amateur radio station for use by the members of a bona fide amateur radio society and licensed to an amateur radio operator acting as the station trustee for the society.

cluster—1. A group of user terminals co-located and connected to one controller, through which each terminal accesses a communication line. 2. A unit of storage that includes one or more sectors of a floppy or hard disk.

clustering—In a computer, the process of grouping things with similar characteristics. A properly programmed computer can take a list of items and group them into clusters.

clutter—Confusing, unwanted echoes that interfere with the observation of desired signals on a radar display.

clutter gating—The technique that provides switching between moving-target indicators and normal video. This results in normal video being displayed in regions with no clutter, and moving-target indicator video being switched in only for the clutter areas.

cm—1. Abbreviation for circular mil. A system for specifying wire size by conductor area. Circular mils are obtained by multiplying the conductor diameter in inches by 1000 and squaring the result. 2. Letter symbol for centimeter.

CML—Abbreviation for current mode logic.

CMOS—*See* complementary metal-oxide semiconductor.

CMR—Abbreviation for common-mode rejection.

CMRR—Abbreviation for common-mode rejection ratio.

CMV—*See* common-mode voltage.

CNC—Abbreviation for computer numerical control. A means of operating production machines (most commonly punch presses and machining equipment) by numerical control instructions generated by CAD/CAM or from a programmable logic controller (PLC).

C network—A network composed of three impedance branches in series. The free ends are connected to one pair of terminals, and the junction points to another pair.

CO—1. Abbreviation for central office. The switching equipment that provides local-exchange telephone service for a given geographical area; designated by the first three digits of the telephone number. 2. *See* cavity oscillator.

coarse-chrominance primary—Also called the Q signal. A zero-luminance transmission primary associated with the minimum bandwidth of chrominance transmission and chosen for its relatively small importance in contributing to the subjective sharpness of the color picture.

coast—On a radar, a memory feature that, when activated, causes the range and/or angle systems to continue to move in the same direction and at the same speed as an original target was moving. Used to prevent lock-up to a stronger target if approached by the target being tracked.

coastal refraction—Bending of the path of a direct radio wave as it crosses the coast at or near the surface. It is caused by differences in electrostatic conditions between soil and water.

coast station—A land-based radio station in the maritime mobile service. It carries on communication with shipboard stations.

coated cathode—In a vacuum tube, a cathode that has been coated with compounds so as to increase its electron emission (e.g., an oxide-coated cathode).

coated filament—A vacuum-tube filament that has been coated with metal oxides to increase electron emission.

coated tape—*See* magnetic-powder-coated tape.

coating—1. The magnetic layer, consisting of oxide particles held in a binder, that is applied to the base film used for magnetic recordings. 2. The magnetizable material on one surface of a recording tape that stores the (audio) signals when recording.

coating thickness—The thickness of the magnetic coating applied to the base film. Modern tape coatings range in thickness from 170 to 650 µin (4.3–16.5 µm), with a preponderance of coatings being approximately 400 µin (10.1 µm) thick. In general, thin coatings give good resolution, at the expense of reduced output at long wavelengths; thick coatings give a high output at long wavelengths, at the expense of degraded resolution.

coax—Abbreviation for coaxial cable.

coaxial—Having a common axis.

coaxial antenna—An antenna comprised of a quarter-wavelength extension to the inner conductor of a coaxial line, and a radiating sleeve that in effect is formed by folding back the outer conductor of the coaxial line for approximately one-quarter wavelength.

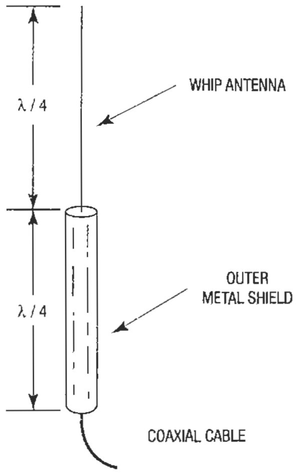

Coaxial antenna.

coaxial cable—Abbreviated coax. 1. A cylindrical transmission line made up of a central conductor (solid or stranded) inside a metallic tube or shield conductor separated by dielectric material in the form of spacers or a solid continuous extrusion. An insulating jacket is optional. 2. A cable consisting of two cylindrical conductors with a common axis. The two conductors are separated by a dielectric. The outer conductor, normally at ground potential, acts as a return path for current through the center conductor and prevents energy radiation from the cable. The outer conductor, or shield, is also commonly used to prevent external radiation from affecting the current in the inner conductor. The outer shield or conductor consists of woven strands of wire or is a metal sheath. 3. Two-conductor wire whose longitudinal axes are coincident; cable with a noise shield around a signal-carrying conductor. *See also* coaxial line.

coaxial cavity—A cylindrical resonating cavity that has a central conductor in contact with its movable pistons or other reflecting devices. The conductor serves to pick up a desired wave in the microwave region.

coaxial conductor—An electric conductor comprising outgoing and returning current paths with a common axis, with one of the paths surrounding the other throughout its length.

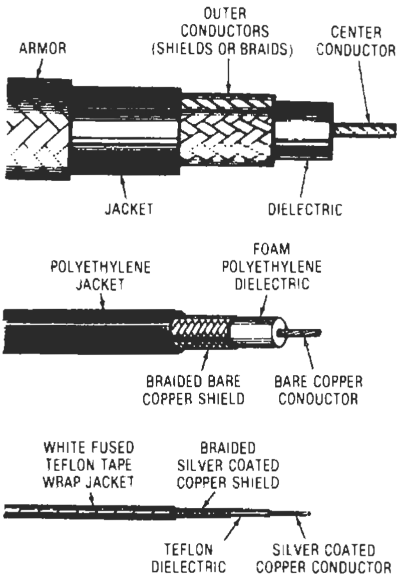
ARMOR
OUTER CONDUCTORS (SHIELDS OR BRAIDS)
CENTER CONDUCTOR
JACKET
DIELECTRIC
POLYETHYLENE JACKET
FOAM POLYETHYLENE DIELECTRIC
BRAIDED BARE COPPER SHIELD
BARE COPPER CONDUCTOR
WHITE FUSED TEFLON TAPE WRAP JACKET
BRAIDED SILVER COATED COPPER SHIELD
TEFLON DIELECTRIC
SILVER COATED COPPER CONDUCTOR

Coaxial cable.

coaxial diode—A diode that has the same outer diameter and terminations as a coaxial cable, into which the diode is designed to be inserted.

coaxial-fed linear array—A beacon antenna having a uniform azimuth pattern.

coaxial filter—A passive, linear, essentially nondissipative network that transmits certain frequencies and rejects others.

coaxial line—Also called coaxial cable, coaxial transmission line, and concentric line. A transmission line in which one conductor completely surrounds the other, the two being coaxial and separated by a continuous solid dielectric or by dielectric spacers. Such a line has no external field and is not susceptible to external fields from other sources.

coaxial-line connector—A device used to provide a connection between two coaxial lines or between a coaxial line and the equipment.

coaxial-line frequency meter—A shorted section of coaxial line that acts as a resonant circuit and is calibrated in terms of frequency or wavelength.

coaxial relay—A type of relay used for switching high-frequency circuits.

coaxial speaker—1. A single speaker comprising a high- and a low-frequency unit plus an electrical crossover network. 2. A tweeter mounted on the axis of and inside the cone of a woofer.

coaxial stub—A short length of coaxial cable joined as a branch to another coaxial cable. Frequently a coaxial stub is short-circuited at the outer end, and its length is so chosen that a high or low impedance is presented to the main coaxial cable at a certain frequency range.

coaxial transistor—A diffused base-alloy emitter, epitaxial mesa germanium semiconductor device with a bandwidth product up to 3 gigahertz, capable of being operated at medium power.

coaxial transmission line—*See* coaxial line.

cobalt doped—Utilizing a combination of "standard" gamma ferric oxide and cobalt as the magnetically active portion of the coating of a recording tape in order to improve maximum output level at low and high frequencies.

COBOL—Acronym for common business-oriented language. A higher-level computer language developed for programming business problems. Used to express problems of data manipulation and processing in English narrative form.

co-channel interference—Interference between two signals of the same type from transmitters operating on the same channel.

Cockcroft-Walton accelerator—A device for accelerating charged particles by the action of a high direct-current voltage on a stream of gas ions in a straight insulated tube. The voltage is generated by a voltage-multiplier system consisting essentially of a number of capacitor pairs connected through switching devices (vacuum tubes). The particles (which are nuclei of an ionized gas, such as protons from hydrogen) gain energies of up to several million electronvolts from the single acceleration so produced. Named for the British physicists J. D. Cockcroft and E. T. S. Walton, who developed this machine in the 1930s.

codan—Acronym for carrier-operated device, anti-noise. An electronic circuit that keeps a receiver inactive except when a signal is received.

codan lamp—A visual indication that a usable transmitted signal has been received by a particular radio receiver.

code—1. A communications system in which arbitrary groups of symbols represent units of plain text of varying length. Codes may be used for brevity or security. 2. System of signaling by utilizing dot-dash-space, mark-space, or some other method in which each letter or figure is represented by prearranged combinations. 3. A system of characters and rules for representing information in a language capable of being understood and used by a computer. Code can be in the form of alphanumeric characters or binary data that can be executed directly by a computer. 4. The language that translates bits into characters in the word processor; it is basically the number of bits the code uses to make up a character and the pattern in which those bits are arranged. For example, ASCII, the industry standard for non-IBM machines, uses seven bits to make up a character, whereas EBCDIC, the IBM system, uses eight bits to make up a character. Systems with different codes will interpret each other's communications incorrectly. 5. A system of using symbols to represent other information. 6. A specific way of using symbols and rules to represent information.

codec—Abbreviation for coder/decoder. 1. A combination device for either digitally coding an analog signal (such as a voice signal) preliminary to transmission operations or decoding a digital signal preparatory to analog processing. 2. A coding-decoding device that converts voice signals into digital bit streams and back again. 3. A complex mixture of analog and digital functions, including (as a minimum) a band-limiting filter, S/H circuit, and nonlinear coder/decoder. 4. The related pair of digital coding and decoding functions in a pulse code modulation (PCM) telephone channel. Two codec pairs are required to complete a bidirectional channel. Each pair permits analog voice and tone signals to be multiplexed, switched, and transmitted in digitally encoded form. The subscriber loop is separated into individual transmit and receive paths by a two- to four-wire hybrid that is part of the subscriber loop interface circuit (SLIC). 5. A single integrated circuit that contains both an analog-to-digital converter (ADC) and a digital-to-analog converter (DAC). A codec performs both pulse code modulation and demodulation.

code character—One of the elements that makes up a code and that represents a specified symbol or value to be encoded. For example, dot-dot-dot-dash is the Morse code character for the letter V.

code conversion—The changing of the bit grouping for a character in one code into the corresponding bit grouping in another code.

code converter—1. A device for translating one code to another. Examples: ASCII to EBCDIC, Gray to BCD, Hollerith to EBCDIC, etc. 2. A decision-making type of digital building block that converts information received at its inputs to another digital code that is transmitted at its outputs. (Also called encoder or decoder.) 3. A ROM or equivalent device that changes inputs into equivalent forms in a desired output format, such as ASCII, IBM Selectric, or Baudot.

coded-alarm system—An alarm system in which the source of each signal is identifiable. This is usually accomplished by means of a series of current pulses that operate audible or visible annunciators or recorders or both to yield a recognizable signal. This technique is usually used to allow the transmission of multiple signals on a common circuit.

coded decimal digit—A decimal digit expressed in terms of four or more 1s and 0s.

code-division multiple access—Abbreviated CDMA. A spread-spectrum scheme, transmitting on a bandwidth much larger than needed but at a much lower level. PN (pseudorandom noise) code is used for retrieving the information. *See also* CDMA.

coded passive reflector antenna—An object intended to reflect Hertzian waves and having variable reflecting properties according to a predetermined code for the purpose of producing an indication on a radar receiver.

coded program—A description of a procedure for solving a problem with a digital computer. It may vary in detail from a mere outline of the procedure to an explicit list of instructions coded in the machine's language.

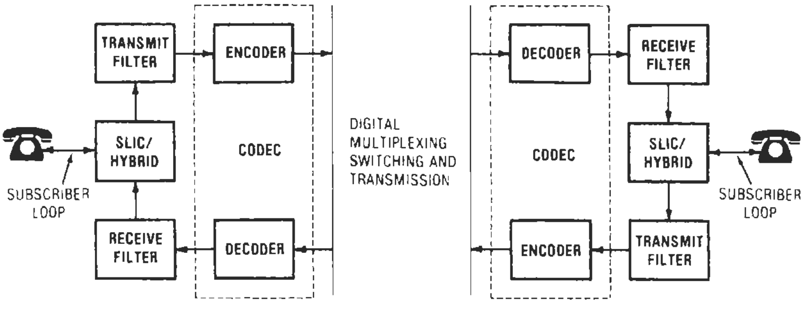

Codec.

code element — 1. One of the finite set of parts of which the characters in a code may be composed. 2. One of the discrete conditions or events in a code, such as the presence or the absence of a pulse.

code holes — The holes in perforated tape that represent information, as opposed to the feed holes or other perforations.

code-practice oscillator — An audio oscillator with a key and either headphones or a speaker, used to practice sending and receiving Morse code.

coder — 1. A device that sets up a series of signals in code form. 2. A beacon circuit that forms the trigger-pulse output of a discriminator into a series of pulses and then feeds them to a modulator circuit. 3. A person who prepares instruction sequences from detailed flowcharts and other algorithmic procedures prepared by others, as contrasted with a programmer, who prepares the procedures and flowcharts.

coder/decoder — *See* codec.

code ringing — In a telephone system, a method of ringing in which the number and/or duration of rings indicate which station on a party line is being called.

code set — 1. The entire set of unique codes that represent specific characters. Different code sets are employed in different equipment. 2. A specific set of symbols and rules used to represent information. 3. All possible values of a code. For an n-bit code, the code set consists of 2^n values. For example, EBCDIC, which is an eight-bit code, has 256 (2^8) possible values. The number of characters a code set can represent may exceed the total number of possible values.

code translation — In telephone operation, the changing of a directory code or number into a predetermined code that controls the selection of an outgoing trunk or line.

coding — 1. Converting program flowcharts into the language used by the computer. 2. The assignment of identification codes to transactions, such as a customer code number. 3. A method of representing characters within a computer. 4. Changing a communications signal into a form suitable for transmission or processing. 5. The preparation of a set of instructions or symbols that, when used by a programmable controller, have a special external meaning.

coding delay — Arbitrary time delay between pulse signals sent by master and slave transmitters.

coding disk — A disk with small projections that operate contacts to generate a predetermined code.

coding form — A form on which the instructions for programming a computer are written. Also called a coding sheet.

coding line — A single command or instruction that directs a computer to solve a problem, usually written on one line.

codiphase radar — A radar system including a phased-array radar antenna and signal-processing and beam-forming techniques.

codistor — A multijunction semiconductor device that provides noise-rejection and voltage-regulation functions.

coefficient — 1. The ratio of change under specified conditions of temperature, length, volume, etc. 2. A number (often a constant) that expresses some property of a physical system in a quantitative way.

coefficient of coupling — *See* coupling coefficient.

coefficient of expansion — The fractional change in dimension of a material for a unit change in temperature.

coefficient of reflection — The square root of the ratio of the power reflected from a surface to the power incident on the same surface.

coefficient of thermal expansion — The average expansion per degree over a specified temperature range, expressed as a fraction of the original dimension. The coefficient may be linear or volumetric.

coercive force — Symbolized by H. The magnetizing force that must be applied to a magnetic material in a direction opposite the residual induction in order to reduce the induction to zero.

coercivity — 1. The property of a magnetic material measured by the coercive force corresponding to the saturation induction for the material. 2. A measure of the amount of applied magnetic field (of opposite polarity) that is necessary to restore a magnetized tape to a state of zero magnetism. High-coercivity tapes exhibit less tendency toward self-erasure and thus have enhanced high-frequency response characteristics, but they require more current through the erase head for full erasure of a recorded signal.

cofire — To place circuits onto an unfired ceramic and fire both circuits and ceramic simultaneously.

cofired ceramic — A material used in making IC packages; its construction is of layers of alumina laminated and fired together to produce a high-strength monolithic structure.

cofiring — Processing the thick-film conductors and resistors through the firing cycle at the same time.

cogging — 1. Nonuniform angular velocity. The armature coil of a motor tends to speed up when it enters the magnetic field produced by the field coils, and to slow down when leaving it. This becomes apparent at low speeds; the fewer the coils, the more noticeable it is. 2. In a motor, the effect caused by improper ratio of stator to rotor slots for a particular speed. It is caused by a magnetic interaction between rotor and stator teeth. In dc motors it is caused by the rich ripple content of the power supply rectifier. The term *cogging torque* is usually applied to reluctance-type synchronous motors. This is the maximum torque of such a synchronous motor before starting to cog or slip, as if a gear were slipping one tooth at a time.

coherence — 1. The property of a set of waves by which their phases are completely predictable along an arbitrarily specified surface in space; also, the relation between a set of sources by which the phases of their respective radiations are similarly predictable. 2. A term used to denote various forms of temporal or statistical phase correlations of electromagnetic fields at different spatial positions; the more extensive the correlations, the greater the coherence. *See* laser; maser. 3. The property of laser light in which the beam emitted is largely of one color or frequency. In the case of the helium-neon laser, this color is red, at 632.9 nm.

coherent bundle — 1. A fiber-optics bundle in which each fiber maintains its relative location throughout the bundle. Thus, an image introduced at one end is transmitted to the other without being scrambled. 2. A bundle of optical fibers in which the spatial coordinates of each fiber are the same or bear the same spatial relationship to each other at the two ends of the bundle. Also called aligned bundle.

coherent carrier — A carrier, derived from a cw signal, the frequency and phase of which have a fixed relationship to the frequency and phase of the reference signal.

coherent carrier system — A transponder system in which the interrogating carrier is raised to a definite multiple frequency and retransmitted for comparison.

coherent detection — A method of deriving additional information from the phase of the carrier.

coherent detector — A detector that gives an output signal amplitude dependent on the phase rather than the strength of the echo signal. It is required for a radar display that shows only moving targets.

coherent display—In random-sampling oscilloscope technique, a plot of a group of samples in which the time sequence of signal events is maintained.

coherent echo—A radar echo that has relatively constant phase and amplitude at a given range.

coherent electroluminescence device—*See* diode laser.

coherent emitter—A source of power that provides a high degree of spectral purity, near-perfect beam collimation, and enormous power densities. Lasers are coherent emitters. *See* incoherent emitter.

coherent fiber bundle—A rigid or flexible fiber bundle that is capable of transmitting an image from one end of the bundle to the other.

coherent interrupted waves—Interrupted continuous waves occurring in wave trains in which the phase of the waves is maintained through successive wave trains.

coherent light—1. A single frequency of light. Light having characteristics similar to a radiated radio wave that has a single frequency. 2. Light of but a single frequency that travels in intense, nearly perfect, parallel rays without appreciable divergence. 3. Light that has the property that, at any point in time or space, particularly over an area in a plane perpendicular to the direction of propagation, or over time at a particular point in space, all the parameters of the wave are predictable and are correlated. 4. Light in which all waves are of exactly the same frequency and exactly in phase. It can therefore act as a carrier and can be modulated for the transmission of information. 5. Light in which the phase relationship between successive waves is such that the beam consists of parallel rays that provide a high concentration of energy.

coherent light communications—Communications using amplitude or pulse-frequency modulation of a laser beam.

coherent light detection and ranging—*See* colidar.

coherent light source—A light source that is capable of producing radiation with waves vibrating in phase. The laser is an example of a coherent light source.

coherent oscillator—An oscillator within some radar sets that furnishes phase references for target returns during intervals between transmitter pulses and that has its output compared with the returns so that the echo becomes coherent video. Coherent video is applied to a cancellation circuit that eliminates nonmoving targets, and only moving targets are supplied to the indicator.

coherent-pulse operation—The method of pulse operation in which a fixed phase relationship is maintained from one pulse to the next.

coherent radar—A type of radar containing circuits that make possible comparison of the phase of successive received target signals.

coherent radiation—1. A form of radiation in which definite phase relationships exist between radiation at different positions in a cross section of the radiant beam. 2. Radiation in which the phase between any two points in the radiation field has a constant difference, or is exactly the same throughout the duration of the radiation. 3. Single-frequency energy such that there is reinforcement when portions of a signal coincide in phase and cancellation when they are in phase opposition.

coherent reference—The reference signal, usually of stable frequency, to which other signals are phase-locked to establish coherency throughout a system.

coherent reflector—Simple or complex surface (such as a corner reflector) from which reflected wave components are coherent with respect to each other and thus combine to yield larger effective power than would be observed from a diffuse scattering surface of the same area.

coherent source—A theoretically ideal light source (fiber optic) that emits a very narrow, unidirectional beam of light of one wavelength (monochromatic). All light emitted from a coherent source is in phase. A laser approximates a coherent source. Contrast *incoherent source.*

coherent system of units—Also called absolute system of units. A system of units in which the magnitude and dimensions of each unit are related to those of the other units by definite simple relationships in which the proportionality factors are usually chosen to be unity.

coherent transponder—A transponder, the output signal of which is coherent with the input signal. Fixed relations between frequency and phase of input and output signals are maintained.

coherent video—A video signal resulting from the combination of a radar echo signal with the output of a continuous-wave oscillator. After delay, the signal so formed is detected, amplified, and subtracted from the next pulse train to give a signal that represents only moving targets.

coherer—An early form of detector used in wireless telegraphy.

coil—Also called inductance and inductor. 1. A number of turns of wire wound around an iron core or onto a form made of insulating material, or one that is self-supporting. A coil offers considerable opposition to the passage of alternating current, but very little opposition to direct current. 2. A number of turns of wire used to introduce inductance into an electric circuit, to produce magnetic flux, or to react mechanically to a charging magnetic flux. In high-frequency circuits, a coil may be only a fraction of a turn. The electrical size of a coil is called inductance and is expressed in henrys. The opposition that a coil offers to alternating current is called impedance and is expressed in ohms. The impedance of a coil increases with frequency.

coil dissipation—The amount of electrical power consumed by a winding. For most practical purposes, this equals the I^2R loss.

coil effect—The inductive effect exhibited by a spiral-wrapped shield, especially above audio frequencies.

coil form—An insulating support of ceramic, plastic, or cardboard onto which coils are wound.

coil loading—As commonly understood, the insertion of coils into a line at uniformly spaced intervals. However, the coils are sometimes inserted in parallel.

coil neutralization—*See* inductive neutralization.

coil rating—*See* input-power rating.

coil resistance—The total terminal-to-terminal resistance of a coil at a specified temperature.

coil serving—A covering, such as thread or tape, that serves to protect a coil winding from mechanical damage.

coil temperature rise—The increase in winding temperature above the ambient temperature when energized under specified conditions for a given period of time, usually that required to reach a stable temperature.

coil tube—A tubular coil form. *See also* spool.

coincidence amplifier—An amplifier that produces no output unless two input pulses are applied simultaneously to the circuit.

coincidence circuit—1. A circuit that produces a specified output pulse when and only when a specified number (two or more) or combination of input terminals receives pulses within a specified time interval. 2. A circuit that has an output signal only when all input signals are present. 3. An AND circuit.

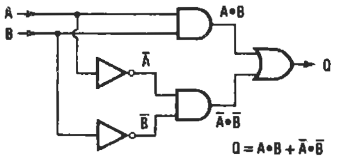

Coincidence circuit.

coincidence counting—The use of electronic devices to detect when two or more pulses from separate counters occur within a given time interval. This is done to determine whether the pulses were produced by the same particle, for example, in scintillation counting, or whether they correspond to the same event. Also called coincidence setting.

coincidence gate—A circuit with the ability to produce an output that is dependent on a specified type or the coincident nature of the input.

coincidence setting—The use of electronic devices to detect when two or more pulses occur within a given time interval. This is done to determine whether the pulses were produced by the same particle, for example, in scintillation counting, or whether they correspond to the same event. Also called coincidence counting.

coincident-current selection—The selection of a magnetic core, for reading or writing, by the simultaneous application of two or more currents.

coincident resonance—A condition in which two systems are joined together (or coupled) and both systems are vibrating near the resonant frequencies at the same time. It often happens that the amplified output from the first system turns out to be the input to the second system, which amplifies that input a second time.

cold—Idiomatic term generally used to describe electrical circuits that are disconnected from voltage sources and are at ground potential. Opposite of the term *hot.*

cold-blade stripper—Device for removing insulation utilizing a cold knife blade.

cold boot—Starting a computer by turning on the power.

cold cathode—1. A cathode whose operation does not depend on the temperature being above the ambient temperature. 2. A cathode that emits electrons, not with the influence of heat radiation, but by means of high-voltage gradient at its surface.

cold-cathode tubes—Tubes in which no external source is used for heating the cathode. These include tubes such as photoelectric cells, gas glow tubes, and mercury rectifiers.

cold cleaning—An organic-solvent cleaning process in which liquid contact accomplishes the solution and removal of residues after soldering.

cold emission—*See* field emission.

cold light—Light that is produced without attendant heat, as from the ionization of a gas by high voltage (neon bulbs, fluorescent lamps), or by electroluminescence, bioluminescence, or cathodoluminescence.

cold-pressure welding—A method of making an electrical connection in which the members to be joined are compressed to the plastic range of the metals.

cold rolling—Rolling a magnetic core alloy into the form of a rod so that the metallic grains are oriented in the long direction of the rod.

cold solder joint—A solder connection exhibiting poor wetting and a grayish, porous appearance due to insufficient heat, inadequate cleaning prior to soldering, or to excessive impurities in the solder solution.

cold weld—A joint between two metals (without an intermediate material) produced by the application of pressure only.

colidar—Acronym for coherent light detection and ranging. An optical radar system that uses the direct output from a ruby laser source without further pulse modulation.

collate—To combine two or more similarly ordered sets of items to produce another ordered set. Both the number and the size of the individual items may differ from those of the original sets or their sums.

collating sequence—1. In digital computers, the sequence in which the characters acceptable to a computer are ordered. The British term is *marshalling sequence.* 2. An ordering assigned to a set of items such that any two sets in that assigned order can be collated.

collator—1. A device to collate sets of punched cards or other documents into a sequence. 2. A device for determining and indicating the coincidence or noncoincidence of two signals.

collection—The mechanism whereby the high-potential gradient and intense electric (drift) field present within the depletion layer of a reverse-biased pn junction can cause the depletion layer to collect any carriers of appropriate type that happen to diffuse it from the adjacent semiconductor regions.

collector—1. In a transistor, the region into which majority carriers flow from the base under the influence of a reverse bias across the two regions. 2. The external terminal of a transistor that is connected to this region. 3. In certain electron tubes, an electrode to which electrons or ions flow after they have completed their function.

collector capacitance—Depletion-layer capacitance associated with the collector junction of a transistor.

collector-coupled logic—*See* current-sourcing logic.

collector current—The direct current flowing in the collector of a transistor.

collector cutoff—The operating condition of a transistor when the collector current is reduced to the leakage current of the collector-base junction.

collector cutoff current—1. The minimum current that will flow in the collector circuit of a transistor with zero current in the emitter circuit. 2. The reverse current of the base-collector junction when the emitter is open circuited, the reverse voltage being specified.

collector-diffusion isolation—Abbreviated CDI. A technique for fabrication of bipolar ICs; the collector diffusion is used to isolate transistors on the same silicon chip electrically, thus reducing the number of photolithographic masking steps required.

collector efficiency—The ratio, usually expressed in percentage, of useful power output to final-stage power-supply power input of a transistor.

collector family—Set of transistor characteristic curves in which the collector current and voltage are variables.

collector junction—The semiconductor junction between the base and collector regions of a transistor. In normal transistor operation, it is reverse biased to collect carriers injected by the emitter (base-to-emitter) junction. In general, the collector junction is designed for a high breakdown voltage, and the emitter junction is designed for a high emitting efficiency.

collector resistance—Resistance in series with the collector lead in the common-T equivalent circuit of a transistor.

collector ring—The collector electrode in an iconoscope.

collector rings—Metal rings suitably mounted on an electric machine that serve, through stationary brushes bearing thereon, to conduct current into or out of the rotating member. *See also* slip ring.

collector transition capacitance—The capacitance across the collector-to-base transition region of a transistor.

collector voltage—The dc collector supply voltage applied between the base and collector of a transistor.

collimated light—1. Parallel light rays, as opposed to converging or diverging rays. 2. A bundle of light rays in which the rays emanating from any single point in the object are parallel to one another, such as the light from an infinitely distant real source or apparent source, such as a collimator reticle.

collimated transmittance—Transmittance of an optical waveguide, such as an optical fiber or integrated optical circuit, in which the light wave at the output has coherency related to the coherency at the input.

collimation—1. The process of adjusting an instrument so that its reference axis is aligned in a desired direction within a predetermined tolerance. 2. The process of making light rays parallel. 3. The property of laser light that keeps the beam from spreading out as it moves away from the laser.

collimation equipment—Equipment designed specifically for aligning optical equipment.

collimation tower—A tower supporting a visual and a radio target used to check the electrical axis of an antenna.

collimator—1. An optical device that creates a beam made up of parallel rays of light, used in testing and adjusting certain optical instruments. It may be used to simulate a distant target, align the optical axes of instruments, or prepare rays for entry into the end of an optical fiber, fiber bundle, or optical thin film. 2. A lens or lens assembly that focuses light into a beam.

collinear array—An antenna array in which half-wave elements are arranged end to end on the same vertical or horizontal line.

collision—Overlapping transmissions when two or more nodes attempt to transmit messages simultaneously or almost simultaneously.

collision-avoidance system—A group of sensors and related instruments placed on board an aircraft to help the pilot detect points of collision and take the appropriate maneuvers to avoid them.

collision detection—The ability of a transmitting node to detect simultaneous transmission attempts on a shared medium.

colocation—The placing of several satellites near each other in orbit. Colocation allows a single fixed receiving antenna to receive signals from all the satellites without moving from one satellite to another (tracking).

color—1. The everyday term for hue. One of the two constituents of the quality chroma (the other being white light). The mind-eye sensation produced by different wavelengths within the visible spectrum produces the colors, some of which are simply identifiable by such names as violet, blue, green, yellow, orange, and red. (Black, white, and gray are not properly described as colors.) 2. Embodies the characteristics of light, other than brightness or luminance, by which a human observer may distinguish between two structure-free patches of light of the same size and shape.

color balance—The adjustment of electron-gun emissions to compensate for the difference in the light-emitting efficiencies of the three phosphors on the screen of the color picture tube.

color bar-dot generator—An rf signal generator that develops a bar or dot pattern on the screen of a color TV picture tube used for test and alignment purposes.

color bars—A test pattern of specifically colored vertical bars used as a reference to test the performance of a color television and transmission paths.

color-bar signal—A test signal used in checking chrominance functions of color TV systems. Typically, it contains bars of six colors, yellow, cyan, green, magenta, red, and blue.

color breakup—Any fleeting or partial separation of a color picture into its display primary components because of a rapid change in the condition of viewing. For example, fast movement of the head, abrupt interruption of the line of sight, and blinking of the eyes are illustrations of rapid changes in the conditions of viewing.

color burst—Also called reference burst. In NTSC color systems, normally refers to a burst of approximately 9 Hz of the 3.579545-MHz unmodulated color subcarrier on the back porch (during the blanking pulse) of the composite video signal. This serves as a color synchronizing signal to establish a frequency and phase reference for the chrominance signal.

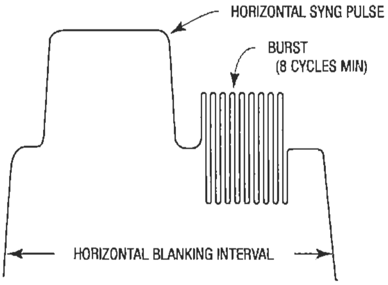

Color burst.

color carrier—*See* chrominance subcarrier.

color-carrier reference—A continuous signal of the same frequency as the color subcarrier and having a fixed phase relationship to the color burst. It is used for modulation at the transmitter and demodulation at the receiver.

colorcast—A color television broadcast.

color code—1. A system of colors for specifying the electrical value of a component part or for identifying terminals and leads. Also used to distinguish between cable conductors. 2. A color system for component identification by use of solid colors, tracers, braids, surface printing, etc.

color coder—Also called color encoder. In a color TV transmitter, that circuit or section which combines the camera signals and the chrominance subcarrier to form the transmitted color picture signal.

color coding—A system of identification of terminals and related devices through the use of colored markings.

color contamination—An error in color rendition due to incomplete separation of the paths carrying different color components of the picture. Such errors can arise in the optical, electronic, or mechanical portions of a color television system.

color-coordinate transformation—Computation of the tristimulus values of colors in terms of one set of primaries from the tristimulus values of the same colors in another set of primaries. Such computation may be performed electrically in a color television system.

color decoder—A section or circuit of a color television receiver used for deriving the signals for the color display device from the color picture signal and the color burst.

color-difference signal—The signal produced when the amplitude of a color signal is reduced by an amount equal to the amplitude of the luminance signal. Color-difference signals are usually designated R-Y, B-Y, and G-Y. In a sense, I and Q signals are also color-difference signals because they are formed when specific proportions of R-Y and B-Y color-difference signals have been combined.

color edging—1. Spurious color at the boundaries of differently colored areas in a picture. 2. Extraneous colors that appear along the edges of video pictures but have no color relationship to those areas.

color encoder—A device that produces an NTSC color signal from separate R, G, and B video inputs. *See* color coder.

color facsimile transmission—The transmission of a facsimile of a color photograph by separating the colors into varying intensities of red, blue, and green and then sending separate transmissions of the three-color reading to a receiving station, which recombines the three signals into the original print.

color fidelity—The degree to which a color television system is capable of faithfully reproducing the colors in an original scene.

color flicker—The flicker that results from fluctuation of both the chromaticity and the luminance.

color fringing—Spurious chromaticity at the boundaries of objects in a color TV picture. It can be caused by the change in relative position of the televised object from field to field, or by misregistration. Color fringing may cause small objects to appear separated into different colors.

color generator—A special rf signal generator for adjusting or troubleshooting a color TV receiver.

colorimeter—An optical instrument designed to compare the color of a sample with that of a standard sample or a synthesized stimulus. (In a three-color colorimeter, the synthesized stimulus consists of three colors of contrast chromaticity but variable luminance.)

colorimetric—Pertaining to the measurement of color characteristics, particularly wavelength and primary-color content.

colorimetric photometer—A photometer that uses a set of color filters to measure the intensity of light in various regions of the spectrum.

colorimetry—The technique of measuring color and interpreting the results.

color killer—A stage designed to prevent signals in a color receiver from passing through the chrominance channel during monochrome telecasts.

color match—The condition in which the two halves of a structureless photometric field look exactly the same.

color media—Transparent colored materials that can be placed in front of an instrument to color the emitted light. These materials are often referred to as gels (for gelatin), but glass or plastic also may be used.

color mixture—Color produced by the combination of lights of different colors. The combination may be accomplished by successive presentation of the components, provided the rate of alternation is sufficiently high; or the combination may be accomplished by simultaneous presentation, either in the same or in adjacent areas, provided the components are small enough and close enough together to eliminate pattern effects.

color oscillator—In a color television receiver, the oscillator operating at the burst frequency of 3.579545 MHz. Its frequency and phase are synchronized by the master oscillator at the transmitter.

color phase—1. The difference in phase between a chrominance primary signal (I or Q) and the chrominance carrier reference. 2. The proper timing relationship with a color signal. Color is considered to be in phase when the hue is reproduced correctly on the screen.

color-phase alternation—Periodic changing of the color phase of one or more components of the color television subcarrier between two sets of assigned values.

color-phase diagram—A vector diagram that denotes the phase difference between the color-burst signal and the chrominance signal for each of the three primary and complementary colors. This diagram also designates vectorially the peak amplitude of the chrominance signal for each of these colors, and the polarities and peak amplitudes of the in-phase and quadrature portions required to form these chrominance signals.

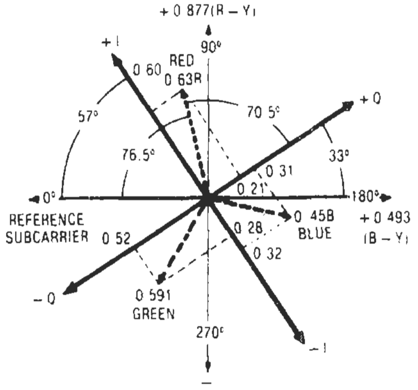

Color-phase diagram.

color picture signal—A signal that represents electrically the three color attributes (brightness, hue, and saturation) of a scene. 2. A combination of the luminance and chrominance signals, excluding all blanking and synchronizing signals.

color picture tube—An electron tube that provides an image in color by scanning a raster and by varying the

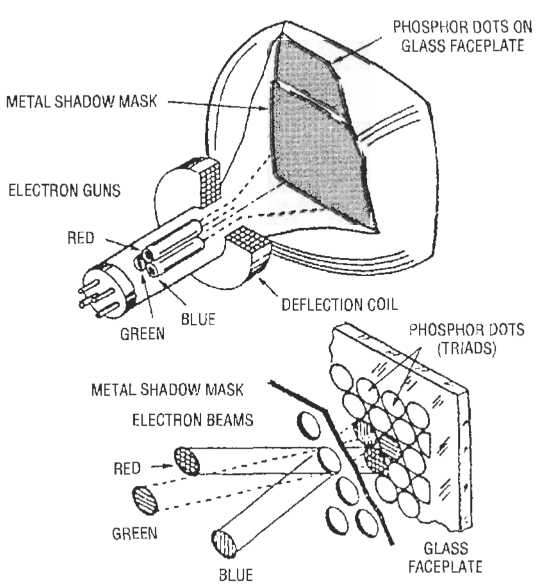

Color picture tube.

intensity at which it excites the phosphors on the screen to produce light of the chosen primary colors.

color primaries—In the color receiver, the saturated colors of definite hue and variable luminance produced by the receiver. These color primaries, when mixed in proper proportions, form other colors.

color purity—Freedom of a color from white light or any colored light not used to produce the desired color. In reference to the operation of a tricolor picture tube, the term refers to the production of pure red, green, or blue illumination of the phosphor-dot faceplate.

color-purity magnet—A magnet placed in the neck region of a color picture tube to alter the electron-beam path and thereby improve color purity.

color registration—The accurate superimposing of the red, green, and blue images that are used to form a complete color picture in a color television receiver.

color rendering index—Abbreviated CRI. A number that approximately represents the effect of a light source on the appearance of colored surfaces. A measure of the degree to which the perceived colors of objects illuminated by the source conform to those of the same objects illuminated by a reference source for specified conditions.

color sampling rate—In a color television system, the number of times per second that each primary color is sampled.

color saturation—1. The degree to which white light is absent in a particular color. A fully saturated color contains no white light. If 50 percent of the light intensity is due to the presence of white light, the color is said to have a saturation of 50 percent. 2. A measure of the amount of white light in a hue. High saturation means that there is no white-light component and that the color is of good quality.

color sensitivity—The spectral sensitivity of a light-sensitive device such as a phototube or camera tube.

color signal—Any signal at any point in a color television system, used for wholly or partially controlling the chromaticity values of a color television picture. This is a general term encompassing many specific connotations, such as are conveyed by the words "color picture signal," "chrominance signal," "carrier color signal," "monochrome signal" (in color television), etc.

color subcarrier—1. A subcarrier that is added to the main video signal to convey the color information. In NTSC systems, the color subcarrier is centered on a frequency of 3.579545 MHz, referenced to the main video carrier. 2. A monochrome signal to which modulation sidebands have been added to convey color information.

color sync burst—A burst of 8 to 11 cycles in the 4.43361875 MHz (PAL) or 3.579545 MHz (NTSC) color subcarrier frequency. This waveform is located on the back porch of each horizontal blanking pulse during color transmissions. It serves to synchronize the color subcarrier's oscillator with that of the transmitter in order to recreate the raw color signals. *See also* color burst.

color sync signal—The series of color bursts (pulses of subcarrier reference signal) applied to the back porch of the horizontal-sync pedestal in the composite video signal.

color television receiver—A standard monochrome receiver to which special circuits have been added. Phosphors capable of glowing in the three primary colors are used on the special screen. By using these primary colors and mixing them to produce complementary colors, and by varying their intensity, it is possible to reproduce an image in somewhat the original colors.

color television signal—The complete signal used to transmit a color picture. Included are horizontal-, vertical-, and color-sync components.

color temperature—1. A way of describing the color of a radiating source in terms of the temperature (in degrees Kelvin) of a blackbody radiating with the same dominant frequency as the source. Certain high-end monitors offer the possibility of setting the color temperature to any desired value. By setting the color temperature, one can often achieve more realistic screen colors. 2. The temperature to which a perfectly black body must be heated to match the color of the source being measured. Color-temperature measurements begin at absolute zero and are expressed in kelvins.

color temperature of a light source—The absolute temperature at which a blackbody radiator must be operated to have a chromaticity equal to that of the light source.

color transmission—1. The transmission of color-television signals that can be reproduced with different values of hue, saturation, and luminance. 2. The transmission of a signal that represents both the brightness values and the color values in a video picture.

color triad—One cell of a three-color phosphor-dot screen of a phosphor-dot color picture tube. Each triad contains one dot of each of the three color-producing phosphors.

color triangle—A triangle drawn on a chromaticity diagram to represent the entire range of chromaticities obtainable when the three prescribed primaries are added. These are represented by the corners of the triangle.

color video—Usually, the technique of combining the three primary colors of red, green, and blue to produce color pictures within the usual spectrum

Colpitts oscillator—A sinusoidal oscillator using a three-terminal active element, such as a tube, transistor, etc., and a feedback loop containing a parallel *LC* circuit. The capacitance of the *LC* circuit consists of two capacitors in series, forming a voltage divider that serves to match the input and output impedance of the active device.

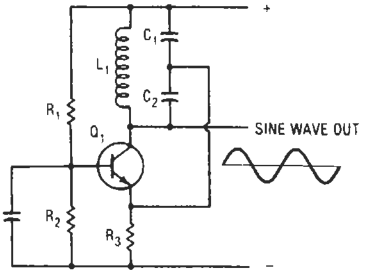

Colpitts oscillator.

column—Also called place. In positional notation, a position corresponding to a given power of the radix. A digit located in any particular column is a coefficient of a corresponding power of the radix.

column binary code—A punched-card code in which successive bits are represented by the presence or absence of punches in adjacent positions in successive columns rather than successive rows. It is used with 36-bit-word computers, in which each group of three columns is used to represent a single word.

column speaker—1. A loudspeaker cabinet of long columnar shape. Usually the loudspeaker is at one end so that the rear of the drive unit is loaded by a column of air. Often columns are made from drain pipes. Also, the name given to a long speaker cabinet containing several loudspeakers for public address work. 2. An array of loudspeakers arranged in a vertical line, having the property of spreading its radiation through a wide angle in the horizontal plane while keeping it in a beam with respect to the vertical plane.

COM — *See* computer-output microfilm printer.

coma — 1. An aberration of spherical lenses, occurring in the case of oblique incidence, when the bundle of rays forming the image is unsymmetrical. The image of a point is comet shaped, hence the name. 2. A cathode-ray tube image defect that makes the spot on the screen appear comet shaped when away from the center of the screen.

coma lobe — A side lobe that occurs in the radiation pattern of a microwave antenna when the reflector alone is moved to sweep the beam. The lobe appears because the feed is not always at the center of the reflector. This scanning method is used to eliminate the need for a rotary joint in the feed waveguide.

comb amplifier — Several sharply tuned bandpass amplifiers whose inputs are connected in parallel and whose outputs are separate. The amplifiers separate various frequencies from a multifrequency input signal. The name is derived from the comblike appearance of the response pattern of various output peaks displayed along a frequency-base axis.

comb filter — 1. A filter for high resolution and picture sharpness that eliminates "dot crawl" and "hanging dots" caused by cross-color and cross-luminance distortion. Comb filters provide exceptional isolation of chrominance and luminance information and increased picture resolution for sharper, clearer, more detail-perfect color picture reproduction. Capable of delivering 700 lines of horizontal resolution, at the S-video input, it greatly exceeds the maximum clarity of the 330 lines of horizontal resolution delivered by broadcast television and easily supports the demands of videodisc, CD-I, and other advanced high-resolution video media. 2. A type of filter network that is, in effect, a multiple-bandpass design that passes only frequencies within a number of narrow bands, or provides outputs corresponding to each of its passbands. Comb filters are so named because their response characteristics have the appearance of a comb. 3. A filter that has a series of very narrow, deep notches where signals are attenuated.

comb generator — Usually a step-recovery diode circuit that converts a single frequency rf input into an rf output signal that contains a large number of spectral lines, each of which is harmonically related to the input frequency.

combination cable — A cable in which the conductors are grouped in combinations such as pairs and quads.

combination microphone — A microphone consisting of two or more similar or dissimilar microphones combined into one.

combination tones — Frequencies produced in a nonlinear device, such as in an audio amplifier, having appreciable harmonic distortion.

combinatorial logic — Digital circuitry in which the states of the outputs from a device depend only on the states of the inputs. *See also* sequential logic.

combined-gate IC — A single IC chip in which several gate circuits are interconnected to form a more complex circuit.

combiner — A circuit for mixing video, trigger, and scan data from the synchronizer for the modulation of a link.

combiner circuit — A circuit that combines the luminance and chroma channels with the sync signals in color-television cameras.

combustible — *See* flammable.

come-back — A point in the stopband of a filter where a spurious response occurs beyond points at which there is proper attenuation. Come-backs usually occur at frequencies much higher than the passband frequencies because of feed-through in parasitic elements.

command — 1. In a computer, one or more sets of signals that occur as the result of an instruction. *See also* instruction, 1. 2. An independent signal from which the dependent signals in a feedback control system are controlled according to the prescribed system relationships. 3. A signal that initiates or triggers an action in the device that receives the signal. *See* operation code. 4. An issue to the PC to execute a function, such as print, file, erase, or send a document.

commandable bug — A surreptitious listening device that can be turned on and off remotely.

command code — *See* operation code.

COMMAND.COM — A disk file that contains the command processor and must be present on the start-up disk for DOS to run.

command control — A system whereby functions are performed as the result of a transmitted signal.

command destruct signal — A radio signal for destroying a missile in flight.

command file — A computer file that will execute a program when the filename is entered.

command guidance system — A missile guidance system in which both the missile and the target are tracked by radar. The missile is guided by signals transmitted to it while it is in flight. *See also* command link.

command language — A computer source language that consists primarily of procedural operators, each of which is capable of invoking a function to be executed.

command link — The portion of a command guidance system used to transmit steering commands to the missile. *See also* command guidance system.

command module — *See* module.

command net — A communication network that connects an echelon of command with some or all of its subordinate echelons for the purpose of command and control.

command pointer — A multiple-bit register that indicates the memory location being accessed in the control store.

command reference — In a servo or control system, the voltage or current to which the feedback signal is compared. As an independent variable, the command reference exercises complete control over the system output.

comment — An expression that explains or identifies a particular step in a routine but which has no effect on the operation of the computer in performing the instructions for the routine.

comment field — In a computer, an area in a record assigned for entry of explanatory comments about a program.

commercial test (measuring) equipment — Devices used as working instruments selected from suppliers' catalogs as suitable for the user's measuring needs and procured as standard (off-the-shelf) items. This includes measuring devices that are installed, intact, and in consoles. Devices procured from suppliers but modified by user-imposed specifications in such a manner as to affect console performance do not fall into this category.

commercial time-sharing — A type of computer use with remote terminal, interactive, and time-sharing characteristics. The user of the machine pays an amount determined by the time used.

committed power and ground — Patterns that have been permanently established to input/output connectors for power and ground plane busing.

common — 1. Shared by two or more circuits. Used to designate the terminal of a three-terminal device that is shared by the input and output circuits. (Thus, a transistor may be operated in a common-base configuration, a common-collector configuration, or a common-emitter configuration.) Vacuum tube connections may be characterized in a similar way, but *grounded* is normally used instead of *common*. 2. A point that acts as the reference potential for several circuits — a ground. 3. *See* tailor-made.

common base — A circuit configuration in which the base terminal is common to the input circuit and to the output circuit and in which the input terminal is the emitter terminal and the output terminal is the collector terminal.

common-base amplifier — Also called a grounded-base amplifier. A transistor amplifier in which the base element is common to both the input and the output circuit. It is comparable to the grounded-grid configuration of a triode electron-tube amplifier.

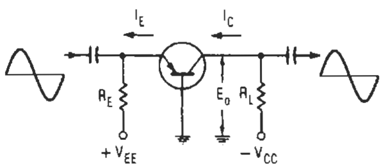

Common-base amplifier.

common-base circuit — A transistor circuit in which the base electrode is common to both input and output circuits.

common-base feedback oscillator — A common-base bipolar transistor amplifier with a feedback network between the collector (output) and the emitter (input) to produce oscillations at a desired frequency.

common-base transistor — Circuit configurations in which the base terminal is common to the input circuit and to the output circuit and in which the input terminal is the emitter terminal and the output terminal is the collector terminal.

common battery — A system of current supply in which all direct-current energy for a unit of a telephone system is supplied by one source in a central office or exchange.

common-battery office — A central office that supplies transmitter and signal current for its associated stations, and for signaling by the central office equipment, from a power source located in the central office.

common business-oriented language — Specific language by which business data-processing procedures may be precisely described in a standard form. Intended not only as a means for directly presenting any business program to any suitable computer for which a compiler exists, but also as a means of communicating such procedures among individuals. Also called COBOL.

common bus system — A set of standard data, address, and control lines available to all computer modules. The use of bus interface circuits makes it possible for a user to tie in and communicate with other users.

common carrier — 1. A government-regulated (by the FCC) private company that furnishes the general public with telecommunications service facilities, for example, a telephone or telegraph company. Specialized common carriers offer private line services. 2. A commercial company that sells communications services to any member of the public. A common carrier cannot provide programming for the channels itself.

common-carrier fixed station — A fixed station that is open to public correspondence.

common-channel interference — Radio interference resulting from two stations transmitting on the same channel. Characterized principally by beat-note generation (heterodyne whistle) and the suppression or capture of the weaker signal by the stronger one.

common-channel signaling — A method of using a single signaling channel to carry signaling information relating to a number of information channels. Common-channel signaling information is sent in packet form.

common collector — A circuit configuration in which the collector terminal is common to the input circuit and to the output circuit and in which the input terminal is the base terminal and the output terminal is the emitter terminal.

common-collector amplifier — Also known as an emitter-follower and a grounded-collector amplifier. A transistor amplifier in which the collector element is common to both the input and the output circuit. This configuration is comparable to an electron-tube cathode follower.

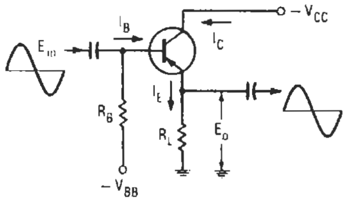

Common-collector amplifier.

common-collector circuit — A bipolar transistor circuit in which the collector is the common (grounded) electrode.

common-collector transistor — Circuit configuration in which the collector terminal is common to the input circuit and to the output circuit and in which the input terminal is the base terminal and the output terminal is the emitter terminal.

common communications carrier — A company recognized by an appropriate regulatory agency as having a vested interest in furnishing communications services.

common emitter — A circuit configuration in which the emitter terminal is common to the input circuit and to the output circuit and in which the input terminal is the base terminal and the output terminal is the collector terminal.

common-emitter amplifier — Also called grounded-emitter amplifier. A transistor amplifier in which the emitter element is common to both the input and the output circuit. This configuration is comparable to a conventional electron-tube amplifier.

common-emitter circuit — A bipolar transistor circuit in which the emitter is the common (grounded) electrode.

common-emitter transistor — Circuit configuration in which the emitter terminal is common to the input circuit and to the output circuit and in which the input terminal is the base terminal and the output terminal is the collector terminal.

common language — A form representing information that a machine can read and that is common to a group of computers and data-processing machines.

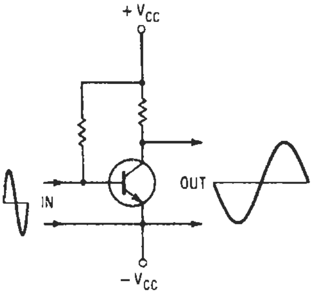

Common-emitter amplifier.

common mode—1. Signals that are identical with respect to both amplitude and time. Also used to identify the respective parts of two signals that are identical with respect to amplitude and time. 2. A high-speed-modem interface name.

common-mode characteristics—The characteristics pertaining to performance of an operational amplifier in which the inverting and noninverting inputs have a common signal.

common-mode coupling—Coupling that results in similar signals with respect to ground on different circuit leads.

common-mode gain—The ratio of the output voltage of a differential amplifier to the common-mode input voltage. The common-mode gain of an ideal differential amplifier is zero. Typical values in op amps are around −30 dB.

common-mode impedance input—The internal impedance between either one of the input terminals of a differential operational amplifier and signal ground.

common-mode input—That signal applied in phase (i.e., common mode) equally to both inputs of a differential amplifier.

common-mode input capacitance—The equivalent capacitance of both inverting and noninverting inputs of an operational amplifier with respect to ground.

common-mode input voltage—The maximum voltage that can be applied simultaneously between the two inputs of a differential amplifier and ground without causing damage.

common-mode interference—Interference that appears between the terminals of the measuring circuit and ground.

common-mode output voltage—The output voltage of an operational amplifier resulting from the application of a specified voltage common to both inputs.

common-mode range—Maximum voltage that can be applied to differential inputs with respect to ground. The maximum difference between inputs is the full-scale input range

common-mode rejection—Abbreviated CMR. 1. Also called in-phase rejection. A measure of how well a differential amplifier ignores a signal that appears simultaneously and in phase at both input terminals (called a common-mode signal). Usually and preferably stated as a voltage ratio, but more often stated in the dB equivalent of said ratio at a specified frequency, e.g., "120 dB at 60 Hz with a source impedance of 1000 ohms." 2. A measure of the ratio of differential-mode gain to the common-mode gain present in all practical amplifiers. When both inputs to the amplifier are raised by the same voltage, the output should, ideally, be unaffected.

common-mode rejection in decibels—Twenty times the log of the common-mode rejection ratio.

common-mode rejection ratio—Abbreviated CMRR. 1. The ratio of the common-mode input voltage to the output voltage expressed in dB. The extent to which a differential amplifier does not provide an output voltage when the same signal is applied to both inputs. 2. The ratio of differential-mode gain to common-mode gain. 3. The ratio of the change of input offset voltage of an operational amplifier to the change in common-mode voltage producing it.

common-mode resistance—The resistance between the input- and output-signal lines and circuit ground. In an isolated amplifier, this is its insulation resistance. (Common-mode resistance has no connection with common-mode rejection.)

common-mode signal—1. The instantaneous algebraic average of two signals applied to a balanced circuit (i.e., two ungrounded inputs of a balanced amplifier), with all signals referred to a common reference. 2. In an amplifier with a differential input, a signal (referred to ground) that appears at both inverting and noninverting inputs with the same phase, amplitude, and frequency. Power-line hum is the most frequently encountered common-mode signal.

common-mode voltage—Abbreviated CMV. 1. The amount of voltage common to both input lines of a balanced amplifier. Usually specified as the maximum voltage that can be applied without breaking down the insulation between the input circuit and ground. (Common-mode voltage has no connection with common-mode rejection.) 2. The voltage component of a two-wire input (control) signal that is common to both lines. The actual control signal is the difference between the two input voltages; for example, if the two input signals are (with respect to a common ground) 110 volts and 100 volts, then the common-mode voltage is 100 volts, and the control voltage is 10 volts. 3. An undesirable signal picked up in a transmission line by both wires making up the circuit, to an equal degree, with respect to an arbitrary "ground."

common-mode voltage gain—The ratio of ac voltage with respect to ground at the output terminal of an amplifier (or between the output terminals of an amplifier with differential outputs) to the common-mode input voltage.

common-mode voltage range—The range of voltage that may be applied to both inputs of an operational amplifier without saturating the input stage. This may limit the output capabilities in the voltage-follower connection.

common pool—A dedicated area of memory used as storage and shared by various processes.

common-user channels—Communication channels that are available to all authorized agencies for transmission of command, administrative, and logistic traffic.

common-user circuit—A circuit shared by two or more services, either concurrently or on a time-sharing basis. It may be a unilateral, bilateral, or joint circuit.

communal chained memory—A technique employed in dynamic storage allocation in a computer.

communicating word processor—*See* electronic mail.

communication—1. The transmission of information from one point, person, or equipment to another. 2. The sensing of a measurement signal or phenomena for display, recording, amplification, transmission, computing, or processing into useful information. 3. Transmission of intelligence between points of origin and reception without alteration of the sequence or structure of the information content.

communication band—The band of frequencies due to the modulation (including keying) necessary for a given type of transmission.

communication channel—Part of a radio or wire circuit, or a combination of wire and radio that connects two or more terminals.

communication control character—A character whose purpose is to control or facilitate data transmission over communication networks.

communication engineer—An electrical engineer who specializes in the design, construction, operation, or maintenance of communication circuits, equipment, or systems whether by wire or radio (wireless).

communication facilities—Installations, equipment, and personnel used to provide telecommunication.

communication-line controller—A hardware unit that performs line-control functions with a modem.

communication link—The physical means of connecting one location to another for the purpose of transmitting and receiving information.

communications common carriers—Companies that furnish communications services to the public, regulated by the FCC or appropriate state agencies.

communications port—A connection on a terminal through which data is input and/or output.

communications program—Software that gives a computer the ability to communicate with other computers over telephone lines.

communications protocol—A set of rules that govern the communications between computers over telephone lines. Both computers must have the same settings and follow the same standards for communication to be successful.

communications receiver—A receiver designed for reception of voice or code signals from stations operating in the communications service.

communications satellite—An orbiting space vehicle that actively or passively relays signals between communications stations.

communications security—The protection resulting from all measures designed to deny unauthorized persons information of value that might be derived from the possession and study of telecommunications or to mislead unauthorized persons in their interpretations of the results of such possession and study.

communication switch—A device used to execute repetitive sequential switching.

communication zone indicator—A device that indicates whether or not long-distance high-frequency broadcasts are successfully reaching their destination.

community antenna television—Abbreviated CATV. A television system that receives and retransmits television broadcasts. Microwave transmitters and coaxial cables are used to bring the television signals to subscribers in a community.

community bulletin board—A dial-up computer used to exchange messages.

community dial office—A small dial-telephone office that serves an exchange area and that operates with no employees located in the building.

community television system—A receiving system by means of which television signals may be distributed over coaxial cables to homes in an entire community.

community TV cable—Coaxial cables used to transmit television signals from a master antenna to a group of receivers in a community.

commutating capacitor—Also called speedup capacitor. 1. In a flip-flop circuit, a capacitor connected in parallel with the cross-coupling resistor to accelerate the transition from one stable state to another. 2. The capacitor connected in parallel between SCR stages to momentarily reverse the current through the SCR so as to cut it off (commutate).

commutating filter—1. A bandpass filter whose center frequency depends only on the frequency with which it is driven. 2. A clocked, switched-capacitance digital filter that uses periodic signal sampling techniques to synthesize discrete approximations of all the classic analog filters.

commutation—1. A mechanical process of converting the alternating current in the armature of direct-current generators into the direct-current generator output. 2. Sampling of various quantities in a repetitive manner for transmission over a single channel. 3. The switching of currents back and forth between various paths as required for operation of some system or device. In particular, a switching of current to or from the appropriate armature coils of a motor or generator. The turning off of an active element at the correct time as in an inverter or power controller. 4. The transfer of load current from one thyristor to another; sometimes used as a means of turning off a thyristor by rapidly reducing its current to zero.

commutation capacitors—1. Cross-connected capacitors in a thyratron inverter. They provide a path such that the start of conduction in one thyratron causes an extinguishing pulse to be applied to the alternate thyratron. Also used in inverter circuits employing semiconductor devices. 2. A specially designed capacitor used in the turn-off (commutation) circuit of an SCR, where it is subjected to exceedingly fast rise time pulses. Thus the capacitor must be capable of discharging large peak currents in very short periods of time.

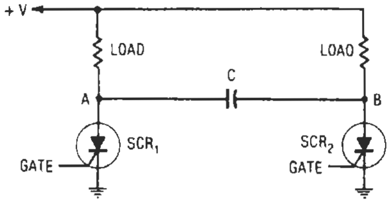

Commutation capacitor.

commutation switch—A device used to carry out repetitive sequential switching.

commutator—1. The part of the armature to which the coils of a motor are connected. It consists of wedge-shaped copper segments arranged around a steel hub and insulated from it and from one another. The motor brushes ride on the outer edges of the commutator bars and thereby connect the armature coils to the power source. 2. Device used in a direct-current generator to reverse the direction of an electric current and maintain a current flowing in one direction. 3. A switch or equivalent device that permits the reversal or exchange of external connections of a transducer to provide a desired sequencing of signals.

compact cassette—A small (4 × 2½ × ½ inch or 10.2 × 6.3 × 1.3 cm) tape cartridge developed by Philips, containing tape about ⅛-inch (3.2 mm) wide, running at 1⅞ ips (12.1 cm/s). Recordings are bidirectional, with both stereo tracks adjacent for compatibility with monophonic cassette recorders, whose heads scan both stereo tracks at once.

compact disc (CD) — 1. A read-only plastic disc that uses optical storage techniques to store large amounts of music (audio) or digitally encoded data. 2. A stereo digital audio system based upon a 12 cm single-sided polycarbonate disc with an internal reflective layer of aluminum or gold. Information is read by sensing the presence or absence of reflected light from a tightly focused laser beam pointing at the pits and bumps in the reflective surface within the disc. The digital audio signal is sampled to 16 bits per channel at a rate of 44.1 kHz. 3. Small plastic disc on which digital information is recorded. The data thus recorded can be played back using a laser. The packing density on the disc is very high, enabling hundreds of megabytes of data to be stored.

compactron — An electron tube based on a building-block concept that involves the standardizing of basic tube sections, diodes, triodes, and pentodes; clipping them together as required; and sealing them in a single envelope.

compander — 1. A combination consisting of a compressor at one point in a communication path to reduce the volume range of signals, followed by an expander at another point to improve the ratio of the signal to the interference entering the path between the compressor and expander. 2. A device consisting of a compressor at the transmitting end and an expander at the receiving end that operate as nonlinear amplifiers to obtain a more advantageous amplitude-quantizing relationship for the reduction of noise. The process of companding is particularly important with audio signals. The principles have also been effectively used in the processing of picture information. 3. A complementary noise-reduction system in which both pre- and postprocessing are used to provide no alteration of signal while reducing noise. It improves the signal-to-noise ratio in systems in which an analog signal is passed through a noisy transmission medium. 4. An electronic circuit that amplifies low-input signal amplitudes more than high amplitudes so that all signal amplitudes are compressed into a narrow range. This compressed signal can be expanded back into the original signal by amplifying large amplitudes more than smaller amplitudes. 5. A device used on some telephone channels to improve transmission performance. The equipment compresses the outgoing speech volume range and expands the incoming volume range on a long-distance telephone circuit.

companding — A process in which compression is followed by expansion. Companding is often used for noise reduction, in which case the compression is applied at the transmitter before the noise exposure and the expansion at the receiver.

compandor — 1. Abbreviation for compressor-expander. An audio-frequency circuit that reduces (compresses) loud sounds and increases (expands) quiet sounds so that the output level stays almost constant. A compressor is used in the transmitter. In the receiver, an expander would follow the demodulator and precede the audio power amp. 2. *See* compander.

companion keyboard — A remote keyboard connected by a multiwire cable to an ordinary keyboard and able to operate it.

comparator — 1. A circuit that compares two signals and supplies an indication of agreement or disagreement. 2. In a computer, a circuit that determines whether the absolute difference between a data sample and the previous sample passed is greater than or equal to a redundancy criterion (which may be a tolerance or a limit). 3. A device that compares two inputs for equality. One type compares voltages and gives one of two outputs: less than or greater than. Another type compares binary numbers and has three outputs: less than, equal to, or greater than. A third type compares phase or frequency and gives a variable voltage depending on the relationship between the inputs. 4. A unit often found in audio showrooms, which, by switch selection, will connect up a combination of speakers, amplifier, tuner, pickup, tape player, etc., for comparing different types. 5. A circuit that compares two signals and provides a difference signal. 6. An active device that provides a logical 0 when its input is below a preset reference value and a logical 1 when its input is above that value. 7. A device that compares two different signals and provides an output when they differ in frequency, phase, voltage, or power level. 8. A circuit that evaluates an output parameter to determine if it falls within some predetermined limits. 9. A device for checking the accuracy of one set of data by comparing it with a second set of data and then noting any variation between the two.

compare — 1. A computer operation in which two quantities are matched for the purpose of discovering their relative magnitudes or algebraic values. 2. A computer instruction that effectively subtracts one word from another and indicates which of the two is larger.

comparison — The examination of how two similar items of data are related. The comparison is usually followed by a decision.

comparison bridge — A type of voltage-comparison circuit resembling a four-arm electrical bridge. The elements are so arranged that if a balance exists in the circuit, a zero error signal is derived.

comparison testing — Real-time comparison between the actual output responses of the device under test and those of a known good reference device when the same input stimulus patterns are applied to both devices in parallel.

compatibility — 1. That property of a color television system that permits unaltered monochrome receivers to receive substantially normal monochrome from the transmitted signal. 2. The property that makes possible use of a stereo system with a monophonic program source, or reproduction of a stereo program monophonically on a monophonic system. 3. The ability of one unit to be used with another without detrimental effect on the signal through mismatch. For example, a compatible pickup will play both mono and stereo records. 4. The ability to run the same software programs and connect the same peripherals and add-on equipment (e.g., boards, printers, modems) as another PC. When a machine is said to be compatible, it is more often than not compatible with the IBM PC. Compatible PCs are also referred to as clones.

compatible — A term applied to a computer system that implies that it is capable of handling both data and programs devised for some other type of computer system.

compatible color — A TV broadcast system that produces a color signal that can be received by either a black-and-white or a color set. The luminance values (the basis of black-and-white reception) and the chrominance values (the basis of color reception) are broadcast as different portions of the total signal so that the luminance values are not dependent on the chrominance values for reproduction.

compatible integrated circuit — A hybrid IC in which the active circuit element is within the silicon planar integrated structure. A passive network, which may be separately optimized, is deposited onto its insulating surface to complete the IC device.

compatible monolithic integrated circuit — A device in which passive components are deposited by thin-film techniques on top of a basic silicon-substrate circuit containing the active components and some passive parts. *See also* all-diffused monolithic integrated circuit.

compensated amplifier— 1. A broadband amplifier whose frequency range is extended by the proper choice of circuit constants. 2. A wideband amplifier made so by the addition of low- and high-frequency compensation.

compensated-impurity resistor— A diffused-layer resistor into which are introduced additional n- and p-type impurities.

compensated-loop direction finder— A direction finder employing a loop antenna and a second antenna system to compensate for polarization error.

compensated semiconductor— A semiconductor in which one type of impurity or imperfection (donor) partially cancels the electrical effects of the other (acceptor).

compensated volume control— *See* loudness control.

compensating filter— A filter used to alter the spectral emission of an emulsion to a specified response to different wavelengths.

compensation— 1. The controlling elements that compensate for, or offset, the undesirable characteristics of the process to be controlled in the system. 2. The shaping of an op-amp frequency response in order to achieve stable operation in a particular circuit. Some op amps are internally compensated, whereas others require some external compensation components in some circuits. 3. The phenomenon whereby extremely small quantities of donor and acceptor impurities present in a semiconductor crystal tend to cancel out each other, so that the material tends to behave according to the dominant impurity only. If both types of impurity are present to an equal extent, the material tends to behave as if it were an intrinsic material.

compensation signal— A signal recorded on a tape, along with the computer data and on the same track as the data; this signal is used during the playback of data to electrically correct for the effects of tape-speed errors.

compensation theorem— An impedance in a network may be replaced by a generator of zero internal impedance, the generated voltage of which at any instant is equal to the instantaneous potential difference produced across the replaced impedance by the current flowing through it.

compensator— 1. In a direction finder, the portion that automatically applies to the direction indication all or part of the correction for the deviation. 2. An electronic circuit for altering the frequency response of an amplifier system to achieve a specified result. This refers to record equalization or loudness correction.

compile— 1. To bring digital-computer programming subroutines together into a main routine or program. 2. To produce a binary coded program in a computer from a program written in source (symbolic) language by selecting appropriate subroutines from a subroutine library as directed by the instructions or other symbols of the source program. The linkage is supplied for combining the subroutines into a workable program, and the subroutines and linkage are translated into binary code.

compiler— 1. An automatic coding system in a computer that generates and assembles a program from instructions written by a programmer. 2. A unit that converts computer programs written in higher-level languages, such as BASIC and C++, into the machine language (object code) of the computer. It is necessary to write an entire program into a compiler's memory before the compiler executes or performs a translation on it. 3. Computer routine that translates symbolic instructions to machine instructions and replaces certain items of input with series of instructions, called subroutines. 4. A high-level, Englishlike programming language that converts instructions into executable machine code. Two examples of such programming languages are COBOL and FORTRAN. 5. A high-level language processor that converts or translates a sequence of source language statements into a corresponding sequence of machine language instructions that may be later loaded into memory and executed by the processor to perform the desired functions. No matter how many times a section of code is used in the assembled program, it will be translated only once and put in its proper place. 6. A program that converts a high-level language into machine language for a specific microprocessor/computer. 7. A program that translates high-level language programs into a series of machine-code instructions for a computer to execute; it may also check the programs' semantic consistency.

compiler language— A computer language system consisting of various subroutines that have been evaluated and compiled into one routine that can be handled by the computer. FORTRAN, COBOL, and ALGOL are compiler languages. Compiler language is the third level of computer language. *See* machine language, 3, for other levels. *See also* high-level language.

compiler program— Software, usually supplied by the manufacturer, to convert an application program from compiler language to machine language.

compiling routine— A routine by means of which a computer can itself construct the program used to solve a problem.

complement— 1. In an electronic computer, a number whose representation is derived from the finite positional notation of another by one of the following rules.

True complement: Subtract each digit from 1 less than the base, then add 1 to the least significant digit and execute all required carries.
Base-minus-1s complement: Subtract each digit from 1 less than the base. (For example, 9s complement in base 10, 1s complement in base 2, etc.)

2. To form the complement of a number. (In many machines, a negative number is represented as a complement of the corresponding positive number.) The binary opposite of a variable or function. The complement of 1 is 0 and the complement of 0 is 1; thus, for example, the complement of 011010 is 100101.

complementary— 1. A term describing integrated circuits that employ components of both polarity types connected in such a way that operation of either is complemented. A complementary bipolar circuit would employ both npn and pnp transistors, and a complementary MOS circuit (CMOS) would employ both n-channel and p-channel devices. In general, complementary devices operate with opposite polarity voltages and currents, which is advantageous in many circuit applications. 2. Two driving-point functions whose sum is a positive constant.

complementary binary or inverted binary— The negative true binary system. It is similar to the binary code except that all binary bits are inverted. Thus, zero scale is all 1s while full scale is all 0s.

complementary circuit— A circuit that provides push-pull operation (sink and source capability) with a single input.

complementary clocks— Two clock signals with opposite phase.

complementary colors— Two colors are complementary if, when added together in proper proportion such as by projection, they produce white light.

complementary constant current logic (C^3L)— A high-density approach to bipolar LSI that has switching speeds of 3 nanoseconds.

complementary metal-oxide semiconductor—Device formed by the combination of a PMOS and an NMOS (p-type and n-type channel semiconductors), exhibiting very low power consumption and high noise immunity. Strictly speaking, CMOS refers to an IC manufacturing technology; the term is almost always used to describe an IC logic family. The CMOS logic family is characterized by very low power dissipation, low circuit density per chip, and moderate speed of operation when compared with other IC logic families. *See also* complementary MOS; ECL; I^2L; Schottky TTL; TTL.

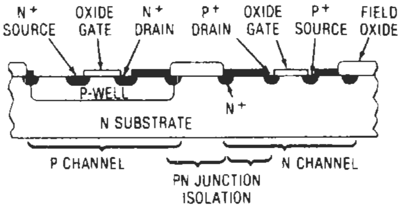

Conventional CMOS.

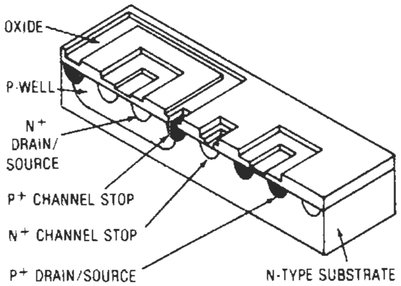

CMOS process.

Complementary metal-oxide semiconductor (CMOS).

complementary MOS—Abbreviated CMOS. 1. Pertaining to n- and p-channel enhancement-mode devices fabricated compatibly on a silicon chip and connected into push-pull complementary digital circuits. These circuits offer low quiescent power dissipation and potentially high speeds, but they are more complex than circuits in which only one channel type (generally p-channel) is used. 2. A digital inverter consisting of a p-channel and an n-channel enhancement-mode field-effect transistor. The transistors are connected in series across the power supply with gates linked together as the input.

complementary operator—The logic operator whose result is the NOT of a given logic operator.

complementary push-pull—A power amplifier in which the output transistors are of complementary polarities (i.e., pnp and npn). In some amplifiers of this kind the driven transistors also constitute a complementary pair.

complementary rectifier—Half-wave rectifying circuit elements that are not self-saturating rectifiers in the output of a magnetic amplifier.

complementary silicon-controlled rectifier—A pnpn semiconductor device that is the polarity complement of the silicon-controlled rectifier.

complementary-symmetry circuit—An arrangement of pnp and npn transistors that provides push-pull operation from one input signal.

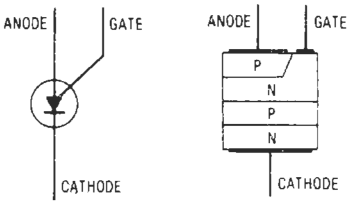

Symbol. Junctions.

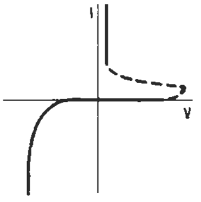

Anode characteristic

Complementary silicon-controlled rectifier.

complementary tracking—A system of interconnection of two or more devices in which one (the master) operates to control the others (the slaves).

complementary transistor amplifier—An amplifier that utilizes the complementary symmetry of npn and pnp transistors.

complementary transistor logic—A digital logic circuit configuration making use of a complementary transistor emitter coupled AND-OR gate. Basically, a two-level diode gate using simultaneous npn and pnp action.

complementary transistors—Two transistors that have opposite conductivity (pnp and npn) and usually have matching electrical characteristics; npn and pnp pairs with similar electrical characteristics.

complementary unijunction transistor—Abbreviated CUJT. 1. An integrated semiconductor structure with characteristics similar to those of a unijunction transistor, but complementary to other unijunction transistors in the way that pnp transistors are complementary to npn transistors. 2. A silicon planar device similar to a unijunction transistor except that the operating currents and voltages are of the opposite polarity. The electrical characteristics are stable, consistent, and predictable over a wide temperature range. The CUJT will operate from a 5-volt supply and is therefore compatible with integrated circuits. Typically, the case is electrically connected to the substrate and must be isolated from the circuit.

complementary wave—A wave brought into existence at the ends of a coaxial cable or two-conductor transmission line, or any discontinuity along the line.

complementary wavelength—The wavelength of light of a single frequency. When combined with a sample color in suitable proportions, the wavelength matches the reference-standard light.

complementing—Changing each binary 1 into a 0 and each 0 into a 1.

complement number—A number that, when added to another number, gives a sum equal to the base of the numbering system. For example, in the decimal system, the complement of 2 is 8. Complement numbers are used in some computer systems to facilitate arithmetic operations.

complement number system—A system of number handling in which operations are performed on the complement of the actual number. The system is used in some computers to facilitate arithmetic operations.

complementor—A circuit or device that produces a Boolean complement. A NOT circuit.

complete carry—A system of executing the carry process in a computer. All carries, and any other carries to which they give rise, are allowed to propagate to completion in this system.

complex components—Indivisible and nonrepairable components having more than one function.

complex function—An integrated device in which three or more circuits are integral to a single silicon chip. In addition, the circuits are interconnected on the chip itself to form some electronic function at a higher level of organization than a single circuit. The interconnection pattern for the function is predetermined by the fixed mask; no wiring discretion is available for yield purposes. The input and outputs of all the circuits are not normally exposed to the package terminals. An example of a complex function would be a full adder or a multibit serial shift register.

complex instruction set computer—*See* CISC.

complex parallel permeability—The complex relative permeability measured under stated conditions on a core with the aid of a coil. The parameter characterizing the induction is the impedance of the coil when placed on the core, expressed as a parallel connection of reactance and resistance. The parameter characterizing the field strength is the reactance the coil would have if placed on a core of the same dimensions but with unity relative permeability, the distribution of the magnetic field being identical in both cases. The coil should have negligible copper losses.

complex permeability—Under stated conditions, the complex quotient of the moduli of the parallel vectors representing induction and field strength in a material. One of the moduli varies sinusoidally with time, and the component chosen from the other modulus varies sinusoidally at the same frequency.

complex series permeability—The complex relative permeability measured under stated conditions on a core with the aid of a coil. The parameter characterizing the induction in the core is the impedance of the coil when placed on the core, expressed as a series connection of reactance and resistance. The parameter characterizing the field strength is the reactance this coil would have if placed on a core of the same dimensions but with unity relative permeability. The coil should have negligible copper losses.

complex steady-state vibration—A periodic vibration of more than one sinusoid. It includes repeating square waves, sawtooth waves, etc., because these waveforms can be expressed in terms of a Fourier series of sinusoidal terms.

complex target—A radar target made up of a number of reflecting surfaces that, taken together, are smaller in all dimensions than the resolution capability of the radar.

complex tone—A sound wave produced by the combination of simple sinusoidal components of different frequencies. A sound sensation characterized by more than one pitch.

complex wave—A periodic wave made up of a combination of several frequencies or several sine waves superimposed on one another.

complex-wave generator—A device that generates a nonsinusoidal signal having a desired repetitive characteristic and waveform.

compliance—1. The reciprocal of stiffness; that is, the ability to yield or flex. 2. The ease with which a phonograph stylus can be deflected by the groove wall. Expressed in microcentimeters per dyne (10^{-8} m/dyne) as the distance through which the stylus will be deflected by a force of 1 dyne. Typical values are from 10 to 50 microcentimeters per dyne. In general, the higher the compliance, the better the low-frequency tracking at a given tracking force. Compliance can be measured in several ways—static and dynamic. 3. The mechanical and acoustical equivalent of capacitance. 4. The flexibility of a speaker cone's suspension. High compliance is important in a woofer for accurate reproduction of low-frequency signals of large amplitude.

compliance extension—A form of master/slave interconnection of two or more current-regulated supplies to increase their output voltage range through series connection.

compliance range—The range of voltage needed to sustain a given constant current throughout a range of load resistances.

compliance voltage—The output voltage of a dc power supply operating in constant-current mode.

compliance voltage range—The output voltage range of a dc power supply operating in a constant-current mode.

component—1. An essential functional part of a subsystem or equipment. It may be any self-contained element with a specific function, or it may consist of a combination of parts, assemblies, accessories, and attachments. 2. In vector analysis, one of the parts of a wave, voltage, or current considered separately. 3. A packaged functional unit consisting of one or more circuits made up of devices, which in turn may be part of an operating system or subsystem. A part of, or division of, the whole assembly or equipment. Normally interchangeable with *unit*. 4. In high fidelity, a specialized item of equipment designed to do a particular part of the work in a sound system. 5. Any of the basic parts used in building electronic equipment, such as a resistor or capacitor.

component density—The number of components contained in a given volume or within a given package or chip. The quantity of components on a printed board per unit area.

component layout—The physical arrangement of the components in a chassis or printed circuit.

component level bus—The set of input and output pins, with defined functions and timing, through which a microprocessor sends and receives signals.

component operating hours—A unit of measurement for the period of successful operation of one or more components (of a specified type) that have endured a given set of environmental conditions.

component part—A term sometimes used to denote a passive device.

component placement equipment—Automatic systems for sorting and placing components onto hybrid circuit substrates or printed circuit boards, consisting of indexing conveyor, sorter, placement heads, missing component detector, programmable electropneumatic control, and options to handle special requirements.

component population—The variety and number of components (transistors, resistors, transformers, etc.) necessary to perform the desired electrical function.

component side—That side of a printed board on which most of the components will be mounted.

component stress—Those factors of usage or test, such as voltage, power, temperature, frequency, etc., that tend to affect the failure rate of component parts.

composite cable—A cable in which conductors of different gages or types are combined under one sheath.

composite circuit—A circuit that can be used simultaneously for telephony and direct-current telegraphy or signaling, the two being separated by frequency discrimination.

composite color signal—The color picture signal plus all blanking and synchronizing signals. Includes luminance and chrominance signals, vertical- and horizontal-sync pulses, vertical- and horizontal-blanking pulses, and the color-burst signal.

composite color sync—The signal comprising all the sync signals necessary for proper operation of a color receiver. Includes the deflection sync signals to which the color sync signal is added in the proper time relationship.

composite conductor—Two or more strands of different metals, such as aluminum and steel or copper and steel, assembled and operated in parallel.

composite controlling voltage—The voltage of the anode of an equivalent diode, combining the effects of all individual electrode voltages in establishing the space-charge limited current.

composited circuit—A circuit that can be used simultaneously for telephony and for direct-current telegraphy or signaling, separation between the two being accomplished by frequency discrimination.

composite dialing—In telephone operations, a method of dialing between distant offices over one leg of a composite set.

composite filter—A filter with two or more sections.

composite guidance system—A guidance system using a combination of more than one individual guidance system.

composite picture signal—The television signal produced by combining a blanked picture signal with the sync signal.

composite signal—The stereo FM broadcast modulation signal consisting of a 19-kHz pilot tone, L + R information and L − R information modulated on a suppressed 38-kHz carrier and (if any) a 67-kHz FM carrier with a ±6-kHz deviation SCA channel.

composite sync signal—The position of the composite video signal that synchronizes the scanning process.

composite TV signal—A combination of video picture, color, audio, and synchronization information.

composite video—A standard type of video in which red, green, and blue signals are mixed together.

composite video signal—1. The complete video signal, containing both picture and sync information. For monochrome, it consists of the picture signal and the blanking and synchronizing signals. For color, color-synchronizing signals and color-picture information are added. 2. The combined signals in a television transmission, including the picture signal, vertical and horizontal blanking, and synchronizing signals. 3. The complete video signal, consisting of the chrominance and luminance information as well as all sync and blanking pulses.

composite wave filter—A combination filter consisting of two or more low-pass, high-pass, bandpass, or band-elimination filters.

composite wire clad—A wire having a core of one metal to which is fused an outer shell of one or more different metals.

composition—The conversion of computer files containing text and typesetting commands into a format for input to an imaging system, such as a phototypesetter.

composition resistor—*See* carbon resistor.

compound-connected transistor—Two transistors that are combined to increase the current amplification factor at high emitter currents. This combination is generally employed in power-amplifier circuits.

compound-filled transformer—A transformer that is contained within a case and in which the structural insulating material is supplemented by submergence in a solid or semisolid insulating material introduced into the case in a fluid state.

compound horn—An electromagnetic horn of rectangular cross section. The four sides of the horn diverge in such a way that they coincide with or approach four planes, with the provision that the two opposite planes do not intersect the remaining planes.

compounding—A form of noise reduction using compression at the transmitting end and expansion at the receiver. A compressor is an amplifier that increases its gain for lower-power signals. The effect is to boost these components into a form having a smaller dynamic

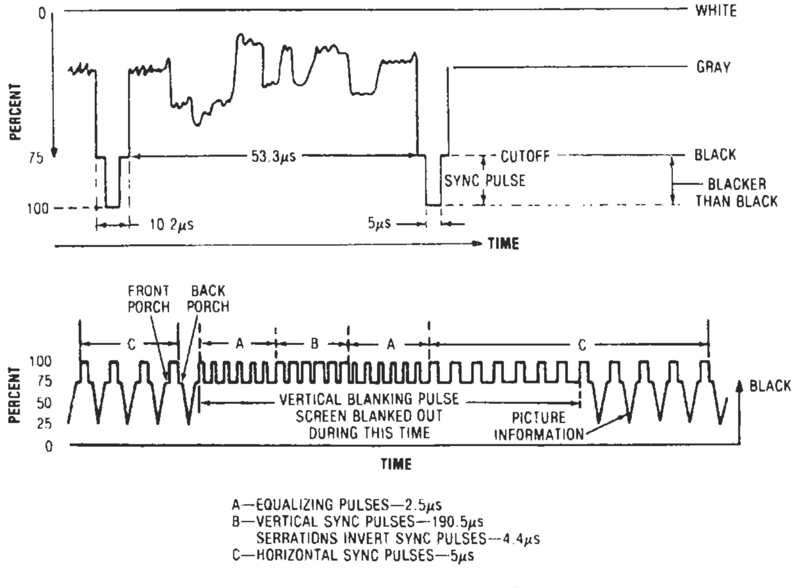

Composite picture signal.

range. A compressed signal has a higher average level and therefore less apparent loudness than an uncompressed signal, even though the peaks are no higher in level. An expander reverses the effect of the compressor to restore the original signal.

compound modulation — *See* multiple modulation.

compound-wound motor — A dc motor having two separate field windings. One, usually the predominant field, is connected in parallel with the armature circuit, and the other is connected in series.

compress — To reduce some parameter of a signal, such as bandwidth, amplitude variation, duration, etc., while preserving its information content.

compressed-air loudspeaker — A loudspeaker that has an electrically actuated valve to modulate a stream of compressed air.

compressed file — A file whose contents have been compressed by a special utility program so that it occupies less space on a disk or other storage device than in its uncompressed (normal) state.

compressed speech — A representation of speech in which some redundant features of the digitized speech have been removed.

compression — 1. A process in which the effective amplification of a signal is varied as a function of the signal magnitude, the effective gain being greater for small than for large signals. In television, the reduction in gain at one level of a picture signal with respect to the gain at another level of the same signal. 2. Electronic reduction of the dynamic range so that quiet sounds are raised and loud sounds lowered. The most common application is an "automatic" recording in which it is important that all sounds recorded are made intelligible when played back. Also used when necessary to avoid overrecording and distortion, or to lift the signal level clear of background noise or hum. 3. A technique used to increase the number of bits per second sent over a data link by replacing often-repeated characters, strings, and command sequences with electronic code. When this compressed data reaches the remote end of the transmission link, data decompression is used to restore the data to its normal form for display. 4. The conversion of information to a format that requires fewer bits and can be reversed to its original state once transferred to a new location.

compressional wave — In an elastic medium, a wave that causes a change in volume of an element of the medium without rotation of the element.

compression driver unit — A speaker driver unit that does not radiate directly from the vibrating surface. Instead, it requires acoustic loading from a horn that connects through a small throat to an air space adjacent to the diaphragm.

compression ratio — The ratio between the magnitude of the gain (or amplification) at a reference signal level and its magnitude at a higher stated signal level.

compression seal — A seal made between an electronic package and its leads. The seal is formed as the heated metal, when cooled, shrinks around the glass insulator, thereby forming a tight joint.

compressor — 1. A device that performs analog compression. 2. A transducer that, for a given amplitude range of input voltages, produces a smaller range of output voltages. In one important type of compressor, the envelope of speech signals is used to reduce their volume range.

compressor expander — *See* compander.

compromise network — In a telephone system, a network used in conjunction with a hybrid coil to balance a subscriber's loop. The network is adjusted for an average loop length, an average subscriber's set, or both, and gives compromise (not precision) isolation between the two directional paths of the hybrid coil.

Compton diffusion — An elastic shock between a photon and an electron. The photon is diffused with a lesser energy and the electron acquires a kinetic energy equal to the energy decrease of the photons.

Compton effect — The elastic scattering of photons by electrons. Because the total energy and momentum are conserved in the collisions, the wavelength of the scattered radiation undergoes a change that depends on the scattering angle.

computational stability — The degree to which a computational process remains valid when subjected to such effects as errors or malfunctions.

compute bound — A program that is speed-limited by the computations being performed.

computer — 1. Any device capable of accepting information, applying prescribed processes to the information, and supplying the results of these processes; sometimes, more specifically, a device for performing sequences of arithmetic and logical operations; sometimes, still more specifically, a stored-program digital computer capable of performing sequences of internally stored instructions, as opposed to calculators on which the sequence is impressed manually (desk calculator) or from tape or cards (card-programmed calculator). 2. A tool for managing data. It can work with numbers and alphanumeric data such as names, words, addresses, and stock numbers. It can be programmed to repeat the same function over and over. It can logically evaluate information given to it, and act on its own findings. It can store huge volumes of data for future use, reference, and updating, and even "converse" with its operator. 3. A machine in which stored instructions operate on other instructions to modify or alter them. Data words and instruction words of the same size, stored in the same medium, differ only in their function. The same word can be both a data word and an instruction word at different times during the execution of a program. 4. An electrical/electronic device that can accept information, process it mathematically (in accordance with previous instructions), and then provide the results of this processing. 5. An electronic device that uses programmed instructions to monitor and control various types of data in order to solve mathematical problems or control industrial applications. Its instructions are executed in various sequences, as required. 6. A calculating device that processes data represented by a combination of discrete data (in digital computers) or continuous data (in analog computers). 7. A device that manipulates data and makes comparisons according to a series of instructions stored in its memory. By changing the instructions the computer can be made to do a completely different

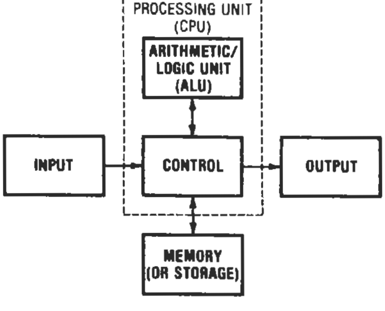

Computer.

task. Instructions and data are stored in the same memory and both can be manipulated by the computer with equal ease. 8. A device that is capable of solving problems or manipulating data by accepting data, performing prescribed operations (mathematical or logical) on the data, and then delivering or applying the results of these operations.

computer access device input—A device that automatically routes to the computer all teletypewriter observation reports that are received in a standard format.

computer-aided design—*See* CAD.

computer-aided engineering—Abbreviated CAE. An umbrella term that covers all uses of computers in engineering applications. Thus, computer-aided design and computer-aided manufacturing are branches of computer-aided engineering. The subject area is not usually considered to include software engineering.

computer-aided manufacturing—Abbreviated CAM. The use of computer technology to manage, control, and operate manufacturing either through direct or indirect computer interface with the physical and human resources of the company.

computer-aided software engineering—*See* CASE.

computer-aided tomography—*See* CAT.

computer architecture—That set of a computer's attributes (such as registers, addressing modes, and instruction set) that are visible to the programmer.

computer assisted tomography—*See* CAT.

computer code—Also called machine language. The code by which data is represented within a computer system. An example is binary-coded decimal.

computer control—The parts of a digital computer that have to do with the carrying out of instructions in the proper sequence, the interpretation of each instruction, and the application of signals to the arithmetic unit and other parts in accordance with this interpretation.

computer control counter—1. A counter that stores the next required address. 2. Any counter that provides information to the control unit.

computer diagnosis—The use of data processing systems for evaluation of raw data.

computer entry punch—A combination card reader and keypunch used to enter data directly onto the memory drum of a computer.

computer-generated hologram—A synthetic hologram produced using a computer plotter. The binary structure is formed on a large scale and is then photographically reduced. The holograms are finally etched into a medium.

computer graphics—1. Computer output in the form of pictorial representation (graphs, charts, drawings, etc.) that is displayed visually, usually by a cathode-ray tube. 2. A person-oriented system that uses a computer to create, transform, and display pictorial and symbolic data.

computer-integrated manufacturing—*See* CIM.

computer interface—1. Peripheral equipment for attaching a computer to scientific or medical instruments. 2. A device designed for data communication between a central computer and another unit such as a PC processor.

computer interfacing—The synchronization of digital data transmission between a computer and one or more external I/O devices.

computerized axial tomograph—*See* CAT.

computerized robot—A servo model run by a computer. The computer controller does not have to be taught by leading the arm-gripper through a routine: new instructions can be transmitted electronically. The programming for such "smart" robots may include the ability to optimize, or improve, its work-routine instructions.

computer language—1. A system of commands used to develop software for computers (e.g., DOS). 2. The method or technique used to instruct a computer to perform various operations. *See* high-level language; machine language.

computer-limited—Having to do with the condition in which the time required for computation is greater than the time required to read inputs and write outputs.

computer literacy—1. Computer and information system comprehension. 2. The ability to use computer technology in a particular discipline.

computer network—Two or more connected computers that have the ability to exchange information.

computer numerical control—Abbreviated CNC. The use of a dedicated computer within a numerical-control unit to perform some or all of the basic numerical-control functions.

computer-output microfilm printer—Abbreviated COM printer. A microfilm printer that will take output directly from the computer, thus substituting for line printer or tape output.

computer polarization holography—A technique used to store wavefront information on thin polarization information-recordable materials (i.e., photochromic crystals) by controlling the polarization angle of a small illuminating spotlight in each sampling cell on a crystal.

computer port—The physical location at which the communication line interfaces to the computer.

computer program—A series of instructions or statements prepared in a form acceptable to the computer, the purpose of which is to achieve a certain result. *See* software.

computer programmer—A person who designs, writes, debugs, and documents computer programs.

computer programming language—A set of precisely defined structures and syntax (representation, conventions, and rules of use and interpretation) devised to simplify communication with a computer, such as BASIC, FORTRAN, C++, and Java. The greater the power of a higher-level language, the greater is the complexity of information that can be precisely conveyed in an efficient manner.

computer science—1. The field of knowledge that involves the design and use of computer equipment, including software development. 2. The science of solving problems with computers.

computer system—The computer and its attached peripherals, such as disk drives, monitor, keyboard, and printer.

computer tape—A high-quality magnetic digital recording tape that must be rated at 1600 fci (flux changes per inch) or 530 flux changes per centimeter, or greater.

computer terminal—Peripheral computer equipment for entering and retrieving data. Sometimes incorporates cathode-ray tube for display.

computer user tape system—*See* CUTS.

computer utility—A network of central computers linked through data communications facilities to remote terminal systems.

computer word—A sequence of bits or characters that is treated as a unit and that can be stored in one computer location. Same as machine word.

computing—Performing basic and more involved mathematical processes of comparing, adding, subtracting, multiplying, dividing, integrating, etc.

computing device—Any electronic device or system that generates and uses timing signals or pulses of more than 10,000 pulses (cycles) per second and uses digital techniques; inclusive of telephone equipment that uses digital techniques or any device or system that generates and uses rf energy for data processing functions, such as electronic computations, operations, transformations, recording, filing, sorting, storage, retrieval, or transfer. (Radio transmitters and receivers; industrial, scientific, and medical equipment; and any other radio-frequency devices specifically subject to an examination requirement are excluded from this definition.)

computing machine—An automatic device that carries out well-defined mathematical operations.

COMSTAR I, II, and III—Three American satellites that can carry video and are operated by AT & T. They are located at 128°, 95°, and 87° west longitude.

concatenate—To unite in a series; to connect together; to chain. Used to describe the action of relating data used by a computer program in some organized manner.

concave—Curved inward.

concentrated-arc lamp—A type of low-voltage arc lamp having nonvaporizing electrodes sealed in an atmosphere of inert gas and producing a small, brilliant, incandescent cathode spot.

concentration gradient—A difference in carrier concentration (holes or free electrons) from point to point in a semiconductor.

concentrator—1. A device that feeds the signals from several data terminals into a single transmission line for input to a computer, or vice versa. 2. An analog or digital buffer switch used to reduce the required number of trunks. 3. A device for combining many low-speed data lines into one high-speed data line. 4. A device that uses hardware and software to perform computer communication functions. (A term first applied to telephone-switching systems that permitted greater economy in use of facilities by combining many phone circuits into one.) 5. An electronic device that interfaces in a store-and-forward mode with multiple low-speed communication lines at a message level and then retransmits those messages to a processing site via one or more high-speed communication lines.

concentric cable—*See* coaxial line.

concentric groove—*See* locked groove.

concentric-lay cable—1. A concentric-lay conductor. 2. A multiple-conductor cable composed of a central core surrounded by one or more layers of helically laid insulated conductors.

concentric-lay conductor—A conductor composed of a central core surrounded by one or more layers of helically laid wires. In the most common type of concentric-lay conductor, all wires are of the same size, and the central core is a signal wire.

concentric line—*See* coaxial line.

concentric strand—A strand that consists of a central wire or core surrounded by one or more layers of spirally laid wires. Each layer after the first has six more strands than the preceding layer and is applied in a direction opposite to that of the layer under it.

concentric stranding—A method of stranding wire in which the final wire is built up in layers such that the inner diameter of a succeeding layer always equals the outer diameter of the underlying layer.

concentric-wound coil—A coil with two or more insulated windings that are wound one over the other.

concurrent processing—The ability of a computer to work on more than one program at the same time.

condensation soldering—The immersion of a part to be reflow soldered into a reservoir of hot saturated vapor. As the vapor condenses on the part, the latent heat of vaporizaton is released to heat the part.

condensed mercury temperature—The temperature of a mercury-vapor tube, measured on the outside of the tube envelope, in the region where the mercury is condensing in a glass tube or at a designated point on a metal tube.

condenser—Obsolete term for capacitor in electronics. A device in an ignition circuit that is connected across the contact points to reduce arcing by providing a storage for electricity as points open. Note: The term condenser is still used in the automotive field.

condenser antenna—*See* capacitor antenna.

condenser microphone—*See* electrostatic microphone.

condenser speaker—*See* electrostatic speaker.

condenser tissue—Kraft paper of 0.002 inch (51 μm) or less nominal thickness used in the manufacture of capacitors with paper or paper/film dielectrics.

conditional—1. In a computer, subject to the result of a comparison made during computation. 2. Subject to human intervention.

conditional breakpoint instruction—A conditional jump instruction that causes a computer to stop if a specified switch is set. The routine then may be allowed to proceed as coded, or a jump may be forced.

conditional jump—Also called conditional transfer of control. 1. An instruction causing a program transfer to an instruction location other than the next sequential instruction only if a specific condition tested by the instruction is satisfied. If the condition is not satisfied, the next sequential instruction in the program line is executed. 2. A computer program transfer that occurs only when the instruction that specifies it is executed and other specified conditions are satisfied.

conditional statement—Also called IF statement. A statement that causes the computer to check something and use that as a basis for choosing among alternative courses of action. Same as a branch.

conditional transfer—A program instruction that causes central control either to process the next instruction in sequence or to jump to some other indicated instruction, depending on the results of some previous operation.

conditional transfer of control—*See* conditional jump.

condition code—In a computer, a limited group of program conditions, such as carry, borrow, overflow, etc., that are pertinent to the execution of instructions. The codes are contained in a condition codes register.

conditioned line—A telephone circuit that has had its frequency response and/or delay characteristics optimized.

conditioning—1. Equipment modifications or adjustments required to provide matching of transmission levels and impedances or to provide equalization between facilities. 2. The addition of equipment to a leased voice-grade channel to provide minimum line characteristics necessary for data transmission. 3. Time-limited exposure of a test specimen to a specified environment(s) prior to testing. 4. Applying electronic filtering elements to a communication line to improve its ability to support higher transmission data rates. *See also* equalization.

condition queue—A queue, declared in the monitor, that lines up blocked processes.

Condor—A cw navigational system, similar to Benito, that automatically measures bearing and distance from a single ground station and displays them in a cathode-ray indicator. The distance is determined by phase comparison, and the bearing by automatic direction finding.

conductance—Symbolized by G or g. 1. In an element, device, branch, network, or system, the physical property that is the factor by which the square of an instantaneous voltage must be multiplied to give the corresponding energy lost by dissipation as heat or other permanent radiation, or by loss of electromagnetic energy from the circuit. 2. The real part of admittance. 3. Reciprocal of resistance; measured in siemens. It is the ratio of current through a material to the potential difference at its ends. 4. The component of sinusoidal current in phase with the terminal voltage of a circuit, divided by that voltage

conducted heat—Thermal energy transferred by thermal conduction.

conducted interference—Any unwanted electrical signal conducted on the power lines supplying the equipment under test, or on lines supplying other equipment to which the one under test is connected.

conducted signals—Electromagnetic or acoustic signals propagated along wire lines or other conductors.

conducted spurious transmitter output—A spurious output of a radio transmitter that is conducted over a tangible transmission path such as a power line, control circuit, radio-frequency line, waveguide, etc.

conductimeter—*See* conductivity meter.

conduction—The transmission of heat or electricity through or by means of a conductor.

conduction band—1. A partially filled energy band in which electrons can move freely, allowing the material to carry an electric current (with electrons as the charge carriers). 2. The band of energy levels occupied by a valence electron when it is liberated from an atom. Electrical conduction in a semiconductor crystal takes place through the transport of electrons in the conduction band.

conduction current—The power flow parallel to the direction of propagation, expressed in siemens per meter.

conduction-current modulation—1. Periodic variations in the conduction current passing a point in a microwave tube. 2. The process of producing such variations.

conduction electrons—The electrons that are free to move under the influence of an electric field in the conduction band of a solid.

conduction error—The error in a temperature transducer due to heat conduction between the sensing element and the mounting to the transducer.

conduction field—Energy that surrounds a conductor when an electric current is passed through the conductor, and that, because of the difference in phase between the electrical field and magnetic field set up in the conductor, cannot be detached from the conductor.

conductive adhesive—An adhesive material that has metal powder added to increase electrical conductivity.

conductive epoxy—An epoxy material (polymer resin) that has been made conductive by the addition of a metal powder, usually gold or silver.

conductive gasket—A special highly resilient gasket used to reduce rf leakage in shielding that has one or more access openings.

conductive level detector—A device with single or multiple probes. A change in level completes an electrical circuit between the container and/or probes.

conductive material—A material in which a relatively large conduction current flows when a potential is applied between any two points on or in a body constructed from the material. Metals and strong electrolytes are examples of conductors.

conductive pattern—1. The arrangement or design of the conductive lines on a printed circuit board. 2. The configuration or design of the conductive material on the base material. Includes conductors, lands, and through connections when these connections are an integral part of the manufacturing process.

conductive pattern-to-board outline—The location of the printed pattern relative to the overall outline dimensions of the printed circuit board.

conductive plastic potentiometer—A potentiometer in which the resistive element consists of a blend of resin (epoxy, polyester, etc.) and processed carbon powder applied to a plastic or ceramic substrate.

conductivity—1. The conductance between opposite faces of a unit cube of material. The volume conductivity of a material is the reciprocal of the volume resistivity. 2. The ability of a material to conduct electric current. It is expressed in terms of the current per unit of applied voltage. It is the reciprocal of resistivity. 3. The ability to conduct or transmit heat or electricity. 4. The ability of a material to allow electrons to flow, measured by the current per unit of voltage applied. It is also the reciprocal of resistivity. 5. Synonym for conductance. 6. The parameter of a material that indicates the extent to which it permits a net electrical current; normally measured in terms of the conductance in reciprocal ohms (siemens) between opposite faces of a cube of the material measuring 1 centimeter on each side. The conductivity of a material is the reciprocal of its resistivity. 6. Property of a material to allow electrical current to flow with very little loss. For natural surfaces, conductivity in general is increased with increased moisture content.

conductivity meter—Also called conductimeter. An instrument that measures and/or records electrical conductivity.

conductivity modulation—1. The change in conductivity of a semiconductor as the charge-carrier density is varied. 2. The process whereby the effective electrical conductivity of a semiconductor region is modified by the injection of excess carriers. Thus, excess majority carriers injected into a lightly doped region can cause the effective conductivity to be increased simply by providing further carriers for current conduction. Conversely, excess minority carriers injected into a heavily doped region can cause the effective conductivity to be reduced by increasing the incidence of recombination and, hence, reducing the number of carriers available for conduction.

conductivity-modulation transistor—A transistor in which the active properties are derived from minority-carrier modulation of the bulk of resistivity of the semiconductor.

conductor—1. A bare or insulated wire or combination of wires not insulated from one another, suitable for carrying an electric current. 2. A body of conductive material so constructed that it will serve as a carrier of electric current. 3. A material (usually a metal) that conducts electricity through the transfer of orbital electrons. 4. A material, such as copper or aluminum, that offers low resistance or opposition to the flow of electric current. 5. A medium for transmitting electrical current. A conductor usually consists of copper, aluminum, steel, silver, or other material. 6. A solid, liquid, or gas that offers little opposition to the continuous flow of electric current. 7. Any material whose valence energy band is only partially filled with electrons, so that empty levels are immediately available for a net electron movement. Such materials conduct electricity readily, even at extremely low temperatures. 8. A substance or body that allows a current of electricity to pass continuously and easily through it. A member of a class of materials that conduct electricity easily, i.e., have a low resistivity (10^{-4} ohm/cm). 9. A single conductive path in a conductive pattern.

conductor side—The side of a single-sided printed board containing the conductive pattern.

conductor spacing—The distance between adjacent edges (not centerline to centerline) of isolated conductive patterns in a conductor layer of a printed circuit.

conductor-to-hole spacing—The distance between a conductor edge and the edge of a component hole.

conduit—1. A tubular raceway designed for holding wires or cables designed and used expressly for this purpose. It may be a solid or flexible tube in which insulated electrical wires are run. 2. Metal sleeve through which electrical wires pass.

conduit wiring—Wiring carried in conduits and conduit fittings.

cone—The diaphragm that sets the air in motion to create a sound wave in a direct-radiator loudspeaker. Usually it is conical in shape.

cone breakup—The inability of a speaker cone to work as a piston at high frequencies, the effect being that the cone is not under the complete control of the voice coil, certain parts of it moving in opposition to other parts like a rippled rope. Responsible for uneven frequency response.

cone of nulls—A conical surface formed by directions of negligible radiation.

cone of silence—An inverted cone-shaped space directly over the aerial towers of some radio beacons. Within the cone, signals cannot be heard or will be greatly reduced in volume.

conference call—A telephone call that interconnects three or more telephones and permits all parties to converse at random.

confetti—Flecks or streaks of color caused by tube noise in the chrominance amplifier. Because of its colors, confetti is much more noticeable than snow in a black-and-white picture. The chrominance amplifier is therefore cut off during a monochrome program.

confidence—1. The likelihood, expressed as a percentage, that a measurement or statement is true. 2. The degree of assurance that the stated failure rate has not been exceeded.

confidence factor—The percentage figure expressing confidence level.

confidence interval—A range of values believed to include, with a preassigned degree of confidence, the true characteristic of the lot.

confidence level—1. The probability (expressed as a percentage) that a given assertion is true or that it lies within certain limits calculated from the data. 2. A degree of certainty.

confidence limits—Extremes of a confidence interval within which there is a designated chance that the true value is included.

configuration—1. The relative arrangement of parts (or components) in a circuit. 2. A listing of the names and/or serial numbers of the assemblies that make up an equipment. 3. The hardware and/or software making up a system. 4. Combination of computer and peripheral devices at a single installation. 5. A general-purpose computer term that can refer to the way a computer is set up. It is also used to describe the total combination of hardware components that make up a computer system and the software settings that allow various hardware components of a computer system to communicate with one another.

configuration file—A file that contains information on the way a system is set up.

configure—The act of changing software or hardware actions by changing the settings in a computer.

confocal resonator—A wavemeter for millimeter wavelengths. It consists of two spherical mirrors that face each other; a change in the spacing between the mirrors affects the propagation of electromagnetic energy between them, making possible direct measurement of free-space wavelengths.

conformal coating—1. A thin nonconductive coating, either plastic or inorganic, applied to a circuit for environmental and/or mechanical protection. 2. A protective coating applied to completed printed circuit boards that conforms to the shape of the components and provides complete electrical as well as environmental insulation.

conformance error—The deviation of a calibration curve from a specified curve line.

confusion jamming—An electronic countermeasure by means of which a radar may detect a target, but the radar operator is denied accurate data regarding range, azimuth, and velocity of the target. This result is accomplished through amplification and retransmission of an incident radar signal with distortion to create a false echo. Also called deception jamming.

confusion reflector—A device that reflects electromagnetic radiation to create echoes for purposes of causing confusion of radar, guided missiles, and proximity fuses.

congestion—A condition in which the number of calls arriving at the various inputs of a communications network are too many for the network to handle at once and are subject to delay or loss. (The concept applies in an analogous way to any system in which arriving traffic can exceed the number of servers.)

conical horn—A horn whose cross-sectional area increases as the square of the axial length.

conical scanning—A form of scanning in which the beam of a radar unit describes a cone, the axis of which coincides with that of the reflector.

conjugate—Either of a pair of complex numbers that are mutually related in that their real parts are identical and the imaginary part of one is the negative of the imaginary part of the other, that is, if $a = x + iy$, then $a = x - iy$ is its conjugate.

conjugate branches—Any two branches of a network in which a driving force impressed on one branch does not produce a response in the other.

conjugate bridge—A bridge in which the detector circuit and the supply circuits are interchanged, compared with a normal bridge.

conjugate impedance—An impedance whose value is the conjugate of a given impedance. For an impedance associated with an electric network, the conjugate is an impedance with the same resistance component as the original and a negative reactive component.

conjugate matching—A condition of source- and loading-impedance matching in which the source impedance and the load impedance have equal resistive parts and equal reactance values with opposite signs. This results in maximum power transfer.

connected—A network is connected if, between every pair of nodes of the network, there exists at least one path composed of branches of the network.

connecting block—A cable-termination block in which access to circuit connections is available.

connection—1. The attachment of two or more component parts so that conduction can take place between them. 2. The point of such attachment.

connection diagram—1. A diagram showing the electrical connections between the parts that make up an apparatus. 2. A pattern illustrating the connections needed to place an electronic system in operation when such a system includes one or more assemblies, power supplies, and devices being controlled.

connector—1. A coupling device that provides an electrical and/or mechanical junction between two cables, or between a cable and a chassis or enclosure. 2. A device that provides rapid connection and disconnection of electrical cable and wire terminations. 3. A plug or receptacle that can be easily joined to or separated from its mate. Multiple-contact connectors join two or more conductors with others in one mechanical assembly. 4. A device consisting of a mating plug and receptacle. Various types of connectors include DIP, card-edge, two-piece, hermaphroditic, and wire-wrapping configurations. 5. Devices designed to provide separable through connections in cable-to-cable, cable-to-chassis, or rack and panel applications. 6. A device that holds two parts of a circuit together so that they make electrical contact.

connector assembly—The combination of a mated plug and receptacle.

connector discontinuity—An ohmic change in contact resistance.

connector flange—A projection that extends from or around the periphery of a connector and incorporates provisions for mounting the connector to a panel.

connector receptacle—1. An electrical fitting with contacts constructed to be electrically connected to a cable, coaxial line, cord, or wire to join with another electrical connector mounted on a bulkhead, wall, chassis, or panel.

connect time—1. The total time required for establishing a connection between two points. 2. In a computer-based data communications assembly, the switching time required to set up a connection between two terminal points.

conoscope—An instrument for determining the optical axis of a quartz crystal.

consequent poles—Additional magnetic poles present at other than the ends of a magnetic material.

consol—*See* sonne.

console—1. A cabinet for a radio or television receiver that stands on the floor rather than on a table. 2. Main operating unit in which indicators and general controls of a radar or electronic group are installed. 3. A part of a computer that may be used for manual control of the machine. The computer operator's control panel or terminal. 4. An array of controls and indicators for the monitoring and control of a particular sequence of actions, as in the checkout of a rocket, a countdown, or a launch.

console operator—A person who monitors and controls an electronic computer by means of a central control unit or console.

consonance—Electrical or acoustical resonance between bodies or circuits not connected directly together.

constant—1. An unvarying or fixed value or data item. 2. Any number not expected to change.

constant-amplitude recording—In disc recording, a relationship between the modulations in the groove and the electrical signals making them so that the width of the groove (the excursions of the cutting stylus) is proportional to the amplitude, or power, of the signal. In playback, a similar relation between the record and the motion of the stylus so that the cartridge produces equal voltage regardless of frequency. Crystal and ceramic pickups have a constant amplitude characteristic on playback.

constant current—1. A current that does not undergo a change greater than the required precision of the measurement when the impedance of the generator is halved. 2. Having to do with a type of power-supply operation in which the output current remains at a preset value (within specified limits) while the load resistance varies, resulting in an output-voltage variation within the voltage range of the power supply.

constant-current characteristic—The relationship between the voltages of two electrodes, the current to one of them as well as all other voltages being maintained constant.

constant-current/constant-voltage supply—A power supply that behaves as a constant-voltage source for relatively large values of load resistance and as a constant-current source for relatively small values of load resistance. The crossover point between these two modes of operation occurs when the value of the critical load resistance equals the value of the supply voltage setting divided by the supply current setting.

constant-current modulation—Also called Heising modulation. A system of amplitude modulation in which the output circuits of the signal amplifier and carrier-wave generator or amplifier are directly and conductively coupled by a common inductor. The inductor has an ideally infinite impedance to the signal frequencies and therefore maintains the common plate-supply current of the two devices constant. The signal-frequency voltage thus appearing across the common inductor also modulates the plate supply to the carrier generator or amplifier, with corresponding modulation of the carrier output.

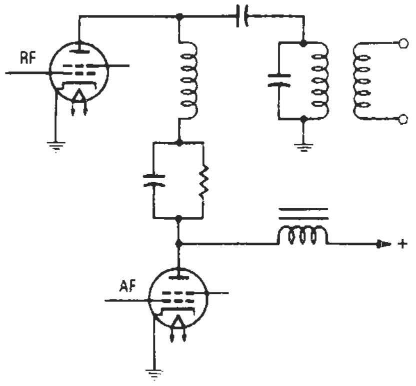

Constant-current modulation.

constant-current power supply—A regulated power supply that acts to keep its output current constant in spite of changes in load, line, or temperature. Thus, for a change in load resistance, the output current remains constant to a first approximation, while the output voltage changes by whatever amount necessary to accomplish this.

constant-current transformer—1. A transformer that automatically maintains a constant current in its secondary circuit under varying conditions of load impedance when supplied from a constant-potential source. 2. A transformer that supplies a constant current to a varying load. Regulation is accomplished by either varying the separation between primary and secondary windings or by the use of a resonant network.

constant-delay discriminator—*See* pulse demoder.

constant K-filter—An image-parameter filter comprising a tandem connection of a number of identical prototype L-section filters. Each adjacent pair of L-sections together forms either a T- or π-network. The product of the series and shunt impedances is a constant that is independent of frequency.

constant K-network—A ladder network in which the product of its series and shunt impedances is independent of frequency within the range of interest.

constant-luminance transmission—A type of transmission in which the transmission primaries are a luminance primary and two chrominance primaries.

constant-power-dissipation line—A line superimposed on the output static characteristic curves and representing the points of collector voltage and current, the product of which represents the maximum collector power rating of a particular transistor.

constant-ratio code—A code in which the combinations that represent all characters contain a fixed ratio of ones to zeros.

constant-resistance network—A network that will reflect a constant resistance to the output circuit of the driving amplifier when terminated in a resistive load. Loudspeakers do not reflect a constant impedance; therefore, an amplifier does not see a constant resistance. This disadvantage may be somewhat compensated for by the use of negative feedback by the amplifier.

constant-ringing drop—Abbreviated CRD. A relay that when activated even momentarily will remain in an alarm condition until reset. A key is often required to reset the relay and turn off the alarm.

constant-velocity recording—In disc recording, a relationship between the wiggles in the groove and the electrical signals making them, whereby the frequency of the signal determines the degree of excursion of the cutter. In playback, a similar relation between the recorded wiggles and the motion of the stylus so that the cartridge produces voltages that vary in strength, or amplitude, as the frequency in the groove varies. Magnetic cartridges have a constant velocity characteristic and must be equalized by special networks during playback.

constant voltage—1. Voltage that does not undergo a change greater than the required precision of the measurement when the impedance of the generator is doubled. 2. Having to do with a type of power-supply operation in which the output voltage remains at a preset value (within specified limits) while the output current is varied within the range of the power supply.

constant-voltage charge—A charge method for a rechargeable battery (storage battery) in which voltage at the battery terminals is held at a constant value throughout the charge cycle.

constant-voltage charger—A battery charger that maintains a constant output voltage so that the charging current tapers off as the battery becomes charged. When fully charged, the battery and the charger supply only minor variations in the load current.

constant-voltage/constant-current cross over—Behavior of a power supply in which there is automatic conversion from voltage stabilization to current stabilization (and vice versa) when the output current reaches a preset value.

constant-voltage/constant-current (cv/cc) output characteristic—A regulated power supply that acts as a constant-voltage source for comparatively large load resistances and as a constant-current source for comparatively small load resistances.

constant-voltage power supply—1. A regulated power supply that acts to keep its output voltage constant in spite of changes in load, line, or temperature. Thus, for a change in load resistance, the output current changes by whatever amount necessary to accomplish this. 2. A power supply capable of maintaining a fixed voltage across a variable load resistance and over a defined input voltage and frequency change. The output is automatically controlled to maintain constant the product of output current times load resistance.

constant-voltage transformer—A transformer delivering a fixed predetermined voltage over a limited range of input-voltage variations (e.g., 95–125 volts).

constructive synthesis—Synthetic voice generation system that builds words from a prescribed set of linguistic or phonetic sound segments, such as phonemes. Each language has its own set of such sound segments.

contact—1. One of the current-carrying parts of a relay, switch, or connector that is engaged or disengaged to open or close the associated electrical circuits. 2. To join two conductors or conducting objects in order to provide a complete path for current flow. 3. The juncture point to provide the complete path. 4. The conducting part of a connector that acts with another such part to complete or break a circuit; contacts provide a separable through connection in a cable-to-cable, cable-to-box, or box-to-box situation. 5. The disc or bar of precious metal on a key, jack, or relay spring that touches another similar contact, thus making a temporary, low-resistance electrical connection through which current can flow.

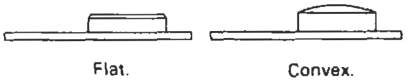

Contact, 1.

contact-actuation time—The time required for any specified contact on a relay to function. When not otherwise specified, it is the initial actuation time. For some purposes, it is preferable to use either the final or effective actuation time.

contact arc—The electrical (current) discharge that occurs between mating contacts when the circuit is being disestablished.

contact area—The common area between two conductors or a conductor and a connector through which the flow of electricity takes place.

contact arrangement—1. The combination of contact forms that make up the entire relay-switching structure. 2. The number, spacing, and positioning of contacts in a connector.

contact bounce—1. The uncontrolled making and breaking of contact when relay contacts are closed. 2. Internally caused intermittent and undesired opening of closed contacts of a relay, caused by one or more of the following: (*a*) Impingement of mating contacts. (*b*) Impact of the armature against the coil core on pickup or against the backstop on dropout. (*c*) Momentary hesitation, or reversal, of the armature motion during the pickup or dropout stroke.

contact bounce time—The time interval from initial actuation of a relay contact to the end of bounce brought about during pickup or dropout or from external causes.

contact chatter—1. The undesired vibration of mating contacts during which there may or may not be actual physical contact opening. If there is no actual opening but only a change in resistance, it is referred to as dynamic resistance and appears as "grass" on the screen of an oscilloscope having adequate sensitivity and resolution. 2. A sustained rapid physical opening and closing of contact points caused by mechanical vibrations.

contact combination—1. The total assembly of contacts on a relay. 2. Sometimes used for contact form.

contact current—The peak current through the contacts of a contact-triggered system at the instant the contacts open.

contact device—A device that when actuated opens or closes a set of electrical contacts; a switch or relay.

contact emf—A small voltage established whenever two conductors of different materials are brought into contact.

contact follow—The displacement of a stated point on the contact-actuating member following the initial closure of a contact.

contact force—1. The amount of force exerted by one of a pair of closed contacts on the other. 2. The force exerted by the moving mercury on a stationary contact or electrode in a mercury switch.

contact gap—Also called contact separation. The distance between a pair of mating relay contacts when they are open.

contact length—The length of travel of one contact while touching another contact during the assembly or disassembly of a connector.

contactless vibrating bell—A vibrating bell whose continuous operation depends on application of an alternating current without circuit-interrupting contacts, such as those used in vibrating bells operated by direct current.

contact load—The electrical power demands encountered by a contact set in any particular application.

contact microphone—1. A microphone designed to pick up mechanical vibrations directly from the sound source and convert them into corresponding electrical currents or voltages. 2. A microphone designed for attachment directly to a surface of a protected area or object; usually used to detect surface vibrations.

contact miss—Failure of a contact mating pair to establish the intended circuit electrically. This may be a contact resistance in excess of a specified maximum value.

contact modulator—Also called electromechanical chopper. A switch used to produce modified square waves having the same frequency as, and a definite phase relationship to, a driving sine wave.

contact noise—The random fluctuation of voltage across a junction through which current is flowing from one solid to another.

contactor—1. A device for the purpose of repeatedly establishing or interrupting an electric power circuit. 2. A heavy-duty relay used to control electrical circuits.

contactor alarm—A signal calling attention to lowered pressure in a cable gas-pressure system.

contact potential—Also called Volta effect. 1. The difference of potential that exists when two dissimilar, uncharged metals are placed in contact. One becomes positively charged and the other negatively charged, the amount of potential depending on the nature of the two metals. 2. The potential difference between the contacting surfaces of two metals that have different work functions.

contact-potential difference—The difference between the work functions of two materials, divided by the electronic charge generated by them.

contact pressure—The amount of pressure holding a set of contacts together.

contact printing—A method of screen printing in which the screen is almost in contact with the substrate. Used for printing with metal mask.

contact rating—The electrical power-handling capability of relay or switch contacts under specified environmental conditions and for a prescribed number of operations.

contact rectifier—A rectifier consisting of two different solids in contact. Rectification is due to the greater conductivity across the contact in one direction than in the other.

contact resistance—1. Total electrical resistance of a contact system, such as the resistance of a relay or a switch measured at the terminals. Usually this resistance is only a fraction of an ohm. 2. The ohmic resistance between the contacts of a switch connector or relay. It may an be extremely small value—typically in the milliohm range. Contact resistance is normally measured from terminal to terminal. 3. The resistance between the wiper terminal and the resistive element of a potentiometer. 4. Electrical resistance of pin and socket contacts when assembled in a connector under typical service use. Electrical resistance of each pair of mated pin and socket contacts in the connector assembly is determined by measuring from the pin to the extreme terminal end of the socket (excluding both crimps) when carrying a specified test current. (Overall contact resistance includes wire-to-wire measurement.) 5. In electronic elements, such as capacitors or resistors, the apparent resistance between the terminating electrode and the body of the device.

contact resistance variation—Abbreviated crv. The maximum instantaneous charge in contact resistance that results from moving the wiper of a potentiometer from one position to another. It is expressed as a percentage of the potentiometer's total resistance.

contact retainer—A device used to retain a contact in an inset or body; it may be either on the contact or in the insert.

contact retention—The minimum axial load a contact in a connector can withstand in either direction while remaining firmly fixed in its normal position in the insert.

contacts—In a relay, the current-carrying parts that engage or disengage to open or close electrical circuits.

contact separation—The maximum distance between the stationary and movable contacts when the circuit is broken.

contact spring—1. A current-carrying spring to which the contacts are fastened. 2. A non-current-carrying spring that positions and tensions a contact-carrying member.

contact symbology diagram—Commonly referred to as a ladder diagram, it expresses the user-programmed logic of the controller in relay-equivalent symbology.

contact wetting—The coating of a contact surface with an adherent film of mercury.

contact wipe—The distance of travel (electrical engagement) of one contact during its engagement with or separation from another or during mating or unmating of the connector halves.

contaminated—Made radioactive by addition of a radioactive material.

content-addressable memory—Memory in which information is retrieved by specifying the data rather than the address at which the data is stored.

content-addressed storage—*See* associative storage.

content indicator—A display device that indicates the content in a computer, and the program or mode in use.

contention—1. A condition that occurs on a multidrop communication channel when two or more locations attempt to transmit simultaneously. 2. Unregulated bidding for a line by multiple users. 3. Competition for use of the same communication facilities; a line-control method in which terminals request or bid to transmit.

contents—The information stored in any part of the computer memory.

contiguous allocation—An allocation method that assigns adjacent sectors to a file.

Continental code—*See* International Morse code.

continuity—1. A continuous path for the flow of current in an electric circuit. 2. In radio broadcasting, the prepared copy from which the spoken material is presented.

continuity check—A test performed on a length of finished wire or cable to determine if electrical current flows continuously throughout the length. Conductors may also be checked against each other to ascertain that there are no shorts.

continuity test—An electrical test for determining whether a connection is broken.

continuity writer—In radio broadcasting, the person who writes the copy from which the spoken material is presented.

continuous carrier—A carrier over which transmission of information is accomplished by means which do not interrupt the carrier.

continuous commercial service—*See* CCS.

continuous-current rating—The designated rms alternating or direct current that a device can carry continuously under specified conditions.

continuous data—Any set of data whose information content can be ascertained continuously in time.

continuous duty—The ability of a device or a control to operate continuously with no off or rest periods.

continuous duty rating—The rating applied to equipment if operated for an indefinite length of time.

continuous load—A load in which the maximum current is expected to continue for three hours or more.

continuously loaded cable—A submarine cable in which the conductors are continuously loaded.

continuous output power—The maximum power (in watts) that an amplifier will deliver from each channel (with all channels operating) without exceeding its rated harmonic distortion. Measured with a 1-kHz signal. Power ratings should include harmonic distortion and the load impedance (4, 8, or 16 ohms). For example, continuous output power 40 W/40 W (at less than 1 percent harmonic distortion, into 8-ohm load).

continuous power—The power an amplifier is capable of delivering for at least 30 seconds with a sine-wave signal.

continuous power output—*See* rated power output.

continuous rating—The rating that defines the load that can be carried for an indefinite length of time.

continuous recorder—A recorder that makes its record on a continuous sheet or web rather than on individual sheets.

continuous scan thermograph—Equipment for presenting a continuous scan image of the thermal pattern (thermogram) of a patient or an object on a cathode-ray tube.

continuous spectrum—The spectrum that exhibits no structure and appears to represent a continuous variation of wavelength from one to the other.

continuous variable—A variable that may assume any value within a defined range.

continuous-wave radar—A system in which a transmitter sends out a continuous flow of radio energy to the target, which reradiates (scatters) the energy intercepted and returns a small fraction to a receiving antenna.

continuous waves—Abbreviated cw. Electromagnetic sine waves generated as a continuous train of identical oscillations. They can be interrupted according to a code, or modulated in amplitude, frequency, or phase in order to convey information.

continuous-wave tracking system—A tracking system that operates by keeping a continuous radio beam on a target and determining its behavior from changes in the antenna necessary to keep the beam on the target.

contour control system—In automatic control of machine tools, a system in which the cutting path of a tool is controlled along two or more axes.

contourograph—A device in which a cathode-ray oscilloscope is used to produce images that have a three-dimensional appearance.

contrahelical—In the wire and cable industry, the term is used to mean the direction of a layer with respect to the previous layer. Thus, it would mean a layer spiraling in an opposite direction from the preceding layer within a wire or cable.

CONTRAN—A computer-programming language in which instructions are written at a compiler level, thereby eliminating the need for translation by a compiling routine.

contrast—1. The actual difference in density between the highlights and the shadows. Contrast is not concerned with the magnitude of density, but only with the difference in densities. 2. Amplitude ratio between picture white and picture black. 3. Ratio between the maximum and minimum brightness values in a picture. 4. In optical character recognition, the differences between the color or shading of the printed material on a document and the background on which it is printed. 5. A noticeable difference in color, brightness, or other characteristics in a side-by-side comparison. 6. The difference in tone between the lightest and darkest areas of a photographic print or television image. 7. The ratio between the brightest and darkest parts of a picture.

contrast control—1. A method of adjusting the contrast in a television picture by changing the amplitude of the video signal. 2. With respect to television, a potentiometer that allows variation of the intensity of the different elements of an image and that can be used to accentuate the highlights and shadows in an image. In a color television system, saturation and hue may also be controlled.

contrast range—The ratio between the whitest and blackest portions of a television image.

contrast ratio—1. Ratio of the maximum to the minimum luminance values in a television picture or a portion thereof. 2. The ratio of total display element of luminance to the background luminance.

control—Also called a control circuit. 1. In a digital computer, those parts that carry out the instructions in proper sequence, interpret each instruction, and apply the proper signals to the arithmetic unit and other parts in accordance with the interpretation. 2. Sometimes called a manual control. In any mechanism, one or more components responsible for interpreting and carrying out manually initiated directions. 3. In some business applications, a mathematical check. 4. In electronics, a potentiometer or variable resistor. 5. In an alarm system, any mechanism that sequences the interrogation of protected site units, resets latched alarms, and performs similar functions.

control ampere turns—The magnitude and polarity of the control magnetomotive force required for operation of a magnetic amplifier at a specified output.

control amplifier—*See* preamplifier.

control block—A storage area through which information of a particular type required for control of the operating system is communicated among the parts of the system.

control cable—A multiconductor cable made for operation in control or signal circuits, usually flexible, relatively small in size, and with relatively small current ratings.

control card—In computer programming, a card containing input data or parameters for a specific application of a general routine.

control center—*See* preamplifier.

control character—1. A character whose occurrence in a particular context initiates, modifies, or halts operation. 2. An element of a character set that may produce some action in a device other than a printed or displayed character. A character may become a control character in some systems by being preceded by a special character or set of characters. 3. Within a code set, a character intended to initiate, modify, or stop a control function.

control characteristic—1. A plot of the load current of a magnetic amplifier as a function of the control ampere turns for various loads and at the rated supply voltage and frequency. 2. The relationship between the critical grid voltage and the anode voltage of a tube.

control circuit—*See* control.

control circuits—In a digital computer, the circuits that carry out the instruction in proper sequence, interpret each instruction, and apply the proper commands to the arithmetic element and other circuits in accordance with the interpretation.

control-circuit transformer—A voltage transformer utilized to supply a voltage suitable for the operation of control devices.

control-circuit voltage—The voltage provided for the operation of shunt-coil magnetic devices.

control compartment—A space within the base, frame, or column of a machine used for mounting the control panel.

control counter—In a computer, a device that records the storage location of the instruction word to be operated on following the instruction word in current use.

control current—Current that occurs in the control circuit when control voltage is applied.

control data—In a computer, one or more items of data used to control the identification, selection, execution, or modification of another routine, record file, operation, data value, etc.

CONTROL DATA or Control Data—A trademark and service mark of Control Data Corporation in respect to data processing equipment and related services.

control electrode—An electrode on which a voltage is impressed to vary the current flowing between other electrodes.

control field—In a sequence of similar items of computer information, a constant location where control information is placed.

control-flow machine—A parallel-processing architecture with a single central sequence of instruction, carried out by many processors.

control grid—The electrode of a vacuum tube, other than a diode, upon which a signal voltage is impressed to regulate the plate current.

control-grid bias—The average direct-current voltage between the control grid and cathode of a vacuum tube.

control-grid plate transconductance—The ratio of the amplification factor of a vacuum tube to its plate resistance, combining the effect of both into one term.

controlled avalanche—A predictable, nondestructive avalanche characteristic designed into a semiconductor device as protection against reverse transients that exceed its ratings.

controlled-avalanche device—A semiconductor device that has very specific maximum and minimum avalanche-voltage characteristics and is also able to operate and absorb momentary power surges in this avalanche region indefinitely without damage.

controlled-avalanche silicon rectifier—A silicon diode manufactured with characteristics such that, when operating, it is not damaged by transient voltage peaks.

controlled-carrier modulation—Also called variable-carrier or floating-carrier modulation. A modulation system in which the carrier is amplitude modulated by the signal frequencies, and also in accordance with the envelope signal, so that the modulation factor remains constant regardless of the amplitude of the signal.

controlled-impedance cable—Package of two or more insulated conductors in which impedance measurements between respective conductors are kept essentially constant throughout the entire length.

controlled rectifier—1. A rectifier employing grid-controlled devices such as thyratrons or ignitrons to regulate its own output current. 2. Also called an SCR (silicon-controlled rectifier). A four-layer pnpn semiconductor that functions like a grid-controlled thyratron.

controller—1. An instrument that holds a process or condition at a desired level or status as determined by comparison of the actual value with the desired value. 2. A device or group of devices that serves to govern, in some predetermined manner, the electric power delivered to the apparatus to which it is connected. 3. A hardware interface that accepts instructions from a computer and reformats them to program an instrument or peripheral.

controller function—Regulation, acceleration, deceleration, starting, stopping, reversing, or protection of devices connected to an electric controller.

control-line timing—Clock signals between a modem and a communication-line controller unit.

control link—Apparatus for effecting remote control between a control point and a remotely controlled station.

control locus—A curve that shows the critical value of grid bias for a thyratron.

control operator—An amateur radio operator designated by the licensee of an amateur radio station to also be responsible for the emissions from that station.

control panel—A panel having a systematic arrangement of terminals used with removable wires to direct the operation of a computer or punched-card equipment.

control point—1. A point that may serve as a reference for all incremental commands. 2. The operating position of an amateur radio station where the control operator's function is performed.

control-power disconnecting device—A disconnective device, such as a knife switch, circuit breaker, or pullout fuse block, used for the purpose of connecting and disconnecting the source of control power to and from the control bus or equipment.

control program—A computer program that places another program and its environment in core memory in proper sequence and retains them there until it has finished operating.

control ratio—1. The ratio of the change in anode voltage to the corresponding change in critical grid voltage of a gas tube, with all other operating conditions maintained constant. 2. Also called programming coefficient. The required range in control resistance of a regulated power supply to produce a 1-volt change in output voltage. Expressed in ohms per volt.

control read-only memory—Abbreviated CROM. A major component in the control block of some microprocessors. It is a ROM that has been microprogrammed to decode control logic.

control rectifier—A silicon rectifier capable of switching or regulating the flow of a relatively large amount of power through the use of a very small electrical signal. These solid-state devices can take the place of mechanical and vacuum tube switches, relays, rheostats, variable transformers, and other devices used for switching or regulating electric power.

control register—Also called instruction register. In a digital computer, the register that stores the current instruction governing the operation of the computer for a cycle.

control section—*See* control unit.

control sequence—In a computer, the normal order of execution of instructions.

control station—Station licensed to conduct remote control of another amateur radio station.

control store—A memory circuit designed to hold the sequence of commands that determines operation of the sequential-state machine. Sometimes referred to as the microprogram store.

control tape—In a computer, a paper or plastic tape used to control the carriage operation of some printing output devices. Also called carriage tape.

control unit—1. That section of an automatic digital computer that directs the sequence of operations, interprets coded instructions, and sends the proper signals to the other computer circuits to carry out the instructions. Also called control section. 2. A preamplifier unit in an audio setup. Signals from audio sources, e.g., tuner or pickup microphone, are fed into it. Equalization (where necessary) is applied, then the signal is fed to the main amplifier. Volume and tone controls are usually incorporated, together with any necessary program-selection switch.

control variable—The plant inputs and outputs that a control system manipulates and measures to properly control them.

control voltage—Voltage applied to control input terminals of a relay.

control-voltage winding—The motor winding that is excited by a varying voltage at a time phase difference from the voltage applied to the fixed-voltage windings of a servomotor.

control winding—In a saturable reactor, the winding used for applying a controlling magnetomotive premagnetization force to the saturable-core material.

control wire and cable—Any wire that carries current to control a tube, device, relay, or to cause any event without actually carrying the energy controlled in the event.

convection—1. The motion in a fluid as a result of differences in density and the action of gravity. 2. The transfer of heat from a high-temperature region in a gas or a liquid as a result of movement of masses of the fluid. 3. A conveying, or transference, of heat or electricity by moving particles of matter.

convection cooling—A method of heat transfer that depends on the natural upward movement of the air warmed by the heat dissipated from the device being cooled.

convection current—The amount of time required for a charge in an electron stream to be transported through a given surface.

convection-current modulation—1. The time variation in the magnitude of the convection current passing through a surface. 2. The process of producing such a variation.

convenience receptacle—An assembly consisting of two or more stationary contacts mounted in a small insulating enclosure that has slots to permit blades on attachment plugs to enter and make contact with the circuit.

convention—A definite formatting method used in electronic diagrams to present the clearest picture of the circuit function. Some common conventions are as follows: (*a*) circuit signal flow from left to right, with inputs on the left and outputs on the right; (*b*) locating various circuit functional stages in the same sequence as the signal flow; (*c*) placing the highest voltage sources at the top of the sheet and the lowest at the bottom; and (*d*) showing auxiliary circuits that are included but are not a main part of the signal flow, such as oscillators and power supplies, on the lower half of the drawing.

convergence—1. The condition in which the electron beams of a multibeam cathode-ray tube intersect at a specified point. 2. Orientation of the three electron beams in a color TV picture tube so they pass through the same hole in a shadow mask at the same time. 3. In optics, the bending of light rays toward each other, as by a convex or positive lens.

convergence coil—One of the two coils associated with an electromagnet, used to obtain dynamic beam convergence in a color television receiver.

convergence control—A variable resistor in the high-voltage section of a color television receiver. It controls the voltage applied to the three-gun picture tube.

convergence electrode—An electrode whose electric field causes two or more electron beams to converge.

convergence magnet—A magnet assembly whose magnetic field causes two or more electron beams to converge.

convergence phase control—A variable resistor or inductance for adjusting the phase of the dynamic convergence voltage in a color TV receiver employing a three-gun picture tube.

convergence surface—The surface generated by the point at which two or more electron beams intersect during the scanning process in a multibeam cathode-ray tube.

conversation—An interactive exchange of information between two systems or systems users.

conversational mode—A mode of computer operation in which a user is in direct contact with a computer, and interaction is possible between human and machine without the user being conscious of any language or communications barrier.

conversational operation—A type of operation similar to the interactive mode, except that the computer user must wait until a question is posed by the computer before interacting.

conversational system—*See* interactive system.

conversion—1. The process of changing from one data-processing method or system to another. 2. The process of changing from one form of representation to another. 3. *See* encode, 2.

conversion efficiency—1. The ratio of ac output power to the dc input power to the electrodes of an electron tube. 2. The ratio of the output voltage of a converter at one frequency to the input voltage at some other frequency. 3. In a rectifier, the ratio of dc output power to ac input power. 4. The ratio of maximum available luminous or radiant flux output to total input power. 5. Of a solar cell, the ratio of the electrical power obtained from the cell to the radiant power falling on the cell.

conversion gain—1. The ratio of the intermediate-frequency output voltage to the input-signal voltage of the first detector of a superheterodyne receiver. 2. The ratio of the available intermediate-frequency power output of a converter or mixer to the available radio-frequency power input.

conversion loss—The ratio of available input power to available output power under specified test conditions.

conversion rate—The speed at which an analog-to-digital converter or digital-to-analog converter can make repetitive data conversions, or the number of conversions performed per second. It is affected by propagation delay in counting circuits, ladder switches, and comparators; ladder *RC* and amplifier settling times; and amplifier and comparator slew rates and integrating time of dual-slope converters. Conversion rate is specified as a number of conversions per second, or conversion time is specified as a number of microseconds to complete one conversion (including the effects of settling time). Sometimes conversion rate is specified for less than full resolution, thus showing a misleading (high) rate.

conversion speed—The measure of how long it takes an analog-to-digital converter to arrive at the proper output code. It is the time delay between the edge of the pulse that starts conversion and the edge of the signal that indicates completion of the conversion.

conversion time—1. The length of time required by a computer to read out all the digits in a given coded word. 2. The time required for a complete conversion or measurement by an analog-to-digital converter, starting from a reset condition. In successive-approximation converters, it ranges typically from 0.8 microsecond to 400 microseconds. 3. Time required for an a/d converter to digitize an input signal. Throughput, the reciprocal of conversion time plus acquisition time, is expressed in channels per second.

conversion transconductance—The magnitude of the desired output-frequency component of current divided by the magnitude of the input-frequency component of voltage when the impedance of the output external termination is negligible for all frequencies that may affect the result.

conversion transducer—A transducer in which the signal undergoes frequency conversion. The gain or loss is specified in terms of the useful signal.

conversion voltage gain (of a conversion transducer)—With the transducer inserted between the input-frequency generator and the output termination, the ratio of the magnitude of the output-frequency voltage across the output termination to the magnitude of the input-frequency voltage across the input termination of the transducer.

convert—1. To change information from one form to another without changing the meaning, e.g., from one number base to another. 2. In computer terminology, to translate data from one form of expression to a different form.

converted data—The output from a unit that changes the language of information from one form to another so as to make it available or acceptable to another machine, e.g., a unit that takes information punched on cards to information recorded on magnetic tape, possibly including editing facilities.

converter—1. In a superheterodyne receiver, the section that converts the desired incoming rf signal into a lower carrier frequency known as the intermediate frequency. 2. A rotating machine consisting of an electric motor driving an electric generator, used for changing alternating current to direct current. 3. A facsimile device that changes the type of modulation delivered by the scanner. 4. Generally called a remodulator. A facsimile device that changes amplitude modulation to audio-frequency-shift modulation. 5. Generally called a discriminator. A device that changes audio-frequency-shift modulation to amplitude modulation. 6. A conversion transducer in which the output frequency is the sum or difference of the input frequency and an integral multiple of the local-oscillator frequency. 7. A device that accepts an input that is a function of maximum voltage and time and converts it to an output that is a function of maximum voltage only. 8. *See* shaft position encoder. 9. A device capable of converting impulses from one mode to another, such as analog to digital, or parallel to serial, or one code to another. 10. Device in a digital system that transforms information coded in one number system to its equivalent in another number system. Typically, conversion is either decimal-to-binary or binary-to-decimal.

converter tube—A multielement electron tube that combines the mixer and local-oscillator functions of a heterodyne conversion transducer.

converter unit—The unit of a radar system in which the mixer of a superheterodyne receiver and usually two stages of intermediate-frequency amplification are located. Performs a preamplifying operation.

converting—Changing data from one form to another to facilitate its transmission, storage, or the manipulation of information.

convex—Curved outward.

Cook system—An early stereo-disc recording technique in which the two channels were recorded simultaneously with two cutters on different portions (bands) of a record as concentric spirals. The playback equipment consisted of two pickups mounted side by side so that each played at the correct spot on its own band.

Coolidge tube—An X-ray tube in which the electrons are produced by a hot cathode.

coordinate digitizer—A device that transcribes graphic information in terms of a coordinate system for subsequent processing.

coordinated indexing—1. In a computer, a system in which individual documents are indexed by descriptors of equal rank so that a library can be searched for a combination of one or more descriptors. 2. A computer indexing technique in which the coupling of individual words is used to show the interrelation of terms.

coordinated transpositions—Transpositions that are installed in either electric supply or communication circuits, or in both, for the purpose of reducing induction coupling, and which are located effectively with respect to the discontinuities in both the electric supply and communication circuit.

coordinate system—A way by which a pair of numbers is associated with each point in a plane (or a triplet of numbers is associated with each point in three-dimensional space) without ambiguity.

coordination—A term describing the ability of the lower rating of two breakers in series to trip before the higher-rating one trips.

coordinatograph—A precision drafting instrument used in the preparation of artwork for mask making.

copperclad—A thin coating of copper fused to an aluminum core. Used in some building wires (No. 12 and larger).

copper-covered steel wire—A wire having a steel core to which is fused an outer shell of copper.

copper loss—*See* I^2R loss.

copper-oxide photocell—An early type of non-vacuum photovoltaic cell consisting of a layer of copper oxide on a metallic substrate, with a thin transparent layer of a conductor over the oxide. Light falling on the cell produces a small voltage between the substrate and the conducting layer. This type of cell is extensively used in exposure meters for cameras because it requires no external source of electric power.

copper-oxide rectifier—A metallic rectifier in which the rectifying barrier is the junction between metallic copper and cuprous oxide. A disc of copper is coated with cuprous oxide on one side, and a soft lead washer is used to make contact with the oxide layer.

copper-sulfide rectifier—A semiconductor rectifier in which the rectifying barrier is a junction between magnesium and copper sulfide.

copperweld—A thin coating of copper fused to a steel core. Used in line wire and cable messengers, and stranded with copper for strength or extending flex life.

coprocessor—1. A device that performs specialized processing in conjunction with the main microprocessor of a system. 2. In a computer, a device that performs specialized processing in conjunction with the main microprocessor. It works in tandem with another central processing unit to increase the computing power of a system. An extra microprocessor to handle certain tasks faster than the main processor. Same as math coprocessor, numeric coprocessor.

copy — 1. To hear a transmission. *See* subject copy. 2. To duplicate a file or program so that one can retain the original and work on the duplicate. Usually refers to duplicating one disk to another. *See also* backup.

copyguard — Also called stop-copy. Trademarked names for processing applied to a prerecorded video tape to prevent unauthorized copying of the recording. Typically the 60-Hz vertical-sync pulses are weakened, with the expectation that when they are further weakened in copying the tape, the image will roll vertically in playback.

copying telegraph — An absolute term for a facsimile system for the transmission of black-and-white copy only.

copy the mail — CB radio term for just listening to the radio without talking much.

Corbino effect — A special case of the Hall effect that occurs when a disc carrying a radial current is placed perpendicularly into a magnetic field.

cord — 1. One or a group of flexible insulated conductors covered by a flexible insulation and equipped with terminals. 2. A small, very flexible insulated cable constructed to withstand mechanical abuse. Generally, a cord is considered to be a size No. 10 and smaller.

cord circuit — A circuit, terminated in a plug at one or both ends, used at a telephone switchboard position in establishing connections.

cordless phone — A communication system that consists of two pieces: a transponder and a portable wireless handset. The transponder answers the telephone call, or processes an outgoing call, and is connected directly to the telephone line. Typically the transponder transmits to the handset on a frequency that is nominally 1.6 MHz (the high end of the broadcast band) by feeding rf into the ac power line. The handset receives the signal through a ferrite bar or loop antenna built into the handset. The handset transmits to the transponder on the radio control band at 49 MHz. The modulation is NBFM (narrow-band FM), which provides essentially noise-free reception both ways over an (approximately) 300-foot operating range.

cordless switchboard — A telephone switchboard in which manually operated keys are used to make connections.

cord sets — Portable cords fitted with any type of wiring device at one or both ends.

cordwood — 1. A sandwich-type construction wherein components lie in a vertical cordwood pattern between horizontal layers. 2. The technique of producing modules by bundling parts as closely as possible and interconnecting them into circuits by welding or soldering leads together.

cordwood module — 1. A high-density circuit module in which discrete components are mounted between and perpendicular to two small, parallel printed circuit boards to which their terminals are attached. 2. A module formed by bundling or stacking parts between a pair of end plates and interconnecting them into circuits by welding or soldering leads together.

core — 1. A magnetic material placed within a coil to intensify the magnetic field. 2. Magnetic material inside a relay or coil winding. 3. In fiber optics, the light-conducting portion of the fiber, defined by the high refractive index region. The core is normally in the center of the fiber, bounded by the cladding. 4. The central, light-transmitting portion of a fiber-optic cable. It must have a higher index of refraction than the cladding. 5. A small magnetic torus of ferrite used to store a bit of information. Core memories can be strung on wires so that organizations of 32K by 18 are possible in a size of $\frac{1}{2}$ inch high by 6 by 6 inches (1.27 by 15.24 by 15.24 cm). Advantages of core memory are that it is nonvolatile and the oldest main storage technology. (This technology is no longer used.)

core instruction set — A complete set of the operators of the instructions of a computer and the types of meanings associated with their operands.

coreless-type induction heater — A device in which an object is heated by induction without being linked by a magnetic core material.

core loss — Also called iron loss. Loss of energy in a magnetic core as the result of eddy currents, which circulate through the core and dissipate energy in the form of heat.

core memory — 1. A magnetic type of memory made up of miniature ferrite toroids, each of which can be magnetized in one direction to represent a 0 and in the other direction to represent a 1. It is a permanent memory, since if the power is removed the stored information remains. Core memory is characterized by low-cost storage and relatively slow memory operating speed. Core memories are nonvolatile, but have destructive readouts. 2. An array of doughnut-shaped ferrite cores whose diameter ranges from 9 to 18 mils (thousandths of an inch) or 37.5 to 75 μm. Core memories are arranged in a stack configuration in which fine wires are strung through the center of the cores, usually by manual methods. The wires supply current, which causes data to be written into and read out from the core. The storage property is a magnetic one in which the orientation of the core molecules is changed to read or write data. Core stacks can contain as few as 1024 bits or as many as 10^6 bits (one megabit). 3. A memory that is characterized by low-cost storage and relatively slow memory operating speed. Core memories are nonvolatile, but have destructive readouts. *See also* internal storage, 1. (This technology is no longer used.)

core plane — A horizontal network of magnetic cores that contains a core common to each storage position.

core rope storage — Direct-access storage in which a large number of doughnut-shaped ferrite cores are arranged on a common axis, and sense, inhibit, and set wires are threaded through individual cores in a predetermined manner to provide fixed storage of digital data. Each core stores one or more complete words instead of a single bit.

core storage — 1. In a computer, a form of high-speed storage that uses magnetic cores. 2. In a calculator, a storage register in which the contents will remain even after the machine has been switched off.

core store — A matrix of small magnetic rings or cores upon which electrical pulses may be stored. The presence of a pulse in a train is recorded by magnetizing a core, the absence of a pulse by leaving a core unmagnetized.

core transformer — A transformer in which the windings are placed on the outside of the core.

core tuning — Adjusting the inductance and thereby the frequency of resonance of a coil by moving a powdered iron or ferrite core in or out of the coil.

core-type induction heater — A device in which an object is heated by induction. Unlike the coreless type, a magnetic core links the induction winding to the object.

core wrap — Insulation placed over a core before the addition of windings.

corner — 1. An abrupt change in direction of the axis of a waveguide. 2. A neighborhood or point at which a curve makes a sharp or discontinuous change of slope.

corner cut — A corner removed, for orientation purposes, from a card to be used with a computer.

corner effect — The rounding off of the attenuation versus frequency characteristic of a filter at the extremes (or corners) of the passband.

corner frequency — 1. The frequency at which the open-loop gain-versus-frequency curve changes slope. For a servo motor, the product of the corner frequency in radians per second and the time constant of the motor is unity. 2. The frequency at which the two asymptotes of the gain-magnitude curves of an operational amplifier intersect. 3. The upper frequency at which 3-dB attenuation occurs in a high-gain amplifier. A cornering circuit usually is introduced to attenuate the high-frequency signals before the natural phase shift of the amplifier becomes greater than 90°. When properly designed, the cornering circuit prevents high-frequency oscillations in feedback amplifiers. The corner frequency is sometimes erroneously referred to as the cutoff frequency.

corner reflector — A reflecting object consisting of two (dihedral) or three (trihedral) mutually intersecting conducting surfaces. Trihedral reflectors are often used as radar targets.

corner-reflector antenna — An antenna consisting of a primary radiating element and a dihedral corner reflector formed by the elements of the reflector.

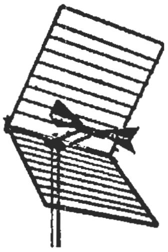

Corner-reflector antenna.

corona — 1. A luminous discharge of electricity, due to ionization of the air, appearing on the surface of a conductor when the potential gradient exceeds a certain value but is not sufficient to cause sparking. 2. Any electrically detectable, field-intensified ionization that occurs in an insulating system but does not result immediately in catastrophic breakdown. (Corona always precedes dielectric breakdown.) 3. The ionization of gases about a conductor that results in a bluish-purple glow due to the voltage differential between a high-voltage conductor and the surrounding atmosphere. 4. A device used in an electrostatic copier to impart an electrical charge (in the dark) to the photoconductive material (zinc-oxide-coated paper) to make it sensitive to the action of light. 5. The small, erratic current pulses resulting from discharges in voids in a dielectric during voltage stress.

corona discharge — A phenomenon that occurs when an electric field is sufficiently strong to ionize the gas between electrodes and cause conduction. The effect is usually associated with a sharply curved surface, which concentrates the electric field at the emitter electrode. The process operates between an inception voltage and a spark breakdown voltage. These potentials and the current-voltage characteristics within the operating range are affected by the polarity of the corona electrodes as well as the composition and density of the gas in which the discharge occurs.

corona effect — The glow discharge that occurs in the neighborhood of electric conductors where the insulation is subject to high electric stress. With an alternating current, the effect produced when two wires, or other conductors having a great difference of voltage, are placed near each other.

corona endurance — Resistance to corona cutting.

corona extinction voltage — Abbreviated cev. The voltage at which discharges preceded by corona cease as the voltage is reduced. The corona extinction voltage is always lower than the corona start voltage.

corona failure — Failure due to corona degradation at areas of high voltage stress.

corona loss — A loss or discharge that occurs when two electrodes having a great difference of pressure are placed near each other. The corona loss takes place at the critical voltage and increases very rapidly with increasing pressure.

corona resistance — Also called ionization resistance, brush-discharge resistance, slot-discharge resistance, or voltage endurance. 1. That length of time that an insulation material withstands the action of a specified level of field-intensified ionization that does not result in the immediate, complete breakdown of the insulation. 2. The ability of a material to withstand sustained high applied voltage.

corona shield — A shield placed around a high-potential point to redistribute electrostatic lines of force and prevent corona.

corona start voltage — Abbreviated csv. The voltage at which corona discharge begins in a given system.

corona voltage level — The voltage at which corona discharge does not exceed a specified level following the application of a specified higher voltage.

corona voltmeter — A voltmeter in which the peak voltage value is indicated by the beginning of corona at a known and calibrated electrode spacing.

correction — An increment that, when added algebraically to an indicated value of a measured quantity, results in a better approximation to the true value of the quantity.

corrective equalization — *See* frequency-response equalization.

corrective maintenance — The maintenance performed on a nonscheduled basis to restore equipment to satisfactory condition.

corrective network — Also called shaping network. An electric network designed to be inserted into a circuit to improve its transmission or impedance properties, or both.

correed — A glass-enclosed miniature reed switch. It is similar to the ferreed except that it is operated only when there is current through its surrounding winding, and releases when current stops.

correed relay — A device consisting of a hermetically sealed reed capsule surrounded by a coil. It is used as a switching device in telephone equipment.

correlated characteristic — A characteristic known to be reciprocally related to some other characteristic.

correlation — 1. The relationship, expressed as a number between −1 and 1, between two sets of data, etc. 2. A relationship between two variables; the strength of the linear relationship is indicated by the coefficient of correlation. 3. A measure of the similarity of two signals.

correlation detection — A method of detection in which a signal is compared, point to point, with an internally generated reference. The output of such a detector is a measure of the degree of similarity of the input and reference signal. The reference signal is constructed in such a way that it is at all times a prediction, or best guess, of what the input signal should be at that time.

correlation direction finder — A satellite station separated from radar to receive a jamming signal. By correlating the signals received from several such stations, the range and azimuth of many jammers may be obtained.

correlation distance — A term used in tropospheric propagation. The minimum spatial separation between antennas that will give rise to independent fading of the received signals.

correlation orientation tracking and range system — A system generally using a parabolic antenna for the analysis of a narrow band of radar energy for tracking and ranging purposes.

correlation tracking and ranging — A nonambiguous short-base-line, single-station, cw phase comparison system measuring two direction cosines and a slant range, from which space position can be computed.

correlation tracking and triangulation — A trajectory-measuring system composed of several antenna base lines separated by large distances and used to measure direction cosines to an object. From these measurements, the space position is computed by triangulation.

correlation tracking system — A system utilizing correlation techniques in which signals derived from the same source are correlated to derive the phase difference between the signals. This phase difference contains the system data.

corrosion — 1. A chemical action that causes gradual destruction of the surface of a metal by oxidation or chemical contamination. Also caused by reduction of the electrical efficiency between the metal and a contiguous substance or the disintegrating effect of strong electrical currents or ground-return currents in electrical systems. The latter is known as electrolytic corrosion. 2. In semiconductors, a defect in or on the aluminum metallization, usually a white crystalline growth. 3. A material's chemical alteration by electrochemical interaction with its environment. Corrosion reflects a metal's proclivity to return to the more stable compound state from which it was refined. Corrosive reaction represents an essential division of chemical kinetics.

corrosive fluxes — Also called acid fluxes. Fluxes consisting of inorganic acids and salts; they are generally required where the condition of the surface is well below the ideal for rapid wetting by molten solder.

cosecant-squared antenna — An antenna that emits a cosecant-squared beam. In the shaped-beam antenna used, the radiation intensity over part of its pattern in some specified plane (usually the vertical) is proportionate to the square of the cosecant of the angle measured from a specified direction in that plane (usually the horizontal).

cosecant-squared beam — A radar-beam pattern designed to give uniform signal intensity in echoes from distant and nearby objects. It is generated by a spun-barrel reflector. The beam intensity varies as the square of the cosecant of the elevation angle.

cosine law — The law which states that the brightness in any direction from a perfectly diffusing surface varies in proportion to the cosine of the angle between that direction and the normal to the surface.

cosmic noise — Radio static whose origin is due to sources outside the earth's atmosphere. The source may be similar to sunspots, or spots on other stars.

cosmic rays — Any rays of high penetrating power produced by transmutations of atoms in outer space. These particles continually enter the earth's upper atmosphere from interstellar space.

COS/MOS — *See* complementary metal-oxide semiconductor.

coulomb — 1. The quantity of electricity that passes any point in an electric circuit in 1 second when the current is maintained constant at 1 ampere. The coulomb is the unit of electric charge in the mksa system. 2. The measure of electric charge, defined as a charge equivalent to that carried by 6.281×10^{18} electrons.

Coulomb's law — Also called law of electric charges or law of electrostatic attraction. The force of attraction or repulsion between two charges of electricity concentrated at two points in an isotropic medium is proportionate to the product of their magnitudes and is inversely proportionate to the square of the distance between them. The force between unlike charges is an attraction, and the force between like charges is a repulsion.

coulometer — An electrolytic cell that measures a quantity of electricity by the amount of chemical action produced.

Coulter counter — An electronic cell-counting instrument operating on the ion-conductivity principle. Designed by J. R. Coulter. *See also* cell counter.

count — In radiation counters, a single response of the counting system.

countdown — A decreasing tally that indicates the number of operations remaining in a series.

counter — 1. A circuit that counts input pulses. One specific type produces one output pulse each time it receives some predetermined number of input pulses. The same term may also be applied to several such circuits connected in cascade to provide digital counting. Also called divider. 2. In mechanical analog computers, a means for measuring the angular displacement of a shaft. 3. Sometimes called accumulator. A device capable of changing from one to the next of a sequence of distinguishable states upon receipt of each discrete input signal. 4. An arrangement of flip-flops producing a binary word that increases in value by 1 each time an input pulse is received. It may also be called a divider, since successive counter stages divide the input frequency by 2. A counter has a maximum count, depending on its size, called a modulus or mod. For example, a mod-8 counter can count up to 7, and on the eighth input it resets itself back to a count of 0. When it resets, it also provides an output pulse, which could be counted by another counter. 5. A device capable of changing stages in a specified sequence upon receiving appropriate input signals; a circuit that provides an output pulse or other indication after receiving a specified number of input pulses. 6. A memory-type digital building block that counts pulses received at its input and transmits the cumulative total at its output. 7. In relay-panel hardware, an electromechanical device that can be wired and preset to control other devices according to the total cycles of one on and off function. 8. An instrument that detects and records the occurrence of events either for as long as the counter remains energized or over some predetermined period.

counterbalance — A weight, usually adjustable, fitted at the pivot end of a pickup arm. It counters the weight of the pickup head and cartridge unit and allows adjustment of the stylus pressure to the desired value.

counter circuit — A circuit that receives uniform pulses representing units to be counted and produces a voltage in proportion to their frequency.

counterclockwise polarized wave — *See* left-handed polarized wave.

counter-countermeasures — Use of anti-jamming techniques and circuits designed to decrease the effectiveness of electronic countermeasure activities on electronic equipment.

counterelectromotive cell—A cell of practically no ampere-hour capacity used to oppose the line voltage.

counterelectromotive force—Abbreviated counter EMF. A voltage developed in an inductive circuit by an alternating or pulsating current. The polarity of this voltage is at every instant opposite that of the applied voltage.

countermeasures—That part of military science dealing with the employment of devices and/or techniques intended to impair the operational effectiveness of enemy activity.

counterpoise—A system of wires or other conductors, elevated above and insulated from ground, forming a lower system of conductors of an antenna. Used to capacitively couple a radio transmitter to the ground when the ground resistance is high.

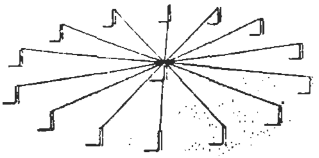

Counterpoise.

counters (Geiger and scintillation)—Instruments used to detect ionizing radiation having very short wavelength (about one-thousandth the wavelength of visible light). Natural sources of this radiation are radium, uranium isotopes, cosmic rays, and ores in which these elements are present; man-contrived sources are the atomic bomb, nuclear reactors used for generating electric power, high-voltage radar CRTs, and X-ray machines.

counter tube—Also called radiation counter tube. An electron tube that converts an incident particle or burst of incident radiation into a discrete electric pulse. This is generally done by utilizing the current flow through a gas that is ionized by the radiation.

counting efficiency—In a scintillation counter, the ratio, under specified conditions, of the average number of photons or particles of ionizing radiation that produce counts to the average number of photons or particles incident on the sensitive area.

counting-rate meter—A device for indicating the time rate of occurrence of input pulses averaged over a time interval.

counting-type frequency meter—An instrument for measuring frequency. Its operation depends on the use of pulse-counting techniques to indicate the number and/or rate of recurring electrical signals applied to its input circuits.

counts—Clicking noises made by a radiation-detecting instrument in the presence of radiation. *See* scintillation counter.

counts per turn—The total number of code positions per 360° of encoder shaft rotation.

couple—Two or more dissimilar metals or alloys in electrical contact with each other that act as the electrodes of an electrolytic cell when they are immersed in an electrolyte.

coupled circuit—Any network containing only resistors, inductors (self and mutual), and capacitors and having more than one independent mesh.

coupled modes—In a waveguide, such as an optical fiber, coaxial cable, or metal pipe, coexisting propagation modes whose fields are interrelated and whose energies are mutually interchanged.

coupler—1. A passive device that divides an antenna signal to feed two or more receivers, or combines two or more antenna signals to feed a single down lead. A coupler provides some interset isolation and maintains an impedance match between the antenna and receiver. 2. A component that interconnects a number of optical waveguides (fibers) and provides an inherently bidirectional system by mixing and splitting all signals within the component.

coupling—The association or mutual relationship of two or more circuits or systems in such a way that power may be transferred from one to another.

coupling angle—In connection with synchronous motors, the mechanical-degree relationship between the rotor and the rotating field.

coupling aperture—Also called coupling hole or coupling slot. An aperture, in the wall of a waveguide or cavity resonator, designed to transfer energy to or from an external circuit.

coupling capacitor—1. Any capacitor used to couple two circuits together. Coupling is accomplished by means of the capacitive reactance common to both circuits. 2. A capacitor that allows alternating currents to pass but blocks direct currents. 3. A capacitor intended for the coupling of two or more ac circuits with different dc levels.

coupling coefficient—Also called coefficient of coupling. The degree of coupling that exists between two circuits. It is equal to the ratio between the mutual impedance and the square root of the product of the total self-impedances of the coupled circuits, all impedances being of the same kind.

coupling hole—*See* coupling aperture.

coupling loop—A conducting loop projecting into a waveguide or cavity resonator and designed to transfer energy to or from an external circuit.

coupling loss—1. *See* fiber optics. Those signal losses due to small differences in numerical aperture, core diameter, core concentricity, and tolerances in splicing connectors when two fibers are aligned. Also known as splicing loss and transfer loss. 2. The total optical power loss within a junction, expressed in decibels, attributed to the termination of the optical conductor. 3. Attenuation of the optical signal due to coupling inefficiencies between the flux source and the optical fiber, or between fibers, or between the fiber and the detector in a receiver. Expressed in decibels.

coupling probe—A probe projecting into a waveguide or cavity resonator and designed to transfer energy to or from an external circuit.

coupling slot—*See* coupling aperture.

coupling transformer—A transformer that couples two circuits together by means of its mutual inductance.

courtesy tone—An audible signal transmitted by a repeater that lets users know the repeater had reset at the end of one person's transmission and is available for use by the next person.

covalent bond—A type of linkage between atoms. Each atom contributes one electron to a shared pair that constitutes an ordinary chemical bond.

coverage—A percentage of the completeness with which a braid or shield covers the surface of an underlying insulated conductor or conductors.

covered relay—A relay contained in an unsealed housing.

Covington and Broten antenna—A compound interferometer in which a long line source is adjacent to a two-element interferometer of comparable aperture, in the same straight line.

CPC—Abbreviation for calling party control. A telephone signaling system that notifies the terminating office and any line-connected auxiliary equipment when the calling party has disconnected. This system permits more efficient use of telephone trunk lines by removing the called party from the line as soon as the calling party disconnects. The CPC signal is a pulse to ground potential, usually 100 ms long, equivalent to shorting the two wires of the phone line together.

cpm—Abbreviation for cycles per minute.

C power supply—A device connected in the circuit between the cathode and grid of a vacuum tube to apply grid bias.

cps—1. Abbreviation for cycles per second, an obsolete term. Replaced by the term *hertz*, abbreviated Hz. 2. The number of times per second an electronic event is repeated. 3. Abbreviation for characters per second when speaking of data transmission. A data-rate unit, not to be confused with cycles per second.

cps/bps—The number of characters or bytes per second (bits per track per second) written to or read from a magnetic tape.

CPU—Abbreviation for central processing unit. A primary unit of the computer system that controls interpretation and execution of instructions.

CPU portion—*See* chip sets.

crash—1. A computer condition that causes it to stop working for some reason and need to be restarted by the operator. 2. Abrupt computer failure.

crash-locator beacon—Airborne equipment consisting of various transmitters, collapsible antennas, etc., designed to be ejected from a downed aircraft and to transmit beacon signals to help searching forces to locate the crashed aircraft.

crater lamp—1. A glow-discharge type of vacuum tube whose brightness is proportional to the current passing through the tube. The glow discharge takes place in a cup or crater rather than on a plate as in a neon lamp. 2. A gaseous lamp usually containing neon. Provides a point source of light that can be modulated with a signal.

crazing—Checking of an insulation material when it is stressed and in contact with certain solvents or their vapors.

CRC—Abbreviation for cyclic redundancy check. 1. A method of error detection consisting of a character generated at the transmitting terminal that is matched with a character at the receiving terminal. Matched characters signify correct character reception; unmatched characters indicate an error. 2. An error-checking control technique utilizing a binary prime divisor that produces a unique remainder.

credence—A measure of confidence in a radar target detection; generally it is proportional to the target-return amplitude.

credit balance indicator—In a calculator, warning light to indicate a negative answer.

creepage—The conduction of electricity across the surface of a dielectric.

creepage distance—The shortest distance between conductors of opposite polarities, or between a live part and ground, measured over the surface of the supporting material.

creepage path—The path across the surface of a dielectric between two conductors. Lengthening the creepage path reduces the possibility of arc damage or tracking.

creepage surface—An insulating surface that provides physical separation between two electrical conductors of different potential.

creep-controlled bonding—A method of diffusion bonding in which enough pressure is exerted to cause significant creep deformation at the joint interfaces. The method is characterized by use of intermediate and low unit loads for a period of hours.

creep distance—The shortest distance on the surface of an insulator between two electrically conductive surfaces separated by the insulator.

creep recovery—The change in no-load output occurring with time after removal of a load that had been applied for a specific period of time.

crest factor—1. The ratio of the peak voltage to the rms voltage of a waveform (with the dc component removed). 2. An instrument's dynamic range and ability to respond faithfully to waveform peaks as the rms value approaches full scale. Can also refer to the quality of rms-conversion techniques in general.

crest value—Also called peak value. The maximum absolute value of a function.

crest voltmeter—A peak-reading voltmeter.

crest working off-stage voltage—The highest instantaneous value of the off-state voltage that occurs across a thyristor, excluding all repetitive and nonrepetitive transient voltages.

crest working reverse voltage—The highest instantaneous value of the reverse voltage that occurs across a semiconductor diode or reverse-blocking thyristor, excluding all repetitive and nonrepetitive transient voltages.

CRI—*See* color rendering index.

crimp—1. To compress or deform a connector barrel around a cable so as to make an electrical connection. 2. Final configuration of a terminal barrel formed by the compression of terminal barrel and wire.

crimp contact—A contact whose back portion is a hollow cylinder to allow it to accept a wire. After a bared wire is inserted, a swedging tool is applied to crimp the contact metal firmly against the wire. An excellent mechanical and electrical contact results. A crimp contact often is referred to as a solderless contact.

crimping—A method of attaching a terminal, splice, or contact to a conductor through the application of pressure.

crimping tool—A device used to apply solderless teminals to a conductor.

crimp terminal—A point at which the bared portion of the hookup wire is crimped to either the contact or a tab or pin that mates with the contact terminal.

crimp termination—Connection in which a metal sleeve is secured to a conductor by mechanically crimping the sleeve with pliers, presses, or automated crimping machines. Splices, terminals, and multicontact connectors are typical terminating devices attached by crimping. Suitable for all wire types.

crimp-type termination—Open-barrel or closed-barrel termination in which a stripped wire is inserted into or through a tube that is crimped to the wire with an appropriate tool.

crippled leapfrog test—In a computer, a variation of the leapfrog test in which the test is repeated from a single set of storage locations rather than from a changing set of storage locations. *See also* leapfrog test.

critical area—*See* picture element, 2.

critical angle—1. The maximum angle at which a radio wave may be emitted from an antenna and will be returned to the earth by refraction in the ionosphere. 2. The maximum angle of incidence for which light will be transmitted from one medium to another. Light approaching the interface at angles greater than the critical angle will be reflected back into the first medium. 3. The maximum angle at which light can be propagated within a fiber. The sine of the critical angle equals the ratio of the numerical aperture to the index of refraction of the fiber core. 4. Basically, the least angle of incidence at which total reflection takes place. The angle of incidence in a denser medium, at an interface between the denser and less dense medium, at which the light is refracted along the interface. When the critical angle is exceeded, the light is totally reflected back into the denser medium. The critical angle varies with the indexes of refraction of the two media with the relationship

$$\sin l_c = n'/n$$

where

l_c is the critical angle,
n' is the refractive index of the less dense medium, and
n is the refractive index of the denser medium.

critical characteristic — A characteristic not having the normal tolerance to variables.

critical coupling — Also called optimum coupling. Between two circuits independently resonant to the same frequency, the degree of coupling that transfers the maximum amount of energy at the resonant frequency.

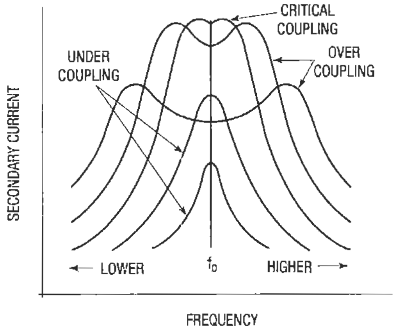

Critical coupling.

critical current — That current, at a specified temperature and in the absence of external magnetic fields, above which a material is normal and below which it is superconducting.

critical damping — The value of damping that provides the most rapid transient response without overshoot. Operation between underdamping and overdamping.

critical dimension — The dimension of a waveguide cross section that determines the cutoff frequency.

critical failure — A failure that causes a system to operate outside designated limits.

critical field — Also called cutoff field. Of a magnetron, the smallest theoretical value of a steady magnetic-flux density, at a steady anode voltage, that would prevent an electron emitted from the cathode at zero velocity from reaching the anode.

critical flicker frequency — The frequency at which a flickering light source is perceived as changing from pulsating to continuous.

critical frequency — Also called penetrating frequency. The limiting frequency below which a magneto-ionic wave component is reflected by an ionospheric layer and above which the component penetrates the layer at vertical incidence. *See also* waveguide cutoff frequency.

critical grid current — The instantaneous value of grid current in a gas tube when the anode current starts to flow.

critical grid voltage — The instantaneous value of grid voltage at which the anode current starts to flow in a gas tube.

critical high-power level — The radio-frequency power level at which ionization is produced in the absence of a control-electrode discharge.

critical inductance — In an inductor input power supply filter, the minimum value of the input inductor needed to ensure that the current drawn through the rectifier never goes to zero.

critical item — *See* critical part.

critical loads — Those loads that will not tolerate a loss of power without incurring major losses in the form of product or equipment damage, or without creation of safety hazards. Critical loads can be subclassified as those that are uninterruptible or those that are short-term interruptible for the time it takes transfer switches to operate.

critical magnetic field — That field intensity below which, at a specified temperature and in the absence of current, a material is superconducting and above which it is normal.

critical part — A part whose failure to meet specified requirements results in the failure of the product to serve its intended purpose. Also called critical item.

critical path — The path that determines a circuit's overall clock rate, typically the longest path in the circuit.

critical potential — *See* ionization potential.

critical race — *See* race

critical rate of rise of off-state voltage — The minimum value of the rate of rise of principal voltage that will cause a semiconductor switching device to switch from the off state to the on state under specified conditions.

critical rate of rise of on-state current — The maximum value of the rate of rise of on-state current that a thyristor can withstand without deleterious effect.

critical resistance value — For a given voltage rating and a given power rating, the one value of resistance that will dissipate full rated power at rated voltage. For values of resistance below the critical value, the maximum (element) voltage is never reached; for values of resistance above the critical value, the power dissipated becomes lower than rated.

critical temperature — That temperature below which, in the absence of current and external magnetic fields, a material is superconducting and above which it is normal.

critical voltage — Also called cutoff voltage. In a magnetron, the highest theoretical value of steady anode voltage, at a given steady magnetic-flux density, at which electrons emitted from the cathode at zero velocity will fail to reach the anode.

critical wavelength — The free-space wavelength corresponding to the critical frequency.

CRO — Abbreviation for cathode-ray oscilloscope.

CROM — Abbreviation for control read-only memory.

Crookes' dark space — *See* cathode dark space.

Crosby system — A compatible multiplex FM stereo broadcast technique in which the right and left signals are combined in phase (sum signal) and transmitted on the main carrier, and also combined out-of-phase (difference signal) and transmitted on the subcarrier. The two signals are combined (matrixed) in the receiving apparatus to restore the right and left channels.

cross alarm — 1. An alarm condition signaled by crossing or shorting an electrical circuit. 2. The signal produced due to a cross alarm condition.

cross assembler — 1. A symbolic language translator that runs on one type of computer to produce machine code for another type of computer. 2. A type of assembler program that generates binary code of a program for a computer other than the model with which it is being used; e.g., an 8080 cross assembler might operate on a PDP-8 minicomputer. 3. Frequently used in conjunction with a down-line load capability for remote control of an unattended microprocessor. *See* assembler; resident assembler.

crossbanding—The use of combinations of interrogation and reply frequencies such that either one interrogation frequency is used with several reply frequencies or one reply frequency is used with several interrogation frequencies.

crossbar—A type of telephone control switching system using a crossbar or coordinate switch. Crossbar switching systems suit data switching because they have low noise characteristics and can handle Touch Tone® dialing.

crossbar switch—1. A switch having a number of vertical paths, a number of horizontal paths, and an electromagnetically operated mechanism for interconnecting any one vertical path with any one horizontal path. An automatic telephone switching system that uses a crossbar switch. 2. A switch with a plurality of vertical paths and electromechanically operated mechanical means for interconnecting any one of the vertical paths with any of the horizontal paths.

crossbar switching system—A method of switching that, when directed by a common control unit, will select and close a path through a matrix arrangement of switches.

crossbar system—An automatic telephone switching system in which, generally, the selecting devices are crossbar switches. Common circuits select and test the switching paths and control the selecting mechanisms. The method of operation is one in which the switching information is received and stored by controlling mechanisms that determine the operation necessary to establish a connection.

cross beat—A spurious frequency that arises as a result of cross modulation.

crosscheck—To check a computation by two different methods.

cross color—The interference in a color-television receiver chrominance channel caused by crosstalk from monochrome signals.

cross-compiler—A compiler that runs on one computer system but generates machine code for another computer system. Typically it runs on a large computer and generates code for a microcomputer, speeding up software development.

cross-correlation function—A measure of the similarity between two signals when one is delayed with respect to the other.

cross coupling—Unwanted coupling between two different communication channels or their components.

cross-current conduction—Transistor turn-off delay that stems mainly from storage in a converter's switching transistors. The phenomenon appears as a stretching out of the saturated conduction period after base drive is removed.

crossed-pointer indicator—A two-pointer indicator used with instrument landing systems to indicate the position of an aircraft with respect to the glide path.

cross-field device—An electronic device in which electrons from the cathode are influenced by a magnetic field that acts at right angles to the applied electric field. When electrons leave the cathode in a direction perpendicular to the magnetic field, this field causes a force to act at right angles to the electron motion. The electrons then spiral into orbit around the cathode instead of moving colinearly with the electric field. Most of the electrons gradually move toward the anode, giving up potential energy to the rf field as they interact with the anode slow-wave structure. The tube structure may be either cylindrical or linear.

crossfield recording—A system in which the bias is not applied to the tape by a recording head but by a separate head on the tape's backing side, so that the bias signal will not partially erase high frequencies as they are being recorded.

crossfire—Interfering current in one telegraph or signaling channel resulting from telegraph or signaling currents in another channel.

crossfoot—1. In a computer, to add or subtract numbers in different fields of the same punch card and punch the result into another field of the same card. 2. To compare totals of the same numbers obtained by different methods.

cross hairs—On a cursor, a horizontal line intersected by a vertical line to indicate a point on the display whose coordinates are desired.

crosshatching—1. In a printed circuit board, the breaking up of large conductive areas where shielding is required. 2. Process of filling in an outline with a series of symbols to highlight part of a design.

cross magnetostriction—Under specified conditions, the relative change of dimension in a specified direction perpendicular to the magnetization of a body of ferromagnetic material when the magnetization of the body is increased from zero to a specified value (usually saturation).

cross modulation—1. A spurious response or form of interference that occurs when the carrier of a desired signal intermodulates with the carrier of an undesired signal. This often happens in early stages of radio receivers, particularly when strong signals from local stations drive these stages into nonlinear operation. 2. A form of interference caused by the modulation of one carrier affecting that of another signal. It can be caused by overloading an amplifier as well as by signal imbalances at the headend.

cross modulation distortion—The amount of modulation impressed on an unmodulated carrier when a signal is simultaneously applied to the rf port of a mixer under specified operating conditions. The tendency of a mixer to produce cross modulation is decreased with an increase in conversion compression point and intercept point.

cross neutralization—A method of neutralization used in push-pull amplifiers. A portion of the plate-to-cathode ac voltage of each tube is applied to the grid-to-cathode circuit of the other tube through a neutralizing capacitor.

cross office switching time—The time required for connection of any input through the switching center to any selected output.

crossover—1. The point where two conductors that are insulated from each other cross. 2. A connection formed between two elements of a circuit by depositing conductive material across the insulated upper surface of another interconnection or element. 3. A point in an integrated or MOS circuit at which an interconnect pattern passes over another conductive part of a circuit but is insulated from it by a thin dielectric layer. *See also* underpass.

crossover distortion—1. Distortion that occurs in a push-pull amplifier at the points of operation where the input signals cross over (go through) the zero reference points. 2. The type of distortion resulting from class B push-pull power amplifiers owing to the lack of coincidence of the two transfer characteristics at the crossover point. The effect is reduced by applying a critical value of biasing to optimize the quiescent current and hence "linearize" the middle portion of the transfer characteristic. The situation is further improved by heavy negative feedback and by circuit design, such that the crossover distortion from hi-fi amplifiers is very small.

crossover frequency — 1. As applied to electrical dividing networks, the frequency at which equal power is delivered to each of the adjacent frequency channels when all channels are terminated in the specified load. *See also* transition frequency. 2. A frequency at which other frequencies above and below it are separated. In a two-way speaker system, for instance, the crossover frequency is the point at which woofer and tweeter response is divided.

crossover network — 1. An electrical filter that separates the output signal from an amplifier into two or more separate frequency bands for a multispeaker system. 2. A circuit (usually employing capacitors and coils) that feeds low notes to a low-frequency speaker (woofer) and high notes to a high-frequency speaker (tweeter). The crossover frequency is that at which frequency bands divide. Sometimes the audio spectrum is divided into more than two bands to drive more than two speakers. In a three-way system, midrange frequencies go to a midrange driver. Frequencies outside the range of each driver are attenuated at a rate determined by the network design. 3. A frequency at which each of two drivers is receiving half the amplifier's power; below or above that point, one speaker will receive more power than the other.

crossover spiral — *See* lead-over groove.

crosspoint — The operated contacts on a crossbar switch.

cross polarization — 1. The component electric field vector normal to the desired polarization component. 2. Describes signals of opposite polarization being transmitted and received. Cross-polarization discrimination refers to the ability of a feed to detect one polarity and reject the opposite-polarity signals.

cross-reference generator — A device that permits symbols (labels, variables, constants) to be correlated with their storage locations in a computer.

cross-sectional area of a conductor — The summation of all cross-sectional areas of the individual strands in the conductor, expressed in square inches or, more commonly, in circular mils.

cross software — Programs that permit a target system to be developed on a host computer with different CPU architecture.

crosstalk — 1. Interference caused by stray electromagnetic or electrostatic coupling of energy from one circuit to another. 2. Undesired signals from another circuit in the same system. 3. Breakthrough of the signal from one channel to another by conduction or radiation. 4. Transient noise induced on a switching signal by interaction with other switching transitions. 5. Audio interference from one track of a stereo tape to another. Poor head alignment often causes this. 6. Leakage of recorded signal from one channel of a stereo device into the adjacent channel or channels. Crosstalk between stereo channels impairs stereo separation; crosstalk between reverse-direction track signals can be heard, backward, during quieter parts of the desired program. 7. A measure of the amount of signal input to an off channel that appears at the output of a multiplexer, superimposed on the signal passed through the on channel. This is a direct function of the frequency of the signals, since semiconductor switches are capacitively coupled within the IC chip. The higher the frequency, the greater the crosstalk. This phenomenon is similar to the feedthrough problem of signal-handling devices. 8. In a transmission line, the noise generated in a passive signal line due to information traveling down an active line in the cable. 9. Measurable leakage of optical energy from one optical conductor to another. 10. Optical coupling between a pair of optical fibers caused by light leakage.

crosstalk coupling — Also called crosstalk loss. Cross coupling between speech communication channels or their component parts. Note: Crosstalk coupling is measured between specified points of the disturbing and disturbed circuit and is preferably expressed in decibels.

crosstalk level — 1. The volume of crosstalk energy, measured in decibels, referred to a base. 2. The effective power of crosstalk, expressed in decibels below 1 mW.

crosstalk loss — *See* crosstalk coupling.

crossunder — 1. A connection of two elements of a circuit by a conductive path deposited or diffused into a substrate. 2. A point in an integrated or MOS circuit at which there is a crossing of two conductive paths, one of which is built into the active substrate for interconnection.

crowbar — 1. A term describing the action that effectively creates a high overload on the actuating member of the protective device. This crowbar action may be triggered by a slight increase in current or voltage. 2. A circuit that shorts the output terminals of a power supply if the output of the power supply rises above a certain preset limit. The crowbar thus brings the output voltage to zero. 3. To place a low-resistance short across the input to a circuit. 4. A protective circuit in a power distribution circuit that shorts the power supply output to ground when an overvoltage condition occurs.

crowbar circuit — 1. An electronic switching system used to protect high-voltage circuits from damage caused by arc currents. The system places a momentary short across the circuit to be protected. 2. A circuit that protects a circuit or system from dangerously high voltage surges by shutting down the power source.

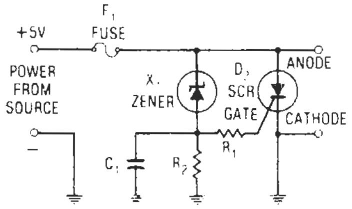

Crowbar overvoltage circuit (SCR).

crowbar protection circuit — A protection circuit that by rapidly placing a low resistance across the output terminals of a power supply initiates action that reduces the output voltage to a low value.

crowbar voltage protector — A separate circuit that monitors the output of a power supply and instantaneously throws a short circuit (or crowbar) across the output terminals of the power supply whenever a preset voltage limit is exceeded. An SCR is usually used as the crowbar device.

CRT — Abbreviation for cathode-ray tube.

CRT display — *See* cathode-ray-tube display.

CRT terminal — A visual-display terminal with an electrostatic deflection-type display. The term sometimes refers to all video "glass" terminals. These terminals' displays use either screen storage or memory-refreshed, raster-scanning techniques to generate visuals.

cryoelectrics—Technology having to do with the characteristics of electronic components at cryogenic temperatures. The branch of electronics that is concerned with applications of cryogenics. A contraction of cryogenic electronics

cryogenic—Of or having to do with temperatures approaching absolute zero.

cryogenic device—A device inteneded to function best at temperatures near absolute zero.

cryogenic motor—A motor that operates at a temperature below −129°C and that uses a cryogenic fluid or gas to cool its windings and bearings.

cryogenics—1. The subject of physical phenomena at temperatures below about −50°C. More generally, the term is used to refer to methods for producing very low temperatures. Also called cryogeny. 2. The study of the behavior of matter at supercold temperatures. 3. The science and technology applied to the creation of low temperatures (i.e., approaching absolute zero).

cryogenic temperature—A temperature close to absolute zero.

cryogeny—*See* cryogenics.

cryosar—A semiconductor device primarily intended for high-speed computer switching and memory applications. This device operates by the low-temperature avalanche breakdown produced by impact ionization of impurities.

cryosistor—A cryogenic semiconductor device in which the ionization between two ohmic contacts is controlled by means of a reverse-biased pn junction. After ionization, the device can act as a three-terminal switch, a pulse amplifier, an oscillator, or a unipolar transistor.

cryostat—A refrigerating unit such as that for producing or utilizing liquid helium in establishing extremely low temperatures (approaching absolute zero).

cryotron—1. A superconductive four-terminal device in which a magnetic field, produced by passing a current through two input terminals, controls the superconducting-to-normal transition—and thus the resistance—between the two output terminals. 2. An electrical switching and binary storage instrument that depends on the oscillating of a superconducting component between low and high resistance levels due to alternation in the magnetic field.

cryotronics—A contraction of cryogenic electronics.

cryptanalyst—A person who solves cryptograms, that is, one who converts secret messages into their original plain-language form without having authorized knowledge of the cryptographic system.

crypto-—1. A prefix used to form words that pertain to the transformation of data to conceal its actual meaning, usually by conversion to a secret code. 2. The science of encrypting data so that only the intended recipient can read it.

cryptochannel—Complete system of cryptocommunications between two or more holders.

cryptogram—*See* ciphertext.

cryptographic system—An algorithm for converting a message in ordinary language into a secret form.

cryptologic—Pertaining to communication intelligence and communication security.

cryptosecurity—Component of communication security that results from the provision of technically sound cryptosystems and their proper use.

crystal—1. A solid in which the constituent atoms are arranged with some degree of geometric regularity. In communication practice, a piezoelectric crystal, piezoelectric crystal plate, or crystal rectifier. 2. A thin slab or plate of quartz ground to a thickness that causes it to vibrate at a specific frequency when energy is supplied. It is used as a frequency-control element in radio-frequency oscillators. 3. Quartz crystal whose piezoelectric vibrational modes provide a highly accurate frequency for clock timing. 4. Solid material in which the atoms or molecules are arranged in a regular three-dimensional lattice array.

crystal anisotropy—A force that directs the magnetization of a single-domain particle along a direction of easy magnetization. To rotate the magnetization of the particle, an applied magnetic field must provide enough energy to rotate the magnetization through a difficult crystal direction.

crystal audio receiver—Similar to a crystal video receiver except for the path direction bandwidth, which is audio rather than video. *See* crystal video receiver.

crystal calibrator—A crystal-controlled oscillator used as a reference to check and set the frequency tuning of a receiver or transmitter.

crystal-can relay—A relay mounted in a can of a specific size and shape; called a crystal can because of its common usage as a mounting for quartz crystals used in frequency-control circuits.

crystal control—Control of the frequency of an oscillator by means of a specially designed and cut crystal.

crystal-controlled oscillator—*See* crystal oscillator.

crystal-controlled transmitter—A transmitter in which the carrier frequency is controlled directly by a crystal oscillator.

crystal counter—An instrument that is used to detect high-energy particles by the pulse of the current formed when a particle passes through a normally insulating crystal to which a potential difference is applied.

crystal cutter—A disc cutter in which the mechanical displacements of the recording stylus are derived from the deformities of a crystal having piezoelectric properties.

crystal detector—A mineral or crystalline material that allows electrical current to flow more easily in one direction than in the other. In this way, an alternating current can be converted to a pulsating current.

crystal diode—1. A two-electrode semiconductor device that makes use of the rectifying properties of a pn junction (junction diode) or a sharp metallic point in contact with a semiconductor material (point-contact diode). Also called crystal rectifier, diode, and semiconductor diode. 2. A diode rectifier using a point contact on a silicon or germanium crystal. Because of its low capacitance, it will rectify at ultrahigh and superhigh frequencies.

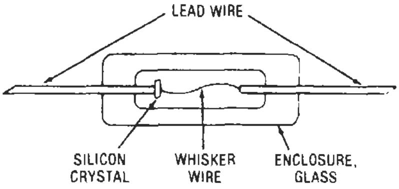

Crystal diode.

crystal field—The electrostatic field acting locally within a crystal as a result of the microscopic arrangement of atoms and ions in the lattice.

crystal filter—1. A highly selective circuit capable of discriminating against all signals except those at the center frequency of a crystal, which serves as the selective element. Resonant mechanical section consists of ceramic discs that vibrate at the band of frequencies to be removed. 2. Electrically coupled, two-terminal

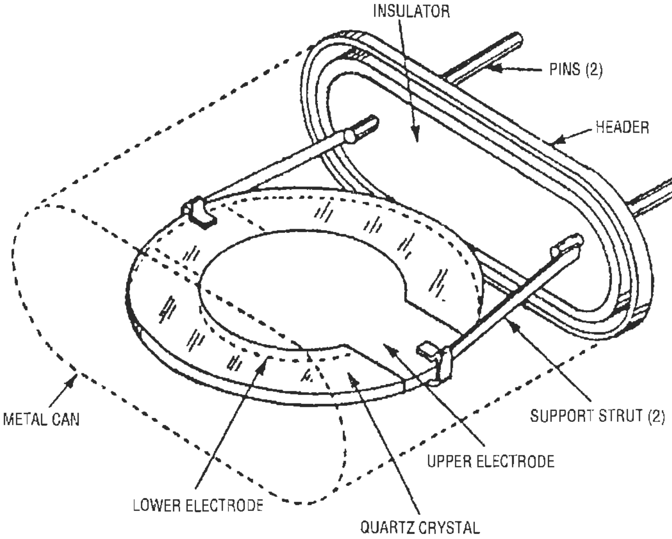

Crystal holder.

electroacoustic resonators (crystals) in ladder and lattice configurations. Monolithic filters with quartz substrates are also called crystal filters. 3. A bandpass filter with piezoelectric crystal components for the passage or impedance of electrical signals of various frequencies.

crystal headphones — Headphones using Rochelle-salt or other crystal elements to convert audio-frequency signals into sound waves.

crystal holder — A case of insulating material for mounting a crystal. External prongs allow the crystal to be plugged into a suitable socket.

crystal imperfection — Any deviation in lattice structure from that of a perfect single crystal.

crystal lattice — A periodic geometric arrangement of points that correspond to the locations of the atoms in a perfect crystal.

crystal loudspeaker — A loudspeaker in which piezoelectric action is used to produce mechanical displacements. Also called piezoelectric loudspeaker.

crystal microphone — Also called piezoelectric microphone. 1. A microphone that depends for its operation on the generation of an electric charge by the deformation of a body (usually crystalline) having piezoelectric properties. 2. A microphone whose generating element is a crystal or ceramic element, which generates a voltage when bent or stressed.

crystal mixer — 1. A mixer circuit with the frequency of the local oscillator being controlled by a crystal. Normally used in superheterodyne radio receivers. 2. A mixer that utilizes the nonlinear characteristic of a crystal diode to mix two frequencies. Frequently used in radar receivers to convert the received radar signal to a lower intermediate-frequency signal by mixing it with a local-oscillator signal.

crystal operation — Operation using crystal-controlled oscillators.

crystal orientation — For MOS devices, the terms ⟨100⟩ and ⟨111⟩ are commonly used. This refers to the angle with respect to crystal facets at which the silicon crystal is sliced. Each has a direct effect on MOS transistor characteristics.

crystal oscillator — Also called crystal-controlled oscillator. An oscillator in which the frequency of oscillation is controlled by a piezoelectric crystal.

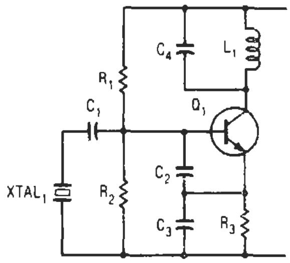

Crystal oscillator.

crystal oven — 1. A container, maintained at a constant temperature, in which a crystal and its holder are enclosed in order to keep their temperature constant and thereby reduce frequency drift. 2. A temperature-controlled container used to stabilize the temperature and resonant frequency of a crystal found in a crystal-controlled oscillator.

crystal pickup — Also called piezoelectric pickup. A phonograph pickup that depends for its operation on the generation of an electric charge by the deformation of a body (usually crystalline) having piezoelectric properties.

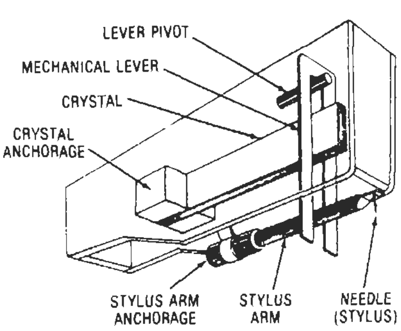

Crystal pickup.

crystal pulling—1. A method of growing crystals in which the developing crystal is gradually withdrawn from a melt. 2. A technique (first developed by J. C. Czochralski) in which a monocrystalline seed is introduced into the top of a body of molten material and then withdrawn slowly to grow, or pull, a large single monocrystal. This technique is used in semiconductor manufacture to produce the uniformly doped monocrystal boules from which most devices are fabricated.

crystal rectifier—1. An electrically conductive or semiconductive substance, natural or synthetic, that has the property of rectifying small radio-frequency voltages. *See* crystal diode.

crystal set—A simple type of radio receiver having no amplifier stages, and only a crystal-detector stage for demodulation of the received signal.

crystal shutter—A mechanical switch for shorting a waveguide or coaxial cable so that undesired rf energy is prevented from reaching and damaging a crystal detector.

crystal slab—A relatively thick piece of crystal from which crystal blanks are then cut.

crystal speaker—Also called piezoelectric speaker. A speaker in which the mechanical displacements are produced by piezoelectric action.

crystal-stabilized oscillator—Abbreviated CSO. A microwave rf source that uses a crystal oscillator operating at some low frequency (usually below 150 MHz) to drive a multiplier to obtain a microwave output frequency. A crystal oscillator can also be used for injection-locked stabilization of a free-running microwave oscillator.

crystal-stabilized transmitter—A transmitter employing automatic frequency control, in which the reference frequency is the same as the crystal-oscillator frequency.

crystal transducer—*See* piezoelectric transducer.

crystal video receiver—A radar receiver consisting only of a crystal detector and video amplifier.

CSA—Abbreviation for Canadian Standards Association, a testing and approval agency for products sold in Canada. This is the Canadian equivalent of Underwriters Laboratories in the United States.

C-scope—A rectangular radar display in which targets appear as bright spots with azimuth indicated by the horizontal coordinate and elevation angle by the vertical coordinate.

CSMA—Abbreviation for carrier sense multiple access. A contention scheme in which stations listen for the presence of a carrier on the channel before sending a packet.

csv—Abbreviation for corona start voltage.

CTCSS—Abbreviation for Continuous Tone-Coded Squelch System. Also called subaudible tones or PL tones (trademarked name by Motorola). This is a tone that is transmitted in addition to a voice signal. When it is equipped with a CTCSS decoder, a repeater will not function unless it receives both the CTCSS tone and the carrier signal.

CT-cut crystal—A natural quartz crystal cut to vibrate below 500 kHz.

CTL—Abbreviation for complementary transistor logic.

CTS—Abbreviation for clear to send. A control signal between a modem and a controller used to initiate data transmission over a communication line.

cubical antenna—An antenna array whose elements are positioned to form a cube.

cue—*See* address.

cue bus—In a mixing console, the bus or channel that is used to feed a program to a performer's headphones. Also called foldback.

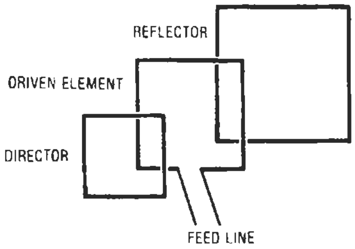

Cubical antenna.

cue circuit—A one-way communication circuit for conveying program control information.

cue control—A switch that temporarily disables a recorder's tape lifters during fast-forward and rewind, so the operator can judge what portion of the recording is passing the heads.

cuff electrode—An electrode in the shape of the letter "C" designed for application of potentials to small circular bodies, such as peripheral nerves.

cuing—1. Locating a particular spot on a recorded tape, preparatory to playing through from that spot. 2. The marking or other identification of particular points or sections of tape to aid in the location of specific, desired selections or portions of the recording. This may be done with grease pencil or other markings directly on the tape, or by making use of a digital index counter, if one exists on the recorder.

cuing device—A lever or control that raises and lowers the tonearm without direct handling by the operator. Usually viscous damped for uniform rise and fall times, no matter how rapidly the control is moved.

CUJT—Abbreviation for complementary unijunction transistor.

cup—A single mechanical section of a potentiometer that may contain one or more electrical resistance elements.

cup core—A core that forms a magnetic shield around an inductor. Usually a cylinder with one end closed. A center core inside the inductor is normally used and may or may not be part of the cup core.

cupping—Curvature of a recording tape in a direction perpendicular to the length of the tape.

cup washer—A washer formed with a recess in one side to retain compression springs or, on binding-post terminals, to prevent escape of connecting wire strands.

curie—A unit of radioactivity; 1 curie (Ci) equals 3.7×10^{10} nuclear transformations per second.

Curie point—In ferroelectric dielectrics, the temperature or temperatures at which peak values of the dielectric constant occur. Also called Curie temperature. The (critical) temperature at which piezoelectric materials lose their polarization and therefore also their piezoelectric properties.

Curie temperature—Temperature in degrees Celsius at which a magnetized sample is completely demagnetized due to thermal agitation. *See* Curie point.

curl—The degree to which a wire tends to form a circle after removal from a spool. An indication of the ability of the wire to be wrapped around posts in long runs.

curls—Extruded material coming out from the edge of a bond.

Curpistor—A subminiature constant-current tube containing two electrodes and filled with radioactive nitrogen.

current — 1. The movement of electrons through a conductor. Current is measured in amperes, milliamperes, microamperes, nanoamperes, or picoamperes. 2. A movement of electrons, positive ions, negative ions, or holes. 3. The rate of transfer of electricity from one point to another.

current amplification — 1. The ratio of the current produced in the output circuit of an amplifier to the current supplied to the input circuit. 2. In photomultipliers, the ratio of the signal output current to the photoelectric signal current from the photocathode.

current amplifier — 1. A device designed to deliver a greater output current than its input current. 2. An amplifier that has a low output impedance and is capable of delivering a heavy current.

current antinode — Also called current loop. The point at which current is a maximum along a transmission line, antenna, or other circuit element having standing waves.

current attenuation — The ratio of the magnitude of the current in the input circuit of a transducer to the magnitude of the current in a specified load impedance connected to the transducer.

current-balance relay — A relay in which operation occurs when the magnitude of one current exceeds the magnitude of another current by a predetermined ratio.

current-carrying capacity — 1. The maximum current a conductor (or braid) can carry without heating beyond a safe limit. 2. The maximum current that can be continuously carried without causing permanent change in the electrical or mechanical properties of a device or conductor. (As applied to phone jacks, it refers to carrying current without interrupting the circuit.) 3. The maximum current an insulated conductor can safely carry without exceeding its insulation and jacket temperature limitations. 4. The maximum current that can be carried continuously without damage to a device, conductor, or machine. It is mainly determined by the temperature which the insulation can withstand and by the ambient temperature.

current-carrying rating — The current that can be carried continuously or for stated periodic intervals without impairment of the contact structure or interrupting capability.

current-controlled oscillator — Abbreviated CCO. A circuit that creates an ac output signal whose frequency is a function of the dc input current.

current density — 1. The amount of electric current passing through a given cross-sectional area of a conductor in amperes per square inch; i.e., the ratio of the current in amperes to the cross-sectional area of the conductor. 2. The ratio of current to surface area.

current echo — The signal that on a transmission line is reflected as the result of some discontinuity.

current feed — The excitation or feeding of an antenna by connecting the feeder at a point of maximum current, as at the center of a dipole or half-wave antenna.

current flicker — Current surges resulting from momentary shorts that can occur within a solid electrolyte capacitor. Under certain conditions, current flicker can avalanche to cause a short, which under low-impedance circuit conditions results in catastrophic destruction of the capacitor.

current foldback — In a dc power supply, a self-resetting protective method that reduces the output current to less than full-rated output under overload or short-circuit conditions.

current generator — A two-terminal circuit element with a terminal current independent of the voltage between its terminals.

current hogging — 1. A condition in which one of several parallel logic circuits takes the largest share of the available current because it has a lower resistance than the other circuits. 2. A condition that exists when several base-emitter junctions are driven from the same output and the input with the lowest base-emitter junction forward potential severely limits the drive current to the other transistor bases.

current-hogging injection logic — Abbreviated CHIL. A logic form that combines the input flexibility of current-hogging logic with the performance and packing density of injection logic.

current limiter — A device that detects current leakage and prevents potential shock hazard by minimizing or interrupting current flow.

current-limiter relay — A relay that opens its contacts when the current in a circuit exceeds a certain preset value.

current limiting (automatic) — 1. An overload-protection mechanism that limits the maximum output current of a power supply to a preset value and automatically restores the output when the overload is removed. 2. In a dc power supply, a self-resetting safeguard that protects the power supply against short-circuit and overload conditions by limiting output current to a predetermined maximum value. *See also* short-circuit protection.

current-limiting fuse — A protective device that anticipates a dangerous short-circuit current and opens the circuit, precluding the development of the peak available current.

current-limiting reactor — A form of reactor intended for limiting the current that can flow in a circuit under short-circuit conditions.

current-limiting resistor — A resistor inserted into an electric circuit to limit the flow of current to some predetermined value. Usually inserted in series with a fuse or circuit breaker to limit the current flow during a short circuit or other fault, to prevent excessive current from damaging other parts of the circuit.

current-limit sense voltage — The voltage across the current-limit terminals required to cause a regulator to current limit with a short-circuited output. This voltage is used to determine the value of the external current-limit resistor when external booster transistors are used.

current loop — 1. Means of communicating data via the presence or absence of current in a two-wire cable. 2. A two-wire transmit/receive interface in which the presence of a 20-mA current level indicates data (a binary 1 or mark) and its absence indicates no data (a 0 or space). Normally used with teletypes, and the only communication method that uses a current signal. *See* current antinode.

current margin — The difference between the steady-state currents flowing through a telegraph receiving instrument that correspond to the two positions of the telegraph transmitter.

current mirror — A circuit that relies on the collector current matching of two transistors (one strapped as a diode) when connected together base to base and emitter to emitter.

current-mode logic — Abbreviated CML. A nonsaturating logic circuit that employs the characteristics of a differential amplifier circuit in its design. Because it is nonsaturating, it is a very fast switching logic design with low logic swings. The gate input element is the base of a transistor, with a separate transistor for each input.

current node — A point at which current is zero along a transmission line, antenna, or any other circuit element that has standing waves.

current noise — Also called excess noise. A low-frequency noise caused by current flowing in a resistor, particularly in film and carbon resistors. The amount of energy varies widely with the type and construction of the resistor. This low-frequency noise is generally measurable only in the region below 100 kHz; the noise power varies inversely with frequency and is a function of both the current in the resistor and the value of the resistor.

current penetration—The depth a current of a given frequency will penetrate into the surface of a conductor carrying the current.

current probe—A type of transformer, usually having a snap-around configuration, used for measuring the current in a conductor.

current pump—A circuit that drives, through an external load circuit, an adjustable, variable, or constant value of current regardless of the reaction of that load to the current, within rated limits of current, voltage, and load impedance.

current rating—1. The maximum continuous current that can be carried by a conductor without degradation of the conductor or its insulation properties. 2. Maximum current that a device is designed to conduct for a specified time at a specified operating temperature. 3. The maximum continuous electrical current recommended for a given wire in a given situation. Expressed in amperes.

current regulator—1. A device that functions to maintain the output current of a generator or other voltage source at a predetermined value, or varies the voltage according to a predetermined plan. 2. A regulating device that limits generator output and prevents excessive generator output by repeatedly inserting a resistance into the generator field circuit.

current relay—A relay that operates at a predetermined value of current. It can be an overcurrent relay, an undercurrent relay, or a combination of both.

current saturation—The condition in which the plate current of a vacuum tube cannot be further increased by increasing the plate voltage.

current-sensing resistor—A resistor of low value placed in series with a load to develop a voltage proportional to the output current. A regulated dc power supply regulates the current in the load by regulating the voltage across this sensing resistor.

current-sensitivity—The current required to give standard deflection on a galvanometer.

current sink—1. A point toward which conventional current flows (electrons flow away from it). 2. An output type of sensor or analog device in which current flows from the load and into the output of the device at a low voltage when it is turned on.

current-sinking logic—Also called input-coupled logic. A logic form that requires that current flow out of the input of a circuit and back into the output of the preceding stage, which serves as a current sink instead of a source.

current source—1. A point from which conventional current flows (electrons flow toward it). 2. An output type of switch or analog device in which current flows from it into the load at a high voltage when it is turned on.

current-sourcing logic—Also called collector-coupled logic. A logic form in which current flows from the output of a circuit and is forced into the input of a similar circuit to activate the circuit that drives.

current-stability factor—In a transistor, the ratio of a change in emitter current to a change in reverse-bias current between the collector and base.

current-transfer ratio—The ratio of an optocoupler's output current to input current at a specified bias. *See* beta.

current transformer—A transformer, intended for measuring or control purposes, designed to have its primary winding connected in series with a circuit carrying the current to be measured or controlled.

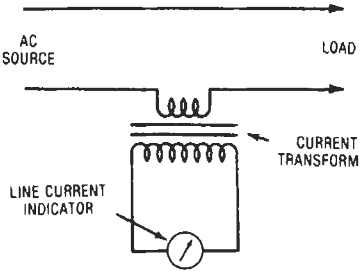

Current transformer.

current-type telemeter—A telemeter in which the magnitude of a single current is the translating means.

cursor—1. A mechanically or electronically generated line that moves back and forth over another surface to delineate accurate readings. 2. A mechanical bearing line on a plan position indicator-type display for reading target bearing. 3. A visual movable pointer on a CRT used by the PC programmer to indicate where an instruction is to be added to the PC program. The cursor is also used during editing functions. 4. A small lighted line or square that appears on a video display screen at the point where the next character will appear. 5. A manually movable illuminated or flashing marker on a computer screen used to indicate the location of the next point at which an action will take place. 6. A means for indicating on a CRT screen the point at which data entry or editing will occur. The intensified element may be at constant high intensity or flashing (alternate high intensity and normal intensity). If flashing, additional data may be necessary to complete the instruction. 7. An indicator that marks the current working position on a display. 8. A blinking character that indicates where the next keyboard stroke will appear. 9. A position marker on a video display.

cursor and update—Circuitry that allows a user to add new material into a video terminal's memory and display. The cursor, which locates the current entry location, is generally a blinking underline, although some are overlines or boxes.

curve tracer—An instrument capable of producing a display of one current or voltage as a function of a second voltage or current, with a third voltage or current as a parameter.

customer set—*See* subscriber set.

cut and paste—To move graphics and/or text from one location to another.

Cutler antenna—A rear feed for a paraboloidal antenna reflector. It consists of a support waveguide with a terminating cavity containing two resonant slots, one on either side of the support waveguide, that face the reflector. Each slot is parallel to the broad faces of the feed waveguide.

Cutler feed—A resonant cavity, at the end of a waveguide, that feeds radio-frequency energy to the reflector of the antenna assembly of some airborne antennas.

cutoff—1. Minimum value of bias that cuts off, or stops, the flow of plate current in a tube. 2. The frequency above or below which a selective circuit fails to respond. 3. The frequency of transmission at which the loss exceeds by 10 decibels that observed at 1000 hertz. 4. The condition when the emitter-base junction of a transistor has zero bias or is reverse biased and there is no collector current. 4. The frequency at which the modulus of measured parameter has decreased to $1/\sqrt{2}$ of its low-frequency value. (For a transistor, the cutoff frequency usually applies to the short-circuit small-signal forward current transfer ratio for either the common-base or common-emitter configuration.)

cutoff attenuator—A variable length of waveguide used below the cutoff frequency of the waveguide to introduce variable nondissipative attenuation.

cutoff current—Transistor collector current with no emitter current and normal collector-to-base bias.

cutoff field—*See* critical field.

cutoff frequency—1. The frequency at which the gain of an amplifier falls below 0.707 times the maximum gain. 2. The frequency that marks the edge of the passband of a filter and the beginning of the transition to the stopband. 3. With respect to a line, the upper frequency limit, usually of a loaded transmission circuit, beyond which the attenuation rises very rapidly. 4. That frequency beyond which no appreciable energy is transmitted. It may refer to either an upper or lower limit of a frequency band.

cutoff limiting—Keeping the output of a vacuum tube below a certain point by driving the control grid beyond cutoff.

cutoff voltage—1. The electrode voltage that reduces the dependent variable of an electron-tube characteristic to a specified value. *See also* critical voltage. 2. The voltage at which the discharge is considered complete. This need not be a very low voltage.

cutoff wavelength—The ratio of the velocity of electromagnetic waves in free space to the cutoff frequency of a waveguide.

cutout—1. An electrical device that interrupts the flow of current through any particular apparatus or instrument, either automatically or manually. 2. Pairs brought out of a cable and terminated at some place other than at the end of the cable.

cutout relay—A device in the circuit between a generator and battery that closes to allow the generator to charge the battery and opens when the generator stops.

cut over—To transfer from one system to another.

CUTS—Abbreviation for computer user tape system. A standard method of recording data in serial form on an audiocassette recorder. Data is recorded at 300 baud by recording 8 pulses at 2400 Hz for a mark, or 4 pulses of 1200 Hz for a space.

cut-signal branch operation—In systems in which radio reception continues without cutting off the carrier, the cut-signal branch operation technique disables a signal branch in one direction when it is enabled in the other to preclude unwanted signal reflections.

cutter—Also called mechanical recording head. An electromechanical transducer that transforms an electrical input into a mechanical output (for example, the mechanical motion that a cutting stylus inscribes into a recording medium).

cut-through—The resistance of a solid material to penetration by an object under conditions of pressure, temperature, etc.

cut-through flow test—A test to measure the resistance to deformation of insulation subjected to heat and pressure.

cutting rate—The number of lines per inch the lead screw moves the cutting-head carriage across the face of a recording blank. Standard rates are 96, 104, 112, 128, 136, and greater, in multiples of 8 lines per inch. For microgroove recordings, 200 to 300 lines per inch are used.

cutting stylus—A special stylus used for cutting of phonograph records in the disc mastering process; it often has a built-in heater element and has special geometry.

CVD—Abbreviation for chemical vapor deposition.

CVD fiber—A process by which a heated gas produces an oxide deposit to fabricate a glass fiber preform. The deposited glass becomes the core.

cw—1 Abbreviation for continuous wave. 2. Abbreviation for clockwise.

cw jamming—Transmission of a constant-amplitude, constant-frequency, unmodulated signal for the purpose of jamming a radar receiver by changing its gain characteristics.

cwp—Abbreviation for communication word processor. A word processor that can communicate, over special lines or regular telephone lines, with another cwp and with mainframe computers, Telexes and TWXs, photocomposers, optical character reading devices, intelligent copiers, and other terminals. The chief limitation is that the communicating systems must be compatible, meaning they must speak the same electronic language in order to understand each other.

cw reference signal—In color television, a sinusoidal signal used to control the conduction time of a synchronous demodulator.

cxr—Abbreviation for carrier. A communication signal used to indicate the intention to transmit data on a line.

cybernetics—1. The study of systems of control and communications in humans and animals, and in electrically operated devices such as calculating machines. 2. The comparative study of the control and intracommunication of information-handling machines and nervous systems of animals and man in order to understand and improve communication. 3. The science that is concerned with the principles of communication and control, particularly as applied to the operation of machines and the functioning of organisms. 4. Study of multiple feedback loop, self-governing systems, usually of great complexity, as are found in living organisms and advanced human-made control systems. 5. A comparative study of the similarities as well as the differences between humans and machines, with respect to their ability to communicate and control.

cyberspace—1. A general term referring to the virtual places that exist only on the Internet. 2. A term coined by William Gibson in his 1984 novel *Neuromancer* to describe a shared virtual environment whose inhabitants, objects, and spaces comprise data that is visualized, heard, and touched. The word *cyberspace* is currently used to describe the whole range of information resources available through computer networks and is often used as a synonym for the Internet. 3. The metaphoric space in which electronic communication takes place. Everything in cyberspace is virtual, that is, not physically real, but a shared experience nonetheless. 4. The "place" one goes to use a computer and modem to communicate with others. Being online puts one in cyberspace.

cyborg—Abbreviation for cybernetic organism. 1. An android capable of heuristic (learning by experience) updating of its own resident intelligence. 2. An organism superior in strength and/or perception to ordinary human beings. 3. A human altered or enhanced by mechanical means—somewhere between a man and a robot. The Borg in Star Trek are the best-known example.

cycle—1. The change of an alternating wave from zero to a negative peak to zero to a positive peak and back to zero. The number of cycles per second (hertz) is called the frequency. *See also* alternation. 2. An off-on application of power. 3. A set of operations carried out in a predetermined manner. 4. In computer terminology, a regularly repeated sequence of operations, or the time required for one such sequence. 5. The complete sequence, including reversal, of an alternating electric current. 6. An interval of time in which an operation or set of events is completed.

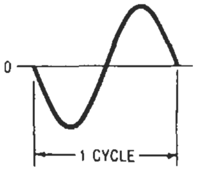

Cycle, 1.

cycle counter—A mechanism or device used to record the number of times a specified cycle is repeated.

cycle criterion—In computer terminology, the number of times a cycle is to be repeated.

cycle index—1. In digital computer programming, the number of times a cycle has been executed. 2. The difference between the number of times a cycle has been executed and the number of times it is desired that the cycle be repeated.

cycle life—For rechargeable cells, the total number of discharge/charge cycles before the cell becomes inoperative.

cycle redundancy check—Abbreviated CRC. 1. An error detection scheme, usually hardware implemented, in which a check character is generated by taking the remainder after dividing all the serialized bits in a block of data by a predetermined binary number. This remainder is then appended to the transmitted data and recalculated and compared at the receiving point to verify data accuracy. 2. Comparison of the checksum derived from data as it was originally written into storage with the checksum derived from the same data as it is being read out of storage. The first checksum is appended to the data as it enters storage. After reading this data, the controller computes a new checksum from it and compares the two. If the checksums match, the data is correct. A checksum error may indicate a damaged area on the memory, data that has changed since written, or erroneous reading of correct data for which a retry may work.

cycle reset—To return a cycle index to its initial value.

cycle shift—In a computer, the removal of the digits of a number of characters from a word from one end of the number or word and their insertion, in the same sequence, at the other end.

cycle stealing—1. A memory cycle stolen from the normal CPU operation for a direct memory access (DMA) operation. 2. A computer memory access technique in which an I/O processor is synchronized to the general processor's memory cycles in such a way that it steals memory cycles between those of the general processor. This is possible on systems in which the memory cycle time is significantly shorter than the processor execution cycle. 3. A characteristic of DMA channels. An I/O device can delay CPU use of the I/O bus for one or more bus cycles while it accesses system memory.

cycle time—1. The interval of time between the occurrence of corresponding parts of successive cycles. 2. The length of time required to obtain information from a memory and then write information back into the memory. Also called read-write cycle time, since it is normally equal to the sum of the write time and the read time. If system memory is core, the read cycle time includes a write-after-read (restore) subcycle. Cycle time is often used as a measure of computer performance, since this is a measure of the time required to fetch an instruction. 3. Time interval during which any set of operations is repeated regularly in the same sequence. 4. The time needed for a CPU to go through a complete operation cycle. 5. The total time required by a device (usually a memory) to complete its cycle and become available again. Typically, the access time will be shorter than the cycle time, though they may sometimes be equal.

cycle timer—A controlling mechanism that opens or closes contacts according to a preset cycle.

cyclically magnetized condition—The condition of a magnetic material after being under the influence of a magnetizing force varying between two specific limits until, for each increasing (or decreasing) value of the magnetizing force, the magnetic-flux density has the same value in successive cycles.

cyclic code—Positional notation, not necessarily binary, in which quantities differing by one unit are represented by expressions that are identical except for one place or column, and the digits in that place or column differ by only one unit. Cyclic codes are often used in mechanical devices because no ambiguity exists at the changeover point between adjacent quantities.

cyclic decimal code—A four-bit binary code word only one digit of which changes state between one sequential code word and the next, and which translates to decimal numbers. It is categorized as one of a group of unit-distance codes.

cyclic memory—A memory that continuously stores information but provides access to any piece of stored information only at multiples of a fixed time called the cycle time.

cyclic shift—A shift in which the data moved out of one end of the storing register is reentered into the other end, as in a closed loop.

cycling—1. A rhythmic variation, near the desired value, of the factor under control. 2. A periodic change from one value to another of the controlled variable in an automatic control system. 3. A periodic oscillation in an automatically controlled system between the high limit and the low limit over which the controls operate.

cycling vibration—Sinusoidal vibration applied to an instrument and varied in such a way that the instrument is subjected to a specified range of vibrational frequencies.

cycloconverter—A step-down static frequency converter that produces a constant or a precisely controllable output frequency from a variable-frequency ac power input. In general, the frequency ratio chosen is 3 to 1 or greater.

cyclogram—An oscilloscope display obtained by monitoring two voltages having a direct cyclic relationship to each other.

cyclograph—A device in which an electron beam moves in two directions, at right angles.

cyclometer register—A set of four or five wheels numbered from 0 to 9 inclusive on their edges, and enclosed and connected by gearing so that the register reading appears as a series of adjacent digits.

cyclotron—1. A device consisting of an evacuated tank in which positively charged particles (for example, protons, deuterons) are guided in spiral paths by a static magnetic field while being highly accelerated by a fixed-frequency electric field. 2. Type of accelerator of nuclear particles (protons or deuterons) that uses an oscillating electric field and a fixed magnetic field to accelerate the particles. *See* accelerator.

cyclotron frequency—The frequency at which an electron traverses an orbit in a steady, uniform magnetic field and zero electric field. Given by the product of the electronic charge and the magnetic flux density, divided by 2π times the electron mass.

cyclotron-frequency magnetron oscillations—Those oscillations having substantially the same frequency as that of the cyclotron.

cyclotron radiation—The electromagnetic radiation emitted by charged particles orbiting in a magnetic field at speeds relatively slow compared with the speed of light. It arises from the centripetal acceleration of the particle moving in a circular orbit, as in a cyclotron.

cyclotron resonance—The effect characterized by the tendency of charged carriers to spiral around an axis in the same direction as an applied magnetic field, with an angular frequency determined by the value of the applied field and the ratio of the charge to the effective mass of the charge carrier.

cylindrical (or **circular**) **connectors**—Separable plugs and receptacles, both housed in cylindrically shaped metal or plastic shells. Shells may stand alone or may be mounted on panels, bulkheads, or walls. When mated, the shells are joined by a coupling ring of threaded, bayonet, or push-pull design. The receptacle (female) contacts are imbedded in an insert of plastic or vitreous insulating material (usually for hermetically sealed connectors) in the shell. The plug (male) contacts are cantilevered cylindrical pins, also partly imbedded in an insert of insulating material. Contact terminations are either soldered or crimp connected to lead-in wires.

cylindrical-film storage—A computer storage device, each storage element of which is a short length of glass tubing with a thin film of nickel-iron alloy on its outer surface. Wires running through the tubing act as bit and sense lines, and conducting straps at right angles to the tubing function as word lines.

cylindrical reflector—A reflector that is part of a cylinder, usually parabolic.

cylindrical wave—A wave whose equiphase surfaces form a family of coaxial cylinders.

Czochralski crystal—A crystal grown by slowly withdrawing a seed crystal from a melt while the melt is held slightly above the melting point of the material. High-quality, low-dislocation-density crystals of germanium and silicon are grown in this manner.

Czochralski technique—A method of growing large, single crystals by pulling them from a molten state. Usually used for growing single crystals of germanium and silicon.

D

D—Symbol for electrostatic flux density, deuterium, dissipation factor, or drain electrode.

DAA—Abbreviation for data access arrangement. 1. A direct interface attachment that connects a data communications device to a telephone line. It protects the public telephone network from a sudden surge of power or interference from the device that is coupled to the line. 2. A telephone-switching-system protective device used to attach uncertified non-telephone-company-manufactured equipment to the carrier network.

DABS—Abbreviation for discrete address beacon system. A sophisticated ground surveillance equipment and complementary airborne transponder. DABS includes a two-way automatic ground-to-air data link.

DAC—Abbreviation for digital-to-analog converter. Also abbreviated dac or d/a converter.

d/a converter—*See* digital-to-analog converter.

d/a decoder—A device that changes a digital word to an equivalent analog value.

Dag—Trademark of Acheson Industries, Inc. Abbreviation for Aquadag.

daisy chain—1. In a computer, a bus line that is interconnected with units in such a way that the signal passes from one unit to the next in serial fashion. 2. A method for prioritizing interrupts. Each unit capable of requesting an interrupt can either pass on the processor acknowledge or block it. In this way the unit that is electrically closest to the processor has highest priority. 3. A method of propagating signals along a bus, often used in applications in which devices not requesting a daisy-chained signal respond by passing the signal on. The first device requesting the signal responds to it by performing an action and breaks the daisy-chained signal continuity. This scheme permits assignment of device priorities based on the electrical position of the device along the bus.

daisy wheel—1. A letter-quality, impact-printing mechanism whose printing element consists of a flat, rimless, metal or plastic wheel with letters molded at the ends of the spokes. The wheel rotates until the required character is in position, and a hammer flies out and hits it, impacting the character onto paper. 2. The print element for a daisy-wheel printer. Print wheels are interchangeable, allowing the operator to select an appropriate font.

daisy-wheel printer—1. An impact printer that prints fully formed characters one at a time by rotating a circular print element composed of a series of individual spokes, each containing two characters. 2. A printer that produces letter-quality type when characters, mounted in a circle, strike an inked ribbon. It is rarely used anymore.

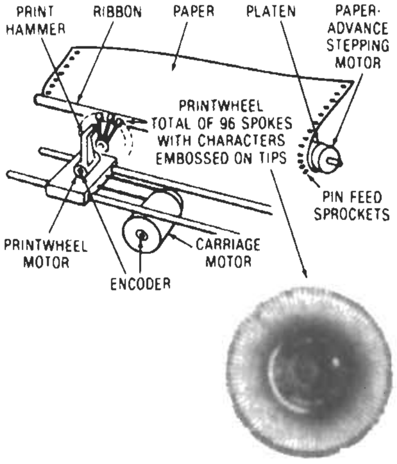
PRINT HAMMER
RIBBON
PAPER
PLATEN
PAPER-ADVANCE STEPPING MOTOR
PRINTWHEEL TOTAL OF 96 SPOKES WITH CHARACTERS EMBOSSED ON TIPS
PIN FEED SPROCKETS
PRINTWHEEL MOTOR
CARRIAGE MOTOR
ENCODER

Daisy-wheel printer.

Damon effect—The change in susceptibility of a ferrite caused by a high rf power input.

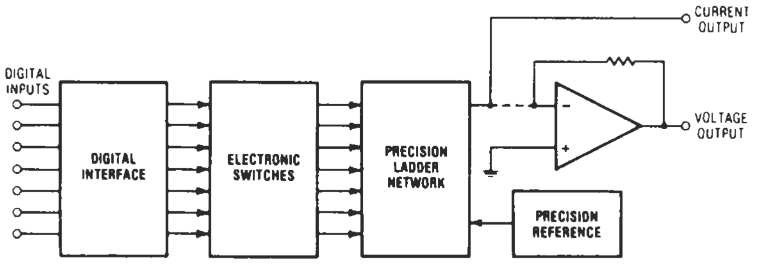
DIGITAL INPUTS
DIGITAL INTERFACE
ELECTRONIC SWITCHES
PRECISION LADDER NETWORK
PRECISION REFERENCE
CURRENT OUTPUT
VOLTAGE OUTPUT
-
+

d/a decoder.

damped natural frequency — 1. The frequency at which a system with a single degree of freedom will oscillate, in the presence of damping, after momentary displacement from the rest position by a transient force. 2. The rate of free oscillation of a sensing element in the presence of damping.

damped oscillation — The oscillation that occurs when the amplitude of the oscillating quantity decreases with time. If the rate of decrease can be expressed mathematically, the name of the function describes the damping. Thus, if the rate of decrease is expressed as a negative exponential, the system is said to be an exponentially damped system.

damped waves — Waves in which successive cycles at the source progressively diminish in amplitude.

dampen — 1. To diminish progressively in amplitude; usually said of waves or oscillations. 2. To deaden vibrations.

damper diode — *See* freewheeling diode.

damper tube — The tube that conducts in the horizontal-output circuit of a television receiver when the current in the horizontal-deflecting yoke reaches its negative peak. This causes the sawtooth deflection current to decrease smoothly to zero instead of continuing to oscillator.

damper winding — Also called amortisseur winding. 1. In electric motors, a permanently short-circuited winding, usually uninsulated, arranged so that it opposes rotation or pulsation of the magnetic field with respect to the pole shoes. 2. A winding of copper bars or rods (squirrel cage) embedded in the pole face of synchronous motors and generators. Used as a starting winding on synchronous motors; acts as the squirrel cage of an induction motor. After the synchronous motor is up to speed, tends to prevent oscillations and hunting — a damping effect that gives it its name.

damping — 1. The reduction of energy in a mechanical or electrical oscillating system by absorption, conversion into heat, or by radiation. 2. Act of reducing the amplitude of the oscillations of an oscillatory system, hindering or preventing oscillation or vibration, or diminishing the sharpness of resonance of the natural frequency of a system. 3. The dissipation of kinetic energy in a system by a controlled energy-absorbing medium. A system can be described as being either critically damped, overdamped, or underdamped. 4. The manner in which the pointer settles to its steady indication after a change in the value of the measured quantity. Two general classes of damped motion are (*a*) periodic, in which the pointer oscillates about the final position before coming to rest; and (*b*) aperiodic, in which the pointer comes to rest without overshooting the rest position. Sometimes referred to as overdamping. 5. The energy-dissipating characteristic that, together with natural frequency, determines the upper limit of frequency response and the response time characteristics of a transducer. 6. The application of a mechanical resistance, such as a rubber or silicone material, to the cantilever pivot of a pickup to reduce the amplitude of the resonance between the tip mass and the compliance of the vinyl record material (which usually occurs between 15 and 30 kHz). 7. The suppression of oscillations, or ringing, in the rotary of a stepping motor caused by sudden changes in velocity. Damping can be accomplished by adding mechanical or viscous fluid dampers or by electronically shaping and controlling the energizing field power.

damping coefficient — The ratio of actual damping to critical damping.

damping constant — The Napierian logarithm of the ratio of the first to the second of two values of an exponentially decreasing quantity separated by a unit of time.

damping diode — A vacuum tube that damps the positive or negative half-cycle of an ac voltage.

damping factor — 1. For any underdamped motion during any complete oscillation, the quotient obtained by dividing the logarithmic decrement by the time required by the oscillation. 2. Numerical quantity indicating ability of an amplifier to operate a speaker properly. Values over 4 are usually considered satisfactory. 3. The ratio of rated load impedance to the internal impedance of an amplifier. 4. The ratio (larger to smaller) of the angular deviations of the pointer of an electrical indicating instrument on two consecutive swings from the equilibrium position. 5. *See* decrement, 1. 6. The ratio of load or speaker impedance to the amplifier's output impedance. Thus, the smaller the output impedance the greater the damping factor. The damping factor increases with increase of voltage negative feedback; with the large amounts of feedback applied to transistor hi-fi amplifiers, the source or output impedance can be as low as 0.1 ohm, giving a damping factor of 80, referred to as an 8-ohm load. 7. The standard damping constant of a second-order feedback system. In a phase-locked loop, it refers to the ability of the loop to respond quickly to an input signal frequency without excessive overshooting.

damping magnet — A permanent magnet and a movable conductor, such as a sector or disc, arranged in such a way that a torque (or force) is produced that tends to oppose any relative motion.

damping ratio — 1. The ratio of the degree of actual damping to the degree of damping required for critical damping. May be affected by changes in ambient temperature. 2. Of a galvanometer, the ratio, expressed as a positive number, of a given deflection to the next deflection in the opposite direction. The greater this ratio, the greater the degree of damping. The natural logarithm of this ratio is called the logarithmic decrement.

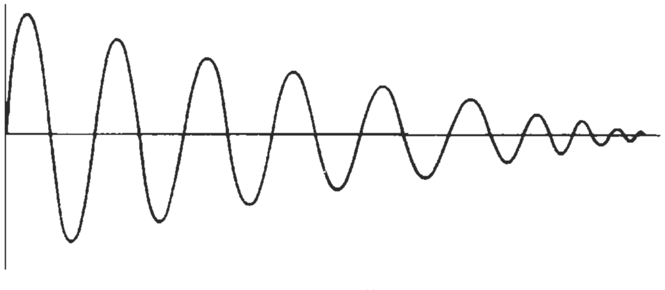

Damped oscillation.

dancer arm—A device that senses tape tension and signals the reel motor to take up or to supply tape.

Daniell cell—A cell having a copper electrode in a copper-sulfate solution and a zinc electrode in a diluted sulfuric acid or zinc-sulfate solution, with the two solutions separated by a porous partition. Generates an essentially constant electromotive force of about 1.1 volts.

daraf—The unit of elastance. It equals the reciprocal of capacitance and is actually farad spelled backward.

dark conduction—Residual electrical conduction in a photosensitive substance in total darkness.

dark current—*See* electrode dark current.

dark discharge—In a gas, an electric discharge that has no luminosity.

dark-field disc—A disc used in the optical electronic type of cell counter that controls light transmission.

dark noise—The current or pulses produced when the photocathode is shielded from all external optical radiation.

dark resistance—The resistance of a photoelectric device in total darkness.

dark satellite—A satellite that does not give information to friendly ground stations, either because it is controlled or because it carries inoperative radiating equipment.

dark space—A nonluminous region of a glow-discharge tube.

dark spot—A phenomenon sometimes observed in a reproduced television image. It is caused by the formation of electron clouds in front of the mosaic screen in the transmitter camera tube.

dark-spot signal—The signal existing in a television system while the television camera is scanning a dark spot.

dark-trace tube—A cathode-ray tube with a screen coated with a halide of sodium or potassium. The screen normally is nearly white, and whenever the electron beam strikes, it turns a magenta color that is of long persistence. The screen can be illuminated by a strong light source so that the reflected image may be made intense enough to be projected.

Darlington amplifier—Also called Darlington pair, double-emitter follower, or β multiplier. 1. A transistor circuit that, in its original form, consists of two transistors in which the collectors are tied together and the emitter of the first transistor is directly coupled to the base of the second transistor. Therefore, the emitter current of the first transistor equals the base current of the second transistor. This connection of two transistors can be regarded as a compound transistor with three terminals. 2. A two-transistor amplifier connected so that the amplification of the amplifier equals the product of the individual transistors' amplification. 3. A composite configuration of transistors that provides a high input impedance and a high degree of amplification.

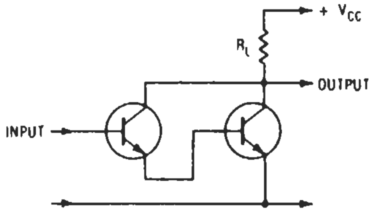

Darlington amplifier.

Darlington-connected phototransistor—A phototransistor whose collector and emitter are connected to the collector and base, respectively, of a second transistor. The emitter current of the input transistor is amplified by the second transistor, and the device has a very high sensitivity to light.

Darlington connection—A form of compound connection in which the collectors of two or more transistors are connected together and the emitter of one is connected to the base of the next. Two transistors connected in this way constitute a Darlington pair.

Darlington pair—*See* Darlington amplifier.

Darlington transistor—A three terminal device that consists of two transistors connected so that the emitter of the first transistor is connected to the base of the second transistor. Such a direct coupled configuration has much higher current gain than can be achieved by a single transistor.

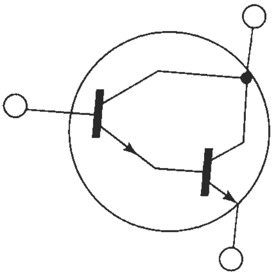

Darlington transistor.

D'Arsonval current—A high-frequency, low-voltage current of comparatively high amperage.

D'Arsonval galvanometer—A dc galvanometer consisting of a narrow rectangular coil suspended between the poles of a permanent magnet.

D'Arsonval instrument—*See* permanent-magnet moving-coil instrument.

D'Arsonval movement—A meter movement consisting essentially of a small, lightweight coil of wire supported on jeweled bearings between the poles of a permanent magnet. When the direct current to be measured is sent through the coil, its magnetic field interacts with that of the permanent magnet and causes the coil and attached pointer to rotate.

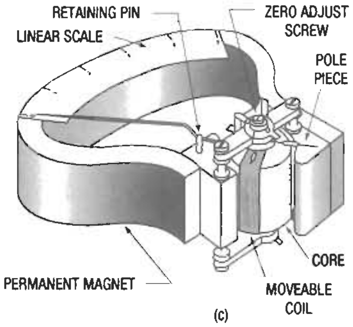

D'Arsonval movement.

DASD — Abbreviation for direct-access storage device. Any storage device utilizing addressing to let users enter or retrieve data without reference to their physical location. For example, a RAM.

dash — Term used in radiotelegraphy. It consists of three units of a sustained transmitted signal followed by one unit during which no signal is transmitted.

dashpot — 1. A device using a gas or liquid to absorb energy from or retard the movement of the moving parts of a circuit breaker or other electrical or mechanical device. 2. A cylinder and piston device using gas or a liquid to retard the movement of a relay or circuit breaker.

DAT — *See* digital audio tape.

data — 1. A general term used to denote any or all numbers, letters, symbols, or facts that refer to or describe an object, idea, condition, situation, or other factors. It connotes basic elements of information that can be processed or produced by a computer. Sometimes *data* is considered to be expressible only in numerical form, but *information* is not so limited. 2. A general term for any type of information. 3. Inputs in the form of a character string that may have significance beyond their numerical meaning. 4. Any representations, such as characters or analog quantities, to which meaning might be assigned.

data access arrangement — A protective connecting arrangement that serves as an interface between a customer-provided modem and the switched network. *See* DAA.

data acquisition — 1. The process by which events in the real world are translated to machine-readable signals. The term usually refers to automated systems in which sensors of one type or another are attached to machinery. 2. The simultaneous collection of data from external sensors, usually in analog form. 3. The function of obtaining data from sources external to a computer system, converting it to binary form, and processing it.

data acquisition and control systems — Assemblies of electronic and mechanical components used to monitor and control complex processes. These systems include the following:

- Process sensors that measure such parameters as temperature, pressure, voltage, and current
- Transmitters that convert measurement data to electrical or pneumatic signals and controls
- Digital computers that test set points, program sequential events, and perform calculations
- Software that provides the computer with instructions and routines
- Process actuators, such as solenoids, relays, valves, and motors, that modify the process in response to computer-generated commands
- Process interface devices, such as analog-to-digital converters, that link transmitters and actuators with digital computers
- Human interface devices, such as printers, keyboards, CRT terminals, switches, alphanumeric displays, chart recorders, and alarms, that facilitate human intervention

data acquisition and conversion system — A method of processing analog signals and converting them into digital form for subsequent processing or analysis by computer or for data transmission.

data acquisition system — 1. A system in which a computer at a central computing facility gathers data from multiple remote locations. 2. System for recording data, usually in digital form, from several sources; can include computing functions.

data bank — A comprehensive collection of libraries of data. For example, one line of an invoice may form an item, a complete invoice may form a record, a complete set of such records may form a file, the collection of inventory control files may form a library, and the libraries used by an organization are known as its data bank. Synonymous with database.

database — Also data base. 1. The entire body of data that has to do with one or more related subjects. Typically, it consists of a collection of data files (such as a company's complete personnel records concerning payroll, job history, accrued vacation time, etc.) stored in a computer system so that they are readily available. 2. A block of computer memory containing information about one given thing. 3. The collection of current variable data elements defined and maintained by the user. 4. A collection of data, consisting of at least one file, that is sufficient for a given purpose or for a given data-processing system. 5. A large and complete collection of information that covers a variety of subject areas. For instance, a medical diagnostic database might contain symptoms for all common diseases or injuries. 6. A collection of data fundamental to a system or to an enterprise. Made up of comprehensive files of information having predetermined structure and organization and able to be communicated, interpreted, or processed by humans or by automatic means. 7. A collection of related data that can be retrieved from memory at will, such as a mailing list or a list of accounts.

database management — 1. A systematic approach to the storage, updating, and retrieval of information stored as data items, usually in the form of records in a file, where many users, or even many remote installations, will use common data banks. 2. A program that enables a computer to store large amounts of information and then sort it in almost any manner. For example, a company's database could give a list of customers by ZIP code, by credit line, alphabetically by name, or by telephone number. The program takes care of managing the storage and retrieval of the data.

database management system — Abbreviated DBMS. A group of programs that allow users to store, alter, and retreive information from a database.

database relations — Linkages within a database that logically bind two or more elements in the database. For example, a nodal line (interconnect) is related to its terminal connection nodes (pins) because they all belong to the same electrical net.

data block — Typically, all the data for one item that is entered into a computer for processing, or the computer output that results from processing. An example of an input data block is an individual shipping list; an example of an output data block is a check to be sent.

data break — A facility that permits input/output transfers to take place on a cycle-stealing basis without disturbing execution of the program by a computer.

data bus — 1. A wire or group of wires used to carry data to or from a number of different locations. 2. The output pins of the MPU chip and associated circuitry used for the transmission of data from one point in the system to another. 3. In fiber optics, an optical waveguide used as a common trunk line to which a number of terminals can be interconnected through optical couplers. 4. A system incorporated into fiber-optic communications characterized by several spatially distributed terminals that are served with the same multiplexed signal.

data catalog — A software tool used to list all of the data elements in a database.

data channel (or communication) equipment — Abbreviated DCE. Equipment that interfaces a transmission facility to a transmitting/receiving device. A modem is a DCE.

data code — A structured set of characters used to stand for the data items of a data element, for example, the numerals 1, 2, ... 7 used to represent the data items Sunday, Monday, ... Saturday.

data collection — Also called data gathering. In a computer, the transferring of data from one or more points to a central point.

data communication equipment — *See* DCE.

data communications — 1. The technology covering the transfer of data over relatively long distances. 2. Transmission of data in both directions between a central location (host computer) and remote locations (terminals) through communication lines. To facilitate this, interfaces such as modems, multiplexers, concentrators, etc., are required at each end of the lines. 3. The movement of encoded information by means of electrical transmission systems.

data communications processor — A small computer used to control the flow of data between machines and terminals over communications channels. It may perform the functions of a concentrator, handshaking, and formatting, but does not include long-term memory or arithmetic functions.

data compression — 1. The process of reducing the number of recorded or transmitted digital data samples through the exclusion of redundant samples or the use of variable-length characters. 2. A technique that provides for the transmission of fewer data bits than originally required without information loss. The receiving location expands the received data bits into the original bit sequence. 3. A method of reducing the number of bits that are needed to represent information. Data compression allows higher communications speeds and also allows more information to be stored on a disk.

data control block — A control block through which the information required by access routines for storage and retrieval of data is communicated to them.

data conversion — 1. The changing of data from one form of representation to another. 2. Changing numerical information from one format to another via an IC processing chip or other means.

data display — Visual presentation of processed data by means of special electronic or electromechanical devices connected (either online or offline) with digital computers or component equipment. Although line printers and punch cards may display data, they usually are categorized as output equipment rather that an displays.

data distributor — An array of simple gates that can accept one or more input lines of data and route the data to specific outputs as determined by the levels of control inputs to the array.

data element — An element that converts data functions into a usable signal. *See also* element, 2.

data element dictionary — A software tool used to describe each data element, i.e., to tell what it is.

Data Encryption Standard — *See* DES.

data entry — 1. Data entered by an operator from a single data device, such as a card or badge reader, numeric keyboard, or rotary switch. 2. Converting data into a form suitable for entry into a computer system, such as by keying from a terminal onto magnetic disk or tape. 3. Entering data directly into a computer system.

data entry device — An electromechanical device to allow manual input of data to a display system. Examples of data entry devices are alphanumeric keys, data tablet, function keys, joystick, mouse, and trackball.

data file — A collection of related data records organized in a specific manner, such as a payroll or inventory file.

data-flow analysis — Graphical analysis of sequential data patterns, e.g., by tools that identify undefined variables.

data flowchart — A flowchart that represents the path of data used in a problem and that defines the major phase of the processing as well as the various data media used.

data-flow diagram — An illustration having a configuration such that it suggests a certain amount of circuit operation.

data-flow machine — A parallel-processing architecture in which each processor acts on instructions when the data needed become available.

data format — The structure and significance of data areas on a storage medium, without reference necessarily to the value of the data contained. Initial values, or limit values, are considered part of a data format definition. The data format itself may have been specified by parameter values at system generation time.

data gathering — *See* data collection.

data generators — Specialized word generators in which the programming is designed to test a particular class of device, in which the pulse parameters and timing are adjustable, and in which selected words may be repeated, reinserted later in the sequence, omitted, etc.

data-handling capacity — The number of bits of information that may be stored in a computer system at one time. The rate at which these bits may be fed to the input either by hand or with automatic equipment.

data-handling system — Semiautomatic or automatic equipment used in the collection, transmission, reception, and storage of numerical data.

data hierarchy — A data structure, made up of sets and subsets, in which every subset of a set is of lower rank than the data of the set.

data highway — A single-cable data link that provides communication among multiple stations that are separate program counters, computers, and data terminals. It eliminates the need for separate, independently wired data links. Whether communicating or not, all stations may function independently.

data identifier — The means of establishing the identity of data by a classification method so that data with close resemblance can be related on a one-to-one or a category basis. Typical identifiers are names, titles, limits or boundaries, classes, and types. Identifiers may also classify the manner of data generation, its completeness, or state of maturity.

data item — The simplest type of information dealt with by a computer system (e.g., a name or employee number). A collection of data items constitutes a record (e.g., payroll information on one employee), and a number of related records constitutes a file (e.g., payroll information on all employees of a company).

data link — 1. Electronic equipment that permits automatic transmission of information in digital form. 2. Equipment, especially transmission cables and interface modules, that permits the transmission of information. 3. A circuit designed to carry digital information, usually by time-division multiplex techniques. Often used to transfer information from one register to another. 4. Any information channel used for connecting data-processing equipment to any input, output, display device, or other data-processing equipment, usually at a remote location.

data logger — A system to measure a number of variables and make a written tabulation and/or record in a form suitable for computer input.

data management — Those control-program functions that provide access to data sets, enforcement of data-storage conventions, and regulation of the use of input/output devices.

data organization—Any of the data-management conventions for determining the arrangement of a data set.

dataphone—1. A trademark of American Telephone and Telegraph Co. to identify the data sets made and supplied for use in the transmission of data over the regular telephone network. 2. A telephone equipped with a modem and appropriate switching for both voice and data transmission.

dataphone digital service—Abbreviated DDS. An all-digital transmission service offered by Bell Telephone that eliminates the need for special digital-to-analog modems.

data pointer—A register holding the memory address of the data (operand) to be used by an instruction. Thus, the register points to the memory location of the data.

data processing—1. The handling of information in a sequence of reasonable operations. 2. The execution of a programmed sequence of operations upon data. A generic term for computing in business and other applications. 3. A term used in reference to operations performed by data-processing equipment.

data-processing center—An installation of computer equipment that provides computing services for users.

data-processing machine—A general name for a machine that can store and process numerical and alphabetical information.

data-processing system—A network of machine components that can accept information, process it in accordance with a plan, and produce the desired results.

data processor—1. An electronic or mechanical machine for handling information in a sequence of reasonable operations. 2. Any device capable of performing operations on data; e.g., a desk calculator, a tape recorder, an analog computer, or a digital computer.

data rate—1. The speed at which data is sent to a receiving computer or device, measured in bits per second. 2. The speed at which digital information is transmitted, expressed in hertz per second or bits per second.

data reduction—1. The process of converting a large quantity of information into a more manageable form, usually including a reduction in volume and a simplification of format. 2. The conversion of raw data into a more useful form, such as graphic representations. 3. The process of transforming masses of raw test or experiment data, usually gathered by instrumentation, into a useful, ordered, or simplified form.

data-reduction system—Automatic equipment employed to simplify the use and interpretation of a large amount of data gathered by instrumentation.

data selector—A decision-making type of digital building block that routes data from any one of several inputs to its output.

data service unit—Abbreviated DSU. A device used on transmission facilities specifically designed to transmit digital data. It is equipped with local and remote testing capabilities (normally provided by a modem on an analog or voice transmission facility) and all other signals that are needed to provide a standard interface to terminals.

dataset—1. A circuit-terminating device used to provide an interface between a circuit and terminal input/output equipment. 2. A modem. A device that converts the signal of a business machine to signals that are suitable for transmission over communications lines and vice versa. It may also perform other related functions. 3. The complete interface unit supplied by the carrier, including a network control and signaling unit, a modem, and devices to protect the network from signals that might interfere with other users. 4. A device used to encode digital data onto voice phone lines.

data sheet—Also called spec sheet. A compilation of terminal information on a specific device defining the electrical and mechanical characteristics of that device.

data signaling rate—The data-transmission capacity, expressed in bits per second, of a set of parallel communication channels.

data sink—1. A communications device that can accept data signals from a transmission device. Also, it may check the received signals and originate signals for error control. 2. A memory or recording device that can store information for future use. *See also* data source.

data source—A communications device that can originate data signals for a transmission device and may accept error control signals. *See also* data sink.

data stabilization—Stabilization of a radar display with respect to a selected reference, regardless of changes in the attitude of the vehicle that carries the radar, as in azimuth-stabilized PPI.

data stream—Data that is flowing from one point to another in a network. A data stream usually consists of a succession of messages (or data blocks).

data stream protocol—The part of a data stream that facilitates the time-critical delivery and synchronization of time-based media.

data switcher—A system used to connect network lines to a specific data-processing computer port.

data synchronizer—A device that controls and synchronizes the transmission of data between an input/output (I/O) device and the computer system.

data tablet—A data entry device consisting of a stylus and a graphic recorder with a coordinate grid similar to the number space of the screen. By pressing the stylus on the tablet, an interrupt is generated and the coordinates of the stylus are stored in special $X - Y$ input registers. The registers are then read by the host computer. The tablet generally replaces the light pen for cursor and tracking symbol movements, and is used extensively in storage-tube display systems, in which light-pen tracking and identification are impossible.

data terminal—1. A common point at which data from various sources is collected and transferred; it may include or connect with several types of data-processing equipment. 2. Equipment at the end of a transmission system for the transmission or reception of data. 3. A class of devices characterized by keyboards and CRT displays.

data termination equipment—Abbreviated DTE. 1. Any device that generates information to be transmitted over a transmission facility. A terminal is a DTE. 2. Equipment that constitutes the data source and/or data sink and provides the communication control function protocol; it includes any piece of equipment at which a communication path begins or ends.

data tracks—Positions of information storage on drum storage devices. Information storage on the drum surface is in the form of magnetized and nonmagnetized areas.

data transfer—Moving data from one (or more) registers (or memory locations) to another, or interchanging data between registers (memory locations) in many different ways.

data transmission—The sending of information from one place to another or from one part of a system to another.

data-transmission equipment—The communications equipment used in direct support of data-processing equipment.

data transmission system — Means for transmitting data; for example, dataphone, radio, etc. *See also* telemetry.

data transmission utilization measure — In a data-transmission system, the ratio of the useful data output to the total data input.

data value — The information contained in data formats. Normally prepared for the generation of specific packages from generic packages.

data word — A computer word that contains or represents the data to be manipulated.

datum reference — A defined point, line, or plane used to locate the pattern or layer for manufacturing, inspection, or for both purposes.

daughterboard — 1. A board that mounts on, and connects to, the motherboard, in a computer. Sometimes used for memory upgrades. 2. A small circuit board that is mounted on and adds capability to another circuit board called a motherboard.

daughter card — Card or board interfaced with a motherboard or backplane.

dB — Abbreviation for decibel. Unit of measurement for power and voltage level.

dBa — Abbreviation for decibels adjusted. Used in conjunction with noise measurements. The reference level is −90 dBm, and the adjustment depends on the frequency-band weighting characteristics of the measuring device.

dBase — A database management program for PCs.

dBi — Decibels of gain relative to an isotropic antenna, or one that radiates equal power in all directions; used to measure antenna gain.

dBj — A unit used to express relative rf signal levels. The reference level is 0 dBj = 1000 microvolts. (Originated by Jerrold Electronics.)

dBk — Decibels referred to 1 kilowatt.

dBm — 1. Abbreviation for decibels above (or below) 1 milliwatt. A quantity of power expressed in terms of its ratio to 1 milliwatt. Power-level measurement unit in the telephone industry based on 600-ohm impedance and 1004-Hz frequency; 0 dBm is 1 mW at 1004 Hz, terminated by a 600-ohm impedance. 2. Decibels referenced to 1 milliwatt; used in communication work as a measure of absolute power values. Zero dBm equals 1 milliwatt.

dB meter — A meter having a scale calibrated to read directly in decibel values at a specified reference level (usually, 1 milliwatt equals 0 dB).

DBMS — *See* Database management system.

dBm0 — The power in dBm measured at, or referred to, a point of zero relative transmission level.

dBr — Decibels relative level. Used to define transmission levels at a point in a circuit, with respect to the level at the zero transmission level reference point.

dBRAP — Decibels above reference acoustical power, which is defined as 10^{-16} watt.

dBRN — Decibels above reference noise. This is a unit used to show the relationship between the interfering effect of a noise frequency, or band of noise frequencies, and a fixed amount of noise power commonly called reference noise. A tone of 1000 hertz having a power level of −90 dBm was selected as the reference noise power because it appeared to have negligible interfering effect and would permit the measurement of interfering effect in positive numbers.

DBS — Abbreviation for direct broadcast satellite.

dBV — The increase or decrease in voltage independent of impedance levels. A unit of electrical pressure. Decibels referred to a standard of 1 volt.

dBW — Decibels referenced to 1 watt.

dBx — Decibels above the reference coupling. Reference coupling is defined as the coupling between two circuits that would be required to give a reading of 0 dBa on a two-type noise-measuring set connected to the disturbed circuit when a test tone of 90 dBa (using the same weighting as that used on the disturbed circuit) is impressed on the disturbing circuit.

dc — Abbreviation for direct current.

D cable — Two-conductor cable, each conductor having the shape of the capital letter D, with insulation between the conductors and between the conductors and the sheath.

dc amplifier — *See* direct-current amplifier.

dc balance — An adjustment of circuitry to avoid a change in dc level when the gain is changed.

dc beta — The dc current gain of a transistor; the ratio of the collector current to the base current that caused it, measured at constant collector-to-emitter voltage.

dc block — 1. A (coaxial) component employed to prevent the flow of dc or video along a transmission line while allowing the uninterrupted flow of rf. The structure is a short section of coaxial line having a capacitance in series with the center and/or outer conductor. The rf flows with negligible reflection or attenuation, while the video frequencies or dc are blocked. 2. A device that stops the flow of dc power but permits passage of higher-frequency signals.

dc breakdown — Voltage at which ionization occurs at a slowly rising dc voltage.

dcc — Abbreviation for double cotton-covered.

dc capacitance — The capacitance of a capacitor calculated from the ratio of the capacitor charge and the direct voltage measured between the terminations.

dc capacitor — A capacitor designed essentially for application with a direct voltage.

dc circuit breaker — A device used to close and interrupt a dc power circuit under normal conditions or to interrupt this circuit under fault or emergency conditions.

dc component — The average value of a signal. In television it represents the average luminance of the picture being transmitted, and in radar it is the level from which the transmitted and the received pulses rise.

dc continuity — A circuit in which an impressed dc current — a reading on a conventional ohmmeter applied across the terminals of a circuit with dc continuity — will result in a deflection of the meter.

dc coupled — The connection by a device that passes the steady-state characteristics of a signal and that largely eliminates the transient or oscillating characteristics of the signal.

dc dump — The withdrawal of direct-current power from a computer. This may result in loss of the stored information.

DCE — Abbreviation for data communication equipment or data channel equipment. Equipment (such as a modem) installed at a user's premises that provides all the functions required to establish, maintain, and terminate a connection and provide signal conversion and coding between the data-terminal equipment and the common carrier's line.

dc generator — A rotating electric device for converting mechanical power into dc power.

dc inserter stage — A television transmitter stage that adds a dc component known as the pedestal level to the video signal.

dcl — Abbreviation for direct current leakage.

dc leakage current — Abbreviated dcl. 1. The relatively small direct current through a capacitor when dc voltage is impressed across it. 2. That direct current which passes through the capacitor dielectric when a capacitor is subjected to a polarized voltage potential across its terminals and has reached a state of charge equilibrium. Usually expressed in microamperes and associated mostly with electrolytic capacitors.

dcnm—Abbreviation for dc noise margin.

dc noise—The noise arising when reproducing a magnetic tape that has been nonuniformly magnetized by energizing the record head with direct current, either in the presence or absence of bias. This noise has pronounced long-wavelength components that can be as much as 20 dB higher than those obtained from a bulk-erased tape.

dc noise margin—Abbreviated dcnm. Noise margin is also called noise immunity. 1. The difference between the normal applied logic levels and the threshold voltage of a digital integrated circuit. 2. The difference between the output voltage level of a driving gate and the input threshold voltage of a driven gate for both the 1 and the 0 states.

dc operating point—The dc values of collector voltage and current of a transistor with no signal applied.

dc overcurrent relay—A device that functions when the current in a dc circuit exceeds a given value.

dc patch bay—Specific patch panels provided for termination of all direct-current circuits and equipment used in an installation.

dc picture transmission—Transmission of the dc component of the television picture signal. This component represents the background or average illumination of the overall scene and varies only with the overall illumination.

dc plate resistance—The value or characteristic used in vacuum-tube computations. It is equal to the direct-current plate voltage divided by the direct-current plate current and is given the symbol R_p.

dc reclosing relay—A device that controls the automatic closing and reclosing of a dc circuit interrupter, generally in response to load circuit conditions.

dc resistivity—The resistance of a body of ferromagnetic material having a constant cross-sectional area, measured under stated conditions by means of direct voltage, multiplied by the cross-sectional area, and divided by the length of the body.

dc restoration—The reestablishment by a sampling process of the dc and the low-frequency components of a video signal that have been suppressed by ac transmission.

dc restorer—Also called clamper or restorer. A clamping circuit that holds either amplitude extreme of a signal waveform to a given reference level of potential.

dc shift—An error in transient response, with a time constant approaching several seconds.

dc short—A coaxial component that provides a dc circuit between the center and outer conductors, while allowing the rf signal to flow uninterrupted. The unit has a high-impedance line shunted across the main coax line. This consequently makes the device frequency dependent.

dc signaling—A transmission method that utilizes direct current.

dc test—A general term for those tests that measure a static parameter such as leakage current.

DCTL—Abbreviation for direct-coupled transistor logic.

dc transducer—A transducer capable of proper operation when excited with direct current. Its output is given in terms of direct current unless otherwise modified by the function of the stimulus.

dc transmission—Transmission of a television signal in such a way that the dc component of the picture signal is still present. This is done to maintain the true level of background illumination.

dcwv—Abbreviation for direct-current working volts. The maximum continuous voltage that can be applied to a capacitor.

DDD—Abbreviation for direct distance dialing.

DDP—Abbreviation for distributed data processing.

DDS—*See* dataphone digital service.

deac—*See* deaccentuator.

deaccentuator—Abbreviated deac. A network or circuit employed in frequency-modulated receivers to deemphasize the higher frequencies in the received signal to restore their proper relative amplitude. *See also* deemphasis.

deactuate pressure—The pressure at which an electrical contact opens or closes as the pressure approaches the actuation level from the opposite direction.

dead—1. Free from any electric connection to a source of potential difference and from electric charge; having the same potential as that of the earth. The term refers only to current-carrying parts that are sometimes alive, or charged. 2. *See* room acoustics.

dead band—Also called dead space, dead zone, or switching blank. 1. In a control system, the range of values through which the measurand can be varied without initiating an effective response. 2. A specified range of values in which the incoming signal can be altered without also changing the outgoing response. 3. The angle within which the rotor of a stepper motor can stop—the error band, or band of position uncertainty caused by bearing and load friction, and rotor and load inertia.

deadbeat—Coming to rest without vibration or oscillation; i.e., the pointer that a highly damped meter or galvanometer moves to a new position without overshooting and vibrating about its final position.

deadbeat instrument—A voltmeter, meter, or similar device in which the movement is highly damped to bring it to rest quickly.

deadbeat response—The response of a critically damped stepper motor that provides rotation from one step to another without overshoot or ringing.

dead break—An unreliable contact made near the trip point of a relay or switch at low contact pressure. As a result, the switch does not actuate, even though the circuit is interrupted.

dead-center position—The place on the commutator of a dc motor or generator at which a brush would be placed if the field flux were not distorted by armature reaction.

dead end—1. In a sound studio, the end with the greater sound-absorbing characteristic. 2. In a tapped coil, the portion through which no current is flowing at a particular bandswitch position.

dead-end tower—An antenna or transmission-line tower designed to withstand unbalanced mechanical pull from all the conductors in one direction, together with the wind strain and vertical loads.

deadlock—1. A situation that occurs when all tasks within a computer system are suspended, waiting for resources that have already been assigned to other tasks that are also waiting for additional resources. Thus, the system can perform no useful work unless tasks are destroyed and their resources reclaimed. 2. A condition in which two processes wait indefinitely for each other. The solution is to allocate resources on a priority basis or to have tie-breaking circuitry.

deadly embrace—A situation in which two processes each unknowingly wait for resources held by the other. *See also* deadlock.

dead range—*See* dead band.

dead room—1. A room for testing the acoustic efficiency or range of electroacoustic devices such as speakers and microphones. The room is designed with an absolute minimum of sound reflection, and no two dimensions of the room are the same. A ratio of 3 to 4 to 5 is usually employed (e.g., 15 ft × 20 ft × 25 ft). The walls, floor, and ceiling are lined with a sound-absorbing material. 2. *See* anechoic.

dead short — A short circuit having minimum resistance.

dead space — 1. An area or zone, within the normal range of a radio transmitter, in which no signal is received. 2. *See* dead band, 2.

dead spot — 1. A geographic location in which signals from one or more radio stations are received poorly or not at all. 2. That portion of the tuning range of a receiver in which stations are heard poorly or not at all because of poor sensitivity.

dead time — 1. The minimum interval, following a pulse, during which a transponder or component circuit is incapable of repeating a specified performance. 2. Any definite delay intentionally placed between two related actions to avoid overlap that could result in confusion or permit another particular event, such as a control decision or switching event, to occur. 3. Time interval in which no clock phase is active.

dead volume — The total volume of the pressure port cavity of a pressure transducer at the rest position (i.e., with no stimulus applied).

dead zone — *See* dead band, 2.

debicon — A high-efficiency microwave generator in which use is made of crossed-field effects.

debouncing — The elimination of accidental bounce signals characteristic of mechanical switches that bounce repeatedly until the contact is finally closed or opened. Debouncing may be performed by hardware (latch) or software (delay).

de Broglie wavelength — The wavelength of radiation that corresponds to a photon whose energy is 1 electron volt: 1.24 micrometer.

debug — 1. To examine or test a procedure, routine, or equipment for the purpose of detecting and correcting errors. 2. To detect, locate, and remove mistakes from a program. Debugging programs are available that test for and isolate errors in another program. 3. A method of fault finding in programs, usually using data dumps and breakpoints. This would be handled in software by a debug routine. 4. Checking the logic of a software program to isolate and remove any mistakes. 5. To detect, locate, and remove all mistakes in a computer program and any malfunctions in the computing system itself.

debugger — 1. An essential program designed to facilitate software debugging. At a minimum, it provides breakpoints, dump facilities, and register and memory examine/modify. 2. Program that facilitates the testing of the object program on a microcomputer and its input-output devices. Debuggers usually accept commands from the user to perform such functions as (*a*) displaying or printing out the contents of the microcomputer memories, or the contents of the registers of the central processing unit, (*b*) modifying the RAM, (*c*) starting execution of the object program from a specific memory location, and (*d*) setting a breakpoint or stopping execution of the program when the instruction at a specific memory location is reached in the program or when a given condition is met. 3. A program that allows the user to observe the program flow and results of the program's operation in a step-by-step mode. It may be used to change data or instructions, alter registers, etc. 4. Program that helps track down and eliminate errors that occur in the normal course of program development.

debugging — 1. Isolating and removing all malfunctions (bugs) from a computer or other device to restore its operation. 2. A process of shakedown operation of each finished material that is performed prior to its being placed in use in order to exclude the early failure period. During debugging, weak elements are expected to fail and be replaced by elements of normal quality that are not subject to early failure. 3. Process of detecting, locating, and correcting mistakes in hardware (system wiring) or software (program).

debugging period — *See* early-failure period.

debugging routines — Programs that aid in the isolation and correction of malfunctions and/or errors in a unit of equipment or another program.

debug monitor — An interactive program that allows the design engineer to intercommunicate in a "friendly" manner (through a terminal device) with the microcomputer system under development and to control closely the execution of an untested microcomputer program in order to check its correct operation.

debug program — A special program used to find errors in a program that is being run on a computer. A debug program allows a programmer to correct programming errors in the program being run.

debunching — Space-charge effect that tends to destroy the electron bunching in a velocity-modulation vacuum tube by spreading the beam due to mutual repulsion of the electrons.

Debye effect — The selective absorption of electromagnetic waves by a liquid made up of molecules with permanent dipole moments.

Debye length — Also called Debye shielding distance or plasma length. A theoretical length that describes the maximum separation at which a given electron is influenced by the electric field of a given positive ion.

Debye shielding distance — *See* Debye length.

decade — 1. The interval between any two quantities having a ratio of 10:1. 2. A group or assembly of 10 units (e.g., a counter that counts to 10 in one column, or a resistor box that inserts resistance quantities in multiples of powers of 10). 3. Ten times a given quantity or range.

decade band — A band having frequency limits related by the equation $f_b - f_a = 10$.

decade box — A special assembly of precision resistors, coils, or capacitors. It contains two or more sections, each having 10 times the value of the preceding section. Each section is divided into 10 equal parts. By means of a 10-position selector switch or equivalent arrangement, the box can be set to any desired value in its range.

decade counter — A logic device that has 10 stable states and may be cycled through these states by the application of 10 clock or pulse inputs. A decade counter usually counts in a binary sequence from state 0 through state 9 and then cycles back to 0. Sometimes referred to as a divide-by-10 counter.

decade resistance box — A resistance box containing two or more sets of 10 precision resistors.

decade scaler — A decade counter, or scale-of-10 counter. A scaler with a factor of 10. It produces one output pulse for every 10 input pulses.

decametric waves — High-frequency band; 3 to 10 MHz.

decay — 1. Gradual reduction of a quantity. 2. Exponential reduction in amplitude. To be reduced in an exponential manner, as the current in a circuit decays when the source of potential is removed. 3. The decrease in the radiation intensity of any radioactive material with respect to time. 4. In a storage tube, a change in magnitude or configuration of stored information by any cause other then erasing or writing.

decay characteristic — *See* persistence characteristic (of a luminescent screen).

decay distance—The distance between an area of wave generation and a point of passage of the resulting waves outside the area.

Decca—A British long-range hyperbolic navigational system that operates in the 70- to 130-kilohertz frequency band. It is a continuous-wave system in which the receiver measures and integrates the relative phase difference between the signal received from two or more synchronized ground stations. One master station and three slave stations are usually arranged in star formation. Operational range is about 250 miles (402 km).

decelerated electrons—Electrons that, after traveling at a great rate of speed, strike a target, become quickly decelerated, and cause the target to emit X-rays.

decelerating electrode—In an electron-beam tube, an electrode to which a potential is applied to slow down the electrons in the beam.

deceleration—The act or process of moving, or of causing to move, with decreasing speed; the state of so moving.

deceleration time—1. In a computer, the time interval between the completion of the reading or writing of a record on a magnetic tape and the time when the tape stops moving. 2. The time required to stop a motor, whether free running or with some braking means.

deception—Deliberate production of false or misleading echoes on enemy radar by the radiation of spurious signals synchronized to the radar or by the reradiation of the radar pulses from extraneous reflectors.

deception device—A device that works to make unfriendly signals either unusable or misleading.

deception jamming—*See* confusion jamming.

deci-—Prefix meaning one-tenth (10^{-1}).

decibel—1. Abbreviated dB. The standard unit for expressing transmission gain or loss and relative power levels. One decibel is one-tenth of a bel. The term *dBm* is used when a power of 1 milliwatt is the reference level. Decibels indicate the ratio of power output (P_o) to power input (P_{in}):

$$dB = 10 \log_{10}(P_o/P_{in})$$

2. A unit of change in sound intensity. One decibel is approximately the smallest change that the ear can perceive. Larger decibel increments reflect the fact that sound intensity must be squared in order for the ear to perceive a doubling of intensity. An increase in intensity is expressed as a positive number of dBs, a decrease as a negative value. No change in intensity is 0 dB, and 0 is also used to indicate a starting point from which changes are measured. 3. A unit used to measure and compare signal levels on a logarithmic scale. 4. A measure of the ratio between two power levels. Doubling or halving the power corresponds to a 3-dB change, and 10 dB corresponds roughly to the audible effect of doubling or halving the loudness of a signal (although it represents a power ratio of 10:1). Decibels are frequently used to specify variation in signal level throughout a range of frequencies (i.e., 20–20,000 Hz ±1 dB) and to specify such other ranges as signal-to-noise ratio. 5. A relative measure of signal or sound intensity or volume. It expresses the ratio of one intensity to another. Can also express voltage and power ratios logarithmically.

decibel meter—Also called dB meter. 1. An instrument for measuring the electric power level, in decibels, above or below an arbitrary reference level. 2. Sound-level indicator.

decibels above or below 1 milliwatt—The unit used to describe the ratio of the power at any point in a transmission system to a reference level of 1 milliwatt. The ratio expresses decibels above or below this reference level of 1 milliwatt.

decibels above or below 1 watt—A measure of power expressed in decibels to a reference level of 1 watt.

decibels above reference noise—An expression used to describe the ratio of the circuit noise level in a transmission system, at any point, to some arbitrarily chosen reference noise. The expression signifies the reading of a noise meter. Where the circuit-noise meter has been adjusted to represent effect under specified conditions, the expression is in adjusted decibels.

decilog—A division of the logarithmic scale used for measuring the logarithm of the ratio of two values of any quantity. The number of decilogs is equal to 10 times the logarithm to the base 10 of the ratio. One decilog therefore corresponds to a ratio of $10^{0.1}$ (i.e., 1.25892+).

decimal—1. Pertaining to a characteristic or property involving a selection, choice, or condition in which there are 10 possibilities. 2. Pertaining to the number representation system with a radix of 10. 3. Pertaining to a system of numerical representation in which there are ten symbols, 0, 1, 2, 3, . . . 9.

decimal attenuator—A system of attenuators arranged so that a voltage or current can be reduced decimally.

decimal-binary switch—A switch by means of which a single input lead is connected to appropriate combinations of four output leads (representing 1, 2, 4, and 8) for each of the decimal-numbered settings of the associated control knob. For example, with the knob in position 7, the input lead would be connected to output leads 1, 2, and 4.

decimal code—A code in which each allowable position has one of 10 possible states. The conventional number system with the base 10 is a decimal code.

decimal-coded digit—One of ten arbitrarily selected patterns of 1s and 0s that are used to represent decimal digits.

decimal digit—One of the numbers 0 through 9 used in the number system with the base 10.

decimal encoder—An encoder in which there are 10 output lines, one for each digit from 0 to 9, for each decade of decimal numbers.

decimal notation—The writing of quantities in the decimal numbering system.

decimal numbering system—The popular numbering system using the Arabic numerals 0 through 9 and thus having a base, or radix, of 10. For example, the decimal number 2345 can be derived in this way:

$$2000 + 300 + 40 + 5 = 2345$$

or

$$2(10^3) + 3(10^2) + 4(10^1) + 5(10^0) = 2345$$

In the decimal system, all numbers are obtained by raising the radix (total number of marks, or 10 in this system) to various powers.

decimal point—In a decimal number, the point that marks the place between integral and fractional powers of 10.

decimal-to-binary conversion—The mathematical process of converting a number written in the scale of 10 into the same number written in the scale of 2.

Decimal	*Binary*	*Decimal*	*Binary*
0	0	10	101
1	1	11	1011
2	10	12	1100
3	11	13	1101
4	100	14	1110
5	101	15	1111
6	110	16	10000
7	111	32	100000
8	1000	64	1000000
9	1001	128	10000000

decimetric waves — 1. Electromagnetic waves having wavelengths between 0.1 and 1 meter. 2. Ultrahigh frequency band; 300 MHz to 3 GHz.

decineper — One-tenth of a neper.

decinormal calomel electrode — A calomel electrode containing a decinormal potassium chloride solution.

decision — In a computer, the process of determining further action on the basis of the relationship of two similar items of data.

decision box — On a flowchart, a rectangle or other symbol used to mark a choice or branching in the sequence of programming of a digital computer.

decision element — In computers or data-handling systems, a circuit that performs a logical operation, such as AND, OR, NOT, or EXCEPT, on one or more binary digits in input information that represent "yes" or "no" and that expresses the result in its output. *See also* gate.

decision table — A table of all contingencies that are to be considered in the description of a problem, together with the actions to be taken. Decision tables are sometimes used in place of flowcharts for problem description and documentation.

deck — 1. In computer usage, a collection of cards, usually a complete set of cards punched for a definite purpose. 2. A term usually applied to a tape machine having no built-in power amplifiers or loudspeakers of its own, but intended rather for feeding a separate amplifier and speaker system, as in a component installation.

declination — The offset angle of an antenna from the axis of its polar mount as measured in the meridian plane between the equatorial plane and the antenna main beam.

declination offset angle — The adjustment angle of a polar mount between the polar axis and the plane of a satellite antenna used to aim at the geosynchronous arc.

decode — 1. In a computer, to obtain a specific output when specific character-coded input lines are activated. 2. To use a code to reverse a previous encoding. 3. To determine the meaning of characters or character groups in a message. 4. To determine the meaning of a set of pulses that describes an instruction, a command, or an operation to be carried out.

decoder — 1. A device for translating a combination of signals into one signal that represents the combination. It is often used to extract information from a complex signal. 2. In automatic telephone switching, a relay-type translator that determines from the office code of each call the information required for properly recording the call through the switching train. Each decoder has means, such as a cross-connecting field, for establishing the controls desired and readily changing them. 3. Sometimes called matrix. In an electronic computer, a network or system in which a combination of inputs is excited at one time to produce a single output. 4. A device that converts coded information into a more usable form, for example, a binary-to-decimal decoder. 5. A circuit that accepts coded input data and activates a specific output(s) in accordance with the code present at the input. 6. A circuit built into an FM tuner to enable it to translate stereo signal information into two matched audio outputs. 7. A means to extract and process recorded quadraphonic sound information from a complex signal into four matched outputs. 8. A device consisting of gates that is usually connected to the output of a counter. It provides an output when the counter is at a specific count or range of counts. 9. A logic device that breaks the code of an incoming binary signal; i.e., converts the coded information into a more usable form. 10. A logic device that converts data from one number system to another (e.g., an octal-to-decimal decoder). Decoders are also used to recognize unique addresses, such as a device address, and bit patterns. *See* code converter. 11. A circuit that restores a signal to its original form after it has been scrambled. 12. A device that reconstructs an encrypted signal so that it can be clearly received. 13. A television set-top device that enables the home subscriber to convert an electronically scrambled television picture into a viewable signal. (This should not be confused with a digital coder/decoder, known as a codec, which is used in conjunction with digital transmissions.)

decoding — 1. The process of obtaining intelligence from a code signal. 2. In multiples, a process of separating the subcarrier from the main carrier.

decoding matrix — A device for decoding many input lines into a single output line.

decoding network — A circuit made so that, when a particular combination of inputs is on, an output appears on one of a number of output lines.

decollimated light — In fiber optics, light rays made nonparallel by striae and boundary defects.

decommutation — The process of recovering a signal from the composite signal previously created by a commutation process.

decommutator — Equipment for separating, demodulating, or demultiplexing commutated signals.

decomposition — Breaking down a software specification, in depth and breadth, to determine all required functions and their relationships.

decoupler — A circuit for eliminating the effect of coupling in a common impedance.

decoupling — 1. The reduction of coupling. 2. To isolate two circuits on a common line. A decoupling network is a low-pass filter (*RC* or *RLC*) that does not isolate equally in both directions.

decoupling circuit — A circuit used to prevent interaction of one circuit with another.

decoupling network — A network of capacitors and chokes or resistors placed into leads that are common to two or more circuits, to prevent unwanted, harmful interstage coupling.

decoy — A reflecting object having reflective characteristics of a target, used in radar deception.

decrement — 1. Progressive diminution in the value of a variable quantity; also the amount by which a variable decreases. When applied to damped oscillations, it is usually called damping factor. 2. A specific part of an instruction word in some binary computers; thus, a set of digits. 3. To reduce the numerical contents of a counter. A decrement of one is usually assumed unless specified otherwise. 4. In an oscillating system with damping (each oscillation has less amplitude than the one preceding it), the ratio of the peak values (voltage, distance, etc.) of two successive half-cycles. It is expressed as a decimal fraction less than 1.

decremeter — An instrument for measurement of the logarithmic decrement (damping) of a wave train.

decryption — The process of "unscrambling" an encrypted or coded message.

dedicated— 1. To set apart for some special use. For example, a dedicated microprocessor is one that has been specifically programmed for a single application, such as weight measurement by scale, traffic light control, etc. (ROMs by their very nature [read-only] are dedicated memories.) 2. A piece of equipment that is assigned to one particular use only. Minicomputers are often dedicated. Microprocessors are intended to be dedicated.

dedicated computer— A computer whose use is reserved for a particular task.

dedicated line— 1. A communication line that isn't dialed, also called a leased or private line. 2. A communication line for voice and/or data rented from a communication carrier. 3. Full-term line allocated to one subscriber at a specific degree of conditioning. 4. Leased line, wired directly between communicating systems. It has faster transmission than regular telephone lines and does not require acoustic couplers (types of modems). 5. A permanent circuit for private use. The leased line physically connected between locations, or through a central office, without using the switching equipment. 6. A very common type of transmission facility for data communications. Each line is separate from the dial-up network (DDD) and its associated equipment and has a discrete, permanent path to its destination. Users pay a flat monthly fee for this service. 7. A telephone line that has a continuous connection, maintained by the telephone company.

dedicated machine— A PC or other microcomputer designed to handle a special, usually single, task. A dedicated server PC could only be used to service user requests, not as both server and workstation.

dedicated register— A register in a computer exclusively used to contain a specific item.

dee— A hollow, D-shaped accelerating electrode in a cyclotron.

dee line— A structural member that supports the dee of a cyclotron and together with the dee forms the resonant circuit.

deemphasis— Also called postemphasis or postequalization. 1. Introduction of a frequency-response characteristic that is complementary to that introduced in preemphasis. 2. Reduction of the level of the higher audio frequencies during FM reception or tape replay so that they compensate for the preemphasis applied to the transmission. This restores an overall uniform response. 3. A form of equalization used in FM tuners, complementary to a preemphasis used in transmission. The purpose is to improve the overall signal-to-noise ratio while maintaining a uniform frequency response. It is expressed in the form of a time constant or product of a resistance and capacitance. Standard FM broadcasts use a 75-µ time constant in the United States, and 50 µs in Europe, whereas Dolby B transmissions use a 25-µs time constant.

deemphasis network— A network inserted into a system to restore the preemphasized frequency spectrum to its original form.

deenergize— 1. To disconnect a device from its power source. 2. To stop the current in a circuit or to remove electrical potential from a circuit, as by opening a switch.

deep discharge— The withdrawal of all available electrical energy before recharging a cell or battery.

deep space net— A combination of radar and communications stations in the United States, Australia, and South Africa so located as to keep a spacecraft in deep space under observation at all times.

default— The value(s) or option(s) that are assumed during operation when not specified.

default value— The value that a database element assumes if the user does not specify another value.

defeat— The frustration, counteraction, or thwarting of an alarm device so that it fails to signal an alarm when a protected area is entered. Defeat includes both circumvention and spoofing.

defect— 1. A condition considered potentially hazardous or operationally unsatisfactory and therefore requiring attention. 2. Any nonconformance with the normally accepted characteristics for a unit. *See also* major defect; minor defect.

defect analysis— The process of examining technical or management (nontechnical) data, manufacturing techniques, or material to determine the cause of variations of electrical, mechanical, or physical characteristics outside the limitations established at any manufacturing checkpoint.

defect condition— Hole conduction in the valence band in a semiconductor.

Defense Electronic Supply Center— *See* DESC.

deferred addressing— An indirect addressing mode in which the directly addressed location contains the address of the operand, rather than the operand itself.

deferred entry— In a computer, an entry into a subroutine as a result of a deferred exit from the program that passes control to the subroutine.

deferred exit— In a computer, the transfer of control to a subroutine at a time controlled by the occurrence of an asynchronous event rather than at a predictable time

defibrillator— An electronic device that applies a brief high-voltage potential to the heart by means of electrodes placed on the chest wall. The defibrillator is used to restore regular rhythm to a heart in ventricular fibrillation.

definite-purpose relay— A relay with some electrical or mechanical feature that distinguishes it from a general-purpose relay.

definition— 1. The fidelity with which the detail of an image is reproduced. When the image is sharp (i.e., has definite lines and boundaries), the definition is said to be good. 2. The degree with which a communication system reproduces sound images or messages. 3. The fidelity with which the pattern edges in a printed circuit (conductors, inductors, etc.) are reproduced relative to the original master pattern. 4. The sharpness of a picture subjectively evaluated in terms of its resolution. 5. The sharpness of a screen-printed pattern; the exactness with which a pattern is printed.

deflecting coil— An inductor used to produce a magnetic field that will bend the electron beam a desired amount in the cathode-ray tube of an oscilloscope, television receiver, or television camera.

deflecting electrode— An electrode to which a potential is applied in order to deflect an electron beam.

deflecting torque— *See* torque of an instrument.

deflection— Movement of the electron beam in a cathode-ray tube as electromagnetic or electrostatic fields are varied to cause the light spot to traverse the face of the tube in a predetermined pattern.

deflection circuit— The circuit that regulates an electron beam's deflection in a CRT.

deflection coil— One of the coils in the deflection yoke.

deflection factor— *See* deflection sensitivity.

deflection focusing— The progressive defocusing of a cathode-ray-tube display image that occurs when the deflected electron beam impinges on the CRT screen at a slant.

deflection plane— A plane perpendicular to the cathode-ray-tube axis and containing the deflection center.

deflection plates— Two pairs of parallel electrodes, the pairs set one forward of the other and at right angles to each other, parallel to the axis of the electron stream

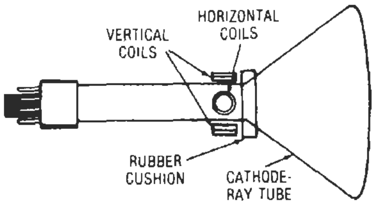

Deflection coils.

within an electrostatic cathode-ray tube. An applied potential produces an electric field between each pair. By varying the applied potential, this field may be varied to cause a desired angular displacement of the electron stream.

deflection polarity—The relationship between the direction of displacement of an oscilloscope trace and the polarity of the applied signal wave.

deflection sensitivity—Also called deflection factor. The displacement of the electron beam at the target or screen of a cathode-ray tube divided by the change in magnitude of the deflection field. Deflection sensitivity is usually expressed in millimeters (or inches) per volt applied between the deflecting electrodes, or in millimeters (or inches) per ampere in the deflection coil.

deflection voltage—The voltage applied to the electrostatic plates of a cathode-ray tube to control the movement of the electron beam.

deflection yoke—An assembly of one or more electromagnets for deflecting one or more electron beams.

defocus-dash mode—A method of storage of binary digits in a cathode-ray tube. Initially, the writing beam is defocused to excite a small circular area on the screen. For one kind of binary digit the beam remains defocused, and for the other kind of binary digit it is suddenly focused into a concentric dot, which traces out a dash on the screen during the interval of time before the beam is cut off and moved to the next position.

defocus-focus mode—A variation of the defocus-dash mode in which the focused dot is not caused to trace a dash.

defragmentation—A process in which all the files on a hard disk are rewritten so that all the parts of each file are written to contiguous sectors.

defruiting—Method of eliminating asynchronous returns in radar beacon systems.

degassing—The process of driving out and exhausting the gases of an electron tube occluded in its internal parts.

degauss—To neutralize the existing magnetic field. The term was coined during World War II to describe the neutralizing of a ship's magnetic field by a grid of cables generating an equal but opposite magnetic field.

degausser—Also called automatic degausser and bulk eraser. 1. A device to clarify the color picture by means of coils within the set. The coils deactivate the magnetization that builds up around a color TV set when it is moved around or when other electrical devices are brought too close to the receiver. 2. Any device for neutralizing magnetism, as in a recorder head or in a separate unit. Also called a tape eraser, for use with a complete tape recording on its reel. 3. Demagnetizer. *See* bulk eraser; head demagnetizer.

degaussing—Girdling a ship's hull with a web of current-carrying cable that sets up a magnetic field equal in value and opposite in polarity to that induced by the earth's magnetic field, thus rendering the ship incapable of actuating the detonator of a magnetic mine.

degeneracy—The condition in which two or more modes have the same resonant frequency in a resonant device.

degenerate modes—A set of modes having the same resonance frequency (or propagation constant). The members of a set of degenerate modes are not unique.

degenerate parametric amplifier—An inverting parametric device for which the two signal frequencies are identical and equal to one-half the frequency of the pump. (This exact but restrictive definition is often relaxed to include cases in which the signals occupy frequency bands that overlap.)

degeneration—*See* negative feedback.

degradation—1. A gradual decline of quality or loss of ability to perform within required limits. The synonym *drift* is often used for electronic devices. 2. A condition in which the system continues to operate but at a reduced level of service. Unavailability of major equipment subsystems or components is the usual cause. 3. A gradual deterioration in performance as a function of time.

degradation failure—Failure of a device because a parameter or characteristic changes beyond some previously specified limit.

degree day—The measure of the deviation of the mean daily temperature from a given standard, with each variance from the standard during a single day recorded as one degree day.

degree of current rectification—The ratio between the average unidirectional current output and the root-mean-square value of the alternating-current input from which it was derived.

degree of membership—The confidence or certainty, expressed as a number from 0 to 1, that a particular value belongs in a fuzzy set.

degree of voltage rectification—The ratio between the average unidirectional voltage and the root-mean-square value of the alternating voltage from which it was derived.

degree rise—The amount of increase in temperature caused by the introduction of electricity into a unit.

degrees of freedom—In a vibrating system, the coordinates necessary to locate the position of the vibrating element at any time. For example, a single-degree-of-freedom system can move along only one axis, in both directions. A two-degrees-of-freedom system will require two coordinates to describe the position of the elements. A multiple-degree-of-freedom system generally has many elements that can move along many axes.

deinstall—To remove a program or hardware device from active service.

deion circuit breaker—A circuit breaker built so that the arc that forms when the circuit is broken is magnetically blown into a stack of insulated copper plates, giving the effect of a large number of short arcs in series. Each arc becomes almost instantly deionized when the current drops to zero in the alternating-current cycle, and the arc cannot reform.

deionization—The process by which an ionized gas returns to its neutral state after all sources of ionization have been removed.

deionization potential—The potential at which ionization of the gas within a gas-filled tube ceases and conduction stops.

deionization time—The time required for the grid of a gas tube to regain control after the anode current has been interrupted.

deionized water—Water that has been treated to remove ions. Deionized water is required in certain electronic applications to prevent contamination of parts coming in contact with the water. *See* demineralized water.

dekahexadecimal—*See* sexadecimal notation.

Dekatron—A cold-cathode counting tube.

delamination—1. Separation of conductive patterns from the substrate, or separation of layers of the base material. 2. A separation between any of the layers of a base material or between the laminate and the conductive foil, or both.

delay—1. The time required for a signal to pass through a device or conductor. 2. The time interval between the instants at which any designated point in a wave passes any two designated points of a transmission circuit.

delay circuit—1. A circuit that delays the passage of a pulse or signal from one part of a circuit to another. 2. An electronic time-delay device that can introduce time delays from a few milliseconds to about 100 milliseconds without significantly degrading signal quality. Can be used to restore acoustics of a large auditorium to recorded programs heard in a normal-sized room.

delay coincidence circuit—A coincidence circuit actuated by two pulses, one of which is delayed a specific amount with respect to the other.

delay counter—In a computer, a device that can temporarily delay a program a sufficient length of time for the completion of an operation.

delay distortion—Also called envelope-delay distortion, phase-delay distortion, or phase distortion. 1. Phase-delay distortion; i.e., departure from flatness in the phase delay of a circuit or system over the frequency range required for transmission, or the effect of such departure on a transmitted signal. 2. Envelope distortion; i.e., departure from flatness in the envelope delay of a circuit or system over the frequency range required for transmission, or the effect of such departure on a transmitted signal. 3. The amount of variation in delay for various frequency components of the facsimile signal, usually expressed in microseconds from an average delay time. 4. The difference between the maximum and minimum phase delay within a specified band of frequencies. 5. Distortion caused by the fact that the higher-frequency components of a signal travel slower over a transmission facility than the lower-frequency components and therefore arrive later and out of phase. Numerically, it is the maximum difference in transmission time between any two frequencies in a specified frequency band, expressed in microseconds. Measured in microseconds of delay relative to the delay at 1700 Hz. Delay distortion doesn't affect voice communication but can seriously impair data transmissions.

delayed automatic volume control—Abbreviated delayed AVC. An automatic volume-control circuit that acts only on signals above a certain strength. It thus permits reception of weak signals even though they may be fading, whereas normal automatic volume control would make the weak signals even weaker.

delayed AVC—*See* delayed automatic volume control.

delayed contacts—Contacts that are actuated a predetermined time after the start of a (timing) cycle.

delayed PPI—A PPI (plan-position indicator) in which the initiation of the time base is delayed.

delayed repeater satellite—A satellite that stores information obtained from a ground terminal at one location and, upon interrogation by a terminal at a different location, transmits the stored message.

delayed sweep—1. In a cathode-ray tube, a type of sweep that is not allowed to begin for a while after being triggered by the initiating pulse. 2. A sweep that has been delayed either by a predetermined period or by a period determined by an additional independent variable.

delay equalizer—1. A device that adds delay at certain frequencies to a circuit in a way to reduce the delay distortion. 2. A corrective network that is designed to make the phase delay or envelope delay of a circuit or system substantially constant over a desired frequency range. 3. A network that introduces an amount of phase shift complementary to the phase shift in the circuit at all frequencies within the desired band. 4. Selective delaying of various frequency components of the received signal to match the delay of other frequency components caused by envelope delay distortion.

delay/frequency distortion—That form of distortion which occurs when the envelope delay of a circuit or system is not constant over the frequency range required for transmissions.

delay line—1. A real or artificial transmission line or equivalent component that is used to delay a signal, either linear or digital, for a predetermined length of time. The delay time is defined as the duration of time between the leading edge of the input pulse and the 50-percent point on the leading edge of the output pulse. 2. A specially constructed cable used in the luminance channel of a color receiver to delay the luminance signal. 3. A sequential logic element that has one input channel and in which the state of an output channel at any instant is the same as the state of the input channel at the instant $t - n$, where n is a constant time interval for a given output channel (the input sequence undergoes a delay of n time units). 4. A device that can cause the transmission of one unit of information to be retarded until another unit can synchronize with it. 5. A device capable of causing an energy impulse to be retarded in time from point to point, thus providing a means of storage by circulating intelligence-bearing pulse configurations and patterns. Examples of delay lines are material media such as mercury, in which sonic patterns may be propagated in time; lumped constant electrical lines; coaxial cables; transmission lines; and recirculating magnetic drum loops. 6. A cable made to provide a very low velocity of propagation with long electrical delay for transmitted signals.

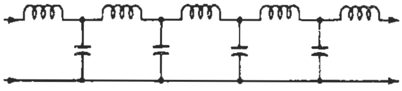

Lumped-constant line.

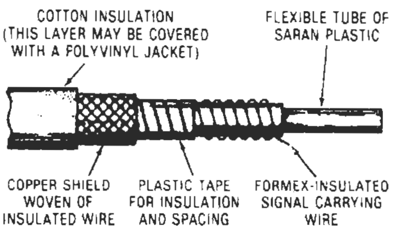
COTTON INSULATION
(THIS LAYER MAY BE COVERED WITH A POLYVINYL JACKET)
FLEXIBLE TUBE OF SARAN PLASTIC
COPPER SHIELD WOVEN OF INSULATED WIRE
PLASTIC TAPE FOR INSULATION AND SPACING
FORMEX-INSULATED SIGNAL CARRYING WIRE

Early-model delay line.

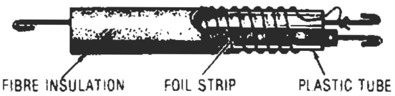

Modern delay line.

Delay lines, 1.

delay-line memory—*See* delay-line storage.

delay-line register—An acoustic or electric delay line in an electronic computer, usually one or an integral number of words in length, together with input, output, and circulation circuits.

delay-line storage—Also called delay-line memory. In an electronic computer, a storage or memory device consisting of a delay line and a means for regenerating and reinserting information into it.

delay multivibrator—A monostable multivibrator that produces an output pulse a predetermined time after it is triggered by an input pulse.

delay on break—A term used to describe a mode of operation relative to timing devices. The delay begins when the initiate switch is opened (delay on break of initiate switch).

delay on energization—Also known as delay on make. A term used to describe a mode of operation relative to timing devices. The delay begins when the initiate switch is closed, or on application of power to the input.

delay on make—Same as delay on energization.

delay PPI—A radar indicator in which the start of the display sweep is delayed after the trigger so that distant targets are displayed on a short-range scale that gives an expanded presentation.

delay relay—Also called time-delay relay. A relay in which there is a delay between the time it is energized or deenergized and the time the contacts open or close.

delay time—1. The amount of time one signal is behind (lags) another. 2. Of a switching transistor, the time interval between the application at the input terminals of a pulse that is switching the transistor from a nonconducting to a conducting state and the appearance at the output terminals of the pulse induced by the charge carriers. The time is usually measured between points corresponding to 10 percent of the amplitude of the applied pulse and of the output pulse. 3. Measurement of the interval between direction of signal to an LED and attainment of 10-percent output current in the photodetector.

delay timer—A term sometimes used to designate a timer that is primarily used for energizing (or deenergizing) a load at the end of a timed period. *See* time-delay relay.

delay unit—The unit of a radar system in which pulses may be delayed a controllable amount.

deleting—Removing something from a computer. It could be text from a word processing window, or a file from a hard disk drive.

deletion record—In a computer, a new record to replace or remove an existing record in a master file.

delimiter—Also called separator. 1. In a computer, a character that limits a string of characters and therefore cannot be a member of the string. 2. A flag that separates and organizes items of data. 3. Any means used to separate data items at input. Most frequently used are spaces or commas. The delimiter for the two integers 123 567 is the space between them. 4. A character that separates and organizes elements of a program in a computer. 5. A punctuation character, such as blackslash or comma, that separates one section of a computer command from another. 6. A text character that marks the beginning and/or end of a unit of data or separates different data components. For example, periods are used as delimiters in domain names, hyphens and parentheses are used in phone numbers and social security numbers, and blank spaces and commas are used in written text.

Dellinger effect—*See* radio fadeout.

delta—1. The Greek letter delta (Δ) represents any quantity that is much smaller than any other quantity of the same units appearing in the same problem. 2. In a magnetic cell, the difference between the partial-select outputs of the same cell in a 1 state and in a 0 state. 3. Brainwave signals whose frequency is approximately 0.2 to 3.5 Hz. The associated mental state is usually a deep sleep or a trancelike state.

delta circuit—A three-phase circuit in which the windings of the system are connected in the form of a closed ring, and the instantaneous voltages around the ring equal zero. There is no common or neutral wire, so the system is used only for three-wire systems or generators.

delta connection—In a three-phase system, the terminal connections. So called because they are triangular, like the Greek letter delta.

delta-delta monitor—A monitor in which the red, green, and blue color guns are arranged in a triangle, and a shadow mask with round holes aligns each gun to the proper phosphor dots.

delta match—*See* Y match.

delta matched antenna—Also called Y antenna. A single-wire antenna (usually one half-wavelength long) to which the leads of an open-wire transmission line are connected in the shape of a Y. The flared part of the Y matches the transmission line to the antenna. The top of the Y is not cut, giving the matching section its triangular shape of the Greek letter delta, hence the name.

delta matching transformer—An impedance device used to match the impedance of an open-wire transmission line to an antenna. The two ends of the transmission line are fanned out so that the impedance of the line gradually increases. The ends of the transmission line are attached to the antenna at points of equal impedance, symmetrically located with respect to the center of the antenna.

delta modulation—1. A means of encoding analog signals in control and communication systems. The output of the delta encoder is a single weighted digital pulse train that may be decoded at the receiving end to reconstruct an original analog signal. 2. A type of waveform encoding in which the differences between individual digitized speech samples are encoded. Data point values are determined by changes from preceding data point amplitudes.

delta modulator—A closed-loop sampled data control system that transmits binary output pulses whose polarity depends on the difference between the input signal being sampled and a quantized approximation of the preceding input signal. Delta modulation affords a simple, efficient method of digitizing voice for secure, reliable communications and for voice I/O in data processing.

delta network—A set of three branches connected in series to form a mesh.

delta pulse code modulation—A modulation system that converts audio signals into corresponding trains of digital pulses to provide greater freedom from interference during transmission over wire or radio channels.

delta tune—A control or switch similar in function to a clarifier, found on many AM transceivers. It compensates for signals off the center frequency of a CB channel. Although its effective range is about the same as that of a clarifier, adjustment is not as critical.

delta wave—A brain wave whose frequency is below 4 hertz.

dem—Abbreviation for demodulator.

demagnetization—Partial or complete reduction of residual magnetism.

demagnetization coefficient—*See* permeance coefficient.

demagnetization curve—In the second quadrant of a hysteresis loop, the portion that lies between the residual induction point, B_r, and the coercive force point, H_c.

demagnetization effect — 1. A decrease in internal magnetic field caused by uncompensated magnetic poles at the surface of a sample. 2. The portion of the normal hysteresis loop in the second quadrant showing the induction, B (gauss), in a magnetic material as related to the demagnetizing field, H (oersted). Points on this curve are usually designated by the coordinates B_d and H_d. This curve describes the characteristics of a permanent magnet (as contrasted with an electromagnet).

demagnetizer — A device that removes residual magnetism from recording or playback tape heads. This magnetism, if not removed, can introduce noise on recordings and cause high-frequency loss.

demagnetizing coefficient — *See* permeance coefficient.

demagnetizing field — A magnetizing force applied in such a direction as to reduce the remanent induction in a magnetized body.

demagnetizing force — A magnetizing force applied in such a direction that it reduces the residual induction in a magnetized body.

demand-driven machine — A parallel-processing architecture in which processors carry out instructions when the results of a processing step are demanded.

demand factor — The ratio of the maximum demand of an electrical system, or part of a system, to the total connected load of a system or that part of a system under consideration.

demand load — The load that is drawn from the source of supply at the receiving terminals, averaged over a suitable and specified interval of time, expressed in kilowatts, amperes, etc.

demarcation strip — A physical interface, usually a terminal board, between a business machine and a common carrier. *See also* interface, 1.

Dember effect — Also known as the photodiffusion effect. The production of a potential difference between two regions of a semiconductor specimen when one is illuminated. This phenomenon is related to the photoelectromagnetic effect, except there is no magnetic field. H. Dember discovered that when an illuminated metal plate producing electrons is bombarded by other electrons from an outside source, the photoelectric emission increases because, in addition to photoelectrons, secondary electrons are also knocked out by bombardment.

demineralized water — Water that has been treated to remove the minerals that are normally present in hard water. Demineralized water is required in some electronic applications in which extreme precautions must be taken to prevent contamination. *See* deionized water.

demodulation — Also called detection. 1. Operation on a previously modulated wave in such a way that it will have substantially the same characteristics as the original modulating wave. 2. The process by which a wave corresponding to the modulating wave is obtained from a modulated wave. 3. The process of retrieving digital (computer) data from a modulated analog (telephone) signal.

demodulator — Abbreviated dem. 1. A device that operates on a carrier wave to recover the wave with which the carrier was originally modulated. 2. A facsimile device that detects an amplitude-modulated signal and produces the modulating frequency as a direct current of varying amplitude. This type of unit is used to provide a keying signal for a frequency-shift exciter unit for radio facsimile transmission. 3. A device that receives tones from a transmission circuit and converts them to electrical pulses, or bits, that may be accepted by a business machine. 4. Circuitry that plays back a CD-4 disc's four signals after reprocessing the base and carrier bands inscribed in each side of the record groove. 5. A functional section of a modem that converts received analog line signals to digital form. 6. A device that separates information from the carrier.

demodulator probe — A probe designed for use with an oscilloscope, for displaying modulated high-frequency signals.

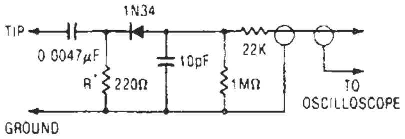

*R may be increased in value to reduce circuit loading

Demodulator probe.

demon — A computer program that waits until an event occurs before running; widely used to circumvent copy-protection procedures.

de Morgan's theorem — A theorem that states that the inversion of a series of AND implications is equal to the same series of inverted OR implications, or the inversion of a series of OR implications is equal to the same series of inverted AND implications. In symbols, $\overline{A \cdot B \cdot C} = \bar{A} + \bar{B} + \bar{C}$, or $\overline{A + B + C} = \bar{A} \cdot \bar{B} \cdot \bar{C}$.

demountable tube — A high-power electron tube having a metal envelope with porcelain insulation. Can be taken apart for inspection and for renewal of electrodes.

demultiplexer — 1. A device used to separate two or more signals that were previously combined by a compatible multiplexer and transmitted over a single channel. 2. A circuit that directs information from a single input to one of several outputs at a time in a sequence that depends on the information applied to the control inputs. 3. A device that reverses the action of a multiplexer, deriving a group of separate channels from the complex multiplex signal. 4. A logic circuit that can route a single line of digital information to other lines. The demultiplexer acts to switch information to many different points. 5. A circuit that applies the logic state of a single input to one of several outputs.

demultiplexing — Abbreviated demux. The process of separating a multiplexed signal into its separate intelligence signals.

demultiplexing circuit — A circuit that is used to separate the signals that have been combined for transmission by multiplex.

demux — *See* demultiplexing.

denary band — A band having frequency limits with the ratio of $f_b/f_a = 10$.

dendrite — A semiconductor crystal with a heavily branched, treelike structure that grows from the nucleus as the metal becomes solidified.

dendritic growth — 1. A technique of producing semiconductor crystals in long, uniform ribbons with optically flat surfaces. 2. The electrolytic transfer of metal from one conductor to another, similar to electroplating except that the dendritic growth usually, though not always, forms from cathode to anode. The dendrite resembles a tree in appearance, and when it touches the opposite conductor, there is an abrupt rise in current.

dense binary code — A binary code in which all the possible states of the pattern are used.

densitometer — An instrument for measuring the optical density (photographic transmission, photographic reflection, visual transmission, and so forth) of a material.

density — 1. A measure of the light-transmitting or reflecting properties of an area expressed by the common logarithm of the ratio of incident to transmitted or reflected light flux. 2. The mass per unit volume. The specific gravity of a body is the ratio of a density to the density of a standard substance. Water and air are commonly used as the standard substances. 3. Amount per unit cross-sectional area (e.g., current, magnetic flux, or electrons in a beam). 4. The logarithm of the ratio of incident to transmitted light. *See* opacity.

density modulation — Modulation of an electron beam by varying the density of the electrons in the beam with time.

density packing — The number of magnetic pulses (representing binary digits) stored on tape or drum per linear inch on a single head.

density step tablet — A facsimile test chart consisting of a series of areas that increase in steps from a low value of density to a maximum value of density.

dentonphonics — The technique of using electronics in broadcasting speech from the mouth. The principle is the same as that of a throat microphone, in which a transducer responds to sound energy transmitted through the tissues as a person speaks.

dependent linearity — Nonlinearity errors expressed as a deviation from a desired straight line of fixed slope and/or position.

dependent mode — In network analysis, a node having one or more incoming branches.

deperming — Another name for demagnetization.

depletion field-effect transistor — An active semiconductor device in which the main current is controlled by the depletion width of a pn junction.

depletion layer — Also called barrier layer. 1. In a semiconductor, the region in which the mobile-carrier charge density is insufficient to neutralize the net fixed charge density of donors and acceptors. 2. That region in the immediate vicinity of a semiconductor pn junction that becomes exhausted or *depleted* of current carriers in order to set up the internal potential barrier involved in either the balance between diffusion and drift currents present in the equilibrium case, or the imbalance between these currents present in a nonequilibrium situation. Being depleted of carriers, the depletion-layer region is virtually composed of "intrinsic" material, irrespective of the doping levels of the p-type and n-type materials from which it is formed. 3. A zone, several atoms thick, at the junction of n-type and p-type semiconductor materials, in which there are no current carriers, either free electrons or holes, unless biased by a direct voltage. Free electrons in the n-type material are repelled by negative charges in the p-type material, and the holes in the p-type material are repelled by the positive nucleus of atoms in the n-type material.

depletion-layer capacitance — Also called barrier capacitance. Capacitance of the depletion layer of a semiconductor. It is a function of the reverse voltage.

depletion-layer rectification — Also called barrier-layer rectification. The rectification that appears at the contact between dissimilar materials, such as a metal-to-semiconductor contact or a pn junction, as the energy levels on each side of the discontinuity are readjusted.

depletion-layer transistor — Any of several types of transistors that rely directly for their operation on the motion of carriers through depletion layers (for example, a spacistor).

depletion-mode device — A field-effect transistor or IC that passes maximum current at zero gate potential, and a decreasing current with applied gate potential.

depletion-mode field-effect transistor — 1. A field-effect transistor that exhibits substantial device current (I_{DSS}) with zero gate-to-source bias ($V_{GS} = 0$ V). 2. An MOS transistor normally on with zero gate voltage applied (channel formed during processing). A voltage of the correct polarity applied to the gate will force majority carriers from the channel, thus depleting it and turning the transistor off.

depletion-mode operation — The operation of a field-effect transistor such that changing the gate-to-source voltage from zero to a finite value decreases the magnitude of the drain current.

depletion region — Also referred to as space-charge, barrier, or intrinsic region. The region, extending on both sides of a reverse-biased semiconductor junction, in which all carriers are swept from the vicinity of the junction; that is, the region is depleted of carriers. This region takes on insulating characteristics and is capable of isolating semiconductor regions from each other. Depletion regions make planar bipolar integrated circuits possible.

depletion-type field-effect transistor — A field-effect transistor having appreciable channel conductivity for zero gate-source voltage; the channel conductivity may be increased or decreased according to the polarity of the applied gate-source voltage.

depolarization — The process of preserving the activity of a primary cell by the addition of a substance to the electrolyte. This substance combines chemically with the hydrogen gas as it forms, thus preventing excessive buildup of hydrogen bubbles.

depolarize — To make partially or completely unpolarized.

depolarizer — A chemical substance, usually manganese dioxide, added to a dry or primary cell to remove the polarizing chemical products resulting from discharge, and thus to keep the discharge rate constant; to prevent formation of hydrogen bubbles at the positive electrode.

deposited carbon — Resistive element made of a thin film of crystalline carbon or a carbon alloy sputtered onto a ceramic rod.

deposition — The application of a material to a substrate through the use of chemical, vapor, electrical, vacuum, or other processes.

depth of cut — The depth to which the recording stylus penetrates the lacquer of a recording disc.

depth of field — 1. The in-focus range of a lens or optical system. It is measured from the distance behind an object to the distance in front of the object when the viewing lens shows the object to be in focus. 2. The distance between the first object in focus and the last object in focus within a scene as viewed by a particular lens; it can vary with the quality and focal length of the lens or with its f-stop setting.

depth of heating — The depth at which effective dielectric heating can be confined below the surface of a material when the applicator electrodes are placed adjacent to only one surface.

depth of modulation — In a radio-guidance system obtaining directive information from the two spaced lobes of a directional antenna, the ratio of the difference in total field strength of the two lobes to the field strength of the greater lobe at a given point in space.

depth of penetration — The thickness of a layer extending inward from the surface of a conductor and having the same resistance to direct current as the whole conductor has to alternating current of a given frequency. *See also* skin depth.

de-*Q* — To reduce the Q of a tuned circuit, as generally applied to carrier-current transmission systems.

derate — 1. To use a device at a lower current or voltage than it is capable of handling in order to reduce the probability of failure or to permit its use under a condition of high ambient temperature. 2. To change the rating or value to a lower rating in consideration of other affecting factors.

derating — 1. The reduction in rating of a device or component, especially the maximum power-dissipation rating at higher temperatures. 2. Deliberately understressing components so as to provide increased reliability. (This requires the selection of components of higher stress than is required for normal operation.) For example, using a $\frac{1}{2}$-watt resistor in circuit conditions demanding a $\frac{1}{4}$-watt dissipation. This process is recognized as an effective and well-established method of achieving reliable designs. The ratio of applied stress to rated stress is the stress ratio; thus, in the example above this would be 0.25/0.5, or 50 percent derated in terms of power. 3. The intentional reduction of stress-to-strength ratio in the application of an item, usually for the purpose of reducing the occurrence of stress-related failures.

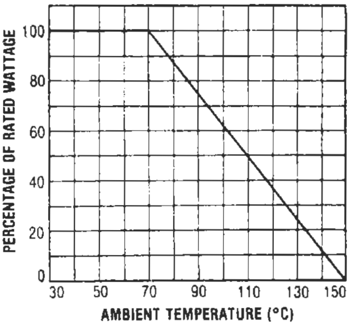

Derating curve.

derating factor — The factor by which the ratings of component parts are reduced to provide additional safety margins in critical applications or when the parts are subjected to extreme environmental conditions for which their normal ratings do not apply.

derivative action — *See* rate action.

derivative control — Automatic control in which the rate of correction is determined by the rate at which the error producing it changes.

derived center channel — A monophonic composite signal derived from the sum or difference of the left and right stereo channels, often fed to an extra speaker to fill in an aural hole between the left and right speakers. The signal from a voltage-derived center-channel output must be fed to an external power amplifier before it can drive a speaker. A power-derived center channel can drive a speaker directly.

DES — Abbreviation for Data Encryption Standard. A scheme approved by the National Bureau of Standards that encrypts data for security purposes. For use within the United State by the NSA (National Security Agency); DES is the data-communications encryption standard specified by Federal Information Processing Systems (FIPS) Publication 46.

DESC — Abbreviation for Defense Electronic Supply Center. 1. The agency that sets military specifications for all electronic components and verifies performance to these requirements. 2. Government agency that controls procurement policies and monitors quality for military electronics contracts.

desensitization — 1. The saturation of one component (an amplifier, for instance) by another so that the first cannot perform its proper function. 2. The effect on a receiver section tuned to one channel that is caused by a strong signal on another channel. It is an AGC-type effect whereby the desired signal's strength appears to be decreased by the presence of a nearby signal. This effect influences a receiver's overall selectivity. 3. The tendency of a receiver to fail to recognize valid DTMF signals in the presence of such factors as dial tone, pilot signals, or data signals. 4. The reduction in sensitivity of a receiver caused by either noise or rf overload from a nearby transmitter.

desiccant — A substance used as a drying agent because of its affinity for water.

design-center rating — Limiting values of operating and environmental conditions that should not be exceeded under normal conditions in a bogey electronic device of a specified type as defined by its published data. These values are chosen by the device manufacturer to provide acceptable serviceability of the device in average applications, taking responsibility for normal changes in operating conditions due to rated supply-voltage variation, equipment component variation, equipment control adjustment, load variation, signal variation, environmental conditions, and variations in the characteristics of all electronic devices.

design compatibility — Electromagnetic compatibility achieved by incorporating in all electromagnetic radiating and receiving apparatus (including antennas) characteristics or features for elimination or rejection of undesired self-generated or external signals in order to enhance operating capabilities in the presence of natural or human-made electromagnetic noise.

design engineer — An engineer who has been assigned to design a specific product for a specific application.

design for maintainability — Those features and characteristics of design of an item that reduce requirements for tools, test equipment, facilities, spares, highly skilled personnel, etc., and improve the capability of the item to accept maintenance actions.

design-maximum ratings — Limiting values of operating and environmental conditions applicable to a bogey electronic device of a specified type as defined by its published data; they should not be exceeded under the worst probable conditions.

design-maximum rating system — *See* design-center rating. These values are chosen by the device manufacturer to provide acceptable serviceability of the device, taking responsibility for the effects of changes in operating conditions due to variations in the characteristics of the electronic device under consideration.

Desk-Fax — Trade name of Western Union Telegraph Co. for a small facsimile transceiver employed principally in short-line telegraph service.

desktop computer — 1. A self contained, totally integrated system that comes complete with central processor, read/write memory, external mass storage in the form of a cassette tape or diskette system, keyboard, computer language and operating system in firmware (usually in the form of read-only memory), connectors for external devices, and an output display such as a CRT, LED readout, or printer. When the computer is turned on, it is ready to solve problems. The operating system and language do not have to be loaded. Compiling is done automatically. Because of this integrated structure, a user can edit and execute complex commands by way of short, simple instructions. The internal operating system will recognize typing or syntax errors and give an easy-to-understand explanation on the display or printout. Also, the computer can prompt the user to supply subsequent inputs so that steps are not missed or implemented out of sequence.

Especially significant in design applications is the unit's ability to accept changes in problem parameters while a program is running. 2. A complete microprocessor-based system that is fairly easy to use by noncomputer experts. 3. A computer designed for scientific/engineering problems that are too complex for personal calculators or that require peripherals or interfacing capabilities for data acquisition or instrumentation control. Desktop computers have full, high-level programming languages, much larger memories than personal calculators, built-in mass storage devices, and I/O ports for interfacing to peripherals or other instruments. Some have impressive graphics capabilities for plotting, drawing, graphing, and lettering. 4. A complete, highly integrated system ready to use as it comes out of the box. It has an input device (a keyboard), a display device (a cathode-ray tube or single-line display), mass storage (magnetic cards, cassettes, floppy discs, CD-ROMs, etc.), a processor, memory, connectors for external I/O (input/output) devices, a power supply, and a language that resides in ROM (read-only memory) all housed in a single package that literally sits on a desk. Of course, the desktop computer may also have special peripherals, such as plotters, digitizers, and other instruments. One can simply connect the external I/O devices and peripherals, if any, turn the power on, and use the machine.

desktop publishing — Abbreviated DTP. The creation on personal computers of publication-quality printed documents that combine text and graphics.

desoldering — Process of disassembling soldered parts in order to repair, replace, inspect, or salvage them. Typical desoldering methods are wicking, pulse vacuum (solder sucker), heat and pull, and solder extraction.

destaticization — Treatment of a material to minimize the accumulation of static electricity and, as a result, the amount of dust that adheres to the material because of such static charges.

destination address — In computer systems having a source-destination architecture, the destination address is the address of the device register or memory location to which data is being transferred.

destination register — In a computer, a register into which data is being placed.

Destriau effect — Sustained emission of light by suitable phosphor powders embedded in an insulator and subjected only to the action of an alternating electric field.

destructive readout — 1. The destruction of data in a storage device by the act of reading the data. For example, reading data from a core memory clears the addressed location. When destructive readouts are used, data (modified or unmodified) is written back into the same location. 2. A characteristic of a memory. The memory is said to have a destructive readout if information retrieved from memory must be written back in immediately after it is used or else it is lost. A core memory has destructive readout. Computers with destructive readouts contain special circuits to write information back into memory after readout.

destructive-readout memory — *See* DRO memory.

destructive test — Any test resulting in the destruction or drastic deterioration of the test specimen.

detail — A measure of the sharpness of a recorded facsimile copy or reproduced image. Generally related to the number of lines scanned per inch. Defined as the square root of the ratio between the number of scanning lines per unit length and the definition in the direction of the scanning line.

detail contrast — The ratio of the amplitude of the high-frequency components of a video signal to the amplitude of the reference low-frequency component.

detail enhancement — Also called image enhancement. A system in which each element of a video picture is analyzed in relation to adjacent horizontal and vertical elements. When differences are detected, a detail signal is generated and added to the luminance signal to enhance it.

detection — *See* demodulation.

detection range — The greatest distace at which a sensor will consistently detect an intruder under a standard set of conditions.

detectophone — An instrument for secretly listening in on a conversation. A high-sensitivity, nondirectional microphone is concealed in the room and connected to an amplifier and headphones or recorder remotely located. Sometimes the microphone feeds into a wireless transmitter that broadcasts over power lines, to permit the listener to be farther away.

detector — 1. A device for effecting the process of detection or demodulation. 2. A mixer or converter in a superheterodyne receiver; often referred to as a first detector. 3. A device that produces an electrical output that is a measure of the radiation incident on the device. 4. A rectifier tube, crystal, or dry disc by which a modulation envelope on a carrier or the simple on-off state of a carrier may be made to drive a lower-frequency device. 5. A device that converts light signals from optical fibers to electrical signals that can be further amplified to allow reproduction of the original signal. 6. A device that converts optical power to other forms. *See* photodetector.

detector balance bias — A controlling circuit used in radar systems for anticlutter purposes.

detector circuit — That portion of a receiver which recovers the modulation signal from the rf carrier wave.

detector diode — A diode, often associated with microwave circuits, that converts rf energy into dc or video output.

detector power efficiency — The ratio of the change in dc power in the load resistance produced by the ac signal to the available power from a sinusoidal voltage generator when the diode is operated under specified conditions.

detector probe — A probe containing a high-frequency rectifying element such as a crystal diode or a tube. Used with an oscilloscope, vacuum-tube voltmeter, or signal tracer for recovering the modulation from a carrier.

detector quantum efficiency — The ratio of the number of carriers generated to the number of photons absorbed.

detector voltage efficiency — The ratio of the dc load voltage to the peak sinusoidal input voltage under specified circuit conditions.

detent — 1. A stop or other holding device, such as a pin, lever, etc., on a ratchet wheel. 2. Switch action typified by a gradual increase in force to a position at which there is an immediate and marked reduction in force. 3. A bump or raised section projecting from the surface of a spring or other part. 4. A device that holds a part, control, or assembly in a given position. For example, some connectors use locking detents on the plug half and indents on the cap half to hold the halves together in proper mated position.

detent torque — A measure of the maximum torque that can be applied to the shaft of a deenergized stepper motor before it begins to rotate.

deterministic signal — A signal whose future behavior can be predicted precisely.

detritus — Loose material dislodged during resistor trimming but remaining in the trimmed area.

detune — 1. To change the inductance and/or capacitance of a tuned circuit and thereby cause it to be resonant at other than the desired frequency. 2. To adjust a circuit so that it does not respond to (is not resonant at) a particular frequency.

detuning stub—A quarter-wave stub for matching a coaxial line to a sleeve-stub antenna. The stub tunes the antenna itself and detunes the outside of the coaxial feed line.

Deutsches Institut für Normung—*See* DIN.

deviation—1. The difference between the actual and specified values of a quantity. 2. The difference, usually the absolute difference, between a number and the mean of a set of numbers, or between a forecast value and the actual datum. 3. In FM transmissions and reception, the increase or decrease of signal carrier frequency from the nominal; also applied to drifting. Standard maximum deviation rating is ±75 kHz for FM radio. 4. A departure from specification requirements for which approval is obtained from the consumer prior to occurrence of the departure.

deviation absorption—Absorption that occurs at frequencies near the critical frequency. Occurs in conjunction with the slowing up of radio waves near the critical frequency, upon reflection from the ionosphere.

deviation distortion—Distortion caused by inadequate bandwidth, amplitude-modulation rejection, or discriminator linearity in an FM receiver.

deviation ratio—In frequency modulation, the ratio of the maximum change in carrier frequency to the highest modulating frequency.

deviation sensitivity—The smallest frequency deviation that produces a specified output power in FM receivers.

device—Also called item. 1. A single discrete conventional electronic part such as a resistor or transistor, or a microelectronic circuit. 2. Any subdivision of a system. 3. A mechanical, electrical, and/or electronic contrivance intended to serve a specific purpose. 4. The physical realization of an individual electrical element in a physically independent body, which cannot be further reduced or divided without destroying its stated function. This term is commonly applied to active devices. Examples are transistors, pnpn structures, tunnel diodes, and magnetic cores, as well as resistors, capacitors, and inductors. It is not, for example, an amplifier, a logic gate, or a notch filter. 5. An electronic part consisting of one or more discrete active or passive elements. 6. A unit of processing equipment in a computer system external to the CPU; synonymous with the term *peripheral.*

device adaptor—*See* interface adaptor.

device channel—A dedicated channel associated with a device; connects a file variable to that device.

device complexity—The number of circuit elements within an integrated circuit.

device cutoff—The condition of an electronic device in which its conduction is either zero or relatively insignificant. With semiconductor devices such as FETs, bipolar transistors, and thyristors, cutoff is normally that condition in which the device passes only leakage currents.

device independence—In a computer, the ability to request input/output operations without regard to the nature of the input/output devices.

device register—An addressable register used to store status and control information or data for transfer into or out of a device.

device under test—Abbreviated DUT. A finished assembled device, or an untested die on a wafer.

Dewar flask—A container with double walls. The space between the walls is evacuated, and the surfaces bounding this space are silvered.

dewetted surface—1. A surface that was initially wetted, i.e., a surface on which the solder flowed uniformly. 2. A condition that results when molten solder has coated a surface and then receded, leaving irregularly shaped mounds of solder separated by areas covered with a thin solder film; base metal is not exposed.

dew point—1. The temperature at which condensation first occurs when a vapor is cooled. 2. The temperature at which moisture begins to condense out of a vapor. The relative humidity is then 100 percent.

df—Abbreviation for direction finder or dissipation factor.

df antenna—Any antenna combination included in a direction finder for obtaining the phase or amplitude reference of the received signal. May be a single or orthogonal loop, an adcock, or spaced differentially connected dipoles.

df antenna system—One or more df antennas and their combining circuits and feed systems, together with the shielding and all electrical and mechanical items up to the receiver input terminals.

D flip-flop—A flip-flop whose output is determined by the input that appeared one pulse earlier; for example, if a 1 appeared at the input, the output after the next clock pulse would be 1.

DHG—Abbreviation for digital harmonic generation.

DI—Abbreviation for dielectric isolation. A fabrication technique by which components in an integrated circuit are electrically isolated from each other by an insulator (dielectric material). DI surrounds the sides and bottom of each transistor with a layer of silicon dioxide (glass). DI has proven particularly advantageous for fabricating high-performance analog ICs. The conventional DI fabrication process for bipolar ICs begins with a wafer of n-type silicon. The side of the wafer that will eventually be the bottom is deeply etched (in V-shaped grooves) to form the sidewall pattern, then silicon dioxide and polycrystalline silicon are grown to fill the etched moats and to thicken the eventual DI substrate. The opposite side of the wafer is polished until the insulating sidewalls appear at the wafer surface. Conventional diffusion and metallization processes follow to complete the IC.

diac—1. Two-lead alternating-current switching semiconductor. 2. *See* three-layer diode. 3. A bidirectional breakdown diode that conducts only when a specified breakdown voltage is exceeded.

diagnostic—1. Having to do with the detection and isolation of a malfunction or error in a computer. 2. A message output by a compiler or assembler indicating that a computer program contains a mistake. 3. Software designed to locate either a fault in the equipment or an error in programming.

diagnostic code—An alphanumeric or word display that indicates a system condition such as a malfunction. The code is either self-explanatory or used to refer to further instructions that are explained in an operator guide.

diagnostic function test—A program for testing overall system reliability.

diagnostic program—1. Special program for checking computer's hardware for proper operation. For example, there are CPU diagnostic checks, memory diagnostic checks, and so on. 2. A test program to help isolate hardware malfunctions in the programmable controller and application equipment. 3. A troubleshooting aid for locating hardware malfunctions in a system, or a program to aid in locating coding errors in newly developed programs.

diagnostic routine—1. An electronic-computer routine designed to locate a malfunction in the computer, a mistake in coding, or both. 2. *See* debug.

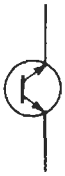

Symbol.

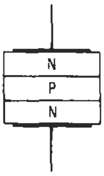

Junctions.

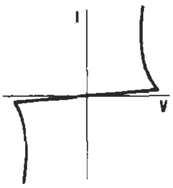

Characteristic.

Diac.

diagnostics—1. Information on what tests a device failed and how they were failed, used to aid in troubleshooting. 2. An output from a tool, indicating software discrepancies and other attributes. 3. Methods used for detecting and isolating faults in a unit under test. 4. Computer system support tasks (usually supplied with an operating system) that test memory, interrupts, timers, and peripheral devices. 5. Programs or procedures used to test a piece of equipment, a communications link or network, or any similar system.

diagnostic test—A test performed for the purpose of isolating a malfunction in the unit under test.

diagnostic testing—Testing designed to locate and identify errors in a computer routine, hardware component, or communication network.

diagonal horn antenna—A horn antenna whose cross sections are all square and in which the electric vector is parallel to a diagonal. The radiation pattern in the far field has almost perfect circular symmetry.

diagonal pliers—Pliers with cutting jaws at an angle to the handles to permit cutting off wires close to terminals or printed circuit boards.

diagram—1. Schematics, prints, charts, or any other graphical representation, the purpose of which is to explain rather than to represent. 2. A schematic representation of a sequence of subroutines designed to solve a problem. 3. A coarser and less symbolic representation than a flowchart, frequently including descriptions in English words. 4. A schematic drawing showing an electrical circuit, or a logical drawing showing logical arrangements within a circuit.

dial—1. A means for indicating the value to which a control knob has been set. 2. A calling device that generates the required number of pulses in a telephone set and thereby establishes contact with the party being called.

dial cable—*See* dial cord.

dial central office—A telephone or teletypewriter office at which is located the automatic equipment necessary for connecting two or more user stations together by wires for communication purposes.

dial cord—Also called dial cable. A braided cord or flexible wire cable connected to a tuning knob so that turning the knob will move the pointer or dial that indicates the frequency to which an analog radio receiver is tuned. Also used for coupling two shafts together mechanically.

dialer—1. A device that detects and reports emergencies by automatic dialing of telephone numbers. When an emergency is detected, the dialer usually begins playing a prerecorded tape containing the telephone number or numbers to be called (in the form of a series of pulses) and the emergency message. When the number has been dialed, the tape continues to play the prerecorded message. 2. *See* automatic dialer.

dialing area—The area within which a telephone company customer may make calls by dialing without using area codes.

dialing key—A dialing method in which a set of numerical keys instead of a dial is used to originate dial pulses. Generally, it is used in connection with voice-frequency dialing.

dial jacks—Strips of jacks associated with and bridged to a regular outgoing trunk jack circuit so that connections between the dial cords and the outgoing trunks can be made.

dial key—The key unit of the subscriber's cord circuit that is used to connect the dial to the line.

dial leg—The circuit conductor brought out for direct-current dial signaling.

dial light—A small pilot lamp that illuminates the tuning dial of a radio receiver.

diallyl phthalate—A thermosetting resin that has excellent electrical insulation properties.

dial office—Telephone central office operating on dial signals.

dialog—Interchange of information between program and user. The program gives prompts or messages, and the user responds by entering appropriate data.

dialog box—1. An on-screen message box in a graphical user interface that allows users to input more specific information than standard commands. 2. An on-screen box that conveys or requests information from the user.

dial pulse—A momentary interruption in the direct current in the loop of a calling telephone, produced by the opening and closing of the dial pulse springs of a calling telephone in response to the dialing of a digit. The current in the calling-line loop is interrupted as many times as there are units in the digit dialed; i.e., dialing the digit 7 generates seven dial pulses (interruptions) in current flowing through the loop of the calling telephone.

dial pulsing—The transmission of telephone address information by the momentary opening and closing of a dc circuit a specified number of times, corresponding to the decimal digit that is dialed. This is usually accomplished, as with an ordinary telephone dial, by manual operation of a finger wheel.

dial register—*See* standard register of a motor meter.

dial telephone system—Telephone system in which telephone connections between customers are ordinarily established by electronic and mechanical apparatus controlled by manipulations of dials operated by the calling parties.

dial tone—1. A hum or other tone employed in a dial telephone system to indicate that the line is not busy and that the equipment is ready for dialing. 2. A tone indicating that automatic telephone switching equipment is ready to receive dial signals, tones, or pulses.

dial-up—1. The use of a dial or push-button telephone for initiating a station-to-station call. 2. Two simplex transmission paths between central offices and a full-duplex path in each subscriber loop. 3. Use of a phone to initiate a call.

dial-up connection—A connection from a computer using a modem to a host computer over standard telephone lines.

dial-up line—Also called switched line. A communication line accessible via dial-up facilities, typically the public telephone network.

diamagnetic—1. Term applied to a substance with a negative magnetic susceptibility. 2. Bars of certain elements, such as zinc, copper, lead, and tin, when freely suspended in a magnetic field arrange themselves at right angles to the lines of force of the magnetic field, i.e., they are magnetized in the opposite direction to the magnetizing field. These elements are said to be diamagnetic.

diamagnetic material—1. A material that is less magnetic than air, or in which the intensity of magnetization is negative. There is no known material in which this effect has more than a very feeble intensity. Bismuth is the leading example of materials of this class. 2. A material having a permeability less than that of vacuum: $\mu < 1$ gauss/oersted. For practical evaluation, a nonmagnetic material.

diamagnetism—1. A phenomenon whereby the magnetization induced in certain substances opposes the magnetizing force. 2. The negative susceptibility exhibited by certain substances. The permeability of such substances is less than unity.

diamond antenna—Also called a rhombic antenna. A horizontal antenna having four conductors that form a diamond, or rhombus.

diamond lattice—The crystal structure of germanium and silicon (as well as a diamond).

diamond stylus—A phonograph pickup with a ground diamond as its point.

diapason—The unique fundamental tone color of organ music.

diaphragm—1. A flexible membrane used in various electroacoustic transducers for producing audio-frequency vibrations when actuated by electric impulses, or electric impulses when actuated by audio-frequency vibrations. 2. In electrolytic cells, a porous or permeable membrane, usually flexible, separating the anode and cathode compartments. 3. In waveguide technique, a thin plate, or plates, placed transversely across the waveguide, not completely closing it, and usually introducing a reactance component. *See also* iris. 4. A sensing element consisting of a membrane placed between two volumes. The membrane is deformed by the pressure differential applied across it.

diathermal apparatus—Apparatus for generating heat in body tissue by high-frequency electromagnetic radiation.

diathermy—1. The therapeutic use of an oscillating electric current of high frequency to produce localized heat in body tissues. 2. The use of radio-frequency fields to produce deep heating in body tissues. The output of a powerful rf oscillator is applied to a pair of electrodes, known as pads, between which the portion of the body to be treated is placed. The body tissues thus become the dielectric of a capacitor, and dielectric losses cause heating of the tissues.

diathermy interference—A form of television interference caused by diathermy equipment, resulting in a horizontal herringbone pattern across the picture.

diathermy machine—A medical apparatus consisting of an rf oscillator frequently followed by rf amplifier stages, used to generate high-frequency currents that produce heat within some predetermined part of the body for therapeutic purposes.

dibit—A group of two bits. In four-phase modulation, each possible dibit is encoded in the form of one of four unique phase shifts of the carrier. The four possible states for a dibit are 00, 01, 10, and 11.

DIC—Dielectrically isolated integrated circuit. Also abbreviated DIIC.

dice—The plural of die.

dichroic—Pertaining to the quality of dichroism.

dichroic filter—1. An optical filter capable of transmitting all frequencies above a certain cutoff frequency and reflecting all lower frequencies, thus being either a high-pass or a low-pass filter. 2. A filter used to selectively transmit light according to its wavelength and not its plane of vibration.

dichroic mirror—1. A special mirror through which all light frequencies pass execpt those for the color that the mirror is designed to reflect. 2. A semitransparent mirror used to selectively reflect light according to its wavelength and not its plane of vibration.

dichroism—Also called dichromatism or polychromatism. 1. A property of an optical material that causes light of some wavelengths to be absorbed when the incident light has its electric-field vector in a particular orientation and not absorbed when the electric-field vector has other orientations. 2. In anisotropic materials, such as some crystals, the selective absorption of light rays vibrating in one particular plane relative to the crystalline axes, but not those vibrating in a plane at right angles thereto. As applied to isotropic materials, this term refers to the selective reflection and transmission of light as a function of wavelength regardless of its plane of vibration. The color of such materials, as seen by transmitted light, varies with the thickness of material examined.

dichromatism—*See* dichroism.

dicing—The process of sawing a crystal wafer into blanks.

dictionary—In digital computer operations, a list of mnemonic code names together with the addresses and/or data to which they refer.

diddle—Automatic transmission of letter or figure characters by a terminal unit if no characters are ready for transmission (most often used with a FIFO memory).

die—1. Sometimes called chip. A tiny piece of semiconductor material, broken from a semiconductor slice, on which one or more active electronic components are formed. (Plural: dice.) 2. A portion of a wafer bearing an individual circuit, or device cut or broken from a wafer containing an array of such circuits or devices. 3. An uncased discrete or integrated device obtained from a semiconductor wafer. *See* chip. 4. A single miniature active or passive component. So named because the circuits are batch fabricated by diffusion processes on a silicon wafer, which is then cut into individual components. Examples: transistors, diodes, integrated circuits, diffused resistors.

die attach—The operation of mounting chips to a substrate. Methods include gold-silicon eutectic bonding, various solders, and conductive (and nonconductive) epoxies.

die bond—1. Attachment of a die or chip to the hybrid substrate. 2. A process in which chips are attached to a substrate (gold, epoxy, wax, etc.). The joint between a die and the substrate.

die bonding—1. The method by which a semiconductor die, or chip, is attached to a mechanical support. 2. Attaching a semiconductor chip to the substrate with an epoxy, eutectic, or solder alloy. 3. The attachment of a die to a gold base, such as a substrate pad, or to a header. Heat, pressure, and a mechanical scrubbing action are used to create a gold-silicon eutectic bond between the die and base.

dielectric—1. The insulating (nonconducting) medium between the two plates of a capacitor. Typical dielectrics are air, wax-impregnated paper, plastic, mica, and ceramic. A vacuum is the only perfect dielectric. 2. A medium capable of recovering, as electrical energy, all or part of the energy required to establish an electric field (voltage stress). The field, or voltage stress, is accompanied by displacement or charging currents. 3. The insulating material between the metallic elements of an electromechanical component or any of a wide range of thermoplastics or thermosetting plastics. 4. Any insulating medium that intervenes between two conductors and permits electrostatic attraction and repulsion to take place across it. 5. A material having the property that energy required to establish an electric field is recoverable, in whole or in part, as electric energy. 6. A material medium in which an electric field can exist in the stationary state. 7. Characteristic of materials that are electrical insulators or in which an electric field can be sustained with a minimum dispersion of power. Such materials exhibit nonlinear properties, such as anisotropy of conductivity or polarization, or saturation phenomena. 8. An insulator. Localized regions of dielectric materials are used in semiconductor devices, for example, to provide electrical isolation between dice, between metal interconnect layers, and between the gate electrode and the channel.

dieletric absorption—Also called dielectric hysteresis (short-term effect), or dielectric soak (long-term effect). 1. A characteristic of dielectrics that determines the length of time a capacitor takes to deliver the total amount of its stored energy. It manifests itself as the reappearance of potential on the electrodes after the capacitor has been discharged. Its magnitude depends on the charge and discharge time of the capacitor. 2. That property of an imperfect dielectric as a result of which all electric charges within the body of the material caused by the application of an electric field are not returned to the field. 3. Reluctance of a capacitor to give up all the electrons stored when the capacitor is discharged. Primarily caused by a polarization effect of dielectric dipoles and to a lesser extent by free electrons in the dielectric requiring a finite time to move to the electrode. The recovery voltage appearing after discharge divided by the charging voltage and expressed as a percentage is called the percent dielectric absorption. 4. The property of a capacitor with slow polarization of its dielectric that results in voltage appearance on the capacitor electrodes after its short-term discharge through a low resistance.

dielectric amplifier—An amplifier employing a device similar to an ordinary capacitor, but with a polycrystalline dielectric that exhibits a ferromagnetic effect.

dielectric analysis—Method of directly monitoring resin cooking, resin staging, and resin curing. Such analysis eliminates many of the variables influencing the selection of a given set of fabrication conditions. Dielectric analyzers consist of a press with heated platens, a clamshell autoclave, and a DTA/dielectric cell.

dielectric anisotropy—The difference between the dielectric along the director and the dielectric perpendicular to the director of a liquid crystal system.

dielectric antenna—An antenna in which a dielectric is the major component producing the required radiation pattern.

dielectric breakdown—1. An abrupt increase in the flow of electric current through a dielectric material as the applied electric field strength exceeds a critical value. 2. A complete failure of a dielectric material characterized by a disruptive electrical discharge through the material due to a sudden and large increase in voltage.

dielectric breakdown voltage—Also called electric breakdown voltage, breakdown voltage, or hi-pot. 1. The voltage between two electrodes at which electric breakdown of the specimen occurs under prescribed test conditions. 2. The voltage at which a dielectric material punctures. 3. The voltage required to cause electrical failure or breakthrough of insulation. Usually expressed as a voltage gradient (volts per mil).

dielectric capacity—The inductivity or specific inductive capacity of a substance, being its ability to convey the influence of an electrified body.

dielectric constant—Also called permittivity, specific inductive capacity, or capacitivity. 1. The ratio of the capacitance of a capacitor with the given dielectric to the capacitance of a capacitor having air for its dielectric but otherwise identical. Symbol: K. 2. That property of a dielectric that determines the electrostatic energy stored per unit volume for a unit potential gradient. 3. The property of a material that determines the amount of electrostatic energy that can be stored when a given voltage is applied.

dielectric crystal—A crystal that is characterized by its relatively poor electrical conductance.

dielectric current—The current flowing at any instant through the surface of an isotropic dielectric that is in a changing electric field.

dielectric dissipation—*See* loss tangent.

dielectric dissipation factor—The cotangent of the dielectric phase angle of a material.

dielectric fatigue—The property of some dielectrics in which the insulating quality decreases after a voltage has been applied for a considerable length of time.

dielectric guide—A waveguide made of a solid dielectric material through which the waves travel.

dielectric heating—A method of raising the temperature of a nominally insulating material by sandwiching it between two plates to which an rf voltage is applied. The material acts as a dielectric, and its internal losses cause it to heat up.

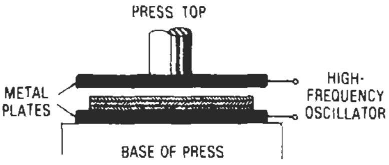

Dielectric heating.

dielectric hysteresis—1. Short-term effect of dielectric absorption. 2. A lagging of an electric field in a dielectric behind the alternating voltage that produces it. It causes a loss similar to that of magnetic hysteresis. 3. *See also* dielectric absorption.

dielectric isolation—1. The electrical isolation of monolithic integrated circuit elements from each other by dielectric material rather than by reverse-biased pn junctions. 2. The use of silicon dioxide barriers created during silicon IC processing to provide isolation between components on a chip.

dielectric lens—A lens used with microwave antennas; it is made of dielectric material so that it refracts radio waves similar to the way an optical lens refracts light waves.

dielectric loss—1. The power dissipated by a dielectric as the friction of its molecules opposes the molecular motion produced by an alternating electric field. 2. The time rate at which electric energy is transformed into heat in a dielectric when it is subjected to a changing electric field.

dielectric loss angle—Also called dielectric phase difference. The complement of the dielectric phase angle (i.e., the dielectric phase angle minus 90°).

dielectric loss factor—Also called dielectric loss index. The product of the dielectric constant of a material times the tangent of the dielectric loss angle.

dielectric loss index—The product of a medium's relative permittivity and the tangent of its dielectric loss angle. *See* dielectric loss factor.

dielectric matching plate—In waveguide technique, a dielectric plate used as an impedance transformer for matching purposes.

dielectric mirror—A highly frequency-selective, multilayer dielectric reflector acting by partial reflection of light at the interface between materials of unequal refractive indices.

dielectric phase angle—The angular difference in phase between the sinusoidal alternating voltage applied to a dielectric and the component of the resultant alternating current having the same period.

dielectric phase difference—*See* dielectric loss angle.

dielectric polarization—*See* polarization, 3.

dielectric power factor—The cosine of the dielectric phase angle, or sine of the dielectric loss angle.

dielectric process—Also called electrographic process. A nonimpact printing technique in which specially treated paper consisting of a conductive base layer coated with a nonconductive thermoplastic material is used to hold an electric charge applied directly by a set of electrode styli. The electric charge corresponds to the latent image of the original. After charging, the image is produced by a toner system similar to that used in electrostatic copying devices. The dielectric process employed on general-purchase nonimpact printers, facsimile devices, and some photocopiers.

dielectric rating—Standard test voltages and frequencies above which failure occurs between specified points in a relay structure.

dielectric resonator—A high-Q, temperature-stable ceramic microwave resonator that is used in microwave oscillator circuits. It can exist in any regular geometrical form and resonates in various modes at frequencies determined by its dimensions and shielding conditions.

dielectric-rod antenna—An antenna in which propagation of a surface wave on a tapered dielectric rod produces an endfire radiation pattern.

dielectric soak—Long-term effect of dielectric absorption. *See* dielectric absorption.

dielectric strength—Also called electric strength, breakdown strength, electric field strength, and insulating strength. 1. The maximum voltage a dielectric can withstand without rupturing. 2. Maximum potential gradient that a dielectric can withstand before it ruptures or a conducting path forms through it. Normally expressed as the ratio of breakdown voltage to the dielectric thickness (volts per mil). 3. The property of a dielectric to withstand an electrical stress. 4. The ratio of dielectric breakdown to the thickness of the insulating materials. 5. The ultimate breakdown voltage of the dielectric or insulation of the resistor when the voltage is applied between the case and all terminals that are tied together. Dielectric strength is usually specified at sea level and simulated high-altitude air pressures.

dielectric susceptibility—The ratio of the polarization in a dielectric to the electric intensity responsible for it.

dielectric tests—1. Tests that consist of the application of a voltage higher than the rated voltage for a specified time for the purpose of determining the adequacy against breakdown of insulating materials and spacings under normal conditions. 2. The testing of insulating materials by the application of a constantly increasing voltage until failure occurs.

dielectric waveguide—A waveguide constructed from a dielectric (nonconductive) substance.

dielectric wedge—A wedge-shaped piece of dielectric material used in one waveguide to match its impedance to that of another waveguide.

dielectric wire—A dielectric waveguide used for short-distance transmission of UHF radio waves between parts of a circuit.

dielectric withstanding voltage—Maximum potential gradient that a dielectric material can withstand without failure.

difference—The signal energy representing the differences in information between the signals in two or more stereo channels. A difference signal is produced when stereo signals differing in electrical polarity or in intensity are mixed together in opposing polarity.

difference amplifier—The basic input stage of most operational amplifiers. An amplifier with an inverting and a noninverting input. The output voltage is a function of the voltage difference between the two inputs. *See* differential amplifier.

difference channel—In a stereophonic sound system, an audio channel that handles the difference between the signals in the left and right channels.

difference detector—A detector circuit in which the output is a function of the difference between the peak or rms amplitudes of the input waveforms.

difference frequency—1. A signal representing, in essence, the difference between the left and right sound channels of a stereophonic sound system. 2. One of the output frequencies of a converter. It is the difference between the two input frequencies.

difference in depth modulation—In directive systems employing overlapping lobes with modulated signals, a ratio obtained by subtracting from the percentage of modulation of the larger signal the percentage of modulation of the smaller signal and dividing by 100.

difference of potential—The voltage or electrical pressure existing between two points. It will result in a flow of electrons whenever a circuit is established between the two points.

difference signal—In a quadraphonic sound system, a signal arrived at by subtracting left or right back-channel signal from its respective front-channel signal. Left front and left back signals, if added to left front minus left back signals, yield a left front channel; if subtracted, they yield a left back channel.

differential—1. A planetary gear system that adds or subtracts angular movements transmitted to two components and delivers the answer to a third. Widely used for adding and subtracting shaft movements in servo systems and for addition and subtraction in computing machines. 2. In electronics, the difference between two levels. 3. A method of signal transmission through two wires that always have opposite states. The signal data is the polarity difference between the wires: whenever either is high, the other is low. Neither wire is grounded. 4. Describing any device whose operation is dependent on the difference between two quantities.

differential amplifier—1. An amplifier with two similar input circuits so connected as to respond to the difference between two voltages (or currents) and effectively suppress voltages or currents that are alike in the two input circuits. 2. A circuit that amplifies the difference between two input signals.

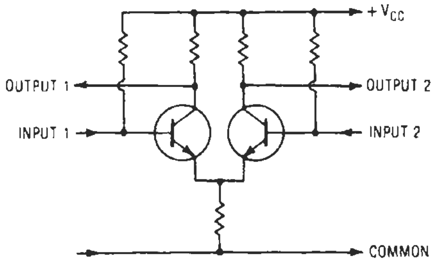

Differential amplifier.

differential analyzer—A mechanical or electrical device primarily designed and used to solve differential equations.

differential angle—The total angle from the operation to the releasing position in a mercury switch.

differential capacitance—The derivative with respect to voltage of a capacitor charge characteristic at a given point on the curve.

differential capacitance characteristic—The function that relates differential capacitance to voltage.

differential capacitor—A variable capacitor having two similar sets of stator plates and one set of rotor plates. When the rotor is turned, the capacitance of one section is increased and the capacitance of the other section is decreased. The sum of the two capacitance values remains substantially constant at all settings, however.

differential comparator—1. A circuit in which differential-amplifier design techniques are applied to the comparison of an input voltage with a reference voltage. When the input voltage is lower than the reference voltage, the circuit output is in one state; when the input voltage is higher than the reference voltage, the output is in the opposite state. Commonly used for pulse-amplitude detector circuits, a-d conversion, and differential receivers for data transmission in noisy environments over a twisted-pair line. 2. A differential circuit for indicating when two input signals are essentially equal, as in a differential pair.

differential cooling—A lowering of temperature that takes place at a different rate at various points on an object or surface.

differential delay—The difference between the maximum and minimum frequency delays occurring across a band.

differential discriminator—A discriminator that passes only pulses having amplitudes between two predetermined values, neither of which is zero.

differential duplex system—A duplex system in which the sent currents divide through two mutually inductive sections of the receiving apparatus. These sections are connected to the line and to a balancing artificial line in opposite directions. Hence, there is substantially no net effect on the receiving apparatus. The received currents pass mainly through one section, or through the two sections, in the same direction and operate the apparatus.

differential equation—A generalization of the algebraic equation in which the unknowns are not simply numbers, but functions of one or more independent variables. Not only the unknown function or functions appear explicitly in the differential equation, but also the first- and possibly higher-order derivatives with respect to the independent variable or variables.

differential flutter—Speed-change errors that occur at different magnitudes, frequencies, or phases across the width of a magnetic tape.

differential gain—1. The ratio of the differential output signal of a differential amplifier divided by the differential input signal causing that output. 2. Variation in the gain of a transmission system with changing modulation.

differential gain control—Also called gain sensitivity control. A device for altering the gain of a radio receiver in accordance with an expected change of signal level in order to reduce the amplitude differential between the signals at the receiver output.

differential galvanometer—A galvanometer having two similar but opposed coils, so that their currents tend to neutralize each other. A zero reading is obtained when the currents are equal.

differential gap—1. The difference between two target values, one of which applies to an upswing of conditions and the other to a downswing. 2. The span between on and off switching points. For example, a room thermostat set for 70° might switch the furnace on at 68° and off at 72°, resulting in a 4° differential. *See also* dead band.

differential gear—In an analog computer, a mechanism that relates the angles of rotation of three shafts. Usually it is designed so that the algebraic sum of the rotations of two shafts is equal to twice the rotation of the third. The device can be used for addition or subtraction.

differential generator—A synchro differential generator driven by a servo system.

differential GPS—A technique using the global positioning satellite (GPS) network, in which a fixed ground GPS receiver (for precisely known three-dimensional coordinates locally) determines corrections to be applied to local aircraft (or other mobile units) using satellite guidance.

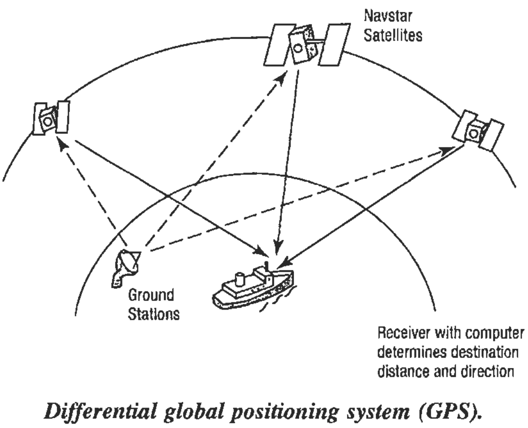
Navstar Satellites
Ground Stations
Receiver with computer determines destination distance and direction

Differential global positioning system (GPS).

differential impedance—The internal impedance observed between the input terminals of an operational amplifier.

differential input—1. An input circuit that rejects voltages that are the same at both input terminals and amplifies the voltage difference between the two input terminals. May be either balanced or floating and may also be guarded. 2. An input applied between two terminals of an operational amplifier, neither of which is at ground (earth) potential.

differential-input amplifier—An amplifier in which the output is ideally a function only of the difference between the signals applied to its two inputs, both signals being measured with respect to a common low or ground reference point.

differential-input capacitance—The capacitance between the inverting and noninverting input terminals.

differential-input impedance—1. The impedance between the inverting and noninverting input terminals of a differential amplifier. 2. The impedance measured between the positive and the negative input terminals of an operational amplifier.

differential-input measurement—Also called floating input. A measurement in which the two inputs to a differential amplifier are connected to two points in a circuit under test and the amplifier displays the difference voltage between the points. In this type of measurement, each input of the amplifier acts as a reference for the other, and ground connections are used only for safety reasons.

differential-input rating—The maximum differential input that may be applied between the two terminals of an operational amplifier.

differential-input resistance—The resistance between the inverting and noninverting input terminals of a differential amplifier. *See* input resistance.

differential-input voltage—The maximum voltage that can be applied across the input terminals of a differential amplifier without damaging the amplifier.

differential-input voltage range—The range of voltages that may be applied between input terminals without forcing the circuit to operate outside its specifications.

differential-input voltage rating—The maximum allowable signal that may be applied between the inverting and noninverting inputs of a differential amplifier without damaging the amplifier.

differential instrument—A galvanometer or other measuring instrument having two circuits or coils, usually identical, through which currents flow in opposite directions. The difference, or differential effect, of these currents actuates the indicating pointer.

differential keying—A method of obtaining chirp-free break-in keying of a cw transmitter by turning the oscillator on quickly before the keyed amplifier stage can pass any signal, and turning it off quickly after the keyed amplifier stage has cut off.

differential linearity—The measure of linearity among digital states in a/d and d/a converters. If the differential linearity is specified as ±½ lsb, the step size from one state to the next may range from ½ to 3/2 of an ideal 1-lsb step.

differential microphone—*See* double-button carbon microphone.

differential-mode gain—Abbreviated DMG. 1. The ratio of the output voltage of a differential amplifier to the differential-mode input voltage. 2. The voltage gain exhibited by an operational amplifier in response to differential-mode signals.

differential-mode input—The voltage difference between the two inputs of a differential amplifier.

differential-mode signal—1. A signal that is applied between the two ungrounded terminals of a balanced three-terminal system. 2. In an amplifier with a differential input, a signal that appears at inverting and noninverting inputs with opposite phase but identical frequency and amplitude. It is not necessarily referred to ground.

differential modulation—A type of modulation in which the choice of the significant condition for any signal element is dependent on the choice for the previous signal element.

differential nonlinearity—The difference between actual analog voltage change and the ideal (1 lsb) voltage change at any code change of a digital-to-analog converter.

differential output voltage—The difference between the values of the two ac voltages that are present in phase opposition at the output terminals of an amplifier when a differential voltage is applied to the input terminals of the amplifier.

differential pair—A pair of transistors sharing a common emitter circuit but with two independent base inputs.

differential permeability—1. The ratio of the positive increase of normal induction to the positive increase of magnetizing force when these increases are minute. 2. The slope of the normal induction curve.

differential phase—1. The difference in phase shift through a television system for a small, high-frequency sine-wave signal at two stated levels of a low-frequency signal on which the first signal is superimposed. 2. In a color TV signal, the phase change of the color subcarrier introduced by the overall circuit, measured in degrees as the picture signal on which it rides is varied from blanking to white level.

differential phase-shift keying—Abbreviated DPSK. A modulation scheme in which the information is conveyed by changes in carrier phase during one interval relative to the preceding interval.

differential pressure transducer—A pressure transducer that accepts simultaneously two independent pressure sources, and the output of which is proportional to the pressure difference between the sources.

differential protective relay—A protective device that functions on a percentage of phase angle or other quantitative difference of two currents or of some other electrical quantities.

differential relay—A relay with multiple windings that functions when the voltage, current, or power difference between the windings reaches a predetermined value. The power difference may result from the algebraic addition of the multiple inputs.

differential resistance—The resistance measured between the terminals of a diode under small-signal and specified bias conditions.

differential resolver—A servo unit with a two-phase stator and a three-phase rotor that is used as a transolver, with the advantage that when connected as a control transformer, the signal does not travel through slip rings.

differential selsyn—A selsyn in which both the rotor and the stator have similar windings that are spread 120° apart. The position of the rotor corresponds to the algebraic sum of the fields produced by the stator and rotor.

differential stage—A symmetrical amplifier stage in which two inputs are balanced against each other so that when there is no input signal, or equal input signals, there is no output signal. An input-signal unbalance, including a signal to only one input, produces an output signal proportional to the difference between the input signals.

differential synchro—*See* synchro differential generator; synchro differential motor.

differential transducer—A device capable of simultaneously measuring two separate stimuli and providing an output proportionate to the difference between them.

differential transformer—Also called linear variable-differential transformer. 1. A transformer used to join two or more sources of signals to a common transmission line. 2. An electromechanical device that continuously translates displacement of position change into a linear ac voltage.

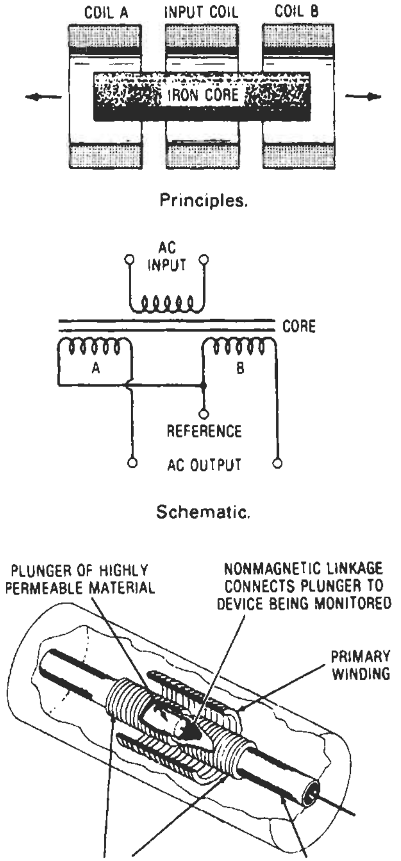
COIL A
INPUT COIL
COIL B
IRON CORE
AC
INPUT
CORE
A
B
REFERENCE
AC OUTPUT
PLUNGER OF HIGHLY PERMEABLE MATERIAL
NONMAGNETIC LINKAGE CONNECTS PLUNGER TO DEVICE BEING MONITORED
PRIMARY WINDING
TWO IDENTICAL, DIFFERENTIALLY CONNECTED SECONDARY WINDINGS
NONMAGNETIC TUBE OR SPOOL

Principles.

Schematic.

Cutaway.

Differential transformer, 1.

differential voltage—For a glow lamp, the difference between the breakdown and maintaining voltage.

differential voltage gain—1. The ratio of the change in output-signal voltage at either terminal of a differential device to the change in signal voltage applied to either input terminal, all voltages measured to ground. 2. The ratio of the differential output voltage of an amplifier to the differential input voltage of the amplifier. If the amplifier has one output terminal, the differential voltage gain is the ratio of the ac output voltage (with respect to ground) to the differential input voltage.

differential winding—A coil winding so arranged that its magnetic field opposes that of a nearby coil.

differential-wound field—A type of motor or generator field having both series and shunt coils connected so they oppose each other.

differentiate—1. To distinguish. 2. To find the derivative of a function. 3. To deliver an output that is the derivative with respect to time of the input.

differentiating circuit—Also called differentiating network and differentiator. A circuit whose output voltage is proportional to the rate of change of the input voltage. The output waveform is then the time derivative of the input waveform, and the phase of the output waveform leads that of the input by 90°. An *RC* circuit gives this differentiating action.

differentiating network—*See* differentiating circuit.

differentiator—*See* differentiating circuit.

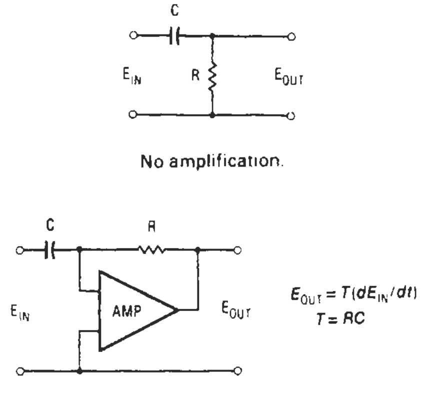

No amplification.

Amplification.

Formula.

Differentiator.

diffracted wave—A radio, sound, or light wave that has struck an object and been bent or deflected other than by reflection or refraction.

diffraction—1. The bending of radio, sound, or light waves as they pass through an object or barrier, thereby producing a diffracted wave. 2. The phenomenon whereby waves traveling in straight paths bend around an obstacle. *See* X-ray crystallography.

diffuse—1. To undergo or cause to undergo diffusion. 2. Light that has been either inadvertently or purposely scattered. Such diffused light propagates in many directions and is not intensely polarized when it illuminates surfaces. With diffused light, a high brightness level may be achieved with minimal glare.

diffused-alloy transistor—Also called drift transistor. A transistor in which the semiconductor wafer is subjected to gaseous diffusion to produce a nonuniform base region, after which alloy junctions are formed in the same manner as for an alloy-junction transistor. It may also have an intrinsic region to give a pnip unit.

diffused-base transistor—Also called graded-base transistor. A type of transistor made by combining diffusion and alloy techniques. A nonuniform base region and the collector-to-base junction are formed by gaseous dissemination into a semiconductor wafer that constitutes the collector region. Then the emitter-to-base junction is formed by a conventional alloy process on the base side of the diffused wafer.

diffused device—A semiconductor device in which a base, usually of silicon, has successive layers of p and n characteristics diffused upon and into the base by means of a series of masks and around which p and n materials, usually phosphorous and boron, adhere to the base by gaseous diffusion in a high-temperature furnace. It is possible to build areas of resistance, capacitance, and active diodes and transistors into the base, creating an entire circuit. Performance is poor in the presence of radiation.

diffused-emitter-and-base transistor — Also called double-diffused transistor. A semiconductor wafer that has been subjected to gaseous dissemination of both n- and p-type impurities to form two pn junctions in the original semiconductor material.

diffused-emitter-collector transistor — A transistor whose emitter and collector are both produced by diffusion.

diffused junction — 1. Type of pn junction, made by using masks to control the diffusion of impurities into monocrystalline semiconductor material. 2. A junction formed by the diffusion of an impurity within a semiconductor crystal.

diffused-junction rectifier — A semiconductor diode in which the pn junction is produced by diffusion.

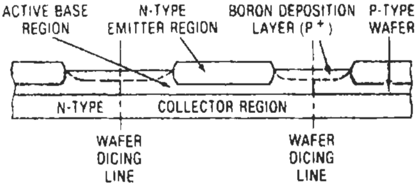

Diffused-junction silicon transistor wafer.

diffused-junction transistor — A transistor in which the emitter and collector electrodes have been formed by diffusion of an impurity into the semiconductor wafer without heating.

diffused-layer resistor — A resistor formed by including an appropriate pattern in the photomask to define diffusion areas.

diffused mesa transistor — A transistor in which the collector-base junction is formed by gaseous diffusion, and the emitter-base junction is formed either by gaseous diffusion or by an evaporated metal strip. The collector-base junction is then defined by etching away the undesired parts of the emitter and base regions, thus producing a mesa.

diffused metal-oxide semiconductor — *See* DMOS.

diffused planar transistor — A transistor made by two gaseous diffusions, but in which the collector-base junction is defined by oxide masking. Junctions are formed beneath this protective oxide layer, with the result that the device has lower reverse currents and good dc gain at low currents.

diffused sound — Sound that has uniform energy density, meaning that the energy flux is equal in all parts of a given region.

diffused transistor — A transistor in which the emitter and collector junctions are both formed by diffusion.

diffused transmission — The total net transmission, by a medium or device, of light that is neither perfectly Lambertian nor parallel. Often used interchangeably with the term *gross transmission*.

diffuser — A device used to scatter or disperse light emitted from a source, usually by the process of diffuse transmission.

diffuse scan — A reflective scan technique in which reflection from a nearby nonshiny surface illuminates a photosensor.

diffuse transmission — Transmission accompanied by diffusion or scatter to the extent that there is no regular or direct transmission.

diffusion — 1. The movement of carriers from a region of high concentration to regions of lower concentration. 2. The tendency of entities such as current carriers to diffuse themselves, or move in directions that increase the uniformity with which their number occupy the available space. Hence, carrier diffusion is a mechanism whereby carriers tend to move downhill along concentration gradients, away from regions of high concentration and toward regions of low concentration. 3. A thermally induced process in which one material permeates another. In silicon processing, doping impurities diffuse into the silicon at elevated temperatures to form the desired junctions. The same impurities penetrate silicon dioxide much more slowly, and therefore silicon dioxide on the surface of the silicon acts as a mask to determine the areas into which diffusion occurs. 4. A high-temperature process involving the movement of controlled densities of n-type or p-type impurity atoms into the solid silicon slice in order to change its electrical properties. 5. The process of adding impurities to a semiconductor material in order to affect its characteristics. 6. One method of modifying the impurity doping of a semiconductor crystal, which makes use of the fact that excited dopant atoms, like carriers, have a tendency to diffuse away from regions of high concentration and toward regions of low concentration. The technique involves prolonged exposure of the semiconductor crystal to a concentrated vapor of the dopant at elevated temperatures, whereupon dopant atoms diffuse into the crystal structure. The resulting doping gradient is roughly exponential, with highest density at the surface.

diffusion and oxidation systems — Equipment in which nonconductive materials are made semiconductive by diffusing controlled amounts of selected impurities into the surface and oxidizing the surface of silicon selectively to provide a protective or insulative layer. Diffusion and oxidation are accomplished by exposing the silicon wafer to specific atmospheres in a high temperature furnace.

diffusion bonding — Formation of a metallurgical joint between similar or dissimilar metals by the process of interdiffusion of atoms across the joint interface in either the solid or liquid state. The term generally is applied to, but is not limited to, solid-state diffusion. The joining surfaces must be brought within atomic distances through the application of pressure.

diffusion capacitance — The capacitance of a forward-biased pn junction.

diffusion constant — The quotient of diffusion-current density in a homogeneous semiconductor, divided by the charge-carrier concentration gradient. It is equal to the drift mobility times the average thermal energy per unit charge of carriers.

diffusion current — 1. The current produced when charges move by diffusion. 2. The flow of a particular type of carrier in a semiconductor due to a concentration difference in that type of carrier. Carriers will flow from an area of high concentration to an area of low concentration.

diffusion furnace — System designed for enclosed elevated temperature processing of solid-state devices and systems in gaseous atmospheres. Diffusion furnaces are operated at temperatures from 1000 to 1300°C to achieve doping of semiconductor substrates by one of a number of processes. Oxidation is a process that puts a protective layer of silicon oxide on the wafer that is used either as an insulator or to mask out certain areas when doping. Deposition systems, of which there are three (liquid, gaseous, solid), are used to deposit impurities on the silicon wafer. Other systems include a drive-in system, used to diffuse impurities into the wafer to a specified level, and an alloy system that is used in a final step of the metallization process.

diffusion length—In a homogeneous semiconductor, the average distance the minority carriers move between generation and recombination.

diffusion process—Doping of a semiconductor material by injection of an impurity into the crystal lattice at an elevated temperature. Usually, the semiconductor crystal is exposed to a controlled surface concentration of dopants.

diffusion transistor—A transistor in which current depends on the diffusion of carriers, donors, or acceptors, as in a junction transistor.

diffusion under the epitaxial film—*See* DUF.

diffusion window—In a semiconductor, the opening etched through the oxide to permit the diffusion of the emitter and base.

digiralt—A system of high-resolution radar altimetry in which pulse-modulated radar and high-performance time-to-digital conversion techniques are combined.

digit—1. One of the symbols, 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9, used in numbering in the scale of 10. One of these symbols, when used in a scale of numbering to the base n, expresses integral values ranging from 0 to $n - 1$ inclusive. 2. A character used to represent a nonnegative integer smaller than the radix, e.g., either 0 or 1 in binary notation or 0 to F in hexadecimal. 3. In a dial telephone system, one of the successive series of pulses incoming from a dial for operation of a switching train.

digital—1. Using numbers expressed in digits and in a certain scale of notation to represent all the variables that occur in a problem or calculation. 2. Of or pertaining to the class of devices or circuits in which the output varies in discrete steps (i.e., pulses or on/off operation). 3. Of or pertaining to an element or circuit whose output is utilized as a discontinuous function of its input. 4. Circuitry in which data-carrying signals are restricted to either of two voltage levels, corresponding to logic 1 or 0. 5. The representation of numerical quantities by means of discrete numbers. It is possible to express in binary digital form all information stored, transferred, or processed by dual-state conditions; e.g., on/off, open/closed, octal, and BCD values. 5. Referring to communications procedures, techniques, and equipment by which information is encoded as either a binary 1 or a binary 0; the representation of information in discrete binary form, discontinuous in time.

digital absorbing selector—A dial switch that sets up and then falls back on the first of two digits dialed; it then operates on the next digit dialed.

digital audio tape—Abbreviated DAT. 1. A digital recording audio format that combines the 16-bit audio quality of compact discs with the recording capability of analog cassettes. 2. A stereo tape format that records sound in digital code and is smaller in size than an audio cassette. A 4-mm DAT drive holds over 1 gigabyte of data and is used as a high-capacity backup medium for computers and as a master source medium for sending data to a CD manufacturer.

digital broadband technology—The technology used in the transmission of digital signals through fiber-optic cables.

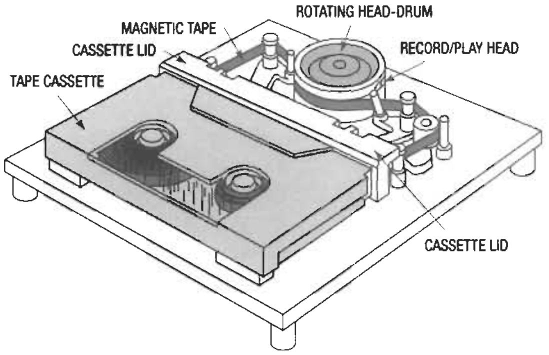

Digital audio tape (DAT) transport system.

digital circuit—1. A circuit that operates like a switch (it is either on or off) and can make logical decisions. It is used in computers or similar decision-making equipment. The more common families of digital integrated circuits (called logic forms) are RTL, DTL, HTL, ECL, and TTL. 2. A circuit that has only two stable states, operating in the manner of a switch, that is, it is either on or off, or high or low (i.e., high voltage or low voltage).

digital clock—A series of synchronized pulses that determine the bit times (data rate) of a digital pattern.

digital communication—1. The transmission of intelligence by the use of encoded numbers—usually uses the binary rather than decimal number system. 2. A system of telecommunications employing a nominally discontinuous signal that changes in frequency, amplitude, or polarity.

digital communications interface equipment—Line interface equipment, including modems.

digital computer—1. An electronic calculator that operates with numbers expressed directly as digits, as opposed to the directly measurable quantities (voltage, resistance, etc.) in an analog computer. In other words, the digital computer counts (as does an adding machine); the analog computer measures a quantity (as does a voltmeter). 2. A computer that processes information in numerical form. Electronic digital computers generally use binary or decimal notation and solve problems by repeated high-speed use of the fundamental arithmetic

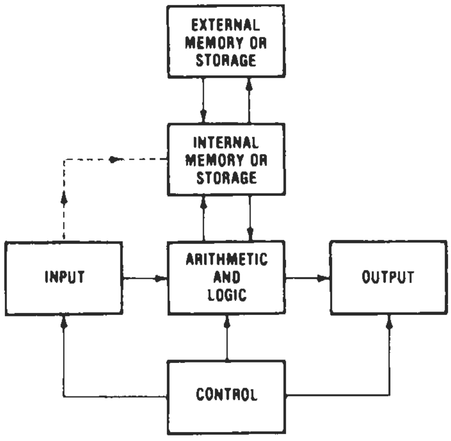

Digital computer block diagram.

processes of addition, subtraction, multiplication, and division. 3. A computer system in which circuit operation is based on specific signal levels. In a binary digital computer, there are two such signal levels, one at or near zero and the other at a defined voltage. 4. A device that performs sequences of arithmetic and logic operations on discrete data. 5. A type of data-processing equipment that counts, utilizing numbers to express the values and quantities. General-purpose digital computers include central storage units and peripheral control units and are designed to solve a wide class of problems. A common feature of general-purpose equipment is the ability to externally modify the program of instructions. Special-purpose digital computers are not intended for a typical commercial physical environment and include rugged computers for military and space applications. An analog computer measures cost or conditions. Hybrid computers utilize both modes. 6. A computer that solves problems by operating on discrete representing variables by performing arithmetic and logic processes on this data.

digital data — 1. Data represented in discrete, discontinuous form, as contrasted with analog data represented in continuous form. Digital data is usually represented by means of coded characters (e.g., numbers, signs, symbols, etc.). 2. Any data that is expressed in digits. The term usually implies the use of binary digits.

digital data-handling system — The electronic equipment that receives digital data, operates on it in a suitable manner, records it in a suitable manner on a suitable medium, and presents it directly to a computer or a display.

digital delay line — *See* active delay line.

digital delay module — *See* active delay line.

digital delay unit — *See* active delay line.

digital device — 1. Typically, an IC that switches between two exclusive states or levels, usually represented by logical 1 or 0. 2. An electronic device that processes electrical signals that have only two states, such as on or off, high or low voltages, or positive or negative voltages. In electronics, *digital* normally means binary or two-state.

digital differential analyzer — A special-purpose digital computer that performs integration and that can be programmed for the solution of differential equations in a manner similar to that of an analog computer.

digital disc recording — An analog disc recording that has been made from a master tape recording that was digitally recorded.

digital filter — 1. A linear computation or algorithm performed on a selected series in the form of an input signal that produces a new series as output. The computational device may be a specifically designed electronic system or a conventional computer. 2. Network that operates on discrete samples of a signal to achieve a desired transfer-function operation on that signal. Digital filters divide into two classes: nonrecursive filters produce an output that is a function of only the previous and present inputs; recursive filters produce an output that is a function of both the past and present inputs and outputs.

digital filtering — 1. A computational process or algorithm by which a sampled signal or sequence of numbers, acting as input, is transformed into a second sequence of numbers called the output. The computational process may correspond to high-pass, low-pass, bandpass, or bandstop filtering, integration, differentiation, or something else. The second sequence can be used for further processing, as in a fast-Fourier-transform analyzer, or it can be converted to an analog signal, producing a filtered version of the original analog signal. 2. The process of smoothing, spectrally shaping, or removing noise from a signal. Digital filters are basically mathematical functions that are performed on the digital data stream; their characteristics can be altered under software control, which adds to their overall flexibility. Finite impulse response (FIR) and infinite impulse response (IIR) are examples of digital filter functions.

digital frequency monitor — A special-purpose digital counter that permits a train of pulses to pass through a gate for a predetermined time interval, counts them, and indicates the number counted.

digital harmonic generation — Abbreviated DHG. The use of circuit elements whose outputs are discontinuous functions of their inputs to produce signals that are an integral multiple of the (fundamental) input signal.

digital image analysis — Technology to measure and standardize the output of a computer-interfaced vidicon system.

digital imaging — The process by which an image that is in electronic form (e.g., a bit-mapped graphic) is altered.

digital information display — The presentation of digital information in tabular form on the face of a digital information display tube.

digital integrated circuit — 1. A switching-type integrated circuit. 2. An integrated circuit that processes electrical signals that have only two states, such as on or off, high or low voltages, or positive or negative voltages. In electronics, *digital* normally means binary or two-state. 3. A monolithic group of logic elements. May be small-scale integration (e.g., SSI gates, flip-flops, latches), medium-scale integration (e.g., MSI decoders, adders, counters), or large-scale integration (e.g., LSI memories, microprocessors). 4. A class of integrated circuits that processes digital information (expressed in binary numbers). The processing operations are arithmetic (such as addition, subtraction, multiplication, and division) or logical (in which the circuit senses certain patterns of input binary information and indicates the presence or absence of those patterns by appropriate output binary signals).

digital integrator — Device for summing or totalizing areas under curves that gives numerical readout. *See also* integrator.

digital logic modules — Circuits that perform basic logic decisions (AND, OR, NOT); used widely for arithmetic and computing functions, flip-flops, half-adders, multivibrators, etc. *See also* logic system.

digitally programmable oscillator — A voltage-controlled oscillator designed to accept a digital tuning word instead of the usual analog signal. Internal digital-to-analog (d/a) converter circuits transform the digital input to an analog voltage. Tuning-curve linearization is usually accomplished through a digital memory. The frequency speed is primarily limited by the d/a circuits.

digital modulation — A method of transmitting human voice or other analog signals using a binary code (0s and 1s). Digital transmission offers a cleaner signal than analog technology.

digital multimeter — Abbreviated DMM. A test instrument used to measure voltage, current, and resistance. The readout of measured values is shown on a digital display which is typically a liquid crystal display (LCD).

digital optical processing — The scanning of photographs or transparencies of images, either by a vidicon camera or flying-spot scanner, for the conversion of the images to digital form for storage on magnetic tape.

digital output — An output signal that represents the size of a stimulus or input signal in the form of a series of discrete quantities that are coded to represent digits in a system of numerical notation. This type of output is to be distinguished from one that provides a continuous output signal.

digital panel meter — Abbreviated DPM. 1. A compact electronic measuring system capable of converting an electrical variable into an unambiguous, accurate numerical reading. Consists of a solid-state analog-to-digital converter and a logic-driven numerical display that can provide readings with accuracies of 0.1 to 0.005 percent of full scale, without the need for eye-straining interpolation. 2. A meter with digital-numerical readout, capable of indicating a range of values from zero to rated maximum, in increments of one least-significant digit.

digital phase shifter — A device that provides a signal phase shift by the application of a control pulse. A reversal of phase shift requires a control pulse of opposite polarity.

digital plotter — An output unit that graphs data via an automatically controlled pen. Data is normally plotted as a series of incremental steps.

digital position transducer — A device that converts motion or position into digital information.

digital programmable delay line — *See* active delay line.

digital radio — 1. Microwave radio in which one or more properties (amplitude, frequency, and phase) of the rf carrier are quantized by the modulating signal. 2. Radio whose instantaneous rf carrier can assume one of a discrete set of amplitude levels, frequency shifts, or phase shifts as a result of the modulating signal. 3. Radio in which the rf carrier, or even the baseband signal, is quantized by means of a modem. In such a radio the rf carrier is still quantized, but baseband or intermediate-frequency filtering can be used to provide some bandwidth control and thereby reduce rf filtering requirements. 4. Any radio that transmits a signal whose informational content is, in whole or in part, digital in format.

digital readout indicator — An indicator that reads directly in numerical form, as opposed to an analog indicator needle and scale.

digital recording — A technique for recording information as discrete points onto magnetic recording media.

digital rotary transducer — A rotating device utilizing an optical sensor that produces a serial binary output as a result of shaft rotation.

digital set top — Also referred to generically as a set-top unit (STU). Television set-top units that accept digital video as well as analog (traditional) video. Some digital set-top units enable interactivity and therefore support shopping, banking, games, etc.

digital signal — 1. An electrical signal with two states — on or off, high or low, positive or negative — such as could be obtained from a telegraph key or two-position toggle switch. *Digital* normally means binary or two-state. 2. Representation of information by a set of discrete values in accordance with a prescribed law. These values are represented by numbers.

digital signals — 1. Discrete or discontinuous signals whose various states are discrete intervals apart. 2. Signals made up of discontinuous pulses whose information is contained in their durations, periods, and/or amplitudes.

digital signature — 1. A numerical representation of a set of logic states, typically used to describe the logic-state history at one device under test output pin during the complete test program. 2. A personal authentication method based on encryption and secret authorization codes used for "signing" electronic documents.

digital simulation — *See* simulation.

digital speech communications — Transmission of voice signals in digitized or binary form.

digital status contact — A logical (on/off) input used mainly to sense the status of remote equipment in process control systems.

digital storage oscilloscope — A special oscilloscope that adapts analog monitoring and recording systems to the capture and analysis of all types of one-shot physical phenomena.

digital switch — 1. A means to interconnect two or more circuits whose information is represented in digital form, using a time-divided network consisting of nonlinear elements. 2. An automatic switching center capable of switching digital signals. It may be a circuit switch or a message switch.

digital switching — Switching of messages digitally by use of integrated electronic circuits for logic and memory, rather than by electromechanical switches.

digital synthesizer — A means of generating several different frequencies without using separate oscillators governed by crystals specially ground for each frequency. A digital synthesizer uses only one reference crystal, a phase-locked loop, and a digital counter to generate a large number of stable frequencies. These circuits are used to reduce dependence on individual crystals, which are relatively expensive.

digital telephone dialer — An automatic telephone dialer that uses a digital code.

digital television — Abbreviated DTV. 1. A television system in which reduction or elimination of picture redundancy is obtained by transmitting only the information needed to define motion in the picture, as represented by changes in areas of continuous white or black. 2. An umbrella term used to describe the digital television system adopted by the FCC in December 1996.

digital thermometer — Electronic temperature-measuring device that reads and/or prints out numerically.

digital-to-analog conversion — The generation of analog (usually variable-voltage) signals in response to a digital code.

digital-to-analog converter — Abbreviated DAC, dac, or d/a converter. 1. A computing device that changes digital quantities into physical motion or into a voltage (i.e., a number output into turns of a potentiometer). 2. A unit or device that converts a digital signal into a voltage or current whose magnitude is proportional to the numeric value of the digital signal. For example:

Digital Input	*Analog Output*
00101 (binary 5)	2 volts
01010 (binary 10)	4 volts
10100 (binary 20)	8 volts

3. A circuit that accepts the discrete, binary outputs of computers and changes them into continuous analog quantities. In general, DACs convert mathematical results into usable electrical quantities. Digital-to-analog converters are used to generate and modulate waveforms, stimulate devices under test, drive motors, or display information. They have applications in process and industrial control systems, cathode-ray-tube displays, and digitally programmed power supplies. Most DACs consist of three major blocks: a precision reference, a set of resistors forming a ladder network, and switches that connect or disconnect the resistor ladder and the reference. Some DACs also include an output amplifier that buffers the current from the ladder and interfaces it to the circuits at the DAC's output. 4. An interface that converts data in a digital form to data in analog form. Used to permit analog output from a digital computer.

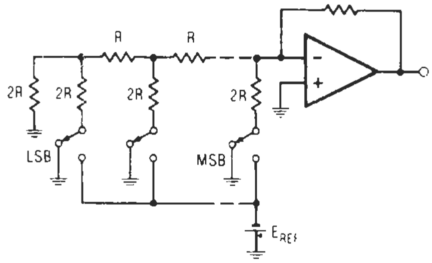

Digital-to-analog converter.

digital-to-disc recording — A recording technique in which a digital master tape is transferred to a conventional master lacquer disc for the manufacturing of phonograph records.

digital transmission — 1. A mode of transmission in which all information to be transmitted over the facility is first converted into digital form and then sent down the line as a stream of pulses. (Such transmission may imply a serial bit stream, but parallel forms are also possible.) When noise and distortion threaten to destroy the integrity of the pulse stream, the pulses are detected and regenerated. 2. The transmission of a signal in which information is represented by a code of discrete elements. Compare with *analog transmission.*

digital versatile disc — *See* DVD.

digital video disc — *See* DVD.

digital voltmeter — 1. An indicator that provides a digital readout of measured voltage rather than a pointer indication. 2. An electronic instrument that converts an analog voltage of unknown magnitude into a digital display of known value.

digit compression — In a computer, any of several techniques used to pack digits.

digitization — 1. The conversion of black-and-white artwork and continuous-tone photographs to a series of binary numbers that can be stored and processed in a computer system. 2. The process of converting analog video, images, or audio into digital format (i.e., 1s and 0s).

digitize — 1. To transform information from the analog (continuous wave-like signal) into a digital format, i.e., into a series of 1s and 0s. 2. To convert an image or signal into digital code for input into a computer; to scan an image, trace a picture on a graphics tablet, or convert camera images. 3. To convert an analog measurement of a physical variable into a number expressed in digits in a scale or notation. 4. To translate a quantitative measurement into a coded numerical equivalent. 5. To convert drawing or picture information into digital form.

digitized image — An image that has been converted into a series of discrete units that are represented in a computer by the binary digits 1 and 0; such an image can then be manipulated for various purposes.

digitized speech — A numerical representation of speech in which the amplitude of the speech waveform has been recorded at regular intervals. Speech is typically sampled from 8000 to 12,500 times per second.

digitizer — 1. A device that converts analog data into numbers expressed in digits in a system of notation. 2. A device that transforms graphical data into planar coordinate information that can be read and understood by a computer. These coordinates are usually presented as *x* and *y* coordinates based on the position of a cursor on the surface, or platen, of the digitizer. The cursor has a viewing area with a crosshair for alignment with the point of interest on the document, and is coupled either mechanically or electrically to a position-sensing device that provides the positional information to the computer. 3. A device that converts coordinate information into numeric form readable by a digital computer. 4. A computer peripheral device that converts an analog electrical signal into numeric form readable by a digital computer. 5. A device that translates input into digital form, to make it possible, for example, to enter sketches into a computer.

digitizing — 1. The process of converting an analog signal to a digital signal. 2. Any method of reducing feature locations on a flat plane to digital representation of *x* and *y* coordinates. 3. The process of converting graphic representations, such as pictures and drawings, into digital data that can be processed by a computer system.

digitron display — In a calculator, a type of display in which all digits appear in the same plane. Similar to mosaic lamp display.

digit selector — In a computer, a device for separating a card column into individual pulses that correspond to punched row positions.

digit-transfer bus — The main wire or wires used to transfer information (but not control signals) among the various registers in a digital computer.

diheptal base — Also called diheptal socket. A vacuum-tube base having 14 pins (such as the base of a cathode-ray tube).

diheptal socket — *See* diheptal base.

DIIC — Dielectrically isolated integrated circuits. Devices isolated from each other by a layer of dielectric insulation, usually glass, rather than by the more conventional reverse-biased pn junction. This "insulated substrate" structure is far more radiation resistant than junction-isolated units, making DIICs valuable in military and aerospace applications.

dimensional stability — 1. The ability of a body to maintain precise shape and size. 2. A measure of dimensional change caused by such factors as temperature, humidity, chemical treatment, age, or stress, usually expressed as units per unit.

dimension ratio *(L/D)* — The ratio of the length of a magnet in the direction of magnetization to its diameter. Or, the ratio of the length of the magnet to the diameter of a circle having an area equal to the cross-sectional area of the magnet. Used as figure of merit to find a magnet's composite permeance coefficient.

diminished-radix complement — *See* radix-minus-one complement.

dimmer — 1. A device for controlling the amount of light emitted by a luminaire. Common types employ resistance, autotransformer, magnetic amplifier, silicon-controlled rectifier or semiconductor, thyratron, or iris control elements. 2. An electric or electronic device that regulates the voltage going to a light source as a means of varying the intensity of the light emitted by the source.

dimmer curve — The performance characteristic of a light dimmer expressed as a graph of the light output of a dimmer-controlled lamp versus the setting of the control in terms of an arbitrary linear scale of 0 to 10.

DIN — 1. The abbreviation for the association in West Germany that determines the standards for electrical and other equipment in that country, Deutsches Institut für Normung (German Institute for Standards). Similar to the American National Standards Institute. 2. A set of standards and specifications promulgated by German manufacturers and covering such audio-related matters as connectors, frequency weighting, measurement techniques, and specifications.

D-indicator — A radar indicator that combines types B and C indicators. The signal appears as a bright spot, with azimuth angle as the horizontal coordinate and elevation angle as the vertical coordinate. Each horizontal trace is expanded vertically by a compressed time sweep to facilitate separation of the signal from noise and to give a rough range indication.

DIN jack—A system of multipin jacks and plugs allowing several connections to be made at once. Named after the German Institute for Standards (DIN).

diode—1. An electron tube having two electrodes: a cathode and an anode. 2. *See* crystal diode. 3. A two-element electron tube or solid-state device. Solid-state diodes are usually made of either germanium or silicon and are primarily used for switching purposes, although they can also be used for rectification. Diodes are usually rated at less than one-half ampere. 4. A two-terminal electronic device that will conduct electricity much more easily in one direction than in the other. 5. A semiconductor device with two terminals and a single junction, exhibiting varying conduction properties depending on the polarity of the applied voltage. 6. A two-terminal semiconductor device exhibiting a nonlinear voltage-current characteristic; it has the asymmetrical voltage-current characteristic exemplified by a single pn junction.

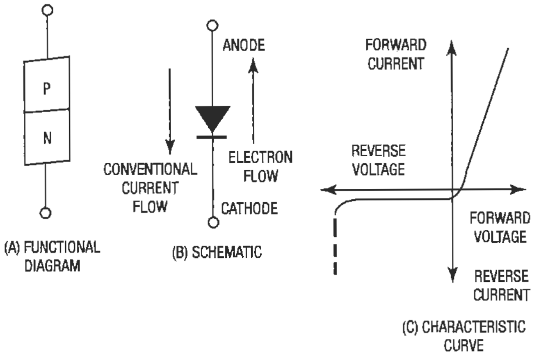

Diode.

diode amplifier—A parametric amplifier that uses a special diode in a cavity. Used to amplify signals at frequencies as high as 6000 MHz.

diode assembly—A single structure of more than one diode.

diode characteristic—The composite electrode characteristic of a multielectrode tube, taken with all electrodes except the cathode connected together.

diode demodulator—Also called diode detector. A demodulator in which one or more semiconductor or electron-tube diodes are used to provide a rectified output that has an average value proportional to the original modulation.

diode detector—*See* diode demodulator.

diode gate—An AND gate that uses diodes as switching elements.

diode isolation—A method in which a high electrical resistance between an IC element and the substrate is obtained by surrounding the element with a reverse-biased pn junction.

diode laser—Also called laser diode, injected laser, coherent electroluminescence device, semiconductor laser. A pn junction semiconductor electron device that converts direct forward-bias electrical input (pump power) directly into coherent optical output power via a process of stimulated emission in the region near the junction.

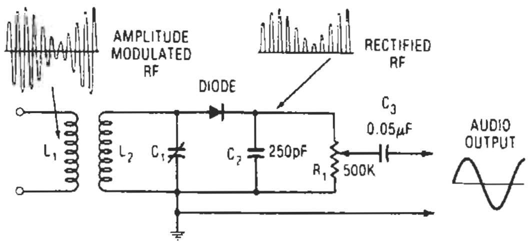

Diode detector.

diode limiter—A circuit employing a diode and used to prevent signal peaks from exceeding a predetermined value.

diode logic—An electronic circuit using current-steering diodes in an arrangement such that the input and output voltages have relationships that correspond to AND or OR logic functions.

diode matrix—1. A two-dimensional array of diodes used for a variety of purposes, such as decoding and read-only memory. 2. A hardware pattern in which diode leads may be inserted to change solid-state control logic.

diode mixer—A diode that mixes incoming radio-frequency and local-oscillator signals to produce an intermediate frequency.

diode modulator—A modulator in which one or more diodes are employed to combine a modulating signal with a carrier signal. It is used chiefly in low-level signaling because it has inherently poor efficiency.

diode pack—A combination of two or more diodes integrated into a solid block.

diode peak detector—A diode used in a circuit to indicate when audio peaks exceed a predetermined value.

diode-pentode—A vacuum tube having a diode and a pentode combined in the same envelope.

diode rectification—The conversion of an alternating current into a unidirectional current by means of a two-element device such as a crystal, vacuum tube, etc.

diode switch—A diode in which positive and negative biasing voltages (with respect to the cathode) are applied in sucession to the anode in order to pass and block, respectively, other applied waveforms within certain voltage limits. In this way, the diode acts as a switch.

diode-transistor logic—Abbreviated DTL. 1. A logic circuit that uses diodes at the input to perform the electronic logic function that activates the circuit transistor output. In monolithic circuits, the DTL diodes are a positive-level logical AND function or a negative-level OR function. The output transistor acts as an inverter to result in the circuit becoming a positive NAND or a negative NOR function. 2. Any logic gate circuit that uses several diodes to perform the AND or OR function, followed by one or more transistors to add power to (and possibly invert) the output. Formerly very popular in digital systems, but now largely superseded by TTL circuits. 3. Logic employing diodes at the input with transistors used as amplifiers and resistor pull-up on the output.

diode-triode—A vacuum tube having a diode and triode combined in the same envelope.

diopter—1. The unit of optical measurement that expresses the refractive power of a lens or prism. 2. A measure of lens power equal to the reciprocal of the lens focal length in meters.

dip—1. A drop in the plate current of a class C amplifier as its tuned circuits are being adjusted to resonance. 2. The angle between the direction of the earth's magnetic field and the horizontal as measured in a vertical plane.

DIP—Abbreviation for dual in-line package.

dip coating—1. A method of applying an insulating coating to a conductor by passing it through an applicator containing the insulating medium in liquid form. The insulation is then sized and passed through ovens to solidify. This medium can be used for magnet wire. 2. A process of applying a relatively thin (less than 50 mils) conformal coating to a part or assembly. The final coating thickness is determined by the viscosity of the coating material, rate of withdrawal, temperature, and the number of coats. 3. Method in which an object is coated by dipping into a plastisol or organosol.

dip encapsulation—A type of conformal coating. An embedding process in which the insulating material is applied by immersion and without the use of an outer container. The coating conforms generally with the contour of the embedding part or assembly.

diplexer—1. A coupling unit that allows more than one transmitter to operate together on the same antenna. 2. A device that enables two (radio) transmitters operating at different frequencies to use the same antenna simultaneously.

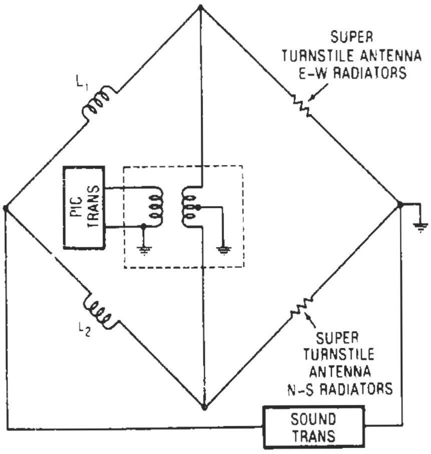

Diplexer circuit.

diplex operation—1. The simultaneous transmission or reception of two messages from a single antenna or on a single carrier. 2. The operation of two radio transmitters on different frequencies into the same antenna simultaneously.

diplex radio transmission—Simultaneous transmission of two signals by using a common carrier wave.

diplex reception—The simultaneous reception of two signals having some feature in common, for example, a single receiving antenna or a single carrier frequency.

dipole—1. A molecule that has an electric moment. For a molecule to be a dipole, the effective center of the positive charges must be at a different point than the center of the negative charges. 2. A form of speaker that radiates in approximately equal amounts to the rear and the front. 3. *See* dipole antenna.

dipole antenna—Also called dipole. A straight radiator usually fed in the center. Maximum radiation is produced in the plane normal to its axis. The length specified is the overall length.

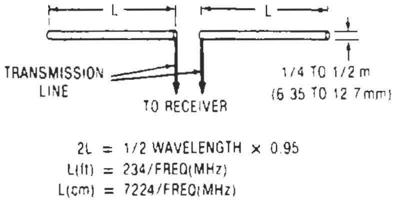

Dipole antenna.

dipole disc feed—An antenna, consisting of a dipole near a disc, used to reflect energy to the disc.

dipping—The process of impregnating or coating insulating materials or windings by the simple method of immersion in the liquid insulating material. A step in the process of treating insulating materials or electrical components by immersion in a liquid insulation, followed by draining and curing to provide increased electrical and mechanical protection.

dip soldering—1. The process of soldering component leads, terminals, and hardware to the conductive pattern on the bottom of a printed circuit board by dipping that side into molten solder or floating it on the surface. 2. A process of joining metals, previously cleaned and fluxed, by immersing them wholly or partially into molten solder. The filling of the joint is by capillary attraction. 3. A process whereby items to be soldered are brought in contact with the surface of a static pool of molten solder for the purpose of soldering the entire exposed conductive pattern in one operation. 4. Soldering by dipping fluxed and fixtured parts in a solder pot. 5. The simplest form of mass soldering. It involves the lowering of a prefluxed assembly onto a solder bath surface. The assembly is then submerged sufficiently for the solder to spread and form the required joints but not flow over the top surface of a printed wiring board.

dip solder terminal—The terminals on a connector that are inserted into holes in the printed circuit board and then soldered in place.

direct—A method of expressing an absolute address in an MPU instruction where the actual address would be specified in hexadecimal in the instruction.

direct access—The process of storing data in, or getting data from, a storage device in such a manner that surrounding data need not be scanned to locate the desired data. The time required to get desired data from the storage device is independent of the location of the data.

direct-access device—*See* random-access device.

direct-access file—A file in which each record may be accessed directly, regardless of its relative position in the file.

direct-access storage—Also called random-access storage. Pertaining to the process of obtaining data from or placing data into storage in which the time required for such access is independent of the location of the data most recently obtained or placed in storage.

direct-acting recording instrument—An instrument in which the marking device is mechanically connected to or directly operated by the primary detector.

direct address—An address that specifies the location in a computer of an instruction operand.

direct addressing—The standard addressing mode in a computer. It is characterized by an ability to reach any point in main storage directly. Direct addressing is sometimes restricted to the first 256 bits in main storage.

direct broadcast satellite—Abbreviated DBS. 1. A satellite that allows the use of inexpensive home reception dishes to receive its high-power signals, including software and data. 2. A term commonly used to describe Ku-band broadcasts via satellite directly to individual end users. The DBS band ranges from 11.7 to 12.75 GHz.

direct capacitance — 1. The capacitance between two conductors, excluding stray capacitance that may exist between the two conductors and other conducting elements. 2. The capacitance measured directly from conductor to conductor through a single insulating layer.

direct-connect modem — A modem that connects directly to a phone line via modular connectors, rather than going through a telephone headset and an acoustic coupler.

direct-coupled amplifier — 1. A direct-current amplifier in which the plate of one stage is coupled to the grid of the next stage by a direct connection or a low-value resistor. 2. An amplifier in which the output of one stage is connected to the input of the next stage without the use of intervening coupling components.

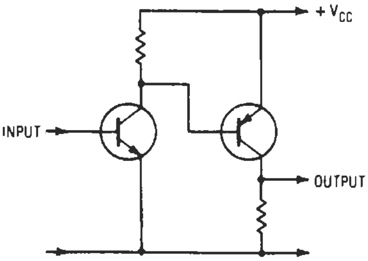

Direct-coupled amplifier.

direct-coupled transistor logic — Abbreviated DCTL. A NOR-gate type of bipolar logic in which the output of one gate is coupled directly to the input of the succeeding gate. This form of logic evolved into resistor-transistor logic because of the difficulty in mass producing transistors within the close tolerances necessary for direct coupling.

direct coupling — 1. The association of two or more circuits by means of an inductance, a resistance, a wire, or a combination of these so that both direct and alternating currents can be coupled. 2. Interstage coupling or speaker coupling with no intervening transformer or capacitor. To the speaker it ensures that the damping factor remains high at low frequencies (but increasing power supply impedance at low frequencies can influence this), while direct coupling generally minimizes low-frequency phase shift and encourages enhanced bass performance. 3. The connection of two circuits in such a way that both ac and dc currents can flow. Neither a transformer nor a capacitor can be used in series with the signal-carrying wires since these components do not pass dc.

direct current — Abbreviated dc. 1. An essentially constant-value current that flows in only one direction. 2. A flow of continuous electric current in one direction as long as the circuit is closed (as opposed to alternating current). 3. A current that flows in only one direction in an electric circuit. It may be continuous or discontinuous, and may be constant or varying.

direct-current amplifier — Also called dc amplifier. An amplifier capable of boosting dc voltages. Resistive coupling only is generally employed between stages, but sometimes will be combined with other forms.

direct-current erasing head — A head that uses direct current in magnetic recording to produce the magnetic field required for erasure. Direct-current erasing is achieved by subjecting the medium to a unidirectional field. Such a medium is therefore in a different magnetic state from one erased by alternating current.

direct-current generator — 1. A rotating machine that changes mechanical into electrical energy in the form of direct current. This is accomplished by commutating bars on the armature. The bars make contact with stationary brushes, from which the direct current is taken. 2. A rotary machine having a commutator that generates direct current when its armature is rotated in a magnetic field by an engine or motor. The commutator reverses the alternating current generated in the armature winding to produce direct current at the machine's output terminals.

direct-current resistance — Abbreviated DCR. The resistance offered by any circuit to direct current.

direct-current restorer — The means by which a direct-current or low-frequency component is reinserted after transmission. Used in a circuit incapable of transmitting slow variations, but capable of transmitting components of higher frequency.

direct-current transmission — Of television, that form of transmission in which a fixed setting of the controls makes any instantaneous value of signal correspond to the same value of brightness at all times.

direct data entry — The entry of data directly into a computer through machine-readable source documents or through the use of online terminals.

direct digital control — Time-sharing, or multiplexing, of a computer among many controlled loops.

direct distance dialing — Abbreviated DDD. 1. A telephone exchange service that enables the telephone user to call other subscribers outside his or her local area without operator assistance. 2. Direct distance dialing by subscribers over the nationwide intertoll telephone network. *See also* area code.

direct drive — A drive system used to rotate a turntable, in which the platter is driven directly by the motor shaft at the exact speed required. These designs usually include electronic motor control.

direct-drive torque motor — A servoactuator that can be directly attached to the load it is to drive. It converts electrical signals directly into sufficient torque to maintain the desired accuracy in a positioning or speed control system.

direct electromotive force — A unidirectional electromotive force in which the changes in values are either zero or so small that they may be neglected.

direct grid bias — The dc component of grid voltage; commonly called grid bias.

direct-insert subroutine — *See* open subroutine.

direct inward dialing — A service allowing outside parties to call directly to extensions on the customer's premises.

direction — The position of one point in space with respect to another.

directional — Having radiative characteristics that vary with direction.

directional antenna — An antenna that radiates radio waves more effectively in some directions than in others. (The term is usually applied to an antenna whose directivity is larger than that of a half-wave dipole.)

directional beam — An electromagnetic wave that is concentrated in a given direction.

directional coupler — 1. A junction consisting of two waveguides coupled together in such a manner that a traveling wave in either guide will induce a traveling wave in the same direction in the other guide. 2. A bilateral electrical network that may be used as a hybrid power splitter, power adder, or mixer.

directional filter—Also called directional separation filter. A filter used to separate the two frequency ranges in a carrier system in which one range of frequencies is used for transmission in one direction and another range of frequencies for transmission in the opposite direction.

directional gain—*See* directivity index.

directional homing—The procedure of following a path in such a way that the target is maintained at a constant relative bearing.

directional hydrophone—A hydrophone having a response that varies significantly with the direction of incidence of sound.

directional lobe—*See* lobe.

directional microphone—1. A microphone whose response varies significantly with the direction of sound. *See also* bidirectional microphone; semidirectional microphone; unidirectional microphone. 2. A microphone that is more sensitive to sounds coming from certain directions than to sounds coming from other directions. Such microphones can be aimed so their most sensitive sides face the sound source, while their least sensitive sides face sources of noise or other undesired sound. *See* cardioid microphone; figure-8 microphone.

directional pattern—Also called radiation pattern. A graphical representation of the radiation or reception of an antenna as a function of direction. Cross sections are frequently given as vertical and horizontal planes, and principal electric and magnetic polarization planes.

directional phase shifter—A passive phase-shifting device in which the phase change for transmission in one direction is different from the phase change for transmission in the opposite direction.

directional power relay—A device that functions on a desired value of power flow in a given direction, or upon reverse power resulting from arcback in the anode or cathode circuits of a power rectifier.

directional relay—A relay that functions in conformance with the direction of power, voltage, current, pulse rotation, etc. *See also* polarized relay.

directional separation filter—*See* directional filter.

direction angle—The angle between the antenna base line and a line connecting the center of the base line with the target.

direction cosine—The cosine of the angle between the base line and the line from the center of the base line to the target.

direction finder—Abbreviated df. Also called radio compass. Apparatus for receiving radio signals and taking their bearings in order to determine their points of origin.

direction finding—The principle and practice of determining a bearing by radio means, using a discriminating antenna system and a radio receiver so that the direction of an arriving wave, and ostensibly the direction or bearing of a distant transmitter, can be determined.

direction of lay—1. The lateral direction in which the strands of a conductor run over the top of the cable conductor as they recede from an observer looking along the axis of the conductor or cable. Also applies to twisted cable. 2. The direction, either right-hand (clockwise) or left-hand (counterclockwise), in which a conductor or group of conductors spiral around a cable core as they travel away from the observer.

direction of polarization—For a linearly polarized wave, the direction of the electrostatic field.

direction of propagation—At any point in a homogeneous, isotropic medium, the direction of the time-average energy flow. In a uniform waveguide, the direction of propagation is often taken along the axis. In a uniform lossless waveguide, the direction of propagation at every point is parallel to the axis and in the direction of time-average energy flow.

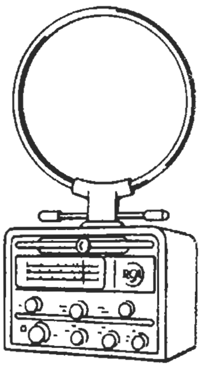

Direction finder.

direction rectifier—A rectifier that supplies a direct-current voltage, the magnitude and polarity of which are determined by the magnitude and relative polarity of an alternating-current selsyn error voltage.

directive gain—In a given direction, 4π times the ratio of the radiation intensity to the total power radiated by the antenna.

directivity—1. The property that causes an antenna to radiate or receive more energy in some directions than in others. 2. The value of the directive gain of an antenna in the maximum-gain direction. 3. The ability of an antenna to pick up signals from one general direction (usually from the front) and effectively reject those from other directions (usually from the back and sides). The front-to-back ratio is one measure of an antenna's directivity. 4. A tendency for some microphones to respond less strongly to sounds arriving from the sides and/or rear. Directional microphones are useful in discriminating on the basis of direction between wanted sounds (musical instruments) and unwanted sounds (audience noises). Directivity is typically graphed on a polar pattern, and is thus classed as nondirectional (omnidirectional), bidirectional (figure-8), or unidirectional (cardioid), supercardioid, or hyperdirectional.

directivity diagram of an antenna—The graphical representation of the gain of an antenna in the different directions of space.

directivity factor—1. In acoustics, the directivity factor is equivalent to directivity, as applied to an antenna. 2. Of a transducer used for sound emission, the ratio of the intensity of the radiated sound at a remote point in a free field on the principal axis to the average intensity of the sound transmitted through a sphere passing through the remote point and concentric with the transducer. 3. Of a transducer used for sound reception, the ratio of the square of the electromotive force produced in response to sound waves arriving in a direction parallel to the principal axis of the transducer to the mean square of the electromotive force that would be produced if sound waves having the same frequency and mean square pressure were arriving at the transducer simultaneously from all directions with random phase. The frequency should be specified in both cases.

directivity index—Also called directional gain. A measure of the directional properties of a transducer. It is the ratio, in decibels, of the average intensity or response over the whole sphere surrounding the projector or hydrophone to the intensity or response on the acoustic axis.

directivity of a directional coupler—Ratio of the power measured at the forward-wave sampling terminals with only a forward wave present in the transmission line to the power measured at the same terminals when the forward wave reverses direction. This ratio is usually expressed in dB and would be infinite for a perfect coupler.

directivity of an antenna—The ratio of the maximum field intensity to the average field intensity at a given distance, implying a maximum value.

directivity pattern—A plot of the response of an electroacoustic transducer as a function of direction.

directivity signal—A spurious signal present in the output of any coupler because its directivity is not infinite.

direct light—Light from a luminous object such as the sun or an incandescent lamp, as opposed to reflected light.

direct lighting—A system of lighting that delivers a majority of light in useful directions without being deflected from the ceiling or walls. Any lamp equipped with a glass or metal reflector arranged to reflect the light toward the object to be illuminated is classified as direct lighting.

directly grounded—*See* solidly grounded.

directly heated cathode—A wire, or filament, designed to emit the electrons that flow from cathode to plate. This is done by passing a current through the filament; the current heats the filament to the point at which electrons are emitted. In an indirectly heated cathode, the hot filament raises the temperature of a sleeve around the filament; the sleeve then becomes the electron emitter.

direct material—A semiconductor material in which electrons move directly from the conduction band to the valence band to recombine with holes. The process of recombination conserves energy and momentum.

direct memory access—Abbreviated DMA. 1. A technique that permits a peripheral device to enter or extract blocks of data from a microcomputer memory without involving the central processing unit. In some cases, a CPU can perform other functions while the transfer occurs. 2. A method of I/O data transfer that does not alter minicomputer instruction-execution flow. The peripheral "steals" memory or CPU cycles to transfer data. 3. A mechanism that allows an input/output device to take control of the CPU for one or more memory cycles in order to write into memory or read from memory. The order of executing the program steps (instructions) remains unchanged. 4. Direct access to a block of memory by more than one system. 5. The technique generally used to transfer blocks of data between a peripheral and random-access memory. It is called direct because the host does not handle the data during the transfer operation. 6. A method of transferring blocks of data directly between an external device and system memory without the need for CPU intervention. This method significantly increases the data transfer rate and hence system efficiency. *See also* cycle stealing.

direct metal mask—A metal mask made by etching a pattern into a sheet of metal.

direct numerical control—A system connecting a set of numerically controlled machines to a common memory for part program or machine program storage with provision for on-demand distribution of data to the machines. Direct numerical control systems have additional provisions for collection, display, or editing of part programs, operator instructions, or data related to the numerical control process.

director—1. A parasitic antenna element located in the general direction of the major lobe of radiation for the purpose of increasing the radiation in that direction. 2. Equipment in common-carrier telegraph message switching systems, used to make cross-office selection and connection from an input-line to an output-line equipment in accordance with addresses in the message. 3. A telephone switch that translates the digits dialed into the directing digits actually used to switch the call. 4. Electromechanical equipment that is used to track a moving target in azimuth and angular height and that, with the addition of other necessary information from an outside source such as a radar set or a range finder, continuously computes firing data and transmits them to the guns. 5. In a machine-tool or process control system, the part of the system that receives the command signals from a controller and converts and amplifies these signals to make them usable by the control devices in the machine or process. 6. In a liquid crystal system, a local symmetry axis around which the long-range order of the crystal is aligned. For the nematic phase, the molecular long axis is—on average—parallel to the director.

directory—Also called a catalog. 1. A table of contents designed to allow convenient access to specific files. 2. A file containing information concerning the other files on a mass-storage device. 3. In the logical format of a disk or disc, a branch of the information tree containing other directories (subdirectories) and/or files. 4. An index of the files on a disk. A directory can contain individual files in addition to other directories.

direct outward dialing—Abbreviated DOD. 1. Dialing of a call into the city system from a PAX/PABX extension without the help of an operator. Usually accomplished by first dialing the digit 9. 2. The dialing of a call from a local system into a toll network without the help of an operator.

direct pickup—Transmission of television images without resorting to an intermediate magnetic or photographic recording.

direct piezoelectricity—A name sometimes given to the piezoelectric effect in which an electric charge is developed on a crystal by the application of mechanical stress.

direct point repeater—A telegraph repeater in which the relay controlled by the signals received over a line sends corresponding signals directly into another line or lines without the use of any other repeating or transmitting apparatus.

direct radiative transition—A transition that involves photons alone. *See* transition, 1.

direct radiator—A speaker that is not horn loaded. The term usually refers to a cone-type speaker, as opposed to a compression driver/horn assembly.

direct-radiator speaker—A speaker in which the radiating element acts directly on the air instead of relying on any other element such as a horn.

direct recording—1. The production of a visible record without subsequent processing, in response to received signals. 2. Analog recording in which continuous amplitude variations are recorded linearly through the use of ac bias.

direct-recording magnetic tape—A method of recording using a high frequency bias in which the electrical input signal is applied to the recording head without alteration.

direct/reflected speaker—A form of speaker in which a small part of the total output is radiated directly forward, with the major part reflected from the wall behind the speaker.

direct resistance-coupled amplifier—An amplifier in which the plate of one stage is connected either directly or through a resistor to the control grid of the next stage, with the plate-load resistor being common to both stages. Used to amplify small changes in direct current.

direct route — In wire communications, the trunks that connect two switching centers, regardless of the geographical path the actual trunk facilities may follow.

direct scanning — A scanning technique in which the object is illuminated the entire time, and in which picture elements of the object are viewed singly by the television camera.

direct sound wave — A wave emitted from a source in an enclosure prior to the time the wave has undergone its first reflection from a boundary of the enclosure. Frequently a sound wave is said to be direct if it contains reflections that have occurred from surfaces within about 0.05 second after the sound was first emitted.

direct synthesizer — 1. A frequency synthesizer producing an output frequency that is related to the reference frequency by the ratio of two integers. The primary advantage of this type of synthesizer lies in its ability to change output frequencies at a moderately fast rate (typically in the microsecond range) and in a random way. Principal applications include frequency-agile radars, secure communication links, and electronic countermeasures. 2. Derives an output from one or more fixed-frequency reference oscillators, using combinations of frequency division, multiplication, mixing, summing, and filtering.

direct-to-disc recording — A technique in which a live performance is recorded directly onto the master lacquer disc for manufacturing of phonograph records.

direct voltage — Also called dc voltage. A voltage that forces electrons to move through a circuit in the same direction and thereby produce a direct current.

direct wave — 1. A wave that is propagated directly through space, as opposed to one that is reflected from the sky or ground. 2. A radio wave that travels from the transmitting antenna to the point of reception without reflection or refraction.

direct Wiedemann effect — *See* Wiedemann effect.

direct-wire circuit — A supervised protective signaling circuit usually consisting of one metallic conductor and a ground return and having signal-receiving equipment responsive to either an increase or a decrease in current.

direct writing galvanometer recorder — Recorder using a pen attached directly to a galvanometer movement for direct writing of signals of frequencies up to about 300 Hz.

disable — To prevent the passage of binary signals by application of the proper signal to the disable terminal of a device.

disassemblers — Programs that do the opposite of compiler programs. Given a machine-code program listing, the disassembler turns it back into an assembly listing, with mnemonic representations, for troubleshooting purposes.

disassembly — 1. Retranslation of machine language into mnemonics during debugging. 2. Translation of binary machine code into assembly-language statements.

disc — 1. A phonograph record. 2. The blank used in a recorder. *See also* disk.

disc capacitor — A small disc-shaped capacitor with a ceramic dielectric, generally used for bypassing or for temperature compensation in tuned circuits.

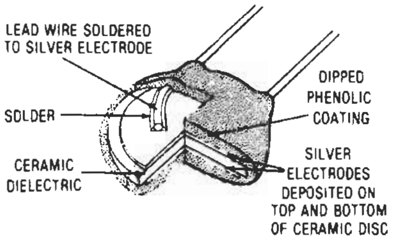
LEAD WIRE SOLDERED TO SILVER ELECTRODE
SOLDER
CERAMIC DIELECTRIC
DIPPED PHENOLIC COATING
SILVER ELECTRODES DEPOSITED ON TOP AND BOTTOM OF CERAMIC DISC

Disc (ceramic) capacitor.

disc files — A type of storage medium consisting of numbers of discs that rotate; each disc has a special coating for retaining stored information.

disc generator — A capacitive-charge type of voltage generator.

discharge — 1. In a storage battery, the conversion of chemical energy into electrical energy. 2. The release of energy stored in a capacitor when a circuit is connected between its terminals. 3. The conversion of dielectric stress of a capacitor into an electric current.

discharge breakdown — Breakdown of a material as a result of degradation due to gas discharges.

discharge key — A device for switching a capacitor suddenly from a charging circuit to a load through which it can discharge.

discharge lamp — A lamp containing a low-pressure gas or vapor that ionizes and emits light when an electric discharge is passed through it. Fluorescent materials are sometimes used on the inside of the glass envelope to

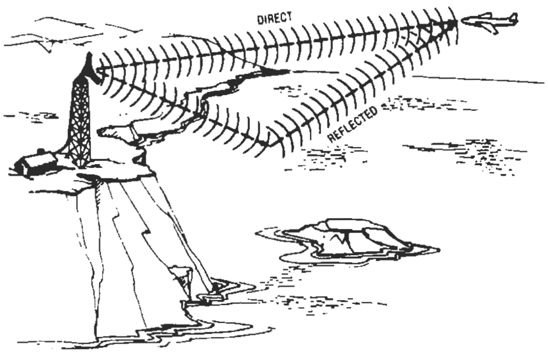

Direct waves.

increase the illumination, as in an ordinary fluorescent lamp.

discharge rate—The amount of current a battery will deliver over a given period of time. A slower discharge rate generally results in more efficient use of a battery.

discharge tube—A tube containing a low-pressure gas that passes a current whenever sufficient voltage is applied.

discharge voltage—Also called clamping voltage. The maximum peak voltage measured across suppressor device terminals when subjected to peak pulse current.

discone antenna—A special form of biconical antenna in which the vertex angle of one cone is 180°.

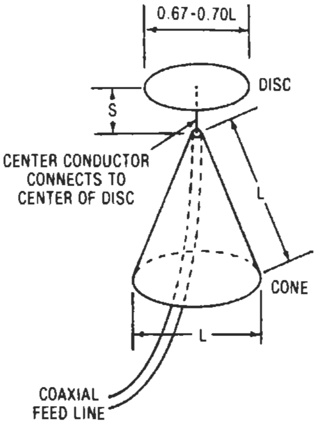

Discone antenna.

disconnect—Also called release. 1. To break an electric circuit. 2. To remove the power from an electrical device (colloquially, "to unplug the device"). 3. To disengage the apparatus used in a connection and to restore it to its ready condition when not in use. 4. A device or group of devices that removes electrical continuity from between the conductors of a circuit and the source of supply.

disconnecting means—A device whereby the current-carrying conductors of a circuit can be disconnected from their source of supply.

disconnector release—A device that disengages the apparatus used in a telephone connection to restore it to the condition in which it exists when not in use.

disconnect signal—1. A signal sent from one end of a trunk or subscriber line to indicate at the other end that the established connection should be released. 2. The on-hook signal in a telephone system by which the calling and called terminals notify the switching equipment that an established connection is no longer needed and should be released.

disconnect switch (motor circuit switch)—A switch intended for use in a motor branch circuit. It is rated in horse power and is capable of interrupting the maximum operating overload current of a motor of the same rating at the rated voltage.

discontinuity—1. A broken connection, or the loss of a specific connection characteristic. 2. The temporary interruption or variation in current or voltage. 3. A point of abrupt change in the impedance of a circuit, where wave reflections can occur.

discontinuous amplifier—An amplifier that reproduces an input waveform on some type of averaging basis.

disc pack—A set of magnetic discs that can be removed from a disc storage as one unit.

disc recorder—A recording device in which the sounds are mechanically impressed onto a disc; as opposed to a tape recorder, which impresses the sound magnetically on a tape.

discrete—1. An individual circuit component, complete in itself, such as a resistor, diode, capacitor, or transistor, and used as an individual and separable circuit element. 2. Pertaining to distinct elements, such as characters, or to representation by means of distinct elements. 3. A term applied to four channels when there are four electrically independent signals, as opposed to matrix. 4. A quad disc or record-playback method that keeps four signals separate, distinct, and independent from recording to playback. 5. Having an individual identity. Fabricated prior to installation and/or separately packaged, not part of an integrated circuit.

discrete circuit—1. A circuit built from separate components that are individually manufactured, tested, and assembled. 2. Electronic circuit built of separate components (transistors, resistors, etc.) connected by wiring or printed-circuit etched conductors.

discrete component—1. A component that has been fabricated prior to its installation (e.g., resistors, capacitors, diodes, and transistors). 2. A circuit component having an individual identity, such as a transistor, capacitor, or resistor.

discrete device—1. A class of electronic components, such as power MOSFETs, bipolar power transistors, MOVs, optoelectronic devices, rectifiers, power hybrid circuits, intelligent power discretes, and transistors. Typically, these devices contain one active element, such as a transistor or diode. However, hybrids, optoelectronic devices, and intelligent discretes may contain more than one active element. In contrast, integrated circuits typically contain hundreds, thousands, or even millions of active elements in a single die. 2. An individual electrical component, such as a resistor, capacitor, or transistor, as opposed to an integrated circuit, which is equivalent to several discrete components.

discrete element—An electronic element, such as a resistor or transistor, fabricated in such a way that it can be measured and transported individually.

discrete part—A separately packaged single circuit element supplying one fundamental property as a lumped characteristic in a given application. Examples: resistor, transistor, diode.

discrete sampling—The lengthening of individual samples so that the sampling process does not deteriorate the intelligence frequency response of the channel.

discrete thin-film component—An individually packaged electronic component having one or more thin films serving as resistive, conductive, and/or insulating elements. Resistors and potentiometers having thin-film metallic resistance elements are examples.

discretionary wiring—The use of a selective metallization pattern in the interconnection of large numbers of basic circuits on a slice of semiconductor material to form complex arrays. The metallization pattern connects only the "good" circuits on the wafers. Discretionary wiring requires a different interconnection pattern for each wafer.

discrimination—1. The difference between losses at specified frequencies, with the system or transducer terminated in specified impedances. 2. In a frequency-modulated system, the detection or demodulation of the imposed variations in the frequency of the carriers. 3. In a tuned circuit, the degree of rejection of unwanted signals.

discrimination ratio — The ratio of the width of the passband of a filter to the width of the stopband of the filter.

discriminator — 1. A device in which amplitude variations are derived in response to frequency or phase variations. 2. A facsimile auxiliary device between the radio receiver and the recorder that converts an audio-frequency-shifted facsimile signal to an amplitude-modulated facsimile signal.

discriminator transformer — A transformer used in FM receivers to convert frequency changes directly to audio-frequency signals.

discriminator tuning unit — A device that tunes the discriminator to a particular subcarrier.

disc-seal tube — Also called lighthouse tube or megatron. An electron tube with disc-shaped electrodes arranged in closely spaced parallel layers to give a low interelectrode capacitance along with a high power output in the UHF region.

dish — 1. A microwave antenna, usually shaped like a parabola, that reflects the radio energy leaving or entering the system. 2. A parabolic type of radio or radar antenna, roughly the shape of a soup bowl. 3. A colloquial expression for a parabolic antenna. 4. Common term for a parabolic microwave antenna.

dish illumination — The area of a dish as seen by the feedhorn.

disk — 1. An electromagnetic storage medium for digital data. 2. High-capacity random-access magnetic storage medium. *See also* disc.

disk cartridge — The flat, round, removable disk pack, containing programs and data, that is placed into a disk drive.

disk drive — 1. Identified floppy, removable, and nonremovable bulk storage for most minicomputer and mainframe systems, and microcomputer systems needing several megabytes of storage. 2. A disk player that rotates the disk, writes data onto it, and reads data from it as instructed by a program.

diskette — *See* floppy disk.

disk operating system — Abbreviated DOS. 1. The software that organizes how a computer reads, writes, and reacts with its disks and talks to its various peripherals (input/output devices), such as keyboards, screens, serial and parallel ports, printers, modems, etc. The most popular operating system for PCs is MS-DOS from Microsoft. 2. An operating system (set of programs) that instructs a disk-based computing system to manage resources and operate related equipment. 3. A set of programs that controls a computer. The DOS performs a variety of tasks, including managing communications between the computer and its peripherals. *See also* operating system.

disk pack — The vertical stacking of a series of magnetic disks in a removable self-contained unit.

disk storage — 1. Random-access auxiliary memory device in which information is stored on constantly rotating magnetic disks. 2. The storage of data on the surface of magnetic disks. 3. A mass storage memory device employing a flat, rotating medium onto which data can be stored via magnetic recording techniques and retrieved by magnetic playback. 4. A method of high-speed bulk storage of programs and data. The medium is a rotating circular plate coated with a magnetic material, such as iron oxide. Data is written (stored) and read (retrieved) by fixed or movable read/write heads positioned over data tracks on the surface of the disk. Addressable portions can be selected for read or write operations.

dislocation — In a crystal, a region in which the atoms are not arranged in the perfect crystal-lattice structure.

dispatcher — In a digital computer, the section that transfers the words to their proper destinations.

dispenser — A device that automatically distributes radar chaff from an aircraft.

disperse — In data processing, to distribute grouped input items among a larger number of groups in the output.

dispersion — 1. Separation of a wave into its component frequencies. 2. Scattering of a microwave beam as it strikes an obstruction. 3. The property of an optical material that causes some wavelengths of light to be transmitted through the material at different velocities, with the velocity a function of the wavelength. (This causes each wavelength of light to have a different refractive index.) 4. In a magnetostrictive delay line, the variation of delay as a function of frequency. 5. The frequency difference that can be analyzed in one sweep by a spectrum analyzer. Dispersion can be considered as that frequency width over which sampling can be performed, and is always equal to or less than the frequency range. 6. The extent to which a speaker distributes acoustical power widely and evenly into the listening area. 7. The undesirable effect of the broadening of optical pulses caused by lengthening of rise and fall times as the pulse travels along the fiber. Sometimes referred to as pulse spreading, it results from either modal or material effects in the fiber that reduce bandwidth. Expressed in nanoseconds per kilometer. 8. A fiber-optic phenomenon that causes pulse widths of transmitted data to lengthen. Dispersion is caused by the arrival of data at the far terminal at different times due to the varying lengths of optical paths in multimode fiber, and by inherent properties in the fiber. Dispersion increases with length of conductor and is caused by the difference in ray path lengths within the fiber core.

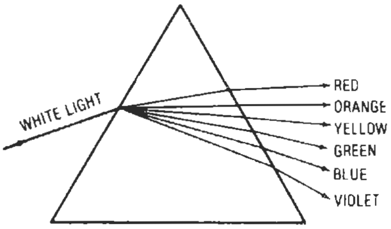

Dispersion, 3 (by a prism).

dispersive medium — A medium in which the phase velocity of a wave is related to the frequency.

displacement — 1. The vector quantity representing change of position of a particle. 2. A number that a computer must add to a base address to form an effective address.

displacement current — A current that exists in addition to ordinary conduction current in ac circuits. It is proportional to the rate of change of the electric field. The current at right angles to the direction of propagation determined by the rate at which the field energy changes.

displacement of porches — The difference in level between the front and back porch of a television signal.

displacement transducer — A device that converts mechanical energy into electrical energy, usually by the movement of a rod or an armature. The amount of output voltage is determined by the amount the rod or armature is moved.

display — Also called readout. 1. Visual presentation of a received signal on a cathode-ray tube or video screen.

2. Row of digits across the top of a calculator, showing input or final answer. In printing-type calculators, referred to as printout. 3. The observable illustration of an image, scene, or data on a screen, such as a console or CRT screen, seen as a graph, report, or drawing. 4. The representation of data in visible form, e.g., on a cathode-ray tube, by lights or indicators on the console of a computer, or a printed report.

display console — A visual display used with a computer to give access to the many elements of data as an array of points. With the display console, an operator may check information in the computer and change it if required.

display-generation time — The time span between the output of data from the host computer and the moment at which the complete display can be viewed.

display generator — An electronic device that interfaces computer-graphics display information with a graphics-display device. Typically, the interface is made between a digital computer and a CRT. In general, a display generator for a rasterscan display contains four subsystems: display controller, display processor, refresh memory, and video driver.

display highlighting — The ability of the word processor to intensify or blink certain portions of the display screen — either the characters themselves or the screen area behind the characters — to emphasize a text segment designated for some special activity such as deleting or moving.

display information processor — A computer used in a combat operations center to generate situation displays.

display loss — *See* visibility factor.

display modes — Each display mode, such as vector, increment, character, point, vector continue, or short vector, specifies the manner in which points are to be displayed on the screen.

display panel — The substrate containing the media for creating an image, including electric connections but excluding the electronic interface.

display primaries — Also called receiver primaries. The red, green, and blue colors produced by a color television receiver and mixed in proper proportions to produce other colors.

display processor — A component of a display generator used to add intelligence. Typically, the device is a microcomputer with stored programs that perform high-level graphics functions.

display-storage tube — A special cathode-ray tube with a long and controllable image presistence and high luminescence.

display unit — A device used to provide a visual representation of data.

display window — The width of the portion of the frequency spectrum presented on panoramic presentation, expressed in frequency units, usually megahertz.

disruptive discharge — The sudden, large current through an insulating medium when electrostatic stress ruptures the medium and thus destroys its insulating ability.

dissector — In optical character recognition, a mechanical or electronic transducer that sequentially detects the level of light in different areas of a completely illuminated sample space.

dissector tube — A camera tube having a continuous photocathode on which a photoelectric emission pattern is formed. Scanning is done by moving the electron optical image of the pattern over an aperture. *See also* image dissector, 1.

dissipation — The undesired loss of electrical energy by conversion into heat.

dissipation constant — A constant of proportionality between the power dissipated and the resultant temperature rise in a thermistor at a specified temperature.

dissipation factor — 1. Symbolized by D. Ratio between the permittivity and conductivity of a dielectric. The reciprocal of the dissipation factor (df) is the storage factor, sometimes called the quality factor (Q). 2. A measure of the ac loss. Dissipation factor is proportional to the power loss (P_L) per cycle (f) per potential gradient squared (E^2) per unit volume (V) as follows:

$$\text{dissipation factor} = (P_L/kE^2fV)$$

where k is a constant. Dissipation factor is approximately equal to power factor when the loss angle is small.

dissipation line — A length of stainless-steel or Nichrome wire used as a noninductive impedance for termination of a rhombic transmitting antenna when power of several kilowatts must be dissipated.

dissonance — The formation of maxima and minima by the superposition of two sets of interference fringes from light of two different wavelengths.

dissymmetrical network — *See* dissymmetrical transducer.

dissymmetrical transducer — Also called dissymmetrical network. A transducer with unequal input and output image impedances.

distance mark — Also called range mark. A mark that indicates, on a cathode-ray screen, the distance from the radar set to a target.

distance-measuring equipment — Abbreviated DME. A radio navigational aid for determining the distance from a transponder beacon by measuring the time of transmission to and from it.

distance protection — The effect of a device operative within a predetermined electrical distance on the protected circuit to cause and maintain an interruption of power in a faulty circuit.

distance relay — 1. A protective relay, the operation of which is a function of the distance between the relay and the point of fault. 2. A device that functions when the circuit admittance, impedance, or reactance increases or decreases beyond predetermined limits.

distance resolution — The ability of a radar to differentiate targets solely by distance measurement. Generally expressed as the minimum distance the targets can be separated and still be distinguishable.

distortion — 1. Undesired changes in the waveform of a signal so that a spurious element is added. All distortion is undesirable. Harmonic distortion disturbs the original relationship between a tone and other tones naturally related to it. Intermodulation distortion (IMD) introduces new tones caused by mixing of two or more original tones. Phase distortion, or nonlinear phase shift, disturbs the natural timing sequence between a tone and its related overtones. Transient distortion disturbs the precise attack and decay of a musical sound. Harmonic and IMD distortion are expressed in percentages; phase distortion in degrees; transient distortion is usually judged from oscilloscope patterns. 2. Unwanted changes in the purity of sound being reproduced or in rf signals. In audio, it generally implies intermodulation and/or harmonic distortion. These are derived from phase differences and/or amplitude distortion in which the amplitude of the output does not bear the same proportion to the input at all frequencies. 3. With a signal frequency (sine wave) signal, distortion appears as harmonics (multiples) of the input frequency. The rms (effective ac point) sum of all harmonic distortion components, plus hum and noise, is known as total harmonic distortion, or THD. When a two-tone test signal is used, distortion components appear at frequencies that are sums and differences of multiples of the input frequencies. Their magnitude is expressed as intermodulation distortion, which is more distressing to hear than THD. The lower the distortion in any form, the better. 4. Any difference in the waveshape after the signal has traversed the transmission circuit. 5. The unwanted changes in signal or signal shape that occur during transmission between two points.

distortion factor — *See* harmonic distortion.

distortion factor of a wave — The ratio of the effective value of the residue after the elimination of the fundamental to the effective value of the original wave.

distortionless line — A transmission line whose propagation constant is independent of frequency. (This is approached in a practical case by adjusting the line parameters, series inductance (l), shunt capacitance (c), series resistance (r), and shunt conductance (g) so that $r/g = l/c$.)

distortion meter — 1. An instrument that measures the deviation of a complex wave from a pure sine wave. 2. An instrument that measures the harmonic content of a sine wave, usually calibrated to read in percent distortion.

distortion tolerance — Of a telegraph receiver, the maximum signal distortion that can be tolerated without error in reception.

distress frequency — A frequency reserved for distress calls, by international agreement. It is 500 kHz for ships at sea and aircraft over the sea.

distributed — Spread out over an electrically significant length, area, or time.

distributed amplifier — A multistage amplifier in which the high-frequency limitation, due to the input and output capacitances of the active element, is circumvented by making these capacitances the shunt elements of lumped-parameter device lines. In this way the overall gain is the sum of the gains of the individual stages rather than the product, thus allowing amplification even when the individual gains are less than unity.

distributed capacitance — Also called self-capacitance. Any capacitance not concentrated within a capacitor, such as the capacitance between the turns in a coil or choke, or between adjacent conductors of a circuit.

distributed computer network — A collection of computers and I/O devices that can communicate with each other. *See* distributed processing.

distributed constants — Constants such as resistance, inductance, or capacitance that exist along the entire length or area of a circuit, instead of being concentrated within circuit components.

distributed data processing — Abbreviated DDP. The functional distribution of certain data-processing activities along logical organizational lines.

distributed-emission photodiode — A broadband photodiode for use in detecting modulated laser beams at millimeter wavelengths.

distributed inductance — The inductance along the entire length of a conductor, as distinguished from the inductance concentrated within a coil.

distributed network — 1. An electrical-electronic device that for proper operation depends on physical size in comparison to a wavelength and physical configuration. 2. A network configuration in which all node pairs are connected either directly or by redundant paths through intermediate nodes.

distributed parameter network — A network in which the parameters of resistance, capacitance, and inductance cannot be taken as being concentrated at any one point in space. Rather, the network must be described in terms of its magnetic and electric fields and the quantities related to the distributed constants of the network.

distributed paramp — A paramagnetic amplifier consisting essentially of a transmission line shunted by uniformly spaced, identical varactors. The varactors are excited in sequence by the applied pumping wave to give the desired traveling-wave effect.

distributed pole — A motor has distributed poles when its stator or field windings are distributed in a series of slots located within the arc of the pole.

distributed processing — 1. A multiprocessing computer technique in which each processor has a specific task or set of tasks to perform. These processors transfer commands and data via a standard communication interface. 2. Performing a data-processing task by performing the needed calculations in a distributed computer network. The efficiency of the data-processing task is improved through the simultaneous performance of operations in several interconnected processors of a distributed computer network. 3. Data-processing tasks performed simultaneously in several interconnected processors of a computer network.

distributing amplifier — An amplifier, either radio frequency or audio frequency, having one input and two or more isolated outputs.

distributing cable — *See* distribution cable.

distributing frame — A structure for terminating permanent wires of a central office, private branch exchange, or private exchange and for permitting the easy change of connection between them by means of cross-connecting wires.

distributing terminal assembly — A frame situated between each pair of selector bays to provide terminal facilities for the selector bank wiring and facilities for cross connection to trunks running to succeeding switches.

distribution — Also called frequency distribution. The number of occurrences of the particular values of a variable as a function of those values.

distribution amplifier — 1. A power amplifier designed to energize a speech, music, or antenna distribution system. Its output impedance is sufficiently low that changes in the load do not appreciably affect the output voltage. 2. A device that provides several isolated outputs from one looping or bridging input, and has a sufficiently high input impedance and input-to-output isolation to prevent loading of the input source.

distribution cable — Also called distributing cable. 1. A cable extended from a feeder cable for the purpose of providing service to a specific area. 2. In a system, the transmission cable from the distribution amplifier to the drop cable.

distribution center — In an alternating-current power system, the point at which control and rotating equipment is installed.

distribution coefficients — Equal-powered tristimulus values of monochromatic radiations.

distribution switchboard — A power switchboard used for the distribution of electrical energy at the voltage common for each distribution within a building.

distributive sort — A sorting procedure that divides data elements into two or more distinct groups or subsets. The partition sort is an example.

distributor — 1. *See* memory register. 2. The electronic circuitry that acts as an intermediate link between the accumulator and drum storage.

distributorless — A semiconductor automotive ignition system that does not utilize breaker contacts to time or trigger the system, nor does it utilize a distributor distribution of the secondary voltage.

disturbance — 1. An irregular phenomenon that interferes with the interchange of intelligence during transmission of a signal. 2. An interruption of a quiet state. 3. Any form of interference with normal communications.

disturbed-one output—A "one" output of a magnetic core to which partial-read pulses have been applied since that core was last selected for writing.

disturbed-zero output—A "zero" output of a magnetic core to which partial-write pulses have been applied since that core was last selected for reading.

disturbing conductor—A conductor carrying energy that creates spurious signals in another conductor.

dither—1. An oscillation introduced for the purpose of overcoming the effects of friction, hysteresis, or clogging. 2. A small electrical signal deliberately injected into an electromechanical device for the purpose of overcoming static friction in the device. In a recording instrument it makes the indicator ready to jump. 3. Constant vibration about a point. 4. The technique of adding controlled amounts of noise to a signal to improve overall system loop control, or to smear quantizing error in an analog-to-digital converter application.

dithering—1. The application of intermittent or periodic acceleration forces sufficient to minimize the effect of static friction with a transducer, without introducing other errors. 2. The creation of additional colors or shades of gray to create special effects or to make hard edges softer.

divergence loss—The part of transmission loss that is caused by the spreading of sound energy.

diverging lens—A lens that is thinner in the center than at the edges. Such a lens causes light passing through to spread out, or diverge.

diversity—1. A form of transmission and/or reception using several modes, usually in space or in time, to compensate for fading or outages in any one of the modes. In the space diversity system, the same signal is sent simultaneously over several different transmission paths, which are separated enough so that independent propagation conditions can be expected. With time diversity, the same path may be used, but the signal is transmitted more than once, at different times. There are other forms of diversity, using different frequencies or different polarizations to provide the separate transmission pole. *See* diversity reception. 2. The practice of constructing a portion of the system or backup system by using a different technology, component, or design, such that the two portions of the total system are not vulnerable to a common-cause failure.

diversity factor—The ratio of the sum of the individual maximum demands of the subdivisions of a system (or part of a system) to the maximum demand of the entire system (or part of the system).

diversity gain—The gain in reception as a result of the use of two or more receiving antennas.

diversity reception—1. A method of minimizing the effects of fading during reception of a radio signal. This is done by combining and/or selecting two or more sources of received-signal energy that carry the same intelligence but differ in strength or signal-to-noise ratio in order to produce a usable signal. *See* frequency-diversity reception; space-diversity reception. 2. A technique for reducing the adverse effects of multipath fading by receiving the same signal on two or more diverse, or different, antenna-receiver combinations with a means of choosing the combination with the strongest signal. A two-channel system is known as dual diversity; a three-channel system, triple diversity. Diversity reception is widely and effectively used in commercial high-frequency installations.

diverter pole generator—A compound-wound direct-current generator with the series winding of the diverter pole opposing the flux generated by the shunt-wound main pole; provides a close voltage regulation.

divide-by-*N* counter—A group of counter stages that can be programmed to divide an input frequency by any number up to *N*.

divide-by-16 counter—A logic device in which four flip-flops count from 0 through 15 and then recycle to 0. All 16 states of the combination of four flip-flops are used. Sometimes referred to as a hexadecimal counter.

divide-by-10 counter—*See* decade counter.

divide-carrier modulation—The process by which two signals are added so that they can modulate two carriers of the same frequency but 90° out of phase. The resultant signal will have the same frequency as the carriers, but its amplitude and phase will vary in step with the variations in amplitude of the two modulating signals.

divide check—In a computer, an indicator that shows that an invalid division has occurred or has been attempted.

divider—*See* counter, 4.

dividing network—Also called speaker dividing network and crossover network. A frequency-selective network that divides the audio-frequency spectrum into two or more parts to be fed to separate devices such as amplifiers or speakers.

D layer—The lowest ionospheric layer, located between about 35 to 55 miles (56 to 88 km) above the earth. Its intensity is proportional to the height of the sun and is greatest at noon. Waves below approximately 3 MHz are absorbed by the D layer when it is present.

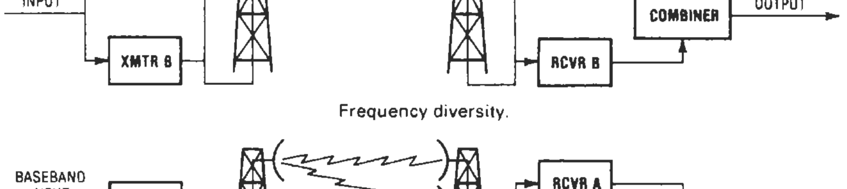

Frequency diversity.

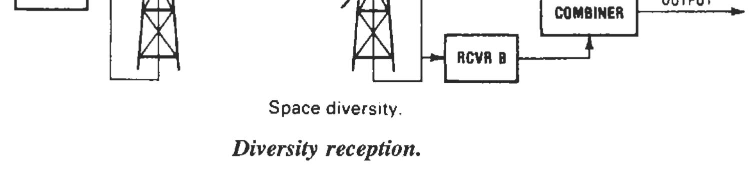

Space diversity.

Diversity reception.

High-angle radiation may penetrate the D layer and be reflected by the E layer.

DMA—Abbreviation for direct memory access. A method in which a device in a computer other than the main processor can access main memory directly. It must first obtain control of the bus from the processor or other DMA devices. Then it can read and write to memory directly. This is a very fast method of transferring large amounts of data from a peripheral to main memory. Using the main processor to handle the data transfer requires more software overhead and reduces system throughput.

DMI—Abbreviation for dual-mode ignition. An adaptation of the Duraspark (Ford Motor) electronic ignition system that allows its use on smaller four-cylinder engines like the 2.3-liter size.

DMM—Abbreviation for digital multimeter.

DMOS—Abbreviation for double-diffused metal-oxide FET semiconductor. A process in which n and p atoms are diffused through the same mask opening to give precisely sized narrow channels. Used on discrete field-effect transistors (not MOS ICs) for ultrahigh gains and frequency performance. A very fast MOSFET fabricated with an extra diffusion step.

DNC—Abbreviation for direct numerical control.

DNL—*See* dynamic noise limiter.

DOC files—Document files. The default file extension for Microsoft Word. Most files with this extension are Microsoft Word files, but there are many that are plain text.

documentation—An orderly collection of recorded hardware and software data such as tables, listings, diagrams, etc.

document reader—A general term referring to OCR or OMR equipment that reads a limited amount of information (one to five lines). Generally operates from a predetermined format and is therefore more restricted in the location of information to be read. The forms involved are generally tab card size or slightly smaller or larger.

doghouse—A small enclosure located near the base of a transmitting-antenna tower and used to house antenna tuning equipment.

Doherty amplifier—A radio-frequency linear power amplifier divided into two sections, the inputs and outputs of which are connected by quarter-wave (90°) networks. As long as the input-signal voltage is less than half the maximum amplitude, section No. 2 is inoperative and section No. 1 delivers all the power to the load. The load presents twice the optimum impedance required for maximum output. At one-half the maximum input, section No. 1 is operating at peak efficiency but is beginning to saturate. Above this level, section No. 2 comes into operation and decreases the impedance presented to section No. 1. As a result, section No. 2 delivers more and more power to the load until, at maximum signal input, both sections are operating at peak efficiency and each section is delivering one-half the total output power.

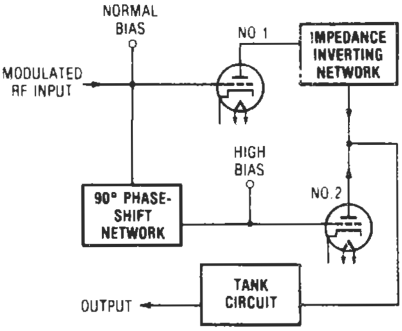

Doherty amplifier.

Dolby—1. A technique that increases the signal-to-noise ratio of a recording medium by raising the volume of quiet passages prior to recording, and lowering them to their original levels during playback. The lowering process automatically reduces any noise that was introduced as a result of the recording or playback processes. 2. Noise-reduction circuit that boosts the recorded signal at the tape hiss frequencies for low levels and reduces the boost progressively as the signal becomes large enough to mask the noise. (The Dolby system has the important advantage that it is standardized and any Dolby tape can be replayed accurately on any other Dolby machine.) 3. Name of a noise-reduction system available as a special circuit on some stereo cassette tape decks. 4. A proprietary electronic device or circuit that reduces the amount of noise (principally tape hiss) introduced during the recording process by boosting—in carefully controlled amounts—the strength of weak high-frequency signals before they are recorded. During playback the signals (and the noise) are cut back by an exactly equivalent amount. The original dynamics are restored, but the noise is reduced by 10 dB.

Dolby A—The original form of the Dolby noise-reduction device, intended for professional use. It has four independently controlled noise-reduction channels, to increase signal-to-noise ratio at low, middle, high and very high frequencies.

Dolby B—1. A simplified version of the original Dolby A, intended primarily for use by nonprofessional recordists. Dolby B functions identically to Dolby A, but has only one controlled frequency band, which is effective primarily on tape hiss. 2. A noise-reduction system widely used in cassette recorders, as well as in some open-reel and cartridge machines and in FM broadcasting. The high-frequency portions of signals being recorded are compressed, with the degree of compression being greater as signal level decreases. An opposite expansion process takes place in playback, restoring the original frequency response, but with a reduction in high-frequency hiss. 3. A complementary noise-reduction system designed to reduce tape (and FM) hiss. A Dolby B circuit boosts low-level high-frequency signals during recording and reduces them, along with the tape's added noise, in a complementary fashion during playback. Noise can be reduced up to 10 db above 5 kHz with the Dolby B system. It is now in virtually universal use in cassette decks.

Dolby B & C noise reduction (cassette deck)—Two systems of noise/hiss reduction invented by Ray Dolby. They work by boosting high frequencies during recording (also called encoding) and attenuating them during playback (also called decoding). Dolby B noise reduction boosts the level of the high-frequency range during recording and tapers the high frequencies during playback, reducing tape hiss by 8 to 10 dB. Dolby C noise reduction uses the same principle as Dolby B, with the addition of a second stage in which the frequencies affected are lower by about one octave. This results in a 15 to 18 dB reduction in tape hiss over an extended frequency range.

Dolby digital—The multi-channel audio encoding formed for DVD.

dolly—A wheeled platform or frame on which the tripod or frame supporting a television camera or other apparatus is mounted to give it wider mobility.

domain—1. In magnetic theory, that region of a magnetic material in which the spontaneous magnetization is all in one direction. In conventional magnetic-tape coatings, this corresponds to one oxide particle. 2. A region within a ferromagnetic substance where the atomic magnets of many atoms tend to orient themselves parallel to each other; the north poles pointing one way act spontaneously. The domains may be treated as small bar magnets of microscopic dimension. 3. In the Internet and other networks, an extension in a host name that identifies the type of host. The seven domains established by the InterNIC are .arpa (ARPANET), .com (company/commercial), .edu (educational institutions), .gov (government), .mil (military), .net (Internet access providers), and .org (organization). Outside the United States, the domain name is a two-letter country code (for example, .fr for France and .ca for Canada).

domain name — The unique name that identifies an Internet site. Domain names always have two or more parts, separated by dots. The part on the left is the most specific, and the part on the right is the most general. A given machine may have more than one domain name, but a given domain name points to only one machine. Usually, all of the machines on a given network will have the same thing as the right-hand portion of their domain names.

domains — *See* particles.

domestic induction heater — A home cooking utensil that is heated by induced currents within it. The unit contains a primary inductor, with the utensil itself acting as the secondary.

dominant mode — Also called fundamental mode or principal mode. In waveguide transmission, the mode with the lowest cutoff frequency. Designations for this mode are $TE_{0,1}$ and $TE_{1,1}$ for rectangular and circular waveguides, respectively.

dominant wave — The guided wave that has the lowest cutoff frequency. It is the only wave that will carry energy when the excitation frequency is between the lowest and the next higher cutoff.

dominant wavelength — 1. Of a color sample, the wavelength of light that matches it in chromaticity when mixed with white light. 2. The wavelength that is a quantitative measure of the apparent color of light as perceived by the human eye.

dominant wavelength of a color — The predominant wavelength of light in a color.

donor — Also called donor impurity. 1. An impurity atom that tends to give up an electron and thereby affects the electrical conductivity of a crystal. Used to produce n-type semiconductors. 2. A chemical that adds electrons to crystal lattices. 3. An impurity from column V of the periodic table, which adds a mobile electron to the conduction band of silicon, thereby making it more n-type. Commonly used donors are arsenic and phosphorous (compare with *acceptor*).

donor impurity — An element or compound whose atoms or molecules have more valence electrons than those of the intrinsic semiconductor material into which they are introduced in small quantities as an impurity or dopant. Because the donor impurity possesses more valence electrons, the material doped with a donor impurity is an n-type semiconductor. *See* donor.

donor-type semiconductor — An n-type semiconductor.

donut — *See* land, 2.

door cord — A short, insulated cable with an attaching block and terminals at each end used to conduct current to a device, such as foil, mounted on the movable portion of a door or window.

doorknob tube — A vacuum tube so called because of its shape designed for UHF transmitter circuits. It has a low electron-transit time and low interelectrode capacitance because of the close spacing and small size, respectively, of its electrodes.

door trip switch — A mechanical switch mounted so that movement of a door will operate the switch.

dopant — 1. An impurity added to a semiconductor to improve its electrical conductivity; any material added to a substance to produce desired properties in the substance. 2. Selected impurity introduced into semiconductor substrates in controlled amounts, the atoms of which form negative (n-type) and positive (p-type) conductive regions. Phosphorus, arsenic, and antimony are n-type dopants for silicon; boron, aluminum, gallium, and indium are p-type dopants for silicon.

dope — To add impurities (called dopants) to a substance, usually a solid, in a controlled manner to cause the substance to have certain desired properties. For example, the number of electrical carriers in silicon can be increased by doping it with small amounts of other semimetallic elements. Ruby is aluminum oxide doped with chromium oxide.

doped junction — A semiconductor junction produced by the addition of an impurity to the melt during crystal growth.

doped region — A layer of an integrated circuit in which impurities have been introduced.

doped solder — Solder to which an element not normally found in solder has been intentionally added.

doping — The addition of controlled amounts of impurities to a semiconductor to achieve a desired characteristic, such as to produce an n-type or p-type material, accomplished through thermal diffusion or ion implantation. Common doping agents for germanium and silicon include aluminum, antimony, arsenic, gallium, and indium.

doping agent — An impurity element added to semiconductor materials used in crystal diodes and transistors. Common doping agents for germanium and silicon include aluminum, antimony, arsenic, gallium, and indium.

doping compensation — The addition of donor impurities to a p-type semiconductor or of acceptor impurities to an n-type semiconductor.

Doppler cabinet — A speaker cabinet in which either the speaker or a baffle board is rotated or moved to change the length of the sound path cyclically and thereby produce a vibrato effect mechanically.

Doppler effect — 1. The observed change of frequency of a wave caused by a time rate of change of the effective distance traveled by the wave between the source and the point of observation. As the distance between a source of constant vibration and an observer diminishes or increases, the received frequencies are greater or less. 2. The apparent change in the frequency of radio wave reaching an observer, due either to motion of the source toward or away from the observer, to motion of the observer, or both. 3. The apparent change in frequency of sound or radio waves when reflected from or originating from a moving object. Utilized in some types of motion sensors. 4. The radiation emitted from a source that moves away from an observer appears to be of lower frequency than the radiation emitted from a stationary source. The radiation emitted from a source moving toward the observer appears to be of a higher frequency than that from a stationary source.

Doppler principle — The theory established by Doppler in 1842 that states that the rate of change in distance between a perceiver and a radiation source determines the change in frequencies.

Doppler radar — A radar unit that measures the velocity of a moving object by the shift in carrier frequency of the returned signal. The shift is proportionate to the velocity of the object as it approaches or recedes.

Doppler ranging—Abbreviated as doran. A cw trajectory-measuring system that utilizes the Doppler shift to measure the distance between a transmitter, missile, transponder, and several receiving stations. From these measurements, trajectory data is computed. In contrast to a similar system, doran circumvents the necessity of continuously recording the Doppler signal by performing the distance measurements with four different frequencies simultaneously.

Doppler shift—1. The change in frequency of a wave reaching an observer or a system, caused by a change in distance or range between the source and the observer or the system during the interval of reception. It is due to the Doppler effect. 2. The change in frequency with which energy reaches a receiver when the source of radiation or a reflector of the radiation and the receiver are in motion relative to each other. The Doppler shift is used in many tracking systems. 3. The magnitude, expressed in cycles per second, of the alteration in the wave frequency observed as a result of the Doppler effect.

Doppler signal—The signal, traveling from transmitter to receiver, that has an altered frequency due to the Doppler effect.

Doppler velocity and position—1. Having to do with a beacon tracking system in which pulses are sent from a tracking station to a receiver in the object to be tracked, and returned to the station on a different frequency. 2. Having to do with a Doppler trajectory-measuring system for determining target position relative to transmitting and receiving stations on the ground.

doran—*See* Doppler ranging.

DOS—Abbreviation for disk operating system.

dosage meter—*See* dosimeter; intensitometer.

dose—A measure of energy actually absorbed in tissue as a result of ionizing radiation.

dosimeter—Also called intensitometer or dosage meter. 1. An instrument that measures the amount of exposure to nuclear or X-ray radiation utilizing the ability of such radiation to produce ionization of a gas. 2. Quartz fiber electrometer that is charged by a battery and discharges when exposed to radiation. Can be direct reading or indirect reading. Measures the total radiation dose received, in rem, and is carried by a person who works with radiation.

DOS prompt—The letter informing DOS system users what drive they're in, followed by the greater-than symbol (C>), which indicates that the system is ready to receive a command.

dot—1. *See* button, 2. 2. Also called bubble. A symbol placed at the input of a logic symbol to indicate that the active signal input is negative. The absence of a dot indicates a positive active signal.

dot AND—*See* wired AND.

dot-bar generator—An instrument that generates a specified output pattern of dots and bars. Used for measuring scan linearity and geometric distortion of TV cameras, video monitors, and TV receivers. Also used for converging cathode-ray tubes as recommended by color monitor and receiver manufacturers.

dot cycle—One cycle of a periodic alternation between two signaling conditions, each condition having unit duration. Thus, two-condition signaling consists of a dot, or marking element, followed by a spacing element. In teletypewriter applications, one dot cycle consists of a mark and a space. The speed of telegraph transmission sometimes is stated in terms of dot cycles per second, or dot speed (half the speed of transmission expressed in bauds).

dot encapsulation—A packaging process in which cylindrical components are inserted into a perforated wafer to form a solid block with interconnecting conductors on both surfaces joining the components.

dot generator—An instrument used in servicing color television receivers. It produces a pattern of white dots so that convergence adjustments can be made on the picture tube.

dot matrix—1. A pattern of dots in a fixed area used for formulation of characters. A method of display character generation in which each character is formed by a grid or matrix pattern of 5 × 7 to 7 × 9 dots, combinations of which form characters on a video or hard-copy medium. For very high quality, 11 × 13 patterns or greater are required. 2. The printing of characters by a matrix pattern of ink dots. 3. A technique for representing characters by composing them out of selected dots from within a rectangular matrix of dots.

dot-matrix character—1. A printed character formed of dots so close together that it gives the impression of having been printed by uninterrupted strokes. The dots are formed by wire ends, jets of ink, electrical charge, or laser beams. 2. A character composed from a rectangular matrix of dots.

dot-matrix display—1. A display format consisting of small light-emitting elements arranged as a matrix. Various elements are energized to depict a character. A typical matrix is 5 × 7. 2. A display composed of dots in close, rectangular array, capable of being individually illuminated to produce alphanumeric characters and graphic displays.

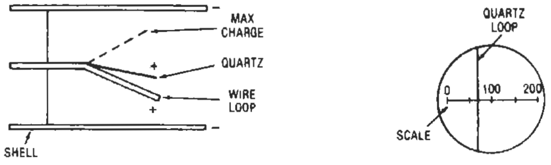

Expanded view of dosage indicator.

Dosage scale

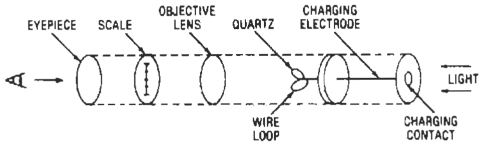

Dosimeter viewer

Dosimeter.

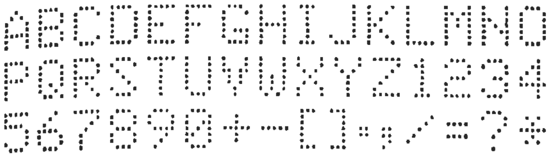

Dot-matrix characters and numerals (5 × 7 matrix).

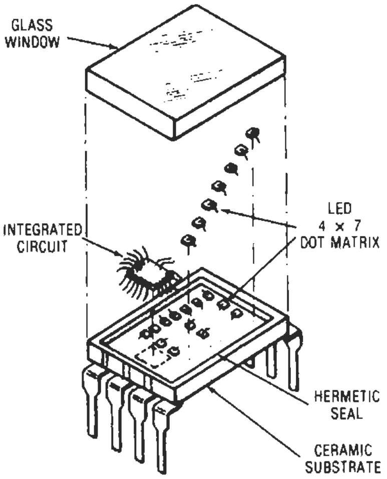

Dot-matrix LED display with on-board IC.

dot-matrix printer—A printer that forms letters by striking the paper with small pins or other means of making a mark (such as a small jet of ink), forming each letter with a pattern of dots. More marks per letter result in a denser image. Some dot-matrix printers squeeze the dots so closely together that they look almost like letter-quality print, which is made by machines that form letters with a single impact, as do traditional typewriters. When dot-matrix printers produce letters that are close to the quality of traditional electric typewriters, the result is often called correspondence-quality or letter-quality print. 2. *See also* wireprinter.

dot pattern—Small dots of light produced on the screen of a color picture tube by the signal from a dot generator. If overall beam convergence has been obtained, the three color-dot patterns will merge into one white-dot pattern.

dot pitch—1. A measure of picture quality or resolution in RGB color monitors. 2. The shortest distance between two phosphor dots of the same color on the screen. The smaller the dot pitch, the better the resolution of the monitor. High-resolution monitors usually have a dot pitch of 0.28 mm. 3. For printers, the number of dots per linear inch; for example, a desktop laser printer prints at least 300 dpi. The larger the number, the higher the resolution. 4. A measure of picture quality or resolution in RGB color monitors, which is more commonly known as pixel resolution. Dot pitch is the distance between screen dots (pixels) measured in millimeters. The shorter the distance, the better the resolution. It is specified in pixels/mm.

dot sequential—Pertaining to the association of the primary colors in sequence with successive picture elements of a color television system. Examples: dot-sequential pickup, dot-sequential display, dot-sequential system, dot-sequential transmission.

double-amplitude-modulation multiplier—A multiplier in which a carrier is amplitude modulated by one variable, and the modulated signal is again amplitude modulated by a second variable. The product of the two variables is obtained by applying the resulting double-modulated signal to a balanced demodulator.

double armature—An armature having two windings and commutators but only one core.

double-base diode—*See* unijunction transistor.

double-base junction transistor—Also called tetrode junction transistor. Essentially a junction triode transistor with two base connections on opposite sides of the central region of the transistor.

double-beam cathode-ray tube—A cathode-ray tube having two electron beams capable of producing on the screen two independent traces that may overlap. The beams may be produced by splitting the beam of one gun or by using two guns.

double-bounce calibration—A method of calibration used to determine the zero set error by using round-trip echoes. The correct range is the difference between the first and second echoes.

double-break contacts—A set of contacts in which one contact is normally closed and makes simultaneous connection with two other contacts.

double-break switch—A switch that opens the connected circuit at two points.

double bridge—*See* Kelvin bridge.

double-button carbon microphone—Also called differential microphone. A microphone with two carbon resistance elements or buttons, one on each side of a central diaphragm. They are connected in parallel to the current source in order to give twice the resistance change obtainable with a single button.

double-channel duplex—A method for simultaneous communication between two stations over two rf channels, one in each direction.

double-channel simplex—A method for nonsimultaneous communication between two stations over two rf channels, one in each direction.

double-checkerboard pattern—*See* worst-case noise pattern.

double-clocking—Incorrect setting of a flip-flop due to bounce in input signal.

double-conversion receiver—A receiver using a superheterodyne circuit in which the incoming signal frequency is converted twice, first to a high IF and then to a lower one.

double-current generator—A machine that supplies both direct and alternating current from the same armature winding.

double density—A type of diskette that allows twice as much data to be stored as single density.

double-diffused epitaxial mesa transistor—*See* epitaxial-growth mesa transistor.

double-diffused metal-oxide FET semiconductor—*See* DMOS.

double-diffused transistor—Also called double-emitter-and-base transistor. A transistor in which two pn junctions are formed in the semiconductor wafer by

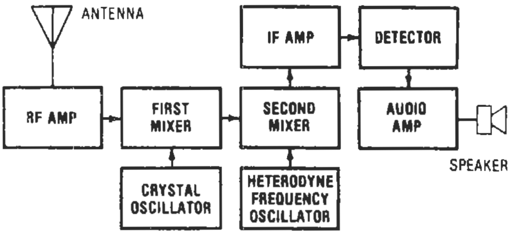

Double-conversion superheterodyne receiver block diagram.

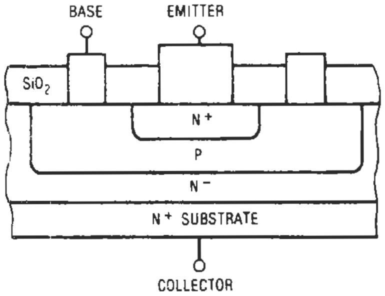

Double-diffused epitaxial planar bipolar npn transistor.

gaseous diffusion of both p-type and n-type impurities. An intrinsic region can also be formed.

double diode—Also called duodiode. A vacuum tube or semiconductor having two diodes in the same envelope.

double-diode limiter—A type of limiter used to remove all the positive signals from a combination of positive and negative pulses or to remove all the negative signals from such a combination of positive and negative pulses.

double-doped transistor—A transistor formed by growing a crystal and successively adding p- and n-type impurities to the melt while the crystal is being grown.

double-doublet antenna—An antenna composed of two half-wave doublet antennas criss-crossed at their centers; one is made shorter than the other to give broader frequency coverage.

double drop—An alarm-signaling method often used in central-station alarm systems, in which the line is first opened to produce a break alarm and then shorted to produce a cross alarm.

double-emitter follower—*See* Darlington amplifier.

double frequency-shift keying—A multiplex system in which two telegraph signals are combined and transmitted simultaneously by frequency shifting among four radio frequencies.

double-grip terminal—A solderless terminal with an extended flared barrel that permits a crimp to be made over the insulation of a wire as well as over the stripped portion.

double image—A television picture consisting of two overlapping images due to reception of the signal over two paths that differ in length so signals arrive at slightly different times.

double insulation—The insulation system resulting from a combination of functional and supplementary insulation.

double-junction photosensitive semiconductor—A semiconductor in which the current flow is controlled by light energy. It consists of three layers of a semiconductor material, with electrodes connected to the ends of each.

double-length number—Also called double-precision number. An electronic computer number having twice the normal number of digits.

double local oscillator—An oscillator mixing system that generates two rf signals accurately spaced a few hundred hertz apart and mixes these signals to give the difference frequency that is used as the reference. This equipment is used in an interferometer system to obtain a detectable signal containing the phase information of an antenna pair, and the reference signal to allow removal of the phase data for use.

double-make contacts—A set of contacts in which one contact is normally open and makes simultaneous connection with two other independent contacts when closed.

double moding—Changing from one frequency to another abruptly and at irregular intervals.

double modulation—1. The process of modulation in which a carrier wave of one frequency is first modulated by a signal wave, and the resultant wave is then made to modulate a second carrier wave of another frequency. 2. A two-step modulation scheme in which an intelligence wave modulates a subcarrier, and then the modulated subcarrier is used to modulate a higher-frequency carrier.

double operand—An instruction type containing two address fields: source operand address field and destination operand address field.

double orthomode coupler—A dish-mounted device that allows reception of both vertically and horizontally polarized signals.

double photoresist—A technique for eliminating pinholes in the photoresist coating during fabrication of microelectronic integrated circuits. The method may consist of two separate applications and exposures of photoresist emulsions of the same or different types.

double-play tape—Tape having half the thickness, and hence double the running time (for a given reel size) of standard 1.5-mil tape.

double pole—A term applied to a contact arrangement to denote that it includes two separate contact forms (i.e., two single-pole contact assemblies).

double-pole, double-throw switch—Abbreviated dpdt. A switch that has six terminals and is used to connect one pair of terminals to either of the other two pairs.

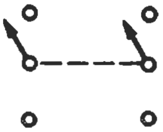

Double-pole, double-throw switch.

double-pole, single-throw switch—Abbreviated dpst. A switch that has four terminals and is used to connect or disconnect two pairs of terminals simultaneously.

double-pole switch—A switch that operates simultaneously in two separate electric circuits or in both lines of a single circuit.

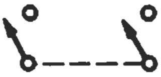

Double-pole, single-throw switch.

double precision—1. Having to do with the use of two computer words to represent one number. 2. The technique of allocating twice the data storage space for representing numeric information in order to achieve twice the accuracy.

double-precision arithmetic—The use of two computer words to represent a single number. This is done where it is necessary to obtain a greater accuracy than a single word of computer storage will provide. It effectively doubles the data word size.

double-precision number—*See* double-length number.

double-pulsing station—A loran station that receives two pairs of pulses and emits pulses at two pulse rates.

double pumping—A technique of pumping a laser for a relatively long time to store energy in subthreshold-level excited states, followed immediately by a very brief second pumping in which the threshold condition is exceeded in some region. This triggers laser oscillation throughout the entire active region and produces peak output powers several times larger than normally seen.

double rail—*See* dual rail.

double-rail logic—Pertaining to self-timing asynchronous circuits in which each logic variable is represented by two electrical lines that in combination can assume three meaningful states: zero, one, and undecided.

double screen—A three-layer screen consisting of a two-layer screen with the additional second long-persistence coating having a different color and different persistence from the first.

double-shield enclosure—A shielded enclosure or room whose inner wall is partially isolated electrically from the outer wall.

double sideband—Amplitude-modulated intelligence that is transmitted at frequencies both above and below the carrier frequency by the audio-frequency value of the intelligence.

double-sideband transmitter—A transmitter that transmits not only the carrier frequency, but also the two sidebands resulting from modulation of the carrier.

double-sided board—A printed board with a conductive pattern on both sides.

double-spot tuning—Superheterodyne reception of a given station at two different local-oscillator frequencies. The local oscillator is adjusted either above or below the incoming signal frequency by the intermediate-frequency value.

double-stream amplifier—A microwave traveling-wave amplifier in which amplification occurs through interaction of two electron beams having different average velocities.

double-stub tuner—An impedance-matching device consisting of two stubs, usually fixed three-eighths of a wavelength apart, in parallel with the main transmission lines.

double superheterodyne reception—Also called triple detection. The method of reception in which two frequency converters are employed before final detection.

double-surface transistor—A point-contact transistor, the emitter and collector whiskers of which are in contact with opposite sides of the base.

doublet—The output voltage waveform of a delay line under linear operating conditions when the input to the line is a current step function.

doublet antenna—An antenna consisting of two elevated conductors substantially in the same straight line and of substantially equal length, with the power delivered at the center.

double tape mark—A delimiter, consisting of two consecutive tape marks, that is used to indicate the end of a volume or of a file set.

double throw—A term applied to a contact arrangement to denote that each contact form included is a break-make.

double-throw circuit breaker—A circuit breaker by means of which a change in the circuit connections can be obtained by closing either of two sets of contacts.

double-throw switch—A switch that alternately completes a circuit at either of its two extreme positions. It is both normally open and normally closed.

double-track recorder—*See* dual-track recorder.

double trigger—A trigger signal consisting of two pulses spaced by a fixed amount for coding.

double triode—*See* duotriode.

double-tuned amplifier—An amplifier in which each stage utilizes coupled circuits having two frequencies of resonance for the purpose of obtaining wider bands than are possible with single tuning.

double-tuned circuit—A circuit in which two circuit elements are available for tuning.

double-tuned detector—A type of FM discriminator in which the limiter output transformer has two secondary windings, one tuned a certain amount above the center frequency and the other tuned an equal amount below the center frequency.

double-V antenna—Also called fan antenna. A modified single dipole that has a higher input impedance and broader bandwidth than an ordinary dipole.

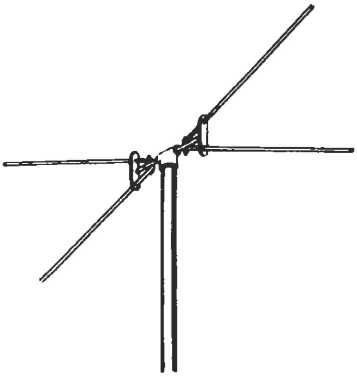

Double-V antenna.

double-winding synchronous generator—A synchronous generator that has two similar windings in phase with one another, mounted on the same magnetic structure but not connected electrically, designed to supply power to two independent external circuits.

doubling—The generation of large amounts of second-harmonic distortion by nonlinear motion of a loudspeaker cone.

doubly balanced modulator—A modulator circuit in which two class A amplifiers are supplied with modulating and carrier signals of equal amplitudes and opposite polarities. Carrier suppression takes place because the two amplifiers share a common plate circuit and only the sidebands appear at the output.

downconversion—The process of converting microwave signals down into a frequency range in which signal processing components are less expensive. Typically, this is a VHF frequency of 70 MHz.

downconverter—1. A type of converter whose input is a radio frequency and whose output is an intermediate frequency. 2. A microwave system (consisting of local oscillators, mixers, and bandpass filters) that accomplishes downconversion. 3. The front end of the satellite TV receiver. 3. The part of a satellite receiving system that converts the downlink signals to a 70-MHz intermediate frequency that is used by the receiver. Although it is sometimes part of the receiver, it is more often externally mounted directly at the LNA so that inexpensive coaxial cable can bring the signal to the receiver.

down lead—Also called a lead-in wire. The wire that connects an antenna to a transmitter, receiver, or downconverter.

downlink—1. A satellite-to-earth microwave channel and related components, such as the earth station receiving equipment. The satellite contains a downlink transmitter; downlink components in the earth station are involved with the reception and processing of satellite-transmitted signals. 2. The communication path from a TV satellite to its ground (earth) stations.

download—1. The process of sending communications instructions, operating software or data from a central computer to individual terminals, including personal computers. To electronically copy a file from one computer to another computer. 2. To receive a file from a remote computer. 3. To transfer information stored in a remote computer system to the user's system. The reverse process is an upload.

downloading—1. The electronic transfer of information from one computer to another, generally from a larger computer to a smaller one, such as a microcomputer. 2. Transferring information from a host computer to another computer. 3. To transfer data stored in one computer to a storage device in another computer. 4. The process of sending configuration parameters, operating software, or related data from a central source to remote stations. 4. Direct transfer of code from a host system (MDS) into a target system or a PROM programmer.

downstream—Outlet side of an instrument.

downtime—1. Any period of time during which a system or device cannot be used as a result of a failure or routine maintenance, but not because of a lack of work or the absence of an operator. 2. The period during which computer or network resources are unavailable to users because of a system or component failure.

downward compatibility—Software that can run on older and/or less powerful versions of a computer it was designed to run on.

downward modulation—Modulation in which the instantaneous amplitude of the modulated wave is never greater than that of the unmodulated carrier.

dpdt—Abbreviation for double-pole, double-throw.

dpi (dots per inch)—A measurement of resolution, usually used with printers and scanners.

DPM—*See* digital panel meter.

dpst—Abbreviation for double-pole, single-throw.

draft mode—A low-quality printing mode available on some printers.

drag—Selecting and moving an on-screen icon via a mouse.

drag angle—A stylus cutting angle of less than 90° to the surface of the record. So called because the stylus drags over the surface instead of digging in. It is the opposite of dig-in angle.

drag cup—A nonmagnetic metal rotated in a magnetic field to generate a torque or voltage proportional to its speed.

drag-cup motor—A small, high-speed, two-phase, alternating-current electric motor having a two-pole, two-phase stator. The rotating element consists only of an extremely light metal cup attached to a shaft rotating on ball bearings. Reversal is accomplished by reversing the connections to one phase. Used in applications requiring quick starting, stopping, and reversal characteristics.

dragging—An interactive technique for repositioning an image on a display screen.

drag magnet—*See* retarding magnet.

drag soldering—A form of mass soldering in which a printed circuit board is mounted on a conveyor and contacts the surface of a static pool of molten solder at a slight angle when entering and exiting the solder bath. The board remains in contact with solder for a defined length (the dragging length) of time, travelling horizontally across the solder surface.

drain—1. The current taken from a voltage source. 2. The working-current terminal (at one end of the channel in a FET) that is the drain for holes or free electrons from the channel. Corresponds to a collector of a bipolar transistor. 3. Terminal that receives carriers from the MOS channel. 4. One of the three regions that form a field-effect transistor. Majority carries that originate at the source and traverse the channel are collected at the drain to complete the current path. The flow between source and drain is controlled by the voltage applied to the gate.

drainage equipment—Equipment used to protect connected circuits from transients produced by the operation of protection equipment.

drain conductor—A conductor in continuous contact with a shield for ground termination.

drain cutoff current—The current into the drain terminal of a depletion-type transistor with a specified reverse gate-to-source voltage applied to bias the device to the off state.

drain terminal—The terminal electrically connected to the region into which majority carriers flow from the channel.

drain wire—1. An uninsulated solid or stranded tinned copper wire that is placed directly under a shield. It touches the shield throughout the cable, and therefore may be used in terminating the shield to ground. It is completely necessary on spiral shielded cables because it eliminates the possibility of induction in a spiral shield. 2. An uninsulated wire, usually placed directly beneath and in electrical contact with a shield. It is used for making shield connections through terminal strips and to ground.

DRAM—Abbreviation for dynamic random-access memory (pronounced "dee-ram"). 1. The least expensive and most popular type of semiconductor read/write memory chip, in which the presence or absence of a capacitive charge represents the state of a binary storage element (0 or 1). The charge must be periodically refreshed. 2. Memory that requires periodic refreshing because of charge leaking from capacitors in the cell circuit. 3. Main memory system of large computers, minicomputers, and even some large microcomputers.

drawing—In the manufacture of wire, pulling the metal through a die or series of dies for reduction of diameter to a specified size.

D-region — The region of the ionosphere up to about 90 kilometers above the earth's surface. It is below the E-region.

dress — 1. The exact placement of leads and components in a circuit to minimize or eliminate undesirable feedback and other troubles. 2. To arrange wire connections, cable ends, or cables so that they present a neat and orderly appearance.

dressed contact — A contact that has a locking spring member permanently attached.

drift — 1. Movement of carriers in a semiconductor as voltage is applied. 2. A change in either absolute level or slope of an input-output characteristic. 3. *See* flutter, 1. 4. *See* degradation, 1. 5. An undesired change in one of the output parameters of a power supply (voltage, current, frequency, etc.) over a period of time. The change is unrelated to all other variables, such as load, line, and environment. Drift is measured over a period of time by keeping all variables (such as line, load, and environment) constant. Specifications usually apply only after a warm-up period. 6. The angular displacement of an aircraft by the wind, generally expressed in degrees. 7. In a dc amplifier, the change in output with constant input, usually measured in terms of the dc input signal required to restore normal output; may be called out as microvolts or millivolts per hour. 8. A change in output attributable to any cause. 9. A change in the properties of an electrical circuit, as a result of aging or temperature changes.

drift current — 1. The flow of carriers in a semiconductor due to an electric field. In the same electric field, holes and electrons will flow in opposite directions due to their opposite charge. 2. The relatively small directional bias that becomes superimposed on the random motion of carriers in an excited crystal lattice under the influence of an applied electric field (drift field).

drifting — An instability in a preset voltage, frequency, or other electronic circuit parameter.

drift mobility — The average drift velocity of carriers in a semiconductor per unit electric field. In general, the mobilities of electrons and holes are not the same.

drift space — 1. In an electron tube, a region substantially free of alternating fields from external sources, in which relative repositioning of the electrons depends on their velocity distributions and the space-charge forces. 2. The distance between the buncher and catcher in a velocity-modulated vacuum tube.

drift speed — Average speed at which electrons or ions progress through a medium.

drift transistor — A type of transistor manufactured with a variable-conductivity base region. Such a base sets up an electric field that speeds up the carriers, thus reducing the transit time and improving high-frequency operation. *See also* diffused-alloy transistor.

drift velocity — Net velocity of charged particles in the direction of the applied field.

drip loop — A loop formed in a transmission line at a point where it enters a building. Condensation of moisture and water that may form on the line will drip off at the loop and thus will not enter the building.

drip-proof motor — A motor in which the ventilating openings are such that foreign matter falling on the motor at any angle not exceeding 15° from the vertical cannot enter the motor either directly or indirectly.

driptight enclosure — An enclosure that is intended to prevent accidental contact with the enclosed apparatus and, in addition, is so constructed as to exclude falling moisture or dirt.

drive — 1. A unit used in mass storage applications to hold and operate the medium (e.g., disk or tape) being used to store data or programs. 2. Also called excitation. The signal applied to the input of a power amplifier. *See* excitation, 2.

drive belt — A belt used to transmit power from a motor to a driven device.

drive circuit — A circuit, usually a printed circuit card or an encapsulated module, that converts an input pulse to the appropriate winding excitation sequence to produce one step of the motor shaft.

drive control — *See* horizontal drive control.

driven element — 1. An antenna element connected directly to the transmission line. 2. An element of an antenna, such as of a Yagi antenna, that is energized directly from the antenna feed line.

driven sweep — A sweep signal triggered by an incoming signal only.

drive pattern — In a facsimile system, an undesired pattern of density variations that result from periodic errors in the position of the recording spot.

drive pin — In disc recording, a pin similar to the center pin but located at one side of it and used to prevent a disc record from slipping on the turntable.

drive pulse — A pulsed magnetomotive force applied to a magnetic cell from one or more sources.

driver — 1. An electronic circuit that supplies input to another electric circuit. 2. A stage of amplification that precedes the power output stage. 3. In a radar transmitter, a circuit that produces a pulse to be delivered to the control grid of the modulator tube. 4. An element coupled to the output stage of a circuit to increase the power- or current-handling capability or fanout of the stage; for example, a clock driver is used to supply the current necessary for a clock line. 5. A device in a logic family controlled with normal logic levels whose output has the capability of sinking or sourcing high current. The output may control a lamp, relay, or a very large fanout of other logic devices. Also a device driving a higher-output device or transistor by supplying power, voltage, or current to it. 6. Any individual speaker within an audio system, such as the woofer, tweeter, etc. 7. A transistor output circuit that has an emitter-follower configuration. 8. A dc driver output module that contains driver output circuits. Each load must be connected between the output and ground as specified in the module data sheet. 9. Amplifier circuit used to reshape signals on a bus when more than one TTL load is present. 10. The low-power oscillator-modulator-amplifier unit that supplies the excitation to a power amplifier. *See also* exciter. 11. A program or routine that controls either external devices or other programs. 12. Typically, an electronic function used to provide amplification to drive high current loads. Term often used to denote bus drivers that rapidly charge and discharge capacitance. Also used to denote the ability to control power, such as when driving a solenoid or other high-current device.

driver element — An antenna array element that receives power directly from the transmitter.

driver gate — An analog switch, usually including two parts: the switch gate and a driver that controls the switch.

driver stage — The amplifier stage preceding the power-output stage.

driving-point admittance — The complex ratio of the alternating voltage for an electron tube, network, or other transducer.

driving-point impedance — 1. At any pair of terminals in a network, the driving-point impedance is the ratio of an applied potential difference to the resultant current at these terminals, all terminals being terminated in any specified manner. 2. The input impedance of a transmission line or of an antenna.

driving power—The power supplied to the grid circuit of a tube where the grid swings positive and draws current for part of each cycle of the input signal.

driving-range potential—The voltage difference between the potential of the electrochemically more active anode and the less active protected metal or cathode. One example of driving potential is the electromotive force in a cathodic protection system that causes current between the protected structure (cathode) and the anode. The driving potential decreases as the electrodes become polarized.

driving signal—Television signals that time the scanning at the pickup point. Two kinds of driving signals are usually available from a central sync generator, one composed of pulses at the line frequency and the other of pulses at the field frequency.

driving spring—The spring driving the wipers of a stepping relay.

DRO memory—Destructive readout memory. A memory in which the contents of a storage location are destroyed in being read. Information must be rewritten after reading, if it is to be returned. An example of a DRO memory is the common computer core memory.

drone cone—An undriven speaker cone mounted in a bass-reflex enclosure.

droop—The decrease in mean pulse amplitude, expressed as a percentage of the 100-percent amplitude, at a specified time following the initial attainment of 100-percent amplitude.

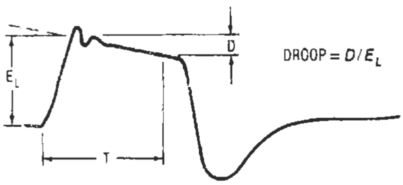

Droop.

drop—1. To develop a specified difference of potential between a pair of terminals as the result of a flow of current. 2. *See* voltage drop.

drop bar—A protective device used to ground a high-voltage capacitor when opening a door.

drop bracket transposition—Reversal of the relative positions of two parallel wire conductors while depressing one, so that the crossover is in a vertical plane.

drop cable—In a cable TV system, the transmission cable from the distribution cable to a dwelling.

drop channel—A type of operation in which one or more channels of a multichannel system are terminated (dropped) at some intermediate point between the end terminals of the system.

drop-down menu—A type of computer menu that drops from the menu bar when requested and remains open without further action until the user closes it or chooses an item from the menu. Same as pull-down menu.

drop-in—The reading of a spurious signal of amplitude greater than a predetermined percentage of the nominal signal amplitude.

drop indicator—An indicator for signaling, consisting of a hinged flap normally held up by a catch. The catch is released by an electromagnet, allowing the flap to drop when a signal is received.

drop-in mike—A surreptitious listening device that transmits phone conversations only, by rf, to a radio receiver. So called because it is "dropped in," replacing the standard carbon microphone in the ordinary telephone handset.

dropout—1. A momentary loss of volume or treble response due to a brief separation of the tape from the surface of the record or play head. A very slight separation causes a treble dropout; more severe loss of head-to-tape contact causes the whole signal to drop out. Dropouts can be caused by buckled or crinkled tape, lumps or pits in the magnetic coating, or detached clumps of oxide passing across the head surface. 2. Short pause in tape replay due to bad tape coating. 3. Momentary loss of signal in a transmission channel. 4. Momentary signal losses due to imperfections in the surface of recording tape or phonograph record.

dropout compensator—Circuitry that senses a signal loss produced by dropout and substitutes missing information with signal from the preceding line; if one line drops out of a picture, it is filled in with the preceding line, resulting in no dropout on the screen.

dropout error—An error, such as loss of a recorded bit, that occurs in recorded magnetic tape because of foreign particles on or in the magnetic coating or because of defects in the backing.

dropouts—Also called keys. Special images inserted at certain points in the array on a photomask used in the production of monolithic circuits.

dropout value—The maximum value of current, voltage, or power that will deenergize a previously energized relay. *See also* hold current; pickup value (voltage, current, or power).

dropout voltage—The input-output voltage differential at which a regulator circuit ceases to regulate against further reductions in input voltage.

dropping resistor—1. A resistor used to decrease a given voltage by an amount equal to the potential drop across the resistor. 2. A resistor placed in series between a voltage source and a load to reduce the voltage supplied to the load.

drop rate—The rate of discharge or decay rate of the sample-and-hold capacitor of a sample-and-hold device. The rate is a function of switch leakage current and the current required by other circuit elements connected to the capacitor. It is expressed as millivolts per second.

drop relay—A relay activated by incoming ringing current to call the attention of an operator to the subscriber's line.

drop repeater—A microwave repeater station equipped for local termination of one or more circuits.

dropsonde—A parachute-carried radiosonde dropped from a high-flying aircraft to measure weather conditions and report them to the aircraft. It is used over water or other areas where ground stations cannot be maintained.

drop wire—A wire suitable for extending an open wire or cable pair from a pole or cable terminal to a building.

dross—Oxide and other contaminants that form on the surface of molten solder.

drum—1. A random-access auxiliary memory device in which information is stored on a revolving drum that is coated with a magnetic material. 2. Rotating magnetic memory that uses the surface of a cylinder.

drum controller—1. A device in which electrical contacts are made on the surface of a rotating cylinder or sector. 2. A device in which contacts are made by the operation of a rotating cam.

drum memory—A rotating cylinder or disk coated with magnetic material so that information can be stored in the form of magnetic spots.

drum parity—A parity error that occurs during the transfer of information to or from drums.

drum plotter — Plotter that draws an image on paper or film mounted on a drum.

drum printer — A type of printer that employs a rotating cylinder. A complete set of characters is embossed on the circumference of the drum for each print position. A set of hammers is used to strike the drum (through the paper and ribbon) and print the proper character each time the drum rotates. (No longer used.)

drum programmer — An electromechanical device that provides stored program logic for control of a sequential operation such as batch processing or machine cycling. It ranks between relay and solid-state systems in the cost/complexity scale.

drum recorder — A facsimile recorder in which the record sheet is mounted on a rotating cylinder. (No longer used.)

drum sequencer — Mechanical programming device that can be used to operate switches or valves.

drum speed — The number of revolutions per minute made by the transmitting or receiving drum of a facsimile transmitter or recorder.

drum storage — A storage device in which information is recorded magnetically on a rotating cylinder; a type of addressable storage associated with some computers.

drum switch — A switch in which the electrical contacts are made on pins, segments, or surfaces on the periphery of a rotating cylinder.

drum transmitter — A facsimile transmitter in which the copy is mounted on a rotating cylinder. (No longer used.)

drum-type controller — 1. A multicircuit timing device, with or without a motor, using a cylindrical carriage into which pins are inserted to program events. 2. A multicircuit timing device intended to be driven from an external rotary power source.

drunkometer — A device measuring the degree of alcoholic intoxication by analyzing the subject's breath.

dry — 1. A condition in which the electrolyte in a cell is immobilized. The electrolyte may be either in the form of a gel or paste or absorbed in the separator material. 2. Said of circuits or contacts that do not carry direct current.

dry battery — Two or more dry cells arranged in series, parallel, or series-parallel within a single housing to provide desired voltage and current values.

dry cell — 1. A voltage-generating cell having an immobilized electrolyte. The commonest form has a positive electrode of carbon and a negative electrode of zinc in an electrolyte of sal ammoniac paste. 2. A source of energy produced by the reaction of an acid or alkaline paste on dissimilar metals or on a metal and a carbon electrode. Normal open-circuit voltage is 1.5 volts. The paste is sealed in a container in normal use, and the cell cannot be recharged.

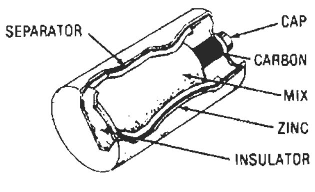

Dry cell.

dry-charged battery — A battery with the plates charged but lacking electrolyte. When ready to be placed in service, the electrolyte is added.

dry circuit — 1. A circuit in which current and voltage are so low that there is no arcing to roughen the contacts. As a result, an insulating film can develop that prevents an electrical closing of the circuit when contacts are brought together, if the proper means are not employed to prevent the initial formation of the film. 2. A circuit in which the open-circuit voltage is 0.03 V or less and the current 200 mA or less. (The voltage is most important because at such a low level it is not able to break through most oxides, sulfides, or other films that can build up on contacting surfaces.) 3. A telephone circuit over which voice signals are transmitted and which carries no direct circuit.

dry-circuit contact — A contact that carries current but neither makes nor breaks while its load circuit is energized. Sometimes erroneously used if referring to low level.

dry contacts — Contacts through which there is no direct current.

dry-disc rectifier — A rectifier consisting of discs of metal and other materials in contact under pressure. Examples are the copper-oxide and the selenium rectifier.

dry dross — Nonmetallic components normally found on still solder pots and drag machines. A dry dross may be mixed in with lux, oil, or any other organic material. The term *dry* refers basically to the absence of any metallic solder.

dry-electrolytic capacitor — An electrolytic capacitor with a paste rather than liquid electrolyte. By eliminating the danger of leakage, the paste electrolyte permits the capacitor to be used in any position.

dry flashover voltage — The voltage at which the air surrounding a clean, dry insulator or shell completely breaks down between electrodes.

dry-reed contact — An encapsulated switch containing two metal wires that serve as the contact points for a relay.

dry-reed relay — A relay that consists of one or more capsules containing contact mechanisms that are generally surrounded by an electromagnetic coil for actuation. The capsule consists of a glass tube with a flattened ferromagnetic reed sealed in each end. These reeds, which are separated by an air gap, extend into the tube so as to overlap. When placed in a magnetic field, they are brought together and close a circuit.

dry shelf life — The length of time that a cell can stand without electrolyte before it deteriorates to a point at which a specified output cannot be obtained.

dry-type forced-air-cooled transformer (class AFA) — A transformer that is not immersed in oil and that derives its cooling by the forced circulation of air.

dry-type self-cooled/forced-air-cooled transformer (class AA/FA) — A transformer that is not immersed in oil and that has a self-cooled rating with cooling obtained by the forced circulation of air.

dry-type self-cooled transformer (class AA) — A transformer that is cooled by the natural circulation of air and that is not immersed in oil.

dry-type transformer — A transformer that is cooled by the circulation of air and that is not immersed in oil.

dsc — Abbreviation for double silk-covered.

D-scope — A radar display similar to a C-scope except that the blips extend vertically to give a rough estimate of the distance.

"D" service — FAA service pertaining to radio broadcast of meteorological information, advisory messages, and notices to airmen.

DSP—Abbreviation for digital signal processor. A specialized chip and/or system that is dedicated to processing real-time signals. Typically used for modems, audio, imaging, and video applications.

DSU—*See* data service unit.

DT-cut crystal—A crystal cut to vibrate below 500 kHz.

DTE—*See* data termination equipment.

DTL—Abbreviation for diode-transistor logic.

DTMF—Abbreviation for dual-tone multifrequency (signaling).

DTP—Abbreviation for desktop publishing.

DTV—Abbreviation for digital television.

D-type flip-flop—A flip-flop that, on the occurrence of the leading edge of a clock pulse, propagates to the 1 output whatever information is at its D (data) conditioning input prior to the clock pulse.

dual—Either of a pair of systems, circuits, etc., that are described by equations of the same form in which the same functional relationships hold provided that the dependent and independent dynamic variables are interchanged between these equations.

dual-beam oscilloscope—An oscilloscope in which the cathode-ray tube produces two separate electron beams that may be individually or jointly controlled.

dual capacitor—Two capacitors within a single housing.

dual capstan—*See* closed loop drive.

dual-channel amplifier—An amplifier that has two channels independent of each other, but similar in design, construction, and output.

dual coaxial cable—Two individually insulated conductors laid parallel or twisted and placed within an overall shield and sheath.

dual cone—Speaker unit containing a main cone for bass and middle frequencies and a smaller, stiffer inner cone for treble frequencies, sometimes called a full-range speaker unit.

dual-diversity receiver—1. A radio receiver that receives signals from two different receiving antennas and uses whichever signal is the stronger at each instant to offset fading. In one arrangement, two identical radio-frequency systems, each with its own antenna, feed a common audio-frequency channel. In another arrangement, a single receiver is changed over from one antenna to the other by electronic switching at a rate fast enough to prevent loss of intelligibility. 2. The operation of combining two identical signals received over diverse paths to obtain an improvement of up to 3 dB in signal-to-noise ratio.

dual-emitter transistor—A passivated pnp silicon planar epitaxial transistor with two emitters; used as a low-level chopper.

dual feedhorn—A feedhorn that can simultaneously receive both horizontally and vertically polarized signals.

dual-frequency induction heater—A type of induction heater in which work coils operating at two different frequencies induce energy, either simultaneously or successively, to material within the heater.

dual-groove record—*See* Cook system.

dual in-line package—Abbreviated DIP. 1. A type of housing for integrated circuits. The standard form is a molded plastic container about $^3/_4$ inch long and $^1/_3$ inch wide (1.9 × 0.8 cm), with two rows of pins spaced 0.1 inch (2.5 cm) between centers. This package is more popular than the flat pack or TO-type can for industrial use because it is relatively inexpensive and is easily dip-soldered into printed circuit boards. 2. Carrier in which a semiconductor integrated circuit or other components, such as transistor, diodes, capacitors, inductors, resistors, or film hybrid circuits, are assembled and sealed. Package consists of a plastic or ceramic body with two rows of seven or more vertical leads that are inserted into a circuit board and secured by soldering. 3. The most common type of integrated-circuit package, which can be either plastic (DIP-plastic) or ceramic (CERDIP). Circuit leads or pins extend symmetrically outward and downward from opposite sides of the rectangular package body. "DIP, side-brazed" is a dual in-line package with leads brazed externally, on the sides of the package.

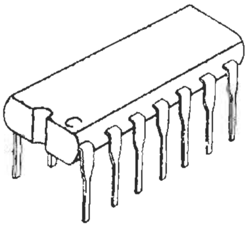

Dual in-line package.

dual meter—A meter constructed so that two aspects of a circuit may be read simultaneously.

dual-mode ignition—*See* DMI.

dual-mode phone—A phone that operates on both analog and digital networks.

dual modulation—The use of two different types of modulation, each conveying separate intelligence, to modulate a common carrier or subcarrier wave.

dual operation—A logic operation whose result is the negation of the result of an original operation when applied to the negation of its operands; for example, the OR operation is the dual of the AND operation. A dual operation is represented by writing 0 for 1 and 1 for 0 in the tabulated values of P, Q, and R for the original operation.

dual pickup—*See* turnover pickup.

dual rail—Also called double rail. Pertaining to a method of transferring data in which the data and the complement of the data are available on different input or output lines or wires.

dual slope converter—An integrating analog-to-digital converter in which the unknown signal is converted to a proportional time interval that is then measured digitally.

dual-tone multifrequency—Abbreviated DTMF. 1. A signaling method in which are employed set pairs of specific frequencies used by subscribers and PBX attendants, if their switchboard positions are so equipped, to indicate telephone address digits, precedence ranks, and end of signaling. 2. A portion of the touch-tone telephone dialing system developed by AT&T that combines two of a set of standard frequencies. The result of the combination is a third or beat frequency (signal) that is the desired or useable signal. DTMF signaling transports precisely defined and matched tone pair signals over sharply tuned amplitude-guarded channels to achieve highly error-immune operation.

dual-tone multifrequency signaling—Abbreviated DTMF signaling. Sending numerical address information from a telephone by sending, simultaneously, a combination of two tones out of a group of eight. The eight frequencies are 697, 770, 852, 941, 1209, 1336, 1477, and 1633 Hz.

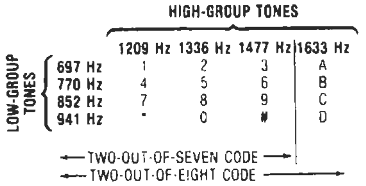

LOW-GROUP TONES	1209 Hz	1336 Hz	1477 Hz	1633 Hz
697 Hz	1	2	3	A
770 Hz	4	5	6	B
852 Hz	7	8	9	C
941 Hz	*	0	#	D

Dual-tone multifrequencies.

dual trace—A mode of operation in which a single beam in a cathode-ray tube is shared by two signal channels. *See also* alternate mode; chopped mode.

dual-track recorder—Normally a monophonic recorder in which the recording head covers slightly less than half the width of a standard quarter-inch tape, making it possible to record one track on the tape in one direction and, after turning the reels over, a second track in the opposite direction. Known also as half-track or two-track recorder.

dual-use line—A communications link that normally is used for more than one mode of transmission (e.g., for voice and data).

dub—Also called rerecording. 1. A copy of a recording. 2. To make a copy of a recording by recording on one machine what another machine is playing.

dubbing—1. In radio broadcasting, the addition of sound to a prerecorded tape or disc. 2. Copying of already recorded material. In tape recording, playing a tape or disc on one machine while recording it on another. The copy is called a dub.

duct—1. An overhead or underground protective tube or pipe used for carrying electrical conductors. 2. In microwave transmission, atmospheric conditions may cause radio waves to follow a narrower path than usual. The narrower path is called a duct. The presence of ducting sometimes causes unusual transmission because the transmission waves do not follow the intended path.

ducting—The trapping of an electromagnetic wave, in a waveguide action, between two layers of the earth's atmosphere, or between a layer of the atmosphere and the earth's surface.

DUF—Abbreviation for diffusion under the epitaxial film. A method for providing a low-resistance path between the active region of an IC transistor and the contact electrode at the surface. A region of high conductance is formed by selective diffusion in the required location prior to deposition of the epitaxial layer.

dumb terminal—1. Conversational slave to a host computer in a communication network. 2. An inexpensive means of interactive computer control that is good for online fixed-program applications. All software is in the host computer mainframe. A dumb terminal contains no user programming or memory for data manipulation. 3. The minimum equipment necessary to communicate with a computer. It consists of a monitor and a keyboard and will do little more than send and receive data. 4. A display terminal with no processing capabilities other than that associated with screen formatting. For processing, the terminal is entirely dependent on the main computer with which it communicates.

dummy—1. A device that can be substituted for another, but which has no operating features. 2. A telegraphy network used to simulate a customer's loop for adjustment of a telegraph repeater. The dummy side of the repeater is the side toward the customer. 3. In a computer, an artificial address, instruction, or other unit of information inserted solely for the purpose of fulfilling such prescribed conditions as word length or block length without affecting operations. 4. A simulating circuit that has no features.

dummy antenna—*See* artificial antenna.

dummying—The process of removing metallic impurities from plating solution with a large-area dummy cathode.

dummy instruction—An artificial instruction or address inserted in a list of instructions to a computer solely to fulfill prescribed conditions (such as word length or block length) without affecting the operation.

dummy load—*See* artificial load.

dummy variable—In a computer program, a symbol inserted at definition time, which will later be replaced by the actual variable.

dump—Also called power dump. 1. To withdraw all power from a computer, either accidentally or intentionally. 2. To transfer all or part of the contents of one section of a digital-computer memory into another section. 3. To transfer information from a register word to a memory position. 4. To transfer all of the information contained in a record into another storage medium. For example, a disc record could be dumped onto tape. 5. Copying contents of memory from one place to another. Same as memory dump.

dump check—Checking a computer by adding all digits as they are dumped (transferred) to verify the sum to make sure no errors exist as the digits are retransferred.

dump circuit—One form of transient suppression circuit, a self-biased snubber. The dump circuit is virtually inactive until the voltage across it exceeds the recent (slow) average value.

dumping resistor—A resistor whose function is to discharge a capacitor or network for safety purposes.

Dunmore cell—*See* lithium chloride sensor.

duodecal socket—A vacuum-tube socket having 12 pins. Used for cathode-ray tubes.

duodecimal—1. Pertaining to a characteristic or property involving a selection, choice, or condition in which there are 12 possibilities. 2. Pertaining to the numbering system with a radix of 12.

duodiode—Also called dual diode. A vacuum tube or semiconductor having two diodes within the same envelope.

duodiode-pentode—An electron tube containing two diodes and a pentode in the same envelope.

duodiode-triode—An electron tube containing two diodes and a triode in the same envelope.

duolateral coil—*See* honeycomb coil.

duopole—An all-pass action with two poles and two zeros.

duotriode—Also called double triode. An electron tube containg two triodes in the same envelope.

duplex—1. The method of operation of a communication circuit in which each end can simultaneously transmit and receive. (Ordinary telephones are duplex. When used on a radio circuit, duplex operation requires two frequencies.) 2. Two-in-one, as two conductors with a common overall insulation or two telegraph transmission channels over one wire. 3. Two conductors twisted together, usually with no outer covering. This term has a double meaning; it is possible to have parallel wires and jacketed parallel wires, and still refer to them as duplex. 4. Two-way data transmission. *Full duplex* describes two data paths that allow simultaneous data transmission in both directions. *Half duplex* describes one data path that allows data transmission in either of two directions, but only one direction at a time.

duplex artificial line—A balancing network simulating the impedance of the real line and distant terminal apparatus; it is employed in a duplex circuit for the purpose of making the receiving device unresponsive to outgoing signal currents.

duplex cable — A cable composed of two insulated stranded conductors twisted together. They may or may not have a common insulating covering.

duplex channel — A communication channel providing simultaneous transmission in both directions.

duplexer — 1. A radar device that, by using the transmitted pulse, automatically switches the antenna from receive to transmit at the proper time. 2. Highly selectable, tunable filters that allow a transmitter and receiver to use one common antenna.

duplexing assembly (radar) — *See* transmit-receive switch.

duplex operation — Simultaneous operation of transmitting and receiving apparatus at two locations.

duplex system — A system with two distinct and separate sets of facilities, each of which is capable of assuming the system function while the other assumes a standby status. Usually both sets are identical in nature.

duplex tube — A combination of two vacuum tubes on one envelope.

duplicate — To copy in such a way that the result has the same physical form as the source. For example, to make a new punched card that has the same pattern of holes as an original punched card.

duplication check — A computer check in which the same operation or program is checked twice to make sure the same result is obtained both times.

Duraspark — Conventional (Ford Motor) electronic ignition for use with high-voltage, high-energy spark timing control.

duration control — A control for adjusting the time duration of reduced gain in a sensitivity time control circuit.

duress alarm device — A device that produces either a silent alarm or local alarm under a condition of personnel stress such as holdup, fire, illness, or other panic or emergency. The device is normally manually operated and may be fixed or portable.

duress alarm system — An alarm system that employs a duress alarm device.

during cycle — The interval while a timer is operating for its preset time period.

dust core — A pulverized iron core consisting of extremely fine iron particles mixed with a binding material for use in radio-frequency coils.

dust cover — A device specifically designed to cover the mating end of a connector so as to provide mechanical and/or environmental protection.

DUT — Abbreviation for device under test.

duty cycle — 1. Ratio of working time to total time for intermittently operated devices. 2. The ratio of on-time to off-time in a periodic on-off cycle. 3. The ratio of operating time to total elapsed time of a device that operates intermittently, expressed as a percentage. 4. In percent, $100(t_o/T)$, where T is the period between pulses and t_o is the pulse width.

duty cyclometer — A test meter that gives a direct reading of duty cycle.

duty factor — 1. In a carrier composed of regularly recurring pulses, the product of their duration and repetition frequency. 2. Ratio of average to peak power. 3. Same as duty cycle except it is expressed as a decimal rather than a percentage. Usually calculated by multiplying pulses per second times pulse width.

duty ratio — In a pulsed system, such as radar, the ratio of average to peak power.

DVD — Abbreviation for the digital versatile disc and digital video disk. A high capacity optical storage medium with improved capacity and bandwidth over the CD for prerecorded multimedia program material with applications in video playback and PC areas.

dv/dt — The rate of change of voltage with respect to time. Proportional to current in a capacitor.

Dvorak keyboard — A keyboard arrangement designed by August Dvorak for increased speed and comfort. It reduces the rate of errors by placing the most frequently used letters in the center of the keyboard for use by the strongest fingers.

DX — 1. Abbreviation for distance. 2. Reception of distant stations. 3. Distant and/or difficult to hear radio stations.

DXer — One who listens to distant or hard-to-hear stations as a hobby.

DX hound — An amateur who specializes in making distant contacts.

DXing — The hobby of listening to distant or otherwise hard-to-hear stations.

dyadic Boolean operator — A Boolean operator that has two operands. The dyadic Boolean operators are AND, exclusive OR, NAND, NOR, and OR.

dyadic operation — An operation on two operands.

dye laser — A laser using a dye solution as its active medium. Its output is a short pulse of broad spectral content, and its achievable gain is high. Dye lasers function at room temperature. Synchronous pumping can be used to produce a continuous train of tunable picosecond pulses for sustained periods.

dynamic — 1. Of, concerning, or dependent on conditions or parameters that change, particularly as functions of time. 2. A speaker drive principle using the interaction between the magnetic field surrounding a voice coil carrying a signal current and a fixed magnetic field to move the coil and the cone to which it is attached. 3. A headphone driver using a voice coil in a magnetic field driving a paper or plastic diaphragm, as in a speaker.

dynamic acceleration — Acceleration in a constantly changing magnitude and direction, either simple or complex motion, usually called vibration. Also measured in gravity units.

dynamically balanced arm — A type of tonearm whose masses are balanced about its pivot, with tracking force applied by a spring. This type of arm does not require that the turntable be level for proper tracking.

dynamic analogies — The similarities in form between the differential equations that describe electrical, acoustical, and mechanical systems that allow acoustical and mechanical systems to be reduced to equivalent electrical networks, which are conceptually simpler than the original systems.

dynamic analysis — Execution of an instrumented program to collect information on its behavior and correctness.

dynamic behavior — The way a system or individual unit functions with respect to time.

dynamic braking — 1. A system of braking of an electric drive in which the motor is used as a generator, and the kinetic energy of the motor and driven machinery is employed as the actuating means of exerting a retarding force. 2. A type of motor braking caused by current being applied to the windings after the power is shut off. This is accomplished either by self-excitation (dc motors) or by separate excitation (ac motors).

dynamic burn-in — High-temperature test with devices subject to actual or simulated operating conditions.

dynamic cell — A memory cell that stores data as charge (or absence of charge) on a capacitor. A typical cell isolates the capacitor from the data line (bit line) with a transistor switch. Thus, when no read or write operation is desired, there is essentially no power required to maintain data. However, normal leakage requires that the charge be periodically restored by a process called refresh. Characteristics of a dynamic cell are very low data retention power, fewer transistors per bit (a one-transistor cell is common), and, usually, less area and lower cost per bit than for static cells.

dynamic characteristics—Relationship between the instantaneous plate voltage and plate current of a vacuum tube as the voltage applied to the grid is varied.

dynamic check—A check used to ascertain the correct performance of some or all components of equipment or a system under dynamic or operating conditions.

dynamic contact resistance—1. In a relay, a change in contact electrical resistance due to a variation in contact pressure on mechanically closed contacts. For example, during wiping motion of sliding contacts during make or prior to break. Also when contact members no longer actually open, as in contact bounce, but members are still vibrating and varying the contact pressure and hence its resistance. 2. A varying contact resistance on contacts mechanically closed.

dynamic convergence—1. The condition in which the three beams of a color picture tube come together at the aperture mask as they are deflected both vertically and horizontally. 2. A composite horizontal and vertical voltage used to ensure correct convergence of the three beams of a tricolor picture tube over the entire surface of the phosphor-dot faceplate. *See also* horizontal dynamic convergence; vertical dynamic convergence.

dynamic crosstalk—A condition (in an amplifier utilizing a single power supply) in which the demands made on one channel will effectively modulate the output of the other channel because the power supply feeding both is pumping up and down. To completely eliminate dynamic crosstalk may require separate and well-shielded power supplies for each channel.

dynamic decay—In a storage tube, decay caused by an action such as that of ion charging.

dynamic demonstrator—A three-dimensional schematic diagram in which the components of the radio, television receiver, etc., are mounted directly on the diagram.

dynamic deviation—The difference between the ideal output value and the actual output value of a device or circuit when the reference input is changing at a specified constant rate and all other transients have expired.

dynamic dump—A dump performed while a program is being executed.

dynamic electrode potential—The electrode potential measured when current is passing between the electrode and the electrolyte.

dynamic equilibrium of an electromagnetic system—1. The tendency of any electromagnetic system to change its configuration so that the flux of magnetic induction will be maximum. 2. The tendency of any two current-carrying circuits to maintain the flux of magnetic induction linking the two at maximum.

dynamic error—An error in a time-varying signal resulting from inadequate dynamic response of a transducer.

dynamic focus—The application of an ac voltage to the focus electrode of a color picture tube to compensate for the defocusing caused by the flatness of the screen.

dynamic headroom—The ability of an amplifier to produce more than its rated power for very short periods of time. An amplifier rated at 100 watts per channel with 3 dB of dynamic headroom can briefly produce 200 watts per channel.

dynamic magnetic field—A field whose intensity is changing and whose lines of force are expanding or contracting. Such change can be periodic or random. Unlike the static field, the dynamic field can transfer energy from one point to another without relative motion between the points.

dynamic memory—1. A type of semiconductor memory in which the presence or absence of an electrical or capacitive charge represents the two states of a binary storage element. Without refresh, the data represented by the electrical charge would be lost. 2. An MOS RAM memory using dynamic circuits. Each bit is stored as a charge on a single MOS transistor. This results in very high-density storage (only one transistor per bit). The charge leaks; therefore, a typical dynamic memory must be refreshed every 2 ms by rewriting its entire contents. In practice, this does not slow down the system, but requires additional memory refresh logic. (Dynamic chips are inexpensive and generally preferred to static ones for sizes over 16 K.)

dynamic memory allocation—Allocation of a limited main memory to successive programs in function of an allocation strategy based on priority, availability, or size.

dynamic microphone—*See* moving-coil microphone.

dynamic MOS array—A circuit made up of MOS devices that requires a clock signal. The circuit must be tested at its rated (operating) speed. Known as clock-rate testing.

dynamic mutual-conductance tube tester—*See* transconductance tube tester.

dynamic noise limiter—Abbreviated DNL. A compatible circuit designed primarily for use with tape recorders. It improves the effective signal-to-noise ratio during replay by selective filtering at low signal levels.

dynamic noise suppressor—An audio filter whose bandpass is adjusted automatically to the signal level. At low signal levels, filtering is highest; at high signal levels, all filter action is removed.

dynamic output impedance—*See* output impedance, 2.

dynamic pickup—A phonograph pickup whose electrical output is the result of the motion of a conductor in a magnetic field.

dynamic plate impedance—The internal resistance to the flow of alternating current between the cathode and plate of a tube.

dynamic plate resistance—*See* ac plate resistance.

dynamic power—*See* music power.

dynamic printout—In a computer, a printout of data that occurs as one sequential operation during the machine run.

dynamic problem check—A dynamic check used to ascertain that the solution determined by an analog computer satisfies the given system of equations.

dynamic programming—1. A procedure used in operations research for optimization of a multistage problem solution in which a number of decisions are available at each stage of the process. 2. A method of sequential decision making in which the result of the decision in each stage affords the best possible answer to exploit the expected range of likely (yet unpredictable) outcomes in the following decision-making stages.

dynamic RAM—*See* DRAM.

dynamic range—1. The difference, in decibels, between the overload level and the minimum acceptable signal level in a system or transducer. 2. The span of volume between the loudest and softest sounds, either in an original signal (original dynamic range) or within the span of a recorder's capability (recorded dynamic range). Dynamic range is expressed in decibels. *See* signal-to-noise ratio. 3. The range of signal amplitudes, from the loudest to the quietest, that can be reproduced effectively by an equipment. Limited by the intrinsic noise of the amplifier and the ambient background noise level of the listening environment and by the power capacity of the amplifier and speaker system. 4. The ratio between the maximum recorded level (usually that which results in 3-percent playback distortion) and the playback noises from a tape recorded with no signal input. Expressed in decibels. 5. The difference between the maximum acceptable signal level and the minimum acceptable signal level. 6. The ratio of the largest to the smallest values of range, often expressed in decibels. 7. The ratio of the maximum output signal to the smallest output signal that can be processed in a system, usually expressed logarithmically in dB. (Dynamic range can be specified in terms of harmonic distortion, signal-to-noise ratio, spurious-free dynamic range, or other ac input-based performance criteria.)

dynamic register — A memory in which the storage takes the form of capacitively charged circuit elements and therefore must be continually refreshed, or recharged, at regular intervals.

dynamic regulator — A transmission regulator in which the adjusting mechanism is in self-equilibrium at only one or a few settings and requires control power to maintain it at any other setting.

dynamic relocation — The ability to move computer programs or data from auxiliary memory to any convenient location in the memory. Normally the addresses of programs and data are fixed when the program is compiled.

dynamic reproducer — *See* moving-coil pickup.

dynamic resistance — Incremental resistance measured over a relatively small portion of the operating characteristic of a device.

dynamic router — A router that automatically broadcasts routing information throughout the Internet at regular intervals. Other dynamic routers use this information to update their routing tables in case any changes have been made to the network.

dynamic run — Also called dynamic test. 1. The test performed on an instrument to obtain the overall behavior and to establish or corroborate specifications such as frequency response, natural frequency of the device, etc. 2. Test based on a time-interval measurement as, for example, the rise time or fall time of a pulse. 3. A test of one or more of the signal properties or characteristics of equipment that is energized or in a nonquiescent state.

dynamic sequential control — A method of operation in which a digital computer can alter instructions or their sequence as the computation proceeds.

dynamic shift register — A shift register in which information is stored by means of temporary charge storage techniques. The major disadvantage of this method is that loss of the information occurs if the clock repetition rate is reduced below a minimum value.

dynamic skew — Short-term misalignment of the read head of a tape player as referenced to a master skew tape. It results from variations in tape-path geometry and tape-path alignment and slitting and "snaking" tolerances of magnetic tape.

dynamic speaker — *See* moving-coil speaker.

dynamic storage — 1. A data-storage device in which the data is permitted to move or vary with time in such a way that the specified data is not always immediately available for recovery. Magnetic drums and disks are permanent dynamic storage; acoustic delay line is a volatile dynamic storage. 2. Information storage using temporary charge storage techniques. It requires a clock repetition rate high enough to prevent loss of information.

dynamic storage allocation — A storage-allocation technique in which program and data locations are determined by criteria applied at the moment of need.

dynamic subroutine — In digital-computer programming, a subroutine that involves parameters (such as decimal point position) from which a properly coded subroutine is derived. The computer itself adjusts or generates the subroutine according to the parametric values chosen.

dynamic test — *See* dynamic run.

dynamic transfer-characteristic curve — A curve showing the variation in output current as the input current changes.

dynamo — 1. Normally called a generator. A machine that converts mechanical energy into electrical energy by electromagnetic induction. 2. In precise terminology, a generator of direct current — as opposed to an alternator, which generates alternating current.

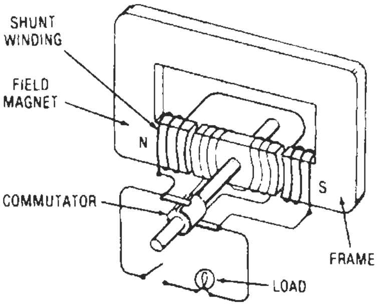

Dynamo.

dynamoelectric — Pertaining to the relationship between mechanical force and electrical energy or vice versa.

dynamometer — 1. An instrument in which the force between a fixed and a moving coil provides a measure of current, voltage, or power. 2. Equipment designed to measure the power output of a rotating machine by determining the friction absorbed by a hand brake opposing the rotation.

dynamotor — Also called a rotary converter or synchronous inverter. 1. A rotating device for changing a dc voltage to another value. It is a combination electric motor and dc generator with two or more armature windings and a common set of field poles. One armature winding receives the direct current and rotates (thus operating as a motor), while the others generate the required voltage (and thus operate as a dynamo or generator). 2 A rotary electrical machine used to convert direct current to alternating current. The machine has a single field structure and a single rotating armature having two windings, one equipped with a dc commutator and the other with ac slip rings.

dynaquad — A germanium pnpn semiconductor switching device that is base controlled and has three terminals. Its operation is similar to that of a flip-flop circuit or latching relay.

dynatron — Also called negatron. A type of vacuum tube in which secondary emission of electrons from the plate causes the plate current to decrease as the plate voltage increases, with the result that the device exhibits a negative-resistance characteristic. Used in oscillator circuits. *See* tetrode.

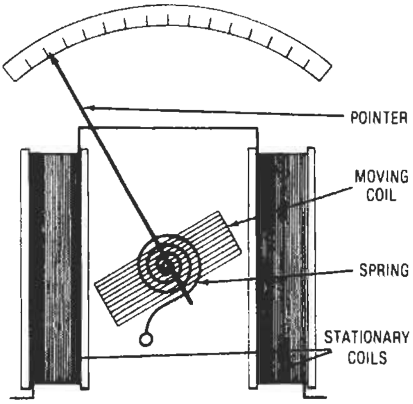

Dynamometer, 1.

dynatron oscillator—A negative-resistance oscillator with negative resistance derived between the plate and cathode of a screen-grid tube operating such that secondary electrons produced at the plate are attracted to the higher-potential screen grid.

dyne—The fundamental unit of force in the cgs system that, if applied to a mass of 1 gram, would give it an acceleration of 1 cm/s^2.

dyne per square centimeter—Also called microbar. The unit of sound pressure. One dyne per square centimeter was originally called a bar in acoustics, but the full expression is used in this field now because the bar is defined differently in other applications.

dynistor—A nonlinear semiconductor having the characteristics of a small current flow as voltage is applied. As the applied voltage is increased, a point is reached at which the current flow suddenly increases radically and will continue at this rate even though the applied voltage is reduced.

dynode—1. An electrode having the primary function of supplying secondary electron emission in an electron tube. 2. The auxiliary electrode that, when functioning within a photomultiplier tube and when bombarded by photoelectrons, gives rise to secondary emission and amplification.

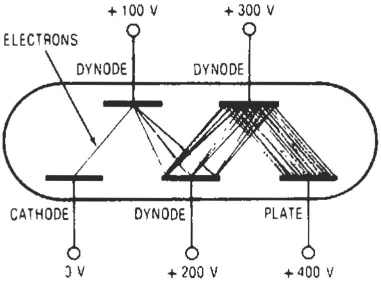

Dynode.

E

E—1. Symbol for voltage or emitter. 2. Abbreviation for illumination.

E and M leads—In a signaling system, the output and input leads, respectively.

E- and M-lead signaling—Communications between a trunk circuit and a separate signaling unit by way of two leads: an M lead over which battery or ground signals are transmitted to the signaling equipment, and an E lead over which open or ground signals are received from the signaling unit.

early failure—A failure that occurs during the initial life phase and is generally caused by initial production, assembly, test, installation, or commissioning errors.

early-failure period—Also called debugging period, burn-in period, or infant-mortality period. The period of equipment life, starting immediately after final assembly, during which equipment failures initially occur at a higher than normal rate due to the presence of defective parts and abnormal operating procedures.

early-warning radar—A radar that usually scans the sky in all directions in order to detect approaching enemy planes and/or missiles at distances far enough away that interceptor planes can be in the air to meet their approach before they are near their target.

EAROM—Abbreviation for electrically alterable ROM. 1. A read-only memory (ROM) that can be erased and reprogrammed any number of times. 2. Device that resembles an EPROM except that electric current rather than ultraviolet light does the erasing. This is an expensive type of PROM because it requires complete read/write control logic. Use is restricted mainly to applications such as machine controllers in which operators must change programs regularly. 3. Similar to the EPROM, the EAROM can be erased by a sort of reversed programming with a high voltage. The EAROM is thus like a RAM that will not lose its data if power is removed.

earphone—Also called receiver. 1. An electroacoustic transducer intended to be placed in or over the ear. 2. An electroacoustic device that transforms electric waves into sound waves. It is intended to be closely coupled, acoustically, to the ear.

earth—Term used in Great Britain for ground.

earth conductivity—The conductance between opposite faces of a unit cube (usually cubic meter) of a given earth material (loam, clay, sand, rock, etc.). The volume conductivity of this earth sample is the reciprocal of its volume (not to be confused with surface) resistivity.

earth current—Also called ground current. 1. Current in the ground as a result of natural causes and affecting the magnetic field of the earth, sometimes causing magnetic storms. 2. Return, fault, leakage, or stray current passing from electrical equipment through the earth.

earthed—A British term meaning grounded.

earth ground—1. A connection from an electrical circuit or equipment to the earth through a water pipe or a metal rod driven into the earth. This connection reduces shock hazards from faulty equipment. Water pipes may no longer be reliable grounds because of the use of transite pipe, neoprene gaskets, and other nonconducting links. Any ground rods driven under the interior of a large building may gradually become ineffective because the building may drive the local water table down so far that the rod is essentially surrounded by dry soil. 2. An actual connection into the surface of the earth by way of a metal or chemical rod or a wire connected to such a conductor.

earth inductor—*See* generating magnetometer.

earth-layer propagation—1. Propagation of electromagnetic waves through layers in the atmosphere of the earth. 2. Propagation of electromagnetic waves through layers below the surface of the earth.

earth oblateness—The slight departure from a perfect spherical shape of the earth and the form of its gravity field.

earth permittivity—The ratio of a capacitor's capacitance using our earth sample as a dielectric to that with air as a dielectric. Permittivity has also variously been termed dielectric constant, specific inductive capacity, and capacitivity.

earth station—The term used to describe the combination of antenna, low-noise amplifier (LNA), downconverter, and receiver electronics; used to receive a signal transmitted by a satellite. Earth station antennas vary in size from the 2-foot to 12-foot (65 centimeters to 3.7 meters) diameter size used for TV reception to those as large as 100 feet (30 meters) in diameter sometimes used for international communications. The typical antenna used for INTELSAT communication today is 13 to 18 meters or 40 to 60 feet in diameter.

EAS—Abbreviation for extensive area service.

EAX—Abbreviation for electronic automatic exchange.

EBCDIC—Abbreviation for extended binary coded decimal interchange code. A coding scheme wherein letters, numbers, and special symbols are represented as unique eight-bit values, allowing for standardization between data communications devices; popularized by IBM. The code can accommodate 256 characters.

E-bend—*See* E-plane bend.

ebiconductivity—Conductivity induced as the result of electron bombardment.

ebmd—Abbreviation for electron-beam mode discharge.

EBS amplifier—Abbreviation for electron-bombarded semiconductor amplifier.

ec—Abbreviation for enamel covered.

eccentric circle—*See* eccentric groove.

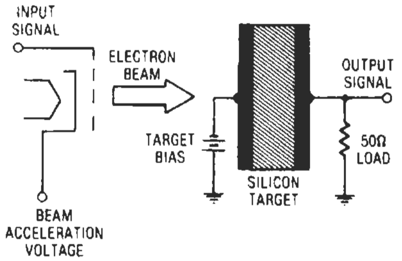

EBS amplifier.

eccentric groove—Also called eccentric circle. An off-center locked groove for actuating the trip mechanism of an automatic record changer at the end of a recording.

eccentricity—In disc recording, the displacement of the center of the recording-groove spiral with respect to the record center hole.

Eccles-Jordan circuit—A flip-flop consisting of a two-stage, resistance-coupled amplifier. Its output is coupled back to its input, two separate conditions of stability being achieved by alternately biasing the two stages beyond cutoff.

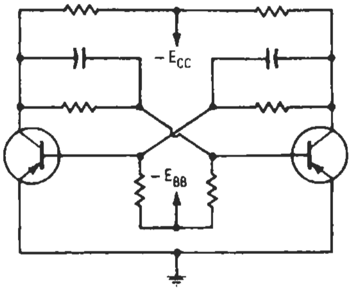

Eccles-Jordan multivibrator.

ECCM—Abbreviation for electronic counter-countermeasures.

ECG—*See* electrocardiogram.

echelon—One of a series of levels of accuracy of calibration, the highest of which is represented by an accepted national standard. There may be auxiliary levels between two successive echelons.

echo—1. In radar, that portion of the energy reflected to the receiver from a target. 2. A wave that has been reflected or otherwise returned with sufficient magnitude and delay to be distinguishable from the directly transmitted wave. 3. In facsimile, a multiple reproduction on the record sheet caused by the arrival of the same original facsimile signal at different times over transmission paths of different lengths. 4. In a radio system, an electronic condition that causes a signal such as a voice signal to be reflected from some point or points in the circuit back to the point of origination of the signal. 5. A delayed repetition (sometimes several rapid repetitions) of the original sound or signal. 6. In tape recording, refers to a provision for picking up some of the sound from a play head while recording, and feeding it back to the record head to produce a rapidly periodic repetition of each sound. Correct echo-volume adjustment causes a decay of the repeated sounds to simulate acoustical reverberation. 7. A special recording effect, in which a portion of the recorded program is taken from the playback head, a short interval after being recorded, and mixed with the incoming program. Principally used at tape speeds greater than 3¾ ips (9.5 cm/s), where the delayed signal is not heard as a separate sound. 8. An instantaneous repetition of the sound heard in playing some tapes or other recordings. This is caused by print through. 9. A portion of the transmitted signal returned from a distant point to the transmitted source with sufficient time delay to be received as interference. 10. A signal that has been reflected at one or more points during transmission with sufficient magnitude and time difference as to be detected as a signal distinct from that of the primary signal. Echoes can either lead or lag the primary signal and appear as reflections or ghosts. 11. Displaying information sent or received on a terminal, to visually detect transmission errors. Remote echo comes from the host computer. Local echo comes from the sender's transmission. 12. The distortion created when a signal is reflected back to the originating station.

echo area—Equivalent echoing area of a radar target (i.e., the relative amount of radar energy the target will reflect).

echo attenuation—In a four-wire (or two-wire) circuit equipped with repeater or multiplex equipment in which the two directions of transmission can be separated from each other, the attenuation of the echo currents (which return to the input of the circuit under consideration) is determined by the ratio of the transmitted power to the echo power received.

echo box—Also called phantom target. A device for checking the overall performance of a radar system. It comprises a resonant cavity that receives a portion of the pulse energy from the transmitter and retransmits it to the receiver as a slowly decaying transient. The time required for this transient response to decay below the minimum detectable level on the radar indicator is known as the ring time and is indicative of the overall performance of the radar set.

echo cancellation—1. A circuit that uses digital signal processing technology in a full-duplex communications node to remove echoes of the transmitted signal from the received signal. 2. The technique used in modems to filter out unwanted signals.

echo canceller—An electronic circuit that attenuates or eliminates the echo effect on satellite telephony links. Echo cancellers are largely replacing obsolete echo suppressors.

echocardiogram—Ultrasound image of the structure and motions of the heart.

echocardiography—A sonarlike noninvasive method of diagnosing cardiac malfunctions. A pulsed ultrasound beam is directed through the chest wall and echoes reflected from differing tissue interfaces (e.g., soft tissue and blood; tissue and bone, etc.) are detected by the barium titanate transducer that originated the pulse. Echoes are recorded against a linear time base to provide a hard-copy record of cardiac structure movement in a given time span, generally defined by the ECG.

echo chamber—A reverberant room or enclosure used for adding hollow effects or actual echoes to radio or television programs.

echo check—A method of checking the accuracy of transmission of data in which the received data is returned to the sending end for comparison with the original data.

echo checking—A method of checking in which transmitted information is reflected back to the transmitting point and compared with what was sent.

echo distortion—1. A modulation-related impairment created when one or more delayed echo signals introduce a phase-sensitive delay ripple across the FM portion of a system. The amplitude of this distortion is a complex function involving the relative magnitude and delay time of the echo signal with respect to the main signal, the level of baseband loading present, and the relative position of the channel in the baseband. 2. A telephone-line impairment caused by electrical reflections at distant points where line impedances are dissimilar.

echoencephaloscope—An ultrasonic instrument for use in brain studies. A transducer that generates a series of ultrasonic pulses and detects the returning echoes is placed against the patient's head. Each pulse is displayed together with its associated echoes on a cathode-ray tube.

echo intensifier—A device, located at the target, that is used to increase the amplitude of the reflected energy to an abnormal level.

echo matching—Rotating an antenna to a position in which the pulse indications of an echo-splitting radar are equal.

echo splitting—In certain radar equipment, the echo return is split and appears as a double indication on the screen of the radar indicator. This splitting is accomplished by special electronic circuits associated with the antenna lobe switching mechanism. When the two echo indications are of equal height, the target bearing is read from a calibrated scale.

echo suppression—1. A control used to disable a responder for a short interval of time so that reception of echoes of the interrogator pulse from nearby targets is prevented. 2. A circuit used to eliminate reflected waves.

echo suppressor—1. A voice-operated device that is connected to a two-way telephone circuit to attenuate echo currents in one direction caused by telephone currents in the other direction. 2. In navigation, a circuit that desensitizes the equipment for a fixed period after the reception of one pulse, for the purpose of rejecting delayed pulses arriving from indirect reflection.

echo talker—A portion of the transmitted signal returned from a distant point to the transmitting source with sufficient time delay to be received as interference.

ECL—Abbreviation for emitter-coupled logic. 1. A type of unsaturated logic performed by emitter-coupled transistors. Higher speeds may be achieved with ECL than are obtainable with standard logic circuits. 2. An IC logic family characterized by its very high speed of operation, low circuit density per chip, and very high power dissipation when compared with other IC logic families. Used mainly in large, very high-speed digital computers and sold mainly on a custom-designed basis. 3. A type of current-mode logic in which the circuits are coupled with one another through emitter followers at the input or output of the logic circuit. 4. A bipolar digital IC family that uses a more complex design than TTL to speed up IC operations. Emitter-coupled logic is costly, power hungry, and difficult to use, but it is four times faster than TTL. 5. A family of nonsaturated logic devices that operate at very high speed. ECL logic dissipates relatively large amounts of power, requires a bias supply, and is characterized by low component density.

ECL bipolar memories—Very high-speed cache, writable control stores, and processing sections of large computers.

ECM—Abbreviation for electronic countermeasures.

ECO—Abbreviation for electron-coupled oscillator.

E-core—The laminated configuration resembling the capital letter E in some transformers and inductive transducers.

ECU—Abbreviation for electronic control unit. 1. Electronic ignition system (Chrysler Corp.) that replaced the breaker-point ignition system. Provides spark timing to the high-energy coil and spark plugs. 2. A high-energy ignition system using a Hall effect semiconductor pickup in the distributor instead of the conventional reluctance pickup.

EDC—Abbreviation for error detection code. 32 bits in each sector that are used to detect errors in the sector data of a CD.

EDCT—Abbreviation for electrochemical diffused-collector transistor. A pnp transistor in which all the mass of p material is etched off and replaced with metal, which acts as a heat sink. It is suitable for high-current, high-speed core driver and computer-memory applications.

eddy current clutch—A device that permits connection between a motor and a load by electrical (magnetic) means—no physical contact is involved. This method is also used for speed control by clutch slippage.

eddy-current heating—Synonym for induction heating.

eddy-current loss—The core loss that results when a varying induction produces electromotive forces that cause a current to circulate within a magnetic material.

eddy currents—Also called Foucault currents. 1. Those currents induced in the body of a conducting mass by a variation in magnetic flux. 2. Circulating currents induced in conducting materials by varying magnetic fields. Usually undesirable because they represent loss of energy and cause heating. 3. Induced currents through an iron core in a transformer. They cause a waste of power.

edge-board connector—Also called card-edge connector. A connector that mates with printed wiring leads running to the edge of a printed circuit board.

edge-board contact—A series of contacts printed on or near any edge of a printed board and intended for mating with an edge connector.

edge connector—1. A one-piece receptacle, containing female contacts, designed to receive the edge of a printed circuit board and interconnect on which the male contacts are etched or printed. The connector may contain either a single or double row of female contacts. Both thermoplastic and thermosetting insulating materials are used. 2. A connector designed to mate with printed circuit boards. May be equipped with a polarizing pin or a key to ensure correct polarity. 3. A row of etched lines on the edge of a printed circuit board that is inserted into a motherboard or an expansion slot of a computer.

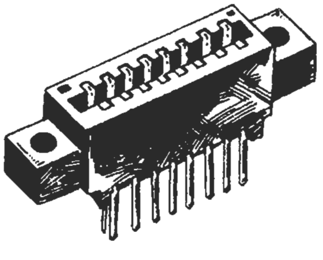
Edge connector.

edge effect—1. *See* following blacks; following whites; leading blacks; leading whites. 2. Nonuniformity of electric fields between two parallel plates caused by an outward bulging of electric flux lines at the edges of the plates. 3. The bulging out of the lines of the electric field at the edges of two parallel electrodes, created by a potential difference between the electrodes.

edge-triggered flip-flop—A type of flip-flop in which some minimum clock-signal rate of change is one necessary condition for an output change to occur.

edging—Undesired coloring around the edges of different-colored objects in a color television picture.

EDI—Abbreviation for electronic data interchange. A format in which business data is represented using national or international standards.

Edison base—Standard screw-thread base used for ordinary electric lamps.

Edison distribution system—A three-wire direct-current distribution system, usually 120 to 240 volts, for combined light and power service from a single set of mains.

Edison effect—Also called Richardson effect. The phenomenon wherein electrons emitted from a heated element within a vacuum tube will flow to a second element that is connected to a positive potential.

Edison storage cell—A storage cell having negative plates of iron oxide and positive plates of nickel oxide immersed in an alkaline solution. An open-circuit voltage of 1.2 volts per cell is produced.

E-display—In a radar, a rectangular display in which targets appear as blips with distance indicated by the horizontal coordinate and elevation by the vertical coordinate.

edit—1. To arrange or rearrange output information from a digital computer before it is printed out. Editing may involve deleting undesired information, selecting desired information, inserting invariant symbols, such as page numbers and typewriter characters, and applying standard processes such as zero suppression. 2. To deliberately modify the user program in a memory.

editing—1. The rearrangement of recorded material to provide a change of content or form, or for replacement of imperfect material. Usually accomplished by cutting and splicing the tape. 2. Revising text with a word processor to create an updated document.

editor—1. A program allowing such text editing functions as addition of a line or character, insertion, or deletion to permit altering of a program. Input data can be anything from programs or reports to raw instrument data. 2. Program that permits data or instructions to be manipulated and displayed. Most common use is in the preparation of new programs. 3. Program that takes the source program, written by the programmer in assembly or high-level language and entered through a keyboard or paper tape, and transfers it to a file in the computer's auxiliary memory, such as magnetic disk or tape. The editor also acts on special commands from the user to add, delete, or replace portions of the source program in the auxiliary memory. Editors can vary significantly in the ease with which they permit a user to make changes in the program. For example, some editors can operate only on entire lines in a program, whereas others can add, delete, or replace arbitrary character strings in the program. However, the less-sophisticated editors are usually easier to learn to use. 4. An interactive software subsystem that allows users to modify test programs directly on an automatic system. 5. A program that permits a user to create new files in symbolic form or to modify existing files. 6. A program that permits a series of mnemonic instructions comprising a user program to be displayed, analyzed, corrected, and otherwise modified quickly and easily on a CRT screen or other terminal device.

EDP—Abbreviation for electronic data processing.

EDP center—*See* electronic data-processing center.

EDPM—Abbreviation for electronic data-processing machine.

EEC—Abbreviation for electronic engine control. Precision control of engine spark timing and exhaust gas recirculation for emissions control and good fuel mileage (Ford Motor). Uses a microprocessor chip.

EEC IV—The fourth progressive version of the electronic engine control (Ford Motor) using the latest technology of microprocessor chips to control engine spark timing, EGR, and feedback carburetor fuel management.

EED—Abbreviation for electroexplosive device. An electroexplosive device for use with an air cushion that automatically inflates to protect the driver when an automobile equipped with such a cushion is involved in a collision.

EEG—*See* electroencephalograph and electroencephalogram.

EEG electrode—Electrode that attaches to scalp for detecting brain waves.

EEPROM or E^{2}PROM—Abbreviation for electrically erasable programmable read-only memory. 1. A field-programmable read-only memory in which cells may be erased electrically and each cell may be reprogrammed electrically. The number of times the E^{2}PROM can be reprogrammed (write/erase cycles) ranges from 10 times to 10^6 times. 2. Similar to PROM, but with the capability of selective erasure of information through special electrical stimulus. Information stored in EEPROM chips is retained when the power is turned off (compare with PROM). *See also* electronically programmable read-only memory.

effective acoustic center—Also called apparent source. The point from which the spherically divergent sound waves from an acoustic generator appear to diverge.

effective actuation time—The sum of the initial actuation time and the contact chatter intervals of a relay following such actuation.

effective address—The address that is actually used in carrying out a computer instruction.

effective ampere—That alternating current which, when flowing through a standard resistance, produces heat at the same average rate as 1 ampere of direct current flowing in the same resistance.

effective antenna length—The length that, when multiplied by the maximum current, will give the same product as the length and uniform current of an elementary electric dipole at the same location, and the same ratio field intensity in the direction of maximum radiation.

effective aperture delay—In a sample-and-hold circuit, the time difference between the hold command and the time at which the input signal is at the held voltage.

effective area—The effective area of an antenna in any specified direction is equal to the square of the wavelength multiplied by the power gain (or directive gain) in that direction, divided by 4π.

effective bandwidth—For a bandpass filter, the width of an assumed rectangular bandpass filter having the same transfer ratio at a reference frequency and passing the same mean-square value of a hypothetical current and voltage having even distribution of energy over all frequencies.

effective capacitance—The total capacitance existing between any two given points of an electric circuit.

effective conductivity—The conductance between the opposite parallel faces of a portion of a material having unit length and unit cross section.

effective confusion area—Amount of chaff whose radar cross-sectional area equals the radar cross-sectional area of a particular aircraft at a particular frequency.

effective current—That value of alternating current that will give the same heating effect as the corresponding value of direct current. For sine-wave alternating currents, the effective value is 0.707 times the peak value.

effective cutoff — *See* effective cutoff frequency.

effective cutoff frequency — Also called effective cutoff. The frequency at which the insertion loss of an electric structure between specified terminating impedances exceeds the loss at some reference point in the transmission band.

effective facsimile band — A frequency band equal in width to the difference between zero frequency and the maximum keying frequency of a facsimile signal.

effective field intensity — Root-mean-square value of the inverse distance fields 1 mile from the transmitting antenna in all directions horizontally.

effective height — 1. The height of the antenna center of radiation above the effective ground level. 2. In loaded or nonloaded low-frequency vertical antennas, a height equal to the moment of the current distribution in the vertical section divided by the input current.

effective irradiance to trigger — The minimum effective irradiance required to switch a light-activated silicon-controlled rectifier from the off state to the on state.

effective isotropically radiated power — *See* EIRP.

effectively grounded — Grounded through a ground connection of sufficiently low impedance (inherent and/or intentionally added) so that fault grounds that may occur cannot build up voltages that are dangerous to connected equipment.

effectiveness — The capability of the system or device to perform its function.

effective parallel resistance — The resistance considered to be in parallel with a pure dielectric.

effective percentage modulation — For a single sinusoidal input component, the ratio between the peak value of the fundamental component of the envelope and the direct-current component in the modulated conditions, expressed as a percentage.

effective radiated power — 1. The product of the radio-frequency power, expressed in watts, delivered to an antenna and the relative gain of the antenna over that of a half-wave dipole antenna. 2. The product of the antenna power (transmitter power less transmission-line loss) times either the antenna power gain or the antenna field gain squared. Where circular or elliptical polarization is employed, the term *effective radiated power* is applied separately to the horizontal and vertical components of radiation. For allocation purposes, the effective radiated power authorized is the horizontally polarized component of radiation only. If specified for a particular direction, it is the antenna power gain in that direction only.

effective radius of the earth — A value used in place of the geometrical radius to correct the atmospheric refraction when the index of refraction in the atmosphere changes linearly with height. Under conditions of standard refraction, the effective radius is one and one-third the geometrical radius.

effective resistance — 1. The average rate of dissipation of electric energy during a cycle divided by the square of the effective current. 2. The equivalent pure dc resistance that, when substituted for the winding of a motor being checked, will draw the same power. It is also equivalent to the impedance of a circuit having a capacitor connected in parallel with the winding and the capacitor adjusted to unity power factor for the circuit.

effective series resistance — A resistance considered to be in series with an assumed pure capacitance inductance.

effective sound pressure — The root-mean-square value of the instantaneous sound pressure at one point over a complete cycle. The unit is the dyne per square centimeter.

effective speed — The speed (less than rated) that can be sustained over a significant period of time and that reflects slowing effects of control codes, timing codes, error detection, retransmission, tabbing, hand keying, etc.

effective speed of transmission — Also called average rate of transmission. The average rate over some specified time interval at which information is processed by a transmission facility. Usually expressed as average characters or average bits per unit time.

effective thermal resistance — Of a semiconductor device, the effective temperature rise per unit power dissipation of a designated junction above the temperature of a stated external reference point under conditions of thermal equilibrium.

effective value — Also called the rms (root-mean-square) value. The value of alternating current that will produce the same amount of heat in a resistance as the corresponding value of direct current. For a sine wave, the effective value is 0.707 times the peak value.

effective wavelength — The wavelength corresponding to the effective propagation velocity and the observed frequency.

efficiency — 1. Ratio of the useful output of a physical quantity that may be stored, transferred, or transformed by a device to the total input of the device. 2. Ratio of the output power to the input power of a power supply, usually expressed as a percentage of the input power measured at nominal line and load conditions. That part of the input not appearing in the output is converted into heat, which must be conducted away from the power supply circuitry. Efficiency is determined to a great extent by the method of regulation and is expressed as a percentage. In the absence of statements to the contrary, it is assumed to be taken at nominal input and output levels and full load conditions. 3. The percentage of the electrical input power to a speaker that is converted to acoustic energy. Varies from a small fraction of 1 percent to as much a 10 percent or more, depending on the design of the speaker. Higher efficiency means that less electrical amplifier power is required for a given listening volume, but it is not directly related to sound quality.

efficiency of a source of light — The ratio of the total luminous flux to the total power consumed. In the case of an electric lamp, it is expressed in lumens per watt.

efficiency of rectification — Ratio of direct-current power output to alternating-current power input of a rectifier.

***E*-field sensor** — A passive sensor that detects changes in the earth's ambient electric field caused by the movement of an intruder. *See also H*-field sensor.

EFTS — Abbreviation for electronic funds transfer system. A payments system in which the processing and communications necessary to effect economic exchange and the processing and communications necessary for the production and distribution of services incidental or related to economic exchanges are dependent wholly or in large part on the use of electronics.

EGA — Abbreviation for enhanced graphics adapter. A color graphics system for IBM PCs and compatibles that supports 16 colors. Provides higher resolution than CGA, lower than VGA.

EHF — Abbreviation for extremely high frequency.

E-H tee — A waveguide junction composed of a combination of E- and H-plane tee junctions that intersect and the main guide at a common point.

E-H tuner — An E-H tee having two arms terminated in adjustable plungers. It is used for impedance transformation.

EIA—*See* Electronic Industries Association.

EIA interface—1. A set of signal properties (time duration, voltage, and current) specified by the Electronic Industries Association for business machine/data set connections. 2. A standardized set of signal characteristics for connection of terminals to modem units.

EIA RS-232-C standard—A set of specifications used throughout the data communications industry to define the interconnection of data terminal equipment (DTE) and data communication equipment (DCE) for the exchange of serial binary data. This standard defines electrical signal characteristics, mechanical interface characteristics, and circuits.

EIC—Abbreviation for electronic instrument cluster (Chrysler Corp.). Digital instrument display for speed, miles, fuel level, fuel consumption clock, etc.

eight-bit chip—A CPU chip that processes data eight bits at a time.

eight-bit color—The color range possible with an eight-bit-graphics system. In such a system, each pixel can display one of 256 colors at any given time.

eight-level code—A code in which eight impulses are utilized for describing a character. Start and stop elements may be added for asynchronous transmission. The term is often used to refer to the U.S. ASCII code.

eight-track—Most commonly, a cartridge tape system having eight narrow tracks on $^1/_4$-inch (6.35-mm) tape wound in a continuous loop around a single hub.

eight-track tape-recording format—Either of two professional tape recording formats (half-track or quarter-track) in whicb eight independent channels can be recorded in the same direction.

E-indicator—A rectangular radar display in which the horizontal coordinate of a target blip represents range and the vertical coordinate represents elevation.

einstein—A unit of energy equal to the amount of energy absorbed by one molecule of material undergoing a photochemical reaction, as determined by the Stark-Einstein law.

Einthoven string galvanometer—A moving-coil type of galvanometer in which the coil is a single wire suspended between the poles of a powerful electromagnet.

E-I pick-off—An assembly of transformer-like laminations, the output coils of which develop a voltage proportional to the displacement of a magnetic element from the neutral position for limited rotary as well as angular travel.

EIRP—Abbreviation for effective (or equivalent) isotropically radiated power. 1. A measure of the signal strength that a satellite transmits toward the earth below. The EIRP is highest at the center of the beam and decreases at angles away from the boresight. 2. The gain of an antenna in a given direction, multiplied by the net power accepted by the antenna from the transmitter.

EKG—Abbreviation for electrocardiograph.

elastance—1. Symbolized by S. In a capacitor, the ratio of potential difference between its electrodes to the charge in the capacitor. It is the reciprocal of capacitance. The unit of measure is the daraf.

$$S\,(\text{daraf}) = V/Q$$

2. A measure of the difficulty of placing an electric charge in a capacitor.

elasticity—The resistance of an electrostatic field. It is the reciprocal of permittivity.

elastic wave—A pure acoustic wave; a moving lattice distortion without a magnetic component.

elastomer—1. A material that has the ability to recover from extreme deformation, in the order of hundreds of percent. It may be thermosetting or thermoplastic. 2. A material that at room temperature stretches under low stress to at least twice its length and snaps back to the original length on release of stress.

E layer—1. One of the regular ionospheric layers, with an average height of about 100 kilometers or 60 to 70 miles. This layer occurs during daylight hours, and its ionization is dependent on the sun's angle. The principal layer corresponds roughly to what was formerly called the Kennelly-Heaviside layer. 2. The lowest layer of the ionosphere that supports long-distance radio communication.

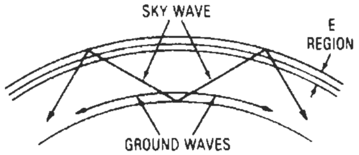

E layer.

elbow—In a waveguide, a bend with a relatively short radius and an angle normally of 90° but sometimes, for acute angles, down to 15°.

el casel—A tape system that uses a cassette similar in plan to the compact cassette, but holding $^1/_4$-inch (6.35-mm) tape running at $3^3/_4$ inches per second (9.5 cm/s); the tape is looped out of the cassette to reach the heads and capstan during recording and playback. In other cassette systems, the heads contact the tape through windows in the cassette shell.

electra—A specific radionavigational aid that provides a number (usually 24) of equasignal zones. Electra is similar to sonne except that in sonne the equasignal zones as a group are periodically rotated in bearing.

electret—1. A permanently polarized piece of dielectric material produced by heating the material and placing it in a strong electric field during cooling. Some barium titanate ceramics can be polarized in this way, and so can carnauba waxes. The electric field of an electret corresponds somewhat to the magnetic field of a permanent magnet. 2. A special plastic piezoelectric element, polarized during manufacture to become the equivalent of a permanently charged capacitor. Generates an amplitude-responsive output voltage, like a ceramic element, but requires less energy from the stylus system. By loading with an appropriate resistance, its output can be converted to the equivalent of a magnetic cartridge's velocity-responding output characteristic.

electric—Containing, producing, arising from, actuated by, or carrying electricity, or designed to carry electricity and capable of so doing. Examples: electric eel, energy, motor, vehicle, wave.

electrical—Related to, pertaining to, or associated with electricity but not having its properties or characteristics. Examples: electrical engineer, handbook, insulator, rating, school, unit.

electrical angle—A quantity that specifies a particular instant in a cycle of alternating current. One cycle is considered to be 360°, so a half cycle is 180° and a quarter cycle is 90°. If one voltage reaches its peak value a quarter cycle after another, the phase difference, or electrical angle between the voltages, is 90°.

electrical bail—A switch action in which, upon actuation of one station, the switch changes the contact position, electrically locks the switch in that position, and releases any station previously actuated.

electrical bias—An electrically produced force tending to move the armature of a relay toward a given position.

electrical boresight—The tracking axis as determined by an electrical indication, such as the null direction of a conical scanning or monopulse antenna system or the beam maximum direction of a highly directive antenna.

electrical bridging—The formation of a conductive path between conductors.

electrical center—1. The point approximately midway between the ends of an inductor or resistor. This point divides the inductor or resistor into two equal electrical values (e.g., voltage, resistance, inductance, or number of turns). 2. The center established by the electromagnetic field distribution within a test coil. A constant-intensity signal, irrespective of the circumferential position of a discontinuity, is indicative of electrical centering. The electrical center may be different from the physical center of the test coil.

electrical charge—The excess on (or in) a body of one kind (polarity) of electricity over the other kind. A plus sign indicates that positive electricity predominates, and a minus sign indicates that negative electricity predominates. Symbol: Q or q.

electrical conductivity—1. The reciprocal of the resistance in ohms measured between opposite faces of a centimeter cube of an aqueous solution at a specified temperature. 2. The property of a fluid or solid that permits the passage of an electrical current as a result of an impressed emf. It is measured by the quantity of electricity transferred across unit area per unit potential gradient per unit time. In sampling and analysis, changes in this property are utilized to measure the presence of certain ions and compounds such as sulfur dioxide.

electrical coupling—Coupling discrete elements with either electrical conductors or reactances.

electrical degree—One 360th part of a cycle of alternating current.

electrical discharge machining—Machining in which metal is removed by a controlled electrical spark in a dielectric.

electrical distance—The distance between two points, expressed as the length of time an electromagnetic wave in free space takes to travel between them.

electrical element—The concept in uncombined form of any of the individual building blocks from which electronic circuits are synthesized. Examples of basic electrical elements are insulation, conductance, resistance, capacitance, and inductance.

electrical erosion—The loss of contact material due to action of an electrical discharge.

electrical forming—The application of electric energy to a semiconductor device in order to permanently modify the electrical characteristics.

electrical gearing—A term used to describe the action of a system in which the output shaft rotates at a different speed from the input shaft, the ratio being established by electrical means.

electrical generator—A machine that converts mechanical energy into electrical energy.

electrical glass insulation—Insulating materials made from glass fibers of varying diameters, lengths, compositions, etc., including yarns, rovings, slivers, cords, and sheets or mats, bounded or treated only as necessary to their manufacture.

electrical ground—The zero voltage reference for the power supply in an electronic device, usually connected to the equipment chassis. May also be connected to the power mains or earth ground.

electrical-impedance cephalography—A method of evaluating blood circulation in the brain by measuring changes in the impedance between two surface electrodes attached to the head. This impedance decreases when the blood volume in the brain increases. The technique is also known as rheoencephalography.

electrical inertia—Inductance that opposes any change in current through an inductor.

electrical initiation—Any source of electrical power used to start a function or sequence.

electrical interlocks—Switches mounted on contactors or other devices and operated by rods or levers. These interlocks open or close depending on the open or closed position of the contact or device with which they are associated, and are used to govern succeeding operations of the same or allied devices.

electrical length—Length expressed in wavelengths, radians, or degrees. Distance in wavelengths × 2π = radians; distance in wavelengths × 360 = degrees.

electrical load—A device (e.g., a speaker) comprising resistive and/or reactive components into which an amplifier, generator, etc., delivers power.

electrically alterable read-only memory—Abbreviated EAROM. An electrically erasable programmable read-only memory built using metal-nitride-oxide-semiconductor (MNOS) technology. The term electrically alterable read-only memory (EAROM) is being replaced by electrically erasable programmable read-only memory (EEPROM or E^2PROM). However, the MNOS technology devices should always be referred to as EAROMs.

electrically connected—Joined through a conducting path or a capacitor, as distinguished from being joined merely through electromagnetic induction.

electrically erasable programmable read-only memory—Abbreviated EEPROM or E^2PROM.

electrically operated rheostat—A rheostat used to vary the resistance of a circuit in response to some means of electrical control.

electrically operated valve—A solenoid- or motor-operated valve used in vacuum, air, gas, oil, water, or similar lines.

electrically variable inductor—An inductor in which the inductance can be controlled by a current or a voltage. It is usually made in the form of a saturable reactor with two windings. One is called the signal, or tuned winding, corresponding to the ac or load winding of a power-handling saturable reactor; the other is the control winding and corresponds to the dc winding of the saturable reactor.

electrical noise—Unwanted electrical energy other than crosstalk in a transmission system.

electrical overstress—The operation of an electronic device beyond its normal range of voltage, current, and power when abnormal electric signals are presented to it.

electrical radian—57.296°, or $1/2\pi$ (or 1/6.28) of a cycle of alternating current or voltage.

electrical ratings—The combinations of voltage and current under which a device or component will operate satisfactorily in specified circuits under standard atmospheric conditions.

electrical reset—A term applied to a relay to indicate that it is capable of being electrically reset after an operation.

electrical resistivity—The resistance of a material to passage of an electric current through it. Expressed as ohms (units of resistance) per mil foot or as microhms (millionths of an ohm) per centimeter cubed (cm^3) at a specified temperature.

electrical resolver — Special type of synchro having a single winding on the stator and two windings whose axes are 90° apart on the rotor.

electrical scanning — Scanning accomplished through variation of the electrical phases or amplitudes at the primary radiating element of an antenna system.

electrical service entrance — A combination of intake wires and equipment including the service entrance wires, electric meter, main switch or circuit breaker, and main distribution or service panel through which the supply of power enters the home.

electrical sheet — Iron or steel sheets from which laminations for electric motors are punched.

electrical shielding — Copper screen, a wire braid, or any conducting material that surrounds a circuit or cable conductors to exclude electrostatic or radio-frequency noises.

electrical switch — A device that makes, breaks, or changes the connections in an electric circuit.

electrical system — The organized arrangement of all electrical and electromechanical components and devices in a way that will properly control the particular machine tool or industrial equipment.

electrical twinning — *See* twinning.

electrical zero — A standard synchro position at which electrical outputs have defined amplitudes and time phase.

electric arc — A discharge of energy through a gas.

electric bell — An audible signaling device consisting of one or more gongs and an electromagnetically actuated striking mechanism.

electric brazing — A brazing (alloying) process in which the heat is furnished by an electric current.

electric breakdown voltage — *See* dielectric breakdown voltage.

electric breeze or wind — The emission of electrons from a sharp point of a conductor that carries a high negative potential.

electric charge — 1. Electric energy stored on the surface of an object. 2. A property of electrons and protons. Similarly charged particles repel one another. Particles having opposite charges attract one another. 3. Electric energy that is stored as stress on the surface of a dielectric.

electric chronograph — A highly accurate apparatus for measuring and recording time intervals.

electric circuit — A continuous path consisting of wires and/or circuit elements over or through which an electric current can flow. If the path is broken at any point, current can no longer flow and there is no circuit.

electric coil — Successive turns of insulated wire that create a magnetic field when an electric current is passed through them. It may also consist of a number of separately insulated sections that lie side by side around the same magnetic circuit.

electric contact — A separable junction between two conductors that is designed to make, carry, or break (in any sequence or singly) an electric circuit.

electric controller — A device that governs the amount of electric power delivered to an apparatus.

electric current — Electricity in motion. In the atoms of metallic substances, there are a number of free electrons, or negatively charged particles that wander in the spaces between the atoms of the metal. The electron movement is normally without any definite direction and cannot be detected. The connection of an electric battery produces an electric field in the metal and causes the free electrons to move or drift in one direction; it is this electron drift that constitutes an electric current. Electrons, being of negative polarity, are attracted to the positive terminal of the battery, and so the actual direction of flow of electricity is from negative to positive, that is, opposite to the conventional direction usually adopted.

electric delay line — A delay line using properties of lumped or distributed capacitive and inductive elements. Can be used as a storage medium by recirculating the information-carrying signal.

electric dipole — Also called a doublet. A simple antenna comprising a pair of oppositely charged conductors capable of radiating an electromagnetic wave in response to the movement of an electric charge from one conductor to the other.

electric-discharge lamp — 1. A sealed glass enclosure containing a metallic vapor or an inert gas through which electricity is passed to produce a bright glow. 2. A lamp in which light (or radiant energy) is produced by the passage of an electric current through a vapor or a gas. Electric discharge lamps may be named after the filling gas or vapor that is responsible for the major portion of the radiation, e.g., mercury lamps, sodium lamps, neon lamps, and argon lamps.

electric displacement — *See* electric-flux density.

electric-displacement density — *See* electric-flux density.

electric eye — 1. The layman's term for a photoelectric cell. 2. The cathode-ray, tuning-indicator tube used in some radio receivers.

electric field — 1. The region about a charged body. Its intensity at any point is the force that would be exerted on a unit positive charge at that point. 2. A condition detectable in the vicinity of an electrically charged body such that forces act on other electric charges in proportion to their magnitudes. 3. Field of force that exists in the space around electrically charged particles. Lines of force are imagined to originate at the protons or positively charged particles and to terminate on electrons or negatively charged particles.

electric-field intensity — A measure of the force exerted at a point by a unit charge at that point.

electric-field strength — The magnitude of the electric field in an electromagnetic wave. Usually stated in volts per meter. *See also* dielectric strength.

electric-field vector — At a point in an electric field, the force on a stationary positive charge per unit charge. May be measured in either newtons per coulomb or volts per meter. This term is sometimes called the electric-field intensity, but such use of the word *intensity* is deprecated in favor of *field strength*, since *intensity* denotes power in optics and radiation.

electric-filament lamp — A glass bulb either evacuated or filled with an inert gas and having a resistance element electrically heated to, and maintained at, the temperature necessary to produce incandescence.

electric filter — 1. Device for rejecting or passing a specific band of signal frequencies. 2. *See* electric-wave filter.

electric-flux density — Also called electric-displacement density or electric displacement. At a point, the vector equal in magnitude to the maximum charge per unit area that would appear on one face of a thin metal plate introduced in the electric field at that point. The vector is normal to the plate from the negative to the positive face.

electric force — Electric field intensity measured in dynes.

electric furnace — A furnace in which electric arcs provide the source of heat.

electric generator — A machine that transforms mechanical power into electrical power.

electric governor-controlled series-wound motor — A series-wound motor having an electric speed governor connected in series with the motor circuit. The governor is usually built into the motor.

electric-heat soldering—Soldering by heating the joint with an electric current.

electric hygrometer—An instrument for indicating humidity by electric means. Its operation depends on the relationship between the electric conductance and moisture content of a film of hygroscopic material.

electric hysteresis—Internal friction in a dielectric field (e.g., the paper or mica dielectric of a capacitor in an ac circuit). The resultant heat generated can eventually break down the dielectric and cause the capacitor to fail.

electrician—A person engaged in designing, making, or repairing electric instruments or machinery. Also, one who sets up an electrical installation.

electric image—The electrical counterpart of an object; i.e., the fictitious distribution of the same amount of electricity that is actually distributed on a nearby object.

electricity—1. The property of certain particles to possess a force field that is neither gravitational nor nuclear. The type of force field associated with electrons is defined as negative and that associated with protons and positrons as positive. The fundamental unit is the charge of an electron: 1.60203×10^{-19} coulomb. Electricity can be further classified as static electricity or dynamic electricity. Static electricity in its strictest sense refers to charges at rest, as opposed to dynamic electricity, or charge in motion. Static electricity is sometimes used as a synonym for triboelectricity or frictional electricity. 2. A basic property of all matter, which consists of negative and positive charges (electrons and protons) that attract each other. 3. The potential energy of electrons at rest. 4. The kinetic energy of electrons in motion. 5. A manifestation of free electrons that can be generated by induction, friction, or chemical action. It is recognized by its magnetic, chemical, and radiant effects.

electric lamp—Any lamp whose emission of radiant energy is dependent on the passage of an electrical current through the emissive medium.

electric light—Light produced by an electric lamp.

electric lines of force—In an electric field, curves whose tangents at any point give the direction of the fields at that point.

electric meter—A device that measures and registers the amount of electricity consumed over a certain period of time.

electric mirror—*See* dynode.

electric moment—For two charges of equal magnitude but opposite polarities, a vector equal in magnitude to the product of the magnitude of either charge by the distance between the centers of the two charges. The direction of the vector is from the negative to the positive charge.

electric motor—A device that converts electrical energy into rotating mechanical energy.

electric network—A combination of any number of electric elements, having either lumped or distributed impedances, or both.

electric oscillations—The back-and-forth flow of electric charges whenever a circuit containing inductance and capacitance is electrically disturbed.

electric potential—A measure of the work required to bring a unit positive charge from an infinite distance or from one point to another (the difference of potential between two points).

electric precipitation—The collecting of dust or other fine particles floating in the air. This is done by inducing a charge in the particles, which are then attracted to highly oppositely charged collector plates.

electric probe—A rod inserted into an electric field during a test to detect dc, audio, or rf energy.

electric reset—A qualifying term indicating that the contacts of a relay must be reset electrically to their original positions following an operation.

electric robot—Programmable machine that is powered by servomotors or stepping motors.

electric service panel—The main cabinet where electricity is brought into a building, then distributed to branch circuits. The panel usually contains the main circuit breaker for shutting down the entire system, and circuit breakers or fuses for shutting down each independent circuit.

electric shield—A housing, usually aluminum or copper, placed around a circuit to provide a low-resistance path to ground for high-frequency radiations and thereby prevent interaction between circuits.

electric strain gage—A device that detects the change in shape of a structural member under load and causes a corresponding change in the flow of current through the device.

electric strength—The maximum electric charge a dielectric material can withstand without rupturing. *See also* dielectric strength; insulating strength. The value obtained for the electric strength will depend on the thickness of the material and on the method and conditions of test.

electric stroboscope—An instrument for observing or for measuring the speed of rotating or vibrating objects by electrically producing periodic changes in the intensity of light used to illuminate the object.

electric tachometer—A tachometer (rpm indicator) that utilizes voltage or electrical impulses.

electric telemeter—A system consisting of a meter that measures a quantity, a transmitter that sends the information to a distant station, and a receiver that indicates or records the quantity measured.

electric transcription—In broadcasting, a disc recording of a message or a complete program.

electric transducer—A device actuated by electric waves from one system and supplying power, also in the form of electric waves, to a second system.

electric tuning—A system by which a radio receiver is tuned to a station by pushing a button (instead of, say, turning a knob).

electric vector—1. A component of the electromagnetic field associated with electromagnetic radiation. The component is of the nature of an electric field. The electric vector is supposed to coexist with, but act at right angles to, the magnetic vector. 2. The electric field associated with an electromagnetic wave and thus with a light wave. The electric vector specifies the direction and amplitude of this electric field.

electric watch—A timepiece in which a battery replaces the mainspring as the prime energy source of the watch, and in which an electromagnet impels the balance wheel through a mechanical switching-contact arrangement.

electric wave—Another term for the electromagnetic wave produced by the back-and-forth movement of electric charges in a conductor.

electric-wave filter—Also called electric filter. 1. A device that separates electric waves of different frequencies. 2. A frequency-selective device, usually passive, made up of resistance, capacitance, and inductance elements having signal transmission characteristics that are a function of frequency. Basically, filters are used to pass desired signals and reject unwanted or interfering signals. Bandpass filters are used when the desired signals encompass a frequency band that does not contain any unwanted signals.

electrification — 1. The process of establishing an excess of positive or negative charges in a material. 2. The process of applying a voltage to a component or device.

electrification time — The time during which a steady direct potential is applied to electrical insulating materials or before the current is measured.

electroacoustic — Pertaining to a device (e.g., a speaker or a microphone) that involves both electric current and sound-frequency pressures.

electroacoustic device — One that employs phonon propagation or vibrations of a material's crystal lattice structure as the basic energy transport mechanism. Electrical energy is converted into acoustic energy by the material's piezoelectric properties.

electroacoustic transducer — A device that receives excitations from an electric system and delivers an output to an acoustic system, or vice versa. A speaker is an example of the first, and a microphone is an example of the second.

electroanalysis — The process of determining the quantity of an element or compound in an electrolyte solution by depositing the element or compound on an electrode by electrolysis.

electrobiology — The science concerned with electrical phenomena of living creatures.

electrobioscopy — The application of a voltage to produce muscular contractions.

electrocardiogram — Abbreviated EKG or ECG. 1. Essentially an electromyogram of the heart muscle. All muscular activity in the body is characterized by the discharge of polarized cells, the aggregate current from which causes a voltage drop that can be measured on the skin. A changing emf will appear between electrodes connected to the arms, legs, and chest, which rises and falls with heart action such that the period of the resulting waveform is the time between heartbeats. Various positive and negative peaks within one cycle of this waveform have been lettered P, Q, R, S, and T, a notation that aids in subsequent analysis and diagnosis. 2. A hardcopy record of heart action potentials obtained by measuring instantaneous potential differences at the surface of the body. In general, the recording describes the depolarization of myocardial muscle cell masses, providing a graphic, but indirect, view of the heart's competence. 3. Graphic tracing of the electric current that is produced by the rhythmic contraction of the heart muscle. Visually, a periodic wave pattern is produced. Changes in the wave pattern may appear in the course of various heart diseases; the tracing is obtained by applying electrodes on the skin of the chest and limbs.

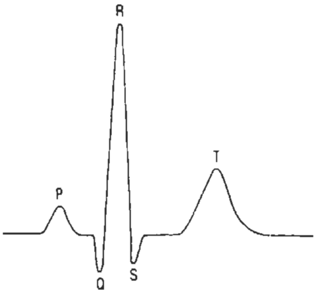

Electrocardiogram for one heartbeat.

electrocardiograph — A medical instrument for detecting irregularities in the action of a human heart. It measures the changes in voltage occurring in the human body with each heartbeat. Abbreviated EKG or ECG.

electrocardiography — Recording and interpretation of the electrical activity of the heart. The voltage generated by the heart is picked up by surface electrodes on the limbs and chest, amplified, and applied to a strip-chart recorder.

electrocardiophonograph — An instrument that records heart sounds.

electrochemical cell — An electrochemical system consisting of an anode and a cathode in metallic contact and immersed in an electrolyte. The anode and cathode may be different metals or dissimilar areas on the same metal surface.

electrochemical deterioration — A process in which autocatalytic electrochemical reactions produce an increase in conductivity and in turn ultimate thermal failure.

electrochemical device — A device that operates on both electrical and chemical principles, e.g., a lead-acid storage battery.

electrochemical diffused-collector transistor — *See* EDCT.

electrochemical equivalent — The weight of an element, compound, radical, or ion involved in a specified electrochemical reaction during passage of a specified quantity of electricity such as a coulomb.

electrochemical junction transistor — A junction transistor produced by etching an n-type germanium wafer on opposite sides with jets of a salt solution such as indium chloride.

electrochemical potential — Also called electrochemical tension. The partial derivative of the total electrochemical free energy of the system with respect to the number of moles of the constituent except that it includes the electrical as well as the chemical contributions to the free energy.

electrochemical recording — A recording made by passing a signal-controlled current through a sensitized sheet of paper. The paper reacts to the current and thereby produces a visual record.

electrochemical tension — *See* electrochemical potential.

electrochemical transducer — A device that uses a chemical change to measure the input parameter, and the output of which is a varying electrical signal proportional to the measurand.

electrochemical valve — Electric valve consisting of a metal in contact with a solution or compound, across the boundary of which current flows more readily in one direction than in the other direction and in which the valve action is accompanied by chemical changes.

electrochemistry — 1. That branch of science concerned with reciprocal transformations of chemical and electrical energy. This includes electrolysis, electroplating, the charge and discharge of batteries, etc. 2. The study of the reversible conversion of chemical energy into electrical energy. Electroplating is an electrochemical process.

electrochromic display — A passive solid-state display that is made from a material whose light-absorption properties are changed by an externally applied electric field. Ordinarily electrochromic materials do not absorb light in the visible range of the spectrum, so they are completely transparent. When a moderate electric field is applied, the material develops an absorption band in the visible spectrum and takes on a color that remains even after the electric field is removed and lasts from minutes to months. The color change can be reversed and the display returned to its original state when the polarity of the applied electric field is simply reversed.

electrocoagulation — The process of solidifying tissue by means of a high-frequency electrical current.

electrocution — Killing by means of an electric current.

electrode — 1. In an electronic tube, the conducting element that does one or more of the following: emits or collects electrons or ions, or controls their movement by means of an electric field on it. 2. In semiconductors, the element that does one or more of the following: emits or collects electrons or holes, or controls their movements by means of an electric field on it. 3. In electroplating, the metal being plated. 4. A conductor by means of which a current passes into or out of a fluid or an organic material, such as human skin; often one terminal of a lead. 5. A metallic conductor such as in an electrolytic cell, in which conduction by electrons is changed to conduction by ions or other charged particles. 6. A conductor, not necessarily metal, through which a current enters or leaves an electrolytic cell, arc, furnace, vacuum tube, gaseous discharge tube, or any conductor of the nonmetallic class. 7. That part of a semiconductor device providing the electrical contact between the specified region of the device and the lead to its terminal. 8. In a spark plug, the center rod passing through the insulator forms one electrode. The rod welded to the shell forms another. They are referred to as the center and side electrodes. 9. A conducting element at whose surface electricity passes into another conducting medium.

electrode admittance — The alternating component of the electrode current divided by that of the electrode voltage (all other electrode voltages maintained constant).

electrode capacitance — The capacitance between one electrode and all the other electrodes connected together.

electrode characteristic — The relationship, usually shown by a graph, between the electrode voltage and current, all other electrode voltages being maintained constant.

electrode conductance — The quotient of the in-phase component of the electrode alternating current divided by the electrode alternating voltage, all other electrode voltages being maintained constant. This is a variational and not a total conductance.

electrode current — Current passing into or out of an electrode.

electrode dark current — Also called dark current. 1. In phototubes, the component of electrode current that flows in the absence of ionizing radiation and optical photons. 2. The current that flows in a photodetector when there is no incident radiation on the detector.

electrode dissipation — The power that an electrode dissipates as heat when bombarded by electrons and/or ions and radiation from nearby electrodes.

electrode drop — The voltage drop produced in an electrode by its resistance.

electrode impedance — The reciprocal of electrode admittance.

electrode inverse current — Current through a tube electrode in the direction opposite to that for which the tube was designed.

electrodeless discharge — A luminous discharge produced by means of a high-frequency electric field in a gas-filled glass tube that has no internal electrodes.

electrodeless discharge tube — Abbreviated EDT. A device consisting of an airtight quartz tube that holds the material to be analyzed. When a high-frequency electrostatic field, generated by microwaves, is applied to the tube, it emits energy of a wavelength identical with that of the contained material.

electrodeposition — Also called electrolytic deposition. *See also* electroplating.

electrode potential — 1. The instantaneous voltage on an electrode. Its value is usually given with respect to the cathode of a vacuum tube. 2. The difference in potential between an electrode and the immediately adjacent electrolyte referred to some standard electrode potential as zero. 3. The potential in volts that an electrode has when immersed in an electrolyte, compared to the zero potential of a hydrogen electrode. The potential depends on the material of which the electrode is made.

electrode reactance — The imaginary component of electrode impedance.

electrode resistance — The reciprocal of electrode conductance. It is the effective parallel resistance, not the real component of electrode impedance.

electrodermography — The recording of the electrical resistance of the skin, which is a sensitive indicator of the activity of the autonomic nervous system.

electrode voltage — The voltage between an electrode and the cathode or a specified point of a filamentary cathode. The terms *grid voltage, anode voltage, plate voltage*, etc., designate the voltage between these electrodes and the cathode. Unless otherwise stated, electrode voltages are measured at the available terminals.

electrodialytic process — A process for producing fresh water by using a combination of electric current and two types of chemically treated membranes.

electrodynamic — Pertaining to electric current, electricity in motion, and the actions and effects of magnetism and induction.

electrodynamic braking — A method of stopping a tape-deck motor gently by the application of a predetermined voltage to the motors.

electrodynamic instrument — An instrument that depends for its operation on the reaction between the current in one or more moving coils and the current in one or more fixed coils.

electrodynamic machine — Electric generator or motor in which the output load current is produced by magnetomotive currents generated in a rotating armature.

electrodynamics — 1. The science dealing with the various phenomena of electricity in motion, including interactions of currents with each other, with their associated magnetic fields, and with other magnetic fields. 2. The study of the generation of electromagnetic power by radiation from high-energy beams.

electrodynamic speaker — A speaker consisting of an electromagnet called the field coil, through which a direct current flows.

electrodynamometer — 1. An instrument for detecting or measuring an electric current by determining the mechanical reactions between two parts of the same circuit. 2. A meter movement consisting of a rotatable (moving) wire coil suspended between two fixed (field) wire coils. The three coils can be connected in various configurations, so that rotation of the moving coil is proportional to applied ac or dc voltage or current, to power, power factor, etc.

electroencephalogram — 1. A waveform obtained by plotting brain voltages (available between two points on the scalp) against time. An electroencephalogram is not necessarily a periodic function, although it can beparticularly if the patient is unconscious. These voltages are of extremely low level and require recording apparatus that displays excellent noise rejection. 2. The tracing of brain waves made by an electroencephalograph.

electroencephalograph — Abbreviated EEG. An instrument for measuring and recording the rhythmically varying potentials produced by the brain by the use of electrodes applied to the scalp.

electroencephalography— 1. Recording and interpretation of the electrical activity of the brain. Voltage (typically 50 microvolts) picked up by electrodes on the scalp is amplified and applied to a strip-chart recorder. 2. Recording of electric currents developed in the brain by means of electrodes applied to the scalp, to the surface of the brain, or placed within the substance of the brain.

electroencephaloscope— An instrument for detecting brain potentials at many different sections of the brain and displaying them on a cathode-ray tube.

electroexplosive device—*See* EED.

electrofluid dynamics generator—Abbreviated EFD. A generator in which the only moving parts are wind-driven charged particles. Their movement from electrode to electrode is analogous to the spinning of an armature. Ideally suited for use at sea or at the seaside, bladeless windmills could be used anywhere there is a source of moisture. Principal advantages of the concept include efficiency and low cost.

electroforming—Also called electrodeposition and electroplating. 1. Making a metal object by using electrolysis to deposit a metal on an electrode. 2. Creating a pn junction by passing a current through point contacts on a semiconductor. 3. The production or reproduction of articles by electrodeposition on a mandrel or mold that is subsequently separated from the deposit. 4. The process of depositing a substance on an electrode by electrolysis, as in electroplating, electroforming, electrofining, or electrotinning.

electroforming process— An electrochemical process of metal fabrication using an electrolyte, an anode to supply the metal, and a control of the electrical current and of the deposition of metal on the matrix of a reflector.

electrogalvanizing—Electrodeposition of zinc coatings.

electrogastrogram—The graphic record that results from synchronous recording of the electrical and mechanical activity of the stomach.

electrograph— 1. A plot, graph, or tracing made by means of the action of an electric current on sensitized paper or other material, or by means of an electrically controlled stylus or pen. 2. Equipment for facsimile transmission.

electrographic process—*See* dielectric process.

electrographic recording—Also called electrostatography. The producing of a visible record by using a gaseous discharge between two or more electrodes to form electrostatically charged patterns on an insulator. *See also* electrostatic electrography.

electrokinetics—The branch of physics concerned with electricity in motion.

electroless deposition—The deposition of conductive material from an autocatalytic plating solution without application of electrical current.

electroless plating— 1. A method of metal deposition by means of a chemical reducing agent present in the processing solution. The process is further characterized by the catalytic nature of the surface, which enables the metal to be plated to any thickness. 2. A chemical process by which certain metals can be plated without electrical current. Tin may be plated onto copper in this manner. 3. The controlled autocatalytic reduction of a metal ion on certain catalytic surfaces.

electroluminescence— 1. Luminescence resulting from a high-frequency discharge through a gas or from application of an alternating current to a layer of phosphor. 2. Direct conversion of electrical energy into light energy in a liquid or solid; for example, photoemission as a result of electron-hole recombination in a pn junction. This is the mechanism employed by the injection laser. The standard abbreviation for the effect is written EL. (This process is not to be confused with the ordinary tungsten filament bulb, where there is an intermediate stage of heat, making the process thermoluminescent.) 3. Light produced in a phosphor that is in an alternating electric field. Consists of a phosphor a few mils thick placed between two metal films, one of which is transparent. Alternating current is applied to the plates through a current-limiting resistor.

electroluminescent display— 1. A display, designated EL, whose segments or elements consist of transparent conductive electrodes separated by a dielectric containing a luminescent phosphor. Application of ac voltage to opposing electrodes causes the dielectric between them to glow with a characteristic blue-green light. 2. The utilization of the light produced when electrical energy is directly converted into light within devices used for visual readout displays or as complex logic-circuit elements. The display devices may be flat, giving a wide viewing angle without parallax, and may have low power needs. Can yield blue, green, yellow, and white colors.

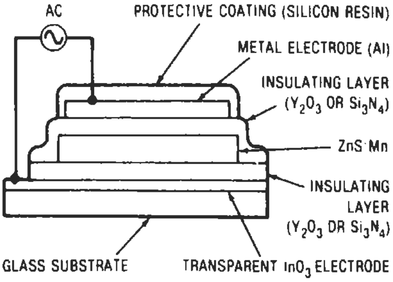
AC
PROTECTIVE COATING (SILICON RESIN)
METAL ELECTRODE (Al)
INSULATING LAYER (Y_2O_3 OR Si_3N_4)
ZnS·Mn
INSULATING LAYER (Y_2O_3 OR Si_3N_4)
GLASS SUBSTRATE
TRANSPARENT InO_3 ELECTRODE

Electroluminescent display.

electroluminescent lamp—A lamp in the shape of a panel that is decorative as well as illuminative. It consists primarily of a capacitor having a ceramic dielectric with electroluminescent phosphor. The amount of illumination is determined by the voltage across the layer and by the frequency applied to it.

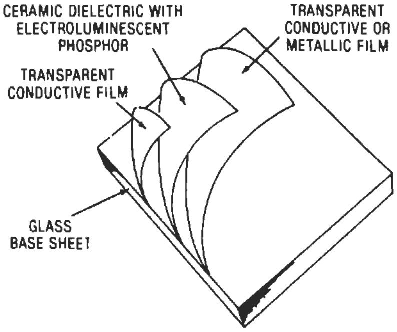
CERAMIC DIELECTRIC WITH ELECTROLUMINESCENT PHOSPHOR
TRANSPARENT CONDUCTIVE OR METALLIC FILM
TRANSPARENT CONDUCTIVE FILM
GLASS BASE SHEET

Electroluminescent lamp.

electroluminescent-photoconductive image intensifier—A panel, made up of electroluminescent and photoconductive (EL-PC) layers, used as either a positive or negative image intensifier, depending on amplitude and phase of its two power supply voltages. The photoconductive layer receives the image and converts it to electrical signals; subsequently, the electroluminescent layer converts the signals to visible, brighter light.

electrolysis — 1. The process of changing the chemical composition of a material (called the electrolyte) by sending an electric current through it. 2. The decay of an underground structure by chemical action due to stray electrical currents. 3. Production of chemical changes of the electrolyte by the passage of current through an electrochemical cell. 4. Conduction of an electric current through a chemical compound in its natural state, solution or as a molten, to decompose the compound.

electrolyte — 1. A substance in which the conduction of electricity is accompanied by chemical action. 2. The paste that forms the conducting medium between the electrodes of a dry cell, storage cell, or electrolytic capacitor. 3. A substance that, when dissolved in a suitable liquid (often water), dissociates into ions, thus rendering the liquid electrically conducting. 4. The current-conducting substance (liquid or solid) between two capacitor electrodes, at least one of which is covered by a dielectric film. 5. A conducting medium in which current is accompanied by movement of matter. Most often an aqueous solution of acids, bases, or salts, but includes many other media, such as fused salts, ionized gases, some solids, etc. 6. A substance that is capable of forming a conducting liquid medium when dissolved or melted.

electrolyte conductivity — Also called specific conductance. A measure of the ability of a solution to carry an electric current. Defined as the reciprocal of the resistance in ohms of a 1-cm cube of the liquid at a specified temperature. The units of specific conductance are the reciprocal ohm-cm (or siemens/cm) and one millionth of this. High-quality condensed steam and distilled or demineralized water have specific conductances at room temperatures as low as or lower than 1 microsiemens/cm.

electrolyte recording — A form of facsimile recording in which ionization causes a chemically moistened paper to undergo a change.

electrolytic — 1. Pertaining to or made by electrolysis; deposited by electrolysis; pertaining to or containing an electrolyte. 2. Said of an electrical device that contains an electrolyte.

electrolytic capacitor — 1. A capacitor consisting of two conducting electrodes, with the anode having a metal oxide film formed on it. The film acts as the dielectric or insulating medium. The capacitor is operable in the presence of an electrolyte, usually an acid or salt. Generally used for filtering, bypassing, coupling, or decoupling. 2. A capacitor in which the dielectric is a film of oxide electrolytically deposited on a plate or slug of aluminum or tantalum. The thinness of the film permits a high capacitance-to-volume ratio. The oxide acts as a dielectric in one direction only. The device is, therefore, polarized. (A nonpolarized electrolytic capacitor is, in effect, two polarized types in series with their like terminals connected together.) 3. A fixed capacitor, having a relatively high capacitance-to-volume ratio due to a very thin electrically formed, nonconducting chemical dielectric (oxide) film. 4. A capacitor in which the electrolytically formed oxide layer on the surface of the anode serves as a dielectric, with a solid or nonsolid electrolyte forming the cathode, thus giving the capacitor polar properties.

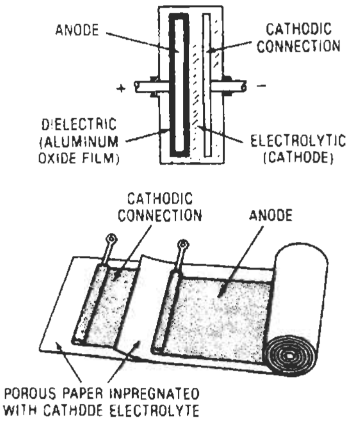
ANODE
CATHODIC CONNECTION
+
−
DIELECTRIC (ALUMINUM OXIDE FILM)
ELECTROLYTIC (CATHODE)
CATHODIC CONNECTION
ANODE
POROUS PAPER INPREGNATED WITH CATHODE ELECTROLYTE

Electrolytic capacitor.

electrolytic capacitor paper — Very pure, porous paper, 17 to 100 micrometers thick, used to separate the metallic electrodes in electrolytic capacitors.

electrolytic cell — 1. In a battery, the container, two electrodes, and the electrolyte. 2. A unit apparatus in which electrochemical reactions are produced by applying electrical energy, or which supplies electrical energy as a result of chemical reactions and which includes two or more electrodes and one or more electrolytes contained in a suitable vessel.

electrolytic cleaning — A process of removing soil, scale, or corrosion products from a metal surface by subjecting it as an electrode to an electric current in an electrolytic bath.

electrolytic conduction — The flow of current between electrodes immersed in an electrolyte. It is caused by the movement of ions from one electrode to the other when a voltage is applied between them.

electrolytic corrosion — Corrosion by means of electrochemical erosion. *See* corrosion.

electrolytic deposition — *See* electrodeposition.

electrolytic development — The method of developing a photographic image by means of an applied electric field. The systems used include electrolysis and photoconductive systems.

electrolytic dissociation — The breaking up of molecules into ions in a solution.

electrolytic interrupter — A device that is tilted to change the current through it.

electrolytic iron — Iron obtained by an electrolytic process. The iron possesses good magnetic qualities and is exceptionally free of impurities.

electrolytic plating — A method of metal deposition employing the work or cathode, the anode, the electrolyte, a solution containing dissolved salts of the metal to be plated, and a source of direct current. The anode metal is dissolved by chemical and electrical means; subsequently, cations are deposited onto the cathode. Electric current is the reducing agent. Copper, nickel, chromium, zinc, brass, cadmium, tin, gold, and silver are the metals most commonly electroplated.

electrolytic potential — The difference in potential between an electrode and the immediately adjacent electrolyte, expressed in terms of some standard electrode difference.

electrolytic rectifier — A rectifier consisting of metal electrodes in an electrolyte, in which rectification of alternating current is accompanied by electrolytic action. A polarization film formed on one of the electrodes permits current in one direction but not in the other.

electrolytic refining — The refining or purifying of metals by electrolysis.

electrolytic shutter — A high-speed shutter, similar to a Kerr cell, that uses the birefringence produced in a liquid during the passage of an electric current through it to change the liquid's optical transmission characteristics.

electrolytic switch—A switch having two electrodes projecting into a chamber containing a precisely measured quantity of a conductive electrolyte, leaving an air bubble of predetermined width. When the switch is tilted from true horizontal, the bubble shifts position and changes the amount of electrolyte in contact with the electrodes, thereby changing the amount of current passed by the switch. Used as a leveling switch in gyro systems.

electrolyzer—An electrolytic cell that produces alkalies, metals, chlorine, or other allied products.

electromagnet—1. A temporary magnet consisting of a solenoid with an iron core. A magnetic field exists only while current flows through the solenoid. 2. A magnet, consisting of a solenoid with an iron core, that has a magnetic field existing only during the time of current flow through the coil. 3. A coil of wire, usually wound on an iron core, that produces a strong magnetic field when current is sent through the coil. 4. A magnet created by inserting a suitable metal core within or near a magnetizing field that is usually formed by passing electric current through a coil of insulated wire. 5. A soft iron core that becomes a magnet temporarily when current flows through a coil of wire that surrounds it.

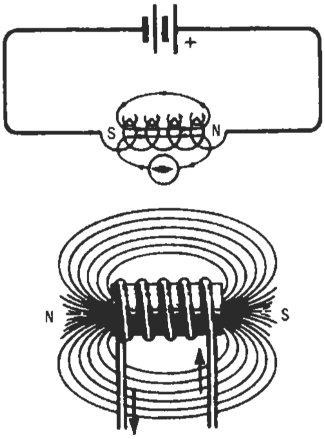

Electromagnet.

electromagnetic—1. Having both magnetic and electric properties. 2. Pertaining to the mutually perpendicular electric and magnetic fields associated with the movement of electrons through conductors, as in an electromagnet. 3. Pertaining to the combined electric and magnetic fields associated with radiation or with movements of charged particles. 4. Pertaining to or caused by the combined electric and magnetic fields that are always associated with an electric current. 5. Pertaining to the relationship between currents and magnetic fields.

electromagnetic amplifying lens—A system made up of a large number of waveguides symmetrically arranged with respect to an excitation medium so that they are excited with equal amplitude and phase in order to provide an effective gain in energy.

electromagnetic bonding—Method for joining thermoplastics in which a metallic preform is placed in the joint area to convert electromagnetic energy into heat for fusion bonding.

electromagnetic cathode-ray tube—A cathode-ray tube that uses electromagnetic deflection to deflect the electron beam.

electromagnetic communications—The electromagnetic wave conductor is space itself. The electromagnetic frequencies available today for communications fall into two categories: frequencies that form "wireless" communications (such as visual light of fairly high frequency), and frequencies that humans use for wireless communications (such as radio, shortwave, and microwave transmitting, of relatively lower frequencies). In communicating by radio, shortwave, and microwave frequencies, translators similar in principle to those used in electrical communications are needed, although the equipment requirement increases.

electromagnetic compatibility—Abbreviated EMC. 1. The ability of electronic devices and communications equipment, subsystems, and systems to operate in their intended environments without suffering or causing unacceptable degradation of performance as a result of unintentional electromagnetic radiation or response. 2. A directive that specifies the acceptable limits for electromagnetic emissions from an electronic device, and how much electromagnetic interference the device should tolerate. 3. Abbreviation for electronic message center (Ford Motor). Digital dashboard electronics displaying digital readouts of speed, miles, fuel, clock, etc.

electromagnetic complex—The electromagnetic configuration of an installation, including all radiators of significant amounts of energy.

electromagnetic coupling—The mutual relationship between two separate but adjacent wires when the magnetic field of one induces a voltage in the other.

electromagnetic crack detector—An instrument for detecting hidden cracks in iron or steel objects by magnetic means.

electromagnetic deflection—The deflection of an electron stream by means of a magnetic field. In a television receiver, the magnetic field for deflecting the electron beam horizontally and vertically is produced by two pairs of coils, called the deflection yoke, around the neck of the picture tube.

electromagnetic deflection coil—A coil around the neck of a CRT, for deflecting the electron beam.

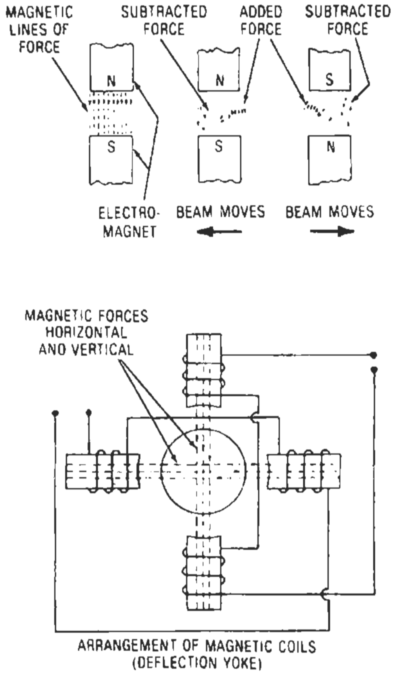

Electromagnetic deflection coil.

electromagnetic delay line—A delay line whose operation is based on the time of propagation of electromagnetic waves through distributed or lumped capacitance and inductance.

electromagnetic energy—Forms of radiant energy, such as radio waves, heat waves, light waves, X-rays, gamma rays, and cosmic rays.

electromagnetic environment—The rf field or fields existing in an area or desired in an area to be shielded.

electromagnetic field—1. The field of influence produced around a conductor by the current flowing through it. 2. A rapidly moving electric field and its associated magnetic field. The latter is perpendicular to both the electric lines of force and their direction. 3. The field associated with radio or light waves, consisting of a magnetic and an electric field at right angles to each other and to the direction of wave propagation.

electromagnetic focusing—In a television picture tube, the focusing produced by a coil mounted on the neck. Direct current through the coil produces magnetic field lines parallel to the tube axis.

electromagnetic horn—A horn-shaped structure that provides highly directional radiation of radio waves in the 100-megahertz or higher frequency range.

electromagnetic induction—The voltage produced in a coil as the number of magnetic lines of force (flux linkages) passing through the coil changes.

electromagnetic inertia—1. The characteristic delay of a current in an electric circuit in reaching its maximum or zero value after application or removal of the source voltage. 2. The property of self-induction.

electromagnetic interference—Abbreviated EMI. 1. Unintentional interfering signals generated within or external to electronic equipment. Typical sources could be power-line transients, noise from switching-type power supplies, and/or spurious radiation from oscillators. EMI is suppressed with power-line filtering, shielding, etc. EMI suppression requirements are frequently specified for military equipment. 2. Electromagnetic phenomena which, either directly or indirectly, can contribute to a degradation in performance of an electronic receiver or system. (The terms *radio interference, radio-frequency interference, noise, emi,* and *rfi* have been employed at various times in the same context.) 3. Disturbances caused by electromagnetic waves (radio, heat, light, etc.) that can impair the reception of the desired transmitted signal. 4. Unwanted electromagnetic emissions, generated by lightning or by electronic or electrical devices, that degrade the performance of another electronic device. Interference may be reduced by shielding. Maximum acceptable levels of EMI from electronic devices are detailed by the Federal Communications Commission.

electromagnetic lens—1. An electron lens in which the electron beams are focused electromagnetically. 2. An electromagnet that produces a suitably shaped magnetic field for the focusing and deflection of charged particles in electron-optical systems. 3. An electron lens consisting of a homogeneous axial electric field and a magnetic field, used in high-quality image tubes for high MTF and small geometrical distortion requirements.

electromagnetic mirror—A surface or region capable of reflecting radio waves, such as one of the ionized layers in the upper atmosphere.

electromagnetic oscillograph—An oscillograph in which a mechanical motion is derived from electromagnetic forces to produce a record.

electromagnetic pollution—The effects of electromagnetic interference (EMI) produced by human-made apparatus. The seriousness of this interference ranges from annoying interference that affects a radio or television channel to interference that causes failure of an important communication channel or a cardiac pacemaker.

electromagnetic pulse—Abbreviated EMP. A reaction of large magnitude resulting from the detonation of nuclear weapons.

electromagnetic radiation—1. That form of energy which is characterized by transversely oscillating electric and magnetic fields and which propagates at velocity *c* in free space. At a sufficient distance from the source, the electric-field vector and the magnetic field vector are at right angles to each other, forming a right-handed (coordinate) system. In an ionized medium, a longitudinal component may be present. 2. A form of power emitted from vibrating charged particles. A combination of oscillating electric and magnetic fields, electromagnetic radiation propagates through otherwise empty space with the velocity of light. This (constant) velocity equals the alternation frequency multiplied by the wavelength; hence the frequency and wavelength are inversely proportional to each other. The spectrum of electromagnetic radiation is continuous over all frequencies. 3. Abbreviated EMR. When discussing shielding, describes radiation generated by electrical means, ranging from a stationary magnetic or electrostatic field to high-frequency changing fields and transmitted plane waves of radio frequency.

electromagnetic reconnaissance—Activity conducted to locate and identify potential hostile sources of electromagnetic radiation, including radar, communication, missile-guidance, and air-navigation equipment.

electromagnetic relay—1. Device that opens or closes contacts by setting "moving" contacts against "fixed" contacts when current passes through an electromagnet. Current sets up a magnetic attraction between the core of the electromagnet and a hinged arm to the tip of which is attached the moving contact. The movement of the arm toward the core of the electromagnet brings moving and fixed contacts together. When current is withdrawn, a spring returns the arm to its original position and the contacts separate. 2. A mechanical switch operated by electric power.

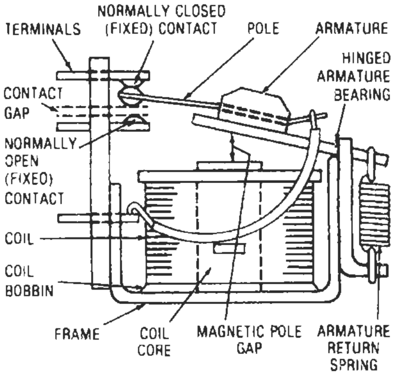

Electromagnetic relay.

electromagnetic repulsion—The repelling action between like poles of electromagnets.

electromagnetics—In physics, the branch concerned with the relationships between electric currents and their associated magnetic fields.

electromagnetic spectrum—1. A chart or graph showing the relationships among all known types of electromagnetic radiation classified by wavelengths. 2. The continuous range of frequencies, from 0.1 to 10^{22} hertz, of which a radiated signal is composed. Spectral dimensions are more conveniently described in terms of wavelength (angstroms), where 1 angstrom is equivalent to 10^{-7} mm. The electromagnetic spectrum includes radio-frequency waves, light waves, microwaves, infrared, X-rays (Roentgen rays), and gamma rays. 3. The ordered array of known electromagnetic radiations, extending from the shortest wavelengths, cosmic rays, through gamma rays, X-rays, ultraviolet radiation, visible radiation, and infrared, and including microwave and all other wavelengths of radio energy. 4. The entire range of wavelengths, extending from the shortest to the longest or conversely, that can be generated physically. This range of electromagnetic wavelengths extends almost from zero to infinity and includes the visible portion of the spectrum known as light. *See also* visible spectrum. 5. The total range of wavelengths or frequencies of electromagnetic radiation, extending from the longest radio waves to the shortest known cosmic rays.

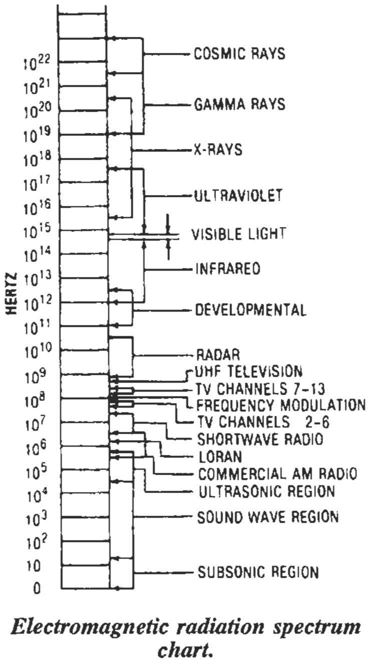

Electromagnetic radiation spectrum chart.

electromagnetic tester—A nondestructive test method for engineering materials, including magnetic materials, that uses electromagnetic energy having frequencies less than those of visible light to yield information regarding the quality of the tested materials.

electromagnetic theory—The theory of propagation of energy by combined electric and magnetic fields included in Maxwell's equations.

electromagnetic theory of light—The theory that states that electromagnetic and light waves have identical properties.

electromagnetic transduction—1. Conversion of the measurand into the output induced in a conductor by a change in magnetic flux. 2. A wave produced by the oscillation of an electric charge. 3. A wave in which there are both electric and magnetic displacements. 4. A transverse wave associated with the transmission of electromagnetic energy.

electromagnetic-type microphones—Microphones in which the voltages are varied by an electromagnet (namely, ribbon or velocity, dynamic or moving-coil, and reluctance or moving-vane microphones).

electromagnetic unit—Abbreviated emu. A unit of electricity based primarily on the magnetic effect of an electric current. The fundamental centimeter-gram-second unit is the abampere. Now considered obsolete.

electromagnetic vibrator—A mechanical device for interrupting the flow of direct current and thereby making it a pulsating current. This is done where a circuit requires an alternating current to operate. A reed within the vibrator is alternately attracted to two electromagnets.

electromagnetic wave—1. The radiant energy produced by oscillation of an electric charge. It includes radio, infrared, visible and ultraviolet light waves, and X-, gamma, and cosmic rays. 2. A wave in which both electric and magnetic displacement are present. 3. Waves of radiation identified by individual fluctuations of electric and magnetic fields. All such waves propagate at the speed of light in free space, which includes most realistic atmospheric conditions. Three material parameters are necessary and sufficient to describe electromagnetic waves in a given medium: dielectric constant (or permittivity), permeability, and conductivity.

electromagnetism—1. The magnetic field around a wire or other conductor when, and only when, current passes through it. 2. Magnetism caused by an electric current in a conductor.

electromanometer—Instrument used for measuring pressure of gases or liquids by electronic methods.

electromechanical—Any device using electrical energy to produce mechanical movement.

electromechanical bell—A bell with a prewound spring-driven clapper that is tripped electrically to ring the bell.

electromechanical breakdown—A mechanical runaway that occurs when the mechanical restoring force fails to balance the electrical compressive force.

electromechanical chopper—*See* contact modulator.

electromechanical diffused-collector transistor—*See* EDCT.

electromechanical energy—Energy present in an induction coil or solenoid.

electromechanical frequency meter—A meter that uses the resonant properties of mechanical devices to indicate frequency.

electromechanical recorder—A device that transforms electrical signals into equivalent mechanical motion that is transferred to a medium by cutting, embossing, or writing.

electromechanical timer—Usually refers to a motor-driven timer, with or without an electrically operated clutch. Can also apply to pneumatic and thermal timers, or slow pull-in or drop-out relays.

electromechanical transducer—A device that transforms electrical energy into mechanical energy and vice versa. A speaker is an example of the first, and a microphone of the second.

electromechanics—That branch of electrical engineering concerned with machines producing or operated by electric currents.

electrometallurgy—That branch of science concerned with the application of electrochemistry to the extraction or treatment of metals.

electrometer—1. An electrostatic instrument that measures a potential difference or an electric charge by the mechanical force exerted between electrically charged surfaces. 2. A dc voltmeter with an extremely high input resistance, usually around 10^{10} megohms, as opposed to 10 megohms or less for a conventional type.

electrometer amplifier—An amplifier circuit having sufficiently low current drift and other noise components, sufficiently low amplifier input-current offsets, and adequate power and current sensitivities to be usable for measuring current variations of considerably less than 10^{-12} A.

electrometer tube—A vacuum tube having a very low control-electrode conductance, to facilitate the measurement of extremely small direct currents and voltages.

electromigration—Motion of ions of a metal conductor (such as aluminum) in response to the passage of high current through it. Such motion can lead to the formation of "voids" in the conductor, which can grow to a size such that the conductor is unable to pass current. Electromigration is aggravated at high temperature and high current density and therefore is a reliability "wear-out" process. Electromigration is minimized by limiting current densities and by adding metal impurities such as copper or titanium to aluminum.

electromotive force—Abbreviated emf. 1. The force that causes electricity to flow when there is a difference of potential between two points. The unit of measurement is the volt. 2. Electrical pressure at the source. Not to be confused with potential difference, which is the voltage developed across a resistance or impedance due to current flowing through it. Both are measured in volts. 3. Electric pressure that causes a current to flow in a circuit; it is the energy put into the circuit by the source per unit electric charge that it supplies to the circuit. The unit of emf is the volt, being the electromotive force required to cause a current of 1 ampere to flow in a resistance of 1 ohm. 4. The difference of electrical potential found across the terminals of a source of electrical energy; more precisely, the limit of the potential difference across the terminals of a source as the current between the terminals approaches zero.

electromotive force series—A list of elements arranged according to their standard electrode potentials, with noble metals, such as gold, being positive and active metals, such as zinc, being negative.

electromotive series—A list of metals arranged in decreasing order of their tendency to pass into ionic form by losing electrons.

electromyogram—A waveform of the contraction of a muscle as a result of electrical stimulation. Usually the stimulation comes from the nervous system (normal muscular activity). The record of potential difference between two points on the surface of the skin resulting from the activity or action potential of a muscle.

electromyograph—An instrument for measuring and recording potentials generated by muscles.

electromyography—Abbreviated EMG. Recording and interpretation of the electrical activity of muscle tissue. A single electrical spike potential is generated when a muscle fiber contracts. The magnitude of the spike potentials is roughly proportional to the amount of muscular tension. Surface detecting electrodes (for many muscle fibers) or needle electrodes (for one or a few fibers) provide a signal that is amplified and displayed on a cathode-ray tube.

electron—An elementary atomic particle that carries the smallest negative electric charge (1.6×10^{-19} coulombs). Electrons are light in mass (1/1840 of the mass of the hydrogen atom, or 9.107×10^{-28} gram), highly mobile, and orbit the nucleus of an atom. Electrons are responsible for the bonds between atoms. Positive electrons, or positrons, also exist.

electronarcosis—1. The induction of unconsciousness by passage of a weak current through the brain. 2. Anesthesia induced by the passage of a precisely controlled electric current through the brain.

electron attachment—Process by which an electron is attached to a neutral molecule to form a negative ion. Often characterized by the attachment coefficient η, which is the number of attachments per centimeter of drift. Also characterized by the ratio $h = \sigma/\theta$, where σ is the attachment cross section and θ the total cross section.

electron avalanche—The chain reaction started when one free electron collides with one or more orbiting electrons and frees them. The free electrons then free others in the same manner, and so on.

electron band—A spectrum band composed of molecules that is usually found in the visible or the ultraviolet because of the electron transition taking place with the molecule.

electron beam—1. A narrow stream of electrons moving in the same direction under the influence of an electric or magnetic field. 2. A stream of electrons, emitted by a single source, that move in the same direction and at the same speed. 3. The electrons emitted by the cathode in a picture tube and focused into a beam that is deflected line by line across the phosphor screen to produce an image.

electron-beam bonding—Process using a stream of electrons to heat and bond two conductors within a vacuum.

electron-beam evaporation—An evaporation technique in which the evaporant is heated by electron bombardment.

electron-beam generator—A velocity-modulated generator, such as a klystron tube, used to generate extremely high frequencies.

electron-beam gun—A device generally used in a cathode-ray or camera tube to emit a stream of electrons moving at uniform velocity in a straight line. It consists of an emitting cathode and an anode, with an aperture for passage of some of the electrons.

electron-beam instrument—Also called a cathode-ray instrument. An instrument in which a beam of electrons is deflected by an electric or magnetic field (or both). Usually the beam is made to strike a fluorescent screen so the deflection can be observed.

electron-beam machining—A process in which controlled electron beams are used to weld or shape a piece of material.

electron-beam magnetometer—An instrument that measures the intensity and direction of magnetic forces by the immersion of an electron beam into the magnetic field.

electron-beam mode discharge—Abbreviated ebmd. A form of discharge produced by a perforated-wall hollow cathode operating under conditions of pressure, voltage, and geometry usually associated with the abnormal glow discharge.

electron-beam recording—The recording of the information contained in a modulated electron beam onto photographic or silicon-resin-coated materials.

electron-beam tube—An electron tube that depends for its operation on the formation and control of one or more electron beams.

electron-beam welding—1. The process of using a focused beam of electrons to heat materials to the fusion point. 2. Process in which a welder generates a stream of electrons traveling at up to 60 percent of the speed of light. It focuses the beam to a small, precisely controlled spot in a vacuum and converts the kinetic energy into an extremely high temperature on impact with the work piece.

electron-bombarded semiconductor amplifier—Abbreviated EBS amplifier. An amplifier consisting of an electron-gun modulation system, semiconductor target, and output coupling network all within a glass or ceramic envelope. The semiconductor target is a pair of silicon diodes, each consisting of two metallic electrodes with a pn junction under the top contact. Amplifier operation is based on the fact that a modulated electron beam can control the current in a reverse-biased semiconductor junction.

electron-bombardment-induced conductivity—In a multimode display storage tube, a process by which the image on the surface of the cathode-ray tube is erased by the use of an electron gun.

electron charge—Also called elementary charge. The charge of a single electron. Its value is 1.602189×10^{-19} coulomb. The fundamental unit of electrical charge.

electron-coupled oscillator—Abbreviated ECO. A circuit using a multigrid tube in which the cathode and two grids operate as a conventional oscillator and the electron stream couples the plate-circuit load to the oscillator.

electron coupling—In vacuum (principally multigrid) tubes, the transfer of energy between electrodes as electrons leave one and go to the other.

electron device—Any device in which the passage of electrons through a vacuum, gas, or semiconductor is the device's principal means of conduction.

electron diffraction—1. The phenomenon or the technique of producing diffraction patterns through the incidence of electrons on matter. 2. The bending of an electron stream that occurs when the stream travels through a medium such as very thin metal foil.

electron-diffraction camera—A special evacuated camera equipped with means for holding a specimen and bombarding it with a sharply focused beam of electrons. A cylindrical film placed around the specimen records the electrons that may be scattered or diffracted by it.

electron drift—The movement of electrons in a definite direction through a conductor, as opposed to the haphazard transfer of energy from one electron to another by collision.

electronegative—Having an electric polarity that is negative.

electronegative developer—A developer containing negatively charged toner particles.

electron emission—The freeing of electrons into space from the surface of a body under the influence of heat, light, impact, chemical disintegration, or a potential difference.

electron emitter—In a cathode tube, the electrode that serves as a source for electrons.

electron filter lens—An electrostatic device that uses an electric potential barrier to allow the transmittance of electrons at or above a set level of energy while stopping the passage of those below it.

electron flow—The movement of electrons from a negative to a positive point in a metal or other conductor, or from a negative to a positive electrode through a liquid, gas, or vacuum.

electron gun—1. An electrode structure that produces and may control focus and may deflect and converge one or more electron beams. 2. A device for producing and accelerating a beam of electrons. 3. The portion of a TV picture tube or cathode-ray tube that produces the stream of electrons and may also focus and center the stream. 4. The source of the electron beam in a picture tube, comprising a cathode plus several focusing electrodes that collimate and focus the electron beam into a spot on the screen. In a color tube there may be three electron guns usually integrated into a single unit (unitized gun), or a single gun for the three colors.

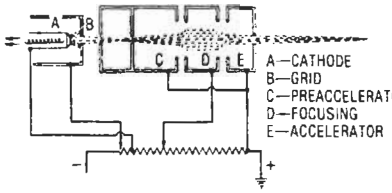

Electron gun.

electronic—1. Pertaining to that branch of science which deals with the motion, emission, and behavior of currents of free electrons, especially in vacuum, gas, or phototubes and special conductors or semiconductors. This is contrasted with electric, which pertains to the flow of large currents in metal conductors. 2. Of or pertaining to devices, circuits, or systems using the principle of electron flow through a conductor. Examples: electronic control, electronic equipment, electronic instrument, electronic circuit.

electronic autopilot—An arrangement of gyroscopes, electronic amplifiers, and servomotors for detecting deviations in the flight of an aircraft and applying the required corrections directly to its control cables.

electronic balance—Weighing balance that uses forces produced by known currents to balance unknown currents, and thereby unknown weights, very accurately to within parts of a microgram.

electronic "bug"—A keying system that converts the Morse signals from a hand key into correctly proportioned and spaced dots and dashes.

electronic bulletin board—A shared file where users can enter information for other users to read or download. Many bulletin boards are set up according to general topics and are accessible throughout a network.

electronic calculator—Electronic device for arithmetic and logarithmic computations; may also include digital printer and computer.

electronic camouflage—Use of electronic means or exploitation of electronic characteristics to reduce, submerge, or eliminate the radar-echoing properties of a target.

electronic carburetor—A fuel-metering actuator in which the air/fuel ratio is controlled by continual variations of the metering rod position in response to an electronic control signal.

electronic charge—The quantity of charge represented or possessed by one electron. It is equal to 1.602189×10^{-19} coulomb.

electronic circuit—A circuit containing one or more electron tubes, transistors, integrated circuits, magnetic amplifiers, etc.

electronic commutator—A type of switch that provides a continuous switching or sampling of a number of circuits by means of a radial-beam electronic tube or electronic switching circuit.

electronic confusion area—Amount of space that a target appears to occupy in a radar resolution cell, as it appears to that radar beam.

electronic control—Also called electronic regulation. The control of a machine or condition by electronic devices.

electronic control unit—*See* ECU.

electronic counter—An instrument capable of counting up to several million electrical pulses per second.

electronic counter-countermeasures—Abbreviated ECCM. 1. Equipment and techniques that allow electronic systems such as radar and communication systems to operate effectively while attempts are made to disrupt or jam their operation. 2. That division of electronic warfare involving actions taken to ensure friendly and effective use of the electromagnetic spectrum despite the enemy's use of electronic warfare. 3. Retaliatory tactics used to reduce the effectiveness of electronic countermeasures.

electronic countermeasures—Abbreviated ECM. 1. All measures taken to reduce the effectiveness of enemy electronic systems such as radar and communications. There are two distinct areas: passive measures, or reconnaissance, and active measures, such as jamming. 2. That division of electronic warfare involving actions taken to prevent or reduce an enemy's effective use of the electromagnetic spectrum. Includes techniques such as chaff and barrage jamming as well as sophisticated methods to deceive the systems without indication to the opposing operators that their systems are being affected. 3. Methods of jamming or otherwise hindering the operation of enemy electronic equipment.

electronic countermeasures control—1. Collection and sorting of large quantities of data for the purpose of measuring and defining radar signals. 2. Examination of the data received in order to determine selection and switching of countermeasure devices with little or no time delay.

electronic coupling—The method of coupling electrical energy from one circuit to another through the electron stream in a vacuum tube.

electronic crowbar—An electronic switching device generally used in a power supply to divert a fault current from more delicate components until a fuse, circuit breaker, or the like has time to respond.

electronic data exchange—*See* EDI.

electronic data processing—Abbreviated EDP. 1. Operations on data carried out mainly by electronic equipment. 2. Use of electronic memories to store, update, and read information automatically, and using that information in accounting, filing, etc. 3. Any computerized information system and the equipment used in that system.

electronic data-processing center—Abbreviated EDP center. A place in which is kept automatically operated equipment, including computers, designed to simplify the interpretation and use of data gathered by instrumentation installations or information-collection agencies.

electronic data-processing machine—Abbreviated EDPM. A machine or its device and attachments used primarily in or with an electronic data-processing system.

electronic data-processing system—Any machine or group of automatically intercommunicating machines capable of entering, receiving, sorting, classifying, computing, and/or recording alphabetical or numerical accounting or statistical data (or all three).

electronic deception—Deliberate radiation, reradiation, alteration, absorption, or reflection of electromagnetic radiations in a manner intended to cause the enemy to obtain misleading data or false indications from his electronic equipment. There are two categories of electronic deception: (*a*) Manipulative deception—the alteration or simulation of friendly electromagnetic radiation to accomplish deception. (*b*) Imitative deception—the introduction into enemy channels of radiations that imitate the enemy's own emissions.

electronic device—1. A device in which conduction is principally by the movement of electrons through a vacuum, gas, or semiconductor. 2. An electronic tube or valve, transistor, or other semiconductor device. This definition excludes inductors, capacitors, resistors, and similar components.

electronic differential—An input or output type of circuit that only amplifies or responds to the difference of two signals, and does not respond to the signal with respect to ground or a supply voltage.

electronic differential analyzer—A form of analog computer using interconnected electronic integrators to solve differential equations.

electronic digital computer—A machine that uses electronic circuitry in the main computing element to perform arithmetic and logical operations on digital data (i.e., data represented by numbers or alphabetic symbols) automatically by means of an internally stored program of machine instructions. Such devices are distinguished from calculators, on which the sequence of instructions is externally stored and is impressed manually (desk calculators) or from tape or cards (card-programmed calculators).

electronic efficiency—The ratio of (*a*) the power at the desired frequency delivered by the electron stream to the oscillator or amplifier circuit to (*b*) the direct power supplied to the stream.

electronic engine control—*See* EEC.

electronic engineering—A branch of electrical engineering that applies the principles of electronics to the solution of practical problems. *See also* electronics.

electronic flash—Also called strobe. 1. The firing of special light-producing, high-voltage, gas-filled glass tubes with a high instantaneous surge of current furnished by a capacitor or bank of capacitors that have been charged from a high-voltage source (usually 450 volts or higher). 2. A device that upon command produces a pulse of luminous energy caused by a discharge of electrical energy through a gas. The term usually implies the use of a flash tube and associated power source and trigger circuit.

electronic flash tube—*See* flash tube.

electronic flash unit—A small xenon-filled tube with metal electrodes fused into the ends. The gas flashes brilliantly when a capacitor is discharged through the tube. The duration of the flash primarily depends on the capacitance of the capacitor. As a rule of thumb, flash time is approximately equal in microseconds to capacitance in microfarads. Sometimes the flash recurs at a specified frequency, which may reach many thousands per second; such a device is called a strobe unit because it

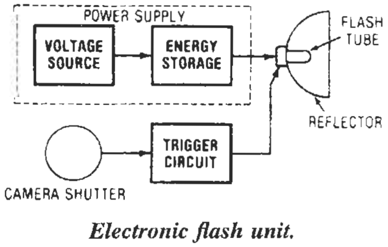

Electronic flash unit.

produces stroboscopic effects and makes rapidly moving parts appear to stand still.

electronic frequency synthesizer—A device that generates two or more selectable frequencies from one or more fixed-frequency sources.

electronic funds transfer system—*See* EFTS.

electronic gate—A device in which diodes and/or transistors provide input-output relations that correspond to a Boolean-algebra function (AND, OR, etc.).

electronic heating—Also called high-frequency heating. A method of heating a material by inducing a high-frequency current in it or having the material act as the dielectric between two plates charged with a high-frequency current.

electronic hookup wires—Wires used to make the internal connections between the various electrical parts of electronic assemblies.

electronic industries—Industrial organizations engaged in the manufacture, design, development, and/or substantial assembly of electronic equipment, systems, assemblies, or the components thereof.

Electronic Industries Association—Abbreviated EIA. A trade association of the electronics industry. Some of its functions are the formulation of technical standards, dissemination of marketing data, and the maintenance of contact with government agencies in matters relating to the electronics industry. The association was originally known as the Radio Manufacturers Association (RMA), and later as the Radio-Electronics-Television Manufacturers Association (RETMA).

electronic instrument—Any instrument that depends for its operation on the action of either one or more electron devices.

electronic instrument cluster—*See* EIC.

electronic intelligence—The technical and intelligence information derived from foreign noncommunications electromagnetic radiations emanating from other than nuclear detonations or radioactive sources.

electronic interference—Electrical or electromagnetic disturbances that result in undesired response in electronic equipment.

electronic jamming—Intentional radiation, reradiation, or reflection of electromagnetic energy for the purpose of reducing the effectiveness of enemy electromagnetic devices or impairing the use of any electronic devices, equipment, or systems being used by an enemy.

electronic keyboard—A keyboard that is used to generate characters through electronic means rather than through mechanical linkages.

electronic keying—A method of keying whereby the dots and dashes are produced solely by electronic means.

electronic line scanning—Facsimile scanning in which a spot on a cathode-ray tube moves across the copy electronically while the record sheet or subject copy is moved mechanically in a perpendicular direction.

electronic mail—Abbreviated e-mail or email. 1. Electronic messages that can be sent over a communications network from one computer to another. 2. Sending messages electronically between computers or terminals.

electronic microphone—A device that depends for its operation on the generation of a voltage by the motion of one of the electrodes in a special electron tube.

electronic mine detector—*See* mine detector

electronic multimeter—A device employing the characteristics of an electron-tube circuit for the measurement of electrical quantities, at least one of which is voltage or current, on a single calibrated scale.

electronic music—The electronic generation and processing of audio signals, or the electronic processing of natural sound, and the manipulation and arrangement of these signals via tape recorders into a finished musical composition.

electronic music synthesizer—An audio signal processor that contains sound generators (oscillators) and additional circuitry such as filters to produce familiar sounds, such as those produced by conventional musical instruments, or to create unique sounds and effects.

electronic news-gathering—Abbreviated ENG. The use of video cameras, recording, and other ancillary electronic gear to collect news stories for TV airing.

electronic organ—The electronic counterpart of the pipe organ. All tones and tone variations, such as vibrato, tremolo, etc., are produced by electronic circuits instead of by pipes.

electronic pacemaker—Also called a pacemaker. An electrical device, usually with electrodes planted in the myocardium, that performs the pacing function in a diseased heart no longer capable of pacing itself. Electronic pacemakers can receive power from implanted batteries, radio-frequency signals, biological energy sources, etc.

electronic packaging—The coating or surrounding of an electronic assembly with a dielectric compound.

electronic part—A basic circuit element that cannot be disassembled and still perform its intended function. Examples of electronic parts are capacitors, connectors, filters, resistors, switches, relays, transformers, crystals, electron tubes, and semiconductor devices.

electronic photometer—Also called photoelectric photometer. A photometer with a photocell, phototransistor, or phototube for measuring the intensity of light.

electronic power supply—A circuit that transforms electrical input energy, that is, alternating or direct current. (Sources operating on rotating machine principles, or deriving electrical power from other energy forms such as batteries and solar cells, are excluded.) Supplies covered by this definition fall into one of four groups: 1) ac in, dc out—most common supplies. 2) ac in, ac out—line regulators, variable-frequency supplies. 3) dc in, dc out—converters. 4) dc in, ac out—inverters.

electronic products—Materials, parts, components, subassemblies, and equipment that employ the principles of electronics in performing their major functions. These products may be used as instruments and controls in communications, detection, amplification, computation, inspection, testing, measurement, operation, recording, analysis, and other functions employing electronic principles.

electronic profilometer—An electronic instrument for measuring surface roughness. The diamond-point stylus of a permanent-magnet dynamic pickup is moved over the surface being examined. The resultant variations in voltage are amplified and then measured with a meter calibrated to read directly in microinches of deviation from smoothness.

electronic raster scanning—Scanning by electronic means so that substantially uniform coverage of an area is provided by a predetermined pattern of scanning lines.

electronic reconnaissance—Search for electromagnetic radiations to determine their existence, source, and pertinent characteristics for electronic warfare purposes.

electronic rectifier—A rectifier using electron tubes or equivalent semiconductor elements as rectifying elements.

electronic regulation—*See* electronic control.

electronic relay—An electronic circuit that provides the functional equivalent of a relay but has no moving parts.

electronics—1. The field of science and engineering concerned with the behavior of electrons in devices and the utilization of such devices. 2. Of or pertaining to the field of electronics, such as electronics engineer, course, laboratory, committee. 3. Name given to that branch of electrical engineering which deals with devices whose operation depends upon the movement of electrons in space as opposed to the movement of electrons in liquids or solid conductors, e.g., radio tubes, photoelectric cells. It includes the study of radio, radar, television, sound films, and control of industrial processes. 4. That branch of science involved in the study and utilization of the motion, emissions, and behaviors of currents of electrical energy through gases, vacuums, semiconductors, and conductors; not to be confused with electrics, which deals primarily with the conduction of large currents of electricity through metals. 5. That branch of science and technology which deals with the study, application, and control of the phenomena of conduction of electricity in a vacuum, in gases, in liquids, in semiconductors, and in conducting and superconducting materials.

electronic search reconnaissance — The determination of the presence, source, and significant characteristics of electromagnetic radiations.

electronic security — Protection resulting from measures designed to deny to unauthorized persons information of value that might be obtained by interception and analysis of noncommunications electromagnetic radiations.

electronic shock absorption — An integrated data bit storage buffer inside a CD portable, which receives information at twice the normal speed but supplies information to the digital-to-analog converter at normal speed, ensuring that any interruption of the data flow caused by shocks or bumps does not result in interruption of play.

electronic shutter — A mechanical shutter with an electronic timing circuit. This circuit allows a wider range of exposure times, can be more accurate, and, placed in a circuit with a photconductive cell, allows automatic setting of shutter speeds.

electronic sky screen equipment — An electronic device for indicating the departure of a missile from a predetermined trajectory.

electronic speed control — 1. A system whereby a motor's speed is controlled by feedback from a frequency-sensing circuit attached to the device being powered; changes from the desired speed cause corrective signals to speed up or slow down the motor. 2. Changes in speed in a record player, whether gross (as from 33 $^1/_3$ to 45 rpm) or small (as an order of ±3 percent), can be made by alternating components in the external speed-regulation circuit, rather than by mechanically shifting belts or idler wheel.

electronic sphygmomanometer — Device that measures and/or records blood pressure electronically.

electronic stethoscope — An electronic amplifier of sounds within a body. Its selective controls permit tuning for low heart tones or high pulmonary tones. It has an auxiliary output for recording or viewing audio patterns.

electronic stimulator — A device for applying electronic pulses or signals to activate muscles, or to identify nerves, or for muscular therapy, etc.

electronic surge arrestor — A device used to switch high-energy surges to ground so as to reduce the transient energy to a level that is safe for secondary protectors (e.g., zener diodes, silicon rectifiers).

electronic switch — 1. A circuit element causing a start and stop action or a switching action electronically, usually at high speeds. 2. An electronic circuit used to perform the function of a high-speed switch. Applications include switching a cathode-ray oscilloscope back and forth between two inputs at such high speed that both input waveforms appear simultaneously on the screen.

electronic switching — Electronic circuits and solid-state devices used to perform most telephone central office switching functions.

electronic switching system — Abbreviated ESS. 1. A telephone switching system that uses a computer with a storage containing program switching logic. The output of the computer actuates reed or electronic switches that establish telephone connections automatically. 2. A system that uses solid-state switching devices and computerlike operations to accomplish switching of telephone calls. 3. A type of telephone switching system that uses a special-purpose digital computer to direct and control the switching operation. ESS permits custom-calling services such as speed dialing, call transfer, and three-way calling.

electronic thermal conductivity — The part of thermal conductivity due to the transfer of thermal energy by means of electrons and holes.

electronic timer — 1. A synchronizer, pulse generator, modulator, or keyer that originates a series of continuous control pulses at an unvarying repetition rate known as the pulse-recurrence frequency. 2. A timer using electronic circuits (either tube or transistor type) to control a time period, in place of a motor or other means.

electronic tube relay — A relay that employs electronic tubes as components.

electronic tuning — 1. Altering the frequency of a reflex klystron oscillator by changing the repeller voltage. 2. Frequency changing in a transmitter or receiver by changing a control voltage rather than circuit components.

electronic video recording — The recording of video images by means of photographic film, or magnetic tape or disk, so that the image's record can be played back in a video format at a later time.

electronic viewfinder — Also called viewfinder monitor. 1. A small TV screen attached to a video camera that allows the operator to view a given scene exactly as it is being viewed by the camera. 2. A small television camera that replaces the reflex viewfinder of a motion picture camera. This permits the image photographed to be viewed simultaneously by a number of people, since the TV image may be transmitted to several receivers.

electronic voltmeter — Also called vacuum-tube voltmeter. A voltmeter that utilizes the rectifying and amplifying properties of electron tubes or semiconductors and their circuits to secure such characteristics as high input impedance, wide frequency range, peak-to-peak indications, etc.

electronic volt-ohmmeter — A device employing the characteristics of an electron-tube or semiconductor circuit for the measurement of voltage and resistance on a single-calibrated scale.

electronic warfare — Abbreviated EW. 1. Military usage of electronics to reduce an enemy's effective use of radiated electromagnetic energy and to ensure our own effective use. 2. Military action involving the use of electromagnetic energy to determine, exploit, reduce, or prevent hostile use of the electromagnetic spectrum, and action that retains friendly use of the electromagnetic spectrum. There are three divisions within electronic warfare: electronic warfare support measures (ESM), electronic countermeasures (ECM), and electronic counter-countermeasures (ECCM).

electronic warfare support measures — Abbreviated ESM. That division of electronic warfare involving actions taken to search for, intercept, locate, and immediately identify radiated electromagnetic energy for the purpose of immediate threat recognition. Thus, ESM provides a source of information required for immediate action involving electronic countermeasures, electronic counter-countermeasures, avoidance, targeting, and other tactical employment of forces.

electronic watch—A timepiece in which a battery replaces the mainspring, and semiconductor elements replace the mechanical switching-contact arrangement.

electronic waveform synthesizer—An instrument using electron devices to generate an electrical signal of a desired waveform.

electron image—A representation of an object formed by a beam of electrons focused by an electron optical system.

electron image tube—1. A cathode-ray tube having a photoemissive mosaic upon which an optical image is projected, and an electron gun to scan the mosaic and convert the optical image into corresponding electrical current. 2. A cathode-ray tube that increases the brightness or size of an image, or forms a visible image from invisible radiation. The focal plane for the optical image is a large, light-sensitive cold cathode. The emission from the cathode is first accelerated through a suitable lens system and then strikes a fluorescent screen, where an image is formed that is an enlarged and brightened reproduction of the original image.

electron lens—1. The convergence of the electrons into a narrow beam in a cathode-ray tube by deflecting them electromagnetically or electrostatically. So called because its action is analogous to that of an optical lens. 2. A system of deflecting electrodes or coils designed to produce an electric field that influences a beam of electrons in the same manner in which a lens affects a light beam.

electron metallurgy—That branch of metallurgy that uses electron microscopic techniques in the examination of the nature of metals.

electron micrograph—A reproduction of an image formed by the action of an electron beam on a photographic emulsion.

electron micrography—The photographic recording of images produced by the electrons from an electron microscope. The electron beam carries the images through an array of lenses, and an enlarged electron image is used to stimulate a fluorescent screen that is photographed by a camera system.

electron microradiography—The photographic recording and later enlarging of very thin specimens, using an electron beam to form the image.

electron microscope—A device utilizing an electron beam for the observation and recording of submicroscopic samples with the aid of photographic emulsions or other short-wavelength sensors. Useful magnification is over 300,000.

electron microscopy—The study of materials by means of the electron microscope.

electron mirror—An electron instrument used to totally reflect an electron beam.

electron multiplier—A vacuum tube in which electrons liberated from a photosensitive cathode are attracted to a series of electrodes called dynodes. In doing so, each electron liberates others by secondary emission and thereby greatly increases the number of electrons flowing in the tube.

electron-multiplier section—A section of an electron tube in which an electron current is amplified by one or more successive dynode stages.

electron optical system—A combination of parts capable of producing and controlling a beam of electrons to produce an image of an object.

electron optics—1. The branch of electronics concerned with the behavior of the electron beam under the influence of electrostatic and electromagnetic forces. 2. The control of free electron movement through the use of electric or magnetic fields, and use of the electron movement in research investigation of electronic diffraction phenomena, directly analogous to the control of light through use of lenses. 3. The area of science devoted to the directing and guiding of electron beams using electric fields in the same manner as lenses are used on light beams. 4. Pertaining to devices whose operation relies on modification of a material's refractive index by electric fields, for example, image-converter tubes and electron microscopes.

electron-pair bond—A valence bond formed by two electrons, one from each of two adjacent atoms.

electron paramagnetic resonance—A condition in which a paramagnetic solid subjected to two magnetic fields, one of which is fixed and the other normal to the first and varying at the resonance frequency, emits electromagnetic radiation associated with changes in the magnetic quantum number of the electrons.

electron probe—A narrow beam of electrons used to scan or illuminate an object or screen.

electron-ray tube—Also called a magic eye. 1. A tube that indicates visibly on a fluorescent target the effects of changes in control-grid voltage applied to the tube. Used as a tuning indicator in receivers. 2. A type of recording-level indicator using a luminous display in a special tube. The display is typically like an "eye" with a keyhole in the middle, and maximum recording level corresponds to the closing-up of a slot at the bottom of the keyhole (largely superseded by meters in current-model recorders).

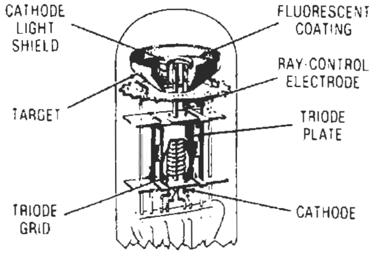

Electron-ray tube.

electron scanning—1. The moving of an electron beam back and forth and/or up and down by deflecting the beam electromagnetically or electrostatically. 2. A deflection of a beam of electrons, at regular intervals, across a CRT screen, according to a definite pattern.

electron spectroscopy—The theory and interpretation of spectra produced by the electron emissions of substances after their irradiation by X-rays.

electron spin—The twirling motion of an electron, independent of any orbital motion.

electron-stream potential—The time average of the difference in potential between a point in an electron stream and the electron-emitting surface.

electron-stream transmission efficiency—With respect to an electrode through which an electron stream passes, the ratio of the average stream current through the electrode to the stream current approaching the electrode. (In connection with multitransit tubes, the electron stream is considered to include only those electrons approaching the electrode for the first time.)

electron telescope—1. An apparatus for seeing through haze and fog. An infrared image is formed optically on the photoemissive mosaic of an electron-image tube and then made visible by the tube. 2. An instrument that serves to produce an enlarged electron image on a fluorescent screen by focusing an infrared image of a distant object on a photosensitive cathode.

electron trajectory—The path of an electron.

electron transit time—The time required for electrons to travel between two electrodes in a vacuum tube. This time is extremely important in tubes designed for ultrahigh frequencies.

electron tubes—Devices used to control the flow of electrons. They may be either gas filled or partially or fully evacuated (vacuum). Common tubes include vacuum tubes, cathode-ray tubes, phototubes, mercury vapor tubes, thyratrons, and microwave tubes.

electron-tube static characteristic—The relationship between two variables of an electron tube, such as the voltage and current of an electrode, with all other variables maintained constant.

electron unit—The unit of charge (negative or positive) equal to the charge on an electron.

electron velocity—The rate of motion of an electron.

electronvolt—Abbreviated ev or eV. The amount of kinetic energy gained by an electron when it is accelerated through an electric potential difference of 1 volt. It is equivalent to 1.602189×10^{-12} erg, or 1.602189×10^{-9} J. It is a unit of energy or work, not of voltage.

electron-wave tube—An electron tube in which streams of electrons having different velocities interact and cause a progressive change in signal modulation along their length.

electro-oculography—Recording and interpretation of the voltages that accompany eye movements. Eye-position voltages from electrodes placed on the skin near the eye are amplified and applied to a strip-chart recorder.

electro-optical detector—A device that detects radiation by utilizing the influence of light in forming an electrical signal. The detector may be a phototube; a photoconductive, photovoltaic, or photojunction cell; a phototransistor; or a thermal detector, such as a thermocouple or bolometer.

electro-optical transistor—A transistor capable of responding in nanoseconds to both light and electrical signals.

electro-optic coefficient—A measure of the extent to which the index of refraction changes with applied high electric fields, such as several parts per ten thousand for applied fields of the order of 20 volts per centimeter. Since the phase shift of a light wave is a function of the index of refraction of the medium in which it is propagating, the change in index can be used to phase modulate the light wave by shifting its phase at a particular point along the guide by changing the propagation time to the point.

electro-optic effect—The change in the index of refraction of a material when subjected to an electric field. The effect can be used to modulate a light beam in a material since many properties—such as light-conducting velocities, reflection and transmission coefficients at interfaces, acceptance angles, critical angles, and transmission modes—are dependent on the refractive indexes of the media in which the light travels.

electro-optic material—A material having refractive indexes that can be altered by an applied electric field.

electro-optic modulator—A device that uses an applied electric field to alter the polarization properties of light.

electro-optic phase modulation—Modulation of the phase of a light wave, such as by changing the index of refraction and thus the velocity of propagation and hence the phase at a point in the medium in which the wave is propagating, in accordance with an applied field serving as the modulating signal.

electro-optic radar—A radar system in which electro-optic instead of microwave techniques and equipment are used to perform the acquisition and tracking operation.

electro-optics—The study of the effects of electric fields on optical phenomena.

electro-optic shutter—A device used to control or block a light beam by means of the Kerr electro-optical effect.

electropad—The part of an electrocardiograph body electrode that makes contact with the skin.

electrophonic effect—The sensation of hearing produced when an alternating current of suitable frequency and magnitude from an external source is passed through an animal or human body.

electrophoresis—1. The movement of particles or ions in solution caused by applying an electric field, as reported by O. Lodge in 1886. 2. The migration of colloidal particles under the influence of an applied electrical field. A colloidal particle, such as a protein molecule, has large numbers of positive and negative radicals that act as if they were on the surface. Thus, since protein molecules carry electric charges, they will migrate when subjected to an electric field. The fractional nature of the net charge makes possible a wide variety of electrophoretic patterns at a given pH. 3. The migration of molecules under the influence of an electric field. 4. The migration of dispersed solid, liquid, or gaseous material to one of two electrodes under the influence of an impressed direct-current voltage. 5. The motivation of particles or ions, suspended in a solution, toward the electrode having the opposite sign, due to the application of an electrical field.

electrophoresis apparatus—An apparatus for causing migration of charged particles (ions) in solution in an electric field. Types include paper, cascading electrodes, high voltage, gel, and thin layer.

electrophoresis scanner—An instrument for reading bands on paper strips or gel, for the purpose of measuring particle movement due to electrophoresis.

electrophoretic display—1. A reflective display that offers a wide choice of colors and has a short-to-medium-term memory that consumes no power. The heart of the display is a suspension of charge pigment particles in a liquid of another color. The suspension, a layer typically 50 micrometers thick, is sandwiched between a pair of electrodes, one of which is transparent. When direct current of the right polarity is applied to the electrodes, the particles are pulled toward the transparent electrode, thus displacing the contrasting liquid and showing their own coloration. When the polarity is reversed, they move to the other electrode and are hidden by the liquid. 2. The movement of suspended particles through a fluid by an electromotive force.

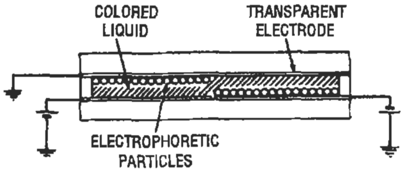

Electrophoretic display.

electrophorus — 1. An early type of static-electricity generator. 2. Simple piece of apparatus used in the laboratory to obtain a number of charges of static electricity from a single initial charge. Typically, it consists of a thick ebonite disc held in a brass sole, and a brass disc with insulated handle. The ebonite disc is charged by rubbing with fur, and the metal disc is brought near and allowed to pick up an induced charge that can be lifted and conveyed where required. 3. A device in which the electric field of an object that has been electrified by friction is used to induce charges in conductors.

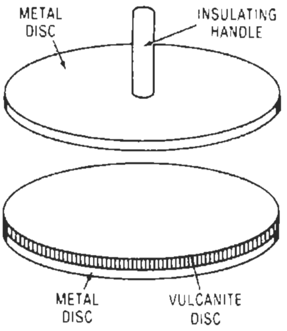

Electrophorus.

electrophotograph — The image formed in electrophotography.

electrophotographic process — The process in which images are formed by various electrical and photographic means. Examples are processes employing selenium-coated drums or zinc-oxide-coated paper.

electrophotography — 1. A term referring to a photographic process in which electrical energy is used to make materials sensitive to light. 2. The photographic recording of an image formed by the alteration in electrical properties of the sensitive materials, induced by the action of light.

electrophotometer — An instrument using a photoelectric sensor for colorimetric determinations.

electrophysiology — The science of physiology as related to electric reactions of the body.

electroplaques — Individual electricity-producing cells in eels and other electric fishes connected in series-parallel arrays, like miniature elements of a battery. They are usually thin waferlike cells, the two surfaces of which differ markedly.

electroplate — 1. To deposit a metal on the surface of certain materials by means of electrolysis. 2. To effect the transfer of one metal to another by means of electrolysis. 3. To apply a metallic coating on a conductive surface by means of electrolytic action.

electroplating — The electrodeposition of an adherent metal coating on a conductive object for protection, decoration, or other purposes, such as securing a surface with properties or dimensions different from those of the basis metal. The object to be plated is placed in an electrolyte and connected to one terminal of a dc voltage source. The metal to be deposited is similarly immersed and connected to the other terminal. Ions of the metal provide transfer to the metal as they make up the current between the electrodes. *See also* electroforming.

electropolishing — 1. The process of producing a smooth, lustrous surface on a metal by making it the anode in an electrolytic solution and preferentially dissolving the minute protuberances. 2. The improvement in surface finish of a metal effected by making it anodic in an appropriate solution.

electropositive developer — A developer containing positively charged toner particles.

electrorefining — 1. The removal of impurities from a metal by electrolysis. 2. The process of anodically dissolving a metal from an impure anode and depositing it cathodically in a purer form.

electroretinograph — An instrument for measuring the electrical response of the human retina to light stimulation.

electroretinography — Recording and interpretation of the voltage generated by the retina of the eye. An electrode fitted to a plastic contact lens is used to pick up voltage from the surface of the eyeball.

electroscope — An electrostatic instrument for measuring a potential difference or an electric charge by means of the mechanical force exerted between electrically charged surfaces.

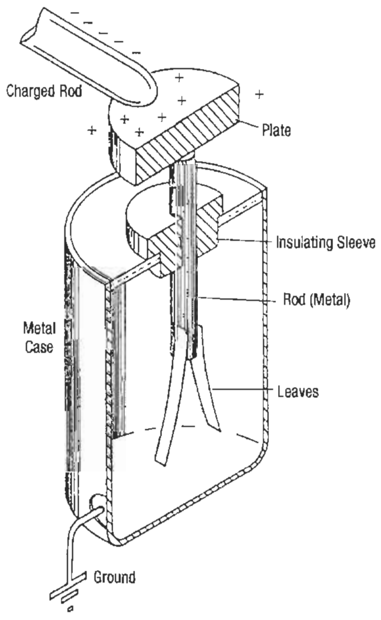

Electroscope.

electrosection — A surgical cutting technique that makes use of a radio-frequency arc.

electrosensitive paper — A paper that turns blue, brown, or black wherever a direct current passes through it. Used with facsimile and high-speed printers.

electrosensitive processor — A facsimile printing process whereby imaging is based on a two-layer paper composed of a white titanium oxide coating and dark underlayer. The paper (which may be sheet or roll fed) is imaged via contact with an electric stylus; as the charged wire touches the paper, the white coating is burned off, line by line, to correspond to the dark image areas of the original.

electrosensitive recording — 1. The passage of electric current into a sheet of sensitive paper to produce a permanent record. 2. A technique that uses the passage of an electric current through a recording medium to produce a permanent image on that medium.

electroshock—A state of shock produced by passing an electric current through the brain. It is useful in the treatment of certain mental disorders.

electroshock therapy—Treatment of certain mental disorders by passing an electric current through the brain.

electrospinograph—A device for detecting and recording electric signals of the spinal cord.

electrostatic—1. Pertaining to static electricity—that is, electricity, or an electric charge, at rest. A constant-intensity electric charge. 2. Applied to speakers and microphones (capacitance type). An electrostatic force is used to activate the diaphragm. The charged diaphragm is suspended between two perforated plates. As an ac signal is applied to the outer plates, the diaphragm vibrates. 3. A form of electrical energy that has the capability of attracting and holding small particles having an opposite electrical charge. 4. A headphone drive system using a thin plastic membrane in a high-voltage electrostatic field, whose variation by the signal voltage moves the entire diaphragm to create a sound pressure wave. 5. The effects produced by electrical charges or fields, alone, without interaction with magnetic influence.

electrostatic actuator—An apparatus comprising an auxiliary external electrode that permits known electrostatic forces to be applied to the diaphragm of a microphone for the purpose of obtaining a primary calibration.

electrostatic capacitor—Two conducting electrodes separated by an insulating material such as air, ceramic, mica, gas, paper, plastic film, or glass. These are generally high-impedance devices.

electrostatic charge—1. An electric charge stored in a capacitor or on the surface of an insulated object 2. The algebraic sum of all positive and negative electric charges present in a specific volume or surface element. 3. An electric charge that is in a state of equilibrium.

electrostatic charge mobility—The property of a barrier material that facilitates or impedes the movement of electrostatic charges internally or on the surface.

electrostatic coating—Process in which the coating material is electrically charged as it leaves the spray gun and is attracted to the part, which has an opposite charge.

electrostatic component—The portion of radiation due to electrostatic fields.

electrostatic-convergence principle—The principle of electron-beam convergence through use of an electrostatic field.

electrostatic copier—A type of copier that employs the principles of photoconductivity and electrostatic attraction.

electrostatic coupling—Method of coupling by which charges on one surface influence those on another through capacitive action.

electrostatic deflection—1. In a cathode-ray tube, the deflection of the electron beam by means of pairs of charged electrodes on opposite sides of the beam. The electron beam is bent toward a positive electrode and bent away from a negative electrode. 2. The deflection of an electron beam by the action of an electrostatic field that has a component perpendicular to the direction of the beam.

elecrostatic discharge—*See* ESD.

electrostatic electrography—That branch of electrostatography which produces a visible record by employing an insulating medium to form latent electrostatic images with the aid of electromagnetic radiation.

electrostatic electrophotography—That branch of electrostatography which produces a visible record by employing a photoresponsive medium to form latent electrostatic images with the aid of electromagnetic radiation.

electrostatic energy—The energy contained in electricity at rest, such as in the charge of a capacitor.

electrostatic field—The vector force field set up in the vicinity of nonmoving electrical charges. The strength of this static field at a point is defined as the force per unit charge on a stationary positive test charge, provided the test charge is so small that it does not disturb the original charge distribution.

electrostatic flux—The electrostatic lines of force existing between bodies at different potentials.

electrostatic focusing—1. The focusing of an electron beam by the action of an electric field. 2. A method of focusing the cathode-ray beam to a fine spot by application of electrostatic potentials to one or more elements of an electron lens system.

electrostatic galvanometer—Galvanometer operated by the effects of two electric charges on each other.

electrostatic generator—A device for the production of electric charges by electrostatic action.

electrostatic headphones—A device held against the ear that reproduces incoming electrical signals as sound. It relies on changes in electrical charge across a diaphragm stretched between two perforated, polarized plates. All parts of the diaphragm experience equal force, and the sound is inherently more linear.

electrostatic induction—1. The process of inducing stationary electric charges on an object by bringing it near another object that has an excess of electric charges. A positive charge will induce a negative charge, and vice versa. 2. Capacitive induction of interfering signals over an air gap separating an instrument (e.g., from its wiring or housing).

electrostatic instrument—An instrument that depends for its operation on the attraction and repulsion between electrically charged bodies.

electrostatic latent image—In an electrostatic copier, the invisible image formed on the zinc-oxide-coated paper by the action of light.

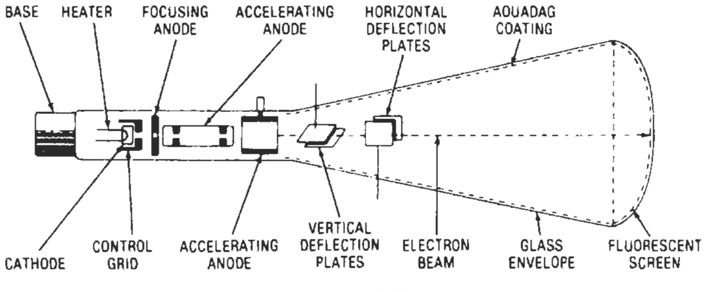

Electrostatic deflection.

electrostatic lens—A lens producing a potential field capable of deflecting electron rays to form an image of an object.

electrostatic loudspeaker—Loudspeaker in which the mechanical forces are produced by the action of electrostatic fields.

electrostatic memory—Also called electrostatic storage. A memory device in which information is retained by an electrostatic charge. A special type of cathode-ray tube is usually employed, together with associated circuitry.

electrostatic memory tube—Also called storage tube. An electron tube in which information is retained by electric charges.

electrostatic meter—A meter with a movement consisting of fixed and movable metal plates interleaved to form a capacitor. Rotation of the movable plates is proportional to the dc or ac voltage applied across the capacitor.

electrostatic microphone—Also called capacitor microphone or condenser microphone. 1. A microphone that contains a metal plate and a thin metal diaphragm set close together. The capacitance of the microphone is thus affected by movement of the diaphragm from air pressure waves. A polarizing voltage is applied to the plates. 2. A microphone whose transduction principle is based on the varying electrical charge across a sound-modulated capacitor. 3. A microphone whose capacitance varies with sound pressure; electronic circuits within the microphone convert this change in capacitance to a varying voltage signal. Electrostatic microphones, unlike other types, require a battery or other voltage source. 4. A type of microphone characterized by its wide frequency range and low distortion. Used for precision measurements and high-quality recording. Can be omnidirectional or cardioid.

electrostatic potential—The voltage that can be measured between any two objects that have different static charges.

electrostatic precipitation—The process of removing smoke, dust, and other particles from the air by charging them so that they can be attracted to and collected by a properly polarized electrode.

electrostatic printer—1. A nonimpact printing technique that forms a copy by attracting toner particles to a static charge on the surface of a photoconductor, then transferring the toner image to the surface of a sheet of copy paper. The image is formed by a laser that develops an electrostatic image charge on the photoconductor according to information being supplied through the input data stream. Each bit of data can be related to a character shape in the memory of the printing system; in most cases, characters are formed by a dot-matrix method similar in concept to that of the matrix printer. Paper is sheet or roll fed. 2. A nonimpact printer that prints dot-matrix characters one at a time by means of wires or pins that supply an electrical charge in the desired patterns onto an aluminum-coated paper; particles of dry ink adhere to the magnetized areas and are then fixed by heat.

electrostatic process—1. A reproduction method in which image formation depends on electrical rather than chemical changes induced by light. 2. A nonchemical, nonimpact imaging process in which a light source, corresponding to the image to be formed, discharges a charged dielectric photoconductive surface to form a latent image. This surface (a photoconductor) containing the latent image is then dusted with dielectric toner powder, which adheres to the charged areas, rendering the image visible.

electrostatic recording—Recording by means of a signal-controlled electrostatic field.

electrostatic relay—A relay in which two or more conductors that are separated by insulating material move because of the mutual attraction or repulsion produced by electric charges applied to the conductors.

electrostatics—The branch of physics concerned with electricity at rest.

electrostatic separator—An apparatus in which a finely pulverized mixture of the materials to be separated is passed through the powerful electrostatic field between two electrodes.

electrostatic series—*See* triboelectric series.

electrostatic shield—1. A shield that prevents electrostatic coupling between circuits, but permits electromagnetic coupling. 2. A metallic enclosure or screen placed around a device so it will not be affected by external electric fields.

electrostatic speaker—Also called capacitor or condenser speaker. A speaker in which the mechanical forces are produced by the action of electrostatic fields.

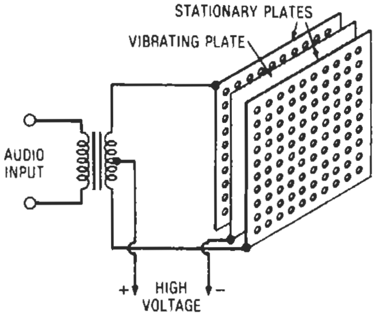

Electrostatic speaker.

electrostatic spraying—A technique of spraying wherein the material being sprayed is given a high electrical charge, while the test piece is grounded.

electrostatic storage—1. The storage of changeable information in the form of charged or uncharged areas, usually on the screen of a cathode-ray tube. 2. Information storage on a dielectric medium that represents the data as those spots on the medium having electrostatic charges, forming an observable record of the data.

electrostatic transducer—A transducer that consists of a capacitor, at least one plate of which can be set into vibration. Its operation depends on the interaction between its electric field and a change in its electrostatic capacity.

electrostatic tweeter—A speaker with a movable flat metal diaphragm and a nonmovable metal electrode capable of reproducing high audio frequencies. The diaphragm is driven by the varying high voltages applied across it and the electrode.

electrostatic unit—An electric unit based primarily on the dynamic interaction of electric charges. Defined as a charge that, if concentrated on a small sphere, would repel with a force of 1 dyne a similar charge 1 centimeter away in a vacuum.

electrostatic voltmeter—A voltmeter depending for its action on electrostatic forces. Its scale is usually graduated in volts or kilovolts.

electrostatography—The process of recording and reproducing visible patterns by the formation and utilization of latent electrostatic charge patterns.

electrostriction — 1. A mechanical deformation caused by the application of an electric field to any dielectric material. The deformation is proportional to the square of the applied field. This phenomenon results from the induced dipole movement caused by the applied field, resulting in the mechanical distortion. 2. Elastic deformation of a dielectric caused by volume force when the dielectric is placed in an inhomogeneous electric field.

electrostrictive effect — The elastic deformation of a dielectric by an electrostatic field.

electrostrictive relay — A relay whose operation is produced by an electrostrictive dielectric actuator.

electrosurgery — The surgical use of electricity in such applications as dissection, coagulation, laser heating, laser welding, diathermy, desiccation of tumors, and hemostasis.

electrosurgical unit — An rf generator whose output is applied to a blade or wire loop used instead of a conventional scalpel for surgical incision or excision.

electrotape — An electronic distance-measuring device.

electrotherapeutics — *See* electrotherapy.

electrotherapy — Also known as electrotherapeutics. 1. The medical science or use of electricity to treat a disease or ailment. 2. Applying electric current to the body for massage or heat treatment.

electrotherapy apparatus — Equipment for applying electric current to the body for massage or heat treatment.

electrothermal — The heating effect of electric current, or the electric current produced by heat.

electrothermal expansion element — An actuating element consisting of a wire strip or other shape and having a high coefficient of thermal expansion.

electrothermal recorder — A recorder in which heat produces the image on the recording medium in response to the received signals.

electrothermal recording — *See* electrothermal recorder.

electrothermic instrument — An instrument that depends for its operation on the heating effect of a current. Examples are the thermocouple and bolometric, hot-wire, and hot-strip instruments.

electrothermics — The branch of science concerned with the direct transformation of electric energy into heat.

electrotinning — Electroplating tin on an object.

electrotyping — The production of printing plates by electroforming.

electrowinning — The process by which metals are recovered from a solution by electrolysis.

element — 1. One of the 104 known chemical substances that cannot be divided into simpler substances by chemical means. A substance whose atoms all have the same atomic number (e.g., hydrogen, lead, uranium). 2. In a computer, the portion or subassembly that constitutes the means of accomplishing one particular function, such as the arithmetic element. 3. Any electrical device (such as an inductor, resistor, capacitor generator, line, or electron tube) with terminals at which it may be connected directly to other electrical devices. 4. The dot or dash of an International Morse character. 5. A radiator, either active or parasitic, that is part of an antenna. 6. The smallest portion of a televised picture that still retains the characteristics of the picture. 7. A portion of a part that cannot be renewed without destruction of the part. 8. A part of an integrated circuit that contributes directly to its electrical characteristics. An active element exhibits gain, such as a transistor; a passive element does not have gain, such as a resistor or capacitor. 9. Lowest level design entity having an identifiable logical, electrical, or mechanical function.

elemental area — *See* picture element.

elemental semiconductor — A semiconductor containing only one element in the undoped state.

elementary charge — A natural unit or quantum into which both positive and negative charges appear to be subdivided. It is the charge on a single electron and has a value of about 4.8037×10^{-10} electrostatic units.

element error rate — The ratio of the number of elements incorrectly received to the total number of elements sent.

elevation — The angular position perpendicular to the earth's surface.

elevation-position indicator — A radar display that simultaneously shows angular elevation and slant range of detected objects.

elevator leveling control — A positioning control used to align the platform of an elevator with the floor level of the building. Metal vanes are mounted in the elevator shaft at each floor level, and an oscillator is mounted on the elevator car. When the elevator is properly leveled, the metal vane is between the plate and the grid coils of the oscillator. A relay connected in the oscillator circuit now energizes. The contacts of this relay are connected in the motor-control circuit of the elevator so that the elevator stops in alignment with the floor level.

ELF — Abbreviation for extremely low frequency.

eliminator — Also called a battery eliminator. A device operated from an ac or dc power line and used for supplying direct current and voltage to a battery-operated circuit.

***E*-lines** — Contour lines of constant electrostatic field strength with respect to some reference base.

elliptically polarized wave — An electromagnetic wave whose electric intensity vector describes an ellipse at one point.

elliptical polarization — Polarization in which the wave vector rotates in an elliptical orbit about a point.

elliptical stylus — A stylus whose cross section, as seen from above, is an ellipse placed across the record groove. Elliptical styli can more readily trace the finer high-frequency modulations of the groove than can spherical styli. Such styli have two radii (e.g., 0.4 × 0.7 mil). *See* biradial stylus.

elliptic function — A mathematical function employed in obtaining the squarest possible amplitude response, or the sharpest passband magnitude rolloff, of a filter with a given number of circuit elements. The elliptic function has a Tchebychev response in both the passband and the stopband. The phase response and transient response of an elliptic-function filter are poorer than for any of the classical transfer functions.

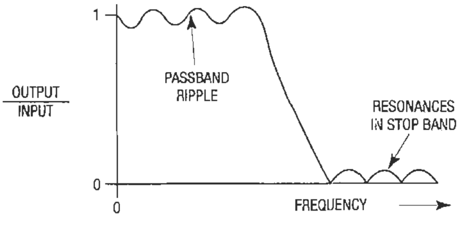

Elliptic-function filter passband.

elongation — Extension of the envelope of a signal as a result of the delayed arrival of certain of the multipath components.

e/m — The ratio of the electric charge to the mass for particles such as electrons and positive rays. For slow-moving electrons, the value of e/m is 1.7588×10^4 coulombs per gram. The value decreases with increasing velocity, however, because of an increase in effective mass.

e-mail — Abbreviation for electronic mail. 1. Messages, usually text, sent from one person to another via computer. E-mail can also be sent automatically to a large number of addresses (mailing list). 2. A system for transmitting messages from one computer terminal to another, where it can be displayed at the receiver's convenience.

embedded code — Machine instructions for checking copy protection that are interspersed with code for other purposes.

embedded software — Computer code that is not visible to the end user.

embedded system — A system into which one or more computing devices (which may be microprocessors or microcontrollers) are incorporated in such a way that the embedded device or devices are not directly accessible to the user of the system.

embedding — A general term for all methods of surrounding or enclosing components and assemblies with a substantial thickness of electrically insulating solid or foam material that substantially fills the voids or interstices between parts.

embedment — The complete encasement of a part or assembly to some uniform external shape. A relatively large volume of a complete package consists of the embedment material.

embossed-foil printed circuits — A printed circuit formed by indenting the desired pattern of metal foil into an insulating base and then mechanically removing the remaining unwanted raised portion.

embossed-groove recording — A method of recording sounds on discs or film strips by embossing sound grooves with a blunt stylus rather than by cutting into them with a sharp stylus. Embossing throws the material up in furrows on each side of the sound groove without actually removing any of the material in the disc or strip.

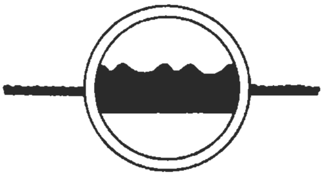

Embossed-groove recording.

embossing stylus — A recording stylus with a rounded tip that forms a groove in the recording medium by merely displacing the material instead of removing it completely.

EMC — Abbreviation for electromagnetic compatibility.

emergency communication — The transmission or reception of distress, alarm, urgent, or safety signals or messages relating to the safety of life or property, or the occasional operation of equipment to determine whether it is in working condition.

emergency radio channel — Any radio frequency reserved for emergency use, particularly for distress signals.

emergency receiver — Receiver immediately available in a station for emergency communication and capable of being energized solely by a self-contained or emergency power supply.

emergency service — The radiocommunication service carrier used for emergency purposes.

emf — Abbreviation for electromotive force.

EMG — Abbreviation for electromyography.

EMI — Abbreviation for electromagnetic interference.

emission — 1. The waves radiated into space by a transmitter. 2. The ejection of electrons from the surface of a material (under the influence of heat, for example).

emission characteristic — The relationship between the emission and the factor controlling it, such as temperature, voltage, or current of the filament or heater. This relationship is usually shown on a graph.

emission current — The current produced in the plate circuit of a tube when all the electrons emitted by the cathode pass to the plate.

emission efficiency — The rating of a hot cathode. Expressed in milliamperes per watt.

emission power — The time rate at which radiant energy is given off in all directions per unit surface area of a radiating body at a given temperature.

emission spectrum — 1. The spectrum showing the radiation emitted by a substance, such as the light emitted by a metal when placed in an electric arc, or the light emitted by an incandescent filament. 2. The spectrum formed by radiation from an emitting source, in contrast to absorption spectra.

emission types — The classification of modes of radio transmission adopted by international agreement. The AM designations are as follows:

Type A0: Unmodulated continuous-wave transmission
Type A1: Telegraphy or pure continuous waves
Type A2: Modulated telegraphy
Type A3: Telephony
Type A4: Facsimile
Type A5: Television

emission-type tube tester — Also called an English-reading tube tester. A tube tester for checking the electron emission from the filament or cathode. The indicating meter is generally calibrated to read "good" or "bad." The tester connects all elements, such as the plate and screen, suppressor, and control grids, together and uses them as an anode.

emission velocity — The initial velocity at which electrons emerge from the surface of a cathode, ranging from zero up to a few volts (attained by very few electrons). This effect accounts for the existence of virtual cathodes and also for the shape of the cutoff region of plate current.

emissive power — The emissivity of a body times the emissive power of a blackbody at the same temperature. For a blackbody, it is the total radiation per unit area of radiating surface.

emissivity — 1. The ratio of the radiant energy emitted by a radiation source to the radiant energy of a perfect (blackbody) radiator having the same area and at the same temperature and conditions. 2. The percentage of infrared energy emitted from a surface for a given temperature compared with the total energy it should emit for that temperature. Instead of percentage, emissivity (E) is expressed in terms of unity. A perfect blackbody has an $E = 1.0$. When the surface has less than 100-percent radiating efficiency, the difference between the E of the surface and 1.0 represents the approximate reflectivity of the surface. Reflectance varies inversely with emissivity. This allows surrounding ambient surfaces to add to or subtract from the temperature.

emitron — A cathode-ray tube developed in England by J. H. Hixenweaver in 1955.

emitron camera — A British television camera tube resembling an iconoscope.

emittance — The power per unit area radiated by a source of energy.

emitter — Also called source. 1. One of the three regions that form a bipolar transistor. Under forward bias of the emitter-base pn junction, the emitter injects minority carriers (electrons or holes) into the base region, where they either recombine or diffuse into the collector. The flow of minority carriers from the emitter to the collector is controlled by the base-emitter pn junction, thereby giving rise to signal amplification. 2. An electrode within a transistor from which carriers are usually minority carriers; when they are majority carriers, the emitter is referred to as a majority emitter. 3. In fiber optics, the source of optical power. *See* coherent emitter; incoherent emitter.

emitter-base and collector-base junction — In a semiconductor, the region where the base and collector and the emitter and base meet. These junctions are defined on the surface of the chip as an oxide step.

emitter bias — The bias voltage applied to the emitter of a transistor.

emitter-coupled logic — *See* ECL.

emitter current — The direct current flowing in the emitter circuit of a transistor.

emitter cutoff frequency — That frequency at which the β of a transistor is down 3 dB from the low-frequency value.

emitter depletion-layer capacitance — The part of the capacitance across an emitter-base junction of a semiconductor that is associated with its depletion layer. The emitter depletion-layer capacitance is a function of the total potential drop across the depletion layer.

emitter follower — A transistor amplifier circuit configuration analogous to a vacuum-tube follower. The circuit is characterized by relatively high input impedance, low output impedance, and a voltage gain of less than unity.

emitter junction — 1. A semiconductor junction normally biased in the low-resistance direction so that minority carriers are injected into the interelectrode region. 2. A junction between the base and emitter regions of a semiconductor normally biased in the forward direction, and through which the charge carriers flow from a region in which they are majority carriers to one in which they are minority carriers.

emitter region — That part of a transistor lying between the emitter junction and the emitter electrode from which carriers flow across the emitter junction.

emitter resistance — The resistance in series with the emitter lead in the common-T equivalent circuit of a transistor.

emitter semiconductor — A junction normally biased in the low-resistance direction to inject minority carriers into an interelectrode region.

emitter series resistance — The resistance between the emitter terminal of a semiconductor and the internal inaccessible emitter point in an equivalent circuit.

emitter terminal — The specified externally available point of connection to the emitter region.

emitter voltage — The voltage between the emitter terminal and a reference point.

EMP — Abbreviation for electromagnetic pulse.

emphasizer — A circuit or device that provides an intentional increase in signal strength at certain audio frequencies.

empire cloth — A cotton or linen cloth coated with varnish and used as insulation on coils and other parts of electrical equipment.

empirical — 1. Based on actual measurement, observation, or experience, as opposed to theoretical determination. 2. Based solely on experiment or observation, rather than on scientific theory. 3. Pertaining to a statement or formula based on experience or observation rather than on deduction or theory.

EMR — Abbreviation for electromagnetic radiation.

emu — Abbreviation for electromagnetic unit.

emulate — 1. To imitate one system with another, such that the imitating system accepts the same data, executes the same programs, and achieves the same results as the imitated system. 2. To imitate a computer system by a combination of hardware and software that allows programs written for one computer to be run on another. *See* simulate.

emulation — 1. The imitation of all or part of one device, terminal, or computer by another, so that the emulating device accepts the same data, performs the same functions, and appears to other network devices as if it were the emulated device. 2. The imitation of a computer system, performed by a combination of hardware and software, that allows programs to run between incompatible systems. 3. For PCs, the process of imitating the behavior of one operating system using a completely different operating system. 4. The generation of one system's code set by another so that the two may communicate. For example, a system with TTY emulation appears like a Teletype system when communicating with another Teletype. 5. The use of hardware or software to generate in real time the expected correct output responses for comparison to the device under test. 6. A hardware model of the target microprocessor used to check out the target system. This can be either the same microprocessor model as used in a target system, or bit-slice architecture that mimics the target microprocessor's function. Using the target microprocessor is called substitutional emulation or in-circuit emulation. *See also* in-circuit emulation.

emulator — 1. A device that is capable of operating in such a manner that it appears to have all of the characteristics of another device. For example, a hardware and software combination that enables one computer to execute programs written for another computer, or a device that produces the same set of outputs for a given set of inputs as does another device. 2. The combination of programming techniques and special machine features that permit a given computing system to execute programs written for another system. 3. A program or a hardware device that duplicates the instruction set of one computer on a different computer, allowing program development for the emulated computer without that computer being available.

emulsion — A suspension of finely divided photosensitive chemicals in a viscous medium, used in semiconductor processing for coating glass masks.

enable — To permit a circuit to be activated by the removal of a suppression signal.

enabling gate — A circuit that determines the start and length of a generated pulse.

enabling pulse — 1. A pulse that opens a normally closed electric gate, or otherwise permits occurrence of an operation for which it is a necessary but not sufficient condition. 2. A pulse that prepares a circuit for some subsequent action.

enameled wire — Wire coated with a layer of baked-enamel insulation.

encapsulant — A material, usually epoxy, used to encase and seal all components in an electronic circuit.

encapsulate—To embed electronic components or other entities in a protective coating, usually done when the plastic encapsulant is in fluid state so that it will set in solid form as an envelope around the work.

encapsulated relay—A relay embedded in a suitable potting compound.

encapsulating—1. Coating by dipping, brushing, spreading, or spraying an electronic component or assembly. An encapsulated unit usually retains its original geometry. 2. Enclosing an article in an envelope of plastic by immersing the object in a casting resin and allowing the resin to polymerize or, if hot, cool.

encapsulating material—A composition primarily adapted for use on or around an electrical device to provide protection from the surrounding environment.

encapsulation—1. A protective coating of cured plastic placed around delicate electronic components and assemblies. It is similar to potting, except the cured plastic is removed from the mold. The plastic therefore determines the color and surface hardness of the finished part. The molds may be made of any suitable material. 2. An embedding process using removable molds or other techniques in which the insulating material forms the outer surfaces of the finished unit. 3. The process of either (*a*) applying a conformal coating by dipping an object in a high viscosity or thixotropic material, or (*b*) using containment and a low viscosity material to provide a relatively thin protective encasement (50 to 100 mils or 1.27 to 2.54 mm) to a part or assembly.

encased control—A self-contained motor speed/torque control completely housed in an enclosure. Switching, indicating, and adjusting devices are provided on the outside of the enclosure. Unit portability, safety, and component protection are leading assets of this design.

encipher—To convert a message from ordinary language into a secret form. *See also* encode

enciphered facsimile communications—Communications in which security is provided by mixing pulses from a key generator with the output of a facsimile converter. Plain text is recovered at the receiving terminal by subtracting identical key pulses. Unauthorized persons are unable to reconstruct the plain text unless they have an identical key generator and they know the daily key setting.

enclosed relay—A relay in which both the coil and the contacts are protected from the environment.

enclosed switch—Switch having internal parts protected by a housing. The enclosed switch can be dust proof, moisture proof, oil or contamination proof, or hermetically sealed.

enclosure—1. An acoustically designed housing or structure for a loudspeaker; also any cabinet for a component, electrical, or electronic device. 2. A surrounding case designed to provide a degree of protection for equipment against a specified environment and to protect personnel against accidental contact with the enclosed equipment.

encode—Also called encipher. 1. To use a code, frequently one composed of binary numbers, to represent individual characters or groups of characters in a message. 2. To change from one digital code to another. If the codes are greatly different, the process usually is called conversion. 3. To substitute letters, numbers, or characters, usually with the intention of hiding the meaning of the message except from persons who know the encoding scheme. 4. The process of converting an event such as a switch closure into a form suitable for transmission over a communication channel.

encoder—1. A device used to electronically alter a signal so that it can only be viewed on a receiver equipped with a special decoder. 2. Any device that modifies information into the desired pattern or form for a specific method of transmission. 3. An electromechanical device that can be attached to a shaft to produce a series of pulses to indicate shaft position; when the output is differentiated, the device is an accurate tachometer. (It is fundamentally oriented to digital rather than analog techniques.) An encoder contains a disc with a printed pattern; as the disc rotates, it makes and breaks a circuit. The more make-and-break cycles per revolution, the better the resolution. 4. A digital-to analog converter. 5. Circuitry in a quadriphonic sound system that, by matrixing in the recording process, turns four signals into two for inscribing, stereo style, on each wall of the record groove. 6. Electromechanical device that transforms analog motion into digital electrical signals. The outputs are incrementally constant for uniform motion characterized by a staircase function, where the output remains constant for a small range of input values. 7. A digital device for converting an input digital signal into its equivalent binary code. *See also* code converter.

encoder accuracy—The maximum positional difference between the input to an encoder and the position indicated by its output; includes both deviation from theoretical code transition positions and quantizing uncertainty caused by converting from a scale having an infinite number of points to a digital representation containing a finite number of points.

encoding—1. Translation of information from an analog or other easily recognized form to a coded form without a significant loss of information. 2. The process of converting an event such as a switch closure into a form suitable for transmission over a communication channel. 3. The scrambling of a signal to prevent viewing of a program by nonsubscribers.

encryption—1. A change made to data, code, or a file such that it can no longer be read or accessed without processing (or unencrypting). 2. The technique of modifying a known bit stream on a transmission line to make it appear like a random sequence of bits to an unauthorized observer.

end-around carry—A computer operation in which the carried information from the left-most bit is added to the results of the right-most addition. It is used for ones complement and nines complement arithmetic.

end-around shift—In a computer, the movement of characters from one end of the register to the other end of the same register.

end bell—An accessory that is similar to a cable clamp and attaches to the back of a plug or receptacle. It serves as an adapter for the rear of connectors. Some angular end bells have built-in cable clamps. Angular end bells up to 90° are available. *See also* end shield.

end bracket—*See* end shield.

end-cell rectifier—A small trickle-charge rectifier for maintaining the voltage of storage-battery end cells.

end cells—Cells that can be switched in series with a storage battery to maintain the output voltage of the battery when it is not being charged.

end central office—The local central telephone office that interconnects customer lines and trunks. It is designated a Class 5 office in the DDD or intertoll network.

end distortion—A shifting of the ends of all marking pulses of start-stop teletypewriter signals from their proper positions relative to the beginning of the start pulse.

end effect—The capacitive effect at the ends of a half-wave antenna. To compensate for this effect, a dipole is cut slightly shorter than a half wave.

end effector — Terminal on a robot arm that carries a hand, welding gun, painting nozzle, or other tool.

end finish — Surface condition at the optical conductor face.

end-fire array — A linear or cylindrical antenna having its direction of maximum radiation parallel to the long axis of the array.

end instrument — A device connected to one terminal of a loop and capable of converting usable intelligence into electrical signals or vice versa. Includes all generating, signal-converting, and loop-terminating devices at the transmitting and/or receiving location.

endless loop recorder — A dictation system in which a nonremovable magnetic tape is sealed in a tank and the tape runs in a continuous loop.

end mark — In a computer, a code or signal used to indicate the termination of a unit of information.

endocardiac electrodes — *See* implantable pacemaker.

endodyne reception — A British term applying to reception of unmodulated code signals. A vacuum-tube circuit having a local oscillator whose frequency is slightly different from that of the carrier signal. Thus, a beat signal in the audio range is produced.

end-of-block signal — A symbol or indicator that defines the end of a block of data.

end of file — Abbreviated EOF. A code placed by a program after the last byte of a file to tell the computer's operating system that no additional data follows.

end-of-file mark — In a computer, a code instruction indicating that the last record of a file has been read.

end of message — The end of data to be transmitted. It can be indicated by a special control code, as in the ASCII code set; by an absence of data for a specified time interval; or by a particular sequence of block gaps and data, as is done on magnetic tape.

end of tape — The point on a computer tape at which the system or operator is given a warning that the physical end of the tape is approaching. It is approximately 25 feet (7.62 m) from the actual end of the tape on $^1/_2$-inch (12.7-mm) computer tape and approximately 50 feet (15.24 m) from the halt marker on $^1/_4$-inch (6.35-mm) tape.

end-of-tape marker — A marker placed on a magnetic tape to indicate the end of the permissible recording area. It may be a photoreflective strip, a transparent section of tape, or a particular bit pattern.

end-of-transmission card — Last card of each message; used to signal the end of a transmission. Contains the same information as the header card, plus additional data for traffic analysis.

endogeneous variable — A variable whose value is determined by relationships included within the model. *See also* exogenous variable.

end-on armature — Of a relay, an armature that moves in the direction of the core axis, with the pole face at the end of the core and perpendicular to this axis.

end-on directional antenna — A directional antenna that radiates chiefly toward the line on which the antenna elements are arranged.

endoradiograph — Equipment for X-ray examination of internal organs and cavities by means of radiopaque materials.

endoradiosonde — Also called radio pill. A device for detecting and transmitting physiological data from the gastrointestinal tract or other inaccessible body cavities.

endothermic — A term describing a chemical reaction in which heat is absorbed.

endothermic reaction — A reaction that is accompanied by the absorption of heat.

end point — The shaft positions immediately before the first and after the last measurable change(s) in output ratio after wiper continuity has been established, as the shaft of a precision potentiometer moves in a specified direction.

end-point control — Quality control by means of continuous automatic analysis. In highly automatic processes, the final product is analyzed, and if any undesirable variations are detected, the control system automatically brings about the necessary changes.

end-point sensitivity — The algebraic difference in electrical output between the maximum and minimum value of the measurand over which an instrument is calibrated.

end-point voltage — The terminal voltage of a cell below which equipment connected to it will not operate or should not be operated.

end resistance — The resistance of a precision potentiometer measured between the wiper terminal and an end terminal, with the shaft positioned at the corresponding end point.

end-resistance offset — In potentiometers, the residual resistance between a terminal and the moving contact, at a position corresponding to full rotation against that terminal.

end-scale value — The value of the actuating electrical quantity that corresponds to end-scale indication of an instrument. When zero is not at the end or at the electrical center of the scale, the higher value is taken. Certain instruments such as power-factor meters, ohmmeters, etc., are necessarily excepted from this definition.

end setting — In a potentiometer, the minimum resistance that is measured between one end of a potentiometer and the wiper, with the wiper mechanically positioned at that end.

end shield — 1. Frequently called end bracket or end bell. In a motor housing, the part that supports the bearing and also guards the electrical and rotating parts inside the motor. 2. In a magnetron, the shield that confines the space charge to the interaction space.

end spaces — In a multicavity magnetron, the two cavities at either end of the anode block that terminate all the anode-block cavity resonators.

end-to-end check — Tests conducted on a completed wire and/or cable run to ensure electrical continuity.

end use — The way the ultimate consumer uses a device.

energize — To apply the rated voltage to a circuit or device, such as to the coil of a relay, in order to activate it.

energized — Also called alive, hot, and live. Electrically connected to a voltage source. Turned on, alive.

energized part — A part at some potential with respect to another part, or the earth.

energy — 1. The capacity for performing work. A particle or piece of matter may have energy because it is moving or because of its position in relation to other particles or pieces of matter. A rolling ball is an example of the first; a ball at rest at the top of an incline is an example of the second. 2. The capacity for doing work and overcoming resistance.

energy conversion — The change of energy from one form to another, e.g., from chemical energy to electrical energy.

energy conversion devices — Devices including primary and secondary cells; fuel cells; photovoltaic systems; electrochemical energy converters; radiation conversion devices; thermionic converters; converters using solar, ionic, or nuclear energy sources; devices for creating a plasma in an interaction space between an emitter and a collector; electrostatic generators for creating an electrical output; organic and inorganic ion exchange and membrane devices; electron volt energy devices; devices for direct conversion of fuel to electricity; and electrical-energy storage-unit devices capable of delivering a power output.

energy density—1. The energy output of a battery, expressed in watt-hours, per unit weight or volume of the battery. 2. The energy per unit volume of a medium.

energy dispersal waveform—A triangular-shaped signal at 30 Hz synchronized with the vertical blanking interval in the TV signal from the satellite, which ensures that the signal will average its power out over the whole channel, even when just the carrier is present. This waveform is removed by the receiver after FM demodulation.

energy efficiency ratio—Term used to gauge the relative electrical efficiency of appliances; found by dividing btu-per-hour output by the number of electrical watts used.

energy gap—The energy range between the bottom of the conduction band and the top of the valence band of a semiconductor.

energy level—A particular value of energy of a physical system, such as a nucleus, that the system can maintain for a reasonably long length of time. Systems on an atomic scale have only certain discrete energy levels and cannot occupy values between these levels.

energy-level diagram—A line drawing that shows the increase or decrease in electrical power as current intensities rise and fall along a channel of signal communications.

energy-measuring equipment—Equipment used to measure energy in electrical, electronic, acoustical, or mechanical systems.

energy of a charge—Represented by $E = 1/2QV$, given in ergs, when the charge Q and the potential V are in electrostatic units.

energy product—The product of the magnetic flux density B in gauss times the magnetic field strength H in oersteds. Used as an index of magnet quality. The larger the maximum energy product, the smaller the required magnet for a given job.

energy-product curve—A curve obtained by plotting the product of the value of magnetic induction B and demagnetizing force H for each point of the demagnetization curve of a permanent magnetic material. Usually shown together with the demagnetization curve.

energy redistribution—A method of finding the duration of an irregularly shaped pulse by considering it as a power curve. The area under the curve can be represented by an equivalent rectangle of the same area and peak amplitude. The original-pulse duration is equal to the rectangle width.

energy state—The position and speed of an electron relative to the position and speed of other electrons in the same atom or adjoining atoms.

energy storage capacitor—Specifically designed capacitor for use in applications wherein the capacitor can be charged over a relatively long period and discharged in a short period, thus increasing the instantaneous power in energy storage systems. Provisions are made to permit the very high currents that accompany the high rate of energy discharge. When the number of discharges is very large, the actual total time during which the capacitor may be subjected to the maximum operating voltage is relatively short.

energy-variant sequential detection—A technique for sequential detection in which a fixed number of transmitted pulses of varying energy are received with a single (upper) threshold device.

ENG—*See* electronic news-gathering.

engineered military circuit—1. Leased long lines of which only the station equipment, local loops, and reserved positions of interexchange channels are paid for continuously. The unresevered portions of leased long lines or interexchange channels are on a steady status and are placed in an operational status and paid for only when required by the command concerned. 2. A standby or on-call circuit that is engineered specifically to meet military criteria.

engineering—A profession in which a knowledge of the natural sciences is applied with judgment to develop ways of utilizing the materials and forces of nature.

English-reading tube tester—*See* emission-type tube tester.

enhanced carrier demodulation—An amplitude-demodulation system in which a synchronized local carrier of the proper phase is added to the demodulator. This has the effect of materially reducing the distortion produced in the demodulation process.

enhancement—Modification of the subjective features of an image to increase its impact on the observer.

enhancement mode—1. An MOS transistor that is normally off with zero gate voltage applied. A gate voltage of the correct polarity attracts majority carriers to the gate area, thus "enhancing" it and forming a current-conducting channel. 2. A device type that is normally off with zero gate voltage. A threshold voltage is then required to turn the device on.

enhancement-mode field-effect transistor—A field-effect transistor in which no device current flows (leakage only) when V_{GS} is zero volts. Conduction does not begin until V_{GS} reaches the threshold voltage.

enhancement mode operation—The operation of a field-effect transistor such that changing the gate-to-source voltage from zero to a finite value increases the magnitude of the drain current.

enhancement MOS transistor—A type of MOS transistor in which no current flows in the absence of an input control signal on the control terminal (called the gate) of the transistor; i.e., a control signal input is required to turn on the device. This reduces power dissipation (power dissipation occurs only when an input signal is present) and results in excellent logic state recognition (full off being one state and on the other).

ensemble—A collection of sample functions of a random process, all of which start from the same zero time.

enterprise number—Also called toll-free number. Unique telephone exchange number that permits the called party to be automatically billed for incoming calls.

entrance box—A metal box that houses overcurrent-protection devices and serves as the point of distribution for the various electrical circuits in a structure.

entrance cable—A cable by means of which electrical power is brought from an outside power line into a building.

entrance delay—The time between actuating a sensor on an entrance door or gate and the sounding of a local alarm or transmission of an alarm signal by the control unit. This delay is used if the authorized access switch is located within the protected area, and permits a person with the control key to enter without causing an alarm. The delay is provided by a timer or timing circuit within the control unit.

entrapped material—Gas or particles bound up in an electrical package so that they cannot escape the package.

entropy—1. A measure of the unavailable energy in a thermodynamic system. 2. The unavailable information in a set of documents. 3. An inactive or static condition (total entropy). 4. A measure of the amount of information in a communication signal, equal to the average number of bits per symbol.

entry—Each statement in a computer programming system.

entry point—In a computer, the programmer-defined instruction at which a task is to begin execution.

enunciation—The act of pronouncing words clearly and distinctly. Articulation.

envelope—1. Also referred to as a bulb. The glass or metal housing of a vacuum tube. The glass housing that encloses an incandescent source. 2. The curve passing through the peaks of a graph and showing the waveform of a modulated radio-frequency carrier signal.

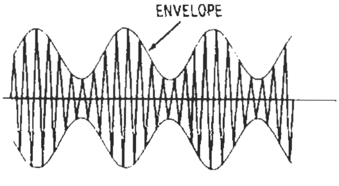

Envelope, 2.

envelope delay—1. The time that elapses as a transmitted wave passes any two points of a transmission circuit. Such delay is determined primarily by the constants of the circuit and is measurable in milliseconds or microseconds. 2. Sometimes called time delay or group delay. The propagation time delay undergone by the envelope of an amplitude-modulated signal as it passes through a filter. Envelope delay is proportional to the slope of the curve of phase shift as a function of frequency. Envelope-delay distortion is introduced when the delay is not the same at all frequencies in the passband. 3. The time difference between the longest delay and the shortest delay for a given band of frequencies. 4. A type of distortion on an analog line in which the signal delay varies with signal frequency. 5. Characteristics of a circuit that result in some frequencies arriving ahead of others, even though they were transmitted together.

envelope-delay distortion—The distortion that occurs during transmission when the phase shift of a circuit or system is not constant over the frequency range.

envelope generator—A circuit in a synthesizer that produces a single, carefully defined waveform. Useful in creating attacks and decays to define notes or for special effects.

environment—The aggregate of all conditions that externally influence a device's performance.

environmental chambers—Test chambers designed to expose the subject being tested to external conditions, such as heat, shock, pressure, and moisture, for the study of their effects of the subject.

environmental conditions—External conditions of heat, shock, vibration, pressure, moisture, etc.

environmentally sealed—Provided with gaskets, seals, potting, or other means to keep out contamination that might reduce performance.

environmental testing—The testing of a system or component under controlled environmental conditions, each of which tends to affect its operation or life.

environmentproof switch—A switch that is completely sealed to ensure constant operating characteristics. Sealing normally includes an O-ring on the actuator shaft and fused glass-to-metal terminal seals or complete potting and an elastomer plunger-case seal.

EOG—Abbreviation for electro-oculography.

episcotister—A device consisting of alternate opaque and transparent discs that rotate at a speed that interrupts light beams at an audio-frequency rate. It modulates the light beam used to excite a photoelectric element.

epitaxial—Pertaining to a single crystal layer on a crystalline substrate, oriented the same as the substrate. In certain semiconductor processes, an expitaxial layer is grown on a silicon substrate during the fabrication of transistors and integrated circuits.

epitaxial deposition—1. The growth of additional material, usually in a thin film, on a substrate. Often the added material has a crystal structure and orientation controlled by matching that of the substrate. 2. Epitaxy. The technique of growing a semiconductor layer on an existing crystal by depositing it directly from reactant vapors, so that the structure of the new layer is isomorphic with, or simply an extension of, that of the original crystal.

epitaxial device—A device constructed in such a manner that the crystalline structure of successive layers is oriented in the same direction as that of the original base material.

epitaxial film—1. A film of single-crystal semiconductor material that has been deposited onto a single-crystal substrate. 2. Any deposited film, provided the orientation of its crystal is the same as that of the substrate material.

epitaxial growth—1. A semiconductor fabrication process in which single-crystal p or n material is deposited and grows on the surface of a substrate. Usually, this material has a different conductivity than the substrate. 2. Crystal growth obtained by depositing a film of monocrystalline semiconductor material on a monocrystalline substrate. 3. The process of producing an additional crystal layer of semiconductor material on a semiconductor substrate. The crystalline structure of the substrate is continued into the epitaxial layer; however, the impurity concentration can be made to differ greatly. 4. The deposition of a single-crystal film on the surface of a single-crystal substrate so that the crystal orientations of the two layers are alike. 5. A process of growing layers of material on a selected substrate. Usually silicon is grown in a silicon substrate. Silicon and other semiconductor materials may be grown on a substrate with compatible crystallography, such as sapphire (silicon-on-sapphire).

epitaxial-growth mesa transistor—A transistor made by overlaying a thin mesa crystal over another mesa crystal.

epitaxial growth process—The process of growing a semiconductor material by depositing it in vaporized form on a semiconductor seed crystal. The deposited layer continues the single-crystal structure of the seed.

epitaxial layer—1. A grown or deposited crystal layer with the same crystal orientation as the parent material and, in the case of semiconductor circuits, of the same basic material as the original substrate. 2. A single-crystal p-type or n-type material deposited on the surface of a substrate. 3. A thin, precisely doped monocrystalline silicon layer grown on a heavily doped thick wafer, into which are diffused semiconductor junctions. In conventional processing of an integrated circuit, the thick wafer is p doped and the epitaxial layer is n doped. 4. A single crystal layer that has been deposited or grown on a crystalline substrate having the same structural arrangement. 5. A layer of silicon grown atop single-crystal silicon that reproduces the same crystallographic orientation as the single-crystal material.

epitaxial material—A material whose atoms are arranged in single-crystal fashion upon a crystalline substrate so that its lattice structure duplicates that of the substrate.

epitaxial planar transistor—A transistor in which a thin collector region is epitaxially deposited on a low-resistivity substrate, and the base and emitter regions are produced by gaseous diffusion with the edges of the junction under a protective oxide mask.

epitaxial process—The process of growing from the vapor phase a single-crystal semiconductor material with controlled resistivity and thickness.

epitaxial transistor—A transistor with one or more epitaxial layers.

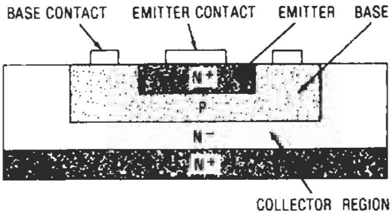

Epitaxial transistor (triple diffused).

epitaxy—1. The controlled growth on a crystalline substrate of a crystalline layer, called an epilayer. In homoepitaxy (e.g., silicon layers on a silicon substrate), the epilayer exactly duplicates the properties and crystal structure of the substrate. In heteroepitaxy (e.g., silicon on sapphire), the deposited epilayer is a different material with a different crystalline structure than that of the substrate. 2. The growth of a crystal on the surface of a crystal of another substance in such a way that the orientation of the atoms in the original crystal controls the orientation of the atoms in the grown crystal.

E-plane—The plane of an antenna containing the electric field. The principal E-plane also contains the direction of maximum radiation.

E-plane bend—Also called E-bend. The smooth change in direction of the axis of a waveguide. The axis remains parallel to the direction of polarization throughout the change.

E-plane T-junction—Also called series T-junction. A waveguide T-junction in which the structure changes in the plane of the electric field.

epoxy—1. Pertaining to a family of thermosetting materials that are widely used for casting and potting and as adhesives. 2. A family of thermosetting resins used in the packaging of semiconductor devices. Epoxies form a chemical bond to many metal surfaces.

EPROM—Abbreviation for erasable programmable read-only memory. A type of nonvolatile memory device whose contents can be erased by exposure to ultraviolet light. *See also* EEPROM; PROM.

epsilon—The Greek letter E, or ε, frequently used to represent 2.71828, which is the base of the natural system of logarithms.

equal-energy source—A source of electromagnetic or sound energy that emits the same amount of energy at each frequency in the spectrum.

equal-energy white—The light produced by a source that radiates equal energy at all visible wavelengths.

equalization—1. The process of reducing the frequency and/or phase distortion of a circuit by the introduction of networks to compensate for the difference in attenuation and/or time delay at the various frequencies in the transmission band. 2. A process of compensating for increases in attenuation (signal loss) with frequency. Different signal frequencies are attenuated differently over a given distance. 3. An intentional departure from response flatness to compensate for complementary characteristics introduced elsewhere in the system (as with discs, tape, and FM broadcasting). Also used to correct for response deficiencies in speakers and other components. 4. Different equalization characteristics are used in the recording and playback amplifiers of a tape recorder, to compensate for the magnetic characteristics of the tape and the heads. Playback equalization is standardized to give flat frequency response with any properly recorded tape, while recording equalization is a property of a particular machine, depending on its head design and the tape for which it was meant. 5. Reshaping the playback characteristics of a recording during playback mode. The simplest way is to adjust the treble and bass controls, but true equalization requires continuous adjustment of the playback frequency response curve at several points. A graphic equalizer is often used for this. 6. The selective amplification or attenuation of certain frequencies. Also refers to recognized industry standards for recording and reproducing characteristics, such as the NAB Standard. 7. The intentional increase in level of certain portions of the audio-frequency spectrum. The term is sometimes misapplied when actually referring to the attenuation of portions of the audio-frequency spectrum. 8. A technique used to compensate for distortions present on a communication channel. Equalizers add loss or delay to signals in inverse proportion to the channel characteristics. The signal response curve is then relatively flat and can be amplified to regain its original form.

equalize—To apply to a circuit an electrical network whose transmission characteristics are complementary to those of the line, so that when the loss (or delay) in the line and that in the equalizer are combined, the overall loss (or delay) is almost the same at all frequencies.

equalizer—1. A passive device designed to compensate for an undesired amplitude-frequency and/or phase-frequency characteristic of a system or component. 2. A series of connections made in paralleled, cumulatively compound direct-current generators to give the system stability. 3. A network, usually adjustable, that corrects the transmission-frequency characteristics of a circuit to permit it to transmit all the frequencies that it passes in a uniform manner. 4. An electronic circuit that introduces compensation for frequency-discriminative effects of elements within a television system. 5. An electronic device that amplifies (boosts) and/or attenuates certain portions of the audio-frequency spectrum. There are many different types of equalizers. 6. A device to allow the frequency response of an audio-signal path to be adjusted in some way.

equalizer circuit breaker—A breaker that serves to control or to make and break the equalizer or the current-balancing connections for a machine field, or for regulating equipment, in a multiple-unit installation.

equalize voltage—A voltage applied to a battery for charging at installation and as a periodic boost charge, common in lead-antimony lead-acid battery systems.

equalizing current—A current circulated between two parallel-connected compound generators to equalize their output.

equalizing network—A network connected to a line to correct or control its transmission frequency characteristics.

equalizing pulses—1. A series of pulses (usually six) occurring at twice the line frequency before and after the serrated vertical TV synchronizing pulse. Their purpose is to cause vertical retrace to occur at the correct instant for proper interlace. 2. In the standard television signal, pulses that minimize the effect of line-frequency pulses on the interlace.

equal-loudness contours—*See* Fletcher-Munson curves.

equation function—As applied to microelectronic circuitry, a combination of electronic elements or circuits capable of solving the electronic-counter portion of a mathematical or Boolean equation. In obtaining the solution, it performs the necessary function within an electronic or electromechanical system.

equation solver—A computer, usually of the analog type, designed to solve systems of linear simultaneous (nondifferential) equations or to find the roots of polynomials.

equilibrium—In a semiconductor context, that state of a semiconductor crystal when there is no net current through the crystal. A crystal is normally in this state when no external voltages or current are impressed on it.

equilibrium brightness—The brightness of the viewing screen when a display storage tube is in a fully written condition.

equilibrium electrode potential—A static electrode potential when the electrode and the electrolyte are in equilibrium with respect to a specified electrochemical reaction.

equiphase surface—In a wave, any surface over which the field vectors at the same instant are either in phase or 180° out of phase.

equiphase zone—In radionavigation, the region in space within which the difference in phase between two radio signals is indistinguishable.

equipment—1. An item having a complete function apart from being a substructure of a system. Sometimes called a set. 2. A general term referring to practically every part of an electrical system, including the parts consuming electrical energy. (Devices are also included in this category.)

equipment augmentation—1. Procuring additional automatic data-processing equipment capability to accommodate increased workload within an established data system. 2. Obtaining additional automatic data-processing equipment capability to extend an established data system to additional sites or locations.

equipment bonding jumper—The connection between two or more portions of the equipment grounding conductor.

equipment chain—A group of units of equipment that are functionally in series. The failure of one or more individual units results in loss of the function.

equipment characteristic distortion—A repetitive display or disruption peculiar to specific portions of a teletypewriter signal. Normally, it is caused by improperly adjusted or dirty contacts in the sending or receiving equipment.

equipment ground—A connection from earth ground to a non-current-carrying metal part of a wiring installation of electric equipment. It reduces shock hazard and provides electrostatic shielding.

equipment life—The arithmetic mean of the cumulative operating times of identical pieces of equipment beginning with the time of acceptance by the ultimate consumer and ending when the equipment is no longer serviceable.

equipotential—A conductor having all parts at a single potential. The cathode of a heater-type tube is equipotential, whereas the filament is not because its voltage varies from one end to the other.

equipotential cathode—*See* indirectly heated cathode.

equipotential line—An imaginary line in space having the same potential at all points.

equipotential surface—A surface or plane passing through all points having the same potential in a field of flow.

equisignal localizer—Also called tone localizer. A type of localizer in which lateral guidance is obtained by comparing the amplitudes of two modulation frequencies.

equisignal radio-range beacon—A radio-range beacon used for aircraft guidance. It transmits two distinctive signals, which are received with equal intensity only in certain directions called equisignal sectors.

equisignal surface—The surface formed around an antenna by all points that have a constant field strength (usually measured in volts per meter) during transmission.

equisignal zone—In radionavigation, the region in space within which the difference in amplitude between two radio signals is indistinguishable.

equivalence—A logic operator having the property that if P is a statement, Q is a statement, R is a statement, etc., then the equivalence of P, Q, R. . . is true if and only if all statements are true or all statements are false.

equivalent absorption—The rate at which a surface will absorb sound energy, expressed in sabins. Defined as the area of a perfect absorption surface that will absorb the same sound energy as the given object under the same conditions.

equivalent absorption area—Area of perfectly absorbing surface that will absorb sound energy at the same rate as the given object under the same conditions. The acoustic unit of equivalent absorption is the sabin.

equivalent binary digits—1. The number of binary digits equivalent to a given number of decimal digits or other characters. 2. The number of binary places required to count the elements of a given set.

equivalent circuit—1. An arrangement of common circuit elements that has characteristics over a range of interest electrically equivalent to those of a different or more complicated circuit or device. 2. A simplified circuit that has the same response to changing voltage and frequency as a more complex circuit. Used to facilitate mathematical analysis.

equivalent circuit of a piezoelectric crystal unit—The electric circuit that has the same impedance as the unit in the frequency region of resonance. It is usually represented by an inductance, capacitance, and resistance in series, shunted by the direct capacitance between the terminals of the crystal unit.

equivalent component density—In circuits in which discrete components are not readily identifiable, the volume of the circuit divided by the number of discrete components necessary to perform the same function.

equivalent conductance—The normal conductance of an atr tube in its mount, measured at its resonance frequency.

equivalent dark-current input—The incident luminous flux required to give an output current equal to the dark current.

equivalent differential input capacitance—The equivalent capacitance looking into the inverting or noninverting inputs of a differential amplifier with the opposite input grounded. *See also* equivalent differential input impedance.

equivalent differential input impedance—The equivalent impedance looking into the inverting or noninverting input, with the opposite input grounded and the operational amplifier operated in the linear amplification region.

equivalent differential input resistance—The equivalent resistance looking into the inverting or noninverting input of a differential amplifier with the opposite input grounded. *See also* equivalent differential input impedance.

equivalent diode—An imaginary diode consisting of the cathode of a triode or multigrid tube and a virtual anode to which is applied a composite controlling voltage of such a value that the cathode current would be the same as the current in the triode or multigrid tube.

equivalent faults—Two or more faults that cause the same output responses and that cannot be isolated from the board output pins and internal nodes being monitored by the tester.

equivalent four-wire system—A transmission system using frequency division to obtain full-duplex operation over only one pair of wires.

equivalent grid voltage—The grid voltage plus plate voltage divided by the mu of the tube.

equivalent height—The virtual height of an ionized layer of the ionosphere.

equivalent input noise current—The equivalent input noise current that would reproduce the noise seen at the output of an operational amplifier if all amplifier noise sources were set to zero and the source impedances were large compared with the optimum source impedance.

equivalent input noise voltage—The equivalent input noise voltage that would reproduce the noise seen at the output of an operational amplifier if all amplifier noise sources and the source resistances were set at zero.

equivalent input offset current—The difference between the two currents flowing into the inverting and noninverting inputs of a differential amplifier when the output voltage is zero.

equivalent input offset voltage—The amount of voltage required at the input to bring the output to zero. Usually this voltage is adjustable to zero by using either a built-in or an external variable resistor (balance control).

equivalent input wideband noise voltage—The output noise voltage of a differential amplifier with the input shorted, divided by the dc voltage gain of the amplifier. This voltage is measured with a true rms voltmeter and is limited to the combined bandwidth of the amplifier and meter.

equivalent loudness—The intensity level of a sound relative to some arbitrary reference intensity, such as a 1000-hertz pure tone, that is judged by the listeners to be equivalent in loudness.

equivalent network—One network that replaces another in a system without altering in any way the electrical operation of the system external to the network.

equivalent noise conductance—The spectral density of a noise-current generator expressed in conductance units at a specified frequency.

equivalent noise input—In a photosensitive device, the value of incident luminous flux that produces an rms output current equal to the rms noise curent within a specified bandwidth when the flux is modulated in a stated manner.

equivalent noise pressure—*See* transducer equivalent noise pressure.

equivalent noise resistance—A measure of the residual noise output of a potentiometer while the slider is being actuated. (The residual noise consists of active components in the form of self-generated voltages arising in the slider contact interface, and passive components in the form of ohmic contact resistance at the point of slider contact.)

equivalent noise temperature—The absolute temperature at which a perfect resistor with the same resistance as the component would generate the same noise as the component at room temperature.

equivalent open-circuit rms noise current—That noise which occurs at the input of the noiseless amplifier due only to noise currents. It is expressed in picoamperes per hertz at a specified frequency or in nanoamperes in a given frequency band.

equivalent periodic line—A periodic line that, when measured at its terminals or at corresponding section junctions, has the same electrical behavior at a given frequency as the uniform line with which it is compared.

equivalent permeability—The relative permeability that a component would have under specified conditions if it had the same reluctance as a component of the same shape and size but different materials.

equivalent plate voltage—The plate voltage plus mu times the grid voltage.

equivalent resistance—The concentrated or lumped resistance that would cause the same power loss as the actual small resistances distributed throughout a circuit.

equivalent series resistance—Abbreviated ESR or R. 1. In a circuit or component, the square root of the difference between the impedance squared and the reactance squared. All internal series resistance of the circuit or component treated as being concentrated in a single resistance at one point. 2. All internal ac series resistance of a capacitor treated as a single resistor. 3. An effective resistance that, if connected in series with an ideal capacitor of a capacitance value equal to that of the capacitor in question, would result in a power loss equal to the active power dissipated in that capacitor at a given frequency.

equivalent time—In random-sampling oscilloscope operation, the time scale associated with the display of signal events.

equivalent time sampling—A method of allowing the storage of repetitive events that occur faster than the maximum digital oscilloscope sampling frequency. The scope acquires the waveform during multiple sweeps by taking samples at various times until scope memory is filled.

equivocation—In a computer, the conditional information contained in an input symbol given an output symbol, averaged over all input-output pairs.

erasable programmable read-only memory—Abbreviated EPROM. A field-programmable read-only memory that can have the data content of each memory cell altered more than once. An EPROM is bulk-erased by exposure to high-intensity ultraviolet light. Sometimes referred to as a reprogrammable read-only memory.

erasable storage—Storage media in a computer that hold information that can be changed.

erase—1. To replace all the binary digits in a storage device by binary zeros. In a binary computer, erasing is equivalent to clearing; in a coded decimal computer, in which the pulse code for decimal zero may contain binary ones, clearing leaves decimal zero, whereas erasing leaves all-zero pulse codes in all storage locations. 2. To remove all information from a register or a memory. Sometimes this consists of writing a zero into all memory positions.

erase head—1. A head on a tape recorder that applies a strong high-frequency alternating magnetic field to the tape so that earlier recordings may be erased as the tape runs past the head. 2. A magnetic tape head that removes previously recorded signals from a tape, usually by applying inaudible high-level, high-frequency bias signals.

erasing speed—In charge-storage tubes, the rate of erasing successive storage elements.

erasure — 1. A process by which a signal recorded on a tape is removed and the tape made ready for rerecording. This may be accomplished by ac erasure, in which the tape is demagnetized by an alternating field that is reduced in amplitude from an initially high value, or by dc erasure, in which the tape is saturated by applying a primarily unidirectional field. 2. The neutralization of the magnetic pattern stored on tape.

E region — The region of the ionosphere about 50 to 100 miles (80 to 160 km) above the earth's surface.

E register — The extension of the computer A-register for use in double-precision arithmetic or logic-shift operations.

$E_r - E_y$ — The resultant color television signal when E_y is subtracted from the original full red signal.

erg — 1. The absolute centimeter-gram-second unit of energy and work. The work done when a force of 1 dyne is applied through a distance of 1 centimeter. 2. Measure of energy. 1 erg $= 10^7$ joules $= 6.25 \times 10^{10}$ eV.

ergonomics — The science of designing office systems to meet the needs of the human body.

ERP — Abbreviation for effective radiated power. The amount of power radiated by an antenna, which may be more or less than the power absorbed by it from the transmitter.

error — 1. In mathematics, the difference between the true value and a calculated or observed value. A quantity (equal in absolute magnitude to the error) added to a calculated or observed value to obtain the true value is called a correction. 2. Any discrepancy between a computed, observed, recorded, or measured quantity and the true, specified, or theoretically correct value or condition. 3. In a computer or data-processing system, any incorrect step, process, or result. In addition to the mathematical usage in the computer field, the term also commonly refers to machine malfunctions, or machine errors, and to human mistakes, or human errors. 4. In data communications, *error* means that a bit transmitted as a 1 was received as a 0 (or vice versa). 5. A deviation occurring during transmission such that a mark signal is received instead of a space signal, or vice versa. 6. A situation in which the data readout of a memory location is different from the data originally stored in that location. The term *error* is normally not applicable to circuits that do not contain memory elements. It can be applied, however, to complex logic arrays that contain memory elements, such as flip-flops or small arrays of RAM. Errors can be classified into three categories: hard error, medium error, and soft error.

error-correcting code — 1. A code in which each acceptable expression conforms to specific rules of construction that also define one or more equivalent nonacceptable expressions, so that if certain errors occur in an acceptable expression, the result will be one of its equivalents and thus the error can be corrected. 2. A computer code using extra bits that automatically detect and correct errors. 3. Code for information transmission that makes it possible to detect errors and to fix them. This involves sending extra information along with each word, with a corresponding reduction in the transmission rate.

error-correcting telegraph system — A system that employs an error-detecting code in such a way that any false signal initiates retransmission of the character incorrectly received.

error-correction routine — A series of computer instructions programmed to correct a detected error condition.

error detecting and feedback system — A system employing an error-detecting code and so arranged that a signal detected as being in error automatically initiates a request for retransmission of that signal.

error-detecting code — In a digital computer, a system of coding characters such that any single error produces a forbidden or impossible code combination.

error detection — 1. An arrangement that senses flaws in received data by examining parity bits, verifying block check characters, or using other techniques. 2. System that detects errors occasioned by transmission equipment or facilities.

error detector — That portion of an automatic control system that determines when the regulated quantity has deviated outside the dead zone.

error rate — 1. A measure of quality of a digital circuit or equipment item. 2. The number of erroneous bits or characters in a sample; it is frequently taken as the number of errors per 100,000 characters. 3. A measure of the ratio of the number of characters of a message incorrectly received to the number of characters of the message received. 4. The ratio of incorrectly received data (bits, elements, characters, or blocks) to the total amount of data transmitted.

error-rate damping — A damping method in which a signal proportional to the rate of change of error is added to the error signal for anticipatory purposes.

error signal — In an automatic control device, a signal whose magnitude and sign are used to correct the alignment between the controlling and the controlled elements.

error tape — A special tape developed and used for writing out errors in order to correct them by study and analysis after printing.

error voltage — A voltage that is present in a servo system when the input and output shafts are not in correspondence. The error voltage, which actuates the servo system, is proportional to the angular displacement between the two shafts.

ESA — Abbreviation for electronic shock absorption.

Esaki diode — *See* tunnel diode.

ESC — Abbreviation for electronic spark control. For General Motors turbo-charged engines, a system that controls timing and engine knock.

escape character — A code-extension character used, sometimes with one or more succeeding characters, to form an escape sequence, which indicates that the interpretation of the succeeding characters is to be different from the code currently in use.

E-scope — A radar display in which targets appear as blips with distance indicated by the horizontal coordinate and elevation by the vertical coordinate.

escutcheon — A backing plate around an opening. Commonly the ornamental metal, wood, plastic, or other framework around a radio tuning dial, control knob, or other panel-mounted part in a radio receiver or television receiver, audio-frequency amplifier, etc.

ESD — Abbreviation for electrostatic discharge. Discharge of a static charge on a surface or body through a conductive path to ground. An electronic component may suffer irreparable damage when it is included in the discharge path.

ESM — *See* electronic warfare support measures.

ESR — Abbreviation for equivalent series resistance.

ESS — 1. Abbreviation for electronic spark selection. An electronic system (General Motors) that controls timing on fuel-injected engines to increase fuel economy and improve emissions. Operates on inputs from engine temperature, rpm, and vacuum sensors. 2. Abbreviation for electronic switching system.

essential loads — Those loads that must be served to keep plant operations at an acceptable level during a prolonged commercial power outage. Such loads might be interruptible for periods of a few seconds to several minutes.

established reliability—A quantitative maximum failure rate demonstrated under controlled test conditions in accordance with a military specification; usually expressed as percent failures per thousand hours of test.

Estiatron—A special type of electrostatically focused traveling-wave tube.

etch—The process of removing material from a wafer (such as oxides or other thin films) by chemical, electrolytic, or plasma (ion bombardment) means. Examples: nitride etch, oxide etch.

etchant—1. A chemical agent that can remove a solid material. For example, a highly selective etchant that acts on silicon dioxide, silicon, or both is employed in semiconductor processing to pattern the silicon surface for diffusion masking. 2. A solution used, by chemical reaction, to remove the unwanted portion of a conductive material bonded to a base. 3. A liquid solution of a chemical or combination of chemicals used to preferentially dissolve metal.

etchback—The controlled removal of all components of base material by a chemical process on the side wall of holes in order to expose additional internal conductor areas.

etched metal mask—A metal mask used for screening wherein the pattern is created in a sheet of metal by the etching process.

etched printed circuit—A type of printed circuit formed by chemically or electrolytically (or both) removing the unwanted portion of a layer of material bonded to a base.

etch factor—A ratio of etched depth to the lateral etch at the boundary of the interface of the photoresist and the substrate.

etching—1. The selective removal of unwanted material from a surface, usually by chemical means. 2. A process using either acids or a gas plasma to remove unwanted material from the surface of a wafer.

etching resist—Material deposited on the surface of a copper-clad base material that prevents the removal by etching of the conductive areas the material covers.

etching to frequency—Finishing a crystal blank to its final frequency by etching it in hydrofluoric acid.

ether—A hypothetical medium that pervades all space (including vacuum) and all matter; assumed to be the vehicle for propagation of electromagnetic radiations.

Ethernet—1. A high-speed network connection. 2. A network standard first developed by Xerox, and refined by DEC and Intel. Ethernet interconnects personal computers and transmits at 10 megabits per second. It uses a bus topology that can connect up to 1024 PCs and workstations within each main branch, and uses a protocol known as carrier sense multiple access with collision detection (CSMA/CD) to regulate communication line traffic. Nodes are linked using coaxial cable, fiber-optic cable, or twisted-pair wiring. Ethernet is codified as the IEEE 802.3 standard. 3. A very common method of networking computers in a LAN. Ethernet will handle about 10 megabits per second and can be used with almost any kind of computer.

E-transformer—A special form of differential transformer employing an E-shaped core. The secondaries of the transformer are wound on the outer legs of the E, and the primary is on the center leg.

ETSI—*See* European Telecommunications Standards Institute.

Ettingshausen effect—Analogous to the Hall effect. The different temperatures found on opposite edges of a metal strip that is perpendicular to a magnetic field and through which an electric current flows longitudinally.

eTV—Abbreviation for educational TV. A nonprofit television station operating to serve community needs in the areas of instruction and cultural development.

Eureka—A ground transponder in the British Rebecca-Eureka radar navigational system.

Euro connector—*See* SCART

European Telecommunications Standards Institute—Abbreviated ETSI. One of the European organizations responsible for establishing common industry-wide standards for telecommunications.

eutectic—1. An isothermal reversible reaction in which a liquid solution is converted into two or more intimately mixed solids on cooling, the number of solids formed being the same as the number of components in the system. 2. An alloy having the composition indicated by the eutectic point on an equilibrium diagram. 3. An alloy structure of intermixed solid constituents formed by a eutectic reaction. 4. Referring to an alloy or solid solution that has the lowest possible melting point, usually below that of its components. 5. The most fusible series of alloys (e.g., 63/37 tin-lead solder). 6. The specific proportions of the constituents of an alloy having the lowest melting point. The system goes from totally molten to totally solid without going through a slushy range at the eutectic composition.

eutectic alloy—1. A combination of two or more metals that has a sharply defined melting point and no plastic range. 2. An alloy with a low and sharp melting point that converts from a solid to a liquid state at a specific recurring point. Used in thermal overload devices.

eutectic bonding—Formation of a metallurgical joint in similar or dissimilar metals through the introduction of a thin film of another metal at the joint interface. Upon application of heat and moderate pressure, the intermediate film and the metals to be joined form a molten eutectic phase, which is then eliminated from the joint by thermal diffusion into the base metals.

eutectic solder—Solder that has the lowest possible melting point for its combination of elements. Eutectic tin-lead solder is composed of 63 percent tin and 37 percent lead. It melts at 361°F. Eutectic tin-silver solder has 96.5 percent tin with 3.5 percent silver, and melts at 430°F.

eV or **ev**—Abbreviation for electron-volt.

evaporation—The deposition in high vacuum of insulation that is thermally liberated from a parent source. Silica films of low optical absorption have been produced by electron bombardment of the parent oxide.

evaporation materials—Metals used for evaporation charges and sputtering targets, including chromium

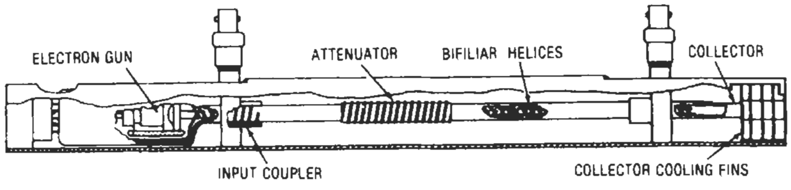

Estiatron.

and its alloys, for (*a*) a thin adhesive layer on IC substrates to allow better deposition of gold or other metal, (*b*) resistor material, and (*c*) vacuum deposition in mask production.

evaporation of electrons—The cooling that occurs on the surface of a cathode during emission. It is analogous to the cooling of a liquid or solid as it evaporates.

evaporation sources—Boats and filaments used as heat sources for vacuum evaporation to form thin layers on substrates. The process is frequently done by resistively heating the evaporant in a ceramic crucible or by self-heating of boats constructed of tungsten, molybdenum, or tantalum.

evaporative deposition—The process of condensing a thin film of evaporated material upon a substrate. Evaporation usually is produced by heating a material in a high vacuum.

E-vector—A vector representing the electric field of an electromagnetic wave. In free space it is perpendicular to the direction of propagation.

even harmonic—Any harmonic that is an even multiple (2, 4, 6, etc.) of the fundamental frequency. The even harmonics of 60 Hz are 120 Hz, 240 Hz, 360 Hz, etc.

even parity—1. Parity bit that is added to a word so that the total number of 1s is even. 2. The condition that occurs when the sum of the number of 1s in a binary word is always even. 3. A dumb-terminal data verification method in which each character must have an even number of 1 bits.

event counter—Instrument that records and totalizes occurring events; can include time of occurrence of events.

event flag—In a computer, an easily implemented synchronization mechanism that can be used for passing messages and data buffers between two cooperating tasks.

EW—Abbreviation for electronic warfare.

E-wave—Designation for TM (transverse magnetic) wave, one of the two classes of electromagnetic waves that can be sent through waveguides.

EX—In a calculator, it is the abbreviation for exchange key. Interchanges the last entry with the preceding value in the calculator.

exalted-carrier receiver—A receiver that counteracts selective fading by maintaining the carrier at a high level at all times.

exalted-carrier reception—A method of receiving either amplitude- or phase-modulated signals in which the carrier is separated from the sidebands, filtered and amplified, and then recombined with the sidebands at a higher level prior to demodulation.

except gate—A gate in which the specified combination of pulses producing an output pulse is the presence of a pulse on one or more input lines and the absence of a pulse on one or more other input lines.

exception—In a computer, a condition that is out of the ordinary in normal task execution; e.g., arithmetic overflow.

excess carriers—Any carriers present in a semiconductor material or region in addition to those present in equilibrium.

excess conduction—Conduction by excess electrons in a semiconductor.

excess electron—An electron introduced into a semiconductor by a donor impurity and available to promote conduction. An excess electron is not required to complete the bond structure of the semiconductor.

excess fifty—In a computer, a representation in which a number N is denoted by the equivalent of $(N + 50)$.

excess meter—An electricity meter that measures and registers the integral, with respect to time, of those portions of the active power in excess of the predetermined value.

excess minority carriers—In a semiconductor, the number of minority carriers that exceed the normal equilibrium number.

excess modified index of refraction—*See* refractive modulus.

excess noise—Also called current noise, bulk noise, and $1/f$ noise. Noise resulting from the passage of current through a semiconductor material.

excess sound pressure—The total instantaneous pressure at a point in a medium containing sound waves, minus the static pressure when no sound waves are present. The unit is the dyne per square centimeter.

excess-three BCD—Abbreviation for excess-three binary-coded decimal. Pertaining to a code based on adding 3 to a decimal digit and then converting the result directly to binary form. Use of this code simplifies the execution of certain mathematical operations in a binary computer that must handle decimal numbers.

exchange—To remove the contents of one storage unit of a computer and place it in a second, at the same time placing the contents of the second storage unit into the first.

exchange cable—A lead-covered, nonquadded, paper-insulated cable used in providing cable pairs between local subscribers and a central office.

exchange code—The three digits following an area code in a telephone number.

exchange key—*See* EX.

exchange line—A line that joins a subscriber or switchboard to a commercial exchange.

exchange plant—Facilities used to serve the needs of subscribers, as distinguished from facilities used for long-distance communication.

exchange register—*See* memory register.

exciplex—From *exc*ited state com*plex*. A chemical reaction occurring in certain lasing materials known as organic dyes and used for adjusting a laser so that it emits light in a color range from near ultraviolet to yellow.

excitation—1. Also called stimulus. An external force or other input applied to a system to cause it to respond in some specified way. 2. Also called drive. A signal voltage applied to the control electrode of an electron tube. 3. In electric or electromagnetic equipment, supplying with a potential, a charge, or a magnetic field. 4. The addition of energy to a system so as to transfer the system from its ground state to an excited state. 5. That energy which is present in a crystalline material as a result of its dynamic interaction with the external environment. This includes the energy acquired by the material in the form of sound, heat, light, and other forms of radiation. 6. Current supplied to energize the field coils of a generator. 7. The signal applied to a power amplifier. 8. The signal applied to a transmitting antenna.

excitation anode—An auxiliary anode of a pool-cathode tube, used to maintain a cathode spot when the output current is zero.

excitation current—The resultant current in the shunt field of a motor when voltage is applied across the field.

excitation energy—The external electrical energy required for proper operation of a transducer.

excitation purity—Also called purity. The ratio between the distance from the reference point to the point representing the sample and the distance along the same straight line from the reference point to the spectrum locus or to the purple boundary, both distances being measured (in the same direction from the reference point) on the CIE chromaticity diagram.

excitation voltage—The voltage required for excitation of a circuit.

excited-field speaker—A speaker in which the steady magnetic field is produced by an electromagnet.

exciter—1. In a directional transmitting antenna system, the part connected directly to the source of power, such as to the transmitter. 2. A crystal or self-excited oscillator that generates the carrier frequency of a transmitter. 3. A small, auxiliary generator that provides field current for an ac generator.

exciter lamp—1. A high-intensity incandescent lamp having a concentrated filament. It is used in making variable-area, sound-on-film recording and in reproducing all types of sound tracks on film, as well as in some mechanical television systems. 2. A light source used in a facsimile transmitter to illuminate the subject copy being scanned. 3. Small incandescent lamp whose intense beam is focused on the optical sound track of a motion picture film. The sound track modulates the beam, which in turn is detected by a photocell that produces an electrical output that is eventually converted into audible sounds.

exciter or dc-generator relay—A device that forces the dc machine-field excitation to build up during starting, or which functions when the machine voltage has built up to a given value.

exciter response—In rotating electrical machinery, the rate of change of the main exciter voltage when the resistance in the main exciter field circuit is suddenly changed. The exciter response may be expressed in volts per second or by the numerical value obtained by dividing the volts per second by some designated value of voltage, such as the nominal collector-ring voltage.

exciting current—Also called magnetizing current. 1. The current that flows in the primary of a transformer when the secondary is open-circuited. This current produces a flux that generates a back emf equal to the applied voltage. 2. The current that passes through the field windings of a generator.

exciton—A mobile, electrically neutral, excited state of holes and electrons in a crystal. An example is a weakly bound electron-hole pair; when such a pair recombines, the energy yielded is the bandgap reduced by the binding energy of the pair.

excitron—A type of rectifier tube used in applications with heavy power requirements and in power distribution systems. It has a single anode and a mercury-pool cathode and is provided with a means for maintaining a continuous cathode spot.

exclusion principle—The principle that states that if particles are considered to occupy quantum states, then only one particle of a given kind can occupy any one state. Particles differ in kind due to their direction of spin momentum, orbit, etc.

exclusive OR—A function that is valid (its value is 1) if one and only one of the input variables is valid. The exclusive OR applied to two variables is valid, or 1, if the binary inputs are different. The term *half-add* is often applied to the exclusive OR with two input variables.

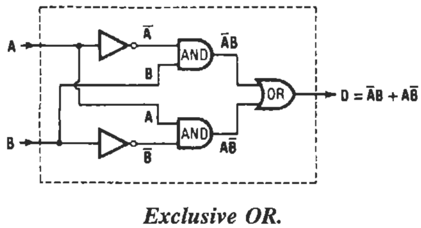

Exclusive OR.

exclusive OR function—A logic operation in which the result is logically true when only one input function is true and false when both inputs are true or false.

exclusive OR gate—1. A type of gate that produces an output when the inputs are the same, but not when they are different. 2. An electronic circuit whose output is a logical 1 if its two inputs are different, and a logical 0 if they are the same. This follows the rules for binary addition if carries are disregarded. For this reason, the exclusive OR gate is sometimes known as a half-adder.

excursion—A single movement away from the mean position in an oscillating or alternating motion.

execute—1. The process of interpreting an instruction and performing the indicated operation (s). 2. To perform a specified computer instruction. To run a program. 3. The third cycle of the three cycles for program instruction execution. During this time, the operation instruction is performed. *See also* decode; fetch.

execution—1. The performance of an operation or instruction. 2. The performance of a specific operation such as would be accomplished through processing one instruction, a series of instructions, or a complete program.

execution time—1. The time required for a computer to execute an instruction, usually several machine cycles. 2. The total time required for the execution of one specific operation, including fetch, decode, and execute steps.

executive instruction—Instruction to determine how a specially written computer program is to operate.

executive routine—In computer operation, a set of coded instructions that controls loading and relocation of other routines, and in some cases employs instructions not known to the general programmer. Effectively, an executive routine is a nonhardware part of the computer itself, except that it is superimposed on all lower-level programs and instructional sets.

exercisers—1. Multioutput data generators; e.g., a memory exerciser produces, at different sets of output terminals, data-input words, addresses, and coincidence/complement signals for an error detector (discriminator), as well as appropriate write and read commands. These instruments are actually small test systems, and might properly be classed with them. 2. A test system or program for a device, such as memory, disk, or tape, designed to detect malfunctions prior to use.

exhaustion—The removal of gases from a space, such as the bulb of a vacuum tube, by means of vacuum pumps.

exit—In a computer, the means of halting a repeated cycle of operation in a program.

exitance—The flux per unit area leaving (diverging) from a source of finite area.

exit angle—The exit angle is the angle between the output radiation vector and the axis of the fiber or fiber bundle.

exit delay—The time between turning on a control unit and the sounding of a local alarm or transmission of an alarm signal on actuation of a sensor on an exit door. This delay is used if the authorized access switch is located within the protected area and permits a person with the control key to turn on the alarm system and to leave through a protected door or gate without causing an alarm. The delay is provided by a timer within the control unit.

exogeneous variable—A variable whose values are determined by considerations outside the model in question. *See also* endogenous variable.

exosphere—The outermost region of the earth's atmosphere, where the atoms and molecules move in dynamic orbits under the influence of the gravitational field.

exothermic—A chemical reaction in which heat is produced.

exothermic reaction—A reaction in which heat is given off.

expand—To spread out part or all of the trace of a cathode-ray display.

expandable breadboarding system—*See* microcomputer development system.

expandable gate—A logic gate whose number of inputs can be increased by the simple addition of an expander block.

expanded contact—In a semiconductor, any pattern that has metallization crossing a diffused junction.

expanded-position-indicator display—An expanded display of a sector from a plan-position-indicator presentation.

expanded-scale meter—1. A meter in which the ratio of deflection per unit of applied energy becomes greater as the energy approaches a specified value. 2. A meter that confines its measurements to a selected narrow range, which occupies the entire scale of the meter. Lesser values are ignored, and no deflection is exhibited until the minimum scale value is reached or exceeded.

expanded scope—A magnified portion of a cathode-ray-tube presentation.

expanded sweep—A preselected portion of a sweep, during which time the electron beam is speeded up in a cathode-ray tube.

expander—1. A transducer that produces a larger range of output voltages for a given amplitude range of input voltages. One important type expands the volume range of speech signals by employing their envelope. 2. A logic block that may be connected easily to an expandable gate to make a larger number of logic inputs available. 3. The part of a compander used at the receiving end of a circuit to restore the compressed signal to its original form. It reduces the amplitude of weak signals and amplifies strong signals. 4. A device used to restore natural dynamic range by counteracting the compression of dynamic range used in the making of recordings and in broadcasting.

expander inputs—Gates used for increasing the number of logic-performing inputs.

expandor—A device that reverses the effect of analog compression. *See also* compander.

expansion—1. A process in which the effective gain applied to a signal is increased for larger signals and decreased for smaller signals. 2. In facsimile transmission, an increase in the contrast between light and dark portions of the transmitted picture.

expansion board—1. A circuit board designed to fit the bus of a computer in order to add additional ports, memory, or functions. 2. A PC board added to a computer and giving it added capabilities, such as increasing the amount of main memory.

expansion slot—A socket on the main board of a personal computer where an accessory card, or expansion board, can be inserted.

experimental model—An equipment model that demonstrates the technical soundness of a basic idea but does not necessarily have the same form or parts as the final design.

experimental station—Any station (except amateur) utilizing electromagnetic waves between 10 kHz and 3000 GHz in experiments, with a view toward the development of a science or technique.

exploring coil—*See* magnetic test coil.

explosion-proof motor—A motor designed and constructed so as to withstand an internal explosion of a specified gas or vapor and to prevent the ignition of the specified gas or vapor surrounding the motor by sparks, flashes, or explosions of the specified gas or vapor inside the motor casing.

explosion-proof switch—A switch (UL listed) that can withstand an internal explosion of a specified gas without causing ignition of surrounding gases.

explosive atmosphere—The condition in which air is mixed with dust, metal particles, or inflammable gas in such proportion that it is capable of igniting or exploding.

exponential—Pertaining to exponents or to an expression having exponents. A quantity that varies in an exponential manner increases by the square or some other power of a factor, instead of linearly.

exponential curve—A curve representing the variation of an exponential function.

exponential damping—Damping that follows an exponential law.

exponential decay—The decay of signal strength, radiation, charge, or some other quantity at an exponential rate.

exponential horn—1. A horn whose cross-sectional area increases exponentially with axial distance. 2. Speaker horn (low or high frequency) in which the flare rate of the horn follows an exponential curve.

exponential quantity—A single quantity that increases or decreases at the same rate as the quantity itself (e.g., the discharge current of a capacitor through a noninductive resistor).

exponential sweep—An electron-beam sweep that starts rapidly and slows down exponentially.

exponential transmission line—A two-conductor transmission line whose characteristic impedances vary exponentially with the electrical length of the line.

exponential waveform—A waveform that is characterized by smooth curves but which possesses pulse properties because it contains numerous constituent frequencies. The exponential waveform undergoes a rate of amplitude change that is either inversely or directly proportional to the instantaneous amplitude.

exposure—1. A measure of the X-ray or gamma radiation at any point; it is used to describe the energy of the radiation field outside the body. 2. The act of subjecting photosensitive surfaces or matter to radiant energy, such as light, to produce an image.

exposure meter—In photography, an instrument that measures scene brightness and indicates proper lens opening and exposure time.

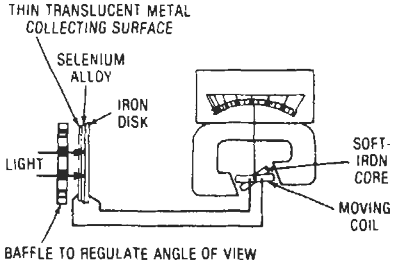
THIN TRANSLUCENT METAL COLLECTING SURFACE
SELENIUM ALLOY
IRON DISK
LIGHT
SOFT-IRON CORE
MOVING COIL
BAFFLE TO REGULATE ANGLE OF VIEW

Exposure meter.

expression—A valid series of constants, variables, and functions that can be connected by operation symbols and punctuated, if necessary, to describe a desired computation.

expression control — In an organ, the control that regulates the overall volume. Usually operated with the right foot.

extended addressing — An addressing mode in a computer that can reach any place in memory. *See also* direct addressing.

extended-area service — Telephone exchange service, without toll charges, that extends over an area where there is a community of interest, in return for a somewhat higher exchange service rate.

extended-cutoff tube — *See* remote-cutoff tube.

extended foil capacitor — *See* noninductive capacitor.

extended-foil construction — A method of fabricating capacitors in which two foil electrodes, separated by a dielectric, are offset so that they may be wound with one foil extending from one end of the winding and the other foil extending from the opposite end.

extended-interaction tube — A microwave tube in which a moving electron stream and a traveling electric field interact in a long resonator. The bandwidth of such a tube is between that of a klystron and a traveling-wave tube.

extended octaves — In an organ, tones above or below the notes on the regular keyboard that can be sounded only when certain couplers are on.

extended play — Abbreviated EP. A 45-rpm record on a 7-inch disc. It provides 8 minutes of playing time, instead of the 5 minutes of a standard 45-rpm disc.

extended-play tape — Recording tape made with thinner base material so that more tape can be spooled onto a given size reel. Generally refers to 1-mil (25.4-µm) or 0.5-mil (12.75-µm) thick tapes, as opposed to standard 1.5-mil (38.1-µm) tape. Extended-play tapes are more susceptible to stretching, breaking, and print-through than standard tapes.

extended speed range — An extension of the basic speed range of a motor by means of field voltage control. When operating in this range, the load must be reduced to maintain a constant horsepower output to prevent overheating of the motor.

extension — 1. A small computer program that can be plugged into a larger one, usually to enhance the main program's capabilities. All operating systems use extensions to enable them to read compact discs. 2. Filenames often end with a period followed by additional characters known as the file extension. An extension is generally a standard abbreviation for a type of file. For example, .text is often used for ASCII files, and .ps for PostScript files. 3. An additional telephone bridged on the same line with the main phone.

extension number — The PBX or Centrex telephone number associated with a particular caller.

extent — The physical positions on input-output devices occupied by or reserved for a particular data set.

external armature — A ring-shaped armature that rotates around the field magnets of a generator or motor.

external circuit — All wires and other conductors that are outside the source.

external control devices — All control devices mounted external to the control panel.

external critical damping resistance — The value of resistance that must be placed in series or in parallel with a galvanometer in order to produce the critically damped condition.

external device — 1. An input or output device, such as a paper-tape reader, line printer, or magnetic tape recorder, under control of a computer. 2. A unit of processing equipment in a computer system external to the CPU.

external feedback — In a magnetic amplifier, the ampere-turns of auxiliary windings that assist in equalizing the load-winding ampere-turns, thereby reducing the control ampere-turns required for control. External feedback may be degenerative to reduce the gain and improve the stability of an amplifier.

externally adjustable timer — A time-delay unit that can be adjusted by varying the resistance of the timer resistor externally, usually by means of a screwdriver.

externally caused contact chatter — That chatter resulting from shock or vibration imposed on the relay by external action.

externally caused failure — A failure caused by an environment outside the design limitations, such as excessive loads, voltages, etc., resulting from operator error, accident, or failure of another part.

externally quenched counter tube — A radiation counter tube that requires an external circuit to prevent it from reigniting.

external memory — An auxiliary storage unit apart from the internal memory of a computer (e.g., magnetic tape).

external photoeffect — The emission of photon-excited electrons from the surface of a material after overcoming the energy barrier at the surface of a photoemissive material.

external photoeffect detector — A photodetector in which the energy of each photon incident on the detector surface is sufficient to liberate one or more electrons; i.e., Planck's constant times the frequency, which is the energy of the photon, is sufficient to overcome the work function of the material, and the liberated electrons move under the influence of an applied electric field. Photoemissive devices make use of the external photoeffect.

external photoelectric effect — The ejection of electrons from the surface of a solid by the absorption of a sufficient amount of photons.

external processor loop — *See* tape monitor.

external Q — In a microwave tube, the reciprocal of the difference between the reciprocal of the loaded Q and the reciprocal of the unloaded Q.

external storage — Storage facilities separate from the computer itself but holding information in a form acceptable to the computer, e.g., magnetic tapes, optical discs, etc.

external wiring — Category of electronic wiring that interconnects various subsystems within the weapons system. It is frequently subjected to severe environments.

extinction potential — The lowest value to which the plate voltage of a gaseous tube can be reduced without cutting off the flow of plate current.

extinguishing voltage — 1. The voltage across a glow lamp when the lamp ceases to glow. 2. The lowest anode voltage at which a gas tube can sustain a discharge.

extraband spurious transmitter output — A spurious transmitter output that lies outside the specified band of transmission.

extra-class license — The highest classification of United States amateur license. Requirements include a code sending and receiving ability of 20 words per minute, a knowledge of advanced theory, and the holding of a general- or conditional-class license for two years.

extract — 1. To remove from a set of items of information all those items that meet some arbitrary criterion. 2. In computer operations, to obtain specific digits from a stored word. 3. To form a new word from selected segments of given words.

extract instruction—In a digital computer, the instruction to form a new word by placing selected segments of given words side by side.

extractor—*See* filter, 3.

extra-high performance macros—Computers that have a throughput of 160 gigabits per second, and a memory size of 256 megabytes.

extramural absorption—The absorption of light, transmitted radically through the cladding of an optical fiber, by means of a dark or opaque coating placed over the cladding.

extramural cladding—A layer of dark or opaque absorbing coating placed over the cladding of an optical fiber to increase internal reflection, protect the smooth reflecting wall of the cladding, and absorb scattered or escaped stray light that might penetrate the cladding.

extraneous emission—Any emission of a transmitter or transponder other than the carrier and those sidebands intentionally added to convey intelligence.

extraneous response—Any undesired response of a receiver, recorder, or other susceptible device due to the desired signals, undesired signals, or any combination or interaction among them.

extraordinary wave—One of two components into which a sky wave is split in the ionosphere. When viewed below the ionosphere in the direction of propagation, it has clockwise or counterclockwise elliptical polarization, depending on whether the magnetic field of the earth has a positive or negative component in the same direction. The extraordinary wave is designated by the letter X and is called the X-wave. The other component is the ordinary wave, or O-wave.

extra play—Tape recording term. Originally, all recording tapes were 1.5-mils (38.1-μm) thick, and a 7-inch (17.8-cm) reel would accommodate about 1200 feet (365 m) of this tape, for a half-hour of continuous recording at 7½ ips. The later extra-play tapes are 1-mil (25.4-μm) thick and allow for 45 minutes of recording, or 15 minutes extra.

extrapolate—1. To estimate the value of a function for variables lying outside the range in which values of the function are known (e.g., to extend the graph of the function beyond the plotted points). 2. To estimate the values of a function that are less than or greater than those already known.

extrapolation—Estimating the future value of some data series based on past observations. Statistical forecasting represents a common example.

extraterrestrial noise—Radio disturbances originating from sources other than those related to the earth; cosmic or solar noise.

extremely high frequency—Abbreviated EHF. 1. The frequency band extending from 30 to 300 GHz. 2. Any of the radio frequencies in the band 30,000 to 300,000 MHz.

extremely low frequency—Abbreviated ELF. A frequency below 300 hertz.

extrinsic base resistance—The resistance between the base terminal of a transistor and the internal inaccessible base point in an equivalent circuit.

extrinsic base resistance-collector capacitance product—The product of the base resistance and collector capacitance of a transistor. It is expressed in units of time, since it is an *RC* time constant, and affects the high-frequency operation of a transistor.

extrinsic conductance—The conductance resulting from impurities or external causes.

extrinsic detector—A photodetector composed of semiconductor material whose responsive properties can be altered by the addition of impurities to the basic material. Copper and mercury-doped germanium are both examples of this semiconductor material.

extrinsic photoconductivity—Photoconductivity due to the addition of impurities or external causes.

extrinsic properties—The properties exhibited by a semiconductor as the result of its modification by imperfections and impurities in the crystal.

extrinsic semiconductor—1. A semiconductor with charge carrier concentration dependent on impurities or other imperfections. 2. The resulting semiconductor produced when impurities are introduced into an otherwise nonsemiconductor crystal. The electrical properties depend on the impurities. 3. A semiconductor whose electrical properties are dependent on impurities.

extrinsic semiconductor material—A semiconductor material with charge carrier concentration dependent on impurities or other imperfections.

extrinsic transconductance—The quotient of a small change in collector current divided by the small change in emitter-to-base voltage producing it, under the condition that other voltages remain unchanged. Thus, if an emitter-to-base voltage change of 0.1 volt causes a collector-current change of 3 milliamperes (0.003 ampere), with other voltages constant, the transconductance is 0.003 divided by 0.1, or 0.03 siemens. For convenience, a millionth of a siemens (which was formerly termed mho), or a microsiemens, is used to express transconductance. Thus, in the example, 0.03 siemens is 30,000 microsiemens.

extruded cables—Cables with conductors that are insulated and formed in a uniform configuration by the application of a homogeneous insulation material in a continuous extrusion process.

extrusion—A method of forcing plastic, rubber, or elastomer material through an orifice in a more or less continuous fashion to apply insulation or jacketing to a conductor or cable.

eyelet—A tubular metal piece having one end (and possibly a second) headed or rolled over at a right angle.

E zone—In the making of frequency predictions, one of three zones into which the earth is divided to show the variations of the F_2 layer with respect to longitude. This zone includes Asia, Australia, the Philippines, and Japan.

F

f—1. Letter symbol for femto (10^{-15}). 2. Symbol for focal length, frequency.

F—1. Symbol for filament, fuse. 2. Abbreviation for Fahrenheit. 3. Letter symbol for farad.

F– (F-minus or **F-negative)**—*See* A– (A-minus or A-negative).

F+ (F-plus or **F-positive)**—*See* A+ (A-plus or A-positive).

fA—Letter symbol for femtoampere (10^{-15} ampere).

fabrication holes—*See* pilot holes.

fabrication tolerance—In the construction or assembly of an equipment or portion of an equipment, the maximum variation in the characteristics of a part that, considering the defined variations of the other parts in the equipment, will permit the equipment to operate within specified performance limits.

Fabry-Perot interferometer—A resonant cavity bounded by two end mirrors separated so as to produce interference for certain allowed optical frequencies. The enhancements and cancellations produced by the internal reflections are, in fact, optical standing waves. The structure can be used as a laser cavity, mode filter, or frequency selector.

face—1. A plane surface on a crystal that stands in a particular and invariable relation to the axes and planes of reference and to other faces. 2. Front, or viewing, surface of a cathode-ray tube. 3. The portion of a meter bearing the scale markings.

face bonding—1. A method of attaching active devices to thin-film passive networks. The semiconductor chips are provided with small mounting pads turned face down and bonded directly to the end of the thin-film conductors on the passive substrate. The term includes ultrasonic, solder-reflow, and solder-ball techniques. 2. Process of bonding a semiconductor chip so that its circuitry side faces the substrate. Flip-chip and beam-lead bonding are two common methods. (Opposite of back bonding.)

faced crystal—A single or twinned mass of quartz bonded in part or entirely by the original growth faces.

face-parallel cut—A Y-cut for a quartz crystal.

face-perpendicular cut—An X-cut for a quartz crystal.

facility—1. A transmission path between two or more locations without terminating or signaling equipment. 2. Anything used or available for use in providing communication service; a communications path.

facom—A long-distance measuring or radionavigational system that derives information of distances by comparing the phases of received and locally generated signals. It is a base-line system operating in the low-frequency band and will work under adverse propagation and noise conditions at ranges of up to 3000 miles (4827 km) from the signal source.

facsimile—1. A process or the result of a process by which fixed graphic material, including pictures or images, is scanned and the information converted into electrical signal waves, which are used either locally or remotely to produce in record form a likeness (facsimile) of the subject copy. 2. Electronic transmission of photographs or documents over a telephone channel. 3. Technology that allows a paper message to be scanned optically, translated into digitally encoded pixels, and sent across the public telephone network to a receiving fascimile (fax) machine, or computer which then reconstructs the original image.

facsimile broadcast station—A station licensed to transmit images of still objects for reception by the general public.

facsimile receiver—An instrument designed to receive facsimile transmissions and translate them into a reproduced image.

facsimile recorder—Apparatus that reproduces on paper the image transmitted by a facsimile system.

facsimile-signal level—An expression of the maximum signal power or voltage created by a scanning of the subject copy as measured at any point in a facsimile system. According to whether the system employs positive or negative modulation, this will correspond to picture white or black, respectively. It may be expressed in decibels with respect to some standard value, such as 1 milliwatt or 1 volt.

facsimile system—An integrated assembly of the elements used for facsimile transmission and reception.

facsimile transmitter—An apparatus employed to convert the subject copy into suitable facsimile signals.

factory calibration—The tuning or altering of a circuit by the manufacturer to bring the circuit into specification. Normally stated as a percent deviation.

fade—1. The gradual lowering in amplitude of a signal. 2. A gradual change of signal strength. 3. To change the strength of a signal gradually. 4. Reducing signal level until it is largely attenuated or completely inaudible (faded away). Short fades occur almost immediately; long fades take from 5 to 30 seconds. Fades are most often done at the end of a recording.

fade chart—A graph of the null areas of an air-search radar antenna; it is used as an aid in estimating target altitude.

fade in—To increase the signal strength gradually in a sound or television channel.

fade out—1. The gradual decrease in signal strength in a sound or television channel. 2. The cessation or near cessation of radio-wave propagation through parts of the ionosphere due to a sudden atmospheric disturbance.

fader—1. A multiple-unit control used in radio for gradual changeover from one microphone or audio channel to another; in television, from one camera to another; and in motion-picture projection, from one projector to another. 2. A control or group of controls for effecting fade in and fade out of video or audio signals.

fading — 1. A drift in the level of received radio signals beyond intelligibility. It is often caused by changes in the upper atmosphere. 2. Variations in intensity of some or all components of a received (radio) signal due to changes in the propagation path. 3. Deliberate slow reduction of signal level by means of the volume control.

fading margin — The number of decibels of attenuation that can be added to a specified radio-frequency propagation path before the signal-to-noise ratio of the channel falls below a specified minimum.

Fahnestock clip — A spring-type terminal to which a temporary connection can readily be made.

Fahrenheit temperature scale — A temperature scale in which the freezing point of water is defined as 32° and the boiling point as 212° under normal atmospheric pressure (760 mm of mercury).

fail hardover — Failure that results in a steady-state maximum system output with no input signal.

fail-safe — 1. Describing a circuit or device that fails in such a way as to maintain circuit continuity or prevent damage. 2. A feature of a system or device that initiates an alarm or trouble signal when the system or device either malfunctions or loses power. 3. A control system that either continues to function safely after a failure or lapses into a predefined condition known to be safe.

fail-safe circuit — A circuit that has an output state that indicates that either a circuit input or the circuit itself has failed. Finds circuit application in complex systems where self-healing subsystems exist. When a subsystem failure is detected, a backup subsystem is automatically inserted.

fail-safe control — A system of remote control for preventing improper operation of the controlled function in event of circuit failure.

fail-safe operation — An electrical system so designed that the failure of any component in the system will prevent unsafe operation of the controlled equipment.

fail soft — 1. An active failure that can be compensated for by the system, or one in which the operator can safely assume control. 2. A method of system implementation that prevents irrecoverable loss of computer usage due to failure of any system resource. It provides for graceful degradation of service.

fail-soft system — A system in a computer that continues to process data despite the failure of parts of the system. Usually accompanied by a deterioration in performance.

failure — 1. The inability of a system, subsystem, component, or part to function in the required or specified manner. 2. The termination of a capability of an item to perform its required function.

failure-activating cause — A stress or force (e.g., shock or vibration) that induces or activates a failure mechanism.

failure analysis — 1. Examination of electronic parts to determine the cause of performance variations outside previously established limits, for the purpose of identifying failure modes and failure-activating causes. 2. The logical, systematic examination of an item or its diagram(s) to identify and analyze the probability, causes, and consequences of potential and real failures.

failure indicator — The observed characteristic that shows that an item is defective.

failure mechanism — 1. A structural or chemical defect, such as corrosion, a poor bond, or surface inversion, that causes failure. 2. The process of degradation or chain of events that results in a particular failure mode. 3. The basic chemical or physical change that results in a catastrophic, degradation, or intermittent failure. 4. Underlying physical or chemical phenomenon that causes a failure. *See* failure mode, 1.

failure mode — 1. Also called failure mechanism. The manner in which a failure occurs, including the operating condition of the equipment or part at the time of the failure. 2. An electrical parameter that is sensitive to degradation and therefore is useful in the identification of failure mechanisms. 3. A catastrophic, degradation, or intermittent failure, usually in the form of opens, shorts, or parameters out of specification.

failure rate — Also called hazard. 1. The average proportion of units failing per unit of time, normally expressed in percent per thousand hours. It is used in assessing the life expectancy of a device. To be meaningful, a statement of the failure rate must be accompanied by complete information regarding the testing conditions, failure criteria, parameters monitored, and confidence level. 2. The statistical probability of the incidence of catastrophic failure in a large number of identical components operated continuously for a given period under stated environmental conditions. Usually expressed as maximum percentage of units predicted to fail under continuous service at maximum rated power (or voltage, if that limit is reached first) at a stated ambient temperature. 3. Expected number of failures in a given time interval under specified conditions.

failure unit — One failure in 10^9 device operating hours.

fall-in — In a synchronous motor, the point at which synchronous speed is reached.

fallouts — *See* transistor seconds.

fall time — 1. The length of time during which a pulse is decreasing from 90 to 10 percent of its maximum amplitude. 2. A measure of time required for a circuit to change its output from a high level (1) to a low level (0). 3. The time required for the pointer of an electrical indicating instrument to move from a steady full-scale deflection to 0.1 (± specified tolerance) of full scale when the instrument is short circuited. 4. The time interval between the instants at which the magnitude of the pulse at the output terminals of a switching transistor reaches specified upper and lower limits, respectively, when the transistor is being switched from its conducting to its nonconducting state. The upper and lower limits are usually 90 and 10 percent, respectively, of the amplitude of the output pulse.

false add — To form a partial sum; that is, to perform an addition without carries.

false alarm — 1. A radar indication of a detected target even though one does not exist. It is caused by noise or interference levels that exceed the set threshold of detection. 2. An alarm signal transmitted in the absence of an alarm condition. These may be classified according to causes: environmental, e.g., rain, fog, wind, hail, lightning, temperature; animals, e.g., rats, dogs, cats, insects; human-made disturbances, e.g., sonic booms, electromagnetic interference, vehicles; equipment malfunction, e.g., transmission errors, component failure; operator error; and unknown.

false course — In navigation normally providing one or more course lines, a spurious additional course-line indication due to undesired reflections or maladjusted equipment.

false-echo device — A device for producing an echo different from that normally observed.

false statement — A statement having a value of 0 in Boolean algebra.

falsing — Extraneous signal or signals that cause a decoding device to operate without the normal input of proper encoding signals.

FAMOS—Abbreviation for floating-gate avalanche-injection MOSFET. A type of MOSFET capable of long-term memory storage; used in EPROMs.

fan—1. The volume of space energized periodically by a radar beam(s) as it repeatedly traverses an established pattern. 2. An air-moving machine used to cool electronic circuitry in applications in which natural convection and radiation cooling are insufficient. Employed in electronics enclosures, fans are capable of moving air past the heat-radiation surfaces of the various active components or their associated heat sinks. The terms *fan* and *blower* are sometimes used interchangeably. In practice, however, the fan is a machine with an axially mounted impeller or fan blade, whereas the blower is one with a paddle-wheel-shaped impeller that draws air in at the side and exhausts it at the end. 3. An axial low-pressure device capable of providing high air flow parallel to the axis of the motor shaft. Available in several types, including the tube-axial with integral venturi porting and, in open construction, without venturi porting.

fan antenna—*See* double-V antenna.

fan beam—A field pattern having an elliptically shaped cross section in which the ratio of major to minor axes usually exceeds 3 to 1.

fan-in—1. The number of inputs that can be connected to a logic circuit. 2. The number of operating controls in a single device that individually or in combination result in the same output from the device. 3. Number of units that can be input-connected to a single similar unit.

fan-in circuit—A circuit that has many inputs feeding to a common point.

fan marker—A radio signal having a vertically directed fan beam that tells the pilot the location of his aircraft while flying along a radio range.

fanned-beam antenna—A unidirectional antenna so designed that transverse cross sections of the major lobe are almost elliptical.

fanning beam—A narrow antenna beam that is scanned repeatedly over a limited arc.

fanning strip—An insulated board, often made of wood, that spreads out the wires of a cable for distribution to a terminal board.

fan-out—1. The number of parallel loads within a given logic family that can be driven from one output mode of a logic circuit. 2. To spread a group of conductors, such as a cable end, apart so that each can be individually tested or identified. 3. The number of standard loads that can be driven by a circuit output in a logic family. A standard load is the current required to switch the basic gate of the family. 4. In a digital circuit, the number of other ICs that can be driven by the device. 5. The number of elements that can be operated in parallel from a single similar element. "Similar" in this case refers to operating parameters such as impedance and does not mean identical function devices.

fan-out circuit—A circuit that has a single output point that feeds many branches.

"fantasy" decoder—A type of descrambler used for satellite descrambling. It uses a 94-kHz sine-wave scrambling plus video inversion. Audio is encoded on a 15.7-kHz carrier that is one-sixth that of the scrambling sine wave. The 94-kHz sine wave is not an exact multiple of the scan rate, in general.

farad—The capacitance of a capacitor in which a charge of 1 coulomb produces a change of 1 volt in the potential difference between its terminals. The farad is the unit of capacitance in the mksa system.

faraday—A unit equal to the number of coulombs (96,500) required for an electrochemical reaction involving one electrochemical equivalent. In an electrolytic process, 1 gram equivalent weight of matter is chemically altered at each electrode for 1 faraday of electricity passed through the electrolyte.

Faraday cage—*See* Faraday shield.

Faraday dark space—The relatively nonluminous region between the negative glow and the positive column in a glow-discharge cold-cathode tube. *See also* glow discharge.

Faraday effect—1. The rotation of the plane of polarization of radio waves as they pass though the ionosphere in the earth's magnetic field. This effect produces a decoupling loss between linearly polarized antennas. 2. The rotation of the plane of polarization that occurs when a plane-polarized beam of light passes through certain transparent substances in a direction parallel to the lines of force of a strong magnetic field. This effect also governs the action of a ferrite rotator in a waveguide.

Faraday rotation—1. Rotation of a signal's polarization caused by the atmosphere's E- and F-layers. 2. The apparent rotation of the plane of polarization of a linearly polarized wave as it passes through a medium (e.g., a ferrite material) that has a different propagation constant for each of the two component waves of opposite rotational sense.

Faraday screen—*See* Faraday shield.

Faraday shield—Also called Faraday screen or Faraday cage. 1. A network of parallel wires connected to a common conductor at one end to provide electrostatic shielding without affecting electromagnetic waves. The common conductor is usually grounded. 2. An electrostatic shield between input and output windings of a transformer. This is done to reduce capacitive coupling between the input and output of the power supply.

Faraday's laws—1. The mass of a substance liberated in an electrolytic cell is proportionate to the quantity of electricity passing though the cell. 2. When the same quantity of electricity is passed through different electrolytic cells, the masses of the substances liberated are proportionate to their chemical equivalents. 3. Also called the law of electromagnetic induction. When a magnetic field cuts a conductor, or when a conductor cuts a magnetic field, an electric current will flow through the conductor if a closed path is provided over which the current can circulate.

faradic current—An intermittent and nonsymmetrical alternating current, such as that obtained from the secondary winding of the induction coil.

faradmeter—An instrument for measuring electric capacitance.

far-end crosstalk—Crosstalk that travels along the disturbed circuit in the same direction as the signals in that circuit. To determine the far-end crosstalk between two pairs, 1 and 2, signals are transmitted on pair 1 at station A, and the crosstalk level is measured on pair 2 at station B.

far field—The space beyond the near field of an antenna, in which radiation is essentially confined to a fixed pattern, and power density along the axis of the pattern falls off inversely with the square of the distance.

far-field region—That region of the field of an antenna where the angular field distribution is essentially independent of the distance from the antenna.

far-infrared radiation—That radiation composed of the wavelengths falling between light and microwaves, ranging from 10 to 2000 micrometers. The portion of the infrared spectrum that contains the longest wavelengths.

Farnsworth image-dissector tube—A special cathode-ray tube for use in television cameras.

far-ultraviolet radiation—That radiation characterized by wavelengths ranging from 0.2 to 0.3 micrometer in the electromagnetic spectrum.

far zone—The region distant from a radio transmitting antenna in which the radiation field is stronger than the induction field and in which the intensity of the radiation field varies inversely with the square of the distance from the antenna.

FAST—Acronym for Fairchild advanced Schottky TTL (Fairchild Camera & Instrument Corp.). A family of high-speed TTL logic that uses very small, low-junction-capacitance transistors.

fast-access storage—In a computer memory or storage, the section from which information may be obtained most rapidly.

fast answerback—*See* calculator mode.

fast automatic gain control—A radar AGC method in which the response time is long compared with a pulse width, but short compared with the time on target.

fastener—A device used to secure a conductor (or other object) to the structure that supports it.

fast forward—1. The provision on a tape recorder permitting tape to be run rapidly through the recorder in the play direction, usually for search or selection purposes. 2. High-speed winding (shuttling) of tape from the supply reel onto the takeup reel.

fast-forward control—A tape-recorder control that permits running the tape through the machine rapidly in the forward direction.

fast Fourier transform—Abbreviated FFT. A computationally efficient mathematical technique that converts digital information from the time domain to the frequency domain for rapid spectral analysis. FFTs generally utilize a time-weighting function to compensate for data records with a noninteger number of samples.

fast groove—Also called fast spiral. In disc recording, an unmodulated spiral groove having a much greater pitch than the recorded grooves.

fast-operate, fast-release relay—A high-speed relay designed specifically for both short operate and short release times.

fast-operate relay—A high-speed relay designed specifically for short operate and long release times.

fast-operate, slow-release relay—A relay designed specifically for short operate and long release times.

fast-release relay—A high-speed relay designed specifically for short release but not short operate time.

fast spiral—*See* fast groove.

fast time constant—An antijamming device used in radar video-amplifier circuits. It differentiates incoming pulses so that only the leading edges of the pulses are used.

fatal error—An error that causes a program to abort.

fatigue—The weakening of a material under repeated stress.

fault—1. A defect in a wire circuit due to unintentional grounding, a break in the line, or a crossing or shorting of the wires. 2. A disturbance (such as lightning) that impairs normal operation. 3. A failure of a hardware or software component in a system that may lead to a system failure or error or some other manifestation that can be detected by a user. 4. An anomaly that prevents the correct operation of the device.

fault current—1. The current that may flow in any part of a circuit or amplifier under specified abnormal conditions. 2. The current in a circuit that results from loss of insulation between conductors or between a conductor and ground.

fault dictionary—A set of fault signatures, each of which indicates the probable faults that could cause the error message matching the signature.

fault electrode current—The peak current that flows through an electrode during a fault, such as an arcback or a load short circuit.

fault finder—A test set for locating troubles in a telephone system.

fault indicator—Equipment that provides an instantaneous alarm, both visual and audible, of failures detected in the various components of its assorted equipment.

fault isolation—1. The process of identifying and locating failures in a unit under test. 2. Determining the cause of a test failure, typically by identifying a defective component or process failure on a board.

fault-isolation resolution—The average number of components to which a fault can be isolated.

fault model—A set of data that logically describes the operations of a device or circuit containing one or more faults.

fault signature—1. Data representing the outputs of a known good unit and used to compare against outputs of a unit under test. 2. A particular output response or set of responses generated when a test program is executed on a device containing a fault. A typical fault signature consists of the incorrect output pin numbers and the test step number at which a test program first detects a fault.

fault simulation—The process of injecting faults during a simulation run to determine the test comprehensiveness of an input pattern.

fault-tolerant—Refers to a computer program or system in which some parts may fail and the program or system will still execute properly.

fault-tolerant circuits—Circuits that are designed so that they could continue to function properly even though part of the circuit failed.

Faure plate—A storage-battery plate consisting of a conductive lead grid filled with active paste material.

fax or FAX—1. Abbreviation for facsimile. 2. A scanner/printer combination that transmits text and graphics over telephone lines. It uses CCITT Group 3 data compression techniques. Small paper documents can be transmitted over long distances very quickly, but the information is not represented as structured data elements, as in EDI.

fax modem—A device connected to a personal computer, giving it the ability to send or receive electronic messages or images over telephone lines.

fc—Letter symbol for footcandle.

FCC—Abbreviation for Federal Communications Commission.

F-connector—A small, metallic, male-type connecting device with internal threads that attach to the end of a coaxial cable to secure and electrically connect the coax to a female F-fitting. The internal threads of the male connector screw onto the external threads of the female connector.

F/D—*See* focal distance-to-diameter ratio.

FDDI—*See* fiber distributed data interface.

F-display—Also called F-scan or F-scope. In radar, a rectangular display in which a target appears as a centralized blip when the radar antenna is aimed at it. Horizontal and vertical aiming errors are indicated by the horizontal and vertical displacement of the blip.

FDM—*See* frequency-division multiplexer and frequency-division multiplexing.

FDMA—Abbreviation for frequency-division multiple access. A method of sharing the capacity of a satellite transponder by assigning each earth terminal a portion of the transponder's bandwidth, into which the terminal places a carrier modulated by its information. The carrier can be received by all other earth terminals.

FDX—*See* full duplex.

feasibility study—An investigation of the advantages and disadvantages of using an alternative approach over the presently used approach.

FEC—Abbreviation for forward error correction. 1. Used to describe equipment that corrects transmission errors at a receiver. The technique provides for transmission of additional information with the original bit stream so that if an error is detected, the receiver can recreate the correct information without a retransmission. 2. A system of data transmission in which redundant bits generated at the transmitter are used at the receiving terminal to detect, locate, and correct any transmission errors before delivery to the data sink. The advantage of this system is that it does not require a feedback channel; therefore, it can be used with a one-way transmission system.

Federal Communications Commission—Abbreviated FCC. A U.S. federal government agency made up of a board of seven commissioners appointed by the president under the Communications Act of 1934, having the power to regulate all interstate and foreign radio communication originating in the United States, including radio, television, facsimile, telegraph, telephone, and cable systems.

federal telecommunications system—System of commercial telephone lines, leased by the government, for use between major government installations for official telecommunications.

feedback—1. In a transmission system or a section of it, the returning of a fraction of the output to the input. 2. In a magnetic amplifier, a circuit connection by which an additional magnetomotive force (which is a function of the output quantity) is used to influence the operating condition. 3. In a control system, the signal or signals returned from a controlled process to denote its response to the command signal. Feedback is derived from a comparison of actual response to desired response, and any variation is used as an error signal combined with the original control signal to help attain proper system operation. Systems employing feedback are termed closed-loop systems; feedback closes the loop. 4. Squeal or howl from a speaker caused by speaker sound entering microphone of same recorder or amplifier. 5. The return of a portion of the output of a circuit or device to its input. With positive feedback, the signal fed back is in phase with the input and increases amplification, but may cause oscillation. With negative feedback, the signal is 180° out of phase with the input and decreases amplification but stabilizes circuit performance and tends to lower an amplifier's output impedance, improve signal stability, and minimize noise and distortion. 6. The process of coupling some of the output of an amplifier back to its input. Negative feedback reduces the gain of an amplifier, but has compensating beneficial results. Positive feedback can be used to boost gain (regeneration), but usually results in oscillation. 7. The flow of information back into the control system so that actual performance can be compared with planned performance. 8. The transfer of a portion of energy from one point in an electrical system to a preceding point. The transfer may be either electrical or acoustical.

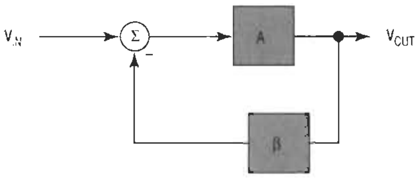

$$\frac{V_{OUT}}{V_{IN}} = \frac{A}{1+A\beta}$$

Feedback.

feedback admittance—In an electron tube, the short-circuit transadmittance from the output electrode to the input electrode.

feedback amplifier—An amplifier that uses a passive network to return a portion of the output signal to modify the performance of the amplifier.

feedback attenuation—In the feedback loop of an operational amplifier, an attenuation factor by which the output voltage is attenuated to produce the input error voltage.

feedback circuit—A circuit that permits feedback in an electronic device.

feedback compensation—The placement of a device, or an additional circuit, into a feedback control system to improve its response in relation to a specific characteristic of a system.

feedback control—1. A type of system control obtained when a portion of the output signal is operated upon and fed back to the input in order to obtain a desired effect. 2. An automatic means of sensing speed variations and correcting to maintain a constant speed or close speed regulation. 3. Guidance technique used by robots to bring the end effector to a programmed point.

feedback control loop—A closed transmission path that includes an active transducer and consists of a forward path, a feedback path, and one or more mixing points arranged to maintain a prescribed relationship between the loop input and output signals.

feedback control signal—That portion of the output signal that is returned to the input in order to achieve a desired effect, such as fast response.

feedback control system—1. A control system comprising one or more feedback control loops; it combines the functions of the controlled signals and commands, tending to maintain a prescribed relationship between the two. 2. A system designed to control the output quantity of a device by returning a portion of its output signal back to its input. This results in the manipulation of the input quantity so that the desired relationship between the input and output signals can be maintained.

feedback cutter—An electromechanical transducer that performs like a disc cutter except it is equipped with an auxiliary feedback coil in the magnetic field. Signals exciting the cutter are induced into the feedback coil, the output of which is fed back in turn to the input of the cutter amplifier. The result is a substantially uniform frequency response.

feedback diode—*See* freewheeling diode.

feedback loop—The components and processes involved in using part of the output as an input for correction or control of the operation of a system.

feedback oscillator—An oscillating circuit, including an amplifier, in which the output is coupled in phase with the input. The oscillation is maintained at the frequency determined by the parameters of the amplifier and the feedback circuits, such as *LC*, *RC*, and other frequency-selective elements.

feedback path—In a feedback control loop, the transmission path from the loop output signal to the loop feedback signal.

feedback regulator—A feedback control system that tends to maintain a prescribed relationship between certain system signals and other predetermined quantities.

feedback sense voltage—The voltage, referred to ground, on the feedback terminal of a regulator while it is operating in regulation.

feedback transfer function—In a feedback control loop, the transfer function of the feedback path.

feedback winding—In a saturable reactor, the control winding to which a feedback connection is made.

feeder—1. A conductor or group of conductors connecting two generating stations, two substations, a generating station and a substation or feeding point, a substation and a feeding point, or a transmitter and antenna. 2. A set of conductors between a main distribution center and secondary ones. A radial feeder supplies energy from one source to a substation or a feeding point that does not receive energy from any other source. The normal flow of energy in such a feeder is in one direction only. 3. The circuit conductors between the service equipment and the branch circuit overcurrent device. 4. A transmission line between an antenna and a radio transmitter or receiver.

feeder cable—Also called trunk cable. 1. A communication cable extending from the central office along a primary route (main feeder cable), or from a main feeder cable along a secondary route (branch feeder cable), and providing connections to one or more distribution cables. 2. In a CATV system, the transmission cable from the head end (signal pickup) to the trunk amplifier.

feeder link—A radio link between an earth station and a satellite, conveying information for a space radiocommunication service other than a fixed-satellite service. In the broadcasting-satellite service, all feeder links are uplinks (from the earth to the satellite), but in the mobile-satellite service, feeder links can be both uplinks and downlinks.

feed-forward—A frequency-compensation technique in operational amplifiers. A small-value capacitor is used to bypass a gain stage that has poor performance at high frequency.

feed function—In automatic control of machine tools, the relative motion between the work and the cutting tool (excluding the motion provided for removal of material).

feed holes—A series of small holes in perforated paper tape that convey no information, but are solely for the purpose of engaging the feed pawls or sprocket that transports the tape over the sensing pins of various reading devices.

feedhorn—1. A satellite TV receiving antenna component mounted at the focal point of a parabolic dish that collects the signal reflected from the main surface reflector and channels this signal into the low-noise amplifier (LNA). 2. A device that collects microwave signals reflected from the surface of an antenna. It is mounted at the focus of all prime focus parabolic antennas.

feed pitch—The distance between the centers of adjacent feed holes in a tape.

feed reel—Also called stock, supply, or storage reel. On a tape recorder, the reel from which the tape unwinds while playing or recording.

feedthrough—1. The accidental or unintentional transfer of a signal from one track to another on a multitrack tape. 2. The use of special connectors to pass conductors through bulkheads or panels. Contacts can be male on one side and female on the other, or they can be male on both sides or female on both sides. Feedthrough connectors differ from rack-and-panel types in that connection can be made on both sides of the panel. 3. A conductor through the thickness of a substrate, thereby electrically connecting both surfaces. 4. The change in the output voltage of a sample/hold circuit in the hold mode caused by a voltage change in the input. Measured in decibels of attenuation.

feedthrough capacitor—1. A feedthrough insulator that provides a desired value of capacitance between the feedthrough conductor and the metal chassis or panel through which the conductor is passing. Used chiefly for bypass purposes in UHF. 2. A coaxial capacitor with a central current-carrying conductor, or a conductor connected with a current-carrying rod surrounded by a capacitor element, that is symmetrically bonded to the center conductor and to the outer casing to form a coaxial construction.

feedthrough insulator—A type of insulator that permits wire or cable to be fed through walls, etc., with minimum current leakage.

feed-thru connection—*See* thru-hole connection.

female—Pertaining to the recessed portion of a device into which another part fits.

female contact—A contact located in an insert or body in such a manner that the mating contact is inserted in the unit. This is similar in function to a socket contact.

femto-—Prefix meaning 10^{-15}. Letter symbol: f.

femtoampere—A unit of current equal to 10^{-15} ampere. Letter symbol: fA.

femtovolt—A unit of voltage equal to 10^{-15} volt. Letter symbol: fV.

fence—1. A line or system of early-warning radar stations. 2. A concentric steel fence placed around a ground radar transmitting antenna to act as an artificial horizon and suppress ground clutter that would otherwise mask weak signals returned from a target at a low angle.

fence alarm—Any of several types of sensors used to detect the presence of an intruder near a fence or any attempt by an intruder to climb over, go under, or cut through the fence.

Fermi-Dirac distribution—A mathematical description of the way in which the current carriers present in a crystalline material have energies distributed above and below the Fermi level; this distribution is a function of the excitation of the material.

Fermi level—The value of electron energy at which the Fermi distribution function is one-half. The average carrier level of a semiconductor region. Hence, a semiconductor crystal in equilibrium has a constant Fermi level throughout.

ferpic—*Fer*roelectric ceramic *pic*ture device. A sandwichlike structure made up of transparent electrodes, a photoconductive film, and a thin plate of fine-grained ferroelectric ceramic. The device stores images in the form of a variation of the birefringence of the ceramic plane (that is, as a variation in the way the plate transmits polarized light).

ferreed—An electromechanical switch that combines the rapid switching of bistable magnetic material with metallic contacts to produce output indications that persist as long as desired without further application of power. Describes any relay whose contact assembly of individual magnetic reeds is sealed in an evacuated glass tube, operated by an external winding. When operated, it is held by magnetism until released by a current pulse.

ferret—An aircraft, ship, or vehicle especially equipped for ferret reconnaissance.

ferret reconnaissance—A form of reconnaissance that detects, locates, and analyzes enemy radars. It is a passive technique that listens for signals transmitted by enemy radars and thus cannot be jammed. The maximum effective range is limited only by the radio horizon. Ferret systems flown at altitudes of a few miles can accurately locate and analyze radars 300 miles (483 km) away. Ferret is an all-weather reconnaissance technique, and no camouflage system works against a properly designated ferret.

ferri- — Prefix indicating a material having a net dipole moment.

ferric oxide (Fe_2O_3) — A red, iron oxide coating for magnetic recording tapes.

ferrimagnetic amplifier — A microwave amplifier utilizing ferrite material in the coupling inductors and transformers.

ferrimagnetic limiter — A power limiter used to replace tr tubes in microwave systems. Its operation is based on a ferrimagnetic material, such as a piece of ferrite or garnet, that exhibits nonlinear properties.

ferrimagnetic materials — Those materials in which spontaneous magnetic polarization occurs in nonequivalent sublattices; the polarization in one sublattice is aligned antiparallel to the other.

ferrimagnetism — A type of magnetism that, because the magnetic moment of neighboring ions tends to align antiparallel, appears microscopically similar to antiferromagnetism. The fact that these moments may be of different magnitudes allows a large resultant magnetization that macroscopically resembles ferromagnetism.

ferristor — A two-winding ferroresonant magnetic amplifier that operates on a high carrier frequency. A two-winding saturable reactor that may be connected as a coincidence gate, current discriminator, free-running multivibrator, oscillator, or ring counter.

ferrite — 1. Also called ferrospinel. A powdered, compressed, and sintered magnetic material having high resistivity, consisting chiefly of ferric oxide combined with one or more other metals. The high resistance makes eddy-current losses extremely low at high frequencies. Examples of ferrite compositions include nickel ferrite, nickel-cobalt ferrite, manganese-magnesium ferrite, yttrium-ion garnet, and single-crystal yttrium-ion garnet. 2. The generic term for a class of nonmetallic, ceramic, ferromagnetic materials (MFe_2O_4) having a spinel-crystal structure. The materials are noted for their high resistivity relative to ferromagnetic metals. Although in its simplest form the M in the formula is iron (Fe), iron is seldom used in actual ferrite manufacture. Manganese, nickel, zinc, magnesium, or copper are more common, and MnZn, NiZn, and MgZn are the most popular combinations.

ferrite bead — A magnetic device for storage of information. It is made of ferrite powder mixtures in the form of a bead fired on the current-carrying wires of a memory matrix.

ferrite circulator — A nonreciprocal microwave network that transmits power from one terminal to another in sequence. Can replace a conventional duplexer, provide isolation of transmitter from receiver, eliminate the requirement for an atr, and isolate the transmitter from antenna reflections.

ferrite core — A core made from iron and other oxides and usually shaped like a doughnut. It is used in circuits and magnetic memories and can be magnetized and demagnetized very rapidly.

ferrite-core memory — A magnetic memory in which read-in and read-out wires are threaded through a matrix of very small toroidal cores molded from a square-loop ferrite.

ferrite isolator — A device either in a waveguide or coax that allows power to pass through in one direction with very little loss, while the rf power in the reverse direction is absorbed. It is useful for maintaining signal source stability and eliminating long-line and frequency-pulling effects in all types of low-power microwave signal sources.

ferrite limiter — A passive low-power microwave limiter that provides an insertion loss of less than 1 dB, with minimum phase distortion, when operating in its linear range. It is used for protecting sensitive receivers from burnout and from blocking by a strong interfering signal.

Ferrite cores.

ferrite phase-differential circulator — A combination microwave duplexer and load isolator that serves as a switching device between a radar antenna and the associated high-power radar magnetron and radar receiver.

ferrite-rod antenna — Also called ferrod or loopstick antenna. An antenna used in place of a loop antenna in a radio receiver. It consists of a coil wound around a ferrite rod.

ferrite rotator — A gyrator, composed of a ferrite cylinder surrounded by a ring-type permanent magnet, that is inserted in a waveguide to rotate the plane of polarization of electromagnetic waves that travel through the waveguide.

ferrites — 1. Chemical compounds of iron oxide and other metallic oxides combined with ceramic material. They have ferromagnetic properties but are poor conductors of electricity. Hence, they are useful where ordinary ferromagnetic materials (which are good electrical conductors) would cause too great a loss of electrical energy. 2. Ceramic structures made by mixing iron oxide (Fe_2O_3) with oxides, hydroxides, or carbonates of one or more of the divalent metals, such as zinc, nickel, manganese copper, cobalt, magnesium, cadmium, or iron.

ferrite switch — A ferrite device that obstructs the flow of energy through a wavelength by causing a 90° rotation of the electric field vector.

ferroacoustic storage — A delay-line type of storage comprising a thin tube of magnetostrictive material, a central conductor that passes through the tube, and an ultrasonic driver transducer at one end of the tube.

ferrod — *See* ferrite-rod antenna.

ferrodynamic instrument — An electrodynamic instrument in which the measuring forces are materially increased by the presence of ferromagnetic material.

ferroelectric — 1. Pertaining to a phenomenon exhibited by certain materials in which the material is polarized in one direction or the other or reversed in direction by the application of a positive or negative electric field of magnitude greater than a certain amount. The material retains the electric polarization unless it is disturbed. The polarization can be sensed by the fact that a change in the field induces an electromotive force that can cause a current. 2. That property of certain materials that determines that they will be polarized in one direction or the other, or reversed in direction of polarization when a positive or negative electric field is applied. The material will remain ferroelectric until it is disturbed.

ferroelectric converter—A device that generates high voltage when heat is applied to it. Its operation is based on the change in the dielectric constant or the permittivity of certain materials, such as barium titanate, when heated. This change reaches maximum at the Curie point.

ferroelectric crystal—A crystal that can be polarized in the opposite direction by applying an electric field weaker than the breakdown strength of the material.

ferroelectric domain—The region of a ferroelectric crystal where spontaneous polarization is uniformly directed.

ferroelectric film—A film in which electric polarization is reversible when influenced by an electric field.

ferroelectricity—A property of certain crystalline materials whereby they exhibit a permanent, spontaneous electric polarization (dipole moment) that is reversible by means of an electric field; the electric analog of ferromagnetism. Materials that show this effect are piezoelectric as well.

ferroelectric materials—Those materials in which the electric polarization is produced by cooperative action between groups or domains of collectively oriented molecules.

ferroelectrics—Pyroelectric materials whose direction of polarization can be reversed by application of an electric field.

ferromagnetic—Pertaining to a phenomenon exhibited by certain materials in which the material is polarized in one direction or the other or reversed in direction by the application of a positive or negative magnetic field of magnitude greater than a certain amount. The material retains the magnetic polarization unless it is disturbed. The polarization can be sensed by the fact that a change in the field induces an electromotive force that can cause a current.

ferromagnetic amplifier—A parametric amplifier based on the nonlinear behavior of ferromagnetic resonance at high rf power levels. In one version, microwave pumping power is supplied to a garnet or other ferromagnetic crystal mounted in a cavity containing a strip line. A permanent magnet provides sufficient field strength to produce gyromagnetic resonance in the garnet at the pumping frequency. The input signal is applied to the crystal through the strip line, and the amplified output signal is extracted from the other end of the strip line. Sometimes incorrectly called a garnet maser, but the operating principle differs from that of the maser.

ferromagnetic material—1. A material having a specific permeability greater than unity, the amount depending on the magnetizing force. A ferromagnetic material usually has relatively high values of specific permeability, and it exhibits hysteresis. The principal ferromagnetic materials are iron, nickel, cobalt, and certain of their alloys. 2. A paramagnetic material that exhibits a high degree of magnetizability. 3. A material that exhibits hysteresis phenomena and whose permeability, greater than 1, is dependent on the magnetizing field.

ferromagnetic oxide parts—Parts consisting primarily of oxides that display ferromagnetic properties.

ferromagnetic resonance—A condition under which the apparent permeability of a magnetic material reaches a sharp maximum at a microwave frequency.

ferromagnetics—The science that deals with the storage of information and the control of pulse sequences through use of the magnetic polarization properties of materials.

ferromagnetic tape—Tape made of magnetic material and used for winding closed cores for toroids and transformers.

ferromagnetism—1. A high degree of magnetism in ferrites and similar compounds. The magnetic moments of neighboring ions tend not to align parallel with each other. The moments are of different magnitudes, and the resultant magnetization can be large. 2. Strong magnetic property of such substances as iron, cobalt, nickel, and certain alloys. Ferromagnetic substances are essential to the construction of such pieces of equipment as speakers, transformers, electric generators, etc. 3. A property of certain metals, alloys, and compounds whereby below a certain critical temperature (the Curie point) the magnetic moments of the atoms tend to align, giving rise to a spontaneous, permanent magnetism (dipole moment) that is reversible by means of a magnetic field. 4. The properties of certain materials that cause them to have relative permeabilities that exceed unity. This permeability permits the materials to exhibit hysteresis.

ferromanganese—An alloy of iron and manganese.

ferrometer—An instrument for making permeability and hysteresis tests of iron and steel.

ferroresonance—Resonance associated with circuits in which at least one of the circuit elements is nonlinear and contains iron.

ferroresonant circuit—A resonant circuit in which one of its elements is a saturable reactor.

ferroresonant transformer—1. A voltage-regulating device that gives regulated ac voltages and incorporates some special advantages. Capable of acting as a step-up or step-down voltage transformer as well as an ac voltage regulator, this component delivers a more or less constant ac output voltage even if the magnitude of the input voltage changes. In addition, the ferroresonant transformer is efficient, inexpensive, rugged, and requires no heat sink. It generates no high levels of electrical noise and provides a degree of protection from transients riding on the ac power line. 2. The principal elements of a constant-voltage power supply. The transformer is specially wound to be tuned to line frequency. The resonant condition established is the key to its operation. Only a simple filter is needed to form a highly reliable and very efficient power supply. It does not require a dissipative regulator. A typical ferroresonant power supply offers line regulation of 2 percent and load regulation of 5 percent; efficiency may reach 80 percent. However, the ferroresonant transformer is considerably larger than the transformer for a linear supply and may be twice as heavy.

ferrospinel—A ceramic-like material containing iron and other elements combined with oxygen. A poor conductor of electricity, it is used in transformers, antenna, loops and television deflecting yokes. *See also* ferrite.

ferrous—Composed of and/or containing iron. A ferrous metal exhibits magnetic characteristics, as opposed to a nonferrous metal, such as aluminum, which does not.

ferrous oxide—The substance on magnetic recording disks and tapes. It can be magnetized, thereby permitting information to be recorded on it magnetically.

ferrule—1. A short tube used to make solderless connections to shielded or coaxial cable. Also molded into the plastic inserts of multiple-contact connectors to provide strong, wear-resistant shoulders on which contact retaining springs can bear. 2. The metal cap around the end of a cartridge-fuse tube that serves as a contact for the fuse.

ferrule resistor—A resistor having ferrule terminals for mounting in standard fuse clips.

FET—Abbreviation for field-effect transistor. Also called unipolar transistor. 1. A transistor controlled by voltage rather than current. The flow of working current through a semiconductor channel is switched and regulated by the effect of an electric charge in a region close to the channel called the gate. A FET has either p-channel or n-channel construction. 2. A transistor whose internal operation is unipolar in nature. The metal-oxide semiconductor FET (MOSFET) is widely used in integrated circuits because the devices are very small and can be manufactured with few steps. 3. A solid-state device in which current is controlled between source terminal and drain terminal by voltage applied to a nonconducting gate terminal. 4. Semiconductor device in which resistance between source and drain terminals is modulated by a field applied to the third (gate) terminal.

fetal cardiotachometer—*See* fetal monitor.

fetal electrocardiograph—*See* fetal monitor.

fetal monitor—An instrument that displays or records the fetal electrocardiogram or other indication of heart action. In some instruments, the maternal electrocardiogram is recorded simultaneously. The instrument may be referred to more definitely as a fetal electrocardiograph, fetal cardiotachometer, or fetal phonocardiograph, depending on its primary purpose.

fetal phonocardiograph—An instrument that provides continuous instantaneous recording of beat-to-beat changes in the fetal heart rate. *See also* fetal monitor.

fetch—1. To go after and return with things. In a microprocessor, the "objects" fetched are instructions that are entered in the instruction register. The next, or a later, step in the program will cause the machine to execute what it was programmed to do with the fetched instructions. Often referred to as an instruction fetch. 2. In a computer, the collective actions of acquiring a memory address and then an instruction or data byte from memory. 3. Reading out an instruction at a particular memory location into the CPU. 4. To obtain a quantity of data from a place of storage. 5. The action of obtaining an instruction from a stored program and decoding that instruction. Also refers to that portion of a computer's instruction cycle during which that action is performed.

fetch cycle—The first cycle in the fetch-decode-execute sequence of instruction execution. During the fetch cycle, the contents of the program counter are placed on the address bus, a read signal is generated, and the program counter is incremented. The data word coming from the memory (the instruction that has been fetched) will then be gated into the instruction register of the control unit.

FET resistor—A field-effect transistor in which, generally, the gate is tied to the drain, and the resultant structure is used in place of a resistor load for a transistor.

"f" factor—The slope of the straight line from which the nonlinearity of a displacement transducer is calculated in microvolts output per volt excitation per unit stimulus.

FFC—Abbreviation for flat flexible cable.

FFT—Abbreviation for fast Fourier transform.

fiber—1. A tough insulating material, generally of paper and cellulose, compressed into rods, sheets, or tubes. 2. A clear glass or plastic optical cable, consisting of a core and cladding, designed to propagate optical energy. The diameter of a fiber can vary from about 10 to 1000 micrometers, depending on type. 3. Glass, silica, or plastic cable by which light is conducted or transmitted. Can be multimode (capable of propagating more than one mode of given wavelength) or single mode (one that supports propagation of only one mode of given wavelength). 4. A thread or threadlike structure such as composes cellulose, asbestos, or glass yarn. Also, a single discrete element used to transmit optical (light-wave) information. Analogous to a single wire used to transmit electrical information. Usually consists of a core that transmits the information and a cladding around the core. 5. The material path along which light propagates; a single discrete optical transmission element.

fiber bundle—1. A consolidated group of single fibers used to transmit a single optical signal. 2. A rigid or flexible concentrated assembly of glass or plastic fibers used to transmit optical images or light. *See* coherent fiber bundle. 3. An assemblage of transparent glass fibers all bundled together parallel to one another. The length of each fiber is much greater than its diameter. By a process of total reflection, this bundle of fibers can transmit a picture from one of its surfaces to the other around curves and into otherwise inaccessible places with an extremely low loss of definition and light.

fiber cable—A cable composed of a fiber bundle or a single fiber, strength members, and a cable jacket.

fiber dispersion—In fiber optics, pulse spreading in a fiber caused by differing transit times of various modes of electromagnetic waves.

fiber distributed data interface—Abbreviated FDDI. A standard for transmitting data on optical fiber cables at a rate of around 100,000,000 bits per second (10 times as fast as Ethernet and about twice as fast as T-3).

fiber metallurgy—The growing of superfine crystal whiskers whose characteristic is relatively great strength in their length-to-diameter ratio.

fiber needle—A playback point or phonograph needle made from fiber. Being softer than a metal or diamond needle, it is less scratchy; however, it has an extremely short life.

fiber-optic field flattener—A plate consisting of fused optical fibers with both surfaces ground and polished, and having the entrance surface curved to match the image curvature of the input system. The plate transmits to the flat exit surface.

fiber-optic rod multiplexer-filter—A graded-index, cylindrically shaped section of optical fiber or rod with a length corresponding to the pitch of the undulations of light waves caused by the graded refractive index, the light beam being injected via fibers at an off-axis end point on the radius, with the undulations of the resulting wave varying periodically from one point to another along the rod with interference layers at the $\frac{1}{4}$-pitch point of the undulations, providing for multiplexing or filtering.

fiber optics—Abbreviated FO. 1. A technology that uses light as a digital information carrier. The transmission medium is made up of small strands of glass, each of which provides a path for light rays that carry the data signal. Fiber-optic technology offers large bandwidth, very high security, and immunity to electrical interference. The glass-based transmission facilities also occupy far less space than other high-bandwidth media, which is a major advantage in crowded underground ducts. 2. Also called optical fibers or optical fiber bundles. An assemblage of transparent glass fibers all bundled together parallel to one another. The length of each fiber is much greater than its diameter. This bundle of fibers has the ability to transmit a picture from one of its surfaces to the other around curves and into otherwise inaccessible places with an extremely low loss of definition and light, by a process of total reflection. 3. The technique of conveying information in the form of light signals through a particular configuration of glass or plastic fibers. 4. A general term describing a light-wave or optical communications system. In such a system, electrical information is converted to light energy, transmitted to another location through optical fibers, and is there converted back into electrical information. 5. A method of communicating analog or digital signals through a noninductive/nonconductive dielectric medium such as glass- or plastic-core cables. The fiber-optic system consists of a light source and

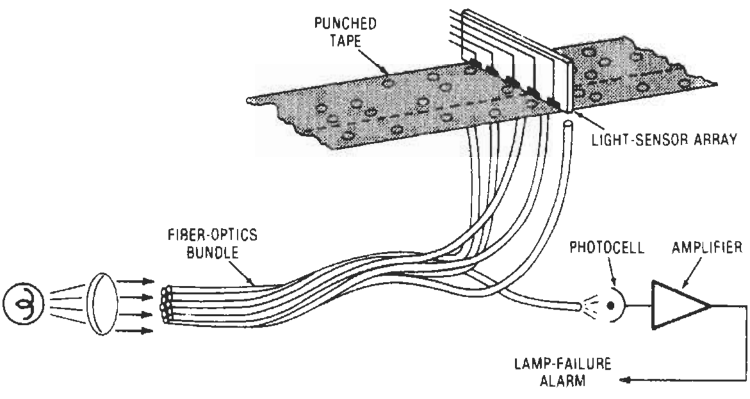

Fiber optics.

light detector, cable-to-semiconductor connectors, and the fiber-optic cable. The transmitter converts the input signal to an analog- or digital-modulated light signal that is communicated through the fiber-optic cable to a light detector. The light detector converts the optical energy back into usable electric form. 6. The technology of guidance of optical power, including rays and waveguide modes of electromagnetic waves, along conductors of electromagnetic waves in the visible and near-visible region of the frequency spectrum, specifically when the optical energy is guided to another location through thin transparent strands. Techniques include conveying light or images through a particular configuration of glass or plastic fibers. Incoherent optical fibers will transmit light, as a pipe will transmit water, but not an image. Coherent optical fibers can transmit an image through perfectly aligned, small (10–12 micrometers) clad optical fibers. Specialty fiber optics combine coherent and incoherent aspects.

fiber-optics bundles — 1. Assemblies of optical fibers. 2. Very fine transparent glass or plastic threads, each of which transmits light.

fiber-optic scanner — A scanner in which a fiber-optic assembly replaces a lens system.

fiber-optics computer interconnection — A means of connecting a computer with a terminal or another computer to transmit electrical signals via fiber-optics cable instead of wire.

fiber-optics multiport coupler — An optical unit, such as a scattering or diffusion solid "chamber" of optical material, that has at least one input and two outputs, or at least two inputs and one output, that can be used to couple various sources to various receivers. The ports are usually optical fibers. If there is only one input and one output port, it is simply a connector.

fiber-optics probe — A flexible probe made up of a bundle of fine glass fibers optically aligned to transmit an image, light, or both.

fiber-optics scrambler — 1. Similar to a fiberscope except that the middle section of loose fiber is deliberately disoriented as much as possible, then potted and sawed. Each half is then capable of coding a picture, which can be decoded by the other half. 2. A device used for coding messages that has a fiber bundle that is aligned at both ends and scrambled in the middle. The resulting halves of the bundles serve as encoders or decoders and the message appears as a random scattering of black dots.

fiber-optics splice — A nonseparable junction joining one optical conduction to another.

fiber-optics system — A light source and light detector; transmitter and receiver; and fiber-optic cable with connectors. The transmitter converts the encoded input signal to dc current that is converted to light energy by the light-emitting diode. This light energy is coupled from the LED into the fiber-optic cable with connectors and is then transmitted through the cable, which is connected to a light detector such as a photodiode (PIN diodes), phototransistor, or Darlington phototransistor. These devices detect the light energy and convert it to a current that may be amplified and decoded to faithfully reproduce the original input signal.

fiber-optics terminus — A device used to terminate an optical conductor that provides a means to locate and contain an optical conductor within a connector.

fiber-optics transmission system — Abbreviated FOTS. A transmission system utilizing small-diameter transparent fibers through which light is transmitted. Information is transferred by modulating the transmitted light. These modulated signals are detected by light-sensitive devices, such as photodetectors.

fiber-optics waveguide — A relatively long, thin strand of transparent substance, usually glass, capable of conducting an electromagnetic wave of optical wavelength (visible or near-visible region of the frequency spectrum) with some ability to confine longitudinally directed, or near-longitudinally directed, light waves to its interior by means of internal reflection.

fiber scattering — In an optical fiber, the coupling, or leaking, of light-wave power out of the core of the fiber by Rayleigh scattering or guide imperfections such as dielectric strain, compositional or physical discontinuities in the core or cladding, irregularities and extraneous inclusions in the core-cladding interface, curvature of the optical axis, or tapering. Scattering losses are measured in all directions as an integrated effect and expressed in decibels per kilometer.

fiberscope — Optical glass fibers, when systematically arranged in a bundle, transmit a full-color image that remains undisturbed when the bundle is bent. By mounting an objective lens on one end of the bundle, and an eyepiece at the other, the assembly becomes a flexible fiberscope that can be used to view objects that would be inaccessible for direct viewing.

fibrillation — A result of loss of synchronization of the heart muscle, causing the individual muscle fibers to contract in a random, uncoordinated sequence. As a result, the heart muscle merely quivers instead of pumping forcefully.

fidelity — 1. The accuracy with which a system or portion of a system reproduces at its output the essential characteristics of the signal impressed on its input. 2. A measure of the exactness with which sound is duplicated or reproduced.

field—1. One of the two equal parts into which a frame of a television image is divided in interlaced scanning. With present U.S. standards, pictures are transmitted in two fields of 262.5 lines each, which are interlaced to form 30 complete frames, or images, per second. 2. That area or space in which a particular geophysical effect, such as gravity or magnetism, occurs and can be measured. 3. A group of characters in a computer that is treated as a single unit of information. 4. A region near an electric charge, a source of electromagnetic radiation, or a magnet in which components or materials may be affected. 5. That silicon area on a chip not used or occupied by active transistors. 6. A group of adjacent bits in a microinstruction. 7. A portion of a microprogram word that represents a group of bits dedicated to controlling a specific piece of hardware. 8. A subdivision of a record, usually consisting of a single item of information related to the rest of the record and serving a similar function in all records of that group. The smallest unit normally manipulated by a database management system.

field application relay—A device that automatically controls the application of the field excitation to an ac motor at some predetermined point in the slip cycle.

field circuit breaker—A device that functions to apply, or to remove, the field excitation of a machine.

field control of speed—The varying voltage applied to the field of a shunt-wound motor to control the motor's speed over the extended range.

field density—*See* magnetic induction.

field-discharge protection—A control function or device to limit the induced voltage in the field when the field current attempts to change suddenly.

field distortion—Distortion between the north and south poles of a generator due to the counterelectromotive force in the armature winding.

field-effect tetrode—A semiconductor device consisting basically of a thin n region adjacent to a similarly thin p region. Two contacts are made to the n side and two to the p side so that currents can be passed through each thin region parallel to the single junction. The two currents remain separate because reverse bias is maintained on the junction. A current in either side affects the resistance of the other side and hence the current in the other side.

field-effect transistor—Abbreviated FET. 1. A transistor in which current carriers (holes or electrons) are injected at one terminal (the source) and pass to another (the drain) through a channel of semiconductor material whose resistivity depends mainly on the extent to which it is penetrated by a depletion region. The depletion region is produced by surrounding the channel with semiconductor material of the opposite conductivity and reverse-biasing the resulting pn junction from a control terminal (the gate). The depth of the depletion region depends on the magnitude of the reverse bias. Because the reverse-biased junction draws negligible current, the characteristics of the device are similar to those of a vacuum tube. 2. A transistor in which the current through a conducting channel is controlled by an electric field arising from a voltage applied between the gate and source terminals.

field-effect tube—A triode with its grid replaced by a nonintercepting control gate. A high positive voltage is applied to this gate in order to draw sufficient current from the cathode. The result is a strong concentration of the electric field at the gap between the gate and the cathode, producing an electron beam passing through the gate to the anode.

field-effect varistor—A passive, nonlinear, two-terminal semiconductor device that maintains a constant current over a wide range of voltage.

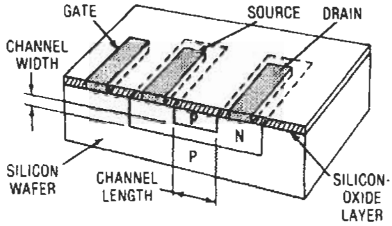

Field-effect transistor.

field emission—Also called cold emission. The liberation of electrons from a solid or liquid by application of a strong electric field at the surface.

field-enhanced photoelectric emission—Increased photoelectric emission resulting from the action of a strong electric field on the emitter.

field-enhanced secondary emission—Incre